Real-Time Emotion Classification using Facial Expression Recognition

Abhishek Sanjay Gurav[1] Prof. Pramila M. Chawan[2]
[1]M. Tech Student [2]Associate Professor
[1,2]Department of Computer Engineering & IT
[1,2]VJTI College, Mumbai, Maharashtra, India

Abstract— Facial Expressions are a key factor for businesses and organizations that consider it detection of emotions crucial for analysing customer satisfaction. Many researchers have tried to solve this issue with varying levels of accuracy for either 5,6 or 7 expressions and some also for specific expressions only. In this paper, we have implemented our own approach using convolution neural networks on FER2013 dataset using different hardware systems.

Keywords: Emotion Classification, Facial Expression Recognition, Convolution Neural Networks

I. INTRODUCTION

Businesses and marketers are looking for quick ways to get effective and efficient feedback results for their potential customers. Therefore, accurately identifying the emotion of others is critical to successful determination of customer satisfaction. When people see different faces, they can easily recognize the same expression, which is called facial expression recognition. The main fields of application of facial expression recognition technology include intelligent control, security, medical, communication, education, fatigue detection, political election and customer satisfaction. Hence recognizing facial expressions is a task so crucial that it needs to be instantaneous and accurate. In this paper we have implemented a solution to handle issues like accuracy and training costs based on our surveys conducted on the existing methods.

II. LITERATURE REVIEW

K. Liu et al. [1], they proposed an average weighting method to avoid potential errors in real-time facial expression recognition based on the traditional convolutional neural network. Using a high framerate camera, the influence of noise from the environments can be reduced. Because of their average weighting method, recognition results become more robust from frame to frame. As a result, the accuracy was improved. But their CNN only gave accurate results on four expressions.

A. Kartali et al. [2] they have presented a pilot study for real-time testing of both the conventional and the deep learning approaches in facial emotion recognition.

Preparatory results showed better generalization power and, better performance in real-time application of fine-tuned AlexNet CNN and Affdex CNN than MLP and SVM approaches. Commercial Affdex CNN had the superior accuracy overall, and AlexNet and SVM had better 'anger' recognition (96.77% vs. 70.97%). As far as FER-CNN was concerned, it had the lowest overall accuracy but high accuracy for "sadness".

S. Kim et al.[3], an FER system based on machine learning is proposed. ROI rearrangement process minimized the environmental change factor and therefore the hierarchical data structure of the person, and countenance classification improved the classification rate. Based on this study, it might be applied to the personalized vehicle interfaces to stop traffic accidents by combining this proposed system with vehicle.

Anh Vo et al. [4], they proposed and implemented a facial expression recognition framework which could be able to improve the accuracy of recognition of anger and sadness expressions. The experiments showed that their proposed system outperforms other state-of-the-art methods on the CK+ database. The model had a training accuracy of 97.7% but only on six expressions.

Y. Qiu et al. [5], had explored, a possible ability for recognizing countenance supported landmarks which, can could prove that the human brain can recognize facial expressions by using only 68 points rather than all pixels of a face image. The test sample result shows that the method based on landmarks also have a performance comparable with CNN-based methods. But, the performance of the proposed method is additionally associated with the precision of landmark extracting algorithm.

Y. Li, S. Wang et al. [6], a hierarchical framework based on Dynamic Bayesian Network for simultaneous facial feature tracking and facial expression recognition was proposed. For six basic expressions recognition, the result wasn't as nearly good as that of state of the art methods, since they didn't use any measurement specifically for expression, and the global expression is inferred directly from AU and the facial feature point measurements and from their relationships. Average training accuracy achieved was 87.43% for CK+ dataset in their research.

M. S. Bartlett et al. [7], the generalization performance for recognition of full facial expressions of emotion in a seven-way choice was 93.3%. The ML-based system presented here is often applied to recognition of any countenance dimension given a training dataset. They also applied the system to fully automated facial action coding, and obtained a mean agreement rate of 94.5% for 18 AU's from the Facial Action Coding System. This is the primary system that we all know for fully automated FACS coding of images without an infrared eye position signal. The expression classifier outputs change smoothly as a temporal function, providing information about expression dynamics that was previously intractable by hand coding.

III. PROPOSED SYSTEM

A. *Problem Statement*

To provide and efficient approach for emotion classification using facial expression detection from training phase to post-deployment without compromising on latency and accuracy of the results.

B. *Problem Elaboration*

Customer satisfaction is a crucial part of face to face communication is a real-time process operating at a time scale in the order of milliseconds. The level of uncertainty at this temporal scale is considerable, making it necessary for humans and machines to depend on sensory rich perceptual primitives instead of slow symbolic inference processes.

In this project we are presenting the real time facial expression recognition of seven basic human expressions: ANGER, DISGUST, FEAR, HAPPY, SAD, SURPRISE, NEUTRAL.

C. *Roposed Methodology*

1) *Convolution Neural Networks:*

In a model where we are trying to distinguish between various facial expressions, the angle of the eyebrows and length of the smile must be explicitly provided as data points. When using CNN, those spatial features are extracted from the image input. This makes CNN ideal when thousands of features have to be extracted. Instead of measuring the individual features, CNN gathers those features on its own. The advantage of using CNN in our implementation, compared to its predecessors is, it detects important countenance features with none human supervision automatically. This is why CNN is an ideal solution to computer vision and image classification problems.

2) *Performance Metrics:*

Since the problem we have proposed a solution for is aimed at businesses and organizations that rely on emotion detection for customer satisfaction, we have decided these two as our performance metrics:

1) Accuracy: Measure of how accurate our model will perform expression prediction.
2) Confusion Matrix: This metric will give the distribution of the prediction of images across each expression.

3) *Dataset Description:*

As suggested by More, Nikhil T. et al [10], the success of any software project depends on the quality of requirements, for this project we have taken the dataset from Kaggle [9], which has a total of 35888 images. In this dataset, images are classified into 7 classes. Each expression is labelled using an index of 0 to 6 in the following order:

0-Angry
1-Disgust
2-Fear
3-Happy
4-Sad
5-Surprise
6-Neutral

IV. IMPLEMENTATION

A. *Technologies Used:*

- Python
- Keras
- TensorFlow
- Matplotlib
- Pandas
- CPU: 2.7Ghz Intel Core i5
- GPU: Nvidia GTX 750Ti
- Google Colaboratory

B. *Data Preprocessing:*

Preprocessing of images is vital to efficiently train the model. It is advised to convert images to grayscale.
Step-1: Load grayscale images.
Step-2: Crop the image to the face.
Step-3: Resize the image to 48*48.
Step-4: Finally save the image.

C. *Model Description:*

Since, our research has provided us with enough insights about the limits with achieving accuracy due to overfitting of data after 80% accuracy, we have decided to train two models. One with original images and another with mirrored images of the same dataset. Then later concatenating those two models and forming another ensemble neural network for the model. This approach helps us overcome the limitations of traditional approach.

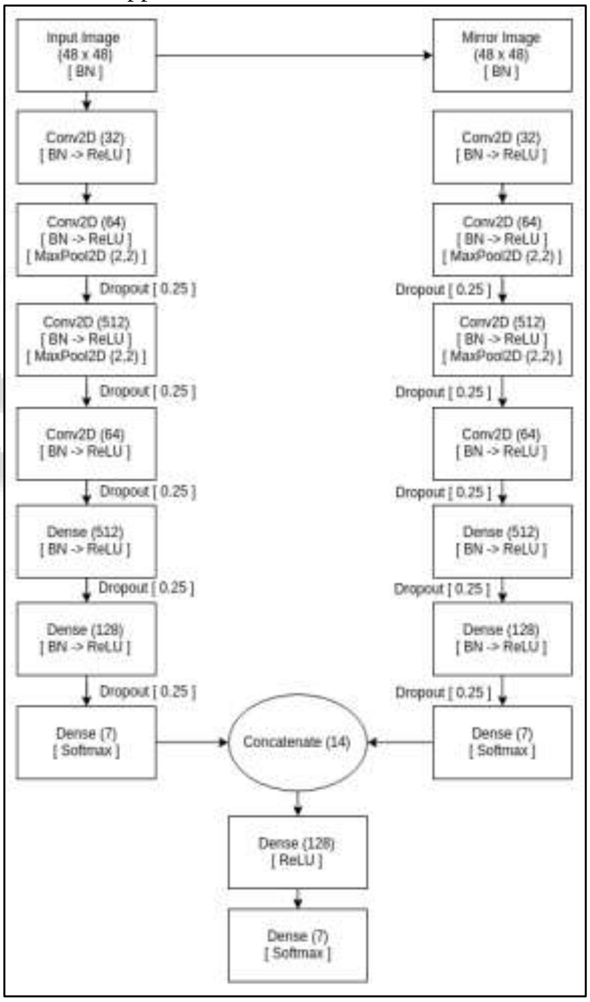

Fig. 1: Neural Network Architecture

Batch size = 32
Number of output classes = 7
Convolution kernel size = 3*3

D. *Video Processing:*

Step-1: Launch the webcam
Step-2: Identify the face using OpenCV HAAR Cascade.
Step-3: Crop the face
Step-4: Dimension the face to 48 * 48 pixels

Step-5: Make a prediction on the face using our pretrained model

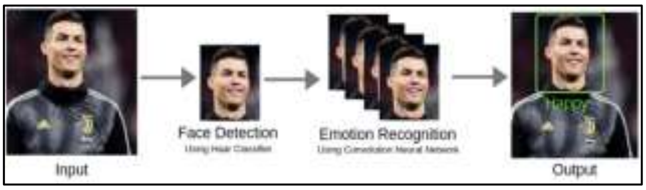

Fig. 2: Run-time visualization

V. RESULTS

We have gone for an 90:10 split, i.e 32300 training samples and 3588 testing samples. An accuracy of 89.4% is achieved after completion 90 Epochs of the model and tested with test data of dataset.

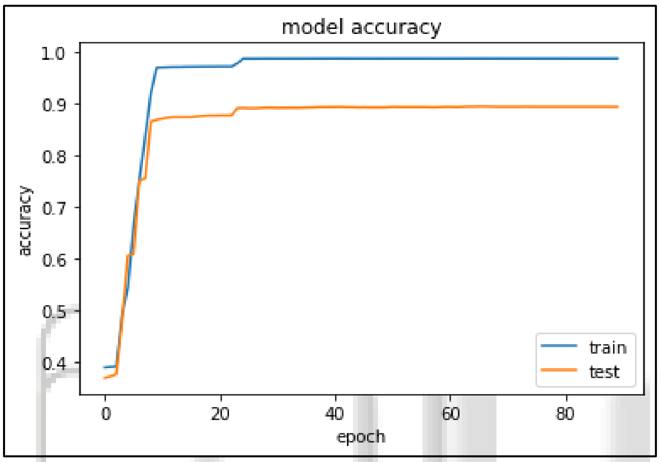

Fig. 3: Accuracy progression while training and testing on dataset

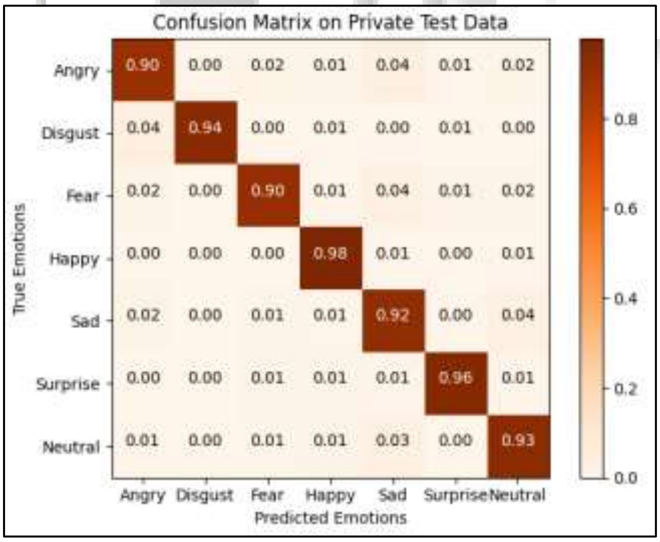

Fig. 4: Confusion Matrix

Model	Accuracy
CNN (Original images)	80.23%
CNN (with Batch Normalization)	85.67 %
CNN-based Ensemble Model	89.40 %

Table 1: Comparison with other methodologies

A. Observed Results:

Following observations were done by manually inputting 100 test images of every expression.
Table below is distributed as follows:

A-Angry
D- Disgust
F- Fear
H- Happy
Sa- Sad
Sr- Surprise
N-Neutral

	Detected As						
Input	A	D	F	H	Sa	Sr	N
A	90	0	0	3	0	0	7
D	0	91	0	6	1	0	2
F	0	2	84	0	3	0	11
H	0	0	0	95	2	0	3
Sa	0	0	1	6	84	0	9
Sr	0	0	10	2	2	83	3
N	0	3	0	3	3	0	91

Table 2: Observed Results in real life

VI. CONCLUSIONS

When using a single convolution network for this problem, the model begins to overfit after achieving 80-85% accuracy, therefore an ensemble neural network made out of the models trained using original and mirrored images was used. This model was able to achieve 89.4% accuracy on FER2013 Kaggle dataset. The confusion in real life analysis with respect to neutral expression comes from its strong resemblance with other expressions. Such level of accuracy will surely help businesses save costs in determining customer satisfaction.

In the future, we will attempt to reduce the confusion caused in real life between ‘neutral’ and other expressions.

REFERENCES

[1] K. Liu, C. Hsu, W. Wang and H. Chiang, "Real-Time Facial Expression Recognition Based on CNN," 2019 International Conference on System Science and Engineering (ICSSE), Dong Hoi, Vietnam, 2019, pp. 120-123, doi: 10.1109/ICSSE.2019.8823409.

[2] A. Kartali, M. Roglić, M. Barjaktarović, M. Đurić-Jovičić and M. M. Janković, "Real-time Algorithms for Facial Emotion Recognition: A Comparison of Different Approaches," 2018 14th Symposium on Neural Networks and Applications (NEUREL), Belgrade, 2018, pp. 1-4, doi: 10.1109/NEUREL.2018.8587011.

[3] S. Kim, G. H. An and S. Kang, "Facial expression recognition system using machine learning," 2017 International SoC Design Conference (ISOCC), Seoul, 2017, pp. 266-267, doi: 10.1109/ISOCC.2017.8368887.

[4] Vo and B. T. Nguyen, "Facial Expression Recognition Based on Salient Regions," 2018 4th International Conference on Green Technology and Sustainable Development (GTSD), Ho Chi Minh City, 2018, pp. 739-743, doi: 10.1109/GTSD.2018.8595551.

[5] Y. Qiu and Y. Wan, "Facial Expression Recognition based on Landmarks," 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chengdu, China, 2019, pp. 1356-1360, doi: 10.1109/IAEAC47372.2019.8997580.

[6] Y. Li, S. Wang, Y. Zhao and Q. Ji, "Simultaneous Facial Feature Tracking and Facial Expression Recognition," in IEEE Transactions on Image Processing, vol. 22, no. 7, pp. 2559-2573, July 2013, doi: 10.1109/TIP.2013.2253477.
[7] M. S. Bartlett, G. Littlewort, C. Lainscsek, I. Fasel and J. Movellan, “Machine learning methods for fully automatic recognition of facial expressions and facial actions,” in IEEE International Conference on Systems, Man and Cybernetics, 2004, pp. 592-597.
[8] Abhishek Gurav, Pramila M. Chawan, “Real-time Emotion Classification using Facial Expression Recognition: A Survey”, International Research Journal of Engineering and Technology (IRJET) Volume 7, Issue 11, November 2020
[9] Kaggle - Challenges in Representation Learning: Facial Expression Recognition Dataset: https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data
[10] More, Nikhil T.; Sapre, Bhushan S.; and Chawan, Pramila M. (2017) "An Insight into the Importance of Requirements Engineering," International Journal of Computer and Communication Technology: Vol. 8 : Iss. 1 , Article 5. DOI: 10.47893/IJCCT.2017.1394