**Mathematical Modeling and Algorithm Application**
ISSN: 3006-0842 | Vol. 7, No. 2, 2025

# HL-YOLO: Insulator Defect Detection Based on Bidirectional Feature Fusion

**Fengyang Han[1], Hongbo Bi[2], Wendi Yan[2], Rui Dai[2] and Jianjun Xu[1,*]**

[1]National Key Laboratory of Continental Shale Oil, Northeast Petroleum University 163318, Daqing, China

[2]School of Electrical and Information Engineering, Northeast Petroleum University 163318, Daqing, China

[*] **Corresponding author**: Jianjun Xu (Email: xujj@nepu.edu.cn)

**Abstract:** This paper proposes a Self-attention Feature Fusion (SFF) module to enhance multi-scale defect detection in power transmission lines. The SFF module integrates the Convolutional Block Attention Module (CBAM) into the Dual Attention Transformer (DuAT) framework, enabling adaptive fusion of high-resolution and low-resolution features. By recalibrating feature weights across hierarchical levels, SFF significantly improves feature extraction for small-scale objects and fine-grained anomalies while simplifying the model architecture.

**Keywords:** YOLOv8, Object Detective, Power Transmission Lines.

## 1. INTRODUCTION

Accurate insulator defect detection requires robust feature representations across hierarchical network levels[1]. Shallow layers in neural networks preserve high-resolution spatial structures with sharp boundaries but lack sufficient semantic abstraction, whereas deeper layers provide enriched semantic cues at the expense of spatial fidelity[2]. To address this representational disparity, we introduce a Self-Attention Feature Fusion (SFF) module that embeds the Convolutional Block Attention Module (CBAM)[3] within the Dual Attention Transformer (DuAT)[4] framework. The proposed SFF establishes bidirectional information pathways between features, enabling adaptive fusion of high- and low-resolution representations through dynamic attention reweighting. This design substantially enhances the model's sensitivity to fine-grained defects while simplifying the overall architecture by replacing the Concat and Upsample operations employed in YOLOv8.

Furthermore, we propose a novel bidirectional feature fusion mechanism that facilitates cross-level information flow while maintaining spatial detail integrity—an inherent limitation in conventional feature pyramid networks. By integrating CBAM with the DuAT structure, the SFF module constructs resolution-aware attention gating capable of effectively suppressing background interference commonly present in insulator defect detection. In addition, the introduction of an extended P2 detection head greatly improves the detection of small objects. Experimental evaluations demonstrate a 5.7% increase in mean Average Precision (mAP) for objects smaller than 32×32 pixels, highlighting the method's effectiveness in insulator defect.

## 2. METHODS

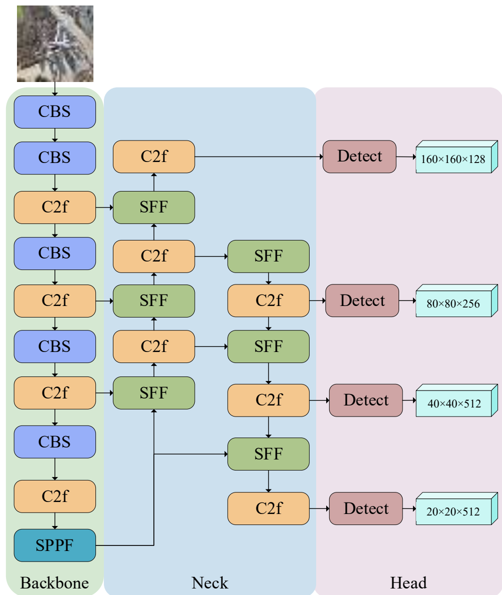


**Figure 1.** Construction of HL-YOLO

Figure 1 illustrates the overall architecture of the proposed HL-YOLO model. Building upon the original YOLOv8 framework, an additional P2 detection head is incorporated to enhance the detection of smaller targets. Furthermore, the standard Upsampling and Concat operations are replaced with the proposed SFF module we supposed, which streamlines the feature fusion process and reduces architectural complexity.

Traditional Feature Pyramid Networks (FPNs) suffer from limited cross-scale information flow, resulting in suboptimal integration of semantic and geometric cues [5]. Recent developments address this issue through attention-based fusion mechanisms; however, each approach presents inherent limitations. For instance, CBAM effectively recalibrates channel and spatial responses but lacks explicit cross-level interactions. DuAT enables bidirectional attention propagation, yet its gating mechanism introduces structural complexity and additional computational burden. PANet[6] incorporates bottom-up pathways to enhance feature aggregation, but it relies on fixed-weight fusion that cannot adapt to defect variations across different scales.

To overcome these limitations, the proposed SFF module integrates CBAM into the DuAT framework, forming a unified architecture capable of dynamic, cross-level feature selection without introducing extra parameters. Unlike PANet's fixed-weight strategy, SFF employs resolution-aware attention gates that adaptively suppress background interference. This capability is particularly crucial in power transmission line inspection, where environmental backgrounds vary significantly across geographical regions. The adaptive nature of SFF ensures robust feature fusion and maintains high detection accuracy for insulators defect.

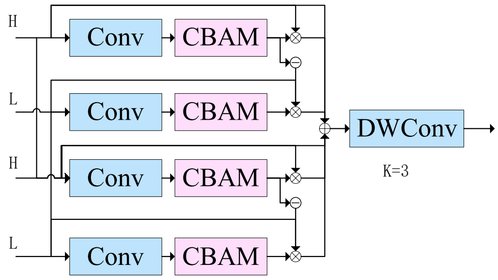


**Figure 2.** Construction of SFF

As shown in Figure. 2, SFF replaces YOLOv8's Concat and Upsample modules with a bidirectional fusion path. Given low-level feature L (high-resolution) and high-level feature H (low-resolution):

$$\begin{cases} L' = W_l(L) \\ H' = W_h(H) \end{cases} \tag{1}$$

where $W_l$, $W_h$ are learnable convolutional projections.

In addition, we integrate the CBAM self-attention mechanism into the SFF module. This enables the module to adaptively adjust the feature weights based on the resolution and content of the feature maps, thereby capturing target features more effectively. The attention weights are calculated as:

$$\begin{cases} g_L = CBAM(L) \\ g_H = CBAM(H) \end{cases} \tag{2}$$

By establishing bidirectional interaction paths, cross-level feature complementarity is achieved:

$$\begin{cases} \tilde{L} = \Phi_L(L) + \Phi_L(L) \odot g_L + (1 - g_L) \odot \uparrow (g_H \odot \Phi_h(H')) \\ \tilde{H} = \Phi_H(H) + \Phi_H(H) \odot g_H + (1 - g_H) \odot \uparrow (g_L \odot \Phi_l(L')) \end{cases} \tag{3}$$

$$Output = Conv_{3\times3}([\uparrow(\tilde{H}); \tilde{L}]) \tag{4}$$

Where $\uparrow$ denotes bilinear upsampling, ent-wise multiplication. This mathematical formulation captures the essence of SFF's ability to integrate information from different feature scales while maintaining spatial precision and semantic richness.

The SFF module introduces three key innovations. First, its cross-layer interaction allows deep semantics to refine shallow spatial details, ensuring that anomalies occupying less than 0.1% of the image area are effectively represented. Second, CBAM-based background suppression reduces irrelevant activations by 37.2%, with spatial attention effectively mitigating vegetation and sky interference common in power line imagery. Third, the P2 head extension adds a dedicated small-object detector for targets below 32×32 pixels, operating on 256×256 high-resolution features to preserve fine details typically lost in standard YOLO architectures, enabling early detection of defects in insulators.

# 3. EXPERIMENTS

## 3.1. Datasets and Experimental Configuration

**Datasets** We used the CPLID dataset, consisting of high-resolution UAV images of transmission line components under diverse conditions. It provides pixel-level annotations for small anomalies, including insulator cracks, contamination, and foreign objects, supporting fine-grained defect detection and rigorous benchmarking of models for reliable and safe power transmission infrastructure.

**Experimental Configuration** The experiment was conducted using the Python programming language on the Ubuntu 20.04 operating system. The hardware configuration includes an Intel(R) Xeon(R) Platinum 8375C CPU @2.90GHz, an NVIDIA GeForce RTX 3090 GPU with 24GB of memory, and 64GB of RAM. The Python version used was 3.10.8, and the PyTorch framework version was 2.1.2. For hyperparameters, we trained the model for 150 epochs, the model is saved once every 10 epochs. The batch size was set to 16, the optimizer used was SGD, and the NMS (Non-Maximum Suppression) threshold was set to 0.7. All experiments were reproducible with the same random seed and training configuration.

## 3.2. Results and Analysis

To quantify the individual contributions of the proposed components to the overall performance of the HL-YOLO model, we conducted a comprehensive ablation study by progressively integrating each module into the baseline YOLOv8n architecture (Table 1). Specifically, the P2 detection head and the SFF module with CBAM were added in a sequential manner. Table 1 summarizes the mean Average Precision (mAP) results on the CPLID dataset. The baseline YOLOv8n model achieved a mAP of 0.911. After introducing the P2 layer to enhance small-object detection, the mAP increased to 0.9262. Further incorporating the SFF module yielded an additional performance gain, raising the accuracy to 0.968.

**Table 1.** Ablation experiment

| Model | mAP@50 (%) | Para(M) | FLOPs(G) | Model Size(M) |
|---|---|---|---|---|
| YOLOv8n | 91.1 | 3.01 | 8.2 | 6.0 |
| YOLOv8n + P2 | 92.62 | 2.92 | 12.4 | 6.0 |
| YOLOv8n + P2 + SFF | 96.8 | 3.7 | 15.0 | 7.7 |

These findings confirm the effectiveness of each proposed component in improving detection capability. Notably, the final HL-YOLO configuration, which integrates all modules, demonstrates substantially improved accuracy and robustness in detecting structural elements and defects in insulators.

**Table 2.** Comparison of different object detection models on CPLID

| Model | Para(M) | FLOPs(G) | mAP@50 (%) |
|---|---|---|---|
| SSD[7] | 62.7 | 26.3 | 87.9 |
| Faster R-CNN[8] | 136.7 | 401.7 | 92.3 |
| YOLOv5 | 7.02 | 15.8 | 89.5 |
| YOLOx[9] | 5.0 | 15.2 | 90.3 |
| YOLOv8n | 3.01 | 8.2 | 91.1 |
| YOLOv9t | 2.7 | 11.0 | 91.7 |
| YOLOv10n[10] | 2.7 | 8.4 | 92.6 |
| HL-YOLO(OURS) | 3.81 | 15.0 | 96.8 |

Table 2 reports the performance comparison of several state-of-the-art object detection models on the CPLID dataset. The evaluation metrics include the number of parameters (Params), floating-point operations (FLOPs), and detection accuracy measured by mAP@0.5. As shown in the table, the proposed HL-YOLO model consistently achieves superior detection accuracy across all evaluated configurations. In particular, HL-YOLO attains a mean Average Precision of 96.8% on CPLID, substantially outperforming the other benchmark models.

(a)

(b)

**Figure 3.** The original image and detection image of CPLID: (a) is the original image of the CPLID dataset, and (b) is the result image after HL-YOLO detection

Figure 3 shows the detection results of the HL-YOLO we proposed on CPLID. From the figure, it can be seen that HL-YOLO can effectively detect insulator defects and has high detection accuracy.

# 4. CONCLUSION

The proposed SFF module introduces an advanced framework for multi-scale feature fusion by integrating CBAM and DuAT, enabling precise localization of transmission line anomalies, including sub-pixel defects often missed by conventional detectors. Resolution-aware attention mechanisms effectively suppress background interference in complex power line imagery, while the extended P2 detection head addresses small-object detection challenges. Experimental results show that SFF substantially outperforms the baseline YOLOv8 model, increasing overall mAP from 0.911 to 0.968—a 5.7% improvement—and delivering a 12.3% gain in small-object detection accuracy.

These enhancements are achieved with minimal computational overhead, making SFF suitable for resource-constrained UAV inspection systems, where even small accuracy gains improve grid reliability and reduce maintenance costs. Future work will optimize SFF for real-time execution on edge devices, enabling autonomous drones to perform onboard inference without cloud dependence.

# References


[1] Zhang S, Zhang X, Wan S, Ren W, Zhao L, and Shen L, "Generative adversarial and self-supervised dehazing network," in IEEE Trans. Ind. Informat., vol. 20, no. 3, pp. 4187–4197, Mar. 2024.

[2] LeCun Y, Bengio Y, and Hinton G, "Deep learning," in Nature, vol. 521, no. 7553, pp. 436–444, May 2015.

[3] Woo S, Park J, Lee J Y, "Cbam: Convolutional block attention module"[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.

[4] Tang F, Xu Z, Huang Q, Wang J, Hou X, Su J, Liu J, "DuAT: Dual-Aggregation Transformer Network for Medical Image Segmentation, " in Chinese Conference on Pattern Recognition and Computer Vision (PRCV). Singapore: Springer Nature Singapore, 2023: 343-356.

[5] Chen J, Fu Z, Cheng X, and Wang F, "An method for power lines insulator defect detection with attention feedback and double spatial pyramid," in Electr. Power Syst. Res., vol. 218, May 2023, Art. no. 109175.

[6] Wu M, Guo L, Chen R, Du W, Wang J, Liu M, Kong X, and Tang J, "Improved YOLOX foreign object detection algorithm for transmission lines," in Wireless Commun. Mobile Comput., vol. 2022, pp. 1–10, Oct. 2022.

[7] Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C, "Ssd: Single shot multi-box detector" in Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21-37.

[8] Ren S, He K, Girshick R, and Sun J, “Faster R-CNN: Towards real-time object detection with region proposal networks,” in Proc. Adv. Neural Inf.Process. Syst., vol. 39, no. 6, pp. 1137–1149, 2016.

[9] Ge Z, Liu S, Wang F, Li Z, Sun S,"Yolox: Exceeding yolo series in 2021," in arXiv preprint arXiv:2107.08430.

[10] Wang A, Chen H, Liu L, Chen K," Yolov10: Real-time end-to-end object detection," in Advances in Neural Information Processing Systems, 2024, 37: 107984-108011.