2020 International Conference for Emerging Technology (INCET)
Belgaum, India. Jun 5-7, 2020

Facial Emotion Detection Using Deep Learning

Akriti Jaiswal, A. Krishnama Raju, Suman Deb
Department of Electronics Engineering
SVNIT Surat, India
{nainajaiswal96, krishnamraju995}@gmail.com, sumandeb@eced.svnit.ac.in

***Abstract*—Human Emotion detection from image is one of the most powerful and challenging research task in social communication. Deep learning (DL) based emotion detection gives performance better than traditional methods with image processing. This paper presents the design of an artificial intelligence (AI) system capable of emotion detection through facial expressions. It discusses about the procedure of emotion detection, which includes basically three main steps: face detection, features extraction, and emotion classification. This paper proposed a convolutional neural networks (CNN) based deep learning architecture for emotion detection from images. The performance of the proposed method is evaluated using two datasets Facial emotion recognition challenge (FERC-2013) and Japaness female facial emotion (JAFFE). The accuracies achieved with proposed model are 70.14 and 98.65 percentage for FERC-2013 and JAFFE datasets respectively.**

***Index Terms*—Artificially intelligence (AI), Facial emotion recognition (FER), Convolutional neural networks (CNN), Rectified linear units (ReLu), Deep learning (DL).**

I. Introduction

Emotion is a mental state associated with the nervous system associated with feeling, perceptions, behavioral reactions, and a degree of gratification or displeasure [1]. One of the current application of artificial intelligence (AI) using neural networks is the recognition of faces in images and videos for various applications. Most techniques process visual data and search for general pattern present in human faces in images or videos. Face detection can be used for surveillance purposes by law enforcers as well as in crowd management. In this paper, we present a method for identifying seven emotions such as anger, disgust, neutral fear, happy, sad, and surprise using facial images. Previous research used deep-learning technology to create models of facial expressions based on emotions to identify emotions [2]. The typical human computer interaction (HCI) lacks users emotional state and loses a great deal of information during the process of interaction. Comparatively, users are more efficient and desired by emotion-sensitive HCI systems [3]. Now a days, interest in emotional computing has increased with the increasing demands of leisure, commerce, physical and psychological well being, and education related applications. Because of this, several products of emotionally sensitive HCI systems have been developed over the past several years, although the ultimate solution for this research field has not been suggested [4].

II. Related Work

A. Human Facial Expressions

The universality of facial expressions and the body language is a key feature of human interaction. Charles Darwin already published on globally common facial expressions in the nineteenth century, which play an important role in nonverbal communication [4]. In 1971, Ekman Friesen declared facial behaviors to be correlated uniformly with specific emotions [5]. Apparently humans but also animals, produce specific muscle movements that belong to a certain mental state. People interested in research on emotion classification via speech recognition are referred to Nicholson et al. [6].

B. Image Classification Techniques

Image classification system generally consists of feature extraction followed by classification stage. Fasel and Luettin provided an extensive overview of the analytical feature extractors and neural network approaches for recognition of facial expression [7]. It may be concluded that both approaches work approximately equally well by the time of writing, at the beginning of the twenty-first century. However, given the current availability of training data and computational power, the expectation is that the performance of models based on neural networks can be substantially improved. Several recent milestones are set out below.

- Krizhevsky and Hinton give a landmark publication on the automatic image classification in general [8]. This work shows a deep neural network that resembles the human visual cortex's functionality. A model to categorize objects from pictures is obtained using a self-developed labeled array of 60,000 images over 10 classes, using the CIFAR-10 dataset. Another important outcome of the research is the visualization of the filters in the network, so that how the model breaks down the images can be assessed.
- In 2010, the launch of the annual Imagenet challenges [9] boosted work on the classification of images, and since then the belonging gigantic collection of labeled data is often used in publications . In a later work by Krizhevsky et al. [10], the ImageNet LSVRC-2010 contest trains a network of 5 convolutional, 3 max pooling, and 3 fully connected layers with 1.2 million high-resolution images.
- In particular, with regard to facial expression recognition, Lv et al. [11] present a network of deep beliefs

978-1-7281-6221-8/20/$31.00 ©2020 IEEE

primarily for use with the JAFFE and expanded Cohn-Kanade (CK+) databases. The results are comparable to the accuracy obtained 95 percentage on the same database by other methods, such as support vector machine (SVM) and learning vector quantization (LVQ).

- The dataset used now is the Facial Expression Recognition Challenge (FERC-2013) [2], organized deep network, obtained an average accuracy of 67.02 percentage on emotion classification.

III. Proposed Model

A. Emotion Detection Using Deep Learning

In this paper we use the deep learning (DL) open library "Keras" provided by Google for facial emotion detection, by applying robust CNN to image recognition [12]. We used two different datasets and trained with our proposed network and evaluate its validation accuracy and loss accuracy. Images extracted from given dataset which have facial expressions for seven emotions, and we detected expressions by means of an emotion model created by a CNN using deep learning. We have changed a few steps in CNN as compared to previous method using a keras library given by Google and also modified CNN architectutre which give better accuracy . We implemented emotion detection using keras with the proposed network.

B. CNN Architecture

The networks are program on top of keras, operating on Python, using the keras learn library. This environment reduces the code's complexity, since only the neuron layers need to be formed, rather than any neuron. The software also provides real-time feedback on training progress and performance, and makes the model after training easy to save and reuse. In CNN architecture initially we have to extract input image of 48*48*1 from dataset FERC-2013. The network begins with an input layer of 48 by 48 which matches the input data size parallelly processed through two similar models that is functionality in deep learning, and then concatenated for better accuracy and getting features of images perfectly as shown in Fig.1 which is our proposed model, **Model-A**. There are two submodels for the extraction of CNN features which share this input and both have same kernel size. The outputs from these feature extraction sub-models are flattened into vectors and concatenated into one long vector matrix and transmitted to a fully connected layer for analysis before a final output layer allows for classification.

This models contains convolutional layer with 64 filters each with size of [3*3], followed by a local contrast normalization layer, maxpooling layer, followed by one more convolutional layer, max pooling, flatten respectively. After that we concatenate two similar models and linked to a softmax output layer which can classify seven emotions. We use dropout of 0.2 for reducing over-fitting. It has been applied to the fully connected layer and all layers contain units of rectified linear units (ReLu) activation function.

First we are passing our input image to convolutional layer which consists of 64 filters each of size 3 by 3, after that it passes through local contrast normalization can remove average from neighbourhood pixels leads to get quality of feature maps, followed by ReLu activation function. Maximum pooling is used to reduce spatial dimension reduction so processing speed will increase. We are using concatenation for getting features of images (eyes, eyebrows, lips, mouth etc) perfectly so that prediction accuracy improved as compared to previous model. Furthermore, it is followed by fully connected layer and softmax for classifying seven emotions. A second layer of maxpooling is added to reduce the number of dimensionality. Here, we use batch normalization, dropout, ReLu activation function, categorical cross entropy loss, adam optimizer, softmax activation function in ouput layer for seven emotion classification.

In JAFEE dataset, input image size is adjusted to that 128*128*3. The network starts with an input layer of 128 by 128 which matches the input data size parallely processed through two similar models as shown in Fig.1. Furthermore, it is concatenated and pass through one more softmax layer for emotion classification and all procedure is same as above.

In **Model-B**, previously proposed by Correa et al. [2], the network starts with a 48 by 48 input layer, which matches the size of the input data. This layer is preceded by one convolutional layer, a local contrast normalization layer, and one layer of maxpooling, respectively. Two more convolutional layers and one fully connected layer, connected to a softmax output layer, complete the network. Dropout has been applied to the fully connected layer and all layers contain units of ReLu.

IV. Experiment Details

We develop a network based on the concepts from [12], [13] and [14] to assess the two models (**Model-A** and **Model-B**) mentioned above on their emotion detection capability. This section describes the data used for training and testing, explains the details of the used data sets and evaluates the results obtained using two different datasets with two models.

A. Datasets

Neural networks, and particularly deep networks, needs large amounts of training data. In addition, the choice of images used for the training is responsible for a large part of the eventual model's performance. It means the need for a data set that is both high quality and quantitative. Several datasets are available for research to recognize emotions, ranging from a few hundred high resolution photos to tens of thousands of smaller images. The two, we will be debating in this work, are the Japanese Female Face Expression (JAFFE) [15], Facial Expression Recognition Challenge (FERC-2013) [16] which contains seven emotions like anger, surprise, happy, sad, disgust, fear, neutral.

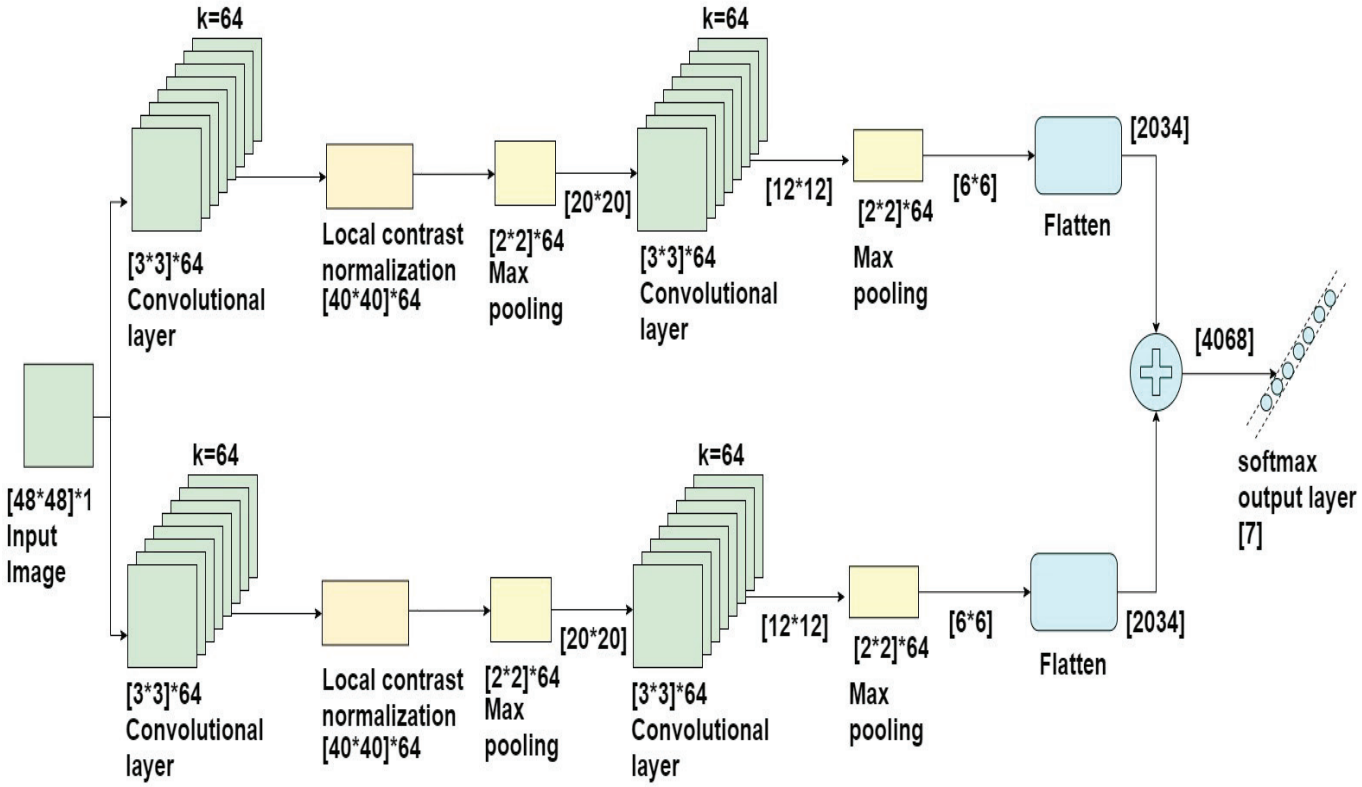

Fig. 1: Network Architecture

The datasets primarily vary in the amount, consistency, and cleanness of the images. For example, the FERC-2013 collection has about 32,000 low-resolution images. It can also be noted that the facial expressions in the JAFFE (i.e. further extended as CK+) are posed (i.e. clean), while the FERC-2013 set displays "in the wild" emotions. This makes it harder to interpret the images from the FERC 2013 set, but given the large size of the dataset, a model's robustness can be beneficial for the diversity.

B. Training Details

We train the network using GPU for 100 epochs to ensure that the precision converges to the optimum. The network will be trained on a larger set than the one previously described in an attempt to improve the model even more. Training will take place with 20,000 pictures from the FERC-2013 dataset instead of 9,000 pictures. The FERC-2013 database also uses newly designed verification (2000 images) and sample sets (1000 images). It shows number of emotions in the final testing and validation set after training and testing our model. The accuracy will be higher on all validation and test sets than in previous runs, emphasizing that emotion detection using deep convolutional neural networks can improve the performance of a network with more information.

C. Results using Proposed Model

In emotion detection we are using three steps, i.e., face detection, features extraction and emotion classification using deep learning with our proposed model which gives better result than previous model. In the proposed method, computation time reduces, validation accuracy increases and loss also decreases, and further performance evaluation achieved which compares our model with previous existing model. We tested our neural network architectures on FERC-2013 and JAFFE database which contains seven primary emotions like sad, fear, happiness, angry, neutral, surprised, disgust.

Fig.2 shows the proportions of detected emotions in a single image of FER dataset. Fig.2(a) shows the image, whereas the detected emotion proportions are shown in Fig.2(b). It is clearly observable that neutral has higher proportion than other emotions. That means, the emotion detected for this image (in Fig.2(a)) is neutral. Similarly, Fig.3 show another image and corresponding emotion proportions. From Fig.3(b), it is observable that happy emotion has higher proportion than others. That suggests that image of Fig.3(a) detects happy emotions.

Similarly, performance is evaluated for all the test images of the dataset. We have achieved 95 percentage for happy, 75

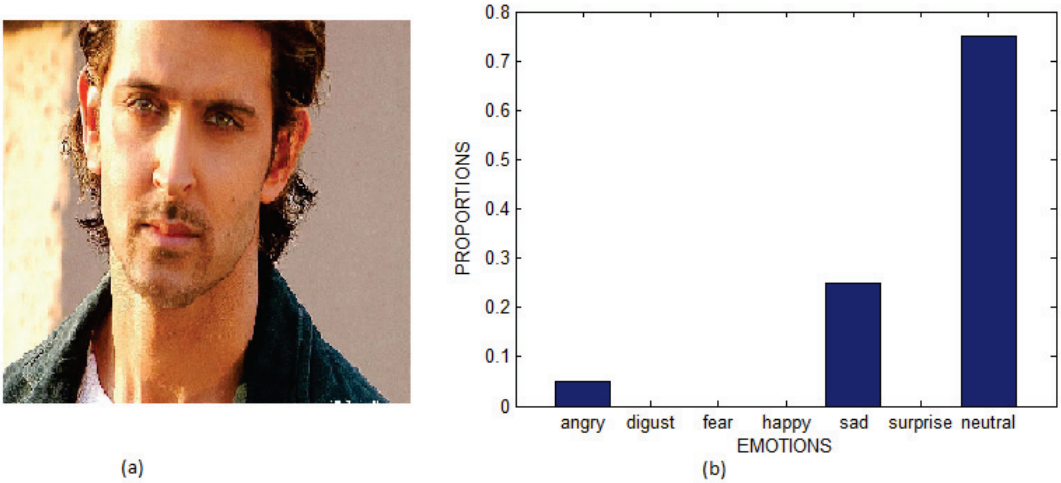

Fig. 2: (a) Image, (b) Proportion of emotions.

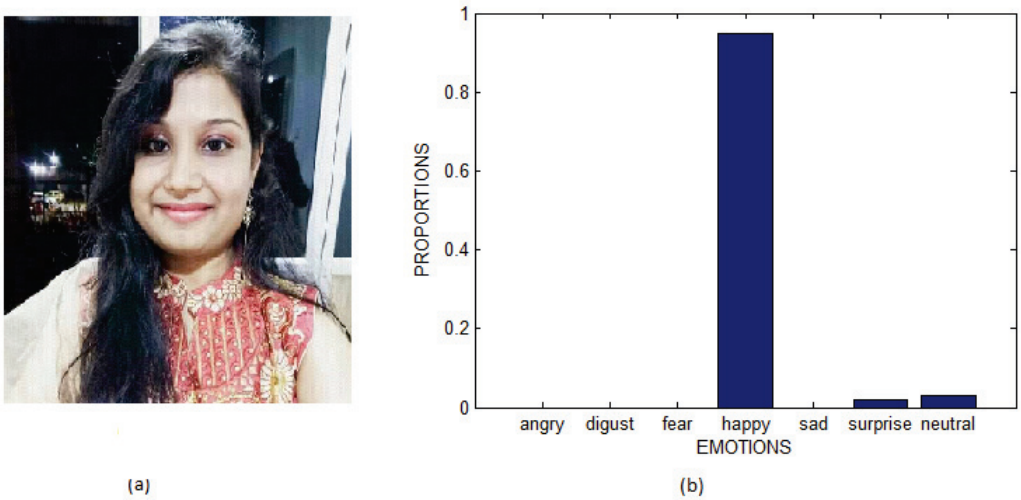

Fig. 3: (a) Image, (b) Proportion of emotions.

percentage for neutral, 69 percentage for sad, 68 percentage for surprise, 63 percentage for disgust, 65 percentage for fear and 56 percentage for angry. On an average we are getting average accuracy of 70.14 percentage using our proposed model.

The confusion matrix of classification accuracy is shown in TABLE I. We get an average validation accuracy of 70.14 percentage using our proposed model in facial emotion detection using FER dataset.

Fig.4 shows the result of test sample related to surprise emotion from JAFFE dataset, and our proposed model also predicted the same emotion with reduced computation time as compared to previous existing model B. Similarly, performance is evaluated for all the test samples of JAFFE dataset. When we are using JAFFE dataset we are getting validation accuracy of 98.65 percentage which is better than previous result and it takes less computational time per step.

Fig. 4: Prediction of emotion.

V. PERFORMANCE EVALUATION

In FER dataset we train on 32,298 samples which is validate on 3589 samples, and in JAFFE dataset we train 833 samples, which is validate on 148 samples for calculation of validation accuracy, validation loss, computational time per step upto to

TABLE I: Confusion Matrix (%) for emotion detection using proposed model

Emotions	Angry	Sad	Happy	Disgust	Fear	Neutral	Surprise
Angry	**56**	12	3	9	8	11	1
Sad	10	**69**	2	6	9	2	2
Happy	0	0	**95**	0	0	3	2
Disgust	7	13	0	**63**	8	5	4
Fear	9	8	3	2	**65**	10	3
Neutral	2	1	8	1	7	**75**	6
Surprise	7	3	11	0	3	8	**68**
Average accuracy = 70.14 (%)							

TABLE II: Qualitative assessment of our proposed model for emotion detection

Model (Dataset)	**Validation accuracy (%)**	**Validation loss**	**Computation time per step (msec.)**
Proposed Model A (FERC-2013)	70.14	1.7577	16 msec.
Model B (FERC-2013)	67.02	2.0389	45 msec.
Proposed Model A (JAFFE)	98.65	0.1694	284 msec.
Model B (JAFFE)	97.97	0.1426	462 msec.

100 and 50 epochs respectively shown in TABLE II. The aim of the training step is to determine the correct configuration parameters for the neural network which are: number of nodes in the hidden layer (HL), rate of learning (LR), momentum (Mom), and epoch (Ep). Different combinations of these parameters have been tested to find out how to achieve the better recognition rate.

From Table II, it is observed that our proposed model shows 70.14% average accuracy compared to the 67.02% average accuracy reported in model B FOR FER dataset. In this case of JAFFE database, we achieved average accuracy 98.65% which is also higher than model B.

VI. Conclusion

In this paper, we have proposed a deep learning based facial emotion detection method from image. We discuss our proposed model using two different datasets, JAFFE and FERC-2013. The performance evaluation of the proposed facial emotion detection model is carried out in terms of validation accuracy, computational complexity, detection rate, learning rate, validation loss, computational time per step. We analyzed our proposed model using trained and test sample images, and evaluate their performance compare to previous existing model. Results of the experiment show that the model proposed is better in terms of the results of emotion detection to previous models reported in the literature. The experiments show that the proposed model is producing state-of-the-art effects on both two datasets.

References

[1] S. Li and W. Deng, "Deep facial expression recognition: A survey," arXiv preprint arXiv:1804.08348, 2018.

[2] E. Correa, A. Jonker,M.Ozo, andR.Stolk, "Emotion recognition using deep convolutional neural networks," Tech. Report IN4015, 2016.

[3] Y. I. Tian, T. Kanade, and J.F.Cohn, "Recognizing action units for facial expression analysis," IEEE Transactions on pattern analysis and machine intelligence, vol. 23, no. 2, pp. 97–115, 2001.

[4] C. R. Darwin. The expression of the emotions in man and animals. John Murray, London, 1872.

[5] P. Ekman and W. V. Friesen. Constants across cultures in the face and emotion. Journal of personality and social psychology, 17(2):124, 1971.

[6] J. Nicholson, K. Takahashi, and R. Nakatsu. Emotion recognition in speech using neural networks. Neural computing applications, 9(4): 290–296, 2000.

[7] B. Fasel and J. Luettin. Automatic facial expression analysis: a survey. Pattern recognition, 36(1):259–275, 2003.

[8] A. Krizhevsky and G. Hinton. Learning multiple layers of features from tiny images, 2009.

[9] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pages 248–255. IEEE, 2009.

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pages 1097–1105, 2012.

[11] Y. Lv, Z. Feng, and C. Xu. Facial expression recognition via deep learning. In Smart Computing (SMARTCOMP), 2014 International Conference on, pages 303–308. IEEE, 2014.

[12] TFlearn. Tflearn: Deep learning library featuring a higher-level api for tensorflow. URL http://tflearn.org/.

[13] Open Source Computer Vision Face detection using haar cascades. URL http://docs.opencv.org/master/d7/d8b/tutorialpyfacedetection.html.

[14] P. J. Werbos et al., "Backpropagation through time: what it does and how to do it," Proceedings of the IEEE, vol. 78, no. 10, pp. 1550–1560, 1990.

[15] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on, pages 94–101. IEEE, 2010.

[16] Kaggle. Challenges in representation learning: Facial expression recognition challenge, 2013.