Chapter 36
Deep Neural Network for Facial Emotion Recognition System

Vimal Singh, Sonal Gandhi, Rajiv Kumar, Ramashankar Yadav, and Shivani Joshi

Abstract Facial emotion recognition is a challenging task to capture and analyze. This paper presents an intelligent approach to detect different facial expressions of a person. In this context, a noval deep learning vgg 19 convolutional neural network architecture is used. Seven different emotions (anger, disgust, fear, happiness, sadness, surprise, and neutral) are considered to recognize emotions. The FER 2013 dataset contains 35,887 different images with these emotions. Results were obtained by using various activation functions. A comparative study of various activation functions has been done.

36.1 Introduction

Intelligent Facial Expression Recognition is a vast and important research topic. Finding emotion from images is difficult and sensitive. The face is the index of our mind. Such a system can understand the situation better to take more fruitful results. The system will also be helpful in human-computer interaction. Since human-computer interaction, interacts with human in uncontrolled environment where the scene lighting, camera view, image resolution, background, user's head pose, gender, and ethnicity can vary significantly.

CNN (Convolutional Neural Network) is a type of neural network which makes an assumption that the input is an image. It contains series of hidden layers that transform the input to output wherein each hidden layer is made up of neurons, and each neuron is connected to all the neurons in previous layers. CNN mainly comprises three types of layers: convolution layer, pooling layer, fully connected layer.

1. Convolution layer
2. Pooling layer
3. Fully connected layer.

The various emotions used in this paper are shown in Fig. 36.1.

V. Singh (✉) · S. Gandhi · R. Kumar · R. Yadav · S. Joshi
G L Bajaj Institute of Technology and Management, Greater Noida, India

| Anger | Happy | Sad | Surprise | Neutral | Fear | Disgust |

Fig. 36.1 Facial emotion used

36.2 Related Work

Gupta et al. [1] have proposed identification of facial expression using CNN, ResNet, and attention block that gives visual perceptibility. The authors have proposed to use a deep self-attention network for facial emotion recognition. They found that the proposed model outperformed the current CNN-based networks by achieving a higher precision of 85.76. Mellouka and Handouzia [2] performed a survey over multiple research publications that employed deep learning techniques. Authors identified that today researchers are restricted by just knowing the six-basic plus neutral emotion rather than creating a larger database in the future. Mehendale [3] proposed a hybrid of CNN and supervised learning for the detection of facial expression. It has been observed that this model gives better results when it operates for different orientations. The correctness was achieved due to the removal in the background. Zhang et al. [4] performed the facial emotion recognition based on an image of facial expression. Authors use biorthogonal wavelet and fuzzy multiclass support vector machine. This work provides an overall accuracy of 96.7%. Jabid et al. [5] presented an approach for facial emotion by using the local directional pattern(LDP) that takes into consideration the facial geometry. The model also identifies the effectiveness of different reduction techniques such as PCA and AdaBoost in terms of cost and accuracy, which shows the superiority of the LDP descriptor to other features of the descriptors. Mollahosseini et al. [6] proposed a deep neural network architecture for the detection of facial expression which consists of two convolution layers and followed by pooling and then four layers of Inception. It also provides a comparison to several state-of-the-art methods in which engineered features and classifier parameters are usually tuned on a very few databases. The proposed method outperformed over conventional CNN methods in terms of accuracy on both subject independent and cross-database evaluation scenarios. Operto et al. [7] performed a survey to determine ER abilities in children and adults with learning difficulties with different learning disorders incorporated in the study without cognitive disabilities and relates to intelligence. The authors concluded that facial emotional recognition outlays are potentially related to difficulties in cognitive control. Shan et al. [8] presented an approach for facial emotion representation based on the conditions of Local Binary Patterns.

36.3   Methodology

The proposed methodology is implemented as follows:

1. Dataset: The dataset deployed for implementation was the FER2013 dataset from the Kaggle challenge on FER. The dataset contains 35,887 grayscale images out of which 32298 are for training purposes, 3589 for testing. Images in the FER2013 dataset comes under one of the seven categories, namely: neutral, happy, fear, surprise, disgust, angry, and sad.
   Emotion labels in the dataset: 0: −4593 images—angry, 1: −547 images—disgust, 2: −5121 images—fear, 3: −8989 images—happy, 4: −6077 images—sad, 5: −4002 images—surprise, 6: −6198 images—neutral. FER 2013 dataset samples are shown in Fig. 36.2.
2. Preprocessing: Resizing of the image into 48 * 48 grayscale images.
3. Grayscaling: Grayscaling is the process of transforming an RGV image input into a grayscale image whose pixel value from 0 to 255 upon the intensity of light on the image. As the pattern of an image does not depend on color and also the processing of color images requires more processing time and resources. Due to this reason grayscale images are used for processing.
4. Normalization: As neural networks are very sensitive to normalized data, normalization of an image is done to remove illumination variations and obtain an improved face image.
5. VGG19 Architecture:

Fig. 36.2 FER 2013 dataset samples

Fig. 36.3 Block diagram of the proposed method

We have conducted extensive experiments to demonstrate the proposed method's effectiveness compared to the most famous classification models, including VGG19 architecture. VGG19 architecture was developed by Simonyan and Zisserman of the University of Oxford with 19 layers, 16 conv, and 3 fully connected. 138 million parameters are eligible for VGG19. VGG19 can train on more than a million images and can classify into 1000 object categories. The detailed methodology is shown in Fig. 36.3.

Block Diagram for Proposed Model:

The VGG19 network consists of sixteen two-dimensional convolutional layers, five max-pooling layers, and three fully connected layers. Max pooling uses the maximum value from each of a cluster of neurons at the prior layer by using a 5 × 5 max-pooling filter. This reduces the dimensionality of the output array. The input to the network is a preprocessed face of 48 × 48 pixels.

36.4 Experimental Results and Discussion

After performing the experiment with the proposed methodology, results are discussed in this section. The experiments are performed on Intel core i5 8400 CPU @ 2.80 GHZ, and python 3. Keeping kernel size 5 * 5 results obtained are shown in Tables 36.1, 36.2 and 36.3. Based on these tables, it is clear that sample size is a very important factor in deep learning. Since the samples of disgust and surprise is low therefore accuracy is poor in these samples. Emotion with happiness performs better compared to others. All others have results that vary in between. Tables 36.1, 36.2, 36.3 and 36.4 show various confusion matrix and results.

Table 36.1 Confusion matrix with Elu activation function

	0	1	2	3	4	5	6
0	333	3	24	13	49	14	59
1	20	20	7	1	3	1	3
2	88	1	207	15	96	48	57
3	15	1	14	785	12	18	54
4	67	0	55	15	300	6	165
5	17	0	35	20	3	317	8
6	28	1	16	35	45	10	485

Table 36.2 Confusion matrix with ReLu activation function

	0	1	2	3	4	5	6
0	332	3	18	14	53	14	61
1	19	21	6	3	2	2	2
2	81	0	187	12	110	52	70
3	8	0	11	791	10	20	59
4	47	0	36	18	342	8	157
5	16	0	34	27	1	315	7
6	16	1	13	26	51	10	503

Table 36.3 Results of Elu and ReLu activation function

Sl. No.	Kernel size (5 * 5)	Activation function	Accuracy
1	5 * 5	Elu	0.682
3	5 * 5	ReLu	0.691

Table 36.4 Comparison with others

References	Methodology	Accuracy(%age)
Mollahosseini et al. [9]	CNN	66.4
Ergen et al. [10]	CNN (ReLu)	57.1
Proposed model	VGG-19 (ReLu, Elu)	69.1

The comparative results on FER 2013 dataset are shown in Table 36.4. Mollahosseini et al. [9] has used a convolution neural network with two convolution layer followed by max pooling and four inception layer and achieved 66.4% accuracy. Tümen et al. [10] achieved 57.1% accuracy, whereas VGG-19 architecture 69.1% accuracy is achieved in the proposed methodology. The proposed architecture is also used in face recognition and pattern recognition [11, 12] applications.

36.5 Conclusions

This paper presents recent development in the facial emotion recognition domain. The paper described VGG 19 architecture with different non-linear activation functions Elu and ReLu. Based on the experimental study, it has been observed that the size of the dataset plays an important role in facial expression recognition. Facial emotions have more samples such as happy, neutral, and angry performed better accuracy than other emotions. Further research based on multimodal deep learning architecture can improve more accuracy in this domain. One challenging issue is recognizing emotions from low-resolution images. It is also observed that the facial expressions are more suitable with the dynamic images in place of static images.

References

1. Guptaa, A., Arunachalama, S., Balakrishnana, R.: Deep self- attention network for facial emotion recognition. Procedia Comput. Sci. Elsevier B.V. **171**, 1527–1534 (2020). https://doi.org/10.1016/j.procs.2020.04.163
2. Mellouka, W., Handouzia, W.: Facial emotion recognition using deep learning: review and insights
3. Mehendale, N.: Facial emotion recognition using convolutional neural networks (FERC). SN Appl. Sci. **2**(3), 1–8 (2020)
4. Zhang, Y.-D., Yang, Z.-J., Lu, H.-M., Zhou, X.-X., Phillips, P., Liu, Q.-M., Wang, S.-H.: Facial emotion recognition based on biorthogonal wavelet entropy, fuzzy support vector machine, and stratified cross validation. IEEE Access **4**, 8375–8385 (2016)
5. Jabid, T., Hasanul Kabir, Md., Chae, O.: Robust facial expression recognition based on local directional pattern. ETRI J. **32**(5), 784–794 (2010)
6. Mollahosseini, A., Chan, D., Mahoor, M.H.: Going deeper in facial expression recognition using deep neural networks. In: 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1–10. IEEE (2016)
7. Operto, F.F., Pastorino, G.M.G., Stellato, M., Morcaldi, L., Vetri, L., Carotenuto, M., Viggiano, A., Coppola, G.: Facial emotion recognition in children and adolescents with specific learning disorder. Brain Sci. **10**(8), 473 (2020)
8. Shan, C., Gong, S., McOwan, P.W.: Facial expression recognition based on local binary patterns: a comprehensive study. Image Vis. Comput. **27**(6), 803–816 (2009)
9. Mollahosseini, A., Chan, D., Mahoor, M.H.: Going deeper in facial expression recognition using deep neural networks. In: 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1–10 (2016). https://doi.org/10.1109/wacv.2016.7477450
10. Tümen, V., Söylemez, Ö.F., Ergen, B.: Facial emotion recognition on a dataset using convolutional neural network. In: 2017 International Artificial Intelligence and Data Processing Symposium (IDAP), pp. 1–5. IEEE (2017)
11. Kumar, R., Sagar, L.K., Awasthi, S.: Human activity recognition from video clip. Intelligent Computing in Engineering, pp. 269–274. Springer, Singapore (2020)
12. Kumar, R., Joshi, S., Dwivedi, A.: CNN-SSPSO: a hybrid and optimized CNN approach for peripheral blood cell image recognition and classification. Int. J. Pattern Recogn. Artif. Intell. 2157004 (2020)