Problem Description

People are often interested in knowing how they will look like in the future. Nowadays, using methods in deep learning, we can utilize technology to predict people's appearance in the future given a photo of them in their younger age. One useful application of this technology is fighting human trafficking and reuniting families. Having a way of predicting future faces given a childhood photo, it would be much easier to recognize a lost child. In this project, we will compare different models of Generative adversarial networks that accomplish this goal.

Related Works

Comparative analysis of CycleGAN and AttentionGAN on face aging application[1]
In this paper the author compared two frequently used image-to-image translation GANs which are CycleGAN (Cycle-Consistent Adversarial Network) and AttentionGAN (Attention-Guided Generative Adversarial Network). CycleGAN has two sets (G and F) of generators and discriminators, and is trained with a cycle-consistency loss which measures the difference between the input (p) and recovered input (F(G(p))). CycleGAN can convert an image from one domain to another without using paired images dataset. AttentionGAN identifies the foreground objects and minimizes the changes in the background. Using attention masks, content masks, and the generated output in one domain, AttentionGAN generates highly realistic images in another domain.
The comparison is quantitatively based on identity preservation, five image quality assessment metrics, while qualitatively based on a perceptual study on generated images, face aging signs, and robustness. The paper concluded that overall CycleGAN performs better than AttentionGAN.

Pivotal Tuning for Latent-based Editing of Real Images[2]
To leverage the generative power of a pre-trained StyleGAN in order to edit an image, one must project the image into the pre-trained generator's domain. However, the author claims that StyleGAN's latent space suffers from tradeoff between distortion and editability. The paper describes a novel technique called Pivotal Tuning Inversion (PTI) that alters the generator to map an out-of-domain image into an in-domain latent code; hence the editing quality of an in-domain latent region can be preserved while changing its portrayed identity and appearance.
The key difference between PTI and other mapping methods lies in the process of finding the latent representation; Performing strictly in the generator's latent space, GAN inversion results in distortion, while one current method that employs an extended latent space denoted as W+ suffers from weaker editability. This paper introduces a novel approach: instead of projecting the input image into the learned manifold, the team first inverts the input image to an editable latent code, and then performs Pivotal Tuning to tune the pre-trained StyleGAN to keep the image's editing qualities even after the generator is slightly modified. As demonstrated in the paper, the pivotal tuning is a local operation in the latent space, shifting the identity of the pivotal region to the desired one with minimal compensation. Since editing requires latent space traversal to find meaningful patterns, this novel method extends the high-quality editing capabilities to images out of its distribution, which demonstrates its superiority when handling challenging inputs such as delicate hairstyle or heavy make-up.
To prepare for edition, this paper proposes finding the closest editable point within the generator's domain, which will be pulled toward the target. The following procedure includes two steps. The first step is to invert the given input to W_p in the native latent space of StypeGAN; the second step is to apply Pivotal Tuning on this pivot code W_p to tune the pre-trained model to produce desired image by simply augmenting appearance-related weights, without affecting the well-behaved structure of StyleGAN's latent space.
The evaluation metrics utilized to test reconstruction quality are pixel-wise distance using MSE, perceptual similarity using LPIPS, and MS-SSIM, etc. To test editing quality, Microsoft Face API and a pre-trained facial recognition network are used to report identity preservation. Overall, PTI achieves the best score even on the task of reconstructing fine-details such as the make-up, lighting, and wrinkles, etc. However, this newly proposed method demonstrates increased quality at the cost of additional computation. The author gives candidate solution that a set of photographs of the individual will be used for PTI to stabilize the notion of personalization of target individual, compared to seeing just a single example.

StyleGAN2 Distillation for Feed-forward Image Manipulation[3]

This paper is based on that StyleGAN2 has a decoupling feature to do image editing. For editing a real image, the process of embedding into the Latent Space of StyleGAN2 is very slow. So in this paper, the author uses pair to get an image-to-image editing method, which doesn't need latent code editing. The method of research is to generate images with StyleGAN, then label them, then get the centers of different classes, and calculate the change vector between the centers of different classes.

Then use the function of StyleGAN decoupled editing, generate a bunch of data by moving to the positive change direction and negative change direction, and then filter to get pair data. After having the training set, the authors trained pix2pix supervised network and compared it with StarGAN and MUNIT, which are unsupervised transformation methods, and found good results and the lowest FID values. The authors also compared it with the Latent-Based method and found that the image-to-image conversion works well. One of the examples, face aging, relevant to our project yielded good results. The authors also point out the limitations of his model. One is the decoupling of StyleGAN2 is not very thorough, although the training set is sieved, actually, it's not very good, and the other is to improve or change pix2pix. Overall, the algorithm is still good.

Proposed Work

By using original face images from CelebA-HQ and FFHQ datasets, use four different kinds of GAN to do the training and generate the original face images in other stage of age and compare the results with the true face images of that age. Here we use CycleGAN, AttentionGAN, StyleGAN, and StyleGAn to do the training for face aging, and compare their performance using different kinds of metrics to assess the quality of samples of GANs, along with outputting the generated face images and evaluting them by ourselves.

Evaluation Metric

For image quality assessment, first of all, we use Frechet Inception Distance (FID) to do the evaluation. It tells how real the generated face images are in comparison to real face images. The lower value represents the better quality of the image. Furthermore, Peak Signal-to-Noise Ratio (PSNR) and Mean-Square Error (MSE)is introduced here to evaluate our results. The MSE represents the cumulative squared error between the compressed and the original image, whereas PSNR represents a measure of the peak error. The lower the value of MSE, the lower the error. Here is the formula:

$$PSNR = 10 \times \log_{10}\left(\frac{max^2}{MSE}\right)$$

Besides, we will print out some images we transformed and evaluate them by ourselves to see if the aging process success or not.[4]

Reference

[1] https://link.springer.com/article/10.1007/s12046-022-01807-4Sec8

[2] https://arxiv.org/pdf/2106.05744.pdf

[3] https://arxiv.org/pdf/2003.03581v2.pdf

[4] https://www.mathworks.com/help/vision/ref/psnr.html