IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 22 December 2023, accepted 3 February 2024, date of publication 13 February 2024, date of current version 26 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3365555

RESEARCH ARTICLE

An Improved Method for Extracting Inter-Row Navigation Lines in Nighttime Maize Crops Using YOLOv7-Tiny

HAILIANG GONG AND WEIDONG ZHUANG
College of Engineering, Heilongjiang Bayi Agricultural University, Daqing 163319, China

Corresponding author: Weidong Zhuang (81nd@163.com)

This work was supported in part by the Monitoring and Analysis Project for the Development of Key Industries in Agricultural Reclamation by the Agricultural Reclamation Bureau of the Ministry of Agriculture and Rural Affairs under Project 18220122, in part by Heilongjiang Provincial Key Research and Development Program under Grant 2023ZXJ07B02, and in part by the "Three Verticals" Basic Cultivation Program under Grant ZRCPY202306.

ABSTRACT In response to the issue of insufficient nighttime illumination in mechanical weeding of maize crops, this study proposes an improved YOLOv7-tiny network model infrared image object detection. The model incorporates the ShuffleNet v1 network to reduce computational complexity, enhance image feature extraction, and obtain more comprehensive semantic information. Additionally, the Coordinate Attention(CA) mechanism module is integrated into the neck network to improve sample detection performance. The EIOU loss function is employed to replace the original loss function, which results in faster model convergence and improved positioning accuracy. The improved YOLOv7-tiny network model is used to detect maize seedlings, with the center point of the detection box serving as the navigation reference point. Subsequently, the least squares method is used to fit the maize rows on both sides, thereby obtaining the inter-row navigation line in the middle of the two rows. Experimental results demonstrate that the improved YOLOv7-tiny network model achieves a detection accuracy of 94.21 % and a detection speed of 32.4 frames per second, enabling accurate identification of maize seedlings at night. The average error between the extracted positioning reference points and the manually labeled midpoint of the maize seedlings is 4.85 cm, meeting navigation requirements of maize crop rows and providing feasibility for deployment on mobile terminal devices.

INDEX TERMS YOLOv7-tiny, object detection, inter-row navigation line, ShuffleNet v1, attention mechanism, loss function.

I. INTRODUCTION

The mechanization of maize cultivation has experienced a paradigm shift with the emergence of intelligent weed control machinery, drastically reducing the dependence on chemical pesticides and fostering the sustainable development of maize crops [1], [2]. At the core of these systems' success is the precise detection and navigation along maize row lines, which facilitates accurate weeding operations and substantially reduces the necessity for manual labor [3], [4], [5]. The ability to perform mechanical weeding nocturnally not only enhances productivity but also serves to safeguard early-stage maize seedlings from harm and to counteract the environmental stressors experienced during daylight hours [6], [7]. Despite the impressive advancements in agricultural automation, intelligent weed control systems continue to face challenges when operating in the less than ideal nocturnal lighting conditions [8], [9]. The diminished intensity of night lighting can lead to a loss of clarity in target edges and details, resulting in decreased accuracy of detection. Insufficient illumination can introduce noise and blur into images, causing targets to merge indistinguishably with the background and thereby adversely affecting the recognition and localization of the targets.

The associate editor coordinating the review of this manuscript and approving it for publication was Zhe Xiao.

Infrared imaging technology has emerged as a promising solution to the low-light challenge, with its ability to render clear images in the absence of sufficient visible light by capturing infrared radiation [10]. Traditional target detection models, however, often underperform in the complex and variable conditions of nighttime agriculture due to their reliance on manual feature extraction and prior knowledge [11]. In contrast, machine learning, and in particular deep learning techniques, have shown superior performance in feature learning for target detection and recognition, with convolutional neural networks leading the charge [12].The literature reveals various efforts to harness these advanced technologies for agricultural applications. Xue et al. [13] and Qing et al [14] have explored the use of infrared imaging in night scenes, with Qing et al. developing lightweight CNN models suitable for mobile and embedded systems. Li [15] the author has advanced this approach by utilizing the YOLOv3 network model to overcome the limitations of conventional cameras, especially in detecting overlapping pedestrians in low-light conditions.

Specifically focusing on agricultural settings, Sa et al. [16], the authors have experimented with pixel-based fusion techniques such as Laplacian Pyramid Transform (LPT) and fuzzy logic to enhance fruit detection using RGB and near-infrared (NIR) images. Tian et al. [17] have proposed a dual-input network to sense human shapes in agricultural fields, combining RGB and far-infrared (FIR) images for improved safety. In the realm of crop row fitting, which is vital for automated weed control, Zhang et al [18] have utilized YOLOv3 network modelfor target extraction of rice seedlings, while Peng et al. [19] have developed an enhanced YOLOv7 network model tailored for detecting navigation lines across diverse orchard and crop row settings, these approaches have yet to comprehensively address the intricacies associated with varying growth stages and environmental conditions. To address issues related to lighting and weed interference, Liu et al. [20] proposed an improved Multi-Scale Efficient Residual Factorized ConvNet (MS-ERFNet) model for the recognition of seedling maize crop rows. The approach also involved utilizing the Least Squares Method to fit the centerlines. Yang et al. [21] combined the YOLOv5 network model with the ExG (excess green) method and Otsu method for navigation line recognition in maize crop rows, achieving line fitting with the least squares method, but still facing limitations related to crop growth and environmental factors.

Within the framework of the aforementioned research context, this study is dedicated to mitigating the impact of inadequate nocturnal illumination on crop identification, with an extended investigation into the recognition of maize crop rows via night-time infrared imaging. By integrating an augmented YOLOv7-tiny network model, we endeavor to employ bounding boxes anchored by the bottom center coordinates coupled with the least squares method for enhanced precision in localization, thereby accommodating the alignment of navigation lines with maize crop rows during nocturnal operations. Our model is meticulously tailored for deployment on resource-constrained mobile terminals, thus fostering efficient night-time weed management, inaugurating novel operational paradigms, and contributing fresh perspectives for the nocturnal weeding machinery workflow.

II. YOLOV7-TINY NETWORK MODEL

Wang et al. [22], the authors proposed the YOLOv7 network model, which is an optimized version of the YOLOv5 network model and represents the latest network model in the YOLO series. YOLOv7 network model exhibits significant improvements in both detection accuracy and speed compared to YOLOv5 network model [23]. However, the complexity of the network architecture and the large number of parameters in YOLOv7 network model make it demanding on device performance, rendering it unsuitable for edge terminal devices [24], [25]. To address this issue, the researchers designed the YOLOv7-tiny network model based on YOLOv7 network model [26], [27], which features a simplified structure specifically tailored for edge GPU devices [28]. YOLOv7-tiny network model consists of three components: the backbone network, the neck network, and the prediction head, as illustrated in Fig. 1.

In the Backbone section, a more concise ELAN is employed instead of E-ELAN, and the convolution operation in MPConv is removed, using only pooling for downsampling. Simultaneously, the optimized SPP structure is retained to provide richer feature maps for utilization in the Neck layer. In the Neck section, the PANet structure is still employed for feature aggregation. In the Head section, standard convolution is used for channel adjustment, replacing REPConv [29]. Compared to YOLOv7 network model, YOLOv7-tiny network model sacrifices some accuracy but gains advantages in terms of speed and lightweight design. Nonetheless, YOLOv7-tiny model still has some limitations [30].

Firstly, in the Backbone section, a significant utilization of ELAN networks is observed, with each ELAN network comprising multiple densely connected standard convolutions. This results in a complex network structure with an excessive number of computations and parameters [31]. Moreover, the network has a relatively limited number of layers, which hampers effective feature extraction.

Secondly, the model employs the Leaky ReLU activation function throughout, which proves to be suboptimal when propagating features downwards. As the model depth increases, the gradient updates at each point become progressively less smooth, thereby adversely affecting classification accuracy.

Lastly, in the Neck section, the continued use of ELAN networks for feature aggregation can lead to redundant features. Therefore, this study proposes a more lightweight approach for feature aggregation, aiming to reduce both parameter count and computational overhead while ensuring the preservation of rich feature representations.

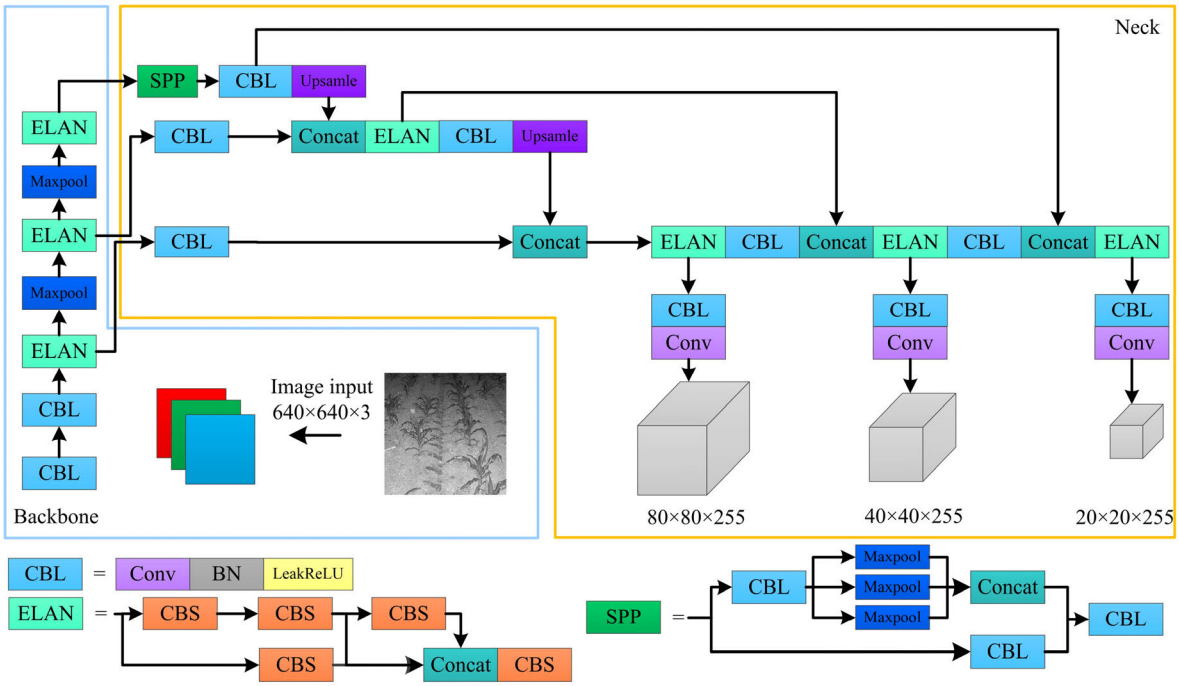

FIGURE 1. YOLOv7-tiny networks model structure.

III. IMPROVED THE YOLOV7-TINY NETWORK MODEL

A. IMPROVEMENT OF THE BACKBONE NETWORK

To address the aforementioned shortcomings of YOLOv7-tiny network model, improvements were made. This study draws inspiration from ShuffleNet v1, a lightweight network for image classification, to improve the Backbone structure by reducing dense connections and increasing network depth. Reducing dense connections helps decrease computational load, while increasing network depth allows for more comprehensive feature extraction [32]. The concept of channel shuffling is introduced, which uniformly shuffles the information from different channels in the input feature map, addressing the issue of limited interaction between groups in group convolution [33]. Furthermore, the network combines group convolution with depth-wise separable convolution to further reduce model parameters, effectively decreasing the computational load of traditional convolutional neural networks. The overall architecture is inspired by residual networks, and the network depth is moderately increased to enhance its learning capacity [34]. However, due to the lightweight design, the main network sacrifices some parameters and may not capture complete semantic information, resulting in room for improvement in terms of accuracy. To further enhance network performance, the proposed approach involves improving the network structure to address this issue.

The calculation of the number of parameters introduced by each type of convolution operation is as follows:

Suppose the width and height of the input feature map are W_i and H_i, respectively, with C_i channels, and the output has C_o channels. The size of the convolution kernel is $K \times K$.

Therefore, the number of parameters under a standard convolution (sconv) is as shown in Equation (1).

$$P_{sc} = C_i \times C_o \times K \times K \tag{1}$$

Group convolution (GConv) builds upon the standard convolution by dividing the convolution kernels and input channels into g groups. Each group of kernels convolves with the feature map independently, resulting in the following parameter count as shown in Equation (2).

$$P_{GC} = C_i/g \times C_o \times K \times K \tag{2}$$

Depth-wise separable convolution (DWConv) involves channel-wise convolution operations on the input feature map, with the number of convolution kernels equal to the number of input channels. The parameter count for depth-wise separable convolution is as follows, represented by Equation (3).

$$P_{DW} = C_i \times K \times K \tag{3}$$

Among these convolution operations, standard convolution has the highest number of parameters. Group convolution has $1/g$ of that, and depth-wise separable convolution has the least, only $1/C_o$ of the standard convolution's parameters. Therefore, combining group convolution with depth-wise separable convolution can significantly reduce the network's parameter count and computational requirements. However, this combination may also lead to a substantial loss of semantic information. Fig. 2 illustrates the basic modules of ShuffleNet v1 network model, consisting of Unit (a) with a stride of 1 and Unit (b) with a stride of 2.

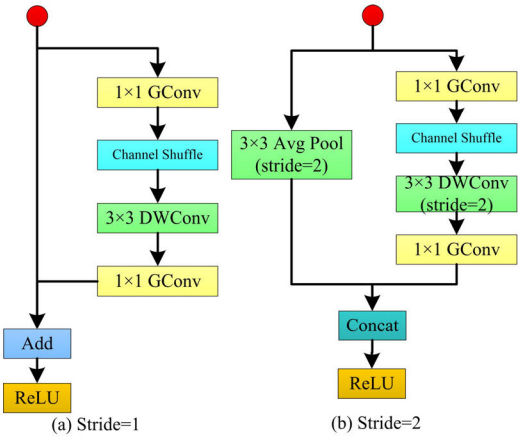

FIGURE 2. ShuffleNet v1 network basic model.

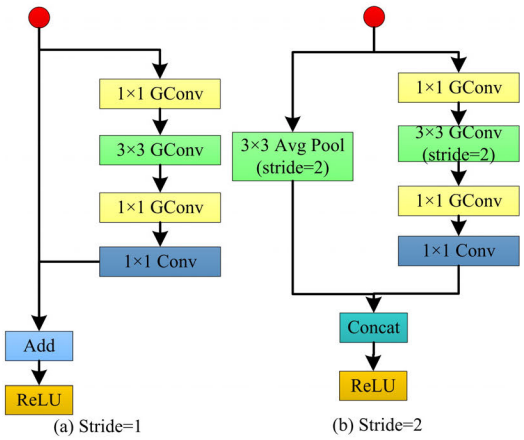

FIGURE 3. Basic moudle of improved ShuffleNet v1.

In this study, the network structure includes two basic modules: Unit (a) and Unit (b). In Unit (a), the network is divided into two branches. The left branch remains unchanged, while the right branch undergoes group convolution and depth-wise separable convolution operations, along with batch normalization and channel shuffling. Finally, the feature maps from both branches are fused through an element-wise addition operation.

In Unit (b), similarly, there are two branches, with downsampling operations applied to both the left and right sides. The left branch reduces the feature map size by half using average pooling, while the right branch performs a convolution operation with a stride of 2. Finally, the feature maps from both branches are concatenated to double the dimensionality. By interleaving the usage of these two basic modules, the ShuffleNet v1 network is constructed, achieving lightweight feature extraction.

Investigative scrutiny of ShuffleNet v1 network's primary modules reveals that the input features, upon entering the right pathway, are first subject to a group convolution (GConv), succeeded by channel shuffling. This is followed by a 3×3 depth-wise separable convolution (DWConv), culminating in a subsequent group convolution. This progression leads to a substantial reduction in parameters within the network, albeit with a commensurate diminution in semantic detail and a slight compromise in accuracy. Depth-wise separable convolution, characterized by per-channel convolutional operations, substantially lowers parameter numbers but lacks inter-channel communication, resulting in an output deficient in feature richness. Channel shuffling aims to ameliorate this by randomly reordering channel positions, yet it does not substantially enhance the semantic interchange among channels.

To address these shortcomings, modifications to the right branch of Fig. 2' s modules have been proposed. Initially, replacing depth-wise separable convolutions with group convolution modules marginally increases parameters while fostering inter-channel information flow. Subsequently, channel shuffling is supplanted by a 1×1 standard convolution placed at the branch's terminus. This standard convolution not only fulfills a comparable role but also amplifies the semantic richness of the feature map without accruing additional parameters. Fig. 3 delineates the revised ShuffleNet v1 module.

B. COORDINATE ATTENTION MECHANISM

The attention mechanism is a data processing method that can be applied to the task of maize crop recognition. By dynamically weighting the inputs, the attention mechanism can emphasize the relevant regions while suppressing the irrelevant background regions. In the context of maize crop recognition, the spatial attention mechanism can be employed to focus on the regions related to maize seedlings. The attention mechanism enables accurate attention calculation for complex target regions, such as occlusions between leaves and residues covering the stems. By leveraging the spatial attention mechanism, the features of maize seedlings can be effectively extracted, leading to accurate recognition.

Contemporary attention frameworks often resort to global max or average pooling, which risks forfeiting the spatial delineation of objects. The Squeeze-and-Excitation Network (SE) attention mechanism module, for instance, is preoccupied with fostering inter-channel dependencies, thus sidelining spatial attributes. The Efficient Channel Attention(ECA) module evolves from SE by advocating a one-dimensional convolution technique to somewhat counteract the data distortion caused by fully connected layers' dimensionality reduction, yet it encounters limitations in managing global dependencies and the interplay of channel and spatial realms. The Convolutional Block Attention Module (CBAM) introduces expansive convolutional kernels for spatial feature extraction, yet it neglects long-range dependencies and is marred by considerable computational demands and augmented complexity. Conversely, the Channel Attention (CA) module contemplates both channel and spatial dimensions, assimilating adaptive channel weights to accentuate pertinent channel information, thereby refining the model's focal precision.

This study proposes the incorporation of the CA module scheme into the YOLOv7-tiny framework, orchestrating

feature maps along both vertical and horizontal axes via global average pooling to forge discrete directional feature maps. These are transmuted into dual attention maps, each ensnaring the extensive spatial correlations along a singular spatial trajectory of the input feature map. The application of these attention maps to the input feature map through multiplication serves to underscore salient representations. This module is proficient in capturing inter-channel data while remaining attuned to positional and directional nuances, which bolsters the model's discernment and pinpointing faculties [35]. Deploying the CA module to the triad of potent feature strata procured from the mainstay network, as well as post-upsampling, amplifies the model's representational acumen, curtails distractions from non-essential targets, and intensifies the recognition and positioning of pertinent targets, thus elevating the network's aggregate detection accuracy. The CA module is adeptly applied to a maize seedling target detection model [36], with the network architecture depicted in Fig. 4. By embedding spatial locational intelligence into channel attention, the CA mechanism module not only curtails computational expenditure but also magnifies the feature extraction potency of targets. The induction of the CA module has been demonstrated to concurrently enhance the precision and expedition of maize seedling recognition.

In the embedding of coordinate information, the input feature map with dimensions $C \times H \times W$ undergoes directional pooling along the X and Y axes to generate attention feature maps Z^h and Z^w with dimensions $C \times 1 \times W$ and $C \times H \times 1$, respectively.

To encapsulate remote spatial interactions with precise locational data, global average pooling is deconstructed as illustrated in Equation (4):

$$\begin{cases} Z_c^h(h) = \dfrac{1}{\mathrm{W}} \displaystyle\sum_{0 \le i < W} x_c(h, i) \\ Z_c^w(w) = \dfrac{1}{\mathrm{H}} \displaystyle\sum_{0 \le i < W} x_c(j, w) \end{cases} \tag{4}$$

where x_c signifies the c-th channel of the input feature X, with h and w denoting the feature map's height and width during model training.

The feature maps Z^h and Z^w are then concatenated, subjected to the F1 operation (utilizing a 1×1 convolution for reduction in dimensionality) followed by an activation function, yielding the feature map f, where ($f \in R^{C/r \times (H+W) \times 1}$), with r representing the downsampling stride to regulate the CA module's dimensionality, as presented in Equation (5):

$$f = \delta(F1([Z^h, Z^w])) \tag{5}$$

In this context, δ denotes the nonlinear activation function.

The feature map f is then divided along the spatial axis into two distinct feature maps $f^h \in R^{C/r \times H \times 1}$ and $f^w \in R^{C/r \times 1 \times W}$. These are subsequently upscaled through two 1×1 convolutions in conjunction with an activation function, culminating in the generation of the attention weight maps

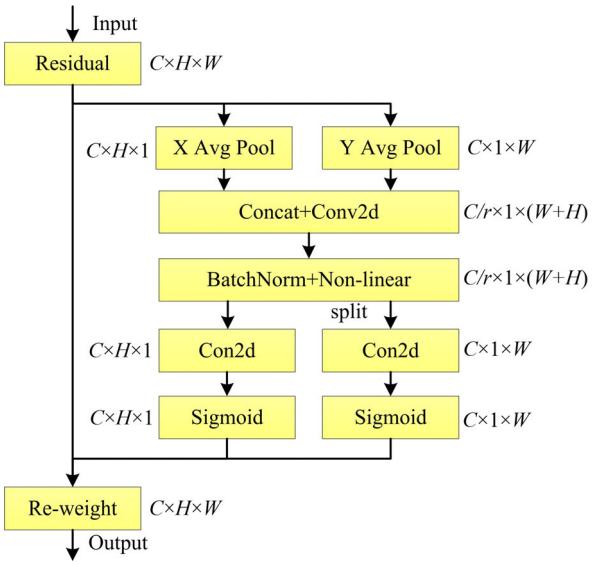

FIGURE 4. Coordinate attention module. Note: C, H, and W for channel number, height, and width, respectively. r is the reduction factor.

(g^w) and (g^h), as indicated in Equation (6):

$$\begin{cases} g^w = \sigma(F_w(f^w)) \\ g^h = \sigma(F_h(f^h)) \end{cases} \tag{6}$$

In the final step, (g^w)) and (g^h) are expanded, and through matrix multiplication, the output is derived as described in the ensuing formula.

C. EIOU LOSS FUNCTION

Crop recognition during nighttime constitutes a formidable task due to the dimly lit conditions which obscure the edges and details of target objects. Hence, accurate target localization is crucial in nocturnal crop recognition [37], [38]. The YOLOv7-tiny network model incorporates a loss function that consists of classification loss, localization loss, and confidence loss. By applying distinct weighting coefficients to these three losses, outcomes of varying emphasis can be achieved. Within YOLOv7-tiny network model, the coordinate loss is computed using the complete intersection over union (CIOU) metric, as shown in Equation (7) and Equation (8).

$$CIOU = IOU - \frac{\rho(b, b^{gt})}{c^2} - \alpha v \tag{7}$$

$$L_{CIOU} = 1 - IOU + \frac{\rho(b, b^{gt})}{c^2} + \alpha v \tag{8}$$

In the equation, $\frac{\rho(b,b^{gt})}{c^2}$ represents the penalty term, while b and b^{gt} respectively denote the center points of the predicted box and target box. The variable ρ represents the Euclidean distance between the two points, and c represents the distance between the diagonal of the minimum enclosing rectangle formed by the predicted box and the target box. The parameter a represents α positive weight balancing factor, and v represents the measurement of consistency in the aspect ratio between the predicted box and the target box.

Regarding parameters α and v, as shown in Equation (9) and Equation (10).

$$\alpha = \frac{v}{1 - IOU + v} \tag{9}$$

$$v = \frac{4}{\pi^2}(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h})^2 \tag{10}$$

However, equation (10) solely reflects the disparity in aspect ratios. While CIOU loss function considers the shape similarity between the predicted and true bounding boxes, it disregards the differences in width and height. Consequently, CIOU loss function may optimize similarity in an unreasonable manner. The Enhanced Intersection over Union(EIOU) loss function addresses this by introducing additional terms for distance loss and aspect ratio loss. Distance loss quantifies the Euclidean distance between the centers of the predicted and true boxes, optimizing the coordinates of the predicted box center. Aspect ratio loss computes the L1 distance between the widths and heights of the predicted and true boxes. Minimizing these discrepancies enables the model to achieve more precise boundary localization of the target. Compared to the inclusion of solely IOU loss function, the EIOU loss function provides a more comprehensive and robust supervisory signal by incorporating distance and aspect ratio losses. This benefits the model's concurrent learning across classification, confidence, and localization. Thus, the EIOU loss function supersedes CIOU loss function as the localization loss function in the YOLOv7-tiny network model, as defined in Equation (11).

$$L_{EIOU} = 1 - IOU + \frac{\rho^2(b, b^{gt})}{(w^c)^2 + (h^c)^2} + \frac{\rho^2(w, w^{gt})}{(w^c)^2} + \frac{\rho^2(h, h^{gt})}{(h^c)^2} \tag{11}$$

The EIOU loss function comprises three components, with the first two inherited from the geometric considerations of CIOU loss function. In the computation of aspect ratio loss, EIOU loss function separately calculates the impact of length and width for the predicted and true boxes, minimizing the differences between their widths and lengths, which accelerates convergence. Given the presence of numerous targets of varying scales in infrared scenes, the EIOU loss function facilitates faster model convergence during training. As it accounts for multiple types of errors, providing a more comprehensive error metric, the model can learn from various perspectives simultaneously. During the inference phase, the EIOU loss function enhances localization accuracy. With more precise predictions of box centers and sizes, the extracted reference points for localization align more closely with manually annotated points, thereby improving the effectiveness of subsequent crop navigation line fitting.

D. INTEGRATION OF IMPROVED MODULES

In summary, the YOLOv7-tiny network model has undergone improvements by integrating a revised network structure, as illustrated in Fig. 5. The original CSPDarknet53 module has been replaced with the lightweight ShuffleNet v1 backbone network. The CA attention mechanism is employed to enhance feature maps of various sizes extracted prior to feature fusion. The EIOU loss function is utilized, and the Leaky ReLU activation function is substituted with SiLU. Following parameter adaptation and adjustment, the integration of the enhanced module with the YOLOv7-tiny object detection network is successfully accomplished.

IV. RESULTS

A. DATA COLLECTION AND PREPROCESSING

The collected dataset consists of infrared images of maize plants captured at the Precision Agriculture Laboratory of the College of Engineering, Heilongjiang Bayi Agricultural University. The HD-SDI6006 infrared onboard camera was used for data acquisition. The dataset covers the entire growth period of maize crops and includes simulated scenarios such as leaf occlusion and missing seedlings. Additionally, images were captured in the maize inter-row spaces at Plot 2-10 of the Second Division of the Friendship

Farm in Heilongjiang Province. The maize was planted in a double-row pattern with large ridges, and the row width was 1.1 m. The camera was mounted on a tractor and positioned in the middle of the maize rows, approximately 1.5 m above the ground. By slowly moving forward, images were captured to obtain information about the inter-row spaces. The dataset was further processed by frame sampling, resulting in a total of 2000 images. The dataset was divided into a training set (1600 images), a validation set (200 images), and a test set (200 images). The images were manually annotated using the LabelImg tool, with the label set as ''maize''.

B. EXPERIMENTAL ENVIRONMENT AND EVALUATION METRICS

The experiments were conducted on a computer system consisting of an Asus machine equipped with an Intel(R) Core(TM) i7-10700H 2.50 GHz processor, 32 GB of RAM, and an Nvidia GeForce RTX 4070 GPU. The operating system used was Windows 11 (64-bit). The deep learning environment was set up with Python 3.9 and Torch 2.0.1. Data augmentation techniques, including random scaling, random cropping, and color enhancement, were employed during the training process to improve the model's generalization capability. The network model was trained and updated using the stochastic gradient descent (SGD) algorithm to optimize the network parameters. A cosine annealing learning rate decay strategy was implemented. The specific hyperparameters used in the experiments are presented in Table 1.

The performance of the improved model was evaluated using precision (P), recall (R), average precision (AP), number of parameters (Params), and frames per second (FPS). Precision measures the ratio of correctly predicted positive instances to all instances predicted as positive, while recall measures the ratio of correctly predicted positive instances to all actual positive instances. The number of parameters

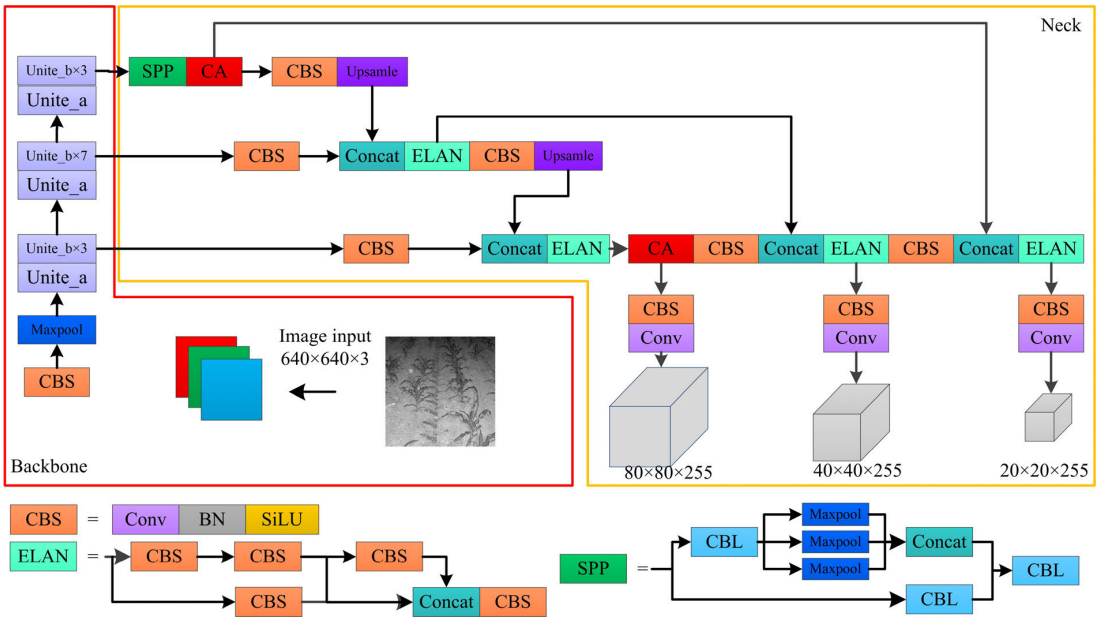

FIGURE 5. Improved the YOLOv7-tiny networks model structure.

TABLE 1. Network training hyperparameters.

Parameters	Value
Initial learning rate	0.01
Minimum learning rate	0.0001
Weight attenuation coefcient	0.0005
Momentum	0.937
Epoch	300
Image input size	640×640×3

TABLE 2. Comparison of different models test results.

Models	Average precision/%	Frames per second/(f·s^{-1})	Params/M
SSD	74.85	11.7	33.2
YOLOv4-tiny	89.05	23.4	7.2
YOLOv5s	93.58	28.9	14.1
YOLOv7-tiny	93.72	27.8	6.2
Improved YOLOv7-tiny	94.21	32.4	6.4

indicates the spatial complexity of the model, and frames per second evaluates the recognition speed of the model. The calculation formulas for these metrics as shown in Equation (12)-(14):

$$P = \frac{T_P}{T_P + F_P} \times 100\% \tag{12}$$

$$R = \frac{T_P}{T_P + F_N} \times 100\% \tag{13}$$

$$AP = \int_0^1 P(R)dR \tag{14}$$

In the equation: T_P represents the number of correctly detected maize by the model, F_P represents the number of background mistakenly detected as maize by the model, F_N represents the number of maize not detected by the model, AP represents the integral area enclosed by the precision-recall curve for a single class detection target.

C. ANALYSIS OF EXPERIMENTAL RESULTS

To demonstrate the improved YOLOv7-tiny model's superior capability in detecting nocturnal maize crops using infrared imagery, we conducted comparative experiments against established models including SSD, YOLOv4-tiny, YOLOv5s, and the unmodified YOLOv7-tiny. Each model underwent training and evaluation on the same hardware setup, with a dataset of 2,000 nocturnal infrared maize crop images. The outcomes, as summarized in Table 2, highlight the performance variances across different metrics.

In our analysis, the SSD model recorded an average precision of 74.85 %, the lowest among the contenders, primarily due to its large parameter count of 33.2 M, which reduces its effectiveness under low-light conditions. Conversely, the improved YOLOv7-tiny model, incorporating an attention mechanism for enhanced feature extraction, achieved the highest average precision of 94.21 %. This represents a significant advancement over the SSD model, with a frame rate of 32.4 fps and a slight parameter increase to 6.4 M, illustrating its efficiency and effectiveness. The YOLOv4-tiny model, while faster at 23.4 fps, offered a lower average precision of 89.05 % due to its reduced capability in handling the intricacies of nocturnal imagery, despite its smaller parameter size of 7.2 M. The improved YOLOv7-tiny model not only surpassed this with a higher precision but also maintained a competitive processing speed, underscoring its superior feature extraction and recognition capabilities.

Although the YOLOv5s model achieved a commendable average precision of 93.58 %, its larger parameter size of 14.1 M hampers its applicability in real-time edge device scenarios. The enhanced YOLOv7-tiny model, on the other hand,

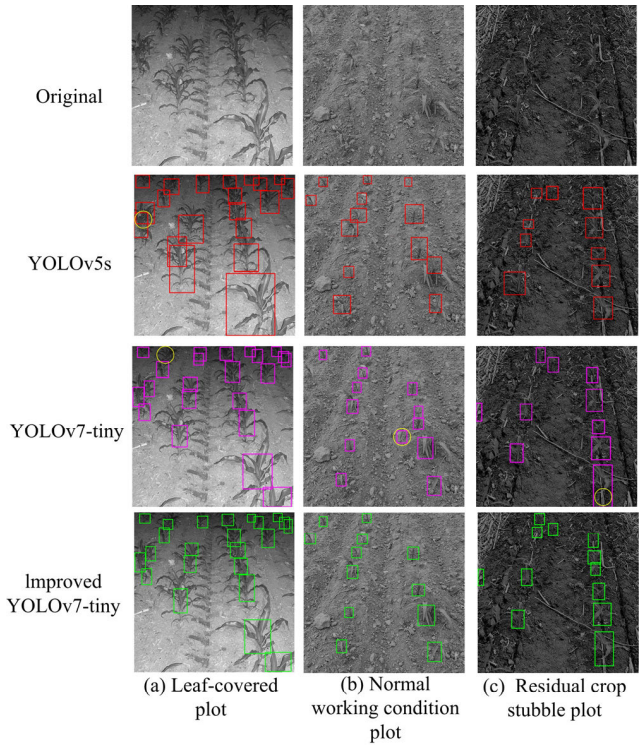

FIGURE 6. Comparison of maize detection model performance. Note: The boxed lines in the figure represent the rectangular bounding boxes detected by the model, and the yellow circles indicate areas of false positives and false negatives.

utilizes ShuffleNet to compress its parameters to 6.4 M while still achieving a high recognition accuracy of 94.21 % and the highest frame rate among the models tested. This makes it exceptionally suited for edge computing applications, where both speed and accuracy are critical.Compared to the original YOLOv7 network model, it improves accuracy by 0.49 % and frame rate by 4.6 fps.

D. COMPARISON OF DETECTION RESULTS

In order to validate the efficacy and universality of the proposed algorithmic enhancements for maize crop row fitting, we compared the detection outcomes of the original YOLOv7-tiny network model with those of the improved YOLOv7-tiny network model. Specifically, we selected three types of scenes from nocturnal infrared imagery of maize crop: images under normal operating conditions, images with maize seedling leaves occluding the view, and images with straw stubble coverage, as depicted in Figure 6.

We analyzed the model's recognition accuracy across these three scenarios. The results indicated that under leaf occlusion conditions, both the original YOLOv7-tiny network model and YOLOv5s network model failed to detect distant crops, and the target bounding boxes for maize crops were inaccurately marked due to insufficient clarity of nighttime crops. Conversely, the improved YOLOv7-tiny network model circumvented these issues, successfully detecting maize crops with less prominent features at a distance. Under normal operating conditions, the original YOLOv7-tiny network model misidentified adjacent leaves as crops, whereas the improved YOLOv7-tiny network model achieved successful detection with greater precision. In areas covered with straw stubble, the dense maize straw stubble led the original YOLOv7-tiny network model to misclassify the stubble and maize crops as the same category, thereby affecting the positioning accuracy.

In contrast, the improved YOLOv7-tiny network model was capable of identifying a greater number of maize crops. The experimental findings demonstrate that the refined model can detect targets within a maize field environment more effectively and accurately. Additionally, the confidence scores output by the improved YOLOv7-tiny network model were generally higher, indicating that the refined network possesses a more robust detection capability and can better focus on the characteristic information of the targets.

V. NAVIGATION LINE DETECTION

A. ACQUISITION OF LOCALIZATION REFERENCE POINTS

The accurate extraction of reference points is crucial for obtaining maize crop inter-row navigation lines. The specific procedure is as follows: Firstly, the improved YOLOv7-tiny object detection network model is used to identify maize seedlings in the field, resulting in the rectangular bounding coordinates for each maize seedling. Then, the rectangular bounding coordinates of each maize seedlings are processed to calculate the centroid coordinates for each maize seedlings. The centroid coordinates (x,y) of a maize seedling can be calculated from the top-left corner coordinates (x_1,y_1) and bottom-right corner coordinates (x_2,y_2) of the rectangular bounding box, as shown in Equation (15):

$$\begin{cases} x = \dfrac{x_1 + x_2}{2} \\ y = \dfrac{y_1 + y_2}{2} \end{cases} \tag{15}$$

Subsequently, we undertake a linear regression analysis to fit a line to the centroid of the maize seedlings. By leveraging techniques such as linear regression, we can effectively model the central line equation that encapsulates the centroid data points of the maize seedlings. Consequently, this equation serves as a reliable reference for extracting key positioning landmarks. These landmarks, such as the starting or ending point of the central line, play a pivotal role in accurately determining the spatial orientation of the maize crop rows. The outcome of this process is illustrated in Fig. 7, where the extracted positioning landmarks are denoted by yellow markers, while the manually annotated centroid points of the maize seedlings are represented by red markers.

B. ANALYSIS OF LOCALIZATION REFERENCE POINT ERRORS

In order to assess the efficacy of the chosen reference points for localization, an error analysis was performed on a randomly selected subset of 100 maize seedling images from the test dataset. Subsequently, a heat map was generated

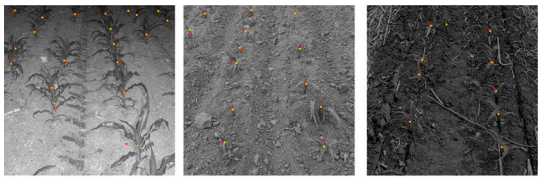

FIGURE 7. Results of localization reference point extraction for manually labeled maize seedling midpoints. Note: The yellow marked points in the figure represent the extracted localization reference points, and the red marked points represent the manually labeled midpoint of the maize seedlings.

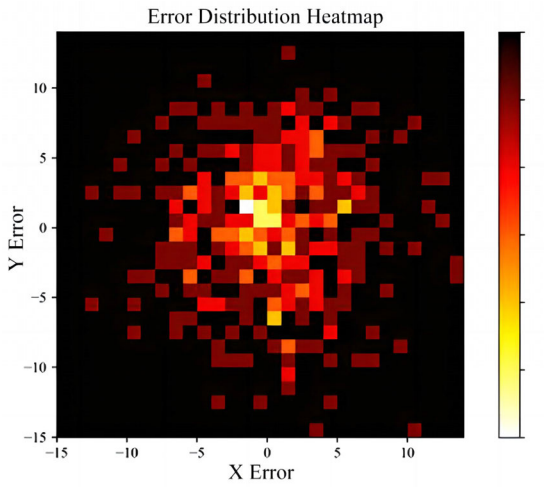

FIGURE 8. Heatmap scatter plot of errors.

to visualize the discrepancies between manually annotated and algorithmically recognized reference points, as depicted in Fig. 8. The analysis revealed that the majority of both horizontal and vertical errors were concentrated within a 10-pixel range. Specifically, approximately 97.5 % of the horizontal errors fell within this threshold, while approximately 98.5 % of the vertical errors exhibited the same characteristic. Furthermore, the average linear error was quantified as 3.43 pixels.

In order to calibrate the camera, the intrinsic matrix of the camera was obtained using Matlab software. This intrinsic matrix allowed us to convert pixel coordinates into camera coordinates, facilitating precise measurements of object positions within the captured images. The resulting average error between the positioning reference points and the manually determined center of the maize crops was calculated to be 4.85 cm. This finding substantiates the effectiveness of utilizing the bottom midpoint of the rectangular frame as a reliable reference point for navigation line positioning.

C. NAVIGATION LINE FITTING

In this study, we utilize the least squares method to fit the rows of maize crops for several compelling reasons:

1. The least squares method is a widely employed regression analysis technique that adeptly characterizes the relationship between a set of data points and a linear model. It operates by directly minimizing the sum of the squares of the residuals-the differences between observed and predicted values. This approach effectively prevents the cancellation of positive and negative errors.

FIGURE 9. Fitting results of maize seedlings on both sides.

2. Considering that the crop rows in the vast agricultural production regions of Heilongjiang reclamation area tend to exhibit an approximately linear distribution, the least squares method is particularly well-suited for this purpose. Compared to alternative approaches such as the Hough transform, the least squares method requires fewer points to fit in our scenario, thereby offering a lower computational complexity. This is advantageous for real-time processing applications.

3. The least squares method yields the equation of the fitted line, which facilitates subsequent extraction of navigational reference points along the line, such as starting and ending points.

Moreover, we integrate the RANSAC algorithm to mitigate the influence of outliers on the fitting results, thereby further enhancing the precision of the fit. The fundamental steps of the RANSAC algorithm are as follows:

1. Randomly select a subset, denoted as n points, from the localization reference points of the maize row on the left side as the candidate inlier set.

2. Employ the least squares method to fit these n points and derive a line (or curve) model. The least squares method operates on the principle that, given a series of (x_i, y_i) points $(i = 1, 2, 3, \ldots, N)$, assuming a linear relationship between x and y, fitting can be performed using the equation $y = Kx + B$. The resulting green lines on both sides, as depicted in Fig. 9.

The fitted maize crop rows, as stated in Equation (16).

$$\begin{cases} y_{left} = K_{left} \times x + B_{left} \\ y_{right} = K_{right} \times x + B_{right} \end{cases} \tag{16}$$

3. Calculate the distance between all other points and the model, and set a threshold to classify points as inliers or outliers. The optimization function aims to minimize the sum of squared differences between the observed values and the fitted values, preventing positive and negative errors from canceling each other out, as stated in Equation (17).

$$f = \sum_{i=0}^{N} (y_i - Kx_i - B)^2 \tag{17}$$

The optimal parameters K and B are obtained when the function f is minimized, indicating the best fit.

4. If the number of inliers exceeds a predefined threshold and the fitting error of the model is smaller than the previous

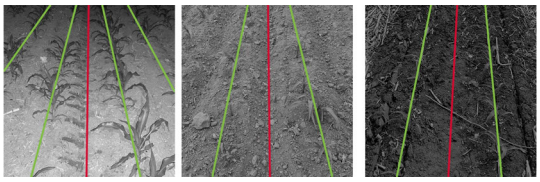

FIGURE 10. Inter-row navigation line fitting results.

best model, update the best model and re-estimate the inlier set.

Repeat steps 1 to 4 until the predefined iteration count is reached. By employing the RANSAC algorithm, a more accurate fitting line for the maize crop rows can be obtained, effectively removing the influence of outliers.

Assuming that 10 coordinates (x_i, y_i) are generated as positioning reference points on one side of the maize crop, i = 1, 2, 3, …, 10, the coordinates (x_l, y_c) and (x_r, y_c) are selected on the left and right sides of the crop row, respectively, with $c = 1, 2, 3, \ldots, 10$. By calculating the average x-coordinate of these 10 points on each side, 10 positioning points (x_a, y_c) are obtained for the navigation fitting line located in the middle of the two maize crop rows, where $a = 1, 2, 3, \ldots, 10$. The least squares method is then applied to these positioning points to calculate the fitted inter-row navigation line, depicted as the red line segment in Fig. 10.

However, the least squares method is not without its limitations: it presumes that the data points adhere to a normal distribution, and deviations from this assumption may compromise the efficacy of the fit. The method tends to underperform when applied to crop rows with significant curvature. Furthermore, a paucity of data points can also diminish the accuracy of the fit. Overall, given the characteristics of our scenario, the least squares method offers a commendable balance between precision and efficiency, rendering it a favorable choice. Nonetheless, it has its constraints, and alternative fitting algorithms that are more apt for future applications may be considered.

The number of frames for three video segments, randomly selected and captured exclusively during nighttime, was recorded. Additionally, the program's start and end times were documented using a timer. The processing duration of the program was determined by subtracting the start time from the end time and then dividing by the number of video frames to obtain the average processing time per frame. The average processing times for these video segments were found to be 0.057, 0.050, and 0.054 seconds per frame, respectively. Consequently, the overall average processing time was calculated to be 0.054 seconds per frame, which satisfies the real-time processing requirements essential for navigation in intelligent weed control machinery.

D. MOBILE PLATFORM DEPLOYMENT

The improved YOLOv7-tiny network model developed in this study has been deployed on a 10-inch industrial tablet computer manufactured by Apache in Chengdu, China. The computer is equipped with an Intel(R) Celeron(R) CPU J1800 processor, 2GB of installed memory, and runs on a 32-bit Windows 7 operating system, as shown in Figure 11. It features a CAN communication interface and supports plug-and-play functionality, greatly simplifying the complexity of traditional RS-232 interface wiring. This computer demonstrates stable operational performance in harsh environments such as agricultural fields, exhibiting high reliability and good stability.

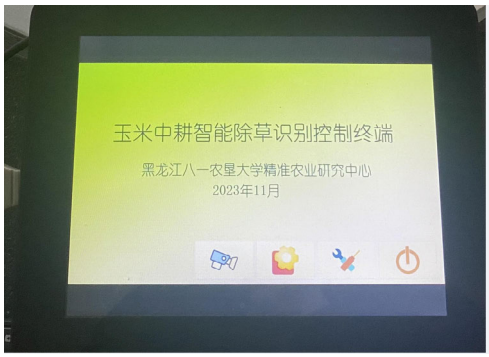

FIGURE 11. Vehicle smart terminal.

TABLE 3. Performance comparison between different terminal devices.

Parameters	Computer	Mobile terminal devices
Average precision /%	94.21	89.5
Frames per second/(f·s^{-1})	32.4	15.2
Hardware Cost /¥	>9000	3000

Following the deployment on the mobile platform, a series of cyclic inference tests were conducted, achieving a detection speed of approximately 15 fps, which meets the requirements of practical engineering applications. To further evaluate the model's real-world performance, inference tests were carried out on a subset of nocturnal maize images from the test set. The results indicated that for close-up image samples, the model maintained effective inference capabilities; for distant image samples with clear and distinct maize crop information, the model also demonstrated high recognition accuracy.

Not only was the algorithm effectively ported and deployed on the mobile platform, but it also retained high precision. Table 3 compares the performance differences in terms of inference speed, model accuracy, and hardware cost between the enhanced YOLOv7-tiny algorithm running on a traditional computer and on a mobile platform. Although there was a slight decrease in model accuracy and a reduction in inference speed due to the limitations of the mobile processor, the inference speed of 15 fps still largely meets the engineering application requirements. Furthermore, the mobile processor's low cost, high integration, and compact structure offer significant economic advantages for engineering applications.

VI. DISCUSSION

The improved YOLOv7-tiny network model proposed in this paper demonstrates significant advantages in the task

of nighttime maize seed detection and inter-row navigation line positioning. Unlike traditional methods reliant on RGB imagery [6], this study employs infrared imaging to bolster the model's target detection capabilities in nocturnal settings. By integrating the ShuffleNet v1 network, our approach not only achieves parameter compression but also enhances feature extraction capabilities compared to the original YOLOv7-tiny network model [22]. The introduced Coordinate Attention mechanism, which considers both channel and spatial dimensions, surpasses conventional channel attention methods in enhancing model recognition precision. Moreover, the adoption of the EIOU loss function [24] over the original loss function accelerates model convergence and improves positioning accuracy. Compared to methods dependent on centerline extraction models [18], our proposed reference point positioning technique identifies the location of inter-row navigation lines more accurately. Experimental results indicate that, among other YOLOv7 network model variants [26], our model exhibits the best balance between recognition accuracy and real-time processing capabilities. Sa et al. [16] explored the fusion of RGB and NIR multimodal images for fruit detection, employing pixel-level fusion techniques. However, the acquisition of multimodal images incurs high costs. Yang et al. [21] introduced a combination of YOLOv5 with the ExG method and Otsu's method for navigation line recognition, successfully segmenting crop rows and background within regions of interest. Yet, their method's performance in nocturnal environments requires enhancement. Peng et al. [19] proposed an improved YOLOv7-based method for navigation line detection within orchards, achieving certain successes in orchard navigation demands. However, this method has not yet fully addressed nighttime maize crop row navigation line recognition. In contrast to the aforementioned studies, our research achieves high-precision detection and extraction of maize crop rows for navigation by comprehensively optimizing the backbone network, attention mechanisms, and loss functions. This approach not only circumvents the high costs associated with multimodal image acquisition but also enhances the robustness and accuracy of crop row detection in various environmental conditions, including at night.

However, potential false positives and false negatives in the target detection model may still lead to inaccuracies in reference point positioning. Future work could consider employing more advanced target detection networks to minimize this source of error [33]. Additionally, human factors in the manual annotation process could introduce errors; future research may reduce this by increasing the number of annotated samples and implementing multiple annotators. Also, the conversion of pixel coordinates to real-world coordinates relies on the camera's intrinsic matrix, where estimation errors could affect the final positioning outcome. As suggested in the literature [37], repeated optimization of the intrinsic matrix through calibration methods can enhance the accuracy of coordinate conversion.

In summary, despite the presence of potential error sources, the improved model presented in this study exhibits outstanding performance in the task of nighttime maize seed detection and navigation line positioning. Future work needs to address these error sources for optimization and further validate and refine the model in actual production environments.

VII. CONCLUSION

This study has successfully developed a maize inter-row navigation line extraction technique based on an improved YOLOv7-tiny network model. By integrating the ShuffleNetv1 backbone network, the CA module, and the EIOU loss function, the model has significantly enhanced detection efficiency and accuracy under nocturnal conditions. Experimental validation reveals that the model achieves a high detection accuracy of 94.21 % on the test dataset and processes at a speed of 32.4 fps, which represents an increase of 0.49 percentage points in accuracy and a speed improvement of 4.6 frames compared to the original YOLOv7-tiny network model. Furthermore, this research adopts the lower midpoint of the detection bounding box as the reference for navigation positioning, achieving a positioning accuracy within an error margin of 4.85 centimeters, thus fulfilling the practical requirements of agricultural machinery navigation.

This technology has also been implemented on mobile devices, providing technical support for automated weeding operations at night. Future research will be dedicated to further enhancing the model's robustness under varying lighting conditions and complex backgrounds, expanding its capability to recognize multiple crop types and discriminate against extraneous matter. This will be achieved by constructing larger datasets to train more powerful deep learning models, thereby improving detection accuracy. In conclusion, this study not only offers an effective technology for intelligent agricultural navigation during nighttime but also lays the groundwork for future applications and research directions in smart mechanized weeding.

REFERENCES

[1] N. E. Korres, N. R. Burgos, I. Travlos, M. Vurro, T. K. Gitsopoulos, V. K. Varanasi, S. O. Duke, P. Kudsk, C. Brabham, C. E. Rouse, and R. Salas-Perez, "Chapter six—New directions for integrated weed management: Modern technologies, tools and knowledge discovery," *Academic Press*, vol. 155, pp. 243–319, 2019, doi: 10.1016/bs.agron.2019.01.006.

[2] A. Venkataraju, D. Arumugam, C. Stepan, R. Kiran, and T. Peters, "A review of machine learning techniques for identifying weeds in corn," *Smart Agricult. Technol.*, vol. 3, Feb. 2023, Art. no. 100102, doi: 10.1016/j.atech.2022.100102.

[3] T. Wang, B. Chen, Z. Zhang, H. Li, and M. Zhang, "Applications of machine vision in agricultural robot navigation: A review," *Comput. Electron. Agricult.*, vol. 198, Jul. 2022, Art. no. 107085, doi: 10.1016/j.compag.2022.107085.

[4] J. H. Westwood, R. Charudattan, S. O. Duke, S. A. Fennimore, P. Marrone, D. C. Slaughter, C. Swanton, and R. Zollinger, "Weed management in 2050: Perspectives on the future of weed science," *Weed Sci.*, vol. 66, no. 3, pp. 275–285, Feb. 2018, doi: 10.1017/wsc.2017.78.

[5] A. Monteiro and S. Santos, "Sustainable approach to weed management: The role of precision weed management," *Agronomy*, vol. 12, no. 1, p. 118, Jan. 2022, doi: 10.3390/agronomy12010118.

[6] H. Zhang, H. Xu, X. Tian, J. Jiang, and J. Ma, "Image fusion meets deep learning: A survey and perspective," *Inf. Fusion*, vol. 76, pp. 323–336, Dec. 2021, doi: 10.1016/j.inffus.2021.06.008.

[7] J. M. Guerrero, J. J. Ruz, and G. Pajares, "Crop rows and weeds detection in maize fields applying a computer vision system based on geometry," *Comput. Electron. Agricult.*, vol. 142, pp. 461–472, Nov. 2017, doi: 10.1016/j.compag.2017.09.028.

[8] A. Peruzzi, L. Martelloni, C. Frasconi, M. Fontanelli, M. Pirchio, and M. Raffaelli, "Machines for non-chemical intra-row weed control in narrow and wide-row crops: A review," *J. Agricult. Eng.*, vol. 48, no. 2, pp. 57–70, Jun. 2017, doi: 10.4081/jae.2017.583.

[9] K. Zheng, X. Zhao, C. Han, Y. He, C. Zhai, and C. Zhao, "Design and experiment of an automatic row-oriented spraying system based on machine vision for early-stage maize corps," *Agriculture*, vol. 13, no. 3, p. 691, Mar. 2023, doi: 10.3390/agriculture13030691.

[10] L. Yang, S. Liu, and Y. Zhao, "Deep-learning based algorithm for detecting targets in infrared images," *Appl. Sci.*, vol. 12, no. 7, p. 3322, Mar. 2022, doi: 10.3390/app12073322.

[11] X. Liang, L. Liu, M. Luo, Z. Yan, and Y. Xin, "Robust infrared small target detection using Hough line suppression and rank-hierarchy in complex backgrounds," *Infr. Phys. Technol.*, vol. 120, Jan. 2022, Art. no. 103893, doi: 10.1016/j.infrared.2021.103893.

[12] Y. Bai, B. Zhang, N. Xu, J. Zhou, J. Shi, and Z. Diao, "Vision-based navigation and guidance for agricultural autonomous vehicles and robots: A review," *Comput. Electron. Agricult.*, vol. 205, Feb. 2023, Art. no. 107584, doi: 10.1016/j.compag.2022.107584.

[13] T. Xue, Z. Zhang, W. Ma, Y. Li, A. Yang, and T. Ji, "Nighttime pedestrian and vehicle detection based on a fast saliency and multifeature fusion algorithm for infrared images," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 9, pp. 16741–16751, Sep. 2022, doi: 10.1109/TITS.2022.3193086.

[14] Q. Kang, H. Zhao, D. Yang, H. S. Ahmed, and J. Ma, "Lightweight convolutional neural network for vehicle recognition in thermal infrared images," *Infr. Phys. Technol.*, vol. 104, Jan. 2020, Art. no. 103120, doi: 10.1016/j.infrared.2019.103120.

[15] W. Li, "Infrared image pedestrian detection via YOLO-V3," in *Proc. IEEE 5th Adv. Inf. Technol., Electron. Autom. Control Conf. (IAEAC)*, vol. 5, Mar. 2021, pp. 1052–1055, doi: 10.1109/IAEAC50856.2021.9390896.

[16] I. Sa, Z. Ge, F. Dayoub, B. Upcroft, T. Perez, and C. McCool, "DeepFruits: A fruit detection system using deep neural networks," *Sensors*, vol. 16, no. 8, p. 1222, Aug. 2016, doi: 10.3390/s16081222.

[17] W. Tian, Z. Deng, D. Yin, Z. Zheng, Y. Huang, and X. Bi, "3D pedestrian detection in farmland by monocular RGB image and far-infrared sensing," *Remote Sens.*, vol. 13, no. 15, p. 2896, Jul. 2021, doi: 10.3390/rs13152896.

[18] Q. Zhang, J. Wang, and B. Li, "Extraction method for centerlines of rice seedings based on YOLOv3 target detection," *Trans. Chin. Soc. Agric. Mach.*, vol. 8, no. 51, pp. 34–43, Jun. 2020, doi: 10.6041/j.issn.1000-1298.2020.08.004.

[19] S. Peng, "Detection of the navigation line between lines in orchard using improved YOLOv7," *Trans. Chin. Soc. Agric. Mach.*, vol. 39, no. 16, pp. 131–138, Sep. 2023.

[20] X. Liu, J. Qi, W. Zhang, Z. Bao, K. Wang, and N. Li, "Recognition method of maize crop rows at the seedling stage based on MS-ERFNet model," *Comput. Electron. Agricult.*, vol. 211, Aug. 2023, Art. no. 107964, doi: 10.1016/j.compag.2023.107964.

[21] Y. Yang, Y. Zhou, X. Yue, G. Zhang, X. Wen, B. Ma, L. Xu, and L. Chen, "Real-time detection of crop rows in maize fields based on autonomous extraction of ROI," *Exp. Syst. Appl.*, vol. 213, Mar. 2023, Art. no. 118826, doi: 10.1016/j.eswa.2022.118826.

[22] C.-Y. Wang, A. Bochkovskiy, and H.-Y.-M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 7464–7475.

[23] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 779–788.

[24] W. Yi and B. Wang, "Research on underwater small target detection algorithm based on improved YOLOv7," *IEEE Access*, vol. 11, pp. 66818–66827, 2023, doi: 10.1109/ACCESS.2023.3290903.

[25] S. Hu, F. Zhao, H. Lu, Y. Deng, J. Du, and X. Shen, "Improving YOLOv7-tiny for infrared and visible light image object detection on drones," *Remote Sens.*, vol. 15, no. 13, p. 3214, Jun. 2023, doi: 10.3390/rs15133214.

[26] L. Ma, L. Zhao, Z. Wang, J. Zhang, and G. Chen, "Detection and counting of small target apples under complicated environments by using improved YOLOv7-tiny," *Agronomy*, vol. 13, no. 5, p. 1419, May 2023, doi: 10.3390/agronomy13051419.

[27] Z. Liu, C. Dai, and X. Li, "Pedestrian detection method in infrared image based on improved YOLOv7," in *Proc. IEEE 3rd Int. Conf. Inf. Technol., Big Data Artif. Intell. (ICIBA)*, Chongqing, China, May 2023, pp. 946–954.

[28] Z. Yang, H. Feng, Y. Ruan, and X. Weng, "Tea tree pest detection algorithm based on improved YOLOv7-tiny," *Agriculture*, vol. 13, no. 5, p. 1031, May 2023, doi: 10.3390/agriculture13051031.

[29] X. Yang, L. Pan, D. Wang, Y. Zeng, W. Zhu, D. Jiao, Z. Sun, C. Sun, and C. Zhou, "FARnet: Farming action recognition from videos based on coordinate attention and YOLOv7-tiny network in aquaculture," *J. ASABE*, vol. 66, no. 4, pp. 909–920, 2023, doi: 10.13031/ja.15362.

[30] Z. Zhang, J. Huang, G. Hei, and W. Wang, "YOLO-IR-free: An improved algorithm for real-time detection of vehicles in infrared images," *Sensors*, vol. 23, no. 21, p. 8723, Oct. 2023, doi: 10.3390/s23218723.

[31] Y. Hua, H. Xu, J. Liu, L. Quan, X. Wu, and Q. Chen, "A peanut and weed detection model used in fields based on BEM-YOLOv7-tiny," *Math. Biosciences Eng.*, vol. 20, no. 11, pp. 19341–19359, Oct. 2023, doi: 10.3934/mbe.2023855.

[32] L. Yang, W. Du, and Y. Zhao, "A lightweight temporal attention-based convolution neural network for driver's activity recognition in edge," *Comput. Electr. Eng.*, vol. 110, Sep. 2023, Art. no. 108861, doi: 10.1016/j.compeleceng.2023.108861.

[33] W. Fang, L. Wang, and P. Ren, "Tinier-YOLO: A real-time object detection method for constrained environments," *IEEE Access*, vol. 8, pp. 1935–1944, 2020, doi: 10.1109/ACCESS.2019.2961959.

[34] H. Liu, Y. Fan, H. He, and K. Hui, "Improved YOLOv7-tiny's object detection lightweight model," *Comput. Eng. Appl.*, vol. 14, no. 59, pp. 166–174, 2023, doi: 10.3778/j.issn.1002-8331.2302-0115.

[35] G. Chen, R. Cheng, X. Lin, W. Jiao, D. Bai, and H. Lin, "LMDFS: A lightweight model for detecting forest fire smoke in UAV images based on YOLOv7," *Remote Sens.*, vol. 15, no. 15, p. 3790, Jul. 2023, doi: 10.3390/rs15153790.

[36] S. Fang, Y. Wang, G. Zhou, A. Chen, W. Cai, Q. Wang, Y. Hu, and L. Li, "Multi-channel feature fusion networks with hard coordinate attention mechanism for maize disease identification under complex backgrounds," *Comput. Electron. Agricult.*, vol. 203, Dec. 2022, Art. no. 107486, doi: 10.1016/j.compag.2022.107486.

[37] P. Liu and H. Yin, "YOLOv7-peach: An algorithm for immature small yellow peaches detection in complex natural environments," *Sensors*, vol. 23, no. 11, p. 5096, May 2023, doi: 10.3390/s23115096.

[38] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2999–3007.

HAILIANG GONG received the master's degree in agricultural engineering from the College of Engineering, Heilongjiang Bayi Agricultural University, Daqing, China, in 2021. He is currently a Lecturer with Heilongjiang Bayi Agricultural University. His research interest includes agricultural smart equipment.

WEIDONG ZHUANG received the Ph.D. degree in agricultural mechanization engineering from the College of Engineering, Heilongjiang Bayi Agricultural University, Daqing, China, in 2011. He is currently a Professor with Heilongjiang Bayi Agricultural University. His research interests include precision agriculture and agricultural information technology.