Research Article

Using Convolution Neural Network for Defective Image Classification of Industrial Components

Hao Wu [1] **and Zhi Zhou** [2]

[1]*School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200444, China*
[2]*Samsung Electronics (China) Research and Development Center, Nanjing 210012, China*

Correspondence should be addressed to Hao Wu; yzwuhao4038@shu.edu.cn

Computer vision provides effective solutions in many imaging relation problems, including automatic image segmentation and classification. Artificially trained models can be employed to tag images and identify objects spontaneously. In large-scale manufacturing, industrial cameras are utilized to take constant images of components for several reasons. Due to the limitations caused by motion, lens distortion, and noise, some defective images are captured, which are to be identified and separated. One common way to address this problem is by looking into these images manually. However, this solution is not only very time-consuming but is also inaccurate. The paper proposes a deep learning-based artificially intelligent system that can quickly train and identify faulty images. For this purpose, a pretrained convolution neural network based on the PyTorch framework is employed to extract discriminating features from the dataset, which is then used for the classification task. In order to eliminate the chances of overfitting, the proposed model also employed Dropout technology to adjust the network. The experimental study reveals that the system can precisely classify the normal and defective images with an accuracy of over 91%.

1. Introduction

With the rapid development of Internet technology and media, the image data spread on the Internet is growing exponentially every day. How to classify these images is a meaningful work. The traditional image classification mode can only be carried out manually; the efficiency is low, but the detection accuracy is not high. It is difficult for massive data image classification to adapt to the manual method to retrieve the target image. So, we need to capture the concerned information from these data through some algorithms. In engineering, it is necessary to detect the defect image from the image of industrial components. Some abnormal images with defects can be extracted from numerous components images to achieve screening and classification. It is used to measure whether the target image meets the relevant detection standards. In order to overcome the shortcomings of artificial image classification, researchers began to use computer tools for defect image classification in recent years. With the development of machine learning and deep learning technology, some algorithms can be applied to defect image classification and detection, which can improve detection accuracy and promote the growth of efficiency.

Deep learning in image classification involves engineering design, biomedicine, transportation exploration, product quality, and media operation. In [1], a large number of images were collected and annotated. The biomedical images were classified by using the transfer and semi-supervised learning model. An AutoML model was proposed to solve the problem of network training superparameters in deep learning. This method has achieved good results on the annotated biomedical datasets. By analyzing the images on social media, the researchers in [2] used computer vision technology and a deep neural network model to classify the media images in real time. It helped people to perceive the information crisis and assess the loss. In [3], the authors used the potential of combining the spectral and spatial characteristics of deep learning analysis data. They used a three-dimensional convolution neural network model combined with the spectral preprocessing method. The pixel-level classification of food is carried out,

which opens up the application of image processing in food engineering. It can be seen that, with the growth of data, it is essential to classify the image. For the industrial field, we often need to take a large number of components images. With the increase in the number, there will be a lot of defective components images. How to classify these images is a complex work. In the early stage, artificial recognition is mainly used to determine whether there are defects in the part image. The advantage of this method is that it can select compelling features, and the classification accuracy is high. However, the time cost is also high, and the efficiency is low. With the development of computer technology and artificial intelligence, various deep learning models can process and calculate data efficiently, making the accuracy of machine classification gradually improve. The machine learning algorithms gradually replace the artificial way.

Deep learning algorithms such as convolutional neural networks, long-term memory networks, graph neural networks, and generative confrontation networks [4–7] have made progress in the field of image processing. In 1986, Rumelhart et al. [8] proposed the backpropagation algorithm of the artificial neural network, which set off a boom of neural networks in machine learning. Compared with boosting, logistic regression, support vector machine, and other shallow model methods based on statistical learning theory [9–11], it has more tremendous advantages. However, there are many problems in the neural network, such as numerous training parameters, easy overfitting, and long training time. In the later research, we are also committed to improving the robustness and efficiency of the model. The improved model is more suitable for digital image processing. Because the manual extraction of sample features does not restrict the deep model, the multihidden layer artificial neural network has excellent feature learning ability. As a result, the learned data can better reflect the essential characteristics of the data, which is conducive to visualization or classification to identify better and predict. Because of the extensive application of deep learning in image processing, this paper uses convolution neural networks and other models in deep learning to classify and detect defects in large-scale industrial components' images taken by cameras. In order to further improve the classification accuracy, we optimize the parameters and network structure in the training process. At the same time, dropout technology is used to avoid overfitting and other typical problems.

The rest of the paper is organized according to the following pattern. First, in Section 2, the related work is studied, followed by the proposed method in Section 3. Then, the experimental setup and result discussion are performed in Section 4. Finally, the paper is concluded in Section 5.

2. Related Work

Image classification is an essential task in computer vision. Researchers have done much research on image processing before. They either use the machine learning method of manual feature extraction or use the neural network of deep learning to construct the network structure for image classification. At the same time, different models are derived to improve the existing methods. The effect of the classification task is getting better and better, which promotes the development of image processing tasks in computer vision. Hai et al. [12] used a support vector machine to classify burn images in medicine. The model used multicolor channel extraction and binarization method based on the adaptive threshold to obtain image features. However, due to the problems of a simple model and unbalanced data, the accuracy of this method can only reach 77.78%. Huang et al. [13] proposed a spectral, spatial hyperspectral image classification method using a support vector machine to get the initial classification probability map. It is used for neighborhood matching and average KNN filtering algorithm to refine the obtained pixel-level probability map, finally using KNN for decision classification.

In computer image processing, the convolutional neural network is the most common because it can build a hierarchical classifier. It can also be used in finely graded recognition to extract image discriminant features for other classifiers to learn. It supports artificial feature extraction and unsupervised learning training. The convolutional neural network is also widely used in image classification. In order to solve the problem of large parameter space of training network, Lawrence [14] proposed a depth architecture generation model based on particle swarm optimization (PSO). It is used to search space effectively and generate an automatic evolution convolutional neural network to classify images. Sun et al. [15] used a genetic algorithm to design a CNN architecture for automatic image classification, which has achieved good results in a wide range of image classification datasets. Ben [16] proposed a group optimization block structure to evolve the CNN model deeply and established a depth network for image classification based on a convolutional neural network. Pritt [17] et al. used a convolution neural network to combine satellite metadata with image features to solve the problems of manual recognition of satellite images. It covers a comprehensive and complex search to automatically identify the targets and facilities in multispectral satellite images, with an accuracy rate of 95% in 15 categories. In order to solve the problem of the limited capacity of softmax function in traditional convolutional neural network model classification, an image classification method combining bionic pattern recognition (BPR) with CNN is proposed, which can classify and recognize objects in high-dimensional feature space by geometric coverage.

3. Proposed Method

A convolution neural network is a kind of deep neural network with a convolution structure. Except for the traditional image processing methods [18–20], the convolution neural network has achieved good image processing tasks and has strong generalization ability. Convolutional neural network is used in many tasks, such as image classification [21], object detection [22–24], instance segmentation [25, 26], and scene understanding [27–30]. This network is

characterized by local perception, weight sharing, and pooling, which can effectively reduce the number of network parameters and quickly capture the deep features of the input image. In this paper, an improved convolution neural network structure is used to classify the defect images of industrial components. This method can train the existing industrial components' image samples, learn the features of data synthesis, build a network structure with stronger expression ability, and automatically mine the feature engineering of data, which can quickly identify the defect images. Solve the problem of qualified components inspection in the industry. See Figure 1, for details.

A convolutional neural network consists of a convolution layer, pooling layer, and fully connected layer. The goal of the convolution layer is to mine more representative input features, and the pooling layer is to reduce the spatial dimension. In contrast, the fully connected layer is used for category prediction. Firstly, the model obtains a new feature map by convolution operation between the input and an automatically learned convolution kernel and then applies a nonlinear activation function to activate the network layer element by element on the convolution result. As an essential part of convolutional neural network, the pooling layer reduces the size of the feature map. Generally, some calculation method is used to fuse the information of a region, such as maximum pooling (using the maximum value of a region to replace the information of the region) and average pooling (using the average value of a region to replace the information of the region). Pooling is a subsampling operation. Its main goal is to reduce the feature space of feature maps or the resolution of feature maps. However, there are too many feature map parameters, and image details are not conducive to high-level feature extraction. The operation mode of pooling is to set a sliding window on the input and send the window contents into the pooling function for calculation. By adding the pooling layer, the image is reduced. As a result, the calculation amount is greatly reduced, and the machine load can be reduced. After several convolution and pooling operation layers, the obtained feature graphs are expanded by rows, connected into vectors, and input into the fully connected network.

For the input spatial two-dimensional image m, its coordinates are (x, y); using the two-dimensional convolution kernel K, the convolution calculation is as follows:

$$S(x, y) = (K * I)(x, y) = \sum_m \sum_n M(m, n) K(x - m, y - n).$$

(1)

Assuming that the size of the convolution kernel is $p * q$ and the kernel weight is w, the convolution process is the sum of all the kernel weights and the brightness of their corresponding elements in the input image:

$$\text{Conv}_{x,y} = \sum_i^{p*q} w_i v_i.$$

(2)

After convolution, bias is usually added, and a nonlinear activation function is introduced. Here, bias is defined as b,

and the activation function is $h(x)$. After activation function, the result is

$$z_{x,y} = h\left(\sum_i^{p*q} w_i v_i + b\right).$$

(3)

The activation function is generally tanh function:

$$h(z) = \tanh(z),$$
$$\tanh(z) = \frac{e^x - e^{-x}}{e^x + e^{-x}}.$$

(4)

In a convolution neural network, a full connection neural network is introduced to classify images. After full connection layer, for the ith neuron in layer L, its output calculation method is as follows:

$$z_i(l) = \sum_{j=i}^{n_{l-1}} w_{i,j}(l) a_j(l-1) + b_i(l).$$

(5)

The training error of a convolutional neural network needs to be measured by an objective function. At present, the more popular objective functions are mean square error and K-L divergence. This paper refers to binary classification the defect images in industrial components, so the K-L divergence method is adopted. In the defect detection task, the image is preprocessed first, input to the convolutional neural network training output, and expressed as a feature map, inputting the feature map to the full connection layer and using sigmoid activation function for binary classification:

$$\text{Dis}(KL) = -\frac{1}{nL} \sum_{j=1}^{nL} \left[r_j \ln\left(\partial_j(L)\right) + \left(1 - r_j\right)\ln\left(1 - \partial_j(L)\right)\right],$$

(6)

where r_j represents the label of the image and $\partial_j(L)$ is the output of the jth neuron in layer L. The weight is updated by gradient descent, where α is the learning rate:

$$W_{ij}(L) = W_{ij}(L) - \alpha \frac{\nabla \text{Dis}(KL)}{\nabla W_{ij}(L)},$$
$$b_i(L) = b_i(L) - \alpha \frac{\nabla \text{Dis}(KL)}{\nabla b_i(L)}.$$

(7)

The weight update process of its image is shown in Figure 2.

4. Experiment

4.1. Experiment Method. Because of the advantages of convolution neural networks in image feature processing, this paper obtains the image data of components captured by a specific industrial camera. These components are the inputs into the deep convolution neural network as a digital image for training and finally select the abnormal image through the classifier. The experiment mainly includes four steps: image acquisition, preprocessing, feature extraction, and data
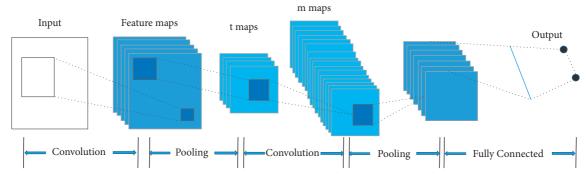
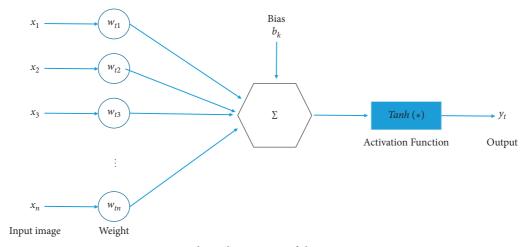FIGURE 1: Structure of convolutional neural network.

FIGURE 2: Weight update process of the processing image.

classification. First, the image comes from a large number of components, and digital images are captured by industrial cameras. Second, the preprocessing uses OpenCV to denoise the image and geometric correction, read the image into the array, and reform the size we need. Then, the preprocessed image is transferred to the model to find the classification attributes to describe the differences between the current image and other graphics. Finally, the features trained by a convolution neural network are sent to the classifier to identify the image of the defective component, and then, the components are screened. See Figure 3, for details.

In this paper, based on the pretrained convolutional neural network and multilayer perceptron in the PyTorch framework, for a large number of image data, the existing industrial components' image annotation training is carried out. In the actual training process, due to the high complexity of the model and the imbalance of data, the overfitting problem will be caused. In this experiment, dropout technology is added to solve the overfitting problem better. Some neurons [31] are discarded randomly in convolutional neural network training, and the dropout rate is set to 0.5.

4.2. Experimental Results and Evaluation

4.2.1. Comparison of Defect Image Classification under Different Algorithms. In order to measure the ability of the

convolutional neural network [32] to capture image features [33–35] of components and the classification of classifier, P (precision), R (recall), and F_1 (F_1 score) are used as evaluation indexes in the experiment:

$$P = \frac{TP}{TP + FP},$$

$$R = \frac{TP}{TP + FN}, \tag{8}$$

$$F_1 = \frac{TP + TN}{TP + TN + FP + FN} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}},$$

where TP is the true positives, TN represents the true negatives, FP names the false positives, and FN names the false negatives.

In the part image classification task of the industrial camera, we compare KNN, SVM, and BP neural networks [36]. These algorithms are commonly used in industrial image classification and verify the effectiveness of the convolution neural network in image classification after processing and fitting.

Table 1 illustrates P, R, and F_1 of five different models for image classification. It can be seen from the table that the performance of the neural network model is better than the traditional machine learning models such as SVM and KNN and convolutional neural networks. These traditional algorithms combined with anti-overfitting
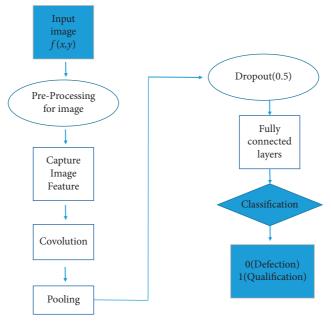
FIGURE 3: Flowchart of using convolution neural network to classify components of the image.

TABLE 1: P, R, and F_1 of image defect classification of components by different methods.

Models	Precision	Recall	F_1
SVM	76.3	84.8	80.32
KNN	79.6	80.2	79.9
BPN	84.7	77.4	80.9
MLP	85.5	76.5	80.8
CNN	**91.4**	**84.9**	**88.0**

technology dropout can more effectively determine the image defect detection. The accuracy rate, recall rate, and F_1 score of convolutional neural network in industrial camera components images are 91.4%, 84.9%, and 88.0%. Thus, to a certain extent, it shows that a convolutional neural network can effectively propose image features and classify them in image processing tasks.

In order to reflect the performance comparison of these five algorithms, several defect images are used as experimental performance comparisons. Then, the average value is taken to evaluate the performance index of the algorithm; specifically, from Precision, Recall, and $F1$, three indexes are evaluated. It can be seen from Figures 4–6 that the performance indexes under F, R, and $F1$ are consistent with the performance shown in Table 1.

4.2.2. Comparison of Classification Algorithms under Different Feature Dimensions.
Based on component defect image processing, SVM, KNN, and CNN are used to calculate the classification accuracy of defect images processed by each preprocessing algorithm and evaluate the effectiveness of quantitative analysis of different preprocessing algorithms. Image processing under continuous dimensions, as shown in Figure 7; with the increase of dimensions, the classification accuracy of defect images is constantly improving. The accuracy under three different algorithms is
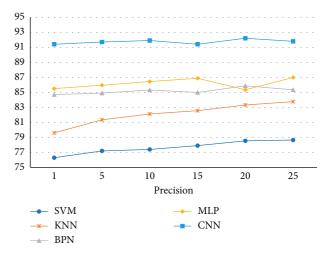

FIGURE 4: Comparison of average value under the Precision index.

also constantly improving. In contrast, the CNN algorithm has the most apparent improvement effect on accuracy. However, CNN's algorithm performance is still the best in running time, as shown in Figure 8.

When the dimension increases, the execution time of the three algorithms decreases, and when the dimension is 11, the execution time of the three algorithms increases. The overall difference in different dimensions is not apparent,
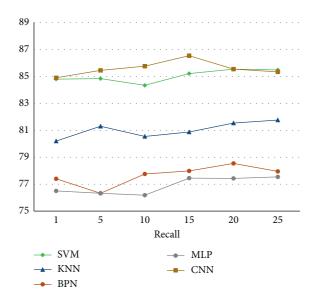
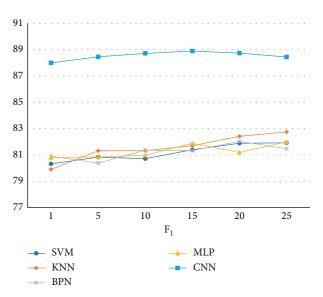Figure 5: Comparison of average value under Recall index.

Figure 6: Comparison of average value under F1 index.
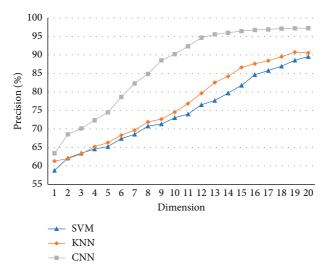

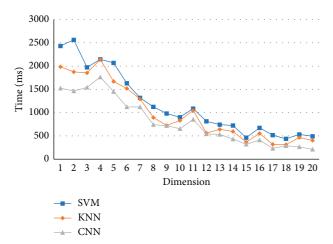Figure 7: Comparison of algorithm accuracy in different dimensions.

Figure 8: Comparison of algorithm time under different dimensions.

and the execution time of CNN performance is still the smallest.

5. Conclusion

In this paper, based on the PyTorch framework, we use the convolutional neural network, combined with dropout technology, to classify and detect industrial camera components' images and screen out the defective components. Firstly, the digital part image captured by the industrial camera is obtained. Next, the image is denoised and geometrically modified by OpenCV to get the appropriate size. Then, the image is input into a convolution neural network for feature extraction and training. Finally, the image is detected and classified by classification function to identify the defective part image. In order to prevent overfitting, the learning rate is adjusted in the training process, and dropout technology is used to discard some neurons randomly. Compared with the standard machine learning models SVM, KNN, BP, and MLP, the results show that P, R and $F1$ indexes of these models can reach 91.4%, 84.9%, and 88.0%, respectively. Thus, it proves that the proposed method is effective in classifying industrial camera components defect detection. Furthermore, through the multi-image and multidimensional performance test, the performance of the CNN algorithm is also the best.

Data Availability

The data used to support the findings of the study are available from the corresponding author upon request.

Conflicts of Interest

The authors declared that they have no conflicts of interest regarding this work.

References

[1] A. Inés, C. Domínguez, J. Heras, E. Mata, and V. Pascual, "Biomedical image classification made easier thanks to transfer and semi-supervised learning," *Computer Methods and Programs in Biomedicine*, vol. 198, Article ID 105782, 2021.

[2] F. Alam, T. Alam, F. Ofli et al., "Social media images classification models for real-time disaster response," 2021.

[3] J. Xu, "Deep spectral-spatial features of near infrared hyperspectral images for pixel-wise classification of food products," *Sensors*, vol. 20, p. 5355, 2020.

[4] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional neural networks," in *Proceedings of the 2013 European Conference on Computer Vision*, Springer International Publishing, Barcelona, Spain, February 2013.

[5] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[6] R. Li, W. Sheng, F. Zhu et al., "Adaptive graph convolutional neural networks," 2018.

[7] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza et al., "Generative adversarial networks," *Advances in Neural Information Processing Systems*, vol. 3, pp. 2672–2680, 2014.

[8] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, *Learning Internal Representations by Error Propagation*, MIT Press, Cambridge, MA, USA, 1988.

[9] T. G. Dietterich, "An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization," *Machine Learning*, vol. 40, no. 2, pp. 139–157, 2000.

[10] P. D. Allison, *Logistic Regression Using the SAS System: Theory and Application*, SAS Publishing, Cary, NC, USA, 1999.

[11] C. Saunders, M. O. Stitson, J. Weston et al., "Support vector machine," *Computer*, vol. 1, no. 4, pp. 1–28, 2002.

[12] T. S. Hai, L. M. Triet, L. H. Thai, and N. T. Thuy, "Real time burning image classification using support vector machine," *EAI Endorsed Transactions on Context-aware Systems and Applications*, vol. 4, no. 12, Article ID 152760, 2017.

[13] K. Huang, S. Li, X. Kang, and L. Fang, "Spectral-spatial hyperspectral image classification based on KNN," *Sensing and Imaging*, vol. 17, no. 1, pp. 1–13, 2016.

[14] T. Lawrence, L. Zhang, C. P. Lim, and E.-J. Phillips, "Particle swarm optimization for automatically evolving convolutional neural networks for image classification," *IEEE Access*, vol. 9, pp. 14369–14386, 2021.

[15] Y. Sun, B. Xue, M. Zhang, G. G. Yen, and J. Lv, "Automatically designing CNN architectures using the genetic algorithm for image classification," *IEEE Transactions on Cybernetics*, vol. 50, no. 9, pp. 3840–3854, 2020.

[16] B. Fielding and L. Zhang, "Evolving image classification architectures with enhanced particle swarm optimisation," *IEEE Access*, vol. 6, pp. 68560–68575, 2018.

[17] M. Pritt and G. Chern, "Satellite image classification with deep learning," 2020.

[18] M. Zhao, A. Jha, Q. Liu et al., "Faster mean-shift: GPU-accelerated clustering for cosine embedding-based cell segmentation and tracking," *Medical Image Analysis*, vol. 71, Article ID 102048, 2021.

[19] Y. Jiang, X. Gu, D. Wu et al., "A novel negative-transfer-resistant fuzzy clustering model with a shared cross-domain transfer latent space and its application to brain CT image segmentation," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18, no. 1, pp. 40–52, 2021.

[20] M. Zhao, Q. Liu, A. Jha et al., "VoxelEmbed: 3D instance segmentation and tracking with voxel embedding based deep learning," 2021, http://arxiv.org/abs/2106.11480.

[21] Y. Jiang, Y. Zhang, C. Lin, D. Wu, and C.-T. Lin, "EEG-based driver drowsiness estimation using an online multi-view and

transfer TSK fuzzy system," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 3, pp. 1752–1764, 2021.

[22] J. Zhang, Y. Liu, H. Liu, and J. Wang, "Learning local-global multiple correlation filters for robust visual tracking with kalman filter redetection," *Sensors*, vol. 21, no. 4, p. 1129, 2021.

[23] J. Zhang, J. Sun, J. Wang, and X. G. Yue, "Visual object tracking based on residual network and cascaded correlation filters," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, pp. 1–14, 2020.

[24] Z. Huang, P. Zhang, R. Liu, and D. Li, "Immature apple detection method based on improved Yolov3," *ASP Transactions on Internet of Things*, vol. 1, no. 1, pp. 9–13, 2021.

[25] X. Zhang, Y. Yang, Z. Li, X. Ning, Y. Qin, and W. Cai, "An improved encoder-decoder network based on strip pool method applied to segmentation of farmland vacancy field," *Entropy*, vol. 23, no. 4, p. 435, 2021.

[26] Q. Liu, I. M. Gaeta, M. Zhao et al., "ASIST: annotation-free synthetic instance segmentation and tracking by adversarial simulations," *Computers in Biology and Medicine*, vol. 134, Article ID 104501, 2021.

[27] X. Ning, K. Gong, W. Li, L. Zhang, X. Bai, and S. Tian, "Feature refinement and filter network for person re-identification," *IEEE Transactions on Circuits and Systems for Video Technology*, 2020, In press.

[28] W. Cai, Z. Wei, R. Liu, Y. Zhuang, Y. Wang, and X. Ning, "Remote sensing image recognition based on multi-attention residual fusion networks," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 1–8, 2021.

[29] R. Liu, X. Ning, W. Cai, and G. Li, "Multiscale dense cross-attention mechanism with covariance pooling for hyperspectral image scene classification," *Mobile Information Systems*, vol. 2021, Article ID 9962057, 15 pages, 2021.

[30] C. Yan, G. Pang, X. Bai et al., "Beyond triplet loss: person re-identification with fine-grained difference-aware pairwise loss," *IEEE Transactions on Multimedia*, 2021, In press.

[31] X. Ning, Y. Wang, W. Tian, L. Liu, and W. Cai, "A biomimetic covering learning method based on principle of homology continuity," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 9–16, 2021.

[32] J. Zhang, W. Wang, C. Lu, J. Wang, and A. K. Sangaiah, "Lightweight deep network for traffic sign classification," *Annals of Telecommunications*, vol. 75, no. 7, pp. 369–379, 2020.

[33] M. Li, G. Zhou, W. Cai et al., "MRDA-MGFSNet: network based on a multi-rate dilated attention mechanism and multi-granularity feature sharer for image-based butterflies fine-grained classification," *Symmetry*, vol. 13, no. 8, p. 1351, 2021.

[34] W. Cai, Z. Wei, Y. Song, M. Li, and X. Yang, "Residual-capsule networks with threshold convolution for segmentation of wheat plantation rows in UAV images," *Multimedia Tools and Applications*, pp. 1-17, 2021, In press.

[35] S. Qi, X. Ning, G. Yang et al., "Review of multi-view 3D object recognition methods based on deep learning," *Displays*, vol. 69, Article ID 102053, 2021.

[36] L. Huang, G. Xie, W. Zhao, Y. Gu, and Y. Huang, "Regional logistics demand forecasting: a bp neural network approach," *Complex & Intelligent Systems*, pp. 1-16, 2021, In press.