Human Emotion Detection and Music Player using Machine Learning

Vaishnavi Borgaonkar, PG Student, Maharashtra Institute of Technology college of engineering, Aurangabad, India,

Prof. Sarita Sanap, Assistant Professor, Maharashtra Institute of Technology college of engineering, Aurangabad, India,

---

Abstract - The model presents the design, implementation, test, and evaluation of a Facial Expression Recognition (FER) and automatic music recommendation system that applies a machine learning algorithm based on Convolutional Neural Networks (CNNs) with the aim of correctly classifying four facial expressions (namely happiness, sadness, anger, and neutral). The CNN module and the FER system were built in python.

Facial expressions convey helpful information that is difficult to detect for an ordinary system. However, being capable of recognizing them could lead to more responsive and intelligent systems that might improve the user experience. By experimenting with Fisherface model and proposed CNN model on facial Cohn-Kanade dataset, the most suitable hyper parameters that yielded a good level of performance were obtained. Additionally, a deep understanding of the strengths and limitations of CNNs was gained.

Results from these experiments show that special care must be taken during different parts of the development process such as architecture selection or hyper parameter tuning. By selecting the correct combination between these two elements, the accuracy of the model and the convergence time improve. The training accuracy for the proposed CNN model is 72.34% which is very similar to the pretrained AlexNet CNN model which is 72.77%. In a Real-Time testing Fisherface is used to detect the face for further facial expression classification

1. INTRODUCTION

Human-Machine Interaction is an emerging field in the latest growing technology of Artificial Intelligence. Like human beings, machine should also have capability to read human expression and then responding according to the requirements. According to the various surveys, there are two basic types of communication include verbal and non-verbal communication. One-third part of communication include verbal and two- third is non-verbal communication. Among various non-verbal components, facial expressions are one of the main information channels in interpersonal communication. Therefore, research in facial expressions is gaining lot of attentions over past decades, not only in the perceptual and cognitive science but also in affective computing and computer animations. The most effective way to recognize the human emotion is to analyse the human facial expressions. The facial expressions contain lot of information of human's emotions or mood [1]. Human can easily understand the expressions of human but to develop accurately this ability in machine is a challenging task. This recognition technique can also be used as a component of Human-Machine Interaction (HMI). Artificial Intelligence will make task easy for Human-Machine Interaction using Facial Expressions.

Image Processing in artificial intelligence is a challenging research area in the field of computer vision. Facial expressions are the fundamental way to expresses the human emotions [3]. The applications of FER are in the field of crime control, health care, virtual education, e-commerce, entertainment, etc. Researchers are still facing problems to behave machine like humans. Even though many algorithms are developed to recognize facial expressions and number of them are applied in machines but there are still gaps to improve accuracy, computational speed and power in algorithm.

At present lot of work is going on in FER domain. Many groups of researchers are having their hands on this domain. For example, Harris corner key point detection technique, Wavelet transform technique, Eigen vectors detection, Landmark or spot detection techniques, Neural Network, Ensemble of Cascaded Regression, Explicit Shape Regression for landmark detection, Project-Out Cascaded Regression for Face Alignment, Principal Component Analysis etc. are used by some researchers. All these techniques were used for Feature Extraction purpose. In combination with all these techniques they have also used various classifiers in it. They have used KNN, SVM etc. classifiers in it. All these researchers sometimes have used combination of some feature extraction techniques and classifiers technique. Due to which it then achieves required output of their approaches.

Latest researchers use the Deep Neural Network (DNN) Domains Especially Convolutional Neural Network (CNN). In this research we have implemented CNN for feature extraction and classification using Adam optimizers. In this paper, implemented facial expression recognition in real time via webcam. We also visualized the activations of hidden layers in convolutional neural network and depending upon the mood of user music will be recommend.

2. LITERATURE SURVEY

2.1 Common Approach for Facial Expression Recognition

The picture depicts the basic structure of a typical approach to facial expression recognition. It consists of five main steps: image acquisition, preprocessing, feature extraction, classification, and postprocessing, which produces the classified face expression's final output.

Figure 2.1: Basic flow diagram for facial expression recognition system

The first phase is picture acquisition, which is responsible for obtaining a video or static image sequence with a different frame of reference for the face. Although image acquisition can be done in two or three dimensions, the sequence of photos yields better results, such as the spatial and temporal aspects of a face expression [1].

Facial Expression Recognition is a difficult problem in general due to the following factors: segmentation difficulty, distortion, illumination, quality, and views. Due to the disarray of the foreground and background objects in real-life photographs, segmentation is difficult. In a real-life image, the distortions are due to optical aberration, which causes physical straight lines in an image to look as curve lines. The source of light that impacts the pixel intensity is directly tied to illumination. Challenges arise as a result of the poor quality of photos obtained from various sources, such as cameras. It also depends on the object class's physical shape. Viewpoints, or diverse frames of reference, that are unable to cope with learning processes [2].

Modification of the image is one of the pre-processing procedures. Scaling, rotation, translation, and normalisation are examples of modifications. One of the approaches for solving the illumination problem is histogram equalizations. The gaussian mixture model and deformable models are employed to solve image segmentation difficulties [1].

2.2 Neural Networks

Neural Networks are a collection of algorithms that attempt to recognise underlying correlation in a piece of data using a method that resembles how the human brain functions [5]. The computational units are referred to as neurons in this context. These networks' basis functions are similar to equation (1), with each basis function being a nonlinear function from a set of input variables to a set of output variables controlled by the vector w of adjustable parameters [5]. According to [4] while the support vector machine has the same generalisation performance as the neural network models, the neural network models are substantially more compact and hence have a faster evaluation time.

3. SYSTEM DEVELOPMENT

3.1 Machine Learning

3.1.1 Fisherface ML algorithm

Fisherface algorithm is an algorithm which work on the basis of LDA and PCA concepts. Linear discriminant analysis (LDA) is a supervised Learning method of machine learning. Now supervised Learning is that where we use such data whose answer is also given to the model to learn it. It work on the concept of dimensionality reduction. Which reduce the execution time among classification.

3.1.2 Principal Component Analysis (PCA) is a one kind of conversion from correlated variables to uncorrelated in the form of mathematical values. It is mostly used for the observing data and from that by some probabilistic calculation generate models. The flow of Fisherface is like it takes classified images then it will reduce the dimension of the data and by calculating it's statistical value according the given categories it stores numeric values in .xml file. While prediction it also calculate the same for given image and compare the value with the computed dataset values and give according result with confidence value.

3.1.3 Resizing images

Whatever the image we have chosen for dataset it mostly related to the size which can give an precise output. The size is chosen such like the model can able to easily distinguish face from image by haarcascade model. And the size what we get from real time scan is not always same as data (very less difference) so, We resize it to the exact model data size. In our case we have chosen 350*350.

Here In this method, we have implemented the cropping of image by given parameters of haarcascade by clahe_image[] and use of cv2's method .resize() to the given size. Finally, We have stored those images in dictionary and after some count(=10) take it to check result.

3.1.4 Gray scaling images

It was the need for the method and because of it's contrast and shaded face, it result in benefit for algorithm to get output.

3.1.5 Face detection

As the given in the code grab_face() methods uses to get the images and do all operation and finally return cropped ,grayed face value in dictionary.

1.1.6 Train and predict methods

This section is use to get prediction and confidence value for given amount of image. Then get the max function with obtained output and final result is shown to the user.

3.2 Playing music

3.2.1 Detected emotions

We have implemented the linking of python with javascript through eel library. Which provide us the privilege to access python methods from js as well as vice versa. Here the striating flow will be in python code as the library is implemented in python then it transfer the control to html, JS. And according to the result we show emoticons.

Sad happy angry neutral

According to which we can classify emotion directory for playing song we have chosen this 4 emotions.

3.2.2Methods for playing songs

In JavaScript file we have implemented too much methods for the switching of song.

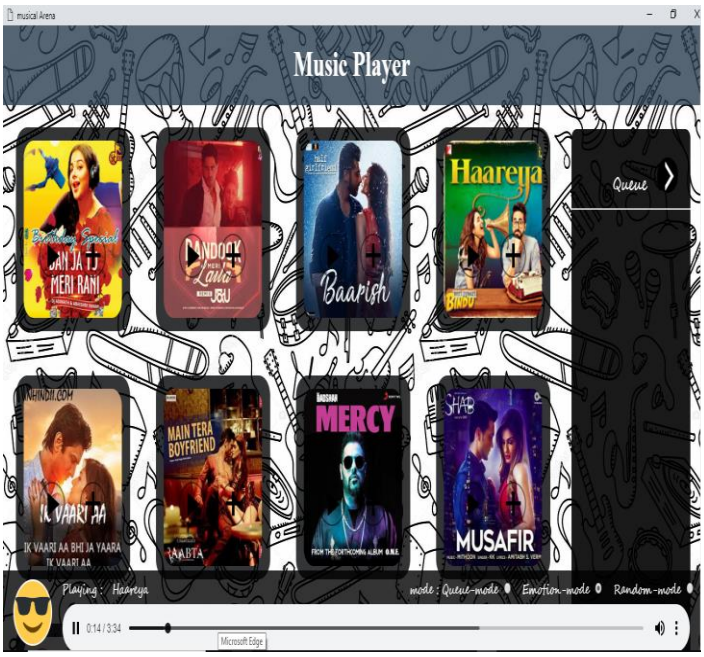
1. Queue
2. Based on Emotion
3. Random

In the first one as queue works it has been implemented. In

second one we call python code to get emotion from user's facial expression and according to that chosen next song which is also randomly and played it. In third one we directly used random function and all the methods are dynamic it can handle as change in number of songs accordingly.

3.2.3 HTML, CSS and JS concepts for online music player.

As we know the css give a great look to communicate and through JS we can interact with user and not look like complicated program run at console and it also give user privilege to choose any song to play.



4. RESULTS

In proposed model, model not stick on one image for testing, While the code will run it will take around 10 images in a short time(1-2 sec) and for all those images it will compute result and according to the average value of that it will give result. Apart from that system have make two codes one work on single face at a time while another work with multiple faces in the image.

After testing model with various users, the accuracy of the proposed system is improved as compared with other models. In the table no. 1, actual emotion and facial emotion results by the system for the 5 users are shown.

User	Actual Emotion	Facial Expression Detected	Accuracy
	Sad	Sad	100
	Happy	Happy	100
	Angry	Neutral	0
	Neutral	Angry	0
	Sad	Sad	100
	Happy	Happy	100
	Angry	Sad	0
	Neutral	Neutral	100
	Sad	Sad	100
	Happy	Happy	100
	Angry	Angry	100
	Neutral	Happy	0
	Sad	Angry	0
	Happy	Happy	100
	Angry	Angry	100

	Neutral	Neutral	100
	Sad	Sad	100
	Happy	Happy	100
	Angry	Angry	100
	Neutral	Neutral	100

5. CONCLUSION

After testing system with various users, here most of the time original emotion get match but sometimes system prediction gets failed. But from the table 2, can conclude that; system have 75% accuracy and accuracy will increase as day by day.

6. REFERENCES

1. Chibelushi, C. and Bourel, F. (2016). Facial Expression Recognition: A Brief Tutorial Overview.
2. Hinton, G. (2012). *Neural Networks for Machine Learning*. [online] Coursera. Available at: https://www.coursera.org/course/neuralnets [Accessed 8 Mar. 2016].
3. Y. Lv, Z. Feng and C. Xu, "Facial expression recognition via deep learning," Smart Computing (SMARTCOMP), 2014 International Conference on, Hong Kong, 2014, pp. 303-308. doi: 10.1109/SMARTCOMP.2014.7043872.
4. Bishop, C. (2006). *Pattern recognition and machine learning*. New York: Springer,pp.227 - 249 and 256 - 272.
5. Cs231n.github.io. (2016). *CS231n Convolutional Neural Networks for Visual Recognition*. [online] Available at: http://cs231n.github.io/convolutional-networks/ [Accessed 15 Apr. 2016].
6. Murphy, K. (2012). *Machine learning*. Cambridge, Mass.: MIT Press, pp. 563 – 579.
7. Cs231n.github.io. (2016). CS231n Convolutional Neural Networks for Visual Recognition. [online] Available at: http://cs231n.github.io/neural-networks-1/ [Accessed 26 Jul. 2016].
8. Poczos, B. and Singh, A. (2016). Introduction to Machine Learning. Deep Learning. CMU.
9. Goodfellow, I., Warde-Farley, D., Mirza, M., Courville, A. and Bengio, Y. (2013). Maxout Networks. JMLR WCP 28, pp.1319-1327. arXiv:1302.4389 [stat.ML].
10. Ufldl.stanford.edu. (2016). Unsupervised Feature Learning and Deep Learning Tutorial. [online] Available at: http://ufldl.stanford.edu/tutorial/supervised/ConvolutionalNeuralNetwork/ [Accessed 15 Apr. 2016].
11. Sebastian Ruder. (2016). An overview of gradient descent optimization algorithms. [online] Available at: http://sebastianruder.com/optimizing-gradient-descent/index.html [Accessed 31 Jul. 2016].
12. Zeiler, M. (2016). Adadelta: An adaptive learning rate method. Google Inc., USA - New York University, USA. arXiv:1212.5701 [cs.LG].
13. Hinton, G. (2016). Lecture 6e. rmsprop: Divide the gradient by a running average of its recent magnitude. [online] Available at: http://www.cs.toronto.edu/~tijmen /csc321/slides/ lecture_slides_lec6.pdf [Accessed 02 Aug. 2016].
14. Cs231n.github.io. (2016). CS231n Convolutional Neural Networks for Visual Recognition. [online] Available at: http://cs231n.github.io/neural-networks-3/ [Accessed 31 Jul. 2016