YOLO-T: Multi-Target Detection Algorithm for Transmission Lines

Shengwen Li, Huabing Ouyang, Tian Chen, Xiaokang Lu, Zhendong Zhao
School of Mechanical Engineering, Shanghai Dianji University, Shanghai 201306, China

***Abstract*—During UAV inspections of transmission lines, inspectors often encounter long distance and obstructed targets. However, existing detection algorithms tend to be less accurate when trying to detect these targets. Existing algorithms perform inadequately in handling long-distance and occluded targets, lacking effective detection capabilities for small objects and complex backgrounds. Therefore, we propose an improved YOLOv8-based YOLO-T algorithm for detecting multiple targets on transmission lines, optimized using transfer learning. Firstly, the model is lightweight while ensuring detection accuracy by replacing the original convolution block in the C2f module of the neck network with Ghost convolution. Secondly, to improve the target detection ability of the model, the C2f module in the backbone network is replaced with the Contextual Transformer module. Then, the feature extraction of the model is improved by integrating the Attention module and the residual edge on the SPPF (Spatial Pyramid Pooling-Fast). Finally, we introduce a new shallow feature layer to enable multi-scale feature fusion, optimizing the model detection accuracy for small and obscured objects. Parameters and GFLOPs are conserved by using the Add operation instead of the Concat operation. The experiment reveals that the enhanced algorithm achieves a mean detection accuracy of 97.19% on the transmission line dataset, which is 2.03% higher than the baseline YOLOv8 algorithm. It can also effectively detect small and occluded targets at long distances with a high FPS (98.91 frames/s).**

***Keywords*—Transmission line inspection; contextual transformer; attention mechanism; ghost convolution**

I. Introduction

Given their role as the main channel for transmitting electricity from power plants to consumers, the importance of transmission lines cannot be overlooked. Conducting timely inspections to detect defects and other problems is crucial to improving line operation safety, extending service life, and reducing the incidence of accidents [1, 2]. With the construction of the smart grid in full swing, the length of overhead transmission lines is increasing. As the demands for intelligent transmission line inspections continue to increase, drones are increasingly replacing manual labor for these inspections [3–5]. However, a large number of images collected by UAV inspection are mainly inspected manually or using traditional image processing techniques, which are inefficient and have poor detection accuracy [6, 7]. With the development of Double-stage detection strategies represented by the R-CNN series [8–10] and Single-stage object detection methods represented by SSD [11] (Single Shot Multi-Box Detector) and YOLO [12–17] (You Only Look Once), the problems of traditional algorithms, such as slow speed and weak robustness, have been solved, providing a new approach for UAV inspection [18].

To achieve a lightweight network and facilitate model deployment on embedded platforms [19], Han et al. [20] proposed an improved Tiny-YOLOv4 for insulator aerial image detection and damage recognition. By incorporating ECA-Net into the multi-scale feature fusion layer, the complexity of YOLOv4 is simplified, balancing detection speed and accuracy. Qiu et al. [21] employed a lighter MobileNets network in place of CSPDarkNet53 in the YOLOv4 model and adjusted the width multiplier in the bi-directional Path Aggregation Network (PANet) for network lightweighting. Based on the YOLOv5 network, Li et al. [22] used BiFPN (bi-directional feature pyramid network) to replace the original PANet to improve feature fusion capability, and used DIoU to substitute the initial CIoU loss function. Through sparse regularization, the scaling factor is used to filter out unimportant channels and prune them. Subsequently, the model's detection accuracy is restored to its prepruning level through secondary training. Although these studies primarily focus on insulators and their defects, they overlook common issues such as the detection of small and occluded targets. Kang et al. [23] introduced an algorithm for detecting multiple defects in insulators by combining a weighted bidirectional feature pyramid (CAT-BiFPN) with an attention mechanism. Despite the model's performance on small targets is improved by constructing a new CAT-BiFPN, adding and subtracting new detection layers and adding a hybrid module of attention and convolution, the mAP is only 93.9%, which still has room for improvement. Moreover, the high parameter count of the model hinders its detection speed.

In response to these challenges, this paper presents a lightweight YOLO-T algorithm for multi-target detection on transmission lines. To enhance the precision and speed of algorithm detection, the following aspects need to be addressed:

1) To achieve lightweighting of the model and improve its detection efficiency, Ghost convolution should replace ordinary convolution in the C2f module of the neck network.

2) To enhance the backbone network's proficiency in extracting critical features, the following measures should be taken: replace the C2f module with the Contextual Transformer feature extraction module, introduce the SE attention module, optimize the structure of the SPPF, and implement other optimization measures.

3) To elevate the model's performance in identifying small and hidden targets, additional measures are required such as adding a new shallow feature layer for combining multi-scale features to boost the neck network and other relevant methods.

II. PROPOSED METHOD

In January 2023 Ultralytics released the YOLOv8 algorithm on GitHub [24], and its network structure follows the previous network structure of YOLOv5, which still includes four parts: Input, Backbone, Neck and Head. Specifically, the Input part mainly adopts Mosaic data enhancement, adaptive anchor frame computation and adaptive gray scale filling, etc. The Backbone part is the backbone network, which mainly composed of Conv, C2f, and SPPF modules. YOLOv8 initially replaced the C3 module with the C2f module. By using more branches connected across layers, the gradient flow of the model is enriched to form a neural network module with superior feature expression capability. The neck section aims to enhance the feature extraction network, and still adopts the FPN-PANet structure to strengthen the network's ability to blend features from varying scales. Head is used as the classifier and regressor of YOLOv8, which decouples the classification and detection processes separately. Meanwhile, the anchor-based approach is replaced by an anchor-free approach. These improvements and optimizations allow the YOLOv8 algorithm to show better performance and accuracy in identifying targets.

YOLOv8 divides the detection network into five versions: n, s, m, l and x according to the dimensions of the network's width and depth. The application in this paper requires a more lightweight model, so this experiment utilizes YOLOv8n as the base model. Although the YOLOv8 algorithm incorporates many new improvements and has made great progress in target detection, it has greatly increased the detection difficulty due to the small target of transmission line inspection, the complex background and the frequent problem of the detection target being obscured. Therefore, to enhance the detection precision and efficiency for long-distance small targets and obscured targets, this study proposes a YOLO-T transmission line multi-target detection model based on the improved YOLOv8, whose network structure is shown in Fig. 1.

As depicted in Fig. 1, in the section of the backbone network, the C2f module is substituted by the Contextual Transformer module for feature extraction purposes, and the SE Attention module is added after it. At the same time, the SPPF structure is optimized to construct the SPPF-C module, to augment the backbone network's feature extraction capabilities for key features. Within the neck network, C2f-G lightens the feature extraction module and replaces the Concat operation with the Add operation to achieve model lightness and enhance detection efficiency. Finally, the PANet-Z multi-scale feature fusion module is constructed by adding new shallow feature layers to enhance the model's capability to detect small-sized targets and objects obscured by occlusion.

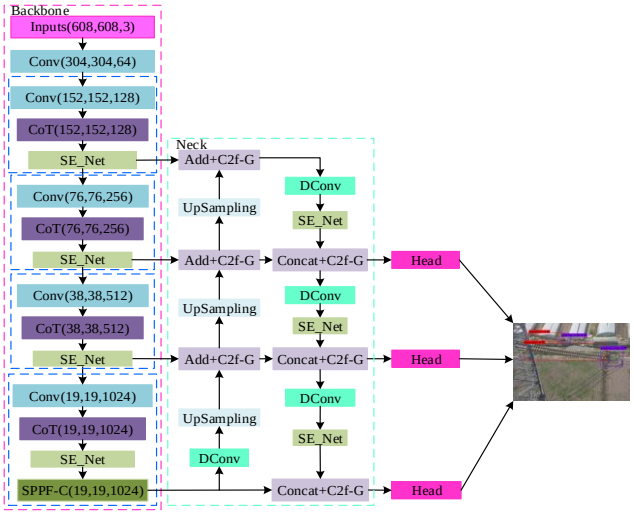


Fig. 1. Structure of the YOLO-T network.

A. C2f-G Lightweight Feature Extraction Module

Although YOLOv8n is a more lightweight model, for embedded devices, further lightweighting of the network is still required to achieve a higher detection speed. The C2f module, as an important part of the YOLOv8 network, can be optimized to achieve the lightweighting of the model. Therefore, the ordinary volume in the C2f module in the neck network is replaced with Ghost Convolution [25] (GConv), and the product constructs the C2f-G module to decrease the model's parameter count and computational overhead. Its structure is shown in Fig. 2.

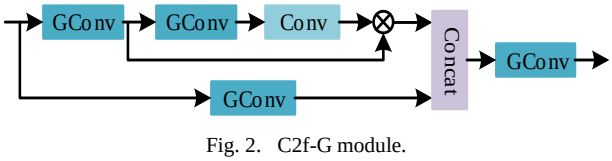


Fig. 2. C2f-G module.

Initially, ghost convolution employs a small set of convolution kernels to extract features from the input feature maps, then performs cheaper linear transformation operations on some of the extracted feature maps, and finally generates the final feature maps through splicing operations. By this method, the model's demand for computational resources is effectively reduced, while accurate feature extraction of the input feature maps is realized. The input feature layer is normally convolved to generate an $m \times H'' \times W''$ feature layer. Taking the Ghost convolution with $m \cdot (s-1) = C'(s-1)/s$ $d \times d$ linear kernels as an example, in order to complete a feature maps, each feature needs to be cheaply linearly transformed once to get s "phantom" feature maps, where $C' = m \times s$, $d \times d << D_K \times D_K$, $s << C$. The structure of ordinary convolution and Ghost convolution is shown in Fig. 3.

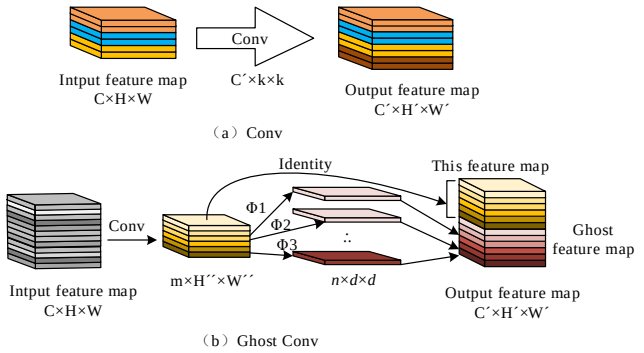


Fig. 3. Traditional Convolutional and Ghost Convolutional. Fig. 3(a) represents the traditional convolution module, while Fig. 3(b) represents the Ghost convolution module.

In Fig. 3, the input feature layer of convolution is $C \times H \times W$, the output feature layer is $C' \times H' \times W'$, and the convolution kernel size is $k \times k$, where C, H, W specify the input feature map's number of channels, height, and width, and C', H', W' denote the number of channels, height, and width of the output feature map, respectively. The formulas for the number of parameters and GFLOPs for Traditional Convolution and Ghost Convolution are shown in Eq. (1)–(4).

$$P_C = C \times C' \times k \times k \tag{1}$$

$$U_C = C \times k \times k \times C' \times H' \times W' \tag{2}$$

$$P_G = C \times m \times k \times k + m \times n \times d \times d \tag{3}$$

$$U_G = k \times k \times C \times H' \times W' \times m + H' \times W' \times m \times d \times d \times (s-1) \tag{4}$$

where, P_C , U_C are the parameters and the computational burden associated with traditional convolution. P_G , U_G are the number of parameters and GFLOPs amount of Ghost convolution, respectively. According to the above formula, the ratio of the number of parameters and GFLOPs amount of Ghost convolution to ordinary convolution can be calculated as:

$$\frac{P_G}{P_C} = \frac{m}{C'} + \frac{m \times n}{C \times C'} \frac{d \times d}{k \times k} \approx \frac{m}{C'} = \frac{1}{s} \tag{5}$$

$$\frac{U_G}{U_C} = \frac{1}{s} + \frac{s-1}{Cs} \cdot \frac{dd}{kk} \approx \frac{1}{s} \tag{6}$$

Here, it can be seen that the ratio of the quantity of parameters and GFLOPs obtained from the traditional convolution to the Ghost convolution is inversely proportional to the count of phantom feature maps, i.e., as the number of the Ghost feature maps increases, Ghost Convolution requires fewer parameters and GFLOPs than Traditional Convolution. The quantity of parameters and GFLOPs is minimized when traditional convolution is skipped and linear operations are directly used to generate the Ghost feature maps.

B. CoT Feature Extraction Module

Transformer structures with self-attention have sparked a revolution in the realm of natural language processing, and have achieved outstanding results in multiple computer vision tasks over the past few years. Nevertheless, the majority of current designs employ self-attention directly on 2D feature maps to acquire attention matrices from isolated query-key pairs at every spatial position, thereby missing the opportunity to leverage the abundant contextual information among adjacent key pairs. In contrast, the CoT (Contextual Transformer) [26] module is a novel Transformer style module that can efficiently address the aforementioned issue. The module structure of CoT is shown in Fig. 4.

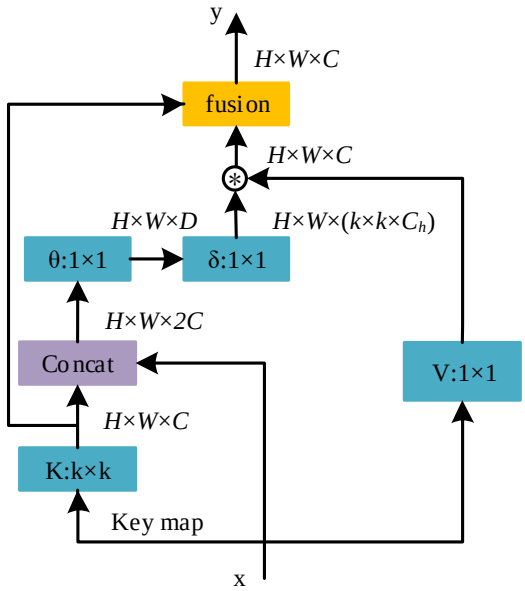


Fig. 4. Contextual Transformer module.

As shown in Fig. 4, the CoT module first encodes the input keys by contextualizing the convolution to obtain a static contextual representation of the input. The encoded keys are then further connected to the input query through two successive convolutions to learn the dynamic multi-head attention matrix. The learned attention matrix is then multiplied by the input values to realize the dynamic contextual representation of the input. Finally, the result of the fusion of static and dynamic contextual representations is used as the final output. In this design, the utilization of contextual information between input keys guides the learning process of the dynamic attention matrix, leading to enhanced visual representation. Therefore, in the backbone network, the C2f module is substituted with the CoT module.

C. SE Attention Module

Deepening of the network layers will lead to partial loss of texture information and contour information of small targets such as insulators at longer distances, which will lead to poor detection of the model. To address the above problems, this article adds a Squeeze-and-Excitation Networks (SE) [27] behind the effective feature layer of the backbone network and down-sampling of the neck network to highlight the key

characteristics and weaken irrelevant information, to improve the characterization capacity of the network and improve the model's ability to detect small targets at a long distance. The SE Attention Module is depicted in Fig. 5.

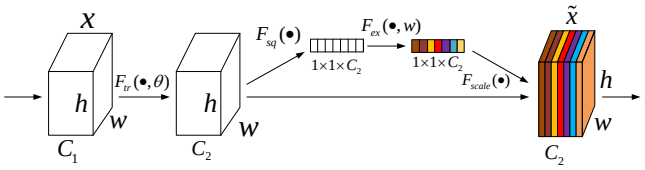


Fig. 5. SE attention module.

The importance of each channel is automatically learned by SE, and enhances the significant features and suppresses the non-significant features according to the importance. To confirm the effectiveness and superiority of the SE attention module added in this paper, the Efficient Channel Attention (ECA) [28] and Convolutional Block Attention Module (CBAM)[29] are inserted into the effective feature layer of the backbone network and behind the down-sampling of the neck network, respectively. Experiments were performed, and the results are displayed in Table I.

TABLE I. EXPERIMENTAL RESULTS OF DIFFERENT ATTENTION MECHANISMS

Attention Module	AP (%)				mAP (%)
	Insulator	*Defect*	*Nest*	*Grading*	
SE	93.97	98.15	95.69	97.20	96.25
ECA	94.21	97.73	95.22	96.72	95.97
CBAM	92.42	97.93	94.78	96.63	95.44

The results in Table I indicate that the inclusion of the SE attention module produces the optimal detection performance. Faced with the four types of detection targets studied in this paper, the AP values of the other three types of targets with the addition of the SE attention module are the highest except for the insulators, and the mAP value reaches 96.25%, which proves the effectiveness of the SE attention module.

D. SPPF-C Module

The SPPF module enhances the model's detection accuracy by applying pooling operations to feature maps of various scales, without altering their size. In this paper, to achieve an even greater level of detection accuracy with the model for the targets to be detected on transmission lines, we introduce a separate convolutional structure based on the SPPF structure and stack it with the results after the SPPF processing. The structure of the improved SPPF-C network is illustrated in Fig. 6.

E. PANet-Z Multiscale Feature Fusion Module

To maximize the use of shallow and deep semantic features, this article designs a PANet-Z feature fusion structure to improve the effectiveness of the model in detecting small and occluded targets at long range. The original PANet structure achieves channel information fusion by stacking two feature maps using the Concat operation to obtain rich semantic features. However, this operation increases the dimension of the feature maps, leading to an increase in computation. In neck networks, whose input feature maps are provided by the backbone feature extraction network, semantic in-formation with high similarity already exists, so in this paper, we utilize the Add operation in place of the Concat operation to save parameters and GFLOPs. In order to enable the Add operation, we use a depth-separable convolution to downscale the 1024 × 1024 feature layer, and also need to adjust the input and output dimensions of the neck C2f module. The PANet-Z structure is shown in Fig. 7.

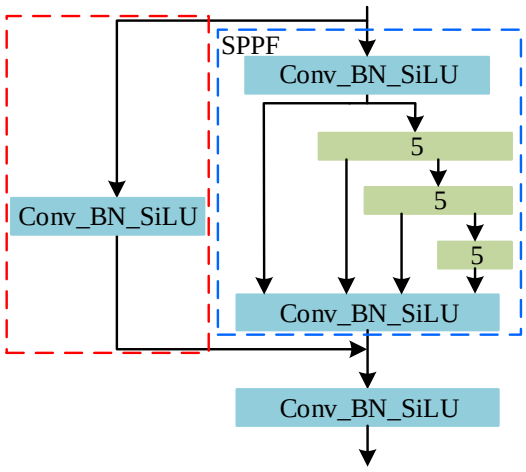


Fig. 6. SPPF-C module.

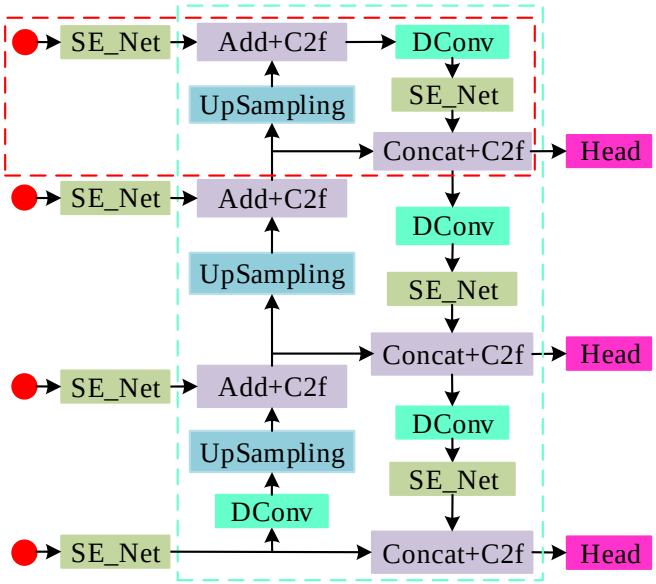


Fig. 7. PANet-Z module.

III. EXPERIMENTAL PREPARATIONS

A. Experimental Environment

The experiments described in this paper utilize the 64-bit Windows 10 operating system, the CORE i5 12490F processor, the RTX 3060 12GB graphics card model, with CUDA11.7 and cudnn8.8.0.121. Using Python3.9.7 programming language, PyTorch2.0 environment of the deep learning

framework, selecting Anaconda3 to configure the development environment and use PyCharm for development.

B. Datasets

In this paper, we study four detection targets, namely, transmission line insulators, insulator defects, voltage equalizing ring dataset, and bird nests in transmission line towers, which are locally enlarged as depicted in Fig. 8.

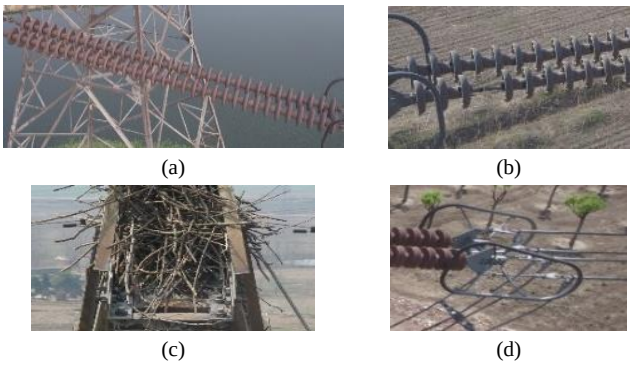


Fig. 8. Detect the target instance. (a) insulator; (b) insulator defect; (c) nest; (d) grading.

There are a total of 1308 images for the four detection targets in Fig. 8, some of which are from the publicly available Chinese Power Line Insulator Dataset [30], and some are images collected online. Because the number of labels for bird nest and insulator defects is small, only 568 and 843, and the defect samples are not balanced, the images containing the labels for bird nest and insulator defects are selected to expand them. By modeling different weather conditions, setting different exposure values and other operations to expand the data set, a total of 3239 images were generated, and the number of each label after expansion is shown in Table II. In the dataset, 20% of the images are randomly chosen to be tested at a later stage, while the remaining 80% are utilized for model training.

TABLE II. NUMBER OF VARIOUS LABELS IN THE DATASET

Label Types	Number of Original Labels	Number of Labels after Expansion
insulator	1935	3540
defect	845	2601
grading	1483	2785
nest	568	2576
total	4831	11,502

C. Evaluation Metrics

The criteria for experimental evaluation are mainly precision (P), recall (R), average precision (AP) and mean average precision (mAP) for each type of target. Precision evaluates the fraction of accurately predicted positive samples out of all samples predicted as positive, Recall is determined by the proportion of accurately predicted targets among all targets. mAP represents the mean of the AP values across classes, with a range between 0 and 1. A greater mAP signifies superior model performance. The calculation formula is specifically outlined as follows:

$$P = \frac{T_P}{T_P + F_P} \tag{7}$$

$$R = \frac{T_P}{T_P + F_N} \tag{8}$$

$$AP = \int_0^1 P(R)\mathrm{d}R \tag{9}$$

$$mAP = \frac{\sum_{i=1}^{n} AP_i}{n} \tag{10}$$

where: TP represents the count of samples identified as positive by the model that are indeed positive samples themselves; FP is the count of samples that the model identifies as positive samples that are themselves negative samples; FN signifies the quantity of samples identified as negative by the model, which are false negative samples; n signifies the total count of classifications; and AP_i is the AP value of the class i label.

D. Setting of Model Parameters

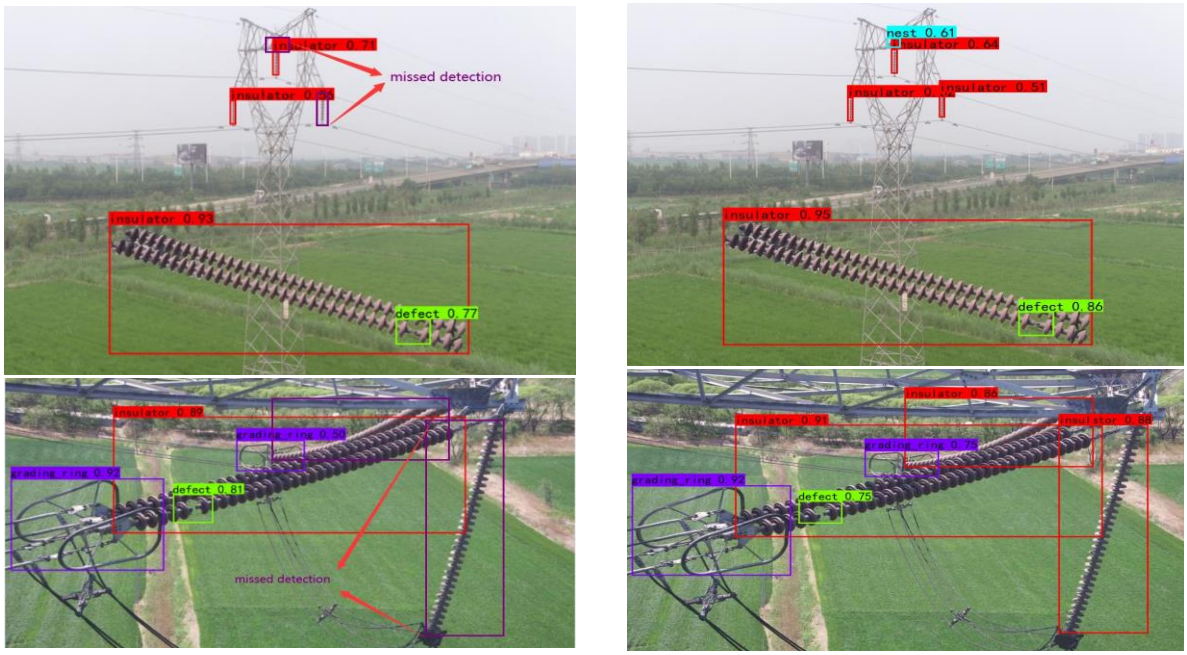
During the training process, transfer learning is employed to expedite model convergence and achieve improved accuracy. The training input image resolution is 608 × 608, the Adam optimizer is employed with an initial learning rate of 0.01, and cosine annealing is utilized to decay the learning rate. A total of 200 epochs are dedicated to training the model, with the freezing training epoch set to 50, and utilizing a batch size of 32. The unfreezing training epoch is set to 150, and the batch size is 16. The proportion of HSV-Saturation enhancement for images is set to 0.7, and the proportion of HSV-Value enhancement is set to 0.4.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

Fig. 9 displays several detection outcomes for small targets and occluded fabric markers, where (a) depicts the detection result of YOLOv8n and (b) illustrates the detection result of YOLO-T. Non-maximum suppression is employed as a post-processing technique during inference, with the confidence threshold set to 0.5 and the IoU threshold set to 0.5. Based on the detection results, it is evident that the original YOLOv8n algorithm exists in the case of leakage and misdetection, and the YOLO-T algorithm can be very well detected by a variety of targets. Furthermore, the FPS of the YOLO-T algorithm on RTX 3060 12G memory can reach 98.91 frames/s, which effectively satisfies the need for real-time detection.

A. Ablation Studies

In order to assess the effectiveness of the refined approach proposed in this paper, multi-group ablation experiments are conducted with the same training set, test set, experimental environment and model training parameters, and the resulting outcomes are presented in Table III.



(a) (b)

Fig. 9. Effect diagram of transmission line target detection. (a) YOLOv8n; (b) YOLO-T.

TABLE III. EXPERIMENTAL ABLATION RESULTS

Models	AP0.5 (%)				$mAP_{0.5}$ (%)	$mAP_{0.75}$ (%)	P (%)	R (%)	Parameter (M)	GFLOPs (G)
	Insulator	*Defect*	*Nest*	*Grading*						
YOLOv8n	92.36	97.79	94.50	95.97	95.16	78.18	98.03	88.43	3.012	7.398
YOLOv8n-A	90.72	98.07	94.22	95.67	94.68	76.44	98.17	85.64	2.756	6.916
YOLOv8n-B	93.57	98.06	95.19	96.92	95.93	78.45	98.31	89.05	2.588	6.486
YOLOv8n-C	94.15	98.23	95.04	96.73	96.04	80.41	98.47	90.15	2.601	6.487
YOLOv8n-D	95.03	98.37	95.78	97.49	96.67	81.19	98.47	91.00	2.602	6.488
YOLO-T	96.19	97.82	96.83	97.93	97.19	82.55	98.49	92.70	2.550	6.715

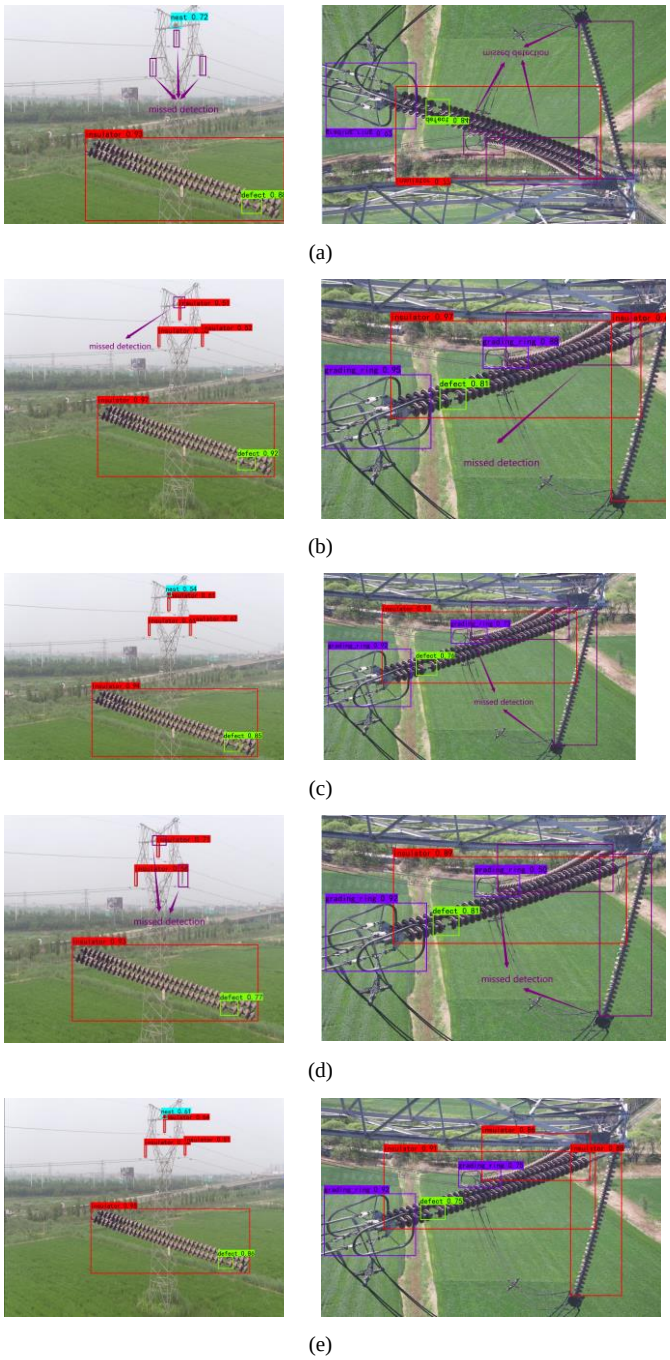
Based on the information provided in Table III, it is evident that YOLOv8n-A is an improvement of the C2f module of the neck network part of the baseline YOLOv8n algorithm, and the C2f-G module is constructed. According to the experimental results, the model's mAP0.5 decreased by 0.48%, but the parameters and GFLOPs were reduced by 0.744M and 0.482G, respectively, thus achieving a lighter model. YOLOv8n-B adds the CoT module based on YOLOv8n-A. Experiments show that the mAP0.5 of the model after adding the CoT module reaches 95.93%, which compensates for the impact of Ghost convolution on the model's detection accuracy. YOLOv8n-C is YOLOv8n-B based on the addition of the SE attention module after the effective feature layer of the output and the down-sampling operation of the neck network, which makes the mAP0.5 of the model reach 96.04% without increasing the complexity of the model almost 96.04%. YOLOv8n-D is constructed on the basis of YOLOv8n-C by improving the SPPF structure. According to the experimental results, the AP value of the YOLOv8n-D algorithm for all kinds of defect detection exceeds 95%, and mAP0.5 reaches 96.67%. Finally, the total parameters and the GFLOPs amount of the YOLO-T (C2f-G + CoT + SE + SPPF-C + PANet-Z) model proposed in this paper are reduced by 0.462M and 0.683G compared to the YOLOv8 baseline algorithm, and the AP values for all types of defects reach more than 96%, and mAP0.5 reaches 97.19%, which is an improvement over the baseline YOLOv8n model by 2.03%. The effectiveness of the improvement measures proposed in this paper was confirmed by the ablation experiments.

B. Comparative Testing of Different Models

To confirm the progress of the algorithm introduced in this paper, four different target detection models, YOLOv5s, YOLOv7, YOLOv8n and YOLOv8s, are also constructed for comparison and trained under the same experimental conditions and parameter settings. The detailed experimental results are presented in Table IV. Fig. 10 shows visualization examples of various algorithms in the transmission line dataset. From the visualization comparison in Fig. 10, it can be clearly observed that YOLO-T has better detection performance for long-distance targets and occluded targets.

TABLE IV. COMPARISON OF DIFFERENT MODEL TEST RESULTS

Models	$mAP_{0.5}$ (%)	$mAP_{0.75}$ (%)	Parameters (M)	GFLOPs (G)
YOLOv5s	91.24	67.10	7.072	14.893
YOLOv7	94.87	78.25	37.211	94.911
YOLOv8s	95.59	80.67	11.137	25.860
YOLOv8n	95.16	78.18	3.012	7.398
Ref. [23]	93.9	-	37.75	-
YOLO-T	97.19	82.55	2.550	6.715



(a) (b) (c) (d) (e)

Fig. 10. Detection results of different algorithms. (a) YOLOv5s; (b) YOLOv7; (c) YOLOv8s; (d) YOLOv8n; (e) YOLO-T.

V. DISCUSSION

The data in Table IV reveal that the YOLOv5s algorithm has more parameters and GFLOPs than YOLOv8n, but the model has a lower mAP. It has a mAP0.5 that is 2.54% lower and a mAP0.75 that is 11.08% lower than that of YOLOv8n. Its mAP0.5 and mAP0.75 are 2.54% and 11.08% lower than those of YOLOv8n, respectively. Although the YOLOv7 algorithm has a similar mAP to the YOLOv8n, but the model contains a larger quantity of parameters and GFLOPs which is nearly ten times larger than YOLOv8n. The YOLOv8 algorithms gradually increase the parameters quantity and GFLOPs as the depth and width of the model become larger, leading to the growth of mAP. However, in the transmission line dataset of this article, YOLOv8s has only a 0.4% increase in mAP0.5 compared to YOLOv8n. Concurrently, the parameters quantity and GFLOPs of the model increases by about four times. In contrast, the YOLO-T algorithm proposed in this paper achieves an mAP0.5 of 97.19%, which is 2.03% higher than the baseline YOLOv8n algorithm and 1.6% higher than the YOLOv8s. Importantly, the model features fewer parameters and the amount of GFLOPs are also lower by 8.587M and 19.145G. Comparative experiments prove that YOLO-T outperforms other algorithms in high transmission line multi-target detection, providing a valuable reference.

The YOLO-T algorithm presented enhances small and occluded target detection at long distances but is limited by the dataset's composition, where such targets are minimal. This limitation results in relatively low detection confidence for these targets, underscoring the necessity for optimization. To address this, expanding the dataset to include a broader spectrum of targets is recommended for future research, which could enrich training and allow for more extensive comparisons, essential for comprehensive high-voltage transmission line inspections. Additionally, since experiments have yet to be conducted with actual mobile drones, forthcoming studies should aim to deploy the refined model on embedded devices for enhanced multi-target detection on transmission lines, further honing the approach through practical application.

VI. CONCLUSIONS

To resolve the problem of low detection accuracy of long-distance small targets and occluded targets encountered in UAV inspection, this paper presents a YOLO-T transmission line multi-target detection algorithm, built upon an improved YOLOv8. Experimentally, it is proved that Ghost convolution can well realize the lightweighting of the model with less loss of detection accuracy. The backbone network's feature extraction capability is improved by using the CoT feature extraction module. By incorporating the SE attention module and adding a residual edge to the SPPF, the network is better able to focus on relevant information. This augments the model's feature extraction from small and occluded targets at a long range. Furthermore, the addition of a new shallow feature layer for multi-scale feature fusion enhances the model's detection accuracy for small targets and occluded objects. Additionally, the Add operation helps save the model's parameters and GFLOPs. The results of the experiments conducted on the transmission line inspection dataset show that

the mAP0.5 of the YOLO-T model can reach 97.19%, which is 2.03% higher than that of the original YOLOv8n algorithm, and the FPS reaches 98.91 frames/s, which can realize the real-time inspection of transmission lines. In addition, the parameter count of the YOLO-T model is only 2.55 M, which lays the foundation for the subsequent deployment on UAV embedded development board.

ACKNOWLEDGMENT

Author Contributions: S.L. designed the research. X.L. and Z.Z. processed the data. S.L. drafted the paper. H.O. and T.C. revised and finalized the paper. All authors have reviewed and approved the final version of the manuscript.

Funding: The research detailed in this article was funded by multiple sources: the Shanghai Local Institutions Capacity Building Program Project (22010501000); the Shanghai Multi-directional Die Forging Engineering and Technology Research Center Funded Project (20DZ2253200); and the Shanghai Lingang New Area Intelligent Manufacturing Industry Institute Funded Project (B1-0299-21-023).

Data Availability Statement: The Chinese Power Line Insulator datasets utilized in this paper are publicly available and can be downloaded from the Internet.

Conflicts of Interest: The authors declare that there are no conflicts of interest.

REFERENCES

[1] Liu, C.Y.; Wu, Y.Q. Research progress on visual detection methods for transmission lines based on deep learning. Chin. J. Electr. Eng. 2023, 43, 7423–7446.

[2] Han G J, Yuan Q W, Zhao F; et al. An Improved Algorithm for Insulator and Defect Detection Based on YOLOv4. Electronics 2023, 12, 933.

[3] Han, G.; Yuan, Q.; Zhao, F.; Wang, R.; Zhao, L.; Li, S.; He, M.; Yang, S.; Qin, L. Application of Deep Learning Object Detection Algorithm in Insulator Defect Detection of Overhead Transmission Lines. High Volt. Technol. 2023, 49, 3584–3595.

[4] Tudevdagva, U.; Battseren, B.; Hardt, W.; Troshina, G.V. Image Processing Based Insulator Fault Detection Method. In Proceedings of the 2018 XIV International Scientific-Technical Conference on Actual Problems of Electronics Instrument Engineering (APEIE), Novosibirsk, Russia, 2–6 October 2018; pp. 579–583.

[5] Zhai Y, Chen R, Yang Q; et al. Insulator fault detection based on spatial morphological features of aerial images. IEEE Access 2018, 6, 35316–35326.

[6] Lee, J.; Cha, S. Automatic object detection and tracking for unmanned aerial vehicle-based power line inspection using deep learning. IEEE Trans. Power Deliv. 2021, 36, 451–461.

[7] Yu, Z.; Yu, J.; Fan, J.; Tao, D. Multi-modal factorized bilinear pooling with co-attention learning for visual question answering. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 1821–1830.

[8] Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 2014; pp. 580–587.

[9] Girshick, R. Fast R-CNN. arXiv 2015, arXiv: 1504.08083.

[10] Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. Adv. Neural Inf. Process. Syst. 2015, 28, 91–99.

[11] Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. arXiv 2016, arXiv:1512.02325.

[12] Redmon J, Divvala S, Girshick R; et al. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, CA, USA, 2016; pp. 779–788.

[13] Redmon, J.; Farhadi, A. YOLO9000, Better, faster, stronger. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 2017; pp. 6517–6525.

[14] Redmon, J.; Farhadi, A. YOLOv3, An incremental improvement. arXiv 2018, arXiv:1804.02767.

[15] Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. YOLOv4, Optimal speed and accuracy of object detection. arXiv 2020, arXiv:2004.10934.

[16] Li, C.; Li, L.; Jiang, H.; Weng, K.; Geng, Y.; Li, L.; Ke, Z.; Li, Q.; Cheng, M.; Nie, W.; et al. YOLOv6, A Single-Stage Object Detection Framework for Industrial. arXiv 2022, arXiv:2209.02976.

[17] Wang, C.-Y.; Bochkovskiy, A.; Liao, H.-Y.M. YOLOv7, Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arXiv 2022, arXiv: 2207.02696.

[18] Lecun, Y.; Bengio, Y.; Hinton, G. Deep learning. Nature 2015, 521, 436–444.

[19] Zhou, W.; Ji, C.; Fang, M. Effective dual-feature fusion network for transmission line detection. IEEE Sens. J. 2023, 24, 101–109.

[20] Han, G.; He, M.; Zhao, F.; Xu, Z.; Zhang, M.; Qin, L. Insulator detection and damage identification based on improved lightweight YOLOv4 network. Energy Rep. 2021, 7, 187–197.

[21] Qiu, Z.; Zhu, X.; Liao, C.; Shi, D.; Qu, W. Detection of Transmission Line Insulator Defects Based on an Improved Lightweight YOLOv4 Model. Appl. Sci. 2022, 12, 1207.

[22] Li, D.P.; Ren, X.M.; Yan, N.N. Research on real-time detection of insulator string loss based on drone aerial photography. J. Shanghai Jiao Tong Univ. 2022, 56, 994–1003.

[23] Kang J, Wang Q, Liu W B. A Multi defect Detection Network for Aerial Insulators Integrating CAT-BiFPN and Attention Mechanism. High Volt. Technol. 2023, 49, 3361–3376.

[24] YOLOv8[EB/OL]. (2023-01-10)[2023-10-11]. Available online: https://github.com/ultralytics/ultralytics (accessed on 11 October 2023).

[25] Han, K.; Wang, Y.; Tian, Q.; Guo, J.; Xu, C.; Xu, C. GhostNet: More Features from Cheap Operations. arXiv 2020, arXiv:1911.11907.

[26] Li, Y.; Yao, T.; Pan, Y.; Mei, T. Contextual Transformer Networks for Visual Recognition. arXiv 2021, arXiv:2107.12292.

[27] Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E. Squeeze-and-excitation networks. IEEE Trans. Pattern Anal. Mach. Intell. 2020, 42, 2011–2023.

[28] Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Effificient Channel Attention for Deep Convolutional Neural Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.

[29] Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; Springer: Mubich, Germany, 2018; pp. 3–19.

[30] Tao, X.; Zhang, D.; Wang, Z.; Liu, X.; Zhang, H.; Xu, D. Detection of Power Line Insulator Defects Using Aerial Images Analyzed With Convolutional Neural Networks. IEEE Trans. Syst. Man Cybern. Syst. 2018, 50, 1486–1498.