

Article

Lightweight U-Net-Based Method for Estimating the Severity of Wheat Fusarium Head Blight

Lei Shi , Zhihao Liu, Chengkai Yang, Jingkai Lei, Qiang Wang, Fei Yin and Jian Wang *

College of Information and Management Science, Henan Agricultural University, Zhengzhou 450046, China; shilei@henau.edu.cn (L.S.); lzh1229@stu.henau.edu.cn (Z.L.); yck03185@stu.henau.edu.cn (C.Y.); ljk2000@stu.henau.edu.cn (J.L.); wqhda@sina.com (Q.W.); yin.fei@foxmail.com (F.Y.)

* Correspondence: leonw63@mail.bnu.edu.cn

Abstract: Wheat Fusarium head blight is one of the major diseases affecting the yield and quality of wheat. Accurate and rapid estimation of disease severity is crucial for implementing disease-resistant breeding and scientific management strategies. Traditional methods for estimating disease severity are complex and inefficient, often failing to provide accurate assessments under field conditions. Therefore, this paper proposes a method using a lightweight U-Net model for segmenting wheat spike disease spots to estimate disease severity. Firstly, the model employs MobileNetv3 as its backbone for feature extraction, significantly reducing the number of parameters and computational demand, thus enhancing segmentation efficiency. Secondly, the backbone network has been augmented with a lightweight Coordinate Attention (CA) module, which integrates lesion position information through channel attention and aggregates features across two spatial dimensions. This allows the model to capture long-range feature correlations and maintain positional information, effectively enhancing the segmentation of wheat spike disease spots while ensuring the model's lightweight and efficient characteristics. Lastly, depthwise separable convolutions have been introduced in the decoder in place of standard convolutions, further reducing the model's parameter count while maintaining performance. Experimental results show that the model's segmentation Mean Intersection over Union (MIOU) reached 88.87%, surpassing the U-Net model by 3.49 percentage points, with a total parameter count of only 4.52 M, one-sixth of the original model. The improved model demonstrates its capability to segment individual wheat spike disease spots under field conditions and estimate the severity of infestation, providing technical support for disease identification research.

Keywords: Fusarium head blight; segmentation; lightweight; Coordinate Attention; severity



Citation: Shi, L.; Liu, Z.; Yang, C.; Lei, J.; Wang, Q.; Yin, F.; Wang, J.

Lightweight U-Net-Based Method for Estimating the Severity of Wheat

Fusarium Head Blight. *Agriculture*

2024, *14*, 938. [https://doi.org/](https://doi.org/10.3390/agriculture14060938)

10.3390/agriculture14060938

Academic Editor: David Guest

Received: 29 April 2024

Revised: 9 June 2024

Accepted: 11 June 2024

Published: 15 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Wheat is one of the world's three major cereal crops, and its cultivation is crucial for global food security [1]. Fusarium head blight (FHB), a prevalent wheat disease, can lead to serious yield losses and pose risks to food safety [2]. The excessive use of pesticides to control FHB can harm the ecosystem. Therefore, accurate and real-time estimation of FHB severity is essential for effective disease management and loss assessment.

Visual estimation is a traditional method for quantifying disease severity. Conventional calculations of FHB severity involve manually counting the number of diseased spikelets on a single wheat spike and calculating the percentage of diseased spikelets in total spikelets to determine the severity level [3]. However, this method is inefficient, time-consuming, and susceptible to subjective bias in practical applications [4]. In contrast, severity assessment based on digital image analysis is accurate, repeatable, and employs established processing procedures. Initially, background noise is removed through image pre-processing or manual operations. Then, color conversion is combined with mathematical morphology operations and thresholding to segment the disease spots. Lastly, the proportion of the disease spot area is calculated to determine the disease severity [5].

Current crop segmentation methods primarily focus on extracting features from disease images [6,7], yielding several research outcomes. For instance, Sarayloo and Asemani [8] extracted texture, color, and shape features from wheat as effective attributes for disease identification, achieving a recognition accuracy of 98.3%. However, traditional image processing and machine learning methods often rely on destructive sampling data and relatively uniform backgrounds [9,10], which do not meet the real-time detection needs of large-scale fields.

With the rapid development of deep learning, convolutional neural network (CNN)-based image processing technologies [11–14] have been extensively applied to crop pest and disease identification and segmentation, showing improved performance [15,16]. Bao et al. [17] used image recognition technology to design a multi-path convolutional neural network that extracts features from the R, G, and B channels of wheat spikes, successfully identifying whether individual wheat spikes were infected with FHB with 100% accuracy. Shi et al. [18] used an improved YOLOv5 algorithm for target detection counting to accurately count wheat spikelets. Li et al. [19] employed image segmentation technology tailored to the characteristics of cucumber diseases, using mixed dilated convolution blocks and attention mechanisms to enhance the U-Net, achieving a segmentation accuracy of 84.97% and successfully segmenting healthy leaves and disease spots.

Image segmentation technology has been continuously refined in recent years [20–23]. However, segmenting Fusarium head blight (FHB) remains a challenging issue due to the computational intensity and complexity required by existing methods. Deng et al. [24] segmented dense overhead images of FHB, achieving an MIoU of 79.9%. Zhang et al. [25] have developed sophisticated models that, while effective, rely on two-stage processes that significantly increase computational demands and complexity in field settings. Wang et al. [26] improved accuracy with a multi-model fusion algorithm but at the cost of increased model complexity and resource requirements. Therefore, it is necessary to explore a more efficient and naturally suitable method for estimating the severity of FHB disease under field conditions.

Model lightweighting is a hot topic in the field of deep learning. Common lightweighting techniques include parameter pruning [27], sparsification, knowledge distillation [28], and the construction of lightweight network architectures. Zhu et al. [29] made lightweight improvements to the DeepLabV3 model used for visual navigation in pitaya orchards, significantly reducing computational complexity and memory consumption by adopting MobileNetV2, thereby also enhancing navigation accuracy. However, the study lacks a comparative analysis with more performant lightweight backbone networks like MobileNetV3. Wu et al. [30] employed a lightweight Coordinate Attention (CA) mechanism, effectively reducing the model's parameter count while ensuring detection accuracy. Despite achieving model lightweighting and enhancing network performance, there is a lack of comparative analysis with similar attention modules and further discussion on improvements.

Building on the experience from previous lightweight models, this study has developed a segmentation model that optimizes the balance between speed and accuracy. The model employs MobileNetv3 as its backbone within the U-Net architecture for efficient feature extraction, and includes an optimized lightweight Coordinated Attention module in the encoder to better segment disease spots. Additionally, it replaces standard convolutions in the decoder with depthwise separable convolutions, reducing the model's computational demands. Comparative experiments with traditional segmentation models and advanced lightweight backbone networks have validated the effectiveness of this model, providing technical support for the intelligent detection of wheat head blight.

2. Material and Methods

2.1. Data Acquisition

Field images of wheat spikes were collected at the Yuan Yang Science and Education Park experimental field of Henan Agricultural University (34°51' N, 113°35' E) from 26 April to 19 May 2023. The experimental field is situated in a warm temperate continental

monsoon climate, characterized by an average annual temperature of 14.4 °C, average annual rainfall of 566.2 mm, and an average frost-free period of 233 days. The soil is loamy with a deep and loosely structured profile, suitable for cultivating wheat.

Images were captured using a smartphone camera with a resolution of 3904×2928 pixels, in JPG format, and set to a 4:3 aspect ratio with macro focus to obtain focused images of individual wheat spikes in the field. A total of 700 original images were captured. To mitigate issues such as camera shake, lens blur, and changes in lighting during the shooting process, 500 high-quality images were selected as experimental data. To simulate fluctuations in lighting conditions caused by environmental factors such as cloud movement and changes throughout the day, Gaussian noise was introduced during the image preprocessing stage. This additive noise model was particularly effective in mimicking uneven lighting conditions, adding random values directly to the pixel values independently of the image content.

Referencing the rules for monitoring and forecasting wheat head blight caused by *Fusarium graminearum* Schw./*Gibberella zeae* (Schw.) Petch (GB/T 15796-2011) [31], the severity of the disease is primarily assessed by calculating the rate of infected spikelets. Given the dense arrangement and symmetrical structure of the spikelets, FHB typically spreads from the stalk, resulting in symmetrically distributed disease spots on both sides of the spike. Therefore, the disease severity can be categorized into five levels based on the ratio of diseased spot area to total spike area, as shown in Table 1. Examples of the data collected are shown in Figure 1.

Table 1. Classification of Fusarium head blight severity.

Grade	0	1	2	3	4
Percentage of Lesion Area	0	0–25%	25–50%	50–75%	75–100%

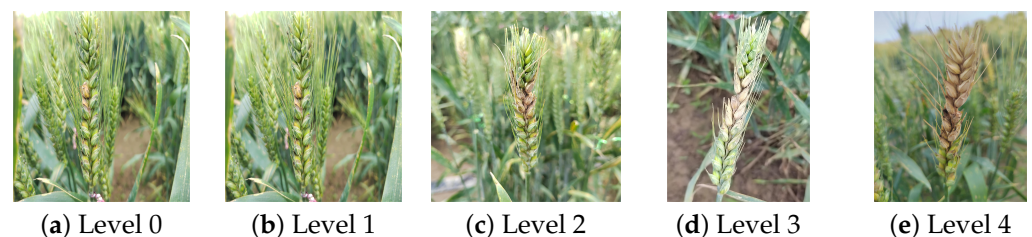


Figure 1. Samples of the dataset with levels 0 to 4.

2.2. Data Annotation and Data Augmentation

In this study, images were manually annotated using LabelMe software (version 3.16.7), employing polygon (Polygons) shapes to produce the annotations. The results were saved in JSON format, which was then converted into 8-bit grayscale images to serve as data label maps. The annotations categorized the images into two types of objects: background and wheat spikes. Blurry wheat spikes within the images were treated as background. An example of the data is shown in Figure 2. During training, the size of the images was adjusted to 512×512 pixels.

In response to the need for robust image data processing, more comprehensive affine transformations such as rotations, translations, scaling, and contrast adjustments were implemented to mimic the various positional and perspective changes as well as lighting conditions that might occur in field images. Gaussian blur was also applied to simulate the defocusing effects that could arise from camera movement or focusing issues. These enhancements, as shown in Figure 3, are crucial for training the model and improving the accuracy and robustness of wheat spike image segmentation.

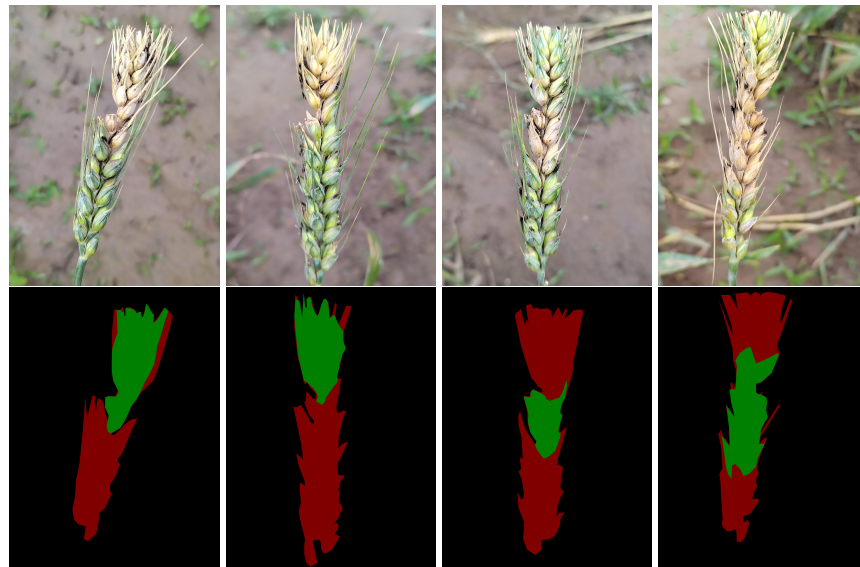


Figure 2. Visualization of wheat spikes and corresponding segmentation masks. Green represents the mask for the diseased areas, and red represents the mask for the wheat spike parts.



Figure 3. Original and enhanced images. (a) Original; (b) rotation transformation; (c) flip horizontal; (d) change contrast; (e) apply blur.

2.3. Lightweight Architecture for Wheat Spike Disease Spot Segmentation

Due to the efficiency of the U-Net architecture in image-to-image mapping tasks, this study has selected it as the basis for model construction. The skip connections between the encoder and decoder in the U-Net architecture provide precise localization information for the output images. This research has developed a lightweight U-Net model for segmenting wheat spike disease spots, as shown in Figure 4. This model utilizes MobileNetv3 as the backbone network for feature extraction, a choice that significantly reduces the model's parameter count and computational load, thereby enhancing its segmentation efficiency. To further enhance model performance, the SE attention module within MobileNetv3 has been replaced with a lightweight Coordinate Attention (CA) module. This CA module not only integrates the positional information of the segmentation target into the channel attention but also aggregates features across two spatial dimensions. This design enables the model to capture long-range feature correlations while preserving essential positional information. Additionally, we have optimized the activation function of the CA module to further improve the model's segmentation accuracy. In the decoder, the use of depthwise separable convolutions instead of standard convolutions further reduces the computational demands, enhancing real-time operational performance. These strategic enhancements not only maintain the model's lightweight design but also significantly improve its efficiency and accuracy in segmenting wheat spike disease spots, thus improving its ability to segment wheat spike disease spots efficiently.

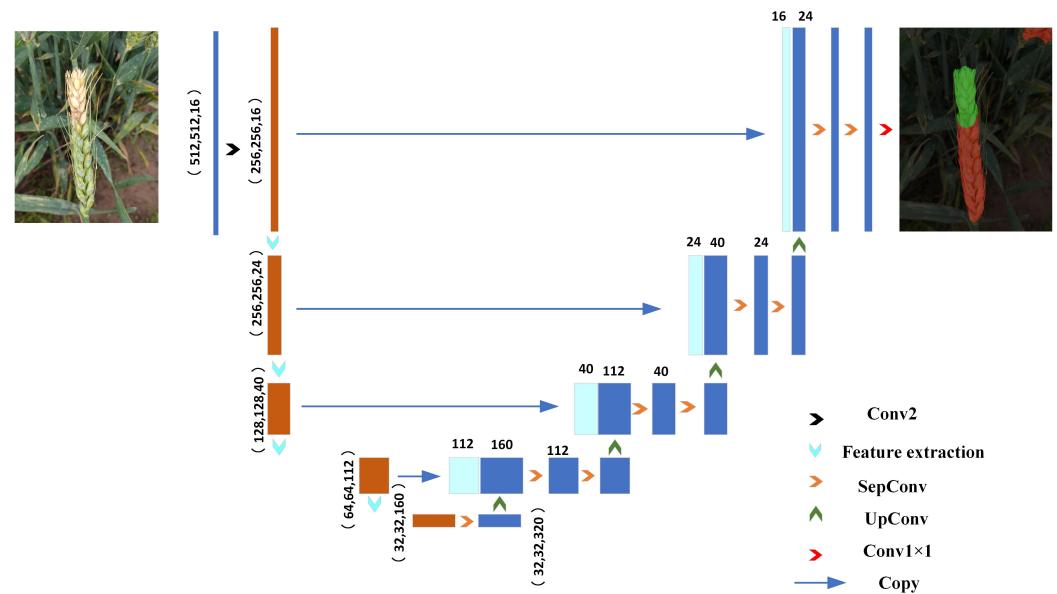


Figure 4. The architecture of U-Net for image segmentation. In the segmentation result image on the right, green represents the segmentation results for the diseased areas, and red represents the segmentation results for the wheat spike parts.

2.4. Backbone Feature Extraction Network

When the model is applied on resource-constrained platforms such as mobile devices, the large parameter size can lead to increased inference time, thereby affecting the device's response speed. To enhance computational efficiency, MobileNetv3 is employed as the backbone feature extraction network. MobileNetv3 incorporates depthwise separable convolutions from MobileNetv1 and the linear bottleneck and inverted residual structures from MobileNetv2, with modifications to these inverted residual structures, as shown in Figure 5.

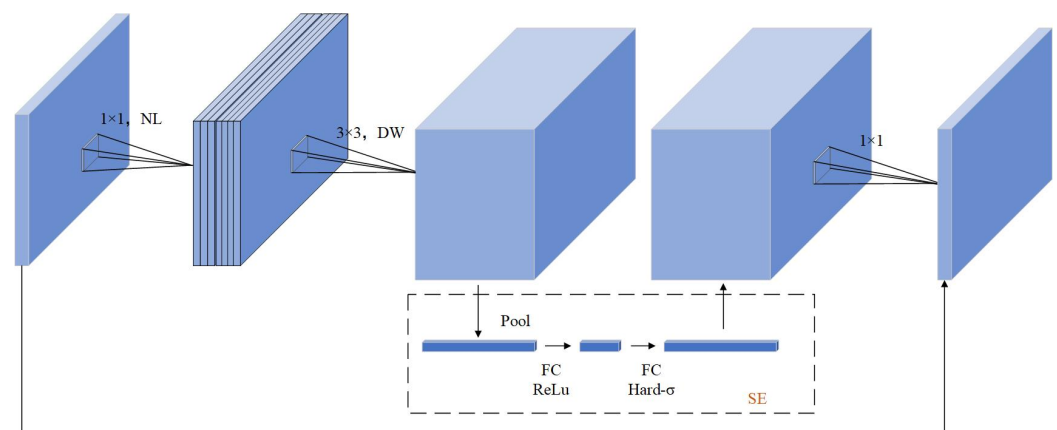


Figure 5. Inverted residual structure.

The inverted residual structure initially performs a 1×1 convolution to expand dimensionality, followed by a 3×3 depthwise convolution (DW) to extract features, and finally a 1×1 convolution to reduce dimensionality. The structure also integrates an SE attention module and replaces the Swish activation function with the h-Swish activation function to ensure accuracy while reducing computational time and enhancing speed. However, the SE module, which uses two fully connected layers to obtain feature channel weights, is characterized by a large parameter size and lacks attention to positional information of features. Consequently, this section proposes substituting the SE attention module in the inverted residual structure of Figure 5 with a CA module. This substitution aims to

differentiate the importance of various channels and segment useful information, thereby achieving precise localization and segmentation of wheat spike disease spots.

2.5. CA Attention Module

The SE (Squeeze-and-Excitation) [32] and CBAM (Convolutional Block Attention Module) [33] attention mechanisms, in their construction, overlook the importance of positional information. To address this issue, the CA mechanism was developed. The core idea of CA is to introduce spatial coordinate information on top of the channel attention mechanism to better capture the relationships between different positions within the image. CA converts the coordinate information of the input feature map into coordinate embedding vectors, which are then combined with the original feature map to form an enhanced feature representation.

The structure of the CA attention module, as shown in Figure 6, processes the input feature map, which has dimensions (C, H, W). It performs global average pooling on the feature map in both the width and height dimensions to obtain feature maps for these two directions. After average pooling in the width direction, the feature layer dimension becomes (C, H, 1), and after pooling in the height direction, it becomes (C, 1, W). These two stages are then merged by transposing the width and height to the same dimension and stacking them, resulting in a feature layer (C, 1, H + W). This layer undergoes convolution, normalization, and activation function processing to produce new features. These new features are then split into two parallel processing stages again, with the width direction feature layer changing to (C, 1, H) and the height direction feature layer to (C, 1, W). After channel adjustment through a 1×1 convolution, sigmoid functions are applied to obtain attention weights for both the width and height dimensions. These attention weights are then multiplied by the original features to produce the final output features. Through these operations, the CA attention module effectively utilizes positional information and encodes the feature map in both spatial directions, thereby enhancing the representation of wheat spike disease spot features. To optimize model performance, we have incorporated the h-Swish activation function, supplanting the conventional sigmoid function. This adaptation facilitates a more efficient and precise processing of features, significantly enhancing the model's performance by enabling smoother gradient flow and reducing computational overhead.

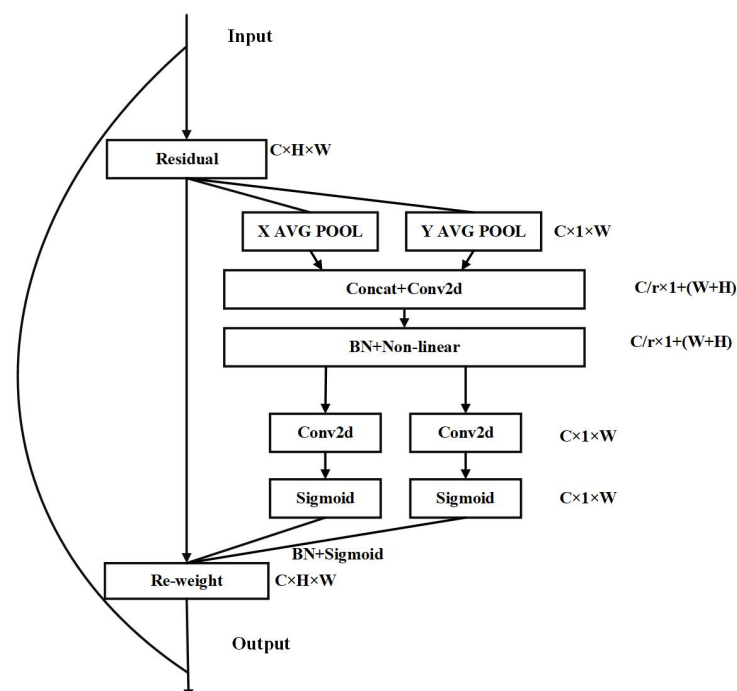


Figure 6. Coordinate Attention.

2.6. Modifications to the Decoder Module

In the decoder module of the U-Net architecture, two 3×3 standard convolutions are traditionally employed to extract features from merged feature maps. However, depthwise separable convolutions, which provide performance comparable to or better than standard convolutions, significantly reduce both the model's parameter size and computational burden. Consequently, after the feature fusion through skip connections in U-Net, depthwise separable convolutions are utilized to perform feature extraction from these fused feature maps. This adjustment not only preserves the effectiveness of feature extraction but also enhances the overall efficiency of the model, making it highly suitable for environments where computational resources are limited. Such improvements are crucial for optimizing the performance of the model without compromising on its operational efficacy.

The structure of a standard convolution is illustrated in Figure 7, where each convolutional kernel operates on all channels of the input feature map, subsequently generating the output feature map by summing the results across all channels. For an input feature map with M channels and a dimension size of D_F , when convolved with a convolutional kernel of size D_K , the output feature map with N channels is produced. Assuming a stride of 1 and that padding is used, the computation of the output features for the standard convolution is demonstrated in the following formula:

$$G_{k,l,n} = \sum_{i,j,m} K_{i,j,m,n} \cdot F_{k+i-1,l+j-1,m} \quad (1)$$

The calculation of the parameter count for standard convolutions is presented in the following equation:

$$D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F \quad (2)$$

The structure of depthwise separable convolutions is illustrated in Figure 8. Depthwise separable convolutions decompose the convolution operation into two steps; first, a depthwise convolution is performed where each convolution kernel operates on a single input channel, and then a pointwise convolution uses 1×1 convolutions to combine the outputs of the depthwise convolutions. This factorization significantly reduces the computational load and the model size. When implementing convolution operations with depthwise separable convolutions, the computation is as outlined in the equation.

$$G'_{k,l,n} = \sum K'_{i,j,m,n} \cdot F_{k+i-1,l+j-1,m} \quad (3)$$

The parameter count for the depthwise convolution is $D_F^2 \times M \times D_K^2$. The parameter count for the pointwise convolution operation is $D_F^2 \times M \times N$. The computational load for the depthwise separable convolution is calculated as shown in the formula.

$$D_F^2 \times M \times D_K^2 + M \times N \times D_F^2 \quad (4)$$

The calculation of the ratio of parameter counts between standard convolutions and depthwise separable convolutions is outlined in the formula.

$$\frac{D_F^2 \times M \times D_K^2 + M \times N \times D_F^2}{D_F^2 \times M \times N \times D_K^2} = \frac{1}{N} + \frac{1}{D_K^2} \quad (5)$$

Additionally, the parameter count for standard convolutions is $N \times M \times D_K^2$, while for depthwise separable convolutions, the depth convolution's parameter count is $M \times D_K^2$ and the pointwise convolution's parameter count is $1 \times 1 \times N \times M$. Therefore, the total parameter count for depthwise separable convolutions amounts to $M(D_K^2 + N)$. From this, it can be deduced that if using a 3×3 convolution kernel, the computational load of depthwise separable convolutions is 8 to 9 times that of standard convolutions. By decomposing the convolution operation into two steps, depthwise separable convolutions can extract features more efficiently, reduce the complexity of the model, accelerate the training and inference

processes, and decrease the risk of overfitting. Consequently, employing depthwise separable convolution modules in the decoder part of the network effectively enhances model performance, particularly suitable for real-time tasks that require efficient processing.

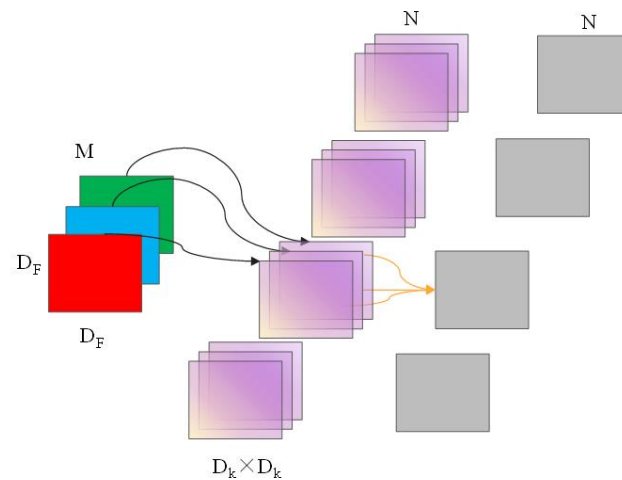


Figure 7. Standard convolution.

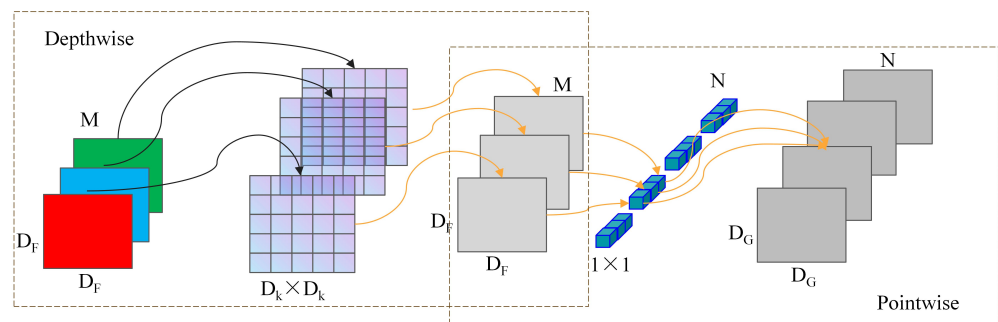


Figure 8. Depthwise separable convolution.

As illustrated in Figure 9, the original encoder in the U-Net architecture employs two standard convolutions ($\text{Conv}3 \times 3$) for feature extraction. This study proposes replacing them with two depthwise separable convolutions ($\text{SepConv}3 \times 3$). While extracting features using depthwise separable convolutions, a residual structure is used to perform a simple identity mapping. This residual structure is combined with the output from the depthwise separable convolutions, addressing the issue of vanishing gradients and enriching the extracted features of the image.

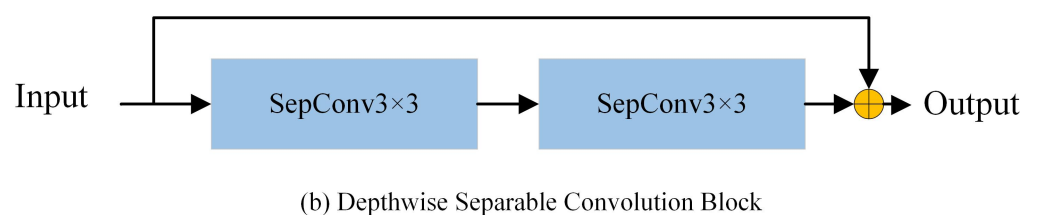
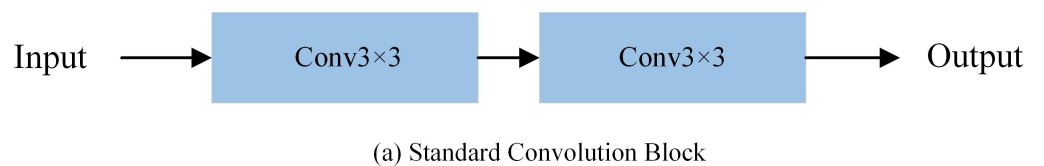


Figure 9. Convolution module.

3. Results

3.1. Selection of Attention Modules

In the U-Net model utilizing MobileNetV3 as the backbone, layers 4, 5, 6, 11, 12, 13, 14, and 15 incorporate the SE attention mechanism, which includes two fully connected layers that perform dimensionality expansion and reduction to obtain the weights for each channel. The spatial complexity of the fully connected layers is closely tied to the size of the input data; larger image dimensions result in increased model volume, computational load, and parameter count. For this experiment, the input image size is set at 512×512 , which results in the SE parameters occupying a significant proportion of the total parameter count in these layers. To reduce the parameter burden, this section substitutes the lighter and more efficient CA module for the SE module. Comparative analysis, as shown in Table 2, reveals that the parameter count of the CA attention module is substantially lower than that of the SE attention module. This substitution not only decreases the model's parameter count but also enhances its performance by reducing computational complexity and preserving essential spatial information more effectively.

Table 2. Number of parameters and percentage of parameters in layers with SE and CA attention modules.

Layer	4	5	6	11	12	13	14	15
SE Params (k)	3.55	7.83	7.83	115.8	226.63	226.63	462.0	1492.61
SE Proportion (%)	0.059	0.131	0.131	1.939	3.795	3.795	7.737	25.324
CA Params (k)	0.872	2.53	2.53	43.26	84.76	84.76	172.92	569.612
CA Proportion (%)	0.017	0.050	0.050	0.0861	1.687	1.687	3.442	11.236

A comparative analysis was conducted on two activation functions within the CA module—Sigmoid and h-Swish—to assess their specific impacts on segmentation performance. The study results show that the h-Swish activation function surpasses Sigmoid in key metrics such as MIoU, MPA, and accuracy (ACC). This indicates that h-Swish is more effective in enhancing the model's segmentation capabilities and also slightly reduces the number of parameters. Detailed comparative results can be seen in Table 3.

Table 3. Performance comparison of CA modules with different activation functions.

CA Module	MIoU (%)	MPA (%)	ACC (%)	FLOPs (GMac)	Params (M)
Sigmoid	86.37	92.34	97.89	3.94	5.02
h-Swish	87.97	93.15	98.36	4.36	4.37

To substantiate the hypothesis regarding the performance enhancements provided by the CA module, a comprehensive comparative analysis of various attention modules was conducted. The results, presented in Table 4, distinctly illustrate the effectiveness of different modules. Compared to other modules such as CBAM, SE, and Eca, the CA module demonstrates superior performance enhancement capabilities, while maintaining lower computational demands and efficiently managing parameters. These improvements demonstrate that the CA module can better integrate features and focus attention, optimizing the segmentation tasks for wheat spike disease spots.

Table 4. Performance comparisons of attention modules.

Attention Module	MIoU (%)	MPA (%)	ACC (%)	FLOPs (GMac)	Params (M)
CBAM	86.65	92.00	98.03	3.94	5.21
SE	87.15	91.52	98.27	3.94	5.97
Eca	86.66	92.32	97.95	3.93	4.46
CA	87.97	93.15	98.36	4.36	4.37

3.2. Ablation Experiment

An ablation study is used to assess the role of different modules in a model by gradually removing or replacing certain modules to observe their impact on model performance. In this experiment, the objective is to evaluate the impact of using MobileNetV3 as the backbone of the U-Net architecture, improvements in the decoder, and the introduction of the CA module on wheat spikelet disease segmentation performance. By comparing various models in terms of MIoU (Mean Intersection over Union), MPA (Mean Pixel Accuracy), ACC (Accuracy), FLOPs (Floating Point Operations), and the number of parameters, we can assess the effect of a lightweight design on model performance. Because the original U-Net model's feature extraction network is similar to VGG, the U-Net with a VGG backbone is chosen as the baseline for this ablation experiment.

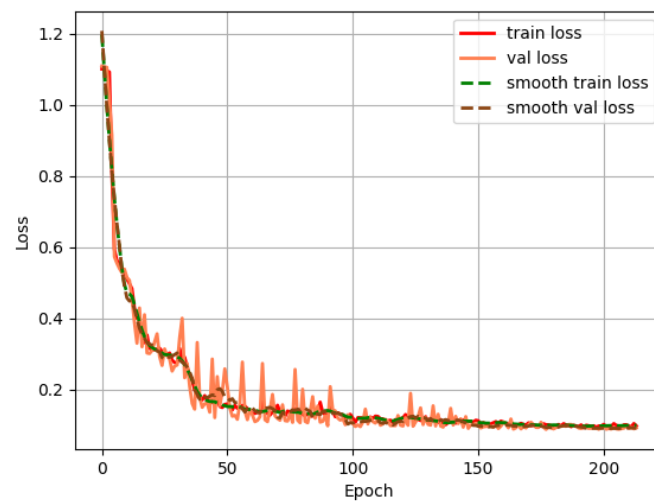
As indicated in Table 5, replacing the encoder with MobileNetv3 resulted in performance enhancements for the model, with mean Intersection over Union (MIoU) increasing to 87.15% and mean Pixel Accuracy (MPA) to 91.52%. There was also a significant reduction in both the model's Floating Point Operations per Second (FLOPs) and parameter count, demonstrating the critical role of MobileNetv3 in lightening the model. Given the need to enhance model segmentation precision while achieving lightweight design, the Squeeze-and-Excitation (SE) attention module was replaced with the more sensitive CA module. This change further improved performance, with MIoU reaching 87.97% and MPA 93.15%. Additionally, while FLOPs slightly increased compared to the previous setup, the model's parameter count was further reduced, showing that the CA attention module not only enhances model performance but also reduces computational complexity and parameter volume.

Table 5. Comparative results of ablation experiments. '✓' indicates that the corresponding improvement was implemented in the experiment.

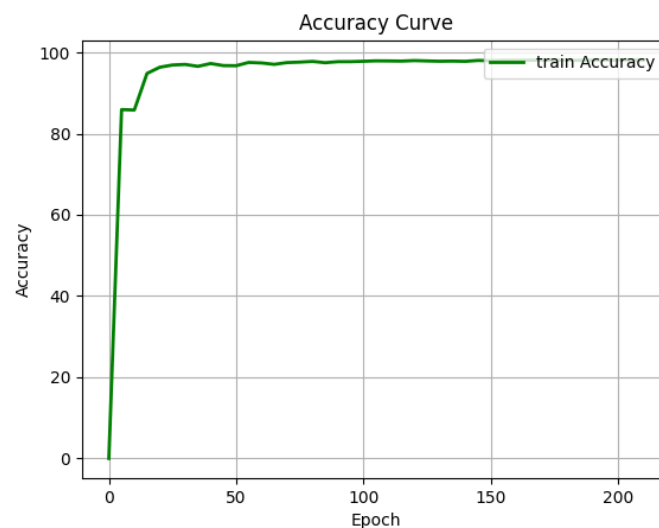
Model	MobileNetv3	Decoder Improvement	CA	MIoU (%)	MPA (%)	ACC (%)	FLOPs (GMac)	Param Count (M)
U-Net	✓	✓		85.08	90.44	98.1	283.97	26.45
				87.15	91.52	98.27	3.94	5.97
				87.97	93.15	98.36	4.36	4.37
	✓	✓	✓	87.67	92.62	98.35	2.66	4.61
			✓	88.57	95.97	98.65	3.51	4.25

The incorporation of depthwise separable convolutions to replace the convolutions in the decoder and the addition of residual connections optimized the structure of the decoder, enhancing overall performance, with MIoU reaching 87.67 and FLOPs being only 2.66 GMac. This confirms that the improvements in the decoder module significantly enhanced model performance. Overall, the enhanced U-Net showed substantial improvements over the original model in terms of performance. Not only did it achieve higher segmentation accuracy with MIoU reaching 88.57%, but it also saw a dramatic reduction in parameter volume and FLOPs, with the model depth being reduced and the parameter count becoming one-sixth of its original size. These lightweight enhancements indicate that the modified model is capable of handling real-time tasks for wheat spike and disease spot segmentation.

The model's loss curve and accuracy curve on the validation set are shown in Figure 10. After training, the model's loss curve converges, and the training accuracy reaches 98.65%. The initial phase of training shows a declining trend in loss, indicating that the model is progressively reducing errors throughout the learning process. As training continues, the loss consistently decreases while the accuracy consistently rises, suggesting continuous performance enhancements throughout the iterations. After 150 training cycles, the loss stabilizes at an extremely low level, and the accuracy reaches a very high value, showcasing the model's exceptional performance and stability in segmenting wheat spike disease spots.



(a) Loss curve



(b) Accuracy curve

Figure 10. Comparison of loss and accuracy curves.

3.3. Performance Comparison among Different Models

To validate the effectiveness of the lightweight model, several proposed models were compared through segmentation experiments. The results of different models on the same wheat spikelet disease segmentation are shown in Figure 11. The figure illustrates how various models perform in segmenting wheat spikelet disease, allowing for an evaluation of each model's segmentation performance by comparing it with the original image (Figure 11a). The segmentation outcomes vary in terms of accuracy and detail. In Figure 11b, the proposed model in this study demonstrates outstanding performance in wheat spikelet disease segmentation. It successfully segments the disease-affected areas, and compared to the original image, the edge treatment is relatively accurate, and the segmentation result appears natural. Figure 11c,d show the segmentation results of PSPNet with MobileNet and ResNet50 as backbone networks, respectively.

Both exhibit similar performance, accurately segmenting the disease-affected regions, though the edge handling is slightly fuzzy, with less detail compared to the proposed model. Figure 11e,f show the results of DeepLabv3 using Xception and MobileNet as backbone networks. Both have fairly accurate segmentation results, but the edge areas still

have some blurriness, indicating that DeepLabv3 falls slightly short of the proposed model in terms of detail. Figure 11g illustrates the segmentation result of VGGU-Net, where the edges are somewhat coarse, and the disease-affected regions lack detail, indicating lower accuracy compared to other models. Figure 11h represents the segmentation result of ResU-Net. Among all the models, ResU-Net has relatively clear edges, but it may exhibit over-segmentation issues. Overall, the proposed model in this study demonstrates high accuracy and stability in wheat spikelet disease segmentation tasks. The edge treatment is relatively fine, and the segmentation result aligns more closely with actual disease-affected areas, enhancing its practical application in wheat spikelet disease segmentation tasks.

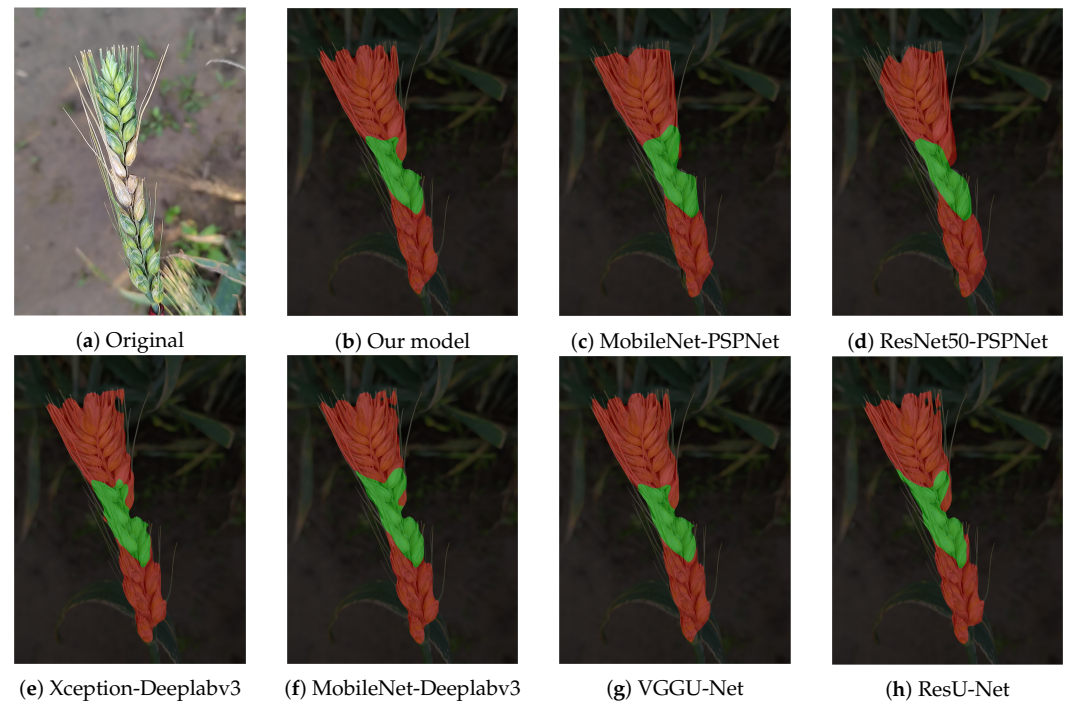


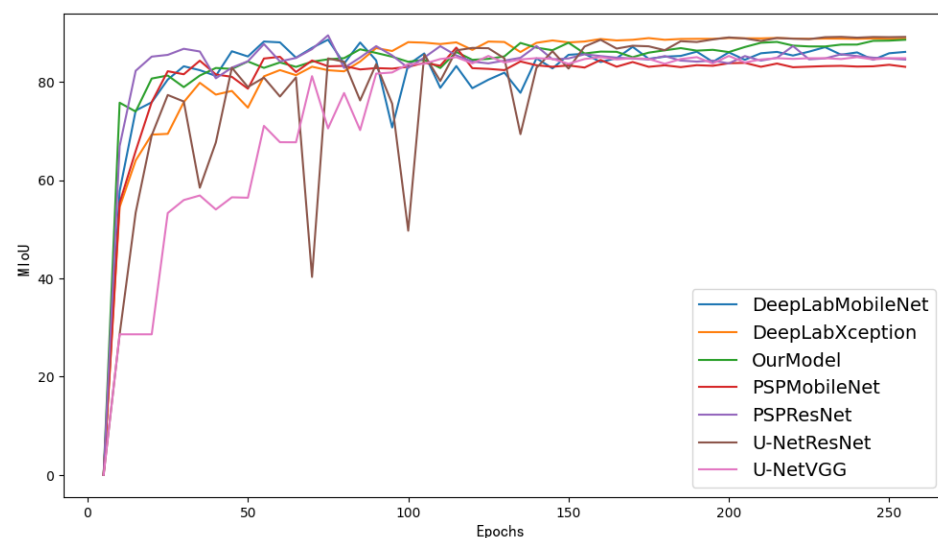
Figure 11. Visualization of wheat spikes and corresponding segmentation masks. Green represents the diseased areas, red indicates wheat spikes, and black denotes the background.

Table 6 compares the performance metrics of different semantic segmentation networks. It is observed that ResU-Net achieves the highest MIOU at 89.16%, indicating superior performance; however, it also has relatively high FLOPs and parameter count. PSPNet with ResNet50 demonstrates lower segmentation MIOU, and both its FLOPs and parameter size are significant. Although Deeplabv3 with Xception and ResU-Net exhibit excellent performance, their high FLOPs and parameter counts may limit their deployment in resource-constrained environments. Deeplabv3 with MobileNet and PSPNet with MobileNet have achieved model lightweighting, but their segmentation accuracy is considerably lower than the other models. The model presented in this paper shows a relative advantage over the others, achieving an MIOU of 88.57%, second only to Deeplabv3 with Xception and ResU-Net, while significantly outperforming these models in terms of FLOPs and parameter count, at 3.51 and 4.25, respectively. This allows it to deliver excellent segmentation performance even in situations with limited computational resources. Therefore, the model presented effectively addresses the issue of limited computing resources while maintaining performance, bringing more flexibility and feasibility to practical application scenarios.

Table 6. Comparative results of different segmentation models.

Model	MIoU (%)	MPA (%)	ACC (%)	FLOPs (GMac)	Param Count (M)
Deeplabv3-MobileNet	85.53	90.78	98.27	26.4	5.81
Deeplabv3-Xception	88.78	93.83	98.42	83.3	54.71
ResU-Net	89.16	93.61	98.43	91.91	46.45
PSPNet-ResNet50	84.55	89.32	98.34	162.16	47.71
PSPNet-MobileNet	83.07	88.70	98.00	8.61	2.38
U-Net-VGG	85.08	90.44	98.10	283.97	26.45
Our Model	88.57	95.97	98.65	3.51	4.25

The training curves of the model, as shown in Figure 12, indicate that most models converge after about 150 training cycles.

**Figure 12.** Comparison of model MIoU curves.

To validate the advantages of the proposed model in terms of lightweight design and high precision, this section conducts comparative experiments with U-Net models using different lightweight backbone networks. In the U-Net architecture, different backbone networks may lead to variations in the number of feature layers. Typically, the number of feature layers depends on the structure and depth of the selected backbone network; deeper backbone networks may generate more feature layers. Therefore, when constructing U-Net with different lightweight networks, efforts are made to maintain the model's segmentation accuracy without deliberately reducing the number of channels used for feature extraction.

According to the training results shown in Table 7, the model discussed in this section demonstrates excellent performance across various metrics such as Mean Intersection over Union (MIoU), Mean Pixel Accuracy (MPA), Accuracy (ACC), Floating Point Operations Per Second (FLOPs), and parameter count. The model achieves an MIoU of 88.57%, with relatively low FLOPs and parameter values, at 3.51 GMac and 4.25 M, respectively.

In comparison, while the model using FasterNet as the backbone achieves an MIoU close to this model, its higher FLOPs and parameter count impact resource efficiency. Models using GhostNet and MobileNetV3 as backbones perform better in terms of FLOPs and parameter count but have slightly lower MIoU than this model. The model with EfficientNet as the backbone reaches an MIoU of 87.15, but has relatively higher FLOPs and parameter count. The model with ShuffleNet as the backbone shows poorer segmentation performance and parameter efficiency, making it less suitable for lightweight disease spot segmentation tasks. Overall, the discussed model maintains high performance while consuming fewer computational resources, making it suitable for use in resource-constrained

environments. Additionally, Figure 13 presents a comparison of the training MIoU curves for models using different lightweight backbones, showing that most models tend to converge after about 150 epochs.

Table 7. Comparative results of different models.

Model	MIoU (%)	MPA (%)	ACC (%)	FLOPs (GMac)	Param Count (M)
ShuffleNet	84.52	90.02	97.96	30.9	10.7
FasterNet	87.97	93.15	98.36	65.15	23.28
GhostNet	85.21	90.94	97.78	3.53	3.00
MobileNetV3	87.15	91.52	98.27	3.94	5.97
EfficientNet	87.15	92.44	98.21	49.28	12.17
MobileNetV2	86.05	91.3	98.07	4.27	4.02
Our Model	88.57	93.59	98.65	3.51	4.25

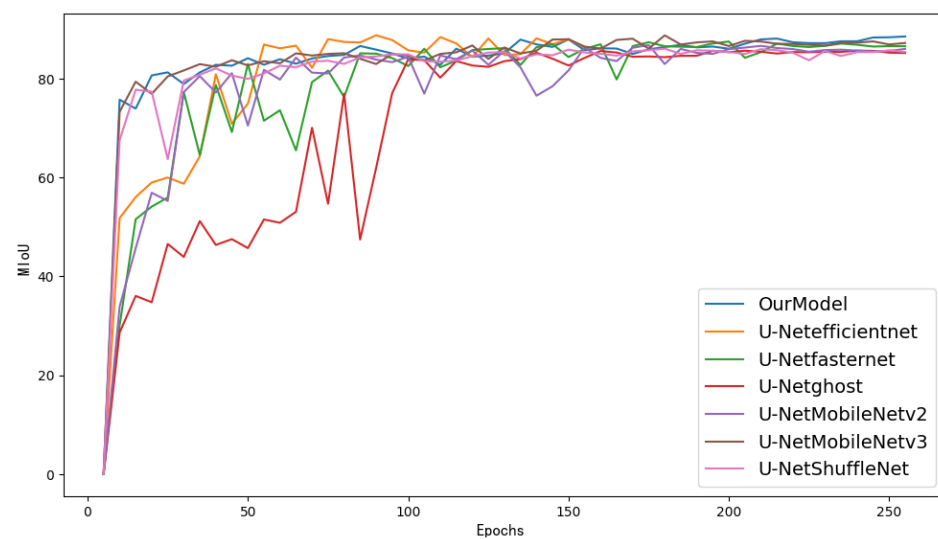


Figure 13. Comparison of model MIoU curves.

The segmentation results of the model are displayed in Figure 14. Under a simple background, the model can accurately segment the wheat spikes and diseased areas. It is evident that the lightweight model presented in this paper can segment disease spots of various sizes and locations, with the edges of the disease spots closely fitting the contours of the spikelets, indicating that the model can achieve precise segmentation of disease spots. By calculating the ratio of pixel points between the diseased area and the wheat spike area (green and red areas, respectively), the severity of the disease affecting the wheat spikes can be determined.

The pixel count of diseased areas, spike pixel count, proportion of diseased area, and severity level of the wheat spikes in the four segmented images in Figure 14 are shown in Table 8. The model accurately calculates the pixel count of diseased areas, spike pixel count, and proportion of diseased area while performing disease segmentation, thereby determining the severity level of the disease affecting the wheat spikes.

Table 8. Lesion segmentation results statistics.

Method	Lesion Pixels	Lesion Voxels	Lesion Area Ratio (%)	Severity Level
(a)	1,673,899	702,447	41.9	2
(b)	1,801,142	628,711	34.9	2
(c)	1,819,883	359,814	19.7	1
(d)	1,783,752	814,988	45.7	2

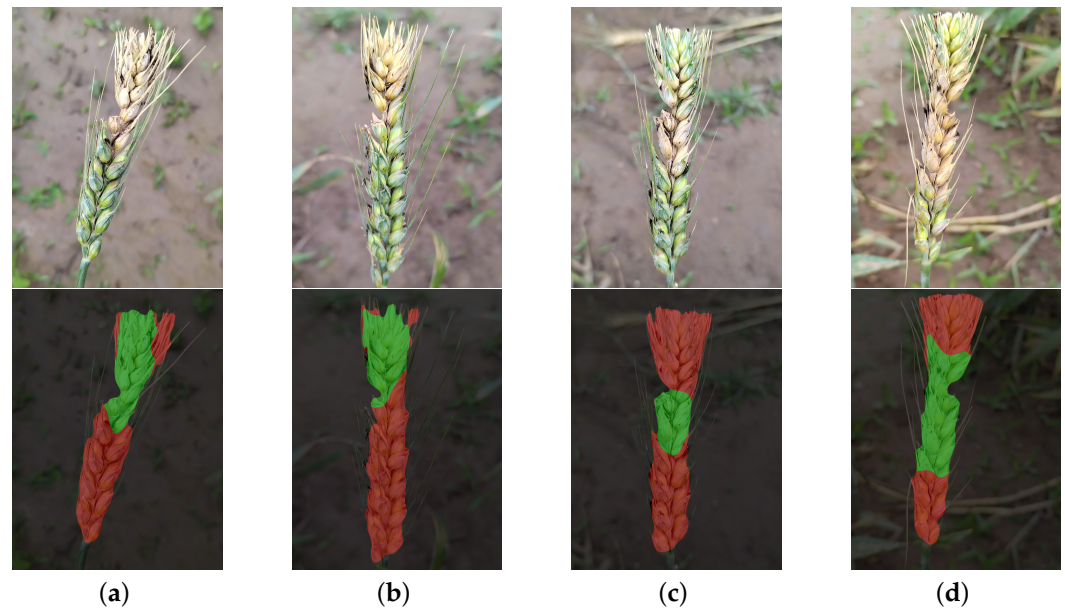


Figure 14. Visualization of wheat spikes and corresponding segmentation masks. Green represents the diseased areas, red indicates wheat spikes, and black denotes the background. (a), (b), (c), and (d) each represent different wheat spikes and their corresponding segmentation results.

4. Discussion

4.1. A Lightweight Model

This study introduces a lightweight U-Net-based model designed to assess the severity of Fusarium head blight (FHB) in wheat. Unlike conventional models, our approach utilizes MobileNetv3 as the backbone for feature extraction and incorporates a CA module to enhance accuracy and efficiency. By integrating these features, the proposed model achieved a segmentation MIoU of 88.87%, which is slightly below that of ResU-Net at 89.16%, but with only 4.52 M parameters, which is one-tenth of those used by ResU-Net, making it highly suitable for real-time segmentation tasks in resource-constrained environments. It can help facilitate quick and easy estimation of the severity of Fusarium head blight in the field.

Compared to traditional U-Net architectures, our model demonstrates substantial improvements in performance and computational efficiency. The combination of MobileNetv3 with CA attention modules enables more effective feature extraction while maintaining a reduced parameter footprint. This lightweight design is essential for applications in real-time field environments, where computational resources are limited. Additionally, the accuracy of our model in disease identification has been validated through rigorous ablation studies, indicating its robustness in real-world applications. When benchmarked against traditional methods such as fully convolutional networks and pulse-coupled neural networks, our model consistently achieved higher segmentation accuracy and stability [25].

4.2. Robustness under Different Conditions

Gao et al. [34] have achieved single-spike wheat severity estimation using image classification, but these approaches generally require significant time and effort to categorize different severity datasets, and the backgrounds in the data are often simplistic. An essential aspect of our model's effectiveness is its robustness across various environmental conditions, including different growth stages and challenging field backgrounds. The model demonstrated strong capabilities in segmenting wheat ears and disease spots, even under shadow and soil effects, indicating its resilience in real-world scenarios. This robustness contributes significantly to the model's potential for large-scale field applications, enabling accurate FHB severity estimation across diverse conditions.

4.3. Future Directions

While the proposed model has shown improvements, there are still areas for further research. Enhancing the CA attention module and exploring additional lightweight methods could lead to even greater accuracy in disease spot segmentation. Furthermore, expanding the diversity of the dataset and improving data augmentation strategies could enhance the model's generalization capabilities. Overall, the presented model offers a promising approach to FHB severity estimation in wheat, with potential applications in agricultural monitoring and precision agriculture. By balancing accuracy and efficiency, this work paves the way for more effective disease detection and control in the field, contributing to improved crop management and food security.

5. Conclusions

In conclusion, this study introduced a lightweight U-Net-based model for segmenting wheat spikes and estimating the severity of Fusarium head blight. Utilizing MobileNetV3 as the backbone for feature extraction and incorporating CA modules, the model demonstrated improved segmentation accuracy and reduced computational complexity. With an increase in the Mean Intersection over Union (MIoU) to 88.57% and a reduction in the model's parameter count to 4.52 M—only one-sixth of the original model's size—this model can efficiently perform disease spot segmentation and severity estimation.

The integration of depthwise separable convolutions and CA attention mechanisms played a significant role in enhancing the model's performance, yielding better segmentation results compared to traditional methods. The lightweight design not only facilitates real-time wheat spike segmentation and disease severity estimation but also offers a practical solution adaptable to resource-constrained environments.

Author Contributions: Conceptualization, L.S. and Z.L.; methodology, L.S. and Z.L.; software, Z.L.; validation, Z.L. and C.Y.; formal analysis, Z.L. and J.L.; investigation, Z.L., C.Y. and J.L.; resources, L.S., Z.L. and F.Y.; data curation, L.S. and Z.L.; writing—original draft preparation, Z.L.; writing—review and editing, Z.L. and Q.W.; visualization, J.L. and J.W.; supervision, L.S. and F.Y.; project administration, L.S. and J.W.; funding acquisition, L.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (NO. 31501225) and Joint Fund of Science and Technology Research and Development Plan of Henan Province (NO. 222301420113), the Natural Science Foundation of Henan Province of China (NO. 232300420186), the Key Scientific and Technological Project of Henan Province (NO. 24210211193), the Major Science and Technology Special Projects of Henan Province (NO. 221100210600).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data will be made available on request.

Acknowledgments: The authors are thankful to Guihong Yin for his strong support in this work. The authors would like to thank the editor and anonymous reviewers for their helpful comments and suggestions.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Shi, J.; Liu, X.; Qiu, J.; Ji, F.; Xu, J.; Dong, F.; Yin, X.; Ran, J. Deoxynivalenol contamination in wheat and its management. *Sci. Agric. Sin.* **2014**, *47*, 3641–3654.
2. Cuperlovic-Culf, M.; Wang, L.; Forseille, L.; Boyle, K.; Merkley, N.; Burton, I.; Fobert, P.R. Metabolic biomarker panels of response to fusarium head blight infection in different wheat varieties. *PLoS ONE* **2016**, *11*, e0153642. [[CrossRef](#)] [[PubMed](#)]
3. Sood, S.; Singh, H. An implementation and analysis of deep learning models for the detection of wheat rust disease. In Proceedings of the 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), Thoothukudi, India, 3–5 December 2020; pp. 341–347. [[Crossref](#)]

4. Liu, B.Y.; Fan, K.J.; Su, W.H.; Peng, Y. Two-stage convolutional neural networks for diagnosing the severity of alternaria leaf blotch disease of the apple tree. *Remote Sens.* **2022**, *14*, 2519. [[CrossRef](#)]
5. Singh, V.; Misra, A.K. Detection of plant leaf diseases using image segmentation and soft computing techniques. *Inf. Process. Agric.* **2017**, *4*, 41–49. [[CrossRef](#)]
6. Patrício, D.I.; Rieder, R. Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review. *Comput. Electron. Agric.* **2018**, *153*, 69–81. [[CrossRef](#)]
7. Liakos, K.G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine learning in agriculture: A review. *Sensors* **2018**, *18*, 2674. [[CrossRef](#)] [[PubMed](#)]
8. Sarayloo, Z.; Asemani, D. Designing a classifier for automatic detection of fungal diseases in wheat plant: By pattern recognition techniques. In Proceedings of the 2015 23rd Iranian Conference on Electrical Engineering, Tehran, Iran, 10–14 May 2015; pp. 1193–1197. [[Crossref](#)]
9. Gewali, U.B.; Monteiro, S.T.; Saber, E. Machine learning based hyperspectral image analysis: A survey. *arXiv* **2018**, arXiv:1802.08701. [[CrossRef](#)]
10. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36. [[CrossRef](#)]
11. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2016; pp. 3431–3440. [[Crossref](#)]
12. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [[Crossref](#)]
13. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788. [[Crossref](#)]
14. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; Proceedings, Part III 18. Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241. [[Crossref](#)]
15. Zhang, B.; Zhang, M.; Chen, Y. Crop pest identification based on spatial pyramid pooling and deep convolution neural network. *Trans Chin Soc Agric Eng* **2019**, *35*, 209–215.
16. Wang, D.; He, D. Recognition of apple targets before fruits thinning by robot based on R-FCN deep convolution neural network. *Trans. CSAE* **2019**, *35*, 156–163.
17. Wenxia, B.; Qing, S.; Qing, S.; Linsheng, H.; Dong, L.; Jian, Z. Image recognition of field wheat scab based on multi-way convolutional neural network. *Trans. Chin. Soc. Agric. Eng* **2020**, *36*, 174–181.
18. Shi, L.; Sun, J.; Dang, Y.; Zhang, S.; Sun, X.; Xi, L.; Wang, J. YOLOv5s-T: A lightweight small object detection method for wheat spikelet counting. *Agriculture* **2023**, *13*, 872. [[CrossRef](#)]
19. Kaiyu, L.; Xinyi, Z.; Juncheng, M.; Lingxian, Z. Estimation Method of Leaf Disease Severity of Cucumber Based on Mixed Dilated Convolution and Attention Mechanism. *J. Agric. Mach.* **2023**, *54*, 231–239.
20. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890. [[Crossref](#)]
21. Chen, L.C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv* **2017**, arXiv:1706.05587. [[Crossref](#)]
22. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [[CrossRef](#)] [[PubMed](#)]
23. Howard, A.; Sandler, M.; Chu, G.; Chen, L.C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V.; et al. Searching for mobilenetv3. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1314–1324.
24. Deng, G.Q.; Wang, J.C.; Yang, J.; Liu, T.; Li, D.S.; Sun, C.M. Identification of Fusarium head blight in wheat ears based on image and improved U-net model. *J. Triticeae Crops* **2021**, *41*, 1432–1440.
25. Zhang, D.; Wang, D.; Gu, C.; Jin, N.; Zhao, H.; Chen, G.; Liang, H.; Liang, D. Using neural network to identify the severity of wheat Fusarium head blight in the field environment. *Remote Sens.* **2019**, *11*, 2375. [[CrossRef](#)]
26. Wang, Y.H.; Li, J.J.; Su, W.H. An Integrated Multi-Model Fusion System for Automatically Diagnosing the Severity of Wheat Fusarium Head Blight. *Agriculture* **2023**, *13*, 1381. [[CrossRef](#)]
27. Li, C.; Li, H.; Gao, G.; Liu, Z.; Liu, P. An accelerating convolutional neural networks via a 2D entropy based-adaptive filter search method for image recognition. *Appl. Soft Comput.* **2023**, *142*, 110326. [[CrossRef](#)]
28. Zhang, X.; Huang, H. Distilling Knowledge from a Transformer-Based Crack Segmentation Model to a Light-Weighted Symmetry Model with Mixed Loss Function for Portable Crack Detection Equipment. *Symmetry* **2024**, *16*, 520. [[CrossRef](#)]
29. Zhu, L.; Deng, W.; Lai, Y.; Guo, X.; Zhang, S. Research on Improved Road Visual Navigation Recognition Method Based on DeepLabV3+ in Pitaya Orchard. *Agronomy* **2024**, *14*, 1119. [[CrossRef](#)]
30. Wu, J.; Dong, J.; Nie, W.; Ye, Z. A lightweight YOLOv5 optimization of coordinate attention. *Appl. Sci.* **2023**, *13*, 1746. [[CrossRef](#)]
31. National Standard GB/T 15796-2011; Technical Specification for Monitoring and Forecasting of Wheat Head Blight. Standards Press of China: Beijing, China, 2011.

32. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2019; pp. 7132–7141. [[Crossref](#)]
33. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19. [[Crossref](#)]
34. Gao, C.; Gong, Z.; Ji, X.; Dang, M.; He, Q.; Sun, H.; Guo, W. Estimation of fusarium head blight severity based on transfer learning. *Agronomy* **2022**, *12*, 1876. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.