Using Generative Adversarial Networks to Create Images and Videos

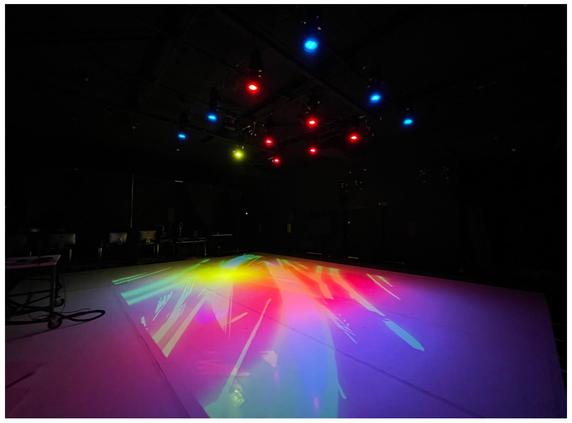
Joshua Pardhe, Computer Systems Engineering, Business (Minor), International Business (Cert.), Cross-Sector Leadership (Cert.)
Mentors: Dr. Ryan Meuth, Dr. Jenny Brian, Dr. Kristen Hermann.
School of Computing and Augmented Intelligence & Barrett, the Honors College

OBJECTIVES AND GOALS

Creative Frameworks was born out of a desire to build inclusive solutions that augmented human creativity. The goal was to build a system for all users to present their creative ideas through images and videos, using an AI model to generate visuals based on text. To make it as accessible as possible, researchers started with the simplest, oldest, and most common computer interface: a keyboard. Success would be defined as a user being able to generate their choice of images or videos through just a keyboard input.

MATERIALS AND METHODS

Researchers leveraged the VQGAN, MIDAS, and DPT libraries [1] and heavily modified the generator to work with serialized input data. The Spotify APIs were used to capture lyrics, timestamps, and tempos for a given track, and Python was used to auto-generate CSV files with the information. Users would then input their desired visuals as a text string for each lyric. A Python script then read the lyric files and generated a JSON string that was used by the custom VQGAN model to generate images. The AI generation allows for two input types: image and video. Image runs the model upon itself, increasing resolution and image detail. Video uses the generated image as an input, slightly transposing it on the X, Y, or Z-axis for each frame, generating a moving sequence of images merged into a video via FFMPEG. Therefore, the same technology could be used for two different generation types. By leveraging AWS and Google Colab, visual generation could be run on high-performance servers to speed up the creation time.



The Creative Frameworks software during a performance.

An image generated by the GAN as a photo.

The same image as above, generated as a painting from the GAN.

RESULTS

As seen on the left, the VQGAN generator produces extremely viable results. The description and formatting of the prompt are key to the quality of the output. Over two dozen users successfully submitted images and videos for generation. Video generation was successful, but due to the nature of the feedback loop, videos were extremely jittery when viewed directly from the program. Premiere Pro was used to interpolate and slow down the video to reduce the jitter and produce a smoother video output.

CONCLUSIONS

The goal for this project was to develop a system for users of all abilities to express themselves creatively through images and videos by using only a keyboard. Through this research, a series of interconnected Python scripts can convert an idea into a high-quality video or image with only a keyboard and simple text-based prompts. Future research can include Python-based interpolation to enhance the quality of the output, and continued development of the VQGAN model for better image interpretation.

ACKNOWLEDGEMENTS

Directors:
Dr. Ryan Meuth, Dr. Jennifer Brian, Dr. Kristen Hermann

Advisors and Special Thanks To:
Dr. Mark Jacobs, Dr. Steven Tepper, Dr. Anne Jones, Ms. Colleen Jennings-Roggensack, Lance Harrop, Nicholas Beck, Connor Rawls, Peter Weisman

FURI

Ira A. Fulton Schools of Engineering
Arizona State University