WOOD RESEARCH doi.org/10.37763/wr.1336-4561/67.3.501510
e-ISSN 2729-8906
67(3): 2022 501-510 pp.

BAMBOO DEFECT CLASSIFICATION BASED ON IMPROVED TRANSFORMER NETWORK

JUNFENG HU, XI YU, YAFENG ZHAO
NORTHEAST FORESTRY UNIVERSITY
CHINA

(RECEIVED NOVEMBER 2021)

ABSTRACT

Deep learning-based methods, especially convolutional neural networks (CNNs), have shown their effectiveness for image classification. In this paper, vision transformer technology is used to classify the surface defects of processed bamboo, which can be more quick and accurate compared with the low efficiency of manual identification. In the first step, we replace the activation function from Gelu to Mish in the encoder part, but the classification performance is not satisfied. Then, to get a better classification results, we keep the original activation function and introduce the DropBlock. Compared with dropout, DropBlock can obtain better classification accuracy. Finally, compared with the results after transfer learning, it is proved that replacing dropout with DropBlock can improve the classification accuracy. The results on the bamboo chip datasets show that the accuracy of this method is 2% higher than the original transformer network whether using transfer learning.

KEYWORDS: Deep learning, transformer, convolutional neural network (CNN), transfer learning.

INTRODUCTION

Many approaches for detecting and classifying image feature flaws have been presented. Mallik et al. (2011) proposed a method based on HOG feature and support vector machine classifier are used to detect whether there are grape leaves in the image, and the sliding window method is used to search leaves (Felzenszwalb et al. 2010). This method has good detection effect for the leaves with positive position, and poor detection accuracy for the leaves with incorrect position and incomplete surface. An online automated visual inspection system based on the genetic algorithm is designed to classify the defects of radiata pine boards (Estévez et al. 2003), the system achieved a classification accuracy of over 80%. A basic k-means algorithm is commonly used as a texture recognition method, however it's difficult to acquire a better

recognition effect since the starting midpoint to reach the local optimal solution is readily impacted (Patgar et al. 2014, Venkateswaran et al. 2013). Chen et al. (2014) proposed a stacked autoencoder for feature extraction. Lin et al. (2012) proposed an approach to represent and recognize objects with a massive number of local image patches, can directly update the feature weights by defining and calculating feature correlations. This method works well with the task of object detection and localization from images. Differential evolution (DE) algorithm is an evolutionary algorithm based on swarm intelligence, which has been used to solve the optimal cluster center and achieved good results (Kwedlo et al. 2011, Kuo et al. 2013). A Kernel-genetic algorithm technique is proposed for feature selection, this method combines the Kernel discriminant analysis (KDA) technique with genetic algorithm (GA) to generate nonlinear wood features and at the same time reduce the dimension of the wood database. The proposed system achieved a classification accuracy of 98.69% (Yusof et al. 2013). The scope of wood surface defects was determined by 3D image (Sioma. 2015), and the defect area was defined accurately, which enhanced the ability of wood surface defect evaluation. Gonzalo et al. (2009) presented an improved low-cost automated visual inspection (AVI) system for wood defect classification. To improve the robustness of binary mode to noise, a method was proposed to segment the pixel difference between the center pixel and adjacent pixels of the image object into local ternary mode (Tan et al. 2010). K-nearest neighbor (KNN) classification algorithm has been proposed to classify the wood knot images (Cetiner et al. 2014), knot images are correctly divided into seven different categories with a correct rate of 98% by the authors. A method based on the calculation of a Gray level cooccurrence matrix and fuzzy BP neural network, the final recognition rate reached 90% (Mu et al. 2015). Hanbay et al. (2016) proposed that the principal curvature of image can not only improve the robustness of classification neural network to image rotation changes, but also obtain the macro structure and micro structure information of target features at the same time, which can improve the effectiveness of classification algorithm. Honghai et al. (2021) studied the influence of long-hop connection based on convolutional neural network on bamboo classification. The results show that the network model with long-hop connection has faster convergence speed and will not overfitting, but the classification accuracy is low in the face of large model.

With the development of deep learning and computer vision technology, the main method for solving bamboo defect detection is to use deep learning technology. CNN, as a technical direction of deep learning, has successfully made a great breakthrough in image classification (Krizhevsky et al. 2012), can accurately extract the characteristics of the object in the image, train a large number of extracted object feature data, then it can classify the characteristics of the object quickly and efficiently. One contribution of this paper is proposed an effective deep learning methodology, which is used to identification bamboo slices. Zhang et al. (2017) adopted a recursive auto-encoder (RAE) as a high-level feature extractor to produce feature maps from the target pixel neighborhoods. Chen et al. (2015) combined deep belief network (DBN) and restricted Boltzmann machine (RBM) for hyperspectral image classification. Wei et al. (2017) proposed a CNN-based feature extractor to finish picture classification by learning discriminative representations from pixel pairs. To boost classification performance, Xin et al. (2019) introduced the DropBlock (Golnaz et al. 2018) as a method of regularize convolutional

networks and optimized the input with STN based on transformer, the result of the method performs well. Leyuan et al. (2019) introduced a novel squeeze multibias network that replaced the typical convolutional layer with a squeeze convolution module, reducing the amount of network parameters while maintaining good classifiability accuracy. The deeper layers of neural network, the higher accuracy of the model achieve in the same datasets (Szegedy et al. 2015). Deep learning has become extremely popular because of its ability to extract features from raw data. It has been applied in computer vision tasks, such as image classification (Kaimign et al. 2016, Yuntao et al. 2018), object detection (Shaoqing et al. 2016), semantic segmentation (Kaiming et al. 2017) and facial recognization (Nagpal et al. 2019). Deep learning has the advantages of strong learning ability, high efficiency, strong adaptability and good portability. However, its disadvantages are obvious, such as large amount of calculation, high cost of hardware and complex model design.

The transformer, a deep learning model, has recently been used for computer vision applications. Transformers were introduced for machine translation and they quickly became the industry standard in many natural language processing (NLP) applications (Vaswani et al. 2017). Dosovitskiy et al. (2021) proposed a direct application of the transformer network to image recognition by vision transformer. Xiang et al. (2021) utilized the 1-D convolution layer to get the embedding of each sequence and adopt the Mish as the activation function, the experiment get a better classification performance. Transfer learning is easy to build a deep layers model and leverage the feature extracting capability of the trained layers (Pan et al. 2010).

The encoder and decoder blocks are the most significant sections of the transformer, as shown in Fig. 1, therefore knowing these two parts is highly useful in comprehending the transformer as a whole. The Transformer's encoder block is made up of numerous identical layers, including the multi-head self-attention and feed-forward network. The residual structure is applied to each sub-layer in order to avoid deterioration during deep neural network training.

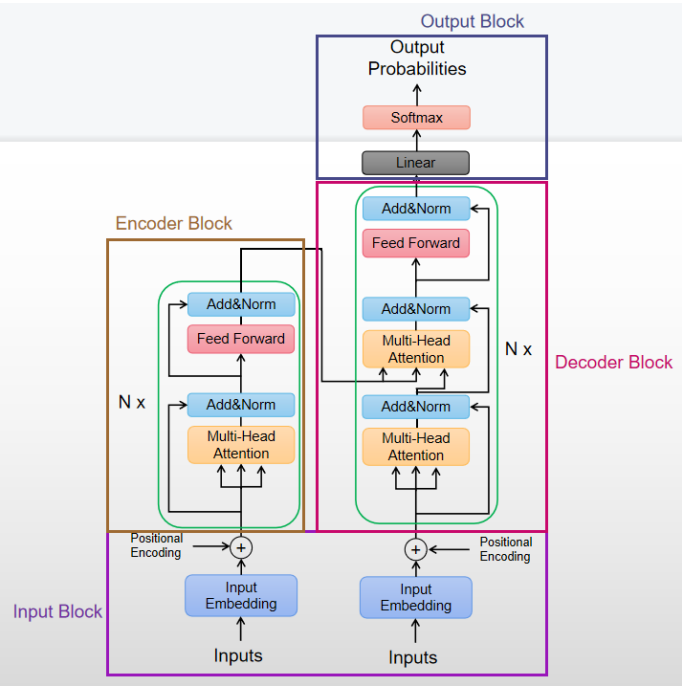


Fig. 1: The structure of transformer (via: arXiv:1706.03762).

The output of each sub-layer is shown in Eq. 1, the multi-head self-attention is defined in Eq. 2, and the attention is shown in Eq. 3:

$$LayerNorm(x+(SubLayer(x))) \tag{1}$$

$$MultiHead(Q,K,V)=Concat(head_1,head_2,\cdots,head_h)W^o \tag{2}$$

$$Attention()=soft\max(\frac{QK^T}{\sqrt{d_k}})V \tag{3}$$

where: $SubLayer(x)$ - the function implemented by the sub-layer,

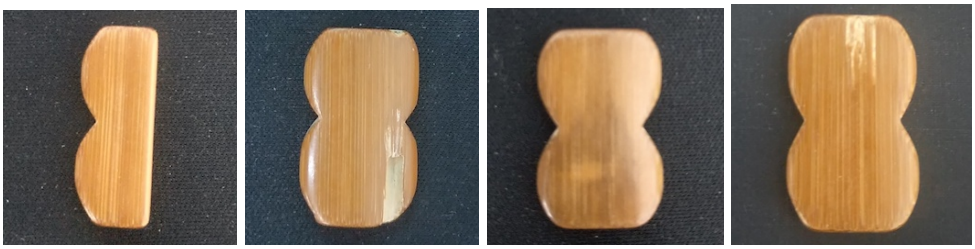
$head_i=Attention(QW_i^Q,KW_i^Q,VW_i^Q)$, $W_i^Q\in R^{d_{model}\times d_q}$, $W_i^K\in R^{d_{model}\times d_k}$, $W_i^V\in R^{d_{model}\times d_v}$, $W_i^O\in R^{d_{model}\times d_v\times h}$ and Q,K - dimension d_k, and V of dimension d_v.

In this paper, we present a lightweight network for bamboo detect image classification, which is inspired by the vision transformer. Fortunately, transfer learning is used in our method, it can classify the bamboo slices effectively. The recognition of bamboo slices made for different varieties and processing technology has strong robustness, which overcomes the shortcomings of traditional methods that need to adjust parameters frequently.

MATERIAL AND METHODS

Datasets

To make sure the result is convinced in the experiment, we select 6360 images after screening out from samples on the premise of ensuring that only the bamboo itself is different and other objective conditions are the same. The datasets are divided into four categories by its feature, type a contains 1700 pictures, type b contains 1700 pictures, type c contains 1700 pictures and type d contains 1200 pictures. We split 80% of the datasets for training and 20% for validation. Each photo is cropped to 224 * 224 pixels, which can be processed by transformer network easily. All the images in the datasets are uniquely numbered according to its defect category, which can be quickly and accurately identified by numbering. The datasets is shown in Fig. 2.



(a) (b) (c) (d)

Fig. 2: Datasets image: (a) Banpiancai, (b) Feipiancai, (c) Huapiancai, (d) Lantoucai.

Experimental environment

In this experiment, we use Pycharm and Pytorch1.8.0 for bamboo defect classification. The experiment adopts windows 10, the CPU is i7-9750H, the RAM is 16GB, the GPU is NVIDIA GeForce GTX 1660Ti, and the CUDA version is 10.2.141.

Research methods

Mish

In the encoder block of the transformer, the network uses Gelu as its activation function and obtains a satisfy result. The Gelu function is shown in Eq. 4:

$$Gelu(x) = x \times 0.5 \times (1 + \tanh[\sqrt{\frac{2}{\pi}}(x + 0.044715 \times x^3)]) \quad (4)$$

In this experiment, we try to replace the Gelu with Mish in the encoder block of the transformer. The Mish function is shown in Eq. 5:

$$Mish(x) = x \times \tanh(softplus(x)) = x \times \tanh(\ln(1 + e^x)) \quad (5)$$

where: x is the input of the two functions.

The difference between Mish and Gelu is shown in Fig. 3. As we all see, the function Mish is similar to Gelu, so we change the activation function from Gelu to Mish in the encoder block.

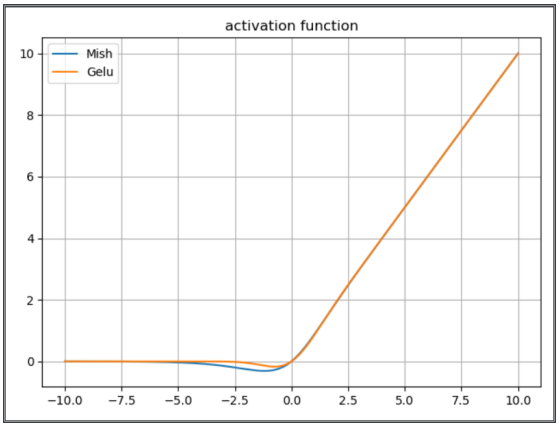


Fig. 3: The difference between Mish and Gelu.

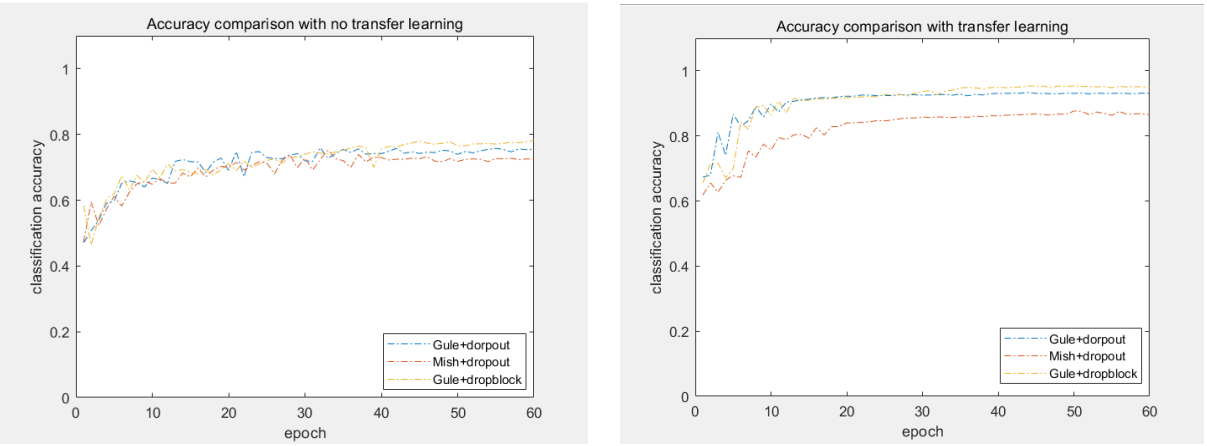
DropBlock

In the deep convolution neural network, overfitting is a common problem. To address this issue, a dropout strategy is presented, which inhibits the cooperation of specific features, so alleviating the overfitting problem. Similarly, if dropout is employed in the convolution layer during the detection phase, it has little impact. Because the convolution layer can learn comparable information near the dropped neurons, the DropBlock module appears to suppress information synergy by randomly inactivating a specific area in order to prevent overfitting in the convolution layer. Its technique throws out the whole adjacent region of the feature map, leaving the remaining units to learn features for the next stages. To validate the bamboo chips

classification performance, the regularization approaches are used in this letter. After all of the fully connected layers, DropBlock-based regularization was applied. There are two parameters of DropBlock, including block-size and γ. The block-size and γ decide to drop the area size and regularize the activation units, resp. In this experiment, we set the value of block-size is 7.

RESULTS AND DISCUSSION

The broken line graph of bamboo fault classification accuracy using transformer and improved transformer networks is shown in Fig. 4. The blue line indicates that the original Transformer model, the red line indicates that the original transformer change the activate function from Gelu to Mish, and the orange line indicates that the Transformer adopts the dropblock construction.



(a) *(b)*

Fig. 4: Classification accuracy of different networks: (a) Comparison of accuracy based on no transfer learning, (b) comparison of accuracy based on transfer learning.

Tab. 1: Classification accuracy of bamboo defects.

Transformer network	Transfer learning	
	no	yes
Gelu + dropout	75.43	93.06
Mish + dropout	72.54	86.63
Gelu + dropblock	77.49	95.39

The classification accuracy in different transformer network models are shown in Tab. 1. It can be seen that the highest accuracy of the three networks is 95.39% when used the transfer learning and DropBlock network. When the basic transformer network is used, the accuracy is 93.06%, compared with the accuracy of 75.43% without transfer learning, the accuracy is improved 17.63%, the improvement effect is obvious. Compared with the original transformer, the accuracy of Mish function under the same conditions is decreased 2.9% and 5.4%, respectively, this proves that our method of changing the activation function does not achieve

satisfactory results. When using the improved transformer network with DropBlock, compared with the transformer before the improvement, the accuracy has increased by 2% at least.

The experimental results show that, regardless of whether transfer learning is utilized or not, the transformer network with DropBlock has the maximum classification accuracy under the identical conditions. When the activation function of the network is changed from Gelu to Mish, the accuracy decreases. But when the improved transformer are used, the accuracy is significantly improved reaches to 95.39%.

Confusion matrix

A confusion matrix is a visual tool that may be used to assess the accuracy of classification. The number of network prediction categories is represented by the total number of columns in the confusion matrix. The overall number of rows reveals the category's real number of classified photographs, while each row indicates the image's accurate classification. The depth of the background color in the matrix image indicates the accuracy of category recognition. The more vibrant the color, the more accurate the model identification.

We split the datasets of training and validation is 8 : 2, so the number of validation of *(a)* banpiancai, *(b)* feipiancai, *(c)* huapiancai and *(d)* lantoucai is 340, 340, 340 and 240, respectively. Fig. 5 shows the confusion matrix of classification networks. In 1260 testing photos, 1202 photos are accurately identified, and the recognition accuracy is 95.39%. A small part of the image classification errors may be due to the similar texture between the bamboo defects.

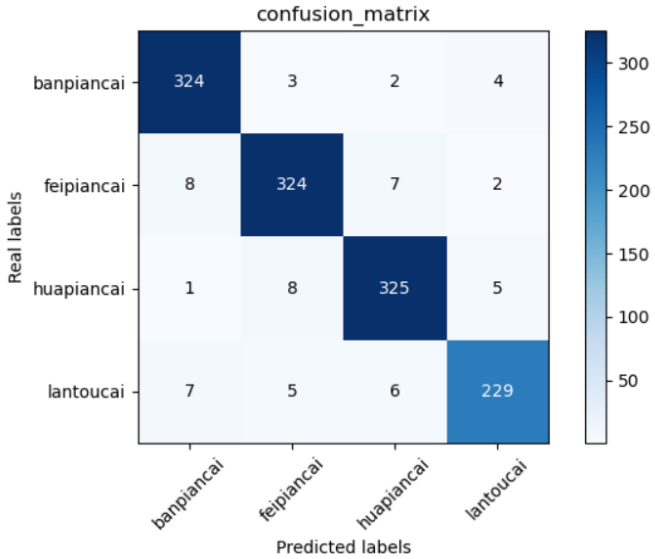


Fig. 5: Confusion matrix of classification networks: Gelu + dorpblock.

CONCLUSION

We present an improved transformer technique for datasets classification in order to achieve accurate bamboo defect classification. The experimental findings reveal that the classification neural networks' accuracy has increased greatly after utilizing the enhanced network compared to the prior ones. The most significant improvement is 3.33%. It has been

demonstrated that the improved method can improve classification accuracy. We can observe from the confusion matrix that there are still some flaws in this experiment. On the one hand, certain images in the datasets include faults that are similar in shape and texture; on the other hand, when there are two types of defects present at the same time, some images must be split into a specific category, which impacts classification accuracy. Due to the aforementioned issues, the author will carefully choose the datasets to ensure that each image contains just one type of flaw. Simultaneously, the author will continue to train and select a new deep learning verification algorithm.

ACKNOWLEDGEMENTS

This work is supported by the Fundamental Research Funds for the Central Universities (No. 2572019BF09), Heilongjiang Provincial Postdoctoral Science Foundation (No.178157).

REFERENCES

1. Mallik, A., Tarrío-Saavedra, J., Francisco-Fernández, M., Naya, S., 2011: Classification of wood micrographs by image segmentation. Chemometrics and Intelligent Laboratory Systems 107(2): 351-362.
2. Felzenszwalb, P., Girshick, R., McAllester, D., Ramanan, D., 2010: Object detection with discriminatively trained part based models. IEEE Transactions on Pattern Analysis and Machine Intelligence 32(9): 1627-1645.
3. Estevez, P.A., Perez, C.A., Goles, E., 2003: Genetic input selection to a neural classifier for defect classification of radiata pine boards. Forest Products Journal 53(7): 87-94.
4. Patgar, S.V., Sharath, K.Y.H., Vasudev, T., 2014: Detection of fabrication in photocopy document using texture features through K-means clustering. Signal Image Processing 5(4): 29-36.
5. Venkateswaran, K., Kasthuri, N., Balakrishnan, K., 2013: Performance analysis of K-means clustering for remotely sensed Images. International Journal of Computer Applications 84(12): 23-27.
6. Chen, Y., Lin, Z., Zhao, X., Wang, G., Gu, Y., 2014: Deep learning-based classification of hyperspectral data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7(6): 2094–2107.
7. Lin, L., Luo, P., Chen, X., Zeng, K., 2012: Representing and recognizing objects with massive local image patches. Pattern Recognit 45(1): 231–240.
8. Kwedlo, W.A., 2011: Clustering method combining differential evolution with the K-means algorithm. Pattern Recognition Letters 32(12): 1613-1621.
9. Kuo, R.J., Suryani, E., Yasid, A., 2013: Automatic clustering combining differential evolution algorithm and K means algorithm. Proceedings of the Institute of Industrial Engineers Asian Conference. Pp 1207-1215.

10. Yusof, R., Khalid, M., Khairuddin, A.S.M., 2013: Application of kernel-genetic algorithm as nonlinear feature selection in tropical wood species recognition system. Computers & Electronics in Agriculture 93(2): 68-77.
11. Sioma, 2015: Assessment of wood surface defects based on 3D image analysis. Wood Research 60(3): 339-350.
12. Gonzalo, A.R., Pablo, E., Pablo, A.R., 2009: Automated visual inspection system for wood defect classification using computational intelligence techniques. International Journal of Systems Science 40(2): 10.
13. Tan, X.Y., Triggs, B., 2010: Enhanced local texture feature sets for face recognition under difficult lighting conditions. IEEE Transactions on Image Processing 19(6): 1635-1650.
14. Cetiner, S., Var, A.A., Cetiner, H., 2014: Classification of KNOT defect types. IEEE Signal Processing & Communications Applications Conference. Pp 1086-1089.
15. Mu, H., Zhang, M., Qi, D., Guan, S., Ni, H., 2015: Wood defects recognition based on fuzzy bp neural network. International Journal of Smart Home 9(5): 143-152.
16. Hanbay, K., Alpaslan, N., Talu, M.F., Hanbay, D., 2016: Principal curvatures based rotation invariant algorithms for efficient texture classification. Neurocomputing 199: 77-89.
17. Honghai, H., Hoanghieu, T., 2021: Improvement for convolutional neural net-works in image classification using long skip connection. Applied Sciences 11: 2092.
18. Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012: ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems 25(2): 1-9.
19. Zhang, X., Liang, Y., Li, C., Huyan, N., Jiao, L., Zhou, H., 2017: Recursive autoencoders-based unsupervised feature learning for hyperspectral image classification. IEEE Geoscience and Remote Sensing Letters 14(11): 1928–1932.
20. Chen, Y., Zhao, X., Jia, X., 2015: Spectral–spatial classification of hyperspectral data based on deep belief network. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 8(6): 2381–2392.
21. Wei, L., Guodong, W., Fan, Z., Qian, D., 2017: Hyperspectral image classification using deep pixel-pair features. IEEE Transactions on Geoscience and Remote Sensing 55(2): 844–853.
22. Xin, H., Yushi, C., 2019: Optimized input for CNN-based hyperspectral image classification using spatial transformer network. IEEE Geoscience and Remote Sensing Letters 16(12): 1884-1888.
23. Golnaz, G., Tsungyi, L., QuocV, L., 2018: DropBlock: A regularization method for convolutional networks. Conference and Workshop on Neural Information Processing Systems. Cambridge, MA, USA. Pp 10727–10737.
24. Leyuan, F., Guangyun, L., Shutao, L., Pedram, G., Jón, A.B., 2019: Hyperspectral image classification with squeeze multibias network. IEEE Transactions on Geoscience and Remote Sensing 57(3): 1291–1301.
25. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015: Going deeper with convolutions. IEEE Computer society conference on computer vision and pattern recognition. Pp 1-9.

26. Kaiming, H., Xiangyu, Z., Shaoqing, R., Jian, S., 2016: Deep residual learning for image recognition. IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA. Pp 770-778.
27. Yuntao, L,. Yong, D., Ruochun, J., Peng, Q., 2018: Visual tree convolutional neural network in image classification. International Conference on Pattern Recognition, Beijing, China. Pp 758–763.
28. Shaoqing, R., Kaiming, H., Ross, G., Jian, S., 2016: Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence 39(6): 1137–1149.
29. Kaiming, H., Georgia, G., Piotr, D., Ross, G., 2017: Mask R-CNN. IEEE International Conference on Computer Vision, Venice, Italy. Pp 2961–2969.
30. Nagpal, C., Dubey, S.R., 2019: A performance evaluation of convolutional neural networks for face anti spoofing. International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary. Pp 1–8.
31. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, AN., Kaiser, L., Polosukhin, I., 2017: Attention is all you need. Conference and Workshop on Neural Information Processing Systems. Cambridge, MA, USA. Pp 5998–6008.
32. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N., 2021: An image is worth 16x16 words: Transformers for image recognition at scale. International Conference on Learning Representations. Pp 1-22.
33. Xiang, H., Wenjing, Y., Hao, W., Yu, L., Yuanxi, P., 2021: A lightweight 1-D convolution augmented transformer with metric learning for hyperspectral image classification. Sensors 21(5): 1751-1770.
34. Pan, S.J., Yang, Q., 2010: A Survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering 22(10): 537-570.

JUNFENG HU, XI YU*
MECHANICAL AND ELECTRICAL ENGINEERING
NORTHEAST FORESTRY UNIVERSITY
HARBIN, HEILONGJIANG 150040
CHINA

YAFENG ZHAO*
INFORMATION AND COMPUTER ENGINEERING
NORTHEAST FORESTRY UNIVERSITY
HARBIN, HEILONGJIANG 150040
CHINA
*Corresponding authors: YXyuxi419@163.com and nefuzyf@126.com