International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249 – 8958 (Online), Volume-9 Issue-4, April 2020

Facial Expression Analysis using Convolutional Neural Networks

Amogh S. Gopadi, Deepak S., Kiran R. B., Naveena A. M., Srividya M. S., Anala M. R.

Abstract*: *Human feelings are mental conditions of sentiments that emerge immediately as opposed to cognitive exertion. Some of the basic feelings are happy, angry, neutral, sad and surprise. These internal feelings of a person are reflected on the face as Facial Expressions. This paper presents a novel methodology for Facial Expression Analysis which will aid to develop a facial expression recognition system. This system can be used in real time to classify five basic emotions. The recognition of facial expressions is important because of its applications in many domains such as artificial intelligence, security and robotics. Many different approaches can be used to overcome the problems of Facial Expression Recognition (FER) but the best suited technique for automated FER is Convolutional Neural Networks(CNN). Thus, a novel CNN architecture is proposed and a combination of multiple datasets such as FER2013, FER+, JAFFE and CK+ is used for training and testing. This helps to improve the accuracy and develop a robust real time system. The proposed methodology confers quite good results and the obtained accuracy may give encouragement and offer support to researchers to build better models for Automated Facial Expression Recognition systems.

Keywords: Convolutional Neural Network, Deep Learning, Facial Expression Recognition, OpenCV DNN.

I. INTRODUCTION

Humans are emotional creatures. We are guided by our emotions in many ways, so understanding them is very important. The best way forward in doing so is to understand facial expressions. The changes in facial muscles together with the emotional state of a person is known as facial expression. Analysis of facial expressions has many applications such as Human Behavior Predictor, Surveillance System and Medical Rehabilitation. It can also be useful in other domains such as robotics, education, automation, etc.

Revised Manuscript Received on April 13, 2020.
* Correspondence Author
Amogh S. Gopadi*, Dept. of Computer Science and Engineering, R.V. College of Engineering, Bengaluru, India. Email:amoghsgopadi@gmail.com
Deepak S., Dept. of Computer Science and Engineering, R.V. College of Engineering, Bengaluru, India. Email: dpks17668@gmail.com
Kiran R. B., Dept. of Computer Science and Engineering, R.V. College of Engineering, Bengaluru, India. Email: kiranrb54321@gmail.com
Naveena A. M., Dept. of Computer Science and Engineering, R.V. College of Engineering, Bengaluru, India. Email: amnaveen471@gmail.com
Srividya M. S., Dept. of Computer Science and Engineering, R.V. College of Engineering, Bengaluru, India. Email: srividyams@rvce.edu.in
Anala M. R., Dept. of Computer Science and Engineering, R.V. College of Engineering, Bengaluru, India. Email: analamr@rvce.edu.in

A facial expression recognition framework could become the core part of socially intelligent robots and fill in as potential applications in various areas of psychological studies, driver lethargy checking, interactive games, versatile portable application to naturally embed feelings in chat and facial nerve regenerating in therapeutic field. Despite the fact that much advancement has been made, perceiving human emotion with a high precision stays to be troublesome because of the intricacy and assortments of facial expressions. Thus, a facial expression recognition system is of utmost importance. In this paper, a new methodology based on the concept of deep learning is proposed. The main aim of this paper is to develop a facial expression recognition system. This system will be used to classify facial expressions into basic emotions namely happy, angry, sad, neutral and surprise. OpenCV Deep Neural Networks is used to detect faces and facial expressions are recognized using Convolutional Neural Networks. To achieve higher accuracy and better real time recognition rate, the model is trained on a combined dataset comprising of images from FER2013, FER+, CK+ and JAFFE datasets. Section II will present the literature survey. In section III, the proposed methodology including the data set description and architecture of the model is explained. In section IV, we have discussed the complete result analysis obtained by the model. Finally, the conclusion is given in section V.

II. LITERATURE SURVEY

A wide range of approaches to different techniques for the purpose of facial expression analysis is discussed in the following literature survey. In [1], an automatic facial expression recognition system using deep CNN is proposed. In this paper, a system is devised for recognizing facial expressions using Deep Convolutional Neural Networks. This system is capable of discovering deep feature representation of facial expressions. In [2], a method for recognition of human expressions using a deep learning framework is proposed. The method uses Gabor filters for feature extraction and Convolutional neural networks for classification. A methodology for facial expression recognition system based on convolutional neural network with data augmentation is proposed in [3]. Data augmentation is required as the CNN architecture requires a large set of previously labelled images which may be difficult to find in public databases. In [4], the problem of emotion detection from images and videos is tackled by using a convolutional neural network approach. The proposed architecture performs better than the previous convolutional neural network

approaches as it is independent of hand-crafted feature extraction.

III. METHODOLOGY

In this research work, a facial expression recognition system is proposed. The detection and recognition of the emotion depends on the dataset used, pre-processing techniques applied, architecture of the model and hyper parameter optimizations. These are explained in detail in the following sub sections:

A. Data Collection and Pre-Processing

A combined dataset is formed by collecting images from different sources. The different datasets used in this research are FER-2013 and FER+ dataset, Extended Cohn Kanade(CK+) database, Japanese Female Facial Expression database(JAFFE) and our own dataset. This is done to improve the generalization ability of the model and to take care that the model is not biased towards a specific group. The datasets used mainly differs in factors like pose, image-quality, alignment, clarity etc. The FER-2013 comprises of 35000 low resolution images of faces [5]. It consists of images in different degrees of angle and dissimilar age groups. The FER-2013 database has many wrongly classified images which decreases the accuracy. To overcome this misclassifications, we use the FER+ dataset. The standard FER2013 dataset is provided a set of new labels by the annotations in FER+ [6]. This is done by labelling images in the original FER dataset by 10 crowd-sourced taggers. In addition to this, the Extended Cohn-Kanade dataset (CK+) has been used which exhibits facial expressions very clearly. The CK+ comprises a total of 593 sequences across 123 subjects [7]. They range from neutral to peak expression. The JAFFE database contains poses of 10 Japanese female models in 7 different facial expressions, totaling about 213 images [8]. In addition to all of these, we have added images captured in real time of 15 subjects in 5 different poses. All these are combined to form the final dataset. The table I shows an exhaustive list of the datasets:

Table I: Number of images in each dataset

	Happy	Sad	Angry	Neu-Tral	Sur-prise	Total
FER+	9422	4625	3259	13457	4947	35710
CK+	69	28	45	593	83	818
JAFFE	31	31	30	30	30	152
OUR	74	70	70	70	75	359
Total	9596	4754	3404	14150	5135	37039

To improve the validation and testing accuracy, a number of pre-processing techniques were applied on the training and testing images, namely Face detection and cropping, scaling, grayscale conversion and data augmentation. To find out an appropriate face detection technique best suited for real time recognition, we tested the Haar Cascade Face Detector and Deep Neural Network Face Detector in OpenCV. The experiments were conducted on a total of 1400 randomly selected images from the training and testing set. The summary of the results with the number of faces detected by each method can be observed in the following table:

Table II: Summary of results on Face detection techniques

Method	Number of faces detected
Haar Cascade Face Detector	997
DNN Face Detector	1040

From our experiments, we found Deep Neural Networks Face Detector in OpenCV to be the best suited for the given task. It is based on the Single Shot Detector framework and a ResNet-10 Architecture is used as its backbone. It can effectively run in real time and has a very accurate face detection rate. It can also work on non-aligned faces with occlusions. After detecting the face, the image is cropped to include the detected face and remove the unwanted elements. This makes the training process more efficient by overcoming the background complexity. Next, these images are resized to 46 X 46 pixels using OpenCV's resize function. After this, all the images in the dataset are converted to grayscale to reduce the pixel complexities using grayscale conversion pre-processing technique. From the numbers in table 1, it can be noted that the dataset is very much biased towards the neutral emotion. To remove this bias and to give equal weightage to all the datasets, Image augmentation is performed. This helps in generating additional images by applying random operations, such as random rotation, shifts, shear, flips etc. on existing image dataset. This will also help the model to become more robust and improve its generalization capacity. The augmentation is done with the help of Keras ImageDataGenerator function. Some of the augmentation operations applied on the dataset are rotation at a certain angle, shifting the height and width, zooming and horizontal flip. The figure 1 shows the data augmentation parameters and the values specified to each of them.

Data augmentation parameters	Values of the specified parameters
Rotation	20
Width shift	0.1
Height shift	0.1
Zoom range	0.1
Horizontal flip	True

Fig. 1. Data Augmentation parameter values

The dataset had a total of 37000 images. After applying augmentation on the dataset, we now got a total of 62000 images with around 11000 images per class. Further, the dataset was split into three sets, training, testing and validation. For model training, 70% of the images had been randomly selected and the remaining 20% for model validation. Another 10% of the images were used for testing the trained model.

B. Architecture of Convolutional Neural Network and Hyper Parameter Optimization

The neural networks with convolution layers is known as Convolutional Neural Networks**.** Equation (1) shows the convolution over an image f (m, n) using a filter h (m, n):

$$f(m,n)*h(m,n)=\sum_{i}\sum_{j}h(i,j)f(m-i,n-j) \quad (1)$$

CNN is best suited for image recognition tasks since they make use of the concept of receptive fields and weight sharing. Thus CNN has been used to overcome the challenges in facial expression classification. To choose the best performing network architecture for the task of facial expression recognition on the given dataset, a comparison of many pre-trained models and novel architectures is carried out. After taking into consideration the accuracies and learning obtained through different models, a novel CNN architecture is designed in this research. The structure of the proposed CNN architecture is given below.

LAYER TYPE	DETAILS
First Convolution Layer	64 filters of size 3x3, ReLU
Second Convolution Layer	64 filters of size 3x3, ReLU
First Max Pooling Layer	Pooling Size 2x2
Dropout Layer	Excludes 50% neurons randomly
Third Convolution Layer	128 filters of size 3x3, ReLU
Fourth Convolution Layer	128 filters of size 3x3, ReLU
Second Max Pooling Layer	Pooling Size 2x2
Dropout Layer	Excludes 50% neurons randomly
Fifth Convolution Layer	256 filters of size 3x3, ReLU
Sixth Convolution Layer	256 filters of size 3x3, ReLU
Third Max Pooling Layer	Pooling Size 2x2
Dropout Layer	Excludes 50% neurons randomly
Seventh Convolution Layer	512 filters of size 3x3, ReLU
Eighth Convolution Layer	512 filters of size 3x3, ReLU
Fourth Max Pooling Layer	Pooling Size 2x2
Dropout Layer	Excludes 50% neurons randomly
First Fully Connected Layer	512 nodes, ReLU
Dropout Layer	Excludes 50% neurons randomly
Second Fully Connected Layer	256 nodes, ReLU
Dropout Layer	Excludes 50% neurons randomly
Third Fully Connected Layer	128 nodes, ReLU
Dropout Layer	Excludes 50% neurons randomly
Fourth Fully Connected Layer	5 nodes, ReLU
Output Layer	5 nodes for 5 classes, SoftMax

Fig. 2. Proposed CNN Architecture

The network is given a pre-processed image of 46 x 46 pixels as input. The network consists of eight two-dimensional convolution layers, four max pooling layers and four fully connected layers. The eight convolution layers are divided into four groups each comprising of two layers. The number of filters in each succeeding group are increased by a factor of 2. Thus, different number of filters are used in each group of the convolution layers progressively starting from 64. Each group is then followed by a single max pooling layer and a dropout layer with a value of 0.5. This is followed by four fully connected layers with 512,256,128 and 5 nodes respectively. After this, a dropout layer has been inserted with a value of 0.5 to exclude 50% of the neurons randomly. At the end, a softmax activation function is used which serves as the output layer. It consists of 5 nodes to predict 5 class output.

To improve the learning process and quality of output, hyper parameter optimization is performed. Hyper-parameters are the parameters which can be arbitrarily set by the user before starting the training process. The approach used for Hyper-parameter optimization is Random Search. In this, we create a grid of hyper parameters and train/test our model on some random combination of these hyper parameters. The optimization is done with the help of Scikit-learn RandomizedSearchCV function. Using the scikit-learn best_estimator_ attribute, the set of hyperparameters which performed best during training are retrieved. The optimised values for the corresponding hyper parameters are as shown below in table III.

Table III: Hyper parameter values

Hyper-parameter	Value
No. of epochs	36
Dropout ratio	0.5
Size of Convolution kernel	3x3
Size of Max Pooling kernel	2x2
Max pooling stride	1
Loss function	Categorical Cross Entropy
Optimiser	Adam
Learning rate	0.01

The optimal number of epochs required to achieve the highest accuracy during training is 36. The ideal Dropout ratio rate is found to be 0.5. Upon tuning the hyper parameters, the highest accuracy is achieved for Adam Optimizer. This is used with a learning rate of .01. The loss function used is Categorical Cross entropy. A number of Callbacks have been included in training. They help to monitor and improve the training process. Some of them are ModelCheckPoint, ReduceLROnPlateau, EarlyStopping and TensorBoard.

IV. RESULTS AND DISCUSSION

A comparison of pre-trained models and proposed architecture is carried out to choose the best performing network architecture for the task of facial expression recognition on the given dataset. The following table IV shows the results of this experiment:

Table IV: Summary of results using different architectures

	Training Accuracy	Testing Accuracy
AlexNet	63.4%	56.2%
VGGNet	86.7%	62.3%
Proposed CNN	84.6%	73.4%

Many other models were used whose results have not been mentioned as they performed poorly on the dataset. From the experiments conducted, we learnt that the given dataset requires a complex CNN architecture rather than a simple one. Based on this, we constructed the proposed network to achieve the highest accuracy. Our model achieved an accuracy of 84.6% on training data and 73.4% on testing data. Figure 3 and figure 4 shows the loss and accuracy curve respectively:

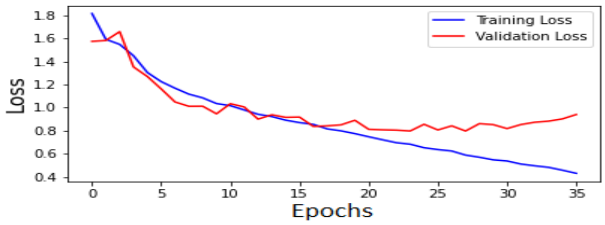

Fig. 3. Graph of loss function vs number of epochs

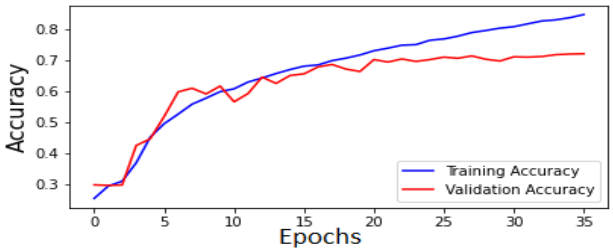

Fig. 4. Graph of accuracy vs number of epochs

The progress of the model accuracy is visualized using TensorBoard. The training process is carried out on Python 3 Google Compute Engine backend(GPU) provided by Google Colab. It took about 960 seconds to obtain the highest accuracy and finish the training process.

To evaluate the model even further and to gain a better understanding on the classification accuracies of each emotion class, Matplot library is used to create a confusion matrix on the images of the test set. The obtained results on the given dataset are shown in figure 5.

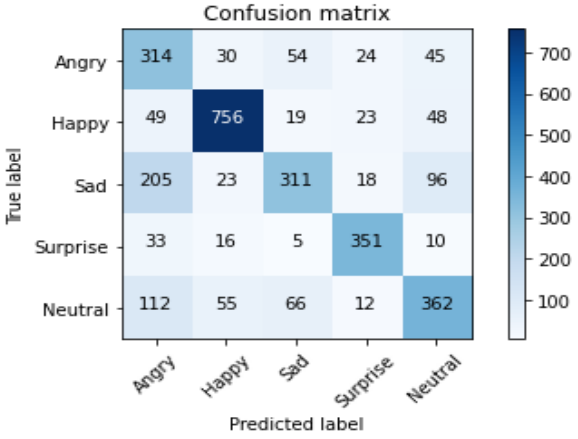

Fig. 5. Confusion Matrix

As shown in figure 5, the actual classes are shown in rows and predicted classes are shown in columns. The model made a total of 3037 predictions where the model predicted angry for 713 times, happy for 880 times, sad for 455 times, surprise for 428 times and neutral for 561 times. However, in reality 467 cases were angry, 895 were happy, 653 were sad, 415 were surprise and 607 were neutral. From this, we observe that Happy has the maximum number of correct classifications and Sad has the most number of misclassifications. To test the model in real time, a small tool is developed using Tkinter library in python. Almost all instances of happiness, surprise and neutral are correctly predicted but sometimes sadness and anger are mispredicted. Another key point to note is that upon detecting a face the model predicted the emotions almost instantaneously with no delays. The figure 6 shows a screenshot of the GUI and realtime results.

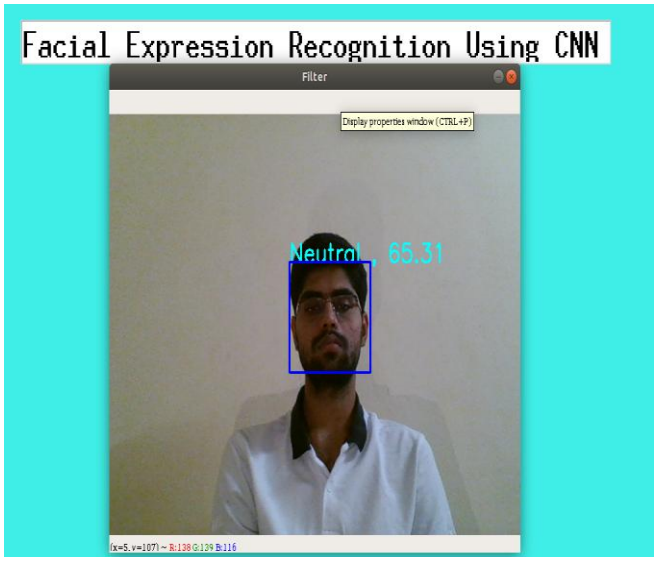

Fig. 6. Screenshot of GUI and Real-time Results

V. CONCLUSION

This work presented a novel methodology for facial expression recognition using convolutional neural network. A novel CNN architecture is proposed to achieve the highest accuracy after experimenting with other existing architectures. Different sources such as FER, FER+, CK+ and JAFFE datasets were utilized for collecting images to build the dataset. The existing bias in the dataset is removed by performing data augmentation. The proposed architecture is then trained on this dataset exhaustively to obtain the trained model. To fine tune the model and achieve better results, hyper parameter optimization is performed. All these preprocessing techniques helped to improve the suitability of the system in real time scenarios. Finally, a real time facial expression recognition system is developed and tested in real world conditions. Some of the improvements that can be made to the existing system are recognizing complex emotions and increasing the practicality of the model. Also our system could be integrated with other components to detect human emotions effectively as facial expressions alone cannot determine the

Retrieval Number: D8335049420/2020©BEIESP
DOI: 10.35940/ijeat.D8335.049420
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

emotion of the person. This paper could also offer support to researchers to carry further work in the field of Facial Expression Analysis.

REFERENCES

1. Shan K., Guo J., You W., Lu D., Bie R. "Automatic Facial Expression Recognition based on a Deep Convolutional-Neural-Network structure" IEEE SERA, 2017
2. Milad Mohammad, Taghi Zadeh, Maryam Imani, Babak Majidi "Fast Facial emotion recognition Using Convolutional Neural Networks and Gabor Filters", IEEE 5th Conference on Knowledge-Based Engineering and Innovation, 2019
3. Tawsin Uddin Ahmed, Sazzad Hossain, Mohammed Shahabad Hossain, Raihan Ul Islam, Karl Anderson "Facial Expression Recognition using Convolutional Neural Network with Data Augmentation", Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV), 2019 and 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR), 2019
4. Peter Burkert, Felix Trier, Muhammad Zeshan Afzal, Andreas Dengel, Marcus Liwickim "DeXpression: Deep Convolutional Neural Network for Expression Recognition", arXiv:1509.05371v2, 17 August 2016.
5. I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, et. al "Challenges in representation learning: A report on three machine learning contests", Neural Networks, Special Issue on "Deep Learning of Representations", 64:59--63, 2015
6. E. Barsoum, C. Zhang, F. Canton, Z. Zhengyou "Training deep networks for facial expression recognition with crowd-sourced label distribution", International Conference on Multimodal Interaction (ICMI), 2016
7. Lucey, Patrick "The Extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression", Computer Vision and Pattern Recognition Workshops (CVPRW), 2010
8. Michael J. Lyons, Shigeru Akemastu, Miyuki Kamachi, Jiro Gyoba. "The Japanese Female Facial Expression (JAFFE) database", 3rd IEEE International Conference on Automatic Face and Gesture Recognition.
9. Abir Fathallah Lotfi, Abdi Ali Douik "Facial Expression Recognition via Deep Learning", IEEE/ACS 14th International Conference on Computer Systems and Applications, 2017
10. G. A. Rajesh Kumar, Ravi Kant Kumar, Goutam Sanyal **"**Facial Emotion Analysis using Deep Convolution Neural Network", International Conference on Signal Processing and Communication (ICSPC), 2017
11. Li M., Xu H., Huang X., Song Z., Liu X. and Li X. "Facial Expression Recognition with Identity and Emotion Joint Learning", IEEE Transactions on Affective Computing, 2018
12. A. Krizhevsky, I. Sutskever, and G. E. Hinton. "Imagenet classification with deep convolutional neural networks", Advances in neural information processing systems, 10971105.
13. Shan Li, Weihong Deng. "A Deeper Look at Facial Expression Dataset Bias", IEEE Transactions on Affective Computing, 2020

AUTHORS PROFILE

Amogh S Gopadi is an Undergraduate Scholar pursuing Computer Science & Engineering from R.V College of Engineering. He is passionate about research and his area of interests are Machine Learning and Artificial Intelligence. He is currently working under the guidance of Prof. Srividya M. S.

Deepak S is an Undergraduate Scholar pursuing Computer Science & Engineering from R.V College of Engineering. His area of interests are Machine Learning, Computer Network and Security. He is currently working under the guidance of Prof. Srividya M.S.

Kiran R. B. is an Undergraduate Scholar pursuing Computer Science & Engineering from R.V College of Engineering. He is passionate about research and his area of interests are Machine Learning and Computer Networks. He is currently working under the guidance of Prof. Srividya M. S.

Naveena A. M. is an Undergraduate Scholar pursuing Computer Science & Engineering from R.V College of Engineering. He is passionate about research and his area of interest is Application Development. He is currently working under the guidance of Prof. Srividya M. S.

Prof. Srividya M. S. is an Assistant Professor at R.V College of Engineering. She has over 8 years of experience in teaching and 8 years of experience in industry. Main area of research interest is Image Processing, Video Processing and Neural Networks. She has guided 5 UG projects and has many publications in international journals and conferences.

Dr. Anala M. R. is an Associate Professor at R.V College of Engineering. She has over 18 years of experience in teaching. Main area of research interest is Computer Architecture, High Performance Computing, Distributed Systems and Parallel programming. She has guided 40 UG projects and 20 PG projects. She has many publications in international journals and conferences
.