Research on the Lip-Print Recognition Based on Multi-scale Feature

Hongcheng Zhou(✉)

School of Electronic and Information Engineering, Jinling Institute of Technology, Nanjing, China
945516882@qq.com

Abstract. To address the problems of difficult feature extraction, small differences between texture information and recognition accuracy to be improved in lip-print recognition tasks, a lip-print recognition algorithm based on grouped multi-scale feature fusion is proposed. The ablation experimental results show that the improved recognition model achieves 98.56% recognition accuracy on the test set, and the model has strong generalization ability and feature refinement expression ability, which can provide support for the application of lip-print recognition technology in the field of identity verification.

Keywords: Lip-print recognition · Deep learning · Multi-scale feature fusion · Attention mechanism

1 Introduction

In recent years, more and more domestic and foreign scholars have paid attention to the field of lip pattern recognition and proposed different lip pattern recognition algorithms. Sandhya et al. [1] compared and analyzed lip pattern recognition methods based on machine learning algorithms, using classification algorithms such as support vector machine (SVM), K nearest neighbor (KNN), ensemble classifier, and artificial neural network (ANN) to classify the extracted features. Doroz et al. [2] proposed a lip pattern recognition algorithm that integrates complex image processing techniques, machine learning, and statistical methods. The proposed method can process partially damaged or incomplete lip patterns, identify and effectively eliminate low-quality areas in the 2 image. Chen Zongyang et al. [3] proposed a method for identifying coating surface defects based on the MobileNetV2 network. Wang Huanxin et al. [4] introduced an efficient channel attention and attention feature fusion module to establish an efficient crop leaf disease recognition model based on improved MobileNetV2. Li Zimao et al. [5] proposed a small sample recognition method for tea diseases. Zhang Dong et al. [6] proposed a counting method based on improved MobileNetV2, which adds volumes and attention modules to the original network to improve the refinement ability of fea-tures. However, its algorithm recognition accuracy still needs to be improved.

X. Chen et al. (Eds.): IoTaaS 2023, LNICST 585, pp. 16–21, 2025.
https://doi.org/10.1007/978-3-031-70507-6_2

2 Data and Preprocessing

2.1 Acquire Image

The collection of lip print images mainly includes contact and non-contact meth-ods. Therefore, combining the characteristics of deep learning algorithms and the ad-vantages of non-contact acquisition methods, a network camera is used to obtain lip print images. To obtain lip print images from different angles, video recording is used. The dataset captured 60 volunteers and ultimately obtained video recordings of each volunteer.

2.2 Data Enhancement

The final dataset consists of 60 people and 18000 images, including images from multiple angles and different noises. The dataset is divided into training, validation, and testing sets in a 7:2:1 ratio. The detailed information is shown in Table 1.

Table 1. Dataset Partitioning Information

Data Set	Training	Validation	Test Set/	Total/
	Set/Sheet	Set/Sheet	Sheet	Sheet
Original data set	2520	720	360	3600
Expanded dataset	12600	3600	1800	18000

3 Lip Pattern Recognition Model

3.1 MobileNetV2 Network

MobileNet is a lightweight convolutional neural network proposed by Google in 2017, which is committed to migrating the deep convolutional neural network model to mobile terminals and embedded devices. The deep separable convolutional structure is shown in Fig. 1. Adjust the data dimensions of input and output, and integrate fea-ture maps generated by different channels [7].

3.2 Improved MobileNetV2 Model

3.2.1 Embedded Attention Mechanism Module

The convolutional block attention module (CBAM) is a hybrid attention mecha-nism that includes channel attention submodules and spatial attention submodules, which are serially connected [8]. The calculation process of the channel attention sub-module is shown in Eq. 1.

$$M_c(F) = (MLP(AvgPool(F)) + MLP(MaxPool(F))) \quad (1)$$

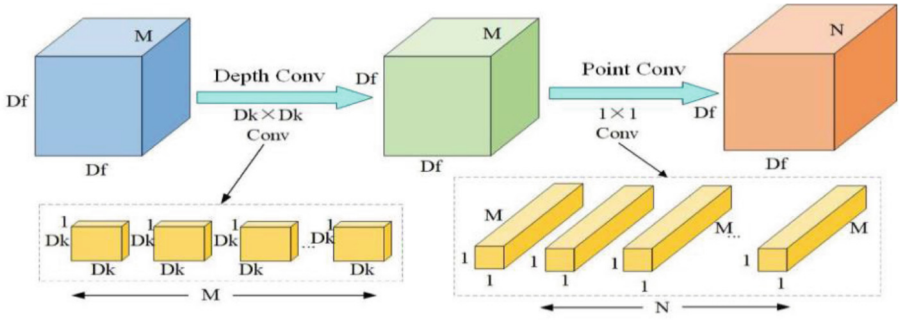

Fig. 1. Deep Separable Convolutional Structure

The calculation process of the spatial attention submodule is shown in Eq. 2. Firstly, this submodule performs an average pooling operation on the output of the channel attention submodule, and then concatenates the pooled feature maps on the channel dimension. This submodule can enable the model to better capture spatial in-formation and improve model performance.

$$M_s(F) = (f_{77}(AvgPool(F),\ MaxPool(F)\,)) \tag{2}$$

3.2.2 Design of Grouped Multiscale Feature Fusion Module

In order to enable the network to fuse multi-scale feature information extraction, it can effectively utilize low dimensional features and adapt to input images of different resolutions [10]. And then uses 3 3, 5 5, 7 7, 9 9 to perform convolution operations on the input feature maps of each group using four different sizes of convolutions to generate feature subgraphs f_1, f_2, f_3, f_4 corresponding to each channel.

$$F = Concat(f\,,\ f_1,\ f_3,\ f_4) \tag{3}$$

Finally, perform channel blending on the feature maps generated from the four sets of channels to increase the exchange and fusion of feature information between different channels, and then perform feature stitching on the four different feature sub-graphs as shown in Eq. 3.

4 Experiment

4.1 Model Evaluation Indicators

All experiments used accuracy, precision, and specificity as indicators to evaluate network performance. Accuracy is a simple and intuitive evaluation indicator in image classification problems, which is the ratio of the sample size correctly predicted by the model to the total sample size. Precision refers to the proportion of positive samples predicted by the model to the positive samples predicted. Specificity refers to the ratio of the predicted negative sample size to the actual negative sample size. The specific calculation process is shown in the following equation.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$Specificity = \frac{TN}{TN + FP} \quad (6)$$

4.2 Experimental Results and Analysis

4.2.1 Experiments Introducing Different Attention Mechanisms

In recent years, the attention modules designed by convolutional neural network include squeeze and excitation (SE), efficient channel attention (ECA), bottleneck at-tention module (BAM), convolutional block attention module (CBAM) and coordinate attention (CA). The results are shown in Table 2.

Table 2. The recognition effect of MobileNetV2 combined with different attention mechanisms

Network	Parameter Quantity /M	Accuracy/%	Average Accuracy /%	Average Specificity/%
mbV2	2.30	96.91	96.97	99.95
mbV2_SE	2.86	95.72	95.78	99.90
mbV2_ECA	2.38	95.5	96.52	99.91
mbV2_BAM	3.67	95.11	95.79	99.91
mbV2_CA	2.75	96.56	96.72	99.93
mbV2_CBAM	3.14	97.39	97.61	99.95

From Table 2, it can be seen that the recognition accuracy of the network has changed to some extent after embedding the attention mechanism. Adding an attention module with convolutional layers will inevitably increase the number of network parameters.

4.2.2 Experiments to Reduce the Number of Network Parameters

Introducing the convolutional attention mechanism module will increase the num-ber of network parameters. The experimental results were analyzed and the value of parameter a in the improved net-work was determined. The experimental results are shown in Table 3.

Compared with the model without introducing parameter a, the recognition accuracy has improved by 0.44%. Although fewer parameters were obtained in other experimental

Table 3. Recognition Effects of Different Parameter Quantity Models

Network	Parameter Quantity/M	Model Size/MB	Accuracy /%	Average Accuracy/%
MobileNetV2	2.30	8.63	96.91	96.97
MobileNetV2_cbam	3.14	11.94	97.39	97.61
MobileNetV2_1234	2.12	8.12	96.36	96.72
MobileNetV2_4567	3.0	11.76	96.72	96.86
MobileNetV2_1357	2.42	9.21	97.83	98.17
MobileNetV2_1246	2.62	10.10	96.42	96.87

groups, priority must be given to the recognition accuracy of the network model. Through the analysis of the ablation experiment results, it was ultimately determined that the bottleneck modules 1, 3, 5, 7 in the improved network were constructed by introducing the true value of parameter a.

5 Summary

We conducted relevant model training and ablation experiments on the lip print dataset, and the experimental results showed that the improved network has better gen-eralization ability and higher recognition rate. The recognition accuracy and average accuracy have been improved by 1.65% and 1.73% respectively on the test set. From the attention heatmap of the network model before and after improvement, it can be concluded that the improved recognition model has better feature refinement and ex-traction capabilities.

Acknowledgment. This work was supported by Cooperative Project of Jiangsu Province production, teaching and research (No. BY2021381), Jin Ling Institute of Technology Ph.D. Startup Fund (jit-b-202314).

References

1. Sandhya, S., Fernandes, R., Sapna, S., et al.: Comparative analysis of machine learning algorithms for Lip print based person identification. Evolut. Intell. **15**, 743–757 (2021)
2. Doroz, R., Wrobel, K., Orczyk, T., et al.: Multidimensional nearest neighbors classification based system for incomplete lip print identification. Expert Syst. Appl. **202**, 117–137 (2022)
3. Zongyang, C.H.E.N., Hui, Z.H.A.O., Yongsheng, L.Y.U., et al.: A recognition method of coating sur-face defects based on the improved MobileNetV2 network. J. Harbin Eng. Univ. **43**(4), 572–579 (2022)
4. Huanxin, W.A.N.G., Zhihao, S.H.E.N., Quan, L.I.U., et al.: Identification of crop leaf diseases based on improved mobileNetV2 model. J. Henan Agric. Sci. **52**(04), 143–151 (2023)
5. Li Zimao, X., Jie, Z.L., et al.: Small sample recognition method of tea disease based on improved DenseNet. Trans. Chin. Soc. Agric. Eng. **38**(10), 182–190 (2022)

6. Zhang, D., Jiang, Y.: Drill pipe counting method based on improved Mo-bileNetV2. J. Mine Autom. **48**(10), 69–75 (2022)
7. Jinhui, L., Di, W., Xiaopan, S.: Research progress on visual image detection based on convolutional neural network. Chinese J. Sci. Instr. **41**(4), 167–182 (2020)
8. Lin, T.Y., Goyal, P., Girshick, R., et al.: Focal loss for dense object detection. IEEE Trans. Patt. Anal. Mach. Intell. **42**(2), 318–327 (2020)
9. Ma, L., Zheng, S.Y., Niu, B.: Action recognition method on regional association adaptive graph convolution. J. Front. Comput. Sci. Technol. **16**(4), 898–908 (2022)
10. Meng, L., Guo, X.Y., Du, J.J., et al.: A lightweight CNN model for image recognition of crop diseases. Jiangsu J. Agricult. Sci. **37**(5), 1143–1150 (2021)
11. Xu, Y.J., Li, C.: Light weight object detection network optimized based on YOLO family computer. Science **48**(11), 265–269 (2021)
12. Selvaraju, R.R., Cogswell, M., Das, A., et al.: Grad-CAM: visual explanations from deep net-works via gradient-based localization. Int. J. Comput. Vis. **128**(2), 336–359 (2020)
13. Lin, T.Y., Goyal, P., Girshick, R., et al.: Focal loss for dense object detection. IEEE Trans. Pattern Anal. Mach. Intell. **42**(2), 318–327 (2020)