GENERATING MUSIC PLAYLIST BASED ON FACIALEXPRESSIONS USING DEEP LEARNING-CNN

Dr. Varanasi Usha Bala[1], B. Bharathi[2], V. Yaswanth Sai[2], T. Sushma Sree[2], B. Jyothi[2]

[1]Assistant Professor, Department of CSE, Anil Neerukonda Institute Of Technology And Sciences (A), Visakhapatnam-531162, India
[2]Final year students of Department of CSE, Anil Neerukonda Institute Of Technology And Sciences (A), Visakhapatnam-531162, India

---

ABSTRACT: Recent studies confirm that humans respond and react to music which music encompasses a high impact on person's brain activity. the common American listens up to four hours of music on a daily basis. Everyone wants to pay attention music of their individual taste, mostly supported their mood. Users always face the task of manually browsing the music and to form a playlist supported their current mood. The proposed project is incredibly efficient which generates a music playlist supported this mood of users. Facial expressions are the most effective way of expressing ongoing mood of the person. the target of this project is to suggest songs forusers supported their mood by capturing facial expressions. Facial expressions are captured through webcam and such expressions are fed into learning algorithm which provides most probable emotion. Once the emotionis recognized, the system suggests a play-list for that emotion, thus saves plenty of your time for a user. Oncethe emotion is detected by CNN then the emotion is employed by Spotify API then the Spotify API generatesa playlist according the emotion of the user. stock prices accurately. The experiment results show that prediction accuracy is over 80%.

KEYWORDS: Image Processing, Face detection, Emotion recognition, Webcam, CNN classification, Spotify API, Music Playlist.

1. INTRODUCTION

Music plays a vital role in our way of life. Users need to face the task of manually browsing the music. Computer vision could be a field of study which encompasses on how computer see and understand digital images and videos. Computer vision involves seeing or sensing a visible stimulus, be of what it's seen and also extract complex information that might be used for other machine learning activities. we are going to implement our use case using the Haar Cascade classifier. Haar Cascade classifier is a good object detection approach which was proposed by Paul Viola and Michael Jones in their paper, "Rapid Object Detection employing a Boosted Cascade of straightforward Features" in 2001. This project recognizes the facial expressions of user and play songs in step with emotion. Facial expressions are best way of expressing moodof an individual. The facial expressions are captured employing a webcam and face detection is finished by using Haar cascade classifier. The captured image is input to CNN which learn features and these features areanalyzed to work out this emotion of user then the music are played in keeping with the emotion. during this project, five emotions are considered for classification which has happy, sad, anger, surprise, neutral. This project consists of 4 modules-face detection, feature extraction, emotion detection, songs classification. Face detection is finished by Haar cascade classifier, feature extraction and emotion detection are done by CNN. Finally, the songs are played consistent with the emotion recognized. CNN may be a sort of deep learning model for processing data that features a grid pattern, like images, which is inspired by the organization of animal visual area and designed to automatically and adaptively learn spatial hierarchies of features, from low- to high-level patterns.

2. LITERATURE SURVEY

Face Expression Recognition Using CNN & LBP, Visual interaction is an efficient means of communication for individuals as social beings. Even a straightforward change in countenance signifies happiness, sorrow, surprise and anxiety. The facial expressions of each person should vary in various contexts like lighting, posture and even background. of these factors still remain a problem while recognizing facial expressions. This paper hopes to bring out a good comparison between two of the foremost commonly used face recognition[FER] techniques and to shed some light on their precision. The methods being employed here are Local binary patterns [LBP] and Convolution neural networks [CNN]. The LBP is supposed as a way just for the aim of extracting features therefore the Support vector machine [SVM] classifier is being utilized for classifying the extracted features from LBP. The dataset used for the aim of testing and training during this paper are CK+, JAFFE and YALE FACE.

A Machine Learning Based Music Player by Detecting Emotions, this paper constitutes the implementation of Convolutional neural network for the emotion detection and thereby playing a song accordingly. so as to get minimal processing, multilayer perceptron is implemented by CNNs. compared to numerous algorithms for image classification, CNNs observed to possess little-processing. this means that the filters utilized in CNNs are advantageous when put next to traditional algorithm. The visualization of features directly are often less informative. Hence, we use the training procedure of back-propagation to activate the filters for better visualization. The multiple actions like capturing, detecting the emotion and classifying the identical can all be confined in concert step through the utilization of CNN.

Emotion-Based Music Player, this paper proposed an emotion-based music player, which is ready to suggest songs supported the user's emotions; sad, happy, neutral and angry. the appliance receives either the user's pulse or facial image from a wise band or mobile camera. It then uses the classification method to spot the user's emotion. This paper presents 2 varieties of the classification method; the center rate-based and therefore the facial image-based methods. Then, the applying returns songs which have the identical mood because theuser's emotion. The user and song emotions during this paper are divided into four types namely: neutral, happy, sad and angry. The experimental results present that detecting the happy emotion is that the most precise with around 98%, while the accuracy of the sad mood detection is that the lowest with 40%.

Automatic facial features recognition using features of salient facial patches, they proposed a system image from database is passed to the facial landmark detection stage to get rid of noise by applying Gaussian Filter or mask. Here itself they used Viola Jones technique of Haar-like features with Adaboost learning for face detection. The feature detection stage consists of Eyebrow corners detector, Eye detector, Noise detector, Lipcorner detector. After these active facial patches are extracted, the classification of features is completed by SVM (Support Vector Machine). While testing it'll take the many images from the database and extract the features and classifies accordingly. They used CK+ (Cohn-Kanade) dataset and JAFEE dataset for training and testing the database. The training database carries with it 329 images in total.

3. DATASET DETAILS

Fer-2013 dataset was prepared by Pierre-Luc Carrier and Aaron Courville, as a part of an ongoing research. They need graciously provided the workshop organizers with a preliminary version of their dataset to use for this contest.

The data consists of 48x48 pixel grayscale images of faces. The faces are automatically registered so the face is more or less centered and occupies about the identical amount of space in each image. The task is to categorize each face supported the emotion shown within the facial features in to at least one of seven categories (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral).

	emotion	pixels	Usage
0	0	70 80 82 72 58 58 60 63 54 58 60 48 89 115 121...	Training
1	0	151 150 147 155 148 133 111 140 170 174 182 15...	Training
2	2	231 212 156 164 174 138 161 173 182 200 106 38...	Training
3	4	24 32 36 30 32 23 19 20 30 41 21 22 32 34 21 1...	Training
4	6	4 0 0 0 0 0 0 0 0 0 0 0 3 15 23 28 48 50 58 84...	Training

The train.csv contains two columns, "emotion" and "pixels". The "emotion" column contains a numeric code starting from 0 to six, inclusive, for the emotion that's present within the image. The "pixels" column contains a string surrounded in quotes for every image. The contents of this string a space-separated pixel values in row major order. test.csv contains only the "pixels" column and your task is to predict the emotion column.

This dataset consists of 35,887 grayscale images. The training set consists of 28,709 examples. the generalpublic test set consists of three,589 examples.

4. METHEDOLOGIES

FACE DETECTION:

Detection of face from a given input image or video is face detection. There are various algorithms for face detection. Viola jones algorithm is used for face detection. The main steps in Viola Jones algorithm are:

HAAR feature:

HAAR features represent some characteristics of the face. Haar features are similar to those convolution kernels which are used to detect the presence of the feature in the given image. Each feature result in a single value which is calculated by subtracting the sum of pixels under white rectangle from the sum of pixels under black rectangle. In the feature, the black region is replaced by plus ones and white region is minus one.

Integral image:

In haar feature calculation, as every time window moves need to sum up all pixels of the black region and those of white region. It is a tedious operation and the solution is integral image. It reduces the computation rather than summing up all pixels under a rectangle with just four corner values of the integral image. To find the value of any pixel just sum the values of pixels to the top and left.

Adaboost:

Viola jones algorithm makes use of 24*24 window as the base window for starting the evaluation of features in any given image. If we consider all possible parameters of haar features like position scale and type we end up calculating 160,000+ features in this window which is practically impossible. So, the basic idea is to eliminate a lot of features which are redundant or which are not useful and select only the features that are very useful. This one by Adaboost eliminating 160 thousand features and narrowing down to only couple of thousands of features which we need to evaluate. The features extracted by Adaboost is weak classifiers.

Adaboost constructs a linear combination of the weak classifier.

Cascading:

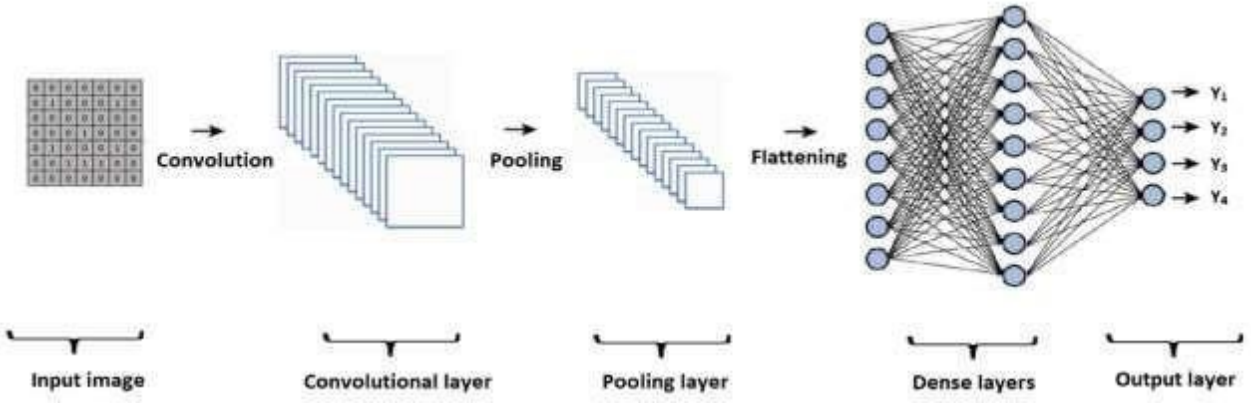
It is the basic principle of Viola Jones face detection algorithm to scan the detector many times through that image itself, each time with a new size. Though an image should contain one or more faces it is clear that an excessive large amount of the evaluated sub-windows might still be negatives. He algorithm should hence concentrate on discarding non-faces quickly. Therefore, a single strong classifier formed out of linear combination of all the best features is not good to evaluate on each window because of computation cost.

FACIAL FEATURE EXTRACTION AND EMOTION DETECTION:CONVOLUTION NEURAL NETWORK:

CNN is an efficient recognition algorithm which is widely employed in pattern recognition and image processing. it's many features like simple structure, less training parameters and flexibility.

A Convolutional Neural Network (ConvNet/CNN) could be a Deep Learning algorithm which may soak up an input image, assign importance (learnable weights and biases) to numerous aspects/objects within the image and be able to differentiate one from the opposite. The pre-processing required in a very ConvNet is way lower as compared to other classification algorithms. The role of the ConvNet is to scale back the photographs into a form which is less complicated to process, without losing features which are critical for getting an honest prediction.

This is important once we are to style an architecture which isn't only good at learning features but is also scalable to massive datasets. A CNN typically has three layers: a convolutional layer, a pooling layer, and a totally connected layer.



Convolution Layer — The Kernel:

The element involved in winding up the convolution operation within the first a part of a Convolutional Layeris termed the Kernel/Filter. the target of the Convolution Operation is to extract the high-level features like edges, from the input image. ConvNets needn't be limited to just one Convolutional Layer. Conventionally, theprimary ConvLayer is chargeable for capturing the Low-Level features like edges, color, gradient orientation, etc. With added layers, the architecture adapts to the High-Level features still. ReLU (Rectified Linear Unit) activation function which is applied after the convolution operation. it's used for bringing non-linearity to the model. It simply converts the negative values present within the feature map to '0'.

Pooling Layer:

The Pooling layer is answerable for reducing the spatial size of the Convolved Feature. this is often to decrease the computational power required to process the information through dimensionality reduction. Furthermore, it's useful for extracting dominant features which are rotational and positional invariant, thus maintaining the method of effectively training of the model. There are two sorts of Pooling: Max Pooling and Average Pooling. Max Pooling returns the utmost value from the portion of the image covered by the Kernel. On the opposite hand, Average Pooling returns the common of all the values from the portion of theimage covered by the Kernel.

Classification — Fully Connected Layer (FC Layer):

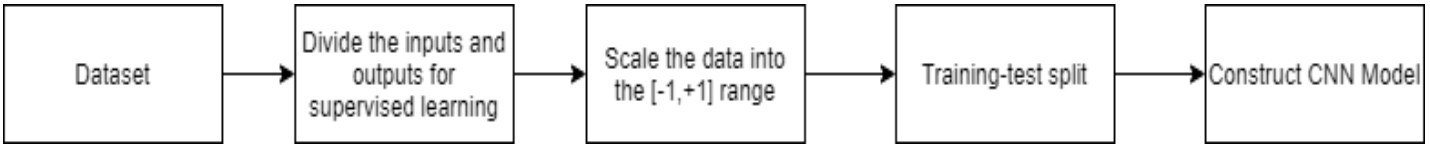
Neurons during this layer have full connectivity with all neurons within the preceding and succeeding layer as seen in regular FCNN. Fully Connected Layer is additionally called as Dense Layer. It provides learning features from all the combinations of the features of the previous layer. The FC layer helps to map the representation betn the input and therefore the output. The flattened output is fed to a feed-forward neural network and backpropagation applied to each iteration of coaching. Over a series of epochs, the model is ready to differentiate between dominating and certain low-level features in images and classify them using the SoftMax Classification technique

PLAYLIST GENERATION:

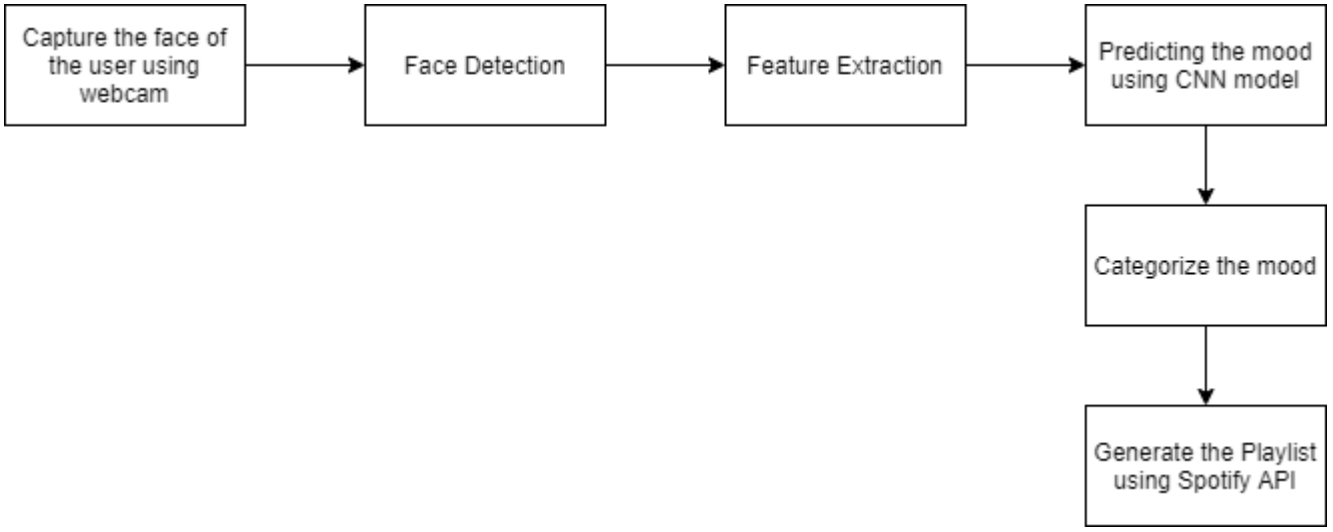
The output of neural network classifier is that the emotion of the user. Then the output emotion is given to Spotify API. Using the detected emotion of the user by CNN the Spotify API generates the playlist according the emotion.

5. SYSTEM ARCHITECTURE

PREPROCESSING OF DATA:



ARCHITECTURE OF THE SYSTEM:



Data Selection: First step is to select data for an organization and split the data as training and testing. In thisproject we have used 75% for training and 25% for testing purpose.

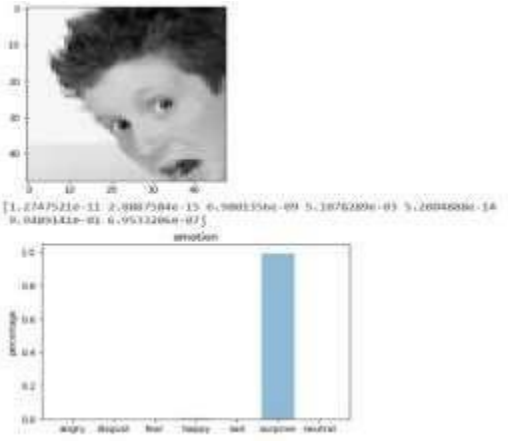
Pre-processing of data: In pre-processing we are selecting attributes required for the algorithm and remainingattributes are neglected. In pre-processing we are using normalization to get values in a particular range.

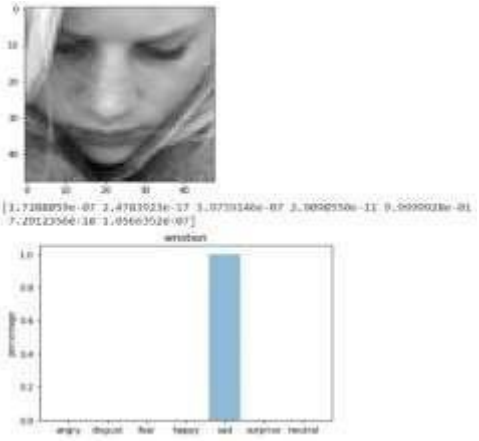
Detection of emotion using CNN: In this system we are using CNN for predicting the emotion of the user. Initially the training data is passed through system and constructing the model. Then in testing phase, using the image of the user the emotion is detected.

Generating playlist using Spotify API: After the emotion is detected by using the Spotify API playlist is generated according to the emotion of the user.

6. EXPERIMENTAL RESULTS

- Number of instances: 35888.
- Instance length: 2304
- 28709 train samples
- 3589 test samples
- Epochs: 100

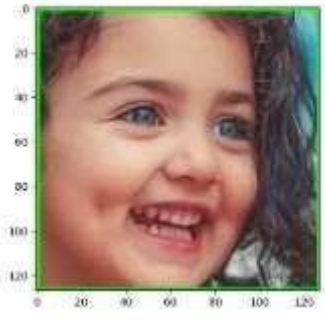


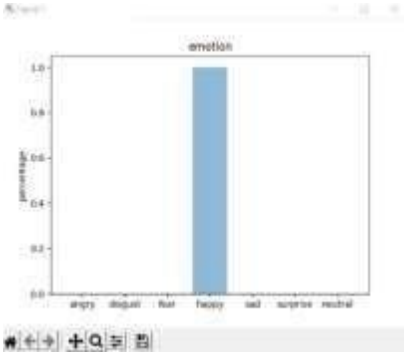


ACCURACY MEASURE:

Emotion Detected: Surprise

Emotion Detected: SadAccuracy: 99.489





OUTPUT:

Emotion Detected: HappyAccuracy: 99.924

Playlist for Emotion: Happy

7. CONCLUSION

In this project, we are generating the playlist according the emotion of the user, we developed an application for predicting the emotion of the user using Convolution neural networks and for generating the playlist we have used Spotify API. We have applied it on various images and achieved an accuracy of above 80%.

8. REFERENCES

[1] S.L. Happy and A. Routray, "Automatic facial expression recognition using features of salient facial patches," in IEEE Transactions on Affective Computing, vol. 6, no. 1, pp. 1-12, 1 Jan.-March 2015.

[2] Rahul Ravi, S.V Yadhukrishna, Rajalakshmi, Prithvi raj, "A Face Expression Recognition Using CNN & LBP",2020 IEEE.

[3] Krittrin Chankuptarat, Raphatsak Sriwatanaworachai, Supannada Chotipant Emotion-Based Music Player, 978-1-7281-0067-8/19/$31.00 ©2019 IEEE.

[4] Karthik Subramanian Nathan, Manasi Arun, Megala S Kannan," EMOSIC — An emotion-based music player for Android, 2017 IEEE.

[5] S.Deebika, K.A.Indira, Jesline, "A Machine Learning Based Music Player by DetectingEmotions", 2019 IEEE.