

# Computer Science Tripos – Part II – Project Proposal

## Music Style Transfer

Charles J.Y. Yoon, King’s College

Originator: Prof. Alan Blackwell

19 October 2018

**Project Supervisor:** Dr. Andrea Franceschini

**Director of Studies:** Dr. Timothy Griffin

**Project Overseers:** Prof. Lawrence Paulson & Prof. Frank Stajano

## Introduction and Description of Work

This project aims to create a system that modifies a given song to mimic the style of another. Specifically, this project will create a system that achieves such goal by one or both of the following methods:

1. Creating a deep neural network that learns the similarity between songs in a playlist of a common stylistic trait. Such style transfer was done for images by Gatys, Ecker, and Bethge, in their paper “A Neural Algorithm of Artistic Style”, in which the network was trained to transfer the style of “Starry Night” by Van Gogh to existing paintings.<sup>1</sup>
2. Training a generative adversarial network<sup>2</sup> that learns the transformation of stylistic trait given a set of pairs of song, the pair being similar in content but different in style. This method has been done for images by Isola, Zhu, Zhou, and Efros, which given a base image yields a more detailed image (e.g. from boxes to images of buildings).<sup>3</sup>

In both cases, styles of music can be perceived differently depending on the individual; therefore, the original dataset will be evaluated to see if the set of songs indeed (1) is a playlist with a common stylistic trait in the case of the first method, or (2) are pairs with the same content (i.e. the “same song”), in the case of the second. Moreover, in the case of second method, similarity between the audio pairs can be analysed by existing mechanisms, that may include the open

---

<sup>1</sup>[arXiv:1508.06576](#) [cs.CV]

<sup>2</sup>[arXiv:1406.2661](#) [stat.ML]

<sup>3</sup>[arXiv:1611.07004](#) [cs.CV]

source library Musly<sup>4</sup> and the mechanisms described by McFee<sup>5</sup> and Zadel<sup>6</sup>. In popular culture this may be referred to as a “remix” or a “cover”; however, this project purports to yield any qualitative change, which may not be as drastic as a remix.

Facebook AI Research team has done a research on transferring musical style previously, and this project largely aims to yield a similar result to their research.<sup>7</sup> This research, however, aimed to transfer a piece of classical music into another instrument; we may aim to target a more general characteristic of music people perceive, and a different genre of songs in this project. Nonetheless, their paper elaborates their evaluation method in detail and this project will consult on their approach to determine the evaluation method of its own.

## Resource Required

### Computing Resource

Since this project contains multiple machine learning experiments, a machine with nVidia CUDA capable GPU would be ideal. The Rainbow group has a machine with a nVidia Titan Xp, which this project will mostly train its network on. As a backup, a personal desktop computer with a nVidia GTX1080 will be used to train networks as well.

### Dataset

A dataset for this project does not seem to be publicly available; therefore, preparing dataset for this project will most likely involve compiling a set of songs. For the first method, it would most likely involve compiling a playlist with a similar stylistic trait. Since there are a number of “chill” playlists publicly available on Spotify, this project will use such playlist. For the second method, acquiring an entire discography of an cover artist such as Boyce Avenue and the original songs for the listed tracks should suffice.

## Starting Point

Neural Networks were covered by the course “Artificial Intelligence” during Part IB; however, most in depth contents such as Convolutional Neural Network and Generative Adversarial Networks will need to be reviewed and learned as the project starts.

---

<sup>4</sup><http://www.musly.org/>

<sup>5</sup><https://escholarship.org/uc/item/8s90q67r>

<sup>6</sup><http://www.music.mcgill.ca/~zadel/611/audiosimilarity/similaritypresentation.pdf>

<sup>7</sup>[arXiv:1805.07848 \[cs.SD\]](#)

# Substance and Structure of the Project

## Previous Work Research

The field of Machine Learning and structures of Neural Networks is rapidly growing. I will need to review the relevant papers and textbooks for an in-depth understanding of different network architectures and systems, and the mathematics behind them. During this time I will also follow the works done for Image and Sound Style Transfer to grab a better sense of neural networks.

## Data Preparation

There is not a publicly available dataset of songs, so I will need to acquire them. To do so, I have contacted Spotify but without response as of now. To start, I will try to implement an audio capturing frontend to acquire the songs, without any storage or redistribution of them. If such methods fail, songs from the public domain can be used, or can be purchased through another medium. After compiling a set of songs, the tracks may be preprocessed or split into fixed size chunks in order to yield a better performance from the classifier.

## Neural Network

The largest work of this project is to architect a neural network structure based on previous works that will effectively learn the common characteristic of the dataset that the project intends to transform. To do so there will be multiple trial-and-error iterations to experiment different structures of neural network.

## Presentation

The entire system will most likely include different preprocessors, and a web service for presentation, so that users can transform their piece of music.

## Evaluation

1. Technicality

Without any human input it is still possible to test the consistency of a neural network. By splitting the dataset into the training set, testing set, and a cross validation set, this project will aim to find a quantitative validation of consistency of the neural network.

2. Experiment

I will conduct an experiment on human perception to test that the network has transferred the style of a network as aimed. In order to do so in a controlled manner, experiment will be devised to meet the following standards:

- To confirm the similarity of content on the dataset, volunteers will be given set of samples of five or more sound tracks, and will be asked to select two that are of the same content.
- To confirm the similarity of content on the synthesized data, users will be given the original recording with multiple synthesized sound samples and will be prompted to select the one that has a similar content.
- To attest to the effectiveness of the transformation, volunteers will be asked to pick one that adheres to the characteristics of the transformation.
- For consistency throughout the track, different sections of each song may be presented to the volunteers for evaluation.
- Volunteers will be asked the same questions that are phrased differently, in order to attest to the consistency and reliability.
- The environment in which the volunteers will be conducting the experiment will also be controlled; all volunteers will be using the same headphones in the same listening environment in order to conduct the experiment.

## Success Criteria

- A successful compilation of music dataset;
- An effective rendering and creation of a neural network architecture;
- Statistically effective behavior of the classifier;
- Successful presentation of the functionalities of the classifier, as described above;
- An agreement of the effectiveness of the classifier by the human perception experiment.

## Timetable and Milestones

Since both of my paper 10 modules are in Michaelmas, a larger portion of the project was allocated to Christmas break and Lent Term.

**Until 2 November (4 weeks)**

**Preliminary Reading**

- Reading up on deep neural networks and generative adversarial networks, using the textbook by Ian Goodfellow<sup>8</sup>.
- Reviewing the papers mentioned above.

#### **Milestone**

- Write up on the Backgrounds and Introduction section.

#### **Until 16 November (2 weeks)**

- Trying different implementation of the Image Style Transfer, familiarise with the machine learning systems that will be installed on the computing machines.
- Starting to prepare dataset, by contacting music providers, implementing an audio capturing software, or purchasing.

#### **Milestone**

- Working Image Style Transfer implementation with an in-depth understanding of the network architecture.
- A sizable dataset to start initial experiments.

#### **Until 30 November (2 weeks)**

- Finishing preparation of dataset.
- Initial Experiments on deciding which method may seem more appropriate.

#### **Milestone**

- A full set of dataset for training, testing, and cross validation.
- A section for the dissertation on methodology chosen and initial experimentation.

#### **Until 14 December (2 weeks)**

- Making a decision on the general architecture of the neural network.
- Validating the choices made and making necessary changes

#### **Milestone**

- A working initial classifier for further enhancement.

---

<sup>8</sup><https://www.deeplearningbook.org/>

**Until 18 January 2018 (Christmas break, 2 weeks)**

- Improving the classifier to start evaluation.
- Preparing the evaluation of the dataset and gathering volunteers.

**Milestone**

- A working transformation engine.
- Fully planned experiment for datasets to be executed in the coming weeks.

**Until 1 February (2 weeks)**

- Starting the coding of the presentation code around the transforming network.
- Conducting experiments on the dataset.

**Milestone**

- Results on the experiments on dataset.
- Working presentation layer.

**Until 15 February (2 weeks)**

- Using the completed network to prepare experiments for the performance of the software.

**Milestone**

- Full experiment prepared for execution in the coming weeks.

**Until 1 March (2 weeks)**

- Conduct experiments prepared and use data in order to start writing up.

**Milestone**

- Initial write up on the implementation and evaluation.

**Until 15 March (2 weeks)**

- Writing up the dissertation, completing each part that may be incomplete.

**Milestone**

- Working draft of the write up.

**Until 29 March (Easter break, 2 weeks)**

- Finish writing up of the dissertation, and send to supervisors for review.

**Milestone**

- Finished initial draft of the dissertation.

**Until 12 April (Easter break, 2 weeks)**

- Start revision on the dissertation after feedback.

**Milestone**

- Revised copy of the dissertation.

**Until 26 April (Easter break, 2 weeks)**

- Preparing for submission.

**Milestone**

- Final copy of the dissertation.