ABSTRACT

Emotion is one of the few terms which does not have a precise definition that is also intelligible. It's ethereal. "Emotion" is a term used to describe a subjective condition expressed through social signals. This differs from, for instance, recognizing the emotional content of a multimedia clip, which is concerned with the feelings that the clip may elicit in its audience. Despite this, practically every decision we've ever made has been influenced by emotion. According to marketing studies, correctly forecasting feelings can be a big source of growth for businesses. Also, can be used for safety measures such as drowsiness detection for a car driver. In the computer vision domain, recognizing automated facial expressions from a facial image is a difficult task that seems to have a wide range of applications, covering driving safety, human-computer interactions, health care, behavioural science, teleconferencing, cognitive science, as well as others. This notion falls under the category of cognitive systems in the area of data and machine learning. The goal is to try to segregate the emotions contained in the data using training data. The FER2013 data consists of pixel images of emotions of faces of seven types of emotions which are Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral. The faces have been automatically registered so that the face is more or less centred and occupies about the same amount of space in each image. The goal is to recognize the emotions in a real-time environment using a webcam. Deep convolutional neural networks have been used to produce numerous advancements in image categorization (CNNs). An automated feature extractor and a classifier are the two major components of these designs. The former generates low-level, mid-level, and high-level characteristics for the item of interest, characterizing basic, moderate, and complicated textures, respectively. In general, a powerful classifier learns the target from a large number of high-level characteristics, hence the network should be trained using a huge amount of data. Deep learning techniques are frequently employed in many picture classification difficulties, such as ImageNet, PASCAL VOC, CIFAR, and the Facial Expression Recognition Challenge 2013. The FER2013 data consists of images of faces of seven types of emotions i.e., Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral. Using the dataset, these networks may be leveraged to produce task-specific outcomes. Using Transfer Learning, we utilized our dataset on a pre-trained network. The pre-trained network is used for both feature extraction and

classification. Transfer Learning takes features and weights from previously trained models and applies them to subsequent models, even when there is less information on the most recent job.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 OUTLINE OF THE PROJECT

Artificial Neural Networks (ANN) is inspired by the human brain and it's can be used for machine learning and artificial intelligence. With these networks, various problems can be solved computer-based. The artificial neural network (ANN) is to some extent modelled on the structure of the biological brain. It consists of an abstracted model of interconnected neurons, whose special arrangement and linking can be used to solve computer-based application problems in various fields such as statistics, technology or economics. The neural network is a research subject of Neuro informatics and part of the artificial intelligence. Neural networks must be trained before they can solve problems.

Image recognition is one of the tasks in deep neural networks. Neural networks are computing systems designed to recognize patterns. Their architecture is inspired by the human brain structure, hence the name. They consist of three types of layers: input, hidden layers, and output. The input layer receives a signal, the hidden layer processes it, and the output layer makes a decision or a forecast about the input data. Each network layer consists of interconnected *nodes (artificial neurons*) that do the computation. Transfer Learning takes features and weights from previously trained models and applies them to subsequent models, even when there is less information on the most recent job.

CHAPTER 2

LITERATURE SURVEY

This section summarises all of the previous work on facial expression recognition. They used MLCNN which automatically selected important mid-level and high-level features based on their contribution and also the video-based facial expression recognition, the temporal model whose backbone is the proposed ensemble of MLCNNs also achieved comparable performance and outperformed in Facial Expression Recognition Using a Temporal Ensemble of Multi-level Convolutional Neural Networks. Facial Expression Recognition Using Deep Convolution Neural Networks Based on Local Gravitational Force Descriptors is an article published in the journal Facial Expression Recognition. The suggested technique beats twenty-five baseline methods by considering the average time and achieves average recognition accuracy of 78 percent, 98 percent, 98 percent, 96 percent, and 83 percent for FER2013, JAFFE, CK+, KDEF, and RAF, respectively. Keith Anderson and Peter McCowan’s A Real-Time Automated System for the Recognition of Human Facial Expressions was not only trained to distinguish between different expressions but also between expressive and non-expressive sequences. However, this system found the classification of fear and anger expressions particularly difficult, and as a result, the accuracy was reduced. Maja Pantic and Ioannis Patras’ Dynamics of Face Expression: Recognition of Facial Actions and Their Temporal Segments from Face Profile Image Sequences automates the segmentation of input image sequences into expressive and expressionless facial activity. Their solution uses a memory-based mechanism that takes into consideration the dynamics of facial expressions to cope with mistakes in facial point tracking, however, the suggested facial expression analyser cannot deal with spontaneously occurring facial behaviour. A Face Emotion Recognition Method Using Convolutional Neural Network and Image Edge Computing, a technique was suggested that can automatically learn pattern features and eliminate the incompleteness caused by artificial design features, as well as direct input the image pixel value via training sample image data. Furthermore, the suggested technique yields a greater recognition rate, although the neural network architecture is extremely complicated in this case. Hyeon-Jung Lee and Kwang-Seok Hong's work on Emotion

Recognition Method and Its Application Using Face Image found that the application performance rate in positive and negative emotions was 72.3 percent, but only 50.7 percent for seven categories, and the neutral category was eliminated. The suggested Human Emotion Recognition using Convolutional Neural Network in the Real-Time model can classify seven different emotions regardless of the number of faces at a time. Although the shallow model, which consists of a single convolution layer, achieves an accuracy of up to 47.94 percent, the model's average performance in identifying the disgust emotion is lower due to a fewer number of pictures in that category in the training dataset. Charvi Jain, Kshitij Sawant, Mohammed Rehman, and Rajesh Kumar's paper Emotion Detection and Characterization Using Facial Features This method only requires two attributes to determine the emotion of a whole face, reducing the amount of storage data required for testing and future applications, but it can only recognize six of the seven emotions. Recognition of Facial Expressions Based on CNN, it provides a method that is very stable and is unaffected by subject-to-subject face physiological changes, but it suffers from overfitting in the training and testing phases due to a lack of facial emotions. Facial Emotion Recognition of Students Using Convolutional Neural Network proposed model by Imane Lasri, Anouar Riad Solh, and Mourad El Belkacemis achieved a 70 percent accuracy rate on the FER 2013 training database, but the actual performance of the model in real-time environment and accuracy on validation dataset is quite low and predicts quite feared faces poorly because it confuses them with sad faces.

CHAPTER 3

AIM AND SCOPE OF THE PRESENT INVESTIGATION

3.1 PURPOSE OF THE PROJECT

The main goal of this project is to build a neural network model which gives an accuracy more than what we have received in previous works as mentioned in the literature review on the same dataset (i.e., FER2013) and be able to work dynamically in real-life scenarios. The dataset chosen for the project is FER2013. The pre-trained models from transfer learning will be used to achieve the target. E.g.: Resnet50, MobileNetV2 etc. The model should be able to predict every emotion with similar accuracy and none of the emotions would be excluded. Here we have tried to build a model that should be able to do this on a real-time basis using a webcam or a camera and the time taken to generate the results be minimized as much as possible to facilitate real-life usage.

Disadvantages:

A few types of images the model tends to do poorly on include:

- The human in an unusual position
- Face appears against a background of a similar colour
- Brightness of the picture
- Scale variation (face is very large or small in image)

Advantages:

The model has higher prediction than the basic machine learning model.
This is simple algorithm built using simple and open-source packages like TensorFlow and SciPy library.

3.2 PROJECT ARCHITECTURE:

ARCHITECTURE OF THE MODEL:

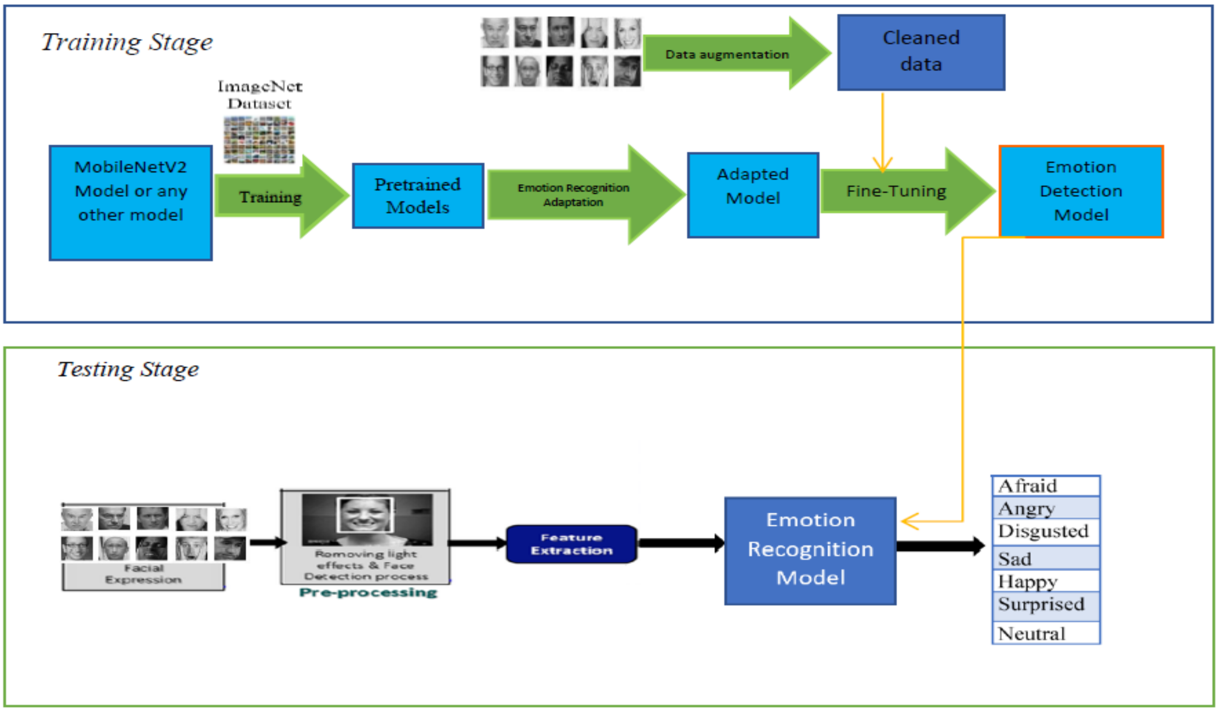

Figure 3. 1

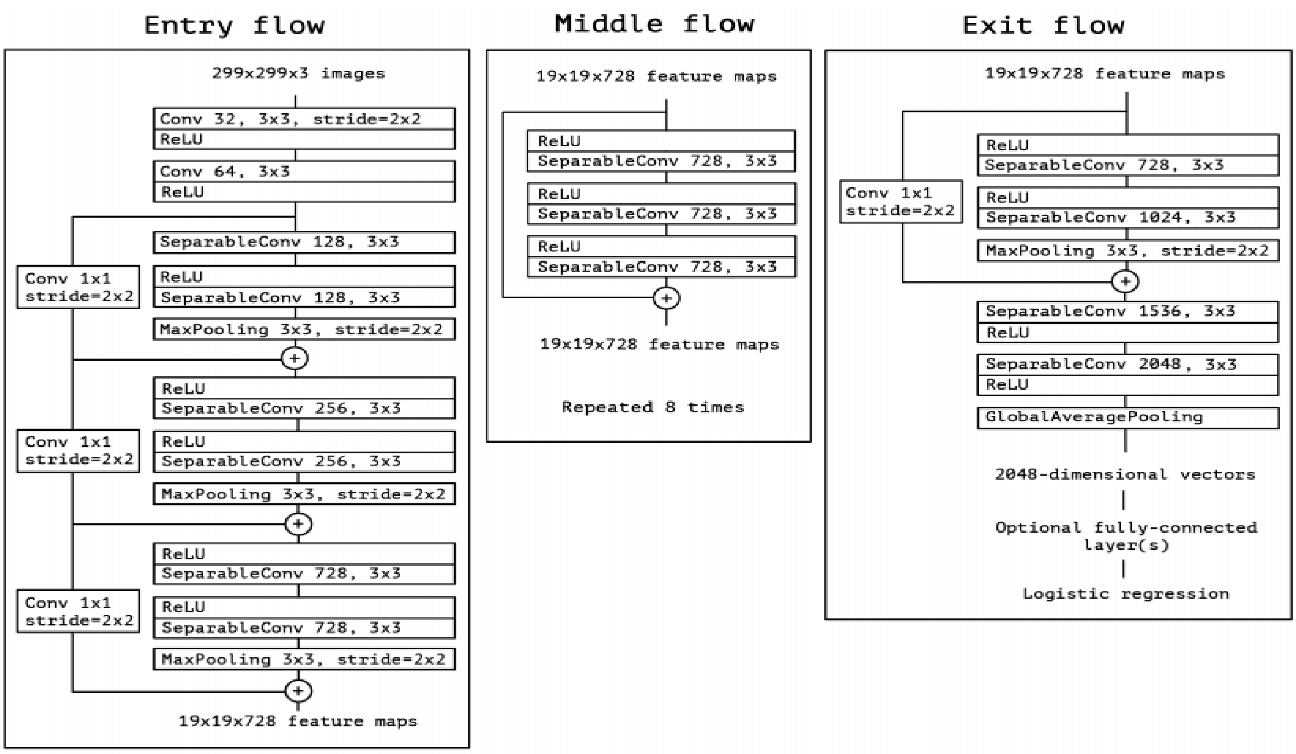

Figure 3.2

CHAPTER 4

EXPERIMENTAL OR MATERIALS AND METHODS

ALGORITHMS USED

4.1 INTRODUCTION

Image recognition is one of the tasks in deep neural networks. Neural networks are computing systems designed to recognize patterns. Their architecture is inspired by the human brain structure, hence the name. They consist of three types of layers: input, hidden layers, and output. The input layer receives a signal, the hidden layer processes it, and the output layer makes a decision or a forecast about the input data. Each network layer consists of interconnected *nodes (artificial neurons*) that do the computation.

Traditional neural networks have up to three hidden layers, deep networks may contain hundreds of them. So, to be able to recognize faces, a system must learn their features first. It must be trained to predict whether an object is X or Z. Deep learning models learn these characteristics in a different way from machine learning models. That's why model training approaches are different as well. Each layer of nodes trains on the output (feature set) produced by the previous layer. So, nodes in each successive layer can recognize more complex, detailed features visual representations of what the image depicts. Such a "hierarchy of increasing complexity and abstraction" is known as *feature hierarchy.*

So, the more layers the network has, the greater its predictive capability.