TABLE OFCONTENTS

LIST OF FIGURES

REAL TIME FACIAL EXPRESSION BASED SMART MUSIC PLAYER

ABSTRACT

The face is an important aspect in predictinghuman emotions and mood.Usuallythe human emotions are extracted with the use of camera. There are many applications getting developed based on detection of human emotions. Few applications of emotion detection are business notification recommendation, e-learning, mental disorder and depression detection, criminal behaviour detection etc. In this proposed system, we develop a prototype in recommendation of dynamic music recommendation system based on human emotions. Based on each human listening pattern, the songs for each emotions are trained. Integration of feature extraction and machine learning techniques, from the real face the emotion are detected and once the mood is derived from the input image, respective songs for the specific mood would be played to hold the users.In this approach, the application gets connected with human feelings thus giving a personal touch to the users.Therefore our projected system concentrate on identifying the human feelings for developing emotion based music player using computervision and machine learning techniques. For experimental results, we use openCV for emotion detection and music recommendation.

CHAPTER 1

INTRODUCTION

People tend to express their emotions, mainly by their facial expressions. Music has always been known to alter the mood of an individual. Capturing and recognizing the emotion being voiced by a person and displaying appropriate songs matching the one's mood and can increasingly calm the mind of a user and overall end up giving a pleasing effect. The project aims to capture the emotion expressed by a person through facial expressions. A music player is designed to capture human emotion through the web camera interface available on computing systems. The software captures the image of the user and then with the help of image segmentation and image processing techniques extracts features from the face of a target human being and tries to detect the emotion that the person is trying to express. The project aims to lighten the mood of the user, by playing songs that match the requirements of the user by capturing the image of the user. Since ancient times the best form of expression analysis known to humankind is facial expression recognition. The best possible way in which people tend to analyze or conclude the emotion or the feeling or the thoughts that another person is trying to express is by facial expression. In some cases, mood alteration may also help in overcoming situations like depression and sadness. With the aid of expression analysis, many health risks can be avoided, and also there can be steps taken that help brings the mood of a user to a better stage.

1.1 PROPOSED ALGORITHMS

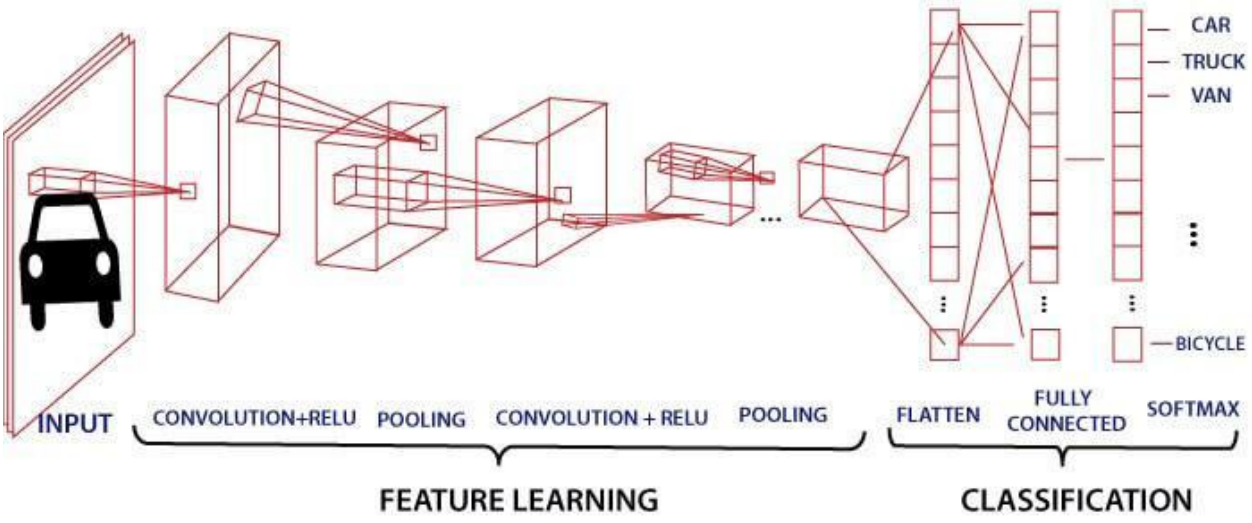
CNN ALGORITHM

Convolutional Neural Network is one of the main categories to do image classification and image recognition in neural networks. Scene labeling, objects detections, and face recognition, etc., are some of the areas where convolutional neural networks are widely used.

CNN takes an image as input, which is classified and process under a certain category such as dog, cat, lion, tiger, etc. The computer sees an image as an array of pixels and depends on the resolution of the image. Based on image resolution, it will see as **h * w * d**, where h= height w= width and d= dimension. For example,

An RGB image is **6 * 6 * 3** array of the matrix, and the grayscale image is **4 * 4 * 1** array of the matrix.

In CNN, each input image will pass through a sequence of convolution layers along with pooling, fully connected layers, filters (Also known as kernels). After that, we will apply the Soft-max function to classify an object with probabilistic values 0 and 1.



Convolution Layer

Convolution layer is the first layer to extract features from an input image. By learning image features using a small square of input data, the convolutional layer preserves the relationship between pixels. It is a mathematical operation which takes two inputs such as image matrix and a kernel or filter.

- The dimension of the image matrix is $\mathbf{h \times w \times d}$.
- The dimension of the filter is $\mathbf{f_h \times f_w \times d}$.
- The dimension of the output is $\mathbf{(h-f_h+1) \times (w-f_w+1) \times 1}$.

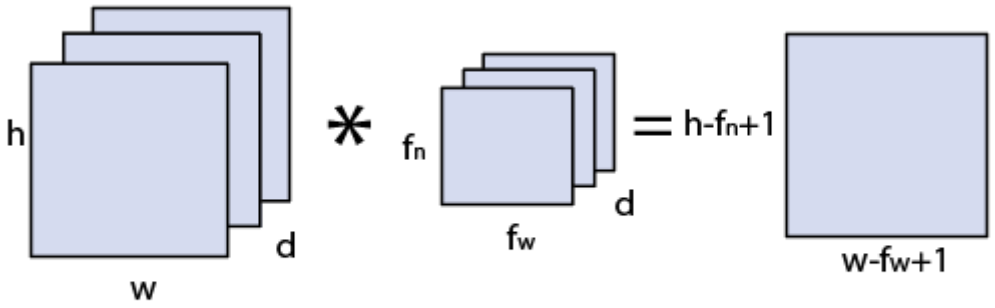


Image matrix multiplies kernl or filter matrix

Let's start with consideration a 5*5 image whose pixel values are 0, 1, and filter matrix 3*3 as:

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

5×5 – Image Matrix 3×3 – Filter Matrix

The convolution of 5*5 image matrix multiplies with 3*3 filter matrix is called "**Features Map**" and show as an output.

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 3 & 4 \\ 2 & 4 & 3 \\ 2 & 3 & 4 \end{bmatrix}$$

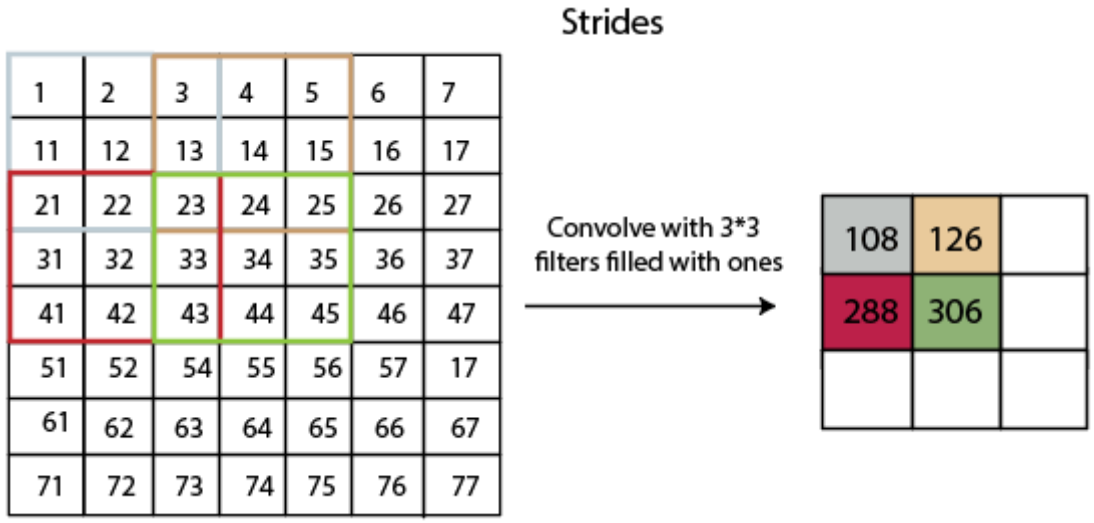
Convolved Feature

Convolution of an image with different filters can perform an operation such as blur, sharpen, and edge detection by applying filters.

Strides

Stride is the number of pixels which are shift over the input matrix. When the stride is equaled to 1, then we move the filters to 1 pixel at a time and similarly, if

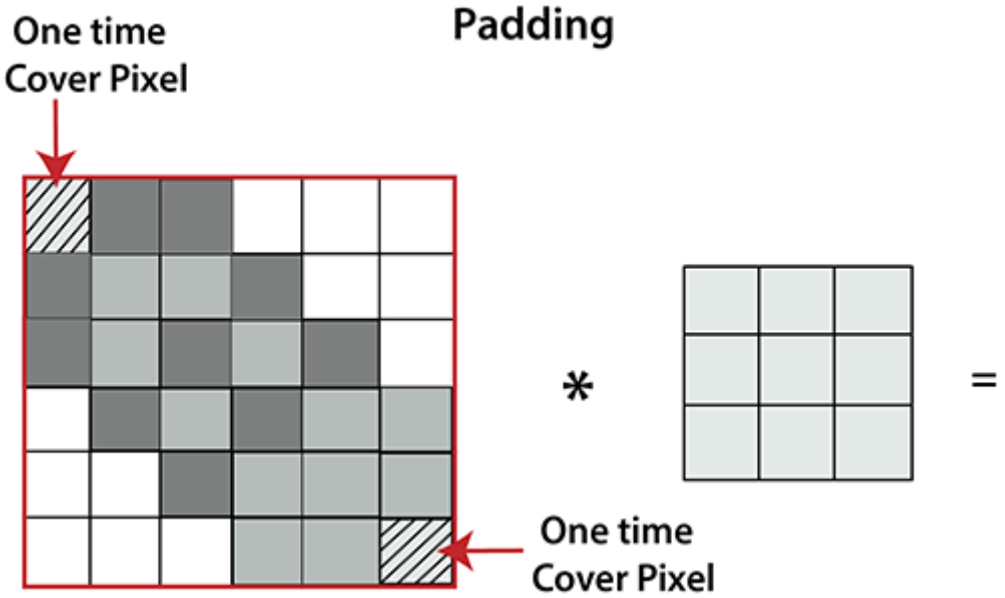
the stride is equaled to 2, then we move the filters to 2 pixels at a time. The following figure shows that the convolution would work with a stride of 2.



Padding

Padding plays a crucial role in building the convolutional neural network. If the image will get shrink and if we will take a neural network with 100's of layers on it, it will give us a small image after filtered in the end.

If we take a three by three filter on top of a grayscale image and do the convolving then what will happen?



It is clear from the above picture that the pixel in the corner will only get covers one time, but the middle pixel will get covered more than once. It means that we have more information on that middle pixel, so there are two downsides:

- Shrinking outputs
- Losing information on the corner of the image.

To overcome this, we have introduced padding to an image. **"Padding is an additional layer which can add to the border of an image."**

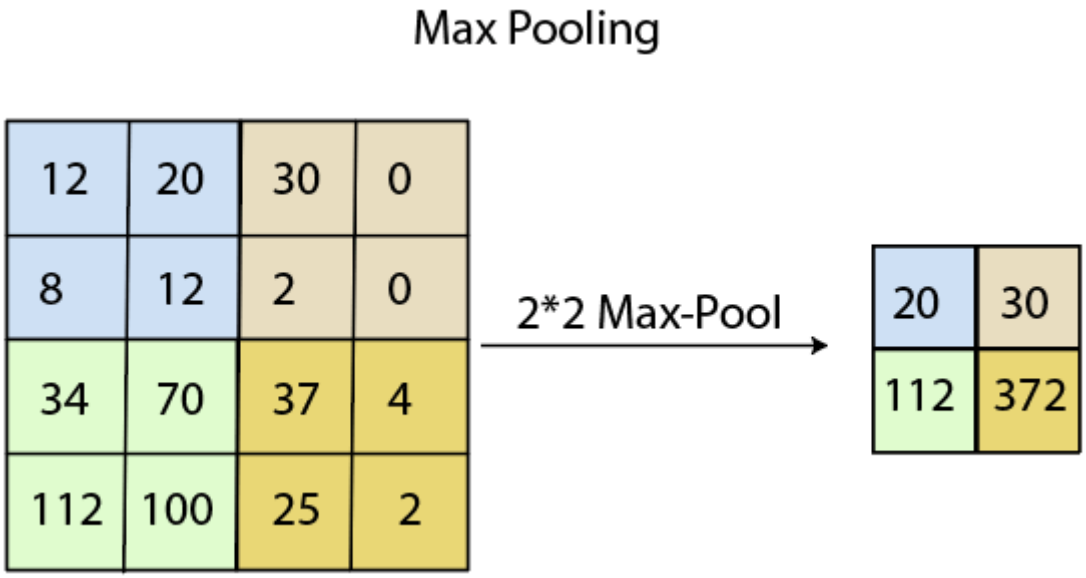
Pooling Layer

Pooling layer plays an important role in pre-processing of an image. Pooling layer reduces the number of parameters when the images are too large. Pooling is "**downscaling**" of the image obtained from the previous layers. It can be compared to shrinking an image to reduce its pixel density. Spatial pooling is also called downsampling or subsampling, which reduces the dimensionality of each map but retains the important information. There are the following types of spatial pooling:

Max Pooling

Max pooling is a **sample-based discretization process**. Its main objective is to downscale an input representation, reducing its dimensionality and allowing for the assumption to be made about features contained in the sub-region binned.

Max pooling is done by applying a max filter to non-overlapping sub-regions of the initial representation.



Average Pooling

Down-scaling will perform through average pooling by dividing the input into rectangular pooling regions and computing the average values of each region.

Syntax

layer = averagePooling2dLayer(poolSize)

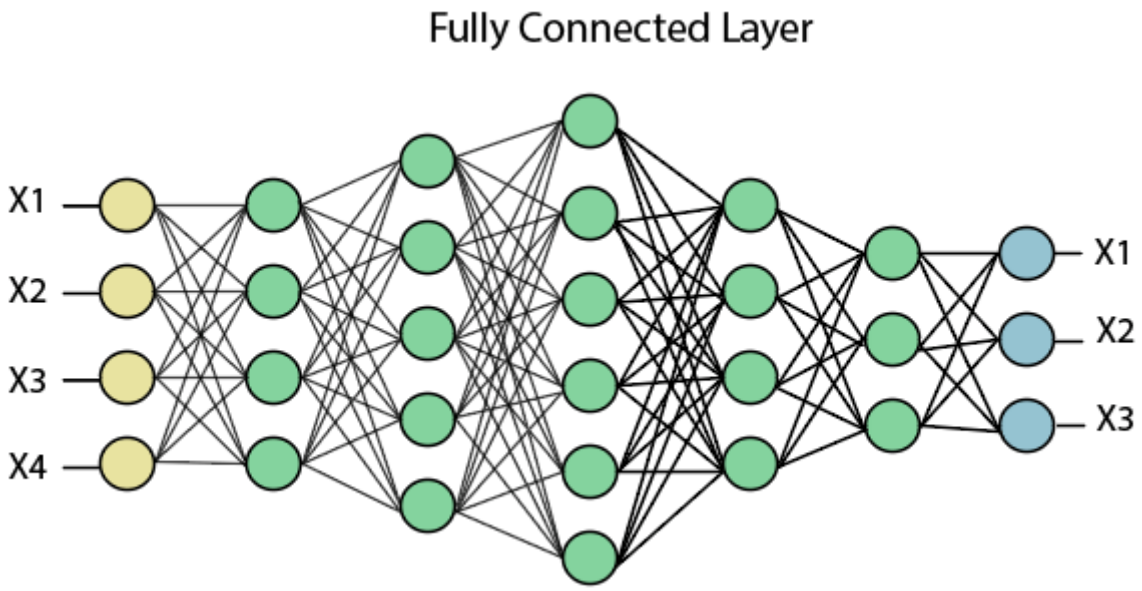
layer = averagePooling2dLayer(poolSize,Name,Value)

Sum Pooling

The sub-region for **sum pooling** or **mean pooling** are set exactly the same as for **max-pooling** but instead of using the max function we use sum or mean.

Fully Connected Layer

The fully connected layer is a layer in which the input from the other layers will be flattened into a vector and sent. It will transform the output into the desired number of classes by the network.



In the above diagram, the feature map matrix will be converted into the vector such as **x1, x2, x3... xn** with the help of fully connected layers. We will combine features to create a model and apply the activation function such as **softmax** or **sigmoid** to classify the outputs as a car, dog, truck, etc.

CHAPTER 2
LITERATURE SURVEY

Smart Music Player Integrating Facial Emotion Recognition

Songs, as a medium, have always been a popular choice to depict human emotions. We validate our models by creating a real-time vision system which accomplishes the tasks of face detection and emotion classification simultaneously in one blended step using our proposed CNN architecture. Reliable emotion-based classification systems can go a long way in facilitating emotions. However, research in the field of emotionbased music classification has not yielded optimal results. In this paper, we present an affective cross-platform music player, EMP, which recommends music based on the real-time mood of the user. EMP provides smart mood-based music recommendation by incorporating the capabilities of emotion context reasoning within our adaptive music recommendation system. Our music player contains three modules: Emotion Module, Music Module and Integrating Module. The Emotion Module takes an image of the user as an input and makes use of deep learning algorithms to identify the mood of the user with an

accuracy of 90.23%.

Mood based Music Recommendation System

A user's emotion or mood can be detected by his/her facial expressions. These expressions can be derived from the live feed via the system's camera. A lot of research is being conducted in the field of Computer Vision and Machine Learning (ML), where machines are trained to identify various human emotions or moods. Machine Learning provides various techniques through which human emotions can be detected. One such technique is to use MobileNet model with Keras, which generates a small size trained model and makes Android-ML integration easier.Music is a great connector. It unites us across markets, ages, backgrounds, languages, preferences, political leanings and income levels. Music players and other streaming apps have a high demand as these apps can be used anytime, anywhere and can be combined with daily activities, travelling, sports, etc. With the rapid development of mobile networks and digital multimedia technologies, digital music has become the mainstream consumer content sought by many young people.People often use music as a means of mood regulation, specifically to change a bad mood, increase energy level or reduce tension. Also, listening to the right kind of music at the right time may improve mental health. Thus, human emotions have a strong relationship with music.In our proposed system, a mood-based music player is created which performs real time mood detection and suggests songs as per detected mood. This becomes an additional feature to the traditional music player apps that come pre-installed in our mobile phones. An important benefit of incorporating mood detection is customer satisfaction. The objective of this system is to analyse the users image, predict the expression of the user and suggest songs suitable to the detected mood.

An Emotion-Aware Personalized Music Recommendation System Using a Convolutional Neural Networks Approach

Recommending music based on a user's music preference is a way to improve user listening experience. Finding the correlation between the user data (e.g., location, time of the day, music listening history, emotion, etc.) and the music is a challenging task. In this paper, we propose an emotion-aware personalized music