Applied Energy 419 (2026) 128093

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apen

DSDINet: Deep semantic decoupling and integration photovoltaic solar cell defect segmentation network

Ziai Zhou [a], Chaoyang Song [a], Shixiong Fang [a], Huimin Lu [b], Jinxia Zhang [a,b,*]

[a] The Key Laboratory of Measurement and Control of CSE, Ministry of Education, School of Automation, Southeast University, Nanjing 210096, China
[b] Advanced Ocean Institute of Southeast University, Nantong 226010, China

HIGHLIGHTS

- A deep semantic decoupling network is proposed for complex PV solar cell defect segmentation.
- A semantic decoupling and aggregation module enhances cross-scale feature fusion.
- A global-local attention module strengthens semantic consistency and detail preservation.
- Extensive experiments demonstrate competitive performance and robustness on public datasets.

ARTICLE INFO

Keywords:
Photovoltaic solar cells
Defect segmentation
Semantic decoupling
Global and local attention
Deep learning

ABSTRACT

Defects in photovoltaic (PV) solar cells can significantly degrade photoelectric conversion efficiency and shorten service life. Solar cell defect segmentation enables accurate localization of defects, thereby supporting the intelligent operation and maintenance of photovoltaic power plants. However, solar cell defect segmentation faces two major challenges. On the one hand, multiscale low-contrast defect information is difficult to extract from the foreground features due to distracting information such as flocculent regions and busbars in the background. On the other hand, high-level semantic information is prone to being lost in the feature fusion process, leading to the missed detection of some edge-located defects and tiny defects. To address these two challenges, a deep semantic decoupling and integration defect segmentation network (DSDINet) is introduced which can effectively extract rich multiscale information from the foreground features and improve the representation capability of high-level features. Specifically, we design a Semantic Information Decoupling and Aggregation (SIDA) module to decouple and process the foreground and background of enriched fine-grained low-level feature maps, and then fuse them with high-level features. For defects of different scales, a Multiscale Feature Exploration (MFE) module is embedded as a built-in submodule of SIDA to mine discriminative features from decoupled foreground defect regions. In addition, to enhance the features of edge-located defects and tiny defects, we propose a Global and Local Attention (GLA) module that simultaneously models global semantic dependencies and local fine-grained details. Experimental results on the public polycrystalline silicon defect segmentation dataset PSCDE demonstrate that DSDINet achieves state-of-the-art performance, demonstrating its effectiveness for PV solar cell defect segmentation. Source code is available at: https://github.com/Taraseu/PV-Defect-Segmentation-DSDINet.

1. Introduction

Photovoltaic power generation has undergone remarkable development over the past decades. By the end of 2023, the total installed capacity of PV systems reached 1.6 TW, accounting for 6.2% of global electricity production [1]. During the manufacturing, transportation, installation and operation of PV solar cells, various defects inevitably occur owing to improper handling, thermal stress, or extreme weather conditions [2]. These defects can significantly reduce photoelectric conversion efficiency and shorten the service life, leading to substantial waste of resources and economic losses [3,4]. Consequently, to ensure the high-standard manufacturing and operation & maintenance (O&M) of PV solar cells, it is essential to precisely segment defects.

Defect segmentation can provide precise details about defects, covering their shape, size, and exact position [8]. The analysis of defect

* Corresponding author at: School of Automation, Southeast University, Nanjing, 210096, Jiangsu, China.
Email address: jinxiazhang@seu.edu.cn (J. Zhang).

https://doi.org/10.1016/j.apenergy.2026.128093
Received 6 February 2026; Received in revised form 8 April 2026; Accepted 20 May 2026
Available online 28 May 2026

information by PV manufacturers allows for the identification of defect sources, leading to the optimization of production processes, the reduction of manufacturing costs, and the enhancement of product quality [9]. The O&M departments of PV power plants can quickly identify and replace damaged PV solar cells, ensuring their stable long-term operation. Deep learning techniques are being employed to steer the PV industry in a more intelligent direction. These advances not only improve the reliability of energy systems but also facilitate the worldwide use and spread of renewable energy, thus making a significant contribution to environmental protection [10].

Over the past few years, deep learning-based semantic segmentation methods have demonstrated significant potential in identifying defects in PV solar cells, which have continuously improved the performance of defect segmentation [11,12]. Despite significant breakthroughs, accurate segmentation remains a challenging task due to inherent technical gaps that have not been fully addressed. Conventional semantic segmentation models, originally designed for natural scenes, struggle to adapt to the unique characteristics of PV solar cell images.

As shown in Fig. 1(a), foreground objects in natural scenes usually exhibit high contrast against backgrounds and are often distributed in large patches in the central region of the image. However, in PV solar cell images, low-contrast defects are distributed against visually complex backgrounds, making them difficult to distinguish. Moreover, defects of various sizes can be distributed anywhere in the image. Therefore, existing semantic segmentation methods generally perform well in segmenting large defects in the central region of PV solar cell images, but struggle with segmenting low-contrast multiscale defects that are typical of real-world PV scenarios. As illustrated in Fig. 1(b), even state-of-the-art semantic segmentation approaches including STDC [5], DDRNet [6], and PIDNet [7] still suffer from discontinuous and missed detection, primarily due to their inability to handle background interference and insufficient multiscale feature extraction. In contrast, the proposed DSDINet achieves precise and complete segmentation of PV solar cell defects, demonstrating its superior adaptability to the unique challenges of PV solar cell defect segmentation compared with natural-scene-based methods.

There are two core challenges of generic natural scene segmentation methods behind these issues: First, these models lack an explicit foreground–background decoupling mechanism, preventing them from effectively extracting rich multiscale information from foreground

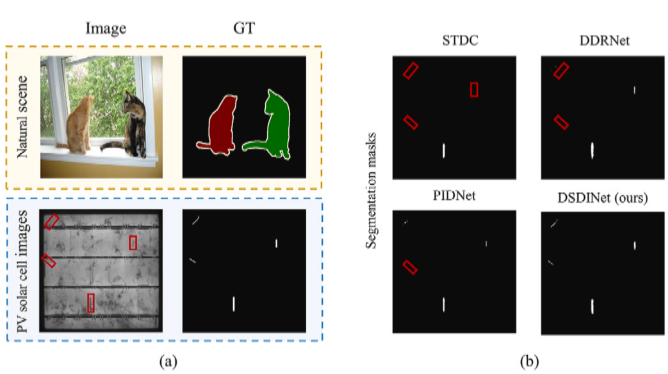

Fig. 1. Key challenges of PV solar cell defect segmentation and superiority of DSDINet: (a) comparison of natural scene images and PV solar cell images. Natural scene images have high-contrast, centrally distributed large foreground patches. PV solar cell images feature low-contrast, multiscale defects randomly distributed across complex backgrounds, increasing segmentation difficulty. (b) Predictions of state-of-the-art semantic segmentation methods and our DSDINet. STDC [5], DDRNet [6], and PIDNet [7] suffer from discontinuous segmentation and missed detection of tiny and edge defects (marked by red bounding boxes) due to background interference and insufficient multiscale feature extraction. Our DSDINet achieves precise, complete PV solar cell defect segmentation.

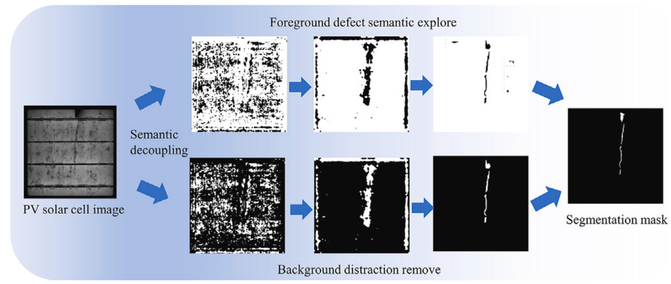

Fig. 2. Schematic diagram of the semantic decoupling and refinement pipeline in DSDINet. The input PV solar cell image is first decoupled into foreground defect features and background interference features via semantic decoupling. Then, foreground defect semantic exploration and background distraction removal are performed separately. Finally, the processed features are fused to generate the precise segmentation mask, which effectively addresses background interference and semantic information loss.

features and leaving them susceptible to background clutter, such as flocculent regions and busbars. Second, their generic feature fusion strategies ignore the semantic discrepancy between low-level details and high-level semantics, causing high-level semantic information to be easily lost during fusion, leading to failures in recognizing edge-located and tiny defects.

To address these two challenges, DSDINet is introduced which can effectively extract multiscale low-contrast defect information from the foreground features and enhance the representation ability of deep features. As illustrated in Fig. 2, the core of DSDINet lies in a semantic decoupling and refinement pipeline that explicitly separates foreground defect features from background interference and refines them in dedicated branches. Specifically, to address the issue of poor foreground distinguishability against complex backgrounds, we introduce the SIDA module. SIDA module decouples foreground and background features by capturing richer low-level details and integrates them into high-level features to enable effective fusion of multi-level semantic information. For defects of different scales, an MFE module is designed to explore the decoupled foreground features. To prevent missed detection of edge-located defects and tiny defects, we propose a GLA module that applies attention to both global and local features to refine the high-level features. Overall, the main contributions of this paper can be outlined as follows:

- A novel end-to-end deep semantic decoupling and integration network is proposed exclusively for PV solar cell defect segmentation. By constructing a systematic decoupling-fusion-enhancement pipeline, DSDINet effectively extracts multiscale low-contrast defect features and strengthens the discriminative representation of deep features, addressing key limitations in existing methods.
- A Semantic Information Decoupling and Aggregation module is presented to explicitly separate foreground defects from background clutter. Assisted by the designed Multiscale Feature Exploration module, it suppresses irrelevant background interference and realizes adaptive multiscale feature extraction and fusion.
- A Global and Local Attention module is developed to simultaneously capture global semantic dependencies and local fine-grained details via combined position and channel attention. It effectively alleviates high-level semantic loss during fusion, thus reducing the missed detection of edge and tiny defects.
- Experimental results on the PV solar cell defect segmentation dataset PSCDE [13] demonstrate that DSDINet outperforms existing state-of-the-art methods. Specifically, DSDINet achieves a 6.36% improvement in F1 score and a 10.45% improvement in defect IoU compared to the second-best method.

2. Related work

This work is related to deep learning methods for PV solar cell defect segmentation, semantic decoupling methods, and attention mechanisms. These related works are briefly presented in sequence.

2.1. Deep learning methods for PV solar cell defect segmentation

Semantic segmentation networks have been widely applied in the segmentation of PV solar cell defects [14]. Fioresi et al. [15] employed the DeepLabv3 [16] model for defect segmentation in monocrystalline and polycrystalline silicon cells. Chen and Rahman et al. [17,18] modified U-Net and pretrained backbones to extract features from electroluminescence images. Han et al. [19] applied a region proposal network in an improved polycrystalline silicon PV solar cell defect segmentation network, thereby improving the accuracy of segmentation. Zhou et al. [20] proposed a segmentation network through the combination of semantic information interaction and multi-level information fusion. Wang et al. [13] used edge detection to better localize defect edges. Zhao et al. [21] designed an improved graph reasoning network to learn hotspot features by the establishment of global long-range dependencies and fully utilizing complementary information across network layers.

Despite these advances, existing PV solar cell defect segmentation works suffer from two core limitations that constitute a critical, unaddressed knowledge gap. On one hand, most prior studies directly retrain or slightly tweak generic semantic segmentation models which are originally designed for high-contrast, centrally distributed natural scene objects before applying them to PV solar cell defect segmentation datasets. They struggle to account for the unique challenges of PV solar cell images such as low-contrast defects scattered randomly across complex backgrounds. This leads to severe background interference and incomplete extraction of foreground defect features. On the other hand, some networks specifically designed for PV solar cell defect segmentation struggle to fully account for the unique characteristics of PV solar cell images in their core architectures, and their code or models have not been made publicly available. This limits reproducibility and practical adoption in the PV industry, creating a critical need for an open, well-validated PV-tailored solution.

2.2. Semantic decoupling methods

Semantic decoupling cues have found applications in multiple visual tasks, including salient object detection [22], camouflage object detection [23], shadow detection [24], and medical image segmentation [25]. Wu et al. [22] proposed a specific network to learn the regions in the foreground features that affect saliency prediction and to erase them from the input image. Zheng et al. [24] established an end-to-end architecture to explicitly learn false positives in the foreground and false negatives in the background, while Mei et al. [23] proposed an implicit strategy for learning this information. Both methods are designed for natural scene tasks and lack adaptability to the low-contrast, multiscale characteristics of PV solar cell defects. Lei et al. [25] decoupled the highest level features into foreground, background, and uncertain regions, gradually reducing the uncertainty during training, but their decoupling strategy fails to explore multiscale features and ignores fine-grained low-level details that are critical for PV solar cell defect segmentation.

Distinct from these semantic decoupling methods, which either target natural scene tasks or neglect low-level detail preservation, our SIDA module is specifically designed for PV solar cell defect segmentation. It decouples foreground and background features by capturing richer low-level details and integrating them with high-level features to enable effective multiscale feature fusion. Additionally, the embedded MFE module adopts a residual-connected dilated convolution branch design tailored to the wide size variation of PV solar cell defects. This design enables adaptive capture of multiscale defect information without losing low-contrast defect semantics which is a critical requirement for PV solar cell images that generic methods struggle to address.

2.3. Attention mechanisms

Attention mechanisms have been widely applied in a variety of visual tasks because they help the deep neural network focus on informative areas while suppressing irrelevant information [26]. Channel attention enhances the representation of important channels by learning the dependencies among channels. SENet [27] is a pioneering work in channel attention, which adaptively learns the importance of each channel, thus enhancing significant features and reducing the emphasis on less relevant ones. Position attention enhances the representation of important spatial locations by modeling dependencies among spatial positions. NLNet [28] is the first to introduce position attention in neural networks, modeling pairwise relationships between pixels on the feature map. CT-HiffNet [29] designs a contour-texture attention module that generates spatial attention weights from multiscale features to highlight boundary- and texture-relevant regions of cropland parcels. S2-IFNet [30] proposes a linear spatial attention module that generates attention weights over spatial positions to highlight forest land areas and suppress irrelevant background.

Multi-attention fusion mechanisms integrate different types of attention mechanisms to fully utilize information from various dimensions. CBAM [31] is the first module to integrate channel and position attention in a sequential manner, thereby significantly enhancing feature representation ability. DANet [32] models the relationships between pixels and channels using parallel channel and position attention. JAM-R-CNN [33] incorporates CBAM [31] into FPN to enhance multiscale features of terraced fields, addressing their narrow and elongated shapes in high-resolution imagery. FPENet [34] employs a mutual embedding upsample module that facilitates mutual guidance between channel-wise semantic features and spatial detail features. TSANet [35] introduces a Spatial-Spectral-Temporal Attention module that captures feature correlations along both the spatial-temporal and spectral-temporal dimensions to enhance agricultural field representation from satellite image time series.

Although these methods model long-range dependencies in both channel and position dimensions to improve network performance across various visual tasks, they do not explicitly balance global context modeling and local detail refinement, which is a critical requirement for PV solar cell defect segmentation. Specifically, they either overemphasize global features or focus on local features. To address this gap, we propose a GLA module that captures global and local contexts simultaneously in both position and channel dimensions. This module directs the network to focus on target defect regions, strengthens the representation ability of high-level semantic features, and effectively mitigates the loss of semantic information for tiny and edge defects during feature fusion.

3. Methodology

This section presents the proposed network for PV solar cell defect segmentation. The existing PV solar cell defect segmentation methods cannot extract rich multiscale information from the foreground features during feature interaction, thereby leading to discontinuous and imprecise segmentation results. In addition, high-level semantic information is easily lost during feature fusion, resulting in the failure to recognize some edge-located defects and tiny defects. To address these two challenges, we present a deep semantic decoupling and integration defect segmentation network, which can effectively extract rich multiscale information from the foreground features and enhance the representation capability of high-level features. Three modules, i.e., the SIDA module, the MFE module and the GLA module are integrated into this network to ensure accurate PV solar cell defect segmentation.

3.1. Overall network structure

We propose a DSDINet based on the encoder–decoder architecture for PV solar cell defect segmentation to improve the accuracy of

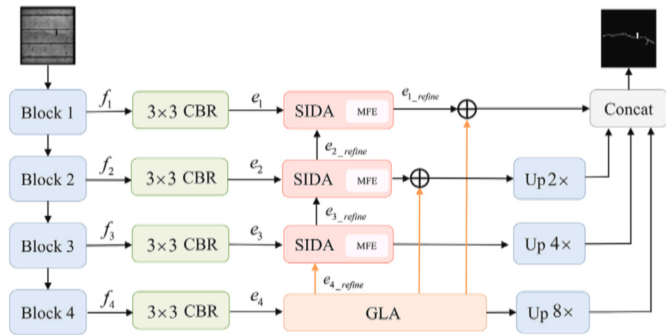

Fig. 3. Overview of the proposed DSDINet. DSDINet uses a MiT backbone to extract multi-level features encoded as e_i. The proposed GLA module refines deep semantic feature map e_4 by modeling long-range channel and position dependencies to enhance edge and tiny defect representation. The proposed SIDA module, with the MFE module embedded as its built-in submodule, decouples neighbor-level features and conducts adaptive multiscale contextual reasoning on the decoupled foreground defect features, then fuses the processed multiscale features with GLA-refined high-level semantic features. The refined features from all levels are aggregated for the final segmentation mask.

segmentation, and a detailed view of the network structure is illustrated in Fig. 3. Specifically, we employ the Mix Transformer (MiT) [36] as the backbone network. Compared with traditional convolutional neural networks (CNNs), MiT combines Transformer's global modeling ability with CNN's local feature extraction ability, excelling in capturing long-range dependencies and global context. It can provide low-resolution coarse features and high-resolution fine-grained features. Feature maps at different levels are downsampled by factors of 4, 8, 16, and 32 relative to the input image, denoted as $f_i (1 \le i \le 4)$. Each feature map is processed by a convolutional layer for encoding, resulting in $e_i (1 \le i \le 4)$.

Subsequently, the feature map e_4 which contains the deepest semantic information is fed to the GLA module for establishment of long-range dependencies in the channel and position dimensions of both local and global contexts. The GLA module enhances the representation ability of high-level semantic features, thereby reducing missed detection of edge-located defects and tiny defects. The refined high-level feature map e_{4_refine} generated by GLA serves as the key high-level semantic input into the subsequent SIDA module, forming the foundation for cross-scale feature fusion. The SIDA module then operates on neighbor-level feature maps to decouple and refine fine-grained low-level information. The MFE module is integrated as a built-in submodule within SIDA, dedicated exclusively to performing adaptive multiscale contextual reasoning on the decoupled foreground defect features, thus ensuring the precise capture of defect information across varying scales. After completing this multiscale enhancement, the SIDA module fuses the processed features with the GLA-refined high-level semantic representations, achieving a robust integration of local details and global context. The resulting feature maps at different levels are concatenated and fused to produce the final feature map. Subsequently, a multilayer perceptron takes the final feature map to predict the segmentation mask.

3.2. Semantic information decoupling aggregation module

The SIDA module aims to utilize the contrastive features of decoupled foreground and background to guide the fusion of adjacent feature layers. This method helps better distinguish the defects from complex background. Specifically, the SIDA module first decouples the low-level features into foreground and background features. As a built-in submodule embedded within the SIDA module, the MFE module then performs multiscale contextual reasoning directly on the decoupled foreground features. Finally, the processed foreground features and the background features are aggregated and then fused with high-level feature maps.

As demonstrated in Fig. 4(a), high-level feature map e_{i+1_refine} is upsampled and then subsequently normalized using a sigmoid layer. Low-level feature map e_i is then multiplied by the normalized map and its inverse to generate foreground feature map e_{if} and background feature map e_{ib}, respectively. Subsequently, the foreground feature map e_{if} is fed into the MFE module for multiscale contextual reasoning. As indicated in Fig. 4(b), MFE module is made up of four branches. The input feature map e_{if} is processed through four dilated convolutions with a dilation rate of r_j, capturing multiscale features $e_{if}(j)$. We set r_j, $j \in \{1, 2, 3, 4\}$ to 1, 2, 4, and 8, respectively. The output of the j^{th},

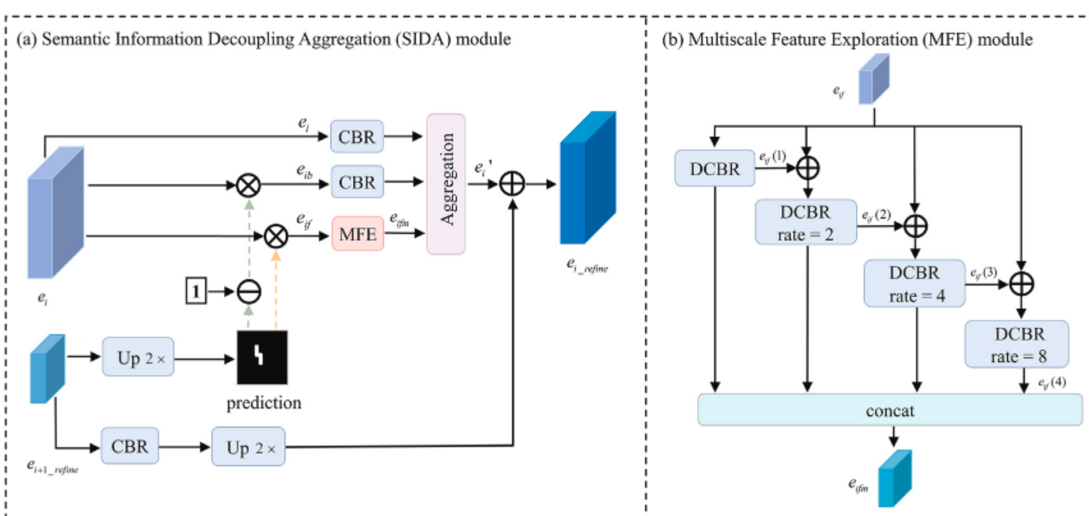

Fig. 4. The structure of SIDA and its building module MFE. (a) SIDA splits low-level feature e_i into foreground e_{if} and background e_{ib} via element-wise multiplication and its inverse using upsampled and sigmoid-normalized high-level feature map e_{i+1_refine}. e_{if} is processed by the proposed MFE module to generate multiscale-enhanced foreground feature e_{ifm}. Finally, e_{ifm}, refined e_{ib} and e_i are aggregated with learnable weights γ, η and δ and fused with upsampled e_{i+1_refine} to output refined feature map e_{i_refine}. (b) MFE processes e_{if} by four dilated convolution branches with residual connections, and concatenates these features to generate multiscale-enhanced foreground feature e_{ifm}.

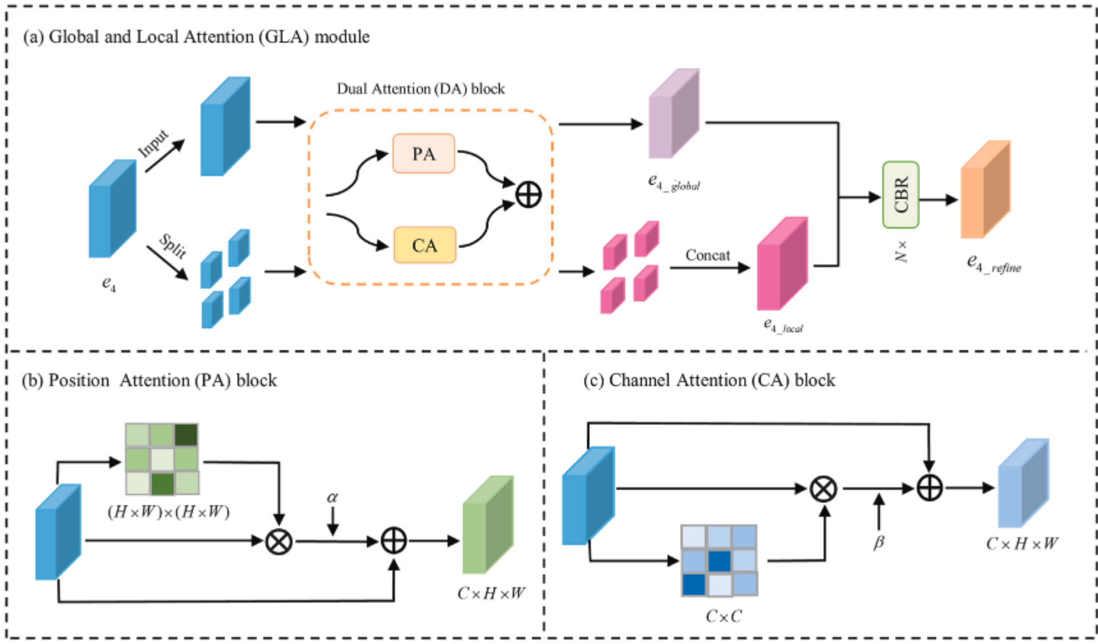

Fig. 5. The structure of GLA its building blocks PA and CA. (a) deep feature e_4 is directly processed by DA to generate global attention feature e_{4_global} while also split into four equal-sized local maps $e_4(k)$ that are DA-processed and concatenated to form e_{4_local}. Finally, e_{4_global} and e_{4_local} are concatenated and refined via NCBR to output enhanced feature e_{4_refine}, boosting tiny and edge-located defect recognition. (b) PA block enhances defect location awareness. (c) CA block emphasizes defect-informative channels.

$j \in \{1,2,3\}$ MFE branch will be fed to the $(j+1)^{th}$ branch for further processing:

$$e_{if}(j) = \begin{cases} DCBR(e_{if}) & \text{if } j = 1 \\ DCBR(e_{if} + e_{if}(j-1)) & \text{if } 1 < j \le 4 \end{cases} \tag{1}$$

where DCBR is the combination of dilated convolution, batch normalization (BN) and ReLU. The foreground feature map e_{ifm} that integrates information from different scales is obtained by concatenating and fusing output features from all four branches:

$$e_{ifm} = CBR\left(concat(e_{if}(j))\right) \tag{2}$$

where CBR represents the combination of convolution, BN and ReLU. This architecture facilitates the network's ability to extract multiscale information from the foreground, thereby enhancing its capacity to segment defects that exhibit large-scale variations. Then, the obtained multiscale enhanced foreground feature map e_{ifm} is integrated with background feature map and image feature map:

$$e_{i'} = CBR(concat(\gamma e_{ifm},\ \eta CBR(e_{ib}),\ \delta CBR(e_i))) \tag{3}$$

where γ, η and δ are the learnable parameters with an initial setting of 1. By integrating adjacent image feature maps, we substantially improve the recognition rate of multiscale low-contrast defects to reduce discontinuous and missed detection. The output refined feature map e_{i_refine} can be calculated as follows:

$$e_{i\_refine} = e_{i'} + U(CBR(e_{i+1\_refine})) \tag{4}$$

where U is the bilinear upsampling.

The SIDA module effectively integrates multiscale enhanced foreground information into each output branch. It simultaneously captures rich defect information from adjacent feature maps. As a result, our segmentation model can learn more defect location information and more robust semantics.

3.3. Global and local attention module

The GLA module refines the highest-level feature map e_4 by simultaneously capturing global and local semantic dependencies, which significantly enhances the representation of tiny and edge-located defects in deep semantic features. Its output e_{4_refine} acts as the high-level semantic guidance for subsequent SIDA modules during cross-scale feature fusion. This upstream-downstream cooperation ensures that the high-level semantic information optimized by GLA can be effectively transmitted to the low-level feature fusion process, and the low-level multiscale features enhanced by SIDA can be guided by high-quality semantic information, thus avoiding the loss of semantic information in cross-scale fusion.

As shown in Fig. 5, the GLA module is composed of a local branch and a global branch. Features from each branch pass through a Dual Attention (DA) block to capture long-range dependencies in channel and position dimensions. Then, the final feature map is obtained by integrating local and global features through a series of convolutions. Specifically, given deep feature map $e_4 \in \mathbb{R}^{C\times H\times W}$, with C, H, and W representing channels, height, and width, respectively. For global branch, e_4 is fed into and processed through the DA block to generate the global attention feature map e_{4_global}:

$$e_{4\_global} = DA(e_4). \tag{5}$$

For the local branch, e_4 is split into four equal-sized local maps $e_4(k) \in \mathbb{R}^{C\times\frac{H}{2}\times\frac{W}{2}}$, where $k \in \{1,2,3,4\}$:

$$e_4(k) = split_h(split_w(e_4)) \tag{6}$$

where $split_h$ and $split_w$ denote the operations of splitting the feature map into two parts equally along the height and width axes, respectively. $e_4(k)$ is processed through the DA block and then concatenated to obtain the local attention feature map e_{4_local}:

$$e_{4\_local} = concat(DA(e_4(k))). \tag{7}$$

The output of the GLA module can be calculated as follows:

$$e_{4\_refine} = NCBR(concat(e_{4\_global}, e_{4\_local})). \tag{8}$$

where NCBR represents the combination of multiple convolutions, BN and ReLU.

DA block consists of a position attention branch and a channel attention branch. As shown in Fig. 5(b) and (c), Position Attention (PA) block strengthens the model's ability to identify and respond to important defect location features while Channel Attention (CA) assigns greater significance to channels that contain more abundant defect information.

The feature map output by the GLA module contains both global and local enhancement information of the highest-level features. Integrating this information into each branch of the network can effectively improve the identification rate of tiny defects and edge-located defects, thus enhancing the accuracy of defect segmentation.

4. Experiments

Under the same experimental conditions, the proposed DSDINet and other state-of-the-art semantic segmentation approaches were trained and tested on the public PV solar cell defect segmentation datasets PSCDE[13] and UCF-EL [15]. Common metrics were used to evaluate the defect segmentation performance. In addition, we also conducted an ablation study to demonstrate the effectiveness of the proposed SIDA module and GLA module.

4.1. Experimental setup

Dataset

PSCDE is a high-quality dataset for PV solar cell defect segmentation. This dataset comprises 700 challenging defect images, featuring various types of defects such as multiscale defects, dense tiny defects, occluded defects and combination defects [13].

To further enhance the model's generalization, this work also incorporates the UCF-EL dataset [15]. This publicly available benchmark provides a substantial collection of 17,064 real-world images of PV solar cells. The dataset is comprehensively annotated for defect segmentation, with each image paired with a pixel-wise mask labeling background, defect-free cell regions, and defects.

Evaluation metrics

Since defective pixels are far fewer than non-defective pixels, the dataset is class-imbalanced. Following previous works, we use four widely adopted metrics for evaluating our method on class-imbalanced image segmentation data, i.e., Precision, Recall, F1 score, and Intersection over Union (IoU). Detailed definitions of the metrics are as follows:

$$\begin{aligned} \text{Precision} &= \frac{TP}{TP+FP} \\ \text{Recall} &= \frac{TP}{TP+FN} \\ \text{F1-score} &= 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \\ \text{IoU} &= \frac{TP}{TP+FP+FN} \end{aligned} \tag{9}$$

where TP represents true positives, FP represents false positives, and FN represents false negatives. Due to the significant variation in the number of defective pixels across different images, the metrics for each image are calculated individually. We aggregate the metrics computed for each image and compute the mean to derive the performance metrics throughout the dataset.

In addition to the above segmentation evaluation metrics, we also introduce four computational efficiency metrics to verify the engineering application of the proposed model, including the number of parameters (Params), computational complexity (GFLOPs), model size (Size), and inference speed (FPS). Params refers to the total number of trainable parameters in the network, which reflects the memory occupation during training and application. GFLOPs quantifies the floating-point operations required for forward inference, representing the computational burden of the model. Size denotes the actual storage space occupied by the trained model file, which is critical for resource-constrained industrial devices. FPS measures the number of frames processed per second during inference, representing the real-time processing capability of the model.

Implementation details

We implemented DSDINet with PyTorch framework. We configured the batch size at 4 and the initial learning rate at 6×10^{-5}. The image resolution is 512×512. We applied random horizontal and vertical flipping for data augmentation. We initialized the backbone parameters with the Mix Transformer [36] pre-trained on ImageNet.

4.2. Comparison with state-of-the-art semantic segmentation methods

To validate the effectiveness of DSDINet, we compared it with 12 semantic segmentation networks, including GCNet [37], ISANet [38], K-Net [39], STDC [5], SegFormer [36], DDRNet [6], PIDNet [7], Seaformer [40], CGRSeg [41], SCTNet [42], SegMAN [43], and GCNet [44]. The selection of these comparison methods is guided by three core criteria to ensure comprehensive, relevant, and fair evaluation.

First, we cover both classic and cutting-edge works. Methods like SegFormer, STDC, and PIDNet are widely adopted for their proven effectiveness in complex scenes, while recent advances such as SegMAN and GCNet represent the state-of-the-art in semantic segmentation, enabling us to verify DSDINet's competitiveness against the latest generic methods. Second, all selected methods are relevant to PV solar cell defect segmentation scenarios. Either they have been previously applied to PV solar cell defect tasks (e.g., SegFormer explicitly employed in prior PV solar cell defect segmentation work) or designed for complex scenes (e.g., PIDNet for traffic scenes), ensuring they serve as meaningful rather than trivial baselines. Third, we prioritize reproducibility and open-source availability. Given the lack of public code for some PV-specific methods, we only selected models with officially released codes.

To ensure fair comparison, prediction maps for all the above networks were obtained by retraining models with their official open-source codes. Additionally, the same evaluation code was used for all prediction maps.

Table 1 presents the comparative evaluation of DSDINet against state-of-the-art methods on PV solar cell defect segmentation datasets: PSCDE and UCF-EL. The results clearly show that DSDINet achieves the best performance on both datasets. Specifically, compared with PIDNet [7], our method shows an improvement of 8.7% in F1 score and 11.48% in IoU on PSCDE.

Performance on the UCF-EL is particularly noteworthy, as this dataset presents different characteristics and challenges. DSDINet achieves an F1 score of 0.6328 and IoU of 0.5087, outperforming the second-best method by significant margins of 6.74% and 11.12% respectively. It is worth noting that the performance gap between DSDINet and other methods is even more pronounced on UCF-EL than on PSCDE, suggesting that our method possesses stronger generalization ability to handle diverse defect patterns. This cross-dataset superiority underscores the practical value of DSDINet for real-world industrial applications.

Moreover, to demonstrate the engineering practicability of DSDINet for real industrial inspection scenarios, we further report its computational efficiency including parameters, GFLOPs, model size and inference speed. As shown in Table 1, DSDINet achieves the smallest model size and the lowest number of parameters among all compared methods, while maintaining a high inference speed. Such a favorable trade-off between segmentation accuracy and computational efficiency confirms that DSDINet is lightweight, efficient and suitable for PV solar cell defect inspection.

In addition, Fig. 6 provides extended qualitative results on both PSCDE and UCF-EL test sets, covering all compared SOTA methods in

Table 1
Defect segmentation results of DSDINet and state-of-the-art semantic segmentation methods on PSCDE and UCF-EL datasets.

Methods	Year	PSCDE-Test				UCF-EL-Test				Computational Efficiency			
		Precision	Recall	F1 score	IoU	Precision	Recall	F1 score	IoU	Params	GFLOPs	Size	FPS
GCNet [37]	2019	0.6131	0.4327	0.5073	0.3884	0.5789	0.5112	0.5432	0.3910	28.08	60.59	112.84	10.74
ISANet [38]	2019	0.6846	0.4860	0.5684	0.4361	0.3789	0.3123	0.3421	0.2118	44.26	41.63	111.58	10.53
K-Net [39]	2021	0.7563	0.4954	0.5987	0.4581	0.5123	0.4456	0.4765	0.3691	37.50	90.34	150.30	24.12
STDC [5]	2021	0.5477	0.3502	0.4351	0.3331	0.4456	0.3789	0.4098	0.2955	12.97	35.99	34.10	122.66
SegFormer [36]	2021	0.8061	0.6218	0.7021	0.5587	0.6012	0.5345	0.5654	0.4109	13.68	40.48	52.20	122.66
DDRNet [6]	2022	0.6489	0.4144	0.5058	0.3910	0.5567	0.4890	0.5209	0.4087	20.18	55.19	77.02	105.28
PIDNet [7]	2023	0.6055	**0.8590**	0.7103	0.5593	0.5678	0.5012	0.5321	0.4241	28.81	68.51	39.58	79.05
Seaformer [40]	2023	0.6897	0.5199	0.5925	0.4625	0.5123	0.4456	0.4765	0.3851	8.65	5.59	31.80	35.74
CGRSeg [41]	2024	0.4780	0.7696	0.5897	0.4433	0.4890	0.4213	0.4524	0.3678	18.10	7.60	87.50	65.30
SCTNet [42]	2024	0.6874	0.3820	0.4911	0.4372	0.2678	0.2011	0.2290	0.1464	17.40	8.50	94.46	260.58
SegMAN [43]	2025	0.5900	0.4087	0.4364	0.3703	0.5901	0.5234	0.5543	0.4571	6.38	**1.01**	73.53	**459.47**
GCNet [44]	2025	0.8109	0.6979	0.7337	0.6196	0.4456	0.3789	0.4098	0.2806	9.20	45.20	169.26	95.22
DSDINet	ours	**0.8243**	0.7721	**0.7973**	**0.6741**	**0.6679**	**0.6012**	**0.6328**	**0.5087**	**3.82**	8.44	**15.67**	140.86

Note: Best results are highlighted in **bold**. Computational efficiency metrics (Params, GFLOPs, Model Size, FPS) are measured under consistent hardware settings.

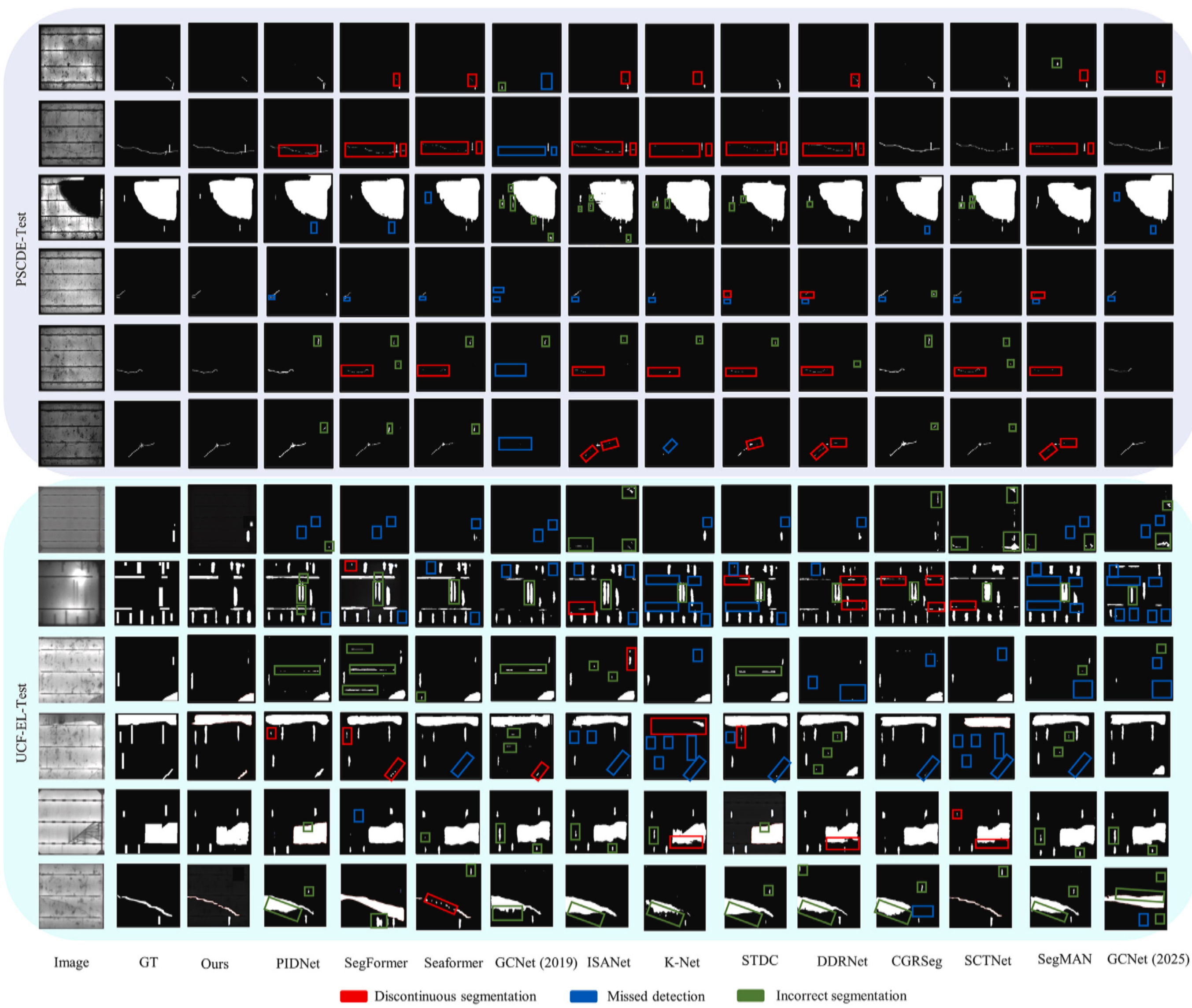

Fig. 6. Qualitative comparison of our DSDINet with state-of-the-art semantic segmentation networks. Our DSDINet achieves continuous segmentation, avoids tiny and edge defect missed detections (marked by blue bounding boxes), alleviates discontinuous defect segmentation (marked by red bounding boxes) and reduces incorrect segmentation of non-defective areas (marked by green bounding boxes) compared to other methods.

Table 2
The effectiveness of GLA and SIDA on PSCDE dataset.

Networks	PSCDE-Test			
	Precision	Recall	F1 score	IoU
(a) Baseline	0.7647	0.6529	0.7044	0.5667
(b) + GLA_G	0.8040	0.6833	0.7388	0.6007
(c) + GLA_L	0.8020	0.6723	0.7314	0.5924
(d) + GLA	0.8056	0.7116	0.7557	0.6232
(e) + SIDA_f	0.7830	0.6908	0.7340	0.5940
(f) + SIDA_fb	0.7386	0.7112	0.7248	0.5861
(g) + SIDA_ff w/o MFE	0.7770	0.6973	0.7350	0.6009
(h) + SIDA_ff	0.7853	0.7346	0.7591	0.6262
(i) + SIDA	0.7882	0.7522	0.7698	0.6356
(j) + GLA + SIDA_f	0.8091	0.7242	0.7643	0.6384
(k) + GLA + SIDA_fb	0.8074	0.7313	0.7675	0.6399
(l) + GLA + SIDA_ff	0.8116	0.7389	0.7735	0.6457
(m) DSDINet	**0.8243**	**0.7721**	**0.7973**	**0.6741**

Note: Best results are highlighted in **bold**.

Table 1. Representative examples were selected to cover tiny defects, edge-located defects, and background-interference cases. Visually, our DSDINet achieves complete and continuous defect segmentation results. In the areas marked by blue bounding boxes, fewer missed detections of tiny and edge defects are observed compared to competing methods. Similarly, the red bounding boxes show that our method may help mitigate discontinuous segmentation. Notably, the green bounding boxes highlight that our model exhibits fewer false positives caused by background noise, demonstrating its potential benefit in suppressing interference. The primary reason is that the GLA module can explore the semantic dependencies from both local and global perspectives on the deep feature maps, providing initial information on the locations of defects for the subsequent fusion of adjacent features. Moreover, the proposed SIDA module can decouple the lower-level features and mine multiscale defect information in the foreground features, enabling the model to continuously segment defects and reduce incorrect segmentation caused by cluttered backgrounds. Overall, DSDINet can more comprehensively and accurately perceive defect features. Therefore, in PV solar cell defect segmentation task, our method can achieve better performance than other segmentation networks.

4.3. Ablation study of DSDINet

To evaluate the effectiveness of the proposed SIDA module and GLA module in DSDINet, comprehensive ablation study is conducted on PSCDE. We replaced both the SIDA module and the GLA module in DSDINet with standard convolution layers to serve as our baseline. The defect segmentation results on the PSCDE test set are reported in Table 2, which quantitatively demonstrate the performance gains from each module.

To further validate these improvements from a qualitative perspective, we visualize the feature refinement process via heatmaps in Fig. 7. As shown in the first row, the raw high-level feature map e_4 initially focuses on background busbars, while mid- and low-level features exhibit significant background noise. After refinement by our GLA and SIDA modules, e_{4_refine} redirects attention to the defect regions. Meanwhile, e_{3_refine} to e_{1_refine} show substantial suppression of background noise and enhanced defect semantic responses. These visual results complement the qualitative ablation study, providing direct evidence that our modules effectively address background interference and semantic information loss.

The effectiveness of GLA

As shown in Table 2, adding global attention (b) and local attention (c) to the base model (a) improves the performance of defect segmentation to some extent, and their combination (d) yields even better results, increasing IoU by 5.65%. The qualitative results shown in Fig. 8 also indicate that the GLA module can effectively direct the model to concentrate more on the defective areas, reducing missed detection of edge-located defects and tiny defects. This confirms that the GLA module is beneficial for accurate PV solar cell defect segmentation.

The effectiveness of SIDA

Based on Table 2(a), the introduction of adjacent feature fusion (e), background feature fusion (f), and foreground feature fusion (g) significantly improves the segmentation results. Incorporating multiscale exploration of foreground features (h) yields additional valuable information. Considering both foreground, background, and multiscale foreground feature mining, i.e., (i), we achieved even better results. For instance, the addition of the SIDA module results in performance improvements of 6.54% for F1 score and 6.89% in IoU which demonstrates

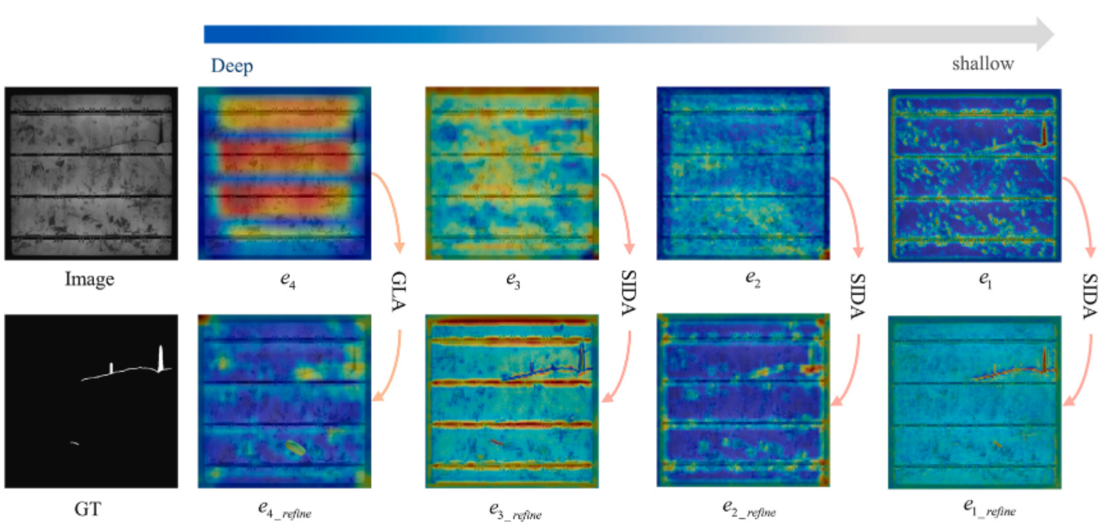

Fig. 7. Heatmap visualization of feature refinement by SIDA and GLA modules. The first row shows the original input PV solar cell image and raw feature maps from different network layers, where red indicates regions of high model attention and blue indicates low attention. The second row presents the ground truth segmentation mask and the refined feature maps after processing by GLA and SIDA modules. Notably, the raw high-level feature map e_4 initially focuses on background busbars, while e_{4_refine} redirects attention to the defect regions. For SIDA-refined features e_{3_refine} to e_{1_refine}, background noise is significantly suppressed and defect semantic responses are enhanced, demonstrating the effectiveness of our modules.

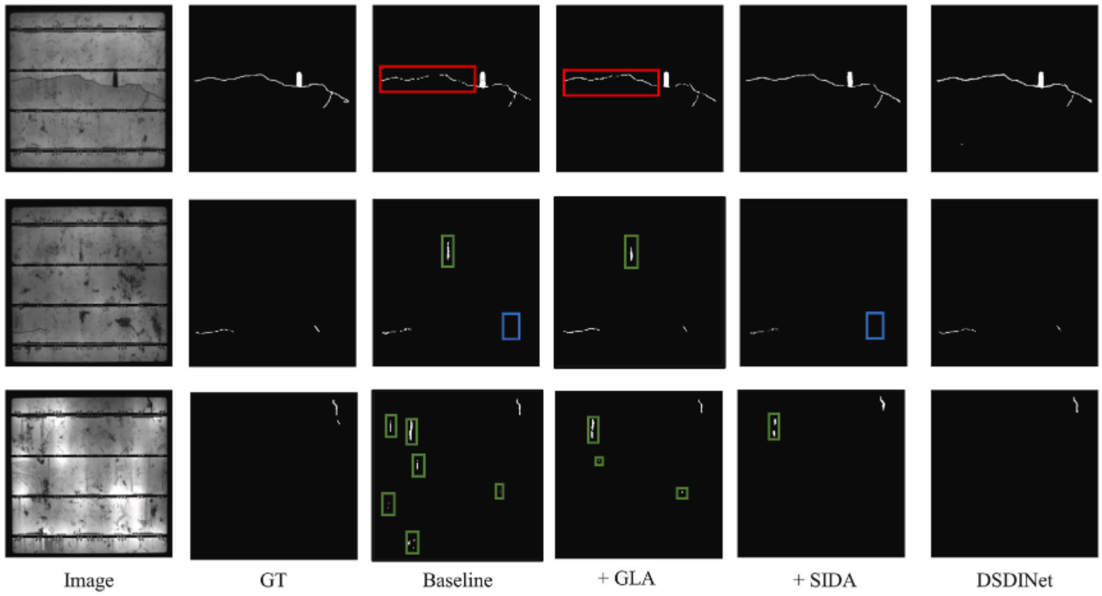

Fig. 8. Qualitative demonstration of the contributions of GLA module and SIDA module in DSDINet. The GLA module guides the model to focus on defective areas, reducing missed detection of edge-located and tiny defects (marked by blue bounding boxes); the SIDA module alleviates discontinuous defect segmentation (marked by red bounding boxes) and reduces incorrect segmentation of non-defective areas (marked by green bounding boxes).

that SIDA module greatly enhances our method's capability for accurate defect segmentation. The results from (j) to (l) demonstrate that the above conclusions remain valid when partial or full SIDA module is added to (d). Furthermore, the qualitative results in Fig. 8 show that the SIDA module significantly alleviates the issue of discontinuous defect segmentation and reduces misclassifications of non-defective areas, which is of great significance for the operation and maintenance of PV power plants.

5. Conclusion

To achieve accurate segmentation of PV solar cell defects, we introduce a defect segmentation network called DSDINet. To address the challenge of low defect contrast and complex backgrounds in PV solar cells, we introduce the Semantic Information Decoupling and Aggregation module to decouple the low-level features into foreground and background, which are then fused with higher-level features. Based on this, to enhance the model's capacity to segment defects of various sizes, we incorporate the Multiscale Feature Exploration module to perform multiscale context perception on the foreground features. Furthermore, to address the challenge of losing high-level semantic information, we design the Global and Local Attention module to obtain more comprehensive defect semantics from the deepest features. The proposed DSDINet achieves competitive performance on public photovoltaic defect segmentation datasets, thereby validating the effectiveness of DSDINet.

CRediT authorship contribution statement

Ziai Zhou: Writing – original draft, Software, Methodology. **Chaoyang Song:** Visualization. **Shixiong Fang:** Funding acquisition. **Huimin Lu:** Project administration. **Jinxia Zhang:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by National Natural Science Fund of China (Grant number 62573123), Research Fund for Advanced Ocean Institute of Southeast University, Nantong (GP202411), and the Fundamental Research Funds for the Central Universities. We also thank the Big Data Computing Center of Southeast University for providing the facility support on the numerical calculations in this paper.

Data availability

Data will be made available on request.

References

[1] Snapshot of global PV markets 2024. International Energy Agency; 2024. [Online]. Available: https://iea-pvps.org/snapshot-reports/snapshot-2024/.
[2] Chen X, Li B, Braid JL, Byford B, Colvin DJ, Glaws A, et al. Open data sets for assessing photovoltaic system reliability. Appl Energy Oct 2025;395(126132). https://doi.org/10.1016/j.apenergy.2025.126132
[3] Dhimish M, d'Alessandro V, Daliento S. Investigating the impact of cracks on solar cells performance: analysis based on nonuniform and uniform crack distributions. IEEE Trans Ind Inform 2022;18(3):1684–93. https://doi.org/10.1109/TII.2021.3088721
[4] Drir N, Chekired F, Mellit A, Blasuttigh N. Hybrid CNN-EML model for fault diagnosis in electroluminescence images of photovoltaic cells. Renew Energy Sep 2025;250(123343). https://doi.org/10.1016/j.renene.2025.123343
[5] Fan M, Lai S, Huang J, Wei X, Chai Z, Luo J, et al. Rethinking bisenet for real-time semantic segmentation. In: IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR); 2021. p. 9716–25.
[6] Pan H, Hong Y, Sun W, Jia Y. Deep dual-resolution networks for real-time and accurate semantic segmentation of traffic scenes. IEEE Trans Intell Transp Syst 2023;24(3):3448–60. https://doi.org/10.1109/TITS.2022.3228042
[7] Xu J, Xiong Z, Bhattacharyya SP. PIDNet: a real-time semantic segmentation network inspired by PID controllers. In: IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR); 2023. p. 19529–39. https://doi.org/10.1109/CVPR52729.2023.01871
[8] Lei X, Zhang J, He Y, Fang S, Zhang K, Wei H. Knowledge distillation based on fast normalization attention and auxiliary multi-label recognition for continual photovoltaic defect segmentation. Renew Energy Jan 2026;256(124505). https://doi.org/10.1016/j.renene.2025.124505
[9] Dong S, Zhang J, Shen Y, Gao D. Referring solar cell defect segmentation in electroluminescence images. IEEE Trans Ind Inform 2025;21(9):6788–98. https://doi.org/10.1109/TII.2025.3567396
[10] Bao J, Yuan X, Wu Q, Lam C-T, Ke W, Li P. CCA-YOLO: channel and coordinate aware-based YOLO for photovoltaic cell defect detection in electroluminescence images. IEEE Trans Instrum Meas 2025;74(5009412). https://doi.org/10.1109/TIM.2025.3541805

[11] Duan R, Wang Y, Chen X, Li S. An enhanced algorithm for cell-level anomaly segmentation in photovoltaic solar panels using electroluminescence imaging. Energy Sep 2025;331(136711). https://doi.org/10.1016/j.energy.2025.136711

[12] Song C, Zhang J, Fang S, Chen L. A photovoltaic defect segmentation framework integrating domain knowledge and fine-tuned SAM. Sol Energy Jan 2026;303(114145). https://doi.org/10.1016/j.solener.2025.114145

[13] Wang C, Chen H, Zhao S. RERN: rich edge features refinement detection network for polycrystalline solar cell defect segmentation. IEEE Trans Ind Inform 2024;20(2):1408–19. https://doi.org/10.1109/TII.2023.3275705

[14] Zhang J, Shen Y, Jiang J, Fang S, Chen L, Yan T, et al. Automatic detection of defective solar cells in electroluminescence images via global similarity and concatenated saliency guided network. IEEE Trans Ind Inform 2023;19(6):7335–45. https://doi.org/10.1109/TII.2022.3211088

[15] Fioresi J, Colvin DJ, Frota R, Gupta R, Li M, Seigneur HP, et al. Automated defect detection and localization in photovoltaic cells using semantic segmentation of electroluminescence images. IEEE J Photovolt 2022;12(1):53–61. https://doi.org/10.1109/JPHOTOV.2021.3131059

[16] Chen L-C, Papandreou G, Schroff F, Adam H. Rethinking atrous convolution for semantic image segmentation.2017, arXiv:1706.05587.

[17] Chen X, Karin T, Libby C, Deceglie M, Hacke P, Silverman TJ, et al. Automatic crack segmentation and feature extraction in electroluminescence images of solar modules. IEEE J Photovolt 2023;13(3):334–42. https://doi.org/10.1109/JPHOTOV.2023.3249970

[18] Rahman MRU, Chen H. Defects inspection in polycrystalline solar cells electroluminescence images using deep learning. IEEE Access 2020;8:40547–58. https://doi.org/10.1109/ACCESS.2020.2976843

[19] Han H, Gao C, Zhao Y, Liao S, Tang L, Li X. Polycrystalline silicon wafer defect segmentation based on deep convolutional neural networks. Pattern Recognit Lett 2020;130(SI):234–41. https://doi.org/10.1016/j.patrec.2018.12.013

[20] Zhou P, Wang R, Wang C, Chen H, Liu K. SIIF: semantic information interactive fusion network for photovoltaic defect segmentation. Appl Energy Oct 2024;371. https://doi.org/10.1016/j.apenergy.2024.123643

[21] Zhao S, Chen H, Wang C, Zhou Y, Zhang Z. RGR-net: refined graph reasoning network for multi-height hotspot defect detection in photovoltaic farms. Expert Syst Appl Jul 2024;245. https://doi.org/10.1016/j.eswa.2023.123034

[22] Wu Z, Su L, Huang Q. Cascaded partial decoder for fast and accurate salient object detection. In: IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR); 2019. p. 3902–11. https://doi.org/10.1109/CVPR.2019.00403

[23] Mei H, Ji G-P, Wei Z, Yang X, Wei X, Fan D-P. Camouflaged object segmentation with distraction mining. In: IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR); 2021. p. 8768–77. https://doi.org/10.1109/CVPR46437.2021.00866

[24] Zheng Q, Qiao X, Cao Y, Lau RWH. Distraction-aware shadow detection. In: IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR); 2019. p. 5162–71. https://doi.org/10.1109/CVPR.2019.00531

[25] Lei M, Wu H, Lv X, Wang X. ConDSeg: a general medical image segmentation framework via contrast-driven feature enhancement. In: AAAI Conf. Artif. Intell. (AAAI); 2025.

[26] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 30; 2017.

[27] Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR); 2018. p. 7132–41. https://doi.org/10.1109/CVPR.2018.00745

[28] Wang X, Girshick R, Gupta A, He K. Non-local neural networks. In: IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR); 2018. p. 7794–803. https://doi.org/10.1109/CVPR.2018.00813

[29] Wu H, Xie J, Deng W, Lin A, Shariff ARM, Akmalov S, et al. CT-HiffNet: a contour-texture hierarchical feature fusion network for cropland field parcel extraction from high-resolution remote sensing images. Comput. Electron. Agric. Dec 2025;239(B). https://doi.org/10.1016/j.compag.2025.111010

[30] Xie J, Zhang M, Wu H, Lin A, Adami M, Shariff ARM, et al. S2-IFNet: a spatial-semantic information fusion network integrated with boundary feature enhancement for forest land extraction from Sentinel-2 data. Int J Appl Earth Obs Geoinf May 2025;139(104505). https://doi.org/10.1016/j.jag.2025.104505

[31] Woo S, Park J, Lee J-Y, Kweon IS. CBAM: convolutional block attention module. In: Eur. Conf. Comput. Vis. (ECCV), vol. 11211; 2018. p. 3–19. https://doi.org/10.1007/978-3-030-01234-2_1

[32] Fu J, Liu J, Tian H, Li Y, Bao Y, Fang Z, et al. Dual attention network for scene segmentation. In: IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR); 2019. p. 3141–9. https://doi.org/10.1109/CVPR.2019.00326

[33] Junyang X, Anqi L, Hao W, Ziwei W, Wenbin W, Qiangyi Y. JAM-R-CNN deep learning network model for remote sensing recognition of terraced fields. Natl. Remote Sens. Bull. 2025;28(12):3136–46.

[34] Liu M, Yin H. Feature pyramid encoding network for real-time semantic segmentation. In: BR. Mach. Vis. Conf. (BMVC); 2019.

[35] Yan S, Yao X, Sun J, Huang W, Yang L, Zhang C, et al. TSANet: a deep learning framework for the delineation of agricultural fields utilizing satellite image time series. Comput. Electron. Agric. May 2024;220(108902). https://doi.org/10.1016/j.compag.2024.108902

[36] Xie E, Wang W, Yu Z, Anandkumar A, Alvarez JM, et al. SegFormer: simple and efficient design for semantic segmentation with transformers. In: Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 34; 2021.

[37] Cao Y, Xu J, Lin S, Wei F, Hu H. GCNet: non-local networks meet squeeze-excitation networks and beyond. In: IEEE Int. Conf. Comput. Vis. (ICCV); 2019. p. 1971–80. https://doi.org/10.1109/ICCVW.2019.00246

[38] Huang L, Yuan Y, Guo J, Zhang C, Chen X, Wang J. Interlaced sparse self-attention for semantic segmentation. [arXiv preprint]. 2019 arXiv:1907.12273.

[39] Zhang W, Pang J, Chen K, Loy CC. K-Net: towards unified image segmentation. In: Adv. Neural Inf. Process. Syst. (NeurIPS); 2021.

[40] Wan Q, Huang Z, Lu J, Yu G, Zhang L. Seaformer: squeeze-enhanced axial transformer for mobile semantic segmentation. In: Int. Conf. Learn. Represent. (ICLR); 2023.

[41] Ni Z, Chen X, Zhai Y, Tang Y, Wang Y. Context-guided spatial feature reconstruction for efficient semantic segmentation. In: Eur. Conf. Comput. Vis. (ECCV); 2024. p. 239–55.

[42] Xu Z, Wu D, Yu C, Chu X, Sang N, Gao C. SCTNet: single-branch CNN with transformer semantic information for real-time segmentation. In: AAAI Conf. Artif. Intell. (AAAI); 2024. p. 6378–86.

[43] Fu Y, Lou M, Yu Y. SegMAN: omni-scale context modeling with state space models and local attention for semantic segmentation. In: IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR); 2025. p. 19077–87.

[44] Yang G, Wang Y, Shi D, Wang Y. Golden cudgel network for real-time semantic segmentation. In: IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR); 2025. p. 25367–76.