IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 16 July 2024, accepted 29 July 2024, date of publication 1 August 2024, date of current version 12 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3436919

RESEARCH ARTICLE

An Insulator Location and Defect Detection Method Based on Improved YOLOv8

ZHONGSHENG LI[1], CHENDA JIANG[2], AND ZHONGLIANG LI[1]
[1]School of Traffic Engineering, Fujian Polytechnic of Water Conservancy and Electric Power, Yongan 366000, China
[2]School of Electric Power Engineering, Fujian Polytechnic of Water Conservancy and Electric Power, Yongan 366000, China

Corresponding author: Chenda Jiang (jcd811107@163.com)

This work was supported by the 2021 Fujian Water Resources and Electric Power Technical College Talent Introduction Research Project "Design of S7-1200 Programmable Logic Controller (PLC) Training Device Based on Fieldbus Technology" under Grant gccrcyjxm_lzl.

ABSTRACT Ensuring the integrity of insulators is critical for the reliability and safety of power transmission systems. To address the need for efficient and real-time inspection of insulator defects on power lines, this paper introduces an advanced defect detection model built upon the YOLOv8 architecture. The model incorporates a novel C2f-Faster-EMA module that modifies the original C2f module used in YOLOv8's backbone for feature extraction. This adaptation employs FasterNet to reduce the model's parameter count, while incorporating an EMA-based attention mechanism to enhance detection accuracy. Additionally, we replace the conventional PANet structure with a BiFPN-P feature fusion module to improve the extraction of shallow features, which is crucial for detecting small-target defects in insulators. Further refinements include the implementation of the Inner-IoU concept to augment the MPDIoU loss function, thus improving the model's ability to learn from challenging samples. Experimental results demonstrate that the proposed Insulator-YOLOv8s model achieves a superior performance over existing mainstream algorithms, with a mean Average Precision (mAP) of 91.5% at an IoU threshold of 0.5. The model is characterized by a parameter count of 5.66M and computational requirements of 21.1 GFLOPs, achieving detection speeds up to 113 frames per second. These enhancements enable the proposed model to identify insulator defects swiftly and with high accuracy, thereby contributing significantly to the safety and maintenance of power transmission infrastructure.

INDEX TERMS Insulator location, defect detection, YOLOv8, lightweight.

I. INTRODUCTION

As technology progresses and the smart industry evolves, the complexity and scale of power networks continue to grow, introducing heightened challenges in the safety inspection of power lines [1]. Insulators are a crucial component of electrical power systems. Their primary functions are to support and secure conductors while ensuring electrical insulation between the conductors and other components. The relationship between insulators and other components encompasses mechanical support, electrical insulation, and safety protection. However, insulators are inevitably susceptible to various failures due to the influence of factors such as lightning, storms, and magnetic fields during their operation. These damages can manifest as spontaneous shattering, string breakages, corrosion, and cracking, posing significant threats to the stability of the power grid [2], [3], [4]. The necessity for regular inspection and timely mitigation of potential risks is paramount. Traditionally, insulator defect detection has relied on manual inspections, which are not only inefficient and costly but also pose substantial safety risks [5]. In response, recent advancements have seen the adoption of computer vision technology for aerial imaging, presenting a more economical and effective method for detecting insulator defects. Nonetheless, achieving accurate detection remains challenging due to the small size of defect pixels relative to the overall image and the complex backgrounds typical of aerial photographs [6].

In recent years, the advancements in deep learning technologies have significantly enhanced the field of object detaction, and its related technologies have also been widely used in all walks of life [7], [8], [9], [10]. Initially, object detection algorithms relied on manually designed features. This methodology has transitioned to the use of

The associate editor coordinating the review of this manuscript and approving it for publication was Prakasam Periasamy.

Convolutional Neural Networks (CNNs), which automate the identification and extraction of image features, thereby improving detection efficiency [11]. Object detection algorithms are primarily categorized into two types: two-stage and single-stage detection methods [12]. Two-stage detection methods, which include generating candidate regions for object detection, are known for their high detection accuracy despite their complexity and slower speeds. Notable examples include R-CNN [13], Fast R-CNN [14], and Faster R-CNN [15]. For instance, Shan et al. [16] employed deep residual networks and cascaded R-CNN operations to detect various objects, including normal insulators, insulator explosions, and bird nests. Similarly, Ghashghaei et al. [17] utilized a concatenated Faster R-CNN model integrated with feature pyramids and region proposal networks, significantly enhancing the detection and recognition accuracy of insulator explosions.

Conversely, single-stage detection methods have gained popularity due to their simplified network architectures and rapid detection capabilities. Noteworthy algorithms include SSD (Single Shot Detector) [18] and YOLO (You Only Look Once) [19]. Wang et al. [20] refined the SSD algorithm for identifying power components in transmission lines, achieving high detection accuracy with fast response times. Additionally, Zhao et al. [21] implemented MobileNet [22] as the backbone for SSD, leveraging its depthwise separable convolutions and bottleneck architecture to reduce the model's parameter count. Zheng et al. [23] developed an improved real-time detection algorithm using YOLOv4 [24], which integrated additional positional and semantic information to enhance multi-scale target detection accuracy and reduce false positives. Chen et al. [25] introduced attention mechanisms and a bidirectional pyramid structure into YOLOv5 [26], improving the detection of insulators and their defects in complex backgrounds. Yi et al. [27] introduced lightweight convolution, designed MaECA attention mechanism, and adopted Mish activation function and SIoU loss function to improve YOLOv5, thus improving the detection accuracy and speed of insulator images. Moreover, Zhai et al. [28] enhanced the detection capabilities of YOLOv7 [29] for small targets by designing the C3GhostNetV2 module and incorporating the CA [30] coordinate attention mechanism, thus reducing the model's complexity.

Despite these advancements, the localization and defect detection of insulators still face numerous challenges. For example, when dealing with complex background environments, most methods struggle to effectively extract features of small and blurry targets, resulting in high rates of missed and false detections of insulator defects. Additionally, due to the memory limitations of mobile hardware devices, effectively balancing detection performance with hardware resource consumption remains one of the challenges in achieving real-time insulator defect detection. In this context, this paper proposes a lightweight detection algorithm, Insulator-YOLOv8s, based on an enhanced version of YOLOv8 [31]. This algorithm is designed to optimize the balance between model detection speed and accuracy. The main contributions of this paper are as follows.

1) A lightweight C2f-Faster-EMA module is designed and integrated into the backbone network, replacing the original C2f module in YOLOv8. This modification significantly reduces the number of parameters and computational load of the model, while further enhancing detection accuracy and speed.
2) A BiFPN-P feature fusion module is introduced in the neck network, which strengthens the feature extraction capability for shallow information. This improvement notably increases the detection accuracy for small insulator defects and effectively reduces both the miss rate and false detection rate.
3) The Inner-MPDIoU loss function is developed to optimize the original bounding box regression loss function. This advancement accelerates the bounding box regression process and enhances the model's generalization ability, resulting in improved precision of anchor box localization for targets.

The primary contributions of our work are summarized as follows: Firstly, the backbone feature extraction network of YOLOv8s incorporates the newly developed C2f-Faster-EMA module, which combines the FasterNet module and EMA attention mechanisms. This innovation not only significantly reduces the parameter count and computational load but also enhances detection accuracy and speed. Secondly, in the neck network, the conventional PANet structure has been replaced with a BiFPN-P feature fusion module. This module improves the network's ability to extract features from shallow layers, thereby increasing the detection precision for small target defects in insulators. Lastly, the Inner-MPDIoU technique has been introduced to refine the loss function. This approach distinguishes among different regression samples and utilizes auxiliary bounding boxes of varying scales to compute the loss, which accelerates bounding box regression speed and boosts the model's generalization capability.

II. ORIGINAL YOLOv8s AND IMPROVED YOLOv8s

The YOLOv8 model is classified into five variants based on network depth and width: n, s, l, m, x. Each variant demonstrates an incremental increase in network size and detection accuracy, with a corresponding decrease in detection speed due to enhanced complexity. Given the computational constraints of drone platforms, this study employs the YOLOv8s model, which optimizes the balance between model size and detection accuracy, thereby fulfilling the demands for real-time detection.

The architecture of YOLOv8 comprises four primary components: input, backbone, neck, and head, as illustrated in Figure 1. The input layer employs Mosaic data augmentation, a technique that amalgamates four randomly cropped and scaled images into a composite image to enrich feature extraction from diverse perspectives and bolster the model's generalization ability. Additionally, YOLOv8 adopts

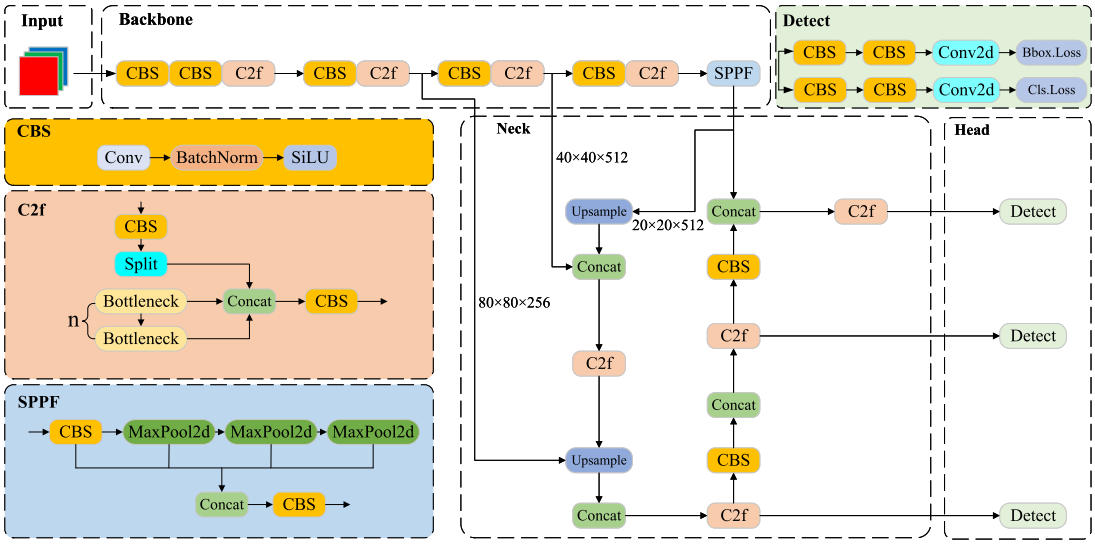

FIGURE 1. The structure of YOLOv8s network.

a strategy from YOLOX [32], where data augmentation is deactivated during the last ten training iterations to acclimate the model to real-world imaging characteristics.

The backbone network incorporates Conv, C2f, and SPPF modules, with the C2f module designed to learn residual features and enhance gradient flow across layers through augmented cross-layer connections. The neck network utilizes a Path Aggregation Network (PAN) structure to enhance feature integration from objects at various scales. The head network segregates classification and detection tasks, focusing on loss calculation and detection box selection. The classification branch employs Binary Cross-Entropy (BCE) for training, while the regression branch integrates Distribution Focal Loss (DFL) [33] and Complete Intersection over Union (CIoU) [34] loss functions, aiming to refine the precision of bounding box predictions.

To enhance the detection accuracy and inference speed of insulator defects, while simultaneously reducing the network's parameter count, this paper presents an improved network model based on YOLOv8s, specifically tailored for insulator defect detection. The proposed model, named Insulator-YOLOv8s, introduces three significant enhancements and optimizations to the YOLOv8s architecture:

1) Backbone Feature Extraction Network Improvement: The original C2f module has been re-engineered into the C2f-Faster-EMA module, which incorporates the FasterNet [35] leveraging PConv convolution. This modification not only accelerates the network's detection speed but also reduces the parameter count. Additionally, by integrating the EMA [36] attention mechanism with the FasterNet module, the network more effectively focuses on areas of interest, thereby enhancing detection accuracy.
2) Neck Network Modifications: The existing PANet in the YOLOv8s neck has been merged with the Bidirectional Feature Pyramid Network (BiFPN) [37] to create a new feature fusion module, BiFPN-P. This module downsamples shallow P2 feature maps from the backbone and merges them with two P3 layers from the neck, enhancing the extraction of shallow semantic information. This development not only increases the detection accuracy of small-target defects in insulators but also standardizes all channels of BiFPN-P to 256, reducing both computational load and model size.
3) Optimization of the Loss Function: The loss function in YOLOv8s has been refined by incorporating the Inner-IoU [38] concept into the MPDIoU function, resulting in the creation of the Inner-MPDIoU [39]. This replacement for the original CIoU loss function enhances the model's bounding box regression precision and improves detection accuracy for challenging samples.

The structure of the enhanced network model, Insulator-YOLOv8s, is depicted in Figure 2. The detection process and feature extraction of the proposed Insulator-YOLOv8s network model are as follows: First, the input insulator images are preprocessed and resized to 640 × 640 pixels, ensuring that all images maintain the same dimensions when entering the neural network. The preprocessed images then pass through a series of convolutional layers, pooling layers, and activation functions in the backbone network to extract multi-level features. Next, the neck network employs the BiFPN-P feature fusion module to combine the extracted features at different scales, enhancing the detection capabilities for both small and large targets. Finally, the detection head converts the extracted feature maps into target detection results. This sequence of steps ensures that YOLOv8 can efficiently extract multi-scale features from the input images and generate accurate target detection results.

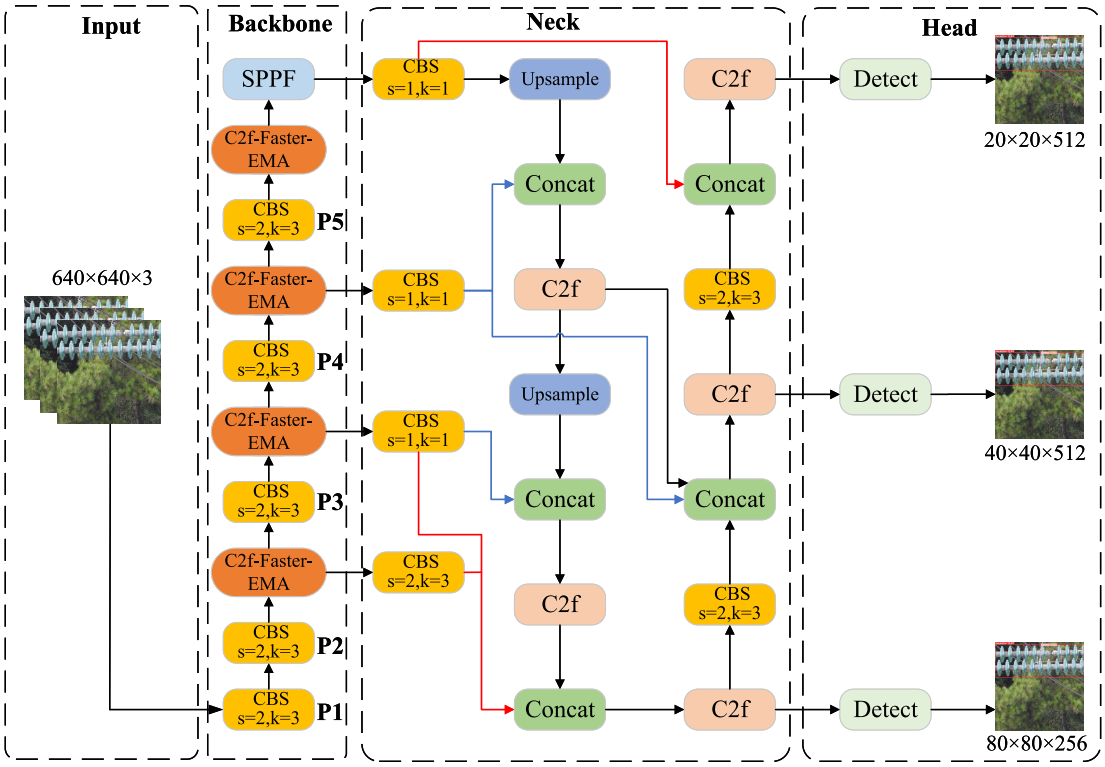

FIGURE 2. The structure of the enhanced Insulator-YOLOv8s network model.

III. METHOD

A. FasterNet AND PConv

In the feature extraction phase, CNNs repeatedly compute and extract semantic information from feature maps. As the network deepens, redundant information within these deeper feature maps increases, leading to potential computational wastage. To address the overhead of parameters and computations, modern lightweight networks often employ techniques such as Deepwise Separable Convolution (DWConv) or Grouped Convolution (GConv). Although these methods have achieved some success, they may increase the frequency of memory access, thereby slowing down the inference speed.

In response, Chen et al. have developed a novel architecture named FasterNet, utilizing a PConv convolution technique to reduce redundancy in calculations and memory usage, and to minimize repetitive feature information, thereby boosting network efficiency. PConv only needs to apply the filter to a part of the input channels for spatial feature extraction, and the rest of the channels remain unchanged. It calculates the first or last consecutive channel as a representative of the entire feature map. This approach allows either the first or the last contiguous set of channels to represent the entire feature map for computation. This selective processing contrasts with traditional Conv convolutions, which demand higher computational resources, as detailed in Equation (1). Assuming a uniform number of channels in both input and output feature maps, the computational load of PConv, where the channel ratio $r = c_p/c = 1/4$, implies that PConv's computational demand is only one-sixteenth that of conventional Conv, as shown in Equation (2).

$$F_c = h \times w \times k^2 \times c^2 \tag{1}$$

$$F_p = h \times w \times k^2 \times c_p^2 \tag{2}$$

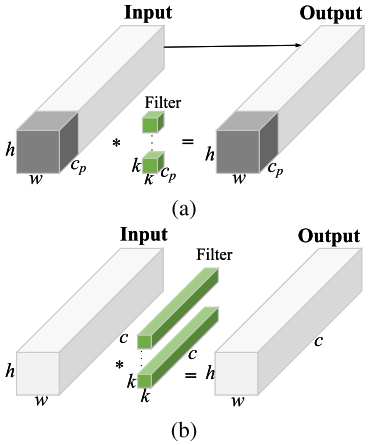

FIGURE 3. Comparison of PConv and Conv structure. (a) PConv module; (b) Conv module.

The comparison of PConv and conventional Conv structures is illustrated in Figure 3. In computational terms, where F_c and F_p represent the computational demands of Conv and PConv respectively, the parameters such as channel height h, channel width w, and the number of channels c for Conv and c_p for PConv, along with the filter size k, determine the overall computational load. This framework is designed to

substantially reduce the computational footprint, particularly in environments demanding high efficiency and speed.

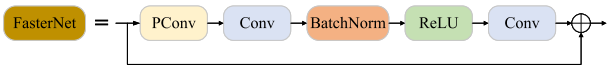

FIGURE 4. Network structure of FasterNet.

The FasterNet architecture prominently features the PConv convolution technique, which plays a crucial role in optimizing computational efficiency. Specifically, each FasterNet module integrates a 3×3 PConv layer flanked by two 1×1 Conv layers. Between these layers, batch normalization (BN) and the ReLU activation function are employed to enhance the network's non-linear expressiveness and improve its generalizability. To further refine network performance, FasterNet incorporates residual connections between the PConv and Conv layers. These connections facilitate the efficient flow of information and the propagation of gradients throughout the network. The detailed structure of the FasterNet module, including the arrangement of PConv and Conv layers along with their interconnections, is illustrated in Figure 4.

B. EMA ATTENTION MECHANISM

The Efficient Multi-Scale Attention (EMA) module is a cross-spatial learning mechanism designed to focus attention on areas of interest while preserving information across network channels. The architecture of the EMA module is depicted in Figure 5.

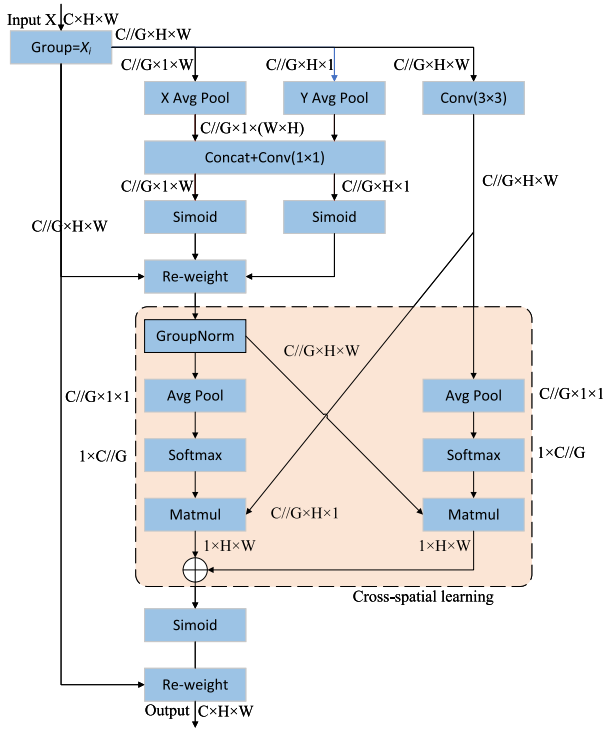

FIGURE 5. EMA attention mechanism.

The operation of the EMA module begins by dividing any given input feature map $X \in \mathbb{R}^{C\times H\times W}$ across the channel dimension into G sub-features, $X = [X_0, X_i, \ldots, X_{G-1}]$, where $X_i \in \mathbb{R}^{\frac{C}{G}\times H\times W}$, and $G = C$. Attention weights for these grouped feature maps are generated independently via 1×1 and 3×3 convolution branches.

In the 1×1 branch, channel information is first encoded through dual one-dimensional global average pooling operations. The resulting features are merged and processed through a 1×1 convolution to produce two parallel one-dimensional feature encoding vectors, each passed through a sigmoid activation function. A multiplication operation then aggregates the attention maps for each channel group, facilitating the interaction of features across different channels.

Conversely, the 3×3 branch captures local cross-channel interactive features by adding a 3×3 convolution layer, enhancing the spatial representation of features. This branch employs a two-dimensional global average pooling to encode global spatial information before jointly activating the channel features. The output of this smaller branch is adjusted to match the dimensions and shape required for activation. The process concludes with the application of a Softmax function on the linear transformations, generating the initial spatial attention map.

This detailed explanation of the EMA module highlights its dual strategy of integrating both global and local feature interactions, thus optimizing attentional focus and enhancing the overall representational capacity of the network.

C. C2f-FASTER-EMA MODULE

The Faster-EMA submodule redesigns the C2f module's Bottleneck structure using PConv convolution and the EMA attention mechanism, as shown in Figure 6. It starts by applying PConv convolution to a quarter of the input channels, leaving the rest unchanged. These channels are then concatenated to minimize redundancy.

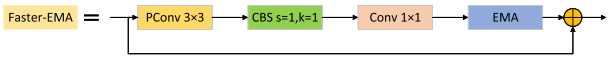

FIGURE 6. Faster-EMA module structure.

Next, the concatenated output is processed by the Channel Boosting and Shrinking (CBS) module, which doubles the number of channels to enhance feature diversity and reduce latency. The CBS module, which includes 1×1 convolutions, normalization, and activation functions, adjusts dimensions and maintains the input channel count. Subsequently, the EMA attention module is integrated to encode global information and improve cross-spatial feature aggregation, enhancing multi-scale feature extraction and inter-layer connectivity.

The C2f-Faster-EMA module, incorporating the Faster-EMA submodule, initially expands the output channels and then splits and concatenates them with multiple Faster-EMA submodules (e.g., $n = 3$), as depicted in Figure 7. This structure not only reduces parameters and computational load but also optimizes gradient flow. The final output feature map

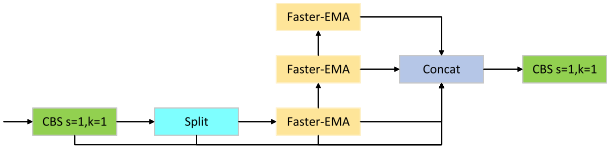

FIGURE 7. C2f-Faster-EMA module structure.

combines these elements to produce a channel count adjusted to c^2 using the CBS module.

D. BiFPN-P FEATURE FUSION MODULE

The YOLOv8 model's neck network employs the PANet architecture, leveraging semantic information from high-level feature maps and positional details from lower-level maps, as shown in Figure 8a. The Bidirectional Feature Pyramid Network (BiFPN) enhances this by introducing efficient bidirectional cross-scale connections and weighted fusion, depicted in Figure 8b. Inspired by BiFPN, this paper proposes the BiFPN-P module, illustrated in Figure 8c, designed to improve the information flow between feature maps of various scales, thus enhancing object detection performance.

BiFPN-P leverages skip-connection capabilities to preserve original feature map information effectively, particularly beneficial for detecting small objects like insulator defects using the larger-scale P2 feature map. The module optimizes the network architecture by maintaining bidirectional connections and streamlining the structure. It eliminates unnecessary branches while preserving P3, P4, and P5 channels as feature map outputs. The P2 channel is downsampled and concatenated with the P3 channel from the neck network, enhancing feature richness. To reduce computational demand, all channels in BiFPN-P are uniformly standardized, simplifying the structure without compromising on feature fusion capabilities. This module employs a weighted fusion method for integrating features, detailed in the subsequent formula (3).

$$O = \sum_i \frac{\omega}{\epsilon + \sum_j \omega_j} I_i \tag{3}$$

where ω represents the weight of the input feature, $\epsilon = 0.0001$ to prevent numerical instability, and I_i represents the eigenvalue of the channel input. Weighted feature fusion is a process where a weight is added to each input feature to balance the features of different channels, allowing for quick normalization of the features.

E. LOSS FUNCTION IMPROVEMENT

The design of the loss function is critical to the performance of an object detection model, particularly the bounding box loss function, which is integral to refining the accuracy of detections. Bounding box regression aims to adjust the predicted window of the detector to more closely align with the actual object's location. The Intersection over Union (IoU) metric, as detailed in Equation (4), has become a widely adopted standard for evaluating the accuracy of predicted

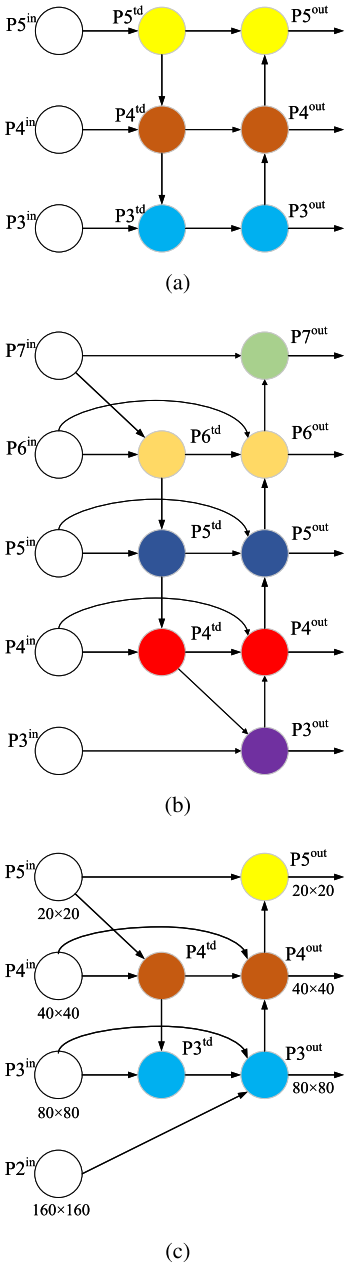

FIGURE 8. PANet, BiFPN and BiPFN-P network structures. (a) PANet, (b) BiFPN, (c) BiPFN-P.

bounding boxes in the field of object detection.

$$\text{IoU} = \frac{|B \cap B^{\text{gt}}|}{|B \cup B^{\text{gt}}|} \tag{4}$$

where, B and $B^g t$ are the center points of the predicted box and the real box respectively.

Over the years, the IoU has undergone continual iterations and updates, resulting in the emergence of many novel bounding box regression loss functions. The bounding box regression loss function used in YOLOv8 employs the CIoU loss. The computational formulas for CIoU are presented in

equations (5), (6), and (7).

$$\mathrm{CIoU} = \mathrm{IoU} - \left(\frac{\rho^2(B^{\mathrm{pred}}, B^{\mathrm{gt}})}{c^2} + \alpha v\right) \tag{5}$$

$$v = \frac{4}{\pi}\left(\arctan\frac{w^{\mathrm{gt}}}{h^{\mathrm{gt}}} - \arctan\frac{w}{h}\right) \tag{6}$$

$$\alpha = \frac{v}{1-\mathrm{IoU}} + v \tag{7}$$

where, ρ is the Euclard distance between two central points, v is a measure of the aspect ratio parameter, c is the diagonal distance of the smallest closure area that can contain both the predicted box and the real box, w^{gt} and hgt represent the width and height of the real box, w and h represent the width and height of the predicted box, and α is the orthogonal balance parameter.

CIoU accounts for the aspect ratio differences between predicted and actual bounding boxes, it does not address the issue of low-quality samples in the dataset. Low-quality samples refer to anchors that have a poor match with actual target objects. These low-quality anchors exhibit the following characteristics: (1) Low IoU values between the anchor and the ground truth bounding box, resulting in inadequate coverage of the target object. (2) Significant differences in size or shape between the anchor and the target object, leading to poor matching with the target object. (3) Considerable positional offset of the anchor from the target object, leading to decreased detection accuracy. In response to this, Inner-IoU introduces the concept of using auxiliary bounding boxes to calculate IoU, thereby enhancing the model's generalizability. The specific calculation process is as follows.

$$b_l^{gt} = x_c^{gt} - \frac{w^{gt}\times \mathrm{ratio}}{2}, \quad b_r^{gt} = x_c^{gt} + \frac{w^{gt}\times \mathrm{ratio}}{2} \tag{8}$$

$$b_t^{gt} = y_c^{gt} - \frac{h^{gt}\times \mathrm{ratio}}{2}, \quad b_b^{gt} = y_c^{gt} + \frac{h^{gt}\times \mathrm{ratio}}{2} \tag{9}$$

$$b_l = x_c - \frac{w\times \mathrm{ratio}}{2}, \quad b_r = x_c + \frac{w\times \mathrm{ratio}}{2} \tag{10}$$

$$b_t = y_c - \frac{h\times \mathrm{ratio}}{2}, \quad b_b = y_c + \frac{h\times \mathrm{ratio}}{2} \tag{11}$$

$$\begin{aligned}\mathrm{inter} = &\left(\min(b_r^{gt}, b_r) - \max(b_l^{gt}, b_l)\right)\\ &\times\left(\min(b_b^{gt}, b_b) - \max(b_t^{gt}, b_t)\right)\end{aligned} \tag{12}$$

$$\mathrm{union} = w^{gt}\cdot h^{gt}\cdot(\mathrm{ratio})^2 \cdot w\cdot h\cdot(\mathrm{ratio})^2 - \mathrm{inter} \tag{13}$$

$$\mathrm{IoU}^{\mathrm{inter}} = \frac{\mathrm{inter}}{\mathrm{union}} \tag{14}$$

where, x_c^{gt} and y_c^{gt} are the horizontal and vertical coordinates of the center point of the real box, x_c and y_c are the horizontal and vertical coordinates of the center point of the prediction box, ratio is the scale factor, and the value range of ratio is [0.5, 1.5].

Inner-IoU quantifies the Intersection over Union (IoU) between auxiliary bounding boxes, enhancing the bounding box regression process. When the ratio is less than one, the auxiliary bounding box is smaller than the actual bounding box. This reduction in size limits the regression range but increases the gradient's absolute value, facilitating faster convergence for samples with high IoU values. Conversely, a ratio greater than one indicates that the auxiliary bounding box is larger than the actual one, which broadens the effective regression range, benefiting samples with low IoU values.

MPDIoU improves upon traditional methods by minimizing the distance between corresponding points at the top-left and bottom-right corners of the predicted and ground truth bounding boxes. This approach is particularly effective for both overlapping and non-overlapping scenarios, enhancing convergence rates as mathematically demonstrated in Equation (15).

$$\mathrm{MPDIoU} = \mathrm{IoU} - \left(\frac{\rho^2(P_1^{\mathrm{pred}}, P_1^{\mathrm{gt}})}{w^2+h^2} - \frac{\rho^2(_2^{\mathrm{pred}}, P_2^{\mathrm{gt}})}{w^2+h^2}\right) \tag{15}$$

where, P_1^{pred}, P_1^{gt}, P_1^{gt}, P_2^{gt} refers to the upper left corner and lower right corner of the prediction box and the ground truth box respectively, and $\rho^2(_2^{\mathrm{pred}}, P_2^{\mathrm{gt}})$ refers to the distance between the corresponding points.

Utilizing the design concept of Inner-IoU, we applied the Inner-IoU loss to the boundary box regression loss function of MPDIoU, thereby designing a new Inner-MPDIoU loss function. The expressions of MPDIoU, Inner-IoU and Inner-MPDIoU loss functions are shown in equations (16), (17) and (18), respectively.

$$LMPDIoU = 1 - \mathrm{MPDIoU} \tag{16}$$

$$L_{\text{Inner-IoU}} = 1 - \mathrm{IoU}^{\mathrm{inner}} \tag{17}$$

$$L_{\text{Inner-MPDIoU}} = L_{\mathrm{MPDIoU}} + \mathrm{IoU} - \mathrm{IoU}^{\mathrm{inner}} \tag{18}$$

IV. EXPERIMENT AND ANALYSIS

A. EXPERIMENTAL ENVIRONMENT AND DATA COLLECTION

The experimental setup for this study was configured as follows: The software environment utilized the PyTorch framework version 2.0 and Python version 3.8 on an Ubuntu 20.04 operating system. The hardware setup included a CPU with 16 vCPUs (Intel Xeon Platinum 8350C @ 2.60GHz) and 42 GB of RAM. The GPU was an NVIDIA GeForce RTX A5000 with 24 GB of memory, using CUDA version 11.7 to accelerate model training. Key parameters were an input image size of 640×640 pixels, a batch size of 16, and an initial learning rate of 0.01. The Stochastic Gradient Descent (SGD) method with a momentum of 0.937 was used for updating network parameters over 200 epochs, with data augmentation disabled in the final 10 epochs.

The dataset used in this study is sourced from the Chinese Power Line Insulator Dataset (CPLID), defect images of glass insulators from Problem B of the 8th "Teddy Cup" competition, and insulator defect images obtained from the internet. The CPLID dataset includes 600 images of

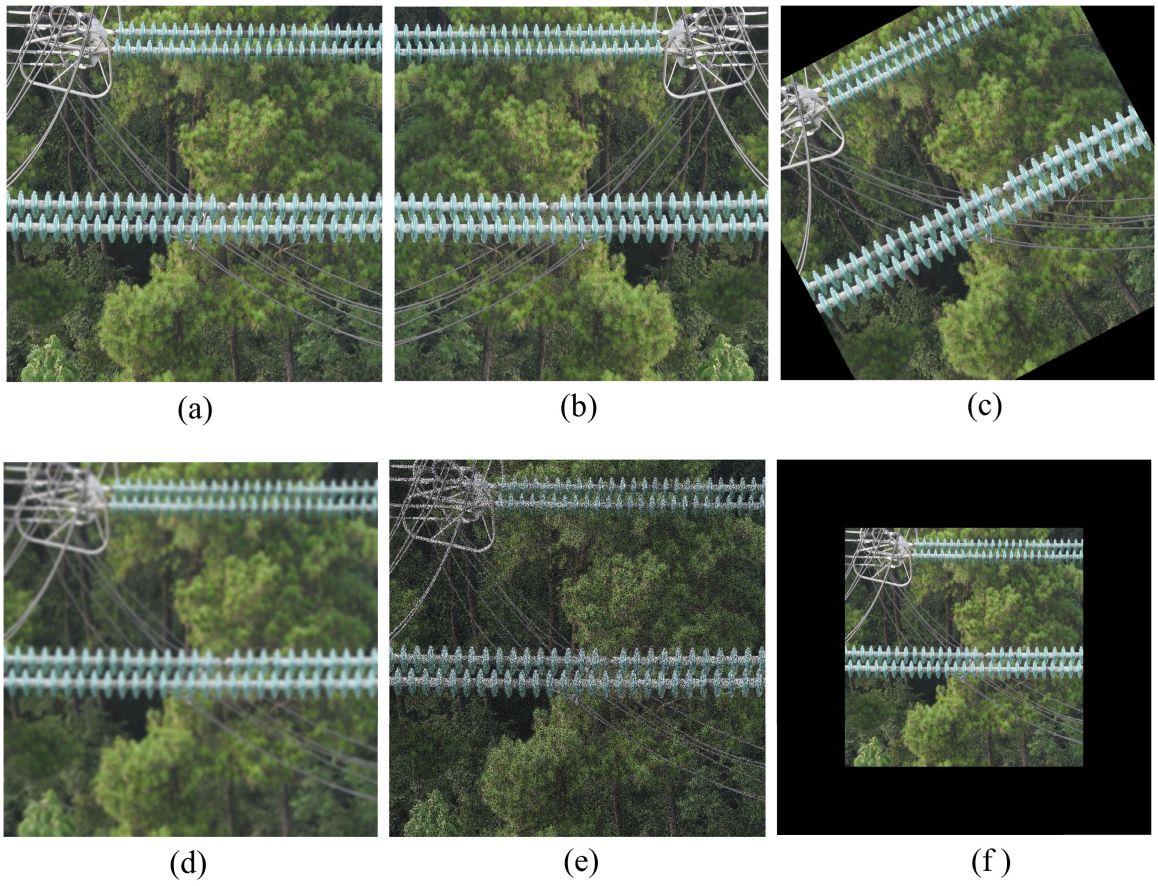

FIGURE 9. Comparison of effects before and after improvement. (a) Detection effects of YOLOv8s, (b) Detection effects of Insulator-YOLOv8s.

normal insulators and 248 images of defective insulators, with a resolution of 1152×864. The "Teddy Cup" Problem B dataset consists of 40 high-resolution insulator images. For this study, these high-resolution images were processed by dividing them into 4×4 grids, and background images without insulators were removed, resulting in 254 usable glass insulator images. Due to the relatively small number of defect images in the insulator datasets, data augmentation was performed. The augmentation techniques included horizontal flipping, rotation, and Gaussian blur, as illustrated in Figure 9. After these augmentations, the dataset expanded to a total of 7,178 images, comprising 4,258 images of normal insulators and 2,920 images of defective insulators. This significantly enhanced the model's robustness in detecting insulator defects under complex conditions. All augmented images were annotated using the LabelImg tool, with annotations categorized as either 'insulator' or 'defect'. The generated annotation files were then saved in a dedicated folder for labels.

For model training and evaluation, the dataset was divided into training, validation, and test sets in a 6:2:2 ratio. The training and validation sets were used for model parameter adjustments, while the test set assessed the model's performance in real-world applications.

B. EVALUATION METRICS

To evaluate the detection performance of different network models, this study utilizes several metrics: Precision (P), Recall (R), mean Average Precision (mAP), parameters, frames per second (FPS), and floating point operations (FLOPs). Precision and Recall are calculated as shown in formulas (19) and (20), respectively. Average Precision (AP) is computed by averaging precision across various recall levels, with higher AP values indicating better average model accuracy, as detailed in formula (21). The mean Average Precision (mAP) is the average AP across all categories, calculated as shown in formula (22). Additionally, mAP@0.5 refers to the mAP calculated for each category at an IoU threshold of 0.5.

$$R = \frac{T_P}{T_P + F_N} \times 100\% \tag{19}$$

$$P = \frac{T_P}{T_P + F_P} \times 100\% \tag{20}$$

$$\mathrm{AP} = \int_0^1 P(R)\, dR \tag{21}$$

$$\mathrm{mAP} = \frac{\sum_{i=1}^{N} \mathrm{AP}_i}{N} \tag{22}$$

where, T_P represents the number of positive samples predicted, F_P represents the number of positive samples wrongly predicted, F_N epresents the number of negative samples wrongly predicted, and N represents the number of categories in the data set.

C. MODEL TRAINING PROCESS

To optimize the training outcomes of the improved YOLOv8s algorithm, it is essential to train the model until the loss curve converges. Under consistent experimental conditions, the insulator defect detection model is trained using training set and verification set. Then, the test set is used to evaluate the overall performance of the model. As illustrated in Figure 10, the loss curve showed a progressively flattening decline with increasing training epochs. By the 150th epoch, the loss fluctuations substantially reduced, and the total loss decreased to below 1.5, indicating convergence and optimal performance.

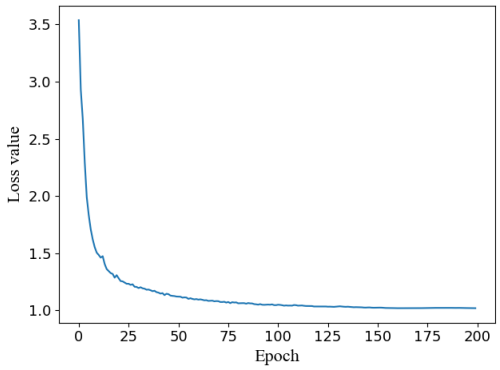

FIGURE 10. Loss value curve.

The training results, presented in Table 1, revealed detection accuracies of 92.3% for intact insulators and 96.4% for defective insulators. Recall rates were 81.0% and 91.9%, respectively, with overall accuracy and recall rates of 94.4% and 86.5%, and an mAP@0.5 of 91.5%. These results demonstrate that the trained model meets practical requirements for detecting insulator defects in UAV imagery.

TABLE 1. Results of model training.

Labels	Precision(%)	Recall(%)	mAP@0.5(%)
insulator	92.3	81.0	87.8
defect	96.4	91.9	95.3
all	94.4	86.5	91.5

D. ABLATION EXPERIMENT

To validate the enhanced detection performance of the modified Insulator-YOLOv8s model, this study sequentially incorporated three improvements into the base YOLOv8s model: the C2f-Faster-EMA lightweight module, the BiFPN-P feature fusion module, and the Inner-MPDIoU loss function. Different modules were added one by one to conduct ablation experiments under consistent experimental conditions, with results presented in Table 2.

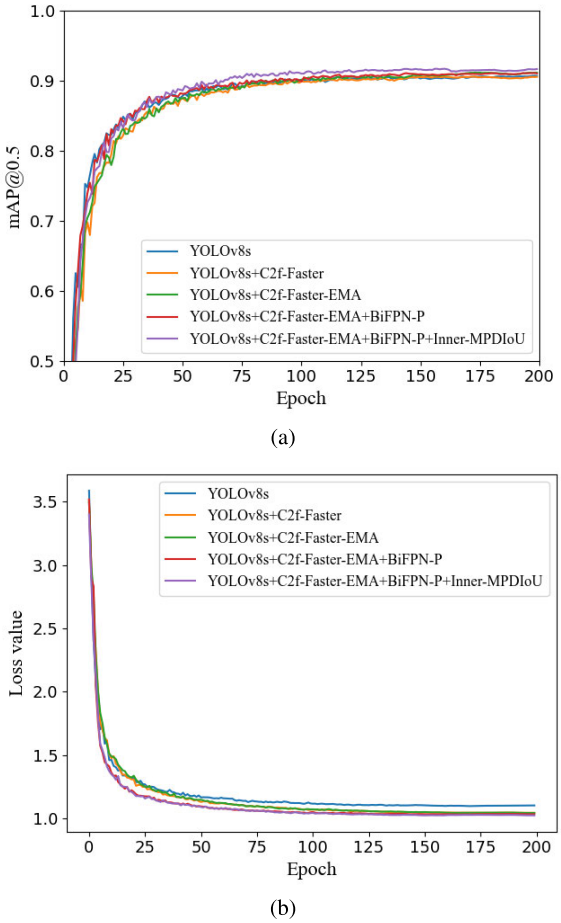

FIGURE 11. Comparison curve of ablation experimental results. (a) The change in curve in mAP@0.5, (b) The change in curve in loss value.

The ablation study reveals the following:

- Group 2: Substituting C2f with the C2f-Faster module resulted in a minor mAP@0.5 decrease of 0.2%, but parameters, FLOPs, and model size were reduced by 25.4%, 24.6%, and 24.9%, respectively, indicating a reduction in network complexity.
- Group 3: Introducing the EMA attention mechanism to form the C2f-Faster-EMA module led to a significant mAP@0.5 increase of 0.7% compared to Group 2, without a substantial increase in parameters or FLOPs, demonstrating enhanced detection accuracy.
- Group 4: Replacing the PANet with the BiFPN-P feature fusion module, setting all channels to 256 and incorporating P2 shallow feature maps, resulted in a slight mAP@0.5 decrease of 0.2% compared to Group 3, but network parameters and model size decreased by 28.7% and 27.8%, respectively, indicating effective model optimization.
- Group 5: Optimizing the CIoU loss with Inner-MPDIoU improved mAP@0.5 by 0.5% compared to Group 4, without altering model size and FLOPs, demonstrating higher localization accuracy.

These results illustrate that the sequential incorporation of these modules and optimizations effectively reduces model

TABLE 2. Results of model training.

Group	Model	mAP@0.5(%)	Parameters(M)	FLOPs(G)	Model Size (M)	FPS(f/s)
1	YOLOv8s	90.7	10.61	28.4	22.5	105
2	YOLOv8s + C2f-Faster	90.5	7.92	21.4	16.9	115
3	YOLOv8s + C2f-Faster-EMA	91.2	7.94	21.8	16.9	117
4	YOLOv8s + C2f-Faster-EMA + BiFPN-P	91.0	5.66	21.1	12.2	113
5	YOLOv8s + C2f-Faster-EMA + BiFPN-P +Inner-MPDIoU	91.5	5.66	21.1	12.2	113

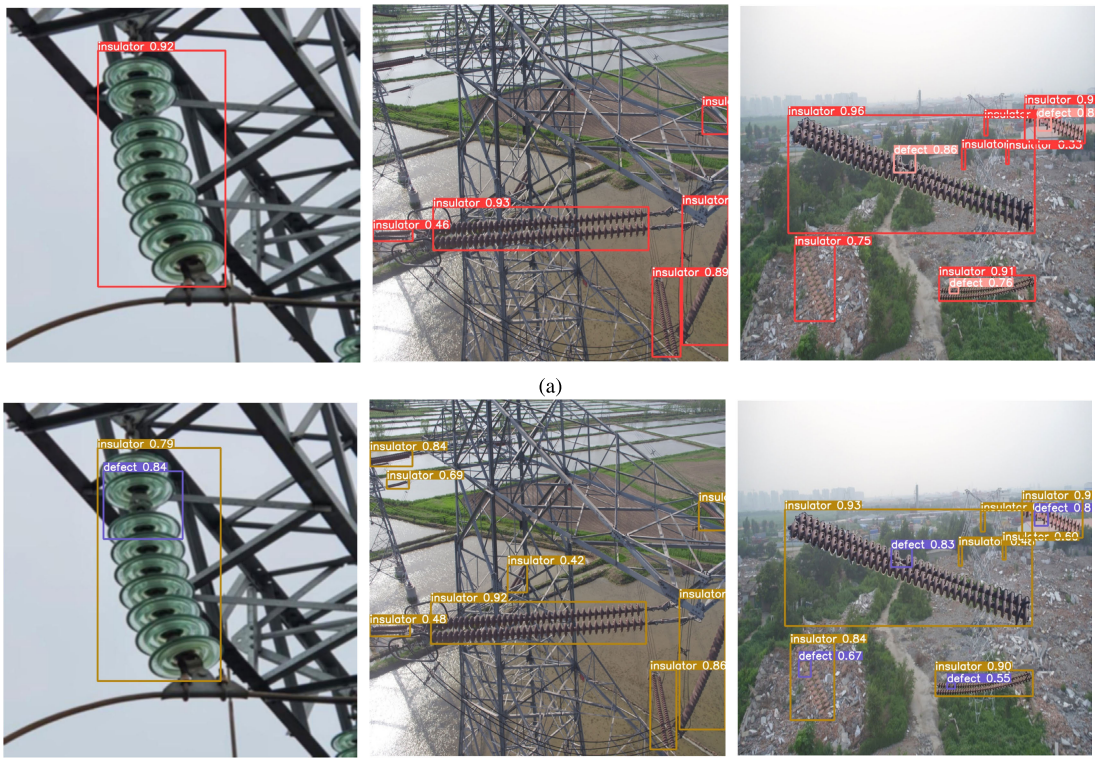

FIGURE 12. Comparison of effects before and after improvement. (a) Detection effects of YOLOv8s, (b) Detection effects of Insulator-YOLOv8s.

complexity while enhancing detection accuracy and performance. Firstly, the C2f-Faster-EMA module leverages PConv convolution and the EMA attention mechanism. PConv convolution improves feature processing efficiency by reducing redundant information during computation, while the EMA attention mechanism focuses attention on areas of interest to enhance detection accuracy. Consequently, this module significantly reduces network complexity while improving the mAP@0.5 metric. Secondly, the BiFPN-P feature fusion module, with its skip connection capability, effectively preserves the initial information in the feature maps and achieves higher-level feature fusion. This reduces the missed detection rate of small insulator targets and decreases the model's parameter count. Lastly, the Inner-MPDIoU loss function distinguishes different regression samples and uses auxiliary bounding boxes of various scales to compute the loss, thereby enhancing the localization accuracy of anchor boxes for detection targets and further improving the accuracy of insulator defect detection.

The results from the ablation studies demonstrate that the proposed improvements significantly enhance the model's performance, increasing accuracy and inference speed in detecting insulator defects. Compared to the baseline YOLOv8s, the enhanced model shows an increase in mAP@0.5 by 0.8%, a reduction in parameters by 46.7%, FLOPs by 25.7%, and model size by 45.8%, along with an increase in FPS by 8 frames. These enhancements ensure that the improved model meets real-time detection requirements for aerial images of insulators.

The comparative curves from the ablation studies are illustrated in Figure 11. Figure 11a shows that each improved method is gradually improved in mAP@0.5 metric, which proves that each module adopted can optimize the model. Figure 11b indicates that the proposed methods achieve the lowest loss values and fastest convergence rates, which verifies that the improved model has higher localization accuracy for detection targets.

E. COMPARISON EXPERIMENT OF DIFFERENT DETECTION MODELS

To validate the enhanced performance of the algorithm improved in this study, several mainstream and classic network models were selected for comparative experiments. The specific results are presented in Table 3. Models denoted

TABLE 3. Results of model training.

Model	Precision(%)	Recall(%)	mAP@0.5(%)	Parameters(M)	FLOPs(G)	Model Size(MB)
Faster R-CNN	94.8	87.8	92.3	136.7	401.7	167.2
SSD	92.5	82.1	87.9	62.7	26.3	94.6
YOLOv5s	95.4	83.9	89.5	7.02	15.8	13.8
YOLOv7-tiny	94.3	84.9	90.1	5.73	13.0	11.7
YOLOv8s	94.5	85.2	90.7	10.61	28.4	22.5
YOLOv8s+MobileNetv3	91.9	81.3	87.3	6.03	15.7	13.0
YOLOv8s+ GhostNet	93.4	82.9	88.7	10.64	19.5	15.8
YOLO-S [25]	95.7	84.5	89.9	6.52	14.9	13.1
IDD-YOLOv7 [27]	94.8	85.3	90.5	30.5	90.3	58.7
Insulator-YOLOv8s	94.4	86.5	91.5	5.66	21.1	12.2

as YOLOv8s+MobileNetv3 and YOLOv8s+GhostNet use MobileNetv3 and GhostNet [40] as the backbone for YOLOv8s, respectively. Compared to SSD, YOLOv5s, YOLOv7-tiny, YOLOv8s, YOLOv8s+MobileNetv3, and YOLOv8s+GhostNet, the improved algorithm demonstrated increases in mAP@0.5 by 3.6%, 2.0%, 1.4%, 0.8%, and 4.2%, respectively, and exhibited higher recall rates. In addition, compared with the insulator defect detection methods mentioned in literature [25] and [27], our algorithm not only has higher detection accuracy, but also has lower parameters and FLOPs. Although our algorithm slightly underperforms Faster R-CNN in mAP@0.5, it requires significantly fewer parameters and FLOPs, making it advantageous for devices with limited memory. These results affirm that the improved algorithm achieves the balance of detection accuracy and lightweight of the model. This is particularly beneficial for real-time detection applications such as drone-based surveillance of insulators, where it shows considerable potential and practical value.

To further demonstrate the detection capabilities of the Insulator-YOLOv8s model, three images of insulators against varied complex backgrounds from the test dataset were selected for evaluation. The results are depicted in Figure 12, with sub-figure (a) illustrating detection outcomes using the YOLOv8s model and sub-figure (b) showing results with the Insulator-YOLOv8s model. During experiments, 'insulator' anchor boxes determined the boundaries of insulators, while 'defect' anchor boxes identified defect boundaries. In sub-figure (a), the first and third images missed two small defect targets, and the second image failed to detect a distant, blurred insulator. Conversely, sub-figure (b) shows that the Insulator-YOLOv8s model effectively addresses these detection issues. Therefore, the proposed insulator defect detection model outperforms the original model.

V. CONCLUSION

To address defect detection in insulators, this study introduces a lightweight detection network model based on Insulator-YOLOv8s. This model effectively resolves issues of complex network structures and slow detection speeds in existing algorithms. Comparative and ablation experiments reveal that replacing the C2f module with the lightweight C2f-Faster-EMA module in the backbone significantly reduces the model's parameter count and computational load, while enhancing detection precision and inference speed. Additionally, the BiFPN-P feature fusion module can effectively extract the shallow layer information of insulator string and defects, and it has a good detection effect on small targets and fuzzy targets. Finally, Inner-MPDIoU, as the boundary box regression function of the model, strengthens the model's learning ability of difficult samples, and further improves the accuracy of insulator defect location.

Experimental data indicate that, compared to the YOLOv8s model, the Insulator-YOLOv8s network reduces parameter count by 46.7%, computational demand by 25.7%, and model size by 45.8%, while improving mAP@0.5 by 0.8% and increasing FPS by 8 frames. Therefore, the improved method can effectively reduce the model complexity and improve the detection accuracy and speed. The improved model can realize the real-time detection of insulator defect image by UAV.

Insulator defect detection tasks currently face numerous challenges. Existing methods often fail to detect insulators in images with particularly complex backgrounds or occluded targets. Additionally, there is a scarcity of publicly available insulator defect images. The dataset used in this paper was primarily expanded through image augmentation techniques, lacking a sufficient variety of defect samples for comprehensive model training and testing. Therefore, one of the future research goals is to introduce various background interferences or use advanced image augmentation techniques to expand the dataset, enabling the model to better adapt to different complex background situations. Moreover, future research could focus on applying techniques such as knowledge distillation or model pruning to make the model more lightweight, thereby further enhancing its detection speed and efficiency.

REFERENCES

[1] K. Li, Z. Yuan, and Y. Li, "Key technology of UAV power line safety inspection," *China High New Technol.*, vol. 5, pp. 48–59, Sep. 2022.

[2] X. Luo, F. Yu, and Y. Peng, "UAV power grid inspection defect detection based on deep learning," *Pow. Syst. Prot. Control*, vol. 50, no. 10, pp. 132–139, 2021.

[3] J. Xuechen, W. Dazhi, N. Yi, and Z. Cuiling, "Query method for optimal diagnosis of power system faults," *High Voltage Eng.*, vol. 43, no. 4, pp. 1311–1316, 2017.

[4] Y. Zhong, R. Hu, Z. Li, and Y. Cai, "Insulator defect detection based on YOLOv4-tiny with improved feature fusion," in *Proc. 12th Int. Conf. Inf. Technol. Med. Educ. (ITME)*, Nov. 2022, pp. 273–277.

[5] V. N. Nguyen, R. Jenssen, and D. Roverso, "Automatic autonomous vision-based power line inspection: A review of current status and the potential role of deep learning," *Int. J. Electr. Power Energy Syst.*, vol. 99, pp. 107–120, Jul. 2018.

[6] M. Chen, Y. Tian, S. Xing, Z. Li, E. Li, Z. Liang, and R. Guo, "Environment perception technologies for power transmission line inspection robots," *J. Sensors*, vol. 2021, pp. 1–16, Mar. 2021.
[7] K. Sharifani and M. Amini, "Machine learning and deep learning: A review of methods and applications," *World Inf. Technol. Eng. J.*, vol. 1, pp. 3897–3904, Oct. 2023.
[8] Y. Liu, Z. Yan, T. Ye, A. Wu, and Y. Li, "Single nighttime image dehazing based on unified variational decomposition model and multi-scale contrast enhancement," *Eng. Appl. Artif. Intell.*, vol. 116, Nov. 2022, Art. no. 105373.
[9] S. Zhang, W. Ren, X. Tan, Z.-J. Wang, Y. Liu, J. Zhang, X. Zhang, and X. Cao, "Semantic-aware dehazing network with adaptive feature fusion," *IEEE Trans. Cybern.*, vol. 53, no. 1, pp. 454–467, Jan. 2023.
[10] S. Zhang, X. Zhang, S. Wan, W. Ren, L. Zhao, and L. Shen, "Generative adversarial and self-supervised dehazing network," *IEEE Trans. Ind. Informat.*, vol. 20, no. 3, pp. 4187–4197, Mar. 2024.
[11] J. Gou, S. Du, and L. Liu, "Self-explosion detection of transmission line insulators based on improved masked area convolutional neural network," *J. Electr. Eng. Technol.*, vol. 38, no. 1, pp. 47–59, 2023.
[12] Z. Wang, Y. Wang, Q. Wang, S. Kang, and V. I. Mikulovich, "Two stage insulator fault detection method based on collaborative deep learning," *Trans. China Electrotechnical Soc.*, vol. 36, pp. 3594–3604, Sep. 2021.
[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.*, 2015, pp. 1440–1448.
[15] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 1–12.
[16] Y. Shan, T. Zhao, Y. Fu, H. Fu, and K. Wang, "Multi-target detection of transmission lines based on four-layer R-CNN," *J. Liaoning Tech. Univ.*, vol. 41, pp. 270–276, Nov. 2022.
[17] S. Ghashghaei and M. Akhbari, "Fault detection and classification of an HVDC transmission line using a heterogenous multi-machine learning algorithm," *IET Gener., Transmiss. Distribution*, vol. 15, no. 16, pp. 2319–2332, Aug. 2021.
[18] 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.*, Oct. 2016, pp. 21–37.
[19] 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)*, Jun. 2016, pp. 779–788.
[20] C. Wang, W. Li, W. Zheng, N. Wang, and H. Zhao, "An improved SSD method for power component identification of transmission lines," *Appl. Sci. Technol.*, vol. 47, pp. 75–81, Aug. 2020.
[21] D. Zhao and X. Yao, "Contact network insulator identification in complex background based on lightweight convolutional networks," *Electroceramic Lightning Arrester*, vol. 1, pp. 172–178, Jul. 2022.
[22] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.
[23] W. Zheng, X. Yang, L. Zhongbin, C. Ren, H. Wu, and C. Wang, "Real-time inspection model for key components of transmission lines based on improved YOLOv4," *Sci. Technol. Eng.*, vol. 21, pp. 10393–10400, 2021.
[24] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.
[25] J. Chen, Z. Fu, X. Cheng, and F. Wang, "An method for power lines insulator defect detection with attention feedback and double spatial pyramid," *Electr. Power Syst. Res.*, vol. 218, May 2023, Art. no. 109175.
[26] (2022). *Ultralytics/YOLOv5: V7.0—YOLOv5*. Accessed: May, 7, 2023. [Online]. Available: https://github.com/ultralytics/yolov5.com
[27] W. Yi, S. Ma, and R. Li, "Insulator and defect detection model based on improved YOLO-S," *IEEE Access*, vol. 11, pp. 93215–93226, 2023.
[28] Y. Zhai, X. Zhao, L. Wang, Y. Wang, X. Song, and H. Zhu, "IDD-YOLOv7: A lightweight method for multiple defect detection of 502 insulators in transmission lines," *J. Graph.*, vol. 42, pp. 90–101, Jun. 2024.
[29] C.-Y. Wang, A. Bochkovskiy, and H.-Y.-M. 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.
[30] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 13708–13717.
[31] A. Aboah, B. Wang, U. Bagci, and Y. Adu-Gyamfi, "Real-time multi-class helmet violation detection using few-shot data sampling technique and YOLOv8," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2023, pp. 5349–5357.
[32] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO series in 2021," 2021, *arXiv:2107.08430*.
[33] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance-IoU loss: Faster and better learning for bounding box regression," in *Proc. AAAI Conf. Artif. Intell.*, Apr. 2020, pp. 12993–13000.
[34] X. Li, W. Wang, L. Wu, S. Chen, X. Hu, J. Li, J. Tang, and J. Yang, "Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, pp. 21002–21012.
[35] J. Chen, S.-H. Kao, H. He, W. Zhuo, S. Wen, C.-H. Lee, and S.-H.-G. Chan, "Run, don't walk: Chasing higher FLOPS for faster neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 12021–12031.
[36] D. Ouyang, S. He, G. Zhang, M. Luo, H. Guo, J. Zhan, and Z. Huang, "Efficient multi-scale attention module with cross-spatial learning," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2023, pp. 1–5.
[37] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and efficient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 10778–10787.
[38] H. Zhang, C. Xu, and S. Zhang, "Inner-IoU: More effective intersection over union loss with auxiliary bounding box," 2023, *arXiv:2311.02877*.
[39] S. Ma and Y. Xu, "MPDIoU: A loss for efficient and accurate bounding box regression," 2023, *arXiv:2307.07662*.
[40] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "GhostNet: More features from cheap operations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1577–1586.

ZHONGSHENG LI was born in 1982. He received the master's degree from Sichuan University, in 2017. He is currently an Associate Professor and a Senior Engineer with the School of Traffic Engineering, Fujian Polytechnic of Water Conservancy and Electric Power, China. His research interests include electrical equipment testing, electrical engineering, and automation.

CHENDA JIANG was born in 1983. He received the master's degree from Fuzhou University, in 2012. He is currently an Associate Professor and a registered Electrical Engineer with Fujian Polytechnic of Water Conservancy and Electric Power, China. His research interests include power system stability, electrical engineering, and automation.

ZHONGLIANG LI was born in 1984. He received the master's degree from Fuzhou University, in 2017. He is currently a Teacher and a Senior Engineer with the School of Traffic Engineering, Fujian Polytechnic of Water Conservancy and Electric Power. His research interests include electrical engineering, industrial automation, and track power supply.