Survey on Facial Emotion Recognition with Convolutional Neural Network

Sarvesh Vedpathak[1], Shailesh Bendale[2]

[1]Student, Dept. of Computer Engineering, NBN Sinhgad School of Engineering, Pune, Maharashtra, India
[2]Head of Dept., Dept. of Computer Engineering, NBN Sinhgad School of Engineering, Pune, Maharashtra, India

---

Abstract - *Facial Expression Recognition is expeditiously developing branches in the Aritificial Neural Network domain. In this paper, we show the arrangement of FER based on static images, utilizing CNNs, without requiring any pre-processing or feature extraction tasks. The paper additionally represents techniques to improve on future accuracy around here by utilizing pre-processing, which incorporates face identification and brightening correction. Datasets like JAFFE and FER2013 were used for performance analysis. Pre-processing strategies like facial landmarks and HOG were consolidated into a convolutional neural network and this has accomplished good accuracy when compared with the existing model.*

Key Words: **Facial Expression Recognition (FER), Convolutional Neural Networks (CNNs), Artificial Intelligence (AI), Facial Action Coding System (FACS)**

1.INTRODUCTION

Facial Expression Recognition an active field in AI and Artificial Neural Network. Some application related this includes Biometric, Information Security, Smart Card, Forensic application, Investigation of Image Databases, Human-Computer Interaction, Verification for identity, Automated Surveillance, Indexing of Videos, Cosmetology and so on. The improvement in this field initially started with the basic standards which were identified with biological nature and then went onto more sophisticated things like machine learning with the help of modern technology. Though a huge advancement has happened in this field, it actually does not have the capacity to offer an definite answer due to the complexity and wide variety of expressions possessed by the people. The majority of the new techniques and the generally existing ones can't in any case ensure that the outcomes are completely exact. The objective of this paper is to arrange images of human faces into discrete emotion classifications using Convolutional Neural Networks, additionally show the pre-processing and feature extraction strategies to improve the accuracy on the FER2013 [1].

Ekman and Friesen tended to this by taking on the framework developed by CarlHerman Hjortsjo called facial activity coding framework (FACS).¨ FACS is a framework based on facial muscle changes and can characterise facial activities to express individual human feelings as defined by Ekman and Friesen in 1978. The Action Units described above show the different movements of facial muscles, which reflect particular momentary changes in facial appearance. Ekman and Friesen had before described the six key fundamental feelings which are happy, anger, fear, surprise, disgust, and sad. This was later on used as the most fundamental feeling to be recognized added with the neutral emotion [2].

2. RELATED WORK

Duchenne de Boulogne was a French neurologist in the nineteenth century, who was keen on Physiognomy and needed to understand how human face muscles work to produce facial expression, as he believed that these were straightforwardly connected to a human's spirit. To do this, he used electric tests to trigger muscle compressions, and afterward took pictures, using recently developed camera innovation, of his subjects' faces showing the twisted expression he was able to make [3].

In 2016, Pramerdorfer and Kampel achieved accuracy of 75.2% on the FER2013, utilizing CNNs. They likewise prepared the engineering for up to 300 epochs and used stochastic angle drop to optimize the cross-entropy loss, with an momentum value 0.9. Different boundaries were fixed, such as learning rate with 0.1, cluster size with 128, and weight decay with 0.0001[4].

Zhang Luo, C.-C. Loy, and X. Tang [5] used a Siamese Network to present a methodology for understanding social relation practices from images and achieved a test precision of 75.1% on the difficult Kaggle look dataset. The authors utilized different datasets, with various names, to increase the training data; they introduced a feature extraction system, just as dealing with feature extraction.

Shekhar Singh and Fatma Nasoz performed competitively and accomplished a FER2013 test accuracy of 61.7%, without including any pre-processing or feature extraction techniques. The test accuracy for the seven facial expression utilizing the ensemble of CNNs was 75.2%. They performed few examination using distinctive batch sizes and epochs yet obtained the best test accuracy utilizing a batch size value of 512 and 10 epochs.

B.- K. Kim, S.- Y. Dong, J. Roh, G. Kim, and S.- Y. Lee [6] proposed ensemble of Convolutional Neural Networks and exhibited that during training and testing it is advantageous to utilize both registered and unregistered types of given face pictures. The authors accomplished a test exactness of 73.73% on the FER2013 dataset. They likewise conducted

Intraface for an ordinary 2-D arrangement, which is freely accessible for milestone detector, and performed brightening standardization. To keep away from the registration error, they performed registration specifically, in view of the consequences of facial milestone recognition.

A. Lopes, E. Aguiar, A. De Souza, and T. Oliveira-Santos [7] proposed their test results show that the blend of normalization techniques made huge upgrade to the accuracy. This methodology uses the area of each one of the eyes while doing the pre-processing strategies, and it was found that it could easily incorporated without impacting the constant idea thought of the framework. The accuracy of the model is supposed to be 96%, but for specific expression like sad, they were essentially prepared to achieve a accuracy of 84%.

3. DATASET

The dataset FER2013 [8] was introduced at the International Conference on Machine Learning 2013 Workshop on Challenges in Representation Learning. FER2013 is a huge dataset, which is freely accessible on Kaggle's FER Challenge. The FER2013 dataset contains 35,887 face crops, including training, validation and testing pictures, with 28,709, 3,589 and 3,589, individually. All pictures are of 48x48 pixel goals and on grayscale. Ian Goodfellow obtained the accuracy of this dataset to be around 65.5%.

4. METHODOLOGY

In this section, we will discuss about our CNNs architecture and methods to additionally work on the accuracy on the dataset FER2013.

1. Pre-processing

Pre-processing can be utilized to improve FER framework execution and can be done previous to the feature extraction process. Picture pre-processing incorporates different processes, like the recognition and arrangement of faces, correction of brightening, occlusion, pose and data augmented.
The picture brightness and contrast vary with brightening and lighting condition of object. Such varieties cause expanded complexity of feature sets and the detection method.

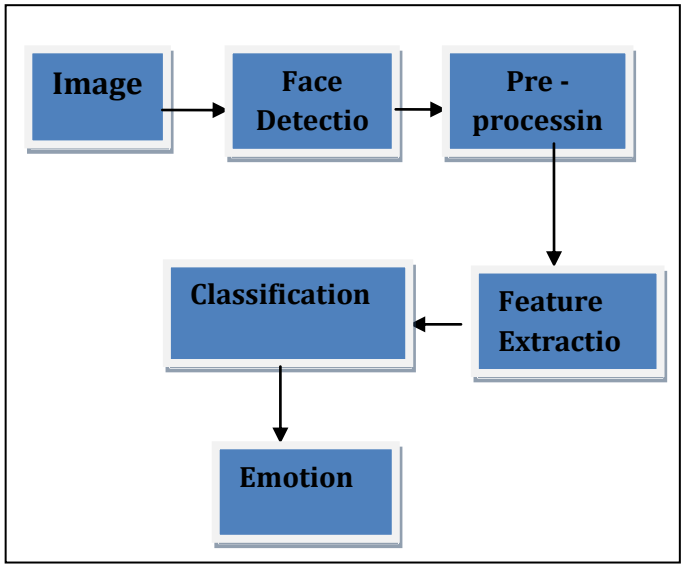

Fig -1: System Architecture of Facial Emotion Recognition

2. Feature Extraction

Appropriately pre-processed data is given as input data into the next module where facial feature extraction is done. Feature extraction selects the relevant information contained in the image so that the detecting the emotion is made easy. This work executes a crossover strategy for facial component extraction by combining the features extracted by CNN with HOG [9] and facial landmarks [10]. HOG depicts local object appearance and shape inside a picture by the circulation of edge bearings. HOG works on local cells; accordingly, it is invariant to changes. These HOG highlights differ for each expression, so recognizing them is more comfortable. This is the reason they chose HOG as the component selector for this framework.
This framework executes a Convolutional Neural Network (CNN) [11] for grouping of emotion utilizing the feature extracted utilizing abilities of CNN and the extra feature extracted in the past module. Proposed CNN has an input layer, four layers of convolution, two layers of pooling, and two fully connected layers for order. Next module portrays the experimental arrangement and discuss the outcomes obtained.

5. EXPERIMENTS AND DISCUSSION

All the tests were conducted utilizing two freely accessible datasets in facial expression recognition field: The JAFFE dataset and FER2013 dataset [12]. JAFFE dataset includes pictures of 7 emotions of 10 Japanese female models. Pictures are 256x256 grey scaled. These pictures were resized since our Convolutional Neural Network model accepts 48x48 pictures as input data. FER2013 dataset contains 35887 pictures, size of 48 x 48 pixels of 7 facial emotions.

To examine the performance of the system utilizing the datasets, Experiment carried out the system with the help of

OpenCV and Keras. Pre-processing is done with OpenCV, and CNN is constructed utilizing Keras.

The proposed framework joins HOG feature and facial milestones with those facial feature extracted by convolutional layers by using a similar CNN architecture, however the HOG feature and facial milestones are added to those exiting the last convolutional layer. The hybrid feature set then, at that point, enters the fully connected layers for further processing.

Table I: Number of images in each emotion of the two datasets [2].

Facial Expression	JAFEE	FER2013
Angry	30	4593
Disgust	29	547
Fear	32	5121
Happy	32	8989
Sad	30	6077
Surprise	30	4002
Neutral	30	6198

Recent works [13] tested their technique independently with each dataset. So they additionally evaluated the framework independently with JAFFE and FER2013 datasets. For evaluation, both datasets are split in the proportion of 80:20 for training and testing individually.

Performance analysis is performed with and without pre-processing. First, The performance of system evaluated utilizing FER2013 dataset. It is clear that without pre-processing steps, the accuracy obtained was 54.2% and when image cropping is applied, it resulted in significantly increase in the accuracy to 74%. At the point when both intensity normalization and cropping is done, then, at that point, the accuracy expanded to 74.4%. Then, at that point, evaluation was performed utilizing JAFFE dataset. Without including pre-processing steps the accuracy obtained was 90.698%. When cropping is applied, it results in a slight expansion in the accuracy to 91.2%. No change in accuracy was there when intensity normalization is applied, since the dataset is very small as compared to FER2013.

6. CONCLUSIONS

In the proposed method, seven unique facial expression from two datasets, JAFFE and FER2013 have been investigated. The facial pictures were preprocessed after it was caught, from which the feature were extracted and the facial expression was identified by the CNN model based on the training. To measure the presentation of the proposed algorithm and check the outcomes, the framework was evaluated utilizing the metric accuracy. The equivalent datasets were utilized for both training and testing by separating the datasets into training and testing samples in the proportion of 80:20 for both JAFFE and FER2013. Experiment results on two datasets, JAFFE and the FER2013 dataset, show that the proposed technique can achieve an excellent performance. A accuracy of 91.2% and 74.4% was obtained on JAFFE and FER2013 datasets separately.

7. REFERENCES

[1] Shekhar Singh and Fatma Nasoz, "Facial Expression Recognition with Convolutional Neural Networks" 2020 IEEE

[2] Dr Ansamma John , Abhishek MC , Ananthu S Ajayan , Sanoop S and Vishnu R Kumar, "Real-Time Facial Emotion Recognition System With Improved Preprocessing and Feature Extraction" IEEE Xplore Part Number: CFP20P17-ART; ISBN: 978-1-7281-5821-1

[3] Duchenne, G.-B. (1862), Mécanisme de la physionomie humaine, ou analyse électro-physiologique de ses différents modes de l'expression. Paris: Archives générales de médecine, P. Asselin; vol. 1, p. 29-47, 152-174.

[4] C. Pramerdorfer and M. Kampel, "Facial expression recognition using convolutional neural networks: State of the art," arXiv preprint arXiv:1612.02903, 2016.

[5] Z. Zhang, P. Luo, C.-C. Loy, and X. Tang, "Learning Social Relation Traits from Face Images," in Proc. IEEE Int. Conference on Computer Vision (ICCV), 2015, pp. 3631–3639.

[6] B.-K. Kim, S.-Y. Dong, J. Roh, G. Kim, and S.-Y. Lee, "Fusing aligned and non-aligned face information for automatic affect recognition in the wild: A deep learning approach," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2016, pp. 48–57.

[7] A. Lopes, E. Aguiar, A. De Souza, and T. Oliveira-Santos, "Facial expression recognition with convolutional neural networks: Coping with few data and the training sample order," Pattern Recognition, vol. 61, 07 2016.

[8] "Challenges in Representation Learning: A report on three machine learning contests." I Goodfellow, D Erhan, PL Carrier, A Courville, M Mirza, B Hamner, W Cukierski, Y Tang, DH Lee, Y Zhou, C Ramaiah, F Feng, R Li, X Wang, D Athanasakis, J Shawe-Taylor, M Milakov, J Park, R Ionescu, M Popescu, C Grozea, J Bergstra, J Xie, L Romaszko, B Xu, Z Chuang, and Y. Bengio. arXiv 2013.

[9] A. Nandi, P. Dutta, and Md. Nasir, "Automatic facial expression recognition using histogram oriented gradients (hog) of shape information matrix," in Intelligent Computing and Communication, V. Bhateja, S. C. Satapathy, Y.-D. Zhang, and V. N. M. Aradhya, Eds. Singapore: Springer Singapore, 2020, pp. 343–351.

[10] M. Bodini, "A review of facial landmark extraction in 2d images and videos using deep learning," Big Data and Cognitive Computing, vol. 3, p. 14, 02 2019.

[11] C. Pramerdorfer and M. Kampel, "Facial expression recognition using convolutional neural networks: State of the art," 12 2016.

[12] Wolfram Research, FER-2013. Wolfram Data Repository, 2018.

[13] Z. Yu and C. Zhang, "Image based static facial expression recognition with multiple deep network learning," November 2015.