Facial expression (emotion) analysis

Mohit Agarwal
San Diego State University
San Diego, United States
magarwal8188@sdsu.edu
GitHub Link

Abstract

Emotion recognition is a popular research topic. It is used in a variety of applications. Robotic vision and interactive robotic communication are two of its most intriguing uses. The face is one of the most important nonverbal communication channels. Facial expression, among other things, conveys emotion, intention, awareness, and pain, governs interpersonal behavior, and transmits mental and biological state. Both verbal and visual modalities can be used to identify human emotions. Facial expressions are an excellent method for identifying a person's emotions. This study describes a real-time technique to emotion recognition. The proposed approach consists of two phases: face detection and emotion classification. The main idea is to detect the face using MediaPipe face detection algorithm, which is based on real-time deep learning. Then, using a pretrained deep convolutional neural networks model, this research seeks to identify the emotion on a person's face into one of seven categories.

Introduction

Facial expression analysis refers to computer systems that attempt to automatically analyze and recognize facial motions and facial feature changes from visual information. Our faces are one of the most powerful markers of our emotions; whether we are laughing or sobbing, our feelings are visible to others. Our expressions mirror our emotions.

In human-machine interaction, an automatic face expression detection system is very crucial. This, however, is not a simple task. Face acquisition, facial feature extraction, and classifier development are usually the three stages of facial expression recognition training.

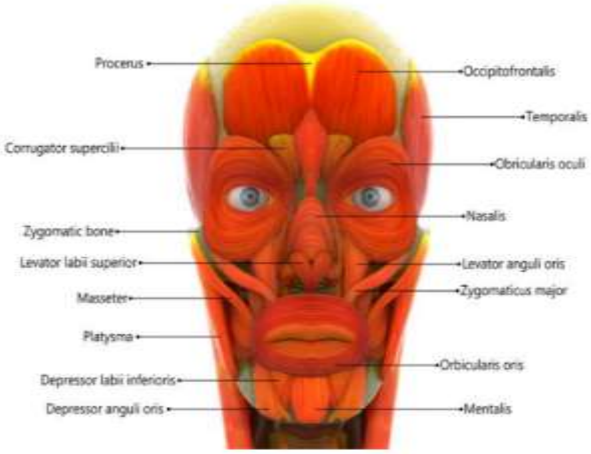

Fig1. forty-three face muscles that regulate our facial expressions

Our system for facial expression recognition performs summarized as follows:

1. MediaPipe Face Detection algorithm is used for Face Detection on the video file input.
2. The bounded box for the detect face is extracted and the corresponding frame area is scaled to 48x48 and fed into the pre-trained model.
3. The model generates a list of SoftMax scores for each of the seven emotion classes.
4. The emotion with the highest score is shown on the screen.

The remainder of this work is structured as follows. Section 2 discusses some related work. Section 3 discuses about the dataset used for training of the facial expression detection model. Furthermore, Section 4 discusses about the MediaPipe open-source model which is being used for face detection in the video file. Section 5 discusses about the model used for face emotion recognition, and Section 6 provides simulation results. Section 7 contains the paper's final observations and conclusions. Finally, Section 8 mentions all the references used for this work.

II Related Work

Several facial expression recognition algorithms have been developed in recent decades, with improved recognition performance. Researchers have made significant progress in constructing automated expression classifiers in recent years [3]. Some expression recognition systems divide the face into emotions like happiness, sorrow, and anger. [4]. Others seek to identify particular muscle movements produced by the face in order to offer an objective description of the face.

Song et al. [5] created a face emotion detection system that operates on a smartphone and using a deep Convolutional Neural Network. There are five layers and 65,000 neurons in the proposed network. According to the scientists, overfitting is typical when utilizing a limited quantity of training data and such a large network.

III Facial Expression Dataset

The FER-2013 database was introduced during the ICML 2013 Challenges in Representation Learning. It was used to train the emotion recognition model. It includes both posed and unposed grayscale headshots with a resolution of 48x48 pixels. The dataset was built by compiling the results of each emotion's Google image search as well as synonyms. There are around 30,000 facially identified photos in the dataset. In FER-2013, each image is assigned to one of seven emotions: happy, sad, angry, afraid, surprise, disgust, and neutral, with happy being the most prevalent emotion, providing a baseline for random guessing of 24.4%.

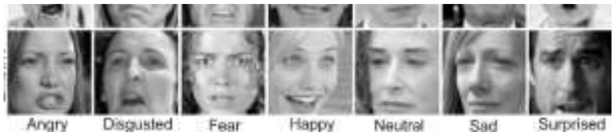

Fig2. Example of FER-2013 Dataset

IV MediaPipe

MediaPipe [2] Face Detection is a lightning-fast face detection system with six landmarks and multi-face capability. It is built on BlazeFace [1] a lightweight and high-performance face detector optimized for mobile GPU inference. Because of the algorithm's super-real-time speed, it may be used in any live experience that demands an accurate facial region of interest as an input for other task-specific models. We can configure the model using certain parameters such as MODEL_SELECTION and MIN_DETECTION_CONFIDENCE.

MODEL_SELECTION

We can set this parameter to either 0 or 1. Use 0 to choose a short-range model that works best for faces within two meters of the camera, and 1 to choose a full-range model that works best for faces within five meters of the camera.

MIN_DETECTION_CONFIDENCE

We can set this parameter within the range of ([0.0, 1.0]). Each face detected is represented as a detection proto message with a bounding box and six key points in a collection of detected faces (right eye, left eye, nose tip, mouth center, right ear tragion, and left ear tragion). The bounding box is made up of x_{min} and width (both normalized to [0.0, 1.0] by the image width) and y_{min} and height (both normalized to [0.0, 1.0] by the image height) (both normalized to [0.0, 1.0] by the image height). Each key point is made up of x and y values that have been normalized to [0.0, 1.0] by the image width and height.

V Proposed Model

The main aim of this project was to analyze the expression (emotion) of all faces that have appeared in a video stream and if multiple faces exist, we needed to identify and keep track of each one.

So, to achieve this, our first goal was to detect the face in the video stream. We used the MediaPipe face detection algorithm. To detect the face in a video file we needed to set certain paraments in the model such MODEL_SELECTION and MIN_DETECTION_CONFIDENCE. In our case we set MODEL_SELECTION to 1 because the faces were within five meters of camera and MIN_DETECTION_CONFIDENCE parameter was set to 0.5, which implies that if a face predicted by the model has a probability greater than or equal to 50%, it is saved as a detected face.

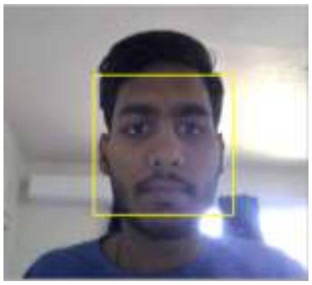

Fig3. Face Detection using MediaPipe

Next, we needed to identify the emotions for the faces detected by the face detection algorithm. For this we used a pre-trained deep neural network model.

V.A Deep Neural Network

A deep neural network (DNN) is a neural network with more than two layers. Deep neural networks use advanced mathematical models to evaluate data in complicated ways. Deep Neural Networks recognize data relationships (from simple to complicated). Consider image recognition as an example. The first hidden layer may be looking for simple operations like identifying the edges in the given picture. As we get further into the network, these simple functions combine to produce more complex jobs, such as discovering more features in a picture to make it more complete.

The last phase of the model is to categorize the provided face into one of the fundamental emotion categories after learning the deep characteristics. Unlike classical approaches, where the feature extraction and classification steps are separate, deep networks may execute from start to finish. A loss layer is added to the network's end to limit the back-propagation error; then, the network may directly output the prediction probability of each sample. The most often used function in CNN is SoftMax loss, which minimizes the cross entropy between the predicted class probabilities and true values.

The pre-trained model was trained on FER-2013 dataset as mentioned above. For the prediction of the emotion of the faces, the frames were first converted to grayscale and were scales to 48x48 pixels as the model was trained on grayscale images. Next, only grayscale region of the predicted face is fed to model for emotion prediction. The model generates a list of SoftMax scores for each of the seven emotion classes. The emotion with the highest score is shown on the screen.

A sample of input image and the corresponding SoftMax scores for all the seven categories has been shown below.

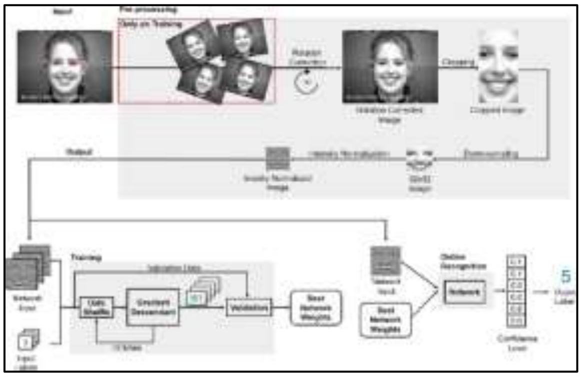

Fig: Complete process for Face Emotion Recognition [6]

VI. Results

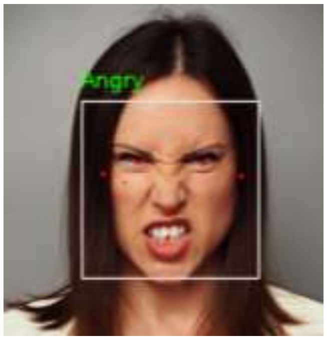

Fig4: Input Image

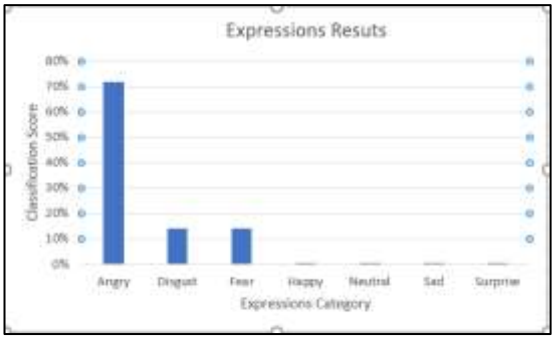

Fig5: Expressions results for the above input image

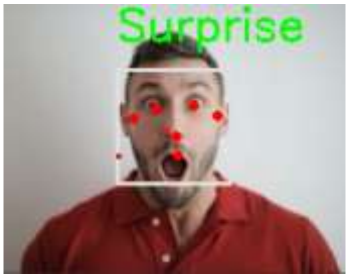

Fig6: Input Image -2

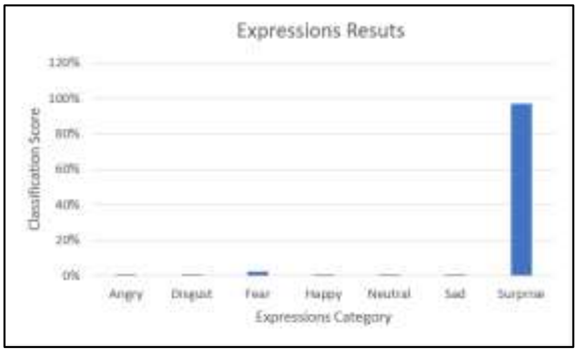

Fig7: Expressions results for the above input image

Finally, a video file was used as an input for the facial emotion analysis, and the outcome of a frame from the video file is displayed below.

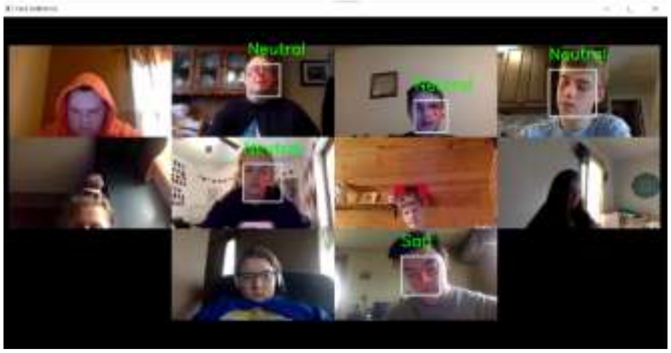

Fig 8. Example of Facial expression Detection on a frame of Video Input File

VII Observation and Conclusion

In this paper, we proposed a pre-trained deep convolutional neural network based facial expression recognition method, with an open-source face detection model called MediaPipe.

Our proposed method achieves good results on the input video file dataset, indicating the considerable potential of our facial expression recognition method. One of the limitations of our model is when there are more faces in a frame our model doesn't performs well because then the faces are more than 5meters apart from the camera and the MediaPipe can detect the faces well which are within 5meters of the camera. As we can see the results shown above that there are 10 faces in the frame, but our model can detect only 5 faces out of them.

Additionally, one of the limitations is our model can't detect any faces for certain frames because in those frames the faces are too far from the camera. One of the frames for which we can't detect any faces is shown below.

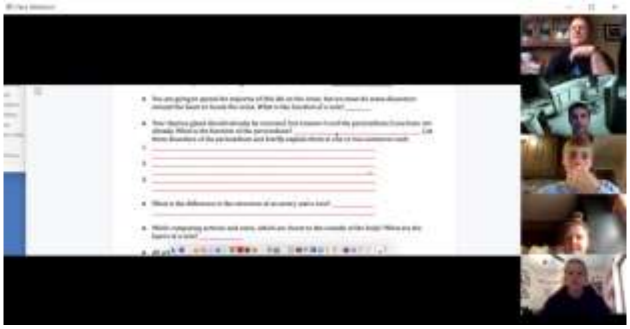

Fig9: No face detected for this frame.

VIII References

[1]. https://arxiv.org/abs/1907.05047

[2].https://google.github.io/mediapipe/solutions/face_detection.html

[3]. Y. Tian, T. Kanade, and J. Cohn. Recognizing action units for facial expression analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(2), 2001.

[4]. G. Littlewort, M. Bartlett, I. Fasel, J. Susskind, and J. Movellan. Dynamics of facial expression extracted automatically from video. Image and Vision Computing, 24(6), 2006.

[5]. I. Song, H.-J. Kim, P.B. Jeon, Deep learning for real-time robust facial expression recognition on a smartphone, in: International Conference on Consumer Electronics (ICCE), Institute of Electrical & Electronics Engineers (IEEE), Las Vegas, NV, USA, 2014.

[6]. Facial Expression Recognition with Convolutional Neural Networks: Coping with Few Data and the Training Sample Order: Andre Teixeira Lopesa, Edilson de Aguiarb, Alberto F. De Souzaa, Thiago Oliveira-Santosa