Section A-Research paper
ISSN 2063-5346

Facial Expression Detection Using CNN with MediaPipe

Mrs.V.Lavanya[1], Mr.K.Azarudeen[2], Dr.G.Vinoth Chakkaravarthy[3], Mrs.CB.Selvalakshmi [4]

[1]Assistant Professor, Department of CSE, Velammal College of Engineering and Technology, Madurai.

[2]Assistant Professor, Department of CSE, Velammal College of Engineering and Technology, Madurai.

[3]Associate Professor, Department of CSE, Velammal College of Engineering and Technology, Madurai.

[4]Assistant Professor, Department of CSE, Velammal College of Engineering and Technology, Madurai.

ABSTRACT

Facial recognition has recently become very important in the fields of artificial intelligence, robotics, security, trade, etc. It can be used in stress analysis in industries by capturing the faces of the employees and by analysing their emotions in order to put forward the suitable therapies for cure in required. Face recognition can also be used in shopping malls to analyse the facial expressions of the users while purchasing various products, thus capturing their interests for purchasing the same. This can then be used together with the other data for the process of data analysis and further promotions for the users in terms of their liking products. In order to analyze facial expressions, a variety of techniques are employed, including support vector machines, the k-nearest neighbor algorithm, and multilayer perceptron. Although there are numerous ways to recognize expressions using machine learning and artificial intelligence techniques, this project makes an attempt to use Convolutional neural networks and MediaPipe to identify various expressions from the faces and classify them appropriately. In order to learn face representation from a smaller data set, this project suggests modifying a Convolutional neural network. Convolutional neural networks are well known to have a higher accuracy when compared to traditional machine learning methods. Our system continuously captures the images of the user using Webcam and analyses the facial expressions accordingly using a convolutional neural network and to identify the mood of the user which can be used for further analyses.

Keyword: CNN, FER, MediaPipe.

Section A-Research paper
ISSN 2063-5346

1. INTRODUCTION

Facial expression detection, in today's world, is one of the most important features of human emotion recognition. Facial Expression can be defined as the facial changes observed in a person in response to the person's internal emotional state, intentions, or social communication. Automated facial expression recognition has a large variety of applications nowadays, such as neuromarketing, data-driven animation, sociable robotics, interactive games and many other human-computer interaction systems. Basically, expression recognition is a task performed by human beings on everyday basis, effortlessly. However, it is not yet easily performed by computers. In specific situations, such as frontal face, controlled settings, and high-resolution photos, several contemporary systems have also demonstrated accuracy levels more than 95%. However, these only seem to be deceptive high-accuracy. Furthermore, most of the issues that face expression recognition systems face in actual situations are not adequately represented by these systems. The use of deep learning in expression detection is mainly for obtaining higher accuracy even for dynamic sequence of input received from webcam live. However, deep learning methods require huge dataset for obtaining such accuracies.

Facial expression data is not as large as required by deep networks. Hence, the proposed system attempts to use a modified CNN and media pipe that can work on even small datasets and ensure that the required accuracy is obtained. The modified CNN, which helps obtain high accuracy even with limited dataset uses data augmentation to recreate large datasets from the available data by processes.

2. LITERATURE SURVEY

Different facial muscles contract to produce facial expressions, which temporarily deform our eyes, noses, lips, cheeks, and other facial features. From person to person, it differs. This survey provides a thorough explanation of the various techniques used for facial expression analysis.

2.1 Judgement based approach

The messages that the facial expressions convey are the main focus of approaches that rely on judgement. Based on the consensus of the experts or coders, or "ground truth," the facial expressions are categorised into a set of emotions. The majority of facial expression analyses fall into one of the six basic emotion classes listed by Ekmen and Friesen in Constants across culture in the face and emotions in 1971.

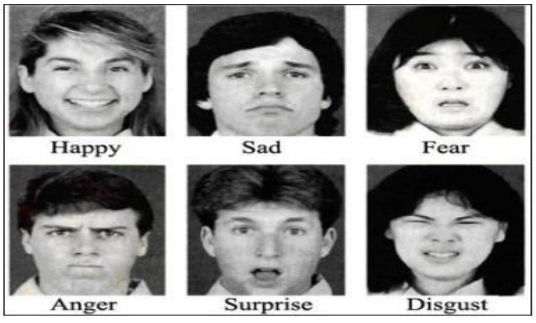


Figure 1 Six Basic Emotions

ISSN 2063-5346

Sign based approaches

Facial expressions are abstracted and described by their location and intensity in sign-based approaches. Therefore, all potential perceptions that might be present on a face make up a comprehensive description framework. Ekman and Friesen's 'A technique for the measurement of facial movement' from 1978 contained the FACS (Facial Action Coding system) proposal. For the description of facial expressions in terms of location and intensity, FACS uses 44 action units. In 1983, Izard, Dougherty, and Hembree proposed a system for identifying affect

expressions by

Figure 2

holistic judgements that included coding schemes like EMFACS, MAX, and AFFEX.

2.2 Reliability of ground truth coding

The labeling of employed databases to recognize or interpret facial expressions has improved recognition accuracy with respect to timing and intensity estimations. Ekman specifies certain points to be addressed while measuring facial expressions in 'Handbook of methods in non-verbal behavioural research' in 1982.

2.3 Automatic face expression analysis

As faces differ from person to person, automatically analysing facial expressions is a difficult task. Automatic facial expression analysis involves facial acquisition, facial feature extraction and analysis.

2.4 Face acquisition

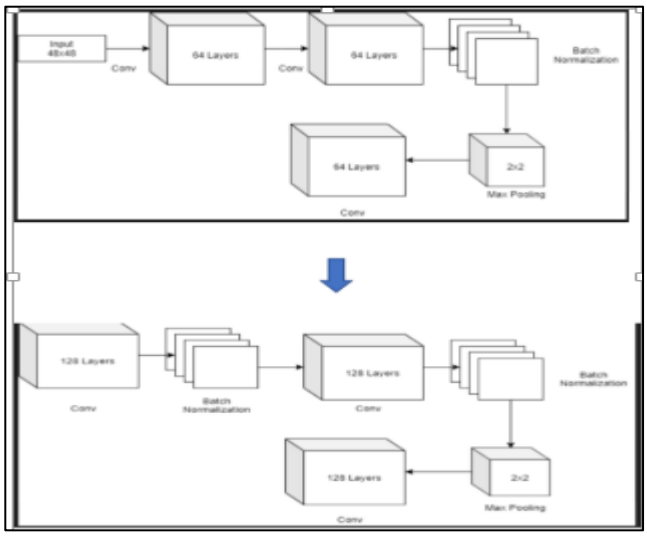
Face recognition is automatic in complicated settings with crowded backgrounds. Using the active appearance models described by Lanitis, Taylor, and Cootes in their 1997 paper "Automatic interpretation and coding of face images using flexible models," this may be accomplished. Hong tracked faces in real time by utilising the steffens person spotter technology. Depending on the viewing angle and distance for a certain face, several facial

expressions will show. In order to normalise test faces in accordance with the general face models stated in "Transactions on pattern Analysis and machine intelligence," distinguishing facial characteristics like the eyes, ears, and nose serve as reference points. Another important factor in face recognition is illumination. According to 'Proceedings of Circuits, Systems, Communications, and Computers' from 1999, Gabor wavelets can be used to lessen changes in light.

3. NEURAL NETWORK DESIGN

The main component in an emotion detection system built using deep learning is the neural network. The neural network deployed in this system will be Convolution Neural Network (CNN). TensorFlow and Keras.The model is designed using TensorFlow and Keras package. Usually, a neural network that is developed is trained by means of forward propagation and backwardpropagation. Computational graphs areconstructed to compute gradients effectively. These gradients are used in the back propagation process while training our model, thus implementing gradient descent. The data structure for the computational graph is implemented by it, as the functions of each of the weights are defined. Hence, while training, TensorFlow makes it easier for the model to compile and run faster and effectively. Fast numerical computing is made possible by the open source library TensorFlow. The Apache 2.0 open source licence was used for its creation and maintenance by Google.Although the C++ API is accessible, the API is formally for the Python programming language. TensorFlow was created for use in both production systems and research and development systems, including but not limited to RankBrain in Google search and the entertaining DeepDream project. This makes it different from other numerical libraries intended for use in Deep Learning like Theano. Mobile devices, large scale distributed systems with hundreds of machines, single CPU systems, GPU systems, and GPU-enabled systems can all run it. Hence, the models can be easily created using Keras, but it uses TensorFlow as its backend. Therefore, there is no issues related to performance when Keras is compared with TensorFlow. Keras has various types of models, layers, activations, optimizers, initializers, and many more tobuild a neural network easily.

3.1 Architecture design of CNN



ISSN 2063-5346

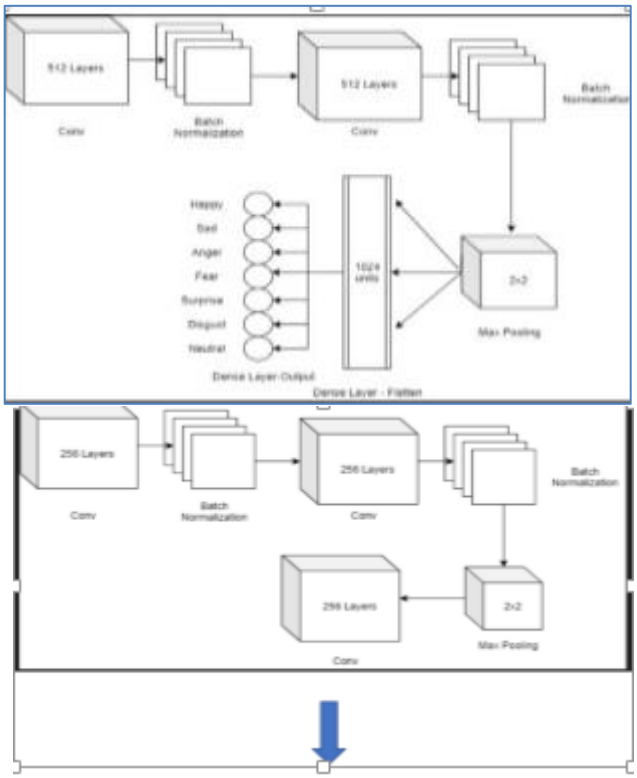


Figure 3 CNN architecture design

The structure of the architecture is asfollows:

- The image input is fed having a dimension of 48*48
- The input is then applied to two convolution layers with 64 filters of the kernel size 3*3.
- Followed by batch normalization of the output from the convolution layer maximum pooling of size 2*2 is then performed with stride 2*2.
- The output from max pooling is fed to another convolution layer of 64filters with 3*3 kernel size.
  All the above layers are illustrated inthe block diagram 1.3(a)
- A convolution layer with 128 filters and a kernel size of 3*3 receives the output from the convolution layer of 64 filters.
- An additional convolution layer with 128 filters and a batch normalisation layer are added after this.
- The output from this process is fed to a convolution layer with 128 filters. Next, a max pooling operation of size 2*2 is carried out.
- All the above layers are illustrated inthe block diagram 1.3 (b)
- The same pattern of layers as in the block diagram 1.3(b) is followed in the next set of layers also but with each convolution layer having 256filters and later with 512 filters.
- The output of the final pooling layer is flattened to create a set of dense layers with 1024 units.

- The final dense layer with seven classification outputs is then applied.

3.2 Dataset

A machine learning platform called Kaggle provided the dataset that was used to train and test the CNN. Anger, disgust, fear, happiness, sadness, surprise, and neutral are the seven emotions that are labelled on each of the 48x48-pixel grayscale images of faces in the FER2013 dataset. The Facial Expression Recognition Challenge on Kaggle was the original competition for which Pierre-Luc Carrier and Aaron Courville produced the dataset. A total of 35,887 images make up the dataset.The FER2013 dataset includes images that are presented in CSV (Comma-Separated Values) format, where each row corresponds to an image with pixel values and an emotion label. There is a training set, a public test set, and a private test set for the FER2013 dataset. 28,709 images make up the training set, while 3,589 images each make up the public and private test sets.

```
Preprocessing Done
Number of Features: 48
Number of Labels: 7
Number of examples in dataset:
```

Figure 4 Dataset Details

3.3 Training and Testing on dataset

Any CNN network needs to be trained and tested before deployment on dataset.For training and testing the datasetmentioned in section 3.2 was used. The testing accuracy of 66.369 % is obtained.

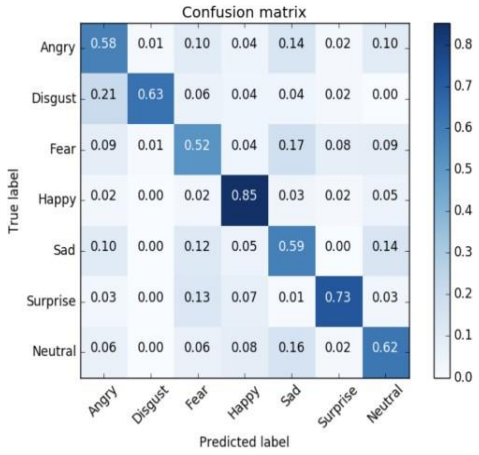


Figure 3.4 Confusion Matrix

Figure 5 Confusion matrix

SYSTEM IMPLEMENTATION

This chapter is about the prototype implementation. It discusses about the various

modules implemented and overall conceptual architecture of the system.

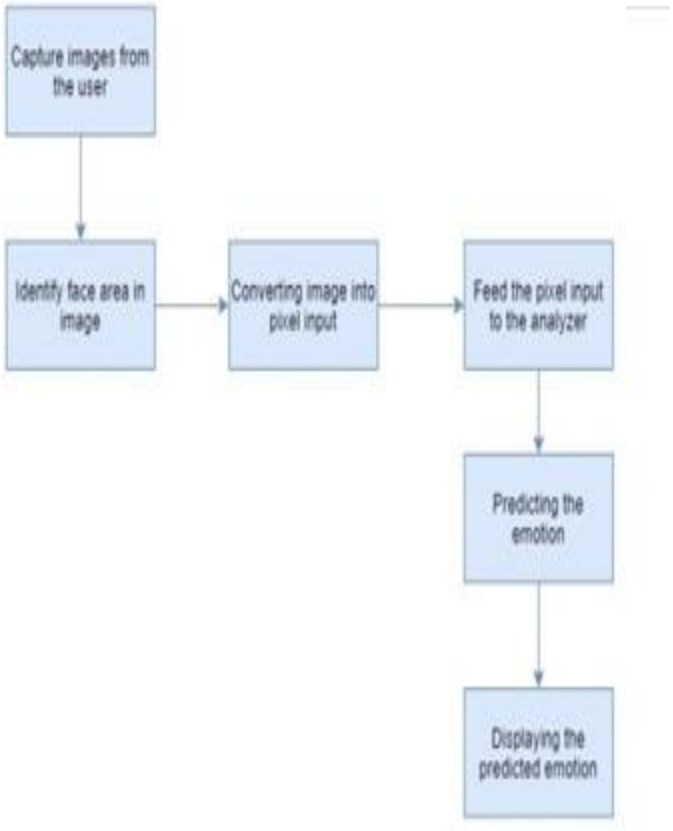


Figure 6 Conceptual Architecture

3.4 Conceptual architecture

The conceptual architecture visually describes the various aspects of the system at a higher level. This Conceptual Architecture's main goal is to make the system understandable. The system takes the image from web camera as input data. The captured image input is given to the pre-processor in which the face area is recognized from the image input. The pre- processed image is converted into pixels and fed into the model for prediction. The model trained using CNN is used for prediction. The model generates the suitable emotion prediction using the trained data. The prediction is then displayed to the user as output.The overall workflow between the various components is described in the following block diagram of the proposed system.

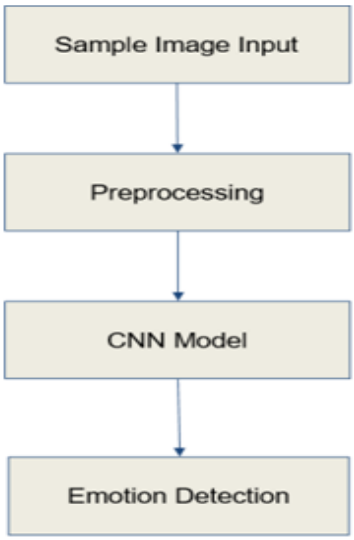


Figure 7 Flow of Components

3.5 Input module

Input module captures facial inputs of the user from the web camera and feeds the input for further processing and for final prediction of the mood of user. This is done using **emotion.py** module in python.

3.6 Pre Processing module

Pre Processing module is used to process theinput image so that it can be fed into the trained model for prediction. The image input generated from the camera is pre- processed in this module. The image is processed, and the facial component is identified. This is done with google MediaPipe face detection module. All this is done using **cv2module** in python.

3.7 Analyzer module

The faces detected using the cascade classifier are resized to 48*48 size and normalized. After completing all this, theresultant array will be provided as the input to the CNN. All this is also done using **cv2 module** and **NumPy** in python.

3.8 Emotion detector

The Emotion Detector module is designed based on **Convolutional Neural Networks** model which is popularly used for imageclassification.

ISSN 2063-5346

4. OUTPUT AND EVALUATION

Output during training

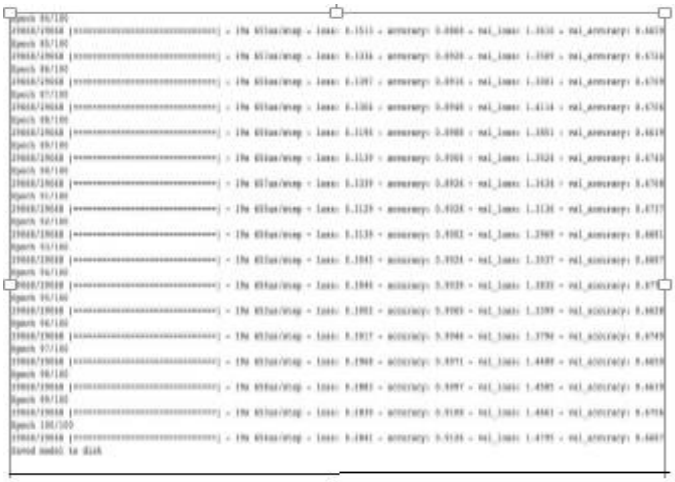
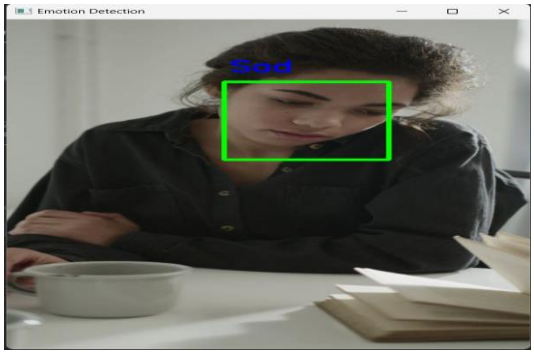


Figure 8 Training Output

Output from the emotion detector

- Output Evaluation on images taken from internet. These images are taken in well lighted conditions in professional camera. This was processed in the implemented system.



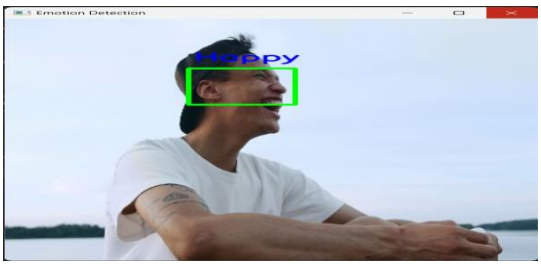


Figure 9 Images from Laptop

APPLICATION AND FUTURE SCOPE:

Application:

The product's scope is multi-fold. It can be used as a part of numerous applications. Some of the scopes of this system are discussed below. The system has the ability to recognize and monitor a user's mental state. To view customer feedback and improve the business, the system is also used in mini-marts and shopping centers. Using the emotional intelligence of a person that this system can identify, clever marketing is possible. The system can be installed in crowded areas like airports, train stations, or bus stops to detect people's faces and facial expressions. The system may sound an internal alarm if it notices any suspicious emotions, such as anger or fear. The system can also be employed for educational purposes, such as receiving feedback based on how a student is acting in class. It can also be used as plugins to monitor mentally disordered patient's behavior at difficult level. System is fully automatic and has the capability to work with video feeds as well as images. It can recognize spontaneous expressions. This makes the system capable of being used even in Digital Cameras where in it is tuned such that the image is captured only when the person smiles or similar emotions. It can feature as added plugin in home automation. For instance, rooms in the house can set the lights, television to a person's current mood by figuring out his/her expression when he/she enters.

Future scope:

The developed system is restricted only to predict the emotion of the facial expression captured at an instance of time by the system. The system can be extended to predict the emotion of the user captured overa sequence of period i.e a video input as a sequence of images to train the model. It can be expanded to forecast emotions based on the user's vocal input, or forecast emotions based on user sound patterns. This system can be delivered as a plug-in feature so that it can be incorporated with many applications that use this feature(emotion detection) in it.

CONCLUSION:

Thus, the purpose of this project has been achieved by efficiently predicting the emotion of the corresponding facial expression captured by the system. This has been achieved by using a trained Convolutional Neural Network architecture. The model has been successfully trained with sufficient input output samples and has produced a decent prediction accuracy as well. The performance has been improved by using increased kernel filters and sufficient dropout percentage so as to ensurethe required features are captured and learnt by the model as well as the model does not overfit for the training samples.

REFERENCES

[1] Jie Wang, Zihao Li, "Research on Face Recognition Based on CNN", 2018 IOP Conf. Ser.: Earth Environ. Sci. 170 032110.

[2] https://towardsdatascience.com/an-intro-to-deep-learning-for-facerecognition.

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

[4] Chalciya Elizabeth Rani P1, Dr. H.P. Mohan Kumar2 "Facial Emotion Detection Using Keras ".

[5] Prathwini, Roshan Fernandes, Anisha P Rodrigues "Emotion Detection in Multimedia Data Using Convolution Neural Network".

[6] Amr Mostafa, Mahmoud I. Khalil, Hazem Abbas " Emotion Recognition by Facial Features using Recurrent Neural Networks ".

[7] " Human Emotion Detection Using Open CV " Mallika Srivastav, Prakhar Mathur, T. Poongodi, Shrddha Sagar, Suman Avdhesh Yadav.

[8] Sarwesh Giri, Gurchetan Singh, Babul Kumar, Mehakpreet Singh, Deepanker Vashisht, Sonu Sharma, Prince Jain " Emotion Detection with Facial Feature Recognition Using CNN & OpenCV "

[9] Ninad Mehendale,"Facial emotion recognition using convolutional neural networks ".

[10] Akriti Jaiswal, A. Krishnama Raju, Suman Deb " Facial Emotion Detection Using Deep Learning ".