Detection Method of Photovoltaic Panel Defect Based on Improved Mask R-CNN

Shuqiang Guo*, Zhiheng Wang, Yue Lou, Xianjin Li, Huanqiang Lin

School of Computer Science, Northeast Electric Power University, China
guoshuqiang@neepu.edu.cn, zhwang95@163.com, lou520yue@163.com, 752674588@qq.com, 2469002151@qq.com

Abstract

To solve the low efficiency and precision of uncrewed inspection in photovoltaic power stations, a segmentation method of improving the defective photovoltaic panels based on improved Mask R-CNN is proposed. The atrous spatial pyramid pooling and spatial attention mechanism were introduced into the extraction network to improve detection accuracy. Uncrewed aerial vehicle infrared video of the panels is used to input the network model for defect detection. As a result, the automatic annotation of the defect position is achieved, significantly improving the efficiency and precision of uncrewed inspection. The precision, recall and FPS are used as performance metrics to evaluate U-Net, PSPNet, Mask R-CNN and the algorithm in this paper. Experiments show that the detection precision of the four models are 77.3%, 82.2%, 84.0% and 89.8% respectively, the recall are 79.4%, 79.0%, 81.6% and 84.4% respectively, and the FPS are 12.5, 9.6, 8.6 and 8.2 respectively. Although the FPS of this algorithm is slightly reduced, the precision and recall have been greatly improved, and can be applied to industry.

Keywords: Photovoltaic panel defect detection, Mask R-CNN, Atrous spatial pyramid, Spatial attention

1 Introduction

At present, photovoltaic (PV) power generation technology is widely used in the whole world, and photovoltaic power generation occupies a large proportion of the total power generation in the world. Photovoltaic panel is the most important component of photovoltaic power generation system, and its good condition has a direct impact on the power generation efficiency of photovoltaic power generation system. As photovoltaic power stations are generally built in areas with relatively harsh environment and affected by bad weather (rain, snow, hail, etc.), the panels are vulnerable to damage. Defective panels can affect the power generation efficiency of the power generation system at best, and can cause safety accidents at worst. Therefore, it is very necessary to carry out defect detection on the panels regularly. Hot spot, hidden crack and breakage are common defects. Because most defects are located in small positions, it is difficult to distinguish them by naked eyes alone. Due to the advantages of low cost and high speed of inspection by Unmanned Aerial Vehicles (UAVs), it is of great research value to take infrared images by UAVs and automatically detect the position and type of defects of panels by computer. At present, deep learning is more and more widely used in industrial defect detection [1-3]. Neural network can automatically extract the features of the input aerial panel image, and fuse the extracted multi-layer feature information. Finally, the network model can learn the complex mapping relationship from the sample image input to the result output. [4-6] Because the Mask R-CNN model has high detection accuracy, this paper improves it and explores an efficient panel defect detection method.

The main contributions of the research work are summarized as follows:

(1) Atrous convolution and attention mechanism are introduced to improve the precision and recall rate of the model.

(2) The proposed model is superior to the existing network models in infrared image defect detection of panels.

2 Literature Reviews

At present, research methods for automatic identification of panel defects are mainly divided into methods based on machine vision and deep learning [7-9].

Machine vision technology is an important branch of artificial intelligence. It relies on manual feature extraction and machine learning classification for recognition and detection. [10] uses Hough transform to locate the suspicious defect area, then calculates and quantifies the characteristics of the suspicious defect area, classifies the statistical data by using support vector machine (SVM) and random forest algorithm, and obtains the defect area and fault type. [11] carries out feature analysis and statistics on the image, uses K-means clustering algorithm to segment the image, quantifies the statistical features in the image, and obtains the histogram. [12] uses OpenCV to construct the template matching algorithm training data set, and detects and classifies the infrared images after obtaining the classifier.

The above research methods have high detection efficiency and accuracy in defect detection of battery panel under specific circumstances. The research of this paper is aimed at the panels of the running power station. The detection method based on machine vision has strong ability to detect single and obvious defects. When the types of defects increase and the defects become subtle, the detection speed and accuracy of this method will be greatly reduced and the generalization ability will become weak.

*Corresponding Author: Shuqiang Guo; E-mail: guoshuqiang@neepu.edu.cn
DOI: 10.53106/160792642022032302018

Instance segmentation [13] is a major branch of computer vision. Mask R-CNN [14-15] has achieved good results in instance segmentation, target detection, human key point detection and other tasks. Due to the large number of panels in the obtained infrared image, the defect is not obvious, and it contains more external environmental interference factors and noise. When using ResNet-101 [16] as the feature extraction network for feature extraction, a large number of subsampling operations in the network will lose the detail information of the defect. Also affected by irrelevant external environmental factors in the image, the subsequent mask prediction network is easy to lose the details of the defects, resulting in inaccurate pixel segmentation. In order to realize the efficiency and high accuracy detection of photovoltaic panel defects in infrared images, an improved detection algorithm based on Mask R-CNN algorithm is proposed in this paper.

3 Methodology

The main improvements are as follows.

(1) For the conv4_x and conv5_x (Table 1) layers in feature extraction of ResNet-101, convolution kernel with different dilation rate was used to replace the convolution kernel in the original network, and the multi-scale feature map extracted by convolution kernel with different dilation rate was superimposed to fuse multi-level feature information. The feature map with a larger receptive field range can be obtained, so that the high-level feature information can be better retained, and the accuracy of mask prediction can be greatly improved.

(2) During the construction of Feature Pyramid Network (FPN) [17], the spatial attention branch network was added, and the fused feature maps were transferred to the spatial attention branch network. The weight vectors representing the importance of each region were outputted, which were superimposed to the original feature map to get a new feature map.

3.1 Atrous Spatial Pyramid Pooling (ASPP)

Compared with the ordinary convolution kernel, the dilation rate is introduced into the atrous convolution [18] kernel, which represents the distance between the effective values in the convolution kernel. The atrous convolution can increase the receptive field without changing the size of the feature image, the larger convolution kernel is used to sample the image at intervals, which can better preserve the local feature information. By using the atrous convolution kernel with different dilation rate, the feature information of receptive fields in different ranges can be obtained. The atrous convolution structure is shown in Figure 1.

Because the shape of the defect location area in the panels is irregular, and the infrared image content factors such as the interference of background is more, when the convolution kernels with large dilation rate, increase distance between sampling points tend to center location information leakage phenomenon, loss of feature information often leads to the follow-up of the pixel segmentation accuracy is low. In order to avoid this phenomenon and strengthen the extraction of feature information near the center point, the fourth neighborhood and the eighth neighborhood at the center of the atrous convolution kernel with dilation rate of 5 and 12 are set as effective sampling points. The modified atrous convolution kernel is shown in the Figure 2.

Using 3×3 convolution check images with expansion coefficients of 1, 2, 5 and 12 for feature extraction, four feature maps with larger receptive fields can be obtained, and they are superimposed together through 1×1 convolution to obtain the multi-scale fused feature map. The pyramid structure of void convolution space is shown in the Figure 3.

3.2 ResNet-101 with ASPP

This paper uses ResNet-101 as the feature extraction network. ResNet-101 uses five convolution layers (Table 1) to extract the features of the input image. Each convolution layer has two basic blocks: Conv Block [19-22] and Identity Block (Figure 4). Except conv1, other convolution layers are composed of Conv Block and Identity Block. The main path of Conv Block is connected in series with 1×1, 3×3, 1×1 convolution block, shortcut uses a 1×1 convolution block, in which the first convolution block of the main path and the convolution block of the shortcut are responsible for changing the dimension of the feature map. Because the continuous down sampling operation of high-level feature map will lose a lot of feature information, 3×3 convolution can not take into account the feature information in a wider range, resulting in the reduction of the accuracy of segmentation results. In this work, we use ASPP instead of 3×3 convolution and cancel the activation function after 3×3 convolution, which will help to extract a wider range of image detail information. Because 1×1 convolution is used in ASPP to fuse the four obtained feature maps, we remove the last 1×1 convolution and BatchNorm block in the main path. Replace the conv4_x and conv5_x layers in the original ResNet-101 with the modified Conv Block. The original ResNet-101 network structure diagram and the modified network structure diagram are shown in Table 1.

3.3 Spatial Attention Mechanism Module

In the process of building a feature pyramid, using ordinary convolution neural network for feature extraction, equal treatment to each pixel in the image, but in fact, different location, different pixels of the importance of the information is not the same. Different pixel treated for feature extraction will have great influence on the final segmentation result. In order to make full use of the importance of feature information in different locations, this paper introduces spatial attention [23-24] module—SANet in the construction of FPN. After the introduction of the spatial attention mechanism module, the feature extraction network will no longer treat every region of the image equally, but pay more attention to the region of interest.

During the construction of feature pyramid, five feature maps of different scales, P2, P3, P4, P5 and P6, were input into the branch of spatial attention module for processing. Firstly, global maximum pooling and global average pooling were carried out on the input feature map respectively to obtain the pooled two feature maps, which

were combined into a feature map, and then dimensionalized into a single-channel feature map through the convolutional layer and input into the activation function to generate the feature map of spatial attention results. At this point, the weight tensor representing the importance of different pixels is obtained, which is multiplied with the input feature graph pixel by pixel to obtain new feature maps Q2, Q3, Q4, Q5, Q6.

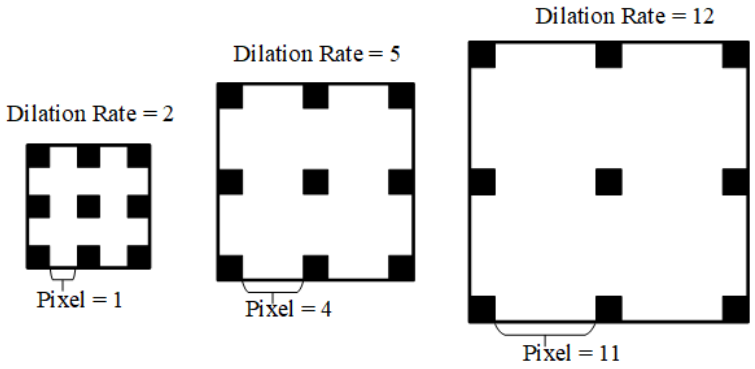


Figure 1. Atrous convolution kernel with different dilation rate

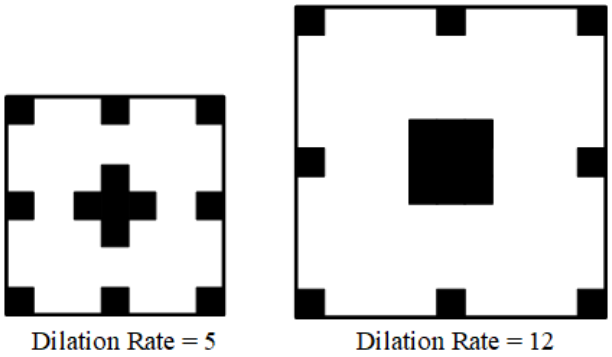


Figure 2. Convolution kernel with dilation rate of 5 and 12

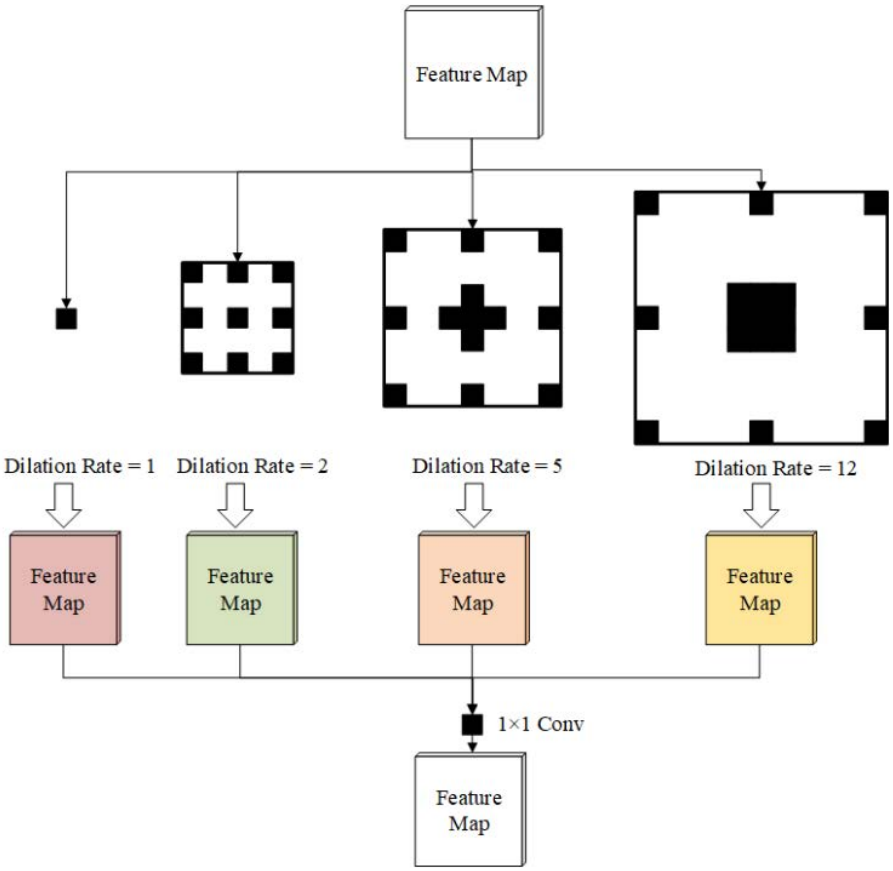
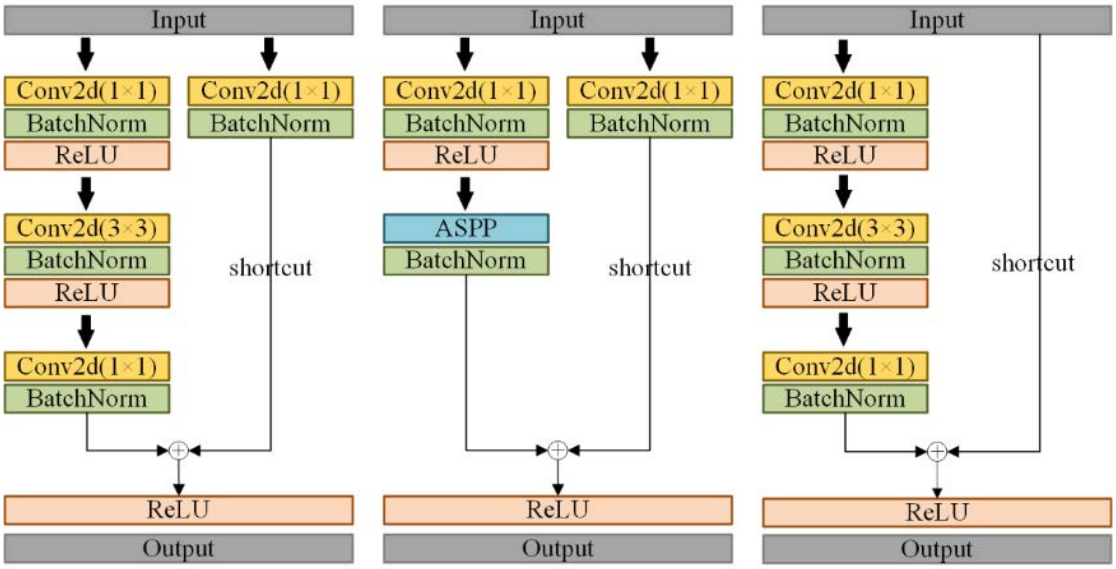


Figure 3. ASPP structure



(a) Original Conv Bloc (b) Conv Block with AS (c) Identity Block

Figure 4. Conv block and identity block

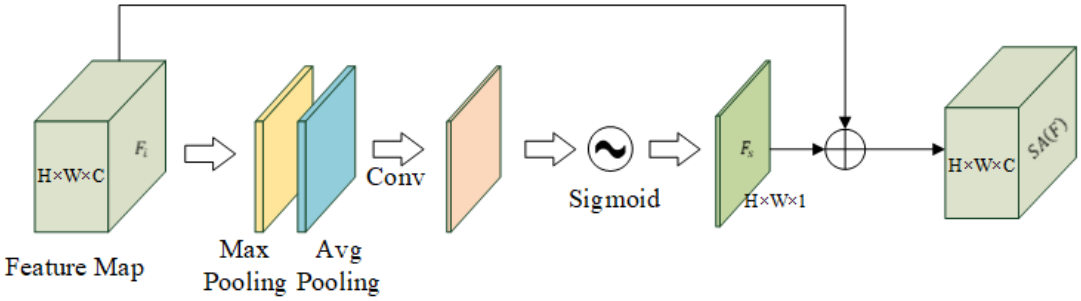


Figure 5. SANet module

Table 1. Architecture of the original and modified ResNet-101

Layer name	Output size	Block name	Origin ResNet-101	Modified ResNet-101
conv1	512×512,64	-	$7 \times 7,64, stride = 2$	$7 \times 7,64, stride = 2$
conv2_x	256×256,256	Conv Block	$\begin{bmatrix} 1 \times 1,64, stride = 2 \\ 3 \times 3,64, stride = 1 \\ 1 \times 1,256, stride = 1 \end{bmatrix} \times 1$	$\begin{bmatrix} 1 \times 1,64, stride = 2 \\ 3 \times 3,64, stride = 1 \\ 1 \times 1,256, stride = 1 \end{bmatrix} \times 1$
		Identity Block	$\begin{bmatrix} 1 \times 1,64, stride = 1 \\ 3 \times 3,64, stride = 1 \\ 1 \times 1,256, stride = 1 \end{bmatrix} \times 2$	$\begin{bmatrix} 1 \times 1,64, stride = 1 \\ 3 \times 3,64, stride = 1 \\ 1 \times 1,256, stride = 1 \end{bmatrix} \times 2$
Conv3_x	128×128,512	Conv Block	$\begin{bmatrix} 1 \times 1,128, stride = 2 \\ 3 \times 3,128, stride = 1 \\ 1 \times 1,512, stride = 1 \end{bmatrix} \times 1$	$\begin{bmatrix} 1 \times 1,128, stride = 2 \\ 3 \times 3,128, stride = 1 \\ 1 \times 1,512, stride = 1 \end{bmatrix} \times 1$
		Identity Block	$\begin{bmatrix} 1 \times 1,128, stride = 1 \\ 3 \times 3,128, stride = 1 \\ 1 \times 1,512, stride = 1 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1,128, stride = 1 \\ 3 \times 3,128, stride = 1 \\ 1 \times 1,512, stride = 1 \end{bmatrix} \times 3$
Conv4_x	64×64,1024	Conv Block	$\begin{bmatrix} 1 \times 1,256, stride = 2 \\ 3 \times 3,256, stride = 1 \\ 1 \times 1,1024, stride = 1 \end{bmatrix} \times 1$	$\begin{bmatrix} 1 \times 1,256, stride = 2 \\ ASPP \end{bmatrix} \times 1$
		Identity Block	$\begin{bmatrix} 1 \times 1,256, stride = 1 \\ 3 \times 3,256, stride = 1 \\ 1 \times 1,1024, stride = 1 \end{bmatrix} \times 22$	$\begin{bmatrix} 1 \times 1,256, stride = 1 \\ 3 \times 3,256, stride = 1 \\ 1 \times 1,1024, stride = 1 \end{bmatrix} \times 22$
Conv5_x	32×32,2048	Conv Block	$\begin{bmatrix} 1 \times 1,512, stride = 2 \\ 3 \times 3,512, stride = 1 \\ 1 \times 1,2048, stride = 1 \end{bmatrix} \times 1$	$\begin{bmatrix} 1 \times 1,512, stride = 2 \\ ASPP \end{bmatrix} \times 1$
		Identity Block	$\begin{bmatrix} 1 \times 1,512, stride = 1 \\ 3 \times 3,512, stride = 1 \\ 1 \times 1,2048, stride = 1 \end{bmatrix} \times 2$	$\begin{bmatrix} 1 \times 1,512, stride = 1 \\ 3 \times 3,512, stride = 1 \\ 1 \times 1,2048, stride = 1 \end{bmatrix} \times 2$

By using global average pooling and maximum pooling and superposition operation, the spatial information of feature map can be fused and mapped into one feature map. The global average pooling operation calculates the average value of all channels at each pixel position, and the global maximum pooling extracts the maximum value of all channels at each pixel position to obtain two H×W×1 feature maps, which are superimposed into an H×W×2 feature map. The formula for calculating channel attention mechanism is as follows:

$$SA(F) = \sigma\left(Conv\left(\left[F_{AvgPool}, F_{MaxPool}\right]\right)\right) \quad (1)$$

Where, $SA(F)$ represents the weight feature map finally obtained, σ represents the Sigmoid activation function, $F_{AvgPool}$ represents the feature map after global average pooling, $F_{MaxPool}$ represents the feature map after global maximum pooling, and $Conv$ represents the convolution operation. The convoluted feature map is obtained through activation function, and the result feature map of attention mechanism is multiplied with the input feature map, and the feature map processed by SANet module is output. The structure of SANet module is shown in Figure 5, where F_i represents the feature map before input module, F_s represents the feature map processed by module, and ⊗ represents the multiplication of corresponding elements. The calculation formula is: $SA(F) = F_i \otimes F_s$.

3.4 Modified FPN Structure

Firstly, after the original image is input to the network, the feature map is extracted through five convolution layers, and the size of the original image is compressed for five times,

each time compressed to 1/2 of the size of the upper feature map, and five feature maps C1, C2, C3, C4 and C5 are obtained. Then, C3, C4 and C5 are up sampled to the dimensions of C2, C3 and C4, respectively fused with C2, C3 and C4 to obtain P2, P3 and P4, C4 is convoluted with 32×32,256 to obtain P5, and P5 is down sampled with MaxPooling convolution with stride size of 2 to obtain P6. Finally, P2, P3, P4, P5 and P6 are input into SANet module to obtain the feature maps Q2, Q3, Q4, Q5 and Q6 processed by attention mechanism. Among them, Q2, Q3, Q4 and Q5 are input into ROI Align layer for regional feature aggregation, and Q3, Q4, Q5 and Q6 are input into RPN to generate anchor boxes. The improved FPN structure is shown in Figure 6.

3.5 Loss Function

The loss function used by the Mask R-CNN model is composed of classification error function, detection error function and segmentation error function:

$$L = L_{class} + L_{box} + L_{mask} \quad (2)$$

The L_{class} and L_{box} loss function is inherited from Faster R-CNN, and the category to which a RoI belongs and its rectangular box coordinate value are predicted by using the fully connected network (FCN). L_{mask} represents the average relative entropy error, and the binary cross entropy loss function is used to calculate the cross entropy of pixels between the mask and the target mask. FCN is used to map each RoI region into a $K_{cls_{num}} \times \mathrm{m} \times \mathrm{m}$ feature layer ($K_{cls_{num}}$ representing the number of categories), and the Sigmoid function is used to calculate the relative entropy of each pixel point, which L_{mask} is obtained. The calculation formula is as follows:

$$L_{mask}\left(K_{cls_{num}}\right) = Sigmoid\left(K_{cls_{num}}\right) \quad (3)$$

4 Experiments

4.1 Experimental Platform

The computer configuration used in this paper is show in Table 2.

Table 2. Configuration of experimental platform

Name	Version
OS	Ubuntu18.04
CPU	2×Intel Xeon(R) Silver 4116
Memory	64G
Graphics card	NVIDIA RTX 2080Ti
TensorFlow	1.13
Keras	2.1.5
CUDA	10.0.130
cuDNN	7.4.2

4.2 Data Enhancement

Because there are few pictures in the training set, in order to avoid over fitting, this paper uses several ways to enhance the data processing: horizontal or vertical flip, random angle rotation, random multiple scaling, adding random salt and pepper noise with different density, so as to enhance the generalization ability of the network. The data enhancement effect is shown in Figure 7.

4.3 Data Set and Training Process

At present, there is no open source infrared image data set of PV panels. We obtained infrared video and images through UAV aerial photography in a domestic PV power station, and obtained about 6000 images after manual processing, including about 700 images with defects. After the data set is enhanced, 3000 images containing different kinds of defects are screened. In order to facilitate training, the images are uniformly cropped to 1024 × 1024. 2200 of them are selected as the training set, and the defect location is marked by *labelme* labeling software. Another 800 images are used as the test set to verify the performance of the model. Set the basic learning rate to 0.001, the number of iterations to 50, and the batchsize to 16. During the training process, the loss function, precision and recall of the training set and the test set change as shown in Figure 8. After 50 epochs of training, the training result model is obtained, and the defect detection and segmentation results of the test set are output, including defect boundary frame, defect mask, defect type and probability of judging as a defect.

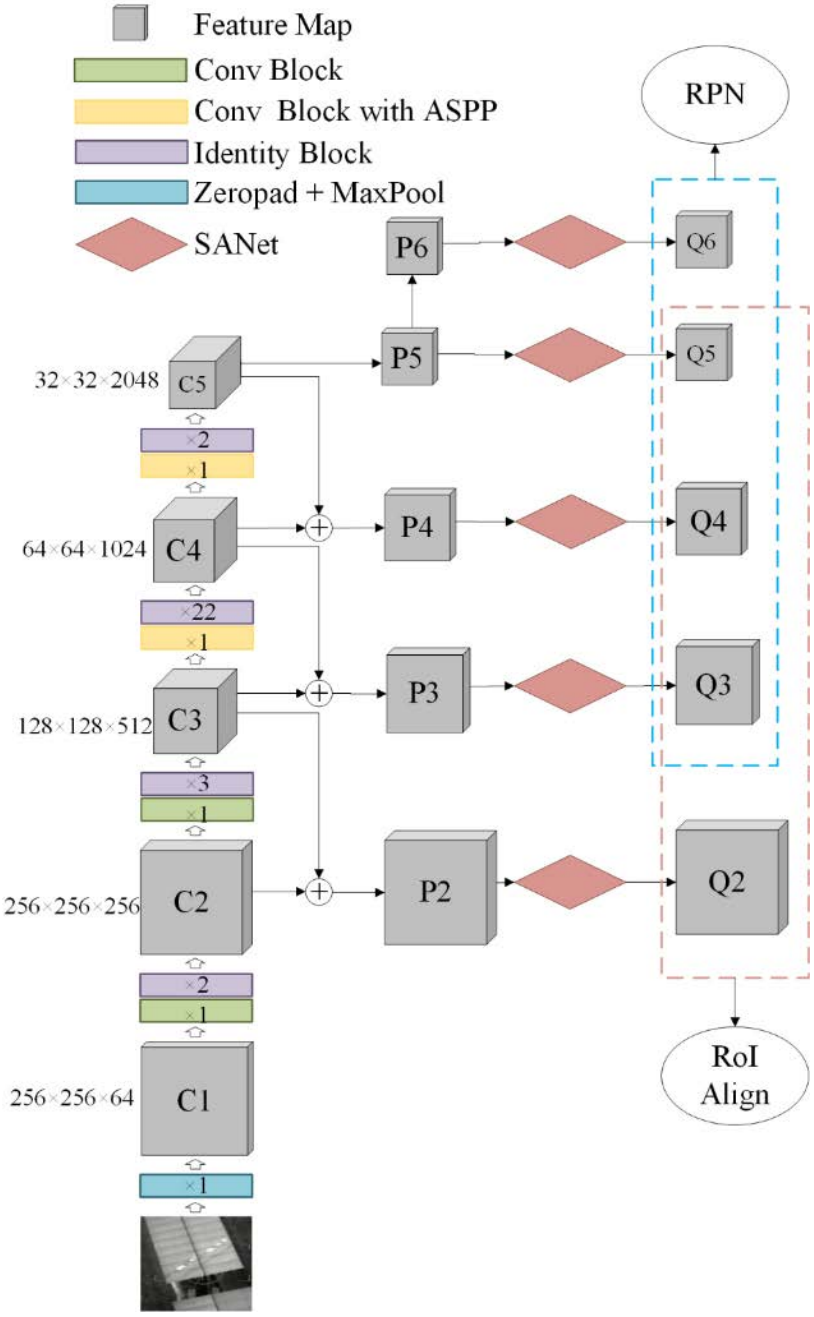
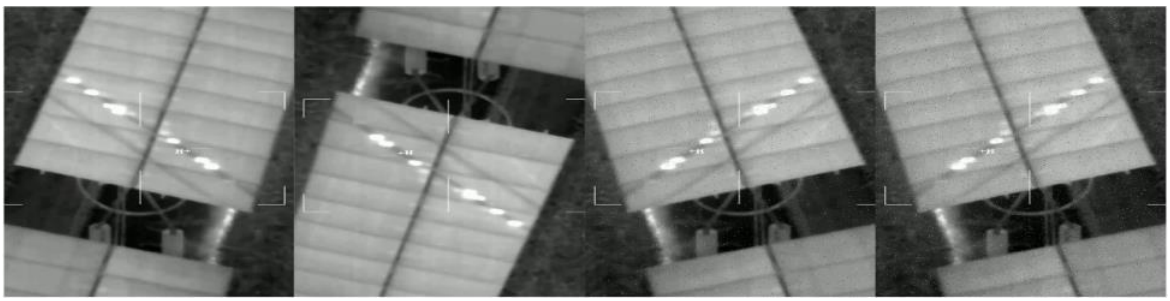
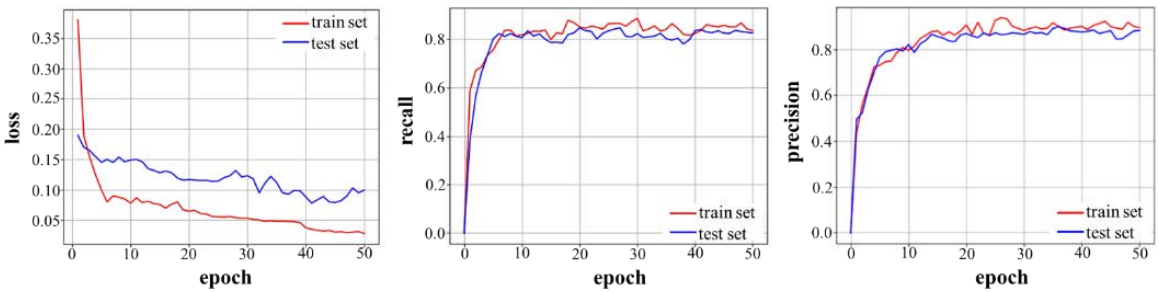


Figure 6. The improved FPN structure



(a) Horizontal flip (b) Vertical flip (c) 0.005 random no (d) 0.01 random noi

Figure 7. Data enhancement effect diagram



(a) epoch-loss curve (b) epoch-recall curve (c) epoch-precision curve

Figure 8. Training and test loss/recall/precision

4.4 The Evaluation Index

Jaccard similarity coefficient is a similarity evaluation index, which can be used to measure the similarity degree of two sets. The cross ratio between the real mask region and the predicted mask can be expressed here:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A|+|B|-|A \cap B|}, \quad 0 \le J(A,B) \le 1 \qquad (4)$$

The pixel coordinates of set A and B are represented as the real labeled mask and the model predicted mask. The average value of Jaccard similarity coefficient of 800 test set images is used as the final evaluation result. Where, $J_r(A,B)$ represents the average Jaccard similarity coefficient, M_i^A and M_i^B represent the real labeling mask and model prediction labeling of the ith test image respectively, and *n* represents the total number of images in the test set.

$$J_r(A,B) = \frac{\sum_{i=1}^{n} J\left(M_i^A, M_i^B\right)}{n} \qquad (5)$$

In order to compare the effects of introducing the ASPP and the spatial attention module in the Mask R-CNN on the segmentation results, the Jaccard similarity coefficient was used as the evaluation index, and the experimental group and the control group were set, where *N* represented the original Mask R-CNN model and *A* represented the model introducing the ASPP structure, *S* represents the model introducing the spatial attention mechanism module, and *AS* represents the model introducing the ASPP structure and the spatial attention mechanism module at the same time. After calculation, the Table 3 is obtained.

By Table 3 can be seen, the introduction of the ASPP structure or attention mechanism module after the segmentation of the original network has brought a little ascension, and at the same time the introduction of the network model segmentation effect is best, ASPP structure and spatial attention mechanism module at the same time improve the feature extraction ability of network and makes the segmentation accuracy of the model for promotion.

Table 3. Jaccard similarity coefficient of different models

Model	$J_r(A,B)$
N	0.469
A	0.522
S	0.597
AS	**0.646**

In order to verify the feasibility of the proposed algorithm, this paper quantifies the accuracy and speed of defect segmentation of the network model through the parameters of precision, recall and FPS, which can accurately reflect the performance of the defect segmentation model. Precision represents the proportion of the number of pixels correctly predicted as defect areas to the number of pixels in all predicted defect areas. The calculation formula [25-26] is as follows:

$$Precision = \frac{TP}{TP+FP} \qquad (6)$$

Recall represents the ratio of pixels in the predicted defect area to the number of pixels in the real defect area. The calculation formula is as follows:

$$Recall = \frac{TP}{TP+FN} \qquad (7)$$

Where, TP is the number of defect pixels correctly predicted, FP is the number of pixels wrongly predicted as defects, and FN is the number of pixels wrongly predicted as background. The average value of al1 image predictions in the test set was used as the final performance evaluation result.

FPS represents the number of image frames that the network model can process per second.

4.5 Results Analysis

This experiment compared the segmentation effects of several common segmentation network models, namely: U-Net [27] model, PSPNet [28] model, Mask R-CNN model, and the improved Mask R-CNN model in this paper. An image with typical defects in the verification set was selected, and the segmentation renderings using Mask R-CNN and the improved Mask R-CNN model in this paper (Figure 9).

Due to the frequent occurrence of hot spot defects and the typical defects, this paper takes an image with many hot spot defects as an example to illustrate the segmentation results. In the segmentation result of Figure 9 (b), there are areas with error detection and inaccurate detection. Use "Error Detection" to mark the areas incorrectly detected as defects, and use "Inaccurate Detection " to mark the areas with inaccurate detection. It can be seen that the areas that were not hot spot defects were detected as hot spot defects in two places, and the segmentation area of hot spot area was too large in another two places (Figure 10).

Segmentation performance evaluation parameters of different network models are shown in the Table 4.

Hot spot defect is a relatively common defect. In Figure 9, "defectA" at the boundary box represents hot spot defect, and number represents the probability of being judged as defect. The original Mask R-CNN model has two error detection areas, and the non-defect area is determined as the defect area. However, the improved algorithm in this paper has no error detection situation.

According to statistics, there are a total of 1240 defect positions in the 800 images of the test set. Different network models are used to detect the correct detection number (P_r), false detection number (P_e) and missed detection number (P_o) of defects respectively. The relative accuracy rate (P_{acc}) is used to measure the detection effect. The calculation formula is as follows: $P_{acc} = \frac{P_r - P_e}{P_a}$. Where P_a represents the total number of defects. After calculation, the Table 5 is obtained.

The correct detection number (P_r) of the improved algorithm in this paper is significantly higher than the other three models (Table 5), and the false detection number (P_e) is far lower than the other three models, and the relative accuracy rate (P_{acc}) is better than the other three models. Although the detection speed of the algorithm in this paper is lower than that of the original network, its accuracy is improved. For the defects in infrared images of panels, the importance of detection accuracy is far greater than detection speed. Therefore, the method of exchanging accuracy for time in this algorithm is of practical application value.

(a) Original image (b) Mask R-CNN (c) Algorithm in this paper

Figure 9. Segmentation effect diagram of different network models

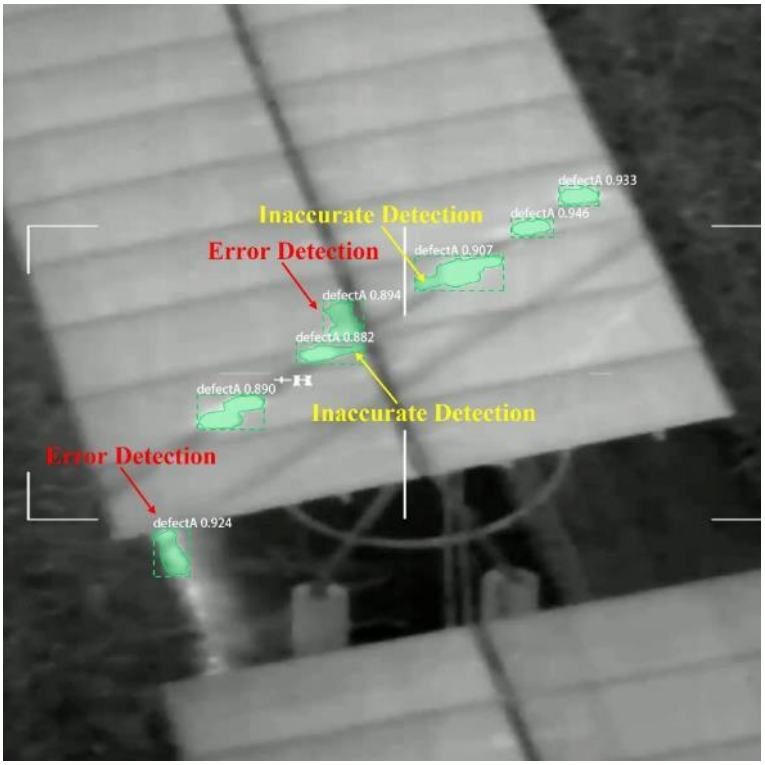


Figure 10. Error detection area and inaccurate detection area

Table 4. Network model evaluation

Model	Precision	Recall	FPS
U-Net	0.773	0.794	12.5
PSPNet	0.822	0.790	9.6
Mask R-CNN	0.840	0.816	8.6
Algorithm in this paper	**0.898**	**0.844**	**8.2**

Table 5. Relative accuracy of defect detection

Model	P_r	P_e	P_o	P_{acc}
U-Net	983	104	257	0.793
PSPNet	1032	79	208	0.769
Mask R-CNN	1132	56	108	0.868
Algorithm in this paper	**1197**	**41**	**43**	**0.932**

5 Conclusion

In this paper, a defect detection method based on the infrared image of PV panels and the Mask R-CNN instance segmentation model is proposed. Based on the original algorithm, a new defect detection and segmentation model is constructed by adding the ASPP and spatial attention mechanism. The UAVs is equipped with an infrared camera to take infrared video of the panel of the PV power station, make the data set and test the model performance, which provides a new feasible scheme for unmanned patrol inspection of the PV power station. The experimental results show that although the detection speed of the improved algorithm is slower than other algorithms, the accuracy is improved by 6.4%, 16.3% and 13.9% respectively compared with Mask R-CNN, U-Net, PSPNet and other algorithms.

6 Limitation and Future Research

In this work, because the detection speed is slow and the algorithm model is relatively large, it is necessary to shoot the video and then use the high configuration computer for detection. The follow-up research focus is to simplify the network model and improve the detection speed without reducing the detection accuracy, and deploy the program on the UAVs to achieve the purpose of real-time detection.

References

[1] I. Zyout, A. Oatawneh, Detection of PV solar panel surface defects using transfer learning of the deep convolutional neural networks, *2020 Advances in Science and Engineering Technology International Conferences (ASET)*, Dubai, United Arab Emirates, 2020, pp. 1-4.

[2] A. Greco, C. Pironti, A. Saggese, M. Vento, V. Vigilante, A deep learning based approach for detecting panels in photovoltaic plants, *Proceedings of the 3rd International Conference on Applications of Intelligent Systems*, Las Palmas de Gran Canaria, Spain, 2020, pp. 1-7.

[3] M. W. Akram, G. Li, Y. Jin, X. Chen, C. Zhu, A. Ahmad, Automatic detection of photovoltaic module defects in infrared images with isolated and develop-model transfer deep learning, *Solar Energy*, Vol. 198, pp. 175-186, March, 2020.

[4] J. J. Vega Díaz, M. Vlaminck, D. Lefkaditis, S. A. Orjuela Vargas, H. Luong, Solar panel detection within complex backgrounds using thermal images acquired by UAVs, *Sensors*, Vol. 20, No. 21, Article No. 6219, November, 2020.

[5] S. Bharati, P. Podder, 1 Performance of CNN for predicting cancerous lung nodules using LightGBM, *Artificial Intelligence for Data-Driven Medical Diagnosis*, De Gruyter, 2021.

[6] S. Bharati, P. Podder, M. R. H. Mondal, V. B. S. Prasath, Medical imaging with deep learning for COVID-19 diagnosis: a comprehensive review, *International Journal of Computer Information Systems and Industrial Management Applications*, Vol. 13, pp. 91-112, July, 2021.

[7] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems 25*, Lake Tahoe, Nevada, USA, 2012, pp. 1097-1105.

[8] S. Bharati, P. Podder, M. Rubaiyat, H. Mondal, Artificial neural network based breast cancer screening: a comprehensive review, *International Journal of Computer Information Systems and Industrial Management Applications*, Vol. 12, pp. 125-137, May, 2020.

[9] S. Bharati, P. Podder, M. R. H. Mondal, N. Gandhi, Optimized NASNet for diagnosis of COVID-19 from lung CT images, *International Conference on Intelligent Systems Design and Applications*, online, 2020, pp. 647-656.

[10] C. Mantel, F. Villebro, G. A. dos Reis Benatto, H. R. Parikh, S. Wendlandt, K. Hossain, P. Poulsen, S. Spataru, D. Sera, S. Forchhammer, Machine learning prediction of defect types for electroluminescence images of photovoltaic panels, *Applications of Machine Learning*, *International Society for Optics and Photonics*, Vol. 11139, Article No. 1113904, September, 2019.

[11] J. Uma, C. Muniraj, N. Sathya, Diagnosis of photovoltaic (PV) panel defects based on testing and evaluation of thermal image, *Journal of Testing and Evaluation*, Vol. 47, No. 6, pp. 4249-4262, November, 2018.

[12] P. Addabbo, A. Angrisano, M. L. Bernardi, G. Gagliarde, A. Mennella, M. Nisi, S. Ullo, A UAV infrared

measurement approach for defect detection in photovoltaic plants, *2017 IEEE International Workshop on Metrology for AeroSpace (MetroAeroSpace)*, Padua, Italy, 2017, pp. 345-350.
[13] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: unified, real-time object detection, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 779-788.
[14] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask r-cnn, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 42, No. 2, pp. 386-397, February, 2020.
[15] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: towards real-time object detection with region proposal networks, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 39, No. 6, pp. 1137-1149, June, 2017.
[16] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770-778.
[17] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection, *30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 936-944.
[18] L. Liu, C. Shen, A. van den Hengel, The treasure beneath convolutional layers: cross-convolutional-layer pooling for image classification, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 2015, pp. 4749-4757.
[19] W. S. Liew, T. B. Tang, C. H. Lin, C. K. Lu, Automatic colonic polyp detection using integration of modified deep residual convolutional neural network and ensemble learning approaches, *Computer Methods and Programs in Biomedicine*, Vol. 206, Article No. 106114, July, 2021.
[20] S. Bharati, P. Podder, M. Mondal, V. B. Prasath, CO-ResNet: optimized ResNet model for COVID-19 diagnosis from X-ray images, *International Journal of Hybrid Intelligent Systems*, Vol. 17, No. 1-2, pp. 71-85, July, 2021.
[21] A. Khamparia, S. Bharati, P. Podder, D. Gupta, A. Khanna, T. K. Phung, D. N. Thanh, Diagnosis of breast cancer based on modern mammography using hybrid transfer learning, *Multidimensional systems and signal processing*, Vol. 32, No. 2, pp. 747-765, April, 2021.
[22] M. R. H. Mondal, S. Bharati, P. Podder, Diagnosis of COVID-19 using machine learning and deep learning: a review, *Current Medical Imaging*, Vol. 17, No. 12, pp. 1403-1418, 2021.
[23] S. Woo, J. Park, J.-Y. Lee, I. S. Kweon, CBAM: convolutional block attention module, *15th European Conference on Computer Vision (ECCV)*, Munich, Germany, 2018, pp. 3-19.
[24] X. Zhu, D. Cheng, Z. Zhang, S. Lin, J. Dai, An empirical study of spatial attention mechanisms in deep networks, *IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, South Korea, 2019, pp. 6687-6696.
[25] S. Bharati, P. Podder, M. R. H. Mondal, Hybrid deep learning for detecting lung diseases from X-ray images, *Informatics in medicine unlocked*, Vol. 20, Article No. 100391, 2020.
[26] M. R. H. Mondal, S. Bharati, P. Podder, CO-IRv2: Optimized inceptionResNetV2 for COVID-19 detection from chest CT images, *PloS one*, Vol. 16, No. 10, Article No. e0259179, October, 2021.
[27] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, *International Conference on Medical image computing and computer-assisted intervention*, Munich, Germany, 2015, pp. 234-241.
[28] J. Zhou, M. Hao, D. Zhang, P. Zou, W. Zhang, Fusion PSPnet image segmentation based method for multi-focus image fusion, *IEEE Photonics Journal*, Vol. 11, No. 6, Article No. 6501412, December, 2019.

Biographies

Shuqiang Guo rceived his M.S. degree in Computer science and technology from School of Information Engineering of Northeast Electric Power University in 2006. He received the D.S degree in information science from Yamagata University in 2010. His main research interests include image processing, pattern recognition, and computer vision.

Zhiheng Wang is currently the Master degree in computer science and technology engineering with Northeast Electric Power University. His main research direction is the application of computer vision in industrial defect detection.

Yue Lou is currently the Master degree in computer technology engineering with Northeast Electric Power University. Her research interests are the areas of image processing and computer technology.

Xianjin Li is currently the Master degree in computer science and technology engineering with Northeast Electric Power University. His research interests are deep learning and digit image processing.

Huanqiang Lin is currently the Master degree in communication engineering with Northeast Electric Power University. His research interests are image processing and computer technology.