REPORT

TOPIC : Automatic Facial Emotion Recognition using CNN

GUIDED BY :

Dr. Priyanka Singh

Dr. Deepak Kachave

Submitted by:

Mohan Thota AP18110010029

Nikhil Gadde AP18110010019

Sushma Nalluri AP18110010008

Darbha Santosh AP18110010020

Lalitha Gadde AP18110010015

AUTOMATIC FACIAL EMOTION DETECTION USING CNN

I. INTRODUCTION

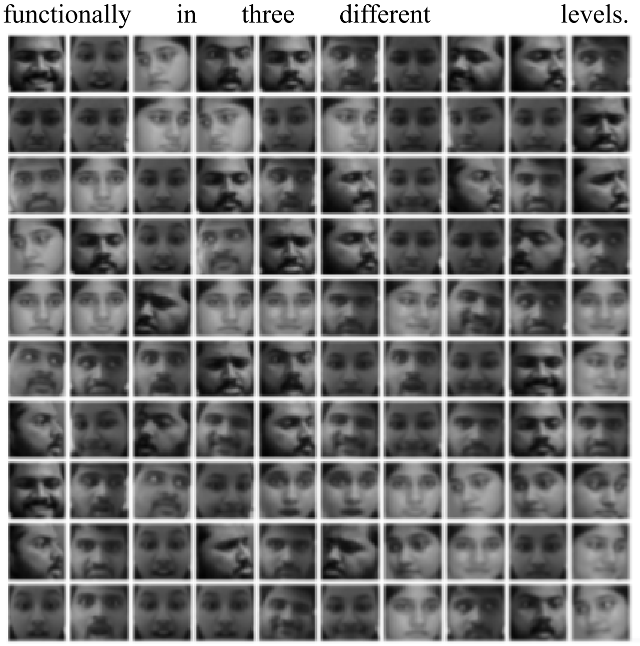
Emotions are considered as an important element of human nature and have been widely studied in psychology and behavioral sciences. An emotion represents the psychological state of a person, which is normally based on internal factors such as mental and physical status of a person. Emotions and related fluctuations in the facial muscles are together known as facial expressions [1]. It supports to judge the existing state of emotion and mood [2] of a person and enables us to make conversation with the other person based on their mood. Six basic categories of emotions such as happy, sad, surprise, anger, fear, and disgust are identified and are referred to as standard facial expressions. Facial expression recognition and the emotional computing system plays a vital role in the area of Human–Computer Interaction. Automatic Facial image recognition has attracted a lot of research interests in recent times. It is used in many real-time applications like gaming applications, criminal interrogations, psychiatry, animations, etc. It is observed that 55% of human communication is through facial expressions[3]. Researchers have presented various techniques for facial emotion recognition with considerable accuracy, however, it is still a challenging task to recognize facial expression for faces captured from different angles and of different nationalities [4]. To recognize and classify the tasks like human emotions, deep learning technique is utilized as it outperforms other methods by its large capabilities of different datasets and fast computation capabilities. Deep learning technique is a standard paradigm to represent the working of the human brain with neurons [5]. This learning usually consists of neural network models where neurons act as inputs and each of them are connected to move as outputs.

II. LITERATURE REVIEW

Automatic facial expression recognition is an interesting and challenging problem which has important applications in many areas like human-computer interaction. Human beings can read and understand the continuous changes of facial reactions,but it is understood that estimating expressions from the human face is a challenging task for a machine.In order to obtain better representation of facial expressions several deep learning techniques, such as Convolutional Neural Networks (CNN) have been developed. Mehendale(2020) proposed a two level CNN framework[6]. Here, the conventional network module is used to extract the primary expression vector. Each convolutional layer consists of four filters, making the output accurate. It works with different orientations due to the unique EV feature matrix. Additional to this, the researcher added a background removal layer in the initial steps making it more accurate to predict the emotion.In 2015, Yu and Zhang used CNN for Facial Emotion Recognition in EmotiW. They used an ensemble of CNNs [7] and randomly perturbed the input image to achieve a 2-3% boost in accuracy. Chuang (2006) have focused on measuring changes in Facial Action Units (FAUs) to capture change in facial expressions. Independent Component Analysis (ICA) has been used as a feature extraction and representation scheme and Support Vector Machine (SVM) has been used as the pattern classifier. The FAUs, i.e., the subtle change of facial expressions are recognized to deal with the comprehensive and heterogeneous subject database [8].Haq et al. (2019) have proposed an algorithm that can handle face pose, non-uniform illumi- nations, and low-resolution factors. The approach uses sixty-eight points[19] to locate the face and use the PCA to extract mean image. The AdaBoost and the LDA methods are used to extract facial features.A softmax classifier with the largest pooling layer is developed by Lei Xu(2019) .With the method having good recognition performance and generalization ability[9],seven types of facial expressions are recognized. Dennis H. et. all in his work[r10], applied a face image to two channels of CNN, the information from the two networks was combined to generate a 94.4% recognition accuracy. A CNN network that uses an ensemble of the outputs of three CNNs for classification is analyzed in [11]. CNNs use the face, mouth and eyes as input. The method when compared to other methods achieved a high recognition rate. Bartlett et al. (2006) have proposed an approach to measure the intensity of AU posed and spontaneous facial expressions [12]. This approach has used distances to the SVM separating hyperplanes as a direct measure of the intensity levels of AUs.Shan et al. (2005) have proposed an approach to obtain complete and concise descriptive facial images by shifting sub-windows on input images[15]. Lin (2006) has used PCA for extracting features from images and used Hierarchical Radial Basis Network (HRBFN) function [16]for classifying facial expressions. The PCA has segmented the facial data into small components to complement the feature extraction procedure.Spherical clustering fused with Online Sequential Extreme Learning Machine (OSELM)[17] has been used for facial expression classification by Ucar et al. (2016). This approach has used both local curvelet transform and sta- tistical features.Yun et al. (2013) have proposed an approach named parametric Kernel Eigenspace Method based on Class (PKEMC) feature[18]. The approach is based on a combination of two methods such as the Parametric Eigenspace Method based on Class feature (PEMC) and Kernel Eigenspace Method based on Class features (KEMC). The approach has used the nearest neighbor classifier along with Eucli- dean distance for classifying facial expressions.Based on the review, it is found that most of the approaches have used various features to estimate emotion from an image . In this paper, a simple CNN architecture to predict basic expressions from video is proposed. This approach uses MobileNetV2 network to train and test the data[13].

III. PROPOSED WORK

Convolution Neural Network (CNN) is the most popular way to classify the contents of different images .The use of which has been literary explosive in the area of visual computing. CNNs are supervised machine learning techniques that can extract deep knowledge from a dataset through rigorous example based training. All the feeded images will be trained by the model. The proposed model is based on the CNN framework used in classifying emotions of a person. The system design model is tested functionally in three different levels.



III.a) Creating a dataset to classify emotions using convolutional neural network:

In this phase , the camera reads the image and captures it based on the open Cv haar cascade and face net algorithm. Facenet helps in extracting features from an image of a person's face.So the biometric facial parts of the person are captured.Hence by facial features the person's face is cropped.The image is converted in to gray scale in order to avoid misclassification with the background and nearby objects.The emotions given in our dataset are happy, sad, neutral, surprise.

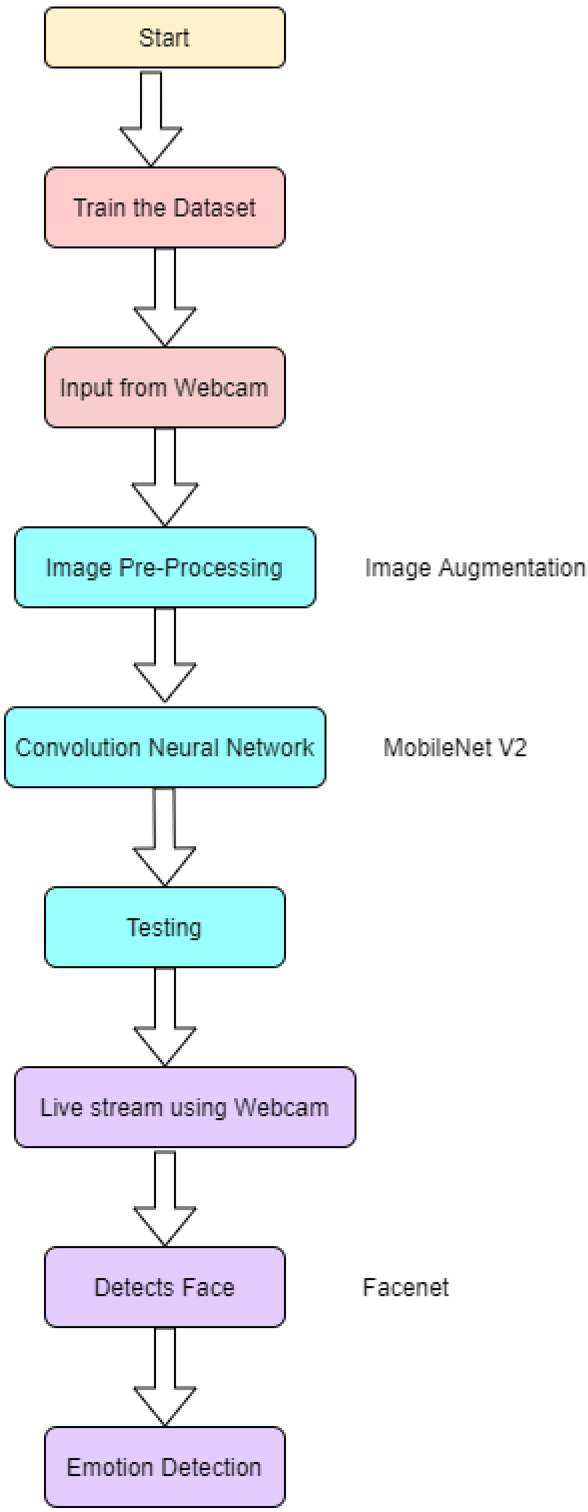
III.b) Training and classifying the dataset

MobileNet_v2 which is imported from keras, is a pre-trained and pre-processed model with over a million images,is used to train and test the input data given by the user. The size of the image used is 224 x 224 which is pre-processed as re-scaling and extracted as a feature vector to train the classifier. Image Augmentation is used to create more images in the training dataset. The images are then converted into numpy arrays. A base Model is created using the mobileNet_v2 network, considering weights from imageNet model/database. The head model is constructed from the baseModel. A final model consisting of baseModel and headModel is generated. The model is trained by passing the training dataset 50 times forward and backward through the network.Finally, the trained model is tested with testing data and all the classification metrics are computed and a plot between loss/accuracy and number of epochs is generated.

III.c) Testing the trained model:

- To the test, a real time input from the video stream is given to the trained model. The trained model therefore uses a pre-trained model called FaceMask which detects the face in the video stream. The result of the classification is displayed on the video frame itself. The extraction of features may be distorted by variance of lighting condition in image.The intraclass noise of lights distort the model in classifying emotions.So light also plays a key factor in recognizing image in classifying emotions[14].

IV. BLOCK DIAGRAM



V. RESULTS

For performing experiments, the classifier was tested with two approaches, one approach contains a mixture of images with 4 subjects having images of four different expressions. and the other approach contains the images of the executor only.

Figure (1) shows the sample images that are present in the dataset and processed by the system. As discussed earlier in *Sect.3.2* mobilenetV2 network is used to train and classify the dataset.When approach one was applied there were multiple scenarios where the actual emotion label of an image was wrongly classified by the classifier.one of such scenarios occurred when the model is given to classify five emotions.

	precision	recall	f1-score	support
happy	0.96	1.00	0.98	26
neutral	0.95	0.84	0.89	25
sad	0.87	0.93	0.90	28
surprise	1.00	1.00	1.00	22
accuracy			0.94	101
macro avg	0.95	0.94	0.94	101
weighted avg	0.94	0.94	0.94	101

[INFO] saving mask detector model...

Figure 2(a): classification metrics for scenario one

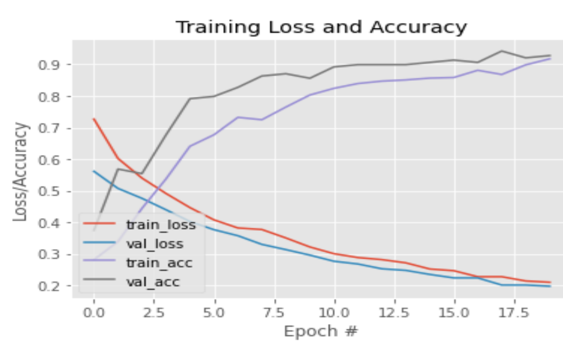


Figure 2(b): plot for scenario one

Based on the output of the experimental results, we restricted the model to classify only four emotions (happy,neutral,sad,surprise).The estimated accuracy of the model is 94%. *Figure 2(a)* shows the classification metrics of scenario one and *figure 2(b)* shows the plot obtained .In approach two,where the training datasets consist of only images of the executor came out with an accuracy of 99% ,which can be the result of less noisy data.

```
              precision    recall  f1-score   support

       happy       1.00      1.00      1.00        24
     neutral       0.96      1.00      0.98        24
         sad       1.00      0.96      0.98        24
    surprise       1.00      1.00      1.00        23

    accuracy                           0.99        95
   macro avg       0.99      0.99      0.99        95
weighted avg       0.99      0.99      0.99        95

[INFO] saving mask detector model...
Training Completed
```

Figure 3(a): classification metrics for scenario two(only executor images)

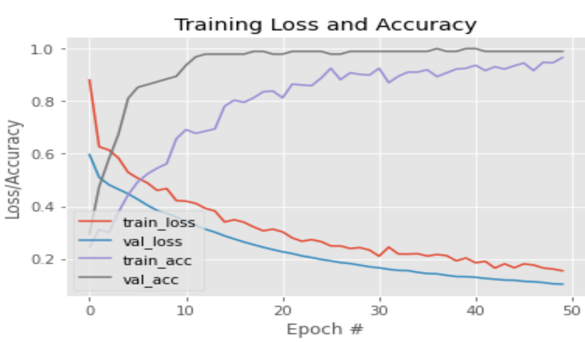


Figure 3(b): plot for scenario two(only executor images)

As mentioned previously, most of the proposed works deal with an image rather from a video frame as the data is continuous in a video it might sometimes be a little tricky to get the required information . This model,not only being a simple technique but can efficiently classify an emotion of a person with less dataset also .

VI. CONCLUSION

Facial emotions might vary from person to person slightly as each individual is structured differently. In this paper we developed a model which detects and classifies four different emotions of a person. For better analysis and interpretation of facial emotions, classification metrics are measured for each training. Two different kinds of self created datasets were used to evaluate this model. The accuracy of the system is around 91+%. The model can be integrated into systems like home automation devices, self driving technology, and in the medical field.

VII. REFERENCES

[1] Ekman, Paul, and Harriet Oster. "Facial expressions of emotion." Annual review of psychology 30.1 (1979): 527-554.

[2] P Forgas, Joseph., and H Gordon. Bower. "Mood effects on person-perception judgments." Journal of personality and social psychology 53.1 (1987): 53.

[3] A. Mehrabian, “Communication without words,” Psychol. Today,vol. 2, no. 4, pp. 53–56, 1968.

[4] Lv Y, Feng Z, Xu C (2014) Facial expression recognition via deep learning. In: Smart Computing (SMARTCOMP), 2014 International Conference on. IEEE[5] Liu, F L C Y, 2015. Improving Steganalysis by Fusing SVM Classifiers for JPEG Images. IEEE, pp. 185-190.

[6] Mehendale, N. (2020, February 18). *Facial emotion* recognition using convolutional neural networks (FERC). SN Applied Sciences.

[7] Z. Yu and C. Zhang. Image based static facial expression recognition with multiple deep network learning. ICMI Proceedings

[8] Kanade, T., Cohn, J. F., & Tian, Y. (2000). Comprehensive database for facial expression analysis. Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'00), Grenoble, France, 46-53.

[9] Lei Xu, Minrui Fei, Wenju Zhou, Aolei Yang (2019) *Face Expression Recognition Based on Convolutional Neural Network,*

[10] Dennis Hamester et al., “Face Expression Recognition with a 2- Channel Convolutional Neural Network”, International Joint Conference on Neural Networks (IJCNN), 2015.

[11] C. Ruoxuan, L. Minyi, and L. Manhua "Facial Expression Recognition Based on Ensemble of Multiple CNNs," CCBR 2016, LNCS 9967, pp. 511-578, Springer International Publishing AG 2016

[12] M. Bartlett, G. Littlewort-Ford, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, "Fully automatic facial action recognition in spontaneous behaviour," in Proc. IEEE Int. Conf. Automatic Face Gesture Recognit., 2006, pp. 223–230.

[13] Sandler, M., & Howard, A. (2018, April 03). MobileNetV2: The Next Generation of On-Device Computer Vision Networks.

[14] Lanqing Zhao, Chieh-En Li(2019)"The Application of CNN on Emotion Recognition".

[15] Shan C, Gong S, Mc Owan PW (2005) Conditional mutual information based boosting for facial expression recognition. In: Clocksin WF, Fitzgibbon AW, Torr PHS (eds), Proceedings of the British machine conference (BMVC), pp 1–10.

[16] Lin DT (2006) Facial expression classification using PCA and hierarchical radial basis function network. J Inf Sci Eng 22:1033–1046

[17]Ucar A, Demir Y, Guzelis C (2016) A new facial expression recognition based on curvelet transform and online sequential extreme learning machine initialized with spherical clustering. Neural Comput Appl 27:131–142

[18] Yun W, Kim D, Park C, Kim J (2013) Multi-view facial expression recognition using parametric kernel Eigenspace method based on class features. In: IEEE conference on systems, man and cybernetics, Manchester, pp 2689–2693

[19] Haq MU, Shahzad A, Mahmood Z, Shah AA (2019) Boosting the face recognition performance of ensemble based LDA for Pose, non- uniform illuminations, and low-resolution images. KSII Trans Internet Inf Syst 13(6):3144–3166