International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Special Issue: 1 Volume No.7 April – 2023
DOI:10.46647/ijetms.2023.v07si01.088 ISSN: 2581-4621

Smart music player based on Facial Micro-expression Recognition

Nalajamgeethanjali[1],T.Rajithareddy[2], V.Sasikala[3], P.Thulasireddy[4], P.Saisahithi[5]
Assistant professor, Student, Department of CSE, AITS, Tirupati

Abstract
Deep learning has gained increasing interest in recent years thanks to the creation and use of large data. Convolutional neural networks, a type of deep learning neural network, are crucial for recognizing faces in images. Convolutional neural network technology for micro expression detection and automatic music transcription are combined in this paper. A recommendation system is created to find a model that can identify facial micro expressions and suggest music based on the associated mood. The facial micro-expression recognition model established with a recognition rate of 62.1%. This research helps to improve the practicality of the music recommendation system, and the related results will also serve as a reference for the application of the music recommendation system in areas such as emotion regulation.
Keywords: Deep learning, Facial expression recognition, CNN, Music recommendation algorithm.

Introduction

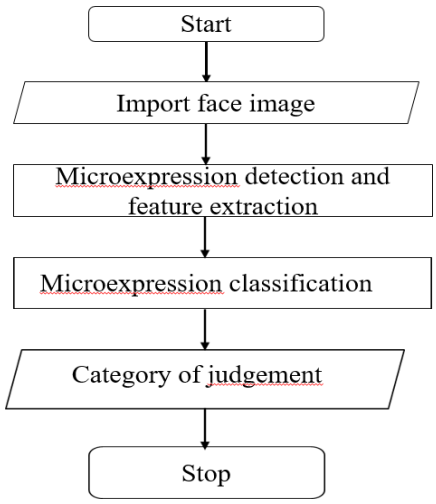
Deep learning is widely employed in picture recognition, image processing, and particularly facial expression detection with the advent of the information age. In the area of human-computer interaction, face recognition has emerged as aresearch hotspot, yet there are still restrictions onhow image processing results can be used. The data in the image is not used for secondary processing, which means that it has not beenfully and effectively utilised in the actual production and life process, and image research frequently focuses on increasing the accuracy of recognition. In this study, a convolutional neural network expression recognition model is created and trained using a deep learning technique. When the outcomes of image processing are integrated with a music recommendation system, the music that best suits the person's mood is suggested. Major music websites' playlists and manually annotated tracks are crawled to build music data collections. The outputs of image processing have a suitably broader range of applications.

1. Facial Emotions Recognition Based On CNN

1.1 Basic Steps For Micro-expression

Recognition The basic process of facial micro-expressionrecognition is as follows:

1. Obtain micro facial expression images ofhuman faces and pre-process the images;
2. Perform micro-expression detection andrelated feature extraction;
3. Perform micro-expression classification.



International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Special Issue: 1 Volume No.7 April – 2023
DOI:10.46647/ijetms.2023.v07si01.088 ISSN: 2581-4621

Image Pre-processing

The pre-processing of facial images is an important step in the facial microexpression recognition system. The quality of the input image (size, pixels, grey value, etc.) is not uniform due to a variety of factors, and the image cannot be directly used in the subsequent face recognition process, which has a significant impact on recognition performance. As a result, image pre-processing has standardised and unified standards, which can eliminate the impact of size, posture, brightness, and darkness on recognition performance. Reduce as much as possible the interference of irrelevant information, thereby increasing the recognition rate of microexpressions. Histogram equalisation, median filtering, grayscale stretching, homomorphic filtering, nearest neighbour algorithm, bilinear interpolation, and other preprocessing methods are available [2-4]. Different micro-expression recognition systems necessitate different image sources, recognition algorithms, and pre-processing techniques. The FER2013 facial micro-expression data set is used in this paper to eliminate the image preprocessing process.

Micro-expressions and Feature Extraction

Feature extraction is an essential component of microexpression recognition. A good feature extraction method can obtain complete and accurate feature information for differentiating between different types of objects. It can also reduce data dimensionality and avoid the interference of irrelevant data. The deep feature extraction method is used in this paper's micro-expression recognition model. Deep learning technology has gradually entered our field of vision with the popularisation and development of big data, and deep learning networks are also of great importance to achieve micro-expression recognition. Traditional facial expression recognition algorithms have poor robustness and insufficient ability to extract expression features. Face expression recognition algorithms based on convolutional neural networks can compensate for these shortcomings. This article employs a convolutional neural network-based feature extraction algorithm.

Convolutional Neural Network (CNN)

CNN neural network stands for convolutional neural network, and it can extract and learn more complex image features. It is similar to other neural networks in that it uses forward propagation to input data and activate the hidden layer, while backpropagation loss values change the hidden layer's parameters. The activation of the hidden layer is determined by the activation function, which introduces nonlinear factors so that the neural network can approach any nonlinear function arbitrarily. CNN differs from other neural networks in that it is a deep neural network. Following the alternate connection of several convolutional layers and pooling layers, the fully connected layer is connected, thereby replacing the hidden layer within the network. However, as the number of network layers and learning parameters increases, so does the risk of fitting during the training process, and the model's generalisation ability. CNN is divided into two parts: the convolution pooling layer, which completes the extraction of image features, and the output, which is classified by the fully connected layer. Convolutional layer: it is composed of several convolutional units, and the parameters of each convolutional unit are optimised using a back propagation algorithm. A convolution operation is performed on the original image and a specific feature filter by the convolution layer. Each operation in multiple convolution processes uses a different filter to map different features.

Pooling layer: Down sampling is a type of pooling. The pooling layer function continuously reduces the data space size, as well as the number of parameters and calculations. It also controls overfitting to some extent.

After flattening the previous results, each node is connected to all nodes in the previous layer, which is used to integrate the previously extracted features.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Special Issue: 1 Volume No.7 April – 2023
DOI:10.46647/ijetms.2023.v07si01.088 ISSN: 2581-4621

Table 1. CNN Model Parameter Settings

Layer Type	Core Size	Output
Input Layer		48*48
Convolutional Layer 1	64&3*3	64&42*42
Convolutional Layer 2	128&3*3	128&36*36
Pooling Layer 1	128&2*2	128&18*18
Convolutional Layer 3	256&3*3	256&12*12
Convolutional Layer 4	256&3*3	256&6*6
Pooling Layer 2	256&2*2	256&3*3
Fully Connected Layer 1		512
Fully Connected Layer 2		7

FER2013 Data Set

For model training, this model employs the FER2013 data set. The FER2013 data set contains 358 86 images of facial expressions. There are 28708 training images, 3589 public test images, and 358 9 private test images among them. Each Among them, x represents the neuron input, y represents t he neuron output, w represents the weight matrix, and b represents the bias matrix. TensorFlow's S oftmaxfunction is used to assign a picture that is composed of a grayscale image with a fixed size of 48 * 48, which is a pose. There are also unposed avatars.

Model Training

CNN is a feedforward neural network, so the training of CNN is divided into two processes of for ward propagation and back propagation [9]. During the forward propagation process, each neuron i n eac-h convolution

After adding a bias, the result is passed through the activation function as an output to form the ne urons of the current layer. These neurons constitute the current layer. Multiple feature maps for dif fere-nt features.

2. Music Recommendation Algorithm

2.1 The Establishment of Song Library

Use Python to crawler music website to store song files and song information. The song information is stored in Excel form. Based on the seven kinds of emotions(anger, dispute, bear, happy, sad, secure, neutral) that can be recognized by the emotion recognition model, the emotion analysis and classification are carried out for the crawled songs, and then the song library corresponding to different emotions is established.

The analysis method of song emotion in this paper can be divided into two steps: lyrics data mining and song emotion analysis.

2.1.1 Data Mining of Lyrics

The steps of data mining for lyrics[13] are as follows (take Chinese songs as an example, the same treatment for English songs):

a. Word segmentation

In this paper, we use the open-source Chinese word breaking system Jieba as a tool to segment the lyrics.

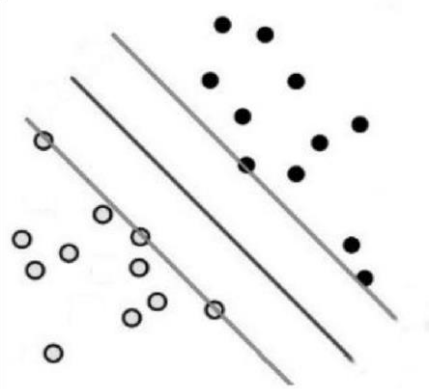
b. Comparison with emotional lexicon

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Special Issue: 1 Volume No.7 April – 2023
DOI:10.46647/ijetms.2023.v07si01.088 ISSN: 2581-4621

In this paper, we draw lessons from the lexicon established by others —— a Chinese emotional lexicon integrated with HowNet and Sogou as a comparison model.
In the lexicon, words are first divided into two sets with opposite meanings, that is, the set containing positive emotional words (such as "happy", "happy", "warm", etc.) and the set containing negative emotional words (such as "degenerate", "abandonment", "despair", etc.). In order to distinguish the emotion of words more accurately, this paper integrates emotion words, degree words (such as "good" in "good happiness") and negative words to compare the lyrics and find the emotional characteristic values in the lyrics.

2.1.2 Emotional Analysis of Songs

This paper uses SVM model in emotion classification. The general principle of SVM is as follows (schematic diagram is shown in Figure3):
We need to classify the black and white points in the graph. SVM will find the boundary between the plane where the black point is and the plane where the white point is, such as the dark line in the graph. The two light lines are used to indicate the distance w between the points and the black line in the two planes, and the goal of SVM is to find the black line with the largest distance Max (w). This paper uses LIBSVM (Chang et al.,2011), a SVM tool set developed by Professor Zhiren Lin of Taiwan University,to classify the training with R language.



Because this paper involves seven emotions, it needs to be classified many times. First, all categories are divided into two subcategories (such as positive emotion and negative emotion), and then the subcategory is further divided into two sub categories, so as to cycle until a separate category is obtained.

2.2 Content-based Recommendation Algorithm (CB) Content-based recommendation is to discover the relevance of items based on the metadata of the recommended items, and then recommend similar items to users based on their past preferences. Under the condition that the song library has been established and facial emotions have been determined, due to the difference in the emotions expressed by different songs in the same emotion, the random recommendation of the existence of the recommended songs cannot relieve negative emotions or promote The role of positive emotions, so the original random recommendation algorithm was improved to a content-based recommendation algorithm.

2.2.1 General Steps of Content-based Recommendation

The general steps of content-based recommendation are as follows:
1. Item Representation: Some features are extracted for each object (in this paper, each song) to represent the object;
2. Profile Learning : The feature data of several objects that the user likes is used as the user's preference feature;
3. Recommendation Generation : By comparing the characteristics of users and candidate objects obtained in the previous step, Recommend the most relevant objects for users. The algorithm flow chart applied in this paper is shown in Figure 4:

DOI:10.46647/ijetms.2023.v07si01.088 ISSN: 2581-4621

2.2.2 Feature Vector Extraction of Songs The user selects three typical songs corresponding to each emotion in the song library. Organize the feature labels of the three songs and record them without repetition as：

A1, A2, A3, ..., An, record the occurrence times of each feature as： a1, a2, a3, ..., an, then get the feature vectors of the user:

[(A1, a1), (A2, a2), (A3, a3), …, (An, an)], the feature vectors of any other song：

[(A1, b1), (A2, b2), (A3, b3), …, (An, bn)]

3. Conclusion In this paper, we proposed a model of facial microe-xpression recognition based on convolutional neural network (CNN). After training the model on FER2013 data set, we got a recognition rate of 62.1%. On the basis of the state that facial expression and emotion were both recognized, the content-based recommendation algorithm was applied to automatically recommend music for users. Compared with the existing algorithms that only recommend music according to the users' past listening preferences, the algorithm proposed in this paper increases the user's emotion recognition, so that the recommended music can better meet the users' listening needs. Therefore this algorithm has a relevantly promising application market. Although we have made some achievements, still some problems need to be solved . For example, the accuracy of microexpression emotion recognition needs to be improved. In the follow-up work, the recognition rate of tags with low recognition will be improved, and the music recommendation algorithm will be further optimized and improved

4. References:

[1] Liu Jianwei, Liu Yuan, Luo Xionglin. Progress in Deep Learning Research [J]. Application Research of Computers, 2014, 31 (7): 1921-1942.
[2] Shen Huijun. Research and implementation of face recognition image preprocessing method [J]. Science and Technology and Innovation, 2014 (18): 119-120.
[3] Zhang Chen. Research on some key technologies of facial micro-expression recognition [D]. 2019.
[4] Liu Mingqi, Ni Guoqiang, Chen Xiaomei. Research on Pretreatment Algorithm of Dorsal Vein Image [J]. Optics Technology, 2007, 33: 255-256.
[5] Li Siquan, Zhang Xuanxiong. Research on Facial Expression Recognition Based on Convolutional Neural Networks [J]. Journal of Software, 2018, v.17; No.183 (01): 32-35.
[6] Hou Yuqingyang, Quan Jicheng, Wang Hongwei. Overview of the development of deep learning [J]. Ship Electronic Engineering, 2017, 4: 5-9.
[7] Liu Sijia, Chen Zhikun, Wang Fubin, et al. Multi-angle face recognition based on convolutional neural network [J]. Journal of North China University of Technology (Natural Science Edition), 2019, 41 (4): 103-108.
[8] Li Huihui. Research on facial expression recognition based on cognitive machine learning [D]. Guangzhou: South China University of Technology, 2019.
[9] Li Yong, Lin Xiaozhu, Jiang Mengying. Facial expression recognition based on cross-connection LeNet-5 network [J]. Journal of Automation, 2018,44 (1): 176-182.
[10] Yao L S, Xu G M, Zhap F. Facial Expression Recognition Based on CNN Local Feature Fusion[J]. Laser and Optoelectronics Progress, 2020, 57(03): 032501.
[11] Xie S, Hu H. Facial expression recognition with FRR-CNN [J]. Electronics Letters, 2017, 53 (4): 235-237.
[12] Zou Jiancheng, Deng Hao. An automatic facial expression recognition method based on convolutional neural network [J]. Journal of North China University of Technology, 2019,31 (5): 51-56.