# Tractor Road Detection Based on Improved YOLOv5m for Grain Truck

Qian Zhang[1*], Wenjie Xu[1], Qingshan Chen[1], Zhenghui Zhao[2], Chunsong Du[2], Xiangqian Xu[3]

[1]School of Agricultural Engineering, Jiangsu University, Zhenjiang, China
[2]School of Electrical & Information Engineering, Jiangsu University, Zhenjiang, China
[3]Shandong Golddafeng Machinery Co., Ltd., Jining, China

*Corresponding author: zhangq_jsu@ujs.edu.cn

**Abstract.** Rapid and accurate detection of tractor roads plays a vital role in the intelligent and autonomous driving of grain trucks. The tractor roads are unstructured, have a complex background and are covered by multiple obstacles. These factors result in low accuracy of tractor road detection based on the existing YOLOv5m. This paper improves YOLOv5m based on the Squeeze-and-Excitation (SE) attention mechanism. After the C3 module in the backbone, the SE attention module is added to optimize the weighting process of local features. This improves the ability of the model to focus on details such as blurred areas at the edge of the tractor roads and obstacles. At the same time, attention optimization of non-full-graph processes also reduces computational redundancy. Experimental verification was carried out on the intelligent grain truck developed by our research group. Experimental results show that compared with the existing YOLOv5m, Deeplabv3+, FCN, and UNet, the $mAP_{50}$ of improved YOLOv5m for tractor road detection is improved by at least 2.7%, and the detection time is 52.1ms. Based on the method proposed in this paper, the onboard vision of intelligent grain truck can quickly and accurately predict the tractor roads, and it will help to further realize the intelligent unmanned operation of the grain truck.

**Keywords:** tractor road detection; grain truck; neural network; vehicle vision

## 1 Introduction

In recent years, the number of grain trucks and the mechanization level of grain transportation in China have increased. During harvest, grain trucks must meet high-intensity, high-efficiency operation requirements, and the complexity of tractor roads directly affects their passability and efficiency. Accurate detection of tractor roads is key to improving transportation efficiency and safety. Real-time perception and detection of tractor roads based on onboard vision provide accurate data. It enables identification of passable areas and avoids unfavorable sections. It improves efficiency and safety [1].

Currently, platforms for tractor road detection mainly include drones, remote sensing satellites, and grain trucks. Drones and satellites are suitable for large-scale detection


X. Li et al. (eds.), *Proceedings of the 2025 International Conference on Agriculture and Resource Economy (ICARE 2025)*, Advances in Biological Sciences Research 46,
https://doi.org/10.2991/978-94-6463-746-5_12

but have low local resolution, are affected by weather, and have poor real-time performance [2]. These limitations make them unsuitable for this paper. Therefore, this paper selects a grain truck onboard platform. The detection methods used are spectral, radar, and visual technologies. Onboard spectroscopy is all-weather capable but lacks real-time performance and is complex. Radar detection is accurate but sensitive to environmental interference, has large data volumes, and is costly. In contrast, Onboard visible light cameras offer high resolution, low cost, and short sampling periods [3]. These advantages make them more suitable for accurate and rapid detection of tractor roads. This is especially true under oblique shooting conditions.

Currently, there are three main methods for vision-based unstructured road detection: (1) methods based on road color, texture, and edge features; (2) methods based on structural models of road edges; and (3) deep learning-based methods [4]. Methods based on road color, texture, and edge features include threshold segmentation, Canny edge detection, etc. These methods are not sensitive to road shape. However, their detection performance drops when road features change or are affected by external factors. The road edge-based structural model assumes that unstructured roads have regular edges. It uses these edges to construct a road model, which is then fitted to identify road boundaries. Common models include the line, parabolic, and spline curve models [5]. This method is robust under conditions like watermarks, shadows, and uneven lighting, but accuracy decreases when road shapes deviate from the assumptions of the model. Deep learning methods, especially convolutional neural networks, offer better robustness and generalization across different environments and road types. Common models for tractor road detection include FCN, UNet, DeepLabV3+, and YOLOv5m. FCN, UNet, and DeepLabV3+ use a two-stage structure to generate candidate regions for classification, while YOLOv5m, a one-stage model, directly predicts object locations and classes, offering higher speed and efficiency in detection [6]. However, rural tractor roads are highly complex, with unstructured surfaces, intricate backgrounds, and many obstacles. The YOLOv5m model shows low detection accuracy in such environments and requires further optimization.

To solve the above problems, this paper optimizes the attention mechanism of the Backbone structure within the existing YOLOv5m framework. This optimization aims to improve tractor road detection accuracy. The effectiveness of this method was validated through experiments on the intelligent grain truck developed by our research group, enabling accurate and rapid detection of tractor roads and advancing the realization of intelligent unmanned operation for the grain truck.

## 2 Construction of the Onboard Vision System and Dataset for the Grain truck

### 2.1 Construction of Onboard Vision System for the Grain truck

A visual data collection system for tractor roads with oblique shooting is installed on the intelligent grain truck to capture tractor road data. As shown in Figure 1, the system uses a ZED 2I binocular camera by Stereolabs, with a 110°(H) × 70°(V) × 120°(D)

field of view, a 2.1 mm focal length, and a 20-meter depth range [7]. The camera is mounted on a bracket beside the cabin at a 23° angle to the ground, enabling detailed image capture of the tractor roads. Image data is processed by an NVIDIA Jetson AGX Orin industrial computer with a high-performance ARM Cortex-A78AE processor and an NVIDIA Ampere GPU, offering strong parallel computing and deep learning acceleration. Efficient data exchange between the industrial machine and main control unit is ensured via a CAN bus, supporting real-time and stable system operation throughout the operation of the grain truck.


**Fig. 1.** Onboard Vision System of the Grain truck

As shown in Figure 2, the coordinate model for visual detection of tractor roads is constructed. $O_{c1}$ - $X_{c1}Y_{c1}Z_{c1}$ and $O_{c2}$ - $X_{c2}Y_{c2}Z_{c2}$ represent the camera coordinate systems of the left and right cameras respectively，where $O_{c1}$ - $X_{c1}Y_{c1}Z_{c1}$ is the base coordinate system of the camera. The angle between the $Z_{c1}$ -axis of the camera coordinate system and the horizontal plane is $\theta$. $O_{i1}$ - $X_{i1}Y_{i1}$ and $O_{i2}$ - $X_{i2}Y_{i2}$ represent the image coordinate systems respectively. $O_{o1}$ - $U_1V_1$ and $O_{o2}$ - $U_2V_2$ represent the pixel coordinate system respectively [8]. $O_w$ - $X_wY_wZ_w$ represents the world coordinate system, $X_w$ and $Y_w$ axes are on the horizontal plane, and $Z_w$ axis is vertically upward. The origin $O_w$ of the world coordinate system and the origin $O_{c1}$ of the camera base coordinate system are on the same axis , which is perpendicular to the horizontal plane, and the distance between the two points is $H$ .

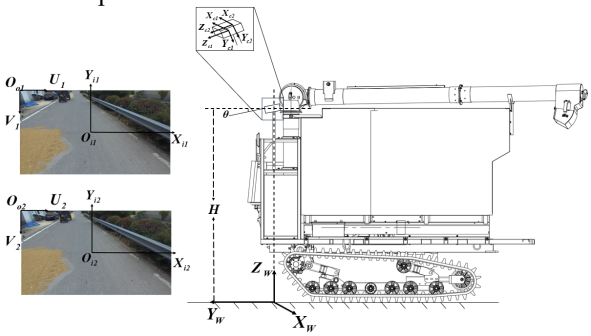


**Fig. 2.** Coordinate Model of the Onboard Vision System

In the pixel coordinate system $O_{o1}$ - $U_1V_1$, the coordinate value of the origin $O_{i1}$ of the image coordinate system $O_{i1}$ - $X_{i1}Y_{i1}$ is $(u_{o1}, v_{o1})$. In the pixel coordinate system $O_{o2}$ - $U_2V_2$, the coordinate value of the origin $O_{i2}$ of the image coordinate system $O_{i2}$ - $X_{i2}Y_{i2}$ is $(u_{o2}, v_{o2})$. The number of pixels per unit length of the left and right cameras is the same, and their length and width are $d_x$ and $d_y$ respectively. The pixel coordinates are $(u_1, v_1)$ and $(u_2, v_2)$, and the image coordinates are $(x_{i1}, y_{i1})$ and $(x_{i2}, y_{i2})$. The conversion relationship between the pixel coordinate system and the image coordinate system of the left and right cameras is shown in formula (1) :

$$\begin{bmatrix} u_n \\ v_n \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{d_x} & 0 & u_{on} \\ 0 & \frac{1}{d_y} & v_{on} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{in} \\ y_{in} \\ 1 \end{bmatrix} \tag{1}$$

The focal lengths of the left and right cameras are both $f$, and the conversion relationship between the image coordinate system and the camera coordinate system is shown in formula (2):

$$\begin{bmatrix} x_{in} \\ y_{in} \\ 1 \end{bmatrix} = \frac{1}{z_{cn}} \cdot \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} x_{cn} \\ y_{cn} \\ z_{cn} \\ 1 \end{bmatrix} \tag{2}$$

The depth value of a point is represented by the Euclidean distance from a point $(x_{c1}, y_{c1}, z_{c1})$ to the origin $O_{c1}$ in the camera base coordinate system $O_{c1}$ - $X_{c1}Y_{c1}Z_{c1}$.

The calibration plate image is collected by the binocular camera to obtain the transformation matrix $T$. The transformation relationship between the camera coordinate system and the world coordinate system can be obtained through formula (3).

$$\begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = T \cdot \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \tag{3}$$

## 2.2 Image Acquisition of Tractor roads and Dataset Construction

In June 2024, tractor road data were collected at Shiyezhou Satellite Village, Zhenjiang City, Jiangsu Province, from 09:00 to 17:00. Image data were captured using a ZED 2I camera (1920×1080 resolution) across multiple experimental areas. Some collected images are shown in Figure 3. The shooting angle was 23° with a height of 2.3 meters.

The data include various types of tractor roads, such as cement, dirt, and gravel roads, representing different road conditions in the operational environment.

**Fig. 3.** Initial dataset of tractor roads

The initial tractor road dataset contains 2,300 sample images. To fully train the network model and improve its robustness and generalization ability, the original dataset was augmented using data augmentation techniques. The specific methods include flipping the images vertically or horizontally with a probability of 50%; randomly adjusting the image brightness within a range from 0.35 to 1.5 times; randomly varying the contrast between 0.2 and 2 times; adding random noise; and applying rotation operations. After data augmentation, part of the VOC_Road dataset is shown in Figure 4.

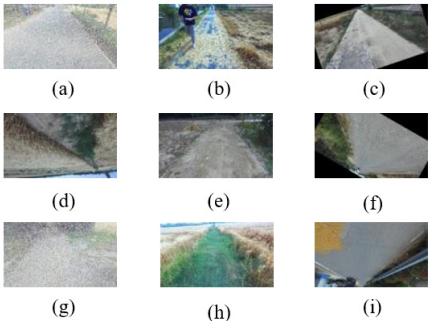


**Fig. 4.** Data augmentation for initial dataset. (a) Image flipping and noise addition. (b) Brightness and contrast adjustment. (c) Brightness adjustment and rotation. (d) Vertical and horizontal flipping. (e) Horizontal flipping and contrast adjustment. (f) Vertical flipping and rotation. (g) Noise addition and contrast adjustment. (h) Brightness and contrast adjustment. (i) Vertical flipping and brightness adjustment.

The augmented dataset was labeled using the Labelme tool, generating the labeled dataset, which was named VOC_Road. The dataset contains 11,927 images, with a total of 1 class label, 'road'. In total, 11,927 semantic labels have been annotated. Figure 5 shows the statistical information of the dataset. For subsequent model training and evaluation, the VOC_Road dataset is randomly divided into training, validation, and test sets in a 8:1:1 ratio. The training set contains 9,536 images, while the validation and test sets each contain 1,192 images.

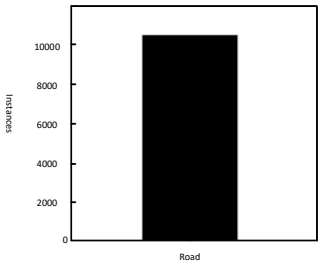


**Fig. 5.** VOC_Road dataset statistics

# 3 Improvement of YOLOv5m Based on Squeeze-and-Excitation Attention Optimization

## 3.1 Existing YOLOv5m

As shown in Figure 6, the YOLOv5m architecture consists of three core modules: Backbone, Neck, and Head. Each module contains three components: CBS (Convolution, Batch Normalization, and SiLU activation), C3 (three CBS modules with a bottleneck structure), and SPPF (Spatial Pyramid Pooling Fast) [9]. In the Backbone, image features are extracted using residual connections and feature reuse. The C3 module learns residual features with three convolutional layers and a bottleneck structure. The SPPF module improves feature capability through three 5×5 max pooling layers. The Neck module includes PAN (Path Aggregation Network) and FPN (Feature Pyramid Network). PANet combines spatial and semantic information at different levels, while FPN fuses multi-scale features through up- and down-sampling. The Head module integrates these multi-scale features to produce the final detection results [10].

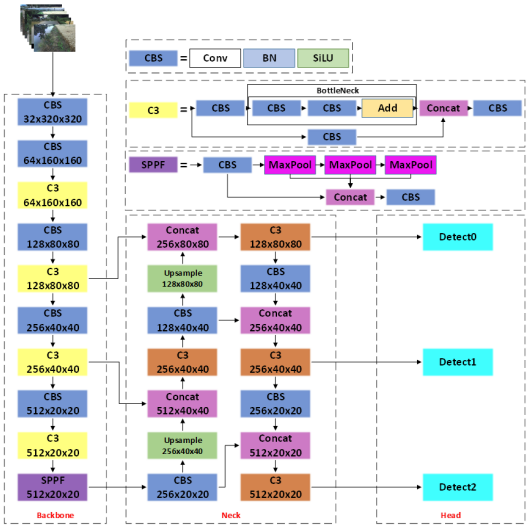


**Fig. 6.** Existing YOLOv5m

### 3.2 Improved YOLOv5m

The existing YOLOv5m Backbone structure lacks an attention optimization mechanism. The feature extraction part pays less attention to the fuzzy region at the edge of the tractor roads and complex background (such as surrounding obstacles, ground, etc.), resulting in low accuracy of tractor road detection. The existing global attention mechanism usually processes all pixels or regions in the image so that the model can focus on the entire image. However, its calculation is relatively redundant [11] and it is difficult to meet the real-time detection requirements of the tractor road in this paper. To solve the above problems, this paper improves the existing YOLOv5m model based on attention optimization to improve the accuracy of tractor road detection in complex backgrounds. This paper adopts Squeeze-and-Excitation (SE) in the Backbone structure of YOLOv5m. This improves the ability of different channels to capture key semantic information, such as fuzzy areas at the edge of the tractor roads and obstacles. At the same time, the attention optimization of the non-full-image process enhances the ability of the model to focus on fuzzy boundaries of the tractor roads. It also suppresses background interference. This approach avoids the computational redundancy of the full-image attention mechanism. It meets the needs of real-time detection [12]. The model structure of the improved YOLOv5m is shown in Figure 7.

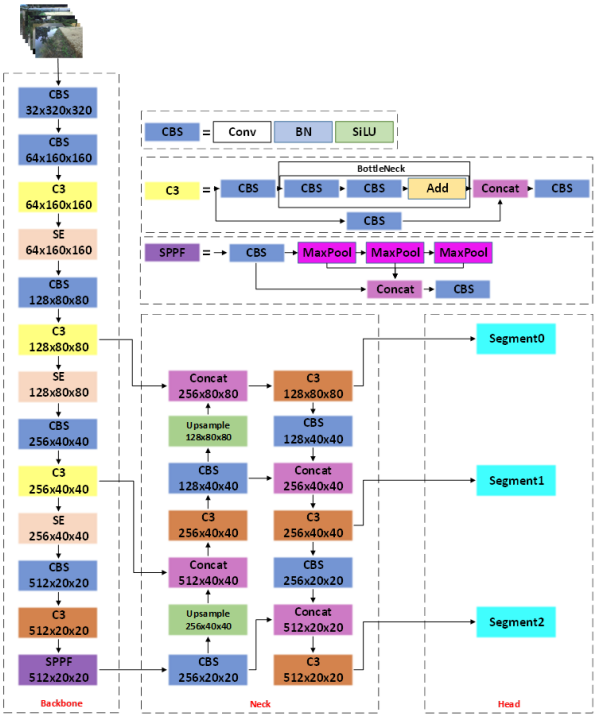


**Fig. 7.** Improved YOLOv5m

To simulate the global image dependency on local feature representations with lightweight computation and memory usage, the SE attention mechanism module (Squeeze-and-Excitation Block) is introduced, as shown in Figure 8. This module is an adaptive attention mechanism. It aims to improve the ability of the model to focus on important features. It does so by weighting and calibrating the channels of feature maps. Its core idea is squeeze and excitation. This module dynamically adjusts the weights of the channel dimensions of the input features, explicitly modeling the dependency on global information, further enhancing key features while suppressing redundant information.

Step 1 Transformation $\mathbf{F}_{tr}$ :Given an input feature map $\mathbf{X}$, it generates a feature map U after operation $\mathbf{F}_{tr}$. The formula is as follows:

$$U_c = V_c * X = \sum_{s=1}^{C'} V_c^s * X^s \tag{4}$$

$X \in \mathbb{R}^{H' \times W' \times C'}$ represents the input feature map, $U \in \mathbb{R}^{H \times W \times C}$ represents the output feature map, $V$ represents the learned set of convolution filters, $V_c$ represents the convolution kernel parameters for generating the $c$-th output channel, $V_c^s$ represents the corresponding 2D spatial convolution kernel, $*$ represents the convolution operation.

Step 2 Squeeze $\mathbf{F}_{sq}(\cdot)$ : The feature map is globally average pooled, and the feature map $W \times H \times C$ containing global information is directly compressed into a feature vector $1 \times 1 \times C$. The channel features of $C$ feature maps are compressed into a single value, so that the generated channel-level statistics $z$ contain contextual information [13]. The formula is defined as follows, where $z_c$ is the $c$-th element of $z$ :

$$z_c = \mathbf{F}_{sq}(\mathbf{u}_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i, j). \tag{5}$$

Step 3 Excitation $\mathbf{F}_{ex}(\cdot, \mathbf{W})$ : Through two fully connected layers $W_1$, $W_2$ processes the vector $z$ obtained in the previous step. The channel weight value $\mathbf{s}$ is obtained. After passing through two fully connected layers, different values in $\mathbf{s}$ represent the weight information of different channels, giving different weights to the channels. The formula is defined as follows:

$$\mathbf{s} = \mathbf{F}_{ex}(\mathbf{z}, \mathbf{W}) = \sigma(g(\mathbf{z}, \mathbf{W})) = \sigma(\mathbf{W}_2 \delta(\mathbf{W}_1 \mathbf{z})), \tag{6}$$

Step 4 Scale $\mathbf{F}_{scale}(\cdot, \cdot)$ : The weight vector $\mathbf{s}$ generated in the third step is used to assign weights to the feature map U to obtain the desired feature map $\tilde{\mathbf{x}}$, which has the same size as the feature map U . The formula is defined as follows:

$$\tilde{\mathbf{x}}_c = \mathbf{F}_{scale}(\mathbf{u}_c, s_c) = s_c \mathbf{u}_c, \tag{7}$$

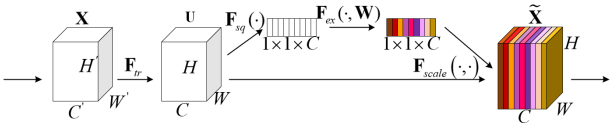


**Fig. 8.** SE structure

## 4 Tractor Road Detection Based on Improved YOLOv5m

To accelerate training, improve model generalization, and reduce overfitting, this paper adopts a transfer learning method. The improved YOLOv5m is pre-trained on the COCO dataset to obtain pre-trained weights. The COCO dataset includes various common objects such as people, animals, and vehicles, with annotations [14]. After pre-training, the weights are fine-tuned on the VOC_Road dataset to learn tractor road features. The retraining parameters are: learning rate = 0.001, weight decay = $5\times10^{(-4)}$, and momentum = 0.937. Training uses a batch size of 16, with the loss curve monitored to ensure it de stops when the loss converges. The trained YOLOv5m is then used to detect tractor roads during grain truck operation.

To verify the feature learning of YOLOv5m for tractor road detection, this paper visualizes the feature map of the YOLOv5m Backbone, as shown in Figure 9. The Conv, C3, and SPPF modules of the trained YOLOv5m model were visualized to assess their effect on extracting features from different tractor road areas. By comparing these features with the background, we observe how YOLOv5m extracts and enhances tractor road features for accurate detection [15]. For effective extraction, the network should distinguish features of different targets, such as obstacles and passable areas. The visualized feature maps from the Backbone show the features extracted at different levels of the network, confirming its learning capability.

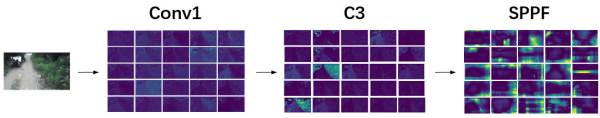


**Fig. 9.** Feature maps in the existing YOLOv5m backbone

To train the improved YOLOv5m, the feature maps of the Conv, C3, and SPPF modules in the Backbone were visualized, as shown in Figure 10. Compared to the original YOLOv5m, the improved YOLOv5m focuses more on the tractor road, resulting in more accurate recognition.

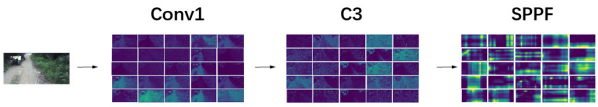


**Fig. 10.** Feature maps in the improved YOLOv5m backbone

# 5 Results and Discussion

## 5.1 Experimental Methods and Evaluation Indexes

This paper conducted tractor road detection experiments using the VOC_Road dataset on YOLOv5m, UNet, FCN, DeepLabV3+, and improved YOLOv5m networks. The VOC_Road dataset consists of 11,927 images with one class, labeled 'road' (11,927 semantic labels). It was randomly split into training (9,536 images), validation (1,192 images), and test sets (1,192 images) in an 8:1:1 ratio. $mAP_{50}$, Precision, and Recall were used as evaluation metrics. $mAP_{50}$ measures average detection accuracy at an IoU threshold of 50%, reflecting the algorithm's classification ability. Precision indicates the proportion of true positives among all predicted positives. Recall shows the proportion of true positives correctly identified out of all actual positives [16].

$$mAP_{50} = \int_0^1 p(r)dr \tag{8}$$

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

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

## 5.2 Detection Comparison Experiments Between Improved YOLOv5m and Existing Neural Networks

Using the VOC_Road dataset, the improved YOLOv5m was compared with the existing YOLOv5m, Deeplabv3+, FCN, and UNet for tractor road detection.

The comparison of test results was shown in Figure 11. As shown in the figure, compared with other neural networks, the improved YOLOv5m has fewer over-detection or missed detections and pays more attention to the pavement of tractor roads.

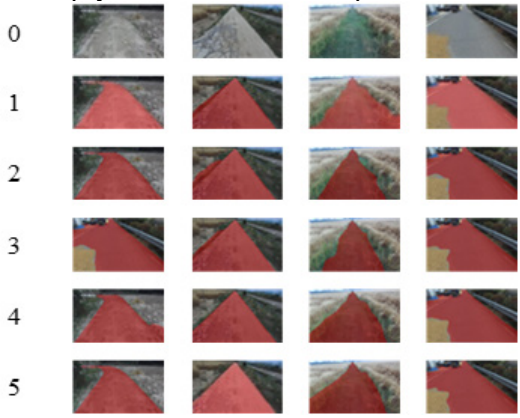


**Fig. 11.** The results of tractor road detection by improved YOLOv5m and existing neural networks

Using formulas (8)-(10), the detection results of the existing YOLOv5m, Deeplabv3+, FCN, Unet, and improved YOLOv5m on the VOC_Road test set were statistically analyzed, and the detection results of the tractor road of the improved YOLOv5m and the existing neural network were compared. The results were shown in Table 1.

**Table 1.** Comparison of detection results by improved YOLOv5m and existing neural networks

| Detection Methods | $mAP_{50}$/% | Pre/% | Re-call/% | time consuming /ms |
|---|---|---|---|---|
| Existing YOLOv5m | 92.4 | 93.2 | 90.7 | 48.2 |
| Deeplabv3+ | 91.2 | 92.3 | 89.4 | 50.2 |
| FCN | 89.8 | 91.0 | 87.6 | 47.8 |
| UNet | 90.5 | 91.5 | 88.2 | 45.3 |
| Improved YOLOv5m | 95.1 | 95.8 | 90.9 | 52.1 |

As shown in Table 1, compared with other existing networks such as YOLOv5m, Deeplabv3+, FCN, and UNet, the improved YOLOv5m improves the $mAP_{50}$ of VOC_Road by 2.7%, 3.9%, 5.3%, and 4.6%, respectively; The precision of tractor road detection increased by 2.6%, 3.5%, 4.8%, and 4.3% respectively; the recall of tractor road detection increased by 0.2%, 1.5%, 3.3%, and 2.7% respectively; Although the average time consumption has increased slightly, the overall results are within the controllable time range of detection, meeting the requirements of high-precision real-time detection of tractor roads by grain trucks.

## 6 Conclusions

To solve the low detection accuracy of the existing YOLOv5m for tractor roads in complex rural environments with blurry edges, this paper proposes an improved YOLOv5m method using attention mechanisms. The key contribution is integrating the Squeeze-and-Excitation attention mechanism into the Backbone of YOLOv5m to optimize feature extraction and enhance the model's focus on tractor road edges, improving accuracy. Results show that the $mAP_{50}$ of the improved YOLOv5m for tractor road detection increased by at least 2.7% compared to YOLOv5m, Deeplabv3+, FCN, and Unet. This improvement boosts the model's ability to recognize tractor road details while maintaining computational efficiency. Detection accuracy in complex environments is significantly enhanced. The method enables intelligent the grain truck to predict tractor roads quickly and accurately, advancing the goal of autonomous operation.

## Acknowledgments

This research was funded by National Natural Science Foundation of China, grant number 52302495; China Postdoctoral Science Foundation, grant number 2024T170354; Zhenjiang Key R&D Plan (Industry Foresight and Common Key Technology) Project, grant number GY2023001.

## References

1. Shang, E., An, X., Li, J., Ye, L., & He, H. (2013). Robust unstructured road detection: the importance of contextual information. Int. J. Adv. Robot. Syst., 10(3), 179.
2. Gao, C., Zhao, F., Zhang, Y., & Wan, M. (2024). Research on multitask model of object detection and road segmentation in unstructured road scenes. Meas. Sci. Technol., 35(6), 065113.
3. Wang, Q., Lyu, C., & Li, Y. (2024). A Method for All-Weather Unstructured Road Drivable Area Detection Based on Improved Lite-Mobilenetv2. Appl. Sci., 14(17), 8019.
4. Ye, J., Shi, S., Li, H., Wang, S., & Yang, S. (2019). Research and implementation of driving concern area detection based on deep learning. J. Syst. Simul., 31(7), 1421-1428.
5. Kao, Y., Che, S., Zhou, S., Guo, S., Zhang, X., & Wang, W. (2024). LHFFNet: A hybrid feature fusion method for lane detection. Sci. Rep., 14(1), 16353.
6. Song, H., He, D., & Xin, X. (2011). Unstructured road detection and obstacle recognition algorithm based on machine vision. Trans. Chinese Soc. Agric. Eng., 27(6), 225-230.
7. Wu, Y., & Lu, Y. (2019). An intelligent machine vision system for detecting surface defects on packing boxes based on support vector machine. Meas. Control, 52(7-8), 1102-1110.
8. Guo, H., & Chen, L. (2024). Multi-object road waste detection and classification based on binocular vision. J. Eng., 2024(5), e12389.
9. Li, M. L., Sun, G. B., & Yu, J. X. (2023). A pedestrian detection network model based on improved YOLOv5. Entropy, 25(2), 381.
10. Ren, M., Zhang, X., Chen, X., Zhou, B., & Feng, Z. (2023). YOLOv5s-M: A deep learning network model for road pavement damage detection from urban street-view imagery. Int. J. Appl. Earth Obs. Geoinf., 120, 103335.
11. Yang, Q., Peng, J., Chen, D., & Zhang, H. (2023). Road Scene Instance Segmentation Based on Improved SOLOv2. Electronics, 12(19), 4169.
12. Zhou, X., Zou, X., Tang, W., Yan, Z., Meng, H., & Luo, X. (2023). Unstructured road extraction and roadside fruit recognition in grape orchards based on a synchronous detection algorithm. Front. Plant Sci., 14, 1103276.
13. Wan, C., Chang, X., & Zhang, Q. (2023). Improvement of Road Instance Segmentation Algorithm Based on the Modified Mask R-CNN. Electronics, 12(22), 4699.
14. Collegio, G. R., Dal Poz, A. P., Filho, A. G. G., & Habib, A. (2024). Convolutional Neural Networks for Road Detection: An Unsupervised Domain Adaptation Approach. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., 48, 65-71.
15. Zhang, Q., & Gao, G. (2020). Prioritizing robotic grasping of stacked fruit clusters based on stalk location in RGB-D images. Comput. Electron. Agric.,*172*, 105359.
16. Guo, Z., Hu, X., Wang, J., Miao, X., Sun, M., Wang, H., & Ma, X. (2024). A duplex transform heterogeneous feature fusion network for road segmentation. Sci. Rep., 14(1), 17438.