Facial Expression Recognition Based on TensorFlow Platform

Xiao-Ling Xia [1,] Cui Xu*[2,] Bing Nan[3]

`College of computer science. Donghua University, Shanghai, China*
[1]*College of computer science. Donghua University, Shanghai, China*
[2]*College of computer science. Donghua University, Shanghai, China*
[1]*sherlysha@dhu.edu.cn,* [2] *2151547@mail.dhu.edu.cn*

Abstract--Facial expression recognition have a wide range of applications in human-machine interaction, pattern recognition, image understanding, machine vision and other fields. Recent years, it has gradually become a hot research. However, different people have different ways of expressing their emotions, and under the influence of brightness, background and other factors, there are some difficulties in facial expression recognition. In this paper, based on the Inception-v3 model of TensorFlow platform, we use the transfer learning techniques to retrain facial expression dataset (The Extended Cohn-Kanade dataset), which can keep the accuracy of recognition and greatly reduce the training time.

1.   Introduction

Facial expression recognition is an important part of human emotion recognition, which is widely used in human-computer interaction, pattern recognition, image understanding, machine vision and other fields. There are more than 10 thousand kinds of expressions, and different people have different ways to express their emotions. In 1971, Paul Ekman [23] the famous psychologist in American proposed that the facial expressions of people from different cultures are much in common, the expression of the six basic emotions of happiness, anger, sadness, disgust, surprise and fear are very similar in many cultures.

Early facial expression recognition is mainly use the common methods in face recognition to classify and recognize facial expressions. Usually, SVM, LBP and Gabor are used to classify and recognize facial expressions according to the features of Haar, Adaboost and neural network. For example, Kobayashi et al. [21] realized the classification and recognition of basic facial expressions based on neural network. Caifeng Shan et al. using SVM classifier based on the LBP features to achieve facial expression recognition [18]. Ioan Buciu et al. using SVM classifier based on ICA and Gabor features for facial expression classification, the facial expression recognition system realized by combining Gabor wavelet and SVM classifier can achieve a higher recognition rate [20]. Xia Mao et al. [17] realized robust facial expression recognition based on RPCA and AdaBoost. Early facial expression recognition methods lack analysis of the unique features of facial expression.

In recent years, as a new recognition method combining artificial neural network and depth learning theory, convolutional neural network has made great progress in the field of image classification. This method uses local receptive field, weights sharing and pooling technology and makes the training parameters greatly reduced compared to the neural network. It also has a certain degree of translation, rotation and distortion invariance of image. It has been widely used in speech recognition, face recognition, handwriting recognition and other applications. Compared with traditional methods, it has higher recognition rate and wider application.

TensorFlow [1] is the second generation of artificial intelligence research and development system developed by Google, which supports convolutional neural network (CNN) [3], recurrent neural network (RNN) and other depth neural network model. This system is widely used in Google's products and services, it has been deployed in more than and 100 machine learning projects, involving more than a dozen field of speech recognition, computer vision, robotics, information retrieval, information extraction, natural language processing, drug testing.

This paper implements an effective facial expression recognition model based on Inception-v3 [8] model under the TensorFlow [1] platform. In this paper, we use the transfer learning technique to retrain the Inception-v3 [8] model on the facial expression dataset, which can reduce the training time as much as possible. The rest of this paper is organized as follows. Section II introduces the related work on facial expression recognition. In section III, we introduce the construction process of

facial expression classification model. We prove the validity of the model through experiments in section IV.

2.   Related Work

In 1971, Paul [23] from the psychological point of view proposed that there are six basic emotions (happiness, sadness, anger, disgust, surprise and fear) across cultures. In 1978, Ekman et al. [22] developed the facial action coding system (FACS) to describe facial expressions. Nowadays, most of the facial expression recognition work is done on the basis of the above work, this paper also chose the six basic emotions and neutral emotions as the standard of facial expression classification.

Lu Guanming et al. [2] proposed a convolutional neural network for facial expression recognition, and the dropout strategy and the dataset expansion strategy are adopted to solve the problem of insufficient training data and the problem of overfitting. C. Shan et al. [16] uses LBP and SVM algorithm to classify facial expressions with 95.10% accuracy. Andre Teixeira Lopes et al. [7] used a depth convolution neural network to classify facial expression, the accuracy reached 97.81%. However, in terms of the convolutional neural network algorithm, low level network model achieves low accuracy, the deep network models often has some problems, such as lack of data, easy to over fitting, long training time and so on.

In the traditional classification and learning, in order to ensure the accuracy and reliability of the classification model trained, need to meet two basic assumptions: the first is the training samples and test samples meet with independent distribution; the second is that it need enough training data. However, in many cases, these two conditions are difficult to meet. The most likely scenario is that the training data is out of date. This often requires us to relabel a large number of training data to meet the needs of our training, but it is very expensive to label new data, which requires a lot of manpower and material resources.

Transfer learning is a new machine learning method based on existing knowledge to solve different but related problems. The goal of transfer learning is to apply the knowledge learned from an environment into a new environment. Compared with the traditional machine learning, transfer learning relaxes the requirements of the two basic assumptions, it cannot require a large amount of training data, or only a small amount of data can be labeled. Transfer learning does not have to be like the traditional machine learning as the training samples and test samples need to be independent and identically distributed. At the same time, compared with the traditional network with random initialization, the learning speed of transfer learning is much faster.

3.   Construction   of   Facial Expression Classification Model

This section focuses on the construction process of facial expression classification model. The process of facial expression can be divided into four steps: image preprocessing, training, validation and testing. The basic steps of facial expression classification model are shown in figure 1.
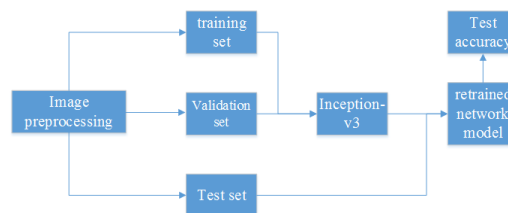

Figure1. Facial expression classification model.

3.1   Image Preprocessing

In order to improve the effect of image classification, image preprocessing is a very important stage. The target image preprocessing includes image format conversion, image clipping.

Image format conversion. Using Python image processing library PIL to achieve the conversion of different image formats.

Image clipping. In order to improve the accuracy of image classification and reduce the interference of other non target information on image classification, the target image can be cut in the process of image preprocessing.

3.2   Inception-v3

Inception-v3 [8] is a network model of Google after Inception-v1 [13], Inception-v2 [9] in 2015. It is a rethinking for the initial structure of computer vision.

Inception-v3 [8] model is mainly based on the basic principles of network design of the Inception structure changes, the main methods are: large size filter convolution decomposition, additional classifier, reduce the size of the feature map, etc.

The Inception-v3 [8] model is trained on the ImageNet datasets (1 million 200 thousand training set, validation set 50000, 100000 test set), containing the information that can identify 1000 classes in ImageNet, the error rate of top-5 is 3.5%, the error rate of top-1 dropped to 17.3%.

3.3   Transfer Learning

Inception-v3 network model is a complex network, it will cost at least a few days or even a few day time if we train the model directly. By using the method of transfer learning, the parameters of the previous layer are unchanged, but only the last layer is trained. The last layer is a softmax classifier. This classifier is 1000 output nodes in the original network (ImageNet has 1000 classes), so we need to remove the last network layer, and then retrain the last layer. The reason for retraining the last layer is to work on the new object class. As a result, identifying 1000 classes of information in the ImageNet is also useful for identifying new classes.

4.   Experiment

This experiment is based on the Inception-v3 [8] model of TensorFlow [1] platform, the hardware platform is MacBook Pro: processor 2.9GHz Intel i7, memory 8GB 1600MHz DDR3. The use of the experimental dataset is CK+ facial expression dataset (The Extended Cohn-Kanade dataset [15], CK+). We select 1004 images of facial expression image, which contains 7 basic facial expressions: happiness (158), sad (155), anger (103), disgust (146), surprised (161), fear (137) and neutral (144)

In this section, the following part is as follows: first, we make a simple introduction on the dataset; then, we introduce the process of the experiment in detail; finally, we verify the effectiveness of the method through the comparison experiment.

4.1   Dataset

The CK+ dataset [15], was released in 2010 and is based on the CK dataset. The image data collected from 210 adults, aged 18-50, of which women accounted for 69%, 81% of Europe and the United States, more than 13% of African Americans. The dataset contains 123 items, 593 image sequences, including 327 image sequences with emotional labels. The dataset is a common dataset for facial expression recognition.

4.2   Experimental Procedure

Image preprocessing. First, the Inception-v3 [8] model is the image of JPG or JPEG format for training, and the image format in the dataset for the PNG format, so to convert image format to PNG format images converted to jpg format. Secondly, because the image in the dataset is the facial expression image captured by the digital camera, and some of the color image, and some gray image, so the image must be converted to grayscale image. Finally, in order to remove the interference of the background and hair improve the accuracy of image classification, we should cut out the face region in the image, and use the clipping image for training, verification and testing.

The Inception-v3 [8] network model is a very complex network, it will cost a lot of time if we train the model directly, and the data of CK+ dataset in [15] is relatively small, the training data is insufficient. Therefore,we use transfer learning technology to retrain the Inception-v3 [8] model.

The last layer of the Inception-v3 [8] model is a softmax classifier, because there are 1000 classes in the ImageNet dataset, so the classifier has 1000 output nodes in the original network. Here, we need to delete the last layer of the network, put the number of output nodes to 7 (the number of facial expressions), and then retrain the network model.

The last layer of the model is trained by back propagation algorithm, and the cross entropy cost function is used to adjust the weight parameter by calculating the error between the output of the softmax layer and the label vector of the given sample category.

4.3   Control Experiment

In order to verify the effectiveness of the proposed method, experiments are carried out to compare the facial expression classification algorithm based on MLP [2], CNN +AD [2], LBP+SVM [16] and CNN [7], and the results are shown in Table I.

Table1. Classification performance comparison of different algorithms in CK+ dataset

Method	Performance
MLP[2]	81.11%
CNN+AD[2]	84.55%
LBP+SVM[16]	95.10%
CNN[7]	97.81%
Our Method	97%

From the experimental results of Table I, we can see that the classification accuracy of this method is higher than that of MLP [2], CNN+AD [2] and LBP+SVM [16] algorithm, but it is less than CNN [7].

Table II shows the number of training layers in CNN [7] and our method.

Table2. The Number of Training Layers

Method	Layers
CNN[7]	5
Our Method	1

We know that the training time will be increased dramatically when we increase one layer. Compared with the method used in CNN [7], our method need to train fewer layers. So Compared with the method used in CNN [7], our method can greatly reduce the training time.

5.   Conclusion

In this paper, based on the Inception-v3 model of Tensor Flow platform, we use the transfer learning technology to train a new facial expression classification model in CK+ dataset. The classification accuracy of the model is 97%, which is higher than that of MLP, LBP and low-level network model. Compared with some deep network models, this paper takes less time. The future work is to study and develop a facial expression recognition model based on dynamic sequences.

References

[1] Martín Abadi, Ashish Agarwal, et al: TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. CoRR abs/1603.04467 (2016)

[2] Lu Guanming, Shi Wanwan, Li Xu, et al: A kind of convolution neural network for facial expression recognition [J]. Journal of Nanjing University of Posts and Telecommunications (NATURAL SCIENCE EDITION). 2016 (01)

[3] Abhineet Saxena: Convolutional neural networks: an illustration in TensorFlow. ACM Crossroads 22(4): 56-58 (2016)

[4] Bo-Kyeong Kim, Jihyeon Roh, Suh-Yeon Dong, Soo-Young Lee: Hierarchical committee of deep convolutional neural networks for robust facial expression recognition. J. Multimodal User Interfaces 10(2): 173-189 (2016)

[5] Zhenhai Liu, Hanzi Wang, Yan Yan, Guanjun Guo: Effective Facial Expression Recognition via the Boosted Convolutional Neural Network. CCCV (1) 2015: 179-188

[6] Mao Xu, Wei Cheng, Qian Zhao, Li Ma, Fang Xu: Facial expression recognition based on transfer learning from deep convolutional networks. ICNC 2015: 702-708

[7] Andre Teixeira Lopes, Edilson de Aguiar, Thiago Oliveira-Santos: A Facial Expression Recognition System Using Convolutional Networks. SIBGRAPI 2015: 273-280

[8] Christian Szegedy, Vincent Vanhoucke, et at: Rethinking the Inception Architecture for Computer Vision. arXiv:1512.00567, 2015.

[9] Sergey Ioffe, Christian Szegedy, et al: Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. ICML2015: 448-456, 2015.

[10] Andre Teixeira Lopes, Edilson de Aguiar, Thiago Oliveira-Santos: A Facial Expression Recognition System Using Convolutional Networks. SIBGRAPI 2015: 273-280

[11] Zhuang Fuzhen, Luoping, Heqing, history Zhongzhi: Transfer learning progress of [J]. Journal of software 2015 (01).

[12] Andre Teixeira Lopes, Edilson de Aguiar, Thiago Oliveira-Santos: A Facial Expression Recognition System Using Convolutional Networks. SIBGRAPI 2015: 273-280

[13] Christian Szegedy, Wei Liu, et al: Going Deeper with Convolutions. arXiv:1409.4842, 2014.

[14] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton: ImageNet Classification with Deep Convolutional Neural Networks. NIPS 2012: 1106-1114

[15] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression,"in Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on. IEEE, 2010, pp. 94–101.

[16] C. Shan, S. Gong, and P. W. McOwan: "Facial expression recognition based on local binary patterns: A comprehensive study," Image and Vision Computing, vol. 27, no. 6, pp. 803–816, 2009.

[17] Xia Mao, Yu-Li Xue, Zheng Li, Kang Huang, ShanWei Lv: Robust facial expression recognition based on RPCA and AdaBoost. WIAMIS 2009: 113-116

[18] Caifeng Shan, Tommaso Gritti: Learning Discriminative LBP-Histogram Bins for Facial Expression Recognition. BMVC 2008: 1-10

[19] Sung-Oh Lee, Yong-Guk Kim, Gwi-Tae Park: Facial Expression Recognition Based upon Gabor-Wavelets Based Enhanced Fisher Model. ISCIS 2003: 490-496

[20] Ioan Buciu, Constantine Kotropoulos, Ioannis Pitas: ICA and Gabor representation for facial expression recognition. ICIP (2) 2003: 855-858

[21] Kobayashi H, Hara F: The recognition of basic facial expression by neutral network. [C], International Joint Conference on Neural Network. 1991:460-466.

[22] E. Friesen and P. Ekman: "Facial action coding system: a technique for the measurement of facial movement," Palo Alto, 1978.

[23] P. Ekman: "Universals and cultural differences in facial expressions of emotion."in Nebraska symposium on motivation. University of Nebraska Press, 1971.