Adaptive lighting based Detection of Cracks and Spalling in concrete structures

Sanjeetha Pennada

Department of Civil and Environmental Engineering

University of Strathclyde

Glasgow, United Kingdom

Thesis submitted in fulfilment for requirements of the degree of

Doctor of Philosophy

May 2025

Declaration of Authenticity and Author's Rights

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.

The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Signed: Sanjeetha Pennada

Date: May, 2025

Abstract

Defects such as cracks and spalling in reinforced concrete can have a detrimental effect on structural integrity if unnoticed and left untreated. Inspection methods have largely evolved over the years from traditional visual inspections to partially and fully automated inspections. One of the most common forms of automated inspections is the use of cameras and application of deep learning algorithms to identify cracks and spalling on the captured images. Automated inspections overcome limitations of visual inspections, mainly as it concerns inconsistency and subjectivity in the identification and interpretation of cracks. Still, they do not come without their own caveats: autonomous vehicles (e.g. drones) require an expert, trained and licenced user, the captured images use diffused lighting that does not necessarily reveal the true geometric characteristics of the crack or even the crack itself, and most deep learning algorithms developed for crack detection and identification are not performing well with low quality images and low light conditions and are mainly trained with diffused light images captured in the lab. This thesis is focusing on addressing all of these challenges in the technology of automated concrete structure inspections in Civil Engineering through the development of both hardware and software.

Utilising the benefits of multi-directional, multi-angle lighting, which has been successfully used in other disciplines (medicine) in highlighting features clearly, an automated platform called ALICS (Adaptive Lighting for Inspection of Concrete Structures) has been developed. It captures images of concrete structures by creating and casting shadows through illuminating light from multiple angles and multiple directions, mimicking what an experienced inspector would do during a visual inspection. The original

design of ALICS has subsequently been modified to field-deployable hardware, suitable for carrying out inspections outside the lab. Following a number of experimental setups where the angle of light was varied and analysis of the captured images using a pretrained VGG-16 neural network model, the optimum angle was identified.

To overcome declined performance of the VGG-16 model, one of the most commonly used models in concrete crack identification in Civil Engineering, in the case of low-quality images, an automated image quality assessment workflow was developed based on BRISQUE (Blind/Referenceless Image Spatial Quality Evaluator) score and incorporated into the VGG-16 model. The traditional VGG-16 and VGG-19 neural network models are developed to satisfy three-channel diffused RGB images. To allow for the analysis of directional lighting, i.e. for five channels, one for the gray scale image in Right, Down, Left, Up and diffused directions each, novel five-channel VGG-16 and VGG-19 were developed, capable to detect and classify cracks.

The VGG-16 five-channel neural network was further improved in terms of evaluation time by utilising the light neural network model, MobileNetV2. The multi-channel MobileNetV2 model, MCNet has been implemented similar to five-channel VGG models. The maximum-intensity fusion technique, which has been successfully used in medical field to combine multiple images has been utilised to implement a fused neural network model, FusedNN. The FusedNN, and MCNet models are developed to satisfy three-channel RGB images, and five-channel images, respectively. The performance of the traditional MobileNetV2 model, FusedNN, the five-channel VGG-16 model, and MCNet were compared for crack detection and classification tasks, and MCNet demonstrated the best performance. For detecting cracks and spalling, a comparison of the four models: traditional, Zoubir, FusedNN, and MCNet, revealed that MCNet outperformed the others. This highlights the advantage of additional information provided by directional lighting to enhance automated concrete crack inspection in Civil Engineering. The performance of traditional model and the FusedNN model was compared under increased exposure values to evaluate the advantages offered by the FusedNN model over traditional model.

Overall, this research offers a comprehensive approach to automated concrete in-

spection, leveraging, for the first time in Civil Engineering, advanced illumination technologies used successfully in other scientific fields. The results in this thesis show that multi-directional lighting, combined with specifically adapted neural networks can successfully identify cracks of widths 0.1mm.

To my husband for his everlasting love and support.

Acknowledgements

Embarking on a PhD journey is an adventure that comes with its fair share of challenges and unexpected twists. The path is often steep and rocky, leaving you feeling lost most of the time, only to reach the end when you least expect it. Like all great journeys, this thesis could not have been completed without the support of many individuals who helped navigate the technical and emotional terrain. I am deeply grateful to those who have been very supportive in this academic experience.

First and foremost, I want to express my heartfelt gratitude to my former primary supervisor, Dr.Marcus Perry, for providing me with this invaluable opportunity and offering unwavering support and guidance throughout my Ph.D. His mentorship was beyond measure, and I cannot adequately express my appreciation for his solutions to research and personal challenges, for pulling me out of moments of doubt, and for believing in me even when I did not.

I express my heartfelt appreciation to my current primary supervisor, Dr. Stella Pytharouli, for her priceless mentorship, illuminating the road ahead, and being a dependable source of support whenever I had any uncertainties. Thank you for being my mentor and helping me bring out my best during the thesis revision process. Thank you for patiently reviewing every sentence of my thesis and helping me shape it into the best version possible. I couldn't have done it without you.

I would like to express my sincere appreciation to Dr.Gordon Dobie, my secondary supervisor, for his consistent support and guidance during my research journey. His insights and advice have been invaluable, and I am grateful for his contributions to the completion of my Ph.D.

Contents

Contents

Contents

Contents

Contents

List of Figures

List of Figures

List of Figures

List of Tables

List of Tables

Glossary of Abbreviations

3D	Three-Dimensional
AI	Artificial Intelligence
ALICS	Adaptive Lighting for the Inspection of Concrete Structures
ANN	Artificial Neural Network
ANOVA	Analysis of Variance
AVI	Automated Visual Inspection
BRISQUE	Blind/Referenceless Image Spatial Quality Evaluator
CCD	Charge Coupled Devices
CFD	Computational Fluid Dynamics
CL	Climbing
CMOS	Complementary Metal-Oxide-Semiconductor
CNN	Convolutional Neural Networks
CPU	Central Processing Unit
CR	Crawling
CV	Computer Vision
DCNN	Deep Convolutional Neural Network
DL	Deep Learning
DNN	Deep Neural Network
DoF	degree-of-freedom
DoF	Depth of Field
DTM	Distance Transform Method
ECG	Electrocardiogram

List of Tables

EEPROM	Electrically Erasable Programmable Read-only Memory
FCN	Fully Convolutional Network
FFT	Fast Fourier Transform
FLIR	Forward Looking Infrared Technology
FN	False Negative
FoV	Field of View
FP	False Positive
FPR	False Positive Rate
FPS	Frames Per Second
FR-IQA	Full-Reference Image Quality Assessment
FusedNN	Fused Neural Network
GB	Ground-Based
GPIO	General-Purpose Input/Output
GPR	Ground penetrating radar
GPS	Global Positioning System
GPU	Graphics Processing Unit
I/O	Input or Output
IoU	Intersection over Union
IPTs	Image Processing Techniques
IQA	Image Quality Assessment
IR	Infrared
ITS	Intelligent Transportation Systems
JPEG/JPG	Joint Photographic Experts Group
LED	Light Emitting Diode
MCC	Matthew's Correlation Coefficient
MCNet	Multi-channel Neural Network
ML	Machine Learning
MSCN	Mean Subtracted Contrast Normalisation
N/A	Not Applicable
NDE	Non-Destructive Evaluation

List of Tables

NDI	Non-Destructive Inspection
NDT	Non-Destructive Testing
NIQE	Naturalness Image Quality Evaluator
NN	Neural Network
NR-IQA	No-Reference Image Quality Assessment
OR	Object Recognition
PPV	Positive Predictive Value
PS	Photometric Stereo
R-CNN	Region-based Convolutional Neural Network
ReLU	Rectified Linear Unit
ResNet	Residual Network
RGB	Red Green Blue
ROI	Region of Interest
RR-IQA	Reduced-Reference Image Quality Assessment
RT	Real-Time
SHM	Structural Health Monitoring
SKCV	Stratified k-fold Cross-Validation
SOTA	State-of-the-art
SRAM	Static Random Access Memory
SSD	Single Shot Detector
sUAS	small Unmanned Aerial Systems
SVM	Support Vector Machine
SVR	Support Vector Regression Model
TIFF	Tag Image File Format
TN	True Negative
TP	True Positive
TPR	True Positive Rate
TuFF	Tubularity Flow Field
UAS	Unmanned Aerial System
UAV	Unmanned Aerial Vehicle

List of Tables

U-BIROS	Ubiquitous Bridge Inspection Robot System
UGV	Unmanned Ground Vehicle
UR	Universal Robot
US	Ultrasound
USA	United States of America
USD	United States Dollar
USB	Universal Serial Bus
UT	Ultrasonic Testing
VGG	Visual Geometry Group
YOLO	You Only Look Once

Chapter 1

Introduction

1.1 Introduction

Civil structures are susceptible to deterioration due to factors ranging from repetitive traffic loads to natural disasters [1]. This gradual deterioration reduces the safety margin and the lifespan of the structure, which in turn poses significant risks to both human life and property [2]. Therefore, early awareness and identification of the defects is crucial to extend lifespan and safety of reinforced concrete structures [3]. The primary forms of defects include cracking and spalling [4].

Routine inspections in reinforced concrete structures are still predominantly conducted visually. Manual visual inspections benefit from the adaptable decision-making of people, but they present significant financial and time costs, often put inspectors at risk, and suffer from low repeatability [5]. Robotics and deep learning are gradually automating some aspects of inspection and image analysis, providing asset managers with remote, consistent and cost-effective tools to maintain concrete structural health and resilience [6].

Most of the automated inspections utilise cameras for image capture. The quality of the captured images is influenced by factors such as lens quality, working distance, brightness and lighting conditions [7, 8]. As it concerns the latter, many automated systems still lack the adaptability to assess diverse defects under varying, low, and uncontrolled lighting conditions [9], with the quality of illumination significantly affecting

the accuracy of image analysis results [10].

Geometrical illumination is commonly applied in surgery [11] to improve feature contrast but has received limited attention in the context of Civil Engineering. Although human inspectors often instinctively use directional lighting with flash-lights, previous studies on crack detection has not fully utilised the additional data that directional lighting could provide [12, 13].

Early studies explored various approaches, such as ring and angled lighting, to identify textural differences associated with concrete cracking [12]. Another investigation examined the impact of brightness and lighting direction on crack recognition in ambient lighting conditions, where they found that low-intensity, low-angle lighting was effective in detecting smaller crack widths [14]. Meanwhile, a separate study used extreme dark-field imaging to detect air voids in concrete, employing nearly parallel lighting to the concrete surface. However, this approach required significant preparation of the substrate [9].

Previous studies on concrete crack detection have primarily focused on analysing laboratory datasets [15]. This tendency is common in concrete defect detection literature, where hardware and software testing often relies on laboratory samples and as such, do not fully capture the complexities of real-world concrete structures [16]. There are no studies on testing and evaluation of directional lighting concrete inspection technology in real-world applications.

Engineering standards typically state that crack widths of 0.2mm to 0.4mm in concrete structures need immediate attention [17]. Previous studies have detected cracks with widths less than 0.1mm on clear backgrounds but faced challenges in identifying such cracks under low-light or uneven illumination conditions [18].

Regarding the image quality, the conventional subjective evaluation method that relies on human perception to assess, is not only slow and costly but also prone to subjective biases [19, 20]. Consequently, objective Image Quality Assessment (IQA) methodologies have been developed to automate image quality prediction through diverse algorithms [21, 22, 23]. Real-time automatic Image Quality Assessment has received little attention in computer-aided Civil Engineering [24, 25].

Previous studies have prioritised defect detection capabilities while neglecting the significance of automating the evaluation of concrete image quality, which is essential for effective crack detection. To add to this, currently there is a lack of image processing algorithms that are trained on datasets comprised of directionally lit images.

This thesis focuses on the development of an automated inspection methodology that leverages multi-angle and multi-directional lighting in low-light or challenging environmental conditions to detect cracks in concrete with widths less than 0.1mm. This research focuses solely on detection and classification, with crack segmentation being beyond the scope of this thesis.

1.2   Thesis Overview

The outline of the remaining of this thesis is as follows:

Chapter 2 focuses on common defects in concrete structures such as, cracks and spalling, exploring how they form, and why they happen. Traditional visual inspection techniques are reviewed, including their advantages, disadvantages, and influencing factors. The chapter also looks into semi-autonomous inspection methods, discussing their benefits over traditional techniques, as well as their drawbacks. The Chapter concludes with the suggestion that automated inspection methods could be a tool to overcome the limitations of manual inspections.

Chapter 3 focuses on automated visual inspection techniques and deep learning models utilised for Civil Engineering infrastructure today. It highlights the advantages of using automated visual inspection methods and introduces black-box and white-box crack analysis methods, with more emphasis on black-box methods. It presents recent studies on image recognition, image classification and semantic segmentation used for automated defect inspections in concrete structures. Further, it reviews illumination methods employed to detect concrete defects in low-light environments, highlighting the limitations present in existing literature and the research gaps that this thesis aims to address.

Chapter 4 outlines the methodology and experimental setup for the implementation

of hardware (ALICS) and development of software (based on VGG-16 model) that allows for the investigation of the potential of directional lighting and suitable optimal illumination angle to provide enhanced inspections of concrete structures.

Chapter 5 presents the implementation of the Blind/Reference-less Image Spatial Quality Evaluator (BRISQUE), a No-Reference IQA (NR-IQA) algorithm in the VGG-16 algorithm. The BRISQUE score threshold (B_T), which significantly improved the overall data quality for deep learning applications is introduced. An automated data cleaning procedure based on a BRISQUE score threshold has been implemented and integrated to ensure that only high-quality images are utilised for further image processing.

Chapter 6 presents a novel approach to concrete crack detection and classification, employing five-channel (multi-channel) neural network models based on the VGG-n architecture (where n=16 and 19) utilising the multi-directional lighting. The study evaluated and compared the performance of these multi-channel deep learning models with their 3-channel versions commonly used for crack detection on Civil infrastructure, for binary image classification tasks. The results show major improvements in crack detection and classification in low-light conditions.

Chapter 7 focuses on multi-class image classification for crack and spalling detection. It implements two novel Convolutional Neural Network (CNN) models, FusedNN and MCNet, which are developed based on the MobileNetV2 architecture using multi-directional lighting approach. The chapter analyses the impact of exposure on diffused and fused images, highlighting the advantages of utilising directional lighting in fused images. It compares the performance of directional lighting-enhanced neural network models with traditional model and advanced model in the literature for both binary and multi-class image classification tasks. It compares the best performing neural network model from Chapter 6 with the best-performing neural network model in Chapter 7, highlighting the advantages of using directional lighting for classifying both cracks and spalling. FusedNN and MCNet are compared to a SOTA model, Zoubir and are found to exhibit superior performance.

Chapter 8 discusses the findings and outlines the future work for this thesis.

Chapter 9 provides the conclusions of this thesis.

1.3 Contributions to Knowledge

The research contributions of the work carried out in this thesis are summarised as follows:

- First study to assess the potential of directional lighting incident at an angle of 10 to 60 degrees on the surface of real world concrete structures, captured under low-light conditions (Chapter 4). Geometrical illumination, commonly applied in surgery to enhance feature contrast, has received limited attention in Civil Engineering, particularly for defect detection in concrete structures. Previous studies have successfully detected cracks smaller than 0.1mm on clear backgrounds but have faced significant challenges when attempting to identify these cracks under low-light or uneven illumination conditions. Based on the literature review presented in Chapters 2 and 3, no other study was identified that utilised directional lighting for crack identification, including those of widths < 0.1mm, in concrete Civil infrastructure.

- Implementation of a novel end-to-end, automated process for data-cleaning via a threshold based image quality assessment algorithm (Chapter 5). Unlike existing methods that might incorporate image quality assessment as a separate pre-processing step after completion of image acquisition, this study incorporates Threshold-based IQA algorithm during image capture, ensuring that only high-quality images are retained for subsequent image processing, and low-quality images are discarded. This enhances the performance of the crack detection algorithm.

- Developed a methodology for the implementation of multi-channel deep learning models for detecting defects like cracks and spalling in concrete Civil Engineering infrastructures (Chapter 6 and Chapter 7). Directional lighting was implemented in this work through sets of 5 images, i.e. 5-channels instead of the commonly

used 3-channels in deep learning algorithms. To achieve this, multi-channel deep learning models were developed. No multi-channel deep learning models used for concrete defect detection exist in the reviewed literature. The proposed methodology for multi-channel deep learning models can be applied to any model.

- Developed the Fused Neural Network (FusedNN), a novel CNN model for detecting cracks in concrete structures by using maximum intensity pixel-level image fusion. This is a technique widely used in medical imaging but not in other fields, including Civil Engineering (Chapter 7). Unlike traditional methods that rely on single-image inputs under challenging lighting conditions, the FusedNN merges multiple images captured from different lighting directions, enhancing contrast and defect visibility in concrete structures. This is the first application of image fusion for concrete defect detection in low-light and real-world conditions where conventional approaches struggle.

- First study to demonstrate the effect of exposure on diffused and fused images, particularly in the context of crack detection (Chapter 7).

In addition to the above contributions, the author of this thesis also contributed to the following:

- The thesis is presenting the design of a lighting apparatus (ALICS rig) that employs illumination techniques such as geometrical illumination (direction and angle) in hardware to aid in automated defect detection/analysis. This is an essential aspect for addressing the current lack of illumination methods in automated inspections. This apparatus is a first in hardware that utilises multi-directional and multi-angle lighting approach to accurately detect concrete defects in low-light environments.

- The author contributed to this hardware by developing and implementing the initial design of the lighting rig along with its hardware components. More specifically, a machine-vision camera and manually operated multi-direction and

multi-angle lighting rig were integrated onto a six-axis robotic arm. This integration was intended to capture thorough datasets of concrete specimen images in controlled laboratory settings.

- Later, in collaboration with two other researchers, the lighting rig was modified to automate multi-angle and multi-directional lighting. Following this, the author took active part in the implementation of the first field deployable ALICS rig.

1.4 Research outputs

The research outputs derived from this thesis and related research work by the author of this thesis are presented below.

1.4.1 Journal Articles

1. **S. Pennada**, M. Perry, J. McAlorum, H. Dow, and G. Dobie, "Threshold-based BRISQUE-assisted deep learning for enhancing crack detection in concrete structures," in J. Imaging 2023, 9, 218. The author carried out the conceptualisation, data pre-processing, methodology, visualisation, validation, software implementation, and original writing. Along with McAlorum and Dow, the author captured the images. Perry and Dobie were involved in the supervision of this work. **https://doi.org/10.3390/jimaging9100218**.

2. J. McAlorum, H. Dow, **S. Pennada**, M. Perry, and G. Dobie "Automated Concrete Crack Inspection with Directional Lighting Platform," in IEEE Sensors Letters. The author contributed to the hardware development (as described in section 1.3), the author captured part of the images in the lab, and implemented the VGG binary classifier outlined in this paper. **https://10.1109/LSENS.2023.3327611**.

3. H. Dow, M. Perry, J. McAlorum, **S. Pennada**, and G. Dobie, "Skeleton-based noise removal algorithm for binary concrete crack image segmentation," in Automation in Construction, vol. 151, p. 104867, 2023. The author contributed to

the development of the hardware (as described in section 1.3) which was utilised to capture the images utilised in this paper, and captured part of the images used in this paper.

https://doi.org/10.1016/j.autcon.2023.104867.

4. H. Dow, M. Perry, **S. Pennada**, R. Lunn, and S. Pytharouli, "3D reconstruction and measurement of concrete spalling using near-field Photometric stereo and YOLOv8," in Automation in Construction, vol. 166, p. 105633, 2024. The author contributed to the hardware (as described in section 1.3) which was utilised to capture the images utilised in this paper.

https://doi.org/10.1016/j.autcon.2024.105633.

1.4.2   Conference Papers

1. **S. Pennada**, M. Perry, J. McAlorum, H. Dow, and G. Dobie, "Performance evaluation of an improved deep CNN-based concrete crack detection algorithm," in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol. 12486, p. 1248615, 2023. The author carried out the conceptualisation, data pre-processing, methodology, visualisation, validation, software implementation, and original writing. McAlorum and Dow helped with capturing the images used in this study. Perry and Dobie were involved in the supervision of this research. https://doi.org/10.1117/12.2657723.

2. J. McAlorum, M. Perry, H. Dow, and **S. Pennada**, "Robotic concrete inspection with illumination-enhancement," in Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2023, vol. 12486,pp. 125–131, SPIE, 2023. The author contributed to the hardware development (as described in section 1.3), the author captured part of the images in the lab, and was the reason for the utilisation of BRISQUE in this study. The author reviewed the final draft of this paper.

https://doi.org/10.1117/12.2655938.

3. H. Dow, M. Perry, J. McAlorum, **S. Pennada**, and G. Dobie, "A novel direc-

tional lighting algorithm for concrete crack pixel-level segmentation," in Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2023, vol. 12486, pp. 344–350, SPIE, 2023. The author contributed to the hardware development (as described in section 1.3), and captured part of the images. **https://doi.org/10.1117/12.2657235**.

4. H. Dow, M. Perry, J. McAlorum, and **S. Pennada**, "Concrete crack pixel-level segmentation: a comparison of scene illumination angle of incidence," in Proceedings of the 11th European Workshop on Structural Health Monitoring, EWSHM 2024, Vol. EWSHM 2024. The author contributed to the hardware development (as described in section 1.3).
   https://doi.org/10.58286/29683.

1.5   Journal Articles (Under Peer Review)

1. **S. Pennada**, M. Perry, J. McAlorum, H. Dow, and G. Dobie, "Concrete Crack and Spalling Detection: A Comparative Study of Conventional and Directional-Lighting-Enhanced Deep Learning Models," ***Submitted to Remote Sensing, MDPI***. The author carried out the conceptualisation, data pre-processing, methodology, visualisation, validation, software implementation, and original writing. McAlorum and Dow helped with capturing the images used in this study. Perry and Dobie were involved in the supervision of this research.

2. H. Dow, M. Perry, J. McAlorum, **S. Pennada**, and G. Dobie, "Comparison of directional and diffused lighting for pixel-level segmentation of concrete cracks," - ***Submitted to Infrastructures***. I contributed to the hardware (as described in section 1.3) which was utilised to capture the images utilised in this paper, and also captured part of the images used in this paper.

1.6 Funding

Chapter 2

Concrete inspection techniques and structural health monitoring

2.1 Introduction

Defects in reinforced concrete structures can lead to significant damage, risking public safety and causing financial cost. Deteriorating infrastructure requires regular inspections and timely maintenance. Structural Health Monitoring (SHM) includes various inspection techniques to assess the condition of these structures over time. This chapter focuses on traditional inspection techniques before conducting a literature review on the state-of-the-art (SOTA) inspection procedures, highlighting their advantages and disadvantages. Following this, a research gap is identified regarding the use of these inspection procedures for detecting cracks and spalling in reinforced concrete structures, particularly in low-light environments.

Concrete is the second most widely consumed material globally after water and the most consumed man-made material in the world, with a consumption rate of 14 billion m^3 in 2020 [26, 27, 28]. The consumption of certain building materials, such as concrete, is unlikely to decrease in the coming years; in fact, with population growth, demand is expected to rise [29].

Design codes are standardised guidelines that outline the rules and specifications

for the design, construction, and maintenance of structures. Various organisations such as the American Concrete Institute (U.S.A), American Society of Civil Engineers (U.S.A), British Standards Institution (U.K), and International Union of Laboratories and Experts in Construction and Structures (France) have established guidance and standards for evaluating the condition and performance of concrete structures [30, 31].

In the UK, structural designs typically follow the Eurocode 2 (EN 1992), a series of European standards adopted by all countries in the European Union, for the design of concrete structures in Civil Engineering. It details material properties, load calculations, structural analysis, and design requirements, ensuring concrete structures are safe, efficient, and durable. Eurocode 2 specifies that maximum allowable crack widths for regular reinforced and prestressed concrete structures range between 0.2mm and 0.4mm. Cracks wider than these limits could compromise the safety and durability of concrete structures [32].

2.2 Concrete Defects

Concrete structures can exhibit various types and levels of structural defects, including minor surface cracks and major issues such as spalling, delamination, and reinforcement bar debonding [33]. Several factors contribute to such defects. For instance, cracks in concrete structures immediately after construction are often due to the surface evaporating water at a faster rate than internal moisture, creating a moisture differential and resulting in tensile strain. After approximately 24 hours, the tensile capacity of concrete prevents this. Thermal strain can also cause early age cracking [34]. Additionally, lack of construction knowledge among contractors, poor cooperation between contractors and designers (such as lack of communication or inconsistencies on project specifications), and contractor negligence can further worsen these issues [35].

Cracks and spalling pose a significant risk to the structural integrity of concrete elements and can result in rework costs, ranging from 6% to 12% of the total construction expenses [36]. Therefore, it is crucial to identify and address such defects early to prevent further deterioration and additional costs. The review of literature will explore

their definition, causes, and underlying mechanisms.

2.2.1 Cracking in Concrete: Definition, Categorisation, and Causes

A crack is a complete or incomplete separation that appears on the surface, dividing the concrete into separate parts. Cracks can occur due to various factors such as elastic deformation, thermal movement, shrinkage, and overloading.

Elastic deformation occurs in concrete when subjected to stress, especially when combined with materials that have different elastic properties, for example concrete and steel. The varying shear stresses where the materials meet can lead to cracks, as the materials respond differently to the applied load due to their varying characteristics.

Fluctuation in temperature can cause concrete to expand and contract, resulting in cracks. For instance, if a concrete slab heats up during the day and cools down at night, it may crack if there is not enough space for it to expand and contract.

Shrinkage occurs as concrete dries and hardens, causing it to lose moisture and contract (shrink). If this shrinkage is restricted (like if it is tightly packed against another structure), it can lead to cracks. This is common in new concrete that has not had enough time to cure properly.

Overloading occurs when a building is subjected to loads exceeding its designed load-bearing capacity. Excessive weight can cause cracks, as the concrete and supporting structures experience too much stress, leading to failure in those areas [37].
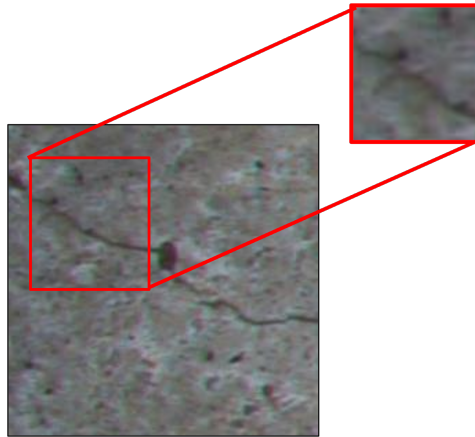
Figure 2.1: Thin concrete crack.
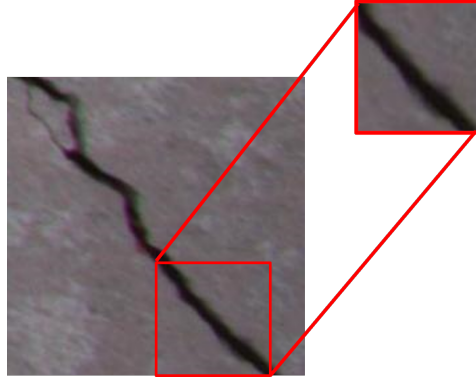
Figure 2.2: Thick concrete crack.

Cracks can vary in shape, size, width, length, and depth. They may also differ in severity ranging from hairline cracks, thin cracks (shown in Figure 2.1) to large, thick and noticeable cracks (shown in Figure 2.2) that affect the structural integrity. These are the sub-images obtained from real-world images captured using ALICS. The severity of a crack is evaluated based on its width [17].

Table 2.1 shows the classification of cracks based on their widths. Thin cracks are typically considered minor and do not significantly affect the structural integrity. Medium cracks indicate moderate distress and needs monitoring or minor repairs, while wide cracks are serious and require immediate attention and possible remedial measures [38].

Table 2.1: Classification of Cracks based on width adapted from [38].

Classification	Crack width
Thin	<1mm
Medium	1mm to 2mm
Wide	>2mm

2.2.2 Spalling: Definition, Categorisation, and Causes

Concrete spalling is a phenomenon where the surface layer of concrete breaks apart and delaminates. An example of concrete spalling is shown in Figure 2.3. Table 2.2 summarises the key causes of concrete spalling.

There are three main types of spalling which occur in concrete [42]:

Figure 2.3: Concrete spalling adapted from [39]. The image dimensions are 780× 440 pixels, and the scale is set at 1cm : 20 pixels.

- Aggregate spalling: This occurs when the connection between the aggregate and the concrete matrix is weakened, often due to freezing and thawing cycles, corrosion, or chemical exposure. As a result, the coarse aggregate in the concrete surface may be lost.

- Corner spalling: This occurs at corners and edges of concrete surfaces, where the concrete is more susceptible to stress and damage. Inadequate reinforcement or a lack of proper waterproofing measures may result in water penetrating the concrete and freezing, causing expansion and damage to the surface.

- Explosive spalling: This kind of spalling arises as a result of high-stress conditions, for example sudden impact, fire, or thermal shock. If the internal pressure in the concrete exceeds its strength, pieces of the concrete break and chip off from the surface.

If left untreated, these types of spalling weaken the integrity of the structure by reducing the load-bearing capacity of key elements such as beams and columns, posing a significant threat to its safety. To effectively identify cracking and spalling structural health monitoring and various inspection techniques are used.

Table 2.2: Causes of Concrete Spalling [40, 41].

Cause of Concrete Spalling	Description
Corrosion of Reinforcement Bars (Rebars)	Moisture and oxygen penetrate the concrete, corroding the steel bars, which expand and push the concrete apart.
Sulfate Attack	Sulfate ions from soil or water react with hydrated compounds in concrete, forming expansive products that create internal stress and lead to surface damage.
Alkali-Silica Reaction (ASR)	Alkali hydroxides in the concrete react with reactive silica in certain aggregates, producing a gel that swells and cracks the concrete.
Freeze-Thaw Cycles	Water trapped in the concrete expands when frozen, creating internal pressure that can crack and degrade the surface.
Thermal Stress	Rapid heating rates and the formation of water vapour and gas pressure within the concrete can cause spalling.
Inadequate Depth of Reinforcement Cover	Insufficient concrete cover over reinforcement can lead to exposure and spalling.
Low-Quality Concrete Cover	Poor quality concrete cover fails to protect the embedded reinforcement, leading to surface damage and spalling.
Improperly Constructed Joints	Faulty joints can allow water ingress and differential movement, causing cracking and spalling.
Improper Water Content	Incorrect water content can lead to weak concrete, making it prone to spalling.
Poor Compaction of Concrete	Insufficient compaction leaves voids and weaknesses in the concrete, making it susceptible to spalling.
Improper Installation Practices	Lack of supervision, failure to vibrate the concrete, or carelessness during installation can lead to spalling.
On-Site Actions by Personnel	Dislodging rebars or carelessly pouring concrete can disrupt the integrity of the structure, leading to spalling.
Structural Parameters	Factors such as hindered thermal expansion, sharp corners, and inadequate reinforcement cover can increase the likelihood of spalling.
High Applied Loads	High loads can create stress and strain in the concrete, making it more susceptible to spalling.

2.3 Structural Health Monitoring and Crack detection in Concrete

SHM involves continuously monitoring the physical and functional condition of structure over time to detect deterioration since the last inspection. With advancements in sensing technologies and data processing techniques, SHM is capable of collecting data on issues like concrete deterioration, steel rebar corrosion, concrete cover delamination, spalling, deflection or settlement, cracks, and changes in geometry. SHM plays a critical role in extending the lifespan of structures and reducing the likelihood of catastrophic failures by providing early warning signs of potential structural failures [43, 44, 45, 46]. Table 2.3 outlines several key SHM techniques to detect cracks. However, implementing SHM poses several challenges [47, 48] as shown in Table 2.4.

Table 2.3: Studies on Structural Health Monitoring Techniques and Crack detection.

Method	Description	References
Signal-Based Methods	Analyses signals from sensors that measure vibrations, accelerations, or strains using signal processing techniques to detect defects, including crack identification in civil structures.	Bakhshi and Tehran [49]
Curvature Mode Shape-Based Techniques	Analyses changes in the curvature of mode shapes to identify damages, particularly cracks by observing deviations from expected patterns.	Wahab and Roeck [50]

Damping-Based Methods	Assesses changes in the damping characteristics to evaluate the performance and detect cracks in civil structures.	Şafak [51]
Vision-Based Monitoring System	A cost-effective system for capturing dynamic responses of civil structures in real-time, utilising consumer-grade cameras to support health monitoring.	Fukuda [52]
Microwave Interferometer for Deflection Measurement	Uses a non-contact microwave interferometer to measure deflections in vibrating stay cables, aiding in the health monitoring of cable-stayed bridges.	Gentile [53]

Table 2.4: Challenges faced by SHM.

Challenge	Description
Environmental Factors	Variations in environmental conditions can alter the dynamic behaviour of structures, making it challenging to accurately detect damage.
Data Inaccuracies	Many damage detection techniques can exhibit inaccuracies, especially in large-scale structures, due to high noise levels and missing measurements.
Complexity of Large Structures	Large structures often have complex behaviours and diverse material properties, complicating the analysis and interpretation of SHM data.

Instrumentation and Technology Limitations	Existing SHM systems face challenges in instrumentation, communication, signal processing, and data storage, which can affect their reliability and efficiency.
Anomalies in Data	Abnormalities in data, such as missing observations and outliers, can hide the true condition of the structure, making damage assessments more difficult.

Figure 2.4 compares two sets of operations: one with SHM deployment and one without. The quality of the structure is determined by its functional value and maintenance cost. When a system does not have SHM, its functional value is initially high but decreases as it ages, which also results in increasing maintenance costs. In contrast, when SHM is deployed, the functional value of the system keeps increasing as it ages with a constant maintenance cost. This means that SHM deployment can enhance the quality and service life of the structural system with standard or lower maintenance costs. Without SHM, a decrease in functional value and an increase in maintenance costs occurs as the structure ages [54].
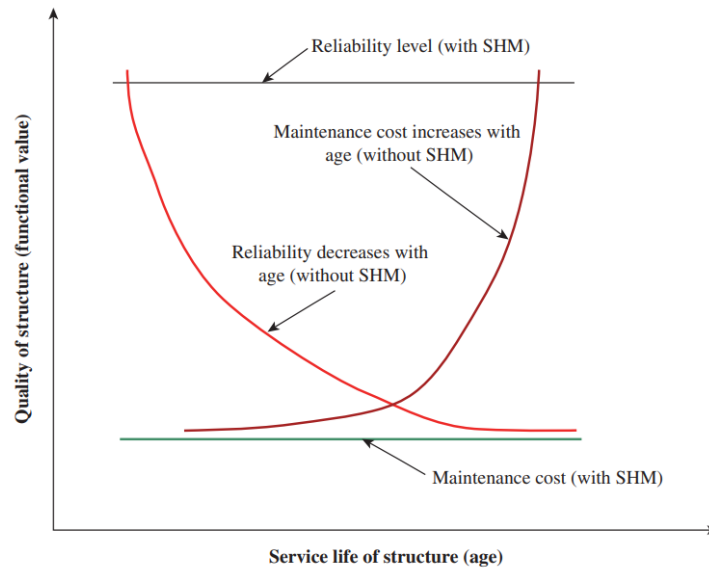
Figure 2.4: Comparison of structures with and without SHM taken from [54].

Chapter 2. Concrete inspection techniques and structural health monitoring

Regular inspections of concrete structures are essential for identifying damage, facilitating repairs, and ensuring safety while maintaining structural integrity. General inspections are carried out every 1 to 2 years, depending on the age of the structure, condition, and use. For example, the bridges in UK are inspected every 1 to 3 years to ensure safety regarding cracks and structural integrity [55]. Principal inspections, which are more detailed, occur approximately every 6 years and involve a thorough examination of the structural elements [56]. There are various levels of inspection ranging from regular inspection techniques to specialised inspections using equipment such as robots, crawlers, drones, etc., which are further described in detail in the following subsections. Inspections are categorised as destructive, semi-destructive, and non-destructive inspections.

Destructive inspections involve testing methods that result in damage to the structure or sample being evaluated. This approach is used to obtain accurate data on the strength of the concrete and its failure characteristics. Common techniques include:

- Load Testing: Applying a controlled load to a structure until failure to assess its load-bearing capacity [57].

- Core Sampling: Removing a cylindrical section of concrete for laboratory analysis, which cannot be reused [57].

Semi-destructive inspection methods cause minor, localised damage to the structure while still allowing for some usability of the tested area. These methods provide valuable information about the properties of the material without significantly compromising the overall integrity.

- Pull-Off Tests: Measuring the bond strength of surface treatments by pulling a small area of the material [58].

- localised Core Sampling: Extracting small samples from the structure that minimise impact on the surrounding concrete [59].

Non-destructive inspections (NDI) involve inspecting a structure to detect defects without causing any damage. There are two key methods within NDI:

- Non-Destructive Testing (NDT), which identifies defects [60], and

- Non-Destructive Evaluation (NDE), which assesses the nature and significance of those defects [61].

Other methods include ultrasonic inspection, magnetic particle inspection, dye penetrant inspection, eddy current inspection, acoustic emission monitoring, and ground penetrating radar (GPR) [62].

One common technique is manual visual inspection, where trained inspectors examine surfaces for signs of wear, corrosion, or other visible defects. This involves physical entry of the person into the field that needs to be inspected. The Bridge Inspection Certification Scheme establishes a regulated framework with structured training for inspectors to possess all the necessary skills and knowledge to conduct thorough inspections, promoting safety and reliability in bridge maintenance [63].

The inspector conducts a visual assessment of the entire structure to identify defects using tools such as rulers, microscopes, magnifying glasses, and cameras, without causing any damage to the structure. The output of a manual visual inspection can be presented in various formats (e.g. documented reports, drawings, or images). For example, based on the severity, a scoring system is applied ranging from 1 (indicating the best condition) to 5 (indicating the worst condition), reflecting the extent of damage in structures including bridges [56]. Table 2.5 below provides a detailed overview of these scores.

Table 2.5: Severity Scores in Bridges.

Score	Description	Condition	Impact
1-2	Minor defects	Structure in good condition	Minimal impact
3	Moderate defects	Structure in fair condition	Potential impact
4	Severe defects	Structure in poor condition	Significant impact
5	Critical defects	Structure in very poor condition	Unsafe

Virtually all structures are manually inspected and this is ultimate method of as-

sessment. The problem is that it is expensive and difficult with large infrastructures. Hence the 'shortcut' methods of SHM etc. Also, these manual inspections often suffer from high error rates, particularly in large-scale infrastructure inspections. These errors occur due to several reasons, including task, environmental, individual, organizational, and social factors, as shown in the Table 2.6 and adapted from [64].

Task refers to the characteristics of the inspection process itself, which affect how easily defects can be identified. Environmental factors involve external conditions like lighting, temperature, and noise, that can either enhance or slow down the ability of inspector to identify defects accurately. Individual factors relate to personal attributes of the inspector, such as experience, intelligence, and fatigue, that affect the performance of the inspector. Organizational factors includes the support and resources provided by the organization, such as management, training, and feedback, which are essential for effective inspections. Lastly, social factors involve interactions like communication, consultation, and social pressure, which can impact inspection outcomes either positively or negatively.

The factors have been reviewed based on practical understanding and experience of the author of this thesis and are categorised into two groups: acceptable (those that strongly influence inspection performance) and non-acceptable (those with minimal impact or considered less relevant).

Table 2.6: Evaluation of factors impacting the performance of inspection: Acceptable and non-acceptable factors.

Factor	Category	Acceptable?	Explanation
Number of Defects	Task	Yes	As number of defects increases, the difficulty of inspection and the likelihood of missing defects increases.

Factor	Category	Acceptable?	Explanation
Defect Type	Task	Yes	Different types of defects vary in difficulty; wider cracks are relatively easy to identify, while finer defects, such as hairline cracks or thin cracks near to edges, require closer and more detailed inspection.
Size of the defect	Task	Yes	Clearly visible defects are easier to detect, but subtle ones such as hairline cracks require more focused inspection.
Defect Location	Task	Yes	It is difficult to identify defects in hard-to-reach or hidden areas during manual inspections.
Pacing of inspections	Task	Yes	The speed of inspections affects the performance; rushing can lead to misclassifications, while going too slowly reduces efficiency.
Multiple Inspections	Task	Yes	Conducting multiple inspections can catch more defects, but too many inspections can lead to fatigue.
Overhead Data	Task	Yes	Additional data can aid inspections, but excessive information may be distracting.
Sex	Individual	No	Sex may influence physical abilities or communication styles, but it is not a major factor that impacts inspection performance.

Factor	Category	Acceptable?	Explanation
Age	Individual	Yes	Age affects visual acuity and stamina, which further impacts inspection performance.
Visual Acuity	Individual	Yes	Good eyesight is crucial for detecting small defects.
Intelligence	Individual	No	A trained inspector is more effective in detecting defects, independent of their intelligence level.
Aptitude	Individual	Yes	Natural ability to notice details and perform technical tasks enhances inspection performance.
Duration of Work/Day	Individual	Yes	More time in the role increases familiarity with common issues, improving inspection skills.
Experience	Individual	Yes	Experienced inspectors are better at identifying defects and managing complex situations based on past experiences.
Visual Lobe	Individual	Yes	A larger field of vision helps in seeing more at once, improving inspection efficiency.
Scanning Strategy	Individual	Yes	A systematic scanning approach helps ensure that defects are not missed.

Factor	Category	Acceptable?	Explanation
Biases	Individual	Yes	Cognitive biases can affect judgement and lead to missed defects. For example, if an inspector has recently encountered a particular type of defect, they may expect to find it again, potentially diverting their attention from recognising other types of defects.
Lighting	Environmental	Yes	Proper lighting is essential for detecting defects clearly; poor lighting can misclassify or miss defects in concrete structures.
Noise	Environmental	Yes	Excessive noise can be distracting and reduce focus during inspections.
Shift Duration	Environmental	Yes	Long shifts can cause fatigue, affecting attention and increasing the risk of missing defects.
Temperature	Environmental	Yes	Extreme temperatures can impact comfort and focus during inspections.

Factor	Category	Acceptable?	Explanation
Time of Day	Environmental	Yes	Performance can vary with the time of day due to natural alertness levels. Its very difficult to carry out inspections in dark or low-light environment. Infrastructures like tunnels must be closed at night for inspections. During these inspections, inspectors use head torches to identify cracks in the harsh, low-light conditions. This can reduce visibility and lead to fatigue, increasing the risk of missing defects [65].
Management Support	Organizational	Yes	Support from management provides necessary resources and motivation, leading to better performance.
Training	Organizational	Yes	Proper training helps inspectors understand procedures and standards, leading to more accurate detection.
Retraining	Organizational	Yes	Ongoing retraining helps inspectors stay updated with current techniques and standards.
Instructions	Organizational	Yes	Clear instructions help inspectors know what to look for, reducing confusion and errors.
Feed-forward Information	Organizational	Yes	Providing information about potential issues before inspections helps inspectors concentrate on critical areas.

Factor	Category	Acceptable?	Explanation
Feedback	Organizational	Yes	Regular feedback helps inspectors learn from their mistakes and improve performance.
Incentives	Organizational	Yes	Incentives can motivate better performance.
Job Rotation	Organizational	Yes	Rotating jobs prevents boredom and fatigue, keeping inspectors more engaged and alert.
Isolation	Social	Partially Accepted	Isolation can improve focus, but too much alone time can reduce motivation and collaboration with others.
Collaboration	Social	Yes	Collaborating with colleagues or experts improves decision-making and accuracy in inspections.
Communications	Social	Yes	Effective communication ensures inspectors have all the necessary information and understand their tasks, reducing mistakes.

Inspection techniques are quite expensive, which can limit their accessibility and usability. For example, the visual inspections of the Brooklyn Bridge in New York cost about 1 million USD. Studies have shown that roughly 56% of the condition ratings of concrete bridges are inaccurately assessed, and 95% of the data is based on visual inspection. This is because inspecting all the bridge components, especially the underside, over a river or stream is challenging [66]. In [67], the consistency of 21 inspection teams was compared across four bridges. The study highlighted significant variations in how inspectors detected concrete cracks and revealed that some experienced inspectors missed cracks due to varying weather and traffic conditions. Therefore, it emphasises the need to improve the consistency in reporting and detection of issues

like cracking. Furthermore, manual inspection techniques tend to be time-consuming and labour-intensive, which can lead to delays in identifying and addressing structural issues promptly [68, 69]. Therefore, there is need to explore alternative approaches that can overcome these limitations and deliver more effective and efficient inspection results [70].

Some inspection tasks are hard for humans to do because they are in places that are difficult to reach, like tight spaces inside air-conditioning ducts, water-filled pipelines, offshore structures, or small gaps in walls. Some tasks are also dangerous for people, such as inspecting tall structures, areas hit by natural disasters, or bridges. This highlights the need for semi-autonomous inspection techniques that involve autonomous vehicle control with human review of data. In these inspection techniques, vehicles such as robots, drones, ground vehicles, or climbers, equipped with cameras or any image capturing device, operate independently while inspectors analyse the collected data to detect defects. Different types of robots that move in various ways have been used to inspect buildings. Examples include unmanned aerial vehicles (UAVs), unmanned ground vehicles (UGVs), marine robots, wall-climbing robots, and cable-crawling robots.

Semi-autonomous inspection is a hybrid between manual and automated inspection techniques. It utilises robotic or automated systems to collect data, while still involving humans for data analysis. They are capable of performing inspections more quicker than manual inspection techniques and also capable of accessing hard-to-reach places more easily. While these inspection techniques often struggle to carry out inspections in low-light conditions, leading to noisy data and increased risk of false positives or missed defects [71]. They have difficulty in detecting sub-surface defects, such as hairline cracks [72]. Also, the success of inspections often relies on the experience and judgement of the operator, which can lead to inconsistencies in identifying defects [73]. Given these limitations, there is an increasing need for a fully automated inspection processes that leverage advanced algorithms for accurate defect classification, particularly in low-light or dark environments, reducing dependence on human decision-making.

2.4   Summary

Concrete, a widely used material in many industries, presents challenges such as cracking, and spalling that affect structural integrity of concrete structures. To ensure the health and safety of concrete structures, effective monitoring and inspection are essential. Traditional manual visual inspections, are often expensive, time-consuming, and inconsistent. Semi-autonomous inspection techniques, which still rely partially on human decision, also face limitations, particularly in varying lighting conditions, low light conditions, and rely on the experience and efficiency of the inspector inspecting the field. Therefore, there is an increasing need for automated inspection processes that leverage advanced algorithms for accurate defect classification, especially in challenging lighting conditions, while minimising reliance on inspector decision. Automated visual inspection is explored in detail in the next chapter.

Chapter 3

Automated Visual Inspection Techniques and Deep learning

This chapter explores the state-of-the-art in automated visual inspection (AVI) techniques with a focus on crack and spalling detection in low-light environments for Civil Engineering infrastructure. It includes the details of AVI techniques, as well as machine learning and deep learning models commonly utilised in the Civil Engineering sector.

3.1 Introduction

Automated visual inspections (AVI) have become increasingly popular amongst Civil Engineering practitioners [74] due to the limitations in traditional inspection methods (discussed in Chapter 2). AVI offers several benefits such as reduced costs, minimal disruptions, and lower risk, as highlighted in [75].

Considerable progress has been made in the last two decades, establishing frameworks for inspecting defects in various types of Civil infrastructure, such as bridges, roads, tunnels, and general facilities using various robots. The use of robots for inspection provides a safer alternative to manual inspection. Automated robotic inspection improves the frequency of inspections and reduces bias in error detection. Robots include UAVs, UGVs, marine vehicles, microbots, wall-climbing robots, cable-suspended robots, cable-crawling robots and legged robots. Most common type of inspection ro-

bots are UAVs and UGVs [76]. Robotic inspection technologies combined with deep learning have been in use for inspecting the built environment for at least the last decade [77].

Computer vision is an interdisciplinary area that enables machines to analyse digital images and videos, thereby improving their capacity to recognise features and patterns using visual information [18]. Deep learning models, particularly those leveraging CNNs, have transformed the field of computer vision [78, 79]. While image processing enables tasks such as image classification and object detection, deep learning further enhances these capabilities, making a significant impact on industrial production [80]. The effectiveness of deep learning depends on the availability of large amounts of data, powerful computing resources, specialised training methods, and advanced networks that enable it to perform tasks intelligently and efficiently [18]. An important area of applications and further development within image classification and segmentation is the detection of structural defects on Civil infrastructure [18].
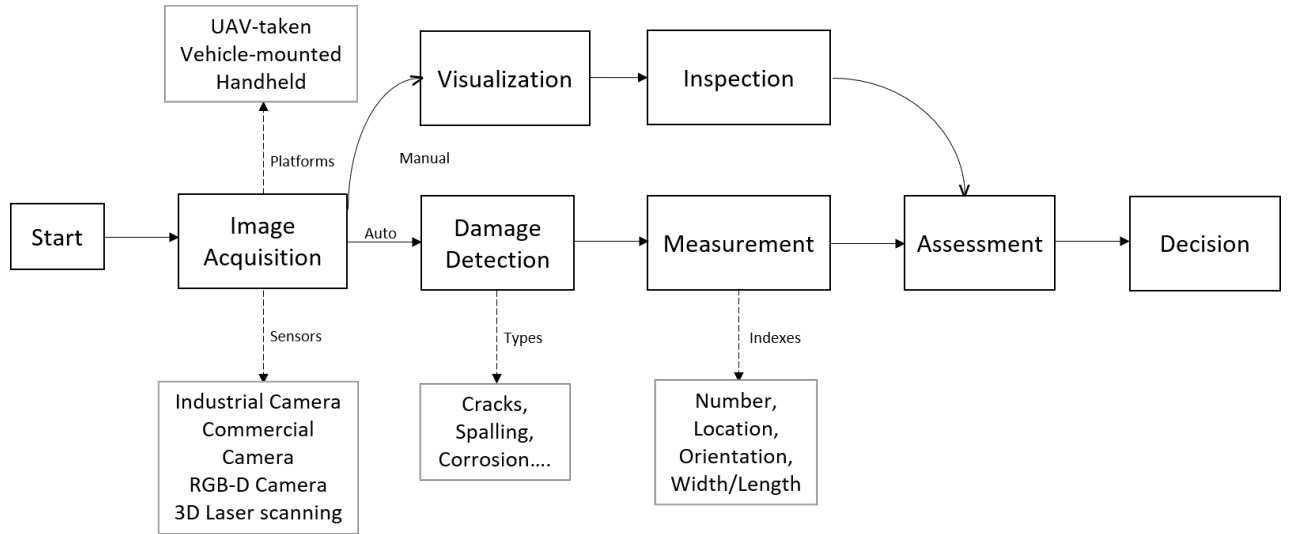
Figure 3.1: Computer vision pipeline designed for the manual or automated inspection of Civil infrastructure modified after [81].

A typical computer-vision-based inspection pipeline for manual or automated inspection of civil infrastructure is shown in Figure 3.1. It comprises of five processes: Acquisition, Detection, Measurement, Assessment, and Decision [81]. The process be-

gins with the acquisition phase, where images or videos of Civil infrastructure are captured using various devices, such as cameras mounted on drones, vehicles, or handheld devices. Once the images are obtained, the detection phase uses algorithms to identify potential cracks or defects, employing either traditional inspection or advanced machine learning methods. After detecting a defect, the measurement phase calculates important details such as the location, width, and depth of the crack. This is followed by the assessment phase, where the severity of the detected cracks is evaluated to determine their potential impact on the infrastructure. Based on this evaluation, decisions are made regarding the necessary maintenance or repairs to ensure the safety of the infrastructure.

3.2 Automated Crack Detection

Automated crack analysis can be categorised into three main types: detection, classification, and segmentation. Detection identifies the specific locations where cracks are present within the image [82]. Classification involves determining whether or not a crack exists in the image [83]. Segmentation provides a more detailed analysis by highlighting the exact crack pixels in that image [84]. Figure 3.2 provides an illustration of this.

Automatic crack detection in concrete structures can be separated into two main approaches: white-box and black-box techniques [75]. The former uses algorithms such as edge detectors and thresholding [85], whereas the latter employs machine learning and artificial neural networks [86]. Both methods have unique benefits that depend on the application. Black-box methods tend to be more effective for detecting cracks initially, while white-box methods are better for detailed, pixel-level segmentation [87].

Black-box methods that employ neural network-based defect detection algorithms typically consist of two main stages: feature extraction and classification. Image processing techniques are first used to extract relevant features from crack images, which are then analysed by classifiers. However, if the extracted features do not accurately represent the cracks, the performance of the classifier can be compromised.
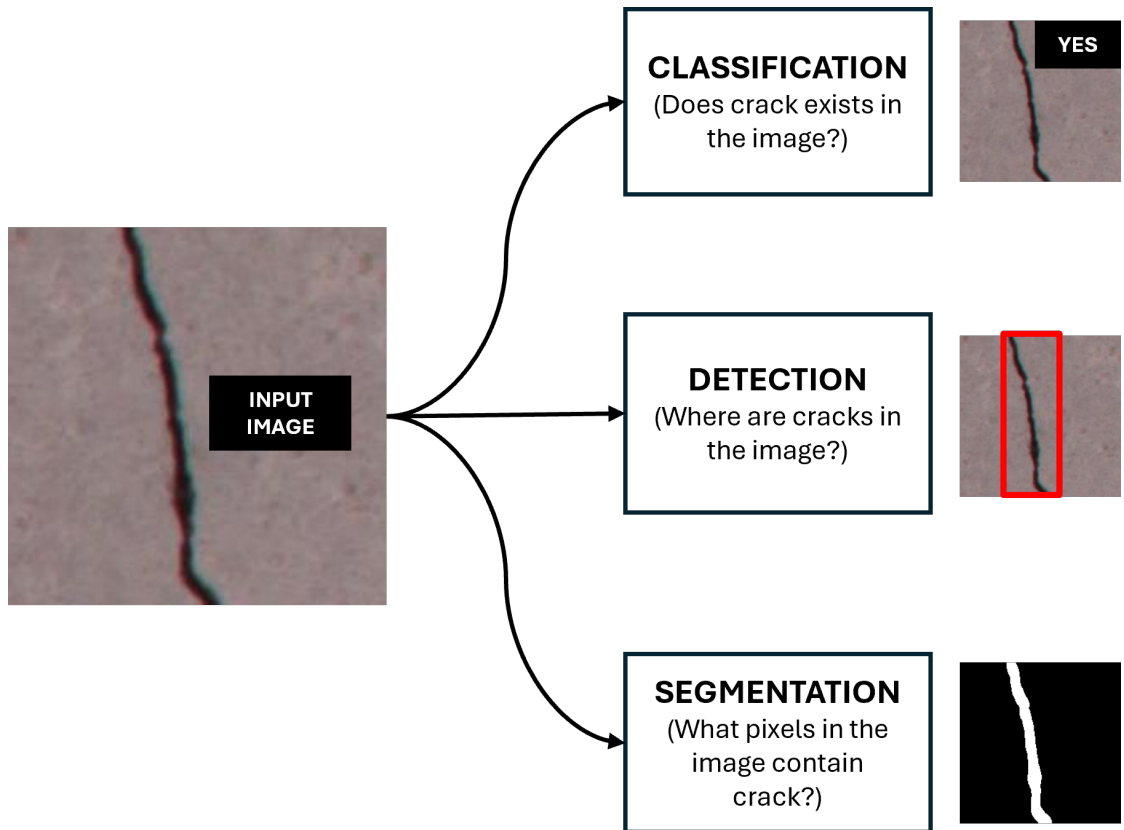
Figure 3.2: Image classification, object detection, and segmentation examples of concrete crack detection.

Deep learning techniques, particularly convolutional neural networks (CNNs), have improved traditional manual inspections by automatically extracting relevant features and showing better performance in crack detection [88]. For example, cracks on pavement surfaces [89], cracks in bridges [90], cracks in pavement structures [91], and pixel-level cracks in concrete structures [92] can be automatically detected. Black-box models are particularly effective at detecting complex patterns in images, enabling tasks such as classification and detection. They learn from vast amounts of training data and can then apply this knowledge to new, unseen data [93].

Machine learning algorithms consist of various methods that enable models to learn from data and make predictions, and they can be classified into four main types: Supervised Learning, Semi-Supervised Learning, Unsupervised Learning, and Reinforcement Learning. Supervised learning involves training a model on labelled data, where the input-output pairs are known, allowing the model to learn. Semi-supervised learning combines both labelled and unlabelled data, taking advantage of the large amounts of unlabelled data to improve accuracy of prediction beyond what can be achieved with labelled data alone. Unsupervised learning focuses on analysing unlabelled data to find patterns, structures, or groupings without human intervention, making it useful for tasks like clustering, dimensionality reduction, and anomaly detection. Lastly, reinforcement learning makes decisions based on rewards or penalties, which helps to optimise its behaviour over time [94].

Black-box algorithms offer several key advantages. They are highly scalable, capable of handling large and dynamic datasets, and perform well on new and unseen data. Their performance improves over time as more data become available provided that the model is periodically retrained or updated with new data. These models can handle binary tasks, such as binary image classification (classifies an image into one of two classes), and more complex tasks, like multi-class image classification (classifies an image into one of three or more classes). Once trained, they require minimum human involvement in decision-making to complete tasks, making them efficient and autonomous [93]. They are particularly effective in specific tasks such as detection, classification, and segmentation, which are explained in detail below.

3.2.1 Object Recognition (Crack Detection)

Crack detection is aimed at identifying cracked areas and highlight these using bounding boxes as shown in Figure 3.2. In the field of computer vision, several architectural families are commonly used for object recognition (OR), including region-based convolutional neural networks (R-CNN) [95], You Only Look Once (YOLO) [96], and Single Shot Detector (SSD) [97]. Among these, R-CNN architectures have been the most widely utilised for crack detection. R-CNN, Fast R-CNN, Faster R-CNN are members of this family, with Faster R-CNN [96] being the most popular choice in the field of crack detection.

In [95], the authors presented the first automated crack detection method using UAVs, deep learning, and high-resolution imaging. They utilised selective search with a pre-trained CNN from the ImageNet dataset. Through transfer learning, they fine-tuned a region-based CNN i.e, R-CNN using 384 manually labelled crack images (3-channel i.e., RGB), allowing the system to identify and locate cracks on concrete surfaces. But the approach has faced several limitations, especially when it was tested to detect hairline cracks under challenging environmental conditions such as shadows, uneven lighting on dark background or dirty concrete surfaces. These issues reduced the visibility of cracks increasing false negatives. Therefore, due to lighting conditions there was a negative effect on the quality of images and in the detection of cracks in concrete structures.

[96] focused on object detection in poor lighting conditions using YOLO and Faster RCNN deep learning models. The images were categorised based on lighting types such as Low, Ambient, Weak, Object, and Single to analyse their impact on object detection. However, low-light settings led to loss of key features, and consequently decreased detection accuracy and an increase in false positives and false negatives.

[97] introduced a large-scale road damage dataset of 9053 annotated images with 15435 instances of eight damage types (based on type of cracks such as linear cracks and alligator cracks, corruptions such as potholes and rutting, along with road damages such as blurring of white lines) captured using a smartphone mounted on a passenger car, and

applied the Single Shot MultiBox Detector with InceptionV2 and MobileNet backbones. The cost-effective image capturing process relied on semi-automated system, with a human driver operating the vehicle. The low-light conditions remained a challenge, particularly in distinguishing cracks from shadows, and reflections. As a result, if an image was affected by shadows, reflections, or blur, it could not be recaptured, since the image acquisition phase is separate from the image processing stage.

[98] developed a modified faster region-based convolution neural network to detect concrete cracks and handwriting scripts in real-world bridge inspection images. By labelling handwriting as a separate object class, the study improved crack detection accuracy in complex backgrounds. As with the previous cases presented above, this network also struggled to classify cracks under poor lighting conditions.

[99] explored low-light object detection by comparing various image enhancement algorithms such as, ResNet-based CycleGAN and U-Net based CycleGAN models. These adaptive image enhancement models also faced limitations in accurately detecting objects under low-light conditions, leading to lower detection performance.

Also, [100] proposed an automated damage detection algorithm for detecting defects such as, cracks, spalling, rebar exposure, spot in concrete structures using EfficientNet and MobileNetV3 architectures. While the improved model demonstrated strong performance under uniform lighting conditions, its accuracy declined significantly when evaluated under varying lighting conditions. Additionally, the model was not robust enough in detecting defects when tested against complex backgrounds.

Even the most recent deep learning algorithms used in crack detection, e.g. MobileNetV3, exhibit reduced performance when the lighting conditions are less than optimal. This does not come as a surprise, as poor illumination introduces noise, lowers contrast, and can obstruct critical features such as cracks, spots, or spalling. As a result, feature extraction becomes less reliable, reducing the accuracy of defect detection. If deep learning algorithms are to become the main tool for crack detection and replace manual inspections, they need to be able to perform even under unfavourable light conditions, and/or complex backgrounds.

3.2.2 Image Classification (Crack Classification)

Image classification makes decisions at the image or image patch level, where the trained architecture determines whether the new input contains a crack or not, as shown in Figure 3.2. The overall architecture consists of two parts: (1) extraction of features from raw RGB images layer by layer, utilising successive convolutional and max-pooling layers, where convolutional layers help in identifying local patterns such as edges and textures and max-pooling layers downsamples the spatial dimensions of the feature maps, (2) fully connected layers are used to classify the features into specific categories (cracked or not cracked) [101].

An early CNN-based crack detection study [83], demonstrated the potential of deep learning for concrete inspections for binary crack classification tasks. However, their approach involved manually capturing images without automated platforms, such as robots or drones, and required labour-intensive manual cropping of training patches. This resulted in limited scalability for large-scale structural inspections. Further, the detector exhibited poor performance on concrete images with stains, where it frequently misclassified stains as cracks or failed to detect cracks near chalk letters.

Light conditions remains a factor that affects performance not only for crack detection but also for crack classification. [91] investigated the effect of network depth on crack classification, concluding that deeper networks are capable of learning more information i.e., the deeper the networks are, the more it learns about detecting cracks. The study also found that networks trained on images from a specific location do not perform well when tested on images from a different location. Images were captured either manually or semi-automatically using smartphones, under ambient lighting conditions, leading to performance degradation (85.6% precision) due to variations in lighting across different locations.

[102] developed a novel method for automatic pavement crack detection and classification, utilising a CNN to detect cracks and principal component analysis to classify them into longitudinal, transverse, and alligator cracks. The study found that pavement images obtained outdoors have large illumination variations that can introduce

challenges in classifying cracks from the background and no specific techniques were proposed to address this drawback.

[87] compared a AlexNet deep convolutional neural network model with edge detectors for image-based crack classification in concrete. The results were obtained using high-quality images taken under good lighting conditions, free from other environmental factors. The extension of these findings to real-world scenarios, where lighting conditions may be poor or uneven, is limited.

[103] conducted a comparative study between fully trained and transfer learning modes using an AlexNet-based deep learning convolutional neural network for crack classification tasks in small Unmanned Aerial Systems (sUAS)-assisted structural inspections of concrete bridge decks and buildings. The study highlights the need for improved handling of poor lighting conditions, as along with image blurriness from sUAS vibrations and low-resolution cameras, as they can negatively impact crack classification accuracy, potentially misleading both human inspectors and conventional image-processing methods.

[104] proposed a CNN architecture for patch-based crack detection in black box images, classifying road surface elements into three categories: crack, road marking, and intact regions. However, the study highlights that image-processing techniques for crack detection, including the proposed CNN, may not perform well under low lighting conditions, which can lead to inaccuracies and misclassifications due to reduced contrast and visibility of cracks.

All the studies mentioned above refer to crack detection and classification on concrete surfaces. There are very few studies for detecting and classifying spalling in concrete. [105] has developed three Mask R-CNNs to detect cracks and spalling in concrete structures. A more recent study by [106] proposed an encoder-decoder-based deep architecture for detecting and classifying the severity level of spalling. [107] combined advanced image processing techniques, including entropy-based segmentation and noise filtering, with machine learning models to accurately classify the spalling damage into severity levels based on area and depth. [108] proposed an automated method for detecting concrete spalling using a piecewise linear stochastic gradient descent logistic

regression model combined with image texture analysis. This method extracts texture features from concrete surface images to classify the image as image samples as spall or no spall under well-lit conditions. To date, no studies have addressed the identification of spalling under low-light conditions.

3.2.3 Semantic Segmentation (Crack Segmentation)

If the classification is performed at the pixel-level i.e., every pixel in an image is labelled with a class then it is known as Semantic Segmentation [109]. The output of any semantic segmentation framework is the input image where crack pixels are highlighted in a different colour, usually white, making them distinguishable from the background pixels, as shown in Figure 3.2. The image processing techniques (IPTs) are applied independently or in combination with object recognition and image classification.

[84] proposed an effective framework (GoogLeNet, ResNet + IPTs) for crack detection and width measurement to automatically extract and localise cracks from concrete images. Various image processing techniques, such as Otsu's thresholding, median filtering, and the Hessian matrix, were then applied to eliminate illumination effects and produce a segmented crack image. The study highlights that illumination plays a crucial role in crack detection accuracy, as variations in lighting can significantly impact crack visibility, feature extraction, and segmentation quality.

In [110], transfer learning-based deep CNN was utilised to identify crack patches. Fast block-wise segmentation was employed to create crack masks and improve crack localisation. However, the study highlights that non-uniform lighting can pose challenges in crack segmentation, leading to misclassification of cracks and sealed cracks, particularly when their intensities and widths are similar.

[111] presents a robust, fully automated hybrid method for concrete crack detection, segmentation, and quantification under complex backgrounds and varying lighting conditions leveraging Faster R-CNN, modified Tubularity Flow Field (TuFF) with CLAHE, and modified DTM. The study used Faster R-CNN algorithm to detect crack regions, which are then localised using bounding boxes. A modified TuFF algorithm is applied to segment the crack pixels within the detected regions. To measure crack thickness and

length, a modified DTM is used, providing precise pixel-level measurements. The study also highlights that variations in illumination can impact accuracy of segmentation.

In [112], AlexNet was developed to classify image patches into crack, sealed crack, and background classes. A Fully connected network was then applied for pixel-level crack segmentation. In [113], a Sobel-edge adaptive sliding window was proposed for efficient crack patch extraction. Non-maximum suppression was applied to retain important patches. The study highlights that these algorithms did not perform well under different lighting conditions.

Existing literature highlights that most studies face significant challenges in detecting defects under low-light conditions. Although numerous image enhancement algorithms have been developed, they frequently fail to perform reliably in such environments due to the inherently poor quality of the captured images [114]. Moreover, these approaches typically train neural network models using standard three-channel RGB images [97, 115] or single-channel gray scale images only [102]. As a result, the potential benefits of leveraging advanced lighting techniques or utilisation of additional information that these lighting techniques can provide still remains unexplored in current defect detection frameworks.

Furthermore, as shown in [97], when images captured in the field are later found to suffer from issues such as reflections, blur, or noise, they cannot be corrected during post-processing stage, and remained as drawback. This is primarily because the image acquisition and image processing phases are carried out separately. In scenarios where inspections take place in hazardous or hard-to-access environments such as in nuclear sectors, returning to the field to recapture images is not feasible. Therefore, there is a critical need to develop an on-board, image-quality-based algorithm capable of detecting and recapturing low-quality, blurred, noisy, or degraded images in real time during the image acquisition stage. This underscores the need for advanced preprocessing techniques, advanced lighting techniques to enhance detection performance under challenging lighting conditions.

3.3 Model Architecture and Optimization Techniques

In this section, the two most commonly identified model architectures in crack detection are presented in more detail. These are the VGG16 and the MobileNetV2 neural network models. Although these models were developed in 2015 [116] and 2018 [117], respectively, VGG16 is still one of the most used network backbones for object classification and detection models [118] and MobileNetV2 is the light weight neural network model suitable for real-world crack detection and classification applications because of its computational efficiency [119, 120]. VGG-16 and its similar, VGG-19 are still commonly used as the benchmark for comparative analysis amongst other, newer deep image models [121]; and have been found to outperform other models used in crack detection such as ResNet-50, and Inception V3 [18].

3.3.1 VGG16 neural network model

The VGG16 deep convolutional neural network, developed by the Visual Geometry Group, is widely used for crack classification and it was shown to achieve high accuracy and efficiency when using pretrained weights from ImageNet [121, 122]. It leverages transfer learning to perform robustly on small datasets [121], making it suitable for the image classification task such as those presented later in this thesis. Choosing a neural network that is widely understood and extensively employed in the Civil Engineering industry for crack detection was preferred over experimenting with newer or state-of-the-art models.

The VGG-16 architecture consists of 16 layers, including 13 convolutional layers and 3 fully connected layers. It accepts an input image of size 224×224×3 (height, width, channels). The convolutional layers use small 3×3 filters with a stride of 1 and same padding to preserve spatial resolution, while max-pooling layers follow each convolutional block with 2×2 filters and a stride of 2 to reduce dimensionality.

The network consists five sequential blocks. The first two blocks each contain two convolutional layers with 64 and 128 filters, respectively, followed by a max-pooling layer. The third block has three convolutional layers with 256 filters, followed by

pooling. The fourth and fifth blocks each contain three convolutional layers with 512 filters and are also followed by pooling layers. After these feature extraction layers, the architecture includes three fully connected layers, the first two have 4096 units with ReLU activation, and the final layer has 1000 units with a sigmoid or softmax activation for classification.

Utilising a widely adopted model that has proven to be successful in crack detection and classification tasks, ensures compatibility with existing practices in the field [122, 123, 124]. The focus was to improve crack inspections in concrete structures and further existing technologies in Civil Engineering using directional lighting. Segmentation-focused models like UNet, which are used for pixel-level crack segmentation tasks [125] were not utilised at this stage of work as the focus is on image-level crack classification (crack and no-crack) tasks. For these reasons, no attempt was made to utilise a different CNN.

Zoubir et al., developed an efficient and automated method for detecting defects in concrete bridges such as cracks, efflorescence, and spalling using VGG-16 and transfer learning [126]. To overcome the challenge of limited training data, the study employed transfer learning by fine-tuning different layers of the pre-trained VGG16 model. Three training configurations were compared: (a) retraining only the classification layers, (b) retraining the classification layers and the last convolutional layer, and (c) retraining the classification layers along with the last two convolutional layers. Among these, configuration (c) achieved the best results with high accuracy and F1 scores across all defect types. Therefore, the model developed in this study sets a strong benchmark for concrete defect classification tasks.

3.3.2 MobileNetV2 neural network model

MobileNetV2, a lightweight convolutional neural network that has been proven effective for crack detection in Civil infrastructure, offering significant advantages for resource-constrained environments [117]. Its architecture, utilising inverted residual blocks, depthwise separable convolutions, and linear bottlenecks, reduces computational complexity by approximately 5.2 times compared to MobileNetV1, enabling deployment

on mobile and embedded devices for real-time inspection of concrete structures, pavements, and bridges.

MobileNetV2-based CNNs have a smaller size and better computational efficiency without sacrificing performance [127]. It has a widespread use in civil engineering applications with high success rate in detecting and classifying cracks in concrete structures[128]. The light weight of the MobileNetV2 architecture combined with its low computational cost and fast processing speed, make it suitable for real-time applications where images are captured using handheld devices, as described in Chapter 4 [119].

MobileNetV1 was developed based on the VGG architecture, and incorporated convolutional layers to enhance accuracy. However, adding more layers led to gradient vanishing issues, where the network struggles to learn effectively as explained in [129]. To address this, ResNet (Residual Network) introduced the residual block, which enhances information flow by incorporating skip connections. These connections allow the input from one layer to bypass one or more intermediate layers and be added directly to the output of a later layer. This design allows easier gradient flow during backpropagation, enabling the effective training of deep neural networks [130].

MobileNetV2 improved upon MobileNetV1 by incorporating residual structure of ResNet and adding a linear bottleneck implementation. The linear bottleneck reduces computational complexity by simplifying convolution calculations, improving efficiency. Additionally, MobileNetV2 uses depth-wise separable convolutions, which split the convolution into depth-wise and point-wise operations, reducing parameters and computational cost. The inverted residual block in MobileNetV2 allows for better information flow and gradient propagation, improving effectiveness of training. These help in improving the performance of MobileNetV2, exceeding the performance of MobileNetV1 in both accuracy and efficiency (Figure 3.3 and Table 3.1 [130]).

One way to transform features from N channels to M channels is by using a block with a specified stride (s) and expansion factor (t). Stride (s) refers to the number of pixels by which the filter moves across the input image during convolution. Expansion factor (t) indicates the ratio of the number of input channels to the number of

Table 3.1: Bottleneck of MobileNetV2 [130].

Input	Operator	Output
HxWxN	1x1 conv2d, ReLU6	HxWxtN
HxWxtN	3x3 dwise s=s, ReLU6	H/s x W/s x tN
H/s x W/s x tN	linear 1x1 conv2d	H/s x W/s x M

output channels in a layer, thereby controlling the width of the network. This block with specified s and t includes a 1×1 convolutional layer followed by the depth-wise convolutional layer, with linear activation instead of non-linear activation, after the point-wise convolutional layer. Down-sampling can be achieved by adjusting the parameter 's' in the depth-wise convolutional layer [130]. MobileNetV2 utilises convolutions and average pooling for processing the input data, with specific configurations denoted by parameters such as c (number of output channels) and n (number of repetitions). The overall network structure of MobileNetV2 is shown in Table 3.2 [130].
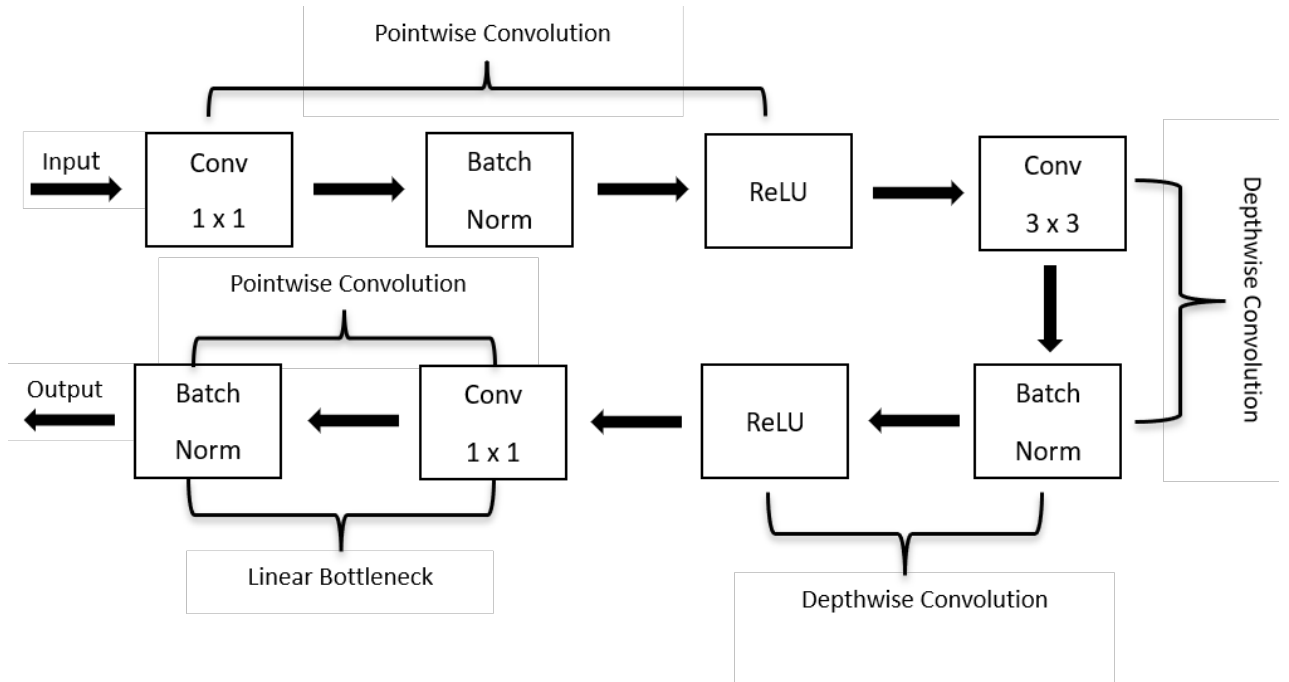
Figure 3.3: Inverted Residual Linear Bottleneck [130].

MobileNetV2 is a network architecture with 19 layers, designed for feature extraction and classification. The conventional MobileNetV2 model is shown in Figure 3.4.
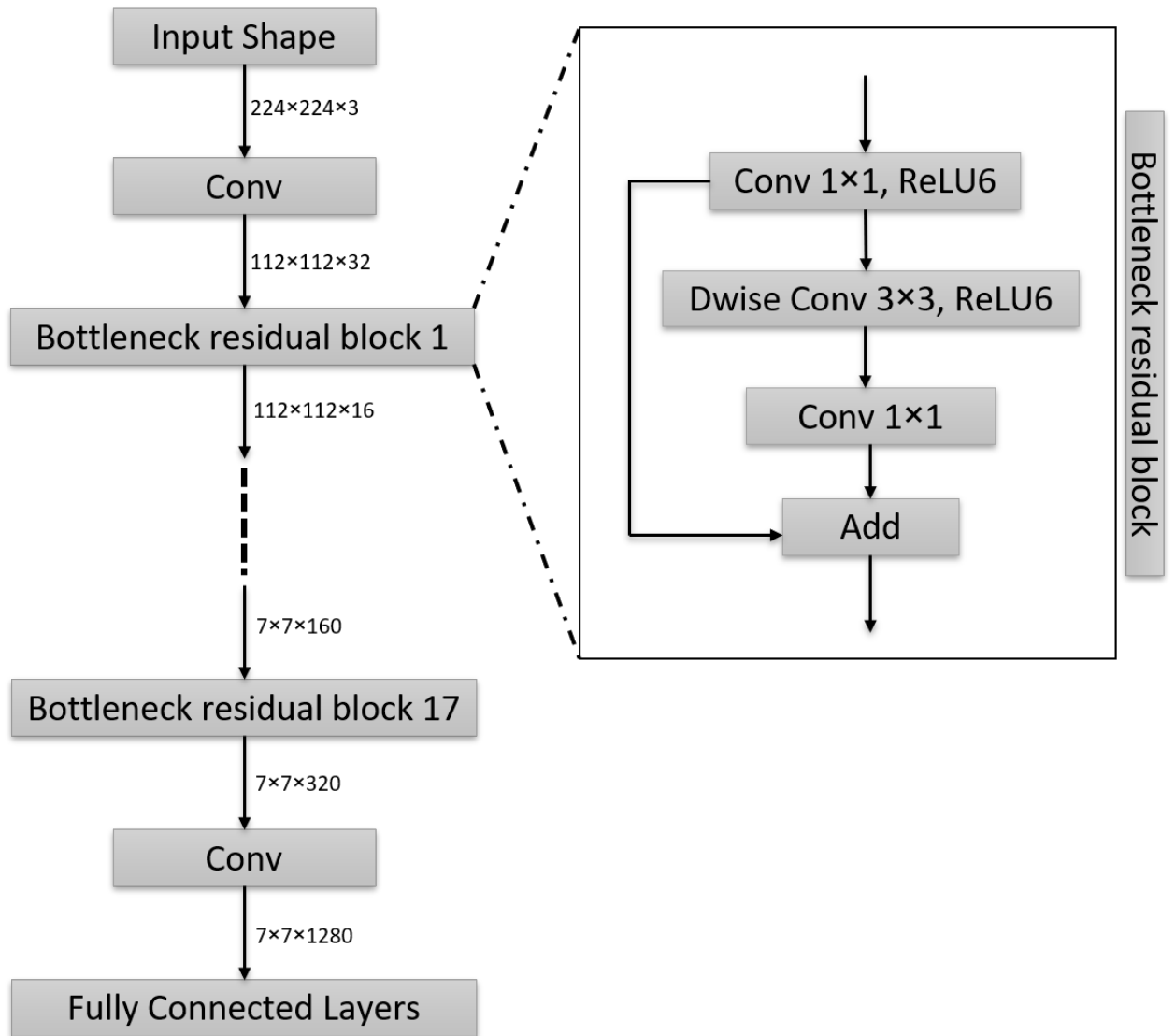
Figure 3.4: Architecture of the traditional three-channel MobileNetV2 model.

Table 3.2: The Overall Network Structure of MobileNetV2 taken from [130].

Input Shape	Operator	t	c	n	s
224 x 224 x 3	conv2d	-	32	1	2
112 x 112 x 32	bottleneck	1	16	1	1
112 x 112 x 16	bottleneck	6	24	2	2
56 x 56 x 24	bottleneck	6	32	3	2
28 x 28 x 32	bottleneck	6	64	4	2
14 x 14 x 64	bottleneck	6	96	3	1
14 x 14 x 96	bottleneck	6	160	3	2
7 x 7 x 160	bottleneck	6	320	1	1
7 x 7 x 320	conv2d 1 x 1	-	1280	1	1
7 x 7 x 1280	avgpool 7 x 7	-	-	1	-
1 x 1 x 1280	conv2d 1 x 1	-	k	-	

3.3.3 Stratified k-fold cross-validation

A stratified k-fold cross-validation (SKCV) shown in Figure 3.5 is utilised in this research. The k value is typically chosen as five or ten for balancing variance and bias [131, 132]. SKCV divides the entire dataset into 'k' equal-sized subsets, or "folds," ensuring that each fold contains nearly same percentage of samples from both majority and minority classes, unlike k-fold cross validation. The process consists of 'k' iterations, each using 'k-1' folds for training the model and one fold for testing the performance of the model [132, 133, 134].

3.3.4 Hyperparameter tuning and Regularization

Hyperparameter tuning refers to the process of optimising the hyperparameters of a machine learning model to improve its performance. They are settings that are external to the model and cannot be learned from the data, for example batch size, and learning rate. Batch size determines the number of samples processed before updating the model, and learning rate controls the step size during optimisation [135]. The goal of hyperparameter tuning is to find the best combination of these parameters that results in the most accurate and efficient model for a given dataset [136].

Overfitting in machine learning refers to a situation where a model fits the training data too closely, resulting in poor generalisation to new, unseen data [137]. Regulariz-
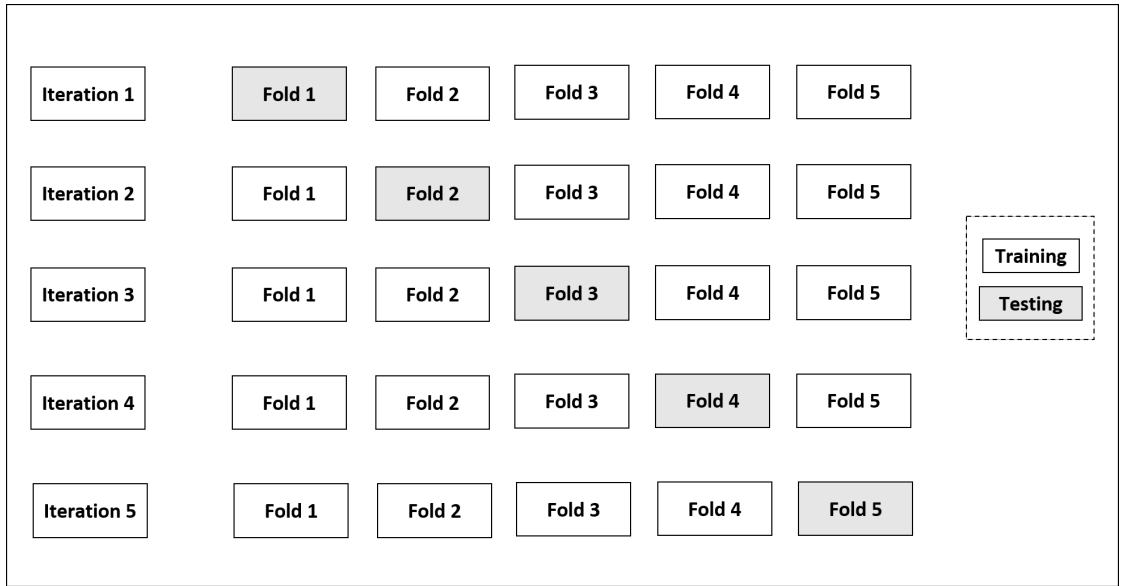
Figure 3.5: Stratified five-fold cross-validation: The gray boxes represent the folds used for testing the model, while the plain boxes represent the folds used for training the model.

ation techniques are used to prevent overfitting by adding a penalty term to the loss function of the model [136]. Early stopping is one such method, as it stops the training process before the model becomes too complex and starts to overfit the training data. By tracking accuracy on the validation set helps in finding the optimal hyperparameter values without overfitting. This strategy ensures a higher level of generalisation and helps in reducing bias while increasing variance, ultimately improving the performance of the model. This approach reduces training time and once training stops, the model with the best validation performance is selected for use [138].

3.4 Image Quality Assessment and Automated Visual Inspections

For major civil and construction assets, semi-autonomous or remote inspections are sometimes carried out with UAVs or wholesale robotic platforms fit with inspection equipment, as reviewed in [139, 140, 141]. In the majority of the reported literature, the

process of capturing images is typically conducted separately from the crack detection phase [142]. In an industrial setting, this would mean that any errors or inaccuracies occurring during image capture, such as improper lighting, could not be corrected or adjusted on-site [7]. An automated image quality assessment method in real time allows for the optional 're-take' of the image and the replacement of the original low quality image with another one of higher quality. This not only reduces the time required for inspection but also enhances the effectiveness of the defect detection algorithms.

Several factors affect the quality of an image when captured using a camera, including lighting conditions, camera settings, motion blur, lens distortion, and noise [7, 8]. Since it is uncertain whether a captured image will be natural, noisy, blurry, or distorted in any other way, assigning a score based on the image quality can help in determining the suitability of the image for training or testing the CNN models.

Image Quality Assessment is the process of evaluating the quality of an image. It is widely used in machine learning, computer vision, neural physiology, image processing and other domains where image quality plays an important role [143]. There are two types of IQA methods.

3.4.1 Subjective Methods

Subjective image quality assessment methods involve gathering individual opinions on the quality of images using a five-level rating scale without participants having full access to the scores or information from others. By averaging the scores provided by the participants, a reference point for image quality is established, and collecting opinions from a larger number of individuals increases the reliability of these results [19].

In practical applications, subjective evaluations are often considered the most accurate for assessing image quality. However, they are expensive, time-consuming, not suitable for real-time applications, and results are influenced by various factors like lighting conditions and brightness [20]. Lighting conditions and brightness play a crucial role in determining image quality and directly impact the performance of deep learning algorithms. Under low-light conditions, images tend to exhibit increased noise,

reduced contrast, and loss of features due to inadequate exposure, which can reduce the performance of neural networks for defect detection tasks. Conversely, excessive brightness may result in overexposed regions where fine important features are lost, increasing the likelihood of false positives. These degradations that occur due to lighting reduce the reliability of automated detection systems and highlight the necessity for models to be robust against illumination.

3.4.2 Objective Methods

The development of accurate mathematical models to evaluate the quality of images similar to human observers is the objective of objective IQA methods. These models aim to replicate image quality predictions similar to humans and provide efficient and cost-effective image quality evaluation [144, 145]. Objective IQA methods extract features from the image and analyse them using a quality score [146]. IQA can be categorised into three frameworks: Full-Reference (FR), Reduced-Reference (RR), and No-Reference (NR) or Blind [147] based on the availability of an undistorted, clear quality image known as the reference image. FR methods require a reference image to evaluate image quality, while RR methods use extracted features from the reference image to evaluate image quality. On the other hand, NR methods can assess image quality without the need for a reference image [148, 149, 150, 151]. Obtaining a reference image in real-world applications is challenging, therefore NR-IQA techniques are more suitable. As the main focus of this thesis is not IQA, no extensive literature review was carried out on existing NR-IQA methods. Instead, the Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) was chosen as the tool to assess the image quality as it predicts the quality of an image without a reference image [152].

3.3.2.1 BRISQUE Method

BRISQUE is effective in evaluating spatial quality making it a valuable tool for assessing image quality in Civil Engineering applications [153]. Additionally, various studies have explored its efficiency in assessing the quality of medical images [144]. Figure 3.6 provides an example of a raw image (natural image) and a degraded image (noisy

image), along with their BRISQUE scores.

BRISQUE employs a support vector regression (SVR) model trained on images with known distortions, such as blur, noise, and artefacts, to compute the image quality score. It compares the distribution of the mean subtracted contrast normalisation (MSCN) of the subject image to the database images. The MSCN of a pixel at location (i, j) is defined by equation (3.1), where I represent the intensity of the pixel, μ is the local mean, and σ is the local variance in intensity of the surrounding pixels.

$$MSCN(i,j) = \frac{I(i,j) - \mu(i,j)}{\sigma(i,j) + 1} \tag{3.1}$$

By analysing the shape and variance of the distribution a histogram of MSCN values is then generated to detect various distortions such as blur, noise, and compression. The model then compares these values to an image database and assigns a score ranging from 1 to 100, with lower scores indicating better image quality. A score close to 0-30 suggests a good(high) image quality, while a score greater than 40 indicates a poor(low) image quality. An example of this is shown in Figure 3.6.
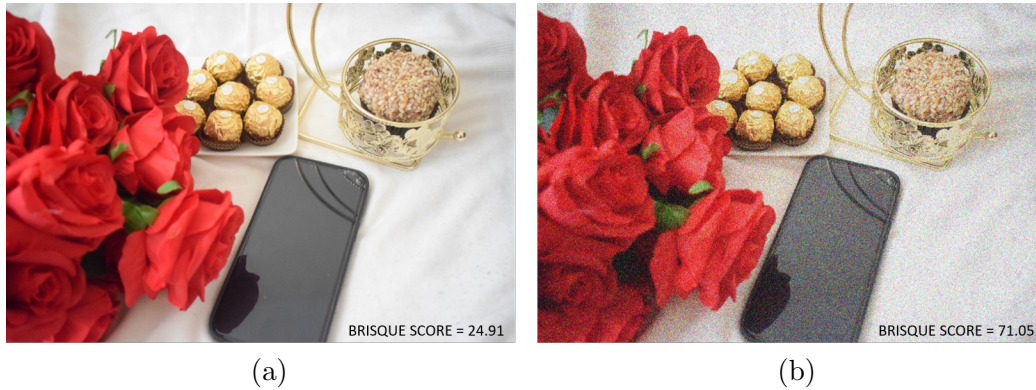
Figure 3.6: Comparison of image quality using BRISQUE scores. a) Natural image with a BRISQUE score of 24.91, indicating better image quality. b) Noisy image with a BRISQUE score of 71.05, indicating poor-image quality. (author's own picture

3.3.2.2 The implementation of BRISQUE in deep learning algorithms

A study carried out in [7] on the evaluation of live road footage for road damage detection using BRISQUE concluded that challenges arose in accurately assessing image

quality due to varying lighting conditions such as harsh sunlight or dark environments. This variability potentially led to inaccuracies in image processing and object detection used for vehicle detection and road crack identification. In low visibility scenarios like poor lighting, or adverse weather, the performance of the system was compromised, affecting its overall accuracy.

[154] utilised BRISQUE to enhance visibility for identifying faulty porcelain insulators in a low-voltage power distribution system. It involves creating a dataset with various backgrounds and light conditions, adding faulty insulators from different angles to evaluate the proposed YOLOV4 model with different object detection models. The study underscores effectiveness of BRISQUE in image quality assessment and also highlights its limitation in accurately evaluating images under low-light conditions due to reduced visibility and clarity. It is essential to address these limitations for real-world applications by developing advanced image quality assessment techniques specifically designed for challenging lighting conditions. This will improve defect detection capability for critical infrastructure maintenance and monitoring.

There are no studies in the reviewed literature that utilise BRISQUE for real-time image quality assessment in low-light conditions, particularly in the context of concrete crack detection or explore the correlation between image quality assessment and concrete crack detection using neural networks. Most existing studies utilise neural network architectures capable of processing noisy, blurred, or otherwise degraded images [10]. However, these image enhancement algorithms still struggle to carry out image recognition tasks under various lighting conditions [114]. Also, all these algorithms are implemented in the post-processing stage. It would be more advantageous to develop image quality-based algorithms integrated into to the image acquisition stage to ensure that only high-quality images are further forwarded for image processing.

3.5 Illumination Techniques and Automated Visual Inspections

Current automated inspection practices lack the adaptability required to assess a diverse number of defects under varying environmental conditions [13]. There may be an overlooked opportunity to utilise advanced illumination techniques to enhance the contrast of defects within the image to improve automatic defect detection methods. Such methods include white-box image processing and black-box neural networks [87], which were discussed earlier. These automatic analysis methods are also replacing human-based analysis due to improved accuracy, consistency and time-efficiency [6].

Commercially available automatic inspection machines are mostly limited to either ground-based (GB) or aerial(drone)-based image capturing platforms [139]. Images are either captured under ambient lighting or with additional lighting during darker environmental conditions. In other research fields, however lighting is utilised as a method of enhancing the contrast of objects: during surgery [155], to improve worker safety [156], and machine vision applications [157, 158]. To the knowledge of the author, there are no examples of adaptive contrast enhancing illumination techniques applied to automated inspections for concrete crack detection.

3.5.1 Conventional Techniques

Illumination is an important variable in both image capture and computer vision tasks. Cameras operate by directing light to a sensor that reconstructs the scene using the Photovoltaic effect. A scene is usually made up of one or more features within some background. The objective of illumination is to create contrast between the feature(s) and the background to make the feature more easily identifiable visually or to a machine. Without adequate illumination, some features in the scene may be lost. Indeed, there are many conventional illumination techniques that are used to create varying contrasts for machine vision [159]. These techniques can be split into 4 categories:

- Geometry
- Spectrum

- Type of light

- Filtering

In geometrical illumination techniques, the physical position, such as the direction or angle of light is varied. This creates contrast by producing directional shadowing of objects in the scene which highlights depth differences and orientation of objects. For example, Figure 3.7 shows images of a pen against a surface captured under varying geometrical illumination conditions. The top row shows lighting at a high angle and bottom row at a low angle, with columns 1 - above, 2 - right and 3 - diffused directions. It is clear that when light is projected perpendicularly to the pen at a lower angle (above, low) there is a longer shadow enhancing the depth difference and orientation of the pen. The insets in column 2 also show that the rough surface is more visible with a lower lighting angle.

Projecting colours onto a scene creates contrast by highlighting objects of the opposite colour on the spectrum. By changing the type of illumination used, such as the light source (LED, fluorescent, halogen, infrared) or the projection method (spot, dome, line), can enhance contrast in some applications. For example, infrared light can be used to reduce reflectivity of materials [160] or for internal defect detection [161]. Filtering, such as polarisation or neutral density, are physical films placed in front of a camera lens in order to modify certain wavelengths of light. These are usually used to provide more flexibility in aperture and exposure settings of the camera when in bright or dark environments or to remove features such as reflections.

3.5.2   Illumination for concrete defect detection

One of the earliest examples of using illumination to improve crack detection was by [12]. In [12], utilised angled lighting to highlight textural differences in concrete, but found that this reduced the accuracy of crack detection neural networks, as the models were trained on images lit using standard diffused lighting. The study acknowledged the challenges posed by ambient lighting, but it lacks a robust solution to avoid its effects, which could compromise detection accuracy in various environmental conditions. Furthermore, the absence of a mobile inspection device limits the practicality of the
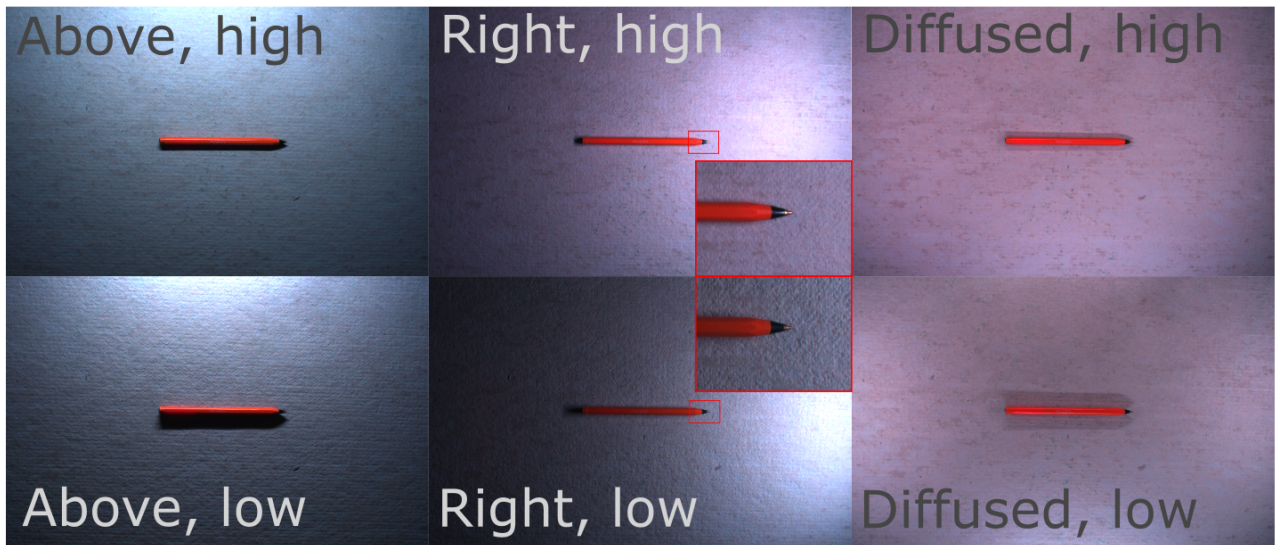
Figure 3.7: Geometrical illumination techniques to enhance contrast of depth and orientation in an image. Inset: zoomed portion showcasing enhanced surface roughness.

method in real-world applications, where adaptability and mobility are essential for effective monitoring of concrete structures.

[9] employed "zero-angle" light to automate the detection of air voids in concrete. By directing light parallel to the air voids (utilising extreme dark-field techniques), the visibility of the perimeter of the voids is significantly improved. This approach enhances the accuracy of void detection and measurement through segmentation. However, it necessitates preparation of the concrete surface by skilled operators before image capture. Additionally, the study faced challenges in accurately segmenting irregularly shaped voids, which limits its effectiveness in low-light environments and practical applications.

The effectiveness of crack detection using the deep learning model is heavily impacted by different lighting conditions as explained in [162]. Photographs were examined under different weather conditions, such as sunny, cloudy, and foggy, as well as at various times of day, including sunset, moonlight, and during specific hourly intervals from 5:00 to 8:00 PM. Detection accuracy declined significantly under low-light conditions, dropping by 50% during sunset and moonlight, about 25% from 6:00 to 7:00 PM, and sharply by 85% from 7:00 to 8:00 PM. However, between 5:00 and 6:00 PM,

no major change was observed, highlighting the importance of adequate illumination for accurate defect detection. These results emphasise the importance of proper illumination for effective defect detection and indicate that improved lighting is required to maintain performance in low-light conditions.

From these examples, it can be inferred that illumination techniques show promise to improve automatic defect detection in concrete. However, there is a lack of research into the different methods of illumination and the existing studies are either too focused on a single method or are application specific.

3.5.3 Lighting in automated visual inspections

Automated visual inspection techniques are effective in capturing detailed information about surfaces, as previously mentioned; however, it is important to note the vital difference between capturing images and detecting cracks. Errors or inaccuracies during image capture cannot be immediately corrected or adjusted, potentially leading to imperfect analysis. Such irregularities could cause significant defects to go unnoticed and require error correction.

Although these platforms can come equipped with lighting, this is to make imaging possible in low-lighting conditions rather than to enhance the contrast with geometrical illumination[163, 164, 165, 166]. To the knowledge of the author, there are no commercially available automated inspection platforms that contain customisable contrast-enhancing illumination methods.

Further, [167] used Light Emitting Diodes (LEDs) in the inspection of pipes in the oil and gas sector. These LEDs are mandatory due to having no ambient light within the pipes, but did not provide other capabilities for enhancing contrast. [92] designed a wall-climbing robot for concrete structure inspection and provided an option to add extra LEDs for compensation when ambient light was insufficient. Similarly, [65, 165, 168] are all examples of autonomous inspection machines in civil sectors that used illumination as a source of light to ensure correctly exposed image capture. Most of these studies focused on capturing images in low-light conditions, yet they have not explored advanced illumination techniques for improving crack detection in such

environments.

3.5.4   Addressing Limitations in Concrete Crack Detection

In the study of concrete crack detection, algorithms are often challenged by uncontrolled or low-light conditions. For example, the Crack Image Analysis System in [77] computes geometric parameters of crack networks under uniform illumination, typically provided by a diffuse and evenly distributed light source. However, this dependence on uniform lighting may limit its accuracy in low-light conditions, reducing the visibility and contrast of crack features. Therefore, adapting the system for low-light conditions could enhance its robustness and reliability.

Another approach employs a dual encoder network structure that combines fast Fourier transform (FFT) with a CNN. The frequency domain branch of the model uses FFT for spatial and frequency domain conversion, reducing the impact of low brightness. The CNN coding branch extracts contextual information, and a feature fusion detection module enhances crack localisation accuracy, making it effective for detection in low-light environments. However, the model may struggle with complex crack patterns (hairline cracks) and capturing target areas in low-light conditions, indicating areas for performance enhancement [169].

Another proposed algorithm improves crack detection in low-light conditions by first removing haze from the image using a contrast stretching method. It then converts the image to gray scale, applies thresholding, and utilises morphological skeletonisation for accurate detection of hairline cracks. This algorithm may struggle to differentiate cracks from background noise, especially in low-light images. Addressing these drawbacks is essential to enhance effectiveness of the algorithm in accurately detecting cracks under challenging lighting conditions [170]. Additional lighting can enhance crack visibility and reduce noise in images, leading to more precise detection. However, even with adequate lighting, image quality remains crucial. Poor image quality, indicated by low resolution and high noise levels, can reduce crack detection accuracy. Therefore, ensuring clear and high-resolution images is essential for optimising detection accuracy.

Advanced image processing methods often assume uniform lighting conditions for

detecting cracks in concrete structures, but they struggled to detect hairline cracks [171]. Therefore, exploring alternative geometrical illumination methods could further enhance the accuracy and efficacy of these algorithms. By addressing the challenges posed by low-light conditions and optimising lighting techniques, concrete crack detection algorithms can be significantly improved, leading to better maintenance and safety of civil structures. This study uses geometrical illumination techniques, leveraging different lighting angles and directions to highlight small differences in texture and surface properties of concrete cracks.

Previous studies mainly focused on laboratory datasets, which do not account for the complexities and variations present in real-world scenarios, such as different illumination conditions and surface irregularities. They did not implement any lighting-based neural network models to enhance defect detection capability [15]. This limitation highlights the need for methods that can effectively handle the challenges posed by real-world crack images.

Subsequent studies aimed to conduct crack detection in real-world settings but encountered challenges with illumination, obtaining clear images, and accurately identifying defects [16]. Additionally, the quality of crack detection is affected by various factors such as camera specifications, surface illumination, and environmental conditions, leading to unclear images and increased false positive rates [87]. In [172], it was found that the height of illumination sources had a significant impact on crack detection accuracy, with higher placements resulting in improved outcomes and a reduction in false positives, suggesting that optimal lighting conditions are essential for effective crack detection in visual inspection systems.

The SOTA automated inspections lack the following capabilities:

- Utilisation of hardware with illumination techniques such as geometrical illumination (direction and angle) to aid in automated defect detection/analysis, both in laboratory-settings and in real-world scenarios.

- Automation of the data cleaning process utilising thresholding techniques, wherein low-quality images are discarded while high-quality images are forwarded for fur-

ther image processing to enhance concrete defect detection capabilities and thus, reduce the need to repeat image collection in the field, i.e. a more cost-effective, in terms of manpower and time, approach.

- Analysis of the influence of lighting direction and angle on the performance of automated defect detection algorithms.

- Analysis of directional lighting utilised defect detection algorithms for binary and multi-class image classification tasks.

The thesis aims to address these limitations in the current literature.

3.6   Summary

The literature review highlighted the following research gaps which this thesis will be focusing on: 1. Lack of suitable hardware to apply geometrical illumination, a technique that resembles the approach followed by experienced inspectors in manual inspections and which has the potential to enhance crack detection and classification. 2. Automated workflow to assess the quality of acquired images of potential concrete cracks in real-time. Higher quality images will ensure higher efficiency in any detection and classification algorithm. 3. All deep learning algorithms used in crack detection and classification in concrete civil infrastructure rely on RGB or gray scale input images and do not explore novel input configurations or advanced lighting strategies to enhance detection and classification under these poor lighting conditions. None can currently accommodate a 5-channel data input (the data obtained by geometrical illumination). A methodology that can adapt a deep learning algorithm to serve this purpose is currently missing.

Addressing the three points above aims to (1) create an automated visual inspection system that employs directional and angled lighting configurations to enhance the visibility of cracks and spalling in concrete in low-light conditions and (2) integrate directional lighting with commonly used and new deep learning algorithms to improve the accuracy of concrete crack detection for Civil infrastructure while at the same time,

implement an image-quality-based algorithm to ensure only high quality images are used for image processing tasks.

Chapter 4

The Conception and Development of ALICS

Chapter 3 discussed the challenges of concrete crack detection in low-light conditions and highlighted the necessity of geometrical illumination techniques to enhance automated defect detection and analysis in Civil Engineering. For this work, the author of the thesis implemented the novel hardware and software platform called ALICS - Adaptive Lighting for the Inspection of Concrete Structures and investigated the potential of directional lighting and suitable optimal illumination angle, which allows for multi-angle and multi-directional lighting for the inspection of concrete structures.

4.1 Introduction

This chapter presents ALICS, a hardware prototype capable of multi-directional, multi-angled illumination of a concrete surface. The objective is to determine whether using geometrical illumination techniques can improve automated defect detection performance using images captured by ALICS.

4.2 The Hardware Development Journey of ALICS and its iterations

ALICS was a large research initiative led by the author along with two collaborators. The author's work focuses on classifying severe defects such as cracks and spalling in concrete structures using black-box techniques only. Figure 4.1 shows author's primary contributions were focused on tasks indicated by the red boxes, which included working on initial prototypes. The green boxes represent areas where the author contributed in completing the final inspection hardware and portable frame field deployable ALICS rig. Conversely, the areas marked with blue boxes indicates the implementation of the field-deployable ALICS version using a shroud design, which was led by other team members. The following subsections will detail the step-by-step process involved in the conception of lab-based and field-based ALICS designs.

4.2.1 Overview of Hardware

Table 4.1 provides an overview of the hardware components used in this research. It represents a simplified hierarchical structure of the hardware components, starting from individual elements like UR10, machine-vision camera, LED strips, ARDUINO, and expanding into further details like interfacing, inspection hardware (ALICS), and field deployable ALICS iterations.

4.2.1.1 Universal Robot UR10

Robotic inspection of the built environment has been made possible with the help of robots equipped with various locomotion types and sensors. There are several types of such robots, like UAVs, UGVs, marine vehicles, wall-climbing (CL) robots, and cable-crawling (CR) robots. The use of robots for inspection provides a safer alternative to manual inspection. Additionally, automated robotic inspection improves the frequency of inspections and reduces subjectivity in detecting errors.

UAVs are popular for inspecting large areas such as bridges, power lines, and buildings because they can quickly cover a lot of ground and adjust their position for a better

Prototype - 1

Initial Prototypes

Prototype - 2

Final ALICS rig design
(Lab-based)

Inspection Hardwares

Field Deployable ALICS rig
(Portable Frame Design)

Field Deployable ALICS rig
(Shroud Design)
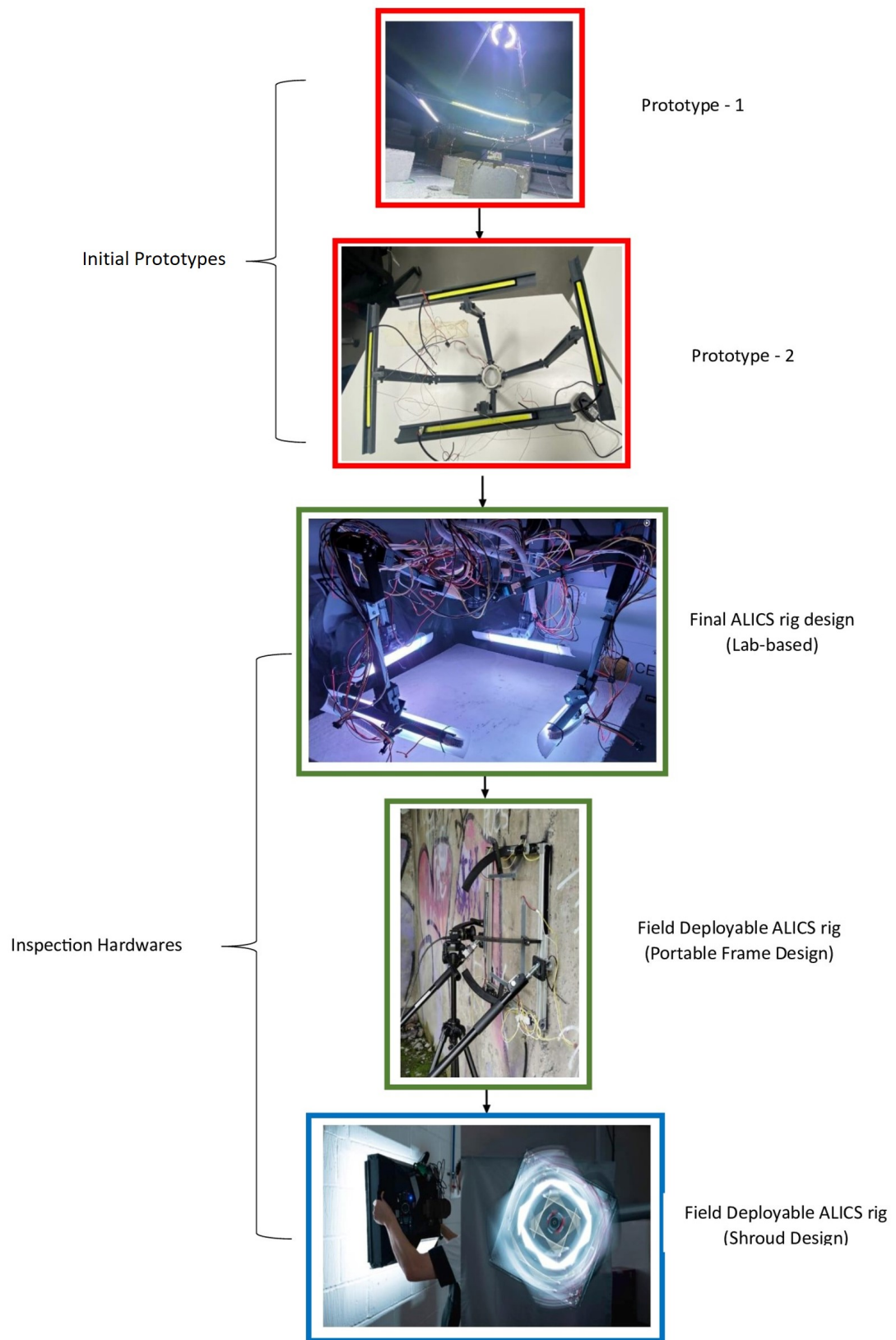
Figure 4.1: The Hardware Development Journey of ALICS. All of these are capable of capturing images of size 5429 × 3458 pixels.

Table 4.1: ALICS Hardware Overview: Components and Specifications.

Component	Description
Universal Robot (UR10)	- Type: Collaborative tabletop robot -360-degree rotation - Axes: 6 (wrist 1, wrist 2, wrist 3, base, shoulder, elbow)
Image capture	- Camera: FLIR Blackfly 1" sensor machine vision camera -Resolution: 5429×3458 pixels - Lens: $f_1 = 8$mm focal length -Working Distance: $D = 360$ mm - FOV: 576mm x 432mm
WS2812B LED Strip	- Type: Individually addressable RGB LED strips -Key Feature: Microcontroller chip in each LED for control -Power Requirement: 5V, 1.5A
ARDUINO	- Type: ARDUINO UNO with ATmega 328 microcontroller - I/O Pins: 14 digital, 6 analog - Operating Voltage: 5V - Clock Speed: 16MHz
Interfacing of ARDUINO with WS2812B LED Strip	- Connection: ARDUINO UNO connected to WS2812B LED strip - Power Source: AC to DC power adapter (5V, 2A) - Additional Components: Capacitor for power stabilisation, resistors for noise reduction.
Inspection Hardware (Lab-based ALICS rig)	- Components: Directional Lighting System with four 3-jointed servo-motorised arms, camera, UR10 robotic arm - Lighting Adjustment: Arms can adjust light incident angles (θ_L) and proximity (P) - Lighting Conditions: Right, Down, Left, Up, Diffused
Field Deployable ALICS rig (Two Iterations)	- First Iteration: Handheld and Portable ALICS Hardware. - Frame Design: Portable and Handheld. - Lighting Control: "Curtain" mechanism for controlled lighting. -Second Iteration: Portable ALICS Hardware with Directional Lighting. - Design: Lightweight aluminium shroud to block ambient light. - Use Cases: Handheld device or can be mounted to a six-axis robot.

view. They are also cost-effective and easy to deploy. However, UAVs have some draw-backs: they can be affected by bad weather, require special licenses and insurance, and are prone to accidents. They also face issues with data security and legal restrictions on where they can fly and what they can inspect. Additionally, their size, weight, and power limitations can make them unsuitable for certain tasks [173, 174].

UGVs offer a simple and robust design, excellent stability for inspection and monitoring tasks. They find applications in building inspections, bridge deck assessments, and floor cleaning. However, UGVs face limitations due to their low profile, which restricts their reach in spaces with high ceilings. They may also struggle on uneven or cluttered terrains. UGVs require collaboration with other robots to overcome these challenges, adding complexity to multi-platform missions [76].

Cobots, or collaborative robots, offer numerous advantages over UAVs, UGVs, crawling, and climbing robots. They are flexible and can perform a wide range of tasks without the need for specialised equipment or modifications. Cobots can also work safely alongside human workers, increasing efficiency and communication [175, 176]. In this research work, a six-axis robot known as a Universal Robot (UR10) is utilised as shown in Figure 4.2.

UR10 is a small yet robust collaborative tabletop robot ideally suited for light assembly tasks. The selection of the UR10 was based on several factors. Firstly, it is certified for safe physical human-robot interaction, ensuring workplace safety. Secondly, it offers a significant cost advantage compared to the current market alternative, MAKO [177], being at least 10 times less expensive. Additionally, its 10 kg payload capacity adequately supports the lighting rig being considered. Furthermore, the UR10's large reach of 1.3 meters exceeds that of many similar collaborative robots, enabling effective scanning of large concrete surfaces [178]. The UR10's 6-axis robot arm is capable of full 360-degree rotation across all six axes: wrist 1, wrist 2, wrist 3, base, shoulder, and elbow. In this particular arrangement, the UR10 plays a crucial role as an important component that is responsible for executing accurate movements and enabling lighting modifications while capturing images. This harmonious integration of the UR empowers the ALICS system with the adaptability necessary to execute concrete struc-

ture inspection tasks.
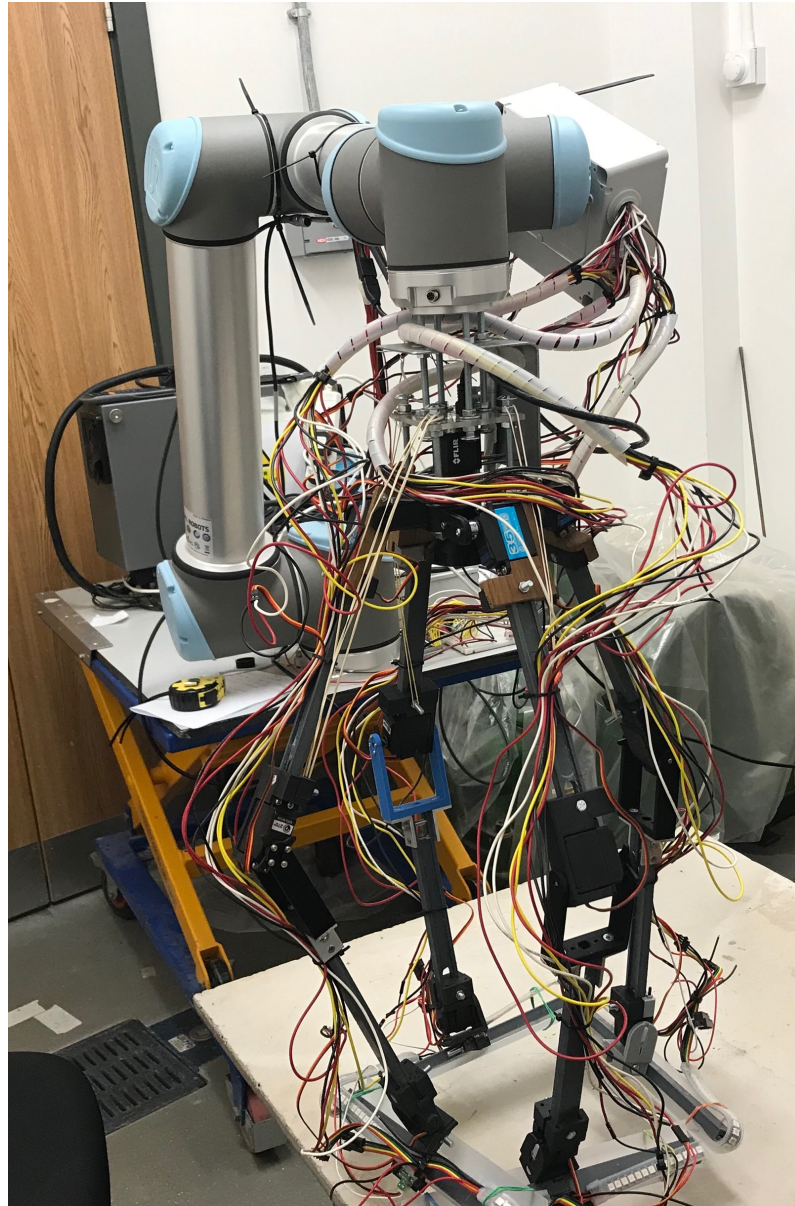
Figure 4.2: Integration of ALICS onto a Six-Axis Universal Robot UR10.

4.2.1.2 Image capture

According to engineering standards, crack widths within the range of 0.2mm to 0.4mm in concrete structures should be identified for further action [17]. Therefore, the objective is to achieve a minimum spatial resolution of ≤ 0.1 mm/pixel during image

capture.

4.2.1.3 Hardware and settings

An illustration of the image acquisition hardware is provided in Figure 4.3. In this work, a FLIR Blackfly 1" sensor machine vision camera, shown in Figure 4.4 was used. The field of view defines the visible area of the object under inspection. The distance from the front or initial surface of the lens to the object under inspection is called as Working Distance (D). The dimension of a camera sensor's active area, often indicated horizontally or vertically is sensor size. The relation between all these parameters is indicated in equation 4.1 and illustrated in Figure 4.5 [179].

$$\text{Focal length} = \frac{\text{sensor size} \times \text{working distance}}{\text{field of view}} \tag{4.1}$$

It has an imaging resolution of 5429 × 3458 pixels and a focal length, $f_l = 8$mm. Together, these provided a feature resolution of ≤ 0.1mm/pixel at a working distance of $D = 350$mm, with a Field of View (FoV), of 574mm × 383mm [180]. For various imaging applications, the lightweight design of FLIR Blackfly® S cameras make them a popular choice. These cameras are equipped with several advanced features that make them highly compatible with machine vision systems. Precise control over exposure, gain, white balance, and colour correction are some of the features that contribute to the effectiveness of the ALICS system. In achieving the desired image resolution and quality for concrete structure inspection applications, selecting the FLIR Blackfly 1" sensor camera and its accompanying lens played a crucial role [181]. The camera specifications are listed in Table 4.2.

The depth of field, DoF, defines the distances at which objects remain in focus:

$$DoF = \frac{2 \cdot F_d^2 \cdot f_n \cdot C}{f_l^2}, \tag{4.2}$$

where $f_n = \frac{f_l}{a_d}$ is the f-number i.e., the ratio of the focal length f_l, to aperture diameter a_d. C is the circle of confusion: a subjective limit that defines an acceptable level of
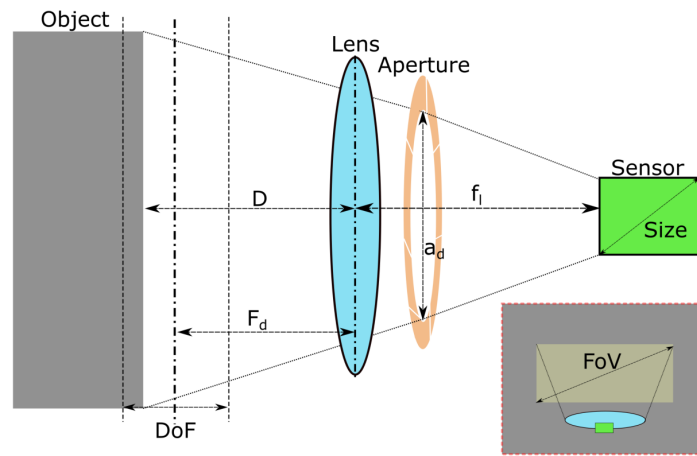
Figure 4.3: Illustration of variables during image capture. Light from an imaged object passes through the lens and aperture onto the camera sensor.

Figure 4.4: FLIR Blackfly 1" sensor machine vision camera.

Table 4.2: FLIR Blackfly S USB3 Camera Specifications taken from [182].

Specification	Value
Firmware Version	1705.0.124.0
Resolution	5429×3458
Frame Rate	18 FPS
Megapixels	20MP
Chroma	colour
Sensor	Sony IMX183, CMOS, 1"
Readout Method	Rolling shutter with global reset
Pixel Size	2.4μm
Lens Mount	C-mount
ADC	10-bit / 12-bit
Minimum Frame Rate	1FPS
Gain Range	0 to 27dB
Exposure Range	69μs to 30s
Acquisition Modes	Continuous, Single Frame, Multi Frame
Partial Image Modes	Pixel binning, decimation, ROI
Image Processing	colour correction matrix, gamma, lookup table, hue, saturation, and sharpness
Sequencer	Up to 8 sets using 5 features
Image Buffer	240MB
User Sets	2 user configuration sets for custom camera settings
Flash Memory	6 MB non-volatile memory
Opto-isolated I/O	1 input, 1 output
Non-isolated I/O	1 bi-directional, 1 input
Serial Port	1 (over non-isolated I/O)
Auxiliary Output	3.3V, 120mA maximum
Interface	USB 3.1
Power Requirements	8 - 24V via GPIO or 5V via USB 3.1 interface
Power Consumption	3W maximum
Dimensions/Mass	29mm x 29mm x 30mm / 36g
Humidity	Operating: 20% to 80% (no condensation), Storage: 30% to 95% (no condensation)
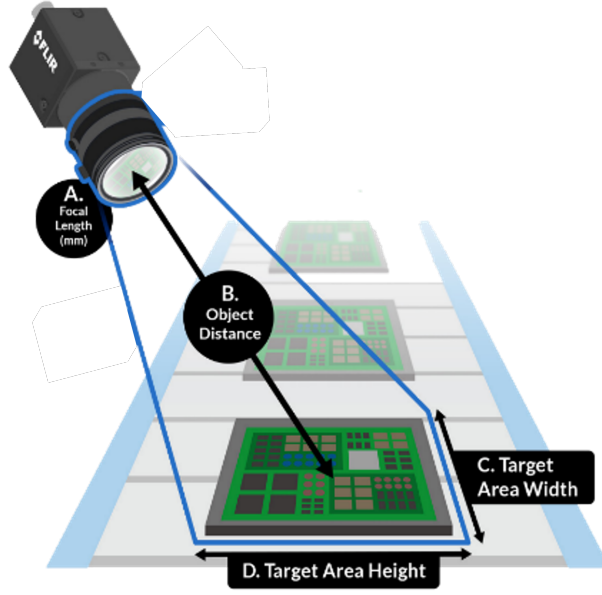
Figure 4.5: Variable Distance from Lens to Concrete, modified from [179].

loss of focus. It is conventionally calculated by dividing the diagonal size of the camera sensor by 1500 [183].

Equation (4.2) shows that controlling the f-number, f_n and focal distance, F_d will vary the available DoF. f_n is typically limited to the following values: 1.4, 2, 2.8, 4, 5.6, 8, 16. When selecting f_n, there is a trade-off for a constant D and F_d: opting for a lower f_n allows for faster image capture (higher exposure rate) and reduces diffraction, but leads to blurred corners and a narrower depth of field (DoF). Diffraction causes a loss of sharpness in an image due to interfering light waves. For consistency, all images captured in this work were taken at a constant f_n of 8 and F_d of 250. This provided a balanced approach for avoiding diffraction and blurred corners, with sufficient DoF = 172 mm, covering D from 200 to 350 mm.

4.2.1.4 WS2812B LED Strips

The implementation of the ALICS platform included the effective integration of WS2812B LED strips to enhance the adaptive lighting system for concrete structure inspection.

Individually addressable RGB LED strips, such as the WS2812B, are widely used for lighting purposes in various hardware setups [184, 185]. This particular strip was chosen due to its simple interface, fast response time, and cost-effectiveness compared to other options [186]. They are known for their unique ability to individually control the brightness and colour of each LED [187].

Figure 4.6 shows an example of this LED strip, which consists of multiple WS2812B LEDs connected in series. What distinguishes these LEDs is that they come equipped with a micro-controller chip embedded within each LED. This chip enables communication through a single-wire interface, allowing it to control numerous LEDs using just one digital pin of devices like an ARDUINO.

Figure 4.6: WS2812B LED Strip.

The microcontroller chip inside each LED manages its colour and brightness. These RGB LEDs can emit a broad range of colours by combining red, green, and blue light, which is a key feature of the WS2812B LEDs. Moreover, these LED strips are highly adaptable and can be cut to various lengths depending on specific requirements.

It was crucial to ensure that the WS2812B LED strip receives adequate power for optimal performance. At full brightness, each LED on the strip consumes approximately 50mA of current. The total current required can be determined using the equation:

$$\text{Total Current (A)} = \text{Number of LEDs} \times \text{Current per LED} \qquad (4.3)$$

For 30 LEDs, the total current was 1.5A.

$$\text{Total Current} = 30 \text{ LEDs} \times 0.05 \text{ A/LED}$$
$$\text{Total Current} = 1.5A$$

To ensure consistent and reliable operation, it is crucial to select a power source that can deliver a minimum of 1.5A of current. In this case, an AC to DC power adapter with a 5V output and a 2A current rating was chosen. This choice exceeds the minimum current requirement, providing a 2A current output, thus guaranteeing that the LED strip is powered effectively. This arrangement allows the LED strip to function optimally without encountering any power-related issues during the experiments.

4.2.1.5 ARDUINO

ARDUINO is a computing platform that is known for its versatility in developing interactive objects and interfacing with different software applications. It is an open-source hardware and software project that features a user-friendly I/O board and utilises the easy-to-learn processing/wiring language within its development environment. It has been utilised in numerous crack detection studies due to its affordability and user-friendly technology, attracting engineers, designers, and individuals alike [188, 189]. At the heart of ARDUINO lies the ATmega328 microcontroller, which is capable of operating at 5V and comes with six analog input pins and fourteen digital I/O pins. The board also features 32KB of flash memory, 2KB of SRAM, and 1KB of EEPROM, and operates at a clock speed of 16MHz. One of the most popular ARDUINO boards is the ARDUINO UNO, which can be powered using either a USB connection or an external power supply. However, it is important to note that the recommended voltage range for safe operation is between 7 volts and 12 volts [187, 190].

The key power pins on an ARDUINO board include VIN, which is the input voltage pin that accepts power from an external source. The 5V pin is used to power the microcontroller and other components on the board, while the 3V3 pin generates a 3.3-volt supply as shown in Figure 4.7. The board also has a GND pin, which is used as a ground reference. Furthermore, ARDUINO has 14 digital pins, each of which can function as an input or output using functions like pinMode(), digitalWrite(), and digitalRead(). These pins operate at 5 volts and can handle up to 40mA of current, with additional internal pull-up resistors. Moreover, the UNO model has six analog inputs (A0 through A5), which expands its capabilities for processing analog signals.
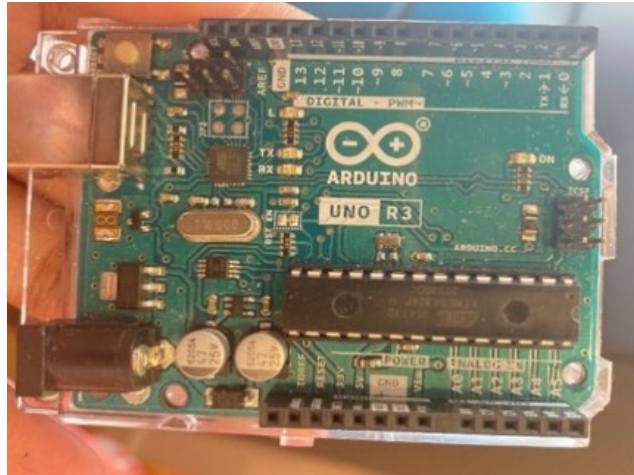
Figure 4.7: ARDUINO Board.

4.2.1.6 Interfacing of ARDUINO with WS2812B LED strip

To power the LED strip, the 5V pin on an ARDUINO UNO board is utilised. In order to stabilise the power supply and minimise electrical noise, a capacitor with a capacitance ranging from 100 to 1000μF was connected between power and ground. Additionally, to further reduce electrical interference, 470-ohm resistors were placed between the digital input pin of the ARDUINO UNO and the data input pin of the LED strip. Upon turning on the circuit, it will operate in accordance with the programmed code. See Figure 4.8 for a visual representation of the setup [187].
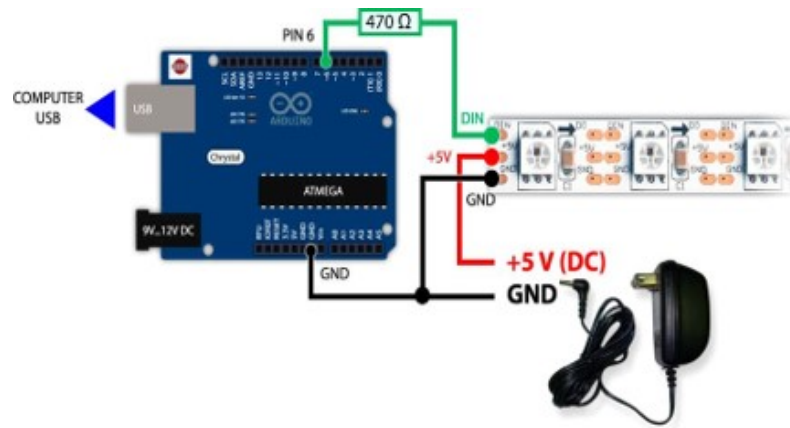
Figure 4.8: Interfacing of WS2812B LED Strip with ARDUINO taken from [187].

4.2.2 Failed Prototypes

The ALICS platform underwent several iterations, as shown in Figure 4.9, and Figure 4.10 to develop adaptive lighting for concrete structure inspection.
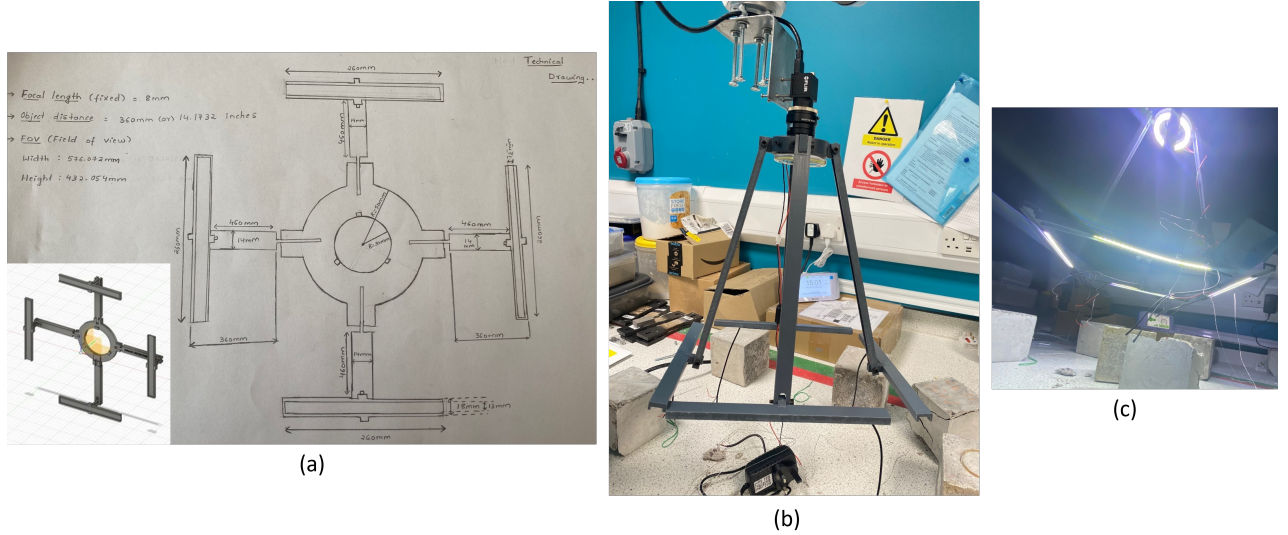
Figure 4.9: Prototype 1 - a) Lighting Rig Sketch and its digital image, b) Lighting Rig design 1, c) A lighting rig consisting of a central ring light surrounded by four bar lights is fitted to the UR10, which is mounted with a machine vision camera.

4.2.2.1 Prototype 1

The author implemented the initial prototype, which involved designing a lighting rig for multi-directional illumination. The Machine Vision Lens calculator helps to determine the field of view, focal length or working distance by providing any of the two parameters. For example, if the sensor size of a machine vision camera is 1" and the focal length is 8mm, and the desired FoV width is 576mm, inputting these values into Focal length of FOV lens calculator in [191], yields a working distance of 360mm and a height of the area to be inspected as 432mm. This provides the target area that needs to be inspected and the working distance. Subsequently, the initial design of the lighting rig was implemented to achieve these parameters as shown in Figure 4.9 (a). For a working distance of 360mm, the boom arms for the LEDs must extend at least 360mm to adequately illuminate the surface. The FOV region thus
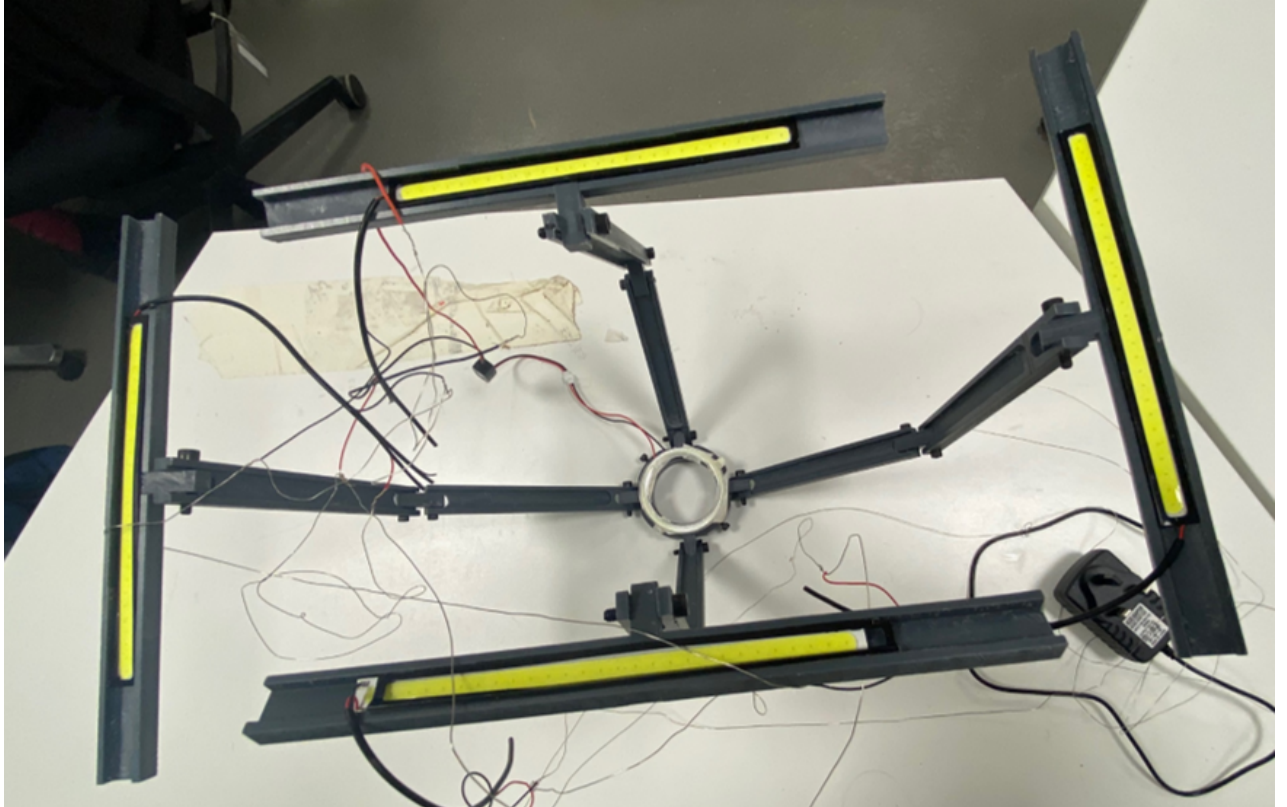
Figure 4.10: Prototype 2: Lighting rig design 1 in Figure 4.9 has only arms, but this modified design consists of shoulder, wrist and elbow allowing multi-angle lighting, with ring light in the centre surrounded by four bar lights.

obtained is (W×H) 576mm×432mm. However, this rig was only an initial prototype, and industrial applications would require hundreds of images to be taken to do one full lateral movement, capturing the entire surface of a structure in a detailed, side-to-side manner. Following this, a multi-angle lighting rig has been developed and was fitted to the Universal robot mounted with a machine vision camera. Each boom arm can be rotated at two points and is held in place with nuts and bolts. This allows to vary lighting angle manually if required. The device is mounted to the camera via an aluminium filter plate which is screwed onto the end of the lens barrel, as shown in Figure 4.9 (b) and Figure 4.9 (c).

The lighting rig consists of one ring light surrounded by four bar lights as shown in Figure 4.9. The ring light is a 12V, 104mm diameter COB LED which can uniformly light the surface. The bar lights are 12V, 170mm long COB LEDs. All lights have a colour temperature of 6000K. This lighting rig was mounted onto machine vision camera (section 4.2.1), which was further mounted onto the universal robot (section 4.2.1) to capture images of cracked concrete slabs. However, this design lacked flexibility for achieving precise multi-angle lighting due to the fixed structure of bar lights. The adaptability of the system was also limited by its fixed working distance, which prevents clear views of thin cracks or surfaces at varying distances. Additionally, the use of a ring light, along with bar lights resulted in bright spots and inconsistent lighting throughout the illuminated area, creating shadows in certain regions. Bar lights have limitations in providing directional lighting, making it difficult to illuminate specific areas accurately and effectively highlighting cracks.

4.2.2.2 Prototype 2

In the second iteration of ALICS, the author concentrated on addressing the limitations of the initial model. The design of the lighting rig was adjusted to include shoulder, elbow, and wrist components, enabling multi-directional and multi-angled illumination as shown in Figure 4.10. The improved lighting rig consisted of one central diffused ring light surrounded by four bar lights placed on the wrists of the lighting rig. This device was mounted on to the camera via an aluminium filter plate which was screwed onto the

end of lens barrel. It was observed that ring lights caused inconsistent lighting across the concrete surface, resulting in uneven illumination. The lack of flexibility in the design of the lighting rig makes it difficult to achieve accurate multi-angle illumination, especially around shoulder, elbow, and wrist joints. Additionally, bar lights may have limited directionality which can make it challenging to precisely illuminate specific areas such as cracks. Both prototypes faced challenges associated with fixed working distances, which may limit adaptability to different scenarios. The drawbacks related to the use of ring lights and bar lights were common to both prototypes and involve issues like bright spots and limited directionality. The movement of the lighting rig is limited. This led to the development of final ALICS device.

4.2.3 Final ALICS device

To address the limitations of Prototype 1 and 2 models, the author replaced bar lights with four LED strips to create an adaptive lighting system and removed the ring light. This modification made the lighting rig adaptable to support multi-directional and multi-angle lighting, an important feature for ALICS functionality. However, the manual implementation of multi-angle, multi-directional lighting in a scene necessitated automation. To achieve this, the author contributed to the engineering of the lighting rig featuring three-jointed servo motorised arms to control the shoulder, elbow, and wrist joints. These arms were designed to manoeuvre the LED strips and project light onto the concrete surface at angles of incidence ranging from $\theta_L = 10°$ to $60°$.

The motorised arm had a maximum limit of 60 degrees for upward movement. Due to its geometric limitations, the lighting rig was incapable of ascending beyond 60 degrees, restricting the movement of the motorised arm. The author's decision to investigate the parameter space to understand how various angles and lighting directions influence the crack inspection process, was driven by the effectiveness of low-angle lighting demonstrated in previous research [9, 192]. The selection of intervals of 10 degrees between the angles was motivated by [193], as coarser intervals simplify data interpretation and provide clear distinctions between angles. Finally, the ALICS platform was mounted onto a robotic arm fitted with machine-vision camera, allowing precise

adjustments using a UR10, as shown in Figure 4.11.
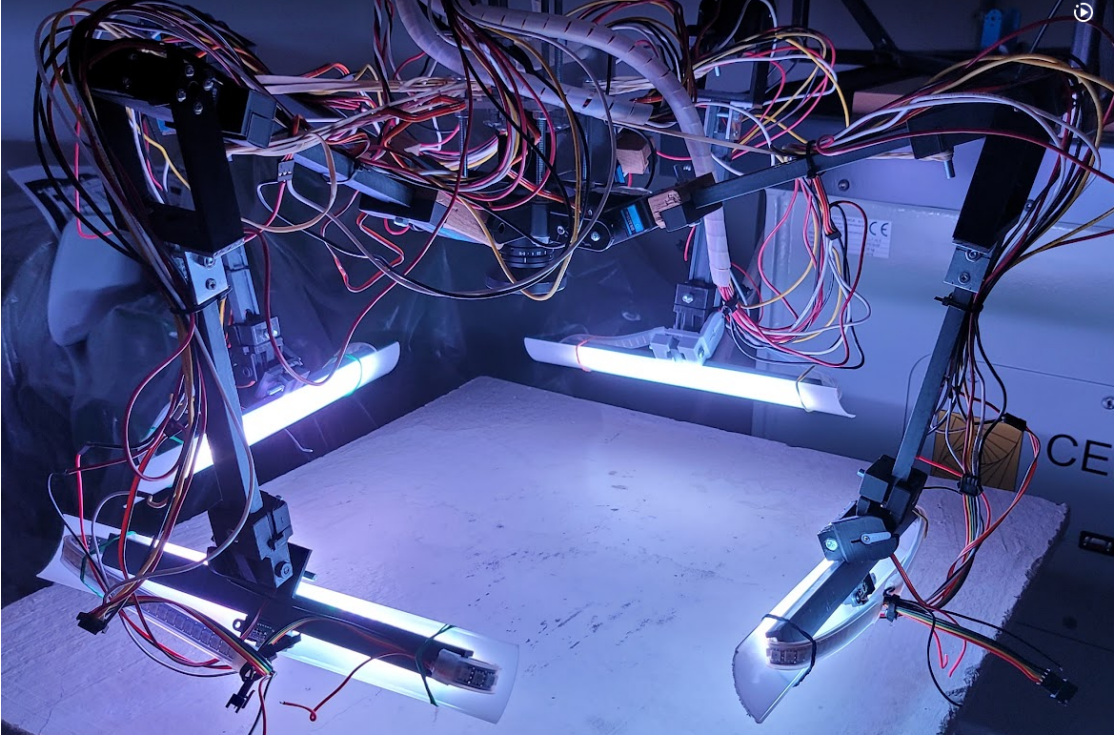
Figure 4.11: ALICS: four 3-jointed servo-motorised arms with WS2812B RGB LED strips surrounding a camera, mounted perpendicularly to a UR10 robotic arm. Each arm can move individually and project light from the LEDs incident onto the surface at angles ranging from 10° to 60° and at a required proximity.

4.2.3.1 Lighting rig and angle control

Figure 4.12 illustrates the design of a single arm. The three servo motors: shoulder (S), elbow (E), and wrist (W), are utilised to incident light on the surface at angle, θ_L and proximity, P. Other variables, such as working distance, D and arm lengths, L1 and L2 also influence the motor angles (Θ_1 and Θ_2). This may lead to several possible solutions for the motor angles, necessitating a least-squares minimization approach. From Figure 4.12, considering the arms as vectors, the horizontal components can be summed as:

$$A + L_1 \cos\theta_1 = L_2 \cos\theta_2 + P \cos\theta_L, \tag{4.4}$$
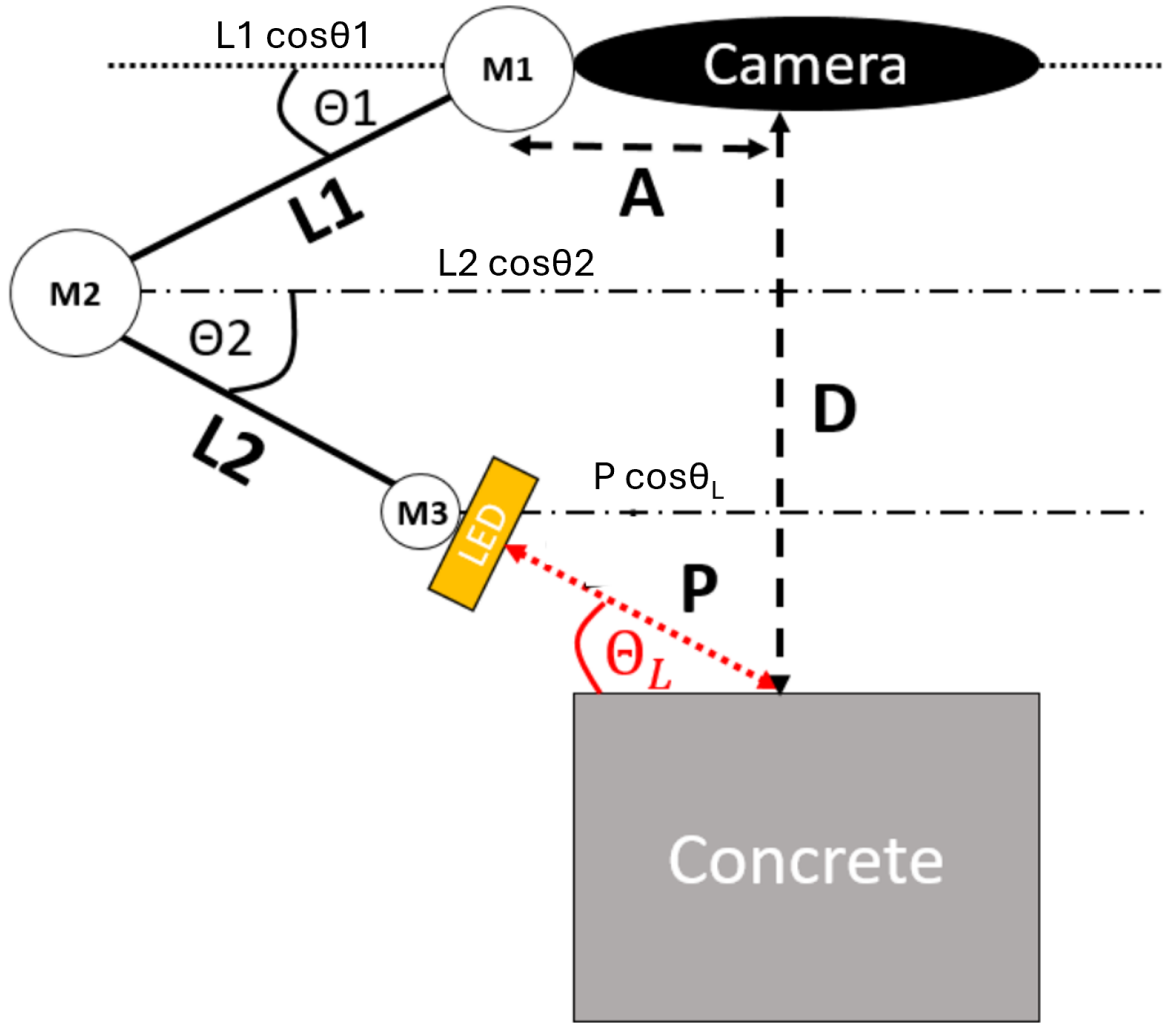
Figure 4.12: Illustration of the 3-jointed motorised LED arms: required angle of incident light, Θ_L, is achieved by moving motors M1, M2 and M3 attached by arms of length L1 and L2. Working distance, D, and LED proximity, P, also influence motor positions.

and the vertical components are:

$$D = L_1 \sin \theta_1 + L_2 \sin \theta_2 + P \sin \theta_L, \qquad (4.5)$$

where $L_1 = 260$ mm and $L_2 = 270$ mm are the lengths of the arms, $A = 80$ mm is the distance from the centre of the lens to the shoulder joint at M1, and D is the working distance of the camera or distance from M1 to the surface. The shoulder (M1) and elbow (M2) joint angles were constrained to $-90° \le \theta_1 \le 90°$ and $0° \le \theta_2 \le 180°$ respectively.

Equations (4.4) and (4.5) were solved using least-squares minimisation to find solutions for θ_1 and θ_2 for set values of θ_L, P and D during image capture. Since these are the angles between the arms and the horizontal, a second calculation must be made to determine the angles of the servo motors. M1 is stationary, so M1 $= \theta_1$. M2 is relative to arm L1, i.e. M2 $= \theta_1 + \theta_2$, which is the angle between L1 and L2. Finally M3 is relative to L2, so M3 $= (180° - \theta_2) + \theta_3$, which is the angle between L2 and the LED projection. An example result of this is given in Table 4.3.

Table 4.3: Example angle calculation by minimisation.

Category	Variable	Value
	D	300 mm
Desired Condition	P	300 mm
	θ_L	20°
	θ_1	-15°
Required Angles	θ_2	80°
	θ_3	20°
	M1	-15°
Motor Angles	M2	65°
	M3	120°

4.2.3.2 Distortion correction

All camera lenses have distortion coefficients that can be calculated and compensated for checker board calibrations. In this work, numerous images of a checkerboard pattern at various distances and angles were captured. The pattern was automatically identified

and used to determine the coefficients required to reconstruct a straight checkerboard pattern. These coefficients were then used to correct for distortion on every image captured throughout this work.

4.2.3.3 Exposure and White Balance

White colour balance and exposure settings can be automatically calculated and adjusted by a camera's on-board algorithms, but this is a slow process relative to the rapid changes in lighting conditions created by ALICS. Having preset values for exposure and white-balance for each lighting condition ensures exposure changes are fast and consistent.

For this work, it was assumed there is no ambient lighting interference. This is achieved using field deployable ALICS rigs using shrouding as described in section B, but was easily adjusted in a laboratory settings by directly mounting it onto the universal robot. The exposure is only dependent on the illumination used.

A sensitivity study was conducted to find the camera's recommended exposure and white-balance settings across all lighting and working distance configurations (e.g. from high angle and diffused at a large working distance, through to low angle and directional at a short working distance). The recommended exposure was found using the in-built auto-exposure algorithm of the camera [181]. In summary, this algorithm compares a histogram of pixel brightness to an optimal mean and variance to find the optimal exposure. While white-balance settings remained fairly constant, exposure settings were found to be dependent on lighting angle when images were illuminated from a single direction. The exposure setting required for a lighting angle θ_L is:

$$E_{\theta_L} = \left[2 - \left(\frac{50 - \theta_L}{50} \right) \right] \cdot E_{50} \tag{4.6}$$

where E_{50} is the exposure setting required during diffused lighting at a $50°$ illumination angle (calibrated once at the beginning of a scan). Essentially, it was found that the auto-exposure for directional lighting was double the diffused exposure at 50-degree lighting angle, and varied linearly with angle. This means a single auto-exposure cal-

culation can be done at 50-degrees during diffused lighting and all other exposures can be computed using equation 4.6.

4.2.3.4 Lab dataset acquisition

With the final ALICS device, it was possible to acquire lab datasets to implement and evaluate image-processing algorithms that leverage directional lighting. Loads were applied to reinforced concrete slabs producing various cracks on the surface with different lengths, widths and orientations. Magnified ruler was used to measure cracks ranging from 0.1 to 1mm in width and 10mm to 500mm in length.

Distinct 5429×3458 pixel images of concrete surfaces were acquired. Each area was captured with single-direction lighting from orientations $O = R, D, L, U, A$, where R, D, L, U, A are lighting from right, down, left, up, and diffused (i.e., all directions) respectively. For each lighting orientation, incident light angles θ_L ranged from $10°$ to $60°$ in steps of $10°$. An example of an area of the slab during varying lighting conditions is shown in Figure 4.13, with insets comparing a single 224×224 pixel block within the larger image. Qualitatively, it is observable that this block exhibits enhanced shadowing during the lower angle of $10°$ when compared to $50°$.

4.2.4   Field deployable ALICS rig

The ALICS design was effective in a controlled laboratory environment, but its limited manoeuvrability, and requirement for it to use in complete darkness conditions made it impractical for outdoor applications. To address these limitations, two new iterations (Portable frame design, shroud design) of ALICS were developed, each enhancing the portability and adaptability, and built upon my original idea of a lighting rig, as shown in Figure 4.1.

The portable frame design incorporated LEDs inside a frame, and utilised a curtain mechanism for regulating lighting conditions. Subsequent development were carried out by another collaborator. The redesign specifically targeted enhancements in manoeuvrability, and adaptability to varying levels of outdoor lighting. The shroud design was developed by two other researchers utilising the author's initial lighting rig design
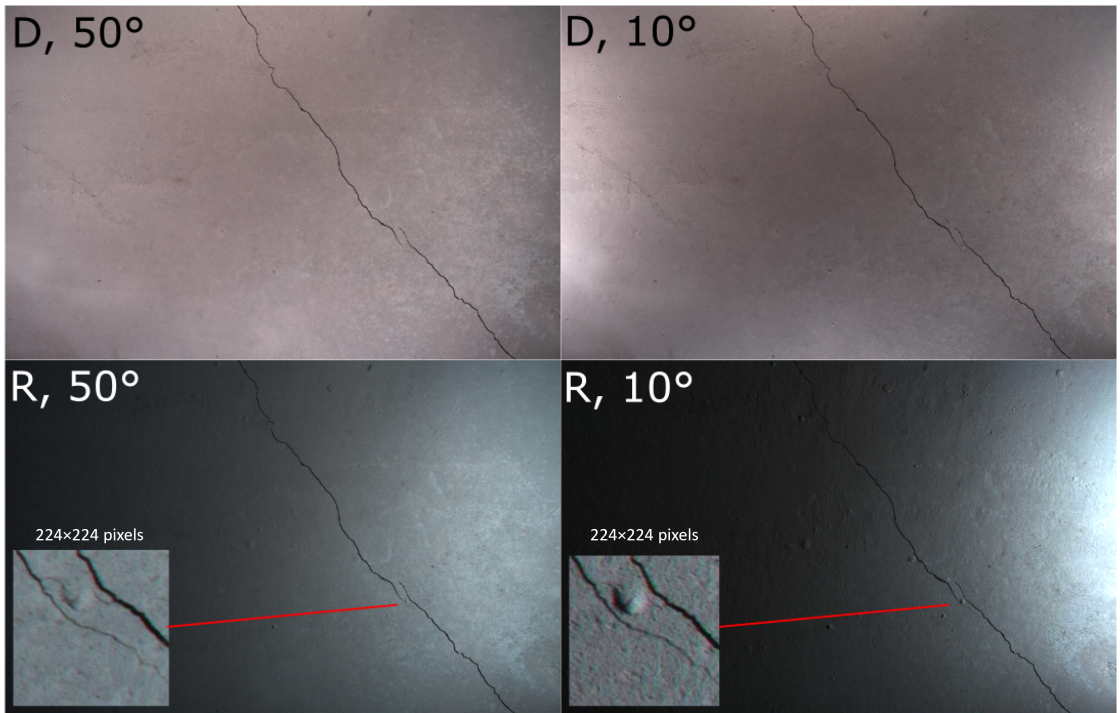
Figure 4.13: Example of lighting conditions at 50° and 10°. Inset: a single 224×224 block indicating enhanced shadowing.

and components, including LED strips for illumination and a machine vision camera positioned on top of the shroud. This shroud effectively blocks out surrounding ambient light, ensuring optimal system performance in both bright outdoor settings or darker environments. Regardless of outdoor lighting levels, the shroud creates a darker setting for image capture, allowing the system to operate efficiently in various challenging lighting situations.

The field deployable ALICS rigs were designed to be used as handheld devices for on-the-spot inspections or mounted onto a six-axis robot for data capture under laboratory-settings.

4.2.5 Verification of ALICS hardware based on in-situ inspections

The aim of real-world deployment was to collect data from concrete assets in public areas, and demonstrate the practical applications of ALICS beyond laboratory settings. Therefore, field experiments were carried out in various locations in Glasgow, including

the Carnbooth area, Kelvin bridge, Bowling Harbour area, and Babcock sites to acquire images using a field-based ALICS rigs. The author, along with other researchers carried out inspections in the Carnbooth area using portable frame design (See Figure 1), and in Babcock using a shroud design (See Figure 3). Inspections at other locations were conducted by other researchers alone and are therefore not detailed in this thesis [194, 195].

Figure 4.14 indicates how images are captured using portable frame design in Carnbooth area of Glasgow. The inspections were conducted under normal weather conditions. However, the setup presented several challenges: one person needed to enter the curtain to block ambient light and manually adjust multiple lighting angles as shown in Figure 4.14, making the process inconvenient and time-consuming. Performing inspections required a minimum of two knowledgeable individuals. Additionally, establishing a connection between the laptop and the portable frame design involved numerous wires, increasing complexity and making management little difficult. Moreover, this design was not suitable for performing inspections within tunnels. Despite the hardware complexities, the captured images were clear and detailed. Concrete images showing thin cracks, thick cracks, and spalling were successfully obtained and subsequently used for the analysis described in Chapters 5, 6, and 7.

The practical challenges of using the portable frame design in real-world conditions led to the development of a shroud-based design and was used to capture images at the Babcock site as shown in Figure 4.15. It uses "dome" shaped aluminium structure to house the LEDs and provide darkness, enabling to capture images without requiring a curtain. While the method of providing darkness varies between the two rigs, the positioning of the inspection hardware (Camera and LEDs) is identical. The design also automates the capture of images at multiple angles, making it ideal for on-site applications. Additionally, the computer is directly integrated onto the device, simplifying the analysis process. Only one person is required to operate it, and even those with no prior knowledge can use it, as it automatically runs all iterations (multiple directions and angles) with the press of a single button, capturing images from various directions and angles. It can also carry out the inspection process in difficult-to-access

Figure 4.14: Conducting fieldwork using a portable ALICS frame at Carnbooth area.

locations like inside of tunnels. The images taken from this site included both cracks and spalling, and these images were utilised for further analysis conducted in the subsequent chapters. Therefore, these campaigns finalised the design of the hardware which was then deemed satisfactory for both lab and field use under any lighting conditions.

Figure 4.15: Conducting fieldwork using a shroud-based ALICS rig at Babcock site.

A total of 45 images were obtained and image acquisition time is 4 seconds per sample. Apart from real-world structures, 5 images were captured under laboratory settings using Figure 4.11. The images captured in the field by the author, along with

those taken by other researchers, have been used to determine the optimal angle, study the effectiveness of directional lighting, develop image-quality based algorithms, and implement advanced deep neural network models utilising multi-directional lighting.

4.2.6 Flowchart showing crack detection, classification and segmentation process

Utilisation of a hybrid approach, outlined in the flowchart in Figure 4.16, involves a combination of techniques for efficient crack detection, classification and segmentation. ALICS is an extensive project which involves author and two other researchers. The red boxes in Figure 4.16 denote primary contributions of the author of this thesis.

4.2.6.1 R-CNN for Feature Detection

The application of Faster Region-Based Convolutional Neural Network (R-CNN), led by McAlorum et al., (highlighted in green in the flowchart), is employed for the detection and localisation of features, such as cracks, against a background. Bounding boxes are generated, the image edges are trimmed, and the remaining portion is cropped into 224×224 pixel blocks. Blocks within bounding boxes (with overlap) are labelled as "feature" and undergo further classification. Blocks outside bounding boxes are designated as "background."

4.2.6.2 Visual Geometry Group-16 (VGG-16) Classification

The author of this thesis led the work, as indicated by the red box in the Figure 4.16. The VGG-16 CNN is utilised when a bounding box containing a crack is identified by R-CNN. The presence of cracks initiate white-box algorithm, which is used for pixel-level segmentation, while negative detections result in the labelling of all pixels as background.

4.2.6.3 White Box Segmentation

Dow et al., (shown in the blue box in the figure) managed the application of white-box techniques to positive blocks i.e., cracks. The goal is to implement pixel-level

segmentation to distinguish "cracked" and "background" pixels.

The hybrid approach involves running the white box on a reduced set of blocks filtered by both R-CNN and VGG-16. This improves accuracy by eliminating background noise in "negative" blocks and reduces false-positive pixels by filtering out non-crack features. The description of white-box techniques is given for context only, and is outside of the scope of this thesis.

4.3 Defining the optimum angle with ALICS

During the performance assessment of the ALICS, laboratory data was collected by applying various loads to reinforced concrete slabs as explained in previous section.

Table 4.4: Auto-Exposure values for different directions.

Direction	Auto-Exposure value (lux-seconds)
Diffused	470,000
Right	1.63 million
Down	1.65 million
Left	1.68 million
Up	1.71 million

The initial dataset included 9 concrete samples captured under diffused lighting conditions in both laboratory using Figure 4.11 and in real-world using Figure 2, with lighting angles ranging from 10 to 60 degrees, comprising 54 samples i.e., 9 images captured under diffused lighting at 6 distinct angles. Following this, the optimal lighting angle was identified, and the same 9 samples were captured using directional lighting (R, D, L, U, and A) in both laboratory and real-world settings, resulting in a collection of 45 images. Consequently, the entire image, measuring $5429{\times}3458$ pixels, was divided into 360 smaller images, each with dimensions of $224{\times}224$ pixels, to meet the input requirements of the model. The dataset was manually labelled as crack (pixels corresponding to cracks), and no crack (pixels without cracks). The initial angle-based data is illustrated in Figure 4.17, while Figure 4.18 shows the directional lighting dataset.
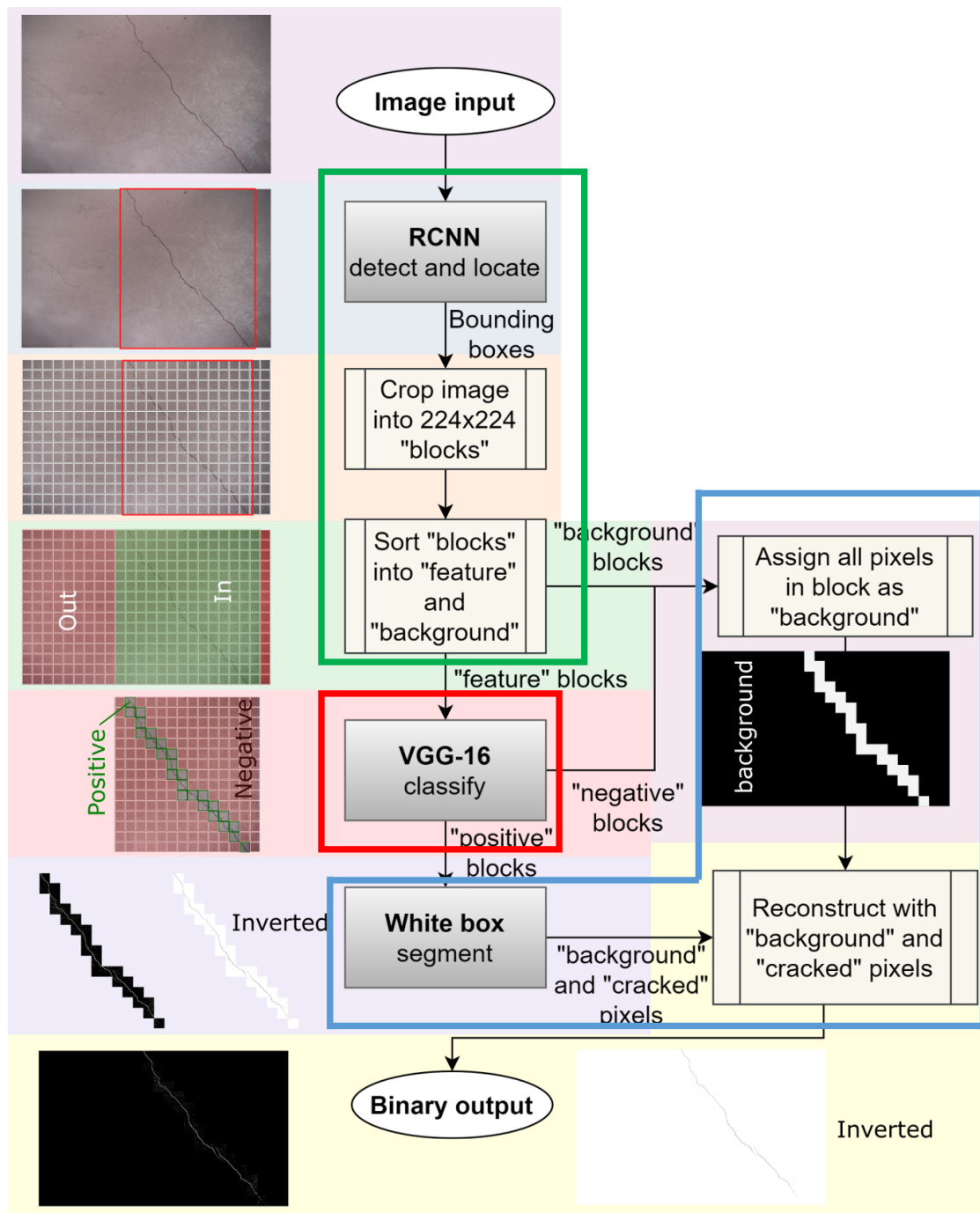
Figure 4.16: Annotated flowchart showing the hybrid crack detection, classification and segmentation approach.
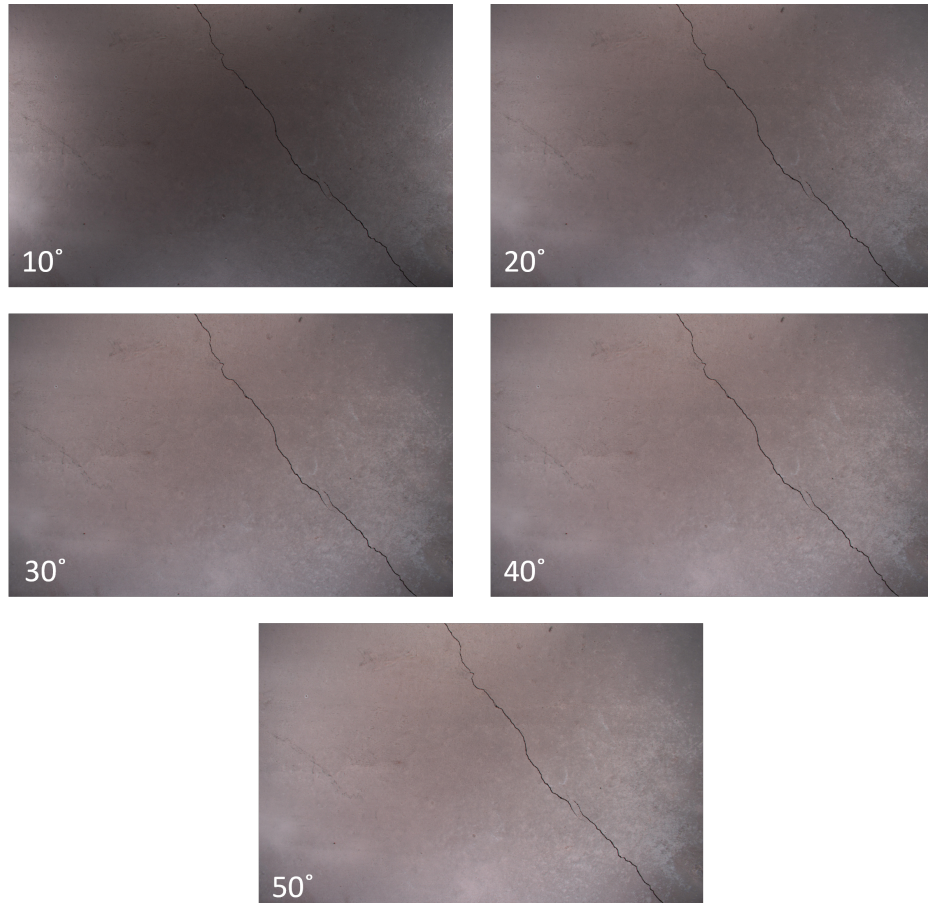
Figure 4.17: Captured image areas (5429×3458 pixels) by ALICS, when light is incident at 50°, 40°, 30°, 20°, and 10° respectively.
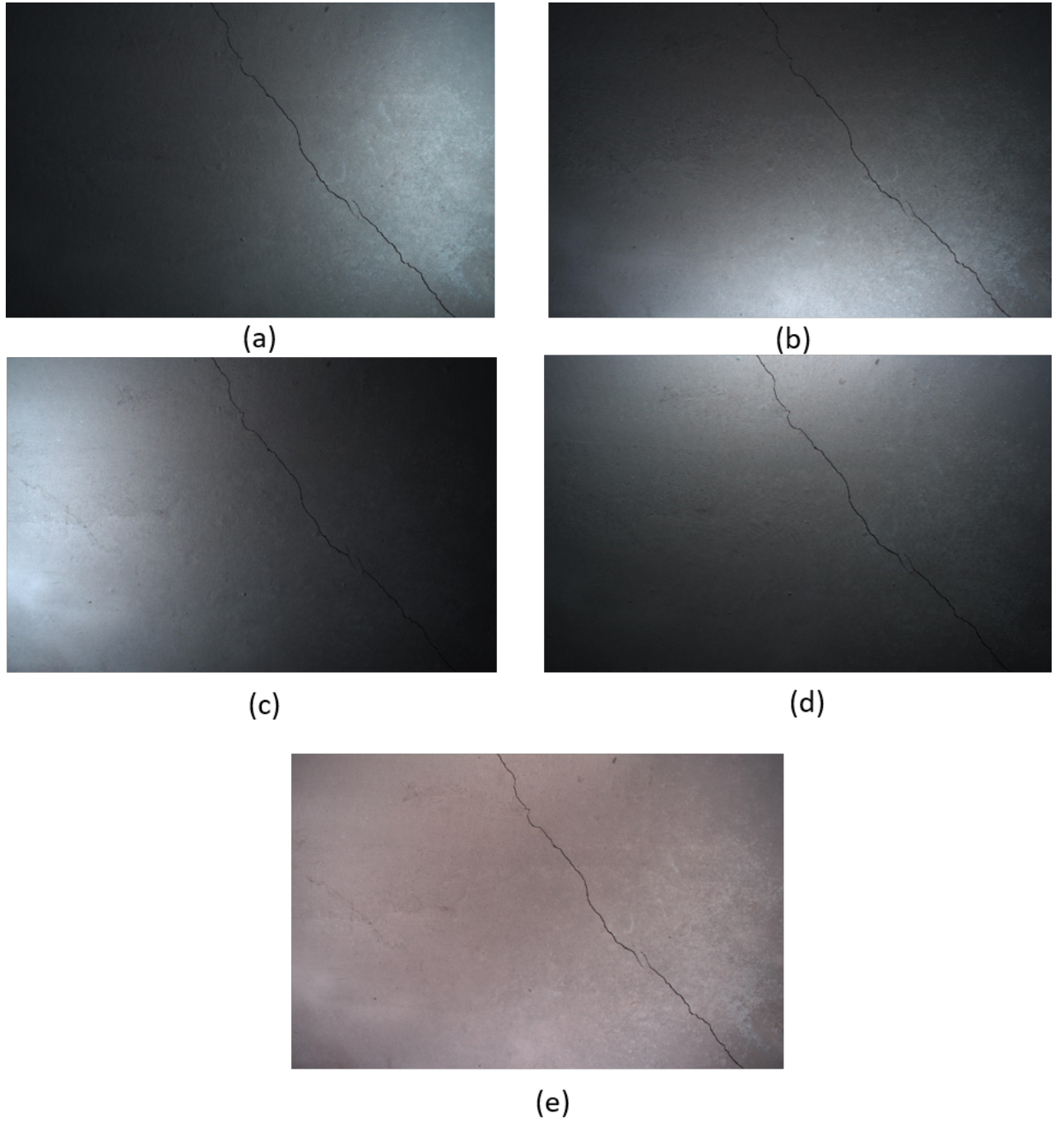
Figure 4.18: Captured image areas (5429×3458 pixels) by ALICS, when light is projected from (a) right(R), (b) Down(D), (c) left(L), (d) up(U) and (e) diffused(A) directions, respectively.

4.4   VGG16 CNN model

The VGG16 neural network, a well-known pre-trained model, was utilised for crack detection on concrete surfaces with ALICS. This model explained in Chapter 3, is trained and tested on the multi-angle dataset. The performance of the trained model is evaluated across different lighting angles ($10°$ to $60°$)to determine its effectiveness. This evaluation helps to identify the optimal lighting angle for effective concrete crack detection. Later, images were captured under auto-exposure settings shown in Table 4.4 from different directions, including right, down, left, up, and diffused, at this optimal angle. A comparison between directional and diffused lighting was conducted using performance metrics. Transfer learning and selective fine-tuning approaches were utilised to enable the model to capture important features from the images and help in more accurate detection [196, 197, 198, 199].

The binary classification of concrete images utilised a pre-trained VGG16 model, previously trained on a large ImageNet dataset [200] as starting point, instead of building the model from scratch. This approach allowed to leverage learned features from the VGG16 model, saving time and resources by avoiding long training on a specific dataset. All layers, except the last six layers of the base model, were frozen to fine-tune it as shown in Figure 4.19.

The Adam optimiser with binary cross-entropy loss and accuracy metrics was used [201]. The model underwent 10 epochs of training, utilising different combinations of batch size and learning rate using Grid search. After experimentation, a learning rate of 0.001 and a batch size of 32 were found to yield the best performance. Subsequently, the model was trained using these optimal values and evaluated on the testing data. To prevent overfitting, dropout was incorporated into the network architecture. Additionally, a dense layer with sigmoid activation was implemented for binary classification. The model performs similarly well on both the training and testing datasets, with no degradation or decline in performance on the testing dataset as training progresses. Performance metrics were computed to evaluate the performance of the model at various angles and directions.
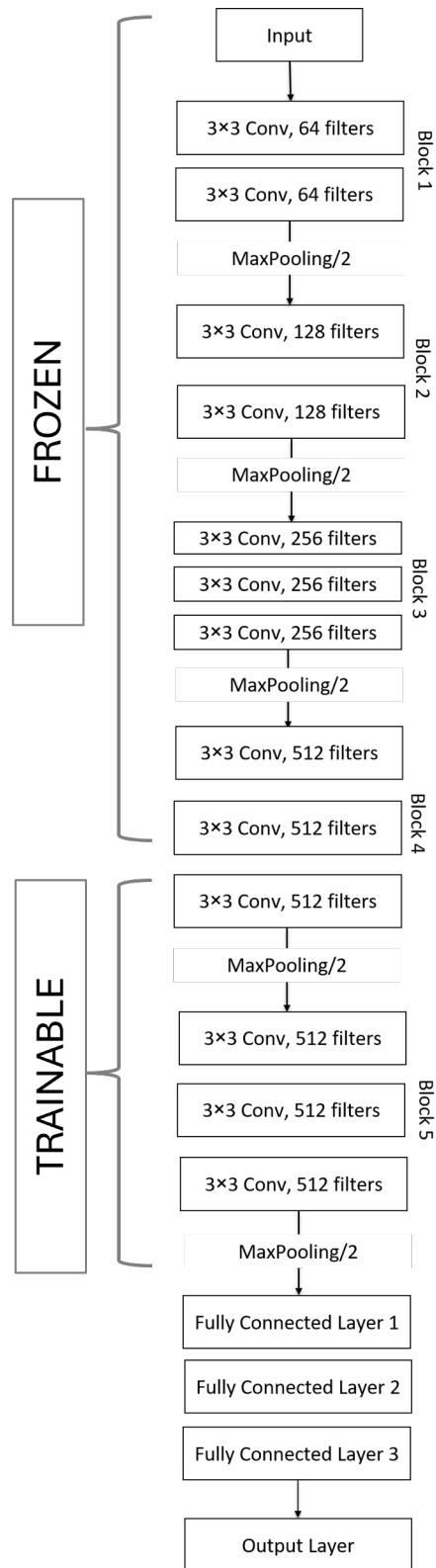
Figure 4.19: VGG16 crack classification model.

4.5 Performance Metrics

A train-test dataset split was employed to divide the entire dataset into two parts: one for training the model to learn features and the other for testing/evaluating the performance of the VGG16 model on multi-angled, multi-directional datasets [202]. The model was trained on the training dataset and was used to generate predictions on the testing dataset. The predicted probabilities were converted to binary class labels using a threshold of 0.5. The performance of the model was assessed using a confusion matrix generated by comparing the predicted labels to the actual labels of the testing dataset.

The confusion matrix provided information about correctly and incorrectly classified instances, represented by true positive (TP), true negative (TN), false positive (FP), and false negative (FN) values. Finally, these values range from 0 to 1, with 1 indicating good and 0 indicating bad, and are utilised to calculate various evaluation metrics as shown in Table 4.5 [203].

Table 4.5: Performance metrics of a binary classifier.

Name	Description	Equation
TPR (Recall)	The ratio of correctly predicted positive instances to the total number of actual positive instances.	$\dfrac{TP}{TP + FN}$
Accuracy	The measure of the model's ability to correctly predict the outcomes.	$\dfrac{TP + TN}{TP + FN + FP + TN}$
MCC	The measure of the correlation between predicted and actual labels, accounting for true or false, both positive and negative predictions.	$\dfrac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}}$
F1 Score	The weighted average of recall and precision.	$\dfrac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$

4.6 Results

4.6.1 Multi-angle lighting

The VGG16 neural network model was trained on 600 sub-images and tested on 120 sub-images. The datasets included images captured at lighting angles ranging from 10 to 60 degrees in 10-degree increments, all under conventional diffused lighting conditions. The Figure 4.20 illustrates the performance of the model across these different angles.
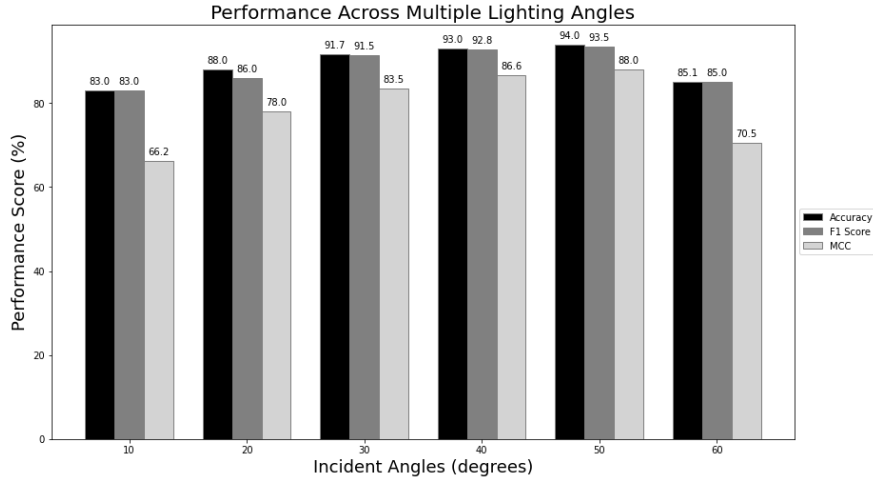
Figure 4.20: Performance Across Multiple Lighting Angles.

The testing accuracy of the model starts at 83% at a 10-degree angle and increases steadily, peaking at 94% at a 50-degree angle as shown in Figure 4.20. This indicates a significant improvement in the ability of the model to accurately predict the presence or absence of cracks as the lighting angle increases up to 50 degrees. However, at 60 degrees, there is a noticeable drop in accuracy to 85.1%. This improvement is also evident in the F1 score, which rises from 83% at 10 degrees to 93.5% at 50 degrees. At 60 degrees, the F1 score drops to 85%, similar to the trend observed in accuracy.

The Matthews correlation coefficient (MCC), a fundamental measure for assessing binary classifications with values between -1 and 1, consistently rises across different lighting angles, with maximum value of 88% at 50 degrees. This indicates that the model performs well across all four confusion matrix categories (TP, TN, FP, FN).

However, at 60 degrees, the MCC value decreases to 70.5%, indicating a reduction in overall classification performance.

In summary, the performance metrics indicated that 50 degrees is the ideal angle. At this particular angle, the model achieved top scores in accuracy, F1 score, and MCC, indicating its optimal overall performance in crack detection under conventional i.e., diffused lighting conditions.

4.6.2 Potential of Directional Lighting

The optimal angle determined based on above analysis is 50 degrees. At this angle, images were captured with illumination from various directions: R, D, L, U, and A, for multiple sets of samples. The primary aim of this analysis was to assess how lighting direction affects the performance of the neural network model. Therefore, the VGG16 neural network model underwent training and tested these directionally-lit images, covering 9 samples across five directions (9 samples × 5 directions = 45 images).

The performance of the VGG16 model in different lighting directions was investigated using two distinct sets. A detailed analysis of Sets 1 and 2 showed trends in the accuracy as shown in Figure 4.21. In Set 1, the model trained on images illuminated from the left direction achieved highest accuracy of 93.1%. Also, the model trained on images illuminated from right and down direction resulted in a higher true positive rate of 96.2%.

Set 2 achieved highest accuracy of 89.4%, when trained and tested on diffused images and the highest true positive rate of 89.7%, when trained and tested on the the images captured in right direction. This shows that in Set 2, diffused lighting is more favourable for overall accuracy, while identification of positive instances is best achieved with lighting from the right direction.

The difference in the lighting orientation for accuracy and TPR emphasises the possibility of optimising directional lighting. In particular, for Set 1, the model shows good performance when trained on images lit from the left direction, whereas for Set 2, optimal performance is achieved with diffused lighting. These results highlight the potential of directional lighting. The choice of the most suitable lighting direction
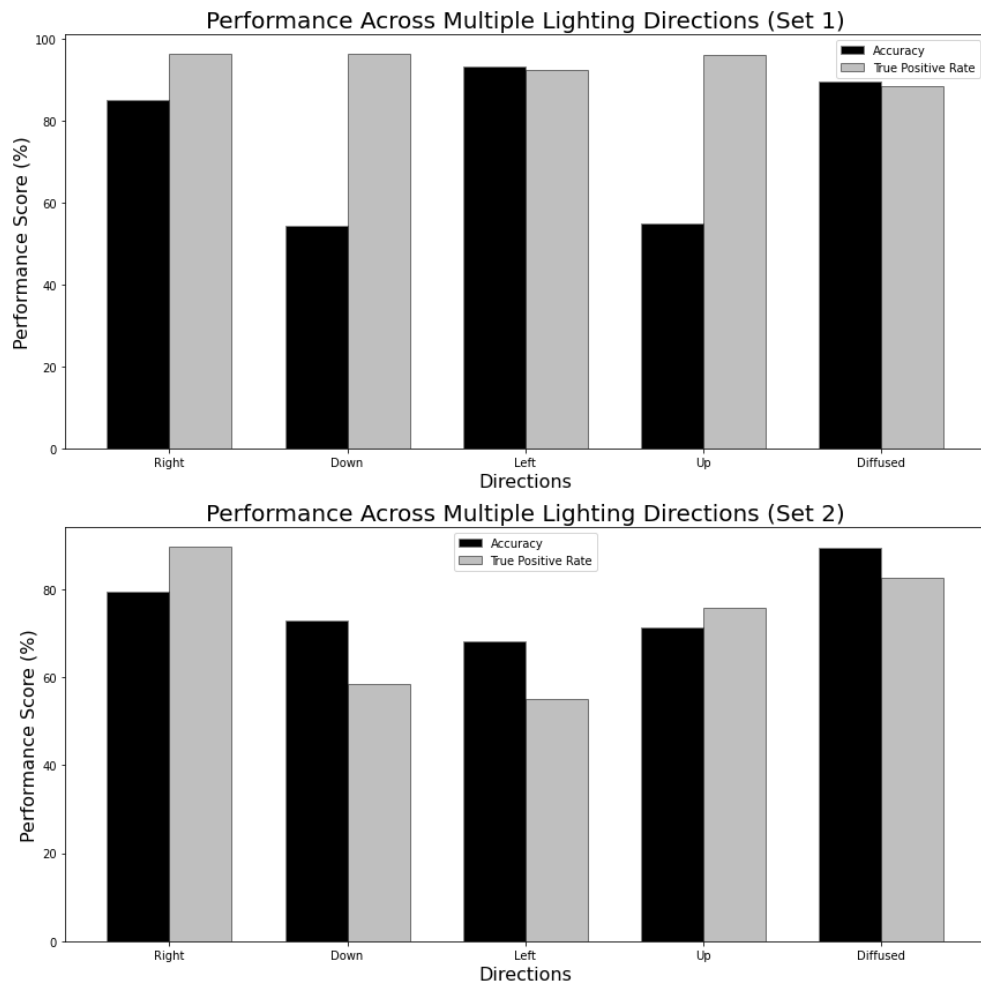
Figure 4.21: Performance Across Multiple Lighting Directions.

impacts the accuracy of the model and its TPR. In practical terms, optimising the lighting direction conditions can improve the robustness of crack detection models in real-world scenarios. The potential of directional lighting in comparison to diffused lighting is evident, yet its full utilisation requires modifications or implementation of new algorithms. Developing new algorithms or approaches could significantly enhance defect detection capabilities in Civil Engineering.

4.7   Summary

ALICS is a novel contribution to automated concrete defect inspection, addressing challenges faced by traditional visual inspections and prior automated techniques in Civil Engineering. It leverages the potential of geometrical illumination techniques, utilising multi-angle and multi-directional lighting approaches to improve the visibility of concrete defects. The ALICS device designed for laboratory use offers a robust framework for conducting testing in controlled environments. Meanwhile, the field-deployable rigs expand their capabilities to real-world applications, capturing real-world data in low-light or dark environments. This ensures that ALICS is not only effective in controlled settings but can also handle the challenges of practical concrete inspection in the field of Civil Engineering.

The images in laboratory and real-world are captured at different angles, ranging from 10 to 60 degrees, and also in diverse lighting conditions i.e., right, down, left, up, and diffused directions. The evaluation showed consistent improvement in accuracy, F1 Score, and MCC as the lighting angle increases, with the best performance observed at 50 degrees, highlighting its significance in achieving overall optimal performance under standard diffused lighting conditions. The multi-directional lighting experiments indicated the potential of directional lighting in enhancing crack detection accuracy compared to diffused lighting. Therefore, utilising these directionally lit images captured at an optimal angle of 50-degrees by ALICS could be further utilised to improve crack detection accuracy, which will be studied in detail in subsequent chapters.

Chapter 5

Threshold-Based BRISQUE-Assisted Deep Learning Model

Chapter 3 presented a summary of the effects that images captured under challenging lighting conditions have on the efficiency of deep learning models in the context of crack detection and identified the need of hardware to allow the capture of images with directional lighting. The hardware described in Chapter 4, is utilised to capture images of concrete samples in low-light environments using directional lighting. However, using robotics and cameras in low-light environments can introduce blur and noise during image capture. These factors can significantly affect the quality of the captured images and consequently affect the recognition capability of the model during the training and testing phases [10]. This drawback poses a major challenge to Civil infrastructure automatic inspections.

This chapter focuses on developing a method to improve the accuracy of deep learning algorithms for detecting cracks in concrete structures. More specifically it focuses on incorporating BRISQUE (see Section 3.4.2) as an image quality assessment during image capture to ensure that only high-quality images are retained for further image processing. The reasoning behind this was to create a closed loop feedback to ALICS

that will enable discarding low quality image and replacing it with a high quality image captured in real time during the image acquisition stage in the field.

5.1   Methodology

In Chapter 4 it was shown that capturing images at an optimal 50-degree lighting angle led to consistent improvements in accuracy, F1 score, and Matthews correlation coefficient when using standard i.e., diffused lighting conditions. Consequently, for analysis in this chapter, all images were captured at this optimal angle under diffused lighting settings in low-light environments.

The workflow of the entire process is clearly shown step-by-step in Figure 5.1. The Imaging Hardware used for data acquisition process is detailed in Section 5.2. The curation of the Pristine dataset, along with the introduction of Gaussian noise and Gaussian blur to generate Noisy and Blurred datasets, is also explained in this section. The sensitivity of BRISQUE to noise and blur is discussed in Section 5.4.1. The formation of testing matrices and the classification metrics used to evaluate them are described in Section 5.3 and Section 5.4.2. The determination of the BRISQUE score threshold is presented in Section 5.4.3. The implementation of an automated BRISQUE-based data cleaning method and its evaluation on images captured outside the lab are explained in Section 5.4.4 and Section 5.4.5, respectively.

Gaussian noise and Gaussian blur are the most common types of image degradations frequently encountered in real-world imaging scenarios [204, 205]. Therefore, these degradations were chosen to explore how they affect the BRISQUE scores. Images captured by cameras often exhibit noise and blur due to various factors such as sensor imperfections, environmental conditions, or motion during image acquisition. These imperfections can arise from factors like sensor limitations, fluctuations in lighting conditions, or unintended camera movements during image capture [7, 8]. As a result, quantifying the effects of Gaussian noise and Gaussian blur is crucial for understanding and improving performance of image processing algorithms in practical applications.

The analysis can be described in the following four steps:
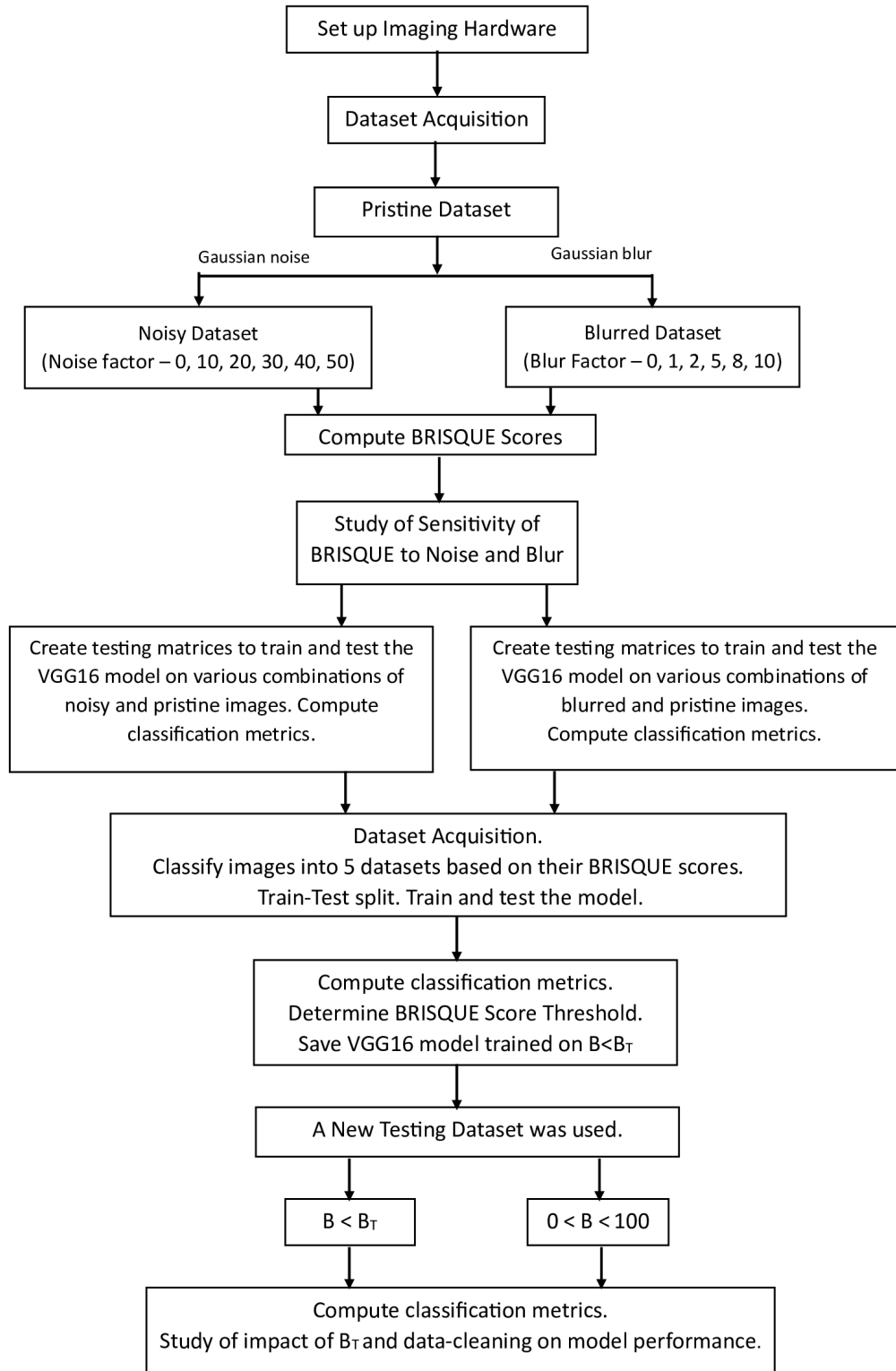
Figure 5.1: Workflow overview of Threshold-Based BRISQUE-Assisted Deep Learning for Enhancing Crack Detection in Concrete Structures.

1. The performance of the model is evaluated on both pristine images and their degraded images generated by artificially introducing Gaussian noise and Gaussian blur to the pristine images. This measures the impact of image degradation on crack classification metrics, including accuracy, F1 score, and MCC.

2. The BRISQUE scores of the degraded images are computed using equation 3.1 to study the sensitivity of BRISQUE to noise and blurred images.

3. The optimal BRISQUE score threshold, B_T is determined by comparing the performance of the VGG16 model across different datasets with varying BRISQUE score threshold values.

4. The impact of B_T and data-cleaning on the performance of the model is studied by evaluating the trained Threshold-based BRISQUE-assisted VGG16 model on an independent real-world testing dataset.

5.2 Dataset Acquisition and Pre-processing

For this study, images from a large dataset comprising 45 high-resolution images of concrete surfaces with visible cracks was employed, as detailed in section 4.3. Out of the 45 images available, only 9 images were captured at 50 degrees, i.e., the angle that allows for the model to achieve highest accuracy (See Figure 4.20) under diffused lighting settings. These 9 images captured in real world, along with one more concrete image captured in laboratory was utilised. The images have either smooth areas or rough areas in the background with cracks (hairline cracks or thick cracks or multiple crack networks).

For example, the image captured in the laboratory as shown in Figure 5.2(a) has a smooth surface but has multiple cracks on it. Of the remaining 9 real-world captured images: four images display irregular cracks (two horizontal, one diagonal, and one vertical) on a rough, greyish background; two images show multiple cracks and one image features a single thin crack running horizontally through the centre on a dark complex background (that has dark stains); and the remaining two images exhibit thin

cracks running through the centre on a smooth concrete surface. Each of these 10 images are further divided into 360 sub-images. This dataset covers different crack types and was collected from various locations and environments, both indoors and outdoors, to ensure that the damages reflect those typically found in Civil infrastructures. After the pre-processing step, some sub-images have thin cracks or hairline cracks, some of them have thick cracks, and some of them have no cracks, and some of them have stains that resemble cracks. In Chapter 7, following the manual annotation of cracks for the binary classification task, an additional preprocessing step was implemented to prepare the data for multi-class image classification to differentiate thin cracks from thick cracks. The Otsu thresholding algorithm was employed to identify thick cracks [206]. In cases where the algorithm failed to form a bimodal histogram then those images were classified as thin cracks [207]. The width of these cracks were found to be within the range of 0.1mm to 0.5mm using crack width ruler.

Overall, a dataset of 3600 sub-images, each corresponding to a size of 224×224 pixels, was curated. Of this dataset, 2100 images displayed cracks, and 1500 images displayed clear concrete surfaces, which were manually labelled by the author. The "baseline" images were degraded by introducing common degradations such as Gaussian noise and Gaussian blur. For each of these pixels, BRISQUE scores were calculated. Subsequently, based on these BRISQUE scores, corresponding colour maps for all the pixels were generated. The colour maps represent the quality of the images, i.e., lighter pixels indicate better image quality, and darker pixels implies low image quality as illustrated in Figure 5.2(b).

Gaussian noise was applied to the original images by increasing the magnitude of the noise level from 10% to 50%; an example of this is shown in Figure 5.3 [208]. In equation (5.1), the coordinates (x,y) represent the pixel coordinates, s(x,y) is the original image, n(x,y) is the added Gaussian noise, and w(x,y) is the resulting noisy image [210].

$$w(x, y) = s(x, y) + n(x, y) \qquad (5.1)$$

(a)   (b)
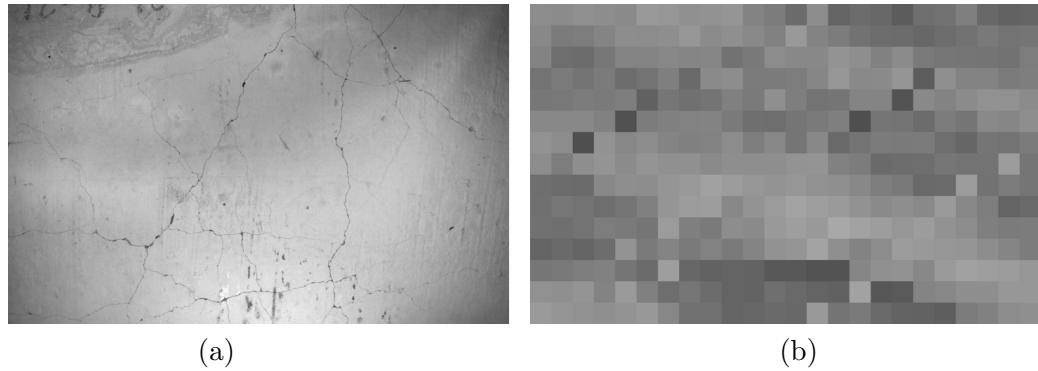
Figure 5.2: a) Baseline (Pristine) image b) colour map of Pristine Image (lighter the pixels, the better the image quality).

(a)   (b)
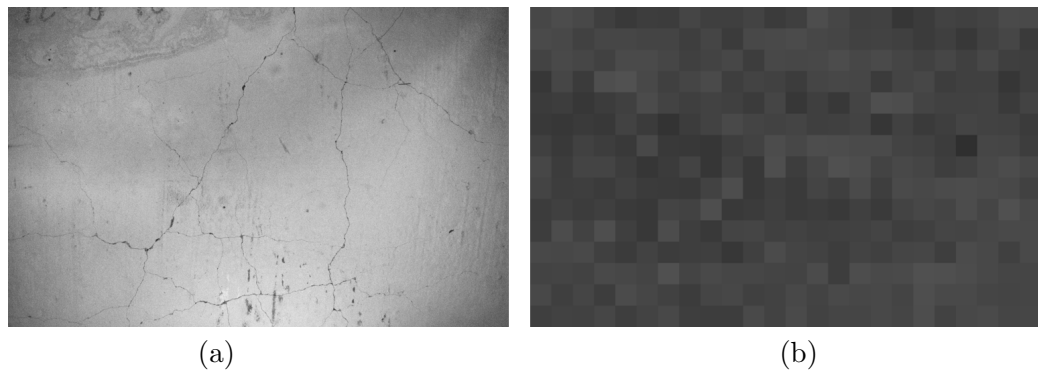
Figure 5.3: a) Noisy image of noise factor 50 b) colour map of Noisy image (lighter the pixels, the better the image quality) [208].

(a)   (b)
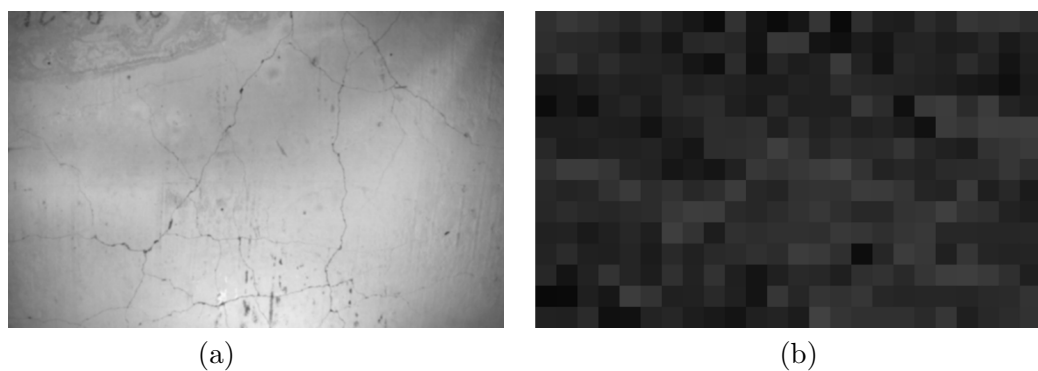
Figure 5.4: a) Blurred image of blur factor 8 b) colour map of Blurred image (lighter the pixels, the better the image quality) [209].

Gaussian blur was applied to the images by adjusting the blur factor from 1% to 10%, aiming to replicate various levels of image blur [209] similar to real-world scenarios such as motion, defocus, or optical irregularities. This range of blur factors helps to explore how Gaussian blur affects images, from slight blurring to more identifiable effects, represented by the equation (5.2),

$$I_{\text{blurred}} = I \otimes G(\sigma^2) \tag{5.2}$$

where I is the original image without the blur, $I_{blurred}$ is the blurred image that contains Gaussian blur, G is a Gaussian kernel with standard deviation σ and $\otimes$ denotes the convolution operation [211].

The Gaussian blur operation was implemented through convolution with a Gaussian kernel. This involves convolving the image with a Gaussian kernel with a specific standard deviation (σ) to create natural blurring found in real-world or practical settings. By systematically varying the blur factor, the study aimed to examine the impact of blur on image quality. Figure 5.4 shows blurred image and its corresponding colour map.

Noise levels up to 50% were chosen because beyond this threshold, cracks in the image become barely visible, making manual labelling challenging. Similarly, the choice of blur levels up to 10% was made because further increase in blur reduces the clarity of the image, making its features nearly invisible to the naked eye. At this stage, changing the noise and blur levels to a point where features are still visible to the naked eye is important for ground-truthing. Further, noise levels up to 50% and blur levels up to 10% were chosen as they reflect real-world imaging conditions. Beyond these limits, such extreme distortions are unlikely to occur naturally, making higher levels impractical for consideration [212, 213, 214].

5.3 Evaluation of the VGG16 model

The VGG16 model was used for this work. The reasoning behind this was because it is one of the most commonly used models for crack detection as described in Chapter

3. The performance evaluation of the VGG16 model, described in Figure 4.19, was conducted on baseline (pristine), noisy, and blurred datasets using evaluation metrics, outlined in Table 4.5. The positive predictive value (PPV), also called as Precision is the ratio of correctly predicted positive instances (TP) to the total number of positive predictions (TP+FP), as described in equation 5.3 was also computed [203].

$$PPV(Precision) = \frac{TP}{TP + FP} \tag{5.3}$$

A testing matrix was created to evaluate the performance of the VGG16 model under different conditions. This matrix outlines the different combinations of training and testing scenarios used during the evaluation as shown in Table 5.1. The rows indicate the testing conditions: noisy, blurred, and pristine images, while the columns represent the training conditions: noisy, blurred, and pristine images. The "x" marks in the matrix indicate specific training and testing combinations. For example, "Train on Noisy, Test on Noisy" means that the model was trained on noisy images and then tested on noisy images. Similarly, "Train on Blur, Test on Pristine" means that the model was trained on blur images and then tested on pristine images.

Table 5.1: Testing Matrix.

	Train on Noisy	Train on Blur	Train on Pristine
Test on Noisy	x		x
Test on Blur		x	x
Test on Pristine			x

5.4   Results

Figure 5.2(a) and Figure 5.2(b) represents baseline (pristine/original) image and its corresponding colour map of BRISQUE scores, which is plotted after data pre-processing as explained in subsection 5.2. Various types of degraded images with different levels

of noise and blur were examined to evaluate their impact on BRISQUE scores within the context of concrete crack detection for Civil infrastructure.

Figure 5.3 and Figure 5.4 display the images with noise level 50 and blur level 8, respectively, along with their corresponding colour maps of BRSIQUE scores. The colour maps clearly show that regions with greater noise or blur have darker pixels, indicating higher BRISQUE scores i.e., lower image quality.

5.4.1 The Sensitivity of BRISQUE to Noise and Blur
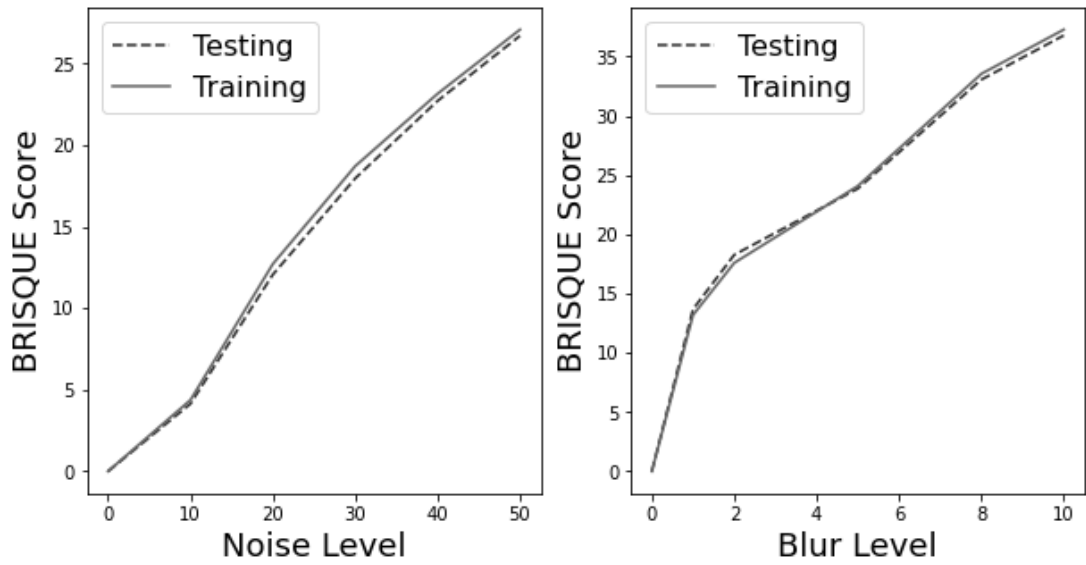
Figure 5.5: Comparison of BRISQUE scores for various levels of noise and blur in image processing.

Figure 5.5 displays the BRISQUE scores of training and testing datasets with different levels of noise (0% to 50% with an increment step of 10%) and blur (0%, 1%, 2%, 5%, 8%, and 10%). Both subplots feature a solid line representing the training data and a dashed line representing the testing data. As seen from the plot, the BRISQUE score increases as the levels of noise and blur increase in the datasets, indicating a decline in image quality.

The colour maps and their corresponding BRSIQUE scores confirms the expected impact of noise and blur on image quality i.e., the baseline image has the lowest score.

As noise and blur levels increase, the BRSIQUE score increases, indicating a decline in image quality. The figure illustrates that while noise levels increased by 10%, the BRISQUE scores displayed a nearly linear relationship. In contrast, a mere 1% increase in blur levels led to a more sensitive and gradually increasing trend in BRISQUE scores. This difference indicates that blur has a more prominent effect on image quality compared to noise levels.

For example, as noise levels increased from 10% to 50%, BRISQUE scores increased by a total of 24, highlighting the sensitivity of BRISQUE to changes in noise. In contrast, the sensitivity to blur levels exhibited a rise of 24 in BRISQUE scores as blur levels increased from 1% to 10%, highlighting the greater impact of blur on BRISQUE scores compared to noise levels. Therefore, it is essential to assess the quality of the images and their impact on the model recognition capabilities. This can be achieved by correlating BRISQUE scores with performance of the model, and develop a thorough methodology for optimising image quality.

5.4.2 Classification metrics on Distorted Images

The VGG16 model underwent training and testing on both original (pristine) datasets and datasets with varying noise or blur levels. The performance metrics are shown in Figure 5.6 and Figure 5.7. For some, F1 and MCC scores were represented as zero in Figure 5.7, suggesting no true positives were identified in the predictions. This could be due to the inability of the model to correctly classify any images in the testing set, indicating overfitting and poor performance on that dataset.

The X-axis in the graphs represents various datasets that were used to train and test the VGG16 model. The Y-axis in the graphs represents the performance of each model, as measured by three metrics - accuracy, F1 score, and MCC. The bar graphs illustrate two scenarios. In the former scenario, the model was trained on a clean dataset and tested on noisy datasets ranging from 0% to 50% noise, where 0% represents pristine images. Conversely, in the latter scenario, the model was trained and tested on datasets with increasing noise levels. Similarly, for blur evaluation, the model was either trained on pristine images and tested on blurred datasets ranging from 0% to 10% blur,
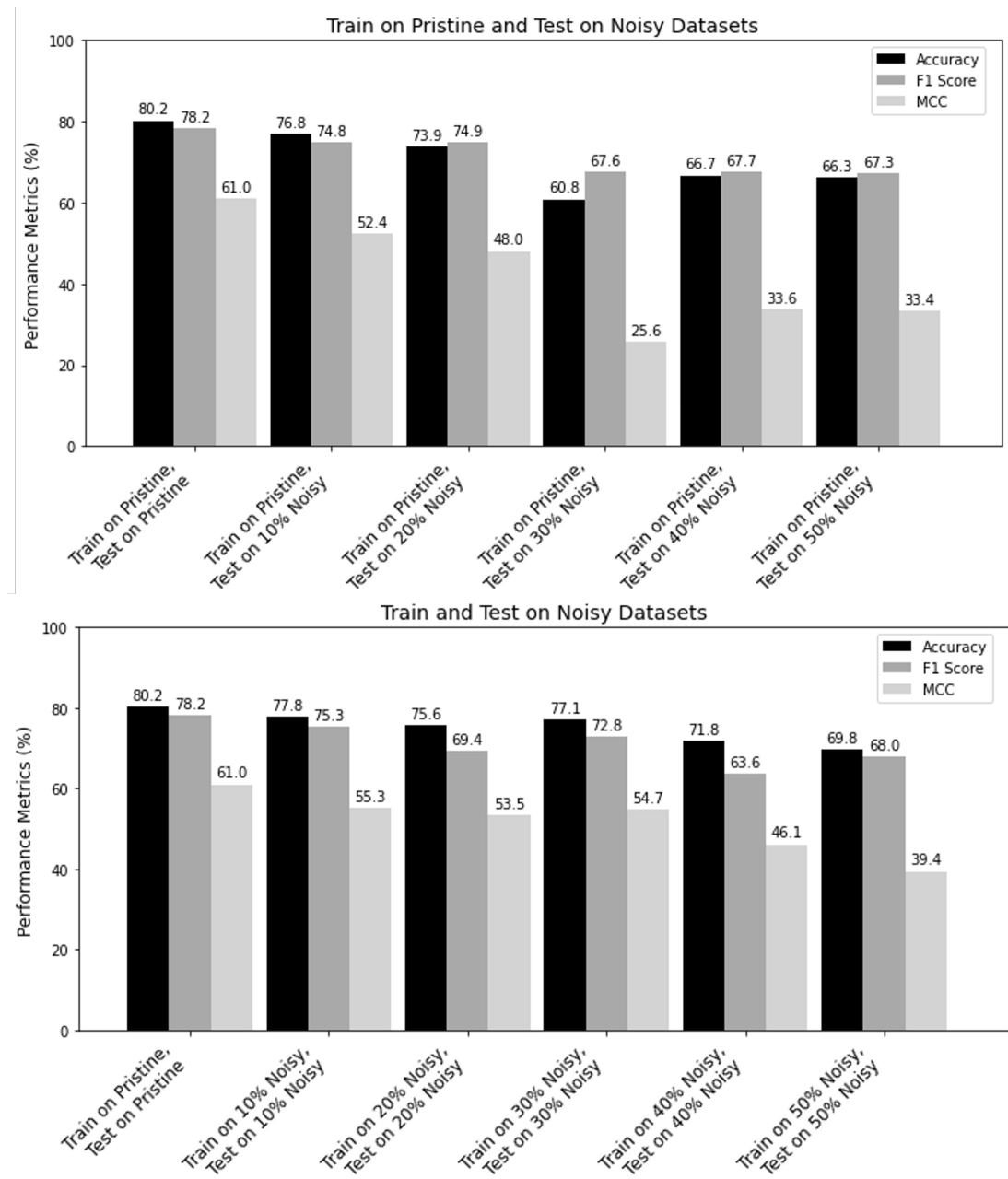
Figure 5.6: Performance comparison of VGG16 neural network model on Noisy images using accuracy, F1 score and Matthew's correlation coefficient metrics.
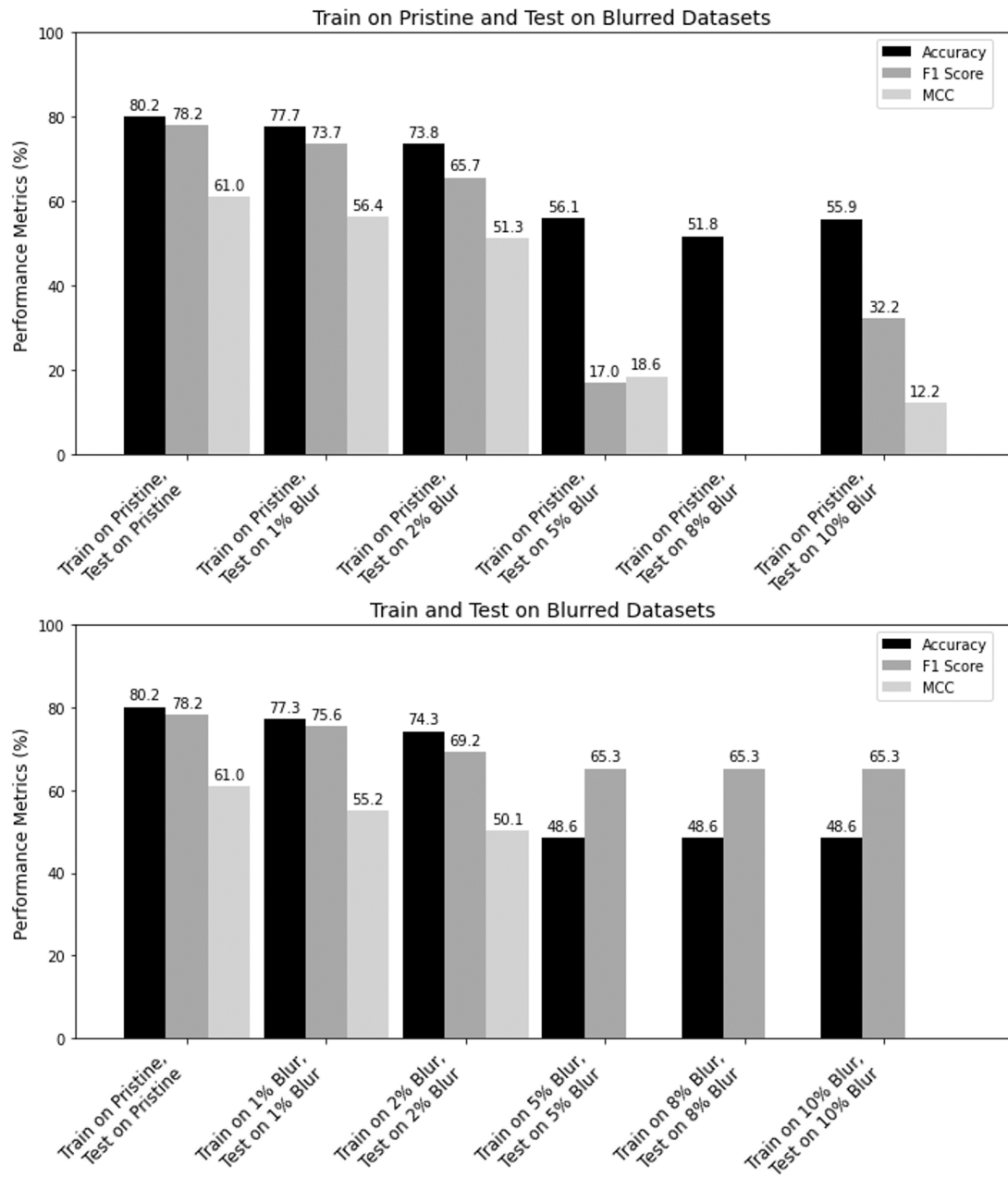
Figure 5.7: Performance comparison of VGG16 neural network model on Blurred images using accuracy, F1 score and Matthew's correlation coefficient metrics.

or trained and tested on datasets with increasing blur levels, where 0% blur denotes pristine images. The black bars represent accuracy, the dark grey bars represent the F1 score, and the light grey bars represent MCC.

In both cases the performance of the model declined with increasing levels of noise and blur, consistently showing that the model trained and tested on pristine images performed the best. For example, when tested on 0% noisy data, the model trained on pristine images achieved an accuracy of 80.2%, whereas the accuracy dropped to 66.3% when tested on 50% noisy data. Similarly, when tested on 0% blur data, the model trained on pristine images achieved an accuracy of 80.2%, while the accuracy decreased to 48.6% when tested on 10% blur data.

The model can classify degraded datasets more effectively when it is trained on degraded images instead of pristine ones. This is expected, the network was trained on degraded images, allowing it to recognise similar degradation during testing, which helps it better identify cracks in such images. On the other hand, a network trained on degraded images could result in reduced performance when tested on pristine images [10] and as such, it could not generalise. An approach that would allow this would be to ensure that the quality of the images used by the model are of high quality, through an image quality assessment workflow.

To determine whether the differences in model performance results were statistically significant, a one-way analysis of variance (ANOVA) was performed [215]. This analysis compares the means of three or more groups to see if there are significant differences among them, based on the obtained p-values. A low p-value, generally below 0.05, suggests strong evidence against the null hypothesis, implying that the observed differences are unlikely to be random and supporting the alternative hypothesis [88]. In our analysis, noisy images produced a p-value of $2.92\mathrm{e}{-}09$, indicating statistically significant results and leading to the rejection of the null hypothesis. Additionally, the ANOVA for blurred images resulted in a p-value of 0.0028, also suggesting support for the alternative hypothesis.

The decline in the performance of the model, ranging from 10.4%-24.3% with increasing levels of noise and blur highlights the critical role of careful data pre-processing

i.e., ensuring clean, relevant, and high-quality data before training the model can reduce the effects of noise and blur in deep learning applications. The next step was to automate this pre-processing stage to ensure that only high-quality images are utilised for model training. This was achieved by establishing a threshold for BRISQUE scores to assess the quality of images considered suitable for training.

5.4.3   Optimising BRISQUE score Threshold for crack detection

This work utilised a dataset comprising 5665 sub-images, containing both the pristine and degraded images. BRISQUE scores of all these were computed and organised into five datasets based on scores ranging between <45 and <85, as outlined in Table 5.2. Subsequently, each dataset was split into train and test sets (70% train and 30% test), manually labelled as positive (cracks) or negative (no cracks), to train and test the VGG16 model.

The analysis revealed that the optimal performance of the model was achieved when the BRISQUE score threshold was set to <45, despite the limited availability of training images as shown in Table 5.2. The performance of the model gradually decreased as the BRISQUE score threshold increased from <55 to <85. This decrease was not consistently linear due to change in the number of training images used. The assumption that performance of the model correlates positively with the number of training images is common, yet the specific amount of training data required for deep learning to achieve satisfactory performance remains uncertain [216]. Notably, increasing the training set size from 720 to 2160 images has improved model performance for BRISQUE score thresholds <55 and <65. However, the performance of the model decreased when the dataset included images with BRISQUE scores <75 and <85. Interestingly, despite having fewer training images (less than nearly 10 times), the model achieved its best performance when the BRISQUE score threshold was set to <45.

The distribution of BRISQUE scores is visually represented in Figure 5.8, which shows distinct score intervals created using histogram binning. It was observed when the BRISQUE score threshold was set to <85 the model achieved accuracy of 0.61, F1 score of 0.69, and MCC of 0.30. Hence, images with high levels of degradation sig-

Table 5.2: Comparison of model performance at different BRISQUE Score Threshold values.

BRISQUE Scores	Number of training images	Accuracy	F1 score	MCC
<45	360	0.8	0.77	0.61
<55	720	0.66	0.73	0.40
<65	2160	0.73	0.76	0.49
<75	3240	0.73	0.66	0.48
<85	3960	0.61	0.69	0.30
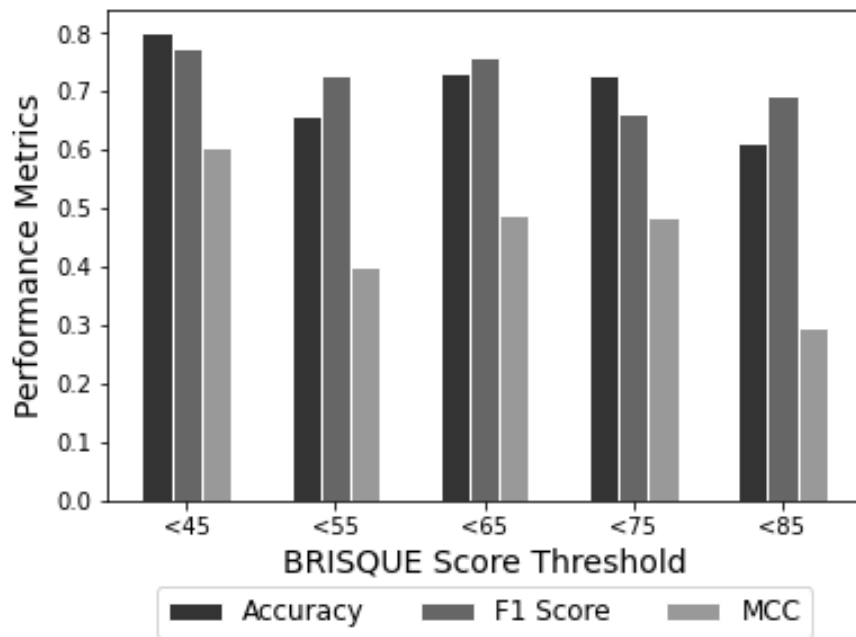
Figure 5.8: Performance comparison of VGG16 neural network model on images with different BRISQUE score thresholds using accuracy, F1 score and Matthew's correlation coefficient metrics.

nificantly degrade the performance of VGG16 model and should be discarded. These findings reveal a previously unexplored relationship between BRISQUE scores, image quality, and performance of neural network models for detecting cracks in low-light environments.
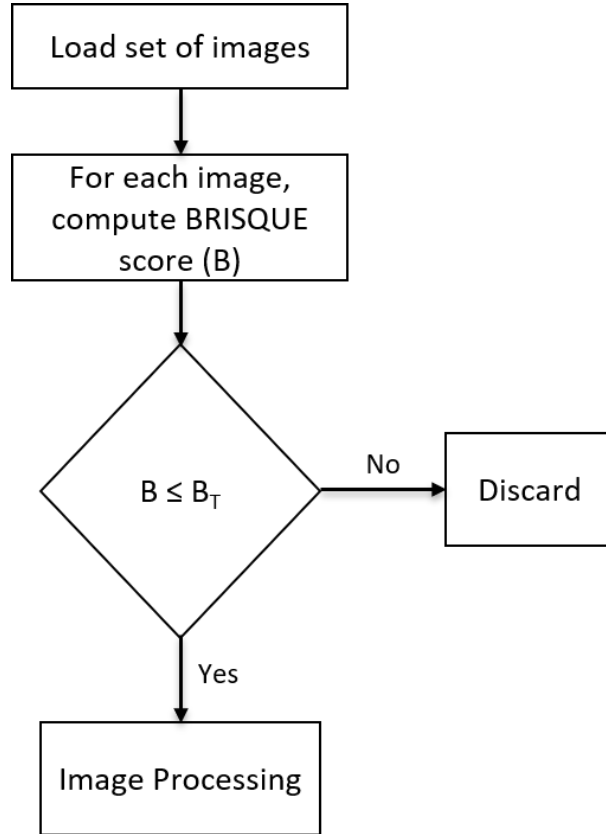
Figure 5.9: BRISQUE-based Data Cleaning, where B and B_T represents BRISQUE score, and BRISQUE score threshold, respectively.

5.4.4   Automated BRISQUE-based Data Cleaning

Figure 5.9 illustrates the process of automated BRISQUE-based data cleaning of images developed as part of this work. The algorithm initially loads images, computes the BRISQUE scores (B) for each image, and compares them to a pre-defined threshold value. In this study, the BRISQUE score threshold (B_T) is set at a constant value of 45 to address the requirements of detecting concrete cracks in low-light environments.
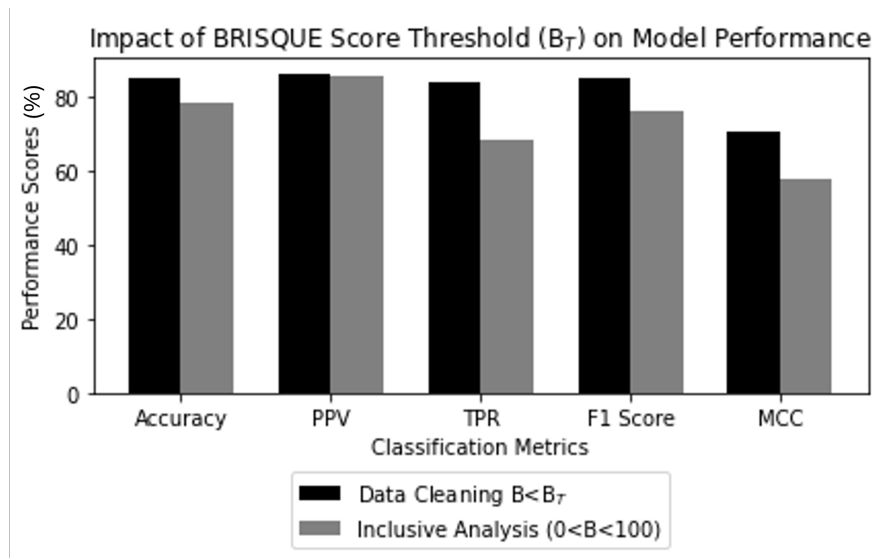
Figure 5.10: Impact of BRISQUE Score Threshold ($B_T <= 45$) on Model Performance - where data cleaning indicates that the trained model is tested on images within BRISQUE Score Threshold i.e., $B<B_T$, and Inclusive Analysis indicate that the testing dataset includes all images, both within and exceeding the BRISQUE score threshold i.e., $0<B<100$.

If the BRISQUE score of the image exceeds the set threshold, the image is discarded; otherwise, it is retained for further image processing. Thus, this allows for the automatic rejection of all low-quality images before VGG16 neural network model is used for crack detection. Based on prior analysis and findings, the author set a BRISQUE score threshold of 45. Therefore, when images are captured using the ALICS device under low-light conditions, images under this BRISQUE score are considered to be of good quality.

Since ALICS is designed for on-site image capture and crack detection, its ability to filter out low-quality images directly improves detection accuracy. Future work will involve the ALICS device capturing a new image if it finds any image with a BRISQUE score above the threshold value, ensuring that only high-quality images are captured. By ensuring that only high-quality images are processed, the algorithm improves the performance without requiring additional post-processing steps. This approach makes ALICS reliable for real-time crack detection applications.

5.4.5   Evaluation of VGG16 using Threshold-based BRISQUE IQA algorithm on images captured outside the lab

To further evaluate the performance of the model, an additional testing was conducted using field data. Specifically, 11 new images captured under diffused lighting settings in the low-light environments at Bowling Harbour area in Glasgow at 50 degrees were utilised. These images had a resolution of 5429×3458 pixels. Each image was subdivided into 360 blocks of size 224×224 pixels, resulting in a total of 3960 images, including both cracked and uncracked samples, which were manually labelled.

The model trained on images with a BRISQUE score <45 was used for testing, as this threshold resulted in superior model performance when images are captured under low-light condition using ALICS, as mentioned in the previous section. The model was tested under two distinct scenarios, one involving data cleaning within a specified BRISQUE score threshold <45 and the other omitting threshold-based data cleaning process to include all images, both within and exceeding the BRISQUE score threshold values i.e., 0<B<100 (includes all BRISQUE scores).

This inclusive dataset covered a broader range of image quality levels, leading to a noticeable impact on the performance metrics of the model. It is worth noting that the performance metrics showed significant improvements when data cleaning was conducted within the BRISQUE score threshold range, as compared to the one without data-cleaning as shown in Figure 5.10. The improvements in classification metrics include:

- 6.9% increase in accuracy, indicates that the model classified concrete cracks and non-cracks more effectively, reducing misclassifications.

- 15.9% increase in TPR indicates more effective crack detection rate.

- 0.6% improvement in PPV, which is relatively modest. However, it signifies a slight reduction in false positives predictions, which further reduces unnecessary inspections.

- 9.1% improvement in the F1 score, signifies an overall better performance of the

model.

- 21.5% increase in MCC, reflects the ability of the model to handle imbalanced datasets.

5.5   Summary

This chapter explored the correlation between the BRISQUE-IQA method and the VGG16 crack classification model. The VGG16 model performed best when trained and tested on clean datasets than degraded datasets containing Gaussian noise and blur, as expected. The evaluation using BRISQUE scores showed that when the model was trained on pristine data, as the noise or blur in the image increased, the BRISQUE score also increased, indicating a decrease in image quality.

Incorporation of transfer learning and fine-tuning to the VGG architecture helped in improving crack detection accuracy, thereby making a valuable contribution to the field of structural health monitoring and image quality assessment. The model achieved effective performance with B_T values below 45, even when trained on 90% less data. This study can also be applied to diverse fields, including medical imaging and face recognition, where high-quality images are critical for accurate diagnoses.

Chapter 6

Comparative Analysis of Multi-channel VGG-n Models implementing Directional Lighting

Chapter 6 explores the advantages of directional lighting by implementing a novel five-channel neural network model. It is customised based on the traditional VGG architecture, which is initially trained on the ImageNet dataset utilising three-channel colour (Red, Green, Blue) images. The objective is to adapt and effectively fine-tune the pre-trained model to accept five-channel images and achieve accurate classification results.

The following sections explore the concepts and methodologies underlying the significance of this method in automated inspection systems.

6.1 Introduction

Numerous studies have demonstrated that brightness and lighting direction significantly influence crack detection accuracy, yet they often fail to fully utilise the advantages of

directional lighting [9] as explained in 3.5.4. [96, 99] employed various deep learning algorithms such as YOLO, Faster RCNN to investigate the effectiveness of CNN techniques under challenging lighting conditions. However, all algorithms struggled to accurately identify cracks due to poor lighting. Poor lighting reduces the contrast between the cracks and the background, making it difficult for feature extraction as explained in Chapter 3.

Directional lighting requires higher number of images than the standard diffused lighting which uses 3-channel (RGB) images. Based on the reviewed literature, there is currently no multi-channel deep learning model with detailed evaluation of directional lighting concrete inspection technology. The development of a novel five-channel deep learning algorithm is the focus of this chapter.

The VGG deep convolutional neural network architecture, commonly used for crack detection for Civil infrastructure, was employed in this study [123, 217]. The five-channel neural network model is implemented based on the VGG-16 and VGG-19 architectures, referred to here as VGG-n, where n = 16 for VGG-16 and 19 for VGG-19.

Conventional CNNs typically use three-channel (RGB i.e., colour images) inputs. However, the novel five-channel model incorporates images captured from five different lighting directions: Right (R), Down (D), Left (L), Up (U), and Diffused (A) captured using the hardware setup described in Chapter 4.

The main objectives of this chapter are as follows:

1. Investigate the potential of directional lighting in binary image classification tasks.

2. Implement a novel five-channel DL model that utilises five-channel dataset, where each channel represents the gray scale version of the image captured in R, D, L, U, and A directions, respectively.

3. The traditional and five-channel neural network models implemented in this chapter employ hyperparameter tuning, regularization techniques, and a stratified cross-validation approach to avoid overfitting, ensuring that the models can generalise well to new, unseen data.

4. Evaluate the performance of the five-channel VGG-n neural network models

118

and compare it with their conventional (traditional) three-channel models that utilise
diffused images alone.

6.2 Methodology

ALICS, described in Chapter 4, is used to capture images in low-light conditions using
a directional lighting approach to extract crack information effectively. The following
subsections provide a detailed description of each step involved in implementing the
VGG-n multi-channel neural network models in detecting cracks from images captured
by ALICS.

6.2.1 Dataset Description and Pre-processing

ALICS captures images of size 5429×3458 pixels in R, D, L, U, and A directions
respectively, as shown in Figure 4.18 and are further cropped into smaller blocks of
size 224×224 pixels to match the input size requirements of traditional VGG-n models.
These images are combined to create a five-channel tag image file format (TIFF) image,
where each channel represents a specific lighting direction.

TIFF is a widely used file format for storing high-quality images, making them
suitable for multi-channel image analysis applications [218]. Transfer learning and
fine-tuning approaches are utilised with the multi-channel VGG-n architectures to ac-
curately classify cracks. For dataset preparation, a well-balanced dataset is created
for training and testing the neural network models, The performance of the model is
evaluated for binary image classification tasks as explained in Section 6.3.2.

6.2.2 Stratified five-fold cross validation, Hyperparameter tuning, and Regularization

In this work, a stratified five-fold cross-validation technique is implemented, where one
subset was selected as the test set in every iteration, while the remaining four were used
for training the model as shown in Figure 3.5. Performance metrics such as accuracy,
precision, F1-score, recall, and MCC are calculated on the test fold for each iteration.

After all the 'k' iterations are complete, the average performance across all iterations is calculated using equations described in Section 6.3.2. This provides an estimate of how well the model will perform on new, unseen data. Repeating the process 'k' times with different training and test fold combinations reduces the performance estimate variance. The study also employs Hyperparameter tuning, Regularization, and early stopping techniques as explained in detail in Chapter 3.

6.2.3 Generation of five channel image

When using the conventional VGG-n architectures, an input image of size $224{\times}224$ pixels, with three colour channels (red, green, and blue) is required. However, the proposed deep learning model utilises a five-channel TIFF input image. To generate a five-channel TIFF image, the initial step involves converting the original three-channel images captured under different lighting directions (R, D, L, U, A), into gray scale images $(224, 224, 1)$ i.e., images of size $224{\times}224$ pixels with single gray channel. These gray scale images are then stacked together along the third dimension to form a single five-channel image $(224, 224, 5)$ i.e., each gray scale image becomes a separate channel in the new multi-channel image. Therefore, each gray scale image corresponds to one channel, allowing the combined image to capture information from all the five lighting directions. Finally, the resulting multi-channel image is saved in TIFF format, denoted by the .tiff extension, as illustrated in Figure 6.1. Since the five-channel generated images are of very high quality and contain image information from all five images, they are stored in TIFF format. These five-channel images constitute the five-channel dataset used for training and testing the novel five-channel VGG-n model, which utilises an input image of dimensions $(224, 224, 5)$.

6.2.4 Implementation of VGG-n multi-channel neural network models

The conventional three-channel VGG-n model accepts an input image of size $(224, 224, 3)$, where 3 represents the RGB colour channels. The five-channel VGG-n model requires an input image of size $(224, 224, 5)$, where 5 channels represents the gray scaled version of the image captured in R, D, L, U, and A, directions respectively. To implement the

Right	Down	Left	Up	Diffused	
Grayscale version	Grayscale version	Grayscale version	Grayscale version	Grayscale version	(224, 224, 3)

5 Grayscale images merged to form five channel TIFF image

(224, 224, 1)

(224, 224, 5)

Output: 5-channel TIFF image
(Grayscale)
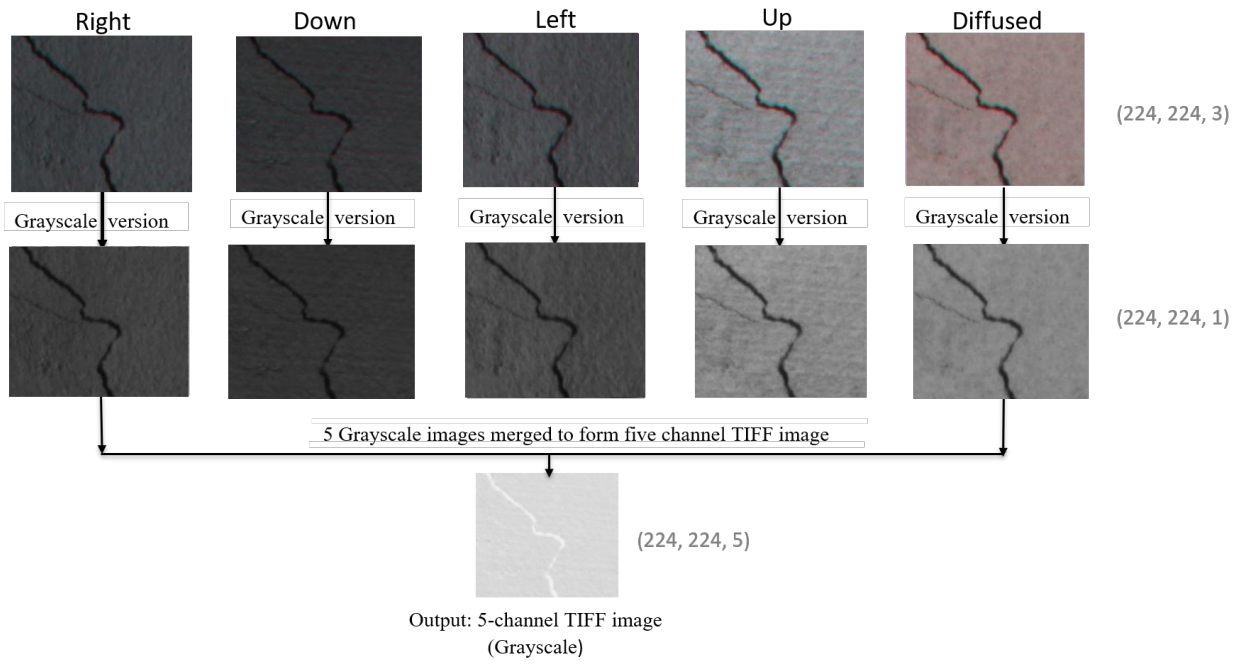
Figure 6.1: Generation of five-channel TIFF image. The three-channel (RGB) colour images $(224, 224, 3)$ are captured in the right, down, left, up, and diffused directions (top row). These five images are further converted to one-channel gray scale images $(224, 224, 1)$ (second row). These images are combined together to form a single five-channel TIFF image of size $(224, 224, 5)$ (bottom row).

five-channel model from the traditional three channel model, the number of channels should be adjusted from 3 to 5. When transitioning from a three-channel input image to a five-channel input image, only the trainable parameters in the first convolutional layer are affected, while the other convolutional layers remain unchanged as shown in Figure 6.2 and Figure 6.3.

Figure 6.2: Summary of VGG-16 three-channel and five-channel models.

The change in the number of trainable parameters from 1792 to 2944 in the first convolutional layer of the proposed model is detailed in Table 6.1. For conventional models, there are 64 filters of size 3×3 kernel, resulting in a total of 576 weights. With three channels, the total number of weights become 1728. Additionally, there

Figure 6.3: Summary of VGG-19 three-channel and five-channel models.

are 64 biases, bringing the total trainable parameters to 1792. In the case of the five-channel models, the values for the number of filters, filter size, and number of weights per filter remain the same. However, with five channels, the total number of weights become 2880. Similar to three-channel, there are 64 biases, resulting in a total of 2944 trainable parameters. This highlights the impact of the number of channels on the total trainable parameters in the first convolutional layer of the VGG-16 and VGG-19 models.

Table 6.1: Parameter Comparison: 3-channel vs 5-channel VGG-16/VGG-19 models in the first convolutional layer.

Parameter	Values for 3-channel	Values for 5-channel
Number of filters	64	64
Filter size	3×3	3×3
Number of weights per filter	9	9
Total weights for all filters	64×9 = 576	64×9 = 576
Number of channels	3	5
Total weights for all channels	576×3 = 1728	576×5 = 2880
Number of biases	64	64
Total trainable parameters	1728 + 64 = 1792	1728 + 64 = 2944

This approach takes the advantage of the pre-trained VGG-n models by transferring their weights to the corresponding layers of the five-channel VGG-n models, except for the first convolutional layer. The weights of the first three channels of the first convolutional layer in are copied as it is, while those of the additional two channels are calculated by averaging across the existing channels, as illustrated in Figure 6.4 and Figure 6.5. This step allows to extend the capability of the proposed models to handle five-channel input, while still benefiting from the pre-existing weights for the initial three channels [75].

To optimise the performance of the proposed model in classifying the five-channel dataset as either crack or no crack, transfer learning and fine-tuning techniques are employed. This involves utilising the pre-trained weights of the VGG-16 and VGG-19 models and fine-tuning them to adapt to the specific features of the new five-channel dataset implemented for this research, enabling the model to efficiently learn the new relevant features. The step-by-step implementation of the customised five-channel VGG-16 and
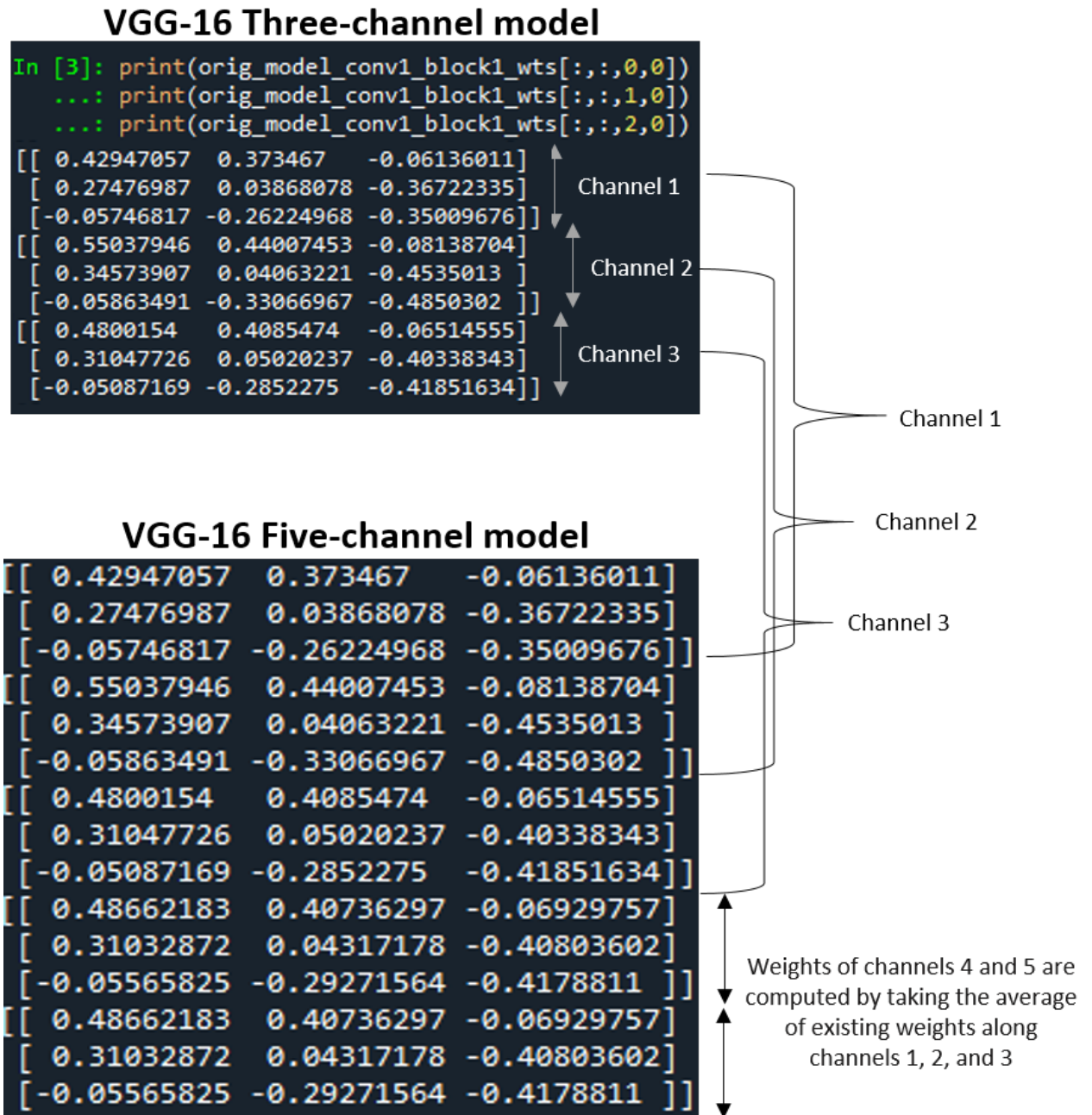
Figure 6.4: Calculating weight values for additional channels in the first convolutional layer of the proposed VGG-16 five-channel neural network model.
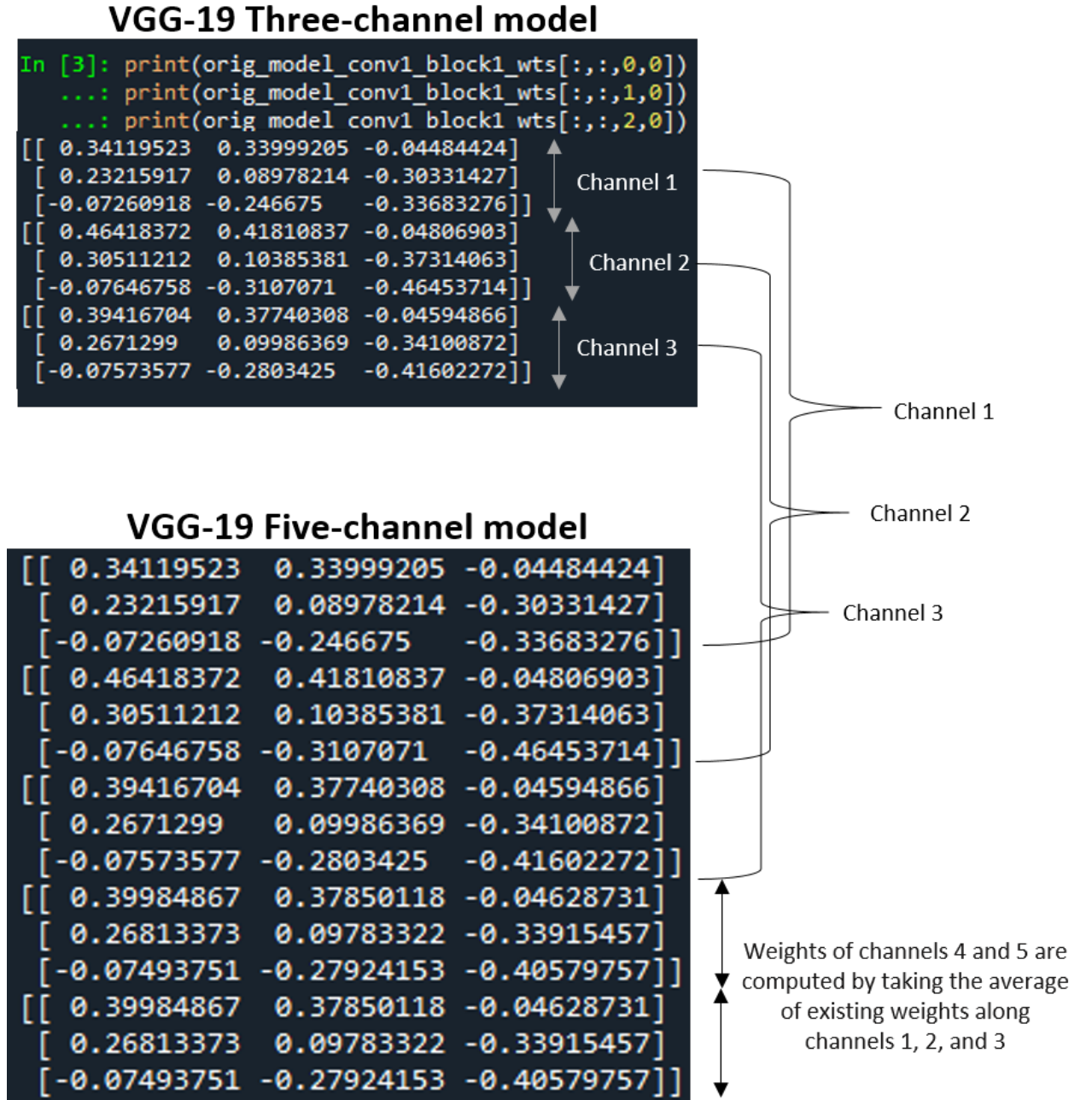
Figure 6.5: Calculating weight values for additional channels in the first convolutional layer of the proposed VGG-19 five-channel neural network model.

VGG-19 models is elaborated in detail in the following section.

6.3   Research Implementation

This section outlines the detailed implementation of both the traditional three-channel
VGG-n architecture and the proposed five-channel VGG-n architecture. Subsequently,
the performance of these models is evaluated on real-world scenarios using various
performance metrics for binary crack classification tasks.

6.3.1   Implementation of Traditional VGG-n Model

The conventional VGG-16 model has 16 weight layers (13 convolutional and 3 fully
connected) as shown in Figure 6.6, and the conventional VGG-19 model has 19 weight
layers (16 convolutional and 3 fully connected) as shown in Figure 6.7. The two tradi-
tional models are trained and tested on three-channel images i.e., the diffused images
utilising stratified five-fold cross-validation, hyperparameter tuning, and early stopping
regularization techniques to prevent overfitting. The hyperparameter tuning is per-
formed using learning rates of 0.1, 0.01, and 0.001, and batch sizes of 16, 32, and 64.
The number of epochs is set to 100 with early stopping criteria.

6.3.2   Implementation of the Five-Channel VGG-n architecture for
    Image Classification

The block diagram in Figure 6.8 showcases the step-by-step process involved in modify-
ing and fine-tuning the VGG-n model for five channel crack classification. The VGG-n
(where n=16, 19) five channel binary classifier models are represented in detail in Figure
6.9 and Figure 6.10, respectively.

   1. Five-channel dataset preparation: To prepare data compatible with the im-
plemented five-channel deep learning model, a dataset of five-channel TIFF images is
created as explained in subsection 6.2.3. The entire five-channel dataset of 1335 images
is split into training and testing sets, and these images are manually annotated as crack
or no crack. For these 1335 images, stratified five-fold cross-validation splits the dataset
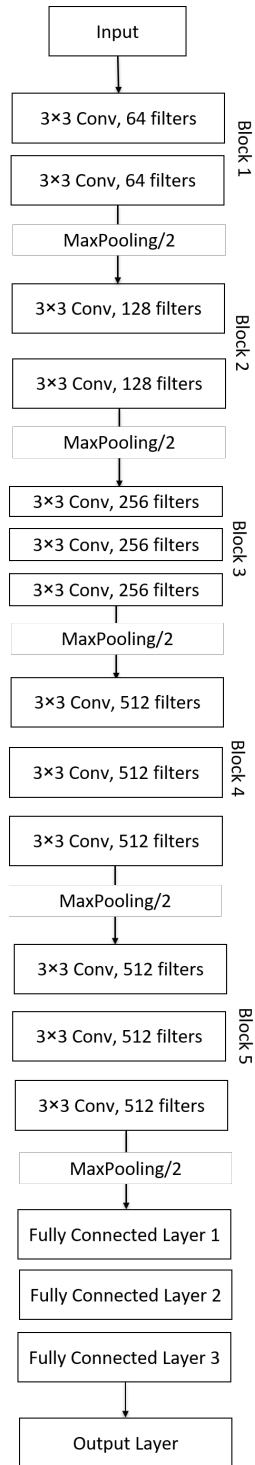
Figure 6.6: Architecture of Traditional VGG-16 3-channel model.
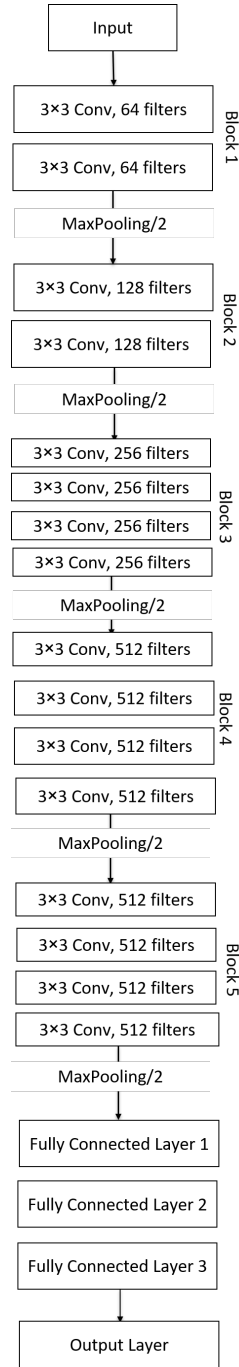
Figure 6.7: Architecture of Traditional VGG-19 3-channel model.
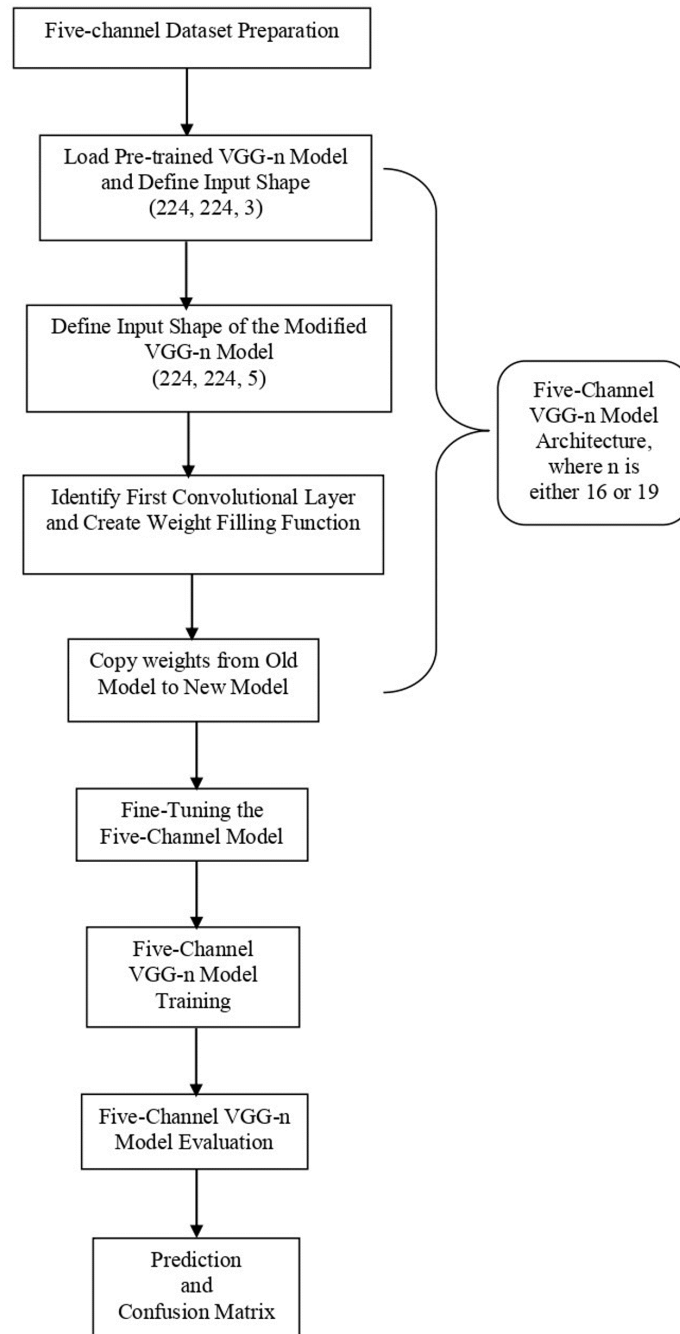
Figure 6.8: Step-by-step Implementation of the five-channel/modified VGG-n Model for crack classification, utilising transfer learning and fine-tuning approaches.

into five folds, each containing 267 images. In each iteration, 1068 images (four folds) are used for training, while 267 images (one fold) are used as the test set. This process repeats five times, ensuring each fold is used as a test set once as shown in Figure 3.5.

2. Five-Channel VGG-n Model Architecture: The five-channel model is implemented using the conventional three-channel model without including the top (fully connected) layers. The input shape of the traditional model is adjusted to accommodate the five-channel images by modifying its configuration as explained in subsection 6.2.4.

3. Fine-Tuning: To fine-tune the model, most of the layers are made non-trainable, except for the last six layers. This means that during training, only the weights of these six layers will be updated. A custom sequential model is built on top of the base model as shown in Figure 6.9 and Figure 6.10.

4. Model Training: The implemented five-channel model was trained on the training data. After experimenting with various combinations of batch size and learning rate, the optimal performance was achieved with a learning rate of 0.01 and a batch size of 32. The model was compiled using the Adam optimiser and the binary cross-entropy loss function.

5. Model Evaluation: After training, the performance of the model is evaluated on testing dataset using evaluation metrics described in section 5.3. The proportion of false positive predictions out of all actual negative instances, known as the False Positive Rate (FPR) shown in equation 6.1, is also computed.

$$PPV = \frac{FP}{TN + FP} \tag{6.1}$$

6. Prediction and Confusion Matrix: The trained model is the used to make predictions and the predicted probabilities are converted into binary class labels (0 or 1). To further assess the performance of the model, a confusion matrix is computed. The performance of the conventional three-channel VGG-n models and five-channel VGG-n model is evaluated using various metrics described in Section 5.3 for binary crack classification tasks. To determine these metrics, the precision, recall, accuracy, F1 score,
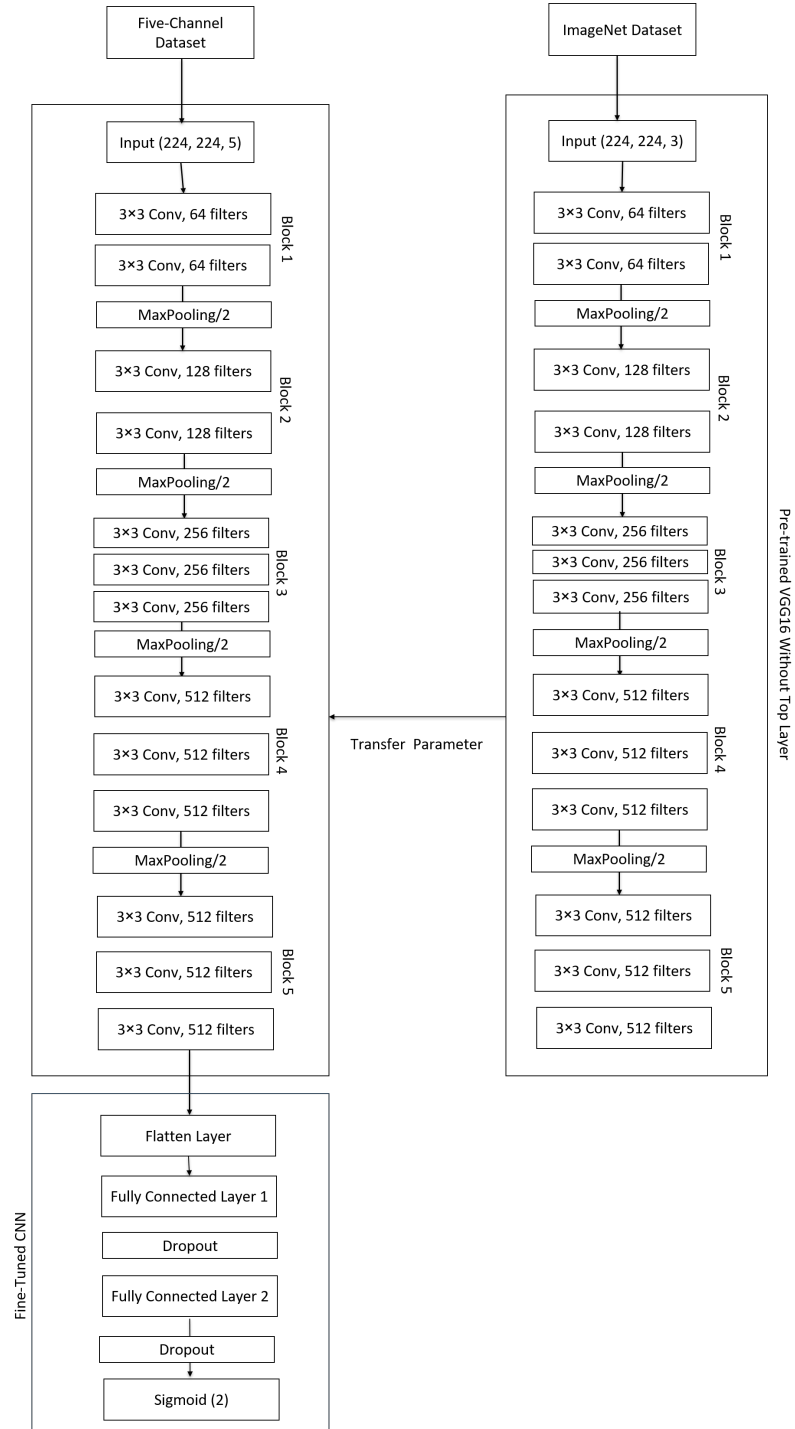
Figure 6.9: Architecture of VGG-16 Five-channel binary classification model.
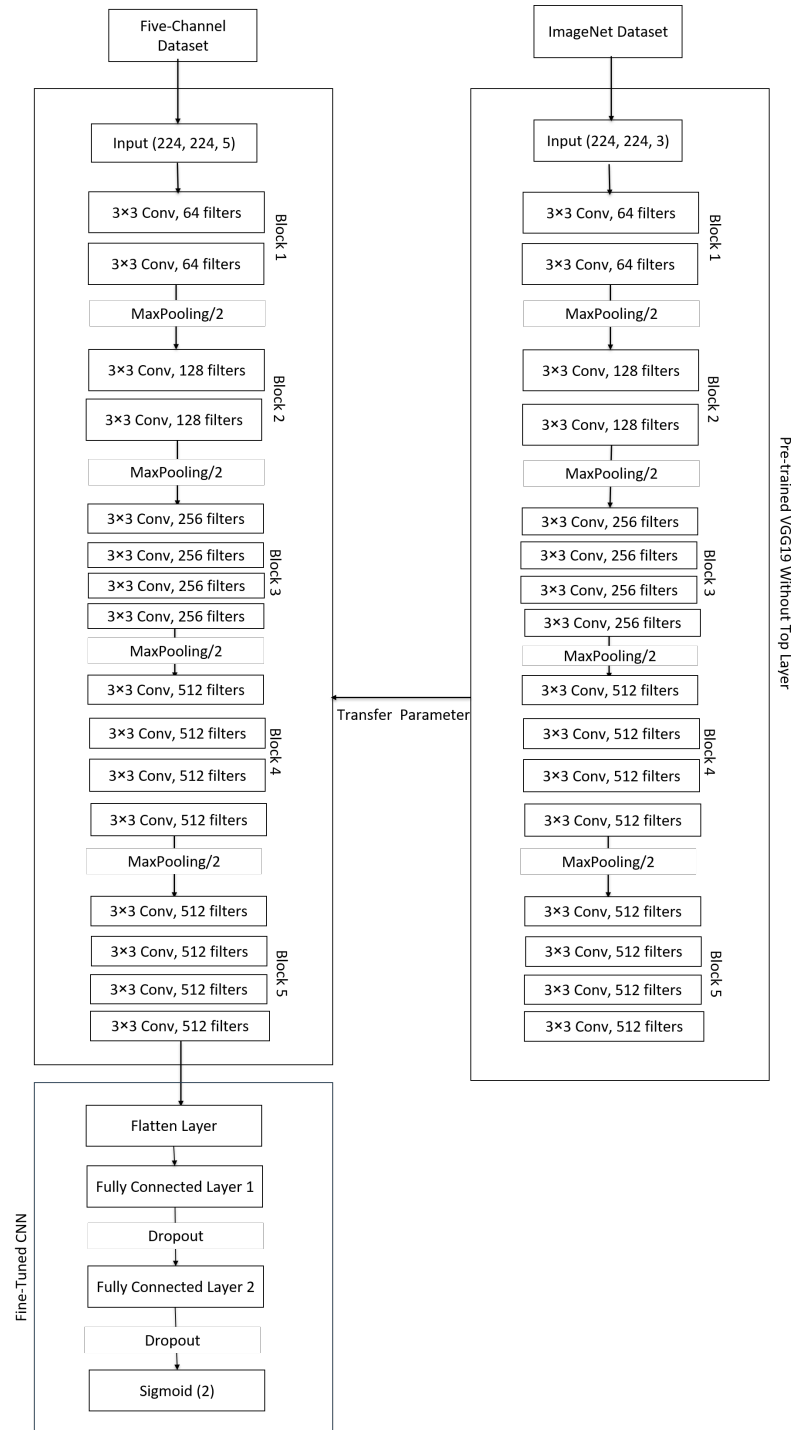
Figure 6.10: Architecture of VGG-19 Five-channel binary classification model.

and MCC are calculated for each of the five folds of the model and then averaged using
equations (6.2) - (6.6).

$$\text{Overall Precision} = \frac{1}{k}\sum_{i=1}^{k}\text{Pre}_i \tag{6.2}$$

$$\text{Overall Recall} = \frac{1}{k}\sum_{i=1}^{k}\text{Rec}_i \tag{6.3}$$

$$\text{Overall Accuracy} = \frac{1}{k}\sum_{i=1}^{k}\text{Acc}_i \tag{6.4}$$

$$\text{Overall F1 score} = \frac{1}{k}\sum_{i=1}^{k}\text{F1}_i \tag{6.5}$$

$$\text{Overall MCC} = \frac{1}{k}\sum_{i=1}^{k}\text{MCC}_i \tag{6.6}$$

where:

- i indicates the range of folds, which vary from 1 to k;

- k indicates the total number of folds in the cross-validation;

- Pre indicates Precision;

- Rec indicates Recall;

- Acc indicates the Accuracy;

- F1 indicates F1 score;

- MCC indicates Matthew's correlation coefficient.

Figure 6.11 provides a visual representation of the data preparation, model archi-
tecture, and crack classification process. This design aims to leverage the capabilities
of the modified model to accurately classify concrete crack specimens based on the
images captured under directional lighting conditions. The VGG-n-five-channel takes

the five-channel images as input and provides a binary classification result as either positive or negative, indicating the presence or absence of a crack, respectively.

6.4 Results

6.4.1 Determination of best performing three and five channel models

The average accuracy across all folds for different combinations of hyperparameters was computed as shown in Figure 6.12 and Figure 6.13. It is evident from the graph that the accuracy of the five-channel models is consistently better compared to the three-channel models in both VGG-16 and VGG-19 architectures.

Figure 6.12 illustrates that the VGG-16 five-channel model achieved the highest accuracy of 95.1% using the hyperparameter combination (32, 0.001), exceeding the VGG-16 three-channel model's highest accuracy of 89% with the same hyperparameters. Similarly, Figure 6.13 shows that the VGG-19 five-channel model, using the hyperparameter combination (64, 0.001), achieved the highest accuracy of 91.4%, outperforming the VGG-19 three-channel model's highest accuracy of 88.4% obtained with the (32, 0.001) hyperparameter combination.

6.4.2 Comparison of VGG-n three and five channel models

Figure 6.14 provides a detailed comparison of the performance metrics between the top-performing VGG-16 and VGG-19 models across both three-channel and five-channel configurations. Additionally, Table 6.2 displays the corresponding confusion matrices for these models. In the case of VGG-16, transitioning from the three-channel to the five-channel configuration resulted in significant improvements across all metrics: accuracy increased from 89% to 95.1%, precision from 89% to 96.3%, recall from 89% to 94.2%, F1 score from 89% to 95.2%, and MCC from 77.5% to 90.3%. Similarly, for VGG-19, the five-channel model showed noticeable improvements compared to the three-channel model: accuracy improved from 88.4% to 91.4%, precision from 90.3% to 92.5%, recall from 87.1% to 90.5%, F1 score from 88.6% to 91.5%, and MCC from 77% to 83%.

```
┌─────────────────────────┐
│     Hardware setup      │
└─────────────────────────┘
             │
             ▼
┌─────────────────────────┐
│   Directional images    │
│  captured in R, D, L, U │
│     and A directions    │
└─────────────────────────┘
             │
             ▼
┌─────────────────────────┐
│     Crop images into    │
│    224×224 blocks       │
└─────────────────────────┘
             │
             ▼
┌─────────────────────────┐
│      Formation of       │
│   five-channel image    │
│     (224, 224, 5)       │
└─────────────────────────┘
             │
             ▼
┌─────────────────────────┐
│    VGG-n five-channel   │
│    binary classifier    │
│     (224, 224, 5)       │
└─────────────────────────┘
        │           │
        ▼           ▼
┌───────────┐  ┌───────────┐
│  Cracks   │  │ No Cracks │
└───────────┘  └───────────┘
```
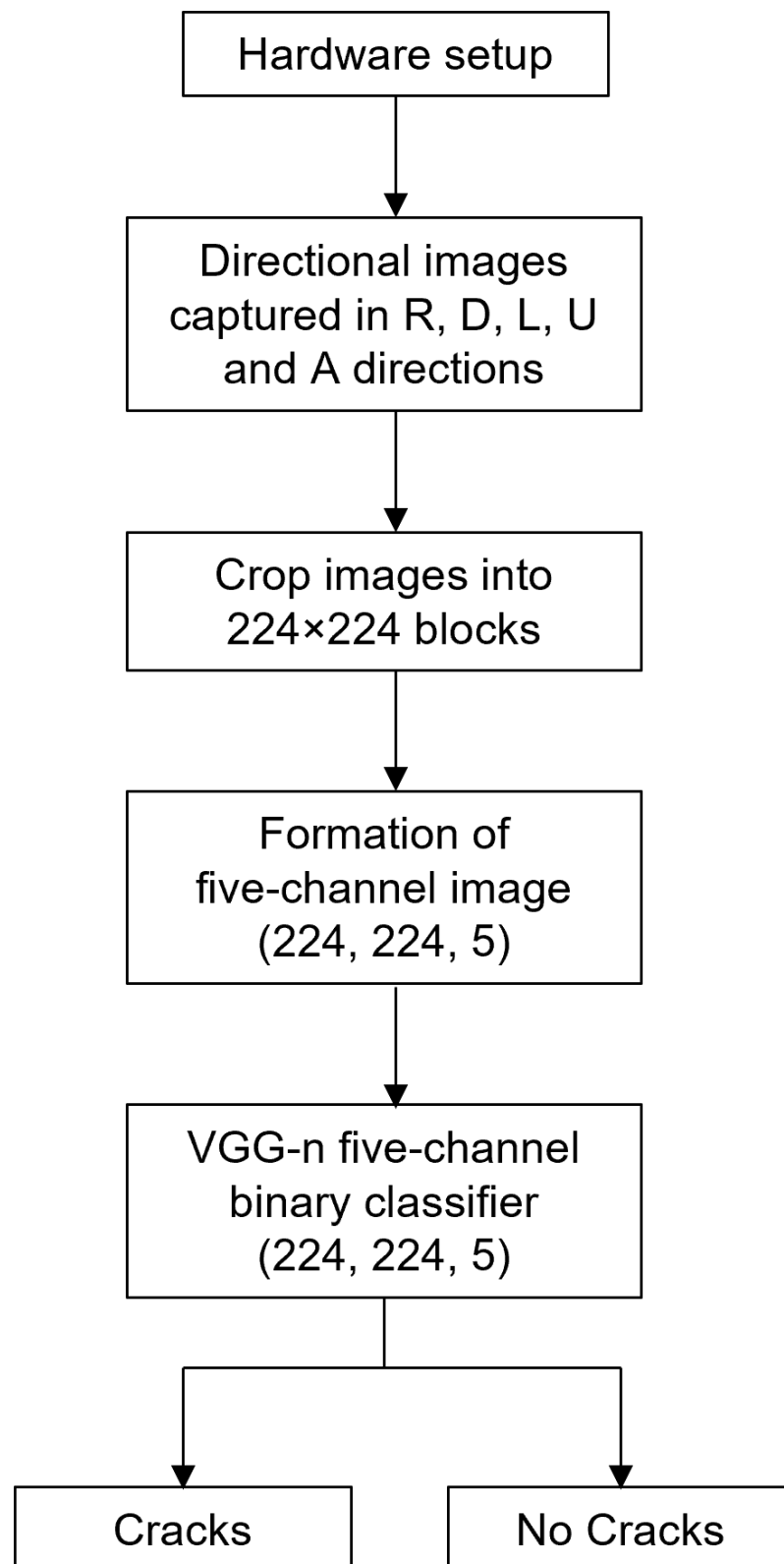
Figure 6.11: Step-by-step Implementation of the Modified VGG-n Model for Five-Channel Image Classification.
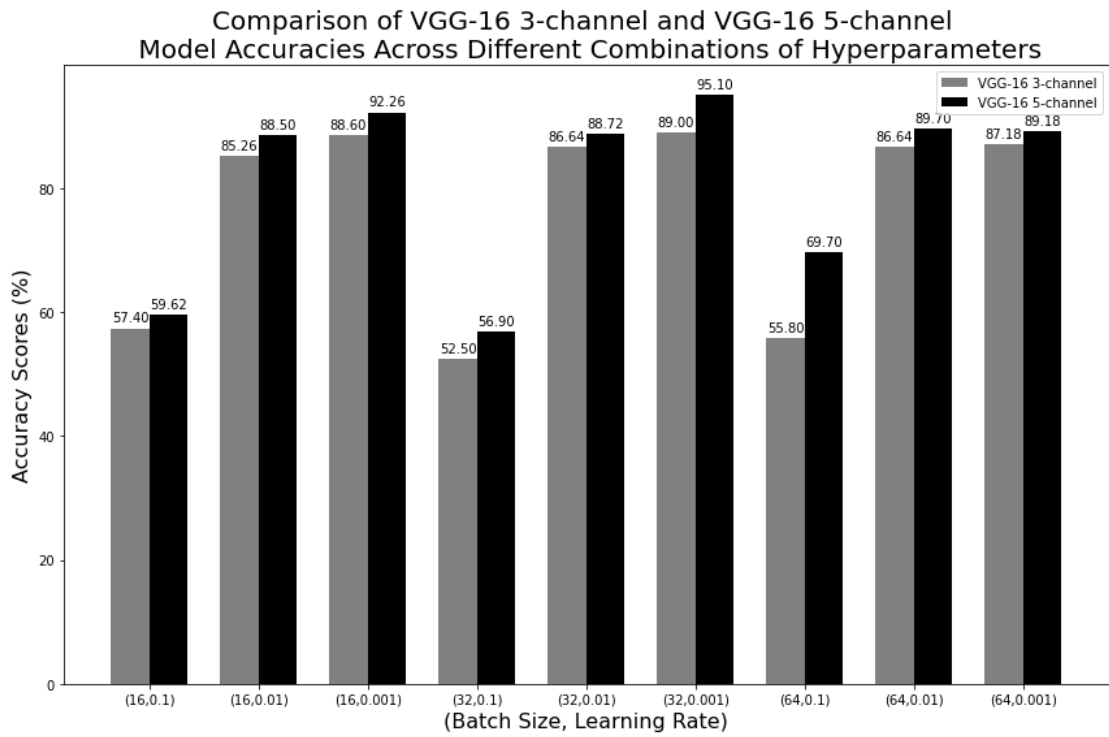
Figure 6.12: VGG-16 three-channel crack classification model vs VGG-16 five-channel crack classification model across different combinations of hyperparameters.
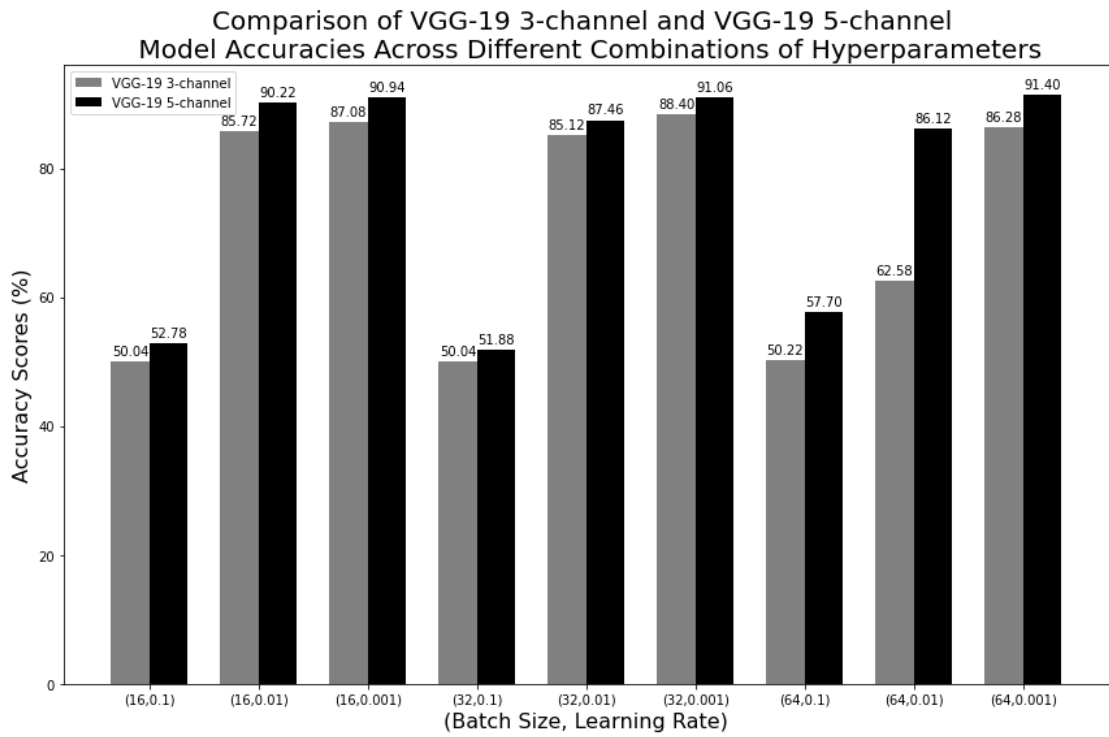
Figure 6.13: VGG-19 three-channel crack classification model vs VGG-19 five-channel crack classification model across different combinations of hyperparameters.
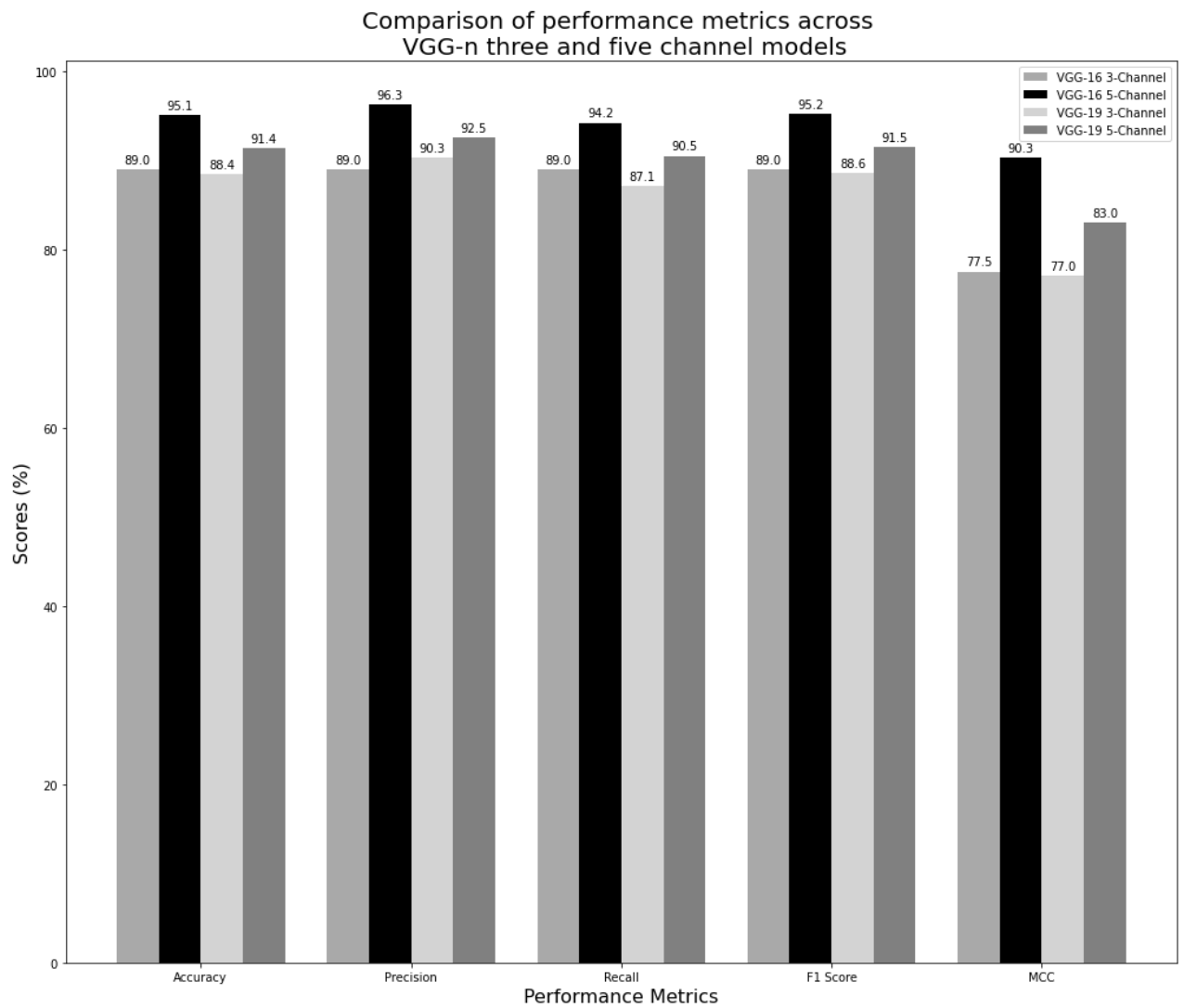
Figure 6.14: Comparison of performance metrics across VGG-n three and five channel models.

Therefore in both architectures, the five-channel models consistently exhibited higher accuracy, indicating their improved ability to accurately classify cracks. They also achieved higher precision, i.e., they make fewer false positive predictions. Additionally, their higher recall rate signifies fewer false negatives, ensuring that most of the actual cracks were detected. Moreover, the higher F1 score demonstrates a balanced performance between precision and recall, indicating accurate classification of both cracks and non-cracks without bias towards either class. Furthermore, the Matthews correlation coefficient reflects improved overall performance in crack detection tasks i.e., the five-channel models perform well across all aspects of classification. The findings suggest that the five-channel models exhibit better performance compared to traditional three-channel models. This is because, incorporation of additional channels into the traditional models provided additional information and features, thereby improving their ability to learn more distinctive representations of concrete cracks.

Table 6.2: Confusion matrices of VGG-n three and five channel models.

Model	TP	FP	FN	TN
VGG-16 3-channel	119	15	15	118
VGG-16 5-channel	129	5	8	125
VGG-19 3-channel	121	13	18	116
VGG-19 5-channel	124	10	13	121

When comparing the five-channel configurations of VGG-16 and VGG-19 across various performance metrics, VGG-16 consistently outperformed VGG-19. Specifically, VGG-16 five-channel demonstrated an accuracy of 95.1%, precision of 96.3%, recall of 94.2%, F1 score of 95.2%, and MCC of 90.3%, whereas VGG-19 five-channel achieved 91.4% accuracy, 92.5% precision, 90.5% recall, 91.5% F1 score, and 83% MCC. These findings indicate that VGG-16 five-channel outperformed VGG-19 five-channel in accurately identifying and categorising cracks from non-crack regions. The improved accuracy in VGG-16 five-channel model is likely due to the additional spatial information captured from different directions, enhancing feature extraction. The VGG16 outperforming VGG19 has been observed before in the international literature by [219]. Traditional models learn features obtained from single image only, while the five-channel

140

model leverages information from five different images captured from different directions, thereby enhancing its robustness and ability to learn significant features. In case of traditional models that utilise only information obtained from one single image, whereas in case of five-channel images the model learn important features from five different images, making model more robust.

6.4.3 Evaluation Time

The evaluation times of various VGG-16 and VGG-19 model configurations were measured to assess their computational efficiency. The results indicate that the five-channel configurations are faster in testing an image compared to their three-channel models. Each channel captures a gray scale version of the image from different lighting directions, allowing the model to better understand texture, and edge details. In contrast, the three-channel RGB model relies on colour variations, which may be less effective in highlighting structural differences.

The trained five-channel and traditional three-channel models were used to evaluate whether an image contained a crack. The five-channel VGG-16 model required 57 milliseconds, and the five-channel VGG-19 model took 65 milliseconds. In contrast, the traditional VGG-16 and VGG-19 models required 87 milliseconds and 107 milliseconds, respectively. These results indicate that the five-channel models, particularly the five-channel VGG-16, achieved faster evaluation times compared to the traditional models.

The improvement is due to the reduced complexity of grayscale images in the five-channel model compared to the RGB images in the three-channel model. In an RGB image, each pixel has three colour values (Red, Green, and Blue), each having its own intensity value, requiring the model to process colour dependencies. In contrast, a five-channel gray scale image consists of five gray scale images, each with a single intensity value per pixel (0-255), eliminating colour information and simplifying data processing per channel. Although the five-channel model has more channels, the gray scale inputs reduce the computational load by focusing on structural features relevant to tasks like crack detection, enabling fewer convolutional operations in early layers and producing more efficient feature maps, thereby leading to faster evaluation times. Additionally,

the simpler gray scale images allow for more efficient memory access, further enhancing processing times. These findings highlight the advantages of utilising additional channels, with VGG-16 five-channel models demonstrating the shortest evaluation times among all configurations used in this chapter. Figure 6.15 visually illustrates these differences, where VGG-16 three-channel model was represented in dark grey, VGG-16 five-channel in black, VGG-19 three-channel in light grey, and VGG-19 five-channel in grey, allowing clear comparison of their respective evaluation speeds.
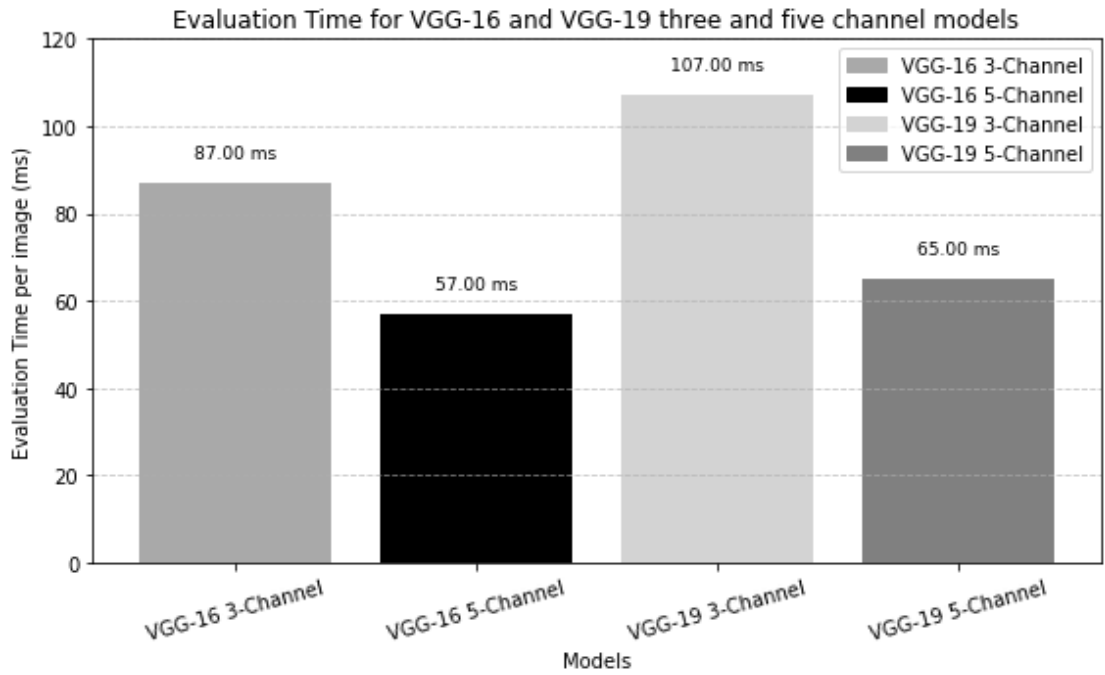
Figure 6.15: Evaluation Time for VGG-16 and VGG-19 three and five channel models.

6.5 Summary

The comparative analysis of VGG-16 and VGG-19 with both three-channel and five-channel configurations revealed that the five-channel models improved crack detection and classification in concrete structures, particularly in low-light environments. The study employed various methodologies such as hyperparameter tuning, regularization techniques, and stratified five-fold cross-validation to enhance performance of the model

142

and minimise the risk of overfitting. These approaches are crucial for improving the robustness of the model on new and unseen data.

When comparing the three-channel and five-channel models, the five-channel models consistently outperformed their three-channel models across different evaluation metrics such as TPR, FPR, precision, F1 score, accuracy, and MCC. Furthermore, a comparison of the two five-channel VGG models showed the superiority of VGG-16 over VGG-19, with performance improvements ranging from 4% to 9% across different evaluation metrics. These findings highlight the effectiveness of the five-channel VGG-16 model and its ability to accurately detect and classify cracks with widths as small as 0.1mm.

The incorporation of directional lighting, transfer learning, and fine-tuning approaches significantly contributed in improving the overall performance of the proposed models. The VGG-16 five-channel model took the shortest time to evaluate an image, requiring nearly 57 milliseconds, compared to both the VGG-19 models and the traditional VGG-16 model. This highlights the potential benefits of utilising additional channels for improved computational efficiency in image processing tasks. As presented in Chapter 3, despite it being developed a decade ago, the VGG16 model is actively used in research for crack detection and this work contributes to this through the development of a new VGG16 model that allows for the implementation of directional lighting in deep learning.

Chapter 7

Directional Lighting Enhanced Crack and Spalling Detection in Concrete Structures

Chapter 6 demonstrated the significance of directional lighting and made such images possible within deep learning by implementing a five-channel neural network model using the VGG architecture. The VGG model exhibited an evaluation time of 57 ms, and was focused on crack detection as the initial dataset contained images captured from laboratory and real-world concrete structures that contained only cracked samples. VGG-based models can be computationally expensive due to their large number of parameters [120, 220]. These models are best suited for use in controlled environments such as research laboratories, where computational resources are readily available and processing time is not a critical constraint. The deeper architecture of VGG model enables the extraction of spatial features, making them particularly suitable for tasks requiring detailed visual analysis [120, 221]. As a next step, this Chapter expands the dataset and methodology to detect additional types of damage, such as spalling in concrete structures.

At the earlier stages of this research, VGG16 was selected due to its well known recognition in various image classification and feature extraction tasks and its simplicity

compared to other models [121, 122]. Its deep architecture made it a favourable choice as an initial baseline model for detecting structural defects in images. However, as the research progressed over time, the computational limitations and relatively large number of parameters in VGG16 became more apparent, especially when aiming for faster speeds and potential deployment in real-time or resource-constrained environments. As future aim of ALICS is to integrate the NN model into the hardware, so that image capture and image processing go simultaneously, it is essential that the model needs be light weight, and faster. MobileNetV2 offered a significant advantage in terms of computational efficiency, achieving comparable performance with a reduced number of parameters compared to VGG16 as demonstrated in [119, 120]. So, MobileNetV2 was chosen in this study.

7.1   Introduction

This research aims to closely replicate the methods used by the human inspectors in automated inspection systems, while at the same time lead to results faster and in a more systematic, less subjective, way. Two models, FusedNN and MCNet, were developed that used images captured under directional lighting conditions. The directional images Right (R), Down (D), Left (L), Up (U), and Diffused (A) are captured using ALICS (described in Chapter 4). The workflow in this chapter is summarised as:

1. Investigate the importance and potentiality of directional lighting in multi-class image classification tasks, as opposed to binary crack classification presented in Chapter 6.

2. Introduce a novel model, FusedNN, which employs pixel-based maximum intensity image fusion technique to obtain a single fused image from the five images captured under different lighting conditions.

3. Develop five-channel neural network model, called multi-channel neural network model (MCNet) to effectively classify cracks and spalling in concrete structures and reduce evaluation time.

4. Compare the performance of the FusedNN and MCNet models with the con-

ventional (traditional) three-channel model and one of the advanced models in the literature i.e., Zoubir's model [126] which was described in Section 3.3.

5. Compare VGG-16 five-channel neural network model (Chapter 6) with the MCNet model developed in this chapter for binary classification. The best-performing model will be further utilised to implement a multi-class classifier.

6. Investigate how increased exposure impacts pixel intensity in fused images and evaluate whether this impact is similar for fused and diffused images and assess the subsequent impact on the performance of neural network model.

7.2 Background

The maximum intensity pixel-level image fusion method [222] is widely utilised in the medical field to merge information from various modalities and improves the quality and clarity of medical images by selecting the pixel with the highest intensity from each corresponding location in the input images. This method can be a powerful tool for detection of defects for civil infrastructure. In this chapter, the author introduces a novel fused neural network model, named FusedNN, which generates a directionally lit fused image by selecting the maximum intensity value for each pixel from five directional images.

This study implements two novel models based on MobileNetV2 architecture explained in Section 3.3. The traditional model shown in Figure 3.4 is trained and tested on three channel diffused images.

7.3 Methodology

The traditional MobileNetV2 model has three-channel input which corresponds to the RGB components of an image. The author proposes two new models, FusedNN and MCNet to utilise images captured in five lighting directions to generate fused and five-channel images, with three and five channel inputs, respectively. The implementation of the three-channel FusedNN model is straightforward and similar to the traditional three-channel model but uses fused images instead. Therefore, the input shape of the

MCNet model should be modified to accept five channel images. The implementation of the multi-channel neural network model, named as MCNet by the author, is similar to the VGG-n five-channel model, as explained in Section 6.2.3 and Section 6.2.4, but utilises MobileNetV2 architecture instead.

7.3.1 Dataset Description and Pre-processing

The same dataset presnted and used in Chapter 4, is used for the work described in this Chapter too. A total of 45 images were captured, by projecting light onto concrete surfaces in R, D, L, U, and A directions, respectively as shown in Figure 4.18 and explained in Section 4.3. The acceptable input size of the MobileNetV2 architecture is 224×224 pixels. Therefore, a pre-processing step was applied to the original images of dimensions 5429×3458 pixels. These images are cropped into multiple 224×224 pixel blocks. The cropped blocks undergo further pre-processing to form five-channel and fused images before being fed into the proposed models.

For the binary classification models, a total of 1335 sub-images were utilised, while for the multi-class image classification task, a dataset of 520 sub-images were used. These images were manually labelled as 'cracked' or 'uncracked' for evaluating the performance of four models in binary classification tasks. For multi-class image classification, the sub-images were annotated as "thick cracks," "thin cracks," "spalling," and "uncracked."

For 1335 images, stratified five-fold cross-validation splits the dataset into five folds, each containing 267 images. In each iteration, 1068 images (four folds) are used for training, while 267 images (one fold) are used as the test set. For 520 images, each fold consists of 104 images, with 416 images used for training and 104 images for testing in each iteration. This process repeats five times, ensuring each fold is used as a test set once.

7.3.2 Generation of five-channel image

MCNet utilises five-channel input image of dimension (224, 224, 5), which was generated as explained in Section 6.2.3. The architecture of the five-channel MobileNetV2 model

i.e., MCNet was modified to accept an input image of dimension (224, 224, 5).

7.3.3   Generation of fused image

The maximum intensity image fusion technique [222] is used to obtain a single fused
image from the five images captured under different lighting conditions. This involves
comparing the pixel values of corresponding pixels from the five directional images at
each pixel position (x,y). The maximum value among the five input images is selected
for each pixel. This process is repeated for each pixel across all five input images to
generate the final fused image, as shown in Figure 7.1. This algorithm selects the
brightest regions from each input image by choosing the greatest intensity value for
each pixel. The heatmap clearly indicates the contributing pixels from each direction.
The maximum intensity fusion technique for the directional images can be expressed
using equation (7.1).

$$F_{xy} = \sum_{x=1}^{5429} \sum_{y=1}^{3458} \max\left(R_{xy}, D_{xy}, L_{xy}, U_{xy}, A_{xy}\right) \tag{7.1}$$

where F_{xy} represents the pixel intensity value at position (x,y) in the output fused
image F. R_{xy}, D_{xy}, L_{xy}, U_{xy}, A_{xy} are the pixel intensity values at position (x,y) in five
directional input images R, D, L, U, and A, respectively.

7.4   Research Implementation

All the models are implemented using hyperparameter tuning, the early stopping reg-
ularization technique, and a stratified five-fold cross-validation approach. The hyper-
parameter tuning is performed using learning rates of 0.1, 0.01, and 0.001, and batch
sizes of 16, 32, and 64. A stratified five-fold cross-validation approach was used to en-
sure that the performance of the model was consistent across multiple (five) balanced
subsets of the data, rather than relying on a single train-test split. This helps to detect
if the model is overfitting to specific data partitions. An early stopping was applied
during training, with monitoring on the validation loss. If the validation loss did not
improve over a predefined number of epochs (patience), training was stopped at that

Right (R) Down (D) Left (L) Up (U) Diffused (A)

$$F_{xy} = \sum_{x=1}^{5429} \sum_{y=1}^{3458} max\ (\ R_{xy}\ ,D_{xy}\ ,L_{xy}\ ,U_{xy},A_{xy}\ )$$

Maximum intensity-based image fusion technique

(224, 224, 3)

Fused image

Right
Below
Left
Up
All

Heatmap of fused image showing contributing pixel directions
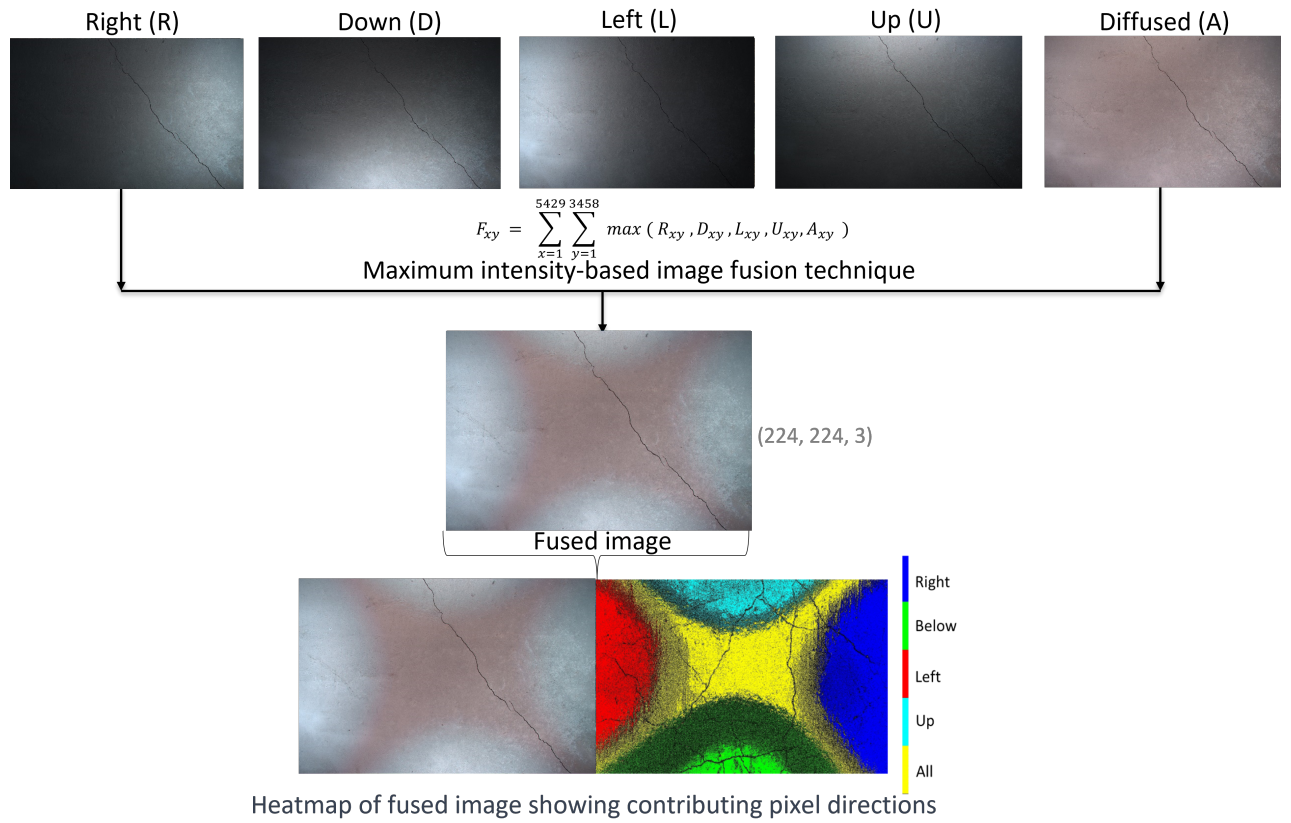
Figure 7.1: The images are captured in R, D, L, U, and A directions, respectively. These images are fused into a single image using the maximum-intensity image fusion technique. The heatmap of the resulting fused image indicates the contributing pixels from each direction.

point. This not only reduces overfitting but also optimizes training time. The number of epochs is set to 100 with early stopping criteria.

The following sections outline the implementation of the proposed models shown schematically in Figure 7.2.

7.4.1 Implementation of the FusedNN Model

The architecture of FusedNN is shown in Figure 7.3. FusedNN that uses MobileNetV2 architecture was initially trained on pre-trained neural network model using ImageNet dataset and further trained on fused image dataset to adapt to the new task. A custom sequential model is built on top of the base MobileNetV2, including two Dense layers with ReLU (Rectified Linear Unit) activation, and a final Dense layer with either a sigmoid activation function for binary classification or softmax for multi-class classification. The model is compiled with the Adam optimiser and uses the binary cross-entropy loss function for binary image classification and sparse categorical cross-entropy loss function for multi-class image classification.

7.4.2 Evaluating the Impact of Exposure Control on FusedNN and Traditional Model Performance

The FusedNN model uses fused images generated through the maximum intensity fusion technique, where all images are captured under auto-exposure settings. A key concern is whether increasing exposure could produce diffused images with brighter pixels that resemble those selected by FusedNN's maximum intensity fusion process. If so, implementing the fusion technique may not offer additional advantage against existing models. Therefore, it is important to evaluate and compare the performance of two models—FusedNN and a traditional MobileNetV2 model across various exposure settings. This investigation aims to determine if adjusting exposure levels of directional images can improve the performance of the traditional model compared to the FusedNN model.

For this examination, two distinct concrete samples generated in the laboratory were chosen. These images were captured under diffused lighting conditions with exposure

Figure 7.2: Step-by-step implementation of FusedNN and MCNet models for binary
and multi-class image classification.
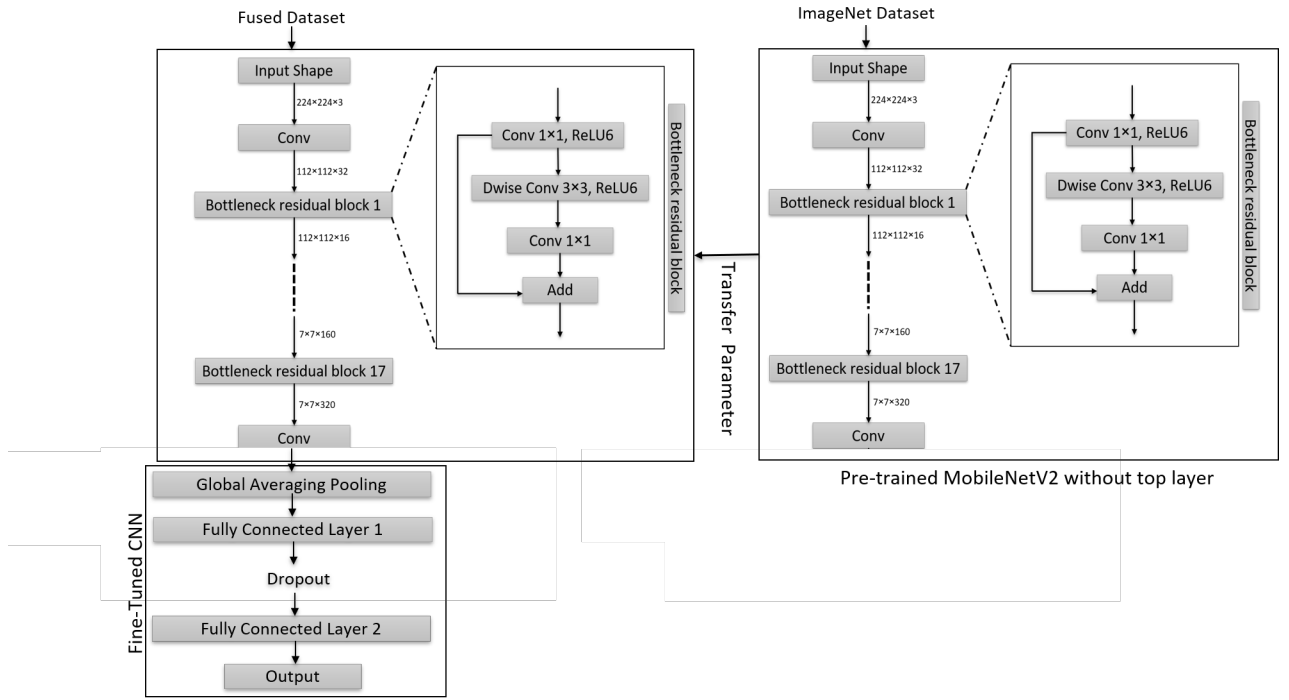
Figure 7.3: Architecture of FusedNN model.

values ranging from 200,000 to 900,000 lux-seconds, and an auto-exposure value of 470,000 lux-seconds was recorded. Additionally, directional images were captured with exposures ranging from 1.2 million to 2.2 million lux-seconds, and specific auto-exposure values were recorded for each direction: right (1.7 million lux-seconds), down (1.65 million lux-seconds), left (1.58 million lux-seconds), and up (1.7 million lux-seconds).

Following image capture, data pre-processing, detailed in Section 7.3.1, was conducted and divided the dataset of 1335 images into positive (cracked) and negative (uncracked) classes for subsequent analysis. To study the influence of exposure under various lighting conditions, the performance of the traditional MobileNetV2 model (detailed in Section 3.3) and the FusedNN model (outlined in Section 7.4.1) were evaluated. Both models were trained on the aforementioned images, aiming to understand how exposure affects the performance of these two models.

The FusedNN model, trained on fused images obtained at increased exposure levels, was compared with the Traditional MobileNetV2 model trained on diffused images

acquired under similar exposure conditions.   Additionally, another comparison was
conducted between the FusedNN model trained on fused images captured at auto-
exposure levels and the traditional MobileNetV2 model trained on diffused images
obtained under similar auto-exposure conditions.   The performance of these models
was evaluated using various evaluation metrics.   This helps to explore how exposure
settings and image fusion techniques influence crack detection accuracy across diverse
lighting conditions.

7.4.3   Implementation of the MCNet Model

The MCNet model was developed to improve defect detection in concrete structures
by utilising a customised five-channel input obtained from gray scale images captured
under various lighting conditions as explained in Section 7.3.2. Table 7.1 shows that
the three-channel models have 32 filters of size 3×3 kernel, resulting in a total of 288
weights. With three channels, the total number of weights are 864. In the case of the
five-channel models, the number of filters, filter size, and number of weights per filter
are the same. However, with five channels, the total number of weights becomes 1440.
Therefore, the number of trainable parameters for the three-channel and five-channel
MobileNetV2 models are 864 and 1440, respectively. This highlights the impact of the
number of channels on the total trainable parameters in the first convolutional layer of
the MobielNetV2 model, similar to that of five-channel VGG models implemented in
Chapter 6.

Therefore, modifications were focused on adjusting the trainable parameters in the
first convolutional layer to accommodate the five-channel input. The weights of the
remaining convolutional layers were directly transferred as it is from a pre-trained Mo-
bileNetV2 model (as they are unchanged). To handle the additional two channels,
weight values were computed by averaging existing channel weights, similar to that of
the multi-channel neural network model discussed in previous chapter, where Mobile-
NetV2 replaced VGG-n to optimise evaluation time. The block diagram in Figure 7.2
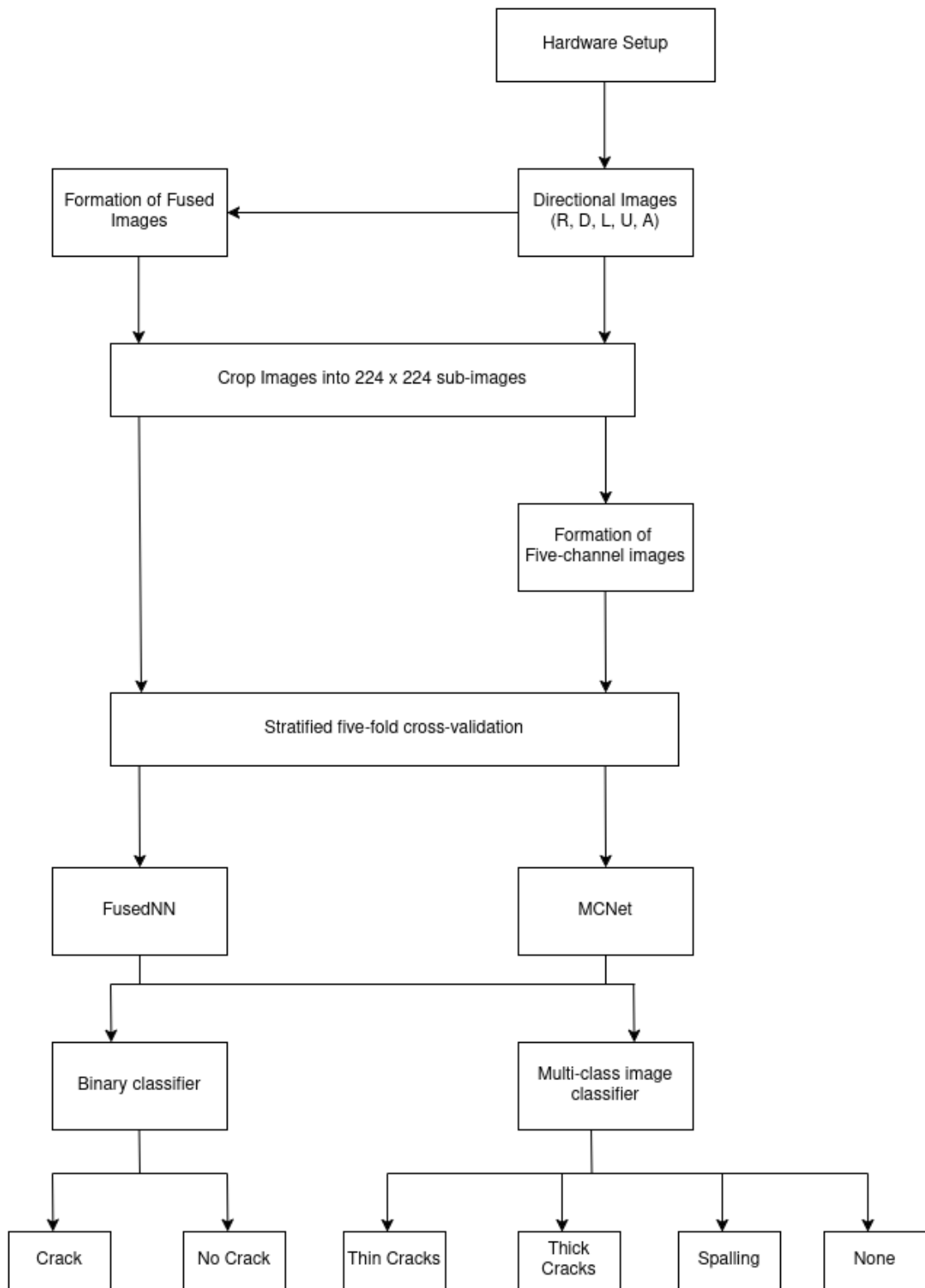showcases the step-by-step implementation of FusedNN and MCNet models for binary
and multi-class image classification. Each block represents a distinct stage or operation

involved in the process, demonstrating the logical flow of tasks.

Table 7.1: Parameter Comparison: 3-channel vs 5-channel MobileNetV2 models in the first convolutional layer.

Parameter	Values for three-channel MobileNetV2	Values for five-channel MobileNetV2
Number of filters	32	32
Filter size	3 x 3	3 x 3
Number of weights per filter	9	9
Total weights for all filters	32 x 9 = 288	32 x 9 = 288
Number of channels	3	5
Total weights for all channels	288 x 3 = 864	288 x 5 = 1440
Total trainable parameters	**864**	**1440**

7.4.3.1 Five-channel dataset preparation

Initially, the focus of this thesis was limited to crack detection due to the availability of a smaller dataset, which only included images of concrete cracks. This formed the basis of the research presented in the previous Chapters. As the research progressed and more annotated data became available, specifically for spalling, the scope of the study was expanded. In this Chapter, additional classes such as spalling were incorporated into the dataset. Consequently, the analysis for the work presented in this Chapter was conducted using an extended dataset, and the newer models such as MCNet and FusedNN were explored in combination with spalling data.

A five-channel dataset of 1335 images and 520 images was prepared to align with the requirements of the MCNet binary and multi-class image classification models, respectively. For binary classification, the labels were positive (crack-1) and negative (no crack-0). For multi-class image classification, the labels were thin crack-0, thick crack-1, spalling-2, and none-3.

7.4.3.2 Five-Channel MobileNetV2 (MCNet) Model Architecture

Figure 7.4 illustrates the architecture of MCNet model. MCNet was built upon the MobileNetV2 architecture, which was originally trained on ImageNet dataset, enabling it to capture low-level patterns and useful features from the images. During fine-tuning,

the weights of the pre-trained model are adjusted through training to accommodate the new five-channel dataset. To preserve the generalisation capabilities of the pre-trained model and prevent overfitting, the early layers were frozen while the later layers were made trainable. Specifically, only the weights of the last six layers are updated during training. This improves the robustness and accuracy of the model in image classification tasks.

The model, built as described in Section 7.4.3, is compiled with the Adam optimizer and uses binary cross-entropy for binary classification, while sparse categorical cross-entropy is used for multi-class classification.
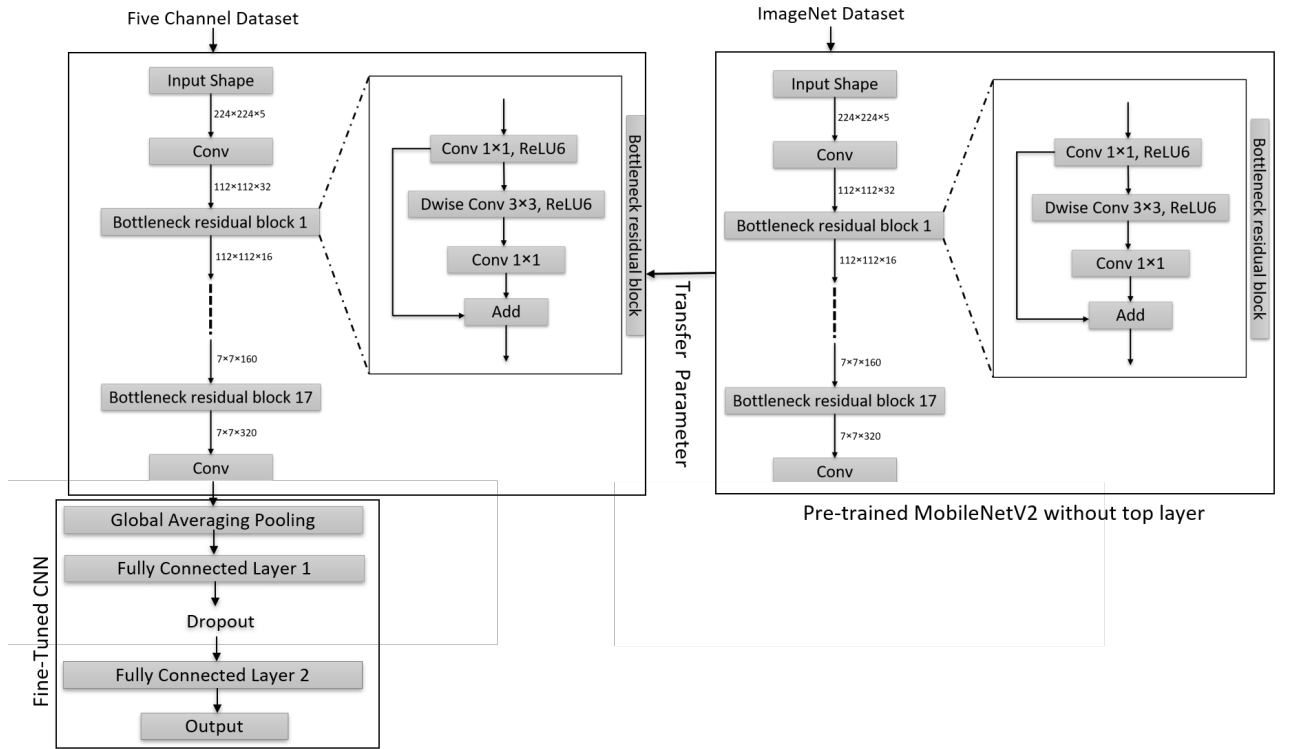
Figure 7.4: Architecture of five-channel MobileNetV2 (MCNet) model.

7.4.4   Benchmarking and Performance Evaluation

Hyperparameter tuning is conducted to identify the optimal learning rate and batch size, ensuring the most accurate and efficient model for the given datasets. The mod-

els also incorporate early stopping to prevent overfitting and employs five-fold cross-validation to ensure robust performance on unseen data. The performance of the conventional three-channel MobileNetV2, FusedNN, MCNet, and Zoubir's model is evaluated using the metrics described in Section 6.3.2 for binary-classification tasks. To determine these metrics, the precision, recall, accuracy, F1 score, and MCC are calculated for each of the five folds of the model and then averaged using equations (6.2) - (6.6). Table 7.2 was utilised to assess the performance of the model for multi-class image classification tasks.

7.5   Results and Discussion

One advantage of using multiple directions for image capture in FusedNN and MCNet is the availability of additional information that can aid in defect detection. To maintain consistency across the models, all four models were trained and tested using the same image captured under different lighting conditions as shown in Figure 4.18. The Zoubir model, similar to traditional model is trained and evaluated on diffused images. For example, an image captured under diffused lighting was used to train and test the MobileNetV2 and Zoubir model, similarly the same image captured under directional lighting (R,D, L, U, and A) was used to form a fused image and a five-channel image as shown in Figure 7.1 and Figure 6.1 to train and test the FusedNN and MCNet models, respectively. This step is crucial to ensure that any observed performance differences are primarily due to architectural variations. For each of the four models, the specific hyperparameter values that yielded the best performance were identified. These optimal hyperparameter values were further used to implement their corresponding multi-class image classification models.

7.5.1   Traditional MobileNetV2 Vs Zoubir Vs FusedNN Vs MCNet models for binary crack classification

Figure 7.5 shows a comparison of two proposed models: FusedNN and MCNet with the traditional and Zoubir models. All models achieved optimal performance with a batch

Table 7.2: Performance metrics of a multi-class image classification, where i ranges from 1 to the total number of classes, l. The average of the metric across all the classes is computed to obtain the macro average metric. M represents macro averaging. β controls the balance between precision and recall.

Metric	Description	Equation
$Accuracy_M$	The accuracy of each class is computed separately, and then take the average of all the accuracies to obtain macro-averaged accuracy.	$\sum_{i=1}^{l} \frac{1}{l} \frac{TP_i+TN_i}{TP_i+TN_i+FP_i+FN_i}$
$Precision_M$	The precision of each class is computed separately, and then take the average of all the precisions to obtain macro-averaged precision.	$\sum_{i=1}^{l} \frac{1}{l} \frac{TP_i}{TP_i+FP_i}$
$Recall_M$	The recall of each class is computed separately, and then take the average of all the recalls to obtain macro-averaged recall.	$\sum_{i=1}^{l} \frac{1}{l} \frac{TP_i}{TP_i+FN_i}$
MCC_M	The MCC of each class is computed separately, and then take the average of all the MCCs to obtain macro-averaged MCC.	$\sum_{i=1}^{l} \frac{1}{l} \frac{TP_i \times TN_i - FP_i \times FN_i}{\sqrt{(TP_i+FP_i)\times(TP_i+FN_i)\times(TN_i+FP_i)\times(TN_i+FN_i)}}$
$F1\text{-}Score_M$	The F1 score of each class is computed separately, and then the average of all the F1 scores is taken to obtain the macro-averaged F1 score.	$\frac{(1+\beta^2)\times\text{Precision}_M\times\text{Recall}_M}{(\beta^2\times\text{Precision}_M)+\text{Recall}_M}$

size of 32 and a learning rate of 0.001. The comparison of the best-performing models, along with their confusion matrices, is shown below.
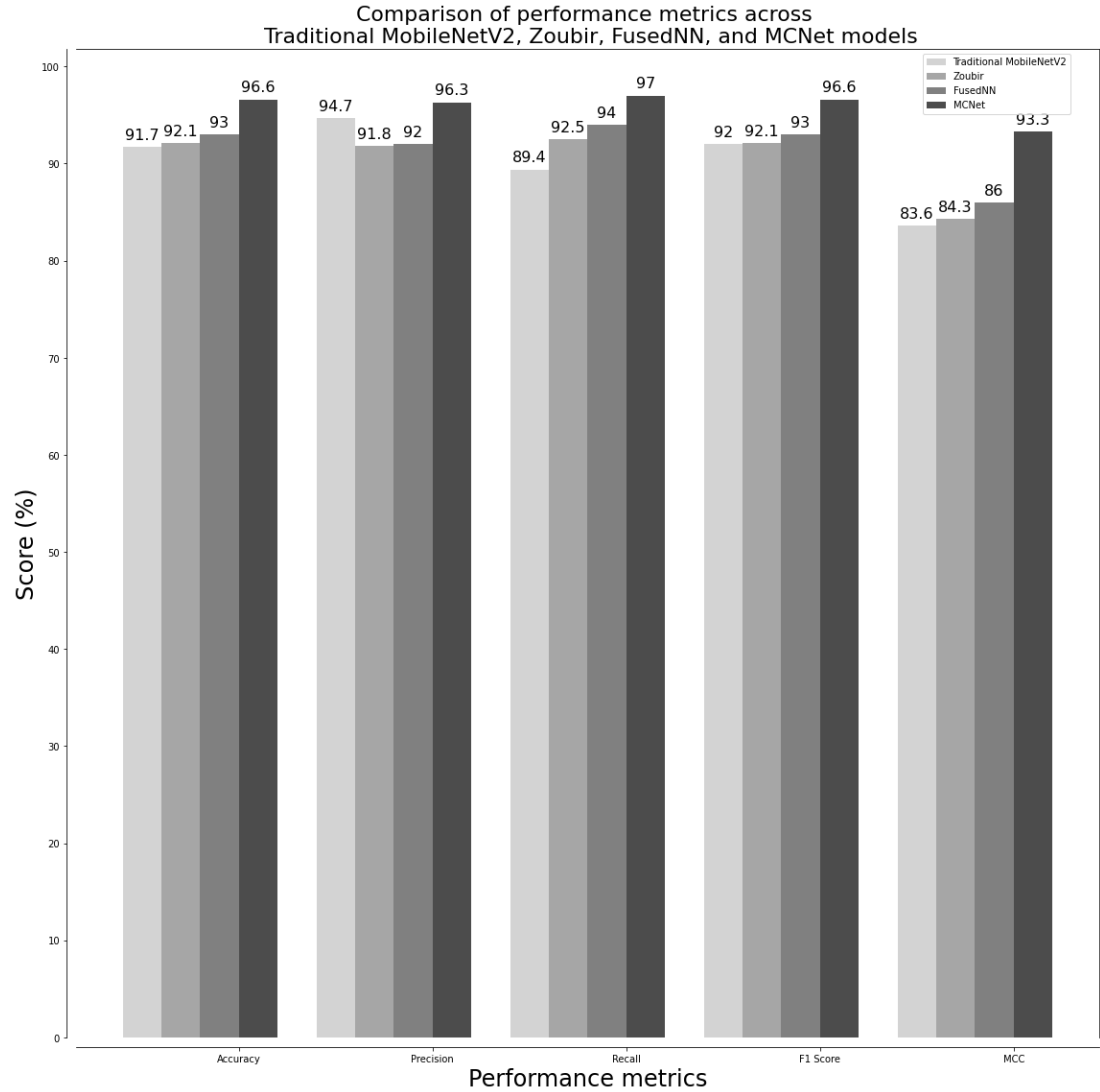
Figure 7.5: Comparison of classification metrics across Traditional MobileNetV2, Zoubir, FusedNN, and MCNet models.

FusedNN achieved a 1.3% higher accuracy, a 4.6% increase in recall, a 1% better F1 score, and a 2.4% improvement in MCC compared to the traditional MobileNetV2 model. Although FusedNN's precision is 2.7% lower than that of Traditional Mobile-NetV2, these metrics indicate that FusedNN is more effective overall, particularly in identifying actual cracks. The improved recall value demonstrates FusedNN's capability

to detect a higher number of true positives, minimising the risk of missing any cracks. The improved F1 score suggests that FusedNN maintains a good balance between precision and recall, ensuring both false positives and false negatives are well identified. The higher MCC value further validates the overall robustness and reliability of FusedNN in crack classification compared to traditional model. Although there is a slight decrease in precision, indicating a marginally higher rate of false positives, this trade-off is considered acceptable given the importance of identifying all potential cracks to prevent serious consequences for Civil infrastructure. This prioritises the identification of all potential cracks, even if it means tolerating a small number of false alarms.

When compared to Zoubir, FusedNN shows a slight edge in performance. It has a 1% higher accuracy and a 1.5% better recall, indicating a greater capability to correctly identify and detect cracks. FusedNN also has a 0.2% higher precision, reflecting a slightly lower rate of false positives. Notably, 0.9% F1 Score higher and 1.7% higher MCC than Zoubir's model, underscoring FusedNN's overall robustness and reliability in crack detection. Overall, the higher accuracy, recall, F1 Score, and MCC of FusedNN compared to Zoubir highlight its superior performance and effectiveness in detecting and classifying cracks. The use of fused images, which involves selecting the maximum intensity pixel value from a directionally lit dataset as shown in Figure 7.1, improved crack visibility within the FusedNN model. This highlights the potential of directional lighting, thereby strengthening FusedNN's ability to accurately detect and classify cracks.

In the comparison between FusedNN and MCNet for crack classification, MCNet consistently demonstrated superior performance across various metrics. MCNet achieved an accuracy of 96.6%, surpassing FusedNN by 3.6%, indicating its ability to classify a higher percentage of cracks correctly. Moreover, MCNet exhibited higher precision at 96.3%, outperforming FusedNN by 4.3%, which signifies its lower rate of false positives and better precision in identifying actual cracks. In terms of recall, MCNet achieved 97.0%, which is 3% higher than FusedNN's recall of 94.0%, highlights MCNet's superior capability to detect more true positive cracks. Additionally, MCNet achieved a higher F1 score of 96.6%, compared to FusedNN's F1 score of 93%, indicating that

MCNet showed a better overall balance between precision and recall. When considering the MCC, which measures the performance of model accounting for true and false positives and negatives, MCNet significantly outperformed FusedNN by 7.3%. While both FusedNN and MCNet utilise directional lighting, MCNet adopts a multi-channel neural network approach that incorporates features from all directional images to enhance crack detection. The confusion matrices of all the models is shown in Figure 7.6.

Figure 7.7 shows qualitative results obtained when four trained models are tested on various samples. When the crack is relatively thick and runs through the centre (Sample 1), all models successfully classified it with high confidence. However, when the crack is located near the extremities, such as along the edges (Sample 4 in Figure 7.7), all models failed to detect it. When the crack is near but not at the edges (Sample 3), both MCNet and FusedNN models were able to detect it. This suggests that directional lighting casts shadows in the defects, and help the models in detecting cracks more effectively compared to the traditional and Zoubir's models. On uneven surfaces where the background is unclear, such as when algae are present (Sample 2), all models except MCNet mistakenly identified the algae as cracks. Additionally, when the sample has algae and the crack is slightly wide and runs through the centre (Sample 5), all models except the traditional model correctly identified the crack. From Figure 7.7 it is shown that MCNet which utilises directionally lit images outperformed the other three models, demonstrating the significance of directional lighting for crack detection on Civil infrastructure. Although the FusedNN model incorporates directional lighting, it relies solely on the pixels with maximum intensity values to form the final image i.e, it only selects one pixel value out of all the five images. In contrast, MCNet utilises each channel to represent gray scale images captured in R, D, L, U, and A, allowing MCNet to learn and extract additional features from all the five images more effectively than FusedNN.

Figure 7.6: Confusion matrices of Traditional MobileNetV2, Zoubir, FusedNN, and MCNet models for binary crack classification models.

Samples	Image Scene (224 x 224 pixels)	Traditional	Zoubir	FusedNN	MCNet	Actual
Sample 1		Crack	Crack	Crack	Crack	Crack
Sample 2		Crack	Crack	Crack	No Crack	No Crack
Sample 3		No Crack	No Crack	Crack	Crack	Crack
Sample 4		No Crack	No Crack	No Crack	No Crack	Crack
Sample 5		No Crack	Crack	Crack	Crack	Crack

Figure 7.7: Model comparison for binary image classification on various samples.

7.5.2 Comparative analysis of VGG-16 five-channel and MCNet models

Based on the above comparison, the top-performing MCNet model is compared with the best-performing VGG-16 five-channel neural network model implemented in Chapter 6. Figure 7.8 shows that MCNet demonstrated superior performance compared to the VGG-16 Five-Channel model across most metrics. MCNet outperformed the VGG-16 five-channel model in accuracy by 1.5%. Both models exhibited identical precision at 96.3%, showing equal effectiveness in minimising false positives. However, MCNet significantly outperformed VGG-16 five-channel in recall by 2.8%, indicating MCNet's efficiency in detecting true positives i.e., cracks. Furthermore, MCNet outperformed the VGG-16 five-channel model by a 1.4% increase in F1 score, indicating a more balance between precision and recall. MCC further highlighted MCNet's superiority, with MCNet achieving a score of 93.3%, while the VGG-16 Five-Channel model scored 90.3%, indicating a 3% improvement. These metrics highlights the enhanced capability of MCNet in accurately identifying cracks, making it a more effective tool for applications, where detection of crack is crucial.
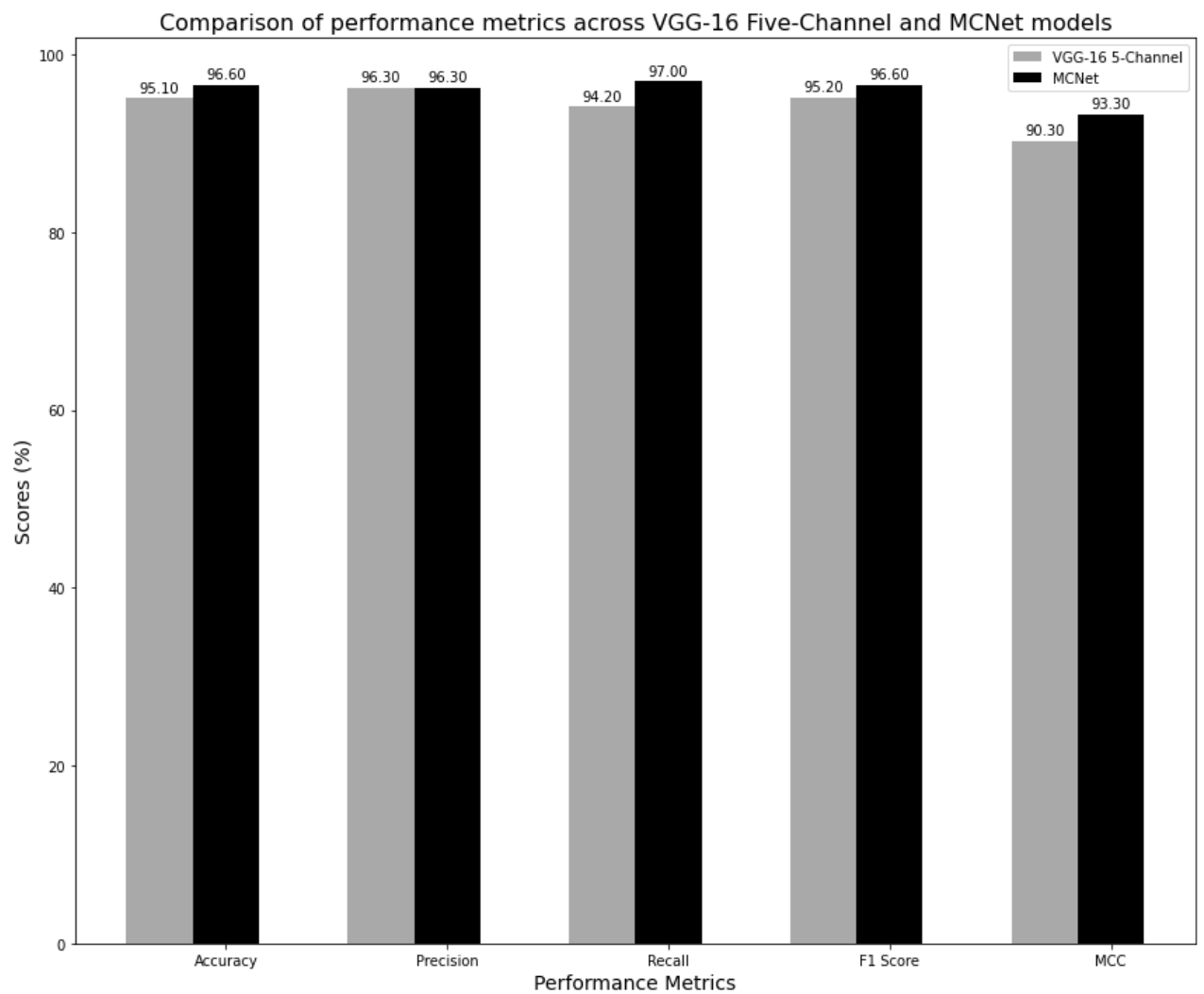
Figure 7.8: Comparative Analysis of VGG16 and MobileNetV2 five-channel models.
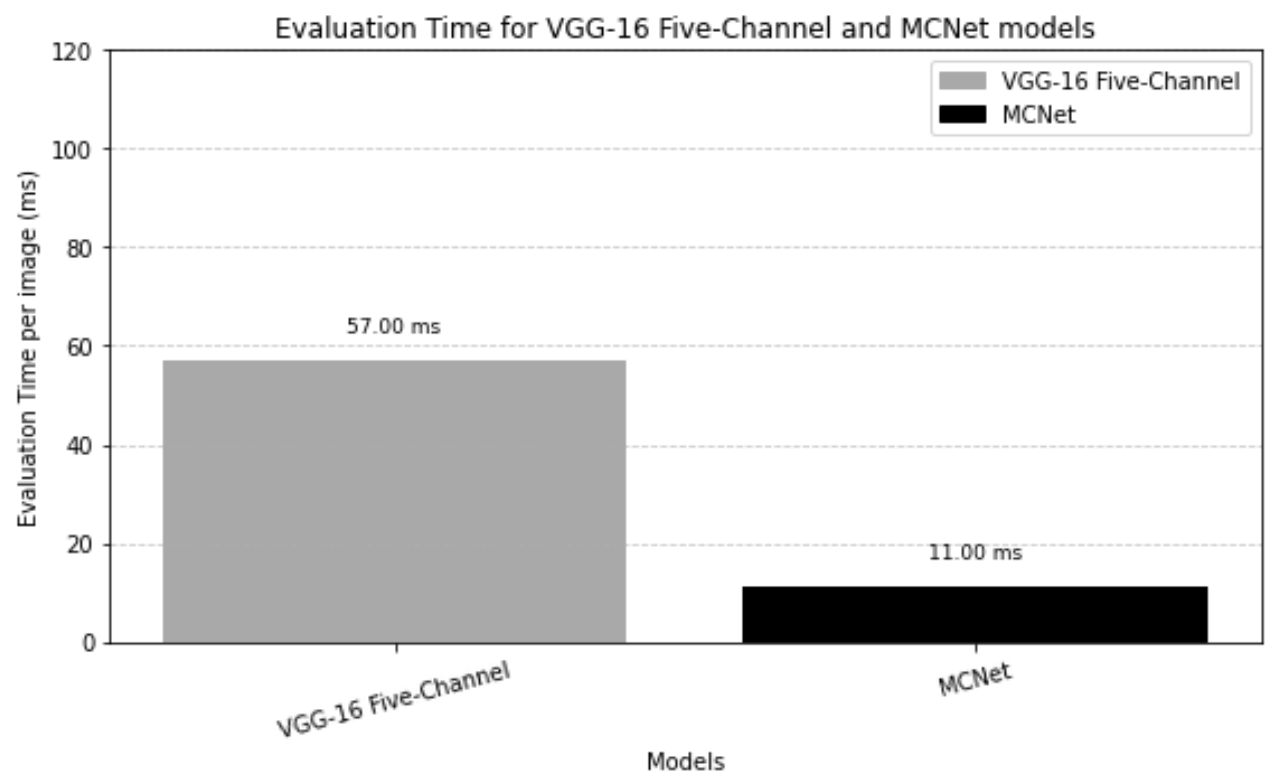
Figure 7.9: Comparison of evaluation times for VGG-16 and MobileNetV2 five-channel crack classification models.

The evaluation times for the VGG-16 Five-Channel and MCNet models, as illustrated in Figure 7.9, highlight a significant advantage of MCNet in terms of processing speed. The VGG-16 Five-Channel model requires 57 milliseconds to evaluate an image, whereas MCNet only needs 11 milliseconds, making MCNet approximately five times faster. Such faster evaluation times are crucial for tasks like concrete structure inspections, where timely and accurate crack detection is essential for maintaining structural integrity and safety. MCNet's ability to deliver fast and reliable results makes it a superior choice for these critical applications in the field of Civil Engineering.

7.5.3 Multi-class Image Classification Metrics

In Figure 7.10, the multi-class image classification evaluation metrics show that accuracy, precision, recall, F1 score, and MCC are almost similar for the different methods, with MCNet performing slightly better, supporting the argument that more advanced models can improve concrete defect detection.

A series of heatmaps are used to visualise these metrics across different classes as shown in Figure 7.11. The heatmap rows depict the models - Traditional MobileNetV2, Zoubir's model, FusedNN, and MCNet models, while the columns represent the target classes - thin cracks, thick cracks, spalling, and none. The colour shades in each cell indicate the magnitude of the associated metric. Lighter shades indicate lower values, while darker shades represent higher values. The evaluation of various defect detection models using multiple performance metrics shows the advantages of the these improved models, MCNet and FusedNN, over the Traditional and Zoubir model.

In the accuracy heatmap, MCNet and FusedNN demonstrated similar performance for detecting thin cracks. However, MCNet exceeded in detecting thick cracks, and achieved highest accuracy of 92%. Notably, MCNet achieved 100% accuracy in identifying spalling. FusedNN outperformed both the traditional model and Zoubir's model in detecting thin cracks. The performance of all models were comparable for detecting thick cracks and spalling. However, FusedNN consistently performed better than the traditional model and Zoubir's model. Zoubir's model and the traditional model exhibited similar metrics, with the traditional model slightly trailing behind.
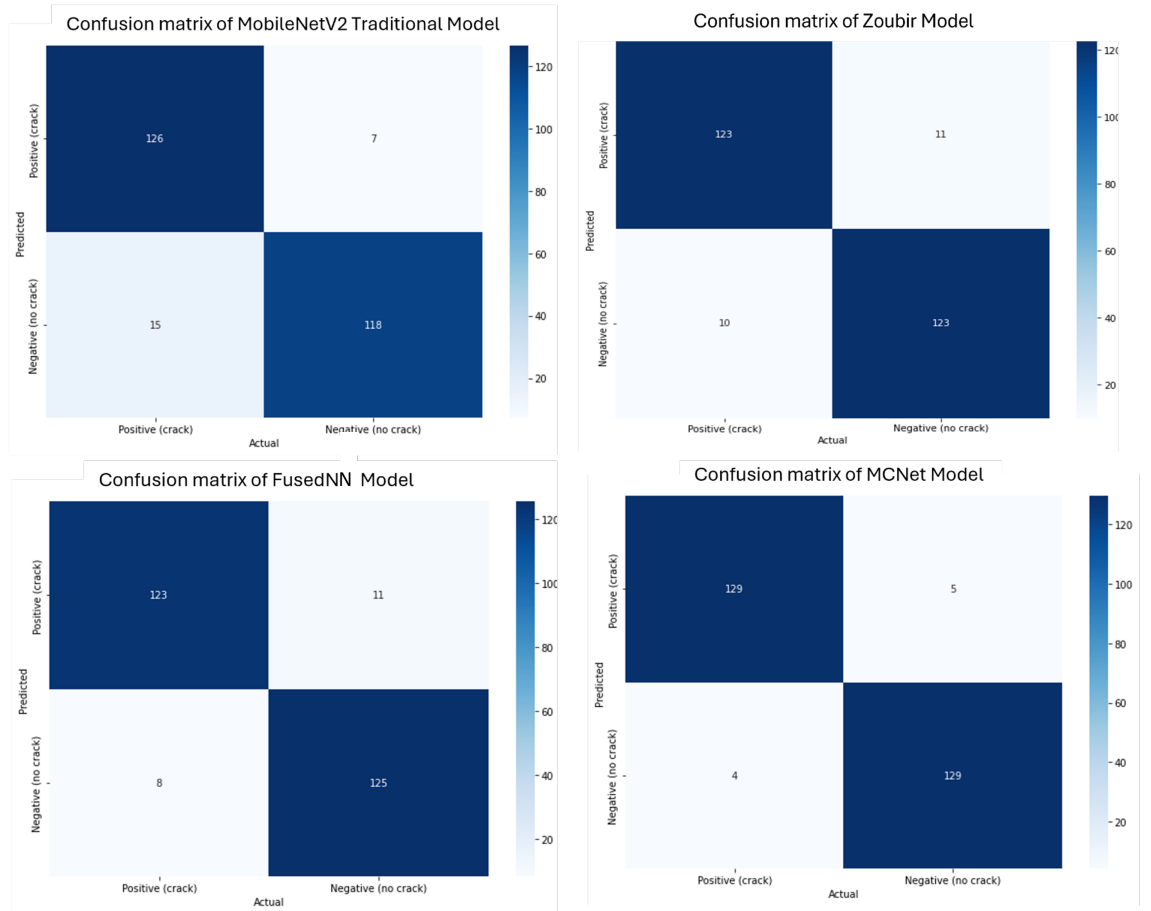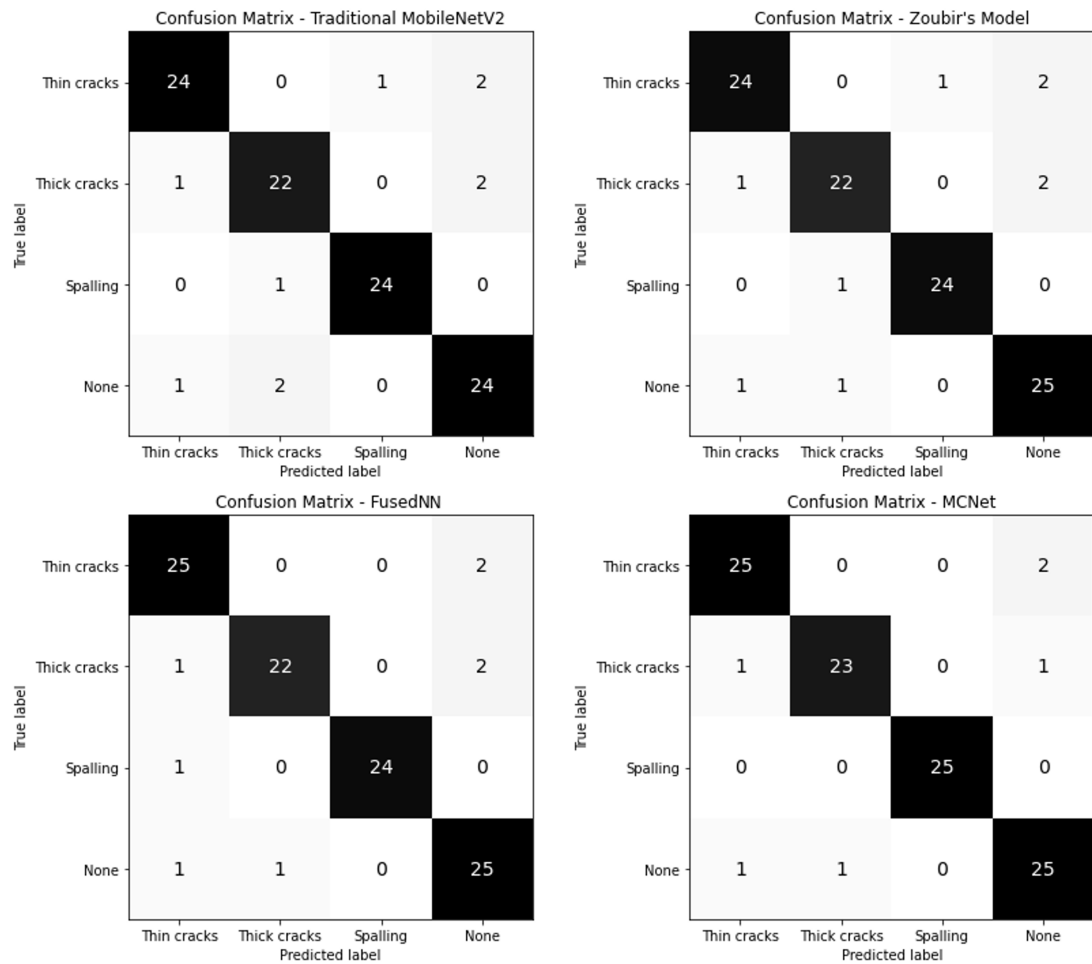
Figure 7.10: Confusion matrices of Traditional MobileNetV2, Zoubir, FusedNN, and MCNet model for multi-class image classification.
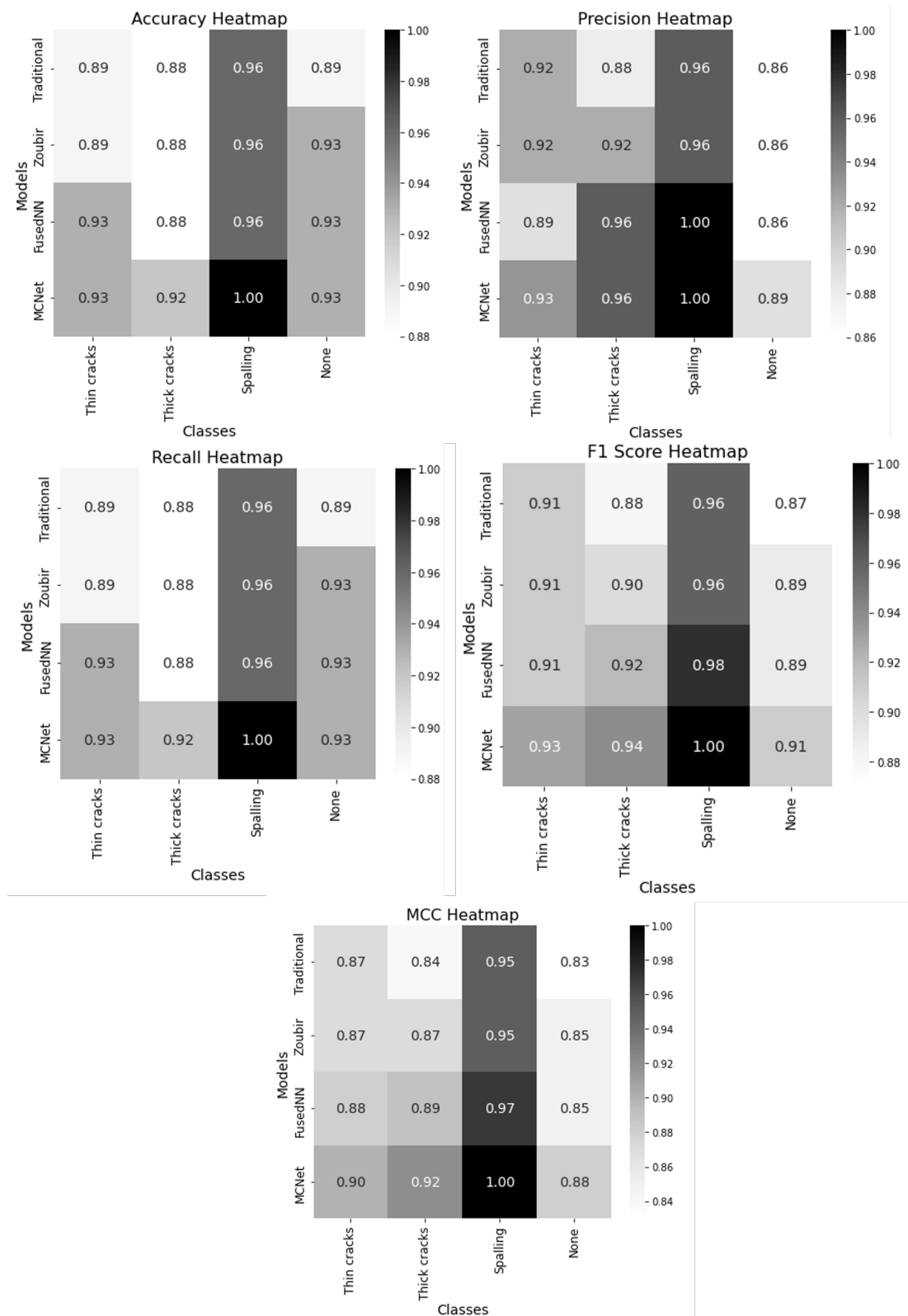
Figure 7.11: Heatmaps for Traditional MobileNetV2 vs Zoubir vs FusedNN vs MCNet models for multi-class image classification

In terms of precision, both MCNet and FusedNN exhibited superior performance in detecting thick cracks, spalling, and the absence of defects compared to the Traditional model and Zoubir's Model. For thin cracks, MCNet achieved an identical precision of 93%, while FusedNN is slightly lower at 89%. Zoubir's Model falls between the Traditional and the proposed models, showing intermediate precision values.

The recall heatmap demonstrates that MCNet achieves the highest recall across most defect categories, indicating its effectiveness in identifying true positives and minimizing missed defects. MCNet and FusedNN both outperformed the traditional model and Zoubir's model in detecting thin and thick cracks, with MCNet showing exceptional recall for spalling. FusedNN also shows high recall values but slightly lags behind MCNet. Zoubir's Model, while performing well, does not match the proposed models in overall recall. FusedNN also performed competitively, particularly for thick cracks and spalling, but with slightly lower scores compared to MCNet. Both the proposed models has shown better F1 scores than the three-channel models.

The MCC heatmap indicates that MCNet achieves the highest MCC scores across most defect categories, highlighting its robustness in predicting both true positives and true negatives while minimizing false positives and false negatives. FusedNN also performed well, especially for thick cracks and spalling, but slightly trails behind MCNet. The traditional model and Zoubir's model has shown moderate MCC scores, with Zoubir's model generally positioned between the traditional model and the advanced models.

Across all metrics, MCNet and FusedNN consistently outperformed the traditional model and Zoubir's model, with MCNet often leading in performance. MCNet is particularly strong in detecting specific types of defects, such as spalling, making it the preferred choice for applications requiring high accuracy in defect detection. For balanced performance across various defect types, both MCNet and FusedNN are suitable, though MCNet holds a slight edge due to its superior performance across multiple metrics. Zoubir's model performs well but does too general and did not reach the performance scores of the advanced models.

Figure 7.12 shows qualitative results obtained when four trained models are tested

on various samples (described in Chapter 4) to detect cracks and spalling. When the crack is located near the extremities, such as along the edges (Sample 1), all models failed to detect it. Both proposed models detected the hairline crack passing almost through the centre (Sample 2). When the crack is thick and passes through the centre (Sample 3), both proposed models and Zoubir's model correctly detected it as thick crack, unlike the traditional model which misclassified it as a thin crack.

Both the proposed models identified spalling in Sample 4. The traditional and Zoubir's models mistakenly identified Sample 5 as spalling. Overall, it is evident that MCNet outperforms the other three models in detecting both cracks and spalling. This superior performance is due to the light being projected onto the surface from all directions, which enhances the visibility of crack extremities and thereby improves defect detection in civil engineering applications. For applications requiring high accuracy in detecting specific types of defects, especially spalling, MCNet would be the preferred choice. However, for a balanced performance across various defect types, both MCNet and FusedNN are suitable, with MCNet having a slight better performance.

The five-channel model demonstrated superior performance compared to FusedNN, but the FusedNN model still outperformed both traditional and advanced models in the literature, such as the Zoubir model, for both binary and multi-class image classification tasks. Therefore, FusedNN can be effectively applied in scenarios where standard neural network models are typically used, particularly in cases where the format of the model input cannot be adapted to accommodate non-traditional inputs like five-channel data.

7.5.4 Effect of exposure on Diffused and Fused Images

7.5.4.1 Analyse the performance of Diffused Images captured under different exposure values

The traditional MobileNetV2 model uses diffused images, while FusedNN creates a fused image by selecting the maximum intensity pixel values from the directional images. So far, the analysis has been conducted under auto exposure settings.

If the images are captured under increased exposure conditions, the resulting images

Samples	Image Scene (224 x 224 pixels)	Traditional	Zoubir	FusedNN	MCNet	Actual
Sample 1		None	None	None	None	Thin Crack
Sample 2		None	None	Thin Crack	Thin Crack	Thin Crack
Sample 3		Thin Crack	Thick Crack	Thick Crack	Thick Crack	Thick Crack
Sample 4		None	None	Spalling	Spalling	Spalling
Sample 5		Spalling	Spalling	None	None	None

Figure 7.12: Model comparison for multi-class image classification on various samples.

will display brighter pixels. Since the FusedNN model selects the maximum intensity pixel values, the pixels from the directional images will also have high exposure values. At a certain point, even though FusedNN selects the maximum intensity pixels, the fused image and the diffused image could become identical due to the uniformly high exposure. If this occurs, the advantage of using FusedNN diminishes, as there would be no significant difference between the fused image and the diffused image. Thus, the potential benefit of FusedNN in enhancing defect detection would be negated under such high exposure conditions.

This scenario highlights the importance of considering exposure settings in evaluating the true efficiency of FusedNN compared to traditional methods. For this reason, an evaluation was conducted to compare the performance of the traditional MobileNetV2 and FusedNN models under different exposure conditions. Various performance metrics were utilised to assess the performance of the model, including accuracy, precision, recall, F1 score, and MCC.

Images of two concrete slabs were captured using auto exposure settings, starting with an initial value of 470,000 lux-seconds. To create underexposed and overexposed

Chapter 7. Directional Lighting Enhanced Crack and Spalling Detection in Concrete Structures

conditions, the exposure value was adjusted between 200,000 and 900,000 lux-seconds. This process was repeated to capture directional images (R, D, L, U), all at an optimal angle of 50 degrees. A stratified five-fold cross-validation technique was then employed for binary classification of these images, training and testing the model on datasets with exposure values ranging from 200,000 to 900,000 lux-seconds. The objective was to identify the optimal exposure value.

Table 7.3: Performance Metrics at Different Exposure Levels.

Exposure Values (in lux-seconds)	Accuracy	F1 Score	MCC
200,000	89.3	88.8	79.4
300,000	90.6	90	81.6
400,000	92	91.9	84.4
470,000 (Auto-Exposure)	92.36	92	85
500,000	92.5	92.2	85.4
600,000	**92.6**	**92.5**	**85.5**
700,000	92.2	91.83	84.9
800,000	91.53	91.3	83.6
900,000	90.3	89.5	81.2

The results of the evaluation as shown in Table 7.3 show that the traditional Mobile-Net model consistently achieved high accuracy across different exposure values, ranging from 89.3% to 92.6%, with the highest accuracy of 92.6% being achieved at an exposure value of 600,000 lux-seconds. The F1 scores, representing the harmonic mean of precision and recall, followed a similar trend, with the highest F1 score of 92.5% observed at the optimal exposure value of 600,000 lux-seconds. Moreover, the MCC values, assessing the quality of binary classifications, maintained a consistent pattern, with the peak MCC of 85.5% achieved at an exposure value of 600,000 lux-seconds. Based on these performance metrics, it can be concluded that an exposure value of 600,000 lux-seconds is the optimal setting for this model when utilising diffused images. The ability of the model to differentiate between cracked and uncracked samples under different lighting conditions is highlighted by its strong performance at this exposure value.
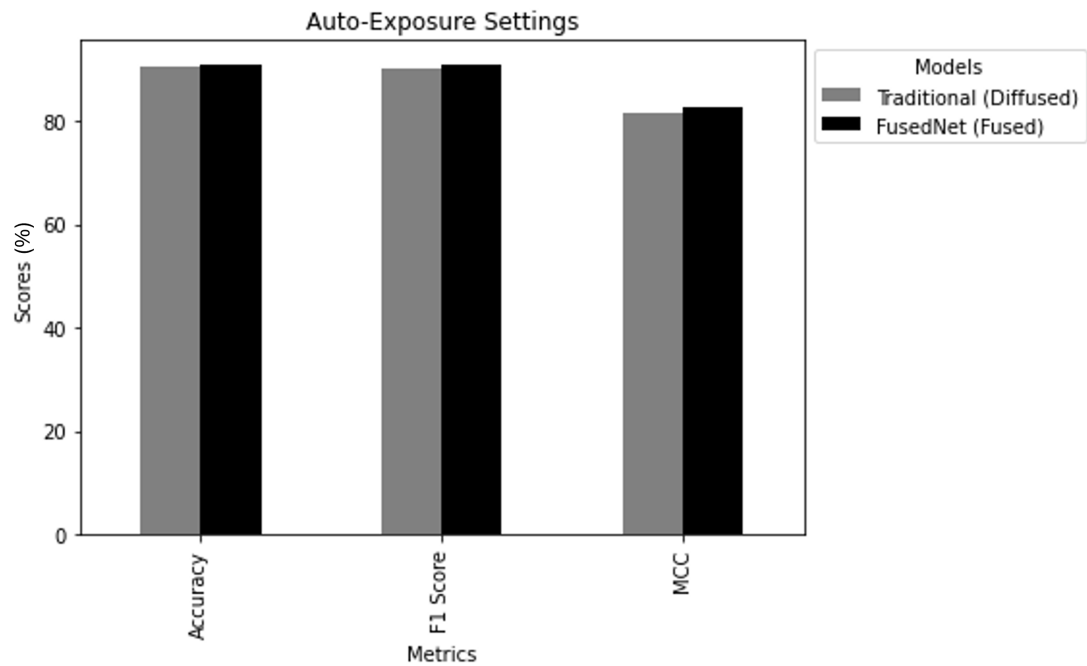
Figure 7.13: Comparison of model performance in auto-exposure settings.
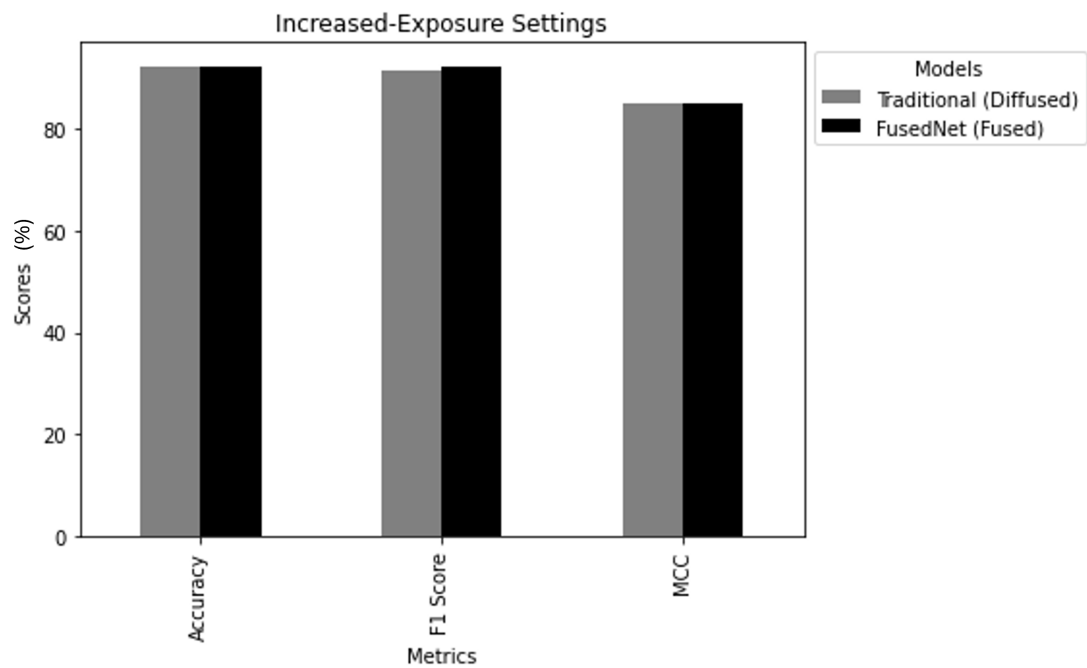
Figure 7.14: Comparison of model performance in increased-exposure settings.

7.5.4.2 Evaluation of Fused and Diffused Images

After obtaining the optimal exposure value, the aim was to explore the potential of fused images. To achieve this, images were captured under different lighting directions, each sharing similar exposure levels. These values were as follows:

- Right Direction: 1.63 million lux-seconds

- Down Direction: 1.65 million lux-seconds

- Left Direction: 1.68 million lux-seconds

- Up Direction: 1.71 million lux-seconds

These images were combined with the diffused image captured at the optimal exposure value of 600,000 lux-seconds to create a fused image. The main objective was to evaluate the performance of these fused images against the diffused image and to study the effect of increased exposure on performance of the model. The research comprised two distinctive scenarios to assess the performance of the model:

i. Auto-Exposure Comparison: In this scenario, traditional MobileNetV2 and FusedNN models were trained and tested using the auto-exposure diffused image and the auto-exposure fused image, respectively. The aim was to understand how well these models performed using images captured at their auto-exposure values. Figure 7.13 draws the following conclusions:

- Accuracy: FusedNN achieved an accuracy of 91.1%, exceeding the traditional model's accuracy of 90.4%.

- F1Score: The F1 score was 91% FusedNN, exceeding traditional model's F1 score of 90.24%.

- MCC: Finally, FusedNN achieved higher MCC of 82.7% compared to the traditional model's MCC of 81.77%.

This indicates that FusedNN had a superior ability to make accurate binary classifications under auto-exposure conditions.

ii. Increased Exposure Comparison: The second scenario focused on the impact of increased exposure on performance of the model. Here, the performance of the MobileNetV2 and FusedNN models was evaluated when trained and tested on diffused and fused images, respectively, under increased exposure conditions. The aim is to understand how increased exposure levels affect the performance of the model for crack classification. Figure 7.14 draws the following conclusions:

- Accuracy: The traditional model achieved an accuracy of 92.1%, while FusedNN reached an accuracy of 92.22%. Both models performed well, with FusedNN maintaining a slight edge.

- F1Score: The F1 score reflects the balance between precision and recall. The traditional model achieved an F1 score of 91.53, while FusedNN obtained high F1 score of 92.3.

- Matthews Correlation Coefficient: The traditional model reached an MCC of 85, while FusedNN achieved an MCC of 85, indicating strong binary classification capabilities.

In the increased exposure scenario, FusedNN still demonstrated superior performance compared to the traditional MobileNetV2 model. This suggests that FusedNN is more reliable and robust in accurately identifying cracked samples, regardless of the exposure conditions.

The study showed that the FusedNN model using fused images (directional images) was effective in auto-exposure and increased exposure settings compared to the traditional model using diffused images alone. This highlights the adaptability and robustness of the fused image approach. This indicates that the dependence of the FusedNN model on maximum-intensity pixel values remains effective under varying exposure settings. In conclusion, this study highlights the potential of the fused image

technique, which outperforms traditional diffused images and maintains its effectiveness under modified exposure levels.
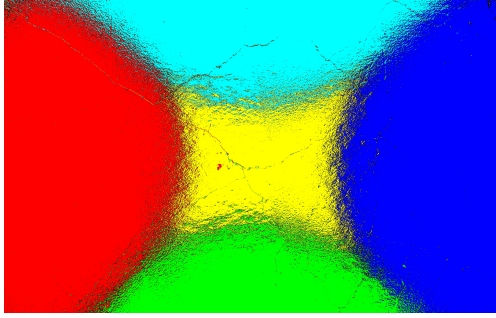
(a) Fused Image at auto-exposure
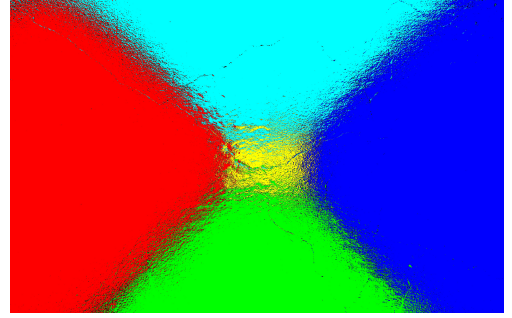
(b) Fused Image at increased-exposure

Figure 7.15: Sample 1: Comparison of fused images generated using directional images captured at auto-exposure and increased-exposure settings.
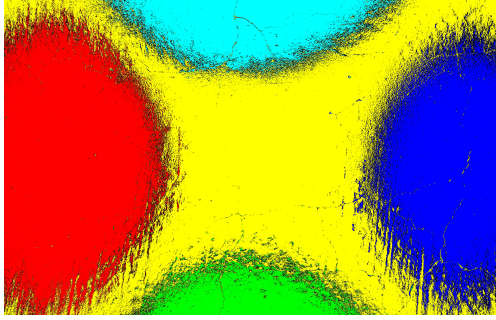
(a) Fused Image at auto-exposure
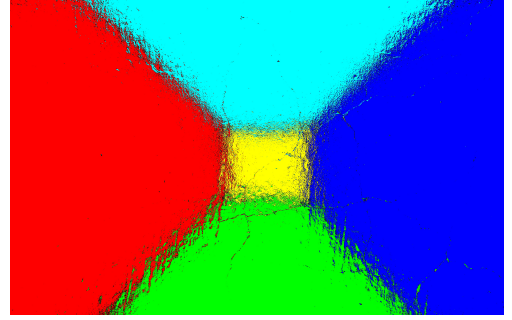
(b) Fused Image at increased-exposure

Figure 7.16: Sample 2: Comparison of fused images generated using directional images captured at auto-exposure and increased-exposure settings.

Figure 7.15 and Figure 7.16 represent the heatmaps of the fused images generated using images captured at auto-exposure and increased-exposure settings for samples - 1 and 2. Even with higher exposure levels, the fused image still depended on selecting the maximum intensity pixels from the directional images. This approach helped in extracting valuable information from varying lighting conditions, leading to enhanced visibility of cracks and an improved classification rate, as shown in Figure 7.13 and Figure 7.14.

7.5.4.3 Comparative Analysis of Traditional MobileNetV2 and FusedNN models under Different Exposure Settings

The comparative analysis of the Traditional MobileNetV2 and FusedNN models, as shown in Figure 7.17, clearly demonstrates the advantage of FusedNN over the Traditional MobileNetV2 model. The X-axis represents various exposure values, including both auto-exposure and increased exposure settings, along with the corresponding F1 scores on Y-axis for both models. The best fit lines indicate the trend in their F1 scores, indicating that as the exposure value increases, the performance of both models also increases. However, FusedNN consistently outperformed Traditional MobileNetV2 across all exposure settings.
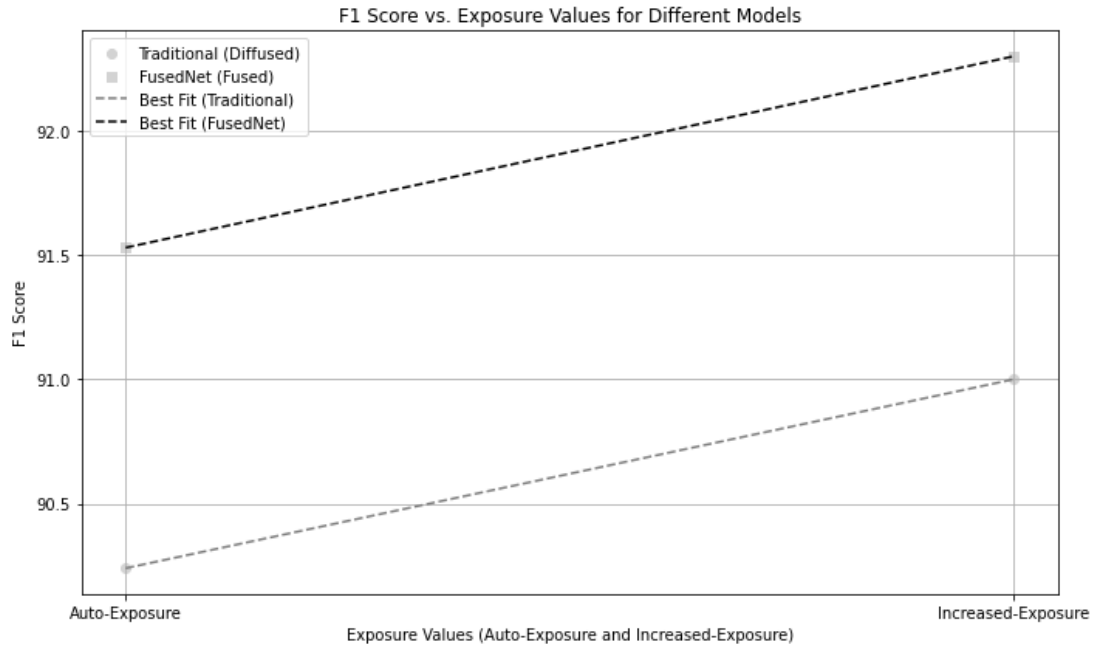
Figure 7.17: Comparative Performance of Traditional MobileNetV2 and FusedNN Models under Auto and Increased Exposure Settings.

The improved performance of FusedNN is due to its novel approach of selecting the maximum intensity pixels, which proves to be advantageous compared to Traditional MobileNetV2 model. In summary, increasing the exposure improves the performance of vision models, particularly in terms of the F1 score. However, FusedNN outperforms Traditional MobileNetV2 as a more efficient model across different exposure settings.

This makes FusedNN an ideal choice for a wide range of applications in the field of Civil Engineering.

7.5.4.4 Examining the Effects of Auto-Exposure on Machine Vision

Auto-exposure is an algorithm that creates brighter images by adapting camera settings like aperture, shutter speed, and ISO. Generated image are designed with human viewers in mind, which can sometimes lead to conflict with machine vision systems need for accurate image data. Under sub-optimal lighting conditions—including low light, uneven light, and back-lighting—auto-exposure can cause severe distortions and degrade the image quality. For example, uneven lighting can result in both overexposed and underexposed regions, impacting human visual perception as well as image recognition capabilities. In the case of high-contrast scenes, auto-exposure algorithms often face challenges as they struggle or try to achieve a balanced exposure throughout the image. This can lead to overexposed shadows and underexposed bright areas. Such imbalance can be harmful when trying to detect features such as cracks in concrete structures [223]. FusedNN, by utilising maximum-intensity pixels in fused images, exceeds auto-exposure algorithms in multiple scenarios. This method maintains the ability of the model to detect fine details, even in challenging lighting environments.

7.5.4.5 Advantages of Fused Image Technique under Varying Exposure Conditions

The findings highlight the strong performance and reliability of the FusedNN model utilising fused image technique when faced with varying exposure conditions. By utilising maximum-intensity pixel values, FusedNN is able to effectively handle different levels of exposure while maintaining high accuracy in distinguishing between cracked and uncracked samples. This approach guarantees that crucial crack details are preserved, ensuring accurate identification of cracks. The adaptability of the fused image technique is further highlighted in these findings. It efficiently handles variations in lighting conditions, demonstrating that FusedNN maintains its effectiveness even when exposed to different illumination levels. This highlights the potential of the fused image

technique, exceeding traditional diffused images under various exposure settings.

7.6 Summary

The methods, FusedNN and MCNet, utilise images captured under right, down, left, up, and diffused directions, respectively. The focus was to improve the time required in automated crack and spalling inspections in concrete structures under challenging lighting conditions by utilising directional lighting.

Stratified five-fold cross validation, hyperparameter tuning and regularization techniques were utilised to avoid overfitting and generalise well on new and unseen data. These models are evaluated both on images captured in both laboratory and real world. In binary crack classification, MCNet outperformed FusedNN by 3.6%, Zoubir's model by 4.5% and traditional model by 4.9%. Similarly, for multi class image classification, MCNet outperformed all the models and achieved highest accuracy in detecting thin cracks, thick cracks, and spalling, respectively. The proposed models are capable of detecting cracks of width as small as 0.1mm. The MCNet model evaluates images five times faster than the VGG-16 five-channel model. These neural network models can be used to detect defects such as cracks and spalling in real-world concrete structures. Also, the proposed models achieve improved performance without a significant increase in evaluation time.

In conclusion, this Chapter highlights the potential of directional lighting techniques for accurate defect detection in concrete structures. It was the first study to evaluate the performance of directional lighting based deep learning models on real world concrete structures. These models offer a valuable solution for automated infrastructure assessment, capable of identifying cracks as narrow as 0.1mm.

Chapter 8

Discussion and Future Work

8.1 Discussion on Research Findings

The aim of this research was to investigate if multi-directional lighting could enhance the detection of cracks in concrete by comparing directional lighting images to diffused lighting images using the same algorithm. The investigation took into account multiple directions of light, to simulate an approach followed by many experienced inspectors in order to enhance manual crack detection and thus supporting the use of directional lighting. Instead of using individual photos per light direction, this research went a step further by combining them all and developed algorithms to do so automatically. The final goal was to innovate how civil inspectors use convolutional neural networks for concrete crack detection. This was achieved by (1) modifying the inputs to CNNs (5 images instead of 3), (2) adjusting those CNNs to accept the new, increased number inputs and (3) demonstrating the benefits of doing so. The choice of which CNN to use is arbitrary. The results presented in this thesis show that images captured with multi-directional lighting can indeed enhance the detection of thin cracks in concrete, while neural networks adapted to achieve the analyses of 5-channel input data remain time and cost-effective.

Eurocode 2 specifies that maximum allowable crack widths for regular reinforced and prestressed concrete structures range between 0.2mm and 0.4mm. Past research indicated that proper illumination is vital for accurate detection of concrete defects,

especially in low-light situations. However, with uniform lighting, the detection of hairline cracks often fails. Most studies have utilised lighting only to improve visibility in low-light conditions, and overlooked the potential advantage that directional lighting can offer for detection of concrete defects on Civil infrastructure. To address this issue, this thesis introduced the implementation of ALICS, the first adaptive lighting platform designed to capture images automatically in multi-directional and multi-angle configurations, enabling the detection of cracks as thin as 0.1mm, exceeding the requirements by the Eurocode 2. The author is one of the developers of ALICS, having developed the very first and second prototypes and heavily contributed to the development of the current ALICS version. ALICS testing extended beyond laboratory experiments, and was successfully utilised in real-world scenarios. From the reviewed international literature, there is no other existing hardware capable of providing images with multi-directional, multi-angle lighting. ALICS was developed to address specifically this technology gap. Since this was a first in its kind, there is no hardware benchmark to evaluate ALICS against. However, the effectiveness of the hardware was manually assessed in the lab and later in applications outdoors as detailed in Chapter 4. The author conducted in-situ tests at the Carnbooth and Babcock sites, and these are included in this thesis. Additional tests using the ALICS hardware were performed by other researchers and are therefore not detailed here.

In this study, the results from the crack detection models were benchmarked against human-defined ground truths i.e., annotated labels (crack/no crack). Although these ground truths may not represent the state-of-the-art, they closely resemble human inspector reports, which are considered the state-of-the-art in visual inspection. The work in this thesis has improved visual inspections of concrete civil infrastructure by providing an automated, faster and reliable solution.

Automated inspections often struggle to detect defects under low-light or challenging environmental conditions. It has been shown in the literature that higher quality images can enhance the performance of neural network models. When using the ALICS device for automatic image capture, there are instances where low-quality images may still be recorded. Rather than relying solely on robust neural network algorithms, such

as noise removal or blur removal algorithms to address this issue, the author proposed a threshold-based BRISQUE-assisted data cleaning process. This method involved setting a threshold value—if an image falls below this threshold, the device re-captures another image ensuring that only high-quality images are forwarded for image-processing. The goal was to focus on capturing quality images initially, rather than trying to improve them during the preprocessing stage. With this algorithm implemented, whenever the ALICS device detected an image that did not meet the threshold BRISQUE score, it automatically discards that image and re-captures another one, guaranteeing that only high-quality images were retained. As a result, training the neural network model with these high-quality images led to enhanced model performance. This approach was deemed more favourable as training a model with noisy images could improve its performance on identifying cracks on noisy images but does not necessarily mean it will perform well on high quality images, as has been shown in the reviewed literature.

The BRISQUE-based crack detection model implemented in this study cannot be tested on public datasets such as [224], due to fundamental differences in data acquisition settings. Public datasets are usually captured under well-lit conditions (i.e., daylight settings), while the datasets used in this study are captured under low-light conditions, with illumination projected from the right, down, left, and up directions. This lighting configuration significantly alters the visual characteristics of the images compared to those captured under standard lighting conditions. As a result, direct comparison with publicly available datasets is not valid. Moreover, the main aim of implementing the BRISQUE-based approach is to evaluate the quality of the images captured in dark or low-light environments, and recapture those images that fall below B_T to ensure that only high-quality images are forwarded for image processing.

Novel five-channel VGG-16 and VGG-19 neural network models were developed and their performance was compared against conventional approaches. The VGG model was chosen as it is well-recognised and commonly employed for detecting cracks in reinforced concrete structures. The use of VGG instead of SOTA models and later the comparison of the five-channel VGG against a 3-channel model do not restrain, limit or invalidate the results of this study, although it was demonstrated in Chapter 3 that

VGG model is still commonly used in research studies today. The choice of the CNN used is arbitrary. The outcome of this research and novelty is the methodology for the incorporation of additional channels in the model, it is not about the actual NN model used. The five-channel models demonstrated a reduction in image evaluation time compared to traditional models used for crack detection in concrete, particularly in low-light environments, thus pushing the discipline boundaries of developed technology further. This thesis developed and fully described an approach that can be modified for any neural network model, using the same methodology utilised to create the five-channel model, which is applicable or can be extended to a wide range of existing models in the literature. Furthermore, this approach can be expanded to n-channel models, if one would like to add additional lighting directions, following the same procedures and implementation strategies.

The evaluation time was further reduced by using MobileNetV2. As mentioned previously, the choice of the model is not what makes this research unique and is not the focus of the study. One could argue that instead of MobileNetV2, the MobileNetV3 or newer versions could have been used. This is true, but it would not change the final goal: whatever the model, it still uses uniform /diffused lighting. By incorporating directional lighting into the model, the crack detection performance increases with less evaluation time. The MobileNetV2 model was not used in the previous, earlier steps of the research because, given the total lack of detection algorithms that worked with more than 3 channels, it was deemed necessary to first start with a well-known, commonly used network, i.e. the VGG.

Later, FusedNN and MCNet were developed for binary and multi-class image classifications tasks, and a comparison between them was provided. Both models included a five-channel approach similar to the VGG-n models, but the FusedNN specifically applied the maximum intensity fusion technique. The MCNet and FusedNN models were evaluated for both binary and multi-class image classification tasks to detect both cracks and spalling in concrete structures. In the binary classification scenario, MCNet outperformed all other models, followed closely by FusedNN. For multi-class image classification, MCNet also showed the best performance. Since FusedNN selects

pixels based on maximum brightness, increasing exposure could result in selected pixels matching those in diffused images captured with higher exposure rates. To investigate this, the author compared performance of both models under increased exposure conditions and found that the fused images still provided advantages over traditional models. Despite the fact that the 5-channel VGG model is outperformed by FusedNN and MCNet, it can still be used for benchmarking future research that incorporates multi-channel approaches given that VGG is still a benchmark model.

Extensive measures were taken to ensure a fair comparison between FusedNN and MCNet models. Although FusedNN and MCNet models demonstrated superior performance, they are more complex than traditional three-channel models and other existing models in literature due to their use of directional images to generate fused and five-channel images. The traditional model uses a single diffused image, while FusedNN and MCNet models consider images from five different directions. Although the input data is different it is still adequate to demonstrate that directional lighting can improve accuracy in real-world scenarios and as such, provide a novel approach that can be integrated into any neural network model.

The Zoubir model outperformed the traditional three-channel MobileNetV2 model; however, it did not surpass the performance of the two novel models, MCNet and FusedNN. This is because Zoubir model utilised diffused RGB images similar to traditional models, and all existing models in the literature. Therefore, the additional information provided by additional channels in FusedNN and MCNet models offers an advantage in extracting more features, providing more information, and there by improving the performance.

The results in this thesis highlighted the potential of directional lighting techniques for accurate defect detection in concrete structures. This research is the first study to develop hardware for and the performance of directional lighting based deep learning models on real world concrete structures. These models offer a solution for automated infrastructure assessment, capable of identifying cracks as narrow as 0.1 mm, much lower than the required abilities set by the Eurocode 2.

8.2 Future Study

The methodologies and results from this research can be applied to various image-based classification problems beyond concrete crack detection, such as autonomous driving and surveillance systems.

Future studies should explore the integration of drones for automated inspections as a potential alternative to the current Universal Robots method. Drones offer greater adaptability, especially when assessing large or complex structures and hard-to-reach areas. The integration of hardware platforms in autonomous navigation systems, allowing real-time analysis of image data, is also an area for future exploration. Optimising computational speed for image processing and analysing multi-modal data, such as infrared imaging, will enhance defect detection and provide a more understanding of structural integrity.

Research on reducing resource requirements without compromising performance of the model, by exploring model compression techniques or hardware solutions that can efficiently handle larger datasets is recommended. Additionally, a cost-benefit analysis of hardware modifications for capturing images from different directions is necessary to evaluate the practicality and economic feasibility of these changes. Time-series data should be analysed to monitor long-term changes in concrete structures, and algorithms developed to detect and track cracks and spalling. This will help us understand how the structural health evolves over time.

Regarding the hardware system, a potential direction would be to explore the possibility of integrating GPS positioning information to allow for revisiting locations and recapture images if the quality falls below the specified threshold. This iterative approach allows to continuously enhance the capability of the system while effectively addressing real-world challenges. Future investigation could explore the use of the threshold-based BRISQUE-assisted data cleaning approach with SOTA models.

Gaussian noise and Gaussian blur are the most common types of image degradations encountered in real-world imaging scenarios; however, they do not account for all possible factors affecting image quality. Future study should explore all other degrad-

ations such as compression artifacts, salt and pepper noise that impact quality of an image.

The FusedNN and MCNet concepts could be further expanded to identify efflorescence, corrosion, delamination in concrete structures. While the models performed well in challenging lighting conditions; they have not been tested under adverse weather conditions, which could be an area for future investigation. Multi-channel and fused neural networks can be applied to SOTA models that primarily use diffused lighting or three-channel images. Applying multi-channel techniques to these models can help in studying their performance and serve as a potential area for future research. Looking into multi-channel and image fusion techniques, including white-box methods, to further increase the performance of the model is also a research direction that should be explored. Additionally, by leveraging more powerful hardware, such as graphics processing units, the performance of the model in developing automated systems for infrastructure assessment and defect detection in Civil Engineering could be further enhanced.

Chapter 9

Conclusions

This thesis implemented Adaptive Lighting for the Inspection of Concrete Structures (ALICS), a novel hardware device that utilises advanced illumination techniques i.e., multi-directional and multi-angle directional lighting, to enhance detection of crack and spalling in reinforced concrete in both laboratory and real-world environments. A first design of a prototype was proposed by the author. This design was subsequently further developed by McAlorum et al. ALICS is mounted onto a Universal Robot, allowing for automatic image capture, while the field-deployable ALICS rigs extend capabilities of ALICS to real-world applications. This setup captures images in multiple directions (right, down, left, up, and diffused) and at various angles (ranging from 10 to 60 degrees), creating an initial dataset for analysis. Using a pre-trained VGG-16 neural network model, it was found that a 50-degree angle consistently provided the best performance metrics. Additionally, the analysis showed that directional lighting has a greater potential for detecting cracks compared to standard diffused lighting, allowing for its use in deep learning models. Since the optimal angle was found to be 50-degrees, so throughout the thesis the images were captured at this angle.

In order to study the existing technology that is prevalent in Civil Engineering, the initial datasets were captured under diffused lighting settings in low-light environments. It was observed that the images are not very clear, as several factors can impact image quality, including lighting conditions, camera settings, motion blur, lens distortion, and noise. Therefore, it was essential to evaluate the quality of the concrete crack images

captured under these conditions, and BRISQUE played a key role in this process. The correlation between image quality, assessed using the BRISQUE algorithm, and the effectiveness of crack inspection algorithms, particularly those utilising the VGG-16 neural network model, was examined. It was found that the sensitivity of BRISQUE increases with higher levels of noise and blur in the images. Consequently, as the noise or blur level increases, the performance of the model declined, consistently showing that a model trained on pristine images performed the best. This decline in the performance of the model highlights the importance of data pre-processing, leading to the implementation of BRISQUE score thresholding technique to identify and discard low-quality images during pre-processing stage. This ensures that only clean, high-quality images are further forwarded for training the neural network model. This automated BRISQUE-based data cleaning technique was applied to real-world images, resulting in a significant improvement in the performance metrics of the model. The threshold-based BRISQUE IQA approach is not restricted to a specific neural network model; however, the author opted for VGG16 due to its widespread use by civil inspectors in various projects. Importantly, this BRISQUE-based data cleaning approach can be incorporated into any SOTA neural network models or other deep learning architectures found in the literature.

The potential of directional lighting was analysed by implementing a novel five-channel VGG-16/19 neural network model, where each channel represented the gray scale version of the image captured in R, D, L, U, and A directions, respectively. The conventional VGG-16/19 model has three-channel diffused RGB images as input. The five-channel model was developed in such a way to accept the five-channel input image, and was trained and tested on five-channel directionally-lit datasets. All the models utilised hyperparameter tuning, regularization techniques and stratified cross-validation approaches to avoid overfitting and to ensure that the models perform well on any unseen data or any data variant from the data they were trained on. The initial comparison between VGG-16/19 traditional and five-channel models demonstrated that the five-channel model consistently outperformed the traditional model across all the combinations of hyper parameters. Additionally, when compared to VGG-16 and

VGG-19 five-channel models, VGG-16 significantly outperformed VGG-19 by 4%-9% in evaluation metrics. The evaluation time of five-channel models is better than that of three-channel models due to the smaller data size of gray scale images used in the five-channel model compared to RGB images in the three-channel model. Smaller data size decreases memory utilisation and speeds up processing times within the model layers. This shows the advantage of using extra channels, with VGG-16 having the shortest evaluation time of all the configurations. These findings underscored the effectiveness of the five-channel model in accurately detecting and classifying cracks of widths as small as 0.1mm.

The maximum intensity fusion technique which has been successful in the field of medical imaging was utilised in this research to develop a novel maximum intensity fused neural network model, called FusedNN, to test its effectiveness in the field of Civil Engineering. Also, the VGG-16 five-channel neural network model implemented earlier was further improved by utilising light neural network model, MobileNetV2 (called as MCNet) to improve evaluation time and was compared with FusedNN, traditional and Zoubir's models. FusedNN uses a pixel-based maximum intensity image fusion technique, where the highest intensity value from the five input images is selected across each pixel to create a fused image. The FusedNN model uses fused images as input and has a structure similar to the traditional MobileNetV2 model. All four models (Traditional, Zoubir, FusedNN, MCNet) were then compared for both binary and multi-class image classification, with MCNet and FusedNN outperforming both the traditional and Zoubir's models, and MCNet being the best. The best-performing VGG-16 five-channel model was compared to MCNet, and MCNet showed a 1.5% to 3% improvement across all metrics while requiring only 20% of the evaluation time needed by the VGG-16 five-channel model. Therefore, by implementing these kind of customised deep learning models, civil engineers can accurately detect potential issues early on, reducing the risk of structural failures and extending the lifespan of Civil infrastructures. The FusedNN model uses a maximum intensity fusion technique, selecting only the brightest pixels. However, if the exposure of the images increases, and the brightest pixels chosen during maximum intensity fusion technique matches the exposure of the diffused image used

in the traditional model, then the advantage of FusedNN could be diminished. To address this, the performance of the traditional model across varying exposure levels was evaluated. It was found that the traditional model performed best at an exposure of 600,000 lux-seconds. So, all the directional images were captured at an exposure similar to that of the diffused image, and the corresponding fused images were generated. A comparison of the traditional and FusedNN models at this exposure level showed that, despite the increased exposure, FusedNN still outperformed the traditional model. This result highlights the adaptability and robustness of FusedNN in accurately identifying cracks, regardless of exposure conditions.

Appendix

A Portable Frame Design

The first iteration of the ALICS system (shown in Figure 1) is designed to be portable and handheld, offering a more flexible and manoeuvrable solution. Unlike the initial design, which relied on complete darkness, this iteration features a frame housing LEDs and employs a "curtain" mechanism to create a controlled lighting environment. Additionally, it has been designed to effectively operate in varying levels of outdoor illuminance.

B Shroud Design

The second iteration of ALICS introduces a lightweight aluminium shroud to address issues related to ambient lighting. This design utilises strips of LEDs to provide the necessary illumination, while the machine vision camera is positioned at the apex of the shroud, directed toward the target surface, as shown in Figure 2 and Figure 3. The shroud effectively blocks out ambient light, ensuring that the system can operate optimally in dark or low-light outdoor environments.

Both iterations of ALICS can function as handheld devices, allowing inspectors to conduct on-the-spot inspections, or can be mounted to a six-axis robot for controlled and precise movements during data capture. Furthermore, image processing can be carried out onboard or on an attached single-board computer, providing immediate analysis capabilities, or data can be streamed to external computing resources for offline processing.

Figure 1: First Iteration of Handheld and Portable ALICS Hardware. Captures images of size 5429×3458 pixels.
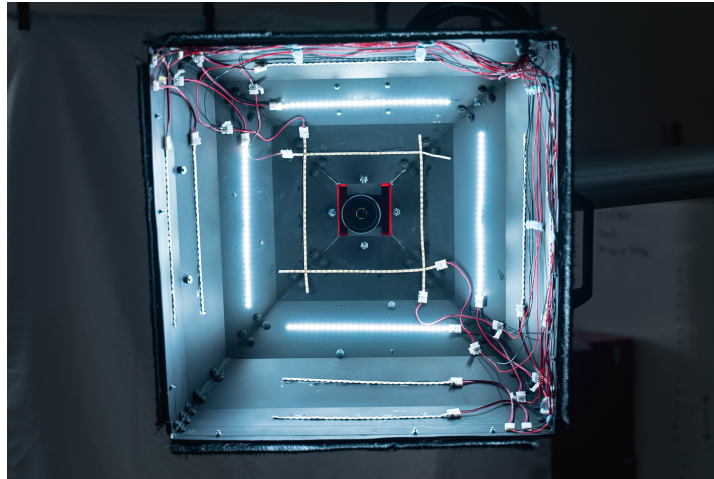
Figure 2: Second iteration of ALICS's hardware, along with the directional lighting prototype, demonstrates its functionality as an inspection platform integrated onto a six-axis robot. Captures images of size 5429×3458 pixels.

Figure 3: Second iteration of ALICS's hardware, along with the directional lighting prototype, demonstrates its utility as a handheld device designed for manual inspection. Captures images of size 5429×3458 pixels.

Chapter 9. Conclusions

Field deployment of these designs has provided the opportunity to capture real-world data sets from concrete assets located in public areas. These iterations help in bringing the ALICS out of the laboratory and into practical, real-world scenarios. Their versatility, and adaptability is useful in various potential applications for automated concrete structure inspection in outdoor and public settings.

Bibliography

[1] A. P. Adewuyi, O. A. Olaniyi, O. S. Olafusi, and A. S. Fawumi, "Compressive and flexural behaviour of unstressed concrete substructure in cassava effluent contaminated soils," Open Journal of Civil Engineering, vol. 05, pp. 239–248, 2015.

[2] S. Fang and J. Huang, "Statics-based model-free damage detection under uncertainties using modal interval analysis," Materials, vol. 13, p. 1567, 2020.

[3] Z. Qi, D. Liu, J. Zhang, and J. Chen, "Micro-concrete crack detection of underwater structures based on convolutional neural network," Machine Vision and Applications, vol. 33, no. 5, p. 74, 2022.

[4] J. K. Chow, Z. Su, J. Wu, P. S. Tan, X. Mao, and Y.-H. Wang, "Anomaly detection of defects on concrete structures with the convolutional autoencoder," Advanced Engineering Informatics, vol. 45, p. 101105, 2020.

[5] K. Laxman, N. Tabassum, L. Ai, C. Cole, and P. Ziehl, "Automated crack detection and crack depth prediction for reinforced concrete structures using deep learning," Construction and Building Materials, vol. 370, p. 130709, 2023.

[6] C. Koch, S. G. Paal, A. Rashidi, Z. Zhu, M. König, and I. Brilakis, "Achievements and challenges in machine vision-based inspection of large concrete structures," Advances in Structural Engineering, vol. 17, no. 3, pp. 303–318, 2014. [Online]. Available: https://doi.org/10.1260/1369-4332.17.3.303

[7] E. Siriwardana, S. K. Amila, S. Kaushalya, S. Chandrasiri, and V. S. Piyaward-

ana, "Driving through a bend: Detection of unsafe driving patterns and prevention of heavy good vehicle rollovers," in 2021 2nd International Informatics and Software Engineering Conference (IISEC), 2021, pp. 1–6.

[8] T. Yamaguchi, S. Nakamura, R. Saegusa, and S. Hashimoto, "Image-based crack detection for real concrete surfaces," IEEJ Transactions on Electrical and Electronic Engineering, vol. 3, no. 1, pp. 128–135, 2008.

[9] Y. Wei, Z. Wei, K. Xue, W. Yao, C. Wang, and Y. Hong, "Automated detection and segmentation of concrete air voids using zero-angle light source and deep learning," Automation in Construction, vol. 130, p. 103877, 2021.

[10] S. Dodge and L. Karam, "Understanding how image quality affects deep neural networks," in 2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX), 2016, pp. 1–6.

[11] J. Shen, H. Wang, Y. Wu, A. Li, C. Chen, and Z. Zheng, "Surgical lighting with contrast enhancement based on spectral reflectance comparison and entropy analysis," Journal of Biomedical Optics, vol. 20, no. 10, pp. 1 – 7, 2015. [Online]. Available: https://doi.org/10.1117/1.JBO.20.10.105012

[12] Z. Chen, R. Derakhshani, C. Halmen, and J. Kevern, "A texture-based method for classifying cracked concrete surfaces from digital images using neural networks," Proceedings of the International Joint Conference on Neural Networks, pp. 2632–2637, 07 2011.

[13] T. Shimano and K. Betsui, "Balanced bright and dark field illumination for remote visual testing to detect cracks on pressure vessel of nuclear reactors," Optical Review, vol. 28, no. 4, pp. 393–402, Jun. 2021. [Online]. Available: https://doi.org/10.1007/s10043-021-00673-5

[14] H.-W. Cho, H.-J. Yoon, and J.-C. Yoon, "Analysis of crack image recognition characteristics in concrete structures depending on the illumination and image acquisition distance through outdoor experiments," Sensors, vol. 16, no. 10, 2016. [Online]. Available: https://www.mdpi.com/1424-8220/16/10/1646

195

Bibliography

[15] L. Zhang, J. Shen, and B. Zhu, "A research on an improved unet-based concrete crack detection algorithm," Structural Health Monitoring, vol. 20, no. 4, pp. 1864–1879, 2021.

[16] Z. Li, J. Yoon, R. Zhang, F. Rajabipour, W. V. Srubar III, I. Dabo, and A. Radlińska, "Machine learning in concrete science: applications, challenges, and best practices," npj computational materials, vol. 8, no. 1, p. 127, 2022.

[17] "EN 1992-1-1: Eurocode 2: Design of concrete structures - Part 1-1: General rules and rules for buildings, Section 7.3.1: Crack Control," 2004. [Online]. Available: https://www.phd.eng.br/wp-content/uploads/2015/12/en.1992.1.1.2004.pdf

[18] R. Ali, J. H. Chuah, M. S. A. Talip, N. Mokhtar, and M. A. Shoaib, "Structural crack detection using deep convolutional neural networks," Automation in Construction, vol. 133, p. 103989, 2022.

[19] H.-I. Lin and F. S. Wibowo, "Image data assessment approach for deep learning-based metal surface defect-detection systems," IEEE Access, vol. 9, pp. 47 621–47 638, 2021.

[20] P. Mohammadi, A. Ebrahimi-Moghadam, and S. Shirani, "Subjective and objective quality assessment of image: A survey," 2014.

[21] D.-E. Lee, G. Hong, M. Maruthi, C.-Y. Yi, Y.-J. Park et al., "Parametric image-based concrete defect assessment method," Case Studies in Construction Materials, p. e02962, 2024.

[22] K. I. Nezhivleva, A. A. Davydova, A. M. Drebuzhan, A. I. Mozhaeva, and A. Balobanov, "Comparing of modern methods used to assess the quality of video sequences during signal streaming with and without human perception," in 2022 Systems of Signal Synchronization, Generating and Processing in Telecommunications (SYNCHROINFO), 2022, pp. 1–6.

[23] A. M. Potashnikov, V. A. Mazin, N. S. Stepanov, A. P. Smirnov, and A. I. Mozhaeva, "Analysis of modern methods used to assess the quality of video se-

quences during signal streaming," in 2022 Systems of Signals Generating and Processing in the Field of on Board Communications, 2022, pp. 1–4.

[24] J. H. Lee, S. Yoon, B. Kim, G.-H. Gwon, I.-H. Kim, and H.-J. Jung, "A new image-quality evaluating and enhancing methodology for bridge inspection using an unmanned aerial vehicle," Smart Struct. Syst, vol. 27, no. 2, pp. 209–226, 2021.

[25] Z. Zhu, D. Huang, X. Zhou, D. Chen, J. Fu, and J. Yang, "Image quality evaluation method for surface crack detection based on standard test chart," Computer-Aided Civil and Infrastructure Engineering, 2023.

[26] S. Dias, J. Almeida, A. Tadeu, and J. de Brito, "Alternative concrete aggregates - review of physical and mechanical properties and successful applications," Cement and Concrete Composites, vol. 152, p. 105663, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0958946524002361

[27] K. Upreti, M. Verma, M. Agrawal, J. Garg, R. Kaushik, C. Agrawal, D. Singh, and R. Narayanasamy, "Prediction of mechanical strength by using an artificial neural network and random forest algorithm," Journal of Nanomaterials, vol. 2022, pp. 1–12, 2022.

[28] P. Wu, A. Liu, J. Fu, X. Ye, and Y. Zhao, "Autonomous surface crack identification of concrete structures based on an improved one-stage object detection algorithm," Engineering structures, vol. 272, p. 114962, 2022.

[29] M. Shahrokhishahraki, M. Malekpour, S. Mirvalad, and G. Faraone, "Machine learning predictions for optimal cement content in sustainable concrete constructions," Journal of Building Engineering, vol. 82, p. 108160, 2024.

[30] D. L. Fillmore, "Literature review of the effects of radiation and temperature on the aging of concrete," 09 2004. [Online]. Available: https://scite.ai/reports/10.2172/910954

Bibliography

[31] British standards institution, Guide to fatigue design and assessment of steel products. BS 7608, 2014.

[32] M. Flah, A. R. Suleiman, and M. L. Nehdi, "Classification and quantification of cracks in concrete structures using deep learning image-based techniques," Cement and Concrete Composites, vol. 114, p. 103781, 2020.

[33] K. Liu, S. Law, and X. Zhu, "A layered beam element for modeling de-bonding of steel bars in concrete and its detection using static measurements," Structural Control and Health Monitoring, vol. 25, p. e2142, 2018.

[34] A. Neville, "Properties of concrete," 1995. [Online]. Available: http://10.250.8.41:8080/xmlui/handle/123456789/48077

[35] G. L. Golewski, "The phenomenon of cracking in cement concretes and reinforced concrete structures: the mechanism of cracks formation, causes of their initiation, types and places of occurrence, and methods of detection—a review," Buildings, vol. 13, no. 3, p. 765, 2023.

[36] M.-K. Kim, H. Sohn, and C.-C. Chang, "Localization and quantification of concrete spalling defects using terrestrial laser scanning," Journal of Computing in Civil Engineering, vol. 29, no. 6, p. 04014086, 2015.

[37] P. Nama, A. Jain, R. Srivastava, and Y. Bhatia, "Study on causes of cracks & its preventive measures in concrete structures," International Journal of Engineering Research and Applications, vol. 5, no. 5, pp. 119–123, 2015. [Online]. Available: https://ssrn.com/abstract=3836421

[38] C. J. C. ., "Study on causes and prevention of cracks in building," International Journal for Research in Applied Science and Engineering Technology, vol. 6, pp. 453–461, 2018.

[39] C. P. Technologies, "Concrete spalling: Causes, effects and repair," https://cp-tech.co.uk/concrete-spalling-causes-effects-and-repair/, Concrete Preservation Technologies, January 28 2017, accessed on April 2023.

[40] "Spalling concrete: The causes and control, repair protection methods," 03 2022, accessed in August 2022. [Online]. Available: https://www.mapei.com/my/en/blog/techical-talk-details/2022/03/16/spalling-concrete-the-causes-and-control-repair-protection-methods

[41] I. Bidžević, S. Džidić, and A. El Sayed, "Spalling of concrete," in International Symposium on Innovative and Interdisciplinary Applications of Advanced Technologies. Springer, 2023, pp. 75–91.

[42] H. N. Desk, "What is spalling in concrete and how to prevent it?" https://housing.com/news/spalling-in-concrete/amp/, 2023, accessed on May 2023. [Online]. Available: https://housing.com/news/spalling-in-concrete/amp/

[43] C. Acquarelli, L. Paliotta, G. D. Bellis, and M. S. Sarto, "Electro-mechanical properties of multilayer graphene-based polymeric composite obtained through a capillary rise method," Sensors, vol. 16, p. 1780, 2016.

[44] J. Chen, X. Wang, S. Liu, and C. Wang, "Application of iterative maximum weighted likelihood estimation in 3-d target localization," Applied Sciences, vol. 9, p. 4921, 2019.

[45] A. Mirzazade, C. Popescu, J. Gonzalez-Libreros, T. Blanksvärd, B. Täljsten, and G. Sas, "Semi-autonomous inspection for concrete structures using digital models and a hybrid approach based on deep learning and photogrammetry," Journal of Civil Structural Health Monitoring, vol. 13, no. 8, pp. 1633–1652, 2023.

[46] P. Ren and Z. Zhou, "Two-step approach to processing raw strain monitoring data for damage detection of structures under operational conditions," Sensors, vol. 21, p. 6887, 2021.

[47] What is structural health monitoring in civil engineering? The Constructor. Accessed on: October, 2023. [Online]. Available: https://theconstructor.org/digital-construction/structural-health-monitoring-civil-engineering/554160/

Bibliography

[48] M. Q. Feng, Y. Fukuda, M. Mizuta, and E. Özer, "Citizen sensors for shm: use of accelerometer data from smartphones," Sensors, vol. 15, pp. 2980–2998, 2015.

[49] A. Bakhshi and H. A. Tehrani, "Structural health monitoring in multi-story frames based on signal processing and rbf neural networks," in Proc., 16th World Conf. on Earthquake Engineering. Chile, Valdivia: Chilean Association on Seismology and Earthquake Engineering, 2017. [Online]. Available: https://api.semanticscholar.org/CorpusID:202752021

[50] M. A. Wahab and G. De Roeck, "Damage detection in bridges using modal curvatures: application to a real damage scenario," Journal of Sound and vibration, vol. 226, no. 2, pp. 217–235, 1999.

[51] E. Safak, "Detection of seismic damage in structures from continuous vibration records," in Proceeding, 9ˆ¡ th¿ International Conference on Structural Safety and Reliability, Rome, Italy, 2005, 2005. [Online]. Available: https://cir.nii.ac.jp/crid/1570009750746272640

[52] Y. Fukuda, M. Q. Feng, Y. Narita, S. Kaneko, and T. Tanaka, "Vision-based displacement sensor for monitoring dynamic response using robust object search algorithm," IEEE Sensors Journal, vol. 13, no. 12, pp. 4725–4732, 2013.

[53] C. Gentile, "Deflection measurement on vibrating stay cables by non-contact microwave interferometer," Ndt & E International, vol. 43, no. 3, pp. 231–240, 2010.

[54] "Structural health monitoring - an overview." [Online]. Available: https://www.worldscientific.com/doi/pdf/10.1142/9789811201097_0001

[55] H. Sun, L. Song, and Z. Yu, "Assessing the fatigue damage of concrete structures using automatically classified crack severity level information," International Journal of Fatigue, vol. 179, p. 108071, 2024.

[56] ADEPT, "Guidance document for performance measurement of highway structures," 2007, accessed: August 2024. [Online]. Available: https:

Bibliography

//www.adeptnet.org.uk/sites/default/files/media/2022-11/Part%20Guidance%
20Document%20for%20Performance%20Measurement-%20A%20Framework.
pdf#:~:text=Cover%20the%20main%20highway%20structure%20types

[57] Z. Orbán and M. Gutermann, "Assessment of masonry arch railway bridges using non-destructive in-situ testing methods," Engineering Structures, vol. 31, no. 10, pp. 2287–2298, 2009.

[58] Ł. Sadowski, "Non-destructive evaluation of the pull-off adhesion of concrete floor layers using rbf neural network," Journal of Civil Engineering and Management, vol. 19, no. 4, pp. 550–560, 2013.

[59] M. Kloiber, M. Drdácký, J. S. Machado, M. Piazza, and N. Yamaguchi, "Prediction of mechanical properties by means of semi-destructive methods: A review," Construction and Building Materials, vol. 101, pp. 1215–1234, 2015.

[60] Guidebook on Non-destructive Testing of Concrete Structures, ser. Training Course Series. Vienna: INTERNATIONAL ATOMIC ENERGY AGENCY, 2002, no. 17. [Online]. Available: https://www.iaea.org/publications/6347/guidebook-on-non-destructive-testing-of-concrete-structures

[61] S. Kim, J. K. Seo, and T. Ha, "A nondestructive evaluation method for concrete voids: frequency differential electrical impedance scanning," SIAM Journal on Applied Mathematics, vol. 69, no. 6, pp. 1759–1771, 2009.

[62] K. Schabowicz, "Non-destructive testing of materials in civil engineering," p. 3237, 2019.

[63] National highways sector scheme 31 for the bridge inspector certification scheme. [Online]. Available: https://www.bridgeforum.org/guidance/national-highways-sector-scheme-31-for-the-bridge-inspector-certification-scheme/

[64] J. E. See, C. G. Drury, A. Speed, A. Williams, and N. Khalandi, "The role of visual inspection in the 21st century," in Proceedings of the Human Factors and

Ergonomics Society Annual Meeting, vol. 61, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 2017, pp. 262–266.

[65] E. Menendez, J. G. Victores, R. Montero, S. Martínez, and C. Balaguer, "Tunnel structural inspection and assessment using an autonomous robotic system," Automation in Construction, vol. 87, pp. 117–126, 2018.

[66] S. K. U. Rehman, Z. Ibrahim, S. A. Memon, and M. Jameel, "Nondestructive test methods for concrete bridges: A review," Construction and building materials, vol. 107, pp. 58–86, 2016.

[67] A. K. Agrawal, G. Washer, S. Alampalli, X. Gong, and R. Cao, "Evaluation of the consistency of bridge inspection quality in new york state," Journal of Civil Structural Health Monitoring, vol. 11, no. 5, pp. 1393–1413, 2021.

[68] M. A. Mohammed, Z. Han, Y. Li, Z. Al-Huda, C. Li, and W. Wang, "End-to-end semi-supervised deep learning model for surface crack detection of infrastructures," Frontiers in Materials, vol. 9, 2022.

[69] Z. Qu, F. Ju, Y. Guo, L. Bai, and K. Chen, "Concrete surface crack detection with the improved pre-extraction and the second percolation processing methods," Plos One, vol. 13, p. e0201109, 2018.

[70] M. Zeeshan, S. M. Adnan, W. I. U. Ahmad, and F. Z. Khan, "Structural crack detection and classification using deep convolutional neural network," Pakistan Journal of Engineering and Technology, vol. 4, pp. 50–56, 2021.

[71] B. Sutter, A. Lelevé, M. T. Pham, O. Gouin, N. Jupille, M. Kuhn, P. Lulé, P. Michaud, and P. Rémy, "A semi-autonomous mobile robot for bridge inspection," Automation in Construction, vol. 91, pp. 111–119, 2018.

[72] C. Eitzinger, S. Ghidoni, E. Menegatti et al., "Thermobot: towards semi-autonomous, thermographic detection of cracks," in International conference on heating by electromagnetic sources, Padua, 2013. [Online]. Available: https://hdl.handle.net/11577/2760678

[73] A. Jimenez-Cano, J. Braga, G. Heredia, and A. Ollero, "Aerial manipulator for structure inspection by contact from the underside," in 2015 IEEE/RSJ international conference on intelligent robots and systems (IROS). IEEE, 2015, pp. 1879–1884.

[74] D. P. Kumar and K. Kannan, "A roadmap for designing an automated visual inspection system," International Journal of Computer Applications, vol. 1, no. 19, pp. 34–37, 2010.

[75] S. Pennada, M. Perry, J. McAlorum, H. Dow, and G. Dobie, "Performance evaluation of an improved deep cnn-based concrete crack detection algorithm," in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol. 12486, 2023, p. 1248615.

[76] S. Halder and K. Afsari, "Robots in inspection and monitoring of buildings and infrastructure: A systematic review," Applied Sciences, vol. 13, no. 4, p. 2304, 2023.

[77] C. Liu, C.-S. Tang, B. Shi, and W.-B. Suo, "Automatic quantification of crack patterns by image processing," Computers & Geosciences, vol. 57, pp. 77–80, 2013.

[78] M. Jogin, Mohana, M. S. Madhulika, G. D. Divya, R. K. Meghana, and S. Apoorva, "Feature extraction using convolution neural networks (cnn) and deep learning," in 2018 3rd IEEE International Conference on Recent Trends in Electronics, Information Communication Technology (RTEICT), 2018, pp. 2319–2323.

[79] P. Patel and A. Thakkar, "The upsurge of deep learning for computer vision applications," International Journal of Electrical and Computer Engineering, vol. 10, no. 1, p. 538, 2020.

[80] M. Saqlain, S. Rubab, M. M. Khan, N. Ali, and S. Ali, "Hybrid approach for shelf monitoring and planogram compliance (hyb-smpc) in retails using deep learning

and computer vision," <u>Mathematical Problems in Engineering</u>, vol. 2022, pp. 1–18, 2022.

[81] D. Ai, G. Jiang, S.-K. Lam, P. He, and C. Li, "Computer vision framework for crack detection of civil infrastructure—a review," <u>Engineering Applications of Artificial Intelligence</u>, vol. 117, p. 105478, 2023.

[82] Y.-J. Cha, W. Choi, and O. Büyüköztürk, "Deep learning-based crack damage detection using convolutional neural networks," <u>Computer-Aided Civil and Infrastructure Engineering</u>, vol. 32, no. 5, pp. 361–378, 2017.

[83] S. Yokoyama and T. Matsumoto, "Development of an automatic detector of cracks in concrete using machine learning," <u>Procedia engineering</u>, vol. 171, pp. 1250–1255, 2017.

[84] F. Ni, J. Zhang, and Z. Chen, "Zernike-moment measurement of thin-crack width in images enabled by dual-scale deep learning," <u>Computer-Aided Civil and Infrastructure Engineering</u>, vol. 34, no. 5, pp. 367–384, 2019.

[85] V. P. Golding, Z. Gharineiat, H. S. Munawar, and F. Ullah, "Crack detection in concrete structures using deep learning," <u>Sustainability</u>, 2022.

[86] S. Sony, K. Dunphy, A. Sadhu, and M. Capretz, "A systematic review of convolutional neural network-based structural condition assessment techniques," <u>Engineering Structures</u>, vol. 226, p. 111347, 2021.

[87] S. Dorafshan, R. J. Thomas, and M. Maguire, "Comparison of deep convolutional neural networks and edge detectors for image-based crack detection in concrete," <u>Construction and Building Materials</u>, vol. 186, pp. 1031–1045, 2018.

[88] S. S. Zadeh, M. Khorshidi, F. Kooban <u>et al.</u>, "Concrete surface crack detection with convolutional-based deep learning models," <u>arXiv preprint arXiv:2401.07124</u>, 2024.

Bibliography

[89] L. Zhang, F. Yang, Y. D. Zhang, and Y. J. Zhu, "Road crack detection using deep convolutional neural network," in 2016 IEEE international conference on image processing (ICIP). IEEE, 2016, pp. 3708–3712.

[90] H. Xu, X. Su, Y. Wang, H. Cai, K. Cui, and X. Chen, "Automatic bridge crack detection using a convolutional neural network," Applied Sciences, vol. 9, no. 14, p. 2867, 2019.

[91] L. Pauly, D. Hogg, R. Fuentes, and H. Peel, "Deeper networks for pavement crack detection," in Proceedings of the 34th ISARC. IAARC, 2017, pp. 479–485.

[92] X. Yang, H. Li, Y. Yu, X. Luo, T. Huang, and X. Yang, "Automatic pixel-level crack detection and measurement using fully convolutional network," Computer-Aided Civil and Infrastructure Engineering, vol. 33, no. 12, pp. 1090–1109, 2018.

[93] H. H. Nakashima, D. Mantovani, and C. Machado Junior, "Users' trust in black-box machine learning algorithms," Revista de Gestão, vol. 31, no. 2, pp. 237–250, 2024.

[94] I. H. Sarker, "Machine learning: Algorithms, real-world applications and research directions," SN computer science, vol. 2, no. 3, p. 160, 2021.

[95] I.-H. Kim, H. Jeon, S.-C. Baek, W.-H. Hong, and H.-J. Jung, "Application of crack identification techniques for an aging concrete bridge inspection using an unmanned aerial vehicle," Sensors, vol. 18, no. 6, p. 1881, 2018.

[96] C. L. S, S. Uma Maheswari, S. S. Rao, N. Mukkapati, C. S. Sundar Ganesh, and A. Kumar, "Deep learning network for object detection under the poor lighting condition," in 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS), 2022, pp. 555–560.

[97] H. Maeda, Y. Sekimoto, T. Seto, T. Kashiyama, and H. Omata, "Road damage detection and classification using deep neural networks with smartphone images,"

Computer-Aided Civil and Infrastructure Engineering, vol. 33, no. 12, pp. 1127–1141, 2018.

[98] J. Deng, Y. Lu, and V. C.-S. Lee, "Concrete crack detection with handwriting script interferences using faster region-based convolutional neural network," Computer-Aided Civil and Infrastructure Engineering, vol. 35, no. 4, pp. 373–388, 2020.

[99] W. Chen and T. Shah, "Exploring low-light object detection techniques," 2021.

[100] Y. Jiang, D. Pang, and C. Li, "A deep learning approach for fast detection and classification of concrete damage," Automation in Construction, vol. 128, p. 103785, 2021.

[101] Y. Hamishebahar, H. Guan, S. So, and J. Jo, "A comprehensive review of deep learning-based crack detection approaches," Applied Sciences, vol. 12, no. 3, p. 1374, 2022.

[102] X. Wang and Z. Hu, "Grid-based pavement crack analysis using deep learning," in 2017 4th international conference on transportation information and safety (ICTIS). IEEE, 2017, pp. 917–924.

[103] S. Dorafshan, R. J. Thomas, C. Coopmans, and M. Maguire, "Deep learning neural networks for suas-assisted structural inspections: Feasibility and application," in 2018 international conference on unmanned aircraft systems (ICUAS). IEEE, 2018, pp. 874–882.

[104] S. Park, S. Bang, H. Kim, and H. Kim, "Patch-based crack detection in black box images using convolutional neural networks," Journal of Computing in Civil Engineering, vol. 33, no. 3, p. 04019017, 2019.

[105] Y. Bai, H. Sezen, and A. Yilmaz, "Detecting cracks and spalling automatically in extreme events by end-to-end deep learning frameworks," ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial

Bibliography

Information Sciences, vol. V-2-2021, pp. 161–168, 2021. [Online]. Available: https://isprs-annals.copernicus.org/articles/V-2-2021/161/2021/

[106] T. Yasmin, C. Le, and H. M. La, "Deep architecture based spalling severity detection system using encoder-decoder networks," in International Symposium on Visual Computing.   Springer, 2022, pp. 332–343.

[107] E. Mohammed Abdelkader, O. Moselhi, M. Marzouk, and T. Zayed, "Entropy-based automated method for detection and assessment of spalling severities in reinforced concrete bridges," Journal of Performance of Constructed Facilities, vol. 35, no. 1, p. 04020132, 2021.

[108] N.-D. Hoang, Q.-L. Nguyen, and X.-L. Tran, "Automatic detection of concrete spalling using piecewise linear stochastic gradient descent logistic regression and image texture analysis," Complex., vol. 2019, pp. 5 910 625:1–5 910 625:14, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:199116568

[109] F. Panella, A. Lipani, and J. Boehm, "Semantic segmentation of cracks: Data challenges and architecture," Automation in Construction, vol. 135, p. 104110, 2022.

[110] K. Zhang, H. Cheng, and B. Zhang, "Unified approach to pavement crack and sealed crack detection using preclassification based on transfer learning," Journal of Computing in Civil Engineering, vol. 32, no. 2, p. 04018001, 2018.

[111] D. Kang, S. S. Benipal, D. L. Gopal, and Y.-J. Cha, "Hybrid pixel-level concrete crack segmentation and quantification across complex backgrounds using deep learning," Automation in Construction, vol. 118, p. 103291, 2020.

[112] K. Zhang, H.-D. Cheng, and S. Gai, "Efficient dense-dilation network for pavement cracks detection with large input image size," in 2018 21st International Conference on Intelligent Transportation Systems (ITSC).   IEEE, 2018, pp. 884–889.

Bibliography

[113] X. Zhang, D. Rajan, and B. Story, "Concrete crack detection using context-aware deep semantic segmentation network," Computer-Aided Civil and Infrastructure Engineering, vol. 34, no. 11, pp. 951–971, 2019.

[114] H. S. Liew, Y. P. Loh, and S. Ong, "Low-light is more than darkness: An empirical study on illumination types and enhancement methods," in 2023 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2023, pp. 651–658.

[115] J. Huyan, W. Li, S. Tighe, J. Zhai, Z. Xu, and Y. Chen, "Detection of sealed and unsealed cracks with complex backgrounds using deep convolutional neural network," Automation in Construction, vol. 107, p. 102946, 2019.

[116] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014. [Online]. Available: https://arxiv.org/abs/1409.1556

[117] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 4510–4520.

[118] S. S. A. Zaidi, M. S. Ansari, A. Aslam, N. Kanwal, M. Asghar, and B. Lee, "A survey of modern deep learning based object detection models," Digital Signal Processing, vol. 126, p. 103514, 2022.

[119] S. Katsigiannis, S. Seyedzadeh, A. Agapiou, and N. Ramzan, "Deep learning for crack detection on masonry façades using limited data and transfer learning," Journal of Building Engineering, vol. 76, p. 107105, 2023.

[120] A. BabaAhmadi, S. Khalafi, M. ShariatPanahi, and M. Ayati, "Designing an improved deep learning-based model for covid-19 recognition in chest x-ray images: a knowledge distillation approach," Iran Journal of Computer Science, vol. 7, no. 2, pp. 177–187, 2024.

[121] V. P. Golding, Z. Gharineiat, H. S. Munawar, and F. Ullah, "Crack detection in

concrete structures using deep learning," Sustainability, vol. 14, no. 13, p. 8117, 2022.

[122] H. Tapamo, A. Bosman, J. Maina, and E. Horak, "Convolutional neural networks for crack detection on flexible road pavements," in International Conference on Soft Computing and Pattern Recognition.   Springer, 2022, pp. 194–204.

[123] Ç. F. Özgenel and A. G. Sorguç, "Performance comparison of pretrained convolutional neural networks on crack detection in buildings," in Isarc. proceedings of the international symposium on automation and robotics in construction, vol. 35. IAARC Publications, 2018, pp. 1–8.

[124] K. Gopalakrishnan, S. K. Khaitan, A. Choudhary, and A. Agrawal, "Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection," Construction and building materials, vol. 157, pp. 322–330, 2017.

[125] H. Li, W. Wang, M. Wang, L. Li, and V. Vimlund, "A review of deep learning methods for pixel-level crack detection," Journal of Traffic and Transportation Engineering (English Edition), vol. 9, no. 6, pp. 945–968, 2022.

[126] H. Zoubir, M. Rguig, M. El Aroussi, A. Chehri, R. Saadane, and G. Jeon, "Concrete bridge defects identification and localization based on classification deep convolutional neural networks and transfer learning," Remote sensing, vol. 14, no. 19, p. 4882, 2022.

[127] S. Meklati, K. Boussora, M. E. H. Abdi, and S.-A. Berrani, "Surface damage identification for heritage site protection: A mobile crowd-sensing solution based on deep learning," J. Comput. Cult. Herit., vol. 16, no. 2, mar 2023. [Online]. Available: https://doi.org/10.1145/3569093

[128] M. Mazni, A. R. Husain, M. I. Shapiai, I. S. Ibrahim, D. W. Anggara, and R. Zulkifli, "An investigation into real-time surface crack classification and measurement for structural health monitoring using transfer learning convolutional

neural networks and otsu method," Alexandria Engineering Journal, vol. 92, pp. 310–320, 2024.

[129] K. Dong, C. Zhou, Y. Ruan, and Y. Li, "Mobilenetv2 model for image classification," in 2020 2nd International Conference on Information Technology and Computer Application (ITCA), 2020, pp. 476–480.

[130] Q. Xiang, X. Wang, R. Li, G. Zhang, J. Lai, and Q. Hu, "Fruit image classification based on mobilenetv2 with transfer learning technique," in Proceedings of the 3rd international conference on computer science and application engineering, 2019, pp. 1–7.

[131] R. Rakshitha, S. Srinath, N. Vinay Kumar, S. Rashmi, and B. Poornima, "Automated detection of cracks in asphalt pavement images using texture descriptors and machine learning classifier," in International Conference on Computer Vision and Image Processing. Springer, 2023, pp. 13–24.

[132] S. Szeghalmy and A. Fazekas, "A comparative study of the use of stratified cross-validation and distribution-balanced stratified cross-validation in imbalanced learning," Sensors, vol. 23, no. 4, p. 2333, 2023.

[133] T. Buckley, B. Ghosh, and V. Pakrashi, "A feature extraction & selection benchmark for structural health monitoring," Structural Health Monitoring, vol. 22, no. 3, pp. 2082–2127, 2023.

[134] S. Widodo, H. Brawijaya, and S. Samudi, "Stratified k-fold cross validation optimization on machine learning for prediction," Sinkron: jurnal dan penelitian teknik informatika, vol. 7, no. 4, pp. 2407–2414, 2022.

[135] M. Wojciuk, Z. Swiderska-Chadaj, K. Siwek, and A. Gertych, "Improving classification accuracy of fine-tuned cnn models: Impact of hyperparameter optimization," Heliyon, 2024.

[136] T. Bartz-Beielstein, "Hyperparameter tuning," in Online Machine Learning: A Practical Guide with Examples in Python. Springer, 2024, pp. 125–140.

[137] X. Ying, "An overview of overfitting and its solutions," in Journal of physics: Conference series, vol. 1168. IOP Publishing, 2019, p. 022022.

[138] T. Miseta, A. Fodor, and Á. Vathy-Fogarassy, "Surpassing early stopping: A novel correlation-based stopping criterion for neural networks," Neurocomputing, vol. 567, p. 127028, 2024.

[139] D. Lattanzi and G. Miller, "Review of robotic infrastructure inspection systems," Journal of Infrastructure Systems, vol. 23, no. 3, p. 04017004, 2017.

[140] Q. Mei and M. Gül, "A cost effective solution for pavement crack inspection using cameras and deep neural networks," Construction and Building Materials, vol. 256, p. 119397, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950061820314021

[141] K.-W. Tse, R. Pi, Y. Sun, C.-Y. Wen, and Y. Feng, "A novel real-time autonomous crack inspection system based on unmanned aerial vehicles," Sensors, vol. 23, no. 7, 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/7/3418

[142] V. Mandal, L. Uong, and Y. Adu-Gyamfi, "Automated road crack detection using deep convolutional neural networks," in 2018 IEEE International Conference on Big Data (Big Data), 2018, pp. 5212–5215.

[143] Z. Wang, "Applications of objective image quality assessment methods [applications corner]," IEEE Signal Processing Magazine, vol. 28, no. 6, pp. 137–142, 2011.

[144] L. S. Chow and R. Paramesran, "Review of medical image quality assessment," Biomedical Signal Processing and Control, vol. 27, pp. 145–154, 2016.

[145] O. Zelmati, B. Bondžulić, B. Pavlović, I. Tot, and S. Merrouche, "Study of subjective and objective quality assessment of infrared compressed images," Journal of Electrical Engineering, vol. 73, no. 2, pp. 73–87, 2022.

Bibliography

[146] T. Chandra Mohan, "Blind image quality assessment of smartphone-captured images in the wild," December 2021. [Online]. Available: http://essay.utwente.nl/89267/

[147] S. Bosse, D. Maniry, K.-R. Müller, T. Wiegand, and W. Samek, "Deep neural networks for no-reference and full-reference image quality assessment," IEEE Transactions on Image Processing, vol. 27, no. 1, pp. 206–219, 2018.

[148] L. Wang, "A survey on iqa," 2022.

[149] S. Xu, S. Jiang, and W. Min, "No-reference/blind image quality assessment: A survey," IETE Technical Review, vol. 34, no. 3, pp. 223–245, 2017.

[150] D. Varga, "No-reference image quality assessment using the statistics of global and local image features," Electronics, vol. 12, no. 7, p. 1615, 2023.

[151] P. Ye, J. Kumar, L. Kang, and D. Doermann, "Real-time no-reference image quality assessment based on filter learning," in 2013 IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 987–994.

[152] J. Park, C.-K. Kang, and Y. Lee, "Quantitative evaluation of the image quality using the fast nonlocal means denoising approach in diffusion-weighted magnetic resonance imaging with high b-value," Journal of the Korean Physical Society, vol. 78, pp. 244–250, 2021.

[153] E. Ichi, F. Jafari, and S. Dorafshan, "Sdnet2021: Annotated nde dataset for subsurface structural defects detection in concrete bridge decks," Infrastructures, vol. 7, no. 9, 2022. [Online]. Available: https://www.mdpi.com/2412-3811/7/9/107

[154] E. U. Rahman, Y. Zhang, S. Ahmad, H. I. Ahmad, and S. Jobaer, "Autonomous vision-based primary distribution systems porcelain insulators inspection using uavs," Sensors, vol. 21, no. 3, p. 974, 2021.

[155] J. Shen, H. Wang, Y. Wu, A. Li, C. Chen, and Z. Zheng, "Surgical lighting with contrast enhancement based on spectral reflectance comparison and entropy analysis," Journal of biomedical optics, vol. 20, no. 10, pp. 105 012–105 012, 2015.

[156] J. Sammarco, B. Macdonald, B. Demich, E. Rubinstein, and M. Martell, "Led lighting for improving trip object detection for a walk-thru roof bolter," Lighting Research & Technology, vol. 51, no. 5, pp. 725–741, 2019. [Online]. Available: https://doi.org/10.1177/1477153518783816

[157] V. Elango and L. Karunamoorthy, "Effect of lighting conditions in the study of surface roughness by machine vision - an experimental design approach," The International Journal of Advanced Manufacturing Technology, vol. 37, no. 1, pp. 92–103, Apr 2008. [Online]. Available: https://doi.org/10.1007/s00170-007-0942-y

[158] S. K. Kopparapu, "Lighting design for machine vision application," Image and Vision Computing, vol. 24, no. 7, pp. 720–726, 2006.

[159] D. Martin, "A practical guide to machine vision lighting," 2021, accessed July 2022. [Online]. Available: https://www.ni.com/en-gb/innovations/white-papers/12/a-practical-guide-to-machine-vision-lighting.html

[160] I. Kuric, J. Klarák, V. Bulej, M. Sága, M. Kandera, A. Hajdučík, and K. Tucki, "Approach to automated visual inspection of objects based on artificial intelligence," Applied Sciences, vol. 12, no. 2, 2022. [Online]. Available: https://www.mdpi.com/2076-3417/12/2/864

[161] C. Ibarra-Castanedo, S. Sfarra, M. Genest, and X. Maldague, Infrared Vision: Visual Inspection Beyond the Visible Spectrum. London: Springer London, 2015, pp. 41–58. [Online]. Available: https://doi.org/10.1007/978-1-4471-6741-9_2

[162] K. Hacıefendioğlu and H. B. Başağa, "Concrete road crack detection using deep learning-based faster r-cnn method," Iranian Journal of Science and Technology, Transactions of Civil Engineering, pp. 1–13, 2022.

Bibliography

[163] L. Yangí, G. Yang, Z. Liu, Y. Chang, B. Jiang, Y. Awad, and J. Xiao, "Wall-climbing robot for visual and gpr inspection," in 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), 2018, pp. 1004–1009.

[164] K. Loupos, A. D. Doulamis, C. Stentoumis, E. Protopapadakis, K. Makantasis, N. D. Doulamis, A. Amditis, P. Chrobocinski, J. Victores, R. Montero, E. Menendez, C. Balaguer, R. Lopez, M. Cantero, R. Navarro, A. Roncaglia, L. Belsito, S. Camarinopoulos, N. Komodakis, and P. Singh, "Autonomous robotic system for tunnel structural inspection and assessment," International Journal of Intelligent Robotics and Applications, vol. 2, no. 1, pp. 43–66, Mar 2018.

[165] H. Kolvenbach, D. Wisth, R. Buchanan, G. Valsecchi, R. Grandia, M. Fallon, and M. Hutter, "Towards autonomous inspection of concrete deterioration in sewers with legged robots," Journal of Field Robotics, vol. 37, no. 8, pp. 1314–1327, 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/rob.21964

[166] E. Menendez, J. G. Victores, R. Montero, S. Martínez, and C. Balaguer, "Tunnel structural inspection and assessment using an autonomous robotic system," Automation in Construction, vol. 87, pp. 117–126, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0926580516303806

[167] H. Crohas and P. Lepert, "Damage detection monitoring method for offshore platforms is field tested," Oil and Gas Journal, 1982.

[168] K. Loupos, A. D. Doulamis, C. Stentoumis, E. Protopapadakis, K. Makantasis, N. D. Doulamis, A. Amditis, P. Chrobocinski, J. Victores, R. Montero et al., "Autonomous robotic system for tunnel structural inspection and assessment," International Journal of Intelligent Robotics and Applications, vol. 2, pp. 43–66, 2018.

[169] J. Cheng, Q. Chen, and X. Huang, "An algorithm for crack detection, segmentation, and fractal dimension estimation in low-light environments by fusing fft and convolutional neural network," Fractal and Fractional, vol. 7, no. 11, p. 820, 2023.

[170] N. Jayanthi, T. Ghosh, R. K. Meena, and M. Verma, "Length and width of low-light, concrete hairline crack detection and measurement using image processing method," Asian Journal of Civil Engineering, vol. 25, no. 3, pp. 2705–2714, 2024.

[171] C. Simler, E. Trostmann, and D. Berndt, "Automatic crack detection on concrete floor images," in Photonics and Education in Measurement Science 2019, vol. 11144.   SPIE, 2019, pp. 191–200.

[172] M. G. Padalkar, C. Beltrán-González, M. Bustreo, A. Del Bue, and V. Murino, "A versatile crack inspection portable system based on classifier ensemble and controlled illumination," in 2020 25th International Conference on Pattern Recognition (ICPR), 2021, pp. 4009–4016.

[173] D. Langenkämper, R. v. Kevelaer, T. Möller, and T. W. Nattkemper, "Gan-based synthesis of deep learning training data for uav monitoring," The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XLIII-B1-2020, pp. 465–469, 2020.

[174] AVIAN. 10 pros & cons of unmanned aerial systems (uav) for surveying & mapping. Accessed on October, 2023. [Online]. Available: https://www.avian.net.au/pros-and-cons-of-uav-surveys/

[175] J. Butt, "Exploring the interrelationship between additive manufacturing and industry 4.0," Designs, vol. 4, p. 13, 2020.

[176] J. Sandoval, M. A. Laribi, S. Zeghloul, and M. Arsicault, "On the design of a safe human-friendly teleoperated system for doppler sonography," Robotics, vol. 8, p. 29, 2019.

[177] R. Tarwala and L. D. Dorr, "Robotic assisted total hip arthroplasty using the mako platform," Current reviews in musculoskeletal medicine, vol. 4, pp. 151–156, 2011.

[178] M. G. Carmichael, S. Aldini, R. Khonasty, A. Tran, C. Reeks, D. Liu, K. J. Waldron, and G. Dissanayake, "The anbot: An intelligent robotic co-worker

for industrial abrasive blasting," in 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2019, pp. 8026–8033.

[179] E. Optics, "Imaging fundamentals," https://www.edmundoptics.com/knowledge-center/application-notes/imaging/6-fundamental-parameters-of-an-imaging-system/, accessed on March 2021.

[180] J. McAlorum, H. Dow, S. Pennada, M. Perry, and G. Dobie, "Automated concrete crack inspection with directional lighting platform," IEEE Sensors Letters, pp. 1–4, 2023.

[181] Teledyne FLIR, "Using auto exposure," 2016, last visited June 2023. [Online]. Available: https://www.flir.co.uk/support-center/iis/machine-vision/application-note/using-auto-exposure/

[182] T. FLIR. Ultra-compact form blackfly s usb3 model: Bfs-u3-200s6c-c: 20 mp, 18 fps, sony imx183, color. Teledyne FLIR. [Online]. Available: https://www.flir.co.uk/products/blackfly-s-usb3/

[183] S. Pennada, M. Perry, J. McAlorum, H. Dow, and G. Dobie, "Threshold-based brisque-assisted deep learning for enhancing crack detection in concrete structures," Journal of Imaging, vol. 9, no. 10, p. 218, 2023.

[184] A. Kaushal and A. K. Sharma, "A novel color-coded light-based communication for autonomous swarm robotics research," in Advances in Energy Technology: Select Proceedings of EMSME 2020. Springer, 2022, pp. 621–629.

[185] K. Blankenbach, R. Isele, D. Ochs, and S. Reichel, "Led-based automotive exterior displays and interior signage for autonomous cars," in Advances in Display Technologies X, vol. 11304. SPIE, 2020, pp. 104–115.

[186] B. HUNKO, "Hardware and software system of light visualization of sound signals," COMPUTER SYSTEMS AND INFORMATION TECHNOLOGIES, 2022.

[187] "Guide for ws2812b addressable rgb led strip with arduino," accessed Feb 2021. [Online]. Available: https://randomnerdtutorials.com/guide-for-ws2812b-addressable-rgb-led-strip-with-arduino/

[188] N. S. S. Rao, M. S. Bhavani, C. H. Reddy, G. V. Durga, R. Likhitha, D. Ganesh, and M. S. Hussain, "Railway track crack detection system."

[189] P. Gawade, S. Kadam, A. Kamble, and V. S. Seethalakshmi, "Iot based robotic car for railway track crack detection system."

[190] "Arduino uno rev3 - arduino documentation," Arduino, accessed: February 2021. [Online]. Available: https://docs.arduino.cc/hardware/uno-rev3

[191] 1st Vision, "Focal length/fov lens calculator," https://www.1stvision.com/lens/fov-lens-calculator, accessed on March 2021.

[192] F.-C. Chen, M. R. Jahanshahi, R.-T. Wu, and C. Joffe, "A texture-based video processing methodology using bayesian data fusion for autonomous crack detection on metallic surfaces," Computer-Aided Civil and Infrastructure Engineering, vol. 32, no. 4, pp. 271–287, 2017.

[193] G. Vogiatzis and C. Hernández, "Practical 3d reconstruction based on photometric stereo," in Computer Vision: Detection, Recognition and Reconstruction. Springer, 2010, pp. 313–345.

[194] H. Dow, M. Perry, S. Pennada, R. Lunn, and S. Pytharouli, "3d reconstruction and measurement of concrete spalling using near-field photometric stereo and yolov8," Automation in Construction, vol. 166, p. 105633, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0926580524003698

[195] H. Dow, M. Perry, J. McAlorum, and S. Pennada, "Concrete crack pixel-level segmentation: a comparison of scene illumination angle of incidence," e-Journal of Nondestructive Testing, 2024.

Bibliography

[196] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," _Journal of Big data_, vol. 3, no. 1, pp. 1–40, 2016.

[197] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" _Advances in neural information processing systems_, vol. 27, 2014.

[198] X. Liu, M. Chi, Y. Zhang, and Y. Qin, "Classifying high resolution remote sensing images by fine-tuned vgg deep networks," in _IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium_. IEEE, 2018, pp. 7137–7140.

[199] Z. Yang, J. Yue, Z. Li, and L. Zhu, "Vegetable image retrieval with fine-tuning vgg model and image hash," _IFAC-PapersOnLine_, vol. 51, no. 17, pp. 280–285, 2018.

[200] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," _Advances in neural information processing systems_, vol. 25, 2012.

[201] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," _arXiv preprint arXiv:1412.6980_, 2014.

[202] J. Tan, J. Yang, S. Wu, G. Chen, and J. Zhao, "A critical look at the current train/test split in machine learning," _arXiv preprint arXiv:2106.04525_, 2021.

[203] N. W. S. Wardhani, M. Y. Rochayani, A. Iriany, A. D. Sulistyono, and P. Lestantyo, "Cross-validation metrics for evaluating classification performance on imbalanced data," in _2019 International conference on computer, control, informatics and its applications (IC3INA)_. IEEE, 2019, pp. 14–18.

[204] J. Shen, Q. Li, and G. Erlebacher, "Hybrid no-reference natural image quality assessment of noisy, blurry, jpeg2000, and jpeg images," _IEEE Transactions on Image Processing_, vol. 20, no. 8, pp. 2089–2098, 2011.

218

Bibliography

[205] E. Eldarova, V. Starovoitov, and K. Iskakov, "Comparative analysis of universal methods no reference quality assessment of digital images," 2021. [Online]. Available: https://libeldoc.bsuir.by/handle/123456789/45954

[206] J. König, M. D. Jenkins, M. Mannion, P. Barrie, and G. Morison, "Weakly-supervised surface crack segmentation by generating pseudo-labels using localization with a classifier and thresholding," IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 12, pp. 24 083–24 094, 2022.

[207] G. Kolappan Geetha, H.-J. Yang, and S.-H. Sim, "Fast detection of missing thin propagating cracks during deep-learning-based concrete crack/non-crack classification," Sensors, vol. 23, no. 3, p. 1419, 2023.

[208] A. L. Ramos, J. Domingo, and D. P. Y. Barfeh, "Analysis of weiner filter approximation value based on performance of metrics of image restoration," in 2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE). IEEE, 2020, pp. 1–6.

[209] S. Srinivasan, R. Rujula Singh, R. R. Biradar, and S. Revathi, "Covid-19 monitoring system using social distancing and face mask detection on surveillance video datasets," in 2021 International Conference on Emerging Smart Computing and Informatics (ESCI), 2021, pp. 449–455.

[210] Takmanman, "Gaussian noise and gaussian filter," https://takmanman.github.io/2020/05/23/gaussian-noise-gaussian-filter.html, 2020.

[211] P. E. Robinson and Y. Roodt, "Blind deconvolution of gaussian blurred images containing additive white gaussian noise," in 2013 IEEE International Conference on Industrial Technology (ICIT), 2013, pp. 1092–1097.

[212] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," IEEE Transactions on Image Processing, vol. 26, no. 7, pp. 3142–3155, 2017.

Bibliography

[213] Z. Yue, H. Yong, Q. Zhao, D. Meng, and L. Zhang, "Variational denoising network: Toward blind noise modeling and removal," Advances in neural information processing systems, vol. 32, 2019.

[214] C. P. Taylor and P. J. Bex, "On the number of perceivable blur levels in naturalistic images," Vision research, vol. 115, pp. 142–150, 2015.

[215] L. St, S. Wold et al., "Analysis of variance (anova)," Chemometrics and intelligent laboratory systems, vol. 6, no. 4, pp. 259–272, 1989.

[216] Y.-S. Lin, P.-H. Huang, and Y.-Y. Chen, "Deep learning-based hepatocellular carcinoma histopathology image classification: Accuracy versus training dataset size," IEEE Access, vol. 9, pp. 33 144–33 157, 2021.

[217] W. Hammouch, C. Chouiekh, G. Khaissidi, and M. Mrabti, "Crack detection and classification in moroccan pavement using convolutional neural network," Infrastructures, vol. 7, no. 11, p. 152, 2022.

[218] C. Kim, J. Lee, S. Jang, M. Lee, and M. Yang, "Image analysis technique for quantifying fluorescein concentration profiles in clays," Journal of Hydrology, vol. 635, p. 131180, 2024.

[219] I. A. Al-Sudani and G. A. Al-Suhail, "A comparative study of vgg16 and vgg19 for oil spill detection in coastal imagery," in 2024 1st International Conference on Emerging Technologies for Dependable Internet of Things (ICETI). IEEE, 2024, pp. 1–8.

[220] H.-Y. Chen and C.-Y. Su, "An enhanced hybrid mobilenet," in 2018 9th International Conference on Awareness Science and Technology (iCAST). IEEE, 2018, pp. 308–312.

[221] O. Ukwandu, H. Hindy, and E. Ukwandu, "An evaluation of lightweight deep learning techniques in medical imaging for high precision covid-19 diagnostics," Healthcare Analytics, vol. 2, p. 100096, 2022.

Bibliography

[222] V. Pandit and R. Bhiwani, "Image fusion in remote sensing applications: A review," International Journal of Computer Applications, vol. 120, pp. 22–32, 06 2015.

[223] W. Wang, X. Yuan, Z. Chen, X. Wu, and Z. Gao, "Weak-light image enhancement method based on adaptive local gamma transform and color compensation," Journal of Sensors, vol. 2021, pp. 1–18, 2021.

[224] S. Dorafshan, R. J. Thomas, and M. Maguire, "Sdnet2018: An annotated image dataset for non-contact concrete crack detection using deep convolutional neural networks," Data in brief, vol. 21, pp. 1664–1668, 2018.