CSS485 DEEP LEARNING

Project 4 Report

Face Generation

Submitted to

Dr. Watthanasak Jeamwatthanachai

By

Ayush Kwatra 6322774645

Chayanan Rattanaboon 6322772433

Sorawee Chirarattanawilai 6322771781

18 December 2023

Table of Contents

1. Datasets

There are 3 datasets that are used in Part 2 codes, which are FFHQ, MetFaces, and AFHQ.

Flickr-Faces-HQ Dataset (FFHQ) is a high-quality image dataset of human faces, originally created as a benchmark for generative adversarial networks (GAN). The dataset consists of 70,000 high-quality PNG images at 1024×1024 resolution and contains considerable variation in terms of age, ethnicity, and image background. It also has good coverage of accessories such as eyeglasses, sunglasses, hats, etc. The images were crawled from Flickr, thus inheriting all the biases of that website, and automatically aligned and cropped using ‘dlib’. Only images under permissive licenses were collected. Various automatic filters were used to prune the set, and finally Amazon Mechanical Turk was used to remove the occasional statues, paintings, or photos of photos.

MetFaces Dataset is an image dataset of human faces extracted from works of art. The dataset consists of 1336 high-quality PNG images at 1024×1024 resolution. The images were downloaded via the Metropolitan Museum of Art Collection API, and automatically aligned and cropped using ‘dlib’. Various automatic filters were used to prune the set.

Animal Faces-HQ dataset (AFHQ) consisting of 15,000 high-quality images at 512×512 resolution. The figure above shows example images of the AFHQ dataset. The dataset includes three domains of cat, dog, and wildlife, each providing about 5000 images by having multiple (three) domains and diverse images of various breeds per each domain.

2. Problem-Solving

In part 1 of the project, we can’t generate faces with control of the features in traditional GAN so in part 2, we decided to use controllable GAN to solve the problem of generating faces with controllable/disentangling features. We have used age, hair color, and facial expressions as controllable features for generating faces in this project.

3. Correlation and Entanglement Feature Solving

- **Disentanglement by contrastive learning:** Training a GAN with characteristics that are clearly disentangled. The outcome is the division of the latent space into sub-spaces, each of which encodes a distinct image.
- **Interpretable explicit control:** An MLP encoder is trained to map control parameter values to an associated latent sub-space for each property. This makes it possible to explicitly control every single property.

4. Pre-Trained Model

We used a pre-trained controllable GAN model with StyleGAN2 architecture for generating faces in Part 2. Here is the link to download the pre-trained model that we used:
https://drive.google.com/file/d/19v0lX69fV6zQv2HbbYUVr9gZ8ZKvUzHq/view?usp=sharing

5. GAN Architecture: StyleGAN2

Generator Architecture:

- **Mapping Network**: StyleGAN2 begins with a mapping network that receives an input of a random noise vector and converts it into an intermediate latent space, often known as a "W" space. This process involves numerous fully connected layers.
- **Style Modulation**: At various stages of the synthesis network, the generator modifies the image's style using the converted latent vectors. Every convolutional layer experiences this modulation, which affects the resulting image in different ways, ranging from tiny details (like hair texture) to coarse features (like facial form).
- **Synthesis Network**: The actual image is generated progressively, starting from low-resolution images and increasing gradually to the final resolution. Each level of this network adds more details to the image.

Discriminator Architecture:

- StyleGAN2's discriminator determines whether an image is generated or real based on the dataset and whether it is a phony image created by the generator.

Advantages of StyleGAN2

- **Improved Image Quality**: StyleGAN2 produces images of higher quality and resolution compared to its predecessor and to the images generated in Part 1. The images are more realistic, with better texture and fewer artifacts.
- **More Control Over Image Generation**: The use of the intermediate latent space and style modulation provides more control over different aspects of the image, enabling nuanced manipulation of the generated content.
- **Better Disentanglement**: StyleGAN2 offers improved disentanglement of features in the latent space. This means that changing one attribute of the

generated image (like age or hairstyle) has less unintended impact on other attributes.

6. Reference

- Alon Shoshan, Nadav Bhonker, Igor Kviatkovsky, Gerard Medioni. (2021). "GAN Control: Explicitly Controllable GANs." Retrieved from: https://assets.amazon.science/bb/c2/15517688468899100f3f815b72cc/gan-control-explicitly-controllable-gans.pdf