Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 10, No. 2, 2023

Minority Clothing Recognition based on Improved DenseNet

Binbin Hu, Fei Tang*, Shengbo Tan

Southwest Minzu University, Chengdu 610225, China

* **Corresponding author**: Fei Tang (Email: sugaarvapeur@gmail.com)

Abstract: In response to the difficulty of minority clothing recognition in China, this paper proposes an SE-DenseNet-LR recognition model. This model replaces the ReLU activation function in each dense connection block with the Leaky ReLU function, which has the advantages of faster convergence speed, better generalization ability, and mitigating gradient disappearance in model training. The SE (Squeeze-and-Excitation) attention mechanism is added after each convolutional layer in each dense connection block to obtain more important feature information. The convolutional layers in the connection layers are replaced with dilated convolutions to increase the receptive field. The model also employs learning rate decay and image dataset augmentation strategies to prevent overfitting. Experimental results show that the SE-DenseNet-LR model achieves an accuracy of 84.35% in recognizing 20 categories of minority clothing, which is 2.35%, 2.66%, and 1.88% higher than the recognition accuracies of ResNet18, ResNet34, and DenseNet models, respectively. This model has strong feature extraction ability and robustness, which lays a good foundation for minority clothing recognition.

Keywords: Minority clothing; Image recognition; DenseNet; Attention mechanism.

1. Introduction

With the continuous development of society, people's attention to multiculturalism has gradually increased, especially in the field of clothing. China is a multi-ethnic country, and each ethnic group has its own unique clothing culture. The fusion and inheritance of this multiculturalism is of great significance to the maintenance and promotion of the world's intangible cultural heritage. However, due to the uniqueness and complexity of different ethnic clothing, its recognition has always been a highly challenging problem.

Each type of ethnic clothing has its own distinctive features. In traditional clothing recognition, it can only rely on manual extraction of features such as color and texture for each type of clothing, which is tedious and complex. Furthermore, the diversity of ethnic clothing makes recognition difficult and inefficient. Methods based on ethnic clothing processing have gradually received widespread attention. In recent years, with the rapid development of deep learning technology, image classification technology has been widely applied in various fields such as medical image classification[1-2], vehicle classification[3], and facial expression recognition [4]. Convolutional neural networks have made breakthrough progress in the field of image classification. Network frameworks such as AlexNet[5], ResNet[6], and VggNet [7] have been used in various types of image classification. Yang Bing[8] and others used DenseNet-BC as the basic network structure, designed and used multi-scale dense connection units, and proposed a local and global attention mechanism to classify and recognize 9 types of ethnic clothing, achieving good recognition accuracy. Hou Hongtao [9] and others used gamma transformation algorithm and Retinex algorithm to enhance the contrast of ethnic clothing images, reduce noise, and achieved an average classification accuracy of 87.81% on 5 types of ethnic clothing. Zhang Yue [10] and others embedded Squeeze-and-Excitation (SE) modules in ResNet, and used feature compression and excitation (Squeeze-and-Excitation) to enable shallow layers to also obtain a global receptive field, improving the recognition rate of ethnic clothing patterns. Zeng Fuliang [11] and others used transfer learning method to train the neural network parameters and weights trained on the large dataset ImageNet on a small dataset, and then changed the fully connected layer of the original neural network to the layer required by the small dataset to classify thangka images. He Qiuge [12] and others improved DeepLabv3+ and used methods such as image stitching, label smoothing, loss function, and cross-entropy loss function to process clothing images and enhance image feature information.

The above methods for identifying minority ethnic costumes mostly consider the characteristics of the costumes themselves, ignoring the impact of lighting and background in the original natural environment. Currently, all research on minority ethnic costumes has only focused on a few categories due to difficulties in collecting datasets and biases in the collected information. In this paper, we address the problem of recognizing minority ethnic costumes by classifying and identifying 20 different categories of such costumes for the first time. Based on the DenseNet [13] network, we propose an SE-DenseNet-LR model, which achieves good results in the recognition of 20 different categories of minority ethnic costumes.

2. Data Processing

2.1. Ethnic Clothing Dataset

This article constructed an ethnic clothing dataset through online search engines, photo-taking, and other methods. The images of ethnic clothing were manually screened to remove

low clarity, low resolution, damaged images, resulting in a total of 20 categories of ethnic clothing images. Some ethnic clothing is shown in Figure 1. Since the images come from multiple sources, there are factors such as inconsistency between the actual ethnic clothing and the target, non-uniform image formats, image damage, and low resolution. After manual screening, a total of 3415 ethnic clothing images in

JPG format were obtained, and Table 1 shows the details of the entire dataset.

(a) Achang (b) Bai (c) Baoan

(d) Buyi (e) Hmong (f) Dai

(g) Han (h) Tu (i) Korean

Figure 1. Some ethnic minority costumes

2.2. Data Augmentation

Due to the limited number of samples for each class, deep neural network training requires a large amount of data, otherwise it may result in overfitting or low generalization ability. In this experiment, to reduce the negative impact of the small dataset, data augmentation strategies were applied to the dataset, including horizontal flipping, vertical flipping, random cropping, and translation. After data augmentation, the dataset was expanded four times, with a total of 13,660 images.

3. Model Construction

Table 1. Details of ethnic minority costumes

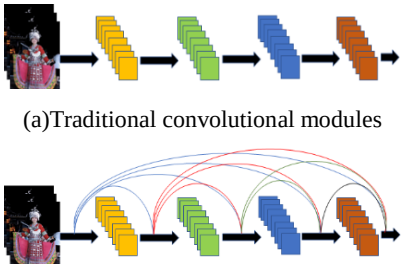
Category	Quantity
Achang	197
Bai	189
Baoan	154
Bulang	160
Buyi	203
Korean	148
Dai	147
Dawoer	142
Deang	156
Dong	149
Dulong	147
Russian	156
Gaoshan	163
Han	166
Hmong	182
Nu	173
Tajiks	192
Tu	213
Yi	163
Yugurs	215
Total	3415

As a novel network structure, DenseNet [13] has become one of the hotspots pursued by researchers due to its advantages in feature reuse and parameter sharing. Based on this, this paper has made a series of improvements: (1) replacing the ReLU activation function in each dense connection block with the Leaky ReLU function, which has the advantages of fast convergence speed, better generalization ability, and alleviating gradient disappearance during model training; (2) adding the SE (Squeeze-and-Excitation) attention mechanism after each convolutional layer in the dense connection block to obtain more important feature information; (3) replacing the convolutional layer in the connection layer with the dilated convolution to increase the receptive field. Each part will be specifically demonstrated below.

3.1. Traditional Networks and DenseNet

Compared to traditional network structures, DenseNet introduces the concept of dense connections, which can improve network performance without increasing parameters and computation. In traditional network structures, there may be information bottlenecks during information transmission because each neuron can only receive information from the previous layer's feature map. In DenseNet, each layer receives feature maps from all previous layers, so the information from the previous layers can be better utilized. Additionally, dense connections in DenseNet can help prevent gradient vanishing, improve network training stability, and generalization ability.

The structure of the traditional convolutional module and the dense connection module are shown in Figure 2a and 2b, respectively.



(a)Traditional convolutional modules

(b)Densely connected modules

Figure 2. Structure diagram of DenseBlock

Traditional networks such as ResNet use the solid and dotted residual connections to add the input and output of each group of convolutional structures, and then input them into the next layer of the network. The mapping formula is:

$$x_l = H_l(x_{l-1}) + x_{l-1}. \quad (1)$$

The output of the l-th layer in DenseNet, denoted as x_l, is obtained by concatenating the outputs of all previous layers, denoted as x_{l-1}, and passing through a nonlinear activation function H_l. The dense connectivity mechanism in DenseNet allows all layers to be mutually connected and concatenates the outputs of all previous layers as the input to the next layer. The mapping formula can be expressed as:

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]). \quad (2)$$

x_l represents the output of the l-th layer, $[x_0, x_1, \dots x_{l-1}]$ represents the concatenation of the output of the $(l-1)$-th layer and the outputs of all previous layers, and H_l is a non-linear transformation function.

The DenseNet network consists of 4 DenseBlocks and 3 Transition Layers, and is divided into different versions based on the number of convolutional and pooling layers, such as DenseNet121, DenseNet169, DenseNet201, etc. In this paper,

the smallest DenseNet121 with the fewest number of layers and parameters is used, as shown in Figure 3. The Transition Layer T_i includes a convolutional layer and a pooling layer, used to connect two Dense Blocks. The convolutional layer is used to reduce the number of features, while the pooling layer compresses the model. Each Dense Block consists of 4 Dense Layers, whose structure is shown in Figure 4. The input is first normalized by BN, then passed through a ReLU activation function, followed by a 1x1 convolution to reduce computational cost. It is then passed through another BN and ReLU, and finally outputted through a 3x3 convolution.

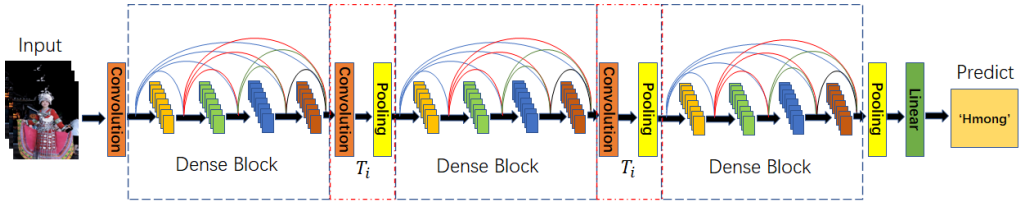


Figure 3. Structure diagram of DenseNet

3.2. Introduction of Attention Mechanism

Attention mechanisms are becoming increasingly popular in the field of computer vision. The main idea behind attention mechanisms is to design a weight distribution for the original feature map and apply it to the feature map. Different features will have different weights, with features with higher weights

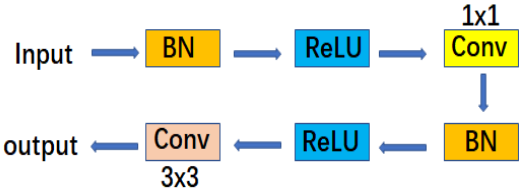


Figure 4. Structure of Dense Layer

subsequent tasks. The SENet[14] used in this paper is designed to assign weights to different positions of the image from the perspective of the channel domain, in order to obtain more important feature information through a weight matrix. The structure diagram of SENet is shown in Figure 5.

$$Z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{i=1}^{W} u_c(i,j) \quad (3)$$

where u_c is the feature map of the C-th channel in the input feature; W and H are the width and height, respectively; Z_c is the value of the feature map at the coordinate position. When SENet is embedded into a certain layer, the received feature map is first subjected to Fsq(·) operation, where Fsq(·) is the global average pooling that reduces the feature map size to 1x1 while keeping C unchanged. Then, Fex(, W) operation is performed, which includes fully connected layers, activation functions, fully connected layers, and sigmoid functions that normalize the specific values of C. Finally, Fscale(·) is applied to perform matrix multiplication between C and the original feature map, and the size remains unchanged after the SENet operation. In this paper, SENet is embedded into each layer after the Dense Layer's convolution operation, and the modified Dense Layer is shown in Figure 6. After embedding SENet, the network combines more attention features, focuses more on important information during feature extraction, ignores irrelevant information such as the background, removes interference, and improves the model's recognition rate.

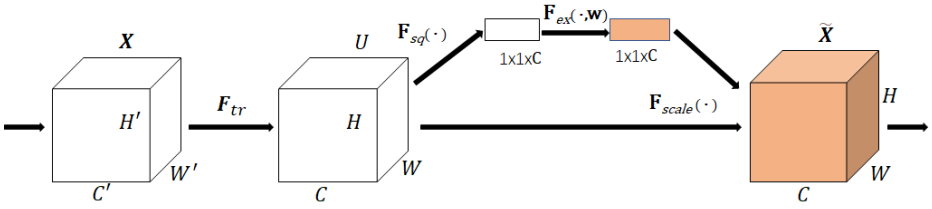


Figure 5. A Squeeze-and-Excitation block

3.3. Replacing Leaky ReLU

Replace ReLU activation function with Leaky ReLU in the Dense Layer. The output of Leaky ReLU has negative features of the image, and the derivative is always non-zero. Therefore, the Leaky ReLU function solves the problem of neuron death when ReLU function enters the negative interval. Its mathematical expression is:

$$f(x) = \begin{cases} x, x \geq 0 \\ \frac{x}{a}, x < 0 \end{cases} \quad (4)$$

When x<0, Leaky ReLU introduces a slope (such as 0.01) to ensure that the gradient is not zero for negative inputs, thereby alleviating the problem of gradient disappearance. ReLU function has no gradient and the neuron "dies" when the input is negative. Therefore, replacing it with Leaky ReLU function can lead to faster convergence speed and better network generalization ability. The modified structure is shown in Figure 7.

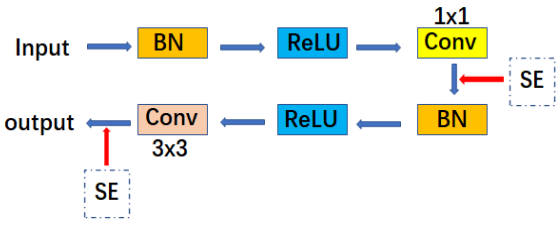


Figure 6. Dense Layer with added SENet

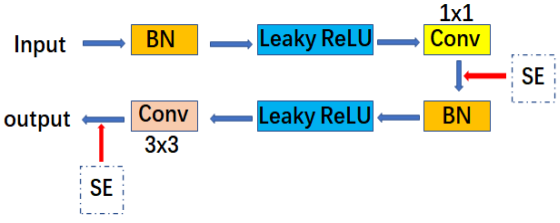
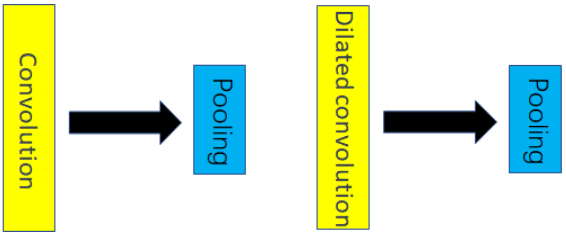


Figure 7. Replace ReLU with Leaky ReLU

3.4. Replacement with Dilated Convolution

The replacement of the convolutional layer with dilated convolutions is achieved by inserting zero values between the convolution kernels, which can increase the effective receptive field of each convolution kernel and enhance the receptive field of the convolutional layer. Dilated convolutions can reduce computational complexity and improve model runtime speed while maintaining the same receptive field. In traditional convolutional networks, pooling layers are often added after convolutional layers to reduce the size of feature maps during downsampling. However, dilated convolutions can directly reduce the size of feature maps by increasing the stride of the convolutional kernel, thereby avoiding the potential loss of information caused by pooling operations. In this paper, all the convolutions in the Transition Layer are replaced with dilated convolutions. The revised structure of the Transition Layer is shown in Figure 8.



(a)Transition (b)Improved Transition

Figure 8. Comparison of Improved Transition Layer

3.5. SE-DenseNet-LR Model

The SE-DenseNet-LR model is based on DenseNet and adds the SE (Squeeze-and-Excitation) attention mechanism after each convolutional layer in the DenseBlock. The ReLU function in the DenseBlock is replaced with the Leaky ReLU function, and all convolutions in the Transition Layer are replaced with the dilated convolutions.

4. Experimental Process

4.1. Experimental environment

The experimental environment is shown in Table 2.

Table 2. Experimental environment

System	Ubuntu 18.04
Language	Python
Framework	Pytorch
GPU	RTX 2080 Ti x2
Memory	12GB x2

4.2. Parameter Setting and Evaluation Metrics

During model training, the Adam optimization algorithm was used with a batch size of 64, 100 epochs, a train-test ratio of 8:2, and a learning rate of 0.01. The cross-entropy loss function was used to optimize the model, which is expressed as:

$$Loss = \sum_{i=1}^{n} y_i \log(\widehat{y_i}). \quad (5)$$

Accuracy is used as the evaluation metric for the model, where a higher accuracy indicates a closer prediction to the ground truth. TP represents the number of true positives, TN represents the number of true negatives, FP represents the number of false positives, and FN represents the number of false negatives in the prediction.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}. \quad (6)$$

4.3. Experimental Results

To verify the performance of the SE-DenseNet-LR model, it was compared with the ResNet18, ResNet34, and DenseNet121 models on the minority clothing dataset. The accuracy changes within 100 epochs were analyzed, as shown in Figure 9. The SE-DenseNet-LR model is significantly superior to ResNet18 and ResNet34 in terms of both accuracy and convergence speed. Compared with DenseNet, the convergence speed of the two models is roughly the same, but the accuracy of the SE-DenseNet-LR model is higher than that of DenseNet, and the curve is smoother.

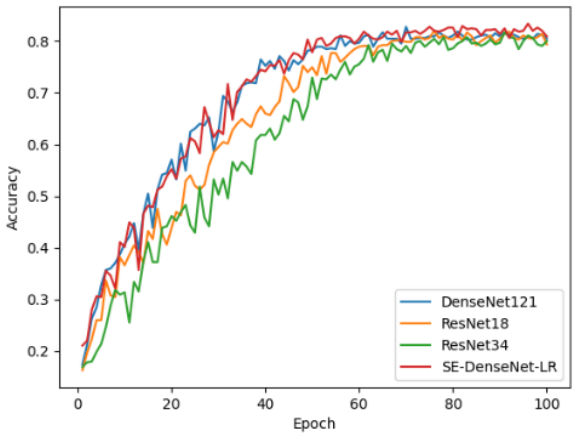


Figure 9. Multi-model comparison

To verify the effectiveness of the SENet, ablation experiments were conducted to compare with models that added ECA attention mechanism and CBAM attention mechanism. Table 3 shows the comparative results: the accuracy of the original DenseNet was 82.47%, and the proving the effectiveness of LeakyReLU. The accuracy of the model with CBAM attention mechanism was lower than that of DenseNet-LR and DenseNet, while the accuracy of the model with ECA attention mechanism was lower than that of DenseNet-LR. The accuracy of the model with SE attention mechanism reached 84.35%, which is higher than that of all models. This experiment proves the effectiveness of SE-DenseNet-LR.

Table 3. Multi-module comparison

Module	Accuracy
ECA-DenseNet-LR	82.61%
CBAM-DenseNet-LR	82.32%
SE-DenseNet-LR	84.35%
DenseNet-LR	83.04%
DenseNet	82.47%

To verify the effectiveness of the dilated convolution, the results of dilated convolution and normal convolution were compared, as shown in Table 4: the accuracy after replacing with dilated convolution is higher than that of normal convolution.

Table 4. Dilated Convolution

Dilated Convolution	Accuracy
Yes	84.35%
No	83.23%

5. Conclusion

A SE-DenseNet-LR recognition model is proposed to

address the difficulty of recognizing ethnic minority costumes in China. By improving the activation function in the Dense Layer and integrating attention mechanisms after the convolutional layer, the convolution in the Transition Layer is replaced by dilated convolution. The model has strong feature extraction capability and robustness, and achieves an accuracy of 84.35% on a self-made dataset of 20 ethnic minority costumes, showing stronger performance compared to other classification algorithms.

References

[1] Zhu Hengde,Wang Jian,Wang ShuiHua, et al. An Evolutionary Attention-Based Network for Medical Image Classification[J]. International Journal of Neural Systems, 2023, 33(3).

[2] Zhang Yuhan,Luo Luyang,Dou Qi, et al. Triplet attention and dual-pool contrastive learning for clinic-driven multi-label medical image classification[J]. Medical Image Analysis, 2023, 86.

[3] Kolukisa Burak,Yildirim Veli Can,Elmas Bahadir, et al. Deep learning approaches for vehicle type classification with 3-D magnetic sensor[J]. Computer Networks, 2022, 217.

[4] Liu Yuanyuan,Peng Jiyao,Dai Wei, et al. Joint spatial and scale attention network for multi-view facial expression recognition[J]. Pattern Recognition, 2023, 139.

[5] Alex Krizhevsky,Ilya Sutskever,Geoffrey E. Hinton. ImageNet classification with deep convolutional neural networks[J]. Communications of the Acm, 2017, 60(6).

[6] Kaiming He,Xiangyu Zhang,Shaoqing Ren, et al. Deep Residual Learning for Image Recognition.[J]. Corr, 2015, abs/1512.03385.

[7] Karen Simonyan,Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition.[J]. Corr, 2014, abs/1409.1556.

[8] Yang B, Xu D, Zhang H, et al. Recognition of ethnic minority clothing based on improved DenseNet-BC[J]. Journal of Zhejiang University (Science Edition), 2021, 48(6): 676-683.

[9] Hou Hongtao, Wang Wei, Shen Hongting, et al. Research on Image Classification of Ethnic Minority Clothing Based on Adaptive Image Enhancement and CNN. Modern Computer, 2022, 28(24): 29-35.

[10] Zhang Yue, He Xiyue, Zhao Chenglong. Design and Implementation of Ethnic Clothing Recognition System Based on SE-ResNet. Electronic Technology and Software Engineering, 2022, (8): 205-208.

[11] Zeng Fuliang, Hu Wenjin, He Guoyuan, et al. Tangka Image Classification Based on DenseNet. Modern Electronics Technique, 2022, 45(6): 153-157.

[12] He Q, Guan M, Gan L. Improved clothing image segmentation algorithm based on DeepLabv3+. Fujian Computer, 2023, 39(2): 21-26.

[13] Gao Huang,Zhuang Liu 0003,Kilian Q. Weinberger. Densely Connected Convolutional Networks.[J]. Corr, 2016, abs/1608.06993.

[14] Hu Jie,Shen Li,Albanie Samuel, et al. Squeeze-and-Excitation Networks[J]. Ieee Transactions on Pattern Analysis and Machine Intelligence, 2019, 42(8).