Single Image Super Resolution for Spatial Enhancement of Hyperspectral Remote Sensing Imagery

Nour Aburaed

Centre for Excellence in Signal and Image Processing
Department of Electronic and Electrical Engineering
University of Strathclyde, Glasgow

Supervised by:

Prof. Stephen Marshall
Dr. Jaime Zabalza
Prof. Hussain Al Ahmad

October 4, 2023

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Abstract

Hyperspectral Imaging (HSI) has emerged as a powerful tool for capturing detailed spectral information across various applications, such as remote sensing, medical imaging, and material identification. However, the limited spatial resolution of acquired HSI data poses a challenge due to hardware and acquisition constraints. Enhancing the spatial resolution of HSI is crucial for improving image processing tasks, such as object detection and classification. This research focuses on utilizing Single Image Super Resolution (SISR) techniques to enhance HSI, addressing four key challenges: the efficiency of 3D Deep Convolutional Neural Networks (3D-DCNNs) in HSI enhancement, minimizing spectral distortions, tackling data scarcity, and improving state-of-the-art performance.

The thesis establishes a solid theoretical foundation and conducts an in-depth literature review to identify trends, gaps, and future directions in the field of HSI enhancement. Four chapters present novel research targeting each of the aforementioned challenges. All experiments are performed using publicly available datasets, and the results are evaluated both qualitatively and quantitatively using various commonly used metrics.

The findings of this research contribute to the development of a novel 3D-CNN architecture known as 3D Super Resolution CNN 333 (3D-SRCNN333). This architecture demonstrates the capability to enhance HSI with minimal spectral distortions while maintaining acceptable computational cost and training time. Furthermore, a Bayesian-optimized hybrid spectral-spatial loss function is devised to improve the spatial quality and minimize spectral distortions, combining the best characteristics of both domains.

Addressing the challenge of data scarcity, this thesis conducts a thorough study on Data Augmentation techniques and their impact on the spectral signature of HSI. A new Data Augmentation technique called CutMixBlur is proposed, and various combinations of Data Augmentation techniques are evaluated to address the data scarcity challenge, leading to notable enhancements in performance.

Lastly, the 3D-SRCNN333 architecture is extended to the frequency domain and wavelet domain to explore their advantages over the spatial domain. The experiments reveal promising results with the 3D Complex Residual SRCNN (3D-CRSRCNN), surpassing the performance of 3D-SRCNN333.

The findings presented in this thesis have been published in reputable conferences and journals, indicating their contribution to the field of HSI enhancement. Overall, this thesis provides valuable insights into the field of HSI-SISR, offering a thorough understanding of the advancements, challenges, and potential applications. The developed algorithms and methodologies contribute to the broader goal of improving the spatial resolution and spectral fidelity of HSI, paving the way for further advancements in scientific research and practical implementations.

Contents

List of Figures

List of Figures

List of Tables

List of Abbreviations

2D Two-dimensional

3D Three-dimensional

AI Artificial Intelligence

ANN Artificial Neural Network

BN Batch Normalization

BOA Bayesian Optimization Algorithm

CC Complex Convolution

CMSE Complex MSE

CNMF Coupled Non-negative Matrix Factorization

CNN Convolutional Neural Network

CReLU Complex Rectified Linear Unit

CRSRCNN Complex Residual Super Resolution Convolutional Neural Network

CSRCNN Complex Super Resolution Convolutional Neural Network

CS Cosine Similarity

CVNN Complex-valued Neural Network

DCNN Deep Convolutional Neural Network

DFT Discrete Fourier Transform

DL Deep Learning

DWT Discrete Wavelet Transform

EDSR Enhanced Deep Super Resolution

ERGAS Erreur Relative Globale Adimensionnelle de Synthèse

FCNN Full Convolutional Neural Network

FFT Fast Fourier Transform

FOV Field of View

FS-3DCNN Frequency Separated 3D Convolutional Neural Networks

GP Gaussian Process

GT Ground Truth

HH High-High

HL High-Low

HR High Resolution

HSI Hyperspectral Imagery

IFOV Instantaneous FOV

LHC Log Hyperbolic Cosine

LH Low-High

LL Low-Low

LR Low Resolution

MAE Mean Absolute Error

ML Machine Learning

MSE Mean Squared Error

MSI Multispectral Imagery

MSLE Mean Squared Log Error

PSF Point Spread Function

PSNR Peak Signal-to-Noise Ratio

ReLU Rectified Linear Unit

RMSE Root Mean Squared Error

ROSIS Reflective Optics System Imaging Spectrometer

RUNet Residual UNet

SAM Spectral Angle Mapper

SISR Single Image Super Resolution

SNR Signal-to-Noise Ratio

SRCNN333 Super Resolution Convolutional Neural Network with kernerl size (3,3,3)

SRCNN915 Super Resolution Convolutional Neural Network with kernerl size (9,1,5)

SRCNN Super Resolution Convolutional Neural Network

SRF	Spectral Response Function
SRGAN	Super Resolution Generative Adversarial Network
SR	Super Resolution
SSIM	Structural Similarity Index Measurement
TC	Transpose Convolution
UCB	Upper Confidence Bound
UIQI	Universal Image Quality Index
VDSR	Very Deep Super Resolution
WRCNN	Wide Residual Convolutional Neural Network
WSRCNN	Wavelet Super Resolution Convolutional Neural Network

Nomenclature

α	Hyperparameter of the hybrid spectral-spatial loss function
η	Binary mask
γ	Parameter that balances ν and ι
ι	Exploration region
$\mathbb{I}(z)$	Imaginary part of a complex number
$\mathbb{R}(z)$	Real part of a complex number
$\widehat{\mathbf{Y}}_{\mathbb{C}}$	Complex estimated high resolution HSI
$\widehat{\mathbf{Y}}$	Estimated high resolution HSI

$\mathbf{Conv}'(\cdot,\cdot)$	TC function
$\mathbf{Conv}(\cdot,\cdot)$	Convolution function
$\mathbf{Conv}_{\mathbb{C}}$	Complex output feature
$\mathbf{I}'$	New $\mathbf{I}$ after applying Data Augmentation
$\mathbf{I}$	A single band from an HSI cube
$\mathbf{K}_{\mathbb{C}}$	Complex kernel
$\mathbf{K}$	Kernel
$\mathbf{X}'$	New $\mathbf{X}$ after applying Data Augmentation
$\mathbf{X}_{\mathbb{C}}$	Complex low resolution HSI
$\mathbf{X}$	Low resolution HSI
$\mathbf{Y}'$	New $\mathbf{Y}$ after applying Data Augmentation
$\mathbf{Y}_{\mathbb{C}}$	Complex ground truth high resolution HSI
$\mathbf{Y}$	Ground truth high resolution HSI
$\mathcal{E}$	Additive noise
$\mu_{(\cdot)}$	Local means
ν	Exploitation region
ϕ	Lowpass filter
ψ	HSI-SISR model
$\sigma_{\mathbf{Y}\widehat{\mathbf{Y}}}$	Cross-covariance

$\sigma_{(\cdot)}$	Standard deviation
$\mathbf{D}$	Downsampling operation
$\mathbf{G}$	Blurring kernel
θ	Rotation angle
φ	Model parameters
ϱ	Highpass filter
$\widehat{\varphi}$	Updated model parameters
$arg(z)$	The angle between the the positive real axis and the line connecting between the origin and z
B	Number of bands in HSI cube
b	Bias
$b_{\mathbb{C}}$	Complex bias
C	Number of bands in a kernel
$f(\cdot)$	Activation function
$f^*(\cdot)$	Objective function
$F_h(\cdot)$	Horizontal flipping function
$F_v(\cdot)$	Vertical flipping function
$G^*(\cdot)$	GP function
$L(\cdot,\cdot)$	Loss function
M, N	Spatial dimension of HSI cube or kernel

$MSE(\cdot, \cdot)$	Function that calculates MSE between 2 images
$PSNR(\cdot, \cdot)$	Function that calculates PSNR between 2 images
$R(\cdot, \cdot)$	Residual error
$ReLU$	ReLU activation function
$ReLU_{\mathbb{C}}$	Complex ReLU activation function
$Rot(\cdot)$	Rotation function
$S_h(\cdot)$	Horizontal shearing function
$S_v(\cdot)$	Vertical shearing function
$SAM(\cdot, \cdot)$	Function that calculates SAM between 2 images
$SSIM(\cdot, \cdot)$	Function that calculates SSIM between 2 images
$UCB^*(\cdot)$	UCB acquisition function
x	Pixel position on the x-axis
x'	New x position after applying Data Augmentation
y	Pixel position on the y-axis
y'	New y position after applying Data Augmentation

Acknowledgments

I would like to express my heartfelt gratitude to everyone who has contributed to the successful completion of this Ph.D. thesis. This journey would not have been possible without the support, guidance, and encouragement of numerous individuals and institutions. Therefore, I would like to extend my sincerest appreciation to the following:

I am deeply grateful to my supervisors [Prof. Stephen Marshall, Dr. Jaime Zabalza, and Prof. Hussain Al Ahmad] for their invaluable guidance, expertise, and unwavering support throughout the entire duration of my research. Their insightful feedback, constructive criticism, and scholarly advice have been instrumental in shaping the direction of this thesis and enhancing the quality of my work. Their commitment, patience, and belief in my abilities have been truly inspiring.

I am indebted to my fellow colleagues and labmates, both at the Hyperspectral Imaging Center as well as Mohammed Bin Rashid Space Centre (MBRSC) Lab, who have contributed to a stimulating research environment. Their intellectual discussions, technical assistance, and collaborative spirit have played a significant role in shaping my ideas and fostering a sense of camaraderie throughout this journey. Their friendship and support have made the challenges more manageable and the victories more memorable. I would like to express my special gratitude to Eng. Mina Al-Saad for her continuous support and valuable input, which played a crucial role in uplifting my motivation and providing me with helpful feedback.

I would like to sincerely thank the individuals who generously volunteered to proofread my thesis, specifically Dr. Mohammed Alkhatib, Dr. Mohammad Sami Zitouni, Eng. Samar Saeed, and my former MSc supervisor Dr. Faisal Shah Khan. Their unique

perspectives and support have played a crucial role in bringing my thesis to its current stage. Without their assistance, this achievement would not have been possible. I would also like to express my gratitude to Dr. Lina Mohjazi for providing significant moral support and valuable career advice throughout my PhD journey.

I am deeply grateful to my family and friends for their unwavering support, understanding, and encouragement during this demanding phase of my academic journey. My siblings' and my mother's love, patience, and belief in my abilities have been a constant source of motivation and strength. I consider myself incredibly fortunate to have such a strong network of support in my life. Although my late father was unable to witness the culmination of this thesis, I hope he is observing my accomplishments with pride.

I extend my gratitude to MBRSC for creating an environment conducive to research and development and for demonstrating confidence in my capabilities by assigning me diverse projects. Their trust and belief in my skills have not only enriched my personal journey but have also paved the path for my career as a researcher, allowing me to contribute to the industry while pursuing my academic aspirations. I am appreciative of their unwavering support and confidence in my abilities.

I would like to express my deep appreciation to all those individuals whose names might not be mentioned here but have played a significant role in shaping my academic and personal development. Your contributions, whether big or small, have been invaluable, and I am truly grateful for your support.

Last, but not least, I want to express my gratitude to my feline companion, Akechi, who has been an extraordinary friend beyond measure. He may be unfamiliar with hyperspectral imagery, but he has been consistently patient throughout the demanding phases of research, writing, and attending conferences, which I am deeply grateful for. His steadfast presence has served as a source of immense support and provided invaluable stress relief during challenging moments.

This thesis is a culmination of the collective effort, encouragement, and assistance provided by everyone mentioned above. Thank you all for your unwavering support and belief in my abilities.

Chapter 0. Acknowledgments

Chapter 1

Introduction

1.1 Scope and Motivation

For several decades, the importance of remote sensing applications has been rapidly increasing. The first applications of remote sensing methods date back to the 1840s, when aviation enthusiasts took pictures of the ground using newly invented cameras [1]. Then, in 1907, Julius Neubronner invented pigeon photography as an aerial photography technique in which pigeons were equipped with a small, lightweight camera [1]. The biggest leap in the field of remote sensing took place with the world's first satellite to orbit the space, Sputnik, was launched in 1957. The field of satellite imagery has been evolving ever since, especially after one of the earliest space-based images was captured by NASA's Explorer 6 in 1959 [2]. The Corona series of satellites came after Explorer 6, which were used for photographic surveillance, and remained active between 1959 to 1972 [3]. Then, Landsat came in 1972 [1], which further raised the interest in using satellites to monitor the earth's surface. Nowadays, the field of remote sensing is vast and technologically advanced, with hundreds of journal papers and conferences to further exploit the full potential of remote sensing instruments. This increasing interest in remote sensing comes from the fact that it covers a wide range of applications. Some of these applications include geology [4–9], vegetation [10–13], Land Cover Land Use (LCLU) [14–17], and oceanography [18–21]. Each application requires different spatial, spectral, and temporal resolutions depending on its objective. Satellites cater to

different resolutions, such as medium and high resolution satellites to facilitate the various requirements of different applications. Additionally, some satellites are designed for specific tasks, such as weather satellites [22], ocean satellites [23], and earth observation satellites [24–26]. One way of exploiting remote sensing data effectively lies in extracting meaningful information from remote sensing imagery, and this is where the role of image processing techniques becomes important. The field of remote sensing applications encompasses a range of tasks that must be executed with precision to ensure significant outcomes. Among these tasks are object detection [27–30], classification [31–33], and semantic segmentation [34, 35]. The accuracy of the results derived from these tasks is directly influenced by the level of detail obtained about an object. This is where the significance of Hyperspectral Imagery (HSI) becomes evident.

The goal of HSI is to obtain the electromagnetic spectrum, typically between 400-2500 nm, for each pixel in the image of a scene, with the purpose of locating objects and generating further meta-information. HSI is built around continuous, very fine-spaced bands as opposed to Multispectral Imaging (MSI), which is more discrete in nature. Furthermore, MSI typically have a few dozens of bands, whereas HSI have hundreds of bands and they capture signals that offer unique signatures to certain objects.

As a result of trade-offs in manufacturing, specifically the challenge of capturing three-dimensional (3D) signals with a two-dimensional (2D) sensor, achieving high spatial and spectral resolutions simultaneously becomes difficult [36]. The specific reason behind this is explained in more detail in Section 2.3 and Section 2.2.2. Consequently, there exists an inherent inverse relationship between spatial and spectral resolutions. While MSI exhibits high spatial resolution, it lacks in spectral resolution, which is precisely the opposite of HSI, as illustrated in Figure 1.1. Also, there exists a trade-off between data volume and Signal-to-Noise Ratio (SNR), which prevents achieving both simultaneously. HSI have low SNR due to reduced illumination by narrow band filters, and they are noisy if a long exposure time is not guaranteed. Practical image processing applications in the context of remote sensing ideally require images having both high spectral and spatial resolution. Therefore, efforts have been invested in the literature to improve the spatial resolution of HSI while exploiting its high spectral res-

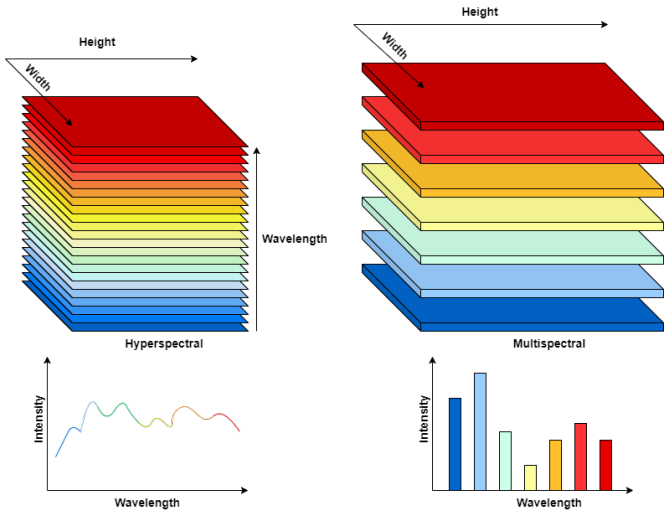


Figure 1.1: An illustration of the difference between HSI and MSI. The height and width represent two dimensions for the spatial domain, and the number of bands represents a third dimension for the spectral domain, which is richer for HSI.

olution simultaneously. One such example is seen in a study conducted by [37], where the authors enhance the spatial resolution of the left imager onboard the Curiosity rover using various pansharpening methods. The goal is to have high spatial and high spectral image cube, which will greatly contribute to the understanding of Mars.

The problem statement at hand revolves around enhancing the spatial resolution of HSI while avoiding spectral distortions. In this context, spectral distortions denote instances where spectral fidelity is not preserved, which is a recurring theme in this thesis.

1.2 Challenges

Within the realm of HSI Single Image Super Resolution (SISR), numerous challenges exist. Nonetheless, this thesis centers on addressing four specific challenges, which are explained in this chapter.

1.2.1 Efficiency of 3D Convolutional Neural Networks

Currently, 3D Convolutional Neural Networks (3D-CNNs) are widely employed as the predominant method for tackling the HSI-SISR problem. However, these networks encounter challenges, such as overfitting and vanishing gradient problems. Additionally, expanding the network size often yields minimal improvements that do not justify the accompanying increase in computational and temporal complexity. Consequently, when designing a network architecture, it is crucial to carefully observe its behavior to prevent overfitting, while also considering the need to minimize processing time and computational demands.

1.2.2 Preserving Spectral Fidelity

The majority of HSI Super Resolution (SR) methods often enhance the spatial resolution of HSI while compromising its spectral resolution. It is vital to minimize spectral distortions since this unique characteristic of HSI enables effective object detection and classification. In this context, spectral distortions means failing to preserve spectral fidelity. Thus, preserving the spectral fidelity of HSI becomes a significant challenge that necessitates careful consideration when developing HSI-SISR approaches.

1.2.3 Data Scarcity

The majority of publicly accessible HSI datasets that are used for testing and benchmarking purposes typically contain a single scene. However, this limited dataset size is insufficient for effectively training and evaluating 3D-CNNs or other learning-based methods. Thus, there is a pressing need to expand the size and diversity of datasets to enable more efficient training of 3D-CNNs and mitigate issues related to overfitting.

1.2.4 Improving State-of-the-art Performance

The current state-of-the-art methods in HSI-SISR still require further improvements in performance. Novel techniques that leverage complex-based approaches can be explored as a possible means of enhancement. These techniques have the potential to enhance the

performance of HSI-SISR by effectively utilizing the spectral-rich information available in the data.

1.3 Problem Statement and Objectives

1.3.1 Problem Statement

The field of HSI-SISR faces a critical challenge in enhancing the spatial resolution of HSI while preserving its spectral fidelity. Existing HSI-SR methods often struggle to achieve balance between spatial and spectral resolution, leading to trade-offs and compromises in the quality of the output images. Furthermore, the limited availability of large-scale and diverse training datasets poses additional obstacles in training accurate and robust HSI-SR models.

The existing HSI-SR techniques frequently suffer from issues such as spectral distortion, spatial artifacts, and inadequate exploitation of the rich spectral information present in HSI [38–40]. These limitations hinder the effective utilization of HSI data for applications, such as target detection, classification, and analysis, where both high spatial and spectral resolutions are crucial [41].

Therefore, there is a pressing need to develop novel HSI SISR approaches that can effectively enhance the spatial resolution of HSI while preserving their spectral details and characteristics. Additionally, the creation of large-scale and diverse training datasets is essential to enable fair comparisons, reproducibility, and advancements in the field of HSI SISR.

Addressing these challenges and developing robust and efficient HSI-SR techniques will significantly contribute to unlocking the full potential of HSI and enable its broader utilization in various applications ranging from remote sensing and agriculture to medical imaging and surveillance.

1.3.2 Objectives

To address the aforementioned technical problems, the aim of this thesis is described in the following objectives:

- To overcome the obstacle of training 3D-CNN models that enhance the spatial resolution of HSI without encountering issues, such as overfitting or vanishing gradient problems
- To minimize spectral distortions while boosting the spatial resolution of HSI
- To tackle dataset scarcity problem by increasing the size and variety of existing datasets
- To explore whether complex-valued or wavelet CNNs can outperform their real-valued counterparts

1.4 Original Contributions and Research Outcomes

1.4.1 Original Contributions

In this thesis, various techniques are introduced for enhancing the spatial resolution of HSI, also known as HSI-SR, primarily through the utilization of SISR approaches. The objective is to address several common challenges associated with HSI-SR, including the efficiency of the approach, spectral distortions, and data scarcity. The contributions are listed as follows:

1. Conducting a thorough literature review on HSI-SR, identify the gap in the research field based on detailed meta-analysis, and propose novel approaches accordingly
2. Devising three novel 3D-CNNs, each with distinct depths: shallow, medium, and deep. The performance of each of these networks is evaluated using three datasets of varying sizes: small, medium, and large. This analysis focuses on investigating the impact of network depth versus dataset size, thereby emphasizing the relationship between these two factors.
3. Boosting the perseverance of HSI spectral fidelity through devising a novel Bayesian-optimized hybrid spectral-spatial loss function. This innovative approach com-

bines the advantages of both spatial and spectral loss functions, aiming to provide the optimal solution for preserving the fidelity of HSI data.

4. Addressing the challenge of data scarcity in HSI-SR by conducting a thorough investigation on diverse Data Augmentation techniques. The objective is to evaluate their impact on HSI-SISR performance and spectral fidelity. Additionally, a novel Data Augmentation technique called CutMixBlur is developed to further mitigate the data scarcity problem.

5. Exploring complex- and wavelet-domain approaches as a potential enhancement of the state-of-the-art HSI-SISR techniques. An in-depth analysis is conducted to evaluate their respective strengths and weaknesses compared to their real-valued counterparts.

1.4.2 Research Outcomes

The work presented in this thesis has led to the publications listed in Table 1.1, which are listed in chronological order according to publication date. It is worth noting that this industrial PhD has been carried out concurrently with other projects while working as a Research Assistant at the Mohammed Bin Rashid Space Centre (MBRSC) Lab based at the University of Dubai. The projects include (but not limited to): Hyperspectral Classification [42, 43], Flood Mapping, Detecting and Counting Palm Trees from UAV Images [44], PolSAR Classification, Airbus ship detection [27], and oil spill segmentation [45]. As these projects are out of the scope of this thesis, the interested reader is referred to the publications listed in Appendix B.

All the developed codes for this PhD project have been shared through the following GitHub repository: `https://github.com/NourO93/SISR_Library`. This repository will be frequently updated. Additionally, the HSI Super Resolution Techniques developed in Chapter 4 of this thesis have been utilized by MBRSC to enhance their MSI.

Table 1.1: A list of publications that resulted as a direct outcome of this research thesis. Each publication is listed with its corresponding contribution from Section 1.4.1 and relevant thesis chapter.

Publication	Contribution	In Thesis
N. Aburaed, M. Q. Alkhatib, S. Marshall, J. Zabalza and H. Al Ahmad, "3D Expansion of SRCNN for Spatial Enhancement of Hyperspectral Remote Sensing Images," *International Conference on Signal Processing and Information Security (ICSPIS)*, Dubai, United Arab Emirates, 2021, pp. 9-12, doi: `https://doi.org/10.1109/ICSPIS53734.2021.9652420`.	(2)	Chapter 4
N. Aburaed, M. Q. Alkhatib, S. Marshall, J. Zabalza and H. A. Ahmad, "SISR of Hyperspectral Remote Sensing Imagery Using 3D Encoder-Decoder RUNet Architecture," *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Kuala Lumpur, Malaysia, 2022, pp. 1516-1519, doi: `https://doi.org/10.1109/IGARSS46834.2022.9883578`.	(2)	Chapter 4
N. Aburaed, M. Q. Alkhatib, S. Marshall, J. Zabalza and H. A. Ahmad, "A Comparative Study of Loss Functions for Hyperspectral SISR," *30th European Signal Processing Conference (EUSIPCO)*, Belgrade, Serbia, 2022, pp. 484-487, doi: `https://doi.org/10.23919/EUSIPCO55093.2022.9909827`.	(3)	Chapter 5
N. Aburaed, M. Q. Alkhatib, S. Marshall, J. Zabalza and H. A. Ahmad, "Complex-valued Neural Network for Hyperspectral Single Image Super Resolution," in *Proc. SPIE 12338 - Hyperspectral Imaging and Applications II*, vol. 123380H, 2023, doi: `https://doi.org/10.1117/12.2645086`.	(5)	Chapter 7
N. Aburaed, M. Q. Alkhatib, S. Marshall, J. Zabalza and H. Al Ahmad, "A Review of Spatial Enhancement of Hyperspectral Remote Sensing Imaging Techniques," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 2275-2300, 2023, doi: `https://doi.org/10.1109/JSTARS.2023.3242048`.	(1)	Chapter 3
N. Aburaed, M. Q. Alkhatib, S. Marshall, J. Zabalza and H. Al Ahmad, "Hyperspectral Data Scarcity Problem From a Super Resolution Perspective: Data Augmentation Analysis and Scheme," *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, California, USA, 2023. [Accepted]	(4)	Chapter 6
N. Aburaed, M. Q. Alkhatib, S. Marshall and H. Al Ahmad, "Attention-infused 3D-SRCNN for Hyperspectral Image Super Resolution," *European Signal Processing Conference (EUSIPCO)*, Helsinki, Finland, 2023. [Accepted]	Future Work	Chapter 8
N. Aburaed, M. Q. Alkahtib, S. Marshall and H. Al Ahmad, "Bayesian Hybrid Loss for Hyperspectral SISR Using 3D Wide Residual CNN," *IEEE International Conference in Image Processing (ICIP)*, Kuala Lumpur, Malaysia, 2023. [Accepted]	(2,3)	Chapters 4 & 5

1.5 Thesis Organization

This PhD thesis is organized into the following chapters to provide a coherent structure and logical progression of the research study:

1.5.1 Chapter 2: Technical Background and Theoretical Framework

This chapter provides the necessary technical background for HSI-SISR. Specifically, it delves into the explanation of CNNs, various image resolution types, and the distinctive features and acquisition modes of HSI. Furthermore, the chapter introduces the characteristics of widely used HSI datasets and outlines the performance metrics that will be employed to evaluate the approaches presented in this thesis.

1.5.2 Chapter 3: Hyperspectral Single Image Super Resolution

This chapter serves as a literature review for HSI-SISR and aims to establish a connection between the past and present research. It provides a thorough overview of HSI-SR in general and then narrows its focus to HSI-SISR approaches, categorizing them into traditional and DCNN methods. Additionally, the chapter includes a statistical meta-analysis of the discussed approaches, shedding light on research trends within the field. This analysis highlights aspects, such as commonly used datasets and evaluation metrics, as well as identifying existing gaps in current research. Subsequently, the chapter introduces a data pre-processing technique that will be consistently applied throughout all chapters. Additionally, it outlines the experimental design, which serves to provide justification for the logical progression of the thesis.

1.5.3 Chapter 4: Extension of 2D-CNNs to 3D-CNNs

In this chapter, three 2D-CNN architectures of different depths are introduced and extended to the 3D domain to investigate their behavior on the following benchmarking datasets: Pavia University, Pavia Center, Indian Pines, and Washington DC Mall. The objective is to develop a network capable of improving the spatial resolution of HSI while preserving its spectral signature. The network should also address challenges, such as overfitting and vanishing gradient problems, which are commonly encountered in 3D-CNNs. Furthermore, it should offer reasonable computation time and memory requirements. Experimental results demonstrate the advantage 3D-CNNs have over 2D-CNNs, in addition to the fact that deep 3D-CNNs are more susceptible to overfitting and vanishing gradient problems, particularly when dealing with limited dataset sizes. Even

with larger datasets, the benefits of deep 3D-CNNs in terms of quantitative metrics do not sufficiently outweigh their high computational time and memory usage. The 3D-CNN with the optimal results is the 3D Super Resolution CNN with filter sizes (3,3,3), which is referred to as 3D-SRCNN333.

1.5.4 Chapter 5: Loss Functions

In this chapter, a thorough review is conducted on the spectral and spatial loss functions commonly employed in SISR. The aim is to identify the most effective functions based on quantitative metrics. The identified best-performing loss functions are then combined into a hybrid loss function, which includes a hyperparameter to regulate the emphasis on the spatial and spectral aspects. The optimization of this hyperparameter is achieved using BOA. Experimental results utilizing the optimal 3D-CNN model developed in Chapter 5.6 indicate that the Charbonnier loss function yields the best spatial loss performance. Furthermore, when paired with the Cosine Similarity (CS) loss function, the hybrid loss function significantly enhances the network's performance compared to the individual use of each function.

1.5.5 Chapter 6: Data Augmentation

In this chapter, the focus is on addressing the issue of data scarcity in HSI by exploring effective Data Augmentation techniques. These techniques aim to increase the size of HSI datasets while minimizing any significant alterations to image features. Each Data Augmentation technique is individually tested to assess its impact on HSI. It is essential not only to maximize the Peak Signal-to-Noise Ratio (PSNR) and Structure Similarity Index Measurement (SSIM) but also to minimize the Spectral Angle Mapper (SAM). Furthermore, a novel Data Augmentation technique called CutMixBlur is introduced. This technique involves blending two images together to create a new data sample while introducing noise only in specific areas of the image. This approach enhances the model's capability to handle image discontinuities and edges while localizing the enhancements. Experimental results indicate that Data Augmentation techniques that lead to the appearance of black borders around the image, such as rotation and

shearing, significantly distort the SAM. Additionally, all effective Data Augmentation methods are combined in two different random manners: Random Stacking and Random Compound. The former applies each Data Augmentation technique to different images, while the latter accumulates the effects of Data Augmentation on the same set of images. The experiments demonstrate that the Random Stacking technique achieves better quantitative and qualitative performance compared to Random Compound.

1.5.6 Chapter 7: Complex- and Wavelet-based Approaches

In this chapter, the HSI-SISR problem is approached from both the complex-domain and wavelet-domain perspectives to explore their potential advantages over real-valued 3D-CNNs. The optimal 3D-CNN model obtained from Chapter 4, which is 3D-SRCNN333, is transformed into a 3D complex-valued CNN, referred to as 3D Complex SRCNN (3D-CSRCNN). Additionally, an alternative version called 3D Complex Residual SRCNN (3D-CRSRCNN) is developed to investigate the impact of an extra residual connection on the network, serving as an ablation analysis. The Fast Fourier Transform (FFT) is utilized to convert the HSI cube into the complex domain. On the other hand, in the wavelet analysis approach, the HSI cube is decomposed into four wavelet components: Low-Low (LL), Low-High (LH), High-Low (HL), and High-High (HH). Each component is then processed using 3D-SRCNN333, and the resulting outputs are merged to form the final enhanced image, leading to the creation of the 3D Wavelet SRCNN (3D-WSRCNN). The research in this chapter draws inspiration from the work presented in [46], using the 3D-SRCNN333 architecture as a benchmark for performance comparison. Experimental results demonstrate that the 3D-WSRCNN does not achieve state-of-the-art performance. However, the 3D-CRSRCNN outperforms both the real-valued 3D-SRCNN333 and the 3D-CSRCNN. It is important to note that the 3D-CRSRCNN incurs a higher computational time compared to the 3D-SRCNN333, highlighting an important trade-off to consider.

1.5.7 Chapter 8: Conclusion and Future Work

This chapter provides a summary of the key objectives of this thesis and presents the research findings from each chapter. It emphasizes the main contributions that have emerged from the conducted research experiments. Furthermore, the chapter outlines potential future directions for this research and identifies unexplored areas that have the potential to yield further advancements in HSI-SISR.

1.6 Summary

This thesis focuses on enhancing the spatial resolution of HSI while preserving its spectral fidelity. It addresses four main challenges: 1) efficiency of 3D-CNNs), 2) preserving spectral fidelity, 3) data scarcity, and 4) improving state-of-the-art performance. Novel 3D-CNNs with varying depths are introduced to efficiently enhance the spatial resolution of HSI with minimal spectral distortions. A Bayesian-optimized hybrid spectral-spatial loss function is devised to preserve spectral fidelity effectively. The problem of data scarcity is addressed through a thorough investigation of Data Augmentation techniques and the development of a novel Data Augmentation technique. Additionally, complex- and wavelet-domain approaches are explored for HSI Super Resolution. The research outcomes include several publications in reputable conferences and journals, in addition to sharing codes with the wider scientific community on Github. This work has practical applications in remote sensing and other fields where HSI data with both high spatial and spectral resolutions are essential.

Chapter 2

Technical Background and Theoretical Framework

2.1 Convolutional Neural Networks

Artificial Neural Networks (ANNs) are a branch of Machine Learning (ML), which is in turn a branch of Artificial Intelligence (AI). ANNs consist of an input layer, one or more hidden layers, and an output layer. When an ANN has one hidden layer, it is known as a shallow neural network. Otherwise, it is known as a deep neural network. A particular class of ANNs that is designed to perform image processing tasks is called Convolutional Neural Networks (CNNs), which was first introduced in the 1990s by Yann LeCunn [47, 48]. CNNs require high computational resources and processing power, which computers could not achieve at that time. Nowadays, with the rapid development of technology, CNNs started gaining more attention, especially after a CNN successfully won ImageNet challenge of classifying 1.2 million images in 2014 [49]. CNNs are now used for various other image processing tasks, including object detection, semantic segmentation, and SR. A CNN that performs SR tasks typically consists of a combination of two or more of the following: convolutional layer, activation function, and pooling layer. CNNs can include other types of layers, such as Batch Normalization (BN). For the context of this study, only the layers relevant to SR will be discussed. For the rest of the thesis, all operations are assumed 2D unless stated otherwise.

Convolution is the product of element-wise multiplication between an image and a filter that consists of one or more kernels. When the filter consists of one kernel, the two terms can be used interchangeably. The filter passes through the image in a specified stride, which refers to the step size by which the convolutional filter moves across the input image when performing convolution operation. This process reduces the size of the image, unless it is padded before applying the convolution filter. The simplest way of padding is by adding zeroes at the border of the image [50]. An example will be discussed in Section 4.3.

For a certain band k of Low Resolution (LR) HSI $\mathbf{X}^k$ of size $N \times N$ and a kernel $\mathbf{K}$ of size $M \times M$, where $N \gg M$, convolution at pixel position (x, y) can be expressed as follows:

$$Conv_{(x,y)} = f\left(\sum_{i=1}^{M}\sum_{j=1}^{M} K_{(i,j)} X^k_{(x+i,y+j)} + b\right), \tag{2.1}$$

where $Conv_{(x,y)}$ is the output feature, $X^k_{(x+i,y+j)}$ is the input that includes the original pixel and the neighboring pixels within the offset range (i, j), $K_{(i,j)}$ is the weight at location (i, j) that corresponds to the input, b is the bias, and f is the activation function. Some of the most commonly used activation functions are Sigmoid and Rectified Linear Activation Unit (ReLU), which are seen in Figure 2.1. According to the literature, ReLU is one of the most suitable activation functions for SR CNNs due to its simplicity and computational efficiency [51]. Additionally, it helps to mitigate vanishing gradient problem [51]. Furthermore, the work in [52] demonstrates that for 3D-CNNs designed for HSI-SISR, ReLU outperforms both Tanh and Parameterized ReLU (PReLU).

The result of the convolution operation is a feature map that summarizes key features of the convolved image. In the case where the image has multiple bands, the filter convolves each band individually. Figure 2.2a illustrates the process of 2D convolution. Convolution causes dimensionality reduction, and the new image size is calculated as follows:

$$N' = N - M + 1 \tag{2.2}$$

Pooling is another key operation in CNNs, which is the process of downsampling an image by selectively discarding features and preserving the important ones. There are

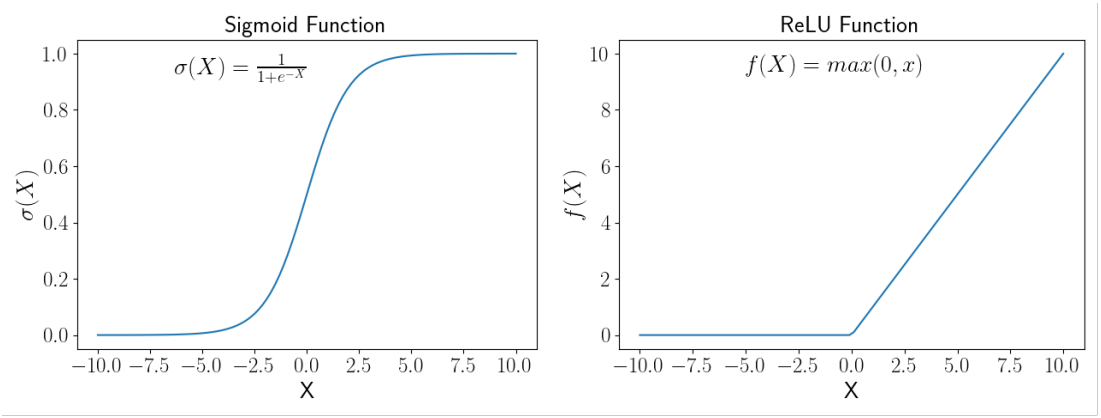
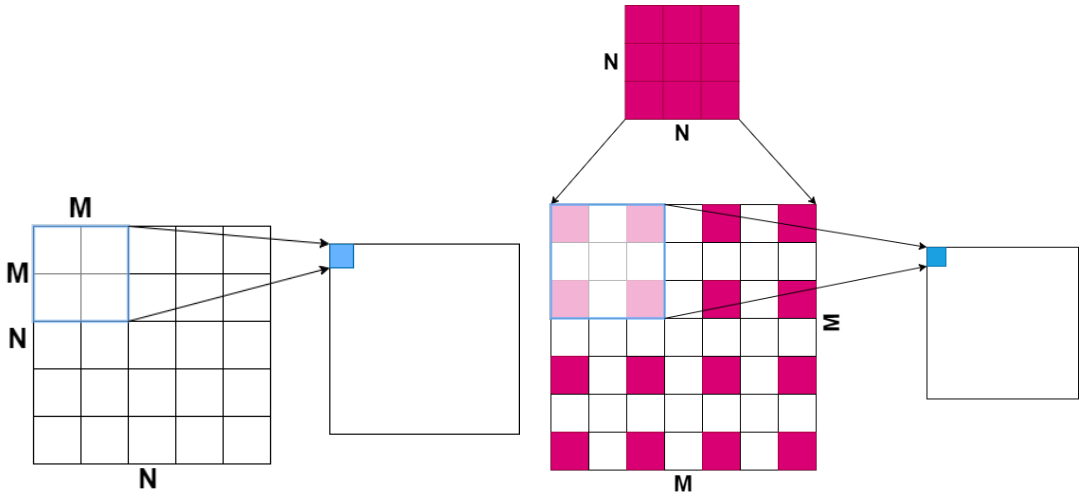


Figure 2.1: Sigmoid and ReLU activation functions.



(a) 2D Convolution (b) 2D Transpose Convolution

Figure 2.2: Illustration of 2D Convolution and 2D Transpose Convolution on an image of a single band.

two types of commonly used pooling; max pooling and average pooling. For example, for **X** of size 6×6 and **K** of size 2×2 with stride 2, the max pooling kernel passes through **X** to produce a feature map by preserving only the highest value and discarding the lower ones. The result is as illustrated in Figure 2.3. Similar to convolution operations, pooling causes images to lose dimensionality if they are not padded.

The counterparts of convolution and pooling operations that reverse their effects are Transpose Convolution (TC) and upsampling, respectively. These operations are commonly used in generative CNNs, such as Autoencoders. Autoencoders learn spatial

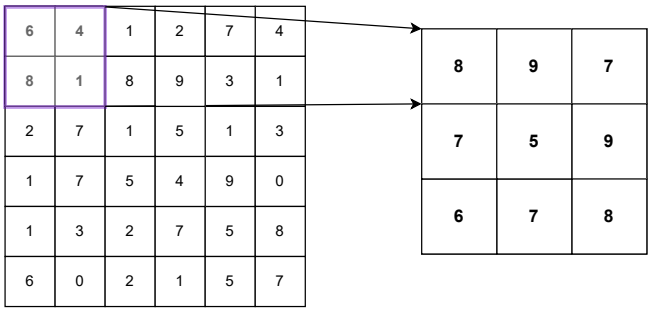


Figure 2.3: An example of pooling a 6×6 image using a max pooling filter of size 3×3 and stride 2.

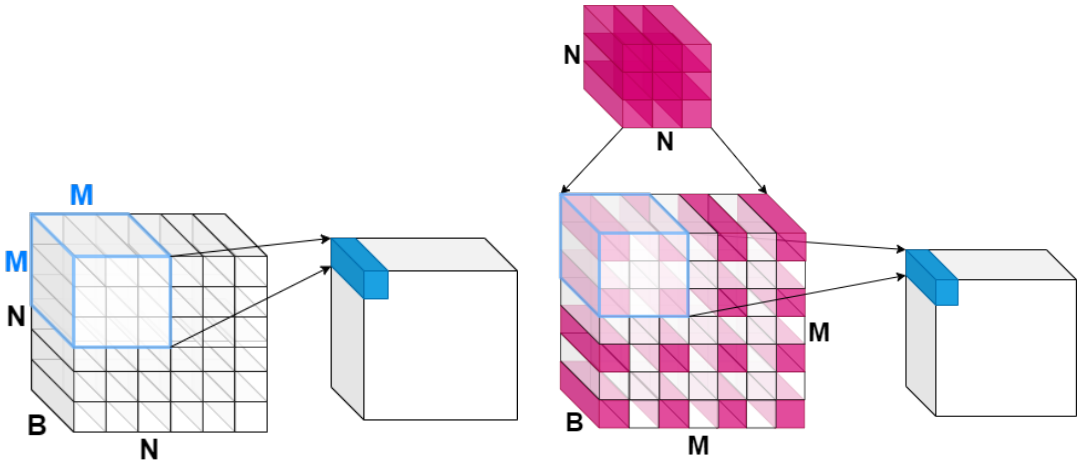
mappings from one image to another, and can be re-purposed to be used for spatial enhancement, as will be seen in Section 4.5. Upsampling can be achieved using any interpolation operation, such as the commonly used bicubic interpolation. TC operates in a similar manner as convolution, however, the kernel in Equation 2.1 is replaced with $\mathbf{X}$ of size $N \times N$, which is convolved with a grid of size $M \times M$, where $M \gg N$. The known input values are spread across the grid and the values in between are set to zero. Figure 2.2b illustrates how TC operates on an image of a single band.

All the aforementioned 2D operations can be extended to 3D such that the calculations are performed over the entire HSI cube rather than processing each band individually. For instance, 3D convolution spans all three directions; height, width, and bands. Therefore, it is an adequate solution to accommodate spectral context. For an LR-HSI denoted $\mathbf{X}$ of size $N \times N \times B$ and a kernel $\mathbf{K}$ of size $M \times M \times C$, 3D convolution at position (x, y, z) can be expressed with the following equation:

$$Conv_{(x,y,z)} = ReLU\left(\sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{C} K_{(i,j,k)} X^{z+k}_{(x+i,y+j)} + b\right) \tag{2.3}$$

Figure 2.4a provides a visual illustration of 3D convolution. Similar to 2D Convolution, 3D Convolution causes dimensionality reduction if the image is not padded. The dimension can be expanded using 3D-TC, which is seen in Figure 2.4b.

Recently, 3D-CNNs have been commonly utilized since they showed effectiveness in HSI-SR, as seen in Section 3.1.3 and Appendix A.3. All the aforementioned layers can be connected together in different topology, such as feed forward [53], skip (or residual)



(a) 3D Convolution (b) 3D Transpose Convolution

Figure 2.4: Illustration of 3D Convolution and 3D Transpose Convolution on an image cube.

connections [54], attention mechanism [55], and Recursive Neural Networks [56], which can enhance the performance of the network depending on its purpose either in terms of output quality or computational complexity.

2.2 Hyperspectral Imaging

The upcoming sub-sections will introduce essential concepts related to HSI system components, spatial and spectral response, acquisition modes, and radiometric calibration. These concepts are pivotal for comprehending the characteristics of the datasets employed in this thesis and the underlying causes behind sensor trade-offs.

2.2.1 HSI System: Components and Response

Figure 2.5 illustrates the typical components comprising an HSI system. These elements exhibit issues that impact the system's performance. To illustrate, the output from the light source varies in terms of spatial position, wavelength, and time. This issue leads to spatial and spectral non-uniformity, along with a drift in the system's behavior over time. Challenges can also arise from the spectral filters or dispersion components and the physical attributes of the imaging optics. For instance, these ele-

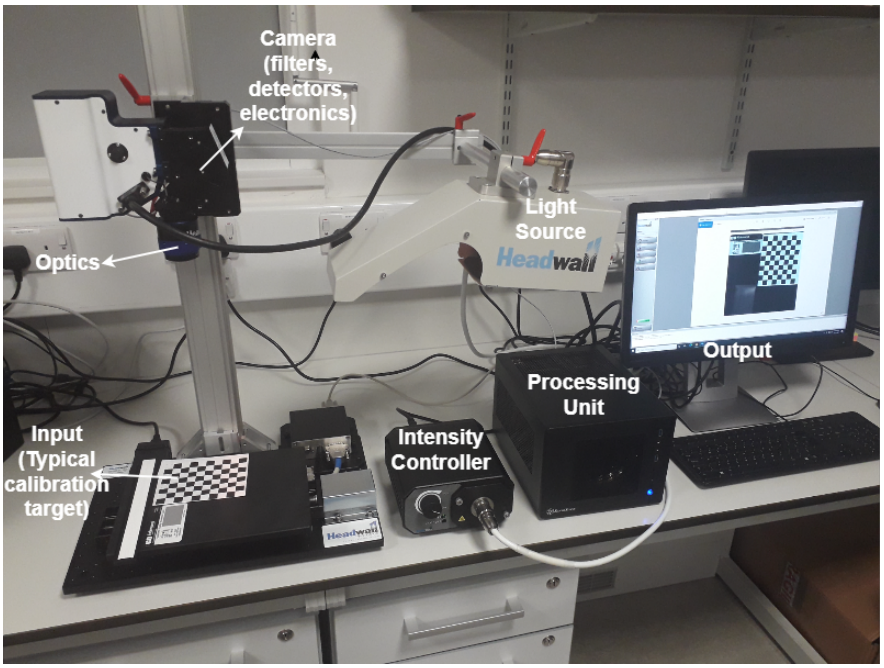


Figure 2.5: Short-Wave Infrared (SWIR) HSI system used in the HSI Center at the University of Strathclyde.

ments may introduce distortions in the geometry and blur the captured scene, [57, 58]. Furthermore, the sensors within the camera are not immune to problems. Typically, a sensor in a Focal Plane Array (FPA) exhibits variations in response based on spatial location, spectral band, and time. The temporal fluctuations in the detector's response are attributable to thermal noise and the quantum efficiency of the detector. In general, detector issues stem from variations in fabrication, detector design, and the inherent quantum properties of current carriers in semiconductor materials [59].

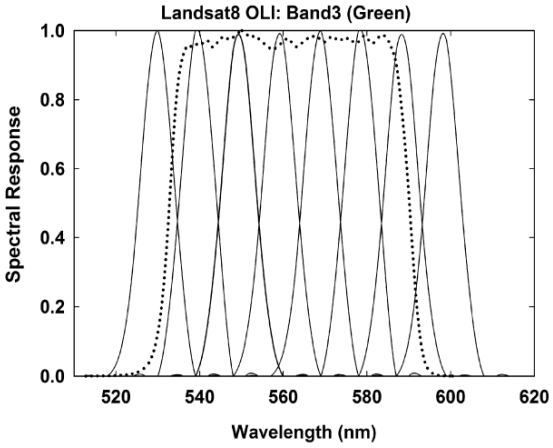


Figure 2.6: SRF comparison of AVIRIS (solid line) and Landsat8 OLI (dashed line) within wavelength range 520 - 600 nm [60].

To gain a better grasp of the HSI system, it is essential to delve into the concept of system response. System response serves as a mathematical depiction of the connection between what goes into the system and what comes out of it. This mathematical relationship serves as a tool for approximating the unknown effects brought about by the various components comprising the system [61]. Interacting with the system can only be accomplished through its input and output terminals. In the case of HSI systems, this input-output relationship must be established as a function of spectral wavelength, spatial location, and, in certain scenarios, time.

The spectral response pertains to the system's sensitivity to radiation at various wavelengths. As an example, Figure 2.6 compares the Spectral Response Function (SRF) between Landsat 8 OLI Band 3 (dashed line) and Airborne Visible / Infrared Imaging Spectrometer (AVIRIS) (solid line) in the wavelength range of 520–600 nm [60]. In systems employing the elements outlined in Figure 2.5, the system's sensitivity is influenced by the transmittance of the spectral filter, the spectral sensitivity of the detector, and the optical system's spectral characteristics. As aforementioned, one or more of these components can cause distortions, which causes the center wavelengths and bandwidths of spectral response functions of HS sensor to shift and broaden [62]. Therefore, spectral calibration is an important step in HSI systems [62].

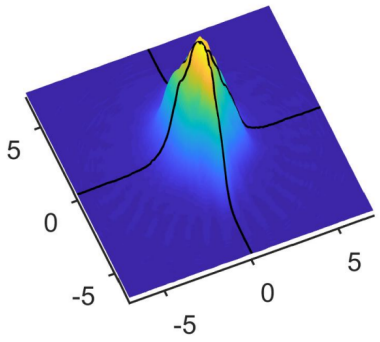


Figure 2.7: 3D plot of a measured PSF, recorded from the SpecIm camera at 699 nm wavelength [63].

The spatial response refers to how the system perceives incoming radiation arriving from various directions. Typically, the system alters the spatial attributes of the scene's radiance that is being captured, introducing distortions and blurring. The sensor's spatial behavior can be characterized using what is known as Point Spread Function (PSF), which illustrates how the response of an individual pixel changes based on its distance from the pixel's center [64]. In the context of HSI, these PSFs typically exhibit a Gaussian-like shape [64]. In a given pixel, all bands should have identical PSFs. Commercial HS cameras commonly assess their spatial coregistration accuracy using a metric known as "keystone" distortion, which primarily measures the shift in pixel centroids caused by any remaining wavelength-dependent magnification variations [63]. Nevertheless, it is important to note that even when keystone distortions are minimal, coregistration challenges can persist due to wavelength-related changes in the width or shape of the PSF [63]. Additionally, sharpness can be measured based on the mean PSF over all bands. Ideally, this mean PSF should coincide with the boundaries of the nominal pixel Field of View (FOV) [63]. Figure 2.7 shows a PSF example measured from SpecIm camera at 699 nm wavelength.

2.2.2 Acquisition Modes

There are several different acquisition modes of HSI, which are classified into four categories; pushbroom, whiskbroom, snapshot, and staring [65]. The whiskbroom mode

obtains data for each band by scanning pixel by pixel and moving the detector in the x-y space. This data is then stored in a Band Interleaved by Pixel (BIP) cube [66]. On the other hand, pushbroom mode follows a similar process, but instead of scanning pixel by pixel, it acquires a complete line of pixels at once, which ultimately forms a Band Interleaved by Line (BIL) cube [66, 67]. Pushbroom systems are usually of the type Charge-Coupled Device (CCD). Pushbroom mode is the most commonly used acquisition mode for HSI, especially in remote sensing, due to several reasons:

- Pushbroom mode allows for rapid data acquisition since it captures an entire line of pixels at once, which results in faster imaging compared to other modes. The continuous array scanning reduces the time needed to acquire a complete image, making it more efficient for large-scale HSI applications [68].

- In pushbroom mode, the detector relies on the motion of the platform, reducing motion artifacts that can degrade image quality. The absence of moving parts also contributes to the overall simplicity and reliability of pushbroom systems [68].

- Pushbroom mode typically provides a higher SNR compared to other acquisition modes. By collecting light from a sequence of pixels forming a line, pushbroom mode averages out the noise, resulting in improved image quality and better spectral accuracy [68]. This is achievable because pushbroom mode has longer dwell time compared to other modes, which is the time spent on each scan line. A longer dwell time implies that the signal strength recorded is high.

- Pushbroom systems can be designed to be compact and lightweight compared to other acquisition modes. This makes them suitable for various applications, including airborne and spaceborne platforms, where size and weight constraints are crucial factors [68].

- Pushbroom mode naturally generates a BIL data cube, which is widely used and compatible with many hyperspectral data analysis and processing techniques. This format simplifies data handling and analysis, facilitating efficient extraction of spectral information from the acquired images [68].

One drawback of pushbroom sensors is the varying sensitivity of the individual detectors in the array (seen in Figure 2.8). In other words, calibration is required for a large number of detectors that make up the sensor system [69]. If the calibration process is not performed accurately or if detectors drift out of calibration over time, it can result in inaccuracies in the data collected by the pushbroom scanner [69]. Also, a very large number of detectors is needed for high resolution images [70]. In addition, the pushbroom scheme requires a wide FOV optics system to obtain the same swath as for a corresponding whiskbroom scanner [70]. An illustration of pushbroom acquisition mode is seen in Figure 2.8. The instantaneous FOV (IFOV) directly affects the image resolution, as will be seen in the next section.

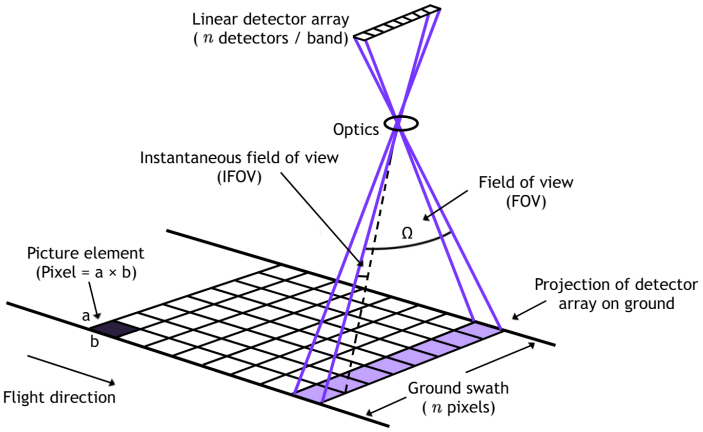


Figure 2.8: Schematic of data acquisition by pushbroom scanner.

2.2.3 Data Characteristics

HSI data have unique characteristics that sets them apart from other spectroscopic approaches. HSI is used to convert radiance (reflected and/or emitted) into the form of data cubes, consisting of two spatial dimensions (rows and columns) and one spectral dimension (bands). This data cube format allows for efficient storage, manipulation, and analysis of the HSI, facilitating the extraction of valuable information from the image [71]. HSI enables spectral unmixing, which is the process of decomposing mixed pixels into their constituent endmembers, also called classes or materials. These end-

members are pure spectral signatures or reference spectra that represent the unique spectral characteristics of materials or substances within a scene or image [72, 73]. By analyzing the spectral signatures of individual endmembers within a pixel, HSI can estimate their abundances or proportions, providing valuable information about the composition of the scene [71]. The coarse spatial resolution of HSI can lead to spectral mixing, which hinders utilizing the full potential of HSI. The high spectral resolution and continuous spectral information of HSI facilitate material identification and classification. By comparing spectral signatures to reference libraries or employing ML algorithms, HSI can identify and classify different materials, such as vegetation types, minerals, and man-made objects [71].

2.2.4 Radiometric Calibration: Radiance and Reflectance

Radiometry is a scientific discipline that explores the generation, propagation, and detection of electromagnetic radiation [74]. It involves the measurement and analysis of numerous radiometric properties, including but not limited to radiance and reflectance [74]. While there are additional radiometric properties, they are not directly pertinent to the context of this thesis.

As aforementioned throughout this chapter, HS sensors used in remote sensing applications capture the spectral characteristics of the Earth's surface in numerous narrow and contiguous bands. When sunlight or solar radiation reaches a surface material, that material reflects the incoming radiation [75]. The amount of energy reflected reveals the spectral properties of the surface material. The radiation incident on the surface and subsequently reflected is referred to as surface reflectance [75]. The sensor, positioned at the Top of the Atmosphere (TOA), measures the radiation that has been reflected, which is known as TOA radiance [76]. TOA radiance would be equal to surface reflectance under ideal conditions. However, atmospheric phenomena like scattering and absorption influence both incident and reflected radiation [75]. This is illustrated in Figure 2.9.

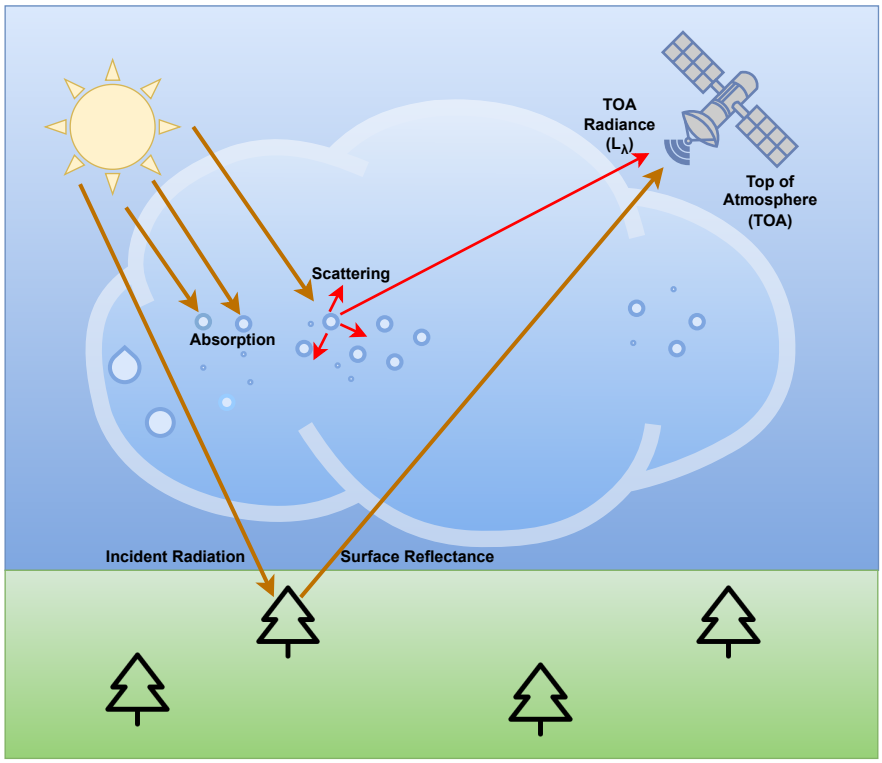


Figure 2.9: Reflectance and radiance in remote sensing.

The raw data recorded by the HS sensors is known as Digital Numbers (DNs) [75,76]. In order to use HSI for meaningful analysis, DN must be converted to TOA radiance and TOA reflectance. This process is known as radiometric calibration [76]. Throughout this thesis, radiance and reflectance will always be assumed to be TOA. Radiometric calibration involves the following steps:

1. The DN values are converted to radiance values by calibrating sensor gain and bias in each spectral band as follows:

$$Rad_k = (DN \times Gain_k) + Bias_k, \tag{2.4}$$

where $Gain_k$ and $Bias_k$ correspond to the gain and bias for spectral band index k [75]. Radiance has units of watt per steradian per square meter ($W \cdot sr^{-1} \cdot m^{-2}$) [76]. The analogous term to radiance in photometry is luminance, which is part of SSIM calculations explained in Section 2.5.2.

2. The radiance is influenced by atmospheric effects, such as absorption and scattering of light. Atmospheric correction aims to remove or compensate for these effects to obtain the surface reflectance values [76]. The conversion from radiance Rad_k to reflectance Ref_k is often performed using the following equation:

$$Ref_k = \frac{\pi Rad_k d^2}{E_0 cos(\theta_{sun})} \tag{2.5}$$

E_0 is the mean extraterrestrial solar irradiance, d the Earth-Sun distance in astronomical units, and θ_{sun} is the solar zenith angle. Reflectance is, thus, the proportion of the radiation striking a surface to the radiation reflected off of it. Therefore, it is dimensionless [76].

In the realm of HSI-SISR, various research studies have shown a preference for working in reflectance space or the radiance space. While reflectance space contains no atmospheric distortion and is considered ideal for HSI analysis, some studies have shown that radiance domain signatures can be sufficient [77]. An intriguing avenue for investigation lies in assessing the robustness of the algorithms presented in this thesis by examining their effectiveness in both radiance and reflectance spaces.

2.2.5 Industrial Importance

HSI is utilized in medical imaging to give precise molecular and functional information about tissues. In order to construct a high-dimensional depiction of the tissue being examined, it captures images of said tissue at various wavelengths and merges them [78]. The identification of numerous biomarkers, the detection of disease, and the tracking of the efficacy of treatments can all be done using this information [78]. For instance, HSI has been used to observe brain blood flow [79], monitor blood oxygenation [80], and find cancer in tissues [81]. HSI is an active area of research in the realm of medical imaging and has the potential to offer significant diagnostic and therapeutic information in a non-invasive manner.

HSI can assess the quality and ripeness of fruits, vegetables, and other agricultural products [82]. By analyzing the spectral signatures of different wavelengths, HSI can

detect internal defects, determine sugar content, identify bruised or damaged areas, and evaluate overall product quality [82]. This information helps in sorting, grading, and ensuring consistency in the agrifood supply chain.

HSI can detect diseases, infections, and pests in plants at an early stage [83]. By capturing detailed spectral information, HSI can identify specific disease signatures or pest infestations that may not be visible to the naked eye. This enables timely intervention, such as targeted treatment or removal of infected plants, reducing the risk of crop loss and improving overall yield and productivity [83].

In terms of remote sensing, HSI enables detailed land cover classification by analyzing the unique spectral signatures of different materials and surfaces. It can differentiate between various types of vegetation, soil types, and water bodies [84, 85]. This information aids in land management, urban planning, and environmental assessments [84]. Moreover, HSI helps monitor and assess environmental parameters such as air pollution [86]. By capturing the full spectrum of light, HSI can detect and quantify pollutants, track changes in vegetation indices, monitor algae blooms, and identify sources of contamination [86]. This data supports effective environmental management and conservation efforts. HSI is also effective in disaster management, as it facilitates rapid and accurate assessment of natural disasters, such as floods [87], wildfires [88], and landslides [89]. It can provide detailed information about the extent of damage, changes in land cover, and the presence of hazards. This data assists in disaster response planning, resource allocation, and post-disaster recovery efforts.

2.3 Image Resolution

An image resolution refers to the amount of details provided by an image. It is a critical factor in determining the quality and clarity of a digital image. Image resolution is particularly important in various fields such as photography, graphic design, and remote sensing. The following list encompasses the different ways an image resolution can be described:

- Pixel count: refers to the total number of pixels in a captured image, typically

expressed as $height \times width \times bands$. For instance, given an HSI with height 64, width 64, and 103 bands, the total pixel count is $64 \times 64 \times 103 = 421,888$ pixels.

- Spatial resolution: is a critical metric used to determine the minimum size of an object that an imaging system can discern [90]. In other words, it gauges the system's ability to distinguish between closely located objects [90]. In the context of remote sensing, spatial resolution pertains to the smallest feature that a pixel in a satellite image can represent [91]. This measure is often conveyed through Ground Sampling Distance (GSD) expressed in meters per pixel (m/pixel) or centimeters per pixel (cm/pixel), which most studies abbreviate as simply "m" or "cm" [92]. Smaller GSD values indicate finer spatial resolution, meaning that the system can capture smaller and more detailed features on the Earth's surface.

- Spectral resolution: describes the ability of a sensor to define fine wavelength intervals. For example, monochrome film captures wavelengths spanning most, if not all, of the visible electromagnetic spectrum [93]. Its spectral resolution is considered low because it does not distinguish individual wavelengths within the visible spectrum [93]. In contrast, color film is responsive to reflected energy across the visible spectrum but exhibits greater spectral resolution. This is because it individually detects reflected energy at specific wavelengths in the blue, green, and red parts of the spectrum [93]. HS sensors operate by detecting hundreds narrow spectral bands across the visible, near-infrared, and mid-infrared regions of the electromagnetic spectrum [93]. Their exceptionally elevated spectral resolution enables precise differentiation among various targets by analyzing their unique spectral reactions within each of these narrow bands. As an example, Reflective Optics System Imaging Spectrometer (ROSIS) sensor yields 103 usable bands between 430-860 nm wavelength, which means that its bandwidth is approximately 4 nm.

- Temporal resolution: refers to the duration required for revisiting and gathering data for the identical location [94]. It is also known as "revisit time". In the context of remote sensing, the timeframe in question depends on both the sensor

platform's orbital features and the characteristics of the sensor itself [94]. Temporal resolution is considered high when the time gap between revisits is brief, and conversely, it is lower when this gap is extended. Typically, temporal resolution is quantified in terms of days [94].

- Radiometric resolution: refers to the sensor's ability to distinguish between the electromagnetic signals reflected by various objects within the same spectral band [95]. It is also known as "dynamic range" or "bit depth", which determines the maximum number of brightness levels available. A sensor with high radiometric resolution is more adept at detecting minor discrepancies in reflected or emitted energy. Imagery data are represented by positive digital numbers between 0 and $2^{\text{bit\_depth}}$. For example, ROSIS sensor's bit depth is 14, hence, it has $2^{14} = 16384$ pixel values available.

It is worth noting that outside the context of remote sensing, some studies use the term "spatial resolution" to refer to pixel count, or the total height and width of an image. However, in this thesis, pixel count and spatial resolution adhere to the definitions mentioned above.

To explain the reason behind why sensors cannot achieve high spectral and spatial resolution simultaneously, the discussion provided in Section 2.2.2 is relevant. A high spatial resolution can be achieved by a decreased IFOV. Less IFOV means less photons, which means that radiometric and spectral resolutions will be affected negatively due to the limited number of photons received by the detector. In order to improve the radiometric and spectral resolutions, IFOV can be increased, which will in turn affect the spatial resolution negatively. Hence, this sensor limitation makes it not possible to have the best of all three resolutions. For this study, only spatial and spectral resolutions are of particular importance, as will be evident throughout the thesis.

2.4 Datasets

HSI are constructed using hyperspectral cameras that are capable of capturing hundreds of imaging bands at different wavelengths for the same spatial area [96]. Typically,

hyperspectral sensors capture images between 400-2500 nm wavelength with regular sampling interval of 4-15 nm. On the other hand, the spatial resolution can be as coarse as 60m, such as NASA's Earth Surface Mineral Dust Source Investigation (EMIT) [97]. Low spatial resolution implies that several objects may be captured within the same pixel, which makes them difficult to identify. This is a key concept behind spectral unmixing, which is discussed in Appendix A.2.1. Section 3.1.4 will illustrate that Pavia University, Pavia Center, Washington DC Mall, and Indian Pines are the most used datasets in the literature, which is why they are used in this study for testing and benchmarking.

2.4.1 Pavia University and Pavia Center

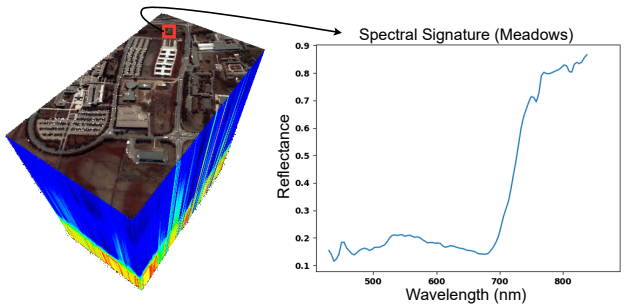


Figure 2.10: Pavia University cube with a sample spectral signature of Meadow class. This dataset is in reflectance domain.

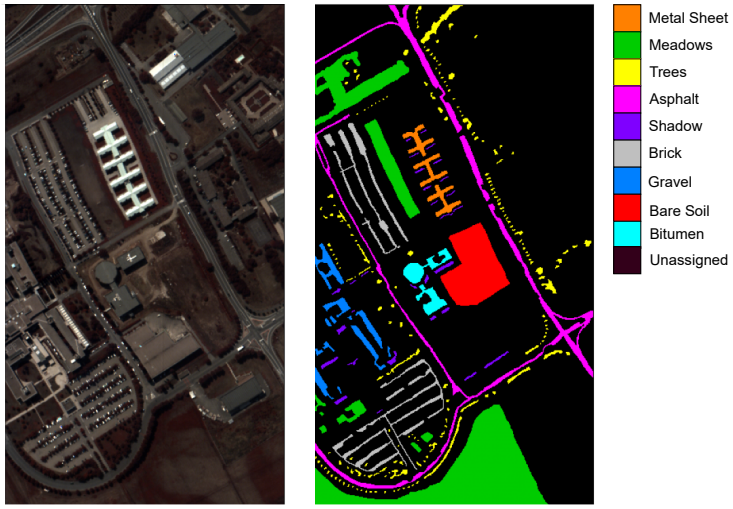


Figure 2.11: RGB composite of Pavia University dataset and its corresponding classes.

Pavia University dataset was captured above an urban region near the University of Pavia in Italy using ROSIS. This dataset comprises 115 bands, with 103 usable bands measuring 610×340 pixels each [98]. It encompasses wavelengths 430-850 nm and has a spatial resolution of 1.3 m [98]. The Ground Truth (GT) of this dataset includes nine categories, predominantly consisting of human-made structures like buildings. Figure 2.10 shows Pavia University cube with a sample spectral signature of "Meadows", and Figure 2.11 shows Pavia University RGB composite and the classes (or endmembers) that exist in the scene. This dataset is available in reflectance domain.

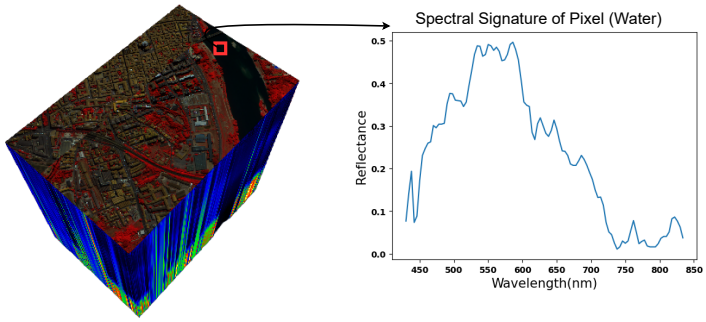


Figure 2.12: Pavia Center cube with with a sample spectral signature of Water class. This dataset is in reflectance domain.

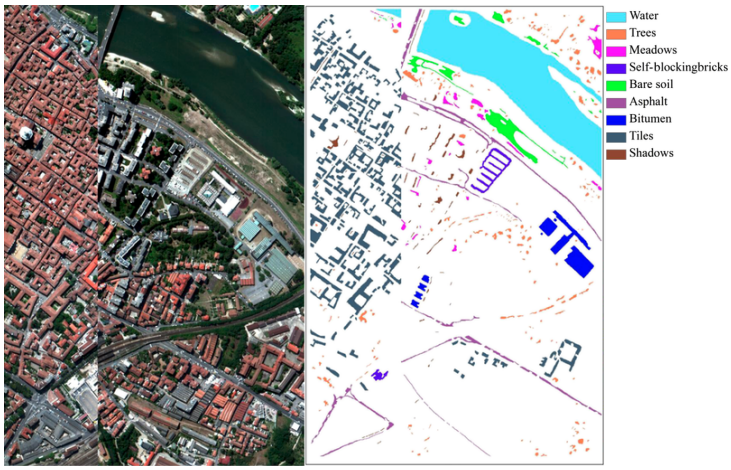


Figure 2.13: RGB composite of Pavia Center dataset and its corresponding classes [99].

A similar dataset to Pavia University is Pavia Center, which was captured using the same sensor and platform, and thus, has the same spectral and spatial resolution as Pavia University [98]. This dataset consists of nine classes, two of which do not exist in Pavia University, which are Water and Tiles. Figure 2.12 shows Pavia Center cube, and Figure 2.13 shows a false color RGB composite of Pavia Center scene and the classes that exist within the scene. This dataset is available in reflectance domain.

2.4.2 Indian Pines

Indian Pines dataset was captured above a farming region in Northwestern Indiana using AVIRIS [100]. The image consists of 220 bands with a 20 m per pixel spatial resolution. Bands 104–108, 150–163, and 220 are corrupted due to water absorption, thus, these bands are discarded, and the remaining number of bands is 200 [101]. Indian Pines cube is shown in Figure 2.14. The Indian Pines scene encompasses 16 land cover categories, the majority of which represent various crop types, as seen in Figure 2.15. This dataset is available in radiance domain.

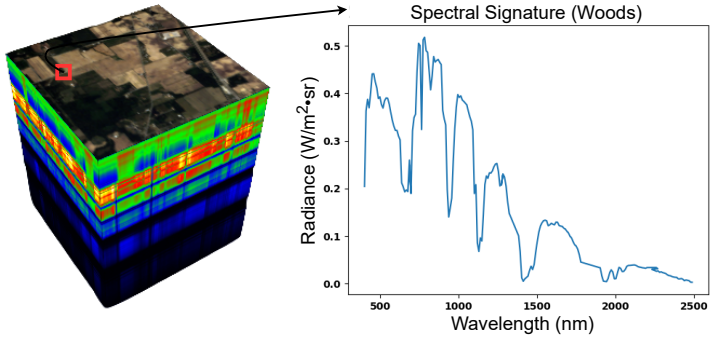


Figure 2.14: Indian Pines cube with a sample spectral signature of Woods class. This dataset is in radiance domain (normalized).

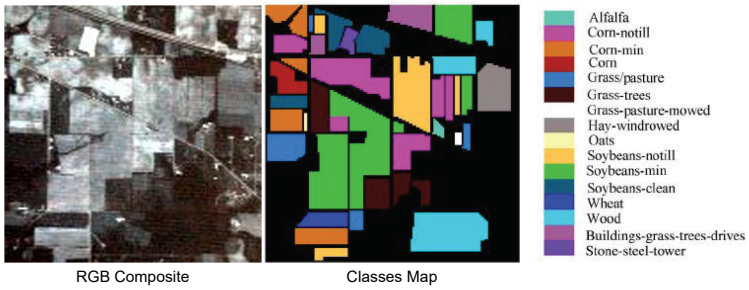


Figure 2.15: RGB composite of Indian Pines dataset and its corresponding classes [102].

2.4.3 Washington DC Mall

Washington DC Mall dataset was captured using Hyperspectral Digital Imagery Collection Experiment (HYDICE) imaging system [103]. It has 1280×307 pixels with 210 bands in the range of 400-240 nm. The spatial resolution is 2 m. There are 191 usable bands, as bands 103-106, 138-148, and 207-210 are corrupted by noise. Washington DC Mall cube is shown in Figure 2.16. This dataset consists of six classes seen in Figure 2.17, and it is available in radiance domain.

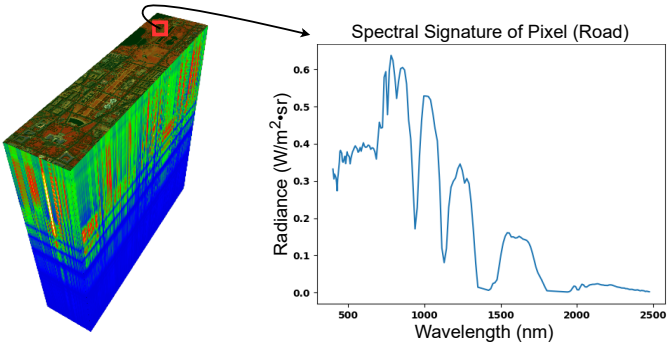


Figure 2.16: Washington DC Mall cube with a sample spectral signature of Road class. This dataset is in radiance domain (normalized).

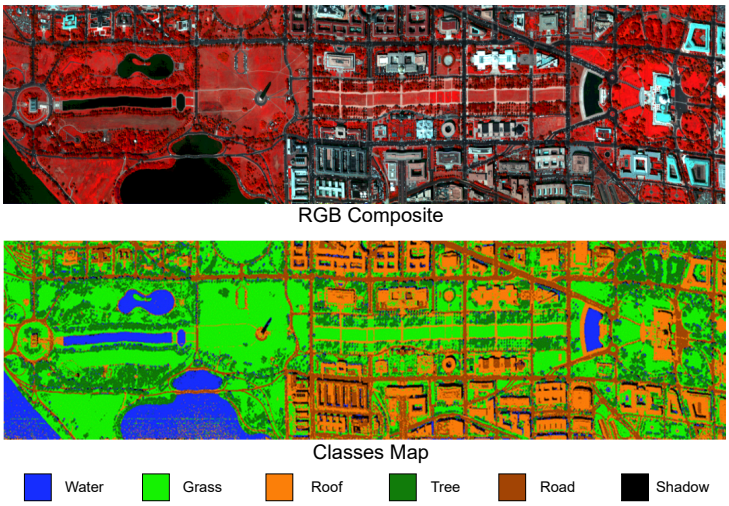


Figure 2.17: RGB composite of Washington DC Mall dataset and its corresponding classes [102].

2.4.4 Others

In addition to the previously discussed datasets, other datasets include NUS [104], Kawakami [73], University of Houston [105], Moffett Field [106], Paris [107], San Francisco [108], and Real Hyperspectral dataset, which consists of Samson, Jasper Ridge, Urban, and Cuprite [109,110]. In 2019, a new HSI dataset called ICONES [111] became publicly available. Due to its recent availability, it has not been used in other research studies thus far, but it is worth mentioning due to its large size and the variety of HSI it provides. Some studies on HSI use CAVE [112] and Harvard [113] datasets, however,

the images in both datasets are not captured using remote sensing devices. Furthermore, CAVE dataset is listed as a multispectral dataset rather than a hyperspectral one [114].

Table 2.1 lists the most widely used HSI datasets and their corresponding characteristics. In addition to the sensors seen in the same table, there are various other sensors, instruments, and data generation models used in the literature to capture or generate HSI datasets, such as Compact High Resolution Imaging Spectrometer (CHRIS) [115], HypXim [116], Compact Airborne Spectrographic Imager (CASI) [117], Digital Imaging and Remote Sensing Image Generation Model (DIRSIG) [118], ASTER [119], PHI [120], Hyperspectral Imager Suite (HISUI) [121], AisaDUAL [122], Airborne Hyperspectral Scanner (AHS) [123] Apex [124], HySpex Visible and Near-Infrared (VNIR) and Short-Wave Infrared (SWIR) cameras [125, 126], EMIT [127], PRSIMA [128], DLR Earth Sensing Imaging Spectrometer (DESIS) [129], and SpecTIR Hyperspectral Airborne Rochester Experiment (SHARE) 2010/2012 [130]

Table 2.1: Examples of the most widely used publicly available HS remote sensing datasets, according to the meta-analysis in Section 3.1.4, and their specifications.

Dataset	Sensor	# of Bands	Spectral range (nm)	Spatial resolution (m)	Platform
Pavia University [98]	ROSIS	103	430-960	3.7	Airborne
Pavia Center [98]	ROSIS	102	430-960	3.7	Airborne
Indian Pines [98]	AVIRIS	200	400-2500	20	Airborne
KSC [98]	AVIRIS	176	400-2500	18	Airborne
Botswana [98]	Hyperion	200	400-2500	30	NASA EO-1 satellite
Chikusei [131]	Headwall Hyperspec-VNIR-C	128	363-1018	2.5	Airborne
Washington DC Mall [132]	HYDICE	191	400-2500	3	Airborne
Cuprite [109]	AVIRIS	224	400-2500	2	Airborne

2.5 Performance Metrics

The quality of enhanced images requires verification for the purpose of evaluation and benchmarking. Verifying the quality of an image with visual inspection is a subjective process that depends on several factors, including screen size and illumination. Therefore, quantitative evaluation is more reliable, in which the quality of the enhanced image (also called estimated image) is assessed by comparing it to the GT image (also called reference or target image).

Performance metrics can be divided into two categories depending on the type

of measured resolution quality; spatial and spectral. The distinction lies in whether the pixel-level comparison is taking place within the same band (spatial) or across several bands (spectral). Examples of spatial metrics include PSNR and SSIM, while examples of spectral metrics include SAM and Erreur Relative Globale Adimensionnelle de Synthèse (ERGAS), which is the relative dimensionless global error.

2.5.1 Peak Signal-to-Noise Ratio

PSNR is one of the most widely used spatial quality metrics, and it is expressed as follows:

$$PSNR(\mathbf{Y}, \widehat{\mathbf{Y}}) = \sum_{k=1}^{B} 10log_{10} \frac{MAX(\mathbf{Y^k})^2}{MSE(\mathbf{Y^k}, \widehat{\mathbf{Y}}\mathbf{^k})} \tag{2.6}$$

$$MSE(\mathbf{Y^k}, \widehat{\mathbf{Y}}\mathbf{^k}) = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} [Y^k_{(i,j)} - \widehat{Y}^k_{(i,j)}]^2 \tag{2.7}$$

where $\mathbf{Y^k}$ is the k^{th} band of GT HSI and $\widehat{\mathbf{Y}}\mathbf{^k}$ is the k^{th} band of the estimated High Resolution (HR) HSI from Low Resolution HSI (LR-HSI, also called source HSI). Both images have height M and width N. $MAX(\mathbf{Y^k})$ refers to the maximum possible value a pixel in the GT HSI can take depending on its radiometric resolution. For instance, the maximum value for images of type 8-bit unsigned integer is 255. Mean Squared Error (MSE) computes the cumulative error between the GT HSI and the estimated HR-HSI, while PSNR computes the maximum possible power of a signal to the power of distortion noise in dB. In ideal cases where both images are identical, PSNR result would be infinite because MSE reaches 0 [133]. Equation 2.6 computes PSNR for each band individually. Throughout this thesis, mean PSNR will be reported, which is the average for all bands. Even though PSNR provides a pixel by pixel comparison, it ignores human visual perception. This can be quantified using SSIM.

2.5.2 Structural Similarity Index Measurement

SSIM [134] allows the inclusion of human visual perception by assessing the errors of three factors; luminance, contrast, and structure. SSIM is expressed as follows:

$$SSIM(\mathbf{Y}, \widehat{\mathbf{Y}}) = \frac{(2\mu_{\mathbf{Y}}\mu_{\widehat{\mathbf{Y}}} + c_1)(2\sigma_{\mathbf{Y}\widehat{\mathbf{Y}}} + c_2)}{(\mu_{\mathbf{Y}}^2 + \mu_{\widehat{\mathbf{Y}}}^2 + c_1)(\sigma_{\mathbf{Y}}^2 + \sigma_{\widehat{\mathbf{Y}}}^2 + c_2)}, \tag{2.8}$$

where $\mu_{\mathbf{Y}}$, $\mu_{\widehat{\mathbf{Y}}}$, $\sigma_{\mathbf{Y}}$, $\sigma_{\widehat{\mathbf{Y}}}$, and $\sigma_{\mathbf{Y}\widehat{\mathbf{Y}}}$ represent local means, standard deviation and cross-covariance for $\mathbf{Y}$ and $\widehat{\mathbf{Y}}$, respectively. The three components of SSIM contribute to the Human Visual System (HVS) as follows:

- Luminance: is a photometric term defined as the intensity of light emitted from a surface area per unit area in a given direction. The SI unit for luminance is candela per square meter ($\mathrm{cd/m^2}$). The luminance component of SSIM is evaluated as follows:

$$l(\mathbf{Y}, \widehat{\mathbf{Y}}) = \frac{2\mu_{\mathbf{Y}}\mu_{\widehat{\mathbf{Y}}} + c_1}{\mu_{\mathbf{Y}}^2 + \mu_{\widehat{\mathbf{Y}}}^2 + c_1}, \tag{2.9}$$

  where $c_1 = (k_1 l)^2$ is added to avoid instability in the denominator, and l is the dynamic range. Typically, $k_1 = 0.01$ [135]. This equation exhibits a qualitative agreement with Weber's law, which is commonly employed to describe light adaptation, also known as luminance masking, within the HVS [134]. Weber's law suggests that the ability to perceive differences between stimuli is not based solely on the absolute magnitude of the stimuli but rather on the relative change in magnitude [136].

- Contrast: is the variation in luminance or color that allows an object to be discerned from its surroundings or another object [137]. Contrast component of SSIM is calculated as follows:

$$c(\mathbf{Y}, \widehat{\mathbf{Y}}) = \frac{2\sigma_{\mathbf{Y}}\sigma_{\widehat{\mathbf{Y}}} + c_2}{\sigma_{\mathbf{Y}}^2 + \sigma_{\widehat{\mathbf{Y}}}^2 + c_2} \tag{2.10}$$

  $c_2 = (k_2 l)^2$ Typically, $k_2 = 0.03$ [135]. Like c_1, c_2 is also added to stabilize the denominator. This measure is consistent with the contrast masking feature of the HVS, which is the phenomenon where the visibility of one element in an image is diminished due to the presence of another element within the same spatial location with similar frequency characteristics [138].

- Structure: refers to the structural information in an image, which is defined as the attributes representing the composition of objects in the scene, irrespective of average luminance and contrast [134]. The structure component of SSIM is calculated as follows:

$$s(\mathbf{Y}, \widehat{\mathbf{Y}}) = \frac{\sigma_{\mathbf{Y}\widehat{\mathbf{Y}}} + c_3}{\sigma_{\mathbf{Y}}\sigma_{\widehat{\mathbf{Y}}} + c_3}, \tag{2.11}$$

  where $c_3 = \frac{c_2}{2}$ [135], and it is added to stabilize the denominator.

The three components are combined, such that:

$$SSIM(\mathbf{Y}, \widehat{\mathbf{Y}}) = l(\mathbf{Y}, \widehat{\mathbf{Y}})^{\alpha} \cdot c(\mathbf{Y}, \widehat{\mathbf{Y}})^{\beta} \cdot s(\mathbf{Y}, \widehat{\mathbf{Y}})^{\gamma} \tag{2.12}$$

The authors of [134] propose that $\alpha = \beta = \gamma = 1$, which leads to the formula in 2.8. SSIM value ranges between 0 and 1, where 0 indicates no similarity and 1 indicates that $\mathbf{Y}$ and $\widehat{\mathbf{Y}}$ are identical. In the special case when $C_1 = C_2 = 0$, SSIM is referred to as Universal Image Quality Index (UIQI) [139], which is considered the predecessor of SSIM. SSIM is preferred over UIQI due to the fact that the latter can lead to unstable results. Equation 2.8 computes SSIM for each band individually. The authors of [134] originally designed SSIM for 2D images, and their recommended approach involves calculating SSIM locally using a sliding window method rather than globally. However, the common convention for computing SSIM in the context of HSI involves calculating SSIM for each band individually and subsequently computing the average across all bands, as seen in various HSI SISR studies [46, 52, 140–143]. Given that SSIM range is limited to 0-1, variations in SSIM across different algorithms may not appear substantial. However, even minute changes, such as 1×10^{-3}, should be taken into account [140, 141].

2.5.3 Spectral Angle Mapper and Cosine Similarity

Verifying the enhanced HSI's spatial quality is insufficient because it is crucial to ensure that its spectral fidelity can be preserved. This can be achieved by measuring the

similarity between the spectra of the GT HSI and the enhanced HSI. SAM [144] offers a quantitative method to measure this similarity, which is expressed as follows:

$$SAM(\mathbf{Y}, \widehat{\mathbf{Y}}) = cos^{-1}\left(\frac{\sum_{i=1}^{B} \mathbf{Y}_i \widehat{\mathbf{Y}}_i}{\sqrt{\sum_{i=1}^{B} \mathbf{Y}^2}\sqrt{\sum_{i=1}^{B} \widehat{\mathbf{Y}}_i^2}}\right). \tag{2.13}$$

SAM value should be as close to 0 as possible. In this thesis, SAM values will be reported in degrees. Cosine Similarity (CS) is an alternative way of reporting spectral similarity, as $CS = cos(SAM)$. Consequently, a perfectly preserved spectral signature should yield CS values of 1.

2.6 Summary

This chapter presented a technical background and theoretical framework for HSI-SISR, aiming to provide a solid foundation for the research discussed in this thesis. It places particular emphasis on the significance of CNNs and their constituent components, such as convolution, pooling, and their counterparts, including TC and upsampling. These elements play a crucial role in constructing various CNN architectures and are vital for the success of HSI-SISR. Considering the unique characteristics of HSI, 3D operations are found to be more suitable for effectively processing HSI data. In this context, the chapter delves into the characteristics of HSI and highlights their relevance in industrial applications. Furthermore, it acknowledges the existence of publicly available HSI datasets in both radiance and reflectance domains that can be utilized for research and benchmarking purposes.

To ensure a thorough evaluation of the enhanced images, quantitative metrics will be utilized, which provide an objective and fair assessment. Among these metrics PSNR, SSIM, and SAM are discussed as reliable indicators for evaluating the quality of the images.

Chapter 3

Hyperspectral Single Image Super Resolution

3.1 Literature Review

3.1.1 Overview

The field of HSI-SR can be broadly divided into two approaches: Fusion and SISR, as illustrated in Table 3.1. Fusion methods involve utilizing an auxiliary MSI to achieve HSI-SR, and they may or may not incorporate prior assumptions about the sensor, such as the Point Spread Function (PSF). Further details about Fusion methods can be found in Appendix A. On the other hand, SISR methods for HSI primarily focus on reconstructing HR-HSI directly from the observed LR-HSI without the need for an auxiliary MSI. Some SISR methods leverage prior assumptions to aid in the reconstruction of HR-HSI. For a GT HR-HSI patch denoted $\mathbf{Y} \in \mathbb{R}^{M \times N \times B}$ and LR-HSI denoted $\mathbf{X} \in \mathbb{R}^{m \times n \times B}$, the mathematical formulation of SISR can be constructed using an observation model, which is expressed as follows:

$$\mathbf{X} = \mathbf{DGY} + \mathcal{E}, \tag{3.1}$$

where $m \ll M$ and $n \ll N$. $\mathbf{D}$ is the downsampling operation, $\mathbf{G}$ is the blurring kernel, and $\mathcal{E}$ is the additive noise. The HR-HSI can be estimated as follows:

Table 3.1: Taxonomy of the main approaches used in the literature to achieve HSI-SR.

HSI-SR	**Fusion**	**Traditional**	**Pansharpening**	CS			[194–201]	
				MRA			[202–211]	
				Variational			[212–218]	
				Hybrid			[219, 220]	
			Method-based	Matrix Factorization			[72, 73, 221–224, 224–252]	
				Tensor			[185, 231, 238, 253, 254, 254–265]	
				Bayesian			[230–232, 232, 266, 266–270]	
		DCNN						[235, 271–283, 283–299]
	SISR	**Traditional**	Interpolation				[178–181]	
			Regularization				[182–186]	
			POCS				[187, 188]	
			SRM				[189–193]	
		DCNN						[46, 52, 145–160, 160–177]

$$\widehat{\mathbf{Y}} = \psi(\mathbf{X}, \varphi) \tag{3.2}$$

where ψ denotes the HSI-SISR model and φ denotes the model's parameters. The objective is to minimize the loss function between $\mathbf{Y}$ and the estimated HR-HSI $\widehat{\mathbf{Y}}$, as follows:

$$\widehat{\varphi} = \text{argmin}\, L(\mathbf{Y}, \widehat{\mathbf{Y}}) \tag{3.3}$$

where $\widehat{\varphi}$ is the updated model's parameters and L is a loss function that determines the error between $\mathbf{Y}$ and $\widehat{\mathbf{Y}}$. It is evident that the choice of loss function is important for the HSI-SISR task. Chapter 5 goes in depth about the various types of loss functions. The most commonly used ones are MSE and Mean Absolute Error (MAE).

The past two decades have witnessed impressive advances in this area of research [300]. In addition to the taxonomy of SISR approaches presented in Table 3.1, it is worth mentioning that SISR approaches that do not assume prior knowledge regarding the degradation kernel in Equation 3.1 are referred to as "blind SISR". The next subsections discuss the advances in HSI-SR by considering two categories; traditional and DCNN approaches, starting from the simple ones and building up to more sophisticated approaches to bridge the gap between the past and the present.

3.1.2 Traditional Methods

The earliest HSI-SR method and the pioneer in this field is the work proposed in [187]. Akgun et al. proposed a system for capturing HSI, and based on the proposed system, they design HSI-SR framework as an inverse problem. Assuming that the degradation kernel is known, Projections Onto Convex Sets (POCS) [301] can be used to estimate the HR-HSI, such that when the estimated HR-HSI is degraded, it will give a result identical to the observed LR-HSI. The proposed POCS gives more accurate results the more additional constraints can be added from prior information or assumptions. POCS can also be used to estimate the PSF of LR images, which aids the estimation

of the HR counterpart, as seen in the work of [188].

Regularization-based methods, also referred to as reconstruction-based methods, reconstruct HR-HSI from LR-HSI in addition to prior assumptions. For instance, in [182], Villa et al. utilize spectral unmixing for SISR-HSI. They extract the endmembers using Vortex Component Analysis (VCA) [302]. Then, they use Fully Constrained Least Squares (FCLS) algorithm to determine the abundance fraction of the endmembers within each pixel. Afterwards, each pixel is divided into sub-pixels according to the required scale factor. The authors assume that each endmember is spatially close to the same family of endmembers in the surrounding pixels. Based on that, they chose Simulated Annealing as a mapping function that minimizes the perimeter of the areas that belong to the same endmember. The authors tested the effectiveness of their method by comparing the classification map of the enhanced HSI to that of the GT. Another reconstruction-based approach was adapted in [183]. The authors presented Maximum a Posteriori-Markov Random Fields (MAP-MRF) based approach. Similar to [182], the first step was to extract the endmembers, and then estimate the abundance maps using FCLS. The reconstruction was performed on the estimated abundance maps using MAP-MRF. The authors considered this approach an improvement to their previous one presented in [184]. Other examples include [303–305].

Tensor-based approaches, while predominantly used in Fusion methods, can also be used for SISR. For example, In [306], Wang et al. argue that HSI can be modeled as a 3D tensor to exploit global (spectral) correlations between HSI bands in addition to local (spatial) correlation among HSI patches. These correlations can be modeled by a non-convex low-rank tensor, which is an optimization problem that can be solved using Local Linear Approximation (LLA) and Alternative Direction Multiplier Method (ADMM). The authors' approach shows superiority against spectral unmixing analysis and various interpolation approaches. Similar examples can be found in [185, 186].

Some SISR approaches rely on Super Resolution Mapping (SRM), a concept that was first introduced by [189]. According to Atkinson, these approaches can be divided into two categories [307]: optimization and learning-based. A dictionary-learning example is presented in [190]. The authors proposed a multi-dictionary based sparse

representation approach, where the proposed feature vector expresses the significant information about spatial dependence. Consequently, multiple distribution dictionaries are learned via sparse representation. The feature vector is then reconstructed by every dictionary. It is also assigned to a class according to reconstruction errors and spectrum distortions. The authors assert that learning-based SRM is robust to noise. Their approach also avoids overfitting problems that can be potentially encountered with neural networks. The recent SRM for HSI-SISR approaches are utilized in conjunction with DCNNs, such as [191–193].

As mentioned in Section 3.1.1, SISR approaches do not require auxiliary MSI, and blind SISR approaches do not impose prior assumptions, especially if the blurring kernel is an unknown function. The simplest blind SISR approach in the literature is interpolation. Interpolation is a term that can be used interchangeably with resampling, in the sense that it involves transforming an image from one coordinate system to another. The accuracy of the interpolation depends on the selection of a proper interpolation kernel. Some of the most common interpolation methods include nearest neighbor, bicubic, and bilinear interpolation [178], and other advanced versions of them [179–181]. Even though interpolation methods are widely used in commercial software, they are not favorable because they introduce artifacts and blurriness, and they are prone to spectral distortions in HSI. Nonetheless, there are several examples in HSI-SR research where bicubic interpolation is used as a benchmark for performance comparison, or as an initial step in the designed approach [50, 308]. The vast majority of SISR approaches nowadays are performed using DCNNs, which have been the predominant approach in HSI SISR from 2017 onward, as will be discussed in the next subsection.

3.1.3 DCNN-based Methods

As discussed in Section 2.1, CNNs consist of automatic feature extractors that omit the requirement of having manual hand-crafted features or human intervention. Instead, CNNs are capable of learning one-to-one mapping between an LR image and its corresponding HR version. In the case of natural images (e.g. RGB images),

many efforts were exerted to improve their spatial resolution via SISR DCNNs. Some of the most prominent methods include Super Resolution Convolutional Neural Network (SRCNN) [309], Very Deep Super Resolution (VDSR) [310,311], Super Resolution Generative Adversarial Network (SRGAN) [312], Enhanced Deep Residual Networks for Single Image Super-Resolution (EDSR) [313], Residual UNet (RUNet) [314], and Autoencoders [315, 316]. Unlike natural images, the spectral fidelity of HSI can be negatively impacted upon spatial enhancement. This is referred to as spectral distortions in this thesis. An example is seen in Figure 3.1. If the spectral fidelity is lost, unmixing the different endmembers will be more challenging, which negatively impacts applications pertaining classification and object detection. Therefore, CNNs developed for HSI-SR must be architected while taking spectral context into consideration.

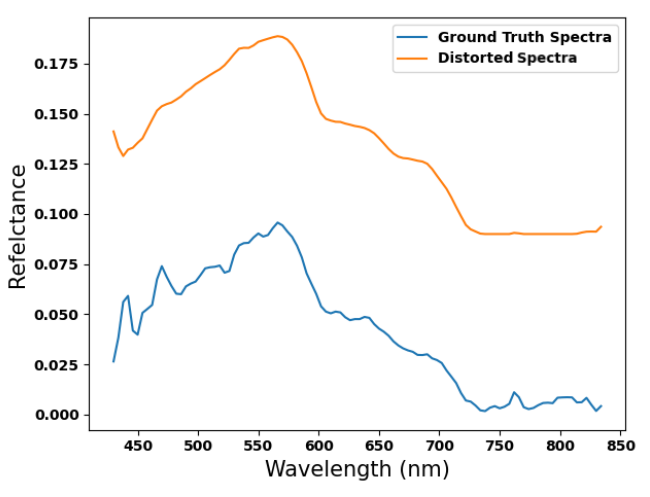


Figure 3.1: An example of spectral distortion that occurred after spatial enhancement.

Some algorithms develop SISR DCNNs while taking inspiration from Fusion methods to minimize spectral distortions. For example, [145] uses transfer learning technique to re-purpose SR DCNN that was originally trained for natural images. Additionally, they utilize Coupled Non-negative Matrix Factorization (CNMF) to capture the spectral relation between spectral bands. The authors compare their results to Fusion methods as well as interpolation methods, and it shows superiority in terms of RMSE, PSNR, SSIM, ERGAS, UIQI, and SAM. Some authors argue that 3D-CNNs capture spectral information better than 2D-CNNs and, thus, they use 3D convolution as their

primary approach for HSI-SR. An example of this can be seen in the work of [146], where the authors used 3D Full CNN (3D-FCNN) to learn both spatial and spectral correlations simultaneously. They further extended their work and improved their algorithm by including one extra convolution layer [52]. Nonetheless, this method is sensor-specific, as it is a way to avoid the necessity of having a large dataset. Therefore, it only works on images acquired by the same sensor. For instance, Pavia Center and Pavia University datasets were acquired by ROSIS sensor, so the algorithm needs to be trained for one of them only. A similar approach was adapted by [147] through using 3D-FCNN with residual connections to enhance spectral and spatial characteristics simultaneously. Another approach that utilizes 3D-FCNN was explored by [46]. The authors decomposed LR-HSI into four groups of wavelet coefficients according to their frequency similarity and, hence, the network is referred to as Frequency Separated 3DCNN (FS-3DCNN). This is done to suppress spectral distortion while maintaining the high-frequency information. The feature cubes are extracted using 3D convolution and the details are reconstructed by 3D deconvolution. The final HR-HSI is obtained by inverse wavelet transformation. This method shows superiority against 3D-FCNN and bicubic interpolation in terms of PSNR, SSIM, and SAM. Another wavelet-based approach was used in [317]. The study presents a method for HSI-SISR using a Multi-scale Wavelet 3DCNN (MW-3D-CNN). Instead of directly reconstructing the HR-HSI, the network predicts the wavelet coefficients of the HR-HSI. The MW-3D-CNN consists of an embedding subnet and a predicting subnet, both built with 3D convolutional layers. The embedding subnet extracts deep spatial-spectral features from the LR-HSI and represents it as feature cubes. These cubes are then fed to the predicting subnet, which has multiple output branches corresponding to wavelet sub-bands. Each branch predicts the wavelet coefficients of the HR-HSI. By applying inverse wavelet transform to the predicted coefficients, the HR-HSI can be obtained.

Some researchers use attention mechanism to amplify the important features extracted from 3D convolution [318]. One example is the study presented in [319], which proposes a Deep Learning (DL) technique called Dual Self-Attention Swin Transformer SR (DSSTSR). This network utilizes the Swin transformer to capture global and local

spatial features and learns spectral sequence information from adjacent bands of the HSI. Additionally, it incorporates an image de-noising module using wavelet transformation to reduce the impact of stripe noise. Experimental results on publicly available datasets demonstrate that DSSTSR outperforms other state-of-the-art HSI-SR methods based on PSNR, SSIM, and SAM [319]. Another example of utilizing attention mechanism is utilized in [148], where the authors devised a 3D Attention-based SRGAN (3DASRGAN) that utilizes SAM as a part of the loss function to guarantee the minimization of spectral distortions. 3DASRGAN prevails over bicubic interpolation, 3D-FCNN, and the original SRGAN. The loss function is an important aspect of CNNs that directly affects spectral fidelity. Other examples of incorporating SAM within the loss function can be seen in Grouped Deep Recursive Residual Network (GDRRN) devised by [149].

In [150], the authors argued that 3D-CNNs are not the optimal choice because the spectral bands of HSI are highly redundant. Therefore, they propose 1D-2D spatial-spectral CNN instead. The 2D path of the network extracts spatial features, and the 1D path of the network utilizes the high similarity between HSI bands. Experimental results on Pavia dataset prove that this network performs better than 3D-FCNN. However, according to [151], the dual 1D-2D CNN does not explore spatial features deeply enough. Therefore, the authors devise a mixed 2D/3D convolution, which they refer to as MCNet. The 2D units help the network to learn the hierarchical features more adaptively, and separable 3D convolution is utilized in the 3D units in order to optimize memory usage while extracting spatial and spectral features. Experimental results on CAVE and Harvard datasets show superiority against bicubic interpolation, 3D-FCNN, and the dual 1D-2D CNN. Using a similar principle, [152] also utilized a mixture of 2D and 3D convolution, and they added Feature Context Fusion (FCF) to combine the features from each band with the preceding one to simplify the network structure and enhance the performance. Section 2.1 discusses that DCNNs can be configured in various topology, and layers can be connected in different strategies. Making efficient use of these strategies highly boosts HSI-SISR performance. This can be observed in several studies in the literature [153–160, 160–175]. Implicit Neural Representations

(INRs) was used in [320] as an effective way of utilizing spectral-spatial features. Another study that utilizes spectral-spatial correlations is presented in [321]. The authors devise a Spatial-Spectral Feature Extraction Network (SSFEN) that consists of three components: a spatial-spectral mapping network, a spatial reconstruction network, and a spatial-spectral fusing network. The authors in [322] address the shortcomings of 2D and 3D networks by devising a Diffused CNN that incorporates spectral convolutions in the Enhanced Convolutional Neural (ECN) block and introduces a series of spectral convolutions in the residual network. Additionally, Histogram of Oriented Gradient (HOG) and Local Binary Pattern (LBP) are utilized to preserve shape and texture information. A feature fusion strategy is employed to effectively combine shallow and deep features, and an image enhancement module is developed to improve the SR image quality. In [323] the authors propose an asymmetric dual-direction quasi-recursive network to explore the features among neighbor and non-neighbor bands separately using forward and backward units. The forward units leverage the high similarity among neighbor bands to thoroughly exploit spatial-spectral features, while the backward units focus on extracting spatial features in non-neighbor bands due to their discontinuous spectra. A global feature context fusion module combines the information from global non-neighbor context and neighbor bands, enhancing information completeness and complementarity. Experimental results on natural and remote sensing HSI datasets demonstrate that the proposed network surpasses state-of-the-art methods in terms of reconstruction quality, noise suppression, and memory efficiency.

All of the aforementioned HSI-SISR approaches suffer from data scarcity. That is, HSI datasets exist as a single scene, which is insufficient to train DCNNs. In SISR approaches, an HSI scene is often divided into patches to train a DCNN. For instance, dividing Botswana dataset into patches of 64×64 yields 92 patches of which 70% is typically used for training and the remaining for testing. Even though this is the most common approach, data scarcity is still a challenge that needs to be overcome. Additionally, the aforementioned HSI-SISR approaches perform well on homogeneous datasets that are captured by the same sensors, and cannot be generalized across different sensors. In 2021, [176] addressed the problem of HSI's high dimensionality and

scarcity of training samples which result in undesirable behaviors, such as overfitting. Their work is built based on the concept of high correlation existence between HSI and their corresponding RGB. Thus, RGB and HSI can be trained jointly, such that RGB-SISR can provide additional supervision. This approach minimizes the amount of HSI dataset required for training, and allows its applicability on heterogeneous datasets. Additionally, the authors also devise a novel Data Augmentation algorithm called Spectral Mixup [324] to increase the amount of training samples. The most recent version of this approach was published by the same authors in 2022 [177]. This method outperforms [150], [151], [153], and [149]. As per our knowledge, this is the only research work that addresses the problem of data scarcity for HSI-SISR.

One of the main drawbacks of the aforementioned algorithms is limiting their experiments to scaling factors of $\times 2$ and $\times 4$. In [151] and [152], the scaling factor goes up to $\times 8$ at most, and HSI-SR requires higher scaling factors in order to be put into practical use. Furthermore, SISR techniques have no unsupervised approaches associated with DCNNs, as the only semi-supervised approach that tackles data scarcity problem is the one presented in [176, 177].

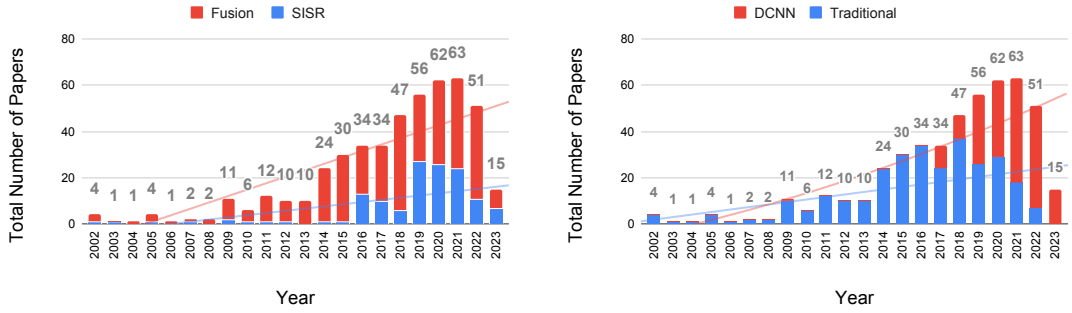
3.1.4 Meta-analysis

Table 3.2: Common disadvantages between HSI-SR methods.

Disadvantage			Spectral distortion	Computational cost	Impractical assumptions	Data scarcity	Require prior knowledge	Scaling factor below 8	Aliasing effects
Fusion	Pansharpening		✓	✓			✓		✓
	Method -based	MF	✓	✓	✓		✓		
		Tensor	✓	✓			✓		
		Bayesian					✓		
	DCNN		✓	✓	✓		✓		
SISR	POCS		✓	✓			✓		
	Regularization		✓					✓	
	SRM		✓					✓	
	Interpolation		✓					✓	✓
	DCNN		✓	✓		✓		✓	

The overall drawbacks of all the studies discussed in both Fusion (Appendix A) and SISR (Sections 3.1.2 and 3.1.3) are summarized in Table 3.2. It can be concluded that the most important drawbacks to consider while designing a new algorithm are spectral distortions, computational cost, scaling factor, in addition to data scarcity, especially for SISR methods.

A web scraping tool was developed using Python programming language to retrieve all the relevant research papers related to HSI-SR. The tool was used to retrieve articles from IEEE Xplore Digital Library related to the following keywords: *hyperspectral super resolution, hyperspectral spatial enhancement, hyperspectral reconstruction, hyperspectral Fusion, hyperspectral SISR*. The tool retrieved various information about the research papers, including title, type of publication (e.g. conference or journal), authors, keywords, DOI, and publication year. The results were later verified with visual inspection and more entries were added manually from other sources, including, but not limited to, SPIE Remote Sensing conference proceedings and MDPI Remote Sensing journal.



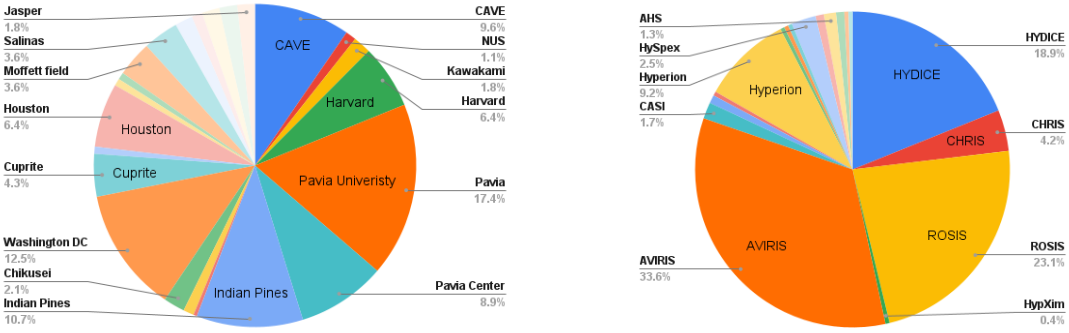
(a) Fusion Vs. SISR

(b) Traditional Vs. DCNN

Figure 3.2: Comparison of the total number of papers between (a) Fusion and SISR, and (b) Traditional Vs. DCNN methods between 2002-2023.

A visual summary of all the retrieved results can be seen in Figures 3.2a and 3.2b. From the yearly total numbers of papers in both figures, it can be observed that the interest in HSI-SR has been increasing over the years. It is expected that the number of published papers will increase further by the end of 2023. Furthermore, although HSI-SISR studies have been increasing for the past 2 decades, Figure 3.2a shows that

in every year, there is a wide gap between the number of Fusion and SISR studies. In fact, the decrease in the total number of papers between 2021 and 2022 is caused by the decrease in SISR papers, as the number of Fusion papers continued to increase. This can be attributed to the data scarcity disadvantage that HSI-SISR methods suffer from. Also, Figure 3.2b shows that the interest in DCNNs for HSI-SR has been gradually increasing since 2017, and the number of publications in DCNN HSI-SR has been exceeding the number of that in traditional methods for the past 4-5 years. Since SISR methods utilize DCNNs much more than traditional approaches, it seems that there is correlation between the rise of DCNN methods and SISR methods.



(a) Datasets

(b) Sensors for dataset collection

(c) Evaluation metrics

Figure 3.3: Most frequently used datasets, sensors, and evaluation metrics in HSI-SR.

According to Figure 3.3a, the five most used datasets are Pavia University, Washington DC Mall, CAVE, Pavia Center, and Indian Pines. As for the most used HS sensor, Figure 3.3b shows that AVIRIS prevails over other sensors by a large margin, followed by ROSIS, HYDICE, and Hyperion. Even though Pavia University is the

most used dataset, AVIRIS remains more widely used than ROSIS because the total number of datasets collected using AVIRIS is more than the number of those collected using ROSIS. Additionally, several studies mention using AVIRIS for dataset collection without specifying any of the standard datasets mentioned in this thesis. This also explains why Hyperion sensor seems to be widely used, but not Botswana dataset. As for the most used metric for quality assessment, Figure 3.3c shows that SAM is used at least 22% of the time, which is an expected result, since it is the simplest formula that gives indication of change in spectral shape. Other metrics with high percentage of usage include PSNR, RMSE, ERGAS, and SSIM. Based on the statistics illustrated in Figure 3.3, the sensors that will be used for testing and benchmarking in this thesis are AVIRIS, ROSIS and HYDICE. Consequently, the corresponding datasets that will be used are Pavia University and Pavia Center from ROSIS, Washington DC Mall from HYDICE, and Indian Pines from AVIRIS. Additionally, the metrics that will be used to evaluate various algorithms in this thesis are PSNR, SSIM, SAM. Despite the frequent utilization of RMSE, several studies state that it is inaccurate and does not provide meaningful insight in terms of error description between images as a numerical value. However, it will be used as an error map to improve the visualization between various methods' outputs for the sake of comparison.

Figure 3.4: Word cloud of the most frequently used terminologies in HSI-SR research.

The word cloud seen in Figure 3.4 gives a visual indication of the mostly used terminologies in HSI-SR research literature. Consistent with Figure 3.2a, the vast majority of the terminologies are related to Fusion methods and their extensions, which further asserts the fact that Fusion methods are currently the center of attention in this research area. Finally, to give a summary of the HSI-SR timeline, Figure 3.5 shows

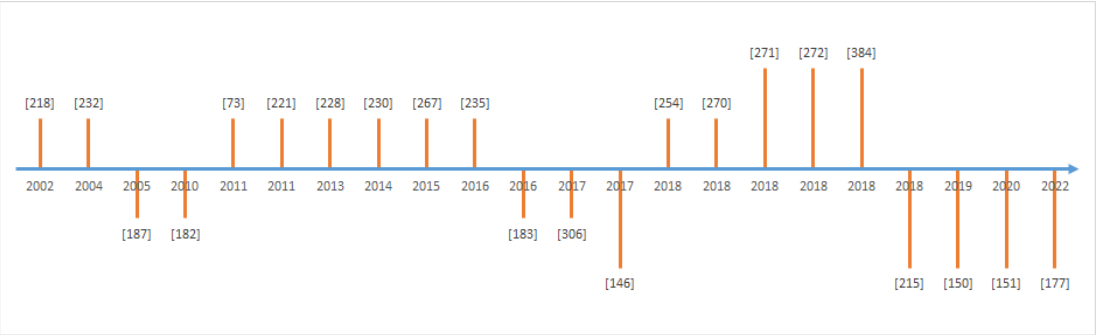
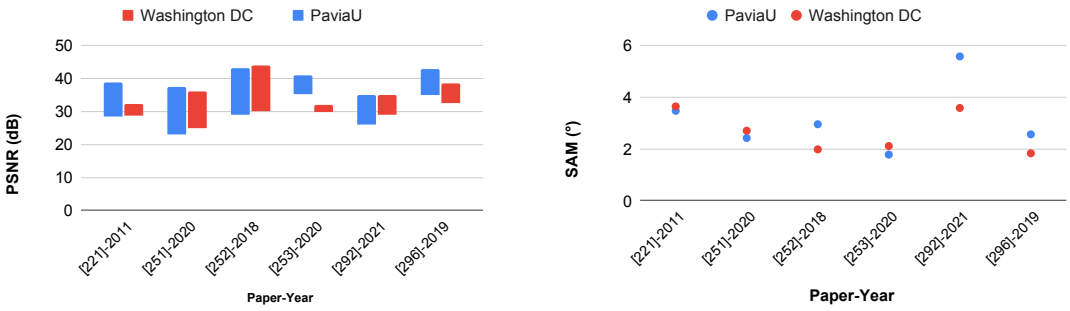


Figure 3.5: A timeline that shows the evolution of HSI-SR techniques throughout the past two decades by highlighting the most representative methods. An upward direction indicates a Fusion method, while a downward direction indicates an SISR method. Short bars indicate a traditional method, while long bars indicate a DCNN method.

the evolution of HSI-SR techniques throughout the years by highlighting the most prominent approaches. This timeline further asserts the fact that DCNNs have been of central interest since 2017. It is worth noting that publications in traditional HSI-SISR techniques stopped around 2016, while the field of traditional Fusion methods remains active.



(a) PSNR (b) SAM

Figure 3.6: Comparison between (a) PSNR and (b) SAM values across different studies on Pavia University and Washington DC Mall datasets.

It is worth observing the quantitative results obtained by the recent studies. To narrow it down, the results are studied for the two most used datasets; Pavia University and Washington DC Mall, and the two most used metrics; PSNR and SAM. Figure 3.6a lists the span of PSNR for Pavia University and Washington DC Mall datasets obtained

by the following studies [221,251–253,292,296]. The PSNR overall ranges between 26dB and 43 dB approximately. A large span of PSNR shows that the obtained result highly depends on certain conditions, or the dataset itself. for instance, study [252] is an example of such large span.

Figure 3.6b shows comparisons between SAM results for the same studies with the same datasets. The worst overall value is 5.57°, while the best is 1.78°. Similar to PSNR case, SAM results can also show inconsistencies across different datasets. For instance, study [292] shows vastly different SAM results for Pavia University and Washington DC Mall datasets.

3.2 Data Pre-processing and Problem Formulation

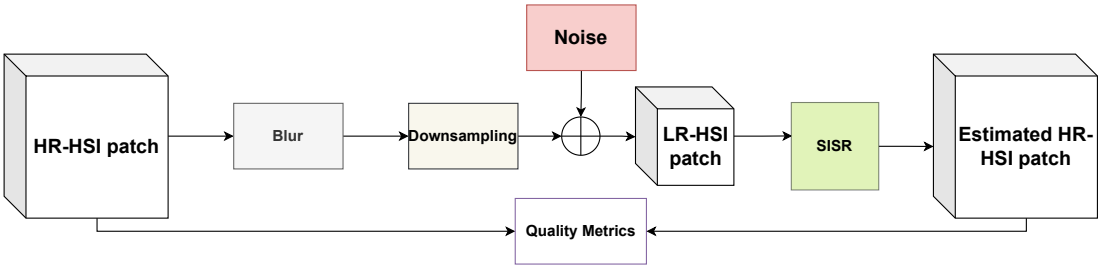


Figure 3.7: Basic framework for HSI-SISR.

The publicly available remote sensing HSI datasets consist of a single scene, which is not enough to train a DCNN. In order to generate enough data for that purpose, the scene is sequentially divided into non-overlapping patches of size 64×64, where each patch is considered as a GT training sample. In the case of Indian Pines and Washington DC Mall datasets, a band-wise normalization procedure is applied to all patches, ensuring that the pixel values fall within the range of 0-1. In contrast, Pavia University and Pavia Center datasets comprise reflectance values, and as such, the values naturally exist within the 0-1 range. Each patch is downsampled according to Equation 3.1. LR-HSI is commonly generated synthetically by applying Gaussian blur and using bicubic interpolation as a downsampling operation [46]. In this study, nearest neighbor interpolation is used to generate LR-HSI instead of bicubic interpolation. That

is, for a scale factor $\times 2$, every other pixel is dropped from the image. For a scale factor $\times 4$, for every set of five pixels, the first pixel is preserved and the remaining four are dropped.

HR-HSI can be estimated by minimizing the error according to Equation 3.3. This is a highly non-linear problem due to the complexity of the HSI cube and the amount of missing information from downsampling, therefore, DCNNs are an ideal framework to obtain a one-to-one mapping between $\mathbf{X}$ and $\mathbf{Y}$.

3.3 Experiment Design

To effectively address the objectives and challenges outlined in Sections 1.4 and 1.2, it is imperative to meticulously design a thorough experimental procedure.

First and foremost, an LR-HSI must be generated according to the scheme in Figure 3.7 after patchifying the scene. Subsequently, the development and testing of 3D-CNNs will be carried out, utilizing the most commonly employed sensors and datasets as indicated by the meta-analysis presented in Section 3.1.4. This exploration of HSI-SISR will encompass various DCNN architectures, encompassing 2D, 3D, and complex variations. During the architecture design phase, all other hyperparameters will be held constant to ensure an equitable comparison. Additionally, the investigation of loss functions will be undertaken, serving as a tool to enhance both spectral and spatial resolution in a hybrid manner. Similar to the architecture design scheme, the hyperparameters for loss functions will remain fixed for the sake of fairness. Lastly, Data Augmentation techniques will be explored within the context of HSI-SISR to address the challenge of data scarcity. Once again, during the study of Data Augmentation techniques, the remaining hyperparameters will be kept consistent to facilitate an equitable comparison.

All the experiments are conducted and tested using Python Keras library with Tensorflow-gpu 2.8.0 as backend and under the same environment to ensure fairness of comparison. By default, this library favors the float64 (double precision) data type. However, to optimize memory and resource usage, float32 (single precision) is employed, as it provides a sufficiently wide bit range to accommodate the values in the datasets without harming the performance [325]. The experiments are trained on NVIDIA

Quadro P6000-24GB Dual GPU and Dual Intel(R) Xeon(R) Gold 6128 at 3.4 GHz,12 core Processor. Furthermore, cloud services and/or GPU are used as needed to speed up the computation process. However, when comparing the processing time of algorithms, consistency is ensured by conducting all the experiments using the same processing unit.

3.4 Summary

HSI-SR approaches can be broadly classified into Fusion and SISR. A thorough analysis of research studies published between 2002 and 2023 indicates a growing interest in this field. The number of papers focusing on Fusion approaches has notably increased, whereas the research on SISR approaches is relatively limited, possibly due to data scarcity challenges. Moreover, there is a growing trend towards utilizing DCNNs in contrast to traditional methods. For this thesis, the Pavia University, Pavia Center, Washington DC Mall, and Indian Pines datasets, which are widely used in the literature, will be employed. The evaluation metrics employed in this study include PSNR, SSIM, and SAM, providing a thorough assessment of the performance. Furthermore, the visualization of error maps through RMSE will enable a better understanding of the quality of the predicted HR-HSI. To ensure fairness in the comparison, all experiments will be conducted under the same experimental setup.

Chapter 4

Extension of 2D-CNNs to 3D-CNNs

4.1 Introduction

Architecting a DCNN for HSI-SISR includes taking several factors into consideration:

- The first one is the fact that convolution and pooling layers reduce the size of the image and in turn cause feature loss that may contain important information [326]. Thus, this effect must be mitigated.
- Second, there are certain layers that are counterproductive to the goal of SISR and should be avoided. Examples of such layers include BN and pixel shuffling. In the case of pixel shuffling, its usage can have a detrimental effect on HSI-SISR by introducing spectral distortions [40].
- Third, residual connections boost the performance of SISR DCNNs, and it is important to incorporate them into the architecture.

Typically, DCNNs that are designed for SISR tasks consist of a feature extraction stage followed by a feature mapping stage, as demonstrated by the architectures presented in this chapter. The objective of this chapter is to demonstrate that extending the conventional 2D-CNNs used for SISR to 3D space can enhance the performance of

HSI-SISR tasks with minimal spectral distortions and improved computational efficiency and/or reduced time complexity compared to existing state-of-the-art methods. The chapter will also explore the relationship between network depth and dataset size. The hypothesis is that small datasets require shallow networks, while large datasets necessitate deeper networks, as shallow networks struggle to effectively learn from extensive datasets, while deep networks may face issues such as overfitting and vanishing gradients when trained on smaller datasets.

To achieve this goal, three novel 3D-CNN architectures, namely shallow, medium-depth, and deep networks, are developed, tested, and evaluated. Each architecture will be thoroughly explained and analyzed by assessing their performance on three datasets of varying sizes; small, medium, and large.

4.2 Experimental Setup

This chapter will conduct individual experiments on each developed 3D-CNN to compare it with its corresponding 2D version. Later in Section 4.6, the performance of all 3D-CNNs will be compared against one another. The objective of this chapter is to conduct a comparative analysis of various architectures. To accurately assess the impact of architecture design, it is imperative to maintain a fixed set of hyperparameters. This approach not only allows isolating the architecture's effect but also guarantees a level of fair and unbiased comparison across all the architectures. These parameters, which are commonly employed for training and testing CNNs, are listed in Table 4.1.

Table 4.1: Training parameters for all the experiments conducted in Chapter 4.

Training parameter	Value		
Optimization function	Adam [327]		
Loss function	MSE		
Epochs	Early Stopping (max 1000)		
Learning rate	Scheduler: [10^{-3}, 10^{-4}, 10^{-5}]		
Dataset	Pavia University	Washington DC Mall	Indian Pines
Training	36	64	2
Validation	4	8	1
Testing	5	8	1

Particularly, MSE and Adaptive Moment Estimation (Adam) have been frequently deployed for HSI-SISR deep learning algorithms, especially when comparing different architectures [40, 328]. Thus, they will be used as a base of comparison between the architectures. Additionally, Table 4.1 provides information on the number of training, validation, and testing images obtained after patchifying the Pavia University, Washington DC Mall, and Indian Pines datasets. Typically, researchers adopt a data split, allocating either 70% for training and reserving the rest for testing and validation, or 80% for training with the remainder reserved for testing and validation [329]. Given the limited supply of available patches, opting for the 80% division appears to be a more equitable choice, ensuring sufficient data for effective training. Indian Pines dataset is the only exception, since the number of patches is insufficient for 80/20% split. Indian Pines patches is split in 50% instead.

An early stopping strategy is implemented to determine the optimal number of epochs. The decision to halt training is contingent on the improvement observed in the selected loss function [330]. If there is no improvement in the validation data, or if the improvement is less than 10^{-5}, for a continuous span of 5 epochs, the training

process concludes [330]. Through experimental observations, it has been noted that the networks typically cease to exhibit improvement around the 500^{th} epoch. Consequently, the training plots seen in this chapter will be from epoch 1 to 500.

As for the learning rate, for float32 data values between 0-1, it is recommended that the learning rate should be less than one and higher than 10^{-6} [331]. Typically, the choice of a learning rate involves picking values approximately on a logarithmic scale within the set $[10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}]$ [332]. The default learning rate for Adam is 10^{-3} [333]. A scheduler is utilized to decide whether to lower the learning rate if no improvement is observed for 3 consecutive epochs, or if the improvement is less than 10^{-5}.

For each network presented in this chapter, a thorough evaluation strategy is employed, involving five experiments. In each of these experiments, the dataset undergoes shuffling, guaranteeing that the patches used for training, validation, and testing are distinct each time. The reported results and plots in this study are a reflection of the averages derived from all five experiments, along with their respective standard deviations.

4.3 3D-SRCNN (Shallow)

4.3.1 Architecture

As mentioned in Section 3.1, the simplest DCNN to execute SISR for MSI and grayscale images is 2D-SRCNN [309]. This network set the standard for SISR, making it a good starting point to explore the spatial improvement of HSI. The most straightforward approach is to extend this network to 3D. 3D-SRCNN consists of three main parts:

1. Patch extraction: which extracts features from $\mathbf{X}$, and it is represented by Equation 2.3.

2. Non-linear mapping: which increases the resolution of the extracted features by utilizing TC, denoted $Conv'$, as visualized in Figure 2.4b.

3. Reconstruction: which constructs the final $\widehat{\mathbf{Y}}$.

The architecture of 3D-SRCNN is illustrated in Figure 4.1. the overall equation of 3D-SRCNN that maps $\mathbf{X}$ to $\widehat{\mathbf{Y}}$ is described as follows:

$$\widehat{\mathbf{Y}} = SRCNN(\mathbf{X}) = Conv(Conv'(Conv(\mathbf{X}))) \tag{4.1}$$

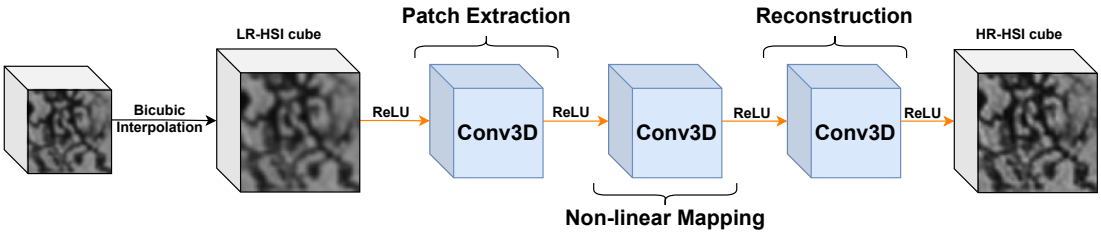


Figure 4.1: Overall architecture of 3D-SRCNN.

2D-SRCNN originally has filters of sizes (9,9), (1,1), and (5,5) [309]. The extended 3D version has filters of sizes (9,9,9), (1,1,1), and (5,5,5). These versions of SRCNN will be referred to as 2D-SRCNN915 and 3D-SRCNN915 throughout the thesis. At each layer, the input is padded in such a way that its height and width do not diminish. The size of the padded image $N' = N + 2p$ after convolution must be equal to the original size N before convolution. Thus, with reference to Equation 2.2, padding p is calculated as follows:

$$\begin{aligned} N' &= N \\ N' - M + 1 &= N \\ N + (2 \times p) - M + 1 &= N \\ p &= \frac{M-1}{2} \end{aligned} \tag{4.2}$$

This can negatively affect the overall quality of the output by causing artifacts to appear around the borders of the image. The intensity of those artifacts increases as the padding size increases. Therefore, modified architectures of both 2D- and 3D-SRCNN are created to reduce padding effects and compare their performance against the original ones [50]. The new filter sizes are all modified to (3,3) in case of 2D, and (3,3,3) in case of 3D. The new modified networks will be referred to as 2D-SRCNN333 and 3D-SRCNN333 throughout the thesis. A comparison between filter sizes of all the

layers is provided in Table 4.2.

Table 4.2: Comparison between the original and modified 2D- and 3D-SRCNN architectures in terms of filter sizes.

Layer #	2D-SRCNN		3D-SRCNN	
	Original (2D-SRCNN915)	Modified (2D-SRCNN333)	Original (3D-SRCNN915)	Modified (3D-SRCNN333)
1	(9,9)	(3,3)	(9,9,9)	(3,3,3)
2	(1,1)	(3,3)	(1,1,1)	(3,3,3)
3	(5,5)	(3,3)	(5,5,5)	(3,3,3)

4.3.2 Results

Figures 4.2 - 4.10 summarize the training results for Pavia University, Washington DC Mall, and Indian Pines datasets. The figures compare the performance between 2D and 3D versions of SRCNN915 and SRCNN333 across all datasets for scale factors $\times 2$ and $\times 4$ in terms of PSNR, SSIM, and CS. In all the figures, the solid line represents the mean outcome of the corresponding network during training, while the dashed line shows the same for validation, and the shaded area represents the uncertainty.

In the case of the Pavia University dataset, the training and validation PSNR for a scale factor of $\times 2$ is depicted in Figure 4.2a. It is evident that the 3D-SRCNN333 network converges more rapidly than the other networks and achieves higher PSNR values compared to all of them. Similar observations can be made for the $\times 4$ scale factor, as shown in Figure 4.2b. Even though the difference between 3D-SRCNN333 and 3D-SRCNN915 appears small, 3D-SRCNN333 shows less standard deviation and, thus, better performance stability across five runs compared to 3D-SRCNN915. Moreover, in both scale factors, the 2D-SRCNN333 consistently outperforms the 2D-SRCNN915 network, and the PSNR of 3D-SRCNN333 is higher than that of 3D-SRCNN915. Consequently, the descending order of performance based on PSNR for both scale factors is as follows: 3D-SRCNN333, 3D-SRCNN915, 2D-SRCNN333, and finally 2D-SRCNN915.

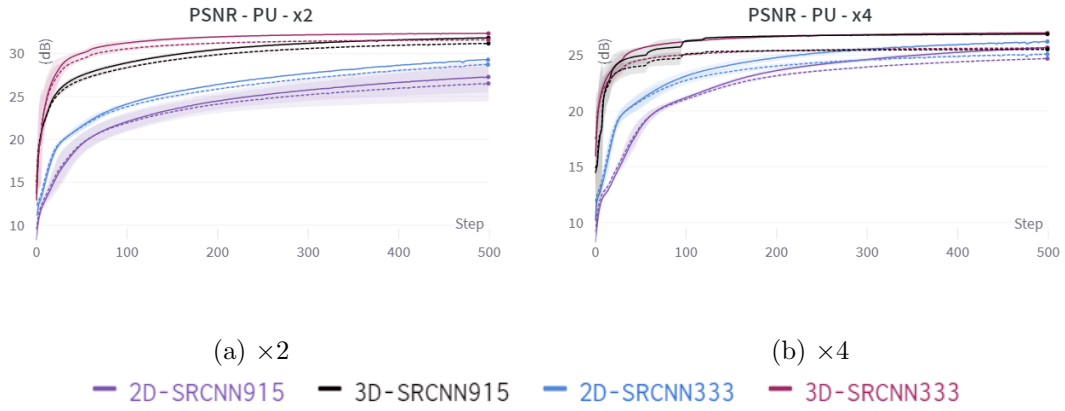


Figure 4.2: Plots of PSNR progression of training 2D- and 3D-SRCNN915 as well as 2D- and 3D-SRCNN333 using Pavia University (PU) dataset. The solid line shows training results, while the dashed line shows validation results. The solid line shows the mean of the training results, while the dashed line shows the mean of the validation results. The shaded area represents the standard deviation of all five experiments.

Similar conclusions can be drawn for the SSIM when observing Figures 4.3a and 4.3b, with the same order of performance as PSNR. Therefore, it can be confirmed that 3D-SRCNN333 effectively improves the spatial quality of the Pavia University dataset.

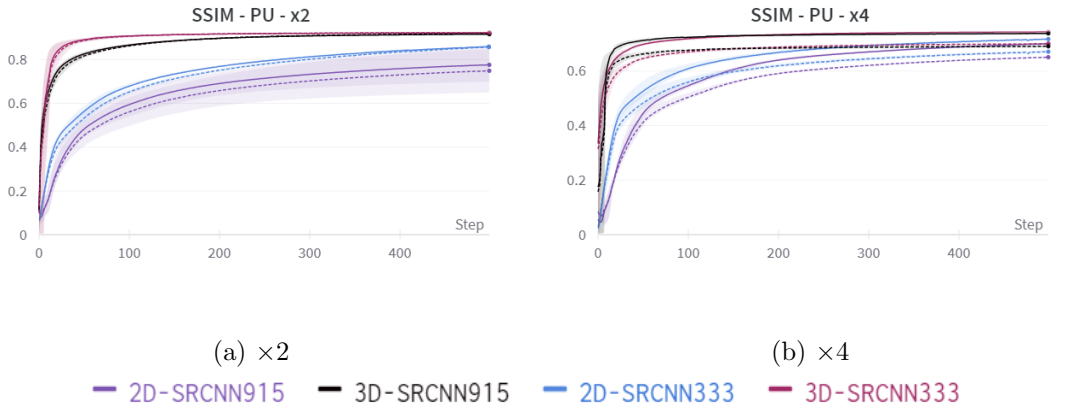


Figure 4.3: Plots of SSIM progression of training 2D- and 3D-SRCNN915 as well as 2D- and 3D-SRCNN333 using Pavia University (PU) dataset. The solid line shows the mean of the training results, while the dashed line shows the mean of the validation results. The shaded area represents the standard deviation of all five experiments.

Regarding spectral quality, Figures 4.4a and 4.4b illustrate the CS between epochs

300 and 500 for better visibility. The figures demonstrate that 3D-SRCNN333 achieves the highest CS among all networks, and the order of network performance is consistent with that of PSNR and SSIM. This provides evidence that SRCNN333 consistently outperforms SRCNN915 and that both spatial and spectral metrics agree that 3D-SRCNN333 delivers better overall performance.

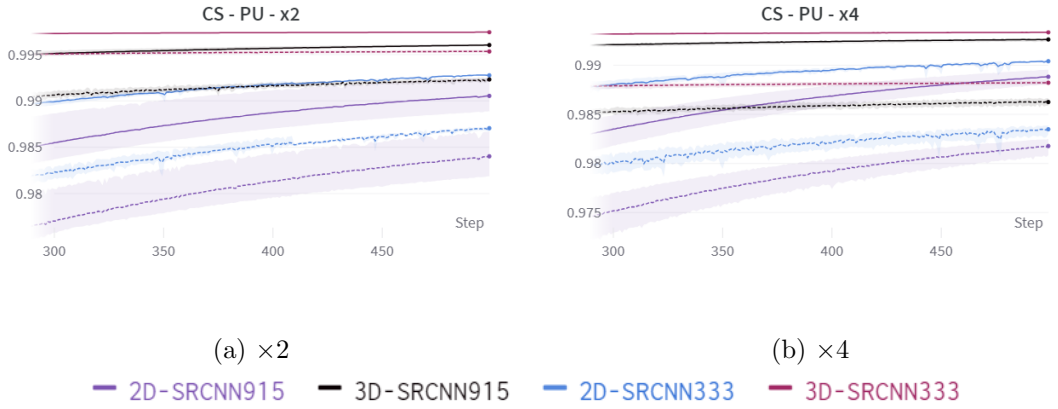


Figure 4.4: Plots of CS progression of training 2D- and 3D-SRCNN915 as well as 2D- and 3D-SRCNN333 using Pavia University (PU) dataset. The solid line shows the mean of the training results, while the dashed line shows the mean of the validation results. The shaded area represents the standard deviation of all five experiments. The plots are shown between epochs 300-500 for better visibility.

Figures 4.5 - 4.7 illustrate the training and validation progression for the Washington DC Mall dataset. It is important to note that Washington DC Mall is a larger dataset compared to Pavia University and may exhibit distinct behavior. This can be observed in Figures 4.5a and 4.5b for PSNR, as well as Figures 4.6a and 4.6b for SSIM. In these figures, it can be seen that 3D-SRCNN915 achieves a performance that is relatively close to that of 3D-SRCNN333, although the latter converges at a faster rate. It is worth mentioning that 2D-SRCNN333 exhibits instability during training and validation for all metrics. The improvement in training and validation is not steady, as the network performance goes through various spikes of ups and downs. This instability can be attributed to the larger size of the Washington DC Mall dataset and the inefficiency of a 2D network in processing hypercubes, although it is not observed in 2D-SRCNN915. Nonetheless, 2D-SRCNN333 still outperforms 2D-SRCNN915.

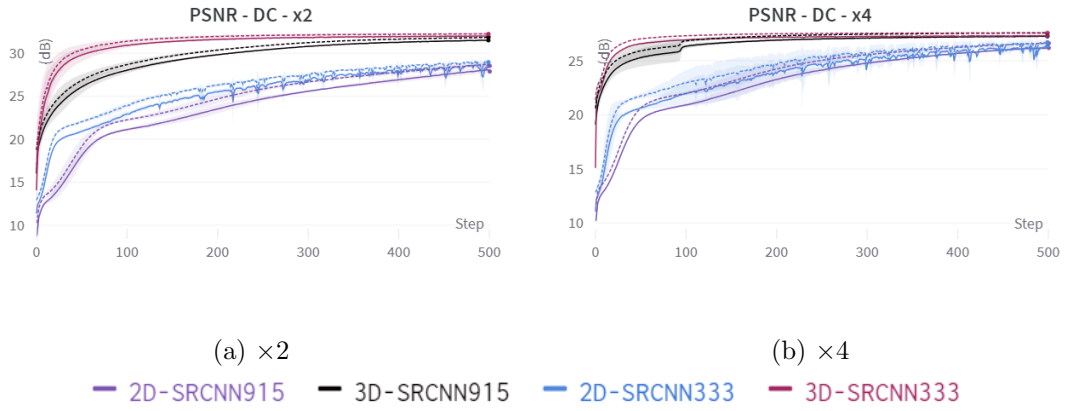


Figure 4.5: Plots of PSNR progression of training 2D- and 3D-SRCNN915 as well as 2D- and 3D-SRCNN333 using Washington DC Mall (DC) dataset. The solid line shows training results, while the dashed line shows validation results. The solid line shows the mean of the training results, while the dashed line shows the mean of the validation results. The shaded area represents the standard deviation of all five experiments.

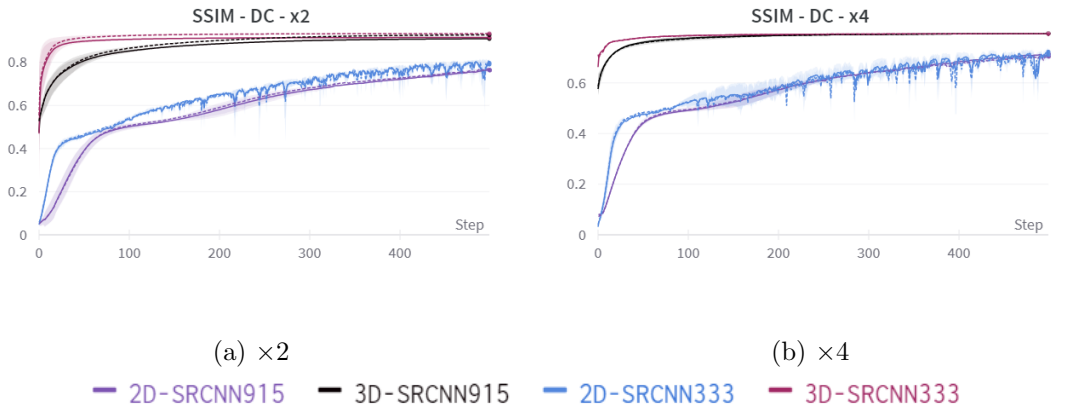


Figure 4.6: Plots of SSIM progression of training 2D- and 3D-SRCNN915 as well as 2D- and 3D-SRCNN333 using Washington DC Mall (DC) dataset. The solid line shows the mean of the training results, while the dashed line shows the mean of the validation results. The shaded area represents the standard deviation of all five experiments.

Overall, 3D-SRCNN333 achieves the best performance across all metrics, and the order of performance is similar to that observed in the Pavia University dataset: 3D-SRCNN333, 3D-SRCNN915, 2D-SRCNN333, and 2D-SRCNN915. Another notable observation regarding the Washington DC Mall dataset is the higher validation perfor-

mance compared to the training performance, especially in terms of CS plots seen in Figure 4.7, which is uncommon. Before drawing definitive conclusions, it is necessary to observe the behavior of the dataset with medium-depth and deep networks.

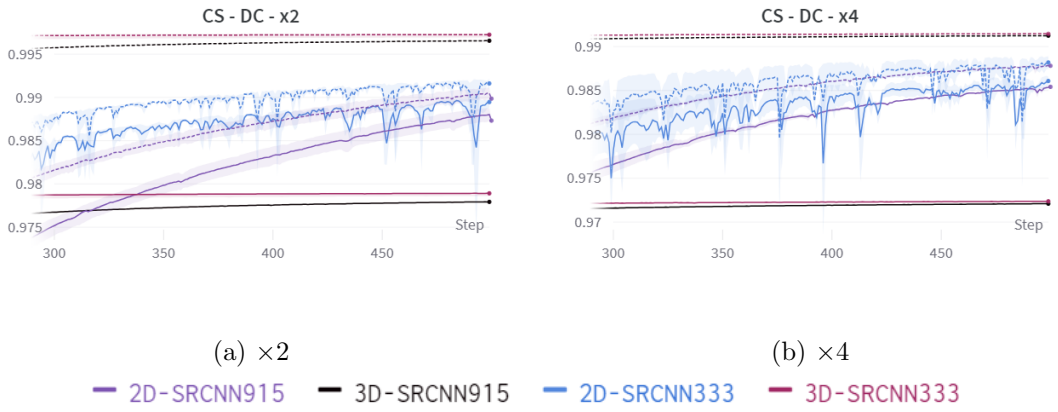
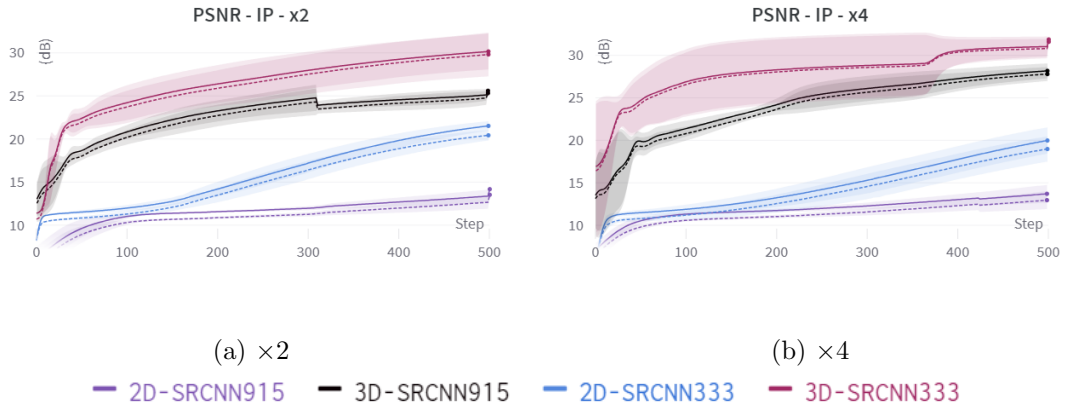


Figure 4.7: Plots of CS progression of training 2D- and 3D-SRCNN915 as well as 2D- and 3D-SRCNN333 using Washington DC Mall (DC) dataset. The solid line shows the mean of the training results, while the dashed line shows the mean of the validation results. The shaded area represents the standard deviation of all five experiments. The plots are shown between epochs 300-500 for better visibility.

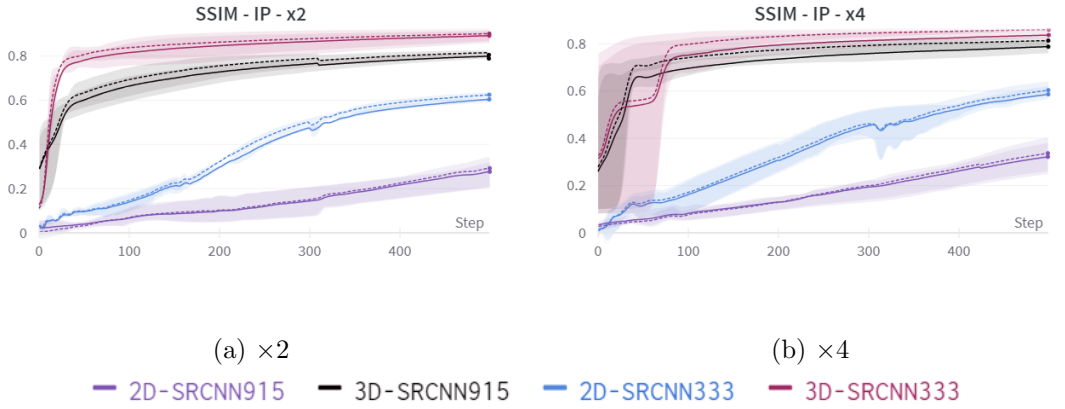
The PSNR training progression of the Indian Pines dataset, as shown in Figure 4.8, present an intriguing scenario due to the dataset's small size. Some outcomes for this dataset deviate from those of Pavia University and Washington DC Mall, particularly the observation that the performance for the ×4 scale factor is better than that for the ×2 scale factor. This is unusual because the former has a larger amount of missing information compared to the latter, and one would expect the ×2 results to be superior. Nevertheless, it is worth noting that 3D-SRCNN still outperforms 2D-SRCNN in both its original and modified versions, indicating that 3D-SRCNN learns effectively despite the extremely limited amount of available data. Furthermore, the superiority of 3D-SRCNN333 over 3D-SRCNN915 reinforces the significance of the modified kernel sizes in enhancing the overall performance of the network, even when dealing with small datasets.



(a) ×2 (b) ×4

Figure 4.8: Plots of PSNR progression of training 2D- and 3D-SRCNN915 as well as 2D- and 3D-SRCNN333 using Indian Pines (IP) dataset. The solid line shows the mean of the training results, while the dashed line shows the mean of the validation results. The shaded area represents the standard deviation of all five experiments.

SSIM results depicted in Figures 4.9a and 4.9b demonstrate that the validation results surpass the training results. This phenomenon could potentially indicate overfitting, considering the small size of the dataset. This behavior will be mitigated in Chapter 6.



(a) ×2 (b) ×4

Figure 4.9: Plots of SSIM progression of training 2D- and 3D-SRCNN915 as well as 2D- and 3D-SRCNN333 using Indian Pines (IP) dataset. The solid line shows the mean of the training results, while the dashed line shows the mean of the validation results. The shaded area represents the standard deviation of all five experiments.

Figure 4.10 displays the results for CS. 3D-SRCNN333 once again outperforms 3D-SRCNN915. The contrast between the performances of SRCNN333 and SRCNN915 is more noticeable in the 2D scenario, where 2D-SRCNN915 falls notably behind 3D-SRCNN333, showcasing a clearer distinction between the two cases.

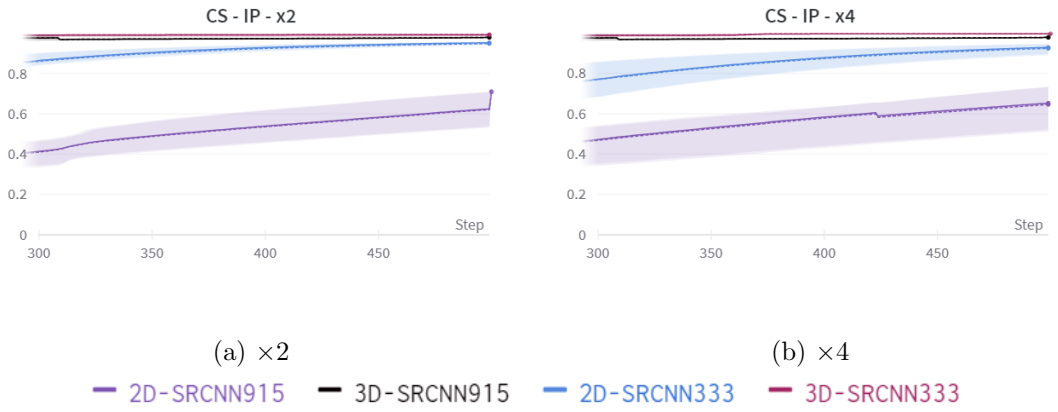


Figure 4.10: Plots of CS progression of training 2D- and 3D-SRCNN915 as well as 2D- and 3D-SRCNN333 using Indian Pines (IP) dataset. The solid line shows the mean of the training results, while the dashed line shows the mean of the validation results. The shaded area represents the standard deviation of all five experiments. The plots are shown between epochs 300-500 for better visibility.

Table 4.3 summarizes all the testing results of all versions of SRCNN on all 3 datasets for scale factors ×2 and ×4. That is, this table shows the performance of the networks on new samples that have not been previously seen during the training and validation phase. The testing results are consistent with those observed in the training and validation figures for all datasets. The unusual behavior due to the limited size of Indian Pines dataset is further highlighted in this table, as it fails to achieve state-of-the-art results.

Table 4.3: Results summary of testing 2D-SRCNN Vs. 3D-SRCNN on datasets of difference sizes: Indian Pines (IP - small), Pavia University (PU - medium), Washington DC Mall (DC - large) in terms of PSNR (dB), SSIM, and SAM (°). The experiments are performed on scale factors ×2 and ×4. Results show that 3D-SRCNN performs better than 2D-SRCNN before and after modifying the filters. 3D-SRCNN333 outperforms all the other networks across all datasets and scale factors.

Dataset	2D-SRCNN				3D-SRCNN			
	Original		Modified		Original		Modified	
	x2	x4	x2	x4	x2	x4	x2	x4
PU (medium)	28.194 ± 0.187 0.8595 ± 0.0022 9.82 ± 0.22	24.704 ± 0.0989 0.6703 ± 0.0058 10.77 ± 0.27	28.248 ± 0.010 0.8622 ± 0.0008 9.17 ± 0.09	25.099 ± 0.117 0.6965 ± 0.0012 10.33 ± 0.14	31.213 ± 0.160 0.9121 ± 0.0017 7.09 ± 0.17	25.479 ± 0.0960 0.6937 ± 0.0004 9.50 ± 0.17	**31.543 ± 0.196** **0.9222 ± 0.0009** **5.31 ± 0.15**	**25.628 ± 0.027** **0.6990 ± 0.0003** **8.54 ± 0.20**
DC (large)	28.774 ± 0.063 0.7712 ± 0.0019 7.26 ± 0.40	26.369 ± 0.059 0.6668 ± 0.0036 9.26 ± 0.31	29.126 ± 0.089 0.7983 ± 0.0025 7.27 ± 0.12	26.685 ± 0.041 0.7038 ± 0.0040 9.05 ± 0.21	31.202 ± 0.019 0.9091 ± 0.0003 4.70 ± 0.05	26.839 ± 0.033 0.7943 ± 0.0052 7.75 ± 0.06	**31.831 ± 0.012** **0.9271 ± 0.0005** **4.50 ± 0.07**	**27.576 ± 0.033** **0.7955 ± 0.0017** **7.46 ± 0.05**
IP (small)	19.779 ± 2.223 0.6446 ± 0.1104 30.46 ± 9.42	4.321 ± 3.886 0.6329 ± 0.1291 19.15 ± 11.25	13.919 ± 1.994 0.37014 ± 0.1566 42.85 ± 10.26	12.831 ± 3.452 0.2441 ± 0.1968 49.08 ± 7.39	24.680 ± 2.952 0.8280 ± 0.0204 11.79 ± 3.15	26.832 ± 2.460 0.7577 ± 0.0184 7.53 ± 3.01	**29.635 ± 2.643** **0.8714 ± 0.1328** **4.78 ± 1.84**	**32.043 ± 2.011** **0.8613 ± 0.0229** **3.83 ± 1.59**

4.4 3D-WRCNN (Medium-depth)

4.4.1 Architecture

The Wide Residual CNN (WRCNN) network is deeper than SRCNN but not as deep as the RUNet discussed in Section 4.5. The network consists of a total of 6 convolution layers with different filter and kernel sizes. It employs the wide activation principle, which involves maximizing the extraction of features before applying the ReLU activation function [334]. The size of each convolution layer is indicated in Figure 4.11. All the convolution layers operate in a 3D manner to extract spatial features while considering spectral context simultaneously. This is considered a deep CNN, which can face challenges such as the vanishing gradient problem and over-parametrization, especially when the dataset size is limited. To address these issues, residual connections are introduced to the architecture. Two forward residual connections are incorporated into the network. However, the possibility of overfitting still exists. Previous studies have shown that the ReLU activation function can impede the propagation of spectral

features throughout the network [334]. The wide activation strategy is employed to alleviate this concern, where the 6 convolutional layers are not followed by ReLU in order to expand the features. This approach enhances performance without compromising computational complexity. The ReLU function is only applied at the end of the network, just before generating the final output.

This network is designed such that the architecture is analogous to the Wide Activation Super Resolution (WDSR) network developed in [334]. This was the first network that used wide activation principle, and creating an architecture with a similar depth provides a basis for comparison. However, a key difference between WRCNN and WDSR is the elimination of BN and pixel shuffling operations. BN has been shown to cause spatial degradation for SISR applications [313]. As for pixel shuffling, while it does not cause spatial degradation, it distorts the spectral signature of HSI [40]. Additionally, pooling is a double-edged sword that can be beneficial for SISR if used within an encoder-decoder architecture [335], as will be seen in RUNet case. Since the proposed network is not very deep and does not follow encoder-decoder topology, pooling layer is avoided.

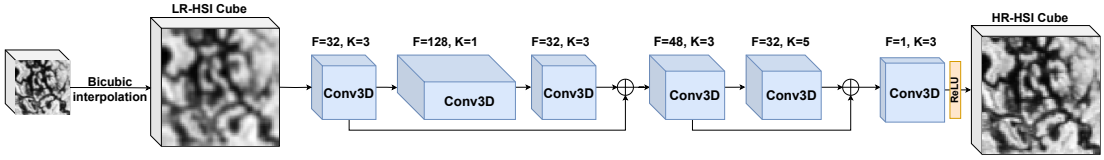


Figure 4.11: Overall architecture of the proposed 3D-WRCNN with the indicated filter (F) and kernel (K) sizes, and ReLU location.

4.4.2 Results

Figures 4.12 - 4.14 summarize the training results for Pavia University, Washington DC Mall, and Indian Pines datasets. The figures compare the performance between 2D- and 3D-WRCNN for scale factors $\times 2$ and $\times 4$. In all the figures, the solid line represents the mean outcome of the corresponding network during training, while the dashed line shows the same for validation, and the shaded area represents the uncertainty.

Figure 4.12a displays the training and validation PSNR for the Pavia University

dataset, covering epochs 0-500. In the case of both scale factors ×2 and ×4, the 3D-WRCNN network exhibits faster convergence compared to the 2D-WRCNN network. However, it is worth noting that the 3D-WRCNN network shows slight instability during training. There are intermittent performance dips occurring randomly between epochs 100-500. This behavior is not observed in the 2D-WRCNN network and interestingly is not present in the 3D-WRCNN network trained on the ×4 scale factor.

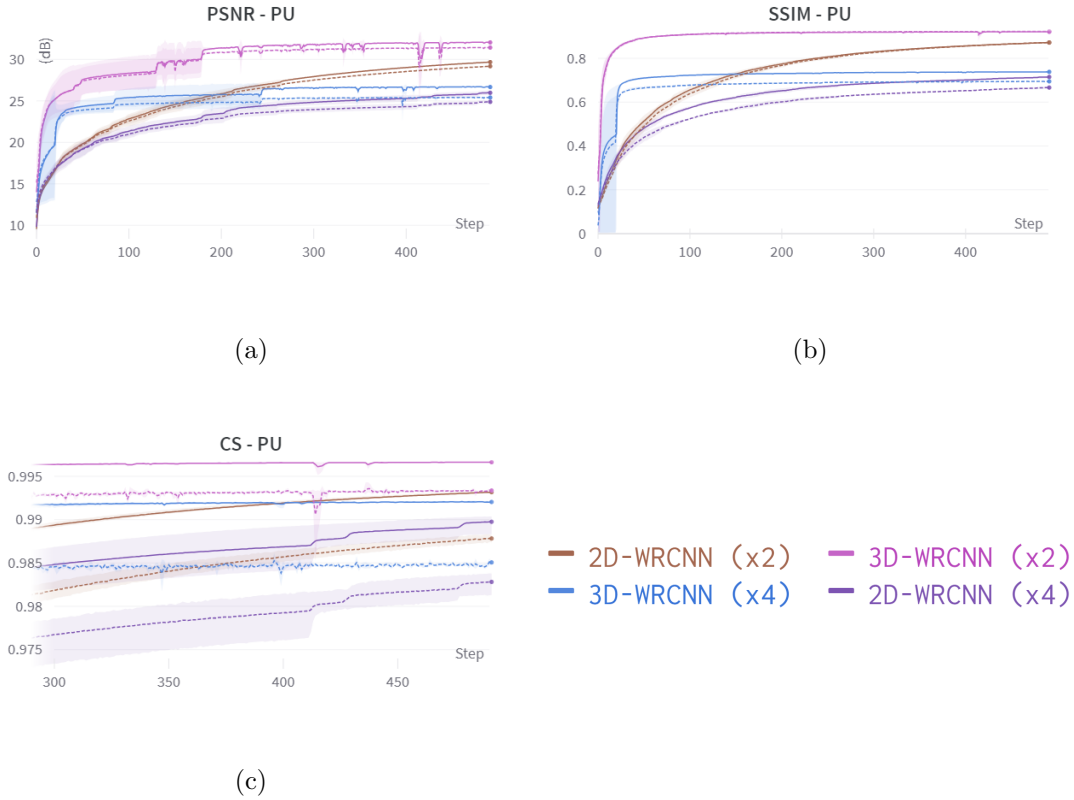


Figure 4.12: This figure shows the progression of training 2D- and 3D-WRCNN on Pavia University dataset. The solid line shows training results, while the dashed line shows validation results, and the shaded area represents the standard deviation. Figures 4.12a and 4.12b show PSNR and SSIM, respectively, for scale factors ×2 and ×4 over 500 epochs. Figure 4.12c shows CS for epochs 300-500 for better visibility.

Regarding SSIM, the 3D-WRCNN network also demonstrates faster convergence and overall stability, with minor fluctuations observed around epoch 400. In contrast, in the case of the CS, as shown in Figure 4.12c, the 3D-WRCNN network displays

random dips and spikes for both scale factors ×2 and ×4. This differs from the observed behavior in the 2D-WRCNN network.

3D-WRCNN exhibits excellent performance on Washington DC Mall dataset, as seen in Figure 4.13. It demonstrates rapid convergence, with only slight instability observed in terms of PSNR. Notably, both the 2D-WRCNN and 3D-WRCNN networks exhibit an interesting behavior where the validation results sometimes surpass the training results. This phenomenon is particularly evident in the CS, as depicted in Figure 4.13c. It is similar to the behavior observed in the 2D- and 3D-SRCNN networks on the Washington DC Mall dataset.

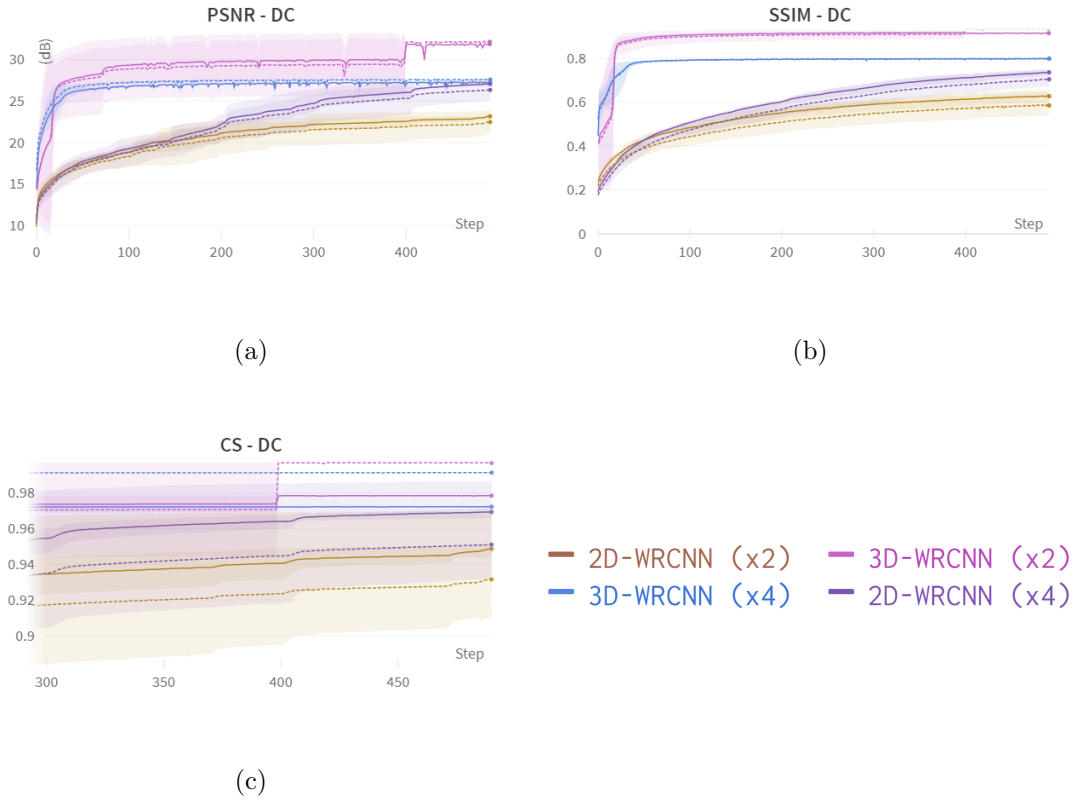


Figure 4.13: This figure shows the progression of training 2D- and 3D-WRCNN on Washington DC Mall dataset. The solid line shows training results, while the dashed line shows validation results, and the shaded area represents the standard deviation. Figures 4.13a and 4.13b show PSNR and SSIM, respectively, for scale factors ×2 and ×4 over 500 epochs. Figure 4.13c shows CS for epochs 300-500 for better visibility.

Indian Pines dataset results seen in Figure 4.14 reveal that 3D-WRCNN converges earlier than its 2D counterpart and in a much faster way compared to Pavia University and Washington DC Mall. Since the dataset is small, it is not enough to achieve a state-of-the-art performance, and all the networks show high uncertainty region during training for all metrics. The SSIM of Indian Pines shows similar behavior to Washington DC Mall in the sense that the validation performance is better than training performance. This happens only for SSIM.

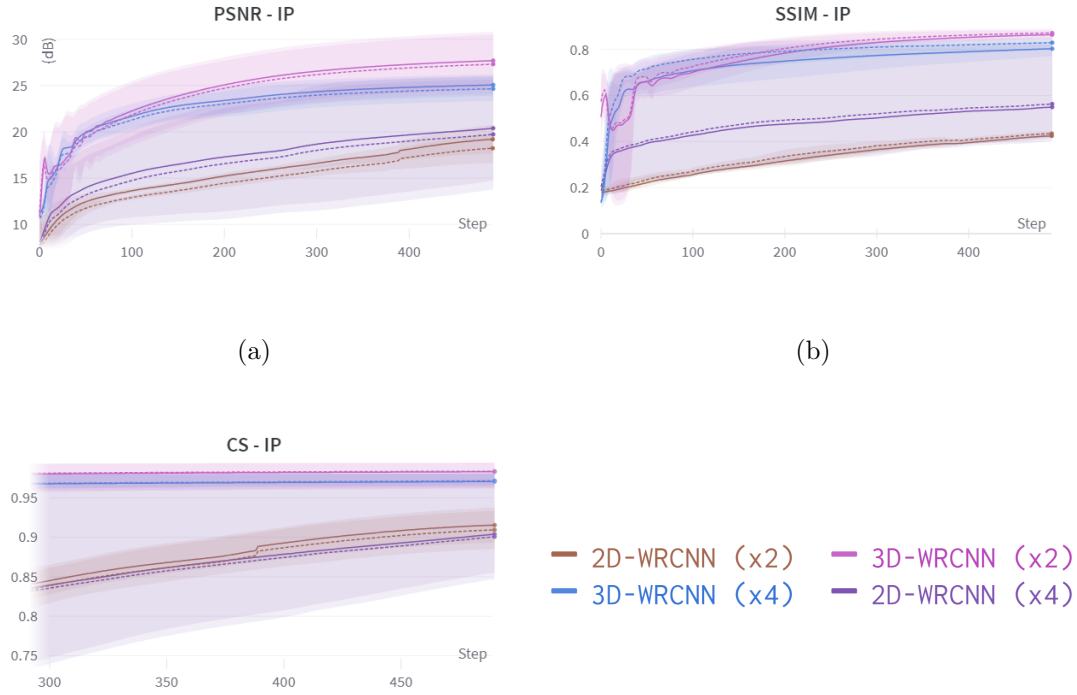


Figure 4.14: This figure shows the progression of training 2D- and 3D-WRCNN on Indian Pines dataset. The solid line shows training results, while the dashed line shows validation results, and the shaded area represents the standard deviation. Figures 4.14a and 4.14b show PSNR and SSIM, respectively, for scale factors $\times 2$ and $\times 4$ over 500 epochs. Figure 4.14c shows CS for epochs 300-500 for better visibility.

Table 4.4 summarizes the performance of 2D- and 3D-WRCNN on testing samples, which are consistent with those observed in the training and validation figures for all

datasets.

Table 4.4: Results of 2D-WRCNN Vs. 3D-WRCNN in terms of PSNR (dB), SSIM, and SAM (°) on 3 different datasets for two scale factors.

Dataset	2D-WRCNN		3D-WRCNN	
	x2	x4	x2	x4
PU (medium)	28.448 ± 0.567 0.8591 ± 0.0015 9.32 ± 1.27	25.501 ± 0.248 0.6565 ± 0.012 10.99 ± 0.81	$\mathbf{31.280 \pm 0.116}$ $\mathbf{0.9127 \pm 0.0001}$ $\mathbf{6.69 \pm 0.17}$	$\mathbf{25.531 \pm 0.128}$ $\mathbf{0.6626 \pm 0.0002}$ $\mathbf{10.08 \pm 0.25}$
DC (large)	27.86 ± 0.264 0.7526 ± 0.0080 9.54 ± 0.43	24.021 ± 0.207 0.6441 ± 0.0056 14.02 ± 0.88	$\mathbf{31.197 \pm 0.052}$ $\mathbf{0.9135 \pm 0.0002}$ $\mathbf{4.65 \pm 0.61}$	$\mathbf{27.582 \pm 0.107}$ $\mathbf{0.7593 \pm 0.0003}$ $\mathbf{8.47 \pm 0.65}$
IP (small)	17.173 ± 2.827 0.4795 ± 0.1596 27.61 ± 5.50	16.955 ± 1.095 0.4687 ± 0.1545 31.52 ± 5.78	$\mathbf{28.954 \pm 1.402}$ $\mathbf{0.8599 \pm 0.0144}$ $\mathbf{6.33 \pm 3.03}$	$\mathbf{24.446 \pm 1.497}$ $\mathbf{0.8442 \pm 0.1749}$ $\mathbf{15.12 \pm 4.11}$

4.5 3D-RUNet (Deep)

4.5.1 Architecture

Autoencoders are a type of generative ANNs that consists of two parts; encoder and decoder. The encoder is specialized in compressing the input into a reduced form represented as a set of features. The decoder then learns how to reconstruct the original form from said features by minimizing the error between the input and the output. Autoencoders have been successfully utilized for anomaly detection [336], segmentation [337], and image de-noising [338]. One example of such network is UNet, which was originally devised for segmenting biomedical images [339]. UNet was later re-purposed to perform SISR tasks for RGB and grayscale images. That is, instead of producing

a binary mask as an output, it produces an output image that is an enhanced version of the input. The work of [314] demonstrates UNet usage for MSI-SISR, particularly RGB images. UNet was later improved further by introducing residual connections to the architecture. This upgraded version of the network is referred to as RUNet, which shows superiority over the original UNet and bicubic interpolation in terms of PSNR, SSIM, and MSE. Autoencoders, including RUNet, thus far have been used to enhance MSI and grayscale images, but not HSI. It is worth investigating RUNet's capability in enhancing HSI and its behavior with small datasets considering the depth of the network.

The original RUNet architecture contains pixel shuffling and BN layers. However, these layers degrade the performance of HSI-SR networks spectrally and spatially. BN, according to the designers of the EDSR [313], destroys image scale information and decreases the flexibility range of activation functions. In addition to improving SR performance, removing BN layers can reduce GPU memory usage, which allows saving resources that can be used to design bigger models if needed [340]. As for pixel shuffling layer, it disturbs the spectral fidelity of HSI [40].

The architecture of the proposed 3D-RUNet differs from the original 2D-RUNet not only due to its 3D operations, but also due to the fact that it encourages symmetry between the encoder and decoder parts. This allows a fully learnable path that deconstructs the LR-HSI to a smaller size down to its important features, and then reconstructs it back to its full size in its HR form. This network utilizes intra- and extra-residual connections. Intra-residual connections exist within layers on the encoder side, and the extra-residual ones extend from the encoder side to the decoder side. These connections enhance the propagation of data throughout the network. There are 4 intra-residual connections between layers on the encoder size, and 4 extra-residual connections that connect each encoder layer to its corresponding decoder layer. The overall network architecture is seen in Figure 4.15.

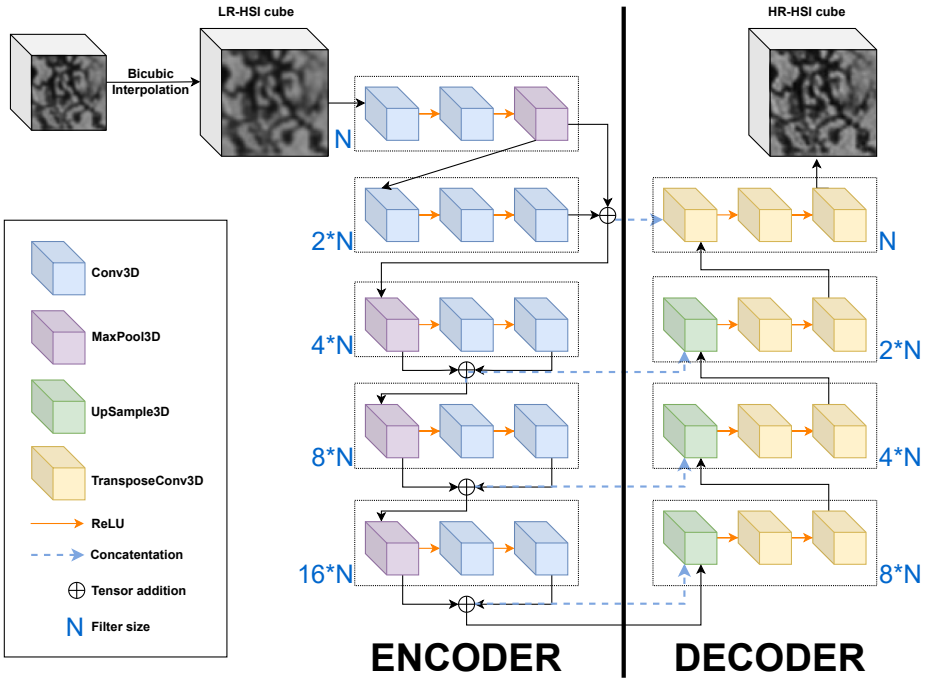


Figure 4.15: Overall architecture of the proposed 3D-RUNet.

4.5.2 Results

The quantitative metrics progression on the Pavia University dataset for epochs 0-500 is depicted in Figure 4.16. The training and validation results for both the 2D- and 3D-RUNet networks at a scale factor of $\times 2$ exhibit relatively close values. This indicates that the networks are not overfitting, which is a positive observation. However, when considering the scale factor of $\times 4$, there is a significant gap between the training and validation results in terms of PSNR, SSIM, and CS. This suggests that the network may not be effectively handling larger scale factors, as the amount of missing information increases with higher scale factors.

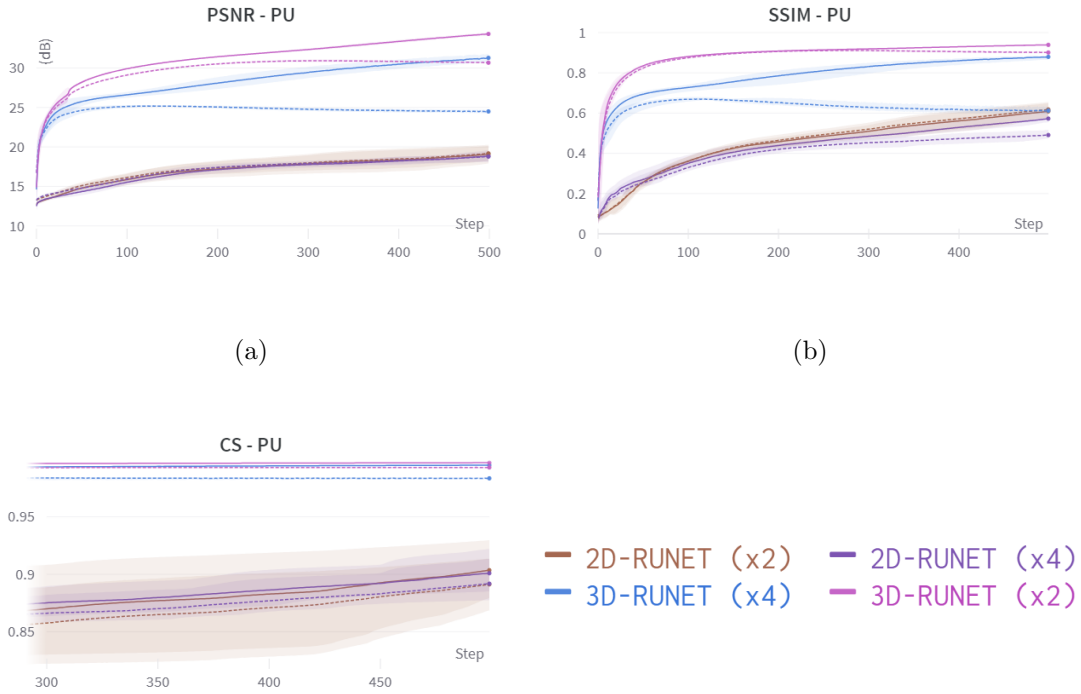
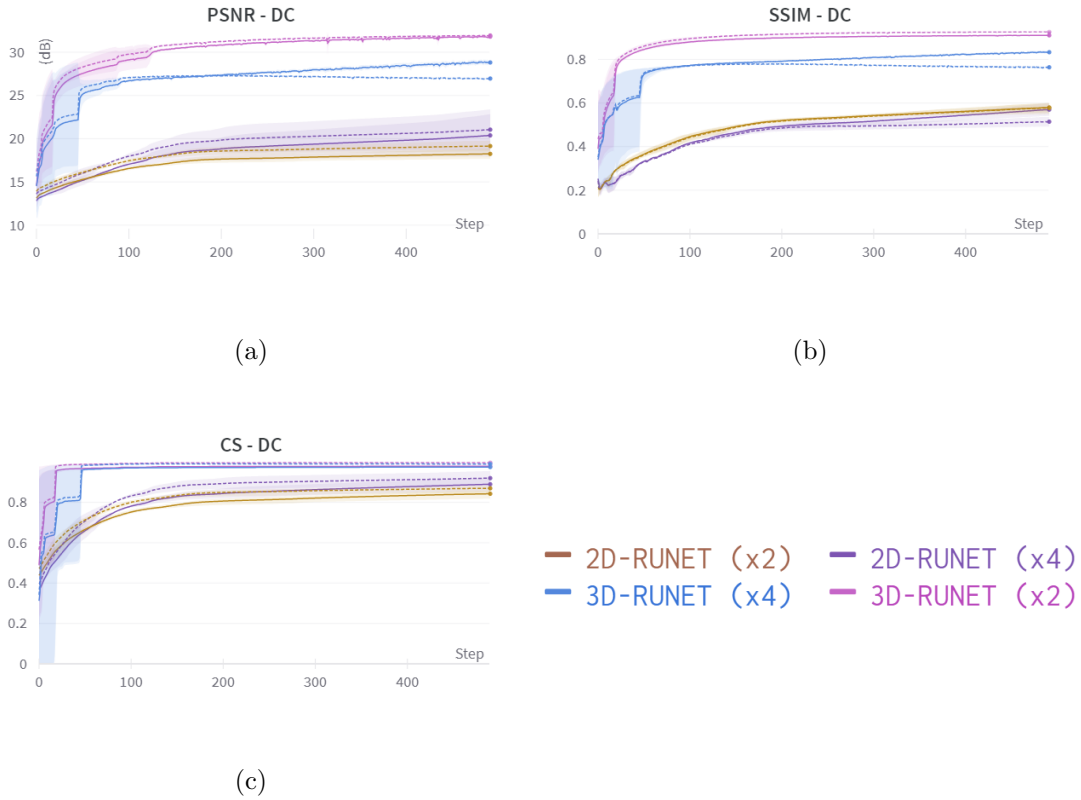


Figure 4.16: This figure shows the progression of training 2D- and 3D-RUNet on Pavia University dataset. The solid line shows training results, while the dashed line shows validation results, and the shaded area represents the standard deviation. Figures 4.16a and 4.16b show PSNR and SSIM, respectively, for scale factors $\times 2$ and $\times 4$ over 500 epochs. Figure 4.16c shows CS for epochs 300-500 for better visibility.

Figure 4.17 provides a summary of the results obtained on the Washington DC Mall dataset. As this dataset is larger than the Pavia University dataset, the issue of the gap between training and validation results appears to be less pronounced, although it is still present. This discrepancy is particularly noticeable in the PSNR and SSIM values of the 3D-RUNet network for the $\times 4$ scale factor. It is possible that the larger dataset allows the capabilities of the RUNet network to be better demonstrated.



(a) (b)

(c)

Figure 4.17: This figure shows the progression of training 2D- and 3D-RUNet on Washington DC Mall dataset. The solid line shows training results, while the dashed line shows validation results, and the shaded area represents the standard deviation. Figures 4.17a and 4.17b show PSNR and SSIM, respectively, for scale factors $\times 2$ and $\times 4$ over 500 epochs. Figure 4.17c shows CS for epochs 300-500 for better visibility.

Figure 4.18 shows that all RUNet versions are unable to achieve acceptable performance on Indian Pines dataset, which can be also observed from Table 4.5. PSNR and SSIM are low, whereas SAM is high. It is interesting to note that 3D-RUNet converges faster for Washington DC Mall, but not for the smaller datasets (Pavia University and Indian Pines).

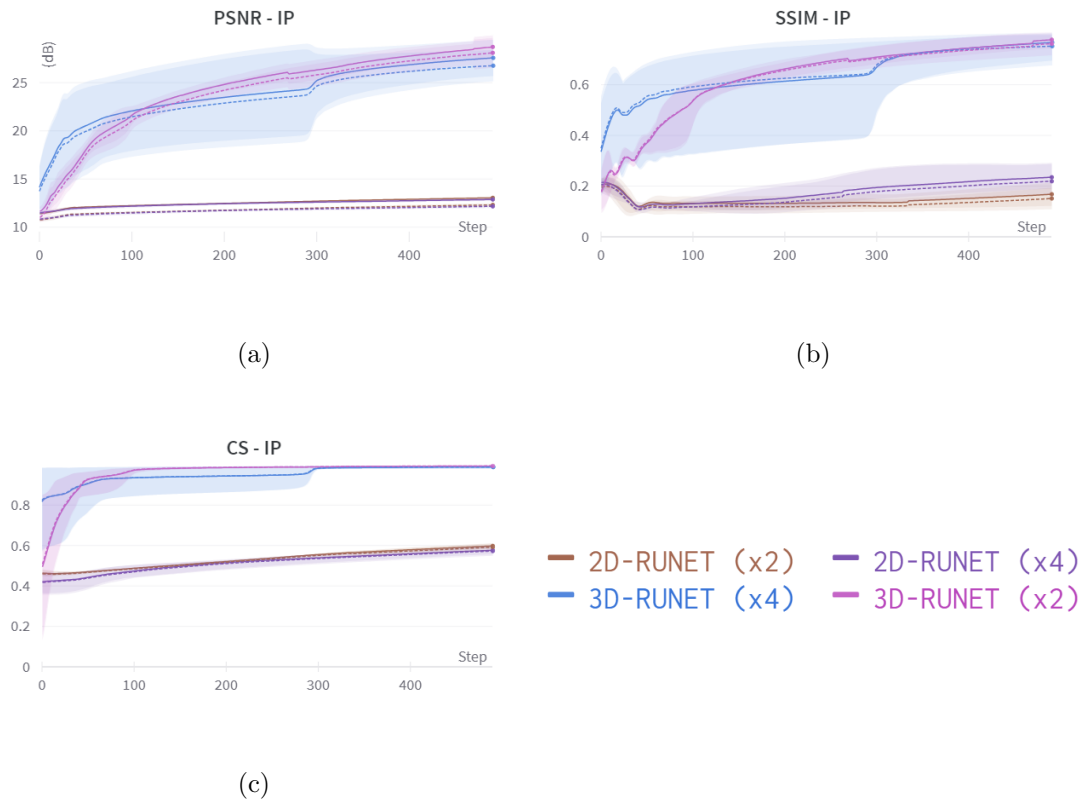


Figure 4.18: This figure shows the progression of training 2D- and 3D-RUNet on Indian Pines dataset. The solid line shows training results, while the dashed line shows validation results, and the shaded area represents the standard deviation. Figures 4.18a and 4.18b show PSNR and SSIM, respectively, for scale factors $\times 2$ and $\times 4$ over 500 epochs. Figure 4.18c shows CS for epochs 300-500 for better visibility.

Upon examining the numerical results in Table 4.5, it can be observed that the 3D-RUNet network achieves satisfactory performance in terms of spatial metrics on the Pavia University dataset. However, the high SAM score indicates that the preservation of spectral signatures was not successful. This phenomenon is not observed in the Washington DC Mall dataset.

Table 4.5: Results of 2D-RUNet Vs. 3D-RUNet in terms of PSNR (dB), SSIM, and SAM (°) on 3 different datasets for two scale factors.

Dataset	2D-RUNet		3D-RUNet	
	x2	x4	x2	x4
PU (medium)	21.612 ± 2.154 0.7204 ± 0.2416 20.54 ± 5.72	19.625 ± 2.397 0.5165 ± 0.2776 24.02 ± 7.35	**30.818 ± 0.260** **0.9077 ± 0.0014** **6.69 ± 0.36**	**24.385 ± 0.215** **0.6082 ± 0.0001** **10.20 ± 0.58**
DC (large)	18.433 ± 2.771 0.5009 ± 0.2035 32.20 ± 6.35	19.389 ± 3.337 0.4918 ± 0.2881 27.75 ± 7.24	**32.007 ± 0.121** **0.9360 ± 0.0010** **4.36 ± 0.11**	**27.598 ± 0.235** **0.7989 ± 0.0015** **7.48 ± 0.21**
IP (small)	11.91 ± 2.47 0.2465 ± 0.1028 54.85 ± 15.27	12.70 ± 2.91 0.2951 ± 0.1492 53.14 ± 13.56	**27.881 ± 1.47** **0.7423 ± 0.1793** **6.49 ± 2.41**	**25.970 ± 2.01** **0.8069 ± 0.0676** **11.28 ± 2.49**

4.6 Results and Analysis

In this section, a detailed analysis and comparison are presented to evaluate the performance of the developed 3D-CNNs. These networks have demonstrated superior performance compared to their 2D counterparts. They will be thoroughly compared not only among themselves but also against other standard methods. The literature features an ongoing discussion regarding the optimal application of deep learning versus traditional methods, with considerations based on dataset size. Given the relatively small size of the datasets in this context, it becomes intriguing to explore whether bicubic interpolation could potentially outperform the developed 3D-CNNs. Furthermore, it is worth noting that the original research paper on 3D-FCNN offers a substantial level of detail, enabling reliable replication. In contrast, some other methods leave certain aspects open to interpretation and may not provide sufficient information for

reliable replication. Consequently, this analysis will also incorporate 3D-FCNN into the analysis.

4.6.1 Pavia University Experiments

Figure 4.19 illustrates the visual outcomes obtained from testing each method on the Pavia University dataset. The top row of each scale factor presents the final output of the method, while the bottom row displays the error map in terms of Root MSE (RMSE). For the $\times 2$ scale factor, bicubic interpolation exhibits the poorest performance, followed by 3D-RUNet, which exhibits some artifacts in the output. 3D-FCNN, 3D-WRCNN, and 3D-SRCNN show only minor discrepancies in the final results, with 3D-SRCNN333 being more effective in reconstructing certain highlighted details within the yellow box in Figure 4.19. The differences among these networks become more pronounced for the $\times 4$ scale factor, where the amount of missing information is greater. 3D-SRCNN333 distinguishes itself by displaying lower error in the RMSE map.

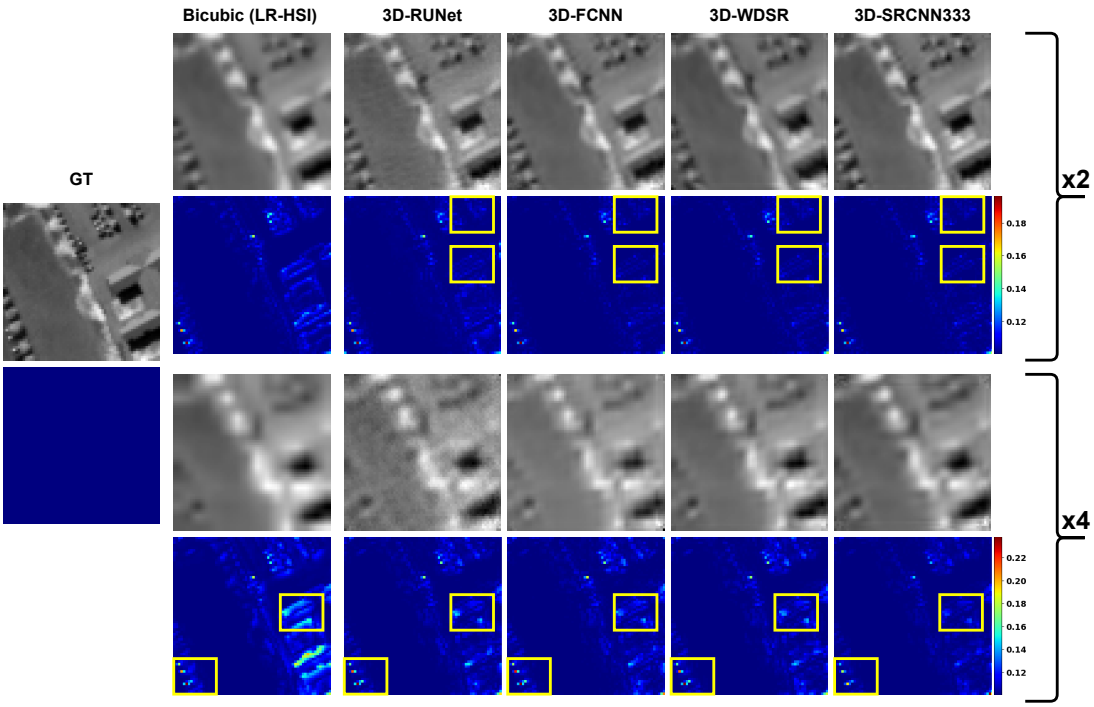
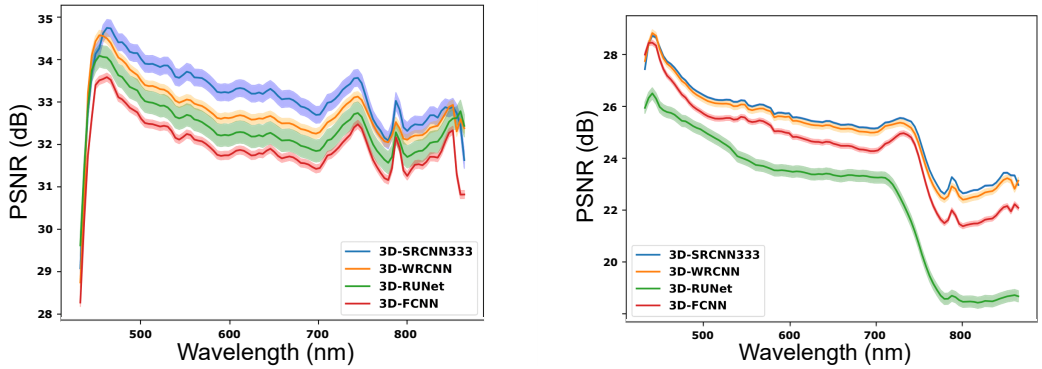


Figure 4.19: Visual results of the predicted HR-HSI for Pavia University by each method. For each scale factor, the top row shows the output of the method, and the bottom row shows the error map in terms of RMSE.

This observation is further supported by examining the PSNR values across the Pavia University wavelength range depicted in Figures 4.20. Based on the findings from both scaling factors of ×2 and ×4, it is evident that 3D-SRCNN333 stands out as the top-performing model in terms of overall PSNR. Figure 4.20b clearly illustrates that 3D-RUNet exhibits the most significant deterioration in PSNR when compared to the other networks; 3D-SRCNN333, 3D-WRCNN, and 3D-FCNN. Additionally, the performances of 3D-SRCNN and 3D-WRCNN are notably close, yet 3D-SRCNN333 maintains a higher PSNR overall.



(a) ×2 (b) ×4

Figure 4.20: PSNR results recorded for Pavia University spectral range, where (a) shows the results for scale factor ×2 and (b) shows the results for scale factor ×4. The solid lines represent the mean, while the shaded areas represent the standard deviation.

The investigation into the fidelity of spectral signatures involves analyzing some of the endmembers that exist within Pavia University scene. As a case study, Metal Sheet, Gravel, and Asphalt have been selected for this purpose. For each endmember, a homogeneous region is carefully selected, enabling the capture of the GT spectral signature's mean. Subsequently, each network's mean spectral signature prediction for the same region is recorded. Since the networks have been trained five times, the prediction step is repeated five times for each network, and the resultant data is summarized through the calculation of both mean and standard deviation of all five tests.

In Figure 4.21, the outcomes of Metal Sheet are displayed for scale factors ×2

and ×4. Figures 4.21b and 4.21d provide a comparison for scale factors ×2 and ×4, respectively, between the spectral signatures generated by each method and the GT for the homogeneous region depicted in Figure 4.21a. Notably, for both scale factors, 3D-RUNet exhibits the most substantial deviation from the GT spectral signature, while 3D-FCNN, 3D-WRCNN, and 3D-SRCNN333 mimic the GT spectral signature more closely in comparison. The visual representations in Figures 4.21c and 4.21e portray the ratio of the predicted signature to the GT. The degree of resemblance to the GT is gauged based on the plot's proximity to the value of 1. In 4.21c, 3D-WRCNN and 3D-FCNN show a large degree of overlap between each other. The difference in performance between all the networks is more pronounced in scale factor ×4, as the networks show more deterioration compared to scale factor ×2, as portrayed by the ratio plot seen in Figure 4.21e. In scale factor ×4, 3D-WRCNN is closer to the value 1 than 3D-FCNN. Overall, 3D-SRCNN333 produces a curve that aligns closest to the ideal value of 1, as seen in both Figures 4.21c and 4.21e, affirming its ability to preserve spectral signature fidelity better than the other networks for this class.

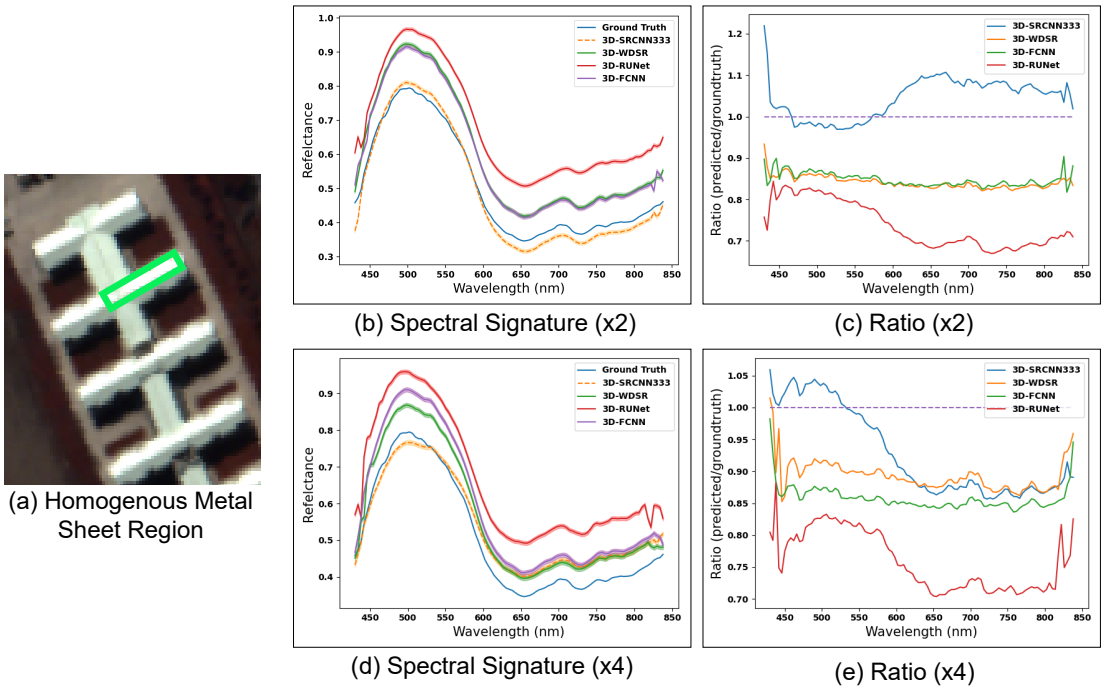


Figure 4.21: The results of testing 3D-SRCNN, 3D-WRCNN, 3D-RUNet, and 3D-FCNN on (a) homogeneous Metal Sheet region, where the spectral signature is plotted in (b) for scale factor $\times 2$, (c) shows the corresponding ratio plots, (d) shows the spectral signature for scale factor $\times 4$, and (e) shows the corresponding ratio plots. The solid lines represent the mean, while the shaded areas represent the standard deviation. The overall order of performance is: 3D-SRCNN, 3D-WRCNN, 3D-FCNN, and 3D-RUNet.

Shifting focus to Gravel class, The results are illustrated in Figure 4.22 for the homogeneous region depicted in 4.22a. At scale factor $\times 2$, the distinctions among the spectral signatures generated by the algorithms are less pronounced compared to the Metal Sheet class, as there is a certain degree of overlap in the uncertainty regions of the predicted results by the networks seen in Figure 4.22b. Nonetheless, upon closer scrutiny of Figure 4.22c, it becomes clear that a more consistent oscillation around the value of 1 is exhibited by 3D-SRCNN333 in comparison to the other networks, further highlighting its proficiency in preserving spectral signature fidelity.

At a scale factor of $\times 4$, the distinctions in the outcomes of each network seen in Figure4.22d become more apparent, as all the networks show larger deviation from the GT compared to scale factor $\times 2$. Figure 4.22e affirms this fact, and shows that 3D-SRCNN333 exhibits less distortions compared to the other networks by being relatively

closer to the value of 1.

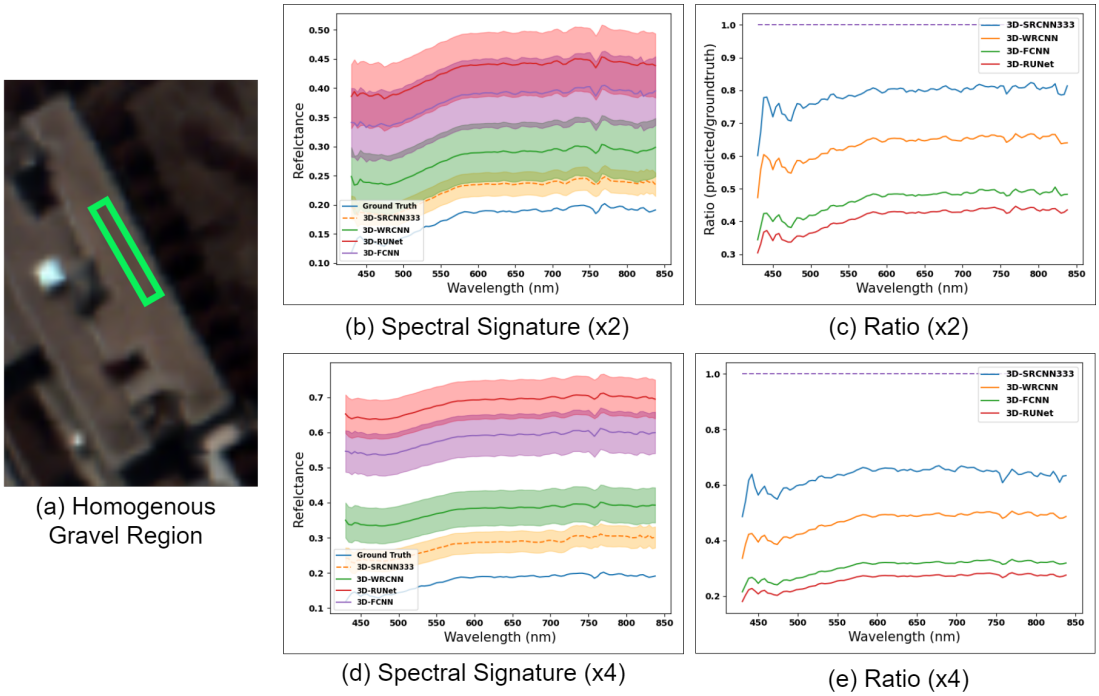


Figure 4.22: The results of testing 3D-SRCNN, 3D-WRCNN, 3D-RUNet, and 3D-FCNN on (a) homogeneous Gravel region, where the spectral signature is plotted in (b) for scale factor $\times 2$, (c) shows the corresponding ratio plots, (d) shows the spectral signature for scale factor $\times 4$, and (e) shows the corresponding ratio plots. The solid lines represent the mean, while the shaded areas represent the standard deviation. The overall order of performance is: 3D-SRCNN, 3D-WRCNN, 3D-FCNN, and 3D-RUNet.

The outcomes pertaining to the Asphalt class are graphically represented in Figure 4.23. Figure 4.23b showcases the mean spectral signature and the associated shaded uncertainty region, as predicted by each network, in comparison with the GT for scale factor $\times 2$. These plots were computed according to the specified region seen in Figure 4.23a. Similar to the Gravel class, the average output generated by 3D-SRCNN exhibits a higher degree of similarity with the GT. However, some degree of overlap is observed between the uncertainty regions of 3D-SRCNN and 3D-WRCNN, as well as between the uncertainty regions of 3D-FCNN and 3D-RUNet. Inspecting Figure 4.23c reveals that 3D-SRCNN333 exhibits the closest proximity to a ratio of 1, while 3D-RUNet demonstrates the most significant deviation from this optimal value.

At a $\times 4$ scale factor for Asphalt, the pattern of observation parallels that observed

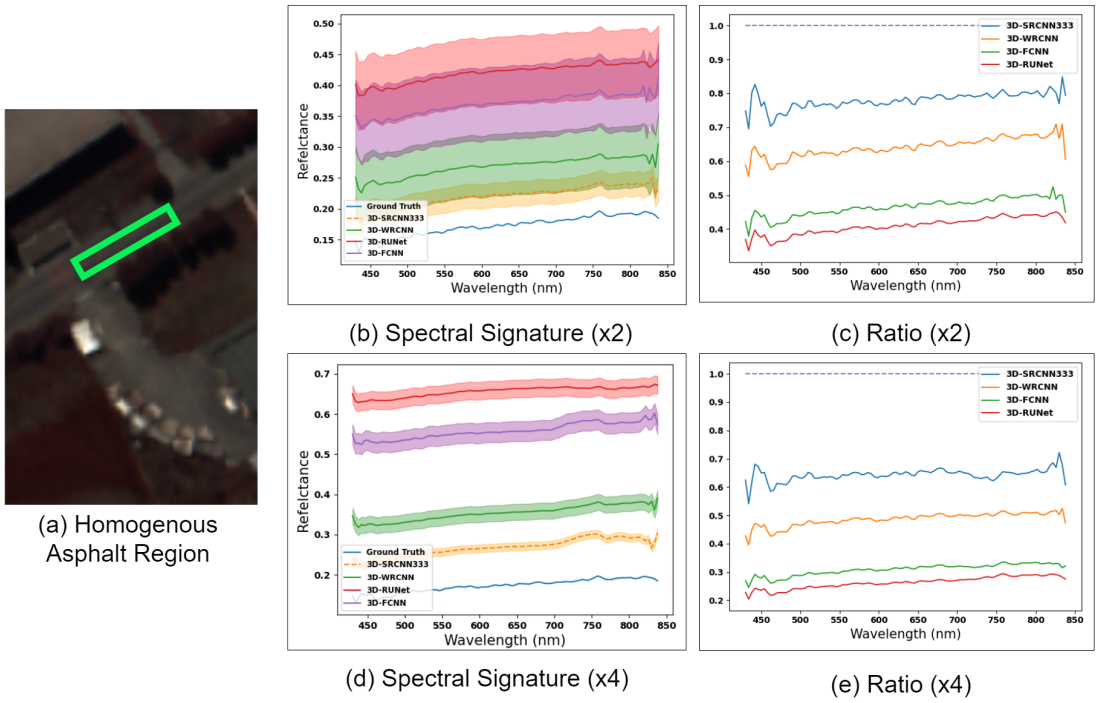


Figure 4.23: The results of testing 3D-SRCNN, 3D-WRCNN, 3D-RUNet, and 3D-FCNN on (a) homogeneous Asphalt region, where the spectral signature is plotted in (b) for scale factor $\times 2$, (c) shows the corresponding ratio plots, (d) shows the spectral signature for scale factor $\times 4$, and (e) shows the corresponding ratio plots. The overall order of performance is: 3D-SRCNN, 3D-WRCNN, 3D-FCNN, and 3D-RUNet.

at a $\times 2$ scale factor, as seen in Figure 4.23d. Specifically, 3D-SRCNN333 adheres to the GT spectral shape, followed by 3D-WRCNN, then 3D-FCNN, and 3D-RUNet. This consistency aligns with the ratio plot presented in Figure 4.23e.

In summary, across the Metal Sheet, Gravel, and Asphalt classes, 3D-SRCNN333 consistently demonstrates the best overall performance, followed by 3D-WRCNN, 3D-FCNN, and, lastly, 3D-RUNet, for both scale factors.

4.6.2 Washington DC Mall Experiments

Washington DC Mall qualitative results are shown in Figure 4.24. 3D-RUNet performance on Washington DC Mall is better than its performance on Pavia University. Furthermore, it is subjectively better than 3D-FCNN. Nonetheless, the performance remains close to that of 3D-WRCNN and 3D-SRCNN333, with only minor differences

between them. The difference is more apparent in the case of scale factor $\times 4$, although their results remain visually similar. This is also reflected in the Wavelength Vs. PSNR plot shown in Figure 4.25, however, it is more apparent that 3D-FCNN performance is lower than the other networks. To get the full picture, the spectral signature must be inspected.

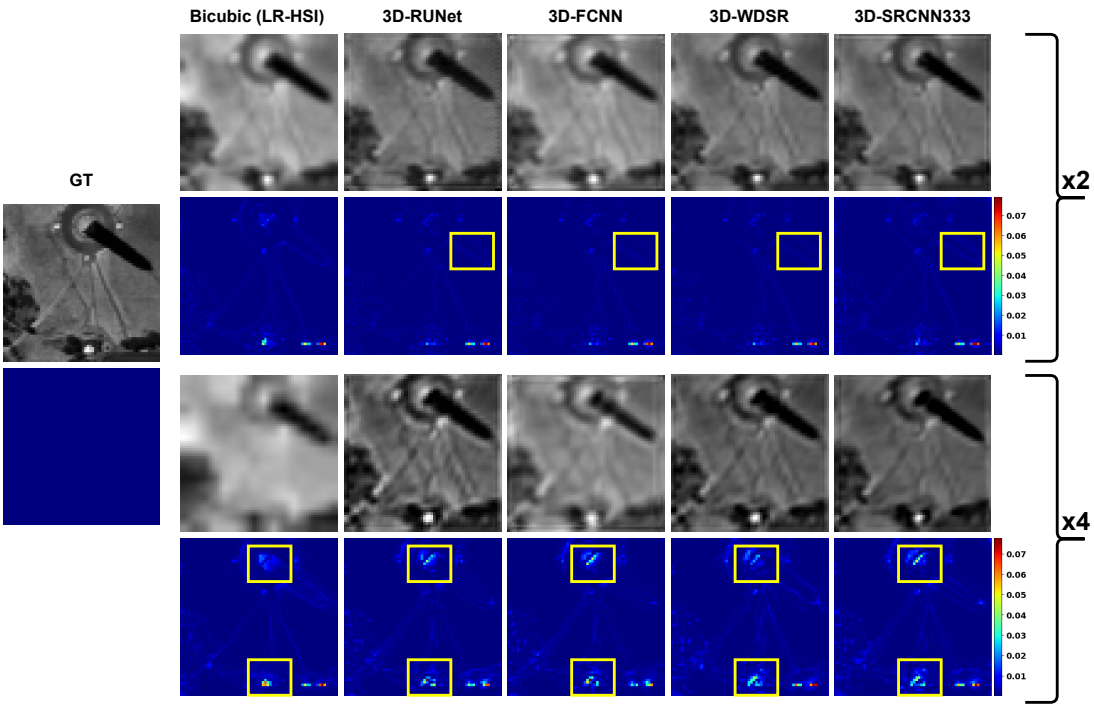
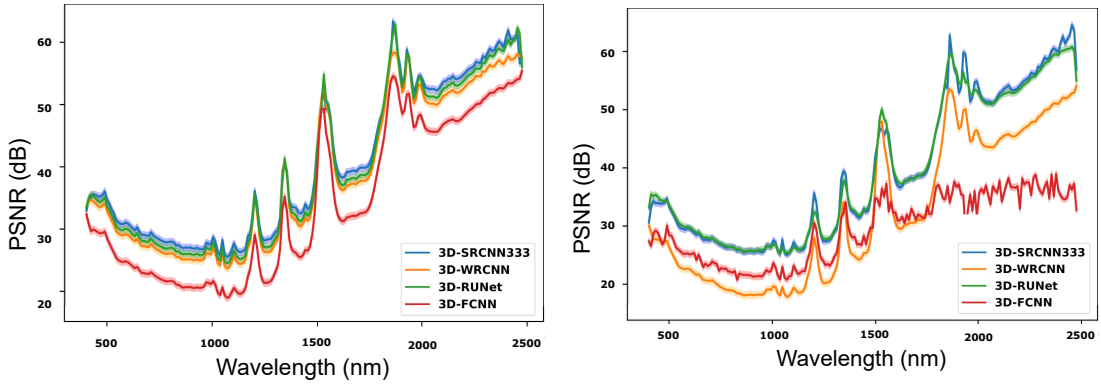


Figure 4.24: Visual results of the predicted HR-HSI for Washington DC Mall dataset by each method. For each scale factor, the top row shows the output of the method, and the bottom row shows the error map in terms of RMSE.



(a) ×2 (b) ×4

Figure 4.25: PSNR results recorded for Washington DC Mall spectral range, where (a) shows the results for scale factor ×2 and (b) shows the results for scale factor ×4. the solid lines represent the mean, while the shaded areas represent the standard deviation.

The fidelity of spectral signatures is investigated by analyzing some of the distinct endmembers within Washington DC Mall scene. To this end, the examined classes are Grass, Roof, and Water. For each endmember, a homogeneous region is carefully selected to facilitate the capture of the mean spectral signature of the GT. Subsequently, the mean spectral signature prediction for the same region by each network is recorded. Since the networks have been trained five times, the prediction step is repeated five times for each network, and the resulting data is summarized through the calculation of both mean and standard deviation based on all five tests.

In Figure 4.26b, the Grass spectral patterns generated by each network are presented for a scaling factor of ×2. The figure provides a visual comparison between the spectral patterns generated by each method and the GT for the uniform area shown in Figure 4.26a. It is evident that most networks manage to capture the overall spectral curve. Upon examining the ratio plot in Figure 4.26c, it becomes apparent that 3D-SRCNN333 and 3D-RUNet are closest to a value of 1, which means they resemble the GT spectral curve the most. On the other hand, 3D-FCNN and 3D-WRCNN exhibit greater deviations from the GT, highlighting the spectral distortions introduced by these networks. The disparity between the outcomes produced by the networks becomes more visible when examining Figure 4.26d. It becomes evident that, up to approximately 1500 nm, both 3D-SRCNN333 and 3D-RUNet outperform 3D-WRCNN and 3D-FCNN. Beyond

this point, their performances nearly converge. This trend is further emphasized in Figure 4.26e, where it is apparent that all networks exhibit a decreasing similarity to the GT after 1500 nm. Nevertheless, 3D-RUNet and 3D-SRCNN333 remain closer to a ratio of 1 when compared to 3D-WRCNN and 3D-FCNN. In summary, all networks display significant spectral distortions in this particular case, with 3D-SRCNN333 and 3D-RUNet showing relatively less distortions.

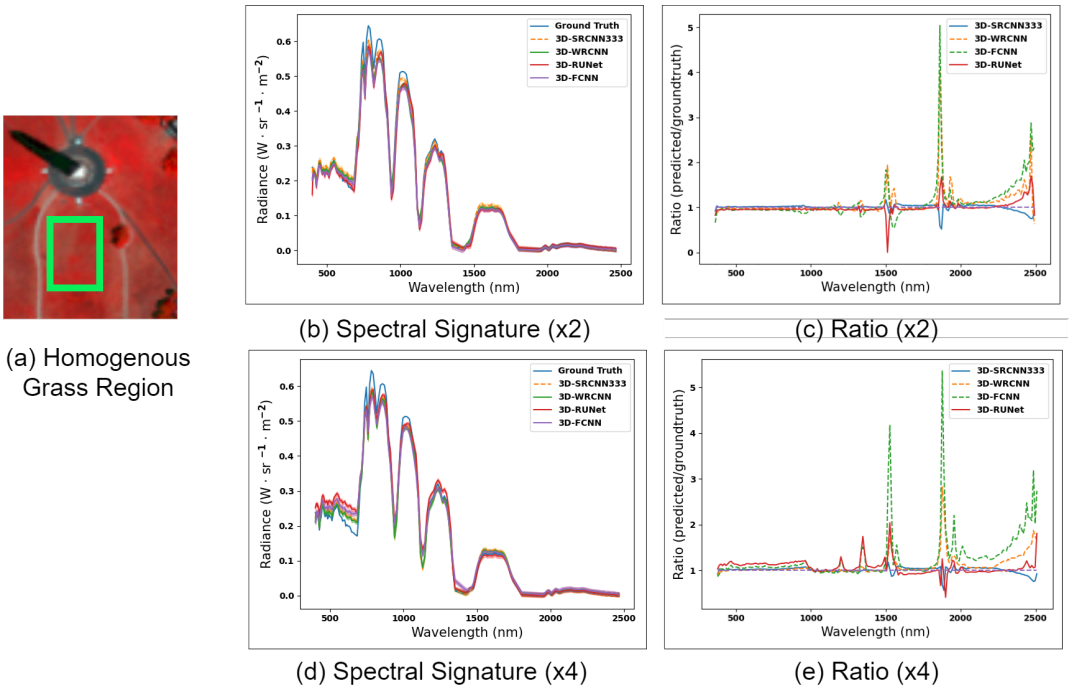


Figure 4.26: The results of testing 3D-SRCNN333, 3D-WRCNN, 3D-RUNet, and 3D-FCNN on (a) homogeneous Grass region, where the spectral signature is plotted in (b) for scale factor $\times 2$, (c) shows the corresponding ratio plots, (d) shows the spectral signature for scale factor $\times 4$, and (e) shows the corresponding ratio plots. The solid lines represent the mean, while the shaded areas represent the standard deviation. The order of performance is: 3D-SRCNN333, 3D-RUNet, 3D-WRCNN, 3D-FCNN.

The Roof class results are depicted in Figure 4.27. The spectral signatures for scale factor $\times 2$ predicted by each method can be observed in Figure 4.27b, specifically for the region indicated in Figure 4.27a. This particular scenario differs from the previous one, as there are more pronounced distortions apparent in 3D-WRCNN and 3D-FCNN, whereas 3D-SRCNN333 and 3D-RUNet maintain relatively closer fidelity. A closer examination of the ratio plot in Figure 4.27c reveals that 3D-RUNet exhibits more

distortions when compared to 3D-SRCNN333.

Similar observations apply to the results at a scale factor of ×4. A closer examination in Figure 4.27d highlights that 3D-FCNN exhibits the most pronounced distortions and its result appears similar to that of 3D-WRCNN, while the other two networks demonstrate a closer adherence to the GT. To discern the differences in their performance more clearly, Figure 4.27e highlights that 3D-SRCNN exhibits less distortions than 3D-RUNet.

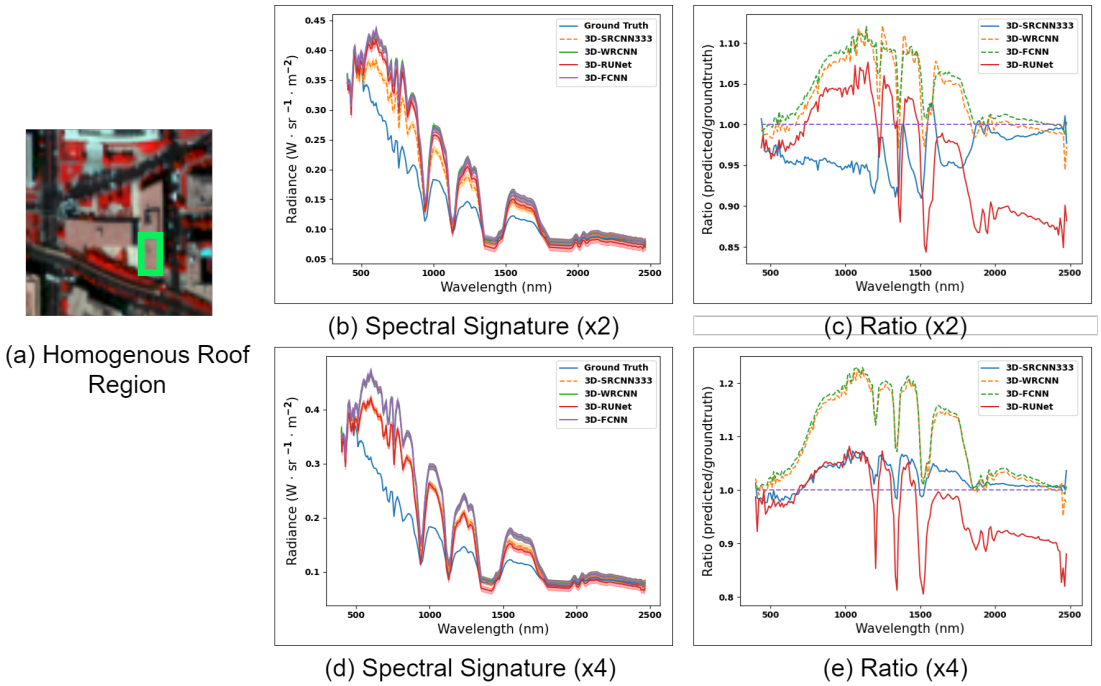


Figure 4.27: The results of testing 3D-SRCNN333, 3D-WRCNN, 3D-RUNet, and 3D-FCNN on (a) homogeneous Roof region, where the spectral signature is plotted in (b) for scale factor ×2, (c) shows the corresponding ratio plots, (d) shows the spectral signature for scale factor ×4, and (e) shows the corresponding ratio plots. The solid lines represent the mean, while the shaded areas represent the standard deviation. The order of performance is: 3D-SRCNN333, 3D-RUNet, then 3D-WRCNN and 3D-FCNN, which exhibit similar performance.

Proceeding to the Water class results, the predicted spectral signatures are recorded for the homogeneous Water region seen in Figure 4.28a. Figure 4.28b reveals that the general spectral shape is effectively captured by 3D-SRCNN333 and 3D-RUNet, albeit with occasional discernible distortions. Figure 4.28c, which displays the ratio

of each network's prediction to the GT, establishes that 3D-RUNet outperforms 3D-SRCNN333 up to approximately 1500 nm. After this point, 3D-SRCNN333 exhibits better performance. For scale factor of ×4, the spectral signatures, as observed in Figure 4.28d, exhibit noticeable deviations from the GT. Figure 4.28e highlights that 3D-SRCNN333 shows less spectral distortions, followed 3D-RUNet, 3D-FCNN, then 3D-WRCNN.

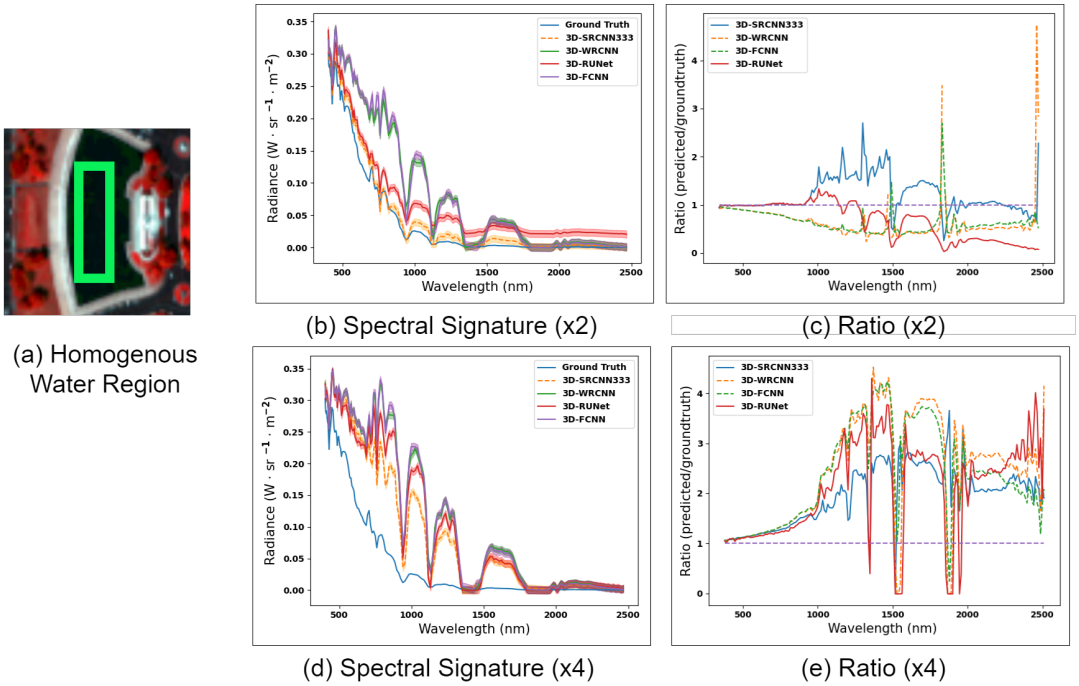


Figure 4.28: The results of testing 3D-SRCNN333, 3D-WRCNN, 3D-RUNet, and 3D-FCNN on (a) homogeneous Water region, where the spectral signature is plotted in (b) for scale factor ×2, (c) shows the corresponding ratio plots, (d) shows the spectral signature for scale factor ×4, and (e) shows the corresponding ratio plots. The solid lines represent the mean, while the shaded areas represent the standard deviation. The order of performance: 3D-SRCNN333, 3D-RUNet, 3D-FCNN, 3D-WRCNN.

In summary, both 3D-SRCNN333 and 3D-RUNet consistently demonstrated comparable performance across Grass, Roof, and Water classes, with occasional instances where 3D-SRCNN333 outperformed 3D-RUNet. In contrast, 3D-WRCNN and 3D-FCNN exhibited notably more distortions in their results when compared to the former two models.

4.6.3 Indian Pines Experiments

The Indian Pines dataset represents a challenging scenario with a very limited size. Figure 4.29 depicts the qualitative outcomes obtained from each method. For the $\times 2$ scale factor, 3D-WRCNN and 3D-SRCNN333 exhibit excellent performance despite the scarcity of training data. On the other hand, 3D-RUNet and 3D-FCNN struggle to capture most of the details and display distortions along the image borders, making them less effective than bicubic interpolation. When it comes to the $\times 4$ scale factor, where more information is missing, the task of reconstructing HR-HSI becomes considerably more challenging for the networks. In this case, 3D-RUNet, 3D-FCNN, and 3D-WRCNN all fail to capture the image details satisfactorily. 3D-SRCNN333 manages to capture a few details, which is relatively better than the other networks, although it does not achieve state-of-the-art performance and is still inferior to bicubic interpolation. At this stage, analyzing PSNR and spectral signature plots would not provide valuable insights since the qualitative results clearly highlight the limitations of the networks when dealing with a limited dataset. This is a common problem in many HSI datasets, as the available datasets are scarce, providing insufficient learning material for the networks. This limitation will be further investigated and discussed in Chapter 6.

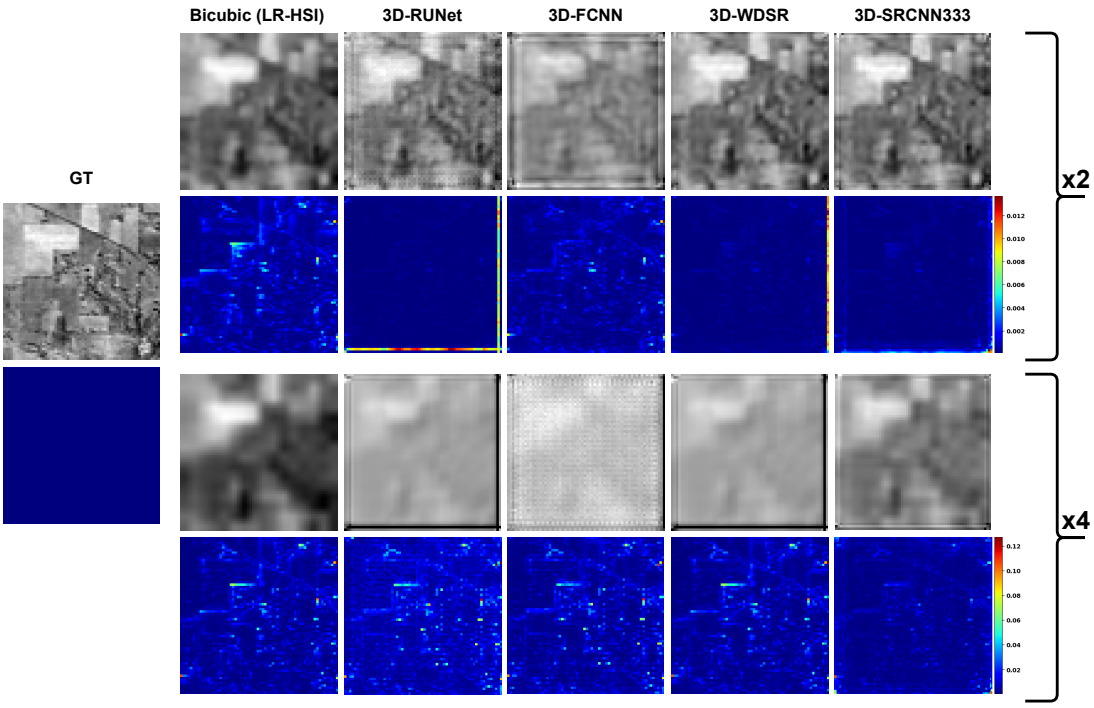


Figure 4.29: Visual results of the predicted HR-HSI for Indian Pines dataset by each method. For each scale factor, the top row shows the output of the method, and the bottom row shows the error map in terms of RMSE. The distorted results caused by the limited size of the dataset highlight the networks' shortcomings without the need to investigate PSNR and spectral signature plots.

4.6.4 Quantitative Comparison

A thorough quantitative summary is presented in Table 4.6, which includes the mean and standard deviation of PSNR, SSIM, and SAM for five experiments, in addition to network size and training time. The table indicates that all quantitative metrics consistently improve or decline across all datasets. This finding will serve as a foundation for the discussion in Section 5.3. In terms of overall performance, 3D-SRCNN333 demonstrates the best results for the Pavia University. For Indian Pines dataset, 3D-SRCNN333 outperforms other networks, but does not outperform bicubic interpolation. This is a case where using traditional methods is more effective than using DCNNs. The performances of 3D-SRCNN333 and 3D-WRCNN are generally very close, but due to its shallower architecture, 3D-SRCNN333 is typically favored since the increase in size from 3D-SRCNN333 to 3D-WRCNN does not yield a significant performance boost.

Among the networks, 3D-FCNN is the smallest, but it is only marginally smaller than 3D-SRCNN333, and the latter offers noticeable improvements in performance.

Table 4.6: Quantitative evaluation between the devised three networks; 3D-SRCNN333, 3D-WRCNN, and 3D-RUNet, compared to other baseline methods in terms of PSNR (dB), SSIM, and SAM (°), in addition to network size and training time in minutes. The results are summarized for Pavia University (PU), Washington DC Mall (CD), and Indian Pines (IP).

Dataset	Bicubic		3D-FCNN		3D-RUNet		3D-WRCNN		3D-SRCNN	
PU (medium)	x2	x4	x2	x4	x2	x4	x2	x4	x2	x4
	27.881 ± 1.411	24.463 ± 3.200	30.940 ± 0.095	25.526 ± 0.014	30.818 ± 0.260	24.385 ± 0.215	31.280 ± 0.116	25.531 ± 0.128	**31.534 ± 0.196**	**25.628 ± 0.027**
	0.75092 ± 0.1479	0.7198 ± 0.2694	0.9145 ± 0.0005	0.6945 ± 0.0002	0.9077 ± 0.0014	0.6082 ± 0.0001	0.9127 ± 0.0001	0.6626 ± 0.0002	**0.9222 ± 0.0009**	**0.6990 ± 0.0003**
	6.03 ± 0.60	9.57 ± 1.45	5.56 ± 0.12	7.07 ± 0.13	6.69 ± 0.36	10.20 ± 0.58	6.69 ± 0.17	10.08 ± 0.25	**5.31 ± 0.15**	**8.54 ± 0.20**
	-		**16.7**		125		83.3		25.6	
DC (large)	29.110 ± 1.68	25.229 ± 1.33	30.509 ± 0.148	27.430 ± 0.121	**32.007 ± 0.121**	**27.598 ± 0.235**	31.197 ± 0.0520	27.582 ± 0.107	31.831 ± 0.012	27.576 ± 0.033
	0.8690 ± 0.0206	0.7120 ± 0.2402	0.9075 ± 0.0012	0.7923 ± 0.0020	**0.9360 ± 0.0010**	**0.7989 ± 0.0015**	0.9135 ± 0.0002	0.7593 ± 0.0003	0.9271 ± 0.0005	0.7955 ± 0.0017
	6.39 ± 1.81	9.941 ± 1.42	6.38 ± 0.22	8.27 ± 0.25	**4.36 ± 0.11**	7.48 ± 0.21	4.65 ± 0.61	8.47 ± 0.65	4.50 ± 0.07	**7.46 ± 0.05**
	-		**58.3**		183.5		108.3		66.7	
IP (small)	35.564 ± 1.482	32.389 ± 1.561	25.394 ± 2.231	22.775 ± 2.942	27.881 ± 1.47	25.970 ± 2.01	28.954 ± 1.402	24.446 ± 1.497	**29.635 ± 2.643**	**32.043 ± 2.011**
	0.9112 ± 0.0211	0.8401 ± 0.0341	0.8551 ± 0.1280	0.8081 ± 0.1154	0.7423 ± 0.1793	0.8069 ± 0.0676	0.8599 ± 0.0144	0.8442 ± 0.1749	**0.8714 ± 0.1328**	**0.8613 ± 0.0229**
	2.73 ± 1.16	3.87 ± 1.82	8.96 ± 2.26	10.23 ± 3.93	6.49 ± 2.41	11.28 ± 2.49	6.33 ± 3.03	15.12 ± 4.11	**4.78 ± 1.84**	**3.83 ± 1.59**
	-		**8.3**		15.2		13.4		10.5	
Network size (# of parameters)			**39,405**		7,350,433		164,609		57,985	

Furthermore, despite 3D-RUNet's impressive performance on the Washington DC Mall dataset, it only slightly surpasses 3D-SRCNN333 while being considerably larger in size, which is not a justifiable trade-off. This observation is depicted in Figure 4.30, illustrating that as the network depth increases, its effectiveness decreases when dealing with small datasets. However, it maintains its efficiency when working with larger datasets. Nevertheless, the marginal improvement in PSNR does not justify the significant increase in network size. Furthermore, it is worth mentioning that 3D-SRCNN showed positive performance more consistently than the other networks in the qualitative evaluation seen in Sections 4.6.2 and 4.6.2. Therefore, 3D-SRCNN333 emerges as the most suitable network due to its balance between quantitative and qualitative performance, compact size, and training time. Since the performance of the network is not perfect, the next chapters will introduce enhancements to the network's parameters and architecture. Additionally, for the Indian Pines dataset, it is of interest to explore whether it is possible to enhance the performance of 3D-SRCNN333 to the extent that it surpasses the performance of bicubic interpolation, particularly considering

the relatively small size of this dataset. This will be explored in Chapter 6.

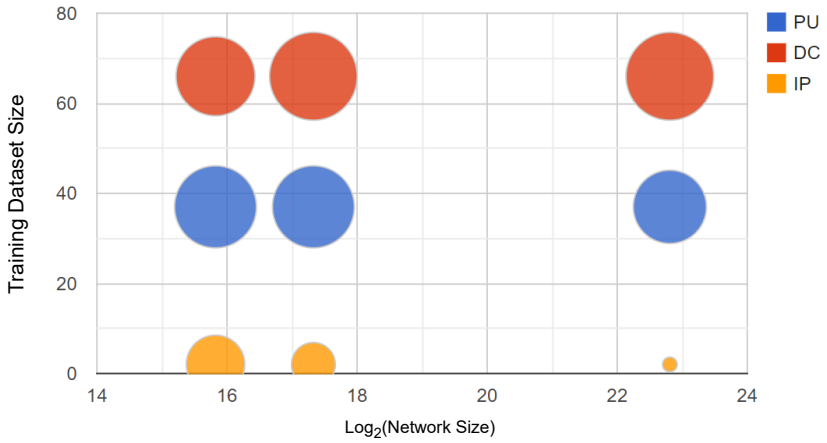


Figure 4.30: Bubble chart of the network size Vs. dataset. The radius of the bubble represents the resulting PSNR by each network for each dataset; Pavia University (PU), Washington DC Mall (DC), and Indian Pines (IP).

4.6.5 Pavia Center Experiments

An essential measure to truly assess the networks' performance is to subject them to testing on a dataset completely independent of their training data. This independence entails introducing new structures or classes that the networks have not encountered during their training phase. In this context, the Pavia Center dataset proves particularly relevant. Captured by the same sensor as the Pavia University dataset, both datasets share similar specifications in terms of spatial and spectral resolution, as detailed in Section 2.4.1 and Table 2.1. Consequently, it becomes intriguing to evaluate how the networks, initially trained on the Pavia University dataset, perform when tested on the Pavia Center dataset. Maintaining consistency, the testing procedure is repeated five times, mirroring the approach applied during training, thereby ensuring the robustness and reliability of the evaluation process.

To commence the analysis, the qualitative results are examined in Figure 4.31. By observing the RMSE maps, it becomes apparent that there is a slightly greater degree of distortion in comparison to Pavia University. Nevertheless, 3D-SRCNN333 continues to demonstrate its capability by capturing more details with relatively few

errors, surpassing the performance of the other networks. Interestingly, the disparities in performance among the networks are more visible in this dataset compared to what was observed with the Pavia University dataset. The performance ranking remains consistent with Pavia University, with 3D-SRCNN333 leading the pack, followed by 3D-WRCNN, 3D-FCNN, and concluding with 3D-RUNet. This order is further corroborated by the PSNR plot presented in Figure 4.32. For the scale factor $\times 2$, there is a substantial overlap between the networks, as evidenced in Figure 4.32a. Nevertheless, 3D-SRCNN333 stands out by achieving higher PSNR values in comparison to the other networks for scale factor $\times 4$ seen in Figure 4.32b.

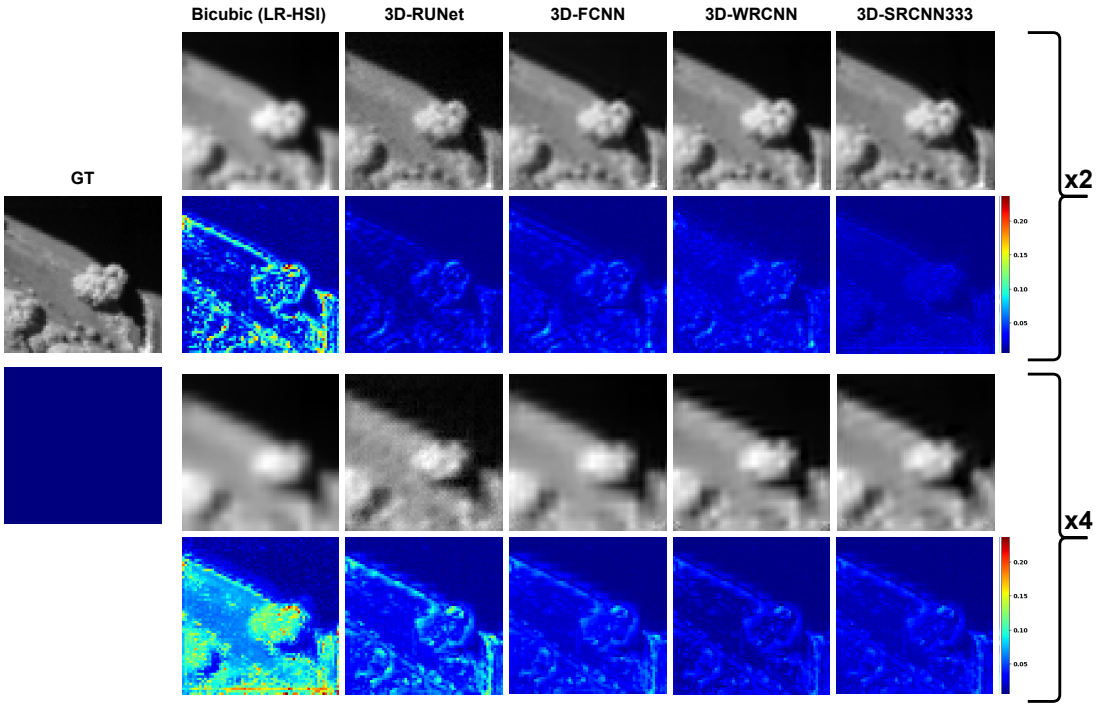
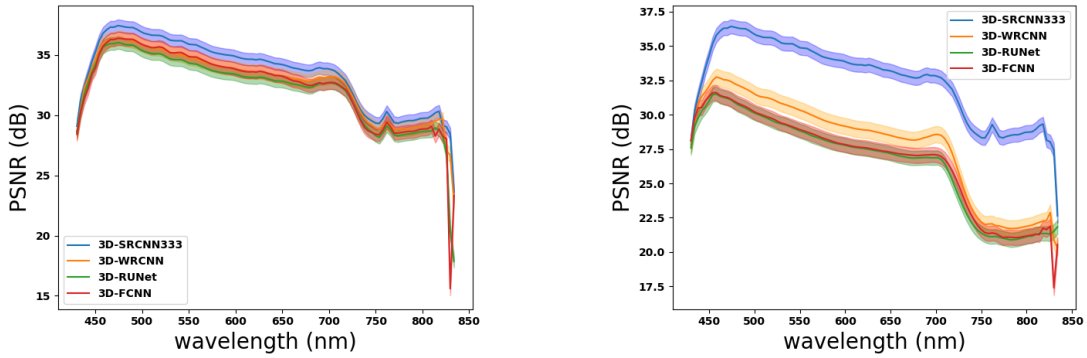


Figure 4.31: Visual results of the predicted HR-HSI for Pavia Center dataset by each method. For each scale factor, the top row shows the output of the method, and the bottom row shows the error map in terms of RMSE. 3D-SRCNN333 shows less distortions compared to the other networks.



(a) ×2 (b) ×4

Figure 4.32: PSNR results recorded for Pavia Center spectral range, where (a) shows the results for scale factor ×2 and (b) shows the results for scale factor ×4. The solid lines represent the mean, while the shaded areas represent the standard deviation.

To assess spectral fidelity, the examination delves into two classes absent from the Pavia University dataset: Water and Tile. Each class is inspected by selecting a uniform region, within which the mean spectral signature is computed. This process is repeated five times for each network, culminating in the recording of mean values and standard deviations.

The results for the Water class are presented in Figure 4.33. Figure 4.33a showcases the region under evaluation. For the scale factor ×2, a slight overlap can be observed between the uncertainty regions of 3D-SRCNN333 and 3D-WRCNN, while a more substantial overlap is evident between 3D-FCNN and 3D-RUNet, as displayed in Figure 4.33b. A deeper analysis via the ratio plot reveals that 3D-SRCNN333 introduces fewer distortions. This trend continues for the scale factor ×4, as seen in Figure 4.33d, where 3D-SRCNN333 stands apart without overlapping with any of the other networks. This distinction is further corroborated by the ratio plot shown in Figure 4.33e. Thus, the order of performance for this class is as follows: 3D-SRCNN333, 3D-WRCNN, 3D-FCNN, and 3D-RUNet.

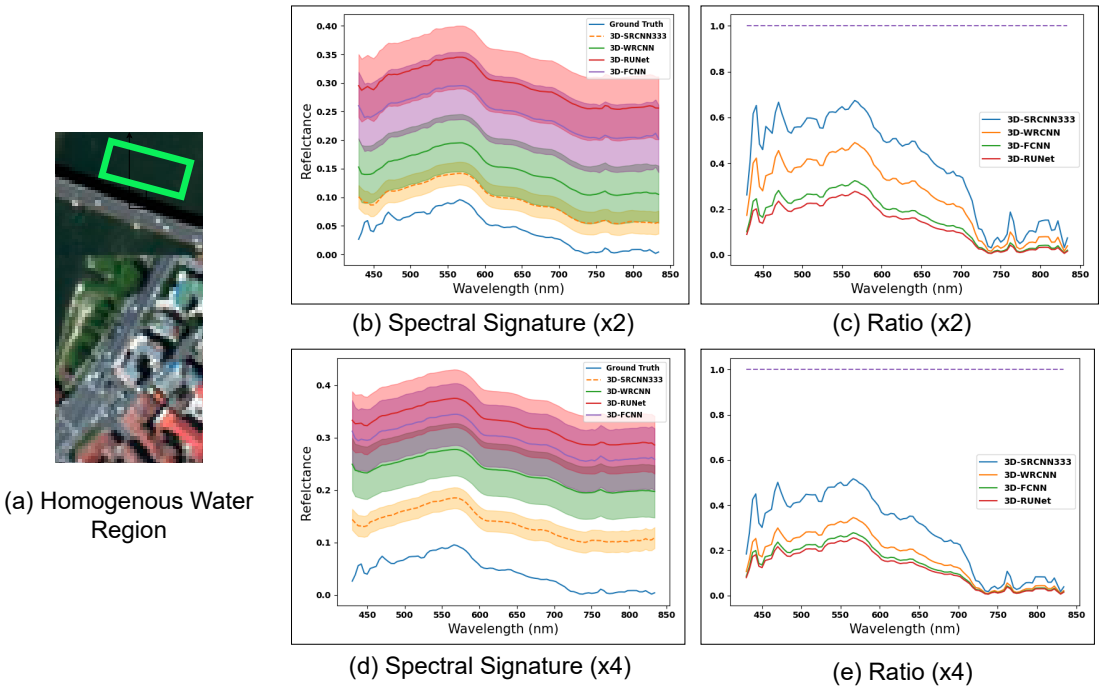


Figure 4.33: The results of testing 3D-SRCNN, 3D-WRCNN, 3D-RUNet, and 3D-FCNN on (a) homogeneous Water region, where the spectral signature is plotted in (b) for scale factor $\times 2$, (c) shows the corresponding ratio plots, (d) shows the spectral signature for scale factor $\times 4$, and (e) shows the corresponding ratio plots. The solid lines represent the mean, while the shaded areas represent the standard deviation. The overall order of performance is: 3D-SRCNN, 3D-WRCNN, 3D-FCNN, and 3D-RUNet.

The second class to be examined is Tile, which is depicted in Figure 4.34a. The observations for the scale factor $\times 2$ results, as shown in Figure 4.34b, echo those of the previous class, with 3D-SRCNN333 exhibiting some overlap with 3D-WRCNN, while a more substantial overlap is observed between 3D-FCNN and 3D-RUNet. Further insights are provided by the ratio plot in Figure 4.34c, which highlights that 3D-SRCNN333 introduces the fewest distortions. Figure 4.34d displays the results for the scale factor $\times 4$, revealing a more significant overlap than scale factor $\times 2$ results. However, the corresponding ratio plot in Figure 4.34e still demonstrates that 3D-SRCNN333 continues to exhibit relatively fewer distortions. In summary, both the Water and Tile classes align with the same performance trend observed in the Pavia University classes, with the order being 3D-SRCNN, 3D-WRCNN, 3D-FCNN, and 3D-RUNet.

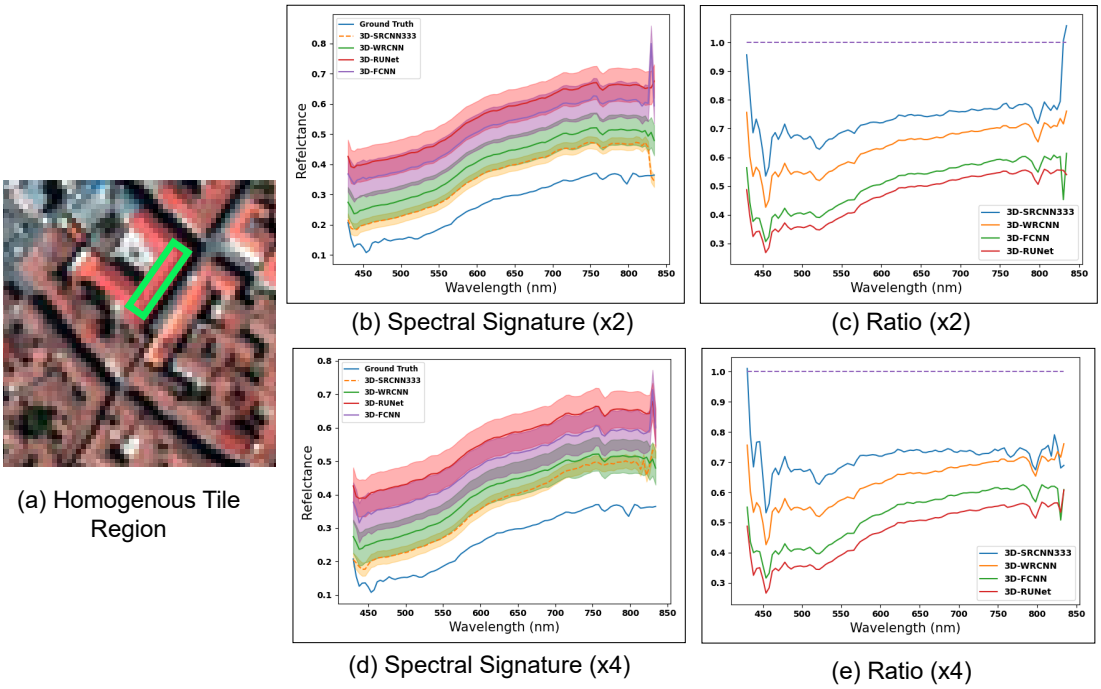


Figure 4.34: The results of testing 3D-SRCNN, 3D-WRCNN, 3D-RUNet, and 3D-FCNN on (a) homogeneous Tile region, where the spectral signature is plotted in (b) for scale factor $\times 2$, (c) shows the corresponding ratio plots, (d) shows the spectral signature for scale factor $\times 4$, and (e) shows the corresponding ratio plots. The solid lines represent the mean, while the shaded areas represent the standard deviation. The overall order of performance is: 3D-SRCNN, 3D-WRCNN, 3D-FCNN, and 3D-RUNet.

Table 4.7 presents the quantitative outcomes of network testing on the Pavia Center dataset, with a focus on PSNR, SSIM, and SAM metrics. These results are in line with the previously discussed qualitative findings and spectral fidelity analysis within this section. 3D-SRCNN333 secures the top position across PSNR, SSIM, and SAM metrics. It is worth noting that all networks exhibit a slight dip in performance when compared to their quantitative results on the Pavia University dataset. While this decline is not substantial, it suggests that the networks cannot achieve perfect generalization.

Table 4.7: Quantitative results of training the networks on Pavia University dataset and testing them on Pavia Center dataset reveal that the order of performance among the networks remains consistent in terms of PSNR, SSIM, and SAM.

Metric	Bicubic		3D-FCNN		3D-RUNet		3D-WRCNN		3D-SRCNN333	
	×2	×4	×2	×4	×2	×4	×2	×4	×2	×4
PSNR (dB)	27.651 ± 1.3684	24.522 ± 2.810	29.822 ± 0.084	25.113 ± 0.022	29.698 ± 0.224	24.006 ± 0.335	30.988 ± 0.012	25.274 ± 0.132	$\mathbf{30.633 \pm 0.151}$	$\mathbf{25.580 \pm 0.121}$
SSIM	0.7538 ± 0.1258	0.6890 ± 0.2557	0.9121 ± 0.0005	0.6879 ± 0.0004	0.9059 ± 0.0011	0.6199 ± 0.0017	0.9105 ± 0.0002	0.6614 ± 0.0004	$\mathbf{0.9180 \pm 0.0001}$	$\mathbf{0.7055 \pm 0.0007}$
SAM (°)	7.84 ± 1.22	9.33 ± 1.42	5.98 ± 0.14	7.11 ± 0.14	6.92 ± 0.45	11.15 ± 0.33	6.54 ± 0.15	9.87 ± 0.20	$\mathbf{5.54 \pm 0.10}$	$\mathbf{8.60 \pm 0.17}$

4.7 Summary

In this chapter, three networks of different depths have been introduced: 3D-SRCNN (shallow), 3D-WRCNN (medium), and 3D-RUNet (deep), with the objective of improving the spatial quality of HSI while preserving their spectral accuracy. These networks were evaluated on datasets of varying sizes: Indian Pines (small), Pavia University (medium), and Washington DC Mall (large). The evaluation involved assessing quantitative metrics such as PSNR, SSIM, and SAM, as well as qualitative analysis by examining the predicted HR-HSI and spectral signatures. Initially, each network was compared against its 2D counterpart, which allowed for observing their behavior with each dataset and demonstrated the superiority of the 3D networks over their 2D counterparts. Subsequently, a comparative analysis among all three networks revealed that 3D-SRCNN333 outperforms the others on Pavia University and Indian Pines datasets, although it slightly lags behind 3D-RUNet on the Washington DC Mall dataset. However, considering the significant difference in computational complexity, the modest improvement does not justify the inefficiency compromise. Consequently, 3D-SRCNN333 emerged as the network with the best overall performance. The limited size of the Indian Pines dataset presented a challenge, as all networks struggled to learn sufficient information from it, resulting in suboptimal performance. However, 3D-SRCNN333 managed to capture more details compared to the other networks. Additionally, 3D-SRCNN333 showed excellent performance on Pavia Center dataset. The issue of data scarcity in Pavia University and Indian Pines datasets is addressed in Chapter 6. Additionally, all the networks employed an MSE loss function that focused solely on spatial

errors and did not account for spectral errors. To address this limitation, Chapter 5 introduces a hybrid loss function tested on 3D-SRCNN333 that addresses both spatial and spectral errors. The findings of this chapter have been published at International Conference on Signal Processing and Information Security (ICSPIS) [50] and IEEE International Geoscience and Remote Sensing Symposium (IGARSS) [335], in addition to an accepted paper at IEEE International Conference on Image Processing (ICIP) 2023 [341].

Chapter 5

Loss Functions

5.1 Introduction

In the preceding chapter, various architectures were investigated; however, the significance of the loss function in addition to the architecture design cannot be overlooked. A well-designed loss function has the potential to enhance the accuracy and efficiency of a model, making it a crucial aspect in ML research and practice. Specifically in the context of HSI-SISR, comprehending the role and types of loss functions is essential. The primary objective of a loss function is to quantify the error between each band of the GT HSI and the estimated HSI, thereby minimizing this error during the training process of the 3D-CNN, which in this case is the 3D-SRCNN333 explained in Section 4.3 of Chapter 4. In SISR-oriented DCNNs, the common choice for loss functions are MSE or MAE. These loss functions measure the spatial discrepancy on a pixel-by-pixel basis between the GT HSI and the estimated HSI, and hence, they are referred to as spatial loss functions in this thesis. In contrast, spectral loss functions pertain to the comparison of spectral signature vectors in a space with a dimensionality equal to the number of bands. Unlike spatial loss functions, spectral loss functions have not been widely employed for HSI-SISR. One example of incorporating spectral loss is seen in [148], where the authors utilized SAM as an integral part of the loss function. However, most of the parameters were set empirically through trial-and-error. Additionally, the authors did not explore other spatial loss functions that could perform better than

MSE and MAE. This gives rise to two important questions: 1) which spatial function achieves the best performance for HSI-SISR, and 2) whether it is feasible to devise a loss function that optimally combines both spatial and spectral aspects. This chapter aims to address these questions by systematically reviewing commonly used spatial and spectral loss functions in the literature, with a focus on evaluating their performance in terms of PSNR, SSIM, and SAM. The outcomes of this review will be utilized to develop a Bayesian-optimized hybrid spectral-spatial loss function that outperforms standard loss functions.

5.2 Conventional Loss Functions for SISR

5.2.1 Spatial Loss Functions

The most straight forward and the most widely used loss function is MSE, which measures the sum of squared differences between every pixel at location (i, j) of band k in $Y^k_{(i,j)}$ and $\hat{Y}^k_{(i,j)}$. MSE is expressed by Equation 5.1. Measuring the absolute differences instead of square differences is the main distinction between MSE and MAE, also known as L1, which is expressed in Equation 5.2. As MSE squares the error, it penalizes large errors more severely than MAE, which in turn makes MSE more sensitive to data outliers.

Mean Squared Log Error (MSLE) is a variation of MSE that calculates the log of the ratio between $\mathbf{Y}$ and $\hat{\mathbf{Y}}$, as seen in Equation 5.3. While MSE and MAE are sensitive to large errors, MSLE is sensitive to small errors. MSLE can be even more sensitive to outliers than MSE because of the logarithm function. Thus, MSLE is particularly useful when the data has a wide range of values and the model needs to be sensitive to small values as well as large values.

$$\begin{aligned} R(\mathbf{Y}_{(i,j)}, \widehat{\mathbf{Y}}_{(i,j)}) &= \sum_{k=1}^{B} Y^k_{(i,j)} - \hat{Y}^k_{(i,j)} \\ L_{MSE} &= \frac{1}{M \times N \times B} \sum_{i=1}^{M} \sum_{j=1}^{N} R(\mathbf{Y}_{(i,j)}, \widehat{\mathbf{Y}}_{(i,j)})^2 \end{aligned} \tag{5.1}$$

$$L_{MAE} = \frac{1}{M \times N \times B} \sum_{i=1}^{M} \sum_{j=1}^{N} \left| R(\mathbf{Y}_{(i,j)}, \widehat{\mathbf{Y}}_{(i,j)}) \right|^2 \tag{5.2}$$

$$L_{MSLE} = \frac{1}{M \times N \times B} \sum_{k=1}^{B} \sum_{i=1}^{M} \sum_{j=1}^{N} \left(log \left(\frac{Y(i,j)+1}{\hat{Y}(i,j)+1} \right) \right)^2 \tag{5.3}$$

Huber [342] is a piece-wise loss function that was devised to be resilient to outliers. It is a hybrid combination of MSE and MAE, such that if the residual error R is less than or equal to a certain threshold δ, the loss is expressed by MSE, otherwise it is expressed by MAE, as seen in Equation 5.4. The challenge is to choose an optimal δ value, as it is dependent on the dataset inliers. Cross-validation is often used to find the optimal δ.

$$L_{Huber} = \frac{1}{M \times N \times B} \sum_{i=1}^{M} \sum_{j=1}^{N} \begin{cases} \frac{1}{2} R(\mathbf{Y}_{(i,j)}, \widehat{\mathbf{Y}}_{(i,j)})^2, & \text{if } |R(i,j)| \leq \delta \\ \delta \left(|R(\mathbf{Y}_{(i,j)}, \widehat{\mathbf{Y}}_{(i,j)})| - \frac{1}{2}\delta \right), & \text{otherwise} \end{cases} \tag{5.4}$$

Log Hyperbolic Cosine (LHC) loss was first proposed in [343], where the authors utilized this function to improve the performance of Variational Auto-Encoders. Even though LHC is not a piece-wise function, its performance is close to that of Huber. However, it lacks the adaptability of Huber, as δ is fixed in LHC. Additionally, LHC differs from MSE in the sense that it does not get affected by the occasional large errors.

$$L_{LHC} = \frac{1}{M \times N \times B} \sum_{i=1}^{M} \sum_{j=1}^{N} log(cosh(R(\mathbf{Y}_{(i,j)}, \widehat{\mathbf{Y}}_{(i,j)})) \tag{5.5}$$

Charbonnier loss function, originally proposed in [344], is a variant of MAE and it was adapted for HSI-SR in [46]. The authors argue that Charbonnier loss is more resilient to outliers and provides more performance improvement over MAE. Charbonnier loss is expressed in Equation 5.6. The standard value of ϵ is $10^{-}3$.

$$L_{Ch} = \sum_{i=1}^{M} \sum_{j=1}^{N} \sqrt{R(\mathbf{Y}_{(i,j)}, \widehat{\mathbf{Y}}_{(i,j)})^2 + \epsilon^2} \tag{5.6}$$

5.2.2 Spectral Loss

As explained in Chapter 1, HSI's distinctive quality lies in their high spectral resolution, which must be preserved while enhancing their spatial resolution. One loss function that takes spectral resolution into consideration is CS, which is expressed in Equation 5.7. CS measures the similarity between the GT vector $\mathbf{y}$ of pixel values at position (i, j) and the estimated vector $\hat{\mathbf{y}}$ of pixel values at the same position across all bands.

$$L_{CS} = -\frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{\sum_{k=1}^{B} Y^k(i,j) \hat{Y}^k(i,j)}{\sqrt{\sum_{k=1}^{B} Y^k(i,j)^2} \sqrt{\sum_{k=1}^{B} \hat{Y}^k(i,j)^2}} \tag{5.7}$$

SAM offers an alternative approach to evaluate spectral loss. As explained in Section 2.5.3, CS is computed as $CS = \cos(SAM)$. Notably, SAM and CS represent the foremost methods for robustly gauging spectral loss in Deep Convolutional Neural Networks (DCNNs) [148]. Also, as stated by [345], CS and SAM exhibit similar behavior. Consequently, the focus centers on L_{CS} as the primary metric for assessing spectral loss. This selection allows delving into the resulting implications when combined with spatial loss, shedding light on the emergent behavior of this fusion.

It is important to acknowledge that CS, and consequently SAM, are not flawless metrics. Their response does not linearly correlate with the difference between the spectra under comparison. Instead, their response reaches a maximum value and saturates at a certain point [345].

5.3 Proposed Bayesian-optimized Hybrid Spectral-Spatial Loss Function

Intuitively, one would argue that a hybrid loss function that combines spatial and spectral losses is the ideal solution to provide the best of both worlds. Therefore, a hybrid loss function is proposed, which combines the top-performing spatial loss with the CS loss. The formulation of this hybrid loss function is presented in Equation 5.8.

$$L_{Hybrid} = \alpha L_s - (1 - \alpha) L_{CS}, \tag{5.8}$$

where the hyperparameter $\alpha \in [0, 1]$ is dataset-dependent. The choice of L_s will depend on the spatial loss function that demonstrates the best quantitative results, as determined by the experimental results discussed in Section 5.5.

α is a parameter that must be tuned in order to obtain the best possible result from the loss function. If $\alpha = 1$, L_{Hybrid} becomes purely spatial, as the spectral loss does not contribute to the function. The opposite is true when $\alpha = 0$. Bayesian Optimization Algorithm (BOA) is one way to find the optimal α. 3D-SRCNN333 with the hybrid loss function can be considered as a black box objective function. Ideally, PSNR and SSIM must be maximized, and SAM must be minimized. However, BOA can maximize or minimize one metric at a time. The experiments and analysis in Chapter 4 reveal that PSNR, SSIM, and SAM mostly improve or deteriorate in unison. Therefore, it is safe to assume that boosting one metric automatically boosts the others. For this study, the focus will be on maximizing PSNR. Thus, BOA can be used as a probabilistic framework to find the α that offers the global maximum.

The fundamental idea behind BOA is to create a rough surrogate model of the objective function, $f^*(\alpha)$, and then use the model to decide on the following point to evaluate. BOA creates a method that can identify the ideal value of a non-convex function with a manageable number of evaluations since it makes use of all the data from prior assessments of $f^*(\alpha)$. Gaussian Process (GP) is the most commonly used probabilistic model for BOA due to its flexibility [346], and it is expressed as follows:

$$f^*(\alpha) \simeq GP(\mu(\alpha), \sigma(\alpha, \alpha')) \tag{5.9}$$

The prior over the functions and the acquisition function are the two key decisions that must be considered during the optimization process. While the acquisition function is used to find the next best point for evaluation, most likely to reduce uncertainty in the function's possible values, the prior conveys assumptions or provides information about the function being optimized. The GP is fit onto $x = \alpha$ and $y = f^*(\alpha)$, such that GP serves as the surrogate model for $f^*(\alpha)$. The acquisition function then uses the GP to predict how $f^*(\alpha)$ varies with α in order to identify which value leads to the largest GP. The most commonly used acquisition function is Upper Confidence Bound

(UCB) [346], which is defined as

$$UCB(\alpha^*) = \mu(\alpha^*) + \gamma^{1/2}\sigma(\alpha^*), \tag{5.10}$$

where $\mu(\alpha^*)$ represents exploitation regions of a specific value α^*, $\sigma(\alpha^*)$ represents exploration regions, and γ is a parameter that balances exploitation and exploration. Often large values of γ are encouraged. In this research $\gamma = 2.0$, such that it provides balance between good values and unexplored areas. This process is iterated 50 times at maximum. Early stopper strategy is adopted, such that the optimization stops if α does not improve PSNR for 10 consecutive iterations. The iterative steps of GP-UCB are defined as follows:

1. Initialize the GP model with prior mean function $\mu(\alpha^*)$ and standard deviation function $\sigma(\alpha^*)$, as seen in equation 5.9.
2. For $t = 1, 2, \ldots, T$:
   (a) Compute the acquisition function UCB(α_t) seen in Equation 5.10 for each point α_t in the search space, where $\mu(\alpha_t)$ and $\sigma(\alpha_t)$ are the mean and standard deviation of the GP model at α_t, respectively.
   (b) Select the next point α_{t+1} to evaluate by maximizing the acquisition function:
   $$\alpha_{t+1} = argmax_{\alpha_t} UCB(\alpha_t) \tag{5.11}$$
   (c) Evaluate the function $f^*(\alpha_{t+1})$.
   (d) Update the GP model with the new data $(\alpha_{t+1}, f^*(\alpha_{t+1}))$.
   (e) Update the current iteration: $t = t + 1$.
3. Return the best point α found so far.

5.4 Experimental Setup

3D-SRCNN333 is trained with the six spatial loss functions explained in Section 5.2, namely MSE, MAE, MSLE, LHC, Huber, and Charbonnier, in addition to CS as a

spectral loss function. The objective is to determine which of the spatial loss functions achieves the optimum PSNR and SSIM, and to prove that CS achieves the best spectral fidelity compared to the spatial loss functions. Based on the results, the hybrid loss function is devised and tested to prove its superiority against the other standard loss functions. BOA is also used to determine the optimal hyperparameter α of the loss function that achieves the best results. These experiments are performed on Pavia University and Washington DC Mall datasets. For a precise evaluation of the influence of the loss functions, it is crucial to keep a consistent set of hyperparameters. This method ensures not only the separation of the effects of each loss function but also assures a fair and impartial comparison among all the functions. The training parameters used for all the experiments in this chapter are listed in Table 5.1.

Table 5.1: Training parameters for all the experiments conducted in Chapter 5. The architecture is 3D-SRCNN333.

Training parameter	Value	
Optimization function	Adam	
Epochs	Early stopper (max 1000)	
Learning rate	Scheduler: [10^{-3}, 10^{-4}, 10^{-5}]	
Dataset	Pavia University	Washington DC Mall
Training	36	64
Validation	4	8
Testing	5	8

Much like the learning parameters applied in Chapter 4 (Section 4.2), a scheduler is employed to fine-tune the learning rate. Initially, it begins with the default learning

rate of 10^{-3} provided by the Adam optimizer. However, if there are no improvements for three consecutive epochs or if the improvement is less than 10^{-5}, the scheduler reduces the learning rate. The minimum allowable learning rate is set at 10^{-5}, and the scheduler will not further decrease it. Additionally, an early stopping strategy is integrated to determine the optimal number of epochs. Training is halted if no improvements are observed or if the improvement is less than 10^{-5} for five consecutive epochs. Furthermore, the data split mirrors that of the previous chapter, with both Pavia University and Washington DC Mall datasets being divided into 80% for training and the remaining 20% for testing and validation.

Each of the loss functions discussed in this chapter undergoes a comprehensive evaluation process, consisting of five separate experiments. Within each experiment, the dataset is shuffled to ensure that the patches allocated for training, validation, and testing differ on every occasion. The reported findings and graphical representations in this study are representative of the means calculated from all five experiments, along with their corresponding standard deviations.

5.5 Results and Analysis

This section showcases the performance of the loss functions on both the Pavia University dataset in the reflectance domain and the Washington DC dataset in the radiance domain. An intriguing aspect to explore is the robustness of the performance between radiance and reflectance. Furthermore, the algorithms trained on the Pavia University dataset are tested on the Pavia Center dataset to assess their behavior when confronted with endmembers that were not encountered during the training process.

5.5.1 Pavia University Experiments

The first set of experiments are performed on Pavia University dataset. Table 5.2 shows a summary of Pavia University quantitative results for scale factors $\times 2$ and $\times 4$. The table shows PSNR, SSIM, and SAM obtained from each loss function.

Table 5.2: Comparison between the performance of spatial and spectral loss functions as well as the devised hybrid loss function on 3D-SRCNN333 trained on Pavia University dataset for scales factors $\times 2$ and $\times 4$. The α values that yielded the optimum performance for $\times 2$ and $\times 4$ for five experiments are indicated with reference to Table C.1-C.5. Δ indicates the difference between the hybrid loss function and the best performing function.

Loss function	$\times 2$ (α = 0.743334, 0.975429, 0.743320, 0.634721, 0.592714)			$\times 4$ (α = 0.970302, 0.683219, 0.843267, 0.970288, 0.927126)		
	PSNR (dB)	SSIM	SAM (°)	PSNR (dB)	SSIM	SAM (°)
MSE	31.534 ± 0.196	0.9222 ± 0.0009	5.56 ± 0.15	25.628 ± 0.027	0.6990 ± 0.0003	8.54 ± 0.20
MAE	31.573 ± 0.161	0.9216 ± 0.0002	5.56 ± 0.10	25.580 ± 0.016	0.6944 ± 0.0011	8.55 ± 0.24
MSLE	31.488 ± 0.128	0.9224 ± 0.0003	5.79 ± 0.12	25.434 ± 0.144	0.6959 ± 0.0001	8.67 ± 0.16
Huber	31.526 ± 0.151	0.9221 ± 0.0003	5.90 ± 0.16	25.533 ± 0.043	0.6955 ± 0.0014	8.85 ± 0.08
LHC	31.531 ± 0.059	0.9219 ± 0.0004	5.90 ± 0.23	25.538 ± 0.083	0.6948 ± 0.0009	8.88 ± 0.19
Charbonnier	$\mathbf{31.605 \pm 0.101}$	0.9248 ± 0.0001	5.67 ± 0.13	$\mathbf{25.689 \pm 0.131}$	$\mathbf{0.6964 \pm 0.0007}$	8.62 ± 0.11
CS	8.867 ± 3.564	0.4702 ± 0.2882	$\mathbf{5.38 \pm 0.09}$	4.108 ± 2.464	0.2587 ± 0.1239	$\mathbf{8.40 \pm 0.04}$
Hybrid	$\mathbf{31.628 \pm 0.011}$	$\mathbf{0.9338 \pm 0.0002}$	$\mathbf{5.40 \pm 0.07}$	$\mathbf{25.696 \pm 0.023}$	$\mathbf{0.6996 \pm 0.0002}$	$\mathbf{8.45 \pm 0.17}$
Δ	0.023 ↑	0.009 ↑	0.16 ↓	0.007 ↑	0.0032 ↑	0.09 ↓

Upon examining the results for scale factor $\times 2$, the variations between the spatial loss functions are generally minor. Charbonnier achieves the highest PSNR and SSIM scores, while MSLE performs the poorest in terms of these metrics. MAE obtains the highest SAM score, but Charbonnier is selected as the spatial loss function for the hybrid loss formulation due to its superior spatial performance. CS, as a spectral loss function, does not excel in spatial metrics but achieves the highest SAM score. Moving on to scale factor $\times 4$, the disparities between the spatial loss functions remain relatively minor. Charbonnier once again emerges as the top spatial loss function, delivering the highest PSNR and SSIM scores. CS maintains its dominance in terms of SAM. Therefore, based on the results from both scale factors, CS and Charbonnier are chosen to construct the hybrid loss function. In Equation 5.8, L_s is substituted with L_{Ch}. While Charbonnier has a lower bound but no upper bound ($0 \leq L_{Ch} < \infty$), CS is a bounded function with $-1 \leq L_{CS} \leq 1$. To align their lower bounds, L_{CS} can be adjusted to $0 \leq L_{CS} + 1 \leq 2$. As a result, Equation 5.8 is revised as follows:

$$L_{Hybrid} = \alpha L_{Ch} + (1 - \alpha)(1 + L_{CS}) \tag{5.12}$$

BOA is then employed to find the optimal value of α that maximizes PSNR, which is expected to result in the highest SSIM and the lowest SAM, as concluded in Chapter 4. Since five experiments are conducted, each experiment produces a different α. All BOA tables related to Pavia University dataset are listed in Appendix C.1. As an example, Table C.1 presents the values of α and their corresponding results for the first 15 iterations of scale factors $\times 2$ and $\times 4$. The table reveals that BOA successfully identifies the optimal α in the fourth iteration for scale factor $\times 2$, resulting in an optimum PSNR of 33.037 dB and an optimum SSIM of 0.9275. However, the SAM value of 4.22° does not exhibit the highest or lowest performance. The best SAM is achieved on the twelfth iteration, but its PSNR and SSIM are inferior to those of the third iteration. This inconsistency in the qualitative metrics contradicts the consistent improvements observed in Chapter 4. Ultimately, the choice between the α values from iterations 3 and 12 depends on the intended purpose of enhancing the spatial quality of HSI. This demonstrates that achieving the best PSNR does not necessarily correspond to obtaining the best SAM. However, it still outperforms the conventional spatial loss functions in terms of all quantitative metrics, as seen in Table 5.2. Regarding scale factor $\times 4$, the highest PSNR and SSIM are attained in the 5^{th} iteration. Unlike the case of scale factor $\times 2$, this result also yields the best SAM. With reference to Tables C.2-C.5, the optimal α values for scale factor $\times 2$ produced from the remaining four experiments are 0.975429, 0.743320, 0.634721, and 0.592714. As for scale factor $\times 4$, the optimal α values are 0.683219, 0.843267, 0.970288, 0.927126. An interesting observation is that the choice of α often leans towards giving more weight to the spatial function, although there are instances where it strikes a nearly equal balance between both functions, as evidenced by the values 0.634721 and 0.592714. Notably, none of the experiments favored a higher emphasis on the spectral loss over the spatial one, as α consistently remained above 0.5. This observation suggests that there are scenarios where the spectral loss alone may not be necessary, and the spatial loss can be sufficient by itself. Nevertheless, the presence of the spectral loss component does not appear

to negatively impact the overall performance of the hybrid loss function. Instead, the proposed hybrid loss function seems to possess the capability to discern when to incorporate spectral loss and when it can be omitted, adapting flexibly to the specific requirements of the task at hand.

Figure 5.1 provides a visual representation of the results obtained using MSE, which was previously employed in Chapter 4, Charbonnier, which achieved the best PSNR and SSIM among the spatial loss functions, and the newly developed Bayesian-optimized hybrid loss function. The discrepancies between the results are most clearly observed in the RMSE map, where it is evident that the hybrid loss function produces fewer errors compared to the other loss functions.

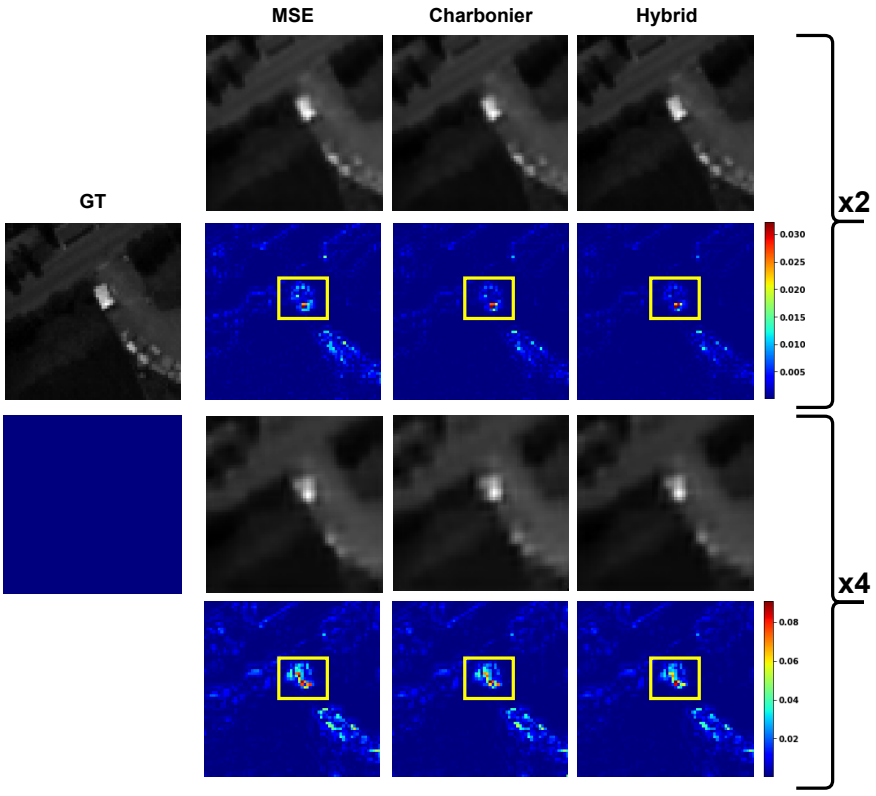
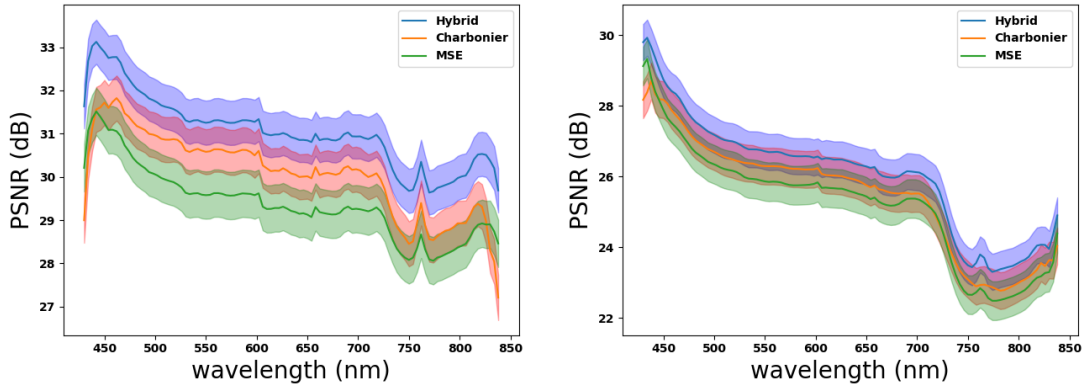


Figure 5.1: Visual results of the predicted HR-HSI for Pavia University using MSE, Charbonnier, and hybrid loss functions. For each scale factor, the top row shows the output of the method, and the bottom row shows the error map in terms of RMSE.

Additionally, the PSNR plots depicted in Figure 5.2 indicate that, for both scale factors $\times 2$ and $\times 4$, the hybrid loss function yields the highest PSNR across all bands,

surpassing MSE and Charbonnier. The PSNR improvements introduced by the hybrid loss function are less apparent in scale factor ×4 compared to scale factor ×2.



(a) ×2 (b) ×4

Figure 5.2: PSNR results recorded for Pavia University spectral range, where (a) shows the results for scale factor ×2 and (b) shows the results for scale factor ×4. The solid lines represent the mean, while the shaded areas represent the standard deviation.

Inspecting the spectral signature requires looking at signature of different classes to observe the performance on each one. As with the previous chapter, the inspected classes are: Metal Sheet, Gravel, and Asphalt. The results are plotted for a homogeneous region, and the mean result of the pixels within that region is recorded. Each method is experimented five times, thus, the evaluation is done the same way five times, and then the average and uncertainty in terms of standard deviation are recorded.

To begin the examination of spectral fidelity, the Metal Sheet class is inspected. The results for scale factors ×2 and ×4 within the homogeneous region, as depicted in Figure 5.3a, are showcased in Figure 5.3b-e. When the scale factor ×2 is considered, the spectral signatures, as observed in Figure 5.3b, reveal a large degree of overlap between the MSE, Charbonnier, and the hybrid loss functions across the entire spectrum range. This overlap persists up to a wavelength of 600 nm, beyond which the hybrid loss function distinguishes itself by remaining closer to the GT.

Figure 5.3c portrays the ratio of the predicted spectral signatures to the GT, affirming the substantial overlap between MSE and Charbonnier. In certain instances, MSE closely approximates 1, while Charbonnier does so in others. Notably, this plot

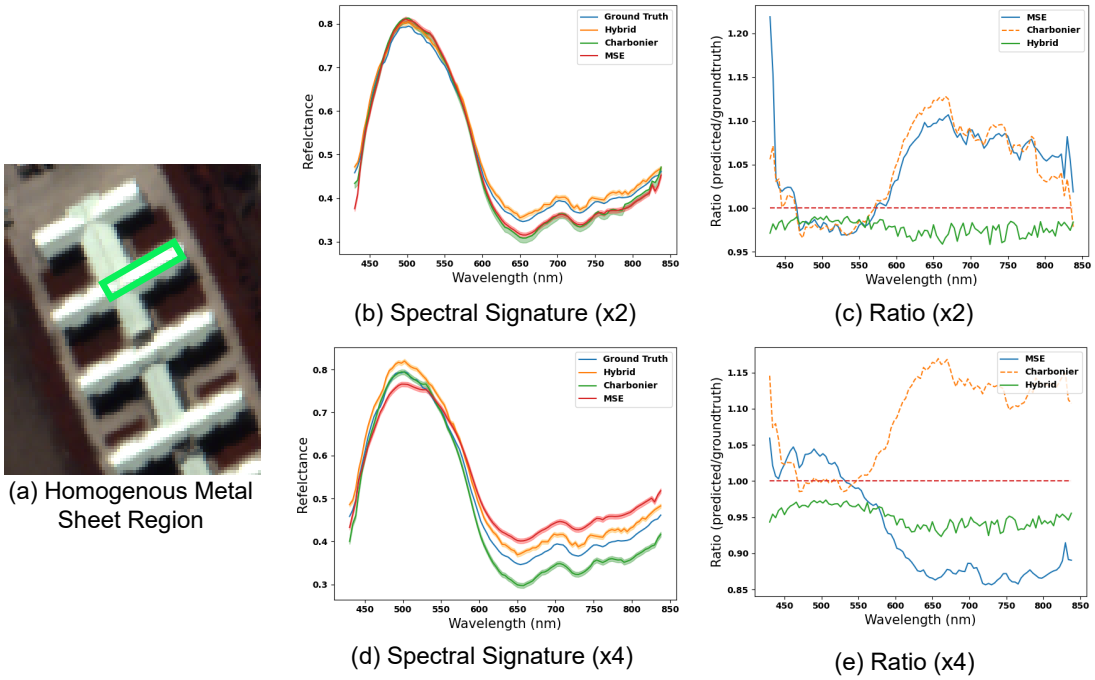


Figure 5.3: The results of testing 3D-SRCNN333 with MSE, Charbonnier, and the proposed hybrid loss function on (a) homogeneous Metal Sheet region, where the spectral signature is plotted in (b) for scale factor $\times 2$, (c) shows the corresponding ratio plots, (d) shows the spectral signature for scale factor $\times 4$, and (e) shows the corresponding ratio plots. The overall order of performance is: hybrid loss, Charbonnier, and MSE.

accentuates the fact that, especially from a wavelength of 600 nm onward, the hybrid loss function closely tracks the GT in contrast to MSE and Charbonnier.

By inspecting the results of scale factor $\times 4$, as illustrated in Figure 5.3d, it becomes apparent that larger distortions and greater disparities between the outcomes of the functions are observed in comparison to the $\times 2$ scale factor. The ratio plot, as depicted in Figure 5.3e, elucidates that the hybrid loss function continues to exhibit proximity to 1 when contrasted with the other two functions, with the exception of the region approximately between 450 nm and 550 nm. In this particular range, Charbonnier aligns more closely with the GT than the hybrid loss function.

In summation, it is noteworthy that the hybrid loss function effectively rectified numerous distortions identified in the performance of 3D-SRCNN333 on the Metal Sheet class, as discussed in Section 4.6.1.

Regarding Gravel class analysis, the outcomes pertaining to the homogeneous re-

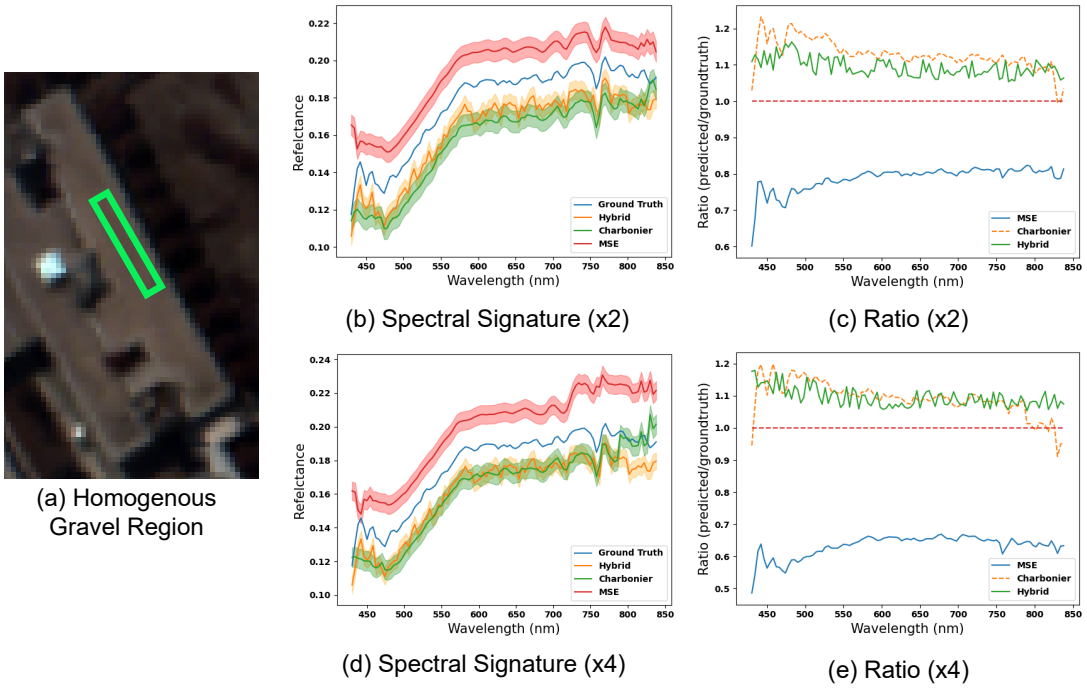


Figure 5.4: The results of testing 3D-SRCNN333 with MSE, Charbonnier, and the proposed hybrid loss function on (a) homogeneous Gravel region, where the spectral signature is plotted in (b) for scale factor $\times 2$, (c) shows the corresponding ratio plots, (d) shows the spectral signature for scale factor $\times 4$, and (e) shows the corresponding ratio plots. This particular class shows high overlap between the performances of the hybrid loss function and Charbonnier, with both of them surpassing MSE.

gion depicted in Figure 5.4a come into view through Figures 5.4b-e. Within this class, a heightened degree of overlap is observed between the Charbonnier and hybrid loss functions for both scale factors, namely $\times 2$ and $\times 4$. Notably, it is worth emphasizing that both the Charbonnier and hybrid loss functions outperform the Mean MSE. This distinction is readily observable when scrutinizing the spectral signature plots found in Figures 5.4b and 5.4d. Moreover, this performance advantage over MSE is further substantiated by the ratio plots presented in Figures 5.4c and 5.4e. Figure 5.4c accentuates that, for scale factor $\times 2$, the hybrid loss function exhibits a slight improvement over the Charbonnier. On the other hand, the overlap between these two functions becomes even more pronounced for scale factor $\times 4$, as observed in Figure 5.4e.

Finally, the performance of the loss functions on the homogeneous Asphalt region depicted in Figure 5.5 is assessed. The resulting evaluations are visually conveyed

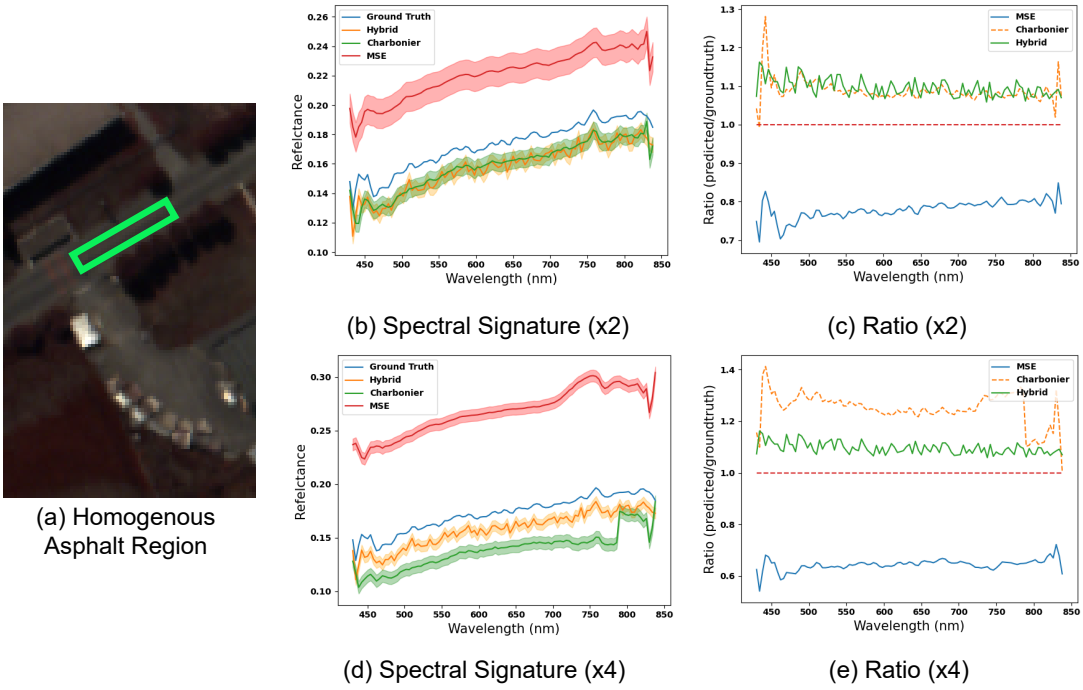


Figure 5.5: The results of testing 3D-SRCNN333 with MSE, Charbonnier, and the proposed hybrid loss function on (a) homogeneous Asphalt region, where the spectral signature is plotted in (b) for scale factor $\times 2$, (c) shows the corresponding ratio plots, (d) shows the spectral signature for scale factor $\times 4$, and (e) shows the corresponding ratio plots. This particular class shows high overlap between the performances of the hybrid loss function and Charbonnier for scale factor $\times 2$, while the hybrid loss function surpasses Charbonnier for scale factor $\times 4$. Both function surpass MSE in terms of performance.

through Figures 5.5b-e. When considering the scale factor $\times 2$, Charbonnier and the hybrid loss functions once again exhibit a substantial degree of overlap, as corroborated by the ratio plot in Figure 5.5c. It is noteworthy that both of these loss functions continue to outperform MSE. In contrast, the advantage of the hybrid loss function becomes more pronounced for scale factor $\times 4$, as evidenced in Figure 5.5d, and this distinction is further reflected in Figure 5.5e. Overall, the performance hierarchy within this class can be listed as follows: hybrid loss, Charbonnier, then MSE.

In summary, the spectral analysis reveals that the hybrid loss function demonstrates acceptable performance enhancement in the reflectance domain, effectively addressing some of the errors that were previously observed when using the 3D-SRCNN333 trained with the MSE loss function.

5.5.2 Washington DC Mall Experiments

The second set of experiments are conducted using the Washington DC Mall dataset, and the results are summarized in Table 5.3 to compare all the spatial and spectral loss functions. In contrast to the Pavia University dataset, the differences between the outcomes of the loss functions are more pronounced. For both scale factors $\times 2$ and $\times 4$, the LHC loss function performs the poorest in terms of PSNR and SSIM. Additionally, the MSE results are noticeably inferior to those of MAE and MSLE, with the latter two achieving higher PSNR by 0.392 dB and 0.406 dB, respectively, for scale factor $\times 2$. Similar observations can be made for scale factor $\times 4$, where MAE and MSLE outperform MSE by 0.104 dB and 0.114 dB, respectively. Similar to the results obtained from the Pavia University dataset, Charbonnier attains the highest PSNR and SAM for both scale factors $\times 2$ and $\times 4$ in the case of the Washington DC Mall dataset. However, in this scenario, Charbonnier also obtains the highest SAM among the spatial loss functions. Hence, it is considered the optimal loss function for constructing the hybrid loss function. As for CS, its PSNR and SSIM are low, which align with expectations, but it achieves the highest SAM. Therefore, Equation 5.12 can be reused for this experiment.

Table 5.3: Comparison between the performance of spatial and spectral loss functions as well as the devised hybrid loss function on 3D-SRCNN333 trained on Washington DC Mall dataset. The α values that yielded the optimum performance for $\times 2$ and $\times 4$ for five experiments are indicated with reference to Table C.6-C.10. Δ indicates the difference between the hybrid loss function and the best performing function.

Loss function	$\times 2$ ($\alpha = 0.975415, 0.507654, 0.628743, 0.621345, 0.722345$)			$\times 4$ ($\alpha = 0.743334, 0.941264, 0.812340, 0.805432, 0.654232$)		
	PSNR (dB)	SSIM	SAM (°)	PSNR (dB)	SSIM	SAM (°)
MSE	31.831 ± 0.012	0.9271 ± 0.0005	4.50 ± 0.07	27.576 ± 0.033	0.7955 ± 0.0017	7.46 ± 0.05
MAE	32.223 ± 0.015	0.9323 ± 0.0002	4.36 ± 0.05	27.680 ± 0.054	0.7988 ± 0.0008	7.22 ± 0.10
MSLE	32.237 ± 0.016	0.9327 ± 0.0001	4.36 ± 0.06	27.692 ± 0.024	0.7983 ± 0.0020	7.25 ± 0.09
Huber	32.134 ± 0.011	0.9314 ± 0.0009	4.44 ± 0.05	27.576 ± 0.039	0.7972 ± 0.0012	7.40 ± 0.06
LHC	31.980 ± 0.026	0.9301 ± 0.0002	4.51 ± 0.02	27.574 ± 0.023	0.7970 ± 0.0013	7.49 ± 0.13
Charbonnier	$\mathbf{32.262 \pm 0.012}$	$\mathbf{0.9330 \pm 0.0002}$	4.29 ± 0.05	$\mathbf{27.854 \pm 0.020}$	$\mathbf{0.8005 \pm 0.00016}$	7.06 ± 0.08
CS	11.788 ± 4.30	0.6673 ± 0.1053	$\mathbf{4.13 \pm 0.01}$	8.903 ± 4.51	0.4235 ± 0.1960	$\mathbf{6.85 \pm 0.04}$
Hybrid	$\mathbf{32.302 \pm 0.022}$	$\mathbf{0.9332 \pm 0.0003}$	$\mathbf{4.28 \pm 0.02}$	$\mathbf{27.990 \pm 0.031}$	$\mathbf{0.8012 \pm 0.00010}$	$\mathbf{6.92 \pm 0.14}$
Δ	$0.04 \uparrow$	$0.0002 \uparrow$	$0.15 \uparrow$	$0.136 \uparrow$	$0.0007 \uparrow$	$0.14 \downarrow$

Since five experiments are conducted, each experiment produces a different α. All BOA tables related to Washington DC Mall dataset are listed in Appendix C.2. As an example, Table C.6 presents the outcomes obtained with different values of α during the initial 15 iterations. For scale factor $\times 2$, the optimal value of α is 0.9754156, which is achieved in the 3rd iteration. In contrast to the Pavia University experiment, this iteration consistently yields the highest PSNR and SSIM, as well as the lowest SAM. Regarding scale factor $\times 4$, the best α value is 0.743334, coinciding with the 3rd iteration once again. Moreover, this value leads to the highest PSNR, SSIM, and SAM. Therefore, in the case of the Washington DC Mall dataset, the devised hybrid loss function, in conjunction with BOA, successfully identified the optimal α value and combined the favorable characteristics of the spatial and spectral loss functions. Referring to Tables C.7-C.10, the optimal α values for scale factor $\times 2$ obtained from the remaining four experiments are as follows: 0.507654, 0.628743, 0.621345, and 0.722345. When considering the scale factor $\times 4$, the optimal α values are 0.941268, 0.812340, 0.805432, and 0.654232. The behavior of α values in the Washington DC Mall case differs from the Pavia University case. In the Washington DC Mall case, the α values

tend to gravitate towards achieving a balance between spectral and spatial losses more often than in the Pavia University case. Specifically, out of the 10 optimal α values, 6 of them fall within the interval (0.5, 0.8), signifying that the spectral loss contributes to the optimization process by approximately $30 - 50\%$. For the remaining 4 values, the emphasis leans more heavily towards the spatial loss, as α falls within the interval (0.8, 1.0). Once again, this highlights the flexibility of the hybrid loss function, which demonstrates the ability to discern when to incorporate spectral loss and when to omit it, thereby adapting to the specific demands of the task at hand.

The visual results depicted in Figure 5.6 align with the quantitative observations. The error map of the output generated by training 3D-SRCNN333 with the hybrid loss function exhibits fewer errors compared to MSE and Charbonnier for both scale factors $\times 2$ and $\times 4$. The differences are more pronounced in the $\times 4$ case, where MSE exhibits more blurriness and artifacts, while the hybrid loss function produces a sharper result.

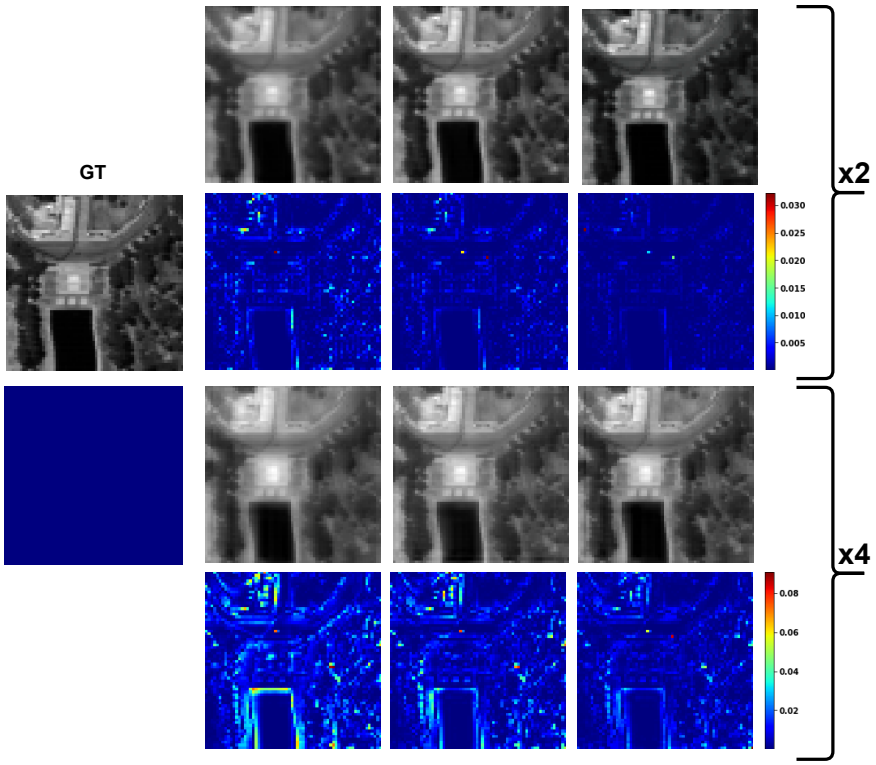
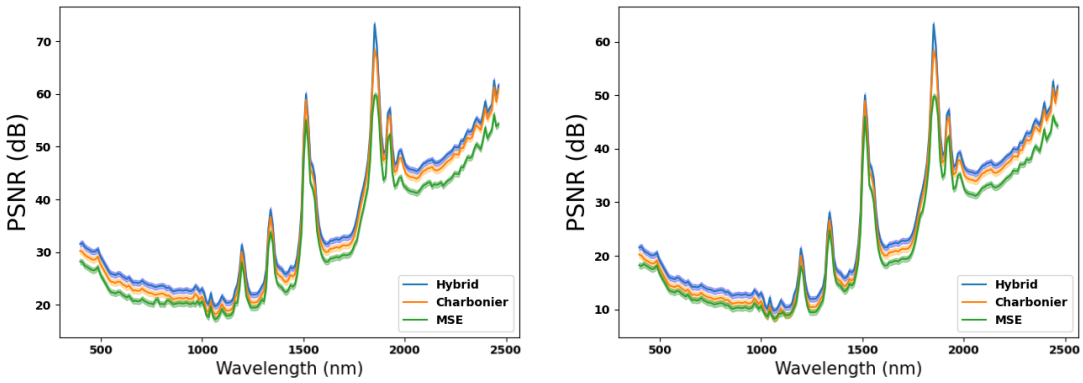


Figure 5.6: Visual results of the predicted HR-HSI for Washington DC Mall using MSE, Charbonnier, and hybrid loss functions. For each scale factor, the top row shows the output of the method, and the bottom row shows the error map in terms of RMSE.

Additionally, the PSNR plots shown in Figure 5.7 indicate that the hybrid loss function achieves the highest PSNR for scale factor $\times 2$, although the disparities between the plots are minimal. In the case of scale factor $\times 4$, the degradation in MSE becomes more noticeable, while the hybrid loss function demonstrates a more stable performance.



(a) $\times 2$ (b) $\times 4$

Figure 5.7: PSNR results recorded for Washington DC Mall spectral range, where (a) shows the results for scale factor $\times 2$ and (b) shows the results for scale factor $\times 4$. The solid lines represent the mean, while the shaded areas represent the standard deviation.

To assess the impact of the newly developed hybrid loss function on spectral fidelity within the Washington DC Mall dataset, an analysis is conducted across three classes, mirroring the approach from the previous chapter: Grass, Roof, and Water. The evaluation involves plotting the results within a uniform region, with subsequent recording of the mean values derived from the pixels within that region. Each method undergoes five experimental runs, ensuring consistency in the evaluation process across these repetitions. The final evaluation is then based on the averaged results with the corresponding uncertainty expressed as the standard deviation.

The outcomes pertaining to the Grass class are depicted in Figure 5.8. Within the region illustrated in Figure 5.8a, the spectral signatures predicted by each network are presented in Figure 5.8b for scale factor $\times 2$. Notably, the significant overlap between these plots necessitates a closer examination through the ratio plot displayed in Figure 5.8c to discern which function yields the most favorable results. Charbonnier loss function exhibits a few spikes deviating from the ideal value of 1, while the performance

of both MSE and the hybrid loss function remains fairly consistent. Transitioning to the scale factor ×4, the overlap in results persists, but the ratio plot in Figure 5.8e reveals that the hybrid loss function predominantly exhibits fewer errors compared to MSE and Charbonnier.

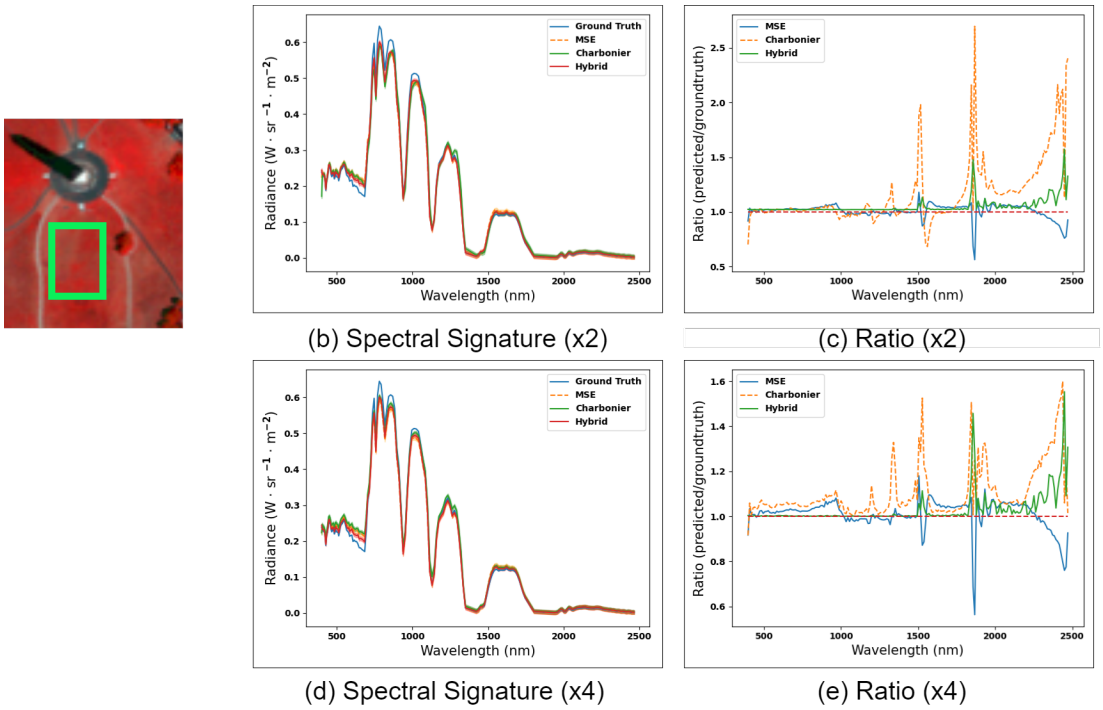
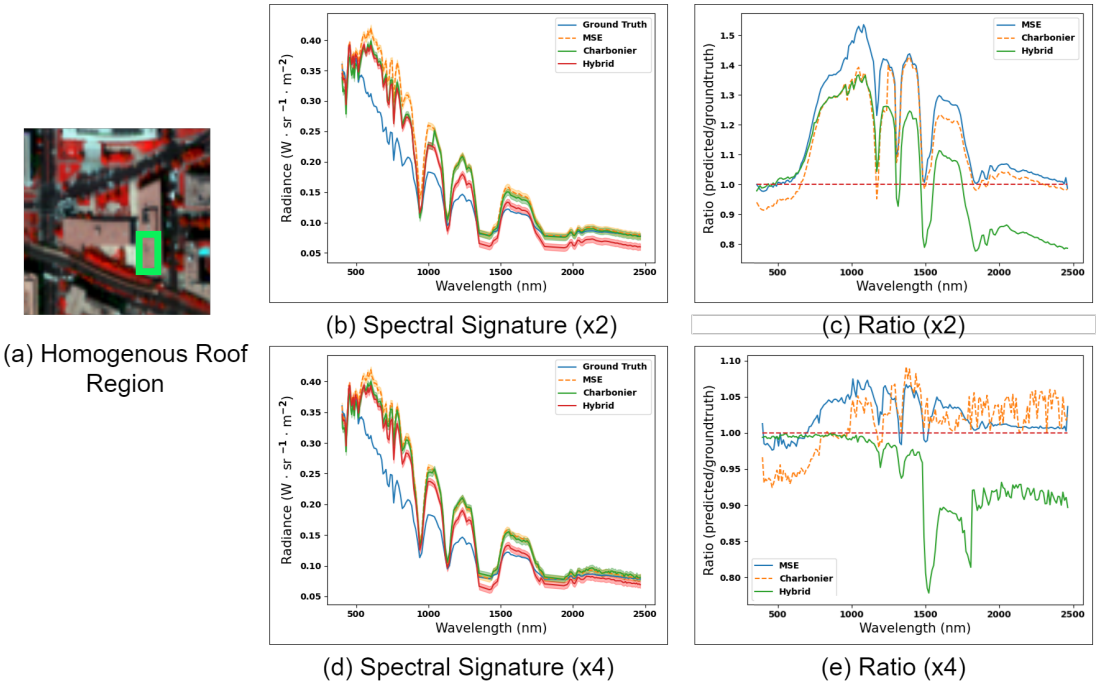


Figure 5.8: The results of testing 3D-SRCNN333 with MSE, Charbonnier, and the proposed hybrid loss function on (a) homogeneous Grass region, where the spectral signature is plotted in (b) for scale factor ×2, (c) shows the corresponding ratio plots, (d) shows the spectral signature for scale factor ×4, and (e) shows the corresponding ratio plots. The high overlap is visible between both scale factors. The ratio plots illustrate that the hybrid loss function shows less distortions than MSE and Charbonnier.

Turning attention to the Roof class, the results are presented in Figure 5.9. There is a reduced level of overlap in this class when compared to the Grass class. For the scale factor ×2 results depicted in Figure 5.9b, it becomes evident that the hybrid loss function closely tracks the GT in comparison to Charbonnier and MSE. This observation is corroborated by the ratio plot in Figure 5.9c. However, it is worth noting that the performance of the hybrid loss function exhibits a decline beyond approximately wavelength 1800 nm, as indicated by the same ratio plot. A similar trend is observed for the scale factor ×4, as illustrated in Figure 5.9d, where the hybrid loss function main-

tains a relatively closer alignment with the GT, albeit with more distortion compared to the scale factor ×2 scenario. The ratio plot in Figure 5.9e further emphasizes this performance trend, showing that the hybrid loss function performs well up to around 1500 nm, after which it experiences a decline in fidelity.



(a) Homogenous Roof Region

(b) Spectral Signature (x2)

(c) Ratio (x2)

(d) Spectral Signature (x4)

(e) Ratio (x4)

Figure 5.9: The results of testing 3D-SRCNN333 with MSE, Charbonnier, and the proposed hybrid loss function on (a) homogeneous Grass region, where the spectral signature is plotted in (b) for scale factor ×2, (c) shows the corresponding ratio plots, (d) shows the spectral signature for scale factor ×4, and (e) shows the corresponding ratio plots. The hybrid loss function shows more distortions in this class, often falling behind MSE and Charbonnier.

Finally, the results for the Water class are presented in Figure 5.10. At a scale factor of ×2, the overlap between the functions remains high, as illustrated in Figure 5.10b. However, upon closer examination via the corresponding ratio plot in Figure 5.10c, it becomes evident that the hybrid loss function introduces fewer distortions overall in comparison to both MSE and Charbonnier. When the scale factor increases to ×4, the overall overlap between the functions is notably reduced, as depicted in Figure 5.10d. The ratio plots further clarify this trend. Except for the wavelength range of approximately 1500 nm to 2000 nm, the hybrid loss function consistently exhibits fewer

distortions than the other two functions.

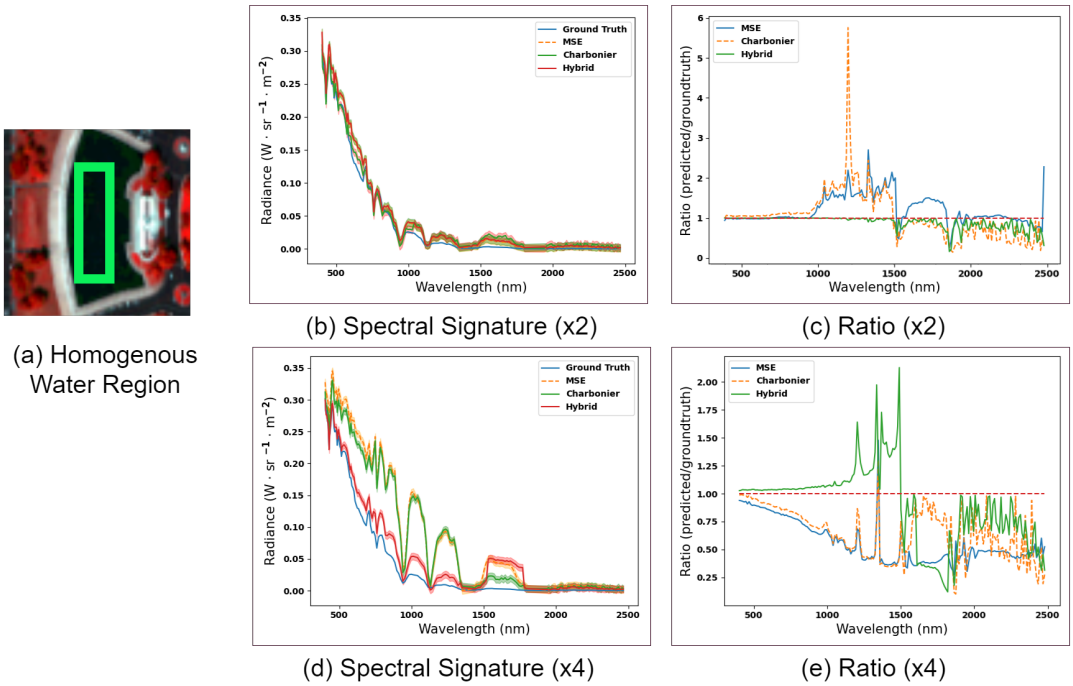


Figure 5.10: The results of testing 3D-SRCNN333 with MSE, Charbonnier, and the proposed hybrid loss function on (a) homogeneous Water region, where the spectral signature is plotted in (b) for scale factor $\times 2$, (c) shows the corresponding ratio plots, (d) shows the spectral signature for scale factor $\times 4$, and (e) shows the corresponding ratio plots. The hybrid loss function shows better performance than MSE and Charbonnier in scale factor $\times 2$, but falls behind between approximately 1200-1500 nm for scale factor $\times 4$.

Overall, the hybrid loss function effectively corrects certain errors that were previously evident in the 3D-SRCNN333 model trained with the MSE loss function. Its ability to do so in the radiance domain highlights its robustness.

5.5.3 Pavia Center Experiments

Similar to the analysis conducted in the previous chapter, it is imperative to subject the loss functions being studied to testing on a dataset that is entirely distinct from their training data. To this end, the Pavia Center dataset will be used as a testing dataset. 3D-SRCNN333 along with the investigated loss functions, originally trained on the Pavia University dataset, are tested on Pavia Center dataset. Recall from Section 2.4.1

and Table 2.1 that Pavia Center dataset and Pavia University dataset share the same characteristics.

The qualitative findings are presented in Figure 5.11. The hybrid loss function showcases improved visual results within the region outlined by the yellow box on the RMSE map, extending its advantage outside of this region as well. However, within areas delineated by the red boxes, it becomes evident that either Charbonnier or MSE manages to capture details more effectively than the hybrid loss function. When examining the PSNR plots in Figure 5.12a, it becomes apparent that both the hybrid loss function and Charbonnier consistently outperform MSE in terms of PSNR. However, Figure 5.12b reveals a high degree of overlap between them, suggesting that the hybrid loss function may encounter challenges in rectifying certain distortions.

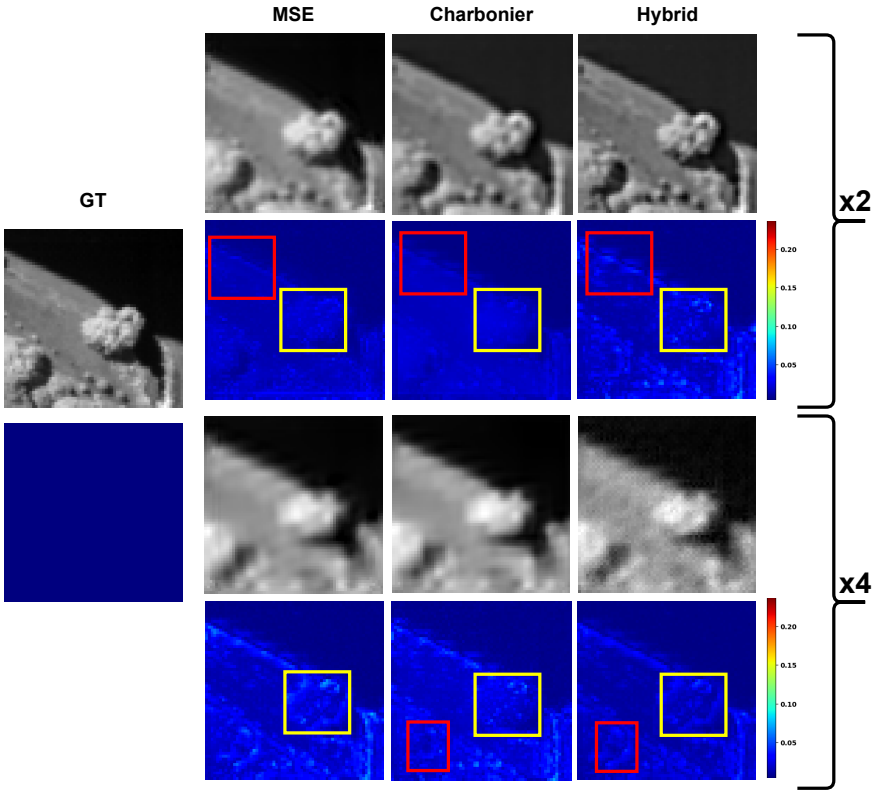
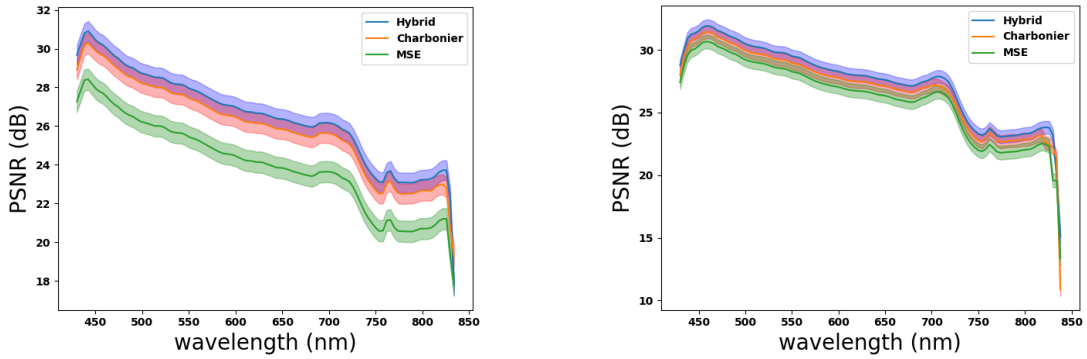


Figure 5.11: Visual results of the predicted HR-HSI for Pavia Center dataset by each method. For each scale factor, the top row shows the output of the method, and the bottom row shows the error map in terms of RMSE. The yellow box shows the region where the hybrid loss function prevails, while the red box shows the region where it falls behind MSE and Charbonnier.



(a) ×2 (b) ×4

Figure 5.12: PSNR results recorded for Pavia Center spectral range, where (a) shows the results for scale factor ×2 and (b) shows the results for scale factor ×4. The solid lines represent the mean, while the shaded areas represent the standard deviation.

In order to evaluate spectral accuracy, this analysis focuses on two categories that are not present in the Pavia University dataset: Water and Tile. For each of these classes, the examination involves the selection of a consistent region where the average spectral signature is calculated. This procedure is conducted five times for each network, resulting in the collection of mean values and standard deviations.

The findings pertaining to the Water class are depicted in Figure 5.13. Figure 5.13a shows the region subject to evaluation. In this class, a noteworthy pattern emerges, demonstrating the closely matched performance of the hybrid loss function and Charbonnier, both surpassing MSE. This trend is evident in both the scale factor ×2 results, as illustrated in Figure 5.13b, and the accompanying ratio plot in Figure 5.13c. A similar observation persists for the scale factor ×4, as depicted in Figures 5.13d and 5.13e. Overall, it becomes evident that the distinction between Charbonnier and the hybrid loss function is relatively subtle within this class.

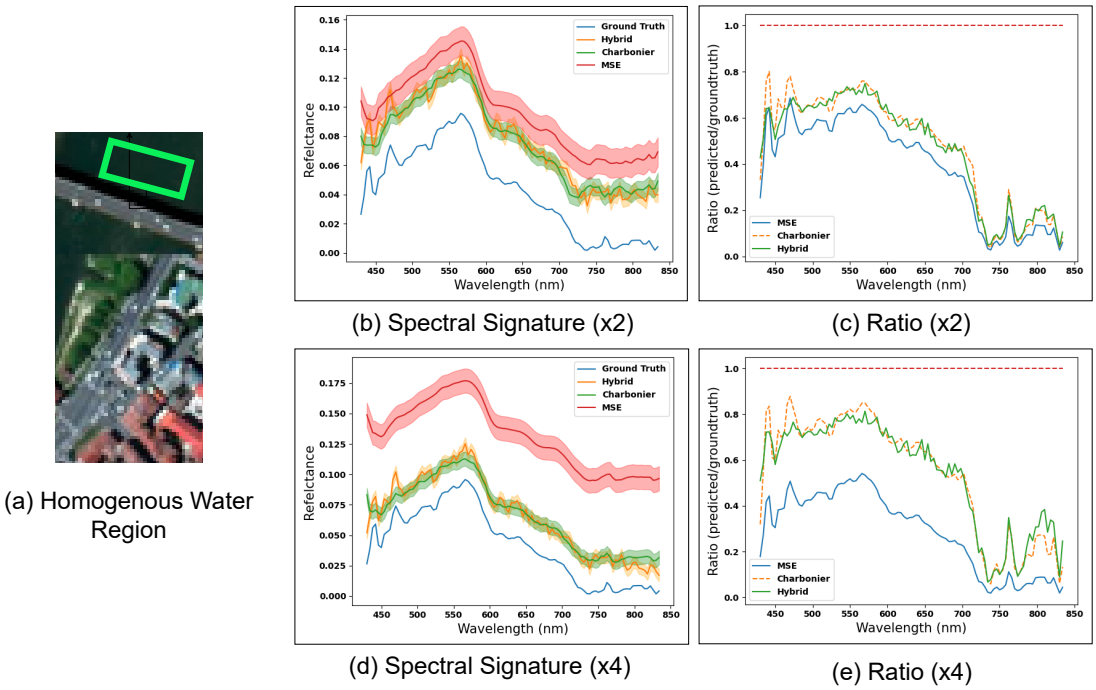


Figure 5.13: The results of testing MSE, Charbonnier, and the hybrid loss function on (a) homogeneous Water region, where the spectral signature is plotted in (b) for scale factor $\times 2$, (c) shows the corresponding ratio plots, (d) shows the spectral signature for scale factor $\times 4$, and (e) shows the corresponding ratio plots. This class shows high overlap between Charbonnier and the hybrid loss function.

The second class to be examined is Tile, which is depicted in Figure 5.14a. Unlike Water class, the hybrid loss function advantage is more visible here. For example, scale factor $\times 2$ results seen in Figure 5.14b show that the hybrid loss function resembles the GT more closely than Charbonnier and MSE. This is confirmed by looking at the ratio plot in Figure 5.14c. The overlap between the hybrid and Charbonnier loss function is higher in scale factor $\times 4$, as seen in Figure 5.14d and confirmed by the ratio plot in Figure 5.14e. The hybrid loss function is only slightly better than Charbonnier, and both rectify most of the errors from MSE.

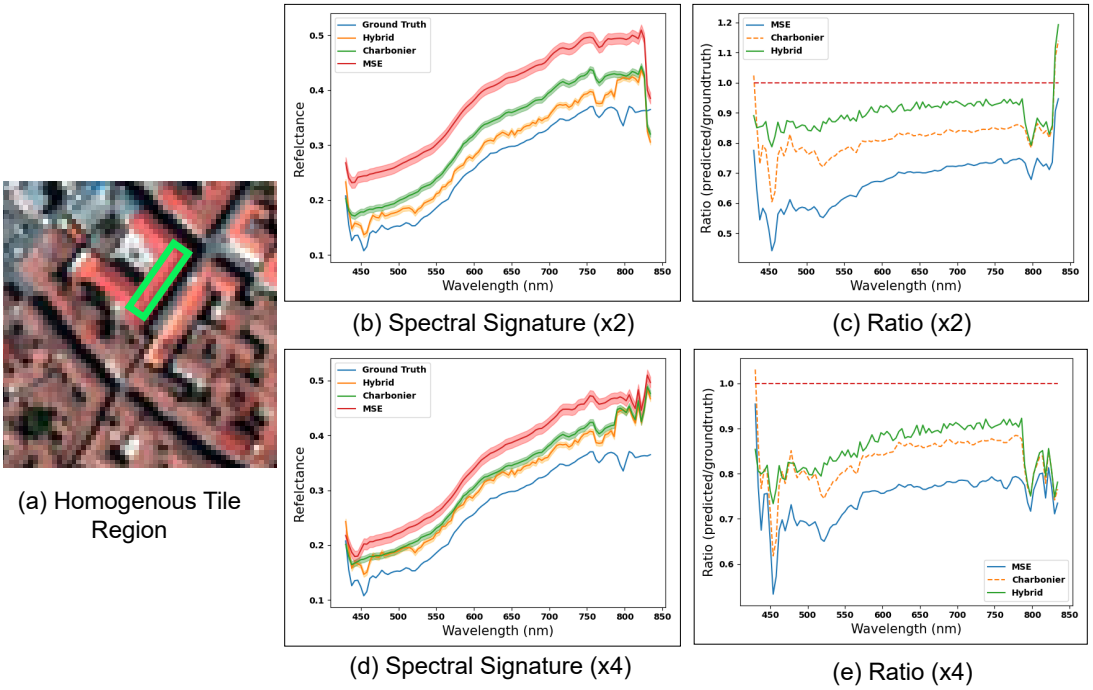


Figure 5.14: The results of testing MSE, Charbonnier, and the hybrid loss function on (a) homogeneous Water region, where the spectral signature is plotted in (b) for scale factor $\times 2$, (c) shows the corresponding ratio plots, (d) shows the spectral signature for scale factor $\times 4$, and (e) shows the corresponding ratio plots. This class shows the advantage of the hybrid loss function more clearly.

Table 5.4 presents the quantitative results of network testing on the Pavia Center dataset, emphasizing PSNR, SSIM, and SAM metrics. These results align with the qualitative observations discussed earlier and the spectral fidelity analysis presented in this section. Overall, the performance overlap between the hybrid loss and Charbonnier is high, but they both surpass MSE. In the table, the bottom row labeled Δ offers insight into the extent of improvement achieved by the hybrid loss function compared to the best-performing loss function for each respective metric. These improvements are modest but higher compared to the enhancements observed in the Pavia University dataset.

Table 5.4: Quantitative results of training 3D-SRCNN333 with MSE, Charbonnier, and Hybrid loss functions using Pavia University dataset and then testing on Pavia Center dataset reveal that the order of performance remains consistent.

Loss Function	×2			×4		
	PSNR (dB)	SSIM	SAM (∘)	PSNR (dB)	SSIM	SAM (∘)
MSE	30.633 ± 0.151	0.9180 ± 0.0001	5.54 ± 0.10	25.580 ± 0.121	0.7055 ± 0.0007	8.60 ± 0.17
Charbonnier	30.921 ± 0.098	0.9200 ± 0.0001	5.41 ± 0.07	27.805 ± 0.118	0.7103 ± 0.0004	8.32 ± 0.20
Hybrid	$\mathbf{31.101 \pm 0.075}$	$\mathbf{0.9251 \pm 0.0001}$	$\mathbf{5.33 \pm 0.10}$	$\mathbf{27.995 \pm 0.103}$	$\mathbf{0.7140 \pm 0.0002}$	$\mathbf{8.28 \pm 0.12}$
Δ	$0.18 \uparrow$	$0.0051 \uparrow$	$0.08 \downarrow$	$0.19 \uparrow$	$0.0037 \uparrow$	$0.04 \downarrow$

5.6 Summary

This chapter presented a Bayesian optimized hybrid spectral-spatial loss function that offers the best of both types of loss functions. The function's goal is to maximize PSNR and SSIM while ensuring minimal spectral distortions by minimizing SAM. The hyperparameter α determines the weight on the spatial loss Vs. the spectral loss. Experiments of standard loss functions on both Pavia University and Washington DC Mall datasets indicate that Charbonnier loss function achieves the highest PSNR and lowest SAM. The only type of spectral loss function available is CS. Thus, the hybrid loss function comprises of Charbonnier as a spatial loss and CS as a spectral loss. BOA optimizes α for each dataset. Experiments on Pavia University dataset show that α achieves the optimal PSNR and SSIM, but not necessarily the best SAM. On the other hand, Washington DC Mall experiments shows that the optimal α leads to the highest PSNR and SSIM, as well as the lowest SAM. In cases where spectral and spatial metrics are in conflict, choosing α depends on the purpose of enhancing the HSI. The hybrid loss function's ability to perform well on Pavia Center dataset and to generally remain consistent within both reflectance and radiance domains highlight its robustness. The findings of this chapters have been published in the European Signal Processing Conference (EUSIPCO) [347]. Furthermore, these findings have been accepted at the upcoming IEEE International Conference on Image Processing (ICIP) in October 2023 [341].

Chapter 6

Data Augmentation

6.1 Introduction

Data Augmentation is a technique employed in ML and DL to artificially expand and diversify a training dataset by generating modified versions of the original data. The objective of Data Augmentation is to enhance the model's ability to generalize effectively by providing supplementary data examples that capture the same underlying concepts as the original data but with minor variations. There are different types of Data Augmentation techniques that can be applied at different levels. Pixel-level Data Augmentation techniques, as discussed in [348], are beneficial for high-level tasks like classification. On the other hand, feature-level Data Augmentation techniques, such as MixUp and DropOut [349], are more suitable for certain tasks but not advantageous for low-level tasks like SISR. In the case of SISR, these techniques can significantly alter the essential features of the data, leading to a decline in performance [348]. Hence, geometric Data Augmentation techniques are typically recommended for SISR. Geometric Data Augmentation involves applying various transformations to the original data, including rotation, scaling, translation, flipping, cropping, adding noise, adjusting contrast, and changing brightness levels [348]. The specific choice of Data Augmentation techniques depends on the type of data and the particular task at hand. By applying these transformations to the original data, new data samples can be generated, which can be used to train the ML model. The additional data samples obtained through

Data Augmentation can help mitigate issues like overfitting and vanishing gradients by providing more diverse examples that can contribute positively to the model's learning procedure.

HSI-SISR faces a significant challenge due to the limited availability of training data, especially in the case of smaller datasets like Indian Pines compared to Pavia University and Washington DC Mall. Therefore, it becomes crucial to address the data scarcity problem in HSI and find ways to increase the dataset size to improve the generalization ability of the network. While Data Augmentation techniques have been extensively used in HSI classification models, their application in the context of HSI-SISR has not been thoroughly investigated. To effectively apply Data Augmentation techniques in HSI-SISR, it is important to conduct systematic experiments and determine which methods do not distort the spectral signature of the HSI data. The evaluation criteria for an effective Data Augmentation method in HSI-SISR involve enhancing PSNR and SSIM metrics while minimizing SAM. This chapter proposes a new Data Augmentation method called CutMixBlur, which will be tested for its effectiveness in enhancing HSI. Additionally, this chapter explores the impact of applying effective Data Augmentation techniques using two randomized approaches: Random Compound and Random Stacking. These approaches aim to further investigate the potential benefits of augmenting the HSI data for improved performance in HSI-SISR.

6.2 SISR Augmentation Techniques

The following subsections introduce image transformation methods that are commonly used for Data Augmentation purposes, explain a new Data Augmentation method named CutMixBlur, and test each method's effect on HSI-SISR in terms of spatial enhancement as well as preserving spectral fidelity.

6.2.1 Flipping/Mirroring

Flipping, also called mirroring [350], can be in the form of horizontal or vertical flips, which are expressed in Equations 6.1 and 6.2, respectively, and illustrated in Figure

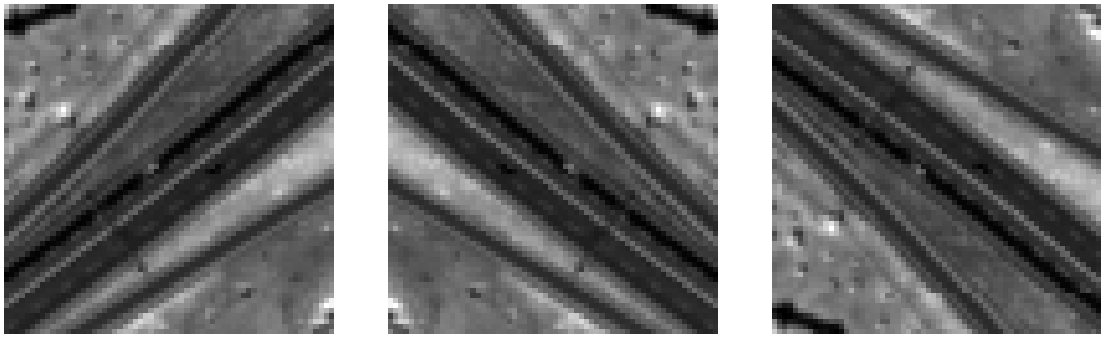
6.1.

$$F_h(\mathbf{I}) = \mathbf{I'}_{(x',y)} = \mathbf{I}_{(x_{max}-x,y)} \tag{6.1}$$

$$F_v(\mathbf{I}) = \mathbf{I'}_{(x,y')} = \mathbf{I}_{(x,y_{max}-y)} \tag{6.2}$$

$\mathbf{I}$ is a single band of an HSI cube, and $\mathbf{I'}$ is the altered band after Data Augmentation. The changes are applied to every band in $\mathbf{X}$ and its corresponding $\mathbf{Y}$. Horizontal flips and vertical flips can be particularly helpful for SISR due to the fact that they maintain the same spatial information as the original image but in a mirrored form. In some situations, this might help the model learn to manage reflections and symmetries in the data. Flipping can also assist the model in learning to handle various rotations and orientations of the objects in the image.

By combining horizontal and vertical flips, SISR models can be trained on a larger and more diverse dataset, which can help them generalize better to new data and improve their overall performance.



(a) Original (b) Horizontal flip (c) Vertical flip

Figure 6.1: A sample that shows (b) horizontal and (c) vertical flipping applied to the 89^{th} band in a patch from Pavia University dataset.

6.2.2 Shearing

Shear involves moving pixels in every HSI band in a certain direction while keeping the image's boundaries constant [351]. In shear augmentation, an angle is selected,

and then the HSI is sheared along the x-axis or the y-axis, as seen in Equations 6.3 and 6.4. This results in the HSI being stretched or compressed in one direction, while maintaining its aspect ratio. The amount of shear applied to the HSI is controlled by a parameter κ that determines the degree of shear. Since the image's boundaries are kept constant, shearing results in empty areas, which can be filled by applying Zero Padding, as seen in Figures 6.2d and 6.2e, or by reusing values from the images itself by, for instance, reflecting the values of the nearest set of pixels, as seen in Figures 6.2b and 6.2c.

$$S_h(\mathbf{I}) = \mathbf{I'}_{(x',y)} = \mathbf{I}_{(x*\kappa,y)} \tag{6.3}$$

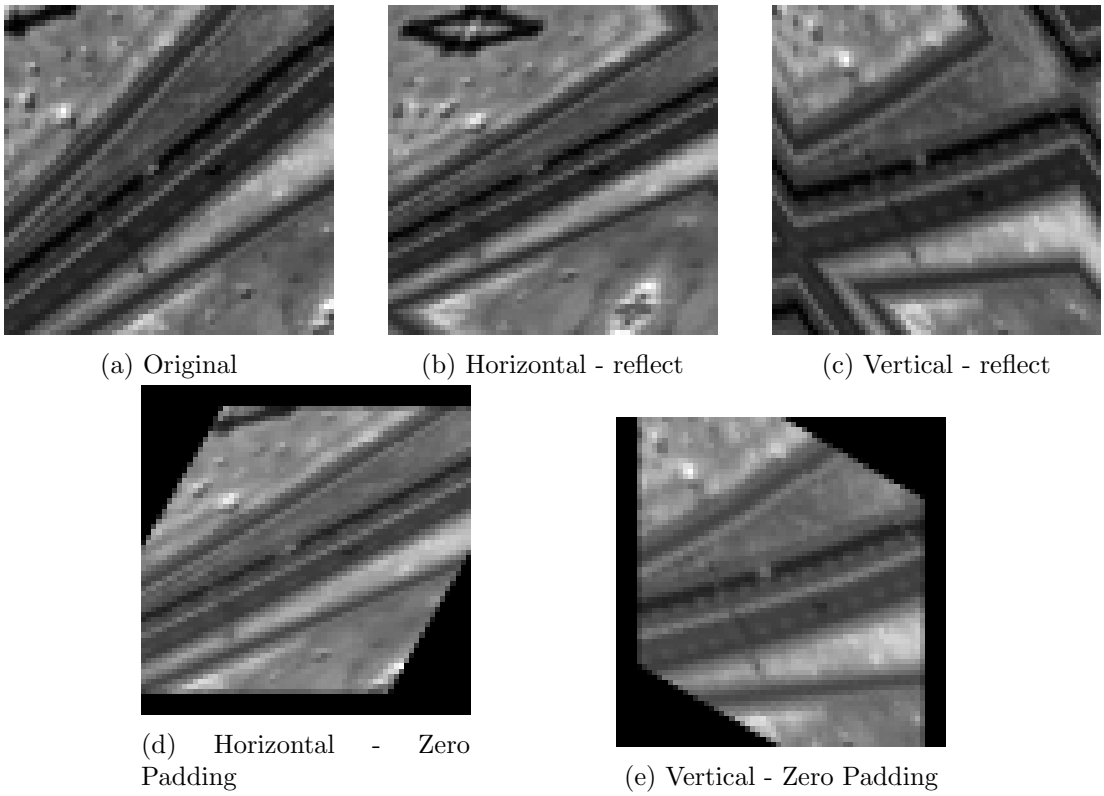
$$S_v(\mathbf{I}) = \mathbf{I'}_{(x,y')} = \mathbf{I}_{(x,y*\kappa)} \tag{6.4}$$

By applying random shear to the training images, the model learns to recognize objects that may appear distorted or slanted in real-world scenarios. This can lead to better accuracy and generalization performance of the model on unseen data. The effect of Zero Padding Vs. applying reflection will be investigated in this chapter.

6.2.3 Rotation

Rotation involves transforming an HSI by a specified angle around its center [350]. This can be done using various algorithms, such as the affine transformation, which involves a linear mapping of the image coordinates.

Rotation can be used to generate new training examples by rotating the original HSI by various angles. This can help to increase the robustness of ML models to variations in orientation or viewpoint. For example, in object recognition tasks, rotating images can help to teach the model to recognize objects from different angles. Typically, rotation is performed with a random angle θ, as seen in Equation 6.5. The angle is chosen from a uniform distribution within a specified range, such as $-45°$ to $45°$, which is the range used for this study. Rotating the image can cause empty areas to appear around its borders, similar to shearing, which can be filled either with Zero Padding or by reflecting the values of the neighboring pixels. Both cases are illustrated in Figure



(a) Original (b) Horizontal - reflect (c) Vertical - reflect

(d) Horizontal - Zero Padding (e) Vertical - Zero Padding

Figure 6.2: A sample that shows shearing applied to the 89^{th} band in a patch from Pavia University dataset. In (b) horizontal and (c) vertical shearing, the image is reflected around the borders, while in (d) horizontal and (e) vertical the image is padded with zeros.

6.3.

$$
\begin{aligned}
Rot(\mathbf{I}) &= \mathbf{I'}_{(x',y')} \\
x' &= x * cos(\theta) - y * sin(\theta) \\
y' &= x * sin(\theta) + y * cos(\theta)
\end{aligned}
\tag{6.5}
$$

6.2.4 Cutting

The basic principle of Data Augmentation Cutting techniques is cropping and mixing regions of the same HSI or two different HSIs. This section discusses three methods: Jigsaw, CutMix, and CutBlur. The newly introduced method, CutMixBlur, also belongs to this Data Augmentation category, which will be described at the end of this

(a) Original (b) Rotation - reflect (c) Rotation - Zero Padding

Figure 6.3: A sample that shows (b) rotation with reflect and (c) rotation with Zero Padding applied to the 89^{th} band in a patch from Pavia University dataset.

subsection.

Jigsaw

Jigsaw image augmentation is a technique commonly used in image classification and object recognition tasks [352]. It involves shuffling and reassembling the pieces of an image in a puzzle-like manner to create new training examples.

The algorithm consists of the following steps:

1. Splitting: The input HSI is divided into a grid of smaller, equally-sized cubes. The number of patches can vary depending on the desired complexity and granularity of the augmented images.

2. Random Shuffling: The cubes are randomly rearranged within the grid. This step introduces spatial disarray to the original HSI.

3. Reassembling: The shuffled cubes are then merged back together to form a new augmented HSI. The original layout and content of the HSI are no longer preserved, creating a novel representation.

By shuffling the cubes and reconstructing the HSI, the HSI-SISR training model is exposed to different spatial relationships and local context. This augmentation technique helps the model become more robust to changes in object arrangement and en-

courages it to learn more invariant features. An example of jigsaw Data Augmentation is seen in Figure 6.4c.

CutMix

CutMix is an image augmentation technique that involves cutting and pasting a portion of one image $\mathbf{Y}$ onto another image $\dot{\mathbf{Y}}$ by selecting two random HSI and randomly cropping a cuboid portion from one HSI. The size and location of the crop are determined by a probability distribution, which can be uniform or non-uniform. The cropped portion is then pasted onto the other HSI at a random location, replacing a cuboid portion of the second HSI. This process creates a new image that contains features from both of the original HSI, effectively mixing them together without distorting the spectral signature of the objects that exist in both HSI. By doing so, CutMix encourages the model to learn more robust features that can handle variations in the input data. CutMix has been shown to be effective in improving the accuracy and robustness of DL models on several image classification tasks. It has also been extended to other tasks, such as object detection and segmentation. A visual example of is seen in Figure 6.4d. Mathematically, CutMix can be expressed as follows:

$$\mathbf{Y}' = \eta * \mathbf{Y} + (1 - \eta) * \dot{\mathbf{Y}}, \tag{6.6}$$

where η is a binary mask that contains the value 1 within the desired region to be cropped and 0 otherwise.

CutBlur

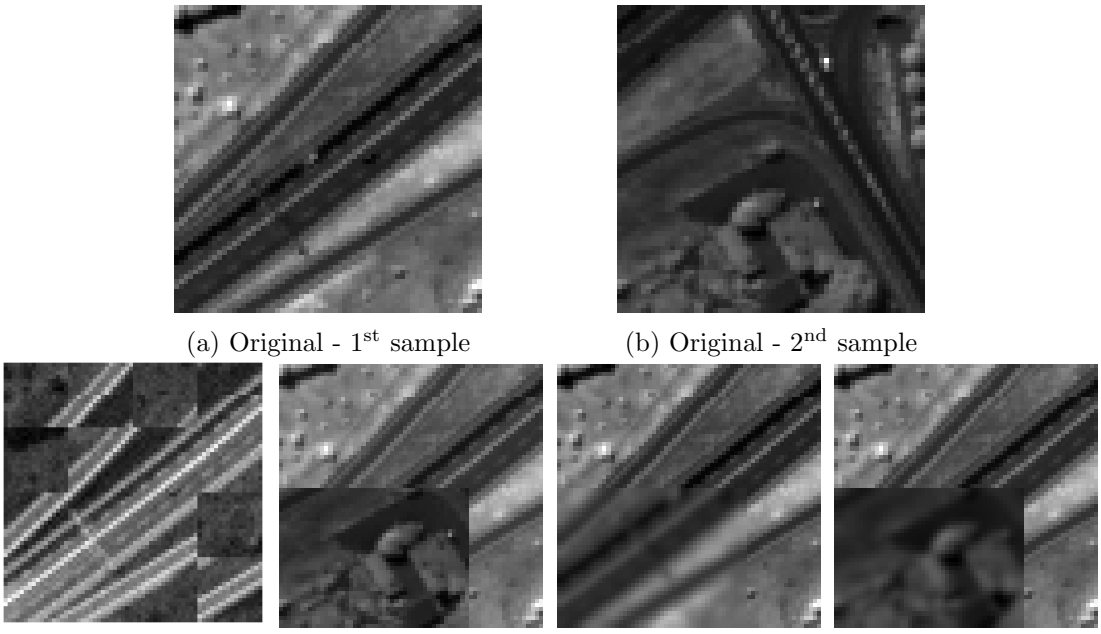
CutBlur is another cutting technique that works by randomly selecting a rectangular region from the LR-HSI based on a specified ratio and maximum size. The selected region is then pasted back into the HR-HSI, replacing the original cutout region. This process is repeated for all input HSI, with each iteration resulting in a new augmented image. CutBlur is represented in Equation 6.7 and visualized in Figure 6.4e.

$$\mathbf{X}' = \eta * \mathbf{Y} + (1 - \eta) * \mathbf{X} \tag{6.7}$$

CutMixBlur

CutMixBlur is a Data Augmentation technique that combines the principles of CutMix and CutBlur. For two samples of HSI, a cuboid of random location and size is cropped from one sample and pasted onto the other. The degradation is applied to the cropped region only, leaving the remainder of the new sample intact. By blending CutMix and CutBlur, CutMixBlur is anticipated to have a similar, if not superior, impact, considering the observed enhancements of CutMix and CutBlur in SISR for MSI [348]. This technique aims to augment the dataset further, enhance the network's capability to localize improvements, and reduce artifacts along the edges. The specific formulation of CutMixBlur is represented by equation 6.8, and it is visualized in Figure 6.4f.

$$\mathbf{X}' = \eta * \mathbf{Y} + (1 - \eta) * \dot{\mathbf{X}} \tag{6.8}$$



(a) Original - 1st sample (b) Original - 2nd sample

(c) Jigsaw (d) CutMix (e) CutBlur (f) CutMixBlur

Figure 6.4: A sample that shows (c) Jigsaw, (d) CutMix, (e) CutBlur, and and (f) CutMixBlur applied by utilizing the 89th band of two samples (a) and (b) from Pavia University dataset.

6.3 Experimental Setup

The 3D-SRCNN333, which was previously developed in Chapter 4 is used to investigate the impact of each Data Augmentation technique. Initially, they are applied individually. The goal is to identify the techniques that improve PSNR and SSIM without causing any deterioration in SAM. These effective techniques will then be combined in a random manner. The randomness is introduced through two different approaches. The first approach is referred to as Random Stacking, where each Data Augmentation technique is applied to the dataset individually, and the augmented data is appended to the original dataset. As a result, the size of the dataset becomes a multiple of the number of Data Augmentation techniques employed. The second approach is denoted as Random Compound, which involves applying all Data Augmentation techniques to each image in the dataset in a compounded manner, effectively doubling the size of the dataset. All experiments are conducted in the same environment to ensure a fair and unbiased comparison.

The experiments are conducted specifically on the Pavia University and Indian Pines datasets because their small sizes make the effects of Data Augmentation more noticeable. Table 6.1 presents the initial and final sizes of the datasets after applying each Data Augmentation technique. Similar to the setup used in Chapter 4 (Section 4.2, MSE is employed as the loss function and Adam is utilized as the optimization algorithm. The learning rate scheduler initiates with a value of 10^{-3} and decreases by power of 10 if no improvement, or an improvement less than 10^{-5}, is observed for three consecutive epochs. The minimum learning rate value allowed is 10^{-5}. Regarding the number of epochs, an early stopping strategy is employed. Training ceases if no improvement, or an improvement less than 10^{-5}, is detected for five consecutive epochs.

Table 6.1: This table shows the resulting dataset size after each Data Augmentation mode. Random Stack (RS) yields bigger data size compared to Random Compound (RC).

DA Mode	Pavia University			Indian Pines		
	Training	Validation	Testing	Training	Validation	Testing
None	36	4	5	2	1	1
Only 1 method	72	9	9	4	2	2
RS	396	49	50	18	9	9
RC	72	9	9	4	2	2

6.4 Results and Analysis

The upcoming sections delve into the performance evaluation of 3D-SRCNN333 trained with Data Augmentation, applied to both the Pavia University dataset in the reflectance domain and the Indian Pines dataset in the radiance domain. Furthermore, the evaluation extends to testing 3D-SRCNN333, trained with Data Augmentation on the Pavia University dataset, on the Pavia Center dataset to assess the effect of Data Augmentation on previously unseen endmembers encountered during the training process.

6.4.1 Pavia University Experiments

Table 6.2 shows a summary of the quantitative results obtained by each Data Augmentation technique in addition to Random Stacking and Random Compound Data Augmentation techniques. The new size of Pavia University dataset after applying a Data Augmentation technique is seen in Table 6.1. Both vertical and horizontal Flipping boost PSNR and SSIM without deteriorating SAM. As for Shearing, both Horizontal and Vertical Shearing with Zero Padding cause large deterioration in SAM. Thus, Shearing with Zero Padding will be excluded when applying Random Compound

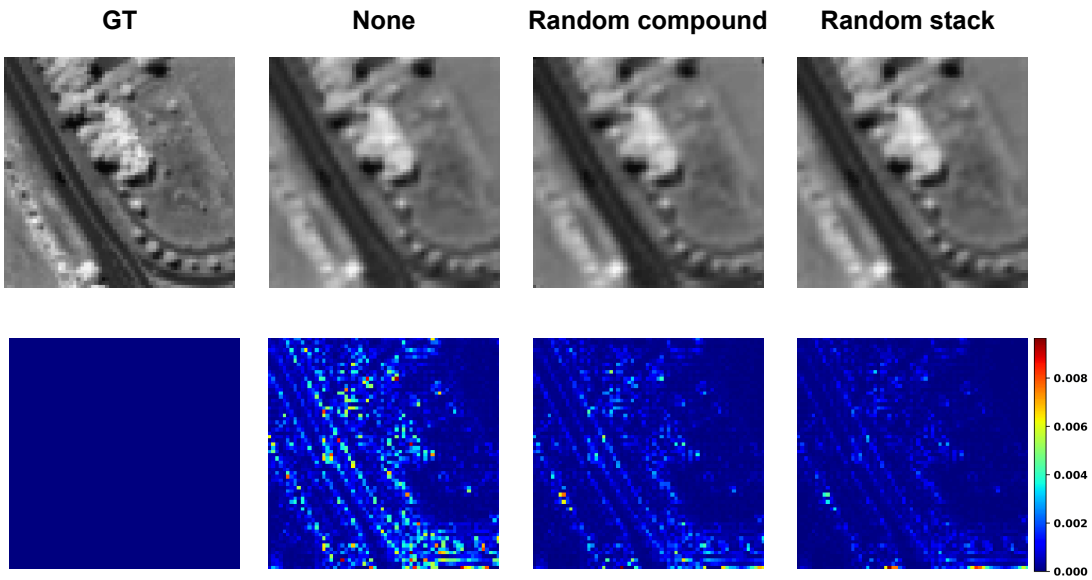
Table 6.2: Results summary of each Data Augmentation (DA) method discussed in Section 6.2 applied on Pavia University dataset. Δ expresses the difference in quality metric between the DA method and the original dataset without DA. The following abbreviations are used: Zero Padding (ZP), Reflect (R). Horizontal (H), Vertical (V), CutMix (CM), CutBlur (CB), CutMixBlur (CMB), Random Compound (RC), and Random Stack (RS).

Metric			PSNR	Δ_{PSNR}	SSIM	Δ_{SSIM}	SAM	Δ_{SAM}
None			31.534 ± 0.196	0	0.9222 ± 0.0009	0	5.56 ± 0.15	0
Flipping		H	32.227 ± 0.150	0.693 ↑	0.9235 ± 0.0005	0.0013 ↑	5.06 ± 0.11	0.50 ↓
		V	32.326 ± 0.145	0.792 ↑	0.9233 ± 0.0004	0.0011 ↑	4.36 ± 0.11	1.20 ↓
Shearing	ZP	H	33.677 ± 0.098	2.143 ↑	0.9412 ± 0.0010	0.019 ↑	18.78 ± 2.32	13.22 ↑
		V	34.175 ± 0.065	2.641 ↑	0.9454 ± 0.0002	0.0232 ↑	14.15 ± 2.77	8.59 ↑
	R	H	32.567 ± 0.102	1.033 ↑	0.9248 ± 0.0001	0.0226 ↑	4.79 ± 0.09	0.77 ↓
		V	32.890 ± 0.143	1.356 ↑	0.9370 ± 0.0002	0.0148 ↑	4.51 ± 0.10	1.05 ↓
Rotation		ZP	34.620 ± 0.132	3.086 ↑	0.9451 ± 0.0002	0.0229 ↑	21.99 ± 3.44	16.43 ↑
		R	33.839 ± 0.085	2.305 ↑	0.9391 ± 0.0001	0.0169 ↑	4.58 ± 0.08	0.98 ↓
Cutting		Jigsaw	32.688 ± 0.096	1.154 ↑	0.9254 ± 0.0003	0.0032 ↑	4.66 ± 0.12	0.90 ↓
		CM	32.793 ± 0.095	1.259 ↑	0.9268 ± 0.0002	0.0046 ↑	4.36 ± 0.05	1.20 ↓
		CB	32.878 ± 0.088	1.344 ↑	0.9276 ± 0.0001	0.0054 ↑	4.50 ± 0.06	1.06 ↓
		CMB	32.620 ± 0.074	1.086 ↑	0.9255 ± 0.0001	0.0033 ↑	4.65 ± 0.05	0.91 ↓
Random		RC	33.005 ± 0.056	1.471 ↑	0.9305 ± 0.0000	0.0083 ↑	3.97 ± 0.07	1.59 ↓
		RS	34.225 ± 0.060	2.691 ↑	0.9443 ± 0.000	0.0221 ↑	3.24 ± 0.04	2.32 ↓

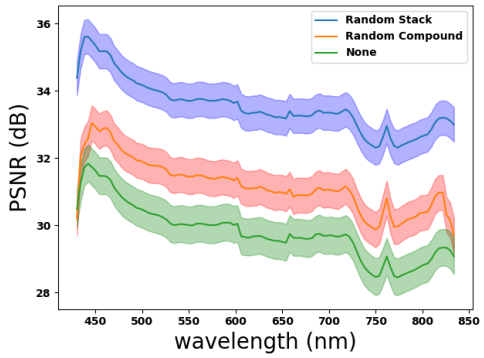
and Random Stacking. Conversely, Shearing with reflection boosts all quality metrics. A similar observation can be made with Rotation results. Rotation with Zero Padding shows the biggest increase in PSNR of +3.086 dB. It also shows a high improvement in SSIM of +0.0229. However, it shows significant deterioration in SAM with an increase of +16.43° compared to the original dataset without Data Augmentation. Thus, Rotation with Zero Padding will be excluded when applying Random Stacking and Random Compound. On the other hand, Rotation with reflection effect boosts PSNR, SSIM, as

well as SAM, as the latter decreases by $-1.09°$ compared to the original. All Cutting methods demonstrate improvements across all quality metrics. CutMixBlur shows the best improvement in SAM, as it decreases by $-1.17°$, while CutBlur shows the best improvement in PSNR and SSIM. The methods that will be applied in a random manner are: horizontal and vertical flipping, Shearing with reflect, Rotation with reflect, Jigsaw, CutMix, CutBlur, and CutMixBlur.

The Random Compound technique, which involves applying all Data Augmentation techniques that showed improvements in all metrics, results in a significant performance boost across all quantitative metrics compared to each individual Data Augmentation technique. This performance improvement is even more noticeable with the Random Stacking technique, which outperforms all other methods, including Random Compound. The difference in Pavia University dataset size after applying Random Compound and Random Stacking can be observed in Table 6.1. The visual evidence of the improvement in PSNR, SSIM, and SAM is apparent in Figure 6.5. The predicted results, along with their RMSE maps in Figure 6.5a, demonstrate that Random Stacking produces fewer errors compared to the original dataset without Data Augmentation and Random Compound Data Augmentation. A similar conclusion can be drawn from Figure 6.5b, where Random Stacking consistently achieves the highest PSNR across the spectral range, followed by Random Compound. Thus, all spatial quantitative and qualitative metrics indicate a substantial improvement in SISR performance for the Pavia University dataset when applying the appropriate Data Augmentation techniques in a Random Stacking manner.



(a) Qualitative results



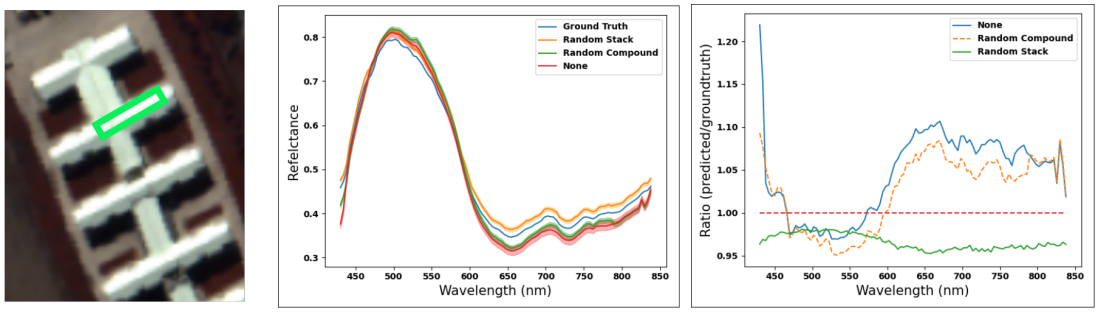
(b) PSNR plot

Figure 6.5: (a) shows the qualitative results of training 3D-SRCNN333 before and after applying random DA techniques on Pavia University dataset. The top row shows the result produced by each method, while the bottom row shows a visualization of the error between the produced results and the GT in terms of RMSE. Random Stack produces the best result visually, and this is evident in its RMSE map. (b) shows PSNR plot for Random Stack, Random Compound, and no Data Augmentation (None), where the solid lines represent the mean and the shaded areas represent the standard deviation.

In line with the previous chapters, the spectral signature analysis of the Pavia University dataset involves three distinct classes: Metal Sheet, Gravel, and Asphalt. The same homogeneous region employed in previous chapters is selected once again. Within this region, the mean spectral signature of the pixels is recorded for both GT and the Data Augmentation methods being evaluated. Moreover, given that each method

was trained and tested five times, the aggregate results encompassing the mean and standard deviation values derived from these repetitions are recorded.

In Figure 6.6a, the Metal Sheet region is shown, with the detailed results presented in Figures 6.6b-c. By examining the spectral signature within Figure 6.6b, it becomes evident that there is an overlapping trend among the results produced by all methods. Delving further into Figure 6.6c, a closer inspection reveals that Random Stack exhibits a notably closer alignment with the GT when compared to Random Compound and the absence of Data Augmentation.



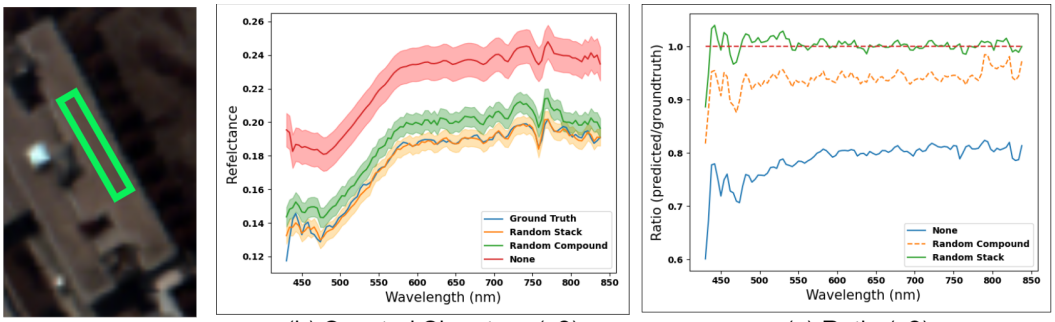
(a) Homogenous Metal Sheet Region

(b) Spectral Signature (x2)

(c) Ratio (x2)

Figure 6.6: The results of testing 3D-SRCNN333 without Data Augmentation (None), Random Compound, and Random Stack Data Augmentation on (a) homogeneous Metal Sheet region, where the spectral signature is plotted in (b), and (c) shows the corresponding ratio plots. The overall order of performance is: Random Stack, Random Compound, None.

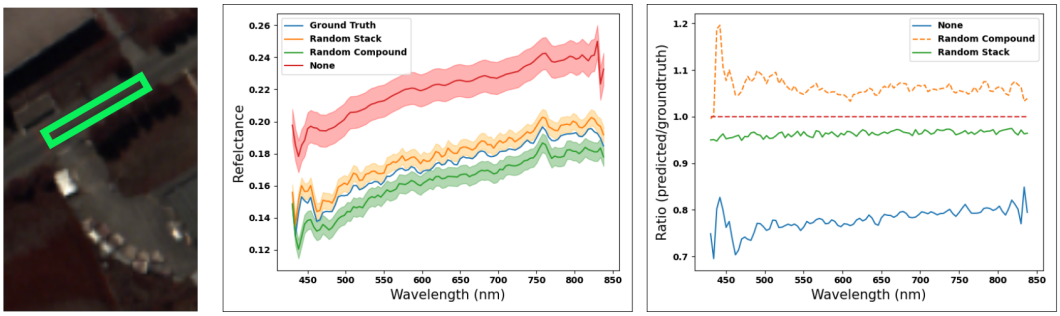
The outcomes for the Gravel region, seen in Figure 6.8a, are detailed in Figures 6.8b-c. Within this particular class, it is noticeable that there is a reduced degree of overlap compared to the Metal Sheet class. As evidenced in Figure 6.8b, the resemblance of Random Stack to the GT is more pronounced than that of Random Compound. This observation is reaffirmed in Figure 6.8c.



(a) Homogenous Gravel Region (b) Spectral Signature (x2) (c) Ratio (x2)

Figure 6.7: The results of testing 3D-SRCNN333 without Data Augmentation (None), Random Compound, and Random Stack Data Augmentation on (a) homogeneous Gravel region, where the spectral signature is plotted in (b), and (c) shows the corresponding ratio plots. The overall order of performance is: Random Stack, Random Compound, None.

The last class to be examined is Asphalt, illustrated in Figure 6.8a. The observations drawn from Figures 6.8b-c closely mirror those observed in the Gravel class. In this class, there is a reduced degree of overlap between the spectral signatures produced by the methods, as depicted in Figure 6.8b. The ratio plots corresponding to these results seen in Figure 6.8c consistently support the same narrative. Notably, Random Stack exhibits a closer proximity to the value of 1 when compared to Random Compound and the no Data Augmentation approach.



(a) Homogenous Asphalt Region (b) Spectral Signature (x2) (c) Ratio (x2)

Figure 6.8: The results of testing 3D-SRCNN333 without Data Augmentation (None), Random Compound, and Random Stack Data Augmentation on (a) homogeneous Asphalt region, where the spectral signature is plotted in (b), and (c) shows the corresponding ratio plots. The overall order of performance is: Random Stack, Random Compound, None.

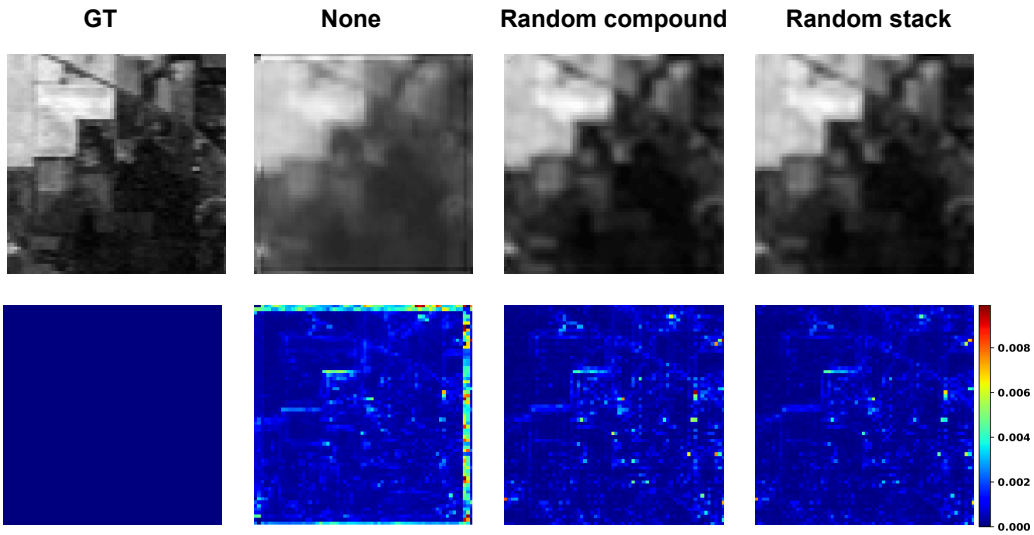
6.4.2 Indian Pines Experiments

The Indian Pines dataset, being the smallest dataset in this study, provides an opportunity to observe the effects of the experiments more prominently compared to the Pavia University dataset. With only 2 trainable images, 1 for validation, and 1 for testing, the impact of the experiments on this dataset is expected to be more evident. Table 6.3 demonstrates that both flipping methods enhance the Indian Pines dataset without causing any deterioration in its spectral signature. Similarly to the Pavia University case, Zero Padding shearing and Zero Padding rotation techniques introduce distortions in SAM despite enhancing PSNR and SSIM, which leads to their exclusion from the Random Stacking and Random Compound techniques. In terms of cutting techniques, Jigsaw exhibits the most significant improvements in PSNR, SSIM, and SAM, followed by CutBlur, CutMixBlur, and then CutMix. For the random Random Stacking and Random Compound techniques, the following methods are employed: horizontal and vertical flipping, shearing with reflection, rotation with reflection, Jigsaw, CutMix, CutBlur, and CutMixBlur.

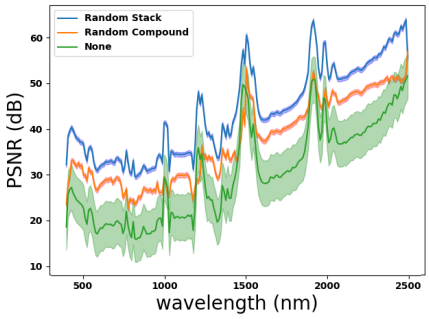
In the Random Compound technique, all the aforementioned Data Augmentation methods are applied together in a compounded manner, effectively doubling the size of the dataset. Random Compound demonstrates significant improvements in all qualitative metrics. Although the dataset size remains the same as when applying each Data Augmentation technique individually, the effect is much more pronounced when using a combination of techniques simultaneously. On the other hand, the Random Stacking technique exhibits more apparent and superior results compared to Random Compound. It applies all the techniques simultaneously while also increasing the size of the dataset. The size of the dataset after applying Random Stacking can be observed in Table 6.1.

Table 6.3: Results summary of each DA method discussed in Section 6.2 applied on Indian Pines dataset. Δ expresses the difference in quality metric between the DA method and the original dataset without DA. The following abbreviations are used: Zero Padding (ZP), Reflect (R). Horizontal (H), Vertical (V), CutMix (CM), CutBlur (CB), CutMixBlur (CMB), Random Compound (RC), and Random Stack (RS).

Metric			PSNR	Δ_{PSNR}	SSIM	Δ_{SSIM}	SAM	Δ_{SAM}
None			29.635 ± 2.643	0	0.8714 ± 0.1328	0	4.78 ± 1.84	0
Flipping		H	33.862 ± 1.24	4.227 ↑	0.9249 ± 0.0022	0.0535 ↑	3.43 ± 0.99	1.35 ↑
		V	33.773 ± 1.05	4.138 ↑	0.9196 ± 0.0020	0.0482 ↑	3.34 ± 0.94	1.44 ↑
Shearing	ZP	H	28.646 ± 1.33	0.989 ↓	0.9102 ± 0.0035	0.0388 ↑	8.56 ± 2.83	3.78 ↓
		V	27.369 ± 1.15	2.266 ↓	0.9266 ± 0.0029	0.0552 ↑	10.64 ± 2.97	5.86 ↓
	R	H	34.124 ± 0.86	4.489 ↑	0.9232 ± 0.0025	0.0518 ↑	3.34 ± 0.81	1.44 ↑
		V	34.639 ± 0.77	5.004 ↑	0.9315 ± 0.0018	0.0601 ↑	3.03 ± 0.70	1.75 ↑
Rotation		ZP	33.563 ± 0.94	3.928 ↑	0.9382 ± 0.0017	0.0668 ↑	14.81 ± 3.29	10.03 ↓
		R	34.462 ± 0.89	4.827 ↑	0.9279 ± 0.0020	0.0565 ↑	3.24 ± 0.72	1.54 ↑
Cutting		Jigsaw	34.454 ± 0.76	4.819 ↑	0.9213 ± 0.0019	0.0499 ↑	3.14 ± 0.71	1.64 ↑
		CM	33.077 ± 0.81	3.442 ↑	0.9074 ± 0.0025	0.0360 ↑	3.71 ± 1.02	1.07 ↑
		CB	33.888 ± 0.79	4.253 ↑	0.9166 ± 0.0029	0.0452 ↑	3.34 ± 0.89	1.44 ↑
		CMB	33.377 ± 1.02	3.742 ↑	0.9111 ± 0.0016	0.0397 ↑	3.62 ± 0.86	1.16 ↑
Random		RC	36.892 ± 0.62	7.257 ↑	0.9443 ± 0.0009	0.0729 ↑	2.56 ± 0.55	2.22 ↑
		RS	37.929 ± 0.55	8.294 ↑	0.9509 ± 0.0009	0.0795 ↑	2.29 ± 0.21	2.49 ↑



(a) Qualitative results



(b) PSNR plot

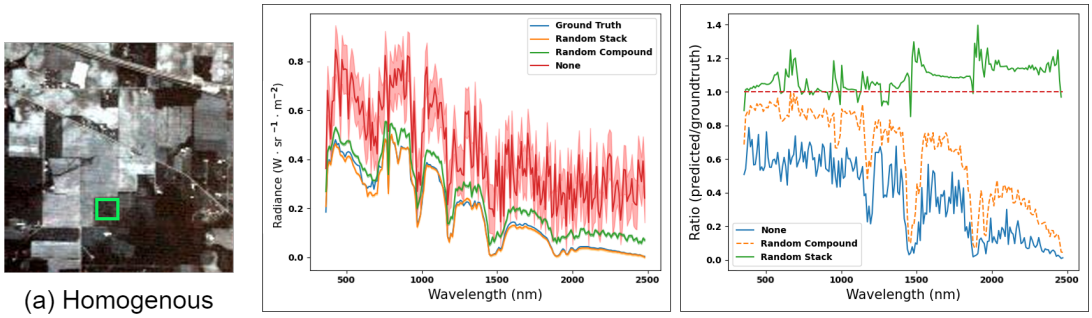
Figure 6.9: (a) shows the qualitative results of training 3D-SRCNN333 before and after applying random DA techniques on Indian Pines dataset. The top row shows the result produced by each method, while the bottom row shows a visualization of the error between the produced results and the GT in terms of RMSE. Random Stack produces the best result visually, and this is evident in its RMSE map. (b) shows PSNR plot for Random Stack, Random Compound, and no Data Augmentation (None), where the solid lines represent the mean, while the shaded areas represent the standard deviation.

The differences in visual quality are evident in Figure 6.9a. The top row showcases the results obtained by training 3D-SRCNN333 without Data Augmentation, with Random Compound Data Augmentation, and then Random Stacking Data Augmentation. The bottom row depicts the RMSE map corresponding to the predicted results obtained from each Data Augmentation method compared to the GT. Random Compound exhibits the least spatial distortions and the closest visual similarity to

the GT, as indicated by both the predicted output and its corresponding RMSE map. Additionally, Figure 6.9b plots the PSNR across the spectral range, demonstrating that Random Stacking achieves the highest PSNR, followed by Random Compound, and lastly the case without Data Augmentation. These results are consistent with the findings from the Pavia University dataset.

To evaluate the spectral fidelity of the Data Augmentation approaches on Indian Pines dataset, three classes are inspected; Grass-trees, Corn-notill, and Wood. As usual, three homogeneous were chosen, each one contains one of the aforementioned class. The mean GT spectral signature is calculated from the pixel within that region. The same thing is done for each Data Augmentation method. Since each method was tried five times, this evaluation within the region is done five times as well, and the mean and standard deviation are calculated.

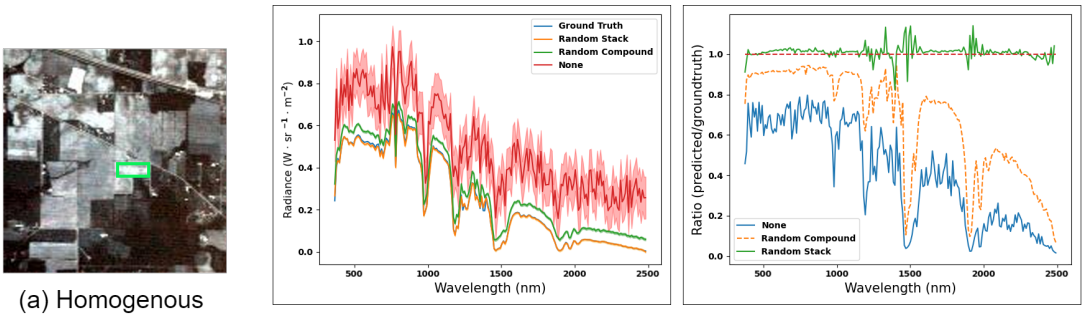
In the case of the Grass-trees class, the examination focuses on the region depicted in Figure 6.10a. The spectral signatures generated by each method can be observed in Figures 6.10b. It is evident that the results obtained without the utilization of Data Augmentation exhibit pronounced distortions, especially since Indian Pines dataset size is very limited. Conversely, both Random Stack and Random Compound yield significantly less distorted results. While the closer resemblance of Random Stack to the GT compared to Random Compound and without Data Augmentation is apparent, a closer examination of the corresponding ratio plots is warranted. As depicted in Figure 6.10c, Random Stack manages to capture the spectral signature more effectively than Random Compound and the no Data Augmentation approach.



(a) Homogenous Grass-trees Region (b) Spectral Signature (x2) (c) Ratio (x2)

Figure 6.10: The results of testing 3D-SRCNN333 without Data Augmentation (None), Random Compound, and Random Stack Data Augmentation on (a) homogeneous Grass-trees region, where the spectral signature is plotted in (b), and (c) shows the corresponding ratio plots. The overall order of performance is: Random Stack, Random Compound, None.

The next class to be examined is Corn-notill illustrated in Figure 6.11a. The findings closely mirror those of the previous class, with stark distortions evident when Data Augmentation is omitted, as depicted in Figures 6.11b. Both Random Stack and Random Compound continue to yield commendable results. In fact, Random Compound exhibits an even closer resemblance to the GT in this case, as indicated by the ratio plot in Figure 6.11c.

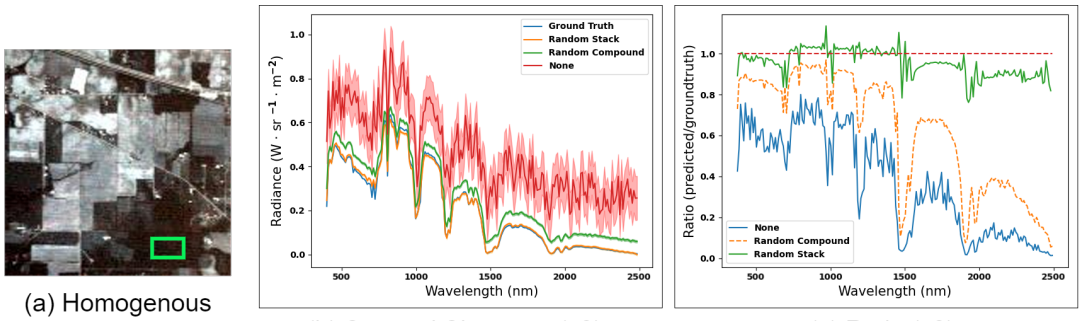


(a) Homogenous Corn-notill Region (b) Spectral Signature (x2) (c) Ratio (x2)

Figure 6.11: The results of testing 3D-SRCNN333 without Data Augmentation (None), Random Compound, and Random Stack Data Augmentation on (a) homogeneous Corn-notill region, where the spectral signature is plotted in (b), and (c) shows the corresponding ratio plots. The overall order of performance is: Random Stack, Random Compound, None.

Regarding the Wood class region shown in Figure 6.12a, the observations align consistently with those made in the previous two classes. Omitting Data Augmentation

continues to lead to substantial distortions in the results. In contrast, both Random Stack and Random Compound prove effective in correcting most of these errors, as evident in the spectral signature plots in Figure 6.12b. As observed in the previous two classes, Random Stack exhibits the closest resemblance to the GT.



(a) Homogenous Wood Region

(b) Spectral Signature (x2)

(c) Ratio (x2)

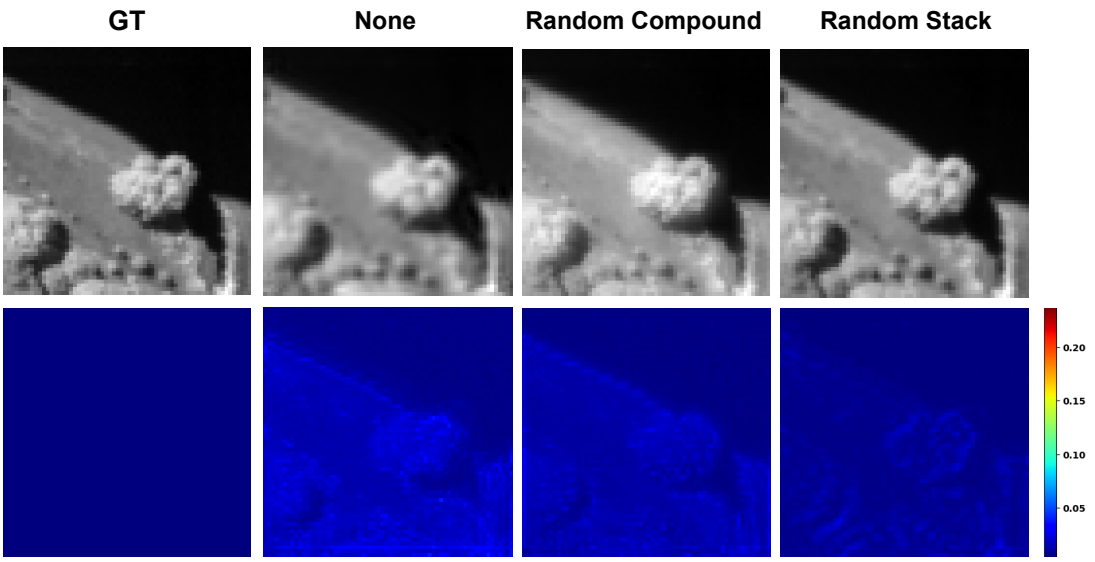
Figure 6.12: The results of testing 3D-SRCNN333 without Data Augmentation (None), Random Compound, and Random Stack Data Augmentation on (a) homogeneous Wood region, where the spectral signature is plotted in (b), and (c) shows the corresponding ratio plots. The overall order of performance is: Random Stack, Random Compound, None.

6.4.3 Pavia Center Experiments

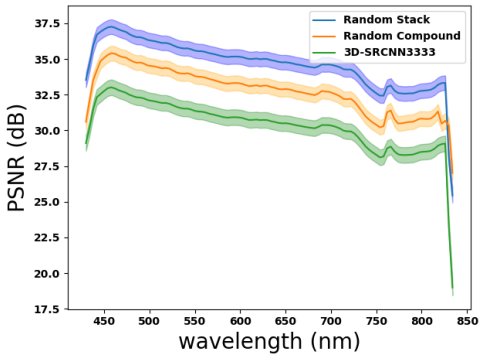
To support the analysis, it is beneficial to examine the performance of 3D-SRCNN333 when trained on the augmented Pavia University dataset and then tested on Pavia Center dataset. With the identification of effective Data Augmentation methods, the evaluation extends to assess how Random Compound and Random Stack behave when applied to the Pavia Center dataset, and whether they yield improvements in this context as they did in Pavia University and Washington DC Mall.

To begin, the qualitative results are observed in Figure 6.13a. While the improvements may be somewhat less pronounced than those in the Pavia University dataset, they remain discernible in both scale factors $\times 2$ and $\times 4$. Random Stack shows visibly less errors compared to Random Compound and the original without Data Augmentation. This trend is substantiated by an examination of the PSNR plots displayed in Figure 6.13b for both scale factors $\times 2$ and $\times 4$. Collectively, these qualitative findings consistently demonstrate that Random Compound surpasses the original model

without Data Augmentation, and Random Stack outperforms them both.



(a) Qualitative results

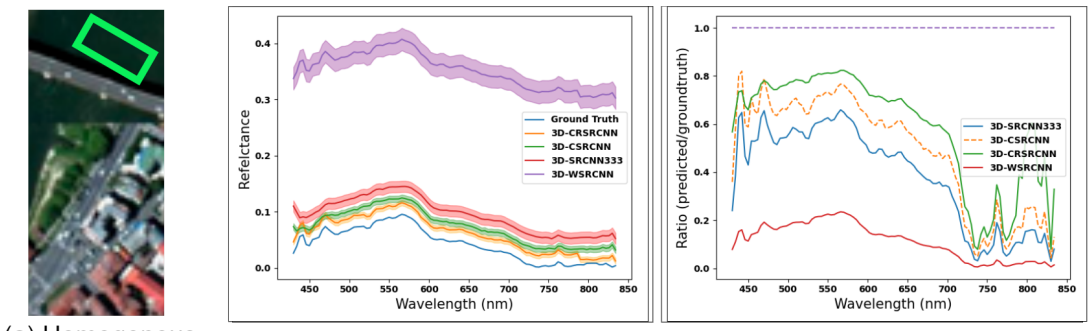


(b) PSNR plot

Figure 6.13: (a) shows the qualitative results of training 3D-SRCNN333 before and after applying random DA techniques on Pavia Center dataset. The top row shows the result produced by each method, while the bottom row shows a visualization of the error between the produced results and the GT in terms of RMSE. Random Stack produces the best result visually, and this is evident in its RMSE map. (b) shows PSNR plot for Random Stack, Random Compound, and no Data Augmentation (None), where the solid lines represent the mean, and the shaded areas represent the standard deviation.

To assess spectral fidelity, the methods are subjected to the examination of some of the classes within Pavia Center scene. In this case, the focus is directed toward the Water and Tiles classes. The results on the uniform region corresponding to the Water class, as depicted in 6.14a, are illustrated in Figures 6.14b-c. Random Stack

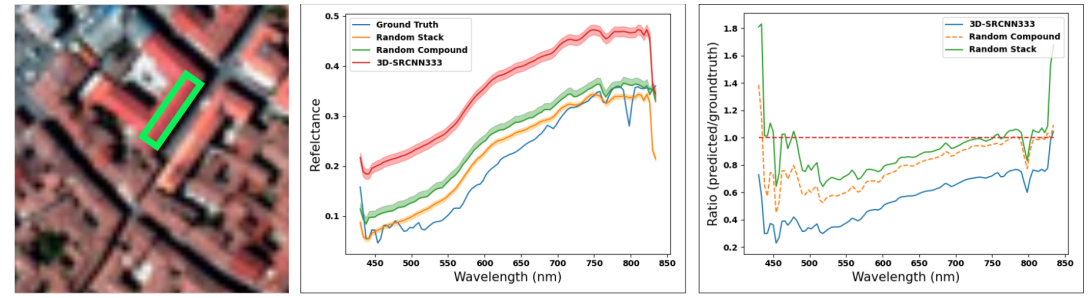
demonstrates exceptional performance in generating a spectral signature that mimics the GT, a trend that is reinforced by the ratio plot displayed in Figure 6.14c. Here, the results produced by Random Stack exhibit the closest proximity to the ideal value of 1.



(a) Homogenous Water Region (b) Spectral Signature (x2) (c) Ratio (x2)

Figure 6.14: The results of testing the original 3D-SRCNN333 without Data Augmentation, Random Compound, and Random Stack Data Augmentation methods on (a) homogeneous Water region, where the spectral signature is plotted in (b) for scale factor $\times 2$, and (c) shows the corresponding ratio plots. The descending order of performance is: Random Stack, Random Compound, and original.

As for Tile class, the results are seen in Figure 6.15. The results produced by Random Stack and Random Compound both show more distortions compared to Water class, but they still outperform the original network without Data Augmentation. This is seen in Figure 6.15b and its corresponding ratio plot in Figure 6.15c.



(a) Homogenous Tile Region (b) Spectral Signature (x2) (c) Ratio (x2)

Figure 6.15: The results of testing the original 3D-SRCNN333 without Data Augmentation, Random Compound, and Random Stack Data Augmentation methods on (a) homogeneous Tile region, where the spectral signature is plotted in (b) for scale factor $\times 2$, and (c) shows the corresponding ratio plots. The descending order of performance is: Random Stack, Random Compound, and original.

To reinforce the analysis, Table 6.4 presents the quantitative results of testing on the Pavia Center dataset, emphasizing PSNR, SSIM, and SAM metrics. These results align with the qualitative observations discussed earlier and the spectral fidelity analysis presented in this section. Overall, the improvements are less than those observed on Pavia University dataset. Nonetheless, they are still notable and show visible enhancements on Pavia Center dataset across all metrics. Random Stack keeps the lead.

Table 6.4: Quantitative results of training 3D-SRCNN333 without Data Augmentation, with Random Compound, and with Random Stack using Pavia University dataset and then testing on Pavia Center dataset. Δ shows the difference between the metric produced by each method and None (without Data Augmentation). The order of performance remains consistent with that of Pavia University in terms of PSNR, SSIM, and SAM.

Loss function	PSNR (dB)	SSIM	SAM (°)
None	30.633 ± 0.151	0.9180 ± 0.0001	5.54 ± 0.10
Random Compound	31.231 ± 0.107 $\Delta = 0.598 \uparrow$	0.9254 ± 0.0002 $\Delta = 0.0021 \uparrow$	4.89 ± 0.12 $\Delta = 0.65 \downarrow$
Random Stack	$\mathbf{31.998 \pm 0.078}$ $\Delta = 1.365 \uparrow$	$\mathbf{0.9300 \pm 0.0001}$ $\Delta = 0.0120 \uparrow$	$\mathbf{4.54 \pm 0.08}$ $\Delta = 1.00 \downarrow$

6.5 Summary

This chapter reviewed various Data Augmentation techniques for the purpose of solving HSI data scarcity problem. The methods are reviewed using Pavia University and Indian Pines datasets, which are very limited in size. To test the efficiency, the datasets before and after Data Augmentation are used to train 3D-SRCNN333, which has previously proven its high performance in Chapter 4. An effective Data Augmentation technique is determined by maximizing PSNR and SSIM, and minimizing SAM compared to the original dataset without Data Augmentation. Experiments on both Pavia University and Indian Pines show that horizontal and vertical flipping, Zero Padding

Shearing, Zero Padding Rotation, Jigsaw, CutMix, CutBlur, and CutMixBlur, which is newly introduced in this work, all show positive enhancements. These methods are then utilized in two randomized approaches, Random Compound and Random Stacking. Both Pavia University and Indian Pines show that Random Stacking is superior to Random Compound, and shows significant enhancements both quantitatively and qualitatively. The findings remain consistent when testing on Pavia Center dataset, which was not involved in the training process. Furthermore, the findings are consistent across datasets in reflectance and radiance domain. This is an important initial step in solving data scarcity problem for HSI-SISR. The findings of this chapter were accepted at The International Geoscience and Remote Sensing Symposium (IGARSS) and are expected to be presented during the month of July 2023 [353].

Chapter 7

Complex- and Wavelet-based Approaches

7.1 Introduction

Researchers continually strive to enhance the performance of existing HSI-SR methods while addressing challenges, such as spectral distortions. One common issue, particularly with CNNs, including 3D-CNNs, is overfitting, which poses a greater difficulty in training when dataset sizes are limited, as exemplified in some cases discussed in Chapter 4. In this thesis, the analysis of HSI has primarily focused on the spatial domain, involving direct manipulation of pixel values arranged in a matrix. However, an alternative representation of images exists in the frequency domain, which characterizes an image based on the rate of change in pixel values. In this representation, high-frequency values indicate sudden changes in pixel values, such as edges. Various mathematical representations for images in the frequency domain have been explored, including Discrete Fourier Transform (DFT) [354], Laplace Transform (LT) [355], and Discrete Wavelet Transform (DWT) [356].

Research findings indicate that the utilization of frequency domain processing provides direct control over high and low frequency components of images. This enables more straightforward manipulation of these components, facilitating the preservation and enhancement of intrinsic image details, such as sharp edges. The integration of

frequency domain analysis with DL techniques has been explored in the context of HSI classification [357–360]. Incorporating frequency components in 3D-CNNs has a promising potential to improve the performance of HSI-SISR tasks.

In this chapter, HSI-SISR is investigated in the complex domain as well as wavelet domain. Their effectiveness is reported and compared in terms of quality metrics as well as resilience to overfitting by incorporating them within 3D-SRCNN333 architecture, which has proven its superiority in Chapter 4.

7.2 Wavelet 3D-SRCNN

DWT has been widely used in image analysis and compression. The advantage it offers over other transformation methods is that it provides insights on both spatial and frequency details. The first level of DWT is performed by passing an image of a single band through low and high filters, and then decomposing it into four sub-bands with approximation coefficients (LL), and detail coefficients, which include horizontal details (LH), vertical details (HL), and diagonal details (HH). Mathematically, DWT can be described with two sets of equations. The first one describes the LL sub-image as follows:

$$W_{ll(x,y)} = \frac{1}{\sqrt{M \times N}} \sum_{i=1}^{M} \sum_{j=1}^{N} I_{(i,j)} \phi_{(i,j)}, \tag{7.1}$$

where $W_{ll}(x, y)$ is the LL sub-image representation of a single HSI band $I(i, j)$ of size $M \times N$. $\phi_{(i,j)}$ is a low-pass filter that halves the bandwidth of $I_{(i,j)}$, which means the size of $W_{ll}(x, y)$ is $floor(\frac{M}{2}) \times floor(\frac{N}{2})$. The second equation describes the details sub-images, LH, HL, and HH, as follows:

$$\begin{gathered} W_{\kappa(x,y)} = \frac{1}{\sqrt{M \times N}} \sum_{i=1}^{M} \sum_{j=1}^{N} I_{(i,j)} \varrho_{\kappa(i,j)} \\ k \in \{lh, hl, hh\} \end{gathered} \tag{7.2}$$

where $W_{\kappa(x,y)}$ with $\kappa \in \{lh, hl, hh\}$ refers to the details sub-images. $\varrho_{\kappa(i,j)}$ is a high-pass filter, and similar to the low-pass filter, the produced sub-image is half the

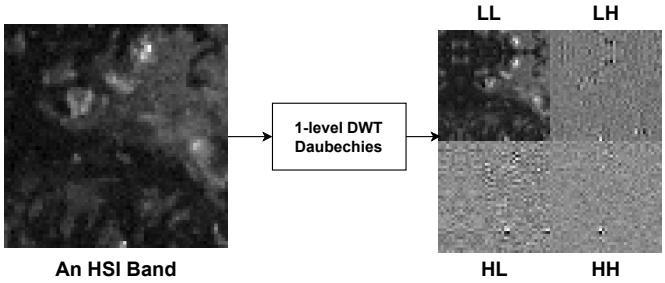


Figure 7.1: Wavelet decomposition of an HSI band.

size of $I_{(i,j)}$. Figure 7.1 demonstrates how each sub-band is produced for a single band from an HSI. Various families of wavelets are available for generating coefficients in the context of DWT. For image de-noising and SR tasks, the most suitable wavelet families to consider are Daubechies, Symlets, and Biorthogonal Spline [361, 362].

The 3D-WSRCNN devised in this work is largely inspired by the research reported in [46,317]. While the architecture is not identical, the principle is similar. That is, the authors in [46, 317] analyzed the wavelet components of the HSI and processed them separately. Each component was processed by a network similar to the 3D-SRCNN915 in terms of depth, filter sizes, and kernel sizes. Thus, the same principle is followed in this study, however, the network follows 3D-SRCNN333, as it has been already proven that 3D-SRCNN333 outperforms 3D-SRCNN915. Additionally, following 3D-SRCNN333 architecture provides a strong basis to compare the performance against 3D-SRCNN333 as well as the 3D-CSRCNN and 3D-CRSRCNN explained in the next section.

After splitting each HSI band into its wavelet sub-components using Daubechies, four HSI cubes are generated: LL, LH, HL, and HH. Each wavelet cube is enhanced using 3D-SRCNN333 in four separate streams. The predicted outputs from each stream are then merged together to apply Inverse DWT (IDWT) and generate the final enhanced HSI. The overall architecture of the 3D-WSRCNN is illustrated in 7.2. Despite 3D-WSRCNN being based on the lightweight architecture of 3D-SRCNN333, the training process is performed four times, which introduces a significant computational burden to take into account.

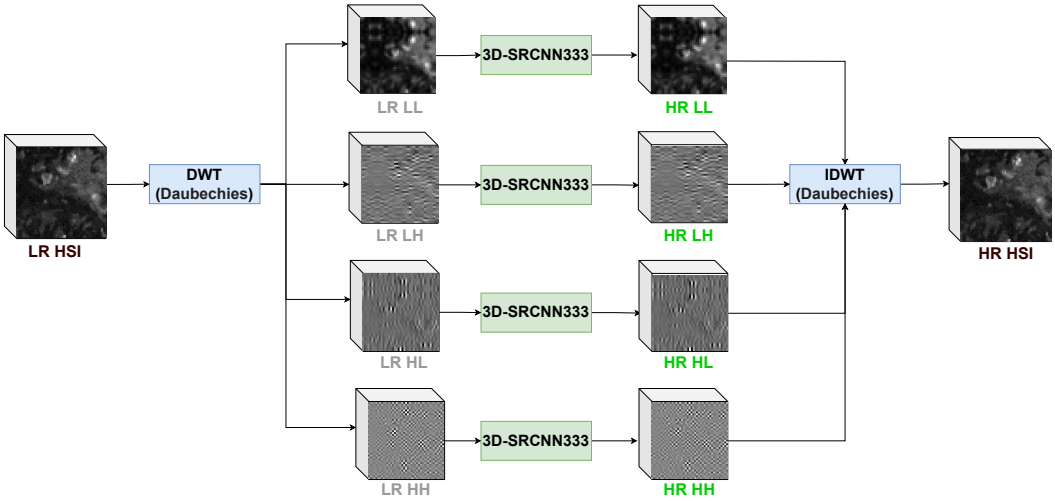


Figure 7.2: Flowchart of the proposed 3D-WSRCNN.

7.3 Complex-valued Convolutional Neural Network

Evidence shows that complex numbers capture image properties better than real numbers [363]. The branch of neural networks that is capable of manipulating complex inputs is referred to as Complex-Valued Neural Networks (CVNNs). Recent research works have demonstrated that CVNNs can boost the performance of HSI classification tasks, and they outperform their real-valued counterparts [357, 358]. Using this type of networks for the purpose of HSI-SISR is a research avenue that has not been explored thus far.

In Chapter 4, the excellent performance of 3D-SRCNN333 was demonstrated on three diverse datasets. Therefore, in this chapter, the architecture of the CVNN is designed to be similar to that of 3D-SRCNN333, providing a solid basis for comparison. The ability to directly manipulate the low and high frequency components of HSI through CVNN is expected to result in superior performance compared to the spatial 3D-SRCNN333. Prior to defining the network architecture, certain components of the network need to be defined in the complex domain, including the Complex Convolution (CC) layer, Complex ReLU (CReLU), and Complex MSE (CMSE). These components will be denoted with $\mathbb{C}$ to avoid confusion with their real-valued counterparts.

- Complex Convolution layer: As stated in Section 3.1 and proven in Chapter 4,

3D operations are more suited for processing HSI than 2D space. For a complex HSI cube denoted $\mathbf{X}_{\mathbb{C}} = [\mathbf{X_r} + i\mathbf{X_{im}}]$, and a Kernel denoted $\mathbf{K}_{\mathbb{C}} = [\mathbf{K_r} + i\mathbf{K_{im}}]$, where $\mathbf{X_r}$, $\mathbf{X_{im}}$, $\mathbf{K_r}$, $\mathbf{K_{im}} \in \mathbb{R}$, a 3D CC layer is defined using Equation 7.3.

$$Conv_{\mathbb{C}(x,y,z)} = \text{ReLU}_{\mathbb{C}} \left(\sum_{i=1}^{M} \sum_{j=1}^{M} \sum_{k=1}^{C} K^{k}_{\mathbb{C}(i,j)} X^{(z+k)}_{\mathbb{C}(x+i,y+j)} + b_{\mathbb{C}} \right), \tag{7.3}$$

in this context, $b_{\mathbb{C}} = [b_r + ib_{im}]$.

- Complex ReLU: CReLU is a complex activation function defined as $max(0, z)$ for some $z \in \mathbb{C}$. To make it analogous to the real-valued ReLU, it assumed that $z = a + ib = \mid z \mid e^{i\theta}$. For z to be considered a positive number, both its real and imaginary parts must be positive, which means the number lies in the first quadrant of the Cartesian plane, and $\theta \in [0, \frac{\pi}{2}]$, as in Equation 7.4.

$$\text{ReLU}_{\mathbb{C}} = \begin{cases} z, & \mathbb{R}(z), \mathbb{I}(z) \geq 0 \\ \\ 0, & \text{otherwise} \end{cases} = \begin{cases} z, & \arg(z) \in [0, \frac{\pi}{2}] \\ \\ 0, & \text{otherwise} \end{cases} \tag{7.4}$$

- Complex MSE: The CMSE is a loss function commonly used in complex-valued neural networks to measure the dissimilarity between the predicted complex-valued HR-HSI denoted $\widehat{\mathbf{Y}}_{\mathbb{C}}$ and the GT complex-valued HR-HSI denoted $\mathbf{Y}_{\mathbb{C}}$. The real MSE is then calculated by computing the square magnitude of CMSE.

It is necessary to translate each band of the HSI cube to the complex domain and convolve it with a complex filter in order to achieve the complex domain equivalent of a traditional real-valued 3D convolution. In this study, Band-wise FFT [364] is used for this purpose.

The proposed 3D-CSRCNN model employs a sequence of three consecutive CC layers, as described in Equation 7.3. Each layer is followed by a CReLU activation. The overall architecture of the 3D-CSRCNN model is depicted in Figure 7.3, indicating the number of filters and kernel sizes. These choices have been made to mimic the effectiveness of the 3D-SRCNN333 model in mitigating artifacts near image borders. Due to the

relatively shallow depth of the 3D-CSRCNN model, pooling and BN layers have been omitted. Additionally, an optional residual connection has been incorporated into the architecture. While the positive impact of residual connections has been demonstrated in real-valued networks [313], their effectiveness in complex-valued networks has not been extensively studied. A comparative analysis will be conducted to evaluate the performance of the network with and without the residual connection. The version of the network with the residual connection will be referred to as 3D-CRSRCNN. This analysis will help determine the contribution of the residual connection in the context of the proposed model.

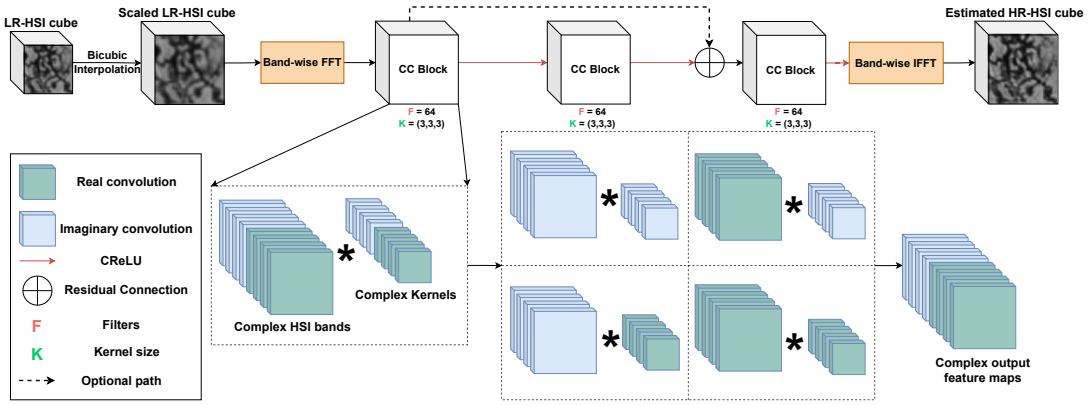


Figure 7.3: Overall architecture of the proposed 3D-CSRCNN with an optional residual connection for 3D-CRSRCNN.

7.4 Experimental Setup

The networks devised in this section, namely 3D-CSRCNN, 3D-CRSRCNN, and 3D-WSRCNN are all trained, tested, and evaluated using Pavia University and Washington DC Mall datasets. Their performances are compared against their real-valued counterpart, which is 3D-SRCNN333. For 3D-CSRCNN and 3D-CRSRCNN, the loss function used is the CMSE, whereas for 3D-SRCNN and 3D-WSRCNN the loss function used is the classical, real-valued MSE.

The learning rate is tuned in a similar way to the previous chapters. Starting from Adam's default value of 10^{-3} the scheduler reduces the learning rate if no improvement

is observed for 3 consecutive epochs, or if the improvement is less than 10^{-5}. The minimum possible value is 10^{-5}. The number of epochs is decided based on early stopping strategy, where the network stops training if no improvement is observed, or if the improvement is less than 10^{-5}, for 5 consecutive epochs. As with the previous chapter, the data division is for both datasets seen in Table 7.1 is 80% for training and the remaining for testing and validation

Table 7.1: Training parameters for all the experiments conducted in Chapter 7. 3D-SRCNN and 3D-WSRCNN use real MSE, while 3D-CSRCNN and 3D-CRSRCNN use complex MSE.

Training parameter	Value	
Optimization function	Adam [327]	
Loss functions	Real MSE, Complex MSE	
Epochs	Early Stopping (max 1000)	
Learning rate	Scheduler: [10^{-3}, 10^{-4}, 10^{-5}]	
Dataset	Pavia University	Washington DC Mall
Training	36	64
Validation	4	8
Testing	5	8

For every network introduced in this chapter, a thorough evaluation approach is employed, involving five separate experiments. In each of these experiments, the dataset is subjected to shuffling, ensuring that the sets of patches used for training, validation, and testing are entirely different in each iteration. The results and plots presented in this study represent the means calculated from all five experiments, accompanied by their corresponding standard deviations.

7.5 Results and Discussion

The following sections present the performance evaluation of four networks, namely, 3D-SRCNN333, 3D-WSRCNN, 3D-CSRCNN, and 3D-CRSRCNN, on both the Pavia University dataset in the reflectance domain and the Washington DC dataset in the radiance domain. It is of interest to investigate whether any significant disparities in performance exist between radiance and reflectance. Additionally, the algorithms initially trained on the Pavia University dataset undergo testing on the Pavia Center dataset to gauge their adaptability when faced with previously unseen endmembers during the training phase.

7.5.1 Pavia University Experiments

The outcomes of training the 3D-SRCNN333, 3D-WSRCNN, 3D-CSRCNN, and 3D-CRSRCNN models on the Pavia University dataset are depicted in Figure 7.4. For each scale factor, the top row shows the predictions of each network, while the bottom row displays the RMSE map. The performance ranking, based on the least to most errors, is as follows: 3D-CRSRCNN, 3D-CSRCNN, 3D-SRCNN333, and 3D-WSRCNN. This order is consistent across both scale factors. Although the visual appearance of 3D-CSRCNN and 3D-CRSRCNN results appears similar, the error map reveals that 3D-CRSRCNN reconstructs more detailed and sharper images compared to 3D-CSRCNN. Conversely, 3D-WSRCNN exhibits blurriness and noticeable degradation, particularly at the $\times 4$ scale factor, where the image borders appear distorted.

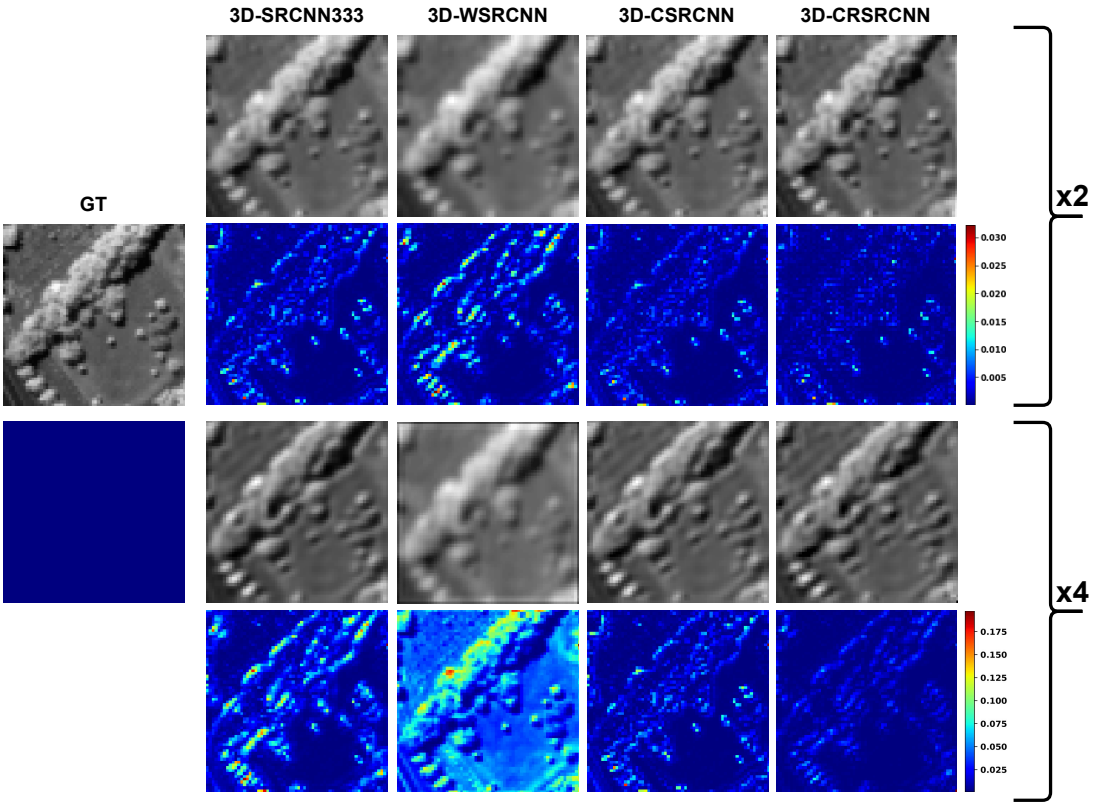
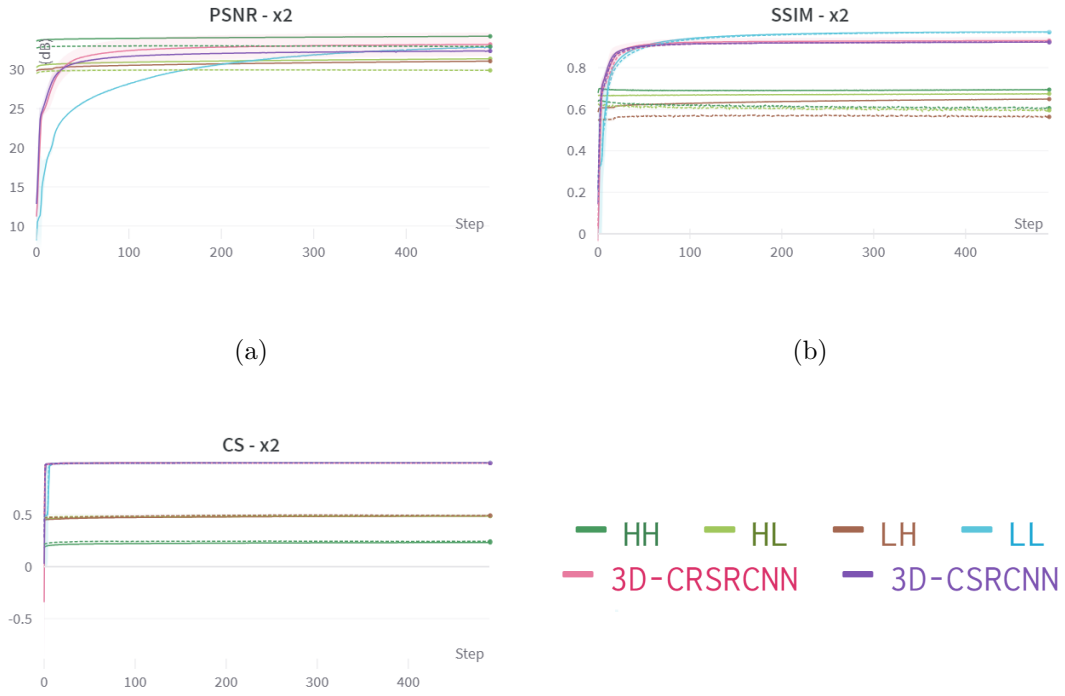


Figure 7.4: Visual results of the predicted HR-HSI for Pavia University dataset by each method. For each scale factor, the top row shows the output of the method, and the bottom row shows the error map in terms of RMSE.

A closer examination of the behavior of each wavelet component during the training process, as shown in Figure 7.5 and 7.6, provides further insights. Components LH, HL, and HH exhibit no signs of improvement throughout epochs 0-500. Although they start with relatively high PSNR, SSIM, and CS values, they remain largely unchanged during training. On the other hand, the LL component shows gradual improvement in all metrics for both scale factors. This suggests that the degradation in the final results originates from the high-frequency components. In contrast to 3D-WSRCNN, the training of 3D-CSRCNN and 3D-CRSRCNN progresses smoothly, with the latter often converging faster than the former.



(a) (b)

(c)

Figure 7.5: This figure shows the progression of training 3D-CSRCNN, 3D-CRSRCNN, and the individual wavelet components of 3D-WSRCNN on Pavia University dataset. The solid line shows training results, while the dashed line shows validation results, and the shaded area represents the standard deviation. Figures 7.5a, 7.5b, and 7.5c show PSNR, SSIM, and CS respectively, for scale factor $\times 2$.

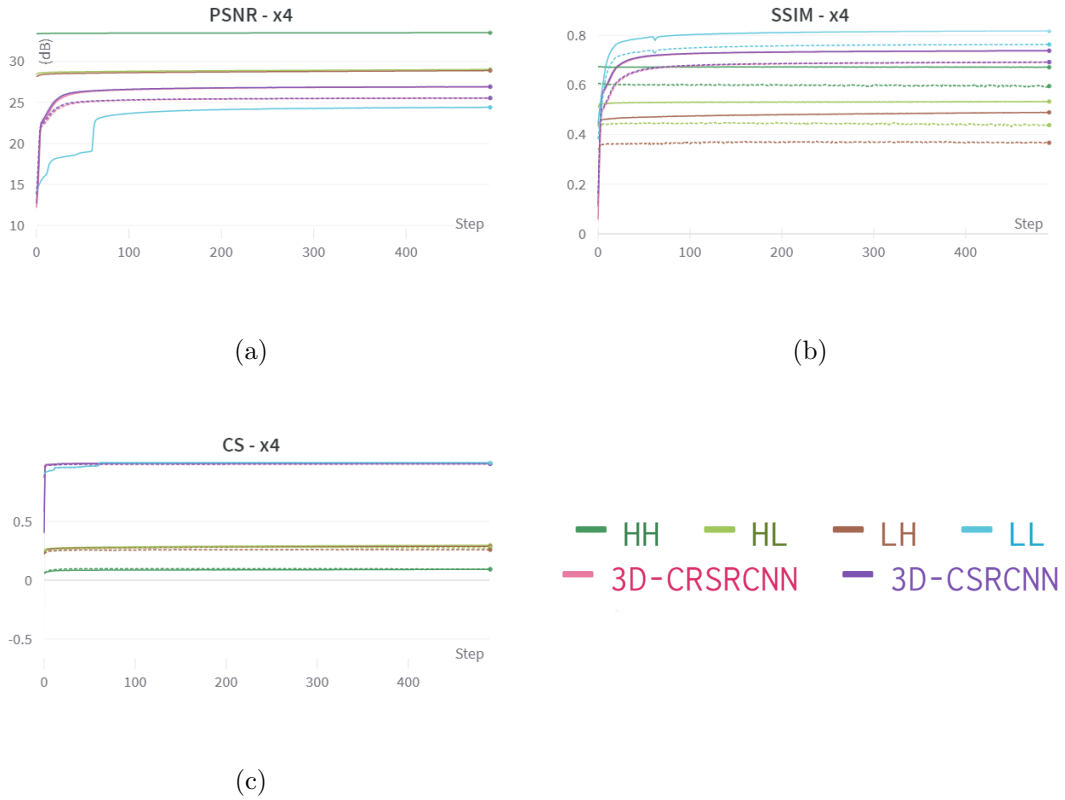
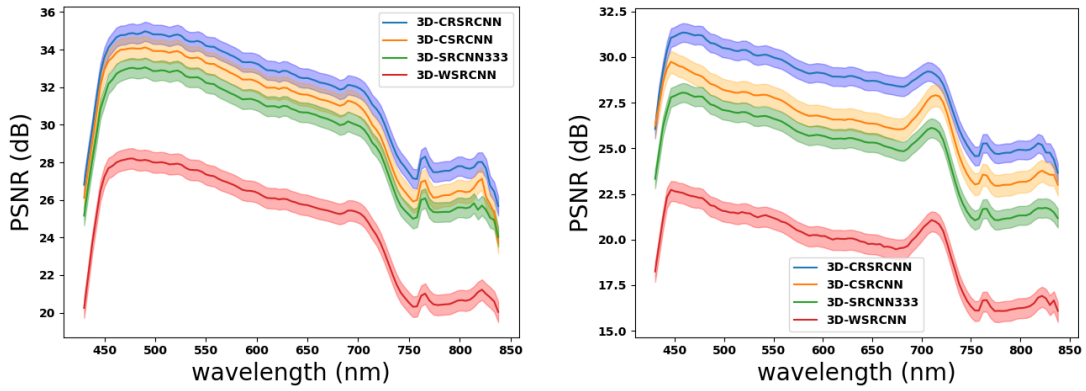


Figure 7.6: This figure shows the progression of training 3D-CSRCNN, 3D-CRSRCNN, and the individual wavelet components of 3D-WSRCNN on Pavia University dataset. The solid line shows training results, while the dashed line shows validation results, and the shaded area represents the standard deviation. Figures 7.6a, 7.6b, and 7.6c show PSNR, SSIM, and CS respectively, for scale factor $\times 4$.

The PSNR values plotted for the spectral range of Pavia University corroborate these findings, as depicted in Figure 7.7a. 3D-CRSRCNN achieves the highest PSNR, followed by 3D-CSRCNN, 3D-SRCNN333, and 3D-WSRCNN, which lags significantly behind, especially from wavelength 700 nm onward. Similar observations can be made from the $\times 4$ scale factor PSNR plot, as shown in Figure 7.7b.



(a) ×2 (b) ×4

Figure 7.7: PSNR results recorded for Pavia University spectral range, where (a) shows the results for scale factor ×2 and (b) shows the results for scale factor ×4. The solid lines represent the mean, while the shaded areas represent the standard deviation.

As in the preceding chapters, the analysis of results encompasses three distinct classes within Pavia University dataset: Metal Sheet, Gravel, and Asphalt. The analysis entails the examination of a homogeneous region, wherein the mean outcome of the pixels within this region is recorded. To ensure robustness, each method undergoes five experimental iterations. The evaluation process is replicated five times, and the outcomes are averaged, with accompanying calculations of uncertainty represented by standard deviation.

The examination commences with an evaluation of the Metal Sheet class, as depicted in Figure 7.8a. Figure 7.8b presents the spectral signatures produced alongside the GT, while Figure 7.8c displays the corresponding ratio plots. These observations reveal that 3D-WSRCNN is unable to accurately mimic the GT spectral signature, in contrast to the other three networks, which exhibit a significant degree of overlap. The ratio plot in Figure 7.8c demonstrates that 3D-SRCNN333 outperforms 3D-CSRCNN, although 3D-CRSRCNN showcases an even more remarkable similarity to the GT. Upon considering the scale factor ×4, Figure 7.8d indicates that 3D-WRCNN introduces even more distortions than at ×2. Additionally, the overlap among the other three networks is reduced compared to ×2, as observed in Figure 7.8d. The corresponding ratio plot in Figure 7.8e reaffirms the higher performance of 3D-SRCNN333

over 3D-CSRCNN, while 3D-CRSRCNN outperforms both networks. In summary, for the Metal Sheet class, the hierarchy of performance is as follows: 3D-CRSRCNN, 3D-SRCNN333, 3D-CSRCNN, and then 3D-WSRCNN, which significantly lags behind.

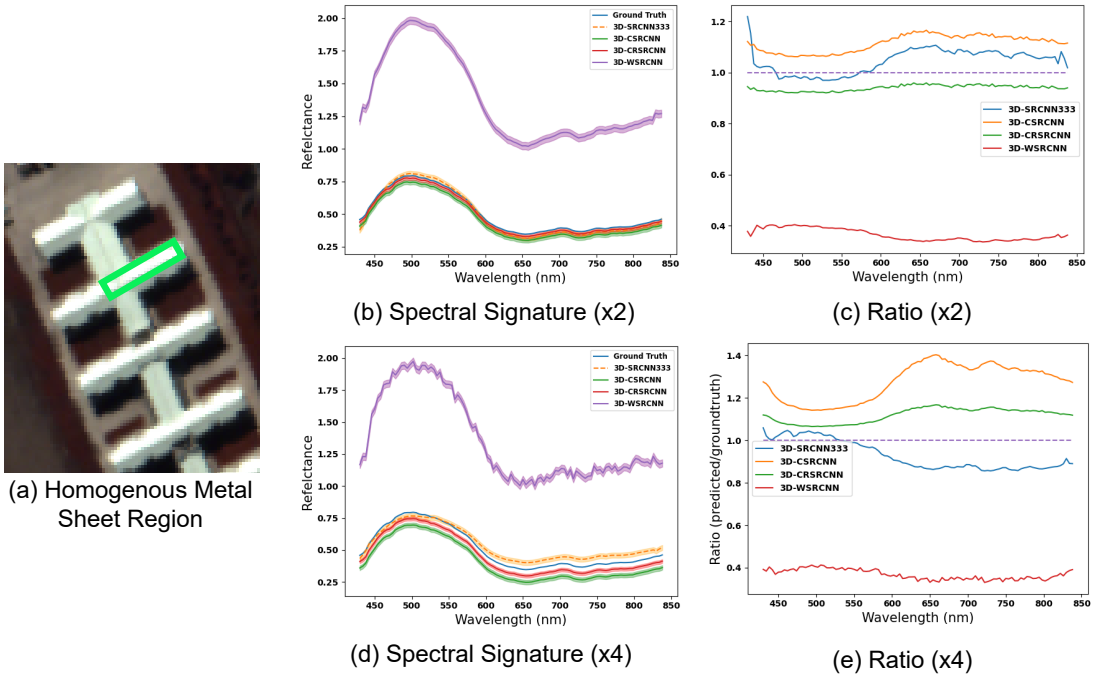


Figure 7.8: The results of testing 3D-SRCNN333, 3D-CSRCNN, 3D-CRSRCNN and 3D-WSRCNN on (a) homogeneous Metal Sheet region, where the spectral signature is plotted in (b) for scale factor $\times 2$, (c) shows the corresponding ratio plots, (d) shows the spectral signature for scale factor $\times 4$, and (e) shows the corresponding ratio plots. The overall order of performance is: 3D-CRSRCNN, 3D-SRCNN333, 3D-CSRCNN, and 3D-WSRCNN.

The next class to be examined is Gravel, as illustrated in Figure 7.9a. In contrast to the Metal Sheet class, distinctions within this category are more discernible for scale factor $\times 2$. Figure 7.9b clearly portrays that 3D-WSRCNN lags behind, while the remaining networks manage to approximate the GT to varying degrees. A clearer perspective is provided by Figure 7.9c, which highlights that 3D-CRSRCNN exhibits the highest similarity to the GT, followed by 3D-CSRCNN and then 3D-SRCNN333, with 3D-WSRCNN displaying the most pronounced distortions. When considering the scale factor $\times 4$, Figure 7.9d reveals an increased overlap among the top-performing networks

compared to the previous scale factor. Figure 7.9e highlights this overlap, particularly between 3D-CSRCNN and 3C-CRSRCNN, as they exhibit a similar pattern. These two networks outperform 3D-SRCNN333 until approximately wavelength 600 nm, at which point all three networks demonstrate highly similar performance.

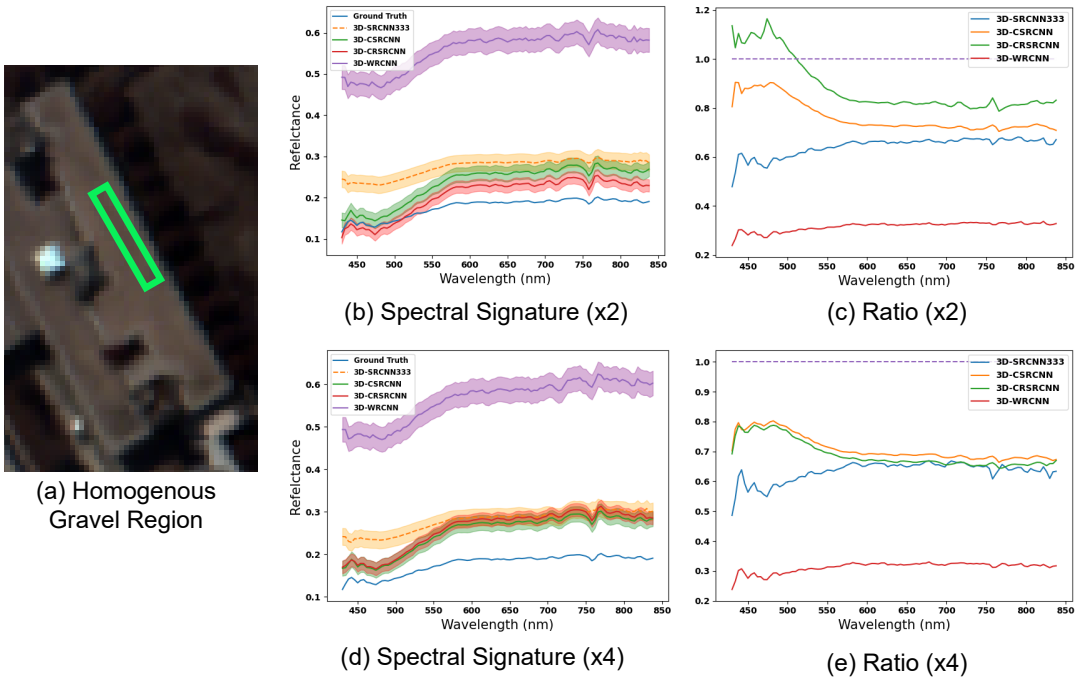


Figure 7.9: The results of testing 3D-SRCNN333, 3D-CSRCNN, 3D-CRSRCNN and 3D-WSRCNN on (a) homogeneous Gravel region, where the spectral signature is plotted in (b) for scale factor $\times 2$, (c) shows the corresponding ratio plots, (d) shows the spectral signature for scale factor $\times 4$, and (e) shows the corresponding ratio plots. The overall order of performance is:3D-CRSRCNN, 3D-CSRCNN, 3D-SRCNN333, and 3D-WSRCNN.

The examination now focuses on the homogeneous Asphalt region seen in Figure 7.10a. In evaluating the results for scale factor $\times 2$, Figure 7.10b distinctly illustrates that 3D-WSRCNN continues to exhibit substantial distortions compared to the other networks. Meanwhile, a noteworthy overlap is observed between 3D-CSRCNN and 3D-CRSRCNN, both mimicking the GT more closely than 3D-SRCNN333. A clearer insight is provided by Figure 7.10c, where initially, 3D-SRCNN333 and 3D-CSRCNN display similar performance up to approximately 500 nm. Subsequently, 3D-CSRCNN gradually approaches an asymptotic value of 1, while 3D-SRCNN333 remains rela-

tively consistent in terms of similarity. In contrast, 3D-CRSRCNN consistently exhibits the highest proximity to 1 among these networks. Upon investigating the results for scale factor $\times 4$, Figure 7.10d highlights an even more pronounced overlap between 3D-CSRCNN and 3D-CRSRCNN. This is reaffirmed in Figure 7.10c, where both networks outperform 3D-SRCNN333. Initially, 3D-CSRCNN slightly surpasses 3D-CRSRCNN, up until approximately 720 nm, after which 3D-CRSRCNN takes the lead. Overall, their performance appears to be roughly comparable in this scenario.

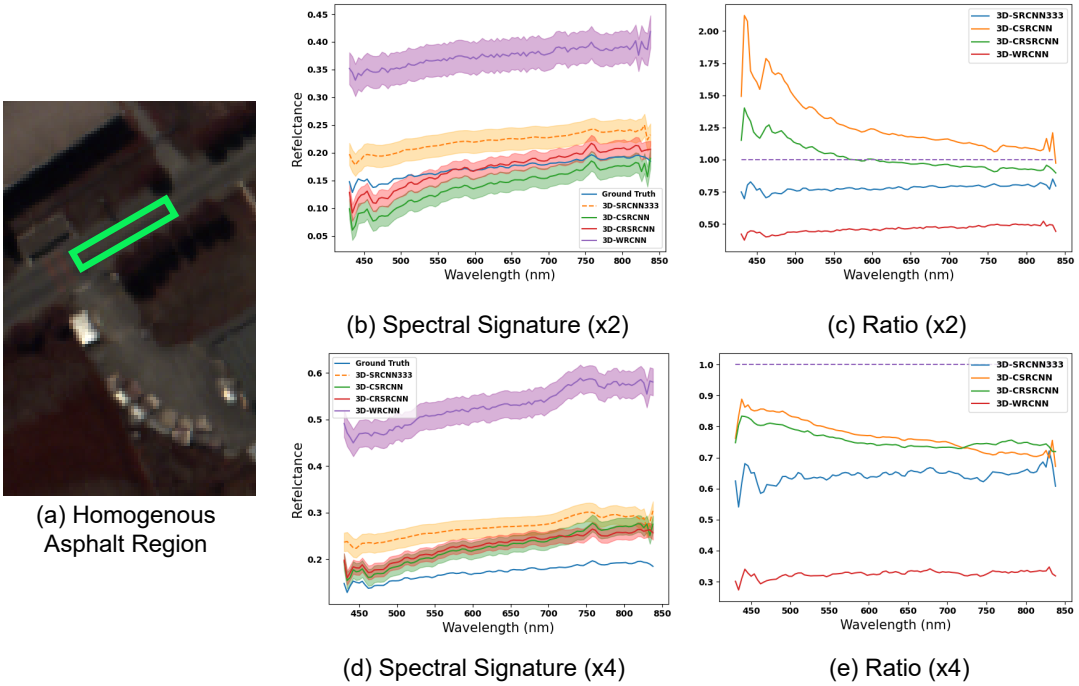


Figure 7.10: The results of testing 3D-SRCNN333, 3D-CSRCNN, 3D-CRSRCNN and 3D-WSRCNN on (a) homogeneous Asphalt region, where the spectral signature is plotted in (b) for scale factor $\times 2$, (c) shows the corresponding ratio plots, (d) shows the spectral signature for scale factor $\times 4$, and (e) shows the corresponding ratio plots. The overall order of performance is: 3D-CRSRCNN, 3D-CSRCNN, 3D-SRCNN333, and 3D-WSRCNN.

In general, the performance of 3D-CSRCNN closely resembles that of 3D-SRCNN333 in certain cases, but 3D-CRSRCNN often mitigates their spectral distortions. As a result, the developed complex networks demonstrates strong performance in the reflectance domain.

7.5.2 Washington DC Mall Experiments

Examining the results obtained from the Washington DC Mall dataset reveals a similar pattern to that of the Pavia University dataset. Figure 7.11 illustrates that 3D-WSRCNN experiences significant performance degradation, particularly at the $\times 4$ scale factor. Conversely, 3D-SRCNN333, 3D-CSRCNN, and 3D-CRSRCNN are capable of capturing more details. At the $\times 2$ scale factor, the differences between the predictions of these three networks are minimal. However, a clear distinction emerges when examining the $\times 4$ scale factor, where 3D-CRSRCNN exhibits fewer errors in its RMSE map.

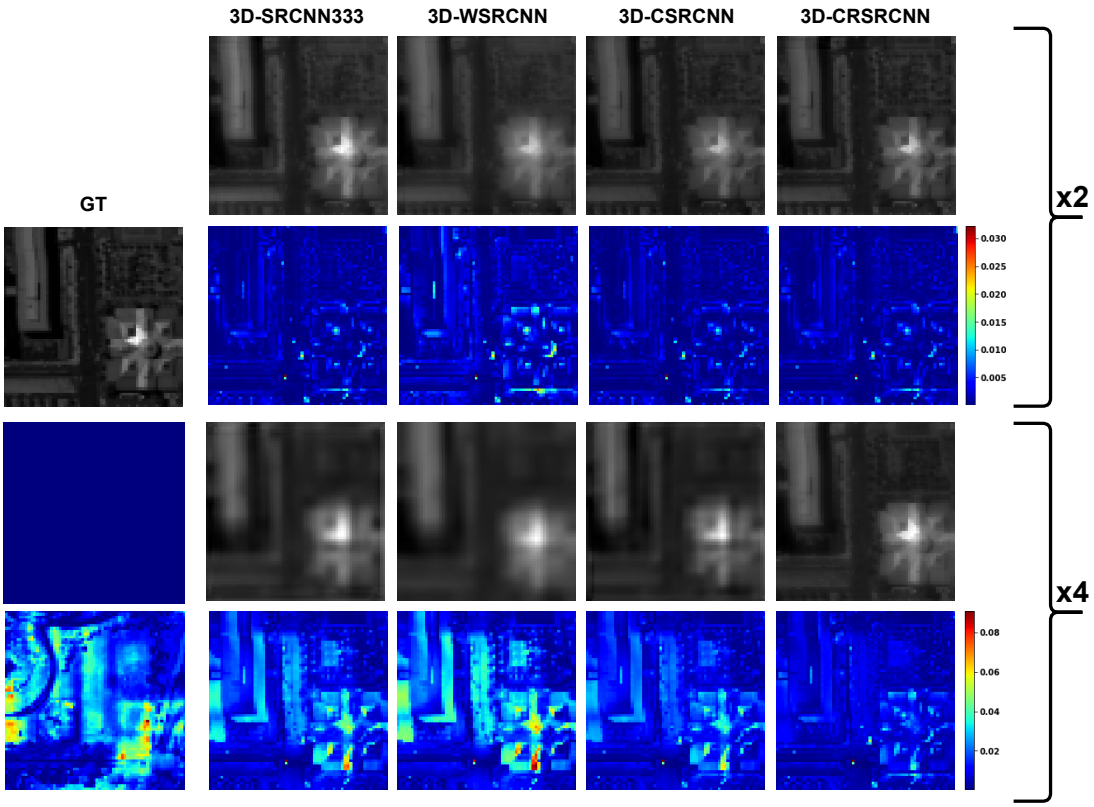
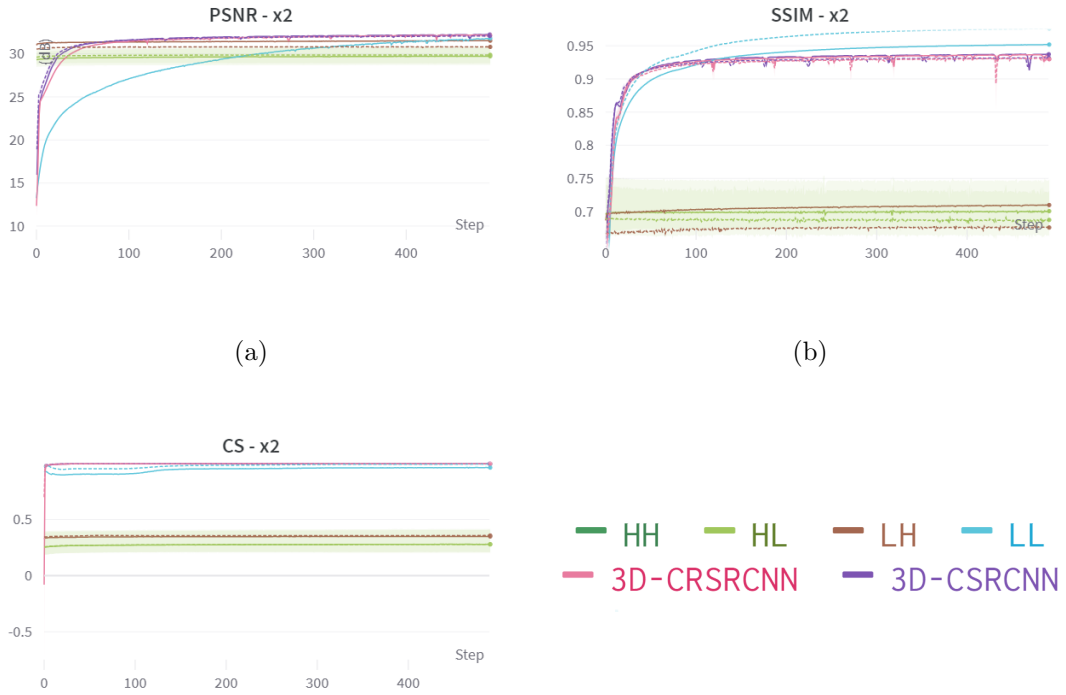


Figure 7.11: Visual results of the predicted HR-HSI for Washington DC Mall dataset by each method. For each scale factor, the top row shows the output of the method, and the bottom row shows the error map in terms of RMSE.

Similar to the Pavia University dataset, the training progress for the Washington DC Mall dataset, as shown in Figures 7.12 and 7.13, indicates that the LH, HL, and HH

components of 3D-WSRCNN do not benefit from the training process, as their PSNR, SSIM, and CS values remain largely unchanged, aside from a few fluctuations.



(a) (b)

(c)

Figure 7.12: This figure shows the progression of training 3D-CSRCNN, 3D-CRSRCNN, and the individual wavelet components of 3D-WSRCNN on Washington DC dataset. The solid line shows training results, while the dashed line shows validation results, and the shaded area represents the standard deviation. Figures 7.12a, 7.12b, and 7.12c show PSNR, SSIM, and CS respectively, for scale factor $\times 2$.

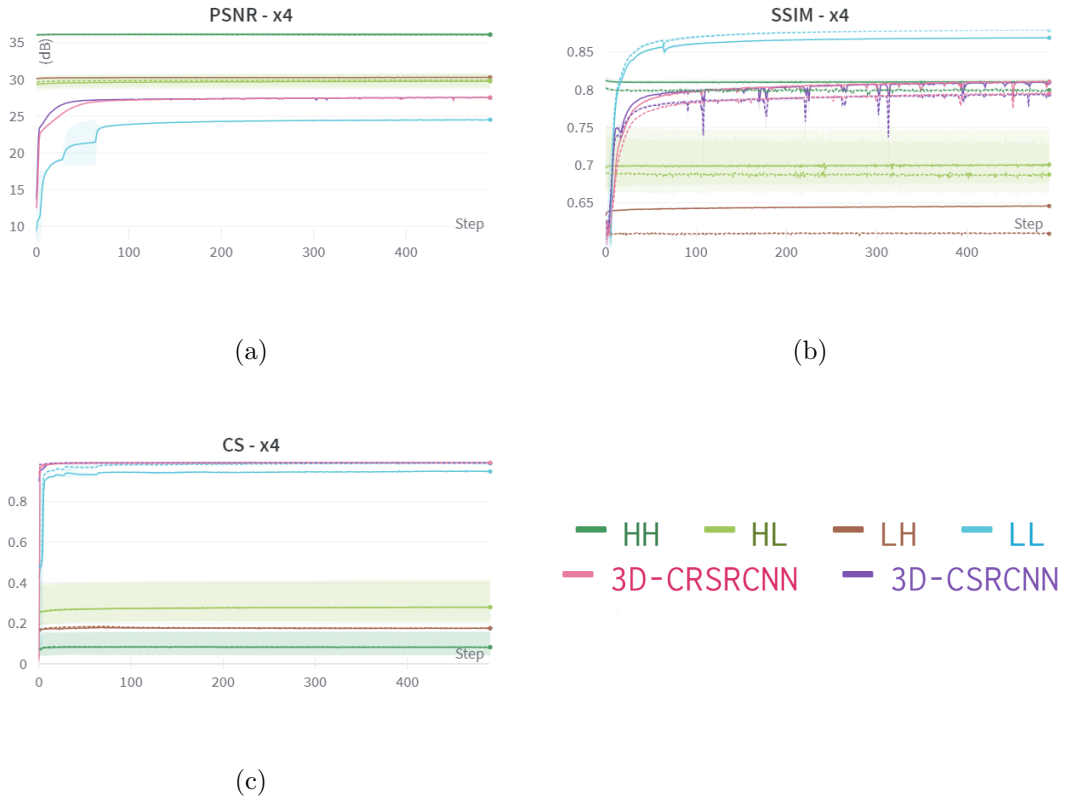
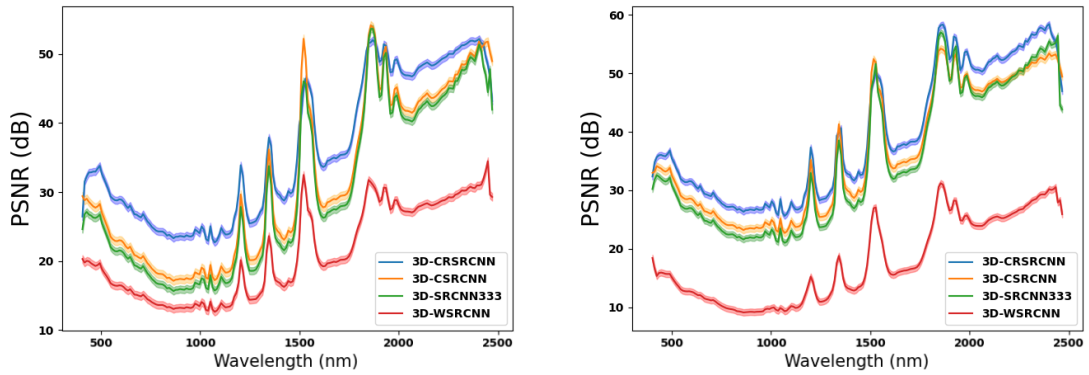


Figure 7.13: This figure shows the progression of training 3D-CSRCNN, 3D-CRSRCNN, and the individual wavelet components of 3D-WSRCNN on Washington DC dataset. The solid line shows training results, while the dashed line shows validation results, and the shaded area represents the standard deviation. Figures 7.13a, 7.13b, and 7.13c show PSNR, SSIM, and CS respectively, for scale factor $\times 2$.

However, the distortions produced by 3D-WSRCNN in the Washington DC Mall dataset are less severe compared to the Pavia University dataset. For instance, Figure 7.14b demonstrates that although the other networks achieve higher PSNR values across the spectral range, 3D-WSRCNN does not lag far behind. The network that achieves the highest PSNR for both scale factors is 3D-CRSRCNN.



(a) ×2 (b) ×4

Figure 7.14: PSNR results recorded for Washington DC Mall spectral range, where (a) shows the results for scale factor ×2 and (b) shows the results for scale factor ×4. The solid lines represent the mean, while the shaded areas represent the standard deviation.

To assess spectral fidelity, the study focuses on evaluating the spectral characteristics of various classes. As in previous chapters, the following classes are examined: Grass, Roof, and Water. For each class, a homogeneous region is selected, and the mean spectral signature is computed. This process is repeated five times for each network under evaluation, and the overall mean and standard deviation results are recorded.

The results for the Grass class are depicted in Figure 7.15, with the region of interest shown in Figure 7.15a. In the case of scale factor ×2, Figure 7.15b reveals a high degree of overlap between 3D-SRCNN333, 3D-CSRCNN, and 3D-CRSRCNN, while 3D-WRCNN exhibits substantial distortions. This observation is corroborated by the ratio plot in Figure 7.15c, which illustrates that 3D-CRSRCNN exhibits the closest resemblance to the ideal value of 1, followed by 3D-CSRCNN and 3D-SRCNN333.

In the case of scale factor ×4, despite the overlap seen in Figure 7.15d, the ratio plot in Figure 7.15e indicates that the networks exhibit more spikes and oscillations compared to scale factor ×2. Nonetheless, 3D-CRSRCNN and 3D-CSRCNN continue to outperform 3D-SRCNN333. Overall, the advantage offered by 3D-CSRCNN and 3D-CRSRCNN in this class is clearly evident.

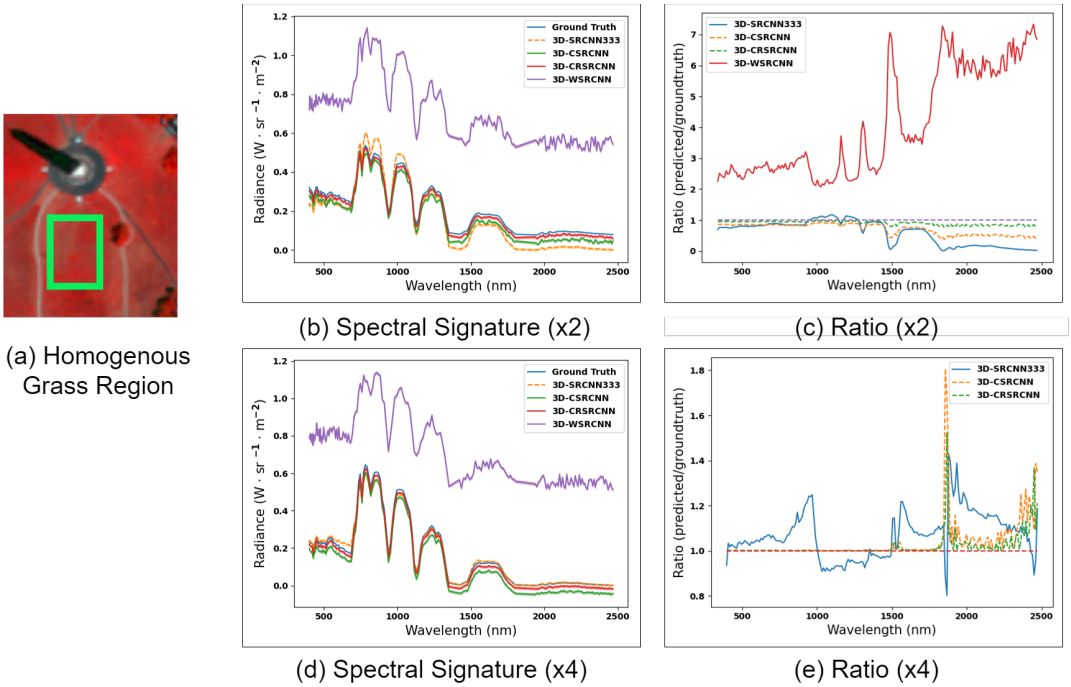


Figure 7.15: The results of testing 3D-SRCNN333, 3D-CSRCNN, 3D-CRSRCNN and 3D-WSRCNN on (a) homogeneous Grass region, where the spectral signature is plotted in (b) for scale factor $\times 2$, (c) shows the corresponding ratio plots, (d) shows the spectral signature for scale factor $\times 4$, and (e) shows the corresponding ratio plots. The overall order of performance is: 3D-CRSRCNN, 3D-CSRCNN, 3D-SRCNN333, and 3D-WSRCNN.

The Roof class is the next subject of examination, focusing on the region depicted in Figure 7.16. In the case of scale factor $\times 2$, the results presented in Figure 7.16b demonstrate the usual performance lag of 3D-WSRCNN. Additionally, it is evident that 3D-CSRCNN and 3D-CRSRCNN effectively mitigate some of the distortions introduced by 3D-SRCNN333. This observation is further substantiated by the corresponding ratio plot in Figure 7.16c. Due to significant distortions, 3D-WRCNN was excluded from this figure, which would otherwise obscure the performance of the other networks. Notably, in the spectral region beyond approximately 1800 nm, 3D-SRCNN333 outperforms 3D-CSRCNN and 3D-CRSRCNN for scale factor $\times 2$. For scale factor $\times 4$, a larger overlap is observed between the top three networks, as depicted in Figure 7.16d. The ratio plots reveal that 3D-CSRCNN effectively mitigates some of the errors from 3D-SRCNN333, while 3D-CRSRCNN further reduces these errors, as illustrated in Figure

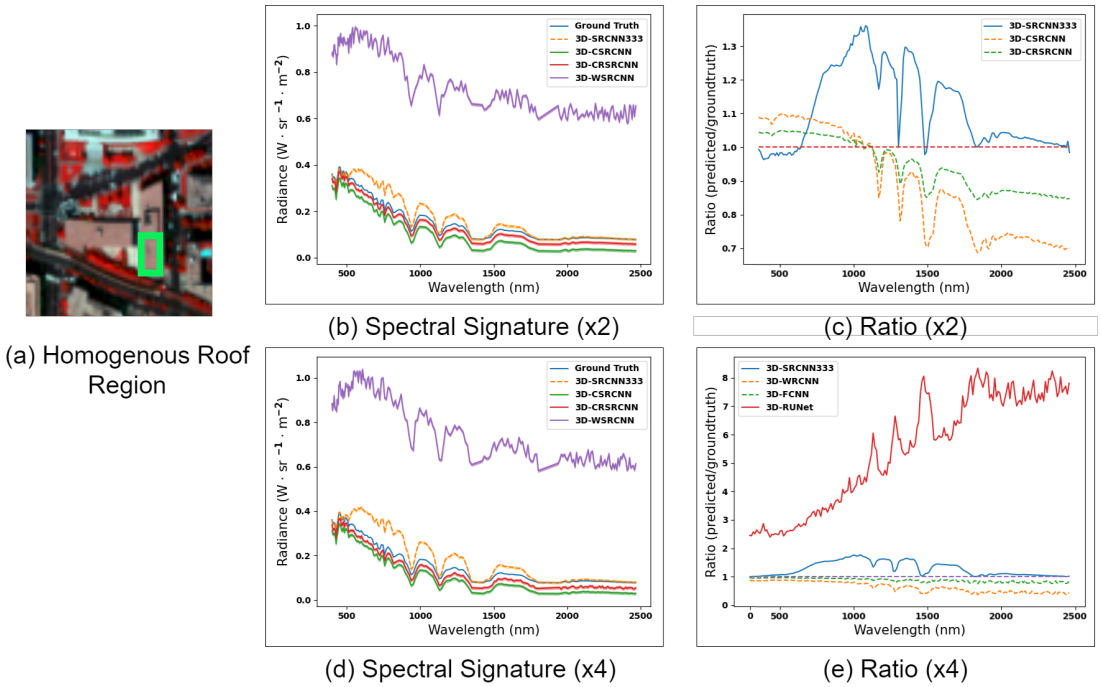


Figure 7.16: The results of testing 3D-SRCNN333, 3D-CSRCNN, 3D-CRSRCNN and 3D-WSRCNN on (a) homogeneous Roof region, where the spectral signature is plotted in (b) for scale factor $\times 2$, (c) shows the corresponding ratio plots, (d) shows the spectral signature for scale factor $\times 4$, and (e) shows the corresponding ratio plots. The overall order of performance is: 3D-CRSRCNN, 3D-CSRCNN, 3D-SRCNN333, and 3D-WSRCNN.

7.16e. Overall, both complex networks show notable enhancements in this class.

Finally, the results for the Water class, corresponding to the region depicted in Figure 7.17a, are displayed in Figures 7.17b and 7.17c for scale factor $\times 2$. These results exhibit a substantial overlap, which is similarly reflected in the ratio plot. However, in this particular scenario, 3D-CSRCNN exhibits greater resemblance to the GT compared to 3D-CRSRCNN. Concerning scale factor $\times 4$, the spectral distortions introduced by 3D-SRCNN333 become more pronounced, but both complex networks manage to mitigate these distortions. However, in this case, the advantage of 3D-CRSRCNN over 3D-CSRCNN is not immediately evident, unlike Grass and Roof classes.

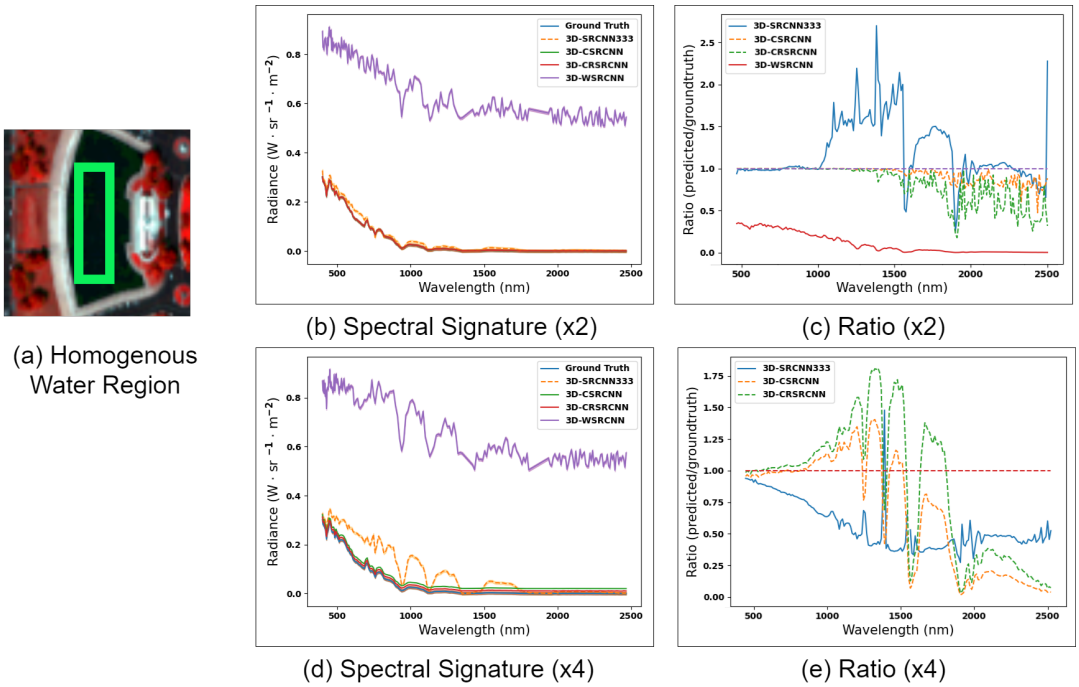


Figure 7.17: The results of testing 3D-SRCNN333, 3D-CSRCNN, 3D-CRSRCNN and 3D-WSRCNN on (a) homogeneous Water region, where the spectral signature is plotted in (b) for scale factor $\times 2$, (c) shows the corresponding ratio plots, (d) shows the spectral signature for scale factor $\times 4$, and (e) shows the corresponding ratio plots. The overall order of performance is: 3D-CSRCNN, 3D-CRSRCNN, 3D-SRCNN333, and 3D-WSRCNN.

Overall, both 3D-CSRCNN and 3D-CRSRCNN manage to correct some of the spectral distortions caused by 3D-SRCNN333, indicating their robustness not only in the reflectance domain but also in the radiance domain. 3D-WSRCNN shows more distortions in the radiance domain compared to reflectance domain.

7.5.3 Quantitative Comparison

The quantitative outcomes for both the Pavia University and Washington DC Mall datasets are summarized in Table 7.2. The performance of 3D-WSRCNN is inadequate, although its performance on the Washington DC Mall dataset is comparatively better than on the Pavia University dataset. It is possible that 3D-WSRCNN requires a larger dataset to fully unleash its potential. On the other hand, 3D-CRSRCNN achieves the highest PSNR, SSIM, and SAM scores. The positive influence of the added residual

connection on the network is evident in the superior performance of 3D-CRSRCNN compared to 3D-CSRCNN. Furthermore, 3D-CRSRCNN outperforms 3D-SRCNN333 by 1.205dB, 0.0041, and 0.73°, respectively, for the ×2 scale factor with the Pavia University dataset. For the ×4 scale factor, the improvement is less pronounced in terms of PSNR, SSIM, and SAM, which amounts to 0.049dB, 0.0005, and 0.21°, respectively. Moreover, despite the larger number of data samples provided by the Washington DC Mall dataset, the performance gap between 3D-SRCNN333 and 3D-CRSRCNN is even narrower compared to what was observed in the Pavia University dataset for the ×2 scale factor. However, the gap widens for the ×4 scale factor. Finally, the training time of 3D-CRSRCNN should be taken into consideration, as it requires approximately 3.6 times more training time than 3D-SRCNN333 for the Pavia University dataset and approximately 4.5 times more training time for the Washington DC Mall dataset. The trade-off between training time and quantitative performance depends on the specific nature of the application.

Table 7.2: Results summary of complex- and wavelet-based CNNs compared to their real-valued counterpart the 3D-SRCNN333 in terms of PSNR (dB), SSIM, and SAM (°), in addition to training time in minutes for scale factors ×2 and ×4. Δ shows the difference in performance between each network and 3D-SRCNN333 for each metric.

Dataset	3D-SRCNN333		3D-WSRCNN		3D-CSRCNN		3D-CRSRCNN	
	x2	x4	x2	x4	x2	x4	x2	x4
PU	31.534 ± 0.196	25.628 ± 0.027	28.625 ± 0.295 Δ = 2.909 ↓	22.953 ± 0.288 Δ = 2.675 ↓	32.727 ± 0.115 Δ = 1.193 ↑	25.659 ± 0.102 Δ = 0.031 ↑	**32.739 ± 0.131** Δ = 1.205 ↑	**25.677 ± 0.012** Δ = 0.049 ↑
	0.9222 ± 0.0009	0.6990 ± 0.0003	0.8533 ± 0.0055 Δ = 0.0689 ↓	0.5020 ± 0.074 Δ = 0.1970 ↓	0.9250 ± 0.0010 Δ = 0.0278 ↑	0.6921 ± 0.0003 Δ = 0.0069 ↓	**0.9263 ± 0.0004** Δ = 0.0041 ↑	**0.6995 ± 0.0001** Δ = 0.0005 ↑
	5.31 ± 0.15	8.54 ± 0.20	10.48 ± 2.28 Δ = 0.0689 ↑	11.50 ± 2.12 Δ = 0.1970 ↑	4.66 ± 0.11 Δ = 0.0278 ↓	8.45 ± 0.08 Δ = 0.0069 ↓	**4.58 ± 0.09** Δ = 0.73 ↓	**8.33 ± 0.12** Δ = 0.21 ↓
	25.6		51.2		83.3		91.7	
DC	31.831 ± 0.012	27.576 ± 0.0330	29.103 ± 0.174 Δ = 2.728 ↓	25.264 ± 0.223 Δ = 2.312 ↓	32.011 ± 0.034 Δ = 0.18 ↑	27.574 ± 0.020 Δ = 0.002 ↓	**32.088 ± 0.021** Δ = 0.257 ↑	**27.580 ± 0.014** Δ = 0.004 ↑
	0.9271 ± 0.0005	0.7955 ± 0.0017	0.89500.0072 Δ = 0.0320 ↓	0.686 ± 0.0067 Δ = 0.1095 ↓	0.9306 ± 0.0022 Δ = 0.0035 ↑	0.7978 ± 0.0004 Δ = 0.0023 ↑	**0.9316 ± 0.0005** Δ = 0.0045 ↑	**0.7960 ± 0.0021** Δ = 0.0005 ↑
	4.50 ± 0.07	7.46 ± 0.05	5.97 ± 0.10 Δ = 0.0689 ↑	8.93 ± 0.32 Δ = 4.43 ↑	4.42 ± 0.05 Δ = 0.08 ↓	7.35 ± 0.13 Δ = 0.11 ↓	**4.39 ± 0.04** Δ = 0.11 ↓	**7.20 ± 0.17** Δ = 0.26 ↓
	66.7		133.4		291.8		300	

7.5.4 Pavia Center Experiments

To gauge the true effectiveness of the developed methods, it is useful to observe the performance of the developed networks when trained on Pavia University dataset and then tested on Pavia Center dataset. Recall from Section 2.4.1 and Table 2.1 that Pavia Center dataset and Pavia University dataset share the same characteristics.

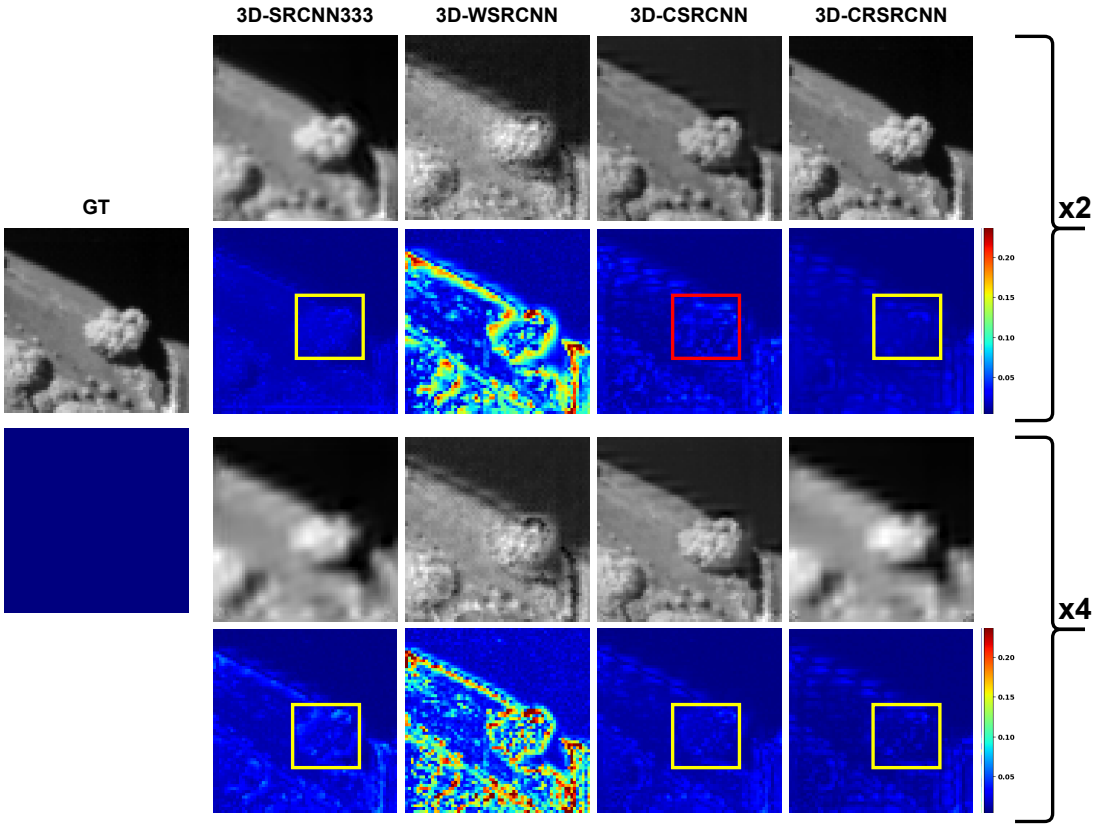
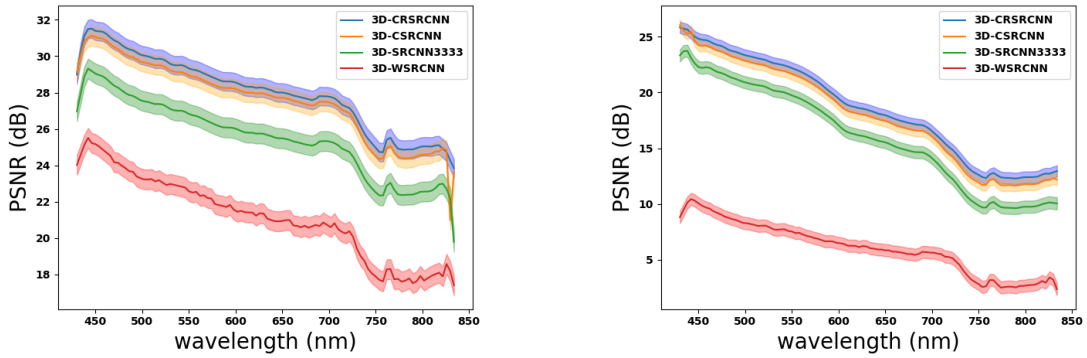


Figure 7.18: Visual results of the predicted HR-HSI for Pavia Center dataset by each method. For each scale factor, the top row shows the output of the method, and the bottom row shows the error map in terms of RMSE. The yellow box shows regions of improvement, while the red box shows regions of deterioration.

First, the qualitative findings are investigated through the visual results seen in Figure 7.18. For scale factor $\times 2$, the 3D-CRSRCNN shows less spectral distortions than 3D-SRCNN333. On the other hand, 3D-CSRCNN shows improvement in some areas outlined by the yellow box, and deterioration in other areas outside of that box. 3D-WSRCNN shows the worst performance among all networks. The PSNR plots for

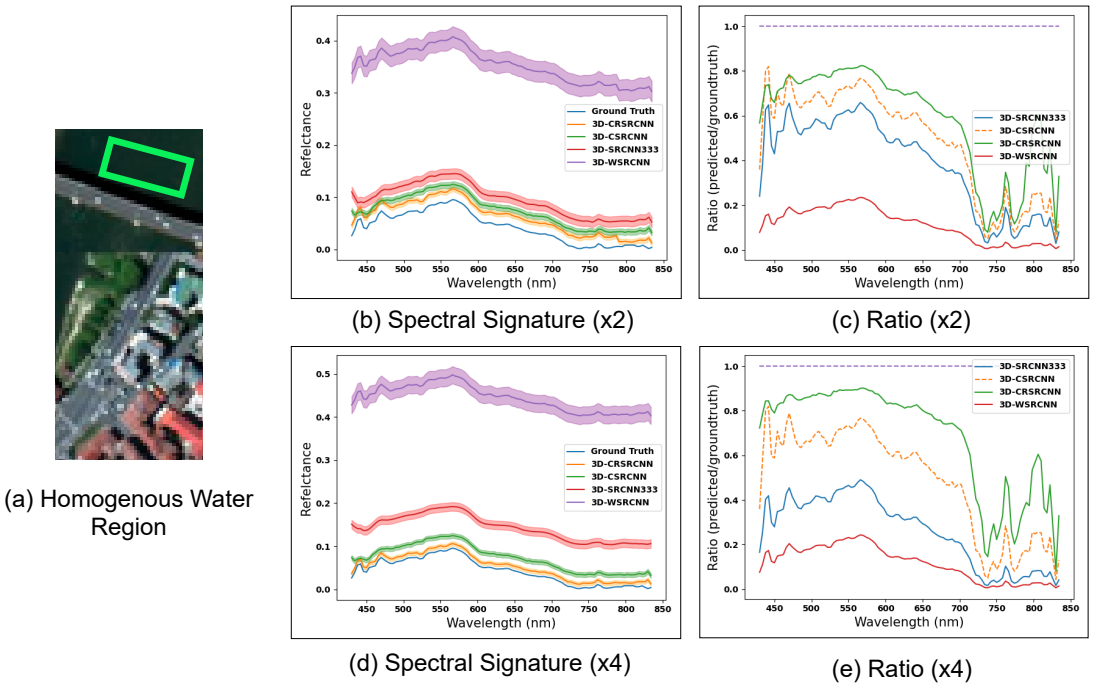
scale factor ×2 shown in Figure 7.19a show similar observation, as it becomes apparent that the 3D-WSRCNN noticeably falls behind, while the other networks show high PSNR with overlapping uncertainty regions. Figure 7.19b show the same observation for scale factor ×4. These results are consistent with Pavia University dataset, however, the overlap is even higher here. The advantage of both complex-domain versions of 3D-SRCNN333 are less evident in comparison to Pavia University.



(a) ×2 (b) ×4

Figure 7.19: PSNR results recorded for Pavia Center spectral range, where (a) shows the results for scale factor ×2 and (b) shows the results for scale factor ×4. The solid lines represent the mean, while the shaded areas represent the standard deviation.

The examination of spectral fidelity for Water class is seen in Figure 7.19. For scale factor ×2, The performance between 3D-SRCNN333, 3D-CSRCNN, and 3D-CRSRCNN overlaps once again, while 3D-WSRCNN falls behind. This is confirmed by the ratio plot, which shows that the top 3 networks have very close performance. The same observation is seen in the spectral signatures of scale factor ×4 seen in Figure 7.20, with the overlap being higher between 3D-CRSRCNN and 3D-CSRCNN.



(a) Homogenous Water Region

(b) Spectral Signature (x2)

(c) Ratio (x2)

(d) Spectral Signature (x4)

(e) Ratio (x4)

Figure 7.20: The results of testing MSE, Charbonnier, and the hybrid loss function on (a) homogeneous Water region, where the spectral signature is plotted in (b) for scale factor $\times 2$, (c) shows the corresponding ratio plots, (d) shows the spectral signature for scale factor $\times 4$, and (e) shows the corresponding ratio plots. 3D-CSRCNN and 3D-CRSRCNN correct the distortions from 3D-SRCNN333.

The results for Tile class are seen in Figure 7.21. For scale factor $\times 2$, the advantage provided by the 3D-CRSRCNN is more evident compared to Water class, as seen in Figure 7.21b. This is confirmed by the corresponding ratio plot seen in Figure 7.21c. As for scale factor $\times 4$, the same observation can be made, but the overlap between the networks is higher.

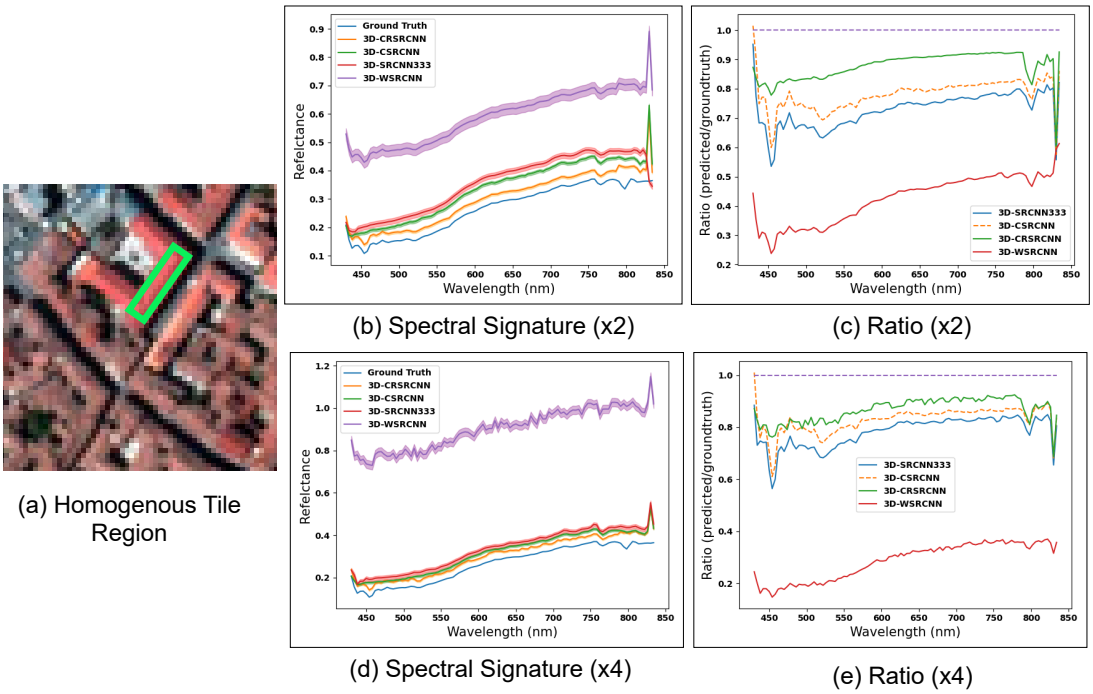


Figure 7.21: The results of testing MSE, Charbonnier, and the hybrid loss function on (a) homogeneous Water region, where the spectral signature is plotted in (b) for scale factor $\times 2$, (c) shows the corresponding ratio plots, (d) shows the spectral signature for scale factor $\times 4$, and (e) shows the corresponding ratio plots. 3D-CSRCNN and 3D-CRSRCNN correct the distortions from 3D-SRCNN333.

Table 7.3 provides the quantitative results of testing the networks on the Pavia Center dataset, focusing on PSNR, SSIM, and SAM metrics. These results corroborate the qualitative findings discussed earlier and the spectral fidelity analysis presented in this section. Within the table, the Δ row sheds light on the extent of improvement or deterioration introduced by each network compared to 3D-SRCNN333 for each respective metric. Notably, the improvements remain consistent across all metrics for 3D-CSRCNN and 3D-CRSRCNN, maintaining the same performance order observed in the Pavia University results, albeit with a lesser degree of improvement compared to Pavia University.

Table 7.3: Quantitative results of training 3D-SRCNN333, 3D-WSRCNN, 3D-CSRCNN, and 3D-CRSRCNN using Pavia University dataset and then testing on Pavia Center dataset reveal that the order of performance remains consistent with that of Pavia University in terms of PSNR, SSIM, and SAM. Δ shows the difference between each network and 3D-SRCNN333 for each metric.

Loss Function	×2			×4		
	PSNR (dB)	SSIM	SAM (°)	PSNR (dB)	SSIM	SAM (°)
3D-SRCNN333	30.633 ± 0.151	0.9180 ± 0.0001	5.54 ± 0.10	25.580 ± 0.121	0.7055 ± 0.0007	8.60 ± 0.17
3D-WSRCNN	25.887 ± 2.29 $\Delta = 4.746 \downarrow$	0.8051 ± 0.0097 $\Delta = 0.1129 \downarrow$	11.86 ± 2.56 $\Delta = 6.32 \uparrow$	21.356 ± 2.35 $\Delta = 4.224 \downarrow$	0.5995 ± 0.0095 $\Delta = 0.106 \downarrow$	12.33 ± 2.44 $\Delta = 3.73 \uparrow$
3D-CSRCNN	30.820 ± 0.120 $\Delta = 0.187 \uparrow$	0.9201 ± 0.0004 $\Delta = 0.0021 \uparrow$	5.30 ± 0.02 $\Delta = 0.24 \downarrow$	25.897 ± 0.195 $\Delta = 0.317 \uparrow$	0.6999 ± 0.0005 $\Delta = 0.0056 \uparrow$	8.50 ± 0.11 $\Delta = 0.10 \downarrow$
3D-CRSRCNN	$\mathbf{30.959 \pm 0.103}$ $\Delta = 0.326 \uparrow$	$\mathbf{0.9223 \pm 0.0002}$ $\Delta = 0.0043 \uparrow$	$\mathbf{5.25 \pm 0.03}$ $\Delta = 0.29 \downarrow$	$\mathbf{26.001 \pm 0.110}$ $\Delta = 0.721 \uparrow$	$\mathbf{0.7015 \pm 0.0003}$ $\Delta = 0.0040 \uparrow$	$\mathbf{8.44 \pm 0.05}$ $\Delta = 0.16 \downarrow$

7.6 Summary

This chapter introduced an approach that utilizes complex and wavelet analysis to construct a CNN for enhancing the spatial resolution of HSI. The architecture of the network is based on the concept of 3D-SRCNN333. The network operates on each wavelet component of the HSI cube individually. Additionally, a complex domain network is designed by transforming the HSI cube using 2D-FFT. Two versions of this network are created: the first one is the 3D-CSRCNN, which does not incorporate a residual connection, and the second one is called 3D-CRSRCNN, which incorporates a residual connection. Experimental results on the Pavia University and Washington DC Mall datasets demonstrate that 3D-WSRCNN performs poorly due to the limited benefit obtained by its high-frequency components during training. It fails to achieve satisfactory performance in terms of PSNR, SSIM, SAM, and visual quality. In contrast, 3D-CSRCNN exhibits excellent performance and outperforms 3D-SRCNN333. The inclusion of a residual connection in the network gives 3D-CRSRCNN an additional advantage over 3D-CSRCNN. However, it should be noted that 3D-CRSRCNN has a relatively longer training time, which can be considered a drawback compared to

the shorter training time of 3D-SRCNN333, depending on the magnitude of the quantitative performance gap between the two networks. Tests on Pavia Center dataset reveal that the 3D-CRSRCNN maintains better performance compared to the other networks, despite being trained on Pavia University dataset. This work has been presented and published in the SPIE Photonex proceedings of Hyperspectral Imaging and Applications II [365].

Chapter 8

Conclusion and Future Work

8.1 Concluding Remarks

This thesis aimed to investigate the topic of HSI-SISR and the several challenges surrounding it. The research encompassed detailed literature review that bridged the past and present, thorough analysis of various devised CNN architectures, loss functions, and pre-processing techniques, which were all evaluated quantitatively and qualitatively. To recapitulate, the research objectives of this thesis aimed to tackle the following challenges:

- Devising a network that offers state-of-the-art performance without compromising spatial and/or time complexity
- Enhancing the spatial resolution of HSI without deteriorating the spectral fidelity
- Solving data scarcity problem in the currently available datasets to overcome overfitting in 3D-CNNs
- Exploring HSI-SISR in complex and wavelet domains

The thesis is divided into four chapters, each chapter tackles one of the challenges listed above. The key findings of the thesis are summarized as follows:

- Extending the traditional 2D-CNNs to 3D offers a solid ground to build an architecture for HSI-SISR. Training three architectures of different depths: 3D-

SRCNN333 (shallow), 3D-WRCNN (medium), and 3D-RUNet (deep) on datasets of different sizes: Indian Pines (small), Pavia University (medium), Washington DC Mall (large) shows that the high complexity caused by deep networks is not worth the compromise, as it does not offer a much better performance compared to shallow and deep networks, especially with small datasets. Furthermore, the modified filter sizes of 3D-SRCNN333 from (9,1,5) to (3,3,3) demonstrates the effect of the filter size on the artifacts caused around the border of the image. 3D-SRCNN333 offers a fair compromise between training time and performance quantitatively and qualitatively.

- The spectral fidelity can be preserved further by introducing a hybrid spectral-spatial loss function that offers the best of both spatial and spectral qualities. Its hyperparameter controls the balance between spatial and spectral quality, which can be tuned using BOA. This function performs better than various traditional loss functions.

- To enhance dataset variety in HSI-SISR, the utilization of Data Augmentation techniques can be beneficial. However, it is crucial to conduct a thorough assessment of these techniques to ensure they do not compromise spectral fidelity. An effective Data Augmentation technique is defined as one that boosts PSNR and SSIM and minimized SAM. The experiments conducted in this chapter indicate that the most effective Data Augmentation techniques for HSI-SISR are those that minimally alter image features. Moreover, Data Augmentation techniques that result in the presence of black borders in the HSI may enhance PSNR and SSIM, but they negatively impact SAM, rendering them ineffective. Notably, techniques such as CutMix, CutBlur, and CutMixBlur exhibit improved localization of enhancements and reduced artifacts near sharp edges. Random Stacking, which involves applying the various effective Data Augmentation techniques in a randomized manner, has demonstrated significant performance enhancements.

- Taking inspiration from the successful outcome of 3D-SRCNN333, a 3D-WSRCNN model was developed to investigate the impact of enhancing wavelet compo-

nents on HSI-SISR. Additionally, a 3D-CRSRCNN model was devised to explore the effects of HSI-SISR in the complex domain. An alternative version of 3D-CRSRCNN was also created, incorporating a residual connection. Detailed performance analysis revealed that 3D-CRSRCNN exhibited excellent performance compared to both the real-valued 3D-SRCNN333 and 3D-CRSRCNN. On the other hand, 3D-WSRCNN demonstrated poor performance, potentially due to inadequate training of high-frequency components. The severity of this poor performance was somewhat mitigated in the Washington DC Mall dataset, suggesting a possible correlation with dataset size. Furthermore, it is important to consider that 3D-CRSRCNN requires nearly four times the training time of 3D-SRCNN333, which should be taken into account when assessing the quantitative and qualitative enhancements provided by 3D-CRSRCNN compared to 3D-SRCNN333.

8.2 Future Work

Based on these findings, various possibilities are open for the future direction of this research. For example, 3D-SRCNN333 can be mixed with 2D components, as some authors argue that combining 3D with 2D operations enhances spectral-spatial correlations. The 2D operations can be in the form of Attention mechanism. For example, Squeeze-and-Excitation (SE) is a powerful attention technique that has been previously used for HSI classification. Preliminary experiments show that introducing SE into the architecture of 3D-SRCNN333 indeed enhances the performance [366].

Although the performance of 3D-WSRCNN was unsatisfactory, the concept remains intriguing as it provides the network with increased control over high and low frequency components. The underperformance may have been due to fully decorrelating these components. Instead of separating them entirely, utilizing wavelet layers and injecting them within 3D-SRCNN333 has the potential to improve the performance and facilitate better network training. Additionally, it would be beneficial to experiment with Data Augmentation techniques on 3D-WSRCNN to determine if they can enhance the training performance of high frequency components.

While 3D-CSRCNN and 3D-CRSRCNN demonstrate impressive performance, further investigation is required to fully unlock the network's capabilities. For instance, exploring a wider range of architectures and incorporating additional techniques like the previously mentioned SE technique could be beneficial. Additionally, the network was evaluated using a complex MSE loss function. However, it would be advantageous to test alternative loss functions adapted to the complex domain, similar to those examined in Chapter 5, which could potentially lead to the development of a hybrid spectral-spatial complex loss function.

It is worth noting that the changes in SSIM observed throughout the thesis are very small compared to PSNR. The reason behind this was stated in Section 2.5.2. In the future, an exploration of improved methods for computing SSIM that better reflect the spatial characteristics of HSI will be undertaken.

Finally, an alternative to Data Augmentation is the utilization of Generative Adversarial Networks (GANs) to generate additional data. However, this approach introduces a fresh set of difficulties since GANs require training on datasets that consist of both pre-distorted and post-distorted HSI samples, which are currently unavailable. Overcoming this challenge could significantly contribute to the advancement of the HSI-SISR field.

Bibliography

[1] "The History of Remote Sensing," Planetek Italia, Feb 2021. [Online]. Available: https://www.planetek.it/eng/training_courses/online_manuals/on_line_course_of_remote_sensing/2_the_history_of_remote_sensing

[2] "Explorer 6," NASA Space Science Data Coordinated Archive, 2022. [Online]. Available: https://nssdc.gsfc.nasa.gov/nmc/spacecraft/display.action?id=1959-004A

[3] K. C. Ruffner, *Corona: America's First Satellite Program*, ser. ADA499635, J. K. McDonald, M. S. MacAuliffe, S. A. Koch, and M. S. Warner, Eds. CENTRAL INTELLIGENCE AGENCY WASHINGTON DC CENTER FOR STUDY OF INTELLIGENCE, 1995.

[4] M. Benson and J. Faundeen, "The U.S. Geological Survey, Remote Sensing, and Geoscience Data: Using Standards to Serve Us All," in *International Geoscience and Remote Sensing Symposium (IGARSS). Taking the Pulse of the Planet: The Role of Remote Sensing in Managing the Environment. Proceedings (Cat. No.00CH37120)*, vol. 3, 2000, pp. 1202–1204 vol.3.

[5] E. O'Connor and A. McDonald, "Applications Of Remote Sensing For Geological Mapping In Eastern Egypt," in *International Geoscience and Remote Sensing Symposium, 'Remote Sensing: Moving Toward the 21st Century'*, vol. 1, 1988, pp. 631–632.

[6] B. Fu, P. Shi, H. Fu, Y. Ninomiya, and J. Du, "Geological Mapping Using Multispectral Remote Sensing Data in the Western China," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2019, pp. 5583–5586.

[7] L. Zhizhong, Y. Rihong, D. Fuxing, D. Peijun, Z. Xianfeng, T. Bingxiang, Z. Huijie, and S. Hongjun, "A Review on the Geological Applications of Hyperspectral Remote Sensing Technology," in *4th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2012, pp. 1–4.

[8] Q. Tan, J. Gao, and X. Li, "Application of SAR Remote Sensing Data to Lithological Mapping: A Case Study in Railway Geological Survey," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 2, 2009, pp. II–342–II–344.

[9] A. B. Pour and M. Hashim, "Remote Sensing Analysis of Geological Structures in Peninsular Malaysia using PALSAR Data," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2016, pp. 6067–6069.

[10] S. Barma, S. Damarla, and S. K. Tiwari, "Semi-Automated Technique for Vegetation Analysis in Sentinel-2 Multi-Spectral Remote Sensing Images Using Python," in *4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, 2020, pp. 946–953.

[11] S. J. Mills, M. P. Gerardo Castro, Z. Li, J. Cai, R. Hayward, L. Mejias, and R. A. Walker, "Evaluation of Aerial Remote Sensing Techniques for Vegetation Management in Power-Line Corridors," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 9, pp. 3379–3390, 2010.

[12] W. Ciężkowski, P. Sikorski, P. Babańczyk, D. Sikorska, and J. Chormański, "Algorithm for Urban Spontaneous Green Space Detection based on Optical Satellite Remote Sensing," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2020, pp. 4430–4433.

[13] M. Piles, G. Camps-Valls, D. Chaparro, D. Entekhabi, A. G. Konings, and T. Jagdhuber, "Remote Sensing of Vegetation Dynamics in Agro-ecosystems Us-

ing SMAP Vegetation Optical Depth and Optical Vegetation Indices," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2017, pp. 4346–4349.

[14] Y. Lin, L. Zhang, and N. Wang, "A New Time Series Change Detection Method for Landsat Land use and Land Cover Change," in *10th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp)*, 2019, pp. 1–4.

[15] D. Yin, X. Chen, and S. Zhao, "Quantitative Typical Land Cover Remote Sensing and its Application in Earthquake Evaluation," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2018, pp. 7304–7307.

[16] O. Bounouh, H. Essid, and I. R. Farah, "Prediction of Land Use/Land Cover Change Methods: A Study," in *International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, 2017, pp. 1–7.

[17] A. Alem and S. Kumar, "Deep Learning Methods for Land Cover and Land Use Classification in Remote Sensing: A Review," in *8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, 2020, pp. 903–908.

[18] G. Zheng, X. Li, and B. Liu, "AI-Based Remote Sensing Oceanography - Image Classification, Data Fusion, Algorithm Development and Phenomenon Forecast," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2019, pp. 7940–7943.

[19] C. Wang, X. Wang, and J. C. B. da Silva, "Studies of Internal Waves in the Strait of Georgia Based on Remote Sensing Images," in *IEEE International Geoscience and Remote Sensing Symposium*, 2020, pp. 3549–3551.

[20] G. K. Devi, B. Ganasri, and G. Dwarakish, "Applications of Remote Sensing in Satellite Oceanography: A Review," *Aquatic Procedia*, vol. 4, pp. 579–584, 2015, international Conference on Water Resources, Coastal and Ocean Engineering

(ICWRCOE'15). [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2214241X15000760

[21] R. G. Scarrott, F. Cawkwell, M. Jessopp, E. O'Rourke, C. Cusack, and K. de Bie, "From Land to Sea, a Review of Hypertemporal Remote Sensing Advances to Support Ocean Surface Science," *Water*, vol. 11, no. 11, 2019. [Online]. Available: https://www.mdpi.com/2073-4441/11/11/2286

[22] B. Thies and J. Bendix, "Satellite Based Remote Sensing of Weather and Climate: Recent Achievements andFuture Perspectives," *Meteorological Applications*, vol. 18, no. 3, pp. 262–295, 2011. [Online]. Available: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/met.288

[23] T. Greicius, "Sentinel-6 Mission Overview," NASA, Sep 2020. [Online]. Available: https://www.nasa.gov/sentinel-6/overview

[24] A. A. Rais, A. A. Suwaidi, and H. Ghedira, "DubaiSat-1: Mission Overview, Development Status and Future Applications," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 5, 2009, pp. V–196–V–199.

[25] A. A. Suwaidi, "DubaiSat-2 Mission Overview," in *Sensors, Systems, and Next-Generation Satellites XVI*, R. Meynart, S. P. Neeck, and H. Shimoda, Eds., vol. 8533, International Society for Optics and Photonics. SPIE, 2012, pp. 220 – 224. [Online]. Available: https://doi.org/10.1117/12.974469

[26] "WMO OSCAR — List of all Satellites," OSCAR. [Online]. Available: https://space.oscar.wmo.int/satellites

[27] M. Al-Saad, N. Aburaed, A. Panthakkan, S. A. Mansoori, H. A. Ahmad, and S. Marshall, "Airbus Ship Detection from Satellite Imagery Using Frequency Domain Learning," in *Image and Signal Processing for Remote Sensing XXVII*, L. Bruzzone and F. Bovolo, Eds., vol. 11862, International Society for Optics and Photonics. SPIE, 2021, pp. 279 – 285. [Online]. Available: https://doi.org/10.1117/12.2600168

[28] N. Aburaed, M. Al-Saad, M. Chendeb El Rai, S. Al Mansoori, H. Al-Ahmad, and S. Marshall, "Autonomous Object Detection in Satellite Images Using Wfrcnn," in *IEEE India Geoscience and Remote Sensing Symposium (InGARSS)*, 2020, pp. 106–109.

[29] N. Aburaed, A. Panthakkan, H. Mukhtar, W. Mansoor, S. Almansoori, and H. A. Ahmad, "Autonomous Building Detection Using Region Properties and PCA," in *International Conference on Signal Processing and Information Security (ICSPIS)*, 2018, pp. 1–4.

[30] Z. Zheng, L. Lei, H. Sun, and G. Kuang, "A Review of Remote Sensing Image Object Detection Algorithms Based on Deep Learning," in *IEEE 5th International Conference on Image, Vision and Computing (ICIVC)*, 2020, pp. 34–43.

[31] Q. Zhu, X. Sun, Y. Zhong, and L. Zhang, "High-Resolution Remote Sensing Image Scene Understanding: A Review," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2019, pp. 3061–3064.

[32] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep Learning for Hyperspectral Image Classification: An Overview," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 9, pp. 6690–6709, 2019.

[33] R. Neware and A. Khan, "Survey on Classification Techniques Used in Remote Sensing for Satellite Images," in *Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, 2018, pp. 1860–1863.

[34] M. Talal, A. Panthakkan, H. Mukhtar, W. Mansoor, S. Almansoori, and H. A. Ahmad, "Detection of Water-Bodies Using Semantic Segmentation," in *International Conference on Signal Processing and Information Security (ICSPIS)*, 2018, pp. 1–4.

[35] B. Cui, X. Chen, and Y. Lu, "Semantic Segmentation of Remote Sensing Images Using Transfer Learning and Deep Convolutional Neural Network With Dense Connection," *IEEE Access*, vol. 8, pp. 116 744–116 755, 2020.

[36] Z. Shi, C. Chen, Z. Xiong, D. Liu, and F. Wu, "HSCNN+: Advanced CNN-Based Hyperspectral Recovery from RGB Images," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018, pp. 1052–10 528.

[37] C. Kwan, B. Budavari, M. Dao, B. Ayhan, and J. F. Bell, "Pansharpening of Mastcam images," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2017, pp. 5117–5120.

[38] N. Aburaed, M. Q. Alkhatib, S. Marshall, J. Zabalza, and H. Al Ahmad, "A Review of Spatial Enhancement of Hyperspectral Remote Sensing Imaging Techniques," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 2275–2300, 2023.

[39] L. Loncan, L. B. de Almeida, J. M. Bioucas-Dias, X. Briottet, J. Chanussot, N. Dobigeon, S. Fabre, W. Liao, G. A. Licciardi, M. Simões, J.-Y. Tourneret, M. A. Veganzones, G. Vivone, Q. Wei, and N. Yokoya, "Hyperspectral Pansharpening: A Review," *IEEE Geoscience and Remote Sensing Magazine*, vol. 3, no. 3, pp. 27–46, 2015.

[40] C. Chen, Y. Wang, N. Zhang, Y. Zhang, and Z. Zhao, "A Review of Hyperspectral Image Super-Resolution Based on Deep Learning," *Remote Sensing*, vol. 15, no. 11, 2023. [Online]. Available: https://www.mdpi.com/2072-4292/15/11/2853

[41] M. B. Stuart, M. Davies, M. J. Hobbs, T. D. Pering, A. J. S. McGonigle, and J. R. Willmott, "High-Resolution Hyperspectral Imaging Using Low-Cost Components: Application within Environmental Monitoring Scenarios," *Sensors*, vol. 22, no. 12, 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/12/4652

[42] M. Q. Alkhatib, M. Al-Saad, N. Aburaed, S. Almansoori, J. Zabalza, S. Marshall, and H. Al-Ahmad, "Tri-CNN: A Three Branch Model for Hyperspectral Image Classification," *Remote Sensing*, vol. 15, no. 2, 2023. [Online]. Available: https://www.mdpi.com/2072-4292/15/2/316

[43] M. Q. Alkhatib, M. Al-Saad, N. Aburaed, S. A. Mansoori, and H. Al Ahmad, "Dimensionality Reduction Techniques with Hydranet Framework for HSI Classification," in *IEEE International Conference on Image Processing (ICIP)*, 2022, pp. 3151–3155.

[44] M. Al-Saad, N. Aburaed, S. A. Mansoori, and H. A. Ahmad, "Autonomous Palm Tree Detection from Remote Sensing Images - UAE Dataset," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2022, pp. 2191–2194.

[45] M. C. El Rai, N. Aburaed, M. Al-Saad, H. Al-Ahmad, S. Al Mansoori, and S. Marshall, "Integrating Deep Learning with Active Contour Models in Remote Sensing Image Segmentation," in *27th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, 2020, pp. 1–4.

[46] L. Wang, T. Bi, and Y. Shi, "A Frequency-Separated 3D-CNN for Hyperspectral Image Super-Resolution," *IEEE Access*, vol. 8, pp. 86 367–86 379, 2020.

[47] Y. LeCun and Y. Bengio, "Convolutional Networks for Images, Speech, and Time Series," *The Handbook of Brain Theory and Neural Networks*, vol. 3361, no. 10, 1995.

[48] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-Based Learning Applied to Document Recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[49] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going Deeper with Convolutions," *CoRR*, vol. abs/1409.4842, 2014. [Online]. Available: http://arxiv.org/abs/1409.4842

[50] N. Aburaed, M. Q. Alkhatib, S. Marshall, J. Zabalza, and H. Al Ahmad, "3D Expansion of SRCNN for Spatial Enhancement of Hyperspectral Remote Sensing Images," in *4th International Conference on Signal Processing and Information Security (ICSPIS)*, 2021, pp. 9–12.

[51] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[52] S. Mei, X. Yuan, J. Ji, Y. Zhang, S. Wan, and Q. Du, "Hyperspectral Image Spatial Super-Resolution via 3D Full Convolutional Neural Network," *Remote Sensing*, vol. 9, no. 11, 2017.

[53] J. Schmidhuber, "Deep Learning in Neural Networks: An Overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.

[54] K. He, X. Zhang, S. Ren *et al.*, "Deep Residual Learning for Image Recognition," *CoRR*, vol. abs/1512.03385, 2015. [Online]. Available: http://arxiv.org/abs/1512.03385

[55] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, "Attention is All you Need," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017. [Online]. Available: https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

[56] A. Chinea, "Understanding the Principles of Recursive Neural Networks: A Generative Approach to Tackle Model Complexity," in *Artificial Neural Networks – ICANN*, C. Alippi, M. Polycarpou, C. Panayiotou, and G. Ellinas, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 952–963.

[57] N. Gat, "Imaging Spectroscopy Using Tunable Filters: A Review," in *Wavelet Applications VII*, H. H. Szu, M. Vetterli, W. J. Campbell, and J. R. Buss, Eds., vol. 4056, International Society for Optics and Photonics. SPIE, 2000, pp. 50 – 64. [Online]. Available: https://doi.org/10.1117/12.381686

[58] D. Ren, *Lenses: Design*. CRC Press, 2015.

[59] A. Rogalski, "Infrared Detectors: Status and Trends," *Progress in Quantum Electronics*, vol. 27, no. 2, pp. 59–210, 2003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0079672702000241

[60] L. Sun, X. Mi, J. Wei, J. Wang, X. Tian, H. Yu, and P. Gan, "A Cloud Detection Algorithm-generating Method for Remote Sensing Data at Visible to Short-wave Infrared Wavelengths," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 124, pp. 70–88, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271616306189

[61] J. G. Proakis and D. K. Manolakis, *Digital Signal Processing (4th Edition)*. USA: Prentice-Hall, Inc., 2006.

[62] K. Tatsumi, N. Ohgi, H. Harada, T. Kawanishi, F. Sakuma, H. Inada, T. Kawashima, and A. Iwasaki, "Retrieval of Spectral Response Functions for the Hyperspectral Sensor of HISUI (Hyperspectral Imager SUIte) by Means of Onboard Calibration Sources," in *Sensors, Systems, and Next-Generation Satellites XV*, R. Meynart, S. P. Neeck, and H. Shimoda, Eds., vol. 8176, International Society for Optics and Photonics. SPIE, 2011, p. 81760S.

[63] H. E. Torkildsen and T. Skauli, "Measurement of Point Spread Function for Characterization of Coregistration and Resolution: Comparison of Two Commercial Hyperspectral Cameras," in *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XXIV*, M. Velez-Reyes and D. W. Messinger, Eds., vol. 10644, International Society for Optics and Photonics. SPIE, 2018, p. 106441F.

[64] D. Inamdar, M. Kalacska, P. O. Darko, J. P. Arroyo-Mora, and G. Leblanc, "Spatial Response Resampling (SR2): Accounting for the Spatial Point Spread Function in Hyperspectral Image Resampling," *MethodsX*, vol. 10, p. 101998, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2215016123000031

[65] Q. Li, X. He, Y. Wang, H. Liu, D. Xu, and F. Guo, "Review of Spectral Imaging Technology in Biomedical Engineering: Achievements and Challenges," *Journal of Biomedical Optics*, vol. 18, no. 10, p. 100901, 2013. [Online]. Available: https://doi.org/10.1117/1.JBO.18.10.100901

[66] T. Adão, J. Hruška, L. Pádua, J. Bessa, E. Peres, R. Morais, and J. J. Sousa, "Hyperspectral Imaging: A Review on UAV-Based Sensors, Data Processing and Applications for Agriculture and Forestry," *Remote Sensing*, vol. 9, no. 11, 2017. [Online]. Available: https://www.mdpi.com/2072-4292/9/11/1110

[67] D. N. Conran and E. J. Ientilucci, "A Vicarious Technique for Understanding and Diagnosing Hyperspectral Spatial Misregistration," *Sensors*, vol. 23, no. 9, 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/9/4333

[68] D. Coulter, P. Hauff, and W. Kerby, "Airborne Hyperspectral Remote Sensing," in *Proceedings of Exploration*, vol. 7, 2007, pp. 375–386.

[69] "Earth Observing - 1," NASA, Nov 2000. [Online]. Available: https://earthobservatory.nasa.gov/features/EO1/eo1_2.php

[70] K. Deilami and M. Hashim, "Very High Resolution Optical Satellites for DEM Generation: A Review," *European Journal of Scientific Research*, vol. 49, no. 4, pp. 542–554, 2011.

[71] G. Vivone, "Multispectral and Hyperspectral Image Fusion in Remote Sensing: A Survey," *Information Fusion*, vol. 89, pp. 405–417, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566253522001312

[72] N. Keshava, "A Survey of Spectral Unmixing Algorithms A Survey of Spectral Unmixing," pp. 55–78, 2003.

[73] R. Kawakami, Y. Matsushita, J. Wright, M. Ben-Ezra, Y. Tai, and K. Ikeuchi, "High-Resolution Hyperspectral Imaging Via Matrix Factorization," in *CVPR 2011*, 2011, pp. 2329–2336.

[74] R. Paschotta, "Radiometry," RP Photonics AG, 2022. [Online]. Available: https://www.rp-photonics.com/radiometry.html

[75] "Hyperspectral Data Correction," MathWorks, 2020. [Online]. Available: https://www.mathworks.com/help/images/hyperspectral-data-correction.html

[76] A. Knudby, "3. Calculations of TOA radiance and TOA reflectance," Pressbooks, 2015. [Online]. Available: https://ecampusontario.pressbooks.pub/remotesensing/chapter/chapter-3-calculations-of-toa-radiance-and-toa-reflectance/

[77] O. Ozdil, A. Gunes, Y. E. Esin, B. Demirel, and S. Ozturk, "Comparison of Target Detection Performance for Radiance and Reflectance Domain in VNIR Hyperspectral Images," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2019, pp. 2186–2189.

[78] R. Cui, H. Yu, T. Xu, X. Xing, X. Cao, K. Yan, and J. Chen, "Deep Learning in Medical Hyperspectral Images: A Review," *Sensors*, vol. 22, no. 24, p. 9790, 2022.

[79] S. Puustinen, H. Vrzáková, J. Hyttinen, T. Rauramaa, P. Fält, M. Hauta-Kasari, R. Bednarik, T. Koivisto, S. Rantala, M. von und zu Fraunberg, J. E. Jääskeläinen, and A.-P. Elomaa, "Hyperspectral Imaging in Brain Tumor Surgery—Evidence of Machine Learning-Based Performance," *World Neurosurgery*, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1878875023004734

[80] Y. H. El-Sharkawy, M. H. Aref, S. Elbasuney, S. M. Radwan, and G. S. El-Sayyad, "Oxygen Saturation Measurements Using Novel Diffused Reflectance with Hyperspectral Imaging: Towards Facile COVID-19 Diagnosis," *Opt Quantum Electron*, vol. 54, no. 5, p. 322, May 2022.

[81] L. Liu, M. Qi, Y. Li, Y. Liu, X. Liu, Z. Zhang, and J. Qu, "Staging of Skin Cancer Based on Hyperspectral Microscopic Imaging and Machine Learning," *Biosensors*, vol. 12, no. 10, p. 790, 2022.

[82] L. M. Dale, A. Thewis, C. Boudry, I. Rotar, P. Dardenne, V. Baeten, and J. A. F. Pierna, "Hyperspectral Imaging Applications in Agriculture and Agro-Food Product Quality and Safety Control: A Review," *Applied*

Spectroscopy Reviews, vol. 48, no. 2, pp. 142–159, 2013. [Online]. Available: https://doi.org/10.1080/05704928.2012.705800

[83] A. F. Cheshkova, "A Review of Hyperspectral Image Analysis Techniques for Plant Disease Detection and Identification," *Vavilovskii Zhurnal Genet Selektsii*, vol. 26, no. 2, pp. 202–213, Mar. 2022.

[84] M. A. Moharram and D. M. Sundaram, "Land Use and Land Cover Classification with Hyperspectral Data: A Comprehensive Review of Methods, Challenges and Future Directions," *Neurocomputing*, vol. 536, pp. 90–113, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231223002436

[85] G. Yang, U. B. Gewali, E. Ientilucci, M. Gartley, and S. T. Monteiro, "Dual-Channel Densenet for Hyperspectral Image Classification," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2018, pp. 2595–2598.

[86] M. B. Stuart, A. J. S. McGonigle, and J. R. Willmott, "Hyperspectral Imaging in Environmental Monitoring: A Review of Recent Developments and Technological Advances in Compact Field Deployable Systems," *Sensors*, vol. 19, no. 14, 2019. [Online]. Available: https://www.mdpi.com/1424-8220/19/14/3071

[87] G. Mateo-Garcia, J. Veitch-Michaelis, L. Smith, S. V. Oprea, G. Schumann, Y. Gal, A. G. Baydin, and D. Backes, "Towards Global Flood Mapping Onboard Low Cost Satellites with Machine Learning," *Scientific Reports*, vol. 11, no. 1, p. 7249, Mar 2021. [Online]. Available: https://doi.org/10.1038/s41598-021-86650-z

[88] K. Thangavel, D. Spiller, R. Sabatini, S. Amici, S. T. Sasidharan, H. Fayek, and P. Marzocca, "Autonomous Satellite Wildfire Detection Using Hyperspectral Imagery and Neural Networks: A Case Study on Australian Wildfire," *Remote Sensing*, vol. 15, no. 3, 2023. [Online]. Available: https://www.mdpi.com/2072-4292/15/3/720

[89] C. Ye, Y. Li, P. Cui, L. Liang, S. Pirasteh, J. Marcato, W. N. Gonçalves, and J. Li, "Landslide Detection of Hyperspectral Remote Sensing Data Based on Deep

Learning With Constrains," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 12, pp. 5047–5060, 2019.

[90] M. Huebschman, R. Schultz, and H. Garner, "IMAGING — Hyperspectral Imaging," in *Encyclopedia of Modern Optics*, R. D. Guenther, Ed. Oxford: Elsevier, 2005, pp. 134–143. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B0123693950007041

[91] "Spatial Analysis." [Online]. Available: https:/cimss.ssec.wisc.edu/sage/remote_sensing/lesson3/concepts.html

[92] Z. Zhang and J. C. Moore, "Chapter 4 - Remote Sensing," in *Mathematical and Physical Fundamentals of Climate Change*, Z. Zhang and J. C. Moore, Eds. Boston: Elsevier, 2015, pp. 111–124. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128000663000048

[93] "Spectral Resolution," Natural Resources Canada, 2015. [Online]. Available: https://natural-resources.canada.ca/maps-tools-and-publications/satellite-imagery-and-air-photos/tutorial-fundamentals-remote-sensing/satellites-and-sensors/spectral-resolution/9393

[94] J. Théau, *Temporal Resolution*. Boston, MA: Springer US, 2008, pp. 1150–1151. [Online]. Available: https://doi.org/10.1007/978-0-387-35973-1_1376

[95] V. S. Frost, "Probability of Error And Radiometric Resolution for Target Discrimination in Radar Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. GE-22, no. 2, pp. 121–125, 1984.

[96] C. González, S. Sánchez, A. Paz, J. Resano, D. Mozos, and A. Plaza, "Use of FPGA or GPU-Based Architectures for Remotely Sensed Hyperspectral Image Processing," *Integration*, vol. 46, no. 2, pp. 89–103, 2013.

[97] J. M. Smith, "Meet EMIT, the Newest Imaging Spectrometer," NASA, May 2023. [Online]. Avail-

able: https://www.earthdata.nasa.gov/learn/articles/meet-emit#:~:text=With%20a%20viewing%20swath%20of,of%20dust%20in%20the%20atmosphere.

[98] M. Graña, M. Veganzons, and B. Ayerdi, "Hyperspectral Remote Sensing Scenes," Grupo de Inteligencia Computacional (GIC). [Online]. Available: http://www.ehu.eus//ccwintco//index.php//Hyperspectral_Remote_Sensing_Scenes

[99] M. Habermann, V. Fremont, and E. H. Shiguemori, "Problem-Based Band Selection for Hyperspectral Images," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2017, pp. 1800–1803.

[100] L. Jiao, R. Shang, F. Liu, and W. Zhang, "Chapter 22 - Deep Neural Network Models for Hyperspectral Images," in *Brain and Nature-Inspired Learning Computation and Recognition*, L. Jiao, R. Shang, F. Liu, and W. Zhang, Eds. Elsevier, 2020, pp. 735–747.

[101] P. Ma, J. Ren, H. Zhao, G. Sun, P. Murray, and J. Zheng, "Multiscale 2-D Singular Spectrum Analysis and Principal Component Analysis for Spatial–Spectral Noise-Robust Feature Extraction and Classification of Hyperspectral Images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 1233–1245, 2021.

[102] X. Yang, X. Zhang, Y. Ye, R. Y. K. Lau, S. Lu, X. Li, and X. Huang, "Synergistic 2D/3D Convolutional Neural Network for Hyperspectral Image Classification," *Remote Sensing*, vol. 12, no. 12, 2020. [Online]. Available: https://www.mdpi.com/2072-4292/12/12/2033

[103] "Remote Sensing Datasets," Remote Sensing Laboratory - School of Surveying and Geospatial Engineering, 2015. [Online]. Available: https://rslab.ut.ac.ir/data

[104] R. M. H. Nguyen, D. K. Prasad, and M. S. Brown, "Training-Based Spectral Reconstruction from a Single RGB Image," in *Computer Vision – ECCV*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham: Springer International Publishing, 2014, pp. 186–201.

[105] "2013 IEEE GRSS Data Fusion Contest – Fusion of Hyperspectral and LiDAR Data," Hyperspectral Image Analysis Lab. [Online]. Available: https://hyperspectral.ee.uh.edu/?page_id=459

[106] "AVIRIS - Airborne Visible / Infrared Imaging Spectrometer - Data," NASA. [Online]. Available: https://aviris.jpl.nasa.gov/data/image_cube.html

[107] "HySure," GitHub. [Online]. Available: https://github.com/alfaiate/HySure

[108] M. A. Veganzones, M. Simões, G. Licciardi, N. Yokoya, J. M. Bioucas-Dias, and J. Chanussot, "Hyperspectral Super-Resolution of Locally Low Rank Images From Complementary Multisource Data," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 274–288, 2016.

[109] R. Clark and A. Swayze, "Evolution in Imaging Spectroscopy Analysis and Sensor Signal-to-Noise: An Examination of How Far We Have Come," in *Summaries of the 6th Annual JPL Airborne Earth Science Workshop*, 1996, pp. 49–53.

[110] F. Zhu, Y. Wang, B. Fan, G. Meng, and C. Pan, "Effective Spectral Unmixing via Robust Representation and Learning-Based Sparsity," *CoRR*, vol. abs/1409.0685, 2014. [Online]. Available: http://arxiv.org/abs/1409.0685

[111] O. Ben-Ahmed, T. Urruty, N. Richard, and C. Fernandez-Maloigne, "Toward Content-Based Hyperspectral Remote Sensing Image Retrieval (CB-HRSIR): A Preliminary Study Based on Spectral Sensitivity Functions," *Remote Sensing*, vol. 11, no. 5, 2019. [Online]. Available: https://www.mdpi.com/2072-4292/11/5/600

[112] F. Yasuma, T. Mitsunaga, D. Iso, and S. Nayar, "Generalized Assorted Pixel Camera: Post-Capture Control of Resolution, Dynamic Range and Spectrum," Technical Report, Department of Computer Science, Columbia University CUCS-061-08, Tech. Rep., Nov 2008.

[113] A. Chakrabarti and T. Zickler, "Statistics of Real-World Hyperspectral Images," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 193–200.

[114] F. Yasuma, T. Mitsunaga, D. Iso, and S. Nayar, "Generalized Assorted Pixel Camera: Post-Capture Control of Resolution, Dynamic Range and Spectrum," in *Technical Report, Department of Computer Science, Columbia University CUCS-061-08*, Nov 2008.

[115] "Explore CHRIS," The European Space Agnency. [Online]. Available: https://earth.esa.int/eogateway/instruments/chris

[116] S. Michel, P. Gamet, and M.-J. Lefevre-Fonollosa, "HYPXIM — A Hyperspectral Satellite Defined for Science, Security and Defence Users," in *2011 3rd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2011, pp. 1–4.

[117] "CASI-1500 VNIR (Hyperspectral) Imagery," Brock University. [Online]. Available: https://brocku.ca/library/casi-1500-vnir-hyperspectral-imagery/

[118] M. K. Jakubowski, D. Pogorzala, T. J. Hattenberger, S. D. Brown, and J. R. Schott, "Synthetic Data Generation of High-Resolution Hyperspectral Data Using DIRSIG," in *Imaging Spectrometry XII*, S. S. Shen and P. E. Lewis, Eds., vol. 6661, International Society for Optics and Photonics. SPIE, 2007, pp. 153–163.

[119] "ASTER Satellite Sensor," Satellite Imaging Corporation. [Online]. Available: https://www.satimagingcorp.com/satellite-sensors/other-satellite-sensors/aster/

[120] G. Dong, Y. Zhang, and C. Zhu, "Spectrum Reconstruction and Spectrum Matching of PHI Hyperspectral Image in City Area," in *Remote Sensing of the Environment: 15th National Symposium on Remote Sensing of China*, Q. Tong, W. Gao, and H. Guo, Eds., vol. 6200, International Society for Optics and Photonics. SPIE, 2006, pp. 242–247.

[121] "ISS Utilization: HISUI (Hyperspectral Imager Suite) - Satellite Missions - eoPortal Directory." [Online]. Available: https://eoportal.org/web/eoportal/satellite-missions/content/-/article/iss-utilization-hisui-hyperspectral-imager-suite-

[122] "AisaDUAL: Combining the aisaEAGLE and aisaHAWK Sensors in One System for Simultaneous Acquisition of VNIR and SWIR Data," AisaDUAL. [Online]. Available: https://www.adept.net.au/cameras/specim/systems/Aisa_dual.shtml

[123] A. Fernández-Renau, J. A. Gómez, and E. de Miguel, "The INTA AHS System," in *Sensors, Systems, and Next-Generation Satellites IX*, ser. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, R. Meynart, S. P. Neeck, and H. Shimoda, Eds., vol. 5978, Oct. 2005, pp. 471–478.

[124] "APEX - Airborne Sensors - eoPortal Directory." [Online]. Available: https://earth.esa.int/web/eoportal/airborne-sensors/apex

[125] HySpex. [Online]. Available: https://www.hyspex.com/?gclid=Cj0KCQjwiNSLBhCPARIsAKNS4_fn6Wtlmq45lSQ2dub-7RCPwzgy-d5pjWuScTlmDmxSX4g40O94XWgaAoyZEALw_wcB

[126] J. M. Meyer, "HySpex by NEO VNIR-1800 and SWIR-384 Imaging Spectrometer Radiance and Reflectance data, with Associated ASD FieldSpec®NG Calibration Data, Collected at Cripple Creek Victor Mine, Cripple Creek, Colorado, 2017," U.S. Geological Survey, Dec 2022. [Online]. Available: https://www.sciencebase.gov/catalog/item/61c204e1d34e2ca389d9e6f1

[127] "Data Products," NASA JPL, 2022. [Online]. Available: https://earth.jpl.nasa.gov/emit/data/data-products/

[128] "PRISMA - Hyperspectral Satellite, Capable of Observing from the Optical to the Near Infrared," Italian Space Agency, 2019. [Online]. Available: https://www.asi.it/en/earth-science/prisma/

[129] "DESIS - Hyperspectral Images - Global," German Aerospace Center (DLR), 2019. [Online]. Available: https://doi.org/10.15489/hxom21uqeo90

[130] J. A. Herweg, J. P. Kerekes, O. Weatherbee, D. Messinger, J. van Aardt, E. Ientilucci, Z. Ninkov, J. Faulring, N. Raqueño, and J. Meola, "SpecTIR Hyperspectral Airborne Rochester Experiment Data Collection Campaign," in *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVIII*, S. S. Shen and P. E. Lewis, Eds., vol. 8390, International Society for Optics and Photonics. SPIE, 2012, p. 839028.

[131] N. Yokoya and A. Iwasaki, "Airborne Hyperspectral Data Over Chikusei," Space Application Laboratory, University of Tokyo, Japan, Tech. Rep. SAL-2016-05-27, May 2016.

[132] "Hyperspectral Images," MultiSpec© — Tutorials. [Online]. Available: https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html

[133] D. Salomon, *Data compression: The complete reference 4th ed.* Springer, 2007.

[134] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image Quality Assessment: from Error Visibility to Structural Similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[135] J. Nilsson and T. Akenine-Möller, "Understanding SSIM," *CoRR*, vol. abs/2006.13846, 2020. [Online]. Available: https://arxiv.org/abs/2006.13846

[136] M. Treisman, "Noise and Weber's Law: The Discrimination of Brightness and Other Dimensions," *Psychological review*, vol. 71, no. 4, p. 314, 1964.

[137] "The Role of Contrast in Ability of Human Vision," ASTRO-GEO-GIS, 2020. [Online]. Available: https://astro-geo-gis.com/the-role-of-contrast-in-ability-of-human-vision/

[138] G. E. Legge and J. M. Foley, "Contrast Masking in Human Vision," *J. Opt. Soc. Am.*, vol. 70, no. 12, pp. 1458–1471, Dec 1980. [Online]. Available: https://opg.optica.org/abstract.cfm?URI=josa-70-12-1458

[139] Z. Wang and A. Bovik, "A Universal Image Quality Index," *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81–84, 2002.

[140] J. Li, Y. Peng, T. Jiang, L. Zhang, and J. Long, "Hyperspectral Image Super-Resolution Based on Spatial Group Sparsity Regularization Unmixing," *Applied Sciences*, vol. 10, no. 16, 2020. [Online]. Available: https://www.mdpi.com/2076-3417/10/16/5583

[141] R. Zhu, F. Zhou, and J.-H. Xue, "MvSSIM: A Quality Assessment Index for Hyperspectral Images," *Neurocomputing*, vol. 272, pp. 250–257, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231217312018

[142] Z. Liu, W. Wang, Q. Ma, X. Liu, and J. Jiang, "Rethinking 3D-CNN in Hyperspectral Image Super-Resolution," *Remote Sensing*, vol. 15, no. 10, 2023. [Online]. Available: https://www.mdpi.com/2072-4292/15/10/2574

[143] Q. Ma, J. Jiang, X. Liu, and J. Ma, "Learning a 3D-CNN and Transformer Prior for Hyperspectral Image Super-Resolution," *Information Fusion*, vol. 100, p. 101907, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566253523002233

[144] F. Kruse, A. Lefkoff, J. Boardman, K. Heidebrecht, A. Shapiro, P. Barloon, and A. Goetz, "The Spectral Image Processing System (SIPS)—iInteractive Visualization and Analysis of Imaging Spectrometer Data," *Remote Sensing of Environment*, vol. 44, no. 2, pp. 145–163, 1993, airbone Imaging Spectrometry.

[145] Y. Yuan, X. Zheng, and X. Lu, "Hyperspectral Image Superresolution by Transfer Learning," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 5, pp. 1963–1974, 2017.

[146] S. Mei, X. Yuan, J. Ji, S. Wan, J. Hou, and Q. Du, "Hyperspectral Image Super-Resolution Via Convolutional Neural Network," in *IEEE International Conference on Image Processing (ICIP)*, 2017, pp. 4297–4301.

[147] W. Liu and J. Lee, "An Efficient Residual Learning Neural Network for Hyperspectral Image Superresolution," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 4, pp. 1240–1253, 2019.

[148] X. Dou, C. Li, Q. Shi, and M. Liu, "Super-Resolution for Hyperspectral Remote Sensing Images Based on the 3D Attention-SRGAN Network," *Remote Sensing*, vol. 12, no. 7, 2020. [Online]. Available: https://www.mdpi.com/2072-4292/12/7/1204

[149] Y. Li, L. Zhang, C. Dingl, W. Wei, and Y. Zhang, "Single Hyperspectral Image Super-Resolution with Grouped Deep Recursive Residual Network," in *IEEE Fourth International Conference on Multimedia Big Data (BigMM)*, 2018, pp. 1–4.

[150] J. Li, R. Cui, B. Li, Y. Li, S. Mei, and Q. Du, "Dual 1D-2D Spatial-Spectral CNN for Hyperspectral Image Super-Resolution," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2019, pp. 3113–3116.

[151] Q. Li, Q. Wang, and X. Li, "Mixed 2D/3D Convolutional Network for Hyperspectral Image Super-Resolution," *Remote Sensing*, vol. 12, no. 10, 2020.

[152] Q. Wang, Q. Li, and X. Li, "Hyperspectral Image Super-Resolution Using Spectrum and Feature Context," *IEEE Transactions on Industrial Electronics*, pp. 1–1, 2020.

[153] J. Jiang, H. Sun, X. Liu, and J. Ma, "Learning Spatial-Spectral Prior for Super-Resolution of Hyperspectral Imagery," *IEEE Transactions on Computational Imaging*, vol. 6, p. 1082–1096, 2020. [Online]. Available: http://dx.doi.org/10.1109/TCI.2020.2996075

[154] Z. Tang, Q. Xu, P. Wu, Z. Shi, and B. Pan, "Feedback Refined Local-Global Network for Super-Resolution of Hyperspectral Imagery," *Remote Sensing*, vol. 14, no. 8, 2022. [Online]. Available: https://www.mdpi.com/2072-4292/14/8/1944

[155] X. Wang, J. Ma, J. Jiang, and X.-P. Zhang, "Dilated Projection Correction Network Based on Autoencoder for Hyperspectral Image Super-Resolution," *Neural Netw.*, vol. 146, no. C, p. 107–119, feb 2022. [Online]. Available: https://doi.org/10.1016/j.neunet.2021.11.014

[156] M. Zhao, J. Ning, J. Hu, and T. Li, "Hyperspectral Image Super-Resolution under the Guidance of Deep Gradient Information," *Remote Sensing*, vol. 13, no. 12, 2021. [Online]. Available: https://www.mdpi.com/2072-4292/13/12/2382

[157] D. Liu, J. Li, and Q. Yuan, "A Spectral Grouping and Attention-Driven Residual Dense Network for Hyperspectral Image Super-Resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 9, pp. 7711–7725, 2021.

[158] Q. Li, Q. Wang, and X. Li, "Exploring the Relationship Between 2D/3D Convolution for Hyperspectral Image Super-Resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 10, pp. 8693–8703, 2021.

[159] X. Wang, J. Ma, and J. Jiang, "Hyperspectral Image Super-Resolution via Recurrent Feedback Embedding and Spatial-Spectral Consistency Regularization," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–13, 2021.

[160] O. Sidorov and J. Y. Hardeberg, "Deep Hyperspectral Prior: Single-Image Denoising, Inpainting, Super-Resolution," in *IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, 2019, pp. 3844–3851.

[161] J. Hauser, G. Shtendel, A. Zeligman, A. Averbuch, and M. Nathan, "SHS-GAN: Synthetic Enhancement of a Natural Hyperspectral Database," *IEEE Transactions on Computational Imaging*, vol. 7, pp. 505–517, 2021.

[162] A. Nikonorov, M. Petrov, S. Bibikov, V. Kutikova, P. Yakimov, A. Morozov, R. Skidanov, and N. Kazanskiy, "Deep Learning-Based Enhancement of Hyperspectral Images Using Simulated Ground Truth," in *10th IAPR Workshop on Pattern Recognition in Remote Sensing (PRRS)*, 2018, pp. 1–9.

[163] J. Hu, Y. Li, X. Zhao, and W. Xie, "A Spatial Constraint and Deep Learning Based Hyperspectral Image Super-Resolution Method," in *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2017, pp. 5129–5132.

[164] S. Sharma, S. Sharma, and K. M. Buddhiraju, "Ant Colony Optimization for Super-Resolution of Hyperspectral Images," in *2016 8th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2016, pp. 1–5.

[165] J. Hu, Y. Tang, and S. Fan, "Hyperspectral Image Super Resolution Based on Multiscale Feature Fusion and Aggregation Network With 3-D Convolution," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 5180–5193, 2020.

[166] X. Li, L. Zhang, and J. You, "Domain Transfer Learning for Hyperspectral Image Super-Resolution," *Remote Sensing*, vol. 11, no. 6, 2019.

[167] Y. Fu, Z. Liang, and S. You, "Bidirectional 3D Quasi-Recurrent Neural Network for Hyperspectral Image Super-Resolution," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 2674–2688, 2021.

[168] J. Li, R. Cui, B. Li, R. Song, Y. Li, Y. Dai, and Q. Du, "Hyperspectral Image Super-Resolution by Band Attention Through Adversarial Learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 6, pp. 4304–4318, 2020.

[169] J. Hu, X. Jia, Y. Li, G. He, and M. Zhao, "Hyperspectral Image Super-Resolution via Intrafusion Network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 10, pp. 7459–7471, 2020.

[170] P. V. Arun, K. M. Buddhiraju, A. Porwal, and J. Chanussot, "CNN-Based Super-Resolution of Hyperspectral Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 9, pp. 6106–6121, 2020.

[171] Q. Wang, Q. Li, and X. Li, "Spatial-Spectral Residual Network for Hyperspectral Image Super-Resolution," *CoRR*, vol. abs/2001.04609, 2020. [Online]. Available: https://arxiv.org/abs/2001.04609

[172] W. Xie, X. Jia, Y. Li, and J. Lei, "Hyperspectral Image Super-Resolution Using Deep Feature Matrix Factorization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 8, pp. 6055–6067, 2019.

[173] K. Zheng, L. Gao, Q. Ran, X. Cui, B. Zhang, W. Liao, and S. Jia, "Separable-Spectral Convolution and Inception Network for Hyperspectral Image super-Resolution," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 10, p. 2593–2607, 2019.

[174] J. Jia, L. Ji, Y. Zhao, and X. Geng, "Hyperspectral Image Super-Resolution with Spectral–Spatial Network," *International Journal of Remote Sensing*, vol. 39, no. 22, pp. 7806–7829, 2018.

[175] C. Wang, Y. Liu, X. Bai, W. Tang, P. Lei, and J. Zhou, "Deep Residual Convolutional Neural Network for Hyperspectral Image Super-Resolution," in *Image and Graphics*, Y. Zhao, X. Kong, and D. Taubman, Eds. Cham: Springer International Publishing, 2017, pp. 370–380.

[176] K. Li, D. Dai, E. Konukoglu, and L. V. Gool, "Hyperspectral Image Super-Resolution with Spectral Mixup and Heterogeneous Datasets," *CoRR*, vol. abs/2101.07589, 2021. [Online]. Available: https://arxiv.org/abs/2101.07589

[177] K. Li, D. Dai, and L. Van Gool, "Hyperspectral Image Super-Resolution with RGB Image Super-Resolution as an Auxiliary Task," in *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2022, pp. 4039–4048.

[178] D. Agrafiotis, "Chapter 9 - Video Error Concealment," in *Academic Press Library in signal Processing*, ser. Academic Press Library in Signal Processing, S. Theodoridis and R. Chellappa, Eds. Elsevier, 2014, vol. 5, pp. 295–321. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780124201491000090

[179] H. Kim, S. Park, J. Wang, Y. Kim, and J. Jeong, "Advanced Bilinear Image Interpolation Based on Edge Features," in *First International Conference on Advances in Multimedia*, 2009, pp. 33–36.

[180] A. Gilman, D. G. Bailey, and S. R. Marsland, "Interpolation Models for Image Super-Resolution," in *4th IEEE International Symposium on Electronic Design, Test and Applications (delta 2008)*, 2008, pp. 55–60.

[181] A. N. A. Rahim, S. N. Yaakob, R. Ngadiran, and M. W. Nasruddin, "An Analysis of Interpolation Methods for Super Resolution Images," in *IEEE Student Conference on Research and Development (SCOReD)*, 2015, pp. 72–77.

[182] A. Villa, J. Chanussot, J. A. Benediktsson, M. Ulfarsson, and C. Jutten, "Super-Resolution: An Efficient Method to Improve Spatial Resolution of Hyperspectral Images," in *IEEE International Geoscience and Remote Sensing Symposium*, 2010, pp. 2003–2006.

[183] H. Irmak, G. B. Akar, S. E. Yuksel, and H. Aytaylan, "Super-Resolution Reconstruction of Hyperspectral Images via an Improved MAP-Based Approach," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2016, pp. 7244–7247.

[184] H. Irmak, G. B. Akar, and S. E. Yuksel, "A MAP-Based Approach for Hyperspectral Imagery Super-Resolution," *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 2942–2951, 2018.

[185] H. Zeng, X. Xie, H. Cui, H. Yin, and J. Ning, "Hyperspectral Image Restoration via Global L1-2 Spatial–Spectral Total Variation Regularized Local Low-Rank Tensor Recovery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 4, p. 3309–3325, Apr 2021. [Online]. Available: http://dx.doi.org/10.1109/TGRS.2020.3007945

[186] S. He, H. Zhou, Y. Wang, W. Cao, and Z. Han, "Super-Resolution Reconstruction of Hyperspectral Images via Low Rank Tensor Modeling and Total Variation Reg-

ularization," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2016, pp. 6962–6965.

[187] T. Akgun, Y. Altunbasak, and R. Mersereau, "Super-Resolution Reconstruction of Hyperspectral Images," *IEEE Transactions on Image Processing*, vol. 14, no. 11, pp. 1860–1875, 2005.

[188] C. Fan, C. Wu, G. Li, and J. Ma, "Projections onto Convex Sets Super-Resolution Reconstruction Based on Point Spread Function Estimation of Low-Resolution Remote Sensing Images," *Sensors*, vol. 17, no. 2, 2017. [Online]. Available: https://www.mdpi.com/1424-8220/17/2/362

[189] P. M. Atkinson, "Mapping Sub-Pixel Boundaries from Remotely Sensed Images," in *Innovations in GIS 4*. Taylor & Francis, 1997, pp. 166–180.

[190] H. Huang, J. Yu, and W. Sun, "Super-Resolution Mapping Via Multi-Dictionary Based Sparse Representation," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 3523–3527.

[191] C. Kwan, J. H. Choi, S. Chan, J. Zhou, and B. Budavari, "Resolution Enhancement for Hyperspectral Images: A Super-Resolution and Fusion Approach," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 6180–6184.

[192] A. Dixit and S. Agarwal, "Super-Resolution Mapping of Hyperspectral Data Using Artificial Neural Network and Wavelet," *Remote Sensing Applications: Society and Environment*, vol. 20, p. 100374, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352938520303189

[193] H. G. Cyril Amala Dhason, I. Muthaia, S. P. Sakthivel, and S. Shanmugam, "Super-Resolution Mapping of Hyperspectral Satellite Images Using Hybrid Genetic Algorithm," *IET Image Processing*, vol. 14, no. 7, pp. 1281–1290, 2020. [Online]. Available: https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-ipr.2018.5108

[194] A. R. Gillespie, A. B. Kahle, and R. E. Walker, "Color Enhancement of Highly Correlated Images. II. Channel Ratio and "Chromaticity"' Transformation Techniques," *Remote Sensing of Environment*, vol. 22, no. 3, pp. 343–365, 1987. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0034425787900885

[195] C. A. Laben and B. V. Brower, "Process for Enhancing the Spatial Resolution of Multispectral Imagery Using Pan-Sharpening," Patent U.S. Patent 6 011 875, 2000.

[196] B. Aiazzi, S. Baronti, and M. Selva, "Improving Component Substitution Pansharpening Through Multivariate Regression of MS +Pan Data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 10, pp. 3230–3239, 2007.

[197] T.-M. Tu, S.-C. Su, H.-C. Shyu, and P. S. Huang, "A New Look at IHS-like Image Fusion Methods," *Information Fusion*, vol. 2, no. 3, pp. 177–186, 2001. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566253501000367

[198] P. C. Jr., S. Sides, and J. Anderson, "Comparison of Three Different Methods to Merge Multiresolution and Multispectral Data: Landsat TM and SPOT Panchromatic," *Photogrammetric Engineering and Remote Sensing*, vol. 57, no. 3, pp. 295–303, 1991.

[199] L. Capobianco, A. Garzelli, F. Nencini, L. Alparone, and S. Baronti, "Spatial Enhancement of Hyperion Hyperspectral Data Through ALI Panchromatic Image," in *2007 IEEE International Geoscience and Remote Sensing Symposium*, 2007, pp. 5158–5161.

[200] R. Restaino, G. Vivone, P. Addesso, and J. Chanussot, "Hyperspectral Sharpening Approaches Using Satellite Multiplatform Data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 1, pp. 578–596, 2021.

[201] J. Choi, K. Yu, and Y. Kim, "A New Adaptive Component-Substitution-Based Satellite Image Fusion by Using Partial Replacement," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 1, pp. 295–309, 2011.

[202] J. G. Liu, "Smoothing Filter-Based Intensity Modulation: A Spectral Preserve Image Fusion Technique for Improving Spatial Details," *International Journal of Remote Sensing*, vol. 21, no. 18, pp. 3461–3472, 2000. [Online]. Available: https://doi.org/10.1080/014311600750037499

[203] F. Murtagh, J. luc Starck, and A. Bijaoui, "Image Restoration with Noise Suppression Using a Multiresolution Support," *Astronomy and Astrophysics, Suppl. Ser*, vol. 112, pp. 179–189, 1995.

[204] Q. Wang and Y. Shen, "Effects of Wavelets Selection on Performances of Hyperspectral Image Fusion," in *IEEE Instrumentation and Measurement Technology Conference Proceedings*, 2006, pp. 812–815.

[205] M. M. Khan, J. Chanussot, L. Condat, and A. Montanvert, "Indusion: Fusion of Multispectral and Panchromatic Images Using the Induction Scaling Technique," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 1, pp. 98–102, 2008.

[206] G. Vivone, R. Restaino, M. Dalla Mura, G. Licciardi, and J. Chanussot, "Contrast and Error-Based Fusion Schemes for Multispectral Image Pansharpening," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 5, pp. 930–934, 2014.

[207] T. Ranchin and L. Wald, "Fusion of High Spatial and Spectral Resolution Images: The Arsis Concept and Its Implementation," *Photogrammetric Engineering and Remote Sensing*, vol. 66, no. 1, pp. 49–61, 2000.

[208] X. Otazu, M. Gonzalez-Audicana, O. Fors, and J. Nunez, "Introduction of Sensor Spectral Response into Image Fusion Methods. Application to wavelet-Based Methods," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 10, pp. 2376–2385, 2005.

[209] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "MTF-Tailored Multiscale Fusion of High-Resolution MS and Pan Imagery," *Photogrammetric Engineering and Remote Sensing*, vol. 72, pp. 591–596, 2006.

[210] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, and L. M. Bruce, "Comparison of Pansharpening Algorithms: Outcome of the 2006 GRS-S Data-Fusion Contest," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 10, pp. 3012–3021, 2007.

[211] J. Lee and C. Lee, "Fast and Efficient Panchromatic Sharpening," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 1, pp. 155–163, 2010.

[212] C. Ballester, V. Caselles, L. Igual, J. Verdera, and B. Rouge, "A Variational Model for P+XS Image Fusion," *Int. J. Comput Vision*, vol. 69, p. 43–58, 2006.

[213] J. Duran, A. Buades, B. Coll, and C. Sbert, "A Nonlocal Variational Model for Pansharpening Image Fusion," *SIAM Journal on Imaging Sciences*, vol. 7, no. 2, p. 761–796, 2014.

[214] J. Duran, A. Buades, B. Coll, C. Sbert, and G. Blanchet, "A Survey of Pansharpening Methods with a New Band-Decoupled Variational Model," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 125, pp. 78–105, 2017.

[215] Z. Huang, Q. Chen, Q. Chen, and X. Liu, "Variational Pansharpening for Hyperspectral Imagery Constrained by Spectral Shape and Gram–Schmidt Transformation," *Sensors*, vol. 18, no. 12, 2018. [Online]. Available: https://www.mdpi.com/1424-8220/18/12/4330

[216] G. Yigitler, O. Ersoy, and T. Ibrekei, "New Transform Techniques for Enhancement and Fusion of Multispectral and Hyperspectral Images," in *ICSC Congress on Computational Intelligence Methods and Applications*, 2005, pp. 6 pp.–.

[217] M. Eismann and R. Hardie, "Resolution enhancement of hyperspectral imagery using coincident panchromatic imagery and a stochastic mixing model," in *IEEE*

Workshop on Advances in Techniques for Analysis of Remotely Sensed Data, 2003, 2003, pp. 282–289.

[218] M. Winter and E. Winter, "Resolution Enhancement of Hyperspectral Data," in *Proceedings, IEEE Aerospace Conference*, vol. 3, 2002, pp. 3–3.

[219] W. Liao, X. Huang, F. Van Coillie, S. Gautama, A. Pižurica, W. Philips, H. Liu, T. Zhu, M. Shimoni, G. Moser, and D. Tuia, "Processing of Multiresolution Thermal Hyperspectral and Digital Color Data: Outcome of the 2014 IEEE GRSS Data Fusion Contest," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 2984–2996, 2015.

[220] W. Liao, X. Huang, F. Van Coillie, G. Thoonen, A. Pižurica, P. Scheunders, and W. Philips, "Two-stage fusion of Thermal Hyperspectral and Visible RGB Image by PCA and Guided Filter," in *2015 7th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2015, pp. 1–4.

[221] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled Non-Negative Matrix Factorization (CNMF) for Hyperspectral and Multispectral Data Fusion: Application to Pasture Classification," in *2011 IEEE International Geoscience and Remote Sensing Symposium*, 2011, pp. 1779–1782.

[222] H. Song, B. Huang, K. Zhang, and H. Zhang, "Spatio-Spectral Fusion of Satellite Images Based on Dictionary-Pair Learning," *Information Fusion*, vol. 18, pp. 148–160, 2014.

[223] A. S. Charles, B. A. Olshausen, and C. J. Rozell, "Learning Sparse Codes for Hyperspectral Imagery," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 5, pp. 963–978, 2011.

[224] B. Huang, H. Song, H. Cui, J. Peng, and Z. Xu, "Spatial and Spectral Image Fusion Using Sparse Matrix Factorization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 3, pp. 1693–1704, 2014.

[225] N. Yokoya and A. Iwasaki, "Effect of Unmixing-Based Hyperspectral Super-Resolution on Target Detection," in *2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2014, pp. 1–4.

[226] G. Licciardi, M. A. Veganzones, M. Simões, J. Bioucas, and J. Chanussot, "Super-Resolution of Hyperspectral Images Using Local Spectral Unmixing," in *2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2014, pp. 1–4.

[227] H. Kwon and Y. Tai, "RGB-Guided Hyperspectral Image Upsampling," in *IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 307–315.

[228] E. Wycoff, T. Chan, K. Jia, W. Ma, and Y. Ma, "A Non-Negative Sparse Promoting Algorithm for High Resolution Hyperspectral Imaging," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013, pp. 1409–1413.

[229] Y. Zhang, S. De Backer, and P. Scheunders, "Bayesian Fusion of Multispectral and Hyperspectral Image in Wavelet Domain," in *IGARSS 2008 - 2008 IEEE International Geoscience and Remote Sensing Symposium*, vol. 5, 2008, pp. V – 69–V – 72.

[230] M. Simões, J. Bioucas-Dias, L. B. Almeida, and J. Chanussot, "Hyperspectral Image Superresolution: An Edge-Preserving Convex Formulation," in *IEEE International Conference on Image Processing (ICIP)*, 2014, pp. 4166–4170.

[231] M. Simões, J. Bioucas-Dias, L. B. Almeida, and J. Chanussot, "A Convex Formulation for Hyperspectral Image Superresolution via Subspace-Based Regularization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 6, pp. 3373–3388, 2015.

[232] R. Hardie, M. Eismann, and G. Wilson, "MAP Estimation for Hyperspectral Image Resolution Enhancement Using an Auxiliary Sensor," *IEEE Transactions on Image Processing*, vol. 13, no. 9, pp. 1174–1184, 2004.

[233] M. Eismann and R. Hardie, "Hyperspectral Resolution Enhancement Using High-Resolution Multispectral Imagery with Arbitrary Response Functions," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 455–465, 2005.

[234] N. Akhtar, F. Shafait, and A. Mian, "Sparse Spatio-spectral Representation for Hyperspectral Image Super-Resolution," in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham: Springer International Publishing, 2014, pp. 63–78.

[235] W. Dong, F. Fu, G. Shi, X. Cao, J. Wu, G. Li, and X. Li, "Hyperspectral Image Super-Resolution via Non-Negative Structured Sparse Representation," *IEEE Transactions on Image Processing*, vol. 25, no. 5, pp. 2337–2352, 2016.

[236] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, p. 1–122, jan 2011. [Online]. Available: https://doi.org/10.1561/2200000016

[237] J. Xue, Y.-Q. Zhao, Y. Bu, W. Liao, J. C.-W. Chan, and W. Philips, "Spatial-Spectral Structured Sparse Low-Rank Representation for Hyperspectral Image Super-Resolution," *IEEE Transactions on Image Processing*, vol. 30, pp. 3084–3097, 2021.

[238] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei, "Nonlocal Patch Tensor Sparse Representation for Hyperspectral Image Super-Resolution," *IEEE Transactions on Image Processing*, vol. 28, no. 6, pp. 3034–3047, 2019.

[239] C. Lanaras, E. Baltsavias, and K. Schindler, "Hyperspectral super-resolution by coupled spectral unmixing," in *IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 3586–3594.

[240] F. A. Mianji, Y. Zhang, and Y. Gu, "Resolution Enhancement of Hyperspectral Images Using a Learning-Based Super-Resolution Mapping Technique," in *IEEE*

International Geoscience and Remote Sensing Symposium (IGARSS), vol. 3, 2009, pp. III–813–III–816.

[241] J. Li, Q. Yuan, H. Shen, X. Meng, and L. Zhang, "Hyperspectral Image Super-Resolution by Spectral Mixture Analysis and Spatial–Spectral Group Sparsity," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 9, pp. 1250–1254, 2016.

[242] C. Prévost, K. Usevich, P. Comon, M. Haardt, and D. Brie, "Performance Bounds for Coupled CP Model in the Framework of Hyperspectral Super-Resolution," in *IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 2019, pp. 201–205.

[243] L. Wang and P. Wang, "Sub-Pixel Mapping for Hyperspectral Imagery Using Super-Resolution then Spectral Unmixing," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2016, pp. 461–464.

[244] R. A. Borsoi, T. Imbiriba, and J. C. M. Bermudez, "Super-Resolution for Hyperspectral and Multispectral Image Fusion Accounting for Seasonal Spectral Variability," *IEEE Transactions on Image Processing*, vol. 29, pp. 116–127, 2020.

[245] M. Şimşek and E. Polat, "Sparse Representation-Based Dictionary Learning Methods for Hyperspectral Super-Resolution," in *24th Signal Processing and Communication Application Conference (SIU)*, 2016, pp. 753–756.

[246] D. Çeşmeci and M. K. Güllü, "Orthogonal Design Based Genetic Algorithm for Spatial Resolution Enhancement of Hyperspectral Images," in *2015 23nd Signal Processing and Communications Applications Conference (SIU)*, 2015, pp. 2442–2445.

[247] R. Wu, W.-K. Ma, X. Fu, and Q. Li, "Hyperspectral Super-Resolution via Global–Local Low-Rank Matrix Estimation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 10, pp. 7125–7140, 2020.

[248] X. Han, J. Yu, J.-H. Xue, and W. Sun, "Hyperspectral and Multispectral Image Fusion Using Optimized Twin Dictionaries," *IEEE Transactions on Image Processing*, vol. 29, pp. 4709–4720, 2020.

[249] X. Li, Y. Yuan, and Q. Wang, "Hyperspectral and Multispectral Image Fusion Based on Band Simulation," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 3, pp. 479–483, 2020.

[250] R. Wu, H.-T. Wai, and W.-K. Ma, "Hybrid Inexact BCD for Coupled Structured Matrix Factorization in Hyperspectral Super-Resolution," *IEEE Transactions on Signal Processing*, vol. 68, pp. 1728–1743, 2020.

[251] J. Liu, Z. Wu, L. Xiao, J. Sun, and H. Yan, "A Truncated Matrix Decomposition for Hyperspectral Image Super-Resolution," *IEEE Transactions on Image Processing*, vol. 29, pp. 8028–8042, 2020.

[252] C.-H. Lin, F. Ma, C.-Y. Chi, and C.-H. Hsieh, "A Convex Optimization-Based Coupled Nonnegative Matrix Factorization Algorithm for Hyperspectral and Multispectral Data Fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 3, pp. 1652–1667, 2018.

[253] Y. Yuan, Q. Wang, and X. Li, "Hyperspectral and Multispectral Image Fusion Using Non-Convex Relaxation Low Rank and Total Variation Regularization," in *IEEE International Geoscience and Remote Sensing Symposium*, 2020, pp. 2683–2686.

[254] S. Li, R. Dian, L. Fang, and J. M. Bioucas-Dias, "Fusing Hyperspectral and Multispectral Images via Coupled Sparse Tensor Factorization," *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 4118–4130, 2018.

[255] F. Ma, S. Huo, and F. Yang, "Graph-Based Logarithmic Low-Rank Tensor Decomposition for the Fusion of Remotely Sensed Images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 11 271–11 286, 2021.

[256] M. Ding, X. Fu, T.-Z. Huang, J. Wang, and X.-L. Zhao, "Hyperspectral Super-Resolution via Interpretable Block-Term Tensor Modeling," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 3, pp. 641–656, 2021.

[257] Y. Qian, F. Xiong, S. Zeng, J. Zhou, and Y. Y. Tang, "Matrix-Vector Nonnegative Tensor Factorization for Blind Unmixing of Hyperspectral Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 3, pp. 1776–1792, 2017.

[258] G. Zhang, X. Fu, K. Huang, and J. Wang, "Hyperspectral Super-Resolution: A Coupled Nonnegative Block-term Tensor Decomposition Approach," 2019.

[259] R. Dian, S. Li, and L. Fang, "Learning a Low Tensor-Train Rank Representation for Hyperspectral Image Super-Resolution," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 9, pp. 2672–2683, 2019.

[260] A. Aidini, M. Giannopoulos, A. Pentari, K. Fotiadou, and P. Tsakalides, "Hyperspectral Image Compression and Super-Resolution Using Tensor Decomposition Learning," in *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, 2019, pp. 1369–1373.

[261] J. Qu, Y. Li, Q. Du, and H. Xia, "Hyperspectral and Panchromatic Image Fusion via Adaptive Tensor and Multi-Scale Retinex Algorithm," *IEEE Access*, vol. 8, pp. 30 522–30 532, 2020.

[262] Y. Xu, Z. Wu, J. Chanussot, P. Comon, and Z. Wei, "Nonlocal Coupled Tensor CP Decomposition for Hyperspectral and Multispectral Image Fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 1, pp. 348–362, 2020.

[263] Y. Bu, Y. Zhao, J. Xue, J. C.-W. Chan, S. G. Kong, C. Yi, J. Wen, and B. Wang, "Hyperspectral and Multispectral Image Fusion via Graph Laplacian-Guided Coupled Tensor Decomposition," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 1, pp. 648–662, 2021.

[264] W. Wan, W. Guo, H. Huang, and J. Liu, "Nonnegative and Nonlocal Sparse Tensor Factorization-Based Hyperspectral Image Super-Resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 12, pp. 8384–8394, 2020.

[265] K. Wang, Y. Wang, X.-L. Zhao, J. C.-W. Chan, Z. Xu, and D. Meng, "Hyperspectral and Multispectral Image Fusion via Nonlocal Low-Rank Tensor Decomposition and Spectral Unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 11, pp. 7654–7671, 2020.

[266] Y. Zhang, S. De Backer, and P. Scheunders, "Noise-Resistant Wavelet-Based Bayesian Fusion of Multispectral and Hyperspectral Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 11, pp. 3834–3843, 2009.

[267] N. Akhtar, F. Shafait, and A. Mian, "Bayesian Sparse Representation for Hyperspectral Image Super Resolution," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 3631–3640.

[268] Q. Wei, J. Bioucas-Dias, N. Dobigeon, and J. Tourneret, "Hyperspectral and Multispectral Image Fusion Based on a Sparse Representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 7, pp. 3658–3668, 2015.

[269] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "An Augmented Lagrangian Approach to the Constrained Optimization Formulation of Imaging Inverse Problems," *IEEE Transactions on Image Processing*, vol. 20, no. 3, pp. 681–695, 2011.

[270] L. Bungert, D. A. Coomes, M. J. Ehrhardt, J. Rasch, R. Reisenhofer, and C.-B. Schönlieb, "Blind Image Fusion for Hyperspectral Imaging with The Directional Total Variation," *Inverse Problems*, vol. 34, no. 4, p. 044003, mar 2018. [Online]. Available: https://doi.org/10.1088/1361-6420/aaaf63

[271] R. Dian, S. Li, A. Guo, and L. Fang, "Deep Hyperspectral Image Sharpening," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 11, pp. 5345–5355, 2018.

[272] X. Han, B. Shi, and Y. Zheng, "SSF-CNN: Spatial and Spectral Fusion with CNN for Hyperspectral Image Super-Resolution," in *25th IEEE International Conference on Image Processing (ICIP)*, 2018, pp. 2506–2510.

[273] X.-H. Han, Y. Zheng, and Y.-W. Chen, "Multi-Level and Multi-Scale Spatial and Spectral Fusion CNN for Hyperspectral Image Super-Resolution," in *IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, 2019, pp. 4330–4339.

[274] Q. Xie, M. Zhou, Q. Zhao, D. Meng, W. Zuo, and Z. Xu, "Multispectral and Hyperspectral Image Fusion by MS/HS Fusion Net," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 1585–1594.

[275] T. Zhang, Y. Fu, L. Wang, and H. Huang, "Hyperspectral Image Reconstruction Using Deep External and Internal Learning," in *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 8558–8567.

[276] X. Wang, J. Chen, Q. Wei, and C. Richard, "Hyperspectral Image Super-Resolution via Deep Prior Regularization with Parameter Estimation," 2021.

[277] M. Vella, B. Zhang, W. Chen, and J. F. C. Mota, "Enhanced Hyperspectral Image Super-Resolution via RGB Fusion and TV-TV Minimization," in *IEEE International Conference on Image Processing (ICIP)*, 2021, pp. 3837–3841.

[278] X. Wang, J. Chen, and C. Richard, "Hyperspectral Image Super-Resolution with Deep Priors and Degradation Model Inversion," 2022. [Online]. Available: https://arxiv.org/abs/2201.09851

[279] L. Zhang, J. Nie, W. Wei, Y. Li, and Y. Zhang, "Deep Blind Hyperspectral Image Super-Resolution," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 6, pp. 2388–2400, 2021.

[280] W. Wei, J. Nie, Y. Li, L. Zhang, and Y. Zhang, "Deep Recursive Network for Hyperspectral Image Super-Resolution," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 1233–1244, 2020.

[281] J. Gao, J. Li, and M. Jiang, "Hyperspectral and Multispectral Image Fusion by Deep Neural Network in a Self-Supervised Manner," *Remote Sensing*, vol. 13, no. 16, 2021. [Online]. Available: https://www.mdpi.com/2072-4292/13/16/3226

[282] W. Chen, X. Zheng, and X. Lu, "Hyperspectral Image Super-Resolution with Self-Supervised Spectral-Spatial Residual Network," *Remote Sensing*, vol. 13, no. 7, 2021. [Online]. Available: https://www.mdpi.com/2072-4292/13/7/1260

[283] W. Sun, K. Ren, X. Meng, C. Xiao, G. Yang, and J. Peng, "A Band Divide-and-Conquer Multispectral and Hyperspectral Image Fusion Method," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–13, 2021.

[284] J. Hu, Y. Li, M. Zhao, and Y. Zhang, "Deep Spatial-Spectral Information Exploitation for Rapid Hyperspectral Image Super-Resolution," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2019, pp. 3109–3112.

[285] J. R. Patel, M. V. Joshi, and J. S. Bhatt, "A Novel Approach for Hyperspectral Image Superresolution Using Spectral Unmixing and Transfer Learning," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2020, pp. 1512–1515.

[286] J. Nie, L. Zhang, C. Wang, W. Wei, and Y. Zhang, "Robust Deep Hyperspectral Imagery Super-Resolution," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2019, pp. 847–850.

[287] Y. Zheng, J. Li, Y. Li, K. Cao, and K. Wang, "Deep Residual Learning for Boosting the Accuracy of Hyperspectral Pansharpening," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 8, pp. 1435–1439, 2020.

[288] J.-F. Hu, T.-Z. Huang, and L.-J. Deng, "Fusformer: A Transformer-Based Fusion Approach for Hyperspectral Image Super-Resolution," 2021.

[289] Z. Zhu, J. Hou, J. Chen, H. Zeng, and J. Zhou, "Hyperspectral Image Super-Resolution via Deep Progressive Zero-Centric Residual Learning," *IEEE*

Transactions on Image Processing, vol. 30, p. 1423–1438, 2021. [Online]. Available: http://dx.doi.org/10.1109/TIP.2020.3044214

[290] C. Ge, Q. Du, W. Sun, K. Wang, J. Li, and Y. Li, "Deep Residual Network-Based Fusion Framework for Hyperspectral and LiDAR Data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 2458–2472, 2021.

[291] J. Xiao, J. Li, Q. Yuan, and L. Zhang, "A Dual-UNet With Multistage Details Injection for Hyperspectral Image Fusion," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–13, 2021.

[292] X. Lu, J. Zhang, D. Yang, L. Xu, and F. Jia, "Cascaded Convolutional Neural Network-Based Hyperspectral Image Resolution Enhancement via an Auxiliary Panchromatic Image," *IEEE Transactions on Image Processing*, vol. 30, pp. 6815–6828, 2021.

[293] J. Yang, L. Xiao, Y.-Q. Zhao, and J. C.-W. Chan, "Variational Regularization Network With Attentive Deep Prior for Hyperspectral-Multispectral Image Fusion," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–17, 2021.

[294] R. Dian, S. Li, and X. Kang, "Regularizing Hyperspectral and Multispectral Image Fusion by CNN Denoiser," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 3, pp. 1124–1135, 2021.

[295] W. Dong, S. Hou, S. Xiao, J. Qu, Q. Du, and Y. Li, "Generative Dual-Adversarial Network With Spectral Fidelity and Spatial Enhancement for Hyperspectral Pansharpening," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, 2021.

[296] F. Zhou, R. Hang, Q. Liu, and X. Yuan, "Pyramid Fully Convolutional Network for Hyperspectral and Multispectral Image Fusion," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 5, pp. 1549–1558, 2019.

[297] Q. Ma, J. Jiang, X. Liu, and J. Ma, "Learning A 3D-CNN and Transformer Prior for Hyperspectral Image Super-Resolution," 2021. [Online]. Available: https://arxiv.org/abs/2111.13923

[298] J.-F. Hu, T.-Z. Huang, L.-J. Deng, T.-X. Jiang, G. Vivone, and J. Chanussot, "Hyperspectral Image Super-Resolution via Deep Spatiospectral Attention Convolutional Neural Networks," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, 2021.

[299] X. Lu, D. Yang, J. Zhang, and F. Jia, "Hyperspectral Image Super-Resolution Based on Spatial Correlation-Regularized Unmixing Convolutional Neural Network," *Remote Sensing*, vol. 13, no. 20, 2021. [Online]. Available: https://www.mdpi.com/2072-4292/13/20/4074

[300] S. M. A. Bashir, Y. Wang, and M. Khan, "A Comprehensive Review of Deep Learning-Based Single Image Super-Resolution," *CoRR*, vol. abs/2102.09351, 2021. [Online]. Available: https://arxiv.org/abs/2102.09351

[301] H. H. Bauschke and J. M. Borwein, "On Projection Algorithms for Solving Convex Feasibility Problems," *SIAM Review*, vol. 38, no. 3, pp. 367–426, 1996. [Online]. Available: http://www.jstor.org/stable/2132495

[302] J. Nascimento and J. Dias, "Vertex Component Analysis: A Fast Algorithm to Unmix Hyperspectral Data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 898–910, 2005.

[303] Z. Guo, T. Wittman, and S. Osher, "L1 Unmixing and Its Application to Hyperspectral Image Enhancement," in *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XV*, S. S. Shen and P. E. Lewis, Eds., vol. 7334, International Society for Optics and Photonics. SPIE, 2009, pp. 579 – 587. [Online]. Available: https://doi.org/10.1117/12.818245

[304] X. Xu, X. Tong, J. Li, H. Xie, Y. Zhong, L. Zhang, and D. Song, "Hyperspectral Image Super Resolution Reconstruction with a Joint Spectral-Spatial Sub-Pixel

Mapping Model," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2016, pp. 6129–6132.

[305] H. Irmak, G. B. Akar, and S. E. Yüksel, "A MAP-Based Approach to Resolution Enhancement of Hyperspectral Images," in *7th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2015, pp. 1–4.

[306] Y. Wang, X. Chen, Z. Han, and S. He, "Hyperspectral Image Super-Resolution via Nonlocal Low-Rank Tensor Approximation and Total Variation Regularization," *Remote Sensing*, vol. 9, no. 12, 2017. [Online]. Available: https://www.mdpi.com/2072-4292/9/12/1286

[307] P. M. Atkinson, "Issues of Uncertainty in Super-Resolution Mapping and Their Implications for the Design of an Inter-Comparison Study," *International Journal of Remote Sensing*, vol. 30, no. 20, pp. 5293–5308, 2009. [Online]. Available: https://doi.org/10.1080/01431160903131034

[308] J. Liu, Z. Wu, L. Xiao, and X.-J. Wu, "Model Inspired Autoencoder for Unsupervised Hyperspectral Image Super-Resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–12, 2022. [Online]. Available: https://doi.org/10.1109%2Ftgrs.2022.3143156

[309] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a Deep Convolutional Network for Image Super-Resolution," in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham: Springer International Publishing, 2014, pp. 184–199.

[310] J. Kim, J. K. Lee, and K. M. Lee, "Accurate Image Super-Resolution Using Very Deep Convolutional Networks," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1646–1654.

[311] V. Romanuke, "An Improvement of the VDSR Network for Single Image Super-Resolution by Truncation and Adjustment of the Learning Rate Parameters,"

Applied Computer Systems, vol. 24, no. 1, pp. 61–68, 2019. [Online]. Available: https://doi.org/10.2478/acss-2019-0008

[312] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. P. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," *CoRR*, vol. abs/1609.04802, 2016.

[313] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced Deep Residual Networks for Single Image Super-Resolution," *CoRR*, vol. abs/1707.02921, 2017. [Online]. Available: http://arxiv.org/abs/1707.02921

[314] X. Hu, M. A. Naiel, A. Wong, M. Lamm, and P. Fieguth, "RUNet: A Robust UNet Architecture for Image Super-Resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019.

[315] A. Heydari and A. Mehmood, "SRVAE: Super Resolution Using Variational Autoencoders," in *Defense + Commercial Sensing*, 2020.

[316] J. Andrew, T. Mhatesh, R. D. Sebastin, K. M. Sagayam, J. Eunice, M. Pomplun, and H. Dang, "Super-Resolution Reconstruction of Brain Magnetic rResonance Images Via Lightweight Autoencoder," *Informatics in Medicine Unlocked*, vol. 26, p. 100713, 2021.

[317] J. Yang, Y.-Q. Zhao, and J. C.-W. Chan, "Hyperspectral Image Super-Resolution Based on Multi-Scale Wavelet 3D Convolutional Neural Network," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2019, pp. 2770–2773.

[318] H.-X. Dou, Z.-C. Wu, Y.-W. Zhuo, L.-J. Deng, and G. Vivone, "Dynamical Fusion Model with Joint Variational and Deep Priors for Hyperspectral Image Super-Resolution," *IEEE Geoscience and Remote Sensing Letters*, pp. 1–1, 2023.

[319] Y. Long, X. Wang, M. Xu, S. Zhang, S. Jiang, and S. Jia, "Dual Self-Attention Swin Transformer for Hyperspectral Image Super-Resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–12, 2023.

[320] K. Zhang, D. Zhu, X. Min, and G. Zhai, "Implicit Neural Representation Learning for Hyperspectral Image Super-Resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–12, 2023.

[321] Y. Li, S. Chen, W. Luo, L. Zhou, and W. Xie, "Hyperspectral Image Super-Resolution Based on Spatial-Spectral Feature Extraction Network," *Chinese Journal of Electronics*, vol. 32, no. 3, pp. 415–428, 2023.

[322] S. Jia, S. Zhu, Z. Wang, M. Xu, W. Wang, and Y. Guo, "Diffused Convolutional Neural Network for Hyperspectral Image Super-Resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–15, 2023.

[323] H. Wang, C. Wang, and Y. Yuan, "Asymmetric Dual-Direction Quasi-Recursive Network for Single Hyperspectral Image Super-Resolution," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1–1, 2023.

[324] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "Mixup: Beyond Empirical Risk Minimization," in *International Conference on Learning Representations*, 2018. [Online]. Available: https://openreview.net/forum?id=r1Ddp1-Rb

[325] T. Dettmers, "Which GPU(s) to Get for Deep Learning: My Experience and Advice for Using GPUs in Deep Learning," 2023. [Online]. Available: http://timdettmers.com/2023/01/30/which-gpu-for-deep-learning/

[326] X. Du, X. Qu, Y. He, and D. Guo, "Single Image Super-Resolution Based on Multi-Scale Competitive Convolutional Neural Network," *Sensors*, vol. 18, no. 3, 2018.

[327] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *CoRR*, vol. abs/1412.6980, 2017. [Online]. Available: https://arxiv.org/abs/1412.6980

[328] T.-H. Lin and C.-H. Lin, "Single Hyperspectral Image Super-Resolution Using Admm-Adam Theory," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2022, pp. 1756–1759.

[329] "Training and Test Sets: Splitting Data," Google, 2022. [Online]. Available: https://developers.google.com/machine-learning/crash-course/training-and-test-sets/splitting-data

[330] J. Brownlee, "A Gentle Introduction to Early Stopping to Avoid Overtraining Neural Networks," Machine Learning Mastery, 2019. [Online]. Available: https://machinelearningmastery.com/early-stopping-to-avoid-overtraining-neural-network-models/

[331] Y. Bengio, "Practical Recommendations for Gradient-Based Training of Deep Architectures," *CoRR*, vol. abs/1206.5533, 2012. [Online]. Available: http://arxiv.org/abs/1206.5533

[332] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[333] "tf.keras.optimizers.Adam," Tensorflow. [Online]. Available: https://www.tensorflow.org/api_docs/python/tf/keras/optimizers/Adam

[334] J. Yu, Y. Fan, J. Yang, N. Xu, Z. Wang, X. Wang, and T. S. Huang, "Wide Activation for Efficient and Accurate Image Super-Resolution," *CoRR*, vol. abs/1808.08718, 2018. [Online]. Available: http://arxiv.org/abs/1808.08718

[335] N. Aburaed, M. Q. Alkhatib, S. Marshall, J. Zabalza, and H. A. Ahmad, "SISR of Hyperspectral Remote Sensing Imagery Using 3D Encoder-Decoder RUNet Architecture," in *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, 2022, pp. 1516–1519.

[336] H. Torabi, S. L. Mirtaheri, and S. Greco, "Practical Autoencoder Based Anomaly Detection by Using Vector Reconstruction Error," *Cybersecurity*, vol. 6, no. 1, p. 1, Jan 2023. [Online]. Available: https://doi.org/10.1186/s42400-022-00134-9

[337] S. Chen and W. Guo, "Auto-Encoders in Deep Learning: A Review with New Perspectives," *Mathematics*, vol. 11, no. 8, 2023. [Online]. Available: https://www.mdpi.com/2227-7390/11/8/1777

[338] K. Bajaj, D. K. Singh, and M. A. Ansari, "Autoencoders Based Deep Learner for Image Denoising," *Procedia Computer Science*, vol. 171, pp. 1535–1541, 2020, third International Conference on Computing and Network Communications (CoCoNet'19). [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050920311431

[339] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.

[340] S. Park, "Is Batch Normalization Harmful? Improving Normalizer-Free ResNets," Jul 2021. [Online]. Available: https://medium.com/geekculture/is-batch-normalization-harmful-improving-normalizer-free-resnets-cf44f2fc0b2e

[341] N. Aburaed, M. Q. Alkhatib, S. Marshall, J. Zabalza, and H. Al Ahmad, "Bayesian Hybrid Loss for Hyperspectral SISR Using 3D Wide Residual CNN," 2023, [Accepted].

[342] J. T. Barron, "A More General Robust Loss Function," *CoRR*, vol. abs/1701.03077, 2017. [Online]. Available: http://arxiv.org/abs/1701.03077

[343] P. Chen, G. Chen, and S. Zhang, "Log Hyperbolic Cosine Loss Improves Variational Auto-Encoder," 2019. [Online]. Available: https://openreview.net/forum?id=rkglvsC9Ym

[344] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a Deep Convolutional Network for Image Super-Resolution," in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham: Springer International Publishing, 2014, pp. 184–199.

[345] H. Deborah, N. Richard, and J. Y. Hardeberg, "A Comprehensive Evaluation of Spectral Distance Functions and Metrics for Hyperspectral Image Processing," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 3224–3234, 2015.

[346] P. I. Frazier, "A Tutorial on Bayesian Optimization," 2018. [Online]. Available: https://arxiv.org/abs/1807.02811

[347] N. Aburaed, M. Q. Alkhatib, S. Marshall, J. Zabalza, and H. A. Ahmad, "A Comparative Study of Loss Functions for Hyperspectral SISR," in *2022 30th European Signal Processing Conference (EUSIPCO)*, 2022, pp. 484–487.

[348] J. Yoo, N. Ahn, and K.-A. Sohn, "Rethinking Data Augmentation for Image Super-Resolution: A Comprehensive Analysis and a New Strategy," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 8372–8381.

[349] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *Journal of Machine Learning Research*, vol. 15, no. 56, pp. 1929–1958, 2014. [Online]. Available: http://jmlr.org/papers/v15/srivastava14a.html

[350] R. Timofte, R. Rothe, and L. Van Gool, "Seven Ways to Improve Example-Based Single Image Super Resolution," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1865–1873.

[351] C. Shorten and T. M. Khoshgoftaar, "A Survey on Image Data Augmentation for Deep Learning," *Journal of Big Data*, vol. 6, no. 1, p. 60, Jul 2019.

[352] P. Chen, S. Liu, and J. Jia, "Jigsaw Clustering for Unsupervised Visual Representation Learning," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 11 521–11 530.

[353] N. Aburaed, M. Q. Alkhatib, S. Marshall, J. Zabalza, and H. Al Ahmad, "Hyperspectral Data Scarcity Problem from a Super Resolution Perspective: Data Augmentation Analysis and Scheme," 2023, [Accepted].

[354] R. Gonzales and R. Woods, *Digital Image Processing*. New Jersey, USA: Addison-Wesley Publishing Company, 1992, pp. 81–125.

[355] R. Haralick and L. Shapiro, *Computer and Robot Vision*, ser. Computer and Robot Vision. Addison-Wesley Publishing Company, 1992, no. v. 1. [Online]. Available: https://books.google.co.uk/books?id=jvRRAAAAMAAJ

[356] F. Xiao and Y. Zhang, "A Comparative Study on Thresholding Methods in Wavelet-Based Image Denoising," *Procedia Engineering*, vol. 15, pp. 3998–4003, 2011, cEIS 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877705811022508

[357] X. Liu, J. Yu, T. Kurihara, C. Wu, Z. Niu, and S. Zhan, "Pixelwise Complex-Valued Neural Network Based on 1D FFT of Hyperspectral Data to Improve Green Pepper Segmentation in Agriculture," *Applied Sciences*, vol. 13, no. 4, 2023. [Online]. Available: https://www.mdpi.com/2076-3417/13/4/2697

[358] M. Peker, "Classification of Hyperspectral Imagery Using a Fully Complex-Valued Wavelet Neural Network with Deep Convolutional Features," *Expert Systems with Applications*, vol. 173, p. 114708, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417421001494

[359] R. Anand, S. Veni, and J. Aravinth, "Robust Classification Technique for Hyperspectral Images Based on 3D-Discrete Wavelet Transform," *Remote Sensing*, vol. 13, no. 7, 2021. [Online]. Available: https://www.mdpi.com/2072-4292/13/7/1255

[360] T. Chakraborty and U. Trehan, "SpectralNET: Exploring Spatial-Spectral WaveletCNN for Hyperspectral Image Classification," 2021.

[361] F. Luisier, T. Blu, B. Forster, and M. Unser, "Which Wavelet Bases are the Best for Image Denoising?" in *Wavelets XI*, M. Papadakis, A. F. Laine, and M. A. Unser, Eds., vol. 5914, International Society for Optics and Photonics. SPIE, 2005, pp. 100–111. [Online]. Available: https://doi.org/10.1117/12.614999

[362] c. P. Dautov and M. S. Özerdem, "Wavelet transform and signal denoising using wavelet method," in *2018 26th Signal Processing and Communications Applications Conference (SIU)*, 2018, pp. 1–4.

[363] N. Guberman, "On Complex Valued Convolutional Neural Networks," *CoRR*, vol. abs/1602.09046, 2016. [Online]. Available: http://arxiv.org/abs/1602.09046

[364] J. W. Cooley and J. W. Tukey, "An Algorithm for the Machine Calculation of Complex Fourier Series," *Mathematics of Computation*, vol. 92, no. 343, pp. 297–301, 1965.

[365] N. Aburaed, M. Q. Alkhatib, S. Marshall, J. Zabalza, and H. A. Ahmad, "Complex-valued Neural Network for Hyperspectral Single Image Super Resolution," in *Hyperspectral Imaging and Applications II*, N. J. Barnett, A. A. Gowen, and H. Liang, Eds., vol. 12338, International Society for Optics and Photonics. SPIE, 2023, p. 123380H. [Online]. Available: https://doi.org/10.1117/12.2645086

[366] N. Aburaed, M. Q. Alkhatib, S. Marshall, J. Zabalza, and H. Al Ahmad, "Attention-infused 3D-SRCNN for Hyperspectral Image Super Resolution," 2023, [Accepted].

[367] B. Zhukov, D. Oertel, F. Lanzl, and G. Reinhackel, "Unmixing-Based Multisensor Multiresolution Image Fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 3, pp. 1212–1226, 1999.

[368] M. Goel, E. Whitmire, A. Mariakakis, T. S. Saponas, N. Joshi, D. Morris, B. Guenter, M. Gavriliu, G. Borriello, and S. N. Patel, "HyperCam: Hyperspectral Imaging for Ubiquitous Computing Applications," in *Proceedings*

of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, ser. UbiComp '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 145–156. [Online]. Available: https://doi.org/10.1145/2750858.2804282

[369] S. W. Oh, M. S. Brown, M. Pollefeys, and S. J. Kim, "Do It Yourself Hyperspectral Imaging with Everyday Digital Cameras," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2461–2469.

[370] T. Takatani, T. Aoto, and Y. Mukaigawa, "One-Shot Hyperspectral Imaging Using Faced Reflectors," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2692–2700.

[371] W. Sun, B. Chen, and D. Messinger, "Nearest-Neighbor Diffusion-Based Pan-Sharpening Algorithm for Spectral Images," *Optical Engineering*, vol. 53, no. 1, p. 013107, 2014.

[372] G. Vivone, L. Alparone, J. Chanussot, M. Dalla Mura, A. Garzelli, G. A. Licciardi, R. Restaino, and L. Wald, "A Critical Comparison Among Pansharpening Algorithms," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 5, pp. 2565–2586, 2015.

[373] R. Dian and S. Li, "Hyperspectral Image Super-Resolution via Subspace-Based Low Tensor Multi-Rank Regularization," *Trans. Img. Proc.*, vol. 28, no. 10, p. 5135–5146, Oct. 2019. [Online]. Available: https://doi.org/10.1109/TIP.2019.2916734

[374] N. Aburaed, A. Panthakkan, M. Al-Saad, M. C. E. Rai, S. A. Mansoori, H. Al-Ahmad, and S. Marshall, "Super-Resolution of Satellite Imagery Using a Wavelet Multiscale-Based Deep Convolutional Neural Network Model," in *Image and Signal Processing for Remote Sensing XXVI*, L. Bruzzone, F. Bovolo, and E. Santi, Eds., vol. 11533, International Society for Optics and Photonics. SPIE, 2020, pp. 305–311.

[375] N. Aburaed, A. Panthakkan, S. Almansoori, and H. Al-Ahmad, "Super Resolution of DS-2 Satellite Imagery Using Deep Convolutional Neural Network," in *Image and Signal Processing for Remote Sensing XXV*, L. Bruzzone and F. Bovolo, Eds., vol. 11155, International Society for Optics and Photonics. SPIE, 2019, pp. 485–491.

[376] Y. Li, J. Hu, X. Zhao, W. Xie, and J. Li, "Hyperspectral Image Super-Resolution Using Deep Convolutional Neural Network," *Neurocomputing*, vol. 266, pp. 29–41, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S092523121730841X

[377] Z. Xiong, Z. Shi, H. Li, L. Wang, D. Liu, and F. Wu, "HSCNN: CNN-Based Hyperspectral Image Recovery from Spectrally Undersampled Projections," in *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, 2017, pp. 518–525.

[378] R. Ran, L.-J. Deng, T.-X. Jiang, J.-F. Hu, J. Chanussot, and G. Vivone, "GuidedNet: A General CNN Fusion Framework via High-Resolution Guidance for Hyperspectral Image Super-Resolution," *IEEE Transactions on Cybernetics*, vol. 53, no. 7, pp. 4148–4161, 2023.

[379] A. Khader, J. Yang, and L. Xiao, "Model-Guided Deep Unfolded Fusion Network With Nonlocal Spatial-Spectral Priors for Hyperspectral Image Super-Resolution," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 4607–4625, 2023.

[380] S. Mei, G. Zhang, N. Wang, B. Wu, M. Ma, Y. Zhang, and Y. Feng, "Lightweight Multiresolution Feature Fusion Network for Spectral Super-Resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–14, 2023.

[381] Y. Wang, W. Li, N. Liu, Y. Gui, and R. Tao, "FuBay: An Integrated Fusion Framework for Hyperspectral Super-Resolution Based on Bayesian Tensor Ring," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, 2023.

[382] W. Dong, J. Qu, S. Xiao, T. Zhang, Y. Li, and X. Jia, "Noise Prior Knowledge Informed Bayesian Inference Network for Hyperspectral Super-Resolution," *IEEE Transactions on Image Processing*, vol. 32, pp. 3121–3135, 2023.

[383] Q. Li, M. Gong, Y. Yuan, and Q. Wang, "RGB-Induced Feature Modulation Network for Hyperspectral Image Super-Resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–11, 2023.

[384] Y. Qu, H. Qi, and C. Kwan, "Unsupervised Sparse Dirichlet-Net for Hyperspectral Image Super-Resolution," 2018.

[385] ——, "Unsupervised and Unregistered Hyperspectral Image Super-Resolution with Mutual Dirichlet-Net," 2019.

[386] L. Zhang, J. Nie, W. Wei, Y. Zhang, S. Liao, and L. Shao, "Unsupervised Adaptation Learning for Hyperspectral Imagery Super-Resolution," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 3070–3079.

[387] V. Lempitsky, A. Vedaldi, and D. Ulyanov, "Deep Image Prior," in *IEEE CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 9446–9454.

[388] K. Zheng, L. Gao, W. Liao, D. Hong, B. Zhang, X. Cui, and J. Chanussot, "Coupled Convolutional Neural Network With Adaptive Response Function Learning for Unsupervised Hyperspectral Super Resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 3, p. 2487–2502, Mar 2021. [Online]. Available: http://dx.doi.org/10.1109/TGRS.2020.3006534

[389] Z. Liu, Y. Zheng, and X.-H. Han, "Deep Unsupervised Fusion Learning for Hyperspectral Image Super Resolution," *Sensors*, vol. 21, no. 7, 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/7/2348

[390] W. Wei, J. Nie, L. Zhang, and Y. Zhang, "Unsupervised Recurrent Hyperspectral Imagery Super-Resolution Using Pixel-Aware Refinement," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–15, 2020.

[391] J. Nie, L. Zhang, W. Wei, C. Ding, and Y. Zhang, "Unsupervised Deep Hyperspectral Super-Resolution With Unregistered Images," in *IEEE International Conference on Multimedia and Expo (ICME)*, 2020, pp. 1–6.

[392] Z. Wang, B. Chen, R. Lu, H. Zhang, H. Liu, and P. K. Varshney, "FusionNet: An Unsupervised Convolutional Variational Network for Hyperspectral and Multispectral Image Fusion," *IEEE Transactions on Image Processing*, vol. 29, pp. 7565–7577, 2020.

[393] Z. Guo, J. Xin, N. Wang, J. Li, X. Wang, and X. Gao, "Unsupervised Across Domain Consistency- Difference Network for Hyperspectral Image Super-Resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–13, 2023.

[394] L. Gao, J. Li, K. Zheng, and X. Jia, "Enhanced Autoencoders With Attention-Embedded Degradation Learning for Unsupervised Hyperspectral Image Super-Resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–17, 2023.

Appendix A

Fusion Techniques for HSI-SR

Image Fusion is the process of combining information from multiple images, such that the final product reveals more information than the individual input images. The pioneer work in Fusion methods dates back to 1999 [367]. Fusion-based methods combine the observed HR-MSI, and LR-HSI of the same scene. According to the literature, using an LR-HSI with the corresponding HR-MSI image to obtain an HR-HSI has shown promising performance. Most approaches use RGB, therefore, HR-MSI and HR-RGB will be used interchangeably. The existing approaches can be roughly divided into two categories. The first one is to design a specific system based on standard RGB cameras. Exploiting time-multiplexed illumination source, multiple color cameras, and a tube of faced reflectors can be used to complete the reconstruction [368–370]. However, this method relies rigorously on environmental conditions and extra equipment, which makes it impractical and costly. Therefore, HR-MSI and LR-HSI Fusion is the favorable approach, but it is considered as an ill-posed problem due to the amount of lost information. Nonetheless, image Fusion is still possible due to the existence of high correlation between RGB and their corresponding HS radiance. In this study, image Fusion approaches are divided into three categories: pansharpening, method-based, and Deep Convolutional Neural Networks (DCNNs). Method-based approaches are further categorized into Matrix Factorization (MF) and spectral unmixing, Tensor-based, and Bayesian-based, as seen in Table 3.1.

A.1 Pansharpening

One example of image Fusion is pansharpening, which transforms an LR-HSI to HR-HSI by fusing it with a PAN band extracted from an MSI. Pansharpening methods can be broadly grouped into four categories: Component Substitution (CS), Multiresolution Analysis (MRA), variational methods, and hybrid approaches.

A.1.1 Component Substitution (CS)

One of the most widely used CS methods for pansharpening is Brovey Transform (BT) [194]. BT is based on spectral modeling and was developed to increase the visual contrast in the high and low ends of the data's histogram, such as shadows, water, and high reflectance areas. It uses a method that multiplies each resampled HS pixel by the ratio of the corresponding PAN pixel intensity to the sum of all the multispectral intensities. It assumes that the spectral range spanned by the PAN image is the same as that covered by the HS bands. The basic procedure of BT first multiplies each HS band by the HR PAN band, and then divides each product by the sum of the MS bands. In the case of RGB, BT can be described using the following equation:

$$DN_{fused1} = \frac{DN_{b1}}{DN_{b1} + DN_{b2} + DN_{b3}} DN_{(HR-PAN)}, \tag{A.1}$$

Where DN is the Digital Number (pixel value), $DN_{b1,2,3}$ are LR-RGB bands, and DN_{fused1} is the resultant HR band. This equation is repeated for each HS band individually. BT is limited to three bands only, and there are constant suitable weights for each band that are different for each satellite.

Gram-Schmidt (GS) pansharpening technique was first invented by Laben and Brower [195]. A synthetic PAN image is acquired by using a GS mode, of which there are three available. In the first mode, the average of HS bands is taken as a synthetic image. In the second one, low-pass filtered version of PAN image is received as a synthetic PAN image. For the last mode, a synthetic PAN image is achieved by using Least Square regression analysis (GS-LS). By subtracting this synthetic PAN image from the original PAN image, spatial details are obtained. Extracted details

are injected into the HS bands, which are upsampled to PAN resolution in order to generate the pansharpened image. The injection gain factor G_k of band number k in GS pansharpening is defined as follows:

$$G_k = \frac{cov(X_k, I)}{var(I)}, k = 1, 2, \ldots, B, \tag{A.2}$$

where *cov* is the covariance, *var* is the variance, and I is the intensity component of the HSI. [196] attempted to enhance this approach by introducing GS Adaptive (GSA).

Intensity Hue Saturation (IHS) [197] method is a standard CS method that was developed based on the assumption that spectral information is mostly contained within the hue and saturation, leaving the spatial information in the intensity component. The basic approach in IHS method is to replace the intensity component with HR PAN image, which is histogram matched to the intensity component, to obtain the spatial detail matrix. These details are injected into each HS band separately in order to obtain the pansharpened image.

PCA technique [198] transforms the HSI to feature space in order to obtain principal components. The first principal component is assumed to contain most of the energy or most of the spatial information. This term is replaced by the PAN image, which is histogram matched to this component. By using inverse PCA transform the pansharpened HSI is obtained. Other works in this area include [199, 200]. Generally, CS algorithms produce remarkable results in terms of spatial resolution, however, they cause spectral distortions, which is their major drawback. In [201], the authors attempted to overcome this drawback regardless of the type of satellite sensor by introducing Partial Replacement Adaptive CS (PRACS). In 2014, Sun et al. introduced a novel pansharpening method that enhances image resolution by utilizing pixel spectra as the fundamental unit of operation [371]. It employs a mixture model to generate high-resolution spectral images, assuming that each spectrum in the fused image results from a weighted combination of neighboring superpixel spectra in the low-resolution spectral image. These weights are determined by a diffusion model derived from the panchromatic image, reflecting the similarity between the pixel of interest and neighboring superpixels. This approach stands out from existing methods, which typically employ

band-wise processing, and it outperforms Gram–Schmidt method [371]. Nonetheless, MRA-based algorithms have better resistance to spectral distortions. These algorithms are discussed in the next sub-section.

A.1.2 Multiresolution Analsis (MRA)

This section discusses some of the most widely used MRA pansharpening methods. MRA-based approaches can be generally conveyed in the following equation:

$$\widehat{Y}_k = \widetilde{Y}_k + G_k \otimes (P - P_L), k = 1, 2, \ldots, B, \tag{A.3}$$

where $\widetilde{Y}_k$ denotes the k^{th} band of the upsampled (interpolated) HSI, $\otimes$ denotes element-wise multiplication, P is the PAN image and P_L is the low pass version of P. Each MRA-based algorithm extracts P_L and G_k differently. For instance, Smoothing Filter Based Intensity Modulation (SFIM) [202] is a pansharpening technique that obtains P_L from P using a Linear Time Invariant (LTI) low pass filter, such as averaging filter. For G_k, High Pass Modulation (HPM) is utilized. The result is then added to each HS band separately.

Other techniques utilize multiresolution image decomposition to obtain P_L. For instance, Wavelet Transform (WT) is used for decomposing an image into its high and low frequency components, which is a powerful technique for multiple image processing tasks, including de-noising and SR. Thus, there is a variation of WT that can be used for pansharpening. In the first step, histogram matching is performed between the PAN image and each HS band. Afterward, WT is applied to the resulting histogram-matched PAN image. The result of this decomposition is categorized into image details extracted through high-pass filtering, and an approximation image extracted through low pass filtering. Only the approximation part is considered, and the other decompositions are set to zero. Hence, inverse WT yields low resolution PAN image. This result is then subtracted from the original PAN image to create the detail matrix. Pansharpened bands are finally obtained by adding this matrix to each HS band [203]. This approach has been explored and expanded by several studies in the literature [204], such as

Decimated Wavelet Transform using additive injection model (Indusion) [205], Additive A Trous Wavelet Transform (ATWT) [206], A Trous Wavelet Transform using the Model 3 (ATWT-M3) [207] and Additive Wavelet Luminance Proportional (AWLP) [208].

Similar to WT, Laplacian Pyramid (LP), which is derived from Gaussian Pyramid, also decomposes the image into its high frequency and low frequency components. LP was improved into what is known as Enhanced LP (ELP), and then Generalized LP (GLP) was derived from ELP. GLP is used as a pansharpening technique by extracting the PAN image details from LP, and then injecting the details into an upsampled version of the HSI. This method has been extended to GLP with Modulation Transfer Function matched filter (MTF-GLP) [209], GLP with MTF-matched filter and Context-Based Decision injection scheme (MTF-GLP-CBD) [210], and Gaussian MTF-matched filter with HPM injection model (MTF-GLP-HPM) [206, 211].

MRA methods perform well in terms of robustness and efficiency, but suffer from aliasing effect and spatial details distortion [215, 372].

A.1.3 Variational Methods

The first variational model framework known as "P+XS Image Fusion" was introduced by [212]. This approach uses the assumption that the geometry of spectral bands of HSI is related to the topographic map of the corresponding PAN image. The goal is to minimize the energy function, which comprises of three components, by using Gradient Descent algorithm. This variational framework carries out by minimizing the sum of integrals HSI and its low-pass filtered version and tangent vector multiplied gradient of HS bands, and also the integral of the sum of subtraction the PAN image from alpha values multiplied HS bands.

$$\mid \bigtriangledown \widehat{Y}_k \mid -\theta * \bigtriangledown \widehat{Y}_k = 0 \tag{A.4}$$

$\widehat{Y}_k$ represents the k^{th} band of the pansharpened image $\widehat{Y}$, and θ denotes the normal vector field of the PAN image. The first component, E_g, that must be minimized is the spatial fidelity term, which is expressed as such:

$$E_g = \sum_{k=1}^{B} \int_{\Omega} (\mid \bigtriangledown \widehat{Y}_k \mid -\theta * \bigtriangledown \widehat{Y}_k) \tag{A.5}$$

where $\Omega \subset R^2$ denotes an open, bounded domain with a Lipschitz boundary. The second component, E_r, that must be minimized is based on the image relation hypothesis. This hypothesis states that PAN image is the product of weighted sums of each MSI band, where each weight corresponds to the energy of the respective spectral band. Thus, the second component is expressed as follows:

$$E_r = \int_{\Omega} (\sum_{k=1}^{B} \alpha_k \widehat{Y}_k - P)^2 \tag{A.6}$$

where α are the weighting coefficients of each band, and P denotes the PAN image. The final component that must be minimized, E_f, is based on the assumption that the LR image is the product of convolution between the HR image and a low-pass filter. Thus, this component is expressed as such:

$$E_f = \sum_{k=1}^{B} \int_{\Omega} \prod_{S} (K_k * \widehat{Y}_k - X_k)^2 dx, \tag{A.7}$$

where $\prod_S$ is a Dirac's comb defined by the grid S, K_k is the convolution kernel of the low-pass filter, and X_k represents the k_{th} band of the LR-HSI. All three components are finally expressed as such [212]:

$$\begin{aligned} E &= \sum_{k=1}^{B} \gamma_k \int_{\Omega} (\mid \bigtriangledown \widehat{Y}_k \mid - \theta * \bigtriangledown \widehat{Y}_k) \\ &= \lambda \int_{\Omega} (\sum_{k=1}^{B} \alpha_k \widehat{Y}_k - P)^2 + \eta \int_{\Omega} \prod_{s} (K_k * \widehat{Y}_k - X_k)^2 dx \end{aligned} \tag{A.8}$$

$\gamma_k, \lambda, \eta > 0$, γ_k allows controlling the relative weight of each band, and λ and η allow controlling weights of each component of the equation.

The most notable methods that extended this framework are Nonlocal Variant (NLV), and Nonlocal Variant with band-decouple (NLVD), both proposed by [213] and [214], respectively. In 2018, [215] enhanced this framework by proposing a variational pansharpening method for HSI constrained by spectral shape and GS transformation. First, the authors utilize the spectral shape feature of the neighboring pixels with a

new weight distribution strategy to reduce spectral distortions caused by the change in spatial resolution. Then, the correlation fidelity term uses the result of GSA to constrain the correlation, thereby preventing the low correlation between the pansharpened image and the reference image. Then, the pansharpening is formulated as the minimization of a new energy function, which produces the final pansharpened image. The authors claim that this method outperforms GSA, guided filter PCA, MTF, SFIM, intensity modulation, the classic and the band-decoupled variational methods. These methods do not limit the number of bands, but suffer from high computational cost and large spectral distortion. Other works in this area include [216–218].

A.1.4 Hybrid Approaches

One of the main challenges for fusing LR-HS and HR-PAN/RGB data is to find an appropriate balance between spectral and spatial preservation. Hybrid approaches use mixture of CS and MRA methods to overcome this challenge. [219, 220] designed a Guided Filter in the PCA domain (GFPCA). Instead of using CS, which may cause spectral distortions, GFPCA uses a high resolution PAN/RGB image to guide the filtering process aimed at obtaining SR. In this way, GFPCA does not only preserve the spectral information from the original HSI, but also transfers the spatial structures of the high resolution PAN/RGB image to the enhanced HSI. GFPCA first uses PCA to decorrelate the bands of the HSI, and to separate the information content from the noise. The first PCA channels contain most of the energy of an HSI, and the remaining PCA channels mainly contain noise. When GF is applied to these noisy channels, it amplifies the noise and causes a high computational cost in processing the data, which is undesirable. Therefore, guided filtering is used to enlarge only the first PCA channels, preserving the structures of the PAN/RGB image, while bicubic interpolation is used to upsample the remaining channels.

A.2 Method-based

A.2.1 Matrix Factorization and Spectral Unmixing

Spectral unmixing is the procedure by which the measured spectrum of a mixed pixel is decomposed into a collection of constituent spectra, or endmembers, and a set of corresponding fractions or abundances, that indicate the proportion of each endmember present in the pixel [72]. The basic principle of MF [73] is to associate the Fusion problem with "linear spectral unmixing"; the data can be described by a linear combination of spectral signals, also called reflectance function basis. Each signal uniquely corresponds to a material present in the scene. Spectral unmixing refers to the problem of finding the number of endmembers in an HSI, their spectral signatures, and their per-pixel abundances. It is the inverse of spectral mixing described as follows:

$$\begin{aligned} r_i &= \sum_{j=1}^{p} w_{ij} + n_i = Hw_i + n_i \\ \sum_{j=1}^{p} w_{ij} &\leq 1, w_{ij} > 0, \quad i = 1, 2, 3, \dots L \end{aligned} \tag{A.9}$$

where $\mathbf{r}_i$ is the spectral vector expressed by a linear combination of several endmember vectors $\mathbf{h}$, p is the number of endmembers in the image, L is the number of pixels, and w_{ij} is a scalar representing the fractional abundance of endmember vector $\mathbf{h}_j$ in the pixel $\mathbf{r}_i$. $\mathbf{H}$ is of size $B \times p$ mixing matrix, where B is the number of bands and $p \ll L$.

An example of this approach for HS unmixing is demonstrated in [221], where the authors proposed Coupled Non-negative Matrix Factorization (CNMF) for HS and MS data Fusion, and studied its effect on HSI classification. Their approach unmixed both sources of data to find the signatures and abundance of the endmembers as described earlier. The relationship between low and high spatial resolution in HS can be described as follows:

$$X = YD_s + \mathcal{E}_s, \tag{A.10}$$

where $\mathbf{D}_s$ is the spatial transform matrix and $\mathcal{E}_s$ is the residual error. D_s is deter-

mined by image registrations and estimation of PSF. Similarly, the relationship between low and high spectral resolution in MS can be described as follows:

$$Z = D_r U + \mathcal{E}_r, \tag{A.11}$$

where $\mathbf{U}$ is the high spectral resolution MSI, $\mathbf{Z}$ is the low spectral resolution MSI, $\mathbf{D}_r$ is the spectral transformation matrix and $\mathcal{E}_r$ is the residual error. D_r is derived from radiometric calibration to obtain Spectral Response Function (SRF). From equation 10, HS and MS can be expressed as follows:

$$Y = W_Y H_Y + E_Y \tag{A.12}$$

$$U = W_U H_U + E_U \tag{A.13}$$

W and H are abundance and endmember matrices, respectively, and E_Y and E_U are residual error matrices. NMF spectral unmixing is commonly performed to minimize the squared Frobenius norm of the residual matrix in the linear spectral mixture model expressed as $\|E_Y\|_F^2$ and $\|E_U\|_F^2$. This principle can be applied to estimate the up-sampled HSI. Assuming HS and MS images capture the same scene, their endmembers should be the same, and the abundance map of LR-HSI data should match that of MSI. The abundance matrix can be extracted from the MSI, and then used to enhance the spatial resolution of HSI [221–224]. HR-HSI can then be approximated as

$$\widehat{Y} \approx W_Y H_U \tag{A.14}$$

CNMF for HSI-SR was extended in [225] to test its effect on target detection. Their experiment showed that CNMF can restore pure spectra, which contributes to accurate target detection. Despite the effectiveness of CNMF approach, the obtained solution is usually not unique, which can lead to an unsatisfactory outcome. In [226], the authors attempted to extend the aforementioned methods where endmembers are extracted directly from down-sampled HSIs. The derived endmembers are used as an input to an unmixing algorithm applied to the MSI. The obtained abundances are then used

to reconstruct the HR-HSI. Another approach that uses spectral unmixing is depicted in [227]. The method is split into two stages; spatial upsampling and spectral substitution. Spatial upsampling is done by estimating an optimal linear combination on exemplar patches for SR reconstruction, followed by evaluation using learned local spectrum dictionary. This approach unmixes HS observation within a pixel using guidance from HR-RGB image. In the spectrum substitution stage, sparse coding is adopted. This stage refines the spectrum obtained in the first stage based on the limited materials assumption within a local region of a scene. Other variations of NMF include Sparse Non-negative Matrix Factorization (SNNMF) [228]. HSI can be unmixed using various mathematical functions, learning algorithms, and probabilistic frameworks, such as K-SVD [224], Bayesian Sparse Representation (BSR) [229,268], HySure [230,231], Maximum a Posteriori (MAP) [232, 233], and GSOM+ [234]. An approach similar to [221] was followed in [223], but the mixing matrix was replaced with a dictionary learned using a non-negative matrix factorization with sparsity regularization code. Another sparse representation dictionary learning method was used in [222], where two dictionaries were learned from the HSI and MSI, and then dictionary-pair learning method was used to establish correspondence between them. Motivated by the successful applications of sparse representation, [235] proposed a Nonnegative Structured Sparse Representation (NSSR) approach for taking consideration of the spatial structure. The authors then conducted optimization procedure with the Alternative Direction Multiplier Method (ADMM) technique [236]. NSSR achieved a large margin on HSI recovery performance compared with the other state-of-the-art approaches. A similar approach was devised recently by [237, 238]. The effectiveness of dictionary approaches strongly depend on how these dictionaries can be obtained. [239] proposed HSI-SR by integrating Coupled Spectral Unmixing (CSU) strategy into HSI-SR and conducted optimization procedure with the proximal alternating linearized minimization method. Other works in this area include [240–252]. One shortcoming of this method and MF approaches in general is that they require good initial points of the two decomposed reflectance signatures to provide satisfactory results. Furthermore, most work generally assumes that the number of the pure materials in the observed scene is smaller than the spectral

band number, which is not always satisfied in the real application. Additionally, MF approaches generally suffer either from spectral distortions or high computation time.

A.2.2 Tensor-based Approaches

A tensor is considered as a generalization form of a matrix. Tensors can be used in the context of HSI-SR by addressing the non-uniqueness of tensor rank. Imposing prior information or regularization are examples of ways to answer this non-uniqueness [185, 253]. HSI are low-rank and self-similar [231, 254, 268]. Therefore, applying low-rank regularization on the core of a tensor addresses this non-uniqueness and avoids the necessity of obtaining an exact value of tensor rank. A tensor-based observation model can be expressed as such:

$$X = YD_1D_2 + \mathcal{E}_s \tag{A.15}$$

$$Z = YD_3 + \mathcal{E}_r \tag{A.16}$$

$\mathbf{D}_1$ and $\mathbf{D}_2$ denote the degradation matrices of the spatial resolution, which can be constructed by downsampling Toeplitz matrix if the PSF is known. $\mathbf{D}_3$ represents the degradation in the spectral resolution, which can be constructed if the SRF of the HS and MS sensors are known. Using this model, tensor decomposition can then be used to estimate $\mathbf{Y}$. For instance, following Tucker decomposition, $\mathbf{Y}$ can be decomposed as follows:

$$Y = TMNB + \mathcal{E}_z, \tag{A.17}$$

where T is the decomposed core tensor, and $\mathcal{E}_z$ is the error term. Consequently, the Fusion model can be formulated as follows:

$$min_Y \frac{1}{2}\|X - YD_1D_2\|_F^2 + \frac{1}{2}\|Z - GD_3\|_F^2 \tag{A.18}$$

Prior information must be incorporated in order to regularize equation A.18. For instance, in [255], the authors take advantage of the similarities between adjacent bands as

well as neighboring pixels, and impose graph regularization on spatial and spectral matrices separately to minimize the effects of distortion. Another example is demonstrated in [373], where Dian and Li propose a Low Tensor Multi-Rank (LTMR) regularization method that exploits high correlation among spectral bands, as well as non-local spatial similarities. In [256], Ding et al. follow a similar strategy, but instead of Tucker decomposition, they use coupled tenosr LL1-based decomposition framework to estimate HR-HSI due to its connection to linear mixture models, as suggested by [257]. In [254], Li et al. considered HR-HSI as a 3D tensor, and formulated the Fusion problem by estimating the core tensor and three dictionary modes through Coupled Sparse Tensor Factorization (CSTF) approach. They also incorporate a regularizer to model the high spectral-spatial correlations. Other examples include [238, 258–265].

A.2.3 Bayesian-based Approaches

The first known approach that utilizes Bayesian Fusion was devised by Zhang et al. in 2009 [266]. The Fusion framework of this approach takes place in the wavelet domain, and is referred to as Wavelet MAP. The authors assume additive noise imaging model for the HSI, and interpolation is used as a priori to bypass the need to estimate the spatial degradation operator and perform SR in a blind manner. Performing MAP [232] to approximate the enhanced image in the wavelet domain rather than the spatial domain allows for scale-specific and subband-specific estimations. The authors compare their approach to spatial domain estimation, in addition to some of the most commonly used pansharpening techniques. Another blind Bayesian-based approach is HySure devised by Simões et al [230, 231].

Most optimization-based approaches rely on explicit parameter turning for each different dataset or sensor. To solve this problem, Akhtar et al. [267] attempt to avoid this problem by utilizing non-parametric Bayesian Sparse Representation (BSR) over four stages. First, the probability distributions and the proportions of the material reflectance spectra are extracted. Second, a dictionary is estimated and transformed based on the spectral quantization of the HR-MSI. Third, the sparse codes of the HR-MSI is computed by using the proposed Bayesian sparse coding. Finally, the HR-HSI

is estimated using the dictionary from the second stage and the sparse coding from the third stage. Another BSR approach was devised in [268], where the authors learn dictionaries from the observed HSI and MSI, and then solve the optimization problem with respect to the target image and the sparse code by using Split Augmented Lagrangian shrinkage algorithm (SALSA) [269], which is an instance of ADMM. According to the authors, "SALSA enables a huge nondiagonalizable quadratic problem to be decomposed into a sequence of convolutions and pixel decoupled problems, which can be solved efficiently" [268]. This approach outperforms MAP [232] and Wavelet MAP [266].

Fusion approaches assume that the HSI and MSI are perfectly co-registered, which is an impractical assumption. In an attempt to overcome this limitation, Bungert et al. [270] devise a blind Bayesian approach with Directional Total Variation that is robust against imprecise registrations between the HSI and MSI.

A.3 Deep Convolutional Neural Networks

DCNNs have recently shown great success in various image processing and computer vision applications. DCNNs have also been applied to RGB image SR and achieved promising performance [36, 374, 375]. Since the correlation between MSI and HSI is highly non-linear, DCNNs have high potential to achieve HR-HS with high accuracy if HR-RGB image is used. Some researchers in the literature focused on utilizing DCNNs to obtain HR-HSI from its LR version only [376], an approach known as SISR. In this case, the CNN is known as Spatial-CNN. However, their enhancement factor is limited to 8 at maximum compared to using observed RGB or, more generally, MS data. HSI require a much higher enhancement factor (e.g. 32). Further elaboration on this can be found in Section 3.1. Other researchers improved the spectral resolution of LR-RGB images using DCNN, in which case it is known as Spectral-CNN [36, 377]. This approach ignores the HS attribute that correlates the narrowband and spectra, which leads to unsatisfactory results. Spatial-DCNN and Spectral-DCNN improve the image in one dimension only. Therefore, it is desirable to design a network architecture that performs enhancements in both dimensions in order to generate an HR-HSI. This

will be explored in the next subsections from supervised and unsupervised learning perspectives.

A.3.1 Supervised

Traditional pansharpening approaches, albeit primitive, can be elevated using DCNNs as well. A notable example of such case is Deep HSI Sharpening (DHSIS) model demonstrated in [271], which learns image priors and incorporates them into the Fusion framework. First, the HR-HSI is initialized by solving Sylvester equation. Then, a one-to-one mapping between the initialized HR-HSI and the reference HR-HSI is learned via deep residual CNN. The priors learned from this network are then utilized in the Fusion framework to obtain the final estimated HR-HSI. This approach shows superiority against traditional pansharpening and MF approaches. A Spatial and Spectral Fusion Network (SSF-Net) for HR-HSI reconstruction was proposed by [272]. The results of the network were promising in spite of the simple concatenation of the upsampled LR-HSI and the HR-RGB image. However, the upsampling of the LR-HSI and the simple concatenation cannot effectively integrate the existing spatial structure and spectral property without high computational cost. In addition, precise alignment is needed for the input of LR-HSI and HR-RGB images, and it is extremely difficult to attain due to the large difference of spatial resolution in the LR-HSI and HR-RGB images.

In 2019, inspired by the success of SSF-Net, Han et al. devised Multi-level and Multi-scale SSF-Net (MS-SSFNet) [273], which fuses LR-HSI with HR-RGB. The authors' proposed DCNN relies on the gradual reduction of the feature sizes of the HR-RGB while increasing the feature sizes of the LR-HSI. Furthermore, DCNNs often suffer from vanishing gradient problem during the training, and the authors alleviate this problem by integrating multi-level cost functions into MS-SSFNet architecture. Other works in this area that also tackle Fusion with spectral-spatial context include [274,275].

All the aforementioned methods assume that the degradation kernels are already known. In order to find a middle-ground between hand-crafted priors and DCNNs that do not assume prior knowledge but need massive training data, many authors used the output of DCNNS as deep prior regularizers [276–278]. For instance, the authors

in [279] attempt to overcome this problem blindly without any prior assumptions. The authors use DCNN to regularize the spatial and spectral degradation instead of using hand-crafted priors. The authors utilize a generator network to model the latent HR-HSI, and the spatial degradation are modeled through a convolutional layer, while the spectral degradation are modeled through a full connected layer. Thus, the network works as an end-to-end pipeline that learns from the LR-HSI and the HR-HSI. This approach outperforms traditional blind Fusion approaches, such as NSSR [235]. A similar blind approach idea is presented in [280], where Wei et al. used a deep recursive residual network to fuse LR-HSI with HR-MSI. Other works in this area include [281–283, 283–299, 318, 378–383].

A.3.2 Unsupervised

Supervised learning algorithms for image Fusion require a large size of HSI dataset perfectly registered with their MSI counterparts, which is unrealistic. Unsupervised learning offers the possibility to bypass this limitation, as it has the potential to achieve remarkable results with small datasets compared to supervised learning approaches. Qu et al. [384] were the first to attempt this task for HSI-SR using CNN. Their network consists of two encoder-decoders that are coupled by the same decoder in order to preserve spectral information. Sparse Dirichlet distribution naturally covers the physical constraints of HSI and MSI. This allows minimizing the angular difference between HSI and MSI representation, which reduces spectral distortions. The resulting network is referred to as uSDN. One of the major challenges that faces this network and image Fusion in general is the assumption that HR-MSI and LR-HSI are accurately registered. The performance of the Fusion typically relies on the registration accuracy. Therefore, [385] attempt to overcome the shortcomings of uSDN by projecting both HR-MSI and LR-HSI into the same statistical space. This representation is assumed to follow Dirichlet representation as well. The authors also exploit Mutual Information (MI) between both images to capture any non-linear statistical dependencies between them. Maximizing MI leads to maximizing spatial correlations, which leads to minimizing spectral distortions. The authors test their approach on CAVE and Harvard

datasets using ERGAS, PSNR, and SAM evaluation metrics. It can be observed that this approach offers an advantage over conventional Fusion methods as well as uSDN.

In a similar approach, Lei et al. [386] take advantage of image prior and utilize it for unsupervised learning that consists of two-stage SR. The first stage is a Fusion model that is pre-trained on synthetic data to generate a general spatial-spectral HSI prior. The second stage is a degeneration model that makes the general HSI prior more specific, which is trained in an unsupervised way. The algorithm performance shows superiority against traditional Fusion algorithms, as well as the results demonstrated by [387] and [274].

Inspired by the recent success of unsupervised DCNNs, Liu et al. [308] embedded NMF into their approach and developed a Model Inspired Autoencoder (MIAE) for unsupervised HSI-SR. NMF's task is to preserve the intrinsicality of the estimated HR-HSI, such that the autoencoder takes each individual HSI pixel as an input sample for the encoder side, and outputs spectral and spatial matrices at the decoder side. However, the value of the input pixel is unknown, so the LR-HSI and the HR-MSI are used as inputs in a pixel-wise manner that is solved by using gradient descent. The loss function of the autoencoder is formulated based on spectral and spatial degradation. Instead of assuming this degradation as shallow priors, the authors propose an additional blind estimation network to estimate the PSF and SRF. The approach outperforms traditional Fusion approaches in addition to [388] and [384]. Other works in this area include [386, 388–394].

Appendix B

Publications from MBRSC Projects

- M. Q. Alkhatib, M. Alsaad, **N. Aburaed**, M. S. Zitouni and H. Al Ahmad, "Pol-SAR Image Classification Using Attention Based Shallow to Deep Convolutional Neural Network," *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Pasadena, California, 2023. [Accepted]

- M. Alsaad, **N. Aburaed**, M. S. Zitouni, M. Q. Alkhatib, S. Almansoori and H. Al Ahmad, "A Robust Change Detection Methodology For Flood Events Using SAR Images," *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Pasadena, California, 2023. [Accepted]

- M. Q. Alkhatib, M. Alsaad, **N. Aburaed**, S. Almansoori, J. Zabalza, S. Marshall and H. Al Ahmad, "Tri-CNN: A Three Branch Model for Hyperspectral Image Classification," *Remote Sensing*, vol. 15, no. 2, p. 316, Jan. 2023, doi: `https://doi.org/10.3390/rs15020316`.

- M. Al-Saad, **N. Aburaed**, S. Al Mansoori, W. Mansoor and H. Al-Ahmad, "A Study on Deep Learning Approaches for Mars Weather Forecasting," *5th International Conference on Signal Processing and Information Security (ICSPIS)*, Dubai, United Arab Emirates, 2022, pp. 76-79, doi: `https://doi.org/10.1109/ICSPIS57063.2022.10002657`.

- **N. Aburaed**, M. Alsaad, S. A. Mansoori and H. Al-Ahmad, "A Study on the Autonomous Detection of Impact Craters" In *Artificial Neural Networks in Pattern Recognition. ANNPR 2022. Lecture Notes in Computer Science*, vol 13739. Springer, Cham. https://doi.org/10.1007/978-3-031-20650-4_15.

- M. Q. Alkhatib, M. Al-Saad, **N. Aburaed**, S. A. Mansoori and H. Al Ahmad, "Dimensionality Reduction Techniques with Hydranet Framework for HSI Classification," *IEEE International Conference on Image Processing (ICIP)*, Bordeaux, France, 2022, pp. 3151-3155, doi: https://doi.org/10.1109/ICIP46576.2022.9897740.

- M. Al-Saad, **N. Aburaed**, S. A. Mansoori and H. A. Ahmad, "Autonomous Palm Tree Detection from Remote Sensing Images - UAE Dataset," *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Kuala Lumpur, Malaysia, 2022, pp. 2191-2194, doi: https://doi.org/10.1109/IGARSS46834.2022.9884126.

- **N. Aburaed**, M. Al-Saad, A. Panthakkan, S. al Mansoori, H. Al-Ahmad and S. Marshall, "A Hybrid Rexception Network for COVID-19 Classification from Chest X-Ray Images," *28th IEEE International Conference on Electronics, Circuits, and Systems (ICECS)*, Dubai, United Arab Emirates, 2021, pp. 1-5, doi: https://doi.org/10.1109/ICECS53924.2021.9665598.

Appendix C

Explored α Values Using BOA

C.1 Pavia University Experiments

Appendix C. Explored α Values Using BOA

Table C.1: Experiment 1: the explored α values using BOA during the first 15 iterations show that the optimum PSNR is achieved at iteration 3 with $\alpha = 0.743334$ for scale factors $\times 2$, and at iteration 5 with $\alpha = 0.970302$ for scale factor $\times 4$.

Iteration	$\times 2$				$\times 4$			
	α	PSNR (dB)	SSIM	SAM (°)	α	PSNR (dB)	SSIM	SAM (°)
0	0.970326	33.013	0.9262	4.21	0.970327	27.764	0.7680	6.45
1	0.592509	32.962	0.9255	4.30	0.592509	27.746	0.7661	6.59
2	0.975416	32.999	0.9263	4.22	0.975416	27.724	0.7662	6.47
3	**0.743334**	**33.037**	**0.9275**	4.22	0.743334	27.745	0.7672	6.56
4	0.727955	32.965	0.9257	4.27	0.727956	27.752	0.7666	6.54
5	0.743320	32.952	0.9256	4.29	**0.970302**	**27.798**	**0.7697**	**6.40**
6	0.743346	33.000	0.9269	4.30	0.970232	27.768	0.7679	6.40
7	0.970308	32.976	0.9260	4.19	0.970281	27.756	0.7679	6.44
8	0.975448	33.023	0.9269	4.20	0.970293	27.753	0.7668	6.47
9	0.970356	32.901	0.9255	4.28	0.970357	27.747	0.7676	6.46
10	0.975429	32.997	0.9267	4.22	0.90624	27.820	0.7704	6.47
11	0.975375	32.950	0.9259	4.23	0.970213	27.783	0.7687	6.43
12	0.975458	32.977	0.9262	**4.18**	0.970147	27.724	0.7648	6.49
13	0.574135	32.963	0.9256	4.33	0.970316	27.754	0.7672	6.45
14	0.425046	33.011	0.9266	4.25	0.970198	27.743	0.7667	6.44

Table C.2: Experiment 2: the explored α values using BOA during the first 15 iterations on Pavia University dataset show that the optimum PSNR is achieved at iteration 8 with $\alpha = 0.975429$ for scale factor $\times 2$ and iteration 11 with $\alpha = 0.683219$ for scale factor $\times 4$.

Iteration	$\times 2$				$\times 4$			
	α	PSNR	SSIM	SAM	α	PSNR	SSIM	SAM
0	0.958278	32.980	0.9259	4.29	0.813497	27.746	0.7667	6.44
1	0.958286	32.951	0.9254	4.27	0.529864	27.753	0.7673	6.54
2	0.743334	32.996	0.9257	4.19	0.975312	27.749	0.7669	6.46
3	0.719829	33.018	0.9257	4.29	0.975318	27.746	0.7677	6.47
4	0.719844	32.966	0.9256	4.22	0.719856	27.726	0.7662	6.47
5	0.74332	32.957	0.9251	4.29	0.975347	27.752	0.7680	6.46
6	0.612487	33.005	0.9250	4.24	0.975328	27.767	0.7661	6.40
7	0.958205	32.979	0.9258	4.20	0.813462	27.784	0.7672	6.43
8	**0.975429**	**33.026**	**0.9263**	**4.19**	0.975361	27.755	0.7703	6.47
9	0.948754	32.912	0.9259	4.28	0.975384	27.759	0.7665	6.47
10	0.948706	32.999	0.9253	4.22	0.936721	27.751	0.7702	6.45
11	0.975375	32.957	0.9260	4.21	**0.683219**	**27.818**	**0.7675**	**6.40**
12	0.970308	32.983	0.9247	4.25	0.975394	27.745	0.7660	6.49
13	0.727955	32.960	0.9252	4.22	0.975326	27.723	0.7677	6.46
14	0.428761	33.012	0.9253	4.22	0.813429	27.799	0.7680	6.58

Table C.3: Experiment 3: the explored α values using BOA during the first 15 iterations on Pavia University dataset show that the optimum PSNR is achieved at iteration 2 with $\alpha = 0.743320$ for scale factor $\times 2$, but it fails to achieve the best SAM, and iteration 14 with $\alpha = 0.843267$ for scale factor $\times 4$.

Iteration	$\times 2$				$\times 4$			
	α	PSNR	SSIM	SAM	α	PSNR	SSIM	SAM
0	0.970326	32.996	0.9258	4.22	0.843219	27.770	0.7666	6.44
1	0.592509	32.978	0.9254	4.29	0.628542	27.744	0.7667	6.46
2	**0.743320**	**33.038**	**0.9258**	4.20	0.956738	27.758	0.7667	6.46
3	0.743334	32.959	0.9257	4.22	0.712394	27.757	0.7672	6.46
4	0.727955	33.023	0.9254	4.27	0.698274	27.746	0.7667	6.43
5	0.975415	32.965	0.9251	4.28	0.956702	27.751	0.7703	6.49
6	0.743346	33.024	0.9265	**4.18**	0.956743	27.728	0.7673	6.47
7	0.97030	32.995	0.9257	4.21	0.843201	27.756	0.7671	6.45
8	0.975448	32.953	0.9264	4.20	0.956719	27.746	0.7669	6.53
9	0.970356	32.954	0.9245	4.28	0.956765	27.753	0.7664	6.46
10	0.975429	33.011	0.9256	4.22	0.889123	27.736	0.7678	6.45
11	0.975375	32.967	0.9247	4.23	0.956731	27.798	0.7664	6.49
12	0.975458	32.912	0.9255	4.18	0.956772	27.781	0.7677	6.58
13	0.574135	32.979	0.9254	4.32	0.956725	27.766	0.7662	6.56
14	0.425046	33.001	0.9258	4.24	**0.843267**	**27.819**	**0.7703**	**6.40**

Appendix C. Explored α Values Using BOA

Table C.4: Experiment 4: the explored α values using BOA during the first 15 iterations on Pavia University dataset show that the optimum PSNR is achieved at iteration 8 with $\alpha = 0.634721$ for scale factor $\times 2$ and iteration 7 with $\alpha = 0.970288$ for scale factor $\times 4$.

Iteration	$\times 2$				$\times 4$			
	α	PSNR	SSIM	SAM	α	PSNR	SSIM	SAM
0	0.865432	32.977	0.9256	4.29	0.970239	27.754	0.7667	6.45
1	0.493217	32.953	0.9259	4.27	0.970312	27.748	0.7672	6.58
2	0.721983	32.998	0.9260	4.20	0.970152	27.751	0.7669	6.46
3	0.634891	33.019	0.9261	4.25	0.970303	27.752	0.7679	6.56
4	0.592348	32.967	0.9257	4.21	0.975412	27.747	0.7663	6.53
5	0.721457	32.954	0.9258	4.29	0.970201	27.7612	0.7681	6.40
6	0.493874	33.006	0.9254	4.23	0.970221	27.727	0.7662	6.42
7	0.865719	32.979	0.9260	4.22	**0.970288**	**27.798**	**0.7703**	**6.40**
8	**0.634721**	**33.025**	**0.9263**	**4.18**	0.970359	27.783	0.7676	6.47
9	0.592716	32.908	0.9254	4.28	0.727951	27.744	0.7666	6.46
10	0.721529	32.997	0.9256	4.22	0.592518	27.761	0.7700	6.46
11	0.634187	32.959	0.9259	4.23	0.970296	27.766	0.7671	6.44
12	0.865198	32.982	0.9252	4.28	0.906210	27.724	0.7674	6.49
13	0.493649	32.962	0.9253	4.19	0.743331	27.753	0.7658	6.45
14	0.592983	33.011	0.9255	4.22	0.592509	27.745	0.7667	6.44

Appendix C. Explored α Values Using BOA

Table C.5: Experiment 5: the explored α values using BOA during the first 15 iterations on Pavia University dataset show that the optimum PSNR is achieved at iteration 13 with $\alpha = 0.592714$ for scale factor $\times 2$, and at iteration 11 with $\alpha = 0.927126$ for scale factor $\times 4$, which fails to achieve the best SAM.

Iteration	$\times 2$				$\times 4$			
	α	PSNR	SSIM	SAM	α	PSNR	SSIM	SAM
0	0.684327	32.940	0.9256	4.29	0.743219	27.746	0.7668	6.45
1	0.752491	33.021	0.9260	4.27	0.592864	27.753	0.7675	6.58
2	0.927136	32.957	0.9258	4.22	0.927431	27.748	0.7678	6.46
3	0.531864	33.041	0.9257	4.28	0.634198	27.744	0.7674	6.56
4	0.826975	32.926	0.9254	4.22	0.865712	27.727	0.7671	6.53
5	0.492317	33.032	0.9259	4.29	0.492713	27.752	0.7687	6.41
6	0.634279	32.961	0.9256	4.23	0.752491	27.767	0.7669	**6.40**
7	0.865712	33.015	0.9253	4.20	0.865231	27.783	0.7669	6.44
8	0.721983	32.935	0.9257	4.25	0.721864	27.755	0.7676	6.47
9	0.592471	33.028	0.9254	4.28	0.531327	27.757	0.7664	6.46
10	0.493126	32.967	0.9260	4.21	0.634975	27.751	0.7696	6.46
11	0.865231	33.004	0.9254	4.20	**0.927126**	**27.819**	**0.7704**	6.42
12	0.927641	32.922	0.9257	4.18	0.592471	27.745	0.7666	6.49
13	**0.592714**	**33.049**	**0.9261**	**4.19**	0.684327	27.723	0.7681	6.45
14	0.634198	32.950	0.9259	4.22	0.493198	27.798	0.7648	6.44

C.2 Washington DC Mall Experiments

Table C.6: Experiment 1: the explored α values using BOA during the first 15 iterations of training 3D-SRCNN333 on Washington DC Mall dataset show that the highest PSNR is achieved at iteration 2 with $\alpha = 0.9754156$ for scale factor $\times 2$, and $\alpha = 0.743334$ for scale factor $\times 4$.

Iteration	$\times 2$				$\times 4$			
	α	PSNR	SSIM	SAM (°)	α	PSNR	SSIM	SAM (°)
0	0.970327	32.223	0.9362	4.54	0.970327	27.440	0.8019	7.80
1	0.592509	32.174	0.9354	4.63	0.592509	27.409	0.7999	7.85
2	**0.975416**	**32.245**	**0.9365**	**4.54**	**0.743334**	**27.476**	**0.8030**	**7.75**
3	0.743334	32.188	0.9363	4.62	0.975416	27.441	0.8029	7.79
4	0.727956	32.182	0.9355	4.62	0.727956	27.438	0.8005	7.85
5	0.743321	32.212	0.9364	4.57	0.975369	27.436	0.8018	7.78
6	0.744372	32.204	0.9362	4.56	0.975426	27.437	0.8020	7.75
7	0.726771	32.230	0.9360	4.53	0.74337	27.460	0.8030	7.80
8	0.593698	32.187	0.9344	4.62	0.743472	27.467	0.8021	7.80
9	0.729336	32.190	0.9356	4.59	0.743502	27.464	0.8020	7.82
10	0.590915	32.223	0.9363	4.57	0.975406	27.435	0.8013	7.78
11	0.589584	32.238	0.9364	4.54	0.743388	27.451	0.8018	7.82
12	0.588235	32.216	0.9360	4.57	0.743525	27.450	0.8019	7.82
13	0.586904	32.208	0.9362	4.59	0.74348	26.172	0.8028	12.71
14	0.730805	32.157	0.9357	4.63	0.970362	27.456	0.8032	7.76

Table C.7: Experiment 2: the explored α values using BOA during the first 15 iterations of training 3D-SRCNN333 on Washington DC Mall dataset show that the highest PSNR is achieved at iteration 12 with $\alpha = 0.507654$ for scale factor $\times 2$, and iteration 2 with $\alpha = 0.941268$ for scale factor $\times 4$, which does not achieve the best SSIM.

Iteration	$\times 2$				$\times 4$			
	α	PSNR	SSIM	SAM	α	PSNR	SSIM	SAM
0	0.862745	33.274	0.9373	4.46	0.935712	27.378	0.8004	9.13
1	0.511235	34.081	0.9373	4.45	0.720145	27.405	0.7985	8.76
2	0.867812	33.191	0.9375	4.41	**0.941268**	**27.481**	0.8009	**7.55**
3	0.628977	34.042	0.9370	4.45	0.789243	27.442	0.8007	8.31
4	0.611235	34.089	0.9372	4.42	0.773591	27.433	0.7990	8.74
5	0.628965	34.032	0.9372	4.43	0.789236	27.437	0.8002	7.87
6	0.630123	34.098	0.9368	4.47	0.790375	27.438	0.8003	7.92
7	0.609877	34.093	0.9370	4.46	0.772123	27.465	**0.8015**	9.04
8	0.512346	34.159	0.9371	4.46	0.721489	27.469	0.8008	8.99
9	0.630988	34.161	0.9373	4.44	0.775622	27.466	0.8003	8.24
10	0.509877	34.166	0.9371	4.48	0.718925	27.431	0.7996	7.88
11	0.508765	34.176	0.9370	4.45	0.717564	27.457	0.8001	8.26
12	**0.507654**	**34.191**	**0.9376**	**4.40**	0.716309	27.450	0.8001	8.28
13	0.506543	34.140	0.9370	4.44	0.715065	26.198	0.8011	12.70
14	0.632099	34.190	0.9376	4.43	0.777834	27.455	0.8016	7.56

Table C.8: Experiment 3: the explored α values using BOA during the first 15 iterations of training 3D-SRCNN333 on Washington DC Mall dataset show that the highest PSNR is achieved at iteration 1 with $\alpha = 0.628743$ for scale factor $\times 2$, and at iteration 5 with $\alpha = 0.812340$ for scale factor $\times 4$.

Iteration	$\times 2$				$\times 4$			
	α	PSNR	SSIM	SAM	α	PSNR	SSIM	SAM
0	0.836459	31.873	0.9351	4.58	0.982345	27.478	0.8012	7.89
1	**0.628743**	**32.965**	**0.9362**	**4.50**	0.734567	27.445	0.7997	7.87
2	0.842617	31.812	0.9348	4.51	0.988765	27.431	0.8006	7.85
3	0.712934	32.754	0.9355	4.54	0.812345	27.434	0.8013	7.94
4	0.697148	32.839	0.9353	4.52	0.796543	27.439	0.7999	7.79
5	0.712927	32.742	0.9352	4.55	**0.812340**	**27.463**	**0.8015**	**7.77**
6	0.714035	32.710	0.9347	4.60	0.813456	27.471	0.8004	7.82
7	0.695821	32.705	0.9350	4.57	0.794321	27.467	0.8019	7.98
8	0.630587	32.861	0.9351	4.58	0.73789	27.430	0.8010	7.92
9	0.699752	32.864	0.9357	4.54	0.798765	27.458	0.8002	7.88
10	0.626418	32.870	0.9352	4.61	0.731234	27.451	0.7995	7.78
11	0.625071	32.881	0.9351	4.55	0.729876	26.211	0.8001	7.95
12	0.623736	32.896	0.9359	4.53	0.728543	27.459	0.8003	7.82
13	0.622409	32.844	0.9351	4.54	0.727219	27.120	0.8007	7.94
14	0.701989	32.895	0.9358	4.56	0.800987	27.086	0.8008	7.78

Table C.9: Experiment 4: the explored α values using BOA during the first 15 iterations of training 3D-SRCNN333 on Washington DC Mall dataset show that the highest PSNR is achieved at iteration 1 with $\alpha = 0.621345$ for scale factor $\times 2$, and at iteration 5 with $\alpha = 0.805432$ for scale factor $\times 4$.

Iteration	$\times 2$				$\times 4$			
	α	PSNR	SSIM	SAM	α	PSNR	SSIM	SAM
0	0.915678	32.467	0.9287	4.64	0.799876	27.385	0.8024	7.41
1	**0.621345**	**33.890**	**0.9311**	**4.50**	0.502345	27.403	0.8005	7.36
2	0.921345	31.574	0.9302	4.51	**0.805432**	**27.481**	**0.8021**	**6.90**
3	0.688765	32.126	0.9284	4.55	0.622109	27.446	0.8027	7.79
4	0.672345	32.213	0.9299	4.53	0.606789	27.432	0.8002	7.83
5	0.68876	32.048	0.9293	4.54	0.622104	27.435	0.8013	7.54
6	0.689876	32.205	0.9286	4.60	0.623456	27.438	0.8006	7.70
7	0.670987	32.982	0.929	4.58	0.604321	27.464	0.8022	7.31
8	0.623456	32.716	0.9287	4.59	0.504321	27.470	0.8015	7.25
9	0.674321	32.172	0.9298	4.54	0.609876	27.468	0.8009	7.44
10	0.618765	31.975	0.9292	4.61	0.498765	27.431	0.7998	7.12
11	0.61789	33.289	0.9291	4.56	0.497654	27.457	0.8003	7.15
12	0.616789	32.874	0.9301	4.52	0.496543	27.452	0.8001	7.33
13	0.615678	33.621	0.9289	4.54	0.495432	26.189	0.8016	7.25
14	0.676543	32.802	0.9295	4.61	0.611234	27.46	0.8018	7.05

Table C.10: Experiment 5: the explored α values using BOA during the first 15 iterations of training 3D-SRCNN333 on Washington DC Mall Mall dataset show that the highest PSNR is achieved at iteration 6 with $\alpha = 0.722345$ for scale factor $\times 2$, and at iteration 14 with $\alpha = 0.654232$ for scale factor $\times 4$.

Iteration	$\times 2$				$\times 4$			
	α	PSNR	SSIM	SAM	α	PSNR	SSIM	SAM
0	0.978532	31.849	0.9317	4.68	0.892317	27.376	0.8003	9.49
1	0.654321	33.267	0.9324	4.52	0.531468	27.408	0.8001	6.87
2	0.984321	30.762	0.9329	4.55	0.898743	27.481	0.8005	6.94
3	0.721098	34.211	0.9321	4.54	0.665432	27.447	0.7999	7.21
4	0.705432	32.985	0.9326	4.54	0.649876	27.432	0.8002	7.53
5	0.721093	35.027	0.9319	4.62	0.665427	27.434	0.8006	7.59
6	**0.722345**	**36.104**	**0.9330**	**4.51**	0.666789	27.439	0.7998	7.87
7	0.702109	32.309	0.9322	4.59	0.646321	27.465	0.8004	7.92
8	0.656789	34.875	0.9328	4.54	0.535678	27.470	0.8007	7.90
9	0.707654	37.029	0.9323	4.69	0.652109	27.468	0.7997	7.78
10	0.649876	30.541	0.9318	4.56	0.528765	27.433	0.8008	7.87
11	0.648765	33.729	0.9316	4.52	0.527654	27.458	0.7996	7.82
12	0.647654	32.416	0.9315	4.54	0.526543	27.453	0.7995	9.48
13	0.646543	36.001	0.9325	4.58	0.525432	26.203	0.8009	8.77
14	0.710987	34.124	0.9327	4.54	**0.654232**	**28.011**	**0.8051**	**6.81**