

An Improved Method of Clothing Image Classification Based on CNN

ZHANG Xuan-yu

College of Computer and Communication, Hunan Institute of Engineering, China
Email: 892565742@qq.com

DENG Zuo-jie

College of Computer and Communication, Hunan Institute of Engineering, China
Email: 1050400089@qq.com

ABSTRACT

In recent years, with the increasing demand for high-precision clothing image classification, people have attached great importance on the clothing image classification method based on CNN. In this paper, we discussed how to improve the accuracy of image classification by preprocessing the data set, combining the data set with CNN structure and improving the loss function. The comparative experiments between different structures of CNN and loss functions have been done. The experimental results show that the method we used can better improve the effect of classification and classification accuracy.

Keywords - CNN, Image classification, Image feature, Loss function

Date of Submission: May 02, 2021

Date of Acceptance: May 18, 2021

I. INTRODUCTION

Based on the artificial intelligence(AI), machine learning, deep learning and other technologies, the clothing image classification technology has gained great improvement, and it has evolved from traditional image classification technology to image classification technology based on deep learning. The traditional method of image feature extraction is to extract according to features of the images. Extracting features of the image with feature extraction factors including Harris [1], SIFT [2], SURF [3] and ORB [4] has been commonly used. Many scholars have done a lot of researches on the image feature extraction, Chen [5] used SURF to extract clothing features, and he applied these features and the machine learning method to clothing classification. Zhang Yong [6] improved the SIFT image feature matching algorithm to improve the efficiency of feature matching.

In 1998, Yann LeCun [7] proposed LeNet which used CNN to recognize images and extract image feature. Even though the CNN has gained rapid development nowadays, most CNN still improve the network accuracy by deepening the network level, the problems of which are over fitting and network degradation [8]. Pan [9] proposed using BP neural network to identify knitted fabrics. Razavai [10] built a deeper CNN by using the structure of CNN to extract image features. Although these methods can successfully process the extraction of image features, when there appears some unknown information in the image samples, it is difficult for network to classify the images, and this kind of problems cannot be effectively solved only by deepening the network level. In order to solve the problems mentioned above, we use the Inception CNN combined with the SENet module, preprocess of the data set, reduce the interference of the unnecessary information from the image and increase the recognition ability of the image by improving the loss function, so as to improve the

image classification accuracy. Our experiments show that the accuracy of this classification method has been significantly improved.

II. RELATED WORK

2.1 INCEPTIONNET

Inception [11] was proposed in 2014. After being improved and updated by researchers, it has evolved from V1(the 1st version), V4(the 4th version) [12] to Inception-ResNet(the latest version). Inception net was the champion of ImageNet competition, and the error rate of Inception net was only 6.67%. The Inception combined in InceptionNet allows the InceptionNet to use convolution cores with different sizes in the same layers of network, which improves the model perception. Meanwhile, the Inception with the batch standardization to alleviate the vanishing gradient. Considering that the V4 has the excellent performance, we also choose it as the network structure.

2.2 SENET

In 2018, JieHu and Li Shen [13] proposed the SENet network structure, the full name of which is Squeeze-and-Excitation Networks. It can also be simply understood as compression and incentive network, which is mainly composed of Squeeze and Excitation.

The compression part changes the original dimension of the feature map, specifically, compressing the $H*W*C$ to $1*1*C$ and the original height and width to one-dimensional height and width, which makes the vision of the receptive area wider. Squeeze will also realize the channel description by Mean-Pooling generating channel information.

After accessing the $1*1*C$ from the compression part, the excitation part uses the fully connected layers to predict each channel C and the feature map is mapped to the

channel. The biggest feature of SENet is that it can easily join other networks to improve the network effect.

III. FEATURE EXTRACTION OF CLOTHING IMAGE

3.1 DATA SET PROCESSING

When the network classifies a data set of clothing image, the images with errors often contain some interference factors except clothing, such as image background, mannequin and so on, while the classification error rate of the images only showing clothing is very low. Therefore, we designed a simple binary classification network to extract the clothing image with interference information, and use the upgraded method to improve the classification accuracy of this kind of image, so as to improve the overall classification accuracy. As shown in Figure 1, we perform a simple binary operation.



Fig 1. Illustration of image data set classification

We get two different types of data sets, one of which is clothing image, the other is clothing image with interference information. After that, we extract the feature of the clothing image with interference information.

3.2 FEATURE EXTRACTION OF CLOTHING IMAGE

BASED ON CNN

When we extract the features of clothing image by using CNN, we mainly simulate human recognition of clothing image by performing repeated convolution and pooling, basing on the effects of convolution and pooling layers in the network structure. The convolution layers is for increasing the original signal of the image to extract the feature of the image; while the pooling layers is for removing some miscellaneous information with high frequency in the image, and these two layers are combined with comprehensive features. Eventually, the clothing image features are output to build the feature library.

In fact, the steps to exact the image feature dominated by CNN for are formulaic, and the slight difference between different CNNs is the size of convolution layers in the network structure. Normally, we upload the clothing image data set to the model, and extract the pixel vector in the image which is used as the basic feature vector data. feature vector data is convoluted by convolution layers to get the image with convolution feature. Here we choose a clothing image with human model as an example. As shown in Figure 2:.

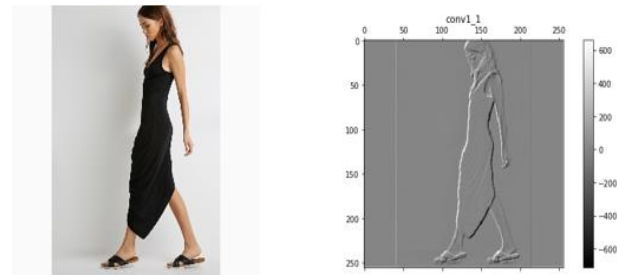


Fig 2. Left picture is the original image; Right picture is the convolution feature image

The linear changes can be adjusted when the model does not have enough linear expression, or we can add the activation function relu with nonlinear factors to solve the problems that the linear model can not solve. As shown in Figure 3:

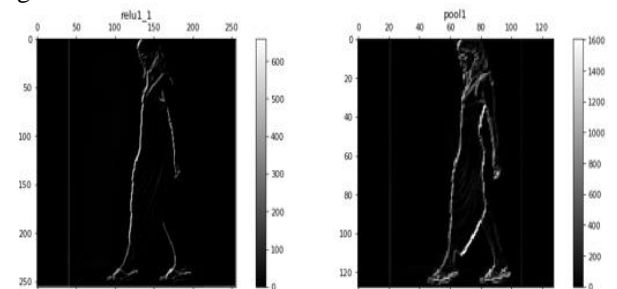


Fig 3. Feature image of Relu Fig 4. Feature image of MaxPooling

After that, we take sampling by MaxPooling for convolution feature image to reduce the amount of feature data. The pooled feature image is shown in Figure 4. After that, we repeatedly use convolution layers, activation function and pooling layers to extract features from the image, and the effect of which is shown in Figure 5.

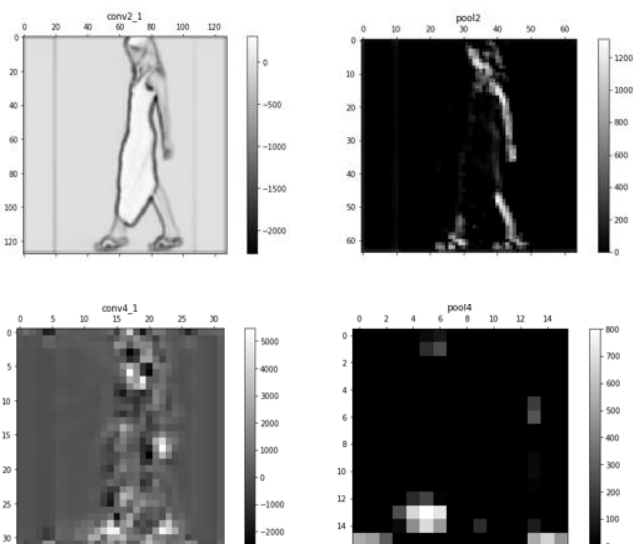


Fig 5. Feature image of multiple use of convolution layers, activation function and pooling layers

Finally, the image features are put together by the fully connected layers to output the image features.

Generally, after the images input to the model are preprocessed, the images are through the convolution layers, pooling layers, fully connected layers, activation function and loss function. Eventually, the feature vectors are output to be as the classification basis. In convolutional neural network, we can extract more image features by increasing the number of layers. However, with the increase of the number of layers, the number of parameters will increase exponentially, so increasing the number of layers is not a good choice.

3.3 LOSS FUNCTION

A loss function is a function that maps an event or values of one or more variables onto a real number intuitively representing some "cost" associated with the event. The loss functions commonly used in classification are 0-1 loss, cross-entropy loss function and softmax loss etc. When defining the loss function, we add the parameters of unknown information to improve the loss function. The weights are as follows:

$$W_i = \begin{cases} 1-x*b, l_i = y \\ x, l_i = a \\ 0, l_i = b \end{cases}$$

The purpose of how the parameters are set is to decrease the classification error rate caused by some unknown information. Adding the parameters of unknown information in the loss function can effectively solve this kind of problems, and to improve the accuracy of classification. The specific implementation process is shown in the figure 6.

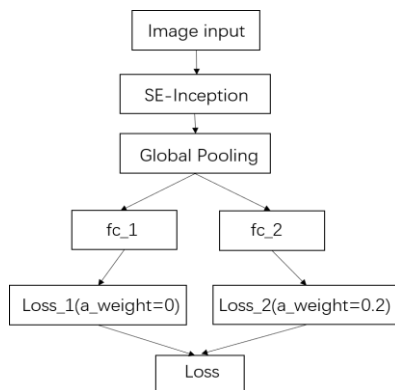


Fig 6. Network flow added with loss

IV. EXPERIMENT AND ANALYSIS

In order to verify the effectiveness of the algorithm, we test it by Anaconda3. Windows 10 is used as the operating system; Python 3.6 is used as the programming language and Tensorflow is used as the deep learning framework. Part of data from Deepfashion are used as the experimental data in this paper. The data set includes different categories of men's and women's clothing on the market. Deepfashion is a large-scale data set supported by the Chinese University of Hong Kong, which contains 800,000 pictures. The image size is 256 * 256. We select 24,605 pictures as training set and 4,921 pictures as test set. We use train_acc and test_acc as classification evaluation indexes.

We carry out the comparative experiments on three different network structures by using the improved loss. As shown in Figure 7, when the network training begins, the training loss and test loss are relatively high due to the insufficient initial fitting. However, with the increase of epoch, the training loss increases, the loss values of Inception and SENet fluctuate, while the loss values of SE-Inception tend to be stable and the parameters basically converge. Therefore, we get a conclusion that the training effect of SE-Inception is obviously better than that of SENet and Inception.

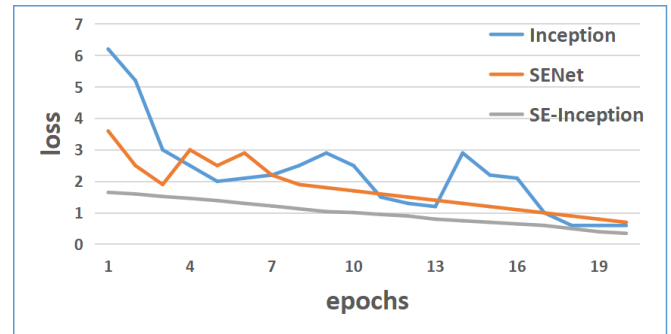


Fig 7. Different network training results

Based on SE-inception, we compare train_acc and test_acc by processing different loss functions with different epochs. It can be seen from the figure that our method has significantly improved in training accuracy and test accuracy.

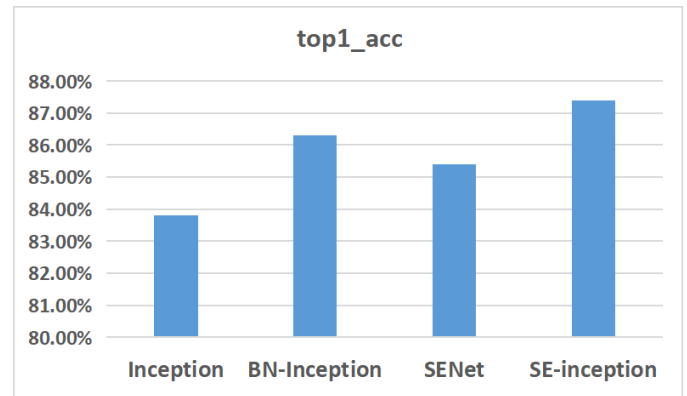


Fig 8. ACC corresponding to different networks

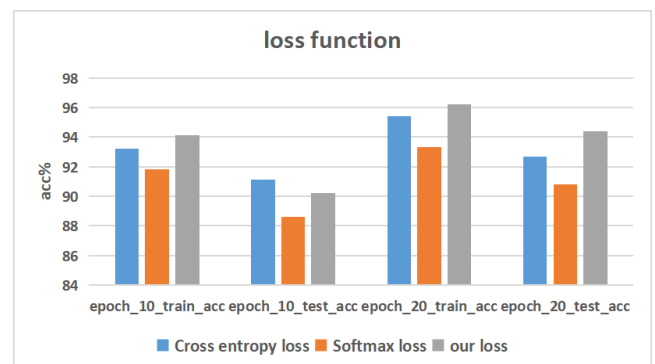


Fig 9. train_acc and test_acc of different loss function

Therefore, we get a conclusion that the accuracy of train_acc and test_acc can be improved obviously by adding unknown information parameters and setting weights to the loss function.

V. CONCLUSION

Through the comparative experiments, the paper analyzes the structure of convolutional neural network and the influence of loss function on the performance of the network. The CNN structure is constructed and the loss function is improved. The test is carried out on part of data sets from DeepFashion. The results show that the classification accuracy of the clothing image classification method based on convolutional neural network has been improved. In conclusion, the method has the advantages of simple training process, less learning parameters, fast convergence and strong generalization ability, which can effectively decrease the high classification error rate from unknown information, and the classification accuracy of which proved to be high. We will continue to study simple and efficient methods of clothing image retrieval.

ACKNOWLEDGMENTS

The research was funded by the Research Foundation of Education Bureau of Hunan Province, China (Grant No. 18 A343)

REFERENCES

- [1] Harris C G, Stephens M. A combined corner and edge detector[C]//Alvey vision conference. 1988, 15(50): 10-5244.
- [2] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International journal of computer vision, 2004, 60(2): 91-110.
- [3] Bay H, Ess A, Tuytelaars T, et al. Speeded-up robust features (SURF)[J]. Computer vision and image understanding, 2008, 110(3): 346-359.
- [4] Rublee E, Rabaud V, Konolige K, et al. ORB: An efficient alternative to SIFT or SURF[C]//2011 International conference on computer vision. Ieee, 2011: 2564-2571.
- [5] Chen L L, Han R P, Meng Q C. Clothing image feature extraction based on SURF[C]//IOP Conference Series: Materials Science and Engineering. IOP Publishing, 2018, 439(3): 032104.
- [6] ZHANG Yong, WU Yujian. Improved SIFT image feature matching algorithm. Computer Engineering and Applications, 2014, 50 (9) : 167-169.
- [7] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [8] Zhou W, Shao Z, Diao C, et al. High-resolution remote-sensing imagery retrieval using sparse features by auto-encoder[J]. Remote Sensing Letters, 2015, 6(10):775-783.
- [9] Pan R, Gao W, Liu J, et al. Automatic recognition of woven fabric pattern based on image processing and BP neural network[J]. The Journal of the Textile Institute, 2011, 102(1): 19-30.
- [10] Razavian A S, Azizpour H, Sullivan J, et al. CNN features off-the-shelf: An astounding baseline for recognition[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2014: 512-519.
- [11] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1-9.
- [12] Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2017, 31(1).
- [13] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.
- [14] Prakash K, Saravanamoorthi P, Sathishkumar R, et al. A study of image processing in agriculture[J]. International Journal of Advanced Networking and Applications, 2017, 9(1): 3311.
- [15] Qureshi I. Glaucoma detection in retinal images using image processing techniques: a survey[J]. International Journal of Advanced Networking and Applications, 2015, 7(2): 2705.
- [16] Gnanambal S, Thangaraj M, Meenatchi V T, et al. Classification algorithms with attribute selection: an evaluation study using WEKA[J]. International Journal of Advanced Networking and Applications, 2018, 9(6): 3640-3644.