# MICRO EXPRESSION DETECTION

*Jyoti Verma (IIT2019202), Mitta Lekhana Reddy (IIT2019204)*
Dhanush Vasa (IIT2019208), Gitka Yadav (IIT2019219)

IIT2019202@iiita.ac.in
IIT2019204@iiita.ac.in
IIT2019208@iiita.ac.in
IIT2019219@iiita.ac.in

## ABSTRACT

Even humans have difficulty recognizing fake facial expressions. In the meantime, the field of computer vision is exploring how to recognize facial expressions in video. In contrast to true emotions, which are expressed through facial micro-expressions, facial expressions are a spontaneous reaction to something. Despite a few attempts made at identifying micro expressions, state-of-the-art methods are not able to identify them with high accuracy. To identify micro-expressions from still images, several CNN based papers were studied in the literature. In this paper what we do is we propose a Micro Expression Recognition CNN model on the MicroExpressions dataset by analyzing the face of the person in question. The proposed CNN model has shown very promising results. As represented by the results at the end of the paper.

***Index Terms—*** Convolution Neural Network, Micro Expressions

## 1. INTRODUCTION

Expressions display what is going on in the human mind at any given moment or time. These expressions are frequently expressed by speech, gestures of a body or by facial expressions. Of all the prevailing modes of expression, facial expressions seem to be the maximum expressive way for humans to display their emotions. Facial expressions are further divided into two classes as “macro-expressions” or “micro-expressions”.

We are working on facial micro-expressions recognition which is based on the methodology i.e.CNN model on the Micro_ Expressions dataset by analyzing the face of the human. Micro-expressions are the hidden feelings or emotions that take region in the period of  and 1/25 seconds. Humans often disguise or suppress their real emotions under excessive situations due to the concern of being caught. As a result, micro-expression recognition research permits extra focus and sensitivity to diffused facial behaviors. Such abilities are beneficial to psychotherapists, interviewers and anyone who is involved in the communications. Recognizing facial expressions has several capacity packages. These expressions can be used by police departments to locate strange behavior of the criminals. Doctors can locate hidden feelings in sufferers to realize whilst extra reassurance is needed. Teachers can realize uneasiness in students and deliver a greater cautious explanation as well as business negotiators also can use happy moments to decide once they propose an appropriate price. Given the low accuracy of human recognition, an opportunity technique for recognizing micro-expressions might be extremely useful.

## 2. RELATED WORK

The development of automatic expression classifiers has progressed considerably in recent years. In some expression recognition systems, the face is classified according to prototypical emotions such as happiness, sadness, and anger [1]. Others attempt to recognize the individual muscle movements that the face can produce in order to provide an objective description of the face[2].

The Facial Action Coding System (FACS) is the best-known psychological framework for describing nearly all facial movements. The FACS system uses Action Units (AU) to classify human facial movements based on their appearance on the face[3]. A facial expression is the result of the accumulation of several AUs that make up the visible facial movement. In addition, new techniques for facial expression recognition have been developed. These include Bayesian networks, neural networks, and hidden Markov models (HMMs)[4,5]. Many of them have drawbacks, such as slow recognition rates. For accurate recognition, it is usually possible to combine two or more techniques, then extract features accordingly. Because illumination and feature extraction are crucial to each technique, pre-processing of the images is essential.

Previously, micro-expressions detection was carried out using non-spontaneous datasets like the YorkDDT. Ongoing research in the field of micro-expression detection uses spontaneous datasets which provide comparatively better results than the former method. These types of datasets are better for the micro-expression detection and interpretation as there is no posed data which makes it difficult to fake the expressions. CASME II[6] , CAS(ME)2 [7] , SAMM [8] are few such examples of spontaneous datasets.

Guo et al. proposed Local Binary Pattern-Three Orthogonal Planes (LBP-TOP) features [9] . They also proposed a nearest neighbor classifier in which Euclidean distance was used to compare the distance between the known and the unknown samples for his micro-expression detection experiment. The above mentioned method achieved the highest accuracy of 63% on the SMIC database. Increase in the facial expression activity further increased the error in this method which invoked a larger computation time to do the classification of emotions making it unsuitable for real time applications.

The precision of detecting micro-expressions in a face data set depends very much on the quality of the registered faces as these figures containing micro-expressions are used to train the CNN. CNN precision in determining the relevant micro-expressions requires to train the model. The new recent advancements in the emotions recognition was achieved by the usage of Convolutional Neural Networks (CNN) as it provides much better accuracy rate. However there is still much to improve on the CNN based algorithms due to the lack of available dataset(s).

Kim et al. proposed a new feature representation for the microexpressions where the spatial temporal states are encoded by a CNN which is then passed to the long short term memory (LSTM).recurrent neural network in which the characteristics which are temporal are being analyzed[10]. The authors evaluated their model on CASME II dataset and achieved an accuracy rate of 60.98%[11]

## 3. DATASET DESCRIPTION

Micro_Expression is the open access dataset used to train and test the proposed method. To avoid over fitting, we are using data augmentation to generate more data. Simultaneously, it can enhance the performance and outcomes of datasets. The specific methods are used to adjust resolution, and angle of the image in the micro expression dataset. For adjustment of resolution of image, specifically each image is blurred. For angle of image, each original image and blurred image is rotated between [-45°,45°] with the increment of 15°.

There are seven kinds of expressions in the Micro_Expression dataset: angry, disgust, fear, happy, neutral, sadness, surprised in both training and testing datasets. There are 7 directories allocated to each seven kinds of expressions on an overall distribution of 7600 images for training and 1860 images for testing.

| Expression | Training | Testing |
|---|---|---|
| angry | 1411 | 350 |
| disgust | 662 | 160 |
| fear | 479 | 120 |
| happiness | 1950 | 480 |
| neutral | 644 | 160 |
| sadness | 1369 | 330 |
| surprise | 1085 | 260 |

## 4. PROPOSED METHODOLOGY

The design of convolutional neural networks allows designers to encode specific qualities into the architecture, as opposed to the regular neural networks which assume the inputs are images. CNN architecture is made up of a series of layers, which include:

1. INPUT
2. CONVOLUTION LAYER
3. RECTIFIER ACTIVATION FUNCTION
4. POOLING LAYER
5. FULLY CONNECTED LAYER

A convolution has filters with a predetermined size that operates on the windowed image to extract features and an INPUT layer contains the raw pixel values of the images. To avoid unequal mapping with filter size, padding is applied to the size of the input picture. Activation function for linear elements that returns a zero value to hidden units. The pooling layer is in charge of down sampling and dimensionality reduction, which minimizes the amount of computer power needed to process data.This layer allows rotational and positional invariant features to be extracted by sliding a kernel or function over the input. Each neuron in the input is connected to each neuron in the output in the Fully connected layer, which computes the score of a specific class based on the N inputs. Based on the class with the highest score, we determine the CNN architecture's anticipated class. Fully connected layers are also called dense layers. As part of the training process, the dropout layer is applied to the convolutional neural networks in order to prevent over fitting. The scaled-up values of the inputs that are kept are used to keep the sum of the inputs constant during training. In order to refine the properties of two-dimensional features to one-dimensional characteristics, Flattening layers are placed before Fully Connected layers. Convolutional Neural Network topologies differ from one designer to the next, and the layer sequence can be changed depending on repeated assessments to achieve maximum recognition accuracy.

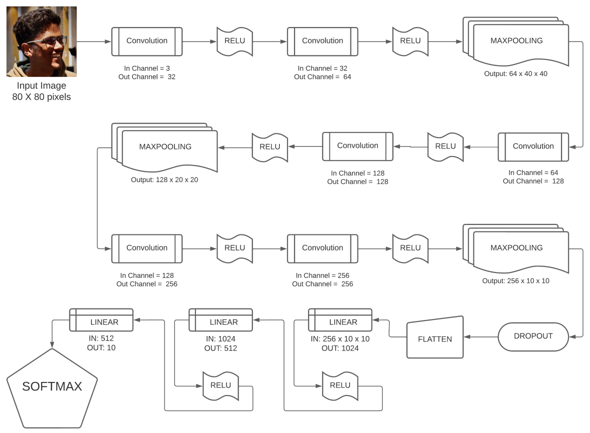
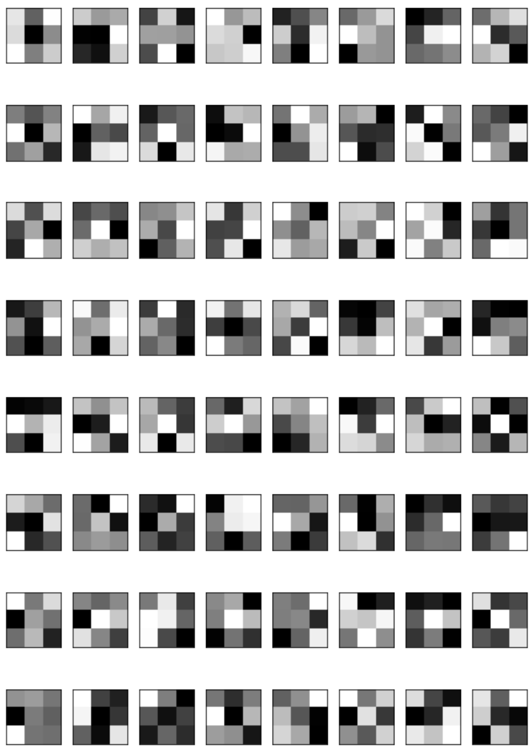



Figure 1:Micro Expression Convolution Neural Network Proposed Methodology

In order to use a training model, the input images needed to be cropped and resized to 80x80 pixels. We can modify these weights by using a back propagation algorithm using a number of epochs in order to get the final weights for the filter. Eventually we will be able to use the weight generated for the classifications phase. With this layer output, a second layer of CONVOLUTION + RELU is added with 32 filters 3X3 pixels in size, thus producing a very large output of 64. With a kernel size of 2, the output from the second CONVOLUTION + RELU layer is given to the MAX-POOLING layer function. Where this process is executed for two more times but with different filter sizes as represented in the diagram. Following the CONVOLUTION layers i will be feeding the data into a DROPOUT with the dropout rate being 0.5 which worked the best. Using the linear layer module and the data obtained, we developed a LINEAR layer feed forward network with n inputs and m outputs using flattened data. Module Ax = B. Inputs are x, outputs are b, and weight is A. This module solves the linear equation Ax = B. Where in the first linear layer we inserted n = 256*10*10 and m = 1024. Next we ran the RELU layer again and repeated the LINEAR + RELU with the linear layer having n = 1024 and m = 512. One more last time but LINEAR with SOFTMAX where Softmax is a kind of squashing function. 0 to 1 will be the output of squashing functions. By interpreting the output in this way, a probability can be determined. Hence with the final linear layer inputs are n = 512 and m = 10. Hence after all this the final out but will be 10X1. Where this generated model has been integrated with our GUI as can be seen in the Results.

## 5. RESULTS

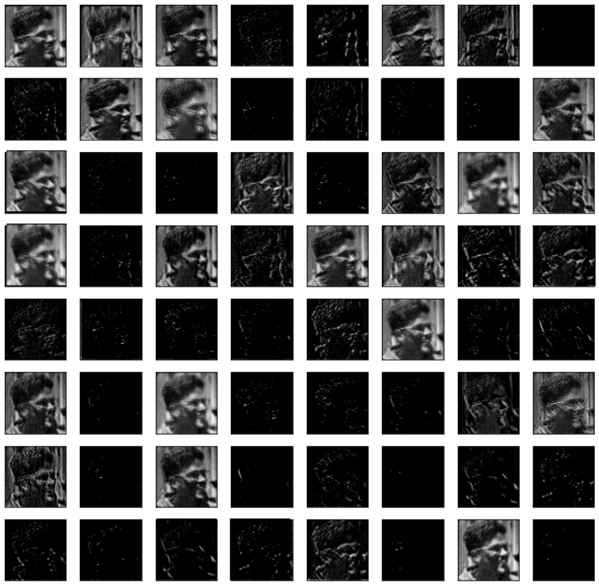
Micro Expression Convolutional Neural Networks were successfully compiled following a successful compilation process. In addition to each individual person's video and picture analysis, a test portion of the Micro_Expression dataset was used to evaluate the performance of the model. As represented in the Data description section, we had a total of 7000 images for the training portion of the dataset. The testing of the models with 0.7478 test loss resulted in an accuracy of 82.00% when combined with rigorous testing with different combinations. In the following images show an elderly Asian man who is currently sad but at a single and sudden frame we can observe that the model was able to detect that he was not only sad but he was angry. This would have been overlooked by other models if it was not looking for micro expressions. Where we were able to identify and point out that he was angry as well about some other scenario. Like this we have various other real time analysis results which have been showing promising results with our Micro Expression Convolution Neural Network.



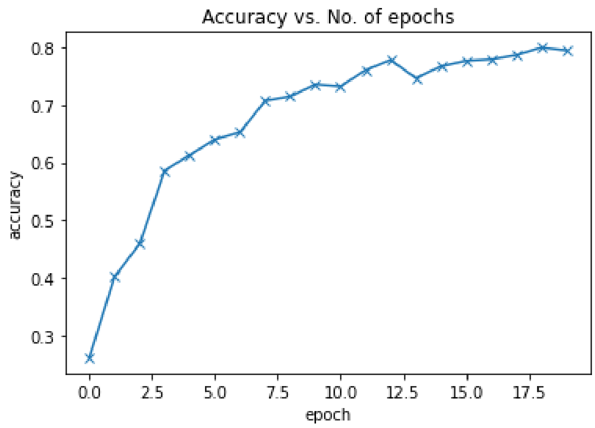
Result 1: 8X8 grid of the 3X3 filters

Result 2: Outputs after the first 2 Convolution Layers

Result 3: Outputs after the 3rd and 4th Convolution Layers



Result 4: Outputs after the 5th and 6th Convolution Layers




Result 5: Epoch vs Accuracy


Figure 2: Representing the Real time output of our Micro Expression CNN work to detect

What we can observe from figure 2 is that for the entire duration of the clip the Asian man was distort about some incident but at one sudden instance we can observe that or model was able to see the inner anger being shown by the grinding of the teeth and squinting of the eyes.

## 6. CONCLUSIONS

To generalize more effectively to situations that a model may encounter in production, the proposed Convolutional Neural Network(CNN) on Micro_Expression dataset has been augmented with additional data. The proposed system and CNN architecture are evaluated by tuning various parameters of CNN to enhance the facial micro expression recognition accuracy of the designed system. Among the future work, there are plans to work with 3D images and videos on top of working on static images to obtain systems capable of recognizing micro-expressions with a higher level of accuracy.

## 7. CITATIONS AND REFERENCES

[1] G. Littlewort, M. Bartlett, I. Fasel, J. Susskind, and J. Movellan. Dynamics of facial expression extracted automatically from video. Image and Vision Computing, 24(6).

[2] M.S. Bartlett, G. Littlewort, M.G. Frank, C. Lainscsek, I. Fasel, and J.R. Movellan. Automatic recognition of facial actions in spontaneous expressions. Journal of Multimedia.

[3] P. Ekman, W. Friesen, Facial Action Coding System: A Technique for the Measurement of Facial Movement, Consulting Psychologists Press.

[4] Cohen, Ira, et al. "Evaluation of expression recognition techniques." Image and Video Retrieval. Springer Berlin Heidelberg. 184- 195.

[5] Padgett, C., Cottrell, G.: Representing face images for emotion classification. In: Conf. Advances in Neural Information Processing Systems. 894900.

[6] W. Yan et al. "CASME II: An Improved Spontaneous Micro-Expression Database and the Baseline Evaluation". In: PLoS ONE 9 .

[7] Fangbing Qu et al. "CAS(ME)2: A Database of Spontaneous Macro-expressions and Micro-expressions". In: HCI.

[8] A. K. Davison et al. "SAMM: A Spontaneous Micro-Facial Movement Dataset". In: IEEE Transactions on Affective Computing 9.1, pp. 116–129.

[9] Y. Guo et al. "Micro-expression recognition based on local binary patterns from three orthogonal planes and nearest neighbor method". In: International Joint Conference on Neural Networks (IJCNN). pp. 3473–3479.

[10] D. H. Kim, Wissam J. Baddar, and Yong Man Ro. "Micro-Expression Recognition with Expression-State Constrained Spatio-Temporal Feature Representations". In: Proceedings of the 24th ACM international conference on Multimedia.

[11] Walied Merghani, Adrian K. Davison, and Moi Hoon Yap. "A Review on Facial Micro-Expressions Analysis: Datasets, Features and Metrics". In: ArXiv abs/1805.02397 .