Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 1205153, 14 pages
https://doi.org/10.1155/2021/1205153

Research Article

YOLO-Highway: An Improved Highway Center Marking Detection Model for Unmanned Aerial Vehicle Autonomous Flight

Zhiwei Zhao,[1] **Jianfeng Han**,[2] **and Lili Song**[2]

[1]*School of Aviation, Inner Mongolia University of Technology, Hohhot 010051, China*
[2]*School of Information Engineering, Inner Mongolia University of Technology, Hohhot 010080, China*

Correspondence should be addressed to Jianfeng Han; hanjianfeng@imut.edu.cn and Lili Song; songlili@imut.edu.cn

Received 22 August 2020; Revised 1 April 2021; Accepted 21 May 2021; Published 9 June 2021

Academic Editor: Bartlomiej Blachowski

Automatic visual navigation flight of an unmanned aerial vehicle (UAV) plays an important role in the highway maintenance field. Automatic highway center marking detection is the most important part of the visual navigation flight of a UAV. In this study, the UAV-viewed highway data are collected from the UAV perspective. This paper proposes a model named the YOLO-Highway that uses an improved form of the You Only Look Once (YOLO) model to enhance the real-time detection of highway marking problems. The proposed model is mainly designed by optimizing the network structure and the loss function of the original YOLOv3 model. The proposed model is verified by the experiments using the highway center marking dataset, and the results show that the average precision (AP) of the trained model is 82.79%, and the frames per second (FPS) is 25.71 f/s. In comparison with the original YOLOv3 model, the detection accuracy of the proposed model is improved by 7.05%, and its speed is improved by 5.29 f/s. Moreover, the proposed model had stronger environmental adaptability and better detection precision and speed than the original model in complex highway scenarios. The experimental results show that the proposed YOLO-Highway model can accurately detect the highway center markings in real-time and has high robustness to changes in different environmental conditions. Therefore, the YOLO-Highway model can meet the real-time requirements of the highway center marking detection.

1. Introduction

The unmanned aerial vehicles (UAVs) have been applied to many fields, including power inspection, building maintenance, and agricultural plant protection. In recent years, highway mileage has rapidly increased worldwide, so highways have been facing a variety of problems, the number of highway maintenance tasks has also rapidly increased, and investment in highway maintenance has increased substantially. Previously, the UAV used for highways worked mainly through human operations, which had high labor costs and low efficiency. Applying UAV based on visual navigation to highway tasks is a very meaningful and feasible way to improve the efficiency of highway inspection, which can get aerial imagery of highway information quickly and conveniently. Visual navigation flight of a UAV in highway inspection is going to be one of the future research topics of the highway maintenance industry.

The ability to detect the highway center marking timely and correctly is a prerequisite for the vision-based navigation flight of UAV. As a basic traffic sign, the highway center marking constrains, as well as guides, the UAV's flight, and it includes almost no interference information. Thus, in this study, the highway center marking is selected as feature information. Due to its small proportion in a highway image, how to better detect the highway center marking is a challenge in the field of visual navigation of UAVs.

In the fields of highway maintenance and UAV navigation, more and more studies have been done in recent years. Li Z. et al. [1] introduced a UAV highway maintenance algorithm based on the random forest classification (RFC) method to evaluate highway health conditions. In 2009, Joseph et al. [2] proposed a data fusion-based visual navigation method for UAVs that solved the problems of purely visual rotational drift and easy following of fast-moving targets. In 2014, Forster et al. [3] designed a

miniature quadcopter called the Nano + equipped with a camera and an embedded onboard computer. They used a semidirect visual odometry (SVO) method combined with the feature point method and the direct method to obtain position information, thus enabling an efficient flight in small indoor scenes with a frame rate of more than 50 fps of positional output. In 2016, Fraundorfer et al. combined the VFH+ (Vector Field Histogram+) algorithm and the boundary exploration algorithm as the main information source to realize autonomous flight and real-time 3D reconstruction of quadrotors indoors and outdoors, but the binocular computation burden was large, and the system was inefficient on the whole, and there were also mapping-accuracy problems [4]. All mentioned visual navigation methods have relatively low efficiency in image feature extraction and detection, relatively poor robustness, and a not-ideal navigation effect.

Currently, a convolutional neural network (CNN) is a popular method in the field of object detection [5,6]. Deep-learning-based detection models can be mainly divided into two categories: two-stage models and one-stage models. The two-stage models mainly include the R-CNN [7], Fast R-CNN [8], and Faster R-CNN [9], which have good detection performance in terms of accuracy. Girshick et al. [10] proposed an R-CNN model that uses a selective search algorithm to calculate the candidate area of images, and then all candidate areas were imported to this model. Compared with the traditional object recognition algorithm, the operation speed of the R-CNN algorithm is still slow, but its accuracy is significantly improved. Inspired by the R-CNN model, a series of excellent two-stage models were proposed to achieve better detection accuracy and speed. These models are more effective, and they include Fast R-CNN and Faster R-CNN using features computed from a single scale. In 2015, using a spatial pyramid pooling structure, HE et al. [11] proposed an effective structure named the SPP-net to eliminate the limitation of fixed network size. The SPP-net can make network input size unlimited, but the detection speed is limited since the training function is stored on the disk. On the basis of the R-CNN, the Fast R-CNN model combines the SPP structure concepts to enhance the detection speed and performance, but the detection speed is still relatively slow. To shorten the extraction time of regions from images, the Faster R-CNN introduces an RPN network instead of the selective search algorithm, but this network detects only seven frames per second in video detection, which does not meet the requirements of real-time detection [12–14]. As a one-stage detection model, in 2016, the YOLO was proposed. This model can directly detect the category and frame of an object in the image without extracting the candidate regions [15]. In addition, a more accurate model called the single-shot multibox detector (SSD) was proposed by Liu et al. [16]. This detector performs detection regression using a multilayer feature map. Compared with the two-stage models, one-stage models have a faster detection speed but similar detection accuracy. In particular, the YOLO has a shorter processing time than SSD.

Many studies have been conducted, and a variety of improved algorithms for visual object detection have been proposed in the UAV field. Han et al. [17] put forward an optimized Faster R-CNN model by performing an integration process and sharing regional characteristics. Zheng et al. [18] optimized the backbone network of YOLOv3 and tested remote sensing images under different lighting conditions. In 2019, Karaduman et al. [19] proposed a new lane detection method by using the k-nearest neighbor and Hough transformation methods. Zhao et al. [20] proposed the YOLOv3-LITE model that can be used in a portable computing platform and can successively detect objects in real-time. At present, the research on highway marking detection is still at a beginning stage, especially in UAV's applications due to the significant increase in requirements for highway marking detection under UAV's visual navigation flight backgrounds, such as high detection performance, short processing time, and enhanced system robustness and generalization ability.

This research aims to obtain a proper balance between the detection accuracy and the detection speed of an object—that is, reducing the processing time while ensuring the detection performance on the UAV platform can meet the detection demand. Thus, a model named the YOLO-Highway that is developed based on the latest research results of the YOLOv3 model using highway images obtained from the UAV perspective is proposed. The proposed model can save memory usage and have a faster processing speed than the YOLOv3 model while ensuring a certain accuracy, which means that it can be implemented in a UAV platform for accurate and efficient object detection.

The main contributions of this work are as follows:

(i) A UAV perspective-based highway center marking detection dataset is created, and data enhancement methods that can be used to optimize the detection result of highway center marking are introduced.

(ii) An improved network structure of YOLOv3 is proposed to improve the detection efficiency and reduce the memory cost, which further enhances the performance in highway marking detection.

(iii) The original loss function of the YOLOv3 model is optimized using the generalized intersection over union (GIoU) metric, which can calculate the bounding box regression loss with high accuracy. The improved loss function can be applied to highway center marking detection under various highway environmental conditions.

The rest of this paper is organized as follows. Section 2 introduces the used materials and the proposed method in detail. Section 3 describes the experimental results of the proposed model and compares the performances of the proposed model and the original YOLOv3 model. Section 4 summarizes the research results and presents future work directions.

2. Materials and Methods

2.1. Data Acquisition. Acquiring suitable data necessary for the training and test of a neural network model is an

indispensable part of deep learning. The highway center markings can be mainly divided into five categories: white solid lines, yellow solid lines, white dotted lines, yellow dotted lines, and yellow double solid lines. The five categories of central highway markings are shown in Figure 1.

There have been several datasets devoted to the highway domain, including the KITTI dataset [21], the CULane dataset, and the BDD100k dataset. However, these datasets seldom contain complete highway marking pictures from the UAV perspective. To obtain a dataset suitable for highway center marking detection, this study creates a new dataset. All images in the new dataset were taken by a low-flying UAV. The images used in this study can be mainly divided into two categories. In the first category, 362 images were selected from the UAV123 and CULane datasets that contain complete highway marking images, as shown in Figure 2(a). The second category included 1528 images of highways containing five major highway center markings, which were captured in Hohhot of China using a DJI drone with a resolution of 1920×1088 pixels. All the images were taken under natural daylight conditions, including environmental condition variation. Figure 2(b) shows the images taken by the DJI drone. Subsequently, the LabelImg software was used to label the highway center markings in the image blocks in the PASCAL VOC [22] format that is suitable for model training, as shown in Figure 2(c).

For the overall data, 75% of images were randomly selected as the training dataset, 5% as the validation dataset, and 20% as the test dataset. The specific division of the data is shown in Table 1.

2.2. YOLOV3 Model. The YOLOv3 [23] is one of the most advanced object detection models that evolved from YOLO and YOLOv2 [24]. Different from the two-stage detection models such as Fast R-CNN, YOLOv3 is a one-stage detector that formulates the object detection problem as a single regression problem. The YOLOv3 model can directly generate the bounding box coordinates and probabilities for each class using the regression without proposed regions. In this way, the YOLOv3 model significantly shortens the processing time compared to the two-stage models such as R-CNN and Fast R-CNN.

The YOLO v3 model is currently the leading state-of-the-art model in the field of real-time object detection. One of the most important features of the YOLOv3 is the introduction of residual network structure to form a deeper network level. The main detection process of the YOLOv3 model is as follows. First, an image is resized to a fixed size of 416×416 pixels and used as model input. Then, the Darknet53, which is the backbone network of the YOLOv3, is employed to extract image features, and finally, a 13×13 tensor is output as a prediction result. The YOLOv3 separates the input image into a number of 7×7 grids, each of which predicts objects that drop in the center of the grid. The detection principle of the YOLOv3 detection model is shown in Figure 3. As shown in Figure 3, each grid calculates the C conditional class probabilities, confidence scores, and their borders. In the model testing process, the class confidence scores are defined as follows:

$$p_r(C_i|\text{Object})\, p_r(\text{Object}) \times IoU_{\text{pred}}^{\text{truth}}, \quad p_r(\text{Object}) \in \{0, 1\}, \tag{1}$$

where $p_r(|C_i|\text{Object})$ denotes the C conditional class probability; $p_r(\text{Object})$ indicates whether there is an object in the grid, and when the grid contains an object, $p_r(\text{Object}) = 1$; otherwise, $p_r(\text{Object}) = 0$; $IoU_{\text{pred}}^{\text{truth}}$ denotes the intersection area between the ground truth box and the predicted bounding box; $p_r(\text{Object}) \times IoU_{\text{pred}}^{\text{truth}}$ is the confidence score that reflects whether the bounding box includes an object and evaluates the accuracy of the predicted bounding box.

2.3. Backbone Networks. The backbone network, as a feature extractor in the object detection model, plays a critical role in object detection, affecting the speed and accuracy of the detection model.

2.3.1. Darknet53. The YOLOv3 uses DarkNet53 as a backbone network. The main structure of the Darknet53 is shown in Figure 4(a). The Darknet53 of YOLOv3 mainly consists of the Residual unit named the Res unit for convenience. The Res unit is derived from ResNet [25], and it can solve the gradient disappearance or gradient explosion problems in the training process. As shown in Figure 4(c), the Res unit consists of CBL blocks, whose structure is shown in Figure 4(b), where it can be seen that CBL includes a convolution sublayer, a batch normalization sublayer, and a Leaky ReLu (Rectified linear unit) activation sublayer. The Res unit enlarges the receiving field by using 3×3 and 1×1 convolutional layers and a shortcut connection. This network consists of several consecutive 1×1 and 3×3 convolutions, with a total of 53 layers.

2.3.2. Cross-Stage Partial Network. In 2019, Wang et al. [26] proposed the cross-stage partial network (CSPNet) that integrates feature maps from the beginning and the end of a network stage to respect the changes in gradients. To enable the detection model to achieve a richer gradient combination and improve the detection performance, the CSPNet first divides the feature map of the base layer into two parts, the short part and the main part, and then merges these parts using the proposed cross-stage hierarchy. By segmenting the gradient flow, the CSPNet enables the gradient flow to spread through different network paths. In this way, the gradient information can have an obvious connection difference. Based on the mentioned, the CSPNet is an efficient model for reducing the memory cost of the training process. It can also achieve a proper balance between detection speed and precision. The CSPNet has been successfully applied to the Darknet53 and ResNet, developing the CSPDarknet53 and CSPResNeXt, respectively [27]. The CSPDarknet53 has a good object detection result on the MS COCO dataset [28].

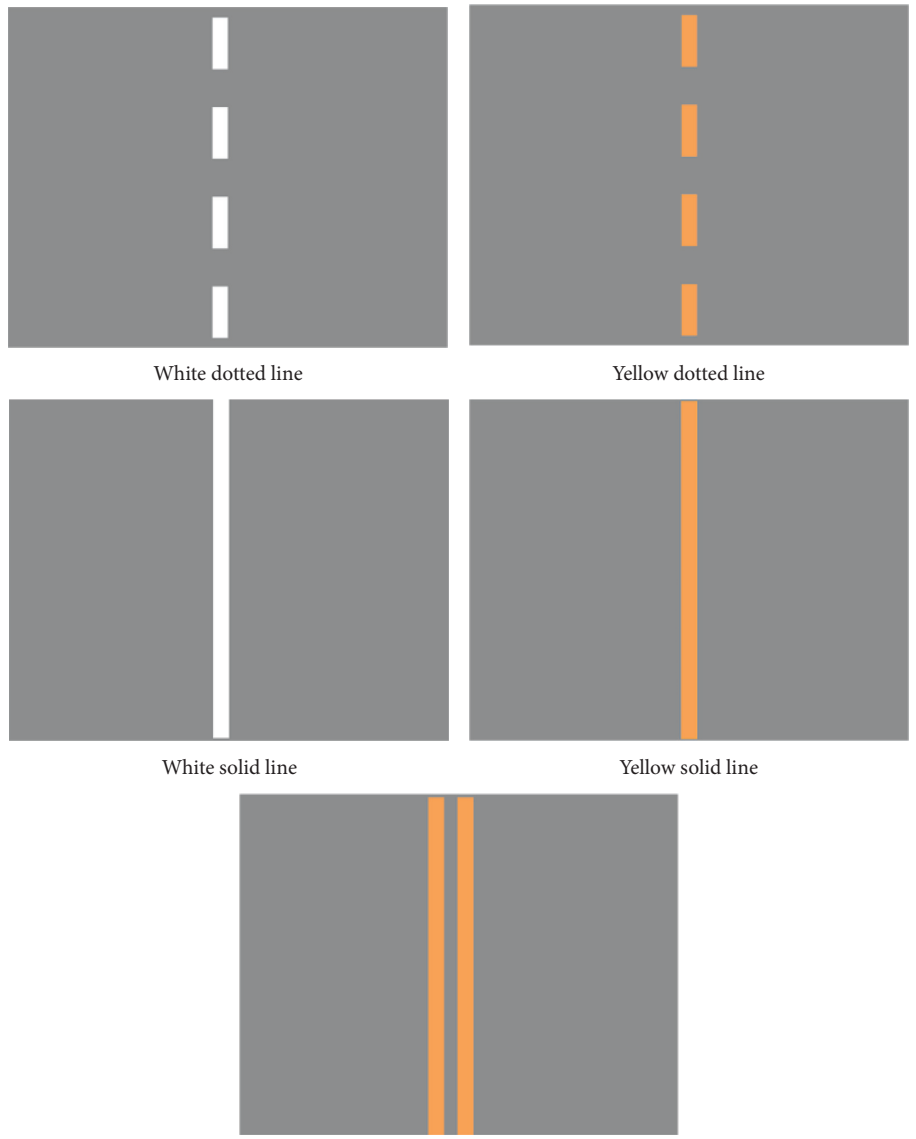
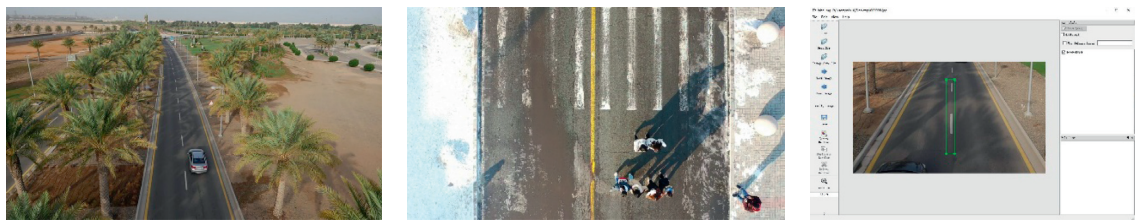


Figure 1: Five categories of central highway markings.



(a) (b) (c)

Figure 2: Dataset used for model development. (a) Image from the UAV123 dataset, (b) image collected by the DJI drone, (c) image processed by the LabelImg software.

Table 1: Data division into three sets.

	Training set	Validation set	Test set
Number of images	3188	213	849

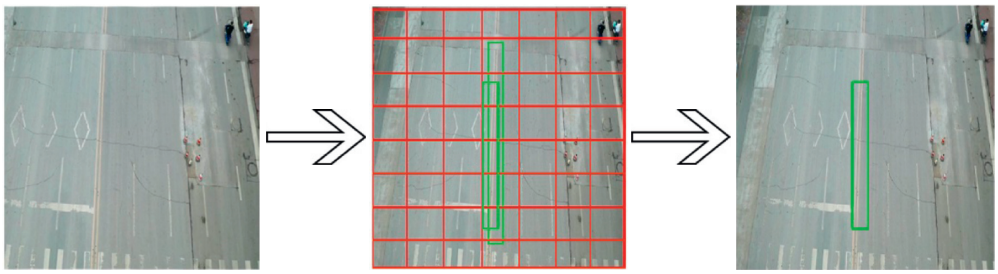


Figure 3: YOLOv3 detection principle.

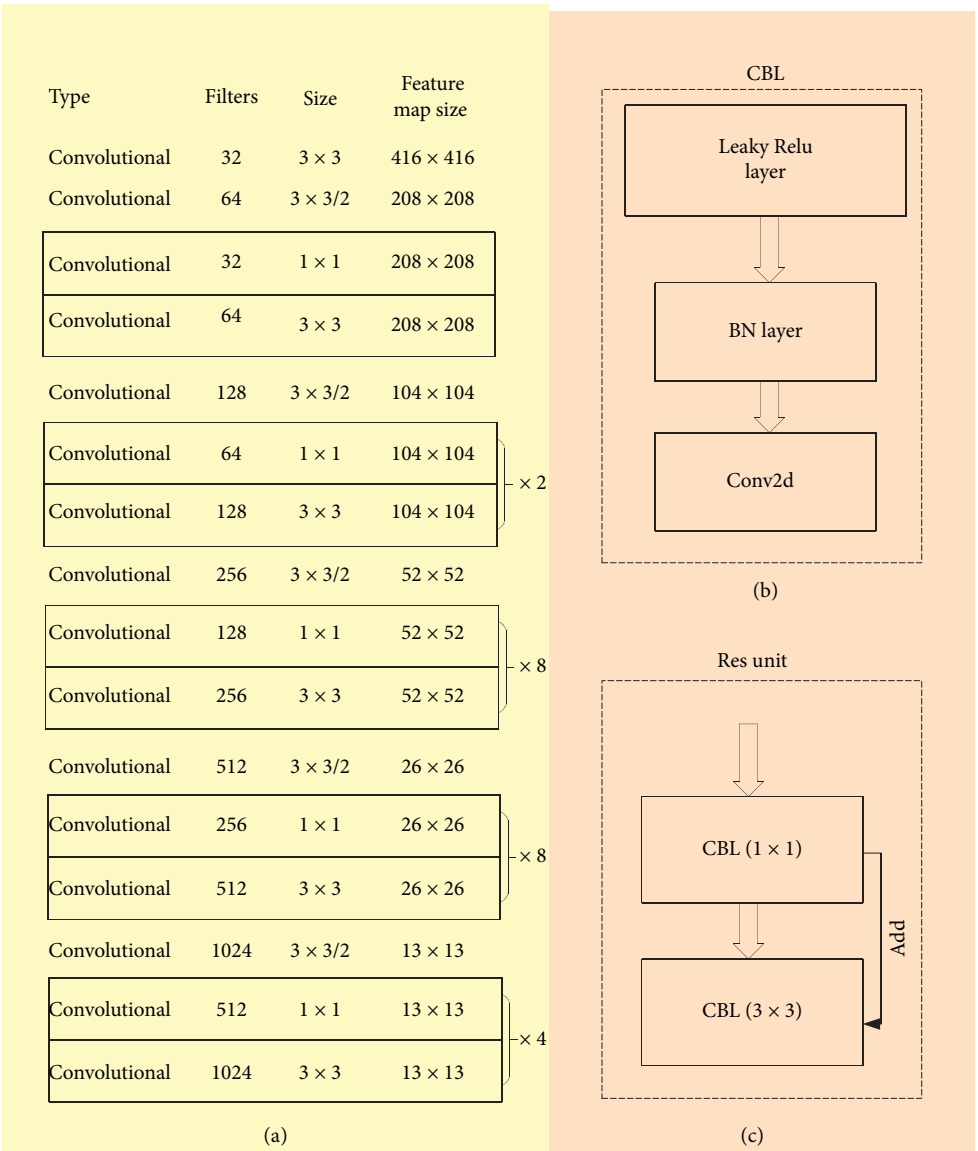


Figure 4: The structure of the Darknet53 convolutional network.

2.4. Method Flow. Due to the particularity of highway center marking images, the detection performance of the original YOLOv3 model cannot meet the real-time and accuracy requirements of the UAV platform. As mentioned previously, this study adopts a new dataset that is suitable for highway center marking detection, and an improved YOLO-Highway model is designed based on the original YOLOv3 model for highway center marking detection by a UAV platform. In detail, the network structure of the YOLOv3 model is improved to reduce memory usage and enhance detection speed and accuracy. Besides, in the optimized YOLOv3 network structure, the original loss function of the

YOLOv3 model is also replaced. The rationality and effectiveness of the improved structure are verified by experiments. The flowchart of the proposed method is shown in Figure 5.

2.5. Improved Yolov3 Model. To extract the feature information more efficiently, an advanced detection model focusing on highway center marking is developed based on the YOLOv3. The developed model can enhance the detection performance of highway center marking while reducing memory usage and detection time. The developed model is named the YOLO-Highway.

In many deep networks, the Mish activation function performs better than the Leaky ReLu activation function and other standard activation functions. The Mish function is a self-regularized, nonmonotonic activation function whose smooth function curve allows penetration of more information into a neural network, resulting in better detection performance and generalization ability. Its function curve is nonmonotonic and uncapped and also smoother than that of the ReLu function, suggesting that it can extract high-level latent features for better generalization, retain relatively small negative inputs, heighten the interpretability, and improve the gradient flow. The Mish activation function is given by equation (2), and its curve is shown in Figure 6. In the proposed model, in addition to the backbone network, the Leaky ReLu sublayer is used as an activation sublayer.

$$\text{Mish} = x * \tanh\left(\ln\left(1 + e^{x}\right)\right). \tag{2}$$

The structure of the proposed YOLO-Highway model is shown in Figure 7. Similar to the original YOLOv3 model, the YOLO-Highway model simultaneously predicts several bounding boxes and class probabilities for these boxes by using a single convolutional network. A resized image having the format of $416 \times 416 \times 3$ is used as the YOLO-Highway input. Then, the backbone network of the YOLO-Highway extracts the input feature. In the verification experiment of the YOLO-Highway, three boxes were predicted at each scale, which means the tensor $N \times N \times [3(4 + 1 + 1)]$ was used for four bounding box coordinates, one confidence score, and one class (highway center marking) object.

To improve the object detection ability of highway center markings, the original backbone network of the YOLOv3 model is improved. The main purpose of the CSPDarknet53 is to enable the model to achieve a richer gradient combination and reduce memory usage and processing time. By learning from the structure of the CSPDarknet53, this study mainly optimizes the Darknet53 to the HDarknet53. The main difference between the HDarknet53 and CSPDarknet53 is that the HDarknet53 has been partially streamlined to obtain better detection speed and less memory usage. Based on the CSPDarknet53, the HDarknet53 is mainly made by CSP unit named CSPX for convenience. The structure of the proposed backbone network is shown in the dotted box in Figure 7(a). The optimized model introduces the CBM block, which is shown in Figure 7(c), where it can be seen that it consists of a convolution sublayer, a batch normalization sublayer, and a Mish Relu activation sublayer. The main components of the CSPX are as follows. First, a CBL block Figure 7(b) is added before and after the Res unit Figure 7(d). Then, the CSPX increases another shortcut connection to concatenate the above structure and a CBL block and two CBM blocks ahead. The CSPX structure is shown in Figure 7(e). In order to reduce memory usage and improve detection speed, the first CBM module of the original CSPDarknet53 is deleted in the proposed model.

Based on the original YOLOv3 model, the YOLO-Highway model uses a scale pyramid structure similar to the FPN network, adopting two upsampling and stitching with the same size feature map in the upper layer of the network. Since the highway marking information has a small scale in a highway image, the proposed model tends to abandon a 13×13 layer of the original model to reduce the parameter

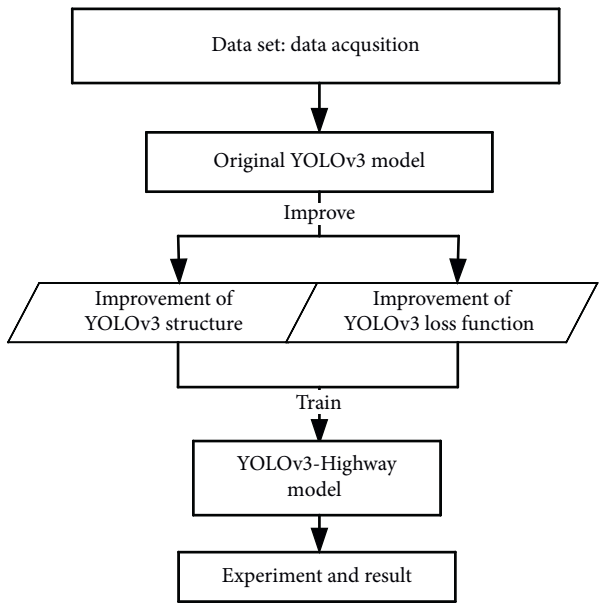


Figure 5: The proposed method flowchart.

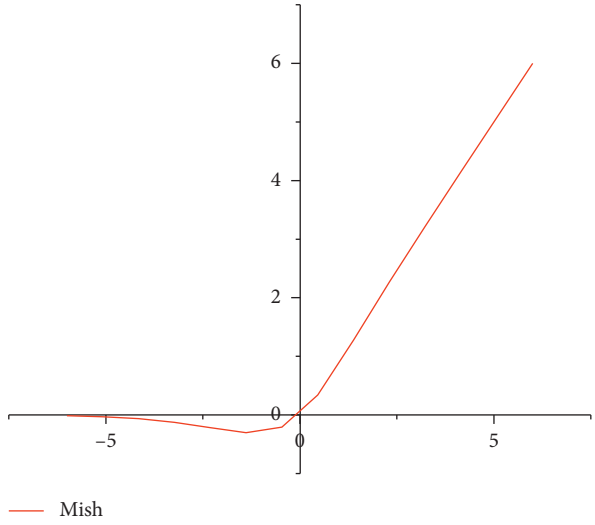


Figure 6: Mish activation function curve.

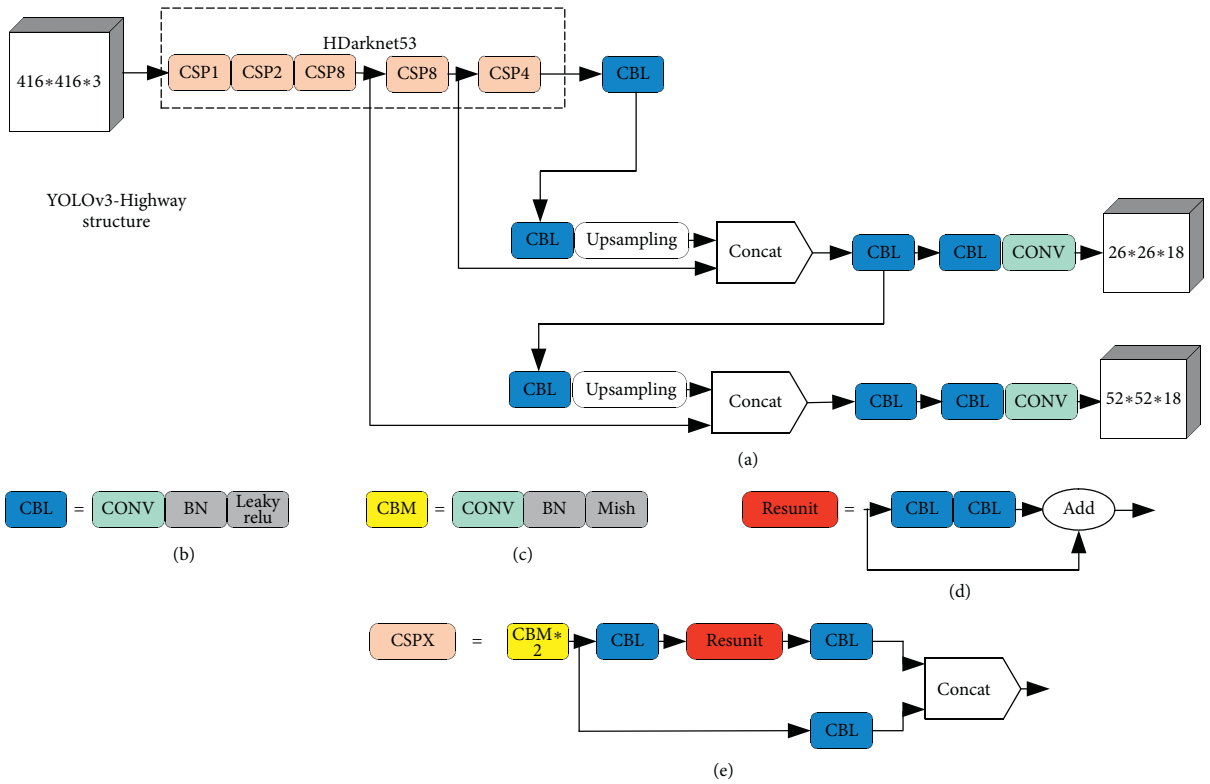


Figure 7: The structure of the proposed YOLO-Highway.

number. To use the features better, the original six layers are pruned to two layers before each detection layer by removing the first four layers. The comparison of the backbone network parameters between the YOLOv3 and YOLO-Highway is presented in Figure 8. The black solid rectangle in Figure 8(a) denotes the Res unit, and the dotted rectangle in Figure 8(b) represents the CSP unit. As shown in Figure 8, the parameters of the YOLO-Highway model are more complex than those of the original YOLOv3 model, but the YOLO-Highway model can achieve the network permutation complexity without extra processing time.

To sum up, the YOLO-Highway is developed by optimizing the original YOLOv3 from two aspects: (i) using the HDarknet53 instead of the Darknet53; (ii) modifying the hierarchical structure of the YOLOv3 model to reduce model redundancy. Accordingly, the proposed model can greatly improve the speed and accuracy of highway center marking detection of the UAV platform.

2.6. Loss Function. The YOLOv3 model uses the intersection over union (IoU) function to evaluate the object detection performance. In detail, the IoU indicates the overlapping degree between the boxes predicted by the detection model and the real bounding boxes of an object. The traditional IoU loss function has two main disadvantages. First, when the IoU value is zero, it cannot be clearly known whether objects are adjacent to each other or far apart. Second, when the IoU values of two objects are equal, the overlapping degree between the two rectangular boxes cannot be indicated. Based on the mentioned, the IoU cannot accurately reflect the overlapping degree between two objects. For instance, consider an example presented in Figure 9. Although the IoU values of the three cases shown in Figure 9 are equal, the overlapping cases are completely different in Figure 9, and the regression effect of the leftmost graph is better than that of the rightmost graph [29]. Therefore, the IoU loss function cannot clearly indicate the real overlapping degree between two objects.

In highway marking detection, accurately assessing the overlapping degree between the predicted and real bounding boxes directly determines the success rate of object detection. Therefore, this study introduces the GIoU to solve the problem of IoU. In order to indicate the overlapping degree between two objects better, different from the IoU, which focuses only on overlapping areas of two boxes, the GIoU is used, and it focuses on all areas. The calculation formulas of the IoU and GIoU are given in equations (3) and (4), respectively, and the bounding box regression loss is calculated by equation (5).

$$IOU = \frac{|B \cap G|}{|B \cup G|}, \tag{3}$$

$$GIOU = IOU - \frac{|C/(B \cup G)|}{|C|}, \tag{4}$$

$$\text{Coord}_{\text{loss}} = 1 - GIOU. \tag{5}$$

In equations (3)–(5), C denotes the minimum enclosing rectangle of the predicted bounding box B and real bounding box G. Equation (4) shows that the value of the GIoU

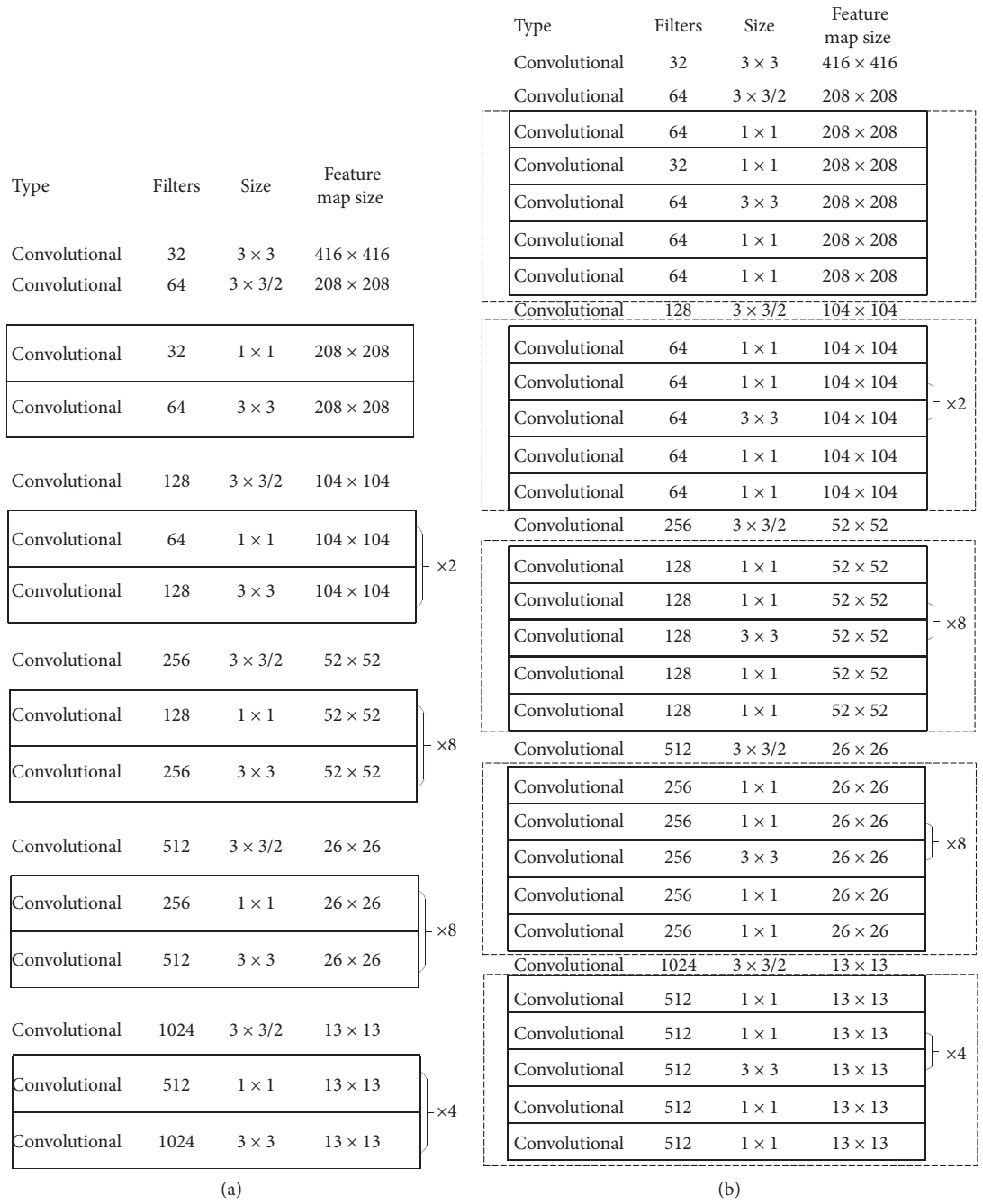
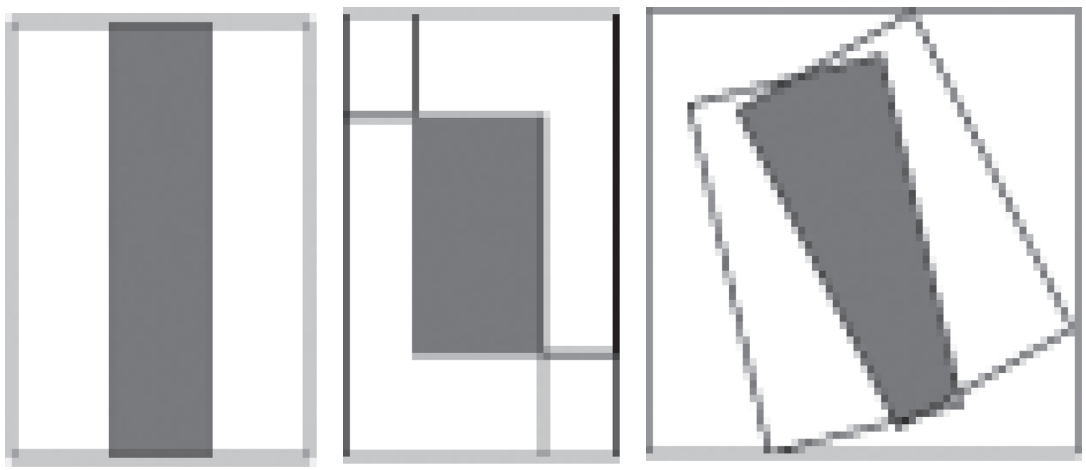


Figure 8: The backbone structures of the YOLOv3 and YOLO-Highway models. (a) Original YOLOv3 structure. (b) The proposed YOLO-Highway structure.

function is always lower than or equal to the value of the IoU function. Based on equation (3), it holds that $0 \leq \mathrm{IoU} \leq 1$, so from equation (4), it can be obtained that $-1 \leq \mathrm{GIoU} \leq 1$, which means that only when the predicted bounding box GioU = 1, it completely agrees with the real frame. Further, when the predicted and real bounding boxes' IoU values are both equal to zero, GIoU is converted to the following equation:



(a) (b) (c)

Figure 9: Three different overlapping situations under the same IoU loss value.

$$GIOU = -1 + \frac{|(B \cup G)|}{|C|}. \tag{6}$$

At this time, when the predicted bounding box differs significantly from the real bounding box, the GIoU is close to −1, and $B \cup G$ remains unchanged. To optimize the GIoU, C must be gradually reduced, that is, the predicted and real boxes must be as close as possible. The confidence loss, classification loss, and the final loss are, respectively, defined by the following equation:

$$\begin{aligned} \text{Conf}_{\text{loss}} &= \sum_{i=0}^{s^2} \sum_{j=0}^{B} 1_{ij}^{\text{obj}} \left[\left(C_i - \hat{C}_i \right)^2 \right] \\ &\quad + \lambda_{\text{noobj}} \sum_{i=0}^{s^2} \sum_{j=0}^{B} 1_{ij}^{\text{noobj}} \left[\left(C_i - \hat{C}_i \right)^2 \right], \\ \text{Class}_{\text{loss}} &= \sum_{i=0}^{s^2} 1_{ij}^{\text{obj}} \sum_{c \in \text{classes}} \left(p_i(c) - \hat{p}_i(c) \right)^2, \end{aligned} \tag{7}$$

where i indicates the i th square, j indicates the j th bounding box predicted by the square, and 1_{ij}^{obj} indicates whether the $i$square contains an object; obj refers to the existence of an object, and noobjrefers to the absence of an object; C_i indicates the class of the predicted object, $\hat{C}_i$ indicates the class of the real object, and λ_{noobj} is the penalty coefficients. Then, the loss function can be expressed as follows:

$$\text{Loss} = \text{Coord}_{\text{loss}} + \text{Conf}_{\text{loss}} + \text{Class}_{\text{loss}}. \tag{8}$$

3. Results

In this section, the evaluation procedure of the proposed model's performance, including the ablation evaluation with different improvement strategies and the detection performance analysis on the test set, is presented. Moreover, the ablation experiments are conducted using the proposed model and YOLOv3 model under different environmental conditions. The experimental results show different improvements in the object detection model. Finally, the detection results on the proposed YOLO-Highway model are presented and compared with those of the original YOLOv3 model.

3.1. k-Means Clustering Method. The k-means clustering method [30] is used to determine the anchor box dimension in the UAV-viewed data. The k-means algorithm divides all samples into k clusters that are typically chosen to be sufficiently far apart from each other in space based on the Euclidean distance to produce valid data mining results. The dimension and number of anchor boxes are obtained by the final cluster centers calculated by the k-means algorithm.

It should be noted that larger labels will produce larger errors, which will have a bad influence on clustering results. Therefore, in this work, the average IoU value is used as a measure of similarity index. The distance is calculated by the following equation:

$$\text{d}(\text{box}, \text{centroid}) = 1 - U_{\text{Avg}IoU}(\text{box}, \text{centroid}), \tag{9}$$

where box refers to the sample, centroid represents the center of the cluster, and $\text{U}_{\text{Avg}IoU}$ (box, centroid) reflects the intersection of the cluster's center box and the cluster box [31]. Clustering analysis of the highway center marking is performed using the proposed dataset, continuously increasing the number of cluster centers starting from one to obtain the relationship between the number of cluster centers k and the average IoU value. The result of the k-means clustering is shown in Figure 10.

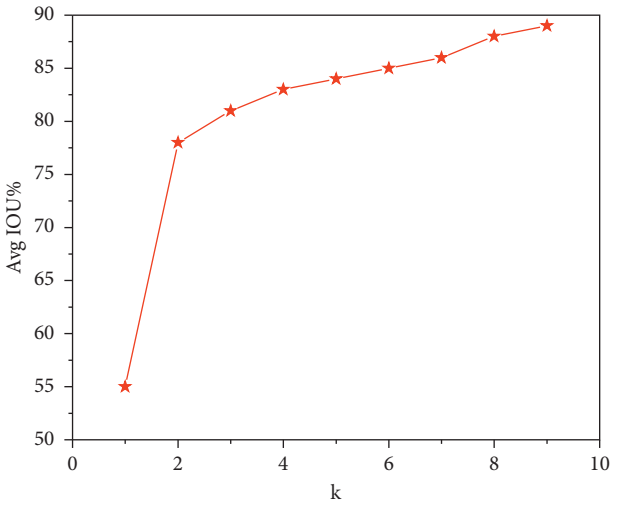


Figure 10: The result of the k-means algorithm.

As shown in Figure 10, when k reached a value of four, the curve gradually became flat, and the clustering results were similar. At that time, clustering was adopted to avoid redundancy. When the number of centers was four, the corresponding anchor served as a prediction frame for detecting the highway marking. Considering that relatively large labels that occupied a certain proportion in the dataset, a group of anchors with a larger size were added on the basis of the three anchors obtained by clustering to match the labels. Finally, (8, 374), (9, 242), (9, 120), (11, 111), (13, 245), and (69, 259) were selected as the anchor parameters.

3.2. Evaluation Indicators. In this study, the average precision (AP) and precision recall (PR) were used to evaluate the YOLO-Highway model's performance quantitatively. In addition, the *F1* score [32] and frames per second (FPS) were used to indicate the detection accuracy and speed comprehensively.

In a binary classification problem, according to the actual and predicted classification results, samples can be divided into four categories [33]: true positive (TP), fault positive (FP), true negative (TN), and fault negative (FN). Precision (P) and Recall rate (R) reflect the relationship between the samples that are predicted to be positive and the samples that are truly positive, and they are respectively defined as follows:

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}},$$
$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}. \tag{10}$$

The *PR* curve shows the proportion of true positive samples in all positive samples determined by the detector [34]. The average precision (AP) represents the average precision score of the recall rate from zero to one, and it is calculated by the following:

$$AP = \int_0^1 P(R)\mathrm{d}R. \tag{11}$$

In addition, the F1 score that denotes the weighted harmonic average of precision and recall was also used in the performance evaluation. The higher the F1 value is, the better the model performance is. The F1 score is defined as follows:

$$F1 = \frac{2PR}{P + R}. \tag{12}$$

The FPS refers to the detection speed, and it shows the number of frames transmitted per second. When the FPS value is higher, there are more frames per second, and the display effect is smoother. The FPS was used as an index of the detection processing speed, and it was defined as the number of images detected per second, in f/s.

3.3. Training Results Comparison. First, the YOLOv3 network model was trained, and then the proposed improved model YOLOv3-CSP, which replaced Darknet53 with CSPdarknet53, was also trained. To achieve a better detection result, the GIoU loss function replaced the original loss function. At the early stage of training, the attenuation coefficient was set to 0.0005, the learning rate was set to 0.001, and the step mode was selected to update the learning rate. When the number of training epochs reached a value of 250 or 300, the learning rate reduced to 10% or 1% of the initial learning rate, respectively. During the training process, the loss curve was used to observe the dynamic process of training. The average loss curves of the training of the three models are presented in Figure 11, where the relationship between the number of training iterations and the loss value during training is presented. The loss change curves indicate that the three models had both similarities and differences. The first similarity was that the trend of losing value in the three models in the first two hundred steps of training decreased sharply. The reason for this phenomenon was because the learning rate in the early stage was relatively large, so the proposed model could reduce the loss value to a relatively low level. Subsequently, with the gradual decrease in the learning rate, the decline speed of the loss value was relatively low, and the changes in the loss value tended to be stable. The main difference between the three models was that the loss value of the YOLO-Highway during the early stage dropped faster than those of the other two models. Because of the introduction of the GIoU loss function, the loss value of the YOLO-Highway model was much lower than those of the other two models. Benefiting from the optimized backbone network and loss function, the proposed model achieved a more comprehensive training result than the other methods while reducing memory usage.

3.4. Ablation Study on Different Modification. For the four different models, the values of *P*, *R*, AP, FPS, and *F1* scores were compared, and the obtained results are listed in Table 2. The results in Table 2 show that the improved YOLO-Highway model achieved a precision of 85.26%, a

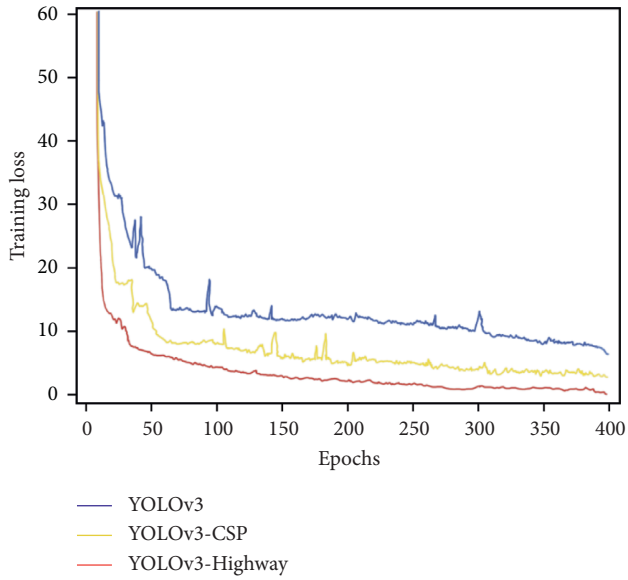


Figure 11: Loss curves on the verification set of three YOLOv3 models.

Table 2: Performance comparison of different models.

Detection model	*P* (%)	*R* (%)	AP (%)	FPS (f/s)	*F*1 (%)
YOLOv3	78	75	79.84	20.22	76.47
Faster-RCNN	80	78	80.21	18.15	78.98
YOLOv3-CSP	81	78	80.43	22.20	79.47
YOLO-highway	85	81	82.79	25.71	82.95

recall rate of 81.26%, an average precision of 82.79%, and an FPS of 25.71 f/s on the test set. By introducing the HDarknet53, the improved model achieved an improvement of 3, 3, 0.59, and 3 in the AP, R, AP, and F1 scores, respectively, with an increase of 1.98 in *FPS*, as shown in Table 2. Besides, the GIoU was used instead of the IoU loss function to calculate the regression loss, and the results showed that in this case, the values of P, R, and F1 were increased by 4, 3, and 3.48, respectively. The proposed model also gained an obvious improvement of 4.25% in AP with a speed increase of 3.51 f/s.

By comparing the original YOLOv3 model and the proposed model, significant improvements in parameters of 7, 6, 2.95, 5.49, and 6.48 were observed. Compared with the two-stage algorithm Faster-RCNN, the improved model improved the parameters by 5, 3, 2.58, 7.56, and 3.97. Benefiting from the optimized backbone network and loss function, the proposed model maintained a high precision while reducing the memory usage and improving the detection speed.

The PR curve is an important indicator of the model's robustness and effectiveness. The comparison results of the PR curve of the three YOLOv3 models used in this study are shown in Figure 12. When the recall value was nearly 0.84, the precision values of the YOLOv3, Faster-RCNN, and YOLOv3-CSP models dropped sharply to approximately 0.85, while the precision value of the YOLO-Highway was almost 0.9. Thus, the YOLOv3-Highway had an obvious advantage in the precision over the other models under the same recall rate, which demonstrated that introducing the GIoU loss function and replacing the Darknet53 with HDarknet53 could provide better model training and enhance the highway center marking detection performance.

3.4.1. Different Environmental Conditions Experiments. To evaluate the performance of the YOLO-Highway under different environmental conditions, the highway markings were divided into three groups corresponding to three typical environmental conditions. More than 100 pictures under different environmental conditions, including partial occlusion, partial damaged, and weak-light conditions, were collected. Then, the values of *P*, *R*, AP, and FPS of the YOLOv3 model and the YOLO-Highway model were compared under different environmental conditions. The experimental results are given in Table 3.

As presented in Table 3, compared to the original YOLOv3 model, the precision performance of the proposed model under three different environmental conditions was improved in the range of 2%–16%, and the average precision was also significantly improved. The results showed that the proposed model could detect the object well under all tested environmental conditions, which is of great practical value for center marking detection of the UAV platform.

3.5. Proposed Model Performance

3.5.1. Detection Performance for Five Highway Center Marking Types. The detection results of three different models were compared on the test set to evaluate the performance of the proposed model more comprehensively. The detection results of proposed model and original YOLOv3 model are shown in Figure 13. In Figure 13, the green boxes denote the detection results of the YOLOv3 model, and the blue boxes denote the detection results of the YOLO-Highway model. Also, in Figure 13, each row contains five main highway center marking types. The pictures in the first row in Figure 13 were obtained by the YOLOv3 model, and the pictures in the second row were obtained by the proposed YOLO-Highway model. Based on the results in Figure 13, it can be concluded that, although the original YOLOv3 model could detect markings, the confidence of the test results was all under 0.9, which was lower than that of the proposed YOLO-Highway model. This demonstrates that the YOLO-Highway model was more fully trained and had a better convergence effect than the original YOLOv3 model. Also, the proposed YOLO-Highway model used a better GIoU loss function, which enhanced the detection performance. Moreover, the more efficient backbone network and more succinct structure of the detection model could further enhance the model's detection performance, achieving excellent precision and generalization of the UAV platform.

3.5.2. Performance under Different Interference Conditions. The interference conditions of a highway marking image can significantly affect the detection performance of a model. In

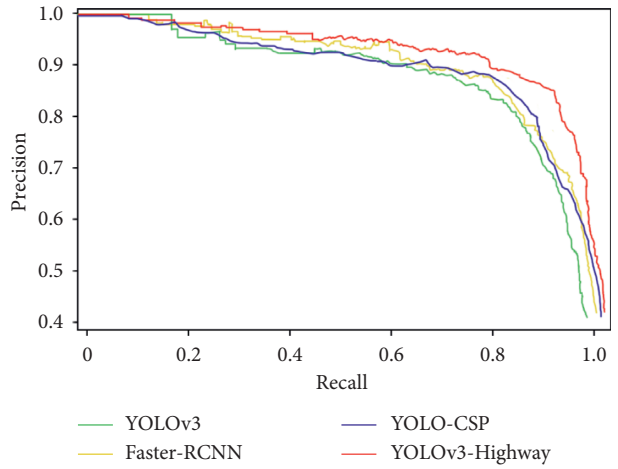


Figure 12: Comparison of the precision recall curves of the three YOLOv3 models.

Table 3: The performance comparison of two YOLOv3 models under different environmental conditions.

Detection model	Partial occlusion (%)				Partial damaged (%)				Weak-light conditions (%)			
	P	*R*	AP	*F*1	*P*	*R*	AP	*F*1	*P*	*R*	AP	*F*1
YOLOv3	65	49	37	56	70	44	36	54	73	55	44	63
YOLO-highway	71	57	49	63	72	58	50	64	86	74	73	80

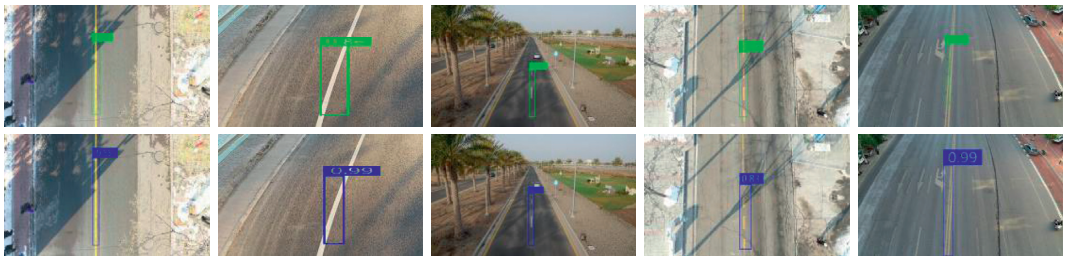


Figure 13: Comparison of detection results for five highway center marking types between the classical YOLOv3 model (the first row) and the proposed YOLO-Highway model (the second row).

this study, the YOLOv3 and YOLO-Highway models were tested under three typical conditions, and the test results are presented in Figure 16. In Figure 16, the images in the first row were obtained by the YOLOv3 model, and the images in the second row were obtained by the proposed YOLO-Highway model; the green boxes denote the detection results of the YOLOv3 model, and the blue boxes denote the detection results of the YOLO-Highway model. The images in Figures 14(a)–14(f) were acquired under different environmental conditions. Under weak-light conditions, the confidence of the test results in Figure 14(d) is higher than in Figure 14(a). In the case of partial damage of the highway making, the model trained on the proposed dataset was able to recognize occluded markings effectively, as shown in Figure 14(b), which is not the case in Figure 14(e). When highway marking was partially occluded by vehicle, the proposed YOLO-Highway model had better recognition performance in Figure 14(f) than in Figure 14(c).

The presented results demonstrate that the proposed model achieved a better performance than the original YOLOv3 model under different environmental conditions; also, the improved model had a stronger generalization ability.

With the continuous development of the UAV industry, UAVs have been applied to many industrial applications. In the future, the UAV systems could be used to achieve highway inspection, which could make them become a useful and meaningful approach for highway inspection. The vision-based navigation of highway marking on a highway is a core problem in highway inspection. In order to acquire the entire highway image, a UAV needs to fly in the center of a highway following the highway markings. Thus, highway marking detection is a core problem in autonomous vision-based navigation flight.

At present, highway marking detection is mainly studied in the autonomous driving field, while the related research in the UAV field is still relatively seldom. Developing a more efficient and faster object detection model to detect highway marking from the perspective of UAV could provide more possibilities for vision-based navigation of UAVs.

Figure 14: Images acquired under different environmental conditions. (a, d) Weak-light conditions, (b, e) partial damage of the highway making, and (c, f) partial occlusion of the highway making.

4. Conclusions and Future Work

In this paper, the YOLOv3 model is used as a basic framework for highway center marking detection, and an improved YOLO-Highway is proposed. First, the highway images obtained from the UAV perspective from the UAV123 datasets and our self-made dataset are combined into a joint dataset. Next, the obtained dataset is enlarged to heighten the robustness of the proposed model. Then, the structure of the original YOLOv3 model is optimized by introducing the CSPDarknet53 backbone network and GIoU loss function. In detail, the HDarknet53 can obtain a richer gradient combination while reducing the computation cost, and the GIoU loss function can make the model more robust. Finally, the k-means algorithm is used to generate parameters of anchor boxes based on the clustering results of highway center marking on the training set. The proposed model is verified by the experiments and compared with the original YOLOv3 model. The experimental results show that, compared to the original YOLOv3, the proposed YOLO-Highway model has improved the AP from 79.84% to 82.79% and the detection speed to 25.71 f/s. Besides, the proposed optimized model could reliably detect highway center marking under a variety of conditions while achieving good detection performance.

However, due to the small number of highway images under different environmental conditions, there are still certain deviations between the detection results and the ground truth. To further improve the detection performance in real-time of the proposed model and to make the proposed model more robust against interference, more highway images acquired under different environmental conditions by the UAV platform are necessary to expand the training dataset, which will be part of our future work.

Data Availability

The data used to support the findings of this study have not been made available because some part of our data were collected on some special highways.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was funded by the Key Technology Research Project in Inner Mongolia Autonomous Region under grant no. 2019GG271 and Key Scientific Research Projects of Colleges and Universities in Inner Mongolia Autonomous Region, China (NJZZ19068).

References

[1] Z. Li, "Identifying asphalt pavement distress using UAV LiDAR point cloud data and random forest classification," *International Journal of Geo-Information*, vol. 8, p. 1, 2019.

[2] S. Joseph, "Implementation of wide-field integration of optic flow for autonomous quadrotor navigation," *Autonomous Robots*, vol. 65, 2009.

[3] C. Forster, M. Pizzoli, and D. Scaramuzza, "Fast semi-direct monocular visual odometry," 2013.

[4] H. Fraundorfer, "Vision-based autonomous mapping and exploration using a quadrotor MAV," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots & Systems IEEE*, Algarve, Portugal, October 2012.

[5] J. Greenhalgh and M. Mirmehdi, "Real-time detection and recognition of road traffic signs," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 4, pp. 1498–1506, 2012.

[6] J. Jin, K. Fu, and C. Zhang, "Traffic sign recognition with hinge loss trained convolutional neural networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 5, pp. 1991–2000, 2014.

[7] R. Girshick, J. Donahue, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," 2014.

[8] R. Girshick, "Fast R-cnn," *Computer Science*, vol. 12, 2015.

[9] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-cnn: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, p. 6, 2015.

[10] R. Girshick, J. Donahue, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," 2014.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," 2014.

[12] Z. Lin, K. Ji, X. Leng, and G. Kuang, "Squeeze and excitation rank faster R-cnn for ship detection in sar images," *IEEE Geoscience and Remote Sensing Letters*, vol. 34, pp. 1–5, 2018.

[13] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-cnn: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[14] J. R. R. Uijlings, K. E. A. Van De Sande, T. Gevers, and A. W. M. Smeulders, "Selective search for object recognition," *International Journal of Computer Vision*, vol. 104, no. 2, pp. 154–171, 2013.

[15] J. Redmon, S. Divvala, R. Girshick, and F. Ali, "You only look once: unified, real-time object detection," 2015.

[16] W. Liu, D. Anguelov, D. Erhan, C. Szegedy et al., "ssd: single shot multibox detector," 2016.

[17] H. Xiaobing, Y. Zhong, and L. Zhang, "An efficient and robust integrated geospatial object detection framework for high spatial resolution remote sensing imagery," *Remote Sensing*, vol. 9, no. 7, p. 666, 2017.

[18] Q. Zhi-, L. Y.-Y. Zheng, P. Chang-Cheng, and L. I. Guo-Ning, "Application of improved yolo V3 in aircraft recognition of remote sensing images," *Electronics Optics & Control*, vol. 47, 2019.

[19] M. Karaduman, A. Çınar, and H. Eren, "UAV traffic patrolling via road detection and tracking in anonymous aerial video frames," *Journal of Intelligent & Robotic Systems*, vol. 92, 2019.

[20] H. Zhao, Y. Zhou, L. Zhang et al., "Mixed yolov3-lite: a lightweight real-time object detection method," *Sensors*, vol. 20, no. 7, p. 1861, 2020.

[21] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The kitti vision benchmark suite," in *Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition 2012*, Long Beach, CA, USA, March 2012.

[22] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, 2010.

[23] J. Redmon and F. Ali, "Yolov3: an incremental improvement," 2018.

[24] J. Redmon and F. Ali, "Yolo9000: better, faster, stronger," 2017.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition 2016*, Las Vegas, NV, USA, June 2016.

[26] C. Y. Wang, I. H. Yeh, and J. W. Hsieh, "Cspnet: a new backbone that can enhance learning capability of cnn," 2019.

[27] T. Y. Lin, M. Maire, S. Belongie, J. Hays, and C. Lawrence Zitnick, "Microsoft coco: common objects in context," 2014.

[28] A. Bochkovskiy and C. Y. Wang, "Yolov4: optimal speed and accuracy of object detection," 2020.

[29] J. Ma, W. Shao, H. Ye, L. Wang, and H. Wang, "Arbitrary-oriented scene text detection via rotation proposals," *IEEE Transactions on Multimedia*, vol. 99, p. 1, 2017.

[30] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," 2014.

[31] G. Ye, Z. Tang, D. Fang, Z. Zhu, and Z. Wang, "Using generative adversarial networks to break and protect text captchas," *ACM Transactions on Privacy and Security (TOPS)*, vol. 43, 2020.

[32] Y. Tian, G. Yang, Z. Wang, H. Wang, E. Li, and Z. Liang, "Apple detection during different growth stages in orchards using the improved yolo-V3 model," *Computers and Electronics in Agriculture*, vol. 157, pp. 417–426, 2019.

[33] D. Liu, G. Hua, V. Paul, and T. Chen, "Integrated feature selection and higher-order spatial feature extraction for object categorization," in *Proceedings of the Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on 2008*, Anchorage, AL, USA, June 2008.

[34] B. Benjdira, K. Taha, A. Koubaa, A. Ammar, and K. Ouni, "Car detection using unmanned aerial vehicles: comparison between faster r-cnn and yolov3," 2018.