Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.7766125

EMOTION BASED MUSIC SUGGESTION SYSTEM USING DEEP LEARNING TECHNIQUES

P Anil Kumar[1],Md Thanveer Khan[2], N Teja[3], K Lokeswari[4]
Raghu Engineering College
Dakamarri (V), Bheemunipatnam (M), Visakhapatnam – 531162
anilkumarprathipati@gmail.com, 19981A0598@raghuenggcollege.in
19981A05A6@raghuenggcollege.in, 19981A0580@raghuenggcollege.in

Abstract
Music plays a vital role in human's mood modulation in day to day life. Music enhances the emotional state of the human by stress reduction, mood regulation and removes distraction. Music regulates the mood by inducing various emotion states such as happiness and relaxation. This further reduces stress as the mood is enhanced by emotional states. The rhythm and tempo of music played can have calming effect on the human stress and reduce distraction to focus on the target work. From the result of many research works that is proven that there are significant relation between human emotion and music. This paper advances a emotional music player that plays music based on human emotion expression. This is based on reinforcement learning for mapping the emotion to the music. Here for the given target emotion, initially the state of the emotion is being is being determined base on the image that is being captured instead of taking emotion from the user in the previous work in the literature. In the advanced system the music player contains two GUI windows which are responsible for all the outputs to view. The initial window is used to capture the image or terminate the process. The final window plays the song and displays the emotion that is being detected. The most effective of these methods in CNN model is 87.02%

Keywords: Mood Modulation, Emotion States, Human Emotion Expression, Reinforcement Learning, CNN

Introduction

The ability of music to induce emotions and control our mood. From normal to pop song, music has been a basic part of human day to day life. The olden times had proven that music had been using as a form of entertainment activity, Communication and treatment. In recent years there is an increase interest to know the connection between music and emotions which paves path for the field of music psychology. Music is an influential tool that can help in managing the emotion ranges, from happiness to anger. The impact of music in case of emotional impact is due to its ability to replicate the brain in several ways. According to the studies it has proven that the music an activate the reward centres of brain which cause of release of dopamine, a chemical substance which is released at the end of nerve fibre responsible for nerve impulse and feeling of pleasure. This make us to feel joy and happiness where as the sad songs are responsible for activation of amygdala part of the brain which cause release of cortisol hormone which causes stress instincts. The impact of music emotionally influenced by personal factors.

For example, particular song may evoke particular emotion in different people depending on one's personal factors.

The impact of emotion in music is not restricted to just one or two emotions. Music can also induce feelings of reminiscence, trepidation and fear. This emotion is influenced by various factors such as cadence, pace, piece of music, and accord. For example initially the pop music is to induce feelings of enthusiasm and power, while deliberate and harmonious music is more likely to induce feelings of peacefulness and relaxation.

The relation between music and emotion is also obvious in the way of music used in a variety of settings. For example, music is frequently used in movies, web series and shows to improve emotional contact of a scene. The correct music can make a scene more theatrical, most important to a powerful impact on the end person.

Music is used in treatment of various psychological health conditions such as depression, anxiety and post traumatic stress disorder. This also helps patients to upscale expressions and emotions by means of music. The music has been shown to have positive impact on mood and advance the mental health.

The connection in-between emotion and music is a complex and interesting area of study. Music has an ability to induce range of emotions and persuade mood. The impact of emotion in music is influenced by various factors such as cadence, pace, piece of music, and accord. The association between music and emotion is obvious based on various factors. Music is a powerful tool that is used to improve emotional and healing senses in human. The study of music psychology continues on the mesmerizing association between music and emotions, and the importance of music in human culture and society.

In order to suggest music for prevailing a target emotion state, here we exploit reinforcement learning techniques which are more effective in case of music suggestion systems. The program is to train an smart agent enough proficient of suggesting songs based on the altering moods from the user from existing emotional state (which is been detected after capturing image).
The implementation of the emotion detection is the initial stage of the music suggestion system. This is based on Convolution Neural Network using Keras framework. The dataset that we had used here is Facial Expression Recognition Challenge (FER2013) which is having 35887 grayscale images if 7 emotions (angry, disgust, fear, happy, neutral, sad, surprise). Here there is ImageDataGenerator function from keras to generate training and validation data with specified batch size and target image size. Each of the four convolutional layers in this implementation is followed by a batch normalization, activation, max pooling, and dropout layer in the CNN architecture. Two fully connected layers and a softmax output layer with seven nodes (one for each emotion class) follow after this.

The model is then compiled with the Adam optimizer and categorical cross-entropy loss function. Then for monitoring validation accuracy, early stopping and reducing learning rate

for validation loss three callbacks are defined. Then model then further trained for 40 epochs by means of fit_generator function.

In the final stage to show the integration between the Emotion detected and the song playing more user-friendly, tkinter library is used to create GUI windows. Here to make the more user-friendly to take the emotion by means of video feed the first GUI window having capture and exit button to capture the image from the video feed. After capturing face, emotion is classified fro the pertained deep learning model and finally displays the result, detected emotion and the name of the song that is being played after its being mapped with the emotion detected.

To get that outcome several python libraries are been imported such as PIL, tkinter, OS, numpy, Playsound, tensorflow and keras. Here we are having two major functions named capture_image and stop_song which drives whole process. The capture_image() function is defined to capture an image from the video feed, detect the face in the image, classify the emotion from the face, and play a song based on the predicted emotion. The function then displays the captured image, predicted emotion, and the name of the song being played on a new GUI window. The stop_song function is used to terminate the song.

Here the root tkinter window is created so that to capture the live feed in order to extract the face from the feed's frame. A capture button, a stop button is labeled in-order capture and terminates the on going process. Here the capture is mapped with capture_image method and the stop button is mapped with stop_song method. The program enter the main event loop by means of root.mainloop() method. Overall the proposed system provides user friendly UI for an emotion based music suggestion player that classifies the emotion from image that is taken from the live feedback and maps the music accordingly based on the emotion that is detected.

Literature Work

Roberto De Prisco et al. The link between emotion and music has therefore been the subject of a number of studies on emotional human-computer interaction and music information retrieval. The purpose of the recommended RL technique is to learn how to select a playlist of musical selections that best "fits" the desired emotional state given the target emotional state and operating under the assumption that a recent playlist of musical selections has had a significant impact on the emotional state.. The results showed that user pleasure, system sensitivity, and suggestion correctness received good scores (up to 4.30, 4.45, and 4.75 on a 5-point, respectively) [1].

H. Immanuel James et al. An individual's mood may often be determined by looking at their face. With a camera, the needed information is immediately collected from a person's face. This facilitates the creation of an appropriate playlist based on a person's emotional characteristics and avoids the need to manually segregate or combine songs into several lists. In order to create an emotion-based music player, the author developed a system that focuses on recognizing human emotions. The methodologies utilized by existing music players to detect emotions are described in the article [2].

Markus Schedl et al. Deep learning (DL) is more and more used in music recommendation systems, just like it is in many other study fields (MRS). For collecting latent characteristics of musical pieces from audio signals or metadata, deep neural networks are utilized in this field. It provides a summary of the phase of the art in music recommendation using deep learning and discusses about dimensions of neural network types [3].

Manjushree K et al. Due to the difficulty of choosing a song from among the thousands that make up a playlist, users are forced to choose songs manually from the playlist in accordance with their mood. As a result, an emotion-based music player has been introduced. The camera takes a user's picture, identifies their expression, gauges their mood, and then plays music in accordance with the completed playlist. The feeling Based Music Player assists the user by identifying their feeling, playing the music, and also recommending similar tracks that will reassure or resonate with their mood [4].

Shravanthi K et al. To create a music recommendation system that uses user data to suggest songs to the user. The most well-liked tunes among all tracks are suggested by this technique. Additionally, it provides tailored recommendations based on previously listened to songs and genres by the user. The technique uses convolution neural networks to categorize songs into various genres. Additionally, the system makes use of a collaborative filtering algorithm to compare songs from user history to all other songs that are comparable. It also consist of Log-in page as Front end [5].

Methodology

Initially the model is trained using FER 2013 data set and the best effiecncey is been stored in model.h5 file. Then the input capturing by using the webcam and then the featured are been extracted from the initial frame. By using the trained model from the dataset the emotional label of the frame is been calculated. By using the label the song is been played using playsound module

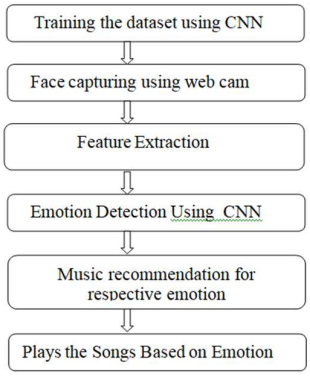

Fig 1| Data Flow Diagram

Dataset

We hare use FER2013 data set which is having 35887 images which are greyscale of seven distinct classes namely Angry, Disgust, Fear, Happy, Neutral, Sad, Surprise. Which is used to depict the emotion from the frames that are been captured from the live feed. The trained model can detect the emotion up to 89.01% accuracy.

Making Training and Validation Data

```
batch_size  = 128

datagen_train  = ImageDataGenerator()
datagen_val = ImageDataGenerator()

train_set = datagen_train.flow_from_directory(folder_path+"train",
                                              target_size = (picture_size,picture_size),
                                              color_mode = "grayscale",
                                              batch_size=batch_size,
                                              class_mode='categorical',
                                              shuffle=True)


test_set = datagen_val.flow_from_directory(folder_path+"validation",
                                              target_size = (picture_size,picture_size),
                                              color_mode = "grayscale",
                                              batch_size=batch_size,
                                              class_mode='categorical',
                                              shuffle=False)
```

```
Found 28821 images belonging to 7 classes.
Found 7066 images belonging to 7 classes.
```

Fig 2 | Training and Validation

Model building

Initially to achieve correct emotion we need to increase the training and testing data given to it. In the proposed system we had used FER 2013 data set. In that there are 35887 grayscale images of seven different classes namely Angry, Disgust, Fear, Happy, Neutral, Sad, and Surprise. Then it is further divided into two for training and testing. The training set consist of 28821 grayscale images of 7 different classes and the test set consist of 7066 grayscale images of 7 different classes. Here to enhance the image in an efficient way they are further focused into resolution with 48x84 pixel density. The ImageDataGenerator() from keras is used to batch the data with real-time data augmentation, here to maintain the efficiency the batch size is fixed to 128. The major task of that function is to allow your model to receive different images at individual epochs.

Model Building

```
from keras.optimizers import RMSprop,SGD,Adam
from keras.callbacks import ModelCheckpoint, EarlyStopping, ReduceLROnPlateau

checkpoint = ModelCheckpoint("./model.h5", monitor='val_acc', verbose=1, save_best_only=True, mode='max')
epochs = 40
early_stopping = EarlyStopping(monitor='val_loss',
                          min_delta=0,
                          patience=3,
                          verbose=1,
                          restore_best_weights=True
                          )

reduce_learningrate = ReduceLROnPlateau(monitor='val_loss',
                              factor=0.2,
                              patience=3,
                              verbose=1,
                              min_delta=0.0001)
```

Fig 3 | Model Building

In order to build the model the initial structure consists of four CNN layers and two fully connected layer. To optimize the code here we had used three major optimizers namely Adam, SGD and RMSprop. This helps in making a model more sequential. Adam stands for Adaptive

Motion Estimation, which is used in case of reducing noise in case of images, this optimizer is generally used in case of dealing with huge datasets with multifarious architecture. The SDG stochastic gradient Descent optimizer is used in case of denting the gradient of loss function while training. The RMSprop that is Root Mean Square propagation optimizer is used to familiarize the rate of learning with magnitude of gradient. Then further the trained model is stored in model.h5 file.

Emotion Extraction

After the data that is been trained using FER 2013 data set of 28821 grayscale images. The trained model in model.h5 file is loaded along with the attributes which are been loaded using haarcascade_frontalface_default.xml file. The emotion labels are here listed under a list naming emotion_labels.

Song mapping and Song Playing

The songs are mapped using a dictionary named songs_dir which contains key as emotion labels and values as the location of the songs according to the predicted emotion. There are two GUI windows named emotion music player and result window. Initially the emotionExteactor extracts the image frame from the live image feed and analyzes the image to get the emotion label as an outcome by means of trained model. Then the emotion label is used as a key to extract the path of the desired emotion and then by using the playbox module in the second window the captured image, detected emotion label and mapped playing song is been displayed

Algorithm

Haar cascade

In day to day life the face detection became as most common functionality. It is usually used to secure device by means of authentication and make the device more secure by verifying user's identification The algorithm which can detect the objects ignoring the resolution of the image is Haar cascade. It's usually works as a classifier. In this model the training is been trained on lots of positive and negative images.

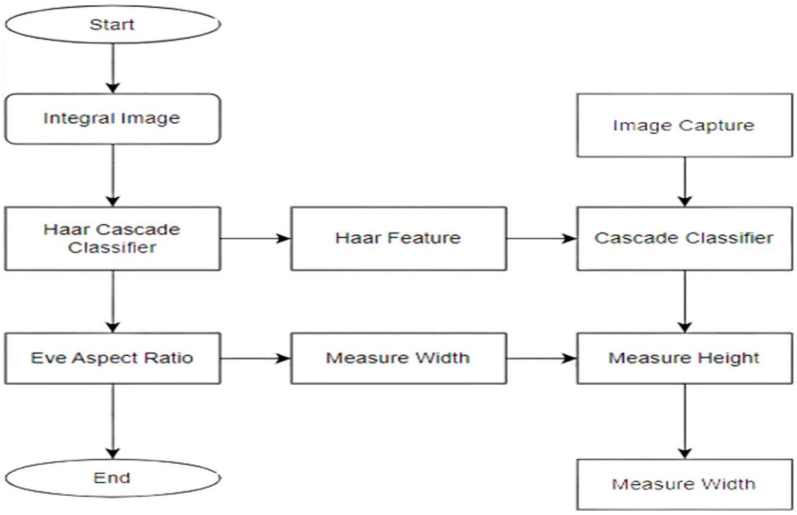

Fig 4 | Haar cascade

CNN

It is actually the extended version of Artificial Neural Networks. CNN mainly deals with layers such as Input, convolution, pooling and fully connected layers. The convolution layer acts as a filter in order to mine (find out) features ,pooling layers reduce the processing by choosing the image, then the final prediction is further analyzed by means of a fully connected layer.

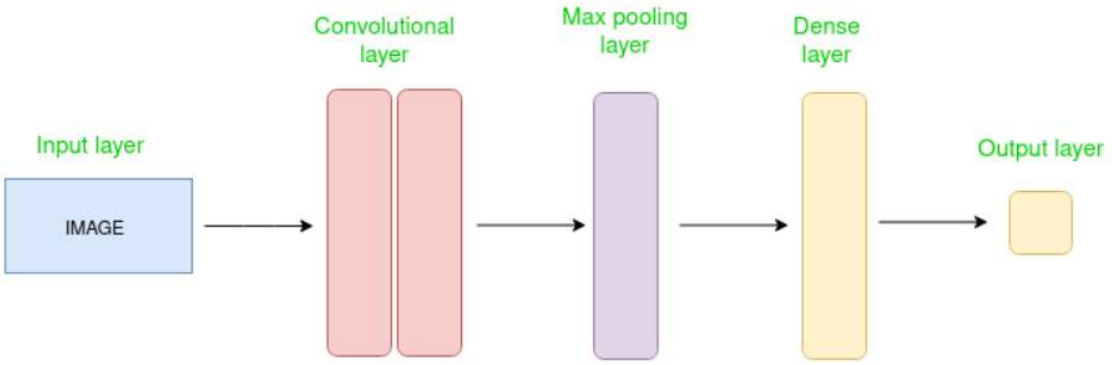

Fig 5 | CNN

Results

System Applications

The tkinter based front end is generally shows how the front end of the proposed system is been connected with the backend. The front end is generally consisting of a Tkinter window which is used in capturing the image from the live feed back using a capture button in the GUI window. The back end of the system is a trained CNN model which is used to extract the emotion from the frame of live feed back and then classifies the emotion based on the seven different classes. After the classification the final result is been displayed in the final tkinter window along with play button and details like detected emotion label, song label which is been mapped to extracted emotion, and the image frame which is extracted from the live feed back. Whenever the user face is detected the emotion is detected by the help of help of pertained model and map the music randomly based on detected emotion and then song is played in background.

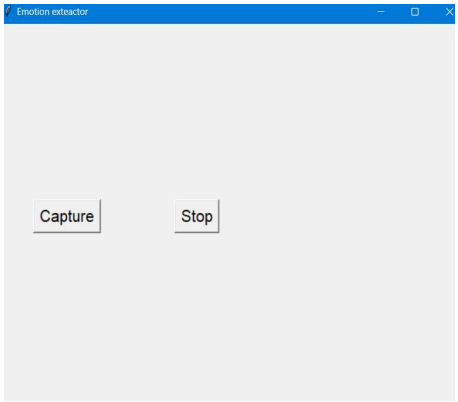

Fig 6 | Emotion Extractor

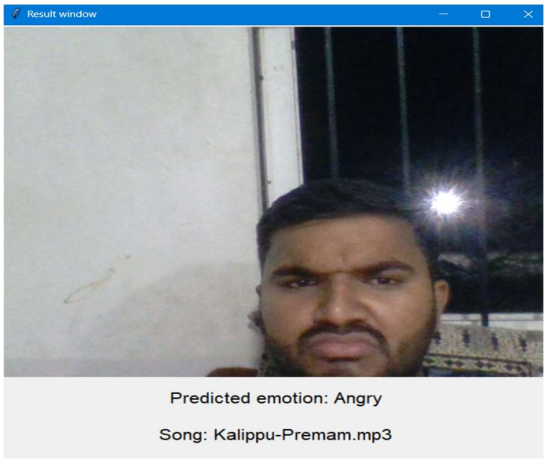

Fig 7 | Result Window

Performance Evolution

To fulfil the Emotion based Music Recommendation System the proposed system is written using Python programming language. In-order to achieve CNN approach for classification of facial emotion by means of live feed captured image tensor flow library is imported. The segment of image is extracted from the live feed and further converted into specific format to make the process smooth. In order to evaluate the proposed system we had selected 35887 image segments of seven different classes. Further these are again divided into 28821 train and 7066 test images. After training the model with 28821 image segments, we had used 7066 segments to evaluate the performance of the proposed system.

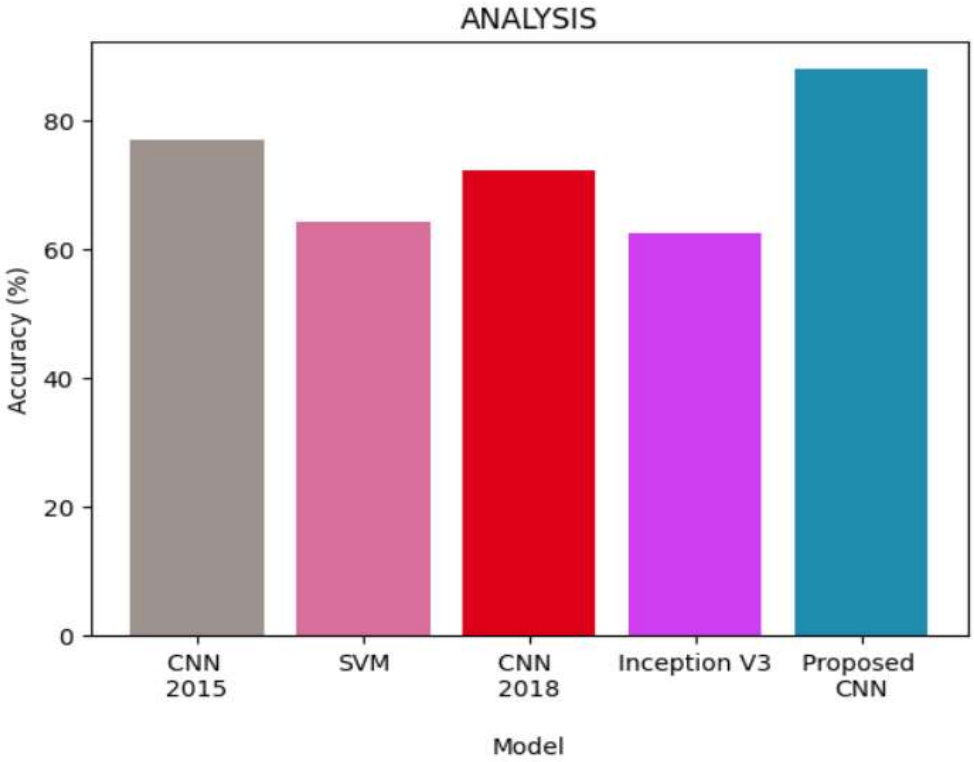

Fig 8 | Comparison of Previous Models

Analysis

We have taken up various classification models like the SVM, CNN, Inception V3and their accuracies are 77.02%, 64.2%, 72.3%, 62.5% respectively. So, we used the CNN model as it predicts the Facial Emotion more accurately when compared to other models.

Authors	Classification Model	Dataset	Accuracy(%)
S. S. Farfade [6]	CNN	FER 2013	77.02
Jeong [7]	SVM	FER 2013	64.2
H Yu [8]	CNN	FER 2013	72.3
P. Sharma [9]	InceptionV3	FER 2014	62.5

Table 1 | Previous model Analysis

Conclusion

We had proposed a emotion based music suggestion system based on CNN and random

approach of playing music. Here CNN is used to classify the emotion by means of FER2013 data set consisting 35887 images in which there are 28821 images for training and 7066 images for testing. As a part of future enhancement we can add connection to online music players such as you tube music and here we can increase the efficiency of the model by adding more number of images to the training and testing models. The efficiency of the model can also be increased by using latest data sets available.

Acknowledgment

A special thanks to my Guide Mr. P Anil Kumar sir, who supported us in completion of the project and paper by providing valuable time, encouragement and all the needy data and resources. We also thank our HOD S Srinadh Raju, for inspiring us. Lastly we would also like to thank all team members and who has supported constantly.

References

1. Roberto De Prisco , Alfonso Guarino , Delfina Malandrino and Rocco Zaccagnino . *" Induced Emotion-Based Music Recommendation through Reinforcement Learning".* Appl. Sci. 2022, 12, 11209. https://doi.org/10.3390/app122111209
2. H. Immanuel James1, J. James Anto Arnold, J. Maria Masilla Ruban, M. Tamilarasan, R. Saranya. *"EMOTION BASED MUSIC RECOMMENDATION SYSTEM".* Volume: 06 Issue: 03 | Mar 2019. e-ISSN: 2395-0056. p-ISSN: 2395-0072.
3. Markus Schedl. "Deep Learning in *Music Recommendation Systems*". Published 29 August 2019. Doi : 10.3389/fams.2019.00044
4. Manjushree K, Shree Lakshmi Shetty C, Spoorthi M, Deeksha R, Sahana C Shekar. *"Emotion Based Music Player".* Volume: 09 Issue: 08 | Aug 2022. e-ISSN: 2395-0056. p-ISSN: 2395-0072
5. Shravanthi K, Sneha Rao S, Swati Joshi, Vandana R. *"A Personalized Music Recommendation System".* Volume: 06 Issue: 06 | June 2019. e-ISSN: 2395-0056. p-ISSN: 2395-0072
6. S. S. Farfade, M. Saberian, and L. J. Li, "*Deep Convolutional Neural Networks for Facial Expression Recognition*," IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2015.
7. Jeong, A. (2018). "*Emotion Recognition using Facial Landmarks, Python, DLib and OpenCV*". Proceedings of the 2018 IEEE International Conference on Consumer Electronics (ICCE), 1-5. https://doi.org/10.1109/ICCE.2018.8326204
8. Yu, H., Wu, Y., & Zhou, W. (2018) "*Convolutional Neural Networks for Facial Expression Recognition Using Multiple Databases*". Frontiers in Psychology, 9, 1445. https://doi.org/10.3389/fpsyg.2018.01445
9. Sharma, P., & Rai, P. (2019). *"A Comparison of Deep Learning Techniques for Emotion Recognition on Static Images".* 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), 2725-2730. https://doi.org/10.1109/SMC.2019.8914258
10. Hanjalic, A.; Xu, L.Q. Affective video content representation and modeling. IEEE Trans. Multimed. 2005, 7, 143–154.

11. Lu, L.; Liu, D.; Zhang, H.J. Automatic mood detection and tracking of music audio signals. IEEE Trans. Audio Speech Lang. Process. 2005, 14, 5–18. [CrossRef]
12. Anagha S.Dhavalikar and Dr. R. K. Kulkarni, "Face Detection and Facial Expression Recognition System" 2014 Interntional Conference on Electronics and Communication System (ICECS -2014).
13. Yong-Hwan Lee , Woori Han and Youngseop Kim, "Emotional Recognition from Facial Expression Analysis using Bezier Curve Fitting" 2013 16th International Conference on Network-Based Information Systems.
14. Arto Lehtiniemi and Jukka Holm, "Using Animated Mood Pictures in Music Recommendation", 2012 16th International Conference on Information Visualisation.
15. Dorfer M, Henkel F, Widmer G. Learning to listen, read, and follow: score following as a reinforcement learning game. In: Proceedings of the 19th International Society for Music Information Retrieval Conference (ISMIR 2018), Paris (2018). p. 784–91.