PV-MambaSeg: Geometry-Enhanced Vision Mamba With Omni-Local Scan for Photovoltaic Cell Defect Segmentation

Ziai Zhou, *Student Member, IEEE*, Chaoyang Song, Xiaotong Kong, Shixiong Fang, Huimin Lu, *Senior Member, IEEE*, and Jinxia Zhang, *Member, IEEE*

***Abstract*—Segmenting various defects in photovoltaic (PV) modules at the pixel-level is critical for the intelligent operation and maintenance of large-scale PV plants. However, existing approaches face two major challenges: 1) they struggle to model geometrically complex and arbitrarily oriented defects, and they cannot effectively suppress background noise, thus resulting in discontinuous and imprecise segmentation; 2) they suffer from excessive computational complexity and substantial memory footprint, making them incompatible with resource-constrained scenarios. To tackle these challenges, we introduce PV-MambaSeg, a lightweight yet effective Mamba-based network designed specifically for PV defect segmentation. We introduce the vision Mamba architecture into the field of PV defect segmentation for the first time and enhance its capability through a novel omni-local scan (OLS) strategy and the global local visual state space (GLVSS) module. OLS performs multidirectional scanning to comprehensively capture structural continuity, thereby enabling superior geometry-aware modeling of irregular defect patterns. In PV defect segmentation task, OLS significantly outperforms existing Mamba scanning strategies. GLVSS enhances local representation by aggregating OLS features with multiple window sizes and calibrates global–local feature responses through a channel attention mechanism. Compared with CNN and transformer-based segmentation models, PV-MambaSeg attains a superior balance between performance and efficiency. Comprehensive experiments on two real-world PV defect datasets demonstrate that our method outperforms other state-of-the-art approaches, achieving the highest mIoU with only 2.77 M parameters and 0.78 GFLOPs. This work provides a powerful and practical solution for automated PV defect segmentation. The corresponding code will be made publicly accessible.**

***Index Terms*—Lightweight network, photovoltaic defect segmentation, scan strategy, vision state space model.**

Received 25 May 2026; accepted 22 June 2026. This work was supported in part by the National Natural Science Foundation of China under Grant 62573123 and in part by the Research Fund for Advanced Ocean Institute of Southeast University, Nantong, under Grant GP202411. Paper no. TII-26-4768. *(Corresponding author: Jinxia Zhang.)*

Ziai Zhou, Chaoyang Song, Xiaotong Kong, and Shixiong Fang are with the Key Laboratory of Measurement and Control of CSE, Ministry of Education, School of Automation, Southeast University, Nanjing 210096, China (e-mail: 230238499@seu.edu.cn; 230248980@seu.edu.cn; 230259964@seu.edu.cn; sxfang@seu.edu.cn).

Huimin Lu and Jinxia Zhang are with the Key Laboratory of Measurement and Control of CSE, Ministry of Education, School of Automation, Southeast University, Nanjing 210096, China, and also with the Advanced Ocean Institute of Southeast University, Nantong 226010, China (e-mail: luhuimin@ericlab.org; jinxiazhang@seu.edu.cn).

Digital Object Identifier 10.1109/TII.2026.3707683

I. Introduction

WITH the global energy structure accelerating its transition toward cleanliness and sustainability, photovoltaic (PV) power generation has emerged as a cornerstone of renewable energy systems [1], [2]. However, as the core components of PV power plants, PV cells are inevitably exposed to harsh outdoor environments during long-term operation, leading to the occurrence of various defects [3], [4]. These defects may degrade the power generation efficiency of individual modules and trigger cascading failures in the entire PV array, leading to substantial economic losses and potential safety hazards [5]. Therefore, accurate and efficient defect detection and localization in PV modules are crucial to ensuring the reliable operation and cost-effective maintenance of PV power plants [6], [7].

In recent years, computer vision-based PV defect segmentation methods have gradually replaced traditional manual inspection due to their higher accuracy, efficiency, and automation [8], [9], [10], [11]. They can precisely locate the positions, sizes, and shapes of defects, providing detailed and reliable information for subsequent maintenance. However, despite the remarkable advancements achieved in pixel-level segmentation of PV defects, there are still two major challenges.

The first challenge lies in modeling geometrically complex and arbitrarily oriented defects while suppressing background noise, which often leads to discontinuous and imprecise segmentation. PV module defects exhibit high geometric variability. As shown in Fig. 1, crack may appear as thin, continuous linear structures with random orientations, corrosion as irregularly shaped low-brightness regions, and contact as short, parallel strips. Meanwhile, PV module backgrounds contain extensive interference because of crystal dislocation that further obscures defect features. Traditional convolutional neural network (CNN)-based methods rely on local receptive fields, struggling to capture global contextual information of defects and distinguish geometrically complex patterns from background noise. Although transformer-based methods introduce self-attention to model global dependencies, they frequently neglect the fine-grained local semantics of irregular defects, resulting in failure to segment small, arbitrarily oriented defects.

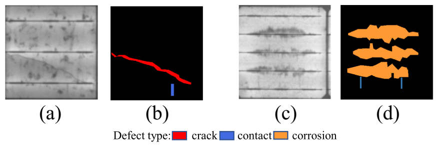

Fig. 1. Examples of typical PV cell defects. (a) and (b) are the original image with crack and contact defects, and its corresponding defect segmentation mask. (c) and (d) are the original image with corrosion and contact defects, and its corresponding segmentation mask.

Furthermore, high computational cost poses a significant challenge for PV power plant O&M. While CNN-based methods offer lower computational complexity, this efficiency is often achieved at the cost of segmentation accuracy; conversely, transformer-based methods deliver superior accuracy but their substantial parameter counts and heavy computational requirements make them unfit for resource-constrained scenarios. This tradeoff between segmentation quality and resource efficiency remains a key bottleneck for PV defect segmentation.

To tackle these challenges, this article introduces PV-MambaSeg, a lightweight yet effective Vision Mamba-based network. Notably, this work pioneers the application of the Vision Mamba architecture in PV defect segmentation. As a sequence modeling framework, Mamba leverages an efficient selective state update mechanism to balance computational efficiency and feature modeling capability. However, directly applying vanilla Vision Mamba to PV defect segmentation has significant limitations: first, its scan order prioritizes sequential token dependencies but misses the irregular spatial distribution of PV defects, failing to fully model defect geometric continuity; second, vision state space module lacks aggregation of local fine-grained details and global context, leading to ambiguous features for low contrast or mixed defects. To enhance Mamba's capacity, we propose the OLS strategy and the lightweight GLVSS module. OLS captures spatial correlations across local patch sequences through its multidirectional scanning design, while GLVSS refines cross-scale feature representations by integrating multiwindow aggregation and adaptive channel attention calibration.

Specifically, OLS introduces three key capabilities aiming to better capture the structural continuity and arbitrary orientations of PV defects compared with existing scanning strategies: it adopts multidirectional full-dimensional scanning to capture the structural continuity of arbitrarily oriented defects; it employs continuous patch scanning to preserve spatial correlations between adjacent regions, mitigating the omission of small defects caused by fragmented scanning; it further leverages differentiated forward–backward paths to extract fine-grained semantics, reducing feature redundancy while enabling complete defect feature decomposition. Complementing OLS, GLVSS refines the extracted features by first aggregating multiwindow features to preserve fine defect patterns and suppress background noise, then fusing these enhanced representations with convolutional fine-grained local features. Finally, the fused outputs are calibrated via an adaptive channel attention mechanism that dynamically balances global context and local details. Extensive experiments verify that OLS outperforms existing Mamba scanning strategies, and the integrated OLS-GLVSS framework achieves robust and accurate PV cell defect segmentation.

PV-MambaSeg simultaneously captures geometrically complex defect features, while maintaining low resource consumption. As shown in Fig. 2, PV-MambaSeg achieves the highest mIoU while preserving a significantly smaller parameter size and GFLOPs when compared to other state-of-the-art methods. The main contributions are presented as follows:

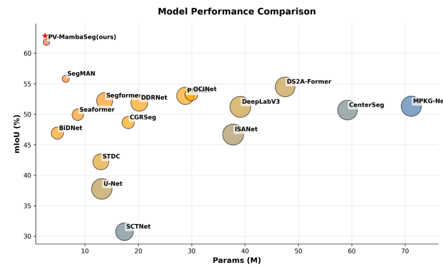

Fig. 2. Comparison of model performance and efficiency. Segmentation accuracy (mIoU, %) on the UCF-EL dataset [3] is plotted against model size (Parameters, M). Each marker represents a model, and its size is proportional to its computational cost (GFLOPs). Our model, PV-MambaSeg, attains a leading balance between accuracy and computational resources.

1) We introduce PV-MambaSeg, the first Mamba-based network tailored for PV multiclass defect segmentation, which effectively models geometrically complex defects, suppresses background noise, and achieves low computational complexity.
2) We introduce a novel OLS strategy that enhances Mamba's ability to model spatial dependencies. OLS introduces a multidirectional and continuous scanning mechanism to capture the holistic structure of arbitrarily oriented defects, effectively mitigating information loss. Its differentiated forward–backward pathways further remedy the feature redundancy inherent in conventional scanning methods.
3) We introduce a novel GLVSS module that synergistically fuses multiscale local features with global context. Through adaptive global–local feature calibration, GLVSS dynamically balances feature responses to significantly improve the segmentation of defects with varying scales and locations, particularly small and edge-confined ones.
4) Comprehensive experiments on two PV defect segmentation datasets verify that PV-MambaSeg establishes a new benchmark in PV defect segmentation, attaining the highest mIoU with merely 2.77 M parameters and 0.78 GFLOPs. It delivers a superior tradeoff between segmentation accuracy and computational efficiency when compared to existing approaches.

II. Related Works

A. Deep Learning PV Defect Segmentation Methods

Deep learning has driven significant advancements in PV defect segmentation, transitioning the field from manual inspection to automated, high-efficiency defect localization and laying a solid foundation for intelligent PV plant operation and maintenance. Existing works have explored diverse technical paths to address defect segmentation challenges, with notable progress in specific scenarios and task settings.

A major subset of deep learning works focuses on segmenting targeted PV defect types. Qian et al. [12] presented a microcrack segmentation approach for PV modules that fuses short-term and long-term deep features. Zhao et al. [13] proposed the refined graph reasoning network which leverages the complementary information among layers and global long-range relational dependencies to learn discriminative feature representations of hotspots. Su et al. [14] developed a residual channelwise attention gate architecture to achieve multiscale feature merging, complex background suppression, and defect feature highlighting for PV farm hotspot detection. These studies have advanced the state-of-the-art in segmenting specific defect categories, demonstrating effective strategies for targeted defect characteristics.

Another line of research has focused on general defect region segmentation from PV cell images. Sovetkin et al. [15] utilized the encoder–decoder deep neural network architecture to perform defect segmentation on thin-film module electroluminescence (EL) images. Tang et al. [16] presented a linear defect segmentation approach for polycrystalline silicon PV cells, leveraging Hessian matrix-based defect feature extraction and multiscale line detector-enabled defect feature enhancement. Zhou et al. [17] presented a segmentation architecture based on interactive semantic information fusion. Wang et al. [1] adopted edge detection technology to investigate the precise segmentation of PV defects. These works have contributed to reliable defect region extraction, supporting basic quality control requirements.

Recent defect segmentation studies have begun to involve the differentiation of multitype defects by using existing methods from the natural domain. Fioresi et al. [3] utilized a Deeplabv3 segmentation model with a ResNet-50 backbone to segment PV defects. Mahboob et al. [11] proposed a defect detection network based on Segformer to segment different PV defects. These efforts have expanded the scope of defect segmentation tasks, moving toward more comprehensive defect characterization.

Overall, existing works have significantly promoted the automation and precision of PV defect segmentation, covering targeted defect segmentation, general defect segmentation, and multitype defect segmentation. However, practical industrial applications still require further improvements: balancing high-precision multitype defect discrimination with lightweight design for resource-constrained scenarios, enhancing adaptability to geometrically complex, arbitrarily oriented, and small-scale defects, and optimizing computational efficiency.

To tackle these challenges, we propose the first PV defect segmentation approach based on the visual state space model (SSM). It efficiently models complex defect structures with low computational overhead, achieving class-aware, high-precision segmentation suitable for practical PV defect analysis.

B. Visual SSM

SSMs have risen as a viable and promising alternative to transformers. Mamba is one of the most representative SSMs and was the first to achieve breakthroughs in natural language processing. Mamba's linear computational complexity relative to sequence length and selective state update mechanism enable efficient handling of long sequences. However, Mamba blocks were originally designed for 1-D sequences, a design constraint that creates a barrier for direct application to 2-D image data. To adapt SSMs to visual tasks, researchers have focused on designing scanning strategies that convert 2-D image patches into 1-D sequences while preserving spatial-semantic information. Vision Mamba [18] employs a bidirectional zigzag scanning strategy, which traverses image patches in both forward and backward zigzag paths to capture cross-region dependencies. VMamba [19] extended the scanning strategy by adding two vertical scanning paths, which enhanced the model's capacity to capture vertical spatial correlations. PlainMamba [20] further optimized this approach by proposing a continuous 2-D scanning method that maintains the spatial and semantic continuity of image patches.

These visual state space models (VSSMs) have demonstrated their strengths by retaining the linear computational complexity while achieving enhanced modeling capabilities for long-range relational dependencies, and they have demonstrated strong performance in general visual tasks [21]. Nevertheless, when applied to PV defect segmentation, existing VSSM scanning strategies still have limitations. PV defects exhibit geometrically complex and randomly oriented structures and are often embedded in noisy backgrounds which demands a customized, comprehensive multidirectional scanning strategy.

Most existing VSSM-based segmentation methods adopt single-directional or simple bidirectional scanning. However, these paradigms are ill-suited for capturing the structural continuity of multioriented PV defects. For instance, unidirectional scanning can fracture the semantic integrity of elongated, diagonal cracks, while conventional bidirectional approaches often introduce feature redundancy rather than complementary information. To better handle these challenges, we propose the OLS strategy. OLS introduces a multidirectional and differentiated scanning mechanism specifically designed to preserve holistic spatial continuity while mitigating feature redundancy. This approach enables a more robust semantic decomposition of PV defects, better adapting to their geometric complexity and the presence of background noise.

C. Mamba-Based Defect Segmentation Methods

In recent years, Mamba and SSM based architectures have gained significant traction as a compelling alternative to transformers in industrial defect segmentation, due to their inherent linear computational complexity and superior long-range dependency modeling capabilities. Existing research in this field has evolved along two distinct technical trajectories.

The first trajectory centers on adapting the standard scanning paradigms from foundational vision SSMs to specific industrial inspection scenarios. Most early Mamba-based defect segmentation works directly to adopt the conventional scanning paradigms from Vision Mamba, VMamba and PlainMamba, and tailor the overall network architecture to meet the specific requirements of diverse industrial applications. Bian et al. [22] integrated SSMs with convolutional layers to capture both global contextual information and local fine-grained details, achieving accurate segmentation of tiny industrial defects. Du et al. [23] designed an OSS Block to extract long-range spatial features under linear complexity, enabling efficient processing of large-scale bridge pier images. Huang [24] constructed a lightweight Mamba segmentation framework tailored for navel orange defect detection. Bao et al. [25] embedded a multidirectional visual state space module into YOLOv8 to realize real-time fabric defect inspection. Zhao et al. [26] combined visual SSMs with a multiaxis interactive feature pyramid to enhance feature representation for surface microdefect detection.

The second trajectory aims to mitigate the inherent limitation of vanilla SSMs in fine-grained local detail modeling by developing hybrid architectures that synergistically fuse

Mamba-derived global contextual features with CNN-derived local detail features. Wang et al. [27] built a strip steel defect detection network that merges global VMamba features and local CNN features through prototype interaction and dynamic multiscale attention. Zhao et al. [28] presented a Mamba-based multimodal fusion framework that couples dual-path encoders and hybrid scanning strategies to fuse RGB and point cloud data for industrial anomaly detection. Li et al. [29] developed a CNN-Mamba hybrid paradigm for fabric defect segmentation, where Mamba models low-frequency structural information and CNNs supplement high-frequency detail features.

Despite the promising results attained by these two research directions in diverse industrial scenarios, they still exhibit two critical limitations that impede their performance when applied to PV cell defect segmentation. First, the majority of existing methods lack task-specific scanning designs tailored to the unique spatial geometric characteristics of industrial defects, notably arbitrary orientations and discontinuous structural patterns. Second, insufficient research attention has been paid to the adaptive fusion and calibration of local and global features within visual state space modules, which severely limits their ability to suppress complex background interference induced by crystal dislocations in PV cells and accurately localize tiny, low-contrast defects.

To address these aforementioned limitations, we propose an optimized multidirectional continuous scanning strategy to effectively model the structural continuity and arbitrary orientation distribution of PV cell defects. Furthermore, we develop a GLVSS module that achieves adaptive aggregation of multiscale global and local features via multiwindow grouping and channel-aware calibration. Notably, the entire PV-MambaSeg framework is developed under lightweight design constraints, rendering it highly suitable for PV inspection scenarios.

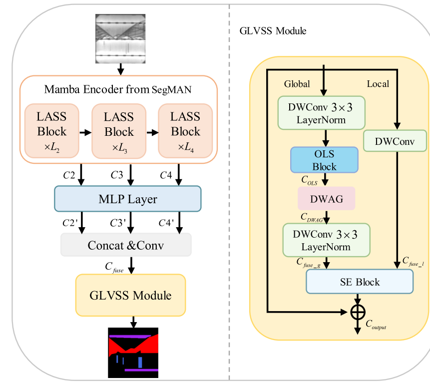

Fig. 3. Overall architecture of the Our PV-MambaSeg. The input image undergoes processing via a lightweight mamba encoder [21] to extract a set of feature maps. These feature maps are then fed into our proposed GLVSS to produce the final segmentation mask. The OLS block shown in the GLVSS module's global branch utilizes our proposed multidirectional scanning strategy.

III. Methodology

A. Preliminary

SSMs represent a class of sequential modeling frameworks that map input data into a continuous state space for selective feature update. They enable linear computational complexity ($O(N)$) while capturing long-range dependencies, which is an advantage over transformers with quadratic complexity ($O(N^2)$) [30]. The core of SSMs lies in discrete state transition and output projection, given by

$$h(t) = \overline{A}h(t-1) + \overline{B}x(t) \tag{1}$$

$$y(t) = Ch(t) + Dx(t). \tag{2}$$

Here, $\overline{A}$ and $\overline{B}$ are discretized parameters that control historical information persistence and input integration, respectively. C denotes the output projection matrix that transforms the hidden state to the output dimension and D is the direct input projection matrix that allows the input to bypass the state transition and contribute directly to the output. Though powerful for sequence modeling, traditional SSMs were designed for NLP and lack adaptability to 2-D visual data.

Vision Mamba [18] addresses these limitations by adapting SSMs to visual tasks. It retains linear complexity while introducing visual-friendly optimizations. It first translates 2-D feature maps into 1-D sequences with convolutional priors to preserve local spatial correlations. Then, it employs input-adaptive gating layers to refine selective state updates, focusing on task-relevant features while ignoring background clutter. For PV cell defect segmentation, Vision Mamba's efficiency and long-range modeling capability make it promising. However, existing scanning strategies, including single-directional or simple bidirectional with reversed order, struggle to capture multioriented defect continuity, suffer from feature redundancy and lack targeted noise suppression.

Building on the Vision Mamba paradigm, our work extends this direction toward PV defect segmentation. We propose PV-MambaSeg, the first Mamba-based network for this task. It is enhanced by the proposed OLS strategy and GLVSS module. OLS substantially redesigns the scanning process, replacing the standard linear scan with a multidirectional mechanism to effectively capture the holistic structure of arbitrarily oriented defects. Concurrently, GLVSS introduces a sophisticated fusion and calibration process that adaptively balances global context with fine-grained local details, which is crucial for identifying defects of varying scales. Collectively, these innovations allow PV-MambaSeg to inherit the high efficiency of SSMs while overcoming limitations of Vision Mamba in this complex industrial domain.

B. Overall Architecture

PV-MambaSeg network is built upon a lightweight encoder–decoder architecture, designed to achieve an excellent balance between high segmentation precision and computational efficiency. This streamlined pipeline is specifically optimized to capture geometrically complex defect patterns while suppressing the background noise in industrial PV cell images. As illustrated in Fig. 3, the input image is fed into a lightweight Mamba encoder. Specifically, we adopt the SegMAN encoder [21] as our backbone, which hierarchically extracts feature maps $C2$, $C3$, and $C4$ at different scales. Following the configuration in [21], the encoder consists of four stages, each comprising a 3×3 convolution for spatial downsampling followed by a series of local attention and state space (LASS) blocks. Each LASS block integrates neighborhood attention for local detail encoding and a 2-D selective scan (SS2D) for global context modeling, both operating in linear time. The resulting multiscale feature maps $C2$, $C3$, $C4$ correspond to the outputs of stages 2,

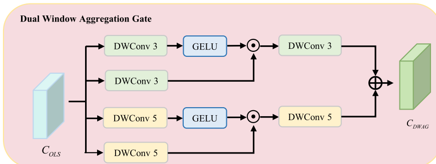

Fig. 4. Architecture of the DWAG. This module is designed to adaptively recalibrate the multidirectional features from the OLS block. It employs two parallel branches with varying kernel sizes for capturing multiscale contextual information. This information is then used to generate a pixelwise attention gate, which enhances features from informative scanning paths while suppressing noise.

3, and 4, respectively, with spatial resolutions of $H/8$, $H/16$, and $H/32$ of the input image. Subsequently, an MLP layer is adopted to align the channel dimensions and spatial sizes of these feature maps, ensuring they are compatible for subsequent fusion. After concatenation, the unified features are further fused via a 1×1 convolution to generate a fused feature map C_{fuse}, which aggregates multiscale information in a compact manner:

$$C_{\text{fuse}} = conv(concat(C2', C3', C4')). \tag{3}$$

Subsequently, C_{fuse} is fed into the proposed GLVSS module, where the proposed OLS strategy is instantiated as a dedicated OLS block in the module's global branch to drive high-quality global feature extraction. As an innovative scanning strategy, OLS performs multidirectional continuous scanning to comprehensively capture the structural continuity of arbitrarily oriented defects, thereby enabling improved geometry-aware modeling of irregular defect patterns that conventional scanning methods often struggle to capture. The GLVSS module further refines the global features via multiwindow aggregation and adaptive channel attention calibration, dynamically balancing the OLS-enhanced global context and fine-grained local details. This synergy effectively suppresses background noise while securing precise localization of small, edge-confined defects. Finally, the output from the GLVSS module C_{output} is processed, resulting in the final segmentation map. This streamlined architecture ensures that the network utilizes minimal computational resources while effectively handling the challenges of PV defect segmentation.

C. Global Local Visual State Space Module

The GLVSS module is devised to synergistically fuse global and local features, thereby enhancing defect representation while suppressing background noise. As shown in Fig. 3, it consists of two parallel branches, a feature fusion block, and a residual connection.

The global branch consists of depth-wise convolution (DWConv) followed by LayerNorm, an OLS block, and a dual window aggregation gate (DWAG). Built on the Mamba architecture, the OLS block effectively captures long-range structural dependencies of defects through powerful global feature extraction, while maintaining high computational speed and low memory usage. While OLS effectively generates multidirectional feature sequences, directly summing these sequences treats all scanning paths equally, which may amplify noise from less informative paths and dilute critical features. Accordingly, we introduce the DWAG to adaptively recalibrate the importance of each path. As shown in Fig. 4, DWAG employs parallel 3×3 DWConv branch and a 5×5 DWConv branch to create an attention gate, allowing it to dynamically aggregate multidirectional information based on varying defect scales. The output C_{DWAG} is computed as:

$$G_k = GELU(DWConv_k(C_{\text{OLS}})) \tag{4}$$

$$C_{\text{DWAG}} = \sum_{k \in \{3,5\}} DWConv_k(C_{\text{OLS}} \odot G_k) \tag{5}$$

where k denotes the kernel size, $\odot$ is elementwise multiplication, and DWConv_k represents depthwise convolution with a $k \times k$ kernel. Finally, C_{DWAG} undergoes a 3×3 DWConv followed by LayerNorm to generate the final global branch output $C_{\text{fuse\_g}}$:

$$C_{\text{fuse\_g}} = LayerNorm(DWConv(C_{\text{DWAG}})). \tag{6}$$

The local branch runs in parallel and is designed to preserve fine-grained details, such as defect edges and textures. It consists of a single DWConv layer that processes the initial input feature C_{fuse} to strengthen local feature representation. The local branch output $C_{\text{fuse\_l}}$ is computed as

$$C_{\text{fuse\_l}} = DWConv(C_{\text{fuse}}). \tag{7}$$

The outputs of the global branch $C_{\text{fuse\_g}}$ and the local branch $C_{\text{fuse\_l}}$ are then fused synergistically. A squeeze-excitation (SE) block is utilized to adaptively recalibrate the channels of the concatenated features, enhancing defect-relevant information while suppressing background noise. Finally, a residual connection is introduced to accelerate network convergence and reduce information loss during feature propagation. The final output C_{out} is computed as

$$C_{\text{output}} = C_{\text{fuse}} + SE(concat(C_{\text{fuse\_g}}, C_{\text{fuse\_l}})). \tag{8}$$

D. Omni-Local Scan Strategy

To comprehensively model geometrically complex and arbitrarily oriented defects, the GLVSS module incorporates the OLS block. As illustrated in Fig. 5, the OLS block partitions the input feature map into multiple image patches. It then performs parallel scanning and diagonal scanning on these patches. The sequences generated from these multidirectional scans are subsequently processed using S6 equations from the Mamba block [30] and aggregated via summation. The output feature C_{OLS} exhibits improved geometric awareness for irregular defect patterns and fully captures the structural continuity of defects. Specifically, for scanning sequences R_i $(i = 0, 1, \ldots, 7)$, the calculation of the output feature map C_{OLS} is given by

$$C_{\text{OLS}} = \sum_{i=0}^{7} S6(R_i). \tag{9}$$

As illustrated in Fig. 6, to comprehensively model various PV defects with improved geometric perception in multiple directions, the OLS strategy consists of four parallel scan paths and four diagonal scan paths. In Fig. 6, image patches of the same color denote continuous local semantics, visually reflecting the semantic coherence that the scanning pathways aim to capture.

For parallel scanning, a local-parallel scan (LPS) is proposed. As shown in Fig. 6(a), the bidirectional zigzag parallel scan adopted in VMamba [19] fails to consider the semantic continuity of adjacent image patches, leading to potential performance degradation. PlainMamba [20] employs continuous bidirectional parallel scanning, yet the forward and backward paths are identical, resulting in redundant feature extraction and loss of local information. In contrast, LPS utilizes differentiated

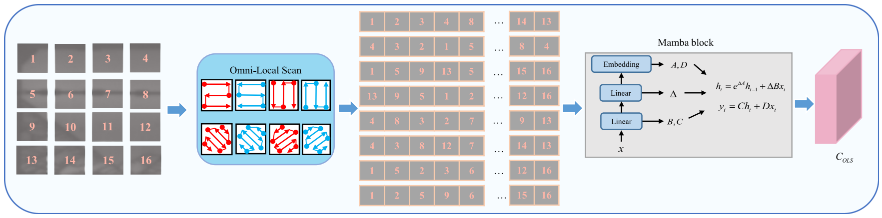

Fig. 5. Operational mechanism of the OLS block. OLS converts the 2-D feature map into multiple 1-D sequences by traversing the image patches in eight distinct directions. These sequences are then modeled using the S6 equations, and their outputs are fused through summation to create a unified representation that captures holistic structural continuity.

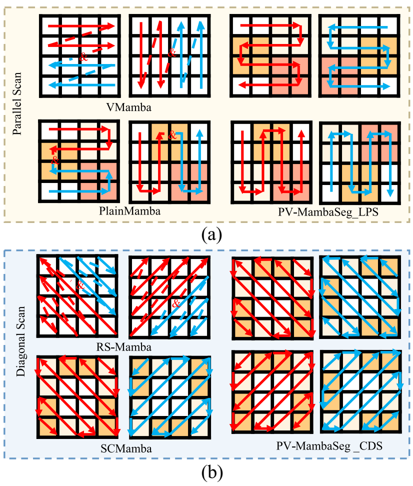

Fig. 6. Visual comparison of different Mamba scanning strategies. (a) Parallel Scan methods: Unlike VMamba's discontinuous scan and PlainMamba's redundant bidirectional scan, our proposed LPS uses differentiated continuous paths to ensure comprehensive and nonredundant feature coverage. (b) Diagonal Scan methods: Compared to RS-Mamba's discontinuous zigzag scan and SCMamba's unidirectional approach, our CDS performs continuous bidirectional scanning on both main and minor diagonals to capture holistic structural information. Patches of the same color represent a continuous semantic region.

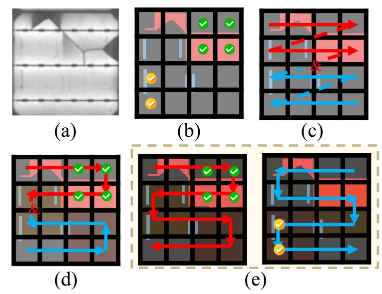

Fig. 7. Visualization of defect semantic capture capability of different scanning strategies on PV cell images. (a) Original PV cell image. (b) Defect mask patches, where the marked patches represent local defect semantics. (c) Zigzag scan of VMamba [19] fails to capture continuous defect semantics due to discontinuous path design. (d) Bidirectional scan of PlainMamba [20] adopts identical forward and backward paths, which can only capture the defect semantics in the upper-right region but misses the defect semantics in the lower-left region. (e) Our proposed OLS leverages differentiated forward and backward scanning pathways, which effectively captures the defect semantics in both the upper-right and lower-left regions, reduces feature redundancy, and enhances the detection of edge defects.

forward and backward scanning pathways to extract distinct local semantics, avoiding feature redundancy and effectively mitigating the missed detection of edge defects.

For diagonal scanning, a cross-diagonal scan (CDS) is proposed. As shown in Fig. 6(b), the bidirectional zigzag scan in RS-Mamba [31] neglects semantic continuity. SCMamba [32] uses continuous unidirectional diagonal scanning, which loses information from reverse scanning. CDS performs continuous scanning in both the main diagonal and minor diagonal directions, enabling it to comprehensively capture the structural continuity of defects.

In contrast to conventional bidirectional scanning schemes that rely on symmetric reversal, our OLS employs differentiated forward and backward scanning paths that generate two structurally independent feature sequences. Mathematically, let the forward scanning sequence be denoted as $S_{\mathrm{fw}} = [s_1, s_2, \ldots, s_N]$, where N represents the total number of image patches and s_i denotes the feature of the i_{th} patch. For existing methods, the backward scanning sequence S_{bw} is merely the elementwise reversal of the forward sequence, i.e., $S_{\mathrm{bw}} = [s_N, s_{N-1}, \ldots, s_1]$. This results in a trajectory that strictly retraces the forward path, leading to significant feature redundancy. In contrast, our designed backward scanning sequence satisfies $S_{\mathrm{bw}}^{\mathrm{ours}} \neq \mathrm{Reverse}(S_{\mathrm{fw}})$, following a differentiated traversal rule that breaks the spatial symmetry of conventional bidirectional scanning. This design ensures that $S_{\mathrm{bw}}^{\mathrm{ours}}$ captures complementary spatial contexts distinct from S_{fw}, while maintaining local semantic continuity.

The visualization results in Fig. 7 further verify the superiority of the proposed differentiated bidirectional scanning in capturing defect semantics. In contrast to VMamba's zigzag scan and PlainMamba's identical bidirectional scan, both of which struggle to fully capture continuous defect semantics or miss partial regions, our OLS leverages differentiated forward and backward scanning pathways to accurately capture defect semantics in all target regions, illustrating its effectiveness in reducing feature redundancy and enhancing edge defect detection.

IV. Experimental Settings and Results

A. Experimental Setup

1) Dataset: To evaluate the proposed method, experiments are conducted on two public PV cell defect segmentation datasets which are UCF-EL and PSCDE. The UCF-EL dataset [3] contains 17 064 real PV cell images including 4114 defect-free images and the rest with at least one type of defect (crack, contact, interconnect (Inter.), corrosion (Corr.)) with each image accompanied by a pixel-level annotation mask. Following the standard experimental protocol in the original work [3], the UCF-EL dataset is randomly divided into training,

TABLE I
PV Cell Defect Segmentation Results on UCF-EL and PSCDE Datasets.

Method	Ref.	UCF-EL Test						PSCDE Test			Computational Efficiency			
		mIoU	**Crack**	**Contact**	**Inter.**	**Corr.**	**Backg.**	**mIoU**	**Defect**	**Backg.**	**Params ↓**	**GFLOPs ↓**	**Model size ↓**	**FPS ↑**
U-Net [35]	MICCAI 2015	0.3776	0.5244	0.2766	0.0153	0.1277	0.9441	0.7185	0.4952	0.9418	13.41	382.07	51.16	238.46
DeepLabv3* [33]	Arxiv 2017	0.5121	0.6165	0.2934	0.0500	0.6437	0.9571	0.7262	0.5003	0.9520	39.12	498.26	149.27	93.01
Segformer* [34]	NIPS 2021	0.5202	0.6055	0.4087	0.0215	0.6077	0.9575	0.7575	0.5587	0.9562	13.68	40.48	52.20	124.46
STDC [36]	CVPR 2021	0.4220	0.4412	0.1995	0.0000	0.5408	0.9286	0.6402	0.3331	0.9473	12.97	35.99	34.10	122.66
DDRNet [37]	TIP 2022	0.5188	0.6537	0.3488	0.0135	0.6185	0.9592	0.6708	0.3910	0.9506	20.18	55.19	77.02	105.28
PIDNet [38]	CVPR 2023	0.5302	0.6367	0.3236	$\underline{0.1340}$	0.6020	0.9545	0.7556	0.5593	0.9518	28.81	68.51	39.58	79.05
Seaformer [39]	ICLR 2023	0.4992	0.5884	0.3337	0.1244	0.4935	0.9558	0.7074	0.4625	0.9523	8.65	5.59	$\underline{31.80}$	35.74
CGRSeg [40]	ECCV 2024	0.4856	0.6237	0.2796	0.0449	0.5232	0.9568	0.7033	0.4433	$\underline{0.9633}$	18.10	7.60	87.50	65.30
SCTNet [41]	AAAI 2024	0.3074	0.5857	0.0000	0.0000	0.0000	0.9513	0.6960	0.4372	0.9548	17.40	8.50	94.46	160.58
SegMAN [21]	CVPR 2025	$\underline{0.5580}$	$\underline{0.6716}$	0.4036	0.0839	$\underline{0.6693}$	0.9618	0.6635	0.3703	0.9567	$\underline{6.38}$	$\underline{1.01}$	73.53	$\underline{459.47}$
GCNet [42]	CVPR 2025	0.4178	0.6078	0.2844	0.0029	0.2271	$\underline{0.9668}$	$\underline{0.7742}$	$\underline{0.6196}$	0.9314	9.20	45.20	169.26	95.22
BiDNet [43]	TIM 2025	0.4690	0.5855	$\underline{0.4137}$	0.0003	0.4223	0.9232	0.7134	0.4831	0.9437	4.82	7.35	57.49	58.25
DS^2A-Former [44]	TIM 2025	0.5448	0.6387	0.3952	0.0896	0.6432	0.9557	0.7626	0.5912	0.9341	47.51	242.56	188.32	14.87
OCINet [45]	TIM 2025	0.5326	0.5776	0.4062	0.1530	0.5734	0.9530	0.7817	0.6274	0.9360	29.95	9.60	120.84	185.62
MPKG-Net [46]	TGRS 2026	0.5131	0.6012	0.3644	0.0682	0.5833	0.9486	0.7426	0.5287	0.9462	71.16	321.04	684.80	9.05
CenterSeg [47]	TGRS 2026	0.5068	0.5924	0.3512	0.0725	0.5718	0.9463	0.7386	0.5214	0.9458	59.20	236.81	31.73	15.44
PV-MambaSeg	Ours	**0.6181**	**0.7844**	**0.4162**	**0.1750**	**0.7451**	**0.9696**	**0.8664**	**0.7654**	**0.9673**	**2.77**	**0.78**	**31.73**	**548.04**

DeepLabv3* [33] and segformer* [34] are existing models adapted for multiclass PV cell defect segmentation. Best results are bold, second-best are underlined.

validation, and test sets with a ratio of 8:1:1 to follow the common experimental convention, while maintaining sufficient training samples for model optimization and reserving adequate data for fair and reliable performance evaluation. Meanwhile the PSCDE dataset [17] is a high-quality dataset consisting of 700 challenging defect images with diverse defect types and fine-grained annotations, which is further split into training, validation, and test sets with a ratio of 4:2:4 to maintain a sufficiently challenging test set while retaining enough validation data for model selection.

2) Implementation details: PV-MambaSeg network was implemented using the PyTorch framework. All experiments were conducted on a machine equipped with an NVIDIA GeForce RTX 3090 GPU and an Intel Core i9-10980 XE CPU. The AdamW optimizer is adopted, with an initial learning rate of 6×10^{-5}, a weight decay of 0.01, a warm-up period of 1500 iterations, and a batch size set to 16. The input image size is uniformly set to 224.

3) Evaluation Metrics: For a comprehensive assessment of the proposed PV-MambaSeg network's performance, we use a combination of segmentation accuracy metrics and model complexity metrics.

For segmentation accuracy, the mean Intersection over Union (mIoU) is used as the metric, which measures the average overlap between predicted defect regions and ground truth masks. Mathematically, mIoU is computed as

$$mIoU = \frac{1}{K+1} \sum_{k=0}^{K} \frac{\mathrm{TP}_k}{\mathrm{TP}_k + \mathrm{FP}_k + \mathrm{FN}_k} \tag{10}$$

where K denotes the number of defect classes, TP_k is the count of pixels correctly classified as class k, FP_k is the count of pixels incorrectly classified as class k, and FN_k is the count of pixels belonging to class k but misclassified.

To evaluate model efficiency, we further employ four complexity metrics. floating point operations (FLOPs) measures computational load by quantifying the number of arithmetic operations required for inference. Params (Parameters) quantifies the number of trainable parameters to reflect the model's storage and optimization complexity. Model size represents the memory footprint when the model is stored. Frames per second (FPS) evaluates the model's inference speed by measuring the number of images processed per second. These metrics collectively offer a balanced evaluation of both segmentation accuracy and practical efficiency.

B. Comparison With SOTA Methods

Table I presents a comprehensive comparison of the proposed PV-MambaSeg with 16 state-of-the-art semantic segmentation methods, evaluated using mIoU and class-specific IoU scores. Among these methods, DeepLabv3 [33] and Segformer [34] are existing models adapted for PV cell defect segmentation. We further include three specialized defect segmentation methods, namely BiDNet [43], DS^2A-Former [44], and OCINet [45], to verify the effectiveness of our approach on industrial defect scenarios. The remaining 11 methods are representative natural-scene semantic segmentation models fine-tuned for PV cell defect segmentation task.

Overall, PV-MambaSeg achieves a leading mean IoU of 0.6181 on the UCF-EL dataset, outperforming the second-best method, SegMAN, by a substantial margin of 6.01% points. It also attains superior results on the PSCDE dataset with a mean IoU of 0.8664, surpassing the next-best GCNet by 9.22% points. This consistent superiority stems from two core designs tailored for PV defects: the OLS strategy enables robust geometric modeling of complex defect morphologies, while the GLVSS module with DWAG adaptively fuses global–local features to suppress background noise and enhance defect representation.

For crack segmentation, PV-MambaSeg achieves an IoU of 0.7844, effectively modeling thin, irregular crack structures through the OLS strategy, which captures structural continuity and multiorientation features. For contact defects, it achieves the highest IoU of 0.4162, outperforming Segformer and demonstrating robust feature discrimination via multiscale local feature aggregation. For the challenging interconnect defect class, PV-MambaSeg sets a new state-of-the-art with an IoU of 0.1750, outperforming PIDNet by 4.10% points, as the DWAG mechanism dynamically emphasizes defect-rich regions and mitigates noise interference. For corrosion defects, it achieves 0.7451, confirming its capability to capture large, irregular defect regions through global–local feature fusion. On the PSCDE dataset,

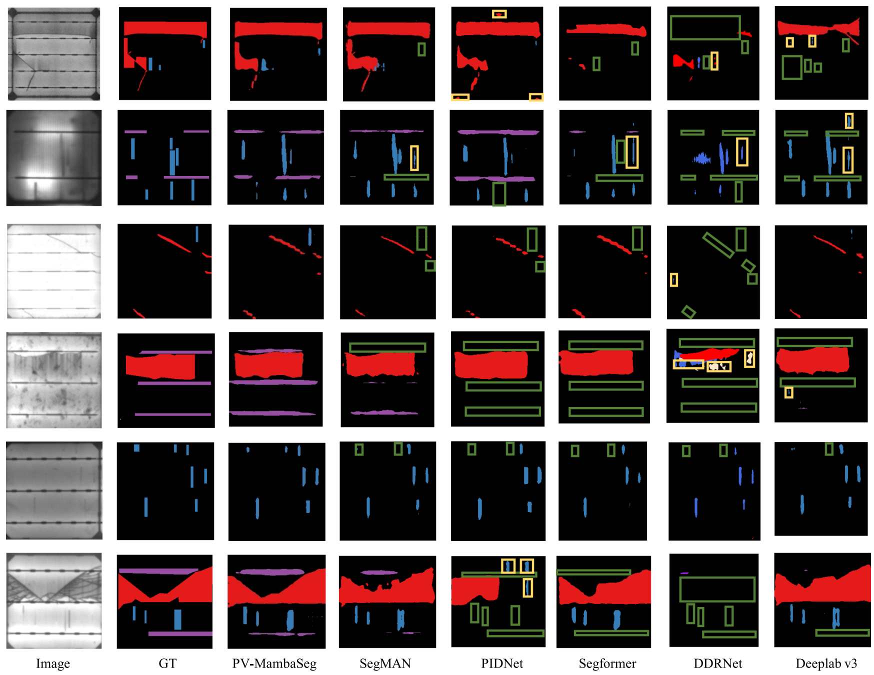

Fig. 8. Visualization comparison of segmentation results between our PV-MambaSeg and five SOTA segmentation methods on PV cell defect images. From left to right: original images, ground truth, PV-MambaSeg results, and other SOTA outputs. Green boxes mark missed detections and yellow boxes mark misidentified regions. PV-MambaSeg accurately segments multitype, multiorientation PV cell defects, thanks to the OLS strategy and GLVSS module's DWAG. It outperforms other methods in reducing missed small/arbitrarily oriented defects and cluttered background false positives, verifying its effectiveness.

PV-MambaSeg achieves competitive defect and background IoU scores, suggesting effective mitigation of background noise. These results indicate that the integration of OLS and GLVSS offers a feasible way to tackle prevalent challenges in PV defect segmentation, such as low contrast, scale variation, irregular morphology, and background interference, and yields relatively consistent performance improvements across different defect categories.

Fig. 8 presents a qualitative visualization comparison between our PV-MambaSeg and five SOTA methods. As shown in the comparison results, other methods suffer from obvious missed detections (marked by green boxes) and false positives (marked by yellow boxes), especially for tiny defects, arbitrarily oriented defects, and defects in cluttered background regions. In contrast, PV-MambaSeg generates more complete, continuous, and accurate segmentation masks, effectively reducing missed detections and false recognitions. These visualization results intuitively confirm the advantages of the proposed method in handling multitype and multiorientation PV cell defects under complex background conditions.

C. Complexity Analysis

In Table I, we conduct a comparison of the computational complexity between PV-MambaSeg and SOTA methods using four metrics: Params, FLOPs, model size, and FPS to assess engineering applicability. PV-MambaSeg achieves remarkable efficiency with only 2.77 M parameters, which is 56.6% fewer than the lightest existing method SegMAN. In terms of computational load, it requires merely 0.78 GFLOPs, outperforming SegMAN by 22.8% and far surpassing heavier methods like DeepLabv3 and U-Net. For model size, PV-MambaSeg occupies 31.73 MB, slightly outperforming Seaformer's underlined 31.80 MB. In terms of inference speed, PV-MambaSeg achieves an outstanding 548.04 FPS, outperforming the second-best method SegMAN by 19.3%.

This efficiency stems from two key designs: 1) The OLS block replaces dense self-attention with directional scanning, reducing redundant computations; 2) The DWAG uses lightweight depthwise convolutions instead of heavy fully connected layers for adaptive feature weighting. Thus, PV-MambaSeg balances high segmentation accuracy and low computational cost, meeting the practical requirements of PV cell defect segmentation.

D. Industrial Inspection Performance Evaluation

To further verify the practical effectiveness of the proposed PV-MambaSeg in real industrial PV cell inspection, we conduct an evaluation dedicated to industrial-oriented metrics and realistic data distribution. Considering that actual production lines contain far more normal PV cells than defective ones, we optimize both the training and test sets to better match engineering practice. Specifically, we enrich the original UCF-EL training set by adding 2000 real-world normal PV cell images, enabling all models to learn the appearance of defect-free samples. Meanwhile, we construct a new private test set that consists of 50 defective images and 500 normal images, yielding a normal-to-defect ratio of 10:1, which simulates defect rate in practical PV manufacturing. Its 50 defective images are randomly selected from the UCF-EL dataset, while 500 normal images are collected from the public ELPV dataset [55]. All comparison methods and our PV-MambaSeg are retrained and tested under the same experimental settings for fair evaluation.

For evaluation, we employ important industrial metrics including missed detection rate (MDR) and false alarm rate (FAR). Lower MDR indicates fewer missed defects, while lower

TABLE II
DEFECT INSPECTION PERFORMANCE COMPARISON ON AN INSPECTION-ORIENTED PV CELL TEST SET

Method	MDR ↓	FAR ↓	Precision ↑	Recall ↑	F1-Score ↑
U-Net [35]	0.1400	0.0531	0.6143	0.8600	0.7168
DeepLabv3* [33]	0.0400	0.0374	0.7164	0.9600	0.8204
Segformer* [34]	<u>0.0200</u>	<u>0.0321</u>	0.7538	<u>0.9800</u>	0.8522
STDC [36]	0.0800	0.0492	0.6479	0.9200	0.7604
DDRNet [37]	0.0400	0.0361	0.7273	0.9600	0.8276
PIDNet [38]	<u>0.0200</u>	0.0301	<u>0.7656</u>	<u>0.9800</u>	<u>0.8596</u>
Seaformer [39]	0.0400	0.0361	0.7273	0.9600	0.8276
CGRSeg [40]	0.0600	0.0453	0.6417	0.9400	0.7834
SCTNet [41]	0.1000	0.0689	0.5625	0.9000	0.6924
SegMAN [21]	<u>0.0200</u>	0.0335	0.7424	<u>0.9800</u>	0.8450
GCNet [42]	0.0400	0.0472	0.6667	0.9600	0.7869
BiDNet [43]	0.0600	0.0532	0.6351	0.9400	0.7581
DS^2A-Former [44]	0.0400	0.0394	0.7059	0.9600	0.8136
OCINet [45]	0.0400	0.0433	0.6857	0.9600	0.8000
MPKG-Net [46]	0.0400	0.0512	0.6486	0.9600	0.7742
CenterSeg [47]	0.0600	0.0472	0.6620	0.9400	0.7770
PV-MambaSeg	**0.0000**	**0.0098**	**0.9091**	**1.0000**	**0.9524**

Best results are bold, second-best are underlined.

FAR reduces unnecessary maintenance costs caused by false warnings. As shown in Table II, our method achieves 0.0000 MDR and 0.0098 FAR, remarkably outperforming all competing approaches. In addition, PV-MambaSeg demonstrates favorable precision, recall, and F1-score. These results demonstrate that the proposed method can accurately identify defective samples without missing real defects, while effectively suppressing false alarms on normal samples. The favorable performance on industrial metrics validates that PV-MambaSeg is well-adapted to real PV inspection scenarios.

E. Ablation Studies

To validate the effectiveness of key components in the proposed PV-MambaSeg, especially the design rationale of the GLVSS module and its internal components, we conduct systematic ablation experiments on the representative UCF-EL dataset [3]. We also perform comparative evaluations between our proposed OLS strategy and other scanning strategies. All experiments share the same encoder and training settings to ensure fairness.

1) Ablation study of GLVSS module: To validate the effectiveness of the proposed GLVSS module as a segmentation head, we compare it with four representative segmentation heads FCNHead [48], ASPPHead [33], FPNHead [49], and UPerHead [50] under the same experimental setup. All segmentation heads share the same lightweight encoder to ensure a fair comparison, with results summarized in Table III. As demonstrated in Table III, the GLVSS module achieves the highest performance across all metrics, with a mean IoU of 0.6181, which is 11.11% higher than the second-best FPNHead. This significant improvement is consistent across all defect categories. In terms of efficiency, GLVSS maintains the lowest computational cost which requires only 2.77 M parameters and 0.78 GFLOPs, slightly outperforming FCNHead and far exceeding heavier decoders like UPerHead. Its model size of 31.73 MB is also the smallest. These results confirm that the GLVSS module achieves a better balance between segmentation accuracy and efficiency in comparison to mainstream segmentation head designs, validating its suitability for PV cell defect segmentation task.

2) Ablation study of components in GLVSS module: For the purpose of further exploring the individual and synergistic contributions of the core components within the GLVSS module namely the OLS block, DWAG, and SE fusion module, we conduct incremental ablation experiments. Table IV summarizes the results, where $\surd$ indicates the presence of a component and ✗ indicates its absence. All configurations share the same lightweight encoder and training settings, with similar computational costs. When used individually, the OLS yields the highest mIoU, highlighting its dominant role in capturing multi-directional defect features. The OLS+DWAG combination achieves the highest mIoU among two-component setups, indicating that DWAG's adaptive multiscale aggregation complements OLS's feature extraction by suppressing noise in complex backgrounds. The full GLVSS module (OLS+DWAG+SE) achieves the highest mIoU of 0.6181, and optimal performance across all defect categories. These results validate that all three components OLS, DWAG, and SE are indispensable and function synergistically: OLS lays the foundation for capturing geometrically complex defects, DWAG enhances feature aggregation robustness, and SE optimizes global–local feature alignment. Their integration enables GLVSS to achieve improved segmentation accuracy without increasing computational overhead.

To further verify the rationality of the DWAG module's multiscale convolution kernel design and the effectiveness of the local branch feature enhancement strategy, additional ablation experiments were conducted, with results presented in Table V. All experimental settings were consistent with the aforementioned component ablation to ensure the fairness and comparability of the results.

For the DWAG module, ablation experiments were performed on different convolution kernel sizes and their pairwise combinations to explore the optimal kernel configuration for PV cell defect segmentation. As shown in the DWAG Ablation group, single-kernel configurations exhibit limited performance, with 3×3 Conv outperforming 5×5, 7×7, and 1×1 Conv. This is because 3×3 convolution can effectively capture fine-grained defect details common in PV cells, while 5×5 convolution complements it by capturing slightly larger defect regions and contextual information. Among all pairwise combinations, the $3 \times 3 + 5 \times 5$ Conv combination balances performance and rationality for PV cell defects, effectively capturing both fine and moderate-scale defects without redundant computation. Notably, all DWAG configurations maintain similar computational costs, confirming that our selected $3 \times 3 + 5 \times 5$ kernel combination is suitable for PV cell defect segmentation.

To evaluate the effectiveness of our local branch feature enhancement strategy, we compared the convolutional block in local branch against four representative feature enhancement methods, including SE [51], CBAM [52], CA [53], and SKNet [54] to verify its effectiveness for local PV cell defect extraction. Results show that our convolution-based local branch, when integrated into the full GLVSS module, outperforms all comparative methods. This demonstrates that our local branch can effectively extract local defect features without additional computational burden, whereas traditional attention-based methods introduce unnecessary complexity, leading to potential noise amplification and higher overhead. Notably, in our GLVSS architecture, the SE [51] module is deliberately positioned after the fusion of global and local branches. Since local features contain rich high-frequency details, premature enhancement prior to fusion could lead to detail loss. By applying

TABLE III
ABLATION STUDY OF GLVSS MODULE

Method	mIoU↑	Crack	Contact	Inter.	Corr.	Backg.	Params ↓	GFLOPs ↓	Model size ↓	FPS ↑
FCNHead [48]	0.4853	0.6060	0.2074	0.0651	0.6005	0.9475	2.81	0.79	32.24	386.56
ASPPHead [33]	0.4084	0.5609	0.1061	0.0003	0.4313	0.9433	16.94	1.36	194.33	244.46
FPNHead [49]	0.5070	0.6313	0.3514	0.0071	0.5854	0.9598	3.22	1.13	37.38	266.81
UPerHead [50]	0.4775	0.6207	0.3764	0.0002	0.4318	0.9585	29.65	40.60	439.86	110.37
GLVSS	**0.6181**	**0.7844**	**0.4162**	**0.1750**	**0.7451**	**0.9696**	**2.77**	**0.78**	**31.73**	**548.04**

Optimal results are shown in bold, and the second-optimal results are underlined.

TABLE IV
ABLATION OF DIFFERENT COMPONENTS IN GLVSS MODULE

OLS	DWAG	SE	mIoU	Crack	Contact	Inter.	Corr.	Backg.	Params ↓	GFLOPs↓	Model size↓	FPS ↑
✓	✗	✗	0.5304	0.6568	0.3703	0.0314	0.6329	0.9606	**2.68**	**0.70**	**30.69**	559.85
✗	✓	✗	0.5141	0.6530	0.3333	0.0237	0.6010	0.9596	2.70	0.72	30.89	**641.66**
✗	✗	✓	0.5194	0.6455	0.3228	0.0483	0.6211	0.9594	2.75	0.71	31.48	637.17
✓	✓	✗	0.5656	0.6714	0.3969	0.1019	0.6953	0.9623	2.71	0.75	30.90	555.63
✓	✗	✓	0.5625	0.6750	0.3743	0.1524	0.6482	0.9624	2.75	0.74	31.48	550.22
✗	✓	✓	0.5632	0.6745	0.3791	0.1213	0.6789	0.9622	2.77	0.76	31.70	630.03
✓	✓	✓	**0.6181**	**0.7844**	**0.4162**	**0.1750**	**0.7451**	**0.9696**	2.77	0.78	31.73	548.04

Optimal results are shown in bold, and the second-optimal results are underlined.

TABLE V
ABLATION STUDIES AND COMPARISONS OF DWAG MODULE AND LOCAL BRANCH

Group	Configuration	mIoU↑	Crack	Contact	Inter.	Corr.	Backg.	Params ↓	GFLOPs ↓	Model size ↓	FPS ↑
DWAG Ablation	Only 1×1 Conv	0.5612	0.7012	0.3615	0.1013	0.6914	0.9508	**2.70**	**0.72**	**31.02**	**562.30**
	Only 3x3 Conv	0.5722	0.7123	0.3806	0.1124	0.7025	0.9532	2.73	0.75	31.20	559.10
	Only 5×5 Conv	0.5681	0.7089	0.3724	0.1081	0.6987	0.9523	2.93	0.75	31.25	556.60
	Only 7×7 Conv	0.5643	0.7052	0.3645	0.1002	0.6959	0.9556	2.74	0.76	31.35	552.10
	1×1 + 3×3 Conv	0.5775	0.7162	0.3885	0.1206	0.7082	0.9540	2.74	0.76	31.35	554.40
	1×1 + 5×5 Conv	0.5807	0.7195	0.3976	0.1215	0.7103	0.9544	2.74	0.76	31.40	552.70
	1×1 + 7×7 Conv	0.5773	0.7138	0.3861	0.1172	0.7159	0.9537	2.75	0.77	31.45	550.20
	3×3 + 7×7 Conv	0.5793	0.7211	0.3848	0.1236	0.7122	0.9549	2.75	0.77	31.45	549.50
	5×5 + 7×7 Conv	0.5819	0.7280	0.3923	0.1220	0.7126	0.9546	2.75	0.77	31.50	548.80
Local Branch Ablation	SE [51]	0.5599	0.7035	0.3652	0.0814	0.6985	0.9510	2.82	0.83	32.10	538.30
	CBAM [52]	0.5341	0.6379	0.3710	0.0066	0.7012	0.9537	2.91	0.88	33.05	526.30
	CA [53]	0.5262	0.6612	0.3308	0.1003	0.5864	0.9523	2.87	0.85	32.60	532.10
	SKNet [54]	0.5564	0.6885	0.3934	0.1051	0.6429	0.9520	2.95	0.91	33.52	521.70
Ours		**0.6181**	**0.7844**	**0.4162**	**0.1750**	**0.7451**	**0.9696**	2.77	0.78	31.73	548.04

Optimal results are shown in bold, and the second-optimal results are underlined.

TABLE VI
ABLATION STUDY OF DIFFERENT SCANNING STRATEGIES

Method	Paradigm	Directions ↓	mIoU ↑	Crack	Contact	Inter.	Corr.	Backg.	Time(s)↓
Vision Mamba [18]	Zigzag	2	0.5289	0.6577	0.3514	0.0457	0.6290	0.9609	**0.0043**
VMamba [19]	Zigzag	4	0.5442	0.6951	0.3813	0.0258	0.6566	0.9622	0.0058
PlainMamba [20]	Continuous	4	0.5647	0.7223	0.4005	0.0672	0.6697	0.9640	0.0058
RS-Mamba [31]	Zigzag	8	0.5610	0.7344	0.4039	0.0198	0.6810	0.9661	0.0058
SCSegamba [32]	Continuous	4	0.5791	0.7492	0.4019	0.0479	0.7292	0.9671	0.0058
OLS	Continuous	8	**0.6181**	**0.7844**	**0.4162**	**0.1750**	**0.7451**	**0.9696**	0.0071

Optimal results are shown in bold, and the second-optimal results are underlined.

SE post-fusion, the model leverages global context to guide local feature screening, thereby achieving more discriminative local detail enhancement from a global perspective compared to simple local enhancement.

3) Ablation study of different scan strategies: We further compare the proposed OLS strategy with five representative mainstream scanning strategies from existing Mamba-based models, including Vision Mamba [18], VMamba [19], PlainMamba [20], RS-Mamba [31], and SCSegamba [32] to validate their effectiveness in capturing geometrically complex PV cell defects. These strategies encompass both zigzag and continuous scanning paradigms, as well as varying numbers of directions, enabling a comprehensive evaluation of OLS. Importantly, RS-Mamba also uses eight directions, yet OLS substantially outperforms it, confirming that the gain comes from the differentiated continuous design rather than direction quantity. All comparative experiments are conducted under identical experimental settings to ensure a fair evaluation. As presented in Table VI, the OLS strategy attains the optimal overall performance, outperforming the second-best SCSegamba by 3.90%. This superiority

TABLE VII
ABLATION OF DIFFERENT SCAN ROUTES, WHERE R0-R7 DENOTE DIFFERENT SCAN ROUTES IN THE OLS

R0	R1	R2	R3	R4	R5	R6	R7	mIoU ↑	Crack	Contact	Inter.	Corr.	Backg.	Time(s)↓
✗	✓	✓	✓	✓	✓	✓	✓	0.6101	0.7762	0.4108	0.1702	0.7389	0.9689	0.0061
✓	✗	✓	✓	✓	✓	✓	✓	0.6094	0.7750	0.4095	0.1698	0.7377	0.9688	0.0061
✓	✓	✗	✓	✓	✓	✓	✓	0.6083	0.7736	0.4089	0.1686	0.7368	0.9687	0.0062
✓	✓	✓	✗	✓	✓	✓	✓	0.6097	0.7745	0.4101	0.1695	0.7380	0.9688	0.0041
✓	✓	✓	✓	✗	✓	✓	✓	0.6109	0.7768	0.4110	0.1705	0.7392	0.9689	0.0062
✓	✓	✓	✓	✓	✗	✓	✓	0.6089	0.7740	0.4096	0.1691	0.7375	0.9688	0.0062
✓	✓	✓	✓	✓	✓	✗	✓	0.6112	0.7771	0.4115	0.1709	0.7395	0.9690	0.0062
✓	✓	✓	✓	✓	✓	✓	✓	**0.6181**	**0.7844**	**0.4162**	**0.1750**	**0.7451**	**0.9696**	0.0071

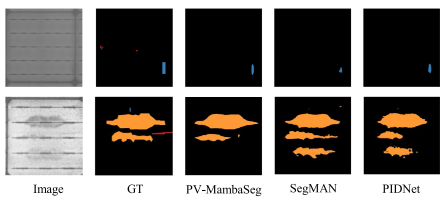

Fig. 9. Representative failure cases. From left to right, the columns are original images, ground truth, predictions of PV-MambaSeg and other methods. The top row shows missed segmentation of tiny, low-contrast cracks. The bottom row illustrates missed segmentation of crack and interconnect in a noisy background.

is consistent across all defect categories. In terms of efficiency, OLS requires 0.0071 s per complete scan, which is slightly longer than Vision Mamba's underlined 0.0043 s but comparable to other strategies.

We also conducted ablation experiments on OLS's scan routes. Table VII indicates that removing any single route of OLS leads to a performance drop compared with the complete 8-route OLS, and the full OLS achieves the best overall performance. This indicates that each scan route contributes meaningful defect information and that the eight routes are complementary rather than redundant.

Overall, the quantitative comparisons and ablation results consistently verify that the proposed OLS outperforms existing Mamba-based scanning strategies for PV cell defect segmentation. By comprehensively capturing structural continuity across more directions, OLS contributes to handling the geometric complexity of PV cell defects, validating its role as a core innovation in PV-MambaSeg.

F. Failure Analysis

Despite the superior performance of PV-MambaSeg, there remain failure cases in specific PV cell defect scenarios. We analyze two representative cases in Fig. 9 to identify limitations and provide insights for future improvements. The first case involves missed segmentation with tiny, low-contrast cracks which account for less than 0.1% of the image pixels and exhibit minimal grayscale variation from the background. Though the multidirectional scanning of OLS has alleviated this issue by enhancing the capture of fine-grained structural continuity, it still struggles to extract sufficient discriminative defect features from these minute regions. The second case is missed segmentation of crack and interconnect in a noisy background, where mixed multitype defect features and background noise cause confusion. The proposed DWAG module has mitigated the interference by adaptively calibrating global and local feature weights, but it still cannot fully resolve the feature confusion, resulting in occasional missed segmentation. These challenges are common across existing PV cell defect segmentation methods, not exclusive to our approach.

These cases highlight that enhancing the model's sensitivity to tiny defects and robustness to feature confusion in noisy background are key directions for future work. Potential solutions include dedicated tiny-defect enhancement modules and noise-aware attention mechanisms to better distinguish defects from complex backgrounds.

V. CONCLUSION

This article presents PV-MambaSeg, which is the first Mamba-based network tailored for PV multiclass defect segmentation. Our model focuses on better modeling complex geometries and balancing accuracy with efficiency through two main components: the OLS strategy for comprehensive multidirectional feature capture and the GLVSS module for adaptive feature fusion.

Extensive experiments on two public PV cell defect segmentation datasets demonstrate the superiority of PV-MambaSeg. It achieves a leading mIoU while maintaining exceptional efficiency, outperforming both CNN-based and transformer-based methods in balancing accuracy and computational cost. Our future efforts will focus on enhancing the model's sensitivity to tiny defects as well as its robustness in highly noisy environments, further expanding its practical applicability.

ACKNOWLEDGMENT

The authors also thank the Big Data Computing Center of Southeast University for providing the facility support on the numerical calculations in this article.

REFERENCES

[1] C. Wang, H. Chen, and S. Zhao, "RERN: Rich edge features refinement detection network for polycrystalline solar cell defect segmentation," *IEEE Trans. Ind. Inform.*, vol. 20, no. 2, pp. 1408–1419, Feb. 2024.
[2] A. Mellit, C. Moussaoui, S. Pastore, and A. Massi Pavan, "An end-to-end prototype for fault diagnosis of solar photovoltaic modules using infrared thermographic images and lightweight deep CNNs," *IEEE Trans. Ind. Appl.*, vol. 62, no. 1, pp. 446–455, Jan. 2026.
[3] J. Fioresi et al., "Automated defect detection and localization in photovoltaic cells using semantic segmentation of electroluminescence images," *IEEE J. Photovolt.*, vol. 12, no. 1, pp. 53–61, Jan. 2022.
[4] W. Huang et al., "MRFP-TDG: A detection transformer with hybrid encoder and position-aware decoder for photovoltaic cell defect detection," *Inf. Sci.*, vol. 735, Apr. 2026, Art. no. 123092.
[5] F. M. Talaat, M. Salem, and W. M. Shaban, "AI-driven fault detection and classification in photovoltaic systems using deep learning techniques," *Sci Rep*, vol. 16, no. 1, Mar. 2026, Art. no. 37958.
[6] G. Chen, K. Zhan, H. Zhao, S. He, X. Fan, and L. Yu, "An efficient algorithm for photovoltaic defect detection with curvature-adaptive and multi-branch convolutions," *Sol. Energy*, vol. 310, May 2026, Art. no. 114401.

[7] Z. Lv et al., "AM-DETR: Photovoltaic cell surface defect detection network based on adaptive multi-scale detection transformer," *Measurement*, vol. 266, Mar. 2026, Art. no. 120508.

[8] G. Wu, Y. Zhang, L. Deng, J. Zhang, and T. Chai, "Cross-modal learning for anomaly detection in complex industrial process: Methodology and benchmark," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 35, no. 3, pp. 2632–2645, Mar. 2025.

[9] Z. Zhang, M. Cai, H. Wang, G. Wu, T. Chai, and X. Zhu, "CostFilter-AD: Enhancing anomaly detection through matching cost filtering," in *Proc. Int. Conf. Mach. Learn.*, vol. 267, 2025, pp. 74540–74564.

[10] S. Dong, J. Zhang, Y. Shen, and D. Gao, "Referring solar cell defect segmentation in electroluminescence images," *IEEE Trans. Ind. Inform.*, vol. 21, no. 9, pp. 6788–6798, Sep. 2025.

[11] Z. Mahboob, M. A. Khan, E. Lodhi, T. Nawaz, and U. S. Khan, "Using segformer for effective semantic cell segmentation for fault detection in photovoltaic arrays," *IEEE J. Photovolt.*, vol. 15, no. 2, pp. 320–331, Mar. 2025.

[12] X. Qian, J. Li, J. Cao, Y. Wu, and W. Wang, "Micro-cracks detection of solar cells surface via combining short-term and long-term deep features," *Neural Netw.*, vol. 127, pp. 132–140, Jul. 2020.

[13] S. Zhao, H. Chen, C. Wang, Y. Zhou, and Z. Zhang, "RGR-Net: Refined graph reasoning network for multi-height hotspot defect detection in photovoltaic farms," *Expert Syst. Appl.*, vol. 245, Jul. 2024, Art. no. 123034.

[14] B. Su, H. Chen, K. Liu, and W. Liu, "RCAG-Net: Residual channelwise attention gate network for hot spot defect detection of photovoltaic farms," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–14, 2021.

[15] E. Sovetkin, E. J. Achterberg, T. Weber, and B. E. Pieters, "Encoder-decoder semantic segmentation models for electroluminescence images of thin-film photovoltaic modules," *IEEE J. Photovolt.*, vol. 11, no. 2, pp. 444–452, Mar. 2021.

[16] W. Tang, Q. Yang, X. Hu, and W. Yan, "Convolution neural network based polycrystalline silicon photovoltaic cell linear defect diagnosis using electroluminescence images," *Expert Syst. Appl.*, vol. 202, Sep. 2022, Art. no. 117087.

[17] P. Zhou, R. Wang, C. Wang, H. Chen, and K. Liu, "SIIF: Semantic information interactive fusion network for photovoltaic defect segmentation," *Appl. Energy*, vol. 371, Oct. 2024, Art. no. 123643.

[18] L. Zhu, B. Liao, Q. Zhang, X. Wang, W. Liu, and X. Wang, "Vision mamba: Efficient visual representation learning with bidirectional state space model," in *Proc. Int. Conf. Mach. Learn.*, 2024, pp. 62429–62442.

[19] Y. Liu et al., "VMamba: Visual state space model," *Adv. Neural Inf. Process. Syst.*, vol. 37, pp. 103031–103063, 2024.

[20] C. Yang et al., "PlainMamba: Improving non-hierarchical mamba in visual recognition," in *Proc. Brit. Mach. Vis. Conf.*, 2024, Art. no. 0755.

[21] Y. Fu, M. Lou, and Y. Yu, "SegMAN: Omni-scale context modeling with state space models and local attention for semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2025, pp. 19077–19087.

[22] J. Bian et al., "Locally aware visual state space for small defect segmentation in complex component images," *IEEE Trans. Ind. Inf.*, vol. 21, no. 8, pp. 5965–5976, Aug. 2025.

[23] H. Du et al., "Automated intelligent detection of defects on bridge piers using a climbing operation robot and vision mamba," *IEEE Trans. Instrum. Meas.*, vol. 74, pp. 1–18, 2025.

[24] Y.-W. Huang and W. Wang, "PVMNet: A navel orange defect detection algorithm based on mamba structure," *Expert Syst. Appl.*, vol. 300, Mar. 2026, Art. no. 130412.

[25] N. Bao, J. Lin, Y. Fan, R. Bao, and A. Simeone, "FabricMamba: A fabric surface defect detection system based on large kernel attention and visual state space," *Eng. Appl. Artif. Intell.*, vol. 162, 2025, Art. no. 112558.

[26] D. Zhao et al., "DC-Mamba for surface micro defect classification on large aperture optics with multi axis attention," *Sci Rep*, vol. 15, no. 1, Oct. 2025, Art. no. 37958.

[27] X. Wang, L. Bao, X. Zhou, L. Xia, and X. Xu, "GLNet: Global-local fusion network for strip steel surface defects detection," *IEEE Signal Process. Lett.*, vol. 32, pp. 1256–1260, 2025.

[28] G. Zhao, L. Tan, M. He, and Q. Wu, "HFMM-Net: A hybrid fusion mamba network for efficient multimodal industrial defect detection," *Information*, vol. 16, no. 12, pp. 1–16, Nov. 2025.

[29] M. Li, P. Ye, Z. Li, S. Cui, P. Zhu, and J. Liu, "An efficient CNN-Mamba hybrid network for fabric defect segmentation," *Results Eng.*, vol. 28, Dec. 2025, Art. no. 107493.

[30] A. Gu and T. Dao, "Mamba: Linear-time sequence modeling with selective state spaces," in *Proc. Conf. Comput. Lang. Model.*, 2024, pp. 1–31.

[31] S. Zhao, H. Chen, X. Zhang, P. Xiao, L. Bai, and W. Ouyang, "RS-Mamba for large remote sensing image dense prediction," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–14, 2024.

[32] H. Liu, C. Jia, F. Shi, X. Cheng, and S. Chen, "SCSegamba: Lightweight structure-aware vision mamba for crack segmentation in structures," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2025, pp. 29406–29416.

[33] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1–9.

[34] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "SegFormer: Simple and efficient design for semantic segmentation with transformers," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 12077–12090.

[35] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Med. Image Comput. Comput. Assist. Intervent.*, vol. 9351, 2015.

[36] M. Fan et al., "Rethinking bisenet for real-time semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 9716–9725.

[37] H. Pan, Y. Hong, W. Sun, and Y. Jia, "Deep dual-resolution networks for real-time and accurate semantic segmentation of traffic scenes," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 3, pp. 3448–3460, Mar. 2023.

[38] J. Xu, Z. Xiong, and S. P. Bhattacharyya, "PIDNet: A real-time semantic segmentation network inspired by PID controllers," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 19529–19539.

[39] Q. Wan, Z. Huang, J. Lu, G. Yu, and L. Zhang, "Seaformer: Squeeze-enhanced axial transformer for mobile semantic segmentation," in *Proc. Int. Conf. Learn. Represent.*, 2023, pp. 1–16.

[40] Z. Ni, X. Chen, Y. Zhai, Y. Tang, and Y. Wang, "Context-guided spatial feature reconstruction for efficient semantic segmentation," in *Proc. Eur. Conf.Comput. Vis.*, 2024, pp. 239–255.

[41] Z. Xu, D. Wu, C. Yu, X. Chu, N. Sang, and C. Gao, "SCTNet: Single-branch CNN with transformer semantic information for real-time segmentation," in *Proc. AAAI Conf. Artif. Intell.*, 2024, pp. 6378–6386.

[42] G. Yang, Y. Wang, D. Shi, and Y. Wang, "Golden cudgel network for real-time semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2025, pp. 25367–25376.

[43] J. Pan, D. Zeng, Z. Wu, and S. Xie, "BiDNet: A real-time semantic segmentation network with antifeature interference and detail recovery for industrial defects," *IEEE Trans. Instrum. Meas.*, vol. 74, 2025, Art. no. 5017916.

[44] H. Zhu et al., "DS2A-Former: Battery surface defect segmentation via dual-stream spatial attention transformer network," *IEEE Trans. Instrum. Meas.*, vol. 74, 2025, Art. no. 2530917.

[45] G. Li, X. Zhou, and H. Li, "Ordered cross-scale interaction network for no-service rail surface defect segmentation," *IEEE Trans. Instrum. Meas.*, vol. 74, 2025, Art. no. 5033210.

[46] Y. Zhang, J. Cheng, Z. Xia, W. Yin, S. Liu, and N. Tashi, "Multipatch perception and knowledge-guided network for semantic segmentation of large-scale remote sensing image," *IEEE Trans. Geosci. Remote Sens.*, vol. 64, 2026, Art. no. 5606219.

[47] W. Zhang et al., "Center-guided classifier for semantic segmentation of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 64, 2026, Art. no. 4404916.

[48] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.

[49] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit*. 2017, pp. 936–944.

[50] T. Xiao, Y. Liu, B. Zhou, Y. Jiang, and J. Sun, "Unified perceptual parsing for scene understanding," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 432–448.

[51] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2011–2023, Aug. 2020.

[52] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, vol. 11211, 2018, pp. 3–19.

[53] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 13708–13717.

[54] X. Li, W. Wang, X. Hu, and J. Yang, "Selective kernel networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 510–519.

[55] S. Deitsch et al., "Automatic classification of defective photovoltaic module cells in electroluminescence images," *Sol. Energy*, vol. 185, pp. 455–468, Jun. 2019, doi: 10.1016/j.solener.2019.02.067.