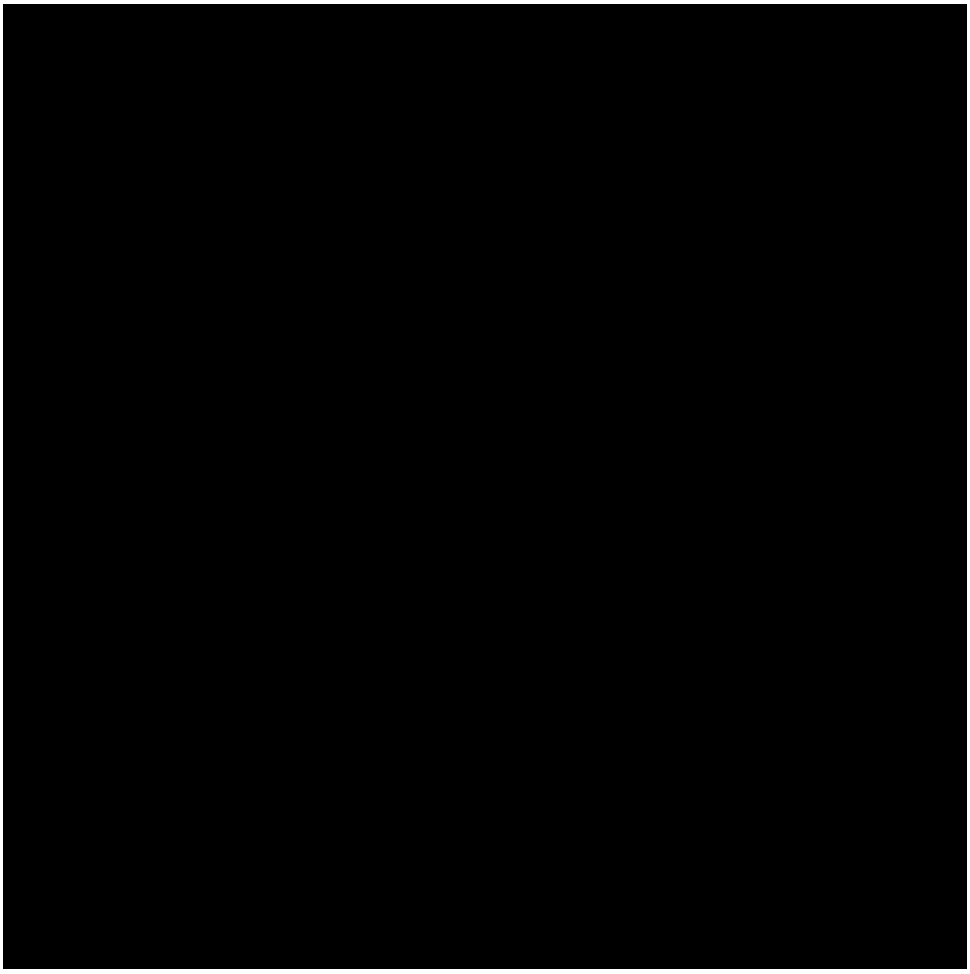
ART AND MACHINE LEARNING
CMU 2023 SPRING
PROJECT 3

Metamorphosis



Members: Joong Ho Choi, Kyle Yang, Zun Wang, Aditti Ramsisaria

DESCRIPTION

This project delves into the philosophical implications of mirrored metamorphosis through the application of machine learning techniques. We use a music transformer and text-to-video generator to create a visually captivating representation of transformation, chronicling the journey of a flower as it blossoms into a magnificent plant with angelic wings, only to return to its humble beginnings. Through experimentation with different mediums such as audio, video, and language, we explored the intricacies and nuances of transformation over the passage of time, and how viewing a mirror of the growth process can offer a unique perspective on your journey.

Concept

Transformation is cyclical. When we think about growth and change from a typical forward perspective, we tend to focus on the future - on where we want to go and what we want to become. The mirrored process of growth and change, which unfolds backward from the end to the beginning, provides a unique perspective on the journey of transformation, prompting contemplation on fundamental questions about the nature of growth and the relationship between the past and the future, beginnings and endings. Ultimately, we invite viewers to reflect on the philosophical implications of growth and change, encouraging a deeper understanding of the transformative power of perseverance and the inherent potential for beauty in every stage of the metamorphosis, no matter the direction from which it is viewed.

Our concept stemmed from our exploration of different video generation technologies, where in the process of generating a video of a plant decaying and dying, the model instead gave us an output of a plant growing angel wings in death. We found this to be very intriguing and decided to explore the symbolism of this imagery further. The sight of a flowering plant growing angel wings can be a powerful symbol of hope and inspiration, as it suggests that even the most humble and ordinary beings can achieve something extraordinary if they persist and grow. This can be a message of encouragement for people who are facing challenges or difficulties, reminding them that they too have the potential to soar beyond their limitations and achieve greatness. We also decided to explore time reversal on sequential data, using it as a way to illustrate the future as a reflection of the past.

Techniques

1. **Text-to-video**

We ended up using the AI text-to-video generator Kaiber. We first gave it an initial image generated by Midjourney, and then used text prompting and tuned settings to generate the video output. We then mirrored the video output so that it plays backwards as well. Unfortunately, like Midjourney, there is only speculation on what machine learning model Kaiber AI uses for its video generation capabilities, with no code or reference paper available.

2. **Music Generation**

We used a PyTorch implementation of Google Brain's Music Transformers and trained it on the MAESTRO v2.0 dataset which contains close to 1300 midi files. The Music Transformer model takes as input a sequence of musical notes or events and generates a sequence of new notes or events that follow the input sequence in a musically coherent manner. The model uses a relative self-attention model that allows it to attend to different parts of the input sequence when generating each new output element, and to capture long-range dependencies and structures in the music. We trained the model with 6 decoder layers, a batch size of 4, with a max sequence size of 1024 for 100 epochs. We then used different primer inputs to generate outputs, and chose the one that fit our theme the best. This was using a primer seed of an opening theme of Card Captor Sakura taken from here.

Process

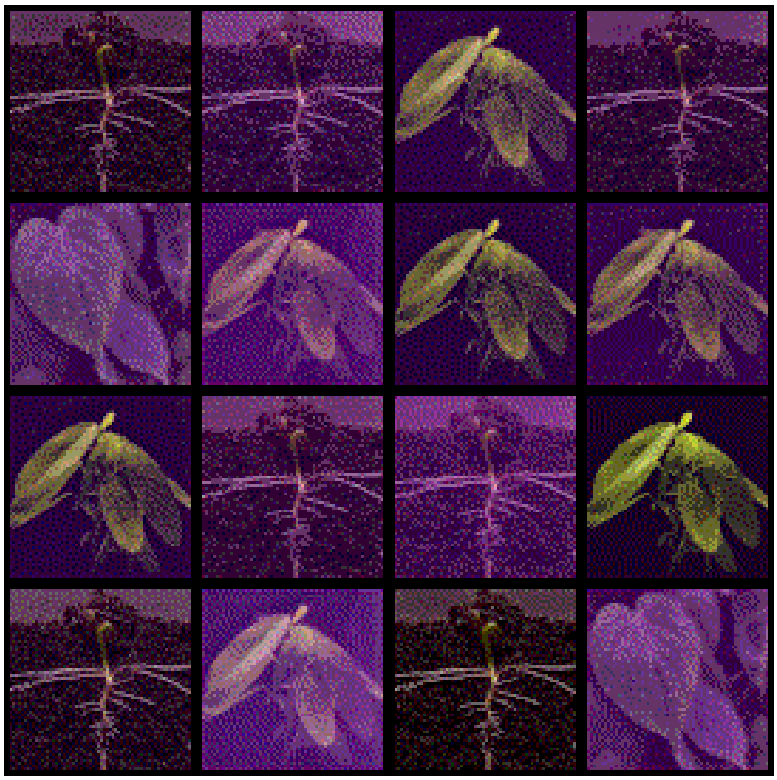
We experimented with multiple techniques for each output modality, and tried to illustrate different combinations of concepts in each modality to tell different stories.

1. **Video Generation Techniques:**

    a. ***Stable Diffusion Videos***

    The particular video diffusion model we used was essentially an extension of stable diffusion but trained to model fixed numbers of frames at fixed spatial resolutions. It uses an extension of U-Net, which is a neural network that is built on 2D convolutional residual blocks, but it extends this network by taking into account

space and time. They call this specialized extension 3D U-Net. Trained over four 10 frame gifs of plants growing of size 64 by 64 for 3000 steps with a batch size of 32.



b. GenMo AI

GenMo AI is a generative media platform with a text-to-video model that can generate immersive live artwork from text prompts or any image. The platform uses a proprietary blend of deep learning models, including Generative Adversarial Networks (GANs), to generate high-quality videos that are realistic and engaging. The key feature that attracted us to this application is that the user can input a series of edits to describe different scenes to set the narrative for the generated video output. We used GenMo to generate the following sequence based on a random thought about having hair made of spider silk as a first attempt:

Initial prompt: a woman covered in spider silk and bubble wrap, realistic, sad
Edit 1: make the silk turn into her hair
Edit 2: add an expression of sorrow
Edit 3: make it psychedelic

As the results looked promising, we tried GenMo for our final concept as well. However, the final results were not great. This time we tried an image generated by Midjourney (on the left) as the initial input.

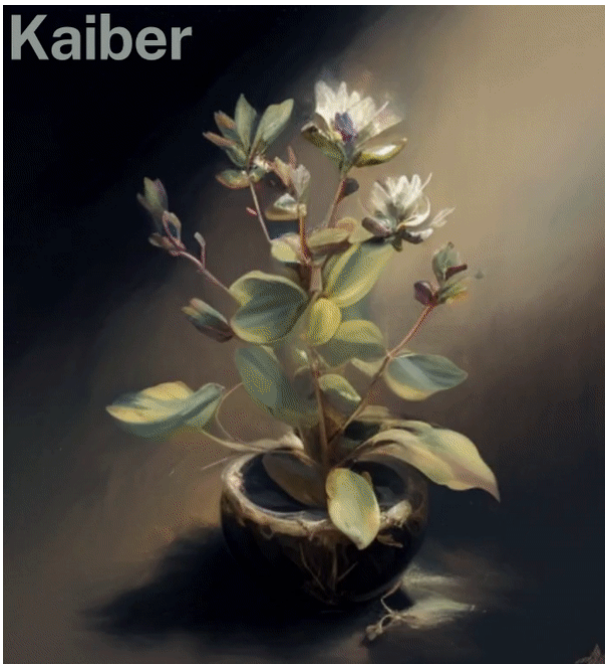
Edit 1: make the plant grow beautiful flowers
Edit 2: make the plant slowly age and die over time

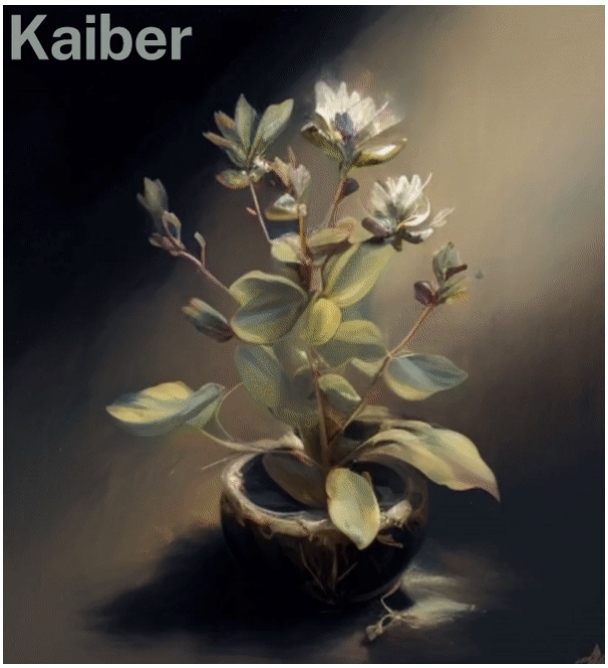
The results we generated were not exactly what we had in mind, so we continued exploring different technologies.

c. Kaiber AI

Inspired by Linkin Park's new music video for "Lost", we decided to try out Kaiber AI. Similar to GenMo AI, Kaiber is a text-to-video generation application. We used a prompt to generate a video of a plant decaying over time in a cinematic style, but interestingly this resulted in a video of a plant growing bigger, growing flowers and eventually growing angel wings.

Subject: A beautiful plant with flowers dying and decaying over time
Style: Cinematic (pre-defined)





We think that the result included angel wings due to the use of the "Cinematic" style. Looking into it more, it seems like that particular style adds the phrase "*exquisite lighting setup, 8k ultra fine detail, private press, associated press photo, angelic photograph, masterpiece*" to the prompt. The "angelic photograph" aspect is likely the reason why angel wings were introduced in the video. To confirm, we retried the same prompt with manually defined style, which got rid of the angel wings - but we decided to use the first one as it was visually and conceptually more interesting.

2. **Music Generation Techniques:**

    a. **Midi-RNN**

    We trained the model using a dataset of about 300 files of classical piano music from composers such as Beethoven, Bach, Rachmaninoff, etc. We tuned the parameters (mainly batch size, epoch, and temperature) and received interesting results but all were far too depressing for our theme. We tried adjusting the temperature and modifying the dataset to select music that were more upbeat but the results remained unsatisfactory as they were either too simple or too depressing. The outputs can be found here.

    b. **Music Transformers**

    We first trained the model using the same parameters on a custom midi dataset of about 450 files of piano pieces (Alex Original). However, we found this to be insufficient data to train this model and the results were not great. We then decided to train on the MAESTRO v2.0 dataset, which generated much better results. For generating our outputs, we used different anime sequences as seeds into the model. The resulting outputs can be found here.

    c. **Music Editor**

    After deciding on the music we wanted to work with, we made some edits through the Audacity music editing software. We first trimmed the music to match the length of our evolving art work. But since the art work we picked reverses back to its root, we decided to do the same to the music by reversing it. So the final product is a piece of music that crescendos before it reverses back to the beginning.

3. **Sentence Generation:**

    a. **Hidden Markov Model**

    Our first exploration choice was the Hidden Markov Model (HMM), which is a statistical model that can be used for generating poetic sentences. The model works by analyzing a corpus of existing poetry and learning the patterns of language, such

as the probability of certain words following others. Once trained on this data, the HMM can generate new poems by selecting words based on the probabilities it has learned. To generate a line of poetry, the model starts with a randomly chosen word and then selects subsequent words based on their probability of following the previous words. This process continues until the desired number of lines or stanzas is generated. By adjusting the parameters of the model, such as the number of hidden states or the size of the training corpus, the HMM can generate poetic sentences that are more or less constrained by the patterns of language it has learned.

Unfortunately, the content of the sentences generated was too incoherent, so we needed to switch to another method.

b. ChatGPT

We used the existing chatGPT platform to generate sentences of both positive and negative sentiment. We would then later incorporate them into rains of text displayed on the pygame screen. Positive sentences would be displayed for uplifting music, whereas negative sentences would be displayed for chaotic music.

Reflection

Our initial idea was to design a screen with continuous rain droplets of positive, encouraging statements, which are generated by AI, that would drop from top to bottom. We would also have AI-generated music play in the background, and its vibe would be uplifting and merry. At the bottom of the end, we would have gifs of plants—generated by diffusion models—growing as they intake positive sentence rain drops.

However, due to the nature of the dataset, the music generated instead sounded tumultuous. As a result, we decided to pivot our idea to raindrops of critical, judgemental, anxiety-inducing sentences, as we play chaotic piano music in the background. We would replace gifs of growing plants with gifs of plants dying gradually. After we created this design, we felt that we were straying from the original message that we had wanted to convey. We were fitting our message to our means, and it should be the other way around.

Hence, we pivoted once again and it led us to our final design covered in the result section later on. We decided to not incorporate the rains of text into our final design, as we felt that covering the screen with the AI-generated artwork gif was more visually and emotionally provocative. Although we could still have rain of texts on the gif, we felt it would diminish the audience's experience. We also used different software to customize the music generated by AI to suit our message more.

Therefore, through this assignment, we had meaningful experience of exploring different ML means and learning how to shape their constrained/limited outputs to depict the message we want to convey.

RESULT

We chose the following clip of changing artwork with music in the background to be our final submission work. The sight of a flowering plant growing angel wings can be a powerful symbol of hope and inspiration, as it suggests that even the most humble and ordinary beings can achieve something extraordinary if they persist and grow. This can be a message of encouragement for people who are facing challenges or difficulties, reminding them that they too have the potential to soar beyond their limitations and achieve greatness.

CODE

https://github.com/JC-78/artML/tree/main

REFERENCES

https://github.com/lucidrains/video-diffusion-pytorch

https://arxiv.org/pdf/2204.03458.pdf
https://colab.research.google.com/drive/15HKMeyS_vU_zzMEvSM6-ttZTorAA-Ihv?usp=sharing
https://github.com/gwinndr/MusicTransformer-Pytorch
https://colab.research.google.com/drive/1E6HxRHWGcHOEJ9MIbYtKechRJ75_I8vM
https://app.kaiber.ai/create
https://alpha.genmo.ai/