IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 31 August 2023, accepted 27 September 2023, date of publication 2 October 2023, date of current version 9 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3321385

APPLIED RESEARCH

DF-YOLO: Highly Accurate Transmission Line Foreign Object Detection Algorithm

SHAO JIA LI, YAN XIA LIU JR., (Member, IEEE), MIAO LI, AND LU DING
College of Urban Rail Transit and Logistics, Beijing Union University, Beijing 100101, China

Corresponding author: Yan Xia Liu Jr. (yanxia.liu@163.com)

This work was supported in part by the Academic Research Projects of Beijing Union University under Grant ZK20202302, in part by the Beijing Natural Foundation under Grant L221015, and in part by the National Natural Science Foundation of China under Grant 62102032.

ABSTRACT Ensuring the promptly removal of foreign objects from transmission lines is crucial for electricity safety. However, the existing object detection algorithms exhibit low precision and recall due to factors such as uncertainty in the type of foreign objects, imbalance of positive and negative samples, and the complexity of aerial photography backgrounds. Therefore, in this paper, an algorithm called DF-YOLO (Deformable Faster-You Only Look Once) for transmission line foreign object detection is proposed to resolve these problems. The algorithm is based on YOLOv7-Tiny and tailored to the dataset's characteristics to achieve high precision along with excellent recall. The Focal-DIoU loss function is utilized to balance positive and negative sample proportions during training. Additionally, the algorithm incorporates the deformable convolution (DCN) module and the SimAM attention mechanism to enhance model performance, particularly in terms of foreign object recall and detection accuracy. Moreover, we optimized the network inference speed with the improved SPPCSPC_S-F module. Experimental results demonstrate that the improved DF-YOLO network achieves a 2.04% increase in mAP@.5 compared to the original YOLOv7-Tiny network. The recall rate also improves from 89% to 91.51%. Additionally, the inference speed of the network rises from 130 to 140 FPS, which enhances detection effectiveness and reduces the frequency of transmission line leaks triggered by foreign object incursion.

INDEX TERMS DF-YOLO, focal-DIoU, SPPCSPC_S-F, DCN, SimAM.

I. INTRODUCTION

Electricity serves as the lifeblood of both modern civilization and the national economy. As the smart grid construction progresses, the scale of transmission lines-vital conduits for electric energy-continues to expand. This expansion poses significant challenges to ensuring the power grid's secure operation [1], [2], [3]. Prompt detection and removal of foreign substance intrusions in transmission lines are imperative to mitigate potential hazards like short-circuit discharges, fault challenges, and other risks to the power grid's operational safety [4].

Traditional manual inspection labor intensity, low efficiency, and inspection results are difficult to digitize does not meet the development direction of the smart grid [5]. In recent years, with the improvement of GPU arithmetic, drone inspection has become a hot spot in the industry [6], [7], [8], but how to make full use of the massive image and video data, relying on manual defect finding is not ideal [9], [10], [11].

The rapid advancement of deep learning in object detection has led to the practical implementation of related technologies on the ground. Object detection algorithms can be broadly categorized into two main types: two-stage networks that utilize candidate frames, and one-stage networks that rely on regression [12]. Representative two-stage networks mainly include R-CNN [13], Faster R-CNN [14] and Mask-RCNN [15]. Liang et al. [16], [17] introduced a transmission line foreign object detection approach based on Faster R-CNN. It's noted that the current datasets exhibits limitations in terms of limited categories and insufficient sample size. Zhang et al. [18] introduced a novel transmission line foreign object detection model, RCNN4SPTL. This approach enhances both detection accuracy and speed in

The associate editor coordinating the review of this manuscript and approving it for publication was Andrea F. Abate.

comparison to Faster R-CNN. Although the aforementioned two-stage network achieves elevated detection accuracy and a diminished leakage rate, its processing speed is sluggish, rendering it unsuitable for real-time detection scenarios.

Representative one-stage networks mainly include SSD [19], YOLOv3 [20], YOLOv5 [21], YOLOX [22] and YOLOv7 [23]. Chen et al. [24] introduced a transmission line foreign object detection method based on the YOLOv3 model, and proposed a new attention module E-CBAM to make the model accuracy reach 88.79%. Li et al. [25] replaced Darknet-53, the backbone of the original YOLOv3 network, with Mobilenetv2, and reduced the network parameters using depth-separable convolution, and the accuracy of the network reached 83.2%, but the leakage rate of the network was larger for some foreign objects with strange morphology. Song et al. [26], [27] introduced a high-voltage line foreign object intrusion detection system based on YOLOv4. They incorporated k-means clustering and the DIoU NMS method to enhance YOLOv4's performance, resulting in an 8% average improvement in network accuracy, reaching 81%. Huang et al. [28], [29] introduced a transmission line foreign object detection algorithm based on YOLOv5s using Ghost lightweight convolution module and KL scatter distribution loss function, and finally deployed the model into the edge device, the model accuracy only reached 84.75%. Liu et al. [30] presented a transmission line foreign object detection model founded on the YOLOX network. Their approach amplifies detection accuracy by incorporating the attention mechanism and ASPP module. This strengthens the model's feature extraction capabilities and enhances sensitivity to foreign objects across diverse scales. However, there is a high rate of missed targets for the trash category. Yu et al. [31] introduced a transmission line foreign object detection model based on YOLOv7. They employed hyper-parameter optimization and depth-separable convolutional SPD to enhance the accuracy of detecting small targets. Additionally, they replaced the original CIoU loss with DIoU loss. As a result, the final average accuracy of the network achieved 92.2%. But the detection speed is too slow. Both the aforementioned one-stage and two-stage networks exhibit superior performance in detecting foreign objects on transmission lines. However, the one-stage network boasts a swifter detection speed and heightened efficiency, significantly curbing the utilization of computational resources and aligning with the demands of real-time detection scenarios.

Through the above literature research, it is known that YOLOv7, as a newer network architecture in the one-stage network, is more accurate and faster than object detection networks such as YOLOv5 under the same volume. However, due to the existence of positive and negative sample imbalance in the transmission line foreign object intrusion dataset, in which the shape of the rubbish category is irregular and the intra-class variance is large, resulting in a high model leakage rate and the real-time performance needs to be further improved, for this reason, this paper is based on the improvement of YOLOv7-Tiny, and puts forward a transmission line foreign object detection model, DF-YOLO, with the main contributions as follows:

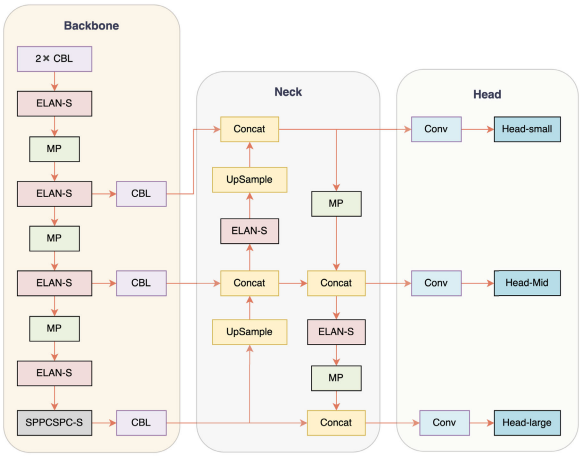


FIGURE 1. YOLOv7-Tiny network structure.

1) In the underlying structure of the backbone network, deformable convolutional DCN is used to replace some of the standard convolutional layers to solve the problem of large intraclass variance in the dataset, and to increase the capability of feature extraction of invasive foreign objects with large intraclass variance.

2) Drawing on the improvement idea of the SPPF module, the SPPCSPC_S module of YOLOv7-Tiny is improved to the SPPCSPC_S-F module, the parallel structure of the pooling part is changed to serial structure, and the size of the pooling kernel is uniformly set to 5, which retains the detail information in the feature map, and improves the inference speed of the network while ensuring that the accuracy is not affected.

3) Added SimAM Parameter-free Attention to select and aggregate informative features by maximizing the information entropy in the feature graph. It improves the expressiveness and performance of the network while keeping the number of network parameters constant.

4) Based on the notion that the Focal Loss loss function addresses the imbalance between positive and negative samples during training through a penalty term, an enhanced Focal-DIoU loss is devised to replace the original CIoU loss in YOLOv7-Tiny. This modification effectively resolves the issue of sample imbalance during training while concurrently reducing computational complexity.

In Chapter 2, we present the original model of the network as well as the proposed method and implementation. Chapter 3 elucidates a comparative analysis between the proposed approach and the original method. Chapter 4 provides a systematic summary of this research and concludes.

II. MODELS AND METHODS

The YOLOv7-Tiny is a lightweight version of the YOLOv7 and consists of three main parts: the backbone network, the neck structure, and the detection head section, as shown in Fig 1.

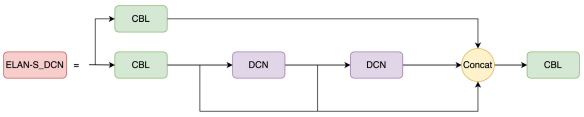


FIGURE 2. ELAN_S-DCN structure.

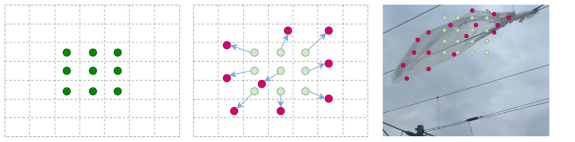


FIGURE 3. Deformable convolutional feature sampling method.

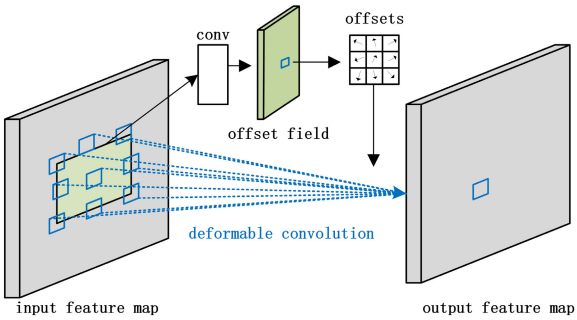


FIGURE 4. Deformable convolution module.

The 2 × CBL in Backbone denotes two normalized convolutions; ELAN-S as the main module for backbone feature extraction is a lightweight version of the ELAN module in the YOLOv7 model, replacing the SiLU activation function with Leaky ReLU and cropping the six layers of convolutions therein to four; the MP module constitutes a dual-channel architecture comprising of maxpooling and convolutions, which allows for the adjustment of the number of channels in the network; SPPCSPC_S is also a lightweight version of the SPPCSPC module in the YOLOv7 network, which utilizes the spatial pyramid structure to achieve multi-scale sensory fields and better capture feature information at different scales in the input feature map. The Neck part uses down-up and up-down bi-directional feature fusion and consists of the ELAN-S module and the UpSample module as well as the Concat module, which is also a kind of feature pyramid structure that achieves a multi-scale object detection effect. In this paper, the ELAN-S module and the SPPCSPC-S module of the YOLOv7-Tiny network are improved in four aspects concerning the characteristics of the transmission line foreign object intrusion dataset.

A. ELAN-S_DCN STRUCTURE

Geometric variations caused by factors such as scale and attitude have been the main challenge for object detection, and in the transmission line dataset the rubbish class has different shapes of attitude and variations, which makes it difficult for the network to learn its features. To solve this problem ELAN-S_DCN structure is designed and deformable convolutional DCN is introduced in the ELAN-S module in the backbone network as shown in Fig 2.

In this case, the second and third layers of convolutional operations CBL are changed to deformable convolutional (DCN). DCN can get the ability of geometric deformation modeling by learning the position of sampling points for better extraction of rubbish category features with irregular shapes and large intra-class variance as shown in Fig 3.

In this case, the sample points for the standard convolution operation are shown in Fig 3(a). The objective of achieving deformable convolution is not to render the convolution kernel deformable itself, but rather to introduce a trainable offset to the sampling points of the convolution kernel. The convolution sampling points with offset are illustrated in Fig 3(b, c).

For a standard convolutional kernel $\mathcal{R}$ the receptive field size and dilation rate are defined. For example, a 3 × 3 convolutional kernel $\mathcal{R} = (-1,-1),(-1,0), \ldots,(0,1),(1,1)$. For the standard convolution R, the P_0 output feature map y on the input feature map x is defined as shown in Eq 1.

$$y(p_o) = \sum_{p_n \in \mathcal{R}} w(p_n) \cdot x(p_0 + p_n) \tag{1}$$

where P_0 is a point on the feature map, P_n is a point in $\mathcal{R}$, and W is the feature point weight.

The DCN is added to the sampling process with an offset Δp_n, $\{\Delta p_n \mid n = 1,2, \ldots,N\}$, $N = |R|$. Thus, DCN can be defined as shown in Eq 2.

$$y(p_o) = \sum_{p_n \in \mathcal{R}} w(p_n) \cdot x(p_0 + p_n + \Delta p_n) \tag{2}$$

The process of extracting features by DCN is shown in Fig 4, where offsets are obtained by adding a convolution layer to the input feature layer, and the convolution kernel and offsets used to generate the output features are learned simultaneously during the training process. The information about the sampled points after the offset is implemented by bilinear interpolation. By introducing DCN, the network can better learn and extract feature information of irregular objects.

B. SPPCSPC_S-F MODULE

SPP(Spatial Pyramid Pooling) module can effectively avoid the distortion problem after image cropping and scaling and solves the problem of repeated extraction of features by convolutional neural networks [32], which has attracted much attention and several improvements since it was proposed in 2015. SPPF is improved from the SPP module by replacing the original CBL (Conv+BN+Leaky ReLU) module with the CBS (Conv+BN+SiLU) module and changing the original Maxpools structure from a parallel structure to a series structure, which improves the speed of the network and reduces the computation of the model [33]. The SPPCSPC_S-F module changes the three parallel Maxpools similar to the inception block in SPPCSPC_S to a series structure similar to the Resnet block, so that the deeper

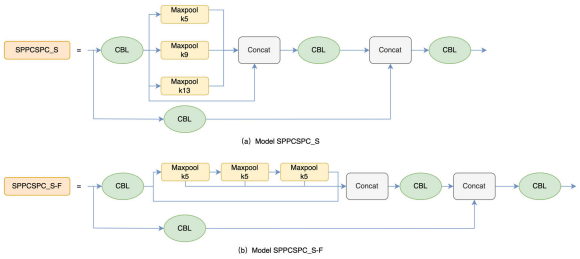


FIGURE 5. Comparison of SPPCSPC_S and SPPCSPC_S-F structures.

network is conducive to improving the feature representation of the rubbish category (morphologically variable); and sets the size of the pooling kernel to 5, which retains more detailed information in the feature map. The module obtains a speedup while keeping the sensory field unchanged, as shown in Fig 5.

C. FOCAL DIoU LOSS FUNCTION

The calculation of the CIoU loss function used in the original network is relatively complex, as shown in Eq 3, the CIoU loss adds the calculation of the bounding box aspect ratio compared to the DIoU loss, however, thousands of bounding boxes will be generated during the training process, which leads to a large computational overhead during the calculation. Therefore, we used the DIoU loss function instead of the CIoU loss function as shown in Eqs 4–5 to reduce the computational overhead of the loss process.

$$L_{CIoU} = 1 - IoU + \left(\frac{\rho^2\left(b, b^{gt}\right)}{c^2}\right) + \alpha\upsilon \tag{3}$$

where b,b^{gt} denotes the centroid of the prediction frame and ground truth, respectively, ρ^2 signifies the Euclidean distance, while c denotes the diagonal distance of the bounding region capable of containing both the real and prediction frames. Additionally, α serves as the weight parameter, and υ is utilized to quantify the similarity of aspect ratios.

$$DIoU = IoU - \frac{\rho^2\left(b, b^{gt}\right)}{c^2} \tag{4}$$

$$L\_DIoU = 1 - DIoU \tag{5}$$

During model training, the amount of negative samples generated significantly surpasses the quantity of positive samples, resulting in an imbalance in the ratio of positive and negative samples, and a large number of negative samples are involved in the calculation of the model loss during the backpropagation process, which affects the decrease in the value of the loss function. Based on this, we combine Focal Loss with DIoU loss. As shown in Eq 6, Focal Loss adds the adjustment factor $(1 - p_t)^\gamma$ balances the proportion of positive and negative samples in the loss calculation process. Introducing this method into the DIoU loss becomes the Focal DIoU loss, as shown in Eq 7, adding a penalty term, IoU^γ, to the DIoU loss function. As the confidence level of the sample gets larger both the IoU value approaches 1, the penalty term is infinitely close to 1, and the Focal DIoU loss becomes a DIoU loss that does not affect the positive sample. When the confidence level of the sample is very small both IoU value is close to 0, which means that the sample is a negative sample of the possibility is extremely high, the penalty term is infinitely close to 0, then the value of the Focal DIoU is infinitely close to 0, which solves the problem of low-quality samples caused by the loss of the value of the violent oscillation of the problem of balancing the proportion of positive and negative samples.

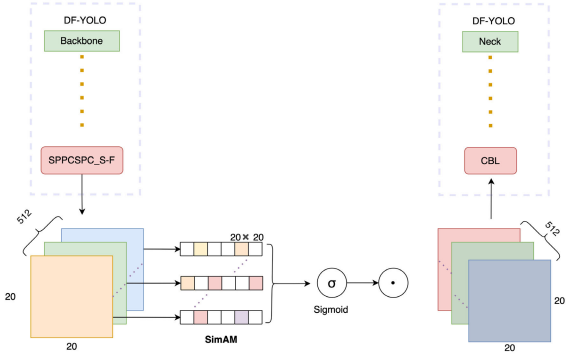


FIGURE 6. Structure of the SimAM model.

$$L_{Focal} = -\alpha_t\left(1 - p_t\right)^\gamma log\left(p_t\right) \tag{6}$$

where p_t denotes the probability of correct classification and the modulation factor γ is used to reduce the weight of easily classified samples and increase the weight of difficult-to-classify samples, taking the value of $\gamma = 0.5$.

$$L_{FocalDIoU} = IoU^\gamma L_{DIoU} \tag{7}$$

D. SimAM PARAMETER-FREE ATTENTION

In the field of neurology, neurons that are abundant in information frequently exhibit distinct firing patterns compared to the neighboring neurons, and they also exert inhibitory effects on the surrounding neural population. Therefore, we should assign higher weights to information-rich neurons. The core idea of SimAM (Similarity Attention Module) is to compute the similarity between different positions in the input feature map to generate a vector of attentional weights for the corresponding positions, i.e., to evaluate the importance of each neuron. Compared to traditional attention such as the SE channel attention and the CBAM spatial attention, the SimAM model does not introduce additional parameters and has a faster computational speed. The module structure is shown in Fig 6.

E. IMPROVED NETWORK MODEL

In this paper, we mainly improve the backbone network of the original YOLOv7-Tiny model by adopting DCN convolution in the last two ELAN_S modules, adding the SPPCSPC_S-F module and SimAM parameter-free attention module in the last layer, and proposing the Focal DIoU loss function. The improved overall network structure is shown in Fig 7.

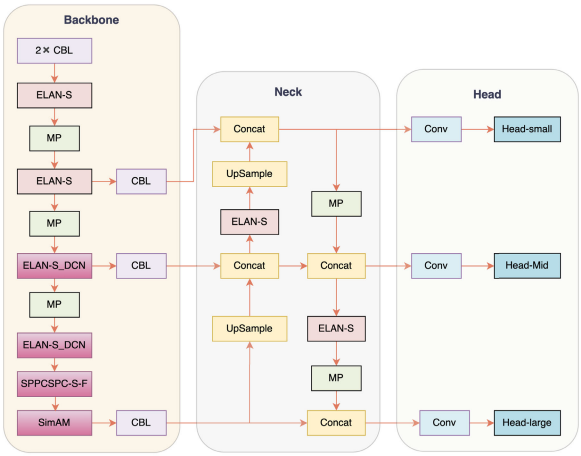


FIGURE 7. Structure of the DF-YOLO network.

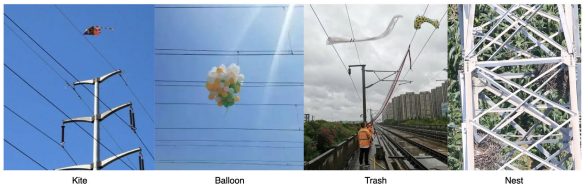


FIGURE 8. Sample dataset.

III. EXPERIMENTS AND ANALYSES

A. DATA ENHANCEMENT

Given that there is no publicly available dataset on transmission line foreign bodies and the complex environment surrounding transmission lines, a total of 1,942 images were obtained through manual collection and online collection methods. As shown in Fig 8, the acquired dataset contains a total of four categories: bird nests, balloons, kites, and trash. The dataset contains 1,325 labels in the birds nest category, 344 labels in the kite category, 240 labels in the balloon category, and 67 labels in the trash category. However, the number of labels for the three categories of kites, balloons, and trash is so small that the model can only be trained with a large amount of data to fully learn the features of the target to be detected. So, we augmented the data for these three types of targets using cropping, flipping, rotating, greying out with 10% probability, randomly increasing or decreasing hue, saturation, exposure, and brightness with a 10% probability, and mosaic enhancement. The final bird's nest, kite and trash categories were enhanced to 1,480, 935, and 766 labels, respectively. The overall dataset reached 2,739 sheets.

The implementation of data augmentation serves to mitigate the occurrence of model overfitting during the training phase, subsequently enhancing the generalization capacity and robustness of the model. By disrupting the order of the dataset, 70% of the dataset is selected as the training set and 30% of the dataset is selected as the validation set.

B. EVALUATION METRICS

mean Average Precision (mAP), Recall(R) and Precision(P) are uniformly used as evaluation metrics during model training and validation. For target detection, mAP is usually an important measure of model performance. The mAP is subdivided into mAP@.5 and mAP@.5-.95, mAP@.5 is the average of the APs for each category computed when the IoU is set to 0.5. mAP@.5-.95 is the average of the APs for each category computed when the IoU threshold is set from 0.5 to 0.95 with a step of 0.05. Assuming that there are K categories and $K > 1$, the mAP is calculated as shown in Eq 8 and Eq 9.

$$AP = \sum_{i=1}^{n-1} (r_{i+1} - r_i)\, P_{interp}\,(r_i + 1) \tag{8}$$

where $r_1, r_2, \ldots, r_n$ is the Recall value corresponding to the first interpolation at the first interpolation of the Precision interpolation segment in ascending order.

$$mAP = \frac{\sum_{i=1}^{K} AP_i}{K} \tag{9}$$

Although the recall rate is an important evaluation index for anomaly detection, in transmission line foreign object detection, once the leakage occurs it will cause serious safety accidents, so the recall rate is equally important for transmission line foreign object detection. Recall is the proportion of all positive samples that are correctly predicted, i.e., the coverage of correct predictions. Precision is the proportion of all positive results that are truly correct. The formulas for recall and precision are shown in Eq 10 and Eq 11.

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

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

TABLE 1. Experimental data of the DCN module.

Module	Class	Precision	mAP@.5
YOLOv7-Tiny	Trash	0.922	0.918
YOLOv7-Tiny+DCN	Trash	0.941	0.924

TABLE 2. SPP, SPPF, SPPCSPC_S and SPPCSPC_S-F model inference speed test.

Modules	parameters	Inference time (ms)
SPP	394240	82.30
SPPF	394240	43.46
SPPCSPC_S	657408	136.50
SPPCSPC_S-F (ours)	657408	99.70

C. COMPARATIVE EXPERIMENTS AND ANALYSES

To verify the effectiveness of the DCN, SPPCSPC_S-F and SimAM modules, the following comparison experiments were done for these three modules respectively.

As depicted in Table 1, significant enhancements in accuracy are witnessed for the "trash" category, known for its pronounced morphological feature variations, upon the inclusion of the DCN module. This augmentation elevates the accuracy of the "trash" category from 92.2% to 94.1%, marking an improvement of 1.9 percentage points.

TABLE 3. SimAM, CBAM, SENet Module Comparison Tests.

Modules	parameters	mAP@.5
YOLOv7-Tiny	6023106	0.9314
YOLOv7-Tiny+SimAM (ours)	6023106	0.9349
YOLOv7-Tiny+CBAM	6031396	0.9303
YOLOv7-Tiny+SENet	6031297	0.9291

TABLE 4. Comparison of different model detection experiments for transmission line foreign body datasets.

Module	Input size	Precision	Recall	mAP@.5	mAP@.5-.95
YOLOv3	640*640	0.4150	0.7430	0.6070	0.2750
YOLOv5-s	640*640	0.9720	0.7510	0.8150	0.5410
YOLOv7-Tiny	640*640	0.9410	0.8900	0.9314	0.5630
YOLOv8-s	640*640	0.9600	0.7820	0.8190	0.5470
Ours	640*640	0.9584	0.9151	0.9518	0.5761

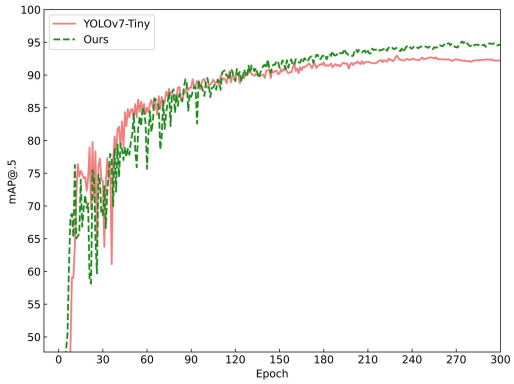


FIGURE 9. Comparison of mAP value curves for DF-YOLO and YOLOv7-Tiny models.

Moreover, mAP@.5 sees a marginal advancement of 0.06%. Within Table 2, four modules (SPP, SPPF, SPPCSPC_S, and SPPCSPC_S-F) undergo assessment. Notably, the SPPCSPC_S-F model showcases an inference time 36.8ms shorter than the SPPCSPC_S module in YOLOv7-Tiny. Meanwhile, Table 3 reveals that the addition of the SimAM attention mechanism has no impact on the number of network model parameters yet enhances mAP@.5 by 0.35%. Conversely, the integration of CBAM and SENet attention mechanisms leads to parameter increases of 8290 and 8191, respectively. However, these additions result in marginal mAP@0.5 reductions of 0.11% and 0.23%.

To validate the efficacy of DF-YOLO, the enhanced model was subjected to a comparative evaluation against the YOLO series algorithm. To ensure the credibility and consistency of the experimental findings, the identical dataset containing transmission line foreign objects was utilized for training and validating both models. The pre-training weights on the COCO dataset were applied to all models using the migration learning method, optimized using the SGD optimizer, the network batch size was uniformly set to 32, the network training epoch was set to 300, the learning rate was set to 0.01, and the input image size for all the models was a 640 × 640 RGB image. The same validation set is used for validation after each round of training. The experimental results are shown in Table 4.

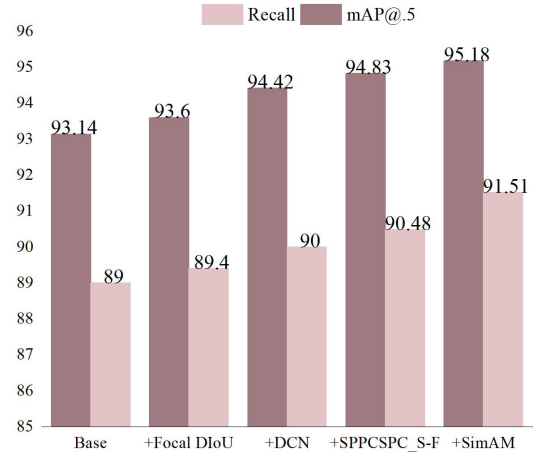
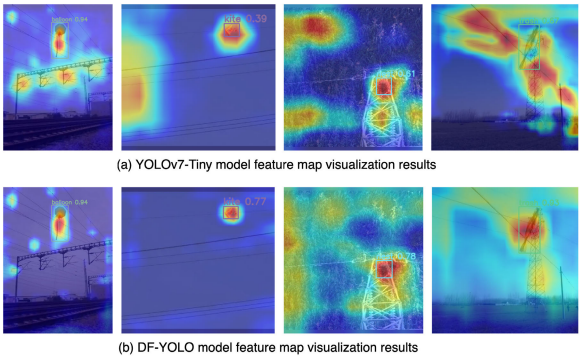


FIGURE 10. Statistical plots of mAP@.5 and Recall data for the DF-YOLO model.

TABLE 5. DF-YOLO model detection results for each category.

Class	Precision	Recall	mAP@.5
Balloon	0.950	0.919	0.953
Kite	0.953	0.900	0.954
Nest	0.945	0.935	0.952
Trash	0.938	0.906	0.948



(a) YOLOv7-Tiny model feature map visualization results

(b) DF-YOLO model feature map visualization results

FIGURE 11. Comparison of feature visualization before and after model improvement.

As can be seen from Table 4, the proposed model outperforms the other models in terms of mAP and recall. The mAP@.5 reaches 0.9518, which is 2.04 percentage points higher than the original network. In terms of recall, our network is also higher than all the other models, with a 2.51% increase in recall compared to the original network and possesses a much lower miss detection rate. The overall mAP@.5 plot for all categories is shown in Fig 9.

The dashed line indicates the mAP@.5 change curve of the improved network, and the solid line indicates the mAP@.5 curve of YOLOv7-Tiny. From the figure, it can be seen that with the increasing number of training rounds, the mAP@.5 curves of the models between 200-300 epochs all tend to be stable, indicating that the models have all converged, and the average accuracy of the improved model is significantly higher than that of the original YOLOv7-Tiny model.

From the results of the above comparison experiments, it can be seen that our proposed method has a better detection effect, the recognition accuracy, recognition effect

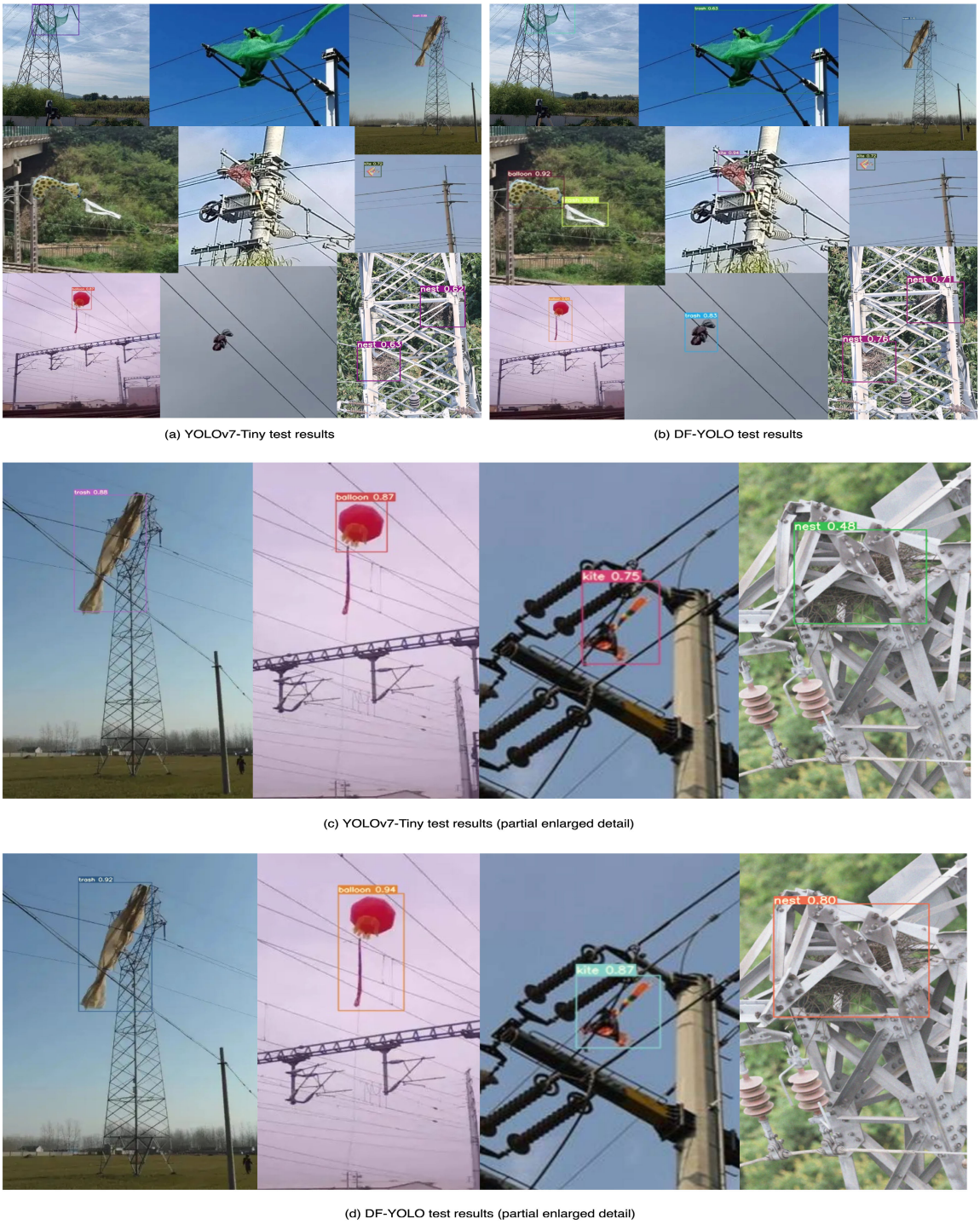


FIGURE 12. DF-YOLO and YOLOv7-Tiny model detection results.

and detection speed have been improved, and it can well meet the needs of transmission line foreign object detection.

D. ABLATION STUDY

Ablation experiments were conducted to verify the validity and optimization of each improvement point in the DF-YOLO model. The experimental results are shown in Fig 10.

Due to the concentration of transmission line foreign object intrusion detection data, there are very few intruding foreign objects within the sample image, which is 1. A large number of negative sample anchors will be generated during the training process, resulting in an imbalance of positive and negative samples, which makes the negative samples dominate in the loss function and affects the model detection accuracy. The change of the original network's loss function

to our proposed Focal DIoU balances the proportion of positive and negative samples in the loss function, resulting in a 0.46% improvement in the model's mAP@.5, and the recall rate also improved by 0.4%. The ELAN-S_DCN module, which replaces some of the convolutional operations with DCN, enables the network to fully learn the features for samples with too large a gap within the class by adding learnable offsets to the sampling points during training, and the mAP@.5 and recall were also boosted by 0.82% and 0.6% as a result. Our proposed SPPCSPC_S-F module further increases the mAP@.5 by 0.41% and the recall by 0.48% while keeping the number of sensing fields and model parameters constant. The network mAP@.5 and recall continues to increase by 0.35% and 1.03% with the addition of the SimAM parameter-free attention module at the last layer of the backbone.

Although target detection usually uses mAP@.5 as an evaluation index, however, once a foreign object is missed it may lead to phenomena such as partial discharge and fault tripping, and the recall rate, as an evaluation index for defect detection, is also very important for this task. The recall of the improved model increases from 89% to 91.51%, which is an improvement of 2.51 percentage points.

In conclusion, the improved network model DF-YOLO achieves a mAP@.5 of 95.18% compared to the original network model YOLOv7-Tiny, an improvement of 2.02%. The inference speed of the model is also improved from 130 FPS to 140 FPS, and the detection effect of the improved model for each category is shown in Table 5.

The data in Table 5 demonstrates that the enhanced model achieves a detection precision, recall, and mAP@.5 exceeding 90% for foreign objects on transmission lines. Particularly, for garbage-like foreign objects with varying shapes and postures, the mAP@.5 and Recall attain 94.8% and 90.6%, respectively. This significantly mitigates the risk of foreign object leakage in transmission lines.

For a comprehensive comparison between the DF-YOLO and YOLOv7-Tiny models, as illustrated in Fig 11, visualizations of their respective feature maps are presented. Observing these visualizations, it becomes evident that during feature extraction, the DF-YOLO model places heightened emphasis on the target itself. In contrast, the original model directs its attention not solely to the foreground of the feature maps, but also to their background, thereby impacting the model's accuracy.

Fig 12 displays the detection results of both the YOLOv7-Tiny model and the improved model. The comparison of parts (a) and (b) in the figure reveals that the DF-YOLO model successfully detects targets in the kite, balloon, and trash categories, which were previously missed by YOLOv7-Tiny due to their complex and low-quality backgrounds. This detection improvement reduces the leakage rate, and the confidence level of the detection results surpasses that of the original model. In the case of foreign objects such as bird's nests, both YOLOv7-Tiny and the improved model exhibit nearly identical target localization results for the bird's nest category. However, the DF-YOLO model displays higher confidence in identifying them as bird's nests compared to the original model, as depicted in Fig 12(c) and (d).

IV. CONCLUSION

Aiming at transmission line safety, a transmission line foreign object dataset was constructed according to the standard of the COCO dataset, with a total of 2739 RGB images. An improved network model based on YOLOv7-Tiny is proposed to improve the detection performance of the model. Deformable convolution is introduced into the backbone network of the model, which improves the model's learning of the target feature information and solves the problem of large intra-class gaps in some categories in the dataset. The introduction of the SimAM parameter-free attention module improves the expressiveness and performance of the network while ensuring that the speed of the model does not decrease. A new spatial pyramid structure, SPPCSPC_S-F, is proposed to improve the detection speed of the network while the number of parameters remains constant. Drawing on the idea of using a penalty term in Focal Loss to balance the proportion of positive and negative samples in the training process, the Focal DIoU loss function is designed to balance the proportion of positive and negative samples in the training process, and to solve the problem of oscillating loss values caused by low-quality samples. The final network achieved a mAP@.5 of 95.18% and a single image inference speed of 140 FPS. Compared to the YOLOv7-Tiny model, the DF-YOLO has improved in terms of precision, recall, and speed. The mAP@.5 is improved by 2.02%, the recall is improved by 2.51%, the leakage problem of the model is improved, and the inference speed of a single image is improved by 10 FPS. In conclusion, the proposed model helps to improve the efficiency of the transmission line inspection process, which is of great significance in the inspection task of the environmental conditions of transmission lines.

REFERENCES

[1] C. Su and Q. Yang, "Overhead transmission line fault cause discrimination based on multi-view sparse feature selection," *Smart Power*, vol. 51, no. 3, pp. 96–103, 2023.

[2] R. Huang, "Research on transmission line fault diagnosis strategy based on neural network," *Electr. Ind.*, vol. 269, no. 4, pp. 27–30, 2023.

[3] P. Liu, Y. Zhang, K. Zhang, P. Zhang, and M. Li, "An improved YOLOv3 algorithm and intruder detection on transmission line," in *Proc. China Autom. Congr. (CAC)*, Nov. 2022, pp. 5736–5741.

[4] Y. Luo, "Evaluation model of the effectiveness of the prevention and control of bird damage faults on overhead transmission lines," *Inf. Technol.*, vol. 47, no. 3, pp. 133–138, 2023.

[5] M. Wu, L. Guo, R. Chen, W. Du, J. Wang, M. Liu, X. Kong, and J. Tang, "Improved YOLOX foreign object detection algorithm for transmission lines," *Wireless Commun. Mobile Comput.*, vol. 2022, Oct. 2022, Art. no. 5835693.

[6] L. Shi, Y. Chen, G. Fang, K. Chen, and H. Zhang, "Comprehensive identification method of bird's nest on transmission line," *Energy Rep.*, vol. 8, pp. 742–753, Sep. 2022.

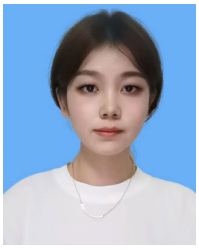
[7] R. Kang, C. Ma, and B. Li, "Real-time monitoring method for uninterrupted operation of transmission line drone automatic inspection system," *Comput. Technol. Autom.*, vol. 42, no. 1, pp. 97–102, 2023.

[8] O. M. Butt, M. Zulqarnain, and T. M. Butt, "Recent advancement in smart grid technology: Future prospects in the electrical power network," *Ain Shams Eng. J.*, vol. 12, no. 1, pp. 687–695, Mar. 2021.
[9] Y. Zhang, A. Wang, and H. Zhang, "Overview of smart grid development in China," *Power Syst. Protection Control*, vol. 49, no. 5, pp. 180–187, 2021.
[10] L. Guo, C. Ye, Y. Ding, and P. Wang, "Allocation of centrally switched fault current limiters enabled by 5G in transmission system," *IEEE Trans. Power Del.*, vol. 36, no. 5, pp. 3231–3241, Oct. 2021.
[11] Z. Yang, X. Xu, K. Wang, X. Li, and C. Ma, "Multitarget detection of transmission lines based on DANet and YOLOv4," *Sci. Program.*, vol. 2021, Dec. 2021, Art. no. 6235452.
[12] Y. Zhang, D. Huang, W. Wang, and J. He, "A review of deep learning based target detection algorithms research and application," *Comput. Eng. Appl.*, vol. 59, no. 18, pp. 1–13, 2023.
[13] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.
[14] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.
[15] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.
[16] H. Liang, C. Zuo, and W. Wei, "Detection and evaluation method of transmission line defects based on deep learning," *IEEE Access*, vol. 8, pp. 38448–38458, 2020.
[17] F. Li, J. Xin, T. Chen, L. Xin, Z. Wei, Y. Li, Y. Zhang, H. Jin, Y. Tu, X. Zhou, and H. Liao, "An automatic detection method of bird's nest on transmission line tower based on faster_RCNN," *IEEE Access*, vol. 8, pp. 164214–164221, 2020.
[18] W. Zhang, X. Liu, J. Yuan, L. Xu, H. Sun, J. Zhou, and X. Liu, "RCNN-based foreign object detection for securing power transmission lines (RCNN4SPTL)," *Proc. Comput. Sci.*, vol. 147, pp. 331–337, Jan. 2019.
[19] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot MultiBox detector," in *Proc. 14th Eur. Conf.*, 2016, pp. 21–37.
[20] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.
[21] X. Zhu, S. Lyu, X. Wang, and Q. Zhao, "TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on dronecaptured scenarios," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 2778–2788.
[22] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO series in 2021," 2021, *arXiv:2107.08430*.
[23] C.-Y. Wang, A. Bochkovskiy, and H.-Y. Mark 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.
[24] Q. Li and G. Chen, "Improved transmission line fault detection with YOLOv3-SPP," *J. Shanghai Electr. Eng. Inst.*, vol. 25, no. 6, pp. 318–324, 2022.
[25] H. Li, L. Liu, J. Du, F. Jiang, F. Guo, Q. Hu, and L. Fan, "An improved YOLOv3 for foreign objects detection of transmission lines," *IEEE Access*, vol. 10, pp. 45620–45628, 2022.
[26] Y. Song, Z. Zhou, Q. Li, Y. Chen, P. Xiang, Q. Yu, L. Zhang, and Y. Lu, "Intrusion detection of foreign objects in high-voltage lines based on YOLOv4," in *Proc. 6th Int. Conf. Intell. Comput. Signal Process. (ICSP)*, Apr. 2021, pp. 1295–1300.
[27] Z. Hui, Z. Jian, C. Yuran, J. Su, W. Di, and D. Hao, "Intelligent bird's nest hazard detection of transmission line based on RetinaNet model," *J. Phys., Conf. Ser.*, vol. 2005, no. 1, Aug. 2021, Art. no. 012235.
[28] Y. Huang, Z. Chen, and Chen, "Real-time transmission line fault detection based on edge computing and improved yolov5s algorithm," *Power Construct.*, vol. 44, no. 1, pp. 91–99, 2023.
[29] H. Li, Y. Dong, Y. Liu, and J. Ai, "Design and implementation of UAVs for bird's nest inspection on transmission lines based on deep learning," *Drones*, vol. 6, no. 9, p. 252, Sep. 2022.
[30] B. Liu, J. Huang, S. Lin, Y. Yang, and Y. Qi, "Improved YOLOX-S abnormal condition detection for power transmission line corridors," in *Proc. IEEE 3rd Int. Conf. Power Data Sci. (ICPDS)*, Dec. 2021, pp. 13–16.
[31] C. Yu, Y. Liu, W. Zhang, X. Zhang, Y. Zhang, and X. Jiang, "Foreign objects identification of transmission line based on improved YOLOv7," *IEEE Access*, vol. 11, pp. 51997–52008, 2023.
[32] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.
[33] H. Tang et al., "A visual defect detection for optics lens based on the YOLOv5-C3CA-SPPF network model," *Opt. Express*, vol. 31, no. 2, pp. 2628–2643, 2023.

SHAO JIA LI was born in Shaanxi, China, in 1998. He received the bachelor's degree in computer science and technology from the Computer Science and Technology College, Xianyang Normal University, in 2021. He is currently pursuing the Graduate degree with the School of Urban Rail Transit and Logistics, Beijing Union University. His research interests include image recognition and deep learning.

YAN XIA LIU JR. (Member, IEEE) received the Ph.D. degree from the School of Automation and Electrical Engineering, University of Science and Technology Beijing, in 2013. She is currently a Professor with the College of Urban Rail Transit and Logistics, Beijing Union University. Her current research interests include pattern recognition, computer vision, deep learning, and intelligent instruments.

MIAO LI was born in Shandong, China, in 1998. She received the bachelor's degree in computer science and technology from the College of Information Science and Engineering, Shandong Agricultural University, in 2022. She is currently pursuing the Graduate degree with the School of Urban Rail Transit and Logistics, Beijing Union University. Her research interests include image recognition and object detection.



LU DING was born in Gansu, China, in 1999. She received the bachelor's degree in communication engineering from the Smart City College, Beijing Union University, in 2021, where she is currently pursuing the Graduate degree with the School of Urban Rail Transit and Logistics. Her research interests include image recognition and deep learning.