Facial Expression Recognition using Deep Learning Techniques

Paras Singh
Student
Department of Information Technology
Maharaja Agrasen Institute of Technology, Rohini, Delhi, India

Abstract— Our facial expressions play an important role in everyday communication between people. This automatic detection of facial expressions has long been studied for potential applications in various fields such as service robots, driver fatigue monitoring, and intelligent training systems. With the advent of human-computer interaction (HCI) systems such as social robots, visual interactive games, and data-driven animations, facial expression recognition (FER) has become a popular research area in recent years. Facial expressions convey 55% of messages conveyed, more than those conveyed by a combination of voice and language [5]. Face, voice, EEG and even text can be used to perform emotion recognition. Of these properties, facial expression is one of the most important for a number of reasons. It contains many useful features for recognizing emotions, it stands out, and it is easy to collect a large set of face data compared to other emotion recognition features. Facial expressions can be divided into six categories: anger, disgust, fear, surprise, sadness, and happiness. More recently, deep learning, especially Convolutional Neural Networks (CNNs), can be used to extract and train many features for proper face recognition systems. However, most of the clues come from different parts of the face, such as the mouth, nose, and eyes, while other parts, such as the hair, ears, and forehead, play a small role in the output. This means that, ideally, the machine learning system should focus only on the important parts of the face and be less sensitive to other areas of the face. In this paper, we propose a deep learning-based facial expression recognition framework that considers only important facial features and ignores other non-critical areas of the face.

Keywords: Facial Expression Recognition, Deep Learning Techniques

I. INTRODUCTION

Emotions are an important part of communicating with others. He doesn't say anything and gives us clues about our current state of mind. Facial expression recognition (FER) has become an active research area due to its applications in medicine, e-learning, monitoring, entertainment, marketing, human-computer interaction, and more. Therefore, it is necessary to develop a mechanism to detect emotions. Traditionally, manual functions have been used in conjunction with machine learning algorithms to solve this problem. However, the recent success of convolutional neural networks (CNNs) in deep learning, particularly in tasks such as object recognition, face recognition, and object detection, has prompted researchers to investigate these techniques in the field of expression recognition [1][2]. Despite the excellent results obtained, reliable face recognition remains a challenge for existing deep learning techniques because in vivo images vary significantly with poses, backgrounds, etc. For emotion recognition tasks that make deep network learning difficult, it is necessary to develop a system that can accurately determine a person's emotional state under given constraints. In this study, we aim to artificially increase the size of the data set using data augmentation techniques [4]. Motivated by the fact that human observers pay close attention to where representations are most common, they decide to focus only on the main parts of the image and ignore extraneous parts such as background detailsas they contribute little or no information [3].

II. RELATED WORK

One of the first studies on emotion recognition was done by Ekman [6]. They identified six basic emotions: happiness, sadness, anger, surprise, fear, and disgust. This work laid the groundwork for all other future work in the field of emotion recognition. Most of the initial work on facial expressions involved a two-step machine learning process. In the first step, facial features are extracted by human or computer software, and in the second step, emotions are detected using classifiers such as SVM or random forest. Some of the traditional and well-known approaches to extracting features from images are Gabor wavelets, direction gradient histograms, Haar features, etc. Although these techniques worked well for limited data sets, they were not able to generalize to larger data sets and data sets. change.

The recent success of convolutional neural networks for deep learning, particularly object recognition and other visual problems, has led several researchers to develop deep learning-based facial expression recognition (FER).

Several deep learning approaches for face recognition, especially CNN methods, have been developed over the past few decades. Some recent techniques focus on advanced network building and model training, multi-structure fusion and selection of fusion parameters, and optimization of classification algorithms.

Mayya et al. [7] proposed a Deep Convolutional Neural Network (DCNN) approach to identify facial expressions. They extracted facial features using ImageNet, a popular DCNN architecture. The last layer of the network provided the dimension vectors we used to test it with a support vector machine (SVM) classifier to recognize facial expressions. Their experiments were performed with two separate databases, CK+ and JAFFE, and achieved accuracies of 96.02% and 98.12% for the 7 grades, respectively. Despite their competitive accuracy, their approach is not end-to-end and is difficult and time consuming. Zhang et al. [8] proposed a well-designed novel CNN that can minimize the same emotional change during exercise and maximize the difference in movement. They trained the model using the 2-way softmax function, which requires a high level of knowledge and skill from the researcher. However, their method is for smile detection and can use 4000 images for a single expression, which is much more than the database available for FER.

All of the above tasks are a significant improvement over the existing emotion recognition tasks, but they seem to

lack a simple part to focus on the important parts of the face for emotion detection. In this work, we will try to solve this problem and focus on the most prominent parts of the face.

A. Convolutional Neural Networks

Convolutional neural networks (CNNs) are multilayer neural network architectures [19]. CNN inputs and outputs are array vectors called feature maps. The size of the array depends on the input type. For example, audio inputs include one-dimensional arrays and text inputs. I have a 2D array in my image. The output object map describes the objects retrieved from the input. CNN consists of three main layers: Convolutional Filter Layer, Pool/Sub-sampling Layer, and Classification Layer.

B. Facial Expression Recognition Methods

The major drawback of the function-based approach is that it requires a lot of effort to design and use various human-generated function extraction methods. To overcome these shortcomings, we propose a new approach based on deep learning, a machine-generated function that automatically extracts facial features.

III. PROPOSED METHOD

In this paper, we propose an automatic facial expression recognition method using DCNN. DCNN is a powerful model that can efficiently gather information, especially images. The proposed method consists of 3 steps:

1) pre-processing,
2) training, and
3) prediction.

These steps to recognize facial expressions using the proposed method are shown in Figure 1 and described in Algorithm 1.

Algorithm 1

A. Pre-processing:

- Face detection using the Viola-Jones standard method and rescaling the face image to 140×140 pixels.
- Finding the center of face image using Canny edge detection and extracting ROI image.
- Adding the average intensity of ROI image to it for normalization.

B. Training:

- The proposed DCNN is trained using 70% of normalized ROI dataset images that are rotated to ±45 and ±75 degrees.

C. Prediction:

- The facial emotion of the input image is extracted using pre-trained DCNN.

1) Pre-processing

One of the most difficult problems in machine learning is the problem of overfitting when using small data sets [24]. Most facial expression data sets are small. To overcome the overfitting problem, which is an automatic facial expression recognition method using 155 deep convolutional neural networks, for each image in the data set, each image was rotated clockwise and counterclockwise (±45,±75) at 45°, 75° Rotate to create 4 rotated images. As a result, the number of images in the dataset increases by a factor of 5 (4 rotated images and originals. As in Algorithm 1, face detection is performed using the Viola Jones method as the standard detection method in the preprocessing step. / Faces and faces) Extract Area Extract Face Area [25] This face area will be adjusted to an image of 140 × 140 pixels. Then clever edge detection is done to get the effective area of the face. Find the center row and column of detected edge pixels as the center of the face region image, called AvgRow and AvgCol, respectively. The effective face area is then determined by selecting the pixels from AvgRow60 to AvgRow + 60 in the row and AvgCol50 to AvgCol + 50 in the column of the input image and is called the instance area (ROI). So for this step we have a smaller image of 121x111 pixels. Smaller input images speed up learning. After that, the average ROI is added to the image to normalize it. It is then rotated ±45 and ±75 degrees, resulting in four rotated versions of the normalized ROI image.

2) Training

The DCNN of the proposed method consists of 3 layers, 2 convolutional layers and 1 fully connected layer. Each convolutional layer is followed by a ReLU layer, a max pooling layer, and a regularization level. A fully connected layer is a 7way class predictor that references 7 different facial expressions. DCNN training of the proposed method is performed using 70% of the dataset images. The DCNN input layer is the same size as the 121 × 111 ROI image. The optimization of the proposed network is performed using stochastic gradient descent (sgdm). Training starts at a learning rate of 0.1 and decreases by a factor of 10 when there is no improvement in the accuracy results of the validation set. Architecture of the proposed deep convolutional neural network

3) Prediction

ROI of the test image is extracted using the Viola Jones object detection method and then the extracted ROI image is created using the pre processing step and then it is fed to the pre-trained DCNN to predict its expression.

IV. RESULTS

In this chapter we have described the results we have obtained on the FER2013 dataset and the Jaffe dataset using our deep learning model and using image cropping using OpenFace.

A. FER 2013 Dataset

Fer2013 contains about 30,000 RGB images of faces with different expressions, limited to 48 × 48, with default designations of 7 types: 0 = anger, 1 = disgust, 2 = fear, 3 = happy, 4 = sad. You can share. , 5 = surprise, 6 = neutral. The Disgust representation has at least 600 images, while the other labels contain around 5000 samples each.

B. JAFFE Dataset

The Japanese Female Facial Expressions (JAFFE) dataset contains 213 female facial expressions corresponding to 10 different subjects. Each image is saved with a resolution of 256 × 256 pixels and an 8-bit gray level. Each subject in the dataset is represented by 7 expression categories (Neutral, Happy, Sad, Fear, Anger, Disgust, Surprise), and each subject has 2-4 images per expression. The network is trained for classification using a training subset of 70%, while a test

subset of 30% is used to test the likelihood that a given face image belongs to a particular facial expression class. Average recognition accuracy is used to evaluate network performance. Table 2 shows the recognition accuracy of the proposed method using the training weights obtained with the highest accuracy. The recognition accuracy achieved with this method is 98.59%. The proposed method provides high recognition accuracy of 100% for surprise, disgust, neutrality, and fear, whereas it gives a low accuracy of 96.77% for anger, happiness, and sadness.

Serial No.	Dataset	Algorithm	Accuracy
1.	FER2013	CNN Network	67.5%
2.	Jaffe	CNN Network	78.1%
3.	Jaffe	CNN + Landmark Detection	87.5%

Table 4.1 Accuracies obtained for different models and different datasets

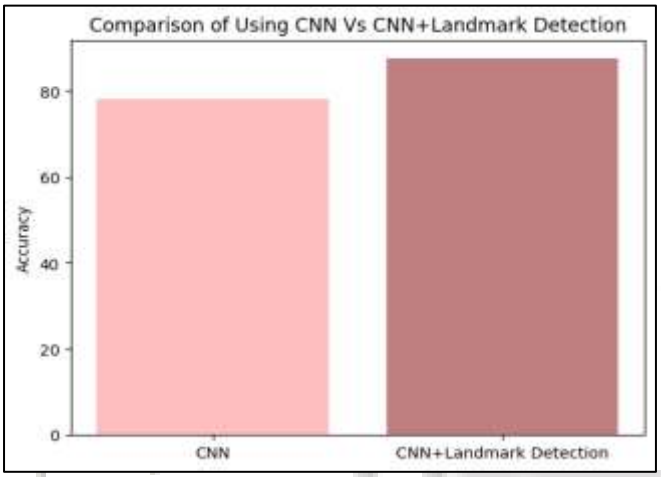

Fig. 4.1: Comparison of the two approaches

We only applied landmark detection to the Jaffe dataset because the image was large and sharp enough to detect the landmarks. On the other hand, for the FER2013 dataset, the images were small and could not be captured by the landmark detection network as shown in Table 4.1. Fig. 4.1 compares two approaches using convolutional neural networks and convolutional neural networks in addition to finding landmarks.

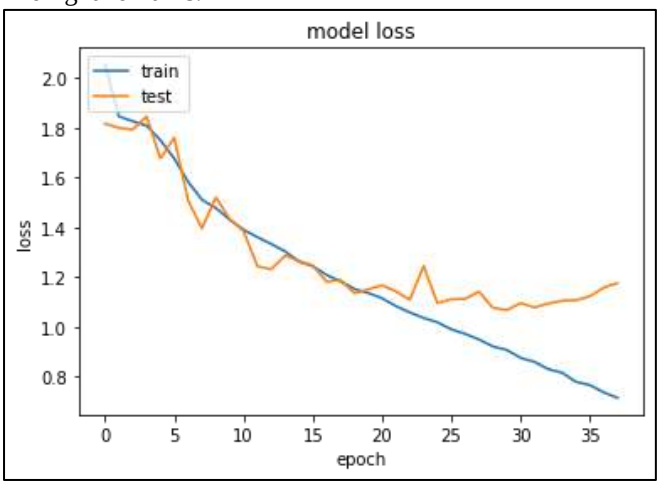

Fig. 4.2: Model Loss Train Vs Test Experiment 1

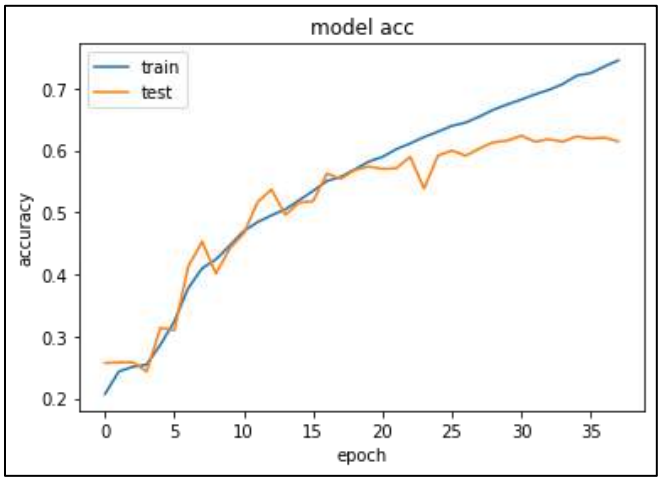

Fig. 4.3: Model Accuracy Train Vs Test Experiment 1

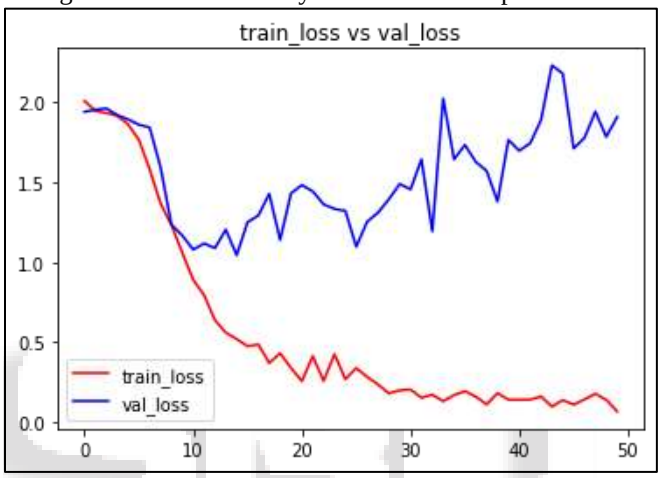

Fig. 4.4: Model Loss Train Vs Test Experiment 2

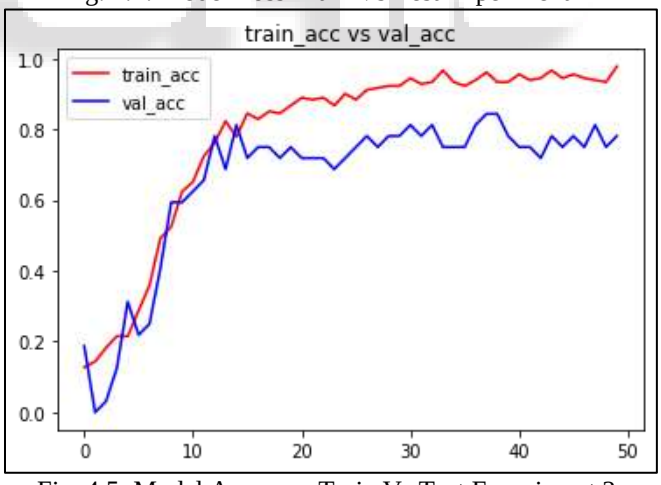

Fig. 4.5: Model Accuracy Train Vs Test Experiment 2

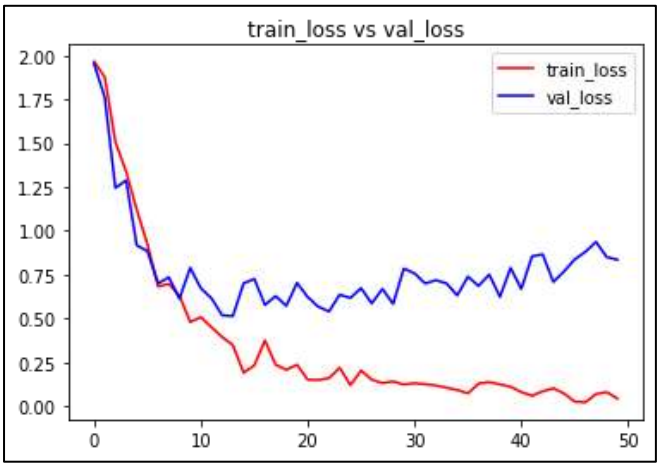

Fig. 4.6: Model Loss Train Vs Test Experiment 3

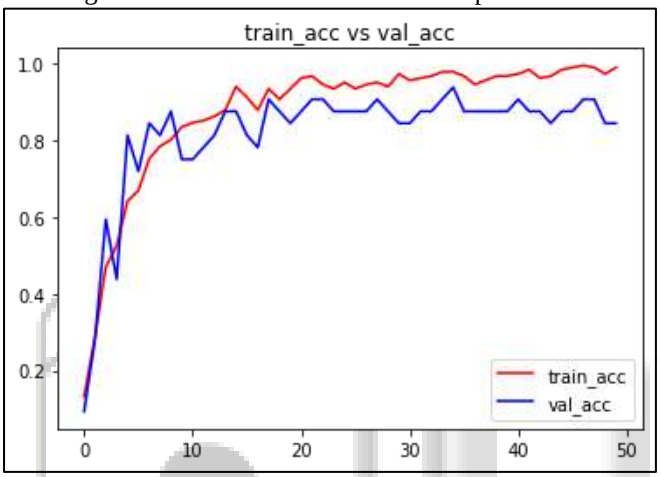

Fig. 4.7: Model Accuracy Train Vs Test Experiment 3

The six diagrams shown on the previous page show the model loss and model accuracy as the number of epochs increases. Fig.s 4.2 and 4.3 show the model loss and accuracy of the first experiment. Fig.s 4.4 and 4.5 show the loss and model accuracy of the second experiment. Fig.s 4.6 and 4.7 show the model loss and accuracy of the third experiment.

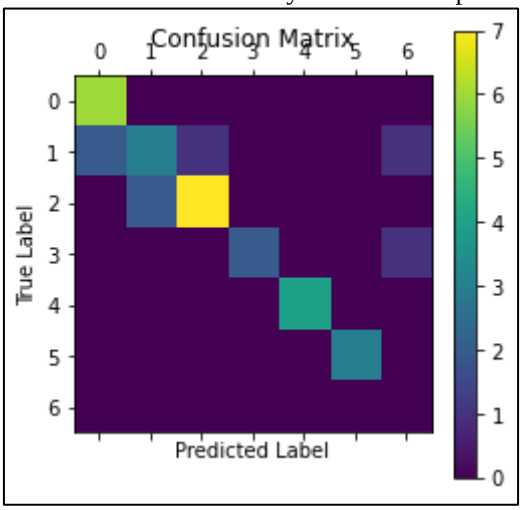

Fig. 4.8: Confusion matrix for experiment 2

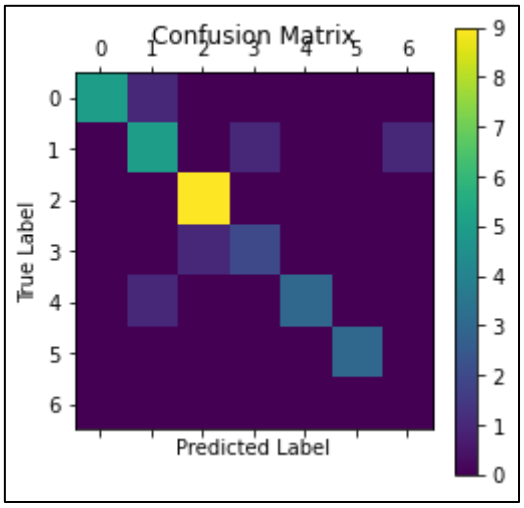

Fig. 4.9: Confusion matrix for experiment 3

Fig. 4.8 and Fig. 4.9 show the confusion matrices for the seven classes of emotions for experiment 2 and experiment 3 respectively.

V. Conclusion and Future Work

In this project, we are solving the problem of facial expression recognition in FER2013 and Jaffe datasets. We propose an efficient approach to FER using facial landmark detection to remove extraneous elements from images such as ears, hair, neck, background, etc., which contain little or no information. We evaluated the effectiveness of our model and the proposed approach, showing that the convolutional neural network and the learning process benefit from the landmark detection step. The results show that the proposed FER approach achieves competitive results in terms of training time, test time, and recognition accuracy. Also, the proposed method can be implemented on a general computer without GPU acceleration.

A. Future Work

- Modify the existing convolutional neural network and make it more sophisticated.
- Use attention convolution network, which can focus on feature-rich parts of the face.
- Region attention networks for tackling pose and occlusion.

References

[1] Khorrami, Pooya, Thomas Paine, and Thomas Huang. "Do deep neural networks study facial action units when recognizing facial expressions?" IEEE International Conference on Computer Vision Seminars. 2015.

[2] Han, Shizhong, Zibo Meng, Ahmed Shehab Khan, and Yang Tong. "Incrementally Accelerated Convolutional Neural Networks for Per-Face Recognition." Advances in Neural Information Processing Systems, pp. 109117. 2016.

[3] Minaee, Shervin, and Amirali Abdolrashidi. "Deepemotion: Facial Expression Recognition Using Convolutional Attention Networks." Preprinted arXiv arXiv: 1902.01019 (2019).

[4] Li, Kuan, Yi Jin, Muhammad Wakar Akram, Ruiz Khan, and Junwei Chen. "Face Recognition Using Convolutional Neural Networks Using New Strategies for Framing and Rotating Faces." Visual computer 36, no. 2 (2020): 391404.
[5] Mehrabian, A.: Communication without words. Communication. Theory, 193-200 (2008)
[6] Ekman, Paul, and Wallace W. presentation. "Faces and emotions are constant in different cultures." Journal of Personality and Social Psychology 17, no. 2(1971): 124.
[7] Maya, Veena, Radhika M. Pai and M. M. Manohara Pai. "Automatic Recognition of Facial Expressions Using DCNN". Rulesia Computer Science 93 (2016): 453461.
[8] Zhang, Kaihao, Yongzhen Huang, Hong Wu, and Liang Wang. "Deep learning-based smile detection of faces." 2015 3rd IAPR Asian Pattern Recognition Conference (ACPR), p.534538. IEEE, 2015.
[9] Zadeh, Amir, Yao Chong Lim, Tadas Baltrusaitis and Louis Philippe Morenci. "Convolutional experts have limited local models for 3D facial landmarks." Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 25192528.2017.

