ORIGINAL RESEARCH PAPER

Computer Science

# RECOGNIZE HUMAN'S FACIAL EMOTIONS USING CONVOLUTIONAL NEURAL NETWORKS

**KEY WORDS:** Emotion recognition, facial expression, Convolutional neural networks (CNN), Deep learning, Data acquisition, TensorFlow, OpenCV

| | |
|---|---|
| **Kejal Vadza*** | Research Scholar, Department of Computer Science, Sabarmati University, Ahmedabad, Gujarat, India *Corresponding Author |
| **Dr. Mukta Agarwal** | Head, Asst.Prof, Department of Computer Science, Sabarmati University, Ahmedabad, Gujarat, India |
| **Dr. Hitesh H. Vandra** | Principal, Shree Swami Narayan Institute of technology- Bhat, Gandhinagar, Gujarat, India |

**ABSTRACT**

Deep learning techniques are now making great strides in various fields, including computer vision. Arguably, you can prepare a convolutional neural network (CNN) model to examine images and recognize facial emotions. In this article, we create a framework to capture the feelings of undergraduate students through their faces. Our framework has her three phases of facial recognition using Haar Cascades, standardization, and emotion recognition using CNNs from the FER 2013 information base, with seven types of articulations. The results obtained show that facial emotion recognition during training is plausible. Therefore, it could help educators modify the show as indicated by studies of sentiment.

## I. INTRODUCTION

The face is the most expressive and open part of the individual [1]. I try to convey a lot of emotions without uttering words. Recognition of facial expressions distinguishes emotions from facial images and is a sign of a person's movements and personality. In the 20th century, American therapists Ekman and Friesen[2] characterized her six basic emotions common to all societies: anger, fear, contempt, pity, shock, and satisfaction.

Appearance recognition has received a lot of attention in recent years due to its implications for clinical practice, friendly machine technology, and education. According to various studies, senses play an important role in training. Going forward, educators will use tests, polls, and perceptions as sources of criticism, but these traditional strategies have always been associated with reduced productivity. Using Understudy Appearance, instructors can modify processes and materials to facilitate understudy learning.

The motivation behind this article is to help students in school by understanding a programmed framework that examines the appearance of college students as a function of a convolutional neural network (CNN), a deep learning computation widely used in image grouping. Achieving emotional awareness. It includes multi-level image processing to remove image representation.

Our framework includes three phases: face recognition, normalization, and emotion recognition. These emotions should be any of his seven emotions: fairness, anger, fear, anger, bliss, shock, and disgust. The rest of this paper is organized as follows. Section 2 provides an overview of related work. Section 3 describes the proposed system. Implementation details are presented in Section 4, followed by experimental results and discussion in Section 5.

## II. Related Works

Many researchers want to use facial emotion recognition (FER) to tackle learning environments. Tang et al. [3] proposed a framework that can decompose understudies to assess the impact of research halls. This framework consists of five phases: information gathering, face recognition, facial recognition, gaze recognition, and post-processing.

This technique uses K nearest neighbors (KNN) for ordering and Uniform Local Gabor Double Pattern Histogram Sequence (ULGBPHS) for design verification. Savva et al. [4] proposed a dynamic close-up and web application that explores the feelings of undergraduates participating in individualized classroom instruction. This application collects live accounts using webcams placed in study rooms and applies AI calculations to them.

[5] Whitehill et al. We proposed a method to perceive engagement based on a researcher's appearance. This approach uses Gabor elements and her SVM calculus to distinguish between intellectual skill as preparation for programming and engagement as understudy that is communicated. Authors received scores from recordings annotated by human reviewers. At this point, the author of [6] then used his vision computer and her AI techniques to see the effects of learning in her PC research facility at the school mechanics.

In [7], the author proposed a framework to detect and review under-learning sentiments and provide additional input for further development of e-learning environments for more salient content delivery. This framework brings her second study into placement in an e-learning environment, using moving examples of relevant data that make sense from the eye and the head. Ayvaz et al. [8] created a facial emotion recognition system (FERS) that senses the enthusiastic state and inspiration of college students in teleconferencing e-learning. This framework uses his four machine learning algorithms (SVM, KNN, Random Forest and Arrangement, and Regression Trees) and the best accuracy was obtained with the KNN and SVM algorithms.

Kimetto Al. The author of [10] perceives emotions in a virtual learning environment according to facial emotion recognition by Haar Cascades method [14], which distinguishes mouths and eyes in her JAFF information base to identify emotions. proposed a model to [11] In Chiou et al. uses innovations in remote sensor networks to develop an intelligent classroom management system that guides educators to quickly change teaching modes to save time.

## III. Proposed Work

In this part, we present a proposed framework for analyzing the appearance of investigated objects using convolutional neural network (CNN) engineering. First, the frame identifies faces from the input image, these detected faces are cropped and normalized to 48 x 48 size. At that point, these facial images will be used to contribute to the CNN. After all, the result is the result of perception (anger, joy, unhappiness, disgust, shock, or fairness). Figure 1 shows the configuration of the proposed approach.


**Fig. 1. The structure of our facial expression recognition system.**

A convolutional neural network (CNN) is an advanced artificial neural network that can distinguish visual samples from an input image with very little preprocessing, in contrast to other image assembly computations. This means that the network learns manually created channels with classical computation [19]. A key unit within a CNN layer is a neuron. They are all connected together such that the output from neurons in one layer is transformed into contributions from neurons in the next layer. Backpropagation computation is used to handle incomplete children in the output. The term convolution implies the use of infoframe channels or sections to create a component map. To be honest, a CNN model contains three types of layers as shown in Figure 2:

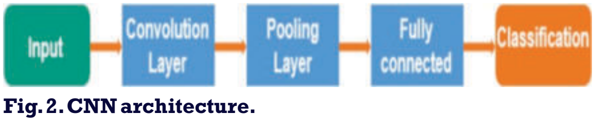



**Fig. 2. CNN architecture.**

Convolutional Network: The main layer for extracting features from the input image. In contrast to ConvNet, the main task of convolution is to separate elements from the information image. Convolution preserves spatial connections between pixels by learning image highlights using small squares with informational information [21]. Play patch elements between the two frameworks. One is the image and the other is the kernel. The recipe for convolution is Equation 1 :

$net(t,f)\ (x\ w)[t,f]\ \sum m \sum n\ x[m,n]w[t\ m,f\ n]$

where net(t, f) is the output of the next layer, x is the input image, w is the filter matrix, and * It's a convolution operation. Figure 3 shows how convolution works.

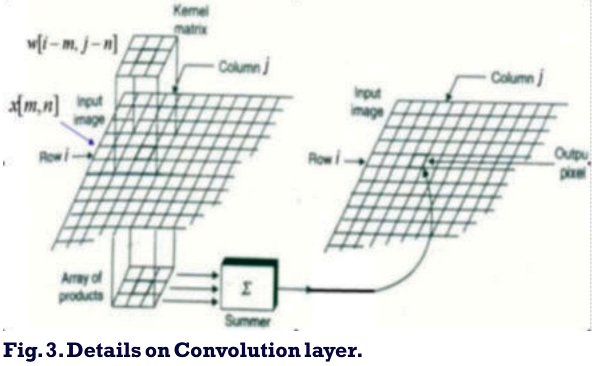



**Fig. 3. Details on Convolution layer.**

Pooling Layer: decreases the dimensionality of every detail map but holds the primary data [21]. Pooling may be of diverse sorts: Max Pooling, Average Pooling, and Total Pooling. The potential of Pooling is to constantly lower the spatial length of the facts portrayal and to make the corporation invariant to little changes, twists, and interpretations withinside the facts picture [21]. In our work, we took the restrict of the rectangular because the unmarried end result to the pooling layer as displayed in Figure 4.

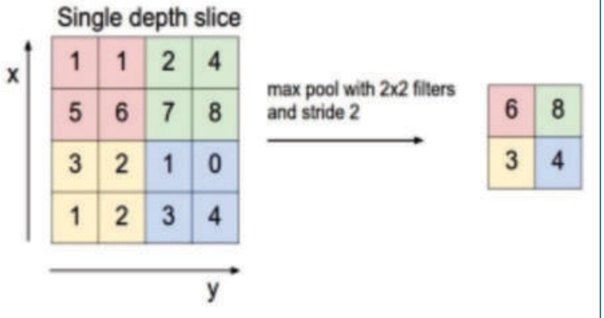



**Fig. 4. Details on Pooling layer [20].**

**Fully Connected Network:**

A traditional multi-layer perceptron with starting work at the result layer. The term "fully connected" means that every neuron in the past layer is associated with every neuron in the next layer. The motivation behind fully connected layers is to use the results of convolutional and pooling layers to classify information images into different classes depending on the prepared dataset. So folding and A pooling layer acts as a feature extractor from the information image and a fully connected layer acts as a classifier. [21].

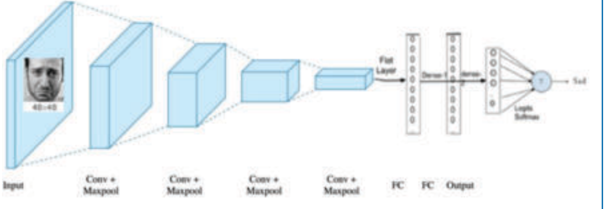



**Fig. 5. Our convolutional neural network model.**

Figure 5 shows the CNN model. It includes 4 layers of convolution with 4 layers of pooling to extract features, 2 fully connected layers, and a softmax layer with 7 class feeling. The input image is a 48x48 sized grayscale face image. For each convolutional layer, we used 3×3 channels in step 2. For pooling layer, we used max pooling layer and 2×2 part in step 2. Therefore, we used the rectified linear unit (ReLU), characterized by Eq. 2, to represent the nonlinearity of the model. This is the most commonly used actuation task these days. $R(z)\ max(0,z)$

As shown in Figure 6, R(z) is zero when z is less than zero, and R(z) is equal to z when z is greater than or equal to zero. Table I shows the model's network configuration.

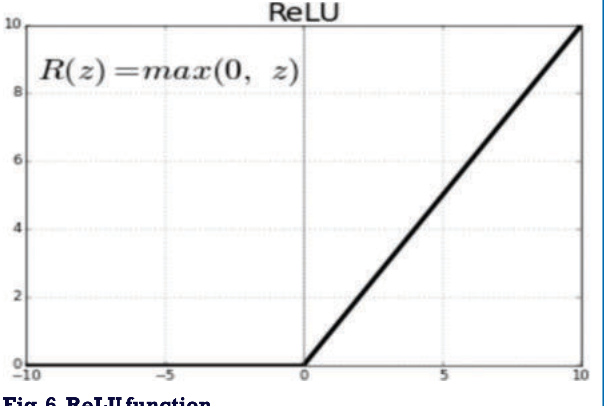



**Fig. 6. ReLU function.**

**TABLE I. CNNCONFIGURATION**

| Layer type | Size | Stride |
|---|---|---|
| Data | 48x48 | - |
| Convolution 1 | 3x3 | 2 |
| Max Pooling 1 | 2x2 | 2 |
| Convolution 2 | 3x3 | 2 |
| Max Pooling 2 | 2x2 | 2 |
| Convolution 3 | 3x3 | 2 |
| Max Pooling 3 | 2x2 | 2 |
| Convolution 4 | 3x3 | 2 |
| Max Pooling 4 | 2x2 | 2 |
| Fully Connected | - | - |
| Fully Connected | - | - |

## Iv. Implementation Details

### A. Data acquisition:

To prepare the CNN architecture, we used the FER2013 [12] data set, as shown in Figure 7. It was built using the Google Image Search API and published during the ICML 2013 challenge. Of course, information-based faces are normalized to 48x48 pixels. The FER2013 information base contains 35887 images (28709 preparatory images, 3589 approval images and 3589 test images) and 7 articulation names. The number of images for each emotion is shown in Table II.

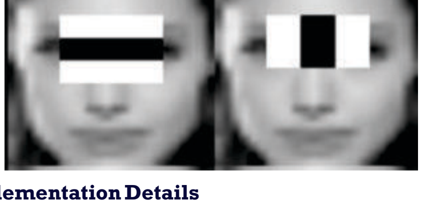

**Fig. 7. Samples from FER 2013 database.**

**Table II. The Number Of Image For Each Emotion Of Fer 2013 Database**

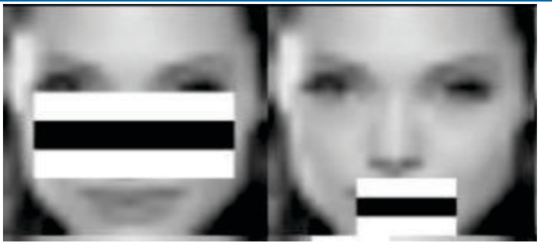
| Emotion label | Emotion | Number of image |
|---|---|---|
| 0 | Angry | 4593 |
| 1 | Disgust | 547 |
| 2 | Fear | 5121 |
| 3 | Happy | 8989 |
| 4 | Sad | 6077 |
| 5 | Surprise | 4002 |
| 6 | Neutral | 6198 |

### B. CNN Implementation:

As shown in Figure 8, we used the OpenCV library [16] to capture the live webcam body and identify the face of interest according to the Haar Cascades strategy [14]. Haar Cascades uses the method described by Freund et al. Developed Adaboost learning calculation. [15] won the Gödel Prize in 2003. Adaboost's learning computation selected a few large elements from a large set to create a convincing set of classifiers. I built a convolutional neural network model using the TensorFlow [18] Keras [17] undeniable API.



## IV. Implementation Details



## V. Experimental Results

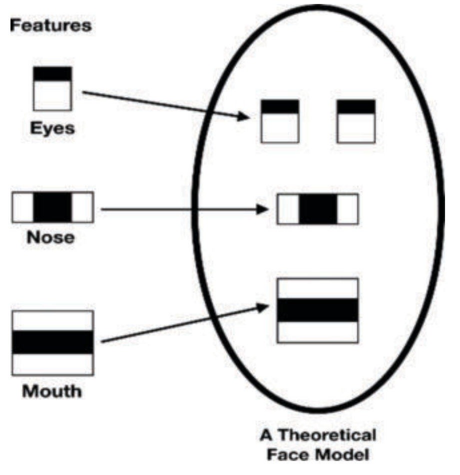



**Fig. 8. Face Detection using Haar Cascades.**

In Keras, I used the ImageDataGenerator class to perform image magnification (see Figure 9). This class allows you to modify the prepared image by panning, moving, shearing, zooming, and mirroring. The arrangement used is: rotation_range=10, width_shift_range=0.1, zoom_range=0.1, height_shift_range=0.1, and horizontal_flip=true.

*Original* *Transformed*

**Fig. 9. Image augmentation using Keras.**

At this point, we characterized the CNN model with 4 layers of convolution, 4 layers of pooling, and 2 fully mapped layers. From this point on

In the CNN model, we applied ReLU capacitance, used group normalization to standardize the activations of the reference point layer at each clump, and applied penalties at different boundaries of the model using L2 regularization. This is how we chose Softmax as our final enactment. Take a vector z of k numbers as information and standardize it to a likelihood cycle. The figure 10 shows how softmax works:

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{k} e^{z_k}} \quad for\ j = 1, \ldots, k$$

**Fig. 10. Softmax function.**

To prepare the CNN model, we split the data set into 80% preparation information and 20% test information. At this point, the model was collected using the Stochastic Angle Drop (SGD) rationalization agent. At each age, Keras checks whether the model performs better than models of previous ages. In that case, the new best model load is saved in the dataset. This allows you to stack loads directly without having to retrain for use in other situations.

## V. Experimental Results

A convolutional neural network model was created using the FER 2013 information base. It includes seven emotions: bliss, anger, misery, contempt, fairness, fear, and shock. The image of his prominent face was resized to 48x48 pixels of his and then switched to a grayscale image, which was used for CNN model posts at the time. With this in mind, his nine young professional deputies on our staff are interested in testing, two of him wearing glasses. Figure 11 shows the emotional consequences of the nine understudy roles. Expected emotional signatures are indicated by red messages, and red bars indicate possible emotions. At the age of 106, he reached an accurate pace of 70%. To assess the productivity and nature of the proposed strategy, we determined clutter grids, accuracy, validation, and F1 scores, as shown in Fig. 12 and Fig. 13, respectively. Our model is best expected to perform happily and wonderfully. In any case, it does not anticipate the emergence of fear well enough, mistaking it for a pitiful face.

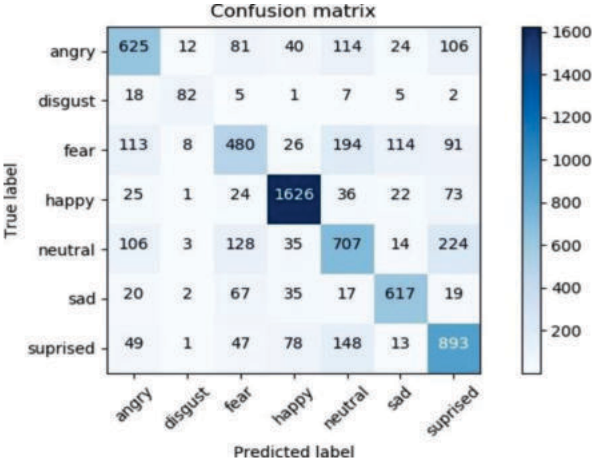



**Fig. 12. Confusion matrix of the proposed method on FER 2013 database.**

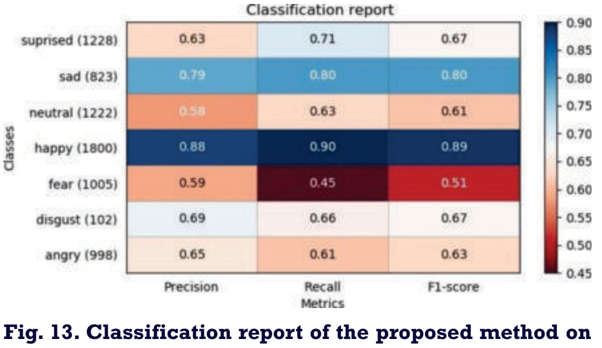



**Fig. 13. Classification report of the proposed method on FER 2013 database.**

## VI. Conclusion and future work

As shown in the figure, we used the OpenCV library [16] to capture live cases from a webcam and differentiate the faces under investigation according to the Haar Cascades strategy [14].

8. Haar Cascades uses the method described by Freund et al. Developed Adaboost learning calculation. [15] won the Gödel Prize in 2003. Adaboost's learning computation selected a few large elements from a large set to create a convincing set of classifiers. I built a convolutional neural network model using the TensorFlow [18] Keras [17] undeniable API.

## VII. REFERENCES

[1]. R. G. Harper, A. N. Wiens, and J. D. Matarazzo, Nonverbal communication: the state of the art. New York: Wiley, 1978.

[2]. P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," Journal of Personality and Social Psychology, vol. 17, no 2, p. 124-129, 1971.

[3]. C. Tang, P. Xu, Z. Luo, G. Zhao, and T. Zou, "Automatic Facial Expression Analysis of Students in Teaching Environments," in Biometric Recognition, vol. 9428, J. Yang, J. Yang, Z. Sun, S. Shan,

[4]. W. Zheng, et J. Feng, Éd. Cham: Springer International Publishing, 2015, p. 439-447.

[5]. A. Savva, V. Stylianou, K. Kyriacou, and F. Domenach, "Recognizing student facial expressions: A web application," in 2018 IEEE Global Engineering Education Conference (EDUCON), Tenerife, 2018, p. 1459-1462.

[6] J. Whitehill, Z. Serpell, Y.-C. Lin, A. Foster, and J. R. Movellan, "The Faces of Engagement: Automatic Recognition of Student Engagementfrom Facial Expressions," IEEE Transactions on Affective Computing, vol. 5, no 1, p. 86-98, janv. 2014.

[7]. N. Bosch, S. D'Mello, R. Baker, J. Ocumpaugh, V. Shute, M. Ventura, L. Wang and W. Zhao, "Automatic Detection of Learning-Centered Affective States in the Wild," in Proceedings of the 20th International Conference on Intelligent User Interfaces - IUI'15, Atlanta, Georgia, USA, 2015, p. 379-388.

[8]. Krithika L.B and Lakshmi Priya GG, "Student Emotion Recognition System (SERS) for e-learning Improvement Based on Learner Concentration Metric," Procedia Computer Science, vol. 85, p. 767-776, 2016.

[9]. U. Ayvaz, H. Gürüler, and M. O. Devrim, "USE OF FACIAL EMOTION RECOGNITION IN E-LEARNING SYSTEMS," Information Technologies and Learning Tools, vol. 60, no 4, p. 95, sept. 2017.

[10]. Y. Kim, T. Soyata, and R. F. Behnagh, "Towards Emotionally Aware AI Smart Classroom: Current Issues and Directions for Engineering and Education," IEEE Access, vol. 6, p. 5308-5331, 2018.

[11]. D. Yang, A. Alsadoon, P. W. C. Prasad, A. K. Singh, and A. Elchouemi, "An Emotion Recognition Model Based on Facial Recognition in Virtual Learning Environment," Procedia Computer Science, vol. 125, p. 2-10, 2018.

[12]. C.-K. Chiou and J. C. R. Tseng, "An intelligent classroom management system based on wireless sensor networks," in 2015 8th International Conference on Ubi-Media Computing (UMEDIA), Colombo, Sri Lanka, 2015, p. 44-48.

[13]. I. J. Goodfellow et al., "Challenges in Representation Learning: A report on three machine learning contests," arXiv:1307.0414 [cs, stat], juill. 2013.

[14]. A. Fathallah, L. Abdi, and A. Douik, "Facial Expression Recognition via Deep Learning," in 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA), Hammamet, 2017, p. 745-750.

[15]. P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, Kauai, HI, USA, 2001, vol. 1, p. I-511-I-518.

[16]. Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," Journal of Computer and System Sciences, vol. 55, no 1, p. 119-139, août 1997.

[17]. Opencv. opencv.org.

[18]. Keras. keras.io.

[19]. Tensorflow. tensorflow.org .aionlinecourse.com/tutorial/machine-learning/convolution- neural- network. Accessed 20 June 2019

[20]. S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in 2017 International Conference on Engineering and Technology (ICET), Antalya, 2017, p. 1-6.

[21]. ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/. Accessed 05 July 2019