2022 6th International Conference on Universal Village (UV) | 978-1-6654-7477-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/UV56588.2022.10185478

Carpet Defect Detection by Transfer Learning Combing Classification and Semantic Segmentation

Tianqing Ren
Jiangnan University
Wuxi, China
rtq1098180986@163.com

Longfei Zhou*
School of Medicine
Duke University
Durham, USA
longfei.zhou@duke.edu

Ke Xu
Hangzhou Dianzi University
Hangzhou, China
xukehdu@outlook.com

Yifan Wang
South China University of Technology
Guangzhou, China
DM872674963@163.com

Siyu Wu
Nanjing University of Information Science and Technology
Nanjing, China
yuiwsy@gmail.com

Yuliang Gai
Jilin University of Finance and Economics
Changchun, China
13180719799@163.com

Jiazheng Chen
Nanjing University of Information Science and Technology
Nanjing, China
202083330028@nuist.edu.cn

Zhichao Gou
Jiangsu University
Zhenjiang, China
Gouzhichao01414@163.com

***Abstract*—Nowadays, with the development of industrial production technology, defect detection has become an indispensable part of industrial production. However, due to various types of products and defects, it can be extremely difficult to identify and locate those defects precisely and accurately. The current major trend in defect detection is using convolutional neural networks and semantic segmentation techniques to better minimize the error rate of human eye recognition and highly improve efficiency. Our work is based on semantic segmentation method and combines it with transfer learning technique enabling our model to train on a relatively small dataset without compromising the performance, and use CNN to firstly classify input images in order to further reduce the number of images to improve computational efficiency and accuracy. Then through incorporating state-of-the-art semantic segmentation model U-Net++, our model achieves the best performance compared to U-Net under transfer learning scenario. We compare our model with the state-of-the-art U-Net. Then we use mIOU and pixel accuracy to measure the models' performance under two scenarios. Results illustrated that through transfer learning scenario, our model achieves the highest scores over other methods.**

Keywords—defects detection, transfer learning, classification, semantic segmentation.

I. Introduction

In the process of industrial production, due to the influence of equipment, process, environment and other factors, defects varying differently often appear on products, such as carpet scratches, leather unevenness, tile shattering and wood cracking. Such defects are usually areas of uneven physical or chemical properties on the surface of the product, which abate products' appearance and usefulness, thus affecting the efficiency of industrial production [1]. Therefore, searching for suitable defect detection methods is of great significance to the development of industry and the whole national economy.

Defect detection methods are mainly divided into two domains: manual inspection and machine vision, of which manual inspection has been gradually superseded by present machine vision techniques due to large labor expense, low detection accuracy, and low efficiency [2]. Machine vision-based detection methods are divided into traditional methods and deep learning methods, in which traditional methods use industrial cameras combined with image processing technology to identify product defects, such as statistical method, filter method and model method [3]. Traditional machine vision methods rely too much on manual feature labeling and specific detection environments, and cannot be applied to large-scale defect detection task for multiple types of industrial products.

In 2012, the proposal of AlexNet [4] promoted the booming development of deep learning, which laid a solid foundation for present deep learning. In the meantime, it was gradually applied in various fields with certain success, such as autonomous driving [5], medical diagnosis [6], and speech recognition [7]. Many scholars began to explore the combination of deep learning and defect detection and carried out a series of research works. Zheng [8] proposed a residual U-structure-embedded encode-decode block of U-Net with a hybrid loss function and a coordinate attention module applying to tire defect detection. Ling [9] present an efficient multihead self-attention method, which can automatically locate single or multiple defect areas of magnetic tile and extract features of the magnetic tile defects. Huang [10] introduced an efficient convolutional neural network which is divided into two parts: segmentation and decision for fabric defect detection.

Most of the existing defect detection methods focus on defects of single product type and require a large number of

*Corresponding author: Longfei Zhou (longfei.zhou@duke.edu)

Authorized licensed use limited to: University of Maryland College Park. Downloaded on August 27,2023 at 16:41:20 UTC from IEEE Xplore. Restrictions apply.

defect samples to train the model. However, in real industrial production scenario, it is impossible to obtain such amount of defect samples to support model training [11]. Since some products have similar types of defects, the model training can be conducted by transfer learning through similar products to reduce the dependence on big data [12].

In this paper, we propose a defect detection model based on transfer learning for multi-variety industrial products, which consists of two parts: classification and segmentation. The proposed model is first trained on the leather dataset and then migrated to the carpet dataset for defect detection. The paper is structured as follows. The second part provides the related knowledge of image classification, semantic segmentation, and transfer learning applied to this paper. The third part details the proposed classification and segmentation model. The fourth part describes the conducted experiments and the analysis of the results. The last part is a summary of this article and possible future work.

II. Related Knowledge

A. Classification

The classification task is used to obtain the category label of an image and determine whether the image has a defect or not and further identify the type of the defect. Most classification tasks in deep learning are implemented by convolutional neural networks. CNNs extract abstract features of images through a series of convolution and pooling layers, and then combine them with fully connected layers to achieve image classification. The common classification networks in defect detection are VGG, GoogLeNet, ResNet, DenseNet, etc. Akram [13] designed VGG-8, VGG-7 and VGG-6 based on VGG-11 network structure to detect surface defects in solar cells. Lu [14] added an attention module to ResNet when performing the task of strip surface defect classification, generating new network structures A-ResNet50 and A-ResNet101.

B. Semantic Segmentation

Semantic segmentation uses pixel-to-pixel analysis to determine whether they are defect targets, and finally the precise shape and location information of the defect can be obtained. Commonly used semantic segmentation models in defect detection include FCN, U-Net series, SegNet, DeepLab series, etc. U-Net is used in the segmentation network of the defect detection model proposed in this paper.

The U-Net network consists of two parts: encoder and decoder, where the encoder extracts image features through continuous convolution and pooling layers, and decoder performs successive upsampling and convolution operations, as well as using jump connections to fuse with features extracted by the encoder to achieve image segmentation. Liu [15] proposed a magnetic tile defect detection and recognition algorithm by combining an improved U-Net model and a classification neural network. The improved U-net model uses the dilated convolution to replace some convolutional layers and pooling layers, and adds more jump connections, which can achieve a prediction accuracy of 93%. Han [16], in their study of poly-silicon wafer defect segmentation, first applied Region Proposal Network to generate underlying defect region, and then segmented these image blocks with the modified U-Net.

C. Transfer Learning

Transfer learning is a machine learning method that transfers the knowledge learned from source data in one domain to another specific target domain. This method allows the model to reduce the dependence on data from the target domain and enables it to achieve better performance even without large amount of training samples. Since sufficient defect images are often not accessible in industrial product defect detection scenario, different products can be considered as different domains and transfer learning can be adopted to reduce the dependence on task-specific data. Liu [17] proposed a defect detection algorithm based on multi-source domain deep transfer learning with Mask R-CNN as the basic framework. Compared with the non-migration approach, the MAP increased by 11.5% in the migration approach. Damacharla [18] built a model based on the U-net framework for steel defect detection, and compared the effect of training with and without pre-training on ImageNet. The result was that the accuracy of detection algorithm based on transfer learning outperformed the traditional detection algorithm by 26%.

III. Method

Accurate defect detection can help improve the efficiency and stability of industrial production. To solve the problem of pixel-level segmentation and a small number of training samples, we divide the problem into three parts. The first part is data expansion. The second part is built around a CNN model. For an image as the input data, we expect the model to be able to judge whether the component contains any types of defects. The third part is built around a U-Net++ model. For the image filtered from the first part that may contain some defects, we expect the output image to be a pixel-level binary image of some defects. The algorithmic framework is shown in Fig. 1.

A. Dataset Expansion

Due to the small size of samples in the original dataset, we use some image enhancement methods, such as flipping, rotation, clipping, and color-shading adjustment. Using the above methods, we expand the size of the dataset to 10 times the original. Finally, we divide the dataset into a training set, a verification set, and a test set with a ratio of 7:2:1. Table I shows the number of examples in each part.

B. Overall Classification

Considering the CNN model, Zhou & Zhang [19] have proposed a model named ECON. We make improvements to this model to fit our dataset (structure of the classification network shows as Fig. 2). The input image of the model is an array of (256, 256, 3). The number of kernels in the Conv2d

TABLE I. The Number of Images

	Train	Validation	Test
Leather	1935	277	553
Carpet	2012	288	575

Authorized licensed use limited to: University of Maryland College Park. Downloaded on August 27,2023 at 16:41:20 UTC from IEEE Xplore. Restrictions apply.

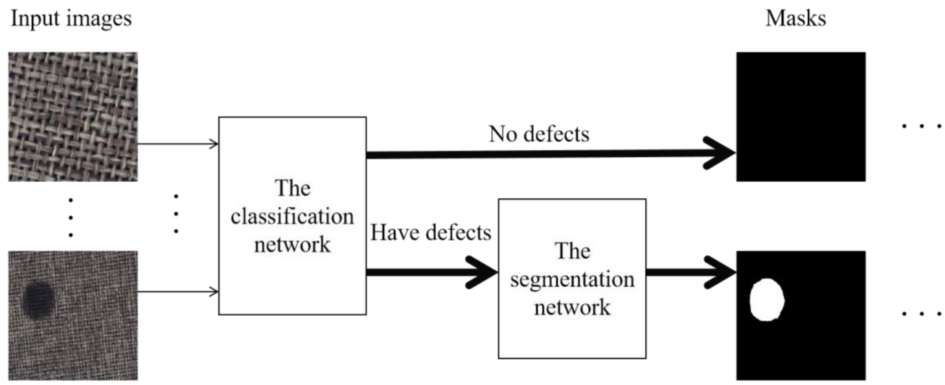


Fig. 1. Algorithmic framework

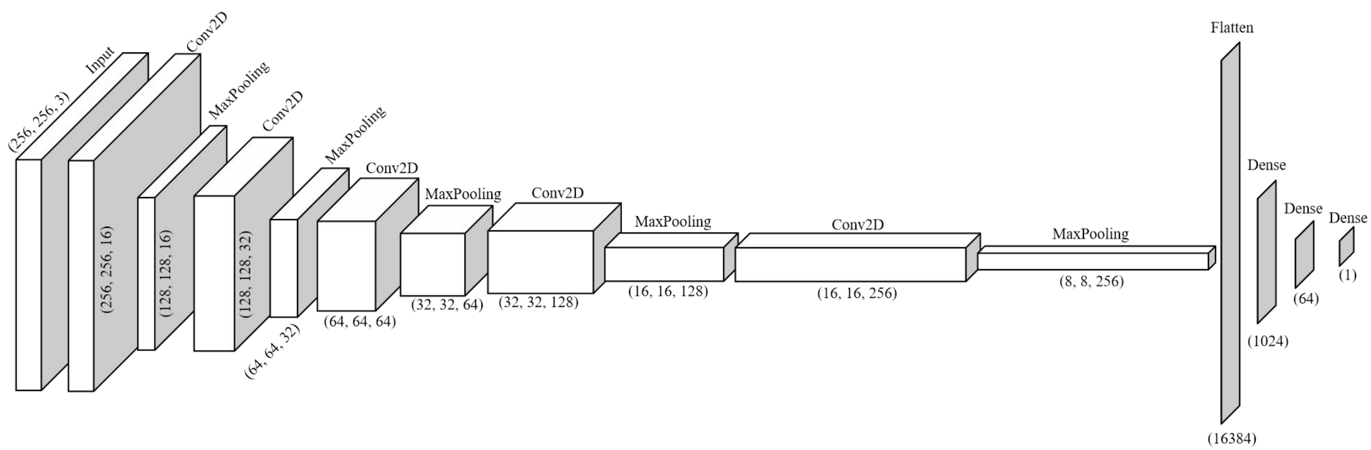


Fig. 2. Classification network

layer is the same as ECON. Next are four Conv2d layers, their sizes are (128, 128, 32), (64, 64, 64), (32, 32, 128), and (16, 16, 256), separately. After each Conv2D layer, there exists a MaxPooling2D layer, which represents an array of (2, 2). After the fifth MaxPooling2D layer, we add a flatten layer, a dense layer (1024), a dropout layer, a dense layer (64), and a dense layer (1) as the final output. In addition, for each batch of training, the numbers of images with defects and images without defects should be as close as possible, which is helpful to improve the convergence efficiency of the model.

The training data we input is all the data in the training set, and we expect the model to output a float probability value(between 0 and 1) to indicate whether the input picture contains defects (0 means no defects, while 1 means have defects), and use 0.3 as the threshold for binary classification.

C. Binary-level Segmentation

The U-Net++ model [20] has 16 Conv2D layers, 10 Conv2DTranspos layers, 4 MaxPooling2D layers, and some activation layers and batch-norm layers, and its residual connections bring better performance and make it easier to prune and speed up the model, as shown in Fig. 3. The number of filters of Conv2D layers in the model are respectively 16, 32, 64, 128, and 256, and the input tensor is also an image of an array (256, 256, 3). Different from the CNN model in the first part, the data we input in U-Net++ is only the image data with defects in the training set. Accordingly, we only filter out the images containing defects in verification set, and test set for training. We expect the model to output the semantic

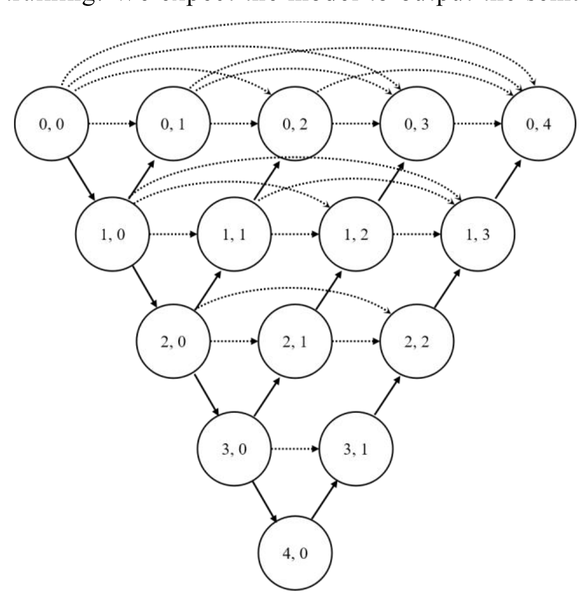


Fig. 3. The network of U-Net++

Authorized licensed use limited to: University of Maryland College Park. Downloaded on August 27,2023 at 16:41:20 UTC from IEEE Xplore. Restrictions apply.

segmentation prediction results and use 0.5 as the threshold for binary classification (0 indicates the background here, while 1 indicates the defects).

For a new image, it will be put into the CNN part firstly. If the result is greater than 0.3, we assert that the image may have some defects, and then send it to the next part for further segmentation. After all, the final output semantic segmentation prediction result is the segmentation of defective parts of the image. This separate training method helps reduce the training burden of the second model and focuses on the semantic segmentation of the defective parts, making the training process easier to implement and can be optimized in the future.

IV. EXPERIMENT

A. Preparation Datasets

We deploy our model on two different datasets to complete the task of pre-classification and segmentation based on transfer learning technique. And we also use U-Net as a comparative model in this scenario. Then in order to demonstrate that transfer learning improves semantic segmentation prediction results, we compare transfer learning scenario with direct training scenario, through our proposed model and U-net. The datasets are leather (source dataset) and carpet (target dataset) from MVTEC anomaly detection dataset (MVTEC AD) [21]. MVTEC AD dataset is for benchmarking anomaly detection methods with a focus on industrial inspection. It contains over 5000 high-resolution images divided into fifteen different object and texture categories, which has 3629 images for training and 1725 images for testing. The training set includes only images without defects. The test set contains both: images containing various types of defects and defect-free images. Each images follows with a according fully annotated ground truth segmentation map for defects (white) and backgrounds (black). The experiments environment is as follows: GPU: NVIDIA Geforce RTX 3060 (laptop), CPU:AMD Ryzen R7-5800h, memory: 16G, pycharm (education edition). The paper uses Adam as an optimization algorithm, and set the initial learning rate to 0.001. Batch size for CNN is 50 and U-Net++ is 4, respectively. Epochs under direct training scenario are 100 and under transfer learning scenario epochs of each state are 50.

B. Results

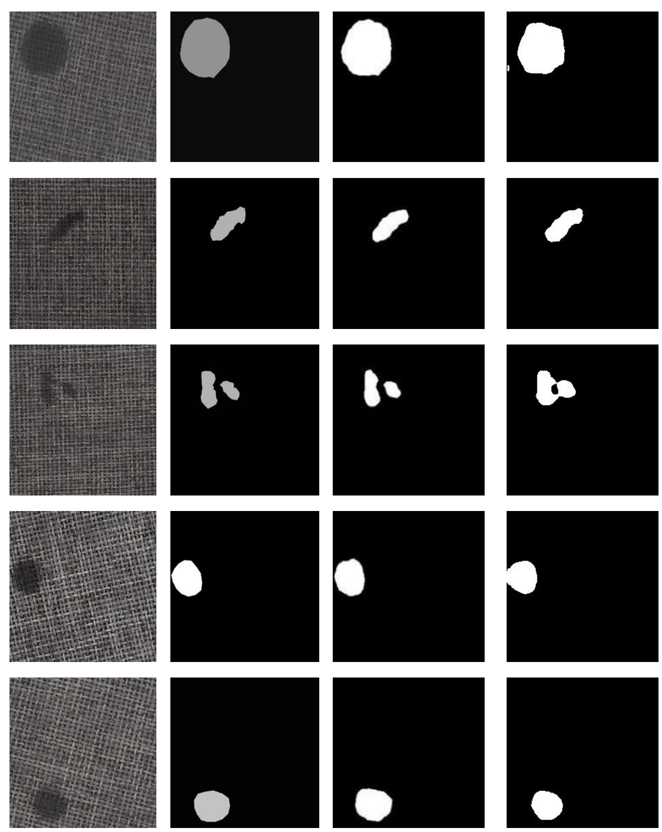
In the field of industrial production, there is still a strenuous procedure to locate various types of defects, which consumes a lot of time and energy. Then scholars racked their brains to come up with a method, semantic segmentation, to better cope with this task. In order to further facilitate this process, we come out with an idea that before letting our model perform semantic segmentation of defects, we pick out the pictures without defects first. Under this circumstance, the model can focus more on capturing the location of defects rather than judging whether images are defective or not, which improves accuracy and efficiency. We use the mean intersection over union (mIOU) and pixel accuracy (PA) of the semantic segmentation map obtained by prediction as the metrics to evaluate models' performance:

$$mIOU = \frac{1}{k+1}\sum_{i=0}^{k} \frac{TP}{FN+FP+TP} \quad (1)$$

$$PA = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

Our model achieves the highest score in transfer learning scenario with mIOU of 0.741 and PA of 0.889 (see Table II). This significantly outperforms U-Net whose mIOU is 0.466 by nearly 0.3 and Dice coefficient is 0.839 by nearly 6%. And the comparison of the results of transfer learning scenario and direct training scenario is also illustrated in Table II, which indicates transfer learning technique can boost efficiency and accuracy of semantic segmentation on defects detection.

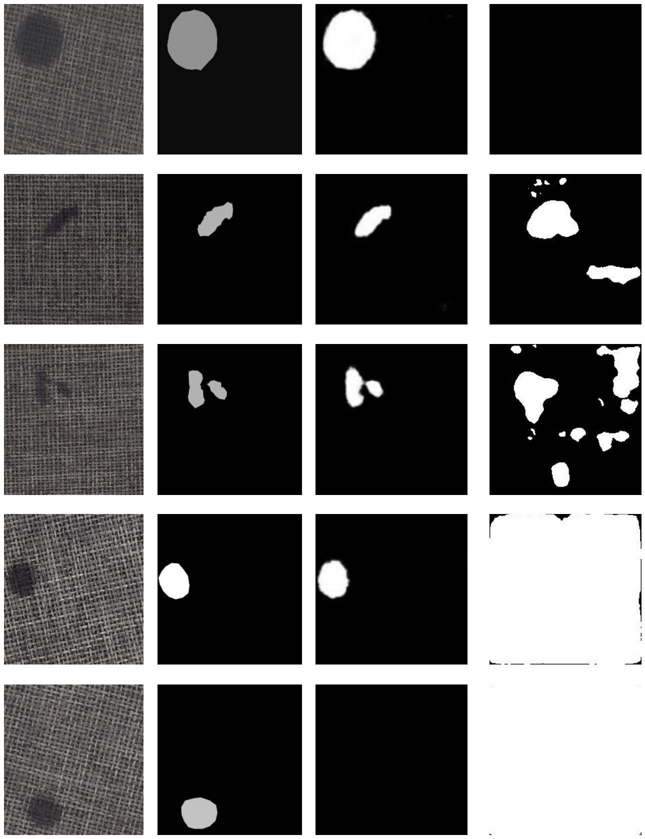
Fig. 4 shows some predication results of defect detection under transfer learning scenario. The first row in Fig. 4 is raw images with defects. It illustrates that the predication results of our model are the closet to the ground truth compared to U-net. And we can also see that in Fig. 5, under direct learning scenario, the predicated pictures have produced many detection errors. The results of U-Net under both scenario have much more detection errors compared to our proposed model. detection performance in semantic segmentation over U-Net under both scenario and the performance of both models under transfer learning scenario are better than that under direct learning scenario.



(a) raw image (b) ground truth (c) segmentation map of our model (d) segmentation map of U-net

Fig. 4. The predication results of transfer learning scenario

Authorized licensed use limited to: University of Maryland College Park. Downloaded on August 27,2023 at 16:41:20 UTC from IEEE Xplore. Restrictions apply.



(a) raw image (b) ground truth (c) segmentation map of our model (d) segmentation map of U-net

Fig. 5. The predication results of direct learning scenario

TABLE II. SEGMENTATION RESULT OF TWO SCENARIOS

Model	**Scenario**	**Mean Intersection over Union**	**Pixel Accuracy**
U-Net	Direct training	0.436	0.787
Proposed model	Direct training	0.675	0.866
U-Net	Transfer learning	0.466	0.839
Proposed model	Transfer learning	**0.741**	**0.889**

V. CONCLUSION

This paper proposes a defect detection model based on transfer learning for multi-variety industrial products, which consists of two parts: classification and segmentation. Our model can effectively pick up images containing various types of defects. Then through semantic segmentation method, it is able to detect defects and produce predication results. In the meantime it has a performance exceeding the state-of-the-art model, U-net. Especially, the proposed model achieves good results under transfer learning scenario compared to direct learning scenario, which better demonstrate that by using transfer learning technique the defect detection efficiency and accuracy will be highly improved. Therefore, our future work should also focus on using different CNN as classifier to further improve our model's performance and applying transfer learning method in more downstream tasks.

REFERENCES

[1] B. Tang, J. Y. Kong, and S. Y. Wu, "Review of surface defect detection based on machine vision," *Journal of Image and Graphics*, vol. 22, no. 12, pp. 1640-1663, 2017.

[2] T. Zhang, Y. T. Liu, Y. N. Yang, X. Wang, and Y. G. Jin, "Review of surface defect detection based on machine vision," *Science Technology and Engineering*, vol. 20, no. 35, pp. 14366-14376, 2020.

[3] H. Su, J. B. Zhang, B. H. Zhang, and W. Zou, "Review of research on the inspection of surface defect based on visual perception," *Computer Integrated Manufacturing Systems*, vol. 14, pp. 1-13, 2021.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84-90, 2017.

[5] A. Carrio, C. Sampedro, A. Rodriguez-Ramos, and P. Campoy, "A review of deep learning methods and applications for unmanned aerial vehicles," *Journal of Sensors and Sensor Systems*, vol. 2017, 2017.

[6] J. Ker, L. Wang, J. Rao and T. Lim, "Deep learning applications in medical image analysis," *IEEE Access*, vol. 6, pp. 9375-9389, 2018.

[7] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A. Mohamed, N. Jaitly et al., "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82-97, Nov. 2012.

[8] Z. Zheng, H. Yang, L. Zhou, B. Yu, and Y. Zhang, "HLU2-net: A residual U-structure embedded U-Net with hybrid loss for tire defect inspection," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-11, 2021.

[9] X. F. Ling, Y. P. Wu, R. Ali, and H. Z. Zhu, "Magnetic tile surface defect detection methodology based on self-attention and self-supervised learning," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 3003810, 2022.

[10] Y. Huang, J. Jing, and Z. Wang, "Fabric defect segmentation method based on deep learning," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-15, 2021.

[11] L. Zhou, L. Zhang, and N. Konz, "Computer Vision Techniques in Manufacturing," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 53, no. 1, pp. 105-117, 2023.

[12] C. L Yang and X. Q. Zhang, "Survey of applications of material defect detection based on machine vision and deep learning," *Materials Reports*, vol. 36, no. 16, pp. 226-234, 2022.

[13] M. W. Akram, G. Li, Y. Jin, X. Chen, C. Zhu, X. Zhao et al., "CNN based automatic detection of photovoltaic cell defects in electroluminescence images," *Energy*, vol. 189, no. 116319, pp. 1-15, 2019.

[14] Y. Lu, B. Chen, D. Chen, S. Yan, and S. Li, "Recognition algorithm of strip steel surface defects based on attention model," *Laser & Optoelectronics Progress*, vol. 58, no. 14, pp. 1-14, 2021.

[15] C. Liu, J. Zhang, and J. P. Lin, "Detection and identification of surface defects of magnetic tile based on neural network," *Surface Technology*, vol. 48, no. 8, pp. 330-339, 2019.

[16] 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 Recognition Letters*, vol. 130, pp. 234-241, 2020.

[17] H. Liu and G. L. Xu, "LCD panel defect detection based on multi source domain depth transfer learning," *Fujian Quality Management*, vol. 2019, no. 6, pp. 250-252, 2019.

[18] P. Damacharla, A. Rao, J. Ringenberg, and A. Y. Javaid, "TLU-Net: A Deep Learning Approach for Automatic Steel Surface Defect

Authorized licensed use limited to: University of Maryland College Park. Downloaded on August 27,2023 at 16:41:20 UTC from IEEE Xplore. Restrictions apply.

Detection," in *International Conference on Applied Artificial Intelligence (ICAPAI)*, May. 2021, pp. 1-6.

[19] L. F. Zhou and L. Zhang, "A novel convolutional neural network for electronic component classification with diverse backgrounds," *International Journal of Modeling, Simulation, and Scientific Computing*, vol. 13, 2022.

[20] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A Nested U-Net Architecture for Medical Image Segmentation," *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, vol. 11045, pp. 3-11, 2018.

[21] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, "MVTec AD — a comprehensive real-world dataset for unsupervised anomaly detection," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 9592-9600.

Authorized licensed use limited to: University of Maryland College Park. Downloaded on August 27,2023 at 16:41:20 UTC from IEEE Xplore. Restrictions apply.