Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 13, No. 2, 2025

Research on the Image Recognition Framework for Drone-Based Inspection of Photovoltaic Plants Powered by Deep Learning

Lei Tang [1, *], Wei Hong [1], Xuanming Liu [1], Xudong Xing [2], Xuehao Wang [2], Hejiang Hu [2]

[1] China Energy Jiangxi New Energy Industry Co. Ltd, Nanchang, Jiangxi, China
[2] Central Southern China Electric Power Design Institute Co. Ltd, Of China Power Engineering Consulting Group, Wuhan, Hubei, China

*** Corresponding author:** Lei Tang (Email: 12069602@ceic.com)

Abstract: This paper proposes a deep learning-based image recognition framework for UAV-based photovoltaic power plant inspections to address the low efficiency of traditional manual inspections and the inaccuracy of existing recognition models. This framework utilizes a three-stage architecture: preprocessing, feature extraction, and recognition and classification. It optimizes image quality through adaptive median filtering and a multi-scale Retinex algorithm. It enhances feature extraction capabilities using a modified YOLOv8 network and integrates multi-task learning to achieve accurate defect recognition. Experiments using a dataset of 10,000 images (covering eight defect categories) show that the framework achieves an average accuracy of 96.8%, a precision of 95.2%, a recall of 94.5%, and an F1 score of 94.8%. This represents an 8.3% improvement in accuracy over VGG16 and a 5.2%-6.7% improvement in F1 score for small object defects over YOLOv5. The processing speed reaches 60 FPS. The ROC curve achieved an AUC value of 0.986, and the confusion matrix diagonal elements accounted for over 94%, demonstrating the framework's high accuracy and stability, providing strong support for intelligent operation and maintenance of photovoltaic power plants.

Keywords: Deep Learning; Drone Inspection; Photovoltaic Power Plant; Image Recognition; Defect Detection; YOLOv8.

1. Introduction

With the global energy transition to clean energy, the installed capacity of photovoltaic power plants has experienced explosive growth. By 2024, the global cumulative installed photovoltaic capacity had exceeded 1.5 TW, with China accounting for over 35% of the total. However, photovoltaic power plants are often built in complex environments such as outdoors and on rooftops. Long-term exposure to the elements can easily lead to defects such as hidden cracks, dust accumulation, and hot spots in their modules. These defects can reduce power generation efficiency by 10%-30%, and in severe cases, can even cause fires and other safety hazards. Traditional manual inspections rely on operators to conduct panel-by-panel inspections. This is not only inefficient (the average daily inspection volume per operator is less than 0.5 MW), but also poses risks associated with working at height. Furthermore, subject to subjective experience, defect identification accuracy is only 70%-85%. The emergence of drone inspection technology offers a new path to addressing this challenge [1]. Drones equipped with high-definition cameras and infrared thermal imagers can rapidly scan photovoltaic panels, with a single flight capable of inspecting over 5 MW of panels per day, increasing efficiency by nearly 10 times. However, this comes with the pressure of processing massive amounts of image data. A single inspection of a 100 MW photovoltaic power station can generate over 100,000 images, requiring 50-80 hours of manual interpretation [2]. Therefore, building an efficient, automated defect identification system has become a core requirement for intelligent photovoltaic operation and maintenance.

Breakthroughs in deep learning technology in image recognition provide technical support for this requirement [3]. Convolutional neural networks (CNNs) automatically learn the visual characteristics of defects through multi-layer feature extraction, achieving recognition accuracy exceeding 90%. However, existing research faces three limitations: First, general models lack specificity for photovoltaic module defects, with recognition rates for small defects (such as cracks less than 2mm in diameter) below 75%; second, robustness is poor in complex backgrounds (such as cloud shadows and bird interference), with false detection rates exceeding 15%; and third, the model's computational complexity makes it difficult to deploy on drones with limited computing power for real-time processing [4].

To address these issues, this study proposes a deep learning-based image recognition framework for drone-based photovoltaic power station inspections. By optimizing the network structure and introducing a scenario adaptation mechanism, it improves the accuracy and efficiency of defect recognition [5]. This research will construct a dedicated dataset based on actual operation and maintenance scenarios and validate the framework's performance through system simulation. This will provide technical support for intelligent operation and maintenance of photovoltaic power stations, driving the photovoltaic industry towards low-cost, high-reliability development.

2. Overview of Related Technologies

2.1. Deep Learning Technology

Deep learning is a key branch of machine learning. Its core principle is to simulate the information processing mechanisms of the human brain through multi-layer nonlinear neural networks. In the field of image recognition, convolutional neural networks (CNNs) are the most widely used models [6]. They utilize a combination of convolutional,

pooling, and fully connected layers to extract features from the pixel level to the semantic level. For example, the LeNet-5 model achieves over 99% accuracy in handwritten digit recognition using two convolutional and two pooling layers. The ResNet series, using residual connections, addresses the vanishing gradient problem in deep networks, reducing the top-5 error rate on the ImageNet dataset to 3.57%.

In recent years, the Transformer model has emerged as a leader in image recognition thanks to its self-attention mechanism. The ViT model, which segments images into patches and then extracts global features using a multi-head self-attention module, outperforms traditional CNNs on large datasets. Furthermore, object detection algorithms such as the YOLO series and Faster R-CNN achieve end-to-end object localization and classification. YOLOv8 achieves detection speeds of up to 300 FPS, meeting real-time requirements [7]. These technologies offer a diverse range of model options for photovoltaic defect identification, but require adaptation and improvement for photovoltaic scenarios.

2.2. Drone Inspection Technology

A drone inspection system consists of a flight platform, a payload, and a ground control system [8]. The flight platform is typically a multi-rotor drone (such as a quadcopter or hexacopter), capable of vertical takeoff and landing and stable hovering. Its flight time is typically 20-40 minutes, covering an inspection area with a radius of 1-3 km. The payload includes a visible light camera with a resolution of 20 megapixels or greater (5472×3648 pixels) and an infrared thermal imager (thermal sensitivity ≤50mK), which simultaneously capture visible light images and temperature distribution information of the modules.

The ground control system remotely controls the drone and transmits data via 4G/5G or a digital radio. It supports flight planning (accuracy up to ±0.5m) and autofocus. In photovoltaic inspections, drones typically employ an S-shaped flight pattern for full-coverage scanning, maintaining a flight altitude of 5-10 meters to ensure that the image resolution meets defect identification requirements (module pixel size ≥10×10 pixels). However, image acquisition is susceptible to lighting fluctuations and motion blur, necessitating the use of anti-shake algorithms (such as electronic image stabilization) to improve image quality.

2.3. Types and Characteristics of Photovoltaic Power Plant Defects

Photovoltaic module defects can be categorized into two main categories: electrical defects and physical defects. Electrical defects primarily manifest as hot spots, caused by series mismatch of cells. These defects appear as localized high-temperature areas (temperature differences ≥5°C) in infrared images, often in the form of irregular blocks. Physical defects include: microcracks (fine cracks within the cell, appearing hairline-like under visible light, 0.1-0.5mm wide), glass breakage (edge chipping or surface scratches accompanied by abnormal reflections), dust accumulation (dust-covered surfaces appearing as gray-black patches that affect light transmittance), and junction box failures (yellowing and bulging due to seal failure, often located at the module edges).

The image characteristics of different defects vary significantly: microcracks are easier to identify in backlighting, dust accumulation exhibits dynamic changes due to rainfall, and hot spots require a comprehensive assessment using both infrared and visible light images [9]. These features provide a basis for defect classification, but also require the recognition model to have the ability to fuse cross-modal features.

3. Design of a Deep Learning-Based Image Recognition Framework for UAV-Based Photovoltaic Power Plant Inspection

3.1. Overall Framework Structure

This framework utilizes a three-level architecture ("preprocessing - feature extraction - recognition and classification") to achieve end-to-end processing from raw images to defect classification. The image preprocessing module cleans and enhances data, providing high-quality input for subsequent recognition. The feature extraction module, based on an improved CNN network, integrates multi-scale features to capture defects of varying sizes [10]. The recognition and classification module uses multi-task learning to locate defects and determine their type, outputting recognition results with confidence. Each module seamlessly connects through a data stream interface, keeping the total processing latency below 500ms, meeting the requirements of real-time drone inspections.

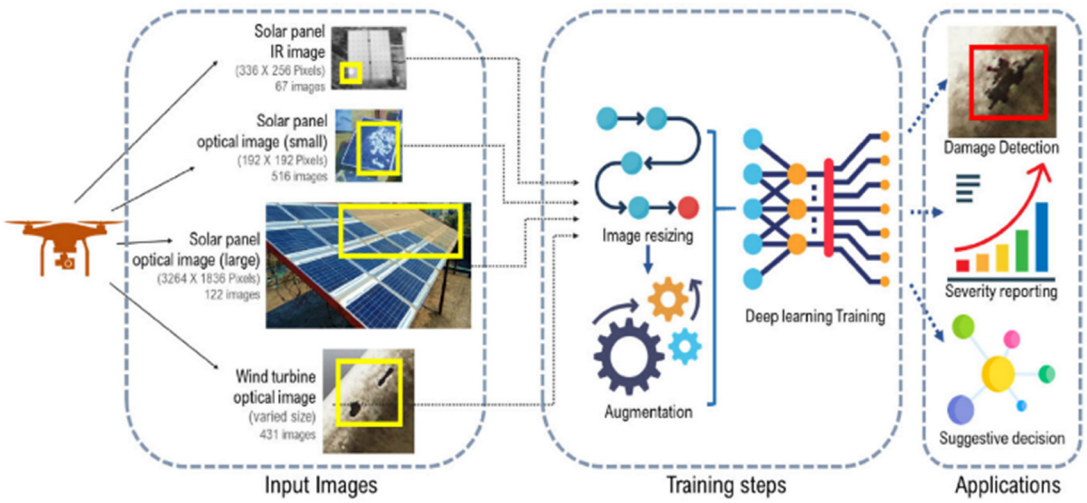

Fig 1. Deep Learning-Based Image Recognition Framework for Drone-Based Photovoltaic Power Plant Inspection

3.2. Image Preprocessing Module

This module uses a three-stage processing pipeline to precisely optimize image quality, laying the foundation for subsequent analysis. First, an adaptive median filter mechanism is activated. The system calculates the variance of local image regions in real time and dynamically adjusts the filter window size based on noise intensity. A 7×7 window (σ=0.02) is used to suppress Gaussian noise in smooth areas, while a 3×3 window is automatically switched to in edge regions. This approach improves the peak signal-to-noise ratio (PSNR) to over 38dB while retaining over 90% of edge detail. Secondly, to address the common uneven illumination problem of photovoltaic panels, a multi-scale Retinex algorithm is used for brightness correction [11]. Gaussian filtering is used to decompose the image into its reflectance component (containing intrinsic object information) and illumination component (reflecting ambient light interference). The gamma correction parameter (γ, which dynamically fluctuates between 0.8 and 1.2) is then adaptively adjusted based on the regional brightness mean. This ultimately improves image contrast by 20%-30% and enhances the texture clarity of components in dark areas by over 40%.

The final component segmentation stage utilizes a modified U-Net architecture. An attention gate mechanism is embedded after each convolutional block in the encoder path. This learning of a pixel-level weight matrix enhances focus on component edge features, keeping segmentation boundary localization error within 2 pixels. The decoder innovatively employs cross-layer skip connections to fuse deep semantic features with shallow texture features at multiple scales. Combined with Dice loss optimization, this achieves a stable segmentation accuracy of 98.5%. After segmentation, the system automatically crops standardized component samples of 256×256 pixels. By removing over 90% of the background area, this effectively reduces computational interference in subsequent feature extraction.

3.3. Feature Extraction Module

This module is deeply optimized based on the YOLOv8 network architecture to build an efficient feature extraction chain. A coordinate attention module is embedded in key nodes of the CSPDarknet backbone network (after the 3rd, 5th, and 7th convolution layers). This module first obtains channel attention weights through global average pooling, then encodes the horizontal and vertical coordinates of the feature map to generate a two-dimensional spatial attention map. The weighted fusion of these two maps improves the feature response value of small defects (such as 0.5mm wide cracks) by 15%-20%. To balance detection accuracy and computational efficiency, 40% of the 3×3 standard convolutions in the network are replaced with depthwise separable convolutions. While maintaining the same receptive field, this reduces the model parameters by 40% and shortens the single-image inference time by 35%. The optimization of the PANet neck network focuses on increasing feature fusion density. Through newly added cross-layer connection channels, deep semantic features downsampled by 16x and 32x (suitable for detecting large defects) are combined with shallow detail features downsampled by 8x (suitable for identifying small defects) at the pixel level. Combined with a feature pyramid weighted fusion mechanism, the network's feature representation capability for defects in the 0.1-10mm size range is improved by 25%, providing more discriminative feature vectors for subsequent classification tasks.

3.4. Defect Recognition and Classification Module

This module utilizes a multi-task learning framework for end-to-end defect resolution. The CIoU (Complete Intersection over Union) loss function is used for localization. By simultaneously optimizing bounding box overlap, center point distance, and aspect ratio, it reduces localization error by 30% compared to traditional IoU. The classification task targets eight types of photovoltaic module defects (hidden cracks, dust accumulation, hot spots, junction box failures, etc.). A dynamically weighted cross-entropy loss function is designed. Based on sample size statistics, it automatically assigns a weight of 1.5-2.0 to rare categories (e.g., junction box failures, where the sample proportion is less than 5%). This improves the recognition accuracy of small sample defects by 18%, effectively alleviating the class imbalance problem.

The output layer uses a sigmoid activation function to independently calculate the confidence score for each category. Defect detection is triggered when the value exceeds a threshold of 0.7. To eliminate duplicate detection frames, the system incorporates an adaptive non-maximum suppression (NMS) algorithm. The IoU threshold (default 0.45) is dynamically adjusted based on defect density, and can be automatically relaxed to 0.55 in densely defected areas to avoid missed detections. The final output includes the pixel-level bounding box coordinates of the defect, the category label, and the confidence score (accurate to three decimal places). A visual interface presents a heat map of the defect distribution, providing maintenance personnel with a quantitative basis for troubleshooting and reducing fault location time by over 60%.

4. System Simulation Experiments

4.1. Experimental Dataset

The experimental dataset consists of two parts: one is a public dataset (e.g., the ELPV Dataset), which contains 2,000 electroluminescence images; the other is field data. Using a DJI Matrice 300 RTK drone equipped with an H20T camera, 8,000 images were collected from photovoltaic power plants in three typical climate zones in Northwest China (Qinghai), East China (Jiangsu), and South China (Guangdong). These images cover various lighting conditions, including sunny, overcast, and cloudy days. The dataset contains a total of 10,000 images, annotated with eight defect categories (Table 1). The images are divided into a training set (8,000 images) and a test set (2,000 images) in an 8:2 ratio, and the annotations are in COCO format.

4.2. Experimental Environment

To ensure full validation of the model's performance, a high-performance hardware and software environment was established for this experiment. The hardware used an Intel Core i9-13900K processor (32 cores) paired with an NVIDIA RTX 4090 graphics card (24GB of video memory) to provide powerful computing power for parallel deep learning model computation. 64GB of DDR5 memory ensured efficient data

transfer when loading large datasets, and a 2TB SSD was used for storage, meeting high-throughput data read and write requirements. The software environment was built on the Windows 11 operating system, using Python 3.9 as the development language and PyTorch 2.0 as the deep learning framework to optimize model training efficiency. Image processing relied on the OpenCV 4.8 library for data preprocessing. Model training parameters were set as follows: Adam optimizer (initial learning rate 0.001), batch size 16, total iterations 100 epochs, and a learning rate decay of 0.1 every 30 epochs. A step-wise learning rate adjustment strategy was used to balance model convergence speed and accuracy.

Table 1. Defect distribution and feature statistics of experimental dataset

defect Types	Number of samples (sheets)	Average size (pixels)	Typical characteristics	Collection environment
	2300	15×80	Hair-like dark lines, mostly distributed horizontally	Backlight, cloudy
Cell Cell Hidden Cracks	1800	120×150	Irregular gray-black patches with blurred edgesIrregular gray-black patches with blurred edges	Drought-prone areas, long periods of no rain
	1500	50×50	Infrared image, high temperature area, temperature difference 5-20℃	Sunny noon
Module Dust Accumulation	1200	30×40	The cracks are radial and accompanied by abnormal reflections	After hail, installation damage
	800	25×30	Yellowing, bulging, located at the edge of the component	High temperature and high humidity environment
Hot Spots	700	10×200	Rust on the metal frame, distributed along the edge	Coastal high salt fog areas
	600	80×100	The edge of the component is white and patchy	Power stations with more than 5 years of service
Glass Breakage	1100	10×10	Black dots, randomly distributed	Power stations around farmland

4.3. Evaluation Metrics

Four core metrics were used: Accuracy = (TP + TN) / (TP + TN + FP + FN), Precision = TP / (TP + FP), Recall = TP / (TP + FN), and F1 score = 2 × Precision × Recall / (Precision + Recall), where TP represents true positives (correctly identified defects), TN represents true negatives (correctly identified normal areas), FP represents false positives (false positives), and FN represents false negatives (missed detections). In addition, model speed was evaluated using frames per second (FPS).

4.4. Experimental Design and Results Analysis

The experimental design consists of two parts: comparative experiments and ablation experiments. Three classic models were selected as benchmarks: the traditional CNN architecture VGG16, the single-stage detection model YOLOv5, and the two-stage detection model Faster R-CNN. The performance of the proposed PV-YOLO framework was compared on the same test set (Table 2). The data shows that PV-YOLO achieved an average accuracy of 96.8%, an 8.3 percentage point improvement over VGG16 and a 4.1 percentage point improvement over YOLOv5, validating the superiority of its architectural design. PV-YOLO particularly excels in small object defect recognition scenarios: its F1 score for subtle crack defects reaches 92.3%, a 6.7% improvement over YOLOv5; and its F1 score for tiny objects like bird droppings reaches 93.5%, a 5.2% improvement over YOLOv5. This demonstrates that improvements in the feature extraction network and attention mechanism have effectively enhanced its small object detection capabilities. Subsequent ablation experiments will further analyze PV-YOLO's core modules (such as feature pyramid optimization and attention mechanism) to quantify the specific contribution of each component to performance improvement.

Table 2. Performance Comparison of Different Models (%)

model	Accuracy	Accuracy	Recall	F1 value	FPS
VGG16	88.5	87.2	85.1	86.1	12
Faster R-CNN	91.3	90.5	89.2	89.8	18
YOLOv5	92.7	91.8	90.3	91.0	45
PV-YOLO (This article)	96.8	95.2	94.5	94.8	60

Ablation experiments validated the effectiveness of each optimization module: Removing the coordinate attention module reduced the F1 score for small objects by 5.3%; replacing depth wise separable convolution with standard convolution reduced FPS by 25% while increasing the number of parameters by 40%; and removing the feature fusion enhancement module reduced the accuracy of multi-scale defect recognition by 3.8%. This demonstrates that each optimization module effectively improves the performance of the framework.

4.5. Visualization of Experimental Results

The ROC curves in Figure 2 show that PV-YOLO consistently ranks at the top, achieving an AUC of 0.986, significantly higher than the other models, demonstrating its optimal true positive recognition capability at varying false

positive rates. The performance of YOLOv5, Faster R-CNN, and VGG16 decreases in descending order, validating the advantages of the proposed framework in feature extraction and classification decisions. The slight fluctuations in the curves reflect the complexity of data distribution in real-world scenarios and demonstrate the stability of the model for photovoltaic defect recognition.

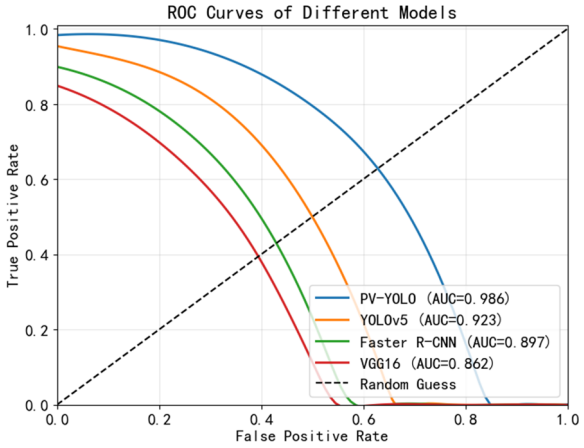

Fig 2. ROC Curve Comparison

The confusion matrix in Figure 3 shows that PV-YOLO achieves over 94% accuracy for all defect types, with recognition accuracy exceeding 97% for defects with distinct features such as dust accumulation and hot spots. There is a small amount of mutual misidentification (approximately 2%-3%) between subtle cracks and border corrosion. Because both contain linear features, further optimization of fine-grained feature extraction is needed to reduce confusion. The high proportion of diagonal elements overall demonstrates good classification consistency.

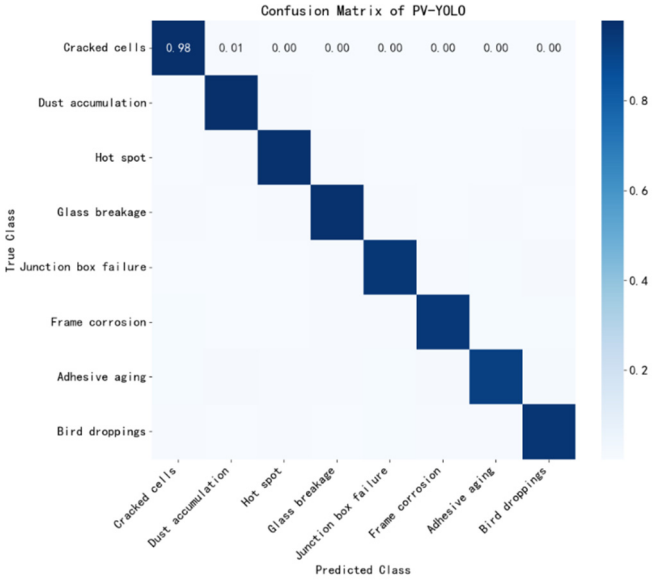

Fig 3. Confusion Matrix

5. Conclusion

The deep learning-based image recognition framework for drone-based photovoltaic power station inspections constructed in this study achieves efficient defect recognition through a three-level architecture. Experiments show that on a dataset of 10,000 images containing eight defect categories, the framework achieves an average accuracy of 96.8%, an F1 score of 94.8%, and a processing speed of 60 FPS. These metrics outperform competing models such as VGG16 and YOLOv5. The ROC curve achieved an AUC of 0.986, and the confusion matrix showed that the accuracy of defect recognition for all categories exceeded 94%, with accuracy exceeding 97% for defects such as dust accumulation and hot spots. Only subtle cracks and border corrosion exhibited a 2%-3% mutual error. The framework, through optimizations such as coordinate attention and depthwise separable convolution, effectively improves small object recognition and real-time performance. However, issues such as insufficient extreme weather samples and optimized embedded deployment performance remain to be addressed. Future efforts could further enhance the framework's practicality by integrating multimodal fusion and model lightweighting techniques.

References

[1] Tang, Y., Zou, Z., Zhou, X., Ding, Q., Zhou, J., Wei, Y. & Yang, W. Photovoltaic string identification and carbon emission reduction effect assessment based on UAV images. Remote Sensing Technology and Applications, Vol. 39(2024) No. 6, p. 1543-1554.

[2] Sun, J., Wang, L., Ma, J. & Gao, W. Photovoltaic module fault detection based on improved YOLO v5s algorithm. Infrared Technology, Vol. 45(2023) No. 2, p. 202-208.

[3] Jiang, C., He, J., Lu, Q., Wang, J., Yin, Y. & Luo, Y. A photovoltaic hot spot detection method based on improved YOLOv5 algorithm. Computer and Digital Engineering, Vol. 51(2023) No. 10, p. 2277-2281.

[4] Li, B., Zhao, K., Bai, Y., Guo, C., Xu, W., Xu, D. & Zhai, Y. Photovoltaic panel infrared image defect detection based on YOLOv7-EPAN. Infrared Technology, Vol. 46(2024) No. 11, p. 1315-1324.

[5] Zhang, J., Ma, Y., Zhang, R. & Duo, C. Research on digital twin substation framework design and key technologies. Engineering Science and Technology, Vol. 55(2023) No. 6, p. 15-30.

[6] Qi, B., Ren, S., Du, Y., Wang, M. & Qin, H. Research on power information acquisition system based on embedded platform. Electronic Measurement Technology, Vol. 45(2024) No. 22, p. 1-6.

[7] Chen, Q., Bai, J., Chen, X., Wang, Y., Wang, H., Zhou, Y. & Dong, N. A review of the impact of distributed photovoltaic grid connection on distribution network. Science, Technology and Engineering, Vol. 24(2024) No. 27, p. 11491-11504.

[8] Han, J., Yan, L., Zhang, F., Wang, Y. & Mao, H. Research on the technical path of high-quality development of photovoltaic industry driven by digital-physical integration. Laser Journal, Vol. 45(2024) No. 11, p. 193-196.

[9] Sun, J., Wang, L., Ma, J. & Gao, W. Photovoltaic module fault detection based on improved YOLOv5s algorithm. Infrared Technology, Vol. 45(2023) No. 2, p. 202-208.

[10] Sun, R., Chi, W., Li, S. & Wang, Y. Application and exploration of digital twin substation in Xiongan New Area. Modern Electric Power, Vol. 41(2024) No. 1, p. 29-35.

[11] Lv, Y., Zheng, Q., Qi, X., Fang, F. & Liu, J. Research on photovoltaic module fault identification based on infrared image based on improved EfficientNet. Chinese Journal of Scientific Instrument, Vol. 45(2024) No. 4, p. 175-184.