Emotion Detection using Convolutional Neural Network

Johanna Freeda G[1], Lavanya C[2], Lekhaa Shree D[3], Nivedhitha M[4], Kavitha C[5]

[1, 2, 3, 4, 5]Department of Computer Science and Engineering, R.M.K College of Engineering and Technology, Chennai, India

Abstract: ***The emotions of a human represent the mental states of feelings that arise without consciousness and effort and are accompanied by physiological changes in facial muscles which implies expressions on face. Some of the spontaneous emotions are happy, sad, anger, disgust, fear, surprise etc. Facial expressions play an important role in nonverbal communication that appears due to internal feelings of a person which reflects on the faces. In order to find the human's emotion through modeling, a extensive research has been carried out in past decades. But still it is far behind from human vision system. In this paper, we are using deep Convolutional Neural Network (CNN) to provide better prediction of human emotions and involving Frames by Frames . FERC-2013 and imdb database has been applied for training in this algorithm. In the proposed system we have experimented the emotion detection based on CNN which provides quite good result and the obtained accuracy may give insights to the researchers for future model of computer based emotion detection system.***

Keywords: emotion detection, deep learning, convolutional neural networks, machine learning, facial expressions

1. Introduction

Emotion can be expressed in various forms such as text, images and videos. A lot of papers are available for text sentiment analysis, but still image emotion analysis is yet to be explored due to its low accuracy rate. Generally image emotion detection can be performed by two ways using machine learning and lexicon based algorithms. Which machine learning algorithm includes SVM (Support Vector Machine), neural networks[6], neighbours, bayesian network, maximum entropy where as the lexicon based algorithm uses statistical and semantic based algorithms. Deep learning is one of the main subfield of machine learning where multiple learning levels are induced by representations or example feeded to the computer that makes the computer intelligence enough to analyse the data and produce the desired output. As deep learning algorithm include multi layer processing which breakdowns the input into many layers, this process reduce the complexity reducing the time and increasing the accuracy performance. Sub sampling layers give better result ,by use of CNN[6].

Generally the facial emotion recognition system have multiple stages of analysis such as face detection, features detection and extraction, and facial expression[7] classification. In this paper we focus on detecting the basic facial expressions using the machine learning algorithm, convolutional neural networks as it provides a better feature extraction. The way neural networks works are very different from other techniques. It is due to the fact that NN isn't "linear" like feature matching or cascades. When we give very complicated tasks like real time face recognition or other difficult patterns it's better to use neural net, because if it train the data first, which produces high precision, and it's faster in prediction as the data is already classified.

2. Literature Review

a) Visual sentiment prediction by merging hand craft and CNN features

Remarkable achievements has been shown in Computer Vision merging [1] CNN along with hand craft features. Color Histogram (CH) and Bag-of-Visual Words (BoVW) which are handcraft features are used along with some local features such as SURF and SIFT to detect sentiment of images. Large amount of training data are sometimes difficult to obtain in the area of visual sentiment and hence both data augmentation and transfer learning from a pre-trained CNN such as VGG16 trained with ImageNet dataset are employed. To attain the improvement in the performance of the visual sentiment prediction framework, End-to-End features from CNN and handshake of hand-craft are used. The results of experiments demonstrate that the visual sentimental prediction framework outperforms the state-of-the-art methods.

b) Facial emotion analysis using deep convolutional networks

The deep-learning [8] algorithms including Convolutional Neural Network(CNN) and recurrent neural network (RNN) are applied to the field of computer vision. These deep-learning-based algorithms have been used for feature extraction, classification, and recognition tasks. The CNN is to completely remove or highly reduce the dependence on physics-based models is the main advantage of CNN. It is also responsible for removal of other pre-processing techniques learning directly from input images and by enabling "end-to-end". CNN has achieved state-of-the-art results in various fields due to the above reasons. It also performs sentiment analysis, face and object recognition and scene understanding.

c) Predicting perceived emotion in animated GIFs with 3D CNN

Using 3D CNN, perceived [3] emotions are predicted in animated GIFs. A human's perceived emotion is not the emotion that they feel when a media sample is presented to them, it is the emotion they think the sample expresses. This way of expressing emotion is called their induced emotion. The features are represented in four different types of calculations: color histograms, facial expressions recognized by a CNN, image-based aesthetics, and a mid-level visual representation called SentiBank. A highest prediction accuracy is reported on 17 categories of human emotions after testing three different methods of regression using the features of facial expression. Facial expression recognition can barely work even when a large proportion of GIFs are made from cartoons or anime.

d) Context sensitive single modality image emotion analysis

Context sensitive emotion recognition[4] has been stated to be largely effective in the literature so far.. In this paper, the largest dataset of images collected are introduced. The datasets are collected from UCF ER which are labeled with emotion and context. A context-sensitive classifier is trained and tested to classify images based on both emotion and context. Experimental results pave way to considerable increase in performance compared to state-of-the-art. The problem in the single-modal domain; i.e. using only still images, remains less attended.

e) Image sentiment analysis using deep learning

Using deep learning, a suitable architecture of CNN for image sentiment analysis[5] is designed. Half a million training samples are obtained by using a baseline sentiment algorithm. The training samples are used to label Flickr images. A progressive strategy is used to fine-tune the deep network [8] and made use of such noisy machine labeled data. Robust and accurate feature learning, that produces the state-of-the-art performance on digit recognition , image classification, musical signal processing and natural language processing is provided by deep learning framework. A huge amount of effort in building powerful neural networks is bestowed by both the academia and industry. From the above studies, deep learning is discovered to be very effective in learning robust features in a supervised or unsupervised fashion. Using different optimization techniques it is tested to achieve the state-of-the-art performance on many challenging tasks.

3. System Architecture

The system architecture defines the working of convolutional neural networks. The basic working of the communication neural network [6] is divided into two parts. First the image is preprocessed where background is eliminated only to focus on the face. Then it is checked whether the datasets are trained. If they are not trained, the training of datasets take place by subsampling and the trained data are classified.

The algorithm convolutional neural networks which is part of deep learning[8] plays a major role in image processing. As the CNN is a part of deep learning, many layers can be applied in featuring and classifying the image frame by frame.

The hidden layers used in CNN are,

- Convolution layer
- Rectified linear units (ReLU) layer
- Pooling layer
- Fully connected layer

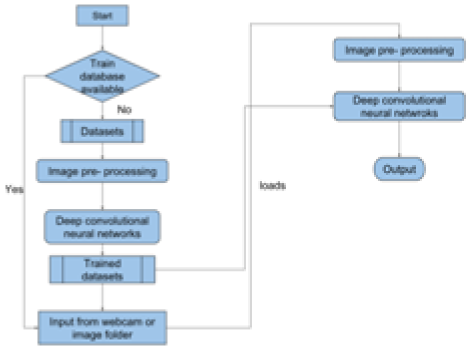


Figure 1: Workflow of the system

- The input is given in the for [64x64x1] . This input will hold the raw pixel values of the image, where in the width is taken as 64 and height as 64 and with three color channels R,G,B.
- Next convolution layer is used to process the output of neurons that are connected to local regions in the input, and the dot product operation is performed between their weights. This may result in volume such as [64x64x12] if we decided to use 12 filters.
- Rectified linear units layer will apply an element wise activation function, such as the max(0,x) thresholding at zero. This leaves the size of the volume unchanged ([64x64x12]).
- And the max pooling layer will perform a downsampling operation along the spatial dimensions (width, height), resulting in volume such as [32x32x12]. The main job of pooling is to reduce the size, so that the complexity is decreased during the computation.
- FC (i.e. fully-connected) layer is used to find out the class scores, resulting in volume of size [1x1x7], where each of the 7 numbers correspond to a class score, such as among the 7 categories of FERC-2013.

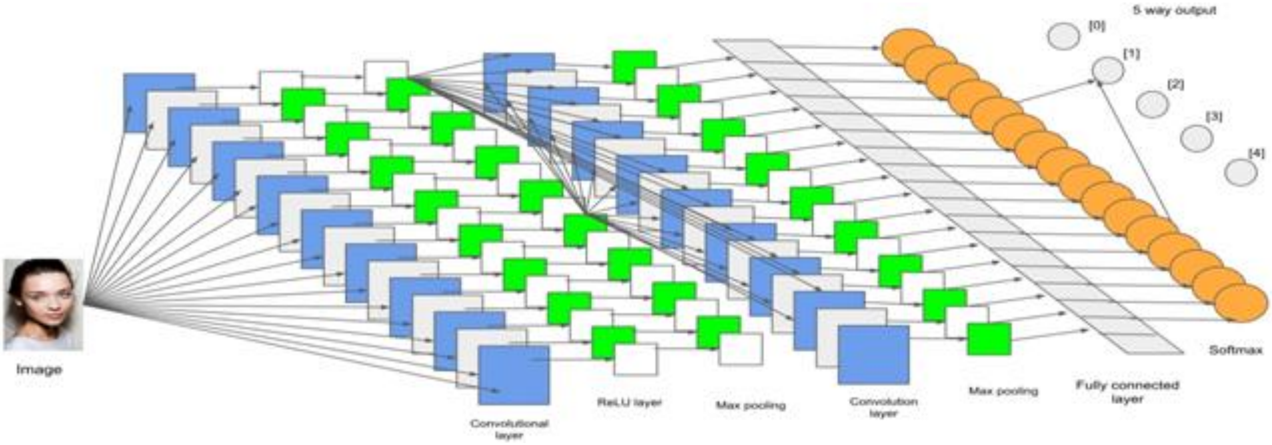


Figure 2: Architecture diagram

- After the fully connected layer, the last function used is Softargmax. The softmax function is mainly used to normalize the pre-processed input from the other layers into probability. The main use of this function is to bring every components in the interval of (0,1).

4. Proposed Algorithms

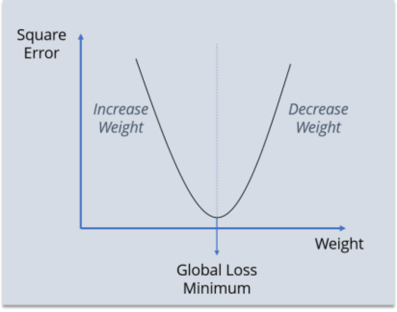
The algorithm used is forward and backward propagation algorithm. These algorithmic steps for the system is,
Step 1: Forward Propagation
Step 2: Backward Propagation
Step 3: Combining all the values processed and update the weights after calculation.

a) Forward propagation

The first step is taking the image to be trained as the input into the network. And then forward propagation[2] phase starts where all the layer including convolution layer, ReLU layer, pooling layer and fully connected layer are used to perform the computation. After the given layers, softmax function is used to normalize the inputs into probabilities. But as the weights for the training image is assigned randomly, the output is also random. Hence the backward propagation is used to correct the errors.

b) Backward propagation

Back propagation algorithm[2] is mainly used to CNN to calculate the weights. The first step is to calculate the error by finding how far is the actual output from the output we obtained. This gives us the error of the model. Then the next step is to check whether the error is minimum or not. If the error is maximum, then new weights are updated to the components and the weights are checked again. This process is repeated until the error is minimum. Once all the errors of the components are reduced, the model is taken into the prediction role.



Step1: Initializing the network weights for each points randomly (smaller values are preferred).

```
Step 2:  do
          forEach
          Training = X
          prediction_value  = neural_net - output(network, X)
          actual_prediction = teacher - output(X)
          compute error (prediction_value- actual_prediction)
          compute Delta w_h for all the predicted weights
         compute Delta w_i for all the predicted weights
         update network weights
```

Step3: Repeat the process until all examples classified correctly or another errors are minimised to the max.

Step4: Return the network

5. Results

When coming to the results in image processing, it is always not accurate or perfect, but we are trying to improve the accuracy by more computational power. The emotions [2] are detected and represented with the window frame as shown in figure 3.

A. Successfully detected emotions

The images in the figure 3 and similar images are detected correctly without errors due to training provided with the datasets. The datasets consists of distinctive facial expressions which when compared with the input image processes the similarity between them which is then represented as intensities.

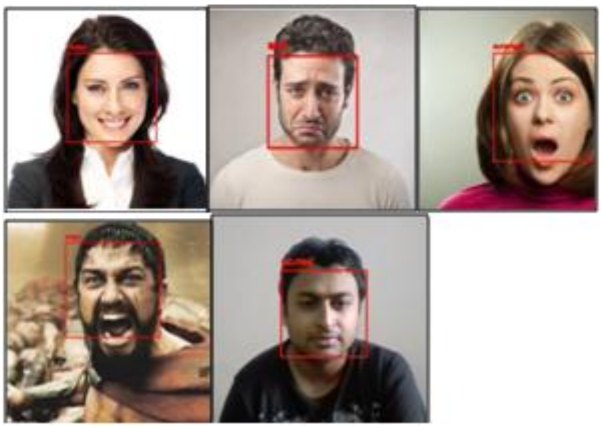


Figure 3: Row wise: (1) is happy face, (2) is sad face, (3) is surprised face, (4) is angry face, (5) is neutral face

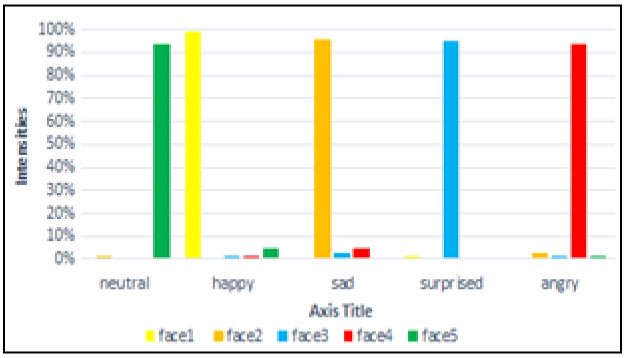


Figure 4: Emotion percentages of above successful test images

The intensities are calculated and the emotion with the maximum intensity as shown in figure 4 is taken into account and displayed in the output.

B. Failures

There are also some failure cases that have been found. The dominant expression is not well defined in the input image itself in most of the failure cases.

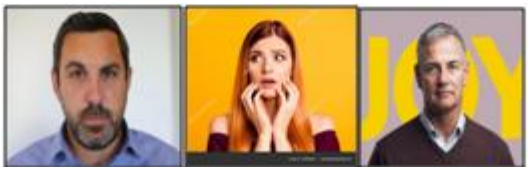


Figure 5: Row wise: (1) is angry face detected as happy, (2) is surprised face detected as sad , (3) is happy face detected as angry.

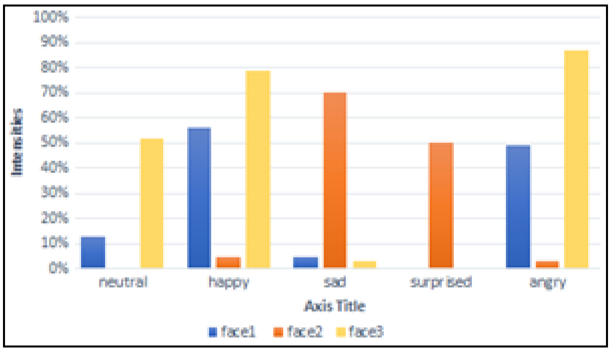


Figure 6: Emotion percentages of above failure test images

The above failures occur due to the dataset imbalance, the FERC-2013 data set contains number of images which are non-uniform to different emotions in the training set.

6. Conclusion

In this project, emotion detection using convolution neural network contains five different layers, for training and classification of standard emotions such as Happy, Sad, Angry, Surprise and Neutral. The real time emotion of the face is recognised with certain accuracy with the help of Python 3.6, OpenCV & Datasets using this training.

A facial expression [7] illustrates the state, mood and current feeling of a person in nonverbal communication. We can understand the emotion of a person in different stages. The percentage of emotion will vary significantly in different stages. In this paper, we have used convolution neural network with 9 layers, for training and classification of 5 types of standard emotions. For better analysis and interpretation of the percentage of emotions in various stages have also been measured with our proposed method. FERC-2013 and IMDB databases have been used in this experiment. For detecting the faces haar cascade classifier has been applied to recognizing emotion. A real-time emotion recognition system using face data is proposed and developed using convolution neural networks and the accuracy of the system we are getting around 70+ %.

References

[1] Wang Fengjiao and Masaki Aono"Visual Sentiment Prediction by Merging Hand-Craft and CNN Features" in *5th International Conference on Advanced Informatics: Concept Theory and Applications (ICAICTA), 2018*

[2] Rajesh Kumar G A, Ravi Kant Kumar and Goutam Sanyal "Facial Emotion Analysis using Deep Convolutional Neural Network " *in International Conference on Signal Processing and Communication (ICSPC), 2017*

[3] Weixuan Chen, Rosalind W. Picard "Predicting Perceived Emotions in Animated GIFs with 3D Convolutional Neural Networks" in *2016 IEEE International Symposium on Multimedia*.

[4] Pooyan Balouchian and Hassan Foroosh "Context sensitive single modality image emotion analysis: A Unified Architecture From Dataset Construction To CNN Classification" *in 2018 25th IEEE International Conference on Image Processing (ICIP).*

[5] Namita Mittal, Divya Sharma, Manju Lata Joshi "Image Sentiment Analysis using Deep Learning " *in 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI).*

[6] Tianyi Liu, Shuangsang Fang, Yuehui Zhao, Peng Wang and Jun Zhang "Implementation of Training Convolutional Neural Networks " *in Computer Science - Computer Vision and Pattern Recognition.*

[7] Shan Li and Weihong Deng "Deep Facial Expression Recognition: A Survey"*in ArXiv 2018 IEEEConference on Computer Vision and Pattern Recognition.*

[8] Deep Learning, *by Yann L., Yoshua B. & Geoffrey H. (2015) (Cited: 5,716)*