Real Time Emotion Recognition Music Player

Anand Pratap(1322871)
Frankfurt University of Applied Sciences
anand.pratap@stud.fra-uas.de

Karishma(1322486)
Frankfurt University of Applied Sciences
karishma.karishma@stud.fra-uas.de

Munshi Fahim Sadi(1322868)
Frankfurt University of Applied Sciences
munshi.sadi@stud.fra-uas.de

Abdullah AlNoman(1323731)
Frankfurt University of Applied Sciences
abdullah.alnoman@stud.fra-uas.de

Sohail Dua(1322512)
Frankfurt University of Applied Sciences
sohail.duap@stud.fra-uas.de

Abstract

The human face is an essential component of a body. It shows behavioural and emotional state of the individual face and it is used to extract the required input from the face that can be done with the help of a camera.The goal of this procedure is to recognize Facial Expression detected by the camera using Tensorflow , Keras and Convolutional Neural Network (CNN).The various type of emotions that are recognized are: angry,disgusted, fearful, happy, neutral,sad,surprised and disgusted. The songs are available in the folder. The songs will be played as per the mood detected by the camera. This research paper deals with CNN algorithm which is one of the Machine Learning techniques which helps in reducing the time to recognize emotions and provides accuracy.

Keywords: Face Recognition, Emotion Detection, CNN, Face Analysis, Tensor flow

1 Introduction

1.1 Motivation

Current research in the field of music has shown that it induces a clear emotional response to its listeners [1]. Musical preferences have been highly correlated with individual moods. The meter, timber, rhythm and pitch of music are managed in areas of the brain that deal with emotions and mood [2]. Undoubtedly, a user's affective response to a music fragment depends on a large set of external factors, such as gender, age [3], culture [4], preferences, emotion and context [5] (e.g. time of day or location).Identifying the emotional state and interpreting correctly using machine learning (ML) techniques has proven to be complicated due the high variability of the samples within each task [6].This leads to models with millions of parameters trained under thousands of samples [7].

1.2 Problem Statement

In the old days, a user had to manually browse the playlist and play the songs that matched their emotions. In today's modern world, various music players have been developed with features like fast forward, variable playback speed, local playback, streaming playback with multi-cast streams and it includes volume modulation, genre classification, etc. However, these features satisfy basic requirements but still the user has to select songs manually based on their emotions. Therefore, we came up with this application that will detect real time emotions of an individual and play music according to it.

1.3 Related Works

- Sang, Cuong and Ha, proposed a discriminative deep feature learning approach with dense convolutional networks (DenseNet) for facial emotion recognition.[8]
- Kumar, Kant and Sanyal proposed a better approach to predict human emotions (frame by frame) using deep convolutional neural network and how emotion intensity expressed by a face changes from low level to high level emotion.[9]
- Wang, Dong and Hu, proposed a deep cascade convolutional network that uses Fast Region-based Convolutional Neural Network (Fast R-CNN) for face detection.[10]
- Li Shang, Qin and Chen, proposed a technique to detect and recognize fish species under water using Fast R-CNN object detector.[11]

1.4 Hypothesis

- Research question : How to detect emotions and play music based on emotions discovered using CNN or Deep Learning?
- Null Hypothesis (H0): Emotions detected are not as per the expectation.
- Alternate Hypothesis (H1): Emotions are as per the expectation and music is being played accordingly.

2 Related Work

Sang, Cuong and Ha, proposed a discriminative deep feature learning approach with dense convolutional networks (DenseNet) for facial emotion recognition. An auxiliary loss is employed in this proposal, to regulate the training process of neural networks. Thus, reducing the intra-class variation of deep features and enhancing the discriminative power of the learned networks. The experimental results show that their proposed approach achieves superior 29 performance in comparison with other recent state-of-the-art methods implemented using the FERC-2013 dataset[8].

Kumar, Kant and Sanyal proposed a better approach to predict human emotions (frame by frame) using deep convolutional neural network and how emotion intensity expressed by a face changes from low level to high level emotion. In their proposed approach, a 9-layer convolutional neural network model is trained to classify 7 facially expressed emotions. The network model was trained using the FER-2013 database [12]. For emotion analysis of micro expressions, classified emotions are measured in percentages as an approach to assess the overall network performance[9].

Wang, Dong and Hu, proposed a deep cascade convolutional network that uses Fast Region-based Convolutional Neural Network (Fast R-CNN) for face detection. With this proposed method, the highest recall rate of true positive against false positive have been achieved on the challenging FDDB benchmark, thus, outperforming the current state-of-the-art methods[30]. The proposed method first utilizes the cascade CNN structure to reject background regions quickly and then uses Fast R-CNN object detector. The Fast R-CNN produces a bounding box and classified each object being the human face[10].

Li Shang, Qin and Chen, proposed a technique to detect and recognize fish species under water using Fast R-CNN object detector. The Fast R-CNN improves the mean average precision (mAP) by 11.2% when compared to deformable parts model (DPM), achieving a mAP of 81.4%, and performing faster than the R-CNN object detector by 80 times. The proposed model is promising for automatic fish identification systems to help marine biologists estimate fish existence and quantity, which would effectively help understand oceanic geographical and biological environments[11].

3 Method

3.1 Problem

The features in the modern world music player satisfy basic requirements but still the user has to select songs manually based on their emotions. Therefore, we have developed an application that will detect real time emotions of an individual and play music according to it.

3.2 Hypothesis

- Research question : How to detect emotions and play music based on emotions discovered using CNN or Deep Learning?
- Null Hypothesis (H0): Emotions detected are not as per the expectation.
- Alternate Hypothesis (H1): Emotions are as per the expectation and music is being played accordingly.

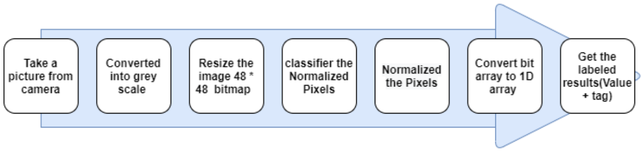
Take a picture from camera
Converted into grey scale
Resize the image 48 * 48 bitmap
classifier the Normalized Pixels
Normalized the Pixels
Convert bit array to 1D array
Get the labeled results(Value + tag)

Figure 1. Image Processing

3.3 Statistical Test

- Dependent Variable:1(Emotion)(angry,disgusted, fearful,happy,neutral,sad,surprised and disgusted)
- Kind of Dependent Variable:Continuous
- Independent Variable :1 (Pixel)
- Kind of Independent Variable: Discrete

3.4 Participants

Table 1. Task Distribution

Sr.no.	Life cycle	Team Members
1	Requirement Analysis	Anand/Karishma
2	Design Implementation	Karishma/Sohail
3	Coding	Sohail/Anand
4	Testing	Munshi/Opu
5	Documentation	The Team

3.5 Material/procedure

3.5.1 Image Processing. In Image processing, the picture is usually captured by the camera and converted into grayscale so that the colour remains the same and only the main area of the face is captured and then the image is resized into 48*48-bit map. Classifier is used to normalize the pixels and then the final result is obtained as shown in Figure[1]

3.5.2 Emotion detection. In order to obtain face detection, at first the RGB image is converted into a binary image which is obtained by calculating the average RGB value for each pixel. The scanning is done only for eyebrows, lips, and nose. The pixels size is 48*48 bitmap image.

3.5.3 Expression recognition. Expression recognition is the procedure which is used to extract eventful attributes. Template matching is done by making use of roll and intercourse coefficient for the higher and faultless matching.The eyebrow, eye, and mouth are stage transcript and output obtained is selected triangle.

3.5.4 Extracting the facial characteristic points. The extraction is done by using the value of the top left angle pixel from the cropped rectangle. It is specified by using height and width of template and enumerates all 30 fcp. These values basically detect the facial enlivenment attribute such as the opening of the eye(or), the height of eyebrow(he), the opening of the mouth(om), the width of the eye(we), the width of mouth(wm).

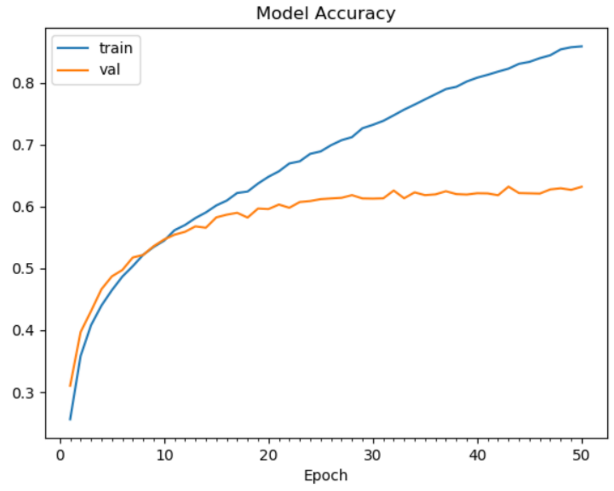

Figure 2. Model Accuracy

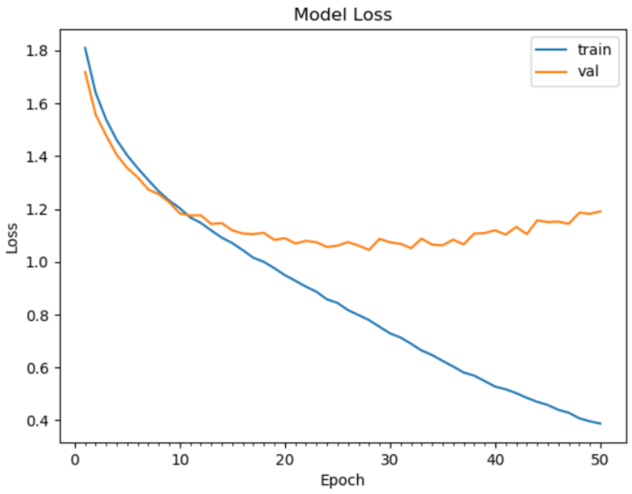

Figure 3. Model Loss

3.5.5 Open CV. Open CV (Open Source Computer Vision) library in Python is mainly used for real-time image processing. It basically supplies a complex number of functions for face recognition and detection. It can be used to train our own classifier for any object like mobile, pen, etc.

3.5.6 CNN Algorithm. Convolutional Neural Networks (CNN) simulates the human brain at the time of analysing visuals. However, optimizing a network for efficient computation is necessary due to computational requirements and complexity of a CNN. CNN constructs a computational model which classifies emotion into 7 moods, namely, angry, disgusted, fearful, happy, neutral, sad, surprised and disgusted.

4 Result

4.1 Descriptive statistics

Result of mood based music system can be seen through accuracy in mood detection and in songs recommendation. It is hard to find accurate human emotion only through only one parameter. But with facial expression it can be detected up to some extents. Result of mood detection through facial expression under proper lightning condition can be seen in figure 3. Model we have used has achieved 63.2% of accuracy. As it is totally computer based system it understands emotions in the way we trained it. System takes that mood and generates music playlist for that mood accurately. System is able to play most of the songs from recommended playlist.

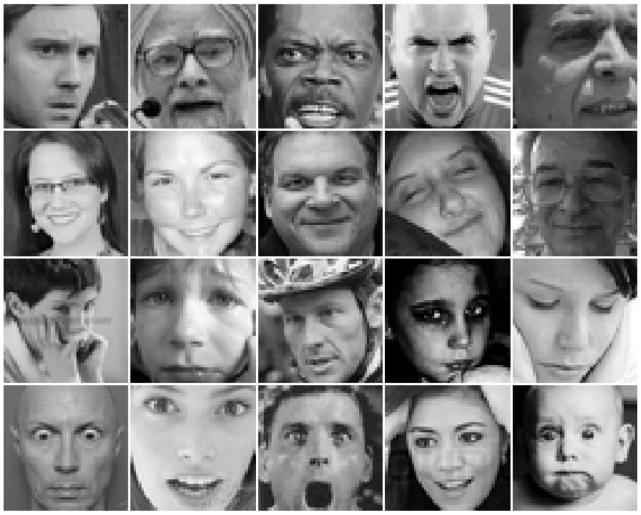

Figure 4. Samples from the FER-2013 dataset

4.2 Inferential statistics

In Figure 5 we provide the confusion matrix results of our emotion classification mini- Exception model.The white areas in Figure 5 correspond to the pixel values that activate a selected neuron in our last convolution layer. The selected neuron was always selected in accordance to the highest activation. We can observe that the CNN learned to get activated by We can observe several common misclassifications such as predicting “sad” instead of “fear” and predicting “angry” instead “disgust”. A comparison of the learned features between several emotions and both of our proposed models can be observed in Figure 4 considering features such as the frown, the teeth, the eyebrows and the widening of one’s eyes, and that each feature remains constant within the same class. These results reassure that the CNN learned to interpret understand- able human-like features that provide generalizable elements. All sub-figures contain the same images in the same order. Every row starting from the top corresponds respectively to the emotions “angry”, “happy”, “sad”, “surprise”

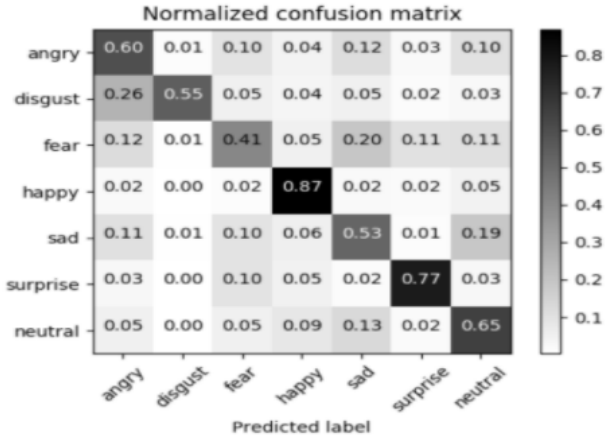

Figure 5. Confusion Matrix

These interpretable results have helped us understand several common misclassifications such as persons with glasses being classified as “angry”. This happens since the label “angry” is highly activated when it believes a person is frowning and frowning features get confused with darker glass frames. Moreover, we can also observe that the features learned in our mini-Xception model are more interpretable than the ones learned from our sequential fully-CNN. Consequently, the use of more parameters in our naive implementations leads to less robust features.

5 Discussion

5.1 Brief summary of findings

In the existing system, no software is used to detect human facial expression. The proposed system has been designed to use the Convolutional Neural Networks (CNN) in order to extract the facial emotion and it uses the Tensor-Flow Machine Learning library.

5.2 Discussion/Interpretation of findings

The existing system uses wearable physiological sensors. On the other hand, live emotion is detected using a webcam in the proposed system which detects, captures and classifies expressions into their respective types. Accordingly, music is being played based on the captured emotion. The Convolutional Neural Network (CNN) model is used by the system for image classification. The model trains itself as per the results and helps in improving the efficiency and accuracy of the system.

5.3 Relation to previous/related work

There is no direct relation between the existing and the proposed system. However, in both the systems we are recognizing human facial expressions but the methodology to extract it differs. Here webcam is used to capture the expressions instead of wearable physiological sensors.

5.4 Limitation of the study

We have found the following limitations in the existing system and therefore proposed a new system to overcome it.

- There is no software to detect human facial expression.
- The current system consumes more time.
- There is no guarantee that the detection result is always correct.

6 Conclusion

In this paper, we proposed an algorithm for web cambased emotion recognition with no manual design of features using a CNN. And various types of advantages like High speed Extraction, and features selection is efficient and easy to gets its result and we work on feature scaling instead of image scaling. The code can be find in the github respository - https://github.com/sdgamer007/RealtimeEmotionRecoginitionMusicPlayer.git

Future Scope for Implementation:

- Facial recognition can be used for authentication purpose.
- Android Development.
- Can detect sleepy mood while driving.
- Can be used to determine mood of physically challenged mentally challenged people.

7 Acknowledgments

The following references are used [1], [2],[3],[4], [5],[6],[7], [8],[9],[10], [11],[12]

References

[1] E. Glenn Schellenberg Swathi Swaminathan. 2015. Current emotion research in music psychology. *Emotion Review*, 7, 2, 189–197.

[2] How music changes your mood. 2017. *Examined Existence[Online]:* http://examinedexistence.com/how-music-changesyour-mood/.

[3] Kyogu Lee and Minsu Cho. [n. d.] Mood classification from musical audio using user group-dependentmodels.

[4] Tillman Weyde Daniel Wolff and Andrew MacFarlane. [n. d.] Culture-aware music recommendation.

[5] Jun-Dong Cho Mirim Lee. 2014. Logmusic: context-based social music recommendation service on mobile device. *Ubicomp '14 Adjunct.* Seattle, WA, USA.

[6] Ian Goodfellow et al. Challenges in Representation Learning. 2013. A report onthreemachinelearningcontests.

[7] Antoine Bordes Xavier Glorot and Yoshua Bengio. 2011. Deep sparse rectifier neural networks. *In Proceedings of the FourteenthInternationalConference on Artificial Intelligence and Statistics,* 315–323.

[8] B. C. Le Tran V. S. Dinh and T. H. Pham. 2018. Discriminative deep feature learning for facial emotion recognition in. *1st International Conference on Multimedia Analysis and Pattern Recognition (MAPR).* Ho Chi Minh City.

[9] K. K. Ravi K. Rajesh and S. Goutam. 2017. Acial emotion analysis using deep convolutional neural network. *International Conference on Signal Processing and Communication (ICSPC),* Coimbatore.

[10] J. Lu I. Unwala X. Yang L. Nwosu H. Wang and T.Zhang. [n. d.] Deep convolutional neural network for facial expression recognition using facial parts,

[11] Q. Hongwei L. Xiu S. Min and C. Liansheng. 2015. Fast accurate fish detection and recognition of underwater images with fast r-cnn. in OCEANS 2015 - MTS/IEEE Washington,

[12] 2013. *Challenges in Representation Learning: Facial Expression Recognition Challenge," Kaggle Inc, 2013. [Online]. Available:* ://www.kaggle.com/c/challenges-in-representation-learning- facialexpression-recognition-challenge.