DEEP LEARNING (DL)-ENABLED SYSTEM FOR EMOTIONAL BIG DATA

A Seminar Report

Submitted

In Partial Fulfillment of the Requirements

for the award of the Degree of

BACHELOR OF TECHNOLOGY

In

Computer Science & Engineering

APJ Abdul Kalam Technological University

Thiruvananthapuram-Kerala

By

ARUN PRASAD M (AIK19CS019)

COMPUTER SCIENCE & ENGINEERING

Albertian Institute of Science and Technology
AISAT Kalamassery –Kochi-682022
2019-2023

Vision

To be a Centre of excellence for professional education and related services creating technically competent and ethically strong innovative minds committed to the growth of the nation and beyond.

Mission

- oi are committed to provide value-based education with ample opportunities for research and industry institution interaction.
- We take very possible step to enhance the skills and bring out quality professionals, providing a friendly and growth-oriented ambience with appropriate resources.
- We improve ourselves through continuous evaluation and updation to meet the challenges and requirements of the modern society.

Motto

We make Engineers, not just Engineering Graduates

PROGRAM OUTCOMES (PO)

At the end of the program, graduate engineers will be able to

PO 1 - Engineering Knowledge: Apply the knowledge of mathematics, science, engineering fundamentals and an engineering specialization for the solution of complex engineering problems.

PO 2 - Problem Analysis: Identify, review research literature, formulate and analyze complex engineering problems, thereby arrive at substantiated conclusions using first principles of mathematics, natural sciences and engineering.

PO 3 - Design/Development of Solutions: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for public health and safety, and cultural, societal, and environment considerations.

PO 4 - Conduct investigations of complex problems: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

PO 5 - Modern tool usage: Create, select and apply appropriate techniques, resources, and modern engineering and IT tools, including prediction and modelling to complex engineering activities with an understanding of the limitations.

PO 6 - The Engineer and Society: Apply reasoning informed by contextual knowledge to assess societal health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

PO 7 - Environment and sustainability: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

PO 8 - Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

PO 9 - Individual and teamwork: function effectively as an individual, and a member in diverse teams and in multi-disciplinary settings.

PO 10 - Communication: Communicate effectively on complex engineering activities with the engineering community and with the society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give

and receive clear instructions.

PO 11 - Project management and finance: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member or leader in a team, to manage projects, and in multi disciplinary environments.

PO 12 - Lifelong learning: Recognize the need for, and have the preparation and ability to engage in independent and lifelong learning in the broadest context of technological knowledge.

COURSE OUTCOMES (COs)

After successful completion of the course, the students will be able to:

CO1: Identify academic documents from the literature which are related to her/his areas of interest (Cognitive knowledge level: Apply).

CO2: Read and apprehend an academic document from the literature which is related to her/ his areas of interest (Cognitive knowledge level: Analyze).

CO3: Prepare a presentation about an academic document (Cognitive knowledge level: Create).

CO4: Give a presentation about an academic document (Cognitive knowledge level: Apply).

CO5: Prepare a technical report (Cognitive knowledge level: Create).

CO-PO Mapping Matrix

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	2	2	1	1		2	1					3
CO2	3	3	2	3		2	1					3
CO3	3	2			3			1		2		3
CO4	3				2			1		3		3
CO5	3	3	3	3	2	2		2		3		3

CERTIFICATE

Certified that the work contained in the seminar titled "**DEEP LEARNING (DL) – ENABLED SYSTEM FOR EMOTIONAL BIG DATA**", by ARUN PRASAD M (AIK19CS019), has been carried out under my supervision and that this work has not been submitted elsewhere for a degree.

Name & Signature of Guide	Name & Signature of HOD
Nisy John Panicker	Dr.Jeswin Roy Dcouth
Assistant Professor	Associate Professor, HOD
Dept. of CSE, AISAT	Dept. of CSE, AISAT

College /Department Seal

APPROVAL & EVALUATION

Presented for the B. Tech Semester VII Seminar Evaluation held on …….....

Evaluation Committee

1. ……

2. ……

3. ……

ACKNOWLEDGEMENT

I would like to express my gratitude and appreciation to Dr. S. Jose, Principal, Albertian Institute of Science and Technology. A heartfelt gratitude to my guide Prof. Nisy John Panicker, Dept. of CSE for her valuable suggestions and encouragement. A special thanks to Prof. Dr. Jeswin Roy Dcouth, HOD, Department of CSE. I would also like to thank all the seminar coordinators Prof. Dr. Suja C. Nair, Prof. Chinnu Edwin A, and Prof. Gayathry V. for their encouragement and support.

Date: -12-2022

Place: Kochi ARUN PRASAD M. (AIK19CS019)

ABSTRACT

Emotion care for human well-being is important for all ages. an emotion care system based on big data analysis for autism disorder patient training, where emotion is detected in terms of facial expression. The expression can be captured through a camera as well as Internet of Things (IoT)-enabled devices. The system works with deep learning techniques on emotional big data to extract emotional features and recognize six kinds of facial expressions in real-time and offline. A convolutional neural network (CNN) model based on MobileNet V1 structure is trained with two emotional datasets, FER-2013 dataset and a new proposed dataset named MCFER. The experiments on three strategies showed that the proposed system with deep learning model obtained an accuracy of 95.89%. The system can also detect and track multiple faces as well as recognize facial expressions with high performance on mobile devices with a speed of up to 12 frames per second.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

With the proliferation of Artificial Intelligence (AI) and Internet of Things (IoT), emotion plays an important role in human life and communication. In the daily life, emotion is an inextricable part of the interaction of human beings, which can be observed by the changes in physiological features and behaviours. Because emotion recognition has a great potential to improve the quality of life, in the past decades, emotion recognition has aroused a lot of attention of many researchers and has been a popular research topic in various fields such as robotics, human-computer interaction, and entertainment, to name a few. Meanwhile, emotion care can be very useful in medical applications when medical staff need to assess the patient's feeling and behaviour during or after the surgery. With the development of big data and deep learning, huge amount of data including emotional data is generated in recent years, which cannot be handled with the traditional techniques. To this end, deep learning techniques has the potential to solve this problem. By using deep learning techniques to analyse the emotional big data, machines can learn and understand emotions to meet human needs, because deep learning techniques can learn and track different physiological features on the body.

Physiological features are closely associated with the generation of emotion and can be used for the recognition of the emotion. However, physiological data are inconvenient to obtain, many researchers pay more attention on other factors, such as facial expressions, gesture, and voice. Among these mentioned factors, facial expressions play the most important role, which contribute 55 percent in the emotion analysis, while the vocal part and verbal part contribute approximately 38 percent and 7 percent, respectively. Therefore, facial expressions are the most significant part in the behaviour analysis of emotion. In addition, with the development of smartphones and wearable devices, the portable healthcare system becomes more and more important. Many smart devices are developed to help people monitor health, such as heart rate, EEG. However, emotion care system working on smart devices is paid less attention, while there are so many emotion care systems based on computer system, which are not user-friendly and portable. In this Report, an

emotion care system based on automatic facial expression recognition (FER) system working on an Android smartphone is proposed.

CHAPTER 2

LITERATURE SURVEY

This section summarizes recent research related to the proposed system.

1. **AUTOMATED FACIAL EXPRESSION RECOGNITION APP DEVELOPMENT ON SMART PHONES USING CLOUD COMPUTING**

   Automated human emotion detection is a topic of significant interest in the field of computer vision. Over the past decade, much emphasis has been on using facial expression recognition (FER) to extract emotion from facial expressions. In this , the proposed system presents a novel method of facial recognition based on the cloud model, in combination with the traditional facial expression system. The process of predicting emotions from facial expression images contains several stages. The first stage of this system is the pre-processing stage, which is applied by detecting the face in images and then resizing the images. The second stage involves extracting features from facial expression images using Facial Landmarks and Center of Gravity (COG) feature extraction algorithms, which generate the training and testing datasets that contain the expressions of Anger, Disgust, Fear, Happiness, Neutrality, Sadness, and Surprise. Support Vector Machine (SVM) classifiers are then used for the classification stage in order to predict the emotion. In addition, a Confusion Matrix (CM) technique is used to evaluate the performance of these classifiers. The proposed system is tested on CK+, JAFFE, and KDEF databases. However, the proposed system achieved a prediction rate of 96.3% when Facial Landmarks and the Center of Gravity (COG) +SVM method are used.

2. **AUDIO-VISUAL EMOTION RECOGNITION USING MULTI-DIRECTIONAL REGRESSION AND RIDGE LET TRANSFORM**

   In this Report [5], an audio-visual emotion recognition system using multi-directional regression (MDR) audio features and ridgelet transform-based face image features. MDR

features capture directional derivative information in a spectro-temporal domain of speech, and, thereby, suitable to encode different levels of increasing or decreasing pitch and formant frequencies.

For video inputs, interest points in a time frame are detected using spectro-temporal filters, and ridgelet transform is applied to cuboids around the interest points. Two separate extreme learning machine classifiers, one for speech modality and the other for face modality, are used. The scores of these two classifiers are fused using a Bayesian sum rule to make the final decision. Experimental results on enter FACE database show that the proposed method achieves accuracy of 85.06 % using bimodal inputs, 64.04 % using speech only, and 58.38 % using face only; these accuracies outnumber the accuracies obtained by some other state-of-the-art systems using the same database.

3. DEEP LEARNING FOR REAL-TIME ROBUST FACIAL EXPRESSION RECOGNITION ON A SMARTPHONE

a real-time robust facial expression recognition function on a smartphone. To this end trained a deep convolutional neural network on a GPU to classify facial expressions. The network has 65k neurons and consists of 5 layers. The network of this size exhibits substantial overfitting when the size of training examples is not large. To combat overfitting, applied an data augmentation and a recently introduced technique called "dropout". Through experimental evaluation over various face datasets, it show that the trained network outperformed a classifier based on hand-engineered features by a large margin. With the trained network, here developed a smartphone app that recognized the user's facial expression. In this method share the experiences on training such a deep network and developing a smartphone app based on the trained network.

CHAPTER 3

EXISTING SYSTEM

1. **REAL-TIME MOBILE FACIAL EXPRESSION RECOGNITION SYSTEM**

   For FER system on smartphone, Suk and Prabhakaran [1] proposed a system that distinguishes between neutral and non-neutral expression frames in video sequences by using facial landmarks. If non-neutral expression is found, the new dynamic features are generated by displacing saved neutral features with current features. Then the new features are fed into SVM models for FER task. The SVM model obtained an accuracy of 86%. The mobile application was tested on Samsung Galaxy S3 with 2.4 frames per second (fps).

2. **DEEP LEARNING FOR REAL-TIME ROBUST FACIAL EXPRESSION RECOGNITION ON A SMARTPHONE**

   Song *et*[2] *a* developed a deep learning FER application with DNN model with an accuracy of up to 99.2%. The DNN has 5 layers and recognize 5 facial expressions (i.e., anger, happy, sad, surprise, and neutral). The smartphone application captures first the user's face, then it sends a request to a server, thus the server predicts facial expressions by a trained model and then sends the prediction to mobile phone. Due to the app uses the client-server architecture, it cannot work offline.

3. **ROBUST FACIAL EXPRESSION RECOGNITION AGAINST ILLUMINATION VARIATION APPEARED IN MOBILE ENVIRONMENT**

   Jo *et al.* [3] proposed a new robust FER system against illumination variation, which utilizes Active Appearance Model (AAM) and NN with a Difference of Gaussian (DOG) to fix illumination variation problems and recognize facial expressions.

4. AUTOMATED FACIAL EXPRESSION RECOGNITION APP DEVELOPMENT ON SMART PHONES USING CLOUD COMPUTING

Alshamsi*et al.* [4] developed an automated FER application with an accuracy of 96.3%, but it uses cloud computing for FER based on facial landmarks and SVM.

The existing smartphone applications are compared in Table 1. While most applications work in real time and offline, none of them supports multiple users for emotion detection. Therefore develop a multi-user Android application running in real time and offline for emotion care and autism disorder patient training.

Table 1: Comparison of existing FER smartphone applications

Method	Method	Multi-user	Real-time	Offline
Suk et al. [31]	SVM	No	Yes	Yes
Song et al. [32]	CNN	No	Yes	No
Jo et al. [33]	AAM	No	Yes	-
Alshamsi et al. [34]	SVM	No	Yes	No
Ours	CNN	Yes	Yes	Yes

CHAPTER 4

PROPOSED SYSTEM

In this Report, DT-FER based emotion care system, where emotion is detected in terms of facial expression. The system contains two parts: model and Android application. In the model part, after image pre-processing for the images from dataset, the deep learning model is trained for emotion classification and deployed on a smartphone. The DT-FER Android application (app) detects the multiple faces in the images captured from the smartphone camera by using face detector and predicts facial expressions from detected face images by using CNN model in real time and offline. The application calculates a score for each facial expression and shows the highest score and emotion on the smartphone screen. And the highest score indicates the performed emotion meets the standard better. And the performed facial expression and predicted results can be saved into device to help carers to observe facial expressions of patients at any time and provide better suggestions to patients.

In addition, a new dataset for emotion recognition is collected to train the CNN model. In summary, the main contributions of this Report are

1) A new low-cost and multi-user framework for emotion detection is proposed. The system is based on an

Android application that uses CNN model to classify facial expressions and works in real time and offline.

2) A new challenging and less-constrained dataset called MCFER is introduced for facial expression classification. The dataset is collected in real scenario with more complex conditions, such as movement and light interference.

CHAPTER 5

SYSTEM ARCHITECTURE

DT-FER based emotion care system, where emotion is detected in terms of facial expression. As shown in Figure 2, the system contains two parts: model and Android application.

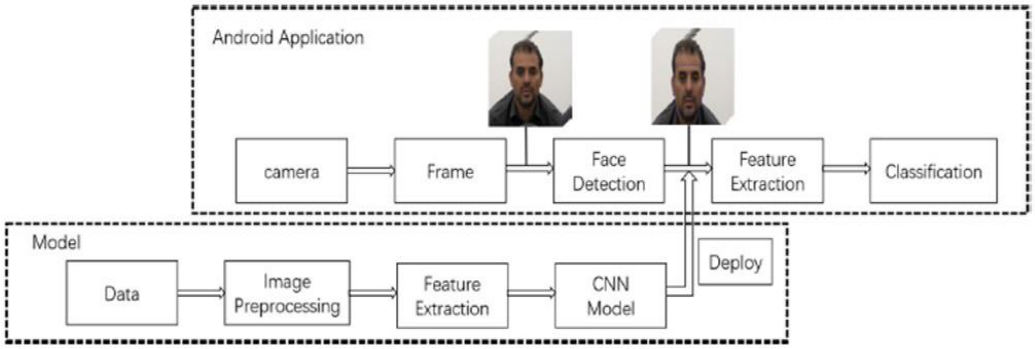


Figure 1: System architecture

CHAPTER 6

METHODOLOGY

1. DATA SET

In this section, there are two popular emotional big data and a new dataset that is named MCFER and its acquisition process. The architecture of CNN model and the framework of application are also described in detail.

A. FER-2013

Facial Expression Recognition 2013 (FER-2013) dateset [10]was prepared in Challenges in Representation Learning: Facial Expression Recognition Challenge, which is hosted by Kaggle. FER-2013 database has seven facial expression categories (e.g., angry, disgust, fear, happy, sad, surprise, and neutral) and three different sets such as training set (28.709 images), validation set (3.589 images), and test set (3.589 images). All images in this dateset are grayscale with 48×48 pixels, thus corresponding to faces with various poses and illumination, where several faces are covered by hand, hair, and scarves. Because of FER-2013 is collected from the Internet and has various real-world conditions, it becomes one of the largest and most challenging database for facial expression recognition.

B. CK+

CK+: the Extended CohnKanade (CK+) database consists of 593 deliberate image sequences from 123 subjects, which is the most used lab-controlled database for evaluation of FER systems. The database was labelled with seven basic facial expressions (anger, contempt, disgust, fear, happiness, sadness and surprise) by adopting the FACS (Facial Action Coding System).

C. MCFER

1) DATA ACQUISITION

Unlike most popular databases (e.g., CK+, JAFFE) that were collected in special lab environments with same and specific environment, new MCFER (Multimedia Communications Research Laboratory Facial Expression Recognition) database is collected in various places of University of Ottawa without particular environment, which makes the data more realistic. Totally, 15 participants (33% female and 67% male) between 23 and 60 years old take part in the experiment. The participants are randomly selected in real life and most of them are students or staff at the University of Ottawa. The experiment is designed as follows: first, the participant is asked to read and sign a consent form to participate in the experiment. Then, the participant reads an instruction about the details of the experiment and is instructed by an experimenter to understand the purpose of the study and the detailed experimental procedure. Once the participant is ready in front of the screen, the researcher turns the camera on and asks the participant to perform a series of facial expressions starting with angry and ended in surprise. The participants start the experiment in a natural place where they are found, instead of performing in a special man-made environment. Moreover, participants can look at the camera at any angle they want.

2) DATA DEDUPLICATION

For every participant, one video with six kinds of facial expressions is collected and processed. A haar cascade classifier proposed by Viola and Jones is used to detect the face from video frame by frame. When a face is detected, the face image is saved into the database and labelled according to the facial expression the participant shows. Because the database has a lot of similar images due to the successive frames, here use difference hash (dhash) algorithm to select representative images from the dataset. The difference hash is one of image fingerprint algorithms, and it creates a unique hash value by calculating the difference between adjacent pixel values. To select images form the dataset, use the difference hash to compute the image fingerprints because of its speed and accuracy, the image was first shrank to a new size, which would match any similar images regardless of how it is stretched by ignoring the original size and aspect ratio. Second, the colourful image was converted to grayscale image which reduce hash from $S2$ C S pixels to a total of $S2 + S$

colours. Then, the dhash algorithm calculated the difference between adjacent pixels, which identifies the relative gradient direction. Here got *S* differences from *S* C 1 pixels per row.

Therefore, the total differences of the image are *S*2 bits. Finally, by compared the brightness between adjacent pixels to get the hash value. If the left pixel is brighter than the right pixel, the bit is set to 1, otherwise 0. To compare two hashes, just counted the number of bits that are different which is the Hamming distance. If the Hamming distance between two images is less than the threshold *D*, one image would be discarded because regarded these two images were the ``same'' images containing repetitive information.

DHASH ALGORITHM

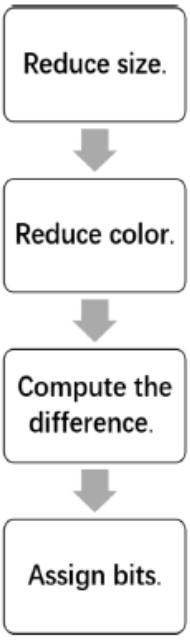


Figure 2: Dhash Algorithm

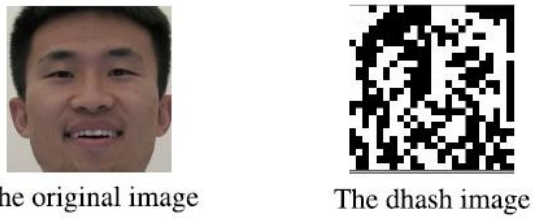


Figure 3: The dhash image is obtained from the original image by using dhash algorithm

The original facial expression was encoded as a new hash value. Select 24 as the image size and 170 as the threshold, which means all images with the distance less than 170 were discarded. The dhash image shows the encoded image using dhash. Then, four volunteers who retain vote power are found to judge whether each image belongs to its corresponding category. Once one of the volunteers cast opposing vote for the image, the image is discarded. Finally obtained 287 images for the proposed MCFER dataset. It depicts some collected images of the MCFER dataset.

EXAMPLES OF FER-2013, CK+, MCFER

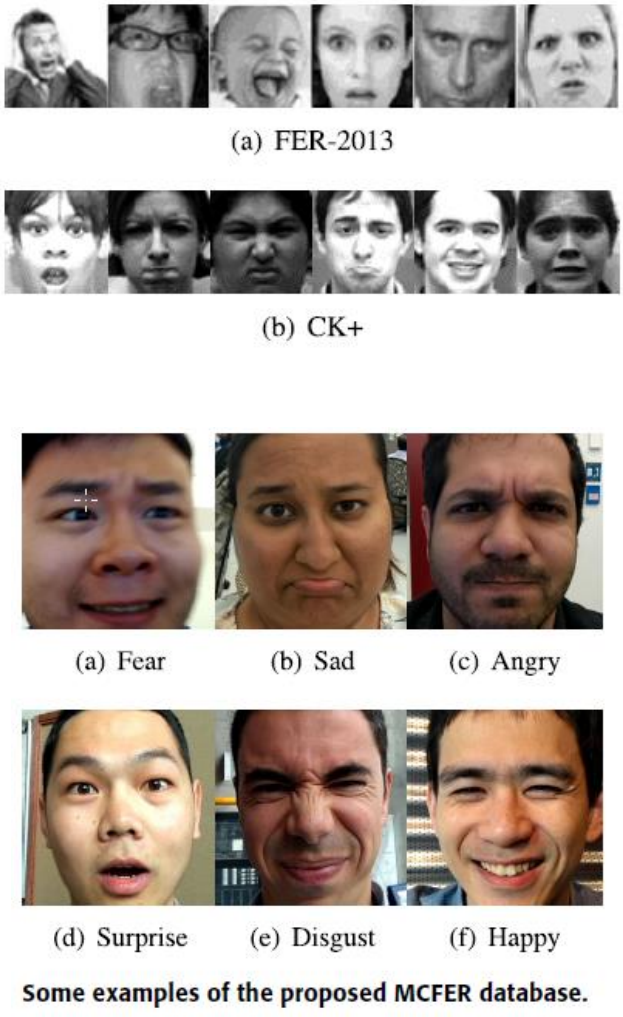


Figure 4: Examples of FER-2013, CK+, MCFER

2. PREPROCESSING

For the system, image Preprocessing is necessary before an image is fed into the CNN model. The image Preprocessing mainly consists of two stages: face detection, data augmentation. A face detector is adopted for face detection in the system. If faces are detected, the four coordinates of region of interest (ROI) of the faces would be returned to the system, the system would crop the faces and discard irrelevant background. Data augmentation is used to process the detected face images and increase the quantity of data, because training process of deep learning model usually needs huge amounts of data. The images are cropped by the random bounding boxes that have different cropped ranges from 0.85 to 1. Then the data are randomly flipped and rotated. The amount of data has increased by 200 times.

3. CNN MODEL

The MobileNet V1 [8], which is used as a model architecture, was proposed in 2017. It is a lightweight deep neural network, which already became in an underlying network structure. Because its key point is to construct a small neural network, it can be widely used on mobile and embedded devices with remarkable speed and good accuracy compared with other architectures. The main idea of the MobileNet V1 is to decouple standard convolution into a 1_1 pointwise and depth wise convolutions to extract features from big data. Depth wise convolution is used to extract features, where those features are combined into new features by pointwise convolution. Thus, Mobilenet V1 architecture has smaller model size and complexity because of fewer number of parameters and fewer additions and multiplications. The reduction of computation between traditional convolution and combination of Depth wise and pointwise convolutions are compared in Equation 1. The computation amount of traditional convolution is represented as denominator and the numerator shows the computation amount of MobilNet V1.

$$\frac{D_F * D_K * D_F + M * N * D_F * D_F}{D_K * D_K * M * N * D_F * D_F} = \frac{1}{N} + \frac{1}{D_K^2} \qquad (1)$$

Where DK is the convolutional kernel size and DF denotes an input feature map, while M and N denote the number of input and output channels, respectively.

In addition, the MobileNet V1 also has two hyper parameters such as the resolution multiplier ρ to control the size of the feature map and the width multiplier α to reduce the calculation amount. Because of the MobileNet V1 can reach a satisfactory trade-off between accuracy and speed compared to other popular CNN structures (e.g. AlexNet, VGG16). select this lightweight CNN model as framework to train a FER classifier.

THE PROPOSED ARCHITECTURE WITH JOINT LOSS

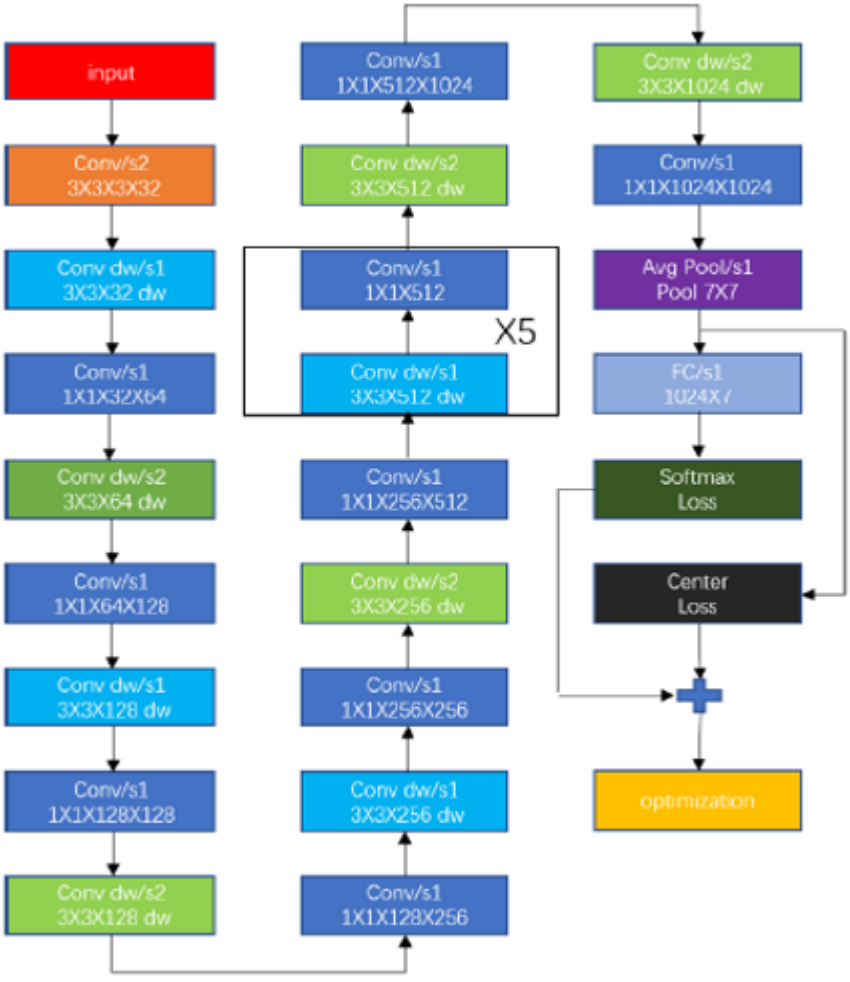


Figure 5: The proposed architecture with joint loss

4. JOINT LOSS

Center loss [9] shown in Equation 2 can be used in training step to increase discriminatory power by reducing the distance constraint between the feature and its corresponding class center. Therefore, it can be used to learn discriminative feature and improve model performance.

$$L_{CL} = \frac{1}{2}\sum_{i=1}^{N}$$

$$L = L_S + \lambda \cdot L_{CL} = -\sum_{i=1}^{N} \log \frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^{m} e^{W_j^T x_i + b_j}} + \frac{1}{2}\sum_{i=1}^{N} \|x_i - c_{y_i}\|_2 \quad (3)$$

where x_i denotes the ith features extracted from y_ith class and cy_i denotes the learned center for the y_ith class. The softmax loss increases the distance between different classes, the center loss reduces the distance within a class. Therefore, the joint loss containing center loss and softmax loss not only enlarges inter-class feature difference, but also reduces intra-class feature variation, as shown in Equation 3.

5. TWO-STAGE STRATEGY

Training a new model from scratch takes a tremendous amount of time and effort and needs a lot of data to achieve high performance. A two-stage approach [6] to train the CNN model in order to deal with the problem of insufficient size of small datasets. After that implement the first stage fine-tuning with FER-2013 dataset by using the model with pre-trained weights from the ILSVRC-2012 (ImageNet) [7]. After the best trained model is obtained from the FER-2013 dataset, the second stage fine-tuning is implemented with new dataset to obtain the final model. For the two-stage strategy, the last fully connected layer is replaced by new fully connected layer with six facial expression classes output. For the first fine-tuning stage, due to the FER-2013 dataset has 48 × 48 pixels images and the input size of original MobileNet V1 is 224 × 224 pixels, the Gaussian distribution is adopted to initialize the parameters of the first convolutional layer. For the second fine-tuning stage, the input image of MCFER dataset has the same size of 48 × 48 pixels as the size of FER-2013 dataset

After pre-processing. The weights of the first convolutional layer are initialized with the weights from the FER-2013.

6. DT-FER APPLICATION

The CNN model is developed using TensorFlow platform, which is an end-to-end open-source platform for machine learning. In order to use CNN model on the mobile phone, the model First needs to be converted to a new data (i.e., Lite) file by Tensor Flow Lite. Tensor Flow lite provides a set of tools to run TensorFlow models on mobile and embedded devices. After the model is converted to a lite file, it is deployed on mobile phone with the TensorFlow Lite interpreter. When the app starts to work, it continuously captures the images from the front camera or rear camera. The haar-like feature is employed to detect faces in the application and then the detected faces are cropped and resized to 48 × 48 size. After normalization and other processing methods, the face images are fed into the model as input. Finally, the prediction and other results are shown on the screen in real time.

WORKFLOW OF THE PROPOSED MOBILE APPLICATION

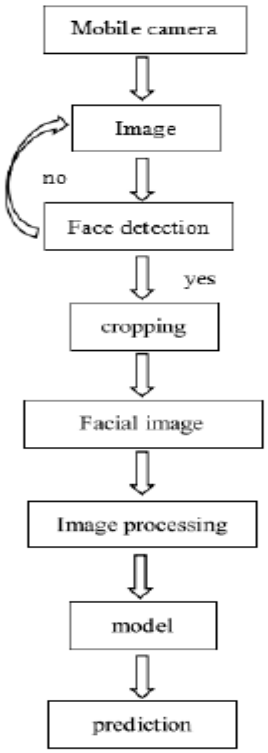


Figure 6: Workflow of the proposed mobile application

CHAPTER 7

RESULTS

1. PERFORMANCE OF CNN MODEL

To train a model for FER, all images of the datasets have to be detected by using the haar cascade classifier in order to determine if there is a face in the image. If the classifier detects the face, this will be cropped. After data augmentation and normalization for the cropped images, the processed images are used to train the model. With 8000 training steps, the model trained on FER-2013 database in the first fine-tuning stage obtained an accuracy of 67.03%. Based on the first stage pre-trained model on FER-2013 dataset, the accuracy of model on MCFER dataset in the second stage is about 95.89%.The model also has a precision of 97.58% and a recall of 100%. In addition, as shown in Figure 9 F1-score of the model is 98.79%.

THE ACCURACY OF THE FIRST STAGE MODEL

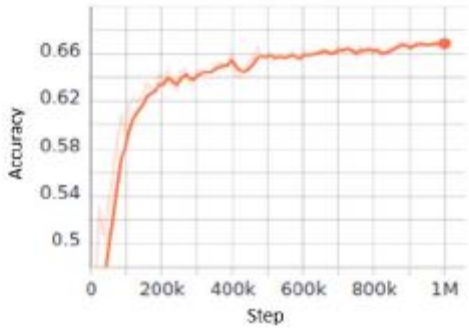
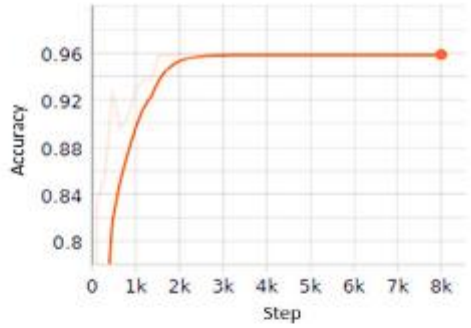
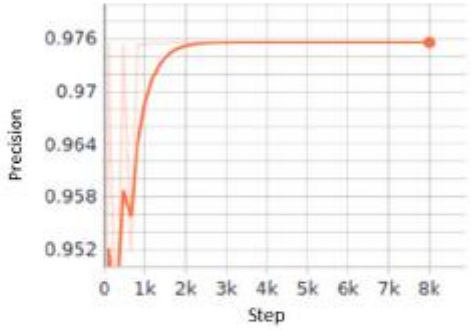


Figure 7: The accuracy of the first stage model with 1 million steps on FER-2013 dataset

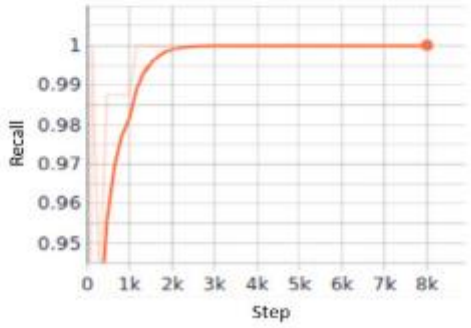
THE PERFORMANCE OF THE SECOND STAGE MODEL



(a) The accuracy in the second stage



(b) The precision in the second stage



(c) The recall in the second stage

Figure 8: The performance of the second stage model with 8000 training steps on MCFER dataset

THE F1-SCORE OF THE SECOND STAGE MODEL

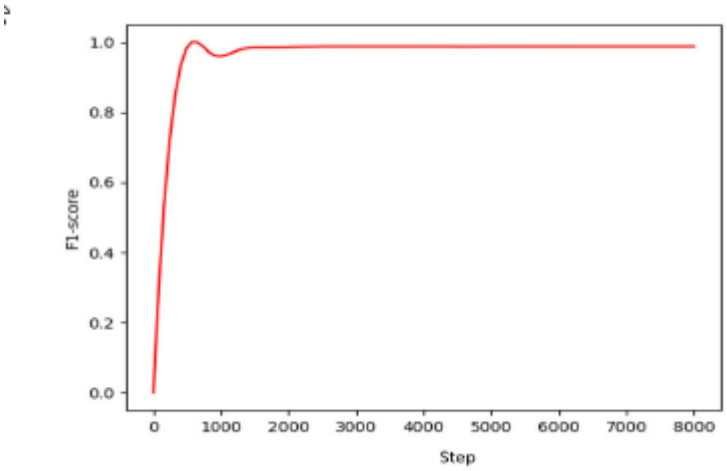


Figure 9: The F1-score of the second stage model with 8000 training steps on MCFER dataset

Test three methods with different training stages and loss functions on two datasets, CK+ and MCFER. One-stage with softmax loss method indicates that the CNN model is fine-tuned on the pre-trained ImageNet model with softmax loss function, the FER-2013 dataset is not used in this method. Two-stage with softmax loss means the model is first fine-tuned on pre-trained ImageNet model with FER-2013, then FER-2013 model is fine-tuned on CKC or MCFER. The last one, two-stage with joint loss, is the method used in the Report, the center loss and softmax loss are combined together to get the joint loss function. Two-stage strategy improves the performance of the model significantly, because large dataset first coarsely adjusts the feature extractor to FER task, then small dataset _ne tunes the model again to get better performance. Meanwhile, the joint loss from softmax loss and center loss also improves the accuracy of the model, as the center loss make the model to update the weights to better learn deep discriminant features from data.

2. PERFORMANCE OF DT-FER APPLICATION

SAVING MODE AND FREE MODE IN THE PROPOSED MOBILE APPLICATION

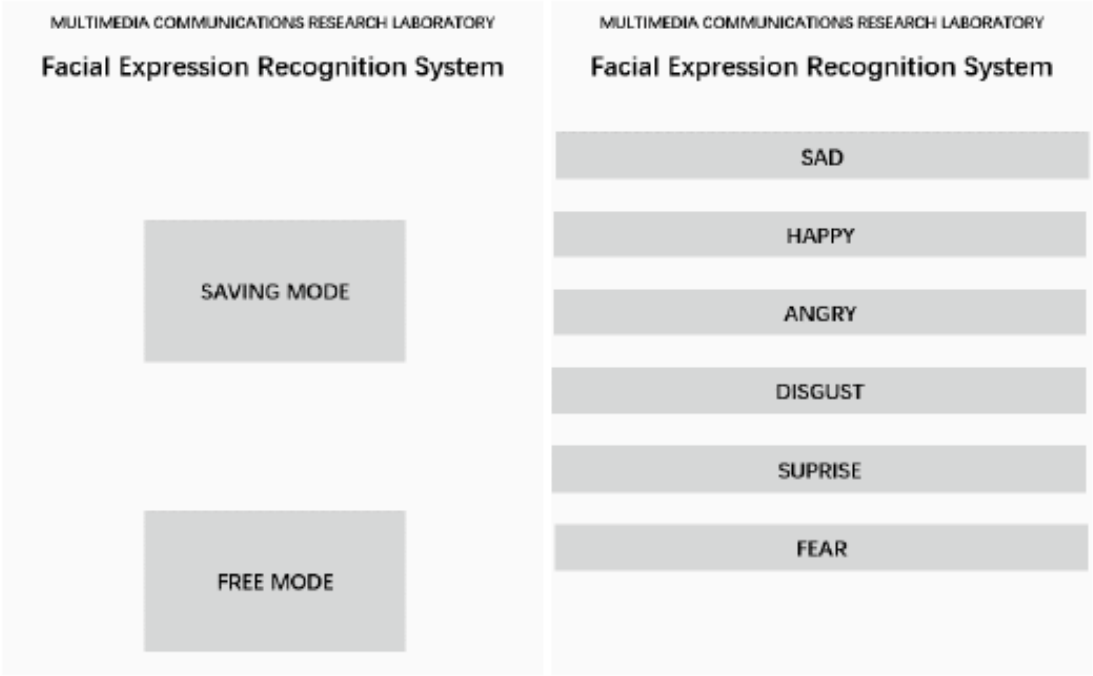


Figure 10: Left: Saving mode and free mode in the proposed mobile application. Right: When the saving mode is selected, the predicted facial expressions will be saved in their corresponding directory

To monitor patients in real time and offline, the model is integrated into the Android application instead of remote server. The application has two kinds of working modes: saving mode and free mode. The Saving Mode means the images and predicted results are saved on the phone, which would help carers to observe behaviors of people with autism at any time and provide better suggestions to patients, but the app would run slower because of saving image. The free mode, in turn, means nothing needs to be saved, so the app runs at higher speed compared with saving mode. In addition, the application support multiple facial expressions recognition, which can detect multiple faces and recognize facial expression for each face at the same time.

When users use the developed system, four results, time spent in prediction, frame per second, predicted facial expression, and prediction confidence, are shown on the screen in real time. Time spent in prediction indicates the running time that CNN model spent in predicting facial expressions. Frame per second shows that the speed of the system. The predicted facial expression shows the users' facial expression detected by the system, and Prediction confidence means confidence score for the predicted result. With the help of the developed system, doctors or parents could monitor the training process of patient. The system is tested on a Huawei G9 VNS AL00 smartphone, which has Qualcomm MSM8952 CPU Processor, 3GB of RAM, and 16GB of ROM. The prediction is shown on the screen, as well as the time taken by the model to make the prediction (e.g., 40 MS,46 MS, etc.). The App can recognize facial expressions with 12 fps in free mode. However, when the system works in saving mode, it is slower with a speed of 9 fps. Meanwhile, the app can detect multi-face and recognize facial expression. But if there are more faces, the speed is lower. For instance, when two faces are detected by the app, the speed would be about 5 fps, which is half of the speed recognizing one face. Because of the limitation of computation power on the mobile device, Here tested the application with saving mode under the condition of full load. The obtained speed was about 4 fps, which is the lowest speed for the application in the Smartphone.

SCREENSHOT OF THE DT-FER APP FOR BASIC EXPRESSIONS

(a) Surprise and Angry

(b) Happy

(c) Fear

(d) Sad

(e) Disgust

Figure 11: Screenshot of the DT-FER app for basic expressions

CHAPTER 8

CONCLUSION

In this Report, a deep learning technique to process emotional big data and develop an emotion care system using facial expression recognition system working on smartphone. The CNN model is trained with two emotional datasets: FER-2013 and a new proposed dataset, MCFER, which was collected at the University of Ottawa. Here employ a two-stage strategy and joint supervision to train the model and test the method on two datasets: CK+ and MCFER, which shows the performance is improved. The model on MCFER obtains a good performance with 95.89% accuracy. The Android application can detect multiple faces and recognize facial expression for every face in real time and offline. Here test the application in a smartphone with two modes, saving mode that saves the predicted results to help doctors or parents monitor autism disorder patient and free mode that make the system work at high speed. The application has a good performance with a speed of up to 12 fps.

CHAPTER 9

FUTURE SCOPE

The Proposed System has a good performance with a speed of up to 12 fps. Because of the weak computation power and limited memory for the embedded system, the model could be quantified to save memory space and improve speed in the future. Emotion-based music player (EMOSIC) which can recognize the mood of the individual and play music accordingly. It could play a crucial role in improving a person's mental state. Through improvement make it into a system that can distinguish lies said by an individual through his/her facial expression and body language.

REFERENCES

[1] M. Suk and B. Prabhakaran, ``Real-time mobile facial expression recognition system_A case study," in *Proc. IEEE Conf. Comput. Vis. PatternRecognit. Workshops*, Jun. 2014, pp. 132_137.

[2] I. Song, H.-J. Kim, and P. B. Jeon, ``Deep learning for real-time robust facial expression recognition on a smartphone," in *Proc. IEEE Int. Conf.Consum. Electron. (ICCE)*, Jan. 2014, pp. 564_567.

[3] G.-S. Jo, I.-H. Choi, and Y.-G. Kim, ``Robust facial expression recognition against illumination variation appeared in mobile environment," in *Proc.1st ACIS/JNU Int. Conf. Comput., Netw., Syst. Ind. Eng.*, May 2011,pp. 10_13.

[4] H. Alshamsi, V. Kepuska, and H. Meng, ``Automated facial expression recognition app development on smart phones using cloud computing," in Proc. IEEE 8th Annu. Ubiquitous Comput., Electron. Mobile Commun.Conf. (UEMCON), Oct. 2017, pp. 577-583.

[5] M. S. Hossain and G. Muhammad, ``Audio-visual emotion recognition using multi-directional regression and ridgelet transform," J. Multimodal User Interfaces, vol. 10, no. 4, pp. 325333, 2016.

[6] Y. Miao, H. Dong, J. M. A. Jaam, and A. E. Saddik, ``A deep learning system for recognizing facial expression in real-time," *ACM Trans. Multi-media Comput., Commun., Appl.*, vol. 15, no. 2, pp. 1_20, Jun. 2019.

[7] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, ``ImageNet large scale visual recognition challenge," Int. J. Comput. Vis., vol. 115, no. 3, pp. 211252, 2015.

[8] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, MobileNets: Efcient convolutional neuralnetworks for mobile vision applications," CoRR, vol. abs/1704.04861,pp. 1-9, Apr. 2017. [Online]. Available: http://arxiv.org/abs/1704.04861

[9] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, ``A discriminative feature learningapproach for deep face recognition," in Proc. ECCV, 2016, pp. 499-515.

[10] I. J. Goodfellow et al., ``Challenges in representation learning: Areport on three machine learning contests," Neural Netw., vol. 64,pp. 5963, Apr. 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0893608014002159