IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

# A Convolutional Neural Network Driven Method Of Facial Sentiment Analysis

[1]Ram Murti Rawat, [2]Yatharth Yadav, [3]Vipin Ranga, [4]Vikas Kumar

[1]Assistant Professor, Department of Computer Engineering, Delhi Technological University, Delhi
[2]Undergraduate Student, Department of Computer Engineering, Delhi Technological University, Delhi
[3] Undergraduate Student, Department of Computer Engineering, Delhi Technological University, Delhi
[4] Undergraduate Student, Department of Computer Engineering, Delhi Technological University, Delhi

***Abstract:*** Human facial expressions are an integral and straightforward means of displaying sentiments. Automatic analysis of these unspoken sentiments has been an interesting and challenging task in the domain of computer vision with its applications ranging across multiple domains including psychology, product marketing, process automation etc. This task has been a difficult one as humans differ greatly in the manner of expressing their sentiments through expressions. Machine learning, specifically deep learning has been instrumental in making breakthrough progress in many fields of research including computer vision. Through this research paper, we hereby introduce a convolutional neural network (CNN) implemented architecture that tackles this problem of facial sentiment analysis. For training and testing purposes we have made use of the FER-2013 public dataset. This task has been undertaken in a series of steps namely, preprocessing of the data followed feature extraction and finally classification by our trained model network. The results of our experiment have been very encouraging and are an improvement in the domain of automated analyzing of facial sentiments.

***Index Terms*** **- Deep Learning, Facial Sentiment Analysis, Convolutional Neural Network, Face detection, Network Architecture.**

## I. INTRODUCTION

Sentiments are an important part of any inter-personal interaction. These can be conveyed through different ways, such as via facial expressions [19], [8], [13], speech [5], [22], gesture and even stance. Among these choices, facial expressions are the most visible and information rich, and can be exploited for sentiment analysis. Additionally, it is comparatively easier to collect and process faces than other means of expressions.

A facial expression is a intricate execution of the facial muscles, which expresses the sentiments or emotional state of the subject to anyone observing it [1]. In basic terms, expression are a message about what a person is feeling inside. For these reasons, a human-computer interaction system [6] for an automatic face recognition or facial sentiment analysis has attracted increasing focus from researchers in psychology, animation [8], human-computer interface [6], linguistics, neuroscience and medicine [3], [4] and security [5].

Today, computer assisted analysis of face and it's expressions is an emerging field. Sentiment analysis consists of associating an emotion to facial image. So the objective is to determine from the face, the internal sentiments of a person. An automated facial sentiment analysis system plays an important role in simplifying human machine interaction. However, this is not simple to do.

For some time now, by making use of deep learning and convolutional neural networks (CNN) [10] which have been inspired from the study of biology, many features of facial expressions can be extracted and analysed for decent sentiment analysis [7], [18]. Motivated by this, through this research, we provide a deep learning based model for facial sentiment analysis. We provide a convolutional network architecture by which, sentiments can be classified into a standard set of seven emotions [9]: Disgust, Fear, Anger, Surprise, Happiness, Sadness and Neutral based on facial features.

The rest of the paper follows the following schema. The previous related works in the field have been mentioned in section II. This is followed by our proposed approach with related background explanations in III. Section IV has the experiment and results followed by conclusions in section V. Lastly, references for the paper have been mentioned at the last.

## II. RELATED WORKS

In an iconic work in emotion analysis by Paul Ekman [9], six principal sentiments were identified which are, sadness, anger, surprise, happiness, fear and lastly disgust. In later works, neutral was added in this list making them the seven basic human emotions for analysis purposes.

Previous works in the field of sentiment analysis have relied on a bidirectional approach. In the first step, features are identified in the raw images which are then followed by a classification methodology (like SVM, vector field convolution VFC [12], decision trees, naïve bayes etc.). These approaches worked acceptably on simpler datasets of images in a controlled environment, but faced limitations on more challenging datasets that had more intra-class diversity. Various image features for facial feature extraction have been used in the previous papers like local binary pattern (LBP) [15], histogram oriented gradients [20], quantization of local phase and scale-invariant feature transform (SIFT) [21].

Recent studies have discovered the use of deep learning models for sentiment analysis. [13] illustrates a Boosted Deep Belief Network (BDBN) by its authors for this analysis in three iterative training stages. Zhang et al. in his paper [16] proposes a novel deep neural network: DNN-based facial feature analyzing model to find the relationship among SIFT features and high-level semantic information.

In the ImageNet challenge which took place in 2014, the top position holders all approached the problem with CNNs. Among these, the GoogleNet architecture had an amazing 6.66% error rate in classification [14]. Another note-worthy architecture that showed impressive results is the AlexNet [17], based on the regular CNN layered approach which is, convolution layers, then max-pooling layers and rectified linear units i.e ReLUs, and at the end some fully connected layers. AlexNet was also among the first ones to suggest using dropout layers in the network to tackle the problem of over-fitting.

## III. PROPOSED APPROACH

With the goal to ameliorate the process of facial sentiment analysis systems, we propose a classification mechanism using a CNN architecture. Due to the need of large data required for training of deep networks, we have made use of the FER2013 dataset which is available publically. In the subsequent section, we list out the features of our chosen dataset, followed by the description of our network architecture and finally the performance measures used for evaluation.

### *Dataset*

The facial expression recognition 2013 i.e. FER2013 database [11] was uncovered in the ICML challenge, 2013 in Representational Learning. This dataset houses a whopping 35,887 images of faces mostly taken in uncontrolled settings. The images are of 48x48 resolution with training set of more than 28k images and each of the testing and validation sets have 3.5k images respectively. The creation of this database was handled using the GoogleAPI for image search with faces marked with one of the 7 basic sentiments. In comparison with other datasets, FER2013 is better, has more variation in way of partially covered faces, low contrast in images, which all makes it better for training our model to be more generalized and prevent overfitting.

Fig 1: Sample images from fer2013

***Network architecture***

Our architecture is pretty straightforward with input of 48x48 passing to convolution layers each followed a pooling layer and ending with dense layers with dropout layer in between. The exact details of the layers is given in Table 1.

Table 1. Network Architecture

| Layer type | Filter/Size | Activation |
|---|---|---|
| Conv2D | 64/(5,5) | relu |
| MaxPool2D | (5,5) | - |
| Conv2D | 64/(3,3) | relu |
| Conv2D | 64/(3,3) | relu |
| AvgPool2D | (3,3) | - |
| Conv2D | 128/(3,3) | relu |
| Conv2D | 128/(3,3) | relu |
| AvgPool2D | (3,3) | - |
| Dense | 1024 | relu |
| Dropout | - | rate=0.2 |
| Dense | 1024 | relu |
| Dropout | - | rate=0.2 |
| Dense | 7 | softmax |

The different layers and activation functions are explained as follows:

1. **Convolution layer**

   The purpose of the Convolution layer is to handle the high-level features of the image. Traditionally, the leading Convolution Layer is used for analysing the low-level features such as edges, colour, gradient orientation, etc. With the subsequent layers, the architecture starts recognising the high-level features also.

2. **Pooling layer**

   Pooling layers are responsible for reducing the size of the Convolved Feature in space. This reduces the computational requirements for processing the data through dimensionality reduction. Another use-case is that it extracts dominant features that show rotational and positional invariance, thus maintaining the efficiency of training. There are various types of Pooling methods: ***Max*** and ***Average*** Pooling. MaxPool outputs the max value from the section of the image matrix covered by the kernel while, the AvgPool layer takes the average of the values covered in the kernel window and returns it.

3. **Dense layer**

   After the convolution and pooling operations have been applied on the input, and it has become sufficiently smaller and relevant with high level features, the 2d matrix is condensed to single dimension and the totally-connected neural layer is necessary as a way of learning non-linear combinations from the high-level features. These layers are updated by back propagation and update of weights. To prevent overfitting, some random neurons are dropped out to keep the model robust. The output layer neurons are equal to the number of classes in the problem.

4. **Activation function**

   These are functions that are used to introduce non-linear properties in the model. In case of neural networks, they convert the input signal to a node to the output signal from it. There are many kinds of activation functions but the most used ones are:

***ReLU*** i.e. rectified linear units also known as ramp function. It has a simple process that is mathematically defined as, relu(x) = max(0,x).

***Softmax*** function, is a function that inputs M real numbers, and normalizes it into M probabilities proportional to the value of the input numbers.

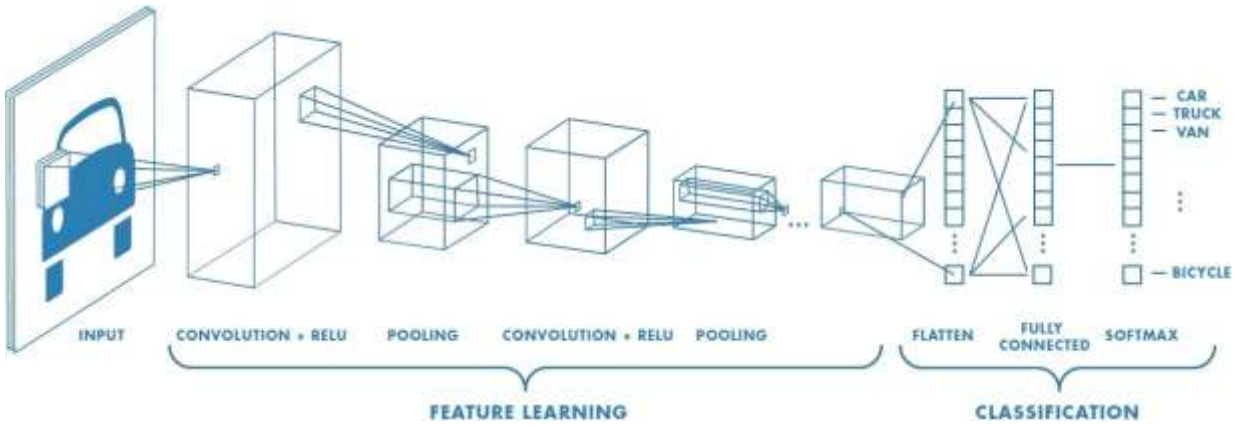


Fig 2: General layers in structure of a CNN

***Performance Metrics***

Since the task at hand is a multi-class classification one, cross entropy is used as a loss function and confusion matrix is used for performance indication.

**1. Categorical cross entropy**

Categorical crossentropy is one of the loss functions which is used for problems having single label categorization i.e. categorical crossentropy is beneficial in classification problems where only single result can be correct. The mathematical equation for this is,

$$L(y,\hat{y}) = -\sum_{j=0}^{M}\sum_{i=0}^{N}(y_{ij} * log(\hat{y}_{ij}))$$

Fig 3: Cross Entropy Equation

**2. Confusion matrix**

It is a table or matrix that is used for description of the performance of a classification model on a dataset for which the true values are known to us.

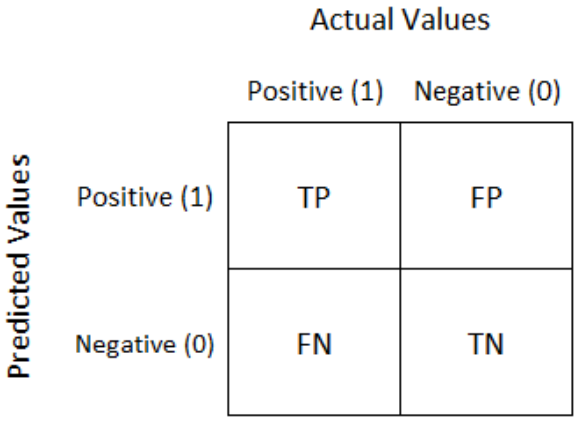


Fig 4: Confusion Matrix format

## IV. EXPERIMENTAL RESULT

The technical specifications of the computing system used for this experiment are :

Table 2. System Specifications

| | |
|---|---|
| CPU | Intel(R) Core(TM) i5-7200U CPU @ 2.7 GHz |
| RAM | 8.00 GB |
| OS | Windows 10 Home 64-bit |
| GPU | NVIDIA® GeForce® 940MX 4GB |

Jupyter-Notebook 6.0.1 along with Anaconda 1.9.7 having Python 3.7.4 is utilized for implementation on this system. Also, due to the need of large computational power for the training of network over such a vast dataset, Google colab notebook has been used. It provides free GPUs (one of Nvidia K80s, T4s, P4s and P100s), upto 12 GB of RAM for a period of 12 hours which came in handy for training purposes.

The distribution graph of the dataset is presented below :

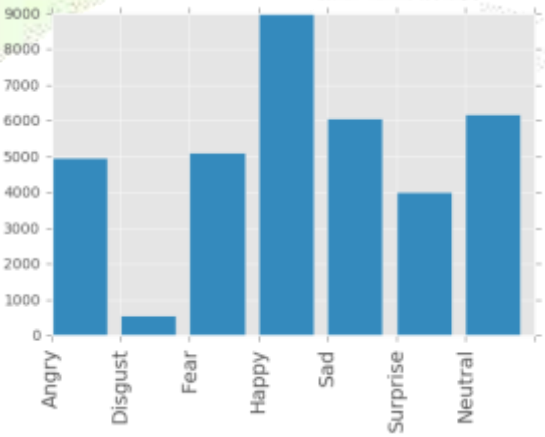


Fig 3: Distribution of fer2013 dataset

The model was trained, tested on the training and testing dataset respectively with the following evaluations metrics :

Training loss : 0.223031098232
Training accuracy : 92.0512731201

Testing loss : 2.27945706329
Testing accuracy : 57.4254667071

The accuracy should not express the right impression for a multi class classification problem as this. The confusion matrix for this model is presented below :

| | Angry | Disgust | Fear | Happy | Sad | Surprise | Neutral |
|---|---|---|---|---|---|---|---|
| Angry | 214 | 9 | 53 | 30 | 67 | 8 | 86 |
| Disgust | 10 | 24 | 9 | 2 | 6 | 0 | 5 |
| Fear | 45 | 2 | 208 | 29 | 89 | 45 | 78 |
| Happy | 24 | 0 | 40 | 696 | 37 | 18 | 80 |
| Sad | 65 | 3 | 83 | 56 | 285 | 10 | 151 |
| Surprise | 7 | 1 | 42 | 27 | 9 | 303 | 26 |
| Neutral | 45 | 2 | 68 | 65 | 88 | 8 | 331 |

Fig 4: Confusion matrix for the prediction set

## V. CONCLUSION & DISCUSSION

In this paper, we explore the field of facial sentiment analysis. We present a convolution neural network for the task of classification of facial images into the seven regular emotions which are, happiness, fear, sadness, anger, surprise, disgust and neutral. The fer2013 dataset has been used for training and testing purposes due to it's extensiveness and robustness. The performance of our network is 57% over the dataset which is very good considering that model with accuracy of 34% came in first place in the fer2013 challenge. This proves it's effectiveness in the field of facial sentiment analysis and is an improvement over other approaches. In the future, we plan on implementing it for real-time analysis and reduce latency. The usecases of Facial Sentiment Analysis are huge and it is a field that will see many more newer and better contributions than ever before.

## REFERENCES


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

[2] M. J. Jones and P. Viola, "Robust real-time face detection," International journal of computer vision, vol. 57, no. 2, pp. 137–154, 2004.

[3] Edwards, Henry J. Jackson, Jane , and Philippa E. Pattison. "Emotion recognition via facial expression and affective prosody in schizophrenia: a methodological review." Clinical psychology review 22.6: 789-832, 2002.

[4] Chu, William Wei-Jen Tsai, Hui-Chuan, YuhMin Chen and Min-Ju Liao. "Facial expression recognition with transition detection for students with high-functioning autism in adaptive e-learning." Soft Computing: 1-27, 2017.

[5] Chlo, Clavel, Iona Vasilescu, Laurence Devillers, Gal Richard, and Thibaut Ehrette. "Fear-type emotion recognition for future audio-based surveillance systems." Speech Communication 50, 2008.

[6] Cowie, Nicolas Tsapatsoulis, Ellen Douglas-Cowie, Roddy, Winfried Fellenz,George Votsis, Stefanos Kollias, and John G. Taylor. "Emotion recognition in human-computer interaction." IEEE Signal processing magazine 18, no. 1: 32-80, 2001.

[7] Pooya, Thomas Huang, Khorrami, and Thomas Paine. "Do deep neural networks learn facialaction units when, doing expression recognition?." in IEEE Conference on Comp. Vision. 2015.

[8] Aneja, Gary Faigin, Deepali, Alex Colburn, Barbara Mones, and Linda Shapiro. "Modeling stylized character expressions via deep learning." In Asian Conference on Computer Vision, pp. 136-153. Springer, 2016.

[9] Paul, Ekman, and W.V. Friesen. "Constants across cultures in the face and emotion." Journal of personality and social psychology, 1971.

[10] LeCun, Y. Generalization & network design strategies. Connectionism in perspective, 143-155, 1989.

[11] Carrier, I. J., Mirza, P. L., Courville, A., Goodfellow, M., & Bengio, Y. FER-2013 database. University de Montreal, 2013.

[12] H. Mliki, H. BenAbdallah, M. Hammami, and N. Fourati, "Data mining-based facial expressions recognition system." in SCAI, 2013, pp. 185–194.

[13] P. Liu, Y. Tong, Z. Meng, and S. Han, "Facial expression recognition via a boosted deepbelief network," in the IEEE Conference on Comp. Vision and Pattern Rec., pp. 1805-1812. 2014.

[14] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al. Imagenet large scale visual recognition challenge. arXiv preprint arXiv:1409.0575, 2014.

[15] S. Moore and R. Bowden, "Local binary patterns for multi-view facial expression recognition," Computer Vision Image Understand. vol. 115, no. 4, pp. 541–558, 2011.

[16] T. Zhang, K. Yan, Z. Cui, Y. Tong, J. Yan, and W. Zheng, "A deep neural network-driven feature learning method for multi-view facial expression recognition," Transaction of IEEE on Multimedia, vol. 18, no. 12, pp. 2528–2536, 2016.

[17] A. Krizhevsky, G. E. Hinton, and I. Sutskever. Imagenet classification with deep convolutional neural networks. In Advances in neural data processing systems, pages 1097–1105, 2012.

[18] Tzirakis, George Trigeorgis, Mihalis A. Nicolaou, Panagiotis, Bjrn W. Schuller, and Stefanos Zafeiriou. "End-to-end multi-modal emotion reco. using neural networks." IEEE Journal of Topics in Signal Processing 11, no. 8: 13011309, 2017.

[19] Mollahosseini, Mohammad H. Mahoor, Ali, and David Chan. "Going deeper in facial expression recognition using deep neural networks." Applications of Computer Vis., Winter Conference on. IEEE, 2016.

[20] M.Dahmaneand J.Meunier, "Prototype-based modeling for facial expression analysis," Trans. Multimedia IEEE, vol. 16, no. 6, pp. 1543–1552, Apr. 2014.

[21] W. Zheng, H. Tang, Z. Lin, and T. Huang, "Emotion recognition from arbitraryview facialimages,"in Proc.11thEur. Conf. Comput.Vis., 2010, pp. 490–503.

[22] Han, Kun, Ivan Tashev, and Dong Yu. "Speech emotion recognition using deep neuralnetwork and extreme learning machine." 15th annual conf. of the ISCA, 2014.