Review Article

A Study on Detection of Emotions with the Help of Convolutional Neural Network

Puneet Joshi

GKM College of Engineering & Technology, G.K.M. Nagar, Chennai India.

I N F O

A B S T R A C T

Facial Expression is a mean which can be employed to find out what is running in the mind of the person whom we talk. It is always an easy task for humans to derive some insights from the expressions. But it is a complicated task for machine to derive it by using Computer Algorithms. But recent developments in the field of Computer Vision and Machine Learning have enhanced the availability of resources so that it become possible to derive certain conclusions from the input we are giving to particular Machine. This paper is planning to propose an approach to find out what emotional stage the particular person is running which is termed as Facial Emotion Recognition using Convolutional Neural Network (FERC). The FERC constitutes two parts: The first part of this Convolutional Neural Network is used to remove background from an image and second part is used to extract features from facial expressions. It consists of a database of around 10,000 images. The final layer which is a perceptron which works in series with two-layer Convolutional Neural Network. The Perceptron is used to adjust weights and exponent values after going through each iteration. Again, a background removal is applied just before generation of Emotional Vector. This will help us to avoid from multiple problems which can occur. Using Two stage CNN model we found out accuracy near about 84.92 percentage which is a better value based on insights from 24 values. The Two Layer CNN works in series where remaining layer is used to adjust weights after each iteration. FERC follows an exactly new approach compared to Single stage CNN and thus enhancing the accuracy. Further it uses a novel background elimination procedure compared to technology of EV earlier existed which could avoid coping with coping with more than one troubles that may occur. FERC is employed with about more than 28K pictures by using FER2013 dataset. We can make use of FERC in many cases including predictive learning, Lie detector etc.

Keywords: Facial Emotion Recognition, Deep Convolutional Neural Network, Convolutional Layer, Pooling Layer, Fully Connected Layer

Introduction

Human Emotions can be derived by using Facial Expression which is considered as vital definitions of human feelings.[1] In most case facial expressed is considered as the effective way to unveil truth through nonverbal communication.[2] Almost in 55 percentage of cases it went correct which is not a bad amount.[3] In the present approach, the author primarily focusses on an investigation through face which finally eliminates misleading features by keeping background intact.[4] This can help during CNN training process so that we can derive more efficient output.[5] In this present manuscript the author primarily focuses on Five basic facial expressions which are mainly anger, Sad, happy, feared and surprised.[6] The Algorithm which is mentioned in this manuscript classifies the images into five basic emotions basedon input given.[7] There are two main approaches which has reported till now.[8] The first type is used to distinguish the expressions which are identified with an explicit classifier.[9] The second one is used to highlight the characteristics of facial expressions. Action units which are marked as expression markers in the Facial Action Coding System (FACS).[10] By Facial Muscle changes AU were discriminable.[11] Facial Emotion Recognition is one of the broadly used era that examine and extract facts concerning feelings from unique sources.[12] The supply of the enter can both be pictureor video.[13] It is an essential part of the 'affective com-putting' own circle of relatives of technologies, that is a multidisciplinary area of have a take a observe on computers' capacity to renowned and understand human feelings and emotional states.[10] It regularly relies upon on Artificial Intelligence era. Facial expressions are nonverbal indicates thatof conversation that screen human feelings.[14] Decryption such emotion expressions has been a have a take a observe subject matter in science, likewise as withinside the Humanpc Interaction sector, for lots years.[15] The great availability of cameras, moreover as current technical improvements in bioscience evaluation, system mastering, and sample reputation, have all motor-assisted the occasion of the FER era.[16] Many organizations, from enterprise behemoths like NEC or Google to small startups like Affective and Eyeris, are fundingwithinside the era, demonstrating its growing relevance.[9] There are also opportunity EU Horizon2020 initiatives1 which might be watching the era' use. Face detection, face expressiondetection, related expression type to an emotion is the three approaches withinside the FER evaluation.[17] The havea take a observe of facial landmark placements is hired to training session feelings (e.g., end of nose, eyebrows).[11] Changes of their postures are tested in films a good wayto spot contractions throughout a fixed of face muscles. Faces may be categorized as simple feelings (anger, disgust, fear, joy, sadness, and surprise) or compound feelings (luckilysad, mirthfully surprised, luckily disgusted, tragically scared, unfortunately angry, unfortunately surprised) relying on the algorithm.[18] In different circumstances, facial expressions canbe linked to a person's physiological or intellectual condition (e.g., tiredness or boredom).[19] The pix or motion pictures used as enter to FER algorithms come again from a range of sources, collectively with police research cameras, cameras settled near marketing and marketing displays in businesses, social media and streaming services, and personal devices.[20]Biometric identity also can be used along with FER.[7] Its accuracy can be expanded through the use of era to examine plenty of sources, together with speech, text, fitness statistics from sensors, and blood go with the drift styles inferred from photographs.[21] Potential programs for FER encompass a huge variety of fields, that are stated underneath in classes consistent with software area.[22] We are Mainly dealing withSeven basic Emotions.[23]

Angry

Nose got elongated, eyebrows form the shape of parabola which is in the form of inverted, Mouth length is decremented are some of the characteristics of Anger emotion.[24]
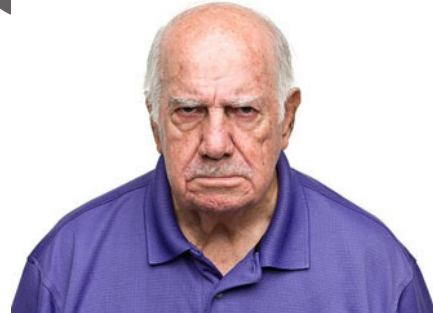

Figure 1. Angry

Disgust

Here it is hard to find eyebrow size. Mouth forms the shape of parabola which is in the shape of inverted which describes some of the features of disgust emotion.[25]

Figure 2. Disgust

Fear

Mouth forms the shape of parabola which is in the shape of inverted. Both the eyes are pointed towards the nose center.These are some characteristics of fear.[26]

Figure 3. Fear

Happy

In this case eyes which is opened more than normal in size. Eyebrows also moves little bit upwards from their original position. These are some characteristics of Happy emotion.[27]

Figure 4. Happy

Sad

Mouth forms the shape of parabola which is inverted in size.Eyebrows move towards the side of Nose.[28]

Surprise

Figure 5. Sad

When a person falls into this type of emotion his eyebrow almost form the shape of circle and Nose is projected outwards.[29]

Figure 6. Surprise

Neutral

This is considered as base emotion from where all further emotions are calculated.[30]

Literature Survey

William Dias a, Fernanda Andalo, Rafael Padilha, Gabriel Bertocco, Waldir Almeida, Paula Costa, Anderson Rocha a[31] proposed models basically based on two methods that is Random patches and triplet loss. In random patches Author takes 9 datasets and tries to implement Facial Emotion Recognition model using that. In second method author tries to implement Face emotion Recognition using triplet loss methodology author tries to explore another direction in a trial to enhance cross-dataset accuracy for emotion recog- nition. they try to learn sensible embeddings by the notion of similarity and unsimilarity of facial expressions. Feature vectors are transported to a hyperspace wherever the gap between samples of the same category is small, whereas the distance between examples of totally different classes is large. this can be accomplished by using a triplet loss function throughout model training. There are using different datasets for the application of this Bosphorous, Child Affective Fa- cial Expression (CAFE), Compound Facial Expression of Emo- tions(CFEE),Cohn Kanade(CK+),Karolinska Directed Emo- tional Faces Dataset, MUG facial expression database, Natural Visible and Infrared Facial expression database (NVIE), Oulu- CASIA NIR and VIS facial expression database, Radbound face database. By going through the results, we found that NVIE dataset is performing Low accuracy. Sumeet Saurav · Prashant Gidde · Ravi Saini · Sanjay Singh.[32] In this approach author gives image input to two Convolutional Neural Network and concatenate the two inputs and performs the detection of emotions. They have done this operation on two datasets FER 2013 and FER Plus datasets. With FER 2013 dataset it got highest accuracy for combined network of Covolutional Neural Network, Bag of Visual Words (BOVW) and Support

Vector Machine (SVM) which is around 75.42and with FER Plus dataset it got highest accuracy for Sparse Covolution Network (SCN). Ninad Mehendale.[33] In this method author proposes to give input as image. The first step then is removal of background, then classify different vectors of face, then extract expression from image,then classify the images into different emotions using CNN and thus generate output. For the procedure we have make use of three datasets Caltech database which contribute images of 450 which gives an accuracy of 85percentage, CMU database which have around 7,50,000 images which gave accuracy around 78 percentage and NIST database which gave accuracy around 96 percentage which have images around 3248. The higher accuracy of first and last may be due to overfitting because of smaller dataset. Tanoy Debnath, Md. Mahfuz Reza1, Anichur Rahman, Amin Beheshti, Shahab S. Band and HamidAlinejad Rokny.[4] In this method the author is using three datasets that is FER2013 dataset, Japanese Female Facial Expressions (JAFFE)and ex- tended Cohn Kanade database. From this model author found out accuracy of 78-96 percentage which is better than other networks such as Alexnet, Resnet, Googlenet and VGG.Saiyed Umer · Ranjeet Kumar Rout · Chiara Pero · Michele Nappi.[5] in this approach the author gives input image. After face detection it is converted into training and testing image and training image after data augmentation, it is given to CNN model and finally it is sent to Expression Recognition Model and thus finally detects emotions. FER datasets are used here and highest accuracy author is getting in this approach is 97.69 percentage

Proposed Approach

We are taking input images from camera or any other picture capturing devices. We are then inputting the pictures into CNN. By using CNN, we remove the background of the image and find out the vector of face using this. Second Layer of CNN is used to classify the image into different emotions such as Anger, Fear, Disgust, Sad, Disgust, Surprise after comparing with the images in the dataset

Dataset

First the requirement for the approach is a dataset which is taken from 2013 challenge of Facial Emotion Recognition. The data includes 48x48 pixel grayscale snap shots of faces. For making face more focused and to make every image occupy equal quantity of area the faces were made registered auto- matically.[5] The Main task is to divide the face into seven classes according to facial vectors (Zero denotes Angry, one denotes Disgust, two denotes Fear, three denotes Happy, fourdenotes Sad, five denotes Surprise and Six denotes Neutral). The given training set comprises of samples having 28,709 examples and the public test set consists of samples having 3,589 examples.[8,34]
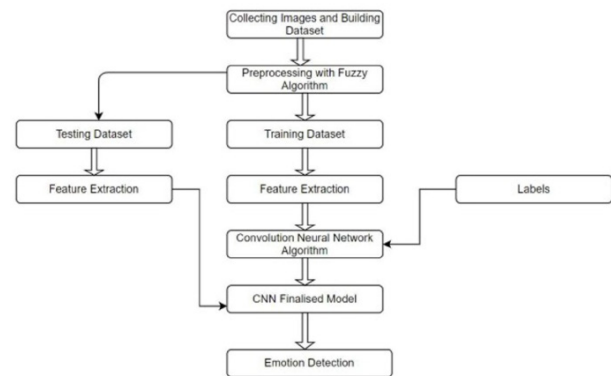

Figure 7. Proposed Approach

Convolutional Neural Network

The technique which is used to evaluate the various features in an image is mainly by Convolutional Neural Net- work (CNN). The difference between Convolutional Neural Network and Multilayer Perceptron lies in its Convolutional Layers or Hidden Layers. A -stage CNN shape underpins the suggested method. Background elimination is that the prelim- binary step, this is hired to extract feelings from a photograph. The number one expressional vector is extracted exploitation the regular CNN community module (EV). By chase down essential face spots, the expressional vector (EV) is created. Changes in expression are intently linked to strength unit. An easy perceptron is used to calculate. There are fundamental elements to a CNN architecture.[35]

(1) Convolutional Layer: The important layer which helps us in extracting essential features of an image is done by this layer. By the use convolutional Mathematical Machine, they filter important features and eliminates the features which is not necessary.[36]. The dot product is used to cleaning input photo by deriving important features from an input image of M rows and M columns. After this step we got a Feature Map which has information about corners and edges. This map which we got is then given for finding what all capabilities particular that image have.[37]

(2) Pooling Layer: After Convolutional layer the input is given to Pooling Layer. Its Main Function is to reduce the size of Feature Map so that the total process can be done with less cost of computation. This is finished with the useful resource of using lowering the connections among layers and walking independently on every characteristic map. Many kinds of Pooling procedures used which relies on the mechanism which we processed. First one is Max pooling in which biggest element is derived from feature map. The second one is Average pooling where average of pooling is taken inorder to reduce the size of feature map. Third one

is sum pooling which finds the sum of all additives inorder to get a more efficient feature map. Mainly Pooling Layer is used as an intermediate layer between Convolutional Layer and Fully Connected Layer.[1]

(3) Fully Connected Layer: The Fully Connected Layer connects the neurons by blending it with Weight and biases. It is placed earlier than Output Layer in Numerous Layers. The input photos which are flattened through previous layers are given to this layer for further process. After that, the flattened vector is despatched thru some more FC levels, in which the mathematical realistic operations are typically performed. The categorization manner receives began out out at this point.[38]
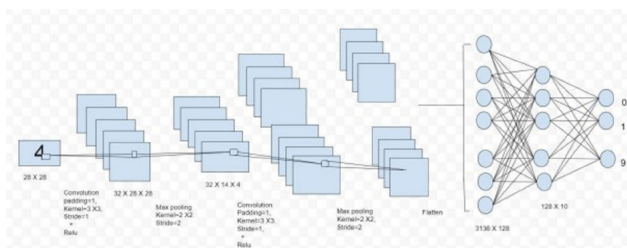

Figure 8. Convolutional Neural Network

Dropout

The dataset for training is quite vulnerable to overfitting when most of the trends are associated with. Overfitting takesvicinity on the equal time as a version plays so nicely on training facts that it has an unfavorable have an impact onits simple general overall performance on the equal time as finished to smooth facts. A dropout layer is used to remedy this problem, wherein some neurons are eliminated from the neural network for the duration of the training process, ensuingin a smaller version. After passing a dropout of 0. three, 30 percentage of the nodes with inside the neural network are randomly eliminated.[39]

Activation Functions

Activation Function is one of the important parts of CNN model. It is used to examine, approximate whether there is anyrelationship which is complex among different set of Variables. In simple words we can say it determines the information about the model in its upcoming steps. It provides some Non Linearity to the Model. There are lot of Activation Functions used .Some are ReLU,Sigmoid,tanH and Soft max. Each of the given activation functions has plenty of uses. Binary class of CNN model is determined bysigmoid and softmax functions , while softmax is mainly employed in multiclass classification problems.[40]

Results and Discussion

Table 1 suggests the findings of the recommended CNN fashions' comparative have a take a observe the FER2013 Dataset. We evaluated the version's overall performance the use of 4 assessment metrics: reputation accuracy, the variety of version parameters (expressed in millions (M)), and the version's reminiscence garage capacity (in megabytes (MB). When we examine the findings in Table 1, we see that at the FER2013 dataset, the proposed DICNN version outperformed the alternative latest FER processes. In addition, in contrast to these days brought neural community fashions, the recommended light-weight DICNN version finished most ex- cellent overall performance and maintained a stability among identity accuracy, parameter variety, and reminiscence garage size. Deep mastering-primarily based totally pc imaginative and prescient algorithms in real-global programs working on a useful resource-limited embedded platform are regularly decided through those considerations. Georgescu et al used the FER method to mix traits derived through 3 CNN fashions. These troubles restrict their FER gadget's capacity to run on embedded systems. In conclusion, our recommended FER gadget has now no longer most effective attained aggressive accuracy at the FER2013, however it's also computationally green. Table 1 suggests the DICNN version's comparative evaluation findings at the FER dataset. We evaluated the overall performance of numerous of these days introduced FER processes to that of the FER2013 database in phrases of reputation accuracy, variety of version parameters, and version size. The recommended DICNN version has finished aggressive reputation accuracy in phrases of reputation accuracy. With an 84.92 percentage identity accuracy, 1.08 million version parameters, and 5.40MB reminiscence garage, the recommended DICNN version gives the most excellent trade- off among reputation accuracy and computational efficiency.

Conclusion

We brought a completely unique light-weight CNN version for real-time FER on this work. To that aim, we recommendeda twin included CNN (DICNN) version that mixes complimentary understanding from proprietary CNN fashions that had been skilled independently at the FER datasets. On the FER dataset, we examined the DICNN version. The recommended DICNN version's software become proved thru comparisons with many latest CNN fashions. In evaluation to preceding CNN fashions, the advanced DICNN version finished a truth- ful stability of reputation accuracy and processing economy.

On a computer PC with a GPU, the entire FER pipeline processed frames at a fee of forty eight frames consistent

with 2nd. The TensorRT optimised version withinside the FER process. As a result, the FER gadget has been advanced to be computationally green and perfect to be used in real-global programs. We need to enforce the proposed FER gadget on anAndroid phone withinside the future, and utilise it as a device to resource visually impaired individuals (VIPs) of their everyday conversation.

References

1. Dias, William, Fernanda Andalo´, Rafael Padilha, Gabriel Bertocco, Waldir Almeida, Paula Costa, and Anderson Rocha. "Cross-dataset emo- tion recognition from facial expressions through convolutional neural networks." Journal of Visual Communication and Image Representation 82 (2022): 103395.

2. Saurav, Sumeet, Prashant Gidde, Ravi Saini, and Sanjay Singh. "Dual integrated convolutional neural network for real-time facial expression recognition in the wild." The Visual Computer 38, no. 3 (2022): 1083- 1096.

3. Srimani, P. K., and R. Hegde. "Analysis of facial expressions with respect to navarasas in bharathanatym styles using image processing." International Journal of Knowledge Engineering 3, no. 2 (2012): 193.

4. Mehendale, Ninad. "Facial emotion recognition using convolutional neural networks (FERC)." SN Applied Sciences 2, no. 3 (2020): 1-8.

5. Debnath, Tanoy, Md Reza, Anichur Rahman, Amin Beheshti, Shahab S. Band, and Hamid Alinejad-Rokny. "Four-layer ConvNet to facial emotion recognition with minimal epochs and the significance of data diversity." Scientific Reports 12, no. 1 (2022): 1-18.

6. Khattak, Asad, Muhammad Zubair Asghar, Mushtaq Ali, and Ulfat Batool. "An efficient deep learning technique for facial emotion recogni- tion." Multimedia Tools and Applications 81, no. 2 (2022): 1649-1683.

7. Umer, Saiyed, Ranjeet Kumar Rout, Chiara Pero, and Michele Nappi. "Facial expression recognition with trade-offs between data augmenta- tion and deep learning features." Journal of Ambient Intelligence and Humanized Computing 13, no. 2 (2022): 721-735.

8. Vu, Hoai Nam, Mai Huong Nguyen, and Cuong Pham. "Masked face recognition with convolutional neural networks and local binary patterns." Applied Intelligence 52, no. 5 (2022): 5497-5512.

9. Kollias, Dimitrios. Abaw: Valence-arousal estimation, expression recog- nition, action unit detection and multi-task learning challenges." arXiv preprint arXiv:2202.10659 (2022).

10. Gera, Darshan, S. Balasubramanian, and Anwesh Jami. "CERN: Com- pact facial expression recognition net." Pattern Recognition Letters 155 (2022): 9-18.

11. Han, Jiayi, Liang Du, Xiaoqing Ye, Li Zhang, and Jianfeng Feng. "The devil is in the face: Exploiting harmonious representations for facial expression recognition." Neurocomputing 486 (2022): 104-113..

12. Han, Jiayi, Liang Du, Xiaoqing Ye, Li Zhang, and Jianfeng Feng. "The devil is in the face: Exploiting harmonious representations for facial expression recognition." Neurocomputing 486 (2022): 104-113.

13. Boughida, Adil, Mohamed Nadjib Kouahla, and Yacine Lafifi. "A novel approach for facial expression recognition based on Gabor filters and genetic algorithm." Evolving Systems 13, no. 2 (2022): 331-345.

14. Fard, Ali Pourramezan, and Mohammad H. Mahoor. "Ad-corre: Adap- tive correlation-based loss for facial expression recognition in the wild." IEEE Access 10 (2022): 26756-26768

15. Fan, Xinqi, Mingjie Jiang, Ali Raza Shahid, and Hong Yan. "Hierarchi- cal scale convolutional neural network for facial expression recognition." Cognitive Neurodynamics (2022): 1-12.

16. Giannopoulos P, Perikos I, Hatzilygeroudis I (2018) Deep learning approaches for facial emotion recognition: a case study on FER 2013. In: Hatzilygeroudis I, Palade V (eds) Advances in hybridization of intelligent methods. Springer, Berlin, pp 1–16

17. Gajarla V, Gupta A (2015) Emotion detection and sentiment analysis of images. Georgia Institute of Technology, Atlanta

18. Wu YL, Tsai HY, Huang YC, Chen BH (2018) Accurate emotion recognition for driving risk prevention in driver monitoring system. In: 2018 IEEE 7TH global conference on consumer electronics (GCCE). IEEE, pp 796–797

19. Zhang K, Huang Y, Du Y, Wang L (2017) Facial expression recognition based on deep evolutional spatial-temporal networks. IEEE Trans Image Process 26(9):4193

20. Jung H, Lee S, Yim J, Park S, Kim J (2015) Joint fne-tuning in deep neural networks for facial expression recognition. In: Proceedings of the IEEE international conference on computer vision. pp 2983–2991

21. Zhao X, Liang X, Liu L, Li T, Han Y, Vasconcelos N, Yan S (2016) Peak- piloted deep network for facial expression recognition. In: European conference on computer vision. Springer, pp 425–442

22. Shan C, Gong S, McOwan PW (2009) Facial expression recognition based on local binary patterns: a comprehensive study. Image Vis Comput 27(6):803

23. Littlewort G, Whitehill J, Wu T, Fasel I, Frank M, Movellan J, Bartlett M (2011) The computer expression recognition toolbox (CERT). In: Face and gesture 2011. IEEE, pp 298–305

24. Hsu RL, Abdel-Mottaleb M, Jain AK (2002) Face detection in color images. IEEE Trans Pattern Anal Mach Intell 24(5):696

25. Cowie R, Douglas-Cowie E, Tsapatsoulis N, Votsis G, Kollias S, Fellenz W, Taylor JG (2001) Emotion recognition

in human– computer interaction. IEEE Signal Process Mag 18(1):32

26. Herna´ndez-Travieso JG, Travieso CM, Pozo-Ban˜os D, Alonso JB et al (2013) Expression detector system based on facial images. In: BIOSIGNALS 2013-proceedings of the international conference on bio- inspired systems and signal processing

27. Kumar KC (2012) Morphology based facial feature extraction and facial expression recognition for driver vigilance. Int J Comput Appl 51:2

28. Ernst H (1934) Evolution of facial musculature and facial expression. J Nerv Ment Dis 79(1):109

29. Kim DH, An KH, Ryu YG, Chung MJ (2007) A facial expression imitation system for the primitive of intuitive human robot interaction. In: Sarkar N (ed) Human robot interaction. IntechOpen, London

30. Xue Yl, Mao X, Zhang F (2006) Beihang university facial expression database and multiple facial expression recognition. In: 2006 Inter- national conference on machine learning and cybernetics. IEEE, pp 3282–3287

31. Kong SG, Heo J, Abidi BR, Paik J, Abidi MA (2005) Recent advances in visual and infrared face recognition—a review. Comput Vis Image Underst 97(1):103

32. Dols JMF, Russell JA (2017) The science of facial expression. Oxford University Press, Oxford

33. Parr LA, Waller BM (2006) Understanding chimpanzee facial expres- sion: insights into the evolution of communication. Soc Cogn Affect Neurosci 1(3):221

34. Mal HP, Swarnalatha P (2017) Facial expression detection using facial expression model. In: 2017 International conference on energy, com- munication, data analytics and soft computing (ICECDS). IEEE, pp 1259–1262

35. Xie S, Hu H (2018) Facial expression recognition using hierarchical fea- tures with deep comprehensive multipatches aggregation convolutional neural networks. IEEE Trans Multimedia 21(1):211

36. Ratyal N, Taj I, Bajwa U, Sajid M (2018) Pose and expression invariant alignment based multi-view 3D face recognition. KSII Trans Internet Inf Syst 12:10

37. Shamshirband, S. et al. A review on deep learning approaches in healthcare systems: Taxonomies, challenges, and open issues. J. Biomed. Inform. 113, 103627 (2021)

38. Javanmard, R. et al. Proposed a new method for rules extraction using artifcial neural network and artifcial immune system in cancer diagnosis. J. Bionanosci. 7(6), 665–672 (2012)

39. Rajaei, P. et al. A user-friendly tool for viral sequence analysis. Genes 12(2), 186 (2021).

40. Alinejad-Rokny, H., Sadroddiny, E. Scaria, V. Machine learning and data mining techniques for medical complex data analysis. Neurocom- puting 276, 1 (2018)