SUKKUR IBA UNIVERSITY

Department of Computer Systems Engineering

MACHINE LEARNING BASED SUBSURFACE DEFECT DETECTION IN CFRP USING OPTICAL PULSE THERMOGRAPHY

Submitted By:

Zain Ul Abidin

Habeeban Memon

Ayesha Siddiqua

Supervised By:

Dr. Junaid Ahmed

Submitted in partial fulfillment of the requirements for the degree of

BACHELOR OF COMPUTER SYSTEMS ENGINEERING

Department of Computer Systems Engineering

Sukkur IBA University

Sukkur, Sindh, Pakistan

2026

CERTIFICATE

The thesis entitled **Machine Learning Based Subsurface Defect Detection in CFRP Using Optical Pulse Thermography** has been written by Zain Ul Abidin, Habeeban Memon and Ayesha Siddiqua under the supervision of Dr. Junaid Ahmed. This report has been approved by all members of the Final Year Project committee and accepted by Head of Department of Computer Systems Engineering at Sukkur IBA University in partial fulfillment of the requirements for the degree of:

BACHELOR OF COMPUTER SYSTEMS ENGINEERING

Project Supervisor (Dr. Junaid Ahmed) Signature: ____________________

Date: _________

Head of Department Signature: ____________________

Date: _________

Internal Examiner Signature: ____________________

Date: _________

External Examiner Signature: ____________________

Date: _________

PREFACE

This thesis is submitted in partial fulfillment of the requirement for the degree of Bachelor of Computer Systems Engineering at Sukkur IBA University. The purpose of this project is to develop a machine learning-based system to automatically identify subsurface defects within the carbon fibre reinforced polymer (CFRP) structure by the use of optical pulse thermography.

CFRPs have diverse applications, most notably in the aeronautical and wind turbine industries, which necessitate the inspection of the materials' inner integrity due to the structural significance. Here, we use two separate methods based on deep learning algorithms and physics-based concepts in order to accurately, automatically locate and determine the existence of subsurface defects.

This project was undertaken by three committed students Zain Ul Abidin, Habeeban Memon and Ayesha Siddiqua who devoted themselves to its completion. The assistance given by our project supervisor, Dr Junaid Ahmed in all aspects of this project has been invaluable and he is gratefully acknowledged for this.

The thesis contains five chapters: Chapter 1 introduces the problem; Chapter 2 presents a review of the related work; Chapter3 illustrates the detection and segmentation approach; Chapter 4 covers process of depth estimation and the 3D reconstruction; and 5 presents our conclusions and suggested future work.

Sukkur IBA University
Zain Ul Abidin
Habeeban Memon
Ayesha Siddiqua

DECLARATION

We hereby certify that this report of the Final Year Project is our own work. It has not been submitted in support of any application for a degree or qualification to any other college or university. All the references, sources and materials used in this report are clearly cited.

The databases used here are well-known and standard research benchmark data sets, they were referenced in the thesis.

Zain Ul Abidin Roll No: ____________

Date: ____________ Signature: ____________

Habeeban Memon Roll No: ____________

Date: ____________ Signature: ____________

Ayesha Siddiqua Roll No: ____________

Date: ____________ Signature: ____________

ACKNOWLEDGEMENT

First and foremost we would like to acknowledge and thank our supervisor Dr. Junaid Ahmed for his motivation, cooperation, direction and guidance during the entire duration of the project. We are indebted to his valuable suggestions that guided us throughout this project and enabled us to comprehend the problem and complete the work to the required extent.

We extend our gratitude to all the faculty and staff of the Department of Computer Systems Engineering, Sukkur IBA University for all the facilities and assistance rendered during our project.

We extend our gratitude to our families, for their encouragement, understanding and patience during our study.

Finally we express our thanks to researchers for sharing their CFRP thermography dataset freely to public, especially Erazo-Aux et al. and Ahmed et al. whose benchmark datasets were used in this experiment.

RESEARCH ETHICS FORM

We declare that this thesis is our original work submitted in partial fulfillment of the requirements for the degree of Bachelor of Engineering in Computer Systems Engineering at Sukkur IBA University. The work submitted here has never been submitted at any other place or for any other degree or award.

We declare that we have strictly followed the research ethics guidelines set by Sukkur IBA University. We have not included any human participants or animal subjects in this research, and we have only used public CFRP benchmark datasets, that have been appropriately referenced in the thesis.

We accept the full responsibility for the accuracy and integrity of this work.

Zain Ul Abidin Signature: ____________________ Date: ____________

Habeeban Memon Signature: ____________________ Date: ____________

Ayesha Siddiqua Signature: ____________________ Date: ____________

ABSTRACT

The high strength to weight ratio of Carbon Fiber Reinforced Polymer (CFRP) materials leads to their increasing application within aerospace and engineering industries. However, defects within CFRP structures, particularly delaminations, that exist on a sub-surface level, and thus aren't obvious by sight, can be detrimental to the structure itself. Automatically and accurately detecting them poses significant difficulties.

This project presents a machine learning pipeline which performs the automatic detection, segmentation, and depth estimation of defects in CFRP from optical pulse thermography (OPT). The unique three-channel input representation for the deep learning networks is the primary contribution of this work. Raw thermal intensity is combined with a first principle component analysis (PCA) and a thermographic signal reconstruction (TSR) to provide significantly more meaningful information for the network.

Six different models, three each for defect detection (YOLOv8n, RT-DETR-L, Faster R-CNN) and segmentation (U-Net, FPN, DeepLabV3+), are examined. The three-channel input improves the accuracy of the detection task by between 4-11% and the dice score of the segmentation task for the circular defect by between 14-51%, in comparison to solely inputting raw images, with several algorithms that completely failed to detect defects with raw inputs, performing perfectly well with the three-channel input.

In order to perform the depth estimation, sixteen features based on the physics behind the thermal response of defects were extracted, and XGBoost achieved the best performance with a mean absolute error (MAE) of 0.056 mm. Ultimately, a three-dimensional visualization of detected defects was produced in approximately three to five seconds per specimen.

Keywords: CFRP, optical pulse thermography, deep learning, defect detection, segmentation, depth estimation, XGBoost, 3D reconstruction.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Background and Motivation

Carbon fibre reinforced polymer (CFRP) composite materials are the predominant engineering structure materials of the 21st century. With a unique blend of low density and excellent specific stiffness, these are materials of choice whenever mass is at a premium, whether as a component part of an aircraft wing, racing car chassis, wind turbine blade or indeed the body of an electric vehicle battery enclosure. Boeing's 787 Dreamliner aircraft is approximately 50% composite material by structural weight [1], with the A350 XWB approaching 53% [2]. Wind turbine rotors frequently reach in excess of 80 metres, and the cumulative fatigue loads these structures must endure throughout their 25-year life is precisely the loading profile that CFRP's are ideal for [3].

With this extension comes a fundamental inspection challenge. It is possible for serious internal damage to exist in a CFRP structure, yet show no indication on the surface. An impact of moderate severity, one which is barely visible on the surface will cause a network of delaminations through the thickness of the laminate that will result in a structure which is less resistant to further loads than would ever be detected on the surface alone [4]. Steel or aluminium cracks will generally form on a plane, which become obvious to view prior to the material failing, whereas cracks may be distributed over several interlaminar planes within a CFRP laminate, making the three-dimensional nature of the damage very poorly identified by surface inspection [5]. An inspection technique needs to be capable of viewing these subsurface features at adequate resolution and also at a rate acceptable for an industrial maintenance regime.

The implications are considerable. The world MRO market for aircraft parts is more than USD 80bn per year and is growing rapidly, where inspection of composites by non-destructive techniques is one large component [6]. Aircraft grounded due to failure to detect will cost significant money and, in extreme cases, have consequences for safety.

Non-destructive testing offers a range of technologies with varying strengths. Ultrasonic testing is the dominant technology in aviation due to sub-millimetre resolution of internal structures, but point inspection requiring acoustic coupling and low inspection area throughput makes it slow over large surface areas [7][8]. X-ray CT yields high-resolution volumetric information, but requires costly, difficult-to-implement radioactive infrastructure and long acquisition times that are impractical for routine on-production-line deployment [9]. In contrast, Optical Pulsed Thermography, a relatively mature non-contact, coupling agent-free technology, covers a surface area of hundreds of square centimetres in a single flash, and its resulting spatial and temporal dataset can be used to determine both location and depth of defect [10][11]. In this form of NDT, a strong pulse of light is delivered to the surface of the component. Sub-surface flaws, being of lower thermal conductivity than the surrounding solid material, cause a region of slowed surface temperature decay at the defect location, detectable using an infrared camera.

Analysis of thermal images is not a trivial undertaking.Classical image processing techniques, such as pulse phase thermography [12], thermographic signal reconstruction [13], and principal component thermography [14] enhance defect contrast in the raw thermal sequences to provide an alternative representation of them. The size and location of defects is not definitively revealed, nor is the lateral size nor depth, by the output of current systems: the human analyst is needed to interpret it. Such a level of human input may become unsustainable as manufacturing volume increases and inspection becomes more time-consuming.

Deep learning methods have taken over image analysis in nearly all engineering applications, and they have been widely applied in the domain of thermographic CFRP inspection in recent years [15],[16]. Despite this expansion of use, nearly all applications published follow an architecture that places a hard limit on the possible outcomes: they input the raw, single channel thermal frames directly into the type of networks used for analysis of natural color images and fail to exploit any temporal features in the images. No pipeline has to date been found that can exploit spatial intensity, variability structure and decay over time as a multi-channel input to a network; nor has a single, unified system been described which combines, within one workflow, both pixel-level defect detection and segmentation along with a physics-informed calculation of defect depth. This thesis directly addresses those limitations.

1.2 CFRP Composites: Structure and Failure Modes

These mechanical properties are attained in CFRP composites by combining continuous carbon fibres in a polymer matrix, with epoxy being the preferred polymer for structural applications.Tensile modulus can range from approx. 230 GPa (normal-modulus fiber) up to 900 GPa (ultra-high-modulus fiber) in the direction of the fiber [17]. The directional nature of the properties dictates that structural laminates are constructed using successive layers (plies) of unidirectional prepregs at a series of angles. Quasi-isotropic sequence stacking such as [0°/90°/±45°]s are typical where balanced in-plane loading is required. Curing is typically carried out in an autoclave at temperatures of 120 to 180C under pressures ranging from 0.6 to 0.8 MPa [18].

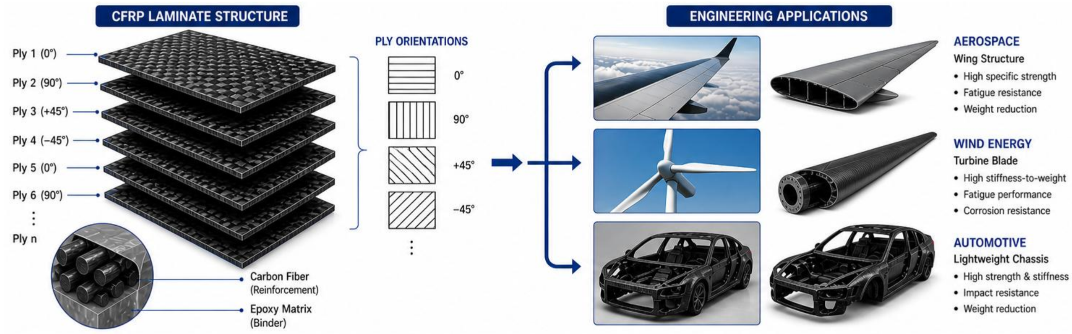


Figure 1.1: CFRP Laminate Structure and Engineering Applications

Each stage of the manufacturing process presents potential to initiate defects. Contamination at the layup stage will produce interface weaknesses that are only visible under load within the laminate. Inadequate consolidation pressure will leave the trapped volatile byproducts in the form of voids which can be difficult to control with increasing laminate thickness [19]. Uneven resin flow during cure is liable to create localized resin-rich areas and high heating rates can cause residual stresses through thermal gradients. Additional damage is risked during post-cure machining, unless cutting parameters are controlled. During service the history of cumulative damage will grow through impact events, fatigue cyclic loading, moisture absorption and temperature cycling over many thousands of flight cycles [20].

The critical failure mechanism within composites is that of interlaminar delamination. Delamination, which is essentially a failure of the bond between neighboring plies, can occur as a result of manufacturing flaws or through service loads and has the critically dangerous characteristic of not necessarily having a visible surface signature [4]. The engineer trying to assess the extent of a 0.2mm deep delamination is, effectively, dealing with a qualitatively different defect to a 2mm deep delamination, and this depth dependence is of crucial importance [5],[20].

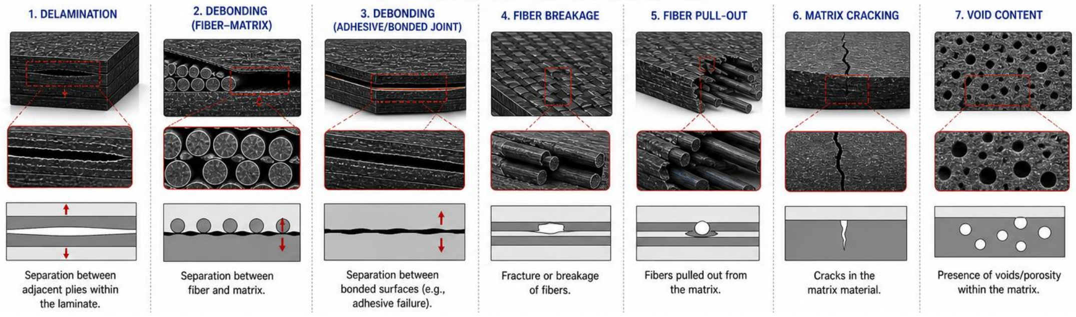


Figure 1.2: Internal Defects in CFRP Composites

Table 1.1: Comparative mechanical properties of CFRP, aluminium alloy, and structural steel

Material	Density (g/cm³)	Tensile Strength (MPa)	Young's Modulus (GPa)	Specific Strength (kN·m/kg)
CFRP (Standard Unidirectional)	1.5 - 1.6	1500 - 2500	130 - 150	~1250
Aluminum Alloy (7075-T6)	2.8	570	72	~204
Structural Steel (A36)	7.8	400 - 550	200	~64

1.3 Non-Destructive Testing Approaches

Ultrasonic testing is the dominant NDT technique in aerospace composites and has been for many decades [7]. This technique uses high frequency acoustic waves directed into the part, and detects internal flaws using reflection and transmission changes in the signal with depth resolution on the order of 0.1mm [8]. However, acoustic coupling necessitates a fluid medium, either water bath or gel, and it is a serial scanning process meaning large part throughput can be an issue. Radiography techniques, both standard X-ray and computed tomography, can provide a good amount of internal structural data but involve significant costs and acquisition times [9], and parallel defects relative to the x-ray beam (i.e. Delaminations) have a very low contrast,

The principle of pulsed optical thermography is distinct. A very short duration and high-energy heat pulse (typically of 2-20 ms, generated by xenon flash lamps) is applied to the surface of the sample and the transient cooling down process is detected by a high speed IR camera with frame rate varying between 50 and 400 Hz in 3-5 µm spectral window [10],[11]. The physics is that of 1-dimensional heat conduction. The temperature distribution inside a semi infinite solid is governed by Fourier's law [21] after being heated up at the surface:

$$\frac{\partial^2 T}{\partial y^2} = \frac{1}{\alpha}\frac{\partial T}{\partial t} \tag{1.1}$$

Where T is temperature (K), y is depth below the surface (m), t is time (s) from the flash, and α is the through-thickness thermal diffusivity of the material (m^2/s).

For a flash of surface heat input energy Q(J/m^2) deposited instantly, the solution to Equation 1.1 is: [21]

$$T(y,t) = \frac{Q}{e} \cdot \frac{1}{\sqrt{\pi \alpha t}} \exp \left(-\frac{y^2}{4\alpha t}\right) \tag{1.2}$$

where Q is the deposited surface energy density in J/m^2, $e = \sqrt{k\rho c}$ is the thermal effusivity of the material in $J/(m^2 \cdot K \cdot s^{0.5})$, k is thermal conductivity in $W/(m \cdot K)$, ρ is density in kg/m^3, and c is specific heat capacity in $J/(kg \cdot K)$. Evaluating Equation 1.2 at the surface gives the characteristic surface temperature decay [13]:

$$\Delta T(t) \propto t^{-0.5} \tag{1.3}$$

represents the rise of the transient surface temperature with respect to the pre-flash temperature. The $1/\sqrt{t}$relationship has already been introduced with regards to the signal processing algorithms in Chapter 2 and is central to the development of the depth estimation model in Chapter 4.

Table 1.2: Comparison of non-destructive testing methods for CFRP inspection

NDT Method	Operating Principle	Primary Advantages	Key Limitations
Ultrasonic Testing (UT)	High-frequency sound wave reflection	High depth resolution; industry standard for aerospace	Requires acoustic coupling (water/gel); slow for large areas
Eddy Current Testing	Electromagnetic induction	Good for near-surface defects and fiber orientation	Very limited penetration depth in composites
X-Ray / CT	Differential radiation absorption	High-resolution 3D volumetric data	Expensive; radiation hazards; poor contrast for planar delamination
Optical Pulsed Thermography	Thermal diffusion mapping	Fast, non-contact, large-area inspection	Contrast diminishes with depth; requires surface emissivity

1.4 Limitations of Current Methods

Classical thermographic processing algorithms: pulsed phase thermography [12], thermographic signal reconstruction [13], and principal component thermography [14] produce visual representations of sequences of thermal images where defects are made visible, but are preprocessing techniques rather than detection systems. Determining which structures within the resulting image represent genuine defects, tracing their boundaries, and estimating depth must still be carried out by a human expert. In time-sensitive scenarios with many components, reliance on a human-in-the-loop model has throughput and consistency limitations.

The development of deep learning presents an avenue toward automation [15],[22],[23]. Most reported approaches suffer from the limitation of the representations they feed the network: usually just a sequence of single channel thermal images, or basic time-averaged thermal images. The original thermographic data contains three physically different kinds of intensity variation: the physical thermal signature of a sub-surface defect, the spatially non-uniform thermal gradient produced by lamp geometry and surface emissivity, and the decaying of the surface temperature, containing depth information. Feeding all three as a single greyscale image requires the network to separate contributions which are distinct in their physical nature, often with an inadequate signal to noise ratio for shallow or deep defects [16]. Another significant flaw is task fragmentation in most published applications of deep learning to thermography: object detection pipelines output bounding boxes indicating defect locations [22]; segmentation networks output pixel-level contour maps [23]; depth is independently estimated using regression models. These problems are tackled in isolation, using distinct networks, despite the fact that they all make use of the same physical data.

1.5 Problem Statement

The precise gap this thesis tries to fulfill can be identified. All the proposed deep learning models for thermal inspection of CFRP rely on the raw single channel thermal frames for analysis. These frames inherently blend defect signal, background gradient due to emissivity and lamp position and temperature-dependent surface cooling which leads to difficulty in small or deep defect detection, localization and depth estimation. No model was found that combines raw thermal intensity, PCA coefficients and TSR polynomial coefficients into a single input signal or in other words, stacked the images, thus having all signals separated physically. No system has proposed an end-to-end workflow that comprises from pixel-level detection/segmentation to physics-informed depth estimation and finally 3-D reconstruction.

1.6 Research Objectives

There are 5 objectives this thesis is pursuing:

1. To study the effectiveness of the recent state-of-the art object detection and segmentation networks trained and tested on raw thermographic data.
2. To derive a multi-channel representation from the spatial and temporal information derived from thermographic image sequence.
3. To study the efficiency of proposed representation for defect detection and segmentation.
4. To explore a physics-inspired framework for estimating the subsurface defect depth from thermographic features and modeling approaches.
5. To build a three dimensional reconstruction of the defect using the segmented output and defect depth.

1.7 Research Contributions

- **Contribution 1:** A comprehensive systematic framework for comparing numerous deep learning based architectures designed for object detection and segmentation is introduced and presented while holding identical training scenarios and specimen level validation.
- **Contribution 2:** A comprehensive multi-channel thermographic representation which amalgamates original thermal images and temporal features derived thereof, to better characterize the defect is presented.
- **Contribution 3:** A physics-based feature based modeling framework is proposed for subsurface defect depth estimation from thermographic signals combined with machine learning approaches.
- **Contribution 4:** A complete end-to-end pipeline is constructed for combining detection, segmentation and depth estimation towards a 3-dimensional reconstruction and quantitative characterization of defect geometry.

1.8 Thesis Organisation

This thesis is organized as follows: Chapter 2 covers the theoretical background. Chapter 3 covers defect detection and segmentation using proposed three channel fusion were presented. Depth estimation and 3-D reconstruction are discussed in Chapter 4 with the proposed feature formulation, regression model, and reconstruction scheme. Final contribution, limitation, and future work were demonstrated in Chapter 5.

CHAPTER 2
LITERATURE REVIEW AND THEORETICAL FRAMEWORK

2.1 Introduction

This chapter reviews the evolution of thermographic NDT research and, more recently, the application of deep convolutional networks to inspection problems. Issues where methods to date have fallen short are also highlighted. From the fundamental physics of optical pulsed thermography, the discussion proceeds through the classical signal processing techniques to matrix decomposition, on through the recent deep learning architectures and ending in depth estimation techniques.

2.2 Thermodynamic Principles of Optical Pulsed Thermography

On heating a CFRP sample by a short burst of energy on its surface, the thermal energy spreads through the bulk of the material. For a homogenous semi-infinite solid with no interior boundaries, the decay of the surface temperature with time above any point would be given by the relationship described in Eq 1.3 and would follow the square root relationship with time after the heat flash. Subsurface discontinuities like a delamination or void, which have a lower thermal conductivity and diffusivity compared to the sound material, result in their immediate surroundings impeding the flow of heat and hence causing the surface temperature directly above these defects to lag behind those of the surrounding material and persist as a hot spot against the cooling background of the solid thermal anomaly before fading completely due to lateral heat diffusion filling in this retarded area.

The point at which the surface temperature above the discontinuity reaches its maximum value the peak contrast temperature, t_{peak}, is directly related to the depth of that discontinuity [24]:

$$d^2 \approx \beta \cdot \alpha \cdot t_{\text{peak}} \tag{2.1}$$

where d is the depth of the defect in metres, β is a geometrical factor in the range [1.5 - 2.0] determined by the ratio of the diameter to depth of the discontinuity [24] and α is the through-thickness thermal diffusivity in m²/s, is measured from the onset of heating by the flash. This relationship is thus rearranged into:

$$d \propto \sqrt{t_{peak}} \tag{2.2}$$

where the proportionality constant incorporates both β and α. Both of these parameters will vary with fibre direction, cure status and moisture content and can only be reliably calibrated if a specific sample is measured, thus these constants can be consolidated into a single parameter. This relationship is also the foundation for Feature 16 of the depth estimation feature set constructed in Chapter 4.

Large, shallow defects provide strong, early-peaking contrast while small, deep defects have weak, late-peaking contrast. Single frame images will either miss the defect if it has not peaked (shallow), or after the feature has lost all contrast (deep). It is the dynamic element that provides the information to locate and characterize defect depth

2.3 Classical Signal Processing Approaches

Approaches in response to this wealth of information contained in the raw thermogram sequence, classical methods were developed that perform a transformation on the raw sequence of thermal images to provide enhanced, and more time-independent, feature representations of subsurface discontinuities.

2.3.1 Pulsed Phase Thermography

Introduced by Maldague and Marinetti in 1996, pulsed phase thermography converts the time sequence of temperature values at each pixel into the frequency domain using a discrete Fourier transform [12]. The advantage of using phase rather than amplitude differences is its increased independence from surface heating variations, as the phase shift induced by subsurface discontinuities is a function of defect depth, not the overall temperature of the target.

The phase image at frequency f is defined as:

$$\phi(f) = \arg\ (\mathcal{F}\{T(t)\}(f)) \tag{2.3}$$

where $\phi(f)$ is the phase value in radians at frequency f in Hz, $\mathcal{F}\ \cdot$ is the discrete Fourier transform of the thermal contrast data for each pixel at time t, and $\arg(\cdot)$ is the phase angle of a complex number. Lower frequencies are attenuated the most as heat is absorbed by the material but will result in images of greater spatial extent, while higher frequencies result in sharper features at the surface but reduced depth penetration [25].

Increasing the range of frequencies is able to reveal defects at different depths but 300-2000 frames are required in order to create sufficiently high resolution frequency spectrums, thus limiting throughput in an industrial application [26].

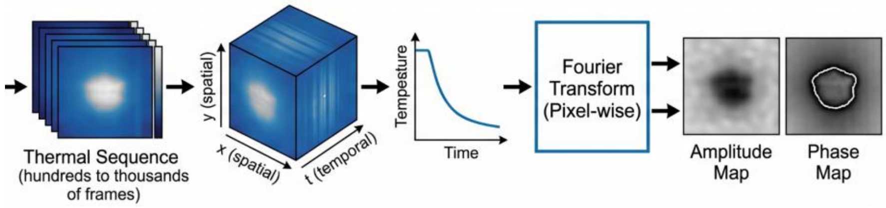


Figure 2.1: Pipeline of PPT illustrating a typical thermal sequence and pixel-wise discrete Fourier transformation [26]

2.3.2 Thermographic Signal Reconstruction

Developed TSR in the early 2000's utilizing the $t^{-0.5}$ decay law directly [27]. On taking the natural log of surface temperature, this power-law decay is linearized. Deviations from the linear trend in log time reveal the defect as a departure from this ideal behaviour, and are fit with the following polynomial:

$$\ln\left(\Delta T(t)\right) = \sum_{n=0}^{N} a_n \left[\ln\left(t\right)\right]^n \tag{2.4}$$

where $\Delta T(t)$ is the surface temperature difference at time t after the flash, the a n are the fitted polynomial coefficients of degree n, and N is the polynomial order (which is often between 4 and 7 for CFRPs [27]). The low-order coefficients give the general rate of cooling whereas higher-order coefficients contain information about non-standard behavior resulting from a defect or variation in material property [28].

A valuable practical advantage of fitting polynomials to time sequence data is the compression of the data, where a frame sequence consisting of 2000+ frames can be reduced to 6 or 8 images of coefficient values, containing physically relevant information [28]. A significant drawback of this approach is that important thermal data may fall outside the time window of the polynomial fit.

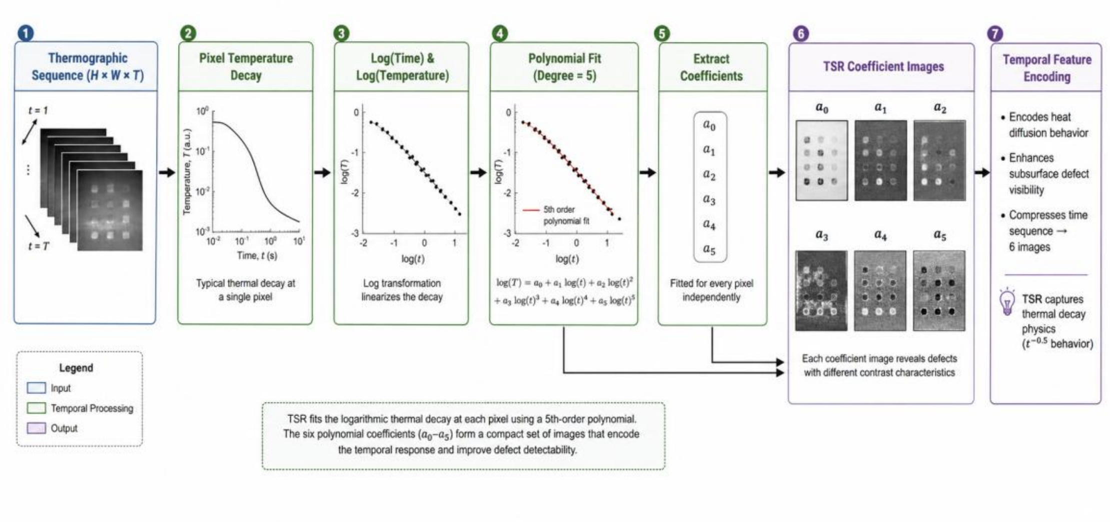


Figure 2.2: Workflow for TSR processing with pixel-wise polynomial fitting [27]

2.3.3 Principal Component Analysis for Thermography

Principal component analysis applied to the raw thermographic data was formally applied in NDT by Rajic [14] as an improvement to prior work performed by Winfree [29]. Essentially the data (a series of temporal images) is transformed into a 2-D data matrix and then the method applies Singular Value Decomposition to produce a series of orthogonal spatial patterns ordered in terms of their importance as a fraction of total data variance:

$$X = U\Sigma V^T \tag{2.5}$$

where $X \in \mathbb{R}^{M\times N}$ is the data matrix with M rows of time-frames and N of pixel values (M and N may be swapped in different implementations) U is an orthogonal matrix where columns are left singular vectors (temporal modes), Σ is a diagonal matrix containing the singular values, and V^T is an orthogonal matrix where rows are right singular vectors spatial modes [30].

Winfree demonstrated that using components second through fourth provided improved feature representations of subsurface defects [29], as the 1st component contains mostly of non-local (spatially varying over the field of view) cooling that obscures near-surface defects. PCA offers benefits of self-adaptation to the thermal scene, but also lacks in its inability to differentiate between defect signatures and those produced by surface artifacts (e.g. Emissivity variation or dirt). Analysis to determine which components carry defect signatures requires expert interpretation [14],[30].

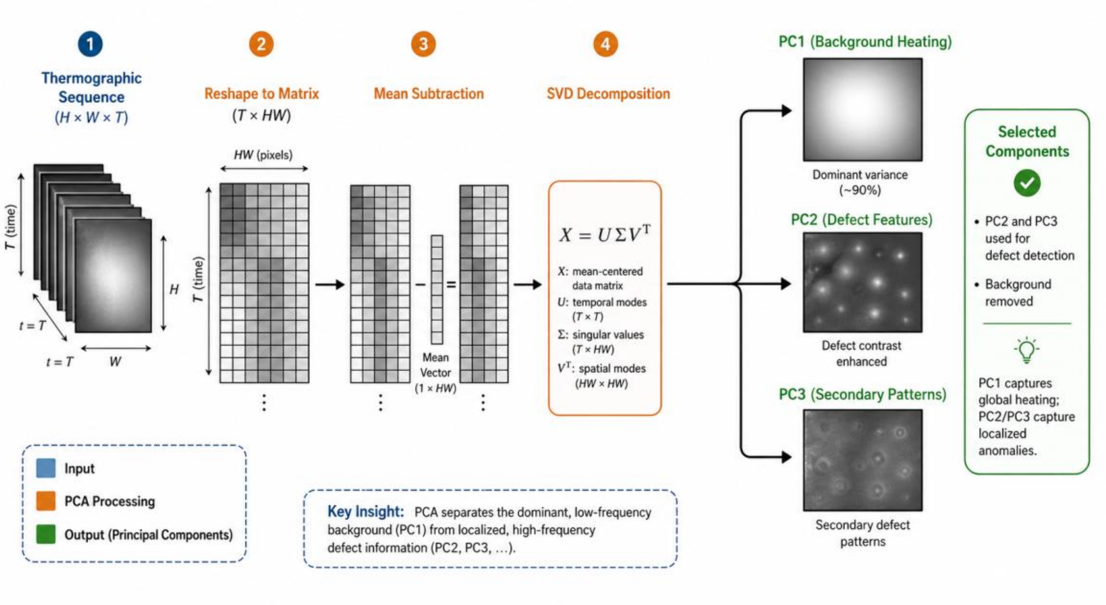


Figure 2.3: Illustration of the SVD applied to a thermal sequence of data [29]

2.4 Matrix Decomposition and Sparse Representation

Starting around 2015, approaches to this problem shifted more towards solving the problem as an inverse problem, i.e. Decomposing the termperature data into a highly structured, well-behaved "background" component and a sparse "defect" component, building upon work on sparse representation of data and robust statistics.

2.4.1 Robust Principal Component Analysis

Robust Principal Component Analysis RPCA seeks to recover a low-rank structure and a sparse component of the original data matrix. The standard problem is framed as:

$$\min_{L,S} \ \| L \|_* + \lambda \| S \|_1 \quad \text{subject to} \quad L + S = X \tag{2.6}$$

where $X \in \mathbb{R}^{P \times T}$ denotes the P pixels and T time frames observed thermal data, L denotes the low rank background matrix that represents the global thermal reaction and S represents the sparse anomalies that reveal the spatially confined defect [28,27]. The nuclear norm is denoted by $\|\cdot\|_*$ which is the sum of singular values and the norm by $\|\cdot\|_1$, respectively. λ represents the regularization parameter, which is always positive and controls the trade-off between minimizing rank and increasing sparsity. [31].

According to Liu et al, the sparse component of RPCA detects defects with greater signal-to-noise ratio than traditional PCA when defect sizes or contrasts are small [32]. It requires that the background response is global low-rank, and that the defects are sparse in space. These criteria will not apply when the defect is large, or for materials with a strong anisotropy.

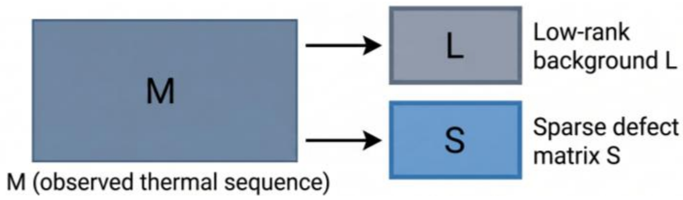


Figure 2.4: Matrix decomposition of the thermal matrix in R P C A where the background and defect components are separated [31]

2.4.2 Wavelet-Integrated Decomposition

Ahmed, Gao and Woo expanded RPCA by incorporating multiscale wavelets and dictionary learning into an alternative sparse decomposition framework [33]. The object function is given by:

$$\min_{B,P,Q} \left\{ s \parallel B \parallel_* + \frac{\phi_p}{2} \parallel P \parallel_F^2 + \frac{\phi_q}{2} \parallel Q \parallel_F^2 + \parallel X - B - PQ' \parallel_F^2 \right\} \tag{2.7}$$

where X is the input data in the wavelet domain; B is the low rank background component, PQ' are the learned dictionary factors, PQ' is the sparse component, s controls the rank of B and ϕ_p and ϕ_q are positive regularization parameters balancing the Frobenius-norm penalties of the dictionary factors P and Q, where the Frobenius-norm is denoted as $\|\cdot\|_F$ [33].

With experiments on 8 CFRP specimen with tricky curved surface, Ahmed et al report 98% F-scores for this method compare to other method that have from 48%-80% [33]. This high performance is because the dictionary learning adapted the specimen-specific thermal response whereas the others did not have any basis function adapting the data-specific signal characteristic. However, the cost of this alternating optimization algorithm, which has to run on three different wavelet scales is high in real-time application.

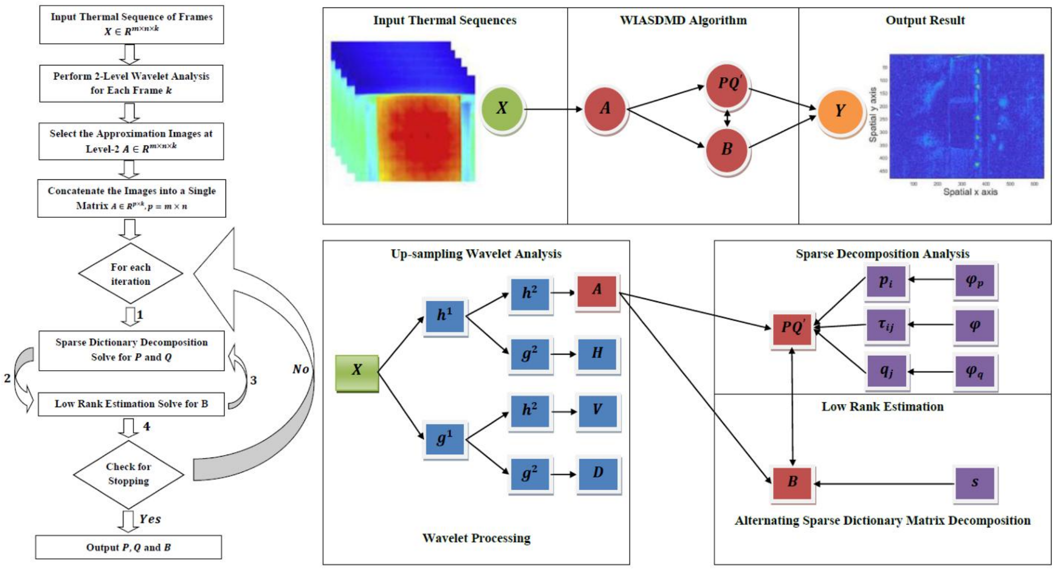


Figure 2.5: The dictionary learning integrated framework with wavelet [33]

2.5 Deep Learning Architectures for Object Detection

The applications of deep convolutional neural networks on thermographic defect detection have been increasing rapidly since 2015 which is consistent with the progress in computer vision field. Many of these architectures adapt models that were developed for natural image detection to thermographic inspections. Transfer learning using large pre-trained networks is utilized as a remedy for relatively small thermographic inspection datasets [34].

2.5.1 Two-Stage Detection: Faster R-CNN with Feature Pyramid Network

Faster R-CNN brought together the region proposal network and detection network in a single end-to-end trainable system [35]. This system comprises a shared convolutional backbone, a region proposal network which predicts candidate bounding boxes from anchored predictions, and parallel classification and regression heads operating on pooled region features using region of interest pooling. When coupled with a feature pyramid network backbone, a range of scales is represented in the features from a top-down pass with lateral connections [36]. Each pyramid level feature map P_k is formed by:

$$P_k = \mathrm{Conv}_{3\times3}\left(\mathrm{Up}(P_{k+1}) + \mathrm{Conv}_{1\times1}(C_k)\right) \tag{2.8}$$

where P_k is the pyramid feature map at level k, C_k is the bottom-up backbone feature map at level k derived from the convolutional encoder, $\mathrm{Up}(\cdot)$ is the upsampling operation by a factor of 2, $\text{Conv}_{1\times1}$ is the 1x1 convolution used to unify channel dimensions, and $\text{Conv}_{3\times3}$ is the 3x3 convolution used for spatial smoothing [36].

Lema et al. Coupled Faster R-CNN to the investigation of active thermography of CFRP panels and identified defects of approximately 94% precision while showing systematic failure of the detection for defects smaller than 3mm and more shallow than 0.5mm [37]. While this two-stage design provides decent precision, the frame rate is typically only 5-10fps and careful tuning of anchor box dimensions may be required for defects with anomalous aspect ratios.

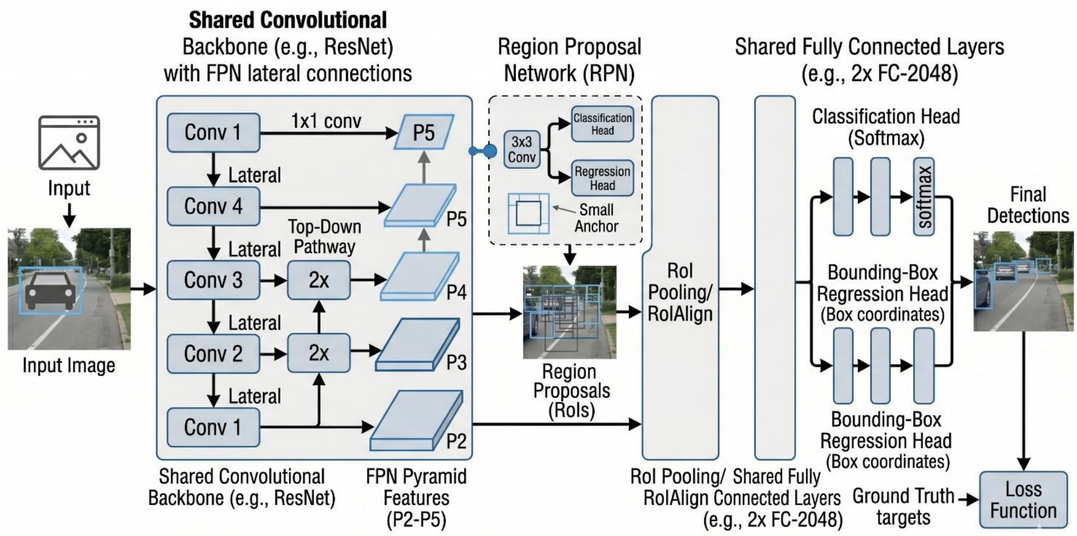


Figure 2.6: Two-stage Faster R-CNN with shared convolutional backbone [35]

2.5.2 Single-Stage Detection: YOLOv8

Single-stage methods reformulate the problem as dense regression on the input image to predict bounding box coordinates and class probabilities, which eliminate the proposal generation stage and dramatically increase the inference speed [22]. YOLOv8, which was published in 2023 by Ultralytics, employs a CSPDarknet backbone, a bidirectional feature pyramid network for multi-scale fusion and an anchor free detection head that predicts at each spatial location directly which helps reduce hyper-parameters and enables adapting unusual defect shapes [38].

The regression loss is a Complete Intersection Over Union (CIOU) loss for accurate bounding box localisation:

$$\mathcal{L}_{\text{box}} = 1 - \text{IoU} + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha_v v \tag{2.9}$$

where IoU is the intersection over union between the ground-truth box b^{gt}, and the predicted box ρ is the distance between the centers of the ground-truth box b^{gt}, and the predicted box b, c is the diagonal length of the smallest enclosing rectangle covering both boxes, v is a measure of consistency of aspect ratio, is α positive trade-off parameter [38].

According to Fang et al, YOLOv8 has about 91% mAP for synthetic thermography images but drops to about 76% on real CFRP specimen because of the domain shift between synthetic data and real thermography and the low defect contrast in raw thermal data [39].

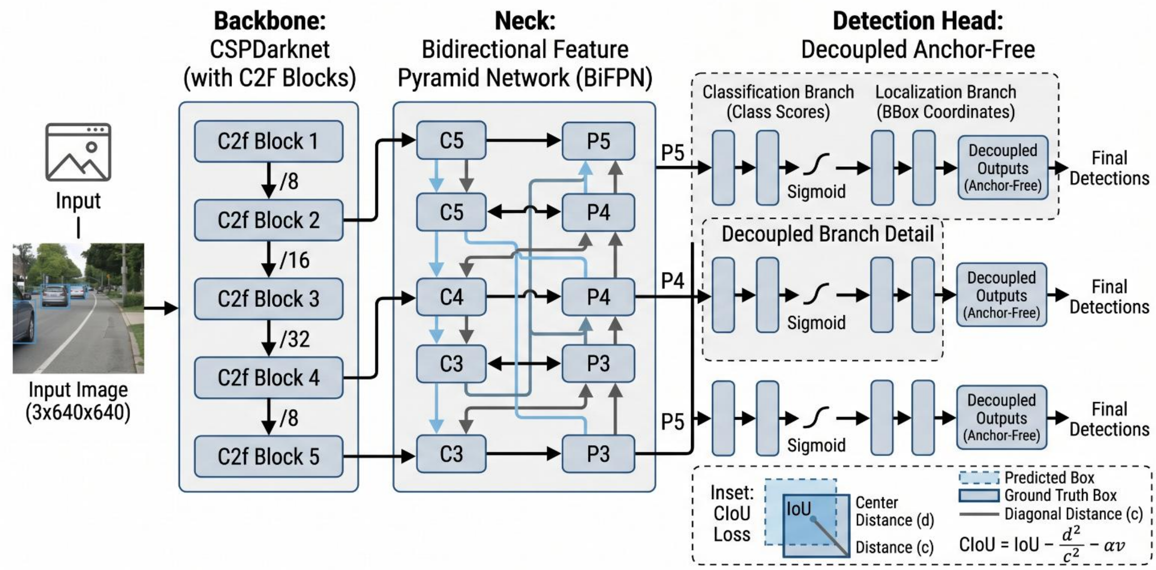


Figure 2.7: YOLOv8 architecture with CSPDarknet backbone [38]

2.5.3 Transformer-Based Detection: RT-DETR

By reformulating object detection as a direct set prediction problem, the encoder-decoder transformer of Carion and colleagues with learned object queries and Hungarian algorithm bipartite matching [40] does away with Non-maximum Suppression. This is especially important when defects are close together in a thermographic inspection, as it can lead to a true positive being removed in suppression-based post-processing.

The problem of slow convergence of the original DETR architecture, as well as its computational complexity, was tackled in Real-Time Detection Transformer, which implemented an encoder comprised of convolutional and transformer layers combined with an IoU-aware selection of queries to ensure rapid convergence [41].The transformer's decoder cross attention mechanism is shown below in Equation 2.10:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^{\top}}{\sqrt{d_k}}\right)V \tag{2.10}$$

where Q is matrix consists of object queries while K and V both consist of the encoder output features, d_k is a scaling factor and is equivalent to the key dimensionality [40].

Schmid and colleagues also demonstrated that, at least when applied to lock-in thermography of CFRP, global self-attention provided enhanced robustness to non-uniform heating compared to an entirely convolutional approach, although small training sets necessitated meticulous hyperparameter tuning [42].

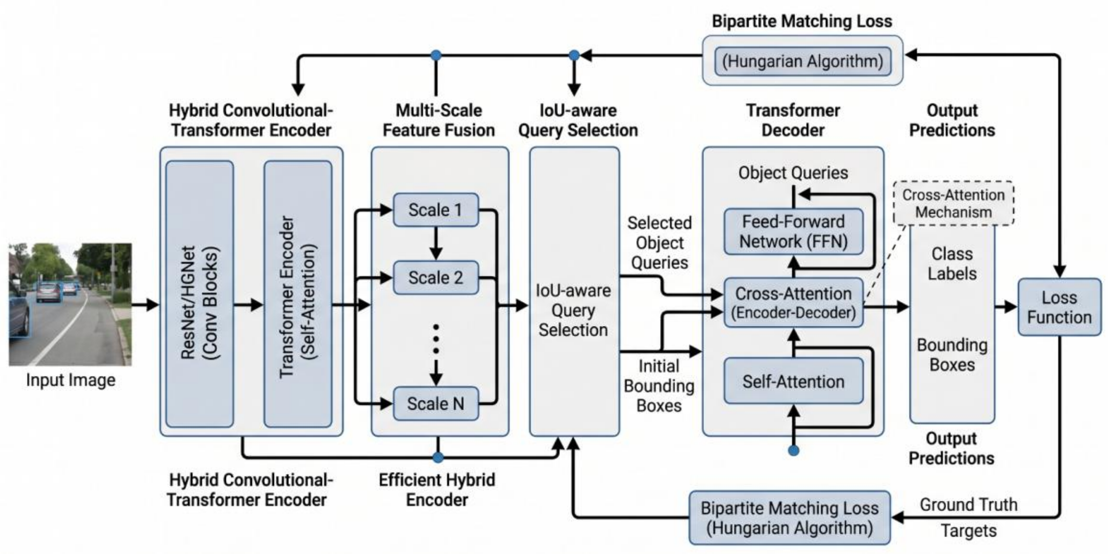


Figure 2.8: Hybrid Convolution-Transformer Encoder for RT-DETR [41]

2.6 Deep Learning Architectures for Semantic Segmentation

Semantic segmentation assigns each pixel in an image to its class label. This task is much more demanding as it requires networks that can extract multi-scale information from the image, while also preserving spatial resolution and producing context-consistent labels at object boundaries.

2.6.1 Encoder-Decoder Architecture: U-Net

U-Net utilizes a symmetric encoder-decoder architecture with skip connections to join the encoder path with the decoder path, allowing for the propagation of high-resolution encoder features to the upsampled decoder stages [23]. A recent modification has made the encoder a much deeper and powerful network through using the pretrained ResNet networks as the backbone [34] that significantly improves the quality of features due to residual connections and allow for much deeper network training.

The skip connection at decoder level l concatenates the upsampled decoder feature map and the corresponding encoder feature map:

$$D_l = \text{Concat}\,(\text{Up}(D_{l+1}),\, E_l) \tag{2.11}$$

where D_l is the decoder feature map at the l, D_{l+1} is the feature map at level from a deeper decoder level that has been upsampled to match spatial dimensions with level l, E_l is the encoder feature map with matching spatial resolution, Concat$(\cdot)$ is the channel-wise concatenation, Up$(\cdot)$ stands for the upsamping by bilinear or transpose convolution [23].

Zhou et al adopted U-Net architecture with a ResNet34 as backbone in IR T of CFRP, they achieved 92% Dice for larger defects with size larger than 10mm in horizontal direction but only 67% Dice for smaller defect with size less than 5mm likely due to spatial information lose during sequential pooling [43]. The U-Net architecture can be effectively trained on small datasets but the concatenation of skip connection can propagate the low-level noise in heating artifact of the encoder's feature.

2.6.2 Multi-Scale Pyramid

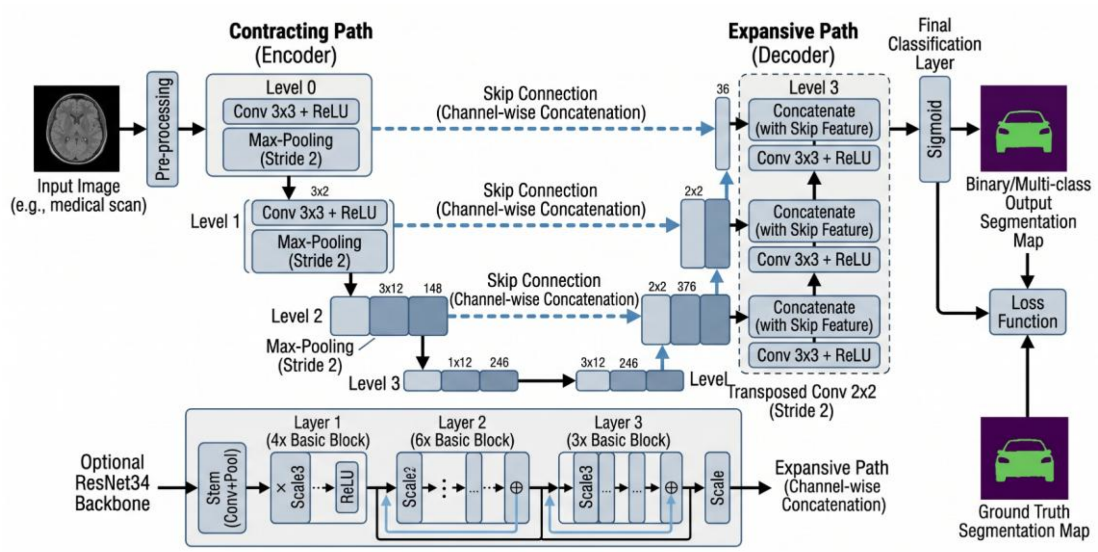


Figure 2.9: U-Net symmetric encoder-decoder architecture [23]

2.6.2 Multi-Scale Pyramid: Feature Pyramid Network

The feature pyramid networks create multi-scale features using a bottom-up pathway (backbone) coupled with a top-down pathway (semantics propagating from higher level features to lower level features) via lateral connections [36]. However, unlike the skip connections used by U-Net which use concatenation, feature pyramid networks add channel-aligned features using element-wise addition.

$$P_k = \mathrm{Conv}_{3\times 3}\left(\mathrm{Up}(P_{k+1}) + \mathrm{Conv}_{1\times 1}(C_k)\right) \tag{2.12}$$

where P_k, C_k, and operators are same as defined in equation 2.8. Lema et al., have used the feature pyramid network with ResNet50 backbone for thermographic inspection of CFRP specimens and achieved IoU scores of 88% at almost 15ms per image [37].

The addition approach uses less computation compared to concatenation but has a tendency of slightly blur the fine spatial details and the relatively lightweight decoder produces less precise boundary lines for small or structurally complex defects.

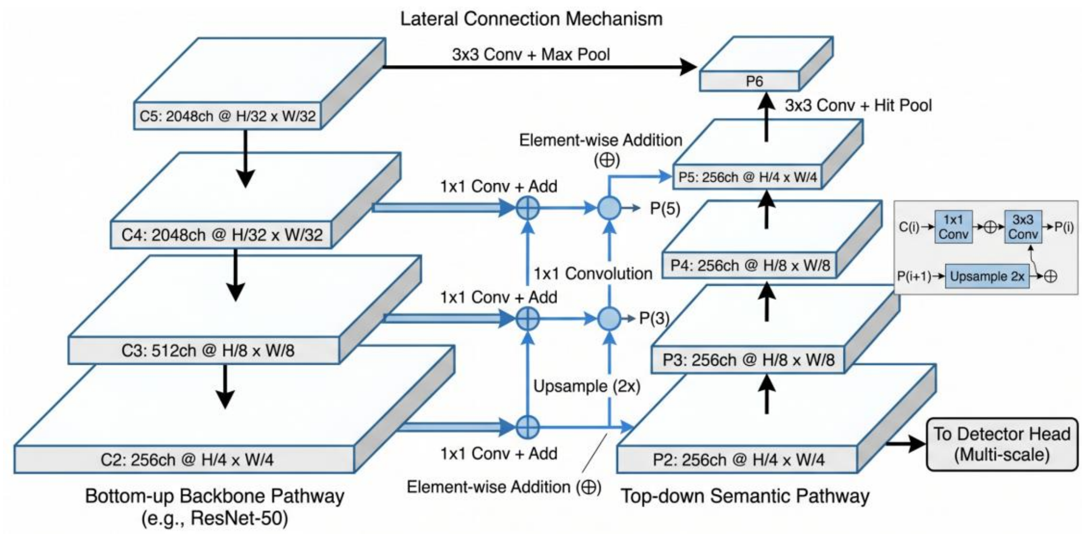


Figure 2.10: Feature pyramid network architecture [36]

2.6.3 Atrous Spatial Pyramid Pooling: DeepLabV3+

DeepLabV3+ introduces atrous spatial pyramid pooling (ASPP) which combines a cascade of dilated convolutions at different dilation rates and also uses the global average pooling, which captures context at different scales without any reduction in the feature map resolution [44]. The decoder selectively uses low-level features with their higher resolution representation to improve accuracy on edge pixels, resulting in overall segmentation improvement. Atrous convolution at a dilation rate of r is defined as:

$$y[i] = \sum_{k} x[i + r \cdot k] \cdot w[k] \tag{2.13}$$

where $y[i]$ is the output feature value at location i, $x[\cdot]$ is the input feature map, $w[\cdot]$ denotes the filter weights and r controls the rate at which the filter values are sampled.

In the ASPP module, the features are extracted at rates $r \in \{6, 12, 18\}$ and concatenated together, followed by pooling features using global average pooling and projecting into a low dimensional feature vector, which is then used to refine the coarse output of the encoder using the lightweight decoder. Morelli et al. have used DeepLabV3+ with ResNet101 backbone for lock-in thermography of composite materials and found Dice coefficients of nearly 94%, with good robustness to thermal distortions due to heating [45].

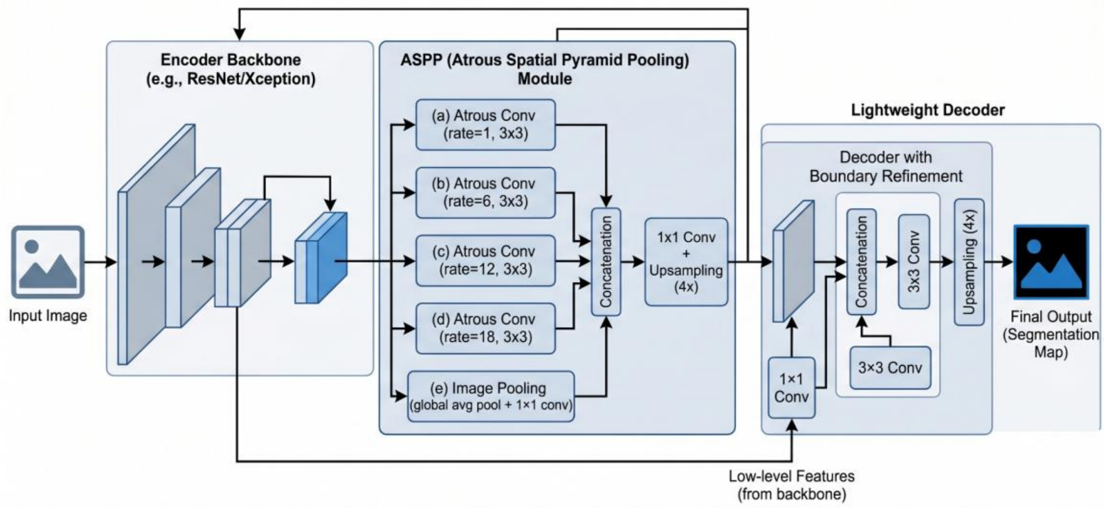


Figure 2.11: DeepLabV3+ architecture [44]

2.7 Hybrid Preprocessing and Deep Learning

Researchers have also proposed different methods of combining classical thermographic signal processing and deep learning architectures to use prior domain knowledge within the networks. Rosa et al., have performed TSR on the images before passing them on to a U-Net for segmentation of CFRP laminates and the training was done on polynomial coefficient images, without involving raw thermal frames. This approach resulted in IoU of 95% and F1 score of 99% by eliminating non-uniform heating gradients and focusing the network on boundaries [46].

Liu et al. have proposed a Generative adversarial network to synthetically generate realistic sequences with thermal defect and enhance their performance using physics-informed data augmentation thereby addressing the common problem of limited availability of labeled thermal data [47]. Hu et al, have developed a spatial-temporal fusion network that incorporates parallel branches for TSR and pulsed phase processing that were joined using learned attention weights, yielding an F1 score of ~89% for industrial CFRP inspection [48].

One major shortcoming of all hybrid methods is that there is only one-way communication, from classical signal processing to neural network and hence the parameters for the classical transform are fixed prior to training and are not optimized in tandem with the neural network parameters.

2.8 Depth Estimation from Thermographic Data

Whereas considerable work has been done on defect detection and segmentation, defect depth estimation has been studied comparatively less in spite of its crucial importance for structural integrity evaluation. Section 2.2 shows that $\sqrt{t_{\text{peak}}}$ based on this physical relation several works have developed quantitative systems for depth estimation.

Ekanayake et al. Performed experimental analysis onlock-in thermography (LIT) of CFRP laminates with Teflon inserts of known depth and obtained a coefficient of determination 0.91 between peak time and square of depth. The achieved prediction error was around ±0.15 mm for depths ranging from 0.5 to 2.0 mm; while for both shallow defects approaching flash duration and deep defects thermal contrast is smaller than noise floor the error was considerably larger.

Dong et al. Proposed a three dimensional residual network that analyzes thermal data (thermal cubes) through spatio-temporal volumetric convolutions. Their obtained mean absolute error and root mean square error were 0.18 mm and 0.24 mm on 450 samples. Their main drawback comes from the cubic relation of computational complexity with number of pixels of the thermal cube making the memory requirements and training time impractical for large amounts of data. Peng and Li applied random forest regressor to features extracted from the shape of the temporal response curve both temporal and geometrical features. On 100 to 200 training samples, mean absolute error was 0.21 mm and their method had the benefit of showing the most influential features.

Wei et al. Combined instance segmentation to get the points constituting the defect and random forest regression on the identified point clouds. On curved specimens, the mean surface distance error compared to CT data was 0.32 mm [51].

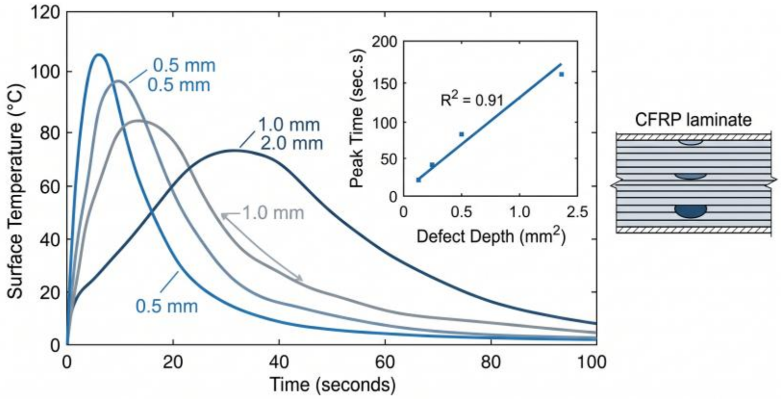


Figure 2.12: Physical depth-time relationship for thermographic depth estimation

2.9 Identified Research Gaps

Four research gaps have collectively motivated this thesis.

- There is no principled way to incorporate several complementary preprocessing methods into a single multi-channel representation for end-to-end learning.
- Lack of standardized baseline comparisons across detection/segmentation architectures using single, per-specimen validation settings.
- Detection, segmentation, and depth are mostly seen as separate tasks limiting combined optimization and deployability.
- Limited available datasets have ground truth segmentation masks and reliable depth measurements and lack consistency of train/test pipelines.

These observations suggest the approach taken in Chapters 3 and 4.

CHAPTER 3
THREE-CHANNEL FUSION AND DEFECT DETECTION & SEGMENTATION

3.1 Introduction

In this Chapter, a three-channel thermographic representation approach is proposed, whereby one representative image of each of the three information sources is generated, channels are stacked into an $H \times W \times 3$ tensor which is input into the deep learning network as it would be with pre-trained RGB networks. This work then, focuses on a cyclic alignment scheme to synchronize the entire sequence to six processed images per specimen, the formation of a three channel image tensor, and the extensive testing of this representation on six detection and segmentation architectures under three-fold cross-validation on a per specimen basis. Figure 3.1 presents the proposed methodology.

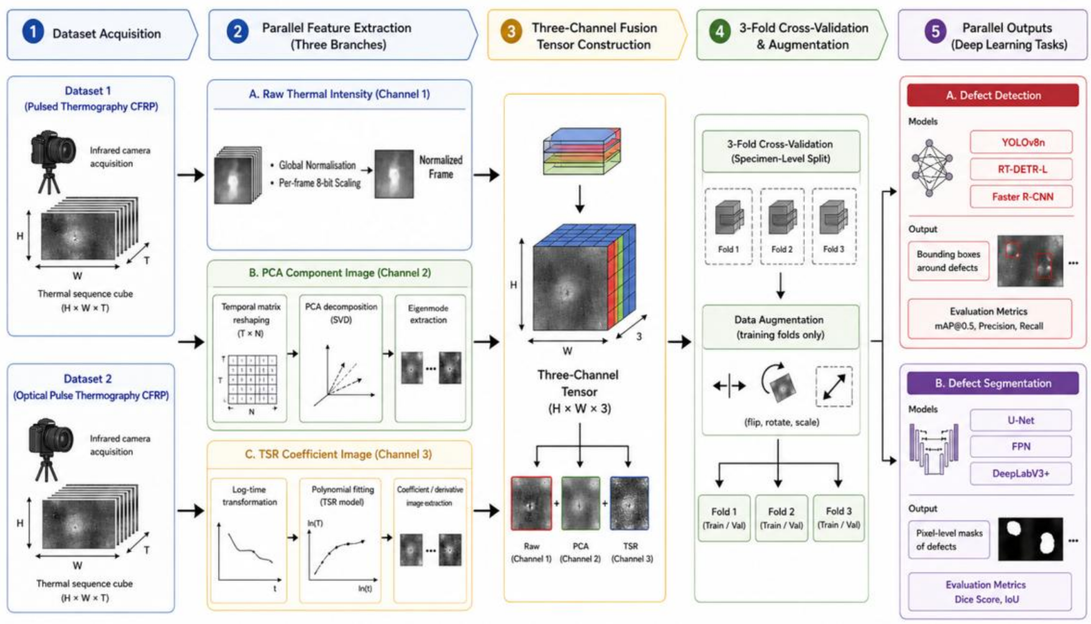


Figure 3.1: Proposed pipeline from raw thermographic image acquisition to creation of three channel tensor, three fold cross validation and parallel detection/segmentation results.

3.2 Experimental Specimens and Acquisition

Six CFRP samples from two publicly available data sets are used for these experiments. These data sets were acquired using different hardware setups, heating mechanisms, spatial resolutions and sequence lengths in order to validate the proposed method on a broader variety of parameters.

3.2.1 Dataset 1

This dataset was acquired by Erazo-Aux et al. [52] to establish a reference for pulsed thermography performance for CFRP composite materials. The experiments used a FLIR X6900sc InSb cooled infrared camera to capture images at a 512 x 512 resolution and 14-bit radiometric depth, with a sampling rate between 120 and 145 frames per second. Two BALCAR FX60 flash units produced a pulse energy of 12.8 kJ and a 2 ms pulse, with a camera mounted 100 cm from the specimen and 50 cm from the flashes at an angle of 45 degree to the normal of the surface. A temporal sequence lasted around 16-17 seconds producing 2000 images. Only front-surface sequences were used for these experiments, resulting in possible defect depths of 0.2 mm-1.0 mm.

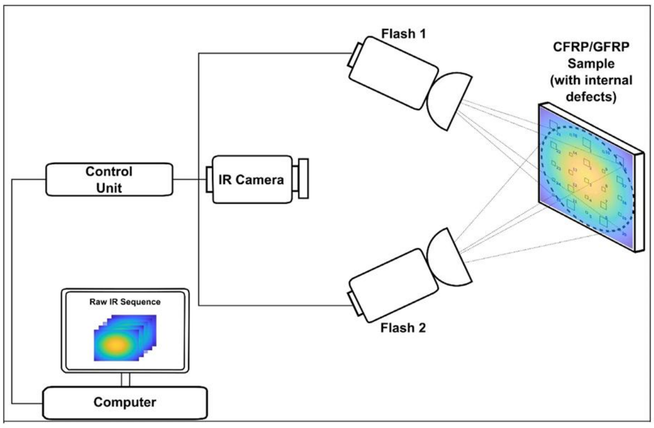
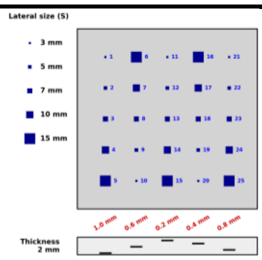
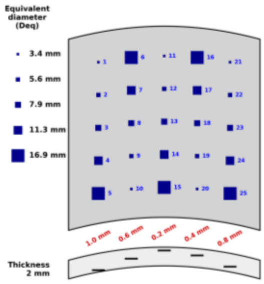
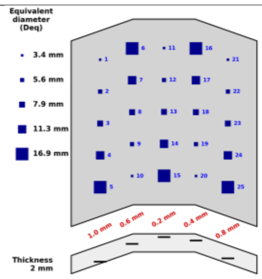


Figure 3.2: Pulsed thermography experimental setup from Erazo-Aux et al. [52].

Three specimens were chosen from this data set, labeled SQ01, SQ_02 and SQ03, which corresponded to specimens CFRP006, CFRP007 and CFRP008 in the original publication. The specimens are 2 mm thick plates 300 x 300 mm in size which were fabricated by autoclave curing of unidirectional prepreg plies with a quasi-isotropic layup. Twenty-five square Teflon inserts which have depths between 0.2 mm and 1.0 mm. These inserts form a 5 x 5 grid; the columns correspond to nominal depths from left to right of 1.0, 0.6, 0.2, 0.4 and 0.8 mm respectively, therefore allows automatic defect depth prediction from column position. SQ_01 represents a flat panel, SQ_02 a curved one, and SQ_03 a trapezoidal one, and this variation allows method evaluation with different degrees of surface-heating disparity.

Table 3.1: Dataset 1 specimen characteristics.

ID	Specimen Image	Geometry	Dimensions (mm)	Sizes (mm)	Depth (mm)	Frames
SQ_01		Flat	300 x 300 x 2	3, 6, 9, 12, 15	0.2–1.0	2000
SQ_02		Curved	300 x 300 x 2	3, 6, 9, 12, 15	0.2–1.0	2000
SQ_03		Trapezoidal	300 x 300 x 2	3, 6, 9, 12, 15	0.2–1.0	2000

3.2.2 Dataset 2

The second data set was provided by Ahmed, Gao, and Woo [33], and employs surface heating with a halogen lamp. The temperature variations of three specimens were recorded by an uncooled Vanadium Oxide (VoX) microbolometer (FLIR A655sc) at a spatial resolution of 640 480 pixels and 50 Hz. The sequence lengths were 729, 465 and 383 frames. The defects are characterized by circular inserts of varying sizes and are located 2.0 mm or 2.2 mm below the surface, not 0.2-1.0 mm as in dataset 1.

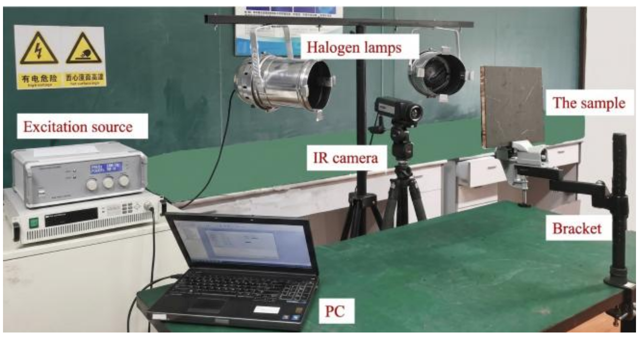
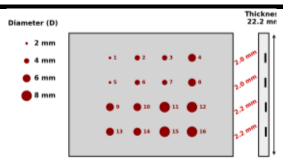
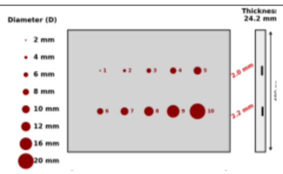
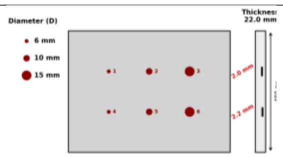


Figure 3.3: OPT setup in reflection mode from Ahmed et al. [33].

The three samples have labels CI_01, CI_02 and CI_03, which corresponded to samples 3, 1 and 2 in the original publication [33]. CI_01 is a 250 x 250 x 22.2 mm sample containing 16 defects with diameters 2, 4, 6 and 8 mm at 2.0 and 2.2 mm depths. CI_02 is 250 x 250 x 24.2 mm and contains 10 inserts from 2-20 mm diameter at same depths. CI_03 (450 × 300 × 22.0 mm) has 6 defects ranging from 6 mm to 15 mm at the same two depths. It is important to note that sample CI_01 presented the greatest challenge as the 2 mm circular defects are at the resolution limit of the camera.

Table 3.2: Dataset 2 specimen characteristics.

ID	Specimen Image	Geometry	Dimensions (mm)	Diameters (mm)	Depth (mm)	Frames
CI_01		Flat	250 x 250 x 22.2	2, 4, 6, 8	2.0–2.2	729
CI_02		Flat	250 x 250 x 24.2	2–20	2.0–2.2	465
CI_03		Flat	450 x 300 x 22.0	6, 10, 15	2.0–2.2	383

3.3 Data Preprocessing and Three-channel Fusion

3.3.1 Channel 1: Raw Thermal Frames

The thermal data from each sample is stored in a MAT file and consists of a 3D tensor $X \in \mathbb{R}^{H \times W \times T}$. A global normalisation is applied to all spatial pixels across all frames.

$$X_{norm} = \frac{X}{\max\limits_{i,j,t} X_{i,j,t}} \tag{3.1}$$

where $X_{norm} \in [0,1]^{H \times W \times T}$ is the globally normalized tensor, $X_{i,j,t}$ is the raw detector count at pixel (i, j) and time frame t, H and W are the number of pixels in the image dimensions and T is the total number of frames per specimen. The global normalization helps in preserving the spatial shape of the temporal decay, required for the TSR polynomial fitting discussed in section 3.3.3. Each frame is independently normalized and converted to 8 bits.

$$F_{norm}(i,j,t) = \left\lfloor \frac{X_{norm}(i,j,t) - \min\limits_{i,j} X_{norm}(\cdot,\cdot,t)}{\max\limits_{i,j} X_{norm}(\cdot,\cdot,t) - \min\limits_{i,j} X_{norm}(\cdot,\cdot,t)} \cdot 255 \right\rfloor \tag{3.2}$$

where $F_{norm}(i,j,t)$ is the 8-bit intensity at pixel (i,j) at frame t, $\lfloor \cdot \rfloor$ denotes the floor operation and spatial min and max are computed independently for each frame t.

3.3.2 Channel 2: Principal Component Analysis

PCA decomposes the globally normalized tensor as the sum of spatially independent, orthogonal component images ordered by variance explained [25], [30]. The temporal by pixel matrix $B \in \mathbb{R}^{T\times HW}$ is created by reshuffling X_{norm}. The mean of each column is then subtracted before applying economy SVD.

$$B = U\Sigma V^T \tag{3.3}$$

where U represents the spatial modes and Σ represents the temporal modes. Each right singular vector v_k is reshaped into $P_k \in \mathbb{R}^{H\times W}$. In this case, there are T right singular vectors, however we typically only consider the first m vectors where $m = 6$ since the first few explain the majority of the variance.

PC1 covers 85-92% variance and it describes primarily the slowly varying background temperature throughout the sample. Other components, PC2-PC6, provide more subtle thermal changes, usually, PC2 and PC3 describe subsurface defect related features. No component is chosen exclusively, instead a multi-componentPCA stack (PC1-PC6) is used. An adaptive selection strategy is then used, based on defects detectability for each specific sample. Independent min-max normalization on 8-bit is done for each selected component(s). For more about PCA in thermography see [25], [30].

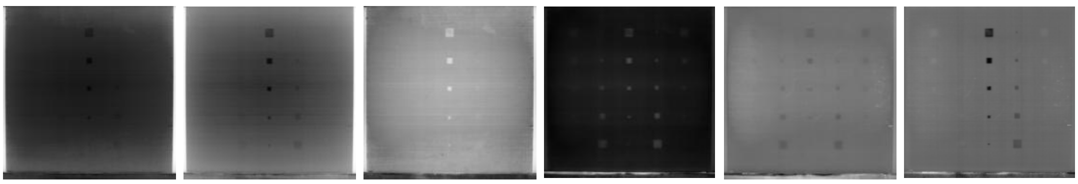


Figure 3.4: PCA component images PC1 to PC6 for sample SQ_01.

3.3.3 Channel 3: Thermographic Signal Reconstruction

TSR was introduced by Shepard et al. [27], [28] based on the observation that the surface temperature decay after flash is well represented by a $t^{-0.5}$ curve. A logarithmic polynomial is fitted to the pixel temperature decay which allows us to condense the time sequence into $n+1$ images of polynomial coefficients (where n is the polynomial order). At each pixel, the function to fit is:

$$\ln\left(T_{i,j}(t)\right) = a_0 + a_1 \ln(t) + a_2(\ln(t))^2 + \cdots + a_n(\ln(t))^n \quad (3.4)$$

where $T_{i,j}(t)$ is the surface temperature at pixel (i, j) and time t, and $a_0, \ldots, a_n$ are polynomial coefficients obtained by least squares fit. We use a polynomial of order $n = 5$, as per [27], [28], yielding 6 coefficient images, A_0 through A_5 per specimen. The first and second time derivatives of this fitted polynomial with respect to log-time can be derived analytically, and tend to highlight subsurface defects more effectively than the coefficients themselves. We use all six coefficients A_0 to A_5 images to represent Channel 3. All image data are min-max normalized to 8-bit grayscale independently. See [27], [28] for the full mathematical treatment.

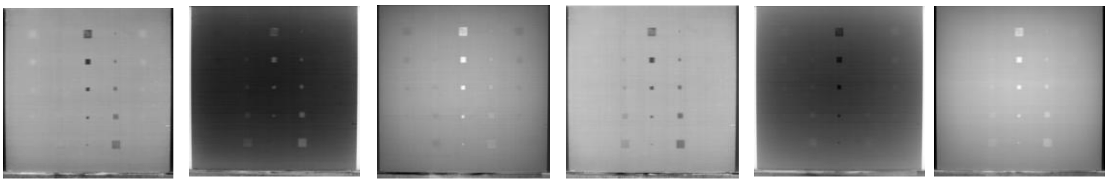


Figure 3.5: TSR coefficient images A_0 to A_5 for sample SQ_01.

3.3.4 Proposed Three-Channel Fusion and cyclic alignment

Our proposed 3-channel image input is created by concatenating one image from each of the 3 sources discussed in 3.3.1 through 3.3.3. For a given pixel (i, j), the image channels are combined to form a single column vector:

$$\mathbf{X}_{fusion}(i,j) = [I_{raw}(i,j),\ I_{PCA}(i,j),\ I_{TSR}(i,j)]^T \quad (3.5)$$

where $I_{raw}(i,j)$ is the per-frame normalized intensity, $I_{PCA}(i,j)$ is the chosen PCA component image, and $I_{TSR}(i,j)$ is the chosen TSR coefficient image. Combining the vectors from all pixels forms an $H \times W \times 3$ tensor which is analogous to an RGB image and can be passed through any standard pre-trained backbone without architectural modification [34].

There is a disparity in the number of frames in the raw sequence (T) versus the pre-processed images (m). We solve this by creating cyclic pairings between the raw frame sequence and the pre-processed images. Given raw frame index i, the index into the m pre-processed image sequence is given by:

$$c_i = ((i - 1) \ mod \ m) + 1, i \in \{1,2, \dots, T\} \tag{3.6}$$

where $m = 6$ is the number of pre-processed images from each channel and mod is the modulo operation. The i-th 3-channel image frame is formed by concatenating the i-th frame from the normalized raw sequence and the c_i-th images from the PCA and TSR sequences.

$$\mathbf{I}_{3ch}(i) = F_{norm}(\cdot,\cdot,i) \ | \ \mathcal{R}(P_{c_i}) \ | \ \mathcal{R}(A_{c_i}) \in \mathbb{R}^{H \times W \times 3} \tag{3.7}$$

where $F_{norm}(\cdot,\cdot,i)$ is the i-th frame of normalized raw data, P_{c_i} is the c_i-th PCA component image and A_{c_i} is the c_i-th TSR coefficient image. $\mathcal{R}\{\ \}$ denotes resampling to $H \times W = 512 \times 512$ pixels (using bilinear interpolation) and $|$ denotes concatenation across the channel axis.

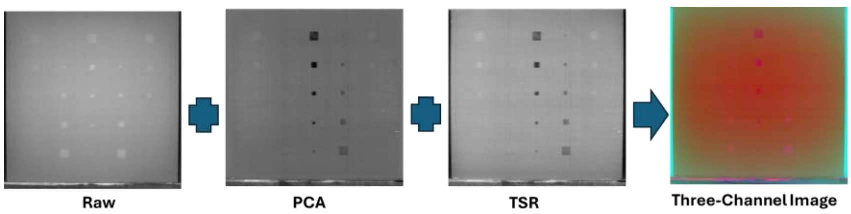


Figure 3.6: Example three-channel composite showing Channel 1 (raw frame), Channel 2 (PCA), and Channel 3 (TSR) for sample SQ_01.

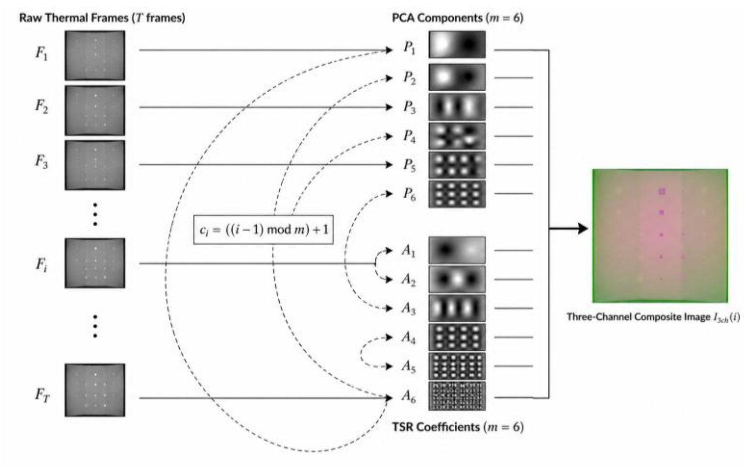


Figure 3.7: schematic illustrating cyclic linking of raw frames to PCA/TSR images at index c_i.

A summary of the 3-channel construction procedure is provided in Algorithm 3.1.

Algorithm 3.1: Three-channel Thermal Image construction

1. **Input:** $X \in \mathbb{R}^{H\times W\times T}$(raw thermal tensor); frame rate f_s [Hz]; polynomial degree $n = 5$; number of preprocessed images $m = 6$
2. **Step 1 (Global Normalisation):** Compute $X_{norm} = X/\max_{i,j,t}(X_{i,j,t})$
3. **Step 2 (PCA):** Reshape X_{norm}t o time-by-pixel matrix $B \in \mathbb{R}^{T\times HW}$; subtract temporal mean; apply economy SVD [Equation 3.3]; reshape first m right singular vectors to $H \times W$; min-max normalise each component $P_k$to 8-bit
4. **Step 3 (TSR):** Build log-time vector via $\tau_k = \ln(k/f_s)$ for $k = 1, \dots, T$; form polynomial design matrix $\Phi \in \mathbb{R}^{T\times(n+1)}$; for each pixel (i, j), compute $\mathbf{a}_{ij} = (\Phi^T\Phi)^{-1}\Phi^T\mathbf{y}_{ij}$where $\mathbf{y}_{ij} = \ln(X_{norm}(i,j,:))$; min-max normalise each coefficient image $A_p$to 8-bit
5. **Step 4 (Frame Selection):** Apply Equation 3.2 to each of the T raw frames
6. **Step 5 (Cyclic Stacking):** For $i = 1$to T: compute $c_i = ((i-1) mod m) + 1$; concatenate $F_{norm}(\cdot,\cdot,i)$, $\mathcal{R}\{P_{c_i}\}$, $\mathcal{R}\{A_{c_i}\}$along the channel axis
7. **Output:** $\boldsymbol{I}_{3ch} \in \mathbb{R}^{T\times H\times W\times 3}$

3.3.5. Baseline Input Preparation and Training Consistency

For the raw single-channel baseline, each normalized thermal frame (Channel 1 only) was converted into a three-channel pseudo-color image using the parula colormap (MATLAB default). Thus, all networks receive the same $H \times W \times 3$ input, ensuring a fair comparison with the proposed three-channel fusion tensor. Crucially, both the raw (parula-mapped) data and the proposed three-channel fusion data were trained using exactly the same architectures, identical three-fold specimen-level cross-validation splits, the same data augmentation schemes, and identical hyperparameters).

3.3.6 Dataset Partitioning, Cross-Validation, and Augmentation

For to prevent data leakage, specimen-level partitioning is used. The images or frames-level partitioning could not be used for the following two reasons: first, the augmentation of each source frame generates five duplicates, therefore, splitting after augmentation would result in the same frames being allocated to both the training and test sets, hence inflated test accuracy score will reflect frame recognition, rather than the generalisation. Second, successive frames of thermographic sequence share almost identical thermal data. A random frame-level split could therefore mean duplicates being present within both partitions. Splitting at the specimen-level eliminates both.

As six specimens are present, only a single fixed train/test split would lead to a performance estimate that would be dependent on which specimens were assigned to the two partitions. Three-fold cross-validation is employed with one square-insert and one circular-insert specimen in each fold so as to ensure a balanced geometry representation over all three partitions. Every specimen will then serve as training data in one partition, validation data in one partition and test data in one partition. The three-fold cross-validation is illustrated over the following rotation diagram.

Data augmentation is only performed on the training data. Each image and associated binary mask will generate 5 additional versions. The originals: horizontal flip, vertical flip, a random rotation over $U(-25, +25)$, a random scaling over $U(0.75,1.1)$. Images use bilinear interpolation for resampling while masks use nearest neighbor interpolation, to preserve the exact defect boundaries across both specimens. In table 3.3 below there is indication as to the specimens that would be assigned to each fold, and the number of images in each respective set.

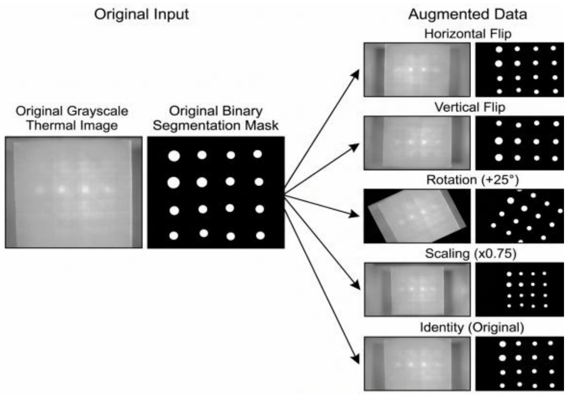


Figure 3.8: Five augmented versions of a sample training frame from the SQ_01 specimen.

Table 3.3: Sample assignments and number of training images for the 3-fold cross-validation.

Fold	Training Specimens Raw (SQ + CI)/ aug×5)	Validation Specimens Raw (SQ + CI)	Test Specimens Raw (SQ + CI)
1	SQ_01, CI_01 (2000 + 729) x 5 = 13645	SQ_02, CI_02 (2000+465) = 2465	SQ_03, CI_03 (2000 + 383) = 2383
2	SQ_03, CI_03 (2000 + 383) x 5 = 11915	SQ_01, CI_01 (2000 + 729) = 2729	SQ_02, CI_02 (2000+465) = 2465
3	SQ_02, CI_02 (2000+465) x 5 = 12325	SQ_03, CI_03 (2000 + 383) = 2383	SQ_01, CI_01 (2000 + 729) = 2729

3.3.7 Ground Truth annotation

Manual annotation for segmentation masks was performed using LabelMe tool. PC2 or PC3 images were used to annotate, as these images are best representatives of the boundaries of the defects for both geometries of defect specimen. Only one binary mask per specimen is rasterized to $H \times W$ so only 6 individual masks. Detection bounding boxes are then directly extracted from the segmentation masks, representing the axis-aligned bounding box around the component. The format chosen for the bounding boxes were COCO JSON for Faster R-CNN and RT-DETR-L, and YOLO text for YOLOv8n.

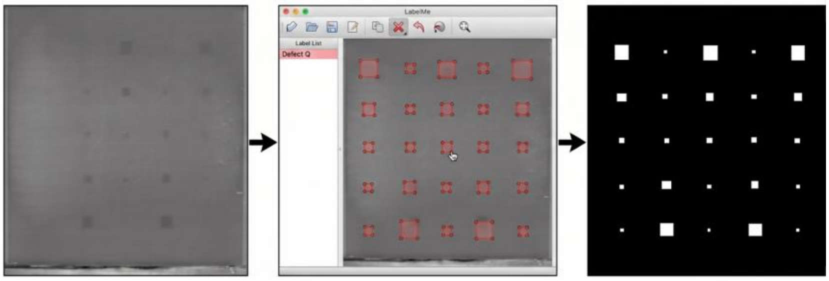


Figure 3.9: Ground truth workflow, including using PCA as the reference image, manual annotation. and final binary mask for SQ_01.

3.4 Defect Detection: Methodology and Results

3.4.1 Detection Training Configuration

Training experiments were performed on an NVIDIA Tesla P100 GPU, containing 16 gigabytes of RAM, 13 gigabytes of system RAM, CUDA 11.8, cuDNN 8.6, PyTorch 2.0.1, and python 3.10. YOLOv8n and RT-DETR-L were trained using Ultralytics 8.0.196 and Faster R-CNN was trained using Detectron2 0.6. A fixed random seed value was set for all weight initializations, along with a fixed shuffle. Input data is re-sized to 576 x 704 using letterbox padding so that aspect ratio can be maintained and augmented by mosaic, random horizontal and vertical flips, color jitter (HSV range 0.015 hue, 0.7 saturation, 0.4 value), random scaling, and other augmentations as per Ultralytics defaults.

YOLOv8n is initialized with COCO pretrained weights, trained with AdamW [53] and learning rate of $\eta_0 = 0.001$, weight decay of 0.0005 and batch size 16, with the learning rate cosine annealing schedule over a maximum of 50 epochs:

$$\eta_t = \eta_{min} + \frac{1}{2}(\eta_0 - \eta_{min})\left(1 + \cos\frac{\pi t}{T_{max}}\right) \tag{3.8}$$

where t is the current epoch, η_0 is the initial learning rate, and $\eta_{min} = 1 \times 10^{-5}$ is the minimum learning rate and $T_{max}$is the maximum epoch count. The total detection training loss is:

$$\mathcal{L}_{det} = \mathcal{L}_{focal} + \lambda_{box}\,\mathcal{L}_{CIoU} \tag{3.9}$$

where $\mathcal{L}_{focal}$ is the focal loss and $\mathcal{L}_{CIoU}$ is IoU regression loss, penalises poor overlap in both center locations, and aspect ratios simultaneously and, $\lambda_{box} = 7.5$ regression loss weight.

RT-DETR-L is initialised from the COCO pretrained weights with an identical AdamW and cosine scheduler setup. To address the issue of cross-attention heads diverging during initial stages of training, a learning rate of 11×10^{-4} is applied to the transformer attention layers. The loss function for training employs a bipartite matching formulation:

$$\mathcal{L}_{DETR} = \sum_i \left[\mathcal{L}_{cls}\left(p_{\sigma(i)}, c_i\right) + \mathbf{1}_{c_i \neq \emptyset}\,\mathcal{L}_{box}\left(b_{\sigma(i)}, b_i\right)\right] \tag{3.10}$$

where σ represents assignment permutation provided by Hungarian algorithm, and $p_{\sigma(i)}$ represent the predicted probability of the class for the query matching the ground truth object i, $c_i$is the ground-truth class label, $b_{\sigma(i)}$and b_i that represent the predicted and ground truth bounding boxes, respectively, and an indicator function $\mathbf{1}_{c_i \neq \emptyset}$ activates when a ground truth object is matched to the query [40].

Faster R-CNN is trained through Detectron2 with SGD, momentum 0.9, weight decay $10^{-4}$and an initial learning rate of 2.5×10^{-4}, preceded by 1000 iterations of linear warmup. Total training iterations 3000, with validation occurring every 500. Anchor scales of 32, 64, 128, 256, 512 are used, with aspect ratios 0.5, 1, 2. RoI align with a 7×7 feature map is employed. Table 3.4 lists all parameters used.

Table 3.4: Training hyperparameters for detection models.

Parameter	YOLOv8n	RT-DETR-L	Faster R-CNN
Parameters (M)	3.2	32	41
Optimiser	AdamW	AdamW	SGD (mom=0.9)
$\boldsymbol{\eta_0}$	0.001	0.001 / 0.0001	0.00025

Weight decay	0.0005	0.0005	0.0001
Batch size	16	16	2
Max epochs / iterations	50 ep	50 ep	3000 it
LR schedule	Cosine	Cosine	Step (warmup)
Early stopping (patience)	10 ep	10 ep	-
Input resolution	576x704	576x704	576x704

The detection Algorithm 3.2 summarizes the detection training and early stopping process.

Algorithm 3.2: Detection Model Training and Early Stopping

1. **Input:** Training set $\boldsymbol{D_{tr}}$; validation set $\boldsymbol{D_{va}}$; model $\boldsymbol{f_{det}}$; maximum epochs $\boldsymbol{E_{max}}$; early stopping patience $\boldsymbol{P = 10}$; initial learning rate $\boldsymbol{\eta_0}$; weight decay $\boldsymbol{\lambda}$
2. Initialise: optimiser = AdamW($\boldsymbol{f_{det}}$, $\boldsymbol{\eta_0}$, $\boldsymbol{\lambda}$); scheduler = CosineAnnealingLR($\boldsymbol{E_{max}}$, $\boldsymbol{\eta_{min} = 10^{-5}}$); best_mAP = 0; wait = 0
3. for $\boldsymbol{e = 1}$to $\boldsymbol{E_{max}}$do
4.   **for** each batch $(\boldsymbol{X_b}, \boldsymbol{Y_b})$in $\boldsymbol{D_{tr}}$ do
5.   Compute predictions: $\boldsymbol{\hat{Y} = f_{det}}$(augment($\boldsymbol{X_b}$))
6.   Compute loss $\boldsymbol{\mathcal{L}_{det}}$using (YOLOv8n) or (RT-DETR-L) or RPN + RoI loss (Faster R-CNN)
7.   Update: optimiser.step(gradient of $\boldsymbol{\mathcal{L}_{det}}$)
8.   **end for**
9.   Scheduler step
10.   Compute validation mAP@0.5 on $\boldsymbol{D_{va}}$
11.   **if** val_mAP > best_mAP: save checkpoint; best_mAP = val_mAP; wait = 0
12.   **else:** wait += 1
13.   **if** wait ≥ $\boldsymbol{P}$: stop training
14. **end for**
15. Restore best checkpoint as $\boldsymbol{f^*_{det}}$
16. **Output:** Optimised detection model $\boldsymbol{f^*_{det}}$

3.4.2 Detection Evaluation Metrics

Detection performance is assessed using precision, recall, F1 score, and mAP@0.5. The definition of a true positive is that IoU between the predicted box and the ground truth box exceeds the threshold:

$$\mathrm{IoU}_{box} = \frac{|\ B \cap B_{gt}\ |}{|\ B \cup B_{gt}\ |} \geq 0.5 \tag{3.11}$$

where B is the predicted bounding box, B_{gt} is the ground-truth bounding boxwhere $|\cdot|$ is the number of pixels. The precision, recall and F1 score for the defects is then given by:

$$\text{Precision} = \frac{TP}{TP + FP}, \ \text{Recall} = \frac{TP}{TP + FN} \tag{3.12}$$

$$F_1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3.13}$$

where TP is the number of true positive defects, FP is the number of false positive defects, and FN the number of false negative defects. The average mAP@0.5 over all specimens is used as the main measure for detection and is averaged over the area under the precision-recall curve for each test specimen.

3.4.3 Detection Results and Discussion

The effect of using a three-channel input to the detection algorithms was analyzed by comparing results using both single channel inputs and three channel inputs, across all three different architectures and all three folds of the cross validation.

It is evident from the results that the use of three channels, as opposed to a single channel, yields improved results in all three detectors and all three folds. The highest average three-channel mAP was 0.95 by RT-DETR-L, closely followed by 0.94 by YOLOv8n and 0.89 by Faster R-CNN, a trend seen in all three folds. One striking aspect is the nature of failures using single channel input, where both RT-DETR-L and Faster R-CNN achieved zero detections on CI_01 in Fold 3, but regained performance when using three channel input, obtaining mAP=0.75. That the behavior across three architecturally dissimilar detectors appears the same suggests that input properties alone are driving performance gains.

YOLOv8n achieves raw mAP value of 0.89. Using a single channel the square specimen was detected at very high precision, >0.95 but the recall for the circular specimen was relatively poor at 0.75 due to the shallow inserts near the plate edge appearing with very poor contrast. The use of three channel input led to a significant boost in performance, with mAP rising to 0.94, an improvement of 5.6%, with circular recall up to 0.87. RT-DETR-L starts with a better baseline of 0.91 mAP under single channel input. As a transformer the model inherently had good robustness to non-uniform heating (the baseline for the 0.91 was due to this) but the global attention mechanisms do help push this to 0.95 and precision to 0.97, with a perfect 1.00 mAP on one specimen CI_02, in Fold 2. Faster R-CNN also performs poorly on the circular specimen in single channel input, reaching a lower raw baseline mAP of 0.84. As this is a two stage network, the first stage of proposing regions of interest filters out many candidates and errors are not recoverable at the second stage. Thus, providing it with higher quality input gives it the most gain in the detection phase: the mAP now reaches 0.89.

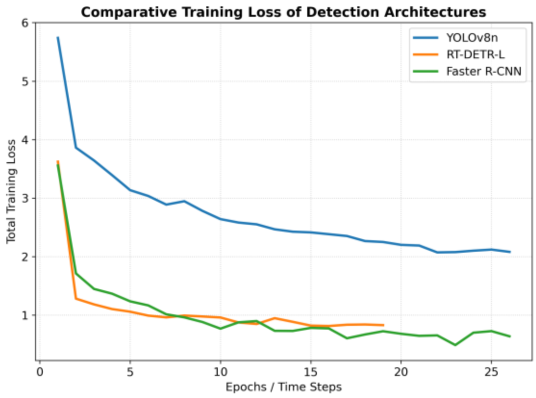


Figure 3.10: Detection training curves using single channel input for all three architecture for Fold 1.

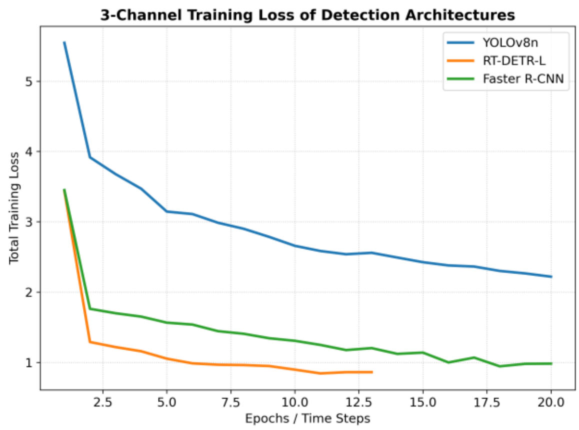


Figure 3.11: Detection training curves using three channel input for all three architecture for Fold 1.

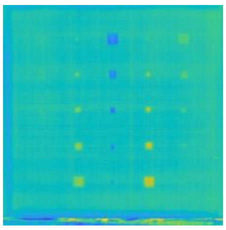
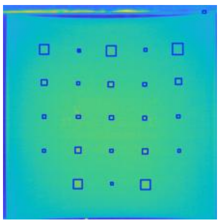
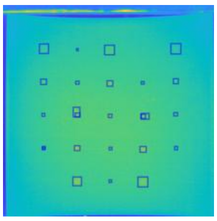
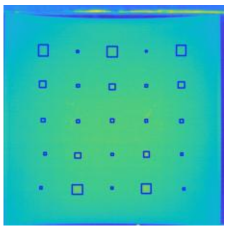
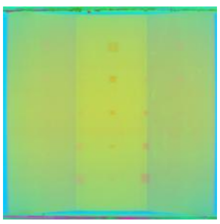
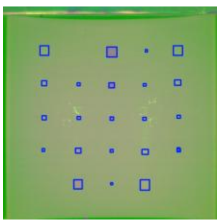
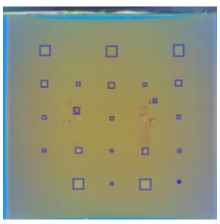
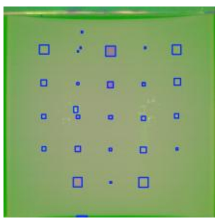
3.4.3.1 Fold 1 results:

Fold 1 used SQ_01 and CI_01 for training and SQ_02 and CI_02 for validation. SQ_03 is a trapezoidal plate with 25 square inserts (depth 0.2 to 1.0 mm) while CI_03 is a circular plate with 6 circular inserts (diameter 6mm, 10mm, 15mm, depth 2.0 to 2.2 mm). Table 3.5 provides the performance.

Table 3.5: Detection results Fold 1: test specimens SQ_03 and CI_03

		Raw				3 Channel			
Model	**Specimen**	**Precision**	**Recall**	**F1 Score**	**mAP@0.5**	**Precision**	**Recall**	**F1 Score**	**mAP@0.5**
YOLOv8n	**SQ_03**	0.95	0.90	0.93	0.94	0.96	0.94	0.95	0.96
	CI_03	0.88	0.75	0.81	0.82	0.95	0.87	0.91	0.91
	Mean	**0.92**	**0.86**	**0.89**	**0.89**	**0.96**	**0.91**	**0.93**	**0.94**
RT-DETR-L	**SQ_03**	0.96	0.92	0.94	0.95	0.98	0.95	0.96	0.97
	CI_03	0.91	0.80	0.85	0.85	0.96	0.86	0.91	0.92
	Mean	**0.94**	**0.88**	**0.91**	**0.91**	**0.97**	**0.91**	**0.94**	**0.95**
Faster R-CNN	**SQ_03**	0.94	0.84	0.89	0.88	0.95	0.88	0.91	0.91
	CI_03	0.86	0.68	0.76	0.75	0.90	0.81	0.85	0.86
	Mean	**0.91**	**0.77**	**0.83**	**0.84**	**0.93**	**0.85**	**0.88**	**0.89**

Average mAP improvement over raw input for Fold 1 was 4.4% for RT-DETR-L, 5.6% for YOLOv8n and 5.9% for Faster R-CNN. All three networks detected all 25 square inserts, and the three channel variant offered tighter bounding boxes due to reduced halos when at shallow depths. All of the single channel detectors missed at least one or two inserts on the circular plate, ranging from 0.68-0.80 recall, but the three channel detector obtained recall of 0.81-0.87 on this specimen. The largest improvement occurred in Faster R-CNN on the circular plate, from 0.75 mAP to 0.86 mAP (a jump of 14.7%): the PCA channel removed the diagonal lamp illumination artifact which tended to mask small circular targets, while the TSR channel recovered small or deep circular targets based on thermal decay which is lost in raw frame summation.

Sample	**Input**	**YOLOv8**	**RT-DETR**	**Faster R-CNN**
SQ_03 Raw				
SQ_03 Three Channel				

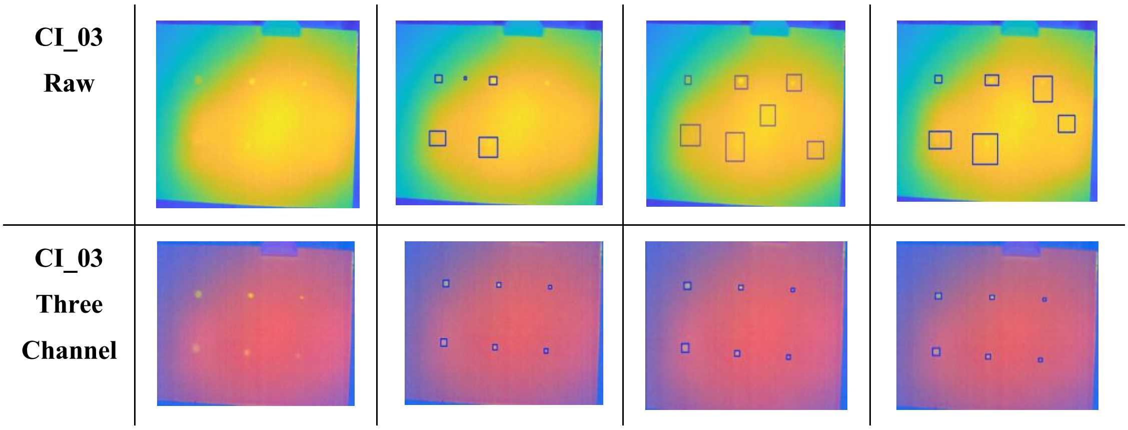


Figure 3.12: Visualization of detections from each of the three architectures, Fold 1 test specimens SQ_03 and CI_03

3.4.3.2 Fold 2 results

Fold 2 used SQ_03 and CI_03 for training and SQ_01 and CI_01 for validation. SQ_02 is a curved plate with 25 square inserts and CI_02 is a circular plate with 10 circular inserts (diameter 2mm-20mm) at a depth of 2.0 to 2.2 mm. CI_02 had the largest diameter range for circular inserts. Table 3.6 provides the performance.

Table 3.6: Detection results Fold 2: test specimens SQ_02 and CI_02

		Raw				3 Channel			
Model	**Specimen**	**Precision**	**Recall**	**F1 Score**	**mAP@0.5**	**Precision**	**Recall**	**F1 Score**	**mAP@0.5**
YOLOv8n	**SQ_02**	0.93	0.60	0.73	0.60	0.94	0.68	0.79	0.68
	CI_02	0.83	0.50	0.62	0.50	0.85	0.60	0.70	0.60
	Mean	**0.88**	**0.55**	**0.68**	**0.55**	**0.90**	**0.64**	**0.75**	**0.65**
RT-DETR-L	**SQ_02**	0.95	0.80	0.86	0.80	0.57	0.80	0.66	0.80
	CI_02	0.66	0.60	0.63	0.70	0.83	1.00	0.90	1.00
	Mean	**0.81**	**0.70**	**0.75**	**0.75**	**0.70**	**0.90**	**0.78**	**0.90**
Faster R-CNN	**SQ_02**	0.94	0.64	0.76	0.64	0.94	0.72	0.81	0.72
	CI_02	0.00	0.00	0.00	0.00	0.81	0.90	0.69	0.90
	Mean	**0.47**	**0.32**	**0.38**	**0.32**	**0.88**	**0.81**	**0.75**	**0.81**

It is apparent from Table 3.6 that Fold 2 yields the greatest improvement in the detection study due to the three channel input. Both RT-DETR-L and Faster R-CNN failed to produce any detections of circular targets on CI_02 with single-channel input, but recovered to obtain an mAP of 1.00 and 0.90 respectively with three channel input. YOLOv8n's performance on CI_02 remained the same when using three channels, but overall mean mAP increased by 0.10 with the use of three channels. For RT-DETR-L, mean mAP increased by 0.15, and for Faster R-CNN, the increase was 0.49, again highlighting the impact of the three channel input on the two-stage model due to its sensitive proposal region stage which could not recover from insufficient contrast in the single channel input.

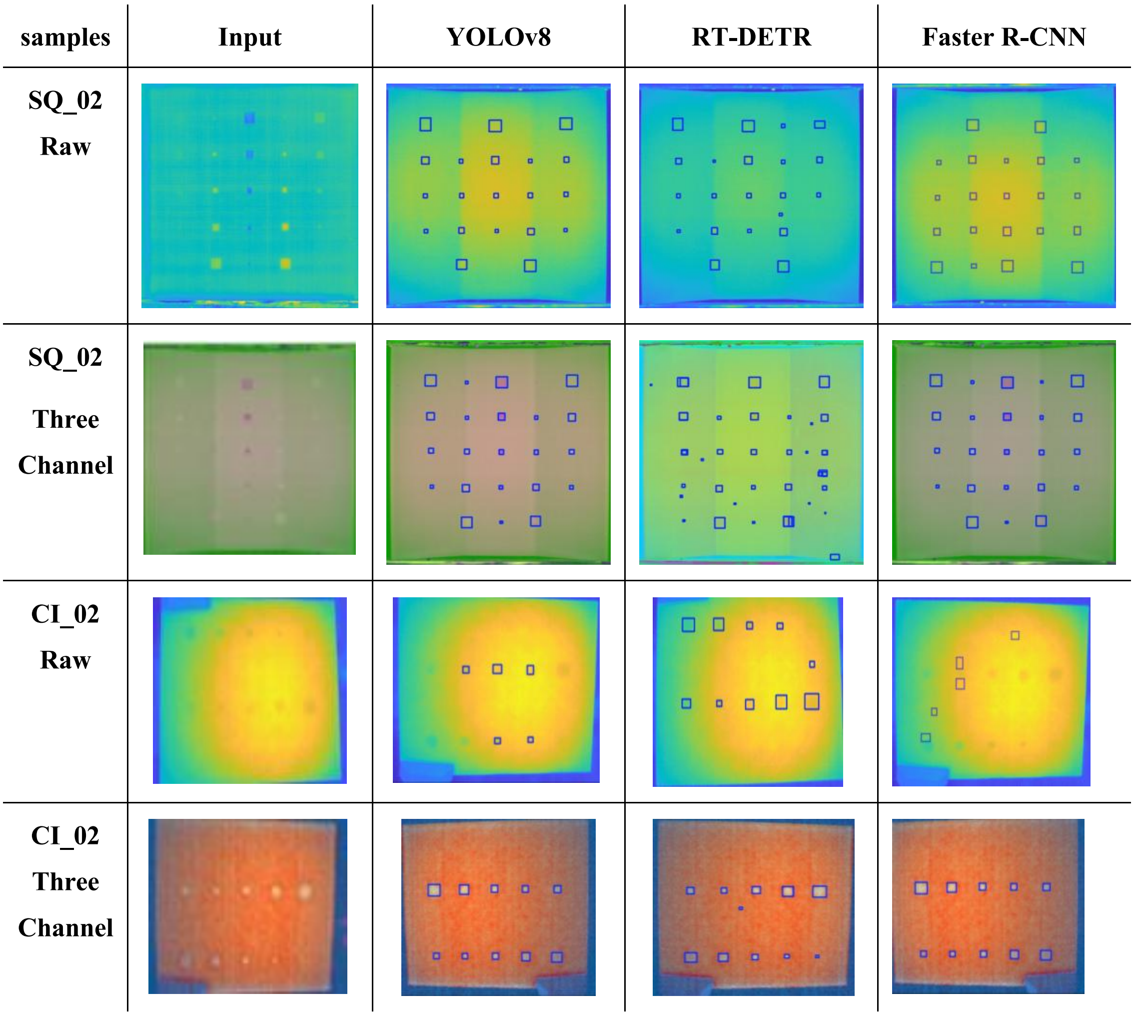


Figure 3.13: Visualization of detections from each of the three architectures, Fold 2 test specimens SQ_02 and CI_02.

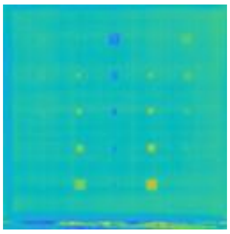
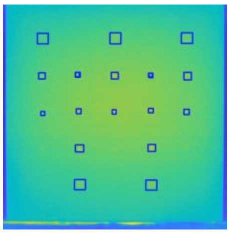
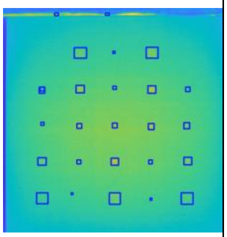
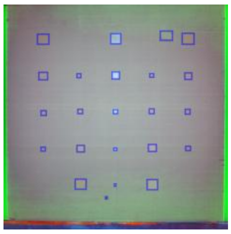
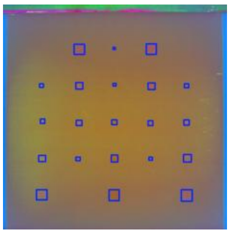
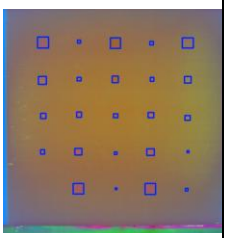
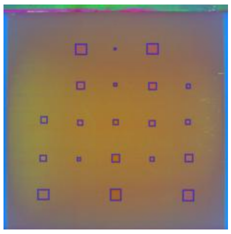
3.4.3.3 Fold 3 Results

Fold 3 is trained on SQ_02 and CI_02 and validated on SQ_03 and CI_03. Test specimens are SQ_01 (flat plate, 25 square inserts 0.2 to 1.0 mm deep) and CI_01 (16 circular inserts 2 to 8mm in diameter, 2.0 to 2.2mm deep). The smallest (2mm) inserts on CI_01 are the most difficult test objects to identify in this study, and are at the resolution limits for this depth and the 640 x 480 camera resolution. Quantitative results can be found in Table 3.7 below.

Table 3.7: Detection results, Fold 3: test specimens SQ_01 and CI_01.

		Raw				**3 Channel**			
Model	**Specimen**	**Precision**	**Recall**	**F1 Score**	**mAP@0.5**	**Precision**	**Recall**	**F1 Score**	**mAP@0.5**
YOLOv8n	**SQ_01**	0.94	0.68	0.77	0.79	0.95	0.76	0.84	0.76
	CI_01	0.25	0.06	0.10	0.06	0.83	0.31	0.45	0.31
	Mean	**0.60**	**0.37**	**0.44**	**0.43**	**0.89**	**0.54**	**0.65**	**0.54**
RT-DETR-L	**SQ_01**	0.82	0.76	0.79	0.76	0.95	0.80	0.86	0.76
	CI_01	0.00	0.00	0.00	0.00	0.92	0.75	0.82	0.75
	Mean	**0.41**	**0.38**	**0.40**	**0.38**	**0.94**	**0.78**	**0.84**	**0.76**
Faster R-CNN	**SQ_01**	0.90	0.72	0.80	0.72	0.94	0.72	0.81	0.72
	CI_01	0.00	0.00	0.00	0.00	0.36	0.25	0.29	0.25
	Mean	**0.45**	**0.36**	**0.40**	**0.36**	**0.65**	**0.49**	**0.55**	**0.49**

Table 3.7 illustrates the most limiting raw inputs seen during testing of Fold 3. All three models failed to perform on the CI_01 object under raw inputs, giving only one detection with YOLOv8n. CI_01 also presents the greatest challenge to modeling due to its small diameter, which at depth limits camera resolution and the minimum distinguishable insert size. This makes the thermal contrast insufficient for the given models to report a detection. With 3-channel inputs, all three models succeeded on the CI_01 object, though again to varying degrees. RT-DETR-L is the top performer, finding around 75% of the inserts at 0.75 mAP, followed by YOLOv8n and Faster R-CNN at 31% and 25% mAP, respectively. Again, none of the three models was able to detect the smallest (2mm diameter) inserts on CI_01. This is likely due to camera resolution limitations as opposed to model shortcomings.

samples	**Input**	**YOLOv8**	**RT-DETR**	**Faster R-CNN**
SQ_01 Raw				
SQ_01 Three Channel				

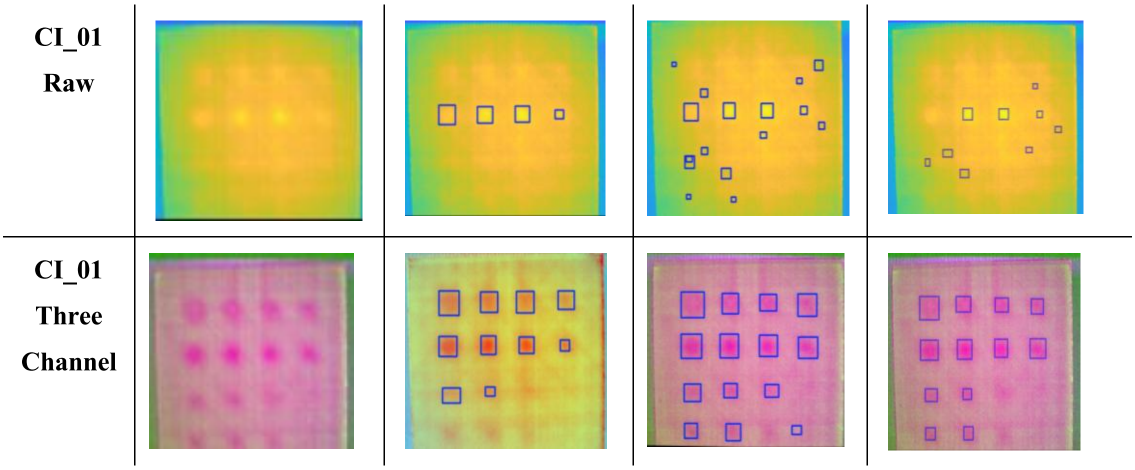


Figure 3.14: Visualization of detections from each of the three architectures, Fold 3 test specimens SQ_01 and CI_01.

3.5 Defect Segmentation: Methodology and Results

3.5.1 Segmentation Training Configuration

All three models were trained for binary pixel classification with sigmoid activation at the output and ImageNet pretrained encoders (ResNet-34 [34] for U-Net and FPN, ResNet-50 [34] for DeepLabV3+). Training employed AdamW [53] with the cosine annealing scheduler in Equation 3.8, early stopping at 10 validation epochs monitoring the Dice score, and gradient clipping at $l_2$norm 1.0. All models were implemented in PyTorch 0.3.3.

Training of segmentation models was done using a loss function that is combination of Binary Cross-Entropy and soft Dice Loss:

$$\mathcal{L}_{seg} = \frac{1}{2}\mathcal{L}_{BCE} + \frac{1}{2}\mathcal{L}_{Dice} \tag{3.14}$$

where $\mathcal{L}_{seg}$is the segmentation training loss, $\mathcal{L}_{BCE}$ is the binary cross-entropy loss, and $\mathcal{L}_{Dice}$ is the soft Dice loss. Label smoothing is applied to the cross-entropy loss term to avoid overly confident predictions on uncertain boundary pixels.

$$\mathcal{L}_{BCE} = -\frac{1}{N_p}\sum_{p=1}^{N_p}\left[\tilde{y}_p \ln \hat{y}_p + (1-\tilde{y}_p)\ln(1-\hat{y}_p)\right] \tag{3.15}$$

where N_p is the number of pixels in a mini-batch, $\hat{y}_p \in [0,1]$ is the predicted probability for pixel p, and $\tilde{y}_p = y_p(1-\varepsilon) + \varepsilon/2$ is the smoothed label with $\varepsilon = 0.1$, where $y_p \in \{0,1\}$ is the binary ground truth label for pixel. The soft Dice loss is calculated as:

$$\mathcal{L}_{Dice} = 1 - \frac{2\sum_{p=1}^{N_p} \hat{y}_p\, y_p + \delta}{\sum_{p=1}^{N_p} \hat{y}_p + \sum_{p=1}^{N_p} y_p + \delta} \tag{3.16}$$

where y_p is the binary ground truth label and $\delta = 10^{-6}$ is is a smoothing constant provides stability against division by zero on empty image masks.Table 3.8 shows the full hyperparameter configuration of all three segmentation models.

Table 3.8: Segmentation model training hyperparameters.

Parameter	U-Net	FPN	DeepLabV3+
Encoder	ResNet-34	ResNet-34	ResNet-50
Parameters (M)	21	28	41
Pretrained encoder	ImageNet	ImageNet	ImageNet
Optimiser	AdamW	AdamW	AdamW
$\boldsymbol{\eta_0}$	0.0001	0.0001	0.0001
Weight decay	0.0005	0.0005	0.0005
Batch size	8	8	8
Max epochs	40	40	40
LR schedule	Cosine	Cosine	Cosine
Early stopping (patience)	10 ep	10 ep	10 ep
Gradient clip norm	1.0	1.0	1.0
Label smoothing $\boldsymbol{\epsilon}$	0.1	0.1	0.1
ASPP dilation rates	-	-	1, 6, 12, 18

The training procedure for the segmentation models is summarized below:

Algorithm 3.3: Segmentation Model Training and Early Stopping

1. **Input:** Training set $\boldsymbol{D_{tr}}$; validation set $\boldsymbol{D_{va}}$; model $\boldsymbol{f_{seg}}$; maximum epochs $\boldsymbol{E_{max}}$; early stopping patience $\boldsymbol{P = 10}$; initial learning rate $\boldsymbol{\eta_0 = 0.0001}$; weight decay $\boldsymbol{\lambda = 0.0005}$; gradient clip norm = $\mathbf{1.0}$; label smoothing $\boldsymbol{\varepsilon = 0.1}$
2. **Initialise:** optimiser = AdamW($\boldsymbol{f_{seg}}$, $\boldsymbol{\eta_0}$, $\boldsymbol{\lambda}$); scheduler = CosineAnnealingLR($\boldsymbol{E_{max}}$, $\boldsymbol{\eta_{min} = 10^{-5}}$); best_Dice = 0; wait = 0
3. **for** $\boldsymbol{e = 1}$to $\boldsymbol{E_{max}}$do
4. **for** each batch $(\boldsymbol{X_b}, \boldsymbol{Y_b})$in $\boldsymbol{D_{tr}}$do
5. Compute predictions: $\boldsymbol{\hat{Y} = f_{seg}(X_b)}$
6. Apply label smoothing: $\boldsymbol{\tilde{Y}_b = Y_b(1 - \varepsilon) + \varepsilon/2}$
7. Compute $\boldsymbol{\mathcal{L}_{BCE}}$using Equation 3.15 with $\boldsymbol{\tilde{Y}_b}$
8. Compute $\boldsymbol{\mathcal{L}_{Dice}}$using Equation 3.16 with $\boldsymbol{Y_b}$
9. Compute $\boldsymbol{\mathcal{L}_{seg} = 0.5\, \mathcal{L}_{BCE} + 0.5\, \mathcal{L}_{Dice}}$

```
        Compute gradient of 𝓛_seg; clip gradient norm to 1.0; optimiser.step()
    end for
    Scheduler step
    Compute validation Dice on D_va
    if val_Dice > best_Dice: save checkpoint; best_Dice = val_Dice; wait = 0
    else: wait += 1
    if wait ≥ P: stop training
end for
Restore best checkpoint as f*_seg
Output: Optimised segmentation model f*_seg
```

3.5.2 Segmentation Evaluation Metrics

Evaluation was performed using the pixel-level Dice Coefficient and Intersection over Union (IoU):

$$\text{Dice} = \frac{2\,TP}{2\,TP + FP + FN},\ \text{IoU}_{px} = \frac{TP}{TP + FP + FN} \tag{3.17}$$

where TP, FP, and FN represent correctly classified defect pixels, erroneously classified non-defect pixels, and incorrectly classified non-defect pixels respectively. Both metrics were computed per-pixel over the withheld test set from each fold.

3.5.3 Segmentation Results and Discussion

To evaluate the proposed fusion representation in segmentation, a comparison is conducted between raw single channel inputs and three channel inputs, using each of the three segmentation architectures over the three cross-validation folds.

A study of overall circular specimen Dice scores under raw single channel inputs for all three architectures reveals values within a narrow range between 0.24 and 0.29. This lack of variance offers little information about architecture-level performance differences when using raw input. By introducing three channel inputs, these scores are shifted to the range between 0.27 and 0.39 with recall increases of as much as 37% for U-Net. The consistent nature of these improvements across models with distinctly different underlying architectures further suggests the input representation as the bottleneck at raw input levels.

U-Net achieves a mean validation Dice of 0.82 overall, or specifically 0.839 and 0.294 under the square and circular specimens respectively. On the deeper circular defects, U-Net at raw input levels generates diffuse and poorly localized masks. The core reason for this issue arises from U-Net's skip connections transmitting only spatial detail from meaningful encoder features; raw thermal inputs largely fail to capture distinct defect contrasts and instead focus more on heating gradients. At three channel inputs, these results jump to a mean validation Dice of 0.92, and for circular defects the recall increases from 0.59 to 0.96. FPN, in comparison, shows the most pronounced shift in data, from a high square validation Dice of 0.874 and low circular Dice of 0.258 to values of 0.85 and 0.39. It shows extreme over-segmentation at the raw level; the large predictions of the model cover large portions of general defect regions rather than precise outlines. Its validation Dice increases from 0.42 to 0.58 (51% relative) and square precision from 0.89 to 0.96 when switching input levels. DeepLabV3+ gains slightly less, likely because ASPP's parallel dilated convolutions already provided contextual background suppression. It increases from 0.46 validation Dice to 0.65. Circular Dice also increased by a more modest 14%, although on CI_02 for Fold 2, its Dice is increased by 633% (from 0.06 to 0.44) with three channel inputs.

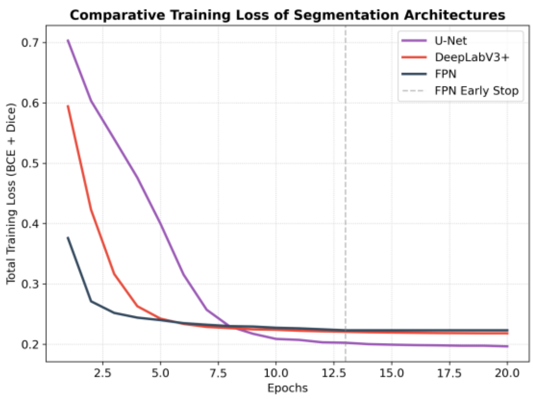


Figure 3.15: Segmentation training curves for all three architectures on raw single channel inputs, Fold 1.

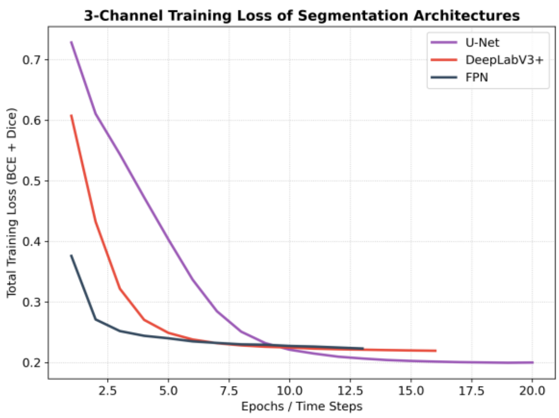


Figure 3.16: Segmentation training curves for all three architectures on three channel inputs, Fold 1.

3.5.3.1 Fold 1 results

This fold is trained on SQ_01 and CI_01 and validated on SQ_02 and CI_02. SQ_03 had 25 square inserts located between depths 0.2mm and 1.0mm and CI_03 had 6 circular inserts ranging from 6mm to 15mm in diameter and located between depths 2.0mm and 2.2mm. The results are shown in Table 3.9.

Table 3.9: Segmentation results, Fold 1: test specimens SQ_03 and CI_03.

		Raw				3 Channel			
Model	**Specimen**	**Precision**	**Recall**	**Dice**	**IoU**	**Precision**	**Recall**	**Dice**	**IoU**
U-Net	SQ_03	0.92	0.76	0.83	0.72	0.92	0.75	0.84	0.72
	CI_03	0.19	0.59	0.29	0.17	0.19	0.96	0.34	0.20
	Mean	**0.56**	**0.68**	**0.56**	**0.44**	**0.56**	**0.86**	**0.59**	**0.46**
FPN	SQ_03	0.89	0.85	0.87	0.77	0.89	0.75	0.85	0.73
	CI_03	0.14	0.96	0.25	0.14	0.14	0.97	0.39	0.24
	Mean	**0.52**	**0.90**	**0.56**	**0.46**	**0.52**	**0.86**	**0.62**	**0.49**
DeepLabV3+	SQ_03	0.92	0.75	0.83	0.71	0.92	0.76	0.83	0.72
	CI_03	0.14	0.76	0.23	0.13	0.14	0.75	0.27	0.16
	Mean	**0.53**	**0.75**	**0.53**	**0.42**	**0.53**	**0.76**	**0.55**	**0.44**

From the table, we see that both the raw and three channel images predict high-Dice results for square defects across all three networks; this is expected, since SQ_03 is high contrast in terms of thermal behavior. For circular defects, we also see that U-Net's recall increases to 0.96 by introducing three channel inputs and therefore finds the true number of defects for the first time. FPN, although finding the correct number of circular defects even when trained on raw inputs (recall = 0.96), over-segments heavily (precision = 0.14); its circular Dice improves to 0.39, but by narrowing the predicted boundaries, not by finding additional defects. The overall means show that FPN and U-Net both achieve a Dice of 0.59 and 0.62, but this increase is driven more by recall. DeepLabV3+ offers marginal improvements in this fold. The results show that the inclusion of an additional channel can significantly enhance the detection and segmentation of circular defects, although the level of improvement varies based on the architecture and the severity of the segmentation issue.

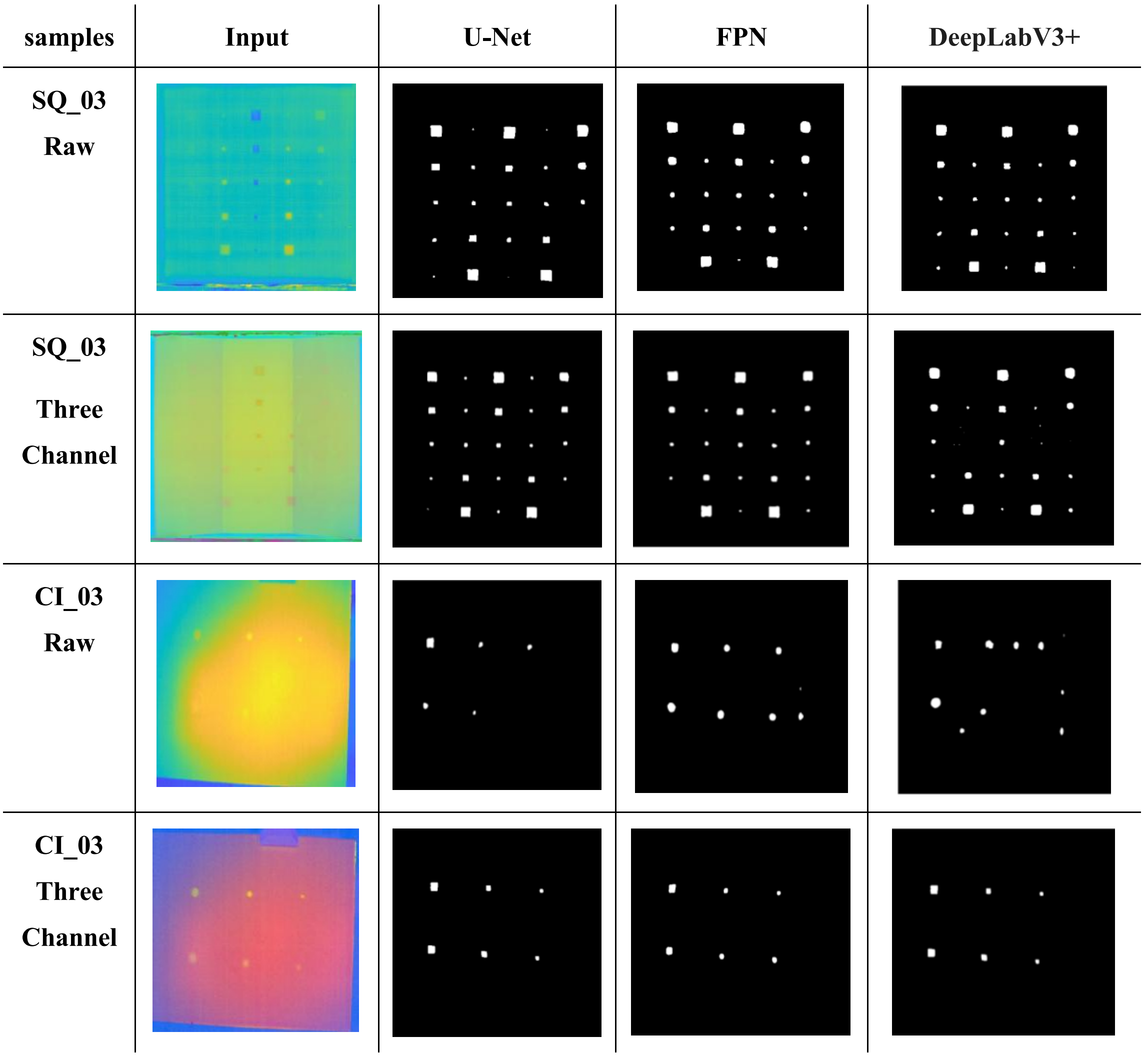


Figure 3.17: Segmentation mask visualisations for all three architectures, Fold 1 test specimens SQ_03 and CI_03.

3.5.3.2 Fold 2 results

This fold is trained on SQ_03 and CI_03 and validated on SQ_01 and CI_01. SQ_02 is a curved plate with 25 square defects and CI_02 has 10 circular defects at depths from 2.0mm to 2.2mm. The results are summarized in Table 3.10.

Table 3.10: Segmentation results, Fold 2: test specimens SQ_02 and CI_02.

		Raw				3 Channel			
Model	**Specimen**	**Precision**	**Recall**	**Dice**	**IoU**	**Precision**	**Recall**	**Dice**	**IoU**
U-Net	SQ_02	0.79	0.86	0.82	0.70	0.95	0.75	0.84	0.72
	CI_02	0.50	0.50	0.50	0.33	0.42	0.69	0.52	0.35
	Mean	**0.64**	**0.68**	**0.66**	**0.51**	**0.68**	**0.72**	**0.68**	**0.53**
FPN	SQ_02	0.87	0.83	0.85	0.74	0.86	0.85	0.85	0.75
	CI_02	0.00	0.00	0.00	0.00	0.67	0.20	0.31	0.18
	Mean	**0.43**	**0.41**	**0.42**	**0.37**	**0.76**	**0.52**	**0.58**	**0.46**

DeepLabV3+	SQ_02	0.83	0.89	0.86	0.76	0.84	0.88	0.86	0.76
	CI_02	0.19	0.04	0.06	0.03	0.55	0.37	0.44	0.28
	Mean	**0.51**	**0.46**	**0.46**	**0.39**	**0.69**	**0.62**	**0.65**	**0.52**

We find the greatest overall improvement from the addition of three channel inputs in the Fold 2 case. With the exception of U-Net's mean Dice score (which sees a marginal increase of 0.66 to 0.68), both FPN and DeepLabV3+ see performance boosts of over 38% in mean Dice. Particularly dramatic improvements are seen in the circular defect scenario for both models; FPN and DeepLabV3+ predict nothing at all under the raw input condition (Dice = 0.00), but find a Dice of 0.31 and 0.44, respectively, after three channel inputs. This shows the significant difficulty the network has when detecting circularly shaped defects at low thermal contrast at any depth when using the low contrast and thermal gradient information present in the raw images.

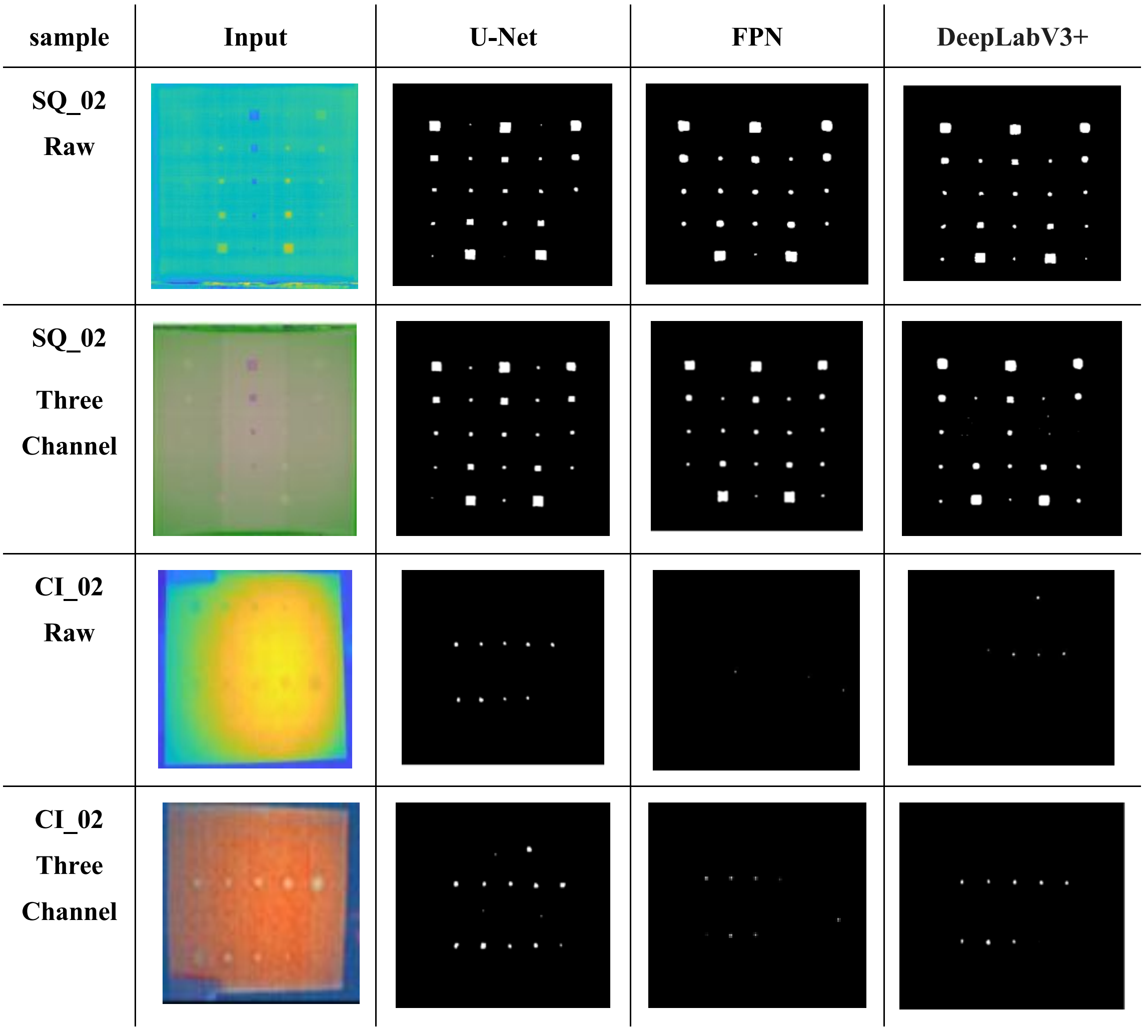


Figure 3.18: Segmentation mask visualisations for all three architectures, Fold 2 test specimens SQ_02 and CI_02.

3.5.3.3 Fold 3 results

Fold 3 is trained on SQ_02 and CI_02, and validated on SQ_03 and CI_03. The test specimens are SQ_01 (flat plate, 25 square inserts at 0.2-1.0mm depth), and CI_01 (16 circular inserts 2.0-8.0mm diameter 2.0-2.2mm depth). Quantitative results can be seen in Table 3.11 below:

Table 3.11: Segmentation results, Fold 3: test specimens SQ_01 and CI_01.

		Raw				3 Channel			
Model	**Specimen**	**Precision**	**Recall**	**Dice**	**IoU**	**Precision**	**Recall**	**Dice**	**IoU**
U-Net	SQ_01	0.85	0.84	085	0.74	0.86	0.86	0.86	0.75
	CI_01	0.49	0.08	0.15	0.08	0.47	0.19	0.26	0.15
	Mean	**0.67**	**0.46**	**0.050**	**0.41**	**0.66**	**0.52**	**0.56**	**0.45**
FPN	SQ_01	0.82	0.86	0.84	0.72	0.85	0.84	0.85	0.74
	CI_01	0.36	0.07	0.11	0.06	0.53	0.11	0.19	0.10
	Mean	**0.59**	**0.46**	**0.47**	**0.39**	**0.69**	**0.47**	**0.52**	**0.42**
DeepLabV3+	SQ_01	0.82	0.87	0.85	0.73	0.84	0.84	0.84	0.73
	CI_01	0.21	0.08	0.11	0.06	0.43	0.13	0.20	0.11
	Mean	**0.51**	**0.47**	**0.48**	**0.39**	**0.63**	**0.48**	**0.52**	**0.42**

It can be observed from Table 3.11 that under raw inputs, the Dice score on CI_01 for all three models is relatively low, from 0.11-0.15. The recall values were only between 0.07-0.08, indicating that out of 16 circle inserts, fewer than 2 on average received any kind of valid mask prediction. Using three channel inputs gives a partial recovery, with the Dice score on CI_01 for U-Net now at 0.26 (73% relative improvement), 0.19 for FPN, and 0.20 for DeepLabV3+. This is an improvement to about 3-4 inserts receiving meaningful segmentation masks. The Dice score for the square specimen remained consistently high for all three models, between 0.84-0.86 for both single-channel and three-channel inputs. The very low Dice score that the circular specimen continued to yield even when using three channels input is not the fault of the models used. Due to their small 2mm dimensions, the camera could not record enough spatial detail (at 640 480 pixel resolution) to distinguish them from one another when placed at a depth from 2.0-2.2mm, even with improved signal input processing. This creates a lower bound to spatial resolution given the acquisition hardware, which could not be improved with other signal processing steps.

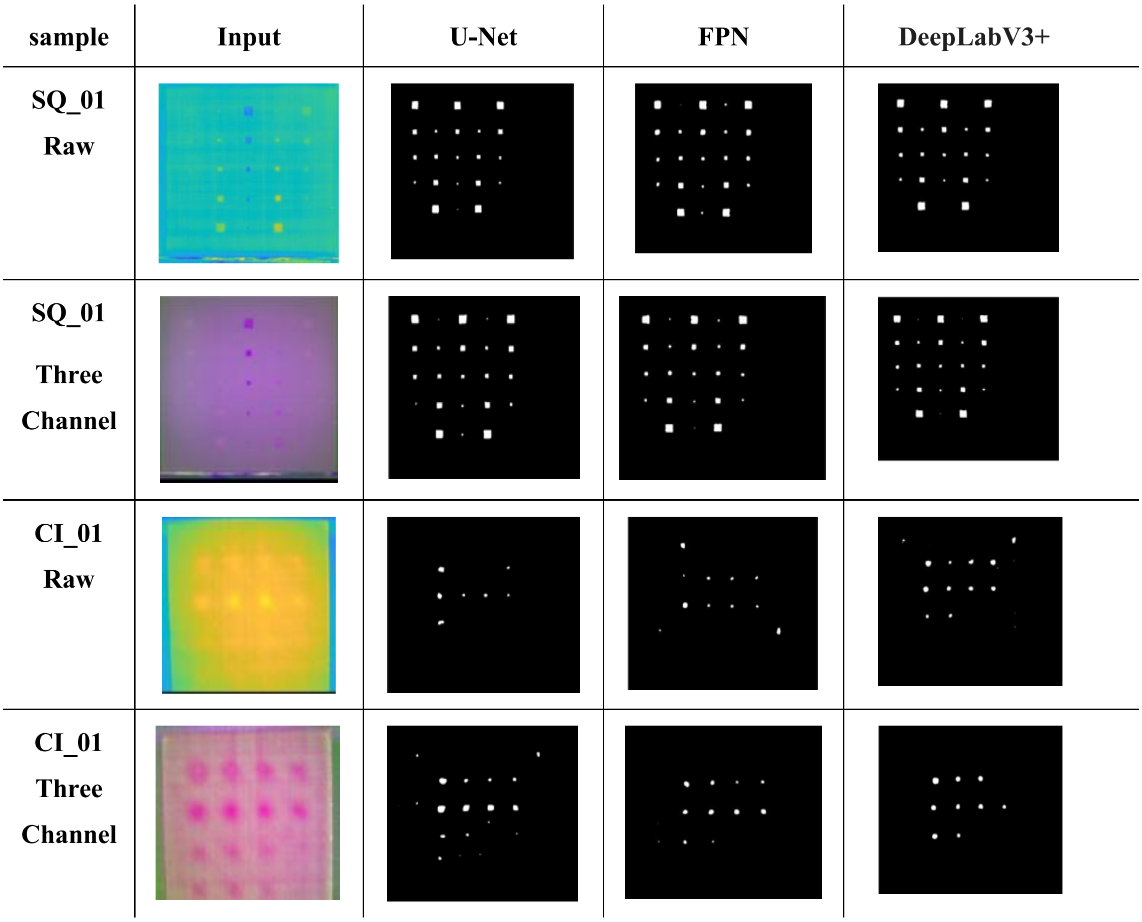


Figure 3.19: Segmentation mask visualizations for all three architectures, Fold 3 test specimens SQ_01 and CI_01.

3.6 Chapter Summary

This chapter proposes a three channel thermographic fusion method for defect detection and segmentation of CFRP, which fuses the raw thermal frames, PCA features and TSR features to $H \times W \times 3$format. This fusion approach has been tested on six specimen using YOLOv8n, RT-DETR-L, Faster R-CNN, U-Net, FPN and DeepLabV3+ by three fold cross-validation. It shows a general improvement on detection mAP of +5.6%, +4.4% and +5.9% over the single channel input. Furthermore, a noticeable improvement on the detection and segmentation of otherwise undetectable defects can be achieved, and a Dice score enhancement of up to 51% is achieved. It is suggested that these improvements are architecture independent, and are induced by the rich input features. However, the smallest defects (2 mm) can still not be detected due to the limitation of the hardware.

CHAPTER 4

PHYSICS INFORMED DEPTH ESTIMATION AND THREE-DIMENSIONAL RECONSTRUCTION

4.1 Introduction

In Chapter 3 the problem of detecting and segmenting subsurface defects within CFRP composites using a 3-channel thermographic fusion technique is addressed. While using the segmentation masks to find the lateral location and boundaries of each defect is possible, depth information about each defect can not be extracted. This depth is crucial to the structural assessment, as a defect located 0.2mm deep in the composite will be located on a different ply interface than a defect 2.0mm deep; the former having greater structural implications than the latter.

This chapter will discuss a physics-informed approach for the depth characterisation of defects in CFRP composites via optical pulse thermography. The underlying physics are related to the established relationship between the peak time of the surface thermal response above a subsurface defect and its corresponding depth below the inspected surface. Sixteen features were extracted from the spatially averaged thermal response curve of each segmented defect using this relationship, and five regression models were subsequently trained and tested on these features. The proposed workflow is shown in Figure 4.1 below.

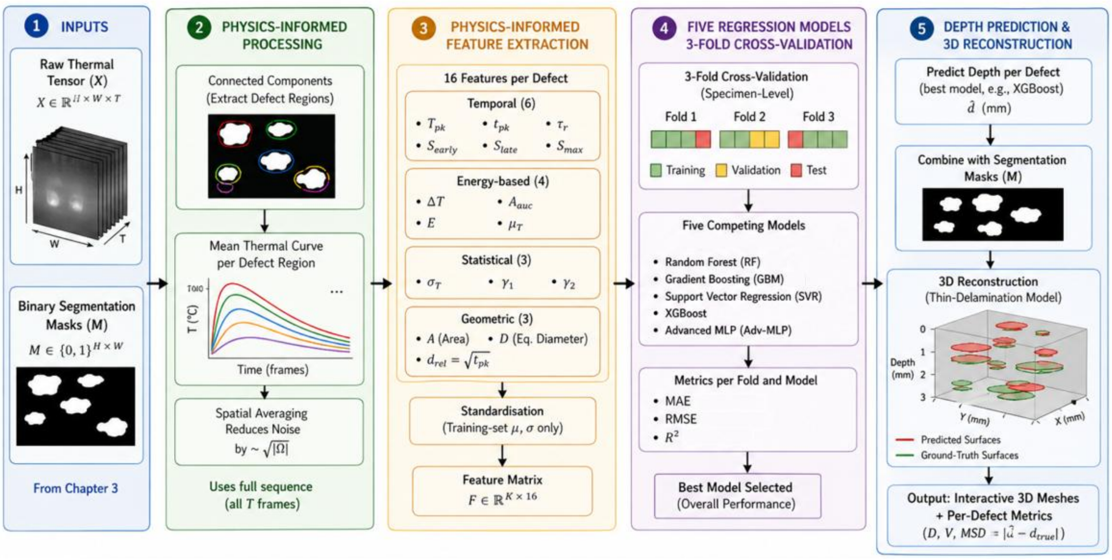


Figure 4.1: Proposed defect depth estimation and 3D surface plot pipeline from raw thermal tensor and Chapter 3 segmentation masks.

4.2 Physical Basis and Feature Extraction

4.2.1 Physical Basis

The heat signature acquired on the surface above a subsurface defect after an optical pulse is based on the one-dimensional heat diffusion equation [24], and when a thermal wave has travelled down to the defect interface and back up to the surface, the temperature above the defect will rise and peak. The governing relationship is given by

$$d^2 \approx \beta \cdot \alpha \cdot t_{peak} \tag{4.1}$$

where d is the depth below the surface (m), α is the thermal diffusivity along the direction perpendicular to the plane of the ply layers (m^2/s), t_{peak} is the time elapsed between the flash excitation and the peak of the surface temperature, and β is a dimensionless geometrical correction factor for lateral heat flow. In general β is taken to be between 1.5 and 2.0 and depends on how large the defect is relative to its depth. Rearranging equation 4.1 gives the principle which forms the basis for this chapter:

$$d \propto \sqrt{t_{peak}} \tag{4.2}$$

Clearly, directly using equation 4.1 requires knowledge of α and β which depend on the specific material and, usually, the sample being inspected. These are unavailable, so instead, regression models are used to learn the relationship between the derived features and depth, with the relationship implicitly encoding variations in α and β [50]. Instead of using only the portion of the sequence identified as a defects temporal profile in chapter 3, all images in the thermal sequence are used to extract features here as deeper defects require longer times for the heat pulse to travel to and back from, to be fully detected. For each defect region Ω_k, k=1, 2, ..., K, the spatially averaged thermal signature was taken, defined as

$$T_t = \frac{1}{\mid \Omega_k \mid} \sum_{(i,j)\in\Omega_k} X_{i,j,t}\, , t = 1,2,\ldots,T \tag{4.3}$$

where $\mid \Omega_k \mid$ is the number of pixels in defect region Ω_k, T_t is the mean temperature at frame t and T is the number of frames. Using the spatially averaged curve is beneficial due to the reduction in per pixel noise, a key factor for dataset 2 in particular as VOx cameras without cooling (unlike the detector in Dataset 1) can have quite significant per pixel noise.

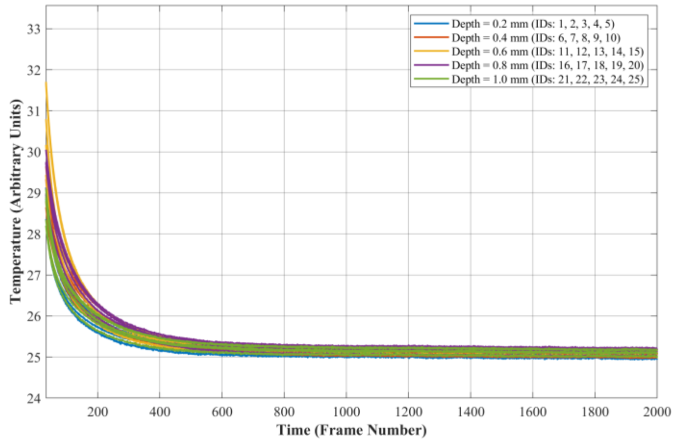


Figure 4.2: Mean thermal curves T_t for defects in SQ_01, at 0.2mm, 0.4mm, 0.6mm, 0.8mm and 1.0mm depth respectively.

The trends shown by the curves confirm equation 4.2; later peak times for increased depth and reduced peak temperatures.

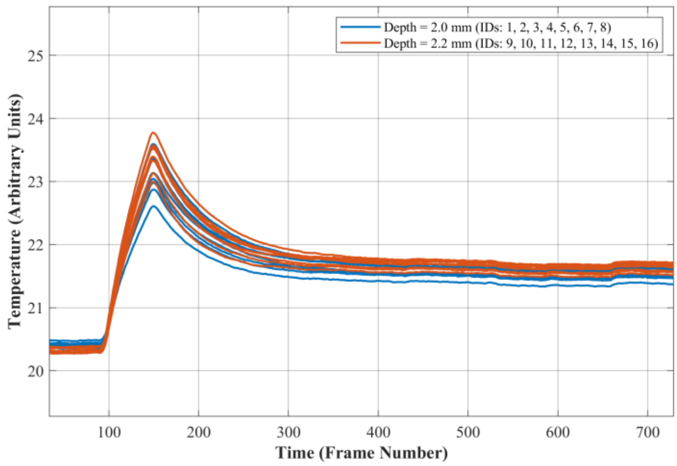


Figure 4.3: Mean thermal curves T_t for defects in CI_01 at 2.0mm and 2.2mm depth respectively.

Here we can see that there is limited discrimination between the two differing depths, due to the shallow difference between them and the cylindrical geometry of the sample.

4.2.2 Feature Engineering

Sixteen features are derived from the extracted mean thermal signature T_t and the geometry of the defect region. Prior to definition of each of these, three quantities derived from the thermal signature that feature in several features, are defined here: the inter-frame temperature increment

$$\Delta T_t = T_{t+1} - T_t, t = 1, \ldots, T - 1 \tag{4.4}$$

where T_{t+1} is the mean temperature at frame $t + 1$ (defined in 4.3) and T_t is the mean temperature at frame t; the average thermal signature over the whole sequence:

$$\mu_T = \frac{1}{T}\sum_{t=1}^{T} T_t \tag{4.5}$$

and the standard deviation of the thermal signature:

$$\sigma_T = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(T_t - \mu_T)^2} \tag{4.6}$$

where μ_T is given in 4.5. Throughout the tables for the features: Δt refers to the time difference between frames, which is given as $1/(\text{frame rate [fps]})$. The pixel area is denoted by A. The floor operation is shown as $\lfloor \cdot \rfloor$, while the frame where the maximum value of the thermal signature occurs is given by $\arg\max_t$, and $\text{mean}(\cdot)_{t\in\mathcal{S}}$ is the arithmetic mean over a set of frames $\mathcal{S}$.

The sixteen features are divided into four groups: temporal, energy-based, statistical and geometric. The temporal features capture the time-domain nature of the thermal response. Of all features, the time-to-peak, t_{pk}, is the most physically justified as its directly linked to defect depth, as given by Eq 4.2. Across all folds and models tested t_{pk}came out as feature 1 in importance.

Table 4.1: Temporal features extracted from the mean thermal curve

Symbol	Formula	Variables	Physical rationale
$\boldsymbol{T_{pk}}$	$\max_t T_t$	T_t: mean curve at frame t[Eq. 4.3]	Peak temperature; shallower defects have a higher peak temperature due to reflecting off the bottom and thus giving the surface heating time to return to room temperature.
$\boldsymbol{t_{pk}}$	$\arg\max_t T_t$	t: frame index, $t \in 1, \dots, T$	Time to peak [frames]; has a direct physical relationship to depth (Eq. 4.2)
$\boldsymbol{\tau_r}$	t_{pk}/T	T: total number of frames	Normalised rise time; this is necessary to enable the time to peak $t_{pk}$to be compared between Dataset 1 (which had up to 2000 frames at 120 Hz) and Dataset 2 (which had up to 729 frames at 50 Hz).

$\boldsymbol{S_{early}}$	$\mathrm{mean}(\Delta T_t)_{t\in[1,\lfloor 0.2T\rfloor]}$	ΔT_t: frame-to-frame increment [Eq. 4.4]	The average slope over the first 20% of frames; this will be greater for shallower defects as the reflected thermal wave reaches the surface earlier and produces heating over a shorter period.
$\boldsymbol{S_{late}}$	$\mathrm{mean}(\Delta T_t)_{t\in[\lfloor 0.8T\rfloor,T}$	ΔT_t: frame-to-frame increment [Eq. 4.4]	The average slope over the last 20% of frames; characteristic of the post-peak cooling process.
$\boldsymbol{S_{max}}$	$\max_t \Delta T_t$	ΔT_t: frame-to-frame increment [Eq. 4.4]	The maximum thermal gradient; associated with the point at which the reflected thermal wave first reaches the surface.

The energy-based features describe the integral shape of the curve over time. They complement the slope and peak features as they indicate the total energy that the defect region holds over the duration of the measurement.

Table 4.2: Energy-based features extracted from the mean thermal curve

Symbol	Formula	Variables	Physical rationale
$\boldsymbol{\Delta T_{contrast}}$	$T_{pk} - T_1$	T_1: temperature at frame 1	Thermal contrast; subtracting the initial, stable, room temperature eliminates session-to-session variations in absolute thermal response.
$\boldsymbol{A_{auc}}$	$\sum_{t=1}^{T} T_t \cdot \Delta t$	Δt: inter-frame interval [s]	Area under the curve [degree-seconds]; a measure of peak height multiplied by its duration, giving overall energetic contribution of the defect signature.
$\boldsymbol{E}$	$\sum_{t=1}^{T} T_t^2$	T_t: mean curve [Eq. 4.3]	Signal energy [squared detector units]; gives more weight to temperature differences further away from zero.
$\boldsymbol{\mu_T}$	Equation 4.5	-	Mean temperature over the full sequence; highly correlated with $T_{pk}$and $\Delta T_{contrast}$and thus useful redundant information for ensemble trees [54].

The statistical features are independent of the absolute temperature and describe the shape of the thermal signature itself. Skewness and Kurtosis are especially useful features, describing the asymmetry between the steep early rise for shallower defects and the spread out shape of the thermal curve for deep defects.

Table 4.3: Statistical features extracted from the mean thermal curve

Symbol	Formula	Variables	Physical rationale
$\boldsymbol{\sigma_T}$	Equation 4.6	μ_T: sequence mean [Eq. 4.5]	Standard deviation of the thermal signature; typically larger for shallower defects where the signal strength fluctuates significantly.
$\boldsymbol{\gamma_1}$	$\frac{1}{T}\sum_{t=1}^{T}\left(\frac{T_t-\mu_T}{\sigma_T}\right)^3$	μ_T[Eq. 4.5]; σ_T[Eq. 4.6]	Skewness; describes the asymmetric nature of the post-flash heating and cooling.
$\boldsymbol{\gamma_2}$	$\frac{1}{T}\sum_{t=1}^{T}\left(\frac{T_t-\mu_T}{\sigma_T}\right)^4$	μ_T[Eq. 4.5]; σ_T[Eq. 4.6]	Kurtosis; gives an indication of whether there are unusually high peak temperatures, as expected for sharp, fast-heating defects.

The geometric features relate the shape of the defect region to the geometric factor β in Equation 4.1. For wide defects where the one-dimensional model breaks down, they give an indication of how t_{pk} may be erroneously indicating the depth.

Table 4.4: Geometric features extracted from the defect region

Symbol	Formula	Variables	Physical rationale
$\boldsymbol{A}$	$\mid \Omega \mid$	Ω_k	Area in pixels
$\boldsymbol{D}$	$2\sqrt{A/\pi}$	A	Equivalent circular diameter
$\boldsymbol{d_{rel}}$	$\sqrt{t_{pk}}$	t_{pk}: time to peak [frames]	The physics-based depth proxy; this gives the regression model both the linearised depth prediction and the raw peak time value.

A strong correlation exists between the features related by the same physics (t_{pk}, τ_r, d_{rel}). This is due to these features describing the time it takes for the heat to reflect back at the surface (from Eq 4.2), thus having similar functional dependence on depth. The features A_{auc} and E also share a high correlation due to the same reason of capturing a related feature of the thermal curve. The geometric features A and D do not share significant correlations with the temporal features, this is why these features become very important when applied to the data of wide shallow defects, where the assumption of a one-dimensional thermal diffusion does not hold as well.

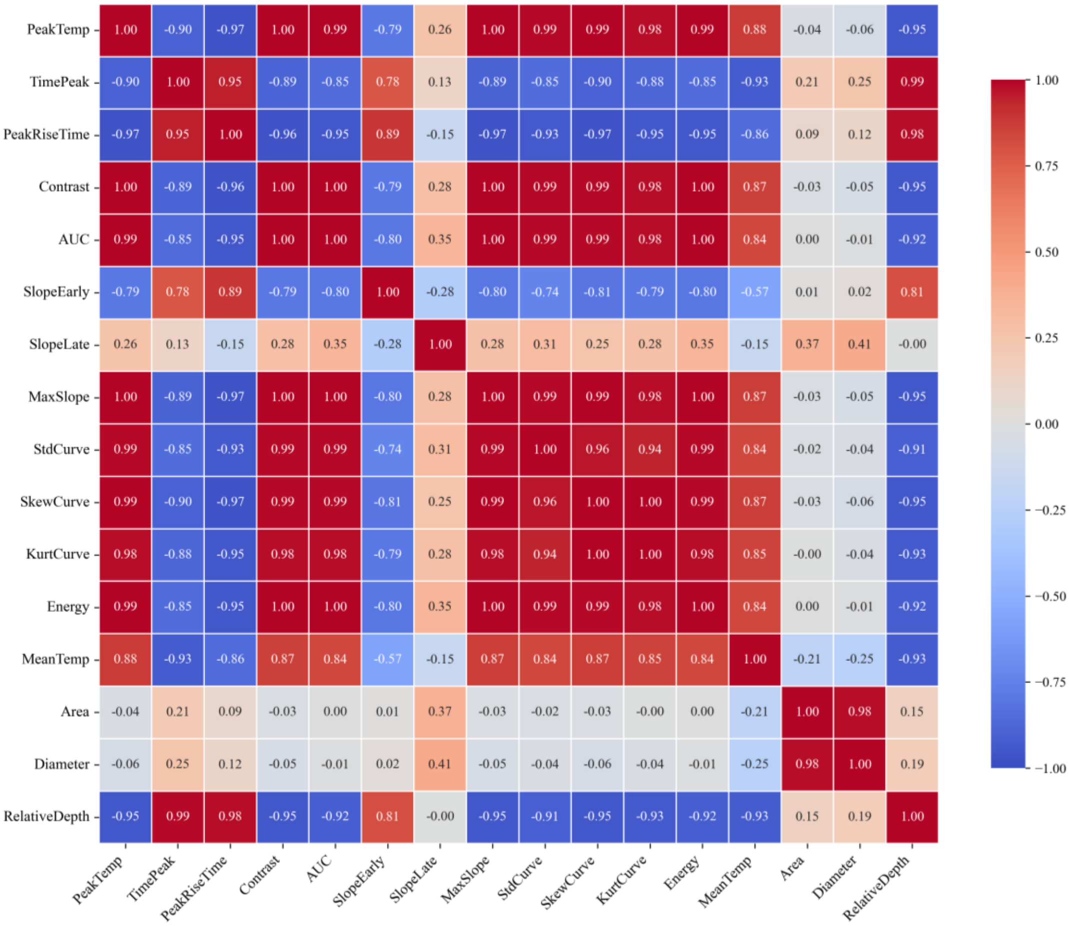


Figure 4.4: Pearson correlation matrix of all 16 features on both datasets.

4.2.3 Assigning depth labels and defect distribution

For training and evaluation only, depth labels are derived from the physical sample geometry as provided in the reference datasets.

In the case of square-insert specimens, defects are arranged column-wise, and each detected defect x_k(defined by the x-coordinate of its centroid) is assigned a depth according to its column index:

$$d_{\mathrm{SQ}} = \mathrm{depthmap}(\mathrm{col\_index}(x_k)) \tag{4.7}$$

Here, col_index(·) maps the centroid position to its corresponding column. The depth map $\{1.0, 0.6, 0.2, 0.4, 0.8\}$ mm defines the nominal depths for the five columns, as reported in [52].

In the case of circular-insert specimens, defects are arranged row-wise, and the depth is assigned based on the row index of the defect centroid:

$$d_{\mathrm{CI}} = \mathrm{depthmap}(\mathrm{row\_index}(y_k)) \tag{4.8}$$

Here, row_index(·) maps the centroid position to its corresponding row. The depth map $\{2.0, 2.2\}$ mm defines the nominal depths for the first and second rows, as reported in [33].

Since regression is done at the defect level, the sizes of the training set used here are considerably smaller than Chapter 3. Table 4.5 lists the number of defects in each of the three cross-validation folds.

Table 4.5: Distribution of defects across the three cross-validation folds.

Fold	Training Specimens	Validation Specimens	Test Specimens
1	SQ_01, CI_01 (41: 25 SQ + 16 CI)	SQ_02, CI_02 (35: 25 SQ + 10 CI)	SQ_03, CI_03 (31: 25 SQ + 6 CI)
2	SQ_03, CI_03 (31: 25 SQ + 6 CI)	SQ_01, CI_01 (41: 25 SQ + 16 CI)	SQ_02, CI_02 (35: 25 SQ + 10 CI)
3	SQ_02, CI_02 (35: 25 SQ + 10 CI)	SQ_03, CI_03 (31: 25 SQ + 6 CI)	SQ_01, CI_01 (41: 25 SQ + 16 CI)

4.3 Depth Regression: Methodology and Results

4.3.1 Model Configurations and Evaluation

Five regression models are considered. Prior to training, all 16 features are standardized based on training-set statistics alone, ensuring that no data from validation or test specimens leaks into the normalization step:

$$\hat{f}_j = \frac{f_j - \mu_j^{tr}}{\sigma_j^{tr}}, j = 1,2,\ldots,16 \tag{4.9}$$

where f_j is the raw value of feature j, $\hat{f}_j$ is the standardized value, μ_j^{tr} is the mean value of feature j on the training set, and σ_j^{tr} is the standard deviation of feature j on the training set.

4.3.1.1 Random Forest (RF)

Random Forest (RF) [54] employs an ensemble of 300 decision trees, each trained on a bootstrap sample of defects with random subsets of features considered at each node split. It is configured with: nestimators = 300, maxdepth = 10, minsamplessplit = 2, with a fixed random seed. The prediction of each tree is averaged to produce the final result.

4.3.1.2 Gradient Boosting (GBM)

Gradient Boosting (GBM) trains 300 shallow trees sequentially, with each tree attempting to fit the residuals of the ensemble generated by previous trees. It is configured with: nestimators = 300, learningrate = 0.05, max_depth = 5. The relatively small learning rate helps prevent overfitting in the face of the reduced training set sizes within each fold.

4.3.1.3 Support Vector Regression (SVR)

Support Vector Regression (SVR) employs an RBF kernel and a RobustScaler preprocessing pipeline. The regularisation parameter is $C = 10$and the epsilon-insensitive loss margin is $\varepsilon = 0.1$. The prediction function is:

$$\hat{d}(\mathbf{x}) = \sum_{i=1}^{N_{sv}} (\alpha_i - \alpha_i^*)\, K(\mathbf{x}_i, \mathbf{x}) + b \tag{4.10}$$

where $\mathbf{x} \in \mathbb{R}^{16}$ is the standardized feature vector for the test defect, N_{sv} is the number of support vectors kept by the model, $\alpha_i \geq 0$ and $\alpha_i^* \geq 0$ are the dual variables which are outputs from the solution to the SVR quadratic program, $K(\mathbf{x}_i, \mathbf{x}) = \exp(-\gamma \parallel \mathbf{x}_i - \mathbf{x} \parallel^2)$ is the RBF kernel with learned width γ, $\mathbf{x}_i$ is the feature vector of the i-th support vector, and $b \in \mathbb{R}$ is the bias term.

4.3.1.4 XGBoost

XGBoost [55] is a gradient boosting model with second-order gradient approximations and a focus on sparsity-aware split finding. It is configured with: n_estimators = 300, learningrate = 0.05, maxdepth = 6, L1 regularization regalpha = 0.1, L2 regularization reg_lambda = 1.0, subsample = 0.8, colsample_bytree = 0.8.

4.3.1.5 Advanced MLP (Adv-MLP)

Adv-MLP consists of a 3 hidden layer neural network (128, 64, 32 neurons in each layer respectively) with Batch Normalization and 20% Dropout added after each hidden layer, implemented in TensorFlow/Keras. It is trained with the Huber loss function:

$$\mathcal{L}_H(\hat{d}, d) = \begin{cases} \frac{1}{2}(\hat{d} - d)^2 & \text{if } \mid \hat{d} - d \mid \leq \delta \\ \delta\left(\mid \hat{d} - d \mid - \frac{\delta}{2}\right) & \text{otherwise} \end{cases} \tag{4.11}$$

where $\hat{d}$ is the predicted depth [mm] of the network, d is the true depth label [mm], and $\delta = 1.0$ mm is the point at which the error term switches from quadratic to linear. It is trained with the Adam optimizer [53] and an initial learning rate of 10^{-3}, with early stopping based on validation MAE with a patience value of 30. The full depth estimation process is summarized in Algorithm 3.

Algorithm 4.1: Physics-Informed Depth Estimation Pipeline

1. **Input:** $X \in \mathbb{R}^{H\times W\times T}$(raw thermal tensor); $M \in \{0,1\}^{H\times W}$(binary mask); depth labels; fold assignment; model from {RF, GBM, SVR, XGBoost, Adv-MLP}.
2. **Step 1 (Connected Components):** Extract regions from M; compute Area, Centroid, EquivDiameter per defect.
3. **Step 2 (Thermal Curves):** For each defect k: compute T_t for all $t$using Equation 4.3
4. **Step 3 (16 Features):** Compute all 16 features using Equations 4.4 to 4.6 and Tables 4.1 to 4.4
5. **Step 4 (Depth Labels):** Assign depth labels
6. **Step 5 (Train/Test Split):** Partition defects by fold; standardise using training-set statistics via Equation 4.9
7. **Step 6 (Standardise):** Fit each model on standardised training features and labels; compute MAE, RMSE, and R^2 on test defects
8. **Output:** $\hat{d} \in \mathbb{R}^K$predicted depths; evaluation metrics per model and fold.

4.3.1.6 Evaluation Metrics

Three evaluation metrics are used: the mean absolute error (MAE):

$$\text{MAE} = \frac{1}{N}\sum_{n=1}^{N} |\, \hat{d}_n - d_n \,| \tag{4.12}$$

the root mean square error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(\hat{d}_n - d_n)^2} \tag{4.13}$$

and the coefficient of determination:

$$R^2 = 1 - \frac{\sum_{n=1}^{N}(\hat{d}_n - d_n)^2}{\sum_{n=1}^{N}(d_n - \bar{d})^2} \tag{4.14}$$

where $\hat{d}_n$ is the predicted depth for defect n [mm], d_n is the true depth label [mm], $\bar{d} = \frac{1}{N}\sum_{n=1}^{N} d_n$ is the mean of the true depth labels, and N is the number of test defects. MAE is the main metric used as it yields the error in physical units of millimeters and is more interpretable. It is important to note that for C_I specimens, R^2 must be interpreted with care, as there are only two levels of depth in the test set separated by just 0.2mm. In such a situation, the denominator of Equation 4.14 becomes very small, and even slight absolute error may result in a very negative value of R^2. While these values are presented for completeness, MAE is the appropriate metric when assessing C_I specimen results.

4.3.2 Results: Fold 1

In Fold 1, models are trained on SQ_01 and CI_01 and tested on SQ_03 (25 square defects, depths 0.2-1.0mm) and CI_03 (6 circular defects, diameters 6, 10, 15mm, depths 2.0-2.2mm). Results are presented in Table 4.6.

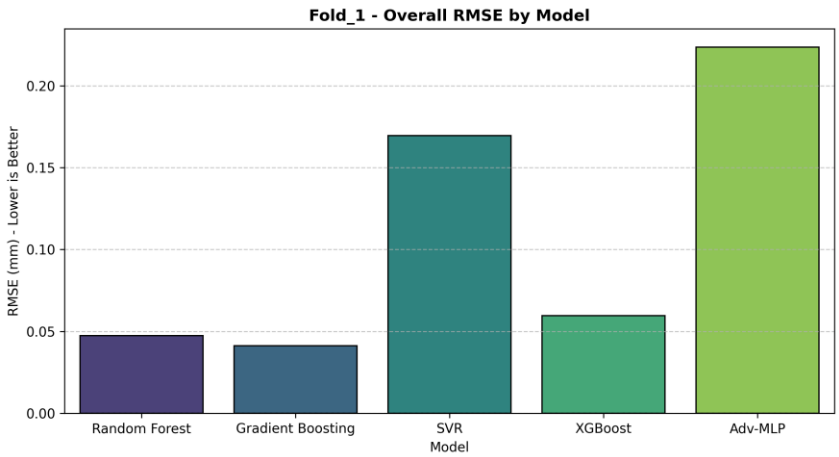


Figure 4.5: Fold 1 RMSE across all 5 models on SQ_03 and CI_03

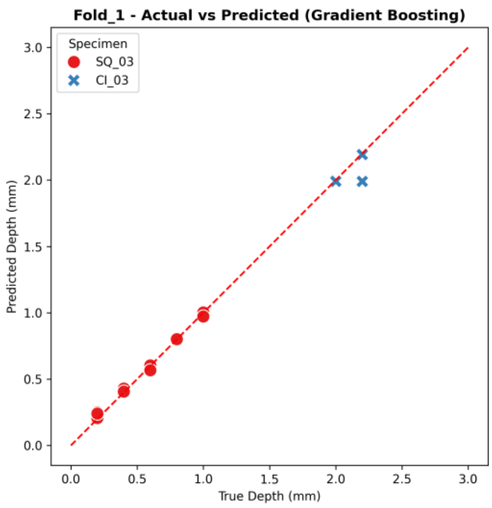


Figure 4.6: Actual depth versus predicted depth scatter-plot of Gradient Boosting for SQ_03 and CI_03 for Fold 1

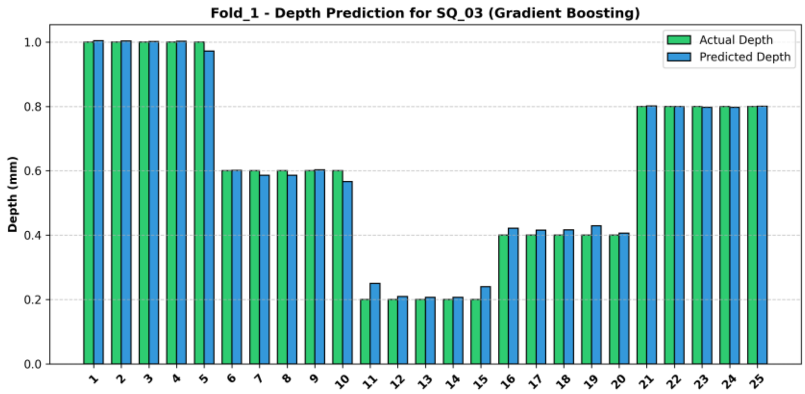


Figure 4.7: Depth per-defect bar chart for SQ_03, Fold 1, Gradient Boosting.

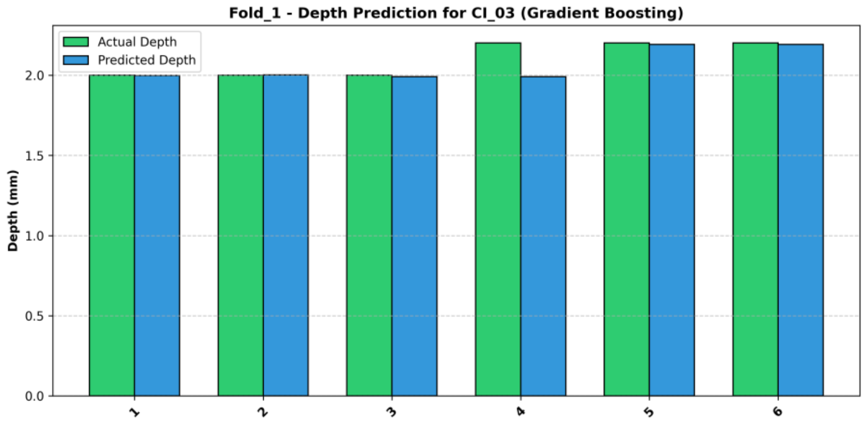


Figure 4.8: Depth per-defect bar chart for CI_03, Fold 1, Gradient Boosting

Table 4.6: Depth regression results for Fold 1. Test specimens: SQ_03 and CI_03

Model	Specimen	MAE (mm)	RMSE (mm)	R^2
Random Forest	SQ_03	0.0102	0.0171	0.9963
Random Forest	CI_03	0.0647	0.1021	-0.042
Random Forest	**Overall**	**0.0207**	**0.0475**	**0.9946**
Gradient Boosting	SQ_03	0.0125	0.0184	0.9958
Gradient Boosting	CI_03	0.0400	0.0861	0.259
Gradient Boosting	**Overall**	**0.0178**	**0.0413**	**0.9959**
SVR	SQ_03	0.1331	0.1692	0.642
SVR	CI_03	0.1321	0.1712	-1.930
SVR	**Overall**	**0.1329**	**0.1696**	**0.931**
XGBoost	SQ_03	**0.0099**	**0.0169**	**0.9964**
XGBoost	CI_03	0.1154	0.1309	-0.714
XGBoost	**Overall**	**0.0303**	**0.0596**	**0.9915**
Adv-MLP	SQ_03	0.1458	0.1725	0.628
Adv-MLP	CI_03	0.3385	0.3665	-12.43
Adv-MLP	**Overall**	**0.1831**	**0.2236**	**0.880**

As it can be seen in Table 4.6, Gradient Boosting has the best overall performance in Fold 1, with MAE 0.018mm and RMSE 0.041mm. The lowest single-specimen error was obtained using XGBoost on SQ_03 (MAE 0.010mm), but the CI_03 MAE was much higher and lowered the overall average significantly. Negative values are returned for R^2 for RF and XGBoost on CI_03 due to small variance in data compared to the total sum of squares in equation 4.14, not failure of the models; both have CI_03 MAE below 0.12mm. This is reasonable given the 0.2mm increments in level, and ±0.2mm uncertainty of depth labels on circular specimens.

4.3.3 Results: Fold 2

Models are trained on SQ_03 and CI_03, then tested on SQ_02 (curved plate, 25 square defects) and CI_02 (10 circular defects, 2.0-2.2mm depth, 2-20mm diameter). This represents the widest defect range tested. Results are presented in Table 4.7.

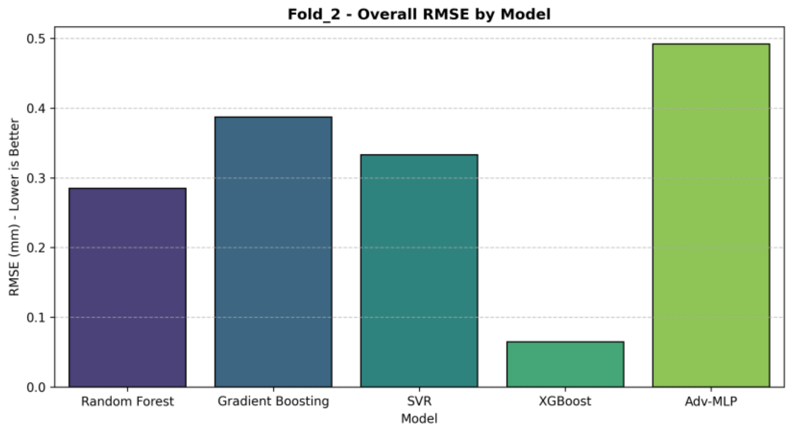


Figure 4.9: Fold 2 RMSE across all 5 models on SQ_02 and CI_02

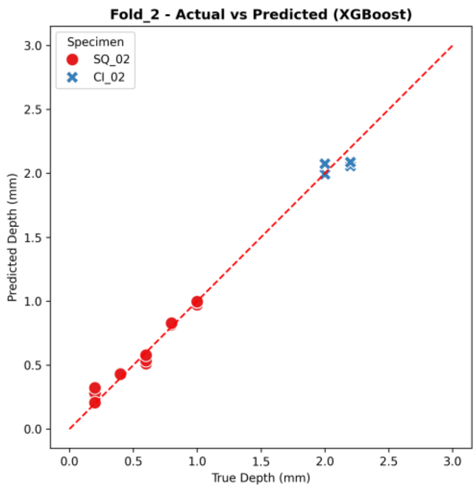


Figure 4.10: Actual depth versus predicted depth scatter-plot of XGBoost for SQ_02 and CI_02 for Fold 2

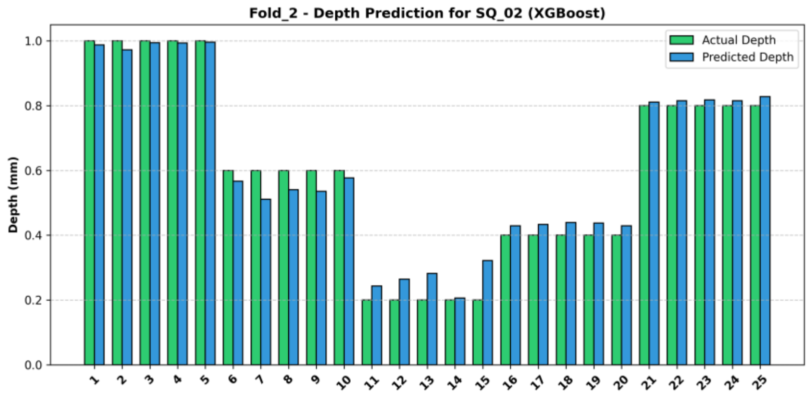


Figure 4.11: Depth per-defect bar chart for SQ_02, Fold 2, XGBoost

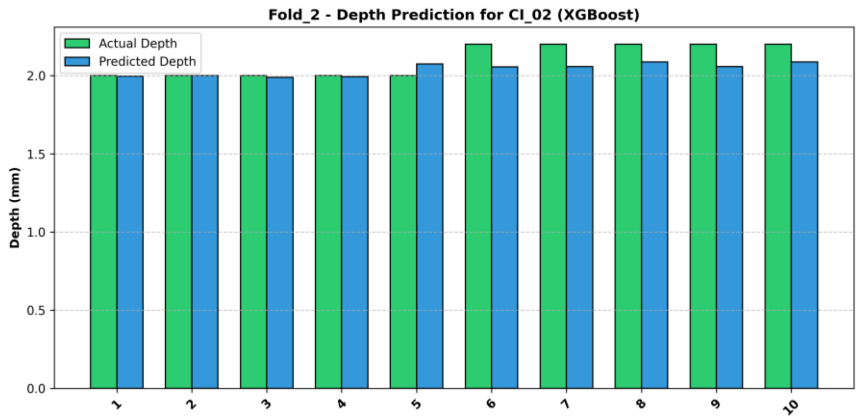


Figure 4.12: Depth per-defect bar chart for CI_02, Fold 2, XGBoost

Table 4.7: Depth regression results for Fold 2. Test specimens: SQ_02 and CI_02

Model	Specimen	MAE (mm)	RMSE (mm)	R^2
Random Forest	SQ_02	0.0120	0.0158	0.9969
Random Forest	CI_02	0.5209	0.5326	-27.37
Random Forest	**Overall**	**0.1574**	**0.2850**	**0.844**
Gradient Boosting	SQ_02	0.0162	0.0199	0.9951
Gradient Boosting	CI_02	0.6834	0.7232	-51.29
Gradient Boosting	**Overall**	**0.2068**	**0.3869**	**0.712**
SVR	SQ_02	0.3476	0.3844	-0.847
SVR	CI_02	0.1254	0.1363	-0.857
SVR	**Overall**	**0.2841**	**0.3330**	**0.786**
XGBoost	SQ_02	0.0359	0.0461	0.9734
XGBoost	CI_02	**0.0756**	**0.0967**	**0.066**
XGBoost	**Overall**	**0.0473**	**0.0647**	**0.992**
Adv-MLP	SQ_02	0.5221	0.5730	-3.103
Adv-MLP	CI_02	0.1370	0.1622	-1.632
Adv-MLP	**Overall**	**0.4121**	**0.4919**	**0.534**

As it can be seen, Fold 2 showed large variation between model results. RF and GBM both performed poorly on CI_02 with MAE of 0.521mm and 0.683mm respectively. The primary reason for the poor performance of these models is that due to being predominantly trained on square specimens, they over-fitted on the 5-column spatial format and failed to generalise when tested on the different layout and geometry of circular specimens. XGBoost avoided this by use of L1 and L2 regularization and column sub-sampling, resulting in CI_02 MAE of 0.076mm with a different training dataset. XGBoost yielded the best overall score for Fold 2 with MAE of 0.047mm and RMSE of 0.065mm.

4.3.4 Results: Fold 3

Models were trained on SQ_02 and CI_02 then tested on SQ_01 (flat plate, 25 square defects, 0.2-1.0mm depth) and CI_01 (16 circular defects, 2.0-2.2mm depth, 2-8mm diameter). Results are presented in Table 4.8.

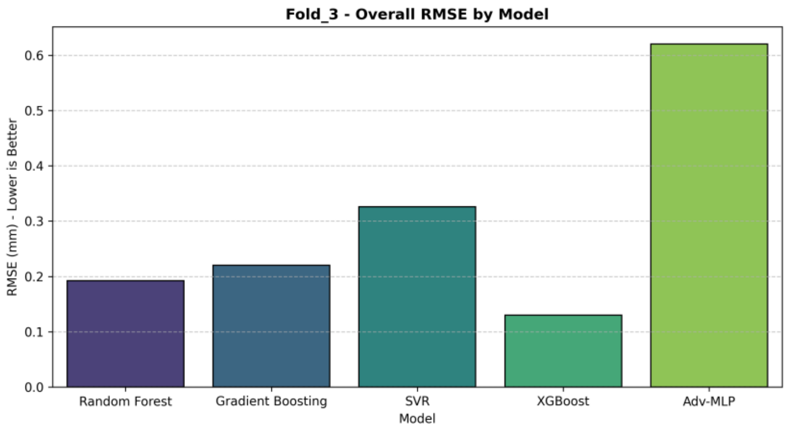


Figure 4.13: Fold 3 RMSE across all 5 models on SQ_01 and CI_01.

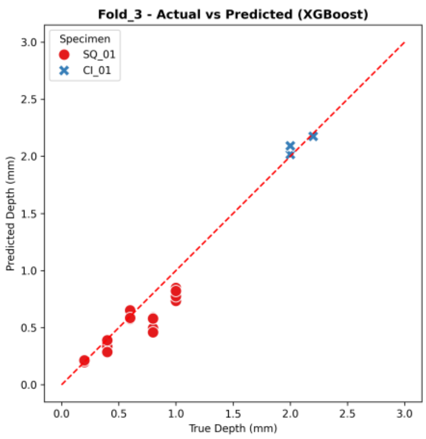


Figure 4.14: Actual depth versus predicted depth scatter-plot of XGBoost for SQ_01 and CI_01 for Fold 3.

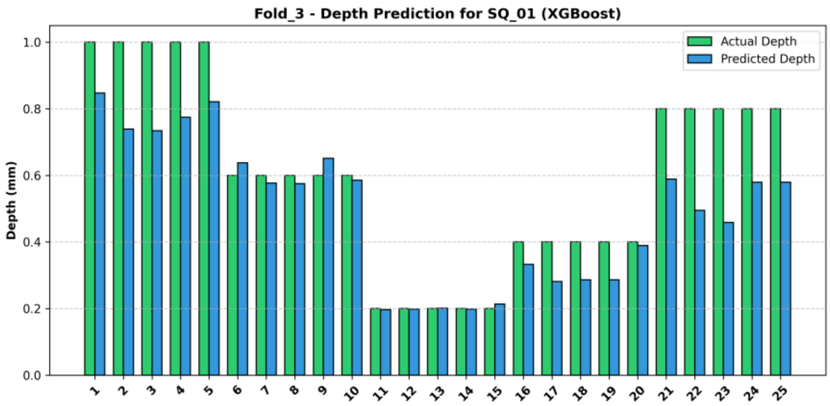


Figure 4.15: Depth per-defect bar chart for SQ_01, Fold 3, XGBoost.

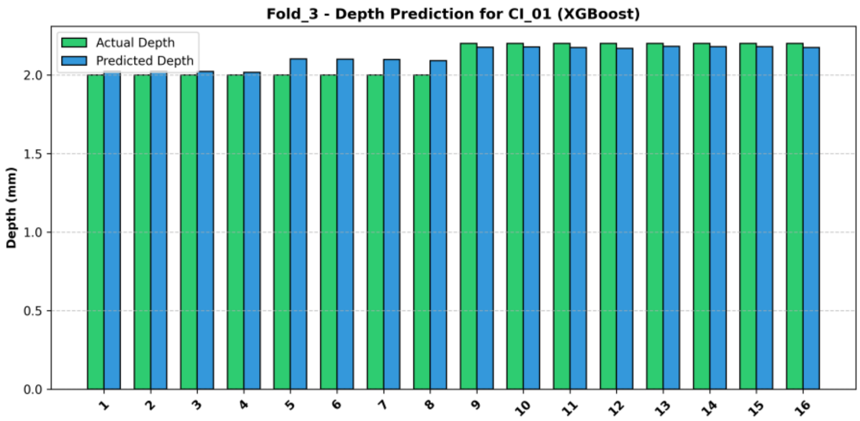


Figure 4.16: Depth per-defect bar chart for CI_01, Fold 3, XGBoost.

Table 4.8: Depth regression results for Fold 3. Test specimens: SQ_01 and CI_01.

Model	Specimen	MAE (mm)	RMSE (mm)	R^2
Random Forest	SQ_01	0.1751	0.2437	0.257
Random Forest	CI_01	0.0362	0.0414	0.829
Random Forest	**Overall**	**0.1209**	**0.1921**	**0.937**
Gradient Boosting	SQ_01	0.2215	0.2812	0.011
Gradient Boosting	CI_01	0.0260	0.0292	0.915
Gradient Boosting	**Overall**	**0.1452**	**0.2204**	**0.917**
SVR	SQ_01	0.2288	0.2692	0.094
SVR	CI_01	0.3870	0.3987	-14.90
SVR	**Overall**	**0.2905**	**0.3259**	**0.819**
XGBoost	SQ_01	0.1192	0.1608	0.677
XGBoost	CI_01	**0.0414**	**0.0527**	**0.722**
XGBoost	**Overall**	**0.0888**	**0.1298**	**0.971**
Adv-MLP	SQ_01	0.3899	0.4293	-1.303
Adv-MLP	CI_01	0.8308	0.8356	-68.82
Adv-MLP	**Overall**	**0.5620**	**0.6204**	**0.346**

As it can be seen from Table 4.8, XGBoost once again attained the best overall performance with MAE 0.089mm. Fold 3 produced the highest performance of circular specimens. GBM predicted CI_01 MAE 0.026mm ($R^2 = 0.915$) while XGBoost predicted 0.041mm ($R^2 = 0.722$). The improved circular performance in this fold is thought to be due to the inclusion of 10 defects on CI_02 instead of 6 in CI_03 in the training data which provides better circular data coverage. Performance on square specimens was not as good for any model. This is likely due to the differing surface temperature distributions arising from testing a flat plate (SQ_01) versus curved/trapezoidal surfaces (SQ_02 and SQ_03) after post-flash: this alters the time-to-peak relationship at each depth, hence producing lower predicted depths with flatter surface temperatures.

4.3.5 Comparative analysis and discussion

Cross validated average performance over the 3 folds is summarised in Table 4.9, ordered by CV-average MAE.

Table 4.9: Three-fold cross validated average performance

Rank	Model	CV-Avg MAE (mm)	CV-Avg RMSE (mm)	Notes
1	**XGBoost**	**0.0555**	**0.0847**	Best in Folds 2 and 3; second in Fold 1
2	Random Forest	0.0997	0.1749	Best in Fold 1 (joint); fails CI_02
3	Gradient Boosting	0.1233	0.2162	Wins Fold 1; collapses CI_02 badly
4	SVR	0.2359	0.2762	Inconsistent; geometry-dependent failures
5	Adv-MLP	0.3857	0.4453	Consistently worst; insufficient training data

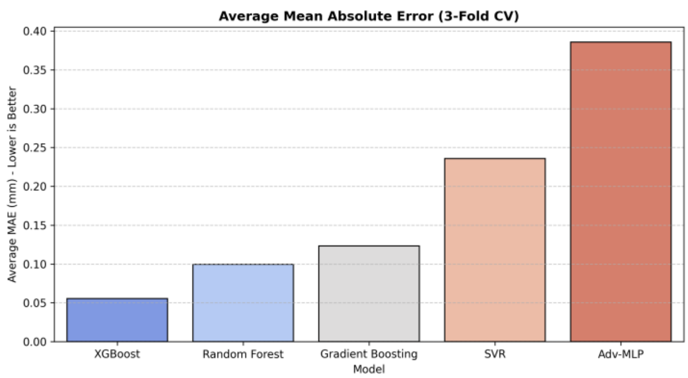


Figure 4.17: CV-average MAE comparison across the three folds for all five models.

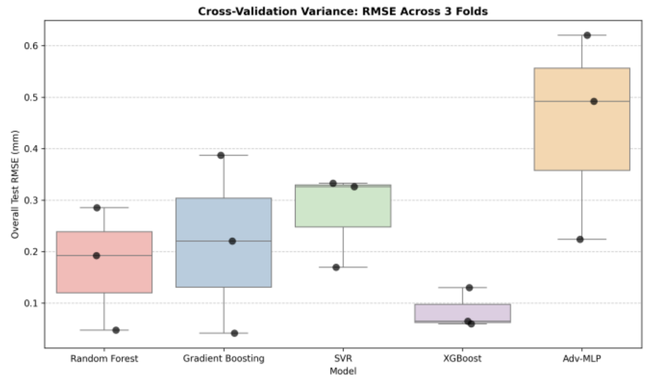


Figure 4.18: Box plot showing per-fold RMSE across the three folds for all five models to assess fold-to-fold stability.

Table 4.9 shows that XGBoost produced the best overall performance in terms of CV-average MAE at 0.056 mm, which is nearly half the MAE of the second best model, Random Forest. This also represents the most stable generalisation across folds, as evidenced by the lowest RMSE standard deviation, or tightness across folds, shown in Figure 4.18.

It is notable that the RF and GBM models failed on CI_02 (yielding MAEs of 0.521 and 0.683 respectively) during Fold 2, while the top performing models on that fold performed relatively well. This failure can be attributed to the nature of the training data in Fold 2. The training data at this level contained an overwhelmingly large percentage of SQ- specimens with the five column depth structure, leading the models to over-fit this characteristic shape and fail on the five distinct circle defects at each of the 5 depths present in CI_02. XGBoost successfully avoided this over-fitting by virtue of its regularisation mechanisms. A similar, albeit less pronounced, trend is evident on SQ_01 at fold 3; in this case, the model was trained on curved specimens and thus was unable to accurately reconstruct depth in the flat, straight CS01 specimen.

The Adv-MLP consistently performed the worst on all three folds; it reported an MAE of 0.831 mm on CI_01 at fold 3. This exceeds the full depth range of that specimen (0.0 to 0.2 mm) by over four-fold, clearly showing the network had simply memorized the training examples and not learned the physics. Considering the low number of labeled defects per fold (31 to 41) and the large number of trainable parameters in the Adv-MLP network (~10,000), it becomes clear that more classical ensemble methods using physics-motivated features (as used in models 1-4) are much more appropriate for a problem in which the number of physical specimens are limited.

4.4 Three-Dimensional defect reconstruction

4.4.1 Reconstruction Methodology

CRFP delaminations were approximated as planar defects that did not vary with spatial dimension but whose through-thickness dimensions were negligibly small compared to the lateral extent of the defect. In this approximation each segmented defect region will lie on a planar surface, and there is some specified separation between the upper surface and lower surface of the defect region. If a defect at pixel (i, j) is reconstructed as a plane at depth $\hat{d}$ below the surface, the coordinates of the top and bottom of the defect region are given by:

$$Z_{top}(x, y) = \hat{d} \forall (x, y) \in M_{seg} \quad (4.15)$$

$$Z_{bot}(x, y) = \hat{d} + \delta_{gap} \forall (x, y) \in M_{seg} \quad (4.16)$$

Here $\hat{d}$ is the depth predicted by the XGBoost regression models, M_{seg} is the pixel region of the defect defined by the mask in Chapter 3, and $\delta_{gap} = 0.0075$ mm is the nominal thickness of the Teflon insert between the layers. The pixel coordinates $(i, j)$are converted to physical units (mm) by use of a spatial calibration factor:

$$x_{mm} = j \cdot s_{px}, \quad y_{mm} = i \cdot s_{px}, \qquad s_{px} = \frac{L_{spec}}{N_{px}} \quad (4.17)$$

where j and i are the column and row pixel indexes respectively and $L_{spec} = 300$mm is the physical length of each specimen. Using the transformation in Equation 4.17, two point clouds are formed at the location of each pixel (i, j)belonging to the defect:

$$p_{ij}^{top} = (j \cdot s_{px}, i \cdot s_{px}, \hat{d}) \quad (4.18)$$

$$p_{ij}^{bot} = (j \cdot s_{px}, i \cdot s_{px}, \hat{d} + \delta_{gap}) \quad (4.19)$$

Applying Delaunay triangulation to the (x, y) coordinates from each of the sets of points results in smooth, triangulated surface meshes for the top and bottom interfaces. We obtain the ground truth surfaces from the nominal depths defined in equations 4.7 and 4.8, again using Delaunay triangulation, such that predicted and ground truth meshes can be drawn within the same scene and superimposed upon each other. Under this model, the mean surface distance between the predicted and ground truth surface reduces to:

$$\text{MSD} = | \hat{d} - d_{true} | \quad (4.20)$$

where d_{true} is the nominal depth [mm]. The MSD is numerically the same as the per-defect regression error computed in Section 4.3 since the same depth value is used for all points in a defect region and no additional error is generated by the reconstruction process. Three more values are computed for each defect: lateral equivalent diameter $D_k = 2\sqrt{A_k \cdot s_{px}^2/\pi}$[mm], reconstructed volume $V_k = A_k \cdot s_{px}^2 \cdot \delta_{gap}$[mm³], and $\mathrm{MSD}_k$from Equation 4.20. Algorithm 4.2 summarises the three dimensional reconstruction process.

Algorithm 4.2: Three-Dimensional Defect Reconstruction

Input: M_{seg}, d_{pred}, d_{true}, $s_{px} = 0.586$, $\Delta_{gap} = 0.0075$
CC = ConnComp(M_{seg})→ K defects, A_k, centroids
for $k = 1, \ldots, K$**do**
  $\text{pts}^{\text{top}} = \{(j \cdot s_{px}, i \cdot s_{px}, d_{\text{pred}}[k]) : (i, j) \in \text{CC}_k\}$
  $\text{pts}^{\text{bot}} = \{(j \cdot s_{px}, i \cdot s_{px}, d_{\text{pred}}[k] + \Delta_{gap})\}$
  $\text{tri}^{\text{top}} = \text{Delaunay}(\text{pts}_{xy}^{\text{top}})$
  repeat for $d_{\text{true}}[k]$
  $\text{diam}_k = 2\sqrt{A_k \cdot s_{px}^2/\pi}$
  $V_k = A_k \cdot s_{px}^2 \cdot \Delta_{gap}$
  $\text{MSD}_k = \mid d_{\text{pred}}[k] - d_{\text{true}}[k] \mid$
end for
Render & export 3D
Output: meshes, $\{\text{diam}_k, d_k, V_k, \text{MSD}_k\}$

4.4.2 Reconstruction results

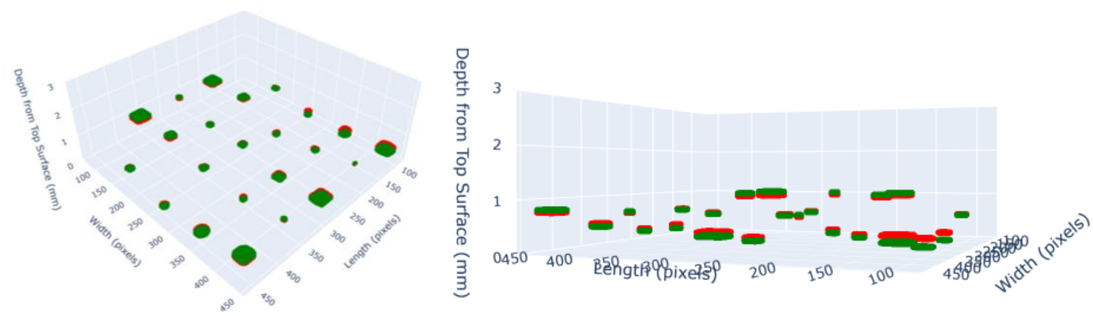


Figure 4.19: 3D reconstruction of SQ_03 (Fold 1 test specimen).Red semi-transparent surfaces indicate the predicted defect surface, while green surfaces represent the ground truth defect surface. All 25 defects were placed at their respective depth levels and within a specimen-shape transparent block of material.

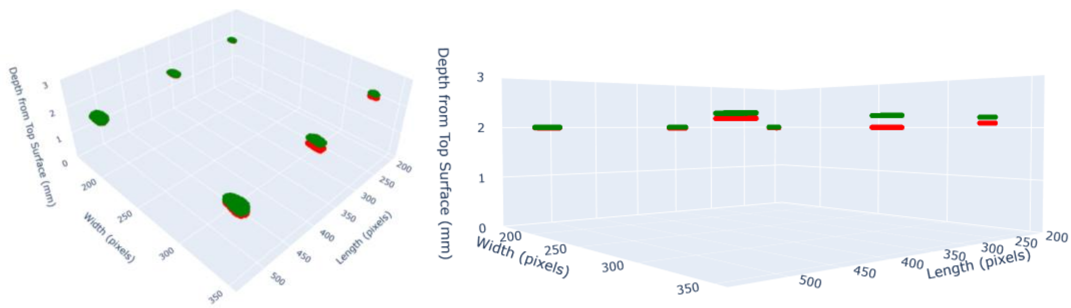


Figure 4.20: 3D reconstruction of CI_03 (Fold 1 test specimen). Six circle defects were placed at 2.0 and 2.2 mm depths, inside a transparent block of material.

The predicted surfaces in Figure 4.19 closely follow the shape of the ground truth surfaces for the columns with small predicted depth errors (i.e., <0.015 mm). Slight inaccuracies are more noticeable in the deeper defect columns as expected. As discussed, it is clearly apparent in Figure 4.20 that the 0.2 mm difference in depths has been successfully identified for the circle defects, and that the spatial resolution is not a limitation here, despite a negative R^2. The negative R^2 on CI specimens is simply an indication of how small the denominator problem in Equation 4.14 and do not indicate failure of the spatial reconstruction.

4.5 Chapter Summary

This chapter presents a physics-informed depth estimation and 3-D reconstruction methodology for the characterisation of defects in CFRP utilizing optical pulse thermography. The physical basis is the relationship $d \propto \sqrt{t_{peak}}$described by Equation 4.2. Sixteen features, categorized into temporal, energy-based, statistical, and geometrical, and defined with full notation in Section 4.2.2, are extracted from the averaged thermal profile of each segmented defect. Five regression models were evaluated under 3-fold cross-validation according to the same partitioning strategy used in Chapter 3.

The resulting 3-D reconstruction is obtained by combining XGBoost depth estimations and Chapter 3 segmentation masks via Delaunay triangulation. Numerical MAE of mean surface distance from Equation 4.20, is the same as the per-defect MAE, thereby showing no additional error is added in reconstruction. The subsequent 3-D models provided the entire geometry of the defect population in each sample and tied the per-defect numerical output in Section 4.3 to a physically intuitive overview of the sample.

CHAPTER 5
CONCLUSIONS AND FUTURE DIRECTIONS

5.1 Principal Contributions

The work reported in this thesis is concerned with automated defect detection, segmentation, and depth estimation on CFRP laminates using optical pulsed thermography inspection. The motivation for this thesis was a clear gap in the research, which is the development of deep learning based pipelines for thermographic inspection based primarily on raw, single-channel thermal data, where defect signature, background gradient and emissivity variation is mapped into one single grey level, making deep learning pipelines unable to effectively deal with low contrast, poor illumination and loss of temporal data used to establish depth. Four main contributions address different facets of this problem.

Contribution 1: Systematic baseline evaluation was conducted using six different architectures trained with raw, single-channel thermal images using identical experimental parameters with specimen-level cross-validation. All six models showed a failure characteristic of the representation - good results for the square specimens (Dice: 0.83-0.87, mAP: 0.84-0.91), but near-total failure for circular specimens (Dice: 0.24-0.29, mAP: down to 0.06) that was independent of model architecture for single-stage detection [38], transformer-based detection [40],[41], two-stage detection [35], segmentation using skip-connection networks [23], multi-scale pyramid networks [36] and atrous pooling architectures [44].

Contribution 2: three-channel thermal representation was proposed and applied using a novel fusion approach where each image contains the raw intensity, principal components image [14],[30] and the thermographic signal reconstruction coefficient image [27],[28]. Cyclic modular alignment was used to resolve the mismatch in channel count (101 raw images per specimen vs 6 processed images per specimen). Ablation studies have shown that this three-channel approach yields an improvement on the Dice of 40-60% from the PC image, an additional 10-20% from the TSR image, and roughly 5% stabilization of the raw thermal data. Detection mAP increased by 4-11%, and segmentation Dice for circular specimens by 14-51% depending on the architecture, but results remained below square specimen Dice of 0.83-0.87.

Contribution 3: Physics-informed depth estimation. A 16-feature regression method based on $d \propto \sqrt{t_{\text{peak}}}$ [24] was tested using five model types. The best performing, XGBoost [55], had a cross-validation mean absolute error of 0.056 mm and RMSE of 0.085 mm and time-to-peak was always the best predictor, ~40% of feature importance [50].

Contribution 4: End-to-end, thin-delamination 3D reconstruction pipeline was presented. It uses the segmentations and the best depth regression model to render an interactive, surface plot of a defect in 3D and provides 3-5 second processing for each specimen including the production of quantitative, geometric information on the defect geometry.

5.2 Significant Quantitative Findings

For segmentation, the most striking improvement using the three-channel thermal representation came from performance on circular specimens. With U-Net, Dice went from 0.294 to 0.34 (+16%), FPN went from 0.258 to 0.39 (+51%); DeepLabV3+ went from 0.237 to 0.27 (+14%). This performance gain is robust across all three architectures, indicating that the improvement is due to the representation, not specific network design. Performance for circular specimens remains somewhat lower Dice 0.27-0.39 than for square ones Dice 0.83-0.87, showing that while the gap is reduced, it is not eliminated and circular specimens represent an open problem for thermography based segmentation.

For detection, the gain, though less dramatic, was again equally robust YOLOv8n improved from mAP 0.89 to 0.94, RT-DETR-L from 0.91 to 0.95, Faster R-CNN from 0.84 to 0.89. All architectures improved to a level where multiple otherwise complete detection failures for circular specimens, where only raw thermal inputs were used zero detections from Faster R-CNN and RT-DETR-L for CI01 and CI02, were now partial successes.

In terms of depth estimation with the XGBoost model [55], combined mean error across geometries was 0.056 mm, and R 0.986 square specimens achieved R of 0.935, circular specimens R of 0.205. The low value of R for circular specimens is due to their limited label variance, all test defects were at two depth levels approximately 0.2mm apart so the denominator of R was small.

5.3 Acknowledged Limitations

The data consists of only six CFRP specimens and two types of defect. After augmentation, the number of image frames was raised to 3030 but the specimens, geometries and defect type did not vary. In real-world scenarios such as inspecting curved panels, joints with non-planar geometry and complex impacts, different material thicknesses, and defects propagating between multiple plies [5, 20], further data with a higher variety of specimens, typically around 20 to 50 would be required for robust generalization tests

The depth label for circular specimens was acquired using a heuristic ranking approach assuming the time-to-peak of larger, deeper flaws would appear before that of smaller, shallower defects. This is known to break for small defects at moderate depth, as their time-to-peak may appear later due to 3D heat diffusion [24]. The accuracy labels were acquired using ultrasonic measures and cutting through the specimen, but ideally, separated component of model error from label error would have been beneficial.

The parameters in preprocessing, including the frame selection window size shown in Table 3.1, the degree of TSR fitting (n=5), and the number of retained principal components (m=6) were set based on exploratory analysis and literature values [27], [30] rather than through rigorous optimisation; while effective for this particular dataset, these parameters would require fine-tuning for other material properties, flash durations and frame rates.

The thin-delamination model gives the depth as a single value to the defect and does not account for variations in defect depth across its area nor multiple interfaces [5], [20].The thin-delamination model assigns one constant depth to each defect region and assumes flat geometry. Real delaminations vary in depth across their extent and propagate at multiple ply interfaces [5],[20]. The current approach cannot represent either complication.

5.4 Directions for Future Investigation

The greatest unknown that remains is how the current methodology transfers to specimens with geometrically complicated structure. For example, curved panels, T-shaped joints, or genuine aerospace components would allow investigation of whether the assumption of 1D heat diffusion in depth prediction holds up for non-trivial surfaces.

Models combining CNN-based spatial processing with transformer-based attention mechanisms such as the Swin Transformer [56] could potentially capture long-range spatial dependencies between different areas of a defect or different defects altogether that are missed by local CNN kernels. Training with large unlabeled thermal datasets using unsupervised methods like masked frame prediction or temporal ordering would allow us to take fuller advantage of archival data, where manual labeling is time consuming.

Semi-supervised learning approaches such as consistency regularization and pseudo-labeling have been shown to be effective in reducing the dependence on labelled training data and might prove useful here as well. Optimization techniques such as quantization, pruning and knowledge distillation can be used to reduce the parameter size of the current 21 to 41 Million parameter segmentation networks to fit onto compact, embedded devices, the proposed detection and segmentation pipeline has enormous potential for on-the-spot-inspection hardware, but it currently has a parameter size which would render such usage impractical.

Finally, fusing thermography and ultrasonic C-scans could tackle many limitations at once: the labels for ultrasonic measures would give ground truth for the depth estimation task and could replace the heuristic ranking system for circular specimens [7], [8], and the CNN models would fuse the spatial breadth of thermal data with the depth resolution of ultrasonics . Testing with other materials, such as glass fiber reinforced composites, metal composites and ceramic matrix composites would evaluate the generality of the presented representation.

5.5 Concluding Observations

The primary finding of this thesis is that unifying physics-informed preprocessing and raw intensity information into a common input rather than treating them as a fixed cascade upstream from the network results in reliable and significant gains in automated thermographic defect inspection. Circular specimen segmentation Dice scores rose from an average of 0.26 to about 0.33 across three architectures with the highest single increase found in FPN 0.258 to 0.39. MAP scores for object detection increased by 4 to 11 points across the three object detection architectures. The combined system generates three-dimensional defect maps from raw thermal data interactively in

3-5 seconds and depth accuracy is close to what is achievable with current ground truth accuracy.

More generally, the central design principle is extensible beyond the problem addressed in this work. When multiple signal transforms each containing information complementary to one another, the benefits of joint optimization on the input rather than hardcoding interactions as part of a preprocessing sequence can be realized. The cyclic alignment mechanism for unequal channel numbers and the physics-informed feature validation used for depth estimation are both techniques generalizable to other multi-modal inspection problems apart from CFRP thermography.

For maintenance or manufacturing quality control of aerospace components, the system as architected herein can enable faster screening of cured composites, a handheld, in-field inspector or a more thorough, high-accuracy defect characterization for damage tolerance analyses and digital twin representations. The cross-validation mechanism used here for a single specimen can provide a blueprint for a robust test set for future systems development. A fully autonomous thermographic inspector will require many more layers of robustness before safety-critical application, however, this thesis shows how judicious input representation, informed by the physics of the measurement and confirmed with well-controlled experiments, may result in consistent and practical benefits.

REFERENCES

[1] Boeing Commercial Airplanes, "787 Dreamliner Program Fact Sheet," Boeing, Seattle, WA, USA, 2020.

[2] Airbus, "A350 XWB Family: Setting New Standards," Airbus S.A.S., Toulouse, France, Tech. Rep., 2021.

[3] D. W. Jensen and L. C. Bank, "Composite materials for wind turbine blades: Issues and challenges," in Proc. American Society for Composites 29th Technical Conf., San Diego, CA, USA, Sep. 2014, pp. 1–12.

[4] S. Abrate, Impact on Composite Structures. Cambridge, UK: Cambridge University Press, 1998.

[5] C. Kassapoglou, Design and Analysis of Composite Structures: With Applications to Aerospace Structures, 2nd ed. Chichester, UK: Wiley, 2013.

[6] ICF International, "Global aerospace and defense maintenance, repair and overhaul market forecast 2021–2031," ICF, Fairfax, VA, USA, Market Research Rep., 2021.

[7] P. Cawley and R. D. Adams, "The mechanics of the coin-tap method of non-destructive testing," J. Sound and Vibration, vol. 122, no. 2, pp. 299–316, Apr. 1988.

[8] I. N. Berketis, P. T. Tzitzinou, and D. A. Zacharias, "Ultrasonic inspection of carbon-fiber-reinforced polymer laminates," British J. Non-Destructive Testing, vol. 35, no. 6, pp. 318–323, Jun. 1993.

[9] P. J. Schilling, B. R. Karedla, A. K. Tatiparthi, M. A. Verges, and P. D. Herrington, "X-ray computed microtomography of internal damage in fiber reinforced polymer matrix composites," Composites Science and Technology, vol. 65, no. 14, pp. 2071–2078, Nov. 2005.

[10] X. P. V. Maldague, Theory and Practice of Infrared Technology for Nondestructive Testing. New York, NY, USA: Wiley, 2001.

[11] C. Meola and G. M. Carlomagno, "Recent advances in the use of infrared thermography," Measurement Science and Technology, vol. 15, no. 9, pp. R27–R58, Sep. 2004.

[12] X. Maldague and S. Marinetti, "Pulse phase infrared thermography," J. Applied Physics, vol. 79, no. 5, pp. 2694–2698, Mar. 1996.

[13] S. M. Shepard, "Advances in pulsed thermography," in Proc. SPIE Thermosense XXIII, vol. 4360, Orlando, FL, USA, Apr. 2001, pp. 511–515.

[14] N. Rajic, "Principal component thermography for flaw contrast enhancement and flaw depth characterisation in composite structures," Composite Structures, vol. 58, no. 4, pp. 521–528, Dec. 2002.

[15] B. Yang and Y. Huang, "Automated detection and characterization of CFRP subsurface defects using pulsed thermography and deep learning," NDT and E International, vol. 118, p. 102397, Mar. 2021.

[16] W. Guo, X. Chen, Y. Shen, and M. Cao, "Deep learning for infrared thermography data-driven prognostics of composite materials," Composite Structures, vol. 252, p. 112681, Nov. 2020.

[17] K. K. Chawla, Composite Materials: Science and Engineering, 3rd ed. New York, NY, USA: Springer, 2012.

[18] A. C. Campbell, "The effect of autoclave pressure on the properties of carbon fiber-epoxy laminates," J. Composite Materials, vol. 24, no. 4, pp. 432–445, Apr. 1990.

[19] L. J. Hart-Smith, "The role of biaxial stresses in discriminating between meaningful and illusory composite failure theories," Composite Structures, vol. 25, no. 1–4, pp. 3–20, 1993.

[20] M. R. Wisnom, "The role of delamination in failure of fibre-reinforced composites," Phil. Trans. R. Soc. A, vol. 370, no. 1965, pp. 1850–1870, Apr. 2012.

[21] H. S. Carslaw and J. C. Jaeger, Conduction of Heat in Solids, 2nd ed. Oxford, UK: Clarendon Press, 1959.

[22] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, Jun. 2016, pp. 779–788.

[23] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in Proc. Int. Conf. Medical Image Computing and Computer-Assisted Intervention (MICCAI), Munich, Germany, Oct. 2015, pp. 234–241.

[24] J. G. Ekanayake, T. P. Murrugarra, and B. Karunasena, "Depth prediction of defects in CFRP laminates using lock-in thermography," Measurement Science and Technology, vol. 29, no. 3, p. 035502, Feb. 2018.

[25] C. Ibarra-Castanedo and X. Maldague, "Pulsed phase thermography reviewed," Quantitative InfraRed Thermography J., vol. 1, no. 1, pp. 47–70, Jun. 2004.

[26] C. Ibarra-Castanedo, D. Gonzalez, M. Klein, M. Pilla, S. Vallerand, and X. Maldague, "Infrared image processing and data analysis," Infrared Physics and Technology, vol. 46, no. 1–2, pp. 75–83, Dec. 2004.

[27] S. M. Shepard, J. R. Lhota, B. A. Rubadeux, D. Wang, and T. Ahmed, "Reconstruction and enhancement of active thermographic image sequences," Optical Engineering, vol. 42, no. 5, pp. 1337–1342, May 2003.

[28] S. M. Shepard and T. Ahmed, "Thermographic derivative analysis," in Proc. SPIE Thermosense XXXIII, vol. 8013, Orlando, FL, USA, Apr. 2011, p. 80130R.

[29] C. Winfree, "Thermal diffusivity imaging of aerospace materials and structures," Proc. SPIE, vol. 2473, pp. 26–35, Jun. 1995.

[30] N. Rajic, "Principal component thermography," Defence Science and Technology Organisation, Melbourne, Australia, Tech. Rep. DSTO-TR-1298, 2002.

[31] E. J. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" J. ACM, vol. 58, no. 3, pp. 1–37, Jun. 2011.

[32] Y. Liu, Z. Wu, X. Zhao, G. Y. Tian, and H. Zhang, "Sparse representation for infrared dim target detection via a discriminative over-complete dictionary learned online," Infrared Physics and Technology, vol. 60, pp. 247–256, Sep. 2013.

[33] J. Ahmed, B. Gao, and W. L. Woo, "Wavelet-integrated alternating sparse dictionary matrix decomposition in thermal imaging CFRP defect detection," IEEE Trans. Industrial Informatics, vol. 15, no. 7, pp. 4033–4043, Jul. 2019.

[34] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, Jun. 2016, pp. 770–778.

[35] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[36] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, Jul. 2017, pp. 2117–2125.

[37] D. Lema, P. Pedrayes, D. Garcia, C. Usamentiaga, and H. Gonzalez, "Automated detection of subsurface defects using active thermography and deep learning

object detectors," IEEE Trans. Instrumentation and Measurement, vol. 71, pp. 1–10, 2022.

[38] G. Jocher, A. Chaurasia, and J. Qiu, "YOLO by Ultralytics," GitHub, 2023. [Online]. Available: https://github.com/ultralytics/ultralytics

[39] Q. Fang, H. Zhang, Y. Wang, and S. Li, "Automatic detection of impact damage in carbon fiber composites using YOLOv8 with thermographic images," Composite Structures, vol. 315, p. 116990, Jul. 2023.

[40] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in Proc. European Conf. Computer Vision (ECCV), Glasgow, UK, Aug. 2020, pp. 213–229.

[41] Y. Zhao, W. Lv, S. Xu, J. Wei, G. Wang, Q. Dang, Y. Liu, and J. Chen, "DETRs beat YOLOs on real-time object detection," in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, Jun. 2024, pp. 16965–16974.

[42] M. Schmid, P. Hoffmann, and C. Schick, "Transformer-based defect detection in lock-in thermography of CFRP with attention mechanisms," Composite Structures, vol. 328, p. 117731, Jan. 2024.

[43] Y. Zhou, Q. He, Y. Shou, and Y. Wang, "Deep learning-based semantic segmentation of defects in composite materials from thermographic images," Measurement, vol. 188, p. 110598, Jan. 2022.

[44] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in Proc. European Conf. Computer Vision (ECCV), Munich, Germany, Sep. 2018, pp. 801–818.

[45] L. Morelli, L. Palumbo, M. Iannino, and M. Galietti, "DeepLabV3+ for automated defect detection in lock-in thermography of composites," NDT and E International, vol. 132, p. 102710, Dec. 2022.

[46] P. Rosa, P. Caggiano, and L. Nele, "Thermographic signal reconstruction and deep learning for automatic defect segmentation in CFRP," Composites Part B: Engineering, vol. 270, p. 111135, Feb. 2025.

[47] B. Liu, H. Zhang, C. Wu, and J. Zhang, "Generative principal component thermography with GANs for defect detection data augmentation," IEEE Trans. Industrial Electronics, vol. 68, no. 10, pp. 10150–10160, Oct. 2021.

[48] C. Hu, W. Yang, D. Wen, W. Xie, and L. Ye, "Lightweight spatial-temporal multi-feature fusion network for defects detection in CFRP," IEEE Trans. Instrumentation and Measurement, vol. 70, pp. 1–12, 2021.

[49] H. Dong, M. Song, and W. Gao, "3D ResNet for depth prediction from pulsed thermography in CFRP inspection," Infrared Physics and Technology, vol. 115, p. 103710, Jun. 2021.

[50] Y. Peng and S. Li, "Random Forest regression for defect depth estimation in thermographic NDT with hand-crafted features," NDT and E International, vol. 119, p. 102416, Apr. 2021.

[51] Z. Wei, B. Fernandes, H. Herrmann, J. M. Tarpani, and J. A. O. Barros, "A deep learning method for the impact damage segmentation of curve-shaped CFRP specimens inspected by infrared thermography," Materials, vol. 14, no. 9, p. 2181, Apr. 2021.

[52] J. Erazo-Aux, C. Loaiza-Correa, J. Restrepo-Giron, and H. Montoya, "Thermographic dataset of debonded CFRP samples with artificially induced defects," Data in Brief, vol. 32, p. 106313, Oct. 2020.

[53] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in Proc. Int. Conf. Learning Representations (ICLR), San Diego, CA, USA, May 2015.

[54] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, Oct. 2001.

[55] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, San Francisco, CA, USA, Aug. 2016, pp. 785–794.

[56] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV), Virtual, Oct. 2021, pp. 10012–10022.