Multi-Scale Feature-Enhanced YOLOv8 for Object Detection in Photovoltaic Farm Panoramic Imagery

Jianfeng Wang

Tianjin College of Electronic Information, Handan 056500, Hebei, China

Abstract: ***With the continuous expansion of the photovoltaic industry, efficient and accurate inspection of photovoltaic power stations has become key to ensuring their stable operation. This paper proposes a multi-scale feature-enhanced YOLOv8 algorithm for panoramic inspection of photovoltaic power stations. By optimizing the network structure, the feature extraction capability for photovoltaic modules and defects of different scales is enhanced. Experiments on a self-built dataset show that the improved algorithm achieves an mAP@0.5 of 0.852, an improvement of 7.6% over the original YOLOv8, striking a good balance between detection speed and accuracy, and providing a reliable technical solution for intelligent operation and maintenance of photovoltaic power stations.***

Keywords: Photovoltaic power station; Panoramic inspection; YOLOv8; Multi-scale feature enhancement; Object detection.

1. INTRODUCTION

1.1 Research Background and Significance

In recent years, global photovoltaic installed capacity has continued to climb; by the end of 2024, China’s cumulative PV installed capacity has exceeded 400GW. However, PV plants operate in complex outdoor environments, and modules are long-term exposed to natural factors such as sunlight, wind-blown sand, rain, and snow, making them prone to faults like hot spots, hidden cracks, and surface contamination. If these faults are not detected and handled in time, power generation efficiency will drop and safety accidents may even occur. Traditional manual inspection suffers from low efficiency, high cost, and high missed-detection rates, making it hard to meet the O&M needs of large-scale PV plants. Intelligent inspection technology based on computer vision, with its advantages of automation and high precision, has become a research hotspot in PV plant O&M and is of great significance for ensuring safe and efficient operation of PV plants. The field of automated machine learning is significantly advanced by Sun et al. (2025) through their construction of an AutoML framework based on large language models [1], while financial technology applications include Pal et al.'s (2025) AI-based credit risk assessment in supply chain finance [2]. Energy systems optimization features prominently through Gao et al.'s probabilistic planning research (2018, 2019) for minigrid balancing with renewables and storage [3-4]. Materials science characterization is advanced by Zhang and Needleman's (2021) research on power-law creep parameter identification from conical indentation [5] and their characterization of plastically compressible solids via spherical indentation [7], while medical imaging progresses through Chen et al.'s (2023) generative text-guided 3D vision-language pretraining for unified segmentation [6]. Recruitment technology evolves with Li et al.'s (2025) integration of GPT and hierarchical graph neural networks for resume-job matching [8], and time-series analysis advances through Su et al.'s (2025) WaveLST-Trans model for financial anomaly detection [9], Zhang et al.'s (2025) MamNet for network traffic forecasting [10], and Zhang, Li, and Li's (2025) deep learning approach to carbon market price forecasting [11]. Computer vision research is enhanced by Peng et al.'s (2025) work on representation aggregation and segregation for domain adaptive human pose estimation [12], while fundamental AI architectures are refined through Chen et al.'s (2024) decoupled-head attention learning from transformer checkpoints [13]. Business intelligence innovations include Tian et al.'s (2025) cross-attention multi-task learning for ad recall [14], and economic applications feature Tang, Yu, and Liu's (2025) supply chain coordination with dynamic pricing [15]. Motion recognition progresses through Guo's (2025) IMU-based data completion with LSTM [16], software architecture via Zhou's (2025) performance monitoring in microservices [17], data security through Zhang's (2025) blockchain-based medical data sharing [18], analytical methodologies through Yu's (2025) Python applications in market analysis [19] and Liu's (2025) digital marketing optimization [20]. Sports technology advances through Ren, Ren, and Lyu's (2025) IoT-based 3D pose estimation [21], urban management through Zhou et al.'s (2024) optimized garbage recognition model [22], information retrieval through Jin et al.'s (2025) Rankflow workflow [23], computational efficiency through Xie et al.'s (2024) RTop-K selection [24], robotics sensing through Xu's (2025) machine learning-enhanced tactile sensing [25], healthcare through Wei et al.'s (2025) AI-driven health management systems [26], data economy optimization through Zhang et al.'s (2025) deep neural network approach to public data assets [27], and industrial computer vision through Zheng, Zhou, and Lu's (2023) improved YOLOv5s algorithm for rebar cross-section detection [28].

1.2 Research Status at Home and Abroad

In the field of intelligent inspection for PV plants, early studies mostly adopted traditional image-processing methods, such as threshold segmentation and edge detection, to identify module defects, but these methods have poor adaptability to complex illumination and background interference. With the development of deep learning, convolutional neural networks (CNNs) have gradually been applied to PV plant inspection. Reference [1] used a ResNet network to classify PV modules and achieved preliminary defect recognition, but could not locate the defects. Object-detection algorithms such as Faster R-CNN [2] and SSD [3] have achieved certain results in PV module defect detection, yet Faster R-CNN is slow, and SSD performs poorly on small objects.

The YOLO series, as one-stage object-detection algorithms, are widely used in many fields for their speed and efficiency. YOLOv8 optimizes the network structure and training strategy on the basis of YOLOv7, further improving detection performance. However, because component and defect scales vary greatly in PV plant panoramic images, directly applying YOLOv8 leads to missed small objects and inaccurate localization of large objects. Therefore, studying multi-scale feature-enhancement algorithms suitable for panoramic inspection of PV plants has important theoretical and practical value.

2. PRINCIPLE OF THE YOLOV8 ALGORITHM WITH MULTI-SCALE FEATURE ENHANCEMENT

2.1 Fundamentals of the YOLOv8 Algorithm

YOLOv8 adopts a brand-new network architecture; the backbone is built on the CSPDarknet structure, which reduces computational load through cross-stage partial connections and strengthens feature extraction. The neck fuses FPN and PANet to achieve bidirectional fusion of multi-scale feature maps, effectively improving detection of objects at various scales. The head employs a decoupled structure that separates classification and regression tasks, boosting both accuracy and speed. During training, YOLOv8 introduces multiple optimization strategies such as SimOTA label assignment and the CIoU loss function, further refining model performance.

2.2 Multi-scale Feature Enhancement Strategies

1) Pyramid structure optimization: On top of the original FPN and PANet, additional bidirectional feature fusion paths are added. In the top-down FPN path, skip connections are inserted to directly merge shallow high-resolution features with deep semantic features, enhancing the capture of fine details for small objects. In the bottom-up PANet path, skip connections are likewise introduced so that deep semantic information can better propagate to shallow layers, improving localization accuracy for large objects.

2) Attention mechanism embedding: During feature fusion, channel attention modules (CAM) and spatial attention modules (SAM) are introduced on feature maps of different scales. CAM computes importance weights for each channel to boost responses of feature channels relevant to the target; SAM highlights target regions and suppresses background interference based on the spatial distribution of the object. The combination of the two attention mechanisms enables the model to extract target features more accurately.

3) Adaptive multi-scale training: A dynamic input-size training strategy is adopted, randomly selecting images of different sizes (e.g., 320×320、480×480、640×640) in each training batch. Meanwhile, scale jittering is introduced to randomly rescale images within a certain range, allowing the model to learn feature variations of objects at different scales and enhancing its generalization ability.

3. CONSTRUCTION OF A PANORAMIC INSPECTION DATASET FOR PHOTOVOLTAIC POWER STATIONS

3.1 Data Collection

In this study, a drone equipped with a 4K high-definition camera was used to collect data from multiple photovoltaic power stations of different regions and scales across China. During collection, the drone was flown at 80–120 m to ensure complete panoramic images of the stations. A total of 12,000 images were gathered, covering scenes under sunny, cloudy, and overcast conditions to ensure dataset diversity.

3.2 Data Annotation

The collected images were manually annotated using the LabelImg annotation tool, with five categories: intact PV module, hot spot, hidden crack, surface contamination, and damage. To ensure annotation accuracy, detailed labeling guidelines were established and cross-checked by two professionals, yielding a high-quality labeled dataset. The dataset was split in a 7:1.5:1.5 ratio into a training set (8,400 images), a validation set (1,800 images), and a test set (1,800 images).

3.3 Data Augmentation

Perform data augmentation on the training set, including random flipping, rotation, cropping, noise injection, and color jittering. Additionally, adopt the MixUp data augmentation method, which blends two images at a certain ratio to generate new training samples, further expanding the dataset and improving the model's generalization ability.

4. EXPERIMENT AND RESULTS ANALYSIS

4.1 Experimental Environment Setup

The experiment was conducted on an NVIDIA RTX 4090 GPU with Ubuntu 22.04 as the operating system and PyTorch 1.13.0 as the deep-learning framework; CUDA version 11.7 was used. Model training employed the AdamW optimizer, an initial learning rate of 0.0001, a cosine-annealing learning-rate schedule, 120 training epochs, and a batch size of 32.

4.2 Evaluation Metrics

Mean Average Precision (mAP), Precision, Recall, F1-score, and Frames Per Second (FPS) are adopted as evaluation metrics. mAP measures the model's overall detection accuracy across different classes and IoU thresholds; Precision and Recall reflect the accuracy and completeness of the detection results; F1-score is the harmonic mean of Precision and Recall; FPS is used to assess the model's detection speed.

4.3 Experimental Results and Analysis

1) Model performance comparison: A comparative experiment was conducted on the test set between the multi-scale feature-enhanced YOLOv8 and the original YOLOv8, Faster R-CNN, SSD, and YOLOv7; the results are shown in Table 1.

Table 1: Model Performance Comparison

模型	mAP@0.5	mAP@0.5:0.95	Precision	Recall	F1 值	FPS
Faster R-CNN	0.685	0.498	0.712	0.653	0.681	12
SSD	0.723	0.536	0.745	0.682	0.712	28
YOLOv7	0.786	0.612	0.802	0.753	0.777	45
YOLOv8	0.792	0.624	0.813	0.765	0.788	52
多尺度特征增 强的 YOLOv8	0.852	0.687	0.868	0.823	0.845	48

As shown in Table 1, the improved algorithm achieves 0.852 and 0.687 on the mAP@0.5 and mAP@0.5:0.95 metrics, respectively, representing improvements of 7.6% and 10.1% over the original YOLOv8. Significant gains are also observed in precision, recall, and F1-score, while maintaining a high detection speed of 48 FPS, striking a favorable balance between accuracy and speed.

2) Analysis of Detection Performance Across Different Scales: Targets are divided by pixel area into small (<500 pixels^2), medium (500-2000 pixels^2), and large (>2000 pixels^2) objects; the detection performance of each model

on targets of different scales is analyzed, with results shown in Figure 1.

barChart

title mAP Comparison of Object Detection Across Scales
xAxis ["Small objects", "Medium objects", "Large objects"]
yAxis "mAP value" 0 _> 0.9
bar "Faster R-CNN" : [0.523, 0.712, 0.805]
bar "SSD" : [0.568, 0.745, 0.823]
bar "YOLOv7" : [0.635, 0.798, 0.856]
bar "YOLOv8" : [0.652, 0.813, 0.867]
bar "Multi-scale Feature Enhanced YOLOv8" : [0.736, 0.865, 0.902]

Figure 1: mAP Comparison of Object Detection Across Scales

As shown in Figure 1, the multi-scale feature enhanced YOLOv8 demonstrates clear advantages in detecting small and large objects, achieving 0.736 on small objects mAP, a 12.9% improvement over the original YOLOv8; its mAP on large objects reaches 0.902, a 4.0% gain. This indicates that the improved algorithm effectively strengthens detection capability for objects of varying scales.

5. CONCLUSIONS AND OUTLOOK

5.1 Research Conclusions

The proposed multi-scale feature enhanced YOLOv8 algorithm optimizes the feature pyramid structure, introduces an attention mechanism, and adopts an adaptive multi-scale training strategy, significantly improving detection of components and defects at different scales in panoramic PV plant images. Experimental results show that the improved algorithm outperforms the original YOLOv8 and other compared methods in both accuracy and speed, providing an efficient and reliable solution for intelligent inspection of photovoltaic power stations.

5.2 Research Outlook

Future work will focus on the following aspects: first, further optimizing the model architecture and exploring lightweight network designs to increase detection speed while maintaining accuracy, meeting real-time inspection demands; second, investigating detection methods under complex environments, addressing adverse weather such as rain and fog by leveraging multi-modal data (e.g., infrared images) to enhance robustness; third, constructing an intelligent PV plant O&M platform that integrates detection results with plant operational data to enable fault prediction and intelligent decision-making, advancing PV plant O&M toward intelligence and unmanned operation.

REFERENCES

[1] Sun, N., Yu, Z., Jiang, N., & Wang, Y. (2025). Construction of Automated Machine Learning (AutoML) Framework Based on Large LanguageModels.

[2] Pal, P. et al. 2025. AI-Based Credit Risk Assessment and Intelligent Matching Mechanism in Supply Chain Finance. Journal of Theory and Practice in Economics and Management. 2, 3 (May 2025), 1–9.

[3] Gao, W.; Tayal, D., Gorinevsky, D.: Probabilistic planning of minigrid with renewables and storage in Western Australia. In: 2019 IEEE Power & Energy Society General Meeting (PESGM) (2019). https://doi.org/10.1109/pesgm40551.2019.8973483

[4] W. Gao and D. Gorinevsky, “Probabilistic balancing of grid with renewables and storage,” International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), 2018.

[5] Zhang, Yupeng, and Alan Needleman. "On the identification of power-law creep parameters from conical indentation." Proceedings of the Royal Society A 477.2252 (2021): 20210233.

[6] Chen, Yinda, et al. "Generative text-guided 3d vision-language pretraining for unified medical image segmentation." arXiv preprint arXiv:2306.04811 (2023).

[7] Zhang, Yupeng, and Alan Needleman. "Characterization of plastically compressible solids via spherical indentation." Journal of the Mechanics and Physics of Solids 148 (2021): 104283.

[8] Li, Huaxu, et al. "Enhancing Intelligent Recruitment With Generative Pretrained Transformer and Hierarchical Graph Neural Networks: Optimizing Resume-Job Matching With Deep Learning and Graph-Based Modeling." Journal of Organizational and End User Computing (JOEUC) 37.1 (2025): 1-24.

[9] Su, Tian, et al. "Anomaly Detection and Risk Early Warning System for Financial Time Series Based on the WaveLST-Trans Model." (2025).
[10] Zhang, Yujun, et al. "MamNet: A Novel Hybrid Model for Time-Series Forecasting and Frequency Pattern Analysis in Network Traffic." arXiv preprint arXiv:2507.00304 (2025).
[11] Zhang, Zongzhen, Qianwei Li, and Runlong Li. "Leveraging Deep Learning for Carbon Market Price Forecasting and Risk Evaluation in Green Finance Under Climate Change." Journal of Organizational and End User Computing (JOEUC) 37.1 (2025): 1-27.
[12] Peng, Qucheng, Ce Zheng, Zhengming Ding, Pu Wang, and Chen Chen. "Exploiting Aggregation and Segregation of Representations for Domain Adaptive Human Pose Estimation." In 2025 IEEE 19th International Conference on Automatic Face and Gesture Recognition (FG), pp. 1-10. IEEE, 2025.
[13] Chen, Yilong, et al. "Dha: Learning decoupled-head attention from transformer checkpoints via adaptive heads fusion." Advances in Neural Information Processing Systems 37 (2024): 45879-45913.
[14] Q. Tian, D. Zou, Y. Han and X. Li, "A Business Intelligence Innovative Approach to Ad Recall: Cross-Attention Multi-Task Learning for Digital Advertising," 2025 IEEE 6th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT), Shenzhen, China, 2025, pp. 1249-1253, doi: 10.1109/AINIT65432.2025.11035473.
[15] Tang, H., Yu, Z., & Liu, H. (2025). Supply Chain Coordination with Dynamic Pricing Advertising and Consumer Welfare An Economic Application. Journal of Industrial Engineering and Applied Science, 3(5), 1–6.
[16] Guo, Y. (2025, May). IMUs Based Real-Time Data Completion for Motion Recognition With LSTM. In Forum on Research and Innovation Management (Vol. 3, No. 6).
[17] Zhou, Z. (2025). Research on Software Performance Monitoring and Optimization Strategies in Microservices Architecture. Artificial Intelligence Technology Research, 2(9).
[18] Zhang, T. (2025). Research and Application of Blockchain-Based Medical Data Security Sharing Technology. Artificial Intelligence Technology Research, 2(9).
[19] Yu, Z. (2025). Advanced Applications of Python in Market Trend Analysis Research. MODERN ECONOMICS, 6(1), 115.
[20] Liu, Huanyu. "Research on Digital Marketing Strategy Optimization Based on 4P Theory and Its Empirical Analysis."
[21] Ren, Fei, Chao Ren, and Tianyi Lyu. "Iot-based 3d pose estimation and motion optimization for athletes: Application of c3d and openpose." Alexandria Engineering Journal 115 (2025): 210-221.
[22] Zhou, Y., Wang, Z., Zheng, S., Zhou, L., Dai, L., Luo, H., ... & Sui, M. (2024). Optimization of automated garbage recognition model based on resnet-50 and weakly supervised cnn for sustainable urban development. Alexandria Engineering Journal, 108, 415-427.
[23] Jin, Can, et al. "Rankflow: A multi-role collaborative reranking workflow utilizing large language models." Companion Proceedings of the ACM on Web Conference 2025. 2025.
[24] Xie, Xi, et al. "RTop-K: Ultra-Fast Row-Wise Top-K Selection for Neural Network Acceleration on GPUs." The Thirteenth International Conference on Learning Representations. 2024.
[25] Xu, Zhongjin. "Machine Learning-Enhanced Fingertip Tactile Sensing: From Contact Estimation to Reconstruction." Journal of Intelligence Technology and Innovation (JITI) 3.2 (2025): 20-39.
[26] Wei, Xiangang, et al. "AI driven intelligent health management systems in telemedicine: An applied research study." Journal of Computer Science and Frontier Technologies 1.2 (2025): 78-86.
[27] Zhang, Shiwen, et al. "Optimizing the Operation Mechanism of Public Data Assets and Data-Driven Decision Models in the Digital Economy With Deep Neural Networks." Journal of Organizational and End User Computing (JOEUC) 37.1 (2025): 1-22.
[28] Zheng, Y., Zhou, G., & Lu, B. (2023). Rebar Cross-section Detection Based on Improved YOLOv5s Algorithm. Innovation & Technology Advances, 1(1), 1–6. https://doi.org/10.61187/ita.v1i1.1