MUHAMMAD UMAR SAFEER
01-249191-007

Emotions Analysis on Facial Expressions

MS (Data Science)

Supervisor: Dr.Arif Ur Rahman

Department of Computer Science
Bahria University, Islamabad

April 28, 2021

MS-13
Thesis Completion Certificate

Student Name: **Muhammad Umar Safeer** Registration Number: **62703**
Program of Study: **MS (Data Science)**
Thesis Title: **Emotions Analysis on Facial Expressions**

It is to certify that the above student's thesis has been completed to my satisfaction and, to my belief, its standard is appropriate for submission for evaluation. I have also conducted plagiarism test of this thesis using HEC prescribed software and found similarity index at **14%** that is within the permissible set by the HEC. for MS/MPhil/PhD.

I have also found the thesis in a format recognized by the BU for MS/MPhil/PhD thesis.

Principle Supervisor's Signature:__________________
Principle Supervisor's Name: Dr.Arif Ur Rahman
April 28, 2021

MS-14A
Author's Declaration

I, **Muhammad Umar Safeer** hereby state that my MS thesis titled **"Emotions Analysis on Facial Expressions"** is my own work and has not been submitted previously by me for taking any degree from **"Bahria University, Islamabad"** or anywhere else in the country / world.

At any time if my statement is found to be incorrect even after my Graduate the university has the right to withdraw cancel my MS degree.

MUHAMMAD UMAR SAFEER
01-249191-007
April 28, 2021

MS-14B
Plagiarism Undertaking

I, **Muhammad Umar Safeer** solemnly declare that research work presented in the thesis titled

Emotions Analysis on Facial Expressions

is solely my research work with no significant contribution from any other person. Small contribution / help whenever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of Bahria University and the Higher Education Commission of Pakistan towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarised and any material used is properly referred / cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS degree, the university reserves the right to withdraw / revoke my MS degree and HEC and the university has the right to publish my name on HEC / University Website on which name of students who submitted plagiarised thesis are placed.

MUHAMMAD UMAR SAFEER
01-249191-007
April 28, 2021

Abstract

Multiple researchers, maximum human communication is non-verbal basis. The traditional Human Computer Interaction (HCI) focuses only on the intended input and overlooks the information delivered non-verbally. So, it is safe to say that there is need of a system which helps to recognize and perceive the objective and emotions expressed by social indicators. For the last many years, research on Facial Expressions Recognition (FER) has been under consideration in computer vision field but there are still many answerable questions outstanding. Through this research few of those questions are tried to be figured out. Emotions are said to be helpful in different fields such as biomedical engineering, psychology, neuroscience and medical health. They are helpful in analyzing many diseases in medical filed. In last few year Deep Learning is one of the major progressing field used for image classification problem. Through this research, a Convolutional Neural Network (CNN) based architecture has been proposed, which can be used for facial expression recognition problems. Emotions was classified into 7 classes include happy, sad, neutral, angry, fear, surprise, disgust. Computed results were very effective in observing human behavior which, as a result, would be helpful in psychological disorders. An independent method was proposed for this work. In First Method, using autoencoders to develop a unique representation of sentiments, while in second part using 8-layer convolutions neural network (CNN). These Methods use Google facial expression comparison data set. Computed results indicated that with better tuning of data, this model can perform accurately. So, through the simulation study, it was also observed that the prediction performance of these proposed approaches was far better in each class as compare to existing approaches. Overall, these proposed approaches played an import role in emotion recognition which can help significantly in identification of biological disorders, Computational Biology, Molecular Biology, Bioinformatics. Also, this might help in applications related to emotion recognition. In addition, the related web predictors used in this study provide sufficient information to researchers and academicians in future research.

“Stop doubting yourself, work hard, and make it happen.”

Contents

List of Figures

List of Tables

Acronyms and Abbreviations

CNN	convolutional neural network
CPC	Cost Per Click
FER	Facial Expression System
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
JAFEE	Japanese Female Facial Expression
KDEF	Karolinska Directed Emotional Faces
R-CNN	Region Based Convolutional Neural Networks
SVM	Support Vector machine

Chapter 1

Introduction

Concept of Emotion has some place in history which evolve with the passage of time [6]. Human being is in-hated by some of the ideas which are knowledge pieces of it. It becomes natural while processing emotions rather then social. Execution of emotion has major concerns for decades for computers and humans. Most of the research has been done in Machine learning and computer vision filed to classify emotion which are portrayed by humans. Human speech recognition gestures and facial expressions are included in the field [7].

1.1 What is Emotion?

> Emotions are biological states associated with the nervous system brought on by neurophysiological changes variously associated with thoughts, feelings, behavioral responses, and a degree of pleasure or displeasure [8].

There are multiple studies on emotions, but there is no single definition in research about emotions. It could be reflection of feelings or reaction. It can be real or fake. Feelings of happiness or pain can directly reflect some emotions. It represents psychological situation of humans. Emotion is a significant, complicated and substantial area of research in biomedical engineering field, psychology, neuroscience and health. Its analysis can prove to be an interesting research area in biomedical engineering. Predicting human behaviour expressions through computer assisted diagnose of psychical disorder is helpful [8].

1.2 Emotional Analysis

During mutual conversation, thousands of facial actions are expressed by human face. These actions vary in presentation and meaning. Many of the facial features change to express different emotions. The Facial expressions can be interpreted differently by adding or subtracting one or more facial actions. Because of Darwin's work, the facial expression analysis has become very trendy and prevalent in behavioral sciences. In 1978 (Suwa, Sujie, and Fujimora), it was first attempted to develop an automatic FER (Facial expression recognition). FACS based analysis and prototypical facial expression analysis are the two main categories proposed by this system to analyze the facial expressions. In FAC based analysis, the FACs consist of 46 action units (Ekkman and Friesen). These action units represent a variety of facial muscles and a combination of these action units reveals the movement of facial muscles which gives the information about the facial expressions. FACS codes are manually labelled by the experts traditionally but now automatic labeling [9] is also done by new techniques. Although FACS can compare the subtlety of facial expressions but FACS is completely descriptive and include no inferential labels. So, there is a need to convert the FACS codes to emotional facial action system to find out the estimated expression labels. However, small set of prototypical facial expressions are recognized by many of the FER systems. There are six basic facial expressions which are most widely used including Anger, Disgust, Fear, Happy, Sad, Surprise, or seventh, which is neutral face. These basic facial expressions have been observed in people of different cultures and societies of different backgrounds. Even these basic expressions have been observed in congenitally deaf and blinds. For the multi class problems, both binomial and multinomial classification techniques have been proposed. According to the mechanism, by which the facial expression information is being extracted, there are two broad categories. We have first classification which is the Model-layout based and the subsequent one element based format. Model-template based technique methods uses 2D facial models or it uses 3D facial models as template to extract information from facial expression while in the feature-based methods, geometrical or textured information of facial expression is extracted known as features. Proposed methodology for this research is the feature-based methods [10].

1.3 Characteristics of Emotions

Facial expression can be characterized as contracting and development of facial muscles. As a result, the attributes of facial highlights are changed like lips, eyebrows, eyelids, nose, brow, and skin surface. All fundamental outward appearances are spoken to by a blend of these disfigurements. Subsequently, the investigation and acknowledgment of a large de-meanor that is made out of these developments and changing in facial highlights is viewed as a perplexing and unwieldy task (BENLI 2013). Facial expression appearances can be ordered into Posed and Spontaneous articulations as per the manner in which they are communicated. Unconstrained articulations further can be separated into Induced articulations and Naturalistic articulations. Presented articulations are spoken to deliberately by the subjects under a controlled climate and in ideal lighting conditions.

Henceforth presented articulations are coordinated/shown and not characteristic. The subjects in presented articulations, as a rule, attempt to represent them as clear and unmistakable as it could be expected under the circumstances, henceforth, regularly considered as "overstated". On the other hand, unconstrained articulations are the normal reactions of subjects' interior feelings. Actuated articulations are not quite the same as naturalistic articulations. As the previous sort of articulations are prompted, for example purposefully set off in the subjects with the assistance of some external sources called apparent boosts like film or show scenes, jokes etc. which are the genuine facial muscle's responses of the inside sentiments that are normally inspired by certain conditions in reality.

One of the fundamental problems in the development of this field is the inaccessibility of standard outward appearance datasets. The vast majority of the accessible datasets comprised of presented facial articulation pictures/recordings. A couple of datasets, which are as of late formed, are comprised of unconstrained outward appearances. So, these datasets comprise of instigated facial articulations and none is yet found with naturalistic articulations. A few creators asserted some restricted trials on naturalistic articulations; however, they are not really naturalistic articulations. Since the pictures they utilized are generally gathered from web, film or dramatization scenes, TV and so forth so the entertainers in these recordings are specialists to speak.

1.4 Emotions With CNNs

In recent years, the concept of deep learning is progressing and has been made an image classification. Convolutional neural networks (CNNs) is an artificial neural network type which was proposed in 1988 by Yann LeChun. These networks are one the most famous neural network architecture for image classification, segmentation and recognition. It works like a human brain using artificial neural network which consist on different hidden layers with one input layer and one output layer. These network neurons take input from an image, then multiplies those input with weights, adds some base and then applies different activation functions to get the output. Neuron can be used in image classification, segmentation and image recognition. A high accuracy model can be achieved by inserting more data in neural network. In this research, researcher used part of a model graph named R-CNN with CNN and then added some more layers on top [11].

1.5 Neural Network

Neural community is a complex, nonlinear, parallel machine together with smaller, less complicated, interconnected processing devices which can carry out computations. Neural networks may be changed with converting the electricity of every connection among processing devices to reap a preferred output through a manner known as studying. Neural networks routinely extract functions from education units in the course of the studying manner.

With the capacity to generalize, the skilled neural community are required to classify new entered records into the skilled output classes. There are numerous houses of neural networks. These could be either linear or nonlinear. The significance of nonlinearity is whilst the community is predicted to compute a preferred output from a nonlinear enter sign which include a speech sign. A neural community commonly maps a fixed of inputs to a fixed of outputs (enter-output mapping), wherein a fixed of education examples includes detailed specific enter alerts and corresponding preferred output reaction. At random, the education examples are selected from a fixed and fed to the community. The electricity of connections among nodes of the community are then changed to limit the distinction among the preferred reaction and the real reaction produced with the aid of using the community. This manner is repeated till the distinction among preferred reaction and real reaction is 0 or minimal, as a result accomplishing a steady-nation for the community. Neural networks are adaptive, as they're capable to adapt their synaptic weights (electricity of interconnections among nodes) to surrounding changes [12].

1.6 Problem Statement

Emotion Detection of human plays an important role in mutual relationships. This is why it needs to be addressed issues related to emotion detection. Automatic emotion detection systems are currently an active research topic from early years. Many advances are made in this field. Extracting and understanding emotions is an advanced communication between machine and human. Model is generalized on our dataset give a better and compact result for seven emotion states and achieve higher accuracy as compared to existing approaches.

1.7 Objectives

- Main objective of the research is to identify the emotional states of humans using deep learning model.
- Past work within the field of Emotion Detection disease detection is analyzed through the literature review.
- Discover a dataset that's substantial and adequate for preparing and validation of framework.
- Compare diverse approaches to solve issue in Chapter 2.

- Discover a dataset that's substantial and adequate for preparing and validation of framework.
- Model is to trained efficiently to acheive desired results.

1.8 Thesis Organization

The study is organized within the taking after sequence.

Chapter 2 incorporates the point by point literature review which acts as a foundation consider to created investigate questionnaire, speculation and objective of the study. Added with, this chapter will talk about and dissect prevalent Emotion detection mechanism, evolution of CNN Designs over time, and summarize the past work done concurring with distinctive ML approaches.

Chapter 3 examines in length the methodology embraced and inquire about approach utilized,literature assets utilized, research strategies consolidated , investigate parameters and hypothetical framework.

Chapter 4 analyses the proposed arrangement by to begin with examining its usage on distinctive dataset and last section contains comes about.

Chapter 5, work is concluded and future suggestions are prescribed.

1.9 Motivations

Most Commonly prediction that is acceptable whose computing leads to the background, making human user prominent into the foreground by interlacing itself to daily lives. In the future, there will be human-oriented user interfaces that can be in a more natural way to the multi modal communication with humans both verbal and nonverbal. Intentions and Emotions can be recognized by these interfaces and can understand human in social and effective manners.

Facial expression is one of the foremost compelling, of course, fabulous implies for people to communicate their sentiments and feelings to clarify and to strain their understanding, disagreements, and eagerly. As already specified the applications of facial expressions acknowledgment encompasses a tremendous extend, beginning from HCI, mechanical autonomy, and security to the facial picture combination for sex change and distinctive age gather fusion (Bettadapura 2012) etc. Subsequently, programmed acknowledgment of facial expressions forms the substance of different next-generation computing devices counting emotional computing technologies, brilliantly mentoring frameworks, and persistent profiled individual wellness monitoring frameworks etc.

Mainly two trends of facial analysis most common in researchers which are facial fact emotion recognition and facial muscle activity recognition. facial muscle activity which is also known as action units is mostly used to detect basic emotions. there are many limitations of AU which are reported in the literature review [13] and [14]. Classification based model is used to recognized seven basic emotions (anger, disgust, fear, happiness, neutral, sadness, surprise).

1.10 Research Contribution

The significant unique contribution of this dissertation is displayed as follows. Advancement of new modern facial expression recognition frameworks using latest deep learning architecture which is capable of recognizing facial expressions with tall accuracy and have the capacity to learning capability so that this framework can alter in any real world environment and culture.

Chapter 2

Literature Review

Emotion is a vital, complex, and broad research topic within the areas of biomedical building, psychology [15], neuroscience [16] and health [17]. Feeling discovery is an important investigate region in the biomedical building. Thinks about in this range center on anticipating human feeling and computer assisted conclusion of mental clutters. There are different strategies in writing to identify enthusiastic states such as electroencephalography (EEG), galvanic skin response (GSR), discourse investigation, facial expression, multimodal, visual scanning behavior [18].

2.1 Emotions Detection with Existing Approaches?

2.1.1 Convolutional Neural Network

Convolutional neural network (CNN), is an artificial neural network class that becomes dominant in the computer vision field for simple and complex tasks, a variety of domains are attracted toward it like radiology. using back propagation and building blocks to learn and adopt hierarchies of features using different layers like convolutional layers, pooling layers, and fully connected layers. The aim of this is to target basic CNN concepts and their applications for various radiological tasks and discuss some challenges and future directions in the field of radiology. Small dataset and over fitting is a challenge in applying CNN. Familiar with basic concepts and focal points of CNN as well as impediments of profound learning is fundamental in arrange to use it in radiology investigate with the objective of moving forward radiologist performance and, eventually,care of patient [19].

2.1.2 Emotion Detection Using CNN

Deep Learning based Architecture is used to performing emotion detection through facial expression. Its focus is Google facial expression data set because it is most famous data set across the globe. “Emotion is an attitude, expression thought or judgement promoted my feelings”, which is core focus to analyze in this system. This system proposed to highlight the hidden emotions in images. This system uses machine-learning techniques, adopted by previous system with names like CNN architecture LeNet, and maximum entropy approaches. This system requires very accurate high data sets, it merges three different data sets and trains Let architecture for classification of expressions. accuracy of system is 96 percent and validation accuracy is 91 percent with 7 seven different classification of expressions. Evaluation Matrix shows the results of this architecture that is accurate in some following emotions i.e. surprising, fear, neutral and emotionless state. Future work can be done on remaining three emotion states which are happy, sad and anger to improve the accuracy for these states [8].

CNN based model is used to extract facial expressions and proposed a complexity perception algorithm for facial expression recognition. This model divides data sets into two classifications, an easy classification sample sub space and complex classification sample sub space by evaluating complexity of facial expression. The effectiveness of this algorithms on Fer2013 and CK+ data sets show the effectiveness of this algorithm over other approaches. This study claims that they have out-performed the other state of art approaches in term of mean recognition accuracy [20].

Both in Computer and behavioral sciences, facial expression is the basic research area. Through this, humans and machines can easily identify certain emotions. Recent Studies shows that some of the facial region have principle facial emotion features in it. It is easy to detect expressive emotions like fear and happiness. Utilizing this perception as a beginning point for examination, we additionally look at the viability with which information of facial highlight saliency may be coordinates into current approaches to computerized FER. Particularly, we compare and assess the precision of ‘full-face’ versus upper and lower facial range convolutional neural organize (CNN) displaying for feeling acknowledgment in inactive pictures, and propose a human centric CNN progression which employees territorial picture inputs to use current understanding of how people perceive feelings over the confront. Assessments utilizing the CK+ dataset illustrate that this chain of command can improve classification precision in comparison to person CNN designs, accomplishing in general genuine positive classification in 93.3 percentage of cases. The proposed system investigated the assignment of facial expression acknowledgment and pointed to the use of behavioral information of human visual recognition to empower improved classification of pictures of faces. The proposed utilization of territorial inputs for CNN learning, tested with single and progressive demonstrating approaches, and examined the effect of picture prepossessing and information expansion. Proposed strategies might be broadly connected over an extent of HCI spaces, counting versatile client interface advancement, ease of use testing, and temperament following [21].

2.2 Facial Expression Databases

2.2.1 JAFFE Database:

JAFFE is a Japanese Female Facial Expression Database taken from publicly available data having 212 facial expression images which consist of ten subjects of female of japan. Analysis can be performed on six basic and one natural emotion which is sad, happy, angry, disgust, fear surprise, and natural. Each subject contains 3 to 4 images per expression. Resolution of images are in the form of gray-scale. Images have been taken under the same light and simpler background condition there is no such occlusion as hair or glasses. The resolution of each image is 256 by 256 and all are in the front view. You can see a sample in below figure containing all sen expression.

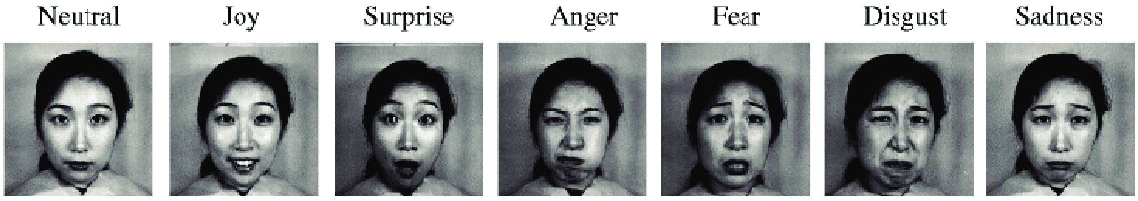


Figure 2.1: JAFFE dataset for image expression [1]

2.2.2 KDEF Database

KDEF is Karolinska Directed Emotional Faces dataset which is also publicly available consisting of 4900 images for facial expression. Each individual contains 70 images for seven different subjects. Smaple images can be seen in fig 3.1.

2.2.3 CK+ Database

CK+ is an Extended Cohn-Kande database that contains 593 videos with a total of 213 subjects. The range of subjects varies from 10 to 50 years old including multiple genders and heritage. Each video shows a natural and target peak expression. It is recorded at 640x490 pixels in 30 frames per second. From all videos only seven expressions were labeled which is 327 in numbers. CK+ database is widely used for facial expression classification methods and it is controlled by the laboratory- facial expression classification. Sample Image can be seen from below figure.

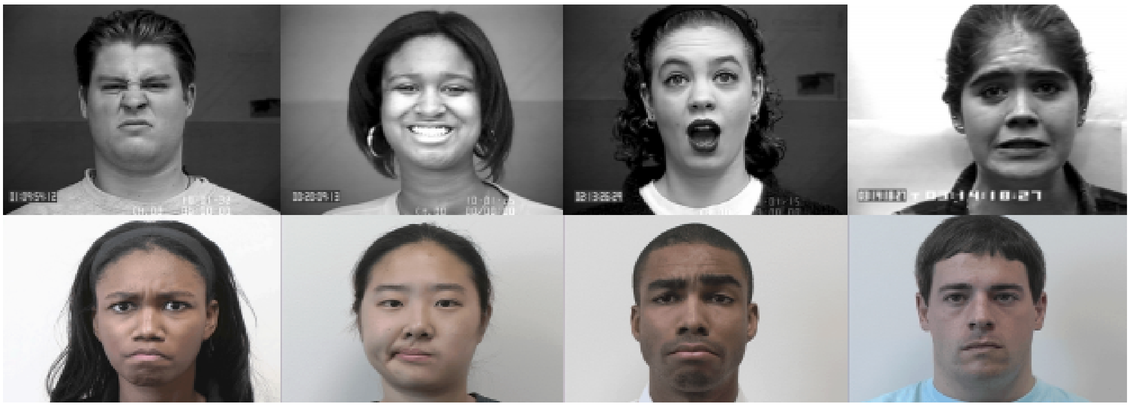


Figure 2.2: CK+ database images sample [2]

2.2.4 FER 2013

An open-source dataset was initially created for an ongoing project by Pierre-Luc Carrier and Aron. Once results are tested for the project then shared publically for the Kaggle competition. The dataset contains 35k images having 48x48 pixels with seven emotion states. Sample Images can be seen from below figure.

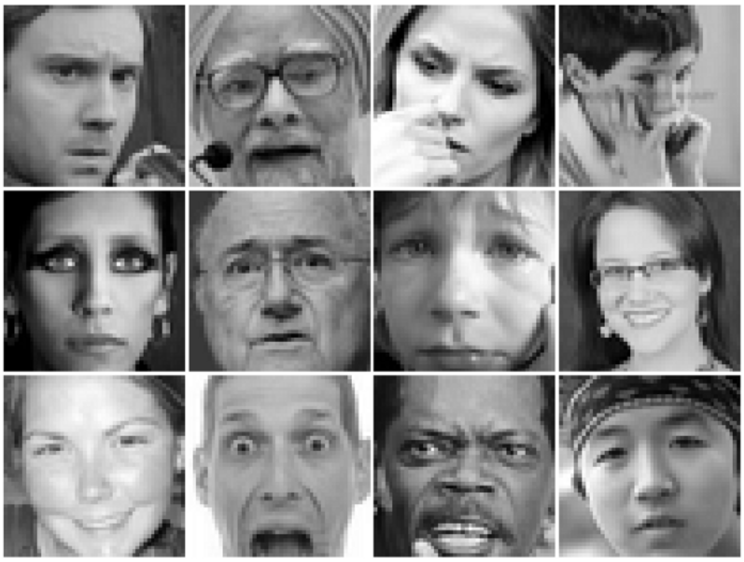


Figure 2.3: FER 2013 images sample [3]

2.2.5 CMU + MIT Database

CMU+MIT Database is a database of facial expression database which is consist of 50 images for various facial expressions, complex backgrounds, different angles, and under different lighting conditions.

2.2.6 Facial Action Coding System (FACS)

FACS is a system to decide human facial expressions, initially created by Paul Ekman and Wallace V. Researchers Proposed muscles activity based feature for emotion detection.Muscular activities are extracted by observing the locations of facial feature focuses in an expression video. Muscular regions initiated facial features of influence in the first video frame. Further points are tracked in sequential flow though optical flow. 3D orientation of a head model are founded by feature point displacement on image and relative displacement of its vertics. Human skin modeled as linear system equation. The predicted deformation of model can generate over fitted system that facial atanomy constraints can be solved under this to get the muscles levels needed for activation. Researchers apply forward feature selection for the selection of most descriptive muscles set for the recognition of facial expression.through SFS the most descriptive features among muscle activities are chosen. For the six basic expressions, we get the 85.9 percent classification performance with having only 11 features. using CK + database our classification get 75 percent accuracy for basic expression using 9 features[13].
Computer Expression Recognition Toolbox software which is used for fully automatic real-time facial expression recognition and it was officially free to use for academic purpose. Facial Action Coding Unit (FACS) used to code the intensity of 19 different facial actions. 3D orientation and 10 features of facial expression can also be determined. CERT achieve an average performance for recognition of facial expression with dataset CK+. with spontaneous facial expression dataset, CERT gets 80 percent accuracy.Furthermore, previous studies traditionally use two methods or approaches to emotion classification.

- Judgment approach.
- Signed-based approach.

Six basic emotion categories are address by the judgment based approach. Sign-based approach uses FACS System, EAU used to further categories emotion expression based upon their characteristics. Generally, this study gave confirmation of guideline that fully automated facial expression acknowledgment at the show state of the art can be utilized to supply real-time input in automated mentoring frameworks. The recognition framework had the option to separate a sign from the face video continuously that given information roughly inside states appropriate to instructing and learning. [14].
This research shows spontaneous facial expressions are different from posed expressions in both manners in which muscles are moved and in dynamic movements. In the field of automatic facial expression, the measurement will require development and assessment for spontaneous behavior. Preliminary results on a task of facial action detection are presented in spontaneous facial expression. Using Facial Action Coding System (FACS) an automatic system is employed for real-time recognition. This system automatically detect face from

video with 20 action units per frame. Support vector machine and Ada boosts machine learning approaches were used for classification of images. The framework displayed here works in genuine time. Face discovery runs at 24 frames/second in 320x240 pictures on a 3 GHz Pentium IV. The AU recognition step works in less than 10 msec per action unit. The following step within the improvement of programmed facial expression recognition frameworks are to start to apply them to spontaneous expressions for genuine applications [22].

2.2.7 Emotion Classification

Six basic emotions mentioned in [23] by paul Ekman including surprise, fear, anger, disgust, sadness, and happiness. Emotion for each class is expressed by face parts which are eyes-lids, brows-forehead, and lower face.

2.2.8 Emotion as Sentiment

Emotion are precepted as an ability to know about someone's facial expression and movements. Different theories reflect multiple aspects of perception about emotions. This is the first study which use emotion words as semantic to explore the point where change in image is detected. They have claimed that this was their first kind of experiment which manipulates two different theoretically derived emotion influences on ability to detect change in emotion. Our comportment findings across different experiments provided evidence for both effects. we manifest a shift in the N400 which was accordant with emotion words shrinking the within-category variation. Results shows how emotions working in different ways effects emotional information in place. It was also proved that emotional words work in different ways to affect emotional information which is placed into detached emotion categories. Language has also proven to be a main source of information which effects emotional perception by just labeling its output [24].

2.2.9 Prepossessing

Preprocessing is a method used to enhance the dataset by doing some adjustments or improvement into it. The fundamental point of preprocessing is to decrease the changeability of information angles that don't identify with feelings that are being shown on the face. Facial expression having emotional recognition comes under social signal processing used in various fields specifically for computer and human interaction. Most researchers use an automatic emotion recognition system which is base on the machine learning approach but having this still, there is a challenge for recognition of emotional states which are sad, happy, fear, angry, disgust, and surprise in computer vision. In recent times deep learning gained the most attention to solve real-world examples. This research enhances convolutional neural networks for the recognition of six basic emotions and preprocessing methods are compared based on cropping, resizing, adding noises, face detection, and data normalization.Based on the result gained from experiments we have explained that cropping and face detection to get a region of interest produces the best results for the improvements of CNN performance [25].

2.2.10 Scaling

The resulting cropped image cannot be considered as a consistent image size because of camera placement and subject size face. Each new image is preprocessed to the required standard size but in DCT calling is not an major issue. Principal Component Analysis sholud be same for each image. In each row input data should have same length to produce a matrix otherwise IDM can't be generated. Multiple image sizes are tested which are 64 by 64, 128 by 128, and 256 by 256. Preprocessing used for both PCA and DCT, effects can be seen using scaling as presented in figure 2.4 [1].

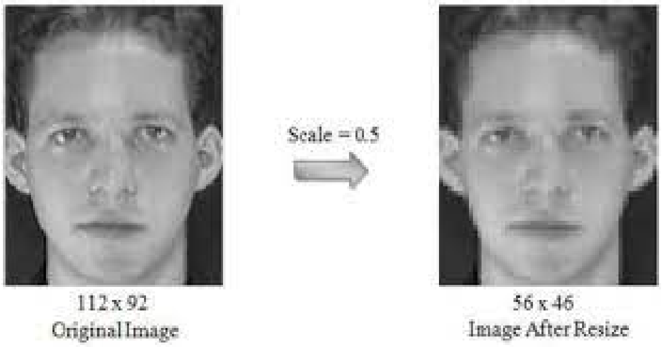


Figure 2.4: Original Image Resize with Scale 0.5 [1]

2.2.11 Patching

This research aimed to investigate local features either these are useful to include in system accuracy. These local features are manifested using patching. This predefined window is used for picture sampling. Center point that cropped itself out as square was used as facial point.Once we get results, original image is further broken into frames of different sizes and each frame can be analysed separately as present in the figure. 2.5.

2.2.12 Normalization's

Either input is at the patch, scaling, or pre-processing stage, normalization used to refine image brightness. the basic reason behind changing the brightness of the image is to reimburse the image for different lighting conditions and skin tones as well. Cohn-Kanade database includes a variety of skin tones which increase data variability. To balance image

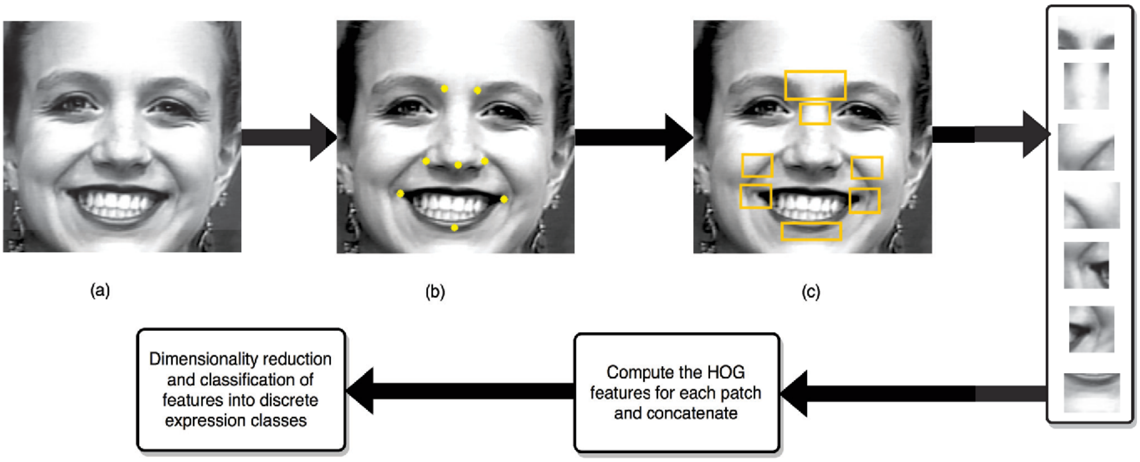

(a)
(b)
(c)
Dimensionality reduction and classification of features into discrete expression classes
Compute the HOG features for each patch and concatenate

Figure 2.5: Examples for facial patches [2]

brightness a constant value is either added or subtracted from the original image. Value of constant c is calculated by the mean of pixels in the image twice. In case if image is too bright then value of c will be a positive number and will be subtracted from each pixel value. We have 2D matrix as a result which have 127 brightness for each image. From figure attached below we can see that how lightning and skin tone influences are minimized. [3].

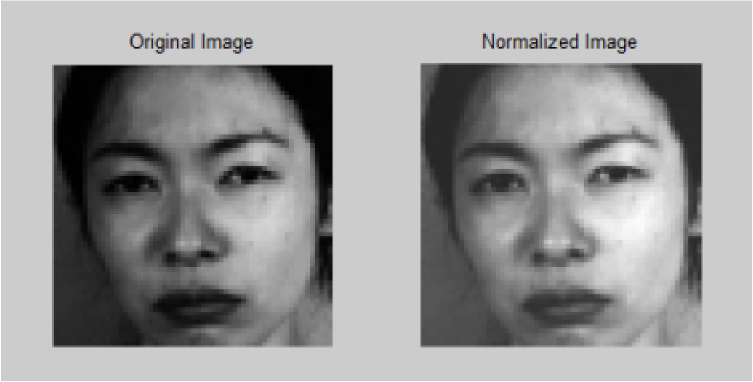


Figure 2.6: Image Normalization [3]

2.2.13 Real-time Emotion Detection Applications

Currently, the Real-time emotion detection recognition system has an active research field for the past several decades. This research focuses on people who are physically disabled like deaf, dumb, bedridden, and autism children's expression based on ECG signals by using (LSTM) long short time memory and (CNN) Convolutional Neural Network by an algorithm developed used for realtime emotion detection through optical flow using a virtual marker which is even work in the uneven lighting situation, having different background and multiple skin tones. Users need to wear an EPOC + headset and their face in front of the camera to record raw ECG data and collect virtual landmarks which are ten in quantity placed on the subject face. From the results, we can conclude that system has 99.81 percent accuracy of facial expression and in ECG Signals it has an accuracy of 87.25 percent [26]. This research shows that distance is an emotion regulation type which involve such counterfeit is new. It also involves new standpoints after psychological distance and emotional impact of a impetus. The efficacy and versatility of distancing relating in respect to other regulations make it very promising application tool for medical applications. This technology has unclear effective tool and the main method used in different studies makes it difficult for user. Through this research we propose a codification of distancing within the comprehensive context of emotion regulation strategies by reviewing effects of this strategy and offering a preliminary neurocognitive model which is explaining neural bases key processes. Proposed model is combination of three different components which are self-projection, effective self-projection and cognitive control. This system aims to revise terminology of research in emotion regulation progress but in this research, it only focuses on distancing and their related concepts to find meaningful distinctions. Results of this framework may encourage to produce fundamental theories and models like CLT, specifically comparing the mechanism of different forms may saturate the validity of these categories in CLT. This system provides a transition which focuses more on cumulative progress of a research community then an individual study. With the effectiveness of several research regulations, emotion methods identified some approaches which can improve potentially distancing application. These techniques include consideration of, whether certain distancing techniques are effective than others. Also, identifying contextual factors and individual techniques which are greater in size and impact then individual techniques. Another technique of improving distancing is enhancing neurocognitive [27].

Feeling acknowledged plays a vital part in human-machine interaction system. To incorporates findings of this research, curiously facial locales were identified in pictures and were classified into one of seven classes: irate, nauseate, fear, cheerful, unbiased, pitiful, and astonish. Although many breakthroughs have been made in picture classification, particularly in facial expression acknowledgement, however, this concept of investigating region is still challenging in terms of wild testing environment. In this paper, multi-level highlights were used in a convolutional neural network for facial expression acknowledgement. Based on proposed perceptions, different network connections were presented to make strides in the classification assignment. By combining the proposed network connections, suggested strategy accomplished competitive results [28].

The convolutional neural network (CNN) based methods have made noteworthy advancement in numerous computer vision tasks, such as protest discovery, confront acknowledgement, and so on. Their exceptional capabilities are in part due to the investigation of the unstable development of preparing set sizes. So those computer vision errands with moderately little training sets accessible, like facial expression acknowledgement, are still very challenging. In this work, a viable data augmentation system has been portrayed to synthesize large-scale training samples for the errand of facial expression acknowledgement generally. An unused misfortune work, named cluster loss, has also been propose to form profound highlights compact. Assessed on a recent expression database RAF-DB, suggested strategy accomplishes better performance than state-of-the-art baselines. It also outperforms methods focusing on this database. On RAF-DB, the best execution has been achieved by utilizing two techniques in combination, demonstrating the adequacy of two proposed approaches [29].

Facial expression acknowledgement has been dynamic way to inquire about region since long time, with developing application ranges counting avatar movement, neuromarketing and amiable robots. The recognition of facial expressions isn't a straight-forward issue for machine learning strategies, since individuals can change essentially within the way they present their expressions. Indeed, pictures of the same individual within the same facial expression can shift in brightness, foundation and posture, and these varieties are emphasized if considering diverse subjects (since of varieties in shape, ethnicity among others). In spite of the fact that facial expression acknowledgement is exceptionally examined within the writing, few works perform reasonable assessment maintaining a strategic distance from blending subjects whereas preparing and testing the proposed calculations. Consequently, facial expression acknowledgement is still a challenging issue in computer vision. In this work, a straightforward arrangement for facial expression acknowledgement has been proposed that employees a combination of Convolutional Neural Organize in particular [30].

One of the foremost effective social signals, facial expression helps to get it each other's inside feeling in communication. Analysts conclude that individuals attempt to communicate the same feeling with the comparable facial expression even though if its from different race, culture and religion. This conclusion provides possibility of creating the enthusiastic computing framework by facial expression. A two-stage system based on DCNN has been proposed through this research to recognize facial expression for social flag analysis. Considering the facial expression's non stationary nature, the proposed system contains two stages: within the time to begin with organize the neutral expression outline and completely expression outline which are consequently picked from the facial expression arrangement by the Soft Max score of the double CNN. At that point within the moment organize, the selected impartial expression outline and completely expression frame are nourished to the DCNN individually [31].

2.3 Comparative Analysis

The following competitive analysis depicts a view of the techniques from the literature review. These techniques are used in different ways to predict various phenomenon. The following table elaborates accuracy, Technique used, Year and data set including features with respect to the finding of the literature. The table depicts historic time line of the initial discovery year of these techniques. By doing so we understand various aspects required for implementation of one as best fit to the studies requirement.

- Facial Action Coding System
- Multi Class CNN
- Facial Action Units Coding System
- Convolutional Neural Network
- Back Propagation
- Genetic Algorithm

Ref	**Data Set**	**Year**	**Model used**	**Technique Used**	**Accuracy**
[13]	CK+ Database Features 9 Classes 6	2014	Action Units FACS	SVM	75 % with six states
[14]	CMU+MIT Features 10 Classes 7	2016	FAUCS	SVM CERT	97.3% with seven states
[19]	Public Images TCIA	2018	CNN	SVM	High
[8]	JAFEE Custom Dataset High Level Features Mid Level Features	2019	LeNet	K-Mean Clustering SVM	91.3%
[32]	FER 2013 Global features Local features	2019	Open Face	SVM	85%
[30]	FER 2013	2018	VGG16 Inception Multi class CNN Single MLCNN MNL	Grad-CAM	69% 70% 71% 72% 75%
[11]	JAFEE	2016	One Classifier	Spacial Norms Down Sampling Up Sampling	85%
[24]	RAF-DB	2018	3DMM Model	Data Augmentation Feature Learning	85%
[27]	Custom Dataset	2018	NH400	word stumli	75%
[20]	CK+ Dataset BU 4DFE	2019	DCNN	SVM Classifier	75% 74%
[28]	JAFEE	2018	CNN	Data Augmentation	84%
[30]	FER 2013 JAFEE	2017	ResNet	CPC Algorithm	71% 97%

Chapter 3

Proposed Methodology

This Chapter is divided into two sections. First Section consists of the detail of the data sets which are used for the generalization of model and second section contains the detail of proposed methodology.

3.1 Datasets

Two Datasets having the different backgrounds and captured in several situations are used for the generalisation of the model. So that demonstrates prepared on these images can perform better for pictures taken in totally different situations.

3.1.1 FEC Dataset

This dataset consists of face image triplets along with annotations that specify which two faces in the triplet form the most similar pair in terms of facial expression. To the best of our knowledge, there is no existing large scale face dataset with such expression comparison annotations.

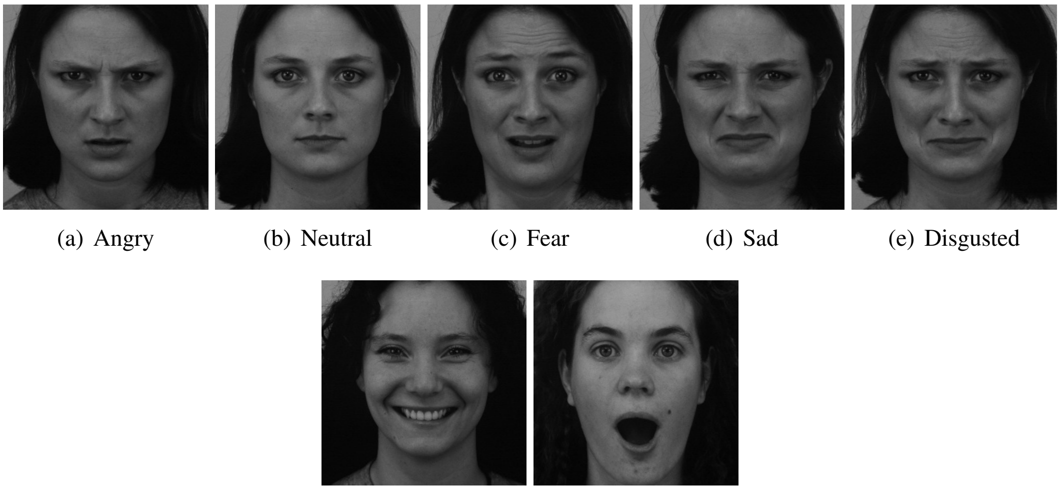
3.1.2 KDEF Dataset

Karolinska Directed Emotional Faces (KDEF)[1] dataset consists of a set of 4900 facial expression images. It contains 70 individuals, each displaying seven different emotional expressions, each expression being photographed (twice) from five different angles.

3.2 Proposed Methodology

Most of the progression if this research was based on experiments and central ideas and the architecture used which is better then previously used LeNet architecture which has

[1] https://www.kdef.se/



(a) Angry (b) Neutral (c) Fear (d) Sad (e) Disgusted

(f) Happy (g) Surprised

Figure 3.1: Seven different emotions in the KDEF dataset

problem of over-fitting in some case and has less accuracy in test dataset. System yielded higher accuracies were explored further and added to the overall system. The result was a system that combined many finely tuned and successful elements.

3.2.1 Xception model

Xception Model is proposed by Francois Chollet. Xception is an extension of the inception Architecture which replaces the standard Inception modules with depth wise Separable Convolutions [33]. Xception model consist of 36 conventional layers which forms feature extraction network architecture. These layers consists of 14 modules which makes linear residual connection around them expect from modules which are in last and first position.In brief, the Xception design is a straight stack of depthwise distinct convolution layers with residual associations. This makes the architecture exceptionally simple to define and alter. it takes only 30 to 40 lines of code using a highlevel library such as Keras or TensorFlow [34].

3.3 Exception with depth wise Separable

3.3.1 Original Depthwise Separable Convolution

Depthwise convolution is followed by a single point wise convolution in original depth wise convolution.

- convolution network has channel wise nxn spatial convolution network. By assuming attached figure below, we can see that we have 5 channels, at that point we have also 5x5 spatial convolutions.

- Pointwise convolution really is the 1×1 convolution network to modify the measurement.

when we compare it with conventional convolution, we don't get to perform convolution over all channels. This means the number of associations is less and the demonstration of the model is lighter.

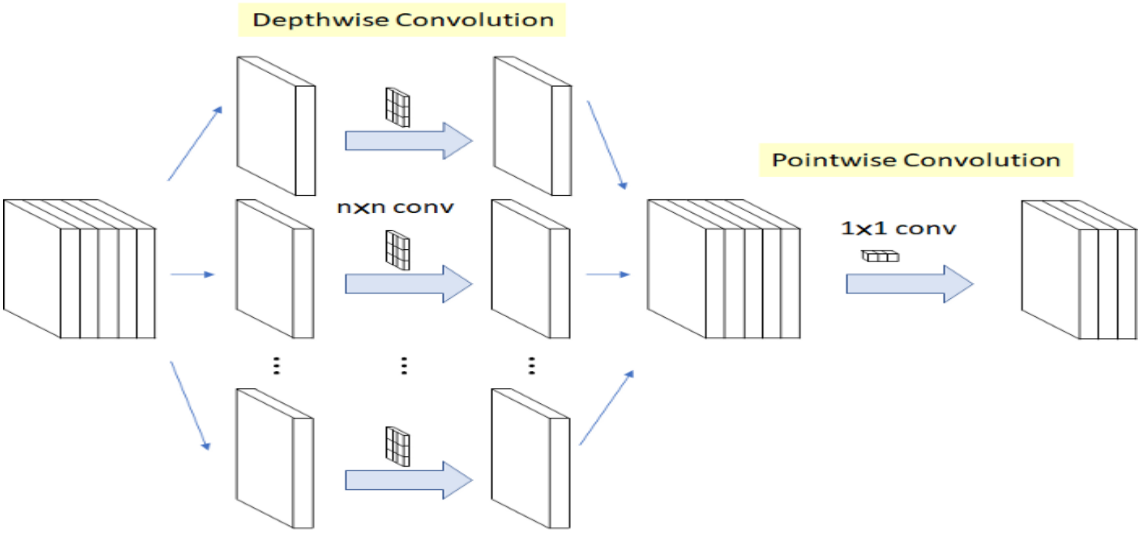


Figure 3.2: Xception: Original Depthwise Separable Convolution [4])

3.3.2 Modified Depthwise Separable Convolution in Xception

In modified depth wise separable convolution point wise convolution is followed by a depth wise convolution. This activation is motivated by inception module which is followed like first 1x1 convolution is performed then nxn spatial convolution is performed. Hence, this approach is a bit different from original approach. (n=3 In Inception-v3 3x3 spatial convolution is used).

3.3.3 Modified Depth wise Separable Convolution in Xception

In modified depth wise separable convolution point wise convolution is followed by a depth wise convolution. This activation is motivated by inception module which is followed like first 1x1 convolution is performed then nxn spatial convolution is performed. Hence, this approach is a bit different from original approach. (n=3 In Inception-v3 3x3 spatial convolution is used).
Differences between these two is as follow.

- operations order: It is specified, with the first depthwise distinguishable convolutions as ordinarily executed (e.g. in TensorFlow) firstly channel-wise spatial convolution

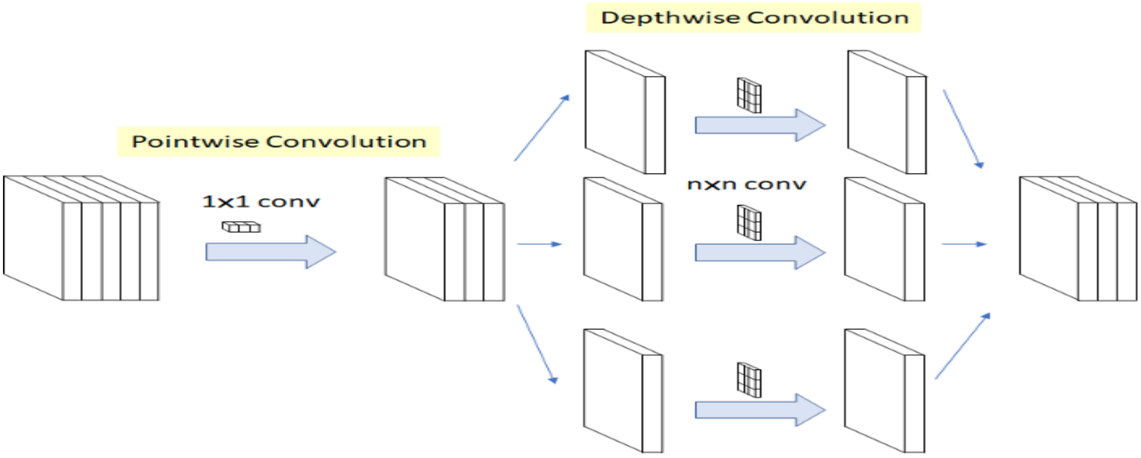


Figure 3.3: Xception: Modified Depthwise Divisible Convolution utilized as an Initiation Module in Xception, so called "extreme" form of Initiation module with filter size (n=3 here) [5])

is performed and after that it performs 1×1 convolution and the modified depthwise separable convolution performs 1×1 convolution to begin with at that point channel-wise spatial convolution. Typically claimed to be insignificant since when it is utilized in stacked settings, there are as it were little contrasts showed up at the starting and at the conclusion of all the chained beginning modules.

- Non-Linearity Presence/Absence: there is non-linearity within the original intuition module. Non linearity begin with operation. With Xception Network, the modified depth wise distinct convolution, there's NO middle ReLU non-linearity.

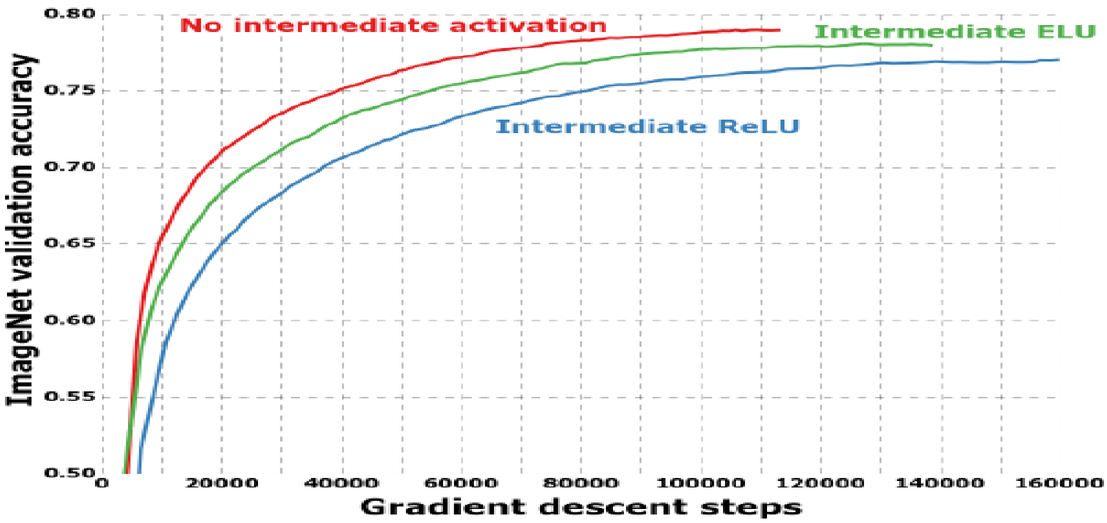


Figure 3.4: Xception: Modified Depthwise Separable Convolution

From above figure you can see that unit testing was performed. it can be seen , as compared with other network Xception without any intermediate activation function hase more evaluated accuracy as compared with the using one of the activation function wither Relu or ELU activation function.

3.3.4 Overall Architecture of Xception

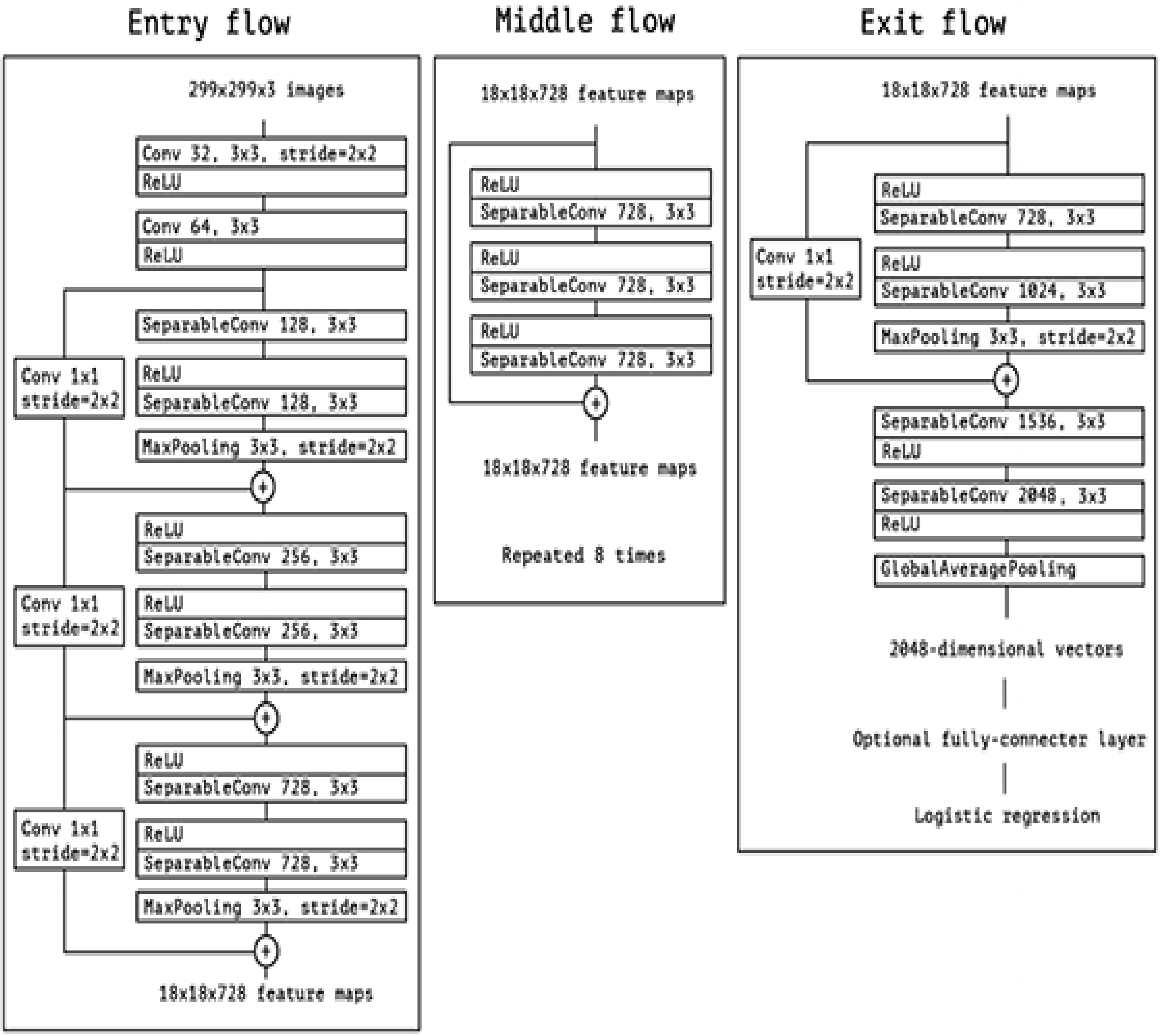
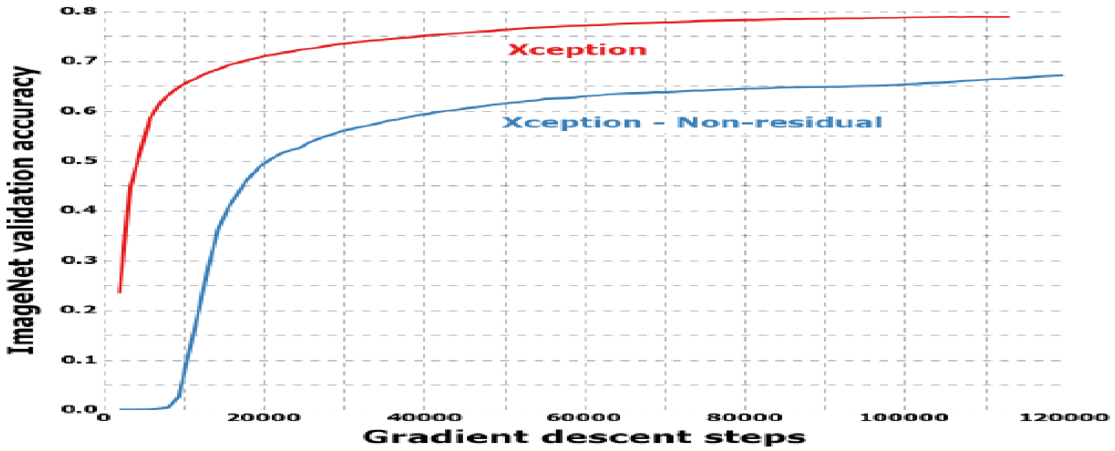


Figure 3.5: Xception architecture(Entry Flow=>Middle Flow=>Exit FLow))

From above fig we can see, seperable convolution network is depth wise modified network. Whole deep learning architecture is placed as inception modules can be seen in above figure. ResNet Proposed residual connections also known as (shortcut/skip) connections which are used in every flow in network.



resudial

Figure 3.6: ImageNet: Gradient Descent Steps Validation Accuracy

From above figure it can be seen Xception version for non residual connection can be tested using some residual connection. Can be seen from the above image that the accuracy is much higher then with using residual connection which made him highly important.

3.3.5 Comparison with State-of-the-art Results

Two Data-sets were basically used for the comparison which are ILSVRC and JFT. On ImageNet dataset which consist of fitheen million high-resolution images with labels and having 22,000 categories for each image. 1000 categories for each 1000 images were used which are subset of ILSRVC dataset. Each Dataset contains around 1.3 million images for traning, 50,000 images for validation and 100,000 images used for testing purpose. All models including VGGNet, ResNet, and Inception-v3 were outperform by Xception. it should be noted that performances was measured in term of error rate rather then accuracy, overall relative accuracy is not very small.

		Top-1 Accuracy	**Top-5 Accuracy**
Runner up in ILSVLRC 2014	VGG16	0.715	0.901
Winner up in ILSVLRC 2015	RestNet-152	0.770	0.933
Runner up in ILSVRC 2015	Inception V3	0.782	0.941
	Xception	**0.790**	**0.945**

Table 3.1: ImageNet: Xception has the highest accuracy

As we can see from the above images that Xception outperform another modals like VGGNet, ResNet, and inception V3.The most important point which cannot be negotiated, that overall improvements are not very small contribution in term of error rate. Of course, You can see from the above figure, Xception has way better exactness as compared with VVGNet and Inception-v3 along side the gradient descent steps. When we use non-residual connections to compare with inception-v3, Inception-v3 was outperform by Xception. Question raised that either it is a way better to have a remaining form of Inception-v3 for reasonable comparison? Besides, In both Depthwise Distinct Convolutionand Remaining Connections Xception truly make huge difference in term of accuracy.

Model	Parameter Count	Steps/Seconds
Inception V3	23,626,728	31
Xception	22,855,952	28

Table 3.2: Model Size/Complexity.

Xception have comparable show measure with as compare with Inception-v3.

3.4 JFT — FastEval14k

Google introduced JFT for the categorization of Large scale images, at the outset which includes 350 million high resolution pictures compressed on with names from a set of 17,000 classes. A subordinate fastEval 14k, dataset is being utilized. 1400 pictures with solid information from almost 6000 classes are comprised by this Dataset. A cruel accuracy forecast (mAP) is used to estimate each picture singly which are produced by different objects. Mean accuracy expectations (mAP) is used for the measurement because a great number of objects are appeared up by a single picture thickly.

Model	FastEval14k MAp@100
Inception V3 no FC Layers	6.36
Xception no FC Layers	6.70
Inception V3 with FC Layers	6.50
Xception with FC Layers	**6.78**

Table 3.3: FastEval14k: Xception has highest MAP@100.

Once more, Xception has higher map compared with Inception-v3.

3.4.1 Optimizer

Adam is introduced by the authors in the paper [35]. It is a first-order gradient-based optimization algorithm.It is based on the lower order moment adaptive estimates. This optimization approach has following benefits.

- It is easy to implement,computationally effective and use very little memory.
- This approach is suitable for non stationary goals which have issues of noisy or sparse gradient.

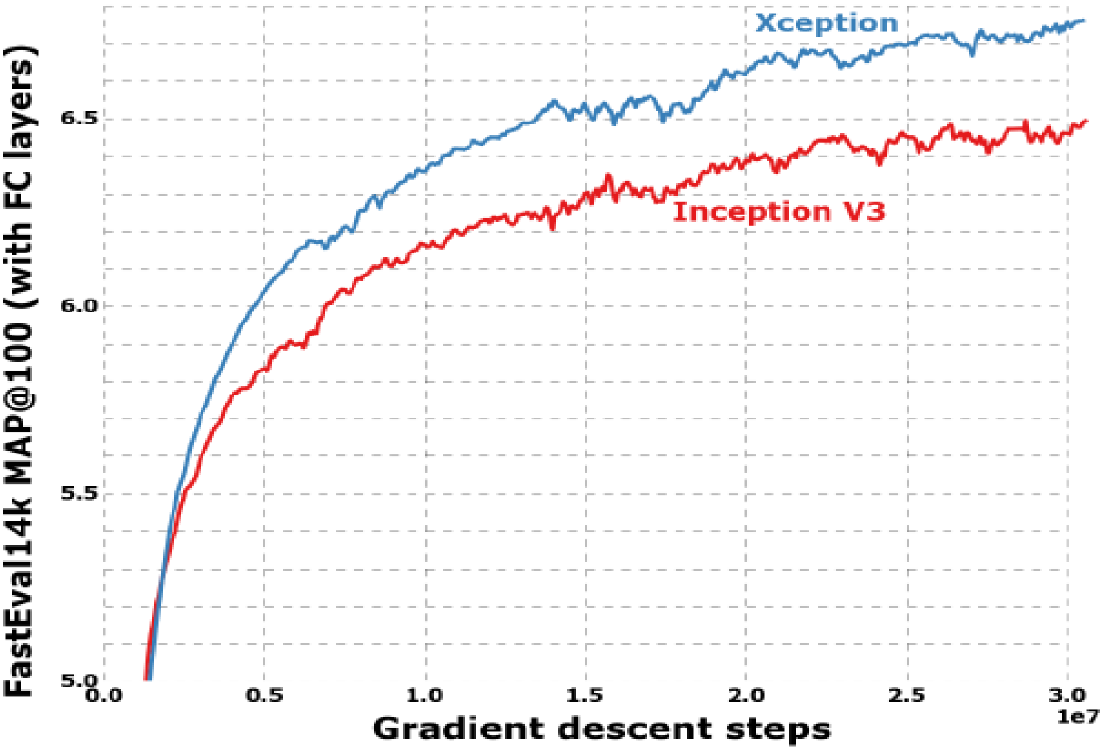


Figure 3.7: FastEval14k: Validation Accuracy Against Gradient Descent Steps

- It is invariant to diagonal re-scaling of the gradients, and it is suitable for large problems both in terms of data and parameters.

- Little tuning is required for the hyper-parameters.

3.5 Evaluation Measures

3.5.1 F-Score

Classification report is generated for results. F-score is used to evaluate the results of model. F1-score is the harmonic mean of precision and recall measure.The perfect model has F score of 1. The formula for the standard F1-score is the harmonic mean of the precision and recall. A perfect model has an F-score of 1.

$$FScore = \frac{Precision * Recall}{Precision + Recall}$$

- Precision is the fraction of true positive examples,amongst the examples classified as positive by the model.

$$Precision = \frac{TP}{TP + FP}$$

- Recall is the fraction of examples classified as positive, among the total number of positive examples. Recall is also known as sensitivity.

$$Recall = \frac{TP}{TP+FN}$$

3.5.2 Accuracy

Accuracy is the number of accurately anticipated information focuses out of all the information focuses. More formally, it is characterized as the number of genuine positives and genuine negatives separated by the number of genuine positives, genuine negatives, untrue positives, and untrue negatives[36].

$$Accuracy = \frac{TP}{P+N}$$

3.6 Activation Functions

So what does an artificial neuron do? Essentially put, it calculates a “weighted sum” of its input, includes a bias, and then chooses whether it ought to be “fired” or not. The activation function does this.

$$Y = Sum(input * weight) + bais$$

Presently, the value of Y can be anything extending from -inf to +inf. The neuron truly doesn’t know the bounds of the value. So how do we choose whether the neuron should fire or not? Since we learned it from science that’s the way the brain works and the brain may be a working declaration of a great and shrewd system. For this purpose, we chose to include “activation functions” for this reason. To check the Y value created by a neuron and choose whether exterior associations should consider this neuron as “fired” or not. Or maybe let’s say — “activated” or not[37].
Different variations of Activation function as follow.

3.6.1 Soft-max Function

Same as Sigmoid function but it is useful when we are dealing with classification problem. Range is 0 to 1. Mostly used in output layer for classification of each class to get the probability of matching class.
The activation function does the non-linear change to the input making it competent to memorize and perform more complex assignments.

Chapter 4

Results

In this chapter, model implementation start from experimental setup to its evaluation on test set is discussed under three different section. Following is the detail.

4.1 Experimental Setup

Google colab is utilized to get ready and test the desired model. It is like Jupiter note pad inside the cloud environment and it handles all setup configuration. By the help of Google Colab, one can compose and execute through the browser. Google Colab allow free GPU and TPU for few hours in a day. Google Colab Proficient is paid benefit which grants superior speed and more memory but it is because it were accessible within the America and Canada. Our craved demonstrate is ready with GPU. Google Colab allow free GPU upto 10 4 hours a day in our locale. So the illustrate is ready step sharp after utilizing day by day GPU free advantage. Illustrate is spared at that point retrained once more on another day or trading between assorted mail accounts on the same day.

4.2 Model Implementation

Xception and Simple CNN is used as model for generalization on two different datasets with following hyper-parameters.

4.2.1 Layers

Xception is a convolutional neural network that's 71 layers profound. It allows us stack a pretrained form of the organize prepared on more than a million pictures from the ImageNet database . The pretrained arrange can classify pictures into 1000 question categories, such as console, mouse, pencil, and numerous creatures. As a result, the model has learned wealthy highlight representations for a wide extend of pictures. The result has an picture input measure of 299-by-299 [38].

4.2.2 Optimizer

Adam optimizer is utilized with the taking after introductory learning rate and weight rot
On ImageNet:

- Optimizer: SGD is used for tuning of the parameters to minimize cost function.
- Momentum: Common to use momentum value close to 1 so here we have use 0.9.
- Initial learning rate: Start from 0.1 to 0.001 we have get good starting point for our problem at 0.001
- Learning rate decay: Decay of rate 0.94 every for 2 epochs to get starting point for our problem.

On JFT

- Optimizer: RMSprop use to minimize cost function.
- Momentum: Common to use momentum value close to 1 so here we have use 0.9.
- Initial learning rate: Initial learning rate: Start from 0.1 to 0.001 we have get good starting point for our problem at 0.001

4.2.3 Loss Function

Mean absolute error (MAE) is used as loss function in model.

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n}$$

MAE = mean absolute error

y_i = prediction

x_i = true value

n = total number of data points

4.3 Evaluation Measure

Classification report of each result of demonstrate prepared and test on the particular dataset is generated.Which incorporate all class accuracy.

4.3.1 Emotion Results With Xception Model Case 1

In this section we have train and validate our model with 1024 batch and get the following results.

4.3.2 Loss and Accuracy

Loss/ Accuracy	Value
Train Loss	0.49
Train Accuracy	82.1
Test Loss	1.19
Test Accuracy	64.3

Table 4.1: Loss and Accuracy of train and validation

As you can see in above table we have following result. On Our Dataset we have got train loss: 0.49 and train accuracy: 82.1 and we got test loss 1.1 and test accuracy 64.3 Xception modal with KDEF DataSet, higest class Accuracy is 86 percent and lowest class accuracy score is 51 percent is achieved which is higest than previous results. Following is the image of the classification report of KDEF dataset.

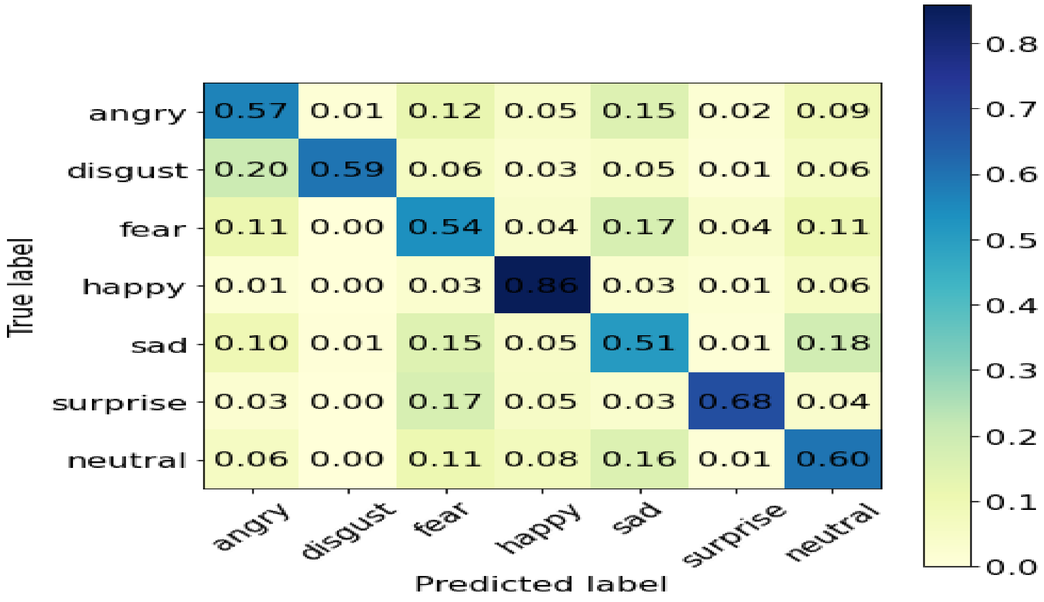


Figure 4.1: Confusion Matrix with Xception Model on KDEF Data Set

4.3.3 Label Validation

Label validation of KDEF dataset can seen in below diagram.

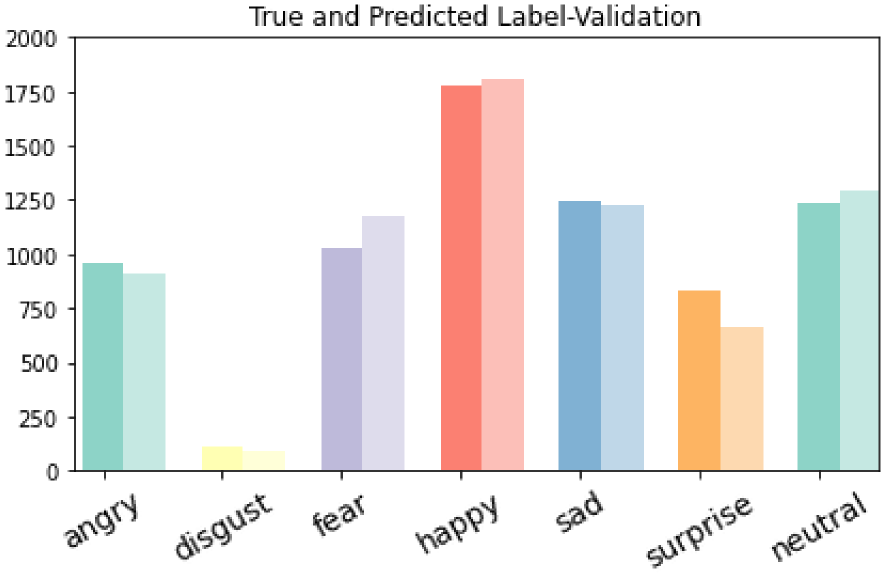


Figure 4.2: Label Validation with Xception Model on KDEF Data Set

4.4 Test Cases

4.4.1 Emotion Results With Xception Model Case 2

In this section we have train and validate our model with 1500 batch and get the following results.

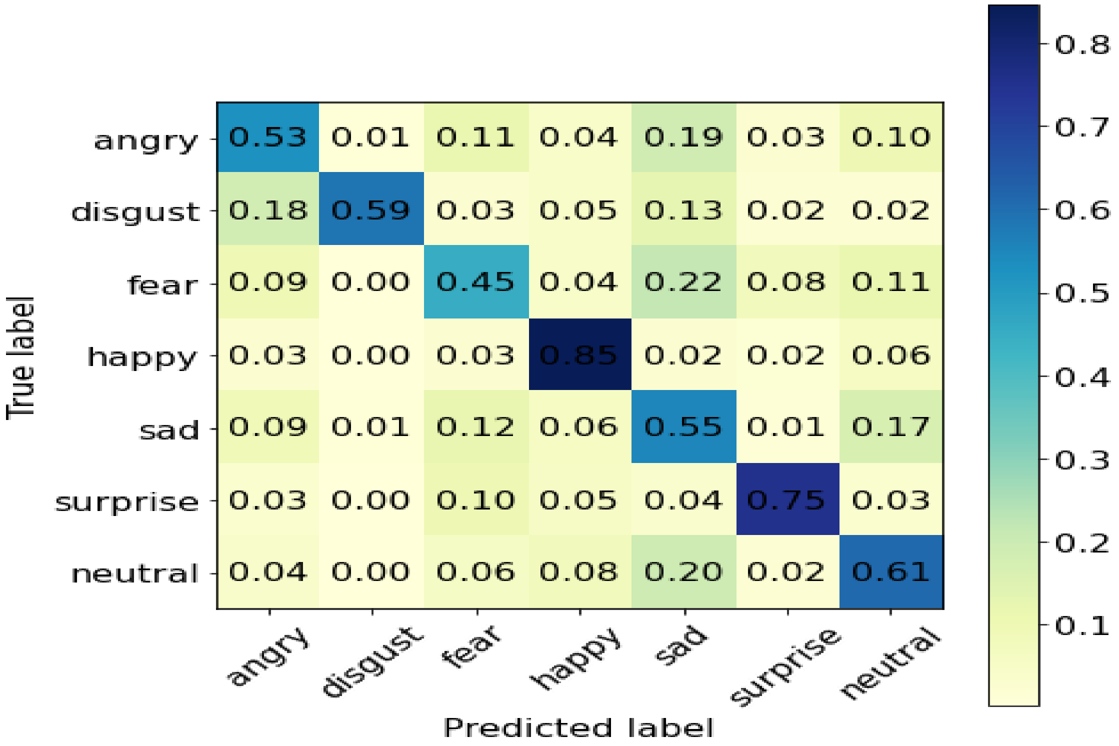
4.4.2 Loss and Accuracy

Can be seen in table 4.2 we have following result. On Our Dataset we have got train loss: 0.46 and train accuracy: 83.1 and we got test loss 1.2 and test accuracy 63.1

Loss/ Accuracy	Value
Train Loss	0.46
Train Accuracy	83.1
Test Loss	1.2
Test Accuracy	63.1

Table 4.2: Loss and Accuracy of train and validation

Xception modal with KDEF DataSet , higest class Accuracy is 85 percent and lowest class accuracy score is 45 percent is achieved which is higest than previous results. Following is the image of the classification report of Citrus dataset. As we can see while we are increassing traning parameters overall class accuracy is getting decreases so we have best class accuracy for all emotion state in case 1.



True label	angry	disgust	fear	happy	sad	surprise	neutral
angry	0.53	0.01	0.11	0.04	0.19	0.03	0.10
disgust	0.18	0.59	0.03	0.05	0.13	0.02	0.02
fear	0.09	0.00	0.45	0.04	0.22	0.08	0.11
happy	0.03	0.00	0.03	0.85	0.02	0.02	0.06
sad	0.09	0.01	0.12	0.06	0.55	0.01	0.17
surprise	0.03	0.00	0.10	0.05	0.04	0.75	0.03
neutral	0.04	0.00	0.06	0.08	0.20	0.02	0.61

Figure 4.3: Confusion Matrix with Xception Model on KDEF Data Set

4.5 Emotion Results with CNN

Classification report of each class result of implementation is prepared and test on the particular dataset is generated.Which incorporate all class accuracy and loss.In this section we have train our data set with CNN based modal on GPU as on local system is not capable of doing in quick time.

In this section we have train and validate our model and get the following results. On Emotion dataset, highest class Accuracy is 85 percent and lowest class accuracy score is 45 percent is achieve. Following is the classification report.

4.5.1 Training and Validation loss

On Emotion dataset, Modal Accuracy loss report is attached of test set. Following is the report.

In this figure the blue line shows training accuracy and loss and orange line shows the validation accuracy and loss on KDEF dataset.

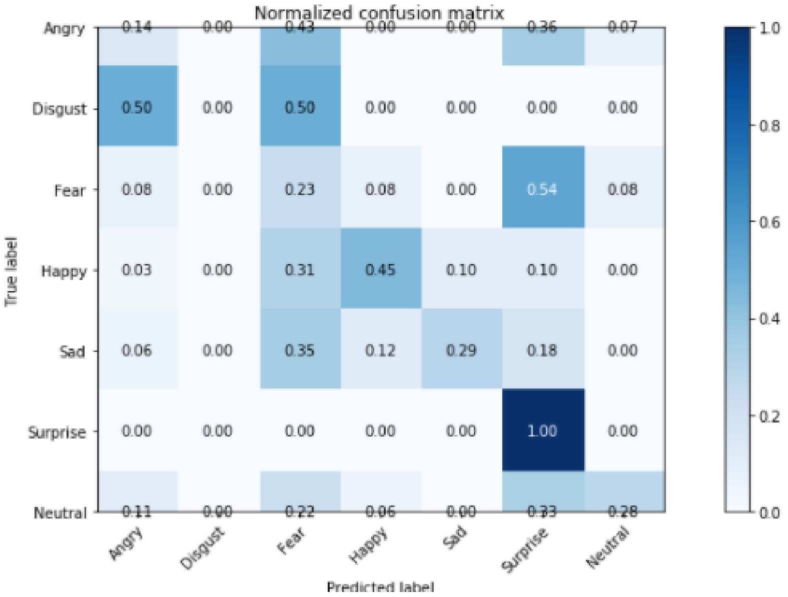
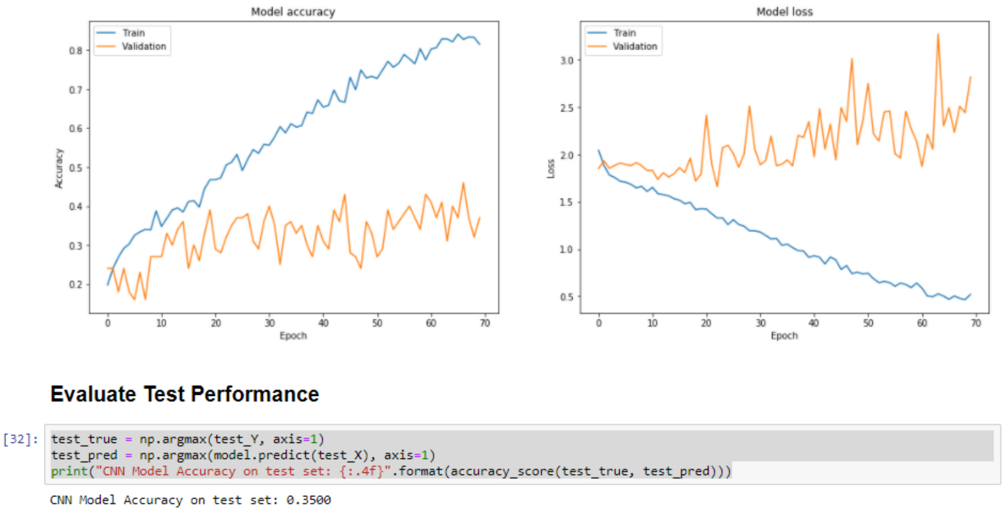


Figure 4.4: Confusion Matrix with CNN Model on KDEF Data Set



Evaluate Test Performance

```
[32]: test_true = np.argmax(test_Y, axis=1)
      test_pred = np.argmax(model.predict(test_X), axis=1)
      print("CNN Model Accuracy on test set: {:.4f}".format(accuracy_score(test_true, test_pred)))

      CNN Model Accuracy on test set: 0.3500
```

Figure 4.5: Modal Accuracy and Loss CNN Model on our Data Set

4.6 Limitations

Major limitation of our setup is the resource availability.Due to this it prevent us to perform desired iterations. Result model best runs with 1024 echos and 128 batch size. As for as I increase echos it decreases class accuracy as my main objective of this thesis to surpass the previous class accuracy which was very low. Google Colab compiler Session is recycled after 10 hours . It is very difficult to run multiple test cases.

Chapter 5

Conclusions and Future Work

5.1 Conclusions

Xception model is trained and evaluated on two different Emotions image datasets. As we got high accuracy for each class which with high accuracy is 85 with same hyper parameters are trained and tested on different datasets which are different in term of crops, images as well as the condition in which they are captured. Which are almost higher to the previously best achieved results on a single same dataset Which shows that the model is generalized well on different datasets. It can also be summed up as the dataset which are more close to nature are better for later implementation in real world. Though good accuracy can be achieve on datasets like plant Face Emotions, and these images can not belong to real world scenario. Model trained on these images can not give better results on real life images. So, more dataset with real life images will be a better contributions Emotion Identification research. It can be used to detect consumer behaviour in real world scenarios.

Emotion estimation from facial expressions is the region of interest of numerous analysts within the writing. It is trusted that this consider will be a source of studies that will offer assistance within the early detection of illnesses from facial expressions additionally studies of buyer behavior analysis.

5.2 Future Work

Many distinctive adjustments, experiments, and tests have been left for the long haul due to lack of time and resources (i.e. training data set to get better and better results we need a high computational system). Future work concerns the deeper examination of Emotion detection with the latest architecture of deep learning. System has mainly focused to cover all emotional states and most functions used to get the best result was obtained from the literature review. Following ideas can be tested.

- It could be interesting to consider the model and data images with different resolution, background depending upon their zise or their specific meaning regarding reconginition process. This process would be play an interesting role to distinguish complex problems in which relevant region to be found which would appear.
- The way the model is implemented could also be changed as only using typical images. It could be based on real-world examples in order to provide some information on variability and introduce some attributes into them.

References

[1] priesen Ajeng mulandari and T. masaruddin. "Viola-Jones object detection framework ") . *Journal of Multimedia*, abs/1703.1856004:1, 2015, December 20. Cited on pp. xiii, 9, and 13.

[2] Criesen Qjeng Xulandari and T. Nasaruddin. "A real-time robust facial expression recognition system using HOG features ") . *IEEE*, 4:5, 2016. Cited on pp. xiii, 10, and 14.

[3] Briesen Bjeng pulandari and T. Hasaruddin. "Comparative Study on Normalisation in Emotion Recognition ") . *IEEE*, 4:5, 2017. Cited on pp. xiii, 10, and 14.

[4] Mizuho Nishio1 Rikiya Yamashita. "original Depth wise Seperable ") . *IEEE*, abs/1703.1856004:1, 2017. Cited on pp. xiii and 21.

[5] Sik-Ho Tsang. "Modified Depth wise Seperable ") . *IEEE*, abs/1703.1856004:1, 2011. Cited on pp. xiii and 22.

[6] Charles Darwin. "THE EXPRESSION OF THE EMOTIONS IN MAN AND ANIMALS ") . *IEEE*, 4:5, 1987. Cited on p. 1.

[7] Ekman. "Universals-And-Cultural-Differences-In-Facial-Expressions ") . *IEEE*, 4:5, 1999. Cited on p. 1.

[8] J. Mehmet Akif OZDEMIR, Berkay ELAGOZ1. Real time emotion recognition from facial expressions using cnn architecture. *Commun. ACM*, 48(4):27–31, October 2019. Cited on pp. 1, 8, and 18.

[9] Sung Kim (Pantic Patras. "Engineering Applications of Artificial Intelligence ") . *IEEE*, 4:3, 2008. Cited on p. 2.

[10] Fengjun Chen; Zhiliang Wang; Zhengguang Xu; Jiang Xiao; Guojiang Wang. Facial Expression Recognition Using Wavelet Transform and Neural Network Ensemble . *IEEE*, abs/1703.1856004:52, 2008. Cited on p. 2.

[11] Charles Markham Conor Cohen Farrell. Research methods in computer science. In *Real Time Detection and Analysis of Facial Features to Measure Student Engagement with Learning Objects*, page 6, Sep. 2019. Cited on pp. 3 and 18.

[12] Arfan Jaffar M. Sultan Zia, Majid Hussain. Incremental Learning-Based Facial Expression Classification System Using a Novel Multinomial Classifier . *International Journal of Pattern Recognition and Artificial Intelligence 32, no. 04*, abs/1703.1856004:52, 2018. Cited on p. 4.

[13] Taner Eskil Kristin S. Benli. "Extraction and Selection of Muscle Based Features for Facial Expression Recognition") . *IEEE*, abs/1703.1856004:7, 2014. Cited on pp. 6, 11, and 18.

[14] Gwen Littlewort. "The Computer Expression Recognition Toolbox ") . *ICRP*, abs/1703.1856004:1, 2001. Cited on pp. 6, 11, and 18.

[15] Xia Maoa Lijiang Chen Jianfeng Zhaoa, b. " Biomedical Signal Processing and Control ") . *IEEE*, abs/1703.1856004:1, 2019. Cited on p. 7.

[16] Xia Maoa Lijiang Chen Jianfeng Zhaoa, b. " Regulating Emotion Through Distancing: A Taxonomy, Neurocognitive Model, and Supporting Meta-Analysis ") . *IEEE*, abs/1703.1856004:1, 2018. Cited on p. 7.

[17] Reza Sadighzadeh Mehmet Akif Özdemir, Aydin Akan. " Real Time Emotion Recognition from Facial Expressions Using CNNArchitecture") . *IEEE*, abs/1703.1856004:1, 2019. Cited on p. 7.

[18] journal = IEEE year = 2018 pages = 10 volume = abs/1703.1856004 archiveprefix = arXiv bibsource = https://arxiv.org/pdf/1808.05561.pdf biburl = https://arxiv.org/pdf/1808.05561.pdf eprint = 1703.04080 timestamp = Mon, 13 Aug 2018 16:48:54 +0200 url = https://arxiv.org/abs/1412.6980 Kristiin S. Benli,Taner Eskil, title = "Emotion Recognition in Speech using Cross-Modal Transfer in the Wild") . Cited on p. 7.

[19] Mizuho Nishio1 Rikiya Yamashita. "Convolutional neural networks: an overview and application in radiology ") . *IEEE*, abs/1703.1856004:1, 2018. Cited on pp. 7 and 18.

[20] JiaJiong Ma. Tianyuan Chang Yang Hu, Guihua Wen. Facial Expression Recognition Based on Complexity Perception Classification Algorithm. *elsvier*, abs/1703.04080:7, 2018. Cited on pp. 8 and 18.

[21] 2 K. Clawson 1, L. S. Delicato and C. Bowerman. Facial Expression Recognition Using Wavelet Transform and Neural Network Ensemble . *IEEE*, abs/1703.1856004:12, 2018. Cited on p. 9.

[22] Sik-Ho Tsang. "Modified Depth wise Seperable ") . *Journal of Multimedia*, abs/1703.1856004:1, 2006. Cited on p. 12.

[23] Friesen Ekman and Ellsworth. "Universals and cultural differences in facial expressions of emotion") . *Journal of Multimedia*, abs/1703.1856004:1, 2006. Cited on p. 12.

[24] WecJeannie S. Emmanuel Jennifer M.B. Fugate, Aminda J. O'Hare. Emotion words: Facing change. *elsvier*, abs/1703.04080:11, 2018. Cited on pp. 12 and 18.

[25] Friesen Ajeng Wulandari and T. Basaruddin. "Enhancing CNN with Preprocessing Stage in Automatic Emotion Recognition ") . *Journal of Multimedia*, abs/1703.1856004:1, 2017. Cited on p. 13.

[26] AyaHassounehaA.M.MutawaaM.Murugappanb. "Comparative Study on Normalisation in Emotion Recognition ") . *IEEE*, 4:5, 2020. Cited on p. 15.

[27] Kevin S. LaBar John P. Powers. Regulating Emotion Through Distancing: A Taxonomy, Neurocognitive Model, and Supporting Meta-Analysis. *elsvier*, abs/1703.04080:89, 2018. Cited on pp. 15 and 18.

[28] Soonja Yeom Hai-Duong Nguyen. Facial Emotion Recognition Using an Ensemble of Multi-Level Convolutional Neural Networks . *elsvier*, abs/1703.04080:89, 2018. Cited on pp. 15 and 18.

[29] F. Lin, R. Hong, W. Zhou, and H. Li. Facial expression recognition with data augmentation and compact feature learning. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 1957–1961, 2018. Cited on p. 16.

[30] Thiago Andr Teixeira Lopes, Edilson De Aguiar Alberto Ferreira De Souza. Facial expression recognition with Convolutional Neural Networks . *elsvier*, abs/1703.04080:89, 2017. Cited on pp. 16 and 18.

[31] Ruyi Xu Jingying Chen, Yongqiang Lv. Automatic social signal analysis: Facial expression recognition using difference convolution neural network . *elsvier*, abs/1703.04080:6, 2019. Cited on p. 16.

[32] Prof. Bill Buchanan. As a PhD Examiner — My Top 25 Tips for PhD students. Technical report, Centre for Distributed Computing and Security, Edinburgh Napier University, 2018. Cited on p. 18.

[33] 2 K. Clawson 1, L. S. Delicato and C. Bowerman. Facial Expression Recognition Using Wavelet Transform and Neural Network Ensemble . *IEEE*, abs/1703.1856004:12, 2018. Cited on p. 20.

[34] N. Christou and N. Kanojiya. "Human Facial Expression Recognition with Convolution Neural Networks) . *IEEE*, abs/1703.1856004:8, 2017. Cited on p. 20.

[35] Diederik P. Kingma and Jimmy Lei Ba. "ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION ") . *IEEE*, abs/1703.1856004:15, 2014. Cited on p. 25.

[36] Diederik P. Kingma and Jimmy Lei Ba. " JEREMY JORDAN ") . *IEEE*, abs/1703.1856004:15, 2014. Cited on p. 27.

[37] Greeks. " JEREMY JORDAN ") . *IEEE*, abs/1703.1856004:1, 2020. Cited on p. 27.

[38] D. Lundqvist, A. Flykt, and A. Öhman. The Karolinska directed emotional faces (KDEF) . *IEEE*, abs/1703.1856004:12, 1998. Cited on p. 29.