Platform to Provide Real Time Emotional Analytics via the Cloud

Subha Dharshini M, Vinoth Raaj J R J, Tousiff T H ,Mohana Gowri S,Siddharthaa R
Students at Sri krishna college of engineering and technology
Kuniamuthur, Coimbatore-641008

Abstract

The proposed system performs face expression recognition of the listener which helps the speaker to determine how the session is being processed by each individual listener wherein he can make changes to his tone of speech and way of conducting so as to make the session equally cherishable for all listeners. This system provides real time emotional analytics via the cloud and utilizes convolution neural network for facial expression recognition. Convolution neural network classify images, cluster all the images by similarity(using photo search) and then performs object recognition. Convolution neural network are algorithms which can be used to identify faces, individuals, street signs and all other visual data aspects.

Keywords - *Convolution neural networks, face expression recognition, Object recognition, Support vector machine*

I. INTRODUCTION

A speaker-listener interaction in any session is the most important thing for a successful session. Most of the listeners delivers only emotional response during sessions, as they are not supposed to deliver feedbacks amidst of session . At the same time, speaker cannot determine all the listener's emotional response when session is going on. This system utilizes convolution neural network for facial expression recognition.

Earlier support vector machine is used to classify the images in facial recognition. SVM is a more generic classifier. SVM simply maps it input to some high dimensional space where the differences between the classes can be revealed. SVMs are shallow architecture when compared to CNNs.
SVMs look to maximize the margin , while CNNs are not. We are CNNs over SVM because and are deep architectures.

In any face expression recognition system, there are 2 steps as follows:

- Face detection(bounded face)in image
- Emotion detection on detected face

For face detection in image we are using Haar Feature-based Cascade Classifiers. In 2001, Paul Viola and Michael Jones in their paper "Rapid Object Detection using a Boosted cascade of Simple features" proposed Haar feature based cascade classifiers. This is the most effective object detection method and a machine learning based approach. Each cascade function is already trained a lot from positive and negative images.

This algorithm needs a image of faces and images without faces for training the classifier. After training the classifier we need to extract features from it. This detects faces in image as well. This is the fastest and better face detector method when compared to other face detectors.

After face detection, the next most important thing to do is emotion detection. For this Xception CNN model is used. After training a classification CNN model architecture, it takes detected face from Haar Feature-based Cascade Classifier as input and gives probabilities of 7 emotions as output. Angry, Disgust, Scared, Happy, Sad, Surprised and neutral are the 7 emotions whose probabilities will be classified on output layer.

II. METHODS

A. Deep neural network architecture

Shallow engineering had the capacity to accomplish 76% exactness as it were. For more accuracy we are using deep neural network architecture .CNN comes under deep neural network architecture. If we train a deep CNN to classify images, then at the first layer itself it will recognize very basic things like edges. At second layer , it will recognize collections of shapes like wheels ,legs, tails, faces. At third layer, higher order features like objects will be recognized. Deep CNN are very harder to train because:

- As the network becomes deeper, data requirement increases.
- As number of parameters(weights) increases then regularization becomes more important in order to do learning of weights from memorization of features towards generalization of features.

B. Face recognition data set

The data for face expression recognition consists of grayscale images of faces. It is of 48*48 pixel. Here the faces are registered automatically and so in each image occupies a same amount of space .We are categorizing faces based on expressions into 7 emotions and numeric codes associated with them:

TABLE I

Code	Emotion
0	Angry
1	Disgust
2	Fear
3	Happy
4	Sad
5	Surprise
6	Neutral

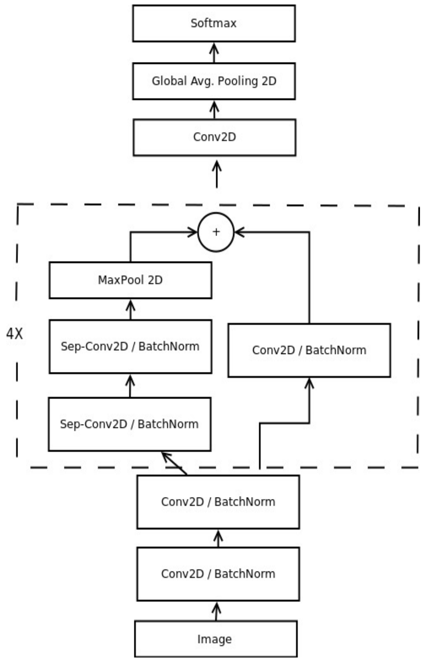
This data is stored in a csv file which consists of 2 columns named emotions and pixels.
The column emotion contains a numeric code ranging from 0 to 6. The pixels column consists of a string surrounded in quotes for each image which is a row of 1*2304. Each string can be converted into image using specific code. After loading the data set we will pre-process the image to feed into the CNN model.

C. CNN Architecture

We are using CNN architecture to classify images into different categories using state of the art network architecture . It includes ImageNet which categorizes image into different categories based on a defined set of words and phrases. To classify emotion of this data-set we are training CNN model:Mini Xception which is small and achieves almost state of art performance. The below architecture which we are using was proposed by Octavio Arragia et al in paper "Real time convolutional neural networks for emotion and gender classification". This architecture combines the deletion of the fully connected layer which was used in common architectures where most of the parameter resides and includes the combined depth-wise seperable convolutions and residual modules.

The reason for why we are using depth-wise seperable convolutions over standard convolutions is that depth-wise seperable convolutions reduce the computation with respect to the standard convolution by a factor of $1/w + 1/D^2$. This engineering is completely convolutional neural system that contains four lingering profundity savvy seperable convolutions where every convolution is trailed by a group standardization task and a ReLU initiation work.

FIGURE I



D. Training the CNN model

There are four techniques which is used for training the CNN model as follows:

1. Data Augmentation

We have ImageDataGenerator class in Keras that is utilized to create groups of tensor picture information with constant information expansion. Utilizing the preparation set by applying changes more information is produced. On the off chance that the preparation set isn't sufficiently adequate to adapt, at that point it is required. By changing the genuine preparing pictures by turn, crop, shifts, shear, zoom, flip, reflection, standardization and so on the picture is created. For this we are utilizing TensorFlow.

2. Kernel Regularizer

Amid advancement it permits to apply punishments on layer parameter. Those punishments are consolidated in the loss of capacity that the system improves. L2 regularization of the loads are contentions in convolution layer. This rebuffs peaky loads and guarantees that all of the information sources are considered. L2 regularization implies that each weight is rotted straightly and that is the reason called weight rot.

3. Batch Normalization

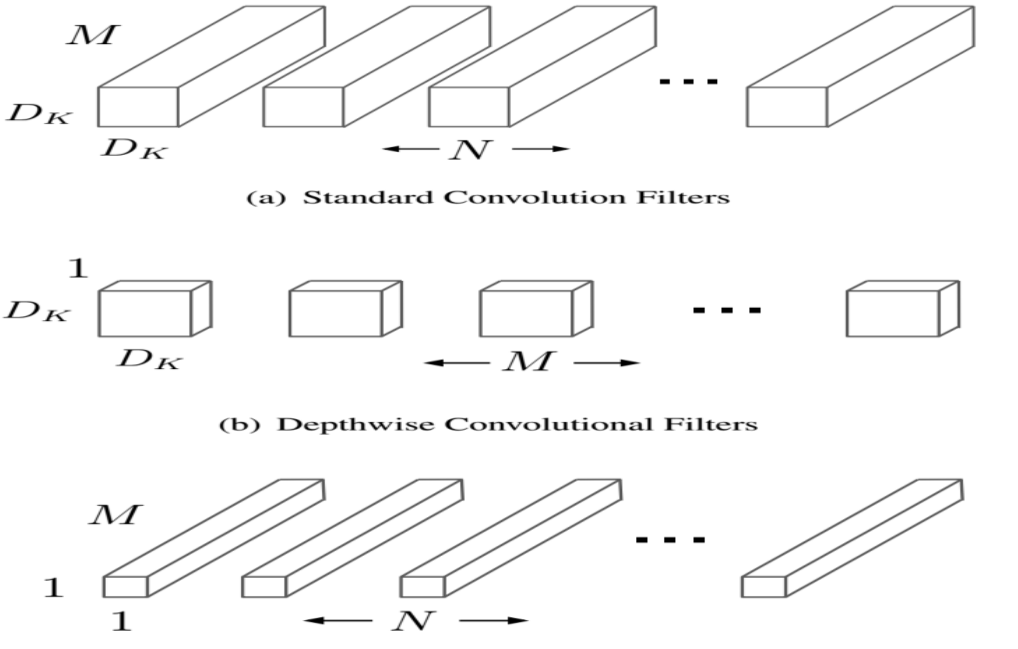
It applies a change that keeps up the mean initiation near 0 and the actuation standard deviation near 1. It tends to the issue of interior covariate shift.It additionally goes about as a regularizer, now and again dispensing with the requirement for Dropout.Batch Normalization accomplishes a similar precision with less preparing advances in this way accelerating the preparation procedure.

4. *Global average pooling*

By taking the normal over all components in the element map it lessens each element map into a scalar esteem. The normal activity powers the system to remove worldwide highlights from the information picture.

5. *Depth-Wise Convolutions*

These convolutions are composed of two different layers: depth-wise convolutions and point-wise convolutions. Profundity astute detachable convolutions diminishes the calculation regarding the standard convolutions by lessening the quantity of parameters.



(a) Standard Convolution Filters

(b) Depthwise Convolutional Filters

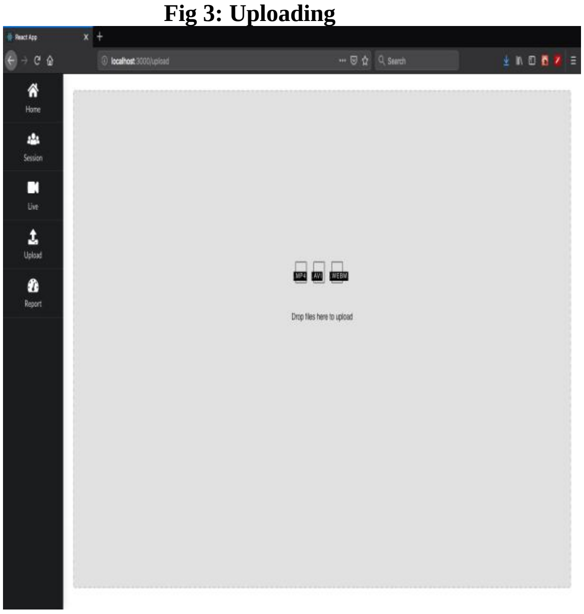
(c) 1 × 1 Convolutional Filters called Pointwise Convolution in the context of Depthwise Separable Convolution

Fig 2: Difference between standard and Depthwise Convolutional Filters

III.TESTING THE MODEL

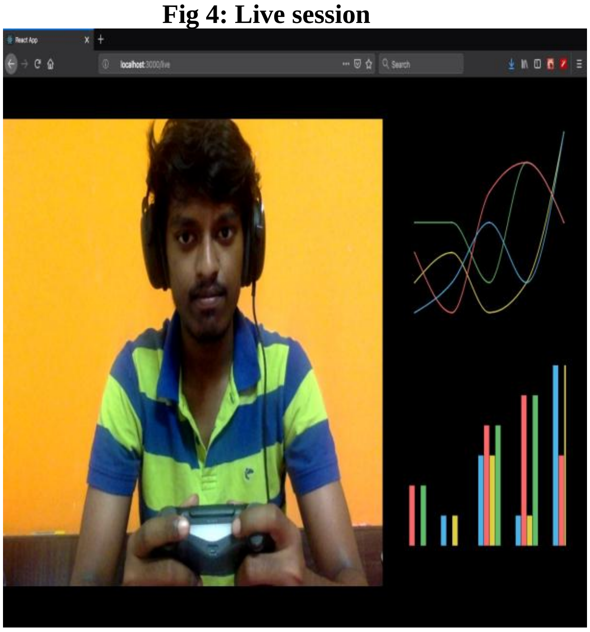
A. *Uploading already captured video for analysis*

Fig 3: Uploading



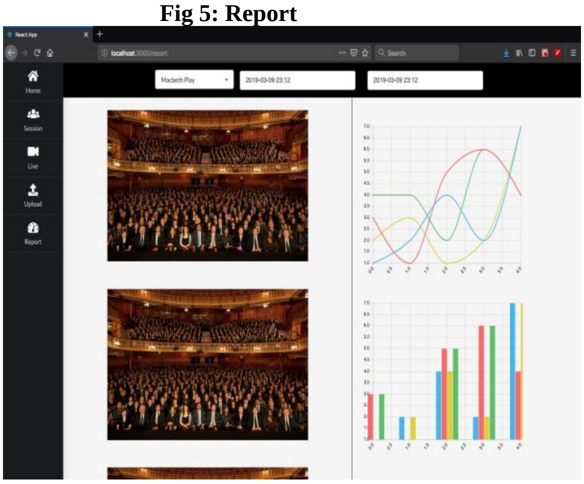
B. *Face expression recognition on live session*

Fig 4: Live session



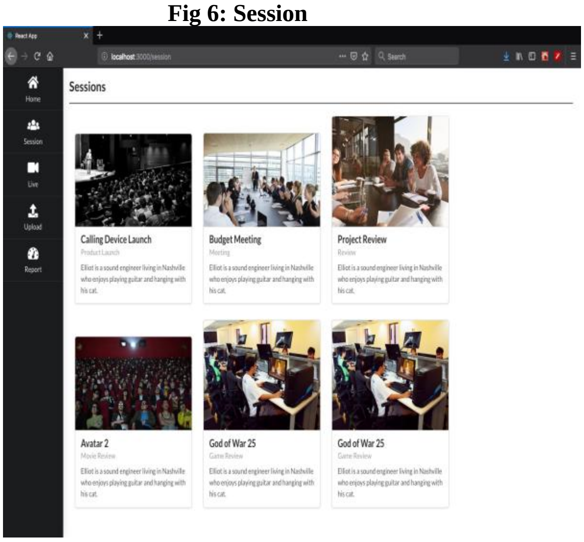
C. *Report*

Fig 5: Report



D. Session

Fig 6: Session



IV. CONCLUSION

We have proposed and tested a general building designs for creating platform for real-time emotional analysis via cloud. Our proposed uses cloud to deliver instant analytics even though user's system doesn't support or not compatible for analytics.

ACKNOWLEDGMENT

This work was carried out as a part of an "International Conference on recent trends and current research in engineering and technology" and attached to the project "Platform to provide real time emotional analytics via the cloud".

REFERENCES

[1] Octavio Arriaga, Paul G Ploger, Matias Valdenegro in "Real time Convolutional neural networks for Emotion and Gender classification" in October 2017

[2] "Facial expression recognition using face regions" by Khadija Lekdioui in December 2018