

Facial Emotion Recognition and Music Recommendation based on Detected Emotion

-Kush Ramesh Jain (2017A3PS0425G)



In Partial Fulfillment of the Course
Study Oriented Project (CS F266)



Introduction

- Inspired by Spotify
- Music is an integral part of everyone's daily life
- Developed a deep learning model to identify the emotion from an image of a person's face through his/her expressions
- Classified songs into emotions based on lyrics

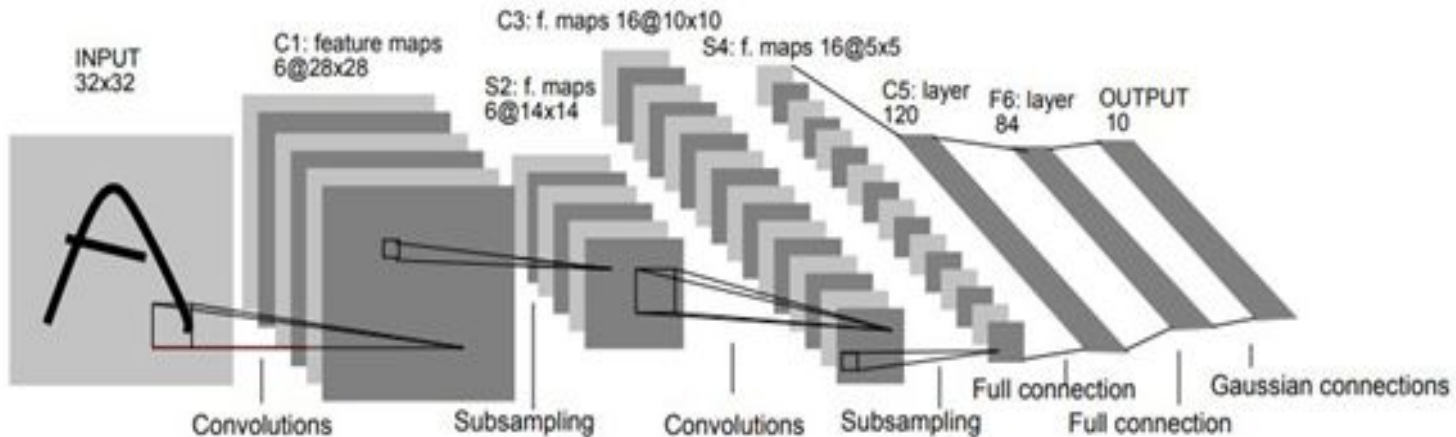


Datasets Used

- The **Karolinska Directed Emotional Faces (KDEF)** has been used for facial emotion recognition
- It is a set of totally 4900 pictures of human facial expressions and has 70 subjects (35 male and 35 female), each expressing seven different emotions
- Emotions expressed: afraid, angry, disgusted, happy, neutral, sad and surprised
- For the music emotion recognition part, the **PMEmo dataset** containing emotion annotations of 794 songs as well as the simultaneous electrodermal activity (EDA) signals

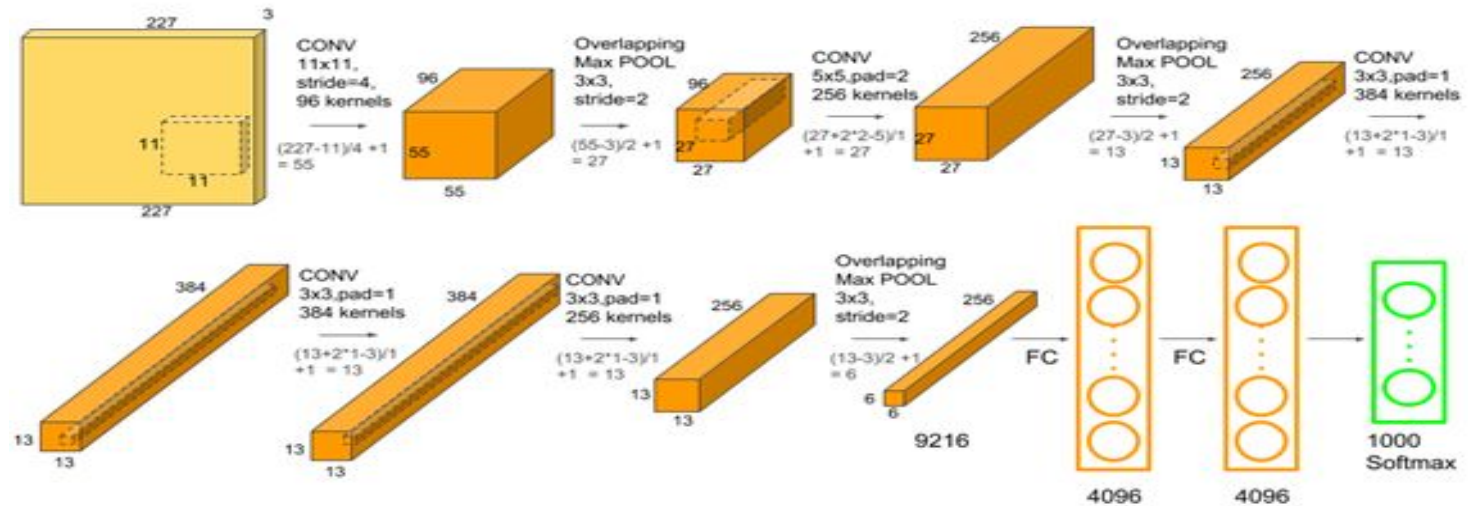
Models Used for Facial Emotion Recognition

1. LeNet - 5 - Validation Accuracy of 14.588 %, not very good



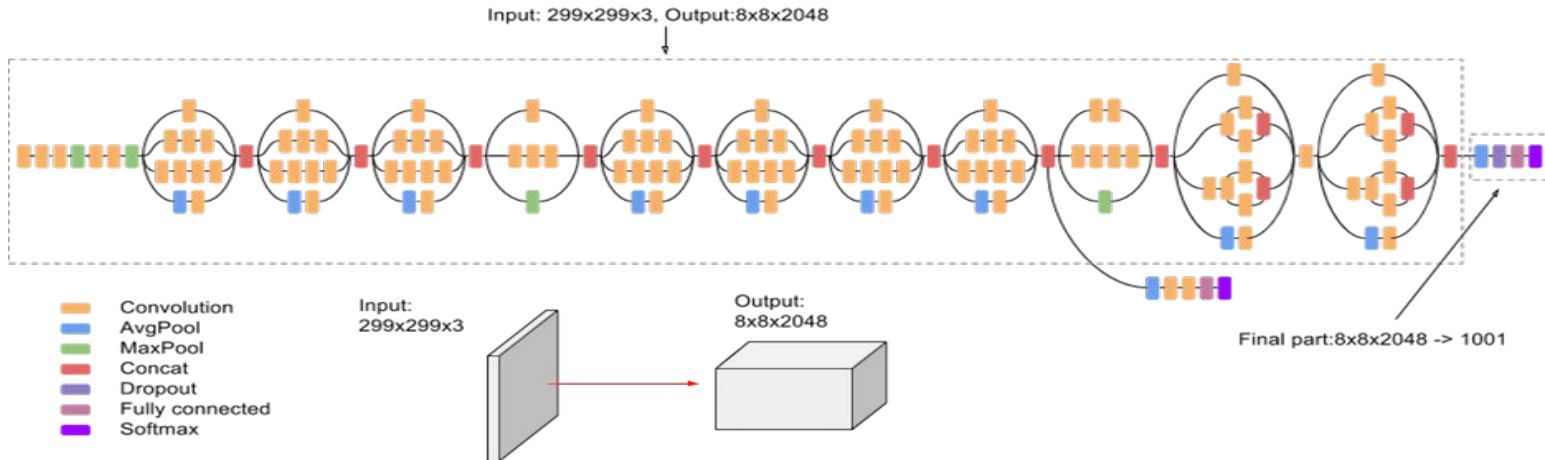
Models Used for Facial Emotion Recognition

2. AlexNet - Validation Accuracy of 58.095 %, better than LeNet - 5



Models Used for Facial Emotion Recognition

3. Inception V3: Validation Accuracy of 94.375 %, Very Good



Labelling the Songs Dataset

- Valence and Arousal values were used, alongwith the following model for the emotion labels of songs
- Since approximate values were used to classify the songs into labels, they are prone to error

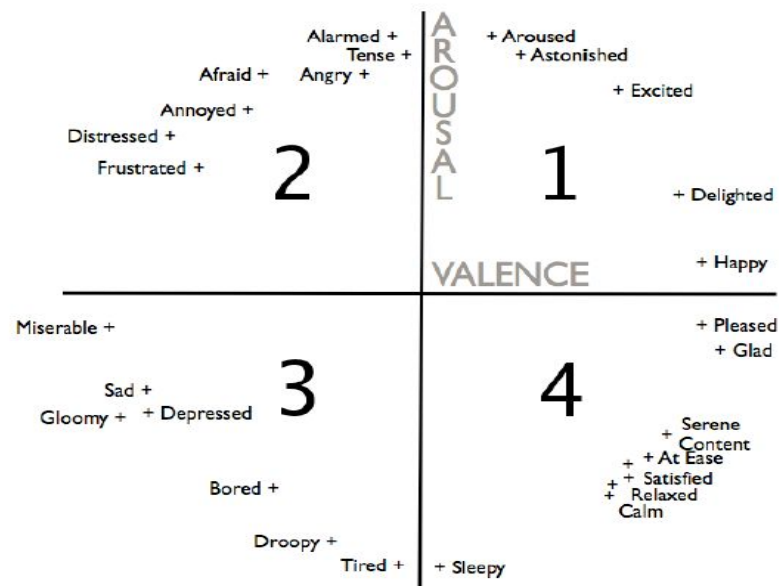


Figure 2. “Circumplex model of affect”, adapted from Russel (1980)

Processing on the Songs Dataset

- Selected songs whose lyrics are available in the PMEmo dataset and those which were selected while emotions were labelled by the Valence-Arousal values graph
- Non-alphabetic characters and symbols were removed from the lyrics using the Regular Expression Operations (re) library
- Tokenized lyrics into words by splitting them at the spaces
- TfidfVectorizer was used to vectorize the words with the maximum number of features set to 500

Models used for Song Emotion Recognition

Model Used	Validation Accuracy (%)
Decision Tree Classifier	36.97
Multi-Layer Perceptron Classifier	32.77
Random Forest Classifier	36.97

Another Approach

- Using Keras text-processing functions (text_to_word_sequence) and GloVe embeddings of 100 dimensions
- Following model was used
- It gave a validation accuracy of 38.202 %

Model: "sequential_1"

Layer (type)	Output Shape	Param #
embedding_1 (Embedding)	(None, 200, 100)	5500
dropout_1 (Dropout)	(None, 200, 100)	0
conv1d_1 (Conv1D)	(None, 196, 64)	32064
max_pooling1d_1 (MaxPooling1D)	(None, 98, 64)	0
dropout_2 (Dropout)	(None, 98, 64)	0
conv1d_2 (Conv1D)	(None, 94, 128)	41088
global_max_pooling1d_1 (GlobalMaxPooling1D)	(None, 128)	0
dense_1 (Dense)	(None, 6)	774
Total params: 79,426		
Trainable params: 73,926		
Non-trainable params: 5,500		

Future Work

- Collect a songs dataset from scratch which has popular songs from recent years and collect better labels for the same
- Test other Natural Language Processing models and approaches to increase validation accuracy
- Try to combine or use a different dataset for Facial Emotion Recognition which is more relevant to real world data
- Develop an application to combine both the models



Thank You