Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12, Issue 04, April 2021: 2257-2264

Research Article

Facio Metrics

Dr. S Radhika, Kavya S, M Swetha, Suchitra S

[1,2,3,4]Department of Computer Science and Engineering, R.M.K Engineering College, R.S.M Nagar, Kavaraipettai

Abstract

According to a survey conducted by a health policy research group, mental health depression has had a negative effect on 53% of adults. Because of its numerous applications in artificial intelligence, detecting an emotion from a human face has become a demand. The aim of this study is to create a facial expression recognition system based on data augmentation and transfer learning. This method allows image data to be classified into seven basic emotions: rage, disgust, fear, happiness, neutrality, sadness, and surprise. Transfer learning with data augmentation achieves a higher level of precision (96.24 percent) and helps to address the shortcomings of existing models

Keywords: *Transfer learning, artificial intelligence, human-computer collaboration, and data augmentation are some of the terms used in this paper*

Introduction

Emotion detection has become a common topic in computer vision in recent years. It is computer software that aids in the identification of human emotions such as happiness, anger, sadness, fear, disgust, and surprise. Furthermore, in public spaces, Emotion Recognition can be more freely conveyed. Extreme arguments between two people, driving vehicles with rage, and so on are only a few examples of human emotions in public places. They will avoid any gruesome act or danger by knowing the person's emotion. Many scientists are interested in analysing facial expressions. Because, as mentioned in the quote, **"one person can be judged by their reactions rather than their behaviour."** In this way, rather than words, a person's mood can be easily discerned by their emotions. Various techniques for automatically detecting human emotions have been developed, and facial expression can be implemented as computerised software. Emotion detection systems are used in a variety of areas, including mobile computing, robotics, and psychology. Any digital camera or video was used to capture an image. Many studies have recently been conducted in the area of facial expression recognition.

Facial emotion detection is a branch of machine learning that is still in its infancy. In this paper, we use mobile net version-2 to predict our model based on Transfer learning (V2). On the aforementioned datasets, the proposed model achieves adequate results in detecting macro facial emotions. Body language makes up 55 percent of the total message in face-to-face contact, while words make up just 7%. The manner in which we deliver the message is critical in comprehending the condition as a whole. Although it is simple for humans to interpret facial expressions, teaching a computer to process data and comprehend human emotions in real-time applications presents a difficult challenge.

With the aid of NLP, machines will understand some simple verbal communication, and understanding facial expressions can help machines understand humans better. Human Computer Interaction (HCI) is a wide field that focuses on creating the most successful interfaces for applications like psychological consultations, medical care or healthcare, tracking, recovery, marketing, advertising, video games, movies, music, and education. In this paper, we propose a new Transfer learning model that recognises seven facial emotions in real time: sorrow, happiness, rage, surprise, disguise, fear, and neutral.

Different people have different facial expressions. The following are the crucial steps in the Facio metrics process:

- Surveillance of the face
- Extraction of Signatures
- Differentiating Expressions

1.1 Surveillance of the face :

Face detection is a technique for detecting human faces in pictures or digital images of some kind. To detect the human face, photographs were first converted to greyscale images. We can see the images in various types like Individual images, large sized images, crowded images, coloured images, and other forms of images are available. To detect facial features in videos and photographs, various algorithms are used. Face recognition algorithms were used to determine the size and position of a face area.

1.2 Extraction of Signatures:

Various features of the human face must be extracted after accurate face detection from the source image. It symbolises a facial feature. The shape of the eyes, eyebrows, width of the mouth, shape of the ears, length of the face, distance between nose and mouth, and so on are some of the main features that should be derived from a human face.

1.3 Differentiating Expressions:

There are two categories of classification. There are both supervised and unsupervised versions of them. When compared to unsupervised classification, supervised classification yields better results. Many classification algorithms have been developed, including c-Means, k-Nearest Neighbor, and neural networks.

Literature Survey

“WIDER FACE: A Face Detection Benchmark”,

The author of this paper explored a potential face recognition dataset. They've released the WIDER FACE dataset. It is wider and contains various types of faces, such as occlusion, poses, and scales. ALFW, FDDB, and PASCAL FACE datasets lack sufficient training data.

They conclude that a WIDER dataset from a diverse perspective improves both negative and positive sample preparation.

“Spatial-temporal recurrent neural network for emotion recognition,”.

They used convolutional neural networks (CNN) and recurrent neural networks (RNN) to integrate spatial connections in EEG signals in this paper. Their methods necessitate a two-dimensional representation of EEG channels on the scalp, which can result in data loss during flattening. The channels are simply laid out in three dimensions. It has trouble learning long term dependencies.

Proposed System

The dataset was collected from FER-2013. The image size of the FER-2013 dataset is 48 x 48 dimensions which is resampled into 224 x 224 dimensions using bi-linear interpolation. The model is trained in MobileNetV2 architecture. It is an efficient network for mobile applications. We replace a completely linked layer with only seven classes on human faces with this network, which can classify images into 1000 object categories. They are angry, disgusted, fearful, happy, sad, surprised and neutral.

3.1 Dataset

We collected the Face Expression Recognition 2013 (FER 2013) dataset from KAGGLE.

The following values are included in the dataset:

0-Angry 1-Disgust 2-Fear 3-Happy 4-Sad 5-Surprise 6-Neutral

FER 2013	0	1	2	3	4	5	6
Trainin g	399 5	43 6	409 6	721 4	483 0	317 1	496 5
Testing	958	11 1	102 4	177 4	124 7	831	123 2

3.2 Problems in Dataset

Imbalance problem:

The solution for **Imbalance** problem is **data augmentation** which increases the dataset. From the table of dataset, we can observe that there are only 436 limited images for disgust whereas for happiness, we have 7214 collections of images. The one way to equal 436 to 7214 images is, we can rotate an image and create another image. Other ways are also available to increase the images. We can scale and zoom in or zoom out images. Various techniques have been used to increase the number of images.

Intra-class variation of FER:

The intra-class variation means image variations occur between different images of one class. With FER - 2013 dataset, we have cartoon, painted and sketch images which are totally different. There are a lot of variations in images. Deep learning architecture must be robust enough to tackle all of them.

Occlusion:

Looking into the figure.1, there are some images with contrast variation, cropped images, covering the face with hands and eyeglasses problems. With these images, the accuracy will be low because these are not an easy dataset.

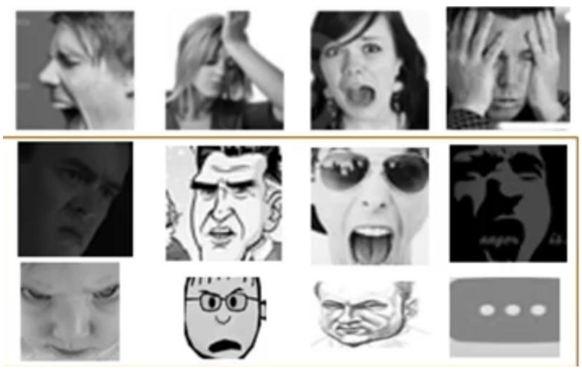

Figure.1 Problems in dataset.

Methodology

General Description

The datasets were collected and preprocessed in order to convert data into such a way that it should be used by machine learning models.

The dataset that we used in our project was FER-2013.

The figure 2.1 shows the dataset folders.

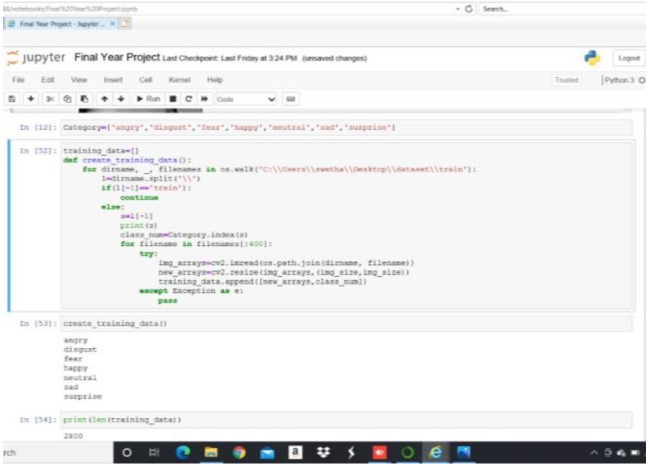

Figure 2 Dataset folders.

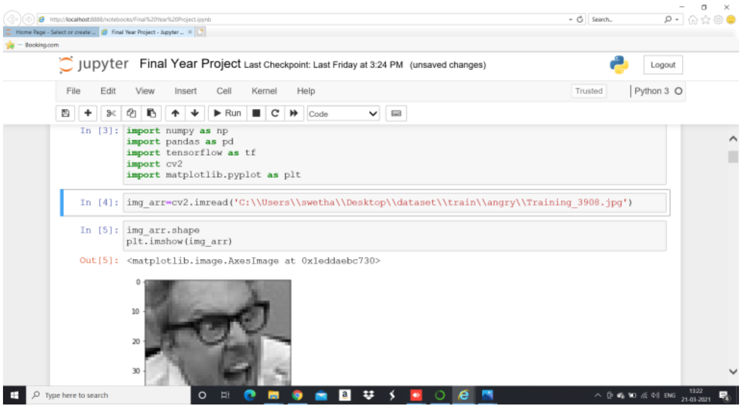

Figure 3 Importing the packages and reading 48 x 48 pixel images.

Bilinear Interpolation:

The FER-2013 dataset was 48 x 48 dimensions. We have to resample into 224 x 224 dimensions using bilinear interpolation. We have to modify these weights for facial emotional connection. The below figure shows how image size is resampled.

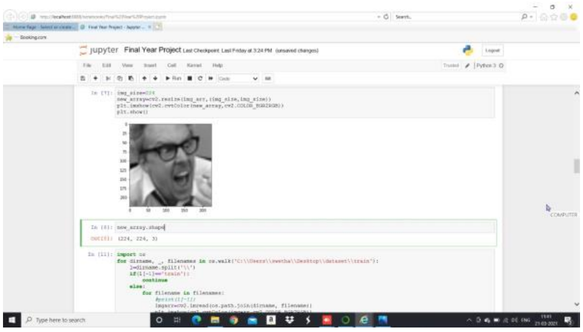

Figure 4 Bilinear interpolation.

The model is trained in MobileNetV2 architecture. It is an efficient Convolutional Neural Networks for mobile applications. This network can classify images into 1000 object categories that we replace with a fully connected layer having only seven classes on human faces.

The number of neurons inside the fully connected layer is always the same as the number of neurons outside the fully connected layer. The network divides the model into 1000 categories, from which we choose seven for output.

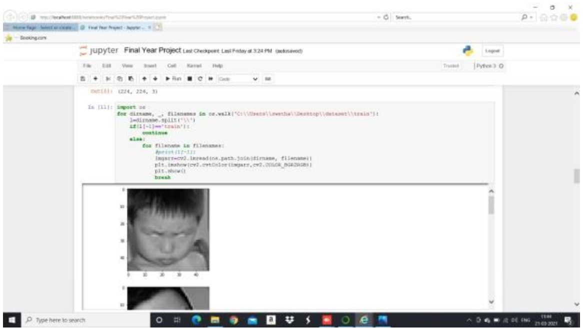

Figure 5 Reading seven classes of images.

Transfer Learning

Transfer learning is a technique in deep learning in which a model is initially developed for one particular task will be reused for another task in future.

MOBILENETV2 :

MobileNet-v2 is 53 layers deep convolutional neural network. This network is capable of classifying images into 1000 object categories from which we are using seven classes for the output image.

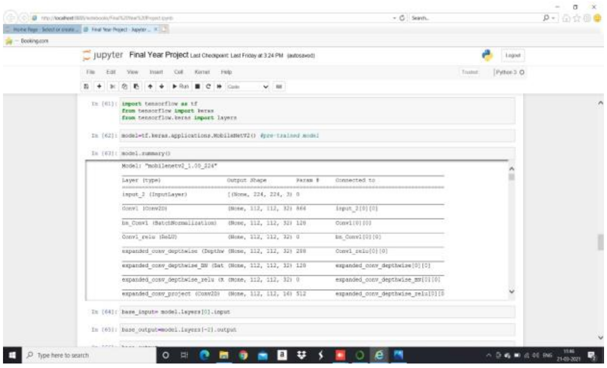

Figure 6 MobileNetV2 model.

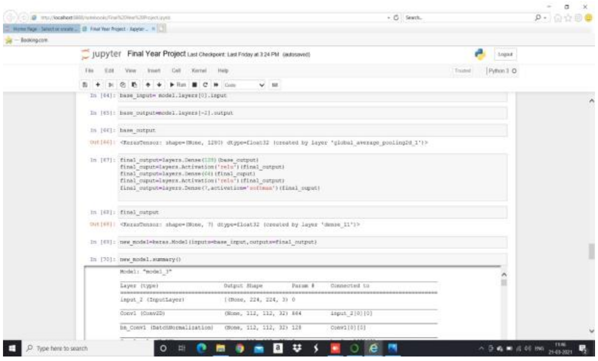

Figure 7 Layer changing code.

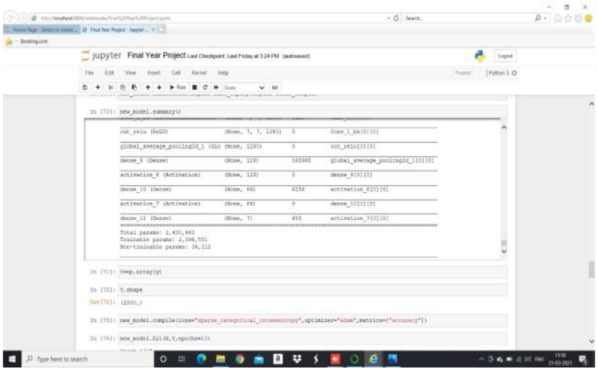

Figure 8 New model.

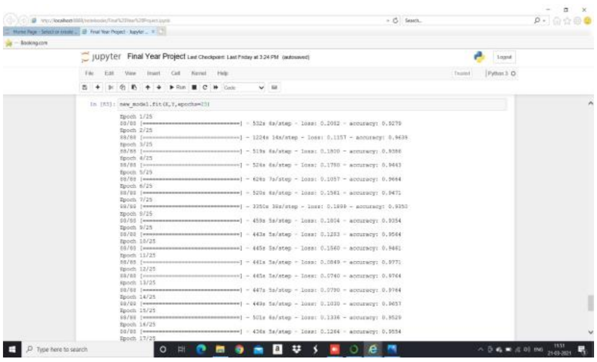

Figure 9 Save and load model.

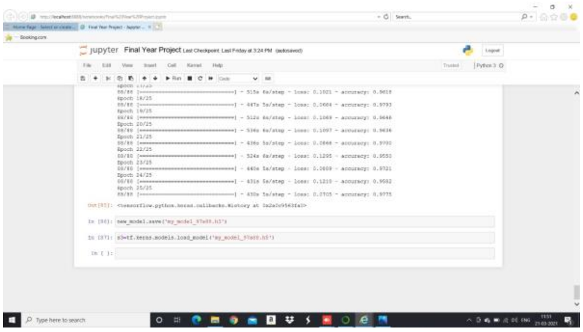

Figure 10 Model training.

Conclusion

We used a Transfer learning technique to assess human emotions in this article. On the aforementioned datasets, satisfactory results were obtained in detecting macro facial emotions. Our project is based on neuroscientific theories of human brain organisation. We also use mobile net version-2(V2) to increase the robustness of our platform in comparison to other models. Extensive testing on two public datasets demonstrates our model's superior output. The results of our model's analysis show that the techniques used produce a consistent and substantial improvement in the model's efficiency.

In future,

1. We planned to range the emotion of human beings on a scale of 10 where 0 starts with low and 10 ends at peak value.
2. Planning to capture the image from a live webcam in order to predict the emotions.

References

[1] T. Zhang, W. Zheng, Z. Cui, Y. Zong, and Y. Li, “Spatial-temporal recurrent neural network for emotion recognition,” .

[2] J. Li, Z. Zhang, and H. He, “Hierarchical convolutional neural net-works for EEG-based emotion recognition,” Cognitive Computation, vol. 10, no. 2, pp. 368–380, 2018.

[3] M. Defferrard, X. Bresson, and P. Vandergheynst, “Convolutional neural networks on graphs with fast localized spectral filtering,”in Advances in Neural Information Processing Systems, 2016, pp.3844–3852.

[4] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A simple way to prevent neural networks from overfitting” The journal of machine learning research, vol.15, no.1,pp.1929-1958, 2014.