PV Module Fault Diagnosis Based on *I–V* Curve Deformation Characteristics and Computer Vision

Lingxiang Zhang, *Student Member, IEEE*, Kun Xia, *Member, IEEE*, Po Xu, Yibo Yu, and Yiming Wang

***Abstract*—Early diagnosis of photovoltaic module faults is crucial to reducing maintenance costs. The *I–V* curve can be intuitively used for diagnosis, but traditionally, only certain feature points on the curve are extracted, neglecting the entire curve's deformation characteristics. This article analyzes the *I–V* curves deformation characteristics based on three-diode model, and presents a novel diagnosis method utilizing *I–V* curve and computer vision. The voltage, current, and power corresponding to each point on the *I–V* curve is mapped to the three channels in RGB format to form a color image. In this way, the complete deformation features of the *I–V* curve can be captured and converted into color features on the image. The improved multiaxis vision transformer (I-MaxViT) is proposed for fault diagnosis on the color image by introducing the scaled cosine attention mechanism and log-spaced continuous position deviation into MaxViT block, which reduces the model's computational complexity. Experimental results demonstrate that the proposed innovative diagnostic method achieves the highest diagnostic accuracy, reaching 94%, when compared to the other six diagnostic algorithms, and shows better universality.**

***Index Terms*—Computer vision, fault diagnosis, *I–V* curve, photovoltaic.**

I. Introduction

PHOTOVOLTAIC (PV) module is the primary components of PV power generation systems and often deployed in harsh outdoor environments. Consequently, early fault diagnosis for these modules is crucial for the overall system maintenance and cost-savings [1].

Received 19 May 2025; revised 22 July 2025; accepted 8 September 2025. Date of publication 10 September 2025; date of current version 23 December 2025. This work was supported by the Open Project of the National Local Joint Engineering Laboratory of High Energy Saving Motor and Control Technology of Anhui University under Grant KFKT202105. Recommended for publication by Associate Editor M. S. ElMoursi. *(Corresponding author: Kun Xia.)*

Lingxiang Zhang is with the School of Automation, Beijing Institute of Technology, Beijing 100081, China, also with the Department of Electrical Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China, and also with the National Local Joint Engineering Laboratory of High Energy Saving Motor and Control Technology, Anhui University, Hefei 230039, China (e-mail: 3120245549@bit.edu.cn).

Kun Xia is with the Department of Electrical Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China (e-mail: xiakun@usst.edu.cn).

Po Xu and Yiming Wang are with Ginlong Technologies Company Ltd., Ningbo 315000, China (e-mail: damon@ginlong.com; office@ginlong.com).

Yibo Yu is with the Mathematical Analytics and Operations Research College of Letters and Science, University of California, Davis, CA 95616 USA (e-mail: yybyu@ucdavis.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TPEL.2025.3608823.

Digital Object Identifier 10.1109/TPEL.2025.3608823

Considering the deployment cost and sunshine conditions of PV arrays, PV arrays are typically built on roofs, water bodies, grasslands, and mountains, which also brings some security risks. Numerous factors can reduce the output power of PV arrays, such as internal failures (e.g., short-circuit and degradation) and the external environmental factors (e.g., dust, leaves, and bird droppings) [2], [3]. Faults caused by the modules themselves are usually harder to detect, while those caused by the external factors may be detected by naked eye. However, manual inspection requires significant time and manpower due to the large number of PV modules and the remote and harsh deployment sites. At present, the data mainly utilized in the fault diagnosis of PV modules includes image data (infrared and electroluminescence images) and electrical data (general electrical data and *I–V* curves) [4], [5], [6].

The infrared imaging method, which utilizes the temperature differences between normal and faulty modules to diagnose faults, can accurately identify faults. Additionally, infrared imaging technology can be combined with autonomous aerial vehicles (AAVs). However, the types of faults that infrared diagnosis can detect are limited, and the high costs of infrared cameras and AAVs restrict their widespread adoption [7], [8], [9], [10], [11]. In [7], an embedded system combined thermal imaging with convolutional neural networks (CNNs) for fault detection. Although it supports automated analysis, the high computational demands of CNNs restrict deployment on resource-constrained embedded platforms. In [8], a diagnosis approach integrating infrared imaging with deep neural networks was introduced to classify faults. Yet, noise and resolution limitations in thermal images could severely affect diagnostic accuracy, with low-quality inputs may lead to misclassification or missed faults. In [9], infrared images are analyzed using a vision transformer model with attention mechanisms for fault detection. This approach can capture global feature relationships but requires a large labeled dataset, is computationally intensive, and is applied offline rather than in real time. In [10], a combination of Lite-UNet segmentation and a cross-channel feature aggregation detection network is applied for hot-spot fault detection. The method enhances detection of dense and small faults, yet it relies on infrared imaging hardware, involves complex network models, and remains limited to hot-spot faults. In [11], a AAV-mounted infrared imaging system is combined with panel segmentation and a detection network enhanced by angle perception and model migration to detect hot-spot faults. While this method effectively addresses variations in viewing angle and scale, it relies on AAV-based infrared acquisition,

results in longer inference times, and is still restricted to hot-spot fault detection.

The electroluminescence (EL) imaging method allows for the visualization of internal structural and electrical anomalies in PV modules when subjected to electrical excitation, enabling high-resolution detection of fine defects, such as microcracks and broken fingers, which may often be overlooked by infrared or visual inspection. However, EL imaging requires controlled conditions and specialized equipment, making it less efficient for large-scale or outdoor applications and increasing costs [12], [13], [14]. In [12], a large-scale dataset of EL images is used to train and evaluate deep learning models for fault detection. This approach relies on EL imaging equipment and is limited to controlled environments, while the detectable fault types in EL images are relatively few, and the data acquisition process is costly. In [13], EL images are analyzed using a weakly-supervised anomaly detection framework based on feature map conversion and hypersphere transformation, which reduces the need for extensive labeled datasets and improves robustness but involves a complex training process and requires prior knowledge. In [14], EL images are analyzed using a hybrid model that combines CNNs with ensemble learning algorithms through a weighted voting mechanism, which improves classification accuracy and robustness but increases computational complexity and requires high-resolution EL images.

The electrical measurement method is to detect faulty modules by comparing and analyzing the operating states or electrical parameters of the modules, which may require a large number of sensors and increase system costs [15], [16], [17], [18], [19]. In [15], a real-time PV monitoring system uses ZigBee to collect electrical parameters and transmits them via 4G to a remote server for processing and storage. While effective for remote monitoring, its performance can degrade due to signal attenuation and interference in large-scale installations. Additionally, limited 4G coverage may restrict use in remote areas. In [16], an online fault detect method was proposed that utilizes dynamic current and voltage without interrupting operation, which enables real-time online detection but is limited to detecting cracks and could not detect other faults. In [17], an improved reflectometry-based fault location method was proposed that utilizes injected step signals and reflected waveforms, which enables accurate fault localization in large-scale PV strings but is limited to electrical faults and shows reduced accuracy for faults located far from the measurement point. In [18], power and voltage are utilized to construct a fault index matrix combined with a neural network, allowing for detection, classification, and localization of multiple faults, but it requires high-quality data and has limited sensitivity to small local faults. In [19], a bipolar junction transistor and a Zener diode are employed to collect electrical data, combined with a semisupervised learning algorithm for fault detection. This approach enables panel-level monitoring without additional traditional sensors, but it requires careful calibration, is sensitive to operating conditions, and incurs additional hardware costs.

With the rapid development of photovoltaics, many PV power stations are now equipped with cloud platform management systems [20]. The *I–V* curve of PV arrays could be easily collected by inverters and uploaded to the cloud in real time for maintenance personnel to monitor and maintain, so the fault diagnosis methods based on *I–V* curves are becoming increasingly popular [21]. However, existing research reports typically use only partial points from the *I–V* curves for diagnosis. For example, open circuit voltage, short-circuit current, and maximum power point are collected and used as input variables of neural network-based fault diagnosis [22], [23], [24], [25], [26]. In [22], a variable prediction model estimated ideal *I–V* curves under normal conditions, with faults detected by comparing predicted and measured values. The method is effective but relies heavily on precise parameter estimation, which may limit its applicability in real-world scenarios. In [23], fault detection based on *I–V* curve convexity and slope was introduced to identify current mismatch. However, its high computation time restricts real-time applications. In [24], a voltage-to-current conversion method was proposed to locate and classify faults. Nevertheless, environmental variations often distort *I–V* characteristics, reducing diagnostic accuracy. In [25], the output power, short-circuit current, and open-circuit voltage extracted from the *I–V* curve are input into an ANN model to diagnose bypass diode faults, achieving high accuracy in identifying both open- and short-circuit conditions. However, the method was validated only on small-scale systems, and its performance in large, diverse field conditions remains to be further verified. In [26], key features extracted from the *I–V* curve, along with the fill factor, are input into machine learning models for fault detection and classification, achieving high accuracy and robustness but relying on high-quality data, are sensitive to noise, and facing challenges in complex and novel situations.

The *I–V* curve contains much richer information beyond the typical feature points such as the open circuit voltage (V_{oc}), short circuit current (I_{sc}), and maximum power point (MPP). For example, when a PV array is aging or experiencing shadow occlusion, the values of the open circuit voltage (V_{oc}), short circuit current (I_{sc}), and maximum power point on the *I–V* curve will change compared with the normal *I–V* curve. However, the V_{oc}, I_{sc}, and MPP on the ageing *I–V* curve and the shadow *I–V* curve may vary similarly. That is, V_{oc}, I_{sc}, and MPP may be the same when aging faults and occlusion faults occur making it difficult to determine, which fault has occurred. But the deformation between these points is different, so the fault diagnosis can be easily achieved from the overall deformation of the *I–V* curve. Additionally, for PV arrays of varying scales, the same V_{oc}, I_{sc}, and MPP values may correspond to entirely different health states, requiring separate model training for each PV station configuration. This significantly limits the generalizability of key-point-based diagnostic methods.

To address the issues discussed above, this article presents a new fault diagnosis method that integrates *I–V* curve with computer vision. First, the deformation characteristics of *I–V* curves under different faults are analyzed. Then, a new mapping strategy is introduced to convert the electrical parameters of each point on the *I–V* curve into corresponding RGB pixels, forming a color image. This transformation enables complex curve deformations to be captured as visual features. These

TABLE I
COMPARISON OF MENTIONED METHODS

Ref.	Data type	Pros	Cons
[7], [8], [9], [10], [11]	Infrared image	• Non-contact	• Limited defect types • Low precision • High cost
[12], [13], [14]	EL image	• High-resolution	• Limited defect types • Strict conditions are required
[15], [16], [17], [18], [19]	General electrical data	• Simple operation	• Many sensors • High cost
[22], [23], [24], [25], [26]	*I–V* curve	• Suitable for multiple faults	• Partial feature • Poor generalization • capability
Method proposed in this article		• Feature extraction integrity • High precision • Small computation • Excellent generalization	Since it was first proposed, it has not been extensively tested in engineering.

images are subsequently classified using the proposed I-MaxViT model. Unlike traditional neural networks constrained by limited receptive fields, I-MaxViT incorporates a self-attention mechanism, enabling global context awareness and enhancing its ability to capture subtle variations in color patterns. This makes it particularly well-suited for fault classification tasks based on color-encoded *I–V* images. Finally, experiments are conducted to validate the proposed method in terms of diagnostic accuracy and generalization performance.

The main contributions of the proposed method are as follows.

1) Converting (U, I, P) corresponding to all data points on *I–V* curve into a pixel point (R, G, B) and forms a color image that the electrical parameters on an *I–V* curve are converted into color image features, so that the complete deformation characteristics on the *I–V* curve could be captured and visualized into a more intuitive color image, allowing more detailed features to be obtained and then be better diagnosed by computer vision.
2) Proposing the I-MaxViT for fault diagnosis of color images, introducing the scaled cosine attention mechanism and log-spaced continuous position deviation into MaxViT block to reduce the computational complexity of model.

The pros and cons of the mentioned literatures are summarized in Table I.

The rest of this article is organized as follows. Section II presents the construction of the PV array model and a thorough analysis of the *I–V* curve characteristics under various fault types. Section III introduces the proposed fault diagnostic method, which integrates full *I–V* curve characteristics with computer vision. Section IV details the experimental verification. Finally, Section V concludes this article.

II. PV MODULE MODEL BUILDING AND FAULT ANALYSIS

To provide clearer understanding of the *I–V* curve characteristics under various fault conditions, this section models PV modules. The PV array was constructed in MATLAB/Simulink using the established PV module model. Through simulation, the *I–V* curves of the PV array under different fault conditions were obtained and analyzed for their characteristics.

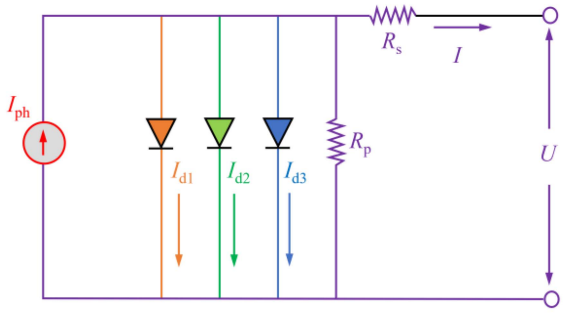

Fig. 1. Equivalent model of three-diode.

A. Proposed PV Three-Diode Model

The mathematical models of PV modules typically employ single-diode, double-diode, and three-diode models (TDM). The single-diode model addresses loss issues in the quasi-neutral region but overlooks composite losses in the space charge region at low irradiance, reducing its accuracy. The double-diode model accounts for losses in both the quasi-neutral and space charge regions, yet it has limitations regarding surface recombination region and other nonideal behaviors [27], [28]. The TDM achieves the highest modeling accuracy because it comprehensively considers the losses of these three regions [29]. In this study, the focus is on analyzing the deformation characteristics of the *I–V* curve across a wide range of temperatures and irradiance levels (including low irradiance), and verifying the universality of the proposed method. Therefore, the TDM was adopted for the modeling of PV modules to ensure the accuracy of the *I–V* curves acquired in extreme environments, as illustrated in Fig. 1.

According to Kirchhoff's current law, the mathematical expressions for current I and branch current $I_{\mathrm{d}i}$ are shown in

$$I = I_{ph} - \sum_{i=1}^{3} I_{oi}\left[exp\frac{(U+IR_s)q}{n_i kT} - 1\right] - \frac{U+IR_s}{R_p} \quad (1)$$

$$I_{\mathrm{d}i} = \sum_{i=1}^{3} I_{oi}\left[exp\frac{(U+IR_s)q}{n_i kT} - 1\right] \quad (2)$$

where I and U are the output current and voltage of the PV module, I_{ph} is the photogenerated current, $I_{\mathrm{d}i}$ is the current in the diodes, $I_{\mathrm{o}i}$ is the reverse saturation current of the diodes, and R_{s} and R_{p} are the equivalent series resistance and parallel resistance, respectively. n_i is the ideality factor of each diode, k is the Boltzmann constant (1.38065e-23 J/K), T is the temperature of the PV module, and q is the electron charge (1.6022e-19 C).

Equation (1) is the transcendental equation of the TDM for the PV module. It could be utilized in MATLAB/Simulink to model a PV module, obtain the relationship between output voltage and current, and generate the *I–V* curve. By identifying the parameters I_{ph}, $I_{\mathrm{o}1}$, $I_{\mathrm{o}2}$, $I_{\mathrm{o}3}$, n_1, n_2, n_3, R_{s}, and R_{p} in the transcendental equation, an accurate PV module model based on TDM can be obtained. Equation (1) is the transcendental equation under standard conditions. As operating conditions change,

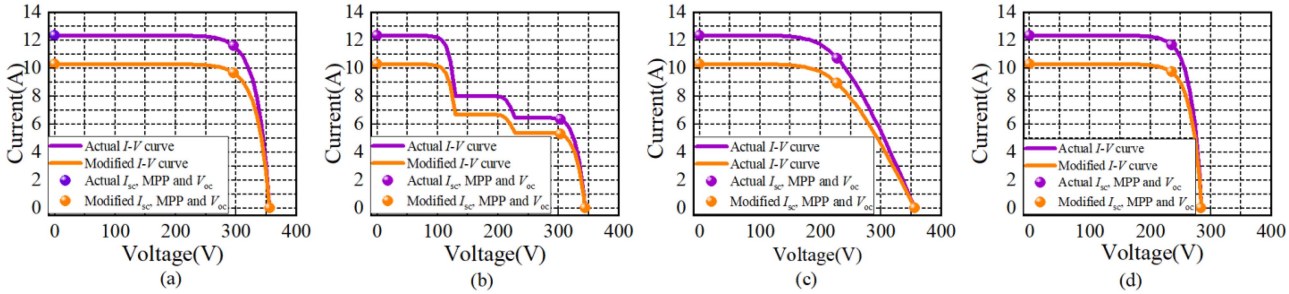

Fig. 2. *I–V* curves before and after correction under varying operating conditions. (a) *I–V* curves under normal conditions. (b) *I–V* curves under shading conditions. (c) *I–V* curves under degradation conditions. (d) *I–V* curves under short-circuit conditions.

I_{ph}, $I_{\mathrm{o}i}$, E_{g}, and R_{p} are affected by variations in temperature and irradiance.

The I_{ph} is determined by the rate at which incident photons generate electron-hole pairs. It is primarily influenced by irradiance, with a slight dependency on temperature. I_{ph} changes linearly with irradiance and is only minimally affected by temperature. The reference photocurrent under standard test conditions is adjusted to the actual irradiance level using the ratio G_e/G_{STC}, while temperature has a minor effect on carrier mobility, the coefficient k_i reflects the variations in carrier mobility and diffusion length with temperature. As the temperature increases, enhanced lattice vibrations can slightly improve carrier collection efficiency. The value of k_i could be obtained from manufacturer data or determined through experimental calibration.

Since the carriers in semiconductors and the energy required to cross the bandgap are temperature-dependent, $I_{\mathrm{o}i}$ exhibits an exponential increase with temperature and is also influenced by E_{g}. As temperature increases, E_{g} decreases linearly. An empirical value commonly used in engineering is that for every 1 K increase in temperature, E_{g} decreases by 0.2677 meV.

The parameter R_{p} represents the parasitic leakage path within the PV cell. Its value is affected by irradiance; as irradiance increases, minority carrier recombination, and dark current are enhanced, leading to a decrease in R_{p}.

I_{ph}, $I_{\mathrm{o}i}$, E_{g}, and R_{p} under actual working conditions need to be further obtained by environmental modified formula, as illustrated in (3), which account for the impact of environment on the parameters [30]. By applying the modified formula, the *I–V* curves under different working conditions are derived, facilitating analysis and evaluation

$$
\begin{aligned}
I_{\mathrm{ph}} &= \frac{G_e}{G_{STC}}\left[I_{\mathrm{ph\cdot STC}} + k_i(T_e - T_{STC})\right] \\
I_{oi} &= I_{oi\cdot STC}\left(\frac{T_e}{T_{STC}}\right)^3 \exp\left(\frac{E_{\mathrm{g\cdot STC}}}{kT_{STC}} - \frac{E_{\mathrm{g}}}{kT_e}\right) \\
E_{\mathrm{g}} &= E_{\mathrm{g\cdot STC}}\left[1 - 0.0002677(T_e - T_{STC})\right] \\
R_p &= R_{p\cdot STC}\frac{G_{STC}}{G_e}
\end{aligned}
\tag{3}
$$

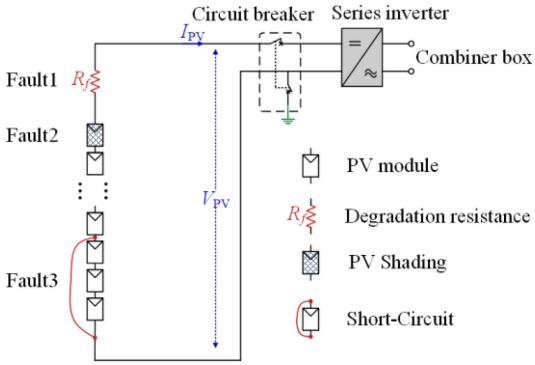

Fig. 3. Fault simulation platform.

TABLE II
PARAMETERS OF UL-280P-60

PV Module	Unit	Value
Maximum power	W	280
Voltage at $P_{\max}$	V	31.2
Current at $P_{\max}$	A	8.98
Open-circuit voltage	V	38.4
Short-circuit current	A	9.36
Temperature coefficient of $P_{\max}$	%/°C	-0.403
Temperature coefficient of V_{oc}	%/°C	-0.330
Temperature coefficient of I_{sc}	%/°C	+0.049

where STC is standard test condition. T_e and G_e are the temperature and irradiance of the environment. k_i is temperature coefficient of short-circuit current. E_g is band gap energy, $E_{g\cdot\mathrm{STC}}$ is1.1557 eV.

The *I–V* curves of the PV modules under different conditions, both before and after correction, were tested at $T_{\mathrm{e}} = 30$ °C and $G_{\mathrm{e}} = 1200$ W/m^2, as illustrated in Fig. 2.

B. Preliminary Analysis of PV Module Faults

In this section, a 2800 W PV array is built within MATLAB/Simulink using the TDM proposed previously, as shown in Fig. 3. The selected PV module is UL-280P-60, which features exceptional corrosion resistance and high PV conversion efficiency, and is widely used in PV power generation systems. For details about the parameters, see Table II. In order to further analyze fault characteristics under various operating conditions,

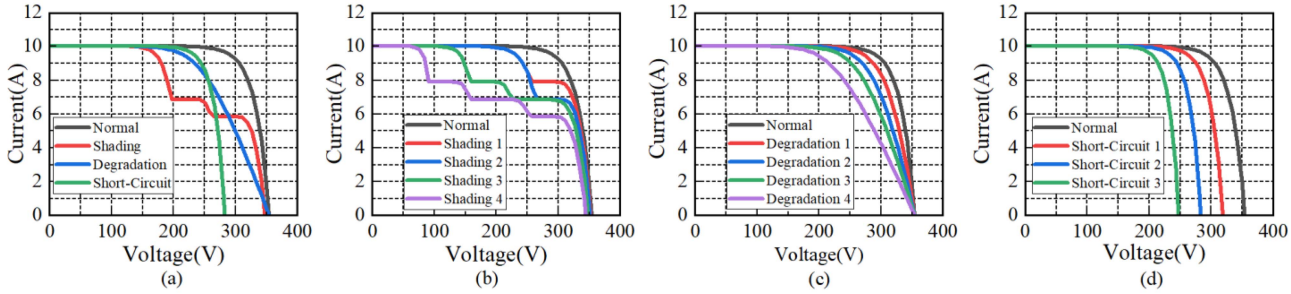

Fig. 4. *I–V* curves. (a) *I–V* curves at different fault conditions. (b) *I–V* curves at different degrees of shading. (c) *I–V* curves at different degrees of degradation. (d) *I–V* curves for different short-circuit numbers of PV modules.

TABLE III
SPECIFIC FAULT SIMULATION SETTINGS

Simulation Type		Operating Conditions of PV Modules
Normal		PV array works normally.
Shading	Shading 1	PV1-PV2 work at 800 W/m^2, and other modules work normally.
	Shading 2	PV1-PV2 work at 700 W/m^2, and other modules work normally.
	Shading 3	PV1-PV2 work at 800 W/m^2, PV3-PV5 work at 700 W/m^2, and other modules work normally.
	Shading 4	PV1-PV2 work at 800 W/m^2, PV3-PV5 work at 700 W/m^2, PV6-PV7 work at 600 W/m^2, and others work normally.
Degradation	Degradation 1	2 Ω resistance in series in PV array.
	Degradation 2	4 Ω resistance in series in PV array.
	Degradation 3	6 Ω resistance in series in PV array.
	Degradation 4	10 Ω resistance in series in PV array.
Short-Circuit	Short-Circuit 1	One PV module is short circuited.
	Short-Circuit 2	Two PV modules are short circuited.
	Short-Circuit 3	Three PV modules are short circuited.
PV1-PV10 are the 10 PV modules in the PV array.		

four typical operating conditions of the PV array (normal, shadowing, degradation, and module short-circuit) were simulated in MATLAB/Simulink, and the *I–V* curves were analyzed, as shown in Fig. 4.

Shading was simulated by adjusting the irradiance of the PV module. Degradation was simulated by placing resistors in series within the PV array. A short-circuit was simulated by directly connecting wires at both ends of the PV module. The specific fault simulation settings are shown in Table III.

Under Shading 1, the *I–V* curve exhibits a step, with the height of this step corresponding to the short-circuit current at an irradiance of 800 W/m^2. Shading 2, in contrast to Shading 1, shows a lower step height. This is because, under Shading 2, the shaded modules are exposed at a lower irradiance of 700 W/m^2. Compared to Shading 2, Shading 3 exhibits an additional step on the curve. This is because, under Shading 3, the modules are subjected to two different levels of shading (700 W/m^2 and 800 W/m^2). The lower step is longer because more PV modules are shaded under 700 W/m^2 than 800 W/m^2. Shading 4, adds two modules operating at 600 W/m^2 compared to Shading 3. Therefore, the *I–V* curve has an additional step again, and the length of the step is the same as those in Shading 1 and Shading 2. The features of *I–V* curve under shading can be summarized as follows: the number of steps on the *I–V* curve corresponds to the number of different shading levels in the array. The length of each step reflects the number of PV modules affected by the respective shading level.

The *I–V* curve demonstrates that as the PV module's degradation increases, the maximum power point on the *I–V* curve is dramatically decreased, meanwhile, the slope of the *I–V* curve steepens in the medium region of the curve. The open-circuit voltage decreases linearly when the PV module is short circuited. For example, if n PV modules are short-circuited in a PV string consisting of 10 modules with identical parameters, the open-circuit voltage of the PV string will decrease by $n/10$.

Through the analysis of the *I–V* curves for the four types of faults, it is observed that under different operating conditions, as the voltage of the *I–V* curve varies from 0 to V_{oc}, the shape, trend, slope of the curve, and ΔI between adjacent points undergo significant changes. Compared to traditional diagnosis for *I–V* curve, which only analyzes specific points such as I_{sc}, V_{oc}, and the maximum power point on the *I–V* curve, focusing on the deformation characteristics between the I_{sc} and the maximum power point, and between the maximum power point and V_{oc} provides a better reflection of the PV module' operating state. Therefore, in this article, *I-V* curve deformation information is utilized as the fault characteristics of PV module, combined with computer vision to complete the task of PV module fault diagnosis.

III. METHODOLOGY

The overall process of the proposed PV array fault diagnosis approach is illustrated in Fig. 5.

Step 1: Obtain the *I–V* curve. *I–V* curves can be easily obtained using an *I–V* curve scanner or an inverter scan and then transferred to the cloud management platform.

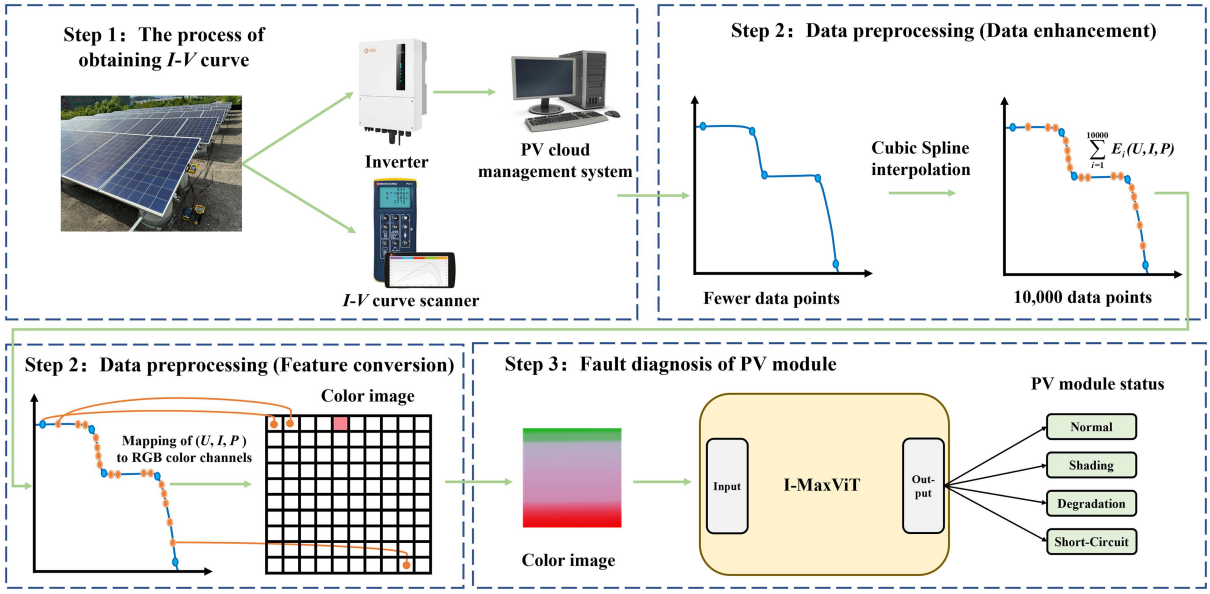

Fig. 5. Overall process of the proposed approach.

Step 2: Preprocess the *I–V* curve. Given that the data points on the initial *I–V* curve are relatively sparse, the pixels converted into color images are also relatively few. Therefore, in order to obtain higher quality color images, data enhancement is employed to increase the number of data points on each curve sufficiently. This ensures that the data is more densely distributed on the curve deformation area while maintaining the original curve shape, thus enhancing the features of the subsequent color image.

Step 3: Convert the *I–V* curve into color image. The electrical parameters (*U*, *I*, *P*) of each point on the curve are mapped to the values of the three-color channels in the RGB color format, and then these pixels are arranged in matrix format to form a color image.

Finally, the color image is input into the I-MaxViT network, which outputs the health status of the PV modules. The specific implementation steps are described in the following sections.

A. *I-V Curve Data Preprocessing*

A novel method for mapping electrical parameters to RGB color image is proposed in the following section. To enhance the features of *I–V* curve in this method, more data should be collected. *I–V* curves are typically acquired in engineering applications using PV inverters or *I–V* curve scanning devices. However, the limitations of the equipment in sample rate and data acquisition accuracy result in a limited amount of collected data. In this study, approximately 30 data points were collected using the *I–V* curve scanner. In order to increase the number of data points on the *I–V* curve to a sufficient amount, the Cubic Spline interpolation was employed. The formula is shown as

$$S_i(x) = a_i(x - x_i)^3 + b_i(x - x_i)^2 + c_i(x - x_i) + d_i \quad (4)$$

where $S_i(x)$is the cubic spline function in the interval $[x_i, x_{i+1}]$. The coefficients a_i, b_i, c_i, and d_i are determined by enforcing the continuity of the function and its first and second derivatives at each knot, which leads to a tridiagonal system of equations. Specifically, the coefficients can be obtained as

$$\begin{aligned} a_i &= \frac{D_{i+1} - D_i}{6h_i} \\ b_i &= \frac{D_i}{2} \\ c_i &= \frac{y_{i+1} - y_i}{h_i} - \frac{h_i}{6}(2D_i + D_{i+1}) \\ d_i &= y_i \end{aligned} \quad (5)$$

where $y_i = f(x_i)$ is the function value at knot x_i, $h_i = x_{i+1} - x_i$ is the interval width, and D_i is the second derivative of the spline at x_i, which is obtained by solving the tridiagonal system from the smoothness conditions. The detailed derivation follows the standard cubic spline formulation, as detailed in [31].

By using cubic polynomials to interpolate between neighboring data points, the Cubic Spline interpolation ensures that the resulting curve is smoother and more continuous over the entire interval. Given a set of data points $(x_0, y_0), (x_1, y_1), \ldots, (x_i, y_i)$, Cubic Spline interpolation fits a cubic polynomial in each adjacent interval $[x_i, x_{i+1}]$ to approximate the data.

Compared to the flat segments on the curve, the electrical parameter characteristics in regions with significant curve deformation are more critical for determining the operational status of the PV array. Nonuniform interpolation is applied to efficiently capture these sections of the *I–V* curve with substantial deformation. More data could be interpolated in the region of the *I–V* curve with obvious slope change. The formula is shown in (6). This method effectively enhances the electrical parameter

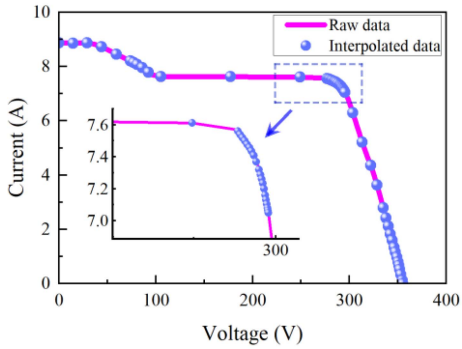

Fig. 6. *I-V* curves with increased data in characteristic regions.

characteristics on the *I–V* curve, as shown in Fig. 6

$$x_{i+1} = k\,|S_i'(x_i)| + x_i \tag{6}$$

where k is the coefficient for nonuniform interpolation.

B. Mapping of I-V Curve to RGB Color Image

For each data point on the *I–V* curve, a vector denoted as $\boldsymbol{E}_i = [U, I, P]$ is constructed, where i is the number of data points on the curve (in this case, i is 10 000). This approach transforms the 10 000 data points on an *I–V* curve into 10 000 data vectors. Each vector encapsulates the three-dimensional (3-D) electrical information of one data point on the *I–V* curve.

Based on the construction of these data vectors, the 10 000 vectors are further arranged into a 100 × 100 matrix, denoted as $\boldsymbol{E}$. Through this approach, the data from the *I–V* curve are represented in matrix form, as shown in (7). In this matrix, the element in the mth row and nth column corresponds to the ((m-1) × n + n)-th data vector on the *I–V* curve. For example, the vector in the second row and first column corresponds to the 101st vector, denoted as $\boldsymbol{E}_{101}$

$$\boldsymbol{E} = \begin{pmatrix} \boldsymbol{E}_{1,1} & \boldsymbol{E}_{1,2} & \cdots & \boldsymbol{E}_{1,99} & \boldsymbol{E}_{1,100} \\ \boldsymbol{E}_{2,1} & \boldsymbol{E}_{2,2} & \cdots & \boldsymbol{E}_{2,99} & \boldsymbol{E}_{2,100} \\ \vdots & \cdots & \boldsymbol{E}_{m,n} & \cdots & \vdots \\ \boldsymbol{E}_{99,1} & \boldsymbol{E}_{99,2} & \cdots & \boldsymbol{E}_{99,99} & \boldsymbol{E}_{99,100} \\ \boldsymbol{E}_{100,1} & \boldsymbol{E}_{100,2} & \cdots & \boldsymbol{E}_{100,99} & \boldsymbol{E}_{100,100} \end{pmatrix}. \tag{7}$$

Linear proportional scaling was employed to normalize the voltage, current, and power, mapping their values to the range (0, 255) before converting the electrical parameters to the RGB mode. This normalizing process enables the subsequent color mapping while preserving the proportional relationships between various data points within the same color channel.

The normalized values of U, I, and P were mapped to the RGB color space, with U, I, and P assigned to red, green, and blue channels, respectively, to form a pixel. The 10 000 vector data from the *I–V* curve were presented in a color image with a resolution of 100 × 100, where the color of each pixel corresponds to a set of (U, I, P) data transformed into RGB colors, as shown in Fig. 7.

This method enables the visual diagnose of fault types based on color features in the image. The various pixel colors and

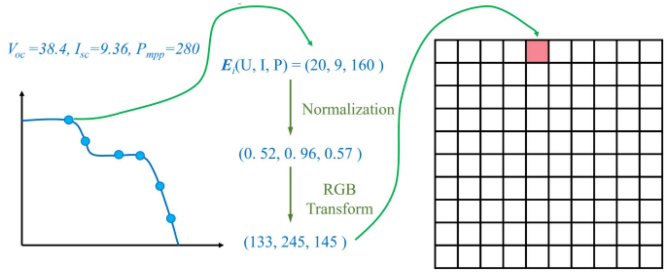

Fig. 7. Mapping electrical parameters to pixels in RGB mode.

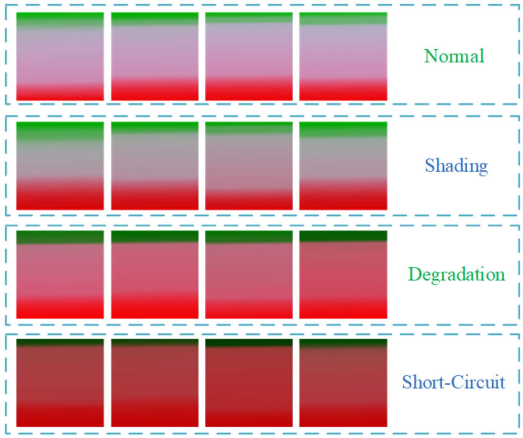

Fig. 8. RGB color images under different operating states.

the overall color change features of the image reveal the differences between the *I–V* curve in different shapes. Variations in color brightness and saturation enable the distinction between *I–V* curves under different operating states, facilitating fault diagnosis. Fig. 8 illustrates the images of the PV array presented in RGB color mode under different operational states.

Fig. 8 demonstrates that the color mapping method produces an effective visualization based on experimental datasets. By mapping 3-D data to RGB colors, an intuitive representation of the curve deformation in a color image is achieved. The color differences on images serve as features that could be utilized by computer vision algorithms for fault diagnostics.

C. Improved Multiaxis Vision Transformer

In this section, the I-MaxViT is introduced, consisting of three main structures, MBconv block, MaxViT block, and output block. The overall operational flow is illustrated in Fig. 9. After mapping the processed electrical parameters to an RGB image, the resolution is reduced through two layers of convolution in the stream module. Subsequently, the multiaxis attention mechanism is employed to sequentially extract features across the four layers of MaxViT block. Finally, the Head block is utilized to classify the detailed features in a stepwise manner.

1) MBConv Block: In the design of MaxViT, the primary convolution operation is employed in the MBConv block, as

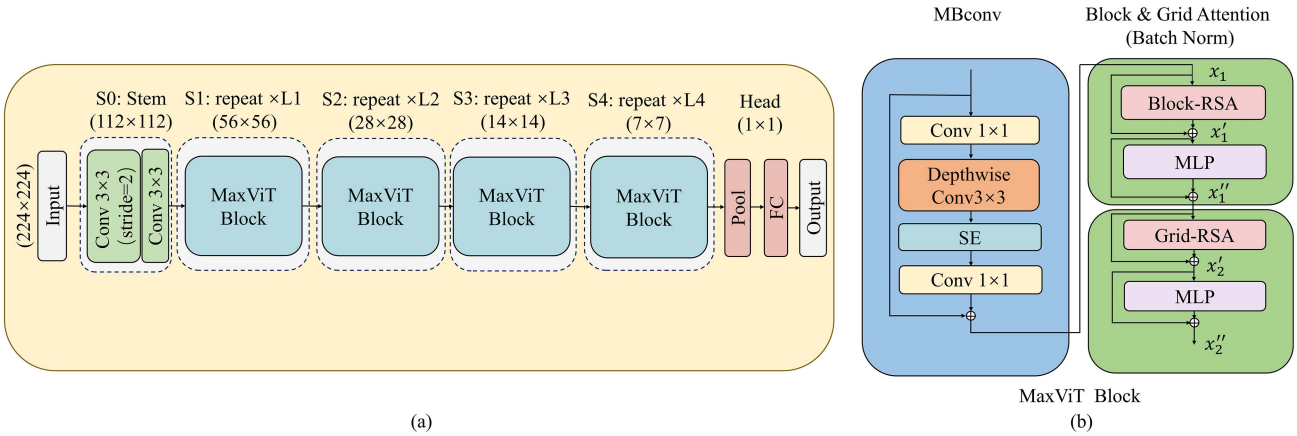

Fig. 9. (a) Overall architecture of MaxViT. L1-L4 denote the number of iterations for different feature layers, with values of [2, 6, 14, 2] adopted in this article. (b) Specific structure of MaxViT. It employs the MBConv block to extract local features and multiscale channel attention features. Then, it utilizes the Block and Grid Attention block to extract rich features, enabling further interaction for image detail information.

shown in

$$x \leftarrow Proj\left(AvgPool2D\left(x\right)\right) or\ x + Proj\left(SE\left(DWConv\downarrow\left(Conv\left(Norm\left(x\right)\right)\right)\right)\right) \quad (8)$$

where x is the output feature, Norm denotes BatchNorm, Conv is the expansion Conv1 × 1, DWConv is the Depthwise Conv3 × 3, SE is the Squeeze-Excitation layer, Proj is the shrink Conv1 × 1 to downproject the number of channels, and AvgPool2D(x) is a component of the first-stage shortcut branch.

It is worth noteworthy that in the first MBConv block of each stage, downsampling is accomplished by applying a Depthwise Conv3 × 3 with a stride of 2, while the shortcut branch incorporates average pooling as one of its components. In the second MBConv block, a conventional shortcut branch is utilized.

2) Relative-Attention Mechanism: For the self-attention mechanism computation, due to the computational complexity associated with global attention mechanisms, a relative attention mechanism is adopted, as demonstrated in (9). The advantage of this approach lies in its ability to incorporate information about the relative distances between different positions, thereby enabling more effective capture of features from related positions. Additionally, relative positions are represented using lower-dimensional encoding, which reduces the model's computational and memory requirements [32]

$$\text{RSAttention}\left(Q,K,M\right)=\text{Softmax}\left(QK^T\sqrt{d}+\boldsymbol{B}\right)M \quad (9)$$

where RSAttention is the relative self-attention mechanism, Q, K, and M stand for feature matrices with different weights at the same positions, respectively, d is the hidden dimension, and $\boldsymbol{B}$ is the position-aware matrix.

The attention weight is jointly determined by a learned static location-aware matrix $\boldsymbol{B}$ and the scaled input-adaptive attention $QK^T/\sqrt{d}$. Simultaneously, by applying Softmax operation to normalize the weights into a probability distribution, the generated weights are assigned to the corresponding value elements V, resulting in the final output vector. Additionally, a trainable parameter matrix $\boldsymbol{B}\in\boldsymbol{R}^{((2H-1)\times(2W-1))}$is utilized. For any spatial positions (x, y) and (x', y'), the corresponding bias parameter $\boldsymbol{B}_{x-x'+H,y-y'+W}$ could be obtained by calculating their relative position indices (Δx, Δy).

3) MaxViT Block: The multiaxis self-attention mechanism could optimize global feature interactions, but it incurs significant computational complexity. To address this, scaled cosine attention mechanism and log-spaced continuous position bias (Log-CPB) [33] were introduced to reduce the complexity of the relative positional bias attention. Additionally, the global-local interaction is achieved by decomposing spatial dimensions into local (Block-RSAttention) and global (Grid-RSAttention) sparse forms. The Block(·) operator, with parameter P, partitions the input image/feature $x\in\boldsymbol{R}^{H\times W\times C}$ into nonoverlapping blocks, each of size $P\times P$. After window partition, the block dimensions are gathered onto the spatial dimension (i.e., -2 axis), as shown in (10). The ReBlock(·) operation is the reverse of the above block partition procedure

$$\text{Block}:(H,W,C)\rightarrow\left(\frac{H}{P}\times P,\frac{W}{P}\times P,C\right)\rightarrow\left(\frac{HW}{P^2},P^2,C\right). \quad (10)$$

Similarly, the Grid(·) operation, with parameter G, partitions the input feature into a uniform G × G grid, with each lattice having an adaptive size of (H/G,W/G). Unlike the block operator, an additional Transpose is applied to align the grid dimension with the spatial axis (i.e., -2 axis), as shown in (11). Its inverse operation, ReGrid(·), restores the gridded input to the normal 2-D feature space

$$Grid:(H,W,C)\rightarrow\left(G\times\frac{H}{G},G\times\frac{W}{G},C\right)\rightarrow\underbrace{\left(G^2,\frac{HW}{G^2},C\right)\rightarrow\left(\frac{HW}{G^2},G^2,C\right)}_{swapaxes(axis1=-2,axis2=-3)}. \quad (11)$$

To more effectively propagate weights trained at low resolutions, the scaled cosine attention is introduced to replace the original dot product attention, and Log-CPB is employed instead

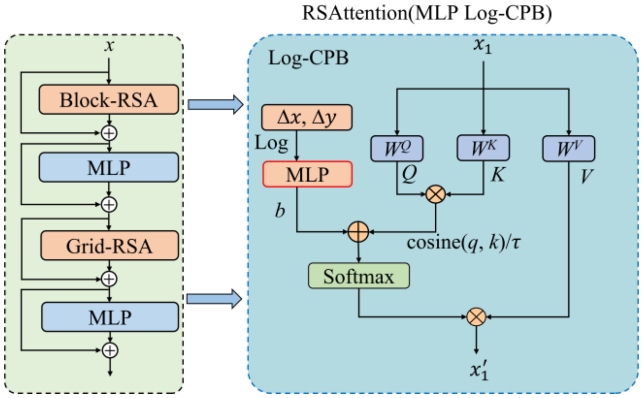

Fig. 10. RSAttention (MLP Log-CPB) mechanism.

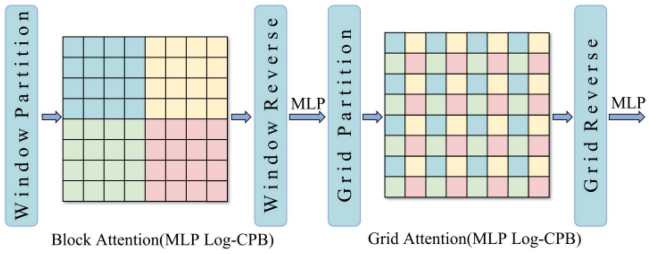

Fig. 11. Calculation process of the multiaxis self-attention mechanism (Max-SA) in the Max-SA module (with a window/grid size of 4 × 4). The window attention module performs self-attention operations within the window, while the grid attention module conducts global attention operations on pixels in the sparse, evenly distributed grid across the entire 2-D space. The complexity of both operations is linearly related to the input size.

of the previous parameterized approach, as shown in Fig. 10. In contrast to standard attention mechanisms, the scaled cosine attention normalizes the query and key vectors before computing their similarity, focusing on directional similarity rather than magnitude, which improves the stability and generalization of the attention computation. Additionally, the Log-CPB is adopted in place of the traditional parameterized relative position bias table with a lightweight MLP that uses logarithmic relative coordinates as input, which enhances the flexibility and efficiency of position encoding. The attention between pixels i and j is computed using the scaled cosine function, as shown in (12). A small network $b(\Delta x, \Delta y)$ is utilized for relative positional coordinates, as shown in (13)

$$Sin(q_i, k_j) = cos(q_i, k_j)/\tau + B_{ij} \tag{12}$$

where B_{ij} is the relative position bias between pixel i and j, τ is a learnable scalar, nonshared across heads and layers, and τ is set to a value greater than 0.01, G is a small network, and a two-layer MLP with ReLU activation function is employed in the middle

$$b(\Delta x, \Delta y) = G(\Delta x, \Delta y). \tag{13}$$

A significant extrapolation of the relative coordinate range is required when switching between window sizes with substantial variations. To address this issue, the logarithmic coordinates are employed instead of the original linear coordinates, as shown in

$$\begin{aligned} \widehat{\Delta x} &= sign(x) \cdot \log(1 + |\Delta x|) \\ \widehat{\Delta y} &= sign(y) \cdot \log(1 + |\Delta y|) \end{aligned} \tag{14}$$

where Δx, Δy, $\widehat{\Delta x}$, and $\widehat{\Delta y}$ are the linear-scaled and log-spaced coordinates, respectively.

The multiaxis attention block is ultimately formed, as shown in Fig. 11. For a given input 2-D electrical parameter image ($x_1^{C\times H\times W} \in \boldsymbol{R}$), the processing through these two attention mechanisms blocks could be expressed as follows.

The local block attention could be expressed as

$$\begin{aligned} x_1^{'} &\leftarrow x_1 + ReBlock(RSAttention(Block(LN(x_1)))) \\ x_1^{''} &\leftarrow x_1 + \text{MLP}\left(LN\left(x_1^{'}\right)\right) \end{aligned}$$

Fig. 12. Experiment platform and fault simulation.

while the global grid attention is formulated as

$$\begin{aligned} x_2^{'} &\leftarrow x_1^{''} + \text{ReGrid}\left(RSAttention\left(Grid\left(LN\left(x_1^{''}\right)\right)\right)\right) \\ x_2^{''} &\leftarrow x_2^{'} + \text{MLP}\left(LN\left(x_2^{'}\right)\right) \end{aligned}$$

where LN is the normalized layer and MLP is a standard multilayer perceptron network utilizing GELU activation function.

IV. Experiments

To further validate the feasibility and performance of the proposed method in practical applications, a 2800 W PV experiment platform was established, and ON-site fault experiments were performed. The PV array constructed in the experiment consists of 10 UL-280P-60 PV modules in series. The PV200 *I-V* curve scanner from U.K. company SEAWARD was employed to collect the data. Cardboard was used to cover 1–3 PV modules to simulate shading. The MC4 interface wire was directly connected at both ends of the PV module to simulate short-circuit fault. Degradation fault was simulated by connecting high-power cement resistors in series within the PV array, as shown in Fig. 12.

TABLE IV
EXPERIMENT SAMPLE NUMBER DIVISION

Type	Total number of datasets	Number of training datasets	Number of test datasets
Normal	114	80	34
Degradation	189	132	57
Shading 1	76	53	23
Shading 2	40	28	12
Short-Circuit 1	32	22	10
Short-Circuit 2	39	27	12

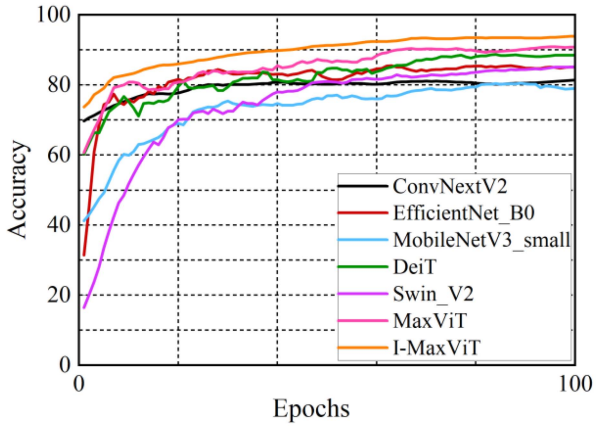

Fig. 13. Accuracy comparison of seven algorithms.

From 21 Aug. 2023 to 31 Aug. 2023, between 10:30 A.M. and 2:30 P.M., the fault experiments were conducted at the experimental base. Due to the poor scanning quality of *I–V* curves at low irradiance, measurements were taken only at irradiances above 600 W/m^2. After excluding the abnormal curves caused by sudden weather changes (temporary clouds), a total of 490 valid *I–V* curves were collected.

To train the I-MaxViT model, 70% of the data samples were randomly selected as the training set. The remaining 30% were used as the test set to evaluate the performance of the fault diagnosis algorithm. The simulated fault types and their corresponding sample counts are detailed in Table IV.

In this experiment, we used workstation model Desktop-JGA1F7G with NVIDIA Quadro P5000 graphics card, which adopts the NVIDIA Pascal architecture, providing powerful computing and rendering capabilities. In this experiment, a Desktop-JGA1F7G workstation equipped with an NVIDIA Quadro P5000 graphics card, based on the NVIDIA Pascal architecture, was utilized, offering robust computing and rendering capabilities.

Upon reaching 100 epochs, the accuracy of I-MaxViT achieved 94%, as shown in Fig. 13. According to the confusion matrix, only 9 samples of Short-Circuit 2 were misclassified as Short-Circuit 1, possibly due to the relatively small proportion of Short-Circuit 2 samples and the high similarity between the two states, which essentially belong to the same fault type. All other fault types were correctly diagnosed, as depicted in Fig. 14.

The proposed I-MaxViT was compared with six representative vision models that represent the three main architectures in

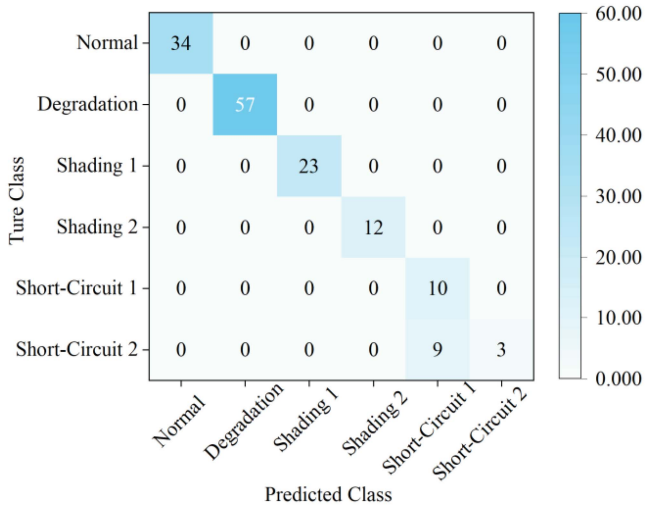

Fig. 14. Confusion matrix of I-MaxViT.

computer vision: CNN-based (ConvNextV2, EfficientNet_B0, MobileNetV3_small), which focus on efficient local feature extraction; Transformer-based (DeiT, SwinV2), which emphasize global context modeling through self-attention; and the hybrid CNN-Transformer model (MaxViT), which combines the strengths of both. To ensure a fair and consistent evaluation, all models were implemented under identical experimental settings. Each model was either trained from scratch or fine-tuned using the same dataset and preprocessing pipeline. Hyperparameters such as learning rate, optimizer, batch size, and number of training epochs were kept consistent across all models. Furthermore, the evaluation metrics were computed under the same testing conditions to provide an objective and comprehensive comparison. The accuracy of those models is shown in Fig. 13. It can be observed that the accuracy of both MaxViT and I-MaxViT tend to stabilize after 60 epochs. In terms of accuracy, the I-MaxViT achieves the highest accuracy.

The evaluation metrics for fault diagnosis results include Precision, Recall, F1-score, Accuracy, area under the receiver operating characteristic curve (AUROC), FLOPs, Memory, Throughput, and Inference Time. These metrics are summarized in Table V and illustrated in Fig. 15. It is observed that while CNN-based models generally exhibit advantages in computational efficiency and Throughput, this comes at the cost of reduced diagnostic Accuracy,, and AUROC, potentially leading to suboptimal fault detection. In contrast, Transformer-based models demonstrate higher diagnostic performance, with improved Precision, Recall, F1-score, and AUROC compared to CNN-based models, but at the expense of increased computational complexity and inference time. Notably, the I-MaxViT achieves the highest performance across all accuracy-related metrics, including Precision, Recall, F1-score, Accuracy, and AUROC, achieving an accuracy of 94%, improving by approximately 3% over MaxViT, while also enhancing throughput. It is evident that the I-MaxViT proposed in this article exhibits superior performance in PV module fault diagnosis.

TABLE V
MODEL COMPARISON ANALYSIS

Model	Type	Precision	Recall	F1-score	Accuracy	AUROC	FLOPs (G)	Memory (MB)	Throughput (img/s)	Inference Time(ms)
ConvNextV2	CNN -based	0.78	0.80	0.76	0.81	0.82	4.50	1100	200	6.50
EfficientNet_B0		0.82	0.85	0.82	0.86	0.86	0.38	715.20	170.65	5.86
MobileNetV3_small		0.78	0.82	0.79	0.83	0.84	0.05	735.90	249.57	4.00
DeiT	Transformer -based	0.92	0.88	0.89	0.90	0.92	4.24	872.52	313.88	3.18
SwinV2		0.82	0.83	0.81	0.86	0.88	4.37	1307.67	101.82	9.82
MaxViT	CNN + Transformer	0.93	0.90	0.88	0.91	0.92	8.54	857.97	53.92	18.54
I-MaxViT		0.96	0.92	0.93	0.94	0.95	4.89	912.20	66.79	14.97

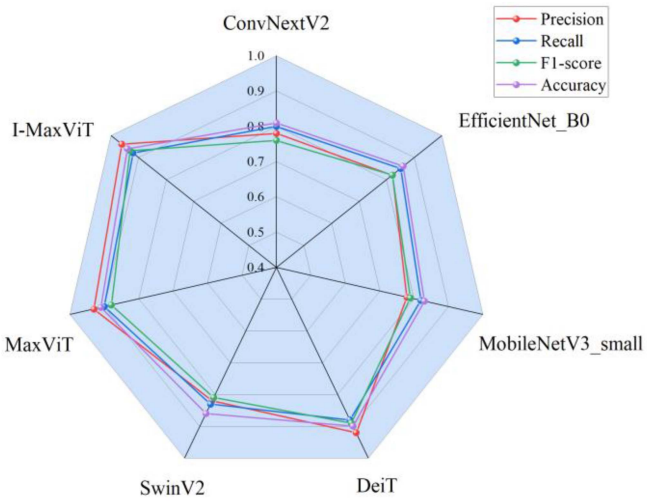

Fig. 15. Cobweb diagrams of evaluation metrics for the seven models.

TABLE VI
ABLATION STUDY RESULTS

Model Variant	Accuracy (%)	FLOPs (G)	Memory (MB)	Throughput (img/s)
Base MaxViT	0.91	8.54	857.97	53.92
MaxViT + Log-CPB	0.92	6.32	875.23	56.31
MaxViT + Scaled Cosine Attention	0.92	6.57	881.64	57.72
Proposed I-MaxViT	0.94	4.89	912.20	66.79

To evaluate the effectiveness of the proposed architectural enhancements in I-MaxViT, an ablation experiment was conducted based on four performance metrics: Accuracy, FLOPs, Throughput, and Inference Time. The experiment focuses on the contributions of Log-CPB and scaled cosine attention, which were, respectively, integrated into the baseline MaxViT.

TABLE VII
EXPERIMENT RESULTS UNDER DIFFERENT PV MODULES AND ARRAYS

PV module model	PV array scale	Number of test datasets	Accuracy
UL-280P-60	2.8 kW 10 × 1	148	0.94
HT72-18X-540	840 W 3 × 1	100	0.93
LR4-72HPH-450M	1.4 kW 5 × 1	100	0.95
UL-280P-60	1.12 kW 2 × 2	100	0.94
UL-280P-60	2.52 kW 3 × 3	100	0.94

As summarized in Table VI, the incorporation of Log-CPB and scaled cosine attention improves accuracy, while reduces FLOPs. In terms of throughput, both modifications lead to better gains. The fully modified I-MaxViT achieves the highest diagnostic accuracy and the best overall tradeoff across all four metrics, validating the effectiveness of this design in RGB-based PV fault diagnosis.

In order to further verify the generalization and accuracy of the proposed method for different PV modules and PV arrays of different scale. Additional PV arrays were constructed using three HT72-18X-540 modules and five LR4-72HPH-450M modules, respectively. A total of 200 *I–V* curves were collected to verifying the accuracy and universality of the proposed method, is shown in Table VII. The results demonstrate that the method exhibits excellent overall performance.

V. CONCLUSION

This article proposes a novel diagnostic method that integrates *I–V* curve deformation characteristics with computer vision. The *I–V* curve characteristics have been converted into image features, enabling better diagnostic results through computer vision, which can capture the complete deformation features of the *I–V* curve while reducing computational complexity by converting the extensive data into a color image. I-MaxViT model is employed for PV fault diagnosis. The relative attention mechanism is improved by introducing scaled cosine attention mechanism and MLP Log-CPB, which decreases the computational complexity. To verify the proposed innovative fault diagnostic method, an experiment platform for a 2.8 kW 10 × 1 PV array is set up. Comparative analysis is performed

with several other diagnostic models. The experiment results demonstrate that, compared to the other six diagnostic algorithms, the proposed innovative diagnostic method achieves the highest diagnostic accuracy, reaching 94% . Additionally, tests on different types of modules and arrays of various sizes demonstrate that the method has excellent universality.

REFERENCES

[1] S. Leva, M. Mussetta, and E. Ogliari, "PV module fault diagnosis based on microconverters and day-ahead forecast," *IEEE Trans. Ind. Electron.*, vol. 66, no. 5, pp. 3928–3937, May 2019.

[2] A. Mehmood, H. A. Sher, A. F. Murtaza, and K. Al-Haddad, "A diode-based fault detection, classification, and localization method for photovoltaic array," *IEEE Trans. Instrum. Meas.*, vol. 70, 2021, Art. no. 3516812.

[3] M.-C. Chang and S.-I. Liu, "An indoor photovoltaic energy harvester using time-based MPPT and on-chip photovoltaic cell," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 67, no. 11, pp. 2432–2436, Nov. 2020.

[4] M. Dhimish and J. Kettle, "Impact of solar cell cracks caused during potential-induced degradation (PID) tests," *IEEE Trans. Electron Devices*, vol. 69, no. 2, pp. 604–612, Feb. 2022.

[5] Y. Wang and B. Ren, "Fault ride-through enhancement for grid-tied PV systems with robust control," *IEEE Trans. Ind. Electron.*, vol. 65, no. 3, pp. 2302–2312, Mar. 2018.

[6] A. Triki-Lahiani, A. B.-B. Abdelghani, and I. Slama-Belkhodja, "Fault detection and monitoring systems for photovoltaic installations: A review," *Renewable Sustain. Energy Rev.*, vol. 82, pp. 2680–2692, 2018.

[7] A. Mellit, "An embedded solution for fault detection and diagnosis of photovoltaic modules using thermographic images and deep convolutional neural networks," *Eng. Appl. Artif. Intell.*, vol. 116, Nov. 2022, Art. no. 105459.

[8] N. Kellil, A. Aissat, and A. Mellit, "Fault diagnosis of photovoltaic modules using deep neural networks and infrared images under Algerian climatic conditions," *Energy*, vol. 263, Jan. 2023, Art. no. 125902.

[9] E. A. Ramadan et al., "An innovative transformer neural network for fault detection and classification for photovoltaic modules," *Energy Convers. Manage.*, vol. 314, Aug. 2024, Art. no. 118718.

[10] S. Hao, T. Qi, X. Ma, J. Li, and T. Li, "RP-CFANet: An adaptive photovoltaic hot-spot fault detection network based on region perception and cross-channel feature aggregation," *IEEE Trans. Instrum. Meas.*, vol. 74, 2025, Art. no. 3530714.

[11] T. He, S. Hao, X. Zhang, X. Ma, S. Sun, and C. Yang, "APM^{2}Det: A photovoltaic hot-spot fault detection network based on angle perception and model migration," *IEEE Trans. Dielectrics Elect. Insul.*, vol. 31, no. 6, pp. 2938–2946, Dec. 2024.

[12] B. Su, Z. Zhou, and H. Chen, "PVEL-AD: A large-scale open-world dataset for photovoltaic cell anomaly detection," *IEEE Trans. Ind. Inform.*, vol. 19, no. 1, pp. 404–413, Jan. 2023.

[13] Z. Chang et al., "Towards more reliable photovoltaic energy conversion systems: A weakly-supervised learning perspective on anomaly detection," *Energy Convers. Manage.*, vol. 316, Sep. 2024, Art. no. 118845.

[14] N. Drir et al., "Hybrid CNN-EML model for fault diagnosis in electroluminescence images of photovoltaic cells," *Renewable Energy*, vol. 250, Sep. 2025, Art. no. 123343.

[15] K. Xia, J. Ni, Y. Ye, P. Xu, and Y. Wang, "A real-time monitoring system based on ZigBee and 4G communications for photovoltaic generation," *CSEE J. Power Energy Syst.*, vol. 6, pp. 52–63, Mar. 2020.

[16] C. Cárdenas-Bravo et al., "Online crack detection on photovoltaic devices using a dynamic response analysis," *Renewable Energy*, vol. 248, Aug. 2025, Art. no. 122990.

[17] W. Su et al., "An improved reflectometry-based fault location scheme for suppressing multiple reflections in large-scale photovoltaic strings," *Sol. Energy*, vol. 270, Mar. 2024, Art. no. 112389.

[18] H. R. Parsa and M. Sarvi, "Online fault diagnosis, classification, and localization in photovoltaic systems," *IEEE Trans. Instrum. Meas.*, vol. 73, 2024, Art. no. 3516208.

[19] U. Kumar, S. Mishra, and K. Dash, "An IoT and semi-supervised learning-based sensorless technique for panel level solar photovoltaic array fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 72, 2023, Art. no. 3521412.

[20] M. Bahloul, L. Breathnach, and S. Khadem, "Design and field implementation of a hierarchical control solution for residential energy storage systems," *IEEE Trans. Smart Grid*, vol. 14, no. 2, pp. 1083–1092, Mar. 2023.

[21] B. Li, C. Delpha, A. Migan-Dubois, and D. Diallo, "Fault diagnosis of photovoltaic panels using full I-V characteristics and machine learning techniques," *Energy Convers. Manage.*, vol. 247, Nov. 2021, Art. no. 114785.

[22] Y. J. Liu et al., "Intelligent fault diagnosis of photovoltaic array based on variable predictive models and I-V curves," *Sol. Energy*, vol. 237, pp. 340–351, May 2022.

[23] M. Ma, Z. Zhang, P. Yun, Z. Xie, H. Wang, and W. Ma, "Photovoltaic module current mismatch fault diagnosis based on I-V data," *IEEE J. Photovolt.*, vol. 11, no. 3, pp. 779–788, May 2021.

[24] X. Chen et al., "Fault prediagnosis, type identification, and degree diagnosis method of the photovoltaic array based on the current–Voltage conversion," *IEEE Trans. Power Electron.*, vol. 39, no. 12, pp. 16708–16719, Dec. 2024.

[25] M. Dhimish and A. M. Tyrrell, "Photovoltaic bypass diode fault detection using artificial neural networks," *IEEE Trans. Instrum. Meas.*, vol. 72, 2023, Art. no. 3507710.

[26] A. Mellit and S. Kalogirou, "Assessment of machine learning and ensemble methods for fault diagnosis of photovoltaic systems," *Renewable Energy*, vol. 184, pp. 1074–1090, Jan. 2022.

[27] Z. J. Duan, H. Yu, Q. Zhang, and L. Tian, "Parameter extraction of solar photovoltaic model based on nutcracker optimization algorithm," *Appl. Sci.*, vol. 13, no. 11, May 2023, Art. no. 6710.

[28] H. Rezk and M. A. Abdelkareem, "Optimal parameter identification of triple diode model for solar photovoltaic panel and cells," *Energy Rep.*, vol. 8, pp. 1179–1188, Apr. 2022.

[29] M. A. K. Magableh, A. Radwan, and Y. A.-R. I. Mohamed, "A novel hybrid approach based on analytical and metaheuristic algorithms for parameters and dynamic resistance estimation of a PV array," *IEEE Trans. Power Syst.*, vol. 38, no. 6, pp. 5459–5474, Nov. 2023.

[30] X. Chen et al., "Research on real-time identification method of model parameters for the photovoltaic array," *Appl. Energy*, vol. 342, Jul. 2023, Art. no. 121157.

[31] S.-C. Lin, K. Chuang, C.-W. Chang, and J.-H. Chen, "Efficient interpolation method for wireless communications and signal processing applications," *IEEE Trans. Microw. Theory Techn.*, vol. 69, no. 5, pp. 2753–2761, May 2021.

[32] Z. Tu et al., "Maxvit: Multi-axis vision transformer," in *Proc. 17th Eur. Conf. Comput. Vis.*, Oct. 2022, pp. 459–479.

[33] Z. Liu et al., "Swin transformer V2: Scaling up capacity and resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 11999–12009.

Lingxiang Zhang (Student Member, IEEE) received the B.S. degree from Xichang University, Xichang, China, in 2021 and the M.S. degree from University of Shanghai for Science and Technology, Shanghai China, in 2024, both in electrical engineering.

His research interests include photovoltaic fault diagnosis and dc–dc power conversion.

Kun Xia (Member, IEEE) received the Ph.D. degree in power electronics and power drives from Hefei University of Technology, Hefei, China, in 2007.

Since 2020, he has been a Professor with the School of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai, China, where he is currently the President with the College of Innovation and Entrepreneurship and the Dean of Public Experiment Center. He was in charge of more than 50 research projects from the government and companies and published more than 80 articles. His research interests include motor and motor control and new energy application.

Po Xu received the B.Sc. degree in power system and automation and the Ph.D. degree in power electronics and power transmission from the Hefei University, Hefei, China, in 2001 and 2006, respectively.

He is the Dean with the Institute of Ginlong Technologies Co., Ltd., as Senior Engineer. His research interests include photovoltaic new energy and photovoltaic system.

Yibo Yu received the B.S. degree in mathematical analytics and operations research from the University of California, Davis, Davis, CA, USA, in 2025. She is currently working toward the M.S. degree in operations research with Columbia University, New York, NY, USA.

Her research interests include machine learning and artificial intelligence.

Yiming Wang received the B.Sc. degree in information technology from the Shanghai Jiao Tong University, Shanghai, China in 2002 and the M.Sc. degree in science by research from the College of Science and Engineering in Electronics and Telecommunication, University of Edinburgh, Edinburgh, U.K., in 2003.

He is the CEO with Ginlong Technologies Company Ltd. He was a Visiting Professor with Ningbo Institute of Technology, Zhejiang University, in 2018.