GEORGE BROWN PRESS
JOURNALS, BOOKS & PROCEEDINGS

Journal of Computer, Signal, and System Research

Article

Deep Learning-Based Defect Detection in Photovoltaic Panels

Lele Liu [1,*]

[1] Anhui University of Science and Technology, Huainan, Anhui, China
* Correspondence: Lele Liu, Anhui University of Science and Technology, Huainan, Anhui, China

Abstract: The defect issues of photovoltaic solar panels are closely related to their efficiency and reliability. To enhance the efficiency and accuracy of defect detection in photovoltaic solar panels, this paper proposes an improved YOLOv8 model for this purpose. The improved C2f module, C2f-MS, is utilized to replace part of the original C2f structure in the model, which not only reduces the computational complexity and parameter count of the model but also enhances the extraction and fusion capabilities of multi-scale features. Additionally, NWD is incorporated into the existing CIOU to improve the detection performance for small targets, making the model's detection capabilities more balanced across various target sizes. Soft-NMS is employed to replace NMS, mitigating the issue of multiple detection boxes for a single target. Experimental results demonstrate that the improved YOLOv8 model achieves enhanced detection accuracy while reducing both the parameter count and computational complexity.

Keywords: defect detection; YOLOv8; NWD; Soft-NMS

1. Introduction

The "carbon peaking by 2030 and carbon neutrality by 2060" goals represent significant strategic decisions in China's energy transition, marking a crucial phase for the country [1]. As of the end of 2021, China's cumulative installed capacity of grid-connected photovoltaic (PV) power generation was approximately 305.9870 GW. In the new context of achieving these "dual carbon" targets, China's solar energy utilization technologies and application scenarios have entered a new period of development [2]. As a crucial renewable energy technology, the widespread application of PV panels has become an important means to reduce dependence on fossil fuels and greenhouse gas emissions. However, during actual production, PV panels may face various defects and issues caused by manufacturing processes, environmental factors, and human factors. Accurate and efficient identification of PV panel defects, identification of their causes, and proposal of solutions are crucial for improving the reliability and efficiency of PV panels.

In the industrial production of PV panels, common defects include cracks, fragmentation, scratches, and broken busbars. Considerable research has been conducted by domestic and foreign research institutions and enterprises to develop more precise and efficient detection methods for these PV defects. Manual visual inspection is a traditional method for detecting PV panel defects, involving direct observation and examination of PV panels to identify surface defects. This method is simple and intuitive but has drawbacks such as strong subjectivity, low efficiency, and inaccuracy. Especially for minute defects or internal issues, manual visual inspection often fails to meet requirements. With the gradual increase in demand for PV panels and the rapid development of computer vision technology, automated detection methods based on machine vision and deep learning have gradually become mainstream. This trend has led to improved production efficiency, reduced costs, and enhanced detection accuracy and reliability.

Previous deep learning-based algorithms for PV panel defect detection have achieved good results in detection tasks. However, there has been limited research on the detection of minute targets among PV panel defects and issues related to model parameter

Received: 01 February 2025
Revised: 04 February 2025
Accepted: 12 February 2025
Published: 13 February 2025

quantity, which are crucial for deployment on mobile devices. Addressing the inadequacy in detecting minute targets in PV panel defect detection and the issue of model parameter quantity, this paper proposes an improved YOLOv8 model to enhance the detection capability for small targets and achieve model lightweighting. Drawing inspiration from depthwise separable convolution and Scale-Aware Modulation (SAM), a new convolutional module, MS Conv (Multi-scale Convolution), is constructed and incorporated into C2f to form a new C2f structure, C2f-MS. This enhances multi-scale feature extraction capabilities and obtains more global information while reducing parameter quantity and computational complexity. Furthermore, NWD is added to the original CIoU loss function to improve the detection capability for small targets, thereby enhancing overall performance. Finally, Soft-NMS is used to replace NMS, addressing the issue of multiple detection boxes for the same target.

2. Introduction to the YOLOv8 Model

YOLOv8 represents the latest version in the YOLO series of object detection algorithms [3]. The network architecture of this algorithm comprises three main components: the backbone feature extraction network, the neck feature fusion network, and the detection head, which collaborate to achieve efficient object detection. Within the network model, the backbone feature extraction network consists of Conv, SPPF, and C2f modules. The C2f module, in particular, draws inspiration from the design concepts of the C3 module in YOLOv5 and the ELAN module in YOLOv7, obtaining richer gradient flow information while maintaining lightweight characteristics. The neck feature fusion network, composed of Upsample, Concat, and C2f, focuses on feature fusion to integrate multi-scale feature information, providing more comprehensive feature representations for subsequent object detection tasks.

The detection head adopts a mainstream decoupled-head structure, separating the classification and detection heads. Additionally, it employs an anchor-free model, directly predicting the center point, width, and height ratio of objects rather than predicting the position and size of anchor boxes. This approach reduces the number of anchor boxes, accelerates non-maximum suppression (NMS), and enhances detection speed and accuracy. This paper improves upon the YOLOv8n model by utilizing three strategies for performance optimization: incorporating a novel C2f-MS module, adding NWD to the original loss function, and replacing NMS with Soft-NMS. The specific improved YOLOv8n architecture is illustrated in Figure 1.

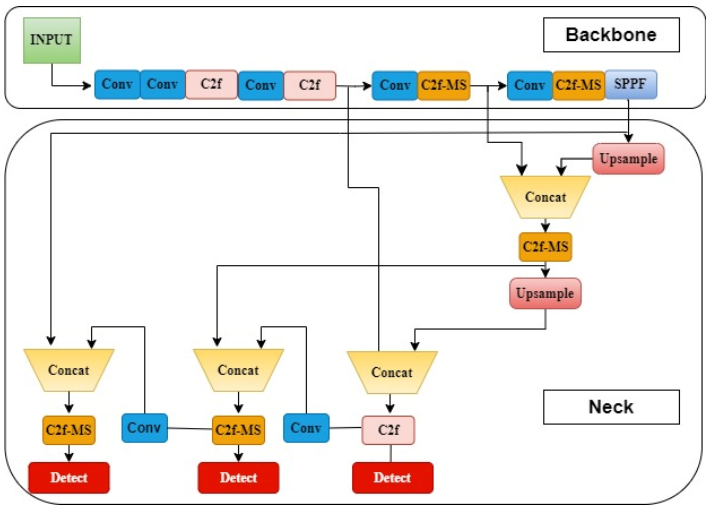

Figure 1. Improved Architecture of YOLOv8.

3. Improvements to the Algorithm

3.1. Improved C2f Module

Depthwise Separable Convolution [4] consists of Depthwise Convolution and Pointwise Convolution. Depthwise Convolution refers to convolution that does not cross channels, meaning each channel of the feature map has an independent convolutional kernel that operates solely on that channel. Depthwise Convolution is employed to perform convolutional feature extraction on the number of channels in each layer. However, these features are extracted on a single feature channel, and the information among channels is independent. Subsequently, Pointwise Convolution (realized by a 1x1 convolution) is used to fuse the features of each channel. The role of Pointwise Convolution is to increase or decrease the dimensionality of feature channels.

Scale-Aware Modulation (SAM) [5] includes two key components: MHMC (Multi-Head Mixed Convolution) and SAA (Scale-Aware Aggregation). MHMC introduces DWConv with different kernel sizes, enabling it to capture spatial features at multiple scales. SAA, designed to enhance information interaction among multiple heads in MHMC, introduces a new lightweight aggregation module. SAA reorganizes and groups the features of different granularities generated by MHMC and then uses 1x1 convolution for cross-group information fusion within and between groups, achieving lightweight and efficient aggregation.

Drawing on the concepts of Depthwise Separable Convolution and SAM, we independently apply convolutions with different kernels to the input channels. The output feature maps of different scales are then integrated and interacted through a 1x1 convolution to achieve feature fusion, constructing a new convolutional module, MS Conv (Multi-scale Convolution). The architecture of MS Conv is illustrated in Figure 2.

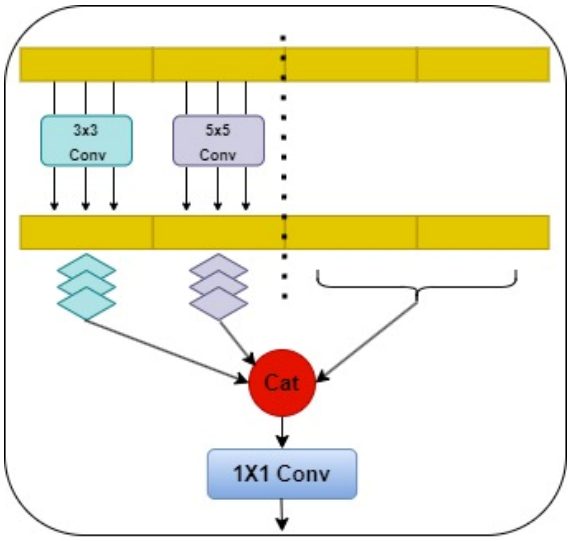

Figure 2. Architecture Diagram of MS Conv.

Half of the input channels undergo no convolution operation, while one-quarter of the input channels are subjected to 3x3 convolution, and another quarter to 5x5 convolution. Subsequently, the feature maps of different scales from these independent channels are concatenated along the channel dimension. Finally, a 1x1 convolution operation is performed to adjust the number of channels. This convolution employs kernels of different sizes, enabling it to capture various details and scale information from the input feature maps, thereby enhancing the model's ability to perceive multi-scale features. Additionally, by processing the channels of the input feature map separately, this convolution avoids redundant multiplication operations and reduces computational load. Furthermore, no convolution operation is applied to the remaining half of the channels, further decreasing computational complexity. In this paper, MS Conv is used to replace some of the Conv in

C2f, and the improved C2f-MS structure is employed to substitute the original C2f structure with more than 512 channels in the model. A comparison of the structures before and after the improvement is illustrated in Figure 3.

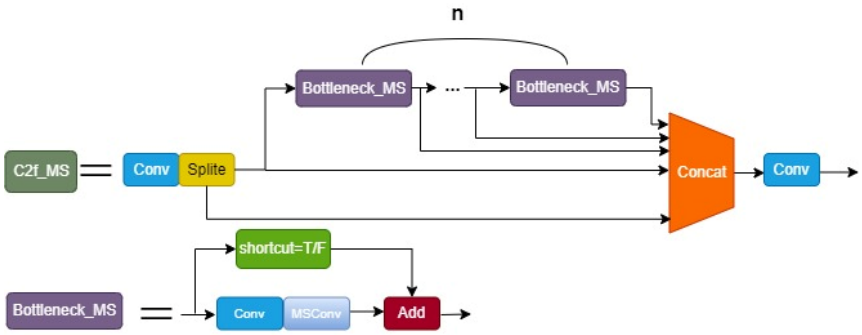

Figure 3. Improved C2f Structure Diagram.

3.2. NWD Loss Function

Intersection over Union (IoU) [6] is a commonly used metric to evaluate the performance of object detection algorithms. It measures the degree of overlap between detection results and ground truth annotations. In object detection tasks, algorithms output a series of bounding boxes to represent the detected object locations. IoU quantifies the similarity between the algorithm-generated bounding box and the ground truth bounding box by calculating the ratio of their intersection area to the union area. IoU serves two main purposes: firstly, as an evaluation metric in object detection tasks, it assesses the algorithm's accuracy in locating objects and the overall effectiveness of object detection. Typically, a detection is considered correct only if its IoU exceeds a threshold (e.g., 0.5 or 0.7). Secondly, in object detection, Non-Maximum Suppression (NMS) is often used to suppress overlapping detection results and retain the most accurate one. IoU is employed to calculate the overlap between two bounding boxes, thereby determining whether a detection result should be retained or suppressed.

Normalized Gaussian Wasserstein Distance (NWD) [7] is a normalized Gaussian distance metric based on the Wasserstein distance, used in object detection to measure the similarity between predicted bounding boxes and ground truth object bounding boxes. It models bounding boxes as two-dimensional Gaussian distributions, calculates the distance between them, and applies normalization to mitigate the impact of factors such as size and spacing.

In the photovoltaic panel defect dataset used in this paper, some targets in the labels are relatively small, leading to a tendency for missed detections during the detection process. However, considering that not all labels contain small targets, and the convergence speed of using NWD is slower than the original CIOU, we only add the NWD loss to the loss function of the original model to enhance the detection capability for small targets. By incorporating the NWD loss, the ability to detect small targets is improved, while the original CIOU [8] loss is retained, ensuring that the convergence speed is not significantly affected.

3.3. Soft-NMS

NMS (Non-Maximum Suppression) [9] is a post-processing module in object detection frameworks, primarily used to eliminate highly redundant object bounding boxes and retain only the box with the highest score within a certain region for each object category. The NMS process begins by selecting the predicted bounding box B1 with the highest confidence score among all candidate boxes as the baseline. Subsequently, all other

bounding boxes with an IOU (Intersection over Union) with B1 exceeding a predefined threshold are removed.

In the context of Soft-NMS [10], rather than directly deleting highly overlapping detected objects with high-confidence targets during the NMS process, their confidence scores are reduced. This approach allows these objects to have the opportunity to be retained as correct detection boxes in subsequent steps, thereby avoiding false detections. When implementing the score reduction strategy, a guiding principle is that the larger the IOU between a detection box and a high-confidence detection box, the greater the magnitude of the confidence score reduction should be.

Soft-NMS is suitable for addressing the issue of missed detections in dense detection scenarios caused by directly deleting highly overlapping objects during the NMS process. In this paper, Soft-NMS is utilized to replace the NMS in the original model. Soft-NMS retains more candidate boxes, which helps capture small, occluded, or densely packed objects that are prone to being overlooked. This improves the detection recall rate, reduces missed and false detections in scenarios with clustered objects, and addresses the issue of multiple detection boxes for the same object, further enhancing detection results.

4. Experimental Results and Analysis

4.1. Experimental Environment

The environmental configuration for this study is as follows: Intel(R) Xeon(R) CPU E5-2686 v4 for CPU, NVIDIA GeForce RTX 3070 Ti with 8GB of VRAM, CUDA version 11.7.0, Ubuntu 20.04.5 LTS as the operating system, PyTorch version 1.13.1, CUDA version 11.7.0 (matched with PyTorch), Python version 3.8.0, an initial learning rate (lr0) set to 0.01, momentum of 0.937, Adamw optimizer, IOU threshold of 0.7, batch size of 32, number of workers set to 8, and image size of 640 x 640 pixels.

4.2. Evaluation Criteria

The evaluation criteria in this paper include the number of parameters, GFLOPs, mAP@0.5 (mean Average Precision with an IOU threshold greater than 0.5), and mAP@[0.5:0.95]. mAP@0.5 reflects the trend of the model's precision as recall varies, with a higher value indicating that the model is more likely to maintain high precision at high recall rates. mAP@[0.5:0.95] represents the average mAP across multiple IOU thresholds, specifically calculated by taking 10 IOU thresholds within the range [0.5, 0.95] with a step size of 0.05, computing the mAP for each threshold, and then averaging these values. A higher mAP@[0.5:0.95] indicates more accurate predicted bounding boxes, as it considers a wider range of high IOU thresholds.

4.3. Ablation Study

In this paper, the network model is improved through three enhancement schemes. To investigate the impact of these enhancements on the research structure, an ablation study is conducted to analyze each enhancement point step-by-step. The specific ablation study data are presented in Table 1. The three enhancements include replacing the C2f structure with more than 512 channels with the improved C2f-MS structure, incorporating NWD into the original CIOU loss function, and replacing the NMS in the original model with Soft-NMS.

Table 1. Ablation Study Results.

C2f-MS	NWD	Soft-NMS	Parameters	GFLOPs	mAP@50	mAP@[0.5:0.95]
×	×	×	3	8.2	87	45.7
✓	×	×	2.7	7.7	88.7	45.3
×	✓	×	a	8.2	88	45.5
×	×	✓	3	8.2	88.8	49.2

✓	✓	×	2.7	7.7	89	46.1
✓	✓	✓	2.7	7.7	89.5	49.8

From the above data, it can be observed that using any single improvement can enhance the model's mAP@50 and improve its overall performance. However, when C2f-MS and NWD are used individually, although mAP@50 increases, mAP@[0.5:0.95] experiences a slight decline. When Soft-NMS is used alone, it achieves the highest improvement in both mAP@50 and mAP@[0.5:0.95] compared to the other two enhancements, with increases of 2.07% and 7.66%, respectively. Soft-NMS reduces overlapping detection results by decreasing the confidence scores between candidate bounding boxes, thereby mitigating missed and false detections in situations where targets are clustered. The combination of all three improvements yields the greatest benefits, with a 9.57% reduction in the number of parameters and a 6.1% decrease in computational complexity (GFLOPs). Under these conditions, mAP@50 increases by 2.87%, from the original 87% to 89.5%. Meanwhile, mAP@[0.5:0.95] rises by 9.0%, from 45.7% to 49.8%. The improved model enhances its overall detection performance by strengthening its multi-scale feature extraction capabilities, optimizing the loss function, and improving its detection ability in cases of target overlap.

4.4. Comparative Experiments

To objectively analyze and compare the performance of different methods in photovoltaic solar panel defect detection, the algorithm proposed in this paper is compared with YOLOv3 [11], YOLOv3-Tiny, YOLOv5n [12], YOLOv7 [13], and YOLOv7-Tiny in terms of the number of parameters, computational complexity, and mAP. The experimental results are presented in Table 2.

Table 2. Comparative Experiments.

Model	Parameters	GFLOPs	mAP@50/%	mAP@[0.5:0.95]/%
YOLOv3	103.7	283.0	89.3	47.9
YOLOv3-Tiny	12.1	19.0	86.9	43.5
YOLOv5n	2.5	7.2	87	44.1
YOLOv7	37.2	105.1	86.5	43.6
YOLOv7-Tiny	6.0	13.2	80.4	38.3
This paper	2.7	7.7	89.5	49.8

From the data in the table, it can be seen that the improved algorithm in this paper achieves the highest mAP@50 and mAP@[0.5:0.95] among all models, with an increase of 9.1% in mAP@50 and 11.5% in mAP@[0.5:0.95] compared to YOLOv7-Tiny. In terms of the number of parameters and computational complexity, the improved YOLOv8 algorithm is only 8.53% and 6.94% higher than YOLOv5n, respectively, and lower than the other algorithms. In summary, the algorithm proposed in this paper outperforms the above-mentioned algorithms in comprehensive effect and demonstrates better detection performance in photovoltaic solar panel defect detection.

5. Conclusion

This paper proposes an improved YOLOv8 model for defect detection in photovoltaic solar panels. The C2f-MS structure is used to reduce the number of model parameters and computational complexity while enhancing multi-scale feature extraction and fusion capabilities. Then, NWD is employed to optimize the original loss function, improving the detection performance for small targets and balancing detection capabilities. Finally, Soft-NMS is used to replace the original NMS, addressing the issue of multiple detection boxes for the same target. Experimental results show that the improved YOLOv8 model exhibits enhanced overall detection performance and reduced model weights, making it more convenient for deployment on mobile devices. Due to the limitations of experiments

and datasets, only three types of defects—scratches, broken grids, and dirt—were tested, and other defects were not studied. In future research, more defects will be investigated to further optimize the defect detection performance of related algorithms.

References

1. H. Sun, Q. Zhi, Y. Wang, Q. Yao, and J. Su, "China's solar photovoltaic industry development: The status quo, problems and approaches," *Appl. Energy*, vol. 118, pp. 221–230, 2014, doi: 10.1016/j.apenergy.2013.12.032.
2. X. Wu, K. Yin, and Y. Lin, "Opportunities and challenges for developing distributed photovoltaic under the emission peak and carbon neutrality goal," in *Proc. Int. Conf. Sustainable Technol. Manag. (ICSTM 2022)*, vol. 12299, pp. 25–34, Nov. 2022, doi: 10.1117/12.2644233.
3. J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91.
4. A. G. Howard, M. Zhu, B. Chen, et al., "Mobilenets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861v1, 2017.
5. W. Lin, Z. Wu, J. Chen, J. Huang and L. Jin, "Scale-Aware Modulation Meet Transformer," *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, Paris, France, 2023, pp. 5992-6003, doi: 10.1109/ICCV51070.2023.00553.
6. Yu, J., Jiang, Y., Wang, Z., et al., "Unitbox: An advanced object detection network," *Proc. 24th ACM Int. Conf. Multimedia*, 2016, pp. 516-520, doi: 10.1145/2964284.2967274.
7. J. Wang, C. Xu, W. Yang, et al., "A normalized Gaussian Wasserstein distance for tiny object detection," arXiv preprint arXiv:2110.13389v2, 2021.
8. Z. Zheng, P. Wang, W. Liu, et al., "Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression," Proc. AAAI Conf. Artif. Intell., vol. 34, no. 7, pp. 12993–13000, 2020, doi: 10.1609/aaai.v34i07.6999.
9. A. Neubeck and L. Van Gool, "Efficient Non-Maximum Suppression," *18th International Conference on Pattern Recognition (ICPR'06)*, Hong Kong, China, 2006, pp. 850-855, doi: 10.1109/ICPR.2006.479.
10. N. Bodla, B. Singh, R. Chellappa and L. S. Davis, "Soft-NMS — Improving Object Detection with One Line of Code," *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 2017, pp. 5562-5570, doi: 10.1109/ICCV.2017.593.
11. J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," arXiv preprint arXiv:1804.02767v1, 201.
12. S. Qiao, L. -C. Chen and A. Yuille, "DetectoRS: Detecting Objects with Recursive Feature Pyramid and Switchable Atrous Convolution," *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, 2021, pp. 10208-10219, doi: 10.1109/CVPR46437.2021.01008.
13. C. -Y. Wang, A. Bochkovskiy and H. -Y. M. Liao, "YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors,"*2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, BC, Canada, 2023, pp. 7464-7475, doi: 10.1109/CVPR52729.2023.00721.