Facial Emotion Recognition using Convolution Neural Network

Shrey Modi
Student of Information Technology,
Devang Patel Institute of Advance Technology and Research(DEPSTAR),
Charotar University of Science and Technology (CHARUSAT), CHARUSAT campus, Changa 388421, India.
18dit037@charusat.edu.in

Mohammed Husain Bohara
Assistant Professor at Computer Engineering Dept.,
Devang Patel Institute of Advance Technology and Research(DEPSTAR),
Charotar University of Science and Technology (CHARUSAT), CHARUSAT campus, Changa 388421, India.
mohammedbohara.ce@charusat.ac.in

***Abstract*—**

With the transition from laboratory-controlled to daunting in-the-wild conditions of facial Emotion recognition (FER) and the recent popularity of deep learning strategies in various fields, deep neural networks[1,2] have rapidly been leveraged to train discriminatory representations for automated FER. The FER will allow us to recognize the Emotion of the human face that is a major blow to recent technological development. Recent FER programmers are typically concentrating on two critical issues: overfitting due to lack of appropriate training evidence and emotion-related differences such as lighting, head posture and identification bias. It encourages to improve the other innovations, such as incorporating FER into the robotic device in order to provide the robot feelings. A systematic survey of deep FER including datasets and algorithms that offer insight into these inherent issues. First, the FER scheme with the relevant context information and recommendations for the implementations to be applied at each level. Introducing the relevant datasets that are commonly used in the literature and including agreed data collection and assessment criteria for these datasets. Competitive results on commonly used metrics that can be used to develop the project to the maximum degree that, in essence, helps more developments and the environment. The main motivation to work is to improve the accuracy for the following model statement which could make an impact on the future work. Along with that a comparative analysis between the used model and transfer learning model will provide a proper novelty for the research work.

Keywords— Facial emotion recognition, convolutional neural network, classification, emotion detection, data augmentation, model accuracy.

I. INTRODUCTION

Facial Emotion Recognition are software tools and techniques providing facial Emotion of the user. The model given is intended to help the communities.

The facial expression is a body language means of expressing feeling, and it can be used as solid proof to determine whether or not someone is telling the truth [1]. In the different Recognition-Processing, such as robotic community can use to give emotions to the robot, blind community to know what is the emotion of the person he is talking to etc Recognition systems have proved to be the required method for users to cope with the flowering of knowledge and have become the most effective and common methods. We use machine learning to build a Facial Emotion Recognition which helps us to know the mood of the various users. Different people have different moods, and this is not reflected when someone cannot see or if someone is facing problems to judge. FER helps users instantly discover their mood and behavior and also their Emotions. In these systems if the user changes mood frequently it will be depicted. The project is to study recognition engines. Action units are used as speech identifiers in the facial action coding scheme (FACS) [2]. Facial muscle modifications were used to distinguish these AUs. In this paper, we introduce a facial expression recognition framework based on Convolutional Neural Networks (CNN). An picture is fed into our framework, and then CNN is used to identify the facial expression mark, which should be one of three options: happy, sad, or neutral.

II. LITERATURE OVERVIEW

Human Emotion can accurately be divided in 7 simple emotions. (happy, anger, sad, fear, surprise, disgust, and neutral). Human facial emotions are conveyed by the manipulation of complex facial muscles. This often delicate and nuanced signs of speech also provide a plentiful of knowledge about the state in which our mind is. Via the detection of facial emotions, we are in a position to calculate the impact of that material and facilities that are made available to the audience/users by a simple and minimum cost, process. Retailers can, for example, use these measures to assess consumer interest.

Many papers using deep learning for facial expression processing have been published in recent years [3], [5]. Image net classification is the first paper, we come across and refer it [4]. Researcher use image net classifier initially in the experiment and do image classification. In the same context, static facial image based classification is done by multiple deep neural network [5]. Which help us to understand the approach in detail. The results of the paper are quite surprising but complex network consume more time. Facial feature identification and quality of feature set was understood by A.

Yao [6]. Possible patterns can be formed on facial expression, study by G. Levi [7].
Health care services can offer improved support by using extra means of input on the mental status of patients during treatment. Entertainment creators should track the participation of the audience in activities in order to reliably generate the content that is wanted. Human beings are well qualified to read the feelings of others, a baby who has just entered the world can tell that you are happy or sad. How can machines do a great job at gaining emotional [8], [9] conditions than we do? But that's why we built a deep neural network that gives machines the opportunity to draw conclusions concerning our mental states. So, we're giving them a powerful tool/organ known as eyes so they are also able to see through it [10],[11]. Using text mining and geographic information systems [12], geospatial and social media analytics is used to analyze the emotions of theme park travelers.
So with help of Facial Emotion recognition we will be able to give the machines the power to see and recognize the feelings which can be a big achievement in the coming future as our world is getting used to Artificial Intelligence [13] and this would be a part of it.

III. PROPOSED METHOD

For this study a dataset of static images are considered and the movement of the proposed model is observed below.

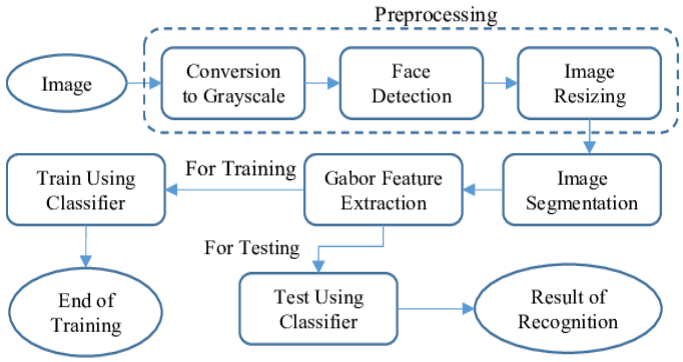

Figure 1 flow of the proposed system

The flowchart explains the entire process in an excellent manner. Firstly the image is taken from the dataset after that the data preprocessing is done which includes the conversion into the grayscale from RGB, then the face detection and after that image resizing is a very essential part of data preprocessing. After that process is finished the flow goes with the image segmentation and Feature extraction , now after the features are extracted from the image some data is sent to the training and some data is sent to the testing where Train/Test is used using the classifier and after that we can get the final results of the recognition.

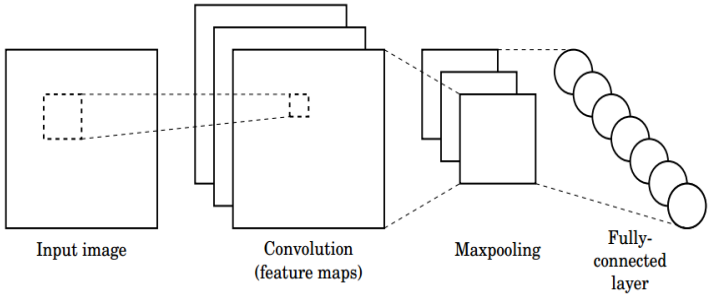

Figure 2 Model Architecture

In this figure the entire Architecture of the model is being displayed firstly the input image after the data preprocessing is shown to the model and then the convolution layers do the feature mapping of the given image which maps the similar features of the image through various layers, after that max pooling is done which reduces the image size assuming the features would be same, and finally we get a fully connected layer which would be helpful in the emotion detection.

IV. METHODOLOGY

The process through which the accuracy has been obtained is by using the dataset that has been described with an aim through which the approach move towards selecting the data, processing the data, training the data to selecting the appropriate model, which would give a good accuracy and remain suitable to the problem and extract the features from the given image in order to make a list to help the model, using activation and generalization for the model to properly respond and finally training the half dataset and the other half of the dataset for the testing purpose.

1) Data Preprocessing

The first thing to start with is to get the dataset for the model, you need to select an appropriate dataset which would be useful for the model to give a proper accuracy which can be also used in further different aspects. In order to perform data preprocessing using Python, we need to import some predefined Python libraries. These libraries are used to perform some specific jobs. Now the dataset selected should be imported into the model. After that we need to handle the missing data into the datasets If our dataset contains some missing data, then it may create a huge problem for our machine learning model. Hence it is necessary to handle missing values present in the dataset. This can be done by two ways -: **By deleting the particular row** or **by calculating the mean.**
Our data consists of 48*48 pixels gray - scale face images. They are the faces of different people and their different feelings. The dataset is split into two "emotion" and "pixel" tables. The "emotion" column includes number code that ranges from 0 up to 6 (included) for the emotion that is there in the picture. Then we load the csv pixels from the file (dataset) in the data frame. Then we apply pre-processing

procedures like resizing, reshaping, converting to grayscale and standardize. Then use vectorization by transforming images to NumPy arrays and panda data frames.

2) Splitting Dataset

The dataset will be divided into training and validation collection in order to verify if the model we have taken is overfitted to the training dataset or not using the validation dataset. This helps us to train as well as test the same database so that we can know the proper accuracy of the model and also figure it out where it is going wrong and can easily correct it.

3) Building the model using CNN

A **Convolutional Neural Network (ConvNet/CNN)** is a class of deep neural networks, most commonly applied to analyzing visual imagery. They are also known as shift invariant or space invariant artificial neural networks, based on the shared-weight architecture of the convolution kernels that scan the hidden layers and translation in variance characteristics.
With most algorithms that handle image processing, the filters are typically created by an engineer based on heuristics. CNNs can learn what characteristics in the filters are the most important. That saves a lot of time and trial and error work since we don't need as many parameters.
A big difference between a CNN and a regular neural network is that CNNs use convolutions to handle the math behind the scenes. A convolution is used instead of matrix multiplication in at least one layer of the CNN. Convolutions take to two functions and return a function.

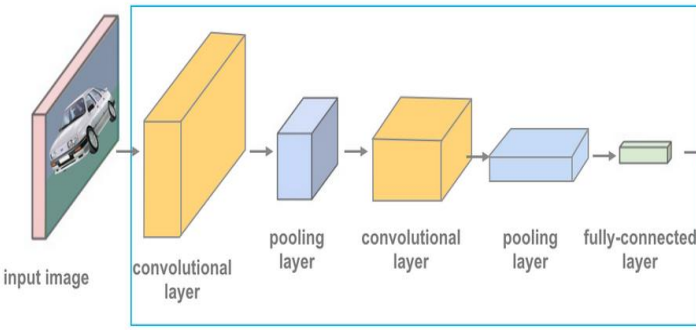

Figure 3 CNN Structure

In the above image how the CNN structure works is displayed. So firstly the input image is sent into the CNN as the preprocessed data , so now the image passes through different layers . convolution is the first layer in which the features are extracted from the input image , with the help of the mathematical operations it takes the image filter and a kernel as inputs from the image. After that the pooling layer will help in reducing the number of parameters when the image is too large so for that we can use max pooling , average pooling or sum pooling . In this model max pooling has been used. The image goes again for the convolution layer in which features are extracted using different techniques ReLU, Leaky ReLU, LeakyReLU is utilized for the proposed model. The last step of the CNN architecture/ structure is the Fully Connected Layer(FC) in which we flattened our matrix into vector and feed it into a fully connected layer like a neural network.

Using TensorFlow for building the back end of the neural network

4) Feature Extraction

In machine learning, image processing, pattern recognition, feature extraction begins with a set of measured data and generates derived values that are informative and non-redundant ,allowing for future learning .The chosen features are to include the appropriate info from the input data, so that the requested task can be accomplished with this reduced exemplification.

For this we use convolution which does matrix multiplication with filter and then proceeds towards the feature detector.

In the feature detector we use a sliding window which each channel slides over and summaries the features.

Normalizing results with the help of batch normalization, as neural networks are very sensitive to unnormalized data.

5) Generalization and Activation

Generalization refers to the ability of the model to better respond to current, previously unknown data from the same distribution as those used to construct the model. That's why we used dropouts in daily interviews. Activation mechanisms are a key aspect of the architecture of the neural network. Choosing the activation feature in the hidden layer will monitor how much the training dataset learns from the network model. Choosing the activation function in the output layer will determine the form of predictions the model will produce. We used 'ELU' as an activation because it prevents the issue of dying ReLU but has also done well relative to LeakyRelu.

6) Model Training and Evaluating

Education is nothing more than a learning cycle. In this section, we describe hyper parameters such as the number of epochs, batch size, learning rate, and so on.I used two callbacks one is `early stopping` for avoiding overfitting training data and other `ReduceLROnPlateau` for learning rate.

7) Testing the Model

So As we used train test split which basically splits the dataset into two different ways one for training the dataset and other for testing the remaining dataset. So now we will test the testing dataset and acquire the results

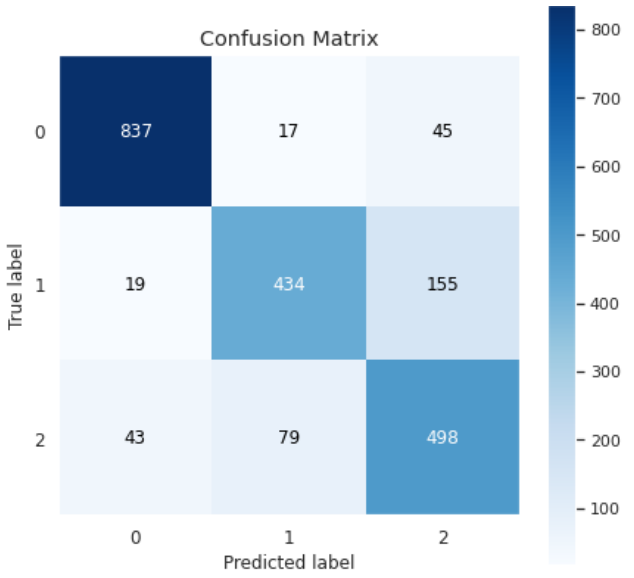

Figure 4 Confusion Matrix

The confusion matrix helps us to find if the model is doing a good job on the classes we have and also tells where the data is lacking and where it is going wrong.

v. Implementation and Results

The implemented model results are-:

Figure 5

Figure 6

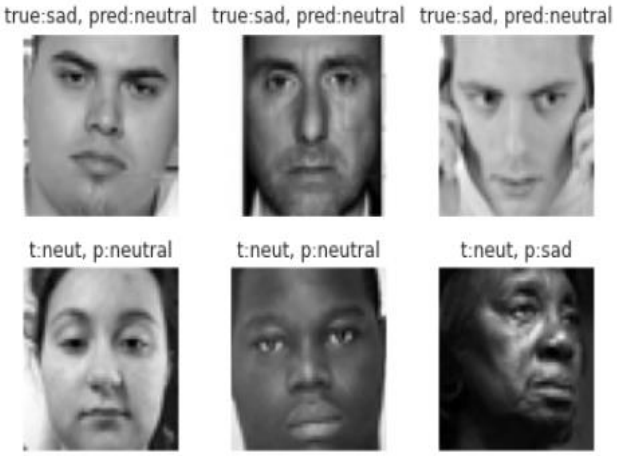

Figure 7

In figure 6 the first image shows a neutral face and our model predicted as neutral one which describes that the model build is well built and is able to make correct predictions

The last image in the second row shows that it is sad and our model predicted that too correctly which again confirmed it is on the right direction towards the prediction.

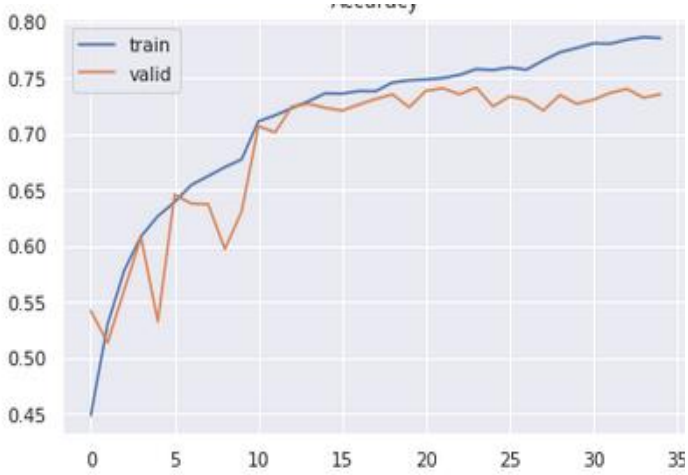

Figure 8 Accuracy with Transfer learning Approach

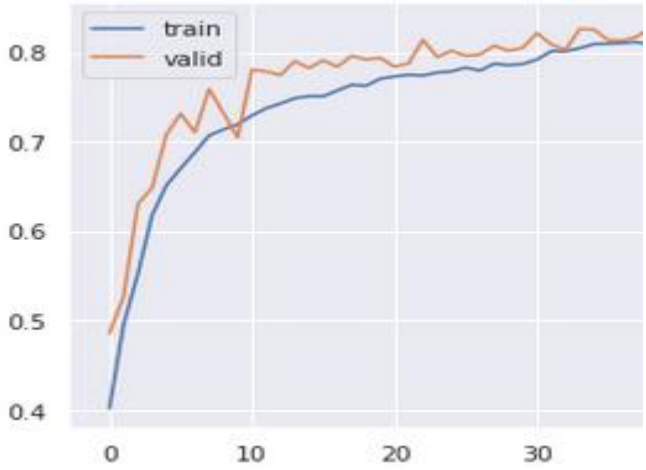

Figure 9 Accuracy with CNN Approach

As a comparative analysis we took our CNN model against the Transfer Learning model, Here, the accuracy comparison is performed between the models for that we have displayed (Figure 8 and Figure 9) the Epoch-Accuracy chart shows the accuracy of a particular epoch. From this we can see that the accuracy of the CNN model is better compared to the transfer learning model.

Table 1 Comparison accuracy of CNN and Transfer Learning

Epoch	CNN method Accuracy	Transfer Learning method Accuracy
1	0.4861	0.5424
5	0.7047	0.5327
10	0.7320	0.6309
15	0.7823	0.7236
20	0.7936	0.7242
25	0.8016	0.7248
30	0.8049	0.7273
35	0.8251	0.7358

In the above Table 1, a detailed accuracy comparison has been performed and the accuracy is shown at different epochs so from here the comparison is concluded that, the CNN model has a better accuracy than the transfer learning model after 35 epochs.

vi. Conclusion

This technology will provide a great boom to many things such as the robotics field, which will provide emotions to them and then to the blind community, where AR technology can be developed to show the real time emotions of a person. It allows individuals with Autism Spectrum Disorder (ASD) to enter PCs by processing. First, the network is described and trained to be capable of classifying the right emotion, after which the trained model is used to identify the emotions in real-time. The model could be utilized for various real-life implementations for successful use in sectors such as health, video game and marketing industry.

vii. Future Work

In this paper we studied about 3 emotions (Happy, sad and Neutral). As a future enhancement we can include other emotions such as anger, fear, and disgust. Further we can use OpenCV and test the model in real time and not on the datasets and also a mobile application can be developed for the same. The model is accurate, but it takes a long time to practice that can be consider as a limitation of our model. We will use other pre-trained architectures like ResNet and MobileNet to prepare the proposed network for comparison with the model in the future.

REFERENCES

[1] Bartlett, Marian, et al. "Data mining spontaneous facial behavior with automatic expression coding." Verbal and Nonverbal Features of Human-Human and Human-Machine Interaction. Springer, Berlin, Heidelberg, 2008. 1-20.

[2] Ekman, Rosenberg. What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS). Oxford University Press, USA, 1997.

[3] Raghuvanshi, Arushi, and Vivek Choksi. "Facial expression recognition with convolutional neural networks." CS231n Course Projects 362 (2016).

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," NIPS, vol. 1, p. 4, 2012.

[5] Z. Yu and C. Zhang, "Image based static facial expression recognition with multiple deep network learning," in Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, ICMI '15, (New York, NY, USA), pp. 435–442, ACM, 2015.

[6] A. Yao, J. Shao, N. Ma, and Y. Chen, "Capturing au-aware
facial features and their latent relations for emotion recognition in the wild," in Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, ICMI '15, (New York, NY, USA), pp. 451–458, ACM, 2015.

[7] G. Levi and T. Hassner, "Emotion recognition in the wild via convolutional neural networks and mapped binary patterns," in Proc. ACM International Conference on Multimodal Interaction (ICMI), November 2015.

[8] B. Kim, J. Roh, S. Dong, and S. Lee, "Hierarchical committee of deep convolutional neural networks for robust facial expression recognition," Journal on Multimodal User Interfaces, pp. 1–17, 2016.

[9] Fan, Yin, et al. "Video-based emotion recognition using CNN-RNN and C3D hybrid networks." Proceedings of the 18th ACM International Conference on Multimodal Interaction. 2016.

[10] Duncan, Dan, Gautam Shine, and Chris English. "Facial emotion recognition in real time." semantic scholar, Computer Science (2016).

[11] Mohammadpour, Mostafa, et al. "Facial emotion recognition using deep convolutional networks." 2017 IEEE 4th international conference on knowledge-based engineering and innovation (KBEI). IEEE, 2017.

[12] Manoharan, Samuel. "Geospatial and Social Media Analytics for Emotion Analysis of Theme Park Visitors using Text Mining and GIS." Journal of Information Technology 2, no. 02 (2020): 100-107.

[13] Tümen, Vedat, Ömer Faruk Söylemez, and Burhan Ergen. "Facial emotion recognition on a dataset using convolutional neural network." 2017 International Artificial Intelligence and Data Processing Symposium (IDAP). IEEE, 2017.

[14] A. Alkuhlani, M. Nassef and I. Farag, Multistage feature selection approach for high-dimensional cancer data, Soft Computing 21(22) (2017), 6895–6906. doi:10.1007/s00500-016-2439-9.

[15] Ko, Byoung Chul. "A brief review of facial emotion recognition based on visual information." sensors 18.2 (2018): 401.

[16] Caffe Framework: Github: https://github.com/BVLC/caffe, last visited on Sept 12, 2018.

[17] K. Bouaziz, T Ramakrishnan, S. Raghavan, K. Grove, A.A.Omari, C Lakshminarayan, " Character Recognition by Deep Learning: An Enterprise solution.", 2018 IEEE Conference on Big Data.

[18] Gervasi, Osvaldo, et al. "Automating facial emotion recognition." Web Intelligence. Vol. 17. No. 1. IOS Press, 2019.

[19] G. Cao, Y. Ma, X. Meng, Y. Gao and M. Meng, "Emotion Recognition Based On CNN," 2019 Chinese Control Conference (CCC), Guangzhou, China, 2019, pp. 8627-8630.doi: 10.23919/ChiCC.2019.8866540.

[20] Lu Lingling liu, "Human Face Expression Recognition Based on Deep Learning-Deep Convolutional Neural Network", 2019 International Conference on Smart Grid and Electrical Automation (ICSGEA)

[21] S. Suresh, H. T. P Mithun and M. H. Supriya, "Sign Language Recognition System Using Deep Neural Network," 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), Coimbatore, India, 2019, pp. 614-618. doi:10.1109/ICACCS.2019.8728411.

[22] Pranav, E., et al. "Facial emotion recognition using deep convolutional neural network." 2020 6th International conference on advanced computing and communication Systems (ICACCS). IEEE, 2020.

[23] Mehendale, Ninad. "Facial emotion recognition using convolutional neural networks (FERC)." SN Applied Sciences 2.3 (2020): 1-8.

[24] Sepas-Moghaddam, Alireza, et al. "Facial emotion recognition using light field images with deep attention-based bidirectional LSTM." ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020.