Research on Surface Defect Detection of Solar Pv Panels Based on Pre-Training Network and Feature Fusion

Huohua Li¹, *, Xiaogang Fu² and Tairu Huang¹

¹School of Electrical Engineering, Shanghai DianJi University, Shanghai, China
²Industrial Technology Center, Shanghai DianJi University, Shanghai, China

*Corresponding author: huohuali@sdju.edu.cn

Abstract. Aiming at the problems of the lack of sample size and the complexity of defect images in defect detection task, based on the idea of transfer learning and hierarchical feature fusion, this paper proposes a deep classification network model of improved vgg19 pre-training network by analyzing the basic principle of feature extraction of convolutional neural network and getting inspiration from feature pyramid network. Then, the model is trained by the small-scale defect images of solar pv panel. Finally, the solar pv panel data set containing four kinds of defects, including cracks, debris, broken gates and black areas, is selected to comprehensively verify the effectiveness of the improved network in the defect detection task. The experimental results show that the proposed method is superior to the classical vgg19 network model in four evaluation indexes, such as accuracy, precision, recall rate and F1 score.

Keywords: Defect detection, transfer learning, feature fusion, vgg19, solar pv panel.

1. Introduction
As a big manufacturing country, China has to produce a large number of industrial products every day. With the improvement of living standards, consumers have higher and higher requirements for product quality. In addition to the performance of products, they also need to ensure the perfect appearance. However, in the process of manufacturing products, surface defects are often inevitable. Surface defects not only affect the beauty and comfort of products, but also affect their performance in different degrees. Therefore, manufacturers attach great importance to the surface defect detection, so as to effectively control the product quality, and analyze some problems in the production process according to the test results, so as to effectively eliminate or reduce the production of defective products and prevent defects. After the products flow into the market, it will cause trade disputes, affect the honor of enterprises, and even endanger personal safety.

For example, in the production process of solar pv(photovoltaic) panels, there will be some defects that are not easy to be detected by naked eyes, such as cracks, debris, broken gates and black areas, which seriously affect the power generation efficiency. Therefore, the defect detection technology of solar pv panel is an important link to ensure the efficiency of pv power generation [1].

At present, manual quality inspection is still the most traditional method of surface defect detection. This method has low sampling rate, low accuracy, poor real-time performance, low efficiency, high
labor intensity, and is greatly affected by the experience of quality inspection workers and subjective factors. Machine vision as a non-contact, non-destructive automatic detection technology, has the outstanding advantages of safety, stability, high efficiency. With the rapid growth of deep learning model represented by convolutional neural network (CNN), it has been successfully applied in many fields of machine vision [2]. Many defect detection methods based on deep learning are also widely used in various industrial scenarios.

However, different from theoretical research, surface defect inspection still faces many problems, such as: 1) we often cannot find large-scale product defect images to complete the training of deep neural network model in engineering design practice, which leads to the phenomenon of model over fitting and falling into local optimal solution. 2) in the real and complex industrial environment, we cannot find large-scale product defect images to complete the deep neural network model training. There may be small difference between defect imaging and background, low contrast, large change of defect scale, variety of types, large amount of noise in defect image, and even a large amount of interference in the imaging of defects under the environment of the factory, as shown in Figure 1.

![Some common surface defect images.](image)

2. Fine tuning strategy of pre-training network based on transfer learning

In the practice of deep learning method, we found that after the effective training of ImageNet data set, the parameters of the deep network models (such as vgg16, vgg19, inception-v3, resnet50 and other classic deep network models) are constantly adjusted, and they often perform well in the test set [2], that is, the model has good generalization ability and robustness. However, if only a small number of training samples are used to train the deep network model, the performance of the model in the test set is poor.

2.1. Transfor learning

Transfer learning [3] refers to the transfer of knowledge from the source domain to the target domain, that is, transferring the previous knowledge to another new domain. In order to transfer the recognition ability of the pre-training network model to the user-defined data set. This paper introduces the idea of transfer learning and lists three fine tuning strategies for classical pre training network, as shown in Figure 2.
2.1.1. Selection rules of fine-tuning strategy. According to the size of the user-defined data set and the similarity with the pre-training data set, the fine-tuning rules are summarized as follows:

1) When the similarity between the user-defined data set and pre-training data set is small, and the scale of the user-defined data set is relatively large, the whole model is often trained on the basis of the original parameters of the pre-training network, that is, Option 1.

2) When the user-defined data set is similar to the pre-training data set to a certain extent, and the scale of the user-defined data set itself is not large, the partial convolution layer of the pre-training network is frozen, and only a small amount of user-defined data is used to train the high-level convolution group, that is, Option 2.

3) When the user-defined data set and the pre-training data set have large similarity and small scale, freeze all convolution layers and train only the full connection layer of the pre-training network, that is, Option 3.

2.2. Classic VGG-19 pre-training network

As shown in Figure 3, the VGG-19 network consists of five convolution groups and three full connection layers. These five convolution groups contain 2, 2, 4, 4, and 4 convolution layers respectively, so there are 19 layers in total.
As pointed out in [4], convolution layers with different depths have different receptive fields, and can extract features of different levels in the image. Shallow convolution kernel can extract sharp and detailed boundary, texture and gray features, which are universal for different images. Therefore, from the theoretical analysis, we can freeze the low-level convolution group, that is, to preserve the shallow weight parameters after updating and adjusting the pre-training data set, which can be directly used to extract the shallow features of user-defined data sets. In addition, the high-level convolution group has larger receptive field, which can extract more abstract semantic information. Therefore, high-level features can be directly sent into the classifier for classification. In order to ensure that the target features are more discriminative, it is necessary to use the user-defined data set to train and adjust the semantic parameters of the model. Finally, the full connection layer is actually a classifier. According to the specific classification task, it needs to adjust adaptively, such as setting the reasonable number of neurons and the number of classification label, and selecting the appropriate loss function and optimizer. In the process of transfer learning, this part can be replaced by the user-defined full connection layer.

3. Hierarchical feature fusion mechanism
As shown in Figure 4, the classic Feature Pyramid Network (FPN) [5] integrates features of different depths organically, strengthens the information flow between layers, and provides more detailed information for visual tasks. By fusing feature maps of different depths, the high-level features of low-resolution and high-level semantic information and the low-level features of high-resolution and low-level semantic information are connected from top to bottom, so that the features at all scales have rich semantic information, so it has been widely used in the fields of target detection and image segmentation.

![Feature Pyramid Network](image)

**Figure 4.** Feature Pyramid Network.

4. Attention mechanism
Attention mechanism has been proposed for the first time in the convolutional neural network Senet (squeeze-and-excitation network) [6]. After adding attention mechanism, the network model can autonomously learn and record the importance of each feature channel. The importance degree is added to feature channels to enhance the importance of useful channels and suppress the importance of useless channels. In fact, this is a weighted fusion process, which allocates appropriate weight coefficients according to the importance of the channel. Senet network greatly improves the efficiency of feature extraction through attention mechanism. Inspired by SENet network, Zhang et al. [7] designed a channel attention network, and its schematic diagram is shown in Figure 5.
Suppose that the size of the input feature map is $H \times W \times C$, where $h$, $W$ and $C$ represent the height, width and channel number of the feature map respectively. Firstly, the HGP of each channel's feature map is pooled globally, and the pooled value is used as the feature descriptor of each channel, thus a $C$-dimensional feature vector can be obtained. Then, a two-layer perceptron network is used to fuse the information of different channels. Where $WD$ is the reduction of the number of channels, $WU$ is the expansion of the number of channels, and the scaling factor of reduction and expansion is $r$. A new feature vector can be obtained, and then the activation function $f$ is used to activate the vector. Finally, the original feature maps are weighted and fused.

5. Improved VGG-19 pre-training network

By introducing the idea of transfer learning and attention mechanism, and getting inspiration from the feature pyramid network, we propose an improved vgg19 deep network model, as shown in Figure 6.

Figure 5. Channel Attention Network.

5.1. Fine tuning strategy of backbone network based on VGG-19

In this study, VGG-19 backbone network is selected. After full training of ImageNet data set, the network has shown good generalization ability in the test set. However, the solar pv panel data set used in this study is small in scale and does not have a large degree of similarity with ImageNet data set. Therefore, according to the fine-tuning rules shown in Figure 1, Option 2 is selected as the fine-tuning strategy for this study.

The specific adjustment methods are as follows: 1) The basic texture parameters of vgg19 and vgg19 are used to extract the original texture features. 2) Training high-level convolution group: using the collected small-scale data set to fully train and adjust the semantic level parameters of the pre training, so that the extracted target features are more discriminative. 3) A user-defined full connection layer is designed to replace the full connection layer of the pre training network. (the parameters are set as follows: full connection layer 1, size = 4096; full connection layer 2, size = 4096; softmax.size = 4, loss = "sparse_categorical_crossentropy"; optimizer="stochastic gradient descent")
5.2. Hierarchical feature extraction module

Inspired by the pyramidal feature fusion module, VGG-19 backbone network has five convolution groups. We can extract a feature map $f_i$ from the last convolution layer of each convolution group (i.e., conv1_2, conv2_2, conv3_4, conv4_4, conv5_4), so a set of feature maps $F = (f_1, f_2, f_3, f_4, f_5)$ can be obtained, as shown in Figure 7.

![Figure 7. Schematic diagram of hierarchical feature extraction.](image)

5.3. Hierarchical feature fusion attention module

Since convolution is a down sampling process, it means that features from deeper layers have lower resolution. If the features extracted from the first convolution group are taken as the standard, the features from the remaining four convolution groups need to be upsampled (deconv) to make the resolution the same. Then, the channel attention mechanism shown in Fig. 5 is introduced into the fusion scheme, five features of different depths are weighted and fused. Finally, the fused feature map is sent to the classifier for classification.

6. Experimental results and analysis

In this experiment, 800 pictures of solar pv panels were collected, which contained four kinds of defects, namely cracks, debris, broken gate and black area. As shown in Figure 8.

![Figure 8. Defects images of solar pv panel.](image)

Using tensorflow2.0 deep learning framework, 680 solar pv panel images are selected as training set. After a series of preprocessing processes such as cutting, tagging and standardization, they are sent to the improved vgg19 deep classification network for training. After the normal convergence of the model, 30 pieces of defect images of four types of solar pv panel were selected as the test set to verify the detection effect of the improved model for the four kinds of defect images. Finally, the classic vgg19 network is used to do the same operation, and the experimental results are compared to verify the advantages of the improved model compared with the traditional model.
The evaluation index was mainly composed of accuracy (AC), recall (R), precision (P), and F1 score.

The accuracy rate is the percentage of correct prediction results in the total sample, and its calculation formula is as follows:

\[
AC = \frac{TP + TN}{TP + TN + FP + FN}
\] (1)

TP: the number of positive classes predicted as positive classes;
FN: the number of positive classes predicted as negative classes;
FP: the number of negative classes predicted as positive classes;
TN: the number of negative classes predicted as negative classes;

The precision rate is the probability that all predicted positive samples are actually positive samples, and its calculation formula is as follows:

\[
P = \frac{TP}{TP + FP}
\] (2)

Recall rate refers to the probability that the actual positive samples are predicted to be positive samples, and the formula is as follows:

\[
R = \frac{TP}{TP + FN}
\] (3)

The F1 score is used to comprehensively evaluate the performance of the model, and the formula is as follows:

\[
F1 = \frac{2 \times P \times R}{P + R}
\] (4)

Under the condition that the training data and test data are the same, the experimental results of VGG-19 model and improved VGG-19 model on four evaluation indexes are shown in Table 1 and table 2.

**Table 1.** Classic VGG-19 model.

| Defect type  | AC (%) | P (%) | R (%) | F1  |
|--------------|--------|-------|-------|-----|
| Cracks       | 90.83  | 85.18 | 76.67 | 80.70 |
| Debris       | 91.67  | 88.46 | 76.67 | 82.14 |
| Broken Gate  | 95.00  | 92.86 | 86.67 | 89.66 |
| Black Area   | 93.33  | 85.71 | 80.00 | 82.76 |

**Table 2.** Improved VGG-19 model.

| Defect type  | AC (%) | P (%) | R (%) | F1  |
|--------------|--------|-------|-------|-----|
| Cracks       | 93.33  | 89.29 | 83.33 | 86.21 |
| Debris       | 94.17  | 92.59 | 83.33 | 87.72 |
| Broken Gate  | 96.67  | 96.55 | 93.33 | 94.91 |
| Black Area   | 95.00  | 92.86 | 86.66 | 89.65 |
By comparing the experimental data in Table 1 and Table 2, it can be found that although the accuracy rate of the classical vgg19 network for the detection of four common defects of the solar pv panel image reaches more than 90%, the accuracy and recall rate are low, especially the crack and fragment type defects, which have low contrast with the normal images, are relatively difficult to accurately distinguish, and the two types of defects are the same. The improved vgg19 network model with feature fusion attention module can theoretically extract more detailed information. Compared with the experimental data, it can be seen that the detection accuracy of crack and fragment is significantly improved, and the recall rate is increased from 76.67% to 83.33%. From the overall analysis, the improved vgg19 network in the accuracy, accuracy, recall rate and F1 score four indicators are effectively improved than the classic vgg19 model, which verifies the effectiveness of the improved model in the defect detection task.

7. Conclusions
In this paper, based on some problems of automatic defect detection technology, such as insufficient samples, low contrast, intra-class difference and inter-class similarity, an improved defect detection model is proposed based on the pre-training backbone network, feature fusion mechanism and attention mechanism. The experimental results show that the deep network model training can be completed only by using small-scale pv panel defect data set, and the evaluation index is better than the classic pre-training network, and the F1 score is more than 90%. However, in order to put the model into practical industrial application, the detection accuracy needs to be further improved. In addition, this study only uses one data set to verify the validity of the model, and only four types of defects are listed. It is far from enough to comprehensively evaluate the effectiveness of the improved model on more defect data sets. However, this paper has a certain reference value for solving the defect detection technology problems faced by similar products.

Acknowledgments
Finally, I would like to thank my instructor for guiding me in my most confused time. I would also like to thank my colleagues in the laboratory for their help. I hope you will be with me in the future and spend a better life in the graduate stage.

References
[1] Yan Yunfei, Zhang Zhen, Zhang Zhien, et al. Solar energy utilization technology and its application [J]. Acta solar Sinica, 2012, 33 (S1): 47-56.
[2] Lecun Y, Bengio Y, Hinton G. Deep learning [J]. Nature, 2015, 521( 7553): 436 - 444 .
[3] Han D, Liu Q, Fan W. A new image classification method using CNN transfer learning and webdata augmentation [J]. Expert Systems with Applications, 2018, 95( 2): 43 - 56.
[4] Zaman Q U, Salyani M. Effects of foliage density and ground speed on ultrasonic measurement of citrus tree volume [J]. Transactions of the ASAE, 2004, 20( 2): 173 - 178 .
[5] A. Mahendran and A. Vedaldi, “Understanding deep image representations by inverting them,” 2015 IEEE Conf. on Comput. Vis. and Pattern Recognit. (CVPR), Boston, MA, 2015, pp. 5188-5196.
[6] Hu J, Shen L, Albanie S, Sun G and Wu E H. 2019. Squeeze-and-excitation networks. IEEE Transactions on Pattern Analysis and Machine Intelligence: 2011-2023 [DOI: 10. 1109 /TPAMI. 2019 . 2913372]
[7] Zhang Y L, Li K P, Li K, Wang L C, Zhong B E and Fu Y. 2018a. Image super-resolution using very deep residual channel attention networks/ /Proceedings of the 15th European Conference on Computer Vision. Munich: Springer: 2472-2481