Are Machine Learning based methods more robust to poor image quality when performing Hyperspectral Imagery anomaly detection compared with conventional signal-processing methods?

A comparative discussion of the impact of image quality on model performance.

Author: Samantha Steer (RegNo: 202354042)

Supervisors: Stephen Marshall, Paul Murray, Jon Piper

A thesis presented for the degree of
Master of Philosophy

Faculty of Engineering
University of Strathclyde
2026

Declaration

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.

Signed: SAMANTHA STEER

Date: 18/04/2026

Acknowledgments

I would like to sincerely thank my supervisors for their guidance and encouragement throughout this project. Their insights were invaluable to the completion of this thesis. I would also like to acknowledge the DSTL Space Programme for their continued support, which made this research possible. Finally, a thank you to my family and friends for their support and motivation.

Abstract

Hyperspectral imaging (HSI) is being increasingly used for the detection of anomalies in various applications, for example, remote sensing for environmental monitoring. However, the quality of HSI sensors can vary significantly due to noise and resolution, which effect the accuracy of anomaly detection methods applied to the data. To address this, this thesis evaluates how different levels of image degradation and enhancement affect the detection capability of a traditional signal processing algorithm (RX detection), and a Machine Learning (ML) based approach (AETNet). Following analysis of the utility of both ML and conventional models, a series of novel experiments were developed to understand the effect of varying image qualities. The results of these experiments demonstrate that image quality affects detection performance at varying levels, depending upon which technique is being used. ML techniques demonstrated slightly higher performance when compared to conventional methods suggesting they may be less degraded by image noise. These findings demonstrate the importance of considering image quality when developing and applying HSI anomaly detection techniques and the approaches that can be taken to ensure optimal pre-processing to achieve the best results.

Contents

List of Figures

List of Tables

Acronyms

ACE Adaptive Coherence Estimator

AE Autoencoder

AETNet General Anomaly Enhancement Network with Random Masks for Hyperspectral Anomaly Detection

CBAD Cluster-Based Anomaly Detection

CAE Convolutional Autoencoder

CFAR Constant False Alarm Rate

CNN Convolutional Neural Network

DBSCAN Density-Based Spatial Clustering of Applications with Noise

DIP Deep Image Prior

DNN Deep Neural Network

GMM Gaussian Mixture Model

GAN Generative Adversarial Network

HR High Resolution

HSI Hyperspectral Image

LR Low Resolution

ML Machine Learning

MAE Mean Absolute Error

MSI Multispectral Image

PCA Principal Component Analysis

RX Reed-Xiaoli

SF Scale Factor

ViT Vision Transformer

ZSSR Zero-Shot Super-Resolution

Chapter 1

Introduction

1.1 Background

In recent years, Machine Learning (ML) techniques have revolutionised imagery analysis across various applications, including in Hyperspectral Imaging (HSI) [1]. In many cases, these methods have demonstrated higher performance when compared to traditional signal processing techniques [2]. Consequently, there is a growing interest in exploring whether ML methods can enhance HSI anomaly detection compared to conventional approaches, as well as in developing new capabilities based on these techniques. This thesis aims to assess the current literature on this topic, evaluate its relevance to the topic, identify existing gaps and to discuss emerging trends that may shape future research directions in this field. Additionally, this thesis presents novel experimental findings that examine how image quality impacts the performance of an ML-based approach in comparison to a conventional signal processing method, providing an insight into their relative effectiveness under varying conditions to ensure an optimal model can be chosen to achieve the best results.

1.2 Thesis Contribution

This study attempts to address current challenges in HSI and anomaly detection by reviewing and building upon methodologies related to image quality.

- It provides a novel evaluation of the impact of image quality on the performance of both conventional signal processing and Machine Learning (ML) techniques. By comparing the robustness of a traditional method with a modern ML approach, this research offers new and valuable insights into their effectiveness under a variety of image conditions.

- The work introduces an original experimental framework designed to assess how different levels of image degradation and enhancement affect anomaly detection performance, thereby advancing the understanding of the role of image quality in automated HSI analysis.
- The thesis identifies the current key trends and challenges in the application of ML to HSI, suggesting pathways for in-depth future research.

Overall, this thesis aims to contribute to the practical application and theoretical understanding of image quality and anomaly detection in HSI, enabling the development of more complete detection algorithms.

1.3 Thesis Plan

The structure of this thesis is as follows:

- Chapter 1 provides an introduction to the research topic by outlining the motivation, objectives, and overarching contributions.
- Chapter 2 details the technical background of the thesis and details the methods, technology and models that underpin the work.
- Chapter 3 contains a literature review that explores current and past research in the field of HSI ML, anomaly detection and the role of image quality in these topics.
- Chapter 4 discusses the methodology used in the experiments conducted, including the metrics used to evaluate the performance of chosen techniques.
- Chapter 5 Presents the results of the experiments and discusses the findings of the study within the context of the wider field by highlighting the key findings and limitations of the conducted research.
- Chapter 6 summarises the main contributions of this thesis and suggests potential areas for future research.

Chapter 2

Technical Background

2.1 Introduction

2.1.1 What is Hyperspectral Imaging?

Hyperspectral Imaging (HSI) is an imaging technique that records a continuous spectrum of light for every pixel in a scene. Whilst standard photography collects only three broad spectral bands (Red, Green, Blue), HSI sensors can collect data across multiple bands (often hundreds). This results in a data cube that contains the spectral signatures for every pixel in an image, enabling the identification of materials based upon their unique spectral properties [3].

2.1.2 What is Anomaly Detection?

In HSI, an anomaly is a pixel or cluster of pixels that are spectrally significantly different to the surrounding local pixels. The anomaly maybe be sub-pixel or larger. An example anomaly would be a man-made metal object within a forest, here the spectral reflectance curve would vastly differ from the vegetation background. Anomaly detection is the process of identifying and extracting these pixels, often through their spectral deviation compared to their backgrounds. In HSI, particularly within HSI remote sensing, anomaly detection is used to identify objects that are of unknown origin. For example, in a defence context, anomaly detection techniques could be used to find camouflaged vehicles that may be indistinguishable in standard RGB imaging. The HSI methods utilise the spectral context of the image and may therefore distinguish objects of a manmade nature when compared to natural backgrounds (e.g. vegetation or desert) [4].

In this chapter, the essential technical concepts that are utilised within this research are examined and discussed. A clear understanding of these principles is required to fully understand the scope and objectives of the study. To aid in understanding, a HSI simulating three material types (water, soil and vegetation) was generated synthetically to show examples of key techniques applied. This 100x100 pixel image was generated in Python using numpy and matplotlib for visualisation. The code simulated a 30-band hyperspectral cube by applying a Linear Mixing Model to spectral signatures of water, soil and vegetation obtained from the USGS Spectral Library [5]. Normalisation was then applied to transform the values into a false colour map.

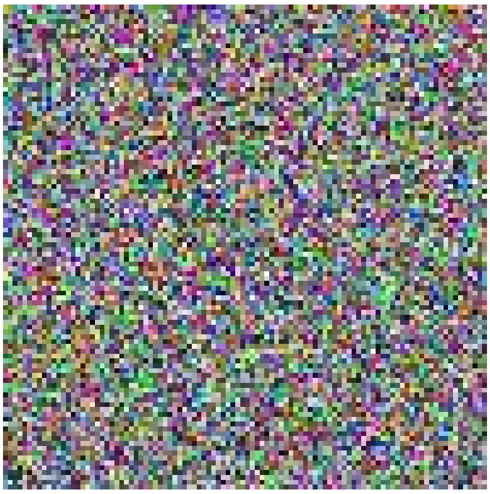


Figure 2.1: Synthetically Generated HSI visualised in RGB

2.2 Image Degradation

The following section describes techniques from literature that can be used to degrade the quality of HSI data with the aim to create datasets of varying image quality to allow algorithm comparison.

2.2.1 Image Downsampling

Image downsampling is the process of reducing the resolution of an image, often by decreasing the number of pixels it contains. This is done by scaling down the image dimensions (spatial or spectral) using a specific scale factor (SF) that determines which data from the matrix is thrown away. For example, if an image is spatially downsampled by a scale factor of 2, the width and height of the image are each halved therefore reducing the overall number of pixels by a factor of four. Downsampling is regularly used to reduce the file size of images and to improve processing speed, as HSI data is often large in size the utilisation of downsampling can aid in managing this to enable more efficient analysis and storage.

However, downsampling HSI introduces important trade-offs. In practice, it is not a simple reduction in spatial resolution through subsampling, but the process often involves spatial and/or spectral aggregation. This process can blur fine spatial structures and mix distinct spectral signatures within a single pixel. As a result, small or subtle anomalies may become less distinguishable, and the spectral integrity of materials can be compromised, leading to reduced accuracy in target detection and classification [6].

The example image downsampled at various SF is given below:

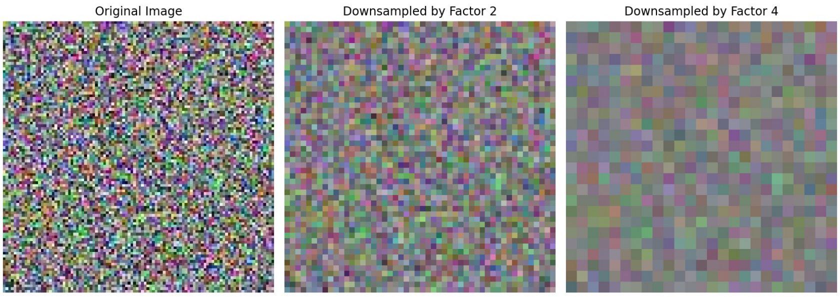


Figure 2.2: Example image downsampled by SF2 and 4

Spatial Downsampling

Spatial downsampling is the process of reducing the spatial resolution of an image. There are multiple techniques that can be used to achieve this, each of the techniques offering different benefits when applied.

The first method of note is the Mean Filtering method. In this method (also known as pixel averaging) a group of adjacent pixels in a square block are averaged to produce a single pixel in the resulting downsampled image. This algorithm is quick and efficient however lacks the smoothness that can be achieved when using other downsampling methods [7].

Subsampling is a sampling technique that involves selecting and keeping every other pixel and discarding the rest. For example, when downsampling by scale factor 2, every second pixel in each row and column is kept whilst the others are discarded. As in most downsampling techniques, when the sampling rate used in this downsampling process gets too low, details of an image can be lost beyond recovery and the image can appear pixelated.

To combat the pixelation effect that can occur in subsampling, a Gaussian filter technique can be employed prior to the downsampling process. In this technique, the Gaussian filter is applied to the image followed immediately by the subsampling of the image, creating a smaller, lower resolution final image. This final image is smoother than an image downsampled using subsampling [8].

All three of the spatial downsampling techniques discussed have similar disadvantages. Firstly, when implementing downsampling techniques on imagery, there will be a reduction in spatial resolution and therefore a loss of smaller details within an image. This is actually a desirable effect for this study as it allows the simulation of low quality data, however this is often unwanted in real-world use cases. Another disadvantage of the use of downsampling can be seen when applying both traditional and ML algorithms to downsampled data. Models that are trained on high resolution data can often suffer performance drops when applied to lower resolution data as there is less information present. This can lead to inaccurate interpretations of results [9].

Spatial vs Spectral Downsampling

It is worth noting that within the context of this thesis, downsampling refers to the downsampling of images in the spatial domain, not the spectral domain. For this study, downsampling was used to simulate lower resolution imagery conditions due to the process causing the loss of image detail. This reduction in spatial resolution leads to the loss of high-frequency components in the image, as fine details are often smoothed out. While this also impacts the spectral domain by reducing the ability to distinguish finer spectral details, the primary focus of this study is on spatial degradation as it is the more significant factor affecting image quality because it directly removes fine structural details that are essential for interpreting an image. The potential spectral loss is secondary in the context of this thesis, as the main concern is how reduced spatial resolution affects the performance of both conventional and ML models. The use of downsampling enabled a consistent and repeatable degradation of images by given scale factors to evaluate which method tends to perform better with lower resolution input images. However, for completeness, a summary of common techniques for spectral downsampling is included within this literature review.

Spectral Downsampling

Spectral downsampling is a process used in both HSI and Multispectral Imaging (MSI) that aims to reduce the number of spectral bands within in a dataset. Whilst the detailed spectral information collected in HSIs is valuable for many tasks it also results in large data sizes that are computationally expensive to process and store. Spectral downsampling can aid in managing this to enable more efficient data processing.

Principal Component Analysis

There is lots of research detailing the use of Principal Component Analysis (PCA) [10] on HSI as a technique to perform dimensionality reduction. HSIs contain many spectral bands which often include noisy and highly correlated data. Having such a large amount of bands increases the computational time required and storage necessary to process this imagery. PCA is used as a method to reduce the dimensionality of this data whilst retaining critical information that is necessary for models to perform accurate classification.

PCA transforms the original HSI into a new set of components called the principal components. These are linear combinations of the original spectral bands that are ordered by the amount of variance that is extracted from the data. The first principal components capture the most significant variance and are then considered the most important features within the image. PCA has been applied to HSI to reduce the number of spectral bands present in an image. This is done by calculating the sample covariance matrix of the spectral bands and then performing eigenvalue decomposition on these. The eigenvectors corresponding to the largest eigenvalues are then identified as the principal components in the format of a feature vector which can be cast along the principal component axis.

By reducing the dimensionality of the HSI data, PCA helps to decrease the computational power required in later processing and has been shown to improve the efficiency of various HSI classification algorithms [11]. PCA has also been shown to reduce noise and improve model accuracy due to it focusing on the most important features within an image. This study required images to be degraded strongly so that anomaly detection methods can be properly tested. However, PCA is not suitable for this purpose because it mainly preserves the most common patterns in the data. As a result, it struggles to significantly degrade areas that are very common in the image (such as background regions), limiting its usefulness here.

$$\boldsymbol{X}' = \boldsymbol{X}\boldsymbol{W} \tag{2.1}$$

Where:

- $\boldsymbol{X}$ is the original data matrix of dimension $n \times d$, where n is the number of samples and d is the number of features,
- $\boldsymbol{W}$ is the matrix of eigenvectors (the principal components) of the sample covariance matrix of $\boldsymbol{X}$,
- $\boldsymbol{X}'$ is the transformed data matrix in the new PCA basis.

$$\boldsymbol{\Sigma} = \frac{1}{n-1}\boldsymbol{X}^T\boldsymbol{X} \tag{2.2}$$

Where:

- $\boldsymbol{\Sigma}$ is the covariance matrix of the data,
- $\boldsymbol{X}^T$ is the transpose of the data matrix $\boldsymbol{X}$.
- $n-1$ is used to produce an unbiased estimate of the true population variance from a finite dataset

$$\boldsymbol{\Sigma}\boldsymbol{W} = \boldsymbol{W}\boldsymbol{\Lambda} \tag{2.3}$$

Where:

- $\boldsymbol{\Lambda}$ is the diagonal matrix of eigenvalues corresponding to the eigenvectors in $\boldsymbol{W}$,
- $\boldsymbol{W}$ is the matrix of eigenvectors (principal components).

In the figure below, PCA has been applied to the synthetic example image. The example image (original) is followed by grayscale images representing the first five eigenvectors. PC 1 captures the most significant patterns/features, whereas PC 5 captures much finer, less ”important” details.

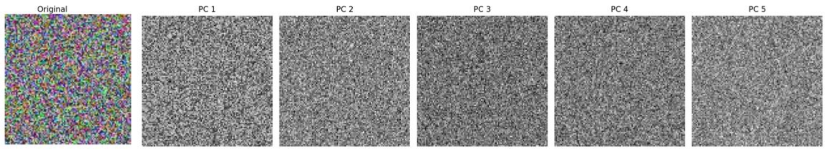


Figure 2.3: PCA applied to example image

2.2.2 Band Clustering

Similar to PCA, band clustering is a technique used in HSI to reduce the number of spectral bands within an image. By grouping similar bands together, this approach decreases the image quality in the spectral domain.

The process of band clustering involves grouping bands based on how similar their spectral profiles are. This similarity calculation is often done using statistical measures such as linear prediction or orthogonal subspace projection [12]. Following this, clustering algorithms such as k-means or hierarchical clustering are employed to group similar bands together [13]. After the clustering there are two paths that may be taken. Firstly, the bands within clusters are combined. The simplest method to achieve this is by averaging pixel values across the bands to create one singular final band. The second method is to select just one representative band from each cluster.

However, band clustering does have negative effects on image quality. It causes the loss of spectral detail from an image and reduces the spectral resolution, this can cause lower classification performance of models as it can be difficult to now distinguish between spectrally similar bands. However, similar to PCA, band clustering can help reduce noise within an image due to the removal of random variations within similar bands [14].

This technique can be useful if there is a requirement to represent data from sensors that collect fewer bands and can therefore be used to extrapolate model results to future-proof concept demonstrators when designing sensors or processing techniques. However, as the band clustering is performed in the spectral domain it is of limited use within this thesis.

Following investigation into the techniques described above, the final degradation method used in this thesis was spatial downsampling with a Gaussian filter. While this is slightly slower than simple subsampling, it produces more realistic results by reducing aliasing effects [15]. It was chosen as a good balance between simplicity and quality, while the other methods were included for completeness.

2.3 Super Resolution

In order to evaluate the impact of image quality on model performance, both image enhancement and degradation should be considered. To enhance images, a common technique used is super resolution. Super resolution is an image processing technique that enhances the resolution of an image by increasing the number of pixels present and aiming to improve its visual quality. This process generally involves generating a high resolution (HR) image from low resolution (LR) images. The technique is alleged to enable more accurate detection of small features and complex details that may be missed in low resolution images and can also make images appear sharper and more defined. In the figure below Super resolution was applied to the example HSI. In image SR SF2, the image was enhanced using a Scale Factor of 2. This means the output has twice the pixel density of the original. For SR SF4, the image was enhanced using a Scale Factor of 4, resulting in four times the pixel density of the original image.

Figure 2.4: Example image enhanced using super resolution with SF 2 and 4

Three techniques were investigated and implemented when researching super resolution within this study following from their identification as relevant in the literature review that follows this chapter.

2.3.1 Technique 1: Deep Image Prior

Deep Image Prior (DIP) [16] is an image processing technique that utilises CNNs to perform super resolution. DIP states that a randomly initialised CNN can capture the internal structure of an image without the requirement for large amounts of training data and can then apply this knowledge to generate further images with similar statistical properties to the original.

Firstly, a CNN is randomly initialised and a random noise vector is fed into the model and an output image is generated. This output image is then compared to the target image and the loss function of the model calculates the statistical difference between the output and target images. Following this calculation, the CNN parameters (weights) are iteratively updated in order to minimise this difference causing the output image to become increasingly similar to the provided target image. DIP utilises early stopping in order to prevent the overfitting of the model to the target image.

For use within super resolution, the problem is framed to find the HR image that, when downsampled, is the same as the input LR image. Within this, there are an almost infinite number of HR images that following downsampling reduce to the original LR image. DIP combats this through regularisation, where the CNN's architectural bias and training dynamics select the most plausible HR image.

Advantages of the DIP technique are summarised below:

- A key advantage of DIP is the lack of a requirement for large scale training datasets. The CNN used performs optimisation using random initialisation to exploit the internal statistics of an input image. This is extremely useful in scenarios where labelled datasets are unavailable or input image types are rare [16].
- DIP is easy to implement and is designed to perform multiple image processing tasks within the same framework. This makes the technique versatile and generalisable across multiple real-world scenarios [16].
- The use of early stopping in DIP reduces the possibility that the network will overfit to specific input images, this maintains a balance between fitting to the known input data and preserving an overall image structure. This approach enables the model to generalise well to different kinds of noise and image artefacts [16].

Despite the usability of DIP, it has disadvantages that are summarised here:

- DIP requires individually optimising a CNN for every image. This can be computationally expensive and time-consuming as the network is trained from scratch and the higher the resolution of an image, the longer the process takes. DIP is therefore not suitable for scenarios that require the quick processing of a large number of images [16].

- Due to DIPs reliance on traditional CNN network structure, it can struggle when performing processing on images that contain complex blur or noise. The network may not be able to generalise well when blur and noise are not random and it may bias overall towards smooth and structured image outputs [16].
- In contrast to the above advantages, there is also a disadvantage of utilising early stopping. The network depends on the perfect timing of early stopping and finding this optimal point can be particularly challenging especially for real-world use cases. Incorrect early stopping can lead to model underfitting, leading to output images still being blurred and containing artefacts [16], [17].

In conclusion, DIP is an effective algorithm that utilises the properties of CNNs to perform super resolution on HSIs without the need for large quantities of training data. Whilst the technique offers clear advantages in terms of its flexibility and ease of use, challenges such as computational cost, over- or under- fitting, and scalability must be balanced when applying to real-world use cases.

2.3.2 Technique 2: KernelGAN

KernelGAN [18] is a technique for single image super resolution that is focused on estimating the downscaling kernel that is used to create a LR image from a HR one. KernelGAN aims to learn this kernel directly from the LR image, without the need for additional training data. Traditional super resolution methods often assume a fixed downscaling process, e.g. bicubic downscaling [19], however this often does not match the real-world scenario. KernelGAN addresses this by estimating the actual downscaling kernel, which is usually unknown. The technique uses a generative adversarial network (GAN) containing two competing networks, a generator and a discriminator.

The generator creates HR patches that, when downscaled with the learned kernel, should resemble LR patches from the initial input image. The discriminator then attempts to distinguish between these generated patches and the actual LR patches. Through this process, KernelGAN learns the most likely downscaling kernel that was used when creating the LR image. Once the downscaling kernel has been learned, it can be used to better model the degradation process during super resolution. By applying this kernel the super resolution process becomes more accurate to a real-world scenario. KernelGAN improved upon previous super resolution results as it enables a more custom and accurate reconstruction of HR images by adapting to the unique features of an input image.

2.3.3 Technique 3: Zero-Shot Super-Resolution

Zero-Shot Super-Resolution using Deep Internal Learning (ZSSR) [20] presents an approach to super resolution that also operates without the need for additional training images. Unlike traditional deep learning methods that rely on large datasets of LR and HR image pairs for training, this method utilises the internal statistics of a single image to complete super resolution.

The basis of the technique is "Deep Internal Learning" which is based on the idea that patches within an image recur within the image across different scales. This method utilises the internal symmetry to train a deep neural network (DNN) using only information from within the input image itself. The model extracts patches from the image and trains the DNN to map these patches to their HR counterparts. The method demonstrates comparative results to traditional super resolution techniques that rely on additional training data, suggesting that ZSSR can achieve well performing super resolution across a wide variety of images. This approach is useful when HR images may be unavailable or when images may contain unique characteristics that general models can not capture, such as fine structures or repetative details.

2.3.4 Super Resolution Drawbacks

Whilst super resolution techniques have multiple applications within image processing, the drawbacks of using such algorithms should be identified and acknowledged [21].

- Computational Cost: The use of super resolution comes with both computational and storage costs. The techniques, especially those involving the use of neural networks, require significant computational resource to run and often require expensive GPUs to perform training and inference in reasonable time due to the sheer amount of parameters that need refining. This cost and time complexity lead to the conclusion that it may be ineffective to use super resolution techniques for real-time applications.

  Related to this specific drawback is the fact that high resolution images require more memory and storage space. When super resolution is applied to data, the storage requirements can significantly increase, meaning the technique may not be suitable for us when systems have limited available storage capacity.

- Artefact Introduction: When super resolution models are applied to datasets, they can artificially introduce artefacts and noise such as blurring. Such artefacts result from the model interpolating details that were not originally present in the input low resolution image. This can lead to unnatural looking images and can occur due to models overfitting to specific patterns within images.

  As super resolution techniques generate pixels from scratch with the aim to fill in missing details, there is an inherent risk that the enhanced output does not authentically represent what a real scene would look like. This can be extremely problematic for applications that require high accuracy and explainability, for example, medical imaging or defence applications.

- Information Recovery: Super resolution techniques can only enhance an image to a certain extent as they rely on the information that is already present in the low resolution image. When the low resolution input is severely degraded or lacking of important details, super resolution techniques cannot recover or create this lacking detailed information. This limitation makes the techniques less effective for situations in which the input is of extremely low quality.

- Generalisation: Super resolution methods are often trained on very specific imagery (e.g. HSI). When the techniques are applied to unseen data that differs from the training data, the models can struggle to extrapolate results to generate correct predictions leading to sub optimal results if utilising trained models.

  To combat this, one can retrain or fine-tune models, however this can be impractical as this process is time consuming and computationally complex. If there is very limited availability of the specific data type required, then super resolution techniques may not be best placed to provide enhancement on these inference images.

Super resolution can be an important technique to enhance low resolution and degraded images however consideration must be given to ensure that this is an appropriate and well performing technique for the specific use case it is being applied to.

2.3.5 Summary

The three methods were investigated to enable a comparative analysis of different super resolution approaches before selecting the most suitable technique for this experiment. Following analysis of the techniques, DIP was selected for use in this experiment due to its straightforward implementation and its ability to be applied to multiple use cases.

2.4 Conventional Processing Methods for Hyperspectral Anomaly Detection

2.4.1 Summary

For this study, it was important to understand both the strengths and limitations of current, conventional HSI anomaly detection techniques. The use of these techniques in this study directly enables a benchmarking experiment to be conducted which allows a comparison of effectiveness of both traditional and newer, ML approaches. Traditional methods also aid in gathering a solid foundational understanding of the basic principles of HSI anomaly detection that can later be applied to new algorithms. Conventional methods have been shown to perform reliably across a variety of use cases and are often much simpler to integrate and apply than ML techniques. Four popular conventional techniques used in HSI anomaly detection are detailed below to enable a comprehensive comparison of techniques.

2.4.2 Adaptive Cosine Estimator

Adaptive Cosine Estimator (ACE) is a detection algorithm that is frequently used with HSI data to identify and locate specific target materials or spectral signatures within a scene [22].

To perform ACE, an input image and a specific, known target spectral signature are required. Following these inputs, the aim of ACE is to calculate how well the input target signature matches a current pixel signature whilst also taking into account the background variability of a scene based upon the squared cosine of the angle between the test pixel spectrum and the target spectrum [23]. Following this calculation, the algorithm returns an “ACE score” for each pixel in the image. The higher this score, the stronger the match between the target spectrum and the pixels’ spectra, suggesting that the given pixel contains the target material.

$$\mathrm{ACE}(\boldsymbol{x}) = \frac{\left(\boldsymbol{s}^T \boldsymbol{\Sigma}_b^{-1} \boldsymbol{x}\right)^2}{\left(\boldsymbol{s}^T \boldsymbol{\Sigma}_b^{-1} \boldsymbol{s}\right)\left(\boldsymbol{x}^T \boldsymbol{\Sigma}_b^{-1} \boldsymbol{x}\right)} \tag{2.4}$$

Where:

- $\boldsymbol{s}$ is the target signature,
- $\boldsymbol{x}$ is the observed pixel spectrum,
- $\boldsymbol{\Sigma}_b$ is the covariance matrix of the background pixels.

The use of ACE is particularly effective for known target detection when image backgrounds are noisy and/or cluttered. As ACE takes into account image background, it can be an effective technique to distinguish between a desired target and the background clutter and presents a robust and tested detection capability in variable scenarios. To counter this point, ACE performs best when background statistics are accurate. The algorithm heavily depends on an accurate estimation of the background covariance matrix, if this is incorrectly calculated, then results of ACE will be poor [24].

Whilst ACE is an extremely useful algorithm for performing known target detection in HSI, it is unsuitable for further comparative study in the context of this thesis due to the desire to focus on unknown material anomaly detection. The requirement of ACE to have a known input signature denotes that if there are multiple targets within an image, the technique must be repeatedly run. Whilst multiple runs are not computationally expensive when running ACE, the lack of known target inputs led to the decision that another technique may be more beneficial within the remit of this thesis. The ACE algorithm was applied to the synthetic image to generate the figure below. Here "ACE - Endmembers" 1, 2, and 3 are detection maps with colours representing the likelihood that a specific signature is present at that pixel.

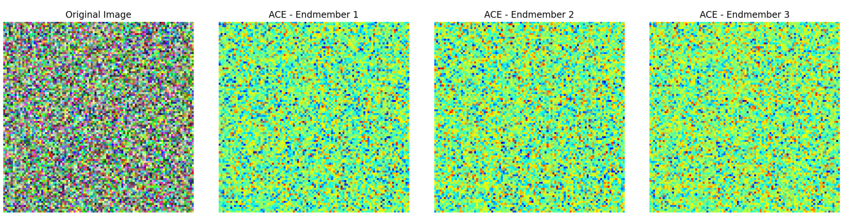


Figure 2.5: ACE applied to example image

2.4.3 Cluster-Based Anomaly Detection

Cluster-Based Anomaly Detection (CBAD) is a technique that utilises unsupervised clustering techniques to perform anomaly detection in HSI [25]. Initially an input image is fed into a clustering algorithm, the most common algorithms being:

- K-Means Clustering: K-Means is a form of partition clustering used to divide data into a pre-defined number of clusters that are determined by data similarity. A number K of random centroids are initialised, each data point within the image is then assigned to its nearest centroid, forming K clusters. The centroid locations are then iteratively updated to be the mean of the assigned points, this process continues until the centroid locations plateau and K clusters are finalised [26]. Outlying points are determined by their distance from the centroid location.

$$J = \sum_{i=1}^{K} \sum_{\boldsymbol{x}_j \in C_i} \|\boldsymbol{x}_j - \boldsymbol{\mu}_i\|^2 \tag{2.5}$$

  Where:

  - K is the number of clusters,
  - C_i is the set of points assigned to cluster i,
  - $\boldsymbol{\mu}_i$ is the centroid of cluster i,
  - $\boldsymbol{x}_j$ represents a data point,
  - $\|\boldsymbol{x}_j - \boldsymbol{\mu}_i\|^2$ is the squared Euclidean distance between $\boldsymbol{x}_j$ and $\boldsymbol{\mu}_i$.

  The figure below displays the synthetic HSI next to the results of applying a K-Means clustering algorithm. Every pixel in the original image has been assigned to a cluster based on similarity, reducing thousands of colours in the synthetic image down to just a few shades (black, grey, and white).

Figure 2.6: K-Means applied to example image

- Density-Based Spatial Clustering of Applications with Noise (DBSCAN): DBSCAN is a clustering technique that does not require the number of clusters to be defined prior to performing the technique. This is done by assigning all points to the categories of: core points, border points or noise points. Clusters are expanded from core points based on the density of points within a distance ϵ and a minimum number of points that are required to form a cluster. Points that do not fit into any clusters are labelled as the outlier noise points [27].

$$N_{\varepsilon}(\boldsymbol{x}_i) = \{\boldsymbol{x}_j \in \boldsymbol{D} \mid \|\boldsymbol{x}_j - \boldsymbol{x}_i\|_2 \leq \varepsilon\} \tag{2.6}$$

$$\text{Core}(\boldsymbol{x}_i) = \begin{cases} 1, & |N_{\varepsilon}(\boldsymbol{x}_i)| \geq \text{MinPts} \\ 0, & \text{otherwise} \end{cases} \tag{2.7}$$

$$\boldsymbol{x}_j \text{ is directly density reachable from } \boldsymbol{x}_i \iff \boldsymbol{x}_j \in N_{\varepsilon}(\boldsymbol{x}_i) \text{ and } \text{Core}(\boldsymbol{x}_i) = 1 \tag{2.8}$$

$$\boldsymbol{x}_i \text{ and } \boldsymbol{x}_j \text{ are density connected if there exists a chain } \boldsymbol{x}_{i_1}, \ldots, \boldsymbol{x}_{i_m} \text{ such that } \boldsymbol{x}_{i_1} = \boldsymbol{x}_i,\ \boldsymbol{x}_{i_m} = \boldsymbol{x}_j, \text{ and each } \boldsymbol{x}_{i_{k+1}} \text{ is directly density reachable from } \boldsymbol{x}_{i_k} \tag{2.9}$$

$$C_k = \{\boldsymbol{x}_i \in \boldsymbol{D} \mid \boldsymbol{x}_i \text{ is density connected to some core point in cluster } k\} \tag{2.10}$$

$$N = \{\boldsymbol{x}_i \in \boldsymbol{D} \mid \boldsymbol{x}_i \notin \bigcup_{k=1}^{K} C_k\} \tag{2.11}$$

DBSCAN has been applied to HSI for band selection [28] and a demonstration of the clustering applied to the example synthetic image is given in the figure below. As DBSCAN is a density-based algorithm if the data is too densely packed then the algorithm may estimate that all points belong to the same cluster, hence the plain second image.

Figure 2.7: DBSCAN applied to example image, only 1 cluster identified

- Gaussian Mixture Model (GMM): GMM is a form of "Fuzzy Clustering", meaning instead of every point being assigned to a cluster, a probability of membership to a cluster is assigned to each point instead. It is assumed that input data is created from a mixture of multiple Gaussian distributions, of which the parameters are unknown. Each of these distributions represents a cluster within the data. The GMM aims to learn the parameters of these distributions that best represent the original dataset and use these to assign clusters to points. The Gaussian parameters are estimated from an iterative expectation-maximisation algorithm [29]. In the E-step, the algorithm calculates the probability of each data point belonging to each cluster, and in the M-step, it updates the parameters (means, data covariances) to maximise the likelihood of the data and repeats this process until data convergence. [30]. Within the context of anomaly detection, the GMM is applied to unseen data and the regions of the images that have low likelihoods are flagged as anomalies. Within the context of HSI, GMMs have been used to perform segmentation and clustering within remote sensing images [31].

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^{N} \log \sum_{k=1}^{K} \pi_k \mathcal{N}(\boldsymbol{x}_i \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \tag{2.12}$$

Where:

- $\mathcal{L}(\boldsymbol{\theta})$ is the log-likelihood function,
- N is the number of data points,
- K is the number of Gaussian components,
- π_k is the weight (mixing coefficient) of the k-th Gaussian component,
- $\mathcal{N}(\boldsymbol{x}_i \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ is the probability density function of a Gaussian distribution with mean $\boldsymbol{\mu}_k$ and covariance $\boldsymbol{\Sigma}_k$,
- $\boldsymbol{x}_i$ represents a data point.

In the following figure, GMM has split the example image into a grayscale representation, grouping pixels based on their probability of belonging to a specific Gaussian distribution.

Figure 2.8: GMM applied to example image

On of the key benefits of the use of clustering techniques to perform anomaly detection is its use of unsupervised learning. Unlike the ML techniques discussed in section 2.3 no labelled data is required for these techniques to perform accurately which is especially useful in the field of HSI where limited labelled examples are readily available. Clustering techniques are also both flexible and scalable due to their ability to perform on different datasets containing clusters that are a variety of shapes and sizes.

However, when compared to other anomaly detection techniques, clustering can be particularly sensitive to noisy and cluttered images which can lead to background being incorrectly classified as anomalous. Multiple clustering algorithms also require parameter setting early on in the implementation process, if these parameters are incorrectly initialised then clusters may be incorrectly assigned further down the pipeline.

2.4.4 Hotelling's T^2 Statistic

Hotelling's T^2 Statistic is a statistical test used to identify anomalies within high-dimensional data [32]. The aim is to determine whether a given point is anomalous based on its deviation from the multivariate mean of the image whilst also taking into account the covariance of the data.

To perform this statistic, firstly the covariance matrix of the image is calculated along with the multivariate mean. Following these calculations, Hotelling's T^2 Statistic is computed for each data point as follows:

$$T^2 = n(\bar{\boldsymbol{x}} - \boldsymbol{\mu})^T \boldsymbol{S}^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu}) \tag{2.13}$$

Where:

- n is the sample size (number of pixels used to model the background of the data),
- $\bar{\boldsymbol{x}}$ is the data point mean vector,
- $\boldsymbol{\mu}$ is the mean vector,
- $\boldsymbol{S}$ is the sample covariance matrix.

The computed value is then compared to an identified threshold value (determined from confidence level and image dimensions). If the calculated T^2 statistic is larger than the threshold value, that data point is considered an anomaly as it is significantly different from the calculated multivariate mean [33].

This statistic is a key tool to aid in identifying changes in the means between multiple data points. It enables the comparison of multiple data points in an image at once to avoid having to run separate tests for each point, saving time and compute resource when performing anomaly detection.

The T^2 statistic has been applied to HSI data to perform "Hierarchical Clustering" [32] with the aim to perform land cover classification and whilst this technique demonstrated successful results, its use within HSI anomaly detection has had limited exploration in existing literature. As a result, it was considered less suitable for the objectives of this study and was not selected for further use. The following figure displays the example image and it's equivalent map of the T^2 statistic for every pixel. The brighter pixels indicate a higher T^2 value, suggesting those specific pixels are statistical outliers compared to the rest of the noise distribution. The darker areas more closely align with the mean of the overall image data.

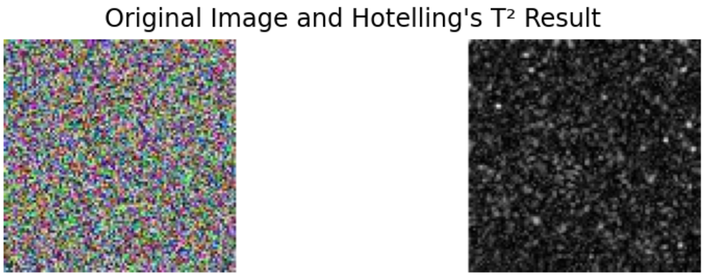


Figure 2.9: Hotelling's T^2 Statistic applied to example image

2.4.5 Reed-Xiaoli Detector

The Reed-Xiaoli (RX) Detector is one of the most commonly used HSI anomaly detection techniques. RX is a statistical technique that starts by modelling the background of an image through the calculation of the mean spectra of an image (the background) and the covariance matrix of this background to gain an understanding of the variability of spectra across the scene. Following these calculations, the Mahalanobis Distance is computed for each pixel in the image to obtain a value for the distance between the pixels' spectral signature and the previously obtained mean background values. The Mahalanobis Distance is calculated by:

$$D_M(\boldsymbol{x}, \boldsymbol{\mu}) = \sqrt{(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{S}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})} \tag{2.14}$$

Where:

- $\boldsymbol{x}$ is the vector of the data point,
- $\boldsymbol{\mu}$ is the mean vector,
- $\boldsymbol{S}^{-1}$ is the inverse covariance matrix of the background data.

The RX score for each pixel is then calculated, a higher RX score indicates that the pixels spectral profile is significantly different from the background suggesting it may be an anomaly [34][35].

The RX anomaly detection score is given by:

$$RX(\boldsymbol{x}) = D_M^2(\boldsymbol{x}, \bar{\boldsymbol{x}}) \tag{2.15}$$

Where:

- $\boldsymbol{x}$ is the feature vector of the pixel being analysed,
- $\bar{\boldsymbol{x}}$ is the sample mean vector,
- $D_M(\boldsymbol{x}, \bar{\boldsymbol{x}})$ is the Mahalanobis distance between $\boldsymbol{x}$ and the mean.

The process of performing RX detection is time and resource efficient however, it is noted in Küçük 2015 [36] that when using local area statistics, the detector infrequently only identifies the centre of anomalous regions and may miss out on larger anomalies. The technique also has the potential to identify sub-pixel targets that may be anomalous which could be useful in remote sensing applications.

The RX detector is well documented, easy to implement, and frequently used. This made it an optimal choice for use within this study to enable a comparison between conventional techniques and ML approaches to HSI anomaly detection.

Hotelling's T^2 statistic, the Mahalanobis distance, and RX detection are all related techniques used to quantify how far a data point deviates from a given distribution. T^2 is a scaled version of the Mahalanobis distance that accounts for sample size, while RX detection directly applies the Mahalanobis distance to HSI for anomaly detection by identifying pixels with high deviations from the background statistics.

The figure below shows the original, synthetic image and the output when that image was run through an RX detector. Here, the bright spots represent pixels that are the most spectrally different from the surrounding background mean.

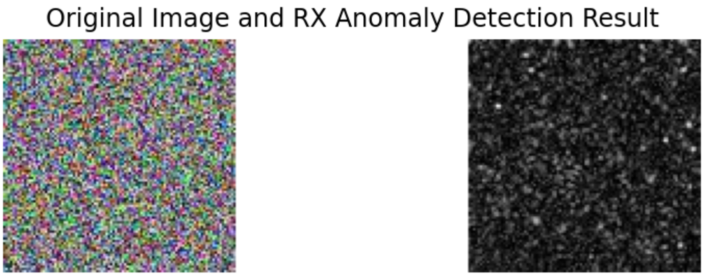


Figure 2.10: RX applied to example image

2.5 Machine Learning Methods for Hyperspectral Anomaly Detection

2.5.1 Summary

In order to fully compare the results of both conventional and ML techniques, the following section will discuss and detail prominent current ML HSI techniques and the challenges faced when implementing these approaches.

2.5.2 Autoencoders

As discussed in [1], Autoencoders (AEs) are a form of unsupervised artificial neural networks that are often utilised with the aim of reconstructing input data to minimise the difference between input and generated output data. The AE aims to learn the underlying structure of unlabelled input data through an encoder and decoder structure as shown below:

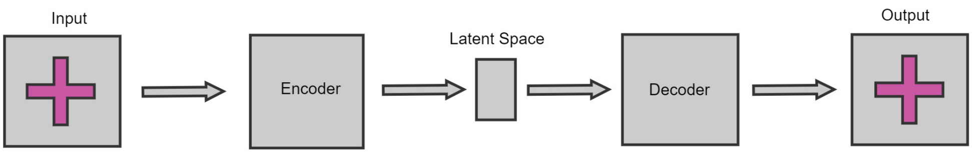


Figure 2.11: Autoencoder Structure

The use of AEs within the field of HSI is well documented as the use of AEs can reduce data dimensionality whilst also preserving essential features [37]. Examples of use cases are detailed below:

- Noise Reduction: AEs have been used alongside HSI data to reduce noise within images, this is done through the learning of a compressed representation of the underlying data structure whilst filtering out image artefacts [37].
- Pre-training: Due to the complexity of HSI, the pre-training and fine-tuning of ML models can save significant amounts of time and compute resource. The use of an AE to pre-train neural network weights allows a user to fine-tune a larger model using a smaller subset of labelled data. This is especially beneficial when labelled data is extremely limited [37].
- Data Augmentation: When training ML models, a common technique employed to avoid overfitting is data augmentation (the process of creating new data samples from existing images). In HSI, AEs can be utilised to generate synthetic images that can increase the diversity of a training dataset with the aim to improve the performance of anomaly detection models [37].

Despite the large amount of research on AEs, there remain many unanswered questions especially when applied to niche datasets such as HSI. One of the key factors behind this research gap is the lack of available labelled HSI benchmarking test datasets. It should also be noted that many HSI based AEs only model either spectral or spatial features in isolation. This separation can lead to the loss of important joint spectral–spatial information during encoding. This directly leads to the inability to accurately reconstruct HSI data during the later decoding process. It is likely that future research in the field will focus on the fusion of both spectral and spatial data.

2.5.3 2D Convolutional Neural Networks

CNNs aim to learn features from input data using a variety of filter operations (convolutions) [38]. As previously introduced in Section 2.3.1 in the context of DIP, these operations form the foundation of CNN architectures. In the context of image processing, this is accomplished through a CNN constructed from various layers listed below:

- Convolutional Layers: The convolutional layers within a CNN apply learnable kernels (filters) to input image data. These filters perform a sliding window operation to calculate the dot product between the defined kernel and the input image. The result of this dot product is a feature map that represents the different details within the input image.
- Activation Functions: Following the convolutional layer, the activation function is then

applied with the aim of introducing non-linearity to the model. This enables the CNN to learn more complex features within the images.

- Pooling Layers: Pooling layers are introduced to the CNN to reduce the spatial dimensions of output feature maps. This operations decreases the number of hyper-parameters within a layer in order to control efficiency. This process also can aid in the prevention of model overfitting. "Max Pooling" and "Average Pooling" are two common pooling operations that select the maximum value and the average values within an image patch respectively.

- Dense Layers: Dense layers, also known as Fully-Connected layers are used to connect layers of neurons within a network together. These layers are used to ensure that information effectively flows throughout a network and to ensure the information extracted from the output of feature maps is taken into account when the CNN is performing classification or regression.

- Dropout Layers: The use of dropout layers in CNN is optional, they are implemented as a technique to prevent overfitting by randomly setting input values to a layer as 0, this forces a network to learn more robustly due to the variation in an image.

Following construction, the network is trained using backpropagation [39]. The process involves a network feeding error rates back through previous layers to update the model weights.

Firstly, a forward pass of data is conducted through the model, immediately followed by an error / loss calculation. This loss value is then propagated back through the network and the amount that each weight and bias value contributed to this loss is computed. The weights of the network are then updated using an optimisation algorithm (often gradient descent) with the aim to reduce the overall loss value. This process is performed on an iterative basis until convergence is reached. The whole process acts as an approach to reduce error in model predictions [40]. The calculation of this error requires a known result and label for every piece of input training data which can be a drawback of utilising these models.

Within a HSI context, multiple papers have explored the use of 2D CNNs and how they can apply to this data. CNNs are regularly used for landscape classification [41], object detection [42] and denoising of images [43] and have shown promising performance and adoption.

A drawback of utilising 2D CNNs on HSI data is that they can only take into account either the spatial data or the spectral data at one time. When processing the spatial data, the CNN can only consider a single spectral band and conversely when processing the spectral bands, only a single spatial pixel can be considered. A way to combat this is to utilise 3D CNN structures which are detailed in the following section.

2.5.4 3D Convolutional Neural Networks

A 3D CNN uses 3D kernels to perform 3D convolution operations in the same process as described for 2D CNNs, however a benefit of 3D networks is that they can extract both spatial and spectral features from an image simultaneously [44]. This feature extraction process enables a deeper understanding of the input data structures and is specifically useful in HSI where a combination of both spectral and spatial information can be used to detect and identify materials (and anomalies).

It has also been demonstrated that the integration of spatial and spectral features can achieve higher classification accuracy when compared to 2D methods, this can mean they are more effective when applied to real-world data. This increased performance is likely due to the networks being more capable of learning complex features and their relationships to each other within an image.

As most image processing (2D) CNNs are not intrinsically designed to handle HSI datacubes, the use of 3D CNNs can reduce the time taken and complexity of image preprocessing. There is no longer a need to perform dimensionality reduction on the HSI and so less manual data engineering process is required.

There are however, downsides to the use of 3D CNNs in HSI processing. As CNN size increases, the computational cost and complexity to perform the convolutional operations increases. The higher dimensionality of the input datacubes can lead to increased compute resource requirements. The issue of a lack of labelled HSI datasets also comes into play, a large amount of labelled data is required to effectively train 3D models and this process can be labour intensive, complex and is not often available in the open-source domain. If a model is trained utilising limited training data, then the model is more likely to overfit and be less generalisable to new and unseen scenes [45].

2.5.5 Vision Transformer Networks

Vision Transformers (ViTs) are a newer type of neural network that were proposed in 2017 [46] and initially developed for use in Natural Language Processing but have been adapted for use within the field of computer vision [47].There are multiple key components that make up a ViT model:

- Patch Embedding: Before being fed into the main transformer module of a ViT, an input image must be divided up into a number of patches (often 16x16 pixels), these patches are then flattened into vector format and act as "tokens" that are fed to the transformer model.
- Positional Embedding: In order to ensure that spatial features are preserved during the above patching process, positional embeddings are assigned to each image patch. This ensures that the transformer model can differentiate between all of the patches based upon their location within the input image.
- Attention: Self-attention (attention) is a process by which the transformer model can weigh up the importance of different features contained within an image relative to each other. Attention allows the model to focus on the important features and patterns across the input image.
- Transformer: The Transformer is made up of multiple layers of attention and feed-forward neural networks. Each layer applies attention to the input image patches which enables the model to learn dependencies and contextual information about the patches.
- Classification: After being processed by the transformer network, a classification layer may be present to assign a class to each image patch to produce a final prediction based upon extracted features [48].

There are several key advantages to the use of ViTs in image processing due to their flexible and scalable nature. Perhaps the most prominent benefit of utilising ViTs is the ability to capture and understand the global context of an image and use this when making predictions. This is in contrast to CNNs which only utilise local context.

Despite the positives detailed, there are drawbacks to the use of ViTs. They can be computationally and resource demanding, more so than CNNs, due to their increased model size making them unpractical for use in limited compute environments. ViTs are also slower to perform both the training and inference process when compared to similar CNNs. Whilst ViTs outperform CNNs when capturing global image information, they can struggle when trying to understand local patterns, hybrid models that combine both CNNs and transformers have been proposed to combat this issue and are an area that is currently undergoing research [49]. To aid comparison, a simple example of both a CNN and ViT architectures are given below.

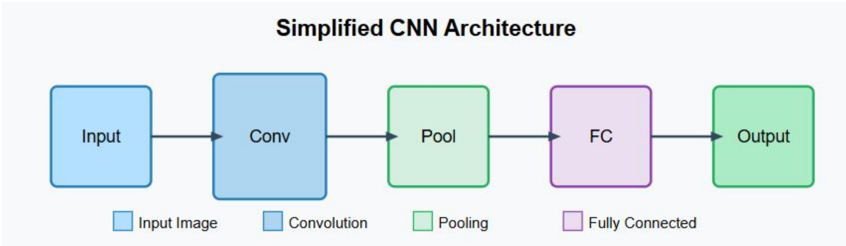


Figure 2.12: Simplified CNN Architecture

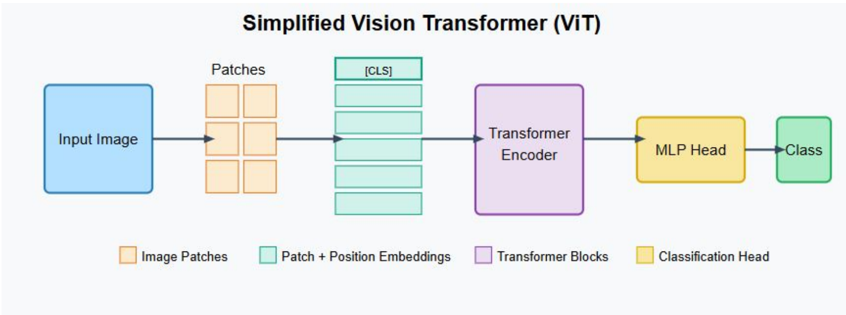


Figure 2.13: Simplified ViT Architecture

Perhaps the most commonly utilised ViT is the SWIN transformer which utilises "Shifted Windows" to build hierarchical feature maps of images for computer vision use cases [50]. In the context of this thesis, multiple studies have applied ViTs to the use case of anomaly detection using HSI to successful results [51], [52], one such example of this is given in the following section.

2.5.6 General Anomaly Enhancement Network with Random Masks for Hyperspectral Anomaly Detection (AETNet)

A key ML technique that is relevant and extremely important in the context of this thesis is AETNet [53]. First published in March 2023, AETNet is a machine learning model for hyperspectral image anomaly detection that aims to address the limitations of conventional methods by generalising across datasets without retraining. The model consists of three modules:

1. Random Masking Module: This creates irregularly shaped masks over training images

based on assumptions that anomalies (when present) are irregularly embedded, spectrally different from backgrounds, and occupy a small area within the image.

2. Image Reconstruction Module: This module employs a Convolutional Autoencoder (CAE) for spatial feature extraction and a Swin UNet with transformer architecture for feature representation learning.

3. Transfer Domain Search: To complete the full detection pipeline.

Following model construction, AETNet is trained to restore masked image sections by minimising the differences between masked inputs and expected outputs. It uses a loss function that considers spatial correlation and penalises structural differences with the aim to help enhance anomalies against backgrounds.

When assessed against baselines by the authors, the model outperformed classical methods even when anomalies were less well defined or backgrounds were cluttered. The model also showed increased generalisability and computational efficiency making it a good candidate for use within this study.

Chapter 3

Literature Review

3.1 Introduction

HSI has become an important data source for anomaly detection in remote sensing due to its high spectral resolution and ability to distinguish between materials with similar spatial characteristics. This capability has led to the use of HSI in applications such as environmental monitoring, agriculture, and defence. Over time, a range of techniques have been developed to try and solve the anomaly detection problem, including conventional signal processing methods and more recent ML approaches.

While considerable research has focused on improving detection accuracy, there has been less attention given to understanding how variations in image quality directly influence the overall performance of these developed techniques. This is particularly relevant in operational contexts where image quality may be constrained by sensor limitations or cost.

The primary goal of this research is therefore to investigate the effect of image quality on both conventional and ML anomaly detection methods when applied to HSI. This chapter reviews relevant literature, with the aim to gain an understanding of the current state of the field and identify capability gaps that influence the work presented in this thesis.

3.2 Scope and Search Strategy

While the primary focus of this research is on HSI, this review will also encompass literature on ML techniques that have been applied to non-HSI imagery, such as Panchromatic, RGB, and MSI as ML has been extensively applied to this imagery due to the availability of large datasets (unlike in HSI). The goal of this is to identify methods in similar fields that could be adapted and applied to HSI, therefore broadening the scope of effective techniques for anomaly detection in HSI

The literature search was conducted using Web of Science, covering publications up to March 2025. Keywords included “Hyperspectral”, “Remote Sensing”, “Anomaly Detection”, “Machine Learning”, and “Signal Processing”, combined using Boolean operators.

Studies were included if they were published in English, focused primarily on remote sensing imagery, and examined anomaly detection using either conventional or ML-based techniques.

3.3 Trends in Hyperspectral Anomaly Detection Research

When conducting a literature review of recent research, a clear distinction was demonstrated between studies focusing on conventional signal processing techniques, such as in [23], [25], [32], [36] and an increasing number developing ML methods, for example, [1], [37], [41], [44], [52].

In recent years there has been a notable shift towards ML based anomaly detection as demonstrated by an increase in publication counts relating to HSI anomaly detection using machine learning techniques, this is demonstrated in the publication count graph below. This trend is likely due to increased data availability, wider access to open-source ML model, and reduced need for specialised ML hardware when developing models. Despite this shift, conventional techniques still remain relevant due to their deterministic nature (compared to probabilistic ML) and relatively low computational requirements, as discussed in [54] and [55].

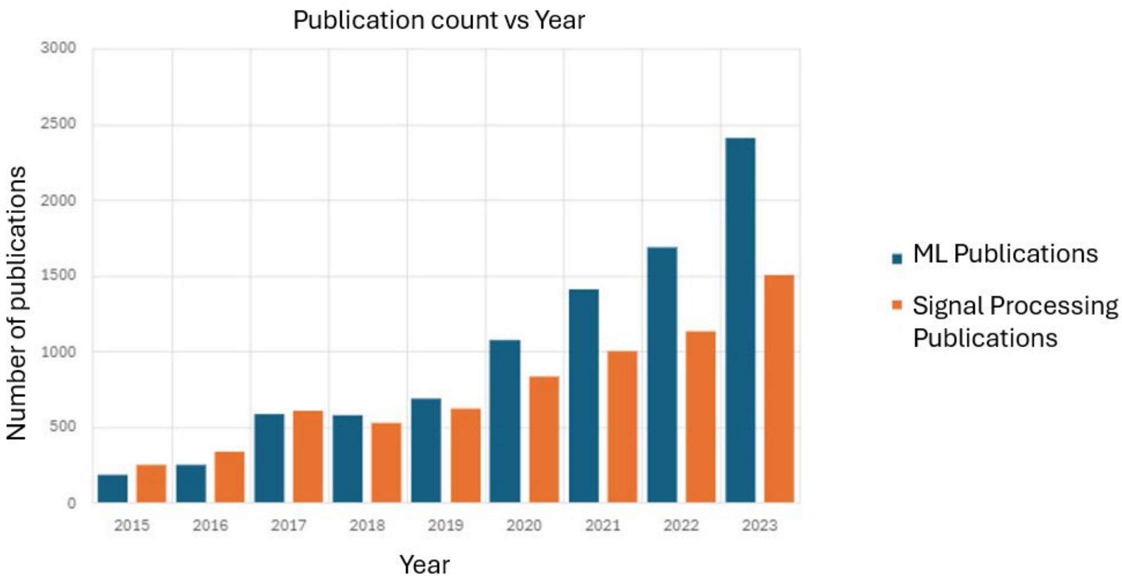


Figure 3.1: Graph of Publication Count vs Year

Publication counts for the graph were obtained using Web of Science and the key works "Hyperspectral" AND "Anomaly Detection" AND "Machine Learning" and "Hyperspectral" AND "Anomaly Detection" AND "Signal Processing" respectively. It was noted that far fewer studies have been conducted into direct comparison efforts between the two approaches in a remote sensing context, with Ghamisi et al [56] being one of relatively few examples, which unlike in this thesis, does not directly assess any effects of image pre-processing on prediction performance

3.4 Image Quality and Preprocessing in Anomaly Detection

Image preprocessing is widely recognised as a key component of anomaly detection. Techniques such as noise reduction, smoothing, dimensionality reduction, and super-resolution have been applied to enhance image quality with the aim of improving imagery analysis [57], [58]. In ML-based approaches, preprocessing affects model performance and generalisation, while in conventional methods it influences statistical estimation and background modelling [54], [55], [58] Studies such as [57] demonstrate that improving image quality through preprocessing can lead to increased ML model performance, particularly when enhancing spatial resolution using super-resolution. However, a large proportion of existing work focuses on optimising model accuracy rather than evaluating the relationship between image quality and anomaly detection model performance. As highlighted in [58], the effectiveness of preprocessing techniques is highly dependent on the specific use case, and inappropriate preprocessing can result in information loss or the introduction of artifacts.

Overall, there is limited literature that directly examines how varying levels of image quality affect both conventional and ML based anomaly detection methods, particularly studies that utilise a controlled experimental framework.

3.5 Conventional Signal Processing Approaches for HSI Anomaly Detection

Conventional anomaly detection techniques in HSI typically focus on statistical modelling of background spectral signatures and the identification of pixels that deviate significantly from this background. The Reed–Xiaoli (RX) detector is perhaps the most widely used approach and has been regularly studied in literature. This method generally assumes a Gaussian or mixture-based background model, which can limit its robustness in complex scenes [59].

Extensions to the traditional RX method have been proposed to address its limitations. For example, kernel-based RX algorithms introduce nonlinear mapping functions with the aim to capture higher-order spectral relationships, resulting in improved detection performance under certain conditions [60]. Other approaches have aimed to incorporate spatial information through subpixel and superpixel representations of features, enabling improved detection of anomalies when compared to single pixel-level approaches [61].

Despite their advantages, conventional methods can be sensitive to noise, background variability, and the high dimensionality of hyperspectral data [54], [55]. Importantly for this thesis, most studies evaluate these techniques using fixed quality imagery, with limited discussion of how degradations and improvements in resolution affect their detection performance.

3.6 Machine Learning Approaches for HSI Anomaly Detection

Machine learning techniques have gained increasing attention in HSI anomaly detection due to their ability to model complex spectral and spatial relationships. Convolutional Neural Networks and other deep learning architectures have demonstrated improved performance compared to traditional methods on benchmark datasets [62]. However, these methods often require large amounts of training data, which can be very difficult to obtain for HSI applications.

More recent work has focused on unsupervised and self-supervised approaches to address the lack of labelled HSI datasets. For example, the ML model presented in [53] introduces a training strategy that enables models to generalise well to new, unseen scenes without the need for retraining. The introduction of the HAD100 dataset in [53] has also enabled more consistent evaluation of ML anomaly detection techniques across HSI remote sensing scenes.

Despite their increased performance, ML approaches often require significant computational resource and may be more sensitive to training data characteristics than conventional techniques. Their performance therefore can degrade when applied to imagery that differs in quality or resolution from the training data, and this resulting limited generalisability can lead to a lack of real-world use cases [62].

3.7 Comparative and Hybrid Approaches

A large amount of existing literature examines conventional and ML anomaly detection techniques in isolation. Comparative studies between the two techniques are relatively rare and often focus on accuracy metrics rather than sensitivity to data quality [63]. Reviews of traditional methods [63] and ML methods [38], [64] demonstrate this gap within the literature.

Hybrid approaches that combine conventional and ML techniques have been proposed as a way to exploit the strengths of both methods. For example, image preprocessing and dimensionality reduction may be performed using traditional techniques before applying ML models to the data [65], [66]. Other approaches have used conventional methods to guide ML models towards anomalous regions, with the aim to improve detection performance and reducing false positives [67], [68].

While these hybrid methods show promise and progress in HSI anomaly detection, their performance regarding image quality conditions has not been thoroughly investigated.

3.8 Industrial and Operational Considerations

Multiple studies related to HSI anomaly detection assume access to a singular high-quality dataset and significant computational resource, these are assumptions that may not be met in real-world industrial and remote sensing applications [69], [70]. In reality, data quality may be limited by sensor limitations, acquisition cost, or sensor bandwidth.

More recently, there is research to suggest that lower-quality data may still be sufficient for certain detection tasks [71]. For example, lower-resolution imagery may be adequate for applications such as agricultural monitoring or rapid environmental assessments, where detection speed and resource efficiency are prioritised over detection quality.

However, existing ML approaches still often assume access to large, high-quality datasets. As demonstrated in [72], certain ML models may struggle to perform well when input imagery does not meet specific size or resolution requirements. This highlights the need for a more flexible, data-driven approach to anomaly detection that considers the multiple trade-offs between image quality, performance, and real world, operational constraints.

3.9 Summary of Research Gaps

This literature review identifies three key gaps in the current field of HSI anomaly detection. First, there is a lack of analysis of how variations in image quality affect the performance of both conventional and ML detection methods. Secondly, comparative studies that directly evaluate these two approaches under experimental conditions are limited. Finally, there is insufficient consideration of industrial and operational constraints, particularly with respect to data quality and computational resources.

This thesis aims to address these gaps by conducting a comparative evaluation of conventional and ML anomaly detection techniques under controlled variations in image quality using a remote sensing HSI dataset. By focusing on practical performance trade-offs, this work aims to provide insights that are relevant to real-world deployment scenarios.

A comparative summary of the HSI anomaly detection approaches discussed in this chapter, including their key characteristics, advantages, and limitations, is provided in the table below.

Approach	Characteristics	Advantages	Disadvantages
RX-based methods	Statistical anomaly detection using background spectral models; assumes Gaussian or mixture distributions	Interpretable and deterministic; low computational cost; proven in standard scenarios	Sensitive to noise and background variation; degraded performance with low image quality
Kernel-based methods	Nonlinear extension of RX using kernel functions to model high-order spectral relationships	Exploits high-order correlations; improved detection; fewer false alarms	Higher computational cost; limited evaluation across noise and quality variations
Spatially enhanced conventional methods	Integration of spectral and spatial context via subpixel, pixel, and superpixel representations	Improved detection using spectral–spatial features; validated on real and synthetic data	Additional preprocessing; sensitive to resolution changes; real vs. synthetic gap
Supervised ML methods	Deep learning models trained on labelled HSI datasets to learn spectral–spatial patterns	High detection accuracy; captures complex patterns; joint spatial–spectral learning	Requires large labelled datasets; sensitive to image quality; high compute demand
Unsupervised / self-supervised ML methods	Learning anomaly patterns from unlabelled data with cross-scene generalisation	Reduced labelling dependence; sensor generalisation; competitive performance	Dependent on training data quality; limited real-world validation
Hybrid methods	Combination of conventional preprocessing with ML-based anomaly detection	Leverages strengths of classical and ML approaches; improved robustness	Increased computational cost; limited sensitivity analysis

Table 3.1: Comparison of HSI Anomaly Detection approaches

Chapter 4

Experimental Design

Building upon the insights gained from the literature review in Chapter 3, this chapter outlines the methodology used to investigate key research questions related to anomaly detection in HSI. The objective of this experimental design is to evaluate the performance of an ML and conventional technique and assess the impact of image quality and preprocessing techniques.

To achieve this objective, three primary research questions were formulated, guiding the development of an experimental framework:

- What is the performance difference between AETNet and RX anomaly detection?
- What is the lowest quality image that can be used to perform HSI anomaly detection?
- What benefit does performing super resolution on the training and testing set add?

4.1 Dataset Acquisition

In order to answer the above questions, the first step was to obtain sufficient data to train and test the AETNet that would also be suitable for utilising RX detection. When searching for a suitable dataset to perform anomaly detection using AETNet multiple options were considered and assessed for suitability which are detailed below.

It should be noted that where HSI images are presented in this thesis, they are displayed in false colour maps where colours are artificially assigned to represent specific wavelengths not visible to the human eye in order to aid visualisation.

4.1.1 ABU Dataset

The ABU dataset [73], published in 2017 consists of 11 scenes captured by NASA's AVIRIS sensor [74]. An example ABU image and ground truth map are displayed below. The images within this dataset are split into three categories: Airport, Beach and Urban. Image size ranges from 80x100 pixels to 400x400 pixels with an average of 185 spectral bands per image. Each image in the ABU dataset has a corresponding ground truth map with anomalies identified to enable to training and testing of models [75]. Whilst this dataset would be useful for more traditional anomaly detection techniques, it does not contain enough images to both train and test the AETNet to allow sufficient comparison with RX detection. AETNet is a deep learning-based method which requires a substantial amount of data for both training and testing to learn meaningful patterns and to generalise well to new, unseen examples, it is likely therefore that the performance comparison with RX (which is typically less data-hungry) would be compromised due to the insufficient dataset size.

Figure 4.1: False colour map and ground truth of an ABU Airport Scene [73]

4.1.2 Target Detection Blind Test Dataset

The Target Detection Blind Test Dataset [76] contains one training and one testing HSI scene, displayed below. The data was collected with the HyMap sensor [77] and contains 126 bands with a resolution of 3m. The classes in the training image are identified and labelled in 5 categories: Buildings, Trees, Grass, Road and Hill Terrain. This dataset provides a useful way to test developed techniques, however it is likely that one image is too few to optimally train an ML model, therefore making it an unsuitable dataset for this thesis.

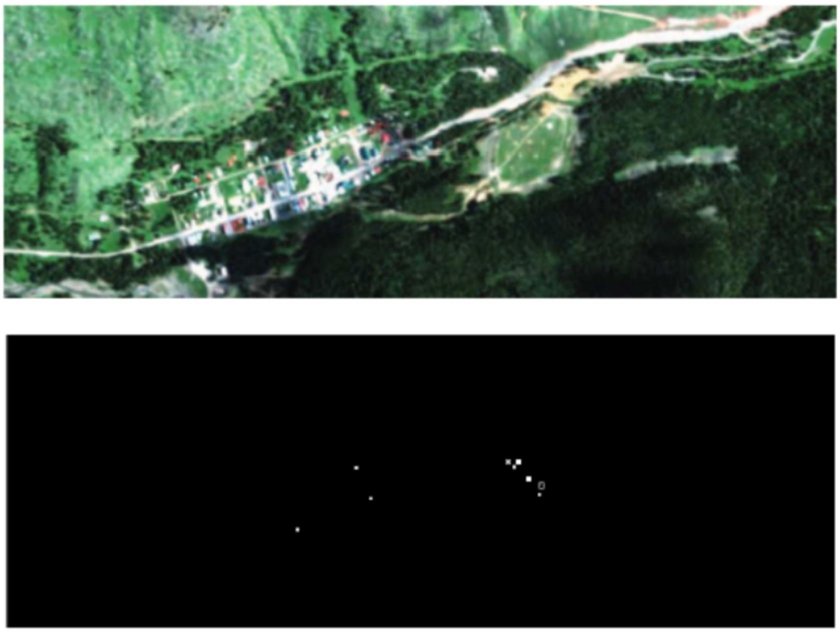


Figure 4.2: False colour map and ground truth of the Target Detection Blind Test Scene [76]

4.1.3 Multi-Platform Optical Remote Sensing Dataset for Target Detection

The Multi-Platform Optical Remote Sensing Dataset for Target Detection [78] utilises NASA's AVIRIS sensor at a spatial resolution of 4m with 370 spectral bands to capture a scene of size 436×1171 pixels that contains 5 targets. A false colour map of the scene is given below. The targets within the scene consisted of fabric sheets of size 10mx10m that were placed atop vegetation and soil backgrounds. Similar to the above datasets, this dataset is useful for testing methods however it does not provide enough data to fully train an ML model.

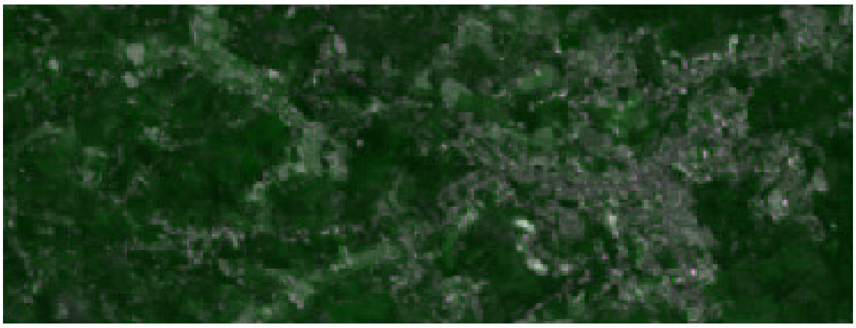


Figure 4.3: False colour map of the Multi-Platform Optical Remote Sensing Scene [76]

4.1.4 HAD100 Dataset

The HAD100 dataset [53] was identified as the most suitable dataset for this study due to its amount of images and available ground truth. The HAD100 dataset consists of 260 anomaly-free HSIs for training and 100 test HSIs obtained from the NASA AVIRIS-NG Sensor [79]. All images are 64x64 pixels in size and the size of ground truth targets ranges from 1 to 69 pixels in various backgrounds, including forest and farmland. The limitation of the use of this dataset is the limited image size. As images decrease in size due to downsampling, the use of RX detection becomes increasingly problematic due to lack of pixels available for covariance calculation. The smallest images used within this thesis were 16x16 pixels which is suboptimal for the use of this technique, however with very limited options available it was deemed that the study should proceed but with this limitation noted.

All training, testing and inference was performed on the first 50 bands of the data as was conducted in the original AETNet paper to enable a clear baseline and comparison. A selection of HAD100 images and ground truth maps are displayed below.

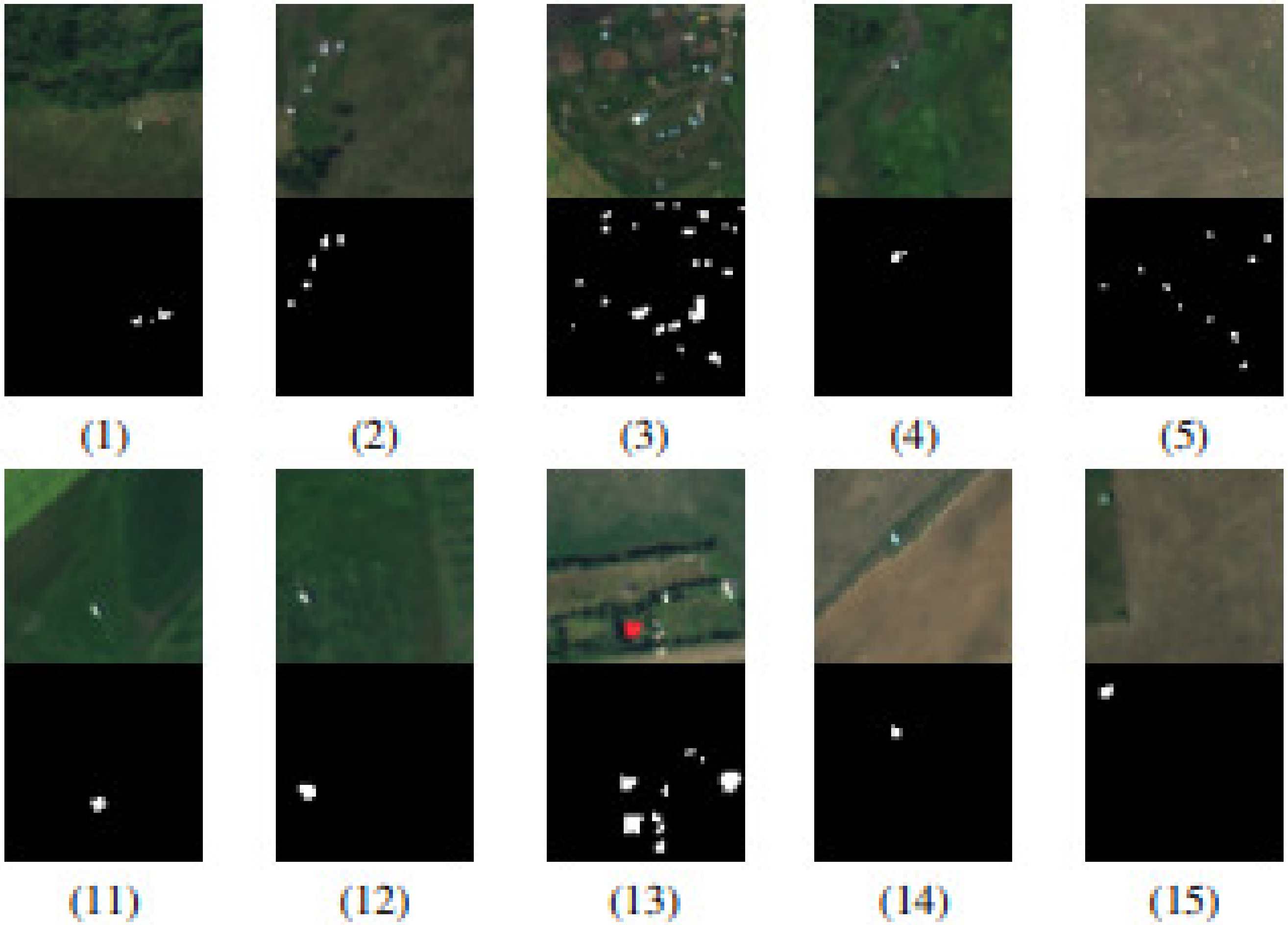
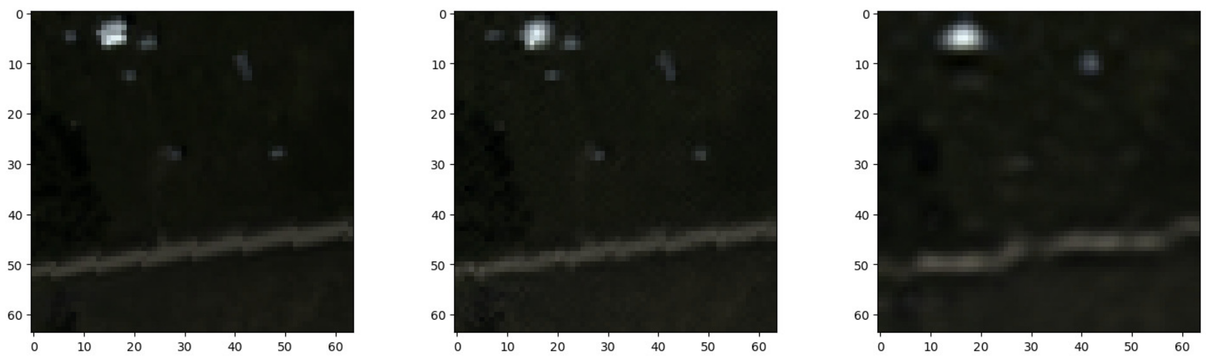


Figure 4.4: False colour maps (top) and respective manually labelled ground truth (bottom) of a selection of test scenes in the HAD100 dataset [53]

4.2 Technique Implementation: DIP

Following acquisition of the HAD100 dataset, super resolution was performed utilising DIP at multiple scale factors. An example of super resolution performed on test image 94 is given below. The code to perform this was obtained from [80].



(a) False colour map of original HAD100 test image 94

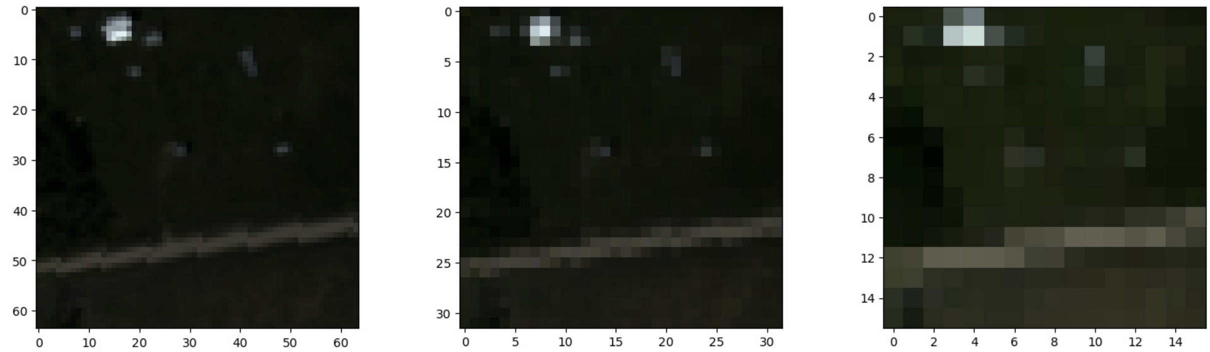
(b) False colour map of HAD100 test image 94 enhanced using DIP with SF 2

(c) False colour map of HAD100 test image 94 enhanced using DIP with SF 4

Figure 4.5: False colour maps of HAD100 test image 94 enhanced using DIP

4.3 Technique Implementation: Downsampling

In order to degrade the quality of the HAD100 dataset, images were downsampled at multiple scale factors using Gaussian filtering and OpenCV in Python. This reflects practical constraints such as limited transmission bandwidth or resource-constrained deployment environments. An example of this downsampling process performed on test image 94 is given below.



(a) False colour map of original HAD100 test image 94

(b) False colour map of HAD100 test image 94 degraded by downsampling with SF 2

(c) False colour map of HAD100 test image 94 degraded by downsampling with SF 4

Figure 4.6: False colour maps of HAD100 test image 94 degraded using downsampling

Chapter 5

Methodology & Results

To address the research questions outlined in the previous chapter, seven experiments were conducted using the HAD100 dataset. The results were evaluated based on the metrics outlined below, with the goal of providing a comprehensive analysis of how image quality impacts the performance of anomaly detection techniques. This chapter presents the experiments, their respective results, the performance metrics used, and a discussion of the findings in the broader context of the thesis.

5.0.1 Metrics

In order to enable a fair comparison between conventional techniques and ML techniques, common evaluation metrics must be defined. When undertaking the literature search detailed previously, six key evaluation techniques stood out as widely used to asses model performance, the precision, recall [81] , mean absolute error [82], F1 scores, precision-recall graphs [83] and F1-threshold graphs [84]. These metrics provide an insight into the performance of a model by measuring its ability to correctly classify ground truth imagery and are detailed below before the results are presented [85].

5.0.2 Precision

The precision of a model is the ratio of true positive predictions to the total number of positive predictions, this measures the accuracy of the positive predictions. A high precision (close to 1.0) indicates that the model makes very few false positive errors [81].

$$\text{Precision} = \frac{TP}{TP + FP} \tag{5.1}$$

Where:

- TP is the number of true positive results,
- FP is the number of false positive results.

5.0.3 Recall

Recall is the ratio of true positive predictions to the total number of actual positive ground truth examples. It measures the model's ability to identify all relevant examples within a dataset. A high recall (close to 1.0) indicates that the model successfully identifies most of the positive instances [81].

$$\text{Recall} = \frac{TP}{TP + FN} \tag{5.2}$$

Where:

- TP is the number of true positive results,
- FN is the number of false negative results.

5.0.4 F1 Score

The F1 score of a model is the harmonic mean of the precision and recall, this provides a single metric that balances the trade-off between the two previous metrics. The F1 Score is particularly useful when the distribution of the classes is heavily unbalanced or when both the presence of false positives and false negatives will cause significant disruption on model output. A higher F1 score (close to 1.0) indicates better performance [83].

$$F_1 = 2 \cdot \frac{TP}{2TP + FP + FN} \tag{5.3}$$

Where:

- TP is the number of true positive results,

- FP is the number of false positive results,
- FN is the number of false negative results.

5.0.5 Mean Absolute Error Score

The Mean Absolute Error (MAE) score is used to measure the accuracy of model predicted pixel values in an image compared to the ground truth pixel values. It is often used to evaluate image processing tasks such as image denoising and super resolution [86]. The MAE score is calculated by averaging the absolute differences between predicted pixel values and the corresponding ground truth values across all pixels in the image. When applied to binary detection maps, MAE measures the proportion of incorrectly classified pixels, where values closer to 0 indicate better agreement between the predicted and ground truth maps [82].

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{5.4}$$

Where:

- n is the number of data points,
- y_i is the true value for the i-th data point,
- $\hat{y}_i$ is the predicted value for the i-th data point.

5.0.6 Precision Recall Curve

Both precision vs recall (PR) curves and plots of F1 score vs model threshold were plotted to provide a graphical representation of technique performance. The use of P-R curves provides a clear picture of model performance by focusing on its ability to correctly identify and classify anomalies to demonstrates the trade-off between precision and recall. This is important in an anomaly detection use case as the consequences of missing anomalies (low recall) or predicting false positives (low precision) can vary depending on the use case. A P-R curve can provide an assessment of a model's effectiveness in detecting the rare anomaly class. By examining a P-R curve, the optimal threshold for model performance can also be identified to balance the trade-off between the two metrics, this ensures a model performs reliably and as expected when applied to real-world datasets [83]. The following graphs are presented per experiment and for clarity, the baseline results from Experiment 0 have been plotted on all graphs to allow performance comparison.

5.0.7 F1 Score vs Model Threshold

Plotting the F1 score vs model threshold enables for evaluation of and optimisation of each model's performance. By plotting F1 score against different thresholds, the effect of threshold on model precision can be visualised. This is especially important in anomaly detection as anomalies are rare and often outnumbered by the "normal" class. The F1 score vs. threshold plot (applied to the test dataset) helps to visually identify the ideal threshold to maximise the F1 score without causing an overwhelming number of false positives [84]. By using this plot, the threshold can be tailored to specific use case requirements, ensuring the model is effective in real-world scenarios.

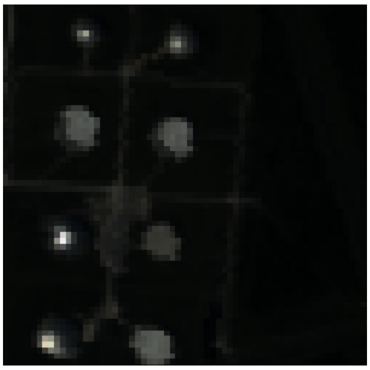
5.0.8 Note

The precision recall curves and F1 vs model threshold graphs are presented in order of experiment with a brief description of the results shown. Following this output prediction results from each experiment are shown alongside the false colour map of HAD100 test image 118 (an image containing multiple clear anomalies) and its ground truth. Full discussion of all results is presented at the end of this chapter. The ground truth labels were constructed by applying the same image transformation pipeline (e.g. downsampling or super resolution) to the original ground truth pixel maps, ensuring spatial consistency with the corresponding input data.

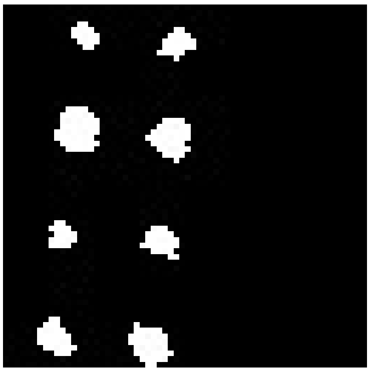
5.0.9 Experiment 0: Baseline Experiment

This experiment serves as the baseline for comparing the performance of a conventional signal-processing method (RX) with an ML approach (AETNet) in the context of HSI anomaly detection. By establishing a standard benchmark on a shared dataset, the experiment allows for a controlled evaluation of both techniques under identical conditions [87]. The results of this baseline experiment provide a foundational reference point, against which the strengths and weaknesses of the different methods can be objectively measured. This comparative analysis is essential for understanding how well an ML approach performs relative to a traditional technique and for identifying areas where an ML approach may offer significant advantages. Experiment

0A involved training an AETNet on the original HAD100 dataset (260 images) and running inference on the original HAD100 test set of 100 images (model identifier AE_0A). Experiment 0B consisted of running RX anomaly detection over the original HAD100 test set (100 images).



(a) False colour map of HAD100 test image 27



(b) Ground Truth

Figure 5.1: Experiment 0 Data Example

5.0.10 Experiment 0 Results

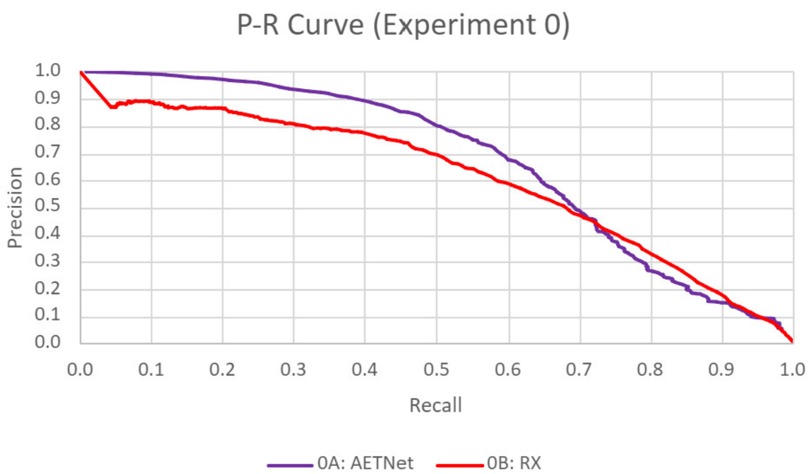


Figure 5.2: Experiment 0: Precision-Recall

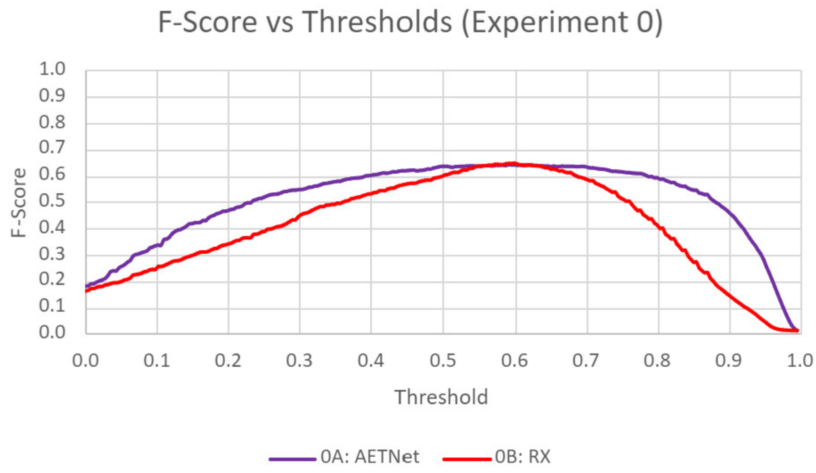
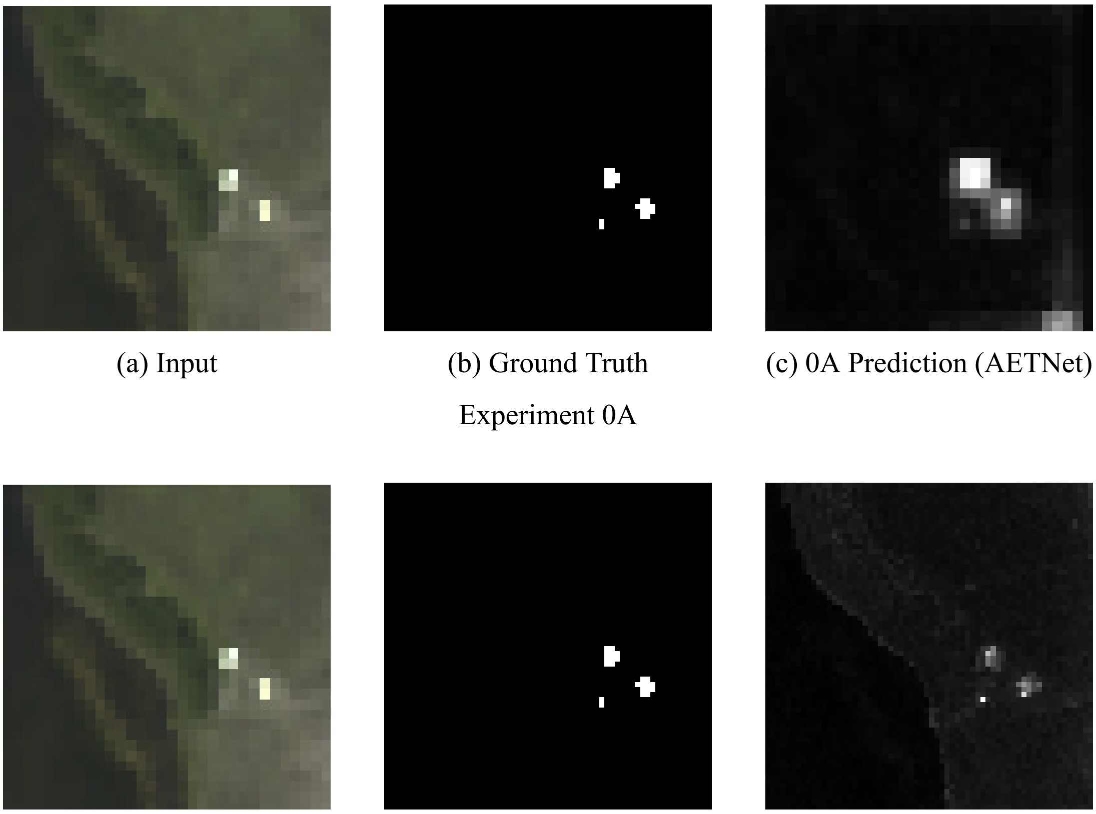


Figure 5.3: Experiment 0: F1 Score



(a) Input (b) Ground Truth (c) 0A Prediction (AETNet)

Experiment 0A

(d) Input (e) Ground Truth (f) 0B Prediction (RX)

Experiment 0B

Figure 5.4: Experiment 0 Results: (0A) AETNet vs (0B) RX anomaly detection on identical test image

Experiment 0 acts as the bench marking experiment. Both techniques were run over original images (no super resolution or downsampling) and the AETNet model used (AE_0A) was trained only on original images. The results of this experiment suggest that the AETNet performs anomaly detection more effectively than RX detection in the majority of cases across all thresholds. However when recall is high, the traditional RX detector tends to slightly outperform the ML AETNet as seen in the figure above.

When visually examining the outputs of Experiment 0A and B there are clear and distinguishable anomalies detected by both techniques.

5.0.11 Experiment 0 Discussion

The results from Experiment 0 suggest that the AETNet outperforms RX detection. This result may be due to the following factors, AETNet utilises a deep neural network whereas RX is a statistical based approach to anomaly detection. The use of the DNN allows the model to extract more complex information and relationships from the HSI compared to RX that relies primarily on statistical deviations from a calculated norm. It could also be suggested that the use of random masking within AETNet enables the model generalise and adapt to unknown features better. It should however again be noted that there were limited pixels available to perform optimal covariance estimations which may affect the RX results.

5.0.12 Experiment 1: AETNet Degraded Test Set

Experiment 1 was the first experiment to introduce image degradation. Both experiment 1A and 1B utilised the model trained in experiment 0A (AE_0A). Inference was run using model AE_0A over a test set that had been downsampled by scale factors 2 and 4 respectively. The goal of this experiment was to evaluate the impact of image degradation on the performance of the AETNet model and to determine the extent to which performance decreases between the different scale factors.

(a) False colour map of HAD100 test image 27 used in Experiment 1A - SF 2

(b) False colour map of HAD100 test image 27 used in Experiment 1B - SF 4

Figure 5.5: Experiment 1 Data Example

5.0.13 Experiment 1 Results

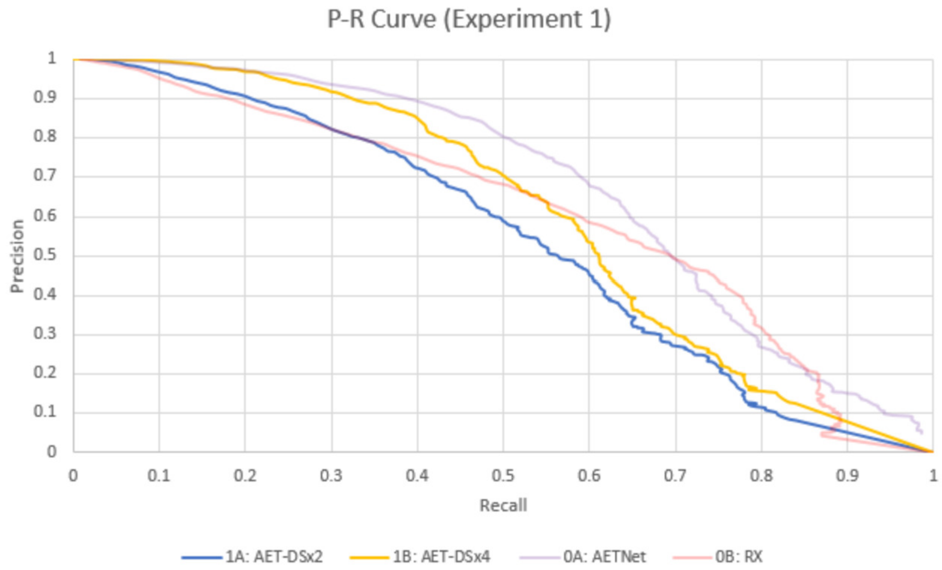


Figure 5.6: Experiment 1: Precision-Recall

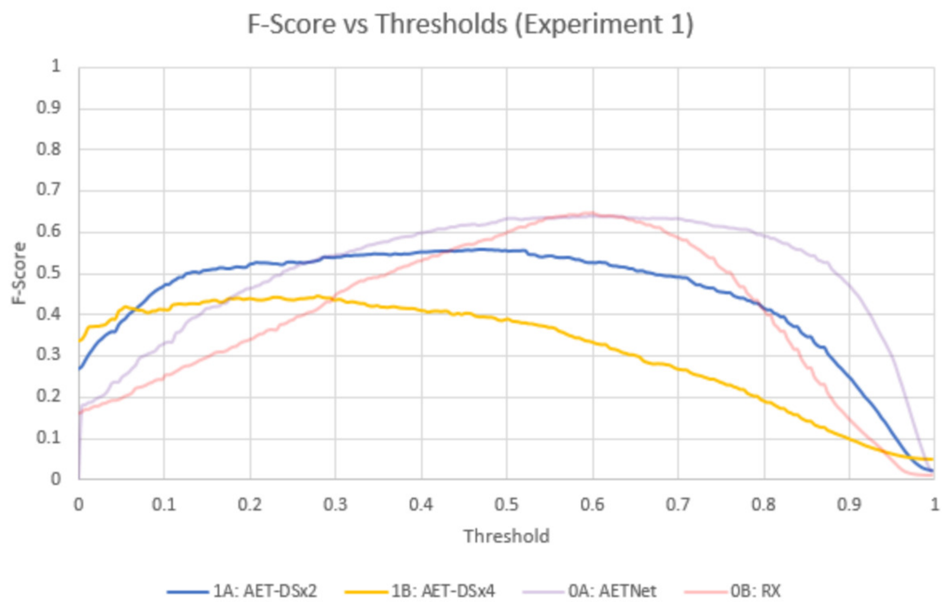
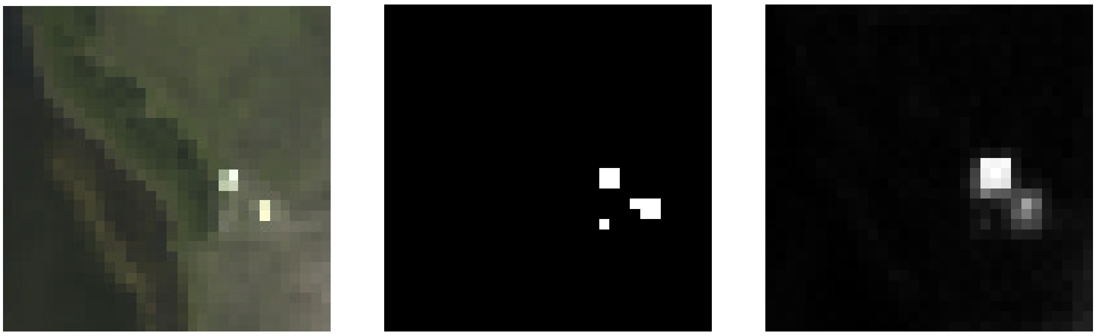


Figure 5.7: Experiment 1: F1 Score

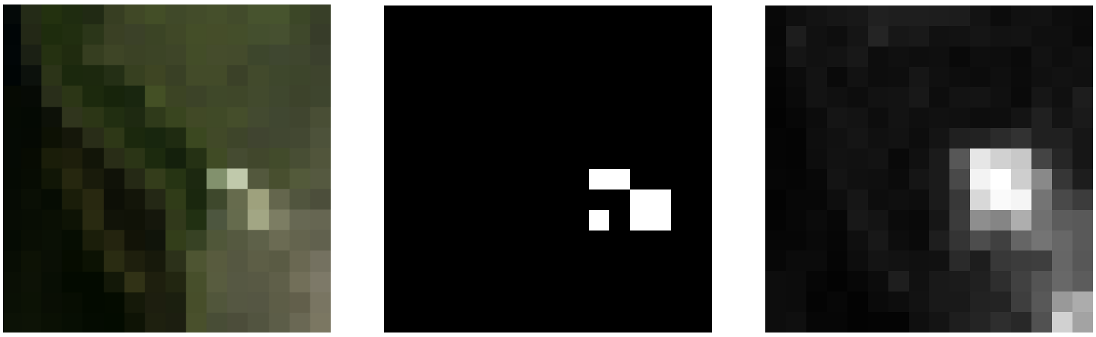


(a) False colour map of HAD100 test image 118

(b) Ground Truth

(c) Model Prediction

Experiment 1A Results



(d) False colour map of HAD100 test image 118

(e) Ground Truth

(f) Model Prediction

Experiment 1B Results

Figure 5.8: Experiment 1 Results: (1A) AETNet vs (1B) RX anomaly detection under degraded test conditions

Results from Experiment 1 suggest that the standard AETNet again outperforms other detectors. When recall is low the performance of the AETNet is superior to RX detection, this occurs even when the AETNet inference is run over lower quality, downsampled test images.

The model predictions of Experiment 1A are identifiable when presented visually, whereas the anomalies predicted in 1B are clustered in the correct area but appear smeared and less differentiable due to the use of downsampling on the original image.

5.0.14 Experiment 1 Discussion

In Experiment 1, the AETNet demonstrates higher performance in anomaly detection when compared to RX detection. Specifically, AETNet outperforms RX in situations where the recall is low, which implies that AETNet may be more effective at correctly identifying anomalies without compromising on precision when only a few anomalies are detected. This improvement can be attributed to AETNet's learned spectral–spatial representations, which present as more robust to noise and resolution loss, whereas RX relies on a fixed global Gaussian model that may be more sensitive to degradation of the underlying covariance structure. AETNet maintains superior performance even with lower quality, downsampled test images, which may indicate that it is capable of detecting anomalies in unideal conditions. Therefore, AETNet may present a more reliable choice for use cases where test data may not always be of a high quality.

5.0.15 Experiment 2: AETNet Degraded Training Set

Experiment 2 was designed to assess how downsampling impacts the performance of the AETNet model when the training data has also been downsampled.

By introducing downsampling during both the training and testing phases, this experiment aims to simulate realistic scenarios where both input and training samples are of lower quality. Understanding the effects of image quality in these conditions is important, as it allows for the evaluation of the model's adaptability to varying levels of image resolution. The insights from experiment 2 are useful for real-world applications where high resolution data may not always be available. Test images used in Experiment 2A and B are the same as Experiment 1A and B respectively.

5.0.16 Experiment 2 Results

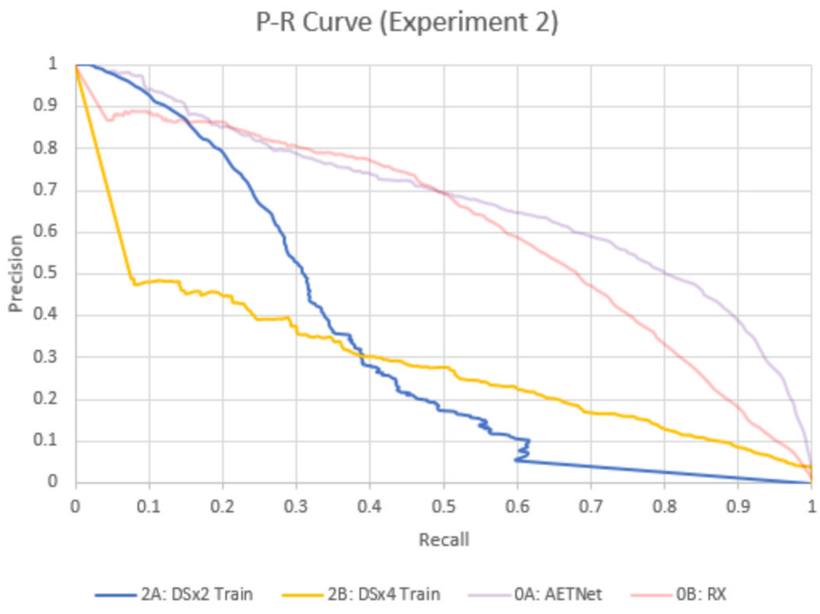


Figure 5.9: Experiment 2: Precision-Recall

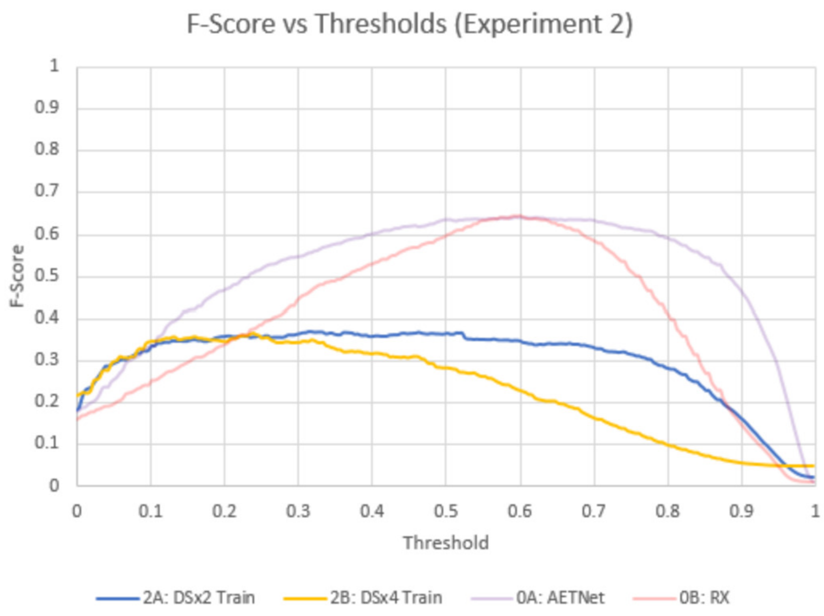
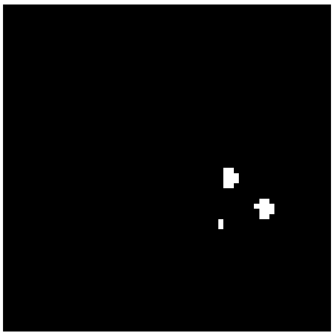
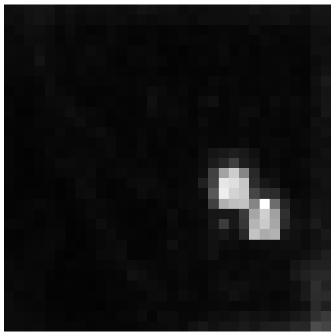


Figure 5.10: Experiment 2: F1 Score

(a) False colour map of HAD100 test image 118

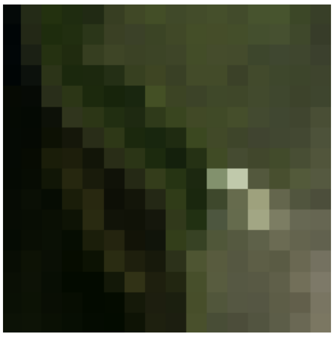


(b) Ground Truth

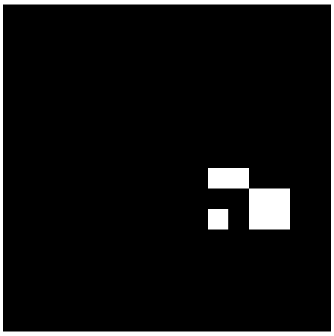


(c) Model Prediction

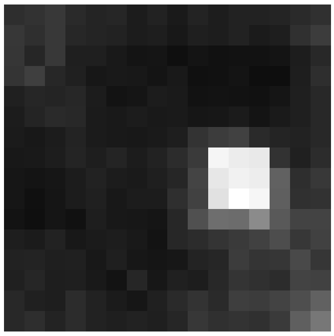
Experiment 2A Results



(d) False colour map of HAD100 test image 118



(e) Ground Truth



(f) Model Prediction

Experiment 2B Results

Figure 5.11: Experiment 2 Results: AETNet trained and evaluated on downsampled data. Test images used in (2A) and (2B) are the same as those used in Experiment 1A and 1B respectively.

Experiment 2 demonstrates a significant decrease in model performance, indicating that ,as expected, downsampling the training set directly and negatively impacts prediction quality. When model threshold is set low, the level of downsampling does not affect the F1 score of the model. At thresholds greater than 0.25, downsampling the data by a scale factor of 2 significantly outperforms the scale factor 4 downsampled set.

Upon visual analysis of the results of Experiment 2, the model downsampled by a factor of 2 tends to identify the ground truth anomalies accurately however it also appears to include a surrounding "buffer" area around the main detected anomalies leading to an increased false positive rate and reduced spatial precision that may be problematic in situations when high precision is required. Experiment 2B generates predicted anomalies that are not clearly defined and are several pixels larger than the ground truth. However, these predictions still tend to cluster around the regions where the actual anomalies are located in the ground truth labels.

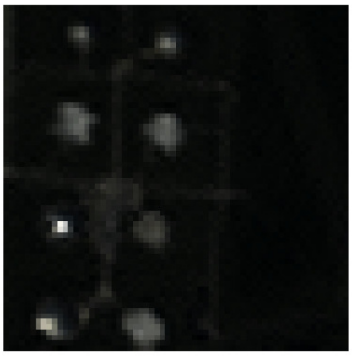
5.0.17 Experiment 2 Discussion

The drastic decrease in model performance suggests that image quality and training data resolution are important when assessing a model's performance. By downsampling the training set, important details and features necessary for accurate anomaly detection are likely lost, leading to poorer model predictions. This highlights the model's need for high quality training data to enable effective learning and accurate predictions. The results of Experiment 2 suggest that without mitigating image enhancements, low quality imagery will lead to poorer model predictions and a traditional method may provide similar results to an ML based one.

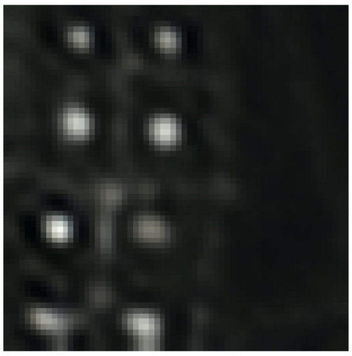
5.0.18 Experiment 3: AETNet Degraded & Enhanced Test Set

Experiment 3 was divided into three parts, all of which used the AE_0A model trained on the original HAD100 dataset. In Experiment 3A, a test set that had been enhanced using DIP with a scale factor of 2 was used. Experiments 3B and 3C involved running inference on a test set that was first downsampled and then enhanced using KernelGAN by scale factors of 2 and 4, respectively. This experiment is conducted to determine whether artificially enhancing low-resolution images improves or degrades the model's performance, thereby assessing its robustness to super resolution in preprocessing and unseen image transformations.

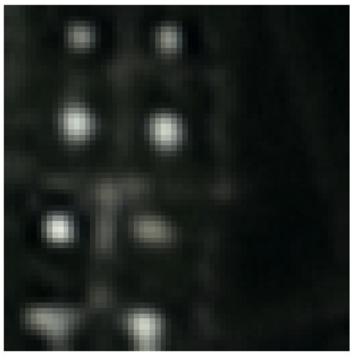
By enhancing the test set using KernelGAN before running inference, the experiment simulates conditions where image quality has been artificially improved allowing for an analysis of how well the model understands unknown enhanced details. Evaluating the model's performance on both downsampled and enhanced images provides insight into how it handles varying image resolutions and quality. This understanding is important for real-world applications where images may be pre-processed to improve resolution before analysis. If the model performs well with super resolution data, it suggests that using enhancement techniques like KernelGAN could be a viable strategy to improve model accuracy in situations where high quality images are not readily available. However, if performance drops, it indicates that the model may struggle with artefacts introduced by super resolution methods.



(a) False colour map of HAD100 test image 27 used in Experiment 3A



(b) False colour map of HAD100 test image 27 used in Experiment 3B



(c) False colour map of HAD100 test image 27 used in Experiment 3C

Figure 5.12: Experiment 3 Data Example

5.0.19 Experiment 3 Results

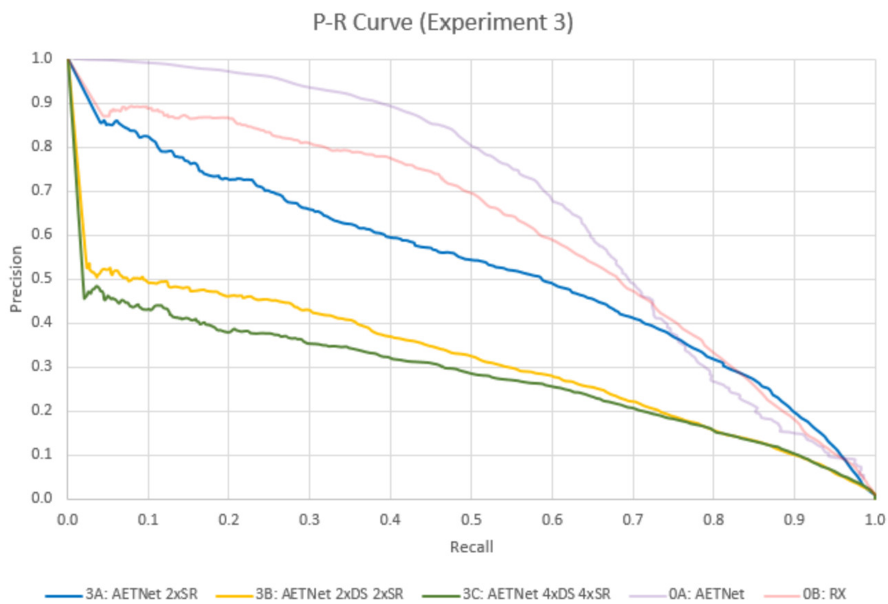


Figure 5.13: Experiment 3: Precision-Recall

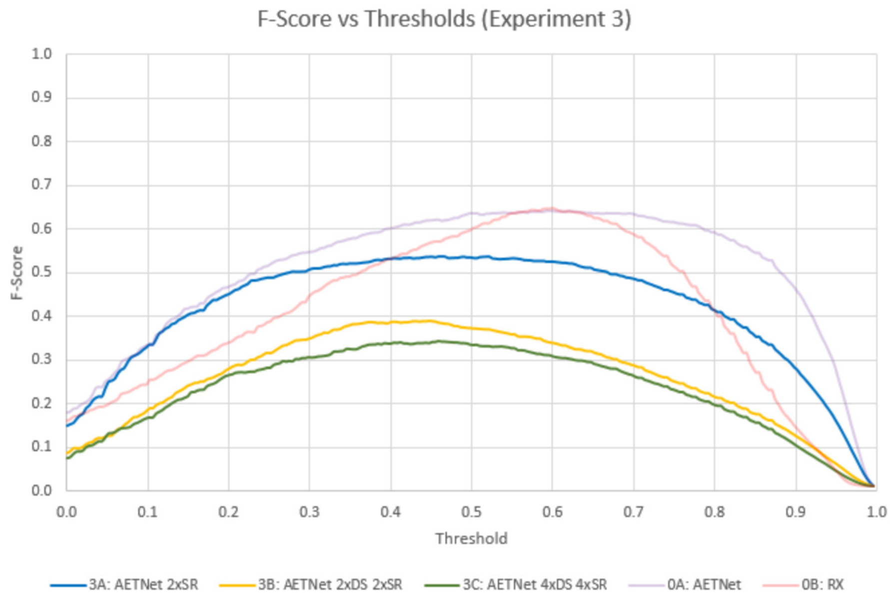
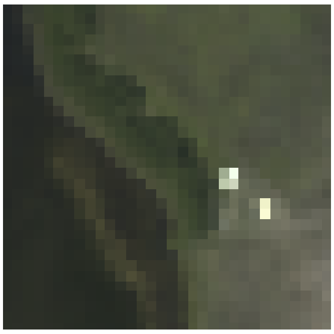
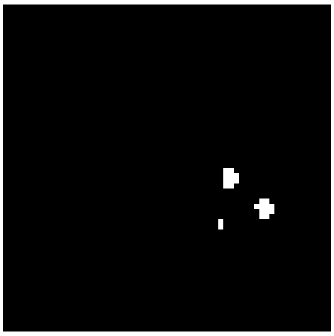


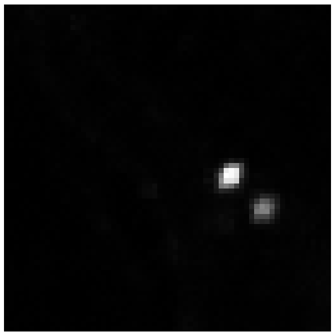
Figure 5.14: Experiment 3: F1 Score



(a) False colour map of HAD100 test image 118

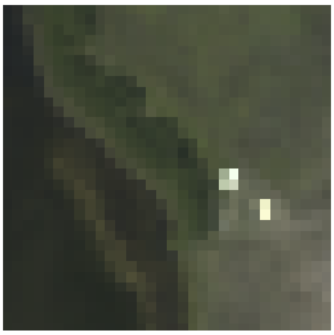


(b) Ground Truth

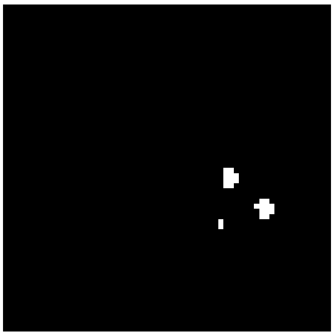


(c) Model Prediction

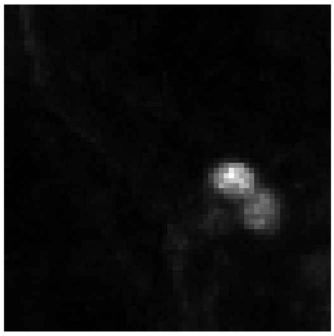
Experiment 3A Results



(d) False colour map of HAD100 test image 118



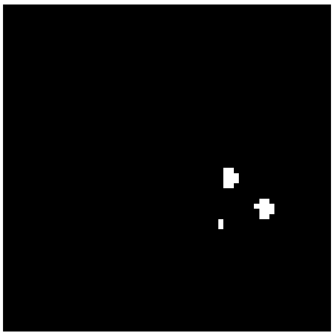
(e) Ground Truth



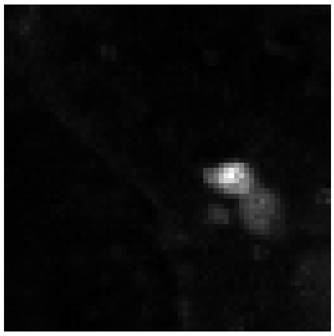
(f) Model Prediction

Experiment 3B Results

(g) False colour map of HAD100 test image 118



(h) Ground Truth



(i) Model Prediction

Experiment 3C Results

Figure 5.15: Experiment 3 Results: AE_0A model evaluated on enhanced and degraded test sets. (3A) test set enhanced using DIP with scale factor 2, (3B) test set downsampled then enhanced using KernelGAN with scale factor 2, and (3C) test set downsampled then enhanced using KernelGAN with scale factor 4.

Experiment 3 introduces super resolution to the HAD100 test set. As expected, test sets that were only enhanced using super resolution performed better than those that were first downsampled and then enhanced. The performance of the original AETNet on test data enhanced by a scale factor of 2 is comparable to that of the RX detector and it outperforms both RX and the original AETNet at when recall levels are high.

When examining the model thresholds, the results suggest that performance consistently declines on the downsampled test sets, however there appears no significant variation in performance based on the specific downsampling scale factor.

From visual examination of Experiment 3 model results, it can be seen that all 3 models perform well and produce distinguishable anomalies. These predictions tend to increase in blurriness with an increasing super resolution scale factor, making detected anomalies appear slightly less distinct.

5.0.20 Experiment 3 Discussion

The finding from Experiment 3 that test sets enhanced using super resolution outperform those that are downsampled before being enhanced suggest that preserving data quality of HSIs is important for increased model performance. This suggests that super resolution is more effective when applied to imagery that retains most of its original, complex details. The second finding is that the original AETNet's performance on test data enhanced by a scale factor of 2 is comparable to the original RX detector which demonstrates that the AETNet may not offer significant performance increase in a scenario where data has not been artificially enhanced. Finally, the decline in performance on the downsampled test sets suggests that the downsampling process causes the loss of important information from data that enhancement techniques, such as super resolution, cannot fully recover. This proves the importance of maintaining data quality to prevent a decline in model performance.

The lack of performance variation between different downsampling factors suggests that once data quality is compromised, the scale factor of downsampling becomes less important suggesting that the act of downsampling itself is detrimental rather than the specific level it occurs at. The results of Experiment 3 demonstrate the importance of minimising data loss before enhancement and suggest that an AETNet trained on standard quality data can perform well on higher quality data.

5.0.21 Experiment 4: AETNet Enhanced Training & Original Test Set

This experiment aims to increase understanding of how the AETNet model responds to being trained with super resolution imagery but tested on the original data. In Experiment 4A, the model was trained on data that had not been downsampled but had been enhanced by a scale factor of 2. Experiment 4B involved training the model on images that had been downsampled and then enhanced by a scale factor of 2, while Experiment 4C followed a similar approach to 4B but with a scale factor of 4.

Training the model using super resolution data offers several advantages in scenarios where image quality, resolution, or detail are important. As super resolution enhances image details, this can improve the model's ability to detect minor features and lead to higher accuracy in a variety of tasks. By providing more data points, super resolution may help the model learn complex patterns that are present within an image. The use of super resolution in model training also introduces variability that can reduce model overfitting alongside the use of standard techniques such as data augmentation to help the model generalise better [88]. Test images used in Experiment 4 are the same as Experiment 0.

5.0.22 Experiment 4 Results

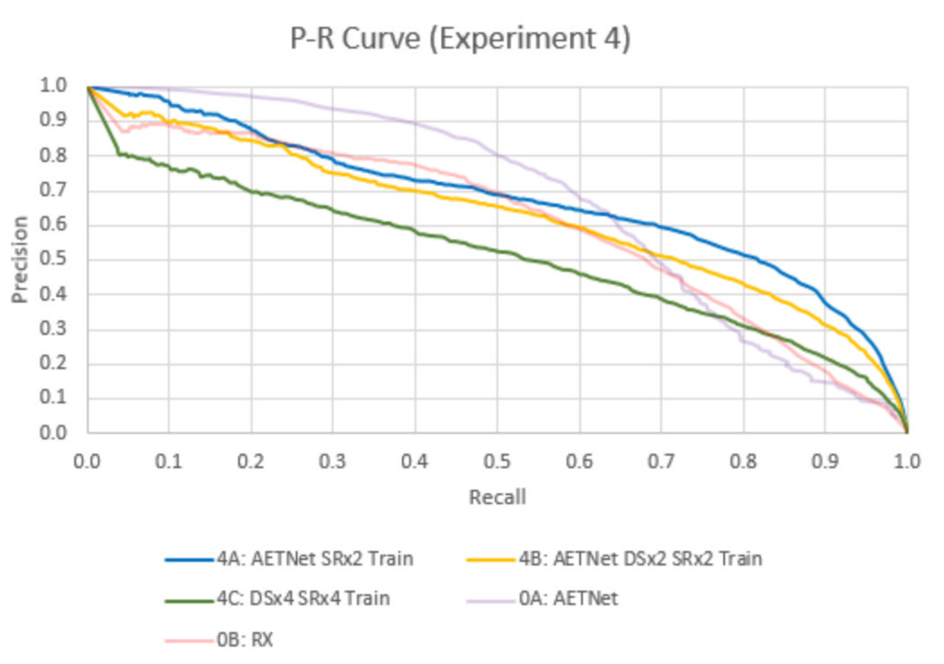


Figure 5.16: Experiment 4: Precision-Recall

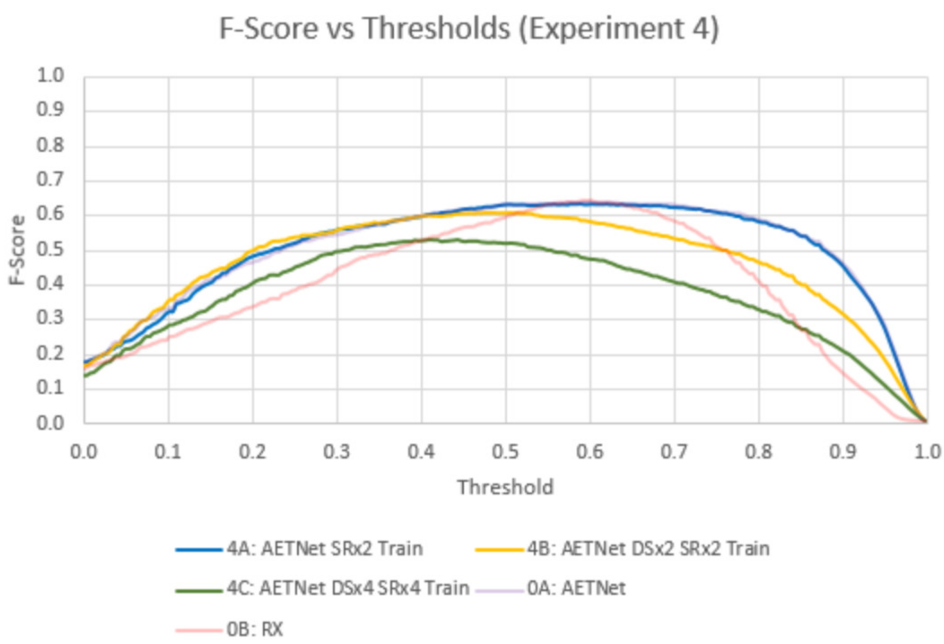
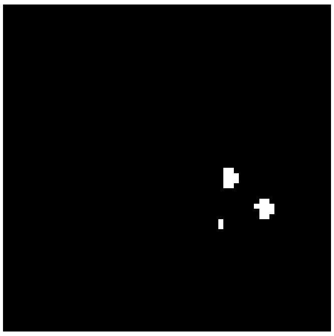
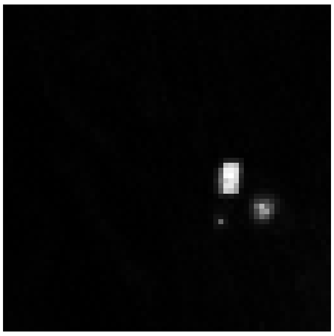


Figure 5.17: Experiment 4: F1 Score

(a) False colour map of HAD100 test image 118



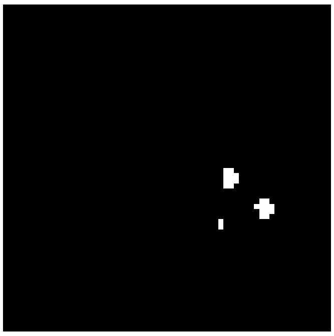
(b) Ground Truth



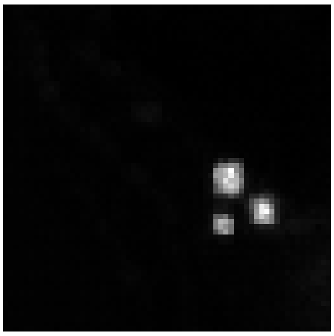
(c) Model Prediction

Experiment 4A Results

(d) False colour map of HAD100 test image 118



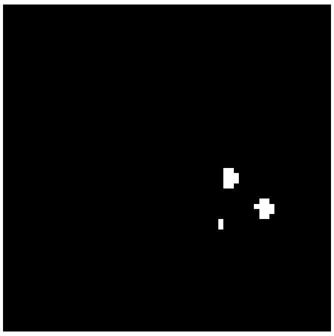
(e) Ground Truth



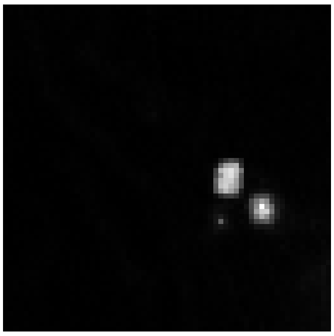
(f) Model Prediction

Experiment 4B Results

(g) False colour map of HAD100 test image 118



(h) Ground Truth



(i) Model Prediction

Experiment 4C Results

Figure 5.18: Experiment 4 Results: AETNet trained on enhanced imagery and evaluated on the original test set. (4A) trained on non-downsampled images enhanced with scale factor 2, (4B) trained on downsampled images enhanced with scale factor 2, and (4C) trained on downsampled images enhanced with scale factor 4.

In Experiment 4, the results show that the AETNet model trained on images enhanced with a scale factor of 2 outperforms the traditional RX detection method and the model trained of images enhanced by scale factor 4. The F1 score of the model used in Experiment 4A surpasses that of all other models tested in this experiment across all thresholds.The P-R curve also suggests that this model is effective at balancing both precision and recall.

When examining the performance of models that were trained on images that were downsampled and then enhanced, a clear gap emerges between different downsampling scale factors. The model that has been downsampled and enhanced using a scale factor of 2 (4B) shows a significant improvement in performance over the model that uses a scale factor of 4 (4C). This improvement is clear in both the P-R curve and the F1 Score vs. threshold graphs. To contrast this result, there is a less significant performance decrease between the enhanced model (4A) and the model that was firstly downsampled and then enhanced with a scale factor of 2 (4B).

When reviewing model predictions for Experiment 4, all three sub-experiments present distinguishable and clear anomalies.

5.0.23 Experiment 4 Discussion

The results of Experiment 4 suggest that enhancing images by a scale factor of 2 and training the AETNet model on these images benefits the performance of the AETNet when considering model precision. Model 4A outperforms both the original AETNet and the conventional RX detection method in precision score. This result suggests that the use of super resolution at a specific level of enhancement may improve the AETNet model's capability to detect anomalies in HSI more accurately.

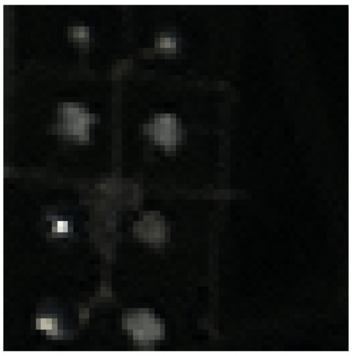
The differences in performance between AETNet models trained on images downsampled and then enhanced at varying scale factors emphasises the importance of implementing effective image preprocessing steps depending on the use case. The significant improvement of the model using a scale factor of 2 over the model using a scale factor of 4 demonstrates that extreme downsampling may result in a loss of important information that is not recovered with enhancement. This loss of information then directly affects the model's ability to accurately detect anomalies in the data. However, the minor performance difference between the enhanced model (4A) and the downsampled then enhanced model with a scale factor of 2 (4B) suggests that moderate levels of downsampling did not compromise the model's capabilities for a remote sensing use case. This allows for the conclusion that it may be worth performing super resolution on imagery that is of significant "bad" quality if aiming to utilise this "bad" data as training data for ML based anomaly detectors.

These results suggest that super resolution can be an effective tool to improving the accuracy of machine learning models in detecting anomalies.

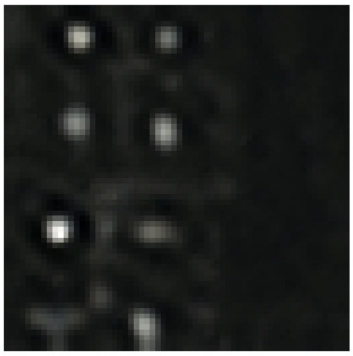
5.0.24 Experiment 5: AETNet Enhanced Test Set

Experiment 5 involved two tests with both being trained on enhanced images. Test 5A used a scale factor of 2 and 5B used a scale factor of 4. Additionally, both tests utilised a test set of enhanced imagery at their respective scale factors.

This experiment provides insight into how the model performs when both the training and testing datasets have been enhanced using different levels of super resolution. By using scale factors of 2 and 4, Experiment 5A and 5B enable the analysis of how varying degrees of enhancement impact model performance. Training and testing the model on super resolution images simulates a scenario where image enhancement techniques are used to improve quality before processing. The results from this experiment can help determine whether larger scale factors lead to increased model accuracy or if this stabilises at a specific point.



(a) False colour map of HAD100 test image 27 used in Experiment 5A



(b) False colour map of HAD100 test image 27 used in Experiment 5B

Figure 5.19: Experiment 5 Data Example

5.0.25 Experiment 5 Results

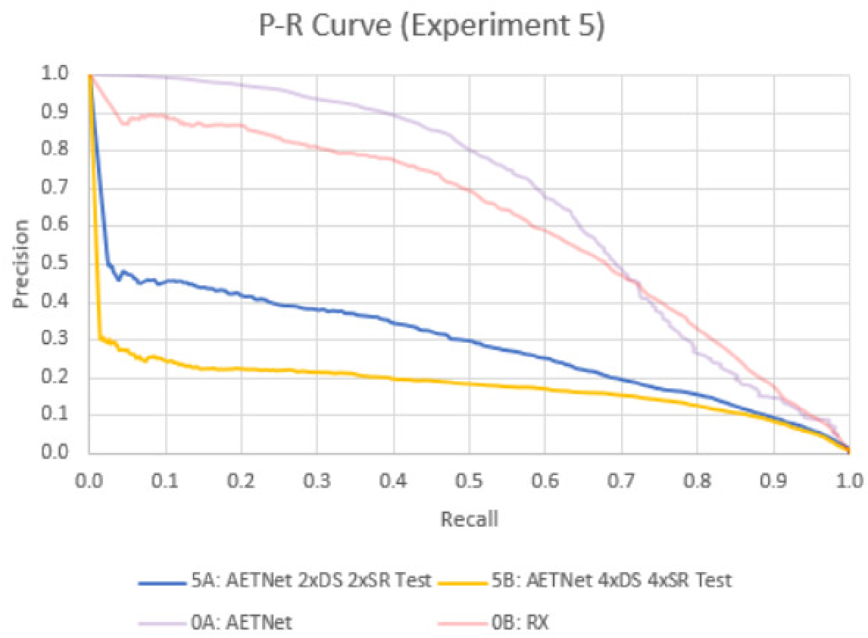


Figure 5.20: Experiment 5: Precision-Recall

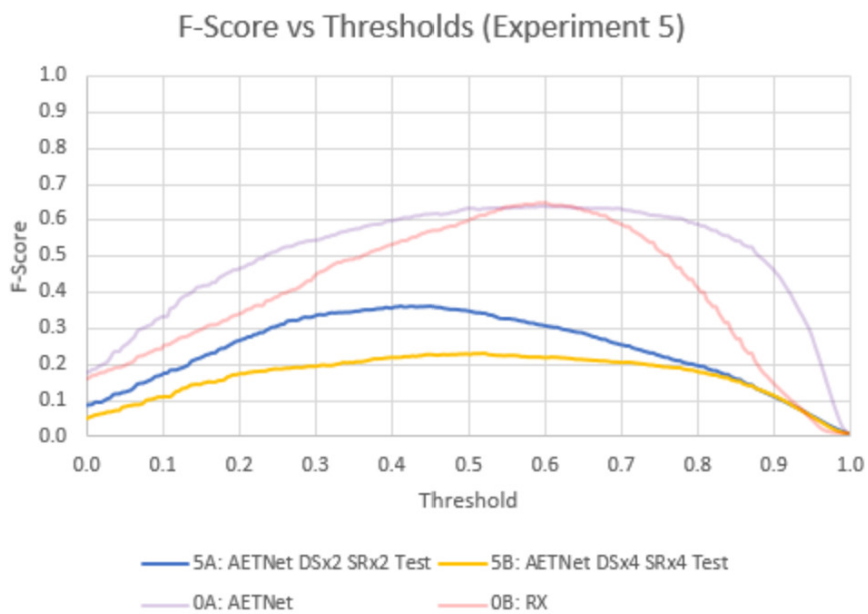
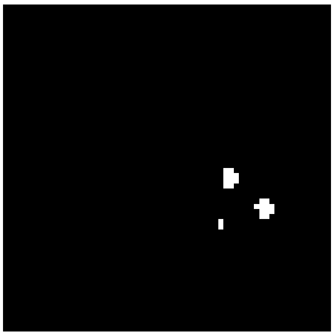
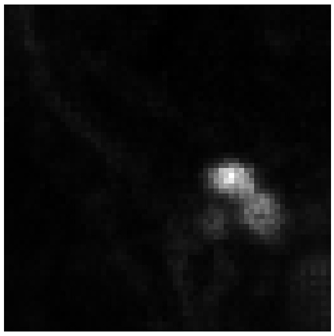


Figure 5.21: Experiment 5: F1 Score

(a) False colour map of HAD100 test image 118



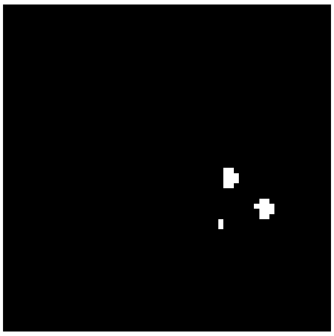
(b) Ground Truth



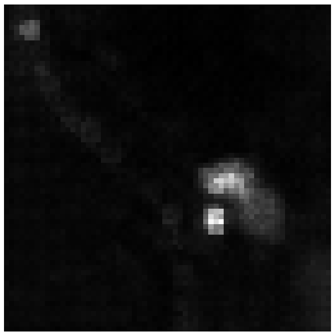
(c) Model Prediction

Experiment 5A Results

(d) False colour map of HAD100 test image 118



(e) Ground Truth



(f) Model Prediction

Experiment 5B Results

Figure 5.22: Experiment 5 Results: both models trained on enhanced images. (5A) scale factor 2, (5B) scale factor 4, with each test evaluated on a matching enhanced test set.

The results of Experiment 5 suggest that the model using data downsampled and then enhanced by a scale factor of 2 performs twice as well as the model sampled and enhanced with a scale factor of 4. However, both models show significantly lower performance when compared to the original AETNet and RX detectors. Both models experience a substantial drop in precision at low recall levels, which then stabilises and converges together at higher recall levels. When examining the F1 scores and corresponding thresholds, both models follow a similar curve pattern.

On a visual inspection, the results of Experiment 5B have noticeable blurring in the areas where anomalies are predicted suggesting the model may finds it difficult to distinguish between anomalies at increasing downsample scale factors.

5.0.26 Experiment 5 Discussion

Experiment 5 suggests that while image enhancement can improve model performance on low quality data, this performance increase directly depends on how significantly degraded the input imagery is. Even when increasing the super resolution factor the loss of data from downsampling is not recovered enough for an ML model to accurately make a prediction. This suggests that for very low data quality, it may be beneficial to use a traditional detector.

The large drop in precision at low recall levels suggests that whilst the AETNet does detect most of the anomalies within an image, it does this with a very high rate of false positives indicating that there must be careful consideration given to decided on a trade off in detection requirements.

The clear blurring of predictions seen in Experiment 5B suggests that the AETNet cannot effectively distinguish between anomalies following this level of downsampling and enhancing. The blurring suggests that details required to make precise and accurate detection's may have been lost in the initial downsampling process and that the super resolution has failed to recover this information. This emphasises the importance of ensuring the initial data is not too low quality that features are completely lost before the super resolution technique is applied, if this is the case then applying the super resolution is unlikely to aid in the accurate detection of anomalies using an ML method.

5.0.27 Experiment 6: RX Degraded Test Set

Experiment 6 examined the RX method for anomaly detection and its sensitivity to downsampling. In Test 6A, RX was applied to test images that had been downsampled by a scale factor of 2, while Test 6B involved downsampling by a scale factor of 4.

This experiment enables a comparison of how image quality impacts a conventional, non-ML technique, providing understanding into the effects of resolution on traditional anomaly detection methods. This evaluation is important to understand the limitations of the RX method when high quality data is unavailable, it also establishes a benchmark at varying scale factors for comparing a conventional technique with an ML technique. Test images used in Experiment 6A and B are the same as Experiment 1A and B respectively.

5.0.28 Experiment 6 Results

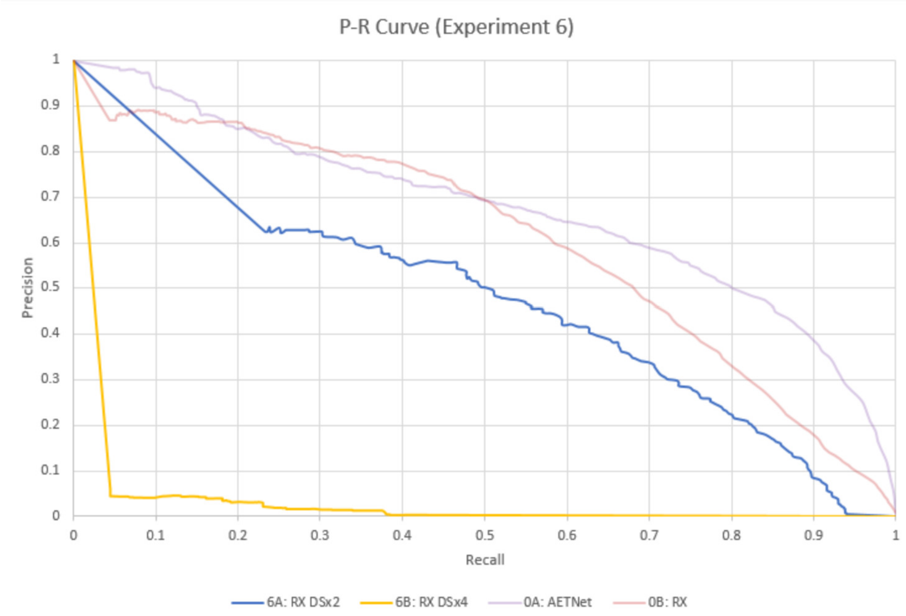


Figure 5.23: Experiment 6: Precision-Recall

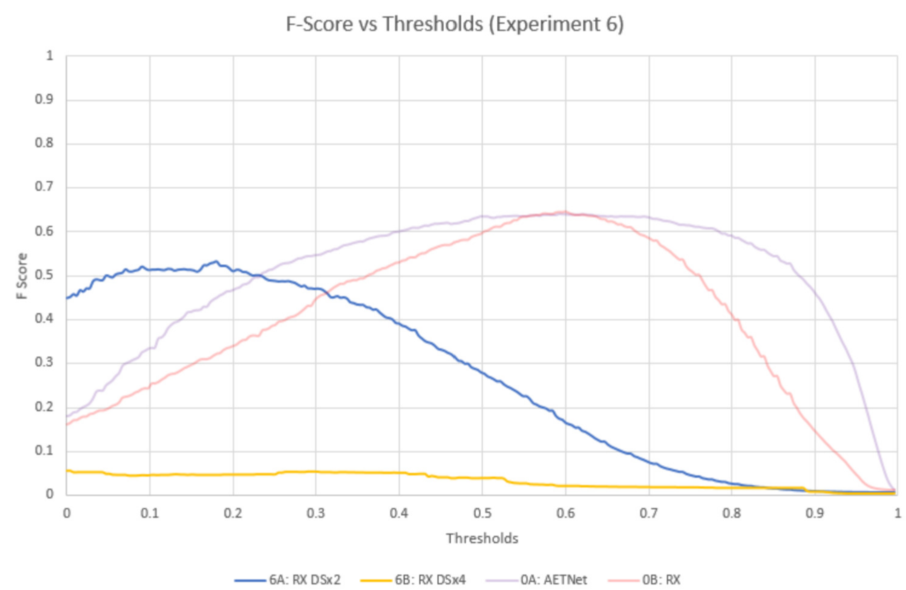
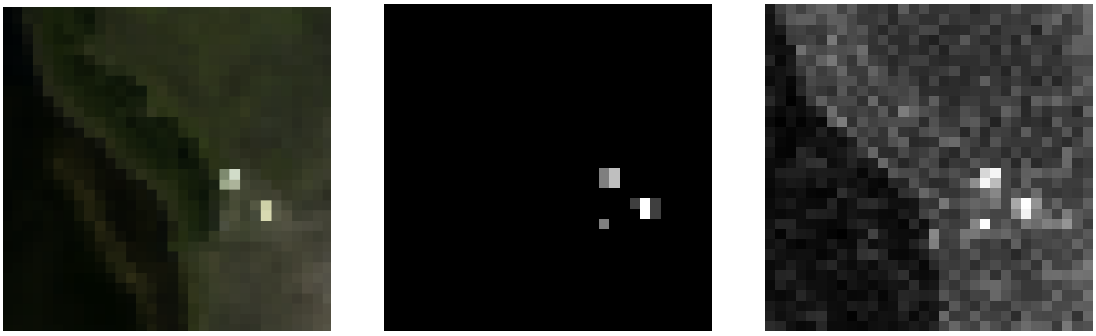


Figure 5.24: Experiment 6: F1 Score

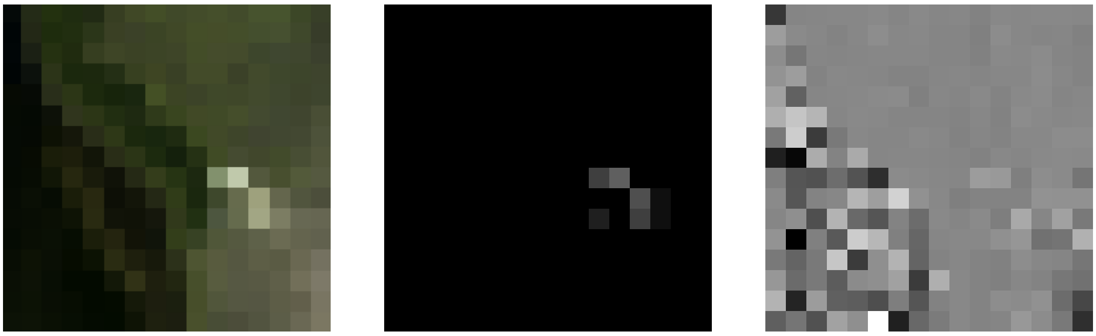


(a) False colour map of HAD100 test image 118

(b) Ground Truth

(c) Model Prediction

Experiment 6A Results



(d) False colour map of HAD100 test image 118

(e) Ground Truth

(f) Model Prediction

Experiment 6B Results

Figure 5.25: Experiment 6 Results: RX detector performance under downsampling with scale factors 2 (6A) and 4 (6B)

When examining the output model predictions of 6A, clear distinction can be made between the predicted anomalies. For 6B, the whole prediction appears cluttered and only background is distinguishable in the output map.

5.0.29 Experiment 6 Discussion

The results of Experiment 6 clearly demonstrate that the RX detector's performance deteriorates significantly with increased downsampling of the input image (with the caveat of suboptimal covariance calculations). Experiment 6A drastically outperformed Experiment 6B across all performance metrics measured and presents the highest recall score of any model assessed. The conclusion can be drawn that in small amounts downsampling can still preserve the information required for effective anomaly detection, whereas downsampling at larger scale factors may result in a loss of critical detail or information. Interestingly, Experiment 6A outperformed both the standard RX detector and the original AETNet model when applied at lower model threshold levels. This indicates that that a certain level of downsampling might actually enhance anomaly detection performance by filtering out unnecessary details whilst also maintaining enough information to accurately detect true anomalies. However use of low model thresholds will inevitably produce increasing false alarm detections and therefore should be used only when the use case deems this acceptable.

When visually analysing the model predictions of Experiment 6, the same conclusions can be drawn. Experiment 6A shows clearly distinguishable anomalies that align with the ground truth maps, whereas 6B cannot differentiate between anomalies and background in an image suggesting a loss of important information in its downsampling process. The results of this experiment highlight that the use of downsampling at appropriate scale factors may maintain and even increase the performance of a conventional anomaly detection technique. Conversely, the use of too much downsampling leads to significant information loss and a lack of accurate predictions.

5.0.30 Experiment 7: RX Enhanced Test Set

The final experiment focused on evaluating the RX detection method's performance on an enhanced test set with scale factors of both 2 and 4. By assessing how RX performs when applied to images that have been enhanced, this experiment provides insight into whether super resolution can improve the accuracy and a traditional anomaly detection technique that does not require advanced training. It enables a direct comparison between RX's performance on enhanced images and its performance on original and downsampled images, helping to determine if super resolution can increase detection performance. Test images used in Experiment 7A and B are the same as Experiment 5A and B respectively.

5.0.31 Experiment 7 Results

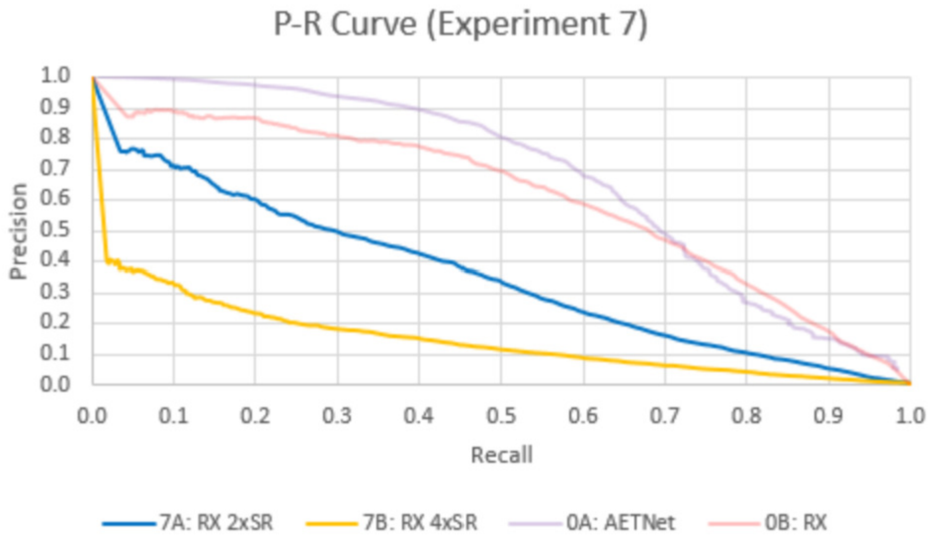


Figure 5.26: Experiment 7: Precision-Recall

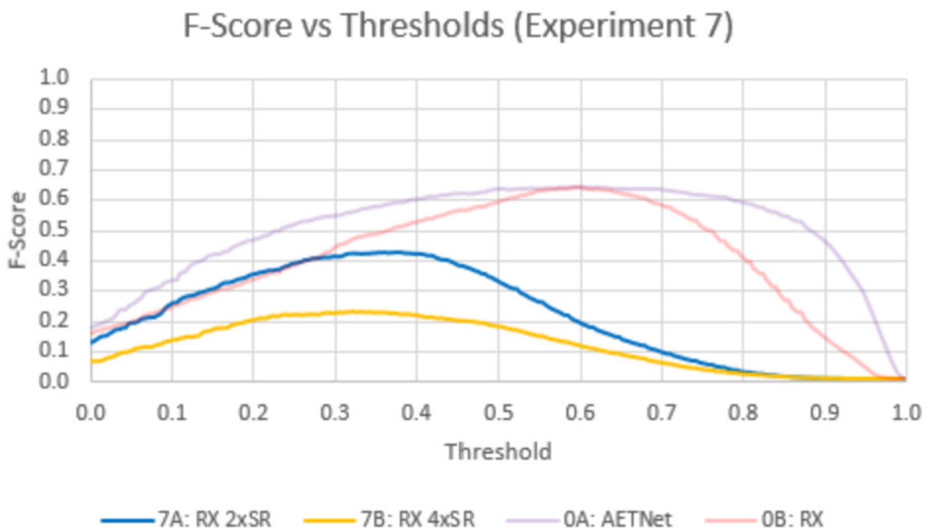
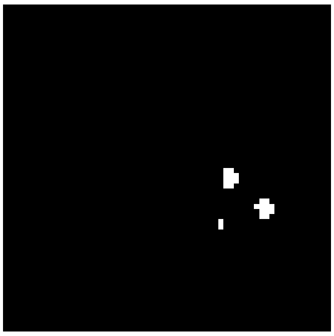
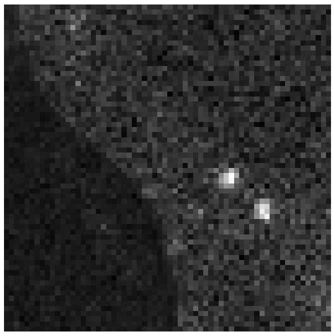


Figure 5.27: Experiment 7: F1 Score

(a) False colour map of HAD100 test image 118



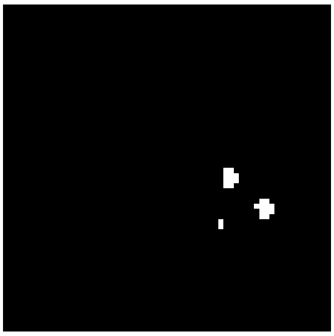
(b) Ground Truth



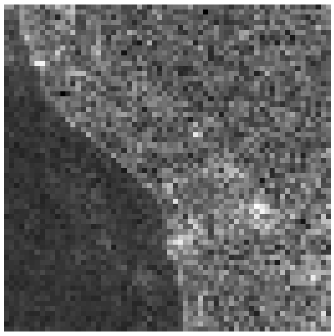
(c) Model Prediction

Experiment 7A Results

(d) False colour map of HAD100 test image 118



(e) Ground Truth



(f) Model Prediction

Experiment 7B Results

Figure 5.28: Experiment 7 Results: comparison of scale factor 2 (7A) and scale factor 4 (7B) enhanced models

Experiment 7 presents similar results to Experiment 6, the model that has been enhanced by scale factor 2 (7A) tends to perform twice as well as scale factor 4 (7B). However the performance decrease between the two scale factors is not as significant. There are also no thresholds at which either enhanced model significantly outperforms the original models.

Upon visual inspection of model outputs, model 7A outperforms 7B. 7A has clear and distinguishable anomalies, whereas 7B tends to be unable to differentiate background and anomalies.

5.0.32 Experiment 7 Discussion

Experiment 7 presents an interesting result in that the more enhancement is applied to the input images, the worse the model gets. The model using a scale factor of 2 (7A) outperforms the model using a higher scale factor of 4 (7B), demonstrating that perhaps excessive use of super resolution can actually degrade the performance of RX anomaly detection. This finding suggests that there is a limit to the benefits of image enhancement through super resolution. Beyond this limit, the added detail may introduce noise or artifacts that impede the model's ability to accurately identify anomalies.

The impact of super resolution scale factors on model accuracy may vary depending on specific experimental conditions, in Experiment 6, downsampling by a factor of 2 increases performance, suggesting that the performance of a model may not be linearly dependent on the scale factor and may vary depending on the preprocessing applied.

Model 7A produces visually clearer and more distinguishable anomalies, whereas model 7B shows cluttered and less distinct outputs and struggles to differentiate between anomalies and the image background. These observations suggest that careful selection of enhancement (or downsampling) levels is important to model performance, as too much super resolution can lead to a reduced model performance due to the obscuring of features that present as important for anomaly detection or due to the addition of increased noise.

5.1 Overarching Discussion

The experimental results from this study indicate that both traditional and ML models are effective at performing anomaly detection in HSI, even when dealing with different levels of image quality. Among the models tested, models AETNet (Original), 4A (No Downsample, SR Train Set, Original Test Set) and 6A (RX Test Set Downsample Factor 2) demonstrated the best performances. Model 4A, an AETNet, significantly outperformed all other experimented AETNet variants, however its F1 score performance was slightly inferior to the original AETNet. This suggests that applying super resolution techniques to data that is already of sufficient quality may not significantly boost the overall performance of ML models. However, model 4A performs at a higher precision than the original AETNet, suggesting that for use cases where a higher precision is necessary (such as in defence, where false positive detections are unacceptable), applying super-resolution to the AETNet model in this way may lead to increased performance.

Conversely, model 6A demonstrated the highest recall value of any model studied suggesting that when lower precision can be tolerated, a conventional model utilising lower quality data may be sufficient for predictions.

The worst performing model in this study was model 6B (an RX detector that utilised data downsampled by a scale factor of 4). This model performed poorly across all performance metrics, this result is likely due to the loss of critical information during the downsampling process. Since the RX detector relies on image statistics to perform anomaly detection, it may struggle to capture more complex features and patterns within an image. When applied to lower quality images, the RX model may not be able to effectively distinguish between normal and anomalous features, leading to the reduced overall performance. In contrast, the AETNet model at an equivalent downsample factor of 4 (model 2B) showed increased performance when compared to the RX detector, indicating that a machine learning technique may be more suitable for performing anomaly detection on lower quality images.

To further emphasise this point, the baseline RX model (0B) outperforms the downsampled RX models (6A & 6B) by larger margins than the AETNet outperforms its downsampled equivalents. When taking all of these factors into consideration, the conclusion can be made that the RX detector is more susceptible to the effects of image quality than the ML AETNet.

Overall, the results of this study highlight the importance of image quality on anomaly detection techniques for use with HSI. The findings suggest that while traditional methods hold value in specific contexts (often use case dependent), ML models may offer higher performance when input images are significantly degraded and their overall performance appears to be less susceptible to changes in image quality.

5.2 Performance Evaluation Table

The following table details the performance metrics obtained from the experiments conducted.

As shown below, the highest F1 Score was obtained in experiment number 0A (original AETNet). and the lowest MAE Score obtained in both experiments 0A and 4A (No Downsample, SR Train Set and Original Test). The highest precision value is exhibited by model 4A and the highest recall value by RX Model 6A (RX Test Set Downsample Factor 2). The lowest F1 Score and highest MAE score were both obtained from experiment number 6B (likely due to sub-optimal covariance estimations which may have also led to it’s outlying 0.710 recall value).

When performing the RX detection, a standard threshold of 0.6 was chosen as this gave the highest accuracy when conducting the baseline test.

Experiment Number	Description	F1 Score	MAE Score	Precision	Recall
0A	AETNet	0.505	0.028	0.688	0.469
0B	RX	0.405	0.080	0.656	0.417
1A	Downsample Test Set Factor 2	0.450	0.070	0.528	0.513
1B	Downsample Test Set Factor 4	0.321	0.196	0.339	0.581
2A	Downsample Train & Test Set Factor 2	0.304	0.090	0.360	0.424
2B	Downsample Train & Test Set Factor 4	0.234	0.274	0.251	0.561
3A	Test Set No Downsample & SR Factor 2	0.419	0.041	0.545	0.473
3B	Test Set Downsample Factor 2 & SR Factor 2	0.262	0.086	0.342	0.437
3C	Test Set Downsample Factor 4 & SR Factor 4	0.236	0.099	0.300	0.437
4A	No Downsample, SR Train Set & Original Test Set	0.503	0.028	0.690	0.394
4B	Downsample Factor 2, SR Train Set Factor 2 & Original Test Set	0.466	0.030	0.611	0.512
4C	Downsample Factor 4, SR Train Set Factor 4 & Original Test Set	0.374	0.038	0.502	0.500
5A	Downsample Factor 2, SR Train & Test Set Factor 2	0.242	0.086	0.304	0.450
5B	Downsample Factor 4, SR Train & Test Set Factor 4	0.169	0.090	0.197	0.388
6A	RX Test Set Downsample Factor 2	0.270	0.173	0.273	0.710
6B	RX Test Set Downsample Factor 4	0.033	0.516	0.032	0.217
7A	RX Test Set SR Factor 2	0.222	0.270	0.343	0.532
7B	RX Test Set SR Factor 4	0.123	0.306	0.179	0.462

Table 5.1: Performance Metrics

5.3 Research Implications

The following section discuss three key emerging implications that have been determined from the work within this thesis.

5.3.1 Storage & Compute

The management of storage and compute resource is a key consideration when using anomaly detection techniques especially when dealing with large datasets such as HSI. As demonstrated in this study, a strategy that can be employed to address storage and compute constraints is the use of downsampling. Downsampling can significantly and predictably decrease the amount of data that needs to be stored and processed, as the number of pixels reduces approximately by a factor of s^2 for a downsampling scale factor s. This makes the datasets more manageable for real-time processing and can reduce the computational complexity of techniques.

Despite the decrease in data size, downsampling can still deliver adequate results in HSI anomaly detection depending on the users requirements (see results of Experiments 1, 2 and 6). By selecting an appropriate downsampling scale factor, the data can retain sufficient information to enable conventional or ML techniques to detect anomalies successfully. In addition, since computational cost for the techniques employed in this study scales approximately linearly with the number of pixels, a scale factor of s results in roughly an $s^2\times$ reduction in compute. Implementing downsampling may therefore enable specific models to accurately detect anomalies whilst also operating effectively on systems that may have limited compute or storage capacities.

5.3.2 Conventional vs ML Techniques

As demonstrated in the experiments of this study, both ML and conventional techniques can be effective at performing anomaly detection in HSI, however their suitability for the task often depends on the specific use case and false positive tolerance. Traditional methods tend to offer more interpretable and explainable results than ML methods as they utilise statistical information to perform anomaly detection. The clarity and ease of use of conventional techniques may make them preferable for use when anomaly and background characteristics are already well defined and when trust and transparency in the detection process is required.

Conversely, ML techniques provide a capability that can more effectively handle complex data and extract meaningful relationships within this data. ML algorithms can apply learning from large HSI datasets to identify non-linear relationships that conventional techniques may be unable to pick up on. The choice of whether to use a conventional or ML technique should be determined by the specific use case taking into account cost of deployment and false positive rate. For scenarios where more false positives can be tolerated (e.g. environmental monitoring), ML methods may be the most suitable due to their increased ability to generalise across scenarios. If false positives must be minimised and increased explainability required (e.g. military contexts), then conventional methods may be preferred due their simplicity and deterministic characteristics.

5.3.3 Super Resolution

The results of this thesis demonstrate that super resolution can be utilised to improve model performance in HSI anomaly detection (see results of Experiments 3, 4, 5 and 7), it may therefore be beneficial in some scenarios to apply this technique to HSI datasets with the goal to improve model accuracy.

The use of super resolution in defence and security contexts can carry increased risk. Models used to perform super resolution essentially "hallucinate" and create details that are not present in the original image to perform the enhancement. Super resolution algorithms interpolate and generate pixels based on statistical predictions in order to creating a higher resolution image that appears more detailed. As a result of these generations, the technique has been shown to produce image artifacts and inaccuracies, in a worse case scenario the model may generate misleading information that could be perceived by an untrained analyst as real. This is particularly problematic in fields where incorrect detail could lead to significant consequences.

This risk of hallucinated details in ML SR can cause moral and ethical concerns if enhanced images are used in a decision making or product development pipeline. Whilst super resolution presents a well documented technique to enhance images, its limitations must be considered for implementation on real-world use cases. Ensuring that there is correct tradecraft and standards implemented for use of the technique are essential steps to ensure reliability of the use of this technique.

Chapter 6

Conclusions

6.1 Summary of Findings

This thesis set out to investigate the impact of image quality on the performance of conventional signal processing and ML techniques when performing Hyperspectral anomaly detection. The primary objective was to develop an innovative experiment plan to compare and contrast the performance of these models at different levels of imagery degradation and enhancement to understand how the performance of anomaly detection models is impacted.

The first key finding of the study, was that the performance of the anomaly detectors was not monotonically dependent on the scale factor by which the images were degraded or enhanced. As demonstrated in Experiments 5, 6 and 7, there is a significant decrease in performance when imagery is degraded and an increase in performance when imagery is firstly downsampled then enhanced, however there appears no straightforward correlation between final performance metrics calculated.

This result indicated that other factors may also be playing a critical role in influencing the models to predict anomalies. This new finding challenges the common assumption that a higher spatial resolution image will always lead to better model performance, emphasising the need for use-case dependent analysis techniques.

In support of this conclusion, the experiments that utilised downsampling to symbolise image quality degradation demonstrated that both the traditional and ML techniques studied in this thesis can still perform effective anomaly detection using lower quality images. However, model performance did improve when enhancement techniques were implemented on these lower quality images, suggesting that marginal gains can be achieved if super resolution is applied at an appropriate scale factor.

A second conclusion that can be drawn from this research is related to the performance comparison of the tradition RX detector and the ML AETNet. At a model threshold of 0.6, both models (trained & tested on original data) perform exactly the same. This result is key when determining which type of model to use for a specific use case. The RX detector is most suitable for situations in which high precision and low recall is preferred, for example, in a defence or targeting context, where the need to accurately identify and detect only specific targets (high precision) is important, even if it means missing some smaller, less obvious targets (low recall). AETNet performs superior when high recall and low precision is needed, for example, in deforestation identification where the aim is to identify all areas of potential deforestation (high recall), even if it means some areas are flagged incorrectly as deforested (low precision). At all other thresholds, the AETNet model outperforms the RX detector at balancing the precision and recall for ideal detection performance.

A third key conclusion that can be drawn from this study comes from the results of Experiment 6A. The results of this experiment demonstrate that the downsampled (scale factor 2) RX detector outperforms both the original RX detector and original AETNet at low model thresholds. The downsampled RX also retains comparable precision and recall performance to the two original models.

This result is notable as it suggests that downsampling may actually enhance detection performance under certain conditions when recall is desired over precision. This reduction in resolution reduces the computational complexity needed to perform anomaly detection whilst still maintaining enough information for a model to perform accurately. This indicates that lower resolution data, which is likely more manageable and less resource-intense, can still achieve sufficient levels of detection performance for specific use cases and may be more appropriate for deployment under varying conditions.

Perhaps the conclusion with the most real-world impact within this thesis is the increased precision value for Model 4A which was trained on data enhanced with super-resolution. For use cases where precision is valued above all else (where any misclassification can cause issue), enhancing imagery before training may provide increased performance as demonstrated by the results of this thesis.

The novel insights provided from this conclusion can be used as guidance when selecting the appropriate detection methods based on the specific needs of a given use case.

6.2 Research Gaps

Despite the results from this study, several research gaps remain that could impact the development and application of both conventional and ML techniques. Examples of some key gaps are detailed in this section.

6.2.1 Target Detection & Identification

This study specifically focuses on anomaly detection in HSI. It does not take into account use cases where specific target detection based on a given target spectra is required.

Anomaly detection (as performed in this study) aims to find anomalies within images without prior knowledge of what these anomalies might be made up of. This detection technique is particularly useful for the detection of unknown and unexpected features within a dataset making it ideal for applications such as environmental monitoring where specific target information is not readily available.

6.2.2 Dataset Availability

One of the key reasons for exploring unidentified anomalies versus known target detection was the lack of labelled HSI datasets available for research use. The scarcity of labelled HSI data presents a major challenge in developing and testing anomaly detection models. Labelled data is often required to train ML algorithms to learn specific features of datasets. Without sufficient labelled data, models lack performance and may overfit to limited examples leading to reduced accuracy.

The process of obtaining labelled HSI data is time consuming and labour intensive and accurately annotating images with specific targets requires specialist knowledge. All of these factors contribute to the lack of availability of labelled datasets in the open source domain resulting in challenges in advancing the development of new algorithms.

To address this challenge, several strategies can be implemented. Research is being conducted into the use of synthetic data [89], unsupervised learning [90] and transfer learning [91] to reduce the fields' reliance on specifically labelled datasets.

6.2.3 Image Degradation Techniques

The aim of this study was to investigate the impact of image quality on the performance of conventional signal processing and ML techniques. An image degradation technique was implemented in order simulate lower quality images however the technique chosen was a simplistic downsampling approach that could be improved in future research.

Implementing multiple image degradation techniques would assist in lowering image quality in various ways, such as adding blurring, noise, or compression artifacts. By applying multiple types of degradation, increased understanding of how detection methods respond to types of decreased image quality could be studied more in-depth.

Only the downsampling technique was utilised in this study to aid in gaining an understanding of how lower spatial resolution affects the performance of anomaly detection models. This focus enables the isolation of this degradation technique and allows the use of super resolution techniques to directly address the effects of downsampling.

6.3 Future Direction

Whilst this study has made contributions to understanding the impact of image quality on both conventional and ML HSI anomaly detection techniques, several avenues remain open for future research. One such area would be studies into how spectral band distribution affects anomaly detection model performance. Research could investigate the optimal band distributions selected for feeding into a model that maximises the accuracy. Understanding the contributions of various bands to the anomaly detection within an image would enable researchers to build more efficient model pipelines and may present a more explainable view of how anomalies were predicted.

To further develop the research from this study, a next step could be incorporating additional super resolution techniques. Research could explore more state of the art super resolution methods used enhance image quality to compare their performance on HSI. Different super resolution techniques present varying strengths and weaknesses, e.g. Method A may perform well when preserving small spatial details, whereas Method B may excel at reducing image noise. By studying a wider pool of super resolution techniques, the model that positively impacts detection performance the most can be selected depending on use case requirements.

A final suggestion for future study would be further comparison between a wider selection of both ML and conventional techniques. By expanding beyond commonly used methods, rarer techniques could reveal insights into optimising automatic anomaly detection performance. Continuing comparative study between ML and traditional methods allows researchers to further understand where the benefits of utilising these respective techniques lie.

By investigating these future directions for research, more robust and effective anomaly detection techniques for HSI can be developed and tailored to specific real-world use cases, building upon the novel research presented in this thesis.

Bibliography

[1] M. Paoletti, J. Haut, J. Plaza, and A. Plaza, "Deep learning classifiers for hyperspectral imaging: A review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 158, pp. 279–317, 2019.

[2] N. R. S. et al, "Hyperspectral anomaly detection: A performance comparison of existing techniques," *International Journal of Digital Earth*, vol. 15, no. 1, pp. 2078–2125, 2022.

[3] J. M. Amigo, "Hyperspectral and multispectral imaging: Setting the scene," in *Data Handling in Science and Technology*, J. M. Amigo, Ed., vol. 32, Elsevier, 2019, pp. 3–16.

[4] K. Tan, Z. Hou, F. Wu, Q. Du, and Y. Chen, "Anomaly detection for hyperspectral imagery based on the regularized subspace method and collaborative representation," *Remote Sensing*, vol. 11, no. 11, p. 1318, 2019.

[5] R. F. Kokaly et al., "Usgs spectral library version 7," U.S. Geological Survey, Data Series 1035, 2017.

[6] G. Licciardi, M. 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.

[7] G. L. D, "Mean, median and mode filtering of images," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 456, pp. 2995–3004, 2000.

[8] C. R. Dyer, "Multiscale image understanding," University of Wisconsin-Madison, Department of Computer Sciences, Tech. Rep. TR679, 1986.

[9] D. Dumitrescu and C.-A. Boiangiu, "A study of image upsampling and downsampling filters," *Computers*, vol. 8, no. 2, 2019.

[10] N. Salem and S. Hussein, "Data dimensional reduction and principal components analysis," *Procedia Computer Science*, vol. 163, pp. 292–299, 2019.

[11] S. Gao, X. Jiang, Y. Zhang, and et al., "Superpixelwise pca based data augmentation for hyperspectral image classification," *Multimedia Tools and Applications*, vol. 83, pp. 81 209–81 229, 2024, Received: 14 September 2023; Revised: 06 January 2024; Accepted: 15 February 2024; Published: 09 March 2024.

[12] C.-I. Chang, "Recursive hyperspectral band processing of orthogonal subspace projection," in *Real-Time Recursive Hyperspectral Sample and Band Processing: Algorithm Architecture and Implementation*. Cham: Springer International Publishing, 2017, pp. 483–503.

[13] O. S. Chander Goud, T. H. Sarma, and C. S. Bindu, "Optimal band selection in hyperspectral images using improved k-means clustering with spectral similarity measures," in *2023 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*, 2023, pp. 268–271.

[14] R. Yang, L. Su, X. Zhao, H. Wan, and J. Sun, "Representative band selection for hyperspectral image classification," *Journal of Visual Communication and Image Representation*, vol. 48, pp. 396–403, 2017.

[15] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd. Upper Saddle River, NJ: Prentice Hall, 2008.

[16] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," *arXiv:1711.10925*, 2017.

[17] C. Zhang and K. S. Yen, "A survey on deep image prior for image denoising," *Digital Signal Processing*, vol. 163, p. 105 235, 2025, Available online 9 April 2025, Version of Record 14 April 2025.

[18] S. Bell-Kligler, A. Shocher, and M. Irani, "Blind super-resolution kernel estimation using an internal-gan," in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds., 2019, pp. 284–293.

[19] M. S. Rad, T. Yu, C. Musat, H. K. Ekenel, B. Bozorgtabar, and J.-P. Thiran, *Benefiting from bicubically down-sampled images for learning real-world image super-resolution*, 2020.

[20] A. Shocher, N. Cohen, and M. Irani, "“zero-shot” super-resolution using deep internal learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 3118–3126.

[21] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Advances and challenges in super-resolution," *International Journal of Imaging Systems and Technology*, vol. 14, no. 2, pp. 47–57, 2004.

[22] X. Jin, S. Paswaters, and H. Cline, "A comparative study of target detection algorithms for hyperspectral imagery," in *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XV*, S. S. Shen and P. E. Lewis, Eds., International Society for Optics and Photonics, vol. 7334, SPIE, 2009, 73341W.

[23] M. L. Pieper, D. G. Manolakis, R. B. Lockwood, T. W. Cooley, P. S. Armstrong, and J. Jacobson, "Hyperspectral detection and discrimination using the ace algorithm," in *Optical Engineering + Applications*, 2011.

[24] E. Truslow, "Performance evaluation of the adaptive cosine estimator detector for hyperspectral imaging applications," Master's Thesis, Northeastern University, Boston, Massachusetts, 2012.

[25] S. Matteoli, M. Diani, and G. Corsini, "A tutorial overview of anomaly detection in hyperspectral images," *IEEE Aerospace and Electronic Systems Magazine*, vol. 25, no. 7, pp. 5–28, 2010.

[26] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza, "Yinyang k-means clustering for hyperspectral image analysis," in *Proceedings of the 17th International Conference on Computational and Mathematical Methods in Science and Engineering (CMMSE 2017)*, Avda. de la Universidad s/n, Spain, 2017.

[27] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, ser. KDD'96, Portland, Oregon: AAAI Press, 1996, pp. 226–231.

[28] A. Chatterjee, S. Ghosh, and A. Ghosh, "Band selection for hyperspectral image classification using dbscan-informed spectral attention network," in *2025 IEEE International Conference on Next-Gen Technologies of Artificial Intelligence and Geoscience Remote Sensing (EarthSense)*, 2025, pp. 1–5.

[29] N. Sammaknejad, Y. Zhao, and B. Huang, “A review of the expectation maximization algorithm in data-driven process identification,” *Journal of Process Control*, vol. 73, pp. 123–136, 2019.

[30] D. A. Reynolds et al., “Gaussian mixture models.,” *Encyclopedia of biometrics*, vol. 741, no. 659-663, 2009.

[31] A. Kartakoullis, N. Caporaso, M. B. Whitworth, and I. D. Fisk, “Gaussian mixture model clustering allows accurate semantic image segmentation of wheat kernels from near-infrared hyperspectral images,” *Chemometrics and Intelligent Laboratory Systems*, vol. 259, p. 105 341, 2025.

[32] M. Veganzones, J. Frontera-Pons, J. Chanussot, and J. Ovarlez, “On the use of the hotelling’s t2 statistic for the hierarchical clustering of hyperspectral data,” in *2013 5th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2013, pp. 1–4.

[33] H. Hotelling, “Multivariate quality control,” in *Techniques of Statistical Analysis*, C. Eisenhart, M. W. Hastay, and W. A. Wallis, Eds., New York: McGraw-Hill, 1947.

[34] C.-I. Chang and S.-S. Chiang, “Anomaly detection and classification for hyperspectral imagery,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 6, pp. 1314–1325, 2002.

[35] J. Frontera-Pons, M. A. Veganzones, F. Pascal, and J.-P. Ovarlez, “Hyperspectral anomaly detectors using robust estimators,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 2, pp. 720–731, 2016.

[36] S. Küçük and S. E. Yüksel, “Comparison of rx-based anomaly detectors on synthetic and real hyperspectral data,” in *2015 7th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2015, pp. 1–4.

[37] G. Jaiswal, R. Rani, H. Mangotra, and A. Sharma, “Integration of hyperspectral imaging and autoencoders: Benefits, applications, hyperparameter tunning and challenges,” *Computer Science Review*, vol. 50, p. 100 584, 2023.

[38] X. Hu et al., “Hyperspectral anomaly detection using deep learning: A review,” *Remote Sensing*, vol. 14, no. 9, 2022.

[39] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors,” *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[40] M. Z. Alom et al., "The history began from alexnet: A comprehensive survey on deep learning approaches," *CoRR*, vol. abs/1803.01164, 2018.

[41] K. Bhosle and V. Musande, "Evaluation of deep learning cnn model for land use land cover classification and crop identification using hyperspectral remote sensing images," *Journal of the Indian Society of Remote Sensing*, vol. 47, no. 11, pp. 1949–1958, 2019, Received 05 April 2019; Accepted 28 August 2019; Published 13 September 2019.

[42] T. T. Pham et al., "Airborne object detection using hyperspectral imaging: Deep learning review," in *Computational Science and Its Applications–ICCSA 2019: 19th International Conference, Saint Petersburg, Russia, July 1–4, 2019, Proceedings, Part I 19*, Springer, 2019, pp. 306–321.

[43] A. Maffei, J. M. Haut, M. E. Paoletti, J. Plaza, L. Bruzzone, and A. Plaza, "A single model cnn for hyperspectral image denoising," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 4, pp. 2516–2529, 2019.

[44] Y. Li, H. Zhang, and Q. Shen, "Spectral–spatial classification of hyperspectral imagery with 3d convolutional neural network," *Remote Sensing*, vol. 9, no. 1, 2017.

[45] M. Ahmad, M. Mazzara, and S. Distefano, "Regularized cnn feature hierarchy for hyperspectral image classification," *Remote Sensing*, vol. 13, no. 12, p. 2275, 2021.

[46] A. Vaswani et al., "Attention is all you need," in *Advances in Neural Information Processing Systems*, vol. 30, Curran Associates, Inc., 2017.

[47] A. Dosovitskiy et al., *An image is worth 16x16 words: Transformers for image recognition at scale*, 2021.

[48] K. Han et al., "A survey on vision transformer," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 1, pp. 87–110, 2023.

[49] L. Deininger et al., *A comparative study between vision transformers and cnns in digital pathology*, Jun. 2022.

[50] Z. Liu et al., *Swin transformer: Hierarchical vision transformer using shifted windows*, 2021.

[51] S. Xiao, T. Zhang, Z. Xu, J. Qu, S. Hou, and W. Dong, "Anomaly detection of hyperspectral images based on transformer with spatial–spectral dual-window mask," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 1414–1426, 2023.

[52] Y. Xu, K. Zhao, L. Zhang, M. Zhu, and D. Zeng, "Hyperspectral anomaly detection with vision transformer and adversarial refinement," *International Journal of Remote Sensing*, vol. 44, no. 13, pp. 4034–4057, 2023.

[53] Z. Li, Y. Wang, C. Xiao, Q. Ling, Z. Lin, and W. An, "You only train once: Learning a general anomaly enhancement network with random masks for hyperspectral anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–18, 2023.

[54] G. Shaw and D. Manolakis, "Signal processing for hyperspectral image exploitation," *IEEE Signal Processing Magazine*, vol. 19, no. 1, 2002.

[55] S. Prasad and J. Chanussot, *Hyperspectral image analysis: advances in machine learning and signal processing*. Springer Nature, 2020.

[56] P. Ghamisi et al., "Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 4, 2017.

[57] S. D and S. I. A., "Preprocessing image data for deep learning," in *Sinteza 2020 - International Scientific Conference on Information Technology and Data Related Research*, 2020, pp. 312–317.

[58] D. Cozzolino, P. Williams, and L. Hoffman, "An overview of pre-processing methods available for hyperspectral imaging applications," *Microchemical Journal*, vol. 193, p. 109 129, 2023.

[59] D. Stein, S. Beaven, L. Hoff, E. Winter, A. Schaum, and A. Stocker, "Anomaly detection from hyperspectral imagery," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 58–69, 2002.

[60] H. Kwon and N. Nasrabadi, "Kernel rx-algorithm: A nonlinear anomaly detector for hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 388–397, 2005.

[61] Z. Huang, L. Fang, and S. Li, "Subpixel-pixel-superpixel guided fusion for hyperspectral anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 9, pp. 5998–6007, 2020.

[62] R. Vaddi and P. Manoharan, “Hyperspectral image classification using cnn with spectral and spatial features integration,” *Infrared Physics & Technology*, vol. 107, p. 103 296, 2020.

[63] M. Ben Salem, K. S. Ettabaa, and M. A. Hamdi, “Anomaly detection in hyperspectral imagery: An overview,” in *International Image Processing, Applications and Systems Conference*, 2014, pp. 1–6.

[64] X. Hu et al., “Hyperspectral anomaly detection using deep learning: A review,” *Remote Sensing*, vol. 14, no. 9, p. 1973, 2022.

[65] Y. Zhou, X. Song, Y. Zhang, F. Liu, C. Zhu, and L. Liu, “Feature encoding with autoencoders for weakly supervised anomaly detection,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 6, 2022.

[66] D. Sculley and C. Brodley, “Compression and machine learning: A new perspective on feature space vectors,” in *Data Compression Conference (DCC’06)*, 2006, pp. 332–341.

[67] F. Luo, L. Zhang, B. Du, and L. Zhang, “Dimensionality reduction with enhanced hybrid-graph discriminant learning for hyperspectral image classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 8, pp. 5336–5353, 2020.

[68] S. Ghaderizadeh, D. Abbasi-Moghadam, A. Sharifi, N. Zhao, and A. Tariq, “Hyperspectral image classification using a hybrid 3d-2d convolutional neural networks,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 7570–7588, 2021.

[69] Z. Li and Y. Zhang, “Hyperspectral anomaly detection based on improved rx with cnn framework,” in *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, 2019.

[70] W. Li, G. Wu, and Q. Du, “Transferred deep learning for anomaly detection in hyperspectral imagery,” *IEEE Geoscience and Remote Sensing Letters*, vol. PP, pp. 1–5, Mar. 2017.

[71] F. Rembold, C. Atzberger, I. Savin, and O. Rojas, “Using low resolution satellite imagery for yield prediction and yield anomaly detection,” *Remote Sensing*, vol. 5, no. 4, pp. 1704–1733, 2013.

[72] P. Xiang, S. Ali, S. K. Jung, and H. Zhou, “Hyperspectral anomaly detection with guided autoencoder,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, 2022.

[73] X. Kang, *Data sets*, Accessed: 2024-08-27, 2024.

[74] Jet Propulsion Laboratory, *AVIRIS: Airborne Visible/Infrared Imaging Spectrometer*, Accessed: 2024-08-27, 2024.

[75] B. Tu, N. Li, Z. Liao, X. Ou, and G. Zhang, "Hyperspectral anomaly detection via spatial density background purification," *Remote Sensing*, vol. 11, p. 2618, Nov. 2019.

[76] D. Snyder, J. Kerekes, I. Fairweather, R. Crabtree, J. Shive, and S. Hager, "Development of a web-based application to evaluate target finding algorithms," in *Proceedings of the 2008 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 2, Boston, MA, 2008, pp. 915–918.

[77] T. Cocks, R. Jenssen, A. Stewart, I. Wilson, and T. Shields, "The hymap airborne hyperspectral sensor: The system, calibration and performance," in *Proceedings of the 1st EARSeL Workshop on Imaging Spectroscopy*, M. Schaepman, D. Schläpfer, and K. I. Itten, Eds., Zurich: EARSeL, Paris, 1998, pp. 37–42.

[78] S. S. Jha, M. Kumar, and R. R. Nidamanuri, "Multi-platform optical remote sensing dataset for target detection," *Data in Brief*, vol. 33, p. 106 362, 2020.

[79] Jet Propulsion Laboratory, California Institute of Technology, *Aviris-ng: Airborne visible infrared imaging spectrometer next generation*, Accessed: 2024-08-22, 2024.

[80] S. Bell-Kligler, A. Shocher, and M. Irani, *Kernelgan: Blind super-resolution kernel estimation using an internal-gan*, `https://github.com/sefibk/KernelGAN`, Accessed: 2024-08-23, 2019.

[81] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, 2010.

[82] Y. Zhang, D. Zhou, S. Chen, S. Gao, and Y. Ma, "Single-image crowd counting via multi-column convolutional neural network," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[83] J. Davis and M. Goadrich, "The relationship between precision-recall and roc curves," in *Proceedings of the 23rd International Conference on Machine Learning (ICML)*, 2006, pp. 233–240.

[84] T.-Y. Lin et al., "Microsoft coco: Common objects in context," in *European Conference on Computer Vision (ECCV)*, 2014.

[85] P. Flach and M. Kull, "Precision-recall-gain curves: Pr analysis done right," in *Advances in Neural Information Processing Systems*, C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, Eds., vol. 28, Curran Associates, Inc., 2015.

[86] J. J. Danker Khoo, K. H. Lim, and J. T. Sien Phang, "A review on deep learning super resolution techniques," in *2020 IEEE 8th Conference on Systems, Process and Control (ICSPC)*, 2020, pp. 134–139.

[87] J. Park, S. Stepney, and I. d'Amico, "A methodology for comparing and benchmarking quantum devices," in *UCNC*, 2023, pp. 28–42.

[88] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, 2016.

[89] Y. Lu, D. Perez, M. Dao, C. Kwan, and J. Li, "Deep learning with synthetic hyperspectral images for improved soil detection in multispectral imagery," in *2018 9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, 2018, pp. 666–672.

[90] S.-S. Chiang, C.-I. Chang, and I. Ginsberg, "Unsupervised target detection in hyperspectral images using projection pursuit," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 7, pp. 1380–1391, 2001.

[91] "Unsupervised transfer learning for target detection from hyperspectral images," *Neurocomputing*, vol. 120, pp. 72–82, 2013, Image Feature Detection and Description.