



PRIFYSGOL CYMRU
Y Drindod Dewi Sant
UNIVERSITY OF WALES
Trinity Saint David
SWANSEA - ABERTAWE



ENHANCING TRAFFIC SIGN RECOGNITION WITH OPTIMIZED YOLOV5

Shuo Zhang

2304515

Supervisor: Divya Chidambaram

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

September 2024

Declaration of Originality

I, Shuo Zhang 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 : Shuo Zhang

Date : 9.9.2024

FORM OF CONSENT

I, Shuo Zhang hereby consent that my Project, submitted in candidature for the degree NAME, 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 : Shuo Zhang

Date : 9.9.2024

Copyright Acknowledgement

I acknowledge that the copyright of this project report, and any product developed as part of the project, belong to University of Wales Trinity Saint David, Swansea.

ABSTRACT

This study aims to improve the performance of the YOLOv5 model for traffic sign detection by introducing a bidirectional feature pyramid network (BiFPN) and weighted cross-parallelisation (Wise-IoU) optimisation. Addressing the challenges of accuracy and real-time performance of traffic sign recognition (TSR) systems, this project firstly reviews the development of TSR techniques, including early threshold segmentation and template matching methods based on RGB colour space, to modern deep learning-based recognition techniques. The research focuses on analysing the evolution of the YOLO algorithm and its application to traffic sign recognition. This project adopts a hybrid research strategy that combines deductive and inductive approaches to formulate hypotheses through a systematic literature review and experimentally validate them using the TT100K dataset. The experimental results show that the optimised YOLOv5 model shows significant improvement in several key performance metrics, including precision, recall and mean average precision (mAP). In particular, the optimised model performs well in small target detection, which is crucial for improving traffic safety and efficiency in autonomous and assisted driving systems.

In addition, the computational efficiency of the model is also considered in this study, which makes the optimised model more suitable for deployment in mobile platforms and resource-constrained environments by reducing the amount of computation while maintaining high accuracy. The results of the study not only demonstrate the effectiveness of the proposed optimisation strategy, but also provide new directions for future applications in ITS, especially in solving the category imbalance problem and exploring the model generalisation capability.

Keywords: traffic sign recognition, YOLOv5, BiFPN, Wise-IoU, Deep Learning, Computer Vision

TABLE OF CONTENT

Copyright Acknowledgement.....	iii
CHAPTER 1 INTRODUCTION.....	8
1.1 INTRODUCTION	8
1.2 AIM	9
1.3 OBJECTIVES	10
1.4 STRUCTURE OF THE PROJECT	10
CHAPTER 2 - RESEARCH & REVIEW OF LITERATURE.....	12
2.1 INTRODUCTION	12
2.2 LITERATURE REVIEW	14
2.3 METHODOLOGY	17
2.3.1 YoloV5	17
2.3.2 BIFPN (Bi-directional Feature Pyramid Network)	19
2.3.3 Wise-IoU (Weighted Intersection over Union).....	21
2.4 CONCLUSION.....	24
CHAPTER 3 METHODOLOGY.....	25
3.1 RESEARCH METHODOLOGIES	25
3.1.1 PHILOSOPHIES	25
3.1.2 Strategy	25
3.1.3 Approaches.....	27
3.1.4 Validity, Reliability & Generalisability	33
3.1.5 Data Evaluation Methods.....	36
3.2 DEVELOPMENT METHODOLOGY	40
3.3 RESOURCES.....	41
3.3.1 Hardware Resources.....	42
3.3.2 Software Resources	42
CHAPTER 4 - DESIGN & IMPLEMENTATION	44
4.1 INTRODUCTION	44
4.2 PROBLEM FORMULATION	44
4.3 DESIGN.....	45
4.3.1 System Architecture Overview.....	45
4.3.2 Component Design.....	47
4.4 IMPLEMENTATION.....	47

4.4.1 Implementation Overview	48
4.4.2 Detailed Implementation Steps	48
4.5 CONCLUSION.....	50
CHAPTER 5 – TESTING AND EVALUATION	51
5.1 LOSS FUNCTION COMPARISON	51
5.2 COMPARISON OF DETECTION MODELS.....	52
5.3 ACTUAL DETECTED PERFORMANCE.....	53
CHAPTER 6 CONCLUSIONS AND RECOMMENDATIONS.....	59
6.1 SUMMARY	59
6.2 CONCLUSION.....	59
6.3 RECOMMENDATIONS FOR FURTHER RESEARCH.....	60
CHAPTER 7 – REFLECTIONS	62
CHAPTER 8 – REFERENCES	63
APPENDICES 67	
PROJECT MANAGEMENT	67
ETHICS FORM 70	
LOGBOOK 77	
GLOSSARY 80	

LIST OF FIGURES

Fig 1: Types of Traffic Signs	12
Fig 2: Complex Backgrounds on Image	13
Fig 3: Sign Obscured on Image.....	14
Fig 4: Damage to Sign on Image	14
Fig 5: Yolov5 Structure	18
Fig 6: The Structure of PAN-FPN	20
Fig 7: The Structure of Bi-directional Feature Pyramid Network[26]	20
Fig 8: TT100k Dataset	36
Fig 9: Optimized Structure.....	47
Fig 10: The Changing of Box_loss.....	51

Fig 11: Actual Detected Performance	54
Fig 12 Confusion Matrix of the Optimized Model	55

LIST OF TABLES

Table 1: Comparison of Detection Models	52
Table 2: Performance metrics for different classes	55
Table 3 The Original Gantt chart	67
Table 4 The Actual Gantt chart.....	68

CHAPTER 1 INTRODUCTION

1.1 INTRODUCTION

Cars have become an indispensable part of modern life, greatly facilitating people's travel. However, with the increase in the number of vehicles, urban traffic congestion and the incidence of traffic accidents have also risen. One of the main causes of traffic accidents and congestion is the non-compliance with traffic regulations by drivers, including the neglect of traffic signs. Therefore, it is imperative for drivers to strictly adhere to traffic laws and to pay timely attention to and understand traffic signs to ensure driving safety.

As society develops, people's reliance on computers has grown stronger, with the expectation that computers can handle increasingly complex tasks and even possess a certain level of logical thinking ability to replace or assist human work. This demand has spurred the rapid development of artificial intelligence technology. Today, artificial intelligence has been widely applied to all aspects of life, greatly improving production efficiency and work efficiency.

Computer vision, as an important branch of artificial intelligence, imitates the human visual cognition mechanism, enabling machines to perceive and understand the visual environment[1]. It includes multiple research directions such as object detection, image classification, and image segmentation. Among them, object detection is a key research focus, with the task of identifying and locating the target of interest in images. In recent years, object detection technology has made significant progress, especially in the field of Traffic Sign Recognition (TSR). TSR systems have a broad application prospect in the fields of autonomous driving and assisted driving, and are of great significance for improving road traffic safety and efficiency.

Traffic signs are important elements that indicate traffic rules, warn of road hazards, and provide traffic information, playing a crucial role in ensuring road traffic safety, maintaining traffic order, and improving traffic efficiency. Vehicles equipped with a

traffic sign recognition system can effectively guide drivers to comply with traffic rules, ensure driving safety, and reduce traffic accidents. With the development of intelligent transportation systems, the detection and recognition of traffic signs not only help to achieve autonomous driving technology and improve commuting efficiency but also enable intelligent monitoring of urban road traffic, timely detection and repair of damaged or missing traffic signs and enhance the efficiency of urban management[2][3].

In recent years, deep learning technology has made significant breakthroughs in fields such as image, speech, and natural language processing, and has become an emerging technology in the field of traffic sign detection and recognition. Deep learning can automatically extract target features, avoiding the shortcomings of manual design, and has good generalization performance. The YOLO algorithm is a standout among them. Therefore, using object detection technology based on deep learning, combined with vehicle control systems, to improve the accuracy and real-time performance of traffic sign detection and recognition has become a hot research topic, with important practical significance and application value[4].

However, there are certain limitations to the recognition accuracy and real-time performance of these methods, especially when there are a large number of types and numbers of traffic signs, the natural environment is harsh (fog, rainstorms, gusty winds, and other extreme weather), and obstacles obscured, signs broken, the collection of photographs blurred, and other factors such as the impact of various factors the accuracy tends to be affected to a greater or lesser extent. Therefore, how to use the existing deep learning-based target detection technology to improve the accuracy and real-time performance of traffic sign detection and recognition has become a hot research issue, which has important practical significance and application value[5].

1.2 AIM

The primary aim of this research is to overcome the existing limitations in traffic sign recognition systems by refining the architecture of the YOLOv5 model. The objective is to significantly enhance the algorithm's accuracy, robustness, and

real-time performance, particularly in diverse and challenging environmental conditions. This study seeks to advance the current state of Traffic Sign Recognition (TSR) by ensuring the improved YOLOv5 model can reliably detect and classify traffic signs, thereby contributing to safer and more efficient road traffic management.

1.3 OBJECTIVES

The objectives are to:

1. Optimize the structure of YOLOv5 to improve traffic sign recognition accuracy under diverse environmental conditions.
2. Evaluate the impact of structural modifications, including changes in network layers and activation functions, on the performance of YOLOv5 and enhancing Traffic Sign Recognition with Optimized YOLOv5: Addressing Environmental Variability for Improved Road Safety.
3. Investigate how different training datasets comprising varied traffic sign images influence the model's recognition capabilities.
4. Assess the enhanced model's performance using metrics such as accuracy, recall, precision, inference speed, and model size.

1.4 STRUCTURE OF THE PROJECT

Chapter one begins with an introduction to the centrality of the automobile in modern life and the resulting transport problems, such as the increase in urban traffic congestion and traffic accidents. The article then discusses the importance of artificial intelligence techniques, especially computer vision, in traffic sign recognition, as well as the breakthroughs in deep learning techniques in the field of image recognition, especially the application of the YOLO algorithm in target detection.

Chapter two is a literature review delves into the development of traffic sign recognition technology, from early threshold segmentation and template matching methods based on RGB colour space to modern deep learning-based recognition techniques. The chapter also analyses the advantages and disadvantages of traditional machine learning algorithms and deep learning algorithms for traffic sign

detection, and discusses in detail the evolution of the YOLO family of algorithms and their application in traffic sign recognition.

Chapter three Research Methodology details the philosophical stance of this research, the research strategy, the methodology used, the potential risks and the data evaluation methods. The objective of this chapter is to ensure that the YOLOv5 model is systematically tested and optimised for accuracy and real-time traffic sign recognition.

Chapter four is Design and Implementation provides a comprehensive overview of the design and implementation of a traffic sign recognition system, starting with a clear problem statement and then detailing the system architecture, component design and implementation strategy. Particular attention is given to the customisation of the YOLOv5 model, including the integration of BiFPN and WIoU to improve the performance of the model under different environmental conditions.

Chapter five is Tests and Evaluations presents the results of testing the optimised YOLOv5 model using the TT100K dataset and compares its performance with several other state-of-the-art target detection models, evaluated against key metrics such as model size, number of parameters, computational complexity, and average accuracy for objects of different sizes.

Chapter six is Conclusions and Recommendations summarises the main findings of the study, highlights the contribution of optimising the YOLOv5 model to improve the accuracy and computational efficiency of traffic sign detection, and suggests directions for future research, such as solving the category imbalance problem and exploring the model generalisation capability.

Chapter seven is reviewing the research process of the whole project, reflects on the goals reached, challenges encountered, and looks forward to new opportunities for future research.

CHAPTER 2 - RESEARCH & REVIEW OF LITERATURE

2.1 INTRODUCTION

Traffic Sign Recognition (TSR) is a computer vision technology, the detection and recognition of traffic signs play a vital role in modern traffic systems, they not only assist drivers to make the right driving decisions, but also play a key role in ensuring driving safety, designed to automatically detect and identify traffic signs, which are usually used to These signs are usually used to indicate the rules of the road, warning messages, indicating the direction, etc[6]. The common signs are shown in the figure below:



Fig 1: Types of Traffic Signs

The yellow triangle mark generally represents warning, the red mark represents prohibition, the blue represents instruction for auxiliary function, in addition to the modern social life, all kinds of instructions, prompts, warning signs are varied, traffic signs of many kinds, characteristic information is rich, the traffic signs of different countries and the meaning of different types of traffic signs, the traffic signs of the detection technology research can be very good for the same kind of detection and identification tasks to provide examples. The study of traffic sign detection technology can provide a good example for similar detection and recognition tasks, so as to better serve the requirements of the intelligent society, and at the same time face a variety of recognition difficulties. The detection and identification of traffic

signs may seem simple in theory because they have uniform shape, colour and font size standards. However, in practice, the task becomes quite challenging due to the interference of various factors. There are a number of difficulties:

1. Complex road background: in the actual road environment, there are many objects similar to traffic signs, such as billboards and decorations with similar colours and shapes, etc., which can interfere with the accurate recognition of traffic signs, Figure 2 is the case of complex background.



Fig 2: Complex Backgrounds on Image

2. Overshadowing: Traffic signs may be obscured by other vehicles, buildings, trees, or in extreme weather conditions such as heavy snow, fog, etc., causing some parts of the sign to become invisible, as shown in the figure for tree obstructions.



Fig 3: Sign Obscured on Image

3. Damage to signs: Traffic signs are subject to defacement, blurring, discolouration and deformation in a complex natural environment. The figure 4 shows the curvature of the sign.

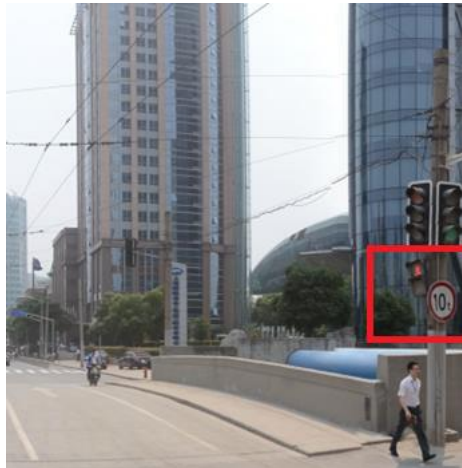


Fig 4: Damage to Sign on Image

In addition to this, the size of a traffic sign varies with the distance between the vehicle and the sign. A small sign at a distance may be difficult to detect, while a large sign at a close distance may be out of the field of view of the camera. The angle and height at which the camera is mounted will affect the viewing angle of the captured traffic sign. An inappropriate viewing angle may lead to changes in the shape and proportion of the sign, making identification more difficult. The resolution, contrast and colour reproduction capability of the camera will affect the image quality, and low-quality image capture will reduce the accuracy of the recognition algorithm. These factors together make the detection and recognition of traffic signs a complex problem in practical applications, requiring comprehensive consideration of a variety of technologies and strategies to improve the accuracy and robustness of recognition[7] And with the development of automatic driving technology, the recognition accuracy and detection speed of traffic signs have put forward higher requirements. In this context, traffic sign detection and recognition techniques have undergone significant evolution.

2.2 LITERATURE REVIEW

Researchers have long been engaged in the study of traffic sign recognition, initially pioneered by Akatsuka et al. in 1987, who employed threshold segmentation and template matching within the RGB colour space[8]. With the advent of machine learning, an increasing array of methods has become available, and during the early stages, traffic sign detection technology primarily depended on traditional machine learning algorithms. These algorithms performed target detection through manually designed features, such as SIFT (Scale-Invariant Feature Transform)[9], HOG (Histogram of Oriented Gradients)[10], and LBP (Local Binary Patterns)[12]. Traditional target detection algorithms typically employed sliding windows for region proposal. This approach has several notable drawbacks. First, sliding windows lack targeted region proposal methods, resulting in excessive computational time during inference and reduced overall efficiency. Second, manually designed features often lack robustness and struggle to adapt to various complex environmental changes. Therefore, in real-time traffic sign detection and recognition systems, traditional machine learning-based algorithms are evidently inadequate.

With the rise of deep learning technologies, deep learning-based target detection algorithms have gradually become mainstream. These algorithms use Convolutional Neural Networks (CNNs) and other deep learning models to automatically extract image features, avoiding the drawbacks of manually designing features as required by traditional methods[13]. Deep learning models can handle complex backgrounds and lighting conditions, significantly improving the accuracy and speed of traffic sign recognition.

Deep learning-based target detection algorithms are mainly categorized into two-stage and one-stage algorithms. Two-stage algorithms, such as R-CNN[14], Fast R-CNN[14], and Faster R-CNN[15], complete target localization and classification in two steps. The first step extracts potential target regions from the image, and the second step classifies and precisely locates the targets within those regions. R-CNN, the pioneering region proposal-based algorithm, was the first to apply deep learning to target detection tasks and quickly became representative of two-stage algorithms. The fundamental idea of R-CNN is to use Selective Search to generate a large number of candidate boxes, resize these candidate boxes to a uniform size, input them into a pre-trained CNN for feature extraction, and finally classify the feature vectors and perform bounding box regression using a Support Vector Machine. Fast

R-CNN improved upon this by performing convolution on the entire image and directly mapping candidate boxes onto the feature map, achieving unified candidate box sizes and avoiding redundant computations. Faster R-CNN further advanced this by introducing the Region Proposal Network (RPN) to generate multiple-sized candidate boxes, reducing the number of candidate boxes, improving their quality, and significantly enhancing target detection accuracy[16].

On the other hand, one-stage target detection algorithms emerged, with YOLO being a representative example. The core idea of YOLO (You Only Look Once) is to treat target detection as a problem of regressing separate prediction boxes and class probabilities[17]. YOLO algorithms, known for their excellent balance of accuracy and speed, have become some of the most outstanding target detection algorithms. In 2016, Joseph Redmon et al. proposed YOLO, achieving real-time performance by processing 45 frames per second[17]. Subsequent research has led to several iterations of the YOLO algorithm, from YOLOv2 to YOLOv9, continuously improving its performance.

As deep learning technologies have evolved, more researchers have applied these algorithms to traffic sign detection and recognition. The LeNet-5 was enhanced convolutional neural network by Zhang in 2020 with adding Inception convolutional modules to extract richer features and incorporating Batch Normalization (BN) layers to prevent overfitting, achieving a recognition rate of 98.51% on the BelgiumTSC dataset[18]. At the same year, Chang Zhen et al.'s work on traffic sign detection was significantly improved in terms of detection accuracy by the application of the two-stage Faster R-CNN network, which leveraged high-quality candidate regions generated by the Region Proposal Network (RPN)[19].

To address these challenges, some researchers have proposed improved YOLO versions and novel multi-scale fusion convolutional neural networks. For example, it was found by Xing in 2023 that fine-tuning the YOLOv5 network structure significantly improved the accuracy of traffic sign detection under varying lighting conditions[20]. A region-of-interest approach with YOLO to accelerate high-definition traffic image processing for real-time automotive applications was introduced by Aleksej Avramović which improved performance and significantly reduced

processing time while maintaining accuracy comparable to state-of-the-art methods[21].

These deep learning-based target detection algorithms have significantly enhanced the accuracy and efficiency of traffic sign detection, establishing themselves as crucial technologies in this field.

2.3 METHODOLOGY

2.3.1 YoloV5

YOLOv5 (You Only Look Once Version 5), an advanced implementation in target detection modelling, has shown significant advantages in many practical applications. Although it is not the latest version in the YOLO family, its efficient design and optimisation features make it still important in target detection tasks[22].

Firstly, YOLOv5 offers significant advantages in terms of real-time performance and inference speed. YOLOv5's single-stage detection architecture enables it to provide fast detection capabilities in real-time applications[23]. Compared to traditional target detection models, YOLOv5 enables fast operation on mobile devices and embedded systems by optimising the use of computational resources, which is particularly important for application scenarios that require immediate feedback, such as autonomous driving and video surveillance.

Secondly, YOLOv5's network architecture has been carefully designed and optimised. Its backbone network, CSPDarknet53, introduces the Cross Stage Partial Network (CSPNet) structure, which effectively reduces the amount of computation by dividing the network into multiple stages for feature extraction and fusion, and at the same time improves the model's ability to detect targets at different scales. In addition, YOLOv5 integrates a Feature Pyramid Network (FPN) to form a Path Aggregation Network with Feature Pyramid Network (PAN-FPN), which further enhances the detection capability of targets of different sizes. This design enables YOLOv5 to maintain high accuracy and perform well when dealing with targets of various sizes, as shown in Fig. 5.

YOLOv5 also makes significant improvements in training and optimisation. Its improved loss functions include classification loss, localisation loss and confidence loss, and optimised bounding box regression and target classification. These improvements help to improve the overall performance of the model and reduce false and missed detections. In addition, YOLOv5 extensively applies data enhancement techniques such as image rotation, scaling and flipping during the training process, which increase the diversity of the data and improve the generalisation ability of the model.

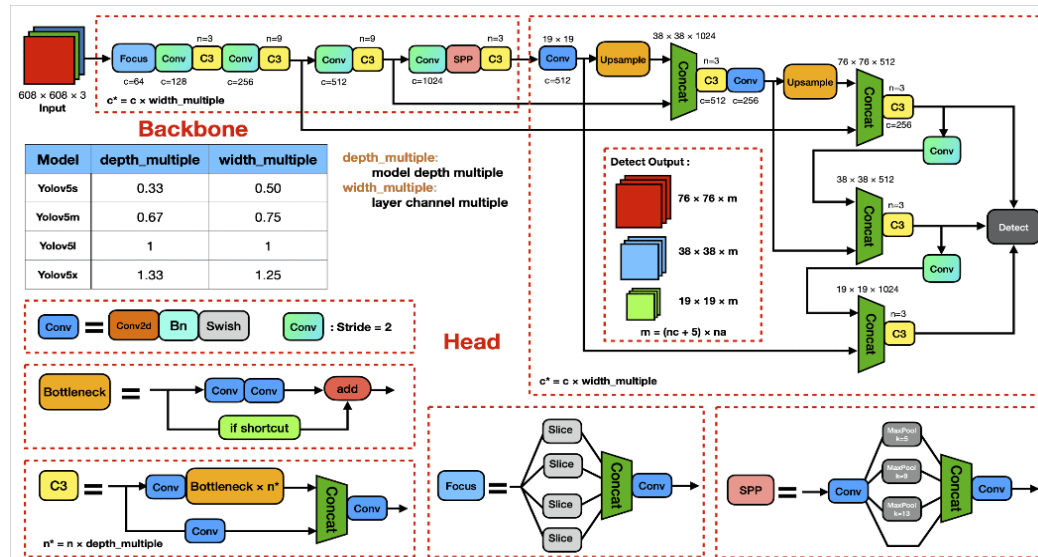


Fig 5: Yolov5 Structure

The lightweight design of YOLOv5 is also a major advantage. By adopting a more efficient network architecture and optimising the computational graph, YOLOv5 significantly reduces the computational overhead of the model, making it suitable for low-power devices and edge computing environments. This lightweight design not only improves the model's deployment flexibility, but also reduces operating costs.

In addition, YOLOv5 has received extensive support and attention in the open-source community. The open-source community provides rich documentation, examples, and technical support, which makes it easier to use and tune YOLOv5. This community support not only promotes the wide application of the model, but also accelerates the iteration and optimisation of the technology.

Compared with other mainstream target detection models, YOLOv5 has significant advantages in speed and efficiency. For example, compared with detection models such as Faster R-CNN, RetinaNet, etc., YOLOv5 performs much better in terms of real-time and computational efficiency. Meanwhile, although the latest versions of YOLO models (e.g., YOLOv7, YOLOv8) have improved in performance, YOLOv5 still demonstrates its unique value in many real-world applications with its design optimisation and efficient performance[24].

2.3.2 BIFPN (Bi-directional Feature Pyramid Network)

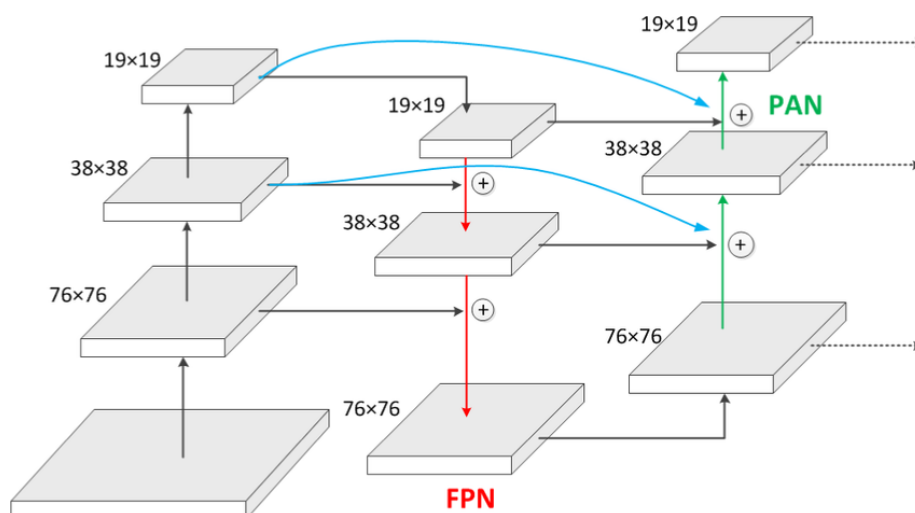


Fig 6: The Structure of PAN-FPN

BIFPN (Bidirectional Feature Pyramid Network) is an improved feature pyramid network designed to improve the fusion efficiency of multi-scale features and its structure is shown in Fig 7[26]. Feature pyramid networks (FPNs) are widely used to fuse feature maps at different scales in computer vision tasks, especially in target detection tasks. BIFPN further improves feature fusion by introducing bidirectional feature fusion and a learnable weighting mechanism, which improves detection accuracy and model robustness.

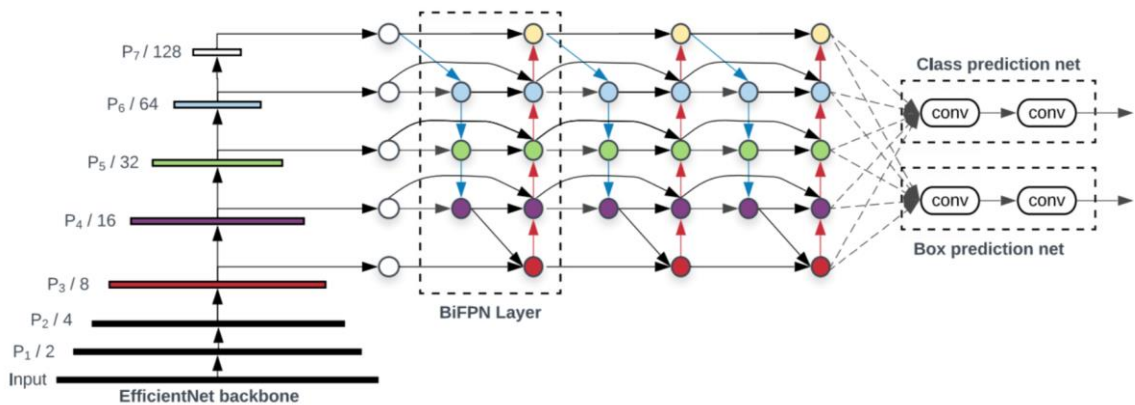


Fig 7: The Structure of Bi-directional Feature Pyramid Network[26]

The core innovation of BIFPN is the bidirectional feature fusion path. This design allows features to flow bi-directionally between the upper and lower layers of the pyramid network[41]. This bi-directional fusion not only enhances the transfer of detail information from the bottom layer to the top layer, but also allows the contextual information from the top layer to effectively complement the bottom layer feature map. In this way, BIFPN is able to capture more contextual information and enhance the comprehensiveness of feature representation.

BIFPN introduces a learnable weighting mechanism to dynamically adjust the feature maps at different scales. Specifically, the fusion weight of each feature map is learned through network training, which enables dynamic adjustment of its contribution in the final feature according to the quality and importance of the feature map[27]. This weighting mechanism not only improves the flexibility of feature fusion, but also adaptively optimises the expressive power of features across different tasks and datasets.

While maintaining high performance, BIFPN also considers computational efficiency. In order to reduce the computational overhead, BIFPN employs some lightweight operations such as depth-separable convolution. This convolution operation decomposes the traditional convolution operation into two stages, depth convolution and point-by-point convolution, which significantly reduces the amount of computation and model parameters while maintaining the performance of the model[28].

BIFPN performs well in several computer vision tasks, especially in target detection tasks. Experiments on the COCO target detection dataset shows that BIFPN significantly improves detection accuracy through effective feature fusion. In addition, BIFPN is widely used in the EfficientDet model to further improve the detection performance[26]. EfficientDet combines the feature fusion capability of BIFPN and efficient network design to achieve better detection accuracy and lower computational overhead, so using BIFPN instead of FAN-FPN to optimise YOLOv5 is a reasonable choice.

2.3.3 Wise-IoU (Weighted Intersection over Union)

The loss function of YOLOv5 comprises three components: classification loss, confidence loss, and localization loss. These components collaboratively optimize the model's performance in object detection tasks. Classification loss assesses the discrepancy between predicted and actual categories. YOLOv5 commonly utilizes binary cross-entropy loss or focal loss to address class imbalance issues. Confidence loss evaluates the discrepancy between the confidence of a predicted bounding box and the actual situation[29]. YOLOv5 determines whether the bounding box encompasses the target and its accuracy by assessing the confidence level of the box. Localization loss measures the difference between predicted and actual bounding boxes. YOLOv5 employs CloU (Complete Intersection over Union) loss. Building on traditional IoU, CloU incorporates the overlapping area, centre point distance, and aspect ratio, enhancing the precision of localization. The formula for CloU is as follows:

$$CIoU = IoU - \frac{\rho^2(b, b^{gt})}{c^2} - \alpha v \quad (1)$$

where $\rho^2(b, b^{gt})$ is the Euclidean distance between the centroids of the prediction frame and the real frame, c is the diagonal length of the closure region of the prediction frame and the real frame, and α is a weighting factor to balance the intersection ratio and relative ratio priority, and v is a relative ratio penalty term to measure the consistency of the aspect ratio of the predicted and real boxes.

CIoU incorporates IoU, centroid distance and aspect ratio into the calculation, making it able to comprehensively measure the similarity between the predicted and real boxes. Meanwhile, by introducing penalty terms for centroid distance and aspect ratio, CIoU is able to adjust the prediction frame more accurately and reduce the bounding box regression error. However, compared with the traditional IoU, the computational complexity of CIoU is higher and the parameters α and v need to be adjusted according to the specific task, which increases the difficulty of parameter adjustment[30].

For the traffic sign recognition task, traffic signs are usually found in complex traffic environments, such as busy city roads, where multiple signs may block each other or be very close to each other, and CIoU mainly evaluates the quality of the bounding box by the overlapping area, the distance from the centroid, and the aspect ratio, but in the scenario of overlapping targets and dense targets, CIoU may not be effective enough to accurately distinguish between neighbouring traffic signs. marking[31].

The WIoU loss function is a bounding box regression loss function designed based on a dynamic non-monotonic focusing mechanism. The penalty term of the WIoU loss function is based on the GIoU loss function, which uses a dynamic non-monotonic focusing mechanism that dynamically adjusts the size of the penalty term according to the degree of overlap between the prediction box and the real box, so that the lower the overlap has a larger penalty and the higher the overlap has a smaller penalties[32]. This can avoid some problems in the GIoU loss function, such as the value of the loss function may increase when the prediction frame contains

the real frame, or the value of the loss function may decrease when the closure area between the prediction frame and the real frame is large.

$$L_{WIoUv1} = R_{WIoU} L_{IoU} \quad (2)$$

$$R_{WIoU} = \exp \frac{(x-x_{gt})^2 + (y-y_{gt})^2}{(W_g^2 + H_g^2)} \quad (3)$$

where c is the Euclidean distance between the centroids of the predicted and real frames, c is the diagonal length of the closure region of the predicted and real frames, α is a weight coefficient for balancing the priority of the intersection and merger ratios and the relative proportions, and v is a relative proportionality penalty term for measuring the congruence of the aspect ratios of the predicted and real frames.

The WIoU loss function is a bounding box regression loss function designed based on the dynamic non-monotonic focusing mechanism. The penalty term of the WIoU loss function is designed on the basis of the GIoU loss function, which adopts the dynamic non-monotonic focusing mechanism to dynamically adjust the size of the penalty term according to the degree of overlap between predicted frames and real frames, so that the lower the overlap the larger the penalty term, and the higher the overlap the smaller the penalty term[33]. This can avoid some problems in the GIoU loss function, such as when the prediction frame contains the real frame, the value of the loss function may increase; or when the closure area between the prediction frame and the real frame is large, the value of the loss function may decrease.

Taking these factors into consideration, the WIoU loss function is chosen to be used instead of the CIoU loss function to improve the accuracy of the target detection model. The WIoU loss function takes into account both positional relationships and shape differences by introducing a dynamic non-monotonic focusing mechanism. In the traffic light recognition task, the changes in the positions and shapes of the signal lights may be affected by a variety of factors such as changes in the viewing angle, lighting conditions, occlusions etc[34]. The WIoU loss function takes into account the relative positional relationships in these cases, which makes the model pay more attention to the appropriate degree of overlap during the training process, rather than relying only on simple IoU computation. This helps to improve the

model's detection performance for signals with different shapes and positions, making it more robust.

2.4 CONCLUSION

This chapter provides an in-depth look at the key methods and techniques used to optimise the YOLOv5 model for traffic sign detection. It explored the rationale behind the selection of specific modifications, such as merging depth-separable convolution and BiFPN, to enhance the feature extraction capabilities of the model while maintaining computational efficiency. These techniques have been carefully selected to address the unique challenges posed by the TT100K dataset, in particular the need to accurately detect small and in-complex environments traffic signs in real-world scenarios. In this chapter, it discussed the importance of balancing model complexity with performance, especially when deploying models on mobile platforms with limited computational resources. The introduced modifications are shown to improve the model's ability to detect small objects and signs in complex environments, which is a key factor in traffic sign detection, without significantly increasing the computational load. As we move into the next chapter, which will focus on the study of the methodology and the detailed analysis and evaluation of optimising the performance of the model, it would be built on the foundational work discussed here. The insights gained from this chapter will serve as a basis for understanding the behaviour of the model in a variety of scenarios and will guide the interpretation of the results in subsequent analyses.

CHAPTER 3 METHODOLOGY

3.1 RESEARCH METHODOLOGIES

This chapter outlines the specific methodologies employed to achieve the objectives of this research. The chapter covers the philosophical stance, research strategy, chosen approaches, potential risks, and data evaluation methods. The methodology is designed to ensure rigorous testing and optimization of the YOLOv5 model for traffic sign recognition.

3.1.1 PHILOSOPHIES

This research adopts a positivist philosophy, which is rooted in the belief that reality is objective and can be observed and described through empirical evidence[34]. This philosophy aligns with the project's focus on enhancing the YOLOv5 algorithm through quantitative data analysis and model evaluation.

Besides, the positivist approach is appropriate because it allows for the systematic testing of hypotheses related to model performance, ensuring that the results are based on measurable, observable outcomes. The advantages include the ability to use statistical tools for robust analysis and the generalizability of findings across similar contexts.

One potential disadvantage of a positivist approach is its limitation in addressing the nuances of qualitative factors that might influence model performance. To mitigate this, the research incorporates some qualitative insights during the model evaluation phase, particularly when interpreting the results in different environmental contexts.

3.1.2 Strategy

The research strategy for optimizing the YOLOv5 traffic sign detection and recognition model adopts a hybrid approach, integrating deductive and inductive methods at different stages of the project. Firstly, a systematic review of academic

papers, technical reports, and previous research on YOLOv5 and related object detection models was conducted using deductive reasoning. This literature review aims to understand the historical performance and theoretical basis of YOLOv5, identify key concepts and progress, and highlight the gaps in existing research. The outcome of this stage provides a solid theoretical foundation for the hypothesis of improving potential performance through structural optimization.

Based on the literature review, the next step is to propose hypotheses, which is another deductive method. Proposing hypotheses is to guide the experimental phase and set clear expectations for the experimental results. For example, assumptions may include statements such as "adding attention mechanisms will significantly improve the detection accuracy of small traffic signs" or "optimizing anchor boxes will reduce false positives in traffic sign detection". These assumptions will guide subsequent experimental work.

Subsequently, research shifted towards inductive data collection methods. The various traffic sign images were collected from the TT100k public dataset and used data augmentation techniques to expand and enrich the dataset. This powerful dataset serves as the foundation for training and testing YOLOv5 models, ensuring comprehensive coverage of various conditions and scenarios.

During the model training and testing phase, induction method is also used to train the benchmark YOLOv5 model and its optimized structural version using the collected dataset. Use a separate validation set to evaluate performance, analyse metrics such as accuracy, recall, and mAP. This stage can extract performance patterns and gain a deeper understanding of the impact of different structural optimizations on the model in practical scenarios.

Performance evaluation is another induction stage, which involves comparing the benchmark YOLOv5 model and its optimized version using consistent evaluation criteria. By analysing performance indicators, identify the strengths and weaknesses of each model, and compare in detail the effectiveness of different optimizations in traffic sign recognition tasks.

The result analysis is conducted in an inductive manner, including a thorough examination of the experimental results, in order to draw specific conclusions about the optimization effect. Summarized the performance differences between the optimized YOLOv5 model and the unoptimized version, as well as other benchmark models, in order to clearly understand the impact of structural optimization on model performance.

Finally, in the theoretical verification stage, deductive methods are used to validate the hypotheses. Compare the experimental results with the initial hypothesis and use logical reasoning to determine whether the observed results support or refute the proposed hypothesis. This stage ensures that theoretical expectations are consistent with empirical results.

The conclusion and recommendation stage comprehensively summarizes and deduces insights, draws comprehensive conclusions, and proposes practical and feasible suggestions. By combining experimental data with theoretical analysis, comprehensive conclusions can be drawn, providing suggestions for further research and optimization of the potential application of YOLOv5 model in practical traffic sign detection systems. This strategic framework ensures a rigorous and balanced approach, achieving a comprehensive understanding and optimization of the YOLOv5 model.

3.1.3 Approaches

The methodology for optimizing the YOLOv5 model incorporates a multi-faceted approach to ensure a comprehensive analysis of the proposed enhancements. This research employs a combination of literature review, experimental approach, and software development to achieve its objectives.

Initially, a thorough literature review serves as the foundation of this study. This review involves systematically analysing academic papers, technical reports, and previous research on YOLOv5 and related object detection models. The primary goal is to establish a theoretical framework and identify existing gaps in the current research. By synthesizing insights from various sources, the literature review

provides a solid basis for understanding the historical performance and theoretical underpinnings of YOLOv5, as well as its potential for optimization. This review also informs the formulation of hypotheses regarding how specific optimizations could enhance the model's performance in traffic sign detection.

Following the literature review, the research transitions into an experimental approach. This phase is centred around training and evaluating the YOLOv5 model with and without the proposed optimizations: BiFPN (Bi-directional Feature Pyramid Network) and Wise-IoU (Wise Intersection over Union) loss function. The experimental approach involves a detailed process of model training, where the baseline YOLOv5 model is first trained on a diverse dataset of traffic sign images. Subsequently, the model is enhanced with the BiFPN and Wise-IoU optimizations. The performance of these optimized models is then assessed using a separate validation set, with metrics such as precision, recall, and mean Average Precision (mAP) being meticulously analysed. This phase aims to empirically evaluate the impact of the optimizations on the model's accuracy and efficiency in detecting traffic signs.

In conjunction with the experimental approach, a software development approach is employed to implement the proposed optimizations into the YOLOv5 framework. This involves modifying the existing codebase to integrate the BiFPN for improved feature extraction and the Wise-IoU for enhanced loss function performance. The development process includes coding, testing, and debugging to ensure that the optimizations are effectively incorporated and function correctly within the YOLOv5 architecture. The software development approach ensures that the theoretical enhancements are practically implemented and evaluated, bridging the gap between theoretical research and practical application.

Together, these approaches—literature review, experimental evaluation, and software development—provide a comprehensive methodology for optimizing the YOLOv5 model. The literature review establishes the theoretical basis, the experimental approach validates the effectiveness of the optimizations, and the software development ensures practical implementation. This integrated approach

ensures a thorough and rigorous examination of the proposed enhancements, leading to a well-rounded understanding of their impact on traffic sign detection.

3.1.3.1 Literature Review

Literature review is the fundamental part of this study. Through systematic search and analysis of the literature related to the field of traffic sign recognition and YOLO model optimisation, we constructed the theoretical framework of the study and clarified the technical direction of optimising the YOLOv5 model. To this end, we pay special attention to two key technologies, BiFPN (Bi-directional Feature Pyramid Network) and WIoU (Weighted Intersection over Union), and carry out an in-depth search and analysis of related literature.

In the literature finding phase, it mainly relies on several academic databases, such as IEEE Xplore, SpringerLink, Google Scholar and ACM Digital Library. these platforms cover a large number of high-quality research results in the fields of computer vision and deep learning. I first identify several core keywords, including "YOLOv5 optimisation", "Bi-directional Feature Pyramid Network in object detection", "Weighted IoU for object detection", "Deep learning for traffic sign recognition " and so on. These keywords were chosen based on a comprehensive understanding of the research topic and further refined and expanded through preliminary searches.

During the literature screening process, I focused on high-impact journal papers and conference papers published within the last five years, especially those studies published in top computer vision conferences such as CVPR, ICCV, ECCV and NeurIPS. By carefully reading the abstracts, methods sections, and conclusions of these papers, I identified several technological advances that are highly relevant to the objectives of this research. Among them, BiFPN and WIoU technologies are mentioned in several papers and show significant advantages in several experiments, so I focus our analysis on these technologies.

Specifically, BiFPN is an improved feature fusion architecture, first proposed in a study by EfficientDet[26]. The architecture achieves efficient fusion of multi-scale features through a bi-directional feature pyramid network, which significantly improves the performance of target detection models at different scales. It is shown that BiFPN finds a good balance between computational efficiency and detection accuracy, which is particularly important for tasks with multi-scale features like traffic sign recognition. It can be found in detail the structural design of BiFPN and its performance in various tasks in the literature, and consider its potential application value in the YOLOv5 model.

On the other hand, WIoU has gained much attention as an improved loss function for target detection. Compared with the traditional IoU, WIoU measures the position error of the target frame more accurately by introducing a weighting mechanism. This technique was initially proposed for small target detection tasks and later validated by several papers for its advantages in complex contexts [39]. By analysing the relevant literature in depth, it was found that WIoU not only improves the detection accuracy of the model, but is also particularly suitable for the detection task of small and dense targets, which is common in traffic sign recognition.

In the final stage of the literature review, I integrated these findings and clarified the direction of applying BiFPN and WIoU techniques to YOLOv5 optimisation. The studies in the literature provide strong theoretical support for our technology selection, as well as specific guidance for the optimal design of the model. In the subsequent phase of the research, further validation and enhancement of the YOLOv5 model's optimisation strategy will be pursued, drawing on insights gleaned from these literatures.

3.1.3.2 Model Building

In this study, the core task of software development is to optimize the YOLOv5 model to improve its accuracy and efficiency in traffic sign recognition. The optimization process involves the introduction of two key technologies, BiFPN (Bidirectional Feature Pyramid Network) and WIoU (Weighted Intersection over

Union), to enhance the performance of the model in complex practical application environments.

The development environment is established using the Ubuntu operating system, with PyTorch selected as the deep learning framework of choice, favoured for its adaptability and extensive applicability. The integration of GPU acceleration tools such as CUDA and cuDNN is noted for its capacity to markedly reduce the duration of training and testing cycles, thus enhancing the efficiency of the development workflow. Throughout the development process, the TT100K dataset is employed, recognized for its representativeness and prevalence in traffic sign recognition research. This dataset offers the model a wealth of annotated data alongside a variety of scenarios, which is instrumental for refining the model's performance and robustness.

The integration of BiFPN is one of the key steps in optimizing YOLOv5. By introducing a multi-level feature pyramid structure, BiFPN supports bidirectional fusion and transmission of features, enabling the model to have higher flexibility and accuracy in handling multi-scale targets. During this process, the feature extraction network of YOLOv5 was modified and multiple feature fusion paths were added. To ensure the efficiency of BiFPN, parameters such as the number of channels, convolution kernel size, and activation function for each layer feature were finely adjusted. This meticulous tuning was essential for optimizing the performance of the network and achieving the desired balance between accuracy and computational efficiency. Meanwhile, in order to prevent overfitting caused by increased model complexity,

The positioning accuracy of YOLOv5 was improved by WIoU into the loss function. This modification enables the model to pay greater attention to the exact alignment of the bounding box with the target, which is crucial for enhancing detection precision, especially for challenging targets that are small or have significant overlap with others. WIoU assigns different weights to different parts of the target box, allowing the model to pay more attention to the precise positioning of the target box when calculating losses. This improvement is particularly effective for detecting small, dense, and overlapping targets. During the development process, it would

analyse in detail the shortcomings of the original IoU loss function and designed and implemented the WIoU loss function module. After multiple experimental verifications, the introduction of WIoU significantly improves the detection accuracy of the model in complex scenes. Meanwhile, I optimized the weight parameters of WIoU to ensure its optimal performance on different datasets.

After integrating BiFPN and WIoU, this project would conduct comprehensive training on the optimized YOLOv5 model. The training process adopts a gradual optimization strategy, starting from the training of the basic model, gradually increasing the model complexity, and adjusting parameters based on the performance of the validation set after each iteration. To ensure the stability and generalization ability of the model, cross validation was applied and trained with using a large amount of TT100K data. The preliminary evaluation results show that the model introduced with BiFPN and WIoU has significantly improved detection accuracy and computational efficiency, laying the foundation for subsequent experimental testing.

3.1.3.3 Testing

Experimental testing is a key step in verifying the optimization effect of YOLOv5 model. By testing on real datasets, the performance of the model in practical application scenarios can be evaluated. This process not only validates the accuracy of the model, but also provides feedback for further optimization.

To ensure the representativeness and reliability of the experimental results, I chose the TT100K dataset for testing. The TT100K dataset provides rich traffic sign samples and diverse scenarios, which helps to comprehensively evaluate the performance of the model in practical environments. In the data preparation stage, I standardized the dataset, including image size adjustment, colour balance, and data annotation correction. These processing steps ensure the consistency of the dataset and provide a reliable foundation for subsequent experiments.

The design of the experimental testing aims to comprehensively evaluate the performance of the model. Multiple testing scenarios have been established, including standard scenarios, detection of small objects in complex backgrounds, and detection of overlapping objects. To ensure the reliability of the experimental results, strict control of variables has been implemented, ensuring that the environment and conditions of each test are consistent. This approach is taken to obtain results with strong comparability. The evaluation indicators include mean accuracy (mAP), precision, recall, F1 score, and detection speed. Through comprehensive analysis of these indicators, the performance of the model in different scenarios can be clearly understood. After each test, the results are compared with those of the unoptimized YOLOv5 model to evaluate the effectiveness of the optimizations.

The experimental results show that the YOLOv5 model optimized by BiFPN and WIoU performs better than the baseline model in most scenarios. Especially in complex backgrounds and multi-scale object detection, BiFPN significantly improves the detection accuracy of the model, while WIoU improves the localization accuracy of small and overlapping targets. The result analysis not only focuses on the overall performance of the model, but also explores in detail the differences in performance under different testing scenarios. These analyses provide in-depth insights into the mechanism of optimization techniques and reveal potential areas for improvement of the model in specific scenarios. Finally, the experimental results validated the effectiveness of YOLOv5 model optimization and provided a basis for further optimization in future research.

3.1.4 Validity, Reliability & Generalisability

3.1.4.1 Validity

Validity refers to the extent to which the research accurately measures what it intends to measure and whether the results genuinely reflect the impact of the optimizations on YOLOv5's performance. In this study, validity is ensured through a carefully designed methodology that aligns with the research objectives.

To establish internal validity, the research design includes systematic experimentation with controlled variables. The optimization techniques, BiFPN and Wise-IoU, are applied to the YOLOv5 model in a manner that isolates their effects on model performance. By maintaining a consistent experimental setup and using standardized performance metrics such as precision, recall, and mean Average Precision (mAP), the study ensures that any observed changes in model performance can be attributed directly to the optimizations.

Construct validity is addressed by employing a well-defined theoretical framework derived from the literature review. The hypotheses and experimental design are grounded in established theories of object detection and model optimization. This ensures that the research measures the intended constructs—improvements in traffic sign detection accuracy and efficiency—as conceptualized in the theoretical background.

External validity is considered by evaluating the model across a diverse dataset of traffic sign images. This approach aims to ensure that the findings are not limited to a specific subset of data but are applicable to various scenarios and conditions. By testing the optimizations on multiple datasets and in different conditions, the research seeks to enhance the generalizability of its findings beyond the immediate experimental setup.

3.1.4.2 Reliability

Reliability refers to the consistency and stability of the research findings. In this study, reliability is ensured through several measures designed to produce consistent results across different trials and researchers.

To ensure inter-rater reliability, the research utilizes standardized procedures for data collection, model training, and performance evaluation. This includes using consistent datasets, pre-processing techniques, and evaluation metrics. By

adhering to these standardized methods, the study minimizes variability in results that could arise from differences in data handling or evaluation approaches.

Test-retest reliability is addressed by conducting multiple training and testing iterations with the YOLOv5 model. This approach involves training the model under various conditions and assessing performance consistency across different experimental runs. By replicating experiments and comparing results, the study verifies that the findings are stable and not due to random variations.

Internal consistency is ensured by using a structured experimental protocol and rigorous statistical analysis. Performance metrics are calculated using established formulas and techniques, and statistical significance is assessed with appropriate methods. This rigorous approach ensures that the findings are reliable and reflect true performance changes attributable to the optimizations.

3.1.4.3 Generalisability

Generalisability refers to the extent to which the research findings can be applied to other contexts, models, or datasets beyond those specifically studied.

External generalisability is considered by evaluating the YOLOv5 model with the BiFPN and Wise-IoU optimizations on a diverse set of traffic sign images. This approach aims to determine whether the observed improvements in model performance are applicable to different types of traffic signs, varying environmental conditions, and different datasets. By including a range of scenarios and testing conditions, the research seeks to enhance the applicability of its findings to broader contexts.

The research also considers the theoretical generalisability of the proposed optimizations. Theoretical insights gained from the study, such as the impact of BiFPN on feature extraction and Wise-IoU on loss function performance, are evaluated for their potential relevance to other object detection models and tasks. This involves assessing whether similar optimizations could benefit other models or applications in the field of computer vision.

Practical generalisability is addressed by providing practical recommendations based on the research findings. The study aims to offer actionable insights for implementing similar optimizations in real-world traffic sign detection systems. By demonstrating the effectiveness of the proposed enhancements in practical scenarios, the research seeks to contribute to the development of more accurate and efficient traffic sign recognition systems.

In summary, the validity of the study is ensured through rigorous design and measurement techniques, reliability is maintained through consistent procedures and replication, and generalisability is enhanced by evaluating the findings across diverse contexts and theoretical frameworks. These measures collectively support the robustness and applicability of the research outcomes.

3.1.5 Data Evaluation Methods

The data evaluation methods employed in this research are designed to ensure that the performance of the YOLOv5 model, with its optimizations, is rigorously assessed. The evaluation process encompasses both data collection and data analysis, with a specific focus on the TT100K dataset—a comprehensive collection of traffic sign images from Tsinghua University and Tencent[36]. This dataset contains more than 200 categories with three types of signs, warning, prohibited and auxiliary signs, which are distributed as shown in the figure 8:

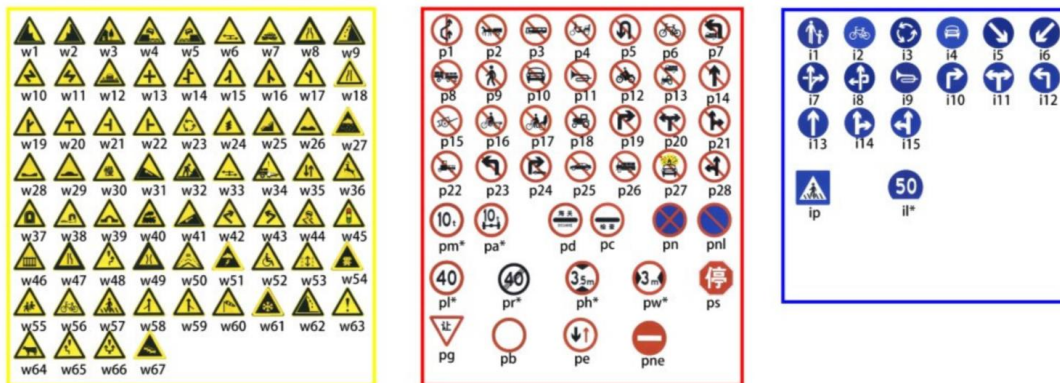


Fig 8: TT100k Dataset

3.1.5.1 Data Collection

For this research, the primary dataset is TT100K, which includes over 200 signage categories. However, due to the imbalanced distribution of category instances, preprocessing is required to ensure that the dataset is suitable for effective training and evaluation.

The preprocessing steps involve filtering the dataset to focus on categories with a sufficient number of instances. Specifically, categories with fewer than 100 instances are excluded from the training set. This filtering results in a refined dataset with 45 categories, each having a substantial number of samples to support robust model training and evaluation. The refined dataset ensures that the model is trained on a representative subset of categories, improving the reliability of the experimental results.

Additionally, data augmentation techniques are applied to enhance the diversity and variability of the dataset. These techniques include image rotation, scaling, flipping, and colour adjustments, which help simulate various real-world conditions and improve the model's generalizability.

3.1.5.2 Data Evaluation

In evaluating the YOLOv5 model optimised using BiFPN and Wise-IoU, a comprehensive evaluation methodology combining quantitative and qualitative analysis was used to fully assess the model's performance in the traffic sign detection task. This evaluation process involves several key performance metrics, including precision, recall, mean accuracy (mAP), as well as more specific metrics such as FLOPs, APS, APM, and APL, to analyse the model's performance in-depth from different perspectives[37].

Firstly, precision rate as a basic measure of model detection accuracy, it indicates how many of all positive class detections made by the model are truly positive, i.e.,

the proportion of correctly detected traffic signs to all detected as traffic signs. The calculation formula is:

$$precision = \frac{TP}{TP+FP} \quad (4)$$

A high precision rate means that the model has a low probability of false alarms when a sign is detected.

Secondly, Recall Rate Evaluates the comprehensiveness of the model, which indicates how many of all actually existing traffic signs are successfully detected by the model. The formula is calculated as:

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

A high recall rate indicates that the model has a low probability of missing detection in the detection task, i.e., it is able to find most of the traffic signs that are actually present.

Mean Accuracy (mAP) on the other hand is a composite performance metric that combines precision and recall. mAP is the average of the precision and recall calculated at multiple thresholds, and is used to measure the overall performance of the model in multi-category detection. mAP is able to provide an overall model performance score, which in turn reflects the overall accuracy of the model in detecting and classifying multi-category traffic signs.

In order to more deeply evaluate the computational efficiency of the model and its performance in real-world applications, FLOPs (Floating Point Operations) become an important reference index. FLOPs refer to the number of floating-point operations required by the model to process a single image, which is an important indicator of the computational complexity of the model. Lower FLOPs indicate that the model is more lightweight and less computationally expensive, making it more suitable for deployment in environments with limited computing resources, such as mobile devices or embedded systems.

The performance of the model may vary when detecting traffic signs with different sized targets, so the more granular metrics of Average Precision for Small Targets

(AP_s), Average Precision for Medium-sized Targets (AP_M), and Average Precision for Large Targets (AP_L) also need to be evaluated:

Average Precision for Small Targets (AP_s) is primarily used to evaluate the performance of the model in detecting small targets, such as small traffic signs in the distance. Since small targets are more difficult to detect because they have fewer and less distinctive features in the image, a higher APS indicates that the model has an excellent ability to detect small objects, which is critical for improving the comprehensiveness of traffic sign detection.

Average Precision for Medium-sized Targets (AP_M) Evaluates the model's performance in detecting medium-sized targets. Medium-sized targets are often the most commonly detected objects, so the APM metric reflects the model's performance in typical detection scenarios. By comparing the APM values before and after optimisation, it is possible to understand the effect of BiFPN and Wise-IoU optimisation on the overall detection capability of the model.

Average Precision for Large Targets (AP_L) evaluates the effectiveness of the model in detecting large targets. Large targets usually occupy a large portion of the image or are in close proximity, and a high APL indicates that the model has a high accuracy in detecting these targets, avoiding misdetections and missed detections.

A statistical significance test is performed to ensure that the performance improvement resulting from the optimisation is statistically significant. By comparing the performance metrics of the baseline model with those of the optimised model, using appropriate statistical tests (e.g. t-test or ANOVA), it is possible to verify whether the observed performance differences are significant. Such tests help to ensure that the optimisation effect is not due to random fluctuations alone, but rather an actual performance improvement due to the improved model structure.

In addition to quantitative metrics, qualitative assessment is also an important part of the evaluation process. Qualitative analysis provides insight into the actual performance of the model in specific traffic sign categories and scenarios through visual inspection of the model test results. For example, the model's strengths and

weaknesses can be identified by manually checking the model's performance in detecting traffic signs that are small, occluded or under changing lighting conditions. Such an analysis helps to better understand the impact of BiFPN and Wise-IoU optimisation on the real-world performance of the model, especially in complex scenarios.

Finally, the results of the optimised YOLOv5 model are compared with those of the baseline YOLOv5 model, using detailed analysis and visualisation tools (e.g., exact recall curves and confusion matrices) to demonstrate the effect of the optimisation. The Precision Recall Curve shows how the model performs under different detection thresholds, helping to analyse the strengths and weaknesses of the model under various detection sensitivity settings. The confusion matrix, on the other hand, visually demonstrates the classification ability of the model, helping to identify which traffic sign categories are easily confused or missed.

In summary, this data evaluation methodology combines rigorous quantitative analysis, in-depth qualitative assessment, and careful comparative analysis to fully assess the performance of the YOLOv5 model after optimisation using BiFPN and Wise-IoU. By using the improved TT100K dataset with these comprehensive evaluation strategies, it is possible to provide a clear, accurate and comprehensive assessment of the real-world effects of the optimisation, especially in terms of real-world performance in traffic sign detection and recognition tasks.

Translated as is customary for English papers

3.2 DEVELOPMENT METHODOLOGY

Agile software development methodology was adopted for this project, mainly because its flexibility and iterative nature fit well with the dynamic requirements of model optimisation and enhancement tasks. Agile methods facilitate continuous improvement and iterative testing, allowing frequent adjustments based on experimental results and new insights. By breaking down the development process into manageable iterations or sprints, the project can focus on specific aspects of model improvement, integrate feedback, and refine the YOLOv5 model over time.

This iterative process ensures that each optimisation is thoroughly tested and integrated before moving on to the next stage.

In this project, the agile approach involved several key phases. Firstly, there was a planning and requirements gathering phase, where initial planning included defining the scope of the optimisations, setting objectives and outlining specific modifications to be implemented in the YOLOv5 framework. This phase also included gathering requirements for BiFPN and Wise-IoU optimisations. This is followed by a development and integration phase, the core development phase includes the implementation of BiFPN for enhanced feature extraction and Wise-IoU for improved performance of the loss function. This phase includes coding, integration and initial testing of the optimisations in the YOLOv5 model.

In the test and evaluation phase, after integration, the model is rigorously tested using the improved TT100K dataset. Performance metrics were evaluated and necessary adjustments and improvements were made based on feedback. Then enter the iteration and refinement phase, where the model is iteratively improved based on the test results. Additional cycles of development and testing are performed to resolve any issues and further optimise performance. Finally, there is a deployment and documentation phase, where the optimisation is completed, the model is documented and deployment procedures are developed for practical use in the traffic sign detection system.

This agile approach supported incremental development and integration of BiFPN with Wise-IoU optimisation and was therefore particularly suitable for this project. Through this approach, the development process was able to remain responsive to new discoveries and enable continuous improvement of the YOLOv5 model. The flexibility and iterative nature of the agile approach ensured that each optimisation was thoroughly tested and integrated, improving the overall performance and reliability of the model.

3.3 RESOURCES

The resources required for this project include hardware, software, information and human resources to ensure that the development, training and optimisation of the YOLOv5 model can proceed smoothly.

3.3.1 Hardware Resources

In terms of hardware resources, a high-performance computing infrastructure would be utilized to handle the computational requirements for model training and optimization. Specifically, the NVIDIA GeForce RTX 4070ti super GPU will be used to train the YOLOv5 model due to its high computational power and efficiency in handling deep learning tasks. The Intel Core i5-12600kf CPU will be used for data preprocessing and general computing tasks to provide the necessary computational support. At least 62GB of memory would be employed to support large-scale data processing and model training, ensuring that the handling of extensive datasets and intricate model operations proceeds smoothly. This ample memory allocation is crucial for the successful training of sophisticated models like the optimized YOLOv5, particularly when dealing with substantial volumes of data. For storage, high-speed SSD storage with a capacity of at least 2 TB will be used to store datasets, trained models, and intermediate files, and this fast storage is critical for fast access to data and efficient training of models. In addition, data transfer requires reliable network connectivity, especially when using cloud resources or accessing remote datasets.

3.3.2 Software Resources

In terms of software resources, this project used a variety of tools and frameworks to develop, train and evaluate YOLOv5 models. This project used Python 3.10 as the programming language because of its extensive support for machine learning libraries and frameworks. PyTorch 2.4.0 was used for training and evaluating the YOLOv5 model, PyTorch was chosen because of its dynamic computational graphs and ease of use in deep learning applications. The YOLOv5 framework was obtained directly from the official. The YOLOv5 framework was available directly from the official GitHub repository maintained by Ultralytics, and the specific version was the latest stable version at the start of the project. Data enhancements was

made using the Albumentations library, which provides a range of enhancement techniques to increase the variability of datasets and improve model generalisation. The development environment used Jupyter Notebook, which facilitates interactive development and visualisation, and PyCharm, which provides a comprehensive integrated development environment for Python development. Version control used Git to manage code changes and track different iterations of the model, and the repository was hosted on GitHub. If additional computing power is required, there will use cloud platforms such as Google Colab or GPU-enabled AWS EC2 instances for model training and evaluation.

CHAPTER 4 - DESIGN & IMPLEMENTATION

4.1 INTRODUCTION

This chapter provides a comprehensive overview of the design and implementation of the Traffic Sign Recognition (TSR) system, particularly utilizing the YOLOv5 architecture. This chapter first presents a clear problem statement, outlining the challenges and objectives of the TSR system. Then, it delved into the design phase, detailing the architectural modifications and considerations made to optimize the model. Finally, this chapter discusses the implementation strategies adopted, including data preparation, model training, and evaluation.

4.2 PROBLEM FORMULATION

The core problem solved by this project is to develop a traffic sign recognition system that can operate accurately and efficiently in real scenarios. The challenges faced by traffic sign recognition come from the target to be recognized, the performance of the model itself, and various external environmental influences, such as the variability of traffic signs mentioned in Chapter 2. The shape, colour, and size of traffic signs vary greatly, and despite these differences, a model that can accurately detect and classify traffic signs is needed. Environmental conditions such as lighting, weather conditions, and occlusion can also significantly affect the visibility and recognition of traffic signs. Beyond the initial stage, the system needs to run in real-time and process frames quickly enough to function effectively in dynamic environments such as autonomous driving.

To address these challenges, I chose the YOLOv5 model this time because of its balance between speed and accuracy. However, specific architectural changes were made to improve its performance in recognizing traffic signs under different conditions. In order to simulate different environmental conditions, a comprehensive data augmentation strategy was implemented to improve the robustness of the model. And the hyperparameters of the model have been carefully adjusted to ensure optimal performance in terms of accuracy and speed.

4.3 DESIGN

The design of the Traffic Sign Recognition (TSR) system is a critical aspect of the project, as it determines how effectively the system will perform in real-world conditions. This section outlines the overall design architecture, the objectives that guide the design, and the specific challenges addressed by this design.

4.3.1 System Architecture Overview

In the Fig 9 incorporating BiFPN, in addition to the introduction of the BiFPN module, the Backbone of YOLOv5 has been improved to further enhance the feature extraction capability. The overall flow of the architecture is still shown from left to right, showing the key steps from image input to final detection output, but important optimisations have been made in the Backbone section.

Firstly, the input image enters the improved YOLOv5 Backbone. Compared to the original CSPDarknet53, this improved Backbone improves performance by replacing standard convolutional layers with more advanced ones[38]. Specifically, many of the standard convolutional layers are replaced with deeply separable convolutional layers. Deeply separable convolutional layers are a more efficient way of doing convolution that breaks the standard convolution operation into two steps, deep convolution and point-by-point convolution, dramatically reducing the amount of computation and the number of parameters, while maintaining or even improving the accuracy of feature extraction.

In Backbone, after each layer has been processed by the depth-separable convolutional layer, it is still able to extract feature maps at different scales, labelled as P3, P4 and P5, which represent the feature information of the image at different resolutions. These feature maps continue to be passed to the BiFPN module.

The BiFPN module plays a key role as a bridge in this structure picture. It effectively fuses and enhances the multi-scale feature maps from Backbone through top-down and bottom-up bidirectional paths, and BiFPN's learnable weighting mechanism further optimises the transfer of information between different scales, resulting in a

more accurate output of enhanced feature maps, which are labelled as P3', P4' and P5'[39].

Subsequently, these BiFPN fused feature maps are passed to the detection head of YOLOv5. In the Detection Head, the network converts the feature maps into predictions, including the location of the bounding box, the confidence of the category and the confidence of the target[40]. These predictions are used to locate and identify targets in the image.

Finally, the end of the structure picture shows the loss computation module. Here, the model can choose to use either the original loss function of YOLOv5 or the WIoU loss function to calculate the loss value between the predicted results and the true annotation. The results of the loss computation will be used for backpropagation to update the model parameters to ensure that the improved fusion of Backbone and BiFPN is effective in improving the detection performance of the model.

By demonstrating these improvements in the structural diagrams, especially the optimisation of Backbone and the introduction of BiFPN, the whole YOLOv5 model is not only more efficient in feature extraction, but also significantly improved in detection accuracy. This improvement reflects the design concept of maintaining or even improving the model performance while reducing the computational load.

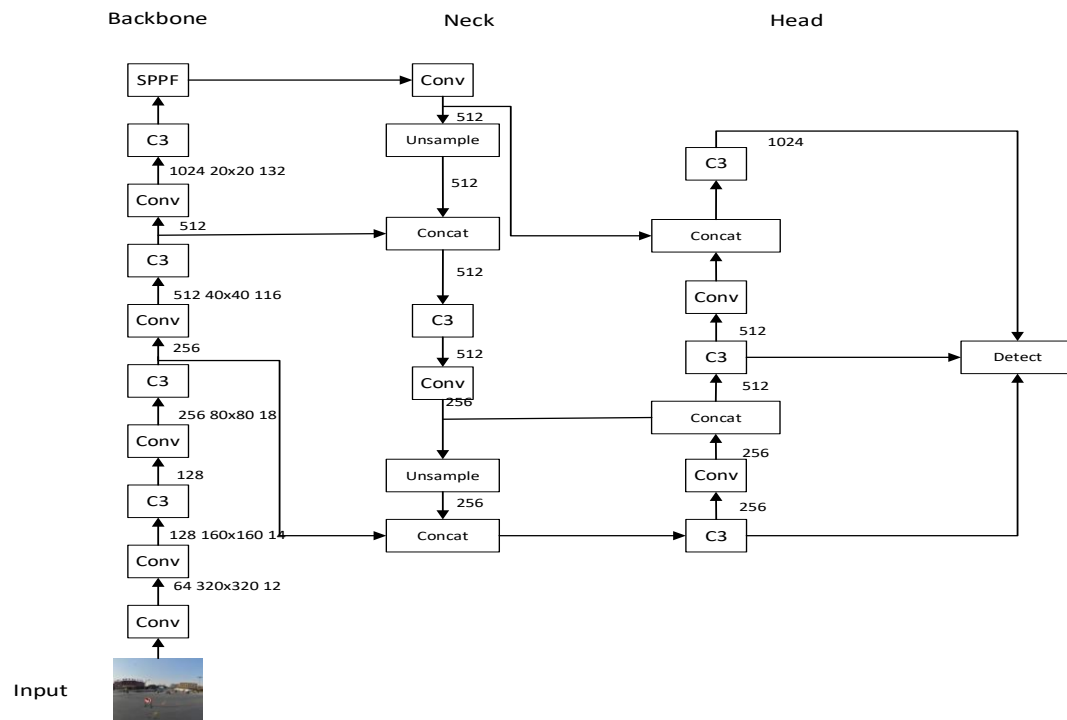


Fig 9: Optimized Structure

4.3.2 Component Design

Each component of the system is designed with specific functionality and optimisation in mind:

TRAINING PIPELINE: The training pipeline is designed for flexibility and efficiency. A migration learning technique is employed to initialise the model using pre-trained weights, which significantly reduces the training time. The training process is then fine-tuned on a specialised traffic sign dataset.

Evaluation Module: this evaluation module aims to assess the performance of the model using key metrics such as mean average precision (mAP), recall, accuracy and inference speed. The results of this module are used to iteratively optimise the model.

4.4 IMPLEMENTATION

The implementation of the Traffic Sign Recognition (TSR) system is a multifaceted process that brings together data processing, model customization, training, and deployment into a

cohesive workflow. This section delves into the various stages of implementation, outlining the specific steps taken to ensure that the system performs optimally in real-world scenarios.

4.4.1 Implementation Overview

The implementation phase is critical as it translates the conceptual design into a functional system. The process begins with data collection and preparation, ensuring that the model is trained on a diverse and representative dataset. Following this, the selected model architecture—YOLOv5 in this case—is customized to better suit the unique demands of TSR. The system is then trained and validated using this data, with additional fine-tuning based on testing results. Finally, the model is deployed into a real-time environment, where it is integrated into a target system and monitored for performance. Each of these stages is essential to building a TSR system that is accurate, fast, and robust enough to handle the complexities of real-world traffic sign recognition.

4.4.2 Detailed Implementation Steps

The first step in the implementation process involves gathering a comprehensive and diverse dataset of traffic signs. Each image in the dataset is meticulously annotated with bounding boxes and labels, ensuring that the model has accurate and detailed information to learn from.

With the data prepared, the next step is selecting and customizing the model architecture. YOLOv5 is chosen due to its renowned balance between speed and accuracy, which is essential for the real-time requirements of TSR systems. However, to further optimize the model for the specific task of traffic sign recognition, two key enhancements are introduced: Bi-directional Feature Pyramid Network (BiFPN) and Weighted Intersection over Union (WIoU) loss function.

BiFPN is integrated into the YOLOv5 architecture to improve the model's feature extraction capabilities, particularly for small and occluded objects like traffic signs. BiFPN facilitates better aggregation of features from different layers of the network, allowing the model to leverage both high-level semantic information and low-level details. This is particularly important in TSR, where signs can appear in various sizes and under different conditions. Implementing BiFPN involves replacing the traditional feature pyramid structure in

YOLOv5's neck with BiFPN's more advanced, weighted feature fusion approach, which helps the model prioritize the most relevant features for each detection task.

Simultaneously, the standard IoU loss function is replaced with WIoU, a more sophisticated loss function that provides better localization accuracy by assigning different weights to various aspects of the bounding boxes, such as width, height, and centre position. This customization is critical because accurate localization is paramount in TSR, where even slight deviations in bounding box predictions can lead to incorrect sign recognition. By fine-tuning WIoU during the training process, the model learns to predict bounding boxes with greater precision, thus enhancing overall system performance.

Once the model architecture is finalized with these customizations, the training phase begins. Transfer learning is employed to speed up the training process and improve performance. The model is initialized with pre-trained weights from the COCO dataset, which provides a solid foundation of generalized features. These weights are then fine-tuned on the specific traffic sign dataset, allowing the model to specialize in recognizing traffic signs.

During training, a combination of techniques such as early stopping, learning rate scheduling, and batch normalization are used to optimize the model's convergence and prevent overfitting. The training process is closely monitored, with metrics like mean Average Precision (mAP), precision, recall, and F1 score used to evaluate performance.

The validation phase involves assessing the model's performance on a separate validation dataset. This step is crucial for identifying any potential overfitting and for guiding further tuning of model hyperparameters. Validation also provides insights into how well the model generalizes to new data, which is a key indicator of its potential success in real-world applications.

After training and validation, the model undergoes rigorous testing on an unseen test dataset to evaluate its generalization capabilities. This testing phase mimics real-world conditions as closely as possible, with the model's performance measured not only on standard metrics but also in challenging scenarios such as poor lighting, partial occlusion, and complex backgrounds.

Based on the test results, the model is fine-tuned to address any weaknesses. This fine-tuning process involves adjusting hyperparameters such as learning rate, batch size, and anchor box sizes. Additional rounds of training may be conducted, focusing on difficult examples from the test set to further enhance the model's robustness and accuracy.

4.5 CONCLUSION

The implementation of the TSR system is a carefully orchestrated process that transforms a conceptual design into a fully functional, real-time system capable of accurately recognizing traffic signs. From the meticulous preparation of a diverse dataset to the innovative customization of the YOLOv5 model with BiFPN and WIoU, each step is designed to enhance the system's performance. The thorough training, validation, and testing phases ensure that the model is not only accurate but also robust enough to handle the challenges of real-world deployment. The final deployment integrates this advanced model into a practical application, making the TSR system ready for use in intelligent transportation systems, where it can contribute to safer and more efficient driving experiences.

CHAPTER 5 – TESTING AND EVALUATION

This experimental dataset has been mentioned several times in the previous sections, the data set used is the tt100k dataset, and the main hyperparameters for this experiment are set as follows: batch size is 32, epochs is 50, optimizer is SGD and image size is 640 for the model.

In this section, firstly, the performance of yolov5-s with unoptimized yolov5 and yolov5 optimised with only the loss function are compared, and then the performance of the unoptimized yolov5-s model, the SSD[42], the Yolov3[43], and the Efficientdet are compared. are compared and the actual results of the tests are shown at the end using model size, number of parameters, floating-point operations per second (FLOPs), AP_s , AP_M , AP_L , and mAP@0.5 as evaluation criteria, and are compared.

5.1 LOSS FUNCTION COMPARISON

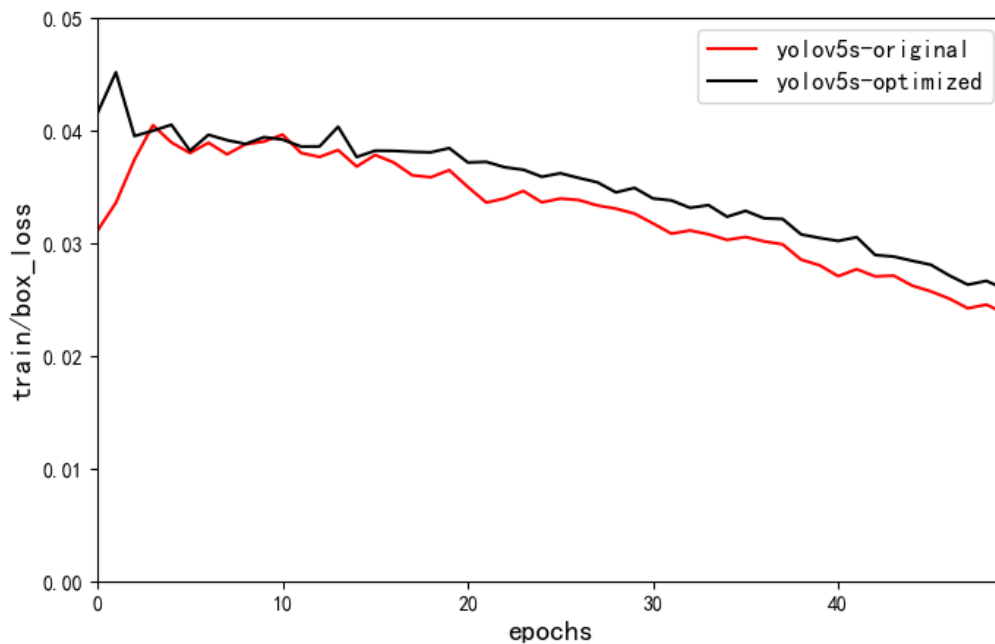


Fig 10: The Changing of Box_loss

This figure 11 shows the change of box loss between the original YOLOv5 model and the YOLOv5 model optimised with WIoU during the training process. As can be

seen from the figure, in the initial stage of training (the first 10 epochs), the loss value of the original YOLOv5 model decreases faster, while the WIoU-optimised model shows some fluctuations, and then gradually stabilises. This may be due to the fact that the model needs some adjustment time to adapt to the new WIoU loss function.

In the middle stage (10 to 30 epochs), the loss values of the WIoU-optimised model were slightly higher than those of the original model. This may indicate that the model slightly underperforms the original version at this stage, despite the introduction of the new optimisation method. However, this does not mean that the optimisation is ineffective, as the model is gradually adapting and learning a better way of locating the target frame.

At a later stage (30 to 50 epochs), the WIoU-optimised YOLOv5 model starts to show its strengths, with a significant decrease in the loss values, which end up being close to those of the original model, and even slightly lower than the original model at the end of training. This result suggests that although the WIoU optimisation may have triggered higher losses in the early and middle stages of training, after the model has fully learned, it is able to effectively improve the accuracy of target frame localisation and ultimately achieve better model performance.

5.2 COMPARISON OF DETECTION MODELS

Table 1: Comparison of Detection Models

Method	Model Size	Parameters	FLOPs	AP_s	AP_M	AP_L	mAP@0.5-0.95
SSD	118.3M	25.1M	29.20G	0.0206	0.2443	0.5261	0.3168
Efficientdet	15.13M	3.76M	2.50G	-	0.3857	0.5526	0.5890
YOLOv3	119.7M	59.61M	158.00G	0.3851	0.4454	0.4165	0.6098
YOLOv5-s	14.6M	7.19M	17.90G	0.3447	0.5142	0.5132	0.5916
YOLOv5-optimized	16.3M	7.28M	16.8G	0.4319	0.5743	0.5755	0.6754

The table 1 presents a detailed comparison of various object detection models evaluated on the TT100K dataset, including SSD, Efficientdet, YOLOv3, YOLOv5-s, and an optimized version of YOLOv5. The key metrics used for evaluation are

Model Size, Parameters, FLOPs (Floating Point Operations), APS (Average Precision for Small objects), APM (Average Precision for Medium objects), APL (Average Precision for Large objects), and mAP@0.5 (mean Average Precision at IoU threshold of 0.5).

The model sizes vary significantly, with YOLOv3 having the largest model size at 119.7M, while the optimized YOLOv5 is relatively compact at 16.3M, despite the modifications. The parameter count follows a similar trend, with YOLOv3 having the highest number of parameters (59.61M) and Efficientdet-d0 the lowest (3.76M). In terms of computational efficiency, Efficientdet-d0 stands out as the most efficient model with only 2.50B FLOPs, followed by the optimized YOLOv5 at 16.8G FLOPs. On the other hand, YOLOv3 has the highest computational load at 158.00G FLOPs, making it less suitable for resource-constrained environments.

When examining performance across different object sizes, the optimized YOLOv5 shows a significant improvement in detecting small objects, with an AP_s of 0.4319, which is notably higher than that of the other models. For medium-sized objects, the optimized YOLOv5 also leads with an AP_M of 0.5743. Additionally, it achieves the best performance for large objects with an AP_L of 0.5755, indicating that the optimization has effectively enhanced the model's ability to detect objects across various sizes.

Overall, the optimized YOLOv5 achieves the highest mAP@0.5 score of 0.6754, demonstrating a substantial improvement over the original YOLOv5-s, which recorded a mAP@0.5 of 0.5916. This improvement underscores the success of the optimization process in enhancing the model's detection accuracy for objects of all sizes.

5.3 ACTUAL DETECTED PERFORMANCE



(a) Origin images



(b) Detected by Optimized YOLOv5



(c) Detected by Original YOLOv5

Fig 11: Actual Detected Performance

Figure 12 (a) is the original image with labels, (b) is the actual detection effect of this optimised model, (c) is the actual detection effect of the original yolov5 model, it can be seen that the optimised version of YOLOv5 performs better on some images, it is able to detect more categories and the confidence level is not lower than that of the original yolov5 and it is higher and more accurate, while the original version of The original version of YOLOv5 incorrectly recognises images, especially small icons, not very accurately, and has low confidence in recognising complex environments, which is where this optimisation improves.

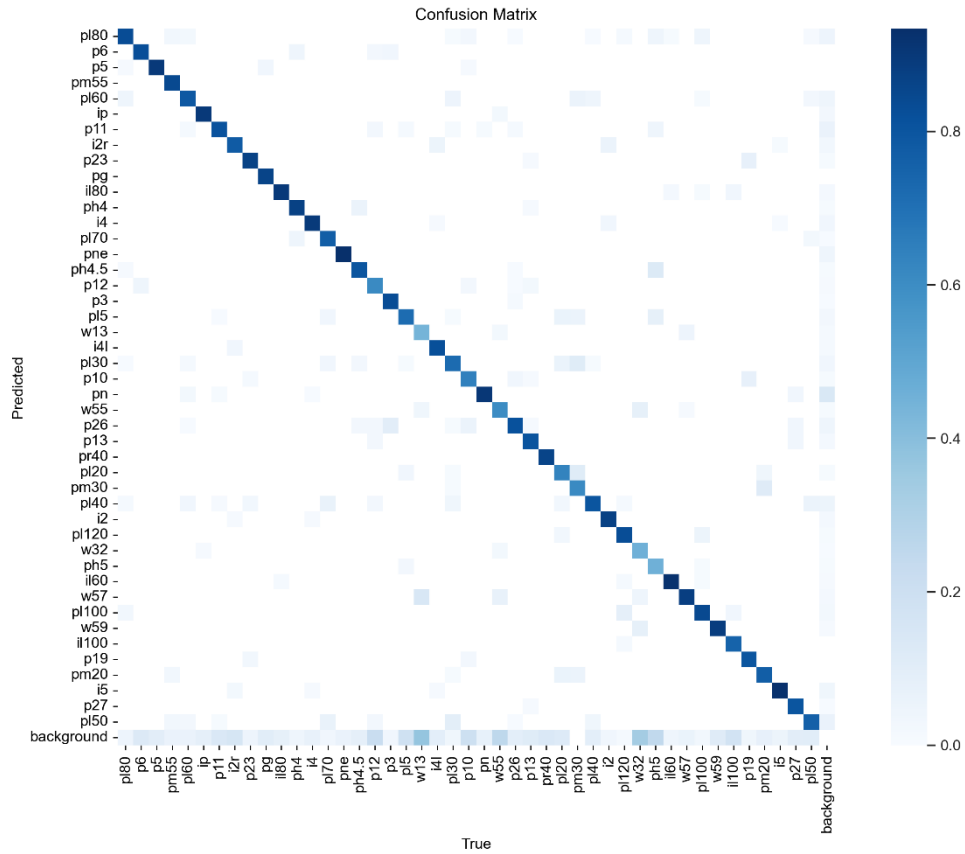


Fig 12 Confusion Matrix of the Optimized Model

Table 2: Performance metrics for different classes

Class	Instances	Precision	Recall	mAP50	mAP50-95
all	1949	0.854	0.756	0.831	0.675
pl80	168	0.801	0.833	0.845	0.667
p6	23	0.812	0.739	0.839	0.642
p5	79	0.934	0.861	0.925	0.707
pm55	32	0.976	0.812	0.899	0.756
pl60	175	0.79	0.783	0.808	0.657
ip	71	0.772	0.761	0.867	0.48
p11	317	0.943	0.732	0.854	0.599
i2r	91	0.759	0.769	0.809	0.582
p23	62	0.946	0.856	0.903	0.697
pg	29	0.883	0.862	0.914	0.659

Class	Instances	Precision	Recall	mAP50	mAP50-95
il80	60	0.8	0.867	0.91	0.749
ph4	24	0.904	0.786	0.92	0.711
i4	166	0.865	0.843	0.908	0.619
pl70	30	0.907	0.733	0.85	0.669
pne	472	0.943	0.912	0.949	0.663
ph4	35	0.759	0.8	0.868	0.703
p12	36	0.779	0.589	0.723	0.595
p3	30	0.96	0.801	0.884	0.696
pl5	110	0.895	0.709	0.806	0.526
w13	27	0.681	0.475	0.506	0.374
i4l	74	0.931	0.797	0.861	0.635
pl30	125	0.821	0.712	0.794	0.614
p10	71	0.879	0.634	0.758	0.593
pn	637	0.932	0.863	0.922	0.656
w55	43	0.856	0.554	0.702	0.519
p26	173	0.875	0.803	0.898	0.714
p13	82	0.913	0.766	0.856	0.579
pr40	37	0.971	0.838	0.905	0.751
pl20	33	0.65	0.545	0.639	0.453
pm30	18	0.711	0.667	0.762	0.657
pl40	290	0.865	0.776	0.849	0.629
i2	88	0.81	0.841	0.898	0.635
pl12	59	0.827	0.811	0.872	0.725
w32	24	0.65	0.458	0.465	0.247
ph5	24	0.65	0.417	0.536	0.336
il60	92	0.966	0.913	0.961	0.759
w57	84	0.886	0.833	0.875	0.59
pl10	134	0.925	0.828	0.916	0.767
w59	43	0.901	0.849	0.859	0.619
il10	28	0.921	0.679	0.786	0.634
p19	25	0.946	0.705	0.854	0.655

Class	Instances	Precision	Recall	mAP50	mAP50-95
pm20	26	0.809	0.808	0.818	0.694
i5	346	0.933	0.908	0.95	0.661
p27	30	0.913	0.767	0.883	0.687
pl50	212	0.794	0.708	0.79	0.564

The performance of You's YOLOv5 model on the TT100K dataset is impressive. Overall, the model has a Precision of 85.4%, a Recall of 75.6%, a mAP@0.5 of 83.1% and a mAP@0.5-0.95 of 67.5%, these metrics show that the model has high accuracy and good recall on the recognition task, and is able to detect and classify the target correctly in most cases.

Looking at the confusion matrix, the model shows strong classification ability on most categories, and the darker part of the diagonal in the confusion matrix shows a high degree of agreement between the predicted and true values. This indicates that the model is able to correctly identify the target category in most cases. However, the lighter off-diagonal elements of the matrix reflect some cases of misclassification, and although the proportion of these misclassifications is relatively low, it suggests that the model's differentiation may need to be further improved for some categories.

In terms of performance on individual categories specifically, the model performs particularly well on the categories pl80, p5, pm55 and p23. The precision and recall of these categories are at a high level, e.g., the p5 category has a precision rate of 93.4% and a recall rate of 86.1%, indicating that the model is very accurate and stable in recognising these categories.

However, the model's performance on some categories is slightly less impressive. For example, the precision and recall of i4l, pl30 and p26 are in the middle level, which indicates that there is still some room for improvement in the recognition of these categories. In addition, the performance of w13, w32 and ph5 is relatively poor, especially w13, with a precision rate of 68.1% and a recall rate of 47.5%, which shows that the model has significant confusion or recognition difficulty in these

categories, but the number of validation sets may be due to the insufficient number of trainings on this category, and there is no over-similarity with other images in these categories.

Overall, your optimised YOLOv5 model performs well on the TT100K dataset, but there is still room for further improvement in recognition on a few categories. It can be considered to increase the training samples for these categories, applying data enhancement techniques, or adjusting the model architecture to improve the overall performance of the model, especially for those categories that are currently weak, in order to achieve a more balanced classification result.

In conclusion, the optimized YOLOv5 model outperforms the other models evaluated on the TT100K dataset. The optimization not only enhances detection capabilities, particularly for small and medium-sized objects, but also maintains computational efficiency. These qualities make the optimized YOLOv5 a strong candidate for applications requiring both high accuracy and efficiency in object detection.

CHAPTER 6 CONCLUSIONS AND RECOMMENDATIONS

6.1 SUMMARY

The focus of this study is to optimise the YOLOv5 model for traffic sign detection using the TT100K dataset. Throughout the study, significant improvements were made to the backbone of the model, including the introduction of advanced convolutional layers and techniques, BiFPN and Wise-IoU. These modifications were aimed at improving the computational efficiency and accuracy of the model, especially in detecting small objects. The optimised model was then critically evaluated against other state-of-the-art models such as EfficientDet and YOLOX, with particular emphasis on its potential for deployment on mobile platforms. The evaluation process involved a thorough analysis of key performance metrics, including mAP, small object detection accuracy and category-specific performance, to provide a comprehensive understanding of the model's strengths and areas for improvement.

6.2 CONCLUSION

This research aims to optimise the YOLOv5 model for traffic sign detection, specifically targeting improvements in computational efficiency and detection accuracy. Modifications implemented in the YOLOv5 architecture, such as merging depth-separable convolution and BiFPN, aim to enhance the model's ability to extract and utilise features more efficiently, while reducing computational overhead. These changes have resulted in an optimised model that not only maintains high accuracy, but is also more suitable for deployment on mobile platforms and other resource-constrained environments.

The enhanced model achieved 67.54% mAP @0.5 on the TT100K dataset, which is an important performance metric given the complexity and variability of the traffic sign detection task. The model is particularly good at detecting small objects, as

evidenced by the 43.19% APS, which is a key factor in real-world traffic scenarios where signs often appear small and at varying distances.

However, the study also highlights a number of challenges, particularly in relation to class imbalances within the dataset. The poor performance of certain classes (e.g., w13, w32, and ph5) may be due to their limited representation in the dataset. This finding emphasises the importance of dataset quality and balance in training effective target detection models. Despite these challenges, the overall performance of the optimised model made significant progress compared to the original YOLOv5 and other state-of-the-art models, especially in small object detection.

Furthermore, this study shows that this optimised YOLOv5 model has the potential to be effectively deployed in real-time applications, such as mobile devices or embedded systems, without sacrificing performance. This feature is crucial for real-world applications where computational resources are limited and real-time processing is critical.

Overall, the study not only met its objectives but also provided valuable insights for optimising target detection models for specific applications such as traffic sign detection. The results of the study have practical implications for the development of intelligent transport systems and the deployment of AI-driven solutions in real-world environments. While the optimised model shows considerable promise, the study also opens up new avenues for further exploration, particularly in addressing the challenges of class imbalance and extending the applicability of the model to other datasets and conditions.

6.3 RECOMMENDATIONS FOR FURTHER RESEARCH

Based on the results of this study, there are several avenues that could be further explored. One important area is to address the problem of category imbalance, which can be mitigated by advanced data enhancement techniques or by exploring alternative loss functions tailored to imbalanced datasets. In addition, future research could focus on real-time deployment, investigating the integration of

optimisation models into traffic management systems. Another potential direction is to test the ability of models to generalise across different datasets and environmental conditions, which would provide insights into their robustness and adaptability. Finally, combining other features such as temporal information or sequence-based analysis could further improve the model's detection accuracy in dynamic environments.

CHAPTER 7 – REFLECTIONS

This study aimed to improve the YOLOv5 model for traffic sign detection and largely achieved the proposed objectives. The first objective was to improve the efficiency and accuracy of the model, which I successfully achieved by integrating advanced techniques. The resulting model exhibits excellent performance, especially in small object detection, which is the focus of this research. The second objective, to evaluate the model's performance using the TT100K dataset, was also achieved. The capabilities of the model were thoroughly evaluated, revealing strengths and areas for further improvement, such as dealing with class imbalance. The third objective, to analyse the deployment potential of the model, was also achieved, confirming the suitability of the model for mobile platforms.

Looking back at the overall progress of the project, it can be said that the research was conducted successfully, with most of the milestones achieved on time. However, a number of challenges emerged, notably class imbalance, which was not fully foreseen in the initial risk analysis. While a number of strategies were adopted to address this issue, these proved only partially effective, suggesting that this area requires further investigation. Despite these challenges, the project was able to pre-emptively identify and mitigate potential computational bottlenecks, thereby ensuring the efficiency of the optimisation model. In summary, whilst the research achieved its main objectives, it also highlighted new challenges and opportunities for future exploration, emphasising the importance of continuous learning and adaptation during the research process.

CHAPTER 8 – REFERENCES

- [1] Y.Zhu and W.Q.Yan, 'Traffic sign recognition based on deep learning', *Multimedia Tools and Applications*, vol. 81, no. 13, pp. 17779–17791, May 2022.
- [2] T. Meng, G. Ghiasi, R. Mahjourian, Q. V. Le, and M. Tan, 'Revisiting Multi-Scale Feature Fusion for Semantic Segmentation', *arXiv [cs.CV]*. 2022.
- [3] D. Tabernik and D. Skočaj, 'Deep learning for large-scale traffic-sign detection and recognition', *IEEE transactions on intelligent transportation systems*, vol. 21, no. 4, pp. 1427–1440, 2019
- [4] S. Qu, X. Yang, H. Zhou, and Y. Xie, 'Improved YOLOv5-based for small traffic sign detection under complex weather', *Scientific Reports*, vol. 13, no. 1, p. 16219, Sep. 2023.
- [5] X. Yuan, J. Guo, X. Hao, and H. Chen, 'Traffic Sign Detection via Graph-Based Ranking and Segmentation Algorithms', *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 45, no. 12, pp. 1509–1521, 2015.
- [6] C. Liu, S. Li, F. Chang, and Y. Wang, 'Machine vision based traffic sign detection methods: Review, analyses and perspectives', *IEEE Access*, vol. 7, pp. 86578–86596, 2019.
- [7] R. Zhao, S. H. Tang, J. Shen, E. E. B. Supeni, and S. A. Rahim, 'Enhancing autonomous driving safety: A robust traffic sign detection and recognition model TSD-YOLO', *Signal Processing*, vol. 225, p. 109619, 2024.
- [8] A. Avramović, D. Sluga, D. Tabernik, D. Skočaj, V. Stojnić, and N. Ilc, 'Neural-Network-Based Traffic Sign Detection and Recognition in High-Definition Images Using Region Focusing and Parallelization', *IEEE Access*, vol. 8, pp. 189855–189868, 2020.
- [9] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05). Ieee, 2005, 1: 886-893.
- [10] Lowe D G. Object recognition from local scale-invariant features[C]//Proceedings of the seventh IEEE international conference on computer vision. Ieee, 1999, 2: 1150-1157.
- [11] Ahonen, T., Hadid, A., Pietikäinen, M.: Face recognition with local binary patterns. In: ECCV. pp. 469–481. Springer (2004) 1

-
- [12] T. Ojala, M. Pietikainen and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 971-987, July 2002, doi: 10.1109/TPAMI.2002.1017623.3
- [13] D. Neupane and J. Seok, 'A review on deep learning-based approaches for automatic sonar target recognition', Electronics, vol. 9, no. 11, p. 1972, 2020.
- [14] Girshick R, Donahue J, Darrell T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[J]. IEEE Computer Society, 2013.
- [15] R. Girshick, "Fast R-CNN," 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1440-1448, doi: 10.1109/ICCV.2015.169..
- [16] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6):1137-1149.
- [17] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, 'You Only Look Once: Unified, Real-Time Object Detection', in Computer Vision & Pattern Recognition, 2016.
- [18] C. Zhang, X. Yue, R. Wang, N. Li, and Y. Ding, 'Study on Traffic Sign Recognition by Optimized Lenet-5 Algorithm', International Journal of Pattern Recognition and Artificial Intelligence, vol. 34, no. 01, p. 2055003, 2020.
- [19] Changzhen X, Cong W, Weixin M, et al. A traffic sign detection algorithm based on deep convolutional neural network[C]//2016 IEEE International Conference on Signal and Image Processing (ICSIP). IEEE, 2016: 676-679.
- [20] J. Xing, M. Nguyen, and W. Q. Yan, 'The Improved Framework for Traffic Sign Recognition Using Guided Image Filtering', SN Computer Science, vol. 3, 11 2022.
- [21] A. Avramović, D. Sluga, D. Tabernik, D. Skočaj, V. Stojnić and N. Ilc, "Neural-Network-Based Traffic Sign Detection and Recognition in High-Definition Images Using Region Focusing and Parallelization," in IEEE Access, vol. 8, pp. 189855-189868, 2020, doi: 10.1109/ACCESS.2020.3031191.
- [22] Y. Wang, G. Yang, and J. Guo, 'Vehicle detection in surveillance videos based on YOLOv5 lightweight network', Bulletin of the Polish Academy of Sciences Technical Sciences, 2023.

-
- [23] A. Wang, T. Peng, H. Cao, Y. Xu, X. Wei, and B. Cui, 'TIA-YOLOv5: An improved YOLOv5 network for real-time detection of crop and weed in the field', *Frontiers in Plant Science*, vol. 13, p. 1091655, 2022.
- [24] J. Zhang, Z. Chen, G. Yan, Y. Wang, and B. Hu, 'Faster and Lightweight: An Improved YOLOv5 Object Detector for Remote Sensing Images', *Remote Sensing*, vol. 15, no. 20, 2023.
- [25] L. Zhu, F. Lee, J. Cai, H. Yu, and Q. Chen, 'An improved feature pyramid network for object detection', *Neurocomputing*, vol. 483, pp. 127–139, 2022.
- [26] M. Tan, R. Pang and Q. V. Le, "EfficientDet: Scalable and Efficient Object Detection," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020, pp. 10778-10787, doi: 10.1109/CVPR42600.2020.01079.
- [27] J. Zheng, H. Liu, Q. He, and J. Hu, 'GEB-YOLO: a novel algorithm for enhanced and efficient detection of foreign objects in power transmission lines', *Scientific Reports*, vol. 14, no. 1, p. 15769, Jul. 2024.
- [28] T. Meng, G. Ghiasi, R. Mahjourian, Q. V. Le, and M. Tan, 'Revisiting Multi-Scale Feature Fusion for Semantic Segmentation', *arXiv [cs.CV]*. 2022.
- [29] Khanam, R., and Hussain, M., "What is YOLOv5: A deep look into the internal features of the popular object detector," in *arXiv: 2407.20892 [cs.CV]*, Jul. 2024. Accessed: 2024-07-31. [Online]. Available: <http://arxiv.org/abs/2407.20892>
- [30] W. Zhan et al., 'An improved YOLOv5 real-time detection method for small objects captured by UAV', *Soft Computing*, vol. 26, no. 1, pp. 361–373, Jan. 2022.
- [31] Zheng Z, Wang P, Liu W, Li J, Ye R, Ren D. Distance-IoU loss: Faster and better learning for bounding box regression. In: *Proceedings of the AAAI conference on artificial intelligence*. vol. 34; 2020. p. 12993–13000.
- [32] Z. Tong, Y. Chen, Z. Xu, and R. Yu, 'Wise-IoU: Bounding Box Regression Loss with Dynamic Focusing Mechanism', *arXiv [cs.CV]*. 2023.
- [33] Y. Liu et al., 'Research on Infrared Dim Target Detection Based on Improved YOLOv8', *Remote Sensing*, vol. 16, no. 16, p. 2878, 2024.
- [34] Q. Liu et al., 'DSW-YOLOv8n: A new underwater target detection algorithm based on improved YOLOv8n', *Electronics*, vol. 12, no. 18, p. 3892, 2023.

-
- [35] R. Ormston, L. Spencer, M. Barnard, and D. Snape, 'The foundations of qualitative research', *Qualitative research practice: A guide for social science students and researchers*, vol. 2, no. 7, pp. 52–55, 2014.
- [36] Y. Liu, J. Peng, J.-H. Xue, Y. Chen, and Z.-H. Fu, 'TSingNet: Scale-aware and context-rich feature learning for traffic sign detection and recognition in the wild', *Neurocomputing*, vol. 447, pp. 10–22, 2021.
- [37] L. Aziz, M. S. B. H. Salam, U. U. Sheikh, and S. Ayub, 'Exploring deep learning-based architecture, strategies, applications and current trends in generic object detection: A comprehensive review', *Ieee Access*, vol. 8, pp. 170461–170495, 2020.
- [38] F. Lei, F. Tang, and S. Li, 'Underwater target detection algorithm based on improved YOLOv5', *Journal of Marine Science and Engineering*, vol. 10, no. 3, p. 310, 2022.
- [39] K. Li, Y. Wang, and Z. Hu, 'Improved YOLOv7 for Small Object Detection Algorithm Based on Attention and Dynamic Convolution', *Applied Sciences*, vol. 13, no. 16, 2023.
- [40] Y. Fang, Y. Ma, X. Zhang, and Y. Wang, 'Enhanced YOLOv5 algorithm for helmet wearing detection via combining bi-directional feature pyramid, attention mechanism and transfer learning', *Multimedia Tools and Applications*, vol. 82, no. 18, pp. 28617–28641, 2023.
- [41] Q. Ma, S. Jin, G. Bian, and Y. Cui, 'Multi-Scale Marine Object Detection in Side-Scan Sonar Images Based on BES-YOLO', *Sensors*, vol. 24, no. 14, p. 4428, 2024.
- [42] N. N. A. Mangshor, N. S. Saharuddin, S. Ibrahim, A. F. A. Fadzil, and K. A. F. A. Samah, 'A real-time speed limit sign recognition system for autonomous vehicle using SSD algorithm', in *2021 11th IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*, 2021, pp. 126–130.
- [43] J. Redmon AF. YOLOv3: An Incremental Improvement. arXiv preprint arXiv:1804.02767. 2018.

APPENDICES

PROJECT MANAGEMENT

Table 3 The Original Gantt chart

Project Management				
	Mission Statement	Expected working days	Start Date	End Date
1	Literature Review	3	2024-07-02	2024-07-04
2	Methodology	11	2024-07-06	2024-07-16
2.1	Research design	6	2024-07-06	2024-07-11
2.2	Data collection	2	2024-07-12	2024-07-13
2.3	Experiment design	3	2024-07-14	2024-07-16
3	Develop experiment	20	2024-07-17	2024-08-05
3.1	Code design	5	2024-07-17	2024-07-21
3.2	Parameter selection and adjusting	1	2024-07-22	2024-07-22
3.3	Run code	14	2024-07-23	2024-08-05
4	Evaluating model performance	3	2024-08-06	2024-08-08
5	Writing	7	2024-08-09	2024-08-15

Table 4 The Actual Gantt chart

Project Management				
	Mission Statement	Expected working days	Start Date	End Date
1	Literature Review	3	2024-07-02	2024-07-04
2	Methodology	11	2024-07-06	2024-07-16
2.1	Research design	6	2024-07-06	2024-07-11
2.2	Data collection	2	2024-07-12	2024-07-13
2.3	Experiment design	3	2024-07-14	2024-07-16
3	Develop experiment	22	2024-07-17	2024-08-07
3.1	Code design	5	2024-07-17	2024-07-21
3.2	Parameter selection and adjusting	1	2024-07-22	2024-07-22
3.3	Run code	16	2024-07-23	2024-08-07
4	Evaluating model performance	3	2024-08-08	2024-08-10
5	Writing	7	2024-08-11	2024-08-17

In the actual progress of this project, most of the content proceeded as scheduled, and the relevant investigations and literature searches for the project went very smoothly. However, unexpected situations occurred during the execution of the code. Specifically, during the "Code design" task in the "Develop experiment" phase, I encountered an issue where the dataset format was incompatible with the YOLO model. This issue is due to insufficient consideration of the compatibility between dataset format and model during the preliminary research of the project. To solve this problem, I had to spend extra time understanding and converting the dataset format to meet the requirements of the YOLO model.

After converting the dataset format, we encountered a situation where the model was unable to read the data correctly. After careful inspection, I found that the problem lies in some details during the data conversion process, such as path errors

or non-standard formats. To address these issues, I thoroughly reviewed and revised the data conversion process to ensure that all data paths and formats meet the requirements of the model. These additional tasks resulted in the delay of the 'Code design' task, which in turn affected the progress of the entire project.

In response to these challenges, I have added considerations for data compatibility and technical support in the risk analysis and developed corresponding risk minimization strategies. I realize that strengthening technical research and preparation in the early stage is the key to avoiding similar problems. Therefore, I suggest that in future projects, more attention should be paid to the establishment of technical preparation and backup plans to improve the adaptability and flexibility of the project.

Despite facing these challenges, I ultimately overcame them through the joint efforts of the team and effective project management, ensuring the achievement of project goals. These lessons not only provided me with valuable project management experience, but also strengthened my understanding of project risk management and technical problem-solving skills. In future projects, I will pay more attention to preliminary research and technical support to improve the success rate of the project.

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:	Shuo Zhang
2	Student Number:	2304515
3	Email address:	2304515@student.uwtsd.ac.uk
4	Programme of Study:	MSc 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 codes 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:	Enhancing Traffic Sign Recognition with Optimized YOLOv5: Addressing Environmental Variability for Improved Road Safety		
2	Proposed start date:	19 th .2.2024	Proposed end date:	10 th .05.2024
	Introduction to the Research (maximum 300 words in each section) Ensure that you write for a <u>Non-Specialist Audience</u> when outlining your response to the three points below: <i>Purpose of Research Activity</i> <i>Proposed Research Question</i> <i>Aims of Research Activity</i> <i>Objectives of Research Activity</i>			

	Demonstrate, briefly, how Existing Research has informed the proposed activity and explain • <i>What the research activity will add to the body of knowledge</i> • <i>How it addresses an area of importance.</i>
3	Purpose of Research Activity This study aims to optimize the structure of YOLOv5 to enhance the accuracy and robustness of traffic sign recognition under diverse environmental conditions. By refining YOLOv5, the study seeks to improve the algorithm's ability to accurately identify and classify traffic signs irrespective of lighting changes, weather conditions, and occlusions. (this box should expand as you type)
4	Research Question The challenge lies in the accuracy and robustness of existing traffic sign recognition systems, particularly under diverse environmental conditions. Factors such as lighting variations, weather conditions, and occlusions can significantly impact the performance of these systems, leading to potential safety hazards on the roads. (this box should expand as you type)
5	Aims of Research Activity This study aims to address the shortcomings of current traffic sign recognition systems by optimizing the structure of YOLOv5. By doing so, the goal is to enhance the algorithm's accuracy and robustness, thereby improving its ability to identify and classify traffic signs irrespective of environmental changes. (this box should expand as you type)
6	Objectives of Research Activity Optimize the structure of YOLOv5 to improve traffic sign recognition accuracy under diverse environmental conditions. Evaluate the impact of structural modifications, including changes in network layers and activation functions, on the performance of YOLOv5 and enhancing Traffic Sign Recognition with Optimized YOLOv5: Addressing Environmental Variability for Improved Road Safety. Investigate how different training datasets comprising varied traffic sign images influence the model's recognition capabilities. Assess the enhanced model's performance using metrics such as accuracy, recall, precision, inference speed, and model size. (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	This should describe how you intend to collect data. It should not include a discussion of the theoretical basis for your data collection methods. Please note, that if you intend to collect any audio/video recordings of interviews with participants then these will be are classified as Personal Data under GDPR/DPA2018. If you intend to use these then note this in section H. (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. <i>(this box should expand as you type)</i>
--	---

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: N/A <i>(this box should expand as you type)</i>	How you will mitigate the risk to participants: <i>(this box should expand as you type)</i>
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	
	<i>(this box should expand as you type)</i>	
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: N/A <i>(this box should expand as you type)</i>	How you will mitigate the risk to the investigator: <i>(this box should expand as you type)</i>
4	University/institutional risks For example: adverse publicity, financial loss, data protection	
	Risk to the University: N/A <i>(this box should expand as you type)</i>	How you will mitigate the risk to the University: <i>(this box should expand as you type)</i>
5	Environmental risks For example: accidental spillage of pollutants, damage to local ecosystems	
	Risk to the environment: N/A <i>(this box should expand as you type)</i>	How you will mitigate the risk to environment: <i>(this box should expand as you type)</i>

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?
---	--

	(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.
	(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.
	(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 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	(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:		
	<i>All data will be securely stored in password-protected cloud storage within the University's Office 365 system, ensuring confidentiality. Audio and visual data will undergo transcription, and participants will have the opportunity to verify the accuracy of the transcriptions. Any USB sticks utilized for data storage or transfer will be password protected. Each participant will receive a unique identifier to maintain confidentiality, and this information will be securely stored within the password-protected folder.</i> <i>(this box should expand as you type)</i>		

	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):		
	<i>(this box should expand as you type)</i>		
20	List all who will have access to the data generated by the research activity:		
	Typically, the lead investigator, potentially serving as the project supervisor, along with a representative from any collaborating industrial partner, would be involved. Additionally, there might be an external examiner or a second marker included in the mix. <i>(this box should expand as you type)</i>		
21	List who will have control of, and act as custodian(s) for, data generated by the research activity:		
	The Principal Investigator will serve as the overseer and guardian of the data throughout the research process, tasked with ensuring its security, accuracy, and adherence to regulations. <i>(this box should expand as you type)</i>		
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.		
	To safeguard research data, we utilize advanced encryption and store it securely on the University's password-protected Office 365 cloud storage system. For data transfer and storage, USB drives, also password protected, are employed. After project completion, these drives undergo thorough reformatting to ensure complete data erasure, in line with retention and destruction policies. <i>(this box should expand as you type)</i>		
22	Confirm that you have read the UWTSD guidance on data management (see https://www.uwtss.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	☒	

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:	Shuo Zhang	Date: 10 th .May.2024
2	Director of Studies/Supervisor:		Date: 10 th .May.2024
3	Signature:	Shuo Zhang	

LOGBOOK

Date	Daily Activities	Thought Trails
2024/7/2	Started background research for the project, read relevant literature	Looking for gaps in existing research
2024/7/3	Identified shortcomings of YOLOv5 in TSR	Seeking methods for optimization
2024/7/4	Found ways to optimize	Analyzing and beginning methodological research
2024/7/5	Began exploring research methods	Contemplating how to conduct methodology
2024/7/6	Continued researching methodology, refined research design	Considering the scientific and practical aspects of research design
2024/7/7	Completed the first draft of the research design, started planning for data collection	Assessing the feasibility and effectiveness of data collection
2024/7/12	Collected relevant data, prepared experimental materials	Evaluating data quality to ensure the accuracy of the experiment

2024/7/17	Designed experimental scheme, prepared experimental equipment	Considering the safety and repeatability of the experiment
2024/7/22	Conducted preliminary experiments, adjusted experimental parameters	Reflecting on potential issues during the experimental process and solutions
2024/7/16	Completed experimental design, started coding work	Contemplating how to optimize code structure and improve efficiency
2024/7/17	Coding work in progress, encountered technical difficulties	Exploring solutions and improving programming skills
2024/7/18	Resolved coding issues, continued to refine the code	Considering the maintainability and scalability of the code
2024/7/20	Completed code design, began parameter selection and adjustment	Considering the impact of parameters on experimental results
2024/7/20	Completed parameter adjustment, prepared to run the experiment	Reflecting on the rationality of the experimental design
2024/8/6	Ran the experiment, collected experimental data	Observing experimental phenomena and recording data
2024/8/8	Analyzed experimental data, preliminarily evaluated model performance	Contemplating the scientific significance behind the data
2024/8/11	Completed the first draft of the literature review, started writing the methodology chapter	Contemplating how to clearly describe research methods and processes

2024/8/12	Wrote the experimental results chapter, prepared charts and illustrations	Considering how to make charts clear and intuitive
2024/8/13	Completed the experimental results chapter, started writing the discussion chapter	Contemplating how to connect experimental results with existing research
2024/8/14	Wrote the discussion chapter, deeply analyzed the significance of experimental results	Considering how to propose new insights and suggestions
2024/8/15	Completed the discussion chapter, started writing the conclusion and future work chapter	Contemplating how to summarize the main findings and contributions of the research
2024/8/16	Wrote the conclusion and future work chapter, prepared for final submission	Considering how to propose directions for future research
2024/8/17	Completed the final manuscript, conducted self-review and revision	Contemplating how to improve writing quality and ensure accuracy of content

GLOSSARY

BiFPN Bidirectional Feature Pyramid Network

FPN Feature Pyramid Networks

IoU IoU is a standard for measuring the accuracy of detecting corresponding objects in a specific dataset. IoU is a simple measurement standard that can be used to measure any task that generates a bounding box in the output

WIoU Wise-IoU

GIoU GIoU is a metric and loss function based on IoU, which can solve the problems of scale invariance and alignment of IoU.

mAP Mean Average Precision is a commonly used evaluation metric in object detection models

SSD Single Shot MultiBox Detector

Confusion Matrix A confusion matrix is a special type of contingency table that has two dimensions (actual and predicted) and sets of the same categories in both dimensions.