International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 06 | June - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

FACIAL GESTURE RECOGNITION

Sumith Yadav. N
B. Tech
School of Engineering
Computer Science-(AI&ML)
Malla Reddy University, India

SuryaKiran.K
B. Tech
School of Engineering
Computer Science-(AI&ML)
Malla Reddy University, India

Suryam. M
B. Tech
School of Engineering
Computer Science-(AI&ML)
Malla Reddy University, India

Swathi Reddy. R
B. Tech
School of Engineering
Computer Science-(AI&ML)
Malla Reddy University, India

Swejana. G
B. Tech
School of Engineering
Computer Science-(AI&ML)
Malla Reddy University, India

Swetha. G
B. Tech
School of Engineering
Computer Science-(AI&ML)
Malla Reddy University, India

Guide: *Prof.Balaiah*
Assistant Professor
School of Engineering
Computer Science-(AI&ML)
Malla Reddy University,India

Abstract: Automatic emotion recognition based on facial expression is an interesting research field, which has presented and applied in several areas such as safety, health and in human machine interfaces. Researchers in this field are interested in developing techniques to interpret, code facial expressions and extract these features in order to have a better prediction by computer.Recognizing emotions from images or video is a trivial task for human eye, but proves to be very challenging for machines and requires many image processing techniques for feature extraction. Several machine learning algorithms are suitable for this job. Any detection or recognition by machine learning requires training algorithm and then testing them on a suitable dataset. This paper explores a couple of machine learning algorithms as well as feature extraction techniques which would help us in accurate identification of the human emotion.

Keywords: Python, Convolution neural networks

I. INTRODUCTION

Although there are many studies in the literature on emotion, emotion is the appearance or reflection of a feeling. Human emotions can be classified as: fear, contempt, disgust, anger, surprise, sad, happy, and neutral. These emotions are very subtle. Facial muscle contortions are very minimal and detecting these differences can be very challenging as even a small difference results in different expressions. We will compare algorithms and the feature extraction techniques .

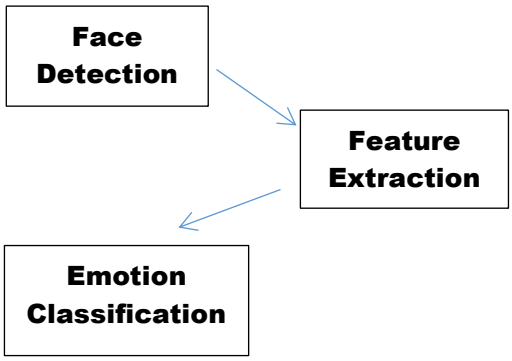
In recent years, with the popularization of deep learning, great progress has been made in image clarification. Convolution Neural Networks(CNNs) are one of the most popular deep learning architectures for image detection, clarification and segmentation.

Convolution Neural Networks built like a human brain with artificial neurons and consist of hidden layers.

Usually before extraction of features for emotion detection, face detection algorithms are applied on the image or the captured frame. We can generalize the emotion detection steps as follows:
1) Dataset pre-processing
2) Face detection
3) Feature extraction
4) Classification based on the features

In this work, we focus on the feature extraction technique and emotion detection higher accuracy rate for emotion recognition through facial expression. Then it was trained with facial dataset for classification of 7 emotion states(happy, sad, fear, angry, disgust, neutral, surprise).With the goal to improve the process of facial sentiment analysis systems, a

classification mechanism is proposed using a CNN architecture. The aim of study is to obtain deep learning model that achieve higher accuracy rate of emotion recognition through facial expression.



II.LITERATURE REVIEW

This paper introduced a multi-posed face detection and expression identification system which is more robust than the other proposed face detection system and facial expression system. This system is based on hybrid-boost multi-class learning algorithm as well as three decision rules which generates higher detection rate and lower false alarm rate.With the goal to improve the process of facial sentiment analysis systems, a classification mechanism is proposed using a CNN architecture. Due to the need of large data required for training of deep networks. The dataset which is available publically is utilized here. In the subsequent section, the features of our chosen dataset are listed out, followed by the description of our network architecture and finally the performance measures used for evaluation.

The methods about how to effectively extract expression features and recognize expression is studied in this paper. Firstly, they segment the face image from each image in the image sequence, and execute the operations of gray and scale normalization, circum rotation revision for the sub-face image. Then a hybrid feature extraction methods is presented in this paper. Experiments show that there method can recognize the six basic expressions such as angry, fear, sad, happy, neutral, disgust, neutral.

There are two different approaches commonly used in computer vision based facial expression recognition so far: recognition using 2D still images and recognition using image sequences. Approaches using image sequence often apply optimal flow analysis to the image sequence and use pattern recognition tools to recognize optimal flow patterns associated with particular facial expression. This approach requires multiple frames of images to recognize expressions and thus has limitations in real-time performance and robustness. Facial expression recognition using still images often use feature based methods for recognition and thus have fairly fast performance but the challenge in this approach is to develop a feature extraction method that works well regardless of variations in human subjects and environmental conditions.

III.Proposed Approach

In this section, we describe our proposed system to analyze students' facial expressions using a Convolutional Neural Network (CNN) architecture. First, the system detects the face from input image and these detected faces are cropped and normalized to a size of 48×48. Then, these face images are used as input to CNN. Finally, the output is the facial expression recognition results (anger, happiness, sadness, disgust, surprise or neutral). Figure 1 presents the structure of our proposed approach.

A Convolutional Neural Network (CNN) is a deep artificial neural networks that can identify visual patterns from input image with minimal pre-processing compared to other image classification algorithms. This means that the network learns the filters that in traditional algorithms were handengineered [19]. The important unit inside a CNN layers is a neuron. They are connected together, in order that the output of neurons at a layer becomes the input of neurons at the next layer. In order to compute the partial derivatives of the cost function the back propagation algorithm is used. The term convolution refers to the use of a filter or kernel on the input image to produce a feature map. In fact, CNN model contains 3 types of layers as shown in Figure 1:

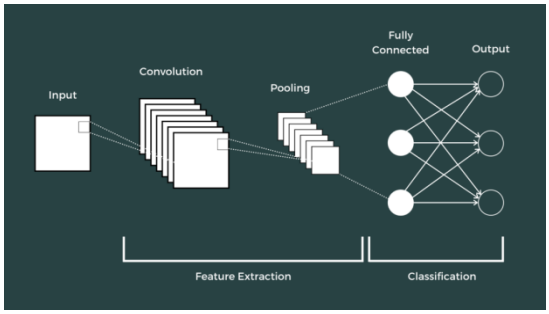


Figure 1: CNN architecture

Convolution Layer: The convolutional layer is the core building block of a CNN, and it is where the majority of computation occurs. It requires a few components, which are input data, a filter, and a feature map. Let's assume that the input will be a color image, which is made up of a matrix of

pixels in 3D. This means that the input will have three dimensions—a height, width, and depth—which correspond to RGB in an image. We also have a feature detector, also known as a kernel or a filter, which will move across the receptive fields of the image, checking if the feature is present.

Pooling layer: Pooling layers, also known as down sampling, conducts dimensionality reduction, reducing the number of parameters in the input. Similar to the convolutional layer, the pooling operation sweeps a filter across the entire input, but the difference is that this filter does not have any weights. Instead, the kernel applies an aggregation function to the values within the receptive field, populating the output array. There are two main types of pooling:

- Max Pooling
- Average Pooling

Fully-Connected Layer: The pixel values of the input image are not directly connected to the output layer in partially connected layers. However, in the fully-connected layer, each node in the output layer connects directly to a node in the previous layer. So the Convolution and Pooling layers act as Feature Extractors from the input image while Fully Connected layer acts as a classifier.

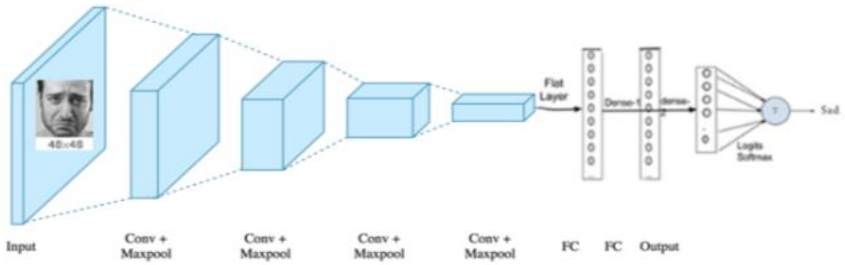


Figure 2 : Our CNN Model

IV.Methodology

A)Image pre-processing: A practical facial expression recognition system is shown in Fig.3 below. The Recognition process begins by first acquiring the image using an image acquisition device like a camera. The image acquired then needs to be preprocessed such that environmental and other variations in different images are minimized.Usually, the image preprocessing step comprises of operations like image scaling, image brightness and contrast adjustment and other image enhancement operations. In this study, an existing image database of human facial expressions is used to train and test the performance of the classifier. The images in the database have already been pre-processed and thus there is no need to apply any image pre-processing operation in this study.

B)Feature Extraction: In order to recognize facial expressions from frontal images, a set of key parameters that best describe the particular set of facial expression needs to be extracted from the image such that the parameters can be used to discriminate between expressions. This set of parameters is called the feature vector of the image and the amount of information extracted from the image to the feature vector is the single most important aspect of successful feature extraction technique.

The next step is to extract relevant features from the preprocessed images. Common feature extraction techniques include the use of Haar cascades, Local Binary Patterns (LBP), or Histogram of Oriented Gradients (HOG).

C)Model Training: After feature extraction, a machine learning model is trained using the extracted features and labeled data. Commonly used models include Convolutional Neural Networks (CNNs), Support Vector Machines (SVMs), or Decision Trees.

D)Model Evaluation: The performance of facial gesture recognition model is evaluated using a separate validation dataset. Various evaluation metrics can be used, such as accuracy, precision, recall to access the model's ability to correctly classify the expressions. Its important to choose evaluation metrics that are appropriate for the specific task and dataset characteristics. Additionally, it is advisable to consider domain-specific factors when interpreting the model.

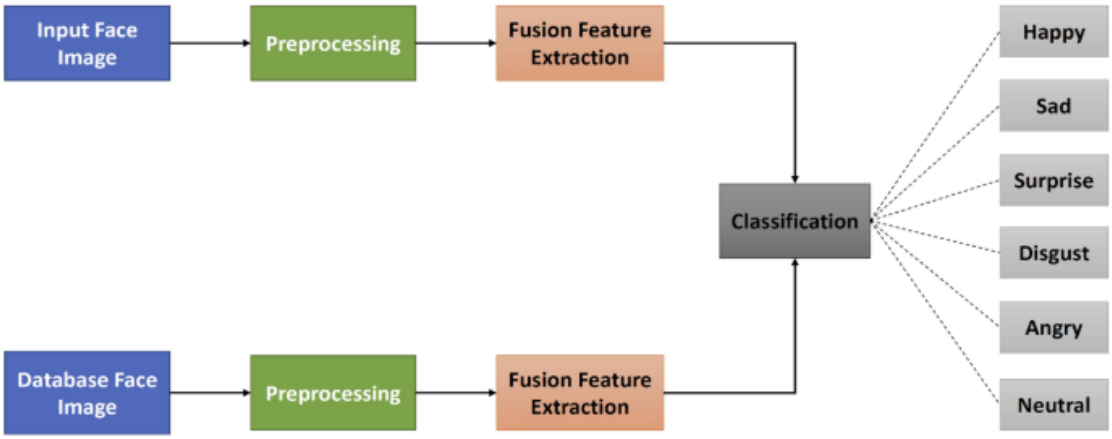


Figure 3: Facial Gesture Recognition system overview

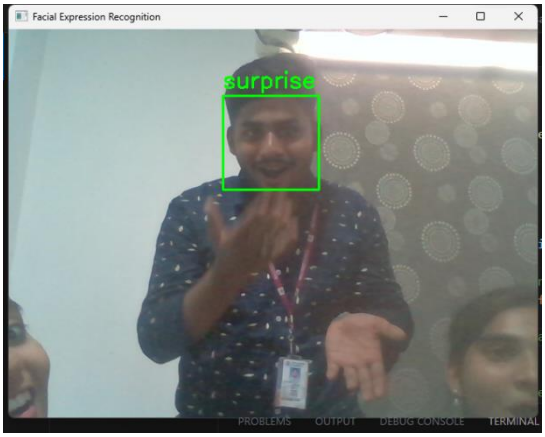
E)Emotion Classification: Here, the device classifies the picture into one of the seven universal expressions as entitled in the dataset—Happy, Sad, Anger, Surprise, Disgust, Fear, and Neutral. The training was carried out using CNN, which is a collection of neural networks. On the training range, the dataset was trained first. Before feeding it into CNN, the process of feature extraction was not performed on the results. The method followed was to experiment on the CNN with various architectures, to obtain better accuracy with the validation set. The step of classification of emotion consists of the following stages:

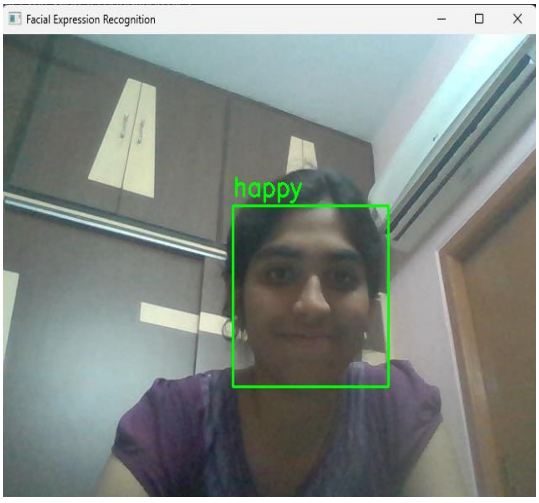
- Data Splitting
- Model training and generation
- Evaluation of model

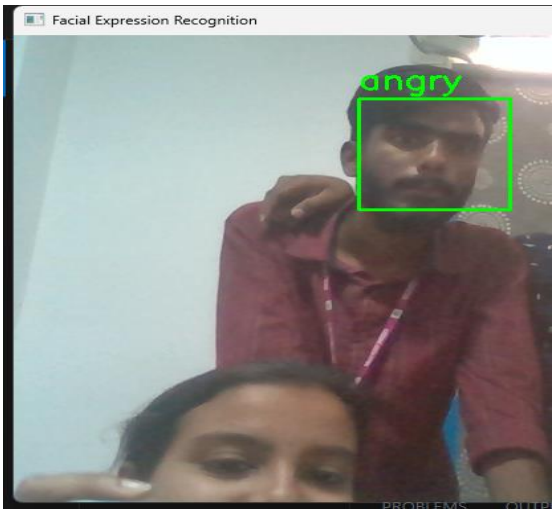
V.Experimental Results

We trained our Convolutional Neural Network model using FER 2013 database which includes seven emotions (happy, anger, sadness, disgust, neutral, fear and surprise). The detected face images are resize to pixels, and converted to grayscale images then were used for inputs to the CNN model. Thus, we students from our faculty participated in the experiment. The Figure 4 shows the emotions' results of 4 students. Working process is explained in following steps:

- The code takes input data, such as images as its input.
- It applies a pre-trained facial gesture recognition model to the input data.
- The model analyzes the facial features and patterns present in the input data.
- Based on the learned patterns, the model predicts the face expressions present in input data.
- The predicted face expressions are outputted by the code.
- Depending on the implementation, the code may display the input data along with the predicated facial expressions or provide the results in a structured format.
- Then as per the expressions text will be displayed as shown in figure4.







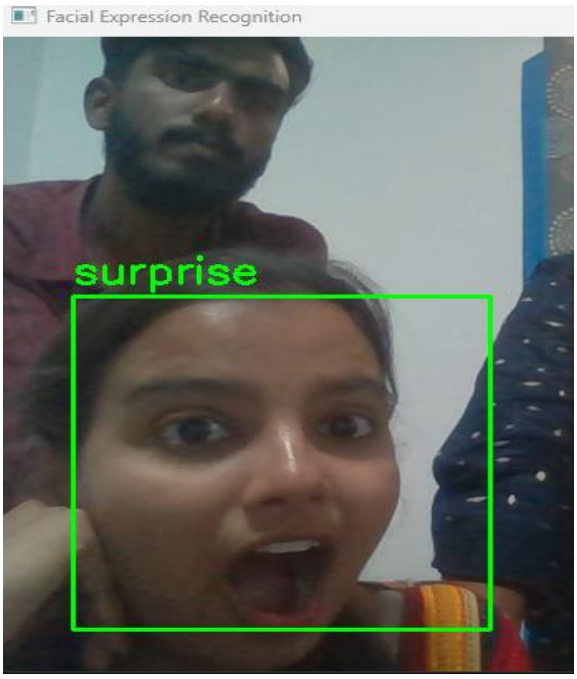


Figure 4: Student face expression results

VI. Conclusion

Facial expression recognition is a rapidly evolving field that has the potential to revolutionize many areas of human-computer interaction, including emotion recognition, human-robot interaction, and facial animation. The development of advanced computer vision and machine learning techniques has enabled accurate and real-time recognition of facial expressions, even in challenging and complex scenarios.

In conclusion, facial expression recognition has many practical applications and potential benefits, including improving the communication and interaction between humans and machines, providing insights into human emotions and behaviors, and enhancing the user experience in various applications. However, there are also challenges and ethical considerations associated with facial expression recognition, including privacy, bias, and cultural differences.

It is essential to address these challenges and develop responsible and transparent methods for developing and deploying facial expression recognition systems.Overall, facial expression recognition is an exciting and rapidly evolving field that has the potential to transform human-computer interaction and enhance our understanding of human emotions and behaviors. Continued research and development in this field are essential to realize the full potential of facial expression recognition and address its associated challenges and ethical considerations.

VII. Future Work

Facial expression recognition has come a long way in recent years, thanks to advances in computer vision and machine learning techniques. However, there is still room for improvement in this field. Here are some future enhancements for facial expression recognition such as multi-model input, continuous expression recognition, cross-cultural adaption, privacy and security and robustness to variations.

VIII. References

[1] Lopes, A. T., & de Carvalho, A. F. (2019). A survey on facial expression recognition. Artificial Intelligence Review, 52(1), 569-582. [Link:https://link.springer.com/article/10.1007/s10462-017-9585-5]

[2] Prabhakar, S., & Ghosal, S. (2020). A review on deep learning approaches for facial expression recognition. Journal of Ambient Intelligence and Humanized Computing, 11(9), 3925-3946. [Link:https://link.springer.com/article/10.1007/s12652-020-02794-2]

[3] Calvo, M., and Lundqvist, D. (2008). Facial expressions of emotion (kdef): identification under different display-duration conditions. Behav. Res. Methods 40, 109–115. doi: 10.3758/BRM.40.1.109[Link:https://github.com/vijuSR/facial-emotion-recognition]

[4] Valstar, M. F., & Pantic, M. (2010). Induced disgust, happiness and surprise: an addition to the MMI facial expression database. In 2010 IEEE International Conference on Multimedia and Expo (pp. 1037-1042). IEEE. [Link:https://ieeexplore.ieee.org/document/5583201]