Research Article Vol. 32, No. 10 / 6 May 2024 / Optics Express 17295
Optics EXPRESS

Deep learning-based method for defect detection in electroluminescent images of polycrystalline silicon solar cells

YUQI LIU, YIQUAN WU,* YUBIN YUAN, AND LANGYUE ZHAO

College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China
*imagestrong@nuaa.edu.cn

Abstract: To achieve defect detection in bare polycrystalline silicon solar cells under electroluminescence (EL) conditions, we have proposed ASDD-Net, a deep learning algorithm evaluated offline on EL images. The model integrates strategies such as downsampling adjustment, feature fusion optimization, and detection head improvement. The ASDD-Net utilizes the Space to Depth (SPD) module to effectively extract edge and fine-grained information. The proposed Enhanced Cross-Stage Partial Network Fusion (EC2f) and Hybrid Attention CSP Net (HAC3) modules are placed at different positions to enhance feature extraction capability and improve feature fusion effects, thereby enhancing the model's ability to perceive defects of different sizes and shapes. Furthermore, placing the MobileViT_CA module before the second detection head balances global and local information perception, further enhancing the performance of the detection heads. The experimental results show that the ASDD-Net model achieves a mAP value of 88.81% on the publicly available PVEL-AD dataset, and the detection performance is better than the current SOTA model. The experimental results on the ELPV and NEU-DET datasets verify that the model has some generalization ability. Moreover, the proposed model achieves a processing frame rate of 69 frames per second, meeting the real-time defect detection requirements for solar cell surface defects.

1. Introduction

In recent years, with the increasing dependence on energy resources, the widespread use and consumption of non-renewable energy sources such as petroleum have led to escalating environmental pollution issues. Solar photovoltaic (PV) technology [1], known for its low production cost, lack of pollution, and renewable nature, has been widely adopted, alleviating the energy crisis in human society [2]. As the export of photovoltaic components continues to grow, the production of solar cells is on the rise. During the manufacturing process of solar cells, defects may occur due to various causes, resulting in a variety of defect types including scratches, damage, stains, and grid breaks, among others [3,4]. The presence of defects significantly diminishes the stability of PV power generation systems, and defect detection in solar cells is a crucial step to ensure their performance and quality. Therefore, to guarantee the safe and efficient operation of PV power stations, automated approaches based on deep learning play a vital role in achieving various defect detections in solar cells [5].

Typically, solar cells can be categorized into monocrystalline silicon and polycrystalline silicon based on different manufacturing materials. Monocrystalline silicon solar cells exhibit uniform background textures, while polycrystalline silicon solar cells feature numerous randomly shaped and sized crystalline particles on the surface. Defect detection in solar cells primarily involves predicting the classification and localization information of multi-scale defects in Electroluminescent (EL) images. The detection network not only possesses classification capabilities but also determines defect positions and sizes through bounding boxes, enabling

#517341 https://doi.org/10.1364/OE.517341
Journal © 2024 Received 28 Dec 2023; revised 3 Apr 2024; accepted 3 Apr 2024; published 26 Apr 2024

Optics EXPRESS

precise identification and localization of defects on solar cell surfaces. Based on network structures, deep learning-based solar cell defect detection networks can be broadly classified into two categories: two-stage networks represented by Faster R-CNN [6], and one-stage networks represented by SSD [7] and YOLO [8]. Two-stage networks predict defect positions and categories based on generated defect proposals regions, but their real-time deployment efficiency is relatively low in resource-constrained scenarios. One-stage networks directly classify and locate defects using extracted features, reducing overall computational demands and making them more deployable in edge devices and embedded systems, exhibiting better adaptability in resource-constrained environments [9]. The comprehensive performance of the YOLO family of networks in the single-stage model is superior and is particularly suitable for application scenarios that require real-time performance, efficiency, and accuracy, making the YOLO family of networks a widely used framework in the field of target detection [10].

Among the existing YOLO series models, YOLOv5 demonstrates efficient deployment on resource-constrained embedded devices and edge computing platforms. Its adaptability is further enhanced by the introduction of various model variants, providing greater flexibility for industrial applications. Due to its real-time processing capabilities, ease of deployment, and the ability to balance detection accuracy, YOLOv5 has emerged as the most widely used model in industrial defect detection tasks [11]. Solar cell defects exhibit linear features and spatial continuity, with non-uniform proportions of defect areas across the entire EL image. This uneven distribution results in targets of varying scales, with some defects occupying a relatively small proportion. YOLO series models still face challenges in achieving high accuracy in multi-scale and small object detection [12]. Therefore, this paper focuses on addressing the detection issues of minute and multi-scale defects in polycrystalline silicon solar cells, proposing a model named ASDD-Net.

The primary contributions of this paper can be summarized as follows:

1. In order to effectively address subtle defects distributed on solar cells, this paper introduces the Space to Depth (SPD) module in the feature extraction section. This module performs downsampling by altering the tensor dimensions and introduces new elements in the channel dimension to fuse features at different scales. It significantly enlarges the receptive field for small target defects, alleviating the issue of potential loss of crucial feature information.

2. To enhance the network's perception of defects with different sizes and shapes, and simultaneously improve feature fusion to emphasize defect areas, this paper designs the EC2f and HAC3 feature fusion modules. The EC2f module, incorporating multi-level feature fusion and multi-head self-attention mechanisms, better integrates information from different spatial and channel dimensions, thereby enhancing the model's global perception capabilities. To improve feature expression and distinctiveness, the HAC3 module introduces ECA-Net and SimAM, further enhancing the detection of multi-scale defects.

3. To enhance the performance of the output detection heads, with a focus on the intended detection of target objects, the ASDD-Net model introduces the MobileViT_CA module before the second detection head. Leveraging its characteristics in global context modeling and channel attention enhancement improves the performance of the detection heads while maintaining the importance of local features.

4. This paper compares the detection performance with other frameworks in the same series and different detection frameworks, demonstrating the superior performance of the proposed method. Furthermore, the ASDD-Net model's generalization capability is verified on two additional datasets.

Optics EXPRESS

The remaining structure of this paper can be divided into the following sections: Section 2 introduces related work on solar cell defect detection, Section 3 presents the ASDD-Net model structure, Section 4 validates the effectiveness of ASDD-Net through analysis of ablation experiments, comparative experiments, and generalization experiments, and Section 5 summarizes the work and outlines future research directions.

2. Related work

Currently, deep learning models trained on extensive datasets have gained widespread popularity, with representative convolutional neural networks being widely employed in areas such as object detection, image classification, and semantic segmentation. Presently, the task of surface defect detection on solar cells is progressively being accomplished through deep learning, holding significant implications for the realization of intelligent manufacturing. However, these deep learning models are typically designed for natural scene images, and their direct application to surface defect detection in solar cell EL images presents several challenges. Consequently, researchers need to adopt task-specific strategies, such as data processing, feature engineering, and innovative neural network architectures, to better address the complexity and uniqueness of solar cell defect detection. Based on the types of industrial inspection tasks, defect detection algorithms leveraging deep learning can be categorized into three types: classification networks, detection networks, and segmentation networks.

Deitsch et al. [13] proposed two deep learning-based methods for the automatic detection of defects in PV cells using Convolutional Neural Networks (CNN) and Support Vector Machines (SVM). Experimental results showed that the CNN classifier achieved high accuracy in defect detection. Pierdicca [14] employed transfer learning with the VGG-16 network to classify remote sensing images of solar cells. However, due to the lower image resolution in the self-constructed photovoltaic module electroluminescence image dataset, the accuracy of the CNN network is only about 70%. Tang et al. [15] designed a CNN-based automatic classification model for EL image defects. They input deep features extracted by CNN into fully connected layers to classify images into four defect classes. However, the model only determines whether defects are present and cannot identify the specific defect locations or types. Sridhar et al. [16] conducted data augmentation on existing unmanned aerial vehicle captured PV images to expand the dataset. They employed a CNN model to classify samples into five fault types and a defect-free category. Their model achieved a notably high level of accuracy. Korkmaz et al. [17] modified a pre-trained architecture to design a novel multi-scale model for detecting various defects in solar panels. This approach exhibited high robustness and classification performance.

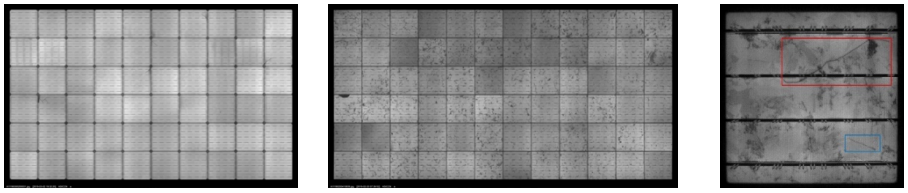
Su et al. [18] proposed a novel object detector for photovoltaic cell defect detection, incorporating a designed bidirectional feature pyramid into the model. This allowed for the effective recognition and detection of hidden cracks, grid breaks, and black spot defects. However, the feature balance factor in the algorithm still requires manual adjustment. To detect hidden cracks and grid break defects within polycrystalline solar panels, Zhang et al. [19] designed a multi-feature region proposal fusion network structure. This network extracts region proposals from different feature layers of convolutional neural networks, but the model has high computational costs and long detection times. Xu et al. [20] introduced a new spatial pyramid pooling operation and channel attention to locate cracks and fragment defects in EL images based on the YOLOv5 model. Chen et al. [21] designed a novel defect object detector that embeds a dual-channel feature pyramid into YOLOv5, significantly improving the model's ability to recognize small target defects. However, the model can detect fewer types of defects on solar cells. Balcıoğlu et al. [22]designed a visual defect detection model based on a new Deep Convolutional Neural Network. In the first stage, it detects solar cell samples containing defects and ranks them based on their level of damage. In the second stage, the selected samples are classified, effectively

Optics EXPRESS

improving the detection performance for small-area defects in complex backgrounds. However, due to cost considerations, the resolution of images in their dataset is relatively low.

Han et al. [23] employed a two-stage approach to segment defects on multicrystalline silicon wafers. The first stage utilized a region generation network to generate potential defect region images. In the second stage, image blocks containing potential defects were processed into appropriate sizes and input into an improved U-Net for the segmentation of scratch and black spot defects. Pratt et al. [24] used a semantic segmentation model based on the U-Net architecture to identify and extract internal hidden crack defects and several other defects simultaneously in monocrystalline and polycrystalline photovoltaic modules. However, the subjectivity and ambiguity in manually annotated labels led to a significant amount of noise and uncertainty in the labeling of solar cell images. Rahman et al. [25] proposed a multi-attention network that incorporated channel attention to extract contextual information and spatial attention to effectively suppress background noise. They combined these two attention mechanisms and integrated them into the U-Net network for the segmentation task of small target defects. However, the dataset used in this study was obtained using PL imaging technology, resulting in relatively low-resolution images. Sohail et al. [26] combined the four models U-Net, attention U-Net, FPN, and LinkNet by ensemble learning technology, which significantly improved the segmentation effect of deep crack and micro-crack defects.

Fig. 1 shows EL images of PV modules made from monocrystalline (a) and polycrystalline (b) solar module and an individual polycrystalline solar cell (c). The monocrystalline silicon solar cell exhibits a uniform background texture, whereas the polycrystalline silicon solar cell surface contains numerous randomly shaped and sized crystalline particles. The focus of defect detection in this study is the polycrystalline silicon solar cell. In Fig. 1 (c), regions with lower grayscale values in the EL image of the polycrystalline silicon solar cell correspond to randomly distributed crystalline grids, forming a complex textured background in the EL image. The red-bordered area indicates a crack defect, and the blue-bordered area represents a background region similar to a crack defect, with both exhibiting relatively similar features.



(a) (b) (c)

Fig. 1. EL images of PV modules. (a) Images of monocrystalline silicon solar module. (b) Images of polycrystalline silicon solar module. (c) Images of polycrystalline silicon solar cells containing micro-crack defects.

In summary, there are still several limitations in solar cell defect detection, including susceptibility to interference from complex backgrounds in polycrystalline solar cell defect detection, the diversity in shapes of similar defects in solar cells, and the gradual disappearance of small defect features as network training progresses and downsampling continues. These issues highlight the need for further research in the field of solar cell defect detection.

3. Structure of the ASDD-Net network

The overall framework of the ASDD-Net network, depicted in Fig. 2, consists of four main components: Input, Backbone, Neck, and Detection Head. The Backbone section utilizes the improved SEDF-Net(Space-Enhanced Dense Feature Network) based on CSPDarknet53 [27] as the feature extraction network. The SPD module accomplishes downsampling by changing the

Optics EXPRESS

tensor's dimensions, the HAC3 module introduces a hybrid attention mechanism, and the EC2f module at the end of the network employs a multi-head self-attention mechanism for feature fusion. The Path Aggregation Network (PANet) [28] effectively integrates feature information from different network layers through top-down path enhancement and lateral connection modules. Based on PANet structure, the neck structure introduces the HAC3 and EC2f modules to integrate high-level semantic information and enhance the emphasis on defect regions more effectively. The Detection Head maps and transforms features of different scales obtained from the Neck into the output of defect targets. The model introduces the MobileViT_CA block before the second detection head to enhance the network's feature expression capability. The role of the YOLO Head is to output information about the class, location, and size of the target.

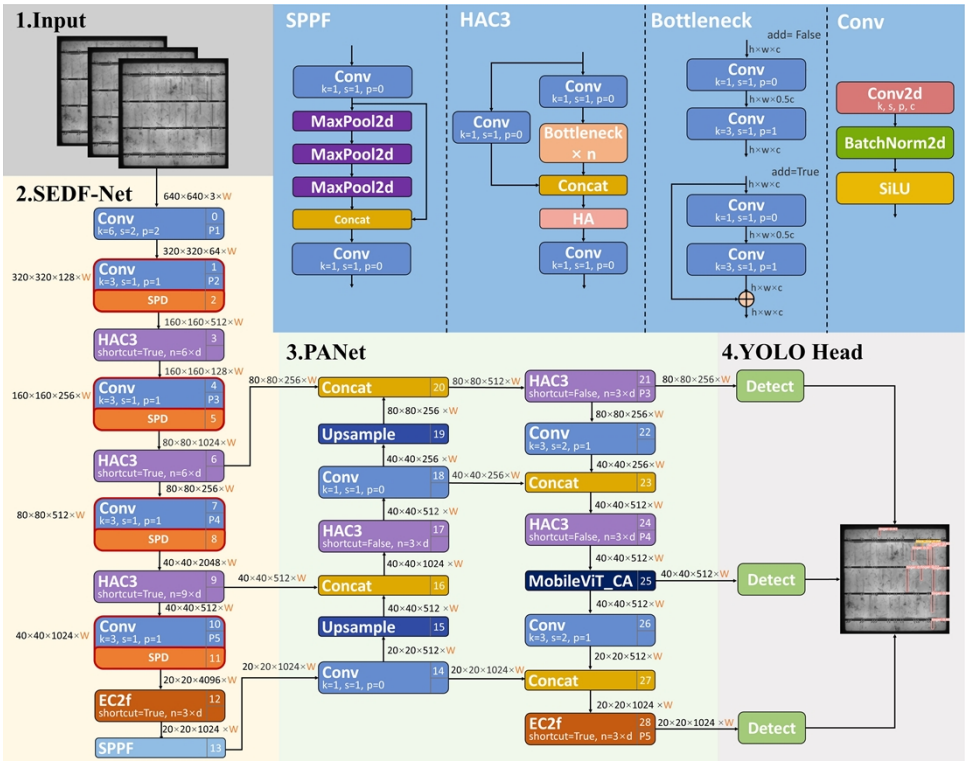


Fig. 2. The architectural diagram of the ASDD-Net network. As a feature extraction network, SEDF-Net includes convolutional layers, SPD, ECAC3, EC2f, and SPPF modules. The SPD module accomplishes downsampling by varying the dimension of the tensor, the HAC3 module comprehensively processes feature maps through a hybrid attention mechanism, and the EC2f module further optimizes feature representation using a multi-head self-attention mechanism. The incorporation of the MobileViT_CA module enhances the head features by leveraging global contextual information, thereby improving detection performance.

3.1. Downsampling adjustment

The defect images of polycrystalline solar cells contain small targets. The use of a large downsampling factor in CSPDarknet53 can result in the loss of crucial feature information for small targets in deep feature maps. Additionally, the residual network structure of CSPDarknet53 employs multiple convolutional modules, which may lead to the loss of edge and fine-grained

Optics EXPRESS

information when extracting defects in solar cells. This loss compromises the accuracy and sensitivity of the network in detecting and locating minor defects. To address this issue, specifically for the characteristic of small defect targets in polysilicon solar cell images, SEDF-Net utilizes the SPD structure for downsampling and combines it with convolution to enhance the detection performance of minor defects like scratches and patches on the surface of solar cells.

The SPD module is a spatial-to-depth convolutional layer that serves to extend the image transformation technique to downsample the feature maps within and across the CNN while preserving all information in the channel dimension [29]. In specific cases, the SPD module is used to alter the dimensions of feature maps, allowing it to capture a broader context of information and mitigating the issue of edge and fine-grained information loss to some extent. Given a feature map X of any size, the SPD module divides the input feature map tensor into non-overlapping sub-features. Each sub-feature map $f(x, y)$ is formed by entries $X(i, j)$ for all i + x and j + y that can be evenly divided by a given ratio, as shown in the following equation.

$$f_{0,0} = X[0 : S : scale, 0 : S : scale], \tag{1}$$

$$f_{1,0} = X[1 : S : scale, 0 : S : scale], \tag{2}$$

$$f_{scale-1,0} = X[scale - 1 : S : scale, 0 : S : scale]; \tag{3}$$

$$f_{0,scale-1} = X[0 : S : scale, scale - 1 : S : scale], f_{0,0} = X[0 : S : scale, 0 : scale] \tag{4}$$

$$f_{1,scale-1} = X[1 : S : scale, scale - 1 : S : scale], f_{1,0} = X[1 : S : scale, 0 : scale], \tag{5}$$

$$f_{scale-1,scale-1} = X[scale - 1 : S : scale, scale - 1 : S : scale]. \tag{6}$$

Among these parameters, "scale" represents the image scale, "f" denotes the sub-feature map, and "X" represents the original feature map. In Fig. 3, which corresponds to a scale of 2, the process is as follows: the feature map "X" is divided into four sub-maps, and these sub-maps are concatenated along the channel dimension to create a new feature map "X." As a result, the spatial dimension of the output features is reduced by a factor of two, while the channel dimension is increased by a factor of the square of the scale.

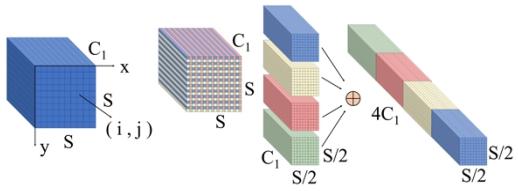


Fig. 3. Structure of the SPD module.

The SPD module achieves downsampling by reorganizing the pixels in the feature map, combining adjacent pixels into larger blocks, and stacking them together. This process can be seen as a form of downsampling. However, unlike conventional convolutional downsampling, the SPD module does not lose information from the original pixels. It can preserve details while enhancing the model's ability to perceive different locations. This is particularly helpful in detecting defects that are widely distributed across solar cell samples.

3.2. Feature fusion optimization

In order to better cope with the defects of different sizes and shapes on the surface of polycrystalline silicon solar cell wafers, we design two feature fusion modules: EC2f and HAC3. The EC2f module introduces a multi-layer feature fusion and a multi-head self-attention mechanism, which contributes to a better integration of the information from different spaces and channels by

Optics EXPRESS

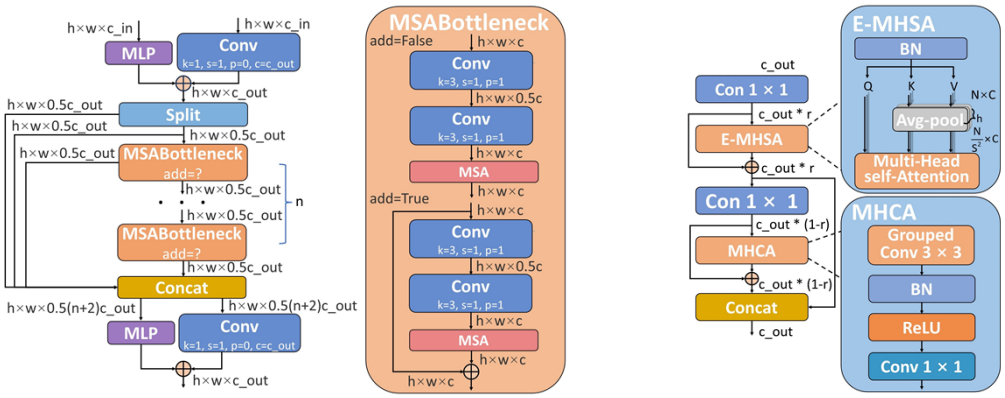
improving the global sensing capability. In contrast, the HAC3 module employs an adaptive hybrid attention mechanism, which allows the model to improve the discriminative and representational capabilities of the extracted features while the number of parameters remains largely unchanged.

Considering that the EC2f module is relatively complex in computation, we apply it at the end of the backbone network and in front of the last detection head to utilize its information integration capability to efficiently process relatively small features. Whereas, the lightweight HAC3 structure is used at other locations to accomplish feature extraction and fusion to maintain the overall computational efficiency. Through this strategy, we fully utilize the information integration capability of the EC2f module and the feature representation capability of the HAC3 module, and thus are able to enhance the overall performance of the surface defect detection network for polycrystalline silicon solar cells.

3.2.1. EC2f

Multi-scale context information is very important in the surface defect detection of polycrystalline silicon solar cells because of the diversity of defect types and sizes and the similarity between different types of defects. In order to understand the features of different defects more comprehensively, we propose a new structure named EC2f to improve the feature representation and generalization performance of the model.

The relatively weak ability of the convolutional layer to perform feature extraction on the input feature maps tends to make the network too linear, which in turn limits the feature expression capability. As shown in Fig. 4(a), the EC2f module processes the input feature x through the MLP and a 1×1 convolutional layer with a step size of 1. The feature maps on the two paths are then summed to form the fusion feature Z, which enables the module to learn richer and more complex feature representations. Next, the fused features are split in the channel dimension and fed into multiple MSABottleneck modules, each of which learns a different feature representation, forming a hierarchical information extraction. The global relevance features are strengthened by splicing the output of the multi-head self-attention mechanism with the original input features in the channel dimension, while retaining certain original feature information. Finally, the spliced features are input to the MLP and a 1×1 convolutional layer with a step size of 1. The two outputs are summed to obtain the final output.



(a) (b)

Fig. 4. EC2f Structure: (a) EC2f module, (b) MSA module.

Transformer block exhibits a robust capability in capturing low-frequency signals, providing essential global information such as overall shapes and structures. However, Transformer blocks may to some extent overlook the local details of the image, especially high-frequency

Optics EXPRESS

signals or local textures [30]. To address this issue, Li et al. [31] proposed a module called Next Transformer Block (NTB), which enables the acquisition of multi-frequency signals in a lightweight mechanism to further enhance the overall modeling capability. Inspired by the NTB module and combined with the characteristics of the Bottleneck structure, this paper introduces the Multi-Scale Attention (MSA) module in each Bottleneck module, thus forming the MSABottleneck structure. This module inherits not only the visual structure and low-level feature extraction advantages of convolutional neural networks but also integrates the Transformer's capability to focus on global information. As illustrated in Fig. 4(b), the MSA structure consists of two key components: Efficient Multi-Head Self-Attention (E-MHSA) and Multi-Head Cross-Attention (MHCA). E-MHSA is responsible for processing the global information, while MHCA focuses on the extraction of local details. This design allows the MSABottleneck structure to focus more comprehensively on various information in the image, thereby improving the model's sensitivity and expressiveness to different features.

To accelerate the inference speed, channel dimension reduction is initially performed using pointwise convolution, followed by the utilization of E-MHSA to capture low-frequency signals. The E-MHSA module conducts attention computation through the spatial reduction self-attention operator SA. This operator maps various tensors to distinct spaces through linear mappings of query (Q), key (K), and value (V):

$$Q = X \bullet W_q \tag{7}$$

$$K = X \bullet W_{\mathrm{k}} \tag{8}$$

$$V = X \bullet W_{\mathrm{v}} \tag{9}$$

In the above equation, X represents the input features, and W_q, W_k, W_v are the weight matrices for query, key, and value, respectively. Before the self-attention computation, it is possible to perform downsampling on the input features. After downsampling, the obtained input features undergo linear transformations and are partitioned into a multi-head form along the channel dimension, denoted as:

$$Z = [z_1, z_2, \ldots, z_h] \tag{10}$$

In Eq. (10), Z denotes the input features. Subsequently, the similarity between queries and keys is computed through dot product operation, followed by the application of the softmax operation to obtain attention weights for each head. This process can be described as:

$$Attention_i = Soft\max(\frac{Q_i \bullet [P_{\mathrm{s}}(K_i)]^T}{\sqrt{d_k}}) \tag{11}$$

In Eq. (11), d_k represents the scaling factor, and i indicates the i-th attention head. The attention weights computed are then utilized to perform a weighted sum with the values, yielding the output of the self-attention operator, denoted as SA:

$$SA(X) = Attention_{\mathrm{i}} \bullet P_s(V_{\mathrm{i}}) \tag{12}$$

Concatenate the attention calculation results, $SA(Z_{\mathrm{i}})$, from each head to obtain a comprehensive representation of multi-head attention. Finally, perform a linear mapping through a projection matrix to obtain the ultimate output:

$$E - MHSA(z) = Concat(SA_1(z_1), SA_2(z_2), \ldots, SA_h(z_h))W^p \tag{13}$$

In Eq. (13), W^P represents the linear projection matrix. Self-attention allows for the consideration of both content information and the features at various positions, as well as their relative distances. This effectively associates information between different objects, including positional information.

Optics EXPRESS

The feature maps, after adjusting the channel dimensions, are then input into the MHCA module to achieve the simultaneous capture of mixed high and low-frequency information, integrating information at different levels of feature extraction. The MHCA module performs multi-head convolution by utilizing group convolution in the convolution operation. The computed results are subsequently processed through normalization layers and activation functions. Finally, the outcomes obtained from each attention head undergo linear projection through the weight matrix W^P. This entire process can be represented as:

$$MHCA(z) = Concat(CA_1(z_1), CA_2(z_2), \ldots, CA_h(z_h))W^P \tag{14}$$

The z in Eq. (14) has the same meaning as in Eq. (10), which is that the feature z is divided into a multi-head form in the channel dimension, which also leads to an increasing number of channel number tables at different stages. the CA is a single-head convolutional attention, which can be defined as:

$$CA(\mathrm{z}) = O(W, (T_m, T_n)) \; where \; T\{m, n\} \; \in \mathrm{z} \tag{15}$$

For each channel a division of Token is performed, where Tm and Tn are the neighboring Token inside this channel, and O is an inner product operation with trainable parameter W for learning the correlation between different Token in the local receptive field. The E-MHSA module first focuses on the information of different locations, while the MHCA module weights the information of different channels in the subsequent stages. This staged attention computation helps the model to gradually focus on more advanced semantic features, while the attention weighting in the channel dimension helps to optimize the feature representation and differentiation ability.

3.2.2. HAC3

In the EL images of polycrystalline silicon solar cells, where the area occupied by defect features is small and there are multi-scale variations, a large amount of redundant background information constitutes an interference for the extraction of defect features. To cope with this problem, a feature fusion module named HAC3 is proposed in this study. The structure of this module HAC3 is divided into two branches: one contains a standard convolutional layer and a stack of multiple Bottleneck blocks, and the other includes only a basic convolutional module. The outputs of these two branches are fused by a concat operation, and the fused feature maps are subsequently fed into the Hybrid Attention (HA) module for stronger feature representation. Finally, the number of channels of the enhanced features is adjusted by a 1×1 convolutional layer to form the final output feature map. The overall structure of the HA module is shown in Fig. 5, and the weights of the ECA-Net and SimAM are dynamically adjusted within the module to adapt to the needs of detecting defects at different scales.

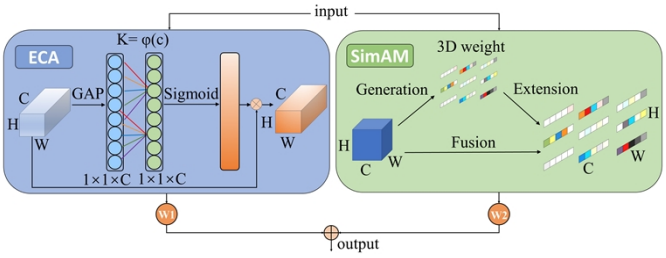


Fig. 5. Structure diagram of HA.

Defects are usually manifested as anomalies or special features in an image, and these features may have different manifestations in different channels. In order to extract these critical features

Optics EXPRESS

more efficiently, we use the Effective Channel Attention (ECA) mechanism in the HA module to pay attention to and adjust the weight of each channel to adapt to the importance of different features. ECA-Net was proposed by Wang et al. [32] in 2020, which adaptively calculates the correlation between the neighboring k channels through a one-dimensional convolution operation, thus avoiding irrelevant information interference.

The ECA attention module begins by applying global average pooling to the input feature map of dimensions $H \times W \times C$. Subsequently, it employs a 1D convolution operation with a convolution kernel of size 'k.' The value of 'k' is determined by an adaptive function based on the input channel count 'C,' as given in Eq. (16), where $|x|_{odd}$ represents the nearest odd integer to 'x'.

$$k = \varphi(c) = \left| \frac{\log_2 c + 1}{2} \right|_{odd} \tag{16}$$

After the convolution operation, weights W for each channel are obtained through the Sigmoid activation function. To further enhance network performance, a shared weight convolution is employed to efficiently capture local interactive channel information, reducing network parameter count. The shared weight method is formulated as in Eq. (17),

$$k = \varphi(c) = \left| \frac{\log_2 c + 1}{2} \right|_{odd} \tag{17}$$

In Eq. (17), σ represents the sigmoid activation operation, W_I denotes the i-th weight matrix obtained by grouping C channels, and W_i^jrefers to the j-th local weight matrix within the i-th weight matrix. Similarly, Y_i^j can be defined. Finally, the obtained weights are multiplied by the original input feature map to obtain feature maps with attention weights. The ECA attention module removes the fully connected layer after global average pooling and utilizes a 1×1 convolutional layer. This avoids dimension reduction and effectively captures inter-channel interaction information, achieving excellent attention effects with fewer parameters.

Given that focusing solely on channel features in the feature fusion structure may result in the algorithm losing positional features of defects, the HA module incorporates the Similarity-based Attention Mechanism (SimAM) [33]. SimAM attention mechanism is different from SE module and CBAM module which can only refine the features by spatial or channel dimension, and its attention weight can be flexibly changed with space and channel, which makes the network more discriminative and sensitive to the location of defects.

The energy function of each neuron is defined as shown in Eq. (18), and the degree of differentiation between the target neuron and the surrounding neurons is inversely proportional to the energy value and directly proportional to the importance, i.e., the greater the degree of differentiation, the greater the importance of the neuron, and the smaller its energy.

$$e_t(w_t, b_t, y, x_i) = \frac{1}{M-1} \sum_{i=1}^{M-1} [-1 - (w_t x_i + b_t)]^2 + [1 - (w_t t + b_t)]^2 + \lambda w_t^2 \tag{18}$$

In Eq. (18), M is the number of all neurons on a single channel; i is the index on the spatial dimension; x_i is the neuron other than the current neuron; t is the current neuron of the input feature in a single channel; w_t and b_t are the weight and bias values under the linear transformations of t and x_i, respectively; and λ is a constant, which is usually taken to be 1E-4. Based on the energy function, we can calculate the energy value of each neuron, and then calculate the important weight value of each neuron.

3.3. Detection head improvements

In the whole detection process, each detection head is responsible for different scales and types of feature detection. In order to improve the network's defect detection performance for polysilicon

Optics EXPRESS

solar cell wafer images, this paper introduces a self-attention module called MobileViT_CA before the second detection head. As shown in Fig. 6, blocks of pixels are represented by blue grids and individual pixels are represented by gray grids. Each of these blue pixels has merged the information of neighboring pixels by convolution operation, and the orange pixel uses the Transformer module to process the eight blue pixels around it, which ultimately makes the orange pixel able to encode all the pixel information in the image. The MobileViT module has the ability to learn the global information, but it is deficient in dealing with some local details. The channel and spatial attention mechanism of the coordinate attention (CA) [34] module enables the model to better learn and pay attention to the important information of different channels, which helps to improve the sensitivity to local details.

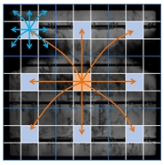


Fig. 6. Image processing of a MobileViT module.

As shown in Fig. 7, the MobileViT_CA module combines the advantages of Convolution and Transformer, and the CA module is introduced to further enhance the global information perception and multi-scale detection accuracy. The global spatial information acquisition of the feature map is realized through the convolution and attention mechanism, and the network performance is improved through channel fusion and shortcut splicing. The MobileViT_CA module makes full use of the global information learning and the channel spatial attention mechanism, which helps to achieve a balance between global perception and local details before the second detection head, and improves the overall image background and defect understanding.

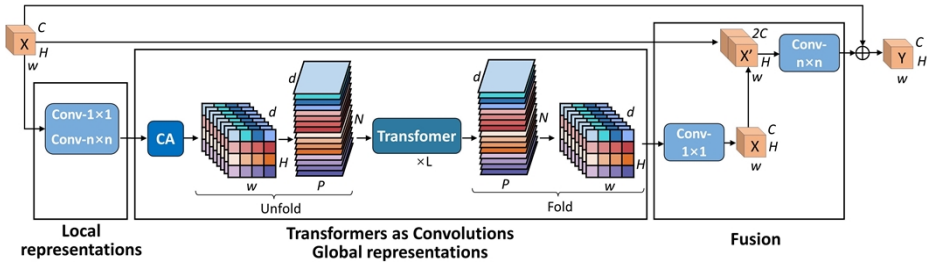


Fig. 7. Structure diagram of MobileViT_CA module.

The CA functions as a lightweight hybrid attention mechanism that simultaneously focuses on channel information and spatial information. In the task of solar cell defect detection, this local feature enhancement enables the network to focus on the target-related feature channels, which is helpful to obtain the defect details on the cell more accurately. The structure of CA is shown in Fig. 8.

The first step embeds the coordinate information into the channel attention. As can be seen from the structure diagram, the input feature maps are first average pooled along the height direction and width direction, respectively, with the expression:

$$z_c^h(h) = \frac{1}{W}\sum_{i=0}^{W-1} x_c(h,i) \tag{19}$$

Optics EXPRESS

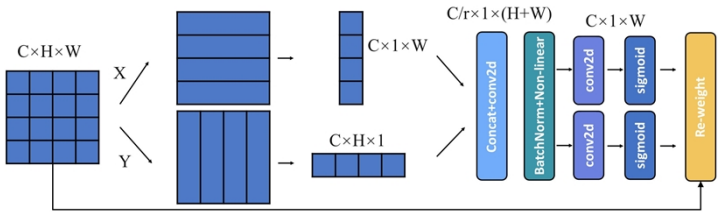


Fig. 8. Structure diagram of CA module.

$$z_c^w(w) = \frac{1}{H}\sum_{j=0}^{H-1} x_c(j,w) \tag{20}$$

In the second step, the two outputs obtained are spliced along the channel direction, and then the spliced feature F1 is sequentially subjected to convolution, regularization and nonlinear activation operations with the expression:

$$f = \delta(F_1([z^h, z^w])). \tag{21}$$

Then the feature map f is divided into two independent tensors f^h and f^w, and they are subjected to convolutional transformation and nonlinear activation, respectively, as in Eqs. (22) and (23). Where F^h and F^w are convolutional operation functions. Finally, these generated attention weights are applied to the input features to obtain the output as in Eq. (24).

$$g^h = Sigmoid(F_h(f^h)), \tag{22}$$

$$g^w = Sigmoid(F_w(f^w)), \tag{23}$$

$$y_c(i,j) = x_c(i,j) \times g_c^h(i) \times g_c^w(j). \tag{24}$$

4. Results

In this section, we conducted experiments on the selected dataset, yielding favorable results. Ablation experiments were carried out to validate the effectiveness of the proposed improvements. Furthermore, we compared our proposed method with several networks to affirm its superiority.

4.1. Datasets and experimental setup

In the production process, polysilicon ingots are first cut into wafers, then the wafers are surface-treated to form the positive and negative electrodes of the cell, and finally processed into solar cell wafers capable of converting light into electricity. The data utilized in the experiment are the images of polycrystalline silicon solar cells with four busbars (ribbon interconnects) obtained by EL imaging technology in the production process. The PVEL-AD dataset [12] contains 1 type of defect-free image and 10 different types of anomalous defects, such as crack, finger, black core, thick line, star crack, fragment, horizontal dislocation, vertical dislocation, printing error, and short circuit. To validate the effectiveness of the proposed method, 4,500 EL defect images with a resolution of 1024×1024 pixels from the above dataset are used for the experiments, which are divided into training, validation, and test sets in the ratio of 6:2:2.

We utilize CSPDarknet53 plus PANet as the base model and improve it to obtain ASDD-Net, with a model depth multiple of 0.33 and width multiple of 0.5. To validate algorithm superiority, we conduct comparative and ablation experiments on the dataset. The experimental platform for this paper is an Intel Core i9-13900K processor and a GeForce RTX3090 graphics card. For each set of experiments, we used some of the same initial training parameters. In the training phase, SGD is used as the optimization function with weight decay and momentum set to 0.0005 and 0.937. The batch size is set to 4 and Epoch is 300.

Optics EXPRESS

4.2. Performance evaluation metrics

In the experiments in this paper, we use Precision, Recall, F1-score, mean average precision (mAP), APs, and Frames Per Second (FPS) as measures of the performance of the experimental models, and the corresponding formulas are listed in Table 1. Frames per second (FPS) is used to synthesize the average inference speed of the model.

Table 1. Evaluation metrics

Name	Formula	Description
Precision	$\text{Precision} = \frac{TP}{TP + FP}$	Percentage of correct number of defective samples that the test set was predicted to be
Recall	$\text{Recall} = \frac{TP}{TP + FN}$	Ratio of the number of defective samples detected in the test set to the number of all defective samples in the predicted data set
F1-score	$\text{F1 - score} = \frac{2\times(Precision \bullet Recall)}{Precision+Recall}$	Harmonized average of precision and recall rates
AP	$\text{AP} = \int_0^1 \text{PdR}$	Area under the P-R curve for a category of defects
mAP	$\text{mAP} = \frac{\sum_{i=1}^{N} \text{AP}_i}{N}$	Average of AP values for all defect categories

4.3. Experimental results

4.3.1. Ablation study

(1) Ablation of Improvements

To demonstrate the effectiveness of the ASDD-Net algorithm in solar cell defect detection and analyze the impact of each improvement on performance, we conducted ablation experiments using consistent hyperparameters and training methods. The baseline model adopts the YOLOv5s architecture, comprising CSPDarknet53 for feature extraction, PANet for feature fusion and saliency enhancement, and YOLO Head for object detection and localization. Table 2 presents the experimental results, where √ denotes the introduction of the module.

Table 2. Results of ablation experiments

Baseline	SPD	EC2f	HAC3	MobileViT_CA	mAP50	F1-score
√					83.88%	81.03%
√	√				85.84%	83.94%
√	√	√			86.36%	85.65%
√	√	√	√		87.01%	85.24%
√	√	√	√	√	88.81%	87.88%

The results of the ablation experiments show that each introduced module has a significant impact on the model performance. First, the introduction of the SPD module performs well for the detection of small-sized defects and improves the mAP50 and F1-score of the model. The introduction of the EC2f module further enhances the model performance, which suggests that the multi-level feature fusion and multi-head self-attention mechanism for the perception of global information are beneficial for improving the detection accuracy. However, the addition of the HAC3 module resulted in some improvement of the model on mAP50, but the F1-score decreased by 0.41% after the introduction of this module. This may be due to the fact that the channel attention mechanism introduced by the HAC3 module may cause the model to focus too much on some of the channels in some cases, thus compromising the balance between accuracy and recall of the model. Finally, the introduction of the MobileViT_CA module has a positive

Optics EXPRESS

impact on improving both mAP50 and F1-score, indicating that it can effectively improve the detection of the model.

To validate the effectiveness of each module more comprehensively, we conducted heatmap experiments on the test set images for the ablation experiments, with Fig. 9 displaying the sample heatmap results. The baseline model exhibits blurring and inaccuracy in defect detection, with the defect regions not prominently highlighted. Upon introducing the SPD module, the brightness of the defect regions gradually increases in the heatmap, indicating improved attention from the model and enhanced detection. The EC2f module further accentuates the defect regions, improving detail perception. However, the HAC3 module leads to some regions becoming overly bright, suggesting an excessive focus on certain channels and affecting precision-recall balance. Finally, the MobileViT_CA module further emphasizes the defect regions, effectively enhancing detection capability.

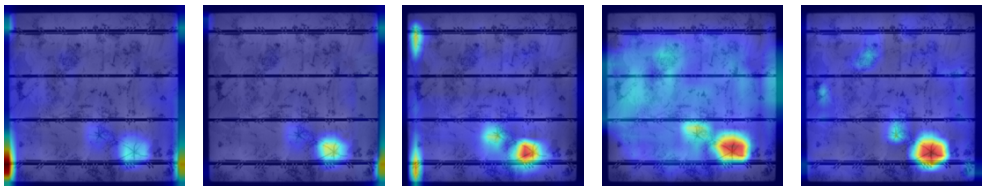


Fig. 9. Heat map of ablation experiment. (a) Baseline; (b) +SPD; (c) +SPD + EC2f; (d) +SPD + EC2f + HAC3; (e) +SPD + EC2f + HAC3+ MobileViT_CA.

In summary, through the gradual implementation of multiple improvement methods, we improved the mAP value of the ASDD-Net model from 83.88% to 88.81%, and the rise of the F1 score indicates that the model has also improved in terms of comprehensive performance. We have continuously improved the performance of the ASDD-Net model and finally achieved better defect detection results.

(2) Comparison Experiments on the Position of MobileViT_CA

The integration of the MobileViT_CA module into ASDD-Net, building upon the SPD, EC2f, and HAC3 modules, not only enhances the model's capacity to capture global information but also amplifies its discriminative capabilities through channel and spatial attention mechanisms. To investigate how the addition of the MobileViT_CA module at different positions affects the performance of solar cell defect detection, we conducted position ablation experiments. These experiments involved placing the MobileViT_CA module before the first, second, and third detection heads separately, followed by performance evaluations under various configurations. Detailed experimental data can be found in Table 3.

Table 3. Results of experiments with MobileViT_CA added at different positions

Head	Feature size	Precision	Recall	mAP50	FPS
1	$80\times80\times128$	**93.90%**	80.42%	**88.84%**	54
2	$40\times40\times256$	92.37%	**83.81%**	88.81%	**69**
3	$20\times20\times512$	91.97%	83.51%	86.94%	63

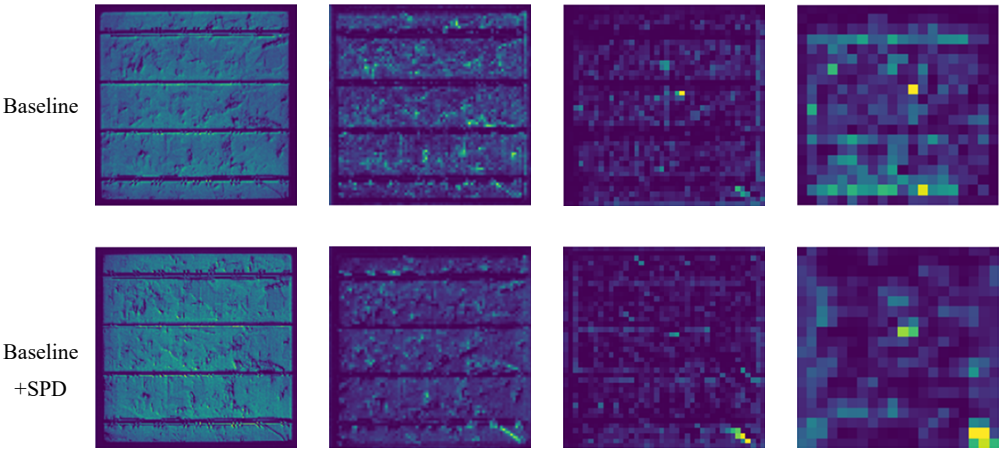
In Table 3, the first column represents the location where MobileViT_CA was added, with 1, 2, and 3 indicating placement before the first, second, and third detection heads, respectively. Analyzing the experimental data, although the mAP value of the model is slightly improved when the MobileViT_CA module is introduced before the first detection head (only 0.03% higher), its detection speed is the lowest among the three sets of experiments. The experimental results show

Optics EXPRESS

that adding MobileViT_CA module in position 2 can make the model achieve the highest recall rate and the fastest detection speed, reaching 69 frames per second.

(3) Comparative analysis of detail preservation of SPD downsampling

To more intuitively demonstrate the effect that downsampling with SPD module can keep more detailed features, we select an EL image with small cracks in the test set, and input it into the baseline model and the improved model with SPD module respectively, and then compare the feature maps generated by the two models under different downsampling layers. The results are shown in Fig. 10. After the first downsampling, both models can clearly display the details of the image, and there is no obvious difference between the feature maps. However, with further downsampling, the feature map extracted from the baseline model is gradually blurred, and the improved model with SPD module can still maintain a high definition. The feature maps extracted from the baseline model become difficult to recognize the defect details of the image after the third downsampling. The feature map of the baseline model has become completely fuzzy after the fourth downsampling, while the defect feature map generated by the improved model still retains the defect features. It is shown that SPD module can effectively preserve the details of defects when dealing with small-size defects, thus improving the detection performance of the model.



(a) First downsampling (b) Second downsampling (c) Third downsampling (d) Fourth downsampling

Fig. 10. Comparison of the feature map details of the two models.

4.3.2. Main result

ASDD-Net demonstrates precise detection of ten defect types, and its comparison with the baseline model regarding the detection accuracy of each defect type is depicted in Fig. 11. Special attention is paid to two complex and challenging types of defects in solar cells, namely, star crack and fragment defects. Star crack defects in solar cells are radial or star-shaped fractures in the silicon, with each branch of the defect being relatively narrow. These defects exhibit varying sizes and irregular shapes, rendering them particularly challenging to detect. Fragment defects refer to areas where the solar cell is broken or missing, appearing as dark regions in EL images. Distinguishing fragment defects from other dark areas in the image, such as black cores or short circuits, poses a certain challenge, especially when they are irregularly shaped. These defects can compromise the structural integrity of the solar panel and result in potential electrical issues. As observed in the figure, ASDD-Net enhances the accuracy of star crack detection by 13.2%

Optics EXPRESS

and improves the accuracy of fragment defect detection by 10.31%. For other types of defects, such as finger interruption, horizontal dislocation, vertical dislocation, printing error, ASDD-Net has also achieved different degrees of performance improvement, and the detection accuracy of most defects has exceeded 50%. The above results show that ASDD-Net can reliably detect these defects that may significantly affect the performance of solar cells, thus contributing to improving product quality and reliability.

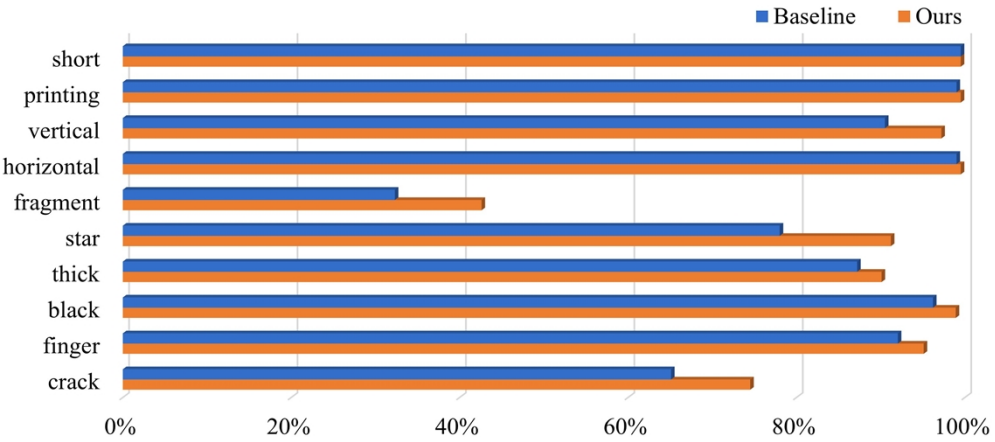


Fig. 11. Comparison of the average precision of the baseline model and the ASDD-Net model in detecting different defect categories.

We selected a representative set of solar cell defect images to compare the detection performance of the original baseline model with the ASDD-Net model. In Fig. 12(a), the detection results of the original baseline model are displayed, while Fig. 12(b) showcases the results obtained using our proposed ASDD-Net network. The results in Fig. 12(a) clearly illustrate that the baseline model is susceptible to missing or incorrectly detecting certain defects, particularly in cases where solar cell images are affected by complex backgrounds or contain multiple defect categories within a single image. In contrast, Fig. 12(b) demonstrates that our ASDD-Net model effectively mitigates the aforementioned issues and enhances detection accuracy. This clear comparison highlights that the ASDD-Net model more accurately captures the critical features of solar cell defects, thus outperforming the original model in terms of detection performance. Furthermore, ASDD-Net achieves an FPS of 69, meeting the real-time inspection requirements of photovoltaic power plants.

In practical engineering applications, the detection process confronts various challenges arising from diverse lighting conditions, shooting angles, and variations in defect characteristics. Therefore, the robustness of the model significantly influences the outgoing quality of solar cells. We randomly selected images covering ten different defect categories from the test set of the dataset and tested them using the ASDD-Net model, and the detection results are shown in Fig. 13. The experimental findings demonstrate that ASDD-Net accurately detects all ten types of defects with high precision, which proves the robustness of the model.

4.3.3. Algorithm comparison experiment

To validate the performance of the proposed improvement methods in solar cell defect detection, a comparative analysis was conducted between the algorithm presented in this paper and nine other object detection models. The experimental results, using mAP as the evaluation metric, are presented in Table 4. During the training process, samples with an IoU greater than or equal to

Optics EXPRESS

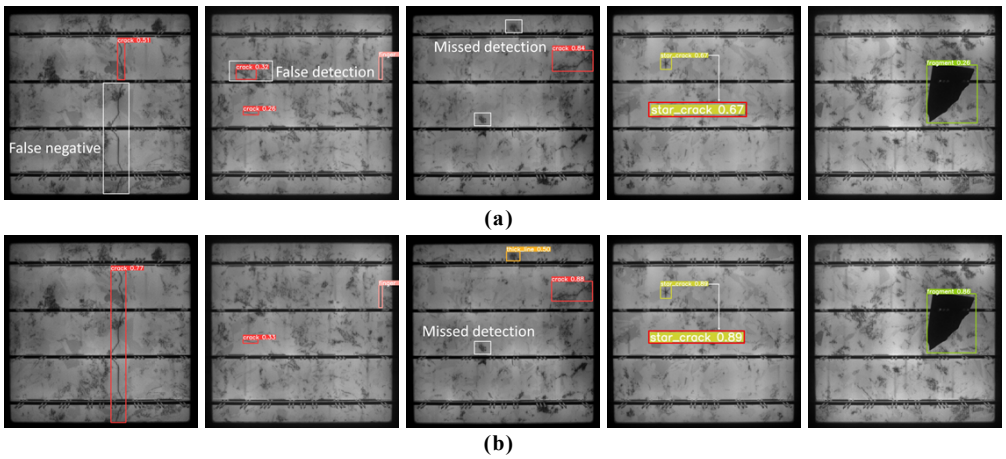


Fig. 12. ASDD-Net and the baseline model detection results comparison. (a) Baseline model assay results. (b) Our assay results.

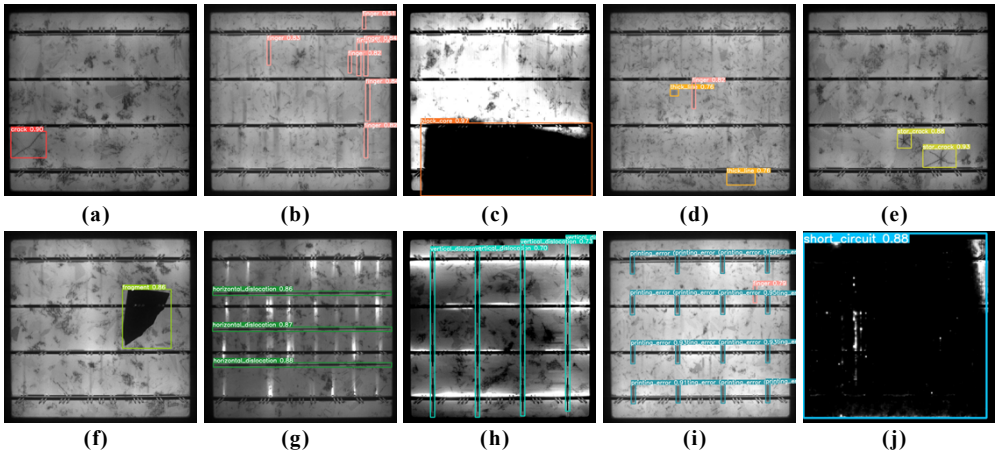


Fig. 13. The detection results of the ASDD-Net on 10 types of defects. (a)–(j) indicate crack, finger, black core, thick line, star crack, fragment, horizontal dislocation, vertical dislocation, printing error, short circuit.

0.5 were designated as positive samples, while those with an IoU less than 0.5 were marked as negative samples.

FoveaBox adopts a bottom-up approach and employs an adaptive framework to generate attention submaps of varying resolutions for the detection of defects at different scales. However, its detection results are relatively poor, possibly due to its structural emphasis on defects of different scales, which may lead to a balance issue between precision and recall. Sparse research (R-CNN) adopts the sparse attention mechanism, which is characterized by focusing only on the information in a specific location to reduce the computational complexity. However, in some cases, the sparse attention mechanism may not be able to obtain the target in complex scenes, which leads to some performance limitations in the defect detection of polycrystalline silicon solar cells. Faster R-CNN and Cascade R-CNN are both two-stage detection algorithms. First, candidate frames are generated, and then they are classified and located. The candidate frames generated in the first stage, however, may be inaccurate, which leads to their poor performance

Optics EXPRESS

Table 4. Experimental results of comparative experiments

Network types	Method	Backbone	Precision	Recall	mAP50
Other	FoveaBox	ResNet50	61.34%	62.33%	64.14%
	Sparse RCNN	ResNet50	87.82%	71.23%	73.31%
	Faster R-CNN	ResNet50	61.74%	61.42%	64.29%
	Cascade R-CNN	ResNet50	58.61%	57.12%	58.12%
YOLO	YOLOv5s	CSPDarknet53	91.30%	72.83%	83.88%
	YOLOv6s	EfficientRep	86.36%	82.12%	82.46%
	YOLOv7	CSPDarknet53	79.23%	61.00%	63.40%
	YOLOv8	CSPDarknet53	81.42%	75.13%	76.97%
	YOLOXs	EfficientRep	91.81%	73.74%	84.99%
	Ours	SEDF-Net	**92.37%**	**83.81%**	**88.81%**

in defect detection of polycrystalline silicon solar cells. In contrast, YOLOv5s and YOLOv6s are two models with better detection effect, and YOLOv6s is an improved Anchor-Free model based on YOLOv5. The overall detection effect of YOLOv7 is poor, while YOLOv8 network needs high computing resources to achieve better detection effect. YOLOXs adopts a single-stage detection structure, which performs well in Precision and mAP, but the recall rate is relatively low. It may be because the Anchor-Free mode is prone to positioning errors, especially in the case of small-size defects.

To sum up, unlike other models that may struggle with complex scenes or require significant computational resources, the proposed algorithm demonstrates robust performance across metrics. The improved algorithm's superior performance can be attributed to its comprehensive consideration of target features in the solar cell defect dataset, leading to targeted enhancements that improve detection accuracy and reliability.

A qualitative comparison was conducted on the detection results of 10 target detection algorithms, as shown in Figs. 14, 15, 16, and 17, where sub-figure (e) in each figure corresponds to the detection results of the baseline model. The results in Figs. 14 and 15 indicate that the YOLO series outperforms other detection networks in detecting defects in complex backgrounds. However, rectangular boxes in the detection results of other networks within the YOLO series may cover excessively large areas, including some interference information. In contrast, the rectangular boxes in the detection results of the ASDD-Net algorithm tightly enclose the defect areas. Results in Figs. 16 and 17 demonstrate that, compared to other detection networks, ASDD-Net can identify smaller defects more effectively. In summary, the ASDD-Net algorithm exhibits excellent performance in detecting defects in polysilicon solar cell wafers, particularly for multi-scale target cracks and small target defects in complex backgrounds.

4.3.4. Generalization experiments

The main goal of the model generalization experiments is to verify the performance stability and generalization ability of the proposed ASDD-Net on different datasets. We selected two representative datasets to validate the generalization performance and retrained various models on these datasets. One is the ELPV [13] dataset containing images of defects in polysilicon and monocrystalline silicon, and the other is the NEU-DET dataset of defects on the surface of hot-rolled steel strips. Both datasets are divided into training, validation, and testing sets in a 6:2:2 ratio. The ELPV dataset was acquired by performing distortion correction, cropping and other operations on 44 different PV modules thereby generating images of solar cell modules with a size of 300×300. In this experiment, the defects are categorized into six categories: Deep crack, Dendritic micro crack, Diagonal micro crack, No fault cell, Parallel to busbar, and

Optics EXPRESS

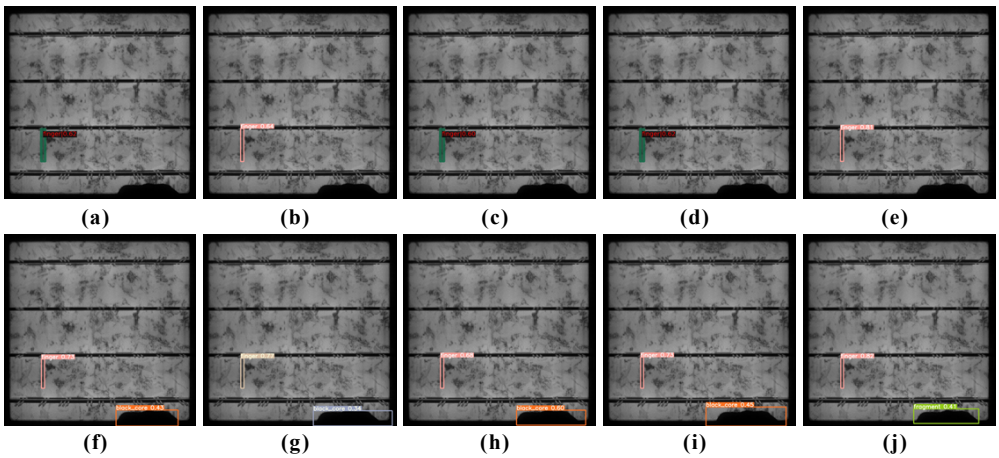


Fig. 14. Detection results in scenario 1: (a)–(j) indicate FoveaBox, Sparse RCNN, Faster R-CNN, Cascade R-CNN, YOLOv5s, YOLOv6s, YOLOv7, YOLOv8, YOLOX and ASDD-Net.

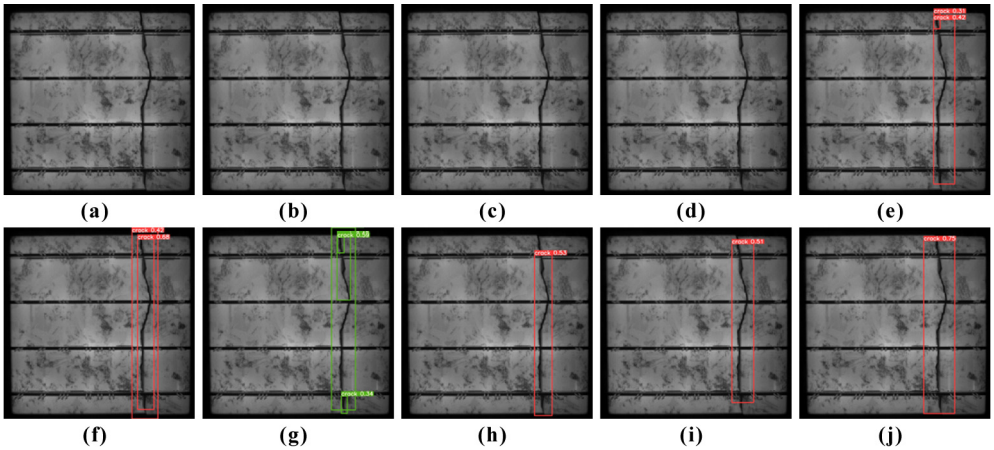


Fig. 15. Detection results in scenario 2: (a)–(j) indicate FoveaBox, Sparse RCNN, Faster R-CNN, Cascade R-CNN, YOLOv5s, YOLOv6s, YOLOv7, YOLOv8, YOLOX and ASDD-Net.

Perpendicular to busbar, and 1500 EL images are labeled with Labelimg were labeled with Labelimg. The detection results are shown in Table 5.

The publicly available NEU-DET dataset of steel surface defects include six typical steel surface defects, namely, rolled-in scale, inclusion, crazing, scratches, patch patches and pitted surface. There are a total of 1800 grayscale maps in the dataset with a resolution of 200×200px. The experimental results are shown in Table 6. This selection aims to validate the generalization performance of ASDD-Net in industrial application scenarios and to examine its adaptability to surface defects in industrial-grade materials.

Comparison tests on these two datasets show that ASDD-Net performs well in model generalization compared to other models, and ASDD-Net not only achieves optimal detection performance on different silicon material datasets such as polycrystalline silicon and monocrystalline silicon, but also demonstrates its generalization adaptability in industrial application scenarios. ASDD-Net provides a highly efficient and generalizable solution for solar cell defect

Optics EXPRESS

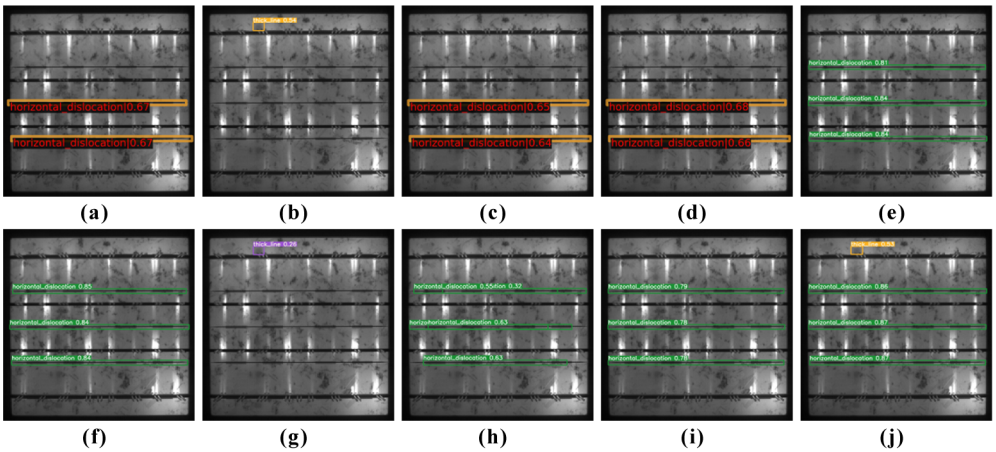


Fig. 16. Detection results in scenario 3: (a)–(j) indicates FoveaBox, Sparse RCNN, Faster R-CNN, Cascade R-CNN, YOLOv5s, YOLOv6s, YOLOv7, YOLOv8, YOLOX and ASDD-Net.

Table 5. Results of comparison experiments on the ELPV dataset

Network types	Method	Backbone	Precision	Recall	mAP50
Other	FoveaBox	ResNet50	43.62%	41.52%	42.34%
	Sparse RCNN	ResNet50	55.73%	56.32%	57.82%
	Faster R-CNN	ResNet50	65.22%	63.74%	61.58%
	Cascade R-CNN	ResNet50	66.46%	66.34%	67.93%
YOLO	YOLOv5s	CSPDarknet53	68.12%	66.30%	68.20%
	YOLOv6s	EfficientRep	67.54%	66.41%	67.63%
	YOLOv7	CSPDarknet53	**70.22%**	66.01%	69.20%
	YOLOv8	CSPDarknet53	61.22%	69.14%	68.47%
	YOLOXs	EfficientRep	66.72%	69.63%	68.17%
	Ours	SEDF-Net	69.57%	**70.52%**	**69.68%**

Table 6. Results of comparison experiments on the NEU-DET dataset

Network types	Method	Backbone	Precision	Recall	mAP50
Other	FoveaBox	ResNet50	57.84%	56.97%	57.72%
	Sparse RCNN	ResNet50	69.31%	70.60%	72.44%
	Faster R-CNN	ResNet50	71.38%	70.66%	76.31%
	Cascade R-CNN	ResNet50	72.14%	71.19%	76.48%
YOLO	YOLOv5s	CSPDarknet53	72.67%	73.52%	76.53%
	YOLOv6s	EfficientRep	73.68%	72.54%	77.27%
	YOLOv7	CSPDarknet53	68.62%	68.74%	73.96%
	YOLOv8s	CSPDarknet53	73.20%	73.27%	77.71%
	YOLOXs	EfficientRep	72.55%	73.56%	77.24%
	Ours	SEDF-Net	**73.99%**	**73.59%**	**78.26%**

Optics EXPRESS

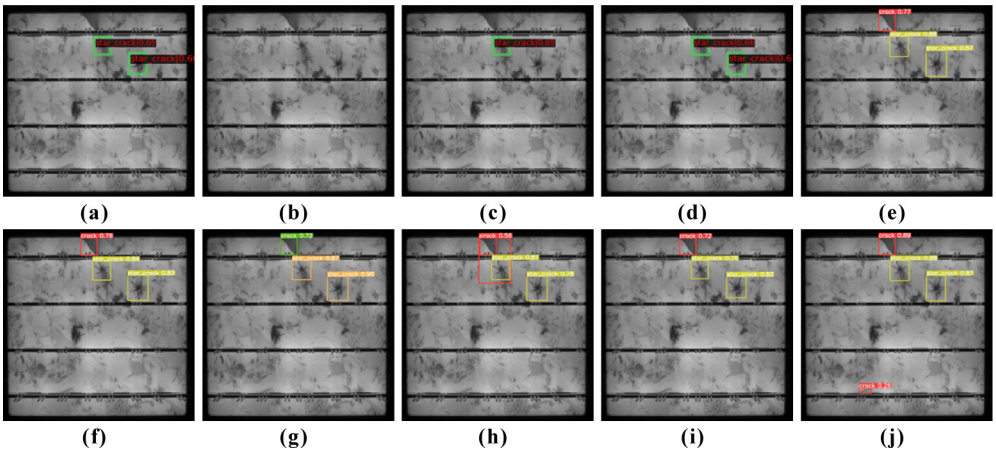


Fig. 17. Detection results in scenario 4: (a)–(j) indicates FoveaBox, Sparse RCNN, Faster R-CNN, Cascade R-CNN, YOLOv5s, YOLOv6s, YOLOv7, YOLOv8, YOLOX and ASDD-Net.

detection through its excellent performance on diverse datasets and validation experiments in real industrial scenarios.

5. Conclusions

In this paper, we propose the ASDD-Net model for the key problem in solar cell defect detection, which can be used to automatically distinguish multi-scale and small target defects in polysilicon solar cell EL images. In the feature extraction stage, we employ the SPD module to ensure the effective extraction of edge and fine-grained information. For feature fusion, we propose two feature fusion modules, EC2f and HAC3.According to the needs of network structure, EC2f and HAC3 modules are used at appropriate positions respectively to make the network more focused on key defect features, thus improving the feature extraction and fusion capability of the overall network. The application of MobileViT_CA module further balances global and local information sensing and enhances the performance of the detection head.

The experimental results show that the ASDD-Net algorithm outperforms other algorithms with a mAP value of 88.81% on the PVEL-AD dataset. And the number of detection frames per second of the model is 69, which meets the demand of real-time solar cell detection. Experiments on ELPV and NEU-DET datasets further confirm the generalizability of the proposed method in this paper to specific application scenarios and surface defects of different materials. Future research will focus on quantitatively grading the detected defects and determining whether the solar cells need to be replaced by evaluating the degree of defects on the solar cells surface. Meanwhile, areas such as small sample learning and 3D visual features will be studied in depth to further improve the accuracy and applicability of defect detection. Furthermore, we plan to explore the application of segmentation algorithms for solar cell defect detection.

Funding. National Natural Science Foundation of China (61573183).

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

Optics EXPRESS

References

1. N. Anvarhaghighi and A. Habibzadeh-Sharif, "Modified transmission line model for grating solar cells," Opt. Express **31**(10), 16315 (2023).
2. Y.-M. Wei, K. Chen, J.-N. Kang, *et al.*, "Policy and Management of Carbon Peaking and Carbon Neutrality: A Literature Review," Engineering **14**, 52–63 (2022).
3. T. Lai, B. G. Potter, and K. Simmons-Potter, "Electroluminescence image analysis of a photovoltaic module under accelerated lifecycle testing," Appl. Opt. **59**(22), G225–G233 (2020).
4. Á. H. Herraiz, A. P. Marugán, and F. Márquez, "Photovoltaic plant condition monitoring using thermal images analysis by convolutional neural network-based structure," Renewable Energy **153**, 334–348 (2020).
5. M. Waqar Akram, G. Li, Y. Jin, *et al.*, "Failures of Photovoltaic modules and their Detection: A Review," Appl. Energy **313**, 118822 (2022).
6. R. Girshick, "*Fast R-CNN*," (2015), pp. 1440–1448.
7. W. Liu, D. Anguelov, D. Erhan, *et al.*, "SSD: Single Shot MultiBox Detector," in *Computer Vision - ECCV*, B. Leibe, J. Matas, N. Sebe, and M. Welling, eds., Lecture Notes in Computer Science (Springer International Publishing, 2016), pp. 21–37.
8. J. Redmon, S. Divvala, R. Girshick, *et al.*, "You Only Look Once: Unified, Real-Time Object Detection," (2016), pp. 779–788.
9. F. Sultana, A. Sufian, P. Dutta, *et al.*, "A Review of Object Detection Models Based on Convolutional Neural Network," in *Intelligent Computing: Image Processing Based Applications*, J. K. Mandal and S. Banerjee, eds., Advances in Intelligent Systems and Computing (Springer, 2020), pp. 1–16.
10. J. Terven and D. Cordova-Esparza, "A Comprehensive Review of YOLO: From YOLOv1 and Beyond," (2023).
11. M. Hussain, "YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection," Machines **11**(7), 677 (2023).
12. B. Su, Z. Zhou, and H. Chen, "PVEL-AD: A Large-Scale Open-World Dataset for Photovoltaic Cell Anomaly Detection," IEEE Trans. Ind. Inf. **19**(1), 404–413 (2023).
13. S. Deitsch, V. Christlein, S. Berger, *et al.*, "Automatic classification of defective photovoltaic module cells in electroluminescence images," Sol. Energy **185**, 455–468 (2019).
14. R. Pierdicca, E. S. Malinverni, F. Piccinini, *et al.*, "Deep convolutional neural network for automatic detection of damaged photovoltaic cells," in *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* (Copernicus GmbH, 2018), Vol. XLII–2, pp. 893–900.
15. W. Tang, Q. Yang, K. Xiong, *et al.*, "Deep learning based automatic defect identification of photovoltaic module using electroluminescence images," Sol. Energy **201**, 453–460 (2020).
16. N. V. Sridharan and V. Sugumaran, "Convolutional Neural Network based Automatic Detection of Visible Faults in a Photovoltaic Module," Energy Sources, Part A: Recovery, Utilization, and Environmental Effects 1–16 (2021).
17. D. Korkmaz and H. Acikgoz, "An efficient fault classification method in solar photovoltaic modules using transfer learning and multi-scale convolutional neural network," Engineering Applications of Artificial Intelligence **113**, 104959 (2022).
18. B. Su, H. Chen, and Z. Zhou, "BAF-Detector: An Efficient CNN-Based Detector for Photovoltaic Cell Defect Detection," IEEE Trans. Ind. Electron. **69**(3), 3161–3171 (2022).
19. X. Zhang, T. Hou, Y. Hao, *et al.*, "Surface Defect Detection of Solar Cells Based on Multiscale Region Proposal Fusion Network," IEEE Access **9**, 62093–62101 (2021).
20. S. Xu, H. Qian, W. Shen, *et al.*, "Defect detection for PV Modules based on the improved YOLOv5s," in *China Automation Congress* (2022), pp. 1431–1436.
21. H. Chen, M. Song, Z. Zhang, *et al.*, "Detection of Surface Defects in Solar Cells by Bidirectional-Path Feature Pyramid Group-Wise Attention Detector," IEEE Trans. Instrum. Meas. **71**, 1–9 (2022).
22. Y. S. Balcıoğlu, B. Sezen, and C. Cubukcu Cerasi, "Solar Cell Busbars Surface Defect Detection Based on Deep Convolutional Neural Network," IEEE Latin Am. Trans. **21**(2), 242–250 (2023).
23. H. Han, C. Gao, Y. Zhao, *et al.*, "Polycrystalline silicon wafer defect segmentation based on deep convolutional neural networks," Pattern Recognition Letters **130**, 234–241 (2020).
24. L. Pratt, D. Govender, and R. Klein, "Defect detection and quantification in electroluminescence images of solar PV modules using U-net semantic segmentation," Renewable Energy **178**, 1211–1222 (2021).
25. M. R. U. Rahman and H. Chen, "Defects Inspection in Polycrystalline Solar Cells Electroluminescence Images Using Deep Learning," IEEE Access **8**, 40547–40558 (2020).
26. A. Sohail, N. Ul Islam, A. Ul Haq, *et al.*, "Fault detection and computation of power in PV cells under faulty conditions using deep-learning," Energy Reports **9**, 4325–4336 (2023).
27. C.-Y. Wang, H.-Y. Mark Liao, Y.-H. Wu, *et al.*, "CSPNet: A New Backbone that can Enhance Learning Capability of CNN," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (2020), pp. 1571–1580.
28. J. Yang, X. Fu, Y. Hu, *et al.*, "PanNet: A Deep Network Architecture for Pan-Sharpening," (2017), pp. 5449–5457.
29. R. Sunkara and T. Luo, "No More Strided Convolutions or Pooling: A New CNN Building Block for Low-Resolution Images and Small Objects," in *Machine Learning and Knowledge Discovery in Databases*, M.-R. Amini, eds., Lecture Notes in Computer Science (Springer Nature Switzerland, 2023), pp. 443–459.
30. N. Park and S. Kim, "How Do Vision Transformers Work?" (2022).

Optics EXPRESS

31. J. Li, X. Xia, W. Li, *et al.*, "Next-ViT: Next Generation Vision Transformer for Efficient Deployment in Realistic Industrial Scenarios," (2022).
32. Q. Wang, B. Wu, P. Zhu, *et al.*, "ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks," (2020), pp. 11534–11542.
33. L. Yang, R.-Y. Zhang, L. Li, *et al.*, "SimAM: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks," in *38th International Conference on Machine Learning* (PMLR, 2021), pp. 11863–11874.
34. Q. Hou, D. Zhou, and J. Feng, "Coordinate Attention for Efficient Mobile Network Design," 13713–13722 (2021).