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

School of
Applied Computing

DETECTION OF FUSARIUM HEAD BLIGHT ON WHEAT SPIKELETS USING A MULTI-SCALE FEATURE FUSION CNN MODEL

Ze Wu

Supervisor Dr. Seena Joseph

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

September 2024

Declaration of Originality

I, Ze Wu 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 : .....

24 Aug, 2024

Date : .....

FORM OF CONSENT

I, Ze Wu, hereby consent that my Project, submitted in candidature for the MSc Software Engineering and Artificial Intelligence, if successful, may be made available for inter-library loan or photocopying (subject to the law of copyright), and that the title and abstract may be made available to outside organisations.

Signed : .....

24 Aug, 2024

Date : .....

ABSTRACT

Fusarium head blight (FHB) is a globally significant fungal disease that severely impacts wheat crops, leading to reduced yields, degraded grain quality, and the accumulation of harmful mycotoxins. These mycotoxins pose severe threats to human and animal health and result in substantial economic losses. Accurate and efficient assessment of FHB disease phenotypes is crucial for developing resistant wheat varieties through breeding. However, current methods for collecting and analyzing wheat phenotypic data are time-consuming, labor-intensive, and often imprecise, particularly at the spikelet level, where disease symptoms are more nuanced and challenging to detect. Hence this study addresses these challenges by proposing a novel lightweight object detection model based on multi-scale feature fusion, specifically designed to detect FHB at the spikelet level in wheat. The proposed model leverages the advanced YOLOv9 framework, integrating the Multi-Scale Feature Enhancement and Fusion (MSFEF) module, to significantly enhance the accuracy of detecting small and subtle disease features in complex field environments. The performance of the developed model is evaluated using a self-constructed dataset comprising 620 annotated RGB images. Results show that the proposed model achieves a mean Average Precision (mAP) of 90.6%, outperforming state-of-the-art models such as YOLOv9-C and YOLOv10-S while maintaining real-time performance with 294 FPS.

Additionally, the model excels in detecting diseased spikelets, achieving an Average Precision (AP) of 92%, and shows high robustness in recognizing healthy

and infected spikelets. Regarding disease phenotype extraction, the model's correlation coefficients with manual detection for disease spikelet rate and FHB severity are 0.81 and 0.75, respectively, underscoring its significant potential for practical application in breeding and phenotype collection. This study introduces a novel, fine-grained detection method for wheat FHB disease and offers practical solutions to enhance the efficiency of wheat resistance breeding and plant disease management.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I, Ze Wu, express my sincere gratitude to my supervisor Dr. Seena Joseph of Trinity Saint David University of Wales for her invaluable support and guidance throughout the project. Her expert advice on academic writing, critical thinking and research methodology significantly helped me with this study. Her continuous feedback and encouragement were essential for me to refine my ideas and complete this complex study, especially in developing the model and analyzing the results.

I would also like to thank the Wuhan Bioinformatics Laboratory for providing access to the data necessary for this project. Their support was essential in enabling me to construct a high-quality research dataset. Additionally, I am grateful to the laboratory for providing a remote server and high-performance graphics cards, which were critical in meeting the computational demands of this study.

My sincere thanks also go to the University of Wales Trinity Saint David for granting me access to its online library, which offered extensive professional literature on Fusarium head blight detection and object detection, making a solid theoretical foundation for my subsequent research.

CHAPTER 1 – INTRODUCTION

1.1 RESEARCH BACKGROUND

Fusarium head blight (FHB), is one of the common fungal diseases in wheat, can cause serious damage to wheat yield [1]. Fusarium graminearum (F. *graminearum*) is one of the main pathogens causing FHB [2], which is known to produce various Fusarium mycotoxins, including deoxynivalenol (DON), trichothecenes, and beauvericin, posing a significant threat to the health of humans and livestock [3]. Accurate identification of infected wheat spikes at the early stage of FHB can greatly contribute to improving the current situation and minimizing associated losses. Currently, FHB management in wheat primarily relies on main approaches including chemical control, biological control, and the development of resistant varieties through plant breeding [4], [5], [6]. It is emphasized that resistance breeding is a core and effective approach in the prevention and control of FHB. However, as a prerequisite for breeding work, the collection and analysis of disease phenotypic information are fundamentally important tasks [7], [8].

1.1.1 FHB Phenotype

When assessing the severity of Fusarium Head Blight (FHB), the Diseased Spikelet Rate (DSR) is a phenotypic indicator in wheat breeding. It measures the level of disease impact on wheat spikelets after FHB infection. The Diseased Spikelet Rate is essential in breeding programs that focus on developing FHB-resistant wheat varieties, as it is used to assess the resistance of breeding populations, particularly in terms of limiting the disease's dissemination within spikelets (known as Type II resistance) [9]. The measurement of DSR allows researchers to assess the efficacy of resistance characteristics, ascertain the severity of the disease following the Chinese Standard GB/T 15796-2011, and analyze the response of wheat plants to FHB. This information is indispensable for comprehending the prevalence of FHB in wheat populations, providing a basis for the enhancement of wheat varieties, and directing disease management practices. In addition, DSR data is advantageous for genome-wide association studies (GWAS) that are designed to identify FHB resistance genes and Quantitative Trait Loci (QTLs) [6].

1.1.2 Wheat FHB Detection Technologies

Image-based phenotypic extraction of plant diseases is becoming increasingly important in plant phenomics research [10]. Digital image processing can be employed to acquire and analyze phenotypic information about plant diseases precisely and efficiently. Furthermore, imaging technology offers dependable data support for agricultural automation. The growth of emerging sensor technologies has driven the update of plant image acquisition devices, with diverse imaging methods including high-resolution imaging, hyperspectral imaging, thermal imaging, etc. These technologies can extract crop disease characteristics from multiple dimensions and utilize modern computational methods such as machine learning for crop monitoring, analysis, and disease diagnosis [11], [12].

Hyperspectral imaging technology is a high-precision, non-destructive tool for disease assessment, capable of extracting key biomarkers based on spectral data [13], [14]. Chlorophyll fluorescence imaging technology is widely applied in the phenotypic characteristic analysis of diseases such as FHB, offering a new dimension for disease monitoring and assessment [15]. Moreover, combined with unmanned aerial vehicle (UAV) technology, it enables efficient crop disease monitoring and identification over large planting areas [16]. However, these high-precision imaging technologies impose higher demands on data acquisition devices, not only increasing costs but also requiring strong dependence on the collection environment. The complexity of manual feature extraction, calibration, and the accuracy of data quality and quantitative analysis are affected by environmental factors.

The development of deep learning introduced by LeCun, et al. [17] provides strong support for the research and application of computer vision in the field of crop phenotype and disease recognition [18]. Image processing technology based on deep learning, with its high precision and high throughput, has become a hot topic in plant disease phenotypic research [19]. The introduction of Convolutional Neural Networks (CNNs), especially in large-scale image data processing and analysis, has demonstrated their powerful capability in identifying abnormalities in crop images and determining potential crop disease symptoms [20]. Deep

convolutional networks like AlexNet [21] are used to extract specific disease feature information and combined with machine learning algorithms such as random forests to classify disease severity [22]. Additionally, the end-to-end Blendmask model proposed in Gao, et al. [23] can simultaneously segment wheat spikes and disease areas, thereby accurately assessing disease severity.

1.1.3 Object Detection on FHB

Object detection algorithms, due to their high accuracy and relatively low dataset annotation difficulty, are widely used in phenotypic extraction. Including both one-stage and two-stage object detection techniques, two-stage object detection algorithms hold great potential in phenotypic extraction applications. One-stage object detection techniques, which can efficiently extract complex image information and are faster and more accurate compared to two-stage detection, are widely used in wheat FHB detection tasks due to their fast inference speed and potential for real-time detection.

The end-to-end improved YOLOv5 model used by Zhang, et al. [24] for extracting wheat spikes, then disease assessment through threshold segmentation and a random forest binary classifier. In Hong, et al. [25], a lightweight model based on the Mobile-Net and YOLOv4 framework can be deployed on UAVs for assessing wheat FHB conditions in field environments. In the study of Bao, et al. [26], an improved YOLOv5s combined with data augmentation strategies was designed for UAV-collected RGB image recognition of FHB disease spikes, YOLOv7-MA network constructed by Meng, et al. [27] with the design of Microscale Detection Layer and integration of CBAM attention mechanism proposed by [28] for spike detection. GSEYOLOX-s combines SimAM proposed by Yang, et al. [29] and ghost convolution designed by Han, et al. [30] to improve YOLOX-s in training and predicting on spike data labeled for disease severity grading [31]. Although end-to-end object detection offers advantages in accuracy and lightweight, the extraction and prediction of disease phenotypic data pose higher demands on such models.
Multi-scale Feature Fusion on Object Detection

Currently, the most widely used real-time detectors for plant disease are still the YOLO series detectors [32], which typically use modified Path Aggregation

Network (PAN) [33] or Feature Pyramid Network (FPN) [34] as the primary mechanisms for feature fusion [35]. However, this feature fusion mechanism suffers from significant information loss in high-level feature maps and consumes much memory during computation, resulting in low processing efficiency [36, 37].

Specifically, Zhao, et al. [38] proposed BiTNet, a lightweight object detection network combining Transformer and Bi-FPN. It designs Efficient Transformer Blocks (ETB) and Efficient Convolution Aggregation Blocks (ECAB) to extract image features, where the bidirectional FPN supplements deep and shallow feature representations from bottom-up and top-down directions. The ECAB, using a single aggregation operation to process multiple feature maps, enhances feature extraction and reduces computational complexity and Memory Access Costs (MAC), thus alleviating the efficiency issues of FPN in feature computation. However, ECAB's drawback is its failure to consider the differing importance of deep and shallow features. ELAN is a structure similar to ECAB but considering this aspect [39]. The critical difference between ECAB and ELAN is that ECAB only aggregates feature maps within layers, emphasizing local feature learning.

In contrast, ELAN aggregates inputs from different convolutional layers, focusing on global feature information. ELAN's structure references VoVNet and CSPVoVNet, combining VoVNet's convolutional block stacking for deep feature extraction and CSPNet's efficient feature processing structure [40], balancing the increased computational cost due to feature fusion in VoVNet and reducing redundant information in the gradient flow through cross-stage partial connections [41]. ELAN is applied in YOLOv9, where the original convolutional blocks are replaced with arbitrary computational blocks, resulting in the Generalized Efficient Layer Aggregation Network (GELAN) network [42].

Despite the significant advancements in the accuracy of current wheat FHB object detectors [43], the detection scale and phenotype extraction remain at the spike level, leading to missed and false detections. In contrast, detecting FHB at the spikelet level can extract more detailed and precise disease phenotypes. However, the smaller size of spikelets and the diversity of infection symptoms impose higher

performance requirements on object detection methods. Consequently, it is imperative to develop a highly precise spikelet-based object detection model.
Hence, this study presents a novel object detection algorithm that utilizes multi-scale feature fusion to identify FHB in wheat spikelets. To further improve the model's performance, this model is constructed on the sophisticated YOLOv9 framework, which includes a multi-scale feature enhancement module and an enhanced feature pyramid structure. This model establishes a strong foundation for identifying and controlling FHB in wheat and disease monitoring and management by accurately extracting disease phenotypes, conducting assessments, and classifying and locating healthy and diseased spikelets.

1.2 RESEARCH PROBLEM STATEMENT

In order to prevent and control FHB, it is essential to have early symptom information about diseases [44]. Nevertheless, the subtle differences and varying susceptibility levels present a significant challenge in the more refined detection of diseased spikelets, thereby increasing the difficulty of recognition [16].

The existing YOLOv9 employs the classical FPN structure, which is constrained by specific feature layers and encounters difficulties in effectively capturing rich contextual information to represent and detect objects with substantial scale variations [42]. Detecting spikelets necessitates multi-scale feature fusion mechanisms to extract spikelet features and enhance contextual information, as wheat spikelets are smaller in scale than entire wheat spikes in field-collected data. In addition, the multi-scale feature hierarchy is restricted to feature layers, which complicates the concurrently incorporating objects' features at varying scales into a feature map [45].

The small size of wheat spikelets, the noise interference of complex field backgrounds, and the diverse disease symptoms can present challenges in accurately detecting wheat in FHB [19]. These detection issues impact the extraction and transmission of fine-grained spikelet features with this detection using CNN models [43].

1.3 RESEARCH AIM

This study aims to develop a light-weight object detection CNN network based on multi-scale feature fusion for detecting FHB infection and phenotypic extraction in the wheat spikelet.

1.4 RESEARCH OBJECTIVES

Obj1 - To critically analyze the current literature to understand trends in multi-scale feature extraction and fusion for plant disease object detection models.
Obj2 - To develop a multi-scale feature fusion object detection model based on CNN for detecting FHB diseased spikelets.
Obj3 - To evaluate the performance of the developed model on self-constructed dataset by comparing it against state-of-the-art models using well-known evaluation metrics.

1.5 CONTRIBUTION OF THE STUDY

- This study introduces a novel object detection method for FHB detection and phenotype extraction focusing on wheat spikelets. This method can accomplish a more precise and detailed identification of the disease phenotype by emphasizing the smaller scale of spikelets rather than entire spikes.
- This study implements a Multi-Scale Feature Enhancement and Fusion (MSFEF) module to identify wheat spikelets in intricate field conditions. This module enhances the model's capacity to differentiate spikelets from noise backgrounds by enhancing and fusing multi-scale features.
- This study integrates the MSFEF module into the advanced YOLOv9 framework, resulting in a resilient model that boosts detection accuracy. The enhanced performance of this integrated model in real-world agriculture scenarios is demonstrated through a systematic evaluation of it against state-of-the-art models using standard evaluation metrics.
- This study provides a dataset of wheat spikes infected with FHB obtained from greenhouse scenes. The dataset consists of an RGB data collection of the infected spikes, and each image is accompanied by meticulous small-scale, fully supervised manual annotations at the spikelet level, which were conducted under the guidance of experts.

1.6 STURCTURE OF THE STUDY

This study is divided into three main parts. Firstly, it involves learning the theoretical knowledge of FHB detection in wheat and object detection algorithms. Thus, by analyzing the current literature for multi-scale feature extraction and fusion. This foundational work will support designing and testing an object detection model capable of detecting FHB in wheat spikelets.

The second part addresses the specific research aims and objectives of the study. It involves the design of the mechanism for multi-scale feature extraction and fusion, the development of an object detection model for identifying FHB in wheat spikelets, and the extraction of disease phenotypes by the proposed model.

The final part discusses the results and predictions of the model, evaluates the application of the algorithm in disease phenotype extraction, and compares it against state-of-the-art models using evaluation metrics. Finally, it analyzes and summarizes the directions for future work and development.

CHAPTER ONE – INTRODUCTION summarizes the background knowledge on FHB detection in wheat, thereby clarifying the research problems for this study. It outlines the research aim, refines the research objectives, and provides a structural plan for the study.

CHAPTER TWO – REVIEW OF LITERATURE focuses on previous research and current advancements in FHB controlling and detection, YOLO detectors, and multi-scale feature fusion of the object detections. This chapter synthesizes and critically analyzes the literature and further clarifies the design of mechanisms and models during the study's implementation.

CHAPTER THREE – RESEARCH METHODOLOGY outlines the research design and methodology employed in the study. It details the architecture of the proposed model, the Multi-Scale Feature Enhancement and Fusion (MSFEF) module, and the evaluation metrics used to assess the model's performance. The chapter also

discusses the data collection process, including the dataset’s sources and the software and hardware used in the experiments.

CHAPTER FOUR – EXPERIMENTS AND RESULTS ANALYSIS presents the experimental setup, the results of the experiments, and a thorough analysis of the findings. It comprises performance comparisons between the proposed and state-of-the-art models and an assessment of the efficacy of various components within the proposed model. Visualizations of the feature distribution, model predictions, and correlation and regression analyses of the extracted phenotypes are also included in the chapter.

CHAPTER FIVE – CONCLUSION, LIMITATION, AND RECOMMENDATIONS highlights the proposed model's contributions to enhancing FHB detection in wheat spikelets, summarizing the study's primary findings. The study's limitations are also discussed, and recommendations for future research are provided, including potential areas for further exploration and development.

CHAPTER SIX – REFLECTIONS demonstrates the challenge-solving process, personal development, and lessons learned that were encountered during the project's completion. The chapter emphasizes the significance of meticulous planning, time management, and the assistance provided by various resources, such as academic guidance and peer communication. It also underscores the importance of adaptability and persistence in surmounting challenges and accomplishing the project's objectives.

CHAPTER 2 – REVIEW OF LITERATURE

The existing literature and advancements in object detection based on Fusarium Head Blight (FHB) are the primary focus of this chapter. Existing approaches to object detection are primarily divided into Machine-Learning and Deep-Learning-based methodologies. End-to-end object detection algorithms, particularly YOLO detectors, have emerged as the most dominant deep detection frameworks due to their rapid inference speed and simplicity of extracting disease phenotypes. Additionally, this chapter highlights the frameworks and research studies associated with detecting FHBs using object detection techniques. The chapter also examines the current structures and research on feature fusion in object detection for spikelet-level detection tasks and the progress in multi-scale feature fusion research. Multi-scale feature fusion methods are critically analyzed, and YOLO is comprehensively explained, along with its timeline.

2.1 FUSARIUM HEAD BLIGHT

Fusarium Head Blight (FHB) is a detrimental fungal disease prevalent among numerous Fusarium fungus species and predominantly caused by Fusarium graminearum. Furthermore, it substantially threatens cereal crops' growth, including wheat, rice, and oats [46]. Symptoms of the disease, including premature withering, ear rot, and whitening, are typically observed during the flowering phase of wheat. This results in substantial reductions in output and substantially impairs wheat quality, occasionally leading to complete crop failure. In wheat fields, the disease is prevalent in semi-humid and humid regions, resulting in 10% to 70% yield reductions and affecting an area exceeding 70,000 hectares [47]. The infection also produces a variety of mycotoxins, including Deoxynivalenol (DON) and Zearalenone (ZEA), which accumulate in wheat and have substantial health implications for both humans and animals [44].

2.2 FHB CONTROLLING AND DETECTION

Chemical and genetic management interventions are the main strategies used to prevent and control FHB in wheat. In chemical control, using fungicides is necessary under specific environmental conditions [48]. Due to the significant

toxicity of Fusarium toxins and their devastating impact on wheat crop yield, fungicides are typically used in combination with crop rotation and the cultivation of resistant crops to protect plants. Conducting research and development on novel techniques for detecting FHB will guarantee systematic and efficient chemical control [49], guiding agricultural practices and facilitating scientifically optimized wheat production activities. Early identification of FHB infection in wheat is essential for identifying the optimal period for applying fungicides.

A crucial genetic control technique for ensuring agricultural production safety is selecting and breeding wheat cultivars resistant to FHB. Effective variety selection necessitates interdisciplinary study encompassing plant breeding, plant pathology, plant phenomics, and other relevant knowledge and practices. Hence, it is imperative to employ accurate and groundbreaking techniques to detect symptoms early and extract disease phenotypes [50].

Presently, the prevailing techniques for detecting FHB encompass optical and biochemical methods [51]. The most straightforward and precise approach for visual evaluation is through human experts [52]. However, the accuracy of human assessment can be reduced by factors such as weariness, external interference, and visual errors [53]. Biochemical techniques include gas chromatography-mass spectrometry analysis [54], polymerase chain reaction [55], and enzyme-linked immunosorbent assay [56]. Although these methods demonstrate high precision, they could be more damaging, time-consuming, and labor-intensive processes [57]. Hence, the pressing objective is to investigate non-invasive and expeditious techniques for identifying FHB in wheat. Several FHB detection methods exist that are mainly non-invasive approaches, such as machine learning-based methods [15, 58-60], and deep learning-based methods [24, 25, 61-65].

2.3 MACHINE LEARNING-BASED PLANT DISEASE DETECTION METHODS

Machine learning plays a vital role in agricultural and plant science research, with wide-ranging applications in plant disease diagnosis [66], crop yield prediction [67, 68], and other areas [69]. Presently, the detection and classification of wheat illnesses are commonly carried out using supervised machine learning algorithms

[58]. Support Vector Machines (SVM) [70], Random Forests [71], and Decision Trees [72] are often employed methods in the feature extraction stage for the classification of wheat illnesses.

The procedure of assessing FHB infection in wheat using digital images consists of two essential stages: segmentation of the wheat spike image and counting the number of spikes. These steps are primarily accomplished by combining digital image processing techniques with machine learning. Zhang, et al. [59] utilized a K-means clustering algorithm in conjunction with image processing to properly segment wheat spikes in photos that feature a cluster of wheat plants. Afterward, a Random Forest classifier was employed to divide the affected areas further, and the FHB severity was determined by computing the area ratio. While this technique is successful in identifying dense clusters of wheat spikes in the field, accurately extracting disease phenotypes using this method depends significantly on the precision of the segmentation algorithm. The technique is vulnerable to noise interference caused by intricate backdrops in real-world settings, requiring additional image pre-processing to remove background noise, resulting in increased computational and time expenses.

Mustafa, et al. [15] employed integrated learning using a Random Forest classifier to identify crucial illness characteristics by merging reflection spectroscopy and chlorophyll fluorescence imaging. Nevertheless, this approach has constraints in extensive applications owing to the challenge of incorporating data from diverse sensors that gather information at various scales. Although machine learning algorithms are elementary to create and may be easily interpreted and comprehended, they are susceptible to underfitting and necessitate feature engineering for pre-processing.

Notably, advanced imaging technologies, such as multispectral and hyperspectral imaging, have shown to be accurate and non-destructive instruments for extracting important biomarkers using spectral data [16]. These methods offer comprehensive data on crops' reflectance or transmittance spectra, allowing for the detection and measurement of disease-related alterations in plants [13]. Therefore, the use of multispectral and hyperspectral imaging technologies can

accurately measure the degrees of resistance that wheat has against fhb [73], making it easier to identify fhb in wheat spikes and crop canopies in the field [74]. These data formats are frequently utilized in conjunction with machine learning for image analysis [75].

Huang, et al. [60] employed the successive projection algorithm (SPA) [76] to choose spectral characteristics from hyperspectral images. They then combined the channels of these images to produce rgb images. Colour features were extracted using colour moments [77], while texture features were obtained using the gray level co-occurrence matrix (GLCM) [78]. Subsequently, a svm model was developed by integrating image and spectral data using the particle swarm optimization (PSO) algorithm [79] to detect fhb. While feasible, this approach requires intricate processing procedures, substantial requirements for image capture, and manual extraction of features. Spectrum imaging technologies offer crucial qualitative and quantitative spectrum data about disorders. However, they encounter difficulties concerning the expenses of detection, technological intricacy, and instrument-related problems [80]. These technologies necessitate high-performance data-gathering equipment, which leads to higher economic expenses and is heavily influenced by the environmental conditions during data collecting.

2.4 DEEP LEARNING-BASED PLANT DISEASE DETECTION METHODS

Deep learning technologies have significantly enhanced research and applications in identifying and phenotyping crop diseases [17, 18]. The combination of deep learning and image processing techniques allows for the efficient and accurate extraction of phenotypic information in plant disease phenotyping research, making it a central focus of study [19]. Deep learning models surpass classic machine learning models' ability to solve computer vision problems, such as plant disease identification, fruit categorization, and crop seed phenotyping. Unlike traditional models, deep learning models do not need manual feature pre-extraction and demonstrate exceptional performance [81]. Convolutional Neural Networks (CNNs) have proven to be highly effective in processing and interpreting large-scale picture data, specifically identifying diseases and probable symptoms in crop images [82]. Convolutional Neural Networks (CNNs) are utilized in

agriculture because they extract important features by employing a combination of layers, taking advantage of the translational invariance of convolutional operators, and considering the spatial correlations between neighboring data. Traditional CNNs, such as LeNet-5 [83], AlexNet [21], ResNet [84], and VGG [85], have been effectively employed for plant disease identification [86].

Deep learning methods for detecting FHB primarily involve training classification models to identify diseases, training segmentation models to separate wheat spikes and diseased regions for assessing severity [61] [62], training object detection models to locate and classify wheat spikes [87], and combining deep learning with machine learning techniques [24]. Girshick introduced the Fast R-CNN and Faster R-CNN network models based on R-CNN [88, 89], which significantly improved the accuracy and speed of the algorithm in wheat spike detection [90], laying a solid technical foundation for the practical application of wheat. Besides, Bernardes, et al. [63] acquired a high detection accuracy of 99% for classifying wheat seed FHB using RGB photographs of individual wheat seeds. They combined hyperparameter optimization through a random search method [91] and fine-tuning with various pre-trained CNNs [92]. Their results proved the efficacy of affordable imaging technology and deep learning models in precisely categorizing wheat seeds infected with FHB. Nevertheless, the symptoms of FHB vary depending on the genotype, which can impact the reliability of detection techniques when applied to different types of wheat.

Qiu, et al. [64] employed an Inception neural network that consisted of larger network branches and various small-size parallel convolutions combined with channel attention modules to analyze Raman spectrum characteristics. This approach resulted in a detection accuracy of 93.62% for wheat grains infected with FHB. However, the practical implementation of this technology is hindered by the intricate structure of grains and the nuanced variations in their spectral properties, which pose difficulties in accurately identifying crucial indications. Additionally, in the study of Wang, et al. [61], two models, Deeplabv3+ [93] and the Hrnet [94], utilized high-resolution sub-networks as the underlying architecture. These models demonstrated exceptional performance in image segmentation and feature recognition. They effectively processed high-throughput wheat spike images and

accurately identified complex areas displaying FHB symptoms. The grading accuracy for FHB severity achieved a remarkable 92.6%. Nevertheless, this two-step processing approach results in significant time and annotation expenses for training data.

Hassan, et al. [95] utilized a combination of machine learning and deep learning techniques to develop two ways for disease identification in maize, potatoes, and tomatoes. The first strategy combined shallow VGG with Random Forest, while the second combined shallow VGG with XGBoost. These approaches achieved an impressive average accuracy of 95.70%. Gu, et al. [65] introduced a method for assessing the severity of FHB using the Relief-F algorithm. They utilized AlexNet to extract deep convolutional features and employed the Random Forest algorithm for classification and recognition. This approach successfully achieved a fusion of high-dimensional and low-dimensional information. This method achieved a precision rate of 94% when assessing the harshness of individual wheat spikes. Nevertheless, it necessitates a gradual execution and distinct feature extraction for categorization and identification, constraining its capacity to swiftly and effectively detect FHB.

2.5 YOLO-BASED FHB DETECTORS

YOLO (You Only Look Once) is a one-stage deep learning-based algorithm that can accurately segment complex images and predict the category and location of objects based on the information from each grid [96]. This fully convolutional network structure can efficiently and quickly process multiple objects simultaneously, making YOLO widely used in object detection [97]. Compared to two-stage algorithms, one-stage algorithms typically have lower precision and recall rates and are less effective at detecting small objects.

However, significant information loss occurs during the feature extraction for high-level semantic information for small-scale targets like wheat spikelets [98]. Thus, optimizing feature fusion mechanisms is necessary to supplement further and enrich the feature representation for these small-scale objects [99] [100] [101]. Zhang, et al. [24] employed the YOLOv5 object detection network and an enhanced DIOU-NMS non-maximum suppression technique to detect wheat

spikes. They then utilized a Random Forest algorithm to categorize colour data that were manually retrieved, resulting in a 96.16% accuracy in detecting FHB. Nevertheless, this method cannot achieve fully automated detection of FHB from start to finish. Gao, et al. [43] enhanced the YOLOv5-S model by substituting the C3 module of the baseline network with SPPF and GhostC3 modules. This modification yielded a compact model with a parameter count of 3.64M and achieved an average precision of 97.15%. Nevertheless, this technique identifies the presence of infection in wheat spikes by analyzing photos that show groups of wheat spikes, but it does not differentiate between specific types of infections. As a result, more precise disease characterization is needed for practical purposes.

Zhang, et al. [102] proposed a modified YOLO detection network that incorporates a straightforward spatial attention network and utilizes gray-coded labels for angle encoding. This modification allows for detecting wheat spikes in any orientation within the image. The spatial attention network employed spatial attention and spatial continuity loss functions to extract the spatial distribution properties of wheat, enabling the identification of each wheat spike in the image. The severity of FHB in individual wheat spikes was evaluated using threshold segmentation and K-means clustering. The approach attained a mean accuracy of 94.44%. Although there have been advancements and enhancements in identifying wheat spikes, detecting FHB still necessitates further post-processing.

Hong, et al. [25] enhanced the YOLOv8 model by substituting the Conv module with GhostConv [30] and employing Focal CIoU as the loss function. In addition, they developed the C-faster module as a substitute for the original C2f module. However, the training and testing data they used could not detect detailed characteristics of infection severity in wheat spikes. As a result, the model could only determine the location and category of infected and healthy wheat spikes without providing precise information about the extent of the infection, such as the rate of infected spikelets.

2.6 FEATURE FUSION OF OBJECT DETECTION

Feature fusion is widely used in the field of object detection. It involves extracting low-level features from shallow layers and high-level features from deep layers

within the network to obtain more precise semantic feature representations. For example, YOLOv2 connects features from different layers at the network bottleneck [103]. Other algorithms, such as Feature Pyramid Network (FPN) [34], focus on more complex fusion mechanisms, extracting and integrating multi-scale features for multi-scale object detection. This approach effectively transmits features and narrows the semantic gap.

2.6.1 Feature Pyramid Network

The Feature Pyramid Network (FPN) achieves feature fusion by facilitating the flow of features between layers, directly merging feature maps from specific layers. During the feature fusion process [104], the feature flow mechanism maintains the coherence of semantic information as it traverses all feature levels. The FPN architecture employs a top-down pyramid structure where higher feature levels undergo multiple down-sampling steps, resulting in smaller feature maps with lower resolution but richer semantic information. However, this process leads to significant information loss [103]. FPN incorporates lateral connections to supplement the missing positional information and mitigate this. This design allows for a more comprehensive feature representation across different scales and has been employed in many detectors [34, 105, 106].

Zhu, et al. [99] designed an innovative feature fusion module and integrated it into the FPN. This module enhances the capability to handle instances of varying sizes by dynamically adjusting the fusion weights based on feature similarity. Additionally, they incorporated a group attention mechanism to enhance the spatial information representation of features across different layers. Their method was integrated into a Faster R-CNN framework with a ResNet-101 backbone, increasing average precision (AP) from 39.7 to 41.4 on the COCO test-dev dataset. However, the inclusion of the similarity-based fusion module and additional attention layers increased the number of parameters, memory usage, and FLOPs.

Wang and Zhong [100] proposed the Adaptive Feature Pyramid Network (AdaFPN), which incorporates Adaptive Up-sampling (AdaUp) that leverages both spatial coordinates and semantic information to predict coordinate offsets for

sampling points, thereby achieving more flexible feature interpolation. Furthermore, by integrating an attention mechanism, they dynamically predict pixel-level fusion weights, balancing high-level semantic and low-level detail information for each pixel. AdaFPN was integrated into both Faster R-CNN and FCOS models, resulting in a 1.2 increase in AP for Faster R-CNN and a 1.0 increase in AP for FCOS on the MS-COCO dataset. However, this method still needs to be improved related to feature misalignment during multi-scale feature fusion, especially in complex scenes such as wheat spikes in fields or objects of diverse scales, highlighting the difficulties in aligning features across different levels.

2.6.2 Multi-Scale Feature Fusion

In multi-scale feature fusion, objects of different scales can be extracted and fused through the improved FPN [34]. Guo, et al. [107] designed a new feature pyramid structure AugFPN to extract scale-invariant contextual semantic information through residual feature enhancement, reducing information loss in the highest-level feature map of the feature pyramid and using MobileNet-v2 as the backbone improved the average precision (AP) by 1.6%.

Zeng, et al. [108] proposed a new multi-scale feature fusion method for defect detection, designing the atrous balanced feature pyramid network (ABFPN). This method enhances contextual information using dilated convolution operators with different dilation rates and applies to skip connections for sufficient feature fusion on top of FPN.

Guan, et al. [109] improved the Faster R-CNN framework for application to the UA-DETRAC car dataset, achieving more accurate localization of small-scale objects. By introducing a feature fusion module, they combined abstract semantic information captured at higher layers with detailed information at lower layers, generating finely resolved feature maps.

Jiang, et al. [110] designed a multi-scale feature extraction module (MSFEM) for UAV aerial image object detection. They extracted rich multi-scale feature information through multiple branches with different convolution operations. They expanded the FPN scale and introduced skip connections using the proposed

bidirectional dense feature pyramid network (BDFPN). Experiments on the VISDrone and UAVDT defect detection benchmark datasets showed better results than state-of-the-art object detection methods.

Du and Liang [101], based on the YOLOv5 model, optimized the FPN using Depth-wise Separable Convolution and Involution operators and spatial attention. Additionally, they designed an adaptive spatial convolutional block attention mechanism (SCBAM), applying self-attention mechanisms to the convolutional block attention module (CBAM) to enhance feature representation capability, resulting in a 7.4% accuracy improvement compared to the original model. Wu, et al. [111], also based on the YOLOv5s network, introduced attention mechanisms into the FPN for feature fusion, embedding coordinate attention (CA) into the original network's C3 module to improve the model's ability to distinguish features of overlapping objects. This network achieved a 3% mAP improvement on the BDD100k road object detection public dataset, reaching a 71.2% mAP.

Cheng, et al. [112] proposed an end-to-end cross-scale feature fusion network based on RCNN. They modeled the relationships between feature maps of different channels by inserting a squeeze and excitation (SE) block at the top of the FPN. They designed a CSFF module to obtain multi-level feature representations. This framework, applied to the DIOR public dataset of natural scenes, improved mAP by 3.0% compared to Faster R-CNN using FPN.

Liu, et al. [113] designed an adaptive multi-scale feature enhancement and fusion module (ASEM). They initially collected multi-scale features through a feature pyramid. They integrated a fine-grained feature extraction module for each level, refining multi-scale features using atrous convolutions with different dilation rates and employing attention mechanisms for feature fusion. Compared to Rotated Faster-RCNN, this achieved a 0.81% mAP improvement on the DOTA-1.0 dataset.

Yang, et al. [114], based on YOLOv3, proposed the MSF-YOLO model, increasing the convolution scale to four scales for the original ResNet and integrating features at each scale to obtain rich hierarchical information. Additionally, they

optimized the anchor box mechanism and training process, resulting in a 31.54% accuracy improvement on an industrial dataset compared to the original YOLOv3.

2.7 YOLO ARCHITECTURE AND TIMELINE

This section primarily introduces the architecture of the YOLO detector and the fundamental principles of its various components. Additionally, it presents the iterative updates and improvements of the YOLO series algorithms through a timeline format.

2.7.1 YOLO Architecture

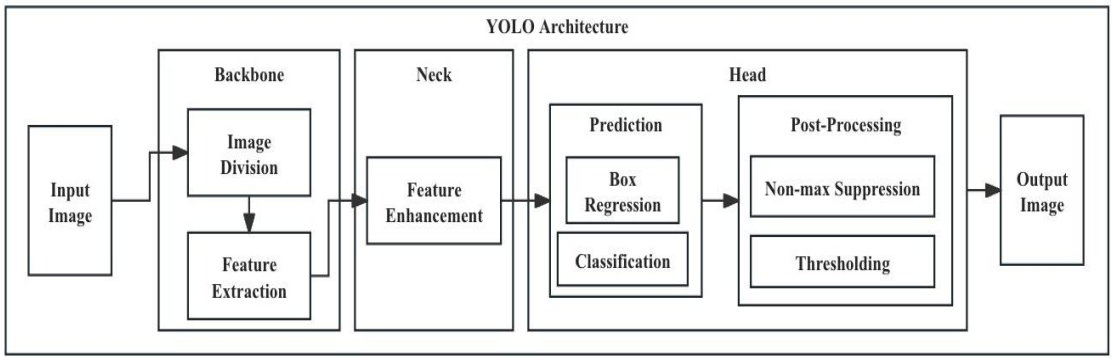

Figure 1: YOLO Network Architecture

As illustrated in **Figure 1**, the YOLO architecture can be divided into several components: Backbone, Neck, and Head [115]. The backbone extracts feature from the input image, and the neck serves as a connection component between the backbone and the head, refining the features extracted by the backbone. The neck enhances semantic information at different scales and often includes feature fusion mechanisms such as feature pyramid networks to enrich feature representation [116]. The head makes predictions based on the features provided by the first two parts of the network, including performing classification, localization, and post-processing such as NMS to retain the highest confidence prediction boxes [117]. YOLO detector completes object detection with a single network forward pass. Its core idea is to convert the detection task into a regression problem, using the entire image as the network input to obtain the positions of the bounding boxes and their corresponding classes [118].

In the Backbone section, the input image is the image to be detected. The input image is first divided into a fixed-size grid, and for each grid cell [119], each is responsible for detecting objects in the image. A CNN extracts features from the image. The primary purpose of this step is to capture essential information from the image, perform dimensionality reduction, and make subsequent detection more efficient. Then, the Neck section further enhances features extracted from the backbone network. Standard methods include using FPN or similar structures to capture multi-scale information better. The Head section consists of two subtasks: Box Regression and Classification. It predicts the bounding boxes of objects in each grid, outputting the coordinates of the boxes and confidence scores and predicting the class of objects in each grid.

Notably, a loss function is used when training to measure the differences in position, confidence, and classification accuracy between the predicted and ground truth boxes [96]. However, the predicted bounding boxes may overlap with each other. YOLO employs the non-maximum suppression (NMS) algorithm [120] to eliminate redundancy, which selects the best prediction boxes based on confidence and overlap [97]. Thus, in the Post-Processing stage, non-maximum suppression is utilized to retain bounding boxes with the highest confidence scores and remove overlapping boxes. Subsequently, detection results with lower confidence scores are filtered based on preset thresholds. The final output includes annotated images with detection boxes and class labels.

2.7.2 YOLO Timeline

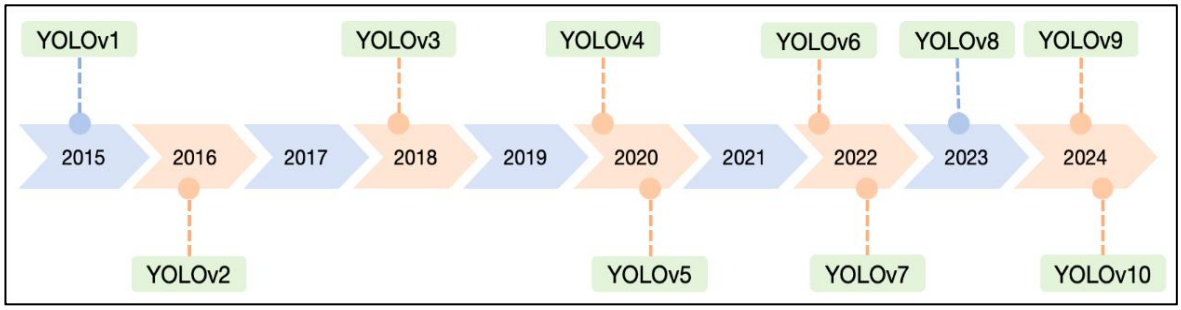

Figure 2: YOLOv1 to YOLOv10 Timeline

As shown in **Figure 2**, YOLOv1 unified the steps of object detection by simultaneously detecting all bounding boxes, dividing the input image into grids of

equal size, and making predictions on each grid. YOLOv2 first introduced k-means clustering anchor boxes in the detection head to enhance bounding box prediction [121]. YOLOv3 incorporated multi-scale predictions at three scales (large, medium, and small) in the neck and head [122]. YOLOv4 employed the path aggregation network (PANet) in the neck for feature aggregation and spatial pyramid pooling (SPP) to enhance the receptive field [123]. YOLOv5 proposed the Spatial Pyramid Pooling Fast (SPPF) based on SPP to improve processing speed and introduced the distribution focal loss function for bounding box prediction [124]. YOLOv6 integrated RepBlocks into PANet, further enhancing model performance through re-parameterization modules [125]. YOLOv7 improved feature fusion in the neck by adopting the ELAN layer aggregation structure, optimizing gradient flow and feature processing efficiency [126]. YOLOv8 eliminated the anchor box mechanism, further enhancing speed [97]. YOLOv9 proposed the Generalized Efficient Layer Aggregation Network (GELAN) based on ELAN in the neck to improve feature fusion mechanisms [42]. YOLOv10 removed NMS in the head, significantly boosting model inference speed [127].

2.8 LITERATURE SUMMARY

Despite the current feature extraction methods, such as AugFPN [107], which reduce information loss through residual feature enhancement, information loss remains when handling the highest-level feature maps. This information loss can lead to suboptimal precision in object detection, particularly for small-scale objects. Although the BDFPN [110] demonstrates excellent performance in multi-scale feature extraction, its complex structure also increases computational overhead, making it unsuitable for resource-constrained environments. While the MSF-YOLO [114] performs well on industrial datasets, its performance on other datasets still requires further validation. This indicates that the generality and robustness of current methods across different application scenarios need improvement.

Although the ASEM [113] enhances feature fusion through atrous convolutions and attention mechanisms, fine-grained feature extraction still has room for improvement when dealing with highly complex or detail-rich images. Existing feature extraction methods still need to be improved in their feature representation capabilities and cannot fully capture all the information of the target objects. For

instance, optimizing FPN based on YOLOv5 [101], which enhances feature representation capabilities by introducing self-attention mechanisms, still needs further enhancement in distinguishing objects within complex backgrounds.

2.9 CHAPTER SUMMARY

This chapter primarily provides a comprehensive analysis of current mainstream machine learning and deep learning research on plant disease detection, mainly focusing on object detection methods and multi-scale feature fusion. Machine learning offers strong interpretability but requires manual feature selection. In contrast, deep learning methods, such as the end-to-end YOLO detector, enable rapid FHB detection. However, the detection target is still at the scale of the wheat spike, preventing the extraction of detailed disease phenotypes. Multi-scale feature fusion mechanisms can capture multi-scale wheat spike features. Therefore, designing an object detection method specifically for spikelets can enhance detection performance and allow the extraction of more detailed FHB phenotypes. Chapter 3 presents the methods and materials for the detection of FHB using YOLO architecture and multi-scale feature fusion module.

integrated RepBlocks into PANet, further enhancing model performance through re-parameterization modules [125]. YOLOv7 improved feature fusion in the neck by adopting the ELAN layer aggregation structure, optimizing gradient flow and feature processing efficiency [126]. YOLOv8 eliminated the anchor box mechanism, further enhancing speed [97]. YOLOv9 proposed the Generalized Efficient Layer Aggregation Network (GELAN) based on ELAN in the neck to improve feature fusion mechanisms [42]. YOLOv10 removed NMS in the head, significantly boosting model inference speed [127].

CHAPTER 3 – RESEARCH METHODOLOGY

This chapter introduces the research methods employed in this study to achieve the aim of the study. The object detection model framework, the proposed module, performance evaluation metrics, research data and materials used for experiments are presented.

3.1 RESEARCH DESIGN

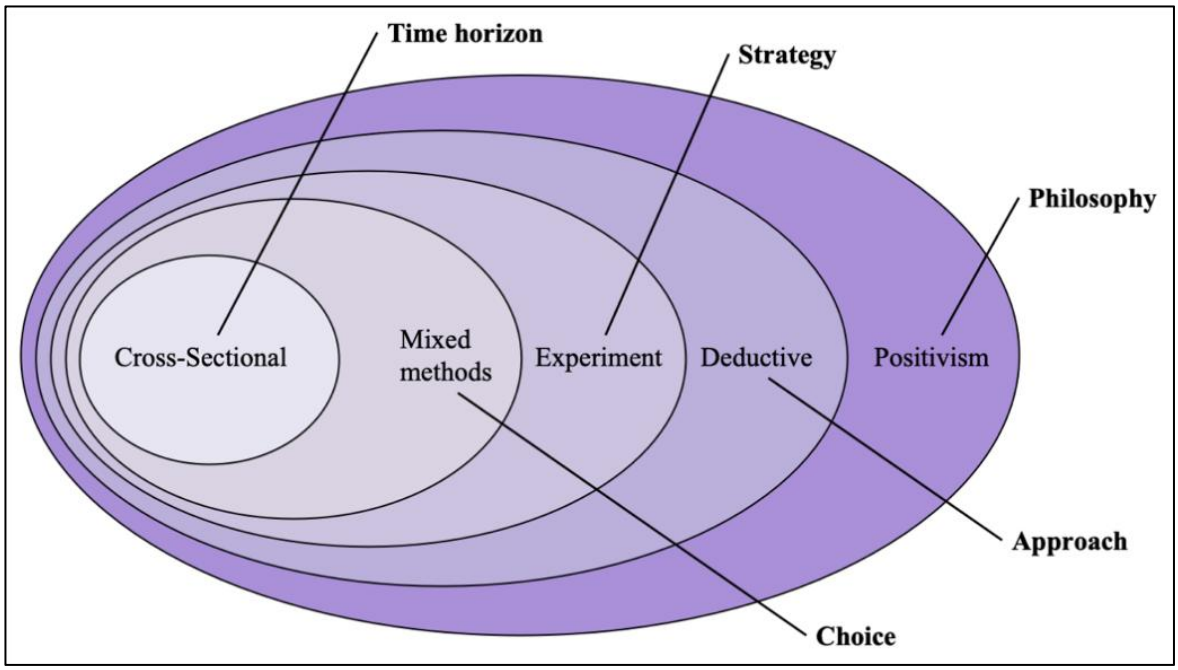

Figure 3: Research Design

Figure 3 shows that this study employs the positivism philosophy and deductive approach [128]. Positivism Philosophy is chosen because this study is highly quantitative, relies on data, experiments, and statistical analysis, and emphasizes objective measurements [129]. The deductive approach is used because the study builds on existing theories and frameworks of deep learning and object detection algorithms. The deductive approach is appropriate as experiments are designed, and data are collected using annotated image datasets to train and test the models, validating the effectiveness of the proposed object detection model for detecting FHB on wheat spikelets [17]. The algorithm learns the relationship between disease features and labels from specific instances of FHB in wheat images for disease detection [130].

The research strategy is experiment-based. To ensure the algorithm's accuracy and robustness, the RGB wheat images must be processed through experiments to train, validate, and test the model [17]. Moreover, the model's performance is evaluated using assessment metrics in experiments to ensure the effectiveness and reliability of the algorithm in detecting FHB in wheat [131].

This study adopts a mixed-methods approach, combining quantitative and qualitative methods commonly used in the current field of plant detection [19, 25, 26, 102]. Object detection algorithms can identify and indicate regions impacted by FHB in images, thereby facilitating the quantification of the disease by calculating the number of infected areas [25]. The object detection algorithm assists in quantifying the severity of FHB by determining the percentage of diseased spikelets in the wheat ear [19]. Furthermore, the model's accuracy and reliability are validated using quantitative metrics, including Precision, Recall, Average Precision (AP), Mean Average Precision at IoU threshold 0.5 (mAP@0.5), and the F1 score. These metrics have become standard evaluation tools in object detection [132-134], widely applied across various studies to validate models' accuracy, stability, and real-world applicability.

The image data analysis in this study enables the identification of diseased regions and the detection of specific features of FHB seen in the photos, such as discoloration and mold patches [13]. The presence of FHB is identified and confirmed by these characteristic features, which are essential for qualitative assessment [135]. The object detection algorithm visually represents the location and extent of the disease by drawing bounding outlines on the images, thereby facilitating the qualitative evaluation of the disease through the visual expression of the image data [136].

In order to identify and classify wheat spikelets infected with FHB, this study implemented a cross-sectional time horizon, utilizing RGB images acquired during the flowering stage. Data collection was conducted at a single point during the flowering period of the wheat crops, as this is the critical period when FHB symptoms are most apparent, thereby enabling more precise measurements [137].

3.2 RESEARCH MODEL

This section introduces the research model architecture utilizing YOLOv9 and outlines the organizational structure of its primary components for the detection of FHB in wheat spikelets.

3.2.1 Proposed Architecture

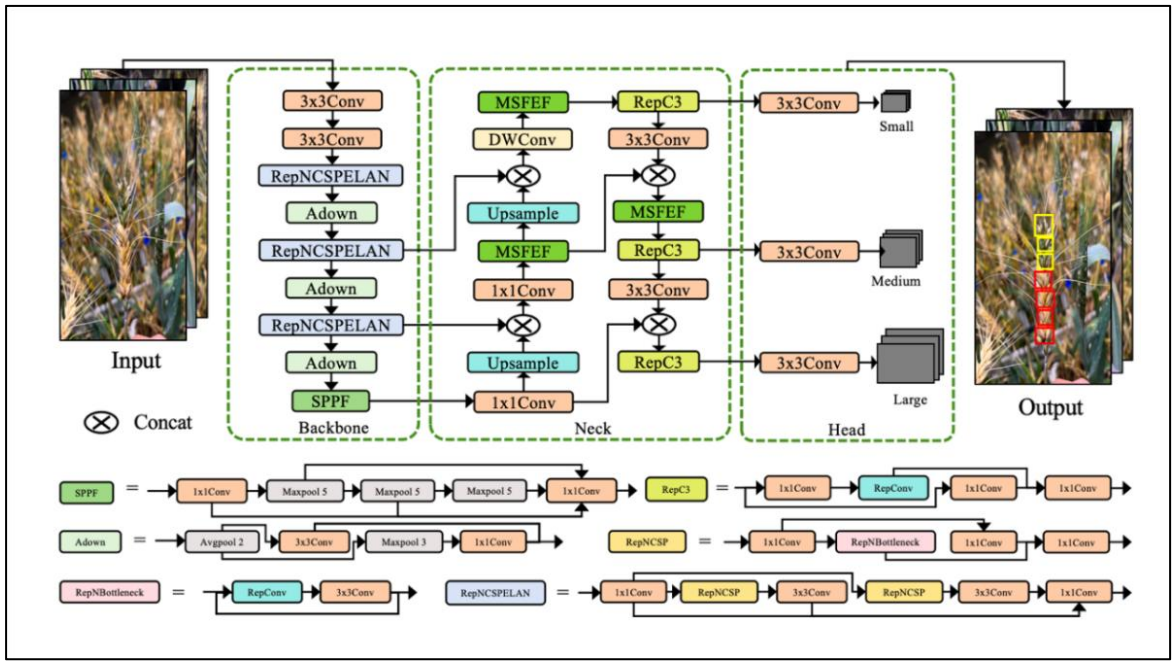

Figure 4: Proposed Model Architecture

The YOLO-based architecture comprises a backbone, bottleneck, and head elements. As illustrated in **Figure 4**, the backbone module is composed of convolutional layers, RepNCSPELAN, Adown, and Spatial Pyramid Pooling Fast (SPPF). The bottleneck module comprises convolutional layers, up-sampling, MSFEF, DWConv layers, and RepC3. The head module integrates object features of large, medium, and small scales through convolutional layers.

The Spatial Pyramid Pooling Fast (SPPF) is initially used at the end of backbone, and a 1×1 convolutional layer is then employed in the bottleneck to integrate the channels of the features extracted by the backbone. The concatenation follows three repetitions of RepNCSPELAN aggregating multi-leveled deep features, then integrated using a 1×1 convolutional layer. Subsequently, the MSFEF is used to supplement and dynamically enhance multi-scale features, fused with feature

maps from the previous RepNCSPELAN aggregations. Considering the increase in parameter count, DWConv is used to reduce computational costs.

The bottleneck harmonizes feature scale and resolution differences, combining deep, semantically rich features with shallow, spatially detailed features. The MSFEF module ensures that features at different scales are dynamically enhanced, providing the network with robustness against changes in object size and complex backgrounds, which is crucial for detecting wheat spikelets appearing at various scales and distances from the camera, and also enhances the model's ability to locate spikelets against complex backgrounds. The network maintains a rich hierarchy of cross-scale features without excessively increasing computational costs.

In proposed network, the RepNCSPELAN module initially splits the feature map into two parts. One part is directly linked to the output through residual connections, while the other part undergoes further feature extraction and fusion operations, subsequently combines the RepNCSP and a 3×3 convolutional layer to extract features progressively. Finally, a 1×1 convolutional layer adjusts the channels and integrates the feature information. The module achieves parameterization through the components RepConv and RepNCSP. RepNCSP is an expansion of the traditional CSP structure, as shown in **Figure 4**, where the original bottleneck structure is replaced with the Repbottleneck, and the conventional convolutional layer typically found in the first layer of the bottleneck is substituted with a RepConv layer.

The Generalized Efficient Layer Aggregation Network (GELAN) in RepNCSPELAN is a highly efficient aggregation structure that integrates CSPNet and ELAN [42]. GELAN balances the computational complexity of the dense architecture with the computational efficiency of CSPNet's feature separation and fusion, optimizing gradient flow through layer-by-layer feature accumulation and merging across various blocks, culminating in a design that is both light-weight and rich in feature representation.

In feature processing sequence shown in **Figure 4**, the Adown module first performs non-padded average pooling to reduce the size of the feature map while preserving features. The feature map is then evenly split along the channel axis, with one part undergoing feature extraction through a 3×3 convolutional layer, while the other part processed through 3×3 maximum pooling, followed by a 1×1 convolution to adjust channels and integrate features. Ultimately, the two parts of the feature map are concatenated along the channel dimension for feature fusion, which enhances the model's feature sensitivity by combining average and maximum pooling operations without increasing the computational burden.

The Spatial Pyramid Pooling Fast (SPPF) is a variant of the Spatial Pyramid Pooling (SPP) module [138], where the processing begins with a 1×1 convolution layer followed by three serially executed maximum pooling layers. The output feature maps from each layer are concatenated, and the sequence concludes with another 1×1 convolution layer, which simplifies the original SPP, streamlining the architecture by employing a fixed-size single pooling kernel instead of multiple sizes. Unlike the parallel execution of three maximum pooling blocks in the original SPP, SPPF preserves historical feature information, enhancing feature interaction. Thus, SPPF operates faster than the original SPP.

Depth-wise separable convolution operations are divided into two sequential steps: depth-wise convolution and point-wise convolution [139]. Depth-wise convolution is performed per channel, using one or more channel groups, which splits the feature map into several sub-channels, allowing each convolution kernel to extract local details from the feature map. Channel expansion of each kernel is specific to certain channels, significantly reducing the number of parameters and simplifying the convolution process. However, depth-wise convolution only considers the feature information of individual channels and fails to capture the complex relationships between different channels, resulting in output feature maps that lack multi-channel dimension correlation. A 1×1 standard convolution is used to process the feature maps after depth-wise convolution, achieving more effective integration of features.

The RepC3 module replaces the bottleneck module by stacking the RepConv layers [140]. RepConv is designed into two-branch convolutional structure with a 3×3 convolution that learns spatial features and a 1×1 convolution that adjusts channels. RepConv merges two branches during the deployment and inference phase to reduce computational complexity, which simplifies the model by calculating biases and using equivalent convolution kernels for replacement. As shown in **Figure 4**, the RepC3 concludes with a 1×1 convolutional layer that alters channels and adjusts the merged features.

3.2.2 Proposed MSFEF Module

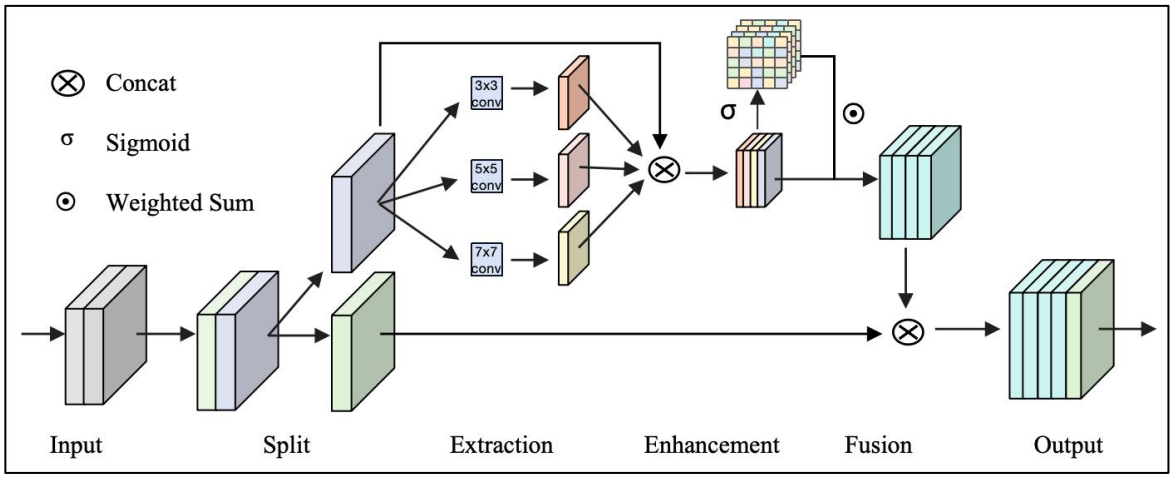

Figure 5: MSFEF Module

Due to the loss of feature details in deep convolutional layers, the subtle differences between wheat spikelets are difficult to be distinguished. The MSFEF module, inspired by the methodologies in [141], balances the simplification of computation and feature fusion by extracting features from only some channels while enhancing the interaction between different scale feature information. As illustrated in **Figure 5**, The module utilizes multi-scale feature extraction to enhance the spatial and semantic information of spike images, improving the model's ability to distinguish between infected spikelets and complex backgrounds.

During the feature extraction, the generated feature maps across different channels exhibit high similarity, resulting in substantial redundancy, which consumes computational resources and hampers efficient feature extraction. By randomly selecting certain channels for conventional convolution while leaving the

rest unprocessed, the design of partial convolution reduces redundant feature computations and lowers memory access, significantly reducing floating-point operations (FLOPs). The principle of dynamic feature-weighted enhancement is illustrated in **Equation 3.1** and **3.2**.

$$W_i = \sigma\left(Conv_{1\times1}(Concat\left(Conv_{k_i}(X_{C_P})\right))\right) \qquad X_{C_P} \in \mathbb{R}^{C\times H\times W}$$

3.1

Where W_i represents the weight maps of features at different scales, k_i represents convolution kernels of various sizes, X_{C_P} represents the feature map processed through convolution.

$$Y = Concat\left(\left(\sum_{i=1}^{n} W_i \odot \left(Conv_{k_i}(X_{C_p})\right)\right), X_{C_u}\right) \qquad Y, X_{C_u}, X_{C_P} \in \mathbb{R}^{C\times H\times W}$$

3.2

Where W_i represents the weight maps of features at different scales, k_i represents convolution kernels of various sizes, Y is the output feature of the module, X_{C_P} represents the feature map processed through convolution, and X_{C_u} represents the unprocessed original feature map. σ denotes a sigmoid operation that determines the weights of features at different scales, ensuring that feature weights lie between 0 and 1. The $\odot$ denotes an element-wise multiplication operation.

The MSFEF module first sets a channel ratio for the input features, randomly splitting the channels into a portion with C_p channels for subsequent processing and another portion as C_u channels of unprocessed original feature maps. The sum of C_p and C_u equals the total number of input channels C. The module dynamically adjusts the weights of features at different levels through a sigmoid function and then performs a weighted fusion to produce dynamically enhanced output features.

To maintain a light-weight design while handling complex features and further enhancing multi-scale feature representation, the design incorporates two 1×1 convolution layers, which introduces residual connections to form an inverse residual block, increasing the channel capacity at intermediate layers to boost feature extraction capabilities.

3.3 RESEARCH DATA ANALYSIS AND EVALUATION METRICS

This section primarily presents the analysis of experimental results related to Fusarium Head Blight detection and the evaluation metrics for the algorithms.

3.3.1 Data Analysis

The diseased spikelet rate (DSR) refers to the proportion of infected spikelets in a single wheat spike. Measuring the disease spikelets rate in samples of different wheat varieties or treatment groups allows for assessing their FHB resistance. The calculation of DSR is defined as in **Equation 3.3**.

$$Diseased\ spikelets\ rate = \frac{Number\ of\ diseased\ spikelets}{Total\ number\ of\ spikelets\ per\ spike} \times\ 100\% \quad 3.3$$

The diseased spikelet rate assesses the severity level of the disease according to the Chinese Agricultural Industry Standard: NY/T 1443.4-2007 published by the Ministry of Agriculture of China, which is classified into five levels as follows: Level 0: No infected spikelets (DSR = 0%); Level 1: Sparse occurrence of infected spikelets (DSR ≤ 25%); Level 2: (25.0% ≤ DSR ≤ 50.0%); Level 3: (51.1% ≤ DSR ≤ 75.0%); Level 4: (DSR ≥ 75%).

The statistical indicators used for comparing model prediction results with manual counting results include R^2, Root Mean Square Error (RMSE), and Relative Root Mean Square Error (rRMSE) [142], as shown in **Equations 3.4**, **3.5**, and **3.6**. RMSE is a commonly used evaluation metric in this study to quantify the difference between the model's disease phenotype predictions and manual observations [143]. rRMSE, the relative form of RMSE, normalizes RMSE by dividing it by the mean of the actual observations, providing a dimensionless error measure [144].

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}(x_i - \overline{x}_i)^2} \quad 3.4$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2} \quad 3.5$$

$$rRMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - y_i}{x_i}\right)^2} \quad 3.6$$

Where n represents the number of test samples, x_i represents the disease severity statistically manually, y_i represents the disease severity calculated automatically, and $\overline{x}_i$ represents the mean of x_i.

The R^2 value can measure the degree of fit between the model's predicted values and the manually recorded values. A high R^2 indicates that the model can accurately predict the diseased spikelet rate, reflecting the model's superiority in quantitatively assessing disease severity. A low RMSE value indicates that the model's predictions for the diseased spikelet rate are very close to the actual values obtained from manual counts, with minimal error, thus providing a more reliable assessment of the disease resistance of different wheat varieties or treatment groups. The rRMSE can measure the proportion of the model's prediction error relative to the actual values, and low rRMSE value indicates that the model has a small prediction error when estimating the diseased spikelet rate, with the predicted results highly consistent with the actual situation, making it suitable for comparison under different levels of disease severity.

3.3.2 Evaluation Metrics

The FHB spikelets detection employs evaluation metrics including Precision, Recall, AP, mAP@0.5, and F1, as illustrated in **Equations 3.8** - **3.12**. Before introducing the precision, the Intersection-over-Union (IoU) refers to the ratio of the intersection and union areas of the ground-truth bounding box and the predicted bounding box, which is an indicator used to measure the overlap between two boxes, determining positive samples and assessing the accuracy of object localization, as defined in **Equation 3.7**.

$$IoU = \frac{S(Predicted\ box \cap GT\ Box)}{S(Predicted\ box \cup GT\ Box)} \quad 3.7$$

Where S denotes the area of the bounding box, when the match between the predicted box and the actual box exceeds the IoU threshold, the predicted box is retained.

$$Precision = (\frac{TP}{TP+FP}) \times 100\% \quad 3.8$$

Precision indicates the proportion of instances predicted as diseased spikelets by the diseased spikelets model. High precision means the model's predictions for diseased spikelets are highly accurate, with few false positives. Specifically, True Positive (TP) refers to instances correctly classified as positive samples, representing instances where the model accurately predicts and localizes healthy and diseased spikelets. False Positive (FP) refers to the number of instances incorrectly classified or predicted with an IoU below the threshold with the GT, including instances where spikelets are incorrectly predicted or inaccurately located, for instance, the objects of background are mistakenly predicted as healthy or diseased spikelets.

$$Recall = (\frac{TP}{TP+FN}) \times 100\% \quad 3.9$$

Recall represents the proportion of actual diseased spikelets that the model correctly detects. High recall indicates that the model can detect most diseased spikelets with few false negatives. FN (False Negative) refers to instances incorrectly classified as negative samples, such as the GT is diseased or healthy spikelets but predicted as background.

$$AP = \int_0^1 P(R)dR \quad 3.10$$

Average Precision (AP) is defined as in Equation (6), representing the PR curve for each class.

$$mAP = \frac{1}{k} \sum_{i=1}^{k} AP_i \quad 3.11$$

The mean Average Precision (mAP) refers to the mean of the sum of Aps across each class, used to evaluate the performance of object detection algorithms, is one of the most widely used evaluation metrics [145], where k represents the total number of classes, and i refers to the current class, with the default IoU threshold set to 0.50, indicating successful detection of objects at this threshold.

The Average Precision and mean Average Precision consider the model's performance at different IoU thresholds. They are used to evaluate the model's overall performance across multiple categories, such as varying degrees of disease severity in spikelets. A high mAP value means the model performs well across all categories, accurately identifying and localizing various diseased spikelets.

$$F1 = \left(\frac{2 \times Precision \times Recall}{Precision + Recall}\right) \times 100\% \quad 3.12$$

F1 calculates the harmonic mean of precision and recall, comprehensively evaluating model performance for detecting diseased and healthy spikelets.

$$FPS = \frac{1000}{t} \quad 3.13$$

In this study, FPS (Frames Per Second) indicates the number of image frames that the spikelet object detection algorithm can process per second. A higher FPS implies a more robust real-time detection capability and faster detection speed of the model. The specific calculation method is shown in Equation 3.13, where ttt represents the time required to process and predict each image, measured in milliseconds.

$$GFLOPs = \frac{FLOPs}{1 \times 10^9} \quad 3.14$$

The calculation method for GFLOPs is shown in Equation 3.14. FLOPs are computed by invoking the Thop library, including the floating-point operations of convolutional, linear, and batch normalization layers. GFLOPs represent billions of floating-point operations per second and are an essential metric for assessing the computational complexity of YOLO-based models in this study. Higher GFLOPs usually indicate higher model complexity and greater computational resource consumption.

3.4 RESEARCH MATERIALS

This section mainly introduces the sources and details of data collection for the experiments, as well as the software and hardware configurations utilized in the experiments.

3.4.1 Research Data

The data source originates from inoculation experiments conducted on wheat spikes during the flowering phase, with investigations and image captures performed 21 days post-experiment. Image acquisition was done using an iPhone 13, utilizing its wide-angle camera (ISO sensitivity 40, focal length 26mm, aperture f1.6). During shooting, the phone was kept parallel to the spikes, with a distance between 15-25 cm, considering the significant variance in growth stages and disease resistance across different wheat varieties.

The phenotypic disparity between healthy and diseased spikelets is minimal, and given the complexity of data collection environments in practical applications, to enhance the generalization capability of the subsequent model training, variations were introduced in aspects such as illumination intensity, shooting distance, background, and the growth state of the wheat selected for photography. For each spike, images of both the front and rear sides (rotated 180 degrees) were captured, with each image having a resolution of 3024×4032 pixels and stored in JPEG format. Each image contains a single spike, and to ensure data quality, images that were blurred or had spikelets severely obstructed by awns, making it difficult to assess disease presence, were excluded. After selection, 620 wheat FHB images from 310 spikes were obtained for the dataset construction. The data collection and construction process were conducted in strict accordance with the guidelines and requirements outlined in **Appendices： APPENDIX A – ETHICS FORM**, ensuring full compliance with academic ethical standards.

3.4.2 Software and Hardware

The LabelImg software (https://github.com/ tzutalin/labelImg.) was employed for manual annotation of the spikelets in the images as either diseased or healthy. The experimental environment operates on a Linux server with an x86_64 architecture (v6.2.0), and running the Ubuntu operating system (v12.2.0), with single video card of GeForce RTX 4090 with 24GB memory (Nvidia AD102).

3.5 CHAPTER SUMMARY

This chapter presents the research methods employed to achieve the study's objectives, focusing on the object detection model framework, the proposed MSFEF module, performance evaluation metrics, research data, and materials used for experiments. The research methodology combines quantitative and qualitative methods to locate and quantify FHB-infected areas in wheat spikes. The study uses a mixed-methods approach and an inductive methodology based on fully supervised learning. An experiment-based strategy is adopted to train, validate, and test the model using RGB wheat images. The chapter details the proposed YOLOv9-based network architecture and its components, emphasizing the integration of modules like SPPF and GELAN to enhance feature extraction and aggregation while maintaining computational efficiency. Data analysis involves calculating DSR and using evaluation metrics including R^2, RMSE, and rRMSE for performance comparison. The chapter concludes by outlining the experimental setup, including the data collection process and the hardware and software configurations used. Chapter 4 presents and analyzes the experimental results, comparing them with state-of-the-art models using the defined evaluation metrics.

CHAPTER 4 – EXPERIMENTS AND RESULTS ANALYSIS

This chapter primarily introduces the basic configurations and parameter selections for training, validating, and testing the proposed model. It then comprehensively compares the performance and complexity of the proposed model with current state-of-the-art (SOTA) models. Additionally, it compares the performance of different component combinations within the proposed model and validates the effectiveness and generality of the proposed MSFEF module on various YOLO detectors. This chapter also presents the feature distribution and final prediction results during the image processing through visualizations. Furthermore, it performs correlation and regression analyses on the extracted phenotypes.

4.1 EXPERIMENT SETUP

The experiment utilized several essential libraries to build, train, and evaluate the proposed model. The primary programming language was Python 3.8. The core deep learning framework used was PyTorch (V1.9.0) and Torchvision (V0.10.0) for handling image transformations and model utilities. Additional libraries were utilized, such as Albumentations (V1.0.3) for advanced data augmentations, Pandas (V1.1.4) and Seaborn (V0.11.0) for data handling and visualization, Matplotlib (V3.2.2) for plotting, Numpy (V1.18.5) for numerical operations, OpenCV (V4.1.1) for image processing, and Thop (V0.1.1) for FLOPs computation.

The dataset consisted of 620 images annotated using LabelImg to generate Pascal VOC format annotations (XML files). These were converted to YOLO format (TXT files) for model training. 496 training images, 62 validation images, and 62 testing images were obtained by dividing the dataset into training, validation, and testing sets in an 8:1:1 ratio. The annotations comprised two categories: "h" for healthy spikelets and "d" for diseased spikelets.

Various metrics, such as Precision, Recall, AP_h, AP_d, mAP, F1 Score, Parameters, GFLOPs, and FPS, were employed to evaluate the model's performance. The

network architecture was configured following the scaling pattern of YOLOv5-S [146] by utilizing a YAML file. The depth and width scaling parameters were set to 0.33 and 0.50, respectively. The base model utilized was the GELAN structure from the YOLOv9 algorithm [42]. The fine-tuning of the model was performed using pre-trained weights obtained from the COCO dataset. The trials were set up for 150 epochs to mitigate overfitting, using a batch size of 8 and 8 workers. The model underwent training to identify two categories of spikelets: healthy and diseased. Moreover, various data augmentation techniques were implemented to improve the robustness of the model during the training process, which was executed using the train.py script. These augmentations encompassed random perspective transformations, copy-paste, cutout, and mix-up [147].

4.2 RESULTS AND ANALYSIS

The proposed model and current SOTA YOLO detectors are this section's primary focus, providing a comprehensive performance comparison. Furthermore, the proposed model's performance was compared under various loss functions. The optimal module combination was determined by conducting performance comparisons of various module combinations within the proposed model architecture. A comparison of the performance of the proposed MSFEF module across various detectors was conducted to verify its generality. In addition, visual analysis was implemented to evaluate the outcomes of feature processing and prediction. Lastly, regression and correlation analyses assessed the extracted phenotypes for accuracy.

4.2.1 Model Performance

The efficacy of the proposed model is evaluated in comparison to a variety of YOLO architecture versions, as illustrated in Table 4.1. Current SOTA detectors were tested in a comprehensive comparison with the proposed model. Precision, recall, AP_h, AP_d, mAP, F1 score, the number of parameters, GFLOPs, and FPS were the evaluation metrics applied for comparison. AP_h denotes the average precision for healthy spikelets, while AP_d represents the average precision for diseased spikelets.

Table 4.1: Model Performance Comparison

Model	Precision (%) ↑	Recall (%) ↑	AP_h (%) ↑	AP_d (%) ↑	mAP (%) ↑	F1↑	Parameters (M) ↓	GFLOPs↓	FPS↑
YOLOv3 [122]	77.1	84.9	86.8	88.9	87.8	80.8	103.7	282.2	110
YOLOv5-S [124]	87.5	82.7	83.8	89.4	86.6	85.0	7.0	15.8	47
YOLOv6 [125]	77.3	83.4	87.3	86.5	86.9	80.2	4.23	11.8	145
YOLOv7 [126]	81.4	75.6	82.8	87.4	85.1	78.4	36.5	103.2	135
YOLOv7-Tiny [126]	79.6	77.5	78.9	87.2	83.1	78.5	6.0	13.0	243
YOLOX-S **[148]**	86.1	81.2	86.7	89.6	88.1	83.5	8.9	26.8	36
YOLOv8-S-Ghost **[149]**	85.0	81.9	88.5	86.9	87.7	83.4	5.9	16.1	62
YOLOv8-S **[149]**	84.0	82.4	88.2	88.0	88.1	83.2	11.1	28.4	58
RT-DETR-L **[150]**	88.0	87.1	91.2	88.4	89.8	87.5	32.0	103.4	29
YOLOv9-C [42]	85.2	83.6	87.2	92.1	89.6	84.4	25.3	102.3	40
GELAN-C [42]	88.4	86.2	88.3	91.2	89.8	87.3	25.4	102.5	192
YOLOv10-S [127]	82.4	84.8	88.9	90.0	89.5	83.6	8.2	25.3	97
Proposed Model-CIoU	85.6	82.7	89.2	92.0	90.6	84.1	6.7	29.4	294

(Note - To visually distinguish performance, the top-3 values are highlighted in red, green, and blue respectively. The metrics are labeled with up arrows to indicate that higher values are indicative of more robust performance, while down arrows indicate that lower values demonstrate superior performance.)

The GELAN-C model obtained the highest precision of 88.4% among the evaluated models, illustrating its effective architecture and robust feature extraction. The efficiency of the RT-DETR-L model in object detection tasks is illustrated by its high AP_h of 91.2% and its highest recall of 87.1%. Furthermore, RT-DETR-L achieved the maximum F1 score of 87.5%, underscoring its balanced precision and recall performance. The YOLOv9-C model was distinguished by its mAP of 89.6% and the maximum AP_d of 92.1%, which are indicative of its advanced architectural enhancements and optimization for various detection scenarios.

The proposed model is highly suitable for real-time applications due to its exceptional FPS of 294. This high FPS indicates the model's optimized design, balancing computational efficiency and complexity optimally. Conversely, the models demonstrating the lowest performance in each metric were also identified. The YOLOv6 model achieved a lower precision of 77.3% and AP_d of 86.5%, which can be attributed to its restricted feature extraction capabilities resulting from fewer parameters. The YOLOv7 model demonstrated the lowest recall rate of 75.6% and F1 score of 78.4%, suggesting difficulties in reliably identifying objects under

different parameters. YOLOv7-Tiny achieved the lowest mAP of 83.1% and AP_h of 78.9% but achieved shorter inference times by reducing the model complexity. The RT-DETR-L model exhibited the lowest FPS value of 29, attributed to its computational burden despite its superb detection capabilities.

The proposed model shows its superior real-time performance with the highest FPS, and robust detection capabilities across multiple metrics. The GELAN-C and RT-DETR-L models also demonstrated strong performance, making them suitable for high-precision and high-recall applications.

4.2.2 Effect of Loss Function on Proposed Model

The definitions of the metrics, arrows, and colour labels used in **Table 4.2** are consistent with those used in the preceding Section 4.2.1. Specifically, the loss function does not alter the model's parameter count or floating-point operations, but it does impact the model's inference speed. Hence, only the FPS evaluation metric has been introduced. **Table 4.2** compares different loss functions—GIoU, DIoU, CIoU, EIoU, and SIoU—applied within the proposed models. The results demonstrate varying performance across different metrics for each loss function.

Table 4.2: Performance Comparisons on Loss Function

Loss Function	Precision (%) ↑	Recall (%) ↑	AP_h (%) ↑	AP_d (%) ↑	mAP (%) ↑	F1↑	FPS↑
Proposed Model-DIoU **[151]**	82.8	77.5	86.3	89.8	88.1	80.1	370
Proposed Model-EIoU **[152]**	85.1	83.7	87.8	92.4	90.1	84.4	286
Proposed Model-GIoU **[153]**	85.7	82.7	85.5	92.6	89.0	84.2	256
Proposed Model-SIoU **[154]**	86.3	82.1	85.7	92.8	89.3	84.1	25
Proposed Model-CIoU **[151]**	85.6	82.7	89.2	92.0	90.6	84.1	294

(Note - To visually distinguish performance, the top-3 values are highlighted in red, green, and blue respectively. The metrics are labeled with up arrows to indicate that higher values are indicative of more robust performance, while down arrows indicate that lower values demonstrate superior performance.)

Regarding Precision, the top three performing loss functions are SIoU at 86.3%, GIoU at 85.7%, and CIoU at 85.6%. SIoU, with its unique incorporation of angle, distance, shape, and IoU cost, effectively reduces false positives, thereby enhancing the alignment precision between predicted and ground truth bounding boxes. Conversely, while GIoU measures geometric discrepancies through the area of the smallest enclosing convex hull, its Precision lags slightly behind CIoU and SIoU, potentially due to its slower gradient updates, particularly in cases where the bounding boxes do not overlap.

For Recall metric, EIoU achieves the highest performance at 83.7%, followed by GIoU and CIoU at 82.7%, with SIoU ranking third at 82.1%. EIoU builds on CIoU by further optimizing the separate aspect ratio differences and processes width and height discrepancies, leading to more efficient regression and, consequently, higher recall rates. Although GIoU and CIoU also consider geometric factors, they do not enhance recall as effectively as EIoU in more complex scenarios.

For AP of Healthy Spikelets, the top performers are CIoU at 89.2%, EIoU at 87.8%, and DIoU at 86.3%. CIoU's introduction of aspect ratio consistency significantly improves the accuracy of bounding box regression, resulting in the highest AP values in this context. EIoU follows closely, further refining the matching of predicted and target boxes through its detailed treatment of width and height differences. DIoU, by accelerating model convergence, also achieves robust performance in the AP of healthy spikelets. In the AP of Diseased Spikelets, the best-performing loss functions are SIoU at 92.8%, GIoU at 92.6%, and EIoU at 92.4%. SIoU excels due to its meticulous control over multiple cost components, especially the angle cost, which enables the predicted box to converge more rapidly to the optimal position.

Regarding mAP (Mean Average Precision), CIoU at 90.6%, EIoU at 90.1%, and SIoU at 89.3% are the top performers. CIoU's balanced approach puts it at the forefront of mAP performance. EIoU, through its detailed handling of aspect ratio discrepancies, further enhances overall model precision, while SIoU improves convergence speed and accuracy through directional control. In the F1 score, EIoU at 84.4% and GIoU at 84.2% exhibit the best performance, with SIoU and CIoU tying for third place at 84.1%. The superior balance of Precision and recall in EIoU and GIoU leads to strong F1 scores, reflecting their effectiveness in correct identification and minimizing false detections.

Finally, regarding FPS (Frames Per Second), DIoU, EIoU, and CIoU stand out, with DIoU achieving a remarkable 370 FPS, which indicates that DIoU's straightforward and efficient calculation of center point distance significantly accelerates model inference speed. At the same time, EIoU and CIoU also

maintain high inference efficiency. However, SIoU shows the poorest performance in FPS, with only 25 FPS, which likely results from the increased computational complexity introduced by multiple cost components, leading to a significant reduction in inference speed. Therefore, although SIoU excels in Precision and mAP, it may not suit applications requiring high real-time performance.

4.2.3 Module Performance of Proposed Model

This section delves into the performance evaluation of the proposed model's individual modules. Understanding how each module contributes to the overall model performance is crucial for identifying the strengths and potential limitations of the architectural design.

Table 4.3: Performance of The Proposed Model with Different Modules

DWConv	MSFEF	SPPF	RepC3	AP_h (%) ↑	AP_d (%) ↑	mAP (%) ↑	Parameters(M) ↓
√				84.3	87.1	85.7	4.6
	√			86.4	92.1	89.2	7.2
	√		√	85.0	89.4	87.2	6.7
√		√		86.3	92.0	89.1	4.9
	√	√	√	87.2	92.4	89.8	6.9
√	√	√	√	88.2	92.4	90.3	6.7

(Note - To visually distinguish performance, the top-3 values are highlighted in red, green, and blue respectively. The metrics are labeled with up arrows to indicate that higher values are indicative of more robust performance, while down arrows indicate that lower values demonstrate superior performance.)

In **Table 4.3**, the performance of the proposed model is systematically compared with different module combinations. It highlights the impacts of integrating various modules like Depth-Wise Convolution (DWConv), Spatial Pyramid Pooling Fast (SPPF), RepC3, and the proposed Multi-Scale Feature Enhancement and Fusion (MSFEF) on key performance metrics such as AP for healthy and diseased spikelets, mAP, and the number of parameters.

The baseline model, which exclusively incorporates DWConv, exhibits a mAP of 85.7% with a parameter count of 4.6 million. Though there is potential for improvement, especially in managing diseased spikelets, this setup provides a respectable mix between accuracy and model complexity with AP_h of 84.3% and AP_d of 87.1%.

Significant enhancements are achieved by incorporating the MSFEF module into the baseline. The mAP experiences a 3.5% increase, from 85.7% to 89.2%. Increases in AP_h to 86.4% and AP_d to 92.1% are the main drivers of this improvement. The efficacy of MSFEF in improving multi-scale feature extraction is underscored by these gains, which allow the model to capture shifts in spikelet conditions more accurately. Nevertheless, this is greeted with an increase in parameters from 4.6M to 7.2M, which suggests a compromise between the performance and the complexity of the model.

The mAP is 87.2% when the MSFEF module is added with RepC3, while AP_h and AP_d are 85.0% and 89.4%, respectively. Although this configuration enhances the baseline, the advances are less significant than those achieved by incorporating MSFEF alone. In contrast to the configuration with MSFEF, the parameter count in this case is 6.7M, indicating that RepC3 contributes to performance enhancement while adding less complexity.

The combination of MSFEF and SPPF results in additional enhancements, including a mAP of 89.8%, AP_h at 87.2%, and AP_d at 92.4%. Feature extraction across scales is improved, and spatial hierarchies are preserved in this configuration, which capitalizes on the assets of both modules. Nevertheless, the number of parameters climbs to 6.9 million, representing a modest increase. The introduction of DWConv, designed to reduce parameters, effectively reduces the count to 4.9M when used in conjunction with SPPF. On the other hand, the mAP of 89.1% is marginally lower than that of MSFEF alone, but it remains above the baseline. Although it compromises some accuracy, especially in APh, this decrease in model size makes it a preferable option when computing efficiency is desired.

The final configuration, which integrates DWConv, SPPF, and RepC3, obtains the highest mAP of 90.3%, with AP_h at 88.2% and AP_d at 92.4%. With a parameter count of 6.7 million, this setup is efficient, considering its excellent performance. In order to optimize feature extraction and computational demands, this configuration

strikes a balance between accuracy and model complexity by utilizing the assets of all modules.

The performance of the proposed model can be substantially improved through the strategic integration of modules. By enhancing multi-scale feature extraction, the MSFEF module offers considerable improvements in mAP, while SPPF improves the model's capacity to manage spatial hierarchies. RepC3 is a valuable module for optimizing computational efficiency despite the minor trade-off in accuracy that occurs when the parameter count is reduced.

4.2.4 Effect of Proposed MSFEF Module

In **Table 4.4**, the performance of various detectors was compared before and after applying the proposed MSFEF module across four distinct groups: YOLOv7, YOLOv8-S, GELAN-C, and GELAN-S. These comparisons emphasize the broad applicability and benefits of the MSFEF module across different networks.

Table 4.4: MSFEF Module Performance

Model	Precision (%) ↑	Recall (%) ↑	AP_h (%) ↑	AP_d (%) ↑	mAP (%) ↑	F1↑	Parameters (M) ↓	GFLOPs↓
YOLOv7 [126]	**81.4**	75.6	**82.8**	87.4	**85.1**	78.4	**36.5**	**103.2**
YOLOv7-MSFEF	78.6	**80.5**	80.5	**89.7**	**85.1**	**79.5**	54.5	211.0
YOLOv8-S [149]	84.0	82.4	88.2	**88.0**	88.1	83.2	**11.1**	**28.4**
YOLOv8-S-MSFEF	**84.5**	**83.7**	**89.4**	**88.0**	**88.7**	**84.1**	11.9	32.5
GELAN-C [42]	**88.4**	**86.2**	88.3	91.2	89.8	**87.3**	**25.4**	**102.5**
GELAN-C-MSFEF	86.5	84.7	**90.9**	**92.7**	**91.8**	85.5	28.4	116.2
GELAN-S [42]	**84.7**	81.7	87.4	89.8	88.6	83.1	**6.7**	**26.7**
GELAN-S-MSFEF	84.1	**84.0**	**88.0**	**92.7**	**90.4**	**84.0**	7.4	30.0

(Note - Values in Bold font indicates better performance in this metric. The metrics are labeled with up arrows to indicate that higher values are indicative of more robust performance, while down arrows indicate that lower values demonstrate superior performance.)

As shown in **Table 4.4**, the introduction of the MSFEF module in YOLOv7 led to a noticeable improvement in Recall, increasing from 75.6% to 80.5%, which indicates that MSFEF effectively enhances the model’s ability to capture more relevant objects, especially in simpler models like YOLOv7. However, Precision saw a slight decrease from 81.4% to 78.6%, suggesting that while MSFEF improves Recall, it might introduce some trade-offs in terms of Precision in this particular model. The mAP remained stable at 85.1%, showing that the MSFEF

module effectively balanced the Precision and Recall improvements. Additionally, the AP for Diseased Spikelets increased from 87.4% to 89.7%, further supporting the module's ability to improve detection accuracy for specific object categories.

By implementing the MSFEF module, the YOLOv8-S model showed enhancements in most measures, although the extent of the effects varied. Precision experienced a slight increase from 84.0% to 84.5%, while Recall displayed a more substantial improvement from 82.4% to 83.7%. This enhancement resulted in a rise in mAP from 88.1% to 88.7%, which illustrates that the MSFEF module improves the detection accuracy and the overall balance between Precision and Recall in more advanced models such as YOLOv8-S. The AP for Healthy Spikelets also increased from 88.2% to 89.4%, demonstrating the efficacy of MSFEF in managing healthy spikelet classifications.

The MSFEF module significantly enhanced most metrics within the GELAN-C group. Despite a slight decrease in Precision from 88.4% to 86.5%, the MSFEF module is notably effective in improving the performance of more complex models like GELAN-C, as evidenced by a rise in mAP from 89.8% to 91.8%. Notably, the AP for Healthy Spikelets and Diseased Spikelets experienced substantial increases, rising from 88.3% to 90.9% and 91.2% to 92.7%, respectively.

The MSFEF module also had a positive impact on the GELAN-S. The precision rate had a marginal decline from 84.7% to 84.1%, but the recall rate increased from 81.7% to 84.0%, resulting in a rise in mAP from 88.6% to 90.4%. In particular, the MSFEF module improves the GELAN-S model's precision-recall balance, as evidenced by the F1 score's rise from 83.1% to 84.0%. Furthermore, the AP_d metric significantly rose from 89.8% to 92.7%, indicating the module's efficacy in enhancing model performance for intricate classes.

The MSFEF module's extensive applicability is illustrated by its implementation in these four networks. The overall improvements in Recall, mAP, and specific AP metrics across different networks show that the MSFEF module effectively improves detection performance and maintains a good balance between Precision

and Recall, making it a valuable addition to various network architectures, even though some trade-offs in Precision were observed, especially in simpler models.

4.2.5 Feature Visualization

Visualization of features is essential for understanding the interpretation and processing of input data by deep learning models. This section visualizes the spatial distribution of features at various layers to investigate the internal feature representations of various models, including the proposed one. In complex scenes where accurate spikelet detection is difficult, these visualizations provide valuable insights into the model's capacity to differentiate between foreground and background objects. The comparison of feature maps across models sheds light on the advantages of the proposed model’s architecture in enhancing feature learning.

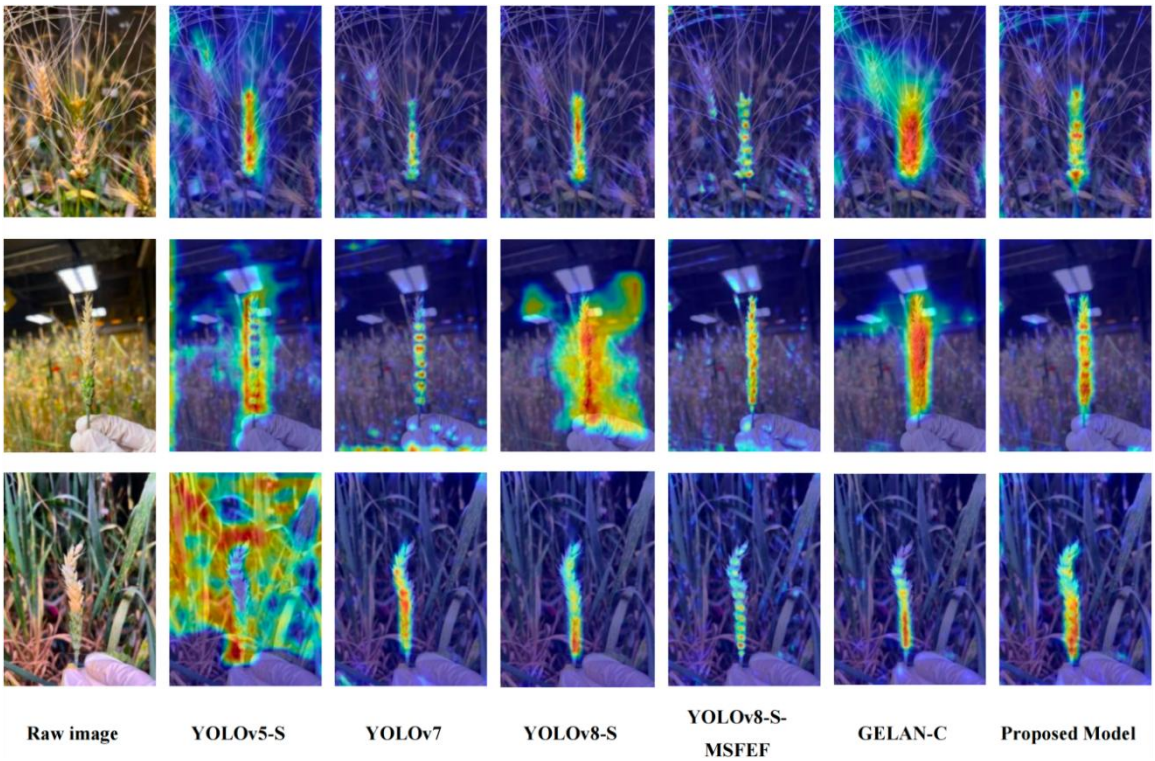

Figure 6: Spatial Distribution of Visualized Features

The Eigen-CAM [155] was utilized to experiment with visualizing features at different layers based on the training weights of different network architectures. As illustrated in **Figure 6**, this experiment selected feature maps generated from layers that process foreground objects in the neck section of each model, enabled understanding how the model classifies and locates important features. The

YOLOv5-S primarily learned through global information perception, benefiting from the backbone of CSPNet's cross-stage feature fusion on background and foreground areas.

The YOLOv7 exhibited the dispersed approach to feature learning, and although the use of re-parameterization techniques enhanced the inference speed, the dynamic structural adjustments introduced instability during the feature learning process. The YOLOv8-S also showed instability in feature transmission in complex backgrounds. The YOLOv8-S-MSFEF utilized multi-scale spatial information to re-utilize spatial feature information, strengthening the distinction between background and foreground, which was sensitive to the morphology of foreground objects with enhanced spatial feature representation by the MSFEF module. The GELAN-C, an efficient aggregated network structure, enhanced feature transmission, offering a more focused and spatially continuous representation of foreground object features. The proposed model fusing the MSFEF module, showed a more detailed perception of wheat spikelet edge contours and morphology than other models, enhancing the focus on global feature information.

4.2.6 Visualization of Model Predictions

Visualizing the predictions generated by various models offers a visible comprehension of their performance in real-world scenarios. This section comprehensively analyzes the prediction outputs from various models, including the proposed model. The practical efficacy of the models can be assessed by examining the detection results, particularly in terms of false positives and false negatives, and the precise identification of diseased versus healthy spikelets.

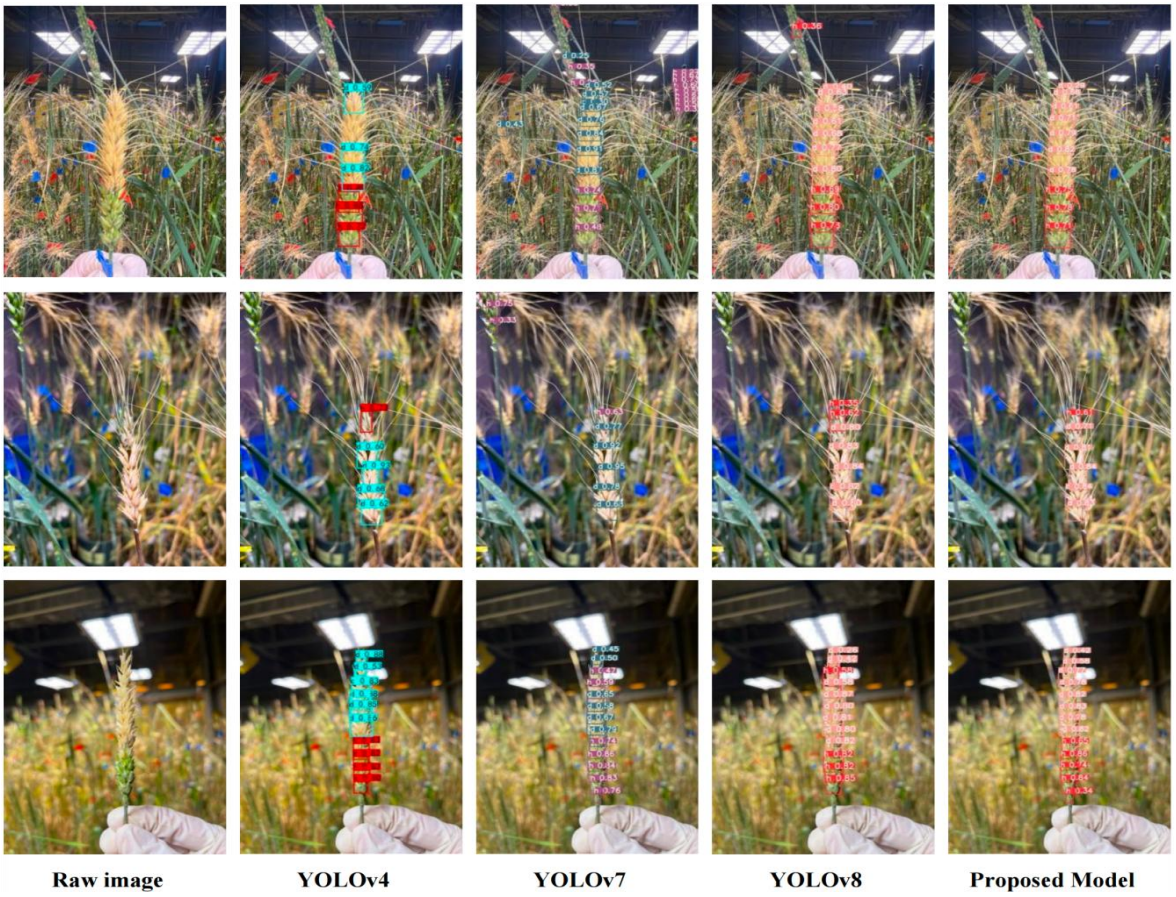

Figure 7: Spikelet Detection Results

As depicted in Figure 7, the YOLOv8-S demonstrated confusion between spikelets and complex backgrounds during prediction, resulting in some false detections. The YOLOv7 displayed more severe perplexity, as it was unable to distinguish between foreground and background spikelets. Conversely, the YOLOv4 experienced significant under-detection as it struggled to learn the features of spikelets, resulting in significant errors in phenotypic extraction. The Proposed Model concentrated on spatial details, utilizing the MSFEF modules to synthesize background and foreground feature information. Consequently, predictions were directed toward the single spikelet as foreground objects. Compared to other models, the proposed model exhibited a reduced rate of missed detections and increased accuracy in classifying infected spikelets, thereby illustrating the potential for precise phenotypic extraction.

4.2.7 Phenotype Assessment

This experiment evaluated the results of manual counting and the proposed model predictions on the test set for phenotypic extraction. Manual Count DSR and Digital DSR represent the DSR determined through manual counting and automatic extraction of proposed model respectively. Manual Count FHB Severity and Digital FHB Severity represent the severity of disease measured manually and extracted by proposed model.

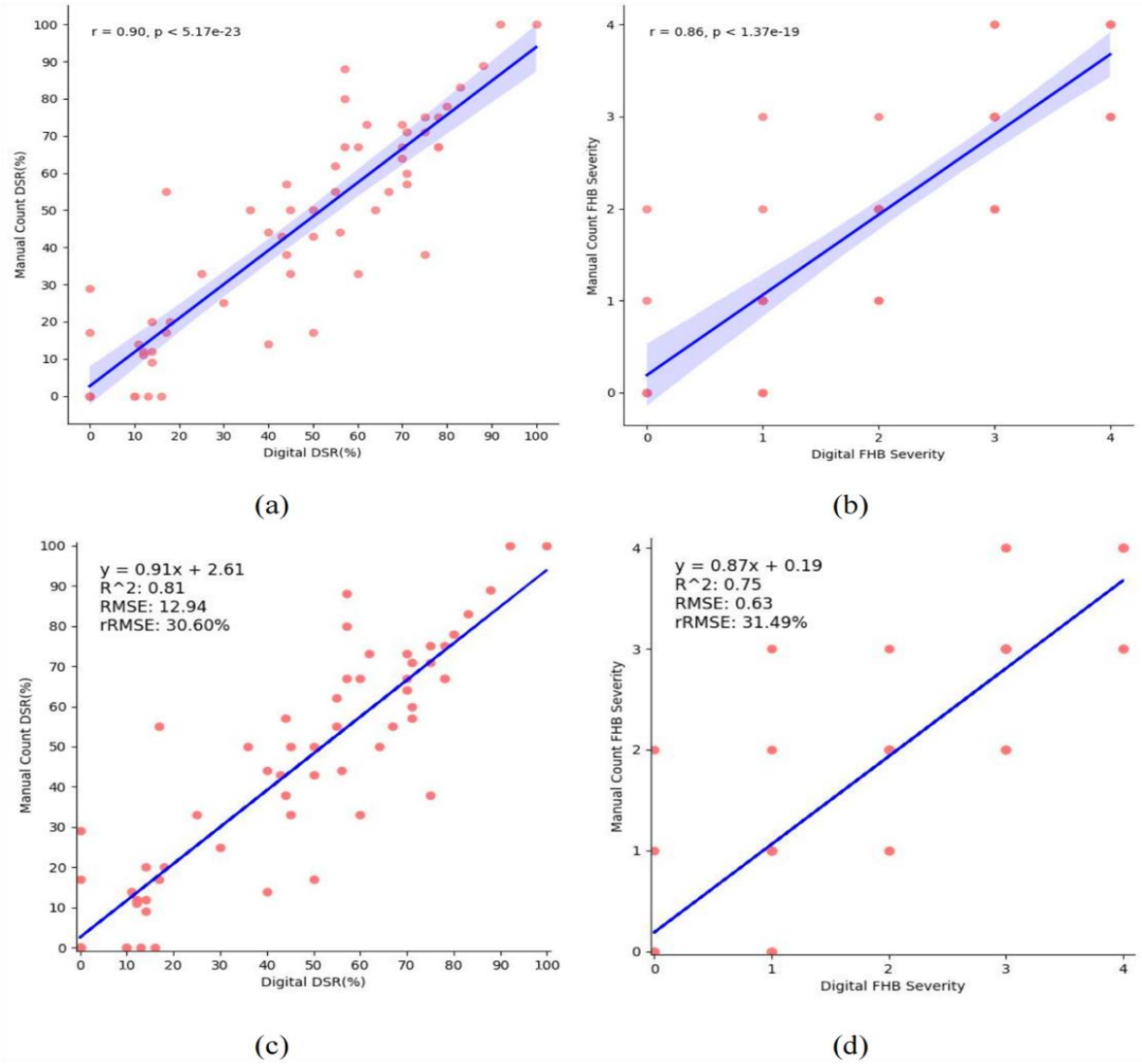

Figure 8: Pearson Correlation and Regression Analyses on DSR and FHB Severity

(Note - Subfigure (a) and subfigure (b) present the correlation analysis results between the DSR phenotype and FHB severity phenotype extracted manually and by the model. Subfigure (c) and subfigure (d) display the regression analysis results between the DSR phenotype and FHB severity phenotype extracted manually and by the model.)

Through **Figure 8** (a) and (b), the Pearson correlation coefficient for DSR and disease severity were 0.90 and 0.86 respectively, indicating a high correlation between manually recorded and automatically extracted DSR and disease severity. Additionally, regression analysis results in (c) and (d) suggested that the automatic extraction was slightly underestimated compared to manual observations, with slopes of 0.91 for DSR and 0.87 for disease severity. Both DSR and disease severity phenotyping methods exhibited strong linear relationships, with coefficients of 0.81 and 0.75, respectively.

Considering the broad range of DSR percentages, the RMSE and rRMSE for DSR extraction were 12.94 and 30.60% respectively. Notably, at higher infection levels, the data points were more dispersed, and accuracy was relatively lower, possibly due to an increase in the number of infected spikelets and the diversity of disease symptoms increasing detection difficulty. The rRMSE for disease severity extraction reaches 31.49%, as the calculation of disease severity is sensitive to the breakpoints between different grades of DSR, where minor differences in DSR might lead to larger discrepancies in disease severity, posing higher demands on the recall.

4.3 DISCUSSION

The experimental results presented in this chapter demonstrate the effectiveness and efficiency of the proposed model, particularly when integrated with the MSFEF module, across various state-of-the-art models.

Firstly, the proposed model exhibits outstanding performance in mean Average Precision (mAP) and Average Precision (AP), especially in detecting healthy and diseased spikelets, which can be attributed to the incorporation of the MSFEF (Multi-Scale Feature Enhancement and Fusion) module. The MSFEF module significantly enhances the model's ability to capture and process multi-scale features [107]. This module allows the model to better distinguish between fine-grained details, crucial in detecting subtle disease symptoms in complex field backgrounds. For instance, the YOLOv8-S-MSFEF variant demonstrated an improvement in both precision and recall, leading to a higher mAP than its base YOLOv8-S counterpart, suggesting that the MSFEF module's ability to enhance

feature representation at multiple scales is a critical factor in the model's better performance.

The challenges in detecting diseased wheat spikelets include distinguishing complex backgrounds from foregrounds due to varied scene compositions and the intra-class variability within the task [156]. Generally, the proposed model exhibits higher average precision in detecting diseased spikelets than healthy ones. It could be attributed to the visible symptom diversity presented by spikelets at different stages of the disease, where varying symptoms lead to significant intra-class variations [157], making the extraction of texture and color features more challenging for the models. Thus, the model continually learned the diverse and complex features of infected spikelets during training. Conversely, healthy spikelets display more uniform colors and textures, with the primary variations stemming from morphological differences across varieties, making it easier for models to learn detection patterns and find shortcuts rather than learning features of healthy spikelets.

From a dataset perspective, the high resolution and diversity of scenes in the dataset used in this study pose a challenge to the generalization capabilities of the models. Some models, like GELAN-C, needed more accuracy when processing these high-resolution, complex background images due to the difficulty balancing global and detailed feature extraction when aggregating features across scales [42]. GELAN-C employs an efficient hierarchical feature aggregation network that captures global features across layers, making it particularly effective in detecting large-scale objects [158]. However, the additional computational complexity introduced during feature aggregation can lead to a decline in accuracy when dealing with small-scale objects, such as spikelets. Therefore, while GELAN-C performs well in specific tasks, its accuracy can be compromised when faced with multi-scale feature extraction and processing challenges.

In performance comparison experiments, the RT-DETR-L model demonstrated exceptional performance in specific vital metrics, particularly in F1 and recall scores. Its unique architecture leverages a Transformer-based feature extraction mechanism that captures global contextual information more effectively, thus

enhancing detection accuracy in complex backgrounds. Compared to other models based on conventional CNNs, RT-DETR-L's Transformer module effectively handles cross-scale features [150], making it more stable in detecting spikelets of varying sizes [159]. However, the model's high precision comes at the cost of increased computational demands, resulting in longer inference times than lighter models like YOLOv7.

YOLOv7's rapid inference capability, aided by re-parameterization techniques [160], simplifies the computational process during inference, significantly boosting the model's speed [126]. Nevertheless, this simplification compromises the stability of feature learning, particularly in complex backgrounds, leading to a decrease in precision. Consequently, while YOLOv7 excels in speed, its ability to accurately recognize features in complex backgrounds is weaker, which explains its higher false positive rate.

The detailed comparison of loss functions also underscores the importance of selecting appropriate loss metrics tailored to specific model requirements. Different IoU loss functions exhibit significant differences in training and inference performance. CIoU provides the best overall performance through comprehensive geometric considerations [151], while DIoU and SIoU demonstrate advantages under specific conditions through different optimization strategies. Although GIoU improves upon the shortcomings of traditional IoU, it may encounter slow convergence issues in complex situations.

Among these, the SIoU loss function introduces directional penalty angles, enabling prioritized alignment along the X or Y axis during distance minimization, thereby accelerating convergence [154]. By reducing the degrees of freedom that might arise during the training process, the convergence path of the bounding boxes becomes more direct. GIoU introduces an additional penalty term to address non-overlapping bounding boxes. GIoU extends the bounding box to the minimum enclosing box, providing a moving gradient even when the boxes do not overlap [153]. However, the GIoU loss function may degrade into IoU loss when handling horizontal or vertical bounding boxes, leading to slower convergence. EIoU, an optimization of GIoU, aims to improve training speed by simplifying the

computational process. EIoU achieves more efficient convergence by reducing unnecessary computational steps, although it may sacrifice some precision [152].

DIoU, the predecessor of CIoU, achieves faster convergence by adding a penalty term for the center point distance on top of IoU loss. DIoU directly minimizes the distance between the predicted and ground-truth box centers, making it particularly effective in handling boxes with different scales and aspect ratios [151]. This characteristic significantly makes DIoU more efficient in processing complex scenes when the boxes do not overlap. However, despite its faster convergence, DIoU's consideration of shape consistency is not as comprehensive as CIoU, resulting in slightly lower mAP than CIoU. CIoU considers the center point distance between the bounding boxes and includes aspect ratio consistency. Through additional angular constraints [151], this extension allows CIoU to provide better geometric matching across all aspects of the bounding boxes, leading to faster and more accurate convergence.

Feature visualization and model prediction analyses further validate the robustness of the proposed model, particularly in its ability to focus on the morphological characteristics of spikelets. The proposed model's superior performance in extracting phenotypic data from images, as demonstrated by the high Pearson correlation coefficients and linear regression analysis results, indicates its potential to replace or supplement manual phenotyping methods. However, the slight underestimation observed in automatic extraction suggests areas for further refinement, particularly in handling high infection levels and diverse symptomatology.

The experimental results reveal the strengths and weaknesses of different models in handling complex agricultural scenarios. However, despite their exceptional resource efficiency and speed, YOLO series models must be improved for delicate detection tasks. The RT-DETR-L exhibits an exceptionally high level of detection accuracy due to its global feature integration capabilities; however, it may sacrifice some inference speed. GELAN-C performs exceptionally well in feature aggregation but displays constraints when effectively processing small-scale objects. Although the MSFEF module has dramatically improved the proposed

model's ability to extract features, there is still an opportunity for more optimization in the loss function design. These results offer essential guidance for the development of future models and indicate potential areas for further optimizing the model's performance in various tasks.

4.4 CHAPTER SUMMARY

This chapter shows the proposed model's experimental setup, results, and analyses compared to those of various state-of-the-art detectors. Performance metrics, such as precision, recall, AP, mAP, GFLOPs, and FPS, were implemented in the experiments. The results indicated that the MSFEF module, when incorporated into the proposed model, substantially enhances real-time performance and detection accuracy, particularly in intricate agricultural scenarios.

The chapter also investigated the influence of various loss functions on the model's performance, concluding that the CIoU loss function is the most effective in preserving a high balance between precision and recall. In addition, the MSFEF module's critical role in improving detection accuracy and feature processing was underscored by the evaluation of the proposed model's performance with various module combinations.

Furthermore, the proposed model's capacity to precisely concentrate on the spatial details of spikelets was demonstrated through feature visualizations and model predictions, which improved its phenotypic extraction capabilities. Lastly, the high correlation between manual and automatic phenotypic extraction was confirmed by the Pearson correlation and regression analyses. Chapter 5 concisely summarizes the research study, acknowledging its limits and offering recommendations.

CHAPTER 5 – CONCLUSION, LIMITATION, AND RECOMMENDATIONS

This chapter offers a concise overview of this study's conclusion, limitations, and future research prospects.

5.1 SUMMARY

The primary focus of this work is to address the difficulty of precisely identifying Fusarium Head Blight (FHB) at the spikelet level in wheat, especially in cases of intricate and diverse field backgrounds. Fusarium head blight (FHB) is a highly destructive disease significantly affecting wheat productivity. It is essential to detect FHB at the spikelet level early and accurately to treat the disease. However, complex field conditions with changing lighting, occlusions, and overlapping spikelets make it harder to identify accurately. Therefore, reliable and sophisticated detection algorithms have to be developed.

The first research objective was to critically review and analyze the literature on multi-scale feature extraction and fusion techniques, pivotal for enhancing object detection models' performance in complex scenarios like FHB detection. A thorough and methodical literature analysis was conducted to achieve the first research objective, identifying and summarizing the main trends, difficulties, and gaps in the state-of-the-art approaches for FHB detection and general object detection in agricultural settings.

The literature review uncovered several significant findings. Firstly, current models generally require scale variance and feature fusion assistance, which is crucial for identifying tiny and tightly clustered spikelets. Furthermore, the review emphasized that multi-scale feature extraction and fusion are essential to enhance detection accuracy, particularly in complex agricultural settings. These observations directly influenced the development of the proposed model, guaranteeing that it was founded on a thorough comprehension of the most recent state-of-the-art methodologies and that it addressed the identified constraints of previous methods.

The second research objective was to develop a novel object detection model, leveraging the GELAN framework of YOLOv9 and incorporating the Multi-Scale Feature Enhancement and Fusion (MSFEF) module. This objective was achieved by developing an advanced object detection model that efficiently incorporates multi-scale feature fusion methods.

The model was developed through an iterative process that involved meticulous testing and fine-tuning with a self-constructed dataset. This dataset was meticulously curated to accurately represent a variety of FHB symptoms and field conditions, guaranteeing that the model could generalize effectively across various scenarios. The model produced a high accuracy level and showed robustness and flexibility in response to different field conditions.

The third research objective involved the rigorous evaluation of the proposed model's performance against several state-of-the-art models, including YOLOv7, YOLOv8-S, and GELAN-C, using the self-constructed dataset. This objective was achieved through a detailed evaluation process. The results indicated that the proposed model outperformed the baseline models in most metrics, particularly in mAP and AP, which are essential for evaluating the accuracy of object detection models.

Additionally, the proposed model exhibited robustness in real-time applications, displaying a competitive FPS score and maintaining high performance without substantial degradation in speed. In this assessment, the self-constructed dataset was instrumental in testing the model's generalizability by providing a realistic and diverse set of scenarios.

The study rigorously evaluated and developed a proposed object detection model to address the critical challenge of accurately detecting FHB at the spikelet level, particularly in complex field backgrounds. The methodologies and findings presented can potentially enhance the management of diseases in crop production, thereby contributing to future advancements in agricultural object detection.

5.2 CONCLUSION

This study introduced a novel object detection model designed to detect FHB infection and extract phenotypic data from wheat spikelets. The proposed model, which incorporates a Multi-Scale Feature Enhancement and Fusion (MSFEF) module, exhibited numerous advantages over state-of-the-art (SOTA) models, particularly regarding real-time performance and accuracy.

For instance, it achieved an AP of 89.2% for healthy spikelets and 92% for diseased spikelets, resulting in a mAP of 90.6%. These results surpass the mAP of competitive models, including YOLOv9-C at 89.6% and GELAN-C at 89.8%. Furthermore, the model demonstrated exceptional real-time performance, achieving a remarkable 294 FPS, considerably higher than other models, including YOLOv8-S at 58 FPS and RT-DETR-L at 29 FPS.

Additionally, the MSFEF module improved the model's detection capabilities, particularly in intricate agricultural environments. The model obtained a more balanced precision-recall trade-off by enhancing feature representation, resulting in a high F1 score of 84.1%. The phenotype extraction results also demonstrated a Pearson correlation coefficient of 0.90 for the DSR and 0.86 for disease severity, suggesting a robust correlation with manual observations and validating its efficacy in real-world applications.

In conclusion, the proposed model has the potential to be a dependable tool for identifying FHB in wheat spikelets and collecting important phenotypic information, which could advance precision agriculture methods since it outperforms current models across various criteria.

5.3 LIMITATION AND RECOMMENDATION

This study has limitations that could affect the generalizability and scalability of the proposed model. These constraints include the dataset's size, the learning paradigm, and the focus on a single wheat spike. The model's applicability across various wheat species, environmental conditions, and larger-scale agricultural settings could be improved by addressing these limitations in future research.

First, the dataset used to train and test the algorithm was constrained to wheat species and particular environmental conditions. This narrow emphasis may limit the model's generalization capacity to different wheat species or environmental conditions. Therefore, further studies should consider enlarging the dataset to incorporate a wider variety of wheat species and photos taken in various environmental settings. Data collection on various wheat types will become a pressing problem. Although this study has demonstrated the potential of methods for distinguishing between healthy and diseased spikelets, the available datasets are restricted in size. Improving performance may necessitate using larger datasets encompassing a wide range of samples to train more generalized large models.

Second, a significant amount of labeled data is needed for the FHB detection strategy used in this study, which is based on a fully supervised learning paradigm. Although effective, this method is time-consuming and expensive during the data labeling phase. Future studies should investigate semi-supervised or alternative learning paradigms that could provide equivalent performance at lower annotation costs to overcome this constraint.

The study's emphasis on individual wheat spikes also poses a drawback because it needs to represent the complexity of actual agricultural settings adequately. The phenotyping of entire wheat plots in field conditions is necessary for advanced breeding research, which demands the identification of disease phenotypes on a larger spatial scale. Particularly in detecting small objects within a larger field, this presents new challenges for the model. Thus, future research may prioritize the creation of automated phenotyping methods that are capable of processing whole wheat plots and the development of disease detection models that are optimized for small object detection.

CHAPTER 6 – REFLECTIONS

It has been a significant journey of growth, learning, and perseverance to complete this endeavor. Particularly considering the difficulty of creating and testing a new model, this procedure needed meticulous preparation and efficient time management. I learned from this experience how important it is to set reasonable objectives, divide work into digestible chunks, and be punctual. Challenges, particularly when experimental results fell short of expectations, highlighted the importance of perseverance and iterative development.

Maintaining a planned calendar, regular check-ins with my supervisor, and clear milestones for each research phase were all essential components of my project management approach. When obstacles arose, this approach enabled quick corrections and helped the project stay on target. I also realized, though, that one should better predict possible delays and create extra buffer time for unanticipated difficulties.

I found university resources, notably access to academic databases and libraries such as ScienceDirect and Web of Science, to be invaluable in establishing my research in the existing literature. The direction of my supervisor on research approaches and academic writing was very important since it helped me to clarify basic ideas and increase the coherence of my academic work. Powerful AI tools like Quillbot and Grammarly also helped me improve the quality of my writing, fix my grammar mistakes, and make my language more professional. But I was wary of their shortcomings since, as my supervisor had stressed, I needed to carefully evaluate what they suggested. Additionally, improvements of my project was significantly influenced by communication with group members. Talking and giving each other feedback produced a supportive learning environment that helped me find gaps and improve the details of my work.

During this study, I learned a lot of useful technical skills, especially in developing deep learning models and planning experiments. Building a self-labeled dataset from unprocessed agricultural images was particularly difficult. From annotating

data to dividing it into training, validation, and test sets, every step needed painstaking precision.

Additionally, I developed my ability to handle the deep learning platforms like PyTorch, which are widely used in the development and optimization of the proposed model. Knowledge of CNNs and how they work for object detection was necessary for integrating complex modules like the MSFEF. Also, I became good at assessing how well models worked using various metrics, which helped guarantee that my findings were thorough and applicable to the study's objectives. During the project's experimental phase, the training environment had to be set up and optimized. This included handling long training sessions and configuring GPU resources.

Planning ahead, always learning, and making the most of what you have are all research essentials, as this study demonstrated. My background in these areas has given me the tools I need to design and execute thorough experiments and interpret the results with confidence. Both my academic and professional pursuits in the future will benefit greatly from these abilities.

REFERENCES

[1] M. P. Reynolds and H.-J. Braun, *Wheat improvement: food security in a changing climate*. Springer Nature, 2022.

[2] C. Dweba *et al.*, "Fusarium head blight of wheat: Pathogenesis and control strategies," *Crop protection,* vol. 91, pp. 114-122, 2017.

[3] E. Mielniczuk and B. Skwaryło-Bednarz, "Fusarium head blight, mycotoxins and strategies for their reduction," *Agronomy,* vol. 10, no. 4, p. 509, 2020.

[4] S. N. Wegulo *et al.*, "Effects of integrating cultivar resistance and fungicide application on Fusarium head blight and deoxynivalenol in winter wheat," *Plant disease,* vol. 95, no. 5, pp. 554-560, 2011.

[5] M. Comby *et al.*, "Screening of wheat endophytes as biological control agents against Fusarium head blight using two different in vitro tests," *Microbiological research,* vol. 202, pp. 11-20, 2017.

[6] B. Steiner, M. Buerstmayr, S. Michel, W. Schweiger, M. Lemmens, and H. Buerstmayr, "Breeding strategies and advances in line selection for Fusarium head blight resistance in wheat," *Tropical Plant Pathology,* vol. 42, pp. 165-174, 2017.

[7] Y. Liu *et al.*, "Genetic mapping and prediction analysis of FHB resistance in a hard red spring wheat breeding population," *Frontiers in plant science,* vol. 10, p. 465465, 2019.

[8] M. Buerstmayr, B. Steiner, and H. Buerstmayr, "Breeding for Fusarium head blight resistance in wheat—Progress and challenges," *Plant breeding,* vol. 139, no. 3, pp. 429-454, 2020.

[9] H. W. Schroeder, J. J. Christensen, J. D. Christensen, J. Platz-Christensen, and H. W. Schroeder, "Factors affecting resistance of Wheat to scab caused by Gibberella zeae," *Phytopathology,* vol. 53, pp. 831-838, 1963.

[10] C. S. Bekkering, J. Huang, and L. Tian, "Image-based, organ-level plant phenotyping for wheat improvement," *Agronomy,* vol. 10, no. 9, p. 1287, 2020.

[11] S. Francesconi, A. Harfouche, M. Maesano, and G. M. Balestra, "UAV-based thermal, RGB imaging and gene expression analysis allowed detection of Fusarium head blight and gave new insights into the physiological responses to the disease in durum wheat," *Frontiers in plant science,* vol. 12, p. 628575, 2021.

[12] M. E. Rieker, M. A. Lutz, A. El-Hasan, S. Thomas, and R. T. Voegele, "Hyperspectral Imaging and Selected Biological Control Agents for the Management of Fusarium Head Blight in Spring Wheat," *Plants,* vol. 12, no. 20, p. 3534, 2023.

[13] D.-Y. Zhang *et al.*, "Integrating spectral and image data to detect Fusarium head blight of wheat," *Computers and electronics in agriculture,* vol. 175, p. 105588, 2020.

[14] M. B. Almoujahed *et al.*, "Detection of fusarium head blight in wheat

under field conditions using a hyperspectral camera and machine learning," *Computers and Electronics in Agriculture,* vol. 203, p. 107456, 2022.
[15] G. Mustafa *et al.*, "Fusarium head blight monitoring in wheat ears using machine learning and multimodal data from asymptomatic to symptomatic periods," *Frontiers in Plant Science,* vol. 13, p. 1102341, 2023.
[16] A.-K. Mahlein, E. Alisaac, A. Al Masri, J. Behmann, H.-W. Dehne, and E.-C. Oerke, "Comparison and combination of thermal, fluorescence, and hyperspectral imaging for monitoring fusarium head blight of wheat on spikelet scale," *Sensors,* vol. 19, no. 10, p. 2281, 2019.
[17] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature,* vol. 521, no. 7553, pp. 436-444, 2015.
[18] H. Tian, T. Wang, Y. Liu, X. Qiao, and Y. Li, "Computer vision technology in agricultural automation—A review," *Information Processing in Agriculture,* vol. 7, no. 1, pp. 1-19, 2020.
[19] R. Qiu, C. Yang, A. Moghimi, M. Zhang, B. J. Steffenson, and C. D. Hirsch, "Detection of fusarium head blight in wheat using a deep neural network and color imaging," *Remote Sensing,* vol. 11, no. 22, p. 2658, 2019.
[20] R. U. Khan, K. Khan, W. Albattah, and A. M. Qamar, "Image-based detection of plant diseases: from classical machine learning to deep learning journey," *Wireless Communications and Mobile Computing,* vol. 2021, pp. 1-13, 2021.
[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems,* vol. 25, 2012.
[22] J. G. Barbedo, "Factors influencing the use of deep learning for plant disease recognition," *Biosystems engineering,* vol. 172, pp. 84-91, 2018.
[23] Y. Gao, H. Wang, M. Li, and W.-H. Su, "Automatic tandem dual blendmask networks for severity assessment of wheat fusarium head blight," *Agriculture,* vol. 12, no. 9, p. 1493, 2022.
[24] D.-Y. Zhang *et al.*, "Assessment of the levels of damage caused by Fusarium head blight in wheat using an improved YoloV5 method," *Computers and Electronics in Agriculture,* vol. 198, p. 107086, 2022.
[25] Q. Hong *et al.*, "A lightweight model for wheat ear fusarium head blight detection based on RGB images," *Remote Sensing,* vol. 14, no. 14, p. 3481, 2022.
[26] W. Bao, C. Huang, G. Hu, B. Su, and X. Yang, "Detection of Fusarium head blight in wheat using UAV remote sensing based on parallel channel space attention," *Computers and Electronics in Agriculture,* vol. 217, p. 108630, 2024.
[27] X. Meng, C. Li, J. Li, X. Li, F. Guo, and Z. Xiao, "YOLOv7-MA: Improved YOLOv7-Based Wheat Head Detection and Counting," *Remote Sensing,* vol. 15, no. 15, p. 3770, 2023.

[28] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 3-19.
[29] L. Yang, R.-Y. Zhang, L. Li, and X. Xie, "Simam: A simple, parameter-free attention module for convolutional neural networks," in *International conference on machine learning*, 2021: PMLR, pp. 11863-11874.
[30] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "Ghostnet: More features from cheap operations," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 1580-1589.
[31] R. Mao, Z. Wang, F. Li, J. Zhou, Y. Chen, and X. Hu, "Gseyolox-s: An improved lightweight network for identifying the severity of wheat fusarium head blight," *Agronomy,* vol. 13, no. 1, p. 242, 2023.
[32] R. Chandana and A. Ramachandra, "Real time object detection system with YOLO and CNN models: A review," *arXiv Prepr. arXiv2208,* vol. 773, 2022.
[33] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8759-8768.
[34] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2117-2125.
[35] F. Sun, N. He, R. Li, X. Wang, and S. Xu, "GD-PAN: a multiscale fusion architecture applied to object detection in UAV aerial images," *Multimedia Systems,* vol. 30, no. 3, p. 143, 2024.
[36] S. Chen *et al.*, "Info-FPN: An Informative Feature Pyramid Network for object detection in remote sensing images," *Expert Systems with Applications,* vol. 214, p. 119132, 2023.
[37] Y. Luo *et al.*, "CE-FPN: enhancing channel information for object detection," *Multimedia Tools and Applications,* vol. 81, no. 21, pp. 30685-30704, 2022.
[38] J. Zhao, H. Zhu, and L. Niu, "BiTNet: a lightweight object detection network for real-time classroom behavior recognition with transformer and bi-directional pyramid network," *Journal of King Saud University-Computer and Information Sciences,* vol. 35, no. 8, p. 101670, 2023.
[39] C.-Y. Wang, H.-Y. M. Liao, and I.-H. Yeh, "Designing network design strategies through gradient path analysis," *arXiv preprint arXiv:2211.04800,* 2022.
[40] Y. Lee, J.-w. Hwang, S. Lee, Y. Bae, and J. Park, "An energy and GPU-computation efficient backbone network for real-time object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2019, pp. 0-0.
[41] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "Scaled-yolov4: Scaling cross stage partial network," in *Proceedings of the IEEE/cvf*

conference on computer vision and pattern recognition, 2021, pp. 13029-13038.
[42] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, "Yolov9: Learning what you want to learn using programmable gradient information," *arXiv preprint arXiv:2402.13616,* 2024.
[43] C. Gao *et al.*, "A fast and lightweight detection model for wheat fusarium head blight spikes in natural environments," *Computers and Electronics in Agriculture,* vol. 216, p. 108484, 2024.
[44] E. Alisaac and A.-K. Mahlein, "Fusarium head blight on wheat: biology, modern detection and diagnosis and integrated disease management," *Toxins,* vol. 15, no. 3, p. 192, 2023.
[45] X. Wang, H. Li, X. Yue, and L. Meng, "A comprehensive survey on object detection YOLO," *Proceedings http://ceur-ws. org ISSN,* vol. 1613, p. 0073, 2023.
[46] J. D. Salgado, L. V. Madden, and P. A. Paul, "Quantifying the effects of Fusarium head blight on grain yield and test weight in soft red winter wheat," *Phytopathology,* vol. 105, no. 3, pp. 295-306, 2015.
[47] J. Miao, G.-P. Zhang, S.-J. Zhang, J.-Q. Ma, and Y.-Q. Wu, "The orange wheat blossom midge promotes fusarium head blight disease, posing a risk to wheat production in northern China," *Acta Ecologica Sinica,* vol. 43, no. 1, pp. 112-116, 2023.
[48] S. R. Pirgozliev, S. G. Edwards, M. C. Hare, and P. Jenkinson, "Strategies for the control of Fusarium head blight in cereals," *European Journal of Plant Pathology,* vol. 109, pp. 731-742, 2003.
[49] A. H. van Bruggen, A. Gamliel, and M. R. Finckh, "Plant disease management in organic farming systems," *Pest Management Science,* vol. 72, no. 1, pp. 30-44, 2016.
[50] R. T. Furbank and M. Tester, "Phenomics–technologies to relieve the phenotyping bottleneck," *Trends in plant science,* vol. 16, no. 12, pp. 635-644, 2011.
[51] M. A. Shahin and S. J. Symons, "Detection of Fusarium damaged kernels in Canada Western Red Spring wheat using visible/near-infrared hyperspectral imaging and principal component analysis," *Computers and electronics in agriculture,* vol. 75, no. 1, pp. 107-112, 2011.
[52] R. Dhariwal, M. A. Henriquez, C. Hiebert, C. A. McCartney, and H. S. Randhawa, "Mapping of major Fusarium head blight resistance from Canadian wheat cv. AAC Tenacious," *International Journal of Molecular Sciences,* vol. 21, no. 12, p. 4497, 2020.
[53] J. G. Barbedo, C. S. Tibola, and J. M. Fernandes, "Detecting Fusarium head blight in wheat kernels using hyperspectral imaging," *Biosystems Engineering,* vol. 131, pp. 65-76, 2015.
[54] F. Martinelli *et al.*, "Advanced methods of plant disease detection. A review," *Agronomy for sustainable development,* vol. 35, pp. 1-25, 2015.
[55] S. Derveaux, J. Vandesompele, and J. Hellemans, "How to do

successful gene expression analysis using real-time PCR," *Methods,* vol. 50, no. 4, pp. 227-230, 2010.
[56] J. Van der Wolf, J. Van Beckhoven, P. Bonants, and C. Schoen, "New technologies for sensitive and specific routine detection of plant pathogenic bacteria," in *Plant Pathogenic Bacteria: Proceedings of the 10th International Conference on Plant Pathogenic Bacteria, Charlottetown, Prince Edward Island, Canada, July 23–27, 2000*, 2001: Springer, pp. 75-77.
[57] M. F. Clark and A. Adams, "Characteristics of the microplate method of enzyme-linked immunosorbent assay for the detection of plant viruses," *Journal of general virology,* vol. 34, no. 3, pp. 475-483, 1977.
[58] N. Jahan, P. Flores, Z. Liu, A. Friskop, J. J. Mathew, and Z. Zhang, "Detecting and distinguishing wheat diseases using image processing and machine learning algorithms," in *2020 ASABE Annual international virtual meeting*, 2020: American Society of Agricultural and Biological Engineers, p. 1.
[59] D. Zhang, Z. Wang, N. Jin, C. Gu, Y. Chen, and Y. Huang, "Evaluation of efficacy of fungicides for control of wheat fusarium head blight based on digital imaging," *IEEE Access,* vol. 8, pp. 109876-109890, 2020.
[60] L. Huang, T. Li, C. Ding, J. Zhao, D. Zhang, and G. Yang, "Diagnosis of the severity of Fusarium head blight of wheat ears on the basis of image and spectral feature fusion," *Sensors,* vol. 20, no. 10, p. 2887, 2020.
[61] Y.-H. Wang, J.-J. Li, and W.-H. Su, "An integrated multi-model fusion system for automatically diagnosing the severity of wheat Fusarium head blight," *Agriculture,* vol. 13, no. 7, p. 1381, 2023.
[62] W.-H. Su *et al.*, "Automatic evaluation of wheat resistance to fusarium head blight using dual mask-RCNN deep learning frameworks in computer vision," *Remote sensing,* vol. 13, no. 1, p. 26, 2020.
[63] R. C. Bernardes *et al.*, "Deep-learning approach for fusarium head blight detection in wheat seeds using low-cost imaging technology," *Agriculture,* vol. 12, no. 11, p. 1801, 2022.
[64] M. Qiu *et al.*, "Raman spectroscopy and improved inception network for determination of FHB-infected wheat kernels," *Foods,* vol. 11, no. 4, p. 578, 2022.
[65] C. Gu, D. Wang, H. Zhang, J. Zhang, D. Zhang, and D. Liang, "Fusion of deep convolution and shallow features to recognize the severity of wheat Fusarium head blight," *Frontiers in Plant Science,* vol. 11, p. 599886, 2021.
[66] S. Ramesh, R. Hebbar, M. Niveditha, R. Pooja, N. Shashank, and P. Vinod, "Plant disease detection using machine learning," in *2018 International Conference on Design Innovations for 3Cs Compute Communicate Control (ICDI3C)*, 2018: IEEE, pp. 41-45.
[67] T. Van Klompenburg, A. Kassahun, and C. Catal, "Crop yield prediction using machine learning: A systematic literature review,"

Computers and electronics in agriculture, vol. 177, p. 105709, 2020.
[68] R. Singh, S. Srivastava, and R. Mishra, "AI and IoT based monitoring system for increasing the yield in crop production," in *2020 International Conference on Electrical and Electronics Engineering (ICE3)*, 2020: IEEE, pp. 301-305.
[69] L. Liu, Y. Dong, W. Huang, X. Du, and H. Ma, "Monitoring wheat fusarium head blight using unmanned aerial vehicle hyperspectral imagery," *Remote Sensing,* vol. 12, no. 22, p. 3811, 2020.
[70] L. Huang, Z. Wu, W. Huang, H. Ma, and J. Zhao, "Identification of fusarium head blight in winter wheat ears based on fisher's linear discriminant analysis and a support vector machine," *Applied Sciences,* vol. 9, no. 18, p. 3894, 2019.
[71] J. Zhao, Y. Fang, G. Chu, H. Yan, L. Hu, and L. Huang, "Identification of leaf-scale wheat powdery mildew (Blumeria graminis f. sp. Tritici) combining hyperspectral imaging and an SVM classifier," *Plants,* vol. 9, no. 8, p. 936, 2020.
[72] W. Haider, A.-U. Rehman, N. M. Durrani, and S. U. Rehman, "A generic approach for wheat disease classification and verification using expert opinion for knowledge-based decisions," *IEEE Access,* vol. 9, pp. 31104-31129, 2021.
[73] E. Alisaac, J. Behmann, M. T. Kuska, H.-W. Dehne, and A.-K. Mahlein, "Hyperspectral quantification of wheat resistance to Fusarium head blight: Comparison of two Fusarium species," *European Journal of Plant Pathology,* vol. 152, pp. 869-884, 2018.
[74] G. Mustafa *et al.*, "Hyperspectral reflectance proxies to diagnose in-field fusarium head blight in wheat with machine learning," *Remote Sensing,* vol. 14, no. 12, p. 2784, 2022.
[75] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science,* vol. 349, no. 6245, pp. 255-260, 2015.
[76] S. F. C. Soares, A. A. Gomes, M. C. U. Araujo, A. R. Galvão Filho, and R. K. H. Galvão, "The successive projections algorithm," *TrAC Trends in Analytical Chemistry,* vol. 42, pp. 84-98, 2013.
[77] N. Keen, "Color moments," *School of informatics, University of Edinburgh,* pp. 3-6, 2005.
[78] B. Sebastian V, A. Unnikrishnan, and K. Balakrishnan, "Gray level co-occurrence matrices: generalisation and some new features," *arXiv preprint arXiv:1205.4831,* 2012.
[79] D. Wang, D. Tan, and L. Liu, "Particle swarm optimization algorithm: an overview," *Soft computing,* vol. 22, no. 2, pp. 387-408, 2018.
[80] S. Mahesh, D. Jayas, J. Paliwal, and N. White, "Hyperspectral imaging to classify and monitor quality of agricultural materials," *Journal of Stored Products Research,* vol. 61, pp. 17-26, 2015.
[81] J. Kotwal, R. Kashyap, and S. Pathan, "Agricultural plant diseases identification: From traditional approach to deep learning," *Materials Today: Proceedings,* vol. 80, pp. 344-356, 2023.

[82] R. U. Khan, K. Khan, W. Albattah, and A. M. Qamar, "Image‐Based Detection of Plant Diseases: From Classical Machine Learning to Deep Learning Journey," *Wireless Communications and Mobile Computing,* vol. 2021, no. 1, p. 5541859, 2021.
[83] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE,* vol. 86, no. 11, pp. 2278-2324, 1998.
[84] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.
[85] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556,* 2014.
[86] S. A. Wagle, "Comparison of Plant Leaf Classification Using Modified AlexNet and Support Vector Machine," *Traitement du Signal,* vol. 38, no. 1, 2021.
[87] J. Ye, Z. Yu, Y. Wang, D. Lu, and H. Zhou, "WheatLFANet: in-field detection and counting of wheat heads with high-real-time global regression network," *Plant Methods,* vol. 19, no. 1, p. 103, 2023.
[88] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440-1448.
[89] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems,* vol. 28, 2015.
[90] L. Li *et al.*, "Development of image-based wheat spike counter through a Faster R-CNN algorithm and application for genetic studies," *The Crop Journal,* vol. 10, no. 5, pp. 1303-1311, 2022.
[91] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *J. Mach. Learn. Res.,* vol. 13, no. 1, pp. 281–305, 2012.
[92] J. Wei, D. Xia, H. Xie, C.-M. Chang, C. Li, and X. Yang, "SpaceEditing: A Latent Space Editing Interface for Integrating Human Knowledge into Deep Neural Networks," presented at the Proceedings of the 29th International Conference on Intelligent User Interfaces, Greenville, SC, USA, 2024. [Online]. Available: https://doi.org/10.1145/3640543.3645211.
[93] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 801-818.
[94] J. Wang *et al.*, "Deep high-resolution representation learning for visual recognition," *IEEE transactions on pattern analysis and machine intelligence,* vol. 43, no. 10, pp. 3349-3364, 2020.
[95] S. M. Hassan, M. Jasinski, Z. Leonowicz, E. Jasinska, and A. K. Maji, "Plant disease identification using shallow convolutional neural network," *Agronomy,* vol. 11, no. 12, p. 2388, 2021.
[96] P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, "A Review of Yolo

algorithm developments," *Procedia computer science,* vol. 199, pp. 1066-1073, 2022.
[97] J. Terven, D.-M. Córdova-Esparza, and J.-A. Romero-González, "A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas," *Machine Learning and Knowledge Extraction,* vol. 5, no. 4, pp. 1680-1716, 2023.
[98] P. K. Sekharamantry, F. Melgani, and J. Malacarne, "Deep learning-based apple detection with attention module and improved loss function in YOLO," *Remote Sensing,* vol. 15, no. 6, p. 1516, 2023.
[99] 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.
[100] C. Wang and C. Zhong, "Adaptive feature pyramid networks for object detection," *IEEE Access,* vol. 9, pp. 107024-107032, 2021.
[101] Z. Du and Y. Liang, "Object Detection of Remote Sensing Image Based on Multi-Scale Feature Fusion and Attention Mechanism," *IEEE Access,* 2024.
[102] D.-Y. Zhang *et al.*, "Enhancing wheat Fusarium head blight detection using rotation Yolo wheat detection network and simple spatial attention network," *Computers and Electronics in Agriculture,* vol. 211, p. 107968, 2023.
[103] J. Deng, S. Bei, S. Shaojing, and Z. Zhen, "Feature fusion methods in deep-learning generic object detection: A survey," in *2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, 2020, vol. 9: IEEE, pp. 431-437.
[104] X. Li *et al.*, "Semantic flow for fast and accurate scene parsing," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16*, 2020: Springer, pp. 775-793.
[105] L. Du, R. Zhang, and X. Wang, "Overview of two-stage object detection algorithms," in *Journal of Physics: Conference Series*, 2020, vol. 1544, no. 1: IOP Publishing, p. 012033.
[106] Y. Gong, X. Yu, Y. Ding, X. Peng, J. Zhao, and Z. Han, "Effective fusion factor in FPN for tiny object detection," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2021, pp. 1160-1168.
[107] C. Guo, B. Fan, Q. Zhang, S. Xiang, and C. Pan, "Augfpn: Improving multi-scale feature learning for object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 12595-12604.
[108] N. Zeng, P. Wu, Z. Wang, H. Li, W. Liu, and X. Liu, "A small-sized object detection oriented multi-scale feature fusion approach with application to defect detection," *IEEE Transactions on Instrumentation and Measurement,* vol. 71, pp. 1-14, 2022.
[109] W. Guan, Y. Zou, and X. Zhou, "Multi-scale object detection with feature fusion and region objectness network," in *2018 IEEE*

International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018: IEEE, pp. 2596-2600.
[110] L. Jiang *et al.*, "MFFSODNet: Multi-Scale Feature Fusion Small Object Detection Network for UAV Aerial Images," *IEEE Transactions on Instrumentation and Measurement,* 2024.
[111] J. Wu, G. Dai, W. Zhou, X. Zhu, and Z. Wang, "Multi-scale feature fusion with attention mechanism for crowded road object detection," *Journal of Real-Time Image Processing,* vol. 21, no. 2, p. 29, 2024.
[112] G. Cheng, Y. Si, H. Hong, X. Yao, and L. Guo, "Cross-scale feature fusion for object detection in optical remote sensing images," *IEEE Geoscience and Remote Sensing Letters,* vol. 18, no. 3, pp. 431-435, 2020.
[113] C. Liu, S. Zhang, M. Hu, and Q. Song, "Object Detection in Remote Sensing Images Based on Adaptive Multi-Scale Feature Fusion Method," *Remote Sensing,* vol. 16, no. 5, p. 907, 2024.
[114] F. Yang, J. Zhou, Y. Chen, J. Liao, and M. Yang, "MSF-YOLO: A multi-scale features fusion-based method for small object detection," *Multimedia Tools and Applications,* pp. 1-22, 2024.
[115] S. Bouraya and A. Belangour, "Deep learning based neck models for object detection: a review and a benchmarking study," *International Journal of Advanced Computer Science and Applications,* vol. 12, no. 11, 2021.
[116] M. G. Ragab *et al.*, "A Comprehensive Systematic Review of YOLO for Medical Object Detection (2018 to 2023)," *IEEE Access,* 2024.
[117] T. Diwan, G. Anirudh, and J. V. Tembhurne, "Object detection using YOLO: Challenges, architectural successors, datasets and applications," *multimedia Tools and Applications,* vol. 82, no. 6, pp. 9243-9275, 2023.
[118] A. Vijayakumar and S. Vairavasundaram, "Yolo-based object detection models: A review and its applications," *Multimedia Tools and Applications,* pp. 1-40, 2024.
[119] H. Upulie and L. Kuganandamurthy, "Real-Time Object Detection using YOLO: A review," *Real-Time Object Detection using YOLO: A review,* 2021.
[120] A. Horzyk and E. Ergün, "YOLOv3 precision improvement by the weighted centers of confidence selection," in *2020 International Joint Conference on Neural Networks (IJCNN)*, 2020: IEEE, pp. 1-8.
[121] J. Sang *et al.*, "An improved YOLOv2 for vehicle detection," *Sensors,* vol. 18, no. 12, p. 4272, 2018.
[122] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767,* 2018.
[123] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934,* 2020.
[124] G. Jocher *et al.*, "ultralytics/yolov5: v6. 2-yolov5 classification models, apple m1, reproducibility, clearml and deci. ai integrations," *Zenodo,*

2022.
[125] C. Li *et al.*, "YOLOv6: A single-stage object detection framework for industrial applications," *arXiv preprint arXiv:2209.02976,* 2022.
[126] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 7464-7475.
[127] A. Wang *et al.*, "Yolov10: Real-time end-to-end object detection," *arXiv preprint arXiv:2405.14458,* 2024.
[128] A. Morbekar, A. Parihar, and R. Jadhav, "Crop disease detection using YOLO," in *2020 International Conference for Emerging Technology (INCET)*, 2020: IEEE, pp. 1-5.
[129] Y. S. Park, L. Konge, and A. R. Artino Jr, "The positivism paradigm of research," *Academic medicine,* vol. 95, no. 5, pp. 690-694, 2020.
[130] M. Shoaib *et al.*, "An advanced deep learning models-based plant disease detection: A review of recent research," *Frontiers in Plant Science,* vol. 14, p. 1158933, 2023.
[131] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A review," *Neurocomputing,* vol. 187, pp. 27-48, 2016.
[132] A. Etienne, A. Ahmad, V. Aggarwal, and D. Saraswat, "Deep learning-based object detection system for identifying weeds using uas imagery," *Remote Sensing,* vol. 13, no. 24, p. 5182, 2021.
[133] A. Rahman, Y. Lu, and H. Wang, "Performance evaluation of deep learning object detectors for weed detection for cotton," *Smart Agricultural Technology,* vol. 3, p. 100126, 2023.
[134] B. N. Naik, M. Ramanathan, and P. Ponnusamy, "Refined single-stage object detection deep-learning technique for chilli leaf disease detection," *Journal of Electronic Imaging,* vol. 32, no. 3, pp. 033039-033039, 2023.
[135] A. Mesterhazy, "Updating the breeding philosophy of wheat to Fusarium head blight (FHB): Resistance components, QTL identification, and phenotyping—A review," *Plants,* vol. 9, no. 12, p. 1702, 2020.
[136] R. P. T. Sundaramurthy, Y. Balasubramanian, and M. Annamalai, "Real-time detection of Fusarium infection in moving corn grains using YOLOv5 object detection algorithm," 2023.
[137] D. Zhang, Q. Wang, F. Lin, X. Yin, C. Gu, and H. Qiao, "Development and evaluation of a new spectral disease index to detect wheat fusarium head blight using hyperspectral imaging," *Sensors,* vol. 20, no. 8, p. 2260, 2020.
[138] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE transactions on pattern analysis and machine intelligence,* vol. 37, no. 9, pp. 1904-1916, 2015.
[139] A. G. Howard *et al.*, "Mobilenets: Efficient convolutional neural

networks for mobile vision applications," *arXiv preprint arXiv:1704.04861,* 2017.
[140] X. Ding, X. Zhang, N. Ma, J. Han, G. Ding, and J. Sun, "Repvgg: Making vgg-style convnets great again," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 13733-13742.
[141] J. Chen *et al.*, "Run, don't walk: chasing higher FLOPS for faster neural networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 12021-12031.
[142] K. Loague and R. E. Green, "Statistical and graphical methods for evaluating solute transport models: overview and application," *Journal of contaminant hydrology,* vol. 7, no. 1-2, pp. 51-73, 1991.
[143] L. Shi *et al.*, "YOLOv5s-T: A lightweight small object detection method for wheat spikelet counting," *Agriculture,* vol. 13, no. 4, p. 872, 2023.
[144] X. Xu *et al.*, "Prediction of wheat grain protein by coupling multisource remote sensing imagery and ECMWF data," *Remote Sensing,* vol. 12, no. 8, p. 1349, 2020.
[145] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The Pascal Visual Object Classes (VOC) challenge," *International journal of computer vision,* vol. 88, pp. 303-338, 2010.
[146] G. Jocher *et al.*, "ultralytics/yolov5: v3. 1-bug fixes and performance improvements," *Zenodo,* 2020.
[147] P. Kaur, B. S. Khehra, and E. B. S. Mavi, "Data augmentation for object detection: A review," in *2021 IEEE International Midwest Symposium on Circuits and Systems (MWSCAS)*, 2021: IEEE, pp. 537-543.
[148] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "Yolox: Exceeding yolo series in 2021," *arXiv preprint arXiv:2107.08430,* 2021.
[149] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics YOLO (Version 8.0. 0)[Computer software]," *URL: https://github. com/ultralytics/ultralytics,* 2023.
[150] Y. Zhao *et al.*, "Detrs beat yolos on real-time object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 16965-16974.
[151] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance-IoU loss: Faster and better learning for bounding box regression," in *Proceedings of the AAAI conference on artificial intelligence*, 2020, vol. 34, no. 07, pp. 12993-13000.
[152] Y.-F. Zhang, W. Ren, Z. Zhang, Z. Jia, L. Wang, and T. Tan, "Focal and efficient IOU loss for accurate bounding box regression," *Neurocomputing,* vol. 506, pp. 146-157, 2022.
[153] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 658-666.

[154] Z. Gevorgyan, "SIoU loss: More powerful learning for bounding box regression," *arXiv preprint arXiv:2205.12740,* 2022.
[155] M. B. Muhammad and M. Yeasin, "Eigen-cam: Class activation map using principal components," in *2020 International Joint Conference on Neural Networks (IJCNN)*, 2020: IEEE, pp. 1-7.
[156] J. T. Leksut, J. Zhao, and L. Itti, "Learning visual variation for object recognition," *Image and Vision Computing,* vol. 98, p. 103912, 2020.
[157] O. Mzoughi, "Automated System for Comprehensive Plant Disease Analysis," in *Applications of Computer Vision and Drone Technology in Agriculture 4.0*: Springer, 2024, pp. 49-64.
[158] R. Sapkota *et al.*, "Comprehensive Performance Evaluation of YOLOv10, YOLOv9 and YOLOv8 on Detecting and Counting Fruitlet in Complex Orchard Environments," *arXiv preprint arXiv:2407.12040,* 2024.
[159] Z. Zhao, S. Chen, Y. Ge, P. Yang, Y. Wang, and Y. Song, "RT-DETR-Tomato: Tomato Target Detection Algorithm Based on Improved RT-DETR for Agricultural Safety Production," *Applied Sciences,* vol. 14, no. 14, p. 6287, 2024.
[160] J. Chen, H. Liu, Y. Zhang, D. Zhang, H. Ouyang, and X. Chen, "A multiscale lightweight and efficient model based on YOLOv7: Applied to citrus orchard," *Plants,* vol. 11, no. 23, p. 3260, 2022.

PROJECT MANAGEMENT

Effective project management was instrumental in guaranteeing the successful conclusion of this research endeavor. In order to enhance the management and monitoring of the project's advancement, I employed a logbook, as shown in **Appendices: Appendix B - LOGBOOK**, to document significant concepts, encountered challenges, and requisite activities to be accomplished at each phase. Additionally, the logbook was a crucial element in the research plan's systematic progression, in addition to serving as a documentation instrument.

Initially, a detailed Gantt chart, as shown in **Figure 9**, was created to provide a roadmap for the project, outlining all planned phases from background reading to the final submission. As the project progressed, it became evident that deviations from the original plan were necessary due to unforeseen challenges and the iterative nature of research. However, each phase of the project was carried out methodically and in alignment with the supervision schedule under the guidance of my supervisor. Consequently, as illustrated in **Figure 10**, a second Gantt chart was developed to reflect the actual path taken during the project. This section discusses the key differences between the planned and actual timelines, the challenges encountered, the risk management strategies employed, the changes made to the project objectives, and justifications for these changes.

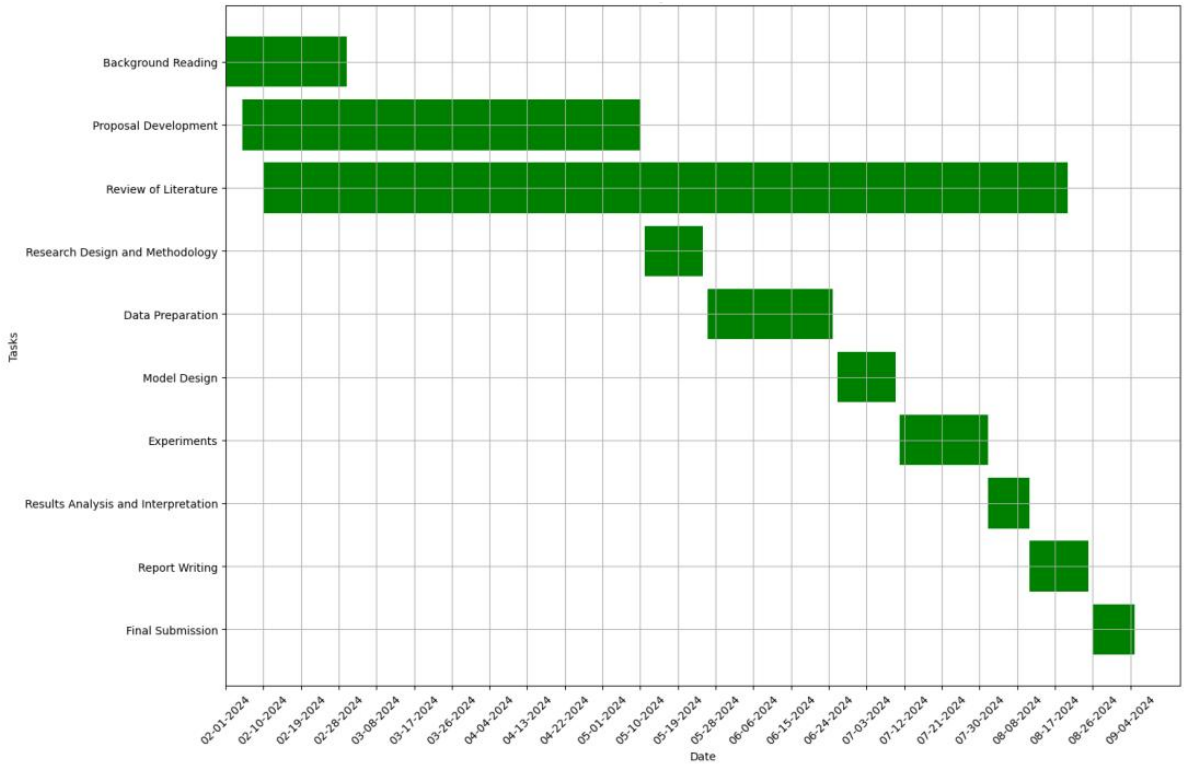

Figure 9: Original Gantt Chart

Comparing the original Gantt chart with the actual Gantt chart reveals several significant differences. The literature review phase, originally scheduled to be completed earlier, was prolonged because additional relevant research was discovered that needed to be included in the project. In order to achieve a more refined and focused approach, it was necessary to revise the research problem and objectives. Nevertheless, the modifications were undertaken following the supervisor's guidance, guaranteeing that the investigation remained on course. Furthermore, the timeline had a significant overlap between the data preparation and model design phases. This overlap was not anticipated in the original plan; however, it became necessary since the data significantly affected the model architecture, necessitating changes to be performed concurrently with data preparation.

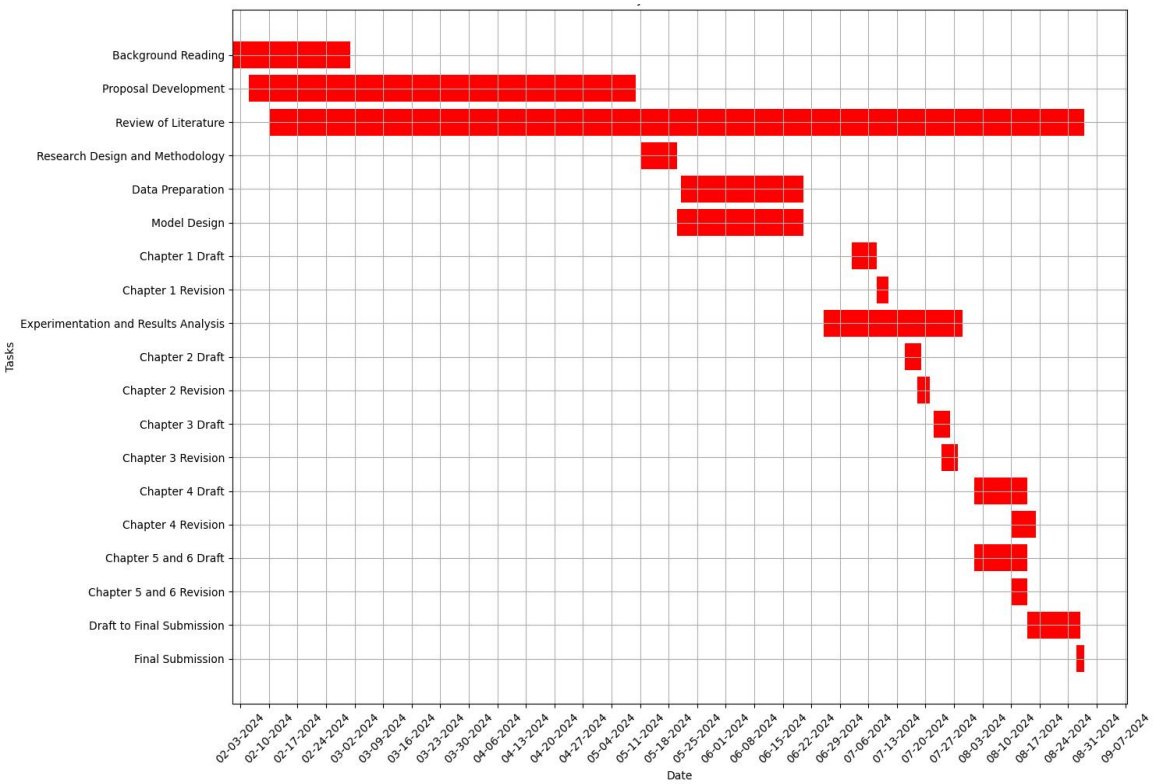

Figure 10: Actual Gantt Chart

An additional significant distinction was the extension of the experimentation and results analysis phase, which exceeded the original timeline. This update was necessary due to unforeseen difficulty in comparing model performances, mainly when dealing with small-scale features in complex backgrounds. These difficulties required several iterations of model development and additional exploration, resulting in the delay of the completion of this phase until late July. The supervision sessions provided critical feedback that guided the necessary adjustments throughout these iterations. Furthermore, the report writing phases, particularly Chapters 1, 2, and 3, were more iterative and overlapped than anticipated. The initial timetable for drafting and revising these chapters was pushed back because of revisions and more research needed in response to supervisor feedback. However, the regular meetings with the supervisor guaranteed that these phases were completed promptly despite the delays.

The primary reason for the deviations from the original plan was the numerous obstacles encountered during the undertaking. During the testing phase, dataset generation became a significant difficulty, with noise in the dataset necessitating

substantial manual selection, which delayed the start of model training and required timeline changes. Further, the model's initial iterations performed differently than anticipated, particularly in detecting small-scale features in complex backgrounds. It led to several rounds of model reworking and more testing, which made the testing phase last longer. Technical issues, like the model's higher-than-expected computational needs, also caused delays because the machine resources had to be optimized to work best with what was available.

A comprehensive risk analysis and minimization strategy was implemented throughout the project to mitigate these risks. Hand selection, double-checking, and labeling under experts' supervision helped reduce problems with dataset construction. Furthermore, the Gantt chart included buffer times for essential tasks, and subsequent tasks were modified according to the actual progress, assuring the project's adherence to the initial timeline despite the difficulties encountered.

In order to accomplish the desired model performance, the experimentation phase was also extended through buffer time to accommodate additional model refinements and testing. These changes were made after careful analysis and discussion with the supervisor, ensuring the research objectives were consistent with the research aim.

Regular self-assessment and project evaluation are also indispensable components of risk management. By reflecting on and reviewing the progress and quality of the study, I can guarantee that the project stays on pace to meet its objectives and make timely modifications to any deviations. Another critical step is to report progress to the supervisor regularly. I present the most recent experimental results in detail and discuss the specific plans for the subsequent phase with the supervisor. Based on the feedback, I made the necessary modifications to improve the quality of the research. My supervisor gave me consistent advice and feedback throughout the project, ensuring that each part was finished quickly, even when things did not go as planned.

APPENDICES

APPENDIX A - ETHICS FORM

ETHICS FORM – STEM MSc STUDENTS ONLY

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:	Ze Wu
2	Student Number:	2304721
3	Email address:	2304721@student.uwtsd.ac.uk
4	Programme of Study:	MSc Software Engineering and Artificial Intelligence
5	Director of Studies/Supervisor:	Dr. Seena Joseph

SECTION B: Internal and External Ethical Guidance Materials

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

SECTION C: Details of Research Activity

1	Indicative title:	Detection of Fusarium Head Blight on Wheat Spikelets via A Multi-Scale Feature Fusion CNN model		
2	Proposed start date:	2024.02.19	Proposed end date:	2024.05.10
	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:** • *Purpose of Research Activity* • *Proposed Research Question* • *Aims of Research Activity* • *Objectives of Research Activity* Demonstrate, briefly, how **<u>Existing Research</u>** has informed the proposed activity and explain			

	• *What the research activity will add to the body of knowledge* • *How it addresses an area of importance.*
3	**Purpose of Research Activity** The object detection algorithms for wheat FHB contribute to enhanced efficiency in agricultural production practices. Farmers and technicians can obtain real-time disease information through mobile devices, facilitating instantaneous data analysis and decision support that optimizes crop management and protection measures, thus improving agricultural production efficiency and resource utilization. In agricultural science research, the rapid and convenient extraction of disease phenotypes aids in improving the efficiency of breeding for scab-resistant varieties. (this box should expand as you type)
4	**Research Question** Compared to disease detection at the scale of wheat spikes, the noise interference of complex field backgrounds and overlapping of spikes can lead to the detection network losing substantial spatial scale and morphological texture semantic information when extracting and transmitting fine-grained spikelet features, making high-precision detection of spikelet FHB challenging. (this box should expand as you type)
5	**Aims of Research Activity** This study aims to develop an object detection network based on multi-scale feature fusion, extracting and enhancing spatial texture and morphological features at different scales of spikelets, capturing subtle differences in disease characteristics caused by varying degrees of spikelet infection, thereby achieving efficient and rapid detection of wheat FHB and phenotypic extraction. (this box should expand as you type)
6	**Objectives of Research Activity** The main objectives of this study are as follows, (i) to develop a multi-scale feature fusion algorithm that can effectively enhance spikelet detection accuracy, (ii) to design an object detection model based on multi-scale feature fusion to achieve high-precision spikelet FHB detection, and (iii) to test the algorithm and validate the model' s accuracy in detecting wheat FHB and its effectiveness in phenotypic extraction. (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. *(this box should expand as you type)*			

SECTION F: Anticipated Risks

	Outline any anticipated risks that may adversely affect any of the participants, the researchers and/or the University, and the steps that will be taken to address them.	
1	**Risks to participants** For example: sector-specific health & safety, emotional distress, financial disclosure, physical harm, transfer of personal data, sensitive organisational information. If you have identified in section D that there are no participants then enter N/A and go skip to question 3.	
	Risk to participants: N/A *(this box should expand as you type)*	*How you will mitigate the risk to participants:* *(this box should expand as you type)*
2	If research activity may include sensitive, embarrassing or upsetting topics (e.g. sexual activity, drug use) or issues likely to disclose information requiring further action (e.g. criminal activity), give details of the procedures to deal with these issues, including any support/advice (e.g. helpline numbers) to be offered to participants. Note that where applicable, consent procedures should make it clear that if something potentially or actually illegal is discovered in the course of a project, it may need to be disclosed to the proper authorities	
	(this box should expand as you type)	
3	**Risks to the investigator** For example: personal health & safety, physical harm, emotional distress, risk of accusation of harm/impropriety, conflict of interest	
	Risk to the investigator: N/A *(this box should expand as you type)*	*How you will mitigate the risk to the investigator:* *(this box should expand as you type)*
4	**University/institutional risks** For example: adverse publicity, financial loss, data protection	
	Risk to the University: N/A *(this box should expand as you type)*	*How you will mitigate the risk to the University:* *(this box should expand as you type)*
5	**Environmental risks** For example: accidental spillage of pollutants, damage to local ecosystems	
	Risk to the environment: N/A *(this box should expand as you type)*	*How you will mitigate the risk to environment:* *(this box should expand as you type)*

SECTION G: Feedback, Consent and Confidentiality

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

1	**Feedback** What de-briefing and feedback will be provided to participants, how will this be done and when?
	You don't need to email your participants with your final report. A good alternative is to set up an email address that they will be able to contact for further details or results. *(this box should expand as you type)*
2	**Informed consent**

	Describe the arrangements to inform potential participants, before providing consent, of what is involved in participating. Describe the arrangements for participants to provide full consent before data collection begins. If gaining consent in this way is inappropriate, explain how consent will be obtained and recorded in accordance with prevailing data protection legislation.
	If you are using a paper questionnaire then you should have the participants sign an appropriate consent form. These forms will count as personal data and should be noted as such in section J. If you are using an online questionnaire, then you should have a screen before the questions start that acts as a consent form, informing participants that by clicking on the NEXT button they are providing consent. *(this box should expand as you type)*
3	**Confidentiality / Anonymity** Set out how anonymity of participants and confidentiality will be ensured in any outputs. If anonymity is not being offered, explain why this is the case.
	Do not collect names unless you really need them. Do not name participants or organisations in any research publications (including the thesis) without their explicit permission. *(this box should expand as you type)*

SECTION H: Data Protection and Storage

	Does the research activity involve personal data (as defined by the General Data Protection Regulation 2016 "GDPR" and the Data Protection Act 2018 "DPA")?	**YES**	**NO**
1	***"Personal data"*** *means any information relating to an identified or identifiable natural person ('data subject'). An identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person. Any video or audio recordings of participants is considered to be personal data.*	☐	☒
	If YES, provide a description of the data and explain why this data needs to be collected:		
2	This includes audio/video data of participants, but can also include IP addresses and usernames. Names, addresses and emails also count, as do consent forms. *(this box should expand as you type)*		
	Does it involve special category data (as defined by the GDPR)?	**YES**	**NO**
3	***"Special category data"*** *means sensitive personal data consisting of information as to the data subjects' –* *(a) racial or ethnic origin,* *(b) political opinions,* *(c) religious beliefs or other beliefs of a similar nature,* *(d) membership of a trade union (within the meaning of the Trade Union and Labour Relations (Consolidation) Act 1992),* *(e) physical or mental health or condition,* *(f) sexual life,* *(g) genetics,* *(h) biometric data (as used for ID purposes),*	☐	☒
	If YES, provide a description of the special category data and explain why this data needs to be collected:		
4	What counts as 'sensitive' will differ between cultures. Any information on behaviour that is not in accordance with cultural norms would count as sensitive personal data. *(this box should expand as you type)*		

	Will data from the research activity (collected data, drafts of the thesis, or materials for publication) be stored in any of the following ways?	YES	NO
5	Manual files (i.e. in paper form)?	☐	☒
6	University computers?	☒	☐
7	Private company computers?	☐	☒
8	Home or other personal computers?	☐	☒
9	Laptop computers/ CDs/ Portable disk-drives/ memory sticks?	☐	☒
10	"Cloud" storage or websites?	☐	☒
11	Other – specify:	☐	☒
12	For all stored data, explain the measures in place to ensure the security of the data collected, data confidentiality, including details of backup procedures, password protection, encryption, anonymisation and pseudonymisation:		
	All data will be stored securely in the OneDrive cloud storage service provided by the University and only I will have access to it. The data will not be shared with anyone else in order to protect the privacy and security of the data. In addition, in order to protect the data from being lost, regular backups of the data will be made to ensure that the data is well preserved. To further enhance data security, all data stored on OneDrive will be encrypted using the Advanced Encryption Standard (AES) to ensure that the data is secure during transmission and storage. This means that even if data is intercepted in transit, the content cannot be deciphered without the correct key. In this way, we can maximize the protection of data from unauthorized access and potential security threats. *(this box should expand as you type)*		

	Data Protection		
	Will the research activity involve any of the following activities:	YES	NO
13	Electronic transfer of data in any form?	☐	☒
14	Sharing of data with others at the University outside of the immediate research team?	☐	☒
15	Sharing of data with other organisations?	☐	☒
16	Export of data outside the UK or importing of data from outside the UK?	☐	☒
17	Use of personal addresses, postcodes, faxes, emails or telephone numbers?	☐	☒
18	Publication of data that might allow identification of individuals?	☐	☒
19	If YES to any question, please provide full details, explaining how this will be conducted in accordance with the GDPR and Data Protection Act (2018) (and/or any international equivalent):		
	This includes data such as drafts of your thesis as well as experimental or survey data. An example of suitable text is given below. *All data will be encrypted and kept in password protected cloud storage on the University Office 365 system which will not be shared. Any USB sticks used to store or transfer data will be password protected. All data transfers will be encrypted and password protected. All participants will be given a unique identifier to ensure confidentiality and this list will be kept securely in the password protected folder. The data will be stored until the completion of the project and then deleted. In accordance with the DPA2018, participants will have the right to ask to see what data is held relating to them, and this data will be deleted immediately if the participant requests this, in which case the data will not be used in the project.* *(this box should expand as you type)*		
20	List all who will have access to the data generated by the research activity:		

	Typically, only the principal investigator and their supervisor will have access to data generated by research activities to ensure research confidentiality and data security. *(this box should expand as you type)*	
21	List who will have control of, and act as custodian(s) for, data generated by the research activity:	
	The Principal Investigator will act as the controller and custodian of the data during research activities and will be responsible for ensuring the security, accuracy and compliance of the data. *(this box should expand as you type)*	
22	Give details of data storage arrangements, including security measures in place to protect the data, where data will be stored, how long for, and in what form.	
	In order to ensure the security of the research data, all data will be encrypted using advanced encryption techniques and stored on the password protected University Office 365 cloud storage system and this data will not be shared with any unauthorized individuals or groups. Any USB drives used during the project, whether for storage or data transfer, will be password protected to prevent unauthorized access. Upon completion of the project, these flash drives will be reformatted to completely destroy all data stored on them. The data will be retained until the completion of the project, after which it will be securely deleted to comply with the relevant policies on data retention and destruction. *(this box should expand as you type)*	
22	Confirm that you have read the UWTSD guidance on data management (see https://www.uwtsd.ac.uk/library/research-data-management/)	☒
23	Confirm that you are aware that you need to keep all data until after your research has completed or the end of your funding	☒

SECTION I: Declaration

	The information which I have provided is correct and complete to the best of my knowledge. I have attempted to identify any risks and issues related to the research activity and acknowledge my obligations and the rights of the participants. In submitting this application I hereby confirm that I undertake to ensure that the above named research activity will meet the University's Research Ethics and Integrity Code of Practice which is published on the website: https://www.uwtsd.ac.uk/research/research-ethics/		
1	**Signature of applicant:**	Ze Wu	**Date:2024.05.10**
2	Director of Studies/Supervisor:		**Date: 2024.05.10**
3	Signature:	Ze Wu	

FOR INTERNAL USE ONLY:

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

APPENDIX B - LOGBOOK

Date	Daily Activities	Thought Trails	Things to Do
2024-02-01	• Started background reading on Fusarium Head Blight (FHB).	• Understanding FHB's impact on wheat crops, focusing on disease symptoms and current detection methods. • Considering how these methods can be enhanced using modern techniques.	• Identify key literature sources for in-depth study. • Organize notes on the disease symptoms and current detection technologies.
2024-02-03	• Continued exploring FHB-related studies, focusing on the impact of environmental conditions on FHB spread.	• Noted that environmental factors greatly influence FHB symptoms; this might require incorporating these variables into the model design.	• Look for studies that explore the impact of environmental conditions on plant diseases.
2024-02-05	• Began Proposal Development. • Initial discussions about the feasibility of the project.	• Considering research aim, objectives, and the feasibility of using CNN for FHB detection. • Initial thoughts on how YOLO might be adapted for this specific problem.	• Draft the research proposal outline. • Develop a preliminary timeline for the project phases.
2024-02-07	• Further refined the proposal, focusing on the objectives and expected outcomes.	• Exploring different model architectures, particularly lightweight ones that can run in real-time scenarios.	• Expand the proposal to include a detailed literature review section.
2024-02-10	• Continued Review of Literature, focusing on FHB detection technologies and YOLO-based object detection models.	• Identified gaps in current detection methods, especially the need for more accurate and efficient phenotypic data collection at the spikelet level. • Consideration of alternative approaches to handle small object detection.	• Collect more papers on deep learning applications in plant disease detection, particularly on multi-scale feature fusion and YOLO models.
2024-02-12	• Started outlining potential modifications to existing YOLO models to better suit FHB detection.	• The complexity of FHB detection at the spikelet level might require custom feature fusion layers.	• Draft a section of the proposal on the proposed model modifications.
2024-02-15	• Detailed proposal drafted, focusing on using YOLO-based CNN for FHB detection.	• Exploring the feasibility of a lightweight model to ensure real-time processing in agricultural settings.	• Prepare a presentation for the proposal presentation. Plan for further deep dives into model architecture studies.

2024-03-01	• Completed background reading and final adjustments to the proposal.	• Finalizing the scope and methodology. • Ensuring the proposal aligns with the project's goals and available resources.	• Submit the proposal and start preparing for literature review consolidation. Review the proposal with peers for additional feedback.
2024-03-10	• Deepened literature review with a focus on multi-scale feature fusion techniques and their application in small object detection.	• Identified potential improvements to existing YOLO models for small-scale object detection, particularly in agricultural settings.	• Start drafting literature review sections in the report. Create a summary of the key papers and how they influence your approach.
2024-03-20	• Conducted a review session to refine the research methodology based on literature findings.	• The integration of feature fusion in YOLO requires careful consideration of computational efficiency vs. accuracy.	• Finalize the literature review draft and start planning for initial experiments.
2024-05-11	• Started research design and methodology, focusing on data collection and preprocessing methods.	• Designed initial experiments and selected datasets for model training. • Deciding on the best dataset sources.	• Set up the experiment environment and organize data collection.
2024-05-21	• Began data preparation and continued model design, implementing preliminary feature fusion layers.	• Experimenting with various CNN architectures and their potential effectiveness in detecting small-scale FHB features. • Considering custom layers to enhance feature extraction.	• Begin pre-processing data for model training. • Fine-tune the model architecture based on initial findings.
2024-06-01	• Completed data preparation, ready for initial model training.	• The quality of the preprocessed data will significantly impact model performance. • Early observations suggest that further data augmentation may be necessary.	• Run initial training cycles to test model performance. • Monitor and log training progress carefully.
2024-06-10	• Finalized model design and began initial experimentation with the preprocessed dataset.	• Considering various CNN architectures and their potential effectiveness in detecting small-scale FHB features. • Evaluating the impact of different feature fusion techniques.	• Run preliminary experiments to evaluate model performance. • Adjust learning rates and other hyperparameters to optimize the training process.
2024-06-15	• Analyzed early experimental results and discussed them.	• Early results suggest promising accuracy but require more data for validation.	• Run experiments. • Consider incorporating feature fusion

			mechanisms into the model.
2024-06-25	• Conducted further experiments on model performance, refining the model design based on early findings.	• The model shows improvement, but overfitting is becoming a concern. • Focused on summarizing the research problem, aim, objectives, and methodology for a clear and concise presentation.	• Continue experiments and test different regularization techniques to reduce overfitting.
2024-07-01	• Completed additional experiments and started integrating findings into the report.	• The results are now more consistent, and the model's accuracy is improving. • Next, focus on optimizing the model performance.	• Begin drafting the experimentation section in the report. • Prepare for the next phase of writing, focusing on chapter 1.
2024-07-02	• Chapter 1 Drafted; early findings from experiments included.	• Drafting report sections concurrently with ongoing experiments helps maintain alignment with research objectives. • The initial draft is too complex and needs simplification for broader accessibility.	• Revise Chapter 1 based on supervisor feedback and finalize. • Simplify technical explanations where necessary.
2024-07-05	• Continued revising Chapter 1, integrating feedback from the supervisor.	• The literature review needs to be more cohesive, tying the reviewed studies closer to the proposed model.	• Finalize Chapter 1 and ensure it provides a strong foundation for the subsequent chapters.
2024-07-09	• Completed Chapter 1 revisions and started outlining Chapter 2.	• Strengthening the connection between the literature review and experimental findings is crucial for coherence.	• Move on to drafting Chapter 2. • Outline the key points to be covered in this chapter.
2024-07-12	• Gathered additional sources for Chapter 2 and began integrating them into the draft.	• Emphasizing the gaps in the literature and justifying the need for the proposed model is critical.	• Continue drafting Chapter 2 with a focus on critical analysis of the literature. • Prepare for a detailed review session with the

			supervisor.
2024-07-15	• Drafted Chapter 2 focusing on literature synthesis and review.	• Emphasizing the gaps and justifying the need for the proposed model. • Including more detailed comparisons of existing models and the proposed approach.	• Prepare a draft of Chapter 2 for review. Make sure all references are properly cited.
2024-07-19	• Revised Chapter 2 after a thorough review session.	• Addressing specific comments regarding the clarity of argumentation in the literature review. • Supervisor suggests adding a section on the limitations of current approaches.	• Finalize Chapter 2 and start drafting Chapter 3. Include a new section on limitations as suggested.
2024-07-22	• Chapter 3 Drafted; focused on research methodology and experiment setup.	• Detailed the technical aspects of the model and the experimental procedures. • Ensuring clarity in the explanation of the methodology to make it reproducible.	• Prepare for another review session with the supervisor. • Double-check the experiment setup details.
2024-07-24	• Conducted a dry run of the experiment to ensure everything is well-documented in Chapter 3.	• Documenting the steps thoroughly is crucial for reproducibility. • Potential issues were identified in the experimental setup documentation.	• Revise the experimental setup in Chapter 3 and add more detail to the methodology.
2024-07-26	• Revised Chapter 3 based on supervisor's feedback. Added more detailed explanations and diagrams where necessary.	• Significant improvements to the explanation of the methodology. • The inclusion of diagrams helps in understanding the model architecture.	• Finalize Chapter 3 and prepare to write about experiments and results.
2024-07-29	• Completed initial experiment analysis and began drafting the results section.	• The model shows strong performance, particularly in detecting small-scale features, which was a primary goal. • Focus on presenting the data clearly.	• Begin drafting Chapter 4 with a focus on presenting experiment results. • Create graphs and tables to visually represent the predictions.
2024-08-01	• Drafted Chapters 4 and 5, incorporating experiment results and discussion.	• Early results show that the proposed model outperforms state-of-the-art techniques in accuracy. • The discussion needs to tie back	• Prepare visuals and tables for the report to clearly illustrate the results.

		to the initial research questions.	
2024-08-05	• Revised draft based • on feedback before finalizing.	• Revised the draft based on supervisor's valuable feedback on the clarity of the results presentation. Some results need further explanation to avoid misinterpretation.	• Refine the results section and incorporate peer feedback. • Start writing the discussion and analysis sections in Chapter 5.
2024-08-10	• Continued work on Chapter 5 and 6, adding reflections, conclusions, and potential implications.	• Emphasizing the model's contributions and potential implications for future research. • Discussing the limitations and possible areas for improvement.	• Prepare the draft report for submission. • Ensure the conclusions are well-supported by the data.
2024-08-14	• Completed the draft report; preparing for final revisions.	• Focus on ensuring the coherence and consistency of the entire report. • Need to ensure that all sections flow logically from one to the next.	• Submit the draft for final review. • Ensure all sections have been covered thoroughly.
2024-08-28	• Final report submission.	• Reflection on the entire project process and outcomes.	• Finalize any loose ends and prepare for the viva. • Start planning for future research directions.

GLOSSARY

FHB Fusarium Head Blight: A fungal disease affecting wheat crops, leading to yield loss and mycotoxin contamination.

CNN Convolutional Neural Network: A deep learning algorithm used primarily for image recognition and analysis.

YOLO You Only Look Once: A one-stage object detection algorithm that predicts object locations and classes in a single pass through the network.

MSFEF Multi-Scale Feature Enhancement and Fusion: A module used in the proposed model to enhance detection accuracy by fusing features at multiple scales.

DSR Diseased Spikelet Rate: A measure used to assess the severity of Fusarium Head Blight by calculating the percentage of infected spikelets in a wheat spike.

SPPF Spatial Pyramid Pooling Fast: A module in the CNN architecture that helps in feature extraction by pooling features at different scales efficiently.

FPN Feature Pyramid Network: A type of network used for feature extraction that allows for multi-scale object detection.

GELAN Generalized Efficient Layer Aggregation Network: A structure that balances computational complexity and efficiency in feature aggregation.