International Core Journal of Engineering Volume 7 Issue 5, 2021
ISSN: 2414-1895 DOI: 10.6919/ICJE.202105_7(5).0065

Facial Expression Recognition Method based on Convolution Neural Network

Guangyuan Zhong, Gaoyuan Liu, Xinglin Du

Shandong University of Science and Technology, Tai'an City, Shandong Province, 271019, China.

Abstract

Aiming at the key and difficult problems in facial expression recognition, such as feature mining and classifier design, this paper takes facial expression image as the object to carry out the research of facial expression recognition based on lightweight convolutional neural network. Based on the construction of spontaneous expression feature view in natural environment, this paper introduces automatic data augmentation technology to obtain rich sample information oriented to deep convolution neural network, and establishes deep learning model of facial expression recognition with the traction of texture information extraction and recognition mechanism of facial expression. The effectiveness of the proposed method is verified on FER2013, FERplus, CK +, SFEW and RAF-DB data sets, and good results are obtained.

Keywords

Convolutional Neural Network; Expression Recognition; Data Augmentation.

1. Introduction

With the continuous breakthrough and improvement of technologies in the Internet, big data and other fields, global informatization has entered a new stage of full penetration and accelerated innovation, which has triggered a new round of scientific and technological revolution and industrial change[1,2,3]. New technologies such as artificial intelligence, machine deep learning and Internet of things are in the boom of development. The real world and the digital world are increasingly converging. Digital, networked and intelligent services are everywhere. With the advent of various intelligent devices, human life and work are more and more closely connected with computers.

Facial expression is one of the most common and natural ways for human beings to convey emotional state and intention. Mehrabian's research shows that in interpersonal communication, the information conveyed by facial expression accounts for a very large proportion, up to 55%, 38% from the speaker's tone, and only 7% depends on the speaker's content[4,5]. It can be seen that facial expression plays an indispensable role in the process of information exchange between people. It is of great practical significance to build a system that can automatically analyze facial expressions in the fields of medical treatment, education and driverless vehicles[6].

With the development of hardware and technology, researchers have almost solved the problem of non spontaneous facial expression recognition in the experimental environment, and began to march towards spontaneous facial expression recognition in the natural environment. In view of the success of deep learning in recent years, many researchers have trained deep models on public data sets[7,8,9]. At present, the research of natural expression recognition focuses on solving two problems. First, the balance between recognition accuracy and computational efficiency of deep learning model. Second, illumination, head posture and occlusion are independent of facial expression.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202105_7(5).0065

2. Related work

General facial expression recognition methods include four steps, namely face detection, face correction, feature extraction and expression recognition[10]. However, due to different expression definition methods and different expression data forms (2D, 3D and thermal images[11]), there are some differences in the execution process, as shown in Figure 1.

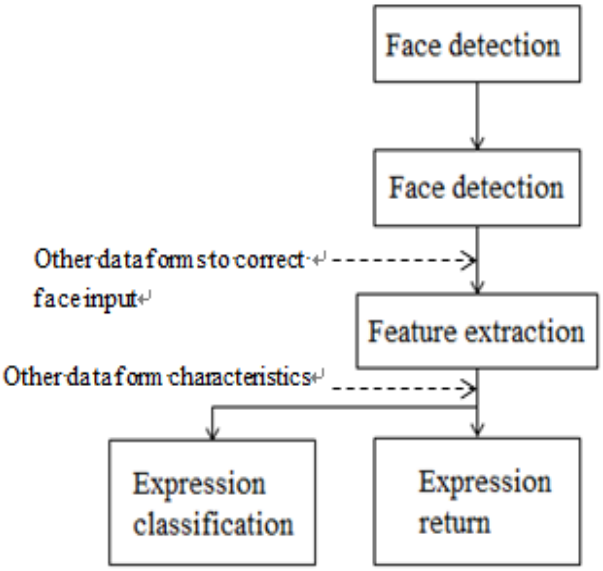

Figure 1. Expression recognition process

There are three modes of facial expression recognition data, namely two-dimensional, three-dimensional and thermal imaging. Facial expression feature extraction is mainly based on texture information, but the utilization rate of color information is low[12]. In addition, 3D images need higher dimensional geometric operations, and thermal images contain insufficient texture information. Therefore, two-dimensional gray image is the most important research object[13].

There are two main definitions of facial expression. The first is continuous action units. There are about 44 action units based on facial coding system, which researchers later supplemented. Through the combination of single or multiple action units, several facial expressions are further formed. Because of the complexity of computation, most researchers take the expression composed of a limited set of action units as the recognition object. The second is the discrete basic expression. At the beginning, there are six basic expressions: anger, disgust, fear, joy, sadness and surprise, as shown in Figure 2. Gao Wen and others[14] described six basic expressions in detail according to the actual characteristics, and built a model based on them. Basic expressions are widely concerned because they are easy to understand and easy to experiment. Later, researchers added the expressionless expression into it to become the seventh basic expression neutral[15,16]. The proposed method is also based on these seven basic expressions.

Figure 2. Six basic emotions. From left to right: disgust, fear, joy, surprise, sadness, anger

ISSN: 2414-1895 DOI: 10.6919/ICJE.202105_7(5).0065

3. Method

Facial expression recognition is an important step towards natural and harmonious human-computer interaction. At present, the application of human-computer interaction is still in its infancy, the machine can not accurately understand human emotions and needs, and can only complete the specified steps according to the steps set by the program. This leads to the weak interaction between the machine and the user, and the essential problem is the poor effect of facial expression recognition algorithm.

3.1 Face detection and correction

In the face detection part, we use support vector machine combined with gradient direction histogram. Firstly, the feature vectors are constructed by calculating the gradient direction of the local region, and then all the feature vectors are input into the classifier. If the output result is positive, the face position is returned, specifically, the coordinates of the upper left corner and the lower right corner of the rectangle are detected.

Compared with other methods, this method balances accuracy and computation speed better, and is more suitable for online recognition applications. The details of calculating the pixel gradient in the image are shown in the formula.

$$f_x(x,y) = f(x+1,y) - f(x-1,y)$$
$$f_y(x,y) = f(x,y+1) - f(x,y-1)$$
$$m(x,y) = \sqrt{f_x(x,y)^2 + f_y(x,y)^2}$$
$$\theta(x,y) = \arctan(f_x(\mathrm{x},\mathrm{y})/f_y(x,y))$$

Where m and θ are magnitude and direction, respectively.

In the face correction part, we use the millisecond set method proposed in, use gradient enhancement to train several regression trees, and then use decision tree set to calculate 68 landmarks including eye contour, nose and mouth contour.

3.2 Dense convolutional neural network

Dense convolutional neural network (DenseNet) is a unique convolutional neural network (CNN) architecture, which can minimize the trainable parameters through dense connection mode and many dimension reduction layers.

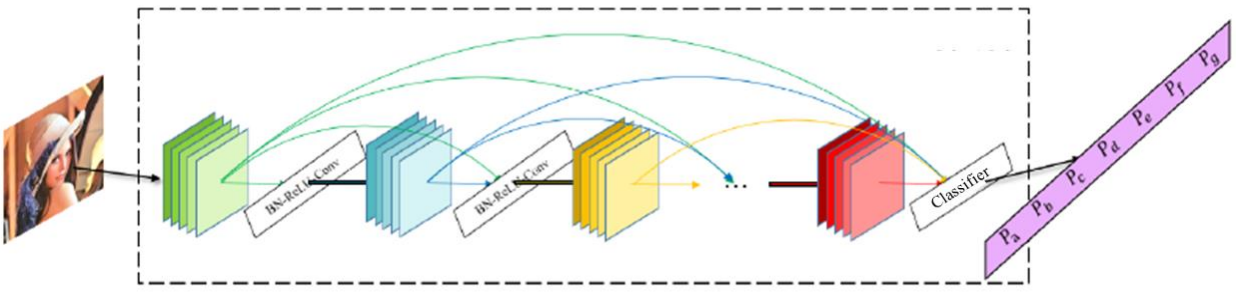

Figure 3. Dense convolutional neural network

In fact, dense convolutional neural network has two key super parameters, namely growth rate *K* and dense block number *n*. The growth rate represents the number of convolution layer filters, which determines the growth rate of the feature graph. We use 2 × 2 average pooling instead of 2 × 2 maximum pooling, because it forces the correspondence between feature maps and categories to be more adaptive to convolution structure. And maximum pooling discards three-quarters of the information, while average pooling takes all the information into account.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202105_7(5).0065

In addition, the average pooling is robust to the input spatial transformation because it sums the spatial information. Mean normalization is actually a generalization function, which can prevent dense connections from falling into the problem of over fitting.

4. Experiment

4.1 Experimental platform

Our model is trained and processed on NVIDIA Titan X graphics card. It has 3584 CUDA units, 12Gb GDDR5X memory, 1531MHz core frequency and 7.0 TFlops single precision floating-point operation. In the software, our algorithm design is based on Python 3.6 and Pytorch deep learning toolkit.

4.2 Experimental results

We trained three DenseNet models on FER2013, FERPLUS and FERFIN datasets. For the discrete case, the setting of super parameters is slightly different.

Table 1. Three densely convolutional neural model architectures.

network layer	feature size	Densenet-1	Densenet-2	Densenet-3
convolution layer	48×48		3×3 conv	
dense block	48×48	(1×1, 3×3) ×12	(1×1, 3×3) ×12	(1×1, 3×3) ×6
transition layer	48×48 24×24		1×1 conv 2×2 AvgPooling	
dense block	24×24	(1×1, 3×3) ×12	(1×1, 3×3) ×12	(1×1, 3×3) ×12
transition layer	24×24 12×12		1×1 conv 2×2 AvgPooling	
dense block	12×12	(1×1, 3×3) ×12	(1×1, 3×3) ×12	(1×1, 3×3) ×24
transition layer	12×12 6×6		1×1 conv 2×2 AvgPooling	
dense block	6×6		(1×1, 3×3) ×12	(1×1, 3×3) ×16
transition layer	6×6 3×3		1×1 conv 2×2 AvgPooling	1×1 conv 2×2 AvgPooling
classification layer	1×1	6×6 7D-Softmax	3×3 10D-Softmax	3×3 7D-Softmax

On the FER2013 dataset, the verifying accuracy of DenseNet-3 is 72.62%, which exceeds 71.16% of the first group in the challenge. We believe that there are two reasons why DenseNet-3 can achieve this result without using any integration method and very few parameters. First of all, feature reuse method increases the input size of subsequent volume layers, and enables subsequent layers to learn new features while accepting prior knowledge of network. Secondly, the setting of dense connection and bottleneck layer greatly reduces the parameters of the network, making the network extract more compact and more distinctive features. On the FERPLUS dataset, the accuracy of DenseNet-2 is 86.58%. This is 0.69% higher than Barsoum's VGG13. The parameter of DenseNet-2 is 41 times less than that of VGG13. When using DenseNet-1, this number is 92 times, and the accuracy is reduced by 0.52%. In FERFIN, the same densenet-2 achieves 86.89% verification accuracy, which validates the assumption that there are noise categories. Because the categories in the database are clearer, dense convolutional neural network has learned more robust features. All results are listed in Table 2.

Table 2. Experiment results of accuracy

models	FER2013	FERPLUS	FERFIN
DenseNet-1	71.911%	85.06%	85.25%
DenseNet-2	72.55%	86.58%	86.89%
DenseNet-3	72.62%	86.67%	86.90%

5. Conclusion

Many researchers believe that dynamic data can extract more useful features to recognize spontaneous facial expression, which is a future research topic worthy of attention. In the future work, we plan to

ISSN: 2414-1895 DOI: 10.6919/ICJE.202105_7(5).0065

detect spontaneous emotions by considering time information, while still introducing lightweight algorithm to ensure the real-time application of the model in practical applications.

References

[1] Wu Haopeng, Lu Zhiying, Zhang Jianfeng, Li Xin, Zhao Mingyue, Ding Xudong. Facial Expression Recognition Based on Multi-Features Cooperative Deep Convolutional Network[J]. Applied Sciences, 2021, 11(4).

[2] Indira DNVSLS, Sumalatha L, Markapudi Babu Rao. Multi Facial Expression Recognition (MFER) for Identifying Customer Satisfaction on Products using Deep CNN and Haar Cascade Classifier[J]. IOP Conference Series: Materials Science and Engineering, 2021,1074(1).

[3] Cai Yongxiang, Gao Jingwen, Zhang Gen, Liu Yuangang. Efficient facial expression recognition based on convolutional neural network[J]. Intelligent Data Analysis, 2021,25(1).

[4] Ahmed Rachid Hazourli, Amine Djeghri, Hanan Salam, Alice Othmani. Multi-facial patches aggregation network for facial expression recognition and facial regions contributions to emotion display[J]. Multimedia Tools and Applications, 2021(prepublish).

[5] H. Sikkandar,R. Thiyagarajan. Deep learning based facial expression recognition using improved Cat Swarm Optimization[J]. Journal of Ambient Intelligence and Humanized Computing,2020(prepublish).

[6] Multimedia; Findings on Multimedia Reported by Investigators at Erciyes University (Static Facial Expression Recognition Using Convolutional Neural Networks Based On Transfer Learning and Hyperparameter Optimization)[J]. Journal of Engineering, 2020.

[7] Shixin Cen, Yang Yu, Gang Yan, Ming Yu, Qing Yang. Sparse Spatiotemporal Descriptor for Micro-Expression Recognition Using Enhanced Local Cube Binary Pattern[J]. Sensors, 2020,20(16).

[8] Jiadi Bao,Shusong Wei,Jingfan Lv,Wenli Zhang. Optimized Faster-RCNN in Real-time Facial Expression Classification[A]. Advanced Science and Industry Research Center.Proceedings of 2019 2nd International Conference on Communication,Network and Artificial Intelligence(CNAI 2019)[C].Advanced Science and Industry Research Center:Science and Engineering Research Center,2019:8.

[9] Qi-di HU, Qian SHU, Ming-ze BAI, Xiao-ming YAO, Kun-xian SHU. FERCaps: A Capsule-Based Method for Face Expression Recognition from Frontal Face Images[A]. Advanced Science and Technology Application Research Center.Proceedings of 2019 International Conference on Power, Energy, Environment and Material Science (PEEMS 2019)[C]. Advanced Science and Technology Application Research Center:Advanced Science and Technology Application Research Center, 2019:6.

[10] Thai Son Ly, Nhu-Tai Do, Soo-Hyung Kim, Hyung-Jeong Yang, Guee-Sang Lee. A novel 2D and 3D multimodal approach for in-the-wild facial expression recognition[J]. Image and Vision Computing, 2019,92.

[11] Signal Processing; Study Data from Guangdong University of Technology Update Understanding of Signal Processing (Occlusion Expression Recognition Based On Non-convex Low-rank Double Dictionaries and Occlusion Error Model) [J]. Electronics Newsweekly,2019.

[12] Yunfang Fu, Qiuqi Ruan, Ziyan Luo, Yi Jin, Gaoyun An, Jun Wan. FERLrTc: 2D+3D facial expression recognition via low-rank tensor completion[J]. Signal Processing, 2019,161.

[13] Pollux Petra M J, Craddock Matthew, Guo Kun. Gaze patterns in viewing static and dynamic body expressions. [J]. Acta psychologica, 2019,198.

[14] Kai Kang, Xin Ma. Convolutional Gate Recurrent Unit for Video Facial Expression Recognition in the Wild[A]. Engineering Society of China, 2019:6.

[15] I. Michael Revina, W.R. Sam Emmanuel. Face Expression Recognition with the Optimization based Multi-SVNN Classifier and the Modified LDP Features[J]. Journal of Visual Communication and Image Representation, 2019,62.

[16] Ying Xiao, Deyan Wang, Ligong Hou. Unsupervised emotion recognition algorithm based on improved deep belief model in combination with probabilistic linear discriminant analysis[J]. Personal and Ubiquitous Computing, 2019,23(3-4).