Advances in Computer, Signals and Systems (2026)
Clausius Scientific Press, Canada
DOI: 10.23977/acss.2026.100111
ISSN 2371-8838 Vol. 10 Num. 1

GMI-DETR: Insulator defect detection network based on GSConv and multi-scale isometric convolution

Xuanyu Liao[1,a], Chengjiang Zhou[1,b,*]

[1]School of Information Science and Technology, Yunnan Normal University, Kunming, Yunnan, China
[a]xuanyuleo@163.com, [b]ChengjiangZhou@ynnu.edu.cn
[*]Corresponding Author

Keywords: Insulator; GSConv; MIC; DETR; Defect detection

Abstract: As a core component of high-voltage transmission lines, the accuracy and efficiency of insulator defect detection are directly related to the safe and stable operation of power systems. Aiming at the problems of insufficient accuracy, weak small-target defect recognition ability and poor adaptability to complex backgrounds in existing insulator detection algorithms, this study introduces a novel insulator defect detection network that integrates GSConv and multi-scale isometric convolution. Firstly, we design the Group Sparse Convolution Cross-stage Fusion Module (GSC), which reduces the number of model parameters and computational overhead while enhancing the feature discrimination ability. Secondly, the Multi-scale Isometric Convolution Module (CRMIC) is constructed to strengthen the recognition capability for small-target defects. Finally, a parallel architecture combining the CRMIC branch and the GSC branch is built to achieve efficient integration of semantic information and spatial details, thereby improving the adaptability to defects with variable scales and complex backgrounds. Experimental results show that, compared with the Baseline, the proposed GMI-DETR achieves 88.0% mAP@50-95 and 99.3% mAP@50 on public datasets, with respective improvements of 0.5% and 0.3%. The model proposed in this paper demonstrates excellent performance in surface defect detection.

1. Introduction

As a core insulating component of high-voltage transmission lines, insulators undertake the dual critical responsibilities of electrical insulation and mechanical fixation, and their operational status directly determines the safety, stability, and power supply reliability of power systems. China's transmission lines span a wide area and cover regions with complex terrain. Exposed to the outdoor natural environment for a long time, insulators are vulnerable to wind and rain erosion, pollution accumulation, sudden temperature changes, mechanical wear, and other factors, resulting in various defects like self-explosion, cracks, damage, and flashover. If such defects fail to be detected and addressed in a timely manner, they will not only reduce the insulation performance of insulators but also may trigger serious accidents such as line tripping, power outages, and even fires, bringing significant economic losses and safety risks to social production and life. Therefore, achieving efficient, accurate, and automated detection of insulator defects has become a core demand and research hotspot in the field of smart grid operation and maintenance.

Traditionally, insulator detection relies on manual inspection and observation with handheld devices. Constrained by factors such as inspectors' experience, harsh operating environments, and visual blind spots, it has inherent drawbacks including low efficiency, high missing and false detection rates, and significant safety risks, making it difficult to address the routine operation and maintenance demands of large-scale transmission lines.

With the rapid development of UAV aerial photography technology and deep learning-based object detection algorithms, automated detection solutions have gradually replaced traditional manual methods and become the mainstream direction for insulator defect detection. Among them, single-stage detection algorithms represented by the YOLO series have been widely applied in the field of power equipment defect detection due to their advantages of fast detection speed and strong real-time performance. Relevant research mostly focuses on model lightweighting, enhancement of feature extraction capability, and optimization of multi-scale target adaptability.

Although deep learning-based insulator detection algorithms have made certain progress, they still face numerous bottlenecks in practical engineering applications. On the one hand, insulator defects are characterized by large scale differences, a high proportion of small-target defects, and strong interference from complex backgrounds. Traditional convolution modules tend to suffer from loss of detailed information and insufficient receptive field adaptability during feature extraction, leading to low detection accuracy for small-target defects. In existing research on transmission line insulator faults, several technologies have been adopted to enrich background information and enhance the model's recognition capability in complex scenarios. For example, Yi et al.[1] constructed the YOLO-S model integrating the lightweight VoV-GSCSP module and the novel MaECA attention mechanism, solving the key problems of insufficient accuracy and large computational complexity of existing insulator defect detection algorithms. Zhou et al.[2] proposed an optimized YOLOv5 method combining two-stage feature extraction and rotated candidate boxes, addressing the key issues of small-target missing detection and excessive redundant background in prediction boxes caused by complex background interference in insulator fault detection. Zhang et al.[3] built the FINet detection framework integrating synthetic fog data augmentation and channel attention mechanism, resolving the core problems of insufficient data and poor model generalization in insulator defect detection under small-sample conditions. Hao et al.[4] constructed the ID-YOLO model fusing the CSP-ResNeSt backbone network and Bi-SimAM-FPN feature pyramid, overcoming the low detection accuracy caused by complex background interference and excessively small insulator defect targets in aerial inspection images. On the other hand, existing algorithms often improve performance by stacking network layers or increasing the number of convolution kernels, resulting in a sharp increase in model parameters and excessive computational overhead. This makes them difficult to adapt to low-computing-power deployment scenarios such as UAVs and edge detection devices, restricting the engineering application of the algorithms. Existing methods have addressed this issue: for instance, Liu et al.[5] built the FA-YOLO model integrating attention mechanism and lightweight backbone, solving the practical application problems of large model computation and fast and accurate recognition under complex backgrounds in insulator defect detection. Tian et al.[6] constructed an improved lightweight YOLOv3 model fusing ResNet50 dense connection and attention mechanism, balancing accuracy and efficiency in insulator defect detection under limited computing resources and complex background interference in UAV power inspection. Tao et al.[7] proposed the SnakeNet model integrating dynamic snake convolution and efficient multi-scale attention, tackling the three core problems of insufficient feature extraction, weak anti-background interference capability, and model lightweighting in small-target insulator defect detection. Jin et al.[8] built the lightweight YOLOv11-ssL network integrating C3k2-star structure and adaptive pruning, solving the key problems of large computational load, difficulty in real-time accurate segmentation and evaluation of insulator pollution detection models on edge devices. Li et al.[9] proposed an improved DETR method integrating multi-scale backbone and self-attention upsampling module, resolving the core issues of low accuracy and unstable matching in small-target insulator defect detection under complex environments. In addition, mainstream feature fusion strategies mostly adopt simple feature concatenation or weighted summation, failing to fully explore the complementary information of features at different scales. Especially in mainstream feature fusion modules such as C2f, the redundancy of convolution operations and the singularity of feature expression further affect the model's recognition robustness for insulator defects in complex scenarios.

Aiming at the aforementioned problems, this paper proposes an improved insulator defect detection scheme integrating GSConv and multi-scale isometric convolution. The core idea is to embed GSConv into the C2f module structure and reconstruct the feature extraction unit to achieve the synergistic improvement of lightweight performance and feature expression capability. Meanwhile, a multi-scale isometric convolution branch is designed to form a parallel output architecture with the C2f module. Through a targeted multi-scale feature fusion strategy, the complementary feature information of different levels and scales is fully explored, enhancing the ability to distinguish small-target defects and insulator features under complex backgrounds. The research work of this paper aims to solve the problems of unbalanced accuracy and efficiency, and insufficient small-target detection performance in existing insulator detection algorithms, providing a more engineering-practical technical solution for the automated inspection of transmission line insulator defects and contributing to the intelligent upgrading of smart grid operation and maintenance. The main contributions are as follows:

1) Aiming at the problems of convolution redundancy and single feature expression in traditional C2f modules, this paper designs a new module GSC (Group Sparse Convolution Cross-stage Fusion Module) by deeply integrating the group sparse characteristics of GSConv with the cross-stage fusion structure of C2f. This module is designed to minimize model parameters and computational overhead for lightweight deployment, while enhancing feature spatial correlation and resolving the imbalance between accuracy and efficiency in insulator detection.

2) To enhance the capability of identifying small-target defects, an independent module named CRMIC based on multi-scale isometric convolution is designed, which forms a parallel output mode together with the optimized GSC module. Through the accurate capture of features at different scales by isometric convolution, this module adapts to the characteristics of insulator defects with large scale variations and a high proportion of small targets, thus significantly improving the feature recognition capability for small-target defects.

3) Through systematic validation conducted on the insulator dataset, the proposed method significantly outperforms mainstream existing methods in key metrics such as mAP@75 and mAP@50-95. This validates the practical value of our approach in detecting subtle insulator defects within complex backgrounds, thereby providing reliable technical support for the intelligent inspection of power equipment.

2. Related works

2.1 GSConv

GSConv is a lightweight convolution operation proposed by Li et al.[10] in 2024. Aiming at the deficiency of Depthwise Separable Convolution (DSC) in feature expression capability while retaining its lightweight advantages, this method proposes a hybrid feature shuffle mechanism based on Standard Convolution (SC), incorporating the efficiency of Depthwise Separable Convolution (DSC), thus providing an efficient convolutional alternative for real-time object detection tasks.

The core structure of GSConv comprises three key steps. The first, dual-branch feature extraction, involves feeding the input feature map simultaneously into two parallel branches. The standard convolution branch uses standard convolution kernels with half the number of output channels for feature extraction, capturing global dependencies between channels and preserving spatial and channel correlation information. In contrast, the depthwise separable convolution branch employs Depthwise Convolution (DWConv) for spatial feature extraction, followed by Pointwise Convolution (PWConv) to adjust the number of channels to half of the output channels, achieving lightweight computation. Second, in the shuffle fusion stage, the feature maps from both branches undergo channel-wise concatenation followed by a uniform shuffle. This operation is designed to disrupt channel isolation within the DSC branch, enabling full fusion of the global information from the SC branch and the local spatial information from the DSC branch, and reconstructing spatial-channel correlation. Third, output and expansion, the shuffled feature map is used as the final output, which can be directly applied to subsequent network layers.

By breaking down standard convolution into depthwise (DWConv) and pointwise (PWConv) operations, DSC significantly reduces computational complexity and parameter count. The core motivation of GSConv is to compensate for DSC's deficiency in feature expression capability while maintaining its lightweight advantages.

2.2 Multi-scale Isometric Convolution

The MIC module is the core component of Multi-scale Isometric Convolution Network (MICN), for long-term time series prediction, we employ the MICN architecture proposed by Wang et al.[11] This module captures both local short-term variations and global long-term trends of time series simultaneously via multi-scale isometric convolution, to overcome the receptive field limitation of traditional convolution for long-term time series forecasting.

The MIC module is composed of two key sub-modules. The first is the Local Module, which extracts local features at different temporal granularities using multi-scale downsampling convolutions (kernel size = 2, 4, 8) to capture short-term variation patterns while reducing computational complexity. The second is the Global Module, which models global feature correlations using Isometric Convolution, maintains feature spatial isometry, avoids information loss caused by scale transformation, and effectively captures long-term trends and periodic patterns. Finally, the outputs of the two modules are fused via channel attention-based weighted fusion to integrate local and global features, highlight key temporal scale information, and improve prediction performance.

The core objective of the MIC module is to maintain the spatial consistency of features at different scales through isometric convolution, realize effective fusion of local and global contexts, and enhance the accuracy of long-term time series prediction.

3. Methdology

3.1 GMI-DETR

To identify small-target defects in complex scenarios, as illustrated in the figure 1, this study proposes an insulator defect detection network named GMI-DETR, which is based on the LW-DETR framework and integrates GSConv with multi-scale equidistant convolution. GMI-DETR consists of an encoder, a redesigned GSC module, a redesigned CRMIC module, a newly developed CRMIC-GSC parallel architecture, and a decoder. This section outlines the architecture of GMI-DETR, followed by a detailed elaboration on GSC and CRMIC—with a specific focus on their customization for the unique demands of the defect detection task.

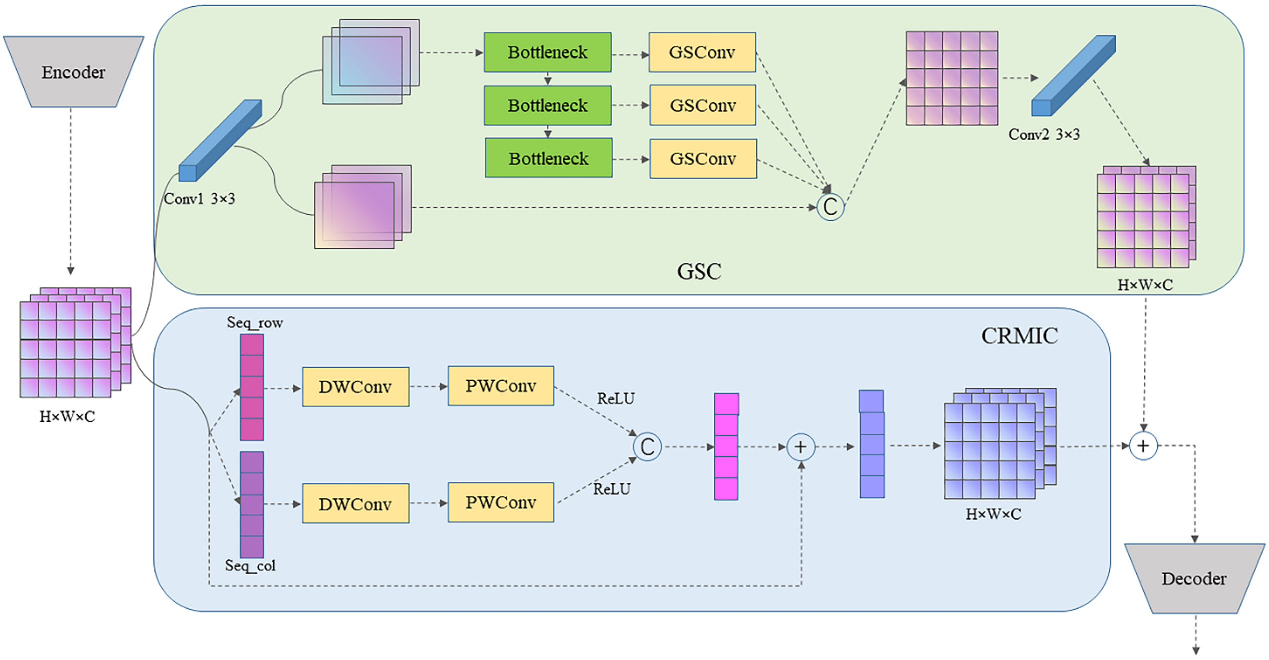

Figure 1. Structure of GMI-DETR.

3.2 GSC

To reduce the convolutional redundancy and parameter count of the C2f module, this paper proposes a Grouped Sparse Convolutional Cross-stage Fusion Module (GSC).

The GSC module adopts a serial-parallel hybrid feature processing pipeline. First, it expands the input feature channels through the cv1 convolutional layer and evenly splits them into two branches to form an initial feature list. Then it enters a cyclic processing stage: the current feature is sequentially extracted from the end of the list, which is successively processed by the Bottleneck module for feature extraction and the GSConv module for lightweight enhancement; the processed result is appended to the end of the feature list, forming progressive feature accumulation. After n iterations, all accumulated features are concatenated along the channel dimension, and finally, the cv2 convolutional layer is used for feature fusion and channel adjustment to output a feature map with target dimensions. The expression is shown as follows:

$$[y_0,y_1]=\text{Split}(\text{cv1}(\text{x}),\text{dim}=1) \tag{1}$$

$$\text{For i}=1,2,3,\ldots\ldots,\text{n}: y_{i+1}=GSConv_i(Bottleneck_i(y_i)) \tag{2}$$

$$\text{Output}=\text{cv2}\left(\text{Concat}(y_0,y_1,y_2,\ldots\ldots,y_{i+1})\right) \tag{3}$$

where $y_i \in \mathrm{R}^{B\times C_k\times H\times W}$. Split evenly splits the features along the channel dimension. Concat concatenates all features along the channel dimension.

This design preserves multi-scale features while enabling the gradual extraction of deep-level features via serial processing, striking a balance between feature diversity and computational efficiency.

3.3 CRMIC

To enhance the recognition capability for small-target defects, this paper proposes a module named CRMIC, which is based on multi-scale equidistant convolution.

This module features a dual-path feature interaction and fusion structure, and its operation pipeline is as follows: First, the module projects the input dual sequences (seq_row and seq_col, both with the shape of [B, L, C]) onto a unified feature space via a shared channel adjustment layer, yielding feature representations with the shape of [B, L, feature_size]. Next, the two types of features are respectively processed by independent depthwise separable convolutions — the horizontal path extracts row-wise local dependencies through a combination of depthwise convolution and pointwise convolution, while the vertical path adopts the same structure to capture column-wise local dependencies; both paths are followed by the ReLU activation function to enhance nonlinearity. Then, the features processed along the two directions are fused via element-wise average to obtain comprehensive features. Finally, the channel recovery layer maps the feature dimension back to the original input channel number, and a residual connection is established with the original seq_row input to generate the final output. The expression is shown as follows:

$$H_r = Linear_a(R),\ H_c = Linear_a(C) \tag{4}$$

$$H_r^{'} = ReLU\left(PWConv_r\left(DWConv_r\left(H_r^{\mathrm{T}}\right)\right)\right)^{\mathrm{T}} \tag{5}$$

$$H_c^{'} = ReLU\left(PWConv_c\left(DWConv_c\left(H_c^{\mathrm{T}}\right)\right)\right)^{\mathrm{T}} \tag{6}$$

$$\mathrm{F} = \frac{H_r^{'} + H_c^{'}}{2} \tag{7}$$

$$\mathrm{O} = \mathrm{R} + \mathrm{Linear}_r(F) \tag{8}$$

where H_r denotes channel adjustment, $H_r^{'}$ denotes horizontal convolution processing, $H_c^{'}$ denotes vertical convolution processing, F denotes feature fusion, and O denotes channel recovery and residual connection.

This process enhances feature interaction through independent modeling and collaborative fusion of bidirectional information, allowing for improved representation without sacrificing computational efficiency.

3.4 Feature Fusion and Output

The input feature map x is first duplicated and fed into the CRMIC branch and GSC branch for processing, respectively. In the CRMIC branch, horizontal and vertical scanning sequences are generated via the cross_scan operation; after bidirectional feature interaction is performed by the CRMIC module, the sequences are restored to 2D feature maps through the inverse_cross_scan operation, and dimension alignment is conducted via the channel_align layer. The processing within the GSC branch begins with the original input undergoing the GSC module for multi-scale feature extraction, followed by layer normalization. In the final step, a fusion operation performs an adaptive weighted summation on the outputs of both branches, using learnable weights normalized by softmax, to produce the fused feature output.

This design enables a complementary fusion between MIC's sequence interaction capability and C2f's multi-scale feature extraction capability, and adaptively adjusts the contribution of the two branches through learnable weights.

4. Experiments and Results

The experiments for this model were conducted on an NVIDIA GeForce RTX 4090 GPU with the Ubuntu 13.2.0, Python 3.8 and PyTorch 2.4 1operating system. The training parameters were set as follows: we set the batch size to 2, trained the model for 200 epochs, and used an initial learning rate of 0.0001.

4.1 Evaluation Index

We compare and analyze the algorithm using evaluation metrics including mean Average Precision (mAP), number of parameters, resource consumption, and Floating Point Operations Per Second (FLOPs). The mean Average Precision (mAP) metric quantifies the overall performance of multi-category object detection models; It integrates detection and localization accuracy by applying an Intersection over Union (IoU) threshold, enabling more robust assessment of object detection across multiple categories and scales. The formula is shown as follows:

$$\text{AP}=\int_0^1 P(R)dR \tag{9}$$

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

4.2 Comparative experiment

To verify the detection performance of GMI-DETR, we conducted a comparative experiment with mainstream detection algorithms on the CPLID dataset. Table 1 shows the experimental results. Analysis of the table demonstrates that our proposed method achieves significant advantages over other models. Specifically, compared with DAB-DETR, DINO-DETR, RT-DETR, LW-DETR, YOLOv10, YOLOv11 and YOLOv12, GMI-DETR improves the mAP@50 by 1.3%, 0.4%, 0.2%, 0.3%, 0.3%, 0.1% and 0.1%, respectively; for mAP@50-95, the corresponding improvements are 18.0%, 3.7%, 1.5%, 0.5%, 0.6%, 0.5% and 0.1%. In terms of mAP@75, GMI-DETR outperforms DAB-DETR, DINO-DETR and RT-DETR by 12.2%, 2.0% and 0.3%, respectively, while only slightly decreasing compared with the remaining models. In addition, GMI-DETR has a much smaller number of parameters than DAB-DETR, DINO-DETR and RT-DETR, and also delivers favorable FLOPs performance.

Table 1. CPILD dataset comparison.

Methods	mAP@50-95(%)	mAP@50(%)	mAP@75(%)	Params(M)	FLOPs
DAB-DETR	70.0	98.0	86.0	43.7	99.6
DINO-DETR	84.3	98.0	96.2	47.5	235.0
RT-DETR	86.5	99.1	97.9	60.9	186.2
LW-DETR	87.5	99.0	98.3	28.2	42.7
Yolov10s	87.4	99.0	98.4	7.2	21.4
Yolov11s	87.5	99.2	98.5	9.4	21.3
Yolov12s	87.9	99.2	98.8	9.2	21.2
GMI-DETR	88.0	99.3	98.2	29.5	45.5

4.3 Ablation experiment

To verify the effectiveness of the proposed method, ablation experiments were conducted on the insulator dataset. We can observe from Table 2 that the performance of Experiment 2 did not improve significantly compared with that of Experiment 1 with the original model after the introduction of the GSC module, but the overall performance basically maintained the same level. After introducing the CRMIC module in Experiment 3, a comparison with Experiments 1 and 2 reveals that both mAP@50 and mAP@75 are significantly improved, indicating that the CRMIC module can effectively enhance the recognition capability for small targets. In comparison with Experiment 1, Experiment 4 achieved improvements of 0.5% and 0.3% in mAP@50-95 and mAP@50, respectively. In summary, these improvements can effectively boost the detection performance.

Table 2. Ablation comparison.

NO.	Methods	mAP@50-95(%)	mAP@50 (%)	mAP@75 (%)	Params(M)	FLOPs
1	Baseline	87.5	99.0	98.3	28.2	42.7
2	Baseline+GSC	87.2	98.8	98.2	28.2	42.8
3	Baseline+ CRMIC	86.7	99.4	98.4	28.1	43.2
4	GMI-DETR	88.0	99.3	98.2	29.5	45.5

5. Conclusion

Aiming at the problems of insufficient detection accuracy, limited capability in small-target defect recognition, and poor adaptability to complex background interference in insulator defect detection for transmission lines, this paper proposes an insulator defect detection scheme integrating GSConv with multi-scale equidistant convolution, which is intended to provide high-performance, easy-to-deploy technical support for automated inspection of smart power grids. This paper primarily contributes in the following three aspects. First, by deeply integrating the grouped sparse convolution characteristics of GSConv with the cross-stage fusion structure of C2f, the GSC module is designed, which substantially reduces the number of model parameters and computational overhead, effectively addresses the issues of convolutional redundancy and single-feature expression in traditional C2f modules, and achieves improved detection accuracy. Second, a parallel architecture consisting of the CRMIC module and the GSC module is constructed, combined with a cross-scale feature fusion strategy. This fully integrates semantic information at different levels and spatial detail features, significantly enhances the model's ability to capture insulator defects with variable scales, especially small-target defects, while mitigating the interference of complex outdoor backgrounds and improving detection robustness. Third, experiments are conducted to verify the superiority of the proposed scheme. Compared with benchmark models and existing improved algorithms, the scheme proposed in this paper exhibits excellent performance in terms of detection accuracy and parameter count, and it can be adapted to low-computing-power inspection scenarios, two prominent examples being unmanned aerial vehicles and edge detection devices, demonstrating strong engineering practicability.

This research provides a new technical approach for automated insulator defect detection and has certain practical significance for promoting the intelligent and efficient upgrading of smart power grid operation and maintenance. Meanwhile, the research has certain limitations; for example, there is still room for improvement in the detection performance of the scheme under extreme weather conditions and scenarios with severe insulator contamination. Future research can be carried out in three directions. First, we can introduce attention mechanisms and adaptive feature calibration modules to further strengthen the feature response of defect regions and improve detection stability under extreme environments. Second, we can combine transfer learning methods to expand the training scope of cross-scenario and cross-type insulator defect datasets, thus strengthening the model's generalization capability. Third, we can explore model quantization and compression technologies to further reduce deployment costs, achieve adaptive integration with more lightweight inspection equipment, and promote the practical application of the technology in large-scale transmission line inspection.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] W. Yi, S. Ma, R. Li, Insulator and defect detection model based on improved yolo-s, IEEE Access 11 (2023) 93215–93226, 10.1109/ACCESS.2023.3309693.

[2] M. Zhou, B. Li, J. Wang, S. He, Fault detection method of glass insulator aerial image based on the improved yolov5, IEEE Transactions on Instrumentation and Measurement 72 (2023) 1–10, 10.1109/TIM.2023.3269099.

[3] Z.-D. Zhang, B. Zhang, Z.-C. Lan, H.-C. Liu, D.-Y. Li, L. Pei, W.-X. Yu, Finet: An insulator dataset and detection benchmark based on synthetic fog and improved yolov5, IEEE Transactions on Instrumentation and Measurement 71 (2022) 1–8, 10.1109/TIM.2022.3194909.

[4] K. Hao, G. Chen, L. Zhao, Z. Li, Y. Liu, C. Wang, An insulator defect detection model in aerial images based on multiscale feature pyramid network, IEEE Transactions on Instrumentation and Measurement 71 (2022) 1–12, 10.1109/TIM.2022.3200861.

[5] J. Liu, M. Hu, J. Dong, X. Lu, The application of a lightweight model fa-yolov5 with fused attention mechanism in insulator defect detection, Frontiers in Energy Research 11 (2023), 10.3389/fenrg.2023. 1283394.

[6] X. Tian, M. Zhang, G. Lu, Power line insulator defect detection using cnn with dense connectivity and efficient attention mechanism, MULTIMEDIA TOOLS AND APPLICATIONS 83 (10) (2024) 28305–28322, 10.1007/s11042-023-15522-7.

[7] Z. Tao, Y. He, S. Lin, T. Yi, M. Li, Snakenet: An adaptive network for small object and complex background for insulator surface defect detection, Computers and Electrical Engineering 117 (2024) 109259, https://doi.org/10.1016/j.compeleceng.2024.109259.

[8] L. Jin, W. Ding, S. Han, J. Wang, A real-time edge inference method for insulator contamination detection with yolov11-ssl, IEEE Transactions on Instrumentation and Measurement 74 (2025) 1–15, 10.1109/TIM.2025.3565254.

[9] D. Li, P. Yang, Y. Zou, Optimizing insulator defect detection with improved detr models, Mathematics .2024; 12 (10).

[10] H. Li, J. Li, H. Wei, Z. Liu, Z. Zhan, Q. Ren, Slim-neck by gsconv: a lightweight-design for real-time detector architectures, Journal of Real-Time Image Processing ,2024, 21 (3), 10.1007/s11554-024-01436-6.

[11] H. Wang, J. Peng, F. Huang, J. Wang, J. Chen, Y. Xiao, MICN: Multi-scale local and global context modeling for long-term series forecasting, in: The Eleventh International Conference on Learning Representations, 2023, https://openreview.net/forum?id=zt53IDUR1U.