BRIDGE PIER SURFACE DEFECT DETECTION BASED ON IMPROVED YOLOV9

Hanzhe Cai

Dr. Divya Chidambaram

Project submitted as part of the requirements for the award of MSc Software Engineering and Artificial Intelligence

September 2024

Declaration of Originality

I, Hanzhe Cai declare that I am the sole author of this Project; that all references cited have been consulted; that I have conducted all work of which this is a record, and that the finished work lies within the prescribed word limits.

This work has not previously been accepted as part of any other degree

submission.

Signed : ....

26AUG,2024

Date : ....

FORM OF CONSENT

I Hanzhe Cai hereby consent that my Project, submitted in candidature for the MSc

Software Engineering and Artificial Intelligence,, if successful, may be made available for inter-library loan or photocopying (subject to the law of copyright), and that the title and abstract may be made available to outside organisations.

Signed : ....

26AUG,2024

Date : ....

Copyright Acknowledgement

ABSTRACT/SYNPOSIS

This research introduces an innovative surface defect detection methodology specifically designed for bridge piers, which integrates state-of-the-art image enhancement algorithms with sophisticated target detection frameworks. This hybrid approach effectively addresses some of the inherent limitations observed in existing deep learning-based defect detection methodologies, particularly under conditions of suboptimal image quality and challenges related to the detection of minute targets. Comparative results demonstrate that this novel technique achieves a 3.9% increase in the mean Average Precision (mAP50) over the baseline model. Furthermore, this is accomplished with a reduction in model complexity, as evidenced by a 9.8% decrease in the number of parameters and a substantial reduction in computational demand, quantified as a 7.5 GFLOPS decrease. This study not only advances the field of structural health monitoring but also enhances the operational efficiency of automated defect detection systems.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I sincerely thank Dr. Divya Chidambaram from the University of Wales Trinity Saint David for her meticulous guidance and valuable advice on this research. Dr. Chidambaram's expertise and insightful perspectives greatly enriched the content and depth of this study. Additionally, I am grateful for the advanced experimental facilities provided by Wuhan University of Technology. These facilities have strongly supported our experimental research, enabling us to achieve significant results in our explorations. This support and resources were crucial for the completion of this study, and I am deeply appreciative of them.

CHAPTER 1 - INTRODUCTION

1.1 Introduction

The safety and structural integrity of bridge piers are paramount for ensuring the overall performance and safety of bridges [1]. High-quality piers are designed to support substantial structural loads and provide enhanced resistance to natural disasters. This robust construction ensures the safety of people and vehicles that traverse the bridge and contributes to lower maintenance costs over time. Furthermore, building bridge abutments to high standards not only ensures compliance with stringent regulatory requirements but also mitigates potential legal liabilities associated with construction and safety failures. Therefore, prioritizing the safety and quality of bridge piers is fundamental to the design and ongoing maintenance of bridges, underscoring their essential role in infrastructure integrity [2].

Surface defect inspection of piers is vital for effective bridge maintenance and safety management [3, 4]. This process facilitates the early detection of critical issues such as cracks and erosion, which pose significant threats to the structural integrity of the bridge. Surface defects are often symptomatic of deeper, more severe internal issues within the bridge's structure. By actively monitoring these defects, the safety of both vehicles and pedestrians using the bridge is significantly enhanced. The timely detection and subsequent repair of these surface defects are critical in preventing serious structural damage. This proactive approach not only extends the bridge's service life but also minimizes the need for extensive and expensive future repairs [5]. Detecting surface defects on bridge abutments is a fundamental aspect of maintenance, essential for ensuring the long-term stability and safety of bridges [6]. This practice is critical in maintaining the structural health of bridges and preventing potential failures.

Czimmermann, et al. [7] extensively discussed a range of vision-based methodologies for detecting and classifying structural defects, delving into advanced artificial vision processing techniques. They highlighted the inherent challenges and limitations associated with traditional visual inspections, which are often subjective and prone to inconsistency. The study presents various deep learning inspection techniques designed to overcome these issues by employing logical and mathematical approaches to image analysis. These methods enhance both the accuracy and reliability of defect detection, showcasing significant advancements in automated inspection technologies.

Tulbure, et al. [8] provided a comprehensive review of recent advancements in defect detection models that utilize Deep Convolutional Neural Networks (DCNN). They detailed the application of these models in object recognition and highlighted their significant advantages over traditional computer vision techniques, specifically in terms of enhanced accuracy and processing speed. The paper extensively discusses several popular object detection frameworks, including region-based Convolutional Neural Networks (CNNs), You Only Look Once (YOLO), and Single Shot MultiBox Detector (SSD). It further explores how these advanced models can be effectively adapted for specialized defect detection tasks, underscoring their versatility and potential in industrial applications.

Hussain [9] provided a detailed overview of the evolution of the YOLO (You Only Look Once) algorithms, emphasizing their transformative role in industrial applications. Since its inception in 2015, continuous enhancements to the architecture of YOLO algorithms have dramatically improved both the speed and accuracy of target detection. These advancements have rendered YOLO particularly suitable for deployment on resource-constrained edge devices, where efficiency is crucial. The article underscores the advantages of utilizing YOLO algorithms for industrial surface defect detection, noting their high efficiency, lightweight framework, and real-time detection capabilities. These features make YOLO an excellent choice for automated quality inspection systems, significantly enhancing both productivity and product quality in manufacturing processes.

1.2 Research problem statement

While the YOLO series has marked significant achievements in target detection, it currently faces a range of challenges that limit its effectiveness in specific scenarios.

Santoso, et al. [10] emphasized that poor image quality and low visibility of defects are major impediments to the performance of target detection models, particularly affecting the accuracy and reliability of detections.

Qiu, et al. [11] mentioned in the article that the targets being detected in the field of defect detection are usually small, which makes the traditional YOLO model miss a portion of the targets when the samples are detected.

In the context of bridge pier surface defect detection, the similarity of defects in color to the background, combined with their small size, significantly hinders the accuracy of defect detection, making it challenging to distinguish defects from their surroundings.

1.3 Aim

The main aim of this report is to propose a method for detecting surface defects on bridge piers that combines an image enhancement method and an improved YOLO model.

1.4 Objectives

The objectives are to:

- Evaluate various image enhancement techniques to identify those that most effectively aid the model in defect detection.
- Modify the original YOLO model's structure to better suit this research.
- Incorporate various modules into the benchmark model to facilitate comparative experiments.
- Analyze and evaluate the final results.

1.5 Structure of the project

The Section 2 of the article comprehensively reviews existing literature on image enhancement, target detection, and defect detection techniques, establishing a theoretical foundation relevant to the focus of this study. Section 3 outlines the research methodology employed in this study, detailing the research process, data collection strategies, and the experimental equipment used, providing a clear framework for the experiments conducted. Section 4 details the experiments conducted to validate the research methodology, describing the specific tests and their configurations. Section 5 then reviews and evaluates both the methodology and the experimental results, ensuring a comprehensive analysis of the study's findings. In addition to reviewing the methodology and results, Section 5 discusses and summarizes key findings, highlights the limitations of the study, and outlines potential directions for future research, making it a crucial component of the study’s overall analysis. Section 6 provides a conclusive overview of the entire study, summarizing the research question, methodology, and experimental demonstrations, and drawing final conclusions from the study's findings.

CHAPTER 2 - RESEARCH & REVIEW OF LITERATURE

This section provides an overview of popular image enhancement methods, target detection methods and defect detection techniques.

2.1 Image enhancement

Image enhancement techniques play a pivotal role in defect detection, significantly boosting both the accuracy and efficiency of inspection systems. These techniques are essential for optimizing the detection process in various industrial and technological contexts. By enhancing the contrast, sharpness, and detail of images, these techniques facilitate a clearer presentation of target object features. This clarity is crucial for detecting small or subtle defects that might otherwise be overlooked, thereby improving the overall reliability of the inspection process [12].

Qi, et al. [13] outlined a comprehensive range of image enhancement techniques, including adaptive methods, multi-scale approaches, detailed edge enhancement, substantial noise reduction, deblurring techniques, and deep learning-based methods. These techniques are versatile, designed to improve image quality across diverse scenarios and applications. In this study, the similarity in color between the defects and the background complicates the target detection model's ability to recognize defects. To address this, it is crucial to employ an image enhancement technique that not only boosts the contrast between the defects and the background but also accentuates the defects' features, thereby facilitating more accurate detection.

Kaur, et al. [14] explained that histogram equalization improves overall image contrast by modifying the histogram to distribute brightness more uniformly across the image. This technique specifically enhances the distinction between darker and lighter areas by broadening the range of more commonly occurring brightness levels, thereby making subtle variations in the image more discernible.

Gupta, et al. [15] provided a comprehensive examination of histogram equalization, beginning with the concept of an image histogram. This histogram quantifies the number of pixels at each of the 256 possible gray levels in an image, depicted in the form of a bar graph, offering a visual representation of pixel distribution across these levels. In histogram equalization, the horizontal axis of the histogram graphically represents the gray levels, while the vertical axis quantifies the pixel count at each level. This technique systematically redistributes these pixel values to achieve a more uniform gray level distribution across the entire image. The result is a significant enhancement in image contrast and a more pronounced presentation of details, making the image visually more distinct and easier to analyze. The procedure for histogram equalization involves several detailed steps: first, calculating a histogram of the original image to understand its gray level distribution; second, deriving a cumulative distribution function from this histogram to assess cumulative pixel density; third, creating a new gray level mapping table based on this cumulative function; and finally, applying this new mapping to the original image. This sequence of actions effectively enhances the image's contrast, bringing subtle details into sharper relief. According to the authors, histogram equalization markedly enhances image contrast and clarity, particularly effective in images with limited gray level variation. Its computational simplicity and straightforward implementation make it well-suited for diverse image processing systems and extensive tasks. Moreover, it can be adapted for use with color images, further broadening its applicability across various image enhancement and preprocessing applications.

Kaur and Singh [16] highlighted the drawbacks of histogram equalization despite its significant benefits in enhancing image contrast and detail presentation. Particularly, in images characterized by uneven grey level distribution or high levels of noise, histogram equalization can inadvertently amplify this noise. This amplification can degrade image quality, detracting from the clarity and sharpness the technique is intended to enhance. Additionally, when applied to color images, treating each color channel with histogram equalization independently can cause distortions in color balance, leading to unnatural color representations. To mitigate these issues, additional processing steps such as histogram matching or conversion to alternative color spaces are often necessary. Furthermore, Kaur and Singh [16] observed that histogram equalization can sometimes result in an overly contrasted image. This over-enhancement can make certain details appear too bright or too dark, leading to a visual output that lacks balance and may obscure critical information rather than revealing it.

Pizer, et al. [17] discussed how Adaptive Histogram Equalization (AHE) specifically enhances the contrast in localized areas of an image, offering a distinct advantage over standard histogram equalization. This targeted approach makes AHE particularly effective for images affected by local shadows and uneven illumination, where it can significantly improve visibility and detail clarity in these challenging conditions. AHE operates by dividing the image into numerous small cell blocks and applying histogram equalization independently to each block. This method allows for precise contrast adjustment tailored to the specific lighting and detail of each segment, thereby enhancing the overall image quality by addressing local contrast variations more effectively.

Lidong, et al. [18] provided an in-depth exploration of Adaptive Histogram Equalization (AHE), a technique that notably enhances both contrast and detail through localized processing. This method specifically targets smaller, distinct areas of an image to tailor contrast enhancements directly to the varying needs of different image sections. Initially, the image is segmented into multiple small local regions, or sub-blocks. For each of these sub-blocks, a unique histogram is calculated independently. Subsequent local histogram equalization on each histogram focuses on optimizing contrast and detail specifically for that region, adapting to the local content's unique characteristics. Once the local regions have been individually equalized, they are reassembled into a single, cohesive image. To ensure a seamless integration, interpolation methods are employed to smooth the transitions between these sub-blocks, effectively preventing any apparent boundaries or visual discontinuities that might detract from the overall image quality.

Thakur and Singh [19] provided a thorough analysis of the advantages and limitations of Adaptive Histogram Equalization (AHE), detailing how it impacts image processing in various conditions. The primary advantage of AHE is its proficiency in significantly enhancing details within low-contrast regions of an image. This capability makes local features more pronounced, proving particularly beneficial in conditions where images are marred by uneven lighting or possess intricate backgrounds. Additionally, AHE exhibits high adaptability, allowing it to be finely adjusted to meet the distinct contrast needs of different areas within an image. This customization enhances its utility across varied imaging requirements. However, AHE also presents certain drawbacks; notably, it can amplify existing noise within an image. This amplification tends to occur particularly in areas already afflicted with noise, becoming more pronounced as contrast is enhanced, which can detract from the overall image quality. Additionally, the computational demands of AHE are considerable, owing to its processing of multiple local regions individually. This complexity can lead to extended processing times, which may become particularly cumbersome in large-scale image processing operations.

Zuiderveld [20] described Contrast Limited Adaptive Histogram Equalization (CLAHE) as a sophisticated improvement upon traditional AHE, specifically engineered to curb the over-amplification of noise by imposing limits on contrast adjustments during the equalization process. CLAHE provides significant advancements over AHE by effectively mitigating issues related to over-enhancement and noise amplification. This is achieved through the strategic incorporation of contrast constraints, which carefully regulate the enhancement process to preserve image quality. CLAHE controls the contrast within individual local regions by trimming the cumulative distribution function of local histograms. This method ensures that the natural appearance of the image is maintained, avoiding the artificial effects often introduced by standard histogram equalization techniques. Furthermore, CLAHE significantly enhances image clarity, especially in environments with substantial noise, by optimizing block sizes to suit varying regional contrast demands. Additionally, it employs advanced interpolation techniques to ensure seamless transitions and minimize any visual discontinuities between sub-blocks. These technical enhancements enable CLAHE to balance local detail with overall image consistency more effectively than traditional methods. The result is a more stable and naturally enhanced image that faithfully represents the original scene.

Musa, et al. [21] provided a thorough analysis of the advantages and disadvantages of Contrast-Constrained Adaptive Histogram Equalization (CLAHE), offering insights into both its efficacy and the challenges associated with its use. The primary advantages of CLAHE include its effective control over image contrast enhancement and its ability to significantly reduce noise amplification. This is achieved by limiting contrast in the cumulative distribution function of local histograms, which results in clearer and more visually natural images, even in challenging lighting conditions. CLAHE excels in maintaining high image quality against complex or noisy backgrounds, facilitating smooth transitions across image areas through carefully optimized block sizes and advanced interpolation techniques. These methods help prevent abrupt changes that can disrupt the visual flow of the image. However, CLAHE presents certain drawbacks, notably its high computational complexity. This complexity stems from the need to process multiple local regions individually and to apply contrast limitations within these areas, which can significantly extend processing times, especially in larger or more detailed images. Although CLAHE surpasses traditional AHE in terms of noise suppression, its approach to local contrast adjustment can sometimes result in a loss of detail or a degradation in overall image quality, particularly when applied to areas with subtle features. Consequently, while CLAHE is effective across a wide range of applications, the optimal performance of this technique requires meticulous adjustment of its parameters to suit specific scene requirements, ensuring the best possible outcome in each unique situation.

Farid [22] elucidated how gamma correction finely adjusts the brightness of an image by applying a nonlinear gamma function to each pixel’s value. This method specifically targets and modifies the luminance levels to suit visual perception more accurately. This technique effectively enhances details in both dimly lit and overly bright areas by non-linearly amplifying or dampening their brightness levels, respectively. Such adjustments result in a significant improvement in the image's overall contrast, making it easier to discern finer details and subtleties.

Rahman, et al. [23] provided an in-depth explanation of gamma correction, a technique designed to enhance visual effects by fine-tuning an image's brightness and contrast. This adjustment is specifically aimed at aligning with the perceptual properties of the human eye, making images appear more natural and visually pleasing. The process involves a non-linear transformation of pixel values, which is controlled by a parameter known as the gamma value. This transformation adjusts the intensity of the pixels based on their original levels, which affects the image's overall luminance. When the gamma value exceeds 1, it results in a darkening of the image, as it compresses the higher grey levels, thereby enhancing both contrast and the visibility of details. On the other hand, a gamma value below 1 causes the image to brighten by expanding the lower grey levels, which improves not only the overall brightness but also makes details more visible. Thus, gamma correction strategically adjusts the range of brightness across an image to optimize the rendering of details and improve visual effects, making it a crucial tool in digital image processing. However, it's important to note that gamma correction primarily modifies brightness and contrast without directly enhancing color saturation or other intrinsic image details. Therefore, it is frequently used in conjunction with other image processing techniques to achieve a more comprehensive optimization of images.

Amiri and Hassanpour [24] comprehensively analyzed gamma correction, highlighting its significant advantages and notable disadvantages in image processing. The primary advantage of gamma correction is its ability to substantially improve the visual effects of an image by adjusting brightness and contrast, which significantly enhances the visibility of details, making it especially useful in images where detail clarity is essential. By adjusting the gamma value, contrast can be effectively enhanced, particularly beneficial in low-contrast or underexposed images, where it improves both overall brightness and the clarity of details, thus transforming the visual presentation. Furthermore, gamma correction is adaptable to the characteristics of different display devices, ensuring that images are consistently presented across various types of screens, which is crucial for applications where uniform visual experience is necessary. However, gamma correction primarily focuses on adjusting brightness and contrast without directly improving color saturation or the fineness of details. This can limit its effectiveness in applications where enhanced color performance and intricate detail are critical. Moreover, excessive application of gamma correction may lead to distortions in brightness or detail, particularly if the gamma value is not appropriately calibrated, potentially compromising image quality. Therefore, gamma correction is typically employed in conjunction with other image processing techniques to achieve a more balanced and comprehensive optimization of images, ensuring that all visual aspects are adequately enhanced.

Each contrast-enhancing image enhancement method offers unique advantages and poses specific limitations. Therefore, the choice of the most suitable method for a given study should be informed by detailed experimental results that evaluate the effectiveness of these methods under various conditions. This data-driven approach ensures the selected method aligns optimally with the specific requirements and challenges of the research.

2.2 Object detection

Object Detection (OD) represents a critical component of computer vision, primarily focused on both identifying and precisely locating objects within images or videos. This process involves more than merely recognizing the objects present; it also entails pinpointing their exact positions within the image, providing both qualitative and quantitative data about the scene [25].

Bai, et al. [26] provided a summary of traditional machine learning-based object detection methods, which relied on hand-designed features and classic algorithms. These methods were the standard before the advent of deep learning technologies, and they involved manually crafted features tailored to specific applications. Despite being less effective in complex scenarios and multi-category detection, these traditional methods still hold value in resource-constrained environments due to their lower computational demands. They extract crucial image features and employ classifiers to detect objects, providing a viable solution where advanced computing resources are limited [27].

Zhao, et al. [28] discussed the significant advancements brought by deep learning-based object detection methods, particularly those utilizing Convolutional Neural Networks (CNNs). These methods have notably improved both the accuracy and speed of object detection, marking a substantial evolution in the field. These deep learning approaches are categorized into two primary types: region proposal methods, which first suggest potential object locations before classifying them, and single-stage methods, which simultaneously detect and classify objects in one pass, streamlining the process.

A foundational approach within the realm of region proposal methods in object detection is represented by the R-CNN family [25]. Girshick, et al. [29] pioneered the development of Regions with Convolutional Neural Network features (R-CNN), an innovative method that synergizes region proposals with CNNs, substantially elevating object detection accuracy. The R-CNN operates through a structured three-step process: First, it generates approximately 2000 candidate regions from the input image via Selective Search. Second, it extracts features from each region using pre-trained CNNs like AlexNet. Finally, it classifies these regions using Support Vector Machines (SVMs) and refines them using bounding box regression, each step building towards accurate object identification. The primary advantage of R-CNN lies in its utilization of deep learning techniques, which markedly improves detection accuracy by effectively learning from vast amounts of visual data. However, R-CNN is hampered by low computational efficiency and a complex training process. Each candidate region is processed independently for feature extraction by a CNN, leading to redundant computations and substantial consumption of computational resources, which can be impractical in real-time applications.

Bharati and Pramanik [30] highlighted that R-CNN significantly boosts object detection accuracy by utilizing Convolutional Neural Networks (CNNs) for feature extraction, achieving remarkable performance improvements across a variety of benchmark datasets. This approach leverages the deep learning capabilities of CNNs to discern and classify intricate features within images. The modular design of R-CNN, which includes separate modules for candidate region generation, feature extraction, classification, and bounding box regression, allows for the independent optimization of each component. This modular approach not only facilitates targeted improvements but also exemplifies the effectiveness of deep learning in refining various aspects of object detection tasks. However, R-CNN faces significant challenges, including low computational efficiency, a complex training process, and substantial memory consumption. These issues stem from the intensive computational demands of processing multiple candidate regions through deep neural networks. The requirement for each candidate region to undergo individual feature extraction by a CNN results in extensive redundant computations, significantly slowing down processing speeds. Additionally, the multi-stage training process complicates the implementation of R-CNN, and its reliance on selective search to generate candidate regions introduces delays that can adversely affect overall detection performance. This reliance on a time-consuming search process can hinder the model's applicability in dynamic environments.

Girshick [31] introduced Fast R-CNN, marking a substantial advancement in the field of object detection. This innovation significantly improved the efficiency and speed of detecting objects while maintaining high accuracy. Fast R-CNN enhances detection speed and operational efficiency by utilizing shared computation across multiple detection tasks, thereby optimizing the overall process and maintaining high accuracy. Fast R-CNN processes the entire input image through a single convolution operation to create a shared feature map, which effectively eliminates the need for repetitive computation across each candidate region. This approach significantly reduces redundancy and increases processing speed. Following the initial convolution, candidate regions are generated either through selective search or a Region Proposal Network (RPN), and then mapped onto fixed-size feature regions using an ROI pooling layer, which standardizes the input size for subsequent classification tasks. Classification and bounding box regression tasks are carried out through fully connected layers and a softmax layer, respectively. A multi-task loss function is employed to optimize these tasks simultaneously, enhancing the model's accuracy and reliability. Although Fast R-CNN requires substantial memory, particularly with high-resolution images and numerous candidate regions, it achieves outstanding accuracies across various benchmarks. This model not only advanced the state of object detection technology but also laid the foundational work for subsequent innovations such as Faster R-CNN and Mask R-CNN.

Ren, et al. [32] introduced Faster R-CNN, a significant evolution in object detection that integrates a Region Proposal Network (RPN) with Fast R-CNN. This integration allows for end-to-end training of both candidate region generation and detection processes, significantly enhancing the system's speed and accuracy. Faster R-CNN starts by extracting features from the entire image using a convolutional neural network (CNN). It then employs an RPN to generate candidate regions directly on this feature map, optimizing the detection workflow by leveraging the extracted features efficiently. The integration with Fast R-CNN's shared convolutional features significantly minimizes redundant computations across the detection process, thereby boosting overall efficiency and streamlining the computational workload. Once resized by the ROI pooling layer to ensure uniformity, candidate regions are processed through fully connected layers for classification and bounding box regression, further refining the detection accuracy. The primary innovation of Faster R-CNN is its unified framework that combines region proposal and object detection into a single streamlined process. This simplification not only reduces computational complexity but also enhances processing efficiency across various detection tasks. Despite its considerable resource demands, Faster R-CNN delivers superior performance across multiple benchmarks. Its extensive utilization in a variety of object detection applications attests to its effectiveness and versatility, setting the stage for future advancements in the field.

Zhao, et al. [28] provided a comprehensive analysis of the significant advantages and notable drawbacks of region proposal methods in object detection, offering a balanced perspective on their efficacy and limitations. Among the advantages, region proposal methods are highly accurate, enhancing target detection by generating high-quality candidate regions. These regions facilitate more effective classification and bounding box regression, crucial for precise object localization. Additionally, region proposal methods are highly adaptable and versatile, suitable for a wide range of image types and detection tasks. This flexibility makes them applicable across diverse applications and scenarios. Both traditional methods and modern deep learning approaches are capable of generating fine-grained candidate regions. This capability is particularly beneficial for detecting complex and fine targets, where detail is paramount. However, the notable drawbacks of region proposal methods include their high computational complexity. This complexity can be a significant barrier in environments where computational resources are limited. Traditional methods such as selective search are particularly slow, rendering them unsuitable for real-time applications where quick processing is essential. The performance of region proposal methods heavily depends on the quality of the candidate regions. Poor quality regions can significantly impact the accuracy of the final detection results, underscoring the importance of precision in the initial stages of detection. Deep learning-based region proposal methods demand significant memory and robust hardware capabilities, particularly when processing high-resolution images and a large number of candidate regions. These requirements can limit their deployment in less capable systems. Additionally, the complex training processes associated with these methods require extensive fine-tuning and a substantial amount of labeled data. This not only increases the difficulty of model development but also escalates the costs associated with it.

The You Only Look Once (YOLO) series stands as a prime example of the single-stage object detection methodology, streamlining the process from image input to detection output [25]. Redmon, et al. [33] introduced YOLO as an innovative single-stage object detection method that accomplishes end-to-end processing by conceptualizing object detection as a regression problem. This approach allows for direct prediction of bounding boxes and class probabilities from full images in a single evaluation. Although YOLO has certain limitations in accurately detecting dense and small targets, its capabilities for real-time and global inference render it highly effective across a wide range of real-world applications. These strengths are particularly valuable in scenarios requiring fast and efficient detection. The YOLO series has undergone continual improvements, achieving significant advances in both detection speed and accuracy. Each iteration of YOLO has introduced enhancements that have incrementally optimized its performance and broadened its applicability [9].

Table.1. Model Performance Comparison

Model	mAP50(%)
R-CNN [29]	58.5
Fast R-CNN [31]	70
Faster R-CNN [32]	73.2
YOLOV3 [34]	79.5
YOLOV5-S [34]	78
YOLOV7 [34]	69.1
YOLOV8-S [34]	83.1

According to Table 1, which showcases various models' performance on the PASCAL VOC 2007 dataset, the YOLO series has significantly outperformed regionally proposed method-based models. This demonstrates the YOLO series' superior efficacy in object detection, marking it as a standout performer in the field [35].

Terven, et al. [36] observed that YOLO has become a pivotal technology in real-time object detection, attributed to its rapid processing speeds and efficient end-to-end training capabilities. These features make YOLO particularly effective for applications requiring immediate detection responses. However, YOLO has shown limitations in accuracy, particularly exhibiting high false detection rates when dealing with small, dense targets set against complex backgrounds. This challenge highlights areas where YOLO's detection algorithm may require further refinement. Subsequent iterations, including YOLOv2, YOLOv3, and YOLOv4, have addressed these accuracy issues to some extent, reducing false detection rates and enhancing overall performance. However, a trade-off between speed and accuracy still persists in some scenarios, indicating ongoing challenges in balancing these critical aspects.

Yung, et al. [37] provided a detailed analysis of the evolutionary progress in the YOLO series, specifically through the advancements presented in YOLOv5, YOLOv6, and

YOLOv7. Each version represents significant technological strides in object detection, showcasing continual improvements in processing efficiency and detection capabilities. YOLOv5 utilizes the CSPNet architecture, which strategically divides and then merges the feature map. This approach significantly reduces computational demands while maintaining a rich representation of features, optimizing both the efficiency and effectiveness of the detection process. YOLOv5 introduces Mosaic data augmentation, a technique that stitches together four images to enhance the model's adaptability. This method improves the system's ability to handle various object sizes and backgrounds, thus broadening its application in diverse scenarios. Additionally, YOLOv5 features dynamic adjustment of anchor frames, which allows it to better accommodate a range of target shapes and sizes, enhancing the model's versatility in detecting diverse objects. YOLOv6 advances further by enhancing its convolutional layers and optimizing sampling strategies, which collectively boost both the speed and accuracy of the detection process. The architecture of YOLOv6 is refined using Network Architecture Search (NAS), a method that identifies optimal configurations, ensuring that the model achieves the best possible performance. Furthermore, YOLOv6 improves the Non-Maximum Suppression (NMS) algorithm, effectively reducing the processing time required after detection, thus streamlining the overall object detection process [38]. YOLOv7 employs a deeper network architecture, enhancing its ability to learn from complex scenes. It also introduces the EvoNorm layer, which contributes to better training stability and expressiveness, facilitating more effective learning and adaptation. Additionally, YOLOv7 utilizes advanced training techniques such as dynamic image resizing, which speeds up model convergence and enhances efficiency, making it more effective in practical applications [39].

In their comprehensive analysis, Wang, et al. [40] explored the capabilities and enhancements introduced in YOLOv8, marking a significant evolution in the YOLO series for real-time object detection. YOLOv8 advances as a highly sophisticated real-time object detection system that not only retains the renowned high speed and user-friendliness of the YOLO series but also significantly improves upon model accuracy and generalization capabilities. These improvements in YOLOv8 stem from the integration of advanced technologies such as deep separable convolution, adaptive feature fusion, and sophisticated data enhancement techniques, each contributing to a more robust and adaptable detection system.

Wang, et al. [41] unveiled YOLOv9, a groundbreaking advancement in object detection technology that significantly boosts detection accuracy and supports real-time, high-precision detection capabilities. YOLOv9 employs a novel architecture, the Generalised Efficient Layer Aggregation Network (GELAN), which is optimized through sophisticated gradient path planning techniques. This architectural innovation allows for enhanced performance through better utilization of computational resources. This innovative approach in YOLOv9 surpasses existing methods by optimizing parameter utilization efficiency using conventional convolution operators, thereby significantly enhancing both the performance and efficiency of the model. As a result, YOLOv9 achieves unparalleled accuracy and operational speed, all while maintaining a lightweight structure, setting a new standard in the field of object detection.

Wang, et al. [42] unveiled YOLOv10, a breakthrough in object detection that achieves top-tier performance while significantly reducing computational overhead. This is accomplished by dispensing with traditional non-maximal suppression (NMS) and through strategic refinements in the model's architecture. YOLOv10 employs a consistent dual allocation strategy that eliminates the need for NMS, thereby lowering inference latency and streamlining the detection process. This strategy optimizes the allocation of computational resources, enhancing real-time processing capabilities. YOLOv10 incorporates several component enhancements to boost both inference efficiency and accuracy. These include lightweight classification headers that reduce computational load, spatial channel decoupling for more efficient downsampling, and a hierarchical bootstrap block design that improves learning effectiveness. Furthermore, YOLOv10 integrates large kernel convolutions and partial self-attention modules, which elevate performance by enhancing feature extraction and attention mechanisms without significantly increasing computational demands. This integration allows for deeper, more nuanced learning of image features.

While the YOLO family is celebrated for its real-time object detection capabilities, it also exhibits a number of limitations that affect its performance across various scenarios. Specific limitations of the YOLO series include inadequate detection of small objects, comparatively lower accuracy than region-based methods, and a marked sensitivity to variations in training data, which can compromise the model's reliability under varying conditions. Furthermore, YOLO often struggles with precise bounding box localization, especially in complex scenes with ambiguous boundaries or intricate backgrounds, where precise object delineation is critical. The design philosophy of YOLO often prioritizes speed over generalization capabilities and positional accuracy. This trade-off can lead to challenges in dynamic environments where adaptability and accuracy are essential [43].

Although each successive generation of YOLO has improved upon its predecessors, these versions continue to exhibit significant shortcomings in handling complex object detection challenges, indicating room for substantial enhancements. These persistent issues underscore the need for further modifications to the YOLO model to enhance its effectiveness and suitability for advanced research applications, aiming to overcome its current limitations.

2.3 Defect detection

Bridge pier surface defect detection is a critical aspect of structural health monitoring (SHM), playing a pivotal role in safeguarding the structural integrity and extending the longevity of bridges. The timely identification and remediation of defects on abutment surfaces are essential for effective bridge maintenance and evaluation. These practices are crucial not only for preventing potential structural failures but also for extending the operational lifespan of bridges, thereby ensuring ongoing safety and functionality. These defects, including cracks, depressions, and other forms of physical damage, typically arise from natural deterioration, environmental influences, or mechanical stress. Each type of damage has its own implications for the structural health of the bridge, making early detection and appropriate remediation critical.

Chen, et al. [44] provided a comprehensive classification of defect detection methods, distinguishing between traditional approaches and modern deep learning techniques, each characterized by distinct methodologies and applications. Traditional defect detection methods encompass manual labeling and various machine learning techniques. Each method offers distinct advantages but also presents unique challenges, making them suitable for specific scenarios based on the requirements of accuracy, efficiency, and complexity.

Xie [45] elaborated on the manual annotation method, a traditional technique in defect detection that heavily relies on the expertise and judgment of skilled professionals. This method remains critical in scenarios where detailed, nuanced understanding of defects is necessary. Initially, the process involves selecting suitable labeling tools and defining the types of defects and the criteria for identification, which sets the groundwork for consistent and accurate data collection. During the inspection phase, trained professionals meticulously identify and mark potential defects, carefully noting each defect's type, size, and location, which are crucial for subsequent analyses and remediation strategies. To ensure accuracy and consistency, the process includes internal audits and feedback mechanisms that help standardize defect identification and reduce subjective discrepancies among different annotators. Once collected, the data is compiled and systematically stored, enabling detailed analysis and facilitating future reference and comparison. Although manual labeling excels in accuracy and adaptability, especially in complex scenarios, it becomes notably inefficient and costly when dealing with large datasets due to its heavy reliance on skilled labor. Additionally, variability in annotator judgment can lead to inconsistencies and challenges in data reproducibility. Prolonged engagement in repetitive tasks may also cause operator fatigue, further compromising the accuracy of the data. Despite these challenges, the evolution of technology is making automated and machine learning methods increasingly viable alternatives, particularly for applications requiring large-scale and rapid processing.

Shahrabadi, et al. [46] detailed a machine learning-based approach to defect detection, specifically designed to enhance both the efficiency and accuracy of the detection process through automation. The process initiates with data preparation, involving the collection of extensive labeled data that includes both defective and non-defective samples. This comprehensive dataset provides a robust foundation for effective training of the machine learning models. Following data preparation, feature engineering takes place, which includes preprocessing and extracting key features from the data, such as image greyscaling and edge detection. These features are critical for enhancing the model’s ability to recognize and classify defects accurately. The selection of an appropriate machine learning model is the next step. Options include support vector machines, random forests, and deep learning models such as convolutional neural networks, each offering unique advantages depending on the complexity and type of defects. These models are employed for their capability to facilitate effective pattern recognition and learning, crucial for accurately detecting and classifying various types of defects. Model training involves the optimization of model parameters using well-labeled datasets, tailored to meet specific defect detection needs and ensure the model performs optimally under various conditions. Validation and testing are conducted with an independent dataset to ensure the model's generalization capability and maintain high accuracy across different scenarios, crucial for reliable deployment. The final stage of the process involves deploying the trained model into production lines or inspection systems for real-time detection. This stage also includes ongoing optimization and adjustments to accommodate new types of defects or changes in the production environment, ensuring the model remains effective over time..

Saufi, et al. [47] conducted a comprehensive analysis of the advantages and disadvantages associated with the application of machine learning in defect detection, providing insights into both its beneficial impacts and the challenges it presents. A primary advantage of machine learning in defect detection is its efficiency. This technology enables the processing of large-scale datasets, which significantly enhances the speed and accuracy of inspections along production lines, streamlining quality control processes. Additionally, machine learning ensures consistent, standardized, and repeatable results in defect detection, significantly reducing the potential for human error and enhancing the reliability of inspections. Furthermore, machine learning offers cost-effectiveness by substantially lowering long-term operational expenses, presenting a financially viable alternative to manual inspections once it is fully implemented and integrated into systems. However, challenges persist with machine learning in defect detection, particularly the need for extensive labeled data for training. This requirement can be resource-intensive and costly initially, posing a significant barrier to implementation. Additionally, while machine learning models generally excel with the training data, they often struggle with generalization when encountering new or unseen types of defects, which can limit their effectiveness in dynamic environments. Lastly, the inherent complexity of these models, especially those based on deep learning, often results in a lack of interpretability. This makes it challenging for users to understand and trust the decision-making process, which is critical for acceptance and effective use in practical settings.

Deep learning-based defect detection has become the predominant method in the field, leveraging extensive labeled datasets and sophisticated neural network models to achieve highly precise and efficient automatic detection. This technological advancement enables superior accuracy and speed in identifying defects, transforming the landscape of industrial quality control [48].

Among deep learning models, YOLO stands out as a leading solution specifically tailored for defect detection. Its real-time processing capabilities and high accuracy make it a preferred choice for industrial applications. YOLO offers real-time detection capabilities, enabling high-precision identification and classification of a wide range of surface defects. This makes it exceptionally useful in scenarios where quick decision-making and accuracy are critical. YOLO enhances defect detection across various sizes and shapes through its implementation of multi-scale feature fusion and attention mechanisms. These technologies allow YOLO to maintain robust performance even in complex environments and under variable lighting conditions, adapting effectively to diverse operational challenges. YOLO's streamlined and lightweight architecture facilitates its easy integration into industrial equipment and embedded systems, significantly enhancing detection accuracy and operational efficiency. This integration is pivotal in advancing smart manufacturing and industrial automation, driving innovations in production processes [49, 50].

In recent years, the YOLO series of models has gained significant prominence in the field of industrial defect detection, becoming a preferred choice for their efficiency and accuracy. Li, et al. [51] developed an innovative real-time method for detecting surface defects on steel strips by enhancing the YOLO network with advanced data enhancement and preprocessing techniques. This approach significantly improves the detection process, making it more efficient and accurate. Experimental results validated the superior accuracy and real-time capabilities of this enhanced YOLO model, confirming its broad applicability across various industrial sectors where quick and accurate defect detection is crucial.

Xu, et al. [52] specifically enhanced the YOLO algorithm to improve the detection of defects on metal surfaces, integrating a feature enhancement module and an attention mechanism to refine detection accuracy. Their version of the YOLO algorithm significantly improves defect detection across different scales and effectively minimizes background noise through the innovative use of a feature enhancement module and an attention mechanism. These enhancements help in distinguishing defects more clearly against noisy backgrounds. Experimental validations of their enhancements showed notable improvements in both accuracy and speed, enabling more efficient identification and classification of metal surface defects. This makes the algorithm particularly effective for real-time industrial applications where timely and accurate defect detection is essential.

Hatab, et al. [53] developed a YOLO-based method for surface defect detection that specifically optimizes the model's structure and parameters, enhancing its ability to efficiently detect and classify a wide range of surface defects. This tailored approach improves the model's applicability and effectiveness in diverse industrial settings. Experimental results confirm significant improvements in detection accuracy and speed, demonstrating that this optimized YOLO model meets the rigorous demands of industrial real-time defect detection and contributes to enhanced production quality control.

These studies not only demonstrate the effectiveness and potential of YOLO models in industrial defect detection but also highlight certain limitations of the original model, including its inadequate detail capture and performance degradation in complex environments. Nonetheless, these shortcomings present opportunities for further research and can be addressed by modifying the YOLO model to better suit specific industrial and research needs, thereby extending its utility and enhancing its performance.

CHAPTER 3 - RESEARCH & REVIEW OF LITERATURE

This section describes the research methodology used in this study, the overall process of the defect detection model, the image enhancement methodology, and the new modules proposed in this work.

3.1 Research method

This study adopts a mixed-methods approach that combines the image enhancement method and object detection method, which are very popular in the current defect detection field [41, 54-57]. The image enhancement method can make the defects on the samples more prominent and obvious, which makes it easier to capture the target detection model; while the object detection method can accurately and quickly capture where the defects are on the surface of the bridge piers, and mark them[9, 58].

An inductive approach based on a fully supervised learning paradigm is used in this study. The detection algorithm relies on many post-labelling image data for training. The algorithm learns the relationship between defect type and performance from specific instances of bridge piers surface defect image samples to derive a method for detecting bridge piers surface defects [10, 59].

The research strategy is experimental. The detection method of this research is completed by two steps of image enhancement and target detection, in order to ensure the accuracy and robustness of the algorithm, it is firstly necessary to experimentally test the validity and superiority of the methods used in each step, in addition to this, it is also necessary to experimentally suggest whether the methods used in the two steps are coordinated, and whether they are paired together to produce better results.

3.2 Research model

This section will present the overall process of bridge abutment surface defect detection including image enhancement methods and new modules proposed for this research.

3.2.1 Overview

In this study, the complete framework of YOLOV9-S is retained and the original RepNCSPELAN4 module is replaced with the newly proposed GhostBottleNeck(GBN) module.

The Fig.1 presents a comprehensive flowchart of a deep learning model designed for detecting surface defects on bridge piers, detailing the process from input to output.

Initially, the surface image of the bridge abutment is captured and input into the target detection model following enhancement by the Gamma correction method [8, 36].

In the Backbone section of the target detection model, the image undergoes initial feature extraction through two convolutional (Conv) layers. Subsequently, features from multiple layers are aggregated via the ELAN layer to enhance their expressiveness, then passed through the AConv module, incorporating average pooling and convolution operations, and finally into the GhostBottleNeck (GBN) module, designed specifically for detecting small targets. This sequence aims to extract valuable information across various scales. Furthermore, the SPPELAN module combines multiple convolution and pooling operations to further aggregate and condense these features, preparing them for advanced processing [41, 57, 60].

In the Neck section, the model integrates multi-scale features from the Backbone using concatenation and up-sampling techniques to create a detailed feature map that combines both deep and shallow features, enhancing the accuracy of the final detection. Repeated applications of the AConv and GBN modules in this section further refine the features, improving defect identification [41].

In the final Head section, the accumulated and optimized features are applied to detect defects. These features are further refined through two convolutional layers and then fused via a concatenation operation to enable effective object detection. The resulting image displays the detected defect within a rectangular box [12, 41].

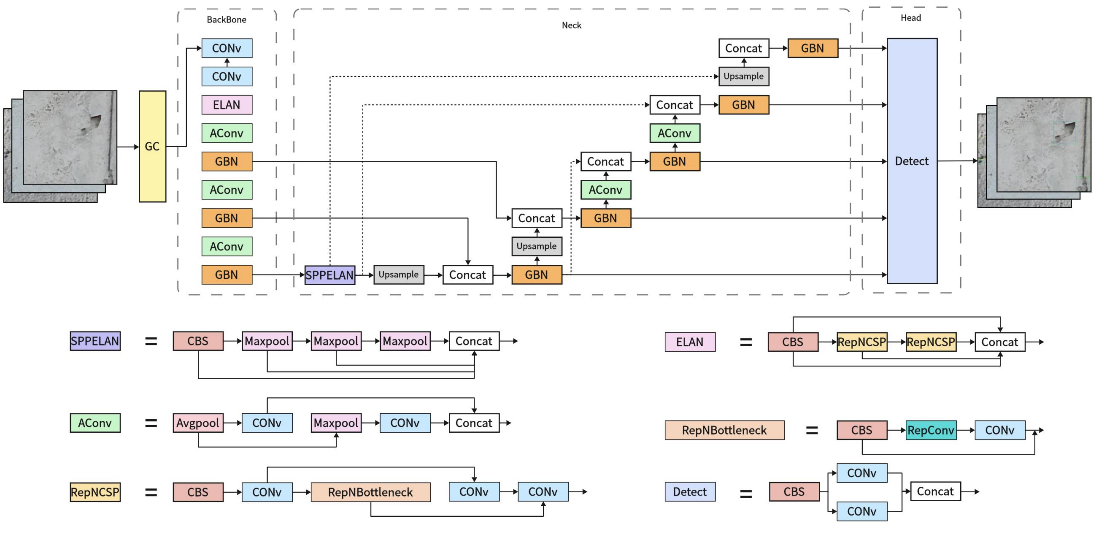

Fig. 1.Overall process for detection of surface defects on bridge piers

Gamma correction enhances the image contrast between defective areas and the background, facilitating defect identification by the target detection model. This model, based on YOLOv9, incorporates a specially designed GhostBottleNeck (GBN) module, replacing the original ELAN module. GBN improves the model's ability to detect small targets through dynamic convolution, thereby enhancing overall performance.

3.2.2 Gamma correction

Gamma correction is a widely employed image enhancement technique that adjusts brightness and enhances contrast, particularly in an image's darker regions. This process not only renders colors more naturally but also enhances the visual effects crucial for computer vision systems [61].

The gamma correction formula adjusts each pixel value of the image I, based on a predefined gamma value γ. The formula is expressed as follows::

$$I_{\text{out}} = I_{\text{in}}^{\gamma} \#(1)$$

In formula (1), I_{in} represents the pixel value of the input image, ranging from 0 to 1. The gamma correction parameter, γ regulates the image's brightness enhancement or reduction. I_{out} denotes the pixel value after gamma correction.

When γ<1, dark regions of the image brighten, enhancing visibility in shadowed details. Conversely, when γ>1, brighter areas darken, increasing image contrast and sharpening bright details [55].

In this study, the surface defects on bridge piers are similar in color to the background, posing significant challenges for the detection model. Adjusting the gamma value to less than 1 enhances the brightness of darker areas without overly brightening lighter areas, thus improving the dynamic range and visual contrast of the image. Gamma correction impacts not only brightness but also indirectly influences color saturation and depth. By applying appropriate gamma correction, color differences between similar targets and backgrounds become more distinct, aiding in distinguishing the target more clearly from its surroundings [45, 53].

3.2.3 GhostBottleNeck

Fig 2 illustrates the overall structure of the GhostBottleNeck (GBN), which resembles the basic residual block found in ResNet. The main pathway of GBN consists of two sequential Ghost Modules: the first module expands the number of channels, while the second reduces them back to the original count of the input channels. The residual connections are similar to those used in ResNet [57, 62].

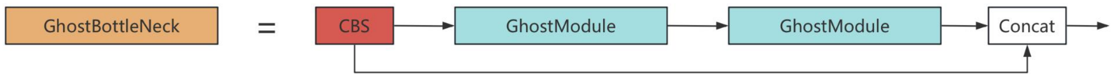

Fig. 2.GhostBottleNeck Overall Structure

GBN does not reduce the height and width of the input feature maps; instead, it increases the network's depth to capture more feature information [57].

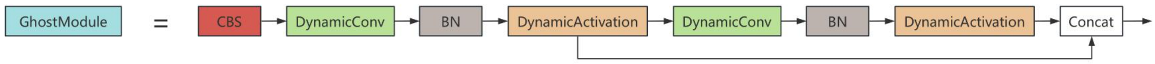

Fig. 3.Internal structure of the GhostModule module

As illustrated in Fig.3, the GhostModule leverages advancements in modern deep learning, including dynamic convolution techniques and dynamic activation layers, to reduce computational demands and enhance efficiency. It begins with a dynamic convolution layer (DynamicConv), which conducts a convolution from the input channel to the initialization channel, followed by a batch normalization layer (BatchNorm2d) and a dynamic activation layer (DynamicActivation). These steps are repeated, and the outputs of these operations are combined using a concatenation operation. The convolution kernels for dynamic convolution are dynamically generated based on the input data, with adjustable parameters for varying inputs, enhancing the network's adaptability [57, 60, 63].

Each input sample is processed by compressing its features into a vector using global pooling, followed by passing through one or more fully connected layers employing functions like softmax to produce a set of weights. These weights, reflecting the relevance of different convolutional kernels, are multiplied with the parameters of each kernel. The accumulated results form the final convolution kernel for the operation, effectively adapting the kernel parameters to suit the current input [64].

The essence of dynamic convolution lies in its ability to adapt convolution kernel parameters to specific inputs, rather than merely selecting the most relevant kernels. This method significantly enhances the network's adaptability and flexibility across diverse inputs, optimizing for particular data features. Such adaptability not only enables the network to better handle input variations but also boosts its overall performance and generalization capabilities [65].

Dynamic activation function selection is a deep learning technique enabling neural networks to dynamically choose the most suitable activation function based on input data or their internal state [66].

Several activation functions are stored in a dictionary and are switched during training based on specific criteria.

The main benefits of dynamic activation function selection include enhanced learning of complex nonlinear patterns through adaptation to various input types, reduced risk of overfitting, and improved model performance on unseen data. This adaptability is crucial for handling diverse or variably distributed data, optimizing processing strategies for each specific scenario [67].

The dynamic convolution technique selects convolution kernels based on input, enhancing the model's focus on varied input features and boosting its defect detection capabilities. Similarly, dynamic activation functions increase the model's flexibility and adaptability, significantly aiding in performance enhancement.

3.3 Data collection

This study utilizes a subset of 1500 images from the SDNET2018 dataset, comprising common defects such as cracks and depressions on bridge piers [68]. Given the challenges of acquiring extensive data in industrial defect detection, this research asserts that effective defect detection is achievable with limited datasets. Therefore, these 1500 images were specifically selected to train and evaluate the model, reflecting the pragmatic constraints in the industrial context [69].

Fig. 4. Example of dataset content

Fig.4 displays various images from the dataset, highlighting the variability in crack thickness and shape, as well as the irregularity and diverse sizes of depressions. These characteristics aim to mirror the real-world conditions of surface defects on bridge piers. Additionally, the inconsistent background colors of construction materials and varying lighting conditions in the images further prepare the model to handle complex environments and backgrounds [70].

3.4 Evaluation metrics

In this study, the performance of the model was evaluated using several key metrics: mean Average Precision (mAP), the number of parameters, and Giga Floating Point Operations Per Second (GFLOPS). The mAP metric was specifically employed to assess the accuracy and effectiveness of the model in detecting and classifying targets. In contrast, the number of parameters and GFLOPS were utilized to evaluate the model's efficiency and computational lightness, indicating how well the model performs under resource constraints [71, 72].

Mean Average Precision (mAP) is calculated by first determining the Average Precision (AP) for each category within a multi-category target detection or classification framework. AP for each category is derived from the area under the curve in a Precision vs. Recall graph. The mAP is then obtained by averaging these AP values across all categories. This metric is crucial for understanding the model's precision in distinguishing between different types of targets, which is especially important in complex scenarios where multiple categories are present [73].

Precision quantifies the proportion of positive identifications made by the model that are actually correct. This metric is calculated using the following formula:$\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP) + False Positives (FP)}} \#(2)\#$

True Positives (TP) represent the number of positive class samples that the model correctly identifies as positive. Conversely, False Positives (FP) are instances where the model incorrectly identifies a negative class sample as positive [74].

Recall measures the proportion of actual positives correctly identified by the model, reflecting the model's ability to detect all relevant instances [75]. It is computed using the following formula:

$$\text{Recall} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP) + False Negatives (FN)}} \#(3)\#$$

The GFLOPS (Giga Floating Point Operations Per Second) metric measures the computational cost of models and hardware, significantly influencing the speed and efficiency of both training and inference phases. Effective evaluation and strategic use of GFLOPS are crucial for optimizing deep learning tasks, ensuring that resources are used efficiently to achieve the best performance [76].

The number of parameters is a fundamental metric that gauges the size and complexity of a neural network. It quantifies the total count of trainable elements within the model, such as weights and biases. A higher number of parameters generally indicates a more complex model, which can significantly affect performance, extend training durations, increase memory demands, and influence inference speeds [77].

3.5 Research material

During the research, Visual Studio Code was employed for writing code, while LabelImg was utilized for annotating the data. For training the model, the study utilized a 2080TI GPU, leveraging CUDA version 11.6, Python version 3.8, and Pytorch version 1.13.1 to ensure efficient computation and compatibility.

CHAPTER 4 - EXPERIMENTAL

This section describes the experiments conducted in the study, including the baseline model experiment, the image enhancement experiment, and the GhostBottleNeck module fusion experiment.

4.1 Baseline model

In this study, to optimize model performance, comparative experiments were conducted using state-of-the-art YOLO series models—YOLOV8-S, YOLOV9-S, GELAN-S, and the newly released YOLOV10—all of which were trained and evaluated on the dataset [9, 41, 42].

Table.2. Baseline Model Performance Comparison

Model	mAP50(%)	Parameters	GFLOPs
YOLOV8-S	74.3	111.3M	28.6
YOLOV9-S	**76.4**	95.9M	38.7
GELAN-S	76.1	71.9M	26.9
YOLOV10-S	71.8	80.6M	24.8

According to the data presented in Table 2, YOLOV9-S stands out as the leading model in terms of detection accuracy, achieving a mean Average Precision (mAP50) of 76.4%. This performance not only highlights its superior capability but also sets a benchmark for other models. Close on its heels is GELAN-S, which showcases a comparable detection prowess with a mAP50 of 76.1%. This slight variance underscores the competitive nature of these models in handling complex detection tasks. Following them, YOLOV10-S demonstrates a commendable mAP50 of 71.8%, proving itself as a robust contender in the realm of object detection. Despite its strengths, YOLOV8-S lags slightly behind, with a mAP50 of 74.3%, indicating room for improvement in future iterations of the model.

When assessing model complexity, YOLOV8-S emerges as the most intricate, featuring 111.3 million parameters. This high parameter count is reflective of its complex architecture designed to capture nuanced features in data. YOLOV9-S, with 95.9 million parameters, presents a slightly simpler yet robust architecture that strategically balances complexity with high performance, making it suitable for varied detection tasks. YOLOV10-S, with 80.6 million parameters, offers moderate complexity and is tailored for efficient performance without overburdening computational resources. Conversely, GELAN-S stands as the most streamlined model with only 71.9 million parameters, emphasizing simplicity and speed, which may benefit applications requiring lower resource consumption.

YOLOV9-S, demanding the most computational resources at 38.7 GFLOPs, reflects its capacity for handling complex tasks at a higher computational cost. YOLOV8-S, using 28.6 GFLOPs, offers a balance, consuming less than YOLOV9-S but more than other models. GELAN-S, at 26.9 GFLOPs, proves to be more resource-efficient than YOLOV8-S, suitable for tasks needing lower GFLOP consumption. YOLOV10-S is exceptionally efficient, requiring only 24.8 GFLOPs, yet achieves a higher mAP50 than YOLOV8-S, demonstrating an excellent balance of low computational demand and high performance.

In industrial defect detection, accuracy is paramount, as the primary objective is to identify defects effectively. Given its superior performance metrics, YOLOV9-S emerges as the most appropriate benchmark model for this specific research task.

4.2 Image enhancement

In this study, to tackle the challenge of distinguishing defective parts from the background due to color similarities, several image enhancement techniques were employed on the original dataset to improve image contrast. These methods included linear transformation, histogram equalization, contrast-limited adaptive histogram equalization (CLAHE), and gamma correction [15, 23, 78, 79]. Each technique aimed to modify the visual aspects of the images, making the defects more discernible against similarly colored backgrounds. Subsequently, the enhanced datasets were used to train and evaluate the original YOLOV9-S model.

Table.3. Comparison of the effect of different image enhancement methods

Model	**IMAGE ENHANCEMENT**	**mAP50(%)**
YOLOV9-S	Linear	69.4
	AHE	75.7
	CLAHE	74.5
	Gamma Correction	**78.9**

As detailed in Table 3, the YOLOV9-S model's performance after applying the linear enhancement method registers at 69.4%, a significant 7% decrease from the baseline model. This reduction underscores that linear enhancement, while straightforward, lacks the complexity needed to effectively highlight essential features within the images. This method fails to optimize the visibility of critical elements that aid the model in accurately recognizing the targets, suggesting that more sophisticated techniques might be necessary for handling such nuanced data scenarios.

Employing the Adaptive Histogram Equalisation (AHE) method elevates the model's performance to 75.7%, closely approaching but still slightly below the benchmark by 0.7 percentage points. This enhancement nears but does not exceed the benchmark model's efficiency, demonstrating AHE's capacity to considerably enhance image contrast and facilitate better target detection. However, it subtly implies that while AHE significantly bolsters contrast and clarity, thereby aiding in target delineation, it still does not fully optimize performance to surpass the highest standards set by the benchmark model. This could point to inherent limitations in how AHE handles data variability and complex scenarios within the dataset.

Contrast-Limited Adaptive Histogram Equalisation (CLAHE) achieves a performance of 74.5%, falling 1.9 percentage points short of the benchmark model. Despite its effectiveness in controlling image noise and preventing excessive enhancement, CLAHE does not reach the high performance standards set by the benchmark. This outcome suggests that while CLAHE is beneficial for maintaining image quality by avoiding over-enhancement, it might not sufficiently highlight the critical features required for optimal defect detection in more complex scenarios.

Gamma correction emerges as the most effective method, achieving a mean Average Precision (mAP) of 78.9%, which surpasses the benchmark by 2.5 percentage points. This superior performance demonstrates gamma correction's ability to significantly enhance image details through precise luminance distribution adjustments, particularly in both dark and bright areas. This enhancement directly contributes to the model's improved capability in accurately detecting targets within complex visual environments.

Collectively, these results underscore gamma correction as the standout performer in these experiments, significantly enhancing model performance. In contrast, other tested methods, while beneficial, do not exceed or only marginally improve upon the baseline model's capabilities. This disparity emphasizes the critical role of selecting appropriate image enhancement techniques tailored to specific model requirements and scene complexities, ultimately impacting the overall effectiveness of defect detection.

4.3 GhostBottleneck

This research investigates the efficacy of substituting the RepNCSPELAN4 module in YOLOV9 with a range of alternative modules specifically designed to enhance the model's capability in detecting small targets. Extensive training and testing were carried out on a dataset that was enhanced for image quality, employing these innovative modules to assess their performance enhancements.

Table.4. Effect of replacing the RepNCSPELAN4 module with a different module

Model	Module	mAP50(%)	Parameters	GFLOPs
YOLOV9-S	RepNCSPELAN4	78.9	95.9M	38.7
	DCNv3 [80]	69.2	102.4M	40.2
	Context Guided Block [81]	79.4	98.4M	39.6
	GhostBottleneck	**80.3**	86.5M	31.2
	DiverseBranchBlock [82]	71.4	98.2M	38.4
	FasterBlock [83]	76.1	87.6M	33.4
	MSBlock [84]	78.6	84.7M	29.6

According to the results detailed in Table 4, the GhostBottleneck module outperformed others with a remarkable mAP50 of 80.3%, closely followed by the Context Guided Block, which achieved a 79.4% score. In comparison, the original RepNCSPELAN4 module reached a score of 78.9%. While other modules like DCNv3 and DiverseBranchBlock showed competent performance, their scores of 69.2% and 71.4% respectively were considerably lower. This differential performance highlights the superior efficacy of the GhostBottleneck and Context Guided Block in enhancing target recognition accuracy within the model.

Regarding model parameters, the GhostBottleneck module is the most streamlined, featuring 86.5 million parameters, in contrast to DCNv3, which holds the highest count at 102.4 million parameters. This variance in parameter count directly influences the model's complexity and the computational resources it demands during operation. This aspect is crucial for understanding the balance between model depth and operational efficiency.

In terms of computational efficiency, the GFLOPs metric reveals that MSBlock leads with a low 29.6 GFLOPs, highlighting its superior efficiency. Conversely, DCNv3 exhibits the highest computational load, demanding 40.2 GFLOPs. Models with lower GFLOPs are generally more efficient, utilizing fewer computational resources, which is advantageous for deploying models in resource-constrained environments.

Collectively, the diverse modules tested showcase varying trade-offs among performance metrics, model size, and computational efficiency. Notably, GhostBottleneck emerges as the most balanced option, excelling in both accuracy and efficiency. This optimal performance profile makes it particularly suitable for applications where both high precision and computational frugality are essential.

4.4 Ablation Study

This study implemented an ablation experiment to rigorously assess the individual and combined impacts of various image enhancement techniques and the GhostBottleneck module on the accuracy of the model.

Table.5. Gamma corrected(GC) and GhostBottleNeck(GBN) ablation experiments

Model	**GC**	**GBN**	**mAP50(%)**
YOLOV9-S			76.4
	√		78.9
		√	78.2
	√	√	**80.3**

According to Table 5, the implementation of gamma correction significantly increases the model’s mean Average Precision (mAP50) from 76.1% to 78.9%. This substantial improvement underscores the effectiveness of enhancing image contrast in boosting the model's ability to recognize and accurately classify targets. Utilizing the GhostBottleneck module independently results in an mAP50 of 77.8%, an improvement over the baseline though slightly less than that achieved with gamma correction. This enhancement highlights the module’s effectiveness in refining the model's capability to detect smaller targets, affirming its role in optimizing component-specific performance. When both gamma correction and the GhostBottleneck module are simultaneously employed, the model achieves an optimal mAP50 of 80.3%. This peak performance demonstrates the synergistic effect of combining these techniques, significantly enhancing the model's overall effectiveness and accuracy in complex scenarios.

Additionally, this study implemented decimation experiments on the dynamic convolution (DC) and dynamic activation (DA) functions within the GhostModule. These experiments were conducted using the YOLOV9-S-GM model, which was equipped with the GhostBottleNeck replacement and tested on gamma-corrected datasets. The objective was to assess how these nuanced, target-specific components—DC and DA—contribute to enhancing the model's performance, particularly in identifying and classifying small targets more effectively.

Table.6. Dynamic convolution and dynamic activation function ablation experiments

Model	**DC**	**DA**	**mAP50(%)**	**Parameters**	**GFLOPs**
YOLOV9-S-GM			78.2	76.8M	31.9
	√		79.2	85.6M	30.9
		√	78.3	77.4M	32.0
	√	√	**80.3**	86.5M	31.2

The experimental results were analyzed across four configurations: The baseline, with no dynamic features enabled, demonstrated a mean Average Precision (mAP50) of 78.2%, utilized 76.8 million parameters, and required 31.9 GFLOPs, serving as a control to gauge the impact of dynamic enhancements on model performance. In the second configuration, with only dynamic convolution activated, the model achieved a higher mAP50 of 79.2% and required more parameters at 85.6 million; however, the computational demand, measured in GFLOPs, decreased to 30.9, underscoring an enhancement in computational efficiency attributed to dynamic convolution. In the third configuration, where only dynamic activation is implemented, there is a marginal decrease in mAP50 to 78.3%, with parameters rising to 77.4 million and GFLOPs increasing to 32.0. This indicates that while dynamic activation adds complexity, it offers minimal gains in performance efficiency. When both dynamic convolution and dynamic activation are employed, the model reaches an optimal performance with an mAP50 of 80.3% and a parameter tally of 86.5 million. Furthermore, this configuration reduces the GFLOPs to 31.2, showcasing a significant boost in performance and computational efficiency through the synergistic effects of these features.

The integration of dynamic convolution and activation significantly enhances the model's accuracy in target detection tasks while optimizing computational resource usage. This balance of performance and efficiency is crucial in resource-limited settings, enabling the model to operate effectively without compromising on accuracy.

4.5 Final Performance Comparison

This study compares the performance of the improved model with current popular defect detection models.

Table.7. Final Model Performance Comparison

Model	**mAP50(%)**	**Parameters**	**GFLOPs**
YOLOV8-S	74.3	111.3M	28.6
YOLOV9-S	76.4	95.9M	38.7
GELAN-S	76.1	71.9M	26.9
YOLOV10-S	71.8	80.6M	24.8
This Study	**80.3**	86.5M	31.2

According to the data presented in Table 7, the model developed in this study significantly enhances detection accuracy, achieving an mAP50 of 80.3%. This represents an impressive gain of 3.9 percentage points over the original YOLOV9-S model, which scored 76.4% in the same metric. This substantial increase not only demonstrates the model's enhanced ability to accurately identify targets but also highlights the effectiveness of the improvements implemented in the model's design and function.

In addition, the model in this study demonstrates an optimized use of parameters, utilizing only 86.5 million parameters compared to the 95.9 million used by the YOLOV9-S. This reduction of approximately 9.4 million parameters signifies a streamlined yet efficient architecture. Such optimization not only preserves high accuracy but also enhances the model's feasibility for deployment and storage, particularly in environments where memory and processing power are limited.

In the realm of computational complexity, measured in GFLOPs, the model developed in this study operates at 31.2 GFLOPs, a substantial reduction from the 38.7 GFLOPs required by the YOLOV9-S model. This decrease of 7.5 GFLOPs suggests a significant enhancement in operational efficiency, including lower energy demands and accelerated processing speeds, which are crucial for real-world application scenarios where quick response times and energy efficiency are valued.

Overall, the model presented in this study not only achieves the highest detection accuracy but also optimizes both the number of parameters and computational complexity. This balance of high precision, streamlined architecture, and reduced computational demands underscores the model's superior performance and efficiency, making it particularly well-suited for deployment in resource-sensitive environments where maintaining operational effectiveness without overburdening system resources is critical.

CHAPTER 5 - DISCUSSIONS

This section will summarise the contributions of this study as well as the shortcomings and future expectations

5.1 Contributions

In this study, we have effectively improved the accuracy and efficiency of bridge abutment surface defect detection by introducing advanced image enhancement techniques and target detection methods, especially Gamma correction and GhostBottleneck (GBN) module. Combined with dynamic convolution and activation techniques, our model shows significant performance advantages in dealing with fine targets and complex backgrounds, and the effectiveness of this study on the task of bridge abutment surface defect detection is demonstrated through a series of experiments.

Among the various image enhancement methods employed, Gamma correction is distinguished by its exceptional capability for contrast adjustment, which markedly influences the overall efficacy of the model. Gamma correction amplifies the details in both dark and bright areas by modulating the luminance distribution within the image. This adjustment is particularly vital for distinguishing between defects and backgrounds that are visually similar. These observations align with extant literature that elucidates the enhancement of target detection performance through image contrast modification.

The GhostBottleneck (GBN) module is ingeniously crafted to minimize computational complexity while preserving efficient information processing capabilities. Empirical comparisons reveal that the GBN module significantly augments the model's proficiency in recognizing small-scale targets, concurrently reducing the parameter count and computational expenditure. This enhancement is of paramount importance for identifying defects on the surfaces of bridge abutments, which are typically minor in size and heterogeneous in form. The superior performance of GBN over traditional ResNet-like modules substantiates the efficacy of deploying deep learning for meticulous target detection.

The integration of dynamic convolution and activation techniques introduces a methodology for tailoring the network's response to varying input data. This adaptive strategy enhances the model's flexibility in accommodating diverse input features, which in turn elevates the precision of defect detection and amplifies the model's generalization capacity. Moreover, the strategic application of dynamic techniques facilitates the optimal utilization of computational resources, an aspect critical in scenarios where such resources are constrained.

This research contributes substantially to the precision and efficiency of defect detection on bridge abutment surfaces, diminishing the reliance on manual inspections and the potential for human errors. Additionally, the model's robust performance renders it suitable for real-time monitoring systems, offering a viable technical solution for the surveillance of bridge integrity.

5.2 Limitations and expectations

Gamma correction, a widely used image enhancement technique, plays a significant role in adjusting the luminance distribution to improve contrast in various imaging tasks. In the context of defect detection on bridge abutment surfaces, gamma correction proves beneficial by enhancing the visibility of details in both dark and bright regions of the image. This adjustment is crucial for distinguishing between subtle variations in defect features and background textures. However, gamma correction may not universally apply to all datasets or defect types due to its fixed parameterization and limited adaptability to varying image conditions.

To address this limitation, it is essential to explore more sophisticated and adaptive image enhancement techniques. Advanced methods, such as adaptive histogram equalization or data-driven enhancement approaches using generative models, could potentially offer more nuanced adjustments tailored to specific image characteristics. These intelligent enhancement techniques can dynamically adapt to the image content and noise levels, thereby reducing the impact of image noise and improving the robustness of the defect detection model.

This study introduces novel modules, including the GhostBottleneck (GBN) and dynamic convolution techniques, to address challenges specific to bridge abutment surface defect detection. The GBN module effectively reduces computational complexity while maintaining robust information processing capabilities, crucial for detecting small and diverse defects. Despite these advancements, there remains room for further improvement in model performance through sophisticated feature fusion strategies.

Feature fusion, which integrates information from multiple sources or scales, can enhance the model’s ability to capture and represent complex defect patterns. Techniques such as multi-scale feature aggregation, attention mechanisms, or cross-modal fusion can potentially improve the model's performance by leveraging complementary information from different feature representations. Incorporating these methods could lead to more accurate and comprehensive defect detection, addressing some of the limitations inherent in the current model architecture.

The large size and computational demands of contemporary target detection models pose significant challenges for real-world deployment, particularly in resource-constrained environments. While this study has made strides in reducing the computational cost by employing the GBN module, further optimization is essential for practical application.

Techniques such as model pruning, quantization, and knowledge distillation can contribute to reducing the model's size and computational requirements. Model pruning involves removing less significant parameters or neurons, while quantization reduces the precision of numerical representations to lower memory and computation needs. Knowledge distillation transfers the knowledge from a larger, more complex model to a smaller, more efficient one. These approaches can further enhance the deployment feasibility of defect detection systems, making them more suitable for real-time monitoring and edge applications.

In summary, while gamma correction and current model innovations significantly contribute to defect detection, exploring more adaptive image enhancement techniques and advanced feature fusion strategies could further refine model performance. Additionally, addressing computational efficiency through pruning and other optimization techniques will be crucial for the practical deployment of defect detection technologies. Future research should focus on integrating these advanced methodologies to achieve more accurate, efficient, and deployable defect detection solutions.

CHAPTER 6 - CONCLUSION

In this comprehensive study, we introduced a deep learning-based method for detecting defects on bridge pier surfaces, which addresses the inherent challenges faced by current methodologies due to the diverse appearances of defects, complex backgrounds, varying lighting conditions, and issues with detecting small targets.

Key to this research was the innovative use of gamma correction as a preprocessing technique to enhance image quality. This method significantly improved visual clarity by enhancing the contrast and sharpness of the images, making darker details more discernible. The application of gamma correction on a dataset of bridge abutment surface defects led to a marked improvement in the model's ability to distinguish and accurately identify features. This enhancement was pivotal in enabling more nuanced feature extraction processes essential for the precise identification of structural anomalies, thereby enhancing the reliability of structural health monitoring systems based on image analysis.

Further advancing our model's capabilities, the GhostBottleneck (GBN) module was integrated to address the challenges posed by diverse defect characteristics and the detection of small targets. The GBN module employs dynamic convolutional layers that adjust their parameters in response to the input image's features, allowing for more accurate feature extraction and improved adaptability of the model. This dynamic adjustment is critical for effectively detecting small defects, which are often overlooked by traditional convolutional networks due to the limitations of fixed convolutional kernels.

The GBN module not only refined the detection process by optimizing kernel adjustments but also enhanced the overall model efficiency by judiciously using computational resources. It adapted the computational complexity according to the necessity for intensive feature extraction, maintaining lower processing levels in less demanding regions. This adaptive feature extraction significantly improved the model's performance in real-time applications by responding adeptly to complex backgrounds and varied defect types.

The integration of the GBN module provided a robust and flexible solution for bridge defect detection, significantly improving the performance of the target detection model through a dynamically tailored deep learning architecture. This approach proved particularly effective in scenarios involving a wide range of defect sizes and complexities.

Finally, the effectiveness of the gamma correction technique and the GBN module was rigorously validated through exhaustive ablation studies. These studies assessed the individual contributions of each component to the model's performance, affirming their significant roles in enhancing detection accuracy and operational efficiency. Moreover, comparative experiments with state-of-the-art models demonstrated the superior performance and practical value of the proposed model, particularly in its adaptability to complex scenarios and its efficiency in processing.

In summary, this research not only demonstrated significant improvements over existing defect detection methods but also laid a strong foundation for future advancements in structural health monitoring, ensuring more reliable and efficient detection of defects in bridge infrastructure.

CHAPTER 7 - REFLECTIONS

The principal objective of this study was to devise a novel technique for detecting surface defects on bridge piers by merging sophisticated image enhancement technologies with an advanced target detection model. The results affirm that this objective has been substantially met, as evidenced by a 3.9% improvement in the mean Average Precision (mAP50) relative to the baseline model. Additionally, the new method demonstrates a reduction in both computational demand and the number of parameters by 9.8% and 7.5 GFLOPS respectively, suggesting enhanced feasibility for practical implementation in future deployments.

Despite these advancements, certain aspects of the project suggest potential areas for further refinement. Specifically, the use of gamma correction has markedly improved recognition capabilities within our dataset; however, its applicability across diverse datasets remains limited. Experiments with various adaptive image enhancement techniques yielded suboptimal results, indicating that future efforts might benefit from exploring deep learning-based image enhancement methods, such as Generative Adversarial Networks (GANs). Such techniques could tailor image quality enhancements more effectively to specific dataset distributions and user requirements, potentially elevating the overall efficacy of the detection system.

Furthermore, enhancements to the model's 'Neck' component, which integrates different scales of features, could yield better focus and detection accuracy for small-scale defects. Implementing more targeted image fusion methods could substantially improve the model’s sensitivity to smaller targets. Unfortunately, time constraints precluded this experimental development during the current phase of research.

In conclusion, while this study marks a significant step forward in the field of structural health monitoring by integrating enhanced imaging with deep learning for defect detection, ongoing research is essential to refine these methodologies further. Future investigations should focus on extending the applicability of image enhancements and optimizing the architectural components of the detection model to maximize performance across varied operational contexts.

REFERENCES

[1] M. Zaid, Z. Yazdanfar, H. Chowdhury, and F. Alam, "A review on the methods used to reduce the scouring effect of bridge pier," *Energy Procedia,* vol. 160, pp. 45-50, 2019.

[2] Z. Chen, W. Zhai, C. Cai, and Y. Sun, "Safety threshold of high-speed railway pier settlement based on train-track-bridge dynamic interaction," *Science China Technological Sciences,* vol. 58, pp. 202-210, 2015.

[3] S. Teng, Z. Liu, and X. Li, "Improved YOLOv3-based bridge surface defect detection by combining High-and low-resolution feature images," *Buildings,* vol. 12, no. 8, p. 1225, 2022.

[4] J. Zhu, C. Zhang, H. Qi, and Z. Lu, "Vision-based defects detection for bridges using transfer learning and convolutional neural networks," *Structure and Infrastructure Engineering,* vol. 16, no. 7, pp. 1037-1049, 2020.

[5] A. Nasr, I. Björnsson, D. Honfi, O. Larsson Ivanov, J. Johansson, and E. Kjellström, "A review of the potential impacts of climate change on the safety and performance of bridges," *Sustainable and Resilient Infrastructure,* vol. 6, no. 3-4, pp. 192-212, 2021.

[6] Z. Zhao, B. Li, R. Dong, and P. Zhao, "A surface defect detection method based on positive samples," in *PRICAI 2018: Trends in Artificial Intelligence: 15th Pacific Rim International Conference on Artificial Intelligence, Nanjing, China, August 28–31, 2018, Proceedings, Part II 15*, 2018: Springer, pp. 473-481.

[7] T. Czimmermann *et al.*, "Visual-based defect detection and classification approaches for industrial applications—a survey," *Sensors,* vol. 20, no. 5, p. 1459, 2020.

[8] A.-A. Tulbure, A.-A. Tulbure, and E.-H. Dulf, "A review on modern defect detection models using DCNNs–Deep convolutional neural networks," *Journal of Advanced Research,* vol. 35, pp. 33-48, 2022.

[9] M. Hussain, "YOLO-v1 to YOLO-v8, the rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection," *Machines,* vol. 11, no. 7, p. 677, 2023.

[10] A. D. Santoso, F. B. Cahyono, B. Prahasta, I. Sutrisno, and A. Khumaidi, "Development of pcb defect detection system using image processing with yolo cnn method," *International Journal of Artificial Intelligence Research,* vol. 6, no. 1, 2022.

[11] Z. Qiu, S. Wang, Z. Zeng, and D. Yu, "Automatic visual defects inspection of wind turbine blades via YOLO-based small object detection approach," *Journal of electronic imaging,* vol. 28, no. 4, pp. 043023-043023, 2019.

[12] Q. Li and S. Ren, "A visual detection system for rail surface defects," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews),* vol. 42, no. 6, pp. 1531-1542, 2012.

[13] Y. Qi *et al.*, "A comprehensive overview of image enhancement techniques," *Archives of Computational Methods in Engineering,* pp. 1-25, 2021.

[14] M. Kaur, J. Kaur, and J. Kaur, "Survey of contrast enhancement techniques based on histogram equalization," *International Journal of Advanced Computer Science and Applications,* vol. 2, no. 7, 2011.

[15] P. Gupta, J. Kumare, U. Singh, and R. Singh, "Histogram based image enhancement techniques: a survey," *Int J Comput Sci Eng,* vol. 5, no. 6, pp. 475-484, 2017.

[16] N. Kaur and K. Singh, "Image enhancement techniques: A selected review," *International Journal of Engineering Research and Technology (IJERT),* vol. 2, no. 3, 2013.

[17] S. M. Pizer *et al.*, "Adaptive histogram equalization and its variations," *Computer vision, graphics, and image processing,* vol. 39, no. 3, pp. 355-368, 1987.

[18] H. Lidong, Z. Wei, W. Jun, and S. Zebin, "Combination of contrast limited adaptive histogram equalisation and discrete wavelet transform for image enhancement," *IET image processing,* vol. 9, no. 10, pp. 908-915, 2015.

[19] V. Thakur and H. Singh, "An artificial neural network based adaptive histogram equalization algorithm for enhancement of low contrast images," in *2021 8th International Conference on Computing for Sustainable Global Development (INDIACom)*, 2021: IEEE, pp. 268-273.

[20] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics gems IV*, 1994, pp. 474-485.

[21] P. Musa, F. Al Rafi, and M. Lamsani, "A Review: Contrast-Limited Adaptive Histogram Equalization (CLAHE) methods to help the application of face recognition," in *2018 third international conference on informatics and computing (ICIC)*, 2018: IEEE, pp. 1-6.

[22] H. Farid, "Blind inverse gamma correction," *IEEE transactions on image processing,* vol. 10, no. 10, pp. 1428-1433, 2001.

[23] S. Rahman, M. M. Rahman, M. Abdullah-Al-Wadud, G. D. Al-Quaderi, and M. Shoyaib, "An adaptive gamma correction for image enhancement," *EURASIP Journal on Image and Video Processing,* vol. 2016, pp. 1-13, 2016.

[24] S. A. Amiri and H. Hassanpour, "A preprocessing approach for image analysis using gamma correction," *International Journal of Computer Applications,* vol. 38, no. 12, pp. 38-46, 2012.

[25] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," *Proceedings of the IEEE,* vol. 111, no. 3, pp. 257-276, 2023.

[26] Q. Bai, S. Li, J. Yang, Q. Song, Z. Li, and X. Zhang, "Object detection recognition and robot grasping based on machine learning: A survey," *IEEE access,* vol. 8, pp. 181855-181879, 2020.

[27] D. M. Ramík, C. Sabourin, R. Moreno, and K. Madani, "A machine learning based intelligent vision system for autonomous object detection and recognition," *Applied intelligence,* vol. 40, pp. 358-375, 2014.

[28] Z.-Q. Zhao, P. Zheng, S.-t. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE transactions on neural networks and learning systems,* vol. 30, no. 11, pp. 3212-3232, 2019.

[29] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580-587.

[30] P. Bharati and A. Pramanik, "Deep learning techniques—R-CNN to mask R-CNN: a survey," *Computational Intelligence in Pattern Recognition: Proceedings of CIPR 2019,* pp. 657-668, 2020.

[31] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440-1448.

[32] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems,* vol. 28, 2015.

[33] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779-788.

[34] H. Lou *et al.*, "DC-YOLOv8: small-size object detection algorithm based on camera sensor," *Electronics,* vol. 12, no. 10, p. 2323, 2023.

[35] K. Tong and Y. Wu, "Rethinking PASCAL-VOC and MS-COCO dataset for small object detection," *Journal of Visual Communication and Image Representation,* vol. 93, p. 103830, 2023.

[36] J. Terven, D.-M. Córdova-Esparza, and J.-A. Romero-González, "A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas," *Machine Learning and Knowledge Extraction,* vol. 5, no. 4, pp. 1680-1716, 2023.

[37] N. D. T. Yung, W. Wong, F. H. Juwono, and Z. A. Sim, "Safety helmet detection using deep learning: Implementation and comparative study using YOLOv5, YOLOv6, and YOLOv7," in *2022 International Conference on Green Energy, Computing and Sustainable Technology (GECOST)*, 2022: IEEE, pp. 164-170.

[38] C. Li *et al.*, "YOLOv6: A single-stage object detection framework for industrial applications," *arXiv preprint arXiv:2209.02976,* 2022.

[39] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 7464-7475.

[40] G. Wang, Y. Chen, P. An, H. Hong, J. Hu, and T. Huang, "UAV-YOLOv8: A small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios," *Sensors,* vol. 23, no. 16, p. 7190, 2023.

[41] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, "Yolov9: Learning what you want to learn using programmable gradient information," *arXiv preprint arXiv:2402.13616,* 2024.

[42] A. Wang *et al.*, "Yolov10: Real-time end-to-end object detection," *arXiv preprint arXiv:2405.14458,* 2024.

[43] T. Diwan, G. Anirudh, and J. V. Tembhurne, "Object detection using YOLO: Challenges, architectural successors, datasets and applications," *multimedia Tools and Applications,* vol. 82, no. 6, pp. 9243-9275, 2023.

[44] Y. Chen, Y. Ding, F. Zhao, E. Zhang, Z. Wu, and L. Shao, "Surface defect detection methods for industrial products: A review," *Applied Sciences,* vol. 11, no. 16, p. 7657, 2021.

[45] X. Xie, "A review of recent advances in surface defect detection using texture analysis techniques," *ELCVIA: electronic letters on computer vision and image analysis,* pp. 1-22, 2008.

[46] S. Shahrabadi, Y. Castilla, M. Guevara, L. G. Magalhães, D. Gonzalez, and T. Adão, "Defect detection in the textile industry using image-based machine learning methods: a brief review," in *Journal of Physics: Conference Series*, 2022, vol. 2224, no. 1: IOP Publishing, p. 012010.

[47] S. R. Saufi, Z. A. B. Ahmad, M. S. Leong, and M. H. Lim, "Challenges and opportunities of deep learning models for machinery fault detection and diagnosis: A review," *Ieee Access,* vol. 7, pp. 122644-122662, 2019.

[48] P. M. Bhatt *et al.*, "Image-based surface defect detection using deep learning: A review," *Journal of Computing and Information Science in Engineering,* vol. 21, no. 4, p. 040801, 2021.

[49] R. Usamentiaga, D. G. Lema, O. D. Pedrayes, and D. F. Garcia, "Automated surface defect detection in metals: a comparative review of object detection and semantic segmentation using deep learning," *IEEE Transactions on Industry Applications,* vol. 58, no. 3, pp. 4203-4213, 2022.

[50] M. Hussain and R. Khanam, "In-Depth Review of YOLOv1 to YOLOv10 Variants for Enhanced Photovoltaic Defect Detection," in *Solar*, 2024, vol. 4, no. 3: MDPI, pp. 351-386.

[51] J. Li, Z. Su, J. Geng, and Y. Yin, "Real-time detection of steel strip surface defects based on improved YOLO detection network. IFAC-PapersOnLine 51 (21), 76–81 (2018)," ed.

[52] Y. Xu, K. Zhang, and L. Wang, "Metal surface defect detection using modified YOLO," *Algorithms,* vol. 14, no. 9, p. 257, 2021.

[53] M. Hatab, H. Malekmohamadi, and A. Amira, "Surface defect detection using YOLO network," in *Intelligent Systems and Applications: Proceedings of the 2020 Intelligent Systems Conference (IntelliSys) Volume 1*, 2021: Springer, pp. 505-515.

[54] M. Ju, C. Ding, D. Zhang, and Y. J. Guo, "Gamma-correction-based visibility restoration for single hazy images," *IEEE Signal Processing Letters,* vol. 25, no. 7, pp. 1084-1088, 2018.

[55] K. Somasundaram and P. Kalavathi, "Medical image contrast enhancement based on gamma correction," *Int J Knowl Manag e-learning,* vol. 3, no. 1, pp. 15-18, 2011.

[56] S.-C. Huang, F.-C. Cheng, and Y.-S. Chiu, "Efficient contrast enhancement using adaptive gamma correction with weighting distribution," *IEEE transactions on image processing,* vol. 22, no. 3, pp. 1032-1041, 2012.

[57] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "Ghostnet: More features from cheap operations," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 1580-1589.

[58] G. Singh and A. Mittal, "Various image enhancement techniques-a critical review," *International Journal of Innovation and Scientific Research,* vol. 10, no. 2, pp. 267-274, 2014.

[59] J. Wang *et al.*, "Toward surface defect detection in electronics manufacturing by an accurate and lightweight YOLO-style object detector," *Scientific Reports,* vol. 13, no. 1, p. 7062, 2023.

[60] K. Han, Y. Wang, J. Guo, and E. Wu, "ParameterNet: Parameters Are All You Need for Large-scale Visual Pretraining of Mobile Networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 15751-15761.

[61] X. Guan, S. Jian, P. Hongda, Z. Zhiguo, and G. Haibin, "An image enhancement method based on gamma correction," in *2009 Second international symposium on computational intelligence and design*, 2009, vol. 1: IEEE, pp. 60-63.

[62] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.

[63] Y. Sun, S. N. Ravi, and V. Singh, "Adaptive activation thresholding: Dynamic routing type behavior for interpretability in convolutional neural networks," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 4938-4947.

[64] F. Wu, A. Fan, A. Baevski, Y. N. Dauphin, and M. Auli, "Pay less attention with lightweight and dynamic convolutions," *arXiv preprint arXiv:1901.10430,* 2019.

[65] C. Tian, X. Zhang, J. Ren, W. Zuo, Y. Zhang, and C.-W. Lin, "A Heterogeneous Dynamic Convolutional Neural Network for Image Super-resolution," *arXiv preprint arXiv:2402.15704,* 2024.
[66] A. C. Marreiros, J. Daunizeau, S. J. Kiebel, and K. J. Friston, "Population dynamics: variance and the sigmoid activation function," *Neuroimage,* vol. 42, no. 1, pp. 147-157, 2008.
[67] T. Chen and H. Chen, "Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical systems," *IEEE transactions on neural networks,* vol. 6, no. 4, pp. 911-917, 1995.
[68] S. Dorafshan, R. J. Thomas, and M. Maguire, "SDNET2018: An annotated image dataset for non-contact concrete crack detection using deep convolutional neural networks," *Data in brief,* vol. 21, pp. 1664-1668, 2018.
[69] A. Saberironaghi, J. Ren, and M. El-Gindy, "Defect detection methods for industrial products using deep learning techniques: A review," *Algorithms,* vol. 16, no. 2, p. 95, 2023.
[70] M. R. Jahanshahi, J. S. Kelly, S. F. Masri, and G. S. Sukhatme, "A survey and evaluation of promising approaches for automatic image-based defect detection of bridge structures," *Structure and Infrastructure Engineering,* vol. 5, no. 6, pp. 455-486, 2009.
[71] Ž. Vujović, "Classification model evaluation metrics," *International Journal of Advanced Computer Science and Applications,* vol. 12, no. 6, pp. 599-606, 2021.
[72] M. Hossin and M. N. Sulaiman, "A review on evaluation metrics for data classification evaluations," *International journal of data mining & knowledge management process,* vol. 5, no. 2, p. 1, 2015.
[73] H. Dalianis and H. Dalianis, "Evaluation metrics and evaluation," *Clinical Text Mining: secondary use of electronic patient records,* pp. 45-53, 2018.
[74] G. Naidu, T. Zuva, and E. M. Sibanda, "A review of evaluation metrics in machine learning algorithms," in *Computer Science On-line Conference*, 2023: Springer, pp. 15-25.
[75] J. Zhou, A. H. Gandomi, F. Chen, and A. Holzinger, "Evaluating the quality of machine learning explanations: A survey on methods and metrics," *Electronics,* vol. 10, no. 5, p. 593, 2021.
[76] J. Fürnkranz and P. A. Flach, "An analysis of rule evaluation metrics," in *Proceedings of the 20th international conference on machine learning (ICML-03)*, 2003, pp. 202-209.
[77] Y. Wang and X. Wang, "A unified study of machine learning explanation evaluation metrics," *arXiv preprint arXiv:2203.14265,* 2022.
[78] A. C. Bovik and S. T. Acton, "Basic linear filtering with application to image enhancement," in *The Essential Guide to Image Processing*: Elsevier, 2009, pp. 225-239.
[79] A. M. Reza, "Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement," *Journal of VLSI signal processing systems for signal, image and video technology,* vol. 38, pp. 35-44, 2004.
[80] H. Li, Y. Zhang, Y. Zhang, H. Li, and L. Sang, "DCNv3: Towards Next Generation Deep Cross Network for CTR Prediction," *arXiv preprint arXiv:2407.13349,* 2024.
[81] T. Wu, S. Tang, R. Zhang, J. Cao, and Y. Zhang, "Cgnet: A light-weight context guided network for semantic segmentation," *IEEE Transactions on Image Processing,* vol. 30, pp. 1169-1179, 2020.

[82] X. Ding, X. Zhang, J. Han, and G. Ding, "Diverse branch block: Building a convolution as an inception-like unit," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 10886-10895.
[83] J. Chen *et al.*, "Run, don't walk: chasing higher FLOPS for faster neural networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 12021-12031.
[84] Y. Chen, X. Yuan, R. Wu, J. Wang, Q. Hou, and M.-M. Cheng, "Yolo-ms: rethinking multi-scale representation learning for real-time object detection," *arXiv preprint arXiv:2308.05480,* 2023.

PROJECT MANAGEMENT

During the actual conduct of this study, a number of unplanned circumstances arose that slightly altered the original progress, and the following is a schedule of the planned process:

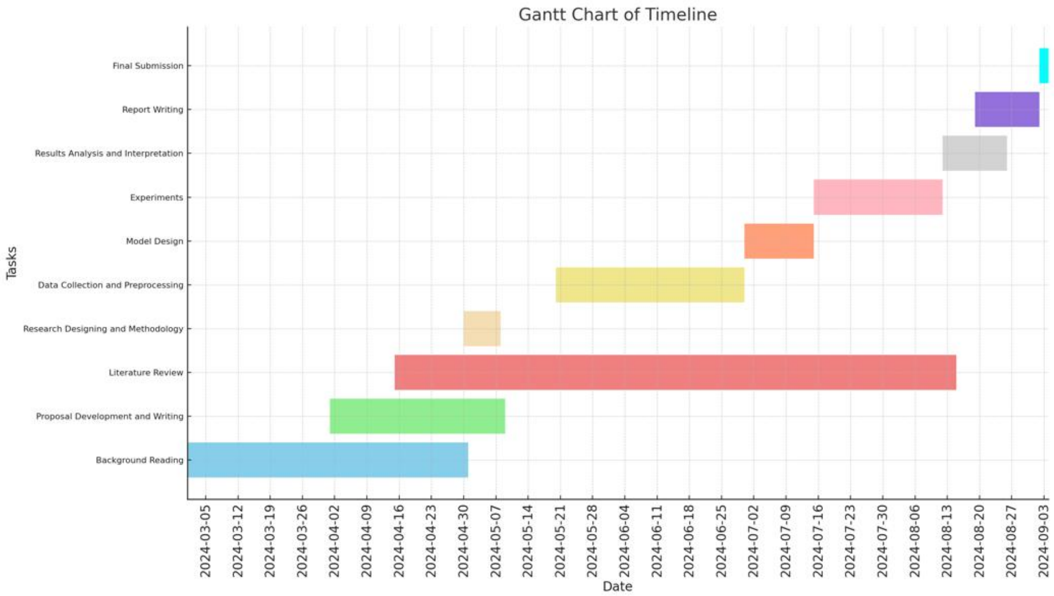

Fig. 5. Planned schedule

Here is the actual schedule:

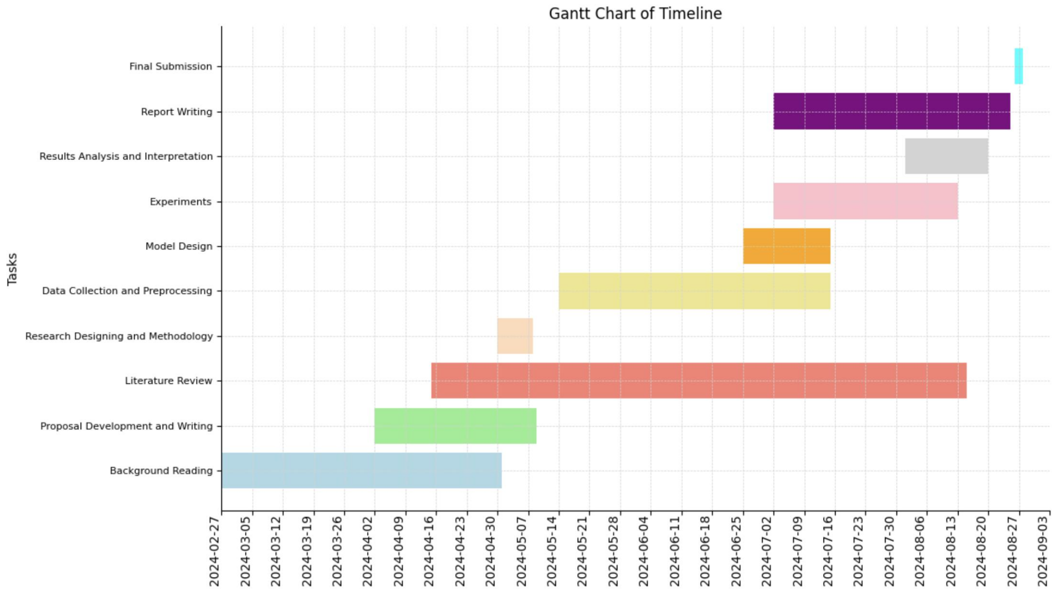

Fig. 6. Actual time schedule

First of all, I changed the title from underwater abutment surface defects detection to abutment surface defects detection, this is because I did not collect the appropriate underwater abutment surface defects dataset, I tried to contact a lot of scholars in the related fields, but I did not get the dataset I wanted, which took me a certain amount of time, in terms of the progress of the data collection, I was planning to collect the data in a short period of time, and then go into the experimentation stage, however, there are not many public datasets related to this study, let alone the labelled dataset. phase, however, there are not many publicly available datasets related to this study, let alone labelled datasets, and it took me a long time to find the datasets that are currently being used in this study and to select and manually label the dataset contents. However, I did not let this process delay my progress as it required more experimentation, and while labelling the data, I have started to design the whole model architecture and think about the experimental design.

Secondly, more time was spent on model design than originally planned, as I did a lot of literature reading on choosing the BASELINE model and inserting the modules, and tried out all these different combinations.

At the same time, the process of experimentation was longer than originally planned, and many combinations of the image enhancement module and the newly inserted module were attempted, and at the same time, during this period, there were three other days in which it was not possible to connect to the server and carry out the experiment due to network reasons, which delayed the progress of the experimentation.

In the end, I started the process early in order to be able to write a higher quality report, especially by writing the first and second parts of the report in advance.

In regard to the data collection phase, the initial plan was to rapidly acquire datasets and advance to the experimental phase. However, the scarcity of publicly available datasets relevant to this research, particularly labeled ones, necessitated extensive searches to identify suitable datasets for this study. Once identified, these datasets required careful selection and manual annotation. Despite these challenges, these preparatory tasks did not impede the overall project timeline, as experimental design and model architecture planning were concurrently executed during the data labeling process.

Furthermore, the model design phase extended beyond initial estimates due to comprehensive literature reviews conducted to select an appropriate baseline model and integrate various modules. This phase involved testing multiple combinations to optimize the model configuration.

The experimental phase also exceeded planned durations, primarily due to the iterative testing of different configurations involving the image enhancement and newly integrated modules. Additionally, there were three instances of server connectivity issues due to network disruptions, further delaying experimental progress.

To ensure the submission of a high-quality report, I initiated the writing phase early, drafting the first and second sections of the report in advance to accommodate these delays.

ETHICS FORM

APPLICATION FOR ETHICAL APPROVAL

In order for research to result in benefit and minimise risk of harm, it must be conducted ethically.

The University follows the OECD Frascati manual definition of **research activity**: "creative work undertaken on a systematic basis in order to increase the stock of knowledge, including knowledge of man, culture and society, and the use of this stock of knowledge to devise new applications". As such this covers activities undertaken by members of staff, postgraduate research students, and both taught postgraduate and undergraduate students working on dissertations/projects.

The individual undertaking the research activity is known as the "principal researcher".

This form must be completed and approved prior to undertaking any research activity.

SECTION A: About You (Principal Researcher)

1	Full Name:	Hanzhe Cai
2	Student Number:	2304714
3	Email address:	2304714@student.uwtsd.ac.uk
4	Programme of Study:	Software Engineering and Artificial Intelligence
5	Director of Studies/Supervisor:	

SECTION B: Internal and External Ethical Guidance Materials

	Please list the core ethical guidance documents that have been referred to during the completion of this form (including any discipline-specific codes of research ethics, location-specific dodes of research ethics, and also any specific ethical guidance relating to the proposed methodology). Please tick to confirm that your research proposal adheres to these codes and guidelines. You may add rows to this table if needed.	
1	**UWTSD Research Ethics & Integrity Code of Practice**	☒
2	**UWTSD Research Data Management Policy**	☒
3		☐

SECTION C: Details of Research Activity

1	Indicative title:	Development of Automatic Detection of Surface Defects In Underwater Bridge Piers		
2	Proposed start date:	March 10th, 2024	Proposed end date:	September 10th, 2024.
	Introduction to the Research (maximum 300 words in each section) **Ensure that you write for a Non-Specialist Audience when outlining your response to the three points below:** • *Purpose of Research Activity* • *Proposed Research Question* • *Aims of Research Activity* • *Objectives of Research Activity* Demonstrate, briefly, how **Existing Research** has informed the proposed activity and explain			

	• *What the research activity will add to the body of knowledge* • *How it addresses an area of importance.*
3	**Purpose of Research Activity** To develop an innovative method for detecting surface defects on underwater bridge piers, combining advanced target detection techniques and image enhancement algorithms, in order to address the limitations of traditional methods applied in complex underwater environments, and thus improve the accuracy and efficiency of the detection. (this box should expand as you type)
4	**Research Question** How to improve the accuracy and efficiency of underwater bridge abutment surface defect detection? This question is triggered by the inadequacy of automated underwater inspection systems due to the poor quality of current underwater images and the lack of effective defect detection algorithms. (this box should expand as you type)
5	**Aims of Research Activity** The aim of this research is to significantly improve the identification accuracy and operational efficiency of automated underwater inspection systems through the integration and optimisation of technologies to provide technical support for bridge safety monitoring. (this box should expand as you type)
6	**Objectives of Research Activity** Evaluate Multiple Underwater Image Enhancement Techniques: Evaluate several underwater image enhancement techniques in a controlled laboratory environment to identify the most effective solution. Integrate improved YOLO framework for defect detection: Integrate selected image enhancement techniques with the enhanced YOLO framework with the expectation of significantly improving recognition accuracy and operational efficiency. Experimental comparison and selection of the most suitable technological solution: based on the requirements of a specific dataset, experimentally compare multiple techniques and select the single technique or combination of techniques that best adapts to the needs of the study. (this box should expand as you type)
	Proposed data collection methods (maximum 600 words) Provide a brief summary of all the methods that **may** be used in the research activity to collect data, making it clear what specific techniques may be used. If methods other than those listed in this section are deemed appropriate later, additional ethical approval for those methods will be needed. You do not need to justify the methods here, but should instead describe how you intend to collect the data necessary for you to complete your project.
7	To overcome the challenge of data scarcity in the detection of surface defects in underwater bridge piers, an innovative approach can be taken to pre-train target detection models using existing datasets of surface defects in pavements, industrial materials and products. Although these datasets do not cover exactly the same areas as the underwater environment, many defect types, such as cracks, corrosion and spalling, are prevalent in construction materials such as bridge piers. By fine-tuning the model, we can adapt it to the specific challenges of the underwater environment. This approach not only extends the diversity of training samples for the model, but also improves the accuracy and reliability of the model in recognising a wide range of defects, thus effectively solving the problem of insufficient dedicated underwater defect datasets. In addition, this cross-domain training strategy provides us with a cost-effective and easy-to-implement solution that makes the development and optimisation of automated underwater defect detection systems more feasible. (this box should expand as you type)

SECTION D: Scope of Research Activity

	Will the research activity include:	YES	NO
1	Use of a questionnaire or similar research instrument?	☐	☒

2	Use of interviews?	☐	☒
3	Use of focus groups?	☐	☒
4	Use of participant diaries?	☐	☒
5	Use of video or audio recording?	☐	☒
6	Use of computer-generated log files?	☐	☒
7	Participant observation with their knowledge?	☐	☒
8	Participant observation without their knowledge?	☐	☒
9	Access to personal or confidential information without the participants' specific consent?	☐	☒
10	Administration of any questions, test stimuli, presentation that may be experienced as physically, mentally or emotionally harmful / offensive?	☐	☒
11	Performance of any acts which may cause embarrassment or affect self-esteem?	☐	☒
12	Investigation of participants involved in illegal activities?	☐	☒
13	Use of procedures that involve deception?	☐	☒
14	Administration of any substance, agent or placebo?	☐	☒
15	Working with live vertebrate animals?	☐	☒
16	Procedures that may have a negative impact on the environment?	☐	☒
17	Other primary data collection methods. Please indicate the type of data collection method(s) below.	☐	☒
	Details of any other primary data collection method: (this box should expand as you type)		

If you have ticked NO to every question then the research activity is (ethically) low risk and you may skip section E and continue to section F.

If YES to any question, then no research activity should be undertaken until full ethical approval has been obtained.

SECTION E: Intended Participants

	Who are the intended participants:	**YES**	**NO**
1	Students or staff at the University?	☐	☒
2	Adults (over the age of 18 and competent to give consent)?	☐	☒
3	Vulnerable adults?	☐	☒
4	Children and Young People under the age of 18? (Consent from Parent, Carer or Guardian will be required)	☐	☒
5	Prisoners?	☐	☒
6	Young offenders?	☐	☒
7	Those who could be considered to have a particularly dependent relationship with the investigator or a gatekeeper?	☐	☒
8	People engaged in illegal activities?	☐	☒
9	Others. Please indicate the participants below, and specifically any group who may be unable to give consent.	☐	☒

	Details of any other participant groups: Complete this only if your participants cannot give consent. This includes animals (this box should expand as you type)		

	Participant numbers and source Provide an estimate of the expected number of participants. How will you identify participants and how will they be recruited?	
10	How many participants are expected?	Ballpark figures are fine, but make sure that you explain how you will identify and contact your participants. *(this box should expand as you type)*
11	Who will the participants be?	*(this box should expand as you type)*
12	How will you identify the participants?	*(this box should expand as you type)*

	Information for participants:	YES	NO	N/A
13	Will you describe the main research procedures to participants in advance, so that they are informed about what to expect?	☒	☐	☐
14	Will you tell participants that their participation is voluntary?	☒	☐	☐
15	Will you obtain written consent for participation?	☒	☐	☐
16	Will you explain to participants that refusal to participate in the research will not affect their treatment or education (if relevant)?	☒	☐	☐
17	If the research is observational, will you ask participants for their consent to being observed?	☒	☐	☐
18	Will you tell participants that they may withdraw from the research at any time and for any reason?	☒	☐	☐
19	With questionnaires, will you give participants the option of omitting questions they do not want to answer?	☒	☐	☐
20	Will you tell participants that their data will be treated with full confidentiality and that, if published, it will not be identifiable as theirs?	☒	☐	☐
21	Will you debrief participants at the end of their participation, in a way appropriate to the type of research undertaken?	☒	☐	☐
22	If NO to any of above questions, please give an explanation			
	You should be able to tick YES for all of these questions. If not, then explain why not in this box. *(this box should expand as you type)*			

	Information for participants:	YES	NO	N/A
24	Will participants be paid?	☐	☒	☐
25	Is specialist electrical or other equipment to be used with participants?	☐	☒	☐
26	Are there any financial or other interests to the investigator or University arising from this study?	☐	☒	☐
27	Will the research activity involve deliberately misleading participants in any way, or the partial or full concealment of the specific study aims?	☐	☒	☐

28	If YES to any question, please provide full details
	You should be able to tick NO for most of these questions. For any cases that you have ticked YES then provide details in this box. If you are using cameras/voice recorders to record interviews then please state that in this box. *(this box should expand as you type)*

SECTION F: Anticipated Risks

	Outline any anticipated risks that may adversely affect any of the participants, the researchers and/or the University, and the steps that will be taken to address them.	
1	**Risks to participants** For example: sector-specific health & safety, emotional distress, financial disclosure, physical harm, transfer of personal data, sensitive organisational information. If you have identified in section D that there are no participants then enter N/A and go skip to question 3.	
	Risk to participants: There are always risks. Do not write N/A unless you have no participants. *(this box should expand as you type)*	*How you will mitigate the risk to participants:* *(this box should expand as you type)*
2	If research activity may include sensitive, embarrassing or upsetting topics (e.g. sexual activity, drug use) or issues likely to disclose information requiring further action (e.g. criminal activity), give details of the procedures to deal with these issues, including any support/advice (e.g. helpline numbers) to be offered to participants. Note that where applicable, consent procedures should make it clear that if something potentially or actually illegal is discovered in the course of a project, it may need to be disclosed to the proper authorities	
	(this box should expand as you type)	
3	**Risks to the investigator** For example: personal health & safety, physical harm, emotional distress, risk of accusation of harm/impropriety, conflict of interest	
	Risk to the investigator: There are always risks. Do not write NA. *(this box should expand as you type)*	*How you will mitigate the risk to the investigator:* *(this box should expand as you type)*
4	**University/institutional risks** For example: adverse publicity, financial loss, data protection	
	Risk to the University: There are always risks. Do not write NA. *(this box should expand as you type)*	*How you will mitigate the risk to the University:* *(this box should expand as you type)*
5	**Environmental risks** For example: accidental spillage of pollutants, damage to local ecosystems	
	Risk to the environment: You may write NA if there are no research-related environmental risks. Driving to the university does not count as a risk. *(this box should expand as you type)*	*How you will mitigate the risk to environment:* *(this box should expand as you type)*

SECTION G: Feedback, Consent and Confidentiality

If you have identified in section D that there are no participants then enter skip this section and continue to section H.

1	**Feedback** What de-briefing and feedback will be provided to participants, how will this be done and when?
	You don't need to email your participants with your final report. A good alternative is to set up an email address that they will be able to contact for further details or results. *(this box should expand as you type)*
2	**Informed consent** Describe the arrangements to inform potential participants, before providing consent, of what is involved in participating. Describe the arrangements for participants to provide full consent before data collection begins. If gaining consent in this way is inappropriate, explain how consent will be obtained and recorded in accordance with prevailing data protection legislation.
	If you are using a paper questionnaire then you should have the participants sign an appropriate consent form. These forms will count as personal data and should be noted as such in section J. If you are using an online questionnaire, then you should have a screen before the questions start that acts as a consent form, informing participants that by clicking on the NEXT button they are providing consent. *(this box should expand as you type)*
3	**Confidentiality / Anonymity** Set out how anonymity of participants and confidentiality will be ensured in any outputs. If anonymity is not being offered, explain why this is the case.
	Do not collect names unless you really need them. Do not name participants or organisations in any research publications (including the thesis) without their explicit permission. *(this box should expand as you type)*

SECTION H: Data Protection and Storage

	Does the research activity involve personal data (as defined by the General Data Protection Regulation 2016 "GDPR" and the Data Protection Act 2018 "DPA")?	**YES**	**NO**
1	***"Personal data"*** *means any information relating to an identified or identifiable natural person ('data subject'). An identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person. Any video or audio recordings of participants is considered to be personal data.*	☐	☒
	If YES, provide a description of the data and explain why this data needs to be collected:		
2	This includes audio/video data of participants, but can also include IP addresses and usernames. Names, addresses and emails also count, as do consent forms. *(this box should expand as you type)*		
	Does it involve special category data (as defined by the GDPR)?	**YES**	**NO**
3	***"Special category data"*** *means sensitive personal data consisting of information as to the data subjects' –* *(a) racial or ethnic origin,* *(b) political opinions,* *(c) religious beliefs or other beliefs of a similar nature,* *(d) membership of a trade union (within the meaning of the Trade Union and Labour Relations (Consolidation) Act 1992),* *(e) physical or mental health or condition,* *(f) sexual life,* *(g) genetics,* *(h) biometric data (as used for ID purposes),*	☐	☒
	If YES, provide a description of the special category data and explain why this data needs to be		

	collected:		
4	What counts as 'sensitive' will differ between cultures. Any information on behaviour that is not in accordance with cultural norms would count as sensitive personal data. *(this box should expand as you type)*		
	Will data from the research activity (collected data, drafts of the thesis, or materials for publication) be stored in any of the following ways?	**YES**	**NO**
5	Manual files (i.e. in paper form)?	☐	☒
6	University computers?	☐	☒
7	Private company computers?	☐	☒
8	Home or other personal computers?	☐	☒
9	Laptop computers/ CDs/ Portable disk-drives/ memory sticks?	☐	☒
10	"Cloud" storage or websites?	☒	☐
11	Other – specify:	☐	☒
12	For all stored data, explain the measures in place to ensure the security of the data collected, data confidentiality, including details of backup procedures, password protection, encryption, anonymisation and pseudonymisation:		
	If possible, save your data on computers that are secure and regularly backed up. Many cloud services only provide GDPR-compliant storage for business customers. An example of suitable text is given below. After I collect data to create a dataset, I store it on my hard drive in accordance with confidentiality requirements and do not share it with others or make it available online. *(this box should expand as you type)*		

	Data Protection		
	Will the research activity involve any of the following activities:	**YES**	**NO**
13	Electronic transfer of data in any form?	☐	☒
14	Sharing of data with others at the University outside of the immediate research team?	☐	☒
15	Sharing of data with other organisations?	☐	☒
16	Export of data outside the UK or importing of data from outside the UK?	☒	☐
17	Use of personal addresses, postcodes, faxes, emails or telephone numbers?	☐	☒
18	Publication of data that might allow identification of individuals?	☐	☒
19	If YES to any question, please provide full details, explaining how this will be conducted in accordance with the GDPR and Data Protection Act (2018) (and/or any international equivalent):		
	This includes data such as drafts of your thesis as well as experimental or survey data. An example of suitable text is given below. The data collected came from different countries and regions so as to ensure that the success of the study can be applied to the different building standards in each country and region. *(this box should expand as you type)*		
20	List all who will have access to the data generated by the research activity:		

	Myself *(this box should expand as you type)*	
21	List who will have control of, and act as custodian(s) for, data generated by the research activity:	
	Myself *(this box should expand as you type)*	
22	Give details of data storage arrangements, including security measures in place to protect the data, where data will be stored, how long for, and in what form.	
	All data will be encrypted and kept in password protected cloud storage on the University Office 365 system which will not be shared. Any USB sticks used to store or transfer data will be password protected, and will be reformatted at the end of the project in order to destroy the data. The data will be stored until the completion of the project and then deleted. *(this box should expand as you type)*	
22	Confirm that you have read the UWTSD guidance on data management (see https://www.uwtsd.ac.uk/library/research-data-management/)	☒
23	Confirm that you are aware that you need to keep all data until after your research has completed or the end of your funding	☒

SECTION I: Declaration

	The information which I have provided is correct and complete to the best of my knowledge. I have attempted to identify any risks and issues related to the research activity and acknowledge my obligations and the rights of the participants. In submitting this application I hereby confirm that I undertake to ensure that the above named research activity will meet the University's Research Ethics and Integrity Code of Practice which is published on the website: https://www.uwtsd.ac.uk/research/research-ethics/		
1	**Signature of applicant:**	Hanzhe Cai	**Date:** 10-May-2024
2	Director of Studies/Supervisor:		**Date:**
3	Signature:		

FOR INTERNAL USE ONLY:

	Ethical approval given		
1	**Signature of assessor:**		**Date:**
2	Name:		
3	Role:		

LOGBOOK

Date Range	Task	Description
2024-02-27	Background Reading	Engaged in extensive background reading to establish a solid foundation for the research. Reviewed scholarly articles, books, and previous studies related to defect detection in bridge piers, focusing on methodologies and technologies previously used. This foundational reading helped identify gaps in current research and shaped the direction of the study by highlighting innovative approaches and potential areas for technological advancement in the field of structural health monitoring.
2024-03-05 to 03-12	Proposal Development and Writing	Developed a comprehensive research proposal that outlined the scope, objectives, and methods of the study. The proposal defined specific goals such as enhancing defect detection accuracy using the YOLOV9 model integrated with advanced image enhancement techniques. It also detailed the methodology, including the types of data to be collected, the experimental design, and the analytical techniques to be used for evaluating the model's performance.
2024-03-13 to 03-26	Literature Review	I I conducted an in-depth review of a broad spectrum of literature encompassing defect detection, image enhancement technologies, and target detection fields. This included foundational theoretical research through to cutting-edge practical applications, such as the use of deep learning algorithms for image analysis and enhancement, and various machine learning models applied in object detection. Special attention was given to studies that provided innovative approaches or improvements to existing techniques, such as employing Convolutional Neural Networks (CNN) to identify and classify defects in images, or developing new image preprocessing algorithms to enhance the accuracy and efficiency of detection systems. Additionally, I reviewed numerous articles on advanced object detection frameworks, such as the YOLO (You Only Look Once) series, exploring their balance between processing speed and accuracy, and their practical application effectiveness in real bridge defect detection scenarios.
2024-03-27 to 04-09	Research Design and Methodology	Based on extensive reading and in-depth analysis of a multitude of documents, I designed a detailed set of research steps and an overall plan. These documents span foundational theories to applied technologies, particularly the latest advancements in image processing, defects detection. Integrating these

Date Range	Task	Description
		research outcomes, I identified key areas and potential innovations for my study. My research plan includes several critical steps: Firstly, experimental design to ensure scientific rigor and operability; secondly, data collection and preprocessing, establishing stringent standards to ensure data quality and experiment effectiveness; thirdly, model development and testing, selecting and optimizing algorithms and models suited to solving practical problems; and finally, results analysis and validation, where experimental outcomes are compared against theoretical expectations to verify the correctness and practicality of the research hypotheses. The formulation of this comprehensive plan aims to ensure the systematic nature of the research process and the reliability of the results, thereby contributing new theoretical and practical knowledge to the advancement of bridge defect detection technology.
2024-04-10 to 04-23	Data Collection and Preprocessing	To better support my research, I conducted an extensive search and organized currently available datasets related to surface defects on bridge piers. These datasets primarily come from various internationally recognized databases and research institutions, encompassing everything from small-scale experimental data to large-scale field-collected data. After a detailed analysis of these datasets, I carefully selected those that include well-defined defect types, high-quality images, and precise annotations. This selection process considered not only the representativeness and diversity of the data but also its practicality and reliability to ensure that the chosen datasets could effectively support the development and testing of image processing and defect detection algorithms.
2024-04-24 to 05-07	Model Design	To delve deeper into and optimize the detection methods for surface defects on bridge piers, I first conducted a comprehensive summary of the currently popular defect detection techniques. These methods include those based on traditional image processing technologies as well as those leveraging deep learning, each possessing unique advantages and limitations. By comparing the performance of these methods in terms of accuracy, speed, and resource consumption, I selected the most suitable object detection framework for our research needs. After selecting the framework, I designed entirely new modules tailored to the specific requirements of our

Date Range	Task	Description
		study to enhance the performance of the target detection framework. These custom modules are aimed at addressing specific issues encountered by traditional models when dealing with complex defects on bridge surfaces, such as variations in lighting, differences in defect sizes, and background noise. The new modules include advanced image preprocessing features to improve the quality of input images and enhanced feature extraction mechanisms to more accurately identify and classify various defects.
2024-05-08 to 05-21	Experiments	To ensure optimal performance of our bridge pier surface defect detection system, I systematically conducted experimental comparisons on a carefully curated dataset using mainstream image enhancement methods and object detection models. These datasets were designed to simulate defect characteristics under various environmental conditions, providing an ideal platform for a comprehensive assessment of model performance. In the experiments, I first applied several different image enhancement techniques, such as contrast adjustment, sharpening, and noise suppression, to improve image quality and enhance model recognition capabilities. Subsequently, I selected several widely used object detection models, such as YOLO, SSD, and Faster R-CNN, and compared their effectiveness and efficiency in handling defects of varying complexity. Each model was evaluated in detail for its accuracy, detection speed, and computational resource consumption to determine the most suitable model for our project needs. Additionally, to further enhance model performance, I attempted to integrate various functional modules into the chosen object detection framework. These modules included new feature extraction layers, more complex classifiers, and task-specific optimization algorithms. Through this approach, I was able to test the impact of each module on the overall performance of the detection system, thereby selecting the most effective combination. Ultimately, through this series of exhaustive experiments and comparative analyses, I not only selected the most suitable image enhancement techniques and object detection models for bridge pier surface defect detection but also optimized the

Date Range	Task	Description
		architecture of the entire detection system, ensuring its efficiency and accuracy in practical applications.
2024-05-22 to 06-04	Results Analysis and Interpretation	After completing a series of experiments with various image enhancement techniques and object detection models, I conducted a detailed comparison of the results to deeply analyze the performance and efficiency of each method in detecting surface defects on bridge piers. Initially, I statistically processed and visualized the data obtained from the experiments, including key metrics such as the accuracy rates, detection speeds, and resource consumption of each model, thus clearly illustrating the performance differences between the techniques. Next, for those experiments that showed significant performance variances, I delved into potential reasons. For instance, some image enhancement methods might have caused loss of detail due to over-processing, while some object detection models might have responded slowly in real-time tasks due to their depth or complexity of structure. Additionally, I considered the characteristics of the datasets, such as the diversity of images and the complexity of defects, which could also impact the final detection outcomes. Through this series of comparative analyses, I not only identified the strengths and limitations of various approaches but also proposed improvement strategies for specific issues. These included optimizing the image preprocessing workflow, adjusting model parameters, or redesigning certain layers of the model to enhance the overall performance and adaptability of the detection system. Ultimately, these analyses helped me develop a more precise and efficient plan for detecting surface defects on bridge piers.
2024-06-05 to 06-18	Report Writing	As the research project drew to a close, I began composing a detailed research report to comprehensively summarize the experimental process, analysis results, and key findings of the study. This report initially outlined the background and objectives of the research, followed by a thorough description of the methods used, experimental design, data collection, and processing procedures. During the writing process, I elaborately explained the reasons for choosing specific technologies and models, and how these choices impacted the experimental results. The analysis section of the report emphasized the

Date Range	Task	Description
		statistical handling and interpretation of experimental data, visually presenting comparative results through charts and data tables, and highlighting differences between experimental and control groups. Additionally, I critically reflected on the challenges and limitations encountered during the experiments and suggested possible improvements for future research. After completing the first draft, I underwent several rounds of revisions and proofreading to ensure the accuracy, clarity, and fluency of the report's content. During this process, I also sought feedback and suggestions from my supervisor and classmates, who provided invaluable input that helped enhance the academic quality and professionalism of the report. Through this rigorous writing and revision process, a detailed and authoritative research report was ultimately produced, contributing new theoretical and practical insights to the field of bridge pier surface defect detection.
2024-06-19 to 07-03	Final Submission	Reviewing and submitting the final report