2022-2023 # Group 02	Image & Sound	Daniel Stulberg Huf Lawson Oliveira Lima Lucas Vitoriano de Queiroz Lira

Generating audio reactive visuals

Resources: List the resources that you identified and that you will use to build your applied course (AC).

- Academic paper on audio reactive visuals:

https://neurips2019creativity.github.io/doc/Stylizing%20Audio%20Reactive%20Visuals.pdf

- Some general explanations and contextualization of the topic can be found in these websites:

https://towardsdatascience.com/the-deep-music-visualizer-using-sound-to-explore-the-latent-space-of-biggan-198cd37dac9a

https://towardsdatascience.com/introducing-lucid-sonic-dreams-sync-gan-art-to-music-with-a-few-lines-of-python-code-b04f88722de1

https://medium.com/analytics-vidhya/how-to-create-a-music-visualizer-7fad401f5a69

- Some YouTube videos capturing the result obtained from the process:

https://youtu.be/B2LTEclhZn4

https://youtu.be/SX_mGV8_3Cs

https://youtu.be/Z6--N968KvQ

Goal: Describe the objectives of your AC; what you want the reader/attendee to discover.

Everyone likes music. As Ray Charles once said: “Music is powerful. As people listen to it, they can be affected. They respond”. Each person has its own musical tastes and/or kinds. But imagine if we could perceive a music track by a different sensorial way, a visual one, which could create a unique representation for all of us. In other words, how can we also see music?

In this AC, we want the reader to help us create and play with a music visualizer. More than that, we want him/her to complete the development of a code which can generate a video that is responsive to the features of a soundtrack.

During this process, we also want to introduce the reader to a powerful tool known as Generative Adversarial Networks (or simply GANs), which represent a class of Deep Learning frameworks that typically produce generative artworks.

Main ideas: List the key concepts covered by your AC, what is the "take-home message".

Firstly, one key concept that will be covered in the beginning is audio preprocessing, more specifically music preprocessing. Music can be analyzed as a combination of sounds, which are simply vibrations defined by frequency and amplitude, i.e., speed and loudness.

Then, the preprocessing is followed by the creation of a "deep music video". That will be done by exploring the latent space of BigGAN, a generative adversarial network created by Google which contains over 300 million parameters and generates images from 1128 inputs (a 1000-unit class vector and a 128-unit noise vector). Interpolating between classes and/or noises in the latent space may lead to interesting results, generally explored by artists to build AI artworks. In our case, we will set BigGan to play with music, building a deep music visualizer, "a latent spaceship with Bluetooth".

Finally, we will have to deal with the outcome of the process. In order to do that, we will combine the frames generated in the output of the network into one single MP4 file, a video containing the visual representation of the music that was set as input.

Applicative context: Describe in what context the method you present can be applied.

It is very well-known that generative art has been making its presence these recent years, just take DALL-E or NFT generators as examples. One can argue that machines have equaled man in terms of artistic proficiency, or go even beyond and ask whether art must imitate reality. Thinking in a positive way, art has been made unequivocally accessible not only to experts, but to all of the common public.

In the context of audio reactive visuals generation, the methods that we are proposing could be used in various ways. The PhD student from Columbia Matt Siegelman (referenced above) points out some of them:

- Build a visualizer that responds to live music in real time;
- Use natural language processing to automatically select ImageNet classes based on semantic associations with song lyrics or audio books;
- Play music to people in fMRI, and interface the class and noise vectors with neural activity to create deep music videos from the brain.

Our group also suggest other applications, such as:

- Use the visualizer to help medical patients dealing with stress, by showing them generated videos according to their musical preferences;
- Study the potential intrinsic visual structure of a song or general sound. Such correlation is sometimes naturally observed by people with synesthesia;
- Entertainment tool directed to the deaf community (techno music for example is specifically appreciated by deaf people because of its strong bass and vibration, now imagine combining that with a visual interface).

Tools: Describe the tools that are required to follow your AC.

In order to follow our applied course, the respective tools are required:

- **librosa** (Python package for music and audio analysis)
- **MoviePy** (Python module for video editing)
- **torch, scipy, pytorch_pretrained_biggan** (libraries to implement the generative art)

Data: Describe the data that are used in your AC.

Data will be gathered in two forms. The first type of data is the music itself, which is an MP3 or WAV file from which we are going to extract its features and properties. The other type of data is the pre-trained BigGAN model that will generate the images for the video.

Organization: This section indicates how you will organize yourself (as a team of students) to build your AC. You have to divide your project (building an AC) into several tasks, each one being associated with an outcome and a deadline. You can also associate different students to different tasks if you think that this is relevant.

Task 1

- Description: Data acquisition, preprocessing and preparation
- Outcome: Data should be gathered and parameters should be adjusted to prepare the inputs of the BigGan network.
- Deadline: 03/03

Task 2

- Description: Model handling
- Outcome: The pre-trained model must be loaded and the class and noise vectors properly initialized.
- Deadline: 10/03

Task 3

- Description: Visualization tool
- Outcome: Frames should be generated in batches and then converted to video (MP4 file).
- Deadline: 17/03

Task 4

- Description: MCQ and do-it-yourself coding of the notebook
- Outcome: Finish MCQ questions and let some cells of the notebook unfinished for the students to complete them.
- Deadline: 24/03