International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Deep Learning based Automatic Facial Emotion Recognition

[1]Shivangi Srivastav, [2]Ashish Khare

Submitted: 03/11/2023 **Revised**: 22/12/2023 **Accepted**: 04/01/2024

Abstract: Automatic feeling acknowledgment reliant upon look is a captivating assessment field, which has presented and applied in a couple of districts like security, prosperity and in human machine interfaces. The task of video synopsis has been playing a significant role in domain of video surveillance based systems. Various automation-based system relies on the identification of human emotions exhibited in running video frames. In this work we have investigated the methods to enhance the performance of deep learning model for the task of facial emotion recognition. Models are trained on Facial Emotion Recognition (FER) database with a keen focus to identify the optimal learning parameters using fit one cycle policy. Transfer learning is applied on popular models viz VGG and Resnet to identify the delineating decision boundary for human emotions. The model has successfully achieved a permissible classification rate of 70.2 % on FER dataset respectively

Keywords: *Face emotion recognition, feature extraction, Deep neural network, Convolutional Neural Network (CNN), Machine learning method.*

1. Introduction

Emotions is difficult to define, and there is no single definition that everyone agrees on. The word “emotion” refers to a person’s feelings and the outward manifestation of those emotions. Researchers have developed a method of classifying words related to emotion, which can be used in a variety of contexts or environments. Face expression is necessity part of human correspondence. Therefore, the accurate classification of the appearance of the image and video data is important in the quest of researchers and the software development industry. In addition to recognizing facial emotions, the field of computer vision (CV) also faces many challenges. Therefore, there is data and competition in these jobs. Large-Scale Visual Recognition Competition (ILSVRC), which focuses on recognizing objects in images. ImageNet has approximately 10 million images in over 1,000 categories. Since the discovery of AlexNet’s 5-layer network and solutions proved successful at ILSVRC in 2012, deep audio networks that can deliver more difficult features have become popular in ‘the field of CV research [13]. For mixed neural networks, the convolutional layer is considered as the feature miner, and the fully connected network is considered as the category. If the network has multiple similarities, the result of the last concurred layer is known as the depth feature. For profound organizations, various layers of common denominators mean a larger solution area, so the deep layer (the boundary produced by the final non-classification plate) has a higher dimension and contains more information from of the introductory image. Therefore, deep work was used to prioritize emotion, and the classification results were improved.

Like ImageNet and ILSVRC, in the face acknowledgment region, there is "feeling acknowledgment" for "challenge (Emotion)" and an Emotion data set intended for this difficulties. The challenge was first performed in 2013. At that time there were two stocks, namely “Wild Facial Expressions” (AFEW) and “Static Facial Expressions” (SFEW). The local orthogonal pattern (LBP-TOP) of the three orthogonal planes is the feature, the vector support machine (SVM) is the category, and the accuracy is 38% [4]. In subsequent competitions, the use of randomized controlled data demonstrated that the use of pre-trained auditory networks to capture visual cues and the use of long-term memory (LSTM) to incorporate long-term effects into accounts are effective. . The winning team at Emotion 2016 performed with three-dimensional convolutional network (C3D) efficiency and achieved the best increase with 59% accuracy [2]. In this review, information from Emotion will be used to set up the model. Meanwhile, the requirements and eventual outcomes of the resistance will be used to evaluate the reasonability of the arrangement model in the model.

2. Related Work

Paul Ekman [21], works in bliss, wretchedness, shock, shock, fear, and unpleasantness were perceived as the six head opinions (other than objective). Ekman later made FACS [22] utilizing this idea, in like manner setting the norm for work on feeling certification from that point forward. Fair-minded was in like manner included later on in generally human acknowledgment datasets, bringing about seven essential feelings. Picture tests of these feelings from three datasets are shown in below figure;

[1]*Centre of Computer Education and Training, Institute of Professional Studies, University of Allahabad Prayagraj, India*
meshiwangi6nov@gmail.com
[2]*Department of Electronics & Communication, JK Institute of Applied Physics and Technology University of Allahabad Prayagraj, India*
ashishkhare@hotmail.com

Fig. 1. (Passed on to right) The six cardinal feelings (joy, trouble, outrage, dread, revulsion, and shock) and impartial. The pictures in the main, second, and the third lines have a place with the FER and JAFFE informational index individually.

Prior chips away at feeling certification depended upon the conventional two-adventure AI approach, where in the hidden development, two or three elements are eliminated from the photos and, in the following turn of events, a classifier (like SVM, cerebrum association, or unpredictable forest area) is utilized to see the opinions. A piece of the prominent hand-made highlights utilized for look assertion join the histogram of coordinated places (HOG) [23,24], neighborhood twofold models (LBP) [25], Gabor wavelets [26], and Haar features[27]. A classifier then, commits the best tendency to the picture. These frameworks appeared to wind up staggering on less problematic datasets, yet with the approach of more testing datasets (which have more intra-class variety), they began to show their limits. To get an unmatched impression of a piece of the reasonable difficulties with the photographs, we suggest the perusers to the photographs in the fundamental line of Figure 1, where the image shows basically a midway face or the face is obstructed with a hand or eyeglasses. With the superb result of enormous learning and, shockingly, more unequivocally convolutional cerebrum networks for picture portrayal and other vision issues [28-35], a few parties made immense learning-based models for look validation (FER). To name a piece of the promising works, Khorrami in [17] explain that CNNs can accomplish a high accuracy in feeling interest and utilized a zero-propensity CNN on the long Cohn-Kanade dataset (CK+) and the Toronto Face Dataset (TFD) to accomplish top level outcomes. Aneja et al. [2] fostered a model of searches for changed vivified characters considering immense learning through setting up a inter phase between to show the presence of human faces, one for that of enlivened appearances, and one to design human pictures into brightened up ones. Mollahosseini [9] consist a mind network for FER using two convolution layers, one max pooling layer, and four "starting" layers,

. They used 10 taggers to relabel each image in the dataset and involved various cost limits with respect to their DCNN, achieving fair accuracy. Han et al. [37] consist a solid supporting CNN (IB-CNN) to deal with the verification of unconstrained looks by aiding discriminative neurons, which showed revives over the best methods by then. Meng in [38] proposed an individual mindful CNN (IA-CNN) that used person and verbalization fragile contrastive occurrences to reduce the blends in learning character and explanation related information. In [39], Fernandez et al. Develop a beginning to end network plan for look affirmation with a thought model.

Khorrami in [17] explain that CNNs can accomplish a high accuracy in feeling interest and utilized a zero-propensity CNN on the long Cohn-Kanade dataset (CK+) and the Toronto Face Dataset (TFD) to accomplish top level outcomes. Aneja et al. [2] fostered a model of searches for changed vivified characters considering immense learning through setting up a inter phase between to show the presence of human faces, one for that of enlivened appearances, and one to design human pictures into brightened up ones. Mollahosseini [9] consist a mind network for FER using two convolution layers, one max pooling layer, and four "starting" layers.

3. Research Model

Deep learning has as of late turned into a hot exploration subject and has accomplished best in class execution for an assortment of utilizations. Deep learning endeavors to catch undeniable level reflections through progressive designs of numerous nonlinear changes and portrayals. In this part, we momentarily present some deep learning strategies that have been applied for FER.

A Deep Neural Network is a complex Neural Network which comprises of a lot more secret layers than an exemplary multi-layer perceptron (MLP). Every one of these layers are prepared independently to learn various degrees of portrayal to sort out input information. Stuhlsatz et al. (2011) frame a methodology which utilizes a DNN engineering and Generalized Discriminant Analysis5 (GerDA). In this concentrate on a Neural Network is straightforwardly prepared with highlights extricated utilizing the OpenEAR tool stash (Eyben et al., 2009). A sum of 6552 information highlights altogether (39 functionals of 56 acoustic LLDs alongside the comparing first and second request delta relapse coefficients) are introduced to the info hubs of the Neural Network.

Simonyan and Zisserman of the University of Oxford prepare a 19-layer (16 conv., 3 completely related) CNN that immovably utilized 3×3 channels with step and

cushion of 1, nearby 2×2 max-pooling layers with stage 2, called VGG-19 model.28,29 Compared to

Alex Net, the VGG-19 (see Fig. 8) is a high essential

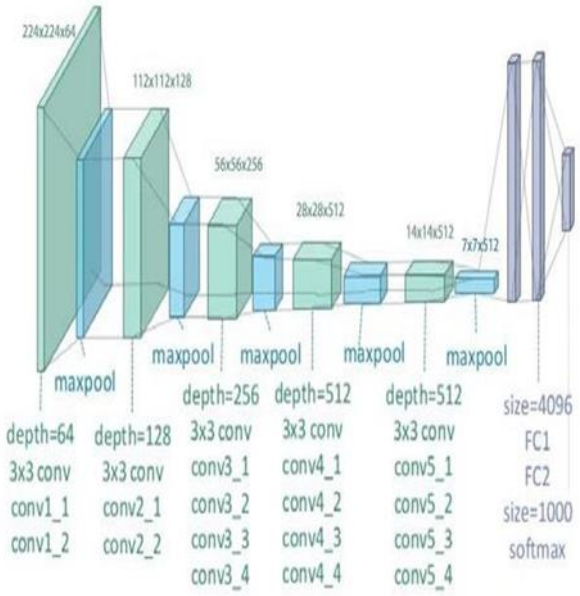

Fig. 2. Architecture of VGG19

CNN with more layers. The VGG-19 model was not the boss of ILSVRC30 2014, notwithstanding, the VGG Net is possibly the most persuading paper since it stayed aware of the probability that CNNs ought to have a massive relationship of layers for this various leveled out portrayal of visual information to work. Keep it gigantic. Keep it direct.

VGG-19 model, a total of 138M limits, was set second all together and first in deterrent in ILSVRC 2014. This model is ready on a subset of the ImageNet27 data base, which is implement in the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC).30 The VGG-19 is ready on more than 1,000,000 pictures and can organize pictures into 1000 article classes, for example, console, mouse, pencil, and various animals. Accordingly, the model has learned rich part depictions for a wide level of images. For the proposed research work complete 25 number of pictures are considered in a solitary bunch which demonstrates the cluster size for the unbiased organization preparing. The batch size is the quantity of tests that will be gone through to the organization at one time. Note that a group is likewise ordinarily alluded to as a scaled down cluster. The batch size is the quantity of tests that are passed to the organization on the double. Presently, review that an age is one single disregard the whole preparation set to the organization. The bunch size and an epoch are not exactly the same thing.

Fig. 3. The six cardinal emotions happy, sad, angry, fear, surprise and neutral in a single batch

4. Results & Analysis

As we know at the time of run of neural network we have to find the learning rate. For the proposed work we found a graph of learning rate on X axis and losses on Y axis. As per learning rate graph figure 4, learning rate is going down with increasing the value of loss, and we will select the learning rate at which we found minimum loss. This is the learning of initial level. At the time of unfreeze the model which means complete model is going to train and we got the confusion matrix. As per table 1 with increasing the epoch the value of train loss, valid loss, error rate decreasing.

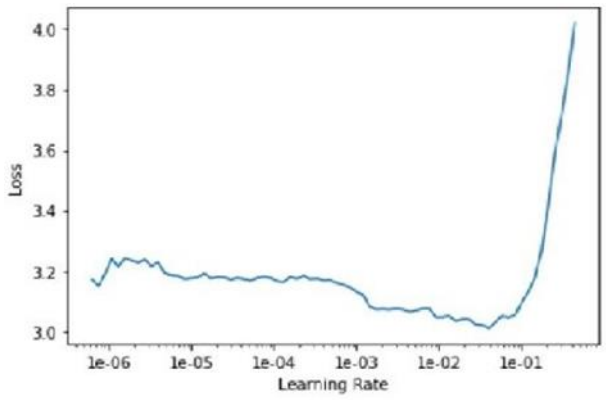

Fig. 4. Learning rate graph

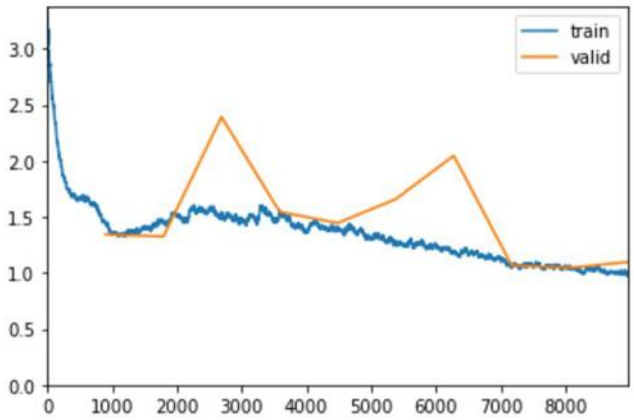

Fig. 6. Losses with different epoch

Table 1. Epoch table with train loss, valid loss and error rate

epoch	train_loss	valid_loss	error_rate	time
0	1.448263	1.341756	0.508432	03:48
1	1.451926	1.324375	0.468014	03:46
2	1.540882	2.390426	0.532125	03:47
3	1.535393	1.545033	0.464808	03:46
4	1.401066	1.443746	0.450035	03:46
5	1.266595	1.657189	0.448502	03:46
6	1.174047	2.044711	0.403624	03:47
7	1.050494	1.064952	0.385784	03:47
8	1.059177	1.044803	0.369895	03:47
9	0.973360	1.096626	0.368780	03:47

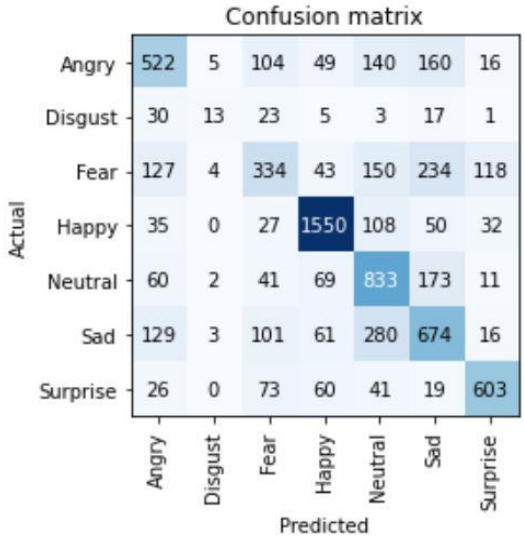

Fig. 7. Confusion matrix

To enhance the performance, learning rate is found in second step as sown in figure 8. Similarly we found confusion matrix with unfreeze model. The diagonal elements of confusion matrix have maximum value which means we found better performance as compare to last step.

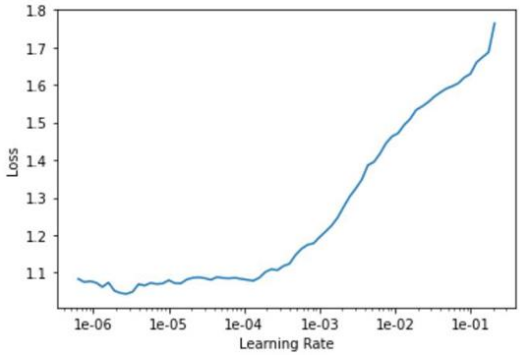

Fig. 8. Learning rate graph

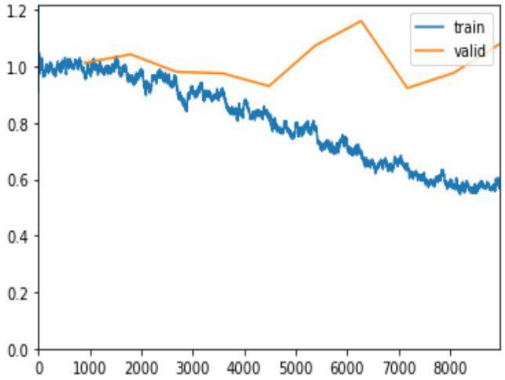

Fig. 9. Losses with different epoch

Table 2. Epoch table with train loss, valid loss and error rate

epoch	train_loss	valid_loss	accuracy	time
0	0.519644	0.936636	0.681672	05:22
1	0.547063	0.946719	0.680000	05:20
2	0.646104	1.016159	0.667038	05:20
3	0.712806	1.026020	0.670941	05:20
4	0.704054	1.028589	0.647387	05:20
5	0.744804	0.966655	0.658676	05:21
6	0.712089	0.977269	0.659094	05:22
18	0.100015	1.485144	0.689895	05:22
19	0.062885	1.445665	0.695192	05:22
20	0.053863	1.515924	0.696585	05:21
21	0.033392	1.498064	0.692683	05:21
22	0.033359	1.521938	0.700209	05:21
23	0.028073	1.540743	0.699512	05:21
24	0.023086	1.551744	0.696725	05:24

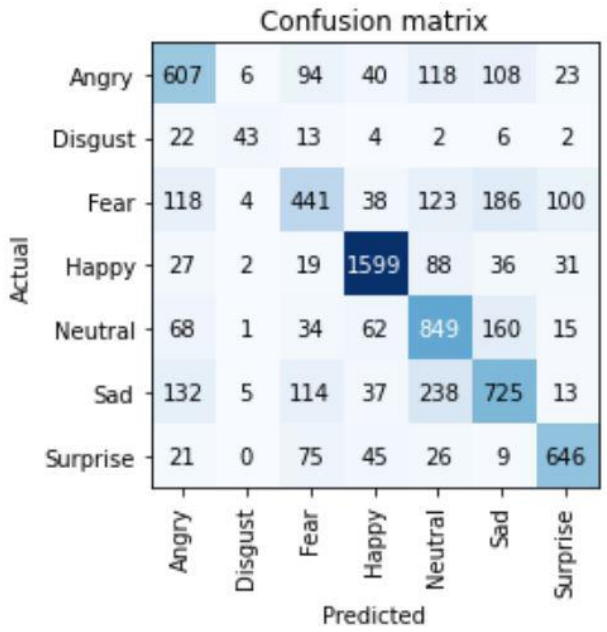

Fig. 10. Confusion matrix

Following figure shows the class of activation map which is used to identify the six cardinal emotions happy, sad, angry, fear, surprise and neutral. Fir the identification of these emotions CAM focuses on the particular part of the face in respect of the emotions. As per figure for the identification of angry emotion, nose part of the face is highlighted. For neutral, happy and sad emotions mouth area is highlighted and for surprise nose and mouth are is highlighted.

Fig.11. Class activation map for six cardinal emotions happy, sad, angry, fear, surprise and neutral

5. Future Scope

Face feeling acknowledgment is a difficult errand including so many subtasks including face location, face following, face acknowledgment, 3D change and others. Of all the sub-tasks referred to under, a predominant face recognizable proof technique got together with face affirmation, could additionally foster the results basically by independent different individuals in a solitary video and imprint their sentiments properly. Feeling is awesome in its real nature and the course of action area of it is outstandingly tremendous and requires a sufficient number of data to learn. Be that as it may, since there are such endless pictures to be checked and stamped actually, not a solitary one of them are used in the readiness cycle. If possible, having even more precisely named data from wild will impact the precision. For the most part, there are still a ton of progress ought to be conceivable as for feeling affirmation. Besides, in a perfect world one day, it will in general do, in reality, circumstance.

References

[1] Chu, H.C.; Tsai, W.W.; Liao, M.J.; Chen, Y.M. Facial feeling acknowledgment with change identification for understudies with advanced chemical imbalance in versatile e-learning. Delicate Comput. 2017, 22, 2973-2999.

[2] Chloé, C.; Vasilescu, I.; Devillers, L.; Richard, G.; Ehrette, T. Fear-type emotion recognition for future audio-based surveillance systems. Speech Commun. 2008, 50, 487–503.

[3] Saste, T.S.; Jagdale, S.M. Emotion recognition from speech using MFCC and DWT for security system. In Proceedings of the IEEE 2017 International Conference of Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 20–22 April 2017; pp. 701–704.

[4] Marco, L.; Carcagnì, P.; Distante, C.; Spagnolo, P.; Mazzeo, P.L.; Rosato, A.C.; Petrocchi, S. Computational assessment of facial expression production in ASD children. Sensors 2018, 18, 3993. [CrossRef]

[5] Meng, Q.; Hu, X.; Kang, J.; Wu, Y. On the effectiveness of facial expression recognition for evaluation of urban sound perception. Sci. Total Environ. 2020, 710, 135484.

[6] Ali, M.; Chan, D.; Mahoor, M.H. Going deeper in facial expression recognition using deep neural networks. In Proceedings of the IEEE 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, USA, 7–10 March 2016.

[7] Liu, P.; Han, S.; Meng, Z.; Tong, Y. Facial expression recognition via a boosted deep belief network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1805–1812.

[8] E. Sariyanidi, H. Gunes, et A. Cavallaro, « Automatic Analysis of Facial Affect: A Survey of Registration, Representation, and Recognition », IEEE Trans. Pattern Anal. Mach. Intell., oct. 2014.

[9] C.-N. Anagnostopoulos, T. Iliou, I. Giannoukos

Features and classifiers for emotion recognition from speech: a survey from 2000 to 2011 Artif. Intell. Rev., 43 (2) (2015), pp. 155-177 févr.

[10] L. Shu A Review of Emotion Recognition Using Physiological Signals Sensors, 18 (7) (2018), p. 2074 juill., doi: 10.3390/s18072074.

[11] J. Kołodziej, H. González-Vélez (Eds.), High-Performance Modelling and Simulation for Big Data Applications: Selected Results of the COST Action IC1406 cHiPSet, Springer International Publishing, Cham (2019), pp. 307H. Alkawaz, D. Mohamad, A.H. Basori, T. Saba Blend Shape Interpolation and FACS for Realistic Avatar 3D Res., 6 (1) (2015), p. 6.

[12] P. V. Rouast, M. Adam, et R. Chiong, « Deep Learning for Human Affect Recognition: Insights and New Developments », IEEE Trans. Affect. Comput., p. 1-1, 2018.

[13] C. Shan, S. Gong, P.W. McOwan Facial expression recognition based on Local Binary Patterns: A comprehensive study Image Vis. Comput., 27 (6) (2009), pp. 803-816.

[14] T. Jabid, M.H. Kabir, O. ChaeRobust Facial Expression Recognition Based on Local Directional Pattern ETRI J., 32 (5) (2010), pp. 784-794.

[15] S. Zhang, L. Li, et Z. Zhao, « Facial expression recognition based on Gabor wavelets and sparse representation », in 2012 IEEE 11th International Conference on Signal Processing, oct. 2012, vol. 2, p. 816-819.

[16] R. Gross, I. Matthews, J. Cohn, T. Kanade, S. Baker Multi-PIE Proc. Int. Conf. Autom. Face Gesture Recognit. Int. Conf. Autom. Face Gesture Recognit., 28 (5) (2010), pp. 807-813.