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.

Our objective is to build a latent space representation of arbitrary music, starting from the sound itself, that is, using the sound wave instead of a note on/off event stream.

From this representation, one can thus interpolate between two of them to get an intermediate music from two existing ones: we can "merge" sounds into a new one.

Furthermore, we will also be able to tweak the latent space 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.

We will use the same architecture as for *Music VAE* but it uses Midi files as music input, that is note on/off events stream. So, upstream that architecture, we will add suitable preprocessing and first layers that one can find for example in music classification techniques.

This will clearly separate our model into two steps. The first processing with mel spectrograms and CNNs, that will produce an efficient, however high-dimensional and time-dependent representation of the music. The second step will be connected in series and will output the lower-dimensional and time-independent latent space representation of the music, using RNNs and fully connected layers.

The overall architecture is summarized in figure 1.

II. USED DATA

We will use the GTZAN dataset, originally built for music classification. As we will train a VAE, we actually only need the music itself, no matter the possible annotation alongside.

If available (it seems that the web API is down, an email has been sent to verify the information), we will also use the Free Music Archive dataset.

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

Reference used for first music processing techniques:

- Automatic Tagging using Deep Convolutional Neural Networks
- Convolutional Recurrent Neural Networks for Music Classification

Reference for second music processing techniques: Magenta research project: Home page and Music VAE directory

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.

Datasets:

- GTZAN dataset
- FMA dataset

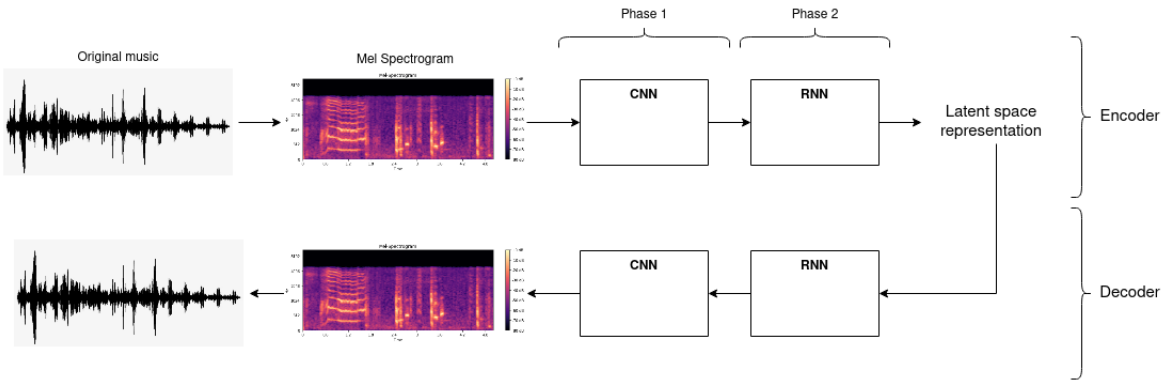

FIG. 1. Architecture of our VAE model