

CS 682 Project Proposal

Group Members: Jacinth David, 1 member

Not shared with another class

Problem:

Generate Music Videos as accompaniment to Music. The project will be interesting because:

- Artistic and Entertainment Value
- Musicians currently have no way to create videos as accompaniment quickly, though creating one really increases the promotional possibilities

An accompaniment video should fit the music in some way. Well created videos match their songs in many respects but I'm going to focus on rhythm or timing.

Proposed Method:

I will attack the problem in 3 phases:

1. Events detection from a collection of Videos
2. Rhythm detection in a song
3. Matching the two in some way to create the music video

Phase 1 can be achieved using the technique shown in the highlight detection paper. I will train the network on the baidu dataset. I plan to experiment with the score function in the paper to obtain highlights suited for use in music. If time permits, I would like to include videos more suited for music, though I will have to label them myself.

Phase 2 can be achieved using the network shown in the Piano transcription paper. The network presented in the paper is trained for transcription which is a harder task than rhythm detection. I plan to extract the rhythm in the MAPS midi labels and train a simpler network on that.

If time permits in Phase 3, I would like to train a RNN to repeat clips/events for repeated sections in the music, otherwise this will be essentially a random mixing of events to sections in the music.

Datasets:

Event Detection -

Baidu's Video Highlights dataset (<https://ai.baidu.com/broad/introduction>)

Rhythm detection -

MAPS dataset

(<http://www.tsi.telecom-paristech.fr/aao/en/2010/07/08/maps-database-a-piano-database-for-multipitch-estimation-and-automatic-transcription-of-music/>)

MusicNet (<https://homes.cs.washington.edu/~thickstn/musicnet.html>)

Background reading planned:

1. Highlight Detection with Pairwise Deep Ranking for First-Person Video Summarization
(https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Yao_Highlight_Detection_With_CVPR_2016_paper.pdf)
2. An End-to-End Neural Network for Polyphonic Piano Music Transcription
(<https://arxiv.org/pdf/1508.01774.pdf>)

Result Evaluation:

I will experiment with the score function in Phase 1 for my need. I will use frame-based F-measure and accuracy to evaluate my network in phase 2. In Phase 3 if an rnn is used, frame-based metrics can be used.

In terms of plots, I will be plotting these metrics per epoch for both training and validation