

Smart Grid Insulator Detection Network Improved based on YOLOv8

Tao Wang , Nan Zhang , Wenqing Yang , Wei Zhang , and Wancai Zhang 

Abstract—Insulators are critical components of power transmission lines. Due to environmental changes, insulators may fail, making timely and effective detection of these failures a pressing issue. However, the detection of inclined insulators faces challenges, such as inadequate fitting of detection frames and excessive background noise within the target frames. To address this, this paper proposes an improved inclined insulator detection network (RCAS-YOLOv8). To resolve issues related to feature sparsity and effectiveness, a non-local module with row and column-level sharing is introduced by considering the correlations between feature points. Finally, the task of locating the four vertices of the insulator is completed by summing the predicted offsets of the target frame's four vertices. Experimental results show that the proposed RCAS-YOLOv8 algorithm has achieved significant improvement in the detection of tilted targets in the Power Line Insulator Dataset (CPLID), with high detection accuracy, in which the APR index of our method reached 0.891.

Link to graphical and video abstracts, and to code:
<https://latam.ieceer9.org/index.php/transactions/article/view/9299>

Index Terms—Insulator, Object Detection, Prediction offset.

I. INTRODUCTION

INSULATORS are specialized devices capable of withstanding voltage and mechanical stress, widely used in overhead transmission lines. These insulators are typically exposed to external air, making them vulnerable to failures due to environmental factors, which pose significant safety risks to power line operations. Historically, insulator inspections were conducted manually through observation and testing. However, as the demand for electricity has grown and the coverage of high-voltage transmission lines has expanded, this method has proven to be time-consuming, labor-intensive, and heavily reliant on the inspector's expertise. Due to its inefficiency, this manual inspection method has largely been abandoned in the power sector.

With the advancement and integration of UAV (Unmanned Aerial Vehicle) technology and computer vision algorithms, this approach has seen widespread application in the inspection of transmission lines. This method not only enhances inspection efficiency but also reduces costs and risks associated with line inspections. Consequently, combining UAV

technology with computer vision for effective detection and defect analysis of insulators has become increasingly crucial for ensuring the reliable operation of power systems.

Upon reviewing previous research, we can categorize the tools used into two main types: traditional insulator detection image techniques and neural network-based detection techniques. Traditional methods primarily rely on manual extraction of insulator attributes, distinguishing the background from the target using common features such as color, shape, position, and texture. Ding et al. [1] proposed an image enhancement method for tempered glass insulators under the HIS mode, based on techniques like image morphology and threshold segmentation. In [2], a semi-local operator was used to extract insulator texture features, overcoming the issue of uneven texture in images. This work introduced a non-convex model within a globally minimized active framework, enabling the extraction of insulator contours from aerial images. Tan et al. [3] proposed a fusion algorithm combining image contour features and grayscale similarity matching to classify and identify defective and normal insulators effectively. Tomaszewski [4] et al. proposed a new method for detecting power line insulator faults by analyzing the transformed color intensity distribution in images captured by inspection cameras. This method enhances the detection of subtle defects such as cracks and corrosion by converting the raw color data into a more informative distribution and then analyzing it using a machine learning algorithm. While these traditional algorithms perform well in specific scenarios, altering the application context or image style can significantly impact their performance, indicating poor generalization ability.

Over the past decade, with the maturation of convolutional neural networks, researchers have increasingly applied deep learning algorithms in the field of computer science, achieving notable success. Some researchers have classified insulator detection under the domain of object detection, modifying general object detection algorithms to suit insulator detection in transmission lines. Current work on insulator detection using convolutional neural networks can be divided into two categories. The first category includes two-stage insulator detection algorithms based on the RCNN series, such as the work by R. Girshick et al. [5] and R. Girshick et al. [6], which separate insulator localization and defect recognition into two distinct steps. The second category consists of single-stage algorithms, such as the SSD algorithm proposed by Liu et al. [7] and the YOLO series, including J. Redmon et al. [8] and Li et al. [9], which directly regress the insulator's position and class in a single step. Zhao et al. [10] proposed an insulator detection method based on binary unsupervised critical points

The associate editor coordinating the review of this manuscript and approving it for publication was Ricardo Arias Velásquez (*Corresponding author: Tao Wang*).

Tao Wang, Nan Zhang, Wenqing Yang, Wei Zhang, and Wancai Zhan are with Information System Integration Branch of NARI Technology Company Ltd., Nanjing, China (e-mails: tao434676@163.com, zhnan0451@sina.com, 1229085103@qq.com, gya467787493@163.com, and 49189433@qq.com).

and local aggregation descriptions for insulator localization using a multi-scale sliding window. In [11], multi-scale and multi-feature descriptors were introduced to represent local features, training the network to obtain multiple spatial feature sequences, thereby achieving background noise elimination and insulator detection. In [12], a two-stage Fast-RCNN algorithm was employed for insulator localization. In [13], the Faster-RCNN algorithm was used for insulator localization, combined with a semantic segmentation algorithm for defect recognition. Detection and segmentation networks were fused in [14]. The ResNeXt [15] network, combined with Online Hard Example Mining (OHEM) [16] technology, was applied from both global and local perspectives to achieve accurate insulator localization. In [17], an improved RestNeSt insulator detection algorithm was proposed, combining a region proposal network to address low detection accuracy and long detection times. A railway insulator fault detection network was proposed in [18], cascading detection and classification tasks to achieve insulator localization and fault classification. In [19], a multi-scale residual neural network was proposed, using convolutional kernels of different sizes for feature extraction and fusion to enhance insulator detection accuracy. Sadykova *et al.* [20] utilized data augmentation to generate insulator images in various natural scenes, using YOLOv2 as the base algorithm for insulator detection. In [21], a convolutional attention module with batch normalization and feature fusion modules was proposed, achieving insulator detection on the ResNet backbone network. In [22], based on the Faster-RCNN network, the feature pyramid module was used to improve the network, achieving insulator localization and defect detection through linear detection and vertical projection. Zhang *et al.* [23] proposed an insulator defect detection algorithm by mining potential representations of normal samples to generate defect samples, achieving high reliability in insulator defect detection. Zhang *et al.* [24] proposed a tightly connected feature pyramid structure based on the YOLOv3 network, which reduces network parameters while maintaining better detection performance. Zheng *et al.* [25] proposed an insulator image detection model based on an improved feature fusion single-shot multibox detector, which is based on the SSD algorithm. Glenn Jocher *et al.* [26] proposed YOLOv5, which has certain limitations in detecting small objects, handling background interference, and minimizing computational consumption—especially in complex backgrounds and on resource-constrained devices. Wang *et al.* [27] proposed YOLOv8, which has shown improvements in accuracy and adaptability. However, its adaptability to small objects remains limited. Additionally, the increased computational requirements and longer training times may hinder its application on edge devices.

In the past five years, researchers have introduced detection models based on the Anchor-Free concept, with notable examples including CornerNet [28], ExtremeNet [29], and CenterNet [30]. While these models are Anchor-Free, they still share the core idea of using the center point and width-height information to detect the target object's rectangular bounding box. However, rectangles, typically oriented horizontally or vertically, often include irrelevant background information,

which can limit the detection of inclined objects. Although some studies have explored inclined object detection, research specifically targeting insulator detection remains limited. Furthermore, on the hardware side, the power industry faces challenges due to the lack of standardized equipment models for circuit inspection. Consequently, various UAV models are used globally for transmission line inspections, leading to inconsistent and region-specific studies. Most existing insulator detection algorithms tend to be rigid, focusing on singular performance metrics such as model size, memory requirements, detection accuracy, and speed. To overcome these limitations and avoid redundant research on region-specific inspection devices, a more versatile solution is needed.

In response to these challenges, this paper introduces a novel inclined insulator detection algorithm, RCAS-YOLOv8. The RCAS-YOLOv8 model is designed to handle the unique characteristics of inclined insulators by enhancing the YOLOv8 framework. This method improves the network structure, optimizes the loss function, and addresses various relevant factors to better detect inclined insulators. The development of this innovative RCAS-YOLOv8 network marks a significant step forward in the field.

The main contributions of this paper are as follows:

- 1) This work developed an end-to-end fault detection model for transmission line insulators, RCAS-YOLOv8.
 - 2) This work proposes non-local module based on row and column-level shared correlations, which further enriches the semantic representation of the feature layers and enhances the correlations between feature points.
 - 3) This work designed a vertex offset prediction loss function, constructing a quadrilateral detection frame for the target area based on the predicted coordinates of the four vertices of the insulator, enabling the detection of inclined insulators.
- The rest of this paper is organized as follows. Section 2 introduces related research works. Section 3 explains the research methodology. In Section 4, we explain the experimental results and discussions in detail. Finally, conclusions are given in Section 5.

II. RELATED WORK

A. Yolo Framework

The YOLO series is a deep learning-based general object detection algorithm designed to predict object locations and categories through a regression-based approach. This method transforms the detection problem into a regression problem, enabling simultaneous object localization and classification, thereby improving detection speed. The advent of the YOLO series marked a significant breakthrough in the field of object detection, particularly excelling in applications requiring high real-time performance. The series is adept at detecting objects of varying sizes, enhancing the detection capabilities for both small and large objects. During the detection process, post-processing techniques such as non-maximum suppression are employed to refine the predictions and ultimately output the positions of the target bounding boxes.

In summary, YOLO has achieved significant improvements in both speed and accuracy compared to earlier versions, making it particularly suitable for real-time detection scenarios. Its

feature extraction and fusion strategies have greatly enhanced the ability to detect objects of different sizes, establishing it as one of the most widely used object detection algorithms today. With these advancements, YOLO has not only increased detection efficiency but also ensured accuracy, providing robust support for various computer vision tasks. Notably, the YOLOv8 algorithm, introduced in 2023, has achieved exceptional precision, surpassing previous iterations. While YOLO is primarily used for recognizing and classifying objects that occupy the entire image, its performance in detecting small-scale objects may be somewhat lower than that of certain contemporary algorithms when operating in environments with specific size configurations.

B. Non-local Network

The non-local network, as an attention mechanism, has been widely applied in video classification and detection tasks, primarily aimed at enhancing model performance by capturing global contextual information. This network enhances the expression of local features by compressing channel features and aggregating global spatial features, enabling the model to more effectively understand and distinguish different scenes and objects in videos. During computation, the network generates weighted values by comparing the similarity between the current pixel and all other pixels within the feature layer, using these weights to amplify the feature information of specific pixels. This process allows the model to capture more intricate details and complex patterns.

This approach effectively establishes long-range dependencies within video frames, allowing the network to capture distant relational information between frames, thereby significantly improving the network's performance in video classification and detection tasks. By introducing global features to augment local feature representation, the non-local network has successfully enhanced overall performance in these tasks. Particularly, the establishment of long-range dependencies enables the network to better capture complex associations between video frames, substantially boosting detection accuracy. This method demonstrates the critical importance of integrating global information to improve the performance of deep learning models in video processing tasks.

III. RESEARCH METHODOLOGY

Although the YOLOv8 algorithm is highly efficient, its detection of inclined insulators suffers from suboptimal bounding box alignment, with the detection frames often containing excessive background noise. In this work, we propose an improvement to the original YOLOv8 network. This section focuses on our proposed network, RCAS-YOLOv8. The structural system diagram is shown in (Fig.1). After feature extraction, the network incorporates a cross-type row and column-level shared module, which enriches the semantic information of the feature map by adding long-range dependencies. This approach encourages the model to learn stable feature information with minimal computational cost, resulting in enhanced performance.

A. Backbone Network

This study focuses on the development of a method named RCAS-YOLOv8, designed to enhance the detection accuracy of inclined insulator targets and adapt to the needs of various edge-side detection devices. In power systems, monitoring the condition of insulators is crucial for ensuring the safety of power supply. Traditional detection methods often struggle with accuracy when dealing with complex backgrounds and insulators at different angles. To address these challenges, this paper proposes an optimized object detection algorithm. The RCAS-YOLOv8 method employs a lightweight model design based on the Darknet-53 architecture, a deep neural network structure widely used in object detection tasks. By streamlining the model structure, RCAS-YOLOv8 reduces computational resource consumption while maintaining model performance, making it efficient for deployment on edge devices. To improve the model's downsampling capability, this study replaces traditional pooling layers with convolutional layers that have a stride of 2. This modification not only retains more feature information but also enhances the model's ability to capture image details, thereby improving detection accuracy. RCAS-YOLOv8 retains the C2f structure within the network, which enhances gradient flow and effectively addresses the problem of vanishing gradients. The C2f structure allows the model to better propagate gradient information during training, further boosting its performance. In the Neck section, the PAN (Path Aggregation Network) structure is adopted. The PAN structure effectively enhances the expression of multi-scale features, enabling the model to excel in detecting objects of varying scales, which is particularly important in complex scenes. In the Head section, a decoupled head structure is designed, separating the classification head from the detection head. By handling these two tasks independently, the model can focus more on each task, improving both classification and detection accuracy and efficiency.

B. Row-Column Associative Sharing (RCAS)

Inspired by the non-local networks used in fields such as video classification and machine translation, which are primarily designed to handle long-range dependencies, we recognize that although there is no temporal dependency in images, the pixels within an image are not isolated. Pixels are inherently related, and when considered collectively, they form a specific object. Based on this understanding, we designed a correlation-sharing module at the row and column levels.

The non-local network module enhances local semantics by establishing semantic correlations between the current position and all other positions. However, its computation is performed at the pixel level. Specifically, if the size of the current feature map is $w \times h \times c$, there are $w \times h$ feature points. Each feature point must calculate its similarity with all other points except itself, requiring $w \times h - 1$ computations, approximately resulting in $(w \times h)^2$ operations. This scales the computation to the square of the current feature map's size, which incurs high computational costs when the image resolution is large, making it unsuitable as a lightweight module.

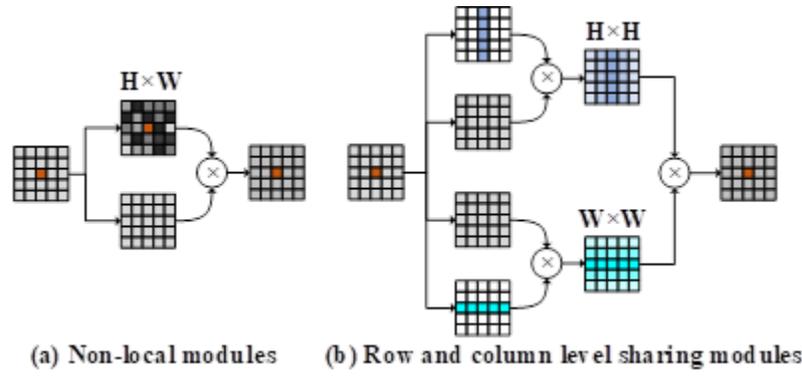


Fig. 2. Comparison of the computational volume of network modules. (a) is the computational process of a non-local network, with a computational volume (HW) of 2; (b) is the computational process of shared row-level and column-level modules..

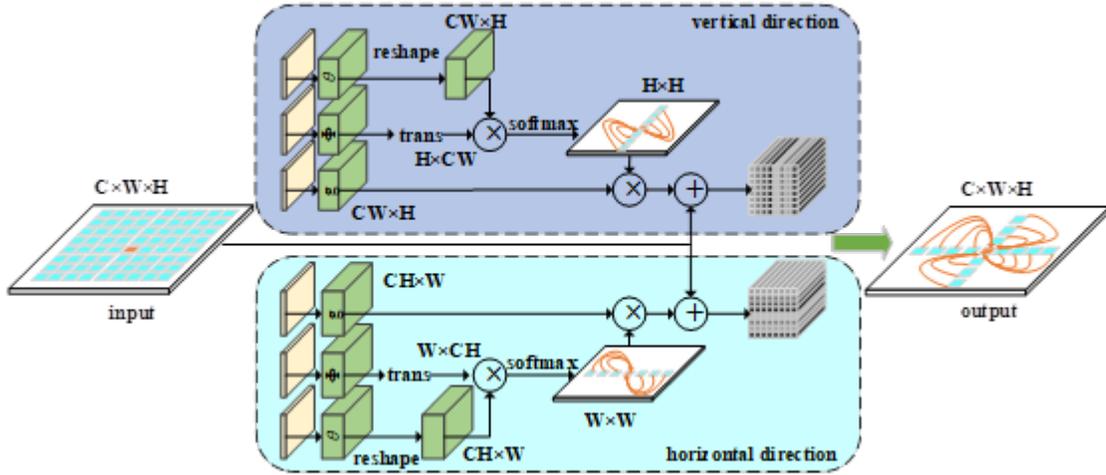


Fig. 3. Row-level and column-level shared modules. The upper part performs column-level similarity calculation, and the lower part performs row-level similarity calculation.

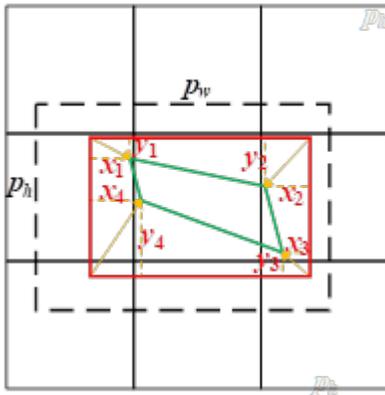


Fig. 4. A graph showing the relationship between detection frames and target locations.

boxes and rectangular boxes used for initial network predictions (coarse detection), respectively. The cyan quadrilateral denotes the actual insulator area. The yellow arrows at the four vertices of the coarse detection rectangle indicate the offsets of the actual insulator's vertices in the corresponding directions. x_n and y_n represent the horizontal and vertical offsets, respectively. The calculation process can be expressed

as follows.

$$x_n = p_w \times e^{t_{x_i}}, i = 1, 2, 3, 4, y_n = p_h \times e^{t_{y_j}}, j = 1, 2, 3, 4, \quad (2)$$

In formula (2), p_w and p_h denote the width and height of the anchor box that best matches the target box. x_n and y_n are the horizontal and vertical offset components of the four vertices, calculated clockwise from the top left corner of the coarse detection rectangle (red rectangular box). Similarly, t_{x_n} and t_{y_n} represent the values corresponding to these four vertex offset components.

$$L_{offset} = \lambda_{coord} \sum_{i=0}^{S \times S} \sum_{j=0}^B \sum_{k=1}^4 I_{ij}^{obj} \cdot \left\{ (2 - \hat{x}_k \times \hat{y}_k) \times \left[(\hat{t}_{x'_k} - t_{x'_k})^2 + (\hat{t}_{y'_k} - t_{y'_k})^2 \right] \right\} \quad (3)$$

In formula (3), S denotes the number of grids in the output layer of the network, resulting in an $S \times S$ grid. B represents the number of candidate boxes generated by each grid, leading to $S \times S \times B$ anchor boxes in total. δ_{ij} indicates whether the j -th anchor box of the i -th grid in the current output layer is responsible for the current target. This is determined by

checking whether the intersection ratio of the anchor box with the target box is the maximum among all intersection ratios of anchor boxes with the current target box. If it is, then δ_{ij} is set to 1, indicating that the current anchor box is responsible for this target; otherwise, δ_{ij} is set to 0.

The loss function of our algorithm comprises the original YOLOv8 loss function and the additional vertex offset prediction loss function. The original loss function includes Localization Loss, Confidence Loss, and Classification Loss. The loss function is expressed as follows: The positioning loss uses CIoU (Complete Intersection over Union) to measure the difference between the predicted bounding box and the true bounding box. CIoU loss formula:

$$\text{CIoU} = 1 - \text{IoU} + \frac{\rho^2(\mathbf{b}, \mathbf{b}^{\text{gt}})}{c^2} + \alpha \cdot \nu, \quad (4)$$

where IoU is the intersection over union of the predicted box and the true box, $\rho(\mathbf{b}, \mathbf{b}^{\text{gt}})$ represents the Euclidean distance between the center points of the predicted box and the true box, c is the diagonal length of the smallest enclosing box containing the predicted box and the true box, α and ν are used to measure the consistency of the aspect ratio. The positioning loss formula is as follows:

$$\mathcal{L}_{\text{loc}} = \sum_{i=1}^N (1 - \text{CIoU}_i), \quad (5)$$

where N represents the number of targets.

Confidence loss is used to measure the difference between the confidence of the predicted target existence and the actual situation, usually using (Binary Cross-Entropy Loss, BCE):

$$\mathcal{L}_{\text{conf}} = - \sum_{i=1}^N [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)], \quad (6)$$

where: y_i is the actual confidence label (1 if the target is present, 0 if it is not present), p_i is the confidence predicted by the model.

The classification loss is used to measure the difference between the category probability distribution predicted by the model and the actual category, and binary cross entropy loss is usually used:

$$\mathcal{L}_{\text{cls}} = - \sum_{i=1}^N \sum_{c=1}^C [y_{i,c} \log(p_{i,c}) + (1 - y_{i,c}) \log(1 - p_{i,c})], \quad (7)$$

where: C is the total number of categories, $y_{i,c}$ is the true label of the i th target belonging to the c th class, and $p_{i,c}$ is the predicted probability that the i th target belongs to the c th class.

The total loss function is a weighted sum of localization loss, confidence loss, classification loss, and vertex offset prediction loss:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{loc}} \cdot \mathcal{L}_{\text{loc}} + \lambda_{\text{conf}} \cdot \mathcal{L}_{\text{conf}} + \lambda_{\text{cls}} \cdot \mathcal{L}_{\text{cls}} + \lambda_{\text{offset}} \cdot \mathcal{L}_{\text{offset}}, \quad (8)$$

Among them, λ_{loc} , λ_{conf} , λ_{cls} and λ_{offset} are the weight coefficients of the corresponding loss parts, which are used to balance the impact of each loss item on the total loss.

The Loss curve is illustrated in Figure 5. The loss curve of the RCAS-YOLOv8 method shows that the loss value

is high in the early stage, but gradually decreases as the training progresses. This shows that the model gradually learns effective feature representations and can reduce training errors. If it continues to decrease in the later stage, it means that the network converges well and the model training is stable. The loss function tends to be stable in the later stage of training, and further optimization is difficult, which means that our model has fully learned the key information in the data.

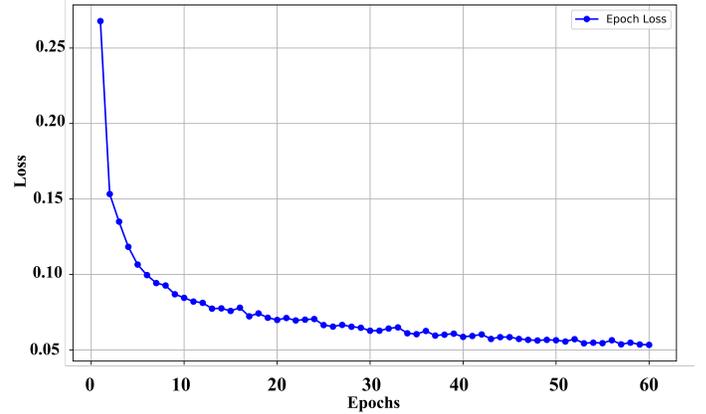


Fig. 5. Loss curve of the RCAS-YOLOv8 method.

IV. EXPERIMENT RESULTS AND DISCUSSION

This section begins by explaining the dataset used, followed by an introduction to the parameters involved in the network training and testing process. The proposed algorithm is then compared with several state-of-the-art algorithms to validate its generalization and effectiveness. Finally, ablation experiments are conducted to discuss and analyze the various modules introduced in the algorithm.

A. Dataset and Implementation Details

The dataset used in this study is the Power Line Insulator Dataset (CPLID) [31] contains two classes: normal and defect, which includes 868 images of insulators captured by drones, which includes 600 images depict normal insulators, while 268 images show defective insulators, with each image having a resolution of 1152×864. To address the issue of data imbalance and account for environmental factors encountered in real-world UAV-based insulator detection, various techniques such as brightness and contrast adjustment, as well as the addition of multiple noise interferences, were applied to augment the dataset. After augmentation, the dataset includes 1,200 images of normal insulators and 1,072 images of defective insulators, totaling 2,271 images. We trained our model using the PyTorch deep learning framework, combined with CUDA 9.0 and cuDNN 7, on an Nvidia RTX 2080Ti GPU. The relevant training parameters used for the network can be referenced in [9]. To ensure that the actual insulator targets were located within the coarse detection boxes, we expanded the bounding box annotations of the insulators by 10 pixels in all directions. Since our algorithm includes the prediction of four vertex offsets, we modified the data format to include



Fig. 6. Experimental comparison chart. The last column is our algorithm, which uses yolov8 as the baseline and adds the vertex prediction this work proposed. The yellow line segment in the figure is the offset of the corresponding vertex, and the cyan line is the insulator position predicted by the algorithm.

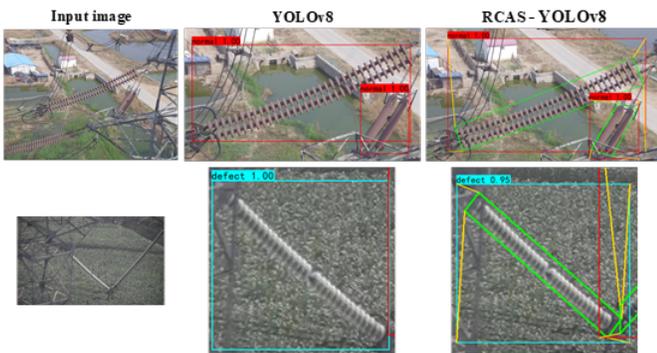


Fig. 7. Detailed Display of Experimental Results

the real coordinates of the four vertices after the original VOC data format. The specific data format is represented as:

$$(X11, Y11, X22, Y22, x1, y1, x2, y2, x3, y3, x4, y4, c), \quad (9)$$

(X11, Y11, X22, Y22): The coordinates of the top-left and bottom-right corners of the annotated insulator bounding box. (x1, y1, x2, y2, x3, y3, x4, y4): The coordinates of the four

vertices of the actual insulator, starting from the top-left corner in a clockwise direction. c: The class label of the detected target, which can be either "normal" or "defective."

To evaluate the effectiveness of our algorithm in detecting inclined insulators, we used Precision, Recall, and AP (Average Precision) as the performance metrics. These three indicators are widely used in the field of object detection. Considering the different focus areas of power companies regarding the above performance metrics, this work proposed an APR metric in this paper, expressed as:

$$APR = \alpha \times \text{Precision} + \beta \times \text{Recall} + \gamma \times \text{AP}, \quad (10)$$

Power companies tend to prioritize the recall rate of images more than detection accuracy. When false positives occur, they can be quickly and easily resolved through manual intervention. However, when insulator detection fails, there is no simple solution, and the costs can be significantly higher. Therefore, our APR metric provides a balanced evaluation that reflects the importance of recall in practical applications.

TABLE I
QUANTITATIVE EXPERIMENTAL RESULTS

Methods	Backbone	R	P	mAP	ARP	F1
Faster R-CNN	ResNet50	0.841	0.854	0.837	0.842	0.847
	ResNet101	0.883	0.871	0.866	0.872	0.887
SSD	ResNet101	0.827	0.842	0.832	0.833	0.834
	VGGNet	0.799	0.811	0.810	0.807	0.805
	MobileNet	0.715	0.729	0.706	0.713	0.722
RetinaNet	ResNet50	0.871	0.874	0.854	0.863	0.872
	ResNet101	0.902	0.898	0.871	0.886	0.900
YOLOv3	Darknet53	0.794	0.833	0.824	0.817	0.813
	MobileNet	0.778	0.815	0.802	0.797	0.796
YOLOv3 tiny	Darknet53 tiny	0.716	0.732	0.713	0.718	0.724
YOLOv4	CSPDarknet	0.844	0.861	0.859	0.855	0.852
YOLOv4 tiny	CSPDarknet-tiny	0.773	0.798	0.763	0.773	0.785
YOLOv5	CSPDarknet	0.796	0.872	0.863	0.874	0.832
CenterNet	ResNet50	0.884	0.892	0.872	0.880	0.888
YOLOv8	CSPNet	0.896	0.895	0.883	0.879	0.895
RCAS-YOLOv8	CSPNet	0.905	0.901	0.889	0.891	0.903

TABLE II
COMPARISON OF PERFORMANCE WITH NON-LOCAL AND IMPROVED NON-LOCAL NETWORKS

Methods	mAP (%)	Up of mAP (%)	Times (ms)	Up of Times (%)
YOLOv8 (CSPNet)	0.824	/	28.65	/
+ Non-local net	0.856	3.88	30.12	-5.13
+ Improved Non-local net	0.853	3.52	29.79	-3.98
Faster R-CNN (ResNet50)	0.837	/	138.46	/
+ Non-local net	0.861	2.87	146.73	-5.97
+ Improved Non-local net	0.866	3.46	141.54	-2.22
CenterNet (ResNet50)	0.872	/	22.17	/
+ Non-local net	0.898	2.98	23.89	-7.76
+ Improved Non-local net	0.902	3.44	22.94	-3.47

B. Experimental Results

In the experiment, we utilized advanced algorithms including Faster R-CNN (ResNet101), YOLOv4 (CSPDarknet), RetinaNet (ResNet101), YOLOv5, YOLOv8, and the Anchor-Free CenterNet (ResNet50) for comparison with our proposed method. The experimental results are illustrated in Figure 6. To enhance the experiment, additional cruise images outside the dataset were introduced for detection, with the results displayed in columns (d) and (e) of Figure 6.

Due to image resolution constraints, Figure 7 selectively magnifies key areas detected by some of the algorithms. Since the detection results of the comparative algorithms are all standard rectangular boxes, we chose to focus on the baseline algorithm YOLOv8 for comparison. This approach enables a clearer understanding of the differences between our proposed method and conventional object detection algorithms. All these algorithms can detect insulators to varying degrees. However, our proposed method demonstrates a superior ability to approximate actual insulator targets, a feat that is challenging for other algorithms. The resulting bounding boxes from our method contain less background information, facilitating finer defect classification. This represents our qualitative comparison approach.

Table 1 gives the quantitative experimental results, comparing different backbone feature extraction networks using SSD, Faster R-CNN, RetinaNet and CenterNet algorithms.

As shown in Table 2, our RCAS-YOLOv8 achieves the best performance, with P and mAP scores of 0.901 and 0.889 respectively, ARP score of 0.891, and F1 score of 0.903 indicating the highest overall performance. The experimental data show that as our algorithm adopts a more complex feature extraction network, its overall performance improves, and this trend is clearly observed in the last few rows, which is consistent with existing knowledge.

Table 3 presents a comparison of different detection models based on key performance metrics: Average Precision (AP), F1 Score, and Frames Per Second (FPS). The Improved YOLOv4 model outperforms others in terms of AP and F1 Score, particularly for defect detection, with an AP of 94.53% and an F1 score of 0.94 for both insulators and defects. This indicates that the Improved YOLOv4 offers a balanced and high-performance detection capability, making it the most accurate model overall. YOLOv3, while achieving a higher FPS (60) compared to other models, shows the lowest AP for insulator detection (79.91%) and a slightly lower F1 score for insulator detection (0.88). This suggests that YOLOv3 is faster but sacrifices some accuracy, particularly in detecting insulators. SSD achieves a solid performance with an AP of 88.16% for insulators and 96.43% for defects, alongside an F1 score of 0.94 for defects and 0.90 for insulators. SSD strikes a good balance between detection accuracy and speed, with 53 FPS, making it suitable for scenarios requiring both high performance and reasonable inference time. Faster

R-CNN, though offering strong detection accuracy with an AP of 87.27% for insulators, lags behind in terms of FPS, achieving only 15 FPS, which makes it less ideal for real-time applications. Overall, Improved YOLOv4 stands out for accuracy, particularly in detecting both insulators and defects, although at the expense of speed, while YOLOv3 is the fastest but less precise. SSD offers a good compromise, and Faster R-CNN, despite its strong accuracy, is slower and less suitable for real-time detection.

C. Ablation Study

To further analyze the impact of our proposed cross non-local structure on algorithm performance, we selected three baseline algorithms: YOLOv8, Faster R-CNN (ResNet50), and CenterNet (ResNet50). These were chosen to compare the performance differences between our cross non-local structure and the non-local network proposed in reference [36]. The three baseline algorithms represent an anchor-based one-stage algorithm, a two-stage algorithm, and the anchor-free CenterNet algorithm, respectively, to validate the generalization capability of our proposed structure.

As shown in Table 2, our proposed cross structure maintains a high degree of consistency in accuracy compared to the original non-local network, even outperforming it in the latter two baseline algorithms. However, in terms of detection speed, our structure sacrifices significantly less speed, achieving nearly double the speed of the original structure. This is because our algorithm employs a shared-weight approach for computing correlations between feature points, rather than the exhaustive computation used in the original method. This approach greatly reduces the computational load, demonstrating that our proposed cross non-local network achieves a better balance of speed and accuracy. Here is the LaTeX code for the table you provided:

V. CONCLUSIONS

This paper proposes a multi-level adaptive oblique insulator detection algorithm, RCAS-YOLOv8, which is an enhanced deep learning algorithm specifically designed for detecting oblique insulators on various mobile detection devices. By introducing row- and column-shared non-local modules to capture the global correlation of feature points, and adding vertex offset prediction loss to generate more accurate quadrilateral detection boxes, the algorithm significantly improves detection accuracy. Coupled with the lightweight Darknet-53 model, RCAS-YOLOv8 maintains high performance while reducing computational resource consumption, making it particularly suitable for real-time detection tasks in complex environments. The algorithm demonstrates significant advantages in insulator detection within power systems and offers valuable insights for future detection technologies.

Our model currently has some accuracy issues when detecting small objects, especially in low-resolution images or scenes with dense objects. This is because the features of small objects are difficult to extract, which limits the performance of the model. To address this issue, in the future we plan to introduce an adaptive feature extraction module to further

improve the accuracy of the model when detecting small objects. In addition, the existing dataset only contains two types of labels: normal and defective. However, in view of the needs of night inspections and infrared image detection in actual applications of power companies, we realize that the types of datasets need to be expanded. Therefore, in the future, we plan to collect a large number of insulator images in different scenarios, including infrared images, night images, and natural disaster scenes, so as to effectively detect insulator defects under a wider range of actual conditions.

ACKNOWLEDGMENTS

This work was supported by the project "Research on Dynamic Perception and Early Warning Technology of Rain, Snow and Freezing Disasters in Transmission Lines by Integrating Multimodal Data", with project number 5500-202424166A-1-ZN

REFERENCES

- [1] J. Ding, Z. Zhang, X. Yu, X. Zhao, and Z. Yan, "A novel moving object detection algorithm based on robust image feature threshold segmentation with improved optical flow estimation," *Applied Sciences*, vol. 13, no. 8, p. 4854, 2023. doi:10.3390/app13084854.
- [2] Q. Wu and J. An, "An active contour model based on texture distribution for extracting inhomogeneous insulators from aerial images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, pp. 3613–3626, 2014. doi:10.1016/j.engappai.2023.106656.
- [3] P. Tan, X.-F. Li, J.-M. Xu, and Y. Ning, "Catenary insulator defect detection based on contour features and gray similarity matching," *J. Zhejiang Univ.-Sci. A*, vol. 21, pp. 64–73, 2020. doi:10.1631/jzus.A1900341.
- [4] M. Tomaszewski, R. Gasz, S. S. Kasana, J. Osuchowski, S. Singh, and S. Zator, "Tcip: Transformed colour intensity profiles analysis for fault detection in power line insulators," *Multimedia Tools and Applications*, pp. 1–32, 2024. doi:10.1109/ACCESS.2024.3474255.
- [5] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587, June 2014. doi:10.1109/CVPR.2014.81.
- [6] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1440–1448, June 2015. doi:10.1109/ICCV.2015.169.
- [7] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A. Berg, "Ssd: Single shot multibox detector," in *Proceedings of the European Conference on Computer Vision*, pp. 21–37, October 2016. doi:10.1007/978-3-319-46448-0_2.
- [8] J. Redmon and A. Farhadi, "Yolo9000: Better, faster, stronger," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6517–6525, July 2017. doi:10.1109/CVPR.2017.690.
- [9] H. Li, L. Deng, C. Yang, J. Liu, and Z. Gu, "Enhanced yolo v3 tiny network for real-time ship detection from visual image," *Ieee Access*, vol. 9, pp. 16692–16706, 2021. doi:10.1109/ACCESS.2021.3053956.
- [10] Z. Zhao, G. Xu, and Y. Qi, "Representation of binary feature pooling for detection of insulator strings in infrared images," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 23, pp. 2858–2866, 2016. doi:10.1109/TDEI.2016.7736846.
- [11] S. Liao and J. An, "A robust insulator detection algorithm based on local features and spatial orders for aerial images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, pp. 963–967, 2015. doi:10.1109/LGRS.2014.2369525.
- [12] F. Gao, J. Wang, Z. Kong, J. Wu, N. Feng, S. Wang, P. Hu, Z. Li, H. Huang, and J. Li, "Recognition of insulator explosion based on deep learning," in *Proceedings of the IEEE 2017 International Computer Conference on Wavelet Active Media Technology and Information Processing*, pp. 79–82, December 2017. doi:10.1109/ICCWAMTIP.2017.8301453.
- [13] Y. Wang, J. Wang, F. Gao, P. Hu, L. Xu, J. Zhang, Y. Yu, J. Xue, and J. Li, "Detection and recognition for fault insulator based on deep learning," in *Proceedings of the IEEE 2018 International Congress on Image and Signal Processing, BioMedical Engineering and Informatics*, October 2018. doi:10.1109/CISP-BMEI.2018.8633245.

- [14] X. Li, H. Su, and G. Liu, "Insulator defect recognition based on global detection and local segmentation," *IEEE Access*, vol. 8, pp. 59934–59946, 2020. doi:10.1109/ACCESS.2020.2982288.
- [15] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5987–5995, July 2017. doi:10.1109/CVPR.2017.634.
- [16] A. Shrivastava, A. Gupta, and R. Girshick, "Training region-based object detectors with online hard example mining," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 761–769, June 2016. doi:10.1109/CVPR.2016.89.
- [17] S. Wang, Y. Liu, Y. Qing, C. Wang, T. Lan, and R. Yao, "Detection of insulator defects with improved resnet and region proposal network," *IEEE Access*, vol. 8, pp. 184841–184850, 2020. doi:10.1109/ACCESS.2020.3029857.
- [18] Z. Wang, X. Liu, H. Peng, L. Zheng, J. Gao, and Y. Bao, "Railway insulator detection based on adaptive cascaded convolutional neural network," *IEEE Access*, vol. 9, pp. 115676–115686, 2021. doi:10.1109/ACCESS.2021.3105419.
- [19] L. She, Y. Fan, J. Wang, L. Cai, J. Xue, and M. Xu, "Insulator surface breakage recognition based on multiscale residual neural network," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–9, 2021. doi:10.1109/TIM.2021.3106112.
- [20] D. Sadykova, D. Pernebayeva, M. Bagheri, and A. James, "In-yolo: Real-time detection of outdoor high voltage insulators using uav imaging," *IEEE Trans. Power Del.*, vol. 35, pp. 1599–1601, 2020. doi:10.1109/TPWRD.2019.2944741.
- [21] Z. Gao, G. Yang, E. Li, and Z. Liang, "Novel feature fusion module-based detector for small insulator defect detection," *IEEE Sens. J.*, vol. 21, pp. 16807–16814, 2021. doi:10.1109/JSEN.2021.3073422.
- [22] W. Zhao, M. Xu, X. Cheng, and Z. Zhao, "An insulator in transmission lines recognition and fault detection model based on improved faster rcnn," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–8, 2021. doi:10.1109/TIM.2021.3112227.
- [23] D. Zhang, S. Gao, L. Yu, G. Kang, X. Wei, and D. Zhan, "Defgan: Defect detection gans with latent space pitting for high-speed railway insulator," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–10, 2021. doi:10.1109/TIM.2020.3038008.
- [24] X. Zhang, Y. Zhang, J. Liu, C. Zhang, X. Xue, H. Zhang, and W. Zhang, "Insudet: A fault detection method for insulators of overhead transmission lines using convolutional neural networks," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–12, 2021. doi:10.1109/TIM.2021.3120796.
- [25] H. Zheng, Y. Sun, X. Liu, C.-L.-T. Djike, J. Li, Y. Liu, J. Ma, K. Xu, and C. Zhang, "Infrared image detection of substation insulators using an improved fusion single shot multibox detector," *IEEE Trans. Power Del.*, vol. 36, pp. 3351–3359, 2021. doi:10.1109/TPWRD.2020.3038880.
- [26] G. Jocher, A. Chaurasia, A. Stoken, J. Borovec, Y. Kwon, K. Michael, J. Fang, Z. Yifu, C. Wong, D. Montes, *et al.*, "ultralytics/yolov5: v7. 0-yolov5 sota realtime instance segmentation," *Zenodo*, 2022. doi:10.5281/zenodo.3908559.
- [27] G. Wang, Y. Chen, P. An, H. Hong, J. Hu, and T. Huang, "Uav-yolov8: A small-object-detection model based on improved yolov8 for uav aerial photography scenarios," *Sensors*, vol. 23, no. 16, p. 7190, 2023. doi:10.3390/s23167190.
- [28] H. Law and J. Deng, "Cornernet: Detecting objects as paired keypoints," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 734–750, September 2018. doi:10.1007/978-3-030-01264-9_45.
- [29] X. Zhou, J. Zhuo, and P. Krähenbühl, "Bottom-up object detection by grouping extreme and center points," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 850–859, June 2019. doi:10.1109/CVPR.2019.00094.
- [30] X. Zhou, D. Wang, and P. Krähenbühl, "Objects as points," *arXiv*, 2019. doi:10.48550/arXiv.1904.07850.
- [31] X. Tao, D. Zhang, Z. Wang, X. Liu, H. Zhang, and D. Xu, "Detection of power line insulator defects using aerial images analyzed with convolutional neural networks," *IEEE Trans. Syst. Man Cybern.: Syst.*, vol. 50, pp. 1486–1498, 2020. doi:10.1109/TSMC.2018.2871750.



Wang Tao received a master's degree in computer technology from Ningxia University, Yinchuan, China, in 2021. Currently working at Nari Group Co., Ltd. His main research interests include salient target detection, image generation in the power field, etc..



Zhang Nan received a master's degree in engineering management from Renmin University of China in Beijing, China in 2024. He currently works at NARI Group Co., Ltd..



Wenqing Yang received his Ph.D. degree in computer science and technology from Nanjing University, Nanjing, China, in 2005. He is currently employed by NARI Group Co., Ltd..



Zhang Wei received a bachelor's degree in software engineering from Nanchang University, Jiangxi, China in 2008. He is currently employed by NARI Group Co., Ltd..



Zhang Wancai received his Ph.D. degree in computer software and theory from Beijing University of Aeronautics and Astronautics, Beijing, China, in 2016. He is currently employed by NARI Group Co., Ltd..