SoC: Morphix Midterm Report

Through my exploration of StyleGAN2-ADA resources provided by my mentors, I have gained a comprehensive understanding of generative adversarial networks and their application in high-quality image synthesis. Here's what I've learned so far in my project journey. This report captures what I've learned so far, from basic understanding to practical implementation and advanced latent space editing.

Week 1: Understanding the basics and setting up the environment

In the first week, I began with the fundamentals of Generative Adversarial Networks (GANs) and the architecture of StyleGAN2. I understood that GANs consist of two neural networks competing against each other - a generator that creates fake images and a discriminator that tries to distinguish between real and fake images. Through the various videos provided by the mentors, I learned that StyleGAN2 represents a significant advancement over traditional GANs.

The concept of latent spaces became central to my understanding. I discovered that StyleGAN2 operates with three distinct latent spaces:

- **Z Space**: The initial latent input, normally sampled from a standard Gaussian distribution.
- **W Space**: A more disentangled latent space, allowing for clearer attribute control.
- **W+ Space**: Provides even finer control by using different latent codes for each layer in the synthesis network.

Watching the provided videos, made the motivation for latent space editing clear: semantic features like age, expression, and gender are encoded in specific directions in this space, making intuitive edits possible.

I also set up the StyleGan2-ADA environment through WSL. Even though WSL was not necessary this project was a great opportunity to try it out. Finally, after the environment was set up, I generated images using the generate.py experimenting with different seeds and truncation values to understand the trade-off between variability and realism.

Week 2: Deep Dive into Latent Spaces

In Week 2, I dove deeper into the latent spaces and their manipulations. I learned how projecting an image into StyleGAN's latent space enables editing real photos using generative techniques.

One major takeaway was the process of latent vector interpolation, which involves smoothly transitioning between two images by interpolating their corresponding latent codes. This process involves taking two latent vectors and creating intermediate points between them, resulting in morphed images that blend characteristics of both source images.

I also explored the concept of image projection, where a real image is "projected" into StyleGAN2's latent space. This is done by optimizing a latent vector such that the output of the generator closely resembles a given real image. Through this, I realized how real-world data could be edited using StyleGAN's generative capabilities.

The main tools I used in this week were:

- **CLIP2STYLEGAN**: for unsupervised latent direction discovery using text prompts.
- **Woctezuma's stylegan2-projecting-images repo**: for reconstructing real images in the GAN's latent space.

Overall, the second week helped me understand the inner workings of StyleGAN2's latent spaces and how they can be manipulated to produce coherent and meaningful image transformations.

Assignment 1:

The necessary files are attached in the github repo.

Week 3 and 4: Semantic Edits & Disentanglement

In these two weeks, the resources were about understanding how to make semantic edits in StyleGAN2's latent space and how different attributes can be controlled independently through disentanglement.

I started by learning about basic attribute manipulation. I learned that various facial features such as age, gender, and expression can be adjusted by moving in specific directions within the latent space. These directions correspond to semantic concepts, and altering them allows for controlled edits. For example, I could make a face appear older or change its expression from neutral to smiling simply by modifying the latent vector along the appropriate direction.

Another key concept I studied was disentanglement, which refers to the separation of different attributes in the latent space. Ideally, modifying one characteristic (like age) should not unintentionally affect others (like gender or pose). I understood how important disentanglement is in achieving precise and isolated image edits.

The key tools I used this week were:

- **InterfaceGAN**: Enables vector arithmetic in latent space for controlled edits.
- **GANSpace**: Uses PCA to find interpretable directions for edits like pose, smile, lighting, etc.

By the end of the week, I had a clearer picture of how StyleGAN2's latent space supports intuitive and interpretable transformations.

Conclusion

So far, learning StyleGAN2-ADA has been interesting. I have learned how high quality images are generated from random noise, how those images are encoded in a semantically meaningful latent space, and how to edit those images using vector arithmetic in that space. Tools like InterfaceGAN and GANSpace have shown me how research translates into intuitive and powerful generative editing tools.