Deep convolutional neural network based facial expression identification using face-parsing method

Raju Yadav Mothukupally[1*], P *Chandra Sekhar* Reddy [2].

[1]PG Scholar, *Computer Science and Engineering*, *GRIET,* Hyderabad, Telangana, India.
[2]Professor, *Computer Science and Engineering*, *GRIET*, Hyderabad, Telangana, India.

Abstract: Face parsing methodology may be a one amongst the advancements in pc vision that analyses the surface synthesis of the external body part, to amass bits of information on options needs correct pixel segmentation of various components of face like (mouth, nose, eyes etc.). Same means the analysis on feeling recognition plays a eventful role in communication and interactions of humanity and additionally relevant to psychological activities. Considering the disadvantage that totally different completely different components of face contain different quantity of knowledge for face expression and also the weighted perform are not an equivalent for various faces. In keeping with analysis, the image classification task ordinarily drives North American country to the notable Convolutional Neural Network (CNN) during which we tend to ar victimization VGG19 model. beyond exploring around however CNN, sometimes performs for greyscale photos, we tend to selected to start from 3 consecutive convolutional layers followed by a most pooling layer, basic exploit work for convolutional layer and "relu" is used, even as an analogous artefact pattern. The highlights to be known victimization the convolutional layer distended to 128 layers from thirty-two, it is suggestable that multi-layered structure (with increasing layers) that performs and results the most effective outcomes for the DNN model. At last, the CNN layer is 1st smoothened and afterwards expertise 2 many dense layers to reach the yield layer during which SoftMax activation perform is used for multiclass classification. We tend to victimization Cohn-Kanadre face expression dataset of seven expressions like contempt, anger, disgust, happiness, fear, disappointment and surprise.

1. Introduction

With the features of current innovative technology, our fits of hunger went high and it ties no limits. In the current time, enormous research and experimentation on in the field of Machine Learning, Deep Learning and Image processing. The Image Processing is an immense region of examination in the current research world and its applications are exceptionally far-reaching. In this field of sign handling where both the info and yield signals are pictures. Perhaps the main utilizations of Image training are Facial acknowledgment. Our feeling is uncovered by the appearances on our countenances. Outward appearances assume a significant part in relational correspondence. Outward appearance is a nonverbal logical signal that gets communicated in our countenances according to our feelings. Programmed acknowledgment of outward appearance assumes a significant part in man-made reasoning and advanced mechanics and in this manner, it takes the age of a person into account. Some application identified with this method and incorporates Identification of individual and control access, Videophone and Teleconferencing, Forensic application, Human-Computer Interaction, Automated Surveillance, Cosmetology, etc.

Identifying the emotion of a human is experimented using numerous methods using an additional resources or data of individual personnel. It very well may be viewed as a subsequent advance to confront identification where we might be needed to set up a second layer of safety, where alongside the face, the feeling is likewise distinguished. This can be valuable to confirm that the individual remaining before the camera isn't only a 2-dimensional portrayal [2]. Another significant area where we see the significance of feeling identification is for business advancements.

---

[*] Corresponding author: rajuyadavmothukupally@gmail.com

Considering the disadvantage that totally different components of face contain different quantity of knowledge for face expression and therefore the weighted perform aren't identical for various faces [1]. Most organizations blossom with client reactions to every one of their items and offers. If a mechanized thinking system can get and perceive progressing sentiments reliant upon customer picture or video, they can make a decision on whether the customer appreciated or despised the thing or offer. We have seen that security is the basic role behind perceiving any person. It will in general be established on finger impression organizing, voice affirmation, passwords, retina acknowledgment, etc. recognizing the objective of the individual can similarly be basic to divert threats. This can be useful in weak regions like air terminals, shows, and significant public social events which have seen numerous breaks lately. Human feelings can be named dread, disdain, nauseate, outrage, shock, misery, bliss, and nonpartisan. These feelings are exceptionally unpretentious. Facial muscle bending's are negligible and recognizing these distinctions can be trying as even a little contrast brings about various articulations [3]. Additionally, articulations of various or even similar individuals may differ for a similar feeling, as feelings are tremendously setting subordinate [4]. While we can zero in on just those regions of the face, which show a limit of feelings like around the eyes and mouth [5], how we are going to extract these emotions and classify them is yet a significant inquiry. Neural organizations and AI have been utilized for these undertakings and have gotten great outcomes. AI calculations have demonstrated to be extremely valuable InDesign acknowledgment and characterization. The main parts of any AI calculation are the highlights. In this research work, we will perceive how the highlights are separated and altered for calculations like SVM [2]. We will think about calculations and the component extraction methods from various research works. The dataset, which consists of human emotions, can be an excellent guide to contemplate the power and nature of order calculations and how they perform for various kinds of datasets. As a rule, before extracting the features or highlights for feeling identification, facial location calculations are pertained on the images. The emotion identification and landmarks extraction process as follows:

- Pre-processing the Data.
- Face detection.
- Landmarks extraction.
- Feature classification.

The primary target of this thesis is to generate the computerized Facial Expression Identification System which takes a human facial images containing some appearance as information and perceive and characterize it into 7 diverse demeanour class, for example, :

1.Anger
2.Contempt
3.Disgust
4.Fear
5.Happiness
6.Sadness and
7.Surprise

Fig. 1. Facial Expressions

2. Conceptual Background Study

In this segment, a depiction of important ideas for this task is introduced. This segment we are giving some insights about the foundation on the points to be discussed in this thesis. This foundation is cultivated by methods for a sequential correction of fields, for example, emotional processing, and AI. To depict the ideas, a top-down approach will be used. Besides, related exploration to the methodology utilized in this venture is presented, too.

2.1 Deep Computing

As portrayed by Rosalind Picard [6], "... full of feeling processing is the sort of registering that identifies with, emerges from, or impacts feelings or other emotional wonders". Full of feeling processing intends to remember feelings for the plan of innovations since they are a fundamental piece of errands that characterize the human experience: correspondence, learning, and dynamic. The primary establishments behind full of feeling figuring is that without feelings, people would not appropriately work as normal dynamic creatures. A few explore show that there is no such an unbelievable marvel as "unadulterated explanation". Feelings are engaged with dynamic since a completely logical methodology would transform into a very tedious cycle, not reasonable for day-by-day

assignments. Investigates around this specific theme have shown that the mind doesn't test every likely choice, yet it is one-sided by feeling to rapidly settle on a choice. A feeling is characterized as a class of characteristics that is inherently associated with the engine framework. At the point when a specific enthusiastic state is set off, the engine framework will give the relating set of guidelines to recreate the specific regulations associated with that class. Up until this point, feelings' significance has been tended to without contemplating human communication. One of the human feeling sympathies which limits that makes everyone mindful and gives us an understanding of what different creatures may be encountering from their present position. Also, sympathy permits us to construct cozy connections and solid networks. Consequently, it is principal towards favorable to social conduct, which incorporates social cooperation and discernment. Accordingly, it is vital for full of feeling registering to create approaches to appropriately gauge these specific tweaks since they can prompt a superior comprehension of a subject's passionate state. The primary two approaches are used for recognizing vocal and facial feelings. Nonetheless, in this thesis experiment, we are just facial feelings were utilized.

2.2 Facial Expression Recognition:

The analysis done by MR. Paul oceanographer has gotten principal to the development of this zone. These days, an oversized portion of the facial feeling acknowledgment examines depend upon Ekman' Facial Action writing [7]. One amongst the strategy conferred a framework that takes the highlights from to characterize of eyebrows, eyes and mouth utilizing a climbable quadrilateral [12]. This framework offers a designing between muscles movement of face and ablaze space. The principal reason behind this framework is to order human facial developments obsessed with the expressions of individuals. This grouping was at first evolved by an expert from Swedish, Mr. Carl-Herman Hjortsj. Figure a pair of displays a bunch of facial feeling units associated to their comparison facial motion. Be that because it might, this designing may confront many difficulties. For example, entertainers fake motions related to facial feelings. The shortage of a real inspiration driving the6 feeling does not forestall folks to counterfeit it. For example, an examination depicts once a patient, who is half-incapacitated, is approached to grin. At the purpose when it is asked, simply an aspect of the mouth raises. Nonetheless, when the patient is conferred to a joke, the two sides of the mouth raise. Henceforth, numerous ways in which to send a sense depend upon the foundation and nature of a selected

feeling. Regarding PCs, numerous prospects emerge to furnish them with the talents to speak and understand feelings. These days, it is possible to repeat the facial units of Mr. Ekman. This can furnish PCs with graphical appearances that provides a lot of characteristic communication [8]. In regard to acknowledgment, PCs have had the choice to perceive some facial classifications: satisfaction, shock, outrage, and nauseate. A lot of knowledge concerning facial feeling acknowledgment is found in dataset segment.

2.3 Facial Expressions:

Inferred various sorts of highlights from the picture and standardize that vector structure. We can utilize different sorts of strategies to recognize the feeling like computing the circles framed and the points between various parts such as nose, mouth, eyes, and so on following are a portion of the unmistakable landmarks are utilized for preparing AI algorithms.

For the numerical selections, the association between facial components is employed to fabricate a phase vector for work [14, 15]. Ghimire and Lee used 2 forms of numerical selections subject to the position and place of fifty a pair of facial accomplishment centers [15]. For the classifier, 2 distinctive ways in which unit of assessment given, either misuse multi-class AdaBoost with dynamic time winding, or employing a SVM on the upheld half vectors.

2.3.1 Facial Expression System (FES):

Facial Expression System is utilized to provide an Identification(Id) for each facialexpression. Each Id is called as an activity unit. A mix of activity units brings about an outward appearance. The miniature changes observed from the face muscles can be characterized by an activity unit which are Id-12 and Id-6. For instance, a grinning face could be characterized as far as activity units, which basically implicates the development of Id-Au6 cheek raiser muscle which brings about a cheerful face. Here Id-6 and Id-12 are the cheek raiser and lip corner puller. A FAS dependent on activity units ares decent frameworks to sort out based on facial muscles are associated with which demeanor. Constant facial models could be produced dependent on them.

Fig. 2. Face movements and actions

2.3.2 Facial Landmarks:

Landmarks of face are very essential inorder to utilize those for faceial identification and acknowledgment. Similar tourist spots can likewise be utilized on account of articulations. The library called Dlib has 68 facial landmarks identifier which provides 68 pin points of the face. In Fig. 2 shows every pin point of total 68 landmarks of the face. Utilizing the library "dlib" extricating the X and Y co-ordinates of every individual facial pin points. These 68 focuses can be partitioned into explicit zones like the nose, right eye, left eyebrow, mouth, left eye, right eyebrow, and jaw.
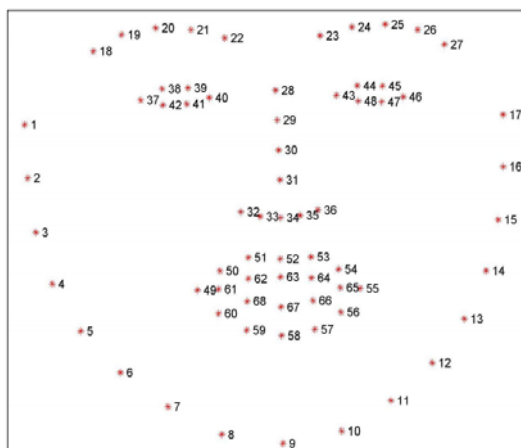
Fig. 2. Landmarks

2.3.3 Feature Recognition Tools:

landmarks are those which help in recognizing the emotion appropriately. Typically, the pictures are recognized based on edges and corners. For discovering edges and corners in pictures, we have many used element indicator calculations using a library called OpenCV, for example, the Harris corner finder. These element indicators consider a lot more factors like forms, structure, and arched. The Key-focuses are corner focuses or edges recognized by the component locator calculation. The component descriptor depicts the zone encompassing the central issue. The portrayal might be a lot which includes (X,Y) co-ordinates or crude pixel forces of a desired region. The descriptor and central issue together structured a nearby component. One illustration of a component was a histogram discriptor used for inclinations. Sphere,SURF, SIFT, and so on are a portion of the component descriptor calculations [9].

2.4 Data Set:

The Extended Cohn-Kanade (**CK+**) dataset contains 593 video arrangements from a sum of 123 distinct subjects, going from 18 to 50 years old with an assortment of genders and legacy. Every video shows a facial shift from the impartial appearance to a designated top demeanor, recorded at 30 FPS with a goal of either 640x490 or 640x480 pixels. Out of these recordings, 327 are named with one of seven articulation classes: outrage, disdain, disdain, dread, satisfaction, trouble, and shock. The CK+ information base is generally viewed as the most widely utilized research facility-controlled look characterization data set accessible, and is utilized in most of look arrangement strategies [10].

JAFFE [Japanese Female Facial Expressions] [13]: The JAFFE information contains 213 photos of ten totally unique female Japanese models showing seven face feelings (six fundamental facial feelings and one unbiased). Each picture was given a score dependent on six emotive words used to misuse sixty Japanese individuals. Every facial picture's essential size is 256 pixels by 256 pixels.

Figure 4: Anger Dataset

Figure 5 Contempt Dataset

Figure 6: Disgust Dataset

Figure 7: Fear Dataset

Figure 8: Happiness Dataset

Figure 9: Sadness Dataset

Figure 10: Surprise Dataset

2.5 VGG-19:

A significant convolutional neural framework in Deep learning is a class of significant, feed-forward significant acknowledging which have been applied precisely to visual imaging assessment. Stood out from various other picture portrayal counts, the VGG19 model uses reasonably minimal pre-preparing. It is viewed as one of the incredible vision model architecture till date. The most special thing about VGG19 is that as opposed to having countless hyper-parameter they zeroed in on having convolution layers of 3x3 filter with a stride 1 and consistently utilized a similar cushioning and max pool layer of 2x2 filter of stride 2. It follows this game plan of convolution and max pool layers reliably all through the entire architecture. Eventually, it has 2 FC (fully connected layers) trailed by a SoftMax for yield. The 16 in VGG19 alludes to it has 16 layers that have loads. This network is a pretty enormous network and it has around 138 million (approx.) parameters.

2.6 Relu:

The activation Function is a unit (neuron) is a fundamental piece of Artificial Neural Network architecture. The utilization of various algorithms has been utilized by analysts since ANN's initial days. the progression function was presented as the initiation function. Be that as it may, the binary idea of the progression function doesn't permit to have a decent blunder estimate. To beat the present circumstance,

sigmoid functions were used. They gave generally excellent execution to little networks.
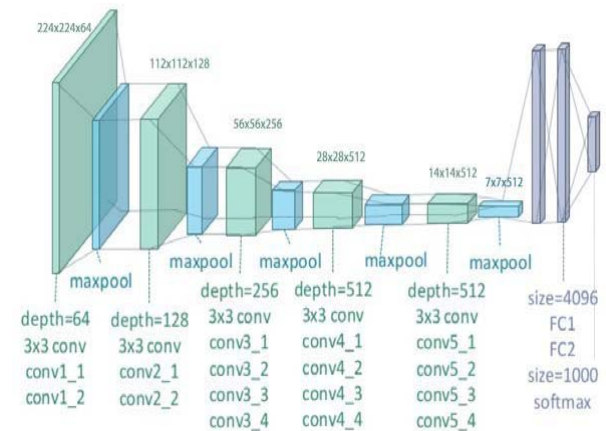

Fig. 2. VGG-19 Model

However, utilizing the sigmoid function demonstrated not very adaptable on enormous networks [11]. The computational expense of the outstanding activity may be truly costly since it can prompt long numbers. The other function instead of utilising the sigmoid function which is the inclination evaporating issue. This implies that the slope an incentive on the bend tails turns out to be too little that it forestalls learning. Under this situation, the corrected straight unit function (ReLU) if benefits contrasted with past regular enactment functions: its computational expense was less expensive, it gave a decent mistake guess and it didn't experience the ill effects of the angle disappearing issue. ReLU is shown on Figure-12, and it is characterized as
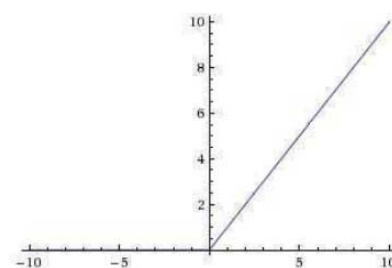

Figure 3: ReLU

$$f(x) = max(0, x)$$

The research done by Mr. Krizhevsky et al. [11] indicates that utilizing ReLU diminished the quantity of ages needed to meet when utilizing SGD divided by factor- 6. Be that as it may, a significant downside when utilizing ReLU is its delicacy when the info dispersion is under nothing. This happens when the

neuron arrives at a moment that it won't be activated by any datapoint again during training.

3 Algorithm:

Step 1: Extracting dataset into folder (In this scenario we preferred properly modified, (1, 48, 48) pixel grayscale images of different expression faces labelled and copied to 7 folders.

Step 2: Pre-processing of images.

Step 3: Detection of a face from each image.

Step 4: The cropped face is converted into grayscale images.

Step 5: The pipeline ensures every image can be fed into the input layer as a (1, 48, 48) NumPy array.

Step 6: The NumPy array gets passed into the Convolution 2D layer.

Step 7: For face detection, inbuilt methods available in dlib library are used. Once the face is detected, the region of interest and important facial features are extracted from it

Step 8: Convolution generates feature maps.

Step 9: Pooling method called MaxPooling2D that uses (2, 2) windows across the feature map only keeping the maximum pixel value.

Step10: During training, Neural network Forward propagation and backward propagation performed on the pixel values.

Step 11: The SoftMax function presents itself as a probability for each emotion class. The model is able to show the detail probability composition of the emotions in the face.

The dataset has a single emotions labelled for each image in a particular folder and doesn't have a connection to each folder. This presents a lot of upheaval if the image names are used as spotlights on setting up a VGG-19 on the solitary picture. Our visual features are in this manner given by a CNN arranged on a blend of two additional feeling datasets of static pictures. For the CNN preparing, we utilized an enormous feeling dataset, Cohn-Kanade Facial Expression Dataset. It comprises of in excess of 2900 pictures of 75 subjects and "Different guys and female Facial Expression Dataset containing 213 pictures, which have seven fundamental articulations: outrage, hatred, appall, dread, satisfaction, pity, and shock. For the pre-handling, we address fluctuating lighting conditions (explicitly, transversely over datasets) we associated histogram evening out. We used the changed appearances gave by the co-ordinators to eliminate features from the CNN. "The course of action incorporates a joined facial central issue's area and the accompanying system explained. Exceptional face area, just as plan methods, have been used for the COHN-KANADE Datasets". Remembering the ultimate objective to be prepared to utilize the extra datasets, we re-changed all datasets to COHN-KANADE using the going with strategy:

1. We recognized seven facial key concentrations for all photos in the COHN-KANADE and setting up a set using the convolutional neural framework course technique.

2. For each dataset, the mean shape has been handled by averaging the bearings of the principle centers.
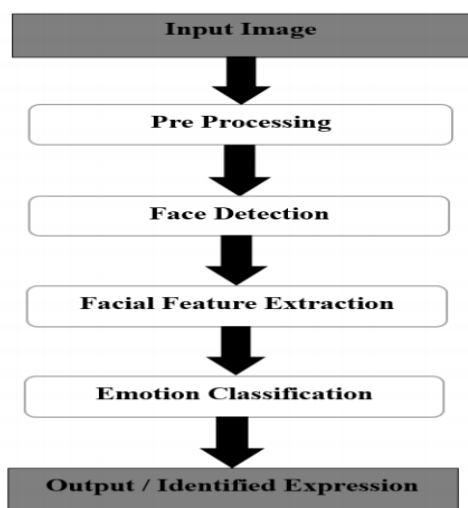

Fig. 4. Process flowchart

4 Proposed Model:

```
1  # Creating model with pre trained imagenet weights
2
3  vgg19 = VGG19(include_top=False, weights='imagenet', input_shape=(48,48,3))
4
5  # Model Summary
6
7  vgg19.summary()
```

Model: "vgg19"

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 48, 48, 3)]	0
block1_conv1 (Conv2D)	(None, 48, 48, 64)	1792
block1_conv2 (Conv2D)	(None, 48, 48, 64)	36928
block1_pool (MaxPooling2D)	(None, 24, 24, 64)	0
block2_conv1 (Conv2D)	(None, 24, 24, 128)	73856
block2_conv2 (Conv2D)	(None, 24, 24, 128)	147584
block2_pool (MaxPooling2D)	(None, 12, 12, 128)	0
block3_conv1 (Conv2D)	(None, 12, 12, 256)	295168
block3_conv2 (Conv2D)	(None, 12, 12, 256)	590080
block3_conv3 (Conv2D)	(None, 12, 12, 256)	590080
block3_conv4 (Conv2D)	(None, 12, 12, 256)	590080
block3_pool (MaxPooling2D)	(None, 6, 6, 256)	0
block4_conv1 (Conv2D)	(None, 6, 6, 512)	1180160
block4_conv2 (Conv2D)	(None, 6, 6, 512)	2359808
block4_conv3 (Conv2D)	(None, 6, 6, 512)	2359808
block4_conv4 (Conv2D)	(None, 6, 6, 512)	2359808
block4_pool (MaxPooling2D)	(None, 3, 3, 512)	0
block5_conv1 (Conv2D)	(None, 3, 3, 512)	2359808
block5_conv2 (Conv2D)	(None, 3, 3, 512)	2359808
block5_conv3 (Conv2D)	(None, 3, 3, 512)	2359808
block5_conv4 (Conv2D)	(None, 3, 3, 512)	2359808
block5_pool (MaxPooling2D)	(None, 1, 1, 512)	0

Total params: 20,024,384
Trainable params: 20,024,384
Non-trainable params: 0

Fig. 5. VGG19 Pre-trained model

3. The datasets are planned by employing a closeness modification among the mean shapes. By handling one modification for every dataset the nose, eyes, and mouth ar by and enormous in a very comparative region holding a small proportion of assortment. we tend to incorporated a loud outer boundary for and COHN-KANADE-faces as appearances were altered even additional immovably diverged from COHN-KANADE.

4. COHN-KANADE-data validation check sets were mapped to use for the amendment understood on the coaching set. what is more, dataset standardization has been performed by utilizing the quality deviation and mean image from the incorporate Cohn-Kanade addresses the example's face feeling data. For the execution and therefore the assessment of the planned model the seventieth of every dataset used for making ready and therefore the rest half-hour for testing.

After splitting the training data and testing data we took a pre-trained VGG19 CNN model along with ImageNet weights to train the data. The model summary which is shown in figure-14 below. After the model summary we have added the extra layers on top of the model to train the images. In this experiment the model is a pretrained one so if we train the whole model again, the model might overfit or underfit which makes the model very uncertain. So, instead of using the whole model to train again we are adding other

layers on top of the VGG19 model instead of disturbing it. So, we added the extra layers and taken all the parameters or neurons into a single vector in other words we call it as flatten layer. After taking it into a single vector we added the dense layer and dropout layer on top of it which is fully convoluted. We trained the dense and dropout layers using ReLU activation and for final layer we have used Softmax function because it is a Multi class classification. we can see the VGG19 model model summary in the below figure-15.

Model: "model"

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 48, 48, 3)]	0
block1_conv1 (Conv2D)	(None, 48, 48, 64)	1792
block1_conv2 (Conv2D)	(None, 48, 48, 64)	36928
block1_pool (MaxPooling2D)	(None, 24, 24, 64)	0
block2_conv1 (Conv2D)	(None, 24, 24, 128)	73856
block2_conv2 (Conv2D)	(None, 24, 24, 128)	147584
block2_pool (MaxPooling2D)	(None, 12, 12, 128)	0
block3_conv1 (Conv2D)	(None, 12, 12, 256)	295168
block3_conv2 (Conv2D)	(None, 12, 12, 256)	590080
block3_conv3 (Conv2D)	(None, 12, 12, 256)	590080
block3_conv4 (Conv2D)	(None, 12, 12, 256)	590080
block3_pool (MaxPooling2D)	(None, 6, 6, 256)	0
block4_conv1 (Conv2D)	(None, 6, 6, 512)	1180160
block4_conv2 (Conv2D)	(None, 6, 6, 512)	2359808
block4_conv3 (Conv2D)	(None, 6, 6, 512)	2359808
block4_conv4 (Conv2D)	(None, 6, 6, 512)	2359808
block4_pool (MaxPooling2D)	(None, 3, 3, 512)	0
block5_conv1 (Conv2D)	(None, 3, 3, 512)	2359808
block5_conv2 (Conv2D)	(None, 3, 3, 512)	2359808
block5_conv3 (Conv2D)	(None, 3, 3, 512)	2359808
block5_conv4 (Conv2D)	(None, 3, 3, 512)	2359808
block5_pool (MaxPooling2D)	(None, 1, 1, 512)	0
flatten (Flatten)	(None, 512)	0
dense (Dense)	(None, 1024)	525312
dropout (Dropout)	(None, 1024)	0
dense_1 (Dense)	(None, 512)	524800
dropout_1 (Dropout)	(None, 512)	0
dense_2 (Dense)	(None, 7)	3591

Total params: 21,078,087
Trainable params: 1,053,703
Non-trainable params: 20,024,384

Fig. 6. Model Summary of other layers along with pretrained model

To compile the model, we have used the ADAM optimizer which implements gradient descent algorithm which is an optimizer used for adjusting the weights of the model which checks also improves the accuracy by making it into optimum state. We have set the hyper parameter for loss function as categorical class entropy as it is a multi-class classification. Using model. Fit, we have initiated the training and the hyper-parameters are set as batch size as 64 and epochs as 20 to pass the whole data set it for twenty iterations. As we preferred to this experiment using the Keras model,

it does a cross validation to see how accurately the model is been trained. The trained model accuracy has shown in figure-16.

Fig. 7. Trained model.

In the above figure- 16 we have achieved a very good accuracy for facial detection which is 94%. As it is a static methodology utilizing removed highlights and emotion recognition utilizing profound learning is utilized in this work. Our execution is isolated into three sections. The initial segment is picture pre-handling and face recognition. For face discovery, inbuilt strategies accessible in dlib library are utilized. When the face is recognized, the district of interest and significant facial highlights are extricated from it. There are different highlights which can be utilized for feeling location.
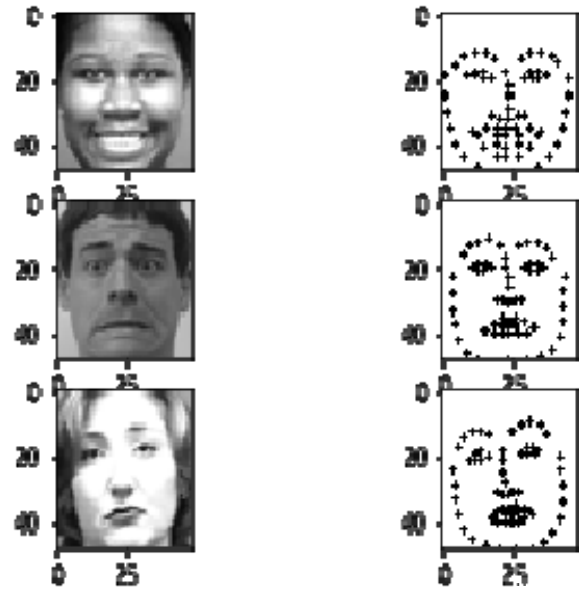
Fig. 8. Land marking images

In this work, the attention is on facial focuses around the eyes, mouth, eyebrows and so forth We have a multi-class order issue and not multi-mark. There is an inconspicuous distinction as a bunch of highlights can have a place with numerous names however just a single one-of-a-kind class. The removed facial highlights alongside VGG19 are utilized to distinguish the multi-class feelings once more. We can find out the VGG19 model which is shown in Figure-18 below which is trained on parsed or land marking images.

```
Epoch 1/20
11/11 [==============================] - 6s 569ms/step - loss: 1.8220 - accuracy: 0.2664 - val_loss: 1.6275 - val_accuracy: 0.4
660

Epoch 00001: val_accuracy did not improve from 0.93827
Epoch 2/20
11/11 [==============================] - 7s 616ms/step - loss: 1.5578 - accuracy: 0.4003 - val_loss: 1.4129 - val_accuracy: 0.4
753

Epoch 00002: val_accuracy did not improve from 0.93827
Epoch 3/20
11/11 [==============================] - 7s 681ms/step - loss: 1.3294 - accuracy: 0.5114 - val_loss: 1.2346 - val_accuracy: 0.5
340

Epoch 00003: val_accuracy did not improve from 0.93827
Epoch 4/20
11/11 [==============================] - 7s 634ms/step - loss: 1.2040 - accuracy: 0.5616 - val_loss: 1.1677 - val_accuracy: 0.5
309

Epoch 00004: val_accuracy did not improve from 0.93827
Epoch 5/20
11/11 [==============================] - 7s 633ms/step - loss: 1.1279 - accuracy: 0.5982 - val_loss: 1.0672 - val_accuracy: 0.5
957

Epoch 00005: val_accuracy did not improve from 0.93827
Epoch 6/20
11/11 [==============================] - 7s 626ms/step - loss: 1.0615 - accuracy: 0.6195 - val_loss: 1.0117 - val_accuracy: 0.6
451

Epoch 00006: val_accuracy did not improve from 0.93827
Epoch 7/20
11/11 [==============================] - 6s 582ms/step - loss: 1.0116 - accuracy: 0.6225 - val_loss: 1.0534 - val_accuracy: 0.5
802

Epoch 00007: val_accuracy did not improve from 0.93827
Epoch 8/20
11/11 [==============================] - 6s 593ms/step - loss: 0.9979 - accuracy: 0.6317 - val_loss: 0.9945 - val_accuracy: 0.5
741

Epoch 00008: val_accuracy did not improve from 0.93827
Epoch 9/20
11/11 [==============================] - 6s 583ms/step - loss: 0.9326 - accuracy: 0.6301 - val_loss: 0.9770 - val_accuracy: 0.6
019

Epoch 00009: val_accuracy did not improve from 0.93827
Epoch 10/20
11/11 [==============================] - 6s 600ms/step - loss: 0.8981 - accuracy: 0.6651 - val_loss: 0.8818 - val_accuracy: 0.6
420

Epoch 00010: val_accuracy did not improve from 0.93827
Epoch 00010: early stopping
```

Fig. 9. Trained model on parsed images

5 Results:

As a part of analysis for facial expression classification using face parsing method is presented in this thesis. We have proposed a new deep CNN architecture for facial expression identification. It takes facial images as the input and classifies them into either of the seven facial expressions: angry, disgust, happy, neutral, sad and surprise. We have tried this model achieved 94% of accuracy with the loss of 22% while detecting the faces and we have achieved 66.51% of model accuracy on parsed images. The comparison graph of accuracy is shown below.
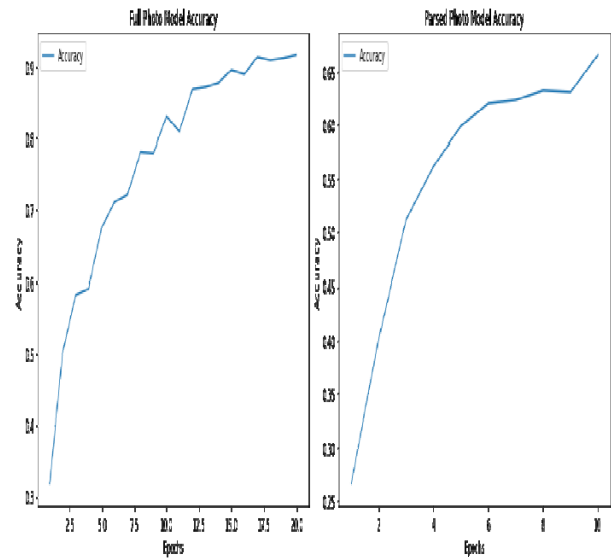

Fig. 10. Accuracy graph

7 Conclusion:

Results and acknowledgment rates demonstrate that our technique outperformed the cutting-edge strategies. For this task, we prepared the model with pictures in which the face was in a single position. In future work, we might want to stretch out our model to various face positions. This will permit us to examine the viability of pre-trained models like VGGNet for facial emotion Identification.

References

1. Y. Lv, Z. Feng, and C. Xu, 2014 International Conference on Smart Computing (2014)
2. W. Swinkels, L. Claesen, F. Xiao, and H. Shen, in *2017 IEEE Conference on Dependable and Secure Computing* (2017), pp. 86–92
3. A. C. L. Ngo, A. C. Le Ngo, Y.-H. Oh, R. C.-W. Phan, and J. See, 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2016)
4. K. M. Rajesh and M. Naveenkumar, in *2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT)* (2016), pp. 1–5
5. H. Ebine, Y. Shiga, M. Ikeda, and O. Nakamura, in *2000 Canadian Conference on Electrical and Computer Engineering. Conference Proceedings. Navigating to a New Era (Cat. No.00TH8492)* (2000), pp. 1091–1099 vol.2
6. Rosalind W. Picard, *Affective computing: challenges*, International Journal of Human-

Computer Studies,**Volume 59**, Issues 1–2, (2003) ,Pages 55-64,ISSN 1071-5819

7. R. W. Levenson, P. Ekman, and W. V. Friesen, Psychophysiology **27**, 363 (1990)

8. I. A. Essa, (1995)

9. Pentland, Alex & Benton, Stephen & Essa, Irfan. (1998)

10. J. P. Mueller and L. Massaron, *Machine Learning For Dummies* (John Wiley & Sons, 2021)

11. J. P. Mueller and L. Massaron, *Artificial Intelligence For Dummies* (John Wiley & Sons, 2018)

12. A. Krizhevsky, I. Sutskever, and G. E. Hinton, Commun. ACM **60**, 84 (2017)

13. S. C. Tai and K. C. Chung, TENCON 2007 - 2007 IEEE Region 10 Conference (2007)

14. M. Lyons, S. Akamatsu, and M. Kamachi, Proceedings Third IEEE (1998)

15. M. Suk and B. Prabhakaran, Proceedings of the IEEE Conference on (2014)

16. D. Ghimire and J. Lee, Sensors **13**, 7714 (2013)

17. T. R. Reddy, B. V. Vardhan, and P. V. Reddy, International Journal of Applied Engineering Research **11**, 3092 (2016)

18. P. Kumar, A. Singhal, S. Mehta, and A. Mittal, Journal of Real-Time Image Processing **11**, 93 (2016)

19. G. Mahalle, O. Salunke, N. Kotkunde, A. K. Gupta, and S. K. Singh, Journal of Materials Research and Technology **8**, 2130 (2019)

20. S. Kumar, P. Reddy, G. Ramesh, and V. Maddumala, Trait. Du Signal **36**, 233 (2019)

21. B. Dhanalaxmi, G. A. Naidu, and K. Anuradha, Procedia Comput. Sci. **46**, 432 (2015)

22. M. Avanthi and P. Chandra Sekhar Reddy, in *Smart Computing Techniques and Applications* (Springer Singapore, 2021), pp. 193–199

23. P. Chandra Sekhar Reddy, P. Vara Prasad Rao, P. Kiran Kumar Reddy, and M. Sridhar, in *Soft Computing and Signal Processing* (Springer Singapore, 2019), pp. 273–280

24. P. C. S. Reddy, S. G. Rao, G. R. Sakthidharan, and P. V. Rao, in *2018 International Conference on Smart Systems and Inventive Technology (ICSSIT)* (2018), pp. 104–109

25. P. Chandra Sekhar Reddy, P. Vara Prasad Rao, P. Kiran Kumar Reddy, and M. Sridhar, in *Soft Computing and Signal Processing* (Springer Singapore, 2019), pp. 273–280