IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A LIGHT WEIGHT CONVOLUTION NEURAL NTWORK FOR FACIAL EMOTION DETECTION

Senthil Kumar M
Associate professor, Department of Computer Science and Engineering
SRM Valliammai Engineering College, Chennai, Tamil Nadu, India

Shantha Kumar S
Department of Computer Science and Engineering
SRM Valliammai Engineering College, Chennai, Tamil Nadu, India

Sharmi G
Department of Computer Science and Engineering
SRM Valliammai Engineering College, Chennai, Tamil Nadu, India

Harsha Vardhan S
Department of Computer Science and Engineering
SRM Valliammai Engineering College, Chennai, Tamil Nadu, India

ABSTRACT - A Lightweight Convolution Neural Network for real-time facial expression detection employs multi-task cascaded convolution networks (MTCNN) to complete face detection and transmit the obtained face coordinates to the facial emotions' classification model. To replace the fully connected layer in the standard deep CNN, this model uses Global Average Pooling. To some extent, each channel of the feature map is associated with the corresponding category, removing the black box features of a completely connected layer. This model combines residual modules and depth-wise separable convolutions, resulting in a smaller number of parameters and a more portable model. Because it has good detection and recognition effects, we choose the FER-2013 dataset.

Keywords - MTCNN, Pooling, Black Box, Fer-2013 dataset

I. INTRODUCTION

People can deliver difficult labor to computers to suit particular life and market needs, thanks to the rapid development of human-computer interaction and pattern recognition, as well as the rapid updating of computer hardware. It is extremely beneficial to humanity. Facial expression recognition is new developing technology used in various fields such as in cellphones, education and psychological health centers, etc. Many cameras now offer smile mode, which means that when the camera detects a smile, a shot is taken automatically without the need to manually press the shutter, improving the user experience. Facial expression recognition is used in several European nations to capture the mood swings of elementary school pupils in class so that their learning status may be analyzed and students can be treated as individuals. Some Toyota models monitor the driver's eyes and facial expressions to detect drowsy driving and prevent traffic accidents.

The classic manual approach and the network model employing deep learning are the two main types of facial expression identification methods now available. Although the traditional approach is frequently utilized, its practical applicability is severely constrained.

II. STUDY ON RELATED WORKS

In 1971, Ekman first divided expressions into six basic forms, including sadness, happiness, fear, disgust, surprise, and anger. A normal expression has been added to the FER-2013 dataset and it is difficult to sort them out manually [1]. Lawrence S propose a hybrid neural-network which uses local image sampling, a self-organizing map (SOM) neural network, and a convolutional neural network in combination for human face recognition [3]. FCM Based Segmentation and Neural Network was used for Classification of Tumor in Brain MRI Images[4]. .Hoo-Chang Shin use deep convolutional neural networks to deal with the computer-aided detection problems, their model involves 5 thousand to 160 million parameters which has high requirements to computer hardware [5]. Chang constructed a convolution neural network for extracting the features of the input images [6]. Uddin proposed a new method for feature extraction known as Local Directional Position Pattern (LDPP), which provides robustness for better facial features. A depth camera-based novel method was put forward by Md. Georgescu combined the automatic features learned by the convolution neural network with the manual features calculated by the bag of the visual word, and used support vector machines as classifiers to predict the class label [7]. Du, chaoben presented a method that realizes segmentation by the multiscale convolutional neural network, edits each input image in multi-scale analysis, obtains the feature mapping of the focus and defocused regions, and it achieves the optimal fusion performance in both qualitative and quantitative aspects [8].

III. MATERIALS AND METHODS

Multi task cascaded convolution network, Light Weight Convolution Network, Pooling

3.1. MTCNN: Multi task cascaded convolution network: Multi-task Cascaded Convolutional Networks (MTCNN) is a framework developed as a solution for both face detection and face alignment. There are three stages of convolutional networks used to recognize faces and parts of face such as eyes, nose, and mouth.

3.2. The Three Stages of MTCNN:

Stage 1: The Proposal Network (P-Net) This completely convolutional network is at the first stage (FCN). A completely convolutional network does not use a dense layer as part of its architecture, which is the difference between a CNN and an FCN. Candidate windows and their bounding box regression vectors are obtained using this Proposal Network. When the goal is to recognize an object of some pre-defined class, such as faces, bounding box regression is a popular technique for predicting the localization of boxes. Following the acquisition of the bounding box vectors, some refining is carried out in order to integrate overlapping regions. This stage reduces the input size and it is easy to vectors. This stage's ultimate result is a list of all candidate windows that have been refined to reduce the number of candidates. In (fig:3.1) P-Net gives a probability of human face to any region with size of 12×12 in the input image. This will be trained first. Then using this as training set, stage two mentioned in (fig:3.2) is carried out.

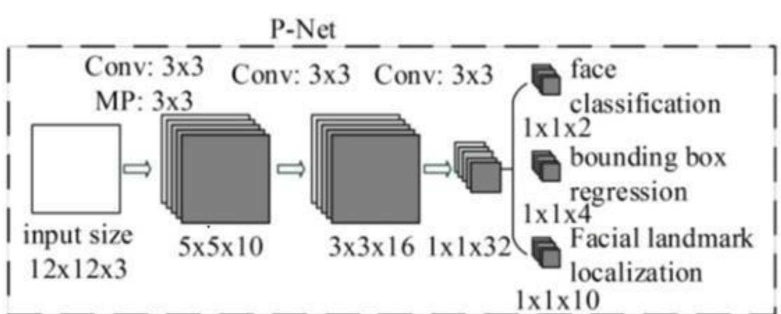


Fig :3.1 P-Net

Stage 2: **The Refine Network (R-Net)** In (fig:3.2) the Refine Network receives all candidates from the P-Net (shown in fig :3.1). Since there is a dense layer at the end of the network architecture, this network is a CNN rather than an FCN like the one before. The R-Net decreases the number of candidates even further, performs calibration with bounding box regression, and merges overlapping candidates using non-maximum suppression (NMS).

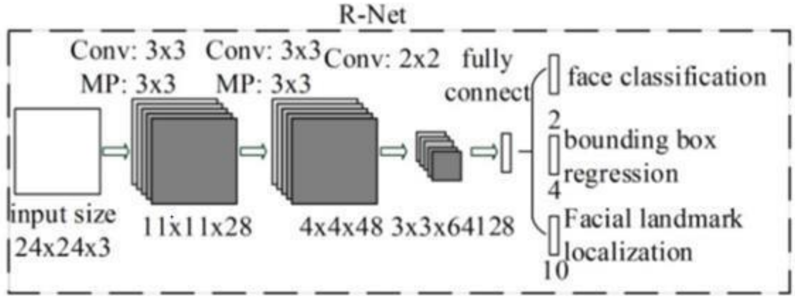


Fig :3.2 R-Net

Stage 3: The Output Network (O-Net). It produces output with five facial landmarks with more detailed description.

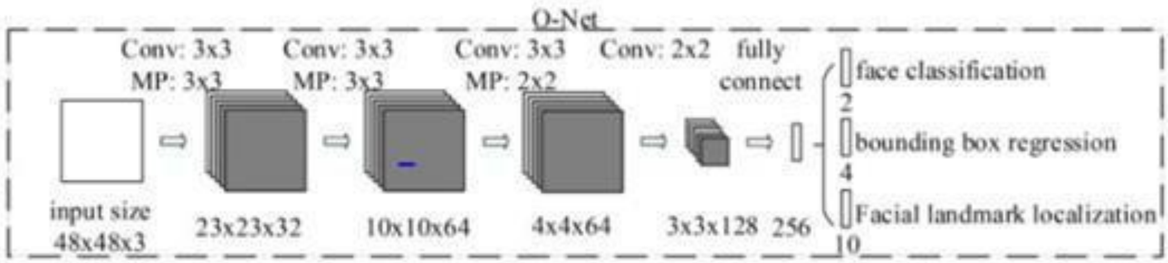


Fig :3.3 O-Net

3.3 THREE TASKS OF MTCNN: Face classification, Bounding box regression, Facial landmark localization

3.4 LIGHT WEIGHT CONVOLUTION NETWORK: The main idea of lightweight model design is to design a "network computing method", for the convolution method. At first, there were various convolution methods were introduced, then six lightweight convolutional neural networks given excellent results in recent years and the innovations of the model are discussed. Following this, the accuracy and parameters of each model on the ImageNet data set are analyzed, and the lightweight techniques of each model are compared.

3.5 POOLING LAYERS: A pooling layer is a new layer added after the convolutional layer. Specifically, after a nonlinearity (e.g. ReLU) has been applied to the feature maps output by a convolutional layer; for example, the layers in a model may look as follows: Input Image, Convolutional Layer, Nonlinearity, Pooling Layer. It is a common pattern for adding a pooling layer next to convolutional layer which issued to ordering layers within a convolutional neural network that may be repeated one or more times in a given model

IV. IMPLEMNETATION:

4.1 Use of Convolution Neural Network: CNN is the most popular way of analyzing images. CNN is different from a multi-layer perceptron (MLP) as they have hidden layers, called Convolution Layers. Two-level CNN framework is the method proposed here. The first layer recommended is background removal used to extract emotions form an image.

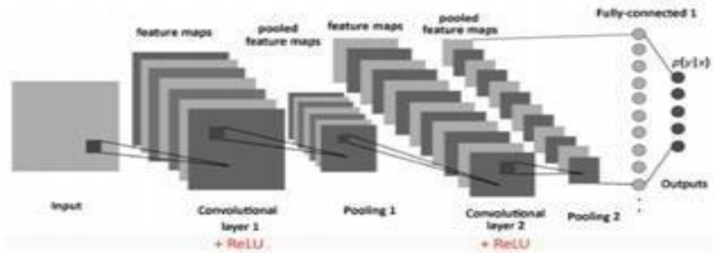


Fig. 4.1 CNN with 2 convolutional, 2 pooling and a fully connected layer

4.2 Two-Level CNN framework: The working of the two-level CNN network is shown in (fig:4.2). In the First layer, CNN network module is used to extract primary expressional vector (EV) is generated by tracking down relevant facial points of importance. EV changes with changes in expression. These vectors are passed in the CNN model form which the results obtained.

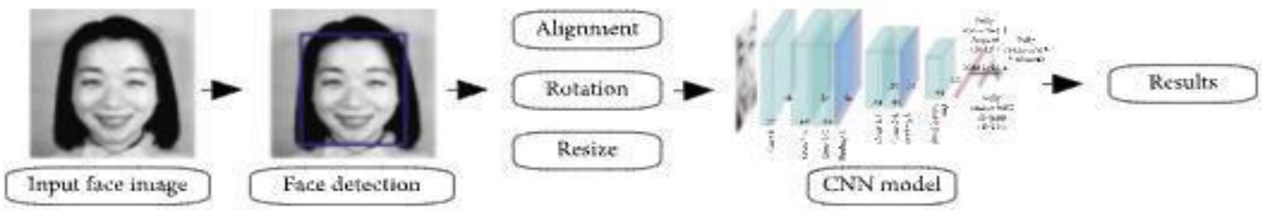


Fig. 4.2 Tracking down the facial points.

4.3 Filters in Convolution Layers Within each layer, four filters were used. The input image fed to the first part CNN generally consists of shapes, edges, textures, and objects along with the face. The edge detector and corner detector are used at the Convolution layer.

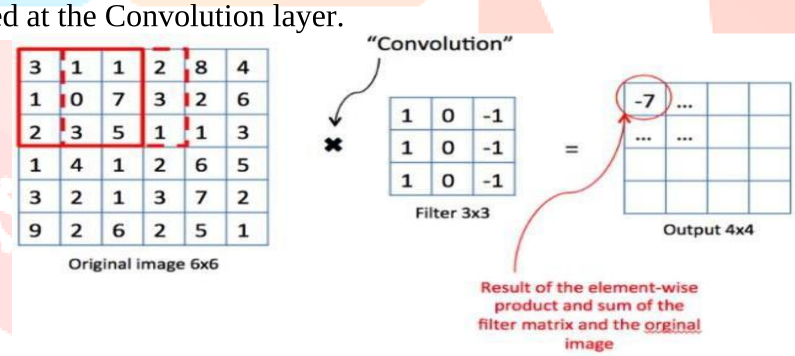


Fig :4.3 Filters

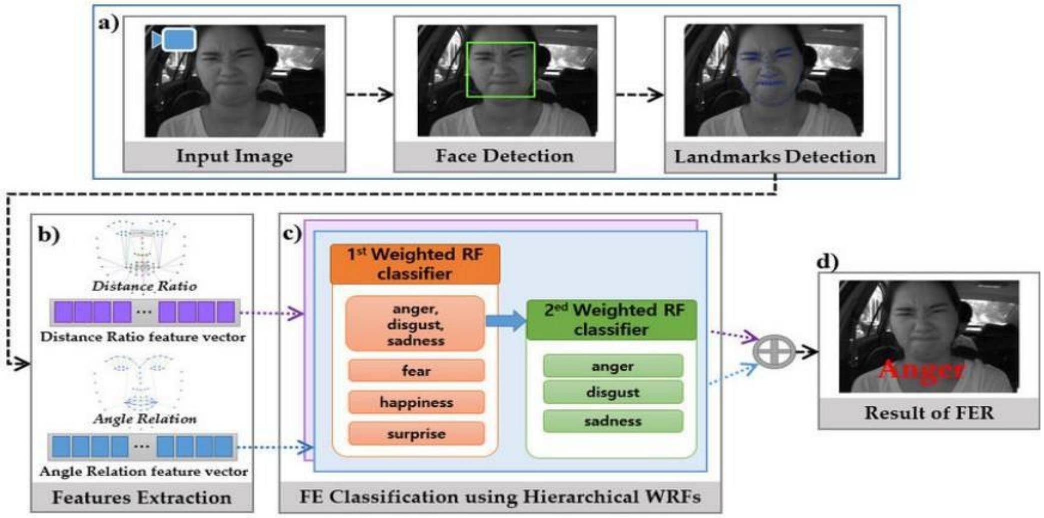
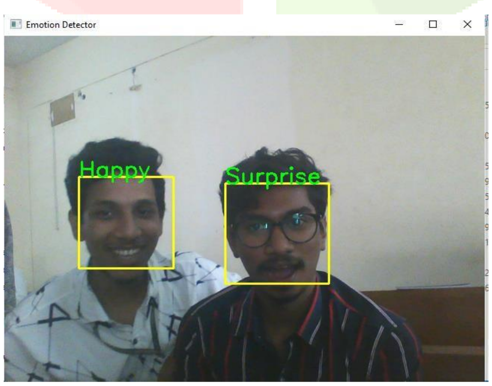


Fig :4.4 Filters

In (fig:4.4) the entire process is shown. First the image is obtained as input and passed through MTCNN which obtains the landmarks and these landmarks are weighed against test data and the most appropriate expression is displayed on the screen.

V. DISCUSSION

Our topic explored in this paper is facial emotion detection with minimal parameters. Anonymous emotion detection for online education is a great tool to evaluate and improve the online student journey. Emotional feedback is used to evaluate a school's course materials, teaching techniques, organization, and layout as students' progress through each module in real time. Find and optimize points of attraction or course stumbling blocks using genuine facial responses and engagement levels. Since it uses very few para meters it makes it easy detect expressions during time and also helps the physically disabled people a great deal. It helps us to understand their mood fluctuations.



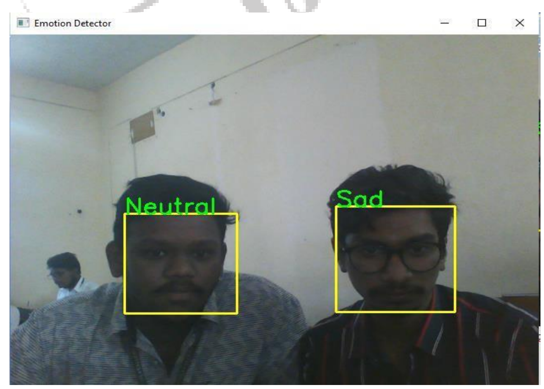


Fig :5.1 **Fig :5.2**

The fig:5.1 and 5.2 shows the results obtained by executing the program.

This is employed with minimal parameters. The expressions expressed by our face is displayed in the screen.

VI. CONCLUSION

Lightweight convolution neural network model reduces the number of parameters in the convolutional layer by eliminating the fully connected layer, combining the residual depth-wise separable convolution, and adding the l2-norm regularization term. And our model has no obviously adverse effect on detection and classification. FERC is a new method of detecting facial emotion that combines the benefits of CNN and supervised learning. Because of the unique 24-digit long EV feature matrix, the FERC algorithm operates with diverse orientations, which is a human benefit. Background elimination improved the accuracy of emotion-based applications such as lie detectors and mood-based learning for students, among other things. Although our model has achieved some results, there may be a lot of noise in the facial expressions captured in real life, such as the images with too strong or too dark lights, blurred images, most of the face is blocked, and other factors that are not conducive to detection. In order to solve this kind of problem, we need to continue our efforts. It can be further implemented for Entertainment purpose such as Recommendation of songs based on our expressions. It can also be implemented for Security purposes.

VII. REFERENCES

[1]. EKMAN, P. “Constants across cultures in the face and emotion,” j pers soc psychol, pp. 124-129. 1971.

[2]. Goodfellow I J, Erhan D, Carrier P L, et al. “Challenges in Representation Learning: A report on three machine learning contests,” Neural,2013,

[3]. Lawrence S, Giles C L. “Face Recognition: A Convolutional Neural Network Approach,” IEEE Trans. Neural Networks., pp. 98-113. 2002.

[4]. FCM Based Segmentation and Neural Network Classification of Tumor in Brain MRI Images, International Conference on Big Data and Cloud Computing, December 2019.

[5]. Shin H C, Roth H R, Gao M, et al. “Deep Convolutional Neural Networks for Computer-Aided Detection: CNN, Architectures, Dataset Characteristics and Transfer Learning,” IEEE Trans. Medical. Imaging. pp. 1285-1298. 2016.

[6] Chang T, Wen G, Hu Y, et al. “Facial expression recognition based on complexity perception classification algorithm,” 2018.

[7] Georgescu M I, Ionescu R T, Popescu M. “Local learning with deep and handcrafted features for facial expression recognition,” IEEE Access, pp. 64827-64836, 2019.

[8] Du C, Gao S. “Image Segmentation-based Multi-focus Image Fusion through Multi-scale Convolutional Neural Network,” IEEE Access, 2017.

[9] H. Ma and T. Celik, “FER-Net: facial expression recognition using densely connected convolutional network,” in Electronics Letters, pp. 184-186, 21 2 2019.

[10] W. Hua, F. Dai, L. Huang, J. Xiong and G. Gui, “HERO: Human Emotions Recognition for Realizing Intelligent Internet of Things,” in IEEE Access, pp. 24321-24332, 2019.

[11] F. Chang, A. Tuan Tran, T. Hassner, I. Masi, R. Nevatia and G. Medioni, “ExpNet: Landmark-Free, Deep, 3D Facial Expressions,” 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition, pp. 122-129,

[12] M. V. Zavarez, R. F. Berriel and T. Oliveira-Santos, “Cross-Database Facial Expression Recognition Based on Fine-Tuned Deep Convolutional Network,” 2017, pp. 405- 412,