Songs recommendation based on facial expressions

Mehul Chaudhari[1], Shreyas Vaidya[2], Tejas Mandre[3], Asha Pawar[4]
[1,2,3,4]*Computer Science and Engineering department, MIT School of Engineering*

Abstract **- Machines are becoming powerful day by day. They are getting more human with time. On that note, we created a fully functional system that could take in the facial expressions of the user and recommend/predict the type of music that they want to listen to. The machine learning model uses concepts of convolutional networks to detect the faces in the camera stream. After detection it analyses the frames in the stream and based on the training data set predicts the mood of that picture. Further, this model is put into production using the REST API architecture.**

Index Terms **- Convolutional neural networks, Machine Learning, Recommendation systems, REST API architecture.**

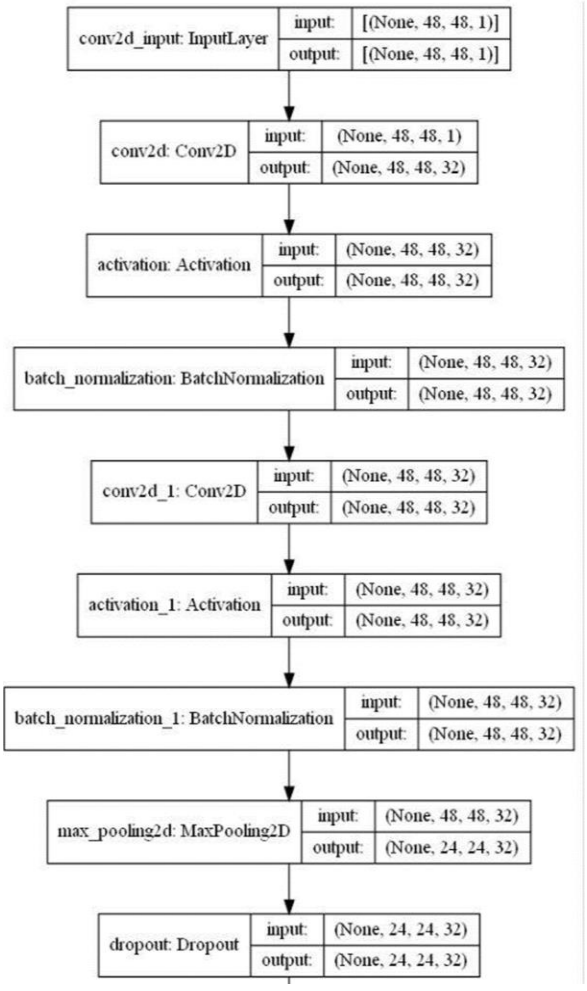
1.INTRODUCTION

Machine learning has proved to be an effective technique to solve most of the problems. We see people are getting drawn more towards a convenient lifestyle be it in any field. Machines are playing a vital role in this. Be it the films you watch or the insurance transactions you do online, machine learning is involved everywhere to provide the user with a better experience. We thought of a similar way to bring changes into the music field. While the current recommendation system is strong enough to predict songs based on the type of music the user has listened to in the past we decided to take it a level up by taking the real live facial expression of the user and recommend/predict the songs they want to listen to. With the use of convolutional neural networks, we created a REST API architecture that could handle the entire data flow right from taking in the camera stream to serving the songs back to the user.

2. CONVOLUTIONAL NEURAL NETWORKS

CNN is an arrangement of many layers of artificial neurons which is aimed at relating the features of an image with some mathematics so that images can be easily differentiated from each other.

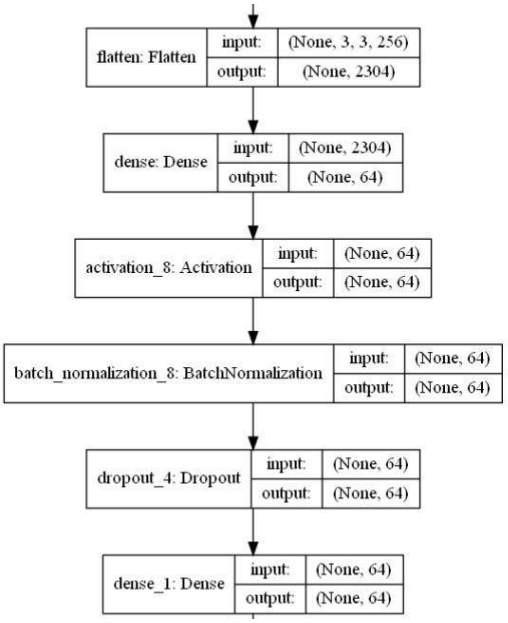
Sequential gom one layer to another layer such that the output of every layer acts as an input for the next layer. But how many layers does our model need? The answer is 7. And how did we end up with that number? An increase in the number of hidden layers means that the model that you have designed will catch the smallest of details and will give out better results. But that also means that at every layer the neuron operations need to be repeated for a large dataset, this will result in a huge increase in the amount of memory that is required for these mathematical relations. The key is to find a sweet spot between accuracy and efficiency.



This is the flow of the first layer of the model. Here 32 is the number of filters that are eventually reduced when using max pooling. (3,3) is the size of the kernel. And the padding has been kept the "same" to protect the dimensionality of the volume. And the activation function used here was "elu".

It was followed by 3 layers with doubling the number of filters in every layer for a more detailed view into the input image.

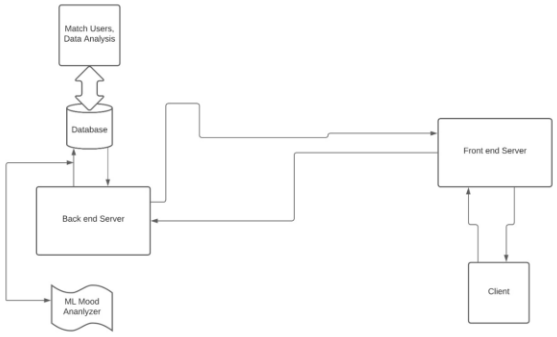
Further, the output is flattened from a 2-D array to a single dimension in the next layer. We apply the dense function so that the neuron in question can get inputs from all the other neurons.



And then the model is put into training giving a validation accuracy of 73%. That is a low number but it works well with whatever real-life examples we have tried.

3. SYSTEM ARCHITECTURE

Real-time camera stream sending to the back end needed socket communication. As we planned to make a web app, WebSockets were used. Websockets connections only need a request initially to activate the connection and the connection does not die out. When the client terminates the session the connection kills itself. Web sockets transmit data in the form of bytes which are decoded on the server-side and formed into frames. The frames are then analysed by the CNN and the appropriate mood along with the list of songs is sent through the REST call.



REST call consists of a GET request to the server with the mood that comes in through the WebSockets. JavaScript helps to build modern front-end layouts which are capable of establishing socket connections and also making asynchronous calls to the server to send the data back and forth. To process these calls and connections from JavaScript, Python forms an excellent choice. We converted the machine learning model into a reusable library that can be imported into any python project. We imported this library to process the requests and analyse facial expressions. Python's ORM is used to manage the database transactions.

Communication protocols used are HTTP and WS. HTTP stands for hypertext transfer protocol and WS stands for web sockets. As a layer of security, these protocols can be converted to https:// and wss:// to encrypt the data that has been sent across the server and client to prevent any man in the middle attacks or packet sniffing.

4. CONCLUSION

The increase stress of exams, career, relationships in life since teenage has contributed to the decrease of happiness and as well as increase in depressions. Music is scientifically proven as mood enhancer as per the genre of that song. Machine learning today is so powerful that it can recognise mood using facial expression. We used CNN to predict/recognise expressions and based on that recommended music to the user. This helps people from sad moods and avoid extra work like choosing the music and genre. Today mental illness is one of the root causes that generations are not productive and depressed at certain phases of life. If we cannot combat the ill effects, generations will surely diminish. More research and innovation can help in protecting public health.

REFERENCES

[1] Facial Expression Recognition via Deep Learning, IEEE -Abir Fathallah, Lotfi Abdi, Ali Douik.
[2] Facial Expression Recognition Using a Hybrid CNN -Tee Connie, Mundher Al-Shabi, Wooi Ping Cheah, Michael Goh.
[3] Facial Expression Recognition with CNN Ensemble,IEEE -Kuang Liu, Mingmin Zhang, Zhigeng Pan.
[4] Deep Learning Models for Facial Expression Recognition, IEEE -Atul Sajjanhar; ZhaoQi Wu; Quan Wen.

Authors
Mehul Chaudhari
Shreyas Vaidya
Tejas Mandre
Asha Pawar