Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2024.048864

ARTICLE

A Novel Foreign Object Detection Method in Transmission Lines Based on Improved YOLOv8n

Yakui Liu[1,2,3,*], Xing Jiang[1], Ruikang Xu[1], Yihao Cui[1], Chenhui Yu[1], Jingqi Yang[1] and Jishuai Zhou[1]

[1]School of Mechanical and Automotive Engineering, Qingdao University of Technology, Qingdao, 266520, China

[2]State Key Laboratory of Electrical Insulation and Power Equipment, Xi'an Jiaotong University, Xi'an, 710049, China

[3]Key Lab of Industrial Fluid Energy Conservation and Pollution Control, Qingdao University of Technology, Ministry of Education, Qingdao, 266520, China

*Corresponding Author: Yakui Liu. Email: liuyakui@qut.edu.cn

Received: 20 December 2023 Accepted: 11 March 2024 Published: 25 April 2024

ABSTRACT

The rapid pace of urban development has resulted in the widespread presence of construction equipment and increasingly complex conditions in transmission corridors. These conditions pose a serious threat to the safe operation of the power grid. Machine vision technology, particularly object recognition technology, has been widely employed to identify foreign objects in transmission line images. Despite its wide application, the technique faces limitations due to the complex environmental background and other auxiliary factors. To address these challenges, this study introduces an improved YOLOv8n. The traditional stepwise convolution and pooling layers are replaced with a spatial-depth convolution (SPD-Conv) module, aiming to improve the algorithm's efficacy in recognizing low-resolution and small-size objects. The algorithm's feature extraction network is improved by using a Large Selective Kernel (LSK) attention mechanism, which enhances the ability to extract relevant features. Additionally, the SIoU Loss function is used instead of the Complete Intersection over Union (CIoU) Loss to facilitate faster convergence of the algorithm. Through experimental verification, the improved YOLOv8n model achieves a detection accuracy of 88.8% on the test set. The recognition accuracy of cranes is improved by 2.9%, which is a significant enhancement compared to the unimproved algorithm. This improvement effectively enhances the accuracy of recognizing foreign objects on transmission lines and proves the effectiveness of the new algorithm.

KEYWORDS

YOLOv8n; data enhancement; attention mechanism; SPD-Conv; Smoothed Intersection over Union (SIoU) Loss

1 Introduction

Due to increasing urbanization, construction machinery operations may damage transmission lines. Additionally, human activities have destroyed many natural habitats, causing birds to seek alternative nesting sites, such as transmission towers. This scenario poses a threat to the safety of transmission lines. Power failure problems caused by external factors not only cause huge economic losses, but also threaten the safety of people's lives, and have now become a hidden problem of the power system that urgently needs to be solved. Therefore, it is necessary to detect faults [1] in

transmission lines to ensure the stable operation of the power system. Traditional inspection methods typically rely on manual inspection, which can be influenced by the environment and subjective judgment. Therefore, it is essential to adopt more efficient and effective inspection methods for transmission lines. The rapid development of unmanned aerial vehicles (UAVs) and machine vision has made this possible. UAVs capture images of transmission lines, ensure the security, transparency, and traceability of drone inspection data transmission through blockchain [2,3] technology, and machine vision is used to identify foreign objects in the images.

Deep learning [4] is a research field of machine learning that is widely used in various applications, including image recognition, speech recognition, and target detection. It is created using artificial neural networks that simulate the neural structure of the human brain. This results in a multi-layer neural network that extracts low-level features from data, such as images, speech, and text. These features are then combined to form more abstract high-level features, which better represent the distributional characteristics of the data. Traditional target detection methods depend on manually designed features, which are inefficient and struggle to utilize the extensive image data available. In recent years, deep learning has emerged as a rapid and powerful tool in image classification and target detection and has gained popularity in agriculture, medicine, remote sensing, and other fields. Its impressive feature learning capability has transformed image processing and target detection. Compared to conventional image processing methods, target detection techniques based on deep learning are characterized by stronger fault tolerance and robustness, as well as a more stable rate of recognition accuracy. Additionally, these techniques possess the benefit of being more economically viable and requiring lower labor costs.

Deep learning-based target detection algorithms can be broadly categorized into two groups. Firstly, there are the two-stage detection algorithms based on candidate regions, which involve the detection and recognition phases. Prominent examples of these algorithms include R-CNN [5], Fast R-CNN [6], Faster R-CNN [7], R-FCN [8], and others. The algorithms utilize feature information, including texture, color and image details. This data is initially divided into a range of proportionate and sized region boxes for detecting target presence. These region boxes are then inputted into the network for target detection. One-Stage detection algorithms, such as YOLO [9–12], SSD [13–15], and OverFeat [16], can determine the location and category of a detected object within a single step. These algorithms do not require the separate screening of candidate boxes to deliver a detection result, resulting in a faster detection speed.

In transmission lines scenarios, real-time and accurate detection and analysis of critical objects and large construction machinery that may cause damage in the transmission lines is required, and YOLO and convolutional neural (CNN) algorithms have the characteristics of fast detection, high accuracy, and strong feature extraction ability, so they are being detected in the field of transmission lines critical objects. Literature [1] proposes a genetic model that conditions the increase of the number and diversity of training images. Literature [17] designs a system based on edge cloud collaboration and reconuration of convolutional neural networks, combining pruned extraction network and compressed sign fusion network to improve the efficiency of multi-scale prediction. However, CNN localization targeting algorithms frequently involve varying parameters, resulting in optimal values differing across different scenarios, and posing challenges in dense area targeting. Literature [18] calculates the shape eigenvalues of insulators and backgrounds, and designed the classification decision conditions to be able to recognize insulators accurately. Literature [19] uses techniques such as CoordConv, DropBlock and Spatial Pyramid Pooling (SPP) to extract insulator features in complex backgrounds and trained the YOLO system with the dataset, which greatly improved the accuracy of aerial insulator detection. Literature [20] enhances the pyramid network by employing the attention mechanism as a feature

extraction network, leading to a significant improvement in prediction accuracy. The identification and detection of transmission lines has been a persistent issue, but literature [21] proposes improvements to the miniature target detection YOLO model by simplifying the feature extraction network (Darknet) and implementing a streamlined prediction anchor structure, resulting in effective transmission lines detection. However, as the network structure of the target detection model deepens, its theoretical performance is expected to gradually improve. Nevertheless, experiments have revealed that adding more layers beyond a certain depth does not lead to performance enhancement. Instead, it slows down the network training convergence process and ultimately reduces detection effectiveness. Empirical evidence suggests that residual networks can effectively solve the aforementioned issues. Facing the problem of low detection precision of some small and medium-sized targets such as detecting bird's nests, a new feature pyramid in feature fusion, path aggregation network and bidirectional feature pyramid [22] are used to improve the precision of small target detection. In the face of the complexity of the environment of the transmission lines, the loss function in the YOLOX algorithm is modified, and the literature adds Convolutional Block Attention Module (CBAM) [23] attention mechanism to the network to improve the feature extraction ability, while modifying the strong feature extraction part and introducing MSR algorithm to further optimize the picture, which significantly improves the recognition effect compared with the traditional YOLOX algorithm [24]. However, the CBAM attention mechanism performs lines and spatial attention operations sequentially, ignoring channel-space interactions and thus losing cross-dimensional information. In terms of the function of fast and accurate identification and localization of dangerous objects in transmission lines, literature [25] takes the YOLOv3 detection model as the basis and improves the bounding box non-maximum value suppression algorithm with reference to the Soft-NMS algorithm and Generalized Intersection over Union (GloU) algorithm to improve the detection model target detection precision rate and recall rate. Tiny remote sensing objects may be incorrectly detected without reference to a long enough range, and the range that needs to be referenced is mostly different for different objects, the introduction of LSKNet [26] in the literature can dynamically adjust the spatial field of view so as to better detect the research objects in different scenarios. In order to solve the limitation of using Complete Intersection over Union (CIoU) Loss [27], literature [28] adopts Smoothed Intersection over Union (SIoU) Loss instead of CIoU Loss function to improve the detection precision of the model and proposes the YOlOv8n-GCBLock-GSConv model, which not only reduces the cost of use, but also can quickly and accurately complete the detection of the target. Literature [29] uses a regression loss combining Weighted Intersection over Union (WIoU) [30] and distributed focusing loss to improve the model convergence ability and model performance superiority. Literature [31] used SIoU Loss instead of the original CIoU Loss in YOLOv7 to speed up the convergence speed and finally used SIoU-NMS to reduce the problem of detection omission due to occlusion. In the actual detection of transmission lines hazards, there are usually occlusions and external interference factors, literature [32] uses SPD-Conv combined with the CBAM attention mechanism so that the model can analyze the density in a specific region.

According to previous studies, deep learning-based algorithms for detecting transmission lines have many limitations, including decreased performance at lower resolutions, smaller objects, and complex environmental backgrounds. In the present paper, the above issues can be addressed by the following improvements:

(1) The SPD-Conv method is utilized to enhance the model's scale and spatial invariance. This is achieved by utilizing spatial pyramid pooling in combination with deep convolutional neural networks to construct a feature pyramid for parameter sharing and convolutional kernel size adaption. As a result, the accuracy and robustness of target detection are improved.

(2) The LSK module efficiently weights the features generated by a series of large deep convolutional kernels that are spatially merged through a spatially selective mechanism. The weights of these convolutional kernels are dynamically determined based on the inputs, enabling the model to adaptively use different large convolutional kernels and adjust the sensory field for each target as needed.

(3) The CIoU Loss is replaced with the SIoU Loss. The SIoU Loss takes into account angular loss, distance loss, and shape loss. It penalizes the size of the target frames and is more reflective of the true similarity between target frames. This replacement will speed up convergence and improve detection accuracy.

The paper is organized as follows: Section 2 introduces the YOLOv8n model. Section 3 details the proposed model, including the LSK Module, SPD-Conv module, and SIoU Loss. Section 4 presents the experimental results and analysis. Section 5 concludes the research.

2 Basic Structure of YOLOv8n

Ultralytics released YOLOv8 in January 2023, following the success of YOLOv5. YOLOv8 offers 5 official configurations, namely YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x, to cater to various scenario requirements. YOLOv8 introduces new features and improvements, such as a new backbone network, decoupled detection heads, and a new loss function, building on the previous version. YOLOv8n adopts a lightweight design that reduces the computational and storage requirements of the algorithm, which can enable the UAV to better process image and video data, improve the real-time and efficiency of the inspection, and enable the UAV to respond quickly and detect problems in a timely manner. Compared with YOLOv7, YOLOv8n improves the detection of small targets by improving the network architecture and training strategy, which increases its usefulness in the inspection process. Overall, the application of YOLOv8n in UAV inspection has higher accuracy, faster speed, and better adaptability and practicality. Therefore, YOLOv8n is selected as the basic training model in this paper.

The YOLOv8n detection model is comprised of four main components: Input, Backbone, Neck, and Head.

(1) Input. The data augmentation technique Mosaic [24] is often utilized in Input, with the anchor-free mechanism being employed to predict the object's center directly in lieu of the offset of the known anchor frames. This results in a reduction in the number of predicted anchor frames, thereby expediting non-maximal suppression NMS [33]. Data augmentation with Mosaic is discontinued in the final ten epochs.

(2) Backbone. The primary purpose of the Backbone is to extract features, and it comprises modules like Conv, C2f, and SPPF. Among them, the Conv module performs convolution, BN, and SiLU activation function operations on the input image. YOLOv8n introduces a new C2f structure as the main module for learning residual features, following YOLOv7s ELAN module. The C2f structure enriches the gradient flow rate by connecting more branches across layers, resulting in a neural network with superior feature representation capability. The SPPF module, also recognized as spatial pyramid pooling, expands the sensory field and captures feature information at various levels within the scene.

(3) Neck. The primary function of the Neck is to merge multi-scale features to produce a feature pyramid. This is achieved by implementing a path aggregation network, which involves using the C2f module to combine the feature maps that are obtained from three distinct phases of the Backbone. These measures facilitate the gathering of shallow data into more profound features.

(4) Head. The current prevalent decoupled header structure is employed by Head to separate the classification and detection headers, thus mitigating any potential disagreements between classification and localization tasks.

3 YOLOv8n Algorithm Improvement Strategy

In light of the lacklustre performance of conventional neural networks in handling low-resolution images and small objects, the SPD-Conv [34] module is applied. This module is capable of dynamically adapting its vast spatial perceptual field to better replicate the range context of diverse targets in a given scene. The Selective Attention LSK module is introduced to enhance the precision of the target detection and accelerate the training process of the neural network. Meanwhile, CIoU Loss is replaced by SIoU Loss to accelerate the convergence and improve detection precision. Based on the above work, the YOLOv8n network model has been improved, as depicted in Fig. 1.

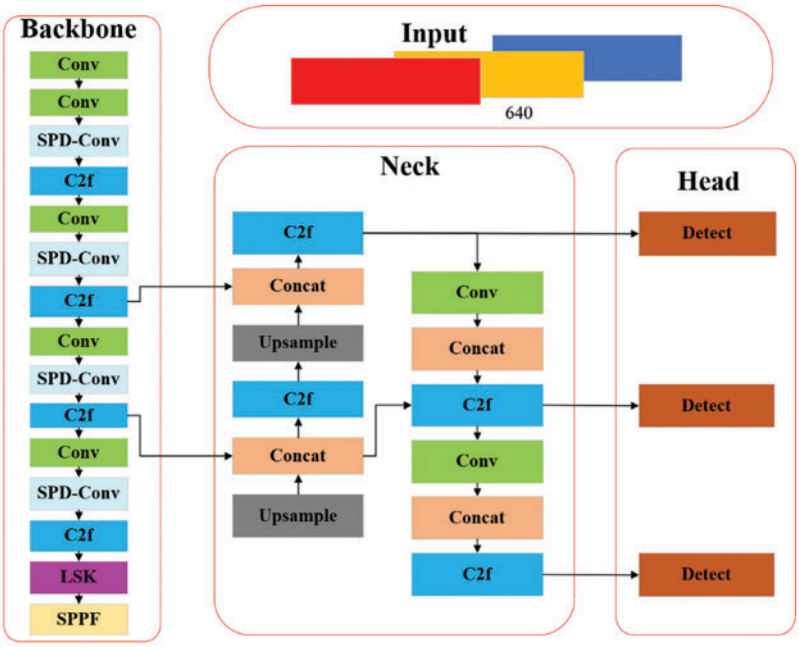

Figure 1: Improved YOLOv8n network mode

3.1 LSK Module

Current improvements for target detection algorithms often ignore the unique a priori knowledge of what occurs in a scene; aerial imagery is often captured in a high-resolution bird's-eye view, and many of the objects in the imagery may be small in size, making it difficult to accurately identify them based on appearance alone. Instead, the recognition of these objects often relies on their context, tiny remotely sensed objects may be mistakenly detected without reference to a sufficiently long range, and the long-range required may vary for different types of objects. But the surrounding background can provide valuable clues as to their shape, orientation and other characteristics. Therefore, this paper introduces the Large Selective Kernel Network (LSKNet) depicted in Fig. 2 [26], which can adaptively modify its expansive spatial sensing field to more accurately represent the remote sensing scene of diverse objects within the scene.

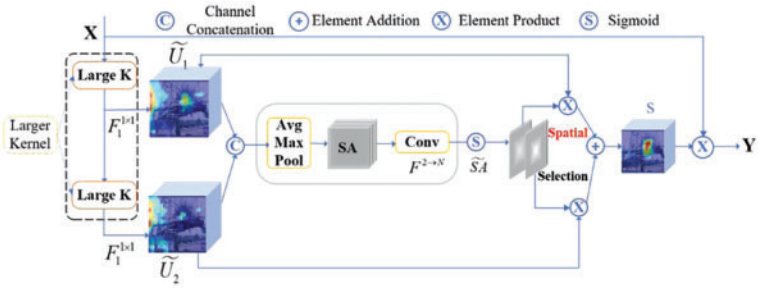

Figure 2: Conceptual drawing of the LSK module

The specific implementation of LSK is as follows: Firstly, two different feature maps are obtained by ordinary convolution and expansion convolution respectively, then the number of channels of the two are converted to the same size by the convolution kernel size of 1∗1, and then the two are stacked to obtain the feature map corresponding to c. Then, average pooling and maximum pooling are carried out on the feature map. Then, the two are stacked, convolved and sigmoid so that the selection weights for different convolution kernel sizes are obtained, and finally, the final output Y is obtained by multiplying and summing the weights with the proposed feature map and multiplying it with the initial input X.

3.2 SPD-Conv Module

Because of the advantages for processing low-resolution images and small target objects, SPD-Conv is introduced to replace the step-size convolution and pooling layers in the traditional CNN architecture. The structure of SPD-Conv is shown in Fig. 3 [34], which consists of a space-to-depth (SPD) layer and a non-step-size convolution (Conv) layer. The input feature maps are first transformed through the SPD layer. Then the convolution operation is performed through the Conv layer. The combination of SPD and Conv layer can reduce the number of parameters without losing information.

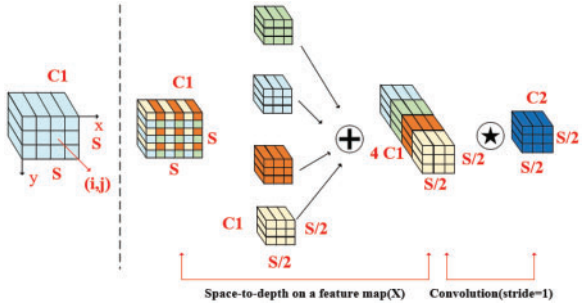

Figure 3: SPD-Conv structure

The process of the SPD-Conv can be summarized as follows: For an intermediate feature map X of arbitrary size, a sub-map consisting of and can be formed by scaling, and each sub-map is down-sampled proportionally to X. When scale = 2, 4 sub-maps are obtained, and the scale of each sample is 1/Scale of the original sample, and then the sub-feature maps are spliced along the lines to obtain the feature map X′, whose scale is $s/2 \times s/2 \times 4C_1$. Then the scale of the feature map X′ is changed to $s/2 \times s/2 \times C_2$, where $C_2 < 4C_1$, by using a non-step-size convolutional layer to preserve the key information as much as possible.

3.3 SIoU Loss

The CIoU Loss is a target detection loss function that integrates bounding box regression metrics. It is used by the traditional YOLOv8n model as its regression loss. Eq. (1) shows the loss function.

$$CIOU = IOU - \left(\frac{\rho^2\left(b, b^{gt}\right)}{c^2}\right) + \alpha\upsilon \tag{1}$$

$$\upsilon = \frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2 \tag{2}$$

$$\alpha = \frac{\upsilon}{(1 - IOU) + \upsilon} \tag{3}$$

The coordinates of the center point of the prediction frame are denoted by b, while b^{gt} denotes the coordinates of the center point of the real frame. The Euclidean distance between the prediction frame and the center point of the real frame is denoted by ρ^2, and c represents the diagonal length of the prediction frame and the real frame with the smallest external frame. The width and height of the frame are denoted by w and h, respectively. Additionally, υ represents the shape loss, and α represents the weight.

However, the approach has three obvious disadvantages: Low convergence and efficiency, due to its complex structure; highly sensitive to changes in target box scale, making it challenging to adapt the model to varying target box sizes; the misleading results when the aspect ratios of different prediction boxes and real frames are equal. To address the above issues, the SIoU Loss is applied as an alternative approach to CIoU Loss. SIoU Loss considers the vector angle between the actual frame and the predicted frame, redefines the penalty indicator in the loss function, and resolves the problem of mismatched directions that occur with un-framed frames in CIoU Loss. Moreover, SIoU Loss helps to avoid the predicted frame from training process of unstable drift during the training process, which improves the convergence speed of the model. SIoU Loss is calculated as:

$$U_{SIO} = U_{IO} - (\Delta + \Omega)/2 \tag{4}$$

where Δ is the distance loss, Ω is the shape loss, U_{IO} is the IoU loss.

4 Result and Analysis

4.1 Preparation before Calculation

4.1.1 Experimental Environment Configuration

The computing is conducted using Python and the Pytorch deep learning framework, the computing environment can be seen in Table 1.

Table 1: Experimental environment configuration

Parameter	Configuration
CPU	AMD Ryzen 758008-Core Processor
GPU	NVIDIA GeForce RTX 3070 Ti
CUDA	12.2

(Continued)

Table 1 (continued)

Parameter	Configuration
Pytorch	1.13
Epochs	200
Workers	0
Batch size	16
Image-size	[640, 640]
Optimizer	auto

4.1.2 Dataset Construction

Currently, the foreign hazards of transmission lines mainly stem from improper construction practices employed by large machinery as well as short circuits caused by avian nesting. This paper presents six target datasets of transmission lines captured through UAV, featuring excavators, trucks, bulldozers, tower cranes, bird nests, and cranes. The issue of a high number of images that were similar due to being captured by the same camera in a single scene was addressed by varying the angle and distance of the camera when taking shots. Images with different poses, including close-up, wide-angle, and side views, were collected. A dataset of 7,425 unique images was ultimately obtained. The categories and amounts of their datasets are displayed in Fig. 4, and some representative images are shown in Fig. 5.

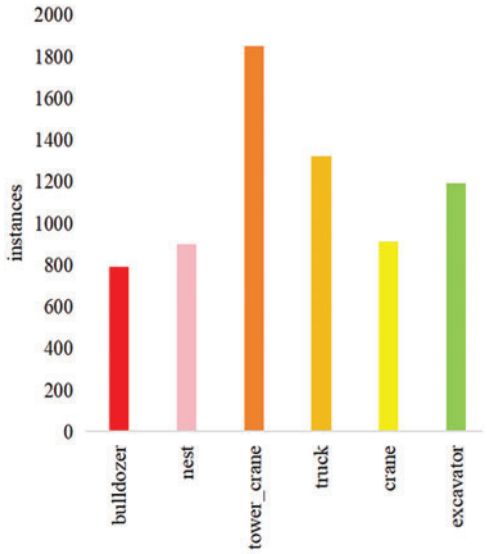

Figure 4: Type and number of data sets

The previous studies suggest that the application of combined data enhancement strategies can effectively improve the performance of machine learning models in tasks requiring precise image recognition capabilities. In this research, a methodological approach combining mosaic data augmentation with traditional data enhancement techniques was utilized to augment the diversity of the target sample dataset. This process entailed a sequence of transformations applied to the image,

encompassing operations such as flipping, scaling, and adjustments in the color gamut. The altered images were then merged with their corresponding bounding boxes. This amalgamation notably contributed to an enhancement in the model's generalization ability and robustness, as demonstrated in Fig. 6.

Figure 5: Partial images of transmission lines from UAV

Figure 6: Data enhancement

To refine the accuracy of the target detection model and assist the neural network in assimilating attribute and locational data of the targets, precise labeling of the objects within the images is imperative. For this purpose, the current study employed the Make Sense web platform for the annotation of the dataset, subsequently acquiring labels in the CoCo format. These labels encompass essential details, including the object name and its spatial coordinates within the image.

Additionally, the dataset underwent a random partitioning in an 8:1:1 ratio, resulting in the formation of a training set comprising 5,940 samples, a validation set comprising 742 samples, and a test set comprising 743 samples. Given the substantial dimensions of tower cranes and the pronounced issues of occlusion they present, a deliberate emphasis was placed on augmenting the representation of tower crane samples within the dataset. This approach is aimed at enhancing the model's capability to accurately identify and analyze such large-scale objects, despite the challenges posed by their size and potential for partial visibility in object detection.

4.2 Evaluation Indicators

To evaluate the model's performance objectively, we introduce several evaluation metrics, including precision, recall, F1, mAP50, mAP50-95, and frames per second transmitted (FPS). Precision, recall, and F1 are calculated as follows:

$$P = \frac{TP}{TP + FP} \tag{5}$$

$$R = \frac{TP}{TP + FN} \tag{6}$$

$$F1 = \frac{2 * P * R}{P + R} \tag{7}$$

where TP represents the count of detection frames that match the actual labelling and are predicted as positive samples; FP represents the count of detection frames that are predicted as positive samples but do not match the real labelling, and FN represents the count of real labelling that cannot be detected.

The graph in question is structured to represent Precision and Recall metrics along the horizontal and vertical axes, respectively. Within this framework, the area enclosed by the plotted curve is indicative of the Average Precision (AP) value for a given category. The mean Average Precision (mAP) is subsequently derived as the mean of the AP values across all categories. Specifically, mAP50 denotes the average of Precision values for all categories at the 50% Intersection over Union (IoU) threshold, while mAP50-95 represents the mean of mAP values calculated at various IoU thresholds, ranging from 50% to 95%.

Furthermore, the calculation of mAP is governed by the following formula, which quantitatively assesses the model's accuracy by averaging the precision across different recall levels and categories, thereby providing a comprehensive evaluation of the model's performance in object detection tasks. This formula encapsulates the integral aspects of precision and recall, offering a robust metric for the assessment of detection algorithms.

$$mAP = \frac{1}{n} \sum_{i=1}^{n} AP_i \tag{8}$$

where P is the proportion of prediction frames that exactly detected the target out of all prediction frames, and R is the proportion of prediction frames that actually detected the target out of all true labelled frames.

FPS indicates the number of frames processed per second, which is used to measure the speed of the detection performance; the higher the value, the faster the detection speed and the better the detection performance.

4.3 Comparison of the Effects of Improved Methods

The effect of common loss functions, including CIoU, SIoU, DIoU, GIoU and WioU, is assessed by comparing the IoU loss function, and the results can be seen in Table 2.

Table 2: Comparison of model performance using different IoU loss functions

Method	Loss function	P	R	mAP50	mAP50-95	FPS	F1
1	CIoU	0.905	0.81	0.873	0.643	60.6	0.855
2	DIoU	0.884	0.809	0.866	0.638	61.73	0.845
3	SIoU	0.886	0.83	0.891	0.647	65.35	0.857
4	GIoU	0.903	0.807	0.871	0.643	60.98	0.852
5	WIoU	0.896	0.817	0.867	0.637	61.35	0.84

As illustrated in Table 2, the YOLOv8n+SIoU model, which uses the SIoU loss function, achieves the highest mAP50 compared to the original YOLOv8n+CIoU model. Although the accuracy decreases slightly, the mAP50, mAP50-95, Recall, FPS, and F1 improve by 1.8%, 0.4%, 2%, 7.8%, and 0.2%, respectively. Compared to the traditional model, the proposed method (YOLOv8n+SIoU Loss) achieves higher detection speed and precision. The use of SIoU effectively improves model fitting and recognition accuracy. Compared to models with different loss functions, the model incorporating SIoU demonstrates superior comprehensive performance and the greatest ability to improve the original model.

To evaluate the efficacy of the LSK attention mechanism, a few other common attention mechanisms are used to compare the detection ability. Specifically, the LSK and CBAM along with the SE [35], and EMA [36] attention mechanisms are added to the final layer of the backbone. The results are shown in Table 3.

Table 3: Performance comparison of models incorporating different attention mechanisms

Method	Attention mechanism	p	R	mAP50	mAP50-95	FPS	F1
1	–	0.905	0.81	0.873	0.643	60.6	0.855
2	CBAM	0.919	0.812	0.88	0.65	59.17	0.862
3	SE	0.881	0.822	0.872	0.642	60.6	0.85
4	EMA	0.905	0.8	0.881	0.647	57.47	0.849
5	LSK	0.918	0.813	0.881	0.647	62.5	0.862

Table 3 shows that, with the exception of the SE attentional mechanism, the accuracy of models incorporating the other three attentional mechanisms improved to varying degrees compared to the original model. The models that integrated the CBAM and LSK attentional mechanisms improved by 1.4% and 1.3%, respectively, in terms of accuracy. In terms of mAP50, the CBAM, EMA, and LSK attention mechanisms improved the model by 0.7%, 0.8%, and 0.8%, respectively. The models that utilized both CBAM and LSK attention mechanisms showed a 0.7% improvement in F1 score. Regarding recall, the models that employed CBAM, SE, and LSK attention mechanisms showed improvements of 0.2%, 0.12%, and 0.3%, respectively. Furthermore, the model that utilized LSK attention mechanism demonstrated the fastest detection speed, with a 3.1% improvement compared to the original YOLOv8n model. In conclusion, the model that incorporates the LSK attention mechanism provides a better trade-off between speed and accuracy and has the best overall performance.

To better illustrate the impact of integrating the attention mechanism on model detection effectiveness, GradCAM [37] heat maps are utilized to visually analyze and compare the detection outcomes of the unimproved model and the model enhanced with the LSK attention mechanism. Fig. 7 displays the detection outcome before enhancement. Conversely, Fig. 8 illustrates the detection outcome after integrating the attention mechanism. The red area highlights the region towards which the model pays more attention while the lighter area shows the opposite. The application of the LSK attention mechanism indicates that the model focuses more on the area nearby to the target, which also helps to suppress the computational power occupied by non-target region.

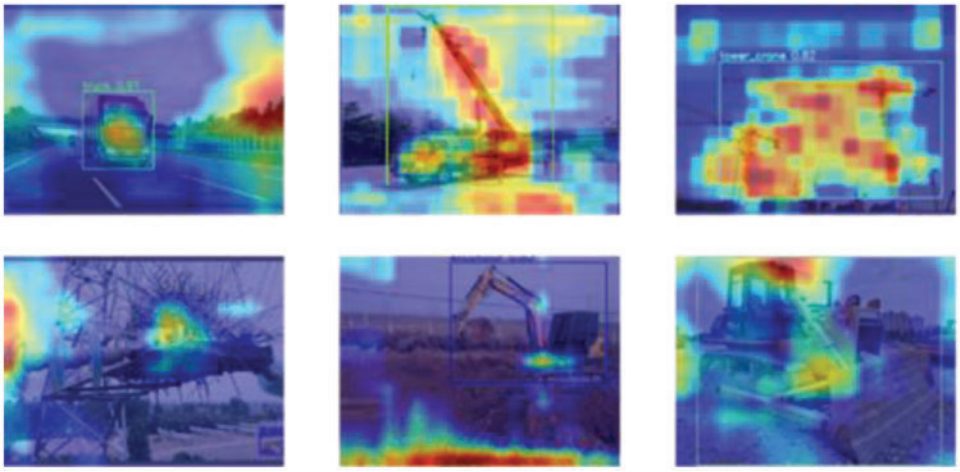

Figure 7: Test results of traditional YOLOv8n

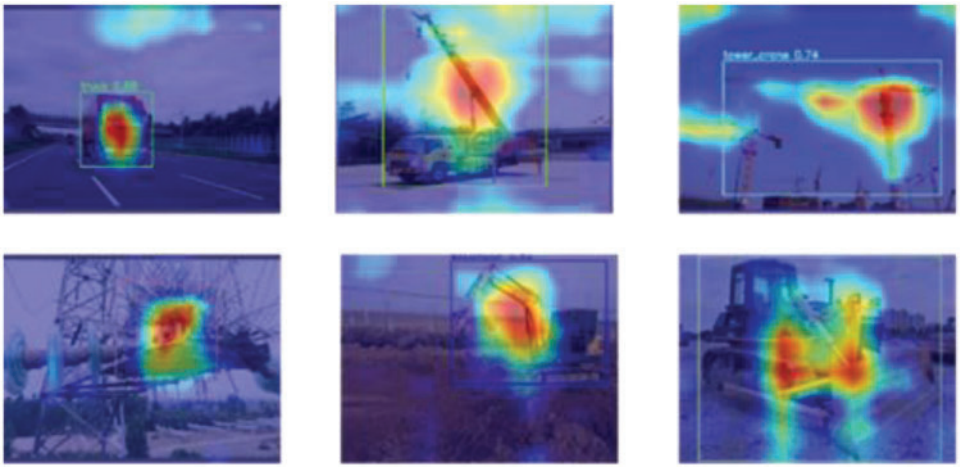

Figure 8: Test result of proposed YOLOv8n

4.4 Ablation Study

To further validate the efficacy of the various improvement methods of the YOLOv8n model, ablation experiments were conducted using different combinations of multiple enhancement modules.

To substantiate the effectiveness of the diverse enhancement methodologies applied to the YOLOv8n model, a series of ablation studies were conducted. These studies involved the utilization of various combinations of enhancement modules, providing a comprehensive evaluation of each method's impact on model performance.

The ablation experiments were conducted using the same training and test sets. YOLOv8n was used as the base framework, and different attentional mechanisms and loss functions were sequentially adopted to obtain new models for training. The results are shown in Table 4.

Upon adding SPD-Conv to the original model, all indexes, except for mAP50, were reduced to varying degrees. The reason for this is unclear. SPD-Conv is a type of spatial depth convolution that can alter the feature map representation, making it difficult for the network to accurately learn target

boundaries or features. To address this issue, this paper proposes adding an attention mechanism to enhance the model's focus on important features, thereby improving target detection accuracy. The table data shows that the model with LSK attention mechanism has significantly improved comprehensive performance compared to the original model. Although there is a slight decrease in detection speed and recall rate, the detection accuracy, mAP50, mAP50-95, and F1 have all improved to varying degrees. Specifically, the detection accuracy has improved by 1.2%, mAP50 by 0.8%, mAP50-95 by 1%, and F1 by 0.3%. Compared to the original model, the model that introduces SIoU Loss has shown improvement in all aspects except for a slight decrease in mAP50-95. The detection accuracy has reached 91.7%, while mAP50 has reached 88.8%. The experimental results indicate that the combination method containing the SPD-Conv module, the LSK attentional mechanism, and the SIoU Loss has achieved the highest accuracy level, demonstrating the effectiveness of the improved model.

Table 4: Ablation study results

Method	SPD	LSK	EMA	SIoU	P	R	mAP50	mAP50-95	FPS	F1
1					0.905	0.81	0.873	0.643	64.93	0.855
2	√				0.918	0.813	0.881	0.647	65.36	0.844
3	√	√			0.917	0.806	0.881	0.653	58.47	0.858
4	√		√		0.874	0.825	0.88	0.641	54.35	0.849
5	√				0.894	0.815	0.875	0.653	59.88	0.853
6	√		√	√	0.915	0.816	0.882	0.653	59.17	0.863
7	√	√		√	0.917	0.817	0.888	0.651	59.17	0.864

4.5 Algorithm Verification

A comparative analysis of the detection performance between the traditional YOLOv8n model and its enhanced counterpart is presented in Fig. 9. The left column shows the traditional model, and the right column shows the improved model. This figure illustrates the detection capabilities of the YOLOv8n model, particularly highlighting the challenges posed by targets with diminutive scales and low contrast against the background. The traditional YOLOv8n model demonstrates variable degrees of detection failures, particularly noted in the imagery of groups a, b, and d, encompassing instances of both leakage and misdetection, especially evident in the images from groups b and c. Conversely, the augmented YOLOv8n model exhibited a consistent ability to accurately detect all targets, even in conditions of low clarity or small target size. The enhancement in detection efficacy is particularly noticeable in scenarios involving targets with ambiguous outlines or reduced scales. This comparative evaluation solidly establishes the superior performance of the enhanced YOLOv8n model, affirming its efficacy in complex detection environments where precision is critical.

(a) detection failures of tower crane

(b) detection failures of truck and misdetection of excavator

(c) misdetection of truck

(d) detection failures of truck and bulldozer

Figure 9: Comparison of YOLOv8n model detection results

5 Conclusion

This study presents an improved version of the YOLOv8n algorithm that is specifically designed to detect foreign objects on transmission lines. The iterative version of the algorithm employs the SPD-Conv module, which replaces the stepping and pooling operations with a space-to-depth convolution followed by a non-stepping convolution. This eliminates the stepping and pooling operations altogether. The model's ability to handle objects with low image resolution or size is enhanced by

downsampling the feature map while preserving distinguishing feature information. Additionally, the selective attention LSK module is included to dynamically adjust its large spatial receptive field to better simulate the ranging context of various objects in the scene and improve the accuracy of detecting small targets. Additionally, substituting the CIoU Loss function with the SIoU Loss function aids in quicker model convergence. The experimental results demonstrate that the SIoU Loss function produces superior detection outcomes when there is significant overlap between target frames. These experimental results confirm the effectiveness of the improved model in object detection and significantly enhance detection accuracy. The data indicate that the enhanced algorithm attains an average detection accuracy of 88.8% and a detection speed of 59.17 frames per second (FPS), demonstrating its potential applicability in identifying foreign objects on transmission lines.

This paper reports progress in recognizing cranes at lower image resolutions or with smaller objects, particularly in the field of leakage detection misdetection. However, the algorithm still has limitations, and tower crane identification accuracy, a significant threat to transmission lines, needs improvement. Tower cranes typically present multiple intersecting lines and angles in their images, increasing the difficulty of accurate bounding box localization. Additionally, tower crane environments are often cluttered with complex scenes, such as construction sites, which can be easily mistaken for surrounding objects. This makes it challenging for algorithms to accurately extract tower crane features from the background. Future research will concentrate on methodological enhancements to improve network performance. To address the identified shortcomings and enable practical applications of the algorithms in complex real-world environments, we will explore the use of larger datasets, increased sensitivity to boundary information, and expanding the sensory field of the model.

Acknowledgement: We would like to express our sincere gratitude to the Natural Science Foundation of Shandong Province and the State Key Laboratory of Electrical Insulation and Power Equipment for providing the necessary financial support to carry out this research project.

Funding Statement: This research was funded the Natural Science Foundation of Shandong Province (ZR2021QE289), and State Key Laboratory of Electrical Insulation and Power Equipment (EIPE22201).

Author Contributions: The authors confirm contribution to the paper as follows: Study conception and design: Yakui Liu, Xing Jiang; data collection: Yakui Liu, Xing Jiang; analysis and interpretation of results: Ruikang Xu, Yihao Cui, Jingqi Yang; draft manuscript preparation: Yakui Liu, Chenhui Yu, Jishuai Zhou. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Data is not available due to commercial restrictions.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] Z. Qian, W. Jing, Y. Lv, and W. Zhang, "Automatic polyp detection by combining conditional generative adversarial network and modified you-only-look-once," *IEEE Sens. J.*, vol. 22, no. 11, pp. 10841–10849, Jun. 2022. doi: 10.1109/JSEN.2022.3170034.

[2] A. A. Khan *et al.*, "A drone-based data management and optimization using metaheuristic algorithms and blockchain smart contracts in a secure fog environment," *Comput. Electr. Eng.*, vol. 102, no. 11, pp. 108234, Sept. 2022. doi: 10.1016/j.compeleceng.2022.108234.

[3] A. A. Khan, A. A. Wagan, A. A. Laghari, A. R. Gilal, I. A. Aziz, and B. A. Talpur, "BIoMT: A state-of-the-art consortium serverless network architecture for healthcare system using blockchain smart contracts," *IEEE Access*, vol. 10, pp. 78887–78898, Jul. 2022. doi: 10.1109/ACCESS.2022.3194195.

[4] W. J. Zhang, G. Yang, Y. Lin, C. Ji, and M. M. Gupta, "On definition of deep learning," in *Proc. 2018 World Autom. Congr. (WAC)*, Stevenson, WA, USA, Jun. 2018, pp. 1–5.

[5] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. 2014 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Columbus, OH, USA, Jun. 2014, pp. 580–587.

[6] R. Girshick, "Fast R-CNN," in *2015 IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Dec. 2015, pp. 1440–1448.

[7] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017. doi: 10.1109/TPAMI.2016.2577031.

[8] J. F. Dai, Y. Li, K. M. He, and J. Sun, "R–FCN: Object Detection via region-based fully convolutional networks," in *Proc. 30th Int. Conf. Neural Inform. Process. Syst.*, 2016, pp. 379–387.

[9] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, 2016, pp. 779–788.

[10] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, 2017, pp. 6517–6525.

[11] J. Redmon and A. Farhadi, "YOlOv3: An incremental improvement," arXiv preprint arXiv:1804.02767, 2018.

[12] A. Bochkovskiy, C. Y. Wang, and H. Y. Liao, "YOLOv4: Optimal speed and accuracy of object detection," arXiv preprint arXiv:2004.10934, 2020.

[13] W. Liu *et al.*, "SSD: Single shot MultiBox detector," in *2016 Eur. Conf. Comput. Vis. (ECCV)*, Amsterdam, The Netherlands, Oct. 2016, pp. 21–37.

[14] C. Y. Fu, W. Liu, A. Ranga, A. Tyagi, and A. C. Berg, "DSSD: Deconvo- lutional single shot detector," arXiv preprint, arXiv:1701.06659, 2017.

[15] Z. X. Li and F. Zhou, "FSSD: Feature fusion single shot multibox detector," arXiv preprint arXiv:1712.00960, 2017.

[16] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks," arXiv preprint arXiv:1312.6229, 2013.

[17] S. Liang *et al.*, "Real-time intelligent object detection system based on edge-cloud cooperation in autonomous vehicles," *IEEE Trans. Intell. Transport. Syst.*, vol. 23, no. 12, pp. 25345–25360, Dec. 2022. doi: 10.1109/TITS.2022.3158253.

[18] C. Yao, L. Jin, and S. Yan, "Recognition of insulators in grid inspection images," (in Chinese), *J. Syst. Simul.*, vol. 24, no. 9, pp. 1818–1822, 2012.

[19] K. Y. Chen, L. G. Xu, and F. X. Chen, "Research on aerial insulator detection method of aerial transmission lines based on YOLOv3," *Sci. Tech. Innov. Appl.*, vol. 13, no. 11, pp. 34–37, 2023. doi: 10.19981/j.CN23-1581/G3.2023.11.009.

[20] X. Y. Jia, H. Q. Wang, Y. Z. Yang, Z. M. Cui, and B. Xiong, "Anchorless frame SAR image ship target detection based on YOLO framework," (in Chinese), *Syst. Eng. Electron.*, vol. 44, no. 12, pp. 3703–3709, 2022.

[21] L. B. Yang, J. N. Yang, and X. M. He, "Transmission line detection based on YOLO target detection algorithm," *Power Inform. Commun. Technol.*, vol. 20, no. 8, pp. 99–105, 2022. doi: 10.16543/j.2095-641x.electric.power.ict.2022.08.011.

[22] S. H. Su, "Research and application of target recognition of transmission lines construction machinery based on improved faster R-CNN," Ph.D. dissertation, Shandong Univ., China, 2021.
[23] S. H. Woo, J. C. Park, J. Y. Lee, and S. K. In, "CBAM: Convolutional block attention Module," in *2018 Eur. Conf. Comput. Vis., Lecture Notes Comput. Sci.*, 2018, pp. 3–19.
[24] Z. J. Zhang, X. H. Cao, L. Meng, and H. J. Zou, "Detection and identification of engineering vehicles in transmission corridor based on improved YoloX," (in Chinese), *Comput. Measur. Control.*, vol. 30, no. 9, pp. 67–73, 2022.
[25] N. Shao, "Research on the key technology of deep learning-based transmission lines hazardous object identification," Ph.D. dissertation, Shandong Univ., China, 2020.
[26] Y. Li, Q. Hou, Z. Zheng, M. M. Cheng, J. Yang, and X. Li, "Large selective kernel network for remote sensing object detection," in *2023 IEEE Int. Conf. Comput. Vis.*, Paris, France, 2023, pp. 16748–16759.
[27] Z. H. Zheng, P. Wang, W. Liu, J. Z. Li, R. G. Ye, and D. W. Ren, "Distance-IoU loss: Faster and better learning for bounding box regression," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 7, pp. 12993–13000, Feb. 2020. doi: 10.1609/aaai.v34i07.6999.
[28] H. C. Yuan and L. Tao, "Detection and identification of fish in electronic monitoring data of commercial fishing vessels based on improved YOLOv8," (in Chinese), *J. Dalian Ocean Univ.*, vol. 38, no. 3, pp. 533–542, 2023.
[29] A. Gao, X. Z. Liang, C. X. Xia, and C. J. Zhang, "An improved dense pedestrian detection algorithm for YOLOv8," (in Chinese), *J. Graphics.*, vol. 44, no. 5, pp. 890–898, 2023.
[30] Z. J. Tong, Y. H. Chen, Z. W. Xu, and R. Yu, "Wise-IoU: Bounding box regression loss with dynamic focusing mechanism," arXiv preprint arXiv:2301.10051, 2023.
[31] R. Zhen, Y. Liu, F. H. Meng, Y. Liu, S. K. Su, and H. T. Zhao, "Low-altitude flying object target detection method based on improved YOLO v7," (in Chinese), *Radio Eng.*, vol. 54, no. 3, pp. 1–14, 2023.
[32] Z. Y. Li and Q. K. Chen, "SPD-based improved YOLOv5 for crowd area density analysis of bus stops," (in Chinese), *J. Chin. Comput.*, pp. 1–9, 2023.
[33] Y. He, C. Zhu, J. Wang, M. Savvides, and X. Zhang, "Bounding box regression with uncertainty for accurate object detection," in *2019 IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, 2019, pp. 2883–2892.
[34] P. H. Hsu, P. J. Lee, T. A. Bui, and Y. S. Chou, "YOLO-SPD: Tiny objects localization on remote sensing based on You Only Look Once and Space-to-Depth Convolution," in *Proc. 2024 IEEE Int. Conf. on Consum. Electron. (ICCE)*, Las Vegas, NV, USA, 2024, pp. 1–3.
[35] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation networks," in *2018 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 7132–7141.
[36] D. Ouyang, "Efficient multi-scale attention module with cross-spatial learning," arXiv preprint arXiv: 2305.13563, 2023. doi: 10.48550/arXiv.2305.13563.
[37] G. M. Zhang, "Design and development of AR home decoration system based on Vuforia SDK development," (in Chinese), *Comput. Knowl. Technol.*, vol. 16, no. 19, pp. 80–81, 2021.