Understanding and Exploring the Latent Space of a Generative Adversarial Network

Project 2

Deep Learning in Computer Vision

June 2023

The main goal of this project is to understand and explore the latent space of a GAN model.

MNIST

You have also been working on creating networks that can generate digits from the MNIST dataset. If you are in need of more content for your poster, you are allowed to also include these on your poster. For this you could: show results from these networks, describe their architectures, the training process, and describe your process of making these choices.

Pre-trained GAN

In this project, you will work with a pretrained StyleGAN2 model which was trained on the FFHQ dataset. The pretrained network can be found here: https://nvlabs-fi-cdn.nvidia.com/stylegan2-ada-pytorch/pretrained/ffhq.pkl. If you want to explore pre-trained StyleGAN2 models on other domains (e.g., cats, dogs, art faces), you can find these weights here: https://nvlabs-fi-cdn.nvidia.com/stylegan2-ada/pretrained/

Useful links

- StyleGAN2-ADA official implementation in pytorch: https://github.com/NVlabs/stylegan2-ada
- Align images to FFHQ: https://github.com/happy-jihye/FFHQ-Alignment
- Pre-trained latent directions (e.g. age, gender, smile): https://hostb.org/NCM
- CLIP and StyleGAN code: https://github.com/vipermu/StyleCLIP
- VQGAN and CLIP (code): https://github.com/nerdyrodent/VQGAN-CLIP
- VQGAN and CLIP (colab):
  https://colab.research.google.com/drive/1ZAus$_gn2RhTZWzOWUpPERNC0Q8OhZRTZ$

Tasks

1. Generate random images from a pre-trained GAN model
2. Reconstruct your own images and get the latent codes using the projector.py script from the official stylegan2 implementation.
   - Do you need to align the images? Does the reconstruction work better?

- Are there examples of real images where the reconstruction works better than others? How does the visual output relate to the loss progression?
- **OPTIONAL:** Can you improve the GAN inversion pipeline (e.g. initialization, losses, encoder)?

3. Interpolate between two real images that you reconstructed.
4. Apply some pre-trained latent directions to your reconstructed images (e.g. aging, smiling, gender).
5. Learn your own latent direction
   - Download a few images that clearly show the direction that you want to learn (e.g. 20 blonde and 20 brunette women or N faces with sunglasses and N without).
   - Align the images and learn their latent codes.
   - Train a linear classifier (SVM) to learn the latent direction.
   - Apply this latent direction to a few real images.
6. Generate images from text prompts using CLIP (e.g. "an image with the face of a woman with purple eyes and blonde hair").
   - Does the initialization (initial latent code) matter?
   - Can you condition the generation on a image and a text prompt?
   - Can you improve this baseline code? (e.g. optimization, learning, initialization, losses)

Hand-in

Your process, performance evaluation, and results should be documented and discussed in a PDF poster to be uploaded on DTU Learn. The deadline for this is Thursday at 23:59 (midnight).

Additional task

You should give peer feedback on 4 other posters during the physical poster session on Monday 19.06.