

Article

Detection of Fusarium Head Blight in Individual Wheat Spikes Using Monocular Depth Estimation with Depth Anything V2

Jiacheng Wang ^{1,2,†}, Jianliang Wang ^{1,2,3,†} , Yuanyuan Zhao ^{1,2}, Fei Wu ⁴, Wei Wu ⁵, Zhen Li ³ ,
Chengming Sun ^{1,2} , Tao Li ^{1,2}  and Tao Liu ^{1,2,*}

- ¹ Jiangsu Key Laboratory of Crop Genetics and Physiology/Jiangsu Key Laboratory of Crop Cultivation and Physiology, Agricultural College of Yangzhou University, Yangzhou 225009, China; mx120240758@stu.yzu.edu.cn (J.W.); dx120240146@stu.yzu.edu.cn (J.W.); dx120220116@stu.yzu.edu.cn (Y.Z.); cmsun@yzu.edu.cn (C.S.); taoli@yzu.edu.cn (T.L.)
- ² Jiangsu Co-Innovation Center for Modern Production Technology of Grain Crops, Yangzhou University, Yangzhou 225009, China
- ³ Faculty of Engineering, Kyushu Institute of Technology, 1-1 Sensui, Tobata-ku, Kitakyushu 804-8550, Japan; li.zhen799@mail.kyutech.jp
- ⁴ Precision Agriculture Lab, School of Life Sciences, Technical University of Munich, 85354 Freising, Germany; feiy.wu@tum.de
- ⁵ Agricultural Information Institute, Chinese Academy of Agricultural Sciences/Key Laboratory of Agro-Information Services Technology, Ministry of Agriculture and Rural Affairs, Beijing 100081, China; wuwei@caas.cn
- * Correspondence: tliu@yzu.edu.cn
- † These authors contributed equally to this work.

Abstract

Fusarium head blight (FHB) poses a significant threat to global wheat yields and food security, underscoring the importance of timely detection and severity assessment. Although existing approaches based on semantic segmentation and stereo vision have shown promise, their scalability is constrained by limited training datasets and the high maintenance cost and complexity of visual sensor systems. In this study, AR glasses were employed for image acquisition, and wheat spike segmentation was performed using Depth Anything V2, a monocular depth estimation model. Through geometric localization methods—such as identifying abrupt changes in stem width—redundant elements (e.g., awns and stems) were effectively excluded, yielding high-precision spike masks (Precision: 0.945; IoU: 0.878) that outperformed leading semantic segmentation models including Mask R-CNN and DeepLabv3+. The study further conducted a comprehensive analysis of differences between diseased and healthy spikelets across RGB, HSV, and Lab color spaces, as well as three color indices: Excess Green–Excess Red (ExGR), Normalized Difference Index (NDI), and Visible Atmospherically Resistant Index (VARI). A dynamic fusion weighting strategy was developed by combining the Lab-a* component with the ExGR index, thereby enhancing visual contrast between symptomatic and asymptomatic regions. This fused index enabled quantitative assessment of FHB severity, achieving an R^2 of 0.815 and an RMSE of 8.91%, indicating strong predictive accuracy. The proposed framework offers an intelligent, cost-effective solution for FHB detection, and its core methodologies—depth-guided segmentation, geometric refinement, and multi-feature fusion—present a transferable model for similar tasks in other crop segmentation applications.

Keywords: wheat; fusarium head blight; monocular image; depth estimation; feature fusion



Academic Editor: Francis Drummond

Received: 20 October 2025

Revised: 17 November 2025

Accepted: 18 November 2025

Published: 19 November 2025

Citation: Wang, J.; Wang, J.; Zhao, Y.; Wu, F.; Wu, W.; Li, Z.; Sun, C.; Li, T.; Liu, T. Detection of Fusarium Head Blight in Individual Wheat Spikes Using Monocular Depth Estimation with Depth Anything V2. *Agronomy* **2025**, *15*, 2651. <https://doi.org/10.3390/agronomy15112651>

Copyright: © 2025 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 most important staple crops globally, extensively cultivated across diverse regions and playing an irreplaceable role in agricultural production. Fusarium head blight (FHB), a climate-associated fungal disease caused by *Fusarium graminearum*, is prevalent in wheat-growing areas worldwide [1]. In severe outbreaks, it can result in complete grain loss, drawing considerable attention across breeding, grain production, and food processing sectors [2]. Accordingly, timely detection and accurate assessment of FHB severity are critical for safeguarding wheat yield. Traditional field-based diagnosis is time-consuming and labor-intensive, often relying heavily on the surveyor's experience, which introduces a high degree of subjectivity [3]. Therefore, developing a rapid, non-destructive, and high-throughput method for FHB detection is of particular importance.

With the ongoing development of digital agriculture, applying computational techniques to detect FHB in wheat has become a prominent area of research [4–6]. In particular, deep learning—a key driver of predictive automation—has accelerated the integration of computer vision into FHB detection workflows. Gao et al. [7] implemented an enhanced lightweight YOLOv5s model for identifying FHB-infected wheat spikes under natural field conditions. By combining MobileNetV3 with the C3Ghost module, the model achieved reduced complexity and parameter count while improving detection accuracy and real-time performance. Ma et al. [8] employed hyperspectral imaging alongside an optimized combination of SBs, VIs, and WFs, and applied the MMN algorithm for normalization, significantly enhancing FHB detection outcomes. Zhou et al. [9] introduced DeepFHB, a high-throughput deep learning framework incorporating a modified transformer module, DConv network, and group normalization. This architecture enabled efficient detection, localization, and segmentation of wheat spikes and symptomatic regions in complex field settings. Evaluated on the FHB-SA dataset, DeepFHB achieved a box AP of 64.408 and a mask AP of 64.966. Xiao et al. [10] proposed a hyperspectral imaging approach for detecting FHB in wheat by integrating spectral and texture features and optimizing the window size of the gray-level co-occurrence matrix (GLCM), which led to a notable improvement in detection accuracy. The study demonstrated that a 5×5 pixel window provided the highest accuracy (90%) during the early stage of infection, whereas a 17×17 pixel window was most effective in the late stage, also yielding an accuracy of 90%.

However, existing approaches for detecting FHB in wheat using RGB or hyperspectral imaging primarily depend on two-dimensional data captured by monocular cameras. These methods often underperform in the presence of complex field backgrounds and require extensive training datasets to reach acceptable accuracy levels [11]. Segmentation networks based on deep convolutional neural networks (DCNNs) also exhibit limitations, particularly in their inability to capture spatial relationships between pixels at different depths within an image [12]. In response, several studies have investigated crop detection methods that incorporate depth information. Wen et al. [13] employed a binocular camera system with parallel optical axes to generate disparity maps and compute 3D point clouds of wheat surfaces, thereby enabling the detection of lodging. By analyzing the angle between wheat stems and the vertical axis, lodging was categorized into upright, inclined, and fully lodged conditions and quantified using point cloud height. Liu et al. [14] proposed a method for pineapple fruit detection and localization that integrates an enhanced YOLOv3 model with binocular stereo vision, where YOLOv3 performs rapid 2D object detection and the stereo system supplies depth data to form a complete end-to-end detection and localization pipeline. Ye et al. [15] proposed a soybean plant recognition model utilizing a laser range sensor. By analyzing structural differences between soybean plants and weeds—specifically in diameter, height, and planting spacing—the model demonstrated the applicability of laser range sensing in complex field environments, providing a reliable solution for real-

time soybean identification and automated weeding. Jin et al. [16] introduced a method for maize stem–leaf segmentation and phenotypic trait extraction based on ground LiDAR data. A median-normalized vector growth (MNVG) algorithm was developed to achieve high-precision segmentation of individual maize stems and leaves. This approach enabled the extraction of multiscale phenotypic traits, including leaf inclination, length, width, and area, as well as stem height and diameter, and overall plant characteristics such as height, canopy width, and volume. Experimental results indicated that the MNVG algorithm achieved an average segmentation accuracy of 93% across 30 maize samples with varying growth stages, heights, and planting densities, with phenotypic trait extraction reaching an R^2 of up to 0.97.

Therefore, crop and plant segmentation methods based on depth information have become an important direction for improving segmentation accuracy. However, traditional stereo cameras and laser sensors face challenges such as high costs, complex calibration, and maintenance requirements, which limit their widespread adoption in large-scale applications [17]. To overcome these limitations, this study proposes an innovative method for wheat ear segmentation and scab detection based on a depth estimation model, aiming to advance depth-guided crop and plant segmentation technology through algorithmic innovation. (1) By reconstructing 3D depth information from single-view RGB images using a depth estimation model (Depth Anything V2 [18]), we address the limitation of traditional 2D images in providing spatial information, thereby enhancing the accuracy of wheat ear segmentation. This approach not only improves segmentation precision but also innovatively utilizes depth information to guide the segmentation process. (2) We propose a geometry-based localization method using sudden changes in stalk width to accurately eliminate redundant stalks. This innovative algorithm significantly optimizes crop segmentation accuracy through morphological feature analysis, overcoming the limitations of traditional methods in handling redundant structures within complex backgrounds. (3) By combining optimized color indices, we developed an efficient scab detection method. This approach enhances the identification accuracy of scab-infected regions through the fusion of color indices and depth information, demonstrating how depth-guided feature fusion can improve the performance of agricultural disease monitoring technologies. Experimental results show that the proposed method achieves an IoU of 0.878 for wheat ear segmentation and an R^2 of 0.815 for scab detection, outperforming traditional depth-guided methods that rely on stereo cameras or LiDAR. In comparison, our method only requires images captured by conventional RGB cameras, significantly reducing hardware costs while avoiding the complex processes of multi-sensor calibration and point cloud computation, thereby improving detection efficiency. Through these algorithmic innovations, this research not only proposes a low-cost, highly feasible precision agriculture solution but also makes significant algorithmic contributions to the field of depth-guided crop and plant segmentation.

2. Materials and Methods

2.1. Field Experiment

The experiment was conducted from October 2023 to June 2024 at the experimental fields of the Jiangsu Provincial Key Laboratory of Crop Genetics and Physiology, Yangzhou University, located in Yangzhou, Jiangsu Province, China (119°42' E, 32°38' N). The site is characterized by a northern subtropical monsoon climate, with an annual average temperature of 13.2 °C to 16.0 °C, annual precipitation of 1518.7 mm, and an annual sunshine duration of 2084.3 h. The soil is classified as light loam, with the 0–20 cm layer containing 65.81 mg kg^{−1} of hydrolyzable nitrogen, 45.88 mg kg^{−1} of available phosphorus, 101.98 mg kg^{−1} of available potassium, and 15.5 g kg^{−1} of organic matter.

Over twenty winter wheat varieties (lines) commonly grown in the middle and lower Yangtze River region were selected as experimental materials. A randomized block design was adopted, with each line sown in five rows measuring 2 m in length and spaced 0.2 m apart. Each row was uniformly sown with 30 seeds, and each variety was replicated three times (Figure 1A).

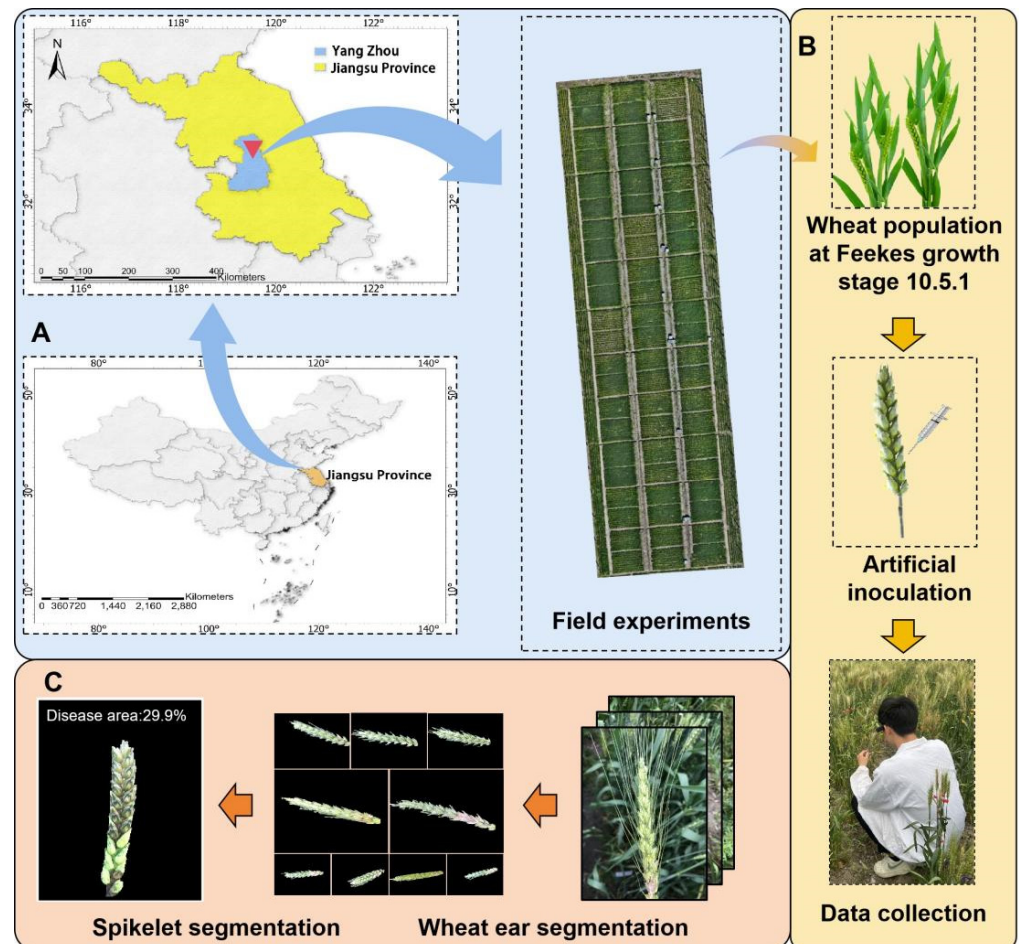


Figure 1. Technical route. Note, (A) shows the field experiments (The red triangle icon indicates the location of the test site on the map.), (B) shows the inoculation of FHB and data collection, (C) shows the data processing and analysis.

FHB was induced using the single-floret injection method to artificially inoculate the wheat lines. A mixture of highly virulent *Fusarium* strains (F0301, F0609, F0980) was used as the inoculum. These strains were first cultured on PDA plates at 25 °C for 5–7 days to ensure hyphal purity and viability. They were subsequently transferred to liquid medium and incubated under constant shaking at 25–28 °C for 3 days to prepare the spore suspension. Spore concentration was determined using a hemocytometer and adjusted to 1×10^5 /mL, followed by mixing the three strains in equal volumes (1:1:1). At Feekes growth stage 10.5.1, twenty wheat spikes at a uniform developmental stage were selected, and 10 μ L of the spore suspension was injected into a single floret at the center of each spike using a syringe. Artificial misting was applied immediately after inoculation to maintain humidity (Figure 1B).

2.2. Image Acquisition of Single-Spike Scab

Image acquisition was carried out from the time of FHB inoculation until Feekes growth stage 11.1 (early grain-filling stage). The imaging device selected for this study was

a pair of AR smart glasses, a choice justified by the requirement for a mobile, hands-free data acquisition system that mimics a natural plant inspection scenario while ensuring high image quality. The specific model used was a pair of AR smart glasses equipped with a high-resolution camera (Changzhou Jinhe New Energy Technology Co., Ltd., Changzhou, China), with a maximum resolution of 2448×3264 pixels (Telephoto Camera, 8 MP, $5\times$ optical zoom, 120 mm focal length). The device features voice control and autofocus capabilities, and image data were transmitted to a server via a mobile phone with cellular connectivity using the Wi-Fi protocol.

Data collection was conducted between 10:00 and 14:00 under natural light conditions to ensure high image quality and accurate color fidelity. The specific procedures for image acquisition were as follows:

- (1) The operator gently holds the wheat stem with one hand to stabilize the target plant.
- (2) The target spike is positioned 30–50 cm from the camera lens to ensure an appropriate field of view.
- (3) The spike is centered within the frame and oriented perpendicular to the camera.
- (4) Autofocus and image capture are triggered via voice command or by lightly tapping the touch-sensitive module on the side of the glasses frame. This hands-free operation minimized camera shake and ensured consistent framing across different samples.

Approximately over 30 images of FHB-infected wheat ears were collected for each test variety. After image acquisition, all images are manually reviewed. Those that are blurry, overexposed, severely occluded or have frame differences are excluded as they are not suitable for analysis. In addition, overly repetitive images are also removed. A total of 300 most representative images of different varieties, different times and different degrees of disease were retained, forming the basic dataset for subsequent digital image analysis and model training. In addition, 200 images were collected in the same way in another experimental field about 10 km away to construct the final validation dataset for FHB detection, to ensure that there are environmental and variety differences from the base dataset.

2.3. Wheat Ear Segmentation

This study employed three distinct deep learning models—Depth Anything V2, DeepLabv3+, and Mask R-CNN—to segment FHB-infected wheat spikes and eliminate background interference. Depth Anything V2 is a depth estimation model that performs segmentation based on reconstructed depth information, whereas DeepLabv3+ and Mask R-CNN are computer vision models that segment wheat spikes by directly identifying them within RGB images.

2.3.1. Depth Estimation Model

Depth Anything V2 is a high-performance monocular depth estimation model designed to infer the 3D structure of a scene from a single image. The model is trained using a large corpus of synthetic images to develop a DINOv2-G-based teacher model, which in turn supervises the student model through large-scale pseudo-labeled real-world images as a bridging dataset (Figure 2). This training strategy enables strong generalization, allowing the model to perform robustly even when applied to plants with complex and variable morphological structures [19].

For robust wheat ear segmentation under challenging field conditions (e.g., occlusion, varying illumination), we used the pre-trained Depth-Anything-V2 model without fine-tuning. The model produces a depth map in which pixel intensity reflects the estimated distance from the camera. The optimal binarization threshold was empirically set to 50% of the maximum depth value, determined through a grid search over a validation set

with thresholds ranging from 30% to 70%. Performance was evaluated using the Dice coefficient and false positive rate. The 50% threshold provided the best balance, minimizing both background inclusion (common at lower thresholds) and mask erosion (prevalent at higher thresholds), thereby maximizing overall segmentation accuracy. The resulting binary mask often contained non-target structures such as awns and stems. To address this, a morphological opening operation was applied using a circular structuring element with a radius of 15 pixels, selected as the optimal value from a tested range of 5 to 25 pixels based on a parameter sensitivity analysis. This radius was found to be the smallest size capable of effectively removing most awns while causing negligible erosion to the central region of the wheat ear. Smaller radii were ineffective in awn removal, whereas larger radii led to noticeable distortion of the ear morphology. Finally, connected component analysis was employed to filter out any residual noise.

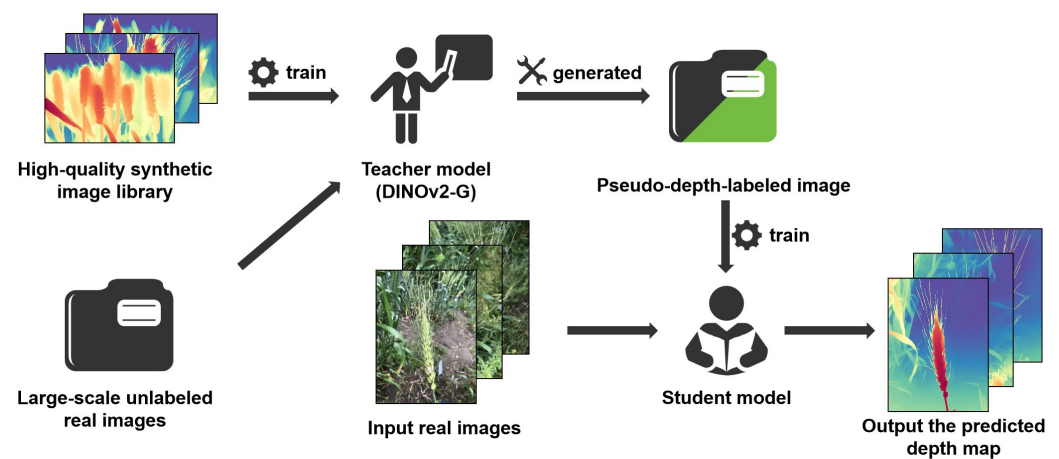


Figure 2. The overall framework of Depth Anything V2.

However, for wider stems, excessive morphological erosion and dilation can lead to the loss of critical edge information from the wheat spike. To address this, a stem localization method based on abrupt changes in width is proposed. The procedure is as follows (implemented in MATLAB R2024a/Python 3.8):

- (1) **Edge scanning:** The spike mask is scanned row by row from the edge near the base of the spike toward the opposite side, recording the mask width as the number of white pixels in each row.
- (2) **Abrupt change detection:** A threshold for the rate of width change is defined, and a sudden increase in row width is used to identify the junction between the spike and the stem.
- (3) **Precise cropping:** The mask on the spike side is retained, while the stem side is masked with black pixels to remove the stem region.

2.3.2. Semantic Segmentation Model

We employ two benchmark models: DeepLabv3+ for semantic segmentation and Mask R-CNN for instance segmentation. The core of DeepLabv3+ lies in its atrous convolution and ASPP module, which capture multi-scale contextual information for accurate pixel-level classification [20]. Mask R-CNN, whereas, introduces RoIAlign to preserve spatial precision, allowing it to not only classify pixels but also separate individual objects, such as distinct wheat spikes in a cluster, proving superior in handling occlusion [21].

To train these two models, we divided the base dataset into a training set (50%), a validation set (25%), and a test set (25%), ensuring that the same variety only appeared in the same type of dataset. We also performed data augmentation on each part, including

operations such as rotation ($\pm 90^\circ$), horizontal and vertical flipping, and Gaussian blur, to expand the dataset. Throughout the entire annotation process, the consistent definition of wheat ears was strictly followed: only the ear body was marked, and the awns and stem segments were clearly excluded. According to this criterion, the wheat ear area was manually labeled using LabelMe software to generate polygonal masks, which were then used to train the DeepLabV3+ and Mask R-CNN networks. During the training process, adjust hyperparameters such as the learning rate, batch size, and the number of iteration rounds. In the prediction stage, the raw segmentation probability maps generated by the trained model often exhibit coarse boundaries. To address this, we introduced a post-processing step incorporating Conditional Random Field (CRF) and Graph Cut (GC). This step does not alter the model's internal parameters but operates on its output. The CRF model leverages color and spatial consistency to refine pixel-level labels, while the GC algorithm globally optimizes the boundary placement. This combination effectively enhances the continuity of segmentation masks and the accuracy of detail representation, transforming the model's probabilistic outputs into sharp, high-quality final results. Table 1 provides detailed parameters and Settings for model training and inference.

Table 1. Configurations, parameters and tools used in the training of the model in this study.

Project	Depth Anything V2	DeepLabv3+	Mask R-CNN
GPU		NVIDIA GeForce RTX 4070	
Operating System		Windows 11	
Data Processing		Python 3.8, Labelme and Matlab R2024a	
Deep Learning Framework		PyTorch 2.9.1	
Backbone or Pretrained Weights	Depth_Anything_v2_vits	Resnet50	Resnet50
Num Classes	N/A	2	2
Batch Size	N/A	8	2
Learning Rate	N/A	1e-4	0.005
Loss Function	N/A	Cross-Entropy Loss	Cross-Entropy + Smooth L1 Loss

2.4. Color Feature Extraction

FHB-infected spikelets are characterized by water-soaked lesions that appear yellow in the early stages and later turn brown, exhibiting a distinct color contrast compared to healthy spikelets [22]. Drawing on prior research [23,24], we compared infected and healthy spikelets across the RGB color space, the Saturation and Value components of the HSV color space, the CIELAB (LAB) color space, and three color indices: Excess Green–Excess Red difference (ExGR) [25], Normalized Difference Index (NDI) [26], and Visible Atmospherically Resistant Index (VARI) [27]. To enhance discriminatory power, a fused index was developed by integrating ExGR with the Lab-a* component. The formulas used to calculate each color index are as follows:

$$ExGR = 3G - 2.4R - B \quad (1)$$

$$NDI = \frac{R - G}{R + G} \quad (2)$$

$$VARI = \frac{G - R}{G + R - B} \quad (3)$$

where R , G , and B denote the digital number (DN) values of the red, green, and blue channels in the image.

2.4.1. HSV Color Space Conversion

In color-based visual perception tasks, the HSV color space is widely adopted for its capacity to decouple color attributes, making it more consistent with the cognitive mechanisms of the human visual system and markedly distinct from the conventional RGB model [28]. The conversion formulas from RGB to HSV are as follows:

$$M = \max(r, g, b), m = \min(r, g, b) \quad (4)$$

$$V = M \quad (5)$$

$$S = \begin{cases} 0 & M = 0, \\ \frac{M-m}{M} & M \neq 0. \end{cases} \quad (6)$$

$$H = \begin{cases} 0 & M = m, \\ 60^\circ \times \left(\frac{G-B}{M-m} \right) & M = R, \\ 60^\circ \times \left(\frac{B-R}{M-m} + 2 \right) & M = G, \\ 60^\circ \times \left(\frac{R-G}{M-m} + 4 \right) & M = B. \end{cases} \quad (7)$$

where r , g , and b represent the normalized red, green, and blue components, each ranging from 0 to 1; S and V correspond to saturation and value (brightness), also within the range [0, 1]; H denotes hue, with values ranging from 0 to 360° .

2.4.2. LAB Color Space Conversion

LAB is a color model grounded in human visual perception and provides distinct advantages for tasks such as precise color quantification, cross-device color consistency, and visual quality assessment. Direct conversion from the RGB color space to LAB is not possible; instead, the transformation must proceed through the intermediate XYZ color space—first from RGB to XYZ, followed by conversion from XYZ to LAB [29]. The corresponding conversion formulas are as follows:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (8)$$

$$L^* = 116f\left(\frac{Y}{Y_n}\right) - 16 \quad (9)$$

$$a^* = 500 \left[f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right) \right] \quad (10)$$

$$b^* = 200 \left[f\left(\frac{X}{X_n}\right) - f\left(\frac{Z}{Z_n}\right) \right] \quad (11)$$

where L^* denotes lightness, with values ranging from 0 to 100; a^* represents the green–red axis, ranging from -128 to 127 ; and b^* represents the blue–yellow axis, also ranging from -128 to 127 . X_n , Y_n , and Z_n represent the reference white values for the respective parameters.

2.4.3. Dynamic Index Fusion

To balance visual distinguishability and segmentation accuracy, this study employs a dynamic index fusion strategy. Based on the severity of infection—manifested as variations in red and yellow pixel intensities within the wheat spike region—the weights of different

indices are dynamically adjusted to minimize segmentation error. The specific weighting formula is as follows:

$$R_{ry} = \frac{1}{N_f} \sum_{(x,y) \in F} I(H(x,y) < 0.2) \quad (12)$$

$$w_a = 0.5 + \alpha(2R_{ry} - 1) \quad (13)$$

$$w_g = 1 - w_a \quad (14)$$

where R_{ry} denotes the proportion of red–yellow pixels; N_f is the total number of pixels within the wheat spike region; F represents the set of pixel coordinates in the spike region; $H(x, y)$ is the HSV hue value of pixel (x, y) , normalized to the range $[0, 1]$; I is the indicator function (equal to 1 if the condition is met, and 0 otherwise); w_a is the fusion weight assigned to the Lab-a* value; w_g is the fusion weight for the inversely normalized ExGR index; and α is the parameter controlling the range of weight variation.

2.5. Evaluation Metrics

2.5.1. Wheat Spike Segmentation

To assess the performance of wheat spike extraction, six evaluation metrics were employed: Accuracy, Precision, Recall, Specificity, Intersection over Union (IoU), and F1 Score. All metrics were calculated at the pixel level. The corresponding formulas are as follows:

$$Accuracy = (TP + TN) / (TP + FP + FN + TN) \quad (15)$$

$$Precision = TP / (TP + FP) \quad (16)$$

$$Recall = TP / (TP + FN) \quad (17)$$

$$Specificity = TN / (TN + FP) \quad (18)$$

$$IoU = TP / (TP + FP + FN) \quad (19)$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (20)$$

where True Positive (TP) denotes the number of wheat spike pixels correctly classified as spike; False Negative (FN) refers to the number of spike pixels incorrectly classified as background; False Positive (FP) represents the number of background pixels incorrectly classified as spike; and True Negative (TN) indicates the number of background pixels correctly classified as background.

2.5.2. Wheat FHB Detection

In addition, R^2 , RMSE, and Cohen's Kappa (k) coefficient were used to evaluate the accuracy of infected area extraction. The corresponding formulas for each metric are as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (21)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (22)$$

$$k = \frac{(Po - Pe)}{(1 - Pe)} \quad (23)$$

where n represents the number of samples; y_i is the true value of the i -th sample; $\hat{y}_i$ is the predicted value of the i -th sample; and $\bar{y}_i$ is the mean of all true values. Observed Agreement (Po) denotes the proportion of samples for which the predicted severity level exactly matches the true severity level. Expected Agreement by Chance (Pe) refers to the

proportion of agreement expected by random chance, calculated based on the marginal distributions of severity levels in both the ground truth and the predicted labels.

2.6. Severity Calculation

Severity is a commonly used metric for quantifying the extent of disease occurrence. In this study, it is defined as the ratio of the infected spikelet area to the total spikelet area. The corresponding calculation formula is as follows:

$$S = \frac{A_I}{A_H + A_I} \quad (24)$$

where S denotes the severity index for an individual wheat spike; A_I represents the number of pixels corresponding to infected spikelets; and A_H represents the number of pixels corresponding to healthy spikelets.

Severity grading is commonly used to assess the extent of disease development. In this study, grading criteria were quantitatively established by examining the size of infected spikelets and entire spikes across different wheat varieties under varying levels of infection. All samples were classified into five severity levels (0–4) according to the Chinese national standard GB/T 15796-2011 Technical Specification “<https://www.sdtdata.com/fx/fmoa/tsLibCard/125622.html> (accessed on 15 May 2024)” for the Monitoring and Forecasting of FHB in Wheat. The severity levels and corresponding symptom descriptions are presented in Table 2.

Table 2. Severity Grading of FHB in Wheat and Corresponding Symptom Descriptions.

Grade	Symptom Description	Severity
0	No spikelets infected.	0
1	Infected spikelets account for less than 1/4 of the total.	$0 < S \leq 0.25$
2	Infected spikelets account for 1/4 to 1/2 of the total.	$0.25 < S \leq 0.50$
3	Infected spikelets account for 1/2 to 3/4 of the total.	$0.50 < S \leq 0.75$
4	Infected spikelets account for more than 3/4 of the total.	$S > 0.75$

3. Results

3.1. Wheat Spike Segmentation Results Using the Depth Estimation Model

The final segmentation results obtained using Depth Anything V2 are presented in Figure 3. Initially, depth-based segmentation effectively localized the main structure of the wheat spike and significantly suppressed interference from complex field backgrounds. However, due to the presence of awns and stems, segmentation accuracy was limited, with an average Precision of only 0.77. To address awn interference, morphological opening combined with connected component analysis was applied, successfully removing most awns while preserving the spike’s edge features (Figure 3A). To further eliminate stem remnants, a stem localization method based on abrupt width variation was introduced. By identifying local discontinuities in width at the spike–stem junction (Figure 3B), redundant stem regions were accurately separated. Compared with conventional morphological approaches (e.g., global erosion), this method avoids manipulating the entire spike structure and preserves the integrity of grain features, thereby yielding high-fidelity segmentation results suitable for downstream phenotypic analysis.

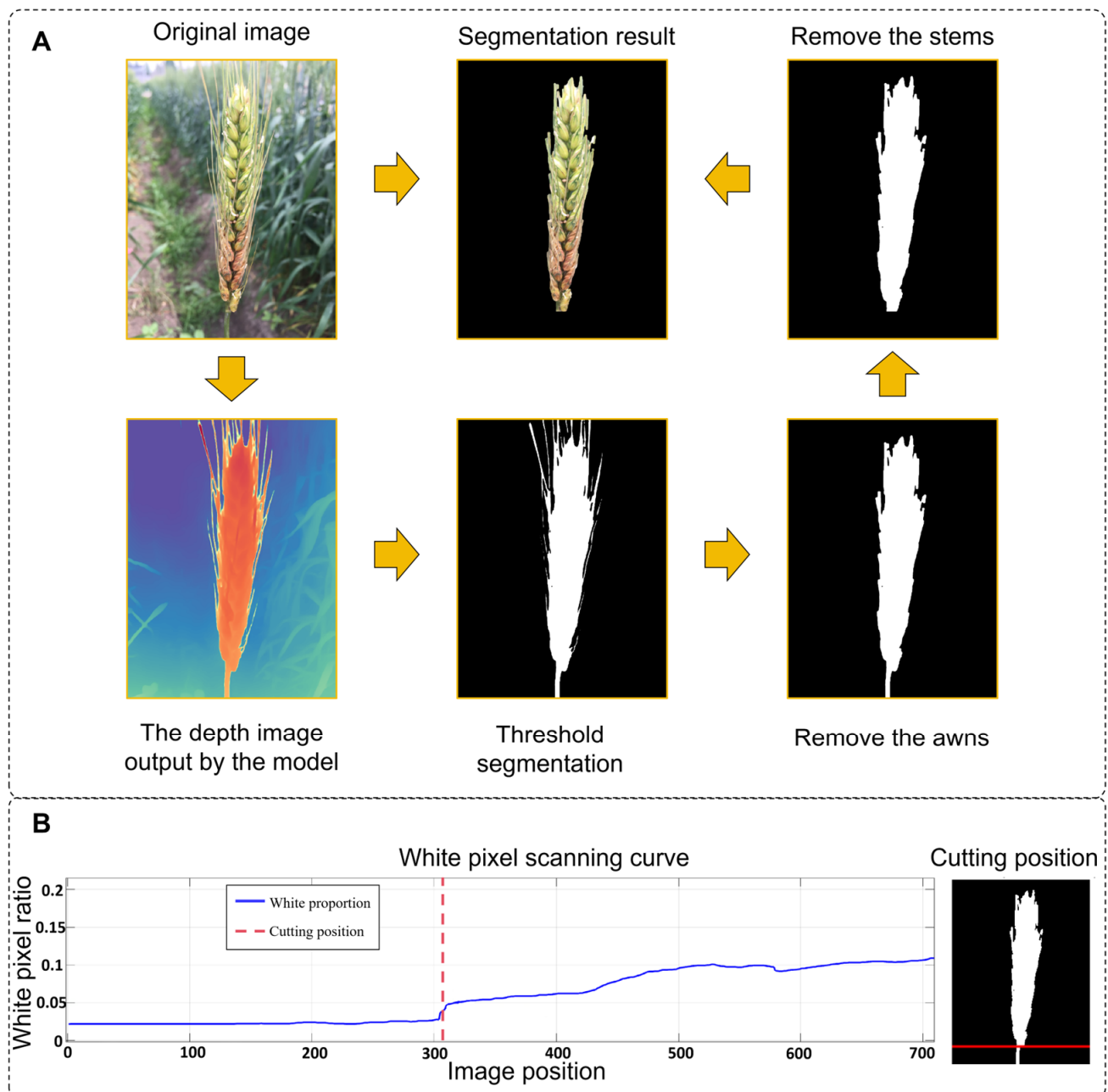


Figure 3. Depth image processing. Note, (A) shows the depth image processing, (B) shows the stem removal.

To assess the performance of the final segmentation results, we conducted a quantitative analysis using Precision–Recall (PR) curves and the Dice similarity coefficient (Figure 4). The PR results indicate that the method achieves an excellent balance between precision and recall, with the majority of sample points clustered in the upper-right region of the PR space (Precision > 0.85, Recall > 0.85). Notably, more than 90% of the samples attained an F1-score exceeding 0.9, demonstrating the model’s stable and reliable performance across wheat spikes with varying morphologies. In terms of segmentation accuracy, the Dice coefficient exhibited a right-skewed distribution, with a median of 0.93 and a maximum value reaching 0.98. These quantitative metrics collectively demonstrate the high reliability and precision of the proposed method in wheat spike segmentation, fulfilling the stringent accuracy requirements of practical agricultural applications.

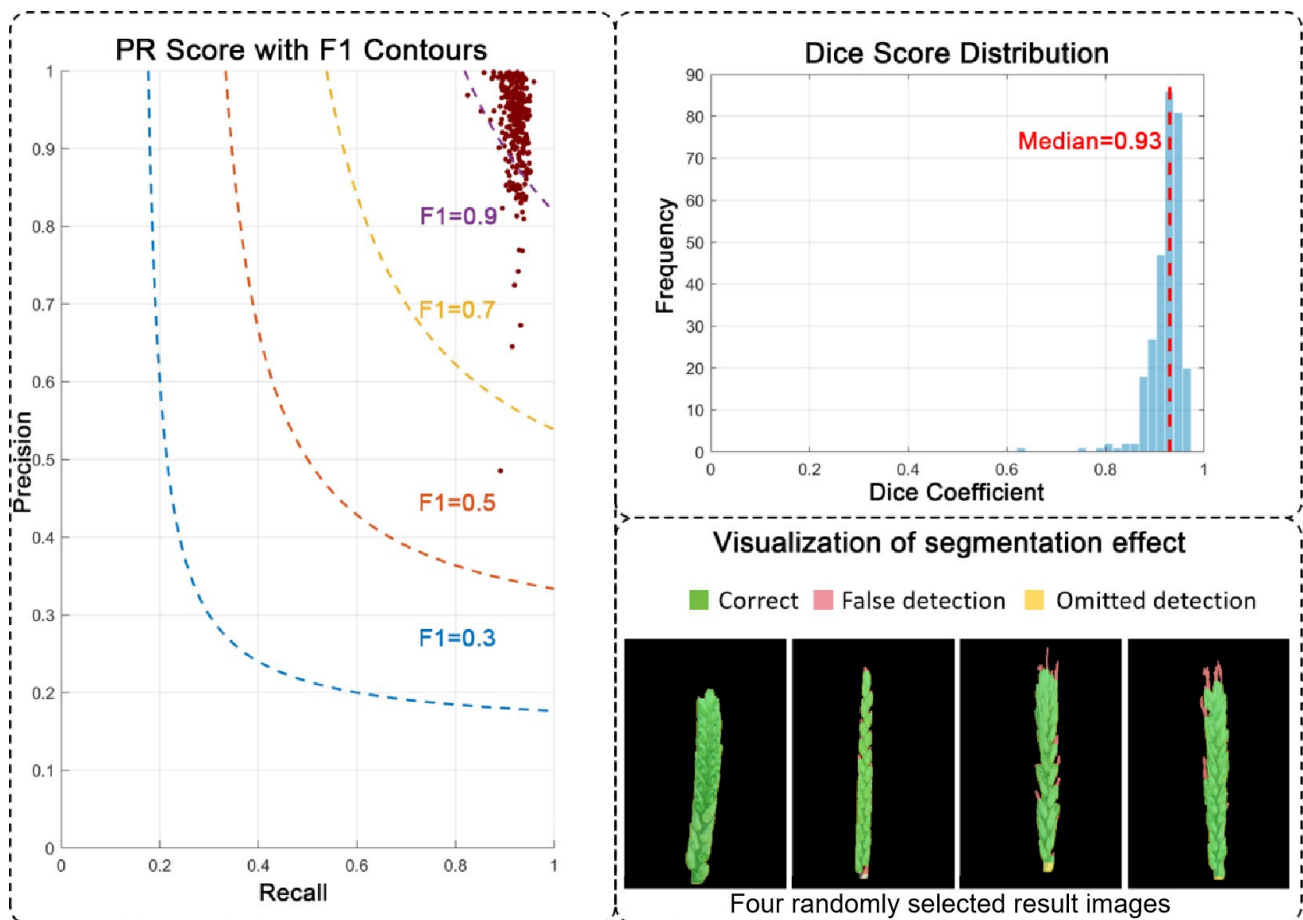


Figure 4. PR score, dice value and visualization results (The red dots represent the PR value of each sample).

3.2. Comparison of Wheat Spike Segmentation Performance

The segmentation performance of three deep learning models was evaluated under varying field conditions and Fusarium head blight severity levels. As shown in Figure 5, Depth Anything V2 achieved the most accurate results, leveraging the inherent robustness of depth information to minimize interference from complex backgrounds such as soil, weeds, and sky. This model effectively suppressed background noise while preserving wheat spike structures. By contrast, both DeepLabv3+ and Mask R-CNN were susceptible to background interference, leading to reduced segmentation precision. Mask R-CNN in particular was prone to false positives due to texture confusion, whereas DeepLabv3+ exhibited a higher rate of false negatives.

A detailed comparison of model segmentation performance is presented in Table 3. All three deep learning models achieved an Accuracy exceeding 0.95. Among them, Depth Anything V2 consistently outperformed the others across all segmentation accuracy metrics, achieving a Precision of 0.945, Recall of 0.926, IoU of 0.878, and F1 Score of 0.935. These values were significantly higher than those of both Mask R-CNN and DeepLabv3+, indicating superior segmentation capability. Although Depth Anything V2 recorded the highest total processing time (2.761 s), which includes essential post-processing steps to generate the final segmentation, this rate still satisfies the requirements for real-time application in many agricultural scenarios where a processing speed of several seconds per image is acceptable. In contrast, Mask R-CNN showed strong Recall (0.940), but its lower Precision (0.825) suggests a higher rate of false positives, and its inference time (1.315 s) offers no speed advantage over Depth Anything V2. DeepLabv3+ was the fastest

(0.152 s) but exhibited the weakest overall performance, particularly in IoU (0.603) and F1 Score (0.737). The results demonstrate a clear trade-off: Depth Anything V2 offers markedly superior accuracy and robustness, making it the most suitable model for applications where segmentation quality is the primary concern, while its processing speed remains within a practical range for real-time use.

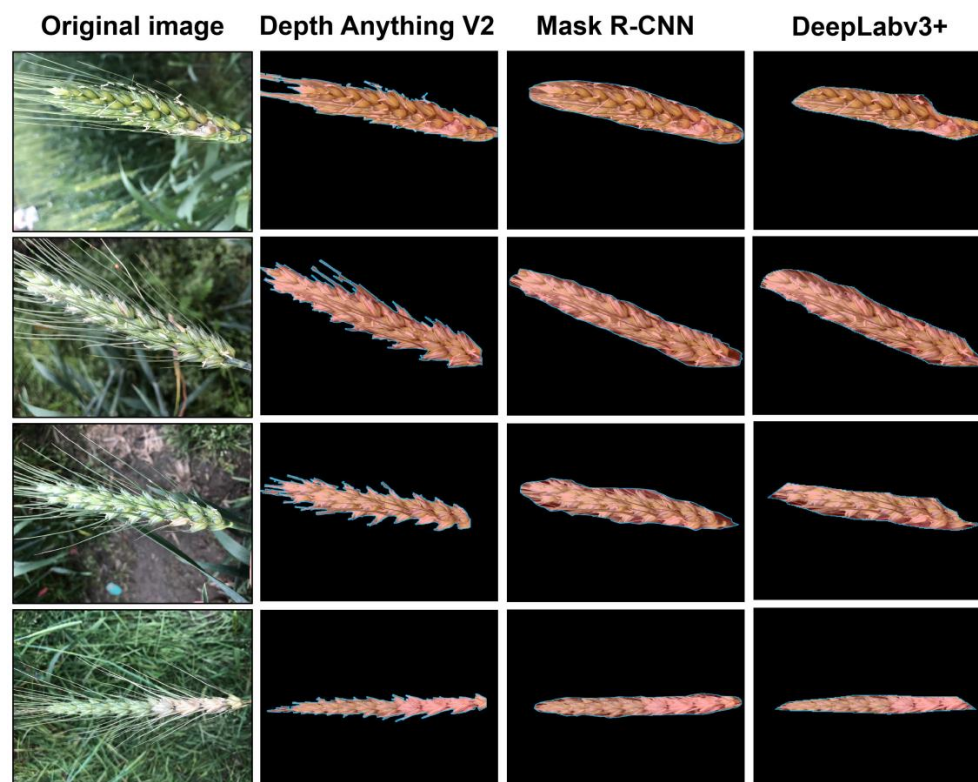


Figure 5. The segmentation effects of different models on four randomly selected images.

Table 3. Results of wheat ear segmentation by different methods.

Method	Accuracy	Precision	Recall	IoU	F1	Time (s)
Depth Anything V2	0.987	0.945	0.926	0.878	0.935	2.761
Mask R-CNN	0.975	0.825	0.940	0.784	0.877	1.315
DeepLabv3+	0.958	0.782	0.702	0.603	0.737	0.152

Note, The time consumption of the Depth Anything V2 model consists of two parts: model reasoning and subsequent processing.

3.3. Segmentation Results of Infected Spikelets

Using depth-estimated wheat spike images segmented by the Depth Anything V2 model, we analyzed the visual characteristics of FHB across different color indices (Figure 6A). The results revealed a pronounced distinction between healthy and infected spikelet regions, particularly in the Lab-a* component and the ExGR index. Lab-a* exhibited high sensitivity to yellow-to-brown discoloration associated with FHB lesions, while ExGR was more responsive to the green tones of healthy tissue—demonstrating a complementary relationship between the two indices.

To leverage this complementarity, we assigned severity-dependent dynamic weights to each index and performed index fusion. The resulting composite index enhanced the visual contrast between symptomatic and healthy areas and effectively reduced transitional zones at lesion boundaries (Figure 6B).

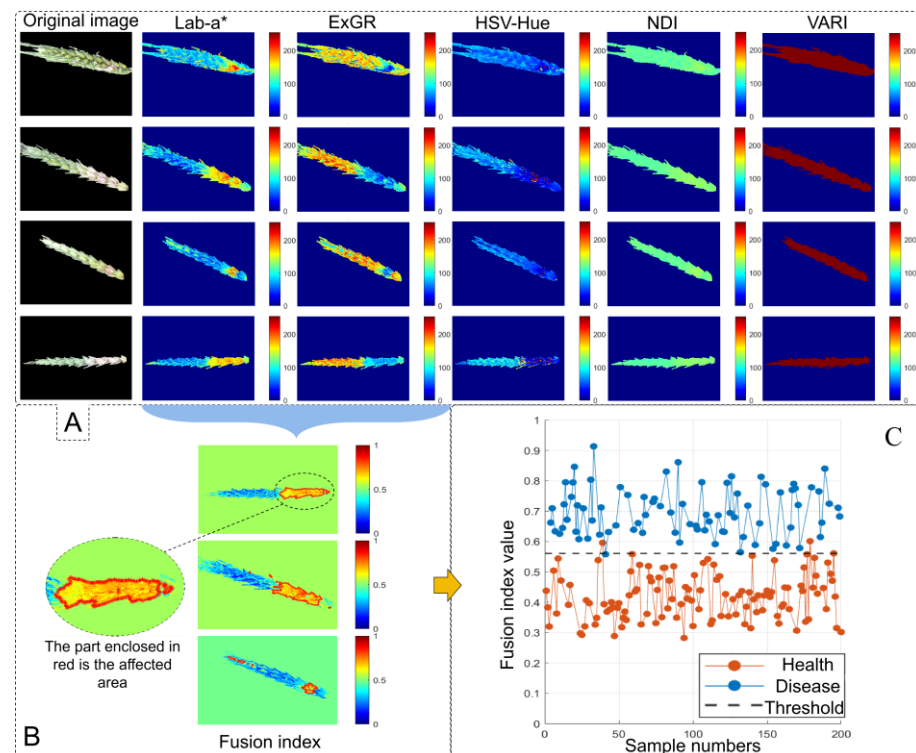


Figure 6. The segmentation result of spikelets. Note, (A) shows the different color indices (Lab-a* represents the a* channel in the LAB color space), (B) shows the exponential fusion result, (C) shows the 200 fusion index samples.

3.4. Ablation Experiment for the Segmentation Method of Diseased Spikelets

To verify the effectiveness and superiority of the proposed fusion index and dynamic weight strategy in wheat ear segmentation and scab detection, we conducted an ablation experiment on the research method in this paper. The experimental results (Figure 7) clearly demonstrate a progressive trend of performance improvement: from a single index (Methods I and II) to fixed weight fusion (Method III), and then to the dynamic weight fusion method of this study (Method IV), the performance metrics (R^2 and RMSE) of the model have all achieved significant improvements. This trend not only demonstrates the feasibility of deeply guided feature fusion, but also highlights the necessity of introducing dynamic adaptive mechanisms. Firstly, the performance of a single exponential model (EXGR or Lab-a*) is relatively limited, mainly due to its inherent limitations. In the complex field environment, changes in light, shading, and the continuous variations in the color of the wheat ears themselves make it difficult for any single feature to stably distinguish between healthy and diseased tissues. For instance, the EXGR index is sensitive to green and healthy tissues, but is susceptible to interference from bright backgrounds. Although the Lab-a* index can better capture the color changes in lesions, it is prone to misjudgment within healthy tissues. Therefore, the performance bottleneck of the method lies in the singularity and fragility of its feature representation ability. Secondly, the significant improvement in the performance of the fixed-weight fusion model (Method III) strongly demonstrates the effectiveness of multi-feature complementarity. By combining the healthy tissue highlighting ability of EXGR and the lesion tissue recognition ability of Lab-a*, the model has obtained a more robust feature representation, thereby being able to better cope with the complexity of the field environment. However, fixed-weight fusion implies a strong assumption that the importance of different features remains constant across the entire dataset. This is clearly inconsistent with the huge intra-class differences in agricultural scenarios (such as different varieties, growth stages, and light conditions). Ultimately, the

core contribution of the highest performance achieved by this method (Method IV) lies in breaking through the limitation of fixed weights and introducing a dynamic weight adaptive mechanism. For instance, in a flat area with abundant sunlight, a higher weight can be assigned to the EXGR index to precisely capture the distribution of chlorophyll. This perceptual ability enables this method to intelligently adjust the contribution of different color indices according to the environment in which each wheat ear is located, thereby achieving more precise and accurate segmentation and recognition.

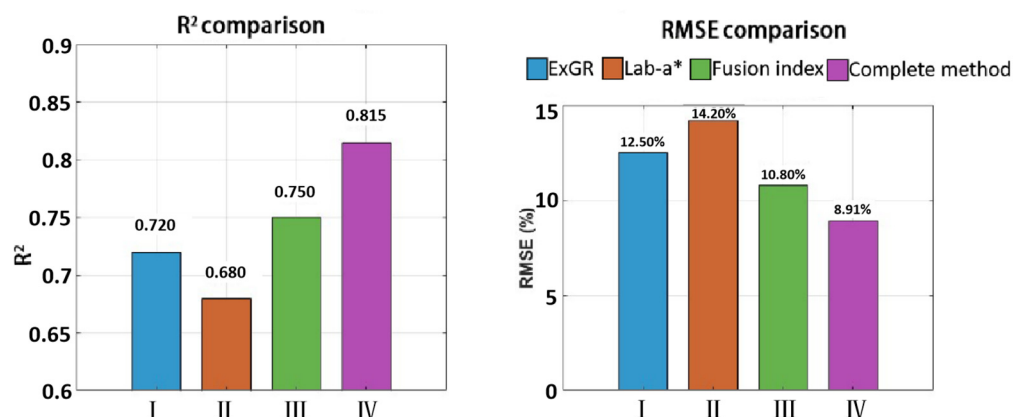


Figure 7. Results of the ablation experiment.

3.5. Infection Rate and Severity Level Calculation Results

The final infection rates calculated for individual wheat spikes were compared against manually measured ground truth values, as illustrated in Figure 8. Detection based on the fused index yielded an R^2 of 0.815 and an RMSE of 8.91%, indicating high accuracy. Further analysis of the prediction error distribution showed that over 75% of the samples fell within a $\pm 10\%$ error range, demonstrating that the proposed fused index provides a reliable estimate of FHB severity.

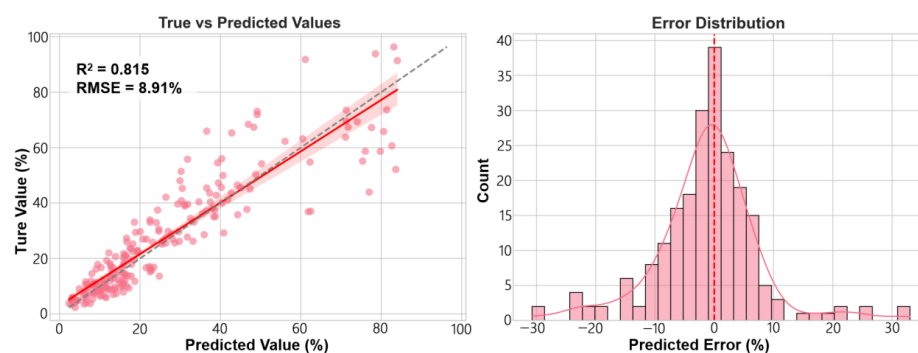


Figure 8. Evaluation of incidence rate results (the curves in the bar chart are trend lines, showing the upward, downward or stable trends of the data).

A detailed classification analysis was conducted across samples with varying levels of disease severity. The proposed detection method demonstrated strong overall performance, achieving a classification accuracy of 89.06% and a Cohen's Kappa coefficient of 0.8334, indicating a high degree of agreement between model predictions and ground truth annotations. Confusion matrix analysis (Figure 9) revealed that diagonal entries accounted for 85.5% of all predictions, while most off-diagonal errors occurred between adjacent severity levels, suggesting the model is effective in limiting misclassifications across non-adjacent grades. Performance varied across severity categories. The model achieved perfect recall (100%) for severity level 1, indicating exceptional detection at mild

infection stages. For level 2, the model demonstrated high precision (93.3%) but was more conservative, resulting in a higher rate of missed detections. As disease severity increased, performance declined, with level 4 showing the lowest precision (73.7%) and level 3 the lowest recall (79.1%). Overall, the coefficients of variation for precision and recall were 8.59% and 9.89%, respectively, reflecting stable classification performance across severity levels. However, further optimization is needed to improve detection accuracy at higher severity stages.

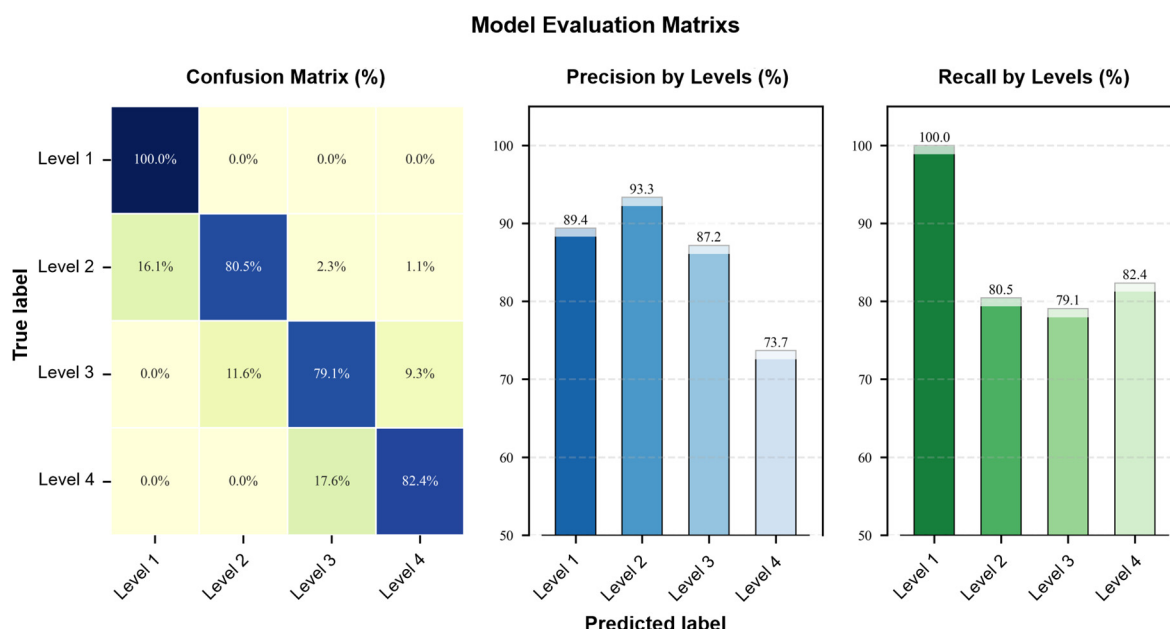


Figure 9. Comparison of different categories.

3.6. External Validation

To evaluate the generalization ability and robustness of the framework proposed in this study, we conducted validation on the validation dataset. Its performance is shown in Table 4. Compared with the results of the base dataset, this study method only observed minor performance degradation in all key indicators. Specifically, the IoU of wheat ear segmentation slightly decreased from 0.878 to 0.862, while the R^2 for estimating the severity of scab slightly decreased from 0.815 to a high level of 0.798. This result indicates that the method proposed in this study has excellent generalization performance. Despite the introduction of new variety variations and natural conditions in the external validation set, the segmentation core based on monocular depth estimation still maintains a highly sensitive recognition ability for the spatial structure of the main body of the wheat ear. Meanwhile, the dynamic weight adjustment mechanism effectively adapts to the possible subtle differences in color characteristics among different varieties, thereby ensuring the stability of disease identification.

Table 4. Performance comparison between the basic dataset and the validation dataset.

Mission	Evaluation Index	Basic Dataset	Verification Dataset
Wheat ear segmentation	IoU	0.878	0.862
	F1	0.935	0.919
FHB detection	R^2	0.815	0.798
	RMSE	8.91%	9.35%

3.7. Robustness Analysis

In this section, we demonstrate the performance of the research method on different wheat varieties. To evaluate the differential impact of morphological opening on awned and awnless wheat spikes, we systematically varied the radius of the structuring element from 3 to 50 during postprocessing and computed the corresponding IoU values for each configuration. As illustrated in Figure 10, the IoU trends differed markedly between the two spike types. For awned spikes, peak IoU values appeared at larger radii, with an optimal structuring element radius of 37 yielding an IoU of 0.894.

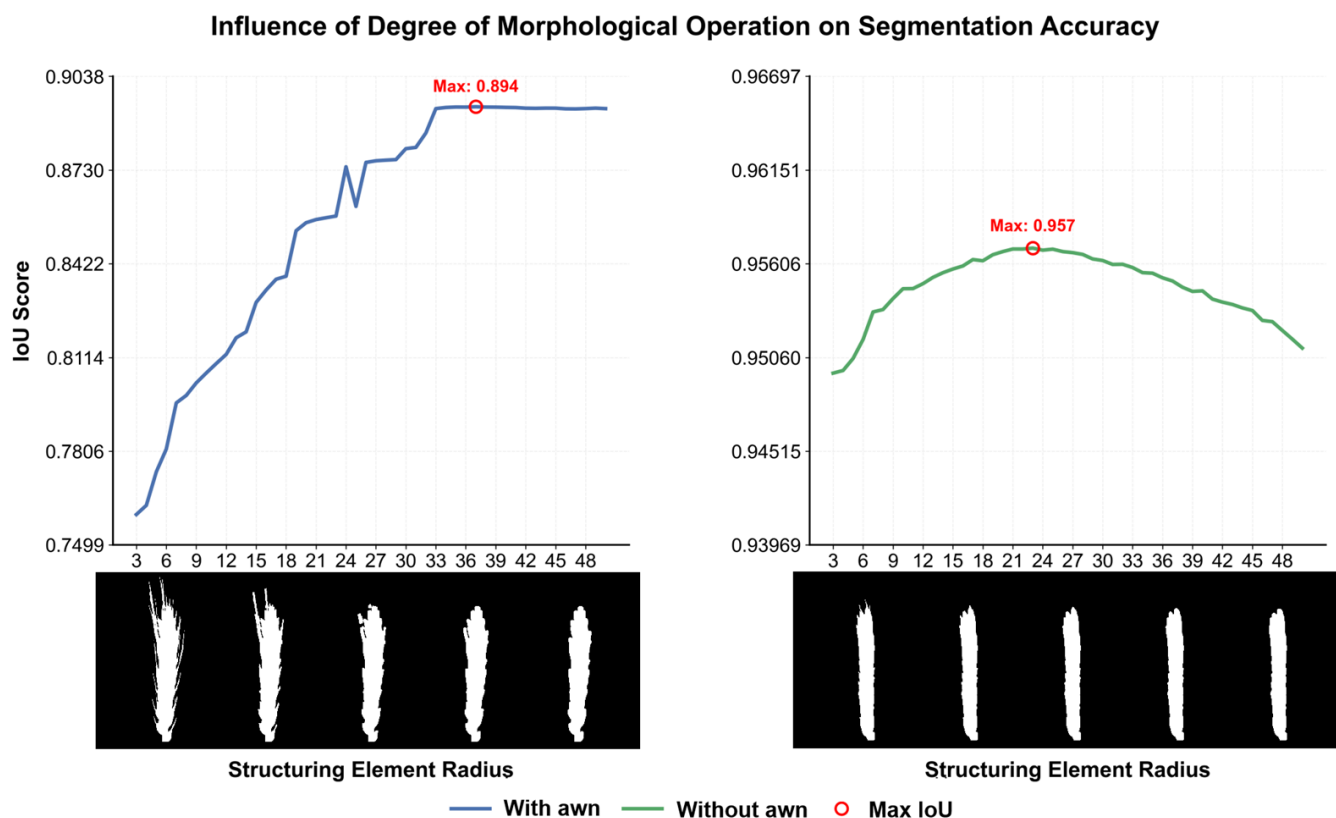


Figure 10. Influence of degree of morphological operation on segmentation accuracy.

In addition, image quality is another critical factor influencing segmentation performance. To evaluate the impact of image degradation—resulting from variations in imaging devices—on segmentation accuracy, we simulated the output of three typical devices (smartphones, low-end cameras, and surveillance cameras) by applying noise, blur, and compression to original images. Segmentation similarity was assessed using Pixel Accuracy (PixelAcc) and IoU, as summarized in Figure 11. Results indicate that decreased image quality leads to reduced segmentation accuracy and increased performance variability. IoU proved more sensitive to degradation, declining by 3.5–9.2%, compared to a 1.2–7.6% drop in PixelAcc. Among the degradation types, Type 3 exhibited the lowest absolute accuracy (PixelAcc: 92.4%, IoU: 91.1%), while Type 2 showed the highest instability, with standard deviations of $\pm 6.5\%$ in PixelAcc and $\pm 9.0\%$ in IoU. Overall, although image degradation diminishes segmentation robustness, its effect remains relatively limited under moderate conditions. Even under severe degradation, segmentation similarity remains above 90%, demonstrating the method's strong cross-platform adaptability and practical utility.

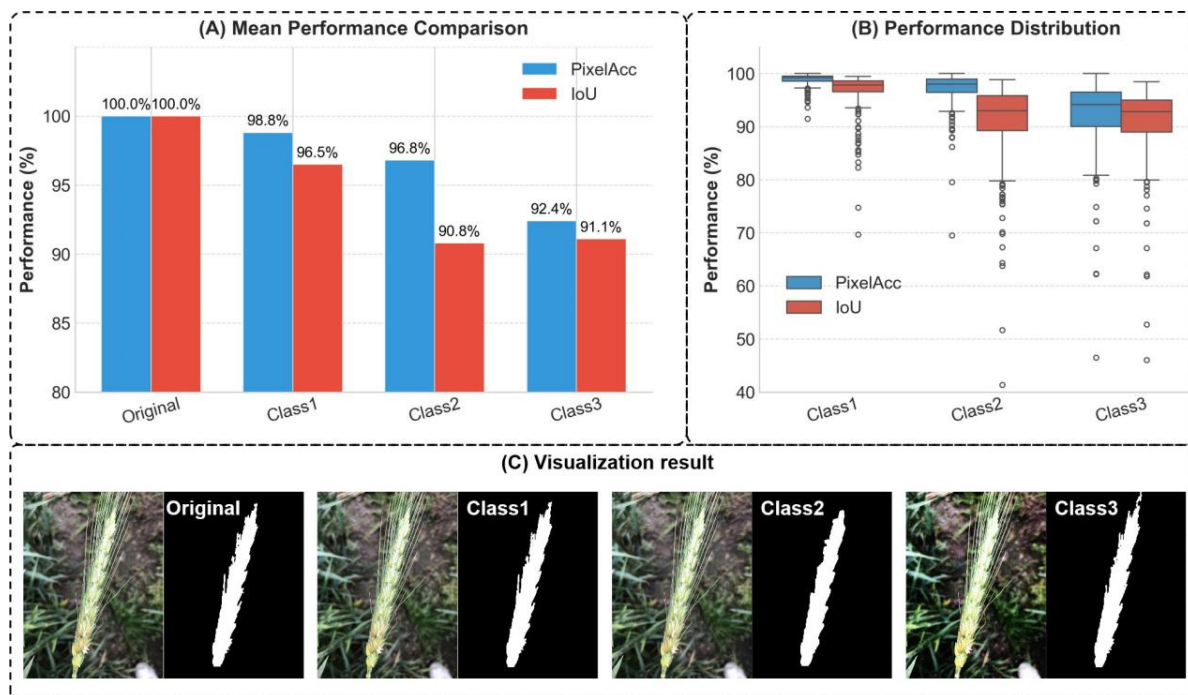


Figure 11. The influence of image quality on the segmentation results of wheat ears. Note, (A) It shows the change in accuracy, (B) shows the change in the stability of the results, (C) shows the visualization results under different treatments.

4. Discussion

4.1. Method Efficacy and Comparative Analysis

This study successfully established a framework integrating monocular depth estimation with dynamic color index fusion, achieving high-precision segmentation of wheat spikes in field conditions ($\text{IoU} = 0.878$) and accurate quantitative assessment of FHB severity ($R^2 = 0.815$), all without the need for annotated data training. Similarly to this study, many existing methods also focus on wheat spike segmentation and model development [30,31]. Zhang et al. [32], based on hyperspectral data from wheat canopies, selected the optimal model from nine candidates and developed two new types of indices combined with model fusion, achieving a wheat FHB detection accuracy of 91.4%. Francesconi et al. [33] utilized UAV-based TIR and RGB image fusion alongside spike physiological data, such as molecular identification of the FHB pathogen, to achieve precise FHB detection in field environments. However, the aforementioned methods, operating under natural field backgrounds, cannot extract information from individual spikes and can only obtain the disease incidence rate at the population level. Wang et al. [34] used structured light scanners mounted on field mobile platforms to acquire point cloud data of wheat spikes and employed an adaptive k-means algorithm with dynamic perspective and Random Sample Consensus (RANSAC) algorithms to calculate 3D data of the spikes. This approach is similar to ours in utilizing three-dimensional data for analyzing spike morphology. Nonetheless, the complexity of acquiring point cloud data may significantly limit its application. Qiu et al. [35] used a Mask R-CNN model to segment collected wheat spike images and employed a new color index to identify diseased spikelets. Similarly, Liu et al. [4] utilized an improved DeepLabv3+ model to segment RGB images and applied dynamic thresholds based on color indices for segmenting diseased spikelets. The rationale behind these methods aligns with that of our study, which involves first segmenting the wheat spikes and then detecting diseases based on the color differences between diseased and healthy spikelets. However, semantic segmentation models typically require large

amounts of data for training to achieve satisfactory performance. The method proposed in this study addresses or circumvents the inherent bottlenecks of traditional approaches. Its convenient and efficient application strikes an excellent balance between accuracy and cost-effectiveness.

4.2. Practical Considerations and Robustness

Although the image processing and deep learning framework proposed in this study offers clear advantages in the high-throughput detection of FHB in wheat, its effectiveness is challenged under real-world field conditions. Factors such as canopy occlusion within wheat populations, weed interference, and variable illumination significantly affect detection performance [36–38]. Additionally, variability in disease severity, cultivar-specific traits, and morphological differences among wheat varieties constitute critical sources of uncertainty [39].

Robustness analysis indicates that our method's uniform morphological post-processing for both awned and awnless wheat spikes imposes a non-negligible constraint on segmentation accuracy. The fundamental morphological differences between these two types mean that a single parameter set is suboptimal for both, which is a primary factor contributing to the performance gap. Consequently, implementing an adaptive post-processing mechanism that automatically tailors parameters based on spike morphological features (e.g., awn presence and density) presents a promising avenue for significantly enhancing accuracy. Although the difference in image accuracy does not affect the effectiveness of this scheme, it is still a factor restricting the final accuracy. To achieve the leap from “effective” to “precise”, subsequent research will focus on developing corresponding online calibration algorithms [40,41]. This algorithm can perform real-time error correction on the segmentation results based on the imaging characteristics of the equipment and the degree of scene degradation, thereby significantly enhancing the accuracy and consistency of this method in different practical scenarios.

In addition, illumination is a critical factor influencing the visual recognition accuracy of FHB in wheat. This study specifically examined the effects of strong direct sunlight and found that under such conditions, wheat spikes often exhibit significant overexposure, leading to whitening and color distortion of surface features [42]. This effect is particularly pronounced in diseased areas. The adverse impact of intense lighting is primarily observed in two aspects: (1) Strong illumination obscures the characteristic pink mold layer associated with FHB, thereby diminishing the color contrast between healthy and infected tissues; and (2) excessive brightness blurs lesion boundaries, reducing overall recognizability. These effects contribute to the decreased accuracy of color index fusion methods under higher infection severity levels. The promising results of GAN-based approaches in mitigating similar issues in agricultural imagery, such as preserving detail under varying light and spectrum [43] strongly support our proposal for future work. Specifically, integrating a learned illumination compensation module, inspired by these advances, represents a highly promising direction to enhance the resilience of our FHB detection framework against harsh sunlight conditions.

4.3. Innovative Contributions and Future Prospects

The fundamental innovation of this study lies in the establishment of a new crop phenotypic model, which is characterized by low implementation cost, minimal hardware dependence and high operational efficiency. This breakthrough was systematically demonstrated through three key contributions: methodologically, it was the first to comprehensively apply advanced monocular depth estimation to wheat ear segmentation, effectively replacing expensive stereo vision systems while maintaining high precision;

Technically, a dynamic weighting mechanism for perceiving the degree of infection has been incorporated to achieve intelligent fusion of image color features, significantly enhancing environmental adaptability. Practically, it offers a quasi-plug-and-play solution, significantly reducing the technical barriers to on-site deployment. It is particularly worth noting that the proposed technical framework integrates “depth-guided segmentation, geometrically optimized morphology and dynamic fusion enhanced recognition”, demonstrating significant portability in various agricultural scenarios.

Looking ahead, we plan to further integrate the method proposed in this paper into the local system of AR glasses to achieve real-time, end-to-end crop monitoring and recognition capabilities, and further enhance the response speed and practicality of the system in field applications. In addition, our research agenda will focus on four strategic directions to advance this technology: first, developing deep learning-powered adaptive morphological processing algorithms to improve generalization capability across diverse spike architectures; second, constructing illumination-invariant feature representation modules to ensure robust performance under varying field conditions; Third, developing a more refined dedicated model for the segmentation of wheat scab to enhance the environmental adaptability of this scheme. Finally, through the technical path of multi-platform integration, explore an automated acquisition scheme suitable for the detection of grain size in ears. Through these coordinated innovations, we anticipate providing an integrated technical solution for achieving unmanned, intelligent, and precision-oriented crop disease monitoring in modern agriculture, thereby contributing substantial technical support for global food security initiatives.

5. Conclusions

This study presents a novel approach for detecting FHB in wheat by integrating monocular depth estimation with a fused color index. In contrast to previous methods that rely on semantic segmentation models, the proposed method does not need to create a dataset for model training. It achieved a segmentation precision of 0.945 and an IoU of 0.878 for wheat spikes, along with an R^2 of 0.815 in disease lesion quantification. The proposed framework not only offers a new perspective for disease detection in wheat but also provides a transferable reference for segmentation tasks in other crop phenotyping applications.

Author Contributions: J.W. (Jiacheng Wang): methodology, resources, writing—original draft. J.W. (Jianliang Wang): methodology, software, writing—review and editing. Y.Z.: formal analysis, investigation. F.W.: formal analysis, software, methodology. W.W.: investigation, validation. Z.L.: funding acquisition, resources, supervision. C.S.: funding acquisition, writing—review and editing. T.L. (Tao Li): software, methodology. Corresponding Author. T.L. (Tao Liu): software, methodology. Corresponding Author. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Key Research and Development Program (Modern Agriculture) of Jiangsu Province (BE2022342), the National Natural Science Foundation of China (32172110), the Yangzhou University Postgraduate International Academic Exchange Fund (YZUF2025116).

Data Availability Statement: The data collected and used in this study contains proprietary information of the Agricultural College of Yangzhou University, Yangzhou and cannot be shared. The code used for analysis is available from the corresponding author, L.T., upon reasonable request.

Conflicts of Interest: The authors declare that they have no known financial interests or personal relationships that might influence the work reported in this paper.

References

1. Zhao, M.; Leng, Y.; Chao, S.; Xu, S.S.; Zhong, S. Molecular mapping of QTL for Fusarium head blight resistance introgressed into durum wheat. *Theor. Appl. Genet.* **2018**, *131*, 1939–1951. [\[CrossRef\]](#)
2. Guo, X.R.; Shi, Q.H.; Liu, Y.; Su, H.D.; Zhang, J.; Wang, M.; Wang, C.H.; Wang, J.; Zhang, K.B.; Fu, S.L.; et al. Systemic development of wheat—*Thinopyrum elongatum* translocation lines and their deployment in wheat breeding for Fusarium head blight resistance. *Plant J.* **2023**, *114*, 1475–1489. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Habib, M.T. Machine vision based papaya disease recognition. *J. King Saud Univ. Comput. Inf. Sci.* **2020**, *32*, 1215. [\[CrossRef\]](#)
4. Liu, T.; Zhao, Y.Y.; Sun, Y.; Wang, J.C.; Yao, Z.S.; Chen, C.; Zhong, X.C.; Liu, S.P.; Sun, C.M.; Li, T. High-throughput identification of fusarium head blight resistance in wheat varieties using field robot-assisted imaging and deep learning techniques. *J. Clean. Prod.* **2024**, *480*, 144024. [\[CrossRef\]](#)
5. Feng, G.Q.; Gu, Y.; Wang, C.; Zhang, D.Y.; Xu, R.; Zhu, Z.W.; Luo, B. Wheat Fusarium head blight severity grading using generative adversarial networks and semi-supervised segmentation. *Comput. Electron. Agric.* **2025**, *229*, 109817. [\[CrossRef\]](#)
6. Gong, Z.; Gao, C.F.; Feng, Z.H.; Dong, P.; Qiao, H.B.; Zhang, H.; Shi, L.; Guo, W. Integrating masked generative distillation and network compression to identify the severity of wheat fusarium head blight. *Comput. Electron. Agric.* **2024**, *227*, 109647. [\[CrossRef\]](#)
7. Gao, C.; Guo, W.; Yang, C.H.; Gong, Z.; Yue, J.B.; Fu, Y.Y.; Feng, H.K. A fast and lightweight detection model for wheat fusarium head blight spikes in natural environments. *Comput. Electron. Agric.* **2024**, *216*, 108484. [\[CrossRef\]](#)
8. Ma, H.; Huang, W.; Dong, Y.; Liu, L.; Guo, A. Using UAV-Based Hyperspectral Imagery to Detect Winter Wheat Fusarium Head Blight. *Remote Sens.* **2021**, *13*, 3024. [\[CrossRef\]](#)
9. Zhou, Q.; Huang, Z.L.; Liu, L.; Wang, F.M.; Teng, Y.; Liu, H.Y.; Zhang, Y.H.; Wang, R.J. High-throughput spike detection and refined segmentation for wheat Fusarium Head Blight in complex field environments. *Comput. Electron. Agric.* **2024**, *227*, 109552. [\[CrossRef\]](#)
10. Xiao, Y.; Dong, Y.; Huang, W.; Liu, L.; Ma, H. Wheat Fusarium Head Blight Detection Using UAV-Based Spectral and Texture Features in Optimal Window Size. *Remote Sens.* **2021**, *13*, 2437. [\[CrossRef\]](#)
11. Zhang, L.; Zhang, Z.; Wu, C.Y.; Sun, L. Segmentation algorithm for overlap recognition of seedling lettuce and weeds based on SVM and image blocking. *Comput. Electron. Agric.* **2022**, *201*, 107284. [\[CrossRef\]](#)
12. Sarkar, C.; Gupta, D.; Gupta, U.; Hazarika, B.B. Leaf disease detection using machine learning and deep learning: Review and challenges. *Appl. Soft Comput.* **2023**, *145*, 110534. [\[CrossRef\]](#)
13. Wen, J.; Yin, Y.; Zhang, Y.; Pan, Z.; Fan, Y. Detection of Wheat Lodging by Binocular Cameras during Harvesting Operation. *Agriculture* **2022**, *13*, 120. [\[CrossRef\]](#)
14. Liu, T.H.; Nie, X.N.; Wu, J.M.; Zhang, D.; Liu, W.; Cheng, Y.F.; Zheng, Y.; Qiu, J.; Qi, L. Pineapple (*Ananas comosus*) fruit detection and localization in natural environment based on binocular stereo vision and improved YOLOv3 model. *Precis. Agric.* **2022**, *24*, 139–160. [\[CrossRef\]](#)
15. Ye, S.; Xue, X.Y.; Sun, Z.; Xu, Y.; Sun, T.; Ye, J.W.; Jin, Y.K. Research and Experiment on Soybean Plant Identification Based on Laser Ranging Sensor. *Agronomy* **2023**, *13*, 2757. [\[CrossRef\]](#)
16. Jin, S.; Su, Y.J.; Wu, F.F.; Pang, S.X.; Gao, S.; Hu, T.Y.; Liu, J.; Guo, Q.H. Stem–Leaf Segmentation and Phenotypic Trait Extraction of Individual Maize Using Terrestrial LiDAR Data. *IEEE Trans. Geosci. Electron.* **2019**, *57*, 1336–1346. [\[CrossRef\]](#)
17. Chen, Y.; Jiang, W.; Luo, Z.; Yang, L. A novel 3D reconstruction method with a binocular-line laser system. *Measurement* **2024**, *227*, 114238. [\[CrossRef\]](#)
18. Yang, L.; Kang, B.Y.; Huang, Z.L.; Zhao, Z.; Xu, X.G.; Feng, J.S.; Zhao, H.H. Depth Anything V2. *arXiv* **2024**, arXiv:2406.09414. [\[CrossRef\]](#)
19. Cao, S.L.; Xu, B.H.; Zhou, W.; Zhou, L.T.; Zhang, J.F.; Zheng, Y.H.; Hu, W.J.; Han, Z.G.; Lu, H. The blessing of Depth Anything: An almost unsupervised approach to crop segmentation with depth-informed pseudo labeling. *Plant Phenomics* **2025**, *7*, 100005. [\[CrossRef\]](#)
20. Peng, H.X.; Zhong, J.R.; Liu, H.A.; Li, J.; Yao, M.W.; Zhang, X. ResDense-focal-DeepLabV3+enabled litchi branch semantic segmentation for robotic harvesting. *Comput. Electron. Agric.* **2023**, *206*, 107691. [\[CrossRef\]](#)
21. He, K.M.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. *arXiv* **2018**, arXiv:1703.06870.
22. Wang, H.; Chen, D.P.; Li, C.L.; Tian, N.; Zhang, J.; Xu, J.R.; Wang, C.F. Stage-specific functional relationships between Tub1 and Tub2 beta-tubulins in the wheat scab fungus *Fusarium graminearum*. *Fungal Genet. Biol.* **2019**, *132*, 103251. [\[CrossRef\]](#)
23. Zhang, D.Y.; Luo, H.S.; Cheng, T.; Li, W.F.; Zhou, X.G.; Wei, G.; Gu, C.Y.; Diao, Z.H. Enhancing wheat Fusarium head blight detection using rotation Yolo wheat detection network and simple spatial attention network. *Comput. Electron. Agric.* **2023**, *211*, 107968. [\[CrossRef\]](#)
24. 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. [\[CrossRef\]](#)
25. Meyer, G.E.; Neto, J.C. Verification of color vegetation indices for automated crop imaging applications. *Comput. Electron. Agric.* **2008**, *63*, 282–293. [\[CrossRef\]](#)

26. Pérez, A.J.; López, F.; Benlloch, J.V.; Christensen, S. Colour and shape analysis techniques for weed detection in cereal fields. *Comput. Electron. Agric.* **2000**, *25*, 197–212. [[CrossRef](#)]
27. Wang, N.; Wei, X.; Zhou, M.T.; Wang, H.J.; Bai, Y.B. UAV-based remote sensing using visible and multispectral indices for the estimation of vegetation cover in an oasis of a desert. *Ecol. Indic.* **2022**, *141*, 109155. [[CrossRef](#)]
28. Philipp, I.; Rath, T. Improving plant discrimination in image processing by use of different colour space transformations. *Comput. Electron. Agric.* **2002**, *35*, 1–15. [[CrossRef](#)]
29. Liu, T.; Wang, J.L.; Chen, J.F.; Zhang, W.J.; Wang, Y.; Zhao, Y.Y.; Sun, Y.; Yao, Z.S.; Wang, J.Y.; Sun, C.M. Detection of the number of wheat stems using multi-view images from smart glasses. *Comput. Electron. Agric.* **2025**, *235*, 110370. [[CrossRef](#)]
30. Feng, G.; Gu, Y.; Wang, C.; Zhou, Y.; Huang, S.; Luo, B. Wheat Fusarium Head Blight Automatic Non-Destructive Detection Based on Multi-Scale Imaging: A Technical Perspective. *Plants* **2024**, *13*, 1722. [[CrossRef](#)] [[PubMed](#)]
31. Liang, K.; Ren, Z.Z.; Song, J.P.; Yuan, R.; Zhang, Q. Wheat FHB resistance assessment using hyperspectral feature band image fusion and deep learning. *Int. J. Agric. Biol. Eng.* **2024**, *17*, 240–249. [[CrossRef](#)]
32. Zhang, H.S.; Zhao, J.L.; Huang, L.S.; Huang, W.J.; Dong, Y.Y.; Ma, H.Q.; Ruan, C. Development of new indices and use of CARS-Ridge algorithm for wheat fusarium head blight detection using in-situ hyperspectral data. *Biosyst. Eng.* **2024**, *237*, 13–25. [[CrossRef](#)]
33. Francesconi, S.; Harfouche, A.; Maesano, M.; Balestra, G.M.; Thermal, U.A.-B. 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. *Front. Plant Sci.* **2021**, *12*, 628575. [[CrossRef](#)] [[PubMed](#)]
34. Wang, F.L.; Li, F.P.; Mohan, V.; Dudley, R.; Gu, D.B.; Bryant, R. An unsupervised automatic measurement of wheat spike dimensions in dense 3D point clouds for field application. *Biosyst. Eng.* **2022**, *223*, 103–114. [[CrossRef](#)]
35. Qiu, R.C.; Yang, C.; Moghimi, A.; Zhang, M.; Steffenson, B.J.; Hirsch, C.D. Detection of Fusarium Head Blight in Wheat Using a Deep Neural Network and Color Imaging. *Remote Sens.* **2019**, *11*, 2658. [[CrossRef](#)]
36. Li, Y.W.; Yuan, X.F.; Wang, B.; Zhang, J. Design of wheat ear detection in natural environment based on improved YOLOv5s. *J. Comput. Methods Sci. Eng.* **2024**, *24*, 3517–3530. [[CrossRef](#)]
37. Guan, Y.J.; Pan, J.Q.; Fan, Q.Q.; Yang, L.L.; Yin, X.; Jia, W.K. CTWheatNet: Accurate detection model of wheat ears in field. *Comput. Electron. Agric.* **2024**, *225*, 109272. [[CrossRef](#)]
38. Li, Z.P.; Zhu, Y.J.; Sui, S.S.; Zhao, Y.H.; Liu, P.; Li, X. Real-time detection and counting of wheat ears based on improved YOLOv7. *Comput. Electron. Agric.* **2024**, *218*, 10867. [[CrossRef](#)]
39. Artemenko, N.V.; Genaev, M.A.; Epifanov, R.U.; Komyshev, E.G.; Kruchinina, Y.V.; Koval, V.S.; Goncharov, N.P.; Afonnikov, D.A. Image-based classification of wheat spikes by glume pubescence using convolutional neural networks. *Front. Plant Sci.* **2024**, *14*, 1336192. [[CrossRef](#)]
40. Tsomko, E.; Kim, H.J.; Izquierdo, E. Linear Gaussian blur evolution for detection of blurry images. *IET Image Proc.* **2010**, *4*, 302–312. [[CrossRef](#)]
41. Flusser, J.; Farokhi, S.; Hoschl, C.; Suk, T.; Zitova, B.; Pedone, M. Recognition of Images Degraded by Gaussian Blur. *IEEE Trans. Image Process.* **2016**, *25*, 790–806. [[CrossRef](#)] [[PubMed](#)]
42. Bao, W.X.; Huang, C.P.; Hu, G.S.; Su, B.B.; Yang, X.J. Detection of Fusarium head blight in wheat using UAV remote sensing based on parallel channel space attention. *Comput. Electron. Agric.* **2024**, *217*, 108630. [[CrossRef](#)]
43. Shubham, R.; Matteo, G. Comparative Evaluation of Modified Wasserstein GAN-GP and State-of-the-Art GAN Models for Synthesizing Agricultural Weed Images in RGB and Infrared Domain. *MethodsX* **2025**, *14*, 103309.

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.