This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2020.3008021, IEEE Transactions on Industrial Informatics

Deep Learning-based Solar-Cell Manufacturing Defect Detection with Complementary Attention Network

Binyi Su, Haiyong Chen, Peng Chen, *Member, IEEE,* Guibin Bian, Kun Liu, Weipeng Liu

***Abstract*—The automatic defects detection for solar cell electroluminescence (EL) images is a challenging task, due to the similarity of defect features and complex background features. To address this problem, a novel Complementary Attention Network (CAN) is designed by connecting the novel channel-wise attention subnetwork with spatial attention subnetwork sequentially, which adaptively suppresses the background noise features and highlights the defect features simultaneously by employing the complementary advantage of the channel features and spatial position features. In CAN, the novel channel-wise attention subnetwork applies convolution operation to integrate the concatenated and discriminative output features extracted by Global Average Pooling (GAP) layer and Global Max Pooling (GMP) layer, which can make fully use of these informative features. Furthermore, a Region Proposal Attention Network (RPAN) is proposed by embedding CAN into Region Proposal Network (RPN) in Faster R-CNN to extract more refined defective region proposals, which is used to construct a novel end-to-end Faster RPAN-CNN framework for detecting defects in raw EL image. Finally, some experimental results on a large-scale EL dataset including 3629 images, 2129 of which are defective, show that the proposed method performs much better than other methods in terms of defects classification and detection results in raw solar cell EL images.**

***Index Terms*—automatic defects detection, solar cell, near-infrared image, attention network, region proposal network**

I. Introduction

THE multicrystalline solar cell defects lead to a seriously negative impact on the power generation efficiency. Thus, defect detection is very crucial to avoid the defective solar cell entering the next production stage during the manufacturing of solar cell. As shown in Fig. 1, finger interruption defect is caused by incomplete printing of the silver paste in visible light (VL) image. The incomplete part will be open-circuited during the energization of the solar cell, which looks like a black vertical strip in the electroluminescence (EL) near-infrared image [1].

In this study, the near-infrared EL images are captured and processed by using the following visual defect detection system, which contains four components: (a) supplying system, (b) image acquisition system, (c) image processing system, (d) sorting system. To make the readers better understand its mechanism, the appearance and internal structure of the solar cell EL near-infrared image defect detection system is shown in Fig. 2.

This work was supported in part by the National Natural Science Foundation of China under Grant 61873315, in part by the Natural Science Foundation of Hebei Province under Grant F2018202078. Paper no. TII-20-1900. (*Corresponding author: Haiyong Chen.*)

B. Su, H. Chen, P. Chen, K. Liu and W. Liu are with the School of Artificial Intelligence and Data Science, Hebei University of Technology, Tianjin 300130, China (e-mail: 18812671934@163.com; haiyong.chen@hebut.edu.cn; holmes83@163.com; liukun03@mails.thu.edu.cn; liuweipeng@hebut.edu.cn)

G. Bian are with the Institute of Automation, Chinese Academy of Sciences, Beijing 100000, China (e-mail: guibin.bian@ia.ac.cn)

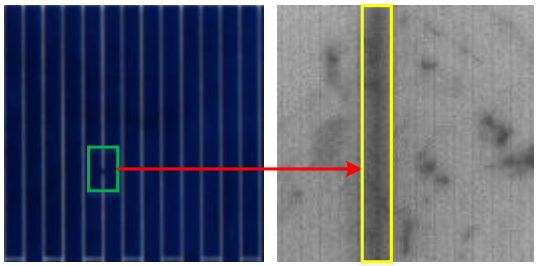


Fig. 1. Broken grid defect in visible-light (VL) image (380-780nm) corresponding to finger interruption defect in near-infrared EL image (850-1150nm).

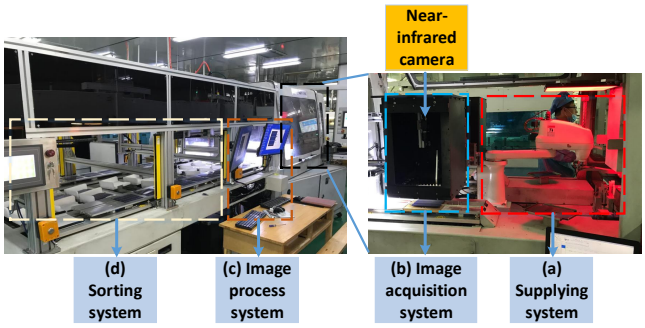


Fig. 2. The appearance and internal structure of the solar cell EL near-infrared image defect detection system.

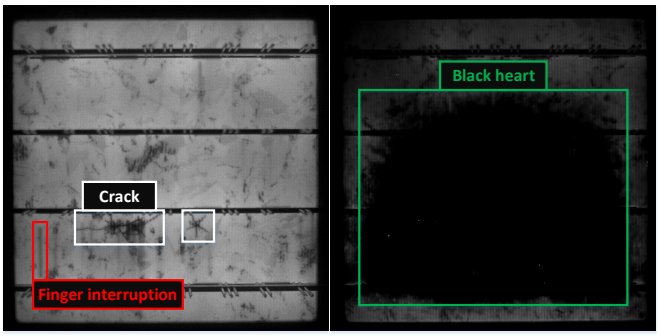


Fig. 3. An original EL near-infrared image with two crack defects in white boxes, one finger interruption defect in red box, one black core defect in green box.

Although, by observing Fig. 3, we can collect a clear EL image by image acquisition system, it can be seen that solar cell EL images show the following characteristics such as the heterogeneously complex background interference, high-resolution image, variable shape and multi-scale defect, which bring great challenges for the automatic defect detection. This paper is dedicated to detecting three types of defects with different shape features: crack, finger interruption and black core [2], which will be used to verify the generality of the

proposed method. The shapes of crack defect mainly include line-like and star-like, which show random texture distribution. The finger interruption defect presents line-like shape and vertical distribution. Black core defect presents blob-like shape and is a cluster of black region in EL image. All of these defects have a negative influence on the power generation efficiency. Thus, automatic defect detection in EL image is extremely valuable for the defective product rejection and desirable quality control during the production.

In fact, computer vision-based method can satisfy the urgent demand of quality monitoring in many fields. Conventional computer vision methods for defect recognition mainly include two aspects: filter-based method [3] and handcrafted feature-based method [4].

For the Filter-based method, textural features are commonly extracted from the spectrum of an image in spectral domain. According to the spectrum difference between defect and background in spectral domain, filter-based method can remove the background spectrum on the basis of retaining the defect spectrum, or remove the defect spectrum on the basis of retaining the background spectrum, then make the difference with the original image to achieve the defect recognition. For instance, Tsai *et al.* [5] proposed an automated visual inspection scheme for multicrystalline solar cell using the mean-shift technique, which can effectively detect finger interruption and contamination defects in the surface of solar cell. Anwar *et al.* [6] presented an improved anisotropic diffusion filter and advanced image segmentation technique for detecting micro-crack in solar cells. Tsai *et al.* [7] proposed a Fourier image reconstruction method to recognize defects such as crack and finger interruption defects in EL images. This method is effective to detect line-like and bar-like shapes defects.

Handcrafted feature-based methods usually extract texture feature from an image. Next, the feature will be used to train and test discriminant classifiers, which are applied to separate some local defects from the background. Tsai *et al.* [8] proposed an unsupervised Haar-like feature extraction and a novel fuzzy k-means algorithm for defect detection. This unsupervised method is efficient to classify the dataset into two classes: defect and non-defect image, and the classification accuracy is 98% in 50 test samples. However, it cannot get the specific category of each defect in the EL images. Subsequently, Demant *et al.* [9] introduced four selected feature descriptors (FDs) to solve crack recognition task. Su *et al.* [10] developed an algorithm for the recognition of crack and finger interruption defects in multicrystalline solar cells, which applied a novel gradient feature descriptor and SVM classifier to accomplish multi-classification of defects and good samples.

Recently, comparing with conventional methods, the deep learning based method has been more and more popular for many fields, such as defect recognition [11][12][13][14][15], fault diagnosis [16] and soft sensing of industrial processes [17][18][19]. Deep learning methods present the advantages of high accuracy, wide versatility, and strong plasticity, which can be classified into three types: autoencoder, deep belief network, and convolution neutral network (CNN). In this paper, we only focus on CNN-based deep learning method for defect recognition. CNN module such as NB-CNN [11] and MS-CNN [12] have shown outstanding performance in the industrial manufacturing defect classification tasks. The success of the deep learning has led to a gradual replacement of conventional pattern recognition for image inspection tasks. Deitscha *et al.* [13] proposed an end-to-end deep CNN to classify defects on solar cell EL images. Chen *et al.* [14] developed a novel solar-CNN architecture to classify defects on solar cell visible-light images. Han *et al.* [15] presented a defect segmentation method for multicrystalline solar cell based on the deep-learning networks. This method applies RPN to generate underlying defect regions, which will be processed to fixed sizes for feeding into an improved segmentation network. However, applying CNN to detect multicrystalline solar cell defect in high resolution image with the complex background disturbance is of great challenge, which has to form a tightly surrounded bounding box around the defects in raw EL image, as shown in Fig. 3.

It is noticed that an important property of human visual system is that one does not attempt to process a whole scene image at once. Instead, humans exploit a sequence of partial glimpses and selectively focus on salient parts in order to capture visual structure better [20]. Recently, the relatively popular terminology of neural network is attention mechanism, which is similar to human visual system. As an important role of CNN module, attention mechanism can be used to select or emphasize weak features of objects and simultaneously suppress noise information from the image. Thus, the attention mechanism is suitable to defect detection in solar cell images under the disturbance of heterogeneously complex background. Moreover, to our best knowledge, there were no works incorporating the attention mechanism in solar cell defect detection, which motivates this study.

Attention model adaptively computes the weights of the channel-wise or spatial attention map that will reweight and multiply with the CNN-extracted feature maps to achieve the suppression of noise interference features and the highlighting of object features, which has been widely applied to various tasks. Wang *et al.* [21] proposed a non-local attention model to boost feature representation ability of the image segmentation task. Shen *et al.* [22] employed channel-wise attention model to suppress the noise features of background and guide multi-scale feature fusion in object detection task. Zhou *et al.* [23] applied spatial attention that can find the object spatial location to boost pedestrian detection. Zhou *et al.* [24] applied visual attention model and wavelet transform to detect defects in glass bottle bottom. As a top-down attention module [25], the proposed Complementary Attention Network (CAN) connects channel-wise attention [26] subnetwork with spatial attention subnetwork together, which not only can suppress noise features of background, but also focuses on spatial locations of the defects in complex background of the EL images.

Channel-wise attention can tell the network to select which feature map of the last-layer output in CNN architectures, so that the useless background information can be suppressed as much as possible. In this paper, the novel channel-wise attention subnetwork applies convolution operation to integrate the concatenated and discriminative output features of GAP layer and GMP layer, which can make fully use of these

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2020.3008021, IEEE Transactions on Industrial Informatics

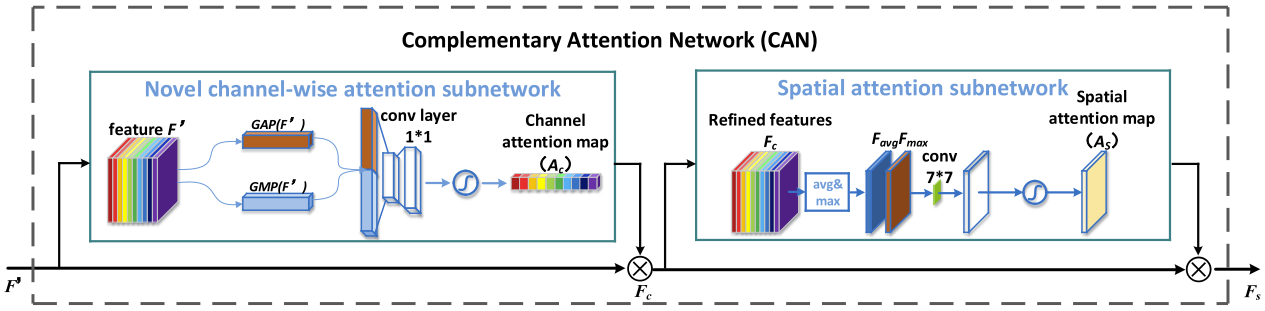


Fig. 4. Overview of the proposed Complementary Attention Network pipeline, which contains a novel Channel-wise Attention Subnetwork and a Spatial Attention Subnetwork. GAP: Global Average Pooling, GMP: Global Max Pooling. avg&max: average & max operation of the refined channel features F_c in channel axis.

informative features. Channel-wise attention reweights the input feature maps with an attention map. If one feature map mainly includes defect information, a high weight will be multiplied with it. On the contrary, if one feature map mainly includes complex background information, a low weight will be multiplied with it. By above calculation, some useful features for defect inspection are retained, and noise features of complex background will be suppressed simultaneously.

As a complementary for the novel channel-wise attention subnetwork, spatial attention subnetwork is designed by reference to CBAM [27] architecture, which can focus on defect spatial locations at the input feature maps. Defect region in the feature map will be assigned a high weight to multiply, and the weight multiplying with the background region is low.

Recently, the Faster R-CNN [28] has been an effective object detection architecture in many fields. In this two-stage network (Faster R-CNN), RPN plays an important role in recommending region proposals of defects, which will be utilized for subsequently more accurate classification and bounding box prediction task. However, the RPN is easily disturbed by the complex background and will recommend a large number of invalid candidate regions of background. In order to suppress the interference of complex background, Region Proposal Attention Network (RPAN) that integrates the novel CAN into RPN is proposed to recommend more refined region proposals of defects.

The contributions of this paper are as follows.

1) we propose a novel Complementary Attention Network (CAN) for adaptive background features suppression and defect features highlighting, which is formed by connecting the novel channel-wise attention subnetwork with spatial attention subnetwork sequentially. In CAN, the novel channel-wise attention subnetwork applies convolution operation to integrate the concatenated and discriminative output features of GAP layer and GMP layer, which can make fully use of these informative features.
2) We propose a novel Region Proposal Attention Network (RPAN) based on integrating attention network (CAN) into RPN, which is used to construct a novel end-to-end Faster RPAN-CNN framework for detecting defects in raw EL images. In Faster RPAN-CNN, RPAN can extract more refined region proposals of defects with the heterogeneous complex background disturbance.
3) The proposed CAN presents good versatility, which not only enhances the classification performance of several classical CNN models, but also improve the detection accuracy of the Faster R-CNN. Furthermore, the detection framework Faster RPAN-CNN can satisfy the real-time quality monitoring in the actual manufacture process of photovoltaic solar cells.

The remainder of this study is organized as follows. Section II gives the details for the proposed methods. Section III presents extensive experiments and related explanations. Finally, conclusion and discussion are presented in Section IV.

II. Methodology

In this section, we mainly introduce the architectures of the proposed Complementary Attention Network (CAN) and Region Proposal Attention Network (RPAN) in details.

A. Complementary Attention Network (CAN)

The proposed Complementary Attention Network (CAN) connects novel channel-wise attention subnetwork with spatial attention subnetwork, which can suppress noise features of background, and focus on defect spatial locations in solar cell EL image with heterogeneous background disturbance. Channel-wise attention tells the network "what" to look, and spatial attention tells the network "where" to look. Thus, the relation of the two attention mechanisms is complementary.

CAN is inspired by the CBAM [27] module. In the channel-wise attention of CBAM, the multi-layer perceptron (MLP) is used to refine the output features of GAP and GMP, and then adds the refined weights of GAP and GMP for feature fusion, which cannot well integrate the informative feature of GAP and GMP. Because, after being refined through MLP, the features of GAP and GMP are more discriminative. Then, addition operation is not a good way for feature fusion, which can't retain the discriminative features of GAP and GMP very well. Thus, we propose that the feature fusion of GAP and GMP can be realized through feature concatenation and convolution operations, which can retain more discriminative features of GAP and GMP. Concatenation operation retains the discriminative features of GAP and GMP, and convolution operation integrates the concatenated features of GAP and

GMP to form the appropriate weights for every channel. That's the reason why the proposed method (CAN) can achieve better performance than CBAM, which is proved by the following quantitative experiments. Otherwise, for sequential connection of two attention modules, there are two manners of connection: channel-first order and spatial-first order. Fig. 4 shows the channel-first order, spatial-first order is opposite placement.

As shown in Fig. 4, the input features $F^{'} \in \mathbb{R}^{H\times W\times C}$ are firstly refined by the channel-wise attention subnetwork, then we will get the channel-wise attention map $A_c \in \mathbb{R}^{1\times 1\times C}$. A_c will be used to reweight each channel of the input features $F^{'}$ to obtain the channel refined feature F_c. Subsequently, the spatial attention subnetwork will be applied to locate the object region in the channel refined feature F_c to obtain the spatial attention map $A_s \in \mathbb{R}^{H\times W\times 1}$. Then, A_s is used to multiply with the channel refined features F_c based on element-wise multiplication. Finally, we will obtain the spatial refined features F_s. CAN like a filter module not only can combine with RPN, but also can be widely used to boost the representation ability of CNNs.

$$F_c = A_c(F^{'}) \times F^{'} \tag{1}$$

$$F_s = A_s(F^{c}) \times F^{c} \tag{2}$$

where W, H denote the width and height of the feature map, C denotes the channel number of the input feature maps.

1) Novel Channel-wise Attention Subnetwork: We propose a novel channel-wise attention subnetwork, which focuses on selecting the channel that includes expected information. To obtain stronger semantic information, we squeeze the spatial dimension of the input features $F^{'}$ by Global Average Pooling (GAP) and Global Max Pooling (GMP). [29] and [30] confirmed that GAP and GMP can gather distinctive features to object recognition. Thus, we concatenate the output features of GAP and GMP, so as to make fully use of the informative features extracted by GAP and GMP. Next, we apply two fully convolution layers to select useful information from the concatenated feature.

As shown in Fig. 4, the implementation details of channel-wise attention are as follows. Firstly, GAP and GMP are applied to reduce spatial dimension information of the features $F^{'}$, generating two feature vectors $GAP(F^{'})$ and $GMP(F^{'})$. Secondly, by concatenating two feature vectors $G = [GAP(F^{'}), GMP(F^{'})] \in \mathbb{R}^{1\times 1\times 2C}$, we fuse the information of $GAP(F^{'})$ and $GMP(F^{'})$ together. Thirdly, G will be passed to the following two fully convolution layers to select useful information. After activation, we will get the weight of each channel $A_c \in \mathbb{R}^{1\times 1\times C}$. The output channels of the first fully convolution layer is C/k, where k is the reduction ratio.

$$A_c(F^{'}) = Sigmoid(conv_{C}^{1\times 1}(conv_{C/k}^{1\times 1}(G))) \tag{3}$$

where $conv_{C}^{1\times 1}$ represents that the convolutional filter kernel size of the fully convolution layer is (1, 1), and the output channel number of the fully convolution layer is C.

Algorithm 1 Defect Detection of Faster RPAN-CNN

Input: a raw IR image with a size of 1024×1024 pixels
1) extract the image feature $F \in \mathbb{R}^{C\times W\times H}$
2) employ RPAN to recommend M suspected defect region proposals $P_i,\ i \in \{1...M\}$
3) proposals $P_i,\ i \in \{1...M\}$ are mapped to feature F, then obtaining proposals feature $P_i',\ i \in \{1...M\}$
4) max pooling layer is employed to dimensionality reduction of $P_i',\ i \in \{1...M\}$, then obtaining one-dimensional vector $P_i'',\ i \in \{1...M\}$
5) fully connected layer applies $P_i'',\ i \in \{1...M\}$ to predict class c_i and location l_i of the defect in proposal $P_i,\ i \in \{1...M\}$
Output: c_i and location l_i of the defect, $i \in \{1...M\}$

2) Spatial Attention Subnetwork: As a supplement for channel-wise attention subnetwork, spatial attention subnetwork is good at retaining more spatial features of defects. We utilize above refined channel-wise features F_c as input of spatial attention subnetwork, which focuses on "where" is the object position. Then, spatial attention map A_s, which carries the location information of the object, will multiply with the refined channel-wise features, as shown in Fig. 4. To compute the spatial attention map A_s, average and max operations along the channel axis are applied to generate two efficient feature maps $[F_{avg}; F_{max}]$, which will be concatenated to form the input features of subsequent convolution operation. The output of the convolution operation will be activated by activation function. Finally, we will obtain the spatial attention map $A_s(F_c) \in \mathbb{R}^{W\times H\times 1}$. Noting that the spatial attention subnetwork is designed by referring to CBAM architecture.

$$F_{avg} = average(F_c) \tag{4}$$

$$F_{max} = max(F_c) \tag{5}$$

$$A_s(F_c) = sigmoid(conv_{1}^{7\times 7}[F_{avg}; F_{max}]) \tag{6}$$

where average & max is the element-wise average & max operation in the channel axis.

B. Defect Detection framework Based on RPAN

In this paper, we use Faster R-CNN as the base detector to detect crack defect, finger interruption defect, and black core defect in EL near-infrared images, while integrating a novel attention mechanism into RPN.

1) Overview of Defect Detection Architecture: The flowchart of our Faster RPAN-CNN architecture is shown in Fig. 5, and the pseudocode is presented in Algorithm 1. The proposed Faster RPAN-CNN architecture mainly includes three parts: feature extraction, proposal extraction (RPAN), classification and detection. Firstly, the input image passes through the backbone network (e. g. VGG16 [31]) to extract features. Secondly, the extracted features will be transmitted to our proposed RPAN module, which will generate many refined region proposals. Thirdly, the features for each proposal are generated by corresponding to the output features of VGG16, next the Region of Interest (RoI) pooling layer will pool the produced features into a fixed-length feature vector, and a sequence of fully connected layers end up with two output layers: one outputs classification scores over K object classes plus a "background" class and the other outputs refined bounding-box positions for each of the K object classes. As

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2020.3008021, IEEE Transactions on Industrial Informatics

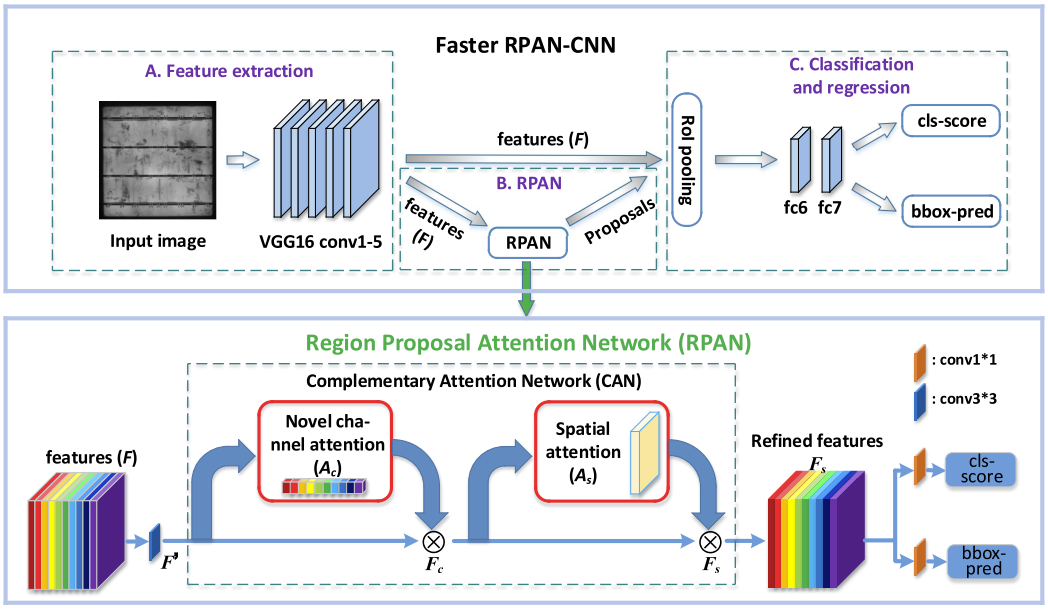


Fig. 5. Defect detection architecture (Faster RPAN-CNN) with the novel region proposal attention network (RPAN). cls-score: classification score, bbox-pred: bounding-box position prediction. conv1*1 & 3*3: the width and height of the convolutional filter kernel is (1, 1), (3, 3).

the first stage of Faster RPAN-CNN architecture, RPAN can be trained by optimizing the following loss function:

$$L(\{p_i\},\{t_i\}) = \frac{1}{N_{cls}}\sum_i L_{cls}(p_i, p_i^*) + \lambda\frac{1}{N_{reg}}\sum_i p_i^* L_{reg}(t_i, t_i^*) \tag{7}$$

where i is the index of an anchor box [28] in a mini-batch and p_i is the predicted probability of anchor box i containing a defect. If the anchor box is predicted to contain a defect, the ground-truth label p_i^* is 1, otherwise p_i^* is 0. t_i is a vector representing the 4 parameterized coordinates of the predicted bounding box, and t_i is that of the ground-truth box associated with a defective anchor box. The term $p_i^* L_{reg}$ means the regression loss is activated only for defective anchor $(p_i = 1)$ and is disabled otherwise $(p_i = 0)$. The outputs of the cls_score layer and bbox_pred layer consist of p_i and t_i respectively. The classification loss L_{cls} is log loss over two classes (defect or not defect). For the regression loss, we use $L_{reg}(t_i, t_i^*) = R(t_i - t_i^*)$ where R is the robust loss function (smooth L1) defined as follows:

$$smooth_{L1}(x) = \begin{cases} 0.5x^2 & if\ |x| < 1 \\ |x| - 0.5 & otherwise \end{cases} \tag{8}$$

The term $p_i^* L_{reg}$ means the regression loss is activated only for defective anchor boxes $(p_i^* = 1)$ and is disabled otherwise $(p_i^* = 0)$. Referring to Faster R-CNN, the two terms are normalized by $N_{cls} = 256$ and $N_{reg} = 2400$, and weighted by a balancing parameter $\lambda = 10$, thus both cls and reg terms are roughly equally weighted.

For bounding box regression, the parameterizations of the 4 coordinates is defined as follows:

$$t_x = \frac{x - x_a}{w_a}, \quad t_y = \frac{y - y_a}{h_a} \tag{9}$$

$$t_w = \log(\frac{w}{w_a}), \quad t_h = \log(\frac{h}{h_a}) \tag{10}$$

$$t_x^* = \frac{x^* - x_a}{w_a}, \quad t_y^* = \frac{y^* - y_a}{h_a} \tag{11}$$

$$t_w^* = \log(\frac{w^*}{w_a}), \quad t_h^* = \log(\frac{h^*}{h_a}) \tag{12}$$

where x, y, w and h donate the center coordinates, width and height of the box. Variables x, x_a, and x^* are for the predicted box, anchor box and ground-truth box, respectively (y, w, h are same as this). Otherwise, this can be thought of as bounding-box regression from an anchor box to a nearby ground-truth box.

2) Proposed Region Proposal Attention Network (RPAN): A novel Region Proposal Attention Network (RPAN) that embeds the proposed Complementary Attention Network (CAN) into RPN is designed to generate more refined region proposals. The CAN learns "what" and "where" to focus on in the input feature maps and guide RPN to predict the underlying defect positions more accurately in complex background.

As shown in Fig. 5, the output features F of VGG16 network will be transmitted to RPAN, which is composed of two branches such as one classification branch and one bounding box prediction branch. These feature maps F are mapped to $F^{'}$(512-d) by a convolutional filter kernel. Then, the features $F^{'}$ are refined by Complementary Attention Network (CAN). Next, the refined features F_s are fed into two sibling fully connected layers including a box-classification layer (cls-score) and a bounding box position prediction layer (bbox-pred). Note that the classification and box-prediction are improved at the same time, because CAN suppresses the interference of complex background and guides RPN to yield more refined region proposals of the defects.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2020.3008021, IEEE Transactions on Industrial Informatics

III. Experiments

The proposed CAN module is modular and independent of application type, which can be easily adapted for image level classification and detection tasks. In order to demonstrate its applicability to image classification and detection, we evaluate the proposed CAN module in the classification task of the cropped EL images and the RPN-based detection task of raw EL images.

A. Dataset

We evaluate the classification effectiveness of our proposed CAN with finger interruption patches of 128×128 pixels. Specifically, 1008 finger interruption patches and 1021 good samples were manually annotated. The patch dataset distribution of training data and testing data is shown in Table I. Moreover, 2129 EL defective images and 1500 good images with raw resolution of 1024×1024 are used to evaluate the classification and detection effectiveness of our proposed Faster RPAN-CNN framework. Table II shows the raw dataset distribution of Faster RPAN-CNN training data and testing data. Due to the difference of power generation efficiency, the brightness of each EL image shows inconsistent characteristic.

TABLE I: DISTRIBUTION OF SOLAR CELL PATCH DATASETS (128×128 PIXELS) IN THE TRAINING AND TESTING SAMPLES. FR PATCH: FINGER INTERRUPTION PATCH.

Dataset (patches)	Fr patch	Good	Total
Training	300	300	600
Testing	708	721	1429
Total	1008	1021	2029

TABLE II: DISTRIBUTION OF SOLAR CELL RAW DATASET (1024×1024 PIXELS) IN THE TRAINING AND TESTING SAMPLES. CK: THE RAW DEFECTIVE IMAGE CONTAINS ONLY CRACK DEFECTS. FR: THE RAW DEFECTIVE IMAGE CONTAINS ONLY FINGER INTERRUPTION DEFECTS. CK&FR: THE RAW DEFECTIVE IMAGE CONTAINS BOTH CRACK DEFECTS AND FINGER INTERRUPTION DEFECTS. BC: THE RAW DEFECTIVE IMAGE CONTAINS ONLY BLACK CORE DEFECTS.

Dataset (raw images)	Defective images				Good images	Total
	Ck	Fr	Ck&Fr	Bc		
Training	260	252	85	250	/	847
Testing	393	521	97	271	1500	2782
Total	2129				1500	3629

B. Evaluation metrics

We evaluate the performance of defect classification by the following indicators such as Precision, Recall and F-measure. Moreover, Average Precision (AP), mean Average Precision (mAP) and Mean Intersection over Union (MIoU) are used to assess the performance of defect detection methods. Noting that AP and mAP are suitable for evaluating the average precision of the single class and multiple classes, respectively.

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

$$F-measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{14}$$

$$IoU = \frac{DetectionResult \cap GroundTruth}{DetectionResult \cup GroundTruth} \tag{15}$$

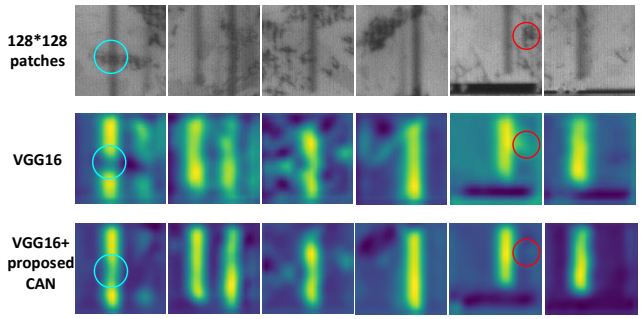


Fig. 6. Gradient Class Activation Mapping (Grad-CAM) visualization results. We compare the visualization results of our proposed method CAN-integrated network (VGG16+CAN) with the baseline network (VGG16) on our patch dataset.

where TP presents the number of true defects, which predicted to be defective; TN presents the number of true non-defect, which is predicted to be non-defective; FP presents the number of true non-defect, which is predicted to be defective; FN presents the number of true defects, which is predicted to be non-defect. $DetectionResult$ is the defect detection box of the detection framework; $GroundTruth$ is the true manual labeling box of the defect.

C. Implementation Details

We use ImageNet-pre-trained weight to finetune the AlexNet [32], VGG16, ResNet50 [33] and DenseNet121 [34] models in the classification evaluation of proposed CAN. In the case of less training samples, the pre-trained model can speed up convergence of the network and improve the accuracy of final target task for small training samples. We fix the reduction ratio k to 2 in channel-wise attention subnetwork. The filter kernel size of convolution layer in spatial attention subnetwork is 7×7. Moreover, if there is no special explanation in this paper, channel-wise attention subnetwork in CAN module is in front of spatial attention subnetwork.

For Faster RPAN-CNN framework, the learning rate of our model is 10^{-3} for 20000 iterations. The number of the proposals is 150. The classification score threshold and non-maximum suppression threshold are both 0.5. It is noteworthy that overlap threshold for a RoI to be considered as background is set to [0.0, 0.3] and other hyperparameters are consistent with the original model. Moreover, solar cell defects do not require strong semantic information, thus we choose the shallow network VGG16 as our basic network, and the ImageNet-pre-trained weight of VGG16 is used to finetune Faster R-CNN and Faster RPAN-CNN in the following experiments.

For the ground truth of different defects (crack defect, finger interruption defect, and black core defect), we use a python dataset annotation library (LabelImg) to make the dataset in VOC2007 format by ourselves, which just needs a box to tightly contain the defect and do not need too much expert experience.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2020.3008021, IEEE Transactions on Industrial Informatics

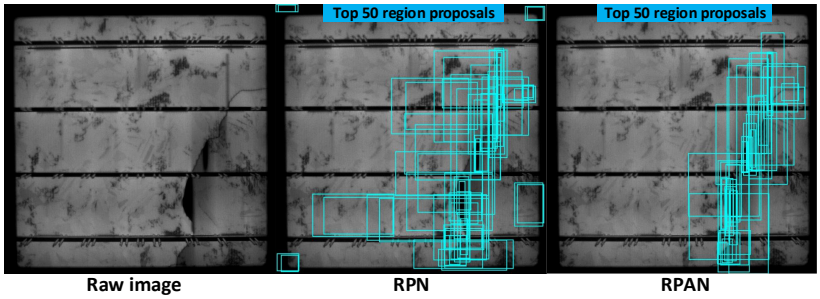


Fig. 7. The top 50 region proposals of RPN and our proposed RPAN for the same original EL image with one crack defect and one finger interruption defect.

TABLE III: CLASSIFICATION RESULTS ON FINGER INTERRUPTION DEFECT PATCHES (DATASET PATCHES), EVALUATED BY PRECISION (P), RECALL (R) AND F-MEASURE (F).

methods	P (%)	R (%)	F (%)
AlexNet	97.16	89.05	92.93
AlexNet+CBAM	97.23	90.56	93.78
AlexNet+proposed CAN	97.74	92.75	95.18
VGG16	93.34	99.89	96.51
VGG16+CBAM	95.20	98.18	96.67
VGG16+proposed CAN	98.24	97.10	97.66
ResNet50	96.67	94.90	95.78
ResNet50+CBAM	99.61	93.47	96.44
ResNet50+proposed CAN	99.31	95.85	97.50
DenseNet121	98.43	93.84	96.08
DenseNet121+CBAM	99.90	93.84	96.77
DenseNet121+proposed CAN	98.82	97.10	97.96

D. Classification Evaluation of the Proposed CAN

1) Visualization analysis with Grad-CAM: To illustrate the effectiveness of proposed CAN module, we apply Gradient Class Activation Mapping (Grad-CAM) [35], which does not need to change the structure of CNN model comparing with CAM [36], to visualize the learned features of the CNN model in finger interruption patches. The base-network is VGG16, and the output feature maps of VGG16 final layer has the strongest semantic information. We embedded CAN between VGG16 backbone and fully connection layers, then using Grad-CAM to visual the output features of CAN. Grad-CAM is formed by weighting and superimposing each feature map of the VGG16 or VGG16+CAN final-layer output. Moreover, Grad-CAM can deepen our understanding about the learned features of convolutional networks, thereby achieving qualitative analysis. As shown in the cyan-blue circle and red circle of Fig. 6, we can clearly see that the Gradient Class Activation Mapping of VGG16+CAN is less disturbed by the complex background and presents more complete defect region than VGG16, which illustrates that our proposed CAN not only can suppress the noise feature of the heterogeneous background, but also can retain more spatial information features of defects, which indeed can guide the CNN to focus on the important defect region.

2) Quantitative Evaluation of the Proposed CAN: As illustrated in Table III, after adopting Complementary Attention Network to refine the feature maps from the final output of CNN models, this paper acquires the most exciting experimental results, which shows that CAN is good at promoting the classification results of many classical CNN models such as AlexNet, VGG16, ResNet50 and DenseNet121 in our patch dataset. Comparing with attention module CBAM [27] that is used to boost the representation ability of CNN models, CAN performs better in the classification task of solar cell near-infrared images, which illustrates that CAN makes more fully use of the discriminative features of GAP layer and GMP layer. These results verify that CAN can be widely used to boost the representation power of CNN architectures. It's worth noting that we harness the pre-training weights to compensate for the shortcoming of insufficient training samples in our experiments. Moreover, proposed CAN is embedded into Region Proposal Network (RPN) in next section, and makes a good performance in defects detection task of crack, finger interruption, and black core with proposed Faster RPAN-CNN.

E. Detection Evaluation of the Proposed RPAN

In this section, we perform detection experiments to evaluate our proposed RPAN module in details. The experiments analysis will be given to verify the effectiveness of the proposed Faster RPAN-CNN model through subjective visualization analysis and quantitative analysis.

1) Visualization of the Top 50 Region Proposals: When training an EL image defect detection model, the proposed RPAN learns to recommend proposals of the image with a score describing the likelihood that the proposal contains a defect. Fig. 7 shows that the top 50 region proposals (top 50 in score rank) of RPN and our proposed RPAN for the same original EL image with one crack defect and one finger interruption defect. It clearly shows that for the case that the defects intersect with the black complex background region, proposed RPAN can recommend more refined defect proposals than RPN of the top 50 region proposals. Interestingly, when combining attention network (CAN) with RPN, CAN suppresses the interference of complex background and guide the RPN to more precisely focus on the defect. Thus, RPAN performs better than RPN to recommend region proposals in EL near-infrared image under complex background disturbance.

2) Visualization of the Detection Results: Visual improvements on raw EL image Dataset are shown in Fig. 8. The crack defect is seriously affected by the complex background, and self-texture structure is complicated. Faster RPAN-CNN

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2020.3008021, IEEE Transactions on Industrial Informatics

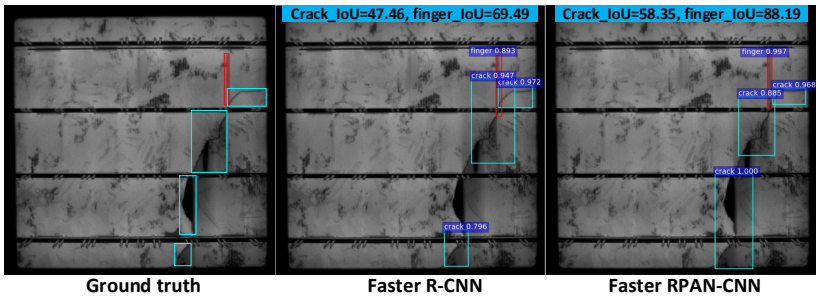


Fig. 8. Visual improvements on raw EL image Dataset. The detection results include four things: prediction box, defect category, classification score of defects and IoU value of defects.

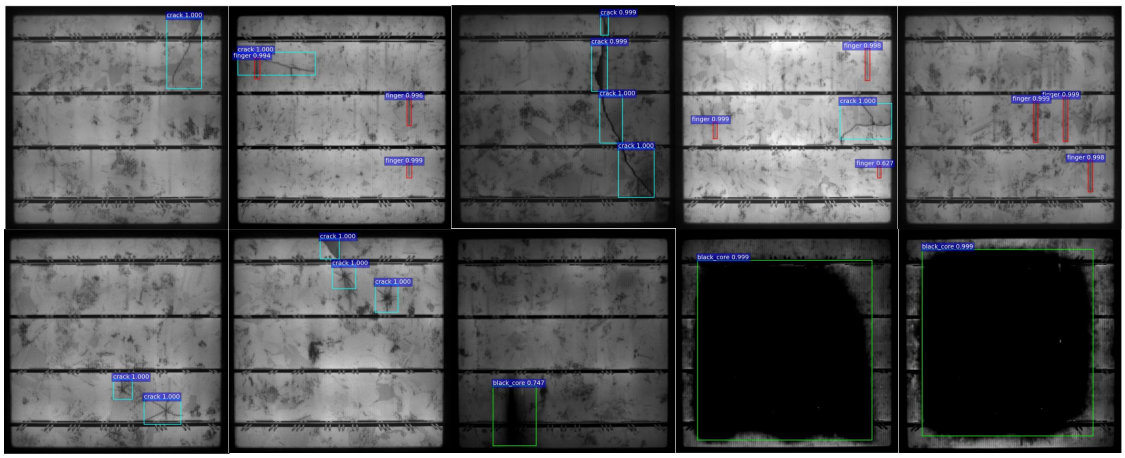


Fig. 9. The visual results of proposed Faster RPAN-CNN framework on raw EL image.

performs better than Faster R-CNN in crack defect location task. The IoU value of the crack defect presents 10.91% higher than the detection result of Faster R-CNN. For finger interruption defect, Faster R-CNN can't tightly surround it with the disturbance of complex background in EL image, while Faster RPAN-CNN corrects this error and the IoU value of the finger interruption defect presents 18.70% higher than the detection result of Faster R-CNN. Moreover, the classification score of the finger interruption defect is also higher. More visual results of Faster RPAN-CNN for raw EL images with various intensity, texture shape and scale are included in Fig. 9. We can easily conclude from above visualization results of our raw EL image dataset that Faster RPAN-CNN framework performs more accurate position prediction than Faster R-CNN under the heterogeneously complex background interference.

3) Quantitative Evaluation of the Proposed Detector: The Precision, Recall, F-measure, mAP and MIoU of different detectors on solar cells raw EL image dataset are shown in Table IV. In this section, we conduct quantitative evaluation between three types of detectors to illustrate the effect of our proposed method (Faster RPAN-CNN). In terms of the defective image classification effect, the F-measure of Faster RPAN-CNN (CAN+RPN, CS) is 98.47%, which performs better than YOLO v3 [37] and Faster R-CNN. Moreover, the recall results of Faster RPAN-CNN (CAN+RPN, CS) is 97.78%, which presents that the defective image is not easy to be missed during the process of intelligent manufacturing.

In order to measure the performance of the learned model in

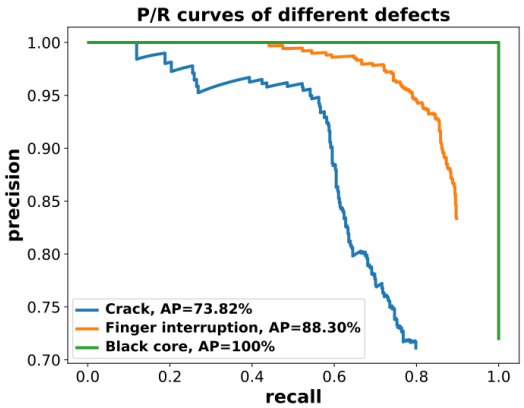


Fig. 10. The P/R curves of different defects using proposed Faster RPAN-CNN (RPN+CAN, CS) detector. Noted that the area enclosed by P/R curve and coordinate axis is the value of AP.

defect detection, we employ mAP and MIoU to demonstrate the effectiveness of our proposed method. We can see that the mAP result of Faster RPAN-CNN (RPN+CAN, CS) is 87.38% in raw EL image dataset, which achieves 8.59% and 4.27% improvement than YOLO v3 and Faster R-CNN respectively, used for comparative experiments. In terms of position prediction, we calculate MIoU between ground truth and prediction box to evaluate the effect of position prediction. As shown in Table IV, we observe that our Faster RPAN-CNN (RPN+CAN, CS) achieves 6.46% and 3.20% improvement

TABLE IV: RESULTS OF OTHER DETECTORS AND OUR PROPOSED FASTER RPAN-CNN USING DIFFERENT ATTENTION NETWORKS ON OUR RAW EL IMAGE DATASET. (RAW IMAGES OF 1024×1024 PIXELS ARE USED TO BE THE INPUTS OF THE DETECTION MODEL.) FOR CAN MODULE, SC: THE CONNECTION SEQUENCE IS SPATIAL ATTENTION IN FRONT AND CHANNEL-WISE ATTENTION IN BACK, CS IS JUST THE OPPOSITE. SINCE AN IMAGE MAY CONTAIN MULTIPLE TYPES OF DEFECTS, THE CLASSIFICATION RESULTS CONSIDER WHETHER AN IMAGE CONTAINS DEFECTS OR NOT AND DO NOT DISTINGUISH THE CATEGORIES OF THE DEFECTS.

Detector (raw dataset)	Classification			Detection		Parameters
	P (%)	R (%)	F (%)	mAP (%)	MIoU (%)	
YOLO v3	86.73	97.30	91.71	78.79	64.96	58.73M
Faster R-CNN	93.53	96.04	94.77	83.11	68.22	260.50M
the proposed Faster RPAN-CNN (RPN+CBAM)	95.15	96.26	95.70	85.78	69.13	261.01M
the proposed Faster RPAN-CNN (RPN+CAN, SC)	96.22	96.84	96.52	86.09	70.68	261.26M
the proposed Faster RPAN-CNN (RPN+CAN, CS)	99.17	97.78	98.47	87.38	71.42	261.26M

than YOLO v3 and Faster R-CNN respectively, on the MIoU evaluation index, which illustrates that Faster RPAN-CNN (RPN+CAN, CS) makes solar cells defect location prediction more accurate under the heterogeneously complex background disturbance. Moreover, it is worth noting that channel-first order Faster RPAN-CNN (RPN+CAN, CS) performs better than the spatial-first order Faster RPAN-CNN (RPN+CAN, SC), and the overhead of parameters and computation is negligible in most cases.

The P/R curves of different defects using our proposed Faster RPAN-CNN (RPN+CAN, CS) are shown in Fig. 10. As can be seen from the Fig. 10, the detection result of the black core defect is the best, the AP value reaches 100%. Black core defect has a single blob-like shape and a large scale, which cause it easy to be detected, so the AP value is very high. The detection result of finger interruption defect, which processes a single vertical shape and small scale, is still easier to be detected. The detection result of crack defect is the worst. This is due to the random shapes and different scales of the cracks, which bring some difficulties to detect crack defects in solar cell EL images.

4) Time-efficiency Evaluation: Fast and accurate defect detection not only reduces production costs, but also improves the production quality of solar cells. Thus, time-efficiency evaluation plays an important role in the process of intelligent manufacture. Time-efficiency evaluation is performed on an Intel CPU (i7-6700K, 4.00GHz) with two NVIDIA GeForce GTX 1080 GPUs. The input of the model is the raw EL image (1024×1024 pixels). Fig. 11 shows the relationship between F-measure, evaluation time and the number of region proposals. Increasing the number of region proposals decreases the chance that a defect will be missed, but it increases the computational time simultaneously. Thus, selecting a proper number of region proposals is important. Based on above results, 150 region proposals are considered to provide a good balance between speed and F-measure.

Moreover, the proposed Faster RPAN-CNN can achieve better results with fewer region proposals than Faster R-CNN, which testifies that proposed RPAN can extract more refined region proposals than RPN. Furthermore, the time-efficiency of proposed Faster RPAN-CNN framework is similar to Faster R-CNN framework, which presents that RPAN module is light-weight.

5) Comparison of some traditional methods: Comparing with traditional methods, CNN has achieved a better performance in image classification task [38]. We compared our approach with some hand-craft feature extraction approaches to further illustrate the effectiveness of the proposed Faster RPAN-CNN, such as GLOH [9], LBP [39], CPICS-LBP [10] and AE-CLBP [40], which are used to extract solar cell raw EL image feature, then the Support Vector Machine (SVM) is employed to train and test the classification model. For traditional methods, 500 good images in raw EL dataset are utilized for training, 1000 good images are utilized for testing. Others are consistent with the dataset distribution (two classes: defective images and good images) in Table II. The kernel function that is applied in SVM is linear function. As shown in the Table V, the proposed Faster RPAN-CNN achieves 23.13%, 19.15%, 16.04% and 10.99% hit rate of F-measure improvement from GLOH+SVM, LBP+SVM, CPICS-LBP+SVM, and AE-CLBP+SVM, respectively. It illustrates that our method is better than traditional methods to classify defective EL image under the disturbance of complex background in multicrystalline solar cells.

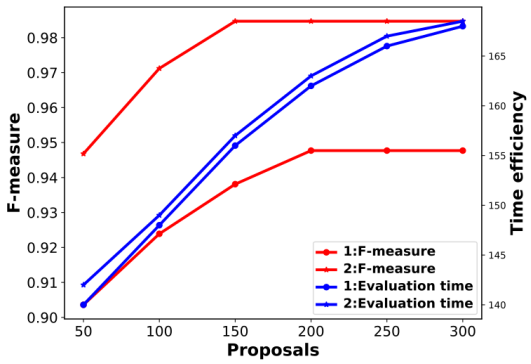


Fig. 11. Relationship between defect detection F-measure, evaluation time, and the number of region proposals. 1: Faster R-CNN, 2: proposed Faster RPAN-CNN.

TABLE V: COMPARISON WITH SOME TRADITIONAL METHODS. PRECISION, RECALL AND F-MEASURE OF DEFECTIVE EL IMAGE.

methods	P (%)	R (%)	F (%)
GLOH+SVM	85.48	67.39	75.34
LBP+SVM	92.07	69.67	79.32
CPICS-LBP+SVM	83.50	82.12	82.43
AE-CLBP+SVM	92.88	82.67	87.48
Faster RPAN-CNN	99.17	97.78	98.47

TABLE VI: THE CLASSIFICATION AND DETECTION RESULTS OF GAP AND GMP IN CHANNEL-WISE ATTENTION SUBNETWORK.

Detector	Classification			Detection	
	P (%)	R (%)	F (%)	mAP (%)	MIoU (%)
Faster RPAN-CNN, GAP	98.21	96.68	97.42	86.45	70.84
Faster RPAN-CNN, GMP	98.06	96.57	97.31	86.12	70.01
Faster RPAN-CNN, GAP&GMP	99.17	97.78	98.47	87.38	71.42

TABLE VII: HYPERPARAMETERS OF FASTER RPAN-CNN

Training	Reduction_ratio k	Weight_decay	Learning_rate	Momentum	batch_size	Max_iteration
	0.5	0.0005	0.001	0.9	256	20000
	Pretrained_model	image_resize	rpn_batch_size	RoI_foreground_threshold	RoI_background_threshold	proposals
	TRUE	(1024, 1024)	64	(0.5, 1.0)	(0.0, 0.3)	150
Testing	classification score threshold			non-maximum suppression threshold		
	0.5			0.5		

IV. Conclusion and Discussion

We can conclude from above experimental results that our proposed CAN not only suppresses the interference of heterogeneously complex background, but simultaneously highlights more complete spatial defects region. The proposed CAN achieves great performance in the EL pitches classification task by integrating into many classical CNN models, which means that our attention module can be widely used to enhance the classification performance of these CNN models. This is encouraging. Moreover, RPAN performs better than RPN to recommend region proposals and makes defect position prediction more accurate in raw EL near-infrared image under complex background disturbance. Our method Faster RPAN-CNN achieves good detection results for line-like and star-like crack, line-like finger interruption, and blob-like black core defects, which shows that our proposed deep learning model has good generality for detecting these shapes of defects in raw solar cell EL images. The code of our proposed CAN module is publicly available to help the readers better understanding our methods. Code is available at: https://github.com/binyisu/Complementary-Attention-Network.

We are also aware that our current study has some limitations. For example, the reduction ratio k in CAN is set manually in EL image defect detection task, which takes time and effort. As a part of future works, automatic machine learning has the capability to automatically learn adaptive reduction ratio k in our attention module.

Appendix A

The classification and detection results of GAP and GMP in channel-wise attention subnetwork are shown in Table VI. The F-measure, mAP and MIoU of Faster RPAN-CNN (GAP&GMP) are 98.47%, 87.38% and 71.42%, respectively, which are better than GAP and GMP alone. Otherwise, the results of GAP are better than GMP, which presents that GAP can extract more informative feature of the EL images than GMP.

The detailed information and hyperparameters of the Faster RPAN-CNN model are presented in Table VII, which can help the readers to better understand our algorithm.

References

[1] T. Fuyuki, A. Kitiyanan, "Photographic diagnosis of crystalline silicon solar cells utilizing electroluminescence," *Appl. Phys. A, Solids Surf.*, vol. 96, no. 1, pp. 189-196, Jul. 2009.

[2] Y. Wang, H. Zhao, D. Song, A. Cai, "Research to the Typical Defects of Crystalline Silicon Photovoltaic Cells based on EL Images" *Int. J. Energy. Sci.*, vol. 3, no. 3, pp. 200-204, Jun. 2013.

[3] S. Mukherjee, S. T. Acton, "Oriented filters for vessel contrast enhancement with local directional evidence," in *Proc. IEEE Int. Symp. Biomed. Imaging*, 2015, pp. 503-506.

[4] D. Mery, C. Arteta, "Automatic Defect Recognition in X-Ray Testing Using Computer Vision," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, Mar. 2017, pp. 1026-1035.

[5] D. Tsai, J. Luo, "Mean Shift-Based Defect Detection in Multicrystalline Solar Wafer Surfaces," *IEEE Trans. Ind. Inform.*, vol. 7, no. 1, pp. 125-135, Feb. 2011.

[6] S. A. Anwar, 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, no. 1, pp. 1-17, Dec. 2014.

[7] D. Tsai, S. Wu, W. Li, "Defect detection of solar cells in electroluminescence images using Fourier image reconstruction," *Sol. Energy Mater. Sol. Cells.*, vol. 99, no. 4, pp. 250-262, Apr. 2012.

[8] D. Tsai, G. Li, W. Li, W. Chiu, "Defect detection in multi-crystal solar cells using clustering with uniformity measures," *Adv. Eng. Inform.*, vol. 3, no. 29, pp. 419-430, Aug. 2015.

[9] M. Demant *et al.*, "Microcracks in silicon wafers I: Inline detection and implications of crack morphology on wafer strength," *IEEE J. Photovolt.*, vol. 1, no. 6, pp. 126-135, Jan. 2016.

[10] B. Su, H. Chen, Y. Zhu, W. Liu, K. Liu, "Classification of Manufacturing Defects in Multicrystalline Solar Cells with Novel Feature Descriptor," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 12, pp. 4675-4688, Dec. 2019.

[11] F. Chen, M. R. Jahanshahi, "NB-CNN: Deep Learning-based Crack Detection Using Convolutional Neural Network and Naïve Bayes Data Fusion," *IEEE Trans. Ind. Electron.*, vol. 65, no. 5, pp. 4392-4400, May. 2018.

[12] G. Jiang, H. He, J. Yan, P. Xie, "Multiscale Convolutional Neural Networks for Fault Diagnosis of Wind Turbine Gearbox," *IEEE Trans. Ind. Electron.*, vol. 66, no. 4, pp. 3196-3207, Apr. 2019.

[13] S. Deitsch *et al.*, "Automatic classification of defective photovoltaic module cells in electroluminescence images," *Sol. Energy.*, vol. 185, pp. 455-468, 2019.

[14] H. Y. Chen *et al.*, "Solar cell surface defect inspection based on multispectral convolutional neural network," *J. Intell. Manuf.*, vol. 31, no. 2, pp.453-468, Feb. 2020.

[15] H. Han, C. Gao, "Polycrystalline silicon wafer defect segmentation based on deep convolutional neural networks," *Pattern Recognit. Lett.*, vol. 130, no. 2, pp. 234-241, Feb. 2020.

[16] Y. Wang, Z. Pan, X. Yuan *et al.,* "A novel deep learning based fault diagnosis approach for chemical process with extended deep belief network," *ISA Trans*, vol. 96, no. 7, pp. 457-467, Jul. 2019.

[17] X. Yuan, B. Huang, Y. Wang, C. Yang and W. Gui, "Deep Learning-Based Feature Representation and Its Application for Soft Sensor Modeling With Variable-Wise Weighted SAE," *IEEE Trans. Ind. Inform.,* vol. 14, no. 7, pp. 3235-3243, Jul. 2018.

[18] X. Yuan, J. Zhou, B. Huang, Y. Wang, C. Yang and W. Gui, "Hierarchical Quality-Relevant Feature Representation for Soft Sensor Modeling: A Novel Deep Learning Strategy," *IEEE Trans. Ind. Inform.,* vol. 16, no. 6, pp. 3721-3730, Jun. 2020.

[19] X. Yuan, L. Li and Y. Wang, "Nonlinear Dynamic Soft Sensor Modeling With Supervised Long Short-Term Memory Network," *IEEE Trans. Ind. Inform.,* vol. 16, no. 5, pp. 3168-3176, May. 2020.

[20] H. Larochelle, G. Hinton, "Learning to combine foveal glimpses with a third-order Boltzmann machine," in *Proc. Int. Conf. Neural Inform. Processing*, 2010, pp. 1243-1251.

[21] X. Wang, R. B. Girshick, A. Gupta, K. He. "Non-local neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7794-7803.

[22] Z. Shen, H. Shi, Y. Jia *et al*., "Improving Object Detection from Scratch via Gated Feature Reuse," in *British Machine Vision Conference*, Jul. 2019.

[23] C. Zhou, M. Wu, S. Lam, "SSA-CNN: Semantic Self-Attention CNN for Pedestrian Detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Feb. 2019.

[24] X. Zhou, Y. Wang, Q. Zhu, J. Mao, C. Xiao, X. Lu, H. Zhang. "A Surface Defect Detection Framework for Glass Bottle Bottom Using Visual Attention Model and Wavelet Transform," *IEEE Trans. Ind. Inform.,* vol. 16, no. 4, pp. 2189-2201, Apr. 2020.

[25] C. Cao *et al*., "Look and think twice: Capturing top-down visual attention with feedback convolutional neural networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, Mar. 2015, pp. 2956-2964.

[26] J. Hu, L. Shen, G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.

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

[28] S. Ren, K. He, R. Girshick, 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, Jun. 2017.

[29] M. Lin, Q. Chen, S. Yan, "Network in network," in *Proc. Int. Conf. Learn. Represent.*, 2014, pp. 1-10.

[30] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, "Learning deep features for discriminative localization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2921-2929.

[31] K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent.*, 2015.

[32] A. Krizhevsky, I. Sutskever, G. E. Hinton. "Imagenet classification with deep convolutional neural networks," In *Proc. Int. Conf. Neural Inform. Processing*, 2012, pp. 1097-1105.

[33] K. He, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770-778.

[34] G. Huang *et al*., "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2261-2269.

[35] R. R. Selvaraju *et al*., "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 618-626.

[36] B. Zhou, A. Khosla, L. A., A. Oliva, A. Torralba, "Learning Deep Features for Discriminative Localization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2921-2929.

[37] J. Redmon, A. Farhadi, "YOLOv3: An incremental improvement," Apr. 2018. [Online]. Available: https://arxiv.org/pdf/1804.02767.

[38] A. S. Razavian, H. Azizpour, J. Sullivan, S. Carlsson, "CNN features off-the-shelf: An astounding baseline for recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 512-519.

[39] T. Ojala, M. Pietikäinen, T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971-987, Jul. 2002.

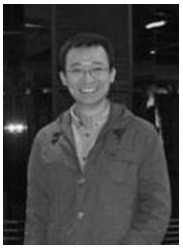
[40] K. Song, Y. Yan, "A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects," *Appl. Surf. Sci.*, vol. 285, no. 21, pp. 858–864, Nov. 2013.

Binyi Su received the B.S. degree from the Hebei University of Technology, Tianjin, China, in 2017, where he is currently pursuing the M.S. degree in automation with the School of Artificial Intelligence and Data Science. He is about to go to School of Computer Science in Beijing University of Aeronautics and Astronautics for a PhD.

He has been studying the automatic detection of photovoltaic solar cell defects in industrial production for three years. He has His current research interests include computer vision and pattern recognition, machine learning and artificial intelligence, industrial image defect detection.

Haiyong Chen received the M.S. degree from the Harbin University of Science and Technology, Harbin, China, in 2005, and the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2008.

He is currently a Professor with the School of Artificial Intelligence and Data Science, Hebei University of Technology, Tianjin, China. He is also an expert in the field of photovoltaic cell image processing and automated production equipment. His current research interests include image processing, robot vision, and pattern recognition.



Peng Chen received M.S. degree and Ph.D. degree from University of science and technology, Beijing, China, in 2009 and 2015. He once studied at Brno Technical University in the Czech Republic and Western Sydney University in Australia.

He is currently a lecturer with the School of Artificial Intelligence and Data Science, Hebei University of Technology, Tianjin, China. His current research interests include intelligent robot, machine vision, 3D environment perception, deep learning and artificial intelligence, machine learning and pattern recognition.

Guibin Bian received the B.S. degree in Mechanical Engineering from North China University of Technology, Beijing, China, in 2004, and the M.S. and Ph.D. degrees in Mechanical Engineering from Beijing Institute of Technology, Beijing, China, in 2007 and 2010 respectively.

He is an associate professor at the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China, where his research interests include design, sensing and control for medical robotics.

Kun Liu received the M.S. degree from the Harbin Institute of Technology, Harbin, China, in 2003, and the Ph.D. degree in automation from Tsinghua University, Beijing, China, in 2009.

She is currently an Associate Professor with the School of Artificial Intelligence, Hebei University of Technology, Tianjin, China. Her current research interests include image processing, computer vision, and pattern recognition.

Weipeng Liu received a M.S. degree in applied mathematics from Hebei University of Technology in 2010, and the Ph.D. degree is received in Control Theory and Control Engineering from the Hebei University of Technology, Tianjin, China, in 2016.

He is currently an Associate Professor with the School of Artificial Intelligence, Hebei University of Technology. His current research interests include image processing, artificial intelligence, robotics, and pattern recognition.