ELEC564 – Spring 2023
Homework 6

Date: April 30th, 2023

Student's Name: Hsuan-You (Shaun) Lin / Net ID: hl116

Link to Colab notebook:
https://colab.research.google.com/drive/1G2OoUcAFNHDC9oRlHLJiJ4OtY88AnggZ?usp=sharing

Please use this provided Colab notebook for the assignment.

Problem 1: StyleGAN (7 points)

In this problem, you will use StyleGAN2 for controlled image generation. Make sure to run the first 3 code cells of the provided Colab notebook. The first cell installs StyleGAN2 and its dependencies. The second cell loads a pre-trained StyleGAN2 model for faces. The third cell provides you with some useful utility functions. Some preliminary code on how to generate synthetic faces using the utility functions is also provided in cell 4.

StyleGAN2's generator converts a vector $z \in R^{512}$ drawn from the standard Normal distribution into a 'style' vector $w \in R^{512}$. The generator then processes the style vector to produce an image $I \in R^{1024x1024x3}$. In this problem, you will find a direction in the style space corresponding to perceived gender and use that direction to alter the perceived gender of synthetic faces.

a. *Interpolating between images*: Choose two random noise vectors `z0`, and `z1`, such that the two generated faces have different perceived genders based on the `face_is_female` function [1]. This function uses a pre-trained face gender classifier to make its prediction.

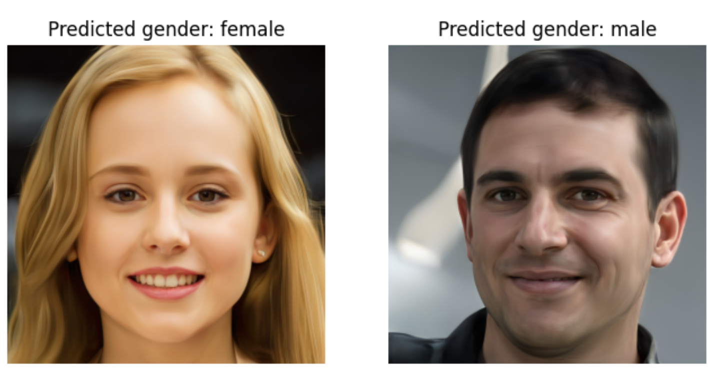

Figure 1. Two random noise vectors z0 and z1 images

i. Interpolate between the **latent vectors** `z0`, and `z1` with 5 intermediate points. Show a strip of 7 faces along with the classifier predictions in your report.

Figure 2. Image of 7 faces interpolated in latent space

ii. Interpolate between the **style vectors** `w0`, and `w1` with 5 intermediate points. Show a strip of 7 faces along with the classifier predictions in your report.

Figure 3. Image of 7 faces interpolated in style space

iii. *Question*: What differences do you notice when interpolating in latent space versus style space? Do the intermediate faces look realistic?

My Answer:

When interpolating in latent space, the intermediate faces look more realistic than when interpolating in style space. I think it's because the latent space is a more direct representation of the underlying structure of the face, while the style space is a more abstract representation of the face.

As you can see from *Figure 2*, the faces in the strip gradually change from female to male as we move from left to right. The faces look more realistic than the faces in the strip interpolated in style space.

b. *Image manipulation with latent space traversals*

i. Sample 1000 random `z` vectors, convert them to style vectors `w`, and get their corresponding perceived genders using the trained classifier. This may take a few minutes.

ii. Train a linear classifier (use scikit-learn's [linear SVM]) that predicts gender from the style vector. The model's coefficients (attribute `coef_`) specify the normal vector to the hyperplane used to separate the perceived genders in style space. Remember to convert your cuda tensors to numpy arrays before sending to scikit-learn's functions.

iii. Sample 2 random w vectors. For each w vector, display a strip of 5 images. The center image will be the image generated by w. The two images to the left will correspond to moving toward the “more male” direction, and the two to the right will correspond to “more female”. To generate the latter 4 images, move along the SVM hyperplane’s normal vector in both directions using some appropriate step size.

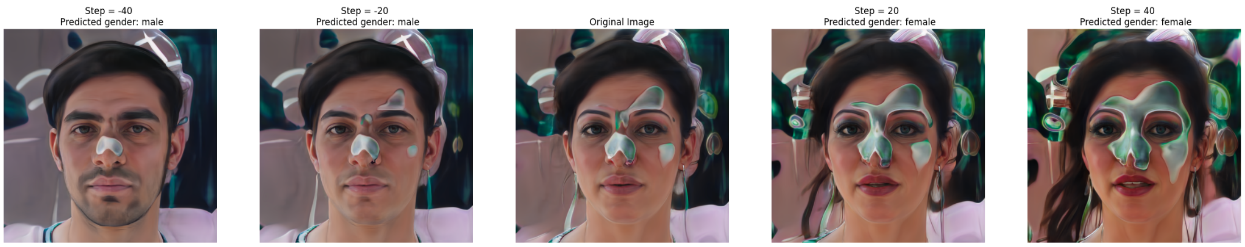

Figure 4. Image of 5 faces generated from left correspond to “more male” and right correspond to “more female” (Step: 20)

Figure 5. Image of 5 faces generated from left correspond to “more male” and right correspond to “more female” (Step: 35)

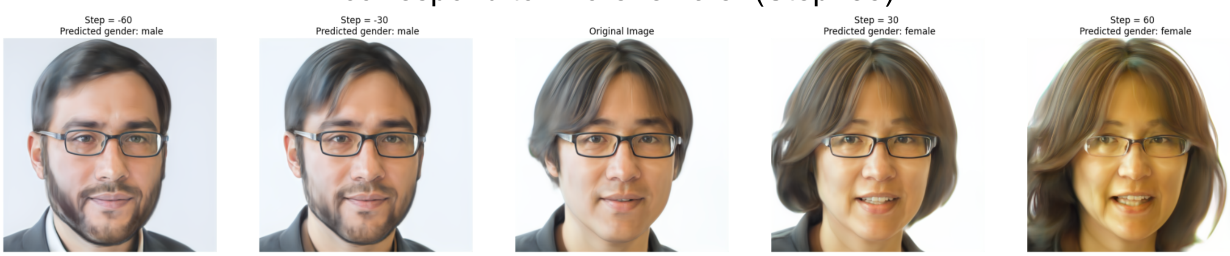

Figure 6. Image of 5 faces generated from left correspond to “more male” and right correspond to “more female” (Step: 30)

iv. *Question*: Do you notice any facial attributes that seem to commonly change when moving between males and females? Why do you think that occurs?

My Answer:

Sure, here are some facial attributes that seem to commonly change when moving between males and females:

- Jawline: Male faces tend to have wider, more angular jawlines than female faces.

- Eyebrows: Male eyebrows tend to be thicker and more bushy than female eyebrows.
- Nose: Male noses tend to be larger and more prominent than female noses.
- Cheeks: Male cheeks tend to be more hollow than female cheeks.
- Mouse: Mal’s mouse tends to be closer than female’s mouse. So females look happier.

These are just a few of the facial attributes that can change between males and females. The reason why these attributes change is due to the effects of hormones. Testosterone and estrogen, the female sex hormone, play a role in the development of secondary sexual characteristics, which are the physical features that distinguish males from females.

The changes in facial attributes that occur between males and females are due to a combination of genetic and environmental factors. Genetics plays a role in determining the basic structure of the face, while hormones play a role in influencing the growth and development of the face. Environmental factors, such as diet and exercise, can also affect the appearance of the face.

Problem 2: Using CLIP for Zero-Shot Classification (5 points)

In this problem, you will use Contrastive Language-Image Pre-Training (CLIP) to perform zero-shot classification of images. You can read more about CLIP in [this blog post](), and check out [the example]() in the official GitHub repository. We will reuse the CIFAR dataset introduced in Assignment 4. Download that dataset as one .npz file [here]() and place it in your Google Drive folder.

a. Perform classification of each test image (last 10,000 images of the dataset) using CLIP. To do so, create 10 different captions (e.g., “An image of a [class]”) corresponding to each of the 10 object classes. Then, for each image, store the label that provides the highest probability score. Report overall accuracy.

```
100%|██████████| 10000/10000 [04:17<00:00, 38.79it/s]Overall accuracy: 89.05999999999999
```

Figure 7. Overall accuracy

b. ELEC/COMP 546 ONLY (3 points). Engineer the caption prompts to try to obtain better accuracy. To do so, give a set of possible captions per class instead of just one. For example, “A bad photo of a [class]” or “A drawing of a [class]”. Report your accuracy.

```
100%|██████████| 10000/10000 [07:25<00:00, 22.43it/s]Overall accuracy: 88.41
```

Figure 8. Overall accuracy