

YOUTUBE VIDEO RECOMMENDATION BASED ON FACIAL EXPRESSION

Krishna A N ¹, Vedavathi B Shriyan ², Vipthi V Jain ³, Yashaswini Dinesh ⁴, Yukthi M Bhat ⁵

Department of Computer Science & Engineering, SJB Institute of Technology, Bengaluru 560-060

Abstract - Human faces play a significant role in determining a person's mood. Emotion plays an important role in various fields, including biomedical engineering, brain science, neuroscience, psychological wellness, and mental health. Emotions are not only used for determination of the state of the human brain but too utilized as a proposal framework to help people in finding things that coordinate their needs and inclinations. This motivated us to create a framework that can viably and proficiently recognize feelings from the facial expressions of the user and recommend the YouTube videos based on that emotion. Emotions are bought by the neurophysiological changes that are associated with thoughts, feelings and behavioral changes in a person. Every emotion needs to be treated in a right way and research suggests that watching YouTube videos influences your mood, and in our experiment, we try to influence it in a good way. We can detect changes in emotion by observing facial expressions, body language, and tone of voice. There are many indicators of a person's mood, but facial expressions are one of the most important. In our experiment, we used the Haar Cascade Classifier to detect the face and a convolution neural network to analyze the facial expression. Performance of our proposed model is 95.66% for face detection and 61.88% for emotion detection.

Key Words: Emotion Detection, FER-2013, HAAR Cascade, YouTube Recommendation, CNN.

1. INTRODUCTION

Recognition of facial expressions is a technique used to identify the most basic human emotions. Expressions convey essential information about human emotions. Programs and systems focusing on the interaction of visuals may play a key part in the future generation of computer vision systems. Happiness, sadness, anger, surprise and neutral emotions are the most frequently detected emotions [1]. Analyzing the facial expressions of a person allows one to detect these basic emotions. A per-son's facial expressions are a way of trying to communicate their emotions. Today increase in technology for digital signal processing and other effective feature extraction algorithms, automated emotional detection in multi-media attributes like music or movies is growing rapidly and this system can play an important role in many potential applications like human-computer interaction systems and entertainment [2]. The suggested system analyses a person's emotions; if the person is experiencing a negative emotion, a set of YouTube videos will be displayed that includes the most relevant genres of videos to help him feel better. If the emotion

is positive, the user will be directed to a set of YouTube videos, which contains a variety of videos that will amplify the pleasant feelings. Haar-like features are applied on real-time images to detect faces using Haar cascade classifiers. For emotion detection, an 8-layer Convolutional Neural Network is used. The emotion that is detected either happy, sad, anger, surprise, neutral, dis-gust, fear is taken as a keyword for recommending the YouTube videos.

Problem Statement: Develop a system that detects the emotion in the real time and gives suggestions of YouTube videos.

2. BACKGROUND

As of late computer vision has upraised the level of advancement within the field of face detection and framework and face-expression acknowledgement to bolster automatic emotion recognition framework. The detection of faces based on skin color is a widely used and useful technique. Using this method, each pixel is classified as skin or non-skin based on its color composition. Although this method works well, it has some drawbacks. Images are often scaled down due to a number of factors in order to reduce computational time. This can affect the overall result. This method also uses an RGB classifier, which is much slower than the new generation classifiers [3]. In the field of emotion detection, many researchers have progressed with many different approaches leading to different results for face expression analysis. Raghuvanshi and Choksi [4] tested Deep CNN architectures and methods, such as fractional max pooling and finetuning, and achieved 48% accuracy on Kaggle's facial expression recognition challenge dataset. Application proposed by Dusi, Sriharshini, Kommuri Krishna Teja, Guntupalli Manoj Kumar [5] give results on image and video files to accurately distinguish between the seven emotions. Their fitted MTCNN model generated the results more accurately than the pre-trained model that exists in deep learning. The results of the emotion recognition algorithm showed that emotions can be identified with up to 60% accuracy. Saravanan, Akash, Gurudutt Perichetla, and Dr KS Gayathri [6] conducted experiments with different methods which include decision trees, feed forward networks and smaller convolutional networks to arrive at their proposed model. The accuracy of 60% was obtained by usage of the Adam optimizer with changed hyperparameters. After going through few papers, we found that Neural Networks are the finest as they are less complex compared to other strategies.

3. METHODOLOGY

The purpose of the system benefits us to perform interaction between the user and the YouTube application. The proposed system is to capture the face properly with the web camera. Real-time Captured images are fed to Haar-cascade classifier to obtain face in the image. This face is then passed to Convolutional Neural Network which predicts the emotion. Then emotion derived from the captured image is used to get a recommendation. YouTube video recommendation based on facial expression system design is shown in Fig.1.

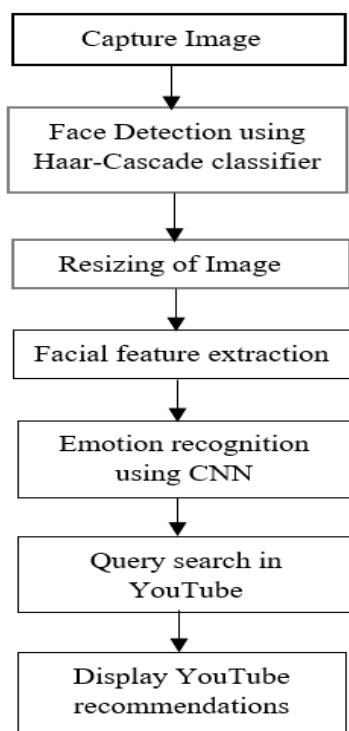


Fig.1. System design

3.1 Dataset:

The dataset used is the FER2013 dataset. This dataset consists of a huge amount of face images which are grayscale images of 48x48 pixels. The total number of images is 35,887. Different types of emotion and number of images used for experiment is given in Table.1. The sample images of FER 2013 dataset are shown in Fig.2.

Emotion	Number of Images
Angry	4953
Disgust	547
Fear	5121
Happy	8959
Neutral	6198
Sad	6077
Surprise	4022

Table.1 Number of images of each emotion.



Fig.2. FER2013 dataset sample images.

3.2 Face Detection:

We are using a webcam to capture continuous images since the primary objective of this session is to obtain real-time camera feed. It starts the execution process by accessing the camera stream and capturing using Open cv library. For face detection, Haar Cascade is used. It is an object detection algorithm used to identify faces in images or real-time videos. This algorithm uses the edge or line detection function proposed in the research paper by Viola and Jones [1]. Based on this approach, a face detection system was developed that is approximately 15 times faster than prior approaches. Object Detection Algorithms such as Haar-Cascade are used to identify faces in images or videos. This detection can be real-time from a video frame or images. The algorithm detects what's either a face (1) or not a face (0) in an image. The Haar-cascade classifier identifies the coordinates of the face found in the image and then the coordinates of that face is passed to the emotion detection module.

3.3 Emotion Detection:

The image obtained from the Haar-cascade classifier is converted to grayscale and then passed to the convolution neural layers. A con-volutional neural network is a type of neural network made up of convolutional layers that lift data in a computationally costly way Convolutions are performed. Convolution is a two-dimensional math operation. To make a third function, combine two functions. As for what the computer sees, the image is simply a matrix of numbers. Using both fully connected layers and convolution layers, convolution operations are done on these integers. All nodes are connected to all other neurons in a completely connected layer. In feedforward neural networks, these are the layers that are used. The convolutional layer is not connected to each neuron, unlike the fully connected layer. Between the discrete locations, a connection is established. A movable "window" moves on the image. The size of this window is called the kernel or filter. These filters help in identifying patterns in the data.

- eight convolutional layers using "RELU" as an activation function.
- four max-pooling using pool size (3,3).
- seven drop out with value 0.25 .
- one flattened layer and four dense layers: three dense layer with 'RELU' and the other with 'SoftMax' as an activation function.
- total parameters and trainable parameters are 14,376,711 and 14,368,135 million, respectively.

3.4 YouTube Recommendation:

The web browser module in python provides a high-level interface for users to view web-based documents. The open function of web browser module opens the URL passed to that function. The queries made for respective expression are shown in Table.2.

Table.2 YouTube query for different expressions.

Emotion	Query Search
Happy	Happy songs
Angry	Anger Management
Sad	Stand-up comedy
Fear	Anxiety relief
Surprise	Relaxing songs
Neutral	Art and crafts
Disgust	Motivational video

4. RESULTS

The OpenCV frontal-face Haar-cascade classifier was tested for a set of 50 images consisting of various number and orientation of faces. It was observed to give an accuracy of 95.66%. Some of the test results are as follows:

Table.3 Test results for Haar-cascade classifier

No. of Faces in an image	No. of Faces detected	Accuracy %
1	1	100
2	2	100
3	3	100
4	4	100
5	5	100
6	4	66.66
7	7	100

The networks acquired for emotion detection are robust in nature, operate at real-time speeds, and can be integrated with the real world. This eight-layer CNN architecture has achieved 61.88% test accuracy and 76.68% of training accuracy on FER2013 dataset with 35,887 for 7 emotions without the need

for preprocessing or feature extraction techniques. Test accuracy of 70.4% accuracy was obtained for 5 emotions excluding fear and sad. The learning rate and number of epochs of the model are 0.001 and 20 respectively. Fig.3 shows Loss graph for training and validation data of CNN model and Fig.4.

shows Accuracy graph for training and validation data of CNN model.

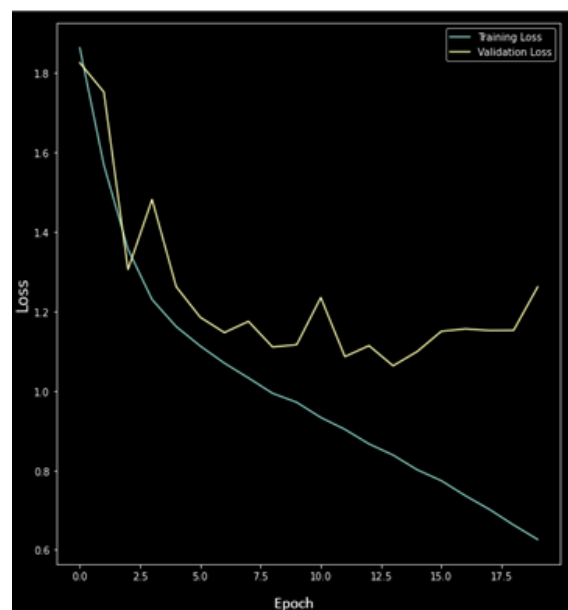


Fig.3. Loss graph for training and validation data of CNN model

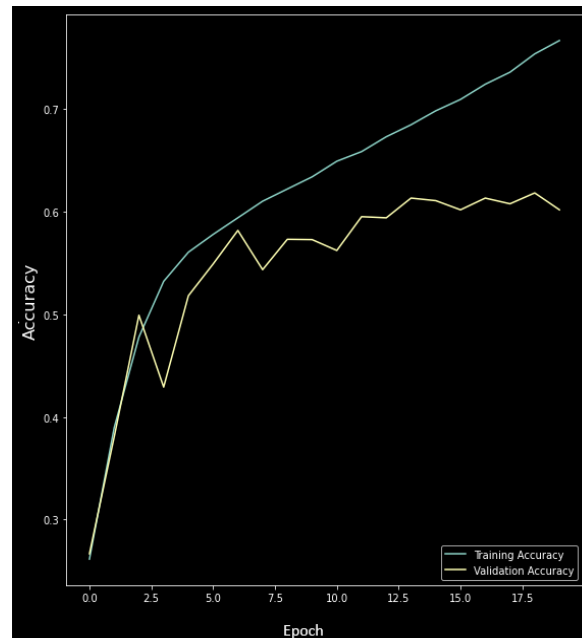


Fig.4. Accuracy graph for training and validation data of CNN mode

Table.4. Comparison with previously available emotion recognition models.

Model Name	Training Accuracy %	Testing Accuracy %
Deeper CNN [4]	60	48
CNN model [5]	72.5	-
CNN model [6]	-	60
This paper	76.68	61.88

As in the Table.4, it shows that our proposed CNN model has performed comparatively better than the previously available emotion recognition models.

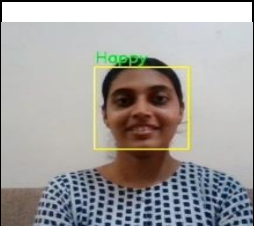
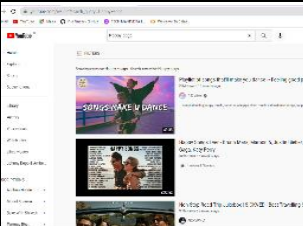
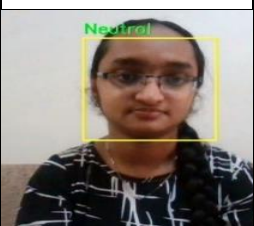
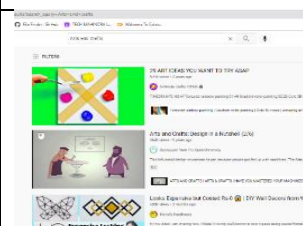
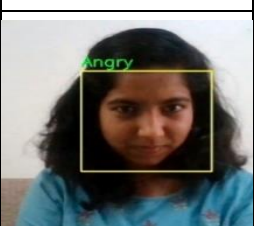
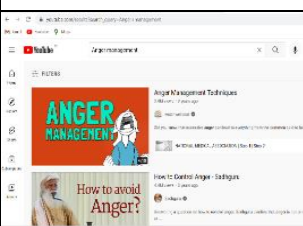
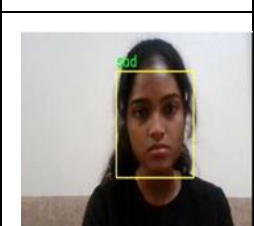
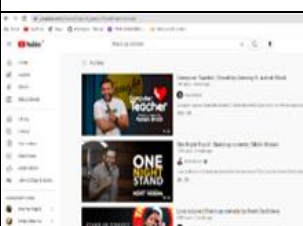
Type of Expression	Face + Emotion Detection	Recommendation
Happy		
Neutral		
Angry		
Sad		

Fig.5 Videos recommended in YouTube for corresponding expression.

Figure.5 shows the input and output of the YouTube video recommendation based on facial expression model for various expressions.

CONCLUSION AND FUTURE ENHANCEMENTS

This paper presents a method that uses a Haar-cascade object detection algorithm to detect face, a customized CNN model with 8 convolution layers to recognize the emotion and recommend YouTube videos in real-time. It is designed for the purpose of enhancing the interaction between YouTube and the user, since videos, songs, and motivational speakers are helpful for changing a user's mood, and for some people, it is a stress reliever. So the presented system presents a Face (expressions)-based recognition system so that it can detect emotions and recommend videos based on those. Performance of our proposed model is 95.66% for face detection and 61.88% for emotion detection. Although the system is fully functional, it can be improved in the future. Future work should attempt to combine our technique with other modalities such as audio modality, including working with other data sets.

REFERENCES

1. Viola, Paul, and Michael Jones. "Rapid object detection using a boosted cascade of simple features." *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*. Vol. 1. IEEE, 2001.
2. Singh, Shekhar, and Fatma Nasoz. "Facial expression recognition with convolutional neural networks." *2020 10th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, 2020.
3. Minu, M. S., et al. "Face recognition system based on haar cascade classifier." *International Journal of Advanced Science and Technology* 29.5 (2020): 3799
4. Raghuvanshi, Arushi, and Vivek Choksi. "Facial expression recognition with convolutional neural networks." *CS231n Course Projects* 362 (2016).
5. Dusi, Sriharshini, Kommuri Krishna Teja, and Guntupalli Manoj Kumar. "REAL-TIME FACIAL EXPRESSION AND EMOTION RECOGNITION FOR RECOMMENDATION OF YOUTUBE VIDEOS." *Turkish Journal of Physiotherapy and Rehabilitation* 32: 2.
6. Saravanan, Akash, Gurudutt Perichetla, and Dr KS Gayathri. "Facial emotion recognition using convolutional neural networks." *arXiv preprint arXiv:1910.05602* (2019).