Clothing image classification based on random erasing and residual network

Zongyan Gao*, Lixin Han
School of Computer and Information, Hohai University, Nanjing, Jiangsu, 211100, China

*Corresponding author’s e-mail: 1425804141@qq.com

Abstract. The traditional clothing classification method mainly consists of manually extracting obvious features such as color, texture and edge of the image. These artificial feature extraction methods are cumbersome and feature recognition rate is not high. In recent years, Deep Residual Network (ResNet) has been widely used in various fields by increasing network depth to obtain higher recognition accuracy. In this paper, the ResNet model is applied to the classification of clothing images, and on this basis, its data pooling layer is improved so that it can learn more rich features of image data. Clothing images are easy to be deformed and occluded. In this paper, a random erasing data enhancement algorithm is used to integrate and improve the model to improve the generalization ability of the ResNet model to such data. The final experimental results show that the classification accuracy of the improved residual model on clothing data in this paper is improved by 2.43%. At the same time, after integrating the random erasure data enhancement algorithm, the generalization ability of the model has been further improved.

1. Introduction
With the development of the Internet industry, you can buy your favorite clothes without leaving the house, which brings great convenience to people's daily life, but at the same time, the amount of clothing image data is also increasing. The classification of image data becomes particularly important. On the one hand, accurate classification of clothing images is conducive to the storage of massive data. On the other hand, it also facilitates people's subsequent work on automatic clothing identification, clothing retrieval, and the construction of clothing recommendation systems. However, in the face of such a large amount of data, manual labeling is time-consuming and laborious, and the results of labeling also lose their accuracy because they are easily affected by human subjective judgment. Deep learning technology can learn the deeper features of clothing images. Applying it to the classification of clothing images not only frees people's hands, but also ensures the accuracy of classification.

In recent years, deep learning (Deep Learning, DL) [1] has become an important development direction in the current field of machine learning. The use of Convolutional Neural Networks (CNN) [2] to solve problems in image classification has gradually become mainstream. Liu et al. [3] annotated the complete large clothing dataset DeepFashion, and constructed a new convolutional neural network. Lu et al. [4] proposed a clothing classification method that extracts local image information and performs light-weight processing of "block convolution" on the model. Eshwar et al. [5] collected five categories and 5,093 clothing images, and used the Inception v3 GoogleNet model to classify them in a pre-training manner. Alexander et al. [6] used the network structure built by themselves to identify clothing types, and used pixel extraction and the "jump" algorithm [7] to
identify clothing colors. Dong et al. [8] fine-tuned the VGG-19 network structure and eliminated the concerns of input image size by adding a pyramid pooling layer between the last convolutional layer and the fully connected layer of the network. Lu et al. [9] proposed an improved residual network to classify and recognize clothing images. This paper improves the convolutional layer in the residual block, adjusts the arrangement in the batch normalization layer and the activation function layer Order, meanwhile, the SE attention module is introduced to improve the recognition and classification effect of the model.

ResNet can solve the problem of gradient dispersion with the deepening of network level [10]. In this paper, the data pooling layer of the original ResNet model is adjusted, and at the same time, the generalization ability of ResNet model is improved by integrating random erasing data enhancement algorithm, and then clothing images are classified.

2. Optimized residual network model

2.1. Residual network

Deep network can naturally integrate the low, middle and high-level features and classifiers of an image into an end-to-end network, and the richer the number of stacked networks, the richer the feature levels learned [11]. However, with the deepening of the network level, there will be problems such as gradient disappearance/explosion and network degradation. For this reason, He et al. [11] proposed a deep Residual Network. Its main idea is to add an identity mapping to the network structure to directly transmit the input of the previous layer to the subsequent convolutional network, and to construct the basic residual learning unit by adding shortcut connections branches when constructing the convolutional neural network.

In this paper, considering the experimental configuration and computing power, this paper finally chose the ResNet50 network structure for the experiment, and applied ResNet to the clothing image with more fine-grained classification, hoping to get a better classification effect than on other networks.

2.2. Improved data pooling layer based on Inception module

In the traditional deep residual network, the pooling layer of data input is realized by the convolutional layer and the pooling layer, and the convolutional layer is a relatively large convolutional block. The most direct way to improve network performance is to increase the network depth and Width, depth refers to the number of network layers, and width refers to the number of neurons. However, large convolutional blocks are prone to the following problems [12]: (1). The larger the network, the more parameters will result in larger calculations and more difficult to train. (2). In the case of limited training set, too many parameters are prone to overfitting. (3). Increasing the depth of the network is prone to the phenomenon of gradient dispersion (even disappearing), and it is difficult to optimize the model. The method to solve the above problem is to increase the depth and width of the network while reducing the parameters, Inception is born under such circumstances [13]. This paper learns from the optimization module Inception of GoogleNet to improve the data pooling layer of the traditional ResNet network.

Inception network structure, the main idea is to replace some large convolution blocks with some small convolution blocks stacked together, on the one hand, the number of weight parameters is reduced; on the other hand, replacing large convolution blocks with small convolution blocks can output feature maps with complex feature patterns of different scales. Inception module is introduced in the residual network data pooling layer, and the large convolution block is changed to construct a small convolution block, and then the convolution result is batch regularized, and then merged in dimensions as the number of layers. Use three small convolutional blocks to replace the large convolutional blocks in the original ResNet. The improved data pooling layer is shown in figure 1.
2.3. Random Erasing (RE)
Occlusion is a key factor affecting image classification. In the ResNet model training process, you can use RE-based data enhancement algorithms [14] to increase the diversity of training picture data to reduce the risk of ResNet model overfitting and ultimately improve the generalization ability of the model. In the training process of the ResNet model, random erasing has a certain probability of execution. For small batches of images, assume that the probability of image I being randomly erased is $p$, and the probability of not being erased is $1-p$. In this process, training images with different degrees of occlusion are generated. Random erasing algorithm randomly selects a rectangular area $I_e$ in the image and erases it with random values. Each pixel in the selected area $I_e$ is assigned a random value in the range of 0-255.

In this paper, the RE data enhancement algorithm is used to process the image data of the training model and integrate it with the ResNet model, so as to achieve the improvement of the clothing image classification method based on the ResNet model.

3. Experiment and result analysis

3.1. Experimental environment and methods
The experiment in this article is completed on the server. Its configuration is: UBUNTU 18.04 operating system, the graphics card is NVIDIA GeForce RTX 2080 Ti GPU, 16GB memory. Adopt TensorFlow platform for training.

In the experiment in this article, TensorFlow's advanced API tf.estimator.Estimator is used in the form of pre-training. Pre-training can greatly shorten the time required for model training. In the model training process, the cross-entropy loss function and Adam algorithm are used as the optimizer for optimization. Only some optimization methods are used in network training, so the network cannot achieve the highest accuracy.

3.2. Dataset
The experiments in this paper were conducted on the public dataset FashionMNIST. This dataset covers 70,000 positive images of clothing products and is divided into ten categories [15]. There are 60,000 images as the training set and 10,000 images as the test set, in addition, we extracted 2% of the images in the training set as the validation set for cross-validation during the experiment, and each image is a 28*28 grayscale picture.
3.3. Comparative test
In this paper, the following sets of comparative tests have been made. Because the result of a single experiment may be accidental, the experimental results of this article are obtained by taking the average of ten times.

(1) The influence of different models on the classification effect.
This article uses the dataset converted in tfrecord format to compare the performance of the ResNet network structure with the current classic convolutional neural networks VGG16 and Inception_v3.

(2) Using the RE algorithm to process the effect of the dataset on the classification effect.
This article uses the RE algorithm described above to process the FashionMNIST dataset, and put the processed dataset into the ResNet network structure for training, and then the results obtained are compared with those obtained by ResNet network structure training without RE algorithm in (1).

(3) Improving the data pooling layer of ResNet network on the classification effect.
Furthermore, we put the FashionMNIST dataset processed by RE algorithm into the ResNet network structure improved by the data pooling layer for training, and compare the obtained results with the experimental results obtained by the model in (2) which only uses RE algorithm to process the data set but does not improve the ResNet data pooling layer.

3.4. Experimental results

(1) The influence of different models on the classification effect.
ResNet, because of its unique residual learning unit structure, can solve the problem of gradient dispersion and degradation in the training process to a great extent, so it can increase the depth of the network so that it can learn deeper features of clothing images. It can be seen from table 1, the classification accuracy of the model obtained by pre-training ResNet on the FashionMNIST dataset is improved by 2 to 3 percentage points compared with VGG16 and Inception_v3, and the classification effect is better.

(2) Using the RE algorithm to process the effect of the dataset on the classification effect.
It can be seen from table 1 that the classification accuracy of the ResNet model that has been processed by the RE algorithm and then trained is higher than the untreated ResNet model by a few percentage points. This is because the dataset that has been randomly erased is trained. The model's generalization ability is stronger, and it has higher adaptability to the results.

(3) Improving the data pooling layer of ResNet network on the classification effect.
From the classification accuracy results obtained in table 1, it can be seen that the ResNet model improved by the data pooling layer has an accuracy improvement of 1.76% compared to the traditional residual network. The reason is that through multiple small convolution kernels, The ResNet model can obtain richer features of the image data set, thereby making the results more accurate.

Table 1. Accuracy comparison of different model classification experiment results.

| Classification model | Accuracy (%) |
|----------------------|--------------|
| VGG16                | 89.61        |
| Inception_v3         | 90.48        |
| ResNet               | 92.83        |
| ER+ResNet            | 93.50        |
| ER+ Improved ResNet  | 95.26        |

4. Conclusion and prospect
In this paper, the residual network is used for the classification of clothing images. Experiments show that because it can train deeper networks and extract deeper features of clothing images, it has better classification effects than other networks. In view of the characteristics of easy occlusion and deformation of clothing images, this paper proposes a random erasing data enhancement algorithm to process the training data to train the model, which enhances the generalization ability of the model and
improves the classification accuracy to a certain extent. In addition, this paper also improves the data pooling layer of the residual network, so that the classification accuracy is further improved. The improved deep residual network performs better than the traditional deep residual network. Therefore, the method in this paper has certain application value for the classification of clothing images. However, there are many types and styles of clothing, so it is more difficult than ordinary image classification, and the data enhancement method is not only a random erasing, there are more methods to consider, these are the research directions in the next step.

References

[1] Zhao, H.R, Zhang, Y., Liu, G.Z. (2019) Study on the Fine-grained image classification algorithm based on RPN and B-CNN. Computer applications and software, 36(3): 216-219+270.
[2] Rawat, W., Wang, Z. (2017) Deep convolutional neural networks for image classification: A comprehensive review. Neural computation, 29(9):2352-2449.
[3] Liu, Z., Luo, P., Qiu, S., Wang, X., Tang, X. 2016. Deepfashion: Powering robust clothes recognition and retrieval with rich annotations. In: Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas, NV, USA.pp.1096-1104.
[4] Luo, M.Y., Liu, Y.F. (2019) Clothing classification method based on lightweight convolutional neural network. Computer system application,28(3):223–228.
[5] Eshwar, S.G., Rishikesh, A.V., Charan, N.A., Umadevi, V. 2016. Apparel classification using convolutional neural networks. In: International Conference on ICT in Business Industry & Government. Indore, India.pp.1-5
[6] Ivanov, A.Y., Borzunov, G. I., Kogos, K. 2018. Recognition and identification of the clothes in the photo or video using neural networks. In: 2018 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering, EIConRus 2018. St. Petersburg and Moscow, Russia.2018-January:1513-1516.
[7] Catherine A.S., Gareth M.J. (2003) Finding the number of clusters in a data set: An information-theoretic approach. Journal of the American Statistical Association, 98 (January): 750–763.
[8] Lu, J.B., Xie, X.H., Li, W.T. (2020) A garment image recognition model based on improved residual network. Computer Engineering and Applications, pp.1-7.
[9] Pan S.J., Yang Q. (2009) A survey on transfer learning[J]. IEEE Transactions on knowledge and data engineering,22(10): 1345-1359.
[10] Li, B.Q., He, Y.Y. (2018) An Improved ResNet Based on the Adjustable Shortcut Connections. IEEE Access, Volume 6:18967-18974.
[11] He, K., Zhang, X., Ren, S., Sun, J. 2016. Analysis of ResNet Model for Malicious Code Detection. In: 14th International Computer Conference on Wavelet Active Media Technology and Information Processing. Chengdu, China.pp.770-778.
[12] Tuo. (2018) Summary of Inception V1, V2, V3 and V4 Models. https://zhuanlan.zhihu.com/p/52802896
[13] Al-Qizwini, M., Barjasteh, I., Al-Qassab, H., Radha, H.2017.Deep learning algorithm for autonomous driving using GoogLeNet. In: 2017 IEEE Intelligent Vehicles Symposium (IV). Los Angeles, CA, USA.pp.89-96.
[14] Mikolajczyk, A., Grochowski, M. 2018. Data augmentation for improving deep learning in image classification problem. In: 2018 international interdisciplinary PhD workshop (IIPhDW). Swinoujście, Poland. pp. 117-122.
[15] Kayed, M., Anter, A., Mohamed, H. 2020. Classification of Garments from Fashion MNIST Dataset Using CNN LeNet-5 Architecture. In: 2020 International Conference on Innovative Trends in Communication and Computer Engineering (ITCE). Aswan, Egypt. pp. 238-243.