Real Time Emotion Based Music Player Using Image Processing

Mohammed Yasin, Akshit Tomar, Mrs R Amutha (Professor)
Department of Information Science and Engineering
AMC Engineering College
Bangalore, Karnataka

ABSTRACT

This paper implements the use of Haar Cascade Classifier for face detection and Convolutional neural network for the emotion detection and thereby playing a song accordingly. In order to obtain minimal processing, multilayer perceptron are implemented by CNNs. In comparison to various algorithms for image classification, CNNs observed to have little-processing. This implies that the filters used in CNNs are advantageous when compared to traditional algorithm. The model includes two convolutional neural networks (CNN) models: a five-layer model and a global average pooling (GAP) model. We combined these CNN models with transfer-learning models. For our transfer-learning models, we used three pre-trained models: ResNet50; SeNet50; VGG16. Our results are comparable with the state-of-the-art models; however, our models are more efficient in performance. The visualization of features directly can be less informative. Hence, we use the training procedure of back-propagation to activate the filters for better visualization. The multiple actions such as capturing, detecting the emotion and classifying the same can all be confined as one step through the use of CNN. The slow performances of the real-time approaches could be enhanced by regularizing the methods and by visualizing the hidden features. Hence the proposed approach could enhance the accuracy and the computation speed..

Keywords:– *Musicplayer, Convolutional Neural Network, OpenCV, Emotion, Machine learning, CNN Model, Haar Cascade Classifier.*

---

1. INTRODUCTION

Music plays an important role in enhancing an individual's life as it is an important medium of entertainment for music lovers and listeners and sometimes even imparts a therapeutic approach
. Emotion and recognition of faces are one of the basic capabilities of human beings. Extending this capability to machines is of great interest in many application areas. Deep learning-based facial expression recognition is one of these methods to detect emotion state (e.g., anger, fear, neutral, happiness, disgust, sadness and surprise) of human. This method aims to detect facial expressions automatically to identify the emotional state with high accuracy. In this method, labelled facial images from facial expression dataset are sent to CNN and CNN is trained by these images.

In this concept music is recommended to the user by detecting the real time capturing of user's emotions. Existing technique were using collaboration technique which will use previous user data to recommend music and This technique requires lot of manual work so, we proposed a system to arrange different music in different categories such as happy, sad or angry etc. Emotion-Basedmusic-player It's music player with chrome as front-End which has the capability to detect emotions I.e, the face of user with the help of machine learning algorithm using python. Based on the detected user's mood song list will be displayed/recommend to the user. In this application image of a person is captured using a real time machine that has the access to the local machinery. and depending on the captured image it compares the database data sets that already saved in the local device through processing it defines the present mood of the user in numerical form based on this music will be played other than that we have some common features that are queue playlist so that we can have a individual playlist and the last one is random it uses python Eel library so that it can pick a random song with out any order. for this we have used libraries like OpenCV, EEL, numpy etc.this system is mainly proposed because music play a vital role in recent times that is to reduce stress.so,in order to detect the emotion we are using face as a main source of data because normally face expression defines the Emotion so according to the mood we play the music that it can change the user's mood.

2. LITERATURE REVIEW

Emotion detection is a complex process as extracting visual cues from the face is complicated since they are not always obvious. They can be very subtle, and even in some cases, non-existent. There exist models that can detect emotion with great accuracy, but they only do that in a controlled environment. In a real environment, the problem becomes much more challenging as we have to factor in lighting, different facial structures, occlusions, head pose, etc. However, the last decade has caused this field to improve to perform better than typical humans drastically. This can be mainly attributed to the popularity of deep learning algorithms and computer vision. This has led to the rise of various applications such as medical treatments, social robotics, driver fatigue surveillance, etc.

2.1 TITLE: Emotion-Based Music Player

AUTHOR: Krittrin Chankuptarat; Raphatsak Sriwatanaworachai; Supannada Chotipant

DESCRIPTION: Nowadays, people tend to increasingly have more stress because of the bad economy, high living expenses, etc. Listening to music is a key activity that assists to reduce stress. However, it may be unhelpful if the music does not suit the current emotion of the listener. Moreover, there is no music player which is able to select songs based on the user emotion. To solve this problem, this paper proposes an emotion-based music player, which is able to suggest songs based on the user's emotions; sad, happy, neutral and angry. The application receives either the user's heart rate or facial image from a smart band or mobile camera. It then uses the classification method to identify the user's emotion. This paper presents 2 kinds of the classification method; the heart rate-based and the facial image-based methods. Then, the application returns songs which have the same mood as the user's emotion. The experimental results show that the proposed approach is able to precisely classify the happy emotion because the heart rate range of this emotion is wide.

2.2 TITLE: Convolutional Neural Networks Models for Facial Expression Recognition
AUTHOR: Burhanudin Ramdhani, Esmeralda C. Djamal*, Ridwan Ilyas

DESCRIPTION: Emotion is a psychological representation of an event that arises spontaneously in a short time and can be reviewed of them from facial expressions, which facial expressions can indicate consumer satisfaction. Facial recognition as an image can be viewed as identity, emotion, age, race, and gender. This is what makes the extraction of emotional patterns from other patterns not easy. This research has built an image recognition system of emotion expression Customer Satisfaction Recognition Through Emotions using Convolutional Neural Networks (CNN) by comparing two configurations using batch sizes 8 and 128 with two datasets that are FER-2013, a self-created dataset, and a cross dataset against four emotion expressions related to customer service that is happy, disappointed, angry and natural. The test results showed better results on the configuration made by researchers with batch size 8 which achieved the best results of 73.98% on the dataset made by researchers and 58.25% in the FER-2013 dataset. Whereas when using batch size 128 best accuracy achieved by the previous research configuration with the FER-2013 dataset is 69.10%.

2.3 TITLE: Machine Learning Based Music Player by Detecting Emotions
AUTHOR: Supriya L P; Rashmita Khilar

DESCRIPTION: Affective computing is a form of machinery that enables the machine to respond to a human stimulus in some way, usually associated with sophisticated mood or emotional indications. This emotion based music player project is a novel approach that helps the user play songs automatically based on the user's emotions. This understands user's facial emotions and plays the songs according to their emotions. Music as a major impact on the regular life of human beings and in innovative, progressive technologies. Generally the operator needs to do with the challenge of looking for songs manually navigate through the playlist to choose from. At this point it suggests an effective and precise model, which would produce a playlist constructed on the user's present emotional state and behavior. Existing strategies to mechanize the method of creating the playlist are computationally moderate, less solid and some of the time includes the utilization of additional hardware. Discourse is the foremost antiquated and ordinary way of communicating considerations, feelings, and temperament and it requires tall specialized, time, and taken a toll. This proposed framework is based on extricating facial expressions in real-time, as well as extricating sound highlights from tunes to decipher into a specific feeling that will naturally produce a playlist so that the fetched of handling is moderately low. The Emotions are recognized using Support Vector Machine (SVM) . The webcam captures the user's image. It then extracts the user's facial features from the captured image. The music will be played from the pre-defined files, depending on the emotion.

3. PROPOSED METHODOLOGY

This proposed system uses Haar like Features for face detection which uses modified Haar Cascades for detection. Firstly, the algorithm needs a lot of positive images and negative images to train the Haar cascades

classifier. Positive images are images with clear faces where negative images are those without any faces. For this work, we have used the FER2013 dataset [17].

FER2013 is a popular and complex benchmark dataset usedin many competitions and research. It has a human accuracyof 64±5%. It consists of around 36,000 images, which arenormalized to 48x48 grayscale images. The images are dividedinto seven different classes, with each class representing afacial expression. The different classes available are Happy(8,988), Neutral (6,197), Sad (6,076), Angry (4,954), Surprise(4,001), Fear (5,120) and Disgust (548).

The proposed framework actualizes a Deep CNN model and a classifier for recognizing expressions by building a system.. In this methodology the following steps are present: facial emotion images pre- processing, data augmentation, transfer learning using sequential model in convolutional neural networks, feature extraction and ensemble classification. The model proposed is composed of two modes. A pre-trained start mode which transforms the input images into descriptor vectors. Another mode consists of several classifiers strongly connected together where each classifier will give its output its own prediction.

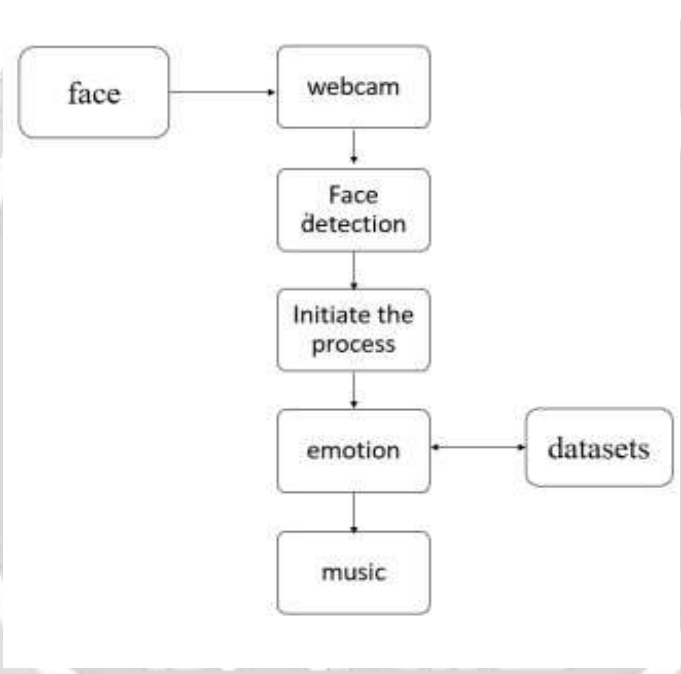

Fig 1. Flow Chart

4. SYSTEM ARCHITECTURE

It consists of the following steps:

1. Data Preparation
2. Feature Extraction using Deep CNN
3. Visualization
4. Training and Testing of Data
5. Identifying the emotions based on the facial expression
6. Based on the emotion opens the specific music file and plays music which describes the emotion of the person

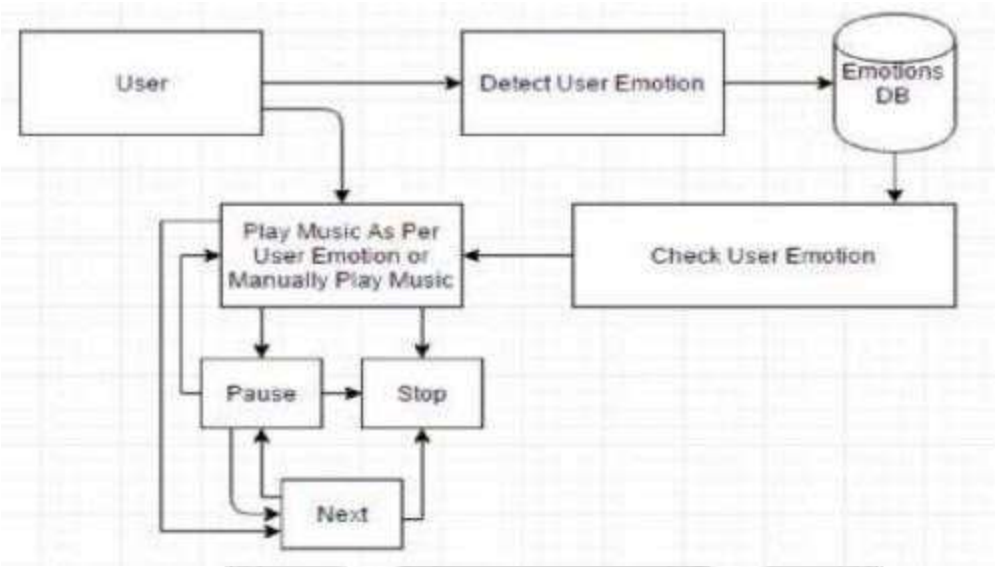

Fig 2. System Architecture

4.1 Face Recognition using Haar Cascade Classifier

Haar_Cascade Classifier is used to detect objects in an image in a set of machine learning algorithms. It was proposed by Paul Viola and Michel for recognizing faces in the given image.
Haar features are different from primary features of images those are color, text and shape. Generally these Haar features are Edge, Line and Rectangle like in convolution networks. It is just like kernel in CNN. Every feature obtained from this is a single value. This can be achieved by subtracting sum of pixels in white area by sum of pixels in black area in one edge.

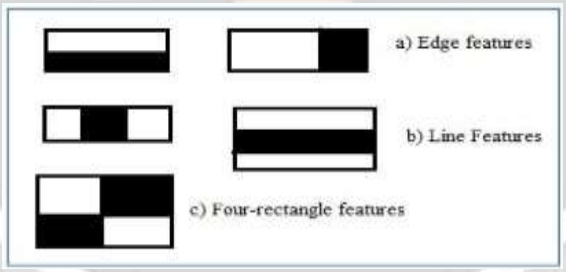

Fig 3. Haar Features

This algorithm passes into four steps. 1. Selection of Haar Features 2. Integral Image Creation 3. Training of input images dataset with ADAboost algorithm. 4. Finally Applying Cascading Classifier for detecting face image and non-face image.

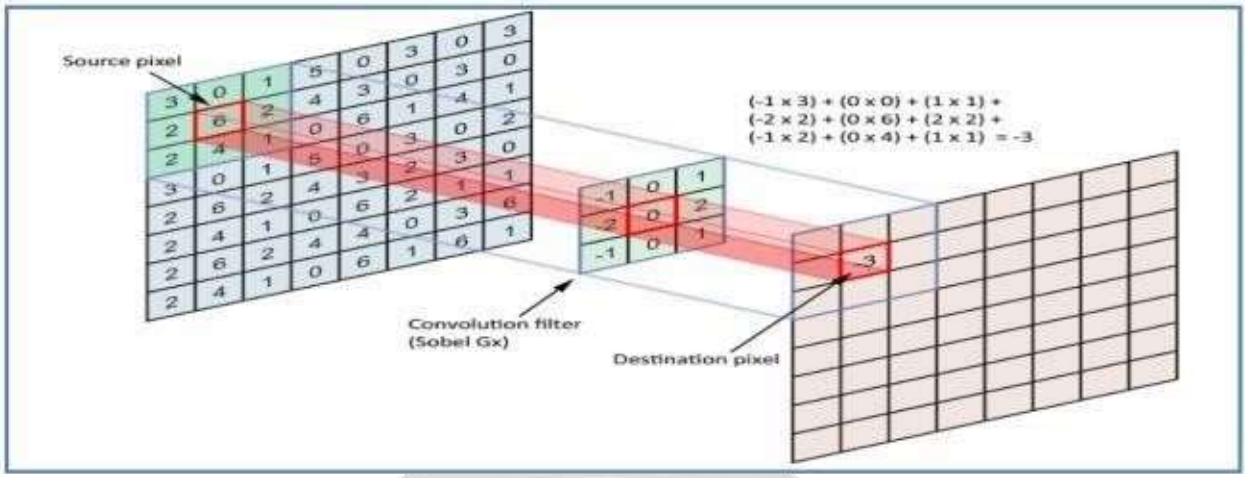

Fig 4. Calculating output one pixel value using four rectangular edge features in the input image

Grayscale conversion and noise removal- Commonly known as smoothening to avoid noise and to get clear data Subtraction operation between the background and the foreground- To differentiate between background and the foreground Apply a threshold to the image resulting from the subtraction- The threshold will ensure that we keep pixels that exceed the defined threshold.

4.2 Emotion Recognition using CNN

CNN is mainly used for image recognition and image classifications. In CNN image classification takes an input image, processes it, and classifies it under certain categories. CNN is another type of neural network that can be used to enable machines to visualize things and perform tasks such as image classification, image recognition, object detection, etc. Image classification is the task of taking an input image and outputting a class (Handgun, Hammer, Axe, Knife, etc) or a probability of classes that best describe the image. CNN is a specialized type of neural network model designed for working with image data.

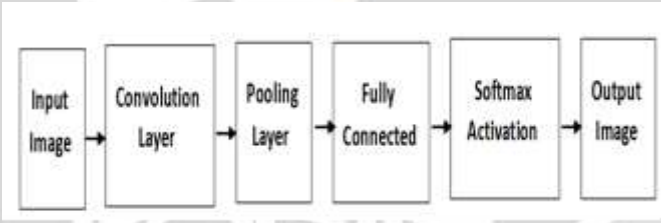

Fig 4. CNN model

The convolutional layer operates called a convolution, hence the neural is called a convolutional neural network. It extracts features from the input image. Convolution is a linear operation that involves the multiple of a set of weights with the input by Paul Viola and Michel for recognizing faces in the given image.(Haar features)

Face emotion recognition implementation using Python, Open CV, CNN. In this, the deep learning model CNN architecture first detects the face and based on facial features, it specifies the various emotions like happy, sad, surprise, angry, neutral, etc

As a programming language, we are using python. For loading an image and to overlink this rectangle and the text data, we use open cv. For face recognition, we use a pretrained models called CNN which contains deep learning architectures which not only detects the facial emotions, but also the race, gender of a person.

Tools we need are anaconda which is a package of multiple libraries and Integrated Development Environment. IDE is a coding tool which allows you to write, test, and debug your code in an easier way, as they typically offer code completion or code insight by highlighting, resource management, debugging tools.

- The computer vision captures the image of the user to consider it to identify the facial expression.
- The matplotlib will plot the possible libraries for the desired output.

- The CNN Haar_Cascade analysis is carried out on the image and predictions are obtained.
- The dominant emotion is identified from the predictions and the rectangle bounding of the detected face is done.
- On the top left of the recognized green frame the dominant emotion is displayed and the emotion recognition is continuous for a given image, video or a live screen.
- Based on the emotion represented by the facial expression opens the specific music file and plays music

5 RESULTS

The system developed was successfully able to identify the dominant emotions such as: 'Happy', 'Sad', 'Surprised' ,'angry','surprised',disgust,and 'Neutral'. And play music accordingly

Fig 6. Detection of Emotion

Fig 7. Emotion based music player

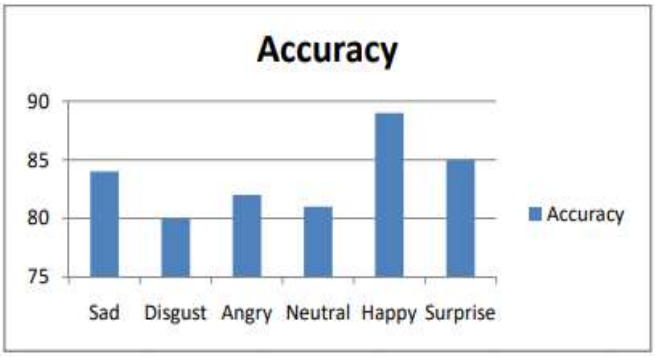

Chart -1. All six expressions are accurate on test data. Identified happy expression is nearly 89% accurate using the Deep CNN algorithm.

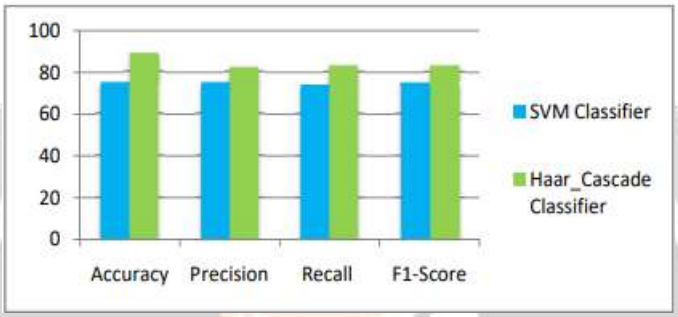

Chart -2. Comparison of Proposed Haar_cascade Classifier with SVM for detecting faces in Images

6 CONCLUSION

Facial Features are one of the most potent channels for Emotion recognition. Convolutional Neural Networks (CNN) can be used as a solution in recognizing Emotion As facial expression recognition technology improves in accuracy, the range of its application will grow, both in the innovation improves in precision, the scope of its application will develop, both in mechanical turn of events and past. In this paper, an emotion detection model is proposed to recommend music based on one's mood. Our work aims to achieve the highest possible accuracy while not compromising the real-time aspect to apply to the real-world scenario. We explored several models built differently, including vanilla CNNs and pre-trained networks based on ResNet50, SenNet50, and VGG16. One model that stood out was the GAP model that managed to achieve an accuracy of 66.54% while reducing the number of parameters by around 80%. This was a breakthrough as such a lightweight model is easily mountable on small devices, which adds to real-world scenarios' applicability. We further solved the challenging class imbalance problem of the FER2013 dataset by using class weighting and data augmentation.

7 REFERENCES

[1] Krittrin Chankuptarat;Raphatsak Sriwatanaworachai;Supannada Chotipant 2019 5th International Conference on Engineering, Applied Sciences and Technology (ICEAST)Year:
2019 | Conference Paper | Publisher: IEEE.
[2] Music player based on emotion recognition of voice signals Sneha Lukose;Savitha S. Upadhya2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)
[3] S. Kumar and M. K. Bhuyan, "Neutral expression modeling in feature domain for facial expression recognition", IEEE Recent Advances in Intelligent Computational Systems (RAICS), 10- 12 Dec. 2015.
[4] A Machine Learning Based Music Player by Detecting Emotions ;S. Deebika;K. A. Indira;Jesline 2019 Fifth International Conference on Science Technology Engineering and Mathematics (ICONSTEM)
[5] Music recommendation system based on user's sentiments extracted from social networks Renata Lopes Rosa;Demóstenes Zegarra Rodríguez;Graça Bressan 2015 IEEE International Conference on Consumer Electronics (ICCE)

[6] U. Ayvaz, H. Gürüler and M. O. Devrim, "USE OF FACIAL EMOTION RECOGNITION IN ELEARNING SYSTEMS", Information Technologies and Learning Tools, vol. 60, no. 4, pp. 95, sept. 2017.
[7] Gerard Pons and David Masip, Supervised Committee of Convolutional Neural Networks in Automated Facial Expression Analysis, IEEE Transaction 2017.
[8] Zirujirao, FengchunQiao, NaimingYao, ZhihaoLi, HuiChen, Hongan Wang, An Ensemble of VGG Networks for Video-Based Facial Expression Recognition, First Asian Conference on
Affective Computing and Intelligent Interaction (ACII Asia),2018
[9] Insha Rafique, Awais Hamid, Sheraz Naseer, Muhammad Asad, Muhammad Awais, Talha Yasir, Age and Gender Prediction using Deep Convolutional Neural Networks, International Conference on Innovative Computing (ICIC),2019
[10] B. Yang, J. Cao, R. Ni and Y. Zhang, "Facial Expression Recognition Using Weighted Mixture Deep Neural Network Based on Double-Channel Facial Images", IEEE Access, vol. 6, pp. 4630- 4640, 2018.
[11] S R Livingstone and FA Russo, "The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic multimodal set of facial and vocal expressions in North American English", PLoS ONE, vol. 13, no. 5, pp. e0196391, 2018.
[12] N. Veeranjaneyulu, Jyostna Devi Bodapati, " Facial Emotion Recognition Using Deep Cnn Based Features ISSN: 2278-3075, vol. 8, no. 7, May 2019.
[13] W. Deng and S. Li, "Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition", IEEE Transactions on Image Processing, vol. 28, no.1, pp. 356-370, 2019.
[14] Moulay Smail Bouzakraoui , Abdelalim Sadiq, Abdessamad Youssfi Alaoui, "Appreciation of Customer Satisfaction Through Analysis Facial Expressions and Emotions Recognition", 9781-7281- 1232-9, IEEE, 2019.