Journal of Advanced Artificial Intelligence, Engineering and Technology

ISSN: 3139-4256

Research Article

WRJAAIET-25-005

A Defect Detection Method for Aerial Insulators Based on an Improved YOLOv8n Model

Yawei Zhao*, Yi Yang, Enze Li, Jialin Liu, Zhihao Tang and Jinlong He

College of Computer and Information Engineering, Hubei Normal University, Huangshi, China

***Correspondence:** Yawei Zhao, College of Computer and Information Engineering, Hubei Normal University, Huangshi, China, E-mail: zyw942688@stu.hbnu.edu.cn; DOI: https://doi.org/10.56147/aaiet.1.1.5

Citation: Zhao Y, Yang Y, Li E, Liu J, Tang Z, et al. (2025) A Defect Detection Method for Aerial Insulators Based on an Improved YOLOv8n Model. J Adv Arti Int Eng & Techn 1: 5.

Abstract

This paper proposes a defect detection method for insulators in transmission lines, addressing the numerous challenges they face in complex outdoor environments, including contamination, fine cracks and missing components. The method is based on an improved YOLOv8n model, which replaces the standard convolutional layers after the first two layers with Adaptive Downsampling layers (ADown) to enhance feature extraction capabilities through multiple convolution and pooling operations. Additionally, Dynamic Snake Convolution (DySnakeConv) is introduced into the C2f module before the Spatial Pyramid Pooling Fast (SPPF) layer, incorporating dynamic adjustment properties to more accurately adapt to and capture the complex shapes and detailed features of targets. To fully utilize the interactivity of the Path Aggregation Network (PAN), a Content-Guided Attention Fusion (CGA Fusion) mechanism layer is added before the detection head, which introduces channel, spatial and pixel-level attention information interactions to help the model better understand and handle the non-uniform fog distribution in images, thereby strengthening attention to key defect areas. Experimental results show that compared to the original YOLOv8n model, the proposed method improves detection accuracy by 5.3%, with a 0.9% increase in mAP at 0.5 IoU. Additionally, the model demonstrates enhanced convergence speed and generalization capabilities, with a parameter size of only 3M, showcasing excellent lightweight characteristics.

Keywords: Defect detection; YOLOv8n; Feature fusion; PAN; ADown

Received date: March 31, 2025; **Accepted date:** April 03, 2025; **Published date:** April 21, 2025

Introduction

Insulators are essential in power systems, supporting conductors and isolating voltage in high-voltage lines to ensure safe operation. Commonly made of ceramic, glass or composite materials, they face harsh environmental conditions, especially at high altitudes, leading to issues like surface contamination, cracks and damage [1]. Traditional manual inspection methods are time-consuming and inefficient, highlighting the need for advanced detection techniques.

To address this, deep learning methods have gained popularity for insulator defect detection. Wu et al. proposed a Faster R-CNN-based algorithm, enhancing detection accuracy for overlapping insulators but with high computational costs [2]. Wang et al. improved YOLOv5s by introducing ECA attention and ASFF for multi-scale feature fusion, boosting accuracy [3]. Other models, like the YOLOv53S-4PH and improved YOLOv7, also advanced detection but had limitations in balancing accuracy and speed [4,5].

Through the optimization and innovation of the model structure mentioned above, the new model improved in this study not only enhances detection accuracy, but also fully considers resource constraints in practice. Its lightweight and efficient characteristics make the model suitable for future inspection tasks combined with drones and it is expected to achieve real-time and accurate detection of insulator defects in complex high-altitude environments.

Journal of Advanced Artificial Intelligence, Engineering and Technology

ISSN: 3139-4256

Network Model Structure and Principles

Model introduction

YOLOv8, an advanced model in the YOLO series, features a modular design that combines a Feature Pyramid Network (FPN) with a Path Aggregation Network (PAN) to enhance multi-scale object detection. This enables the model to capture detailed information, even in complex environments. The anchor-free detection strategy further simplifies the model by eliminating preset anchors, making it more adaptable to objects of various shapes and sizes, particularly beneficial for detecting irregular defects in high-altitude insulator inspections. YOLOv8 also maintains high inference efficiency, making it suitable for real-time applications like drone-based power infrastructure inspections. Additionally, the model incorporates Dynamic Loss Balance to improve training efficiency and accuracy by adjusting classification, localization and regression errors. YOLOv8 demonstrates superior detection performance and speed, reinforcing its effectiveness in real-time defect detection tasks. The algorithm mainly consists of a backbone network, a feature fusion network (neck) and a detection head and the network structure is shown in **Figure 1**.

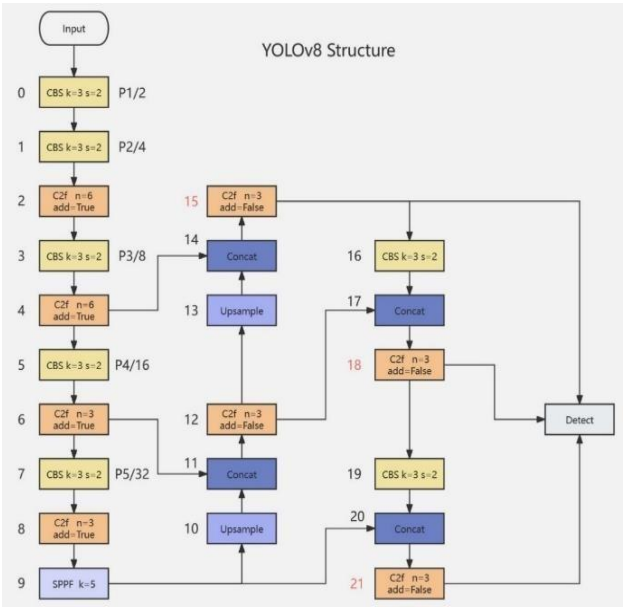

Figure 1: YOLOv8 architecture diagram.

Algorithm improvement analysis

YOLOv8n is a lightweight model within the YOLOv8 series, designed for efficient and rapid inference, making it suitable for real-time processing in resource-constrained environments. However, it has limitations in detecting complex backgrounds and small targets, particularly in high altitude insulator detection, where complex backgrounds and blurred damaged areas complicate recognition. Additionally, the model struggles to capture details such as tiny cracks and the anchor box mechanism may lead to false positives and missed detections. To address these issues, this study optimizes the backbone network, feature fusion module and detection head of YOLOv8n, enhancing the model's adaptability to complex environments and its capability to identify multi-scale targets. These improvements significantly enhance the model's overall performance in insulator detection, allowing for more precise handling of irregular shapes and small defects.

Improvements to the YOLOV8 Model

Improvements to the YOLOv8n model are made from three directions: First, Dynamic Snake Convolution (DySnakeConv) is introduced into the C2f module before the SPPF layer to enhance the model's capability to detect small targets in complex backgrounds, particularly for identifying minor defects in insulators along high-altitude power lines [6]. Second, the network structure is optimized by replacing the convolutional layers responsible for feature extraction at scales P3, P4 and P5, as well as all regular convolutional layers in the model head, with Adaptive Downsampling Layers [7]. This significantly improves multiscale target detection performance by more effectively retaining key information while reducing unnecessary detail loss. Lastly, a Content-Guided Attention Fusion mechanism (CGAFusion) is introduced before the detection head to enhance target detection performance in complex backgrounds, especially for insulator defect detection in environments such as mountains and valleys [8]. This mechanism facilitates the adaptive fusion of features across different scales, enabling the model to capture detailed characteristics of the target area more accurately.

Introduction of DySnakeConv in the bottlene-ck structure of the C2f module

DySnakeConv possesses the capability to dynamically adjust the shape and weights of convolution kernels, enabling it to accurately adapt to the morphology and detailed features of complex targets. This enhances the model's sensitivity to directional changes and nonlinear boundaries in images, thereby improving sensitivity to features from various directions and angles. The underlying principle involves combining standard convolution with snake convolutions along the x and y directions to capture multi-directional features of the input image. DySnakeConv includes three types of convolution operations: Standard convolution, snake convolution along the x direction and snake convolution along the y direction. The corresponding formulas are provided in (1), (2) and (3):

Journal of Advanced Artificial Intelligence, Engineering and Technology

ISSN: 3139-4256

$$y_0^{Cout}(i,j) = \Sigma_{c_in}^{Cin} = 1 \sum_{u=-\left|\frac{k}{2}\right|}^{\left|\frac{k}{2}\right|} \sum_{v=-|z|}^{\left|\frac{k}{2}\right|} k_0^{Cout,Cin}(u,v) \cdot x^{Cin}(i+u, j+v) \ldots (1)$$

$$y_x^{Cout}(i,j) = \Sigma_{c_in}^{Cin} = 1 \sum_{u=-\left|\frac{k}{2}\right|}^{\left|\frac{k}{2}\right|} \sum_{v=-\left|\frac{k}{2}\right|}^{\left|\frac{k}{2}\right|} k_x^{Cout,Cin}(u,v) \cdot x^{Cin}(i+u, j+v+\Delta_x(u,v)) \ldots (2)$$

$$y_y^{Cout}(i,j) = \Sigma_{c_in}^{Cin} = 1 \sum_{u=-\left|\frac{k}{2}\right|}^{\left|\frac{k}{2}\right|} \sum_{v=-\left|\frac{k}{2}\right|}^{\left|\frac{k}{2}\right|} k_y^{Cout,Cin}(u,v) \cdot x^{Cin}\left(i+u+\Delta_y(u,v), j+v\right) \ldots (3)$$

The input feature map x has a shape of $C_{in} \times H \times W$, where Cin is the number of input channels and H and W represent the height and width of the feature map. K_0, K_x and K_y are the convolution kernels with a shape of $C_{out} \times C_{in} \times k \times k$, where Cout is the number of output channels and k is the kernel size. x^{cin} refers to the c_{in}-th channel of the input feature map and $k_0^{Cout,Cin}$ denotes the standard convolution kernel from the c_{in}-th input channel to the c_{out}-th output channel. Δx(u,v) and Δy(u,v) represent the snake-like offsets along the x and y directions, respectively. y_0^{Cout}, y_x^{Cout} and y_y^{Cout} represent the outputs of the standard convolution, the snake convolution along the x direction and the snake convolution along the y direction for the c_{out}-th output channel, respectively. The final output y of the snake convolution is obtained by concatenating y_0^{Cout}, y_x^{Cout} and y_y^{Cout} along the channel dimension. The formula is given in equation (4).

$$y = concat(y_0^{Cout}, y_x^{Cout}, y_y^{Cout}, dim = 1) \ldots (4)$$

dim=1 indicates concatenation along the first channel dimension (height), resulting in a shape of $3 \cdot C_{out} \times H \times W$.

In the C2f Bottleneck module of YOLOv8, the standard convolution is replaced with DySnakeConv, renaming it Bottleneck_DySnakeConv [9]. This module contains three parallel convolution layers: a standard convolution and snake convolutions along x and y directions. The outputs are concatenated along the channel dimension. A 1×1 convolution is added to reduce the concatenated channels back to the target output. The updated structure is shown in **Figure 2**.

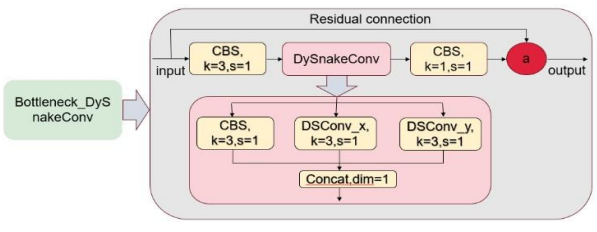

Figure 2: Bottleneck_DySnakeConv structure diagram.

By introducing snake dynamic convolution and an additional 1×1 convolution layer, feature extraction and directional awareness are improved. This enhanced structure captures diverse boundaries and nonlinear shapes more effectively, allowing for better target recognition in complex backgrounds like buildings and trees. The multi-directional extraction in DySnakeConv reduces lighting-related interference in aerial images, enhancing detail sensitivity. As a result, the C2f_DySnakeConv module improves detection accuracy in complex environments, reducing false positives and missed detections. The module's structure is shown in **Figure 3**.

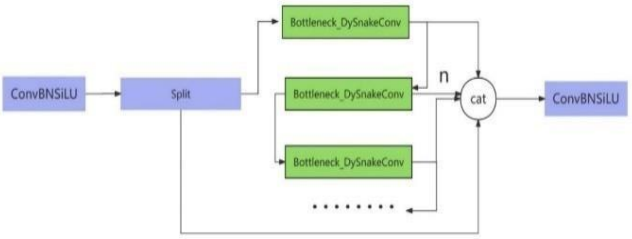

Figure 3: C2f_DySnakeConv structure diagram.

Replacement of standard conv layers with adown layers

The ADown layer adjusts down sampling parameters based on the input feature map, preserving important details, especially for small objects. It first applies average pooling to reduce spatial resolution while keeping key information. The feature map is split into $x1$ and $x2$; $x1$ undergoes a 3×3 convolution with stride 2 to capture high-dimensional features, while $x2$ goes through 3×3 max pooling followed by a 1×1 convolution. These are then concatenated to fuse features at different resolutions, enhancing multi-scale detection. The ADown layer structure is shown in **Figure 4**.

Journal of Advanced Artificial Intelligence, Engineering and Technology

ISSN: 3139-4256

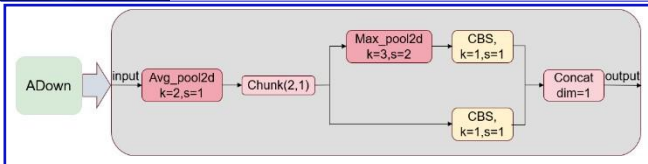

Figure 4: ADown structure diagram.

Complex aerial backgrounds often contain numerous distractions (*e.g.,* power lines and trees), which can introduce noise and lead to the loss of critical information during downsampling in traditional convolutional layers. By using adaptive downsampling layers, the model can automatically identify and filter out background noise, enhancing focus on target areas and preserving features. This reduces false positives and negatives, improving adaptability to complex backgrounds. Specifically ordinary convolutional layers in the model head are replaced with adaptive downsampling layers, enhancing feature map aggregation and ensuring a more accurate final output. Using these layers at scales P3, P4 and P5 effectively integrates multi-scale feature information, allowing the model to better identify and locate insulators in high-altitude power lines. The specific replacement locations are illustrated in **Figure 5**.

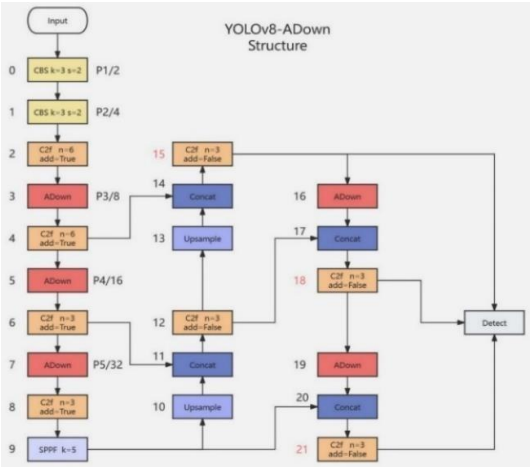

Figure 5: Replacement of CBS with ADown in YOLOv8 diagram.

Introduction of CGAFusion mechanism in the layer before the detection head

The CGAFusion mechanism integrates spatial, channel and pixel level attention to enhance feature representation by consolidating information from various sources [10-12]. It focuses on key spatial regions of the feature map to improve target localization. By initially fusing input feature maps *x* and *y* and applying weighted adjustments through attention mechanisms, CGAFusion enhances the accuracy and flexibility of feature fusion, resulting in more precise and representative features. Its structure is shown in **Figure 6**.

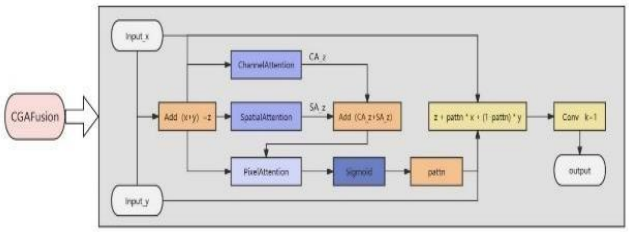

Figure 6: Internal structure of CGAFusion.

Introducing the CGAFusion mechanism before the detection head in the YOLOv8 model improves insulator detection accuracy in complex high-altitude circuits. CGAFusion allows for adaptive content-guided fusion of multi-scale features, enhancing the model's ability to capture details of small, medium and large targets. Features are extracted through P3, P4 and P5 layers and fused *via* upsampling and concatenation. Before reaching the detection head, these features undergo adaptive fusion with the original features, creating a richer multi-scale representation. This process effectively merges different resolutions, enhancing the model's precision and robustness in detecting small defects. The position of the CGAFusion module is shown in **Figure 7**.

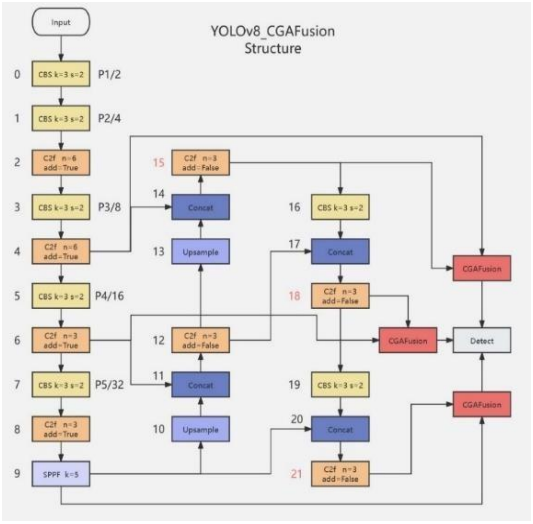

Figure 7: Location of CGAFusion introduction and selection of adaptive fusion layers.

Experiments and Analysis

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

Journal of Advanced Artificial Intelligence, Engineering and Technology

ISSN: 3139-4256

Dataset preparation

The experimental dataset comprises two parts. The first part is the publicly available China Power Line Insulator Dataset (CPLID), which contains 2,150 images divided into training, validation and test sets in a ratio of 7:2:1. The second part consists of 535 images of insulators and their defects collected from a specific region through online downloads. LabelImg was used to annotate the dataset, focusing on nine categories: Glassdirty, Glassloss, Polymer, Polymerdirty, two glass, broken-disc, insulator, pollution-flashover and snow, which cover nearly all common damage and defect scenarios for high-altitude insulators. Sample annotations and the distribution of the insulator dataset, along with anchor box analysis, are shown in **Figures 8 and 9**.

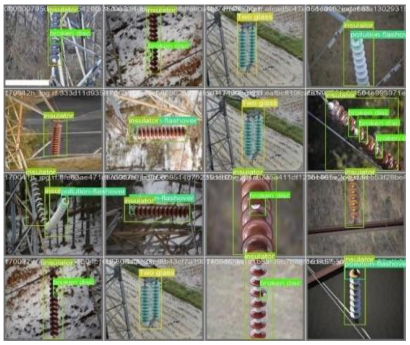

Figure 8: Sample annotation diagram.

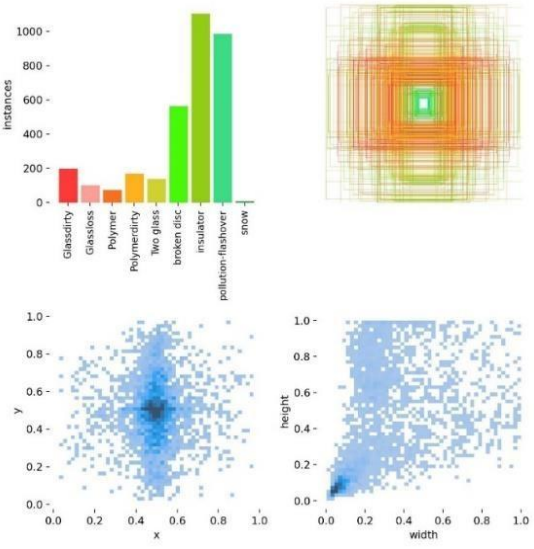

Figure 9: Insulator dataset distribution and anchor box analysis.

Experimental setup

The model training was conducted using the PyTorch framework, operating in a Windows 11 environment. The GPU used was an NVIDIA GeForce RTX 4060 Laptop GPU with 8 GB of VRAM, paired with a 13th Gen Intel® Core™ i7-13700H CPU and 16 GB of RAM. The training consisted of 300 iterations with a batch size of 32. The YOLOv8 version used was 8.1.9, with Python version 3.8.19, PyTorch version 2.0.0 and CUDA version 18. The AdamW optimizer was selected with an initial learning rate of 0.000769 and a momentum of 0.9. Automatic mixed precision was enabled to accelerate training and mosaic augmentation was disabled in the last 60 iterations to mitigate overfitting.

Evaluation metrics

The performance evaluation metrics for the network model primarily include Precision (P), Average Precision (AP), Recall (R) and mean Average Precision (mAP). Among these, mAP is a crucial metric for object detection models, as it considers the detection performance across different classes. A higher mAP value indicates better overall model performance. The calculation formulas for each category metric are as follows:

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \ldots (5)$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \ldots (6)$$

$$AP = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \ldots (6)$$

$$AP = \int_0^1 P(R)\, dR \ldots (7)$$

$$mA_P = \frac{1}{n}\sum_{i=1}^{n} APi \ldots (8)$$

In this study, TP refers to true positives (correctly predicted positive samples), FP refers to false positives (incorrectly predicted positive samples) and FN refers to false negatives (missed positive samples). Average Precision (AP) is derived from the Precision-Recall (PR) curve by calculating multiple precision and recall points at various confidence thresholds, with AP being the area under the curve.

We use mAP@0.5 as a metric for model accuracy, indicating the mean average precision at a 50% IoU threshold. Model performance is also evaluated using parameters (total learnable parameters), model size (overall dimensions) and Frames Per Second (FPS) to assess detection speed.

Ablation experiments and loss comparison experiments

Ablation experiments validated the performance of the improved YOLOv8n model for detecting insulator defects in high-altitude power facilities. Eight combinations were

Journal of Advanced Artificial Intelligence, Engineering and Technology

ISSN: 3139-4256

tested on the CPLID validation set, progressively introducing improvements like ADown, C2f_DySnake and CGAFusion. The modules used in each experimental group are marked with "✓" in the table.

Accurate insulator detection is crucial in power inspection tasks. Results showed that while CGAFusion alone slightly decreased precision, recall improved from 66.5% to 68.5%. Combining ADown and CGAFusion raised precision to 67.6%, but recall and average precision dropped significantly. However, adding the C2f_DySnakeConv module enhanced overall performance, achieving 64.5% average precision and 69.6% recall. Although inference speed decreased from 197.6 FPS to 156.3 FPS, the improvements in precision and recall compensated for this. The model's parameters and size remained low at 3.008M and 6.3M, ensuring efficient operation in resource-limited environments. The results are shown in **Table 1**.

Table 1: Ablation experiment results.

ADown	C2f_DySnakConv	CGAFusion	P/%	R/%	mAP@0.5/%	Param/M	FPS
✓	/	/	66.1	57.2	58	2.594	170
/	✓	/	64.2	61.3	60.4	3.366	109.7
/	/	✓	58.2	68.5	62.3	3.16	157.3
✓	✓	/	51.2	64.7	59.6	2.954	112
/	✓	✓	57.8	68.5	61.3	3.519	100.4
✓	/	✓	67.6	57.4	57.4	2.748	138.4
✓	✓	✓	65	69.6	64.5	3.008	156.3
/	/	/	59.5	66.5	61.6	3.007	197.6

To comprehensively assess the performance of the improved YOLOv8n model in detecting insulator defects in aerial power facilities, we compared the loss curves of the original and improved models during training and validation. The results show that while the improved model's training losses (train/box_loss, train/dfl_loss, train/cls_loss) were slightly higher, its validation losses (val/box_loss, val/dfl_loss, val/cls_loss) were lower [13]. This indicates that although the improved model may converge more slowly, it achieves better generalization and performance in defect detection, especially in complex scenarios. Loss curve comparison is shown in **Figure 10**.

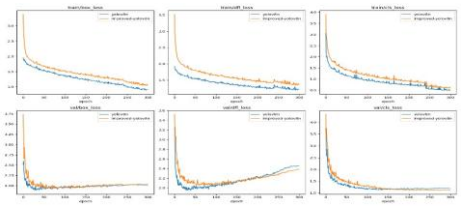

Figure 10: Comparison of loss curves between original YOLOv8n and improved YOLOv8n.

Comparative experiments

To comprehensively evaluate the performance of the improved YOLOv8n model, nine comparison experiments were designed using the test set of the CPLID dataset, including various YOLO versions, Mamba-YOLO and RT-DETR models [14,15]. To ensure validity, all experiments were conducted with lightweight models. The results show that the improved YOLOv8n outperforms mainstream models such as YOLOv6n, YOLOv8n and YOLOv9t in key metrics like precision (72.3%) and mAP@0.5 (64.6%), while maintaining a parameter size of only 3.0M, highlighting its lightweight characteristics [16,17]. In comparison, YOLOv9tiny is competitive in both detection accuracy and model size but has a much lower inference speed of 79.6 FPS, compared to the 156.3 FPS achieved by the improved YOLOv8n. Furthermore, the data reveals that models like YOLOv5n, YOLOv6n, YOLOv8n and YOLOv10n also perform well in terms of inference speed, showcasing the strength of the YOLO series in real-time detection tasks [18,19]. Based on the comparison experiments and analysis of the results, the improved YOLOv8n model strikes an optimal balance between detection accuracy, parameter size, model size and real-time processing capability, demonstrating superiority over other mainstream models. The comparison results are shown in **Table 2**.

Table 2: Comparative experiment results.

Models Structure	P/%	R/%	mAP@0.5/%	Params/M	FPS
Yolov5n	62.8	60.8	58.8	2.504	184.6
Yolov6n	66.7	60	63	4.234	211.7
Yolov7	59.2	68.5	62.3	37.24	42.3
Yolov8n	67	67.1	63.7	3.007	197.6
Yolov9tiny	67.4	68.1	63.5	1.972	79.6
Yolov10n	61	65.3	61.8	2.267	167.5
Mamba-YOLO-T	66.5	59.7	61.2	4.943	89.7
RT-DETR-l	64.5	67.5	62.7	32.002	39.2
Improved Yolov8n	72.3	63.4	64.6	3.008	156.3

Result heatmaps

In further validation experiments, a set of original images containing various damaged insulator defects was selected and defect detection was performed using both the YOLOv8n model and the improved YOLOv8n model. EigenGradCAM heatmaps were generated for comparison [20]. The heatmap generated by the YOLOv8n model shows that while the model was able to identify some damaged regions, the confidence levels were relatively low and the detection boundaries for some damaged areas

Journal of Advanced Artificial Intelligence, Engineering and Technology

ISSN: 3139-4256

were unclear, leading to some missed detections. In contrast, the heatmap produced by the improved YOLOv8n model significantly improved this situation. The model not only accurately identified the damaged regions but also displayed higher confidence within the detection boxes, indicating stronger recognition capabilities. Additionally, the detection boundaries were clearer and more precise. This improvement is primarily attributed to the optimized model architecture and more effective feature extraction, enabling the model to better capture key details in complex scenes, particularly in the task of insulator defect detection for high altitude electrical facilities. The visual comparison of the heatmaps further confirms the superior performance of the improved YOLOv8n model in real-world applications. The original defect images, YOLOv8n-generated heatmaps and improved model-generated heatmaps are shown in **Figures 11, 12 and 13**, respectively.

Figure 11: Original image collection of defects.

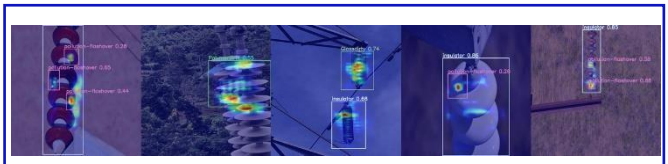

Figure 12: Heatmap generated by the YOLOv8n model.

Figure 13: Heatmap generated by the improved YOLOv8n model.

Conclusion

In light of the complex backgrounds and small defect features present in aerial images of insulators, as well as the common issues of large model sizes and high parameter counts in existing detection models, this paper proposes an improved YOLOv8n model that effectively balances detection accuracy with lightweight performance. The effectiveness of the following enhancements has been validated through ablation and comparative experiments:

The introduction of the Adown layer reduces the number of parameters and the model size, enhancing the model's capability to detect small defects in complex environments.

The DySnakeConv module is applied within the C2f module to improve feature extraction precision, allowing for better detail recognition.

The integration of the CGAFusion mechanism before the detection head enhances information interaction between channel and spatial attention, improving focus on the target area.

Experimental results indicate that the improved model shows enhanced detection accuracy and mAP@0.5 on both the validation and test sets compared to the original YOLOv8n model, with significantly improved convergence speed and generalization capability. Although there is a slight decrease in inference speed, the improvements in detection accuracy and recall compensate for this drawback. Loss comparison results demonstrate that the improved model exhibits lower loss on the validation set, indicating superior generalization ability and practical performance. Comparative experiments further confirm that the improved YOLOv8n model outperforms other mainstream algorithms (such as YOLOv6n and YOLOv5n) in terms of detection accuracy, lower miss rate and fewer parameters. In conclusion, the enhanced YOLOv8n model improves detection accuracy and robustness while maintaining low parameter counts and size, showcasing promising application prospects. Future work will consider deploying this model in UAV image processing systems for broader applications.

References

1. Wei LY, Zou GF, Zhao XY (2023) Two-stage detection method for weak feature defects of insulators in aerial images. Foreign Electronic Measurement Technology 42: 25-34.
2. Wu JP, Tang SB, Li XL (2022) Research on an improved insulator detection algorithm based on convolutional neural networks. Electrical Measurement and Instrumentation 59: 116-122.
3. Wang SZ, Ge RD, Lu JY (2024) Research on transmission line insulator defect detection method based on improved YOLOv5s. Insulators and Surge Arresters: 1-12.
4. Chen K, Liu X, Jia LJ (2024) Insulator defect detection based on lightweight networks and enhanced multi-scale feature fusion. High Voltage Technology 50: 1289-1300.
5. Wu J, Han C, Wang XL (2024) Insulator and grading ring defect detection method based on improved YOLOv7. Optoelectronics & Laser: 114.
6. Yu W, Xie B, Liu D (2024) Casdenet: Cascade automatic road detection network based on dynamic snake convolution and edge branch. IGARSS 2024 IEEE International Geoscience and Remote Sensing Symposium: 9517-9520.
7. Hesse R, Schaub-Meyer S, Roth S (2023) Content-adaptive downsampling in convolutional neural networks. Proceedings of the IEEE/CVF Conference on Computer Vision

and Pattern Recognition: 4544-4553.

8. Chen Z, He Z, Lu ZM (2024) DEA-Net: Single image dehazing based on detail-enhanced convolution and content-guided attention. IEEE Trans Image Process.

9. Varghese R, Sambath M (2024) YOLOv8: A novel object detection algorithm with enhanced performance and robustness. Int Conf Advances in Data Engineering and Intelligent Computing Systems: 1-6.

10. Zhou T, Ruan S, Vera P (2022) A tri-attention fusion guided multi modal segmentation network. Pattern Recognition 124: 108417.

11. Choi H, Yun JP, Kim BJ (2022) Attention-based multimodal image feature fusion module for transmission line detection. IEEE Trans Ind Informatics 18: 7686-7695.

12. Pham TH, Li X, Nguyen KD (2024) Seunet-trans: A simple yet effective unet-transformer model for medical image segmentation. IEEE Access.

13. Can L, Kaige W, Qing L (2023) Powerful-IoU: More straightforward and faster bounding box regression loss with a nonmonotonic focusing mechanism. Neural Networks 170: 276-284.

14. Wang Z, Li C, Xu H (2024) Mamba YOLO: SSMs-based YOLO for object detection. arXiv preprint.

15. Zhao Y, Lv W, Xu S (2024) DETRs beat YOLOs on real-time object detection. Proc IEEE/CVF Conf Comput Vis Pattern Recognit: 16965-16974.

16. Li C, Li L, Jiang H (2022) YOLOv6: A single-stage object detection framework for industrial applications. arXiv preprint.

17. Yaseen M (2024) What is YOLOv9: An in-depth exploration of the internal features of the next-generation object detector. arXiv preprint.

18. Olorunshola OE, Irhebhude ME, Evwiekpaefe AE (2023) A comparative study of YOLOv5 and YOLOv7 object detection algorithms. J Comput Social Informatics 2: 1-12.

19. Wang A, Chen H, Liu L (2024) YOLOv10: Real-time end-to-end object detection. arXiv preprint.

20. Kaczmarek E, Miguel OX, Bowie AC (2023) MetaCAM: Ensemble-based class activation map. arXiv preprint.