Feeling Recognition by Facial Expression Using Deep Learning

Dr. Shalini Gupta¹, Dr. Shubha Jain²

¹Associate Professor, Computer Science and Engineering Department, Kanpur Institute of Technology, KANPUR
²Associate Professor, Computer Science and Engineering Department, Axis Institute of Technology and Management KANPUR
shalinilily2003@gmail.com shubhj@rediffmail.com

Abstract: The visible appearance of the emotion state, personality, purpose, psychological feature activity and psychopathology of an individual is the facial expression. This plays an outgoing role in social affairs. Automatic recognition of facial expressions will be a vital part of natural human-machine interfaces. It should even be applied in activity science and in clinical apply. Fellows in nurturing automatic facial features recognition system must perform detection and placement of faces during a disordered scene, facial feature extraction and facial features classification. Emotion recognition by facial features utilizes a Deep Learning system, which is enforced victimization Convolution Neural Network (CNN). The CNN model of the project relies on LeNet design. Kaggle facial features dataset with seven facial features labels as fear, anger, happy, surprise, sad, neutral and disgust is employed during this project. The system achieved 60.37 accuracy.

Keywords- Facial Expression Recognition, Convolutional Neural Network, Deep Learning, keras

1. Introduction

A facial expression is the visible display of the personality, psychopathology, cognitive activity, and intentions of a person and plays an outgoing role in interactive relations. Human facial expressions are classified into seven basic emotions: fear, anger, happy, surprise, sad, neutral and disgust. The facial emotions are expressed through prompting specific sets of facial muscles. A plentiful amount of information about our state of mind is contained in sometimes subtle, however complex, signals in an expression.

Automatic recognition of facial expressions can be an important constituent of natural human-machine interfaces. This may also be used in behavior science and in clinical training too. This showed progress in recent decades, however, it has been considered for a long period of time. Due to the complexity and varieties of facial expressions, identifying facial expression with a high accuracy remains to be difficult even though much progress has been made.

In day to day life, humans commonly identify emotions by characteristic features, which are presented as a part of a facial expression. For example, happiness is definitely associated with a smile, which is an upward movement of the corners of the lips. In the same way, other emotions are also characterized by other deformations, which is a characteristic of a particular expression. The problems nearby the representation and classification of static or dynamic characteristics of these deformations of face pigmentation is addressed by research into automatic recognition of facial expressions.

A convolutional neural organization (CNN, or ConvNet) is a sort of feed-forward counterfeit neural organization in AI. The availability design between its neurons is enthused by the association of the
creature visual cortex. Individual cortical neurons answer to boosts in a confined locale of room which is known as the responsive field. The responsive fields of various neurons to some degree cover such that they spread the visual field. Inside its open field, the response of an individual neuron to improvements can be assessed numerically by a convolution activity. Convolutional networks were propelled by natural cycles. These organizations might be considered as deviation of multilayer perceptron, which are intended to utilize least measure of pre-handling.

Convolution networks have broad applications in video and image detection suggestive systems and natural language processing. The convolutional neural network is also known as change in variant or space invariant artificial neural network (SIANN). This is named based on its translation in variance characteristics and its shared weights architecture.

LeNet is the example of one of the very first convolutional neural networks. These are the networks which helped in propelling the field of Deep Learning. This innovative work by Yann Le Cun[3][5] was named LeNet5. It was used mostly for character recognition tasks such as reading digits, zip codes etc. The basic architecture of LeNet can be shown as Fig 1.1.

![Fig. 1 A model of CNN](image)

There are four main operations in the Convolution Neural Network as shown above.

2. Objective

The objective of the paper is to design a Convolutional Neural Network deep learning model which in existent time using the facial expressions, can detect emotion of humans. Facial emotion recognition can be used as a base for many real-time applications, which is the one of the useful tasks of it. As humans can easily identify the emotion of other humans with minimum effort, in same way, Automatic detection of emotion of a human face is important due to its use in real-time applications.

3. Literature Review

Two one of a kind processes used for facial features popularity comprise one of a kind methodologies. Dividing the face into separate movement devices or preserving it intact for in addition processing seems to be the primary and the essential difference among the primary processes. In each of those processes, one of kind methodologies, specifically the ‘Geometric based’ and the ‘Appearance-based’ parameterizations may be used.

Rather than following spatial focuses and utilizing development and situating boundaries which differ inside time, shading (pixel) data of related locales of face are treated in Appearance Based Parameterizations. This is so as to get the boundaries that are attempting to shape the component vectors.
Unmistakable highlights, for example, Haar wavelet coefficients, Gabor, along with include extraction and choice strategies, for example, PCA, LDA, and Ada help are utilized inside this structure.

For a classification problem, algorithms like Neural Network Support vector Machine, NAive BAyes and Deep Learning are used.

4. Previous Work

On Robust facial expression recognition using local binary patterns, a model has been proposed[1]. Also, a Facial expression recognition system has been built [2].

Many done the work on "Subject independent facial expression recognition with robust face detection using a convolutional neural network"[4].

A Facial expression recognition system has been built upon recent research to classify images of human faces into discrete emotion categories using convolutional neural network [6]. Facial Expressions Recognition system also has been developed using Convolutional Neural Networks based on Torch model [7].

5. Dataset

The informational index from a Kaggle Facial Expression Recognition Challenge (FER 2013) is utilized for the preparation and testing. It contains pre-trimmed, 48-by-48-pixel grayscale pictures of appearances each marked with one of the 7 feeling classes: (0=Disgust, 1=Happy, 2=Sad, 3=Fear, 4=Angry, 5=Neutral, 6=Surprise). Dataset has preparing set of 35887 facial pictures with outward appearance labels[8]. Dataset contains two segments, "feeling" and "pixels".

The "feeling" segment contains a numeric code extending from 0 to 6, comprehensive, for the feeling that is available in the picture. The "pixels" section contains a string encompassed in cites for each picture. The substance of this string a space-isolated pixel esteems in line significant request. Document contains just the "pixels" section and your errand is to anticipate the feeling segment. The preparation set comprises of 28,709 models. The test set utilized for the comprises of 3,589 models. The approval comprises of another 3,589 models.

| Table 1 Emotion labels in the dataset |
|--------------------------------------|
| 0  | 547 images | Disgust |
| 1  | 8989 images | Happy |
| 2  | 6077 images | Sad |
| 3  | 5121 images | Fear |
| 4  | 4593 images | Angry |
| 5  | 6198 images | Neutral |
| 6  | 4002 images | Surprise |

6. Architecture of CNN

The convolutional neural organization's engineering contains an info layer, some convolutional layers, some completely associated layers, and a yield layer. CNN is designed with some amendments on LeNet Architecture [7].

It has 6 layers without seeing input and output. The architecture of the Convolution Neural Network used here is shown in the figure 1.2.
Fig 2: Modified CNN Architecture

- **Input Layer**

The input layer has pre-determined, fixed dimensions, so the image before feeding into the layer, must be pre-processed. From Kaggle data set, normalized gray scale images of size 48 X 48 pixels are used for training, confirmation and testing. Laptop webcam images are also used for testing purpose, in which face is detected and cropped using Open CV Haar Cascade Classifier and then normalized.

- **Convolution and Pooling (ConvPool)Layers**

Based on batch processing, convolution and pooling is done. Each bunch has N pictures and CNN channel loads are refreshed on those groups. Every convolution layer takes picture bunch contribution of four measurement N x Color-Channel x width x stature. Highlight guide or channel for convolution are likewise four dimensional (channel width, channel tallness, Number of highlight maps in, number of highlight delineates). Four-dimensional convolution is determined between picture group and highlight maps in every convolution layer. After convolution only changing parameter is image height and width.

\[
\text{New image width} = \text{old image width} + 1 - \text{filter width} \\
\text{New image height} = \text{old image height} + 1 - \text{filter height}
\]

For dimensionality reduction, after each convolution layer down sampling / sub sampling is done. This cycle is called Pooling. Max pooling and Average Pooling are two notable pooling techniques. In this paper max pooling is done after convolution. Pool size of (2x2) is taken, which parts the picture into network of squares, which are every one of size 2x2 and takes limit of 4 pixels. In the wake of pooling just stature and width are influenced.

In the architecture, two convolution layer and pooling layer are used. From the start convolution layer size of info picture cluster is Nx1x48x48. The size of picture bunch is N, number of concealing channel is 1 and both picture height and width are 48 pixels. Convolution with incorporate guide of 1x20x5x5 results picture bunch of size Nx20x44x44. After convolution pooling is done with pool size of 2x2, which results picture bunch of size Nx20x22x22. This is followed by second convolution layer with incorporate guide of 20x20x5x5. This results picture gathering of size Nx20x18x18. This is trailed by pooling layer with pool size 2x2, which results picture bundle of size Nx20x9x9.

- **Fully ConnectedLayer**

This layer is spurred by the manner in which neurons communicate signals through the mind. It gets an
enormous number of information highlights and converts highlights through layers associated with teachable loads. Two shrouded layers of size 500 and 300 unit are utilized in completely associated layer. The loads of these layers are gifted by the regressive proliferation of its mistakes and forward engendering of preparing data. The distinction among forecast and genuine worth are assessed, and the weight alteration expected to each layer before is determined by back in the back-spread method. By tuning the hyper-boundaries, for example, network thickness and learning rate, the unpredictability of the design too a straining rate can be controlled. The energy, learning rate, rot and regularization boundary goes under Hyper-boundaries for this layer.

The size of yield is Nx20x9x9 from the second pooling layer and size of information is Nx500of first shrouded layer of completely associated layer. In this way, yield of pooling layer is taken care of to first shrouded layer which is straightened to Nx1620 size. First concealed layer's yield is taken care of to second layer. Second concealed layer's size is Nx300 whose yield is taken care of to yield layer of size equivalent to number of outward appearance classes.

- **OutputLayer**

The Output from the second shrouded layer is associated with yield layer having seven unique classes. Yield is gotten utilizing the probabilities for every one of the seven-class utilizing Softmax initiation work. The anticipated class is the class with the most elevated likelihood.

7. Results And Analysis

CNN architecture for facial expression recognition is implemented using Python. Along with Python programming language, Numpy, Theano and CUDA libraries were used.

For both convolution layer, training image batch size is taken as 30 and filter map is of size 20x5x5. To validate the training process, validation set was used. In last batch of every epoch in validation cost, validation error, training cost, training error are calculated. The image set and corresponding output labels are taken as input parameters for training.

The training process apprised the weights of feature maps and hidden layers which are based on hyper-parameters such as momentum, learning rate, decay and regularization. In this system batch-wise learning rate was used as 10e-5, momentum as 0.99, regularization as 10e-7 and decay as 0.99999.

6000 pictures were utilized to test the model. The classifier gave 56.77 % exactness. The disarray lattice for seven outward appearance classes is appeared in Table 1.
Table 2 Confusion matrix for facial expression recognition

| ACTUAL | Disgust | Happy | Sad | Fear | Angry | Natural | Surprise |
|--------|---------|-------|-----|------|-------|---------|----------|
| Disgust | 10      | 28    | 4   | 2    | 7     | 2       | 4        |
| Happy   | 30      | 0     | 20  | 718  | 40    | 15      | 72       |
| Sad     | 76      | 0     | 2   | 43   | 325   | 4       | 143      |
| Fear    | 52      | 2     | 194 | 18   | 113   | 42      | 78       |
| Angry   | 246     | 1     | 45  | 25   | 79    | 15      | 61       |
| Natural | 50      | 0     | 42  | 42   | 112   | 10      | 351      |
| Surprise| 12      | 0     | 36  | 20   | 12    | 313     | 22       |

A disarray framework is a strategy for summing up the exhibition of arrangement calculation. Grouping precision alone can be misdirecting in the event that you have an inconsistent number of perceptions in each class or on the off chance that you have multiple classes of dataset.

Figuring a disarray network can give a superior thought of what arrangement model is getting right and what kinds of mistakes it is making.

Above tables showed that this model predicts for anger 246 times true out of 457 images whereas it predicted 1 images of anger class as disgust, 45 images of anger as fear and so on for other motions. The true predicted values are the values on the diagonal of matrix for each emotion like for disgust it is 28, for fear it is 194 out of total no of images of that particular emotion.

On the basis of confusion matrix we calculate the error rate and accuracy of our model by finding recall, precision and F1-Score.

The precision, recall and F1-score of each facial expression class is shown in Table 2.

### Table 3 Precision, Recall and F1-Score

|        | Precision | Recall | F1-score |
|--------|-----------|--------|----------|
| Disgust| 0.95      | 0.99   | 0.97     |
| Happy  | 0.69      | 0.68   | 0.68     |
| Sad    | 0.43      | 0.39   | 0.40     |
| Fear   | 0.45      | 0.38   | 0.39     |
| Angry  | 0.39      | 0.42   | 0.41     |
| Natural| 0.45      | 0.49   | 0.47     |
| Surprise| 0.69     | 0.65   | 0.67     |
| Average| 0.57      | 0.57   | 0.57     |

The general exactness and review are 0.57 and 0.57 individually. The model performs truly well on arranging positive feelings bringing about moderately high accuracy scores for cheerful and amazed. Disturb as most elevated accuracy and review as 0.95 and 0.99 as pictures in this class were over examined to address class irregularity. The exactness of Happy is 0.69 and recall is 0.68. The equivalent can be clarified by the models (6500) in the preparation set.

In addition, the precision of surprise is 0.69 and recall is 0.65 but this is with least examples in the training set. Very strong signals must be there in the surprise expressions.

On average, model performance seems to be weaker across negative emotions. Mostly, the emotion sad
has a low precision which is only 0.43 and recall is 0.39. The model usually classified fear, angry and neutral as sad which is wrong classification. In addition, the model it is puzzled most while required to predict neutral and sad faces because these two emotions are perhaps the least expressive ones (excluding faces with crying expressions).

The overall F1-score is also 0.57. F1-score is most noteworthy for disgust due to oversampling of pictures. Happy and surprise higher F1-score as 0.69 and 0.67 separately. Fear has least F1-score as 0.39 and miserable, outrage and unbiased additionally have low F1-score.

8. Conclusion

The outward appearance acknowledgment framework introduced in this examination work contributes a strong face acknowledgment model which depends on the planning of social qualities with the physiological biometric attributes. The natural qualities of the human face with pertinence to different looks, for example, satisfaction, misery, outrage, dread, nauseate and shock are related with mathematical structures. This reestablished as base coordinating format for the acknowledgment framework.

Here a LeNet design based six layer convolution neural organization is actualized to group human outward appearances for example glad, shock, miserable, dread, appall, outrage and unbiased. The framework has been assessed utilizing Accuracy, Precision, Recall and F1-score. The classifier prevailing with exactness of 60.37 %, accuracy of 0.57, review 0.57 and F1-score0.57.

9. Future Scope

In the future work, the model can be extended to color images also. This will allow examining the effectiveness of pre-trained models such as AlexNet or VGGNet for facial emotion recognition.

References

[1] McOwan, P. W.& Shan, C., Gong, S., & (2005, September). Robust facial expression recognition using local binary patterns. In Image Processing, 2005. ICIP 2005. IEEE International Conference on (Vol. 2, pp. II-370).IEEE.
[2] Chibelushi, C. C., & Bourel, F. (2003). Facial expression recognition: A brief tutorial overview. CVonline: On-Line Compendium of Computer Vision, 9.
[3] "Convolutional Neural Networks (LeNet) – DeepLearning 0.1 documentation". DeepLearning 0.1. LISA Lab. Retrieved 31 August 2013.
[4] Matusugu, Masakazu; Katsuhiko Mori; Yusuke Mitari; Yuji Kaneda (2003). Neural Networks. 16 (5): 555–559. doi:10.1016/S0893-6080(03)00115-1. Retrieved 17 November 2013.
[5] LeCun, Yann. "LeNet-5, convolutional neural networks". Retrieved 16 November 2013.
[6] Raghuvanshi, Arushi, and Vivek Choksi. "Facial Expression Recognition with Convolutional Neural Networks." Stanford University, 2016.
[7] Alizadeh, Shima, and Azar Fazel. "Convolutional Neural Networks for Facial Expression Recognition." Stanford University, 2016.
[8] https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge (accessed in June, 2020)