INFO8010: Project Proposal

Alakhir Ahmed,[1] **Charles Romain**,[2] and **Straet François**[3]

[1] *aalakhir@student.uliege.be (s183946)*
[2] *romain.charles@student.uliege.be (s174788)*
[3] *f.straet@student.uliege.be (s181703)*

I. DESCRIPTION

In the deep learning domain, the neural style transfer algorithms were found very efficient in order to create new type of art based on already existent ones. Widely used on images and videos, this might also be usable on another type of art : music. That's what we decided to work on for our project. The principle of the algorithm we want to design and train is very simple : from 2 songs, it should create a third one that will be a mix of the 2.

Concretely, using an encoder-decoder architecture, we can produce a condensed vector that represents a single music. Weighting two of them, one can thus combine pieces of audio.

Furthermore, we will also be able to tweak the condensed vector representation of some music artificially, e.g. maxing out some element, flattening the vector... and exploring those effect in order to artificially create some new cover from an initial piece.

II. USED DATA

For this kind of application, we can simply download royalty-free musics on the web (PyTorch has convenient tools for audio loading).

If we observe that we need significantly more data we will use some datasets like described here or here. We can also browse this wikipedia page if we find nothing.

Also, for this application, the effect of the dataset may be significant, the architecture may not work well for some genres it has not been exposed to. There is a trade-off: if one only uses a specific genre, one will expect better performance in that domain.

III. COMPUTING RESOURCES

As the training on a decent amount of music will be expensive (and that we will need to perform spectrograms on these music), we ask, as proposed during the lecture, for some cloud computing resources.

IV. NICE-TO-HAVES

As one can think of the music as being a language, it raised the idea of taking inspiration from language processing techniques to implement our own encoder and decoders. A word adds information and brings sense to a sentence. In a similar fashion, a note adds some information to the melody and change its "meaning", its tone.

V. RELATED WORK

We can take inspiration from this project hosted on github, and event the works the latter project references to.

Also, for sound processing specifically, we can start from this blog post.