Emotion Detection Module

Vishal Choudhary[1], Umang Arora[2], Rakshit Goel[3], Rachit Jain[4]
[1]Assistant Professor, CSE Department, MIET, Meerut
[2,3,4]B.tech, CSE Department, MIET, Meerut

Abstract:Human expressions can tell us about their feelings and emotional state. Using new technologies we can find out about their emotional state which can be useful in many ways. It can be helpful to find out their mental state or can be helpful in the field of health. With the help of the human emotion we can also guess what they might be thinking or what their next step could be.Humans mainly have seven emotions which are angry, sad, fear, happy, disgust, neutral, and surprised. These emotions can be easily found out from their facial expressions. In the pre-trained model, algorithms extract the features of the face and validate them with the features of the training dataset which is used to train the model, to find out the real time emotion of the person in front of the camera.

Index Terms: CNN, tensorflow, keras, tkinker, ReLU, Activation Function, Pooling

I. INTRODUCTION

Humans can make thousands of expressions, our project's main idea is to capture these expressions and analyse them to find out the emotion of the person. For this analysis we are training a FER (Facial Expression Recognition) system which works on the bases of CNN (Convolutional neural network) model. The connectivity pattern in the architecture of a CNN is analogous to that of neurons in the human brain.

In this CNN model we are using tensorflow as a framework. In tensorflow we implement keras sequential model which is trained by the good quality data. The dataset we are using is the FER2013 dataset, published at the International Conference of Machine Learning. A computer recognizes an image as a pixel array. The dataset consists of 35,685 examples of 48x48 pixel grayscale images of human faces. It is a categorical dataset in which images are classified on the basis of the emotions.

The model is built based on the VGG16 model. VGG16 is a CNN model that was proposed by K. Simonyan and A. Zisserman from Oxford University. We have added three neuron layers. Each neuron of CNN comprises a Convolution layer followed by Pooling. Our aim is to classify the emotion on the face of a person into one of the seven categories i.e. {'angry': 0, 'disgust': 1, 'fear': 2, 'happy': 3, 'neutral': 4, 'sad': 5,'surprise': 6}.

Each convolutional layer consists of weights whose values are to be learned and it has a number of filters which convolves with the input volume to compute activation map. Between these convolutional layers a pooling layer is present. This pooling layer does down-sampling of representation to reduce the number of parameters and computation. Max pooling works better and so is used generally.

II. LITERATURE REVIEW

In facial emotion detection field many approaches have been used to identify and predict the output. The similar thing in many approaches is the detection of eyes, nose and lips and other features on the target face and identification of the expression based on the data.

To identify the facial emotion various techniques are used like Neural Network, Machine is used and this approach is been proven to be much better. Identifying and using of many features at once is proven to be much better then only using one feature at once to identify the facial emotion. It has been proven by research that using deep neural network can generate more discriminating features.

In some methods neural network is used in two ways. One to remove the background and other obstacles from target face and second to concentrate on facial features and gathering the required data to predict the output and to do so FERC algorithm is used with many other scripts. Once background is removed this second part of CNN used various points on face to recognize the facial features. In some system the aim is also to detect expression as well as age and gender and it was assumed that girl and boys have similar facial features and expressions and similar approach is used to identify facial expressions. To obtain facial features CNN network have high branching and both noises and redundancy is non-avoidable in this process.

According to research pre-training is much better then network initialization randomly and pre-trained classifier, such as Random Forest, AdaBoost and Support Vector Machine (SVM) are used for FER based on trained features. Deep-learning approach requires more memory and computation power than that of conventional approaches.

III. METHODOLOGY

Our aim is to classify the emotion on the face of a person into one of the seven categories i.e. angry, disgust, fear, happy, neutral, sad, surprise.

The connectivity pattern in the architecture of a CNN is similar to the connection of neurons in the human brain. A computer recognizes an image as a pixel array. Depending on the resolution, it will see the image as a (height x width x dimension) array.

To build a smart model using CNN we need to follow a step-by-step process which includes data elicitation, data processing, data cleaning, data normalization, splitting into train and test data, building model, training model and testing and validating model.

The entire work process will be carried out in Python language and the features which prompt us to choose this language over any other languages are that python is Object-oriented, Structured Programming, and availability of large libraries which makes it compatible and easy to use. We will use the following libraries: numpy, pandas, matplotlib, OpenCV, Tensorflow and Keras. We will be working in Google colaboratory. Colab, short for colaboratory, is a Jupyter notebook service provided by Google that requires no setup to use.

1 Data Elicitation

Data is the most important aspect in building any Deep Learning model. For recognition of facial emotions using Convolutional Neural Network, we need to train the model using a good quality of data. For this purpose we use the FER2013 dataset, published at the International Conference of Machine Learning. The dataset consists of 35,685 examples of 48x48 pixel grayscale images of human faces. It is a categorical dataset in which images are classified on the basis of the emotions shown in the image.

Next step is that we split data into training and testing sets. The available dataset FER2013 is divided into two parts, the training data set is used to train or fit the model in the learning process and test data is used for validation or evaluation. Test data allows us to check how the model will perform on the unknown data in real time. If we use the whole data set for training there are chances the model might overfit the pattern of the data.

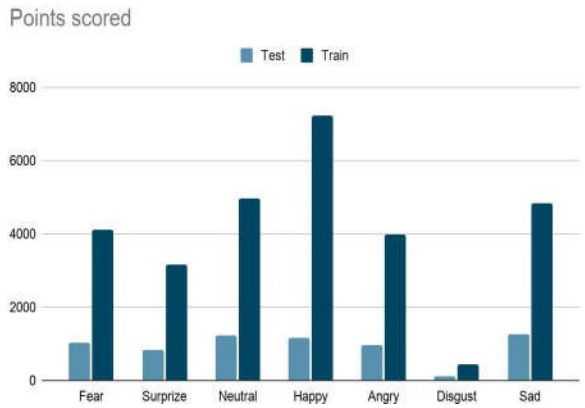

Figure 1. Number of Images in each category

2 Building Model

The proposed model is built based on the VGG16 model. VGG16 is a CNN model that was proposed by K. Simonyan and A. Zisserman from Oxford University. In our module, there are two convolution blocks and each block has two layers. Image of size 48x48 is given as input to the first layer, conv1 after converting to grayscale. In conv1 32 filters are used of size 3x3. The second convolution layer, conv2 has 64 filters with same-padding. Then we have performed Max-pooling over 2x2 pixel window, with number of strides equal to two. Similarly, third and fourth convolution layers conv3 and conv4 have 128 and 256 filters each and have kernel size 3x3. The activation function is Rectifier Linear Unit (ReLU). Padding is 'same' in all layers except for conv4. Then, fully connected layer comprises of 1024 neurons, and the output layer has 7 neurons.

2.1 Convolution Layer

The Convolutional layer has weights which need to be trained. The Convolutional layer has a number of filters whose parameters are to be learned. Each filter has the same dimension as the input volume, but height and weight of the filter is smaller. Each filter convolves with the input volume to compute activation map.

Activation function is used to decide whether a neuron should be activated or not and what should be its impact. Activation function introduces non-linearity in the output of a neuron. They help in limiting the output of a neuron, solve the problem of vanishing gradient, produce values zero-centered (outputs are symmetrical so that the gradient does not shift to particular direction) and reduce computational expense. These all features mainly help in reducing the overfitting of the model.

2.2 Pooling Layer

Pooling layer is present between two convolution layers. Pooling does down-sampling of representation to reduce the number of parameters and computation. Max pooling works better and so is used generally.

3.3 Training Model

Training a model is the most crucial step as the output of the model is defined in this process. Training the model may require multiple attempts for achieving high accuracy. We only proceed to the next step if we are satisfied with the result of training the model. We also need to check and remove over-fitting if it occurs. Our model has been trained for 60 epochs.

3.3 Model Evaluation

Confusion matrix provides a summary of correct and incorrect predictions made by the classification model. We use a confusion matrix to check the performance of the emotion detection model on the set of training and testing data as their true output is known. It is used to measure the variables such as precision and recall.

Classification report gives the details of the predictions made by the emotion detection model. This report shows the classification metrics precision, recall and f1- score on a per-class basis in tabular form. It is calculated using true positive, true negative, false positive and false negative values.

To obtain the real time image we need OpenCV (Open Source Computer Vision Library). OpenCV is an open source computer vision and machine learning library which provides infrastructure for computer vision in our project. It captures the image from the camera and processes it in the form of a 3-D matrix containing pixel values of the image in the form of BGR (blue green red) format. It uses a haarcascade Frontal face classifier to detect the human face in real time. It also provides us the feature for cropping and resizing our image.

The next requirement is to set up an Interactive Graphical User Interface for users. Tkinter, an object oriented layer over Tcl/Tk. Tk is not a section of Python; it is prolonged at the active state. It is an interface for Tk GUI toolkit embedded with Python. Tkinter calls are translated into Tcl commands, making it possible to combine python and Tcl in a single project. Tkinter provides many features like Window, Frames, Buttons, Text Fields and Labels for making a good and interactive GUI.

IV. Result

After training the model, we have achieved an accuracy of 91.03% on the train set, and 66.58% on the test set.

Table 1. Model Evaluation

Data	Accuracy Percentage	val_accuracy	val_loss
Train	91.03	0.9103	0.3737
Test	66.58	0.6658	1.1666

V. Conclusion

A user friendly interface has been designed which incorporates an image capturing and processing window. The camera captures real time images and classifies the facial emotion out of the seven categories {'angry': 0, 'disgust': 1, 'fear': 2, 'happy': 3, 'neutral': 4, 'sad': 5, 'surprise': 6}. The model constructed based on the VGG16 category of CNN achieved an accuracy of 91.03% on train data and 66.58% on test data. The future work can be done on a larger dataset to achieve even higher accuracy and add more categories of emotions as well.

REFERENCES

1. Isabelle H, Sandra B,. Eva C,. Rafael Del-H, . The Emotracker: Visualizing Contents, Gaze and Emotions at a Glance, Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, Geneva, Switzerland. 2–5 September 2013.
2. Linqin C., Hongbo X., Yang Y., Jimin Y., Robust facial expression recognition using RGB-D images and multichannel features. Multimed. Tools Appl. 2 May 2018.
3. Luchao T., Mingchen L., Yu H., Jun L., Guyue Z., Yan Q. C., Robust 3D Human Detection in Complex Environments with Depth Camera. IEEE Trans. Multimedia. 9 September 2018.
4. Sander K., Ioannis P. Fusion of facial expressions and EEG for implicit affective tagging. Image Vis. Comput. February 2013.
5. Pallavi P., Seeja K.R. Emotional state recognition with eeg signals using subject independent approach; Data Science and Big Data Analytic. Springer; Berlin/Heidelberg, Germany. January 2019.
6. Huang Y.-J., Lu H.-C., Yang D.-I., Chen Y.-W., Real-time facial expression recognition based on pixel-pattern-based texture feature. Electron. Lett. 20 August 2007.
7. Ronald D. B., Solange A., David G., Richard J. C. An efficient algorithm for spectral analysis of heart rate variability, IEEE Trans. Biomed. Eng., September 1986.
8. Leonardo F., Alessandro T. A Neural Network Facial Expression Recognition System using Unsupervised, Local Processing, Cognitive Neuroscience Sector SISSA, April 2001.
9. Carlos B., Zhigang D., Serdar Y., Murtaza B., Chul M. L., Abe K., Sungbok L., Ulrich N., Shrikanth S. N. Analysis of Emotion Recognition using Facial Expression , Speech and Multimodal Information, ICMI, USA, January 2004.
10. Myunghoon S., Balakrishnan P. Real-time mobile facial expression recognition system--A case study, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops; USA. June 2014.
11. C. Busso, Z. Deng , S. Yildirim , M. Bulut , C.M. Lee, A. Kazemzadeh , S.B. Lee, U. Neumann , S. Narayanan Analysis of Emotion Recognition using Facial Expression , Speech and Multimodal Information, ICMI'04, PY, USA, 2004.
12. L. Franco, A. Treves. A Neural Network Facial Expression Recognition System using Unsupervised‖ Local Processing, Cognitive Neuroscience Sector, SISSA.