Facial Expression Recognition using Convolutional Neural Network and k-Nearest Neighbors

Shivani Patidar
M. Tech Scholar
Swami Vivekanand College of Engineering,
Indore, India
Shivanipatidar11@gmail.com

Pradeep Baniya
Assistant Professor
Swami Vivekanand College of Engineering,
Indore, India
pradeepbaniya55@gmail.com

Abstract: The importance of facial expression recognition in different applications. To do such task the traditional feature extraction is used which involves in complex processing. Previously various deep neural networks have been used for this task, as well it can be replaced by some improved methods. So, in this paper we have proposed a system for face recognition which uses the convolutional neural network (CNN) and KNN along with the detection of the edges of image. The proposed work flows in two steps in the first the normalization of the facial expression in the image is done, secondly the convolution is done for the extraction of the edges in images. After this the maximum pooling method is used for the dimensionality reduction. At the end the classification of the facial expression is done and the Softmax classifier is used for this classification and the face expression is recognized. The facial expression recognition experiment is done on the Fer-2013 dataset. The results obtained form the proposed approach gets a face recognition rate up to 92.45% for the used dataset. The given method works with lesser number of iterations for the recognition and the system provides approx 1.5 times fast execution as compared to the SDSRN algorithm.

Keywords: Facial Expression Recognition, Image Edge Detection, Convolutional Neural Network, Deep Learning, Maximum Pooling Classification.

I. INTRODUCTION

The interaction seen between the computer and the humans is because of the technology growth in this field. One of the most important field in this is, making the computer work like human. Recently as the growth seen in this field like pattern recognition and artificial intelligence for making the interaction more flexible [1]. The facial expression recognition is seen as the most important way to make any computer intelligent as human. The application of this technique can be seen in various security systems, gaming, distance education etc. This system of facial expression recognition works from extracting the features related to expression from a given set of input and training of the computer is done [2]. Then the input is provided for testing of the facial expression by providing the image to the computer. The expressions on the face like neutral, happiness, sad, surprise, aversion. The field of emotional quantification uses the most of this expression recognition scheme [3]. So, this all the study states that the facial expression recognition technique is very important to study and work on it for various purpose.

The traditional CNN method has some drawbacks as it takes long time for the training of the system. Where a complex background is present there it is having a low recognition capability, so there is a need to reduce such limitations of traditional CNN. Here we also gave a method to recognize the expressions seen in the image of any person. The presented approach uses the CNN method along with the edge detection of the image. Here we are going to extract the edges of input images for all layers present in it. We also used the max pooling method for the dimensionality reduction of the features extracted implicitly. The dataset used here for the experiment is Fer-2013 face expression dataset.

II. FACE EXPRESSION RECOGNITION

The facial expression recognition is a field of intelligence where the computer makes use of the tools of the computer and then the use of the algorithms is done for the recognition of the emotions present in the image [4]. The use of this scheme is seen in various fields such as medical, teaching etc [5]. The medical field uses this scheme for the recognition of the facial expressions of the patients for knowing the depression effect in the human or the drug evaluation can be improved to an extent. This field also uses this scheme for the children treatment and to understand their emotions. The teaching field also uses this technology to read the emotions of the students for the learning interest. The emotions of the students can be captured by using the live images and the student can be guided according. The traffic application of this technology can be used to understand the state of mind of the pilots and drivers by seeing their facial expressions. The use of this technology in daily life of human can make the interaction far better with computer and other application can also make use of this [6].

The facial expression recognition works for knowing the mood of the person present in the image or video [7]. The expression can be of many types such as happy, unhappy, sad, fear, disgust, crying, angry, tensed, surprised etc. All this expression can be studied from the image given as input. The evaluation of expression can be done by giving pre-training to the computer. The already set expressions can be evaluated from the computer by applying the desired algorithm to carry out the results.

Fig1. Facial Expression Recognition

Till now many machine learning algorithms have been applied for the proper recognition of the facial expressions [8]. The methods like SVM, R-CNN, CNN, HMM, ANN have been used and the application of the improved methods for this task is going on.

III. LITERATURE REVIEW

Here the authors have worked on the deep neural networks. Face adversarial effect in the recognition was discussed in the paper. The distortion in image can make the recognition wrong and to prevent this the algorithm was proposed by them. They have presented a face friend safe adversarial method to mis recognize by the enemy's recognition system.[9]

The authors presented their work on the single sample facial recognition. They have presented a method to solve this problem by using the robust features and the generation of the virtual sample. The proposed method by them was based on the multi scale support vector transformation. This method was used to generate virtual samples. They hve verified their method on three different datasets of face recognition.[10]

In this paper the human face recognition system is developed by the authors. They have worked on a robot that takes voice command for any person and after detecting that person the recognition is said to be complete. They have used the PCA and the eigenface.[11]

The authors have presented a face recognition system. They have first worked for the extraction of the face from the video or image and then the recognition is done. The proposed scheme is Bag of Word for feature extraction. Thay implemented their work on the MATLAB application and achieved accuracy rate of 99.21%.[12]

Here the authors have stated various classifiers for face recognition. The PCA is used for the feature extraction in the field of the face recognition. They have worked on the rate of the face recognition. They have used the dataset of ORL for training and the testing of the method.[13]

The authors here presented a scheme for the face recognition system that makes the use of discrete fourier transform along with the logistic chaotic map. They have used the DFT for the encryption of the domain for face recognition. The frequency domain was encrypted by them in their work. The obtained results were good.[14]

Authors here worked on the depth images as they provide an additional mechanism for the three-dimensional information. For a better work in depth image recognition, they have proposed a mechanism based on the CNN. The database used by them is IIITD. The comparison is made with the methods like HOG and LBP.[15]

Authors have presented a mechanism to overcome the recognition limitations seen in the surveillance videos. As this video are blur and not that clear to detect the real expression and body activities. So, the authors have presented a method named as Trunk branch ensemble CNN. They have achieved a higher accuracy in recognition.[16]

IV. PROPOSED APPROACH

4.1 The proposed approach can be carried out by pre-training of the model after that the desired experiments can be carried out. Here we have to carry out our pre-processing in following steps:

1. Find the face in image and cut that face. For this the Haar classifier is used.
2. Normalization of the face image is done for a specific size.
3. Histogram equalization of the image is done for reduction of extra data of image. Grey level equalization is done in this step.
4. Edge of image is extracted at last. Kirsch edge operator is used for this task.

Facial expression recognition based on CNN:

The CNN here used for recognition is feed forward neural network. That may do the feature extraction and network parameters have been optimized by this. There are three layers specifically that are convolution layer, pooling layer and the connective layer.

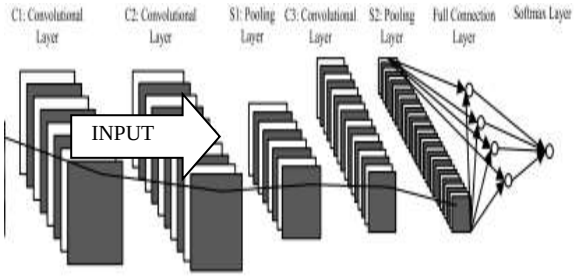


Figure 2. CNN Architecture used in Method

The C1, C2, C3 layers shown in the architecture makes use of the 32,64,128 conv nucli. Conv nuclei here is of 5*5 and the S1, S2 pooling layer of 2*2. Softmax layer used here have 7 number of neurons in it. While connection layer has 300

neurons that is connected to S2. This all layers are fully connected. We carried out a category of expressions that are fear, anger, happy, disgust, neutral, sad and surprise.

Convolution Layer: The image is filtered by convolution and C1 is obtained. We obtain S2 by feature mapping. Now on S2 we perform convolution and obtain the C3. By following the same S4 obtained. The equation used for computation are:

$$y_j^l = \theta(\sum_{i=1}^{N_j^{l-1}} w_{i,j} \otimes x_i^{l=1} + b_j^l), j = 1,2, \dots \dots, M \quad (1)$$

l and l-1 are present and previous layer, feature graph for jth and ith term for present and previous layer and Wij is kernel(convolution), Yj is jth feature graph, jth feature graph bias is Bj, feature maps total number is given by M, N is characteristic connected to present layer.

Pooling Layer: 2*2 pooling layer may perform 50% reduction in the dimensions. The Equation used for pooling is given below:

$$y_j^l = \theta(\beta_j^l down\left(y_j^{l-1}\right) + b_j^l) \quad (2)$$

The j, l, l-1 were as same stated previously. Y is down sampling and B is Bias for respective layers and feature maps.

Full Connection Layer: The previous S2 layer gives a 2D array and we need here 1D array. And after this 128 1D arrays are given as input. Here the feature vector is having 51200 dimensions for the full connection. The output obtained from every neuron is:

$$h_{w,b}(x) = \theta(w^T x + b) \quad (3)$$

Output value is h, eigen vector is x for the neurons as input. b is bias. Theta is used as the function for activation.

Softmax Layer: Here classifier is used as last layer that is softmax. This gives multiple outputs. The output obtained from every neuron is in between 0 and 1 for the input. This 0 to 1 is the probability of the belongings to the class. So, the output with highest probability is taken for classification.

4.2 Facial Expression Classificationalgorithm usingh KNN

Step1. Read the input face image.
Step2. Detect face from input face image.
Step3. If face detected, go to step 4. Otherwise go to step 1.
Step4. Divide a face image into different regions.
Step5. Extract LBP features. Step6. Feature classification (KNN method)
Step7. Recognize expression from classification.

4.2.1 K-Nearest Neighbor Algorithm

The simplest algorithm for identifying a sample from the test set is called the Nearest Neighbor method. The object of interest is compared to every sample in the training set, using a distance measure, a similarity measure, or a combination of measures. The unknown object is then identified as belonging to the same class as the closest sample in the training set. This is indicated by the smallest number if using a distance measure, or the largest number if using a similarity measure. This process is computationally intensive and not very robust.
We can make the Nearest Neighbor method more robust by selecting not just the closest sample in the training set, but by consideration of a group of close feature vectors. This is called the K-Nearest Neighbor method, where, for example, K = 5. Then we assign the unknown feature vector to the class that occurs most often in the set of K-Neighbors. This is still very computationally intensive, since we have to compare each unknown sample to every sample in the training set, and we want the training set as large as possible to maximize success.
We can reduce this computational burden by using a method called Nearest Centroid. Here, we find the centroids for each class from the samples in the training set, and then we compare the unknown samples to the representative centroids only. The centroids are calculated by finding the average value for each vector component in the training set.
The iteration will stop after K-subclasses were obtained and the initial K-vector will become the means of each sub-classes. Figure 4 will show that each class Ci will be divided into number of sub-classes Ci, j represented by the mean _i, j of the data.
According to the Euclidean distance formula, the distance between two points in the plane with coordinates (x, y) and (a, b) is given by equation (4)

$$dist((x,y),(a,b) = \sqrt{(x-a)^2 + (y-b)^2} \quad (4)$$

V. RESULTS EVALUATION

The Database used here for the facial expression recognition is the Fer-2013. The details related to the database is as follows.

- 35685 images
- 48*48-pixel grayscale images
- Categories: happiness, neutral, sadness, anger, surprise, disgust, fear

The Hardware used for the experiment is:

- Intel(R) core I7
- 64-bit OS
- Main frequency CPU of 3.2 GHz
- 8GB RAM

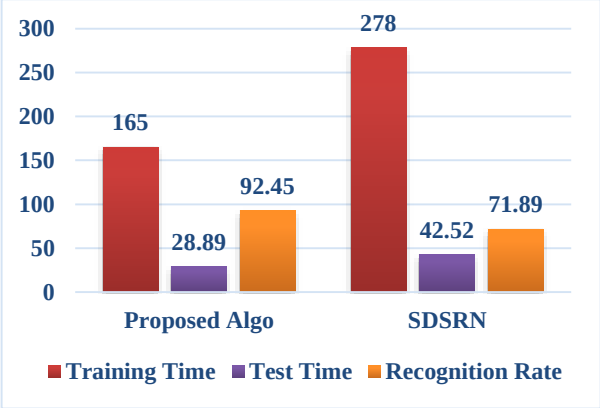
The proposed algorithm is compared with the algorithm that is Sparse-Deep Simultaneous Recurrent Network (SDSRN). Three parameters were analyzed in our experiment that are training time (s), test time (s), Recognition Rate (%).

METHOD	TRAINING TIME (s)	TEST TIME (s)	RECOGNITION RATE (%)
Proposed	165	28.89	92.45

Algorithm			
SDSRN	278	42.52	71.89

Table1. The obtained values for Experiment

The results obtained from the evaluation can be seen in the table shown above. The obtained values shows that we are having an efficient method for expression recognition. The Recognition rate is improved at a great extent. The training and testing time is found very less as compared to the existing algorithm like DSRN. The training time is improved by 113 seconds. Further explanation can be seen in graph given below.



Graph1. Obtained Results

The obtained results can be understood more clearly in graph. The improvement in the training time is by 113 seconds, while the test time is improved by 13.63 seconds. The recognition rate is also improved by 20.56. The experiment has found efficient and accurate for the face recognition.

VI. CONCLUSION

The face expression recognition has become one of the main tasks for various applications. This work is carried out on various algorithms and methods previously. But there is more improvement can be done so here we have presented a mechanism that uses CNN for the recognition of the expressions seen in the image. Also, we have used Softmax Classifier for the classification. The dataset used here is Fer-2013. The results were carried out on three parameters that are training time, test time and the recognition time. The recognition rate found here from the experiment is 92.45%. So, the proposed algorithm can be used for face expression recognition.

Reference:

[1] B. Knyazev, R. Shvetsov, N. Efremova, and A. Kuharenko, "Leveraging large face recognition data for emotion classification," *Proc. - 13th IEEE Int. Conf. Autom. Face Gesture Recognition, FG 2018*, pp. 692–696, 2018, doi: 10.1109/FG.2018.00109.

[2] J. Ueda and K. Okajima, "Face morphing using average face for subtle expression recognition," *Int. Symp. Image Signal Process. Anal. ISPA*, vol. 2019-Septe, pp. 187–192, 2019, doi: 10.1109/ISPA.2019.8868931.

[3] M. R. Reshma and B. Kannan, "Approaches on partial face recognition: A literature review," *Proc. Int. Conf. Trends Electron. Informatics, ICOEI 2019*, no. Icoei, pp. 538–544, 2019, doi: 10.1109/ICOEI.2019.8862783.

[4] P. Shakyawar, P. Choure, and U. Singh, "Eigenface method through through facial expression recognition," in Proceedings of the International Conference on Electronics, Communication and Aerospace Technology, ICECA 2017, 2017, vol. 2017-Janua, doi: 10.1109/ICECA.2017.8212714.

[5] K. Kushwah, V. Sharma, and U. Singh, "Neural network method through facial expression recognition," in Proceedings of the International Conference on Electronics, Communication and Aerospace Technology, ICECA 2017, 2017, vol. 2017-Janua, doi: 10.1109/ICECA.2017.8212721.

[6] Y. Mathur, P. Jain, and U. Singh, "Foremost section study and kernel support vector machine through brain images classifier," in Proceedings of the International Conference on Electronics, Communication and Aerospace Technology, ICECA 2017, 2017, vol. 2017-Janua, doi: 10.1109/ICECA.2017.8212726.

[7] V. S. Tomar, N. Gupta, and U. Singh, "Expressions recognition based on human face," in Proceedings of the 3rd International Conference on Computing Methodologies and Communication, ICCMC 2019, 2019, doi: 10.1109/ICCMC.2019.8819714.

[8] R. He, X. Wu, Z. Sun, and T. Tan, "Wasserstein CNN: Learning Invariant Features for NIR-VIS Face Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 7, pp. 1761–1773, 2019, doi: 10.1109/TPAMI.2018.2842770.

[9] H. Kwon, O. Kwon, H. Yoon, and K. W. Park, "Face Friend-Safe Adversarial Example on Face Recognition System," *Int. Conf. Ubiquitous Futur. Networks, ICUFN*, vol. 2019-July, pp. 547–551, 2019, doi: 10.1109/ICUFN.2019.8806124.

[10] X. Xu, L. Zhang, and F. Li, "MSSVT: Multi-scale feature extraction for single face recognition," *Proc. - Int. Conf. Pattern Recognit.*, vol. 2018-Augus, pp. 1996–2001, 2018, doi: 10.1109/ICPR.2018.8545343.

[11] A. L. Ramadhani, P. Musa, and E. P. Wibowo, "Human face recognition application using PCA and eigenface approach," *Proc. 2nd Int. Conf. Informatics Comput. ICIC 2017*, vol. 2018-Janua, pp. 1–5, 2018, doi: 10.1109/IAC.2017.8280652.

[12] S. Nasr, K. Bouallegue, M. Shoaib, and H. Mekki, "Face recognition system using bag of features and multi-class SVM for robot applications," *2017 Int. Conf. Control. Autom. Diagnosis, ICCAD 2017*, pp. 263–268, 2017, doi: 10.1109/CADIAG.2017.8075668.

[13] E. I. Abbas, M. E. Safi, and K. S. Rijab, "Face recognition rate using different classifier methods based on PCA," *Int. Conf. Curr. Res. Comput. Sci. Inf. Technol. ICCIT 2017*, pp. 37–40, 2017, doi: 10.1109/CRCSIT.2017.7965559.

[14] C. Liu, J. Li, and Y. Duan, "A face image recognition algorithm based on DFT encryption domain," *1st Int. Conf. Electron. Instrum. Inf. Syst. EIIS 2017*, vol. 2018-Janua, pp. 1–6, 2018, doi: 10.1109/EIIS.2017.8298669.

[15] J. Chen, Z. Zhang, L. Yao, B. Li, and T. Chen, "Face recognition using depth images base convolutional neural network," *CITS 2019 - Proceeding 2019 Int. Conf. Comput. Inf. Telecommun. Syst.*, pp. 6–9, 2019, doi: 10.1109/CITS.2019.8862099.

[16] C. Ding and D. Tao, "Trunk-Branch Ensemble Convolutional Neural Networks for Video-Based Face Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 1002–1014, 2018, doi: 10.1109/TPAMI.2017.2700390.