Special Issue - 2021

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
NCETER - 2021 Conference Proceedings

Facial Expression Recognition in Video Call

Sourav Dey
Computer Science and Engineering
Dept. of JIS College Of Engineering, Kalyani, Nadia ,India

Arcoprova Laha
Computer Science and Engineering
Dept. of JIS College Of Engineering, Kalyani, Nadia, India

Mr. Apurba Paul
Asst. Professor of Computer Science and Engineering
Dept. of JIS College Of Engineering, Kalyani, Nadia, India

Shrijoyee Roy
Computer Science and Engineering
Dept. of JIS College Of Engineering, Kalyani, Nadia, India

Suman Paul
Computer Science and Engineering
Dept. of JIS College Of Engineering, Kalyani, Nadia, India

Abstract— **Emotions are a powerful way to communicate. By seeing the emotion of an individual someone can communicate accordingly. Suppose if a person is sad, we can crack a joke to make him happy and there are other endless examples where emotion recognition is needed. Now a human being can easily recognize others' emotions by simply seeing their facial expressions but if a machine understands a person's emotional acts, accordingly it will create a tremendous change in human-computer interaction.**

In this project, we created a video calling application where the machine can understand the emotion of the caller and the receiver and shows the emotion of the caller to the receiver and vice versa. As the emotion changes the machine automatically detects and updates the emotion, which displays as pop up and the machine speaks up the detected emotion.

Keywords— *Artificial Intelligence, Machine learning, Deep Learning, Artificial neural network, Convolution Neural Network, Facial expression recognition, facial emotion. HTML, CSS, JavaScript, NodeJs.*

INTRODUCTION

Today we are living in a technology era. Nowadays everyone use the computer in their daily life. So the machine and human interactions are increasing. Human have the capability to learn from theirexperiences and understand their surroundings environment. But for machine to achieve this capacity, deep learning and machine learning play a key role. Human facial recognition is a basic way to interact with human for machines.

FER has applications in human-computer interaction, Virtual reality, augmented reality, Education, audience analysis in marketing, and entertainment. Also, it has applications in driver safety such as video conferencing, credit card verification, criminal identification, facial action synthesis for the animation industry, and cognitive sciences. Our paper aims to explain the procedures of FER systems.

LITERATURE SURVEY

In the past few years, there are many works have been done on face recognition and have achieved success in real applications. [1] Alpeshkumar Dauda, Nilamani Bhoi proposes Facial expression recognition using PCA and Distance classifier. In this paper, they use PCA for getting the important features of the image and they classified the images.[2] Mingjie Wang, Pengcheng Tan, Xin Zhang, YuKang, Canguo Jin, Jianying Cao in their paper has shown how can effectively recognize facial expression.[4] J. Whitehill, proposed a paper on automatic facial expression recognition in real-time where he makes two approaches to extract features from the faces geometry-based and appearance-based.

METHODOLOGY

The project is divided into two parts, the first is building the model and the deployment of it is the next. In the first part, a dataset of FER (Facial Recognition dataset) is downloaded from Kaggle which was already divided into three parts for training, testing, and validation. Then with the use of different libraries like Tensorflow, Keras, where the algorithms are predefined as classes and we only have to create objects of that class and gives its parameters tocreate a model, a convolution model is created. The model can be done by other machine learning algorithms but CNN stands out from other algorithms because CNN provides other advantages like the preprocessing of data are not required for CNN which on using other algorithms we have to manually process the data but CNN can give importance to various aspects in the image and able to differentiate one from other. As the model is ready, the dataset is fitted to it, and this way the training and validation are done simultaneously. After successful completion of training, the model is saved for deployment.

In the second phase, a python file is written using the OpenCV library to capture the videos from the device.OpenCV is a python module that is used to process images and videos to identify objects or faces. Then with the use of WEBRTC, a communication system is made in javascript. Then the videos captured from the device are passed to the model which then predicts the emotion. Then the predicted emotions are dictated by a speech engine with the help of speech synthesis API

which can speak up any text and the emotion is also displayed as pop-ups on the screen of the callers.

CNN Architecture and Working

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning technique which will absorb associate input image, assign importance (learnable weights and biases) to numerous aspects/objects within the image, and be able to differentiate one from the other. The pre-processing required in a ConvNet is far lower as compared to alternative classification algorithms.

There are three forms of layers that form up the CNN that are the convolutional layers, pooling layers, and fully-connected (FC) layers. Once these layers are stacked, a CNN design are shaped

1. Convolutional Layer

This layer is the initial layer which extracts the varied options from the input images. In this layer, the computation of convolution is performed between the input image and a filter of a specific size MxM. By sliding the filter over the input image, the scalar product is taken between the filter and therefore the parts of the input image with reference to the dimensions of the filter (MxM).

The output is termed as the Feature map which supplies the information about the image such as the corners and edges. Later, this feature map is fed to different layers to find out some other features of the input image.

Image   Convolved Feature
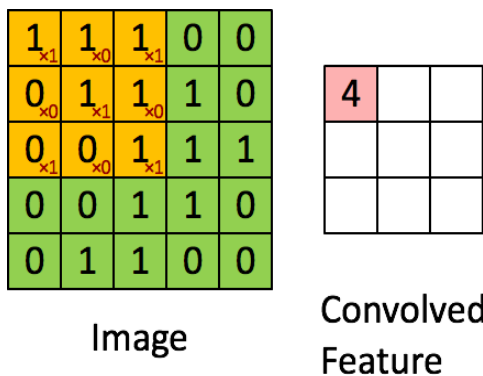
2. Pooling Layer

Convolutional Layer is followed by a Pooling Layer in most of the cases. Primary aim of this layer is to scale down the dimensions of the convolved feature map to scale down the computational costs. This is performed by decreasing the connections between layers and independently operates on each feature map. Depending upon method used, there are several sorts of Pooling operations.

In Max Pooling, the most important element is taken from feature map. Average Pooling computes the average of the elements in a predefined sized Image section. The total sum of the elements in the predefined section is calculated in Sum Pooling. The Pooling Layer usually acts as a bridge between the Convolutional Layer and therefore the FC Layer.
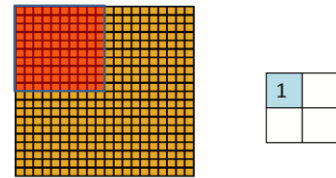
Convolved feature   Pooled feature

1. Fully Connected Layer

Those are basically simple feed forward networks where all the nodes have connection with all the nodes of the next layer. Most of the popular machine learning models have this fully connected layers in the last few layers.

In this, the data from the previous layers are flattened and the obtained vector then undergoes few more FC layers where the mathematical functions operations usually happened and are compiled to form the output layer.

2. Dropout

Usually when all the nodes are connected to each other it may cause overfitting. Overfitting is a condition where the model works very well for the training data but can not perform well in other data. So to avoid this overfitting condition a regularisation method is used where some of the neutrons are ignored or we can say data of those neurons are dropped out , and this technique is called dropout. If in a model dropout is set as 0.2 it means datas of 20% of the neurons are removed randomly.

1. Activation Functions

Activation functions are one of the most important parts of any machine learning models. This functions decides whether the neurons will be activated or not. The purpose of this functions is to add non-linearity to the network.There are many functions like softmax, relu , sigmoid etc. In general relu is used , this function returns zero for any negative input and for any positive value that value is returned.

RESULTS AND DISCUSSION

CNN architecture for facial expression recognition as mentioned above was implemented in Python.A validation set was used to validate the training process. In the last batch of every epoch the validation cost, validation error, training cost, training error are calculated. Input parameters for training are image set and corresponding output labels. The training process updated the weights of feature maps and hidden layers supported hyper- parameters like learning rate, momentum, regularization, and decay. We got an accuracy of 82.33% and which can be considered to be a good model.

CONCLUSION AND FUTURE WORK

In this project, we created a video calling application that recognizes the emotions of the persons calling

Special Issue - 2021

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
NCETER - 2021 Conference Proceedings

and tells the emotion to the other person. There are many previous projects of facial expression but they are rarely been used as a real-world application. In our project we tried to apply the facial expression in video calling app which can be used in real world. The most unique part of our project is that as soon as an emotion is detected the machine speaks up the emotion. So it will be very helpful for the blind people to communicate where they will get to know each other's emotion during calling. Later this project can be improved by using a model which is more accurate as it is challenging to recognize facial expression of peoples with different skin tones, colors and textures

DIAGRAM
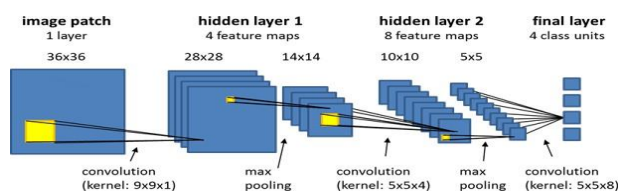
ACKNOWLEDGMENT

I would like to sincerely thank to Prof**. Dr. Dharam Paul Singh**, Head of the Department of Computer Science and Engineering of JIS College of Engineering for his support and guidance and all the respected faculty members of CSE Department for giving the scope of successfully carrying out the project work.

I would also give my gratitude to Prof, **Apurba Paul** for allowing the degree attitude and providing all the guidance related to the project work. His conscription of the topic and all the helpful hints.

Finally, we take this opportunity to thank Prof, **(Dr.) Partha Sarkar** Principal of JIS College of Engineering.

REFERENCES

[1] AlpeshKumar Dauda and Nilamani Bhoi " Facial Expression Recognition using PCA & Distance Classifier", *International Journal of Scientific & Engineering Research*, vol. 5, no. 2, May 2014

[2] Mingjie Wang, Pengcheng Tan, Xin Zhang, YuKang, Canguo Jin, Jianying Cao, "Facial expression recognition based on CNN".

[3] R. Kushwaha and N. Nain, "Facial Expression Recognition", *International Journal of Current Engineering and Technology*, vol. 2, no. 2, pp. 270-278, June 2012

[4] J. Whitehill, "Automatic Real-Time Facial Expression Recognition for Signed Language Translation", May 2006.

[5] Khalid Ibn Zinnah, Nafiz Mahmud, Firoz Hasan, Sabbir Hossain Sagar, " P2P Video Conferencing System Based on WebRTC", International Conference on electrical, computer and communication engineering (ECCE), February 2017

[6] HuayingXue, Yuan Zhang, "A_WebRTC- Based Video Conferencing System with Screen Sharing", IEEE International Conference on computer and communication, 2016

[7] Baluja Rowley, Kanade "Neural Network-based Face Detection" IEEE Transactions on Pattern Analysis and Machine Intelligence. 20 (1) (2015), pp. 23-38