1 Introduction

1.1 Inspiration

Starting off with our class in Digital Creativity with lecturer **Guillaume Massol** at the HSLU, I was already very much inspired by the works of artist Refik Anadol. His approach in creating AI based visuals sourced from various large-scale datasets has had a profound impact on the way I see the possibilities in audiovisual artwork nowadays. Especially in combination with generative AI, the way storytelling based on data is achieved was mesmerizing. As I delved deeper into my research on how Anadol created his art I stumbled upon this blogpost where It is described how he uses software like interactive environments for orchestration of various installments that inhibit objects like buildings, projectors, sensors, etc. This inspired the idea for my project that enables AI generated art in combination with audioreactivity to create an experience for the observer containing both visuals and audio.

1.2 Tools

1.2.1 Touchdesigner

TouchDesigner is a visual development platform that enables users to create interactive media systems. It is widely used for real-time interactive multimedia content, particularly in live performances, installations, and exhibitions. The tool offers a node-based interface where users can connect different components to manipulate and render graphics, video, and audio.

1.2.2 Stable Diffusion

Stable Diffusion is an AI-powered text-to-image generation model. It allows users to create detailed images and artworks from textual descriptions. This model is known for its ability to generate high-quality, coherent, and contextually relevant images based on the input text.

1.2.3 Deforum

Deform is an Extension for Stable Diffusions WebUI that enables AI based creation of Animations. It has various parameters that lets the user create visuals in 2D or 3D, use interpolation and add multiple tweaks and peaks into the output generated.

2 Project

2.1 Stable Diffusion & TouchDesigner

2.1.1 Setup

2.1.1.1 TouchDesigner

2.1.1.1.1 Get License Key

As I needed access to the educational version of TouchDesigner to create the desired visuals and utilize a bigger range of its functionality, I requested a License for the educational version at the HSLU which was thankfully granted.

2.1.1.1.2 Grasp TD Fundamentals

https://learn.derivative.ca/courses/100-fundamentals/

After the successful installation I was ready to get a first grasp on the inner workings of TD. As recommended by my lecturer, I initially looked into the basic tutorials described in the link above to get a first understanding of TD and how the different components work.

2.1.1.1.3 Python in TD:

As it was clear to me that I also wanted to use custom code to trigger certain reactivity and use my own Stable Diffusion API I was adamant on learning to understand how to use python on TD. The above article from Mathew Ragan gives a short introduction on installing external python libraries and point to additional sources on how to work with python in TD. I was also not aware before that for python scripts and dependencies to resolve correctly in TD, that the used python version and the internally set version of TD have to be coherent which therefore made congruent by using a python environment for my local machine.

2.1.1.2 Stability.ai

2.1.1.2.1 Setup Account and get API Key

The first step in using Stability.ai services like API calls are setting up an account at their API Platform. After creation one gets a designated API Key which will be essential for development. At creation every account gets 25 tokens which suffice for about 100 images, depending on parameters like resolution and steps given. To be on the safe side and also have a range to try out, I acquired about 1000 tokens for 10 USD that would suffice to about 5000 images. The main reason why I used their API service was definitely speed, as I could have also run the Stable Diffusion on my machine locally but as I don't have much GPU power the time to resolve even one image would resolve in multiple minutes instead of second with the API call.

2.1.1.2.2 Connect SD API with TD

Based on the tutorial above, I went through the steps of creating a custom stable diffusion API for TouchDesigner. After initial installation of the stability_sdk and setting up the environment, asically, one has to create base comp in TD and create custom component parameters inside of it which correspond the values that are taken by the stable diffusion API. There are various parameters but for ease of use I defined just the most important parameters for text-image and image-image generation. For the development of the custom API the Stable Diffusion API documentation was very helpful as reference as it contains many code snippets to create a connection to the Host, setup the parameters so on and so forth. So after the python scripts have developed, the parameters just have to be linked with the parameters of the base comp with the parameters defined before and voila, the API call can be triggered with the images saved in the specified folder. Important for my specific use case for image-image generation was the Start Schedule parameter, as this defines the influence of the input image to the output. As I didn't want my output image to vary strongly from the input, it was set to about 0.6. This helps in making the images l created to look more transitioned in the following steps of creating an audioreactive visualization and giving the illusion of images evolving from one another.
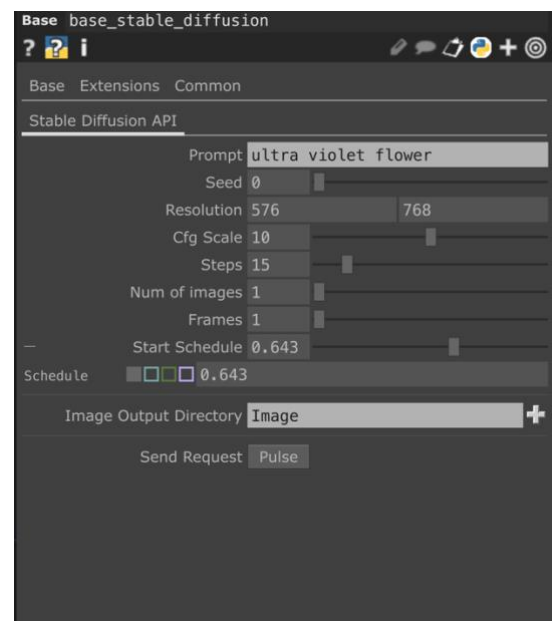
2.1.2 Final Project

Now that the foundational work has been done, we can move on to the fun part and connect the API with the parts of TouchDesigner. For this part I related from following YouTube Tutorial: https://www.youtube.com/watch?v=4wpn_3JNaIc

In TD I created a base comp called independent which creates certain defined shapes and forms which in my case were circles that react to the input audio and run on a separate timeframe so that it does not interfere with the main timeline. In there I get the audio input and extract the spectrum of its waveforms as input the visual TOP that generates forms which then react to the audio. After that I go back one level higher where the base comp independent and the SD_API base comp area set. There I extract the audio spectrum from the independent base comp, trim it and also set a range through a math CHOP and handle it so that through an execute CHOP the generated images will be displayed in a MoviePlaylist and get triggered to change to the next image when the specified range is met through the waveform.

In the end I put the generated images and the audioreactive shapes to one TOP composite and adjust the layover so that it fits to the general theme. This will look like this:
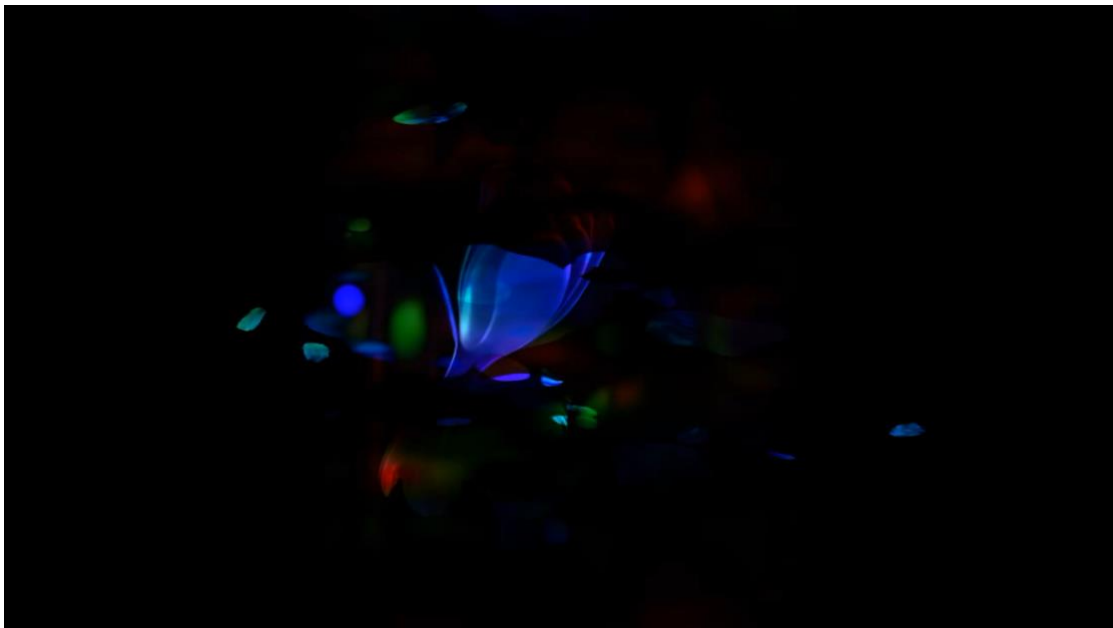
2.2 Stable Diffusion & Deforum Extension

2.2.1 Setup

For setting up Stable Diffusion I used the sd_webUI by automatic1111 provided by our lecturer. I used the HSLU intern GPU Hub to run it as my local machines have very limited GPU power and generation of images and especially videos with higher framerates take far too long. I then linked the Deforum GitHub link as extension in
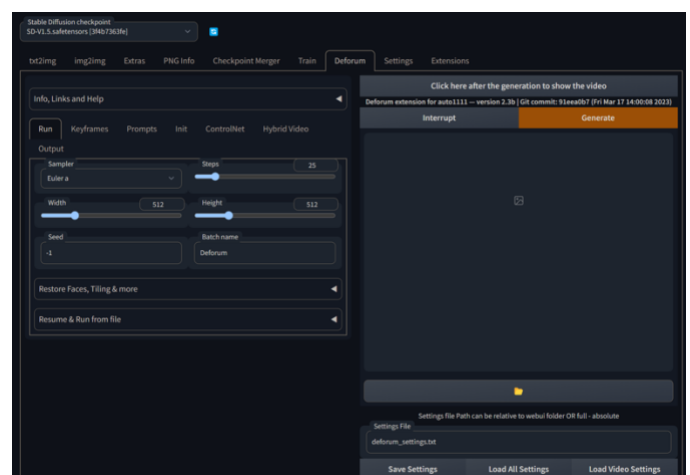
the build of the SD-WebUI and could then use it as extension.

2.2.2   Parameters

In creating an audio-reactive visual, it's crucial for the images to move in sync with the beats of the audio. To mimic such behavior, adjusting the strength parameter in the prompt can yield good results. Since audio files are essentially waveforms that move up or down, the strength parameter can be set for every second, either higher for a more distinct change from the previous image or lower to maintain the illusion of a static visualization. A helpful website has been developed with a tool called Boogie, where you can input the beat and additional parameters like framerate, minimum and maximum strength, etc., to generate a strength plan for every second of the total frames.

2.2.3   Final Product

After setting up the Stable Diffusion WebUI and extending it with the Deforum Extension, as well as generating the necessary strength values, the visuals can be created. The generation time varies depending on factors such as framerate, resolution, steps, etc. For the following visuals, I used a resolution of 1024 for both width and length, 25 steps, 3D, and a framerate of 10 & 15 Frames per Second. I also utilized GPU Hub to run my SD_WebUI, which, in total, took about $6 - 8$ hours for a maximum of 600 frames. For the audio, I used the instrumental version of Omah Lay's song "Soso." However, I encountered some issues in adding the audio directly during generation, even though I specified the audio path. Consequently, I had to manually add it afterward using iMovie. In the end, the result is as follows:
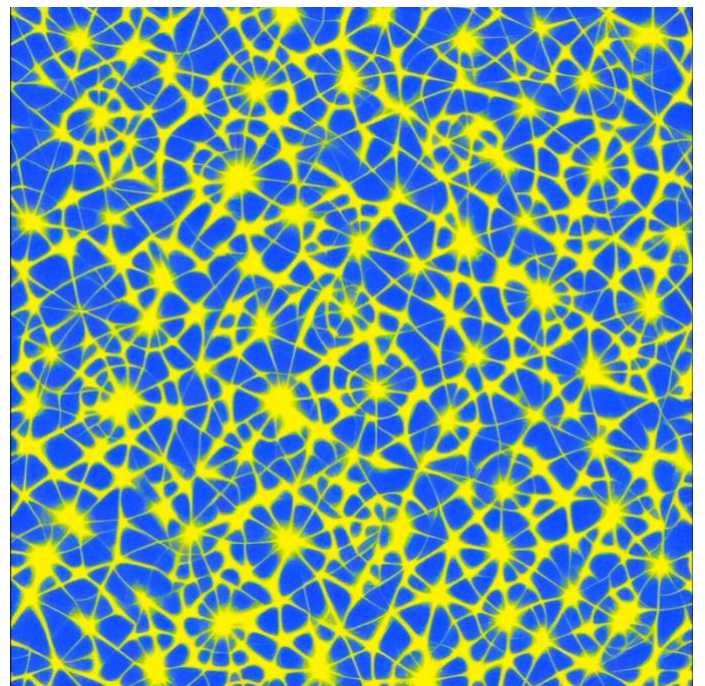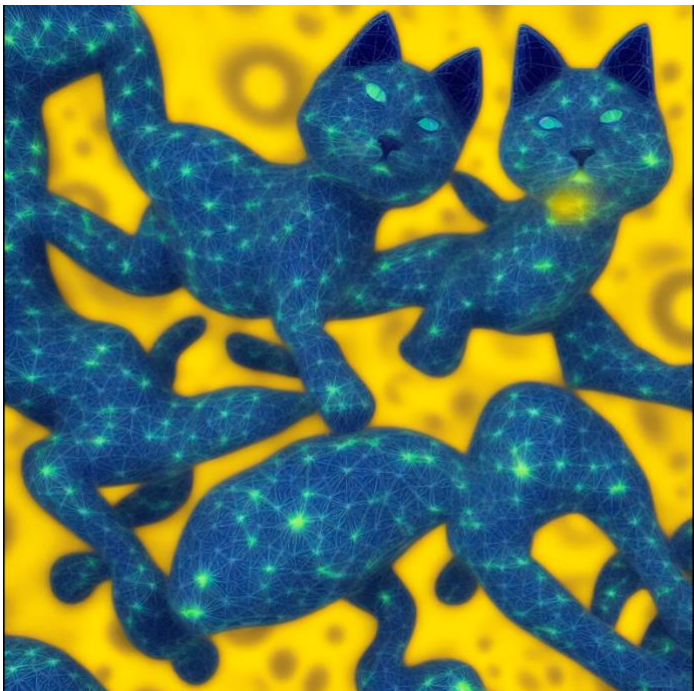
3 Collaboration

3.1 Jakob Fender

While discussing parametrization in Stable Diffusion, particularly with Deforum, Jakob and I explored the impact of certain parameters on the generated visuals and how to ideally define them to create images that look realistic while also conserving GPU power. We concluded that reducing the number of steps and lowering the frames per second to about 10 would yield reasonably good images while keeping the generation time within an acceptable range. Additionally, we delved into the process of generating images. We were amazed at how the parameters and text prompts interacted, producing completely different results even when only a small and seemingly trivial change was made. This provided us with valuable insights and a deeper understanding for our projects.

3.2 Lukas Lottenbach

Lukas and I were discussing the idea of generating content, including short movies and series, with the help of AI. We were particularly interested in 2D animations, envisioning a setup where a static background is complemented by a dynamic second layer on which the storytelling would unfold. This concept also ties in with the notion of audio reactivity, where the use of scores could enhance the tension of a scene. The idea is to employ audio tracks that not only underscore the narrative but also have a tangible impact on the visuals. By doing so, the audio would not just accompany the visuals but actively influence them, creating a more immersive and responsive storytelling experience.

3.3 Team Dave Decker/Jovana Gusak/Manuel Luthiger

At the beginning of the project, we all came together, united by the same goal of creating audio-reactive visuals. During this initial meeting, we discussed and explored the different possibilities for incorporating audio reactivity into visuals, brainstorming how to set up such a project. We had many ideas, including the concept of layering the visuals and assigning each layer a specified range of audio waves so that certain layers would react independently, creating a more fascinating audio-reactive experience. Although this approach seemed feasible with TouchDesigner, in the end, none of us implemented it in our respective projects, as experience has shown that it is not quite as easy as anticipated. Nonetheless, having this brainstorming session was a great starting point for approaching this project.

4 Conclusion/Lessons Learned

To conclude, it is fair to say that this project was a great way to become familiar with AI generation and audio reactivity. I learned a lot about the workings of TouchDesigner, which is an amazing tool. Additionally, understanding how Stable Diffusion works and how extensions like Deforum can be utilized to further specify one's use case was a valuable insight into the world of generative AI. I definitely struggled with many aspects of working with TouchDesigner. While it offers many possibilities, my limited familiarity with it prevented me from realizing its full potential. As a result, I had to find many not-so-elegant workarounds to advance this project. In retrospect, it was still worthwhile, and I would definitely use it again, though I would plan to spend considerably more time studying its fundamentals and inner workings. Learning to work with Stable Diffusion was another achievement. I am amazed by the possibilities it offers and its adaptability to customizable use cases. At this point, I must thank Stability AI for providing such a powerful model to the world for free, and I am excited to develop further projects with it.

Overall, this was a very valuable learning process for me, and I am happy with the progress I made. Although I am still only scratching the surface of generative AI, it has been a good experience, and I will definitely build upon this.

5 Outlook

Looking ahead, I am certain that TouchDesigner will be an essential tool for my upcoming projects, especially those requiring sophisticated interactivity and dynamic visualization. My enthusiasm for further developing the custom Stable Diffusion API is immense, as it opens up a realm of possibilities, like creating more complex generative content and achieving even finer audio reactivity. I am particularly captivated by the prospect of designing immersive visuals that can be seamlessly integrated into live environments, such as concerts or DJ sets, enhancing the audience's experience. The potential of Stable Diffusion in these contexts is undeniable. This model, with its seemingly limitless capabilities in image and video generation, is poised to be a fundamental asset in my toolkit. I look forward to exploring its full spectrum of functionalities and am grateful for the opportunity to incorporate such an innovative technology into my work.