MULTI-FEATURE AWARE POSE AND GEOMETRY BASED FACIAL EXPRESSION RECOGNITION USING DEEP LEARNING

A DISSERTATION

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE AWARD OF DEGREE
OF
MASTER OF TECHNOLOGY
IN
COMPUTER SCIENCE & ENGINEERING

Submitted by:

ANKIT SHARMA
2K20/CSE/05

Under the supervision of
Dr. Manoj Sethi

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

MAY, 2022

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi - 110042

CANDIDATE'S DECLARATION

I, Ankit Sharma, Roll No. 2K20/CSE/05 student of M. Tech (Computer Science and Engineering), hereby declare that the project Dissertation titled **"**Multi-Feature Aware Pose and Geometry Based Facial Expression Recognition Using Deep Learning **"** which is submitted by me to the Department of Computer Science & Engineering, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Technology, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of and Degree, Diploma Associateship, Fellowship or other similar title or recognition.

Place: Delhi　　　　　　　　　　　　　　　　　　　　Ankit Sharma

Date: 25-05-2022　　　　　　　　　　　　　　　　　　2k20/CSE/05

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi - 110042

CERTIFICATE

I hereby certify that the Project Dissertation titled **"Multi-Feature Aware Pose and Geometry Based Facial Expression Recognition Using Deep Learning"** which is submitted by Ankit Sharma, 2K20/CSE/05 Department of Computer Science & Engineering, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Technology, is a record of the project work carried out by the students under my supervision. To the best of my knowledge, this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Place: Delhi — Dr. Manoj Sethi

Date: — Department Of CSE

ACKNOWLEDGMENT

The success of this project requires the assistance and input of numerous people and the organization. I am grateful to everyone who helped in shaping the result of the project.

I express my sincere thanks to **Dr. Manoj Sethi**, my project guide, for providing me with the opportunity to undertake this project under his guidance. His constant support and encouragement have made me realize that it is the process of learning which weighs more than the end result. I am highly indebted to the panel faculties during all the progress evaluations for their guidance, constant supervision and for motivating me to complete my work. They helped me throughout with newideas, provided information necessary and pushed me to complete the work.

I also thank all my fellow students and my family for their continued support.

ANKIT SHARMA

2K20/CSE/05

ABSTRACT

Facial expressions are the primary way to express intentions and emotions. In real life, computers can understand the emotions of humans by analysing their facial expressions. Facial Expression Recognition (FER) plays an important role in human-computer interaction necessity and medical field. In the past, the facial features are extracted manually for recognizing the expressions. In the present, it is an important hotspot in computer vision, Internet of Things, and artificial intelligence fields. Certain processes are involved to recognize the facial expression in an efficient manner such as;

- Pre-Processing
- Segmentation
- Feature Extraction
- Classification

Various algorithms are designed manually for feature extraction and other feature extraction algorithms such as Local Binary Pattern (LBP), Gabor wavelet, Histogram of Oriented Gradient (HOG). Various challenges are involved in the FER to recognize accurate expressions of facial images. For robust classification of facial expression, consideration of illumination and pose of the facial image is important. The poses and facial identity learning are essential to get accurate results. Several existing works faced challenges regarding identity, pose variation, and inter-subject variation. For estimating the pose of the facial images existing methods used hand-crafted features. For detecting the pose of the facial images, pose normalization is performed by considering the angle in the existing works. Previous works also considering one

pixel-based normalization to increase the accuracy for recognizing the facial expression. On the other hand, general illumination effect of the images also affects the accuracy, by contrast, occlusion, etc. Segmentation is one of the major processes to partition the facial images to extract the features. Various existing methods used different techniques to segment the facial images such as bounding box-based segmentation, region-based segmentation, cluster-based segmentation, etc., by using various algorithms such as Discrete Cosine Transform (DCT), K-mean Clustering algorithm, etc. For extracting the features after segmenting the facial images various Machine learning (ML) algorithms such as Support Vector Machines (SVM), Naive Bayes, K-Nearest Neighbours (KNN), and Deep Learning (DL) algorithms such as Convolutional Neural network (CNN), Generative Adversarial Network (GAN), Long-Short Term Memory Networks (LSTM). After extracting the features from the facial images, classification is performed to identify the facial expressions (happy, sad, anger etc.). Various classifiers are used in the previous works for classification processes such as VGG-16, VGG-19, ResNet, etc. These classifiers get the extracted feature as input and classify the expressions. For training and testing purposes various datasets are used but, mostly FER-2013, CK+, are used.

CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

1. FER: Facial Expression Recognition
2. ML: Machine Learning
3. DL: Deep Learning
4. CNN: Convolutional Neural Network
5. DCNN: Deep Convolutional Neural Network
6. CapsNet: Capsule Neural Network
7. GCapsNet: Graph-Based Capsule Neural Network
8. CK+: Cohn- Kanade

CHAPTER 1

INTRODUCTION

1. Introduction

Facial expressions are the primary way to express intentions and emotions[1]. In real life, computers can understand the emotions of humans by analyzing their facial expressions. Facial Expression Recognition (FER) plays an important role in human-computer interaction necessity and medical field. In the past, the facial features are extracted manually for recognizing the expressions. In the present, it is an important hotspot in computer vision, Internet of Things, and artificial intelligence fields. Certain processes are involved to recognize the facial expression in an efficient manner such as,

- Preprocessing
- Segmentation
- Feature Extraction
- Classification

Various algorithms are designed manually for feature extraction and other feature extraction algorithms such as Local Binary Pattern (LBP), Gabor wavelet, Histogram of Oriented Gradient (HOG), etc [4]. Various challenges are involved in the FER to recognize accurate expressions of facial images. For robust classification of facial expression, consideration of illumination and pose of the facial image is important. The poses and facial identity learning are essential to get accurate results. Several existing works faced challenges regarding identity, pose variation, and inter-subject variation. For estimating the pose of the facial images existing methods used hand-crafted features. For detecting the pose of the facial images, pose normalization is performed by considering the angle in the existing works. Previous works also considering pixel-based normalization to increase the accuracy for recognizing the facial expression[9]. On the other hand, general illumination effect of the images also affects the accuracy, by contrast, occlusion, etc.

Segmentation is one of the major processes to partition the facial images to extract the features[4]. Various existing methods used different techniques to segment the facial images such as bounding box-based segmentation, region-based segmentation, cluster-based segmentation, etc., by using various algorithms such as Discrete Cosine Transform (DCT), K-means Clustering algorithm, etc. For extracting the features after segmenting the facial images various Machine learning (ML) algorithms such as Support Vector Machines (SVM), Naive Bayes, K-Nearest Neighbors (KNN), and Deep Learning (DL) algorithms such as Convolutional Neural network (CNN), Generative Adversarial Network (GAN), Long-Short Term Memory Networks (LSTM)[7]. After extracting the features from the facial images, classification is performed to identify the facial expressions (happy, sad, anger etc.). Various classifiers are used in the previous works for classification processes such as VGG-16, VGG-19, ResNet, etc [2]. These classifiers get the extracted feature as input and classify the expressions. For training and testing purposes various datasets are used but, mostly FER-2013, CK+, and JAFFE datasets are used.

1.2 Research Aim and Scope

The aim of this research is to identify the facial expression from the facial images by using Deep learning technique. In addition, this research identifies the problems of considering adequate features, high false-positive rate, less accuracy, and so on.

1.3 Research Objectives

The main objective of this research is to recognize the facial expressions from the facial images with low false positive rate and high accuracy. The remaining research objectives are described as follows,

- To increase the facial image quality preprocessing is initialized by performing normalization to estimate the pose and angle of the facial image and also reduce the illumination effects to increase the accuracy of the facial expression.
- To extract the features efficiently, cluster-based segmentation is performed to estimate the facial objects such as eyebrows, eyes, nose, and mouth which reduces the loss of small expressions.

- Multi-feature is extracted in the feature extraction process by considering both low-level and high-level features of the facial images to increase the accuracy of facial expression recognition.

1.4 Thesis Structure

This thesis is divided into many chapter which are as follows: Chapter 1, Introduction presents the introductory part and the idea behind the aim, scope and objectives of the research. Chapter 2, Literature Review presents the survey of the research papers which are used as a references and the related existing work. Chapter 3, Research methodology presents the overview of the research and the algorithms used. Chapter 4, Implementation and Discussion presents the experiment and the evaluation of the experiment by using various performance metrics. It also presents the detailed explanation of the algorithms used in this research. Chapter 5 presents the conclusion and then references and the publication done so far.

CHAPTER 2

LITERATURE REVIEW

[1] The author proposed an approach to recognize the facial expression from the static facial image by using hybrid deep learning networks. Initially, the expressional features were extracted from the facial image by using spatial attention convolutional neural network (SACNN). In this neural network, VGG-19 is used as a spatial attention module which is used to extract the pixel-based features from the facial images. After extracting the features, landmark detection is performed based on geometry of the facial image by using attention mechanism based on long short-term memory networks (ALSTM), and this mechanism was also used to estimate different landmark regions' importance. Here FER-2013, CK+, and JAFFE datasets are used for experiment. The author used SACNN for extracting the pixel-based expressional features from the facial images. However, SACNN needs huge trained images to extract features effectively that increase the complexity for classification of facial expressions. Here, Batch normalization layer is included in the SACNN to reduce the internal covariance shift but, it needs huge batch size to normalize the outputs of the convolution layer to produce better results and Lack of considering the size of batch leads to high false-positive rate. To discover the facial landmarks from the extracted facial features ALSTM was used and it is also used to evaluate the landmark region's importance adaptively. However, lack of considering the overfitting problem when exploring the facial landmarks that decrease the accuracy of the facial expression classification. [2] In this paper, authors proposed an approach for recognizing the facial expression by using deep convolution neural network (DCNN) with local gravitational force descriptor. Initially, the local features are extracted from the facial image by using the local gravitational force descriptor. After extracting the local features, DCNN is used to classify the facial expressions by dividing it into two major branches.

The first branch of DCNN is used to extract the geometric features such as curves, edges, and lines present in the facial images. The second branch of DCNN is used to extract holistic features and classification. The classification score of facial expressions is computed by using a technique named score level fusion. Here FER-2013, CK+, and JAFFE datasets are used for experiment. Geometric features such as curves, lines, and edges of the facial images are considered by using DCNN. However, these geometric features of the facial images are not enough to classify the facial expressions accurately that leads to high false-positive rate. Here, DCNN is used for both feature extraction and classification of facial expressions from the corresponding features but, DCNN felt difficult to classify the facial expression when the facial image contains rotation or some angle of tilt that leads to less accuracy and more layers were used for the three types of layers present in the DCNN. Especially the number of max-pooling layers present is 5. However, more max-pooling layers present in the DCNN increase the probability of losing the efficient features that leads to high false positive rate. Preprocessing process is performed to align the angle and pose of the facial image. However, lack of considering the illumination parameters leads to face occlusion, contrast, etc., which decreases the classification of facial expressions accuracy. [3] The author proposed an approach for facial expression recognition (FER) by using hybrid deep learning algorithms with pose considering face alignment. Initially, the pose-guided face alignment method is used to decrease the intra-class difference present in the facial images by considering three basic steps such as target pose discovery, template generation, and target matching. Angular symmetry is used for redundant features elimination and selects the efficient template by performing clustering with K-means clustering algorithm. Hybrid deep learning algorithms such as CNN and RNN are used for extracting the facial features and VGG-16 and ResNet are used for classification of facial expressions. Here Oulu-CASIA, CK+, AR, and JAFFE datasets are used for experiment. The results from the pose-guided face alignment method were clustered by using K-means clustering algorithm. However, this clustering algorithm cluster the dataset in k number, and the entire k number of clusters has a single cluster that leads to high complexity to retrieve the extracted features for classification which decreases the performance. In the hybrid deep learning algorithms, RNN also plays an important role in feature extraction. However, the process involved in RNN is difficult which increases the training complexity in classification of facial expression. ResNet is used for classify the facial expression from the extracted features but, the time duration of this network is high that affects the classification with high time complexity. Here, the features are extracted from the clustering output to classify the facial expressions but the clustering of facial images is not enough to extract the features accurately that leads to high false-positive rate. [4] The authors proposed an approach to discover the facial expressions by using ensemble rule with deep learning algorithm. The face expression recognition algorithms are classified into two types such as feature-based algorithm and convolutional neural network-based algorithm. Initially, the facial landmark is extracted from the input image. Perform frontalization by using frontalization algorithm to manage the pose by rotating and manage the brightness of the image. Shortcut CNN is used to extract the features from the facial image and for FER classification adaptive exponentially weighted average ensemble rule was used. FER-2013, JAFFE, and CK+ datasets were used for experiment. In this paper, the landmarks are detected with 68 feature points to perform facial frontalization but, FER needs few landmarks, for instance, the landmarks of jaw are not important to classify accurate facial expression. Considering all facial landmarks for frontalization increases the classification complexity and decreases the accuracy. Here, frontalization algorithm is used to consider the brightness rotation and pose of the facial images but, lack of considering the geometry features increases the false positive rate. CNN is used to extract the features from the frontalized facial image. However, CNN doesn't consider the coordinate frames of the facial images that lead to less accuracy in recognizing the facial expression. [5] The author proposed an approach to detect the facial landmark by performing semantic segmentation. Initially, the architecture of semantic segmentation was designed to segment the facial images by encoding them and feature maps of the facial images were extracted from the encoder and the feature maps are decoded by corresponding feature maps. Secondly, VGG-16 is used to extract facial landmarks from the feature maps and extract the features of the facial image. For classification of images, the feature maps are given to the softmax layer and classify the feature maps based on their weights. For experiment analysis authors created the dataset by own and used VGG-16 for training the facial images for classification of feature maps based on weights to create facial landmarks. However, VGG-16 takes more time to train the images and needs more bandwidth that affects the performance. [6] Authors proposed an approach to identify the facial expression from the facial images by using deep convolutional neural networks. Initially, edge computing is involved to ensure the privacy of the facial images from the cloud. The generative adversarial network (GAN) is modified by including circular consensus to create CycleGAN to train the facial images effectively. Information of class constraints is also included in the CycleGAN to improve style conversion process. The normal classifier of GAN i.e. discriminator is also modified by including an auxiliary expression classifier to classify the facial expressions efficiently. For experiment analysis authors used JAFFE, CK+, and FER-2013 datasets to identify the recognition rate and GAN is modified by adding circular consensus and an auxiliary expression classifier to create CycleGAN for better classification. However, GAN can't have the ability to predict the state of the facial image that leads to less accuracy. [7] Authors proposed an approach to recognize the facial expression by considering the pose and identity invariant of the facial image using dynamic multi-channel metric network (DML-Net). Initially, the DML-Net is used to learn the local and global features from various facial regions by using parallel convolutional networks. To discover the pose and identity-invariant of the facial expressions joint embedded feature learning is used. End-to-End training is performed for the facial images to recognize the facial expressions with low FER loss and overfitting. Here BU-3DFE, Multi-PIE, SFEW 2.0., and KDEF datasets are used to evaluate the facial expressions. DML-Net is used to extracting the features and classification of facial expressions from facial images. However, it is not suitable to classify the facial expressions in unconstrained environments i.e. various datasets. [8] Authors proposed an approach to identify the facial expression by using CapsField technique. It is a technique with the combination of convolutional neural network (CNN) and capsule neural network. Initially, CNN is used to extract the features from the pre-processed array of facial images. Capsule network is used to route the facial features to select the features hierarchically. These methods reduce the redundancy of the features and make the classification process easier with effective classification of facial expressions. Here authors proposed the dataset named light field faces in the wild (LFFW) for experimental analysis. CNN was used with capsule network to extract the features and classify the facial expressions. However, CNN does not consider the alignment and position of the facial images that leads to less accuracy. [9] Authors proposed an approach to recognize facial expression by using frequency neural network (FreNet). Initially, the facial images are pre-processed by considering rotation correction; ensuring the eyes are on horizontal line, and resizing using discrete Fourier transform (DCT). Block-FreNet is introduced to reduce the dimension and effective feature learning. The low-level features are extracted by using learnable multiplication kernel. After extracting the low-level features, high-level features are extracted by using a summarization layer. To classify the facial expressions from the extracted features ANN-based classification layer was used. FER-2013 dataset was used for experimental analysis. DCT is used to perform preprocessing for the facial images considering resizing, rotation correction, etc., but lack of considering the geometrical features leads to high false-positive rate. [10] Authors proposed an approach to identify the facial expression by considering the geometry aware pose invariant using deep learning mechanisms such as generative adversarial network (GAN). Initially, the poses from the facial images were based on angles, and according to facial expressions, the facial landmarks were generated facial landmarks. The identity representation is detached by using the facial shape geometry which was provided by the facial landmarks. GAN is used to extract the features and perform classification for detecting the facial expressions from the extracted features. Here BU-3DFE, SFEW, and Multi-PIE datasets were used for performing experimental analysis. GAN is used to extract the features from the facial images and perform classification for detecting the facial expressions. However, at the same time, GAN trains the generator and discriminator model that increases the training complexity. [11] Authors proposed an approach to recognize the facial expression by considering the alignment and synthesis of facial images by using joint deep learning model. Initially, facial alignment is learned to generate the geometry code to perform facial synthesis. In the facial synthesis, three types of losses are considered such as discriminator loss, content similarity loss, and perceptual loss. Facial features are extracted i.e the expression and geometry code are extracted by considering the pose invariant and corresponding facial landmarks. After extracting the codes the facial expression is classified by using the softmax layer. The datasets used are BU-3DFE, SFEW, and Multi-PIE. Here, normalization is performed for the facial landmarks by considering the inter-ocular distance. However, lack of several illumination parameters such as shadowing effect, contrast, etc., leads to less accuracy. [12] Authors proposed an approach to detect the racial landmarks by using heatmap offset regression technique. Initially, the regression network is divided into two stages such as structural hourglass network (SHN) and global constraint network (GCN). Preprocessing is performed to extract the features accurately. SHN is used to discover the facial landmarks at initial condition by using the heatmap. Improved inception ResNet is used in SHN for contextual feature representation learning. GCN is used to perform offset estimation for efficient landmark location. Loss function is proposed to improve the facial landmarks in a precise manner. Finally, the outputs from SHN and GCN are combined for accurate prediction of facial landmarks. The datasets used are 300W, AFLW, 300-VW, and COFW and heatmap offset regression is used to predict the facial landmarks in a precise manner. However, heatmap doesn’t adapt to the color of the images that leads to less accuracy.

CHAPTER 3

RESEARCH METHODOLOGY

In this research, we have given importance to recognize facial expressions by using hybrid deep learning method. The accuracy of the facial expression is improved by performing preprocessing, segmentation, landmark detection, and feature extraction from the facial images. The important aim of this research is to develop expression recognition with high accuracy based on effective segmentation and classification. For FER, we have taken CK+ (Cohn Kanade) dataset. It have three sequential phases such as,

- Bi-level preprocessing
- Clustering-based Segmentation and Facial Landmark Detection
- Multi-Feature Extraction and Facial Expression Classification

3.1 Bi-level Preprocessing

We take the facial image as input from the given dataset i.e. CK+ dataset. Initially, bi-level preprocessing is performed for the facial images such as illumination normalization and pose normalization. In the illumination normalization occlusion, blurring effect; contrast, and shadowing effect are reduced by using **Grayscale algorithm** to convert the normal facial image to grayscale image to eliminate these respective effects. After performing illumination normalization, facial angle and geometry are normalized by performing pose normalization by using **Geometry-based Polar transformation** to remove the background from the facial image and obtain frontal facial view for better facial expression recognition[11]. The angle of the face is rotated within the image to get the frontal face pose by using polar transformation. At the same time, the position of the face is also important to recognize accurate facial expressions by using geometry transformation.

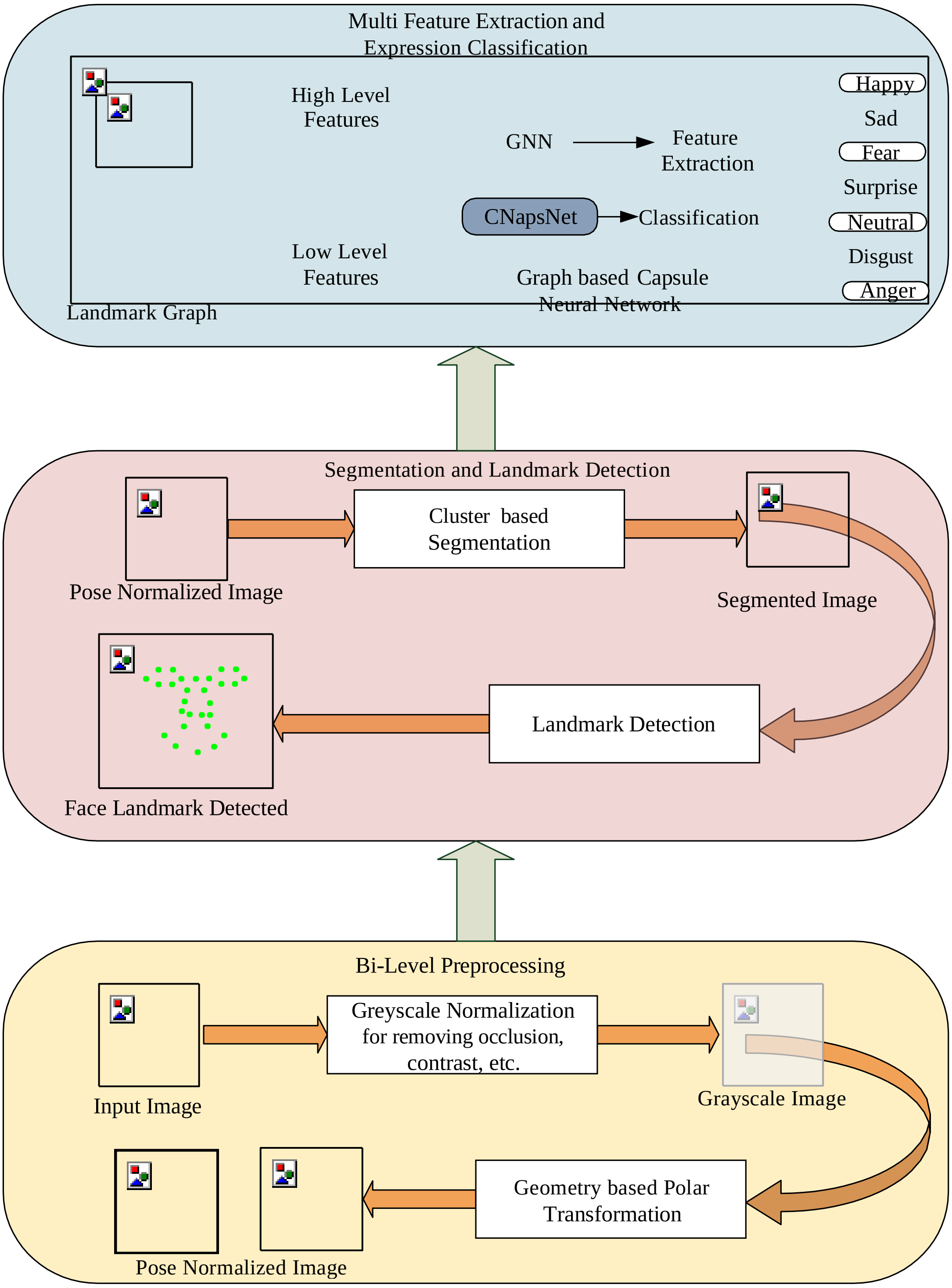


Fig 3.1 Overview

3.2 Clustering-based Segmentation and Facial Landmark Detection

After performing successful bi-level preprocessing, segmentation and facial landmark detection are performed. Segmentation of facial images is mainly performed to recognize the facial objects such as eyebrows, eyes, nose, mouth, and lips by using **Improved Fuzzy C-means clustering algorithm.** Traditional fuzzy c-means clustering algorithm is used for partitioning the images into segments but, it can't determine the center of the cluster. To overcome such drawbacks improved fuzzy c-means clustering algorithm is used to initialize the cluster's center by using **Firefly algorithm** along with fuzzy c-means clustering algorithm. After fulfilling the segmentation process, detection of facial landmarks is also performed effectively. 44 Feature points are used in the facial landmark for better facial expression recognition. 10 feature points are used for eyebrows (5 feature points for each eyebrow), 12 feature points for eyes (6 feature points for each eye), 9 feature points for nose, and 13 feature points for mouth (7 feature points for lips) [13].

3.3 Multi-Feature Extraction and Facial Expression Classification

Various features are extracted from the landmark detected image and it is classified into two levels such as high-level features and low-level features. The high-level features are extracted based on the facial objects such as eyebrows, eyes, nose, lips, and mouth and the corresponding features such as *eyebrow slant, eye size, eye spacing, pupil size, nose length, nose width, nose wrinkle, mouth openness, mouth width, mouth curvature, tight lips, and lips droop*. The low-level features are *shape, texture, and color*. These features are extracted and classified by using deep learning algorithms such as **Graph-based Capsule Neural Network (GCapsNet).** The traditional capsule neural network has high complexity for large number of data sets. To overcome this drawback we integrate the capsule neural network with Graph neural network. Large number of datasets is managed by using Graph neural network and the classification is performed by capsule neural network to select the optimal features to reduce the overfitting and increase the accuracy[14]. Seven important facial expressions are classified depend upon the extracted features; they are *neutral, happy, sad, fear, disgust, surprise, and anger*. [13] Evaluation of the work is performed by considering the following performance metrics,

- Accuracy
- Precision
- Recall
- Confusion matrix
- Facial landmark detection error

CHAPTER 4

IMPLEMENTATION AND DISCUSSION

4.1 Dataset

CK+ Dataset is used for facial expression recognition using deep learning. Cohn-Kanade (**CK+**) dataset having 593 images ranging from 18 to 50 years of age with both the genders and these images are divided into seven expression classes: anger, contempt, disgust, fear, happiness, sadness, and surprise. CK+ dataset is shown in the figure 4.1 while loading of selected image is shown in the figure 4.2 . The facial image is taken as input from the CK+ dataset and the selected image loaded successfully as follows ,

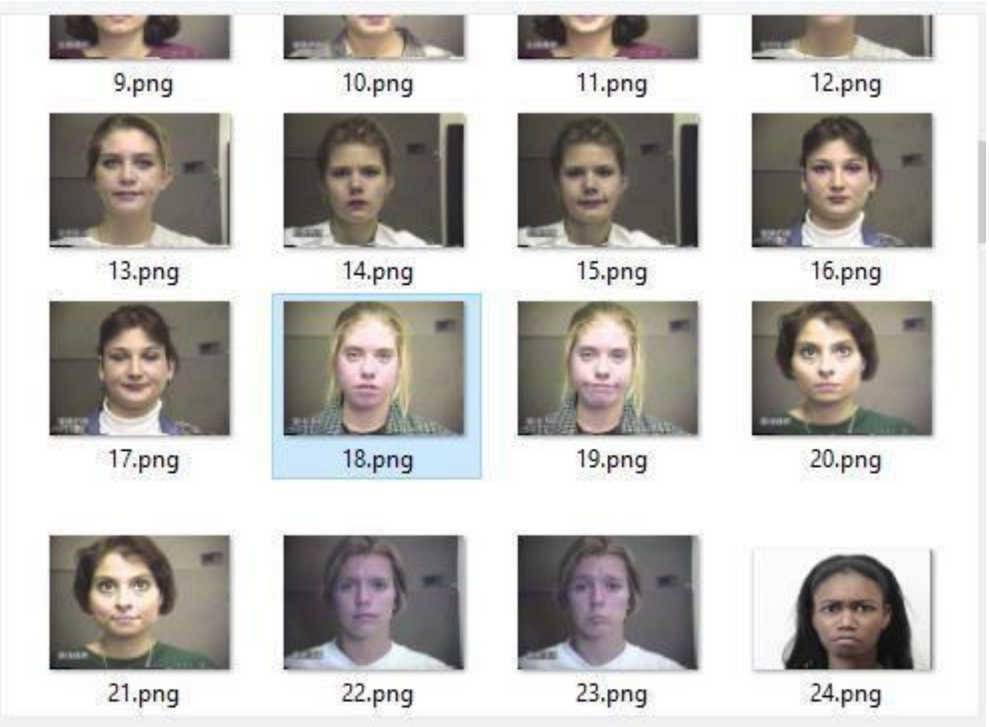


Fig 4.1 Dataset

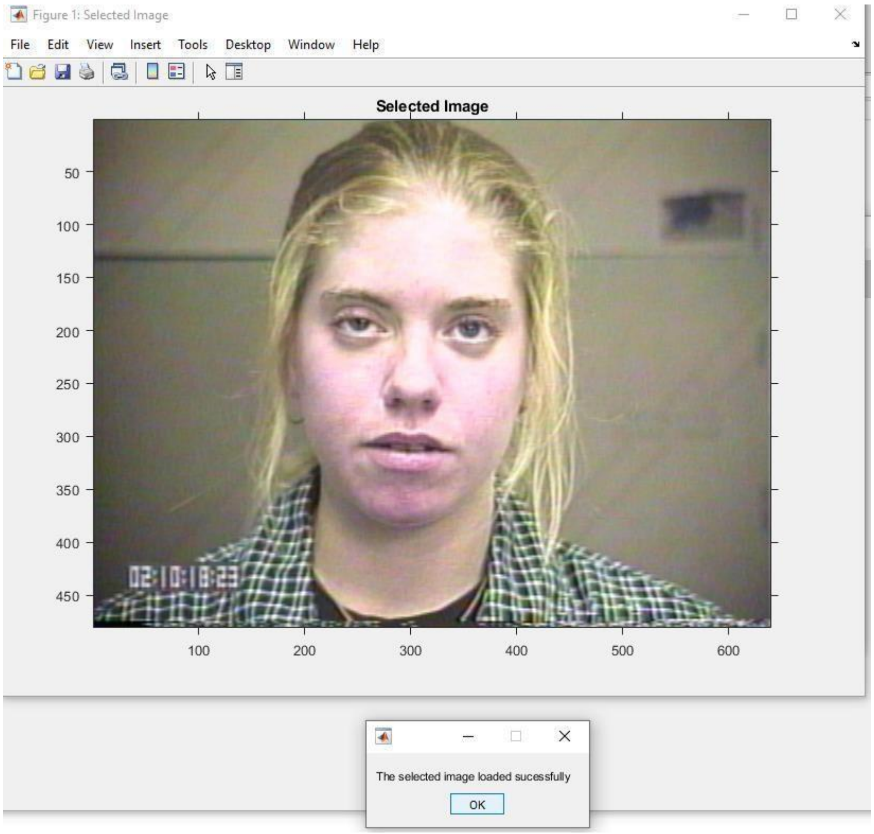


Fig 4.2 Image loaded successfully

4.2 Bi-level Preprocessing

In the bi-level preprocessing, Grayscale algorithm is used for the conversion of facial image which is taken from CK+ dataset in the previous step into the grayscale image. This grayscale algorithm can be considered as the illumination normalization. After performing the Grayscale algorithm, facial angle and geometry are adjusted by doing the pose normalization by performing the Polar transformation to remove the background from the input facial image which was taken as input earlier from the CK+ dataset. Figure **4.3** shows the result of Grayscale algorithm which is Gray image while figure **4.4** shows the bi-level preprocessing which is given below,

Fig 4.3 Grayscale Image

Fig 4.4 Bi-level Preprocessing

4.3 Segmentation

In this step, facial landmark detection and segmentation is performed Segmentation of input facial images is done for the recognization of facial objects such as eyebrows, eyes, nose, mouth, and lips. Here, we used Fuzzy C- means clustering algorithm. Before understanding the Fuzzy C-means clustering algorithm, we first need to understand the basic meaning of clustering. It is basically a mechanism of making the group of the similar data points into different clusters. In simple words, we can say that clustering is a technique, which groups the unlabelled dataset. The datapoints with similarities remain in a group that has less or no similarities with another group. Fuzzy clustering is a type of clustering in which a datapoint can remain in more than one group. It is only the reason that the fuzzy clustering algorithm gives better results in comparison to the K-means clustering algorithm because one datapoint can belong only to the one cluster.

4.3.1 Fuzzy C-means Clustering algorithm

Step 1 – [15] Randomly select the data points and these data points should be initialize into some number of cluster centers. The value of "c" will be the number of cluster centers which we selected randomly.

Step 2 - Now we need to calculate the centroid by using the given formula which is given below,

$$\mathrm{V}_{ij} = (\sum_{1}^{n}(\gamma_{ik}^{m} * x_k) / \sum_{1}^{n} \gamma_{ik}^{m}$$

Step 3 - Now we need calculation to calculate the distance among each data points from the centroid.

Step 4 - Now we need to update the membership values by using the given formula which is given below,

$$\gamma = \sum_{1}^{n} (d_{ki}^{2}/d_{kj}^{2})^{1/m-1}]^{-1}$$

Step 5 - Now we have to repeat the steps (2-4) until the membership values becomes constant.

Step 6 - Now we have to defuzzify the values of membership which are obtained previously.

Here figure 4.4 shows the facial landmark detection while figure 4.5 shows the result of segmentation which is given below,

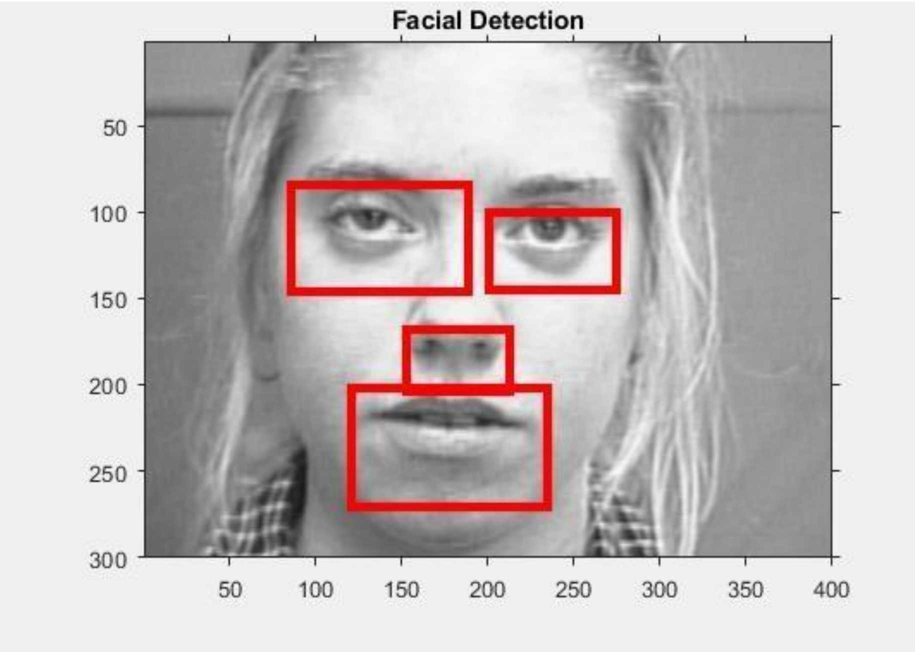


Fig 4.4 Facial Detection

Fig 4.5 Segmentation

4.4 Feature Extraction

There are many features which are find out from the landmark detected image and divided into two levels such as high-level features and low-level features. The high-level features are based on the facial objects such as eyebrows, eyes, nose, lips, and mouth and the corresponding features such as eyebrow slant, eye size, eye spacing, pupil size, nose length, nose width, nose wrinkle, mouth openness, mouth width, mouth curvature, tight lips, and lips droop. The low-level features are shape, texture, and color. Figure 4.6 shows the high level feature extraction while Figure 4.7 shows the low level feature extraction which are given below,

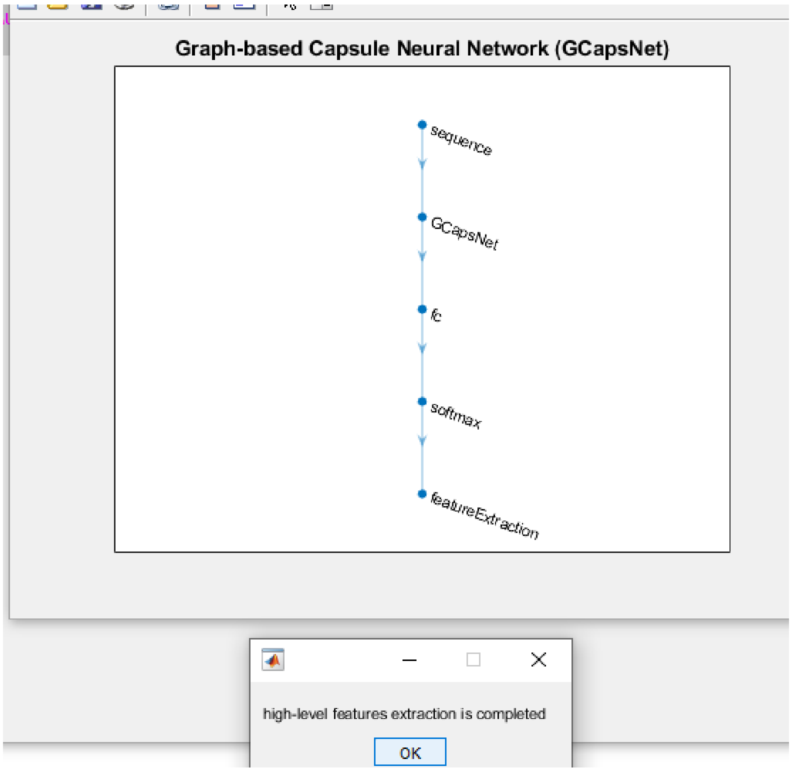


Fig 4.6 High level features extraction

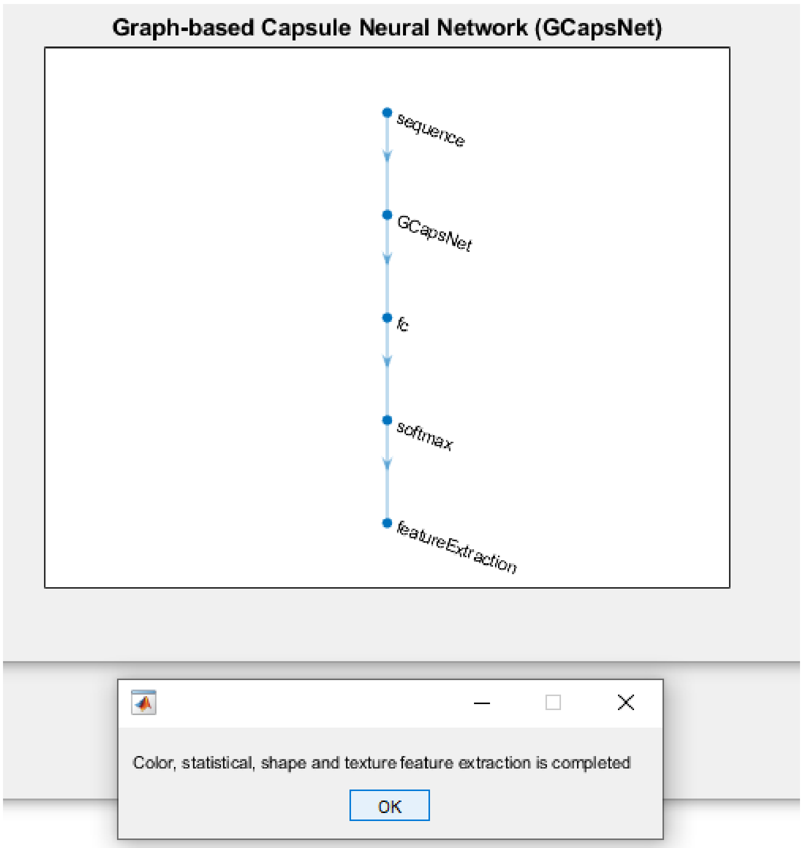
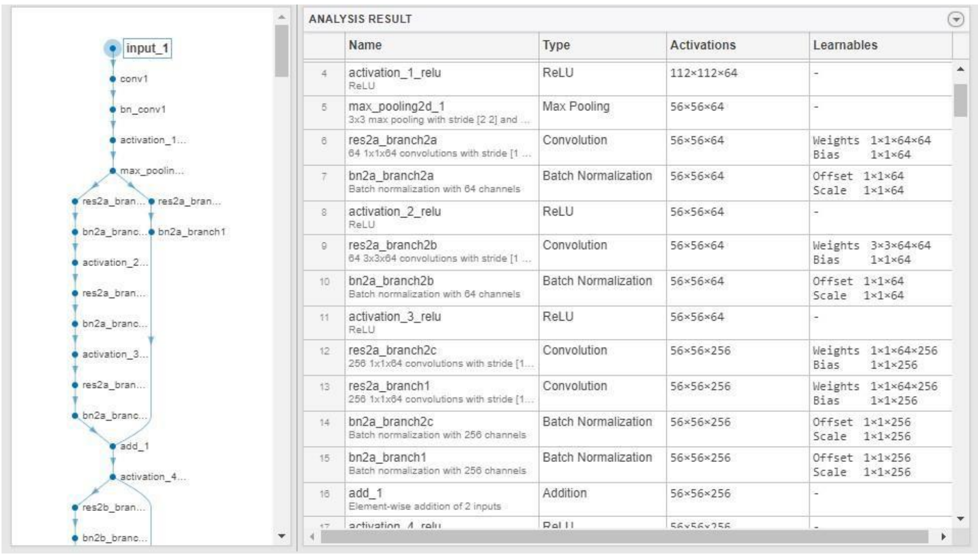


Fig 4.7 Low level feature extraction

4.5 Classification

The deep learning network analyzer for classification is shown in the Figure 4.8 while the deep learning result is shown in the figure 4.9 which are given below,



	Name	Type	Activations	Learnables
4	activation_1_relu ReLU	ReLU	112×112×64	-
5	max_pooling2d_1 3x3 max pooling with stride [2 2] and ...	Max Pooling	56×56×64	-
6	res2a_branch2a 64 1x1x64 convolutions with stride [1 ...	Convolution	56×56×64	Weights 1×1×64×64 Bias 1×1×64
7	bn2a_branch2a Batch normalization with 64 channels	Batch Normalization	56×56×64	Offset 1×1×64 Scale 1×1×64
8	activation_2_relu ReLU	ReLU	56×56×64	-
9	res2a_branch2b 64 3x3x64 convolutions with stride [1 ...	Convolution	56×56×64	Weights 3×3×64×64 Bias 1×1×64
10	bn2a_branch2b Batch normalization with 64 channels	Batch Normalization	56×56×64	Offset 1×1×64 Scale 1×1×64
11	activation_3_relu ReLU	ReLU	56×56×64	-
12	res2a_branch2c 256 1x1x64 convolutions with stride [1...	Convolution	56×56×256	Weights 1×1×64×256 Bias 1×1×256
13	res2a_branch1 256 1x1x64 convolutions with stride [1...	Convolution	56×56×256	Weights 1×1×64×256 Bias 1×1×256
14	bn2a_branch2c Batch normalization with 256 channels	Batch Normalization	56×56×256	Offset 1×1×256 Scale 1×1×256
15	bn2a_branch1 Batch normalization with 256 channels	Batch Normalization	56×56×256	Offset 1×1×256 Scale 1×1×256
16	add_1 Element-wise addition of 2 inputs	Addition	56×56×256	-

Fig 4.8 Deep learning network analyzer

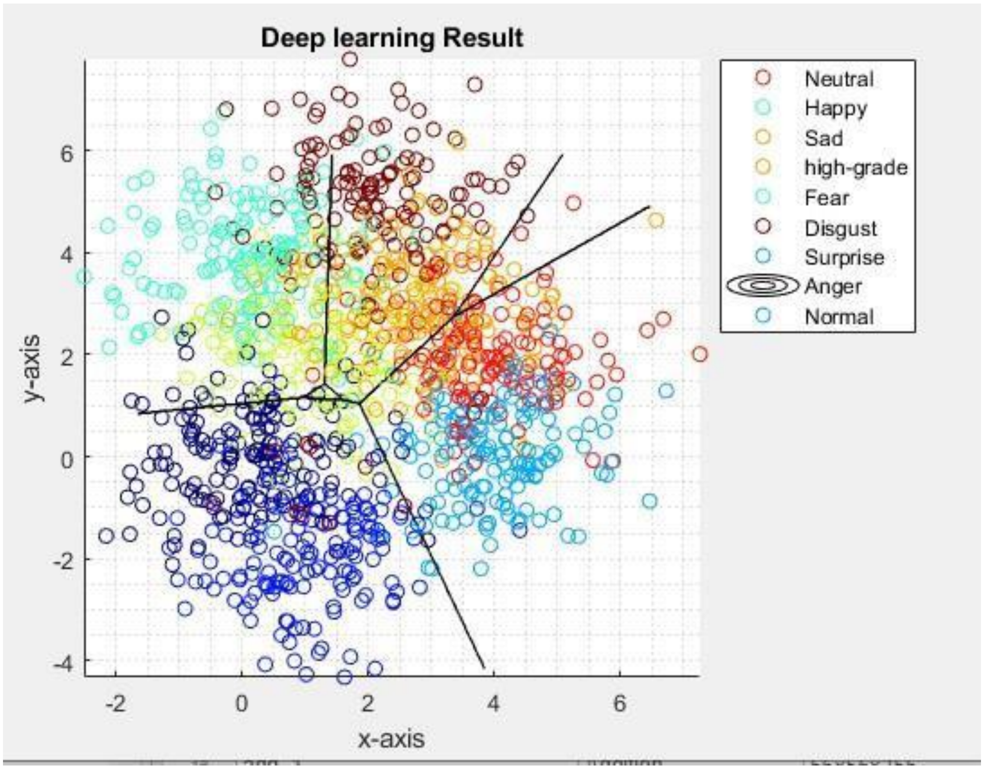


Fig 4.9 Deep Learning Result

Classification is done by using the integration of two deep learning algorithms such as Capsule Neural Network and Graph Neural Network. To overcome from the high complexity of Large number of datasets is managed by using Graph neural network and the classification is performed by capsule neural network to select the optimal features to reduce the overfitting and increase the accuracy. Seven important facial expressions are classified depend upon the extracted features; they are neutral, happy, sad, fear, disgust, surprise, and anger.

The classification result, **" Sad"** is shown in the Figure 4.10 as follows:

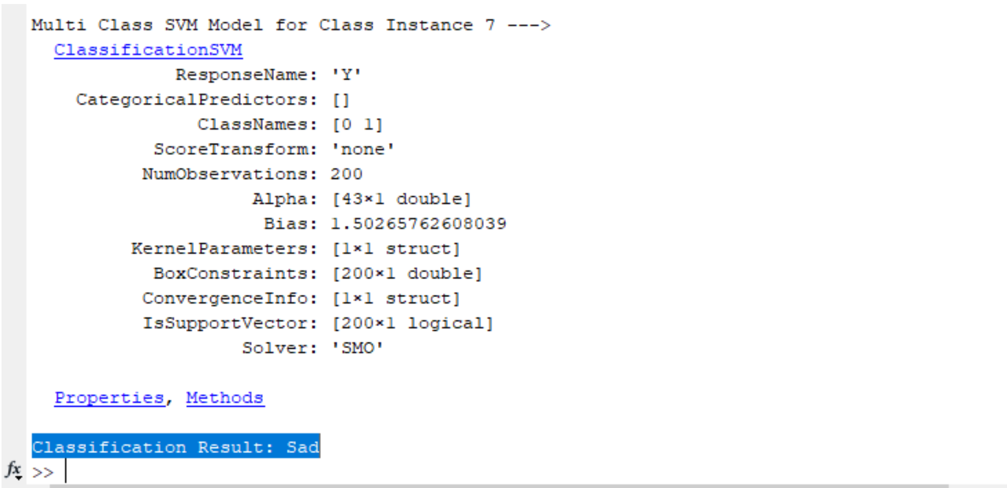


Fig 4.10 Classification result

4.5.1 Capsule Neural Network

[10] Capsule is basically a collection of neurons which gets activated for properties of an object like position and size. CapsNet is a deep learning technique and a type of artificial neural network which overcomes the drawbacks present in the Convolutional neural network. To do so, there will be 40% enhancement of Accuracy in the Capsule Neural Network. It is broadly used in the image recognition field by detecting the features from the pixels of images. Capsule is a vector along with direction as an output. So, it is a vector quantity while neuron is a scalar quantity because it doesn't have any direction. There are basically two major components of a capsule neural network that are Encoder and Decoder. There are six layers in which first

three layers are the part of the encoder while the other three layers are the part of the decoder.[10] The encoder is responsible for taking the image as an input and convert it in the 16-dimensional vector. The decoder is responsible for taking the 16-dimensional vector as an input and tries to reconstruct the input image with the help of data which is fetched by using encoder. To do so, it can predict in the field of image recognition. The initial layer of Encoder is convolutional layer network which helps in the finding of the basic feature of an image. The second and third layer of the Encoder are the PrimaryCaps Network and DigitCaps Network respectively.

4.6 Performance Metrics

Finally, we evaluate the following performance metrics : Accuracy, Precision, Recall, Facial landmark detection error, Confusion matrix. Figure 4.11 shows the Accuracy, Recall metrics while the Precision and Facial landmark detection error can be clearly seen in the figure 4.12 and the Confusion matrix can be seen in the figure 4.13 which are given below,

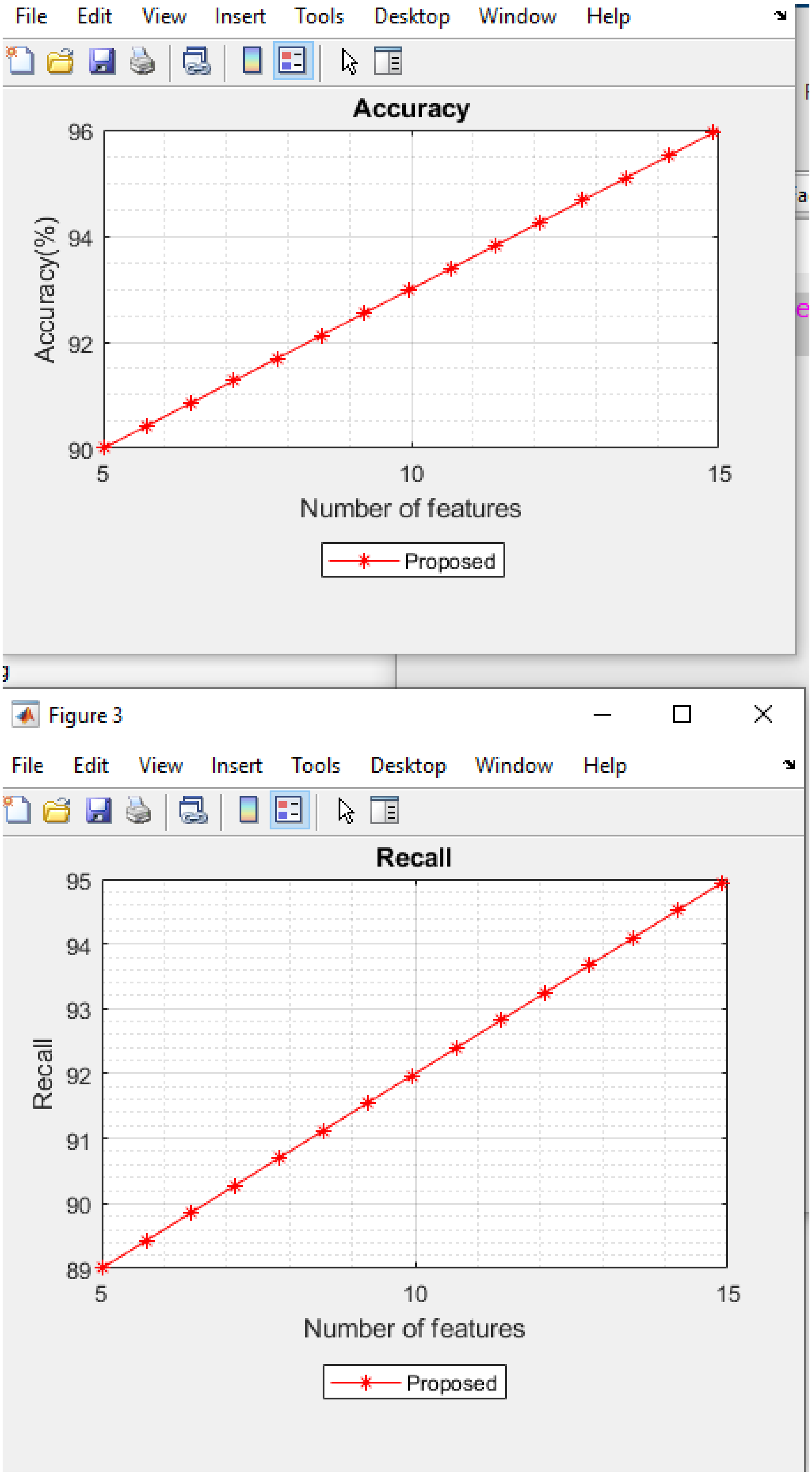


Fig 4.11 Accuracy, Recall metrics

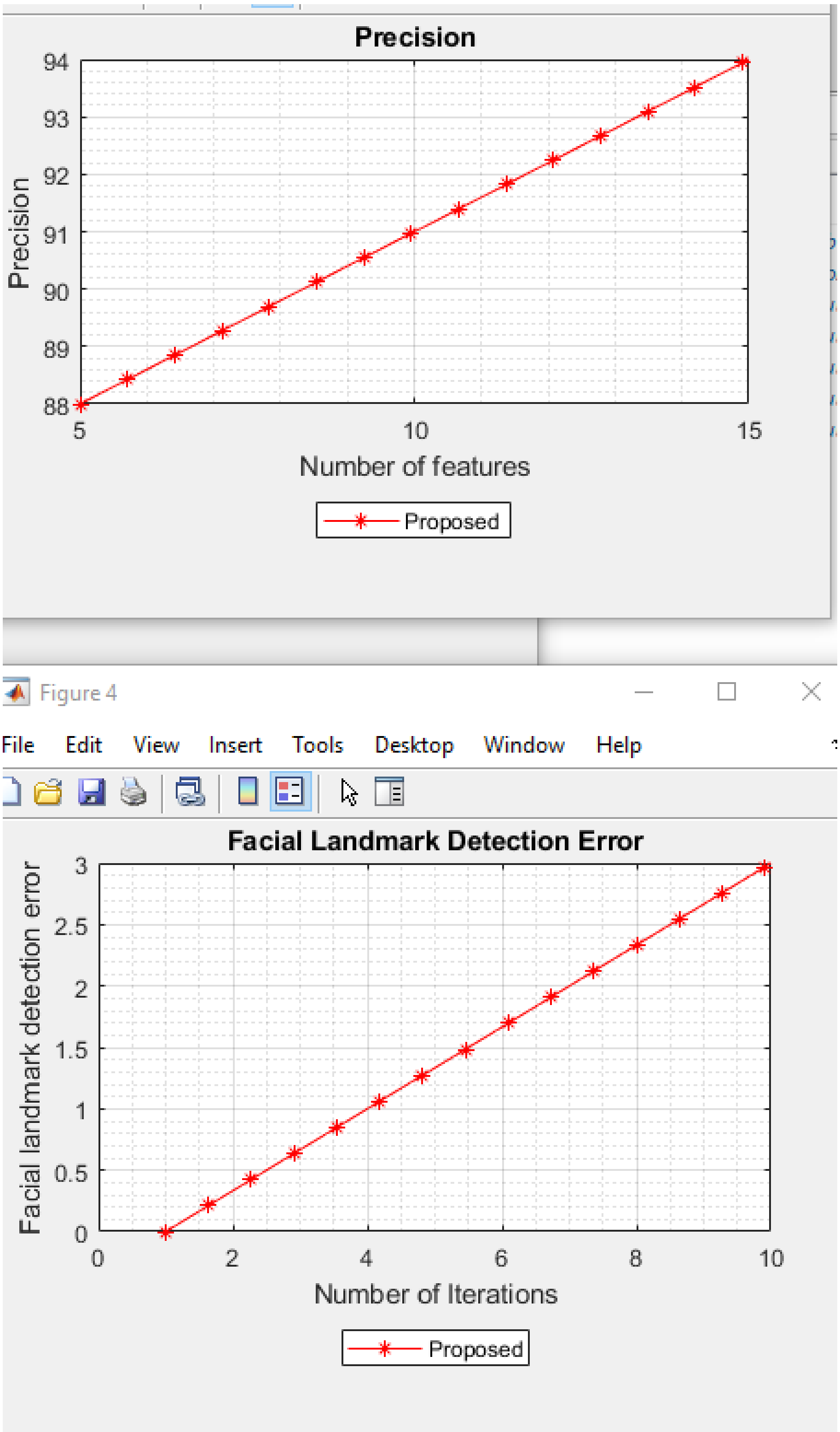


Fig 4.12 Precision and Facial landmark detection error

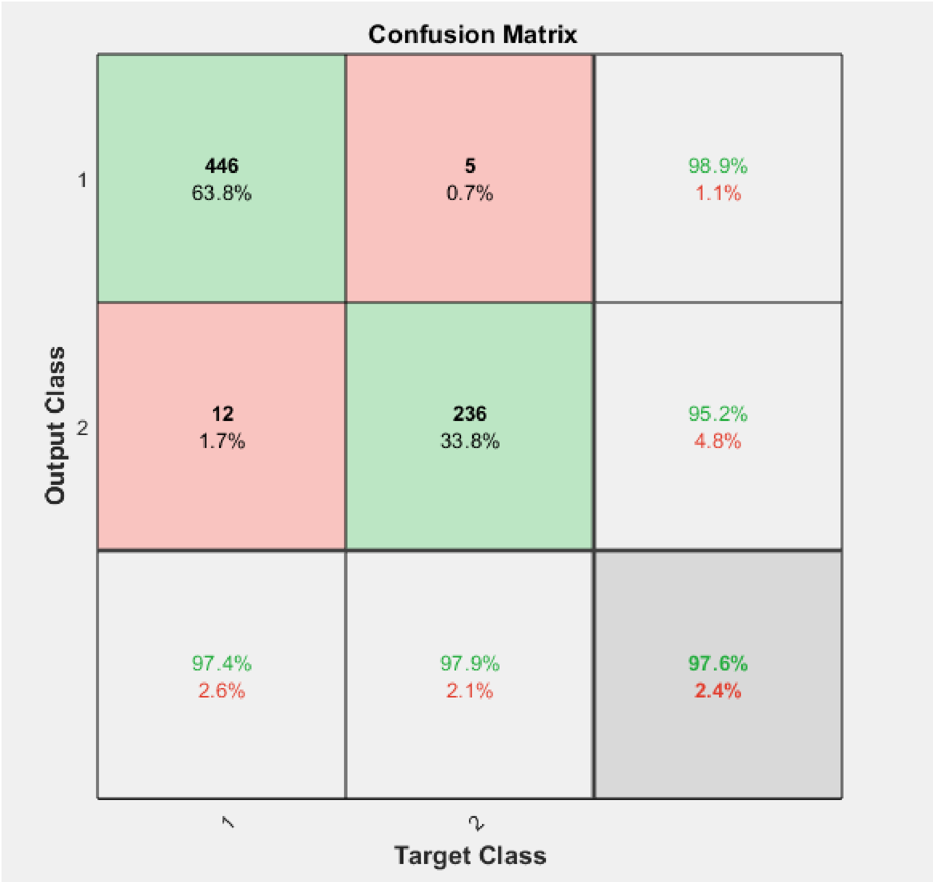


Fig 4.13 Confusion matrix

CHAPTER 5

CONCLUSION

[9]Deep learning covers various information, measurements which is the example of human made consciousness. FER became a challenge in the computer vision and deep learning field. Researchers are improving the accuracy day by day after doing innovative changes by using the deep learning techniques. [11] FER is used for classification of input facial images into seven different expressions such as Sad, Fear, Anger, Happy etc. In this research, we evaluate the performance by using the following Performance Metrics such as Accuracy, Recall, Precision, Facial Landmark Detection Error, Confusion Matrix.

CHAPTER 6

REFERENCES

[1] Liu, C., Hirota, K., Ma, J., Jia, Z., & Dai, Y. (2021). Facial Expression Recognition Using Hybrid Features of Pixel and Geometry. *IEEE Access, 9*, 18876-18889.

[2] Mohan, K., Seal, A., Krejcar, O., & Yazidi, A. (2021). Facial Expression Recognition Using Local Gravitational Force Descriptor-Based Deep Convolution Neural Networks. *IEEE Transactions on Instrumentation and Measurement, 70*, 1-12.

[3] Liu, J., Feng, Y., & Wang, H. (2021). Facial Expression Recognition Using Pose-Guided Face Alignment and Discriminative Features Based on Deep Learning. *IEEE Access, 9*, 69267-69277.

[4] Tsai, K., Tsai, Y., Lee, Y., Ding, J., & Chang, R.Y. (2021). Frontalization and adaptive exponential ensemble rule for deep-learning-based facial expression recognition system. *Signal Process. Image Commun., 96*, 116321.

[5] Kim, H., Kim, H., Rew, J., & Hwang, E. (2020). FLSNet: Robust Facial Landmark Semantic Segmentation. *IEEE Access, 8*, 116163-116175.

[6] Chen, A., Xing, H., & Wang, F. (2020). A Facial Expression Recognition Method Using Deep Convolutional Neural Networks Based on Edge Computing. *IEEE Access, 8*, 49741-49751.

[7] Liu, Y., Dai, W., Fang, F., Chen, Y., Huang, R., Wang, R., & Wan, B. (2021). Dynamic multi-channel metric network for joint pose-aware and identity-invariant facial expression recognition. *Information Sciences, 578*, 195-213.

[8] Tang, Y., Zhang, X., Hu, X., Wang, S., & Wang, H. (2021). Facial Expression Recognition Using Frequency Neural Network. *IEEE Transactions on Image Processing, 30*, 444-457.

[9] Zhang, F., Zhang, T., Mao, Q., & Xu, C. (2020). Geometry Guided Pose-Invariant Facial Expression Recognition. *IEEE Transactions on Image Processing, 29*, 4445-4460.

[10] Zhang, F., Zhang, T., Mao, Q., & Xu, C. (2020). A Unified Deep Model for Joint Facial Expression Recognition, Face Synthesis, and Face Alignment. *IEEE Transactions on Image Processing, 29*, 6574-6589.

[11] Zhang, J., Hu, H., & Feng, S. (2020). Robust Facial Landmark Detection via Heatmap-Offset Regression. *IEEE Transactions on Image Processing, 29*, 5050-5064.

[12] D. Ghimire and J. Lee, "Geometric Feature-Based Facial ExpressionRecognition in Image Sequences Using Multi-Class AdaBoost andSupport Vector Machines," Sensor, vol. 13, pp. 7714–7734, 2013.

[13] X. Feng, M. Pietik¨ainen, and A. Hadid, "Facial Expression RecognitionBased on Local Binary Patterns 1," Pattern Recognit. Image Anal., vol.17, no. 4, pp. 592–598, 2007

[14] A REVIEW ON FACIAL EXPRESSION RECOGNITION SYSTEMSS. Shaul Hammed1 , Dr.A. Sabanayagam2, E. Ramakalaivani3 ISSN-2394-5125 Vol 7, Issue 4, 2020

[15] F. Ahmed, H. Bari, and E. Hossain, "Person-Independent Facial Expression Recognition Based on Compound Local Binary Pattern (CLBP),"Int. Arab J. Inf. Technol., vol. 11, no. 2, pp. 195–203, 2014

[16] Sepas-Moghaddam, A., Etemad, A., Pereira, F., & Correia, P.L. (2021).

CapsField: Light Field-Based Face and Expression Recognition in the Wild Using Capsule Routing. IEEE Transactions on Image Processing, 30, 2627-2642.

CHAPTER 7

PUBLICATION

Dr. Manoj Sethi and Ankit Sharma, “Multi-Feature Aware Pose and Geometry based Facial Expression Recognition using Deep Learning” , Journal of Hunan University Natural Sciences, ISSN 1674-2974, Vol. 49. No. 04. , 2022.

Abstract

Facial expressions are the primary way to express intentions and emotions. In real life, computers can understand the emotions of humans by analysing their facial expressions. Facial Expression Recognition (FER) plays an important role in human-computer interaction necessity and medical field [1]. In the past, the facial features are extracted manually for recognizing the expressions. In the present, it is an important hotspot in computer vision, Internet of Things, and artificial intelligence fields. Certain processes are involved to recognize the facial expression in an efficient manner such as Preprocessing , Segmentation, Feature Extraction, Classification.

Abstracting and Indexing

The Journal of Hunan University Natural Sciences is abstracted and/or indexed by the following:

- **SCOPUS (Elsevier)**
- Web of Science
- MathDi
- Index Copernicus
- Ulrich's Periodicals Directory
- Google Scholar
- ROAD

- EdNA Online Database
- ERIC - Education Resources Information Center
- Education Research Index
- Higher Education Research Data Collection
- Journal Seek
- PsycINFO
- TOC Premier Database
- ZDB - Zeitschriftendatenbank

PAPER NAME

MULTI-FEATURE AWARE POSE AND GEO METRY BASED FACIAL EXPRESSION RE COGNITION USING DEEP LEARNING

AUTHOR

ankitsharma_2K20CSE05.pdf

WORD COUNT

6427 Words

CHARACTER COUNT

36580 Characters

PAGE COUNT

41 Pages

FILE SIZE

832.8KB

SUBMISSION DATE

May 23, 2022 8:26 PM GMT+5:30

REPORT DATE

May 23, 2022 8:27 PM GMT+5:30

9% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 7% Internet database
- 2% Publications database
- Crossref database
- Crossref Posted Content database
- 6% Submitted Works database

Excluded from Similarity Report

- Bibliographic material
- Quoted material
- Small Matches (Less then 11 words)
- Manually excluded text blocks