agriculture

MDPI

Article

Field Obstacle Detection and Location Method Based on Binocular Vision

Yuanyuan Zhang [1], Kunpeng Tian [1], Jicheng Huang [1], Zhenlong Wang [1], Bin Zhang [2,*] and Qing Xie [1,*]

[1] Nanjing Institute of Agricultural Mechanization, Ministry of Agriculture and Rural Affairs, Nanjing 210014, China; 82101222110@caas.cn (Y.Z.); tiankunpeng@caas.cn (K.T.); huangjicheng@caas.cn (J.H.); 82101225604@caas.cn (Z.W.)

[2] Graduate School of Chinese Academy of Agricultural Sciences, Beijing 100083, China

* Correspondence: zhangbin03@caas.cn (B.Z.); xieqing@caas.cn (Q.X.)

Abstract: When uncrewed agricultural machinery performs autonomous operations in the field, it inevitably encounters obstacles such as persons, livestock, poles, and stones. Therefore, accurate recognition of obstacles in the field environment is an essential function. To ensure the safety and enhance the operational efficiency of autonomous farming equipment, this study proposes an improved YOLOv8-based field obstacle detection model, leveraging depth information obtained from binocular cameras for precise obstacle localization. The improved model incorporates the Large Separable Kernel Attention (LSKA) module to enhance the extraction of field obstacle features. Additionally, the use of a Poly Kernel Inception (PKI) Block reduces model size while improving obstacle detection across various scales. An auxiliary detection head is also added to improve accuracy. Combining the improved model with binocular cameras allows for the detection of obstacles and their three-dimensional coordinates. Experimental results demonstrate that the improved model achieves a mean average precision (mAP) of 91.8%, representing a 3.4% improvement over the original model, while reducing floating-point operations to 7.9 G (Giga). The improved model exhibits significant advantages compared to other algorithms. In localization accuracy tests, the maximum average error and relative error in the 2–10 m range for the distance between the camera and five types of obstacles were 0.16 m and 2.26%. These findings confirm that the designed model meets the requirements for obstacle detection and localization in field environments.

Keywords: YOLOv8; binocular vision; field obstacle; autonomous agricultural machinery; detection; localization

Citation: Zhang, Y.; Tian, K.; Huang, J.; Wang, Z.; Zhang, B.; Xie, Q. Field Obstacle Detection and Location Method Based on Binocular Vision. *Agriculture* **2024**, *14*, 1493. https://doi.org/10.3390/agriculture14091493

Academic Editor: Caiyun Lu

Received: 12 July 2024
Revised: 23 August 2024
Accepted: 23 August 2024
Published: 1 September 2024

1. Introduction

Agricultural machinery is a critical tool for implementing precision agriculture, significantly enhancing labor efficiency and reducing labor intensity. With the advancement of intelligent agricultural machinery, autonomous agricultural machines are increasingly attracting the attention of researchers [1]. However, the real-world farmland environment is complex and unstructured, with inevitable obstacles such as humans, utility poles, and stones present during task execution by intelligent agricultural machinery. To ensure the safety and accuracy of autonomous operations in the field, it is crucial to detect surrounding obstacles swiftly and accurately. In recent years, sensor technology has been applied to autonomous farming operations, enabling the precise identification and localization of typical obstacles in the field.

Currently, the detection of field obstacles primarily relies on radar and visual sensors. Radar typically employs 2D/3D LiDAR for data acquisition, determining obstacle positions through LiDAR point cloud processing [2,3]. Considering the cost and the complexity of the farmland environment, visual sensors are more suitable for field obstacle detection than radar. Visual sensors offer the advantage of providing rich information about obstacles, including classification, shape, and texture [4]. Common visual sensors include monocular

and binocular cameras. Monocular cameras estimate target depth using image data from a single viewpoint. In contrast, binocular cameras leverage parallax and matching features from different viewpoints to obtain three-dimensional spatial information, enabling the precise localization of targets [5,6].

Object detection is a fundamental task in computer vision, primarily aimed at locating and identifying targets of interest in images or videos [7]. With the significant advances in feature extraction and object detection achieved by deep learning methods, issues related to the low accuracy and poor generalization capability of image-based methods have been addressed, along with improvements in detection speed. Currently, mainstream object detection algorithms are categorized into two-stage and one-stage methods [8]. R-CNN [9] was the first popular two-stage detection algorithm, decomposing the detection task into two phases: generating candidate regions and then classifying these regions with bounding box regression. Classic two-stage models also include Fast R-CNN [10], Faster R-CNN [11], and Cascade R-CNN [12]. While two-stage models offer commendable detection performance, they are characterized by high complexity, significant computational demands, and a large number of parameters, resulting in slower target detection speeds. These limitations hinder the ability to effectively balance detection speed and accuracy. In contrast, one-stage object detection algorithms simultaneously generate candidate boxes and perform classification and bounding box regression, as exemplified by the YOLO (You Only Look Once) [13] series, SSD (Single Shot MultiBox Detector) [14] model, and RetinaNet [15] model. The SSD model employs convolutional neural networks (CNNs) to extract image features and perform detection across multiple feature layers. This approach allows it to predict bounding boxes and classes on feature maps of different scales, enabling robust object detection. RetinaNet introduces a novel loss function, Focal Loss, to address the class imbalance issue. Focal Loss enhances detection accuracy by reducing the focus on easily classified samples and increasing the emphasis on hard-to-classify ones. In recent studies, Liu [16] utilized the SSD algorithm for pedestrian detection in orchards. Peng [17] leveraged a multi-scale feature fusion-based RetinaNet network to detect fruits in complex field environments. These studies demonstrate the advantages of the SSD and RetinaNet algorithms in object detection, though they still exhibit certain drawbacks in terms of speed and small object detection. The YOLO series, including the classic YOLOv3 [18], YOLOv5 [19], and the subsequent YOLOv6 [20], YOLOv7 [21], and YOLOv8 [22] models, are known for their high computational efficiency, enabling high-precision real-time performance, and are widely used in agriculture. For instance, Li [23] utilized YOLOv3-tiny as the base framework, integrating its shallow features with the second prediction layer features to create a small target detection layer for the third prediction layer. This approach, using a hybrid SEAM and CBAM attention mechanism module, enhanced background interference resistance and improved the average accuracy of the model by 11%. Xue [4] proposed an improved method based on the YOLOv5s algorithm, employing the K-Means clustering algorithm to generate anchor box scales to cover all scales as much as possible, and using the CIoU loss function, combining three geometric measures—overlap area, center distance, and aspect ratio—to reduce missed and false detections, thus improving detection accuracy.

The YOLOv8 architecture, a variant of the YOLO series object detection models, employs deep neural networks for object recognition and localization. Khow [24] introduced CA and Wise-IoU into the YOLOv8 network, resulting in the YOLOv8-CAW model, which not only enhances detection accuracy but also accurately estimates the distance of target objects. Lan [25] aimed to improve the detection accuracy of uncrewed tractors for inclined obstacles in farmland by optimizing the YOLOv8 model. Their enhancements included embedding the SE attention mechanism to improve detection accuracy and utilizing MobileNetv2 and BiFPN to reduce computational load. The improved model achieved a mean average precision (mAP) as high as 98.84%. Although these modified YOLOv8 models have demonstrated excellent performance in object detection tasks, the complex field environment and the limited computational capacity of onboard equipment in uncrewed agricultural machinery make obstacle detection a challenging task. Moreover, the

diverse and variable shapes of field obstacles render YOLOv8 less robust to changes in target shapes, often resulting in inaccurate localization. Additionally, YOLOv8 encounters difficulties in detecting small targets, leading to false negatives and false positives. Field obstacles are often occluded by crops, further contributing to false negatives or positives. To address these issues, this study proposes a novel LPA-YOLOv8n model for the real-time detection of field obstacles, aiming to enhance the accuracy and real-time performance of obstacle detection in agricultural settings.

The main contributions of this paper are as follows:

(1) Introduction of LSKA in the SPPF Module: By incorporating the LSKA module, the network's focus on important features is enhanced, improving the model's capability to extract features in complex field environments. This effectively addresses the issue of robustness to shape variations of various field obstacles.
(2) Replacement of the Bottleneck Block with PKI Block in the Backbone's C2f Section: This substitution serves as the primary gradient flow branch, allowing for a more comprehensive consideration of targets at different scales and enhancing the model's ability to detect small or hard-to-detect objects.
(3) Addition of an Auxiliary Detection Head: By adding an auxiliary detection head to the original detection head, the model gains additional information and feature processing, improving the detection accuracy of occluded targets in agricultural environments.
(4) Fusion of Detected Object Frame Information with Binocular Matching Algorithm: This integration enables accurate identification and precise distance measurement of typical obstacles in complex field environments.

The rest of the paper is structured as follows: Section 2 provides an overview of the YOLOv8 model and details the proposed improvements and network architecture. Section 3 introduces the experimental setup and the design of the obstacle detection and localization system. Section 4 presents and analyzes the experimental results, followed by a discussion in Section 5. Finally, Section 6 concludes this paper by summarizing its contributions.

2. Proposed Improvement Methods

2.1. The YOLOv8 Algorithm

This section first provides an overview of the YOLOv8 algorithm structure, laying the foundation for the design of the improved algorithm. The YOLOv8 network structure consists of three main components: the backbone, neck, and head. The overall structure of YOLOv8 is illustrated in Figure 1. The following detailed explanation focuses on each of the three components of YOLOv8.

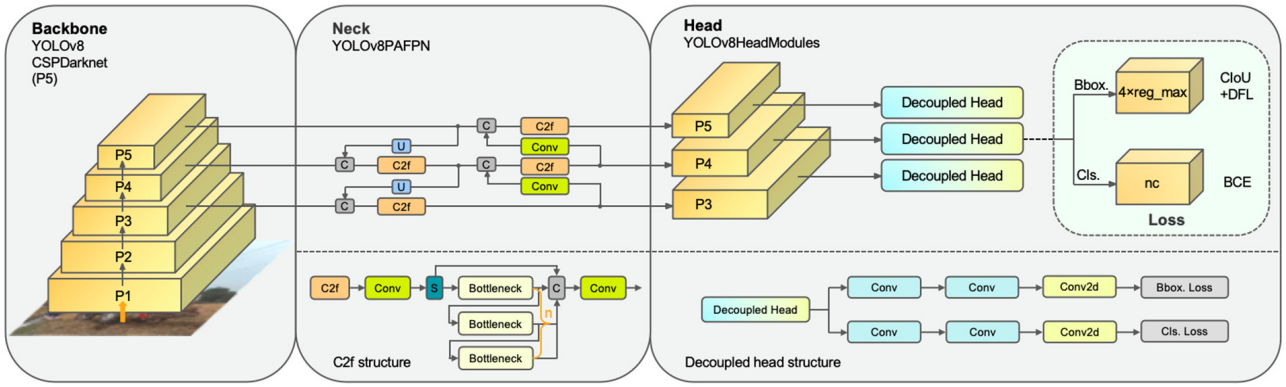


Figure 1. The overall structure of YOLOv8. "C" denotes the concat operation, "U" denotes the upsampling operation, and "S" denotes the split operation.

Backbone: The backbone of YOLOv8 draws on the principles of CSPDarkNet (Cross Stage Partial DarkNet) [26] and is primarily composed of three modules: Conv (convolution), C2f (CSPNet with 2 fused layers), and SPPF (Spatial Pyramid Pooling Fusion). The Conv module consists of Conv2d, BN (BatchNorm2d), and the SiLU (Sigmoid Linear Unit). Inspired by the ELAN (Efficient Layer Aggregation Network) design of YOLOv7, the C2f module offers fewer parameters and superior feature extraction capabilities compared to the C3 (CSP Bottleneck with 3 convolutions) module in YOLOv5, and C2f is shown in Figure 1. The SPPF module employs a cascade of small pooling kernels, which more effectively extracts image features.

Neck: The neck of YOLOv8 integrates concepts from FPNs (Feature Pyramid Networks) [27] and PANs (Path Aggregation Networks) [28]. FPNs convey strong semantic features top-down, facilitating feature extraction from images at various scales. PANs introduce a bottom-up pyramid following the FPN, enhancing multi-scale feature representation capabilities by combining the strengths of both FPN and PAN structures.

Head: YOLOv8 adopts a decoupled head structure, separating classification and detection heads, as illustrated in Figure 1. Additionally, it transitions from anchor-based to anchor-free bounding boxes, improving detection speed and accuracy.

2.2. LPA-YOLOv8 Detection Model

The objective of field obstacle detection is to recognize and locate various classic obstacles (examples include persons, livestock, agricultural machinery, poles, and stones) in agricultural fields. This study proposes an LPA-YOLOv8n detection model based on YOLOv8. The LPA-YOLOv8n model improves the SPPF module in the backbone, and the Bottleneck module in C2f, and introduces an auxiliary detection head in the head section. The modified model's network architecture is illustrated in Figure 2.

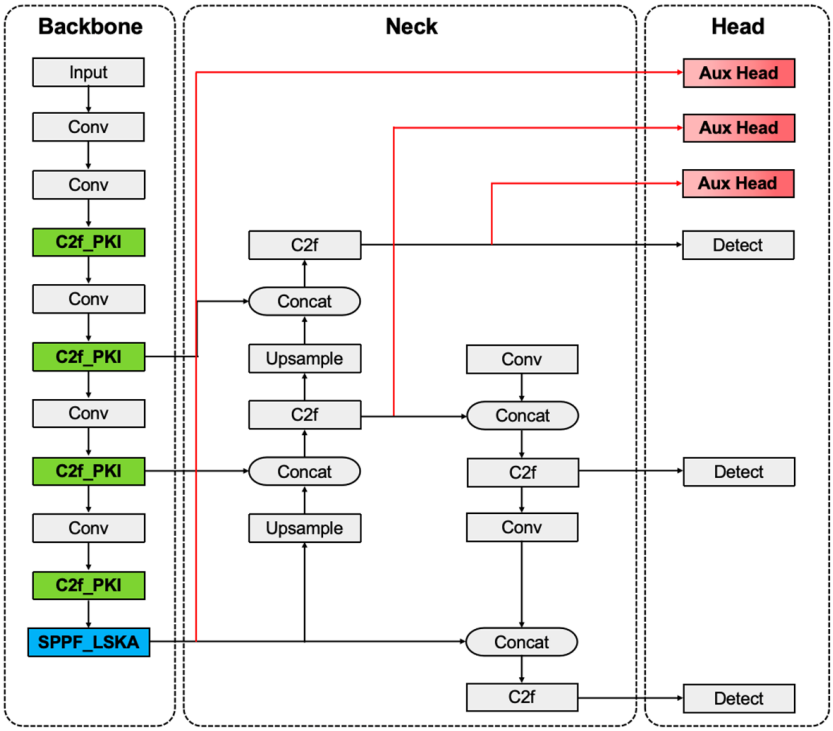


Figure 2. Improved YOLOv8 network model architecture. We incorporated the LSKA module into the SPPF, replaced the Bottleneck component in the C2f with the PKI module, and employed an auxiliary detection head.

2.2.1. SPPF-LSKA

The Vision Attention Network (VAN) [29] with the LKA module has demonstrated exceptional performance across various vision-based tasks. The original LKA module does not utilize dilated depth convolutions but instead employs large convolution kernels in 2D depth convolutions, as illustrated in Figure 3a. To mitigate the high computational cost associated with large kernel sizes in depth convolutions, the large-kernel depth convolutions are decomposed into smaller-kernel depth convolutions, effectively using dilated depth convolutions with relatively large kernel sizes, as shown in Figure 3b. Despite these significant improvements, issues with computational complexity and memory usage persist. Consequently, Lau et al. [30] proposed a novel Large Separable Kernel Attention (LSKA) module, which fundamentally decomposes the 2D convolution kernels of depth convolution layers into cascaded horizontal and vertical 1-D convolution kernels, as depicted in Figure 3c. Specifically, the first two layers of LKA are decomposed into four layers, each consisting of two 1-D convolution layers, as demonstrated in Figure 3d. Compared to the standard LKA design, the LSKA module maintains comparable performance while offering higher computational efficiency and reduced memory usage.

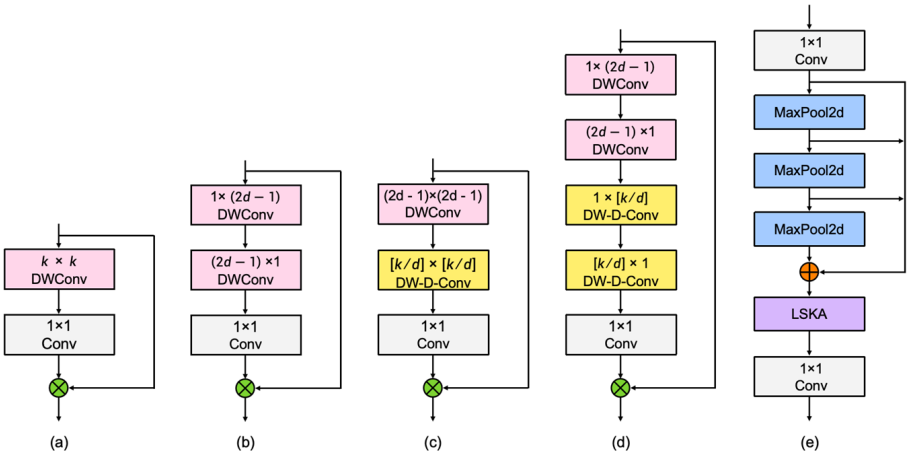


Figure 3. (**a**) LKA-trivial; (**b**) LSKA-trivial; (**c**) LKA; (**d**) LSKA; and (**e**) SPPF-LSKA.

The forward propagation process of the LSKA module primarily involves the following steps: Initially, two convolution layers independently extract features from the input feature map in the horizontal and vertical directions, allowing for a more detailed handling of image features. Subsequently, LSKA employs spatial expansion convolutions with different dilation rates to further extract features. Finally, after a series of convolution operations, a 1×1 convolution fuses the extracted features to generate the final attention map. This attention map is then multiplied element-wise with the original input feature map to enhance the network's focus on important features.

In this study, we integrate the LSKA module into the SPPF module, positioning it after all max-pooling operations and before the second convolutional layer, as shown in Figure 3e. This placement is strategic to leverage large kernel convolutions and spatial expansion convolutions at critical points of feature extraction and fusion, thereby maximizing feature representation capability while controlling the computational load and complexity of the model.

In summary, the LSKA module enhances the model's robustness against variations in the shapes of different field obstacles by capturing extensive contextual information using large separable convolutional kernels and spatially dilated convolutions. By performing operations in both horizontal and vertical directions, LSKA allows for more detailed processing of image features. The attention maps generated through a series of convolutional operations are then used to weight the original features, thereby increasing the network's focus on critical elements. This approach effectively mitigates potential issues related to the model's robustness in dealing with diverse shapes of field obstacles.

2.2.2. C2f-PKI

In the previous subsection, we introduced the LSKA module, which, while expanding the receptive field, introduces significant background noise and overly sparse feature representations. These issues adversely affect the accurate detection of small targets. To address these concerns, we replace the Bottleneck module in the backbone's C2f module with the PKI (Poly Kernel Inception) Block from PKINet [31]. This substitution effectively resolves the aforementioned problems while enhancing multi-scale feature extraction capabilities, emphasizing important features, and ensuring that small target features are not overlooked, thereby improving detection performance. Modifying the backbone allows us to leverage the PKI Block's multi-scale feature learning ability during the initial extraction phase, enhancing the quality of the foundational features.

A PKI Block consists of a PKI module and a CAA module, as shown in Figure 4a. The PKI module utilizes a non-dilated Inception-style depth convolution structure, depicted in Figure 4b. It first employs a small kernel convolution to capture local information, followed by a series of parallel depth convolutions to extract multi-scale texture features at different receptive fields. Finally, a 1×1 convolution integrates features with varying receptive field sizes.

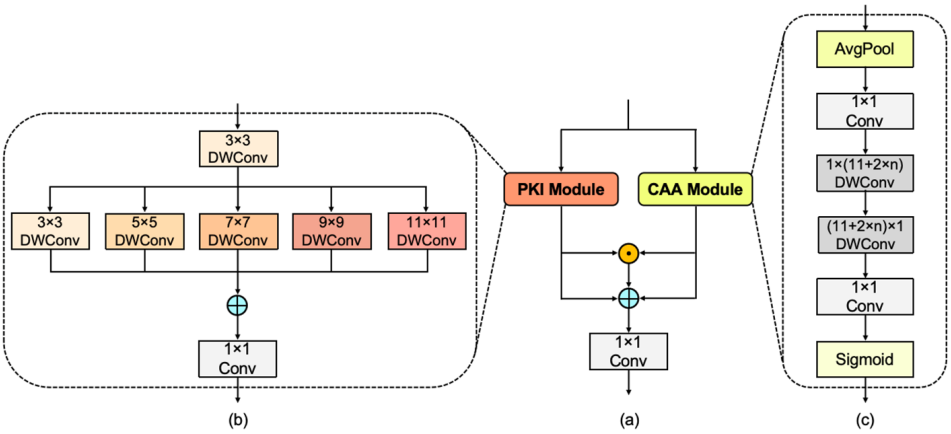


Figure 4. (**a**) PKI Block; (**b**) PKI module; (**c**) CAA module.

The CAA module aims to establish dependencies between distant pixels, as shown in Figure 4c. It first applies average pooling followed by a 1×1 convolution to capture local region features. Next, it fuses the feature maps generated by two depthwise separable convolutions with the original feature map. Finally, the CAA module produces an attention weight to enhance important features by weighting the feature maps at each scale.

In summary, the PKI module focuses on extracting multi-scale local contextual information, while the CAA module captures long-range contextual information. The collaboration of these two components enhances the model's capability to extract feature information in complex field environments and improves the detection of small or hard-to-detect targets.

2.2.3. Aux Head

An Aux head is an auxiliary head added to the intermediate layers before the feature output, as shown in Figure 5a. Its primary purpose is to enable the intermediate layers to learn more information, providing richer gradient information to aid training [32]. In the YOLOv8 model, the Lead head focuses on global features, but it may struggle to effectively handle partially visible features of occluded targets. The Aux head, by performing detection at shallower layers, captures more detailed and local information, enhancing the recognition of occluded targets and compensating for the Lead head's limitations.

During loss computation, the Lead head generates the primary detection results and independently calculates its loss. Meanwhile, the Aux head also contributes to the loss function calculation and assists in backpropagation, updating preceding parameters. The Aux head uses the positive samples matched by the Lead head as its own positive samples

to calculate the corresponding loss. Finally, the losses from the Lead head and Aux head are weighted and combined to form the total loss function, as illustrated in Figure 5b.

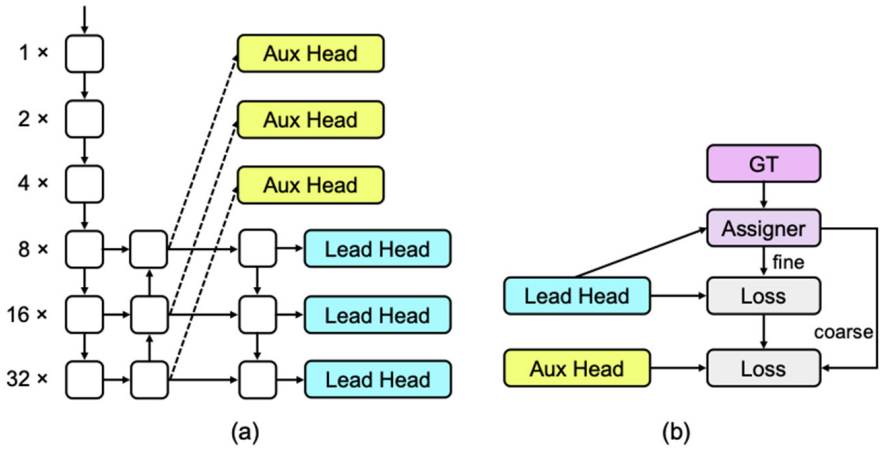


Figure 5. (**a**) Model with auxiliary head; (**b**) coarse-to-fine Lead head-guided assigner.

3. Experimental Setup

3.1. Dataset

For field operations, it is crucial that the training data used for obstacle detection accurately reflect the real-world perspectives encountered by agricultural machinery in the field, ensuring appropriate realism and feasibility. Our dataset comprises images from two sources. The first set of images was captured using an Intel Realsense D435i RGB-D camera (Intel Corporation, Santa Clara, CA, USA), with video sliced into frames using Python scripts, allowing for the acquisition of images from various perspectives and scenes. These images were collected in October 2023 from rice and wheat fields in Huishan District, Wuxi, Jiangsu Province, China, at multiple times during the day, including early morning, noon, and dusk, ensuring a diversity of lighting conditions. The second part of the dataset was assembled using images of various types of obstacles sourced from search engines such as Google and Baidu. These images come from a wide range of backgrounds, resolutions, and shooting angles, thereby increasing the dataset's diversity. Combining these two sources, we assembled a raw dataset of 2378 images, each with a resolution of 640 × 480 pixels. Every image contains at least one target object, ensuring that the model can reliably learn and validate across different datasets. The dataset primarily includes five types of typical field obstacles: person, livestock, agricultural machinery, pole, and stone. Sample images from the dataset are shown in Figure 6a. The raw dataset was randomly split into training, validation, and test sets in an 8:1:1 ratio.

3.2. Data Augmentation

The obstacles were categorized into five types: agricultural machinery, person, livestock, stone, and pole. Using the labeling tool, obstacles in the images were annotated by type, and the resulting txt files were used to store the data label information for each category. Table 1 shows the dataset quantities and label counts for the five categories. Before training the model, we performed image processing on the training set to enhance the model's generalization and robustness. We added weather and environmental effects such as fog, rain, glare, and darkening. Some processed images are shown in Figure 6b. Simulating image processing under various weather conditions enables the model to maintain high detection performance in adverse environments. For example, under low visibility conditions such as dense fog or heavy rain, the model remains capable of effectively identifying field obstacles, thereby ensuring the safe operation of agricultural machinery. Furthermore, the model's adaptability to different lighting conditions, including shadows and reflections, enhances its robustness across various times of day and weather conditions. The classification quantities of the preprocessed dataset are detailed in Table 2.

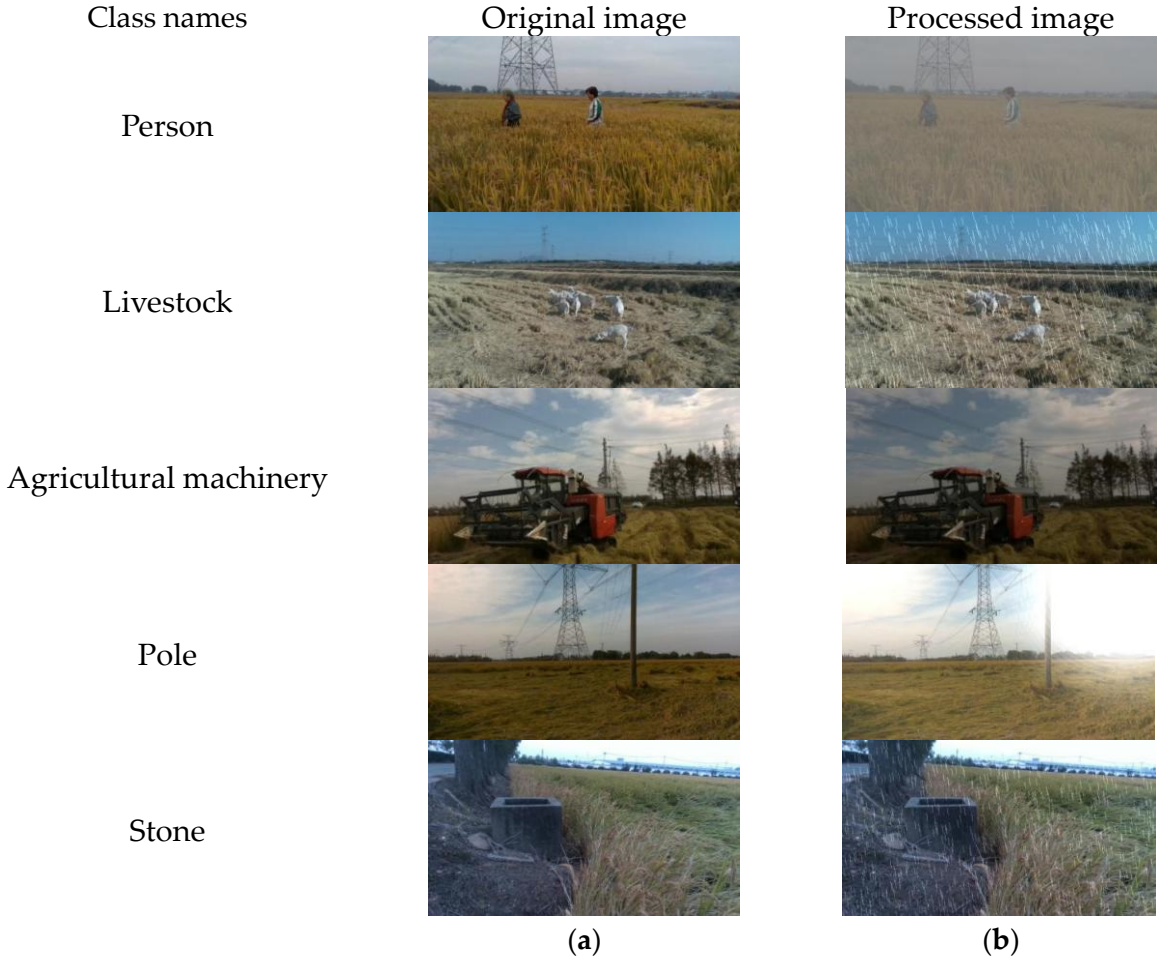


Figure 6. (**a**) Five typical obstacles in rice–wheat fields; (**b**) added conditions of fogging, raining, sunlight, and darkening to images.

Table 1. The number of images and labels for the five categories.

Class	Image	Instance					
		Person	Livestock	Agricultural Machinery	Pole	Stone	All
train	1904	1501	1383	641	627	546	4698
val	237	200	205	67	84	59	615
test	237	182	87	75	116	48	508
all	2378	1883	1675	783	827	653	5821

Table 2. The classification quantities of the preprocessed dataset.

Processing Method	Original	Fogging	Raining	Sunlight	Darkening
Image	878	375	375	375	375

3.3. Obstacle Detection and Localization System

3.3.1. System Framework

Using a binocular camera system to detect and locate obstacles in agricultural fields involves the following specific operational steps: Firstly, RGB images and depth maps of the agricultural field environment ahead of the agricultural machinery operation are captured separately using the RGB camera and two infrared cameras (left and right). Subsequently,

the improved LPA-YOLOv8n detection algorithm is applied to the RGB images to detect obstacles and obtain pixel coordinates of their location points. Using depth information, the three-dimensional coordinates of the obstacle's location points in the camera coordinate system are calculated. These coordinates are then transformed to obtain the absolute coordinates of the obstacles in the world coordinate system, facilitating the detection and localization of typical field obstacles during agricultural machinery operations. A Block diagram of the field obstacle detection and localization system is shown in Figure 7.

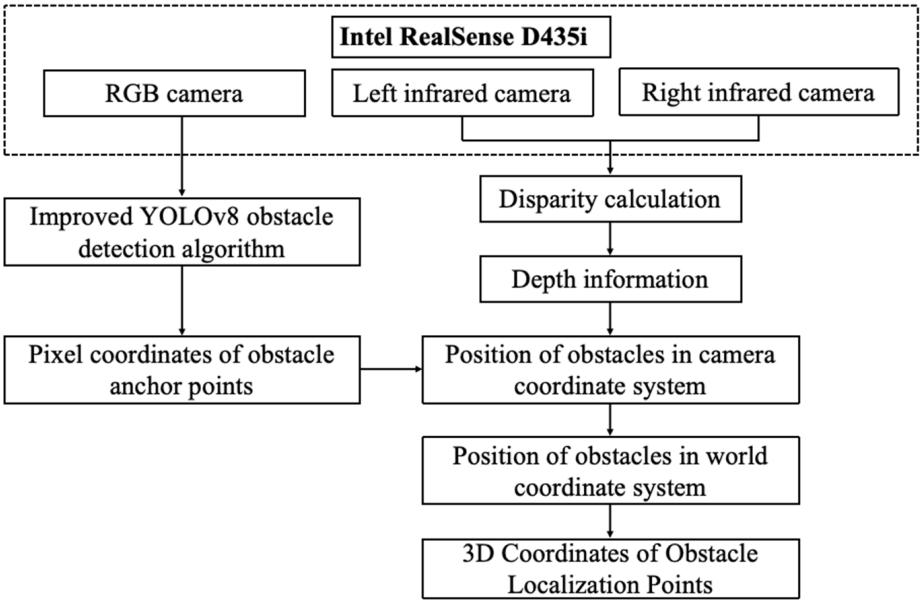


Figure 7. A Block diagram of the field obstacle detection and localization system.

3.3.2. Localization Principle

In the process of distance measurement, the accuracy may vary due to differences in the specifications of binocular cameras. The binocular camera used in this experiment is the Intel RealSense D435i, which measures depth by calculating the distance based on the disparity generated on the imaging planes of the left and right infrared cameras [28]. The principle of disparity is illustrated in Figure 8, where the imaging points of the obstacle's center $P(x_c, y_c, z_c)$ in the left and right cameras are $P_l(x_l, y_l)$ and $P_r(x_r, y_r)$. Since the left and right cameras are on the same horizontal plane, the Y-coordinates in the horizontal direction are identical, $Y_l = Y_r = Y$; f denotes the focal length of the cameras, and B represents the baseline, which is the distance between the origins of the two cameras. According to the principles of triangulation, we derive the equation below.

$$X_l = f\frac{x_c}{z_c} \tag{1}$$

$$X_r = f\frac{(x_c - B)}{z_c} \tag{2}$$

$$Y = f\frac{y_c}{z_c} \tag{3}$$

Disparity is defined as the deviation of the same point in the X-direction between the left and right cameras. To obtain the disparity information $D = X_l - X_r$, the position of point P in the coordinate system of the left camera can be expressed as follows:

$$x_c = \frac{B \cdot X_l}{D} \tag{4}$$

$$y_c = \frac{B \cdot Y}{D} \tag{5}$$

$$z_c = \frac{B \cdot f}{D} \tag{6}$$

The transformation from the camera coordinate system to the world coordinate system is a rigid body transformation, achieved through rotation and translation. The transformation formula is as follows:

$$\begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} \tag{7}$$

where R is the product of the rotation matrices in the X, Y, and Z directions, and T is the translation matrix. In summary, for any point in space, if its imaging points can be found in two cameras, its three-dimensional coordinates and thus the distance to the obstacle can be computed.

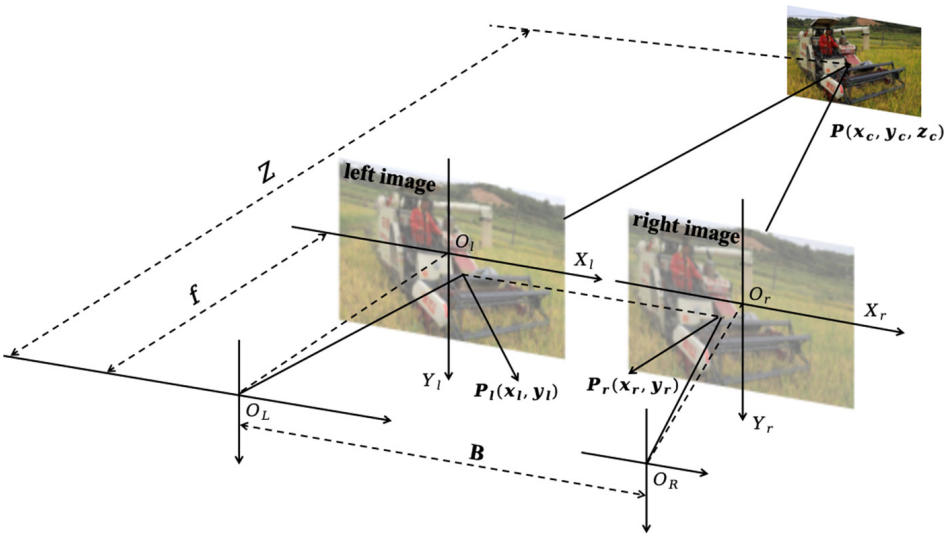


Figure 8. Distance measurement principle.

4. Experimental Results and Analysis

4.1. Implementation Details

This study employed Python 3.8 with PyTorch 1.12.1 as the deep learning framework, utilizing Python as the programming language platform. The experiments were conducted on a desktop computer equipped with an Intel(R) Xeon(R) Gold 5218R CPU @ 2.10GHz (Intel Corporation, Santa Clara, CA, USA) and NVIDIA RTX A4000 (NVIDIA Corporation, Santa Clara, CA, USA), running Windows 11. CUDA 11.3 and CUDNN 8.2.1 were installed for parallel computing and deep neural network libraries, respectively. Hyperparameters were configured with a Batch Size of 32, training epochs set to 300, and a Learning Rate set to 0.001. These settings were chosen to ensure efficient training and validation of the model, aiming to maximize performance and accuracy. It is noteworthy that to prevent overfitting during training, an early stopping mechanism was implemented. Training ceases if the accuracy of the validation set does not improve after a specified number of training epochs.

4.2. Evaluation Metrics

To evaluate the performance of our LPA-YOLOv8 model in detecting field obstacles, we employed three common metrics used in object detection tasks: Recall (R), Precision (P), and mean Average Precision (mAP).

Recall measures the proportion of actual positive samples correctly predicted as positive by the model among all positive samples. Precision indicates the proportion of

samples predicted as positive by the model that are actually positive. Their respective formulas are as follows:

$$R = \frac{TP}{TP + FN} \tag{8}$$

$$P = \frac{TP}{TP + FP} \tag{9}$$

In the formulas, TP (True Positive), FP (False Positive), and FN (False Negative), respectively, denote the number of correctly predicted positive samples, incorrectly predicted negative samples, and incorrectly predicted positive samples.

Average Precision (AP) represents the average precision corresponding to the variation of recall from 0 to 1. mAP is the mean of AP across all categories. mAP@0.5 denotes the average precision across all categories when the IoU threshold is 0.5. The calculation formulas are as follows:

$$AP = \int_0^1 P(R)dR \tag{10}$$

$$mAP = \frac{1}{m}\sum_{\rangle=1}^{m} AP_{\rangle} \tag{11}$$

4.3. Ablation Experiment

To evaluate the effectiveness and feasibility of the improved YOLOv8 algorithm, we conducted ablation experiments. This study introduced three improvements to the baseline YOLOv8 model: Improvement A: integrating the LSKA into the SPPF module; Improvement B: replacing the Bottleneck with the PKI Block; and Improvement C: adding an Aux head to the head section. By incorporating each improvement module individually into the baseline model, we could compare the impact of adding or replacing modules in the original model. The experimental results are shown in Table 3. Due to the early stopping mechanism, the baseline model stopped at 290 epochs, while the other models completed 300 epochs of training.

Table 3. Comparison results of ablation experiments.

A	B	C	P (%)	R (%)	mAP@0.5 (%)	Params (M)	FLOPs (G)
			0.884	0.86	0.891	3.01	8.1
√			0.927	0.836	0.913	3.28	8.3
√	√		0.916	0.847	0.909	3.04	7.9
√	√	√	0.918	0.86	0.917	3.04	7.9

By integrating the LSKA into the SPPF module, precision increased by 4.3%, and mAP@0.5 improved by 2.2%. The inclusion of an attention mechanism effectively enhanced the network's focus on crucial features while suppressing irrelevant information interference. However, recall experienced a slight decrease, and the results for Params and FLOPs indicated a slight increase in model complexity. Conversely, integrating the PKI Block maintained detection accuracy without decline, reducing Params by 8% and FLOPs by 5%, thereby effectively shrinking the model size. Additionally, incorporating the Aux head increased mAP@0.5 by 2.6% without increasing parameter or computational load, and recall improved to match the original model. Compared to the baseline YOLOv8, while the overall parameter count saw a slight increase, actual floating-point operations were reduced. In summary, all three methods positively influenced the detection performance of YOLOv8, validating these improvements through ablation experiments as shown in Figure 9.

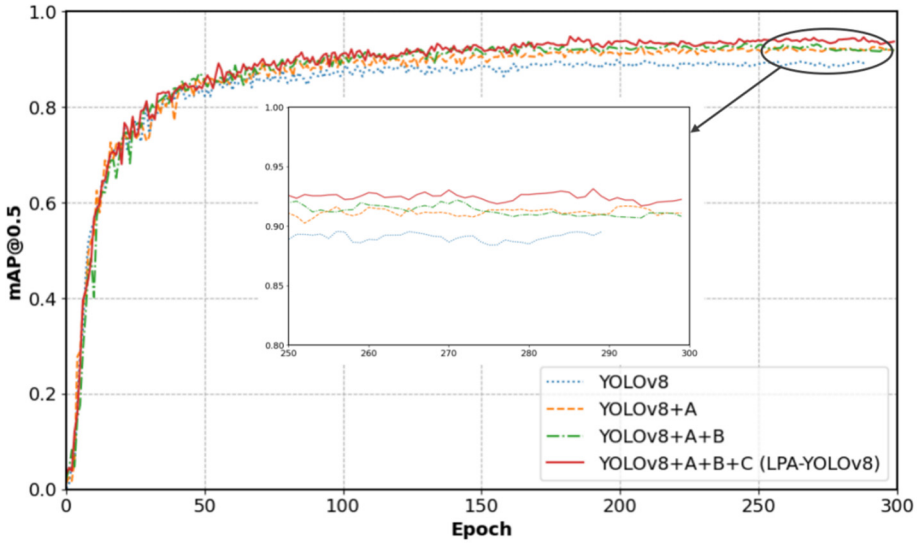


Figure 9. mAP@0.5 convergence curves in ablation experiments.

4.4. Comparative Analysis with Other Models

To evaluate the improved algorithm, we compared the LPA-YOLOv8 model with state-of-the-art object detection models including two-stage Faster R-CNN, Cascade R-CNN, and one-stage RetinaNet, as well as YOLO series models (YOLOv5n, YOLOv6n, YOLOv7, and YOLOv8n). Experimentation and validation were conducted on a custom agricultural obstacle dataset. The results are shown in Table 4, which compares these popular object detection models in terms of average precision (mAP), Floating-Point Operations per Second (FLOPs), and parameter count.

Table 4. Comparison with other popular models.

	Model	Backbone	mAP	Params	FLOPs	FPS
Two-stage	Faster R-CNN	ResNet50	0.636	41.37M	211.9 G	16.2
	Cascade R-CNN	ResNet50	0.72	69.21M	240.6 G	13.6
One-stage	RetinaNet	ResNet50	0.703	36.41M	211.9 G	18.6
	YOLOv5n	CSPDarknet53	0.831	2.50 M	7.1 G	66.7
	YOLOv6n	CSPDarknet53	0.923	4.23 M	11.8 G	69.3
	YOLOv7	CSPDarknet53	0.888	37.22 M	105.2 G	52.1
	YOLOv8n	CSPDarknet53	0.884	3.01M	8.1 G	73.9
Ours	**LPA-YOLOv8n**	**CSPDarknet53**	**0.918**	**3.04 M**	**7.9 G**	**71.5**

From the table, it is evident that despite using the ResNet50 backbone network, two-stage object detection algorithms such as Faster R-CNN and Cascade R-CNN did not achieve high accuracy. Instead, they significantly increased the number of parameters and floating-point operations. RetinaNet, a classic one-stage model utilizing the ResNet50 backbone, exhibits lower accuracy and higher parameter counts compared to YOLO series models. Among commonly used lightweight algorithms, YOLOv5 demonstrates certain advantages in model deployment but lacks prominent detection accuracy. Although YOLOv7 shows relatively good detection accuracy, its large model size makes it less suitable for deployment in perception systems. YOLOv6 manages to reduce model complexity while maintaining good detection accuracy; however, its parameter count and floating-point operations still exceed those of the YOLOv8 model. For an effective object detection system, it is crucial to ensure both accuracy and speed. Therefore, our proposed LPA-YOLOv8n model strikes a better balance between accuracy and model size compared to the aforementioned types of detection models.

To visually assess the performance of the improved algorithms in terms of detection accuracy and robustness, we evaluated lightweight models YOLOv5n, YOLOv6n, YOLOv8n, and LPA-YOLOv8n on the test set of our dataset. The results of the evaluations are presented in Figure 10.

(**a**) (**b**) (**c**) (**d**)

Figure 10. Detection results of five types of field obstacles across different models. (**a**) YOLOv5n; (**b**) YOLOv6n; (**c**) YOLOv8n; and (**d**) LPA-YOLOv8n.

From the comparative results shown in Figure 10, YOLOv5n exhibits relatively lower accuracy in rainy or snowy weather conditions with low visibility. It also shows certain instances of missed detections in livestock images where features are partially obscured. YOLOv6n improves detection accuracy compared to YOLOv5n; however, it demonstrates noticeable false detections in stone images, misidentifying tree trunks as poles. Moreover, it shows errors in detecting agricultural machinery during rainy or snowy weather. YOLOv8, overall, avoids false detections but still has instances of missed detections for obstacles partially obscured. In contrast, the improved LPA-YOLOv8n model accurately detects obstacles in dim lighting, rainy or snowy conditions, and environments with partially obscured features, significantly enhancing precision.

4.5. Accuracy Experiment of Localization System

To further validate the accuracy and robustness of the localization model, we conducted field tests in a real agricultural environment. The experiments were conducted at the Innovation Experimental Base of Academician Zhang Hongcheng in Guangling District, Yangzhou City, Jiangsu Province, China, on 1 June 2024, under partly cloudy to clear weather conditions. During the experiments, the binocular camera lenses were oriented

towards the forward agricultural field to detect and locate obstacles within a distance range of 2 to 10 m. Upon successful detection of obstacles, the three-dimensional coordinates of the obstacle centers were recorded. Actual distances Z_t were used as measurement data for evaluation and analysis. A high-precision laser rangefinder was used to verify and read the real distances Z_a. Distances were varied continuously, with measurements taken at intervals of 0.5 m within the 2 to 10 m range, and the above experiments were repeated and recorded. Finally, absolute errors and relative errors for the five sets of data obtained were analyzed, as shown in Table 5.

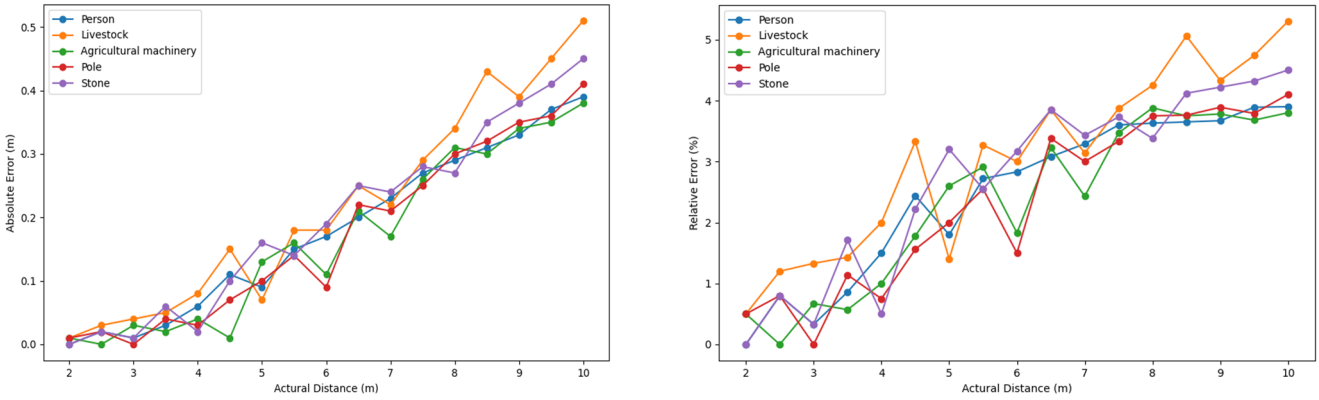
Table 5. Test results of locating accuracy of obstacles in field.

Z_a	Person			Livestock			Agricultural Machinery			Pole			Stone		
	Z_t	E_Z	E_{Z_r}	Z_t	E_Z	E_{Z_r}	Z_t	E_Z	E_{Z_r}	Z_t	E_Z	E_{Z_r}	Z_t	E_Z	E_{Z_r}
2.0	2.00	0.00	0.00	2.01	0.01	0.50	1.99	0.01	0.50	2.01	0.01	0.50	2.00	0.00	0.00
2.5	2.52	0.02	0.80	2.53	0.03	1.20	2.50	0.00	0.00	2.48	0.02	0.80	2.51	0.02	0.80
3.0	2.99	0.01	0.33	2.96	0.04	1.33	3.02	0.03	0.67	3.00	0.00	0.00	3.02	0.01	0.33
3.5	3.53	0.03	0.86	3.55	0.05	1.43	3.52	0.02	0.57	3.54	0.04	1.14	3.55	0.06	1.71
4.0	4.06	0.06	1.50	4.08	0.08	2.00	4.04	0.04	1.00	4.03	0.03	0.75	4.02	0.02	0.50
4.5	4.61	0.11	2.44	4.65	0.15	3.33	4.58	0.01	1.78	4.57	0.07	1.56	4.50	0.10	2.22
5.0	5.09	0.09	1.80	5.07	0.07	1.40	5.13	0.13	2.60	5.10	0.10	2.00	5.17	0.16	3.20
5.5	5.65	0.15	2.72	5.68	0.18	3.27	5.66	0.16	2.91	5.64	0.14	2.55	5.64	0.14	2.55
6.0	6.17	0.17	2.83	6.18	0.18	3.00	6.11	0.11	1.83	6.09	0.09	1.50	6.19	0.19	3.17
6.5	6.70	0.20	3.08	6.75	0.25	3.85	6.71	0.21	3.23	6.72	0.22	3.38	6.75	0.25	3.85
7.0	7.23	0.23	3.29	7.22	0.22	3.14	7.17	0.17	2.43	7.21	0.21	3.00	7.24	0.24	3.43
7.5	7.77	0.27	3.60	7.79	0.29	3.87	7.76	0.26	3.47	7.75	0.25	3.33	7.78	0.28	3.73
8.0	8.29	0.29	3.63	8.34	0.34	4.25	8.32	0.31	3.88	8.30	0.30	3.75	8.27	0.27	3.38
8.5	8.81	0.31	3.65	8.83	0.43	5.06	8.80	0.30	3.75	8.82	0.32	3.76	8.85	0.35	4.12
9.0	9.33	0.33	3.67	9.39	0.39	4.33	9.34	0.34	3.78	9.35	0.35	3.89	9.38	0.38	4.22
9.5	9.87	0.37	3.89	9.95	0.45	4.74	9.85	0.35	3.68	9.86	0.36	3.79	9.91	0.41	4.32
10.0	10.39	0.39	3.90	10.51	0.51	5.30	10.38	0.38	3.80	10.41	0.41	4.10	10.45	0.45	4.50
Mean value		0.18	2.47		0.22	3.06		0.17	2.35		0.17	2.34		0.20	2.71
Max value		0.39	3.90		0.53	5.30		0.38	3.88		0.41	4.10		0.45	4.50

Note: Z_a is the actual distance; Z_t is the test distance; E_Z is the absolute error; and E_{Z_r} is the relative percentage of error.

The experimental results were plotted to create comparative statistical graphs, as shown in Figure 11. The results indicate that within the range of 2 to 10 m, the maximum average absolute error and maximum average relative error in measured distances were 0.22 m and 3.06%, respectively. The maximum absolute error and maximum relative error in measured distances were 0.53 m and 5.30%, respectively. Higher relative errors were observed in the positioning of livestock and stones, primarily due to occlusion by crops and the greater mobility of livestock. Additionally, Table 5 shows that measurement results are relatively accurate at close to medium distances from obstacles, with errors increasing gradually as the distance between obstacles and the camera increases.

Based on the field obstacle localization accuracy test results in Table 5, a comparative statistical graph was plotted as shown in Figure 11. From Figure 11a, it is evident that the test distances for the five types of obstacles fluctuate above the actual distances, and as the distance between obstacles and the camera increases, measurement errors also increase. This is primarily due to the limitations of camera resolution, which results in less precise depth information capture of distant obstacles. Additionally, outdoor lighting conditions affect the camera, leading to some background noise and partial loss of obstacle details, thereby affecting ranging accuracy. As depicted in Figure 11b, the relative error is less than 3% within a range of 5 m and less than 4% within 8 m, indicating that overall measurement errors meet real-time ranging requirements.



(**a**) (**b**)

Figure 11. Statistical comparison graph of localization experiment results. (**a**) Absolute error statistics; (**b**) relative error statistics.

5. Discussion

In this study, we developed the LPA-YOLOv8n model based on the YOLOv8 architecture. The improvements include incorporating the LSKA attention mechanism into the SPPF module of the backbone, replacing the Bottleneck block in the C2f module with the PKI Block, and introducing an auxiliary detection head in the head part. We also utilized binocular cameras to obtain the position information of obstacles. The experimental results demonstrate that compared to the baseline YOLOv8n model, LPA-YOLOv8n improved the precision of obstacle detection in agricultural fields by 3.4%. The issues of missed detections and false positives were effectively mitigated. Measured by mAP@0.5, the model's performance increased from 89.1% to 91.7%, indicating an improvement in the recognition accuracy of five types of obstacles. Furthermore, in the obstacle localization accuracy test, the average absolute error and average relative error for distances between 2 to 10 m were 0.22 m and 3.06%, respectively, confirming that the designed detection and localization system can meet the requirements for detecting and locating common obstacles in agricultural environments.

The LPA-YOLOv8n model presented in this paper demonstrates effective detection and localization accuracy for field obstacles. However, there is still room for improvement and refinement in the system's performance. Currently, the main sources of potential errors and limitations are as follows:

(1) Limited Interpretability. The black-box nature of convolutional neural networks (CNNs) results in a lack of transparency in the model's decision-making process, making it difficult to explain misclassifications and omissions in specific scenarios. This limitation hampers the precise identification of areas for model improvement.
(2) Model Complexity and Computational Resources. Although the model shows improvements in detection accuracy, its high computational complexity may hinder efficient operation on resource-constrained devices. This could impact the real-time performance and deployment feasibility of the model, particularly in autonomous agricultural machinery systems, where real-time obstacle avoidance might not be achievable.
(3) Insufficient Data Diversity. The dataset may lack sufficient diversity in lighting conditions, weather, and seasons, leading to limited generalization ability of the model under various environmental conditions. This can result in decreased detection performance in certain specific environments, affecting the reliability of the model in practical applications.

To address these issues and further enhance the performance and applicability of LPA-YOLOv8n, future research directions include the following aspects:

(1) Enhancing the interpretability of the model. Convolutional neural networks (CNNs) have limited interpretability; thus, heatmaps are utilized to evaluate the pros and cons

of different positions in the field obstacle detection process, as shown in Figure 12. The intensity of red in the heatmap indicates the extent of its impact on target localization. Additionally, heatmaps can be used to analyze the specific impact of enhancement modules on model performance.

(2) Model lightweighting. In the next step, we will deploy the algorithm in an autonomous agricultural machinery system, providing a reference for future research on autonomous obstacle avoidance in uncrewed agricultural vehicles. Thus, the model requires lightweighting. The improved model in this study emphasizes enhanced accuracy, but in resource-constrained environments, the model must be optimized for computational efficiency and reduced memory usage while maintaining high precision and real-time performance.

(3) Expanding dataset quantity. In real-world applications, LPA-YOLOv8n may encounter missed detections, particularly in scenarios with multiple overlapping targets. In such cases, occlusion between targets can prevent the model from accurately detecting and recognizing all targets, thereby affecting detection precision and reliability. Therefore, enriching and expanding the dataset is crucial.

(**a**) (**b**) (**c**)

Figure 12. Heatmap: (**a**) livestock area visualization; (**b**) agricultural machinery area visualization; and (**c**) pole area visualization.

6. Conclusions

This paper presents a field obstacle detection algorithm, LPA-YOLOv8n, based on an enhanced YOLOv8 framework, tailored for detecting typical field obstacles in complex agricultural environments. The algorithm integrates LSKA, PKI, and Aux head structures. The results from model training and validation substantiate the effectiveness of the proposed approach. The improved model utilizes color images captured by RealSense D435i binocular cameras to detect obstacles and obtain their three-dimensional coordinates in the field coordinate system. Experimental findings demonstrate that the LPA-YOLOv8n model achieves an optimal balance between parameter count, computational load, and detection accuracy. It significantly enhances detection precision without increasing parameters, thus reducing computational demands. Moreover, the improved model exhibits promising results in localization accuracy tests, meeting real-time detection and positioning requirements for typical field obstacles in challenging environments with limited computational resources. The enhanced LPA-YOLOv8 model developed in this study significantly improves the autonomous obstacle avoidance capabilities of driverless agricultural machinery, reducing the need for human intervention while increasing operational efficiency and safety. This advancement facilitates the smart and modernized development of agricultural production. The refined model can also be extended to applications in field environment monitoring, such as pest and disease detection and crop growth monitoring, providing intelligent solutions for agricultural management. Furthermore, the designed detection and localization system not only achieves cost savings and efficiency improvements but also reduces the complexity of operating driverless agricultural machinery for farmers. Additionally, it offers valuable insights for obstacle detection and localization in other scenarios.

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

Funding: This research was funded by the Agricultural Science and Technology Innovation Program of the Chinese Academy of Agricultural Sciences (31-NIAM-05) and The Jiangsu Province Science and Technology Support Program, China (BE2023332-2) and The National Key Research and Development Program of China (2022YFD2001404).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in the article.

Acknowledgments: The authors thank the editor and anonymous reviewers for providing helpful suggestions for improving the quality of this manuscript.

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

References

1. Li, S.; Xu, H.; Ji, Y.; Cao, R.; Zhang, M.; Li, H. Development of a following agricultural machinery automatic navigation system. *Comput. Electron. Agric.* **2019**, *158*, 335–344. [CrossRef]
2. Shang, Y.H.; Wang, H.; Qin, W.C.; Wang, Q.; Liu, H.Y.; Yin, Y.X.; Song, Z.H.; Meng, Z.J. Design and Test of Obstacle Detection and Harvester Pre-Collision System Based on 2D Lidar. *Agronomy* **2023**, *13*, 388. [CrossRef]
3. Shang, Y.H.; Zhang, G.Q.; Meng, Z.J.; Wang, H.; Su, C.H.; Song, Z.H. Field Obstacle Detection Method of 3D LiDAR Point Cloud Based on Euclidean Clustering. *Trans. CSAM* **2022**, *53*, 23–32.
4. Xue, J.L.; Cheng, F.; Li, Y.Q.; Song, Y.; Ma, T.H. Detection of Farmland Obstacles Based on an Improved YOLOv5s Algorithm by Using CIoU and Anchor Box Scale Clustering. *Sensors* **2022**, *22*, 1790. [CrossRef] [PubMed]
5. Skoczeń, M.; Ochman, M.; Spyra, K.; Nikodem, M.; Krata, D.; Panek, M.; Pawłowski, A. Obstacle Detection System for Agricultural Mobile Robot Application Using RGB-D Cameras. *Sensors* **2021**, *21*, 5292. [CrossRef]
6. Wen, Y.; Xue, J.L.; Sun, H.; Song, Y.; Lv, P.F.; Liu, S.H.; Chu, Y.Y.; Zhang, T.Y. High-precision target ranging in complex orchard scenes by utilizing semantic segmentation results and binocular vision. *Comput. Electron. Agric.* **2023**, *215*, 108440. [CrossRef]
7. Yang, W.J.; Wu, J.C.; Zhang, J.L.; Gao, K.; Du, R.H.; Wu, Z.; Firkat, E.; Li, D.W. Deformable convolution and coordinate attention for fast cattle detection. *Comput. Electron. Agric.* **2023**, *211*, 108006. [CrossRef]
8. Cao, Y.K.; Pang, D.D.; Zhao, Q.C.; Yan, Y.; Jiang, Y.Q.; Tian, C.Y.; Wang, F.; Li, J.Y. Improved YOLOv8-GD deep learning model for defect detection in electroluminescence images of solar photovoltaic modules. *Eng. Appl. Artif. Intel.* **2024**, *131*, 107866. [CrossRef]
9. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
10. Girshick, R. Fast R-CNN. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
11. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE TPAMI* **2017**, *39*, 1137–1149. [CrossRef]
12. Cai, Z.; Vasconcelos, N. Cascade R-CNN: High quality object detection and instance segmentation. *IEEE TPAMI* **2019**, *43*, 1483–1498. [CrossRef]
13. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
14. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, D.; Reed, S.; Fu, C.Y.; Berg, A.C. *SSD: Singleshot Multibox Detector*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 21–37.
15. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2999–3007.
16. Liu, H.; Zhang, L.; Chen, Y.; Zhang, J.; Wu, B. Real-time Pedestrian Detection in Orchard Based on Improved SSD. *Trans. Chin. Soc. Agri. Mach.* **2019**, *50*, 29–35.
17. Peng, H.X.; Chen, H.; Zhang, X.; Liu, H.N.; Chen, K.Y. Retinanet_G2S: A multi-scale feature fusion-based network for fruit detection of punna navel oranges in complex field environments. *Precis. Agric.* **2024**, *25*, 889–913. [CrossRef]
18. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
19. Glenn, J. YOLOv5 Release v6.1. 2022. Available online: https://github.com/ultralytics/yolov5/releases/tag/v6.1 (accessed on 9 June 2023).

20. Li, C.Y.; Li, L.L.; Jiang, H.L.; Weng, K.H.; Geng, Y.F.; Li, L.; Ke, Z.D.; Li, Q.Y.; Cheng, M.; Nie, W.Q.; et al. YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications. *arXiv* **2022**, arXiv:2209.02976.
21. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv* **2022**, arXiv:2207.02696.
22. Glenn, J. Ultralytics YOLOv8. 2022. Available online: https://github.com/ultralytics/ultralytics (accessed on 26 May 2023).
23. Li, W.T.; Zhang, Y.; Mo, J.Q.; Li, Y.M.; Liu, C.L. Detection of Pedestrian and Agricultural Vehicles in Field Based on Improved YOLOv3—Tiny. *Trans. CSAM* **2020**, *51* (Suppl. S1), 1–8+33.
24. Khow, Z.J.; Tan, Y.F.; Karim, H.A.; Rashid, H.A.A. Improved YOLOv8 Model for a Comprehensive Approach to Object Detection and Distance Estimation. *IEEE Access* **2024**, *12*, 63754–63767. [CrossRef]
25. Yan, X.; Chen, B.; Liu, M.; Zhao, Y.; Xu, L. Inclined Obstacle Recognition and Ranging Method in Farmland Based on Improved YOLOv8. *World Electr. Veh. J.* **2024**, *15*, 104. [CrossRef]
26. Mish, M.D. A self regularized non-monotonic neural activation function. *arXiv* **2019**, arXiv:1908.08681.
27. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.M.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 936–944.
28. Liu, S.; Qi, L.; Qin, H.F.; Shi, J.P.; Jia, J.Y. Path aggregation network for instance segmentation. In Proceedings of the 2018, IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 8759–8768.
29. Guo, M.H.; Lu, C.Z.; Liu, Z.N.; Cheng, M.M.; Hu, S.M. Visual Attention Network. *arXiv* **2022**, arXiv:2202.09741. [CrossRef]
30. Lau, K.W.; Po, L.M.; Rehman, Y.A.U. Large Separable Kernel Attention: Rethinking the Large Kernel Attention design in CNN. *Expert. Syst. Appl.* **2023**, *236*, 121352. [CrossRef]
31. Cai, X.H.; Lai, Q.X.; Wang, Y.W.; Wang, W.G.; Sun, Z.R.; Yao, Y.Z. Poly Kernel Inception Network for Remote Sensing Detection. *arXiv* **2024**, arXiv:2403.06258.
32. Jiang, T.Y.; Li, Z.; Zhao, J.; An, C.G.; Tan, H.; Wang, C.L. An Improved Safety Belt Detection Algorithm for High-Altitude Work Based on YOLOv8. *Electronics* **2024**, *13*, 850. [CrossRef]