Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2024.058063

ARTICLE

A Lightweight Multiscale Feature Fusion Network for Solar Cell Defect Detection

Xiaoyun Chen[1], Lanyao Zhang[1], Xiaoling Chen[1], Yigang Cen[2], Linna Zhang[1,*] and Fugui Zhang[1]

[1]School of Mechanical Engineering, Guizhou University, Guiyang, 550025, China

[2]School of Computer Science and Technology, Beijing Jiaotong University, Beijing, 100044, China

*Corresponding Author: Linna Zhang. Email: zln770808@163.com

Received: 03 September 2024 Accepted: 14 October 2024 Published: 03 January 2025

ABSTRACT

Solar cell defect detection is crucial for quality inspection in photovoltaic power generation modules. In the production process, defect samples occur infrequently and exhibit random shapes and sizes, which makes it challenging to collect defective samples. Additionally, the complex surface background of polysilicon cell wafers complicates the accurate identification and localization of defective regions. This paper proposes a novel Lightweight Multi-scale Feature Fusion network (LMFF) to address these challenges. The network comprises a feature extraction network, a multi-scale feature fusion module (MFF), and a segmentation network. Specifically, a feature extraction network is proposed to obtain multi-scale feature outputs, and a multi-scale feature fusion module (MFF) is used to fuse multi-scale feature information effectively. In order to capture finer-grained multi-scale information from the fusion features, we propose a multi-scale attention module (MSA) in the segmentation network to enhance the network's ability for small target detection. Moreover, depthwise separable convolutions are introduced to construct depthwise separable residual blocks (DSR) to reduce the model's parameter number. Finally, to validate the proposed method's defect segmentation and localization performance, we constructed three solar cell defect detection datasets: SolarCells, SolarCells-S, and PVEL-S. SolarCells and SolarCells-S are monocrystalline silicon datasets, and PVEL-S is a polycrystalline silicon dataset. Experimental results show that the IOU of our method on these three datasets can reach 68.5%, 51.0%, and 92.7%, respectively, and the F1-Score can reach 81.3%, 67.5%, and 96.2%, respectively, which surpasses other commonly used methods and verifies the effectiveness of our LMFF network.

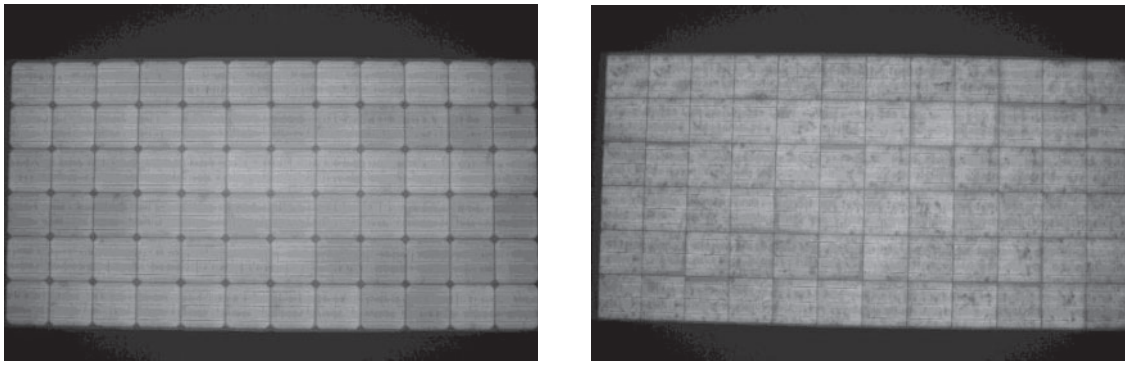
KEYWORDS

Defect segmentation; multi-scale feature fusion; multi-scale attention; depthwise separable residual block

1 Introduction

In recent years, sustainable development has been promoted, and the means of forecasting energy consumption have progressed [1]. Exploiting non-polluting, low-cost renewable energy [2] has gradually become an essential component of the energy strategy. Among these, solar energy is one of the crucial ways to optimize the energy structure with the advantages of high resource potential,

low environmental pollution, and sustainable utilization. Solar cells are one of the main components of photovoltaic power generation, and they can be divided into monocrystalline silicon and polycrystalline silicon according to the material used for production. The surface of monocrystalline silicon exhibits a uniform texture background, while the surface of polycrystalline silicon has more patches and a complex texture background, as shown in Fig. 1. Due to the brittle crystalline silicon material, its production is affected by various factors such as environmental conditions, equipment, and manufacturing processes. These factors can lead to defects, including broken grids, black spots, and cracks. These defects will directly affect the efficiency and safety of PV modules [3,4]. Therefore, defect detection on the cell surface is an essential step in the production process, which is of great significance for developing the photovoltaic power generation industry.



(a) Monocrystalline silicon panel (b) Polycrystalline silicon panel

Figure 1: Electroluminescence (EL) image of PV module cell

Traditional methods for cell defect detection primarily rely on visual inspection. However, the detection efficiency and accuracy are greatly limited by subjective factors, such as the inspector's experience, working environment, and physical condition. Additionally, the locations, shapes, and sizes of surface defects are diverse and random, which makes manual inspection labor-intensive and inefficient. Recently, machine vision and deep learning technologies have been widely adopted for defect detection across various fields, including industrial products defect detection [5–8], fabrics defect detection [9], ultrasound welding defect detection [10], and automotive glass defect detection [11]. To overcome the limitations of traditional detection methods, researchers have begun employing machine vision methods, which mainly rely on electroluminescence (EL) technology with high-resolution infrared cameras to obtain near-infrared images of the battery cell for defect detection. Li et al. [12] proposed a wavelet transform-based defect detection method for polysilicon cells, which used wavelet coefficients of single decomposition layers as features and the differences of wavelet coefficients between coefficients of consecutive layers as weights to distinguish local defects from the crystalline background. This method can effectively detect surface defects in solar cells. Anwar et al. [13] proposed an improved anisotropic diffusion filtering and image segmentation algorithm. Based on the electroluminescence intensity of solar cell micro-crack electroluminescence intensity distribution, Spataru et al. [14] proposed a two-dimensional matched filter-based microcrack detection method and obtained a binary localization map of the microcrack defect location by image post-processing.

However, most machine vision-based methods utilize traditional filtering techniques to reduce the noise in the image during the detection process. These methods often struggle to distinguish the background texture from the real defect when dealing with cell surfaces with complex textures.

Especially for defects with small sizes and complex morphology, the filtering techniques may fail to effectively separate the noise from the effective information, which may lead to misdetections and missed detections.

Deep learning has been introduced into solar cell defect detection to address these issues. Zhang et al. [15] proposed a multi-feature region proposal fusion network MF-RPN to improve the adaptability of scale variations of surface defects in solar cells. However, this network needs to extract candidate regions from different feature layers of the convolutional neural network, which significantly increases the computational cost. Pratt et al. [16] proposed a semantic segmentation model based on the U-Net architecture to locate defective regions accurately, but it performed poorly on polycrystalline silicon cells and microcrack defects. Xie et al. [17] introduced an unsupervised domain adaptive method for detecting defects in EL images of solar cells, which addressed the challenges of defect detection and data labeling on polycrystalline silicon wafers with many impurities. However, this method cannot accurately locate defective regions. While methods based on deep learning can improve detection accuracy to a certain extent, they still face limitations when applied to solar cell defect detection, including large amounts of calculation, complex models, and difficulty in small target detection.

To address the shortcomings of existing methods, this paper focuses on two aspects: lightweight and small target defect detection. For the first aspect, depthwise separable convolution is introduced to construct depthwise separable residual blocks, which replace traditional residual blocks to achieve a lightweight network. For the second aspect, a multi-scale feature fusion module is designed to capture deep and shallow features to improve the network's image representation. Additionally, a multi-scale attention module is employed to focus on finer-grained features for detecting and localizing small target defects. Consequently, the overall architecture of our network is compact and efficient, with the ability to detect all types of defects in monocrystalline and polycrystalline silicon cells. Notably, due to the lack of publicly available solar cell wafer defect detection datasets suitable for training semantic segmentation networks, this paper introduces three segmentation datasets with fine-grained defect labels: two for monocrystalline silicon and one for polycrystalline silicon. These datasets are detailed in Section 4.1.

In summary, the main contributions of this paper are:

(1) Because there are few publicly available solar panel defect detection datasets, three solar cell datasets with refined defect labels are proposed to provide a benchmark for subsequent research on segmentation networks, i.e., SolarCells, SolarCells-S, and PVEL-S. SolarCells and SolarCells-S are monocrystalline silicon panel datasets, while PVEL-S is a polycrystalline silicon panel dataset. The datasets are available on Kaggle: https://www.kaggle.com/datasets/xiaoyunchen666/dataset-of-solar-cells-defect-segmentation, accessed on 25 November 2023.

(2) An end-to-end lightweight solar cell defect detection segmentation network LMFF is proposed. Depthwise separable convolution is used to construct depthwise separable residual blocks to reduce the number of network parameters.

(3) A multi-scale feature fusion module is proposed to fuse the multi-scale features extracted by a feature extraction network. Then, it is combined with the proposed multi-scale attention module to give the network the ability of minor defects segmentation. Experimental results on three datasets verified that our proposed LMFF network achieved high accuracy in solar cell defect detection and localization.

2 Related Work

2.1 Methods Based on Image Classification

Classification-based methods typically divide the input image into blocks with overlap and then classify each block. If an image block contains a certain number of defective pixels, it is marked as defective block. The image block size is determined by the input image size of the deep learning model, with common sizes being 256 × 256, 128 × 128, or 32 × 32. Balzategui et al. [18] proposed a convolutional neural network-based method for detecting surface defects in polysilicon solar cells. This method divided the solar cell EL image into multiple regions by a sliding window, and then a CNN network is used to classify each region as a defective or non-defective region. Additionally, Cha et al. [19] proposed a CNN-based concrete crack detection method to achieve the division of 256 × 256 slab images into defective and non-defective categories. Li et al. [20] fine-tuned the GoogleNet architecture and designed a CNN for crack detection, which achieved binary image classification. However, these methods can only detect defects at the image level, and cannot locate defects at the pixel level.

2.2 Methods Based on Object Detection

Object detection is a basic task in computer vision, which aims to locate defect objects in images and determine their categories. Current object detection methods often build upon classical network architectures such as Faster R-CNN [21], SSD [22], and YOLO [23]. For the defect detection task, defects in the image can be treated as objects such that object detection networks can be used. Zhang et al. [24] proposed a model that fuses Faster R-CNN and R-FCN. In order to address the issue of high false negative rates, this model fused the detection results of two different models and adopted a multi-scale strategy to adjust the region proposal network. In addition, a hard-to-score negative sample mining strategy was used to solve the problem of excessive negative sample space prevalent in solar cell defect images. Similarly, Xu et al. [25] enhanced the Faster R-CNN framework by integrating feature cascade, Adversarial Spatial Dropout Network (ASDN), Soft-NMS, and data augmentation to improve recognition accuracy for road-base defects. Li et al. [26] proposed a surface defect detection method Mobilenet-SSD based on the SSD network combined with MobileNet, which achieved the detection of surface defects such as wear, depressions, and burrs on containers. Although methods based on object detection can achieve good performance, they usually require extensive labeled data for training and can only provide coarse localization of defect targets, which significantly limits their applications in industrial defect detection.

2.3 Methods Based on Image Segmentation

Compared with image classification and object detection methods, segmentation-based approaches can classify each pixel in an image and provide a more accurate localization result. This capability has made segmentation-based approaches widely used in microcrack defect detection. For instance, Yang et al. [27] proposed a crack detection method based on Fully Convolutional Networks (FCN) and achieved good results on pavement and concrete wall image datasets. However, this network could not accurately detect microcracks and cracks near image boundaries. Zhang et al. [28] proposed a pavement crack detection method by combining U-Net and GAN, which can solve the issue of large negative sample space by inputting larger crack images into an asymmetric U-shaped generator. To further improve defect localization, Pratt et al. [16] proposed a semantic segmentation model based on the U-Net architecture. However, this model performed poorly on polysilicon cells and microcrack defects.

To solve the above problems, from the perspective of lightweight, this paper designs a depthwise separable residual block (DSR) based on depthwise separable convolution and residual block to reduce the model parameters, and constructs the overall framework of solar cell defect detection network LMFF based on DSR. Given that features at different scales contain rich image information, this paper uses the proposed feature extraction network to extract multi-scale features. The multi-scale feature fusion module (MFF) is then utilized to fuse the multi-scale features, which makes the fusion features have both shallow texture features and deep semantic features. Additionally, a multi-scale attention module (MSA) is designed in the segmentation network to capture fine-grained key information to improve the detection ability of various defect targets. Compared with the existing models such as U-Net, LMFF is more lightweight, and significantly improves the accuracy and robustness of defect detection through finer multi-scale feature fusion and attention module.

3 Method

This section details the core principles of LMFF. The overall structure of LMFF is illustrated in Fig. 2, which comprises a feature extraction network, a multi-scale feature fusion module (MFF), and a segmentation network. Specifically, the feature extraction network first captures multi-scale features, and the MFF module fuses multi-scale features to obtain multi-scale fusion features. The multi-scale attention (MSA) module in the segmentation network focuses on fine-grained information at various scales, which enables the detection of small target defects. Finally, the segmentation network's output feature map is upsampled to match the resolution of the original image, and the target loss is used to achieve accurate location results of defect areas.

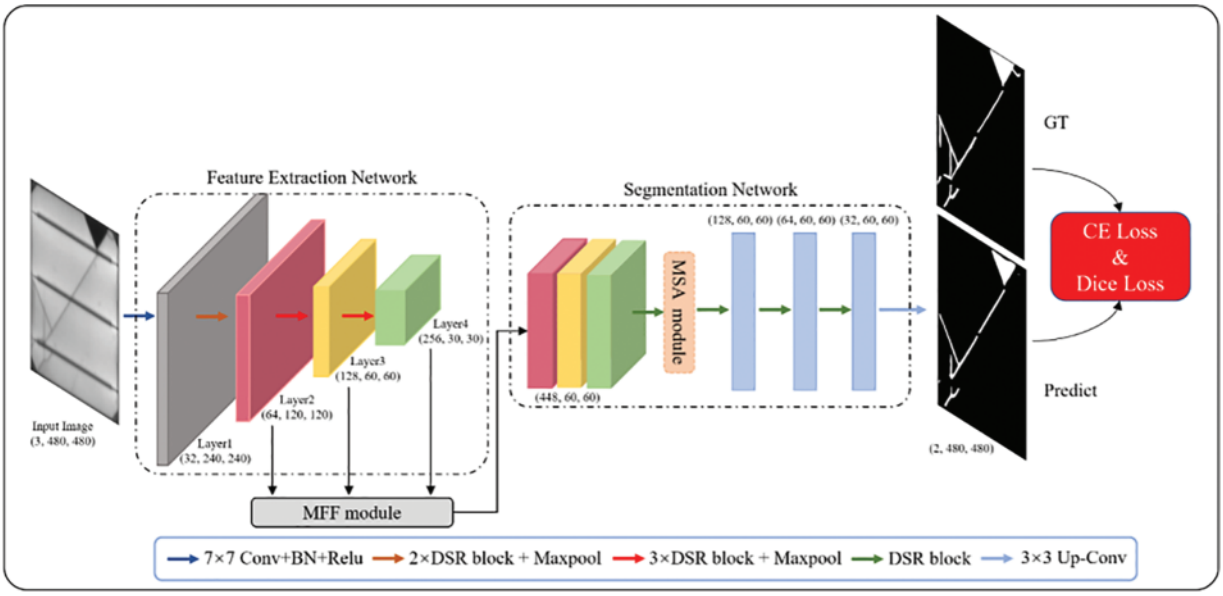


Figure 2: LMFF network structure

3.1 Depthwise Separable Residual Block

In order to make the network more lightweight, we introduce depthwise separable convolution to construct depthwise separable residual blocks. Fig. 3 shows the difference between conventional convolution and depthwise separable convolution. In a conventional convolution operation, when a

three-channel RGB image with a resolution of 5 × 5 (shape of 5 × 5 × 3) is input, after the convolution operation of four 3 × 3 × 3 convolution kernels, the final output is a feature map with a shape of 3 × 3 × 4. Thus, this conventional convolution operation involves 108 parameters (3 × 3 × 3 × 4) and 972 computations (3 × 3 × 3 × 3 × 3 × 4).

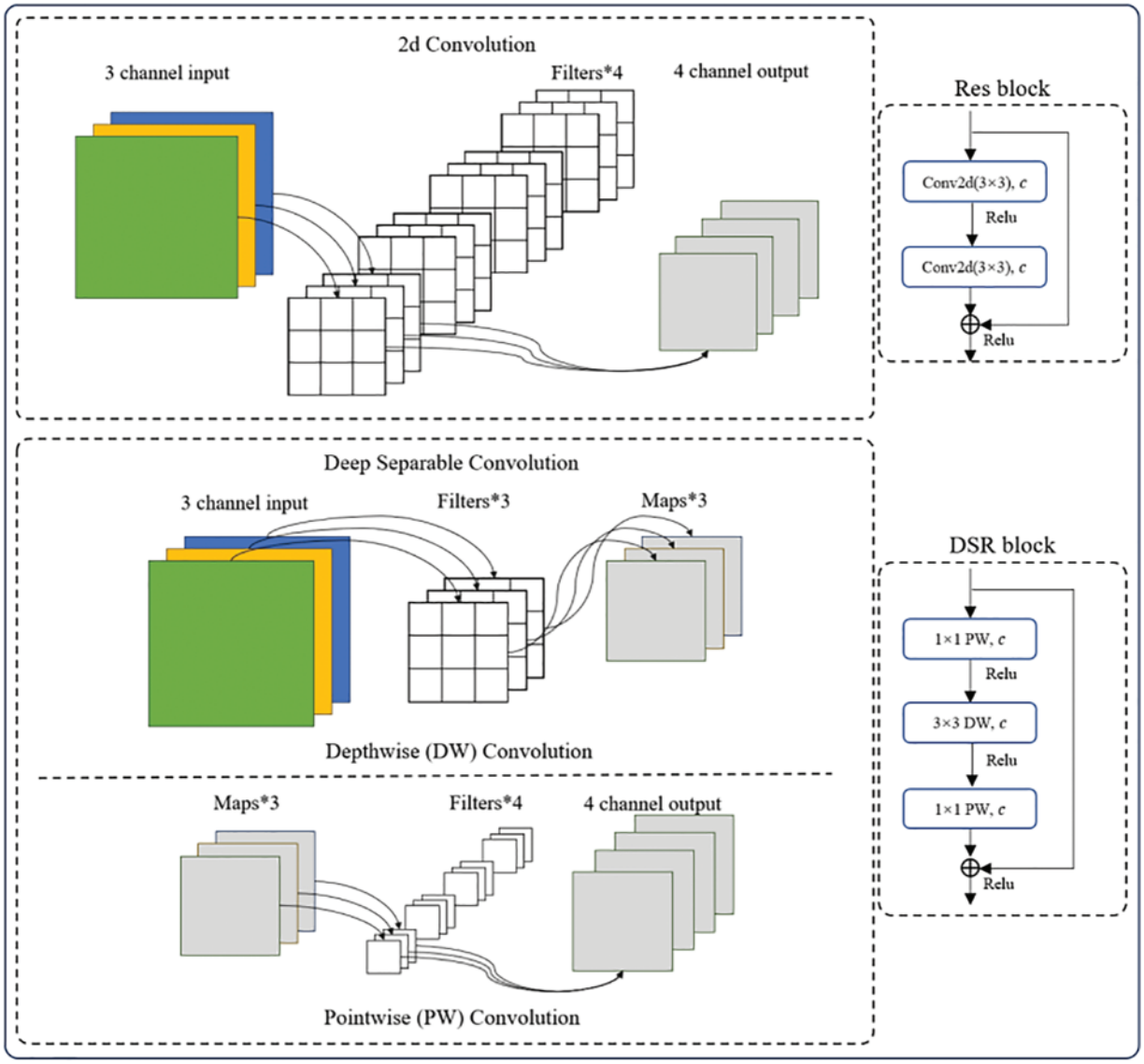


Figure 3: Example of conventional convolution and depthwise separable convolution operations

The depthwise separable convolution operation can be decomposed into two steps: Depthwise (DW) convolution and Pointwise (PW) convolution. In DW convolution, each convolution kernel is responsible for only one channel and the convolution operation is performed on only one channel. Consequently, the number of feature maps generated by DW convolution equals the number of input channels, and the feature map will not be extended. However, DW convolution only operates independently on each channel of the input, which fails to fully utilize the feature information at the same spatial location across different channels. To address this problem, PW convolution with a 1 × 1

convolution kernel is usually performed after DW convolution to fully mix the feature information of different channels. Similar to the conventional 2D convolution, PW convolution combines the feature information from different channels at the same spatial location of the input feature map by weighting to increase the nonlinearity of the output feature. For a three-channel RGB image with a resolution of 5×5 (shape of $5 \times 5 \times 3$), in order to obtain an output of the same shape as conventional 2D convolution, three DW convolutions of $3 \times 3 \times 1$ are first used for convolution calculation, and obtain three feature maps of shape $3 \times 3 \times 1$. These feature maps are then convolved by four $1 \times 1 \times 3$ PW convolutions, and the final output is four feature maps with the shape of $3 \times 3 \times 1$. Therefore, the number of parameters for the depthwise separable convolution in this example is 39 ($3 \times 3 \times 3 + 1 \times 1 \times 3 \times 4$), and the total computation is 351 ($3 \times 3 \times 3 \times 3 \times 3 + 1 \times 1 \times 3 \times 3 \times 3 \times 4$).

The computation and number of parameters for depthwise separable convolution are about one-third of that of conventional 2D convolution. Therefore, this paper introduces depthwise separable convolutions into the residual module to further reduce network parameters. The conventional residual module consists of two 3×3 convolution layers. In this paper, the first 3×3 convolution is replaced by a PW convolution to adjust the number of channels. Following the depthwise separable convolution approach, the second 3×3 convolution is replaced with a combination of DW and PW convolutions to decrease the module's parameter count. The structure of the depthwise separable residual block (DSR) is depicted in the lower right part of Fig. 3.

3.2 Feature Extraction Network

As shown in Fig. 2, the feature extraction network comprises four layers, which can extract the feature information of different levels of photovoltaic cells, respectively. Considered that the size of the output feature matrix of the first layer is $32 \times 240 \times 240$, the feature information contains relative simple features and excessive image detail information. Meanwhile, the network model directly uses the output features of all layers of the feature extraction network to train parameters, which will significantly increase the computation and reduce the detection speed. Therefore, the first layer's information is discarded, and the output features of the last three layers are selected for network model training to alleviate the problem of information redundancy. Specifically, for any input image $x \in R^{w \times h \times c}$, the feature extraction network extracts the feature matrices y_i from the last three layers (layer2, layer3, layer4), as expressed below:

$$f\left(x; \theta_f\right) = y_i \, (i = 2, 3, 4), \tag{1}$$

where θ_f denotes the parameters of the feature extraction network, which need to be optimized during training. $y_i \in R^{w_i \times h_i \times c_i} (i = 2, 3, 4)$ denotes the i-th layer feature matrix.

3.3 Multi-Scale Feature Fusion Module

The different morphologies and sizes of solar cell defect areas increase the difficulty of defect detection. Therefore, we propose the multi-scale feature fusion module (MFF) to fuse the three layers of features extracted by the feature extraction network, which enables the multi-scale fusion features with different receptive fields. The overall structure of MFF is shown in Fig. 4.

Firstly, considering that different feature levels have different receptive fields, shallow features retain more image detail information, and deep features can capture high-level semantic information, so the MFF module uses the features of the last three layers, i.e., y_2, y_3 and y_4 obtained from the feature extraction network for fusion. In order to ensure that different layers have the same resolution for subsequent feature fusion, the resolutions of y_3 and y_4 are uniformly adjusted to a same size $w_2 \times h_2$

as well as the second layer feature y_2. The number of channels c_2, c_3 and c_4 of each layer feature remains constant. The Unfold operation is then used to halve the resolution of each layer feature and expand a dimension of size 4 to increase the perception of local context features. Further, in order to avoid increasing too much computational cost, the number of channels is restored by averaging after the Unfold operation along the amplified dimension. Finally, concatenate operation is performed on the adjusted features of each layer along the channel dimension to obtain the multi-scale fusion feature $\hat{y}$. The above process can be expressed as follows:

$$y_i^{'} = reszie\,(y_i\,(i = 2,\ 3,\ 4)), \tag{2}$$

$$y_i^{''} = Unfold\left(y_i^{'} \in R^{w_2 \times h_2 \times c_i}\,(i = 2,\ 3,\ 4)\right), \tag{3}$$

$$y_i^{'''} = mean\left(y_i^{''} \in R^{w_3 \times h_3 \times c_i \times 4}\,(i = 2,\ 3,\ 4)\right), \tag{4}$$

$$\hat{y} = Cat\left(y_i^{'''} \in R^{w_3 \times h_3 \times c_i}\,(i = 2,\ 3,\ 4)\right), \tag{5}$$

where $\hat{y} \in R^{w_3 \times h_3 \times (c_2 + c_3 + c_4)}$. The $\hat{y}$ obtained by the above steps fuses the features of different receptive fields and forms a dense multi-scale region representation of the input image x. This dense multi-scale feature representation can better capture the defect information of different scales and sizes in the image, which can improve the localization ability of the overall model.

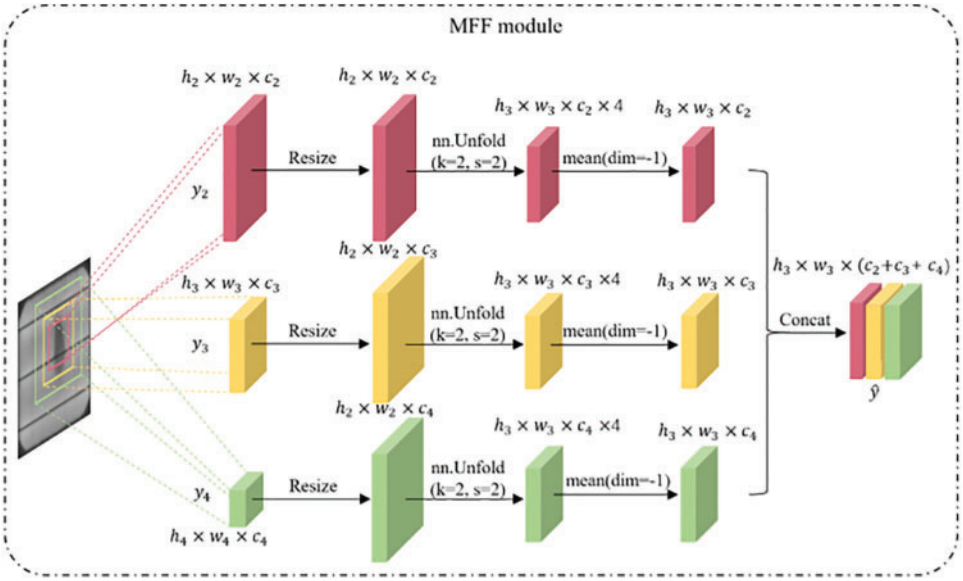


Figure 4: Multi-scale feature fusion module MFF

3.4 Multi-Scale Attention Module

As shown in Fig. 2, the multi-scale fusion feature $\hat{y}$ obtained by MFF module is input into a DSR block to obtain the feature representation $z \in R^{w_3 \times h_3 \times c'}$, and the number of channels is adjusted to c'. In order to enable the model to obtain fine-grained multi-scale feature information while maintaining resolution, this paper proposes a new multi-scale attention module MSA. The overall architecture of MSA refers to the design method of Pyramid Split Attention (PSA) [29]. In the MSA module, the convolution kernels of different sizes are firstly used to extract multi-scale features to enhance the perception of local small defect features and global features. To further enhance MSA ability of

small defect detection, the spatial attention mechanism is introduced to adaptively pay attention to the changes of small areas in the image space, and the channel attention mechanism is used to enhance the dependence between different channel information and key feature information. As shown in Fig. 5, MSA consists of four branches and its implementation can be divided into the following steps:

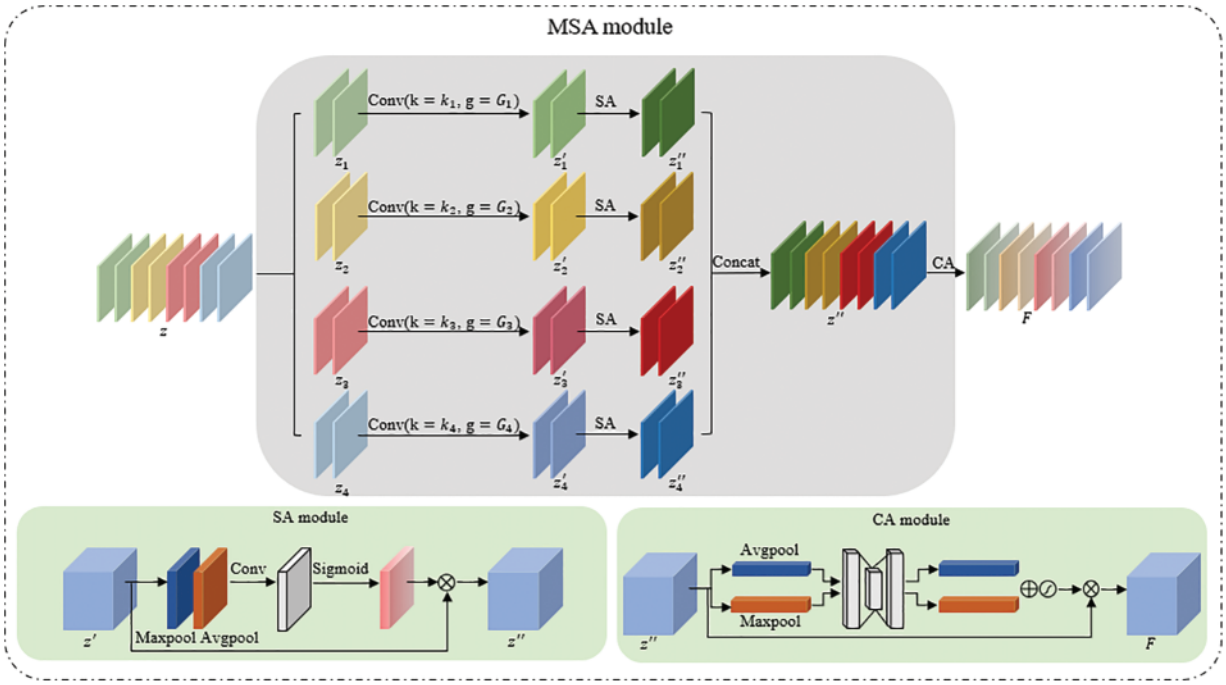


Figure 5: MSA network structure

Firstly, each branch applies convolution kernels of different sizes to extract spatial information of different scales from the input feature map z. The smaller convolution kernel can more acutely capture the local small defect features, while the larger convolution kernel captures the global context. This multi-scale convolution combination is more conducive to extract the feature information of different scales and the change relationship between local defect features and global structure. Thus, the accuracy of small defect detection can be improved. At the same time, referring to PSA, group convolution is used to solve the problem of parameter number increasing caused by the increase of convolution kernels.

$$z_i^{'} = Conv\,(k_i,\ group = G_i)\ (z)\ (i = 1,\ 2,\ 3,\ 4), \tag{6}$$

where the i-th convolution kernel $k_i = 2 \times i + 1$, the ith group is $G_i = 2^{\frac{k_i - 1}{2}}$. $z_i^{'} \in R^{w_2 \times h_2 \times c'/4}$ denotes the feature maps of different scales with the number of channels $c'/4$. It should be noted that the group size $G = 1$ when the convolution kernel $k = 3$.

Secondly, in order to further obtain important information and fine-grained features in the feature space of each branch, a spatial attention mechanism (SA) is introduced to focus on the spatial information on each branch independently. Firstly, global average pooling and maximum pooling along the channel dimension are performed on the input $z_i^{'}$. The results are concatenated along the channel dimension to obtain a $w_2 \times h_2 \times 2$ feature map. After a 7×7 convolution kernel, the number of channels is reduced to 1, followed by the activation function $Sigmoid()$ to obtain the spatial attention

feature map. This feature map is multiplied by the input $z_i^{'}$ to obtain the ultimate output. The specific process can be represented as follows:

$$z_i^{''} = \sigma(Conv(k=7)([Avgpool(z_i^{'}); Maxpool(z_i^{'})]) \times z_i^{'}, \tag{7}$$

Thirdly, the outputs of the four branches are concatenated along the channel dimension. Then the global average pooling (AvgPool) and global maximum pooling (MaxPool) are performed on the concatenated features along the spatial dimension by the channel attention mechanism (CA). Then, the results are processed by the shared fully connected layer (SFC) and activated by the *Sigmoid*() function to obtain the weights of the channels. Finally, the weight vectors are weighted to the feature layers channel by channel.

$$z'' = Cat\left(\left[z_1'' + z_2'' + z_3'' + z_4''\right]\right), \tag{8}$$

$$F = \sigma\left(SFC\left(Avgpool\left(z''\right)\right) + SFC\left(Maxpool\left(z''\right)\right)\right) \times z''. \tag{9}$$

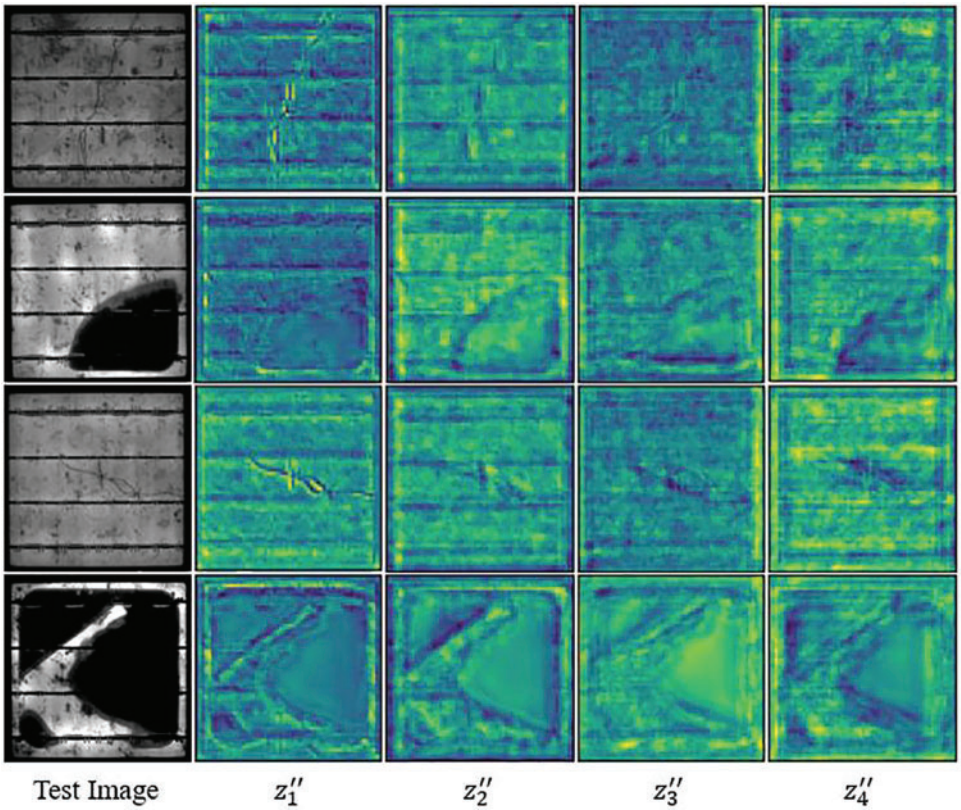


Figure 6: Visualization of the output features of each branch inside the MSA network

As illustrated in Fig. 6, we visualize the four branch outputs z_1'', z_2'', z_3'', z_4'' of the MSA module. The first branch employs a small convolution kernel and receptive field, which can obtain more fine-grained feature output. In the last three branches, with the increase of the convolution kernel size, the receptive field becomes larger and the sampling becomes sparser, which can obtain more abstract feature information and capture the overall trend. Therefore, the multi-scale attention module can

extract richer and more comprehensive feature information. The final output feature F not only contains general global information, but also the detail local information, which enables LMFF to capture various defects in solar cells.

3.5 Target Loss

In the real practice, a common problem is that the negative sample (defect-free images) number is far greater than the positive sample (defective images) number. In order to make the predicted results of LMFF closer to the actual ground truth, this paper jointly trains the network by using the CE loss and Dice loss [30]. Dice loss usually pays more attention to the mining foreground regions, which can well alleviate the problem of too large negative sample space in the image, but there will be loss saturation problems. CE loss can calculate the loss of foreground pixels and background pixels equally. But in the case of imbalanced foreground and background, CE loss alone tends to favor the background, which results in poor performance of model training. On the other hand, the joint usage of CE loss and Dice loss can compensate their both shortcomings such that the network can ensure the similarity between the segmentation results and ground truth.

Specifically, we use (10) and (11) to minimize the CE loss and Dice loss between the model-predicted segmentation results and the ground truth of the image abnormal regions.

$$L_{CE} = -\frac{1}{N}\sum_{i=1}^{N}\left(y^i \log \hat{y}^i + \left(1 - y^i\right)\log\left(1 - \hat{y}^i\right)\right), \tag{10}$$

$$L_{Dice} = \sum_{i=1}^{N}\left(1 - \frac{2\left|\hat{y}^i \bigcap y^i\right|}{\left|\hat{y}^i\right| + \left|y^i\right|}\right), \tag{11}$$

where $\hat{y}$ denotes the model prediction segmentation result. y denotes the ground truth of the image anomaly region, and N denotes the number of training images.

Therefore, the target loss for training LMFF is defined as:

$$L_{total} = \lambda_D L_{Dice} + \lambda_C L_{CE}, \tag{12}$$

where λ_D and λ_C are balanced loss hyperparameters.

4 Experiments

4.1 Datasets

To verify the effectiveness and superiority of the method in this paper, we disclose three solar cell defect detection datasets for semantic segmentation network training, named as SolarCells, SolarCells-S and PVEL-S. These datasets provide refined defect labels. The details are described as follows and the datasets are available on Kaggle: https://www.kaggle.com/datasets/xiaoyunchen666/dataset-of-solar-cells-defect-segmentation, accessed on 25 November 2023.

(1) SolarCells: The SolarCells dataset consists of 190 monocrystalline silicon cell EL defect images collected by our team from the web, with a resolution size of 448 × 448. We divide the dataset according to the ratio of 8:2 between the training set and the test set, with 152 defect images in the training set and 38 defect images in the test set.

(2) SolarCells-S: The SolarCells-S dataset comprises monocrystalline EL defect images collected by our partner company. There are only 36 defect images in total and divided into the dataset according to the ratio of 8:2 between the training set and the test set, with 28 images in the training set and 8

images in the test set. Due to the small number of samples in the dataset, the training set and test set are augmented by rotating 180°, mirroring, darkening, and brightening, respectively. Finally, the training set was increased to 144 images and the test set to 36.

(3) The PVEL-S dataset is a subset of the PVEL-AD [31] dataset, which is jointly published by Hebei University of Technology and Beijing University of Aeronautics and Astronautics for benchmarking PV cell abnormal defect detection methods. 36,543 images of polysilicon abnormal defects with 12 categories, such as cracks, broken grids, black cores, thick lines, scratches, fragments, and broken angles, are included in the PVEL-AD dataset. This dataset is mainly used for target inspection networks. We selected 1200 polysilicon defect images to form the PVEL-S dataset and accurately labeled the defective regions in the pixel-level. The labeled 1200 images were divided according to the ratio of 8:2 between the training set and the test set. The final PVEL-S dataset contains 960 defect images in the training set and 240 defect images in the test set.

4.2 Performance Evaluation Indicators

To better validate the network performance, six metrics such as intersection ratio (IOU), average category intersection ratio (MIOU), F1-Score, foreground pixel accuracy (FPA), background pixel accuracy (BPA), average category pixel accuracy (MPA), are selected in this paper to evaluate the defect segmentation performance comprehensively and reasonably.

$$IOU = \frac{TP}{TP + FP + FN}, \tag{13}$$

$$MIOU = \left(\frac{TP}{TP + FP + FN} + \frac{TN}{TN + FP + FN} \right) /2, \tag{14}$$

$$F1 - Score = \frac{2TP}{2TP + FP + FN}, \tag{15}$$

$$FPA = \frac{TP}{TP + FN}, \tag{16}$$

$$BPA = \frac{TN}{TN + FP}, \tag{17}$$

$$MPA = (FPA + BPA)/2, \tag{18}$$

where TP, FP, TN and FN stand for true positive, false positive, true negative and false negative, respectively. IOU denotes the intersection ratio between the predicted segmentation result and the defective true ground truth, which directly reflects the positioning and segmentation ability of the model. The higher IOU indicates the better the positioning performance. $MIOU$ denotes the average IOU of both foreground and background categories, which is used to measure the average detection performance of the model on different categories. *F1-Score* denotes the reconciled average of the accuracy and recall, which can comprehensively measure the effect of the two to balance the missed and false detections of the model. FPA denotes the ratio of correctly predicted foreground pixels to the total foreground pixels, which reflects the ability of the model to accurately identify non-defect regions. BPA denotes the ratio of correctly predicted background pixels to total background pixels, and the level of BPA is directly related to the reliability of the model in distinguishing background and defect. Especially in the detection task with complex background, BPA is an important performance

measure. *MPA* denotes the average pixel accuracy for both foreground and background categories, and it is used to measure the ability of the model to correctly classify at the pixel level. It focuses on the classification accuracy at the pixel level and helps analyze the performance of the model in fine-grained detection tasks.

4.3 Implementation Details

LMFF uses the feature extraction network to obtain the multi-scale feature outputs of layer2, layer3 and layer4, and resolution size of the input image x is uniformly adjusted to 480×480. The shapes of three features y_2, y_3 and y_4 are $64 \times 120 \times 120$, $128 \times 60 \times 60$ and $256 \times 30 \times 30$, respectively. y_2, y_3 and y_4 are input into the MFF module to obtain the shape of the multi-scale fusion feature $\hat{y}$ as $448 \times 60 \times 60$, and the channel c' of $\hat{y}$ is adjusted to 256 by the DSR block. The MSA module does not change the shape of the input features. The detailed setup of the segmentation network is shown in Fig. 2.

LMFF was optimized by the SGD optimizer with a learning rate of 0.01, momentum = 0.9. The batch size is set to 8. Both λ_D and λ_C are set to 1. The hardware configurations of the devices used for the testing were: Intel(R) Core(TM) i9-10900X CPU@3.70 GHz and NVIDIA GeForce RTX3080Ti.

5 Evaluation Results

In this section, LMFF is first evaluated on SolarCells and SolarCells-S monocrystalline silicon defect detection datasets, and compared with the mainstream segmentation networks FCN [32], Deeplab-v3 [33], U-Net [34] and U^2-Net [35]. Meanwhile, detailed ablation experiments are also conducted on these two monocrystalline silicon defect datasets to analyze the performance of individual modules in the proposed method. Finally, in order to further show the superiority of the proposed method, this paper also makes evaluation on the polycrystalline silicon defect detection dataset PVEL-S.

5.1 SolarCells

Table 1 shows the detection results of LMFF with other segmentation methods on the SolarCells dataset, and it can be seen that LMFF achieves the highest performance in five indicators: IOU, MIOU, F1-Score, FPA and MPA. Compared with the U^2-Net network with the best overall performance, LMFF outperforms it in all indicators, with an increase of 6.5% in IOU, 3.4% in MIOU, 4.8% in F1-Score, 3.0% in FPA, 0.2% in BPA, and 1.6% in MPA. These comparative data strongly illustrate the superiority of LMFF in detecting solar cell defects compared with other methods. Especially, the IOU is improved by 6.5%, which shows that the LMFF method has a strong advantage in accurately locating defects and can realize the segmentation of different types of defects in solar cells.

Table 1: SolarCells dataset detection results

Method	IOU	MIOU	F1-Score	FPA	BPA	MPA
FCN [32]	51.3	74.0	67.8	63.5	**98.7**	81.1
Deeplab-v3 [33]	51.1	73.7	67.6	67.4	98.2	82.8
U-Net [34]	60.6	78.8	75.5	82.2	98.0	90.1
U^2-Net [35]	62.0	79.5	76.5	84.8	97.9	91.4
LMFF (Ours)	**68.5**	**82.9**	**81.3**	**87.8**	98.1	**93.0**

In order to better show the segmentation performance of LMFF, we show the comparison of the predicted segmentation results of different methods on the SolarCells dataset, as shown in Fig. 7. The first column displays the test image, the second column represents the corresponding ground truth and the third to seventh columns represent the prediction segmentation results of different methods, respectively. Fig. 7 clearly shows that FCN and Deeplab-v3 methods exhibit serious leakage segmentation when segmenting defective images, with poor overall effect. In addition, although U-Net and U^2-Net perform better in the overall segmentation results, they still need to be improved in the segmentation details of black spots, fragments and cracks. In contrast, our method performs much better in the overall segmentation performance. Especially, it handles the defect edge details much better and smoother, and the segmentation results are more accurate.

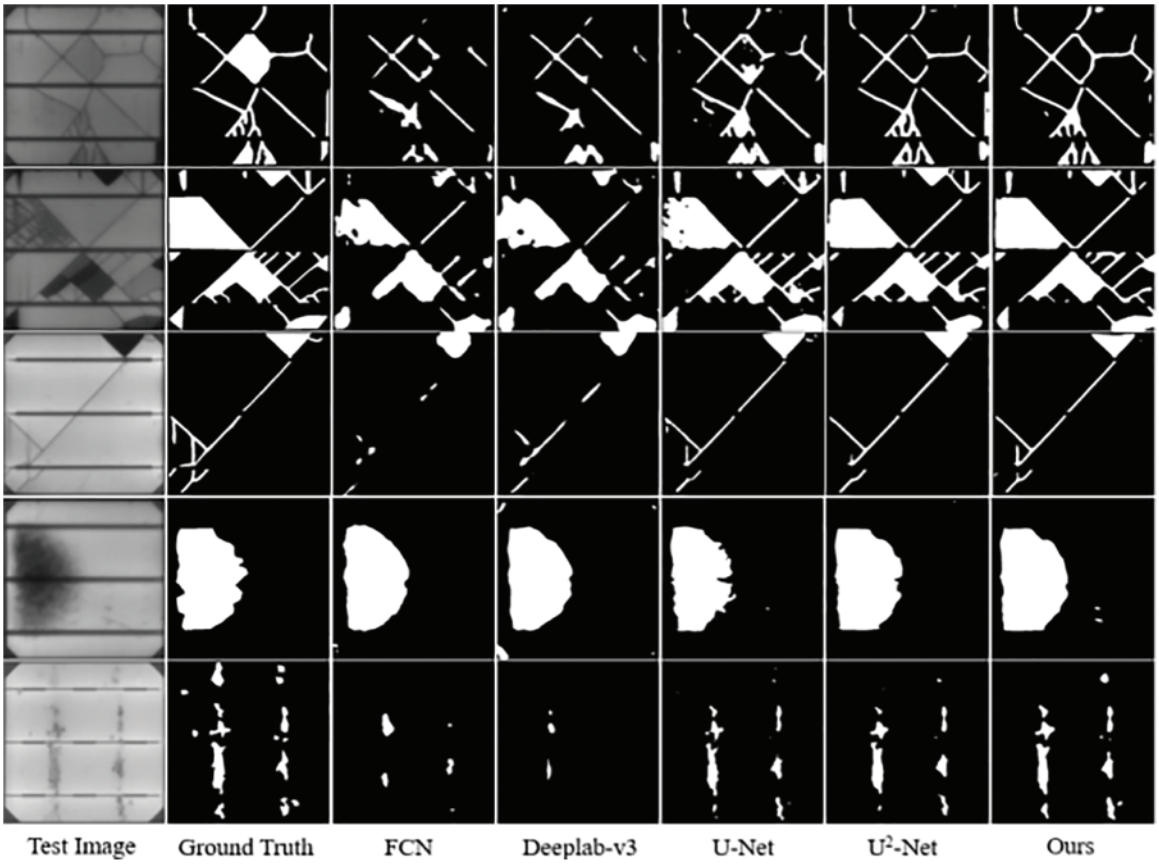


Figure 7: Comparison of segmentation results of different methods on SolarCells dataset

5.2 SolarCells-S

Table 2 shows the detection results of various method on the SolarCells-S dataset. It is evident that FCN and Deeplab-v3 perform relatively well on the BPA metric, but poor on the other five metrics. This shows that they have great limitations in overall performance, especially in terms of segmentation and detection accuracy. In contrast, the U-Net and U^2-Net methods perform relatively well on the whole, and can achieve better results on multiple indicators. Compared with the best U-Net network among the four methods, LMFF achieves 2.1% improvement in IOU, 0.7% improvement in MIOU, 1.7% improvement in F1-Score, 0.8% improvement in BPA. Although the performance on FPA

and MPA metrics of LMFF are lower than that of U-Net and U^2-Net, the significant improvement of LMFF in core indicators (IOU, MIOU, F1-Score) make up for these deficiencies, which further demonstrates the excellent performance of LMFF in detection accuracy and defect location.

Table 2: Detection results of SolarCells-S dataset

Method	IOU	MIOU	F1-Score	FPA	BPA	MPA
FCN [32]	30.5	65.7	45.7	40.6	98.8	72.0
Deeplab-v3 [33]	34.2	68.0	50.1	37.7	**99.5**	71.1
U-Net [34]	48.9	73.3	65.8	**73.6**	97.8	**86.2**
U^2-Net [35]	48.0	72.6	64.9	72.0	98.4	85.3
LMFF (Ours)	**51.0**	**74.0**	**67.5**	67.2	98.6	82.9

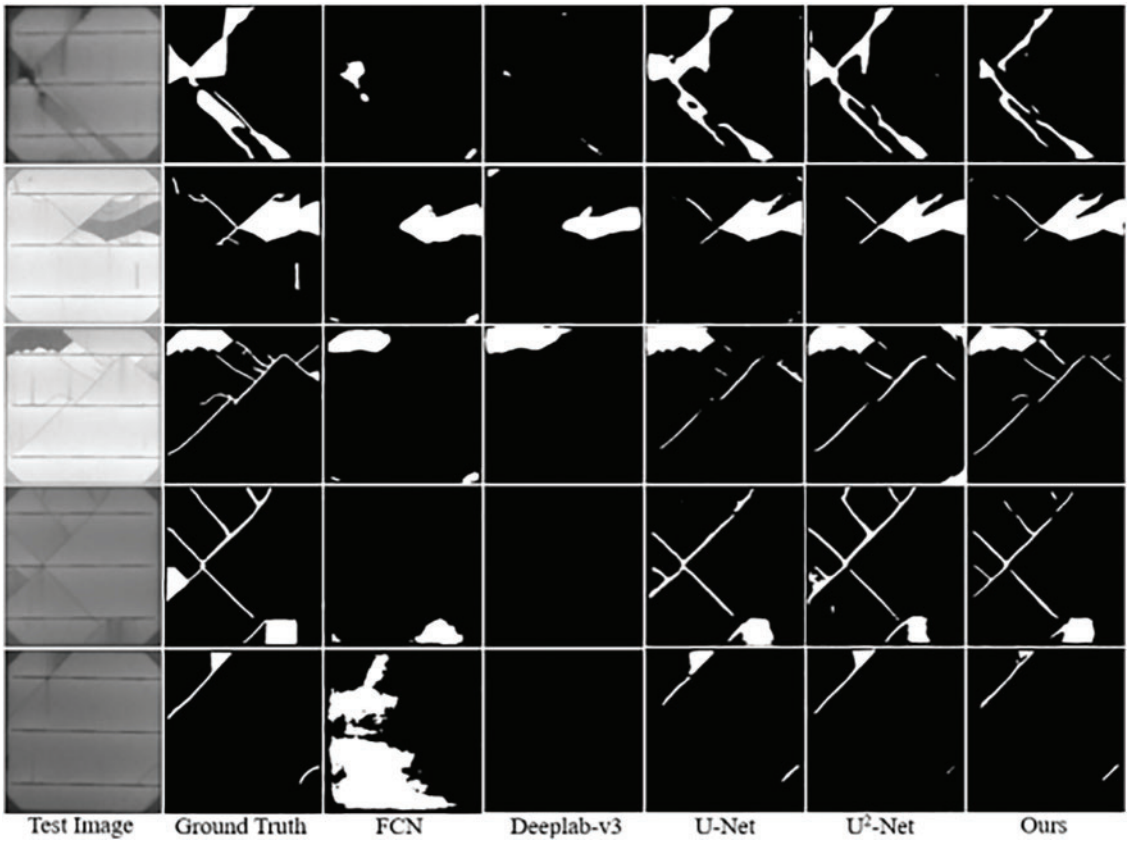


Figure 8: Comparison of segmentation results of different methods on the SolarCells-S dataset

Fig. 8 compares the predicted segmentation results of the LMFF method with other advanced methods on the SolarCells-S dataset. Compared with Fig. 7, it can be seen that the defect segmentation results of each method are lower than those of the SolarCells dataset, and the leakage segmentation phenomenon is more serious. This is due to the smaller training samples in the SolarCells-S dataset. Although the training samples are expanded by data augmentation, more samples with actual defects

are still needed. A small training dataset will increase the difficulty of network training and lead to insufficient training, which generally degrades the defect localization performance. It can be intuitively seen from Fig. 8 that the proposed method can still show relatively stable defect segmentation performance and the good superiority and robustness.

5.3 Inference Speed

In industrial applications, the network model needs to ensure excellent detection performance and efficient inference performance to meet practical demands. Therefore, we comprehensively evaluate the model's inference performance from four aspects: inference time per image, predicted frames per second (FPS), floating point operations and the number of parameters. Table 3 presents the detection results for various methods across these indicators. The LMFF model achieves an inference time of only 5.66 milliseconds per image and a prediction rate of 176.6 FPS, which demonstrates excellent real-time inference capabilities. This result also reflects that LMFF can not only respond quickly, but also maintain efficient inference speed. Additionally, LMFF shows superior performance in terms of network computation and parameter count, with values of 6.51 G and 1.18 M, respectively. Compared with the U-Net network with the best inference performance, LMFF reduces network computation by 29.1 G and the number of parameters by 3.14 M. These significant advantages make LMFF more applicable in complex industrial scenarios. Thus, LMFF not only achieves the best detection performance but also maintains high inference performance.

Table 3: Comparison of different methods in terms of inference time, predicted frames per second (FPS), floating-point operations, and number of parameters

Method	Inference time (ms)	FPS	FLOPs (G)	Param (M)
FCN [32]	13.01	76.87	122.05	32.95
Deeplab-v3 [33]	15.54	64.35	144.25	39.63
U-Net [34]	**5.28**	**189.36**	35.61	4.32
U^2-Net [35]	21.32	46.90	132.37	44
LMFF (Ours)	5.66	176.6	**6.51**	**1.18**

5.4 Ablation Study

To verify the effectiveness of the design strategy, we conducted ablation experiments on SolarCells, SolarCells-S datasets.

(1) Impact of module components: The LMFF network employs the depthwise separable residual block (DSR) as its fundamental module to minimize the number of network parameters.

It can be intuitively found from Table 4 that compared with Res blocks, the detection performance of the model by using DSR has a relatively slight decline on both datasets. We analyzed that DSR first performs independent operations on each input channel based on Depthwise (DW) convolution, and then uses Pointwise (PW) convolution to fuse feature information of different channel dimensions. Such a simple decomption-fusion approach may not be able to fully capture complex context dependencies and feature intersections, which results in performance degradation. However, it can be seen from Table 5 that DSR module performs better in inference performance compared to the Res block. In terms of inference performance, the number of parameters is reduced by 5.86 M, computational load by 35.11 G, frames per second increase by 28.53, and single inference time decreases by 1.05 ms.

Thus, despite a minor reduction in detection performance, the DSR module significantly enhances network inference performance, which makes LMFF more suitable for industrial applications.

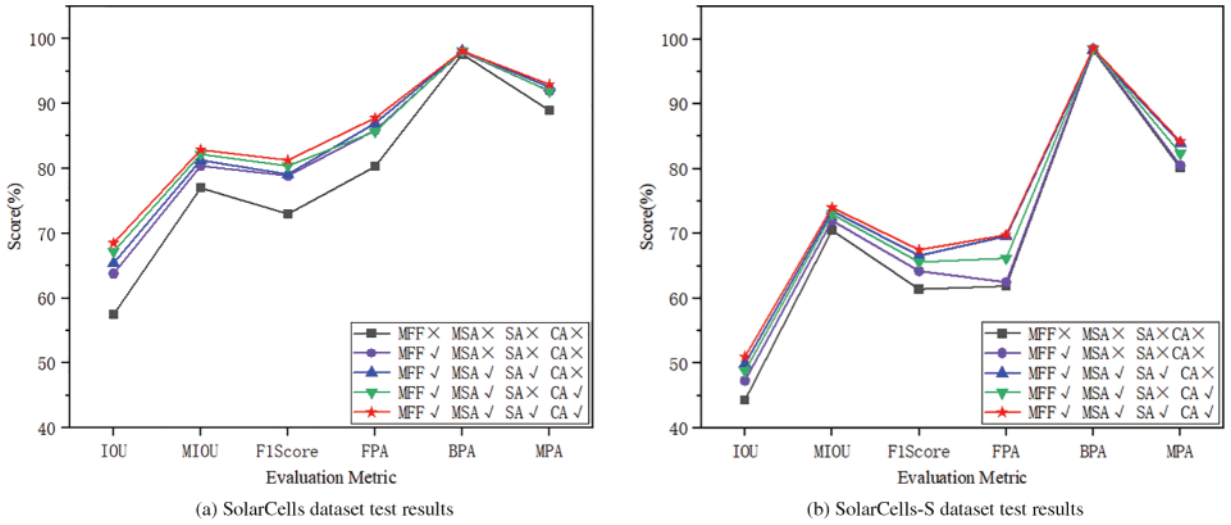
Table 4: Comparison of detection performance between conventional residual blocks and depthwise separable residual blocks

Module	SolarCells		SolarCells-S	
	IOU	MIOU	IOU	MIOU
Res block	68.9	83.2	51.8	74.4
DSR block	68.5	82.9	51.0	74.0

Table 5: Comparison of inference performance between traditional residual blocks and depthwise separable residual blocks

Module	Inference time (ms)	FPS	Flops (G)	Param (M)
Res block	6.71	148.07	41.62	7.04
DSR block	5.66	176.60	6.51	1.18

LMFF uses MFF module to fuse multi-scale features extracted by feature extraction network, which can enable the fused features to contain shallow texture information and deep abstract semantic information. The results of ablation experiments on two datasets are shown in Fig. 9. The comparison between the black and purple lines shows that the introduction of the MFF module significantly improves five metrics of the LMFF network, except for BPA, which verifies the effectiveness of the MFF module.



(a) SolarCells dataset test results (b) SolarCells-S dataset test results

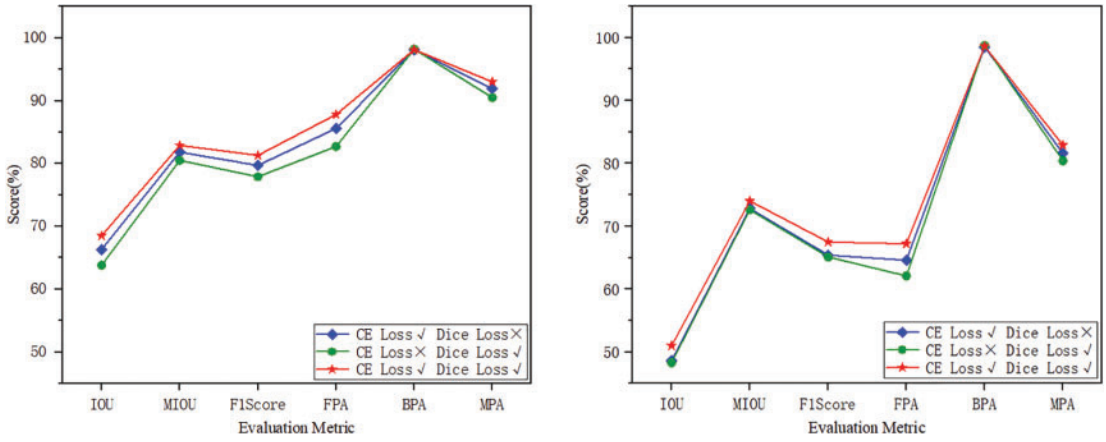
Figure 9: Detection results when using different module components on both datasets

LMFF employs MSA module to extract richer and fine-grained multi-scale information without losing the resolution of fusion features, which can improve the utilization of valuable features and

reduce the missed segmentation caused by information loss. The comparison between the red and purple lines in Fig. 9 shows that the introduction of MSA module further improves the detection results of LMFF on both datasets. Additionally, the Fig. 9 clearly shows that after the introduction of MSA module, the overall performance of LMFF in the six indicators is better than the network performance without MSA module. This confirms the effectiveness of the MSA module in enhancing overall network performance.

To further verify the effectiveness of introducing the Spatial Attention (SA) and Channel Attention (CA) mechanisms within the MSA module, we conduct a comparative analysis by comparing the purple, blue, green and red lines in Fig. 9. The results demonstrate that compared with not using any attention mechanism module, using only one attention mechanism within MSA improves the comprehensive evaluation indicators on both datasets. Furthermore, the simultaneous inclusion of both attention mechanisms significantly enhances the network's performance across all six indicators. These results show that the MSA module more effectively focuses on the dependencies between different channel information in multi-scale features and the small defect features with subtle changes in the feature space by the channel and spatial attention mechanism, so that the MSA module can efficiently grasp the feature change relationship from local to global, and improve the performance of the overall network.

(2) Impact of different losses: In this study, we employ CE loss and Dice loss to jointly train the LMFF. Fig. 10 presents the detection results when different loss functions are used on the two datasets. It can be found that the network performance is relatively poor when using CE loss or Dice loss alone for training. This is mainly due to the fact that CE loss or Dice loss each focuses on different task features and cannot make full use of their respective advantages, which results in poor performance. When the CE loss and Dice loss are used jointly, the network performance is significantly improved, which verifies the effectiveness of the joint loss strategy. In addition, Fig. 11 provides a visual analysis of network defect segmentation maps under different loss functions. When using CE loss or Dice loss independently, the network has some false detection phenomena, while using the joint loss strategy can help the network to learn better anomaly discrimination features, and make the segmentation of defective regions finer and more accurate.



(a) SolarCells dataset test results (b) SolarCells-S dataset test results

Figure 10: Detection results when using different loss functions on the two datasets

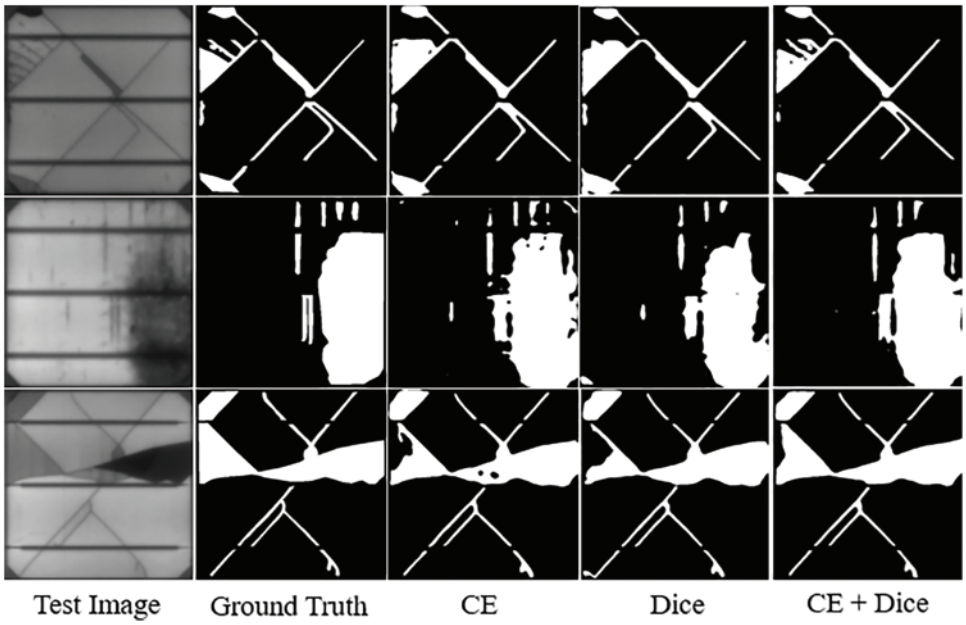


Figure 11: Effect of different loss functions on the performance of network defect segmentation

5.5 Evaluation Results on the PVEL-S Dataset

SolarCells and SolarCells-S are monocrystalline silicon defect detection datasets. The surface texture of monocrystalline silicon cells is uniform, which makes defects easier to be identified. In contrast, the complex textured background of the polycrystalline silicon cells surface contains more impurities, which significantly increases the difficulty of defect detection. Therefore, to further evaluate the defect segmentation performance of LMFF, this paper conducted benchmark tests on the PVEL-S polysilicon dataset. As shown in Table 6, LMFF achieved the best performance on the five evaluation metrics. Meanwhile, combining with the experimental results of Sections 5.1 and 5.2, we found that FCN and Deeplab-v3 performed poorly on the monocrystalline silicon dataset and better on the polycrystalline silicon dataset. Conversely, U-Net and U^2-Net performed much worse, which indicates that robustness and generalization ability of these commonly used segmentation networks are not good and do not perform consistently on different datasets. In contrast, LMFF achieves the highest detection performance on all three datasets with stable performance. This proves the proposed method's effectiveness and robustness.

Table 6: Detection results of the PVEL-S dataset

Method	IOU	MIOU	F1-Score	FPA	BPA	MPA
FCN [32]	90.5	94.3	95.0	96.2	98.8	97.5
Deeplab-v3 [33]	91.5	94.9	95.6	96.0	**99.1**	97.6
U-Net [34]	65.5	78.0	79.2	97.5	90.9	94.2
U^2-Net [35]	62.9	76.1	77.3	97.6	89.8	93.7
LMFF (Ours)	**92.7**	**95.5**	**96.2**	**98.3**	98.6	**98.5**

6 Conclusion

In this paper, we proposed a novel lightweight defect segmentation network LMFF based on the depthwise separable residual block DSR, which can detect various defects in monocrystalline silicon and polycrystalline silicon solar cells. In order to accurately locate the defect areas, we used a feature extraction network to extract the multi-scale features of the input images. Meanwhile, we further considered the different sizes of defects of solar cells and improved the network ability for small defect detection with the well-designed multi-scale feature fusion module and multi-scale attention module. In addition, since there is no publicly available dataset for defect segmentation in the field of solar cells, SolarCells, SolarCells-S, and PVEL-S, were proposed in this paper, where SolarCells and SolarCells-S are monocrystalline silicon datasets while PVEL-S is a polycrystalline silicon dataset.

In future work, our proposed method will be further validated in more scenarios, and the model needs to be further optimized according to the specific requirements of the actual scenarios to improve its performance. Moreover, designing the software-hardware system for actual production processes and deploying the LMFF on the hardware are also important issues.

Acknowledgement: The authors extend their acknowledgment to all the researchers and the reviewers who help in improving the quality of the idea, concept, and the paper overall.

Funding Statement: This work was supported in part by the National Natural Science Foundation of China under Grants 62463002, 62062021 and 62473033, in part by the Guiyang Scientific Plan Project [2023] 48–11, in part by QKHZYD [2023] 010 Guizhou Province Science and Technology Innovation Base Construction Project "Key Laboratory Construction of Intelligent Mountain Agricultural Equipment".

Author Contributions: Study conception and design: Xiaoyun Chen, Lanyao Zhang; data collection: Xiaoyun Chen, Lanyao Zhang; Xiaoling Chen, Fugui Zhang; analysis and interpretation of results: Yigang Cen; draft manuscript preparation: Xiaoyun Chen, Linna Zhang. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: All the study data are included in the article.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

[1] Z. Abbas, M. Waqas, Z. Lun, S. S. Khan, M. Amjad and S. Larkin, "Predicting energy consumption through the LEAP model based on LMDI technique along with economic analysis: A case study," *Energy Explor. Exploit.*, vol. 41, no. 6, pp. 1919–1941, 2023. doi: 10.1177/01445987231202802.

[2] Z. Abbas, M. Waqas, and S. He, "Environmental evaluation of electricity generation from landfill gas using LEAP and IPCC model: A case study of Karachi," *Energy Syst.*, vol. 145, pp. 1–22, 2023. doi: 10.1007/s12667-023-00574-3.

[3] Y. Jiang and C. Zhao, "Attention classification-and-segmentation network for micro-crack anomaly detection of photovoltaic module cells," *Sol. Energy*, vol. 238, pp. 291–304, 2022. doi: 10.1016/j.solener.2022.04.012.

[4] U. Otamendi, I. Martinez, M. Quartulli, I. G. Olaizola, E. Viles and W. Cambarau, "Segmentation of cell-level anomalies in electroluminescence images of photovoltaic modules," *Sol. Energy*, vol. 220, pp. 914–926, 2021. doi: 10.1016/j.solener.2021.03.058.

[5] Y. Chen, B. Chen, W. Xian, J. Wang, Y. Huang and M. Chen, "LGFDR: Local and global feature denoising reconstruction for unsupervised anomaly detection," *Vis. Comput.*, vol. 18, pp. 1–14, 2024. doi: 10.1007/s00371-024-03281-x.

[6] F. Zhang, S. Kan, D. Zhang, Y. Cen, L. Zhan and V. Mladenovic, "A graph model-based multiscale feature fitting method for unsupervised anomaly detection," *Pattern Recognit.*, vol. 138, 2023, Art. no. 109373. doi: 10.1016/j.patcog.2023.109373.

[7] L. Zhang, S. Kan, Y. Cen, X. Chen, L. Zhang and Y. Huang, "A normalizing flow-based bidirectional map residual network for unsupervised defect detection," *Comput. Mater. Contin.*, vol. 78, no. 2, pp. 1631–1648, 2024. doi: 10.32604/cmc.2024.046924.

[8] C. Peng, L. Zhao, S. Wang, Z. Abbas, F. Liang and M. S. Islam, "LightFlow: Lightweight unsupervised defect detection based on 2D flow," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–12, 2024. doi: 10.1109/TIM.2024.3415769.

[9] K. Wu, L. Zhu, W. Shi, and W. Wang, "Automated fabric defect detection using multi-scale fusion MemAE," *Vis. Comput.*, vol. 127, pp. 1–15, 2024. doi: 10.1007/s00371-024-03358-7.

[10] F. Liang *et al.*, "LAD-Net: A lightweight welding defect surface non-destructive detection algorithm based on the attention mechanism," *Comput. Ind.*, vol. 161, 2024, Art. no. 104109. doi: 10.1016/j.compind.2024.104109.

[11] Z. Chen, S. Huang, H. Lv, Z. Luo, and J. Liu, "Defect detection in automotive glass based on modified YOLOv5 with multi-scale feature fusion and dual lightweight strategy," *Vis. Comput.*, vol. 18, pp. 1–14, 2024. doi: 10.1007/s00371-023-03225-x.

[12] W. C. Li and D. M. Tsai, "Wavelet-based defect detection in solar wafer images with inhomogeneous texture," *Pattern Recognit.*, vol. 45, no. 2, pp. 742–756, 2012. doi: 10.1016/j.patcog.2011.07.025.

[13] S. A. Anwar and M. Z. Abdullah, "Micro-crack detection of multicrystalline solar cells featuring an improved anisotropic diffusion filter and image segmentation technique," *EURASIP J. Image Video Process.*, vol. 2014, pp. 1–17, 2014. doi: 10.1186/1687-5281-2014-15.

[14] S. Spataru, P. Hacke, and D. Sera, "Automatic detection and evaluation of solar cell micro-cracks in electroluminescence images using matched filters," in *2016 IEEE 43rd Photovoltaic Spec. Conf. (PVSC)*, Portland, OR, USA, IEEE, 2016, pp. 1602–1607.

[15] X. Zhang, T. Hou, Y. Hao, H. Shangguan, A. Wang and S. Peng, "Surface defect detection of solar cells based on multiscale region proposal fusion network," *IEEE Access*, vol. 9, pp. 62093–62101, 2021. doi: 10.1109/ACCESS.2021.3074219.

[16] L. Pratt, D. Govender, and R. Klein, "Defect detection and quantification in electroluminescence images of solar PV modules using U-net semantic segmentation," *Renew. Energy*, vol. 178, pp. 1211–1222, 2021. doi: 10.1016/j.renene.2021.06.086.

[17] X. Xie, G. Lai, M. You, J. Liang, and B. Leng, "Effective transfer learning of defect detection for photovoltaic module cells in electroluminescence images," *Sol. Energy*, vol. 250, pp. 312–323, 2023. doi: 10.1016/j.solener.2022.10.055.

[18] J. Balzategui *et al.*, "Semi-automatic quality inspection of solar cells based on convolutional neural networks," in *24th IEEE Int. Conf. Emerg. Technol. Fact. Autom. (ETFA)*, Zaragoza, Spain, IEEE, 2019, pp. 529–535.

[19] Y. J. Cha, W. Choi, and O. Büyüköztürk, "Deep learning-based crack damage detection using convolutional neural networks," *Comput. Aided Civ. Infrastruct. Eng.*, vol. 32, no. 5, pp. 361–378, 2017. doi: 10.1111/mice.12263.

[20] S. Li and X. Zhao, "Convolutional neural networks-based crack detection for real concrete surface," in *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems*. SPIE, 2018, vol. 10598, pp. 955–961.

[21] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, 2016. doi: 10.1109/TPAMI.2016.2577031.

[22] W. Liu *et al.*, "SSD: Single shot multibox detector," in *Comput. Vis.-ECCV 2016: 14th Eur. Conf.*, Amsterdam, The Netherlands, Springer International Publishing, Oct. 11–14, 2016, pp. 21–37.

[23] X. Zhu, S. Lyu, X. Wang, and Q. Zhao, "TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios," in *2021 IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Montreal, BC, Canada, 2021, pp. 2778–2788. doi: 10.1109/ICCVW54120.2021.00312.

[24] X. Zhang, Y. Hao, H. Shangguan, P. Zhang, and A. Wang, "Detection of surface defects on solar cells by fusing multi-channel convolution neural networks," *Infrared Phys. Technol.*, vol. 108, 2020, Art. no. 103334. doi: 10.1016/j.infrared.2020.103334.

[25] X. Xu, Y. Lei, and F. Yang, "Railway subgrade defect automatic recognition method based on improved faster R-CNN," *Sci. Prog.*, vol. 2018, no. 1, 2018, Art. no. 4832972. doi: 10.1155/2018/4832972.

[26] Y. Li, H. Huang, Q. Xie, L. Yao, and Q. Chen, "Research on a surface defect detection algorithm based on MobileNet-SSD," *Appl. Sci.*, vol. 8, no. 9, 2018, Art. no. 1678. doi: 10.3390/app8091678.

[27] X. Yang, H. Li, Y. Yu, X. Luo, T. Huang and X. Yang, "Automatic pixel-level crack detection and measurement using fully convolutional network," *Comput. Aided Civ. Infrastruct. Eng.*, vol. 33, no. 12, pp. 1090–1109, 2018. doi: 10.1111/mice.12412.

[28] K. Zhang, Y. Zhang, and H. D. Cheng, "CrackGAN: Pavement crack detection using partially accurate ground truths based on generative adversarial learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 2, pp. 1306–1319, 2020. doi: 10.1109/TITS.2020.2990703.

[29] H. Zhang, K. Zu, J. Lu, Y. Zou, and D. Meng, "EPSANet: An efficient pyramid squeeze attention block on convolutional neural network," in *Proc. Asian Conf. Comput. Vis.*, 2022, pp. 1161–1177.

[30] X. Li, X. Sun, Y. Meng, J. Liang, F. Wu and J. Li, "Dice loss for data-imbalanced NLP tasks," 2019, *arXiv:1911.02855*.

[31] B. Su, Z. Zhou, and H. Chen, "PVEL-AD: A large-scale open-world dataset for photovoltaic cell anomaly detection," *IEEE Trans. Ind. Inform.*, vol. 19, no. 1, pp. 404–413, 2022. doi: 10.1109/TII.2022.3162846.

[32] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.

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

[34] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Med. Image Comput. Comput.-Assist. Interv.-MICCAI 2015: 18th Int. Conf.*, Munich, Germany, Springer International Publishing, Oct. 5–9, 2015, vol. 9351, pp. 234–241.

[35] X. Qin, Z. Zhang, C. Huang, M. Dehghan, O. R. Zaiane and M. Jagersand, "U^2-Net: Going deeper with nested U-structure for salient object detection," *Pattern Recognit.*, vol. 106, 2020, Art. no. 107404. doi: 10.1016/j.patcog.2020.107404.