RCAG-Net: Residual Channelwise Attention Gate Network for Hot Spot Defect Detection of Photovoltaic Farms

Binyi Su, Haiyong Chen, Kun Liu, and Weipeng Liu

***Abstract*—The small hot spot defect detection for photovoltaic (PV) farms is a challenging problem due to the feature vanishing as the network deepens. To solve this challenging problem, a novel residual channelwise attention gate network (RCAG-Net) is proposed by employing a novel RCAG module to achieve multiscale feature fusion, complex background suppression, and defect feature highlighting. In RCAG-Net, the novel RCAG module first realizes feature fusion by adding the features of different scale layers. Next, global average pooling (GAP) and multilayer perceptron (MLP) are used to dimension reduction and refinement of the fused features, then yielding an attention map for channelwise feature reweighting by gate mechanism, which employs selective transmission of the convolution neural network (CNN)-extracted features to achieve informative feature filtering. Moreover, residual connection from the fused features to the final output facilitates the insertion of the new RCAG into some classical pretrained models, without breaking its initial behavior. Finally, the proposed approach is validated through a real defect detection system, and the experimental result clearly verifies its effectiveness for small hot spot detection of PV farms.**

***Index Terms*—Attention network, defect detection, photovoltaic (PV) farms, unmanned aerial vehicle.**

I. Introduction

SOLAR photovoltaic (PV) systems can directly convert solar energy into electrical power. Nowadays, PV electricity energy is becoming more and more interesting and attractive among many kinds of renewable energies. In 2019, China alone accounts for almost 40% of global solar PV expansion. With increasing cost-competitiveness and continuous policy support, global additions will exceed 110 GW per year by 2024. Moreover, the total global power generation of PV systems may reach 1200 GW by 2024 [1].

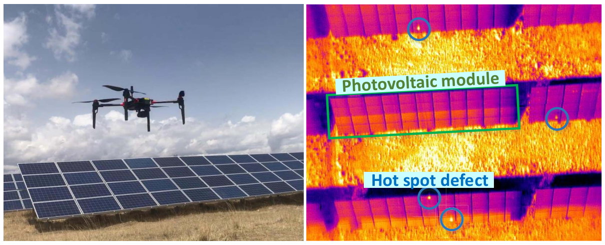


Fig. 1. PV farm and UAV in the left picture, IR image with four small hot spot defects (small object area $\leq 32^2$ pixels, defined in MS COCO data set [9]) in the blue circle on the right picture, and one multicrystalline PV module in the green box.

Thus, defect detection is crucial to the normal operation of PV farms [2] for the timely and accurate fault elimination, which advantages safe and high-efficiency running of the PV farms in the field. As shown in Fig. 1, most of the defects and failures on the multicrystalline PV module are hot temperature regions, named hot spot defect, which presents as blob-shape highlight areas in the infrared (IR) images captured by thermal IR imaging camera mounted on the unmanned aerial vehicle (UAV) [3]. These hot temperature regions associate to some faults, such as dust shielding, broken cells, and circuit fault, which are easily caused by external environment and service life.

Fig. 2 presents the architecture of our UAV-based hot spot detection system, which can diagnose the defect location of global positioning system (GPS) in PV farms. By UAV-based image acquisition system, we can capture relatively clear IR image of hot spot defects, while the characteristics of these defects present small scale, low contrast with background, and the background interference of the suspected defect, which brings some difficulties to the accurate recognition and location of these hot spot defects in PV farms. However, hot spot defects will reduce the efficiency of power generation, produce fire hazards, and then cause irreparable economic losses. Thus, detecting these defects in PV farms is quite necessary for safe and efficient operation during the power generation process.

In fact, computer vision-based method can meet the imperative requirement of the safe and efficient operation in PV farms [4]. Conventional computer vision methods for defect inspection mainly depends on filters [5] or feature descriptors [6], which needs to be designed according to specific applications.

Manuscript received August 30, 2020; revised December 13, 2020; accepted January 16, 2021. Date of publication February 2, 2021; date of current version February 12, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61873315 and Grant 61872023, in part by the Natural Science Foundation of Hebei Province under Grant F2018202078 and Grant F2019202305, and in part by the National Key Research and Development Program of China under Grant 2018YFB2100601. The Associate Editor coordinating the review process was Dong Wang. *(Corresponding author: Haiyong Chen.)*

Binyi Su is with the School of Artificial Intelligence and Data Science, Hebei University of Technology, Tianjin 300130, China, and also with the School of Computing, Beihang University, Beijing 100191, China (e-mail: 18812671934@163.com).

Haiyong Chen, Kun Liu, and Weipeng Liu are with the School of Artificial Intelligence and Data Science, Hebei University of Technology, Tianjin 300130, China (e-mail: haiyong.chen@hebut.edu.cn; liukun@hebut.edu.cn; liuweipeng@hebut.edu.cn).

Digital Object Identifier 10.1109/TIM.2021.3054415

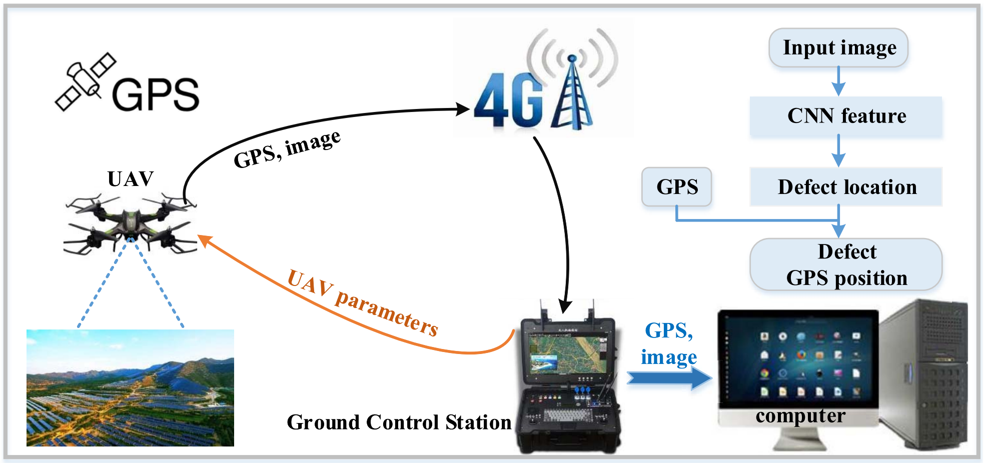


Fig. 2. Our designed UAV-based intelligent automatic defect detection system for hot spot defect in PV farms.

Image filtering is to filter out the specific band frequency in the signal, so as to retain the required band frequency [5]. However, filter-based methods heavily rely on the expertise experience for different applications, which limited its extension. Feature descriptors [6] in conventional methods mainly depends on manually designed extractors, which requires professional knowledge and a complicated parameter adjustment process. At the same time, each method targets a specific application and has poor generalization ability and robustness.

Recently, the deep convolution neural networks (CNNs) have obtained significant success for defect inspection task [7]. Deep learning is a data-driven feature extraction method. According to the learning of massive image samples, the specific feature representation of the data set can be obtained, which are more robust and have more generalization ability. The disadvantage is that the data set used to train the model requires higher computational complexity. Despite its shortcomings, CNN-based methods have achieved excellent performance on the image classification and detection task [8]. Thus, the application of CNN to solve hot spot detection in PV farms shows bright prospect.

For raw PV farm IR image, there are two challenging obstacles: one is the small-scale hot spot defect (small object area $\leq 32^2$ pixels, defined in MS COCO [9]), which just occupies small proportion of pixels in captured images, and another is the complex background disturbance.

In the CNN-based model, as the number of convolutional layers deepens and downsamples, small-scale defect features will continue vanishing, resulting in poor detection performance. As an effective approach, multiscale feature fusion through cross-connection of different scale has great potential to lighten the feature vanish of the small-scale defect, which is employed in this article to improve the effectiveness of small object detection. Li *et al.* [10] employed single shot detection (SSD) algorithm based on the pyramidal feature fusion to detect small object, which achieved a great improvement comparing with original SSD model [11]. Dong *et al.* [12] proposed attention-based feature fusion network for surface small defects' detection, which can fully exploit the high-level features that contain a great deal of semantic information and low-level features that contain rich textural details of small defect. However, low-level features except for rich texture features also contain a large amount of complex background redundant information, which would have a negative effect to the model learned by training.

Based on the above analysis, to suppress complex background feature disturbance during the pyramid feature fusion process, a novel residual channelwise attention gate (RCAG) module is proposed, which achieves feature fusion by adding the features of different scale layers. Then, global average pooling (GAP) and multilayer perceptron (MLP) are used to dimension reduction and refinement of the fused features, which will form an attention map for channelwise feature reweighting by gate mechanism [13] that is responsible for selective transmission of signals or features. Gate mechanism can achieve noise background feature suppression and defect feature highlighting. For example, if one feature map from the fused features mainly contains defect features, a high weight in the attention map will be multiplied with it, which indicates that the feature is allowed to pass. Otherwise, a small weight will be used, which indicates that the feature is prohibited to pass. By above calculation, the useful semantic and texture features from pyramidal layers are preserved, and noise background features will be suppressed as much as possible. Moreover, residual connection from the fused features to the final output allows us to insert a new RCAG module into some classical pretrained models, without breaking its initial behavior, thereby greatly enhancing the versatility of the RCAG module.

Finally, a novel end-to-end regression network (RCAG-Net) is proposed to detect small hot spot defects by employing two RCAG modules to achieve multiscale feature fusion of the final three layers in the backbone feature extraction network through cross-scale connection. The proposed RCAG-Net can greatly improve the detection accuracy of small hot spot defect in PV farm IR images. The contributions of this article are summarized as follows.

1) We propose a novel RCAG module, which can adaptively achieve pyramidal features fusion, background features suppression, and defect features emphasis by employing attention map generated by GAP and MLP operations to reweight with the fused multiscale pyramidal features, and residual connection ensures the versatility of the novel RCAG module.
2) We propose a novel CNN-based defect detection network (RCAG-Net) to detect small hot spot defects, which employs two RCAG modules to select the informative multiscale feature in the backbone final three layers by cross-scale connection. This operation can greatly improve the feature representation ability of the small hot spot defect in PV farm IR images.
3) The proposed RCAG-Net has strong competitiveness in hot spot defect classification and detection performance. Furthermore, the speed of RCAG-Net is fast, which has great potential to be applied in practical hot spot defect location task.

The rest of this study is organized as follows. Section II presents the defect inspection methods, CNN-based detectors, and some attention networks. Section III overviews the designed UAV-based defect detection system. Section IV introduces the details of the proposed approaches. Section V gives and discusses the related experimental results. Finally, conclusion is summarized in Section VI.

II. Related Works

A. Defect Inspection

Many computer vision methods have been proposed to detect defects in various application scenarios. These methods can be roughly classified into the following types: image filtering methods, feature descriptor-based methods, and CNN-based methods.

For image filtering methods, Aghaei *et al.* [4] proposed a filter-based algorithm to detect hot spot defect in IR image of PV farms, which employed the Gaussian filter and the Laplace filter to filter out the complex background information and retained the hot spot region simultaneously. Li *et al.* [5] employed an image matching method based on the Gaussian function to detect surface defects of the modules in PV farms, which achieved desirable performance.

Different from image filtering methods, descriptor-based methods mainly rely on the extracted features and classifiers for defect recognition. Gao *et al.* [6] employed a feature extractor to extract the texture features of the edge, corner, color, and blob in the PV farm IR images, which applied a classifier to divide these images into defective and nondefective. The limit of this method is that the IR image is captured by car-based system, which is slower than UAV-based system to image acquisition used in this article. Liu *et al.* [14] developed an unsupervised machine learning algorithm based on one-class clustering, which applied morphology-based approach to achieve feature extraction, and then clustered all the defects features with similar characteristic into one class to classify the defects in PV farms.

Recently, CNN-based methods have gradually become mainstream for defect detection. Li *et al.* [15] introduced a lightweight neural network for defect inspection of the PV farms, which can be used to classify several types of defects in visible images and achieve significant accuracy improvement comparing with conventional methods. Deitsch *et al.* [16] presented a novel VGG16-based methodology to automatically classify the defective PV module in solar cell electroluminescence images. Alvaro *et al.* [17] developed a two-stage network to build a robust and high-efficiency defect detection module, which first combined the camera GPS position with defect position in IR image to calculate the practical GPS position of the defects in PV farms. However, this method is less time-efficiency: one reason is that this approach is two-stage network [8], [18], which needs a Region Proposal Network to extract the suspected defect proposals in PV farm IR image, and then, another network is applied to classification of these proposals.

B. CNN Detectors

The CNN detectors attempt to classify and locate each object in the image with a right-size bounding box, which can be classified into two types: 1) region-based methods, e.g., Faster R-CNN [8] and Mask R-CNN [18] and 2) regression-based methods, e.g., Retinanet [19] and YOLOv3 [20]. Region-based methods first generate many proposals that contains defects; then, these proposals will be sent to the following network for class division. Region-based methods are more accurate, but it slightly increases computational complexity. The regression-based method is a research hot topic [21] recently, detection accuracy of which can be close to or better than that of region-based methods while maintaining fast speed. The representative regression-based method is YOLOv3, which directly divides an image into small grids to predict bounding boxes by regression. The detection approach proposed in this article is also a one-stage method, which can make a balance between speed and accuracy of defect detection.

C. Attention Network

Attention mechanism has become an indispensable part of the CNN model and has shown excellent effects in suppressing complex backgrounds and highlighting object features. Schlemper *et al.* [13] proposed a novel soft-attention mechanism, which employed gate mechanism and attention module to achieve noise feature suppression and informative feature highlighting in tissue/organ detection. Hu *et al.* [22] proposed a novel channelwise attention module, which is applied to perform dynamic channelwise feature reweighting and makes a good performance in image classification task. Fu *et al.* [23] applied self-attention mechanism to integrate contextual information of objects and suppress the noise background in scene segmentation task. Su *et al.* [24] proposed a novel complementary attention network to detect solar cell PV defects, which can realize the suppression of noise features and the highlighting of defect features by fusing channelwise features and spatial features.

TABLE I
PARAMETERS OF UAV SYSTEM

Parameter	Value
Cruising speed	5-10 m/s
Symmetrical motor wheelbase	643 mm
Size	883×886×398
Maximum signal effective distance	5 km
Mission altitude	⩽35 m
Endurance capacity	0.62 h
Weight	4.69 kg
Power type	Electric drive
Camera sensor size	123.7×112.6×127.1 mm
Frame rate	30 Hz
Field angle of view	60°horizontal, 54°vertical
Maximum resolution	640×512

III. UAV-Based Defect Detection System

At present, in the field of PV farm defect detection, most of the UAV-based systems [5], [25] are only applied to image inspection and cannot obtain the GPS position of defects. Thus, a new UAV system combined with the GPS position of IR thermal camera to calculate the hot spot defect GPS position is designed, which can obtain the rough GPS location information of the defective area and provide a reference for defective PV module maintenance and replacement. Noting that GPS-based precise defect positioning is still a challenging task, which depends on two aspects, one is the defect location in the thermal image, another is the image GPS position in PV farm. In this article, the main target is to improve the defect location accuracy in the thermal image.

Fig. 2 presents our designed UAV-based intelligent automatic defect detection system. When we give the geographic information of the PV farm and the cruising range, the UAV will automatically plan a proper cruise route to cover the whole farm [25]. The speed, altitude, and photographic frequency are controlled by the ground control station (GCS). Every image captured by the thermal IR camera on the UAV is transmitted to GCS by the 4G wireless communication network in real time. Simultaneously, image GPS position, UAV altitude, and image acquisition frequency corresponding to each image will be preserved by GCS. Then, the discriminative features will be extracted for defect location through proposed deep-learning based algorithm in computer server, and this recognition process is offline. We employ the camera GPS position recorded when taking the IR images and the defect location in PV images to calculate the defect GPS position through coordinate system transformation, which is an essential reference to workers to maintain the normal operation of the PV farms.

The PV farm IR images are collected by a digital IR thermal imaging camera (Zenmuse XT2) installed on a lightweight UAV (DJI M200 V2). The sharpness of aerial images and field of view scope of the camera have a great influence on the recognition results. Choosing a suitable field of view scope is essential to hot spot detection. As shown in Fig. 3, the larger the field of view scope is, the more multicrystalline PV modules are contained, but the more blurred the image is. The limitation of the proposed RCAG-Net is that, if the image is taken by UAV at an altitude of about 50 m, the background is very complicated, and the hot spot defect is too small, the proposed RCAG-Net cannot get a good detection effect. Thus, the maximum altitude of 35 m is selected in this article, which is a balance between the view scope and the image sharpness. The specifications of the UAV (DJI M200 V2) show that, when the GPS signal is good, the GPS vertical error of the UAV is ±0.5 m, and the horizontal error is ±1.5 m [26]; thus, the maximum error introduced in the location is estimated to 1.5 m. Noted that other parameters of UAV system are illustrated in Table I. Moreover, season, weather, surface temperature, and other factors should be considered in the process of IR image acquisition.

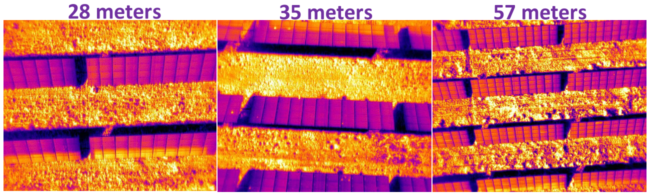


Fig. 3. Images collected by UAV at different altitudes.

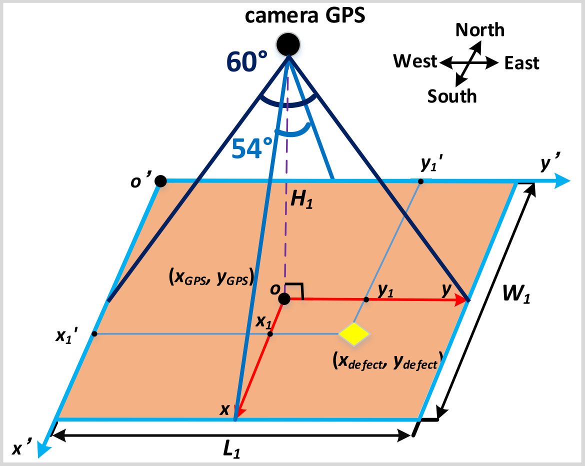


Fig. 4. GPS position of the camera relative to the GPS location of the defect.

As shown in Fig. 4, the angle at which the thermal IR camera takes the image is perpendicular to the ground. The altitude H_1 of the UAV determines the field of view scope (length: L_1, width: W_1), which corresponds to the number of PV modules included in the IR image

$$L_1 = 2H_1 \tan(60^\circ/2) \tag{1}$$
$$W_1 = 2H_1 \tan(54^\circ/2). \tag{2}$$

$x'o'y'$ is the image coordinate system, and the corresponding coordinate of the defect location is (x'_1, y'_1). xoy is the spatial coordinate system in PV farms corresponding to the image. The hot spot defect coordinate (x_1, y_1) in system xoy can be calculated by the following function:

$$x_1 = x'_1/512 \times W_1 - W_1/2 \tag{3}$$
$$y_1 = y'_1/640 \times L_1 - L_1/2 \tag{4}$$

where 512 corresponds to the height and 640 corresponds to the width of the IR image. The hot spot defect GPS position

$(x_{\text{defect}}, y_{\text{defect}})$ is defined as follows:

$$x_{\text{defect}} = x_{\text{GPS}} + x_1/d_{\text{long}} \tag{5}$$

$$y_{\text{defect}} = y_{\text{GPS}} + y_1/d_{\text{lati}} \tag{6}$$

where x_{GPS} and y_{GPS} are the longitude and latitude of the thermal IR camera, respectively, so that the GPS of coordinate origin o is the same. d_{long} is the correspondence between longitude and meter, which is 102834.7426 m per degree in this article. d_{lati} is the correspondence between latitude and meter, which is 111712.6915 m per degree. Noted that the conversion relationship between meter and longitude or latitude is related to geographical location. According to the above calculation, we can get the actual GPS position of the defect in the PV farm. The defect GPS positions are saved as a text file, which will be used to create a measurement history.

IV. Methodology

In this section, we first introduce the channelwise attention module in SENet [22], which motivates our proposed RCAG Module. Then, we present the general and lightweight additional block RCAG. Finally, the defect detection model RCAG-Net is presented, which incorporates several RCAG modules to fuse pyramidal features and suppress the complex background disturbance for small hot spot defect detection in PV farm IR images.

A. Revisiting Channelwise Attention in SENet

We first revisit one of the most popular channelwise attention units in SENet, which improves the representational power of the CNN by measuring channelwise relationships. Given the input feature $X \in \mathbb{R}^{C\times W\times H}$, where C, W, and H are the channel number, width, and height of the feature maps, respectively, the GAP layer will pool it into an 1-D vector $g(X) \in \mathbb{R}^{1\times C}$

$$g(X_k) = \frac{1}{WH}\sum_{i=1,j=1}^{W,H} X_{i,j,k}, \quad k \in \{1,\ldots,C\}. \tag{7}$$

Then, the MLP that is composed of two fully connected (fc) layers (fc-ReLU-fc) is employed to feature refinement. After being activated by the sigmoid function, the channelwise attention map $A \in \mathbb{R}^{1\times C}$ is obtained, which will multiply with the input feature $X \in \mathbb{R}^{C\times W\times H}$ to accomplish the useful feature highlighting and noise feature suppression.

The channelwise attention module in SENet is a single scale feature filter module, and the versatility of this module needs to be further enhanced. To address above two problems, we propose a novel RCAG module, which employs gate mechanism to integrate the fused multiscale feature of different pyramidal layers. Moreover, the versatility of the RCAG module can be greatly improved by the residual connection operation [27].

B. Novel RCAG

When revisiting the single-scale channelwise attention module in SENet, how to design a multiscale, general, and lightweight attention module is very necessary. Thus, we propose the novel RCAG module, which can greatly improve the performance of small hot spot defect detection in the PV farm IR image captured by the UAV-based system.

The schematic of the novel RCAG module is presented in Fig. 5. Given two different scale gate features, the multiscale feature fusion is accomplished by upsampling, convolution, addition, and activation operations. Then, the fused features will be filtered by the following convolution operation, and the GAP layer is employed to extract the global features and accomplish dimension reduction. Next, the MLP layer is combined with a sigmoid function to generate the channelwise attention map, which will multiply with the fused feature for feature reweighting. Finally, the residual connection from the fused features to the final output allows us to insert a new RCAG module into some classical pretrained models, without breaking its initial behavior, and it will improve the efficiency of the network learning channelwise attention map and prevent the gradient from disappearing during the training process.

Specially, the features from different scale layers $p \in \mathbb{R}^{C_p\times W/2\times H/2}$ and $q \in \mathbb{R}^{C_q\times W\times H}$ are transformed into two feature spaces f and g, respectively; after elementwise summation and activation with an ReLU function, the fused feature z is obtained, which can be defined as

$$z = \text{ReLU}(f(\text{upsampling}(p)) + g(q)) \in \mathbb{R}^{C\times W\times H} \tag{8}$$

where $f(p) = W_f p$ and $g(q) = W_g q$. Then, the feature z is fed into a convolutional layer to filter the multiscale features and generates another feature space h, where $h(z) = W_h z$. Next, the GAP layer is employed to dimension reduction and global features extraction, which will pool $h(z)$ into an 1-D vector $g(h(z)) \in \mathbb{R}^{1\times C}$. Next, MLP employs two fully connected layers $fc_{1\times C/r}$ and $fc_{1\times C}$ (where the first layer $fc_{1\times C/r}$ has C/r channels, the second layer $fc_{1\times C}$ has C channels, and r is the reduction ratio) with ReLU function to refine the global feature $g(z)$. Subsequently, the sigmoid function is applied to feature activation to produce the channelwise attention map

$$A = \text{sigmoid}(fc_{1\times C}(\text{ReLU}(fc_{1\times C/r}(g(h(z)))))) \tag{9}$$

where $W_f, W_g, W_h \in \mathbb{R}^{C\times W\times H}$ are weights learned by 1 × 1 convolution operations.

Each channel A_k, $k \in \{1,\ldots,C\}$ in attention map $A \in \mathbb{R}^{1\times C}$ will reweight with the feature map in $h(z) \in \mathbb{R}^{C\times W\times H}$, which is a feature reweighting process. By above calculation, the useful semantic and texture features from different layers are preserved, and the complex background will be suppressed as much as possible. Finally, we perform the residual connection operation that is an pixelwise addition with the fused feature $h(z)$ to acquire the final output feature o_f

$$o_f = \sum_{k=1}^{C} A_k h(z_k) + h(z). \tag{10}$$

The residual connection allows the attention network to learn the weights of each channel at a global view, which can greatly enhance the versatility of the attention module. Otherwise, these weights in the RCAG module can be trained during backpropagation as same as other weights in the network.

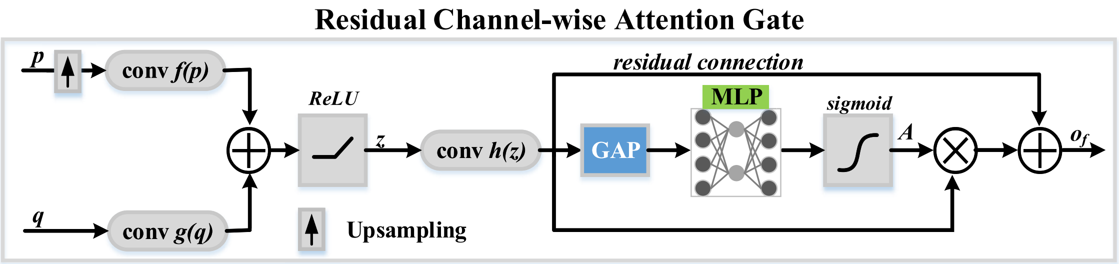


Fig. 5. Schematic of the proposed RCAG module. Small object features are selected by analyzing both the textural and sematic information provided by the gate signals (p and q), which are collected from two different scales.

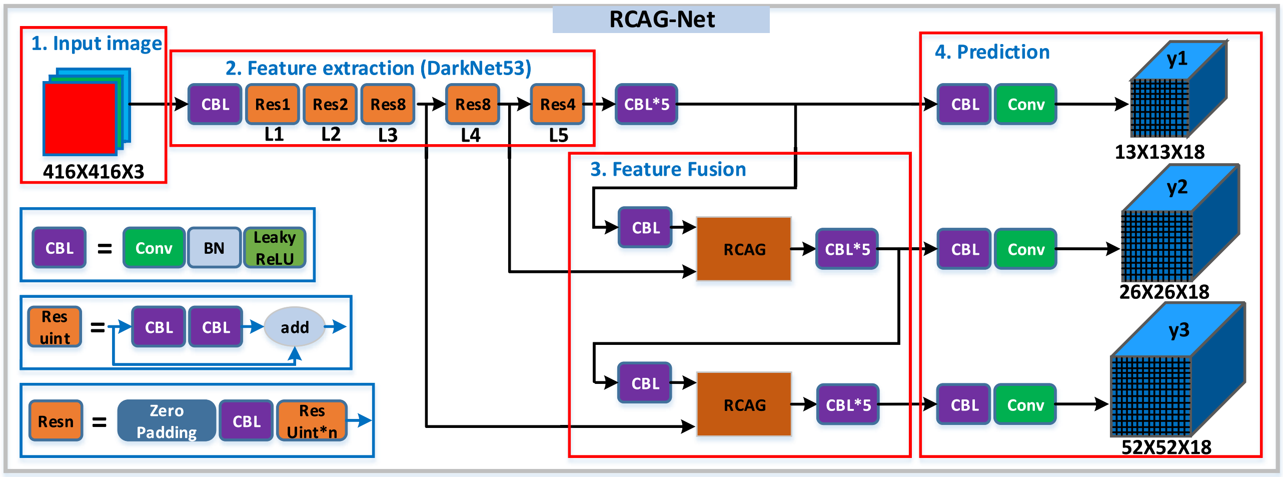


Fig. 6. Architecture of the proposed RCAG-Net. L1–L5 are multiscale feature extraction layers. y1–y3 are multiscale prediction layers.

```
Algorithm 1 Defect Detection of RCAG-Net
Input: a raw IR image with a size of $640 \times 512$ pixels
1) resize the image to $416 \times 416$ pixels
2) extract the image final three-layer features $F_3$, $F_4$ and $F_5$
3) initialize vector $F$ as $F_5$
    for $i = 1$ in range(3)
        i) utilize feature $F$ to predict the defect class $c_i$
           and location $l_i$
        ii) fuse multi-scale features $F = RCAG(F, F_{5-i})$
4) Non-Maximum Suppression (NMS) and Intersection over
Union (IoU) are used to output the suitable defect location
$l_i$
Output: class $c$ and position $l$ of the defect
```

C. Overview of the Proposed Defect Detection Architecture

The RCAG module has a similar effect as the attention module in SENet [22], which can selectively emphasize informative features and suppress disturbed ones. Fig. 6 presents the overall architectures of the proposed RCAG Network (RCAG-Net), and the pseudocode is presented in Algorithm 1. The proposed RCAG-Net employs two novel RCAG modules to obtain better feature representation ability for the small hot spot defect with the complex background disturbance, which can be divided into four major components: i) input image; ii) feature extraction; iii) feature fusion; and iv) prediction. Noted that, except for the RCAG block, the architecture of the RCAG-Net is designed by referring to YOLOv3 [20], which is one of the most popular deep learning object detectors in practical applications as the detection accuracy and speed are well balanced [28].

1) Input Image: The raw IR images of the PV farm and corresponding ground truths are input to the network, which will extract the texture and semantic features. The larger the image size, the higher the computational power and memory required. Compared with the size of the original image (640×512), the processing speed of 416×416 images will be faster, and its accuracy will hardly be affected. In addition, the size of the input model image must be a multiple of 32 ($416/32 = 13$). Because the network downsamples five times, each sampling step is 2, so the maximum step of the network (step refers to the input size divided by the output) is $2^5 = 32$, as shown in Fig. 6. By the above analysis, the input IR image with original resolution 640×512 pixels is resized to 416×416 pixels through bilinear interpolation, and the corresponding ground truths should be adjusted relevantly.

2) Feature Extraction: DarkNet53 [20] as the backbone of the RCAG-Net is employed to extract the informative feature of the IR image in PV farms, which applies the 1×1 and 3×3 convolutional operations [Convolution-BatchNormalization-LeakyReLU (CBL)] with the residual connection to filter out the image features. The architecture of the DarkNet53 is presented in Table II, which includes

TABLE II

ARCHITECTURE OF THE DARKNET53

	Type	Filters	Size	Output
	Convolutional	32	3×3	416×416
	Convolutional	64	3×3/2	208×208
1×	Convolutional	32	1×1	
	Convolutional	64	3×3	
	Residual			208×208
	Convolutional	128	3×3/2	104×104
2×	Convolutional	64	1×1	
	Convolutional	128	3×3	
	Residual			104×104
	Convolutional	256	3×3/2	52×52
8×	Convolutional	128	1×1	
	Convolutional	256	3×3	
	Residual			52×52
	Convolutional	512	3×3/2	26×26
8×	Convolutional	256	1×1	
	Convolutional	512	3×3	
	Residual			26×26
	Convolutional	1024	3×3/2	13×13
4×	Convolutional	512	1×1	
	Convolutional	1024	3×3	
	Residual			13×13

five-scale layers. Every layer is composed of convolution and residual connection to extract features, and the downsampling operation is accomplished by the convolutional layer using a 3×3 kernel with stride = 2. DarkNet53 is much more effective to extract informative features in different pyramidal layers, which will be fused through the following attention-based multiscale feature fusion module (RCAG).

3) Feature Fusion: The output features of the final three pyramidal layers (L3–L5) will be integrated together by the novel RCAG module with recurrent operations. Low-level feature maps contain rich textural details of small objects, while high-level feature maps include much semantic information of the large object. RCAG module utilizes gate mechanism combining with channelwise attention to integrate different pyramidal feature maps, which greatly enhances the feature representation ability for small hot spot defect detection under the complex background disturbance in the PV farm IR images. Noted that, except for the multiscale feature fusion module based on the novel attention module (RCAG), other aspects are similar to YOLOv3.

4) Prediction: Three prediction headers absolutely employ three feature maps with different scales to predict the objects with different sizes. A small object is predicted by the large scale headers, such as $y2$ and $y3$, which includes more details of a small object. A large object is predicted by the small-scale header such as $y1$, which includes more information of large object after feature downsampling.

During prediction, the input image is divided into several grids, the number of which is the product of height and width of final output feature maps. The network divides each thermal image in the training data set into S^2 grids. If the center of the hot spot ground truth falls into a grid, then the grid is responsible for detecting the object. Each grid in the detection header is assigned with three types of anchors (header y1: 10×12, 11×9, and 16×16; header y2: 7×11, 8×10, and 9×8; and header y3: 6×8, 7×8, and 7×9), which are responsible to predict three bounding boxes, respectively. In other words, every grid is applied to predict three defective boxes that consist of four coordinate offsets, one confidence, and one class predictions. Thus, the output result of the prediction header has a size of $S\times S\times(3\times(4+1+1))$, where $S\times S$ represents the resolution of the final convolutional feature map.

5) Loss Function: The loss function is used to evaluate the difference between the predicted value and the ground truth. The smaller the loss function is, the better the performance of the model is. In proposed RCAG-Net, the loss function for each prediction header consists of the coordinate error, the intersection over union (IoU) error, and the classification error, which is denoted as follows:

$$\text{loss} = \text{Error}_{\text{coord}} + \text{Error}_{\text{iou}} + \text{Error}_{\text{cls}}. \tag{11}$$

The coordinate prediction error Error_coord is denoted as follows:

$$\begin{aligned}\text{Error}_{\text{coord}} = {} & \lambda_{\text{coord}}\sum_{i=1}^{S^2}\sum_{j=1}^{B} I_{ij}^{\text{obj}}[(x_i-\widehat{x_i})+(y_i-\widehat{y_i})] \\ & +\lambda_{\text{coord}}\sum_{i=1}^{S^2}\sum_{j=1}^{B} I_{ij}^{\text{obj}}[(w_i-\widehat{w_i})+(h_i-\widehat{h_i})]\end{aligned} \tag{12}$$

where λ_{coord} is the weight of coordinate error, S^2 is the number of grids, whose value is square of height or width of the prediction layers (y1–y3), referring to Fig. 6, $S^2 = 13\times13, 26\times26, 52\times52$ for the three-scales prediction layers, respectively, and B is the number of bounding boxes generated by each grid. $\lambda_{\text{coord}} = 5$ and $B = 3$ are selected in this study. $I_{ij}^{\text{obj}} = 1$ denotes that the hot spot defect falls into the jth bounding box in grid i; otherwise, $I_{ij}^{\text{obj}} = 0$. $(\widehat{x_i}, \widehat{y_i}, \widehat{w_i}, \widehat{h_i})$ are values of the center coordinate, height, and width of the predicted bounding box. (x_i, y_i, w_i, h_i) are the true bounding box values.

The IoU error $\text{Error}_{\text{iou}}$ is defined as follows:

$$\text{Error}_{\text{iou}} = \sum_{i=1}^{S^2}\sum_{j=1}^{B} I_{ij}^{\text{obj}}(C_i-\widehat{C_i})^2+\lambda_{\text{noobj}}\sum_{i=1}^{S^2}\sum_{j=1}^{B} I_{ij}^{\text{obj}}(C_i-\widehat{C_i})^2 \tag{13}$$

where λ_{noobj} is the weight of the IoU error. λ_{noobj} is assigned with 0.5 in this study. $\widehat{C_i}$ is the predicted confidence, which is the likelihood that the grid i contains object. C_i is the true confidence. The classification error $\text{Error}_{\text{cls}}$ is defined as follows:

$$\text{Error}_{\text{cls}} = \sum_{i=1}^{S^2}\sum_{j=1}^{B} I_{ij}^{\text{obj}} \sum_{c\in\text{classes}} (p_i(c)-\widehat{p_i}(c))^2 \tag{14}$$

where c donates the class of the detected object. $p_i(c)$ represents the true probability that the object belongs to class c in grid i. $\widehat{p_i}(c)$ is the predicted value. The $\text{Error}_{\text{cls}}$ for grid i is the total of classification errors for all the objects in the grid. By constraint optimization through the sum of the coordinate error, the IoU error, and the classification error, the deep learning model will gradually converge and achieves the desired performance.

TABLE III
DISTRIBUTION OF THE PV FARM IR IMAGE DATA SET

Dataset	Defective images	Good images	Total
Training	790	\	790
Testing	770	3500	4270
Total	1560	3500	5060

6) Weight Update: The purpose of updating weights through backpropagation is to minimize the loss. When the proposed RCAG-Net, including the RCAG module, is trained, the loss is taken as a function of the weight parameters, and CNN needs to calculate the partial derivative of the loss with respect to each weight parameter and then uses the stochastic gradient descent (SGD) method to iteratively update weights in the direction where the gradient descents fastest, until the conditions for the parameter to stop updating are met. The weight parameters of CNN can be millions or even more than 100 million. The backpropagation algorithm can efficiently calculate the partial derivatives through the reverse derivation mode, which can greatly accelerate the weight learning.

V. EXPERIMENTS

To evaluate the effectiveness of the proposed methods, we carry out extensive experiments on our PV farm IR image data set. The experimental results verify that the proposed RCAG-Net performs much better than the previous approach (YOLOv3) in the challenging small hot spot defect detection task. Especially, the hot spot defect is easy to be disturbed by the complex background, such as poor contrast and the similar background region. In Section V, we introduce our data set and the implementation details, and then, a series of experimental results and discussions are presented to demonstrate the proposed methods.

A. Data Set

Our data set includes 700 defect-free images and 312 defective images, which was collected from a PV farm by a thermal IR camera mounted on a UAV. The resolution of these images is 640 × 512. The average size of the defect ground truth is 112 pixels $\leq 32^2$ pixels, which is defined as a small object in MS COCO data set [9]. These defective images are divided into 158 training images and 154 testing images. The Gaussian blur, contrast normalization, image sharpening, and image mirroring are employed to augment the data set, which can increase the diversity of samples, thereby enhancing the robustness of the hot spot defect detection model. The distribution of the PV farm IR image data set is illustrated in Table III.

B. Evaluation Metrics

The performance of classification is assessed by the following indexes, such as precision (P), recall (R), and F-measure (F). Moreover, average precision (AP), mean IoU (MIoU), the number of parameters, and speed are applied to evaluate the effectiveness of defect detection

TABLE IV
HYPERPARAMETERS OF RCAG-NET

Training	data_format	label_tool	epochs
	VOC2007	LabelImg	500
	input_shape	batch_size	val_split
	(416, 416)	8	0.1
Testing	score	iou	gpu_num
	0.1	0.45	2

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{TN}}, \quad \text{Recall} = \frac{\text{TP}}{\text{TP} + FP} \tag{15}$$

$$F - \text{measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{16}$$

$$\text{IoU} = \frac{\text{detection result} \ \cap \ \text{ground truth}}{\text{detection result} \ \cup \ \text{ground truth}} \tag{17}$$

where TP represents true positive; TN represents true negative; and FP represents false positive. detection result is the prediction box of the defects; ground truth is the manual annotation of the real defect position in training image.

C. Implementation Details

The code of the proposed RCAG-Net architecture is accomplished by Keras (v2.24). We apply the k-means algorithm to calculate the most suitable anchors of our IR image data set, which are 6 × 8, 7 × 8, 7 × 9, 7 × 11, 8 × 10, 9 × 8, 10 × 12, 11 × 9, and 16 × 16. Due to the memory constraint of our server, the batch size is set to 8, and the iterative epochs are set to 500. The proposed algorithm runs on a server with an Intel CPU (i7-6700 K, 4.00 GHz) and two NVIDIA GeForce GTX 1080 GPUs. The input images are resized to 416×416. Each header predicts three boxes at different scales for four bounding box offsets, one confidence prediction, and one class score prediction. Thus, the output dimensions of the final prediction layers (y1, y2, y2) at three different scales are $13 \times 13 \times [3 \times (4+1+1) = 18, 26 \times 26 \times 18, 52 \times 52 \times 18$.

In this study, we make the ground truth of the hot spot defect via a data set annotation software (LabelImg) in the VOC2012 format. A defect corresponds to a tightly enclosed box, without too much expert experience. The standard VOC2012 format data set can ensure a fair comparison between different detectors. The VOC2012 format data set includes three files: 1) Annotations; 2) ImageSets; and 3) JPEGImages. The Annotations file is mainly composed of the xml file, which includes defect ground truth information. The ImageSets file includes the Main file, in which four text files (train.txt, test.txt, trainval.txt, and val.txt) divide the data set. The original thermal images are placed in the JPEGImage file. Moreover, the hyperparameters' information of the RCAG-Net model is presented in Table IV.

D. Evaluation

In this section, the qualitative result presentation and the quantitative experimental analysis are carried out to assess the performance of our proposed algorithm (RCAG-Net).

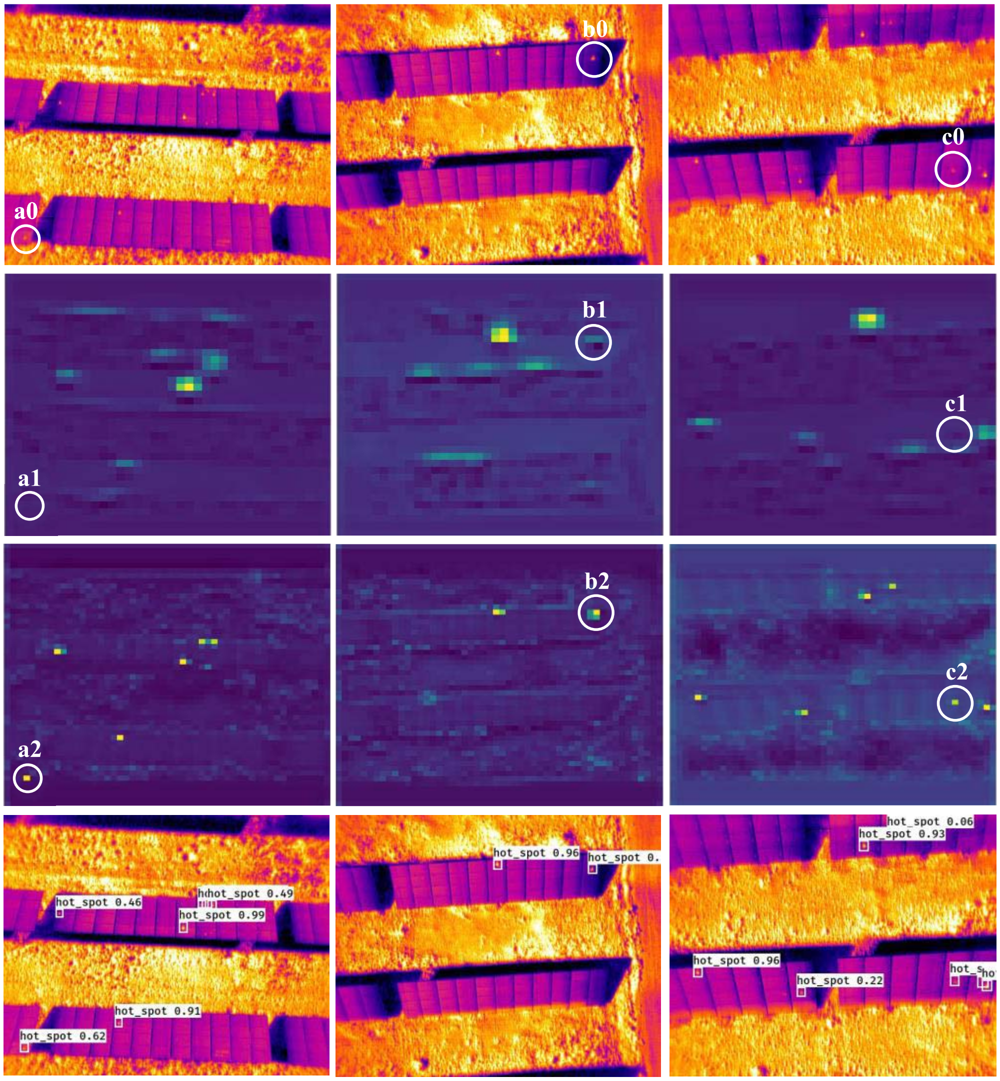


Fig. 7. Visualization of the RCAG output feature maps. First row: raw IR image captured by UAV. Second row: output feature map obtained by the first RCAG module in RCAG-Net (resolution: 26 × 26 pixels resize to 640 × 512 pixels). Third row: output feature maps obtained by the second RCAG module in RCAG-Net (resolution: 52 × 52 pixels resize to 640 × 512 pixels). Fourth row: hot spot defect detection results.

1) Visualization Analysis of the Feature Maps: The RCAG module plays the most important role in the proposed RCAG-Net, which can guide multiscale feature fusion and suppress the noise background features simultaneously. To explore the performance of the proposed RCAG module in the proposed RCAG-Net, we visualize the output feature maps, in which the hot spot defect presents highlight regions, as shown in Fig. 7. Feature maps can help us to better understand what has been learned during the training process.

For example, hot spot defect **a0** as shown in the white circles of Fig. 7 is a relatively small object, the corresponding output feature **a1** in the first RCAG module that is employed to integrate the output feature of **L4** and **L5** has vanished, and the reason is that as the network goes deeper, the feature of small hot spot defect will be weakened by convolution and down-sampling operations until it disappears. Furthermore, feature vanishment will lead to the small defect undetectable, while the second RCAG module, which fuses output features of the first RCAG module and the output features of **L3**, presents complete defect information **a2** that is necessary to the precise hot spot defect detection in the PV farm IR image. Low-level features include more textural and colorful details that are beneficial to small hot spot defect recognition. RCAG receives these details from **L3** and integrates it with the high-level semantic features of the first RCAG module; therefore, rich textural and semantic information is preserved at the same time, which will be applied to accurately predict the small hot spot defect in IR image. The fourth row in Fig. 7 shows the hot spot detection results, which verifies that the proposed RCAG-Net can make an excellent performance on the small hot spot defect detection task, in which the RCAG module plays an important role to guide multiscale feature fusion of the pyramidal layers and improve the feature representation ability of small hot spot defects in PV farm IR images. Moreover, hot spots **b0** and **c0** present similar characteristic with **a0**.

A key problem of YOLOv3 [20] is that if the low-level layer is directly applied to concatenate with the high-level layer, lots

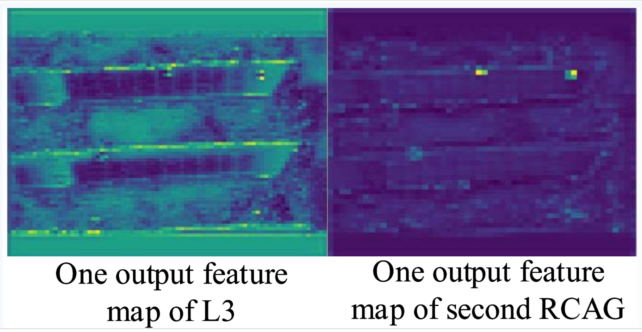

One output feature map of L3
One output feature map of second RCAG

Fig. 8. Difference between the output feature map of L2 and the second RCAG module in the proposed RCAG-Net.

of redundant information in low-level features would have a negative effect on the model learned by training, which can be solved by the proposed RCAG module. Fig. 8 shows the visual difference between the output feature map of **L3** and the second RCAG module in the proposed RCAG-Net. The **L3** output feature map except for the hot spot information also contains many complex background feature, which will have a negative impact on the defect detection results. However, the output feature map from the second RCAG module is less disturbed by the complex background feature, which illustrates that the proposed RCAG module can effectively suppress the disturbance of the complex background during the pyramidal feature fusion process and guide the detection model to focus on the hot spot defects under the complex background disturbance in the PV farm IR images, which is beneficial to improve the detection result of small hot spot defect in PV farm IR image.

2) Visualization of More Detection Results: More visualization results are shown in Fig. 9. The hot spot defects are very difficult to identify by the naked eyes. On the one hand, the defects are very small, and on the other hand, they are seriously interfered with by the complex background. However, the proposed RCAG-Net performs accurate position prediction under the interference of noise background, which illustrates that the RCAG module performs effectively to multiscale feature fusion and complex background suppression of hot spot defect detection in PV farm IR images. As shown in Fig. 9, the detection results include the defective box, class, and confidence. The score threshold of the confidence is set to 0.1 in Table IV, which can ensure that the hot spot defect is not easy to be missed during the process of practical intelligent fault elimination in PV farms. In short, we can conclude from the above presentation results that the proposed RCAG-Net framework performs precisely hot spot defect position prediction under the disturbance of noise background.

3) Quantitative Evaluation: The precision (P), recall (R), F-measure (F), AP, and MIoU of original YOLOv3 and the proposed RCAG-Net with different backbones (DarkNet19 [20], MobileNet [29], VGG16 [30], and DarkNet53 [20]) on our PV farm IR image data set are shown in Table V. Noted that RCAG-Net is designed by referring to YOLOv3; thus, we compare the proposed RCAG-Net with YOLOv3 by

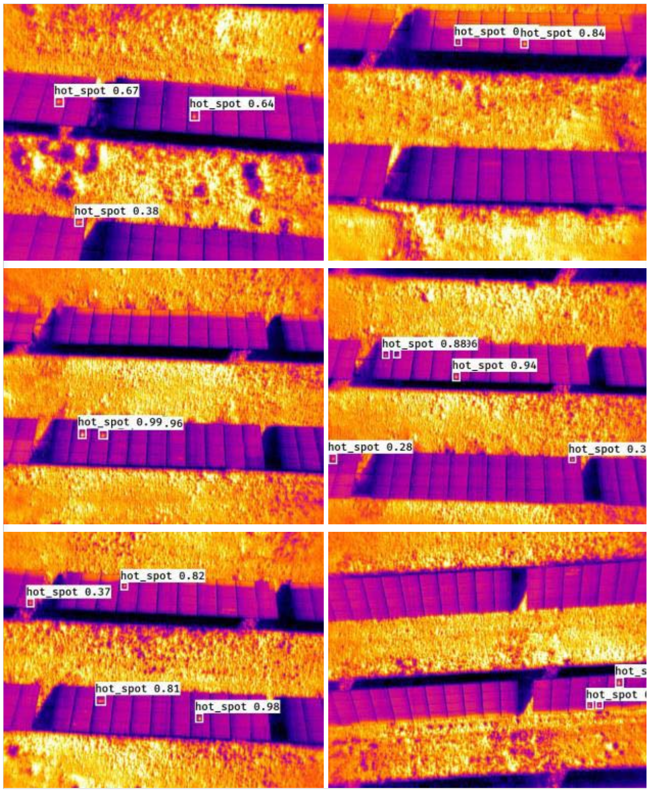


Fig. 9. Visualization results of proposed RCAG-Net on raw PV farm IR images captured by the UAV system.

employing different backbones to illustrate the effectiveness and versatility of the proposed RCAG module, which can be utilized to improve the feature representation ability of small hot spot defect under complex background disturbance. As shown in Table V, the experimental results are presented objectively in terms of image-level classification and detection on our PV farm IR image data set.

For the image-level classification task of small hot spot defect, the RCAG module improves the performance remarkable, which illustrates that the RCAG module can promote the classification results of many popular CNN models, such as DarkNet19, VGG16, MobileNet, and DarkNet53, which are employed to extract the textural and sematical features of PV farm IR image. As shown in Table V, the proposed RCAG-Net achieves 2.40%, 2.99%, 2.95%, and 3.83% hit rates of F-measure improvement from YOLOv3 corresponding to DarkNet19, MobileNet, VGG16, and DarkNet53, respectively, which illustrates that, as a general attention module, the RCAG module can be widely used to fuse multiscale features and boost the feature representation ability of small hot spot defect under complex background disturbance. Moreover, when DarkNet53 is used to be the backbone, the best experimental results are achieved by the proposed RCAG-Net, the performance is improved to 92.61% for F-measure, and the recall of the hot spot defect reaches 97.06%, which means that the defect image is not easy to be missed during the image automatic classification process.

For detection of small hot spot defect, we apply AP, MIoU, the number of parameters, and speed to evaluate

TABLE V

EXPERIMENTAL RESULTS OF THE YOLO-BASED METHODS (YOLOV3 AND THE PROPOSED RCAG-NET) WITH DIFFERENT BACKBONES ON OUR PV FARM IR IMAGE DATA SET

Detector	Backbone	Classification			Detection			
		P(%)	R(%)	F(%)	AP(%)	MIoU(%)	Parameter	Speed(ms)
YOLOv3 [20]	DarkNet19	77.65	79.14	78.38	40.27	18.35	8.27M	25.91
	MobileNet	84.27	91.37	87.68	76.11	43.01	23.06M	34.63
	VGG16	84.09	93.27	88.44	79.08	47.18	42.81M	42.49
	DarkNet53	84.18	93.92	88.78	79.98	48.33	58.72M	88.39
RCAG-Net (ours)	DarkNet19	79.68	81.91	80.78	45.85	23.08	8.21M	25.65
	MobileNet	86.47	95.29	90.67	80.68	46.41	24.08M	35.36
	VGG16	87.46	95.39	91.39	81.36	47.82	43.85M	47.62
	DarkNet53	88.55	97.06	92.61	84.64	51.34	59.74M	90.45

TABLE VI

EXPERIMENTAL RESULTS OF DIFFERENT DETECTION METHODS WITH THE SAME BACKBONE (VGG16)

Detector	Classification			Detection			
	P(%)	R(%)	F(%)	AP(%)	MIoU(%)	Parameters	Speed(ms)
RatinaNet [19]	91.78	65.69	76.57	51.48	32.38	22.32M	38.81
RetinaNet+the proposed RCAG	92.16	71.45	80.49	53.37	35.12	22.83M	39.01
Faster R-CNN [8]	81.23	83.56	82.38	68.70	43.12	260.50M	155.13
RCAG-Net (Ours)	87.46	95.39	91.39	81.36	47.82	43.85M	47.62

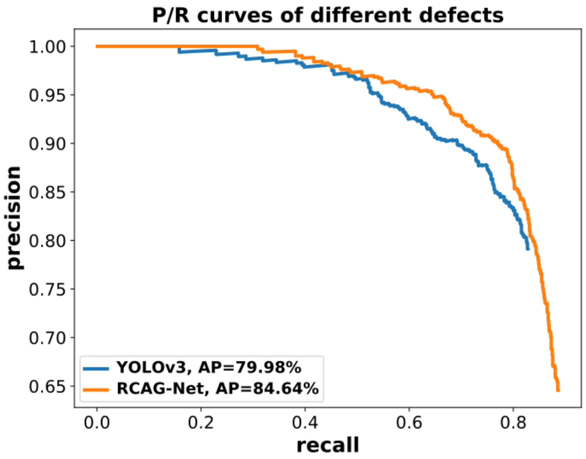


Fig. 10. P/R curves of YOLOv3 and the proposed RCAG-Net with the same backbone (DarkNet53).

the performance of the proposed RCAG-Net architecture. As shown in Table V, in the case of the same backbone, the proposed RCAG-Net performs better than YOLOv3 in hot spot defect detection task. Comparing with the YOLOv3 (DarkNet53), RCAG-Net (DarkNet53) improves AP and MIoU by 4.66 and 3.01 points, respectively. The P/R curves of YOLOv3 and the proposed RCAG-Net corresponding to the same backbone (DarkNet53) are shown in Fig. 10. The AP value is the enclosed area of the curve and axes. From the above comparisons, we can conclude that RCAG-Net is superior to YOLOv3 to detect small hot spot defects, and the RCAG module shows good versatility for different backbones. Moreover, time-efficiency evaluation plays a vital role in the process of intelligent defect detection. With the same input image size (416 × 416 pixels), the RCAG-Net is similar to the previous YOLOv3 in terms of parameter number and speed, which verifies that the proposed RCAG module is lightweight and only slightly increases the complexity and computational burden of the model.

TABLE VII

COMPARED WITH SOME TRADITIONAL METHODS

Method	P (%)	R (%)	F (%)
HOG +SVM [32]	14.81	58.82	23.67
CPICS-LBP+SVM [34]	28.63	64.55	39.66
LBP+SVM [33]	29.31	68.36	41.03
AE-CLBP+SVM [35]	39.53	73.82	51.49
RCAG-Net (DarkNet53)	88.55	97.06	92.61

4) Comparison With Other Detection Methods: In this section, we conduct the quantitative evaluation in hot spot defect detection using RetinaNet [19] and Faster R-CNN [8], which employ the same backbone (VGG16) to extract the feature of PV farm IR image. RetinaNet is a one-stage and multiscale detector, which can be applied to combine with the proposed RCAG module to achieve small hot spot defect detection. As compared in Table VI, RetinaNet+, the proposed RCAG, outperforms the original RetinaNet by 3.92%, 1.89%, and 2.74% in terms of F-measure, AP, and IoU, respectively. The results demonstrate that the proposed RCAG module can be well extended to other object detection frameworks using lower model complexity.

Furthermore, except for RetinaNet, we compare our RCAG-Net with the single-scale detector Faster R-CNN, which is a popular two-stage network and has achieved well performance in defect detection task [24]. In terms of the hot spot classification task, the proposed RCAG-Net achieves 14.82% and 9.01% F-measure improvement than RetinaNet and Faster R-CNN, respectively, which illustrates that the proposed RCAG-Net is the better at classifying defective IR image of the PV farm. In terms of position prediction, we calculate MIoU to evaluate the effectiveness of position prediction. The proposed RCAG-Net achieves 15.44% and 4.70% MIoU improvement than RetinaNet and Faster R-CNN, respectively, which illustrates that the proposed RCAG-Net is more accurate in locating the hot spot defect in PV farm IR image. Otherwise, the recall rate of RCAG-Net (95.39%) is higher than RetinaNet and Faster R-CNN, which means

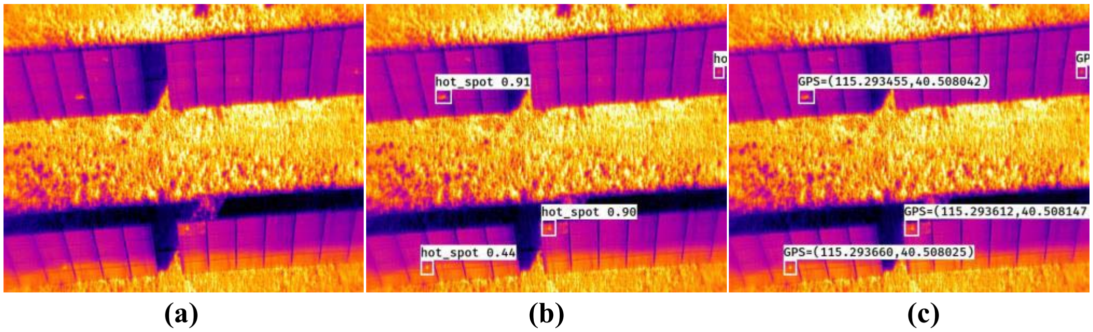

hot_spot 0.91
ho
hot_spot 0.90
hot_spot 0.44
GPS=(115.293455,40.508042)
GP
GPS=(115.293612,40.508147
GPS=(115.293660,40.508025)
(a)
(b)
(c)

Fig. 11. Relative GPS positions of hot spot defects in PV farm IR image. (a) Raw image. (b) Detection results. (c) Relative GPS position.

TABLE VIII
RELATIVE GPS COORDINATES LOCATION

$T/R/B/L$	center coordinate (x1', y1')	GPS Latitude/Longitude	real GPS position	error
173/89/191/108	(182, 98.5)	(115.293410°, 40.508067°)	(115.293414°, 40.508062°)	0.96 m
616/110/635/132	(625.5, 121)	(115.293426°, 40.508318°)	(115.293421°, 40.508321°)	0.87 m
97/453/120/480	(108.5, 466.5)	(115.293660°, 40.508025°)	(115.293668°, 40.508021°)	1.31 m
311/381/337/411	(324, 374)	(115.293612°, 40.508147°)	(115.293606°, 40.508152°)	0.97 m
123/151/152/177	(137.5, 164)	(115.293455°, 40.508042°)	(115.293459°, 40.508050°)	1.27 m

that the undetected error of hot spot defects is low in the practical detection process, which is essential to hot spot defect elimination.

5) Comparison With Some Traditional Methods: For the image classification task, CNN performs much better than traditional methods in recent years [31]. The data set of good images is divided into 800 and 2700 for training and testing the traditional descriptor-based methods, such as histogram of oriented gradient (HOG) [32], local binary patterns (LBPs) [33], Center Pixel Information Center Symmetric LBP (CPICS-LBP) [34], and adjacent evaluation completed LBP (AE-CLBP) [35]. The distribution of the defective images is the same as in Table III. We compared our approach with the above descriptor-based methods to validate the effectiveness of our RCAG-Net. The support vector machine (SVM) with polynomial kernel function is utilized as the base classifier to classify the images. As shown in Table VII, the performance of RCAG-Net is 68.94%, 52.95%, 51.58%, and 41.12% better than HOG+SVM, CPICS-LBP+SVM, LBP+SVM, and AE-CLBP+SVM, respectively, which illustrates that the proposed RCAG-Net performs better than traditional methods in the classification task of the PV farm IR images.

E. GPS Coordinate Transformation

The final GPS coordinates location presents in Table VIII, where $H_1 = 34.99$ m, $x_{\text{GPS}} = 115.293517°$, and $y_{\text{GPS}} = 40.508145°$. The rows denote different defects in the image sample. The first column is the predicted position of the defects in the image coordinate system. T and L represent the coordinates of the top left corner in the IR image detected box, and B and R represent the coordinates of the bottom right corner. The second column defines the center coordinates position of the detected box, which is (x_1', y_1'), where $x_1' = (T + B)/2$, $y_1' = (L + R)/2$. The third column defines the predicted GPS latitude and longitude of the defect. The fourth is the real GPS location of the defects, and the predicted error is shown in the fifth column. If the telemetry data are not accurate enough, the results will be different. Fig. 11 shows the example of final GPS coordinates of the hot spot defects in the PV farm IR image. The detected box in the IR image combines with the GPS location when capturing the image to calculate the defect real GPS position in the PV farm.

The positioning method was verified using a handheld GPS thermal IR camera. We verify all the predicted hot spot GPS positions through manual checking. The measurement error represents the distance from the predicted GPS position to the hot spot defect real GPS position. Referring to Table VIII, the average measurement error is 1.08 m, and all the errors are below the mix estimated threshold of 1.5 m. The length of the PV module cluster is longer than this threshold, and there is a certain distance between different clusters; thus, the location is relatively accurate. Since positioning errors are inevitable, this method cannot guarantee the precise positioning of all defects in PV modules. This measurement error is not a major disadvantage because all detected hot spots must be manually checked before any repairs are performed. In addition, the rough GPS location information of the defective PV module can provide a reference for defective PV module replacement and helps improve the efficiency of maintenance.

VI. CONCLUSION AND DISCUSSION

This article presents a UAV-based intelligent location system, which can feedback the hot spot defect GPS position in PV farms. In addition, we proposed a novel hot spot defect detection framework (RCAG-Net) that adopts a novel attention module (RCAG) to aggregate the pyramidal features of different scales by gate mechanism. The novel RCAG module can effectively fuse multiscale features, suppress the complex background information, and improve the feature representation ability of the CNN models. Experimental results show that RCAG is an extremely lightweight and general module to improve the performance of some deep CNN detectors with different backbones to detect the small hot spot defect in PV farm IR images. Code is available at https://github.com/binyisu/RCAG.

As part of future works, one or more defects appear in adjacent frames multiple times; thus, the repeated areas in adjacent frames need to be processed before outputting the results, which is the focus of our future research.

REFERENCES

[1] International Energy Agency. (2019). *Renewables 2019*. [Online]. Available: https://www.iea.org/reports/renewables-2019

[2] G. Cipriani *et al.*, “Application of thermographic techniques for the detection of failures on photovoltaic modules,” in *Proc. IEEE Int. Conf. Environ. Electr. Eng. IEEE Ind. Commercial Power Syst. Eur.*, Jun. 2019, pp. 321–329.

[3] M. Aghaei, P. B. Quarter, F. Grimaccia, and S. Leva, “Unmanned aerial vehicles in photovoltaic systems monitoring applications,” in *Proc. Eur. Photovolt. Solar Energy 29th Conf. Exhib.*, 2014, pp. 2734–2739.

[4] M. Aghaei, F. Grimaccia, C. A. Gonano, and S. Leva, “Innovative automated control system for PV fields inspection and remote control,” *IEEE Trans. Ind. Electron.*, vol. 62, no. 11, pp. 7287–7296, Nov. 2015.

[5] X. Li, Q. Yang, Z. Chen, X. Luo, and W. Yan, “Visible defects detection based on UAV-based inspection in large-scale photovoltaic systems,” *IET Renew. Power Gener.*, vol. 11, no. 10, pp. 1234–1244, Aug. 2017.

[6] X. Gao, E. Munson, and G. P. Abousleman, “Automatic solar panel recognition and defect detection using infrared imaging,” *Proc. SPIE*, vol. 9476, May 2015, Art. no. 94760O.

[7] H. Han, C. Gao, Y. Zhao, S. Liao, L. Tang, and X. Li, “Polycrystalline silicon wafer defect segmentation based on deep convolutional neural networks,” *Pattern Recognit. Lett.*, vol. 130, no. 2, pp. 234–241, Feb. 2020.

[8] 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, Jun. 2017.

[9] T. Lin *et al.*, “Microsoft COCO: Common objects in context,” in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 740–755.

[10] H. Li, K. Lin, J. Bai, A. Li, and J. Yu, “Small object detection algorithm based on feature pyramid-enhanced fusion SSD,” *Complexity*, vol. 2019, no. 10, Oct. 2019, Art. no. 7297960.

[11] W. Liu *et al.*, “SSD: Single shot multibox detector,” in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2016, pp. 21–37.

[12] H. Dong, K. Song, Y. He, J. Xu, Y. Yan, and Q. Meng, “PGA-net: Pyramid feature fusion and global context attention network for automated surface defect detection,” *IEEE Trans. Ind. Informat.*, vol. 16, no. 12, pp. 7448–7458, Dec. 2020, doi: 10.1109/TII.2019.2958826.

[13] J. Schlemper *et al.*, “Attention gated networks: Learning to leverage salient regions in medical images,” *Med. Image Anal.*, vol. 53, pp. 197–207, Apr. 2019.

[14] S. Liu, L. Dong, X. Liao, Y. Hao, X. Cao, and X. Wang, “A dilation and erosion-based clustering approach for fault diagnosis of photovoltaic arrays,” *IEEE Sensors J.*, vol. 19, no. 11, pp. 4123–4137, Jun. 2019.

[15] X. Li, Q. Yang, Z. Lou, and W. Yan, “Deep learning based module defect analysis for large-scale photovoltaic farms,” *IEEE Trans. Energy Convers.*, vol. 34, no. 1, pp. 520–529, Mar. 2019.

[16] S. Deitsch *et al.*, “Automatic classification of defective photovoltaic module cells in electroluminescence images,” *Sol. Energy*, vol. 185, pp. 455–468, Jun. 2019.

[17] Á. H. Herraiz, A. P. Marugán, and F. P. G. Márquez, “Photovoltaic plant condition monitoring using thermal images analysis by convolutional neural network-based structure,” *Renew. Energy*, vol. 153, no. 2, pp. 334–348, Jun. 2020.

[18] K. He, G. Gkioxari, P. Dollár, and R. B. Girshick, “Mask R-CNN,” in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.

[19] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal loss for dense object detection,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020.

[20] J. Redmon and A. Farhadi, “YOLOv3: An incremental improvement,” Apr. 2018, *arXiv:1804.02767*. [Online]. Available: http://arxiv.org/abs/1804.02767

[21] M. Tan, R. Pang, and Q. V. Le, “EfficientDet: Scalable and efficient object detection,” Jul. 2019, *arXiv:1911.09070*. [Online]. Available: http://arxiv.org/abs/1911.09070

[22] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[23] J. Fu *et al.*, “Dual attention network for scene segmentation,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3146–3154.

[24] B. Su, H. Y. Chen, P. Chen, G.-B. Bian, K. Liu, and W. Liu, “Deep learning-based solar-cell manufacturing defect detection with complementary attention network,” *IEEE Trans. Ind. Informat.*, early access, Jul. 8, 2020, doi: 10.1109/TII.2020.3008021.

[25] P. B. Quater, F. Grimaccia, S. Leva, M. Mussetta, and M. Aghaei, “Light unmanned aerial vehicles (UAVs) for cooperative inspection of PV plants,” *IEEE J. Photovolt.*, vol. 4, no. 4, pp. 1107–1113, Jul. 2014.

[26] DJI. *MATRICE-200-V2 Fligth Controller*. Accessed: May 15, 2020. [Online]. Available: https://www.dji.com/cn/ matrice-200-series-v2

[27] X. Wang, R. Girshick, A. Gupta, and K. He, “Non-local neural networks,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7794–7803.

[28] P. Zhang, Y. Zhong, and X. Li, “SlimYOLOv3: Narrower, faster and better for real-time UAV applications,” Jul. 2019, *arXiv:1907.11093*. [Online]. Available: http://arxiv.org/abs/1907.11093

[29] A. G. Howard *et al.*, “MobileNets: Efficient convolutional neural networks for mobile vision applications,” Apr. 2017, *arXiv:1704.04861*. [Online]. Available: http://arxiv.org/abs/1704.04861

[30] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in *Proc. Int. Conf. Learn. Represent.*, 2015, pp. 1–15.

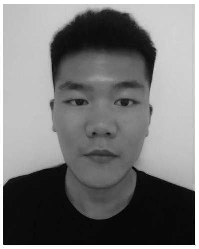
[31] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, “CNN features off-the-shelf: An astounding baseline for recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2014, pp. 512–519.

[32] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2005, pp. 886–893.

[33] T. Ojala, M. Pietikainen, and T. Maenpaa, “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.

[34] B. Su, H. Chen, Y. Zhu, W. Liu, and 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.

[35] K. Song and 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 in intelligent science and technology and the M.S. degree in automation from the School of Artificial Intelligence and Data Science, Hebei University of Technology, Tianjin, China, in 2017 and 2020, respectively. He is currently pursuing the Ph.D. degree with the School of Computer Science, Beihang University, Beijing, China.

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

Haiyong Chen received the M.S. degree in detection technology and automation from the Harbin University of Science and Technology, Harbin, China, in 2005, and the Ph.D. degree in control science and engineering 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.

Kun Liu received the M.S. degree in mechatronic engineering 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 the M.S. degree in applied mathematics and the Ph.D. degree in control theory and control engineering from Hebei University of Technology, Tianjin, China, in 2010 and 2016, respectively.

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.