Jagannath University Journal of Science
Volume 08, Number II, June 2022, pp. 55–66
https://jnu.ac.bd/journal/portal/archives/science.jsp
ISSN 2224-1698

JnUJSci

Image Manipulation Analysis Using StyleFace Generative Adversarial Network

Research Article

Md. Aslam Hossain, Sudipta Kumar Mondol and Md. Abu Layek*
Department of Computer Science and Engineering, Jagannath University, Dhaka-1100, Bangladesh

Received: 20 March 2022
Accepted: 01 May 2022

Abstract: There are many implementations of Generative Adversarial Networks (GANs), one of the key research topics in Artificial Intelligence nowadays. Image manipulation and transformation enhance a sample image's visual appearance by modifying different parameters. StyleFaceGANs are applied to manipulate and transform images, which has a significant impact on many applications dealing with images such as gaming, video conferencing, image-sample generation for research etc. We applied different techniques like semantic diffusion, single attribute manipulation, multiple attributes manipulation, and image interpolation. Applying these image generations, we analyzed the correlation matrix, distance measurement, and (PSNR) value of single and multiple attributes manipulations. We scrutinized StyleFaceGANs to manipulate the face using multiple attributes manipulation parameters like old and young age, eyeglasses and without eyeglasses, different poses, male and female gender, happy and sad expressions. We have achieved a maximum correlation value of 0.9716 when expression and multiple attribute manipulation of the same person take into account.We have gained a maximum value of 0.76 for distance measurement and a maximum PSNR value of 26.41 using the (VGG-FACE) model.

Keywords: *Face generation • Semantic Diffusion • Semantic Manipulation • Image Interpolation • Single and Multiple attributes manipulation • Structural Similarity Index*

1. Introduction

Multiple fields like image generation, video generation (Thies et al., 2016; Wang et al., 2018), and voice generation have addressed Generative Adversarial Networks for many years. For example, generation of image datasets, generation of photos of human countenances (Liu et al., 2020), generation of naturalistic portraits, creation of cartoon characters (Zhang et al., 2017), image-to-image translation (Choi et al., 2018) (Isola et al., 2017) (Richardson et al., 2021; Zhu et al., 2017), text-to-image translation (Gorti & Ma, 2018), Etc are accomplished by using GANs. We aim to introduce GANs in our research field and hope to gain currency in no time. GANs could also come in handy in gaming. In multiplayer games, the designer has many characters to design, and each of the characters has to be unique. It is challenging to make the character's looks, styles, Etc. StyleFaceGAN in image manipulation and transformation plays a significant role. (Liu et al., 2020; Shen et al., 2020) proposed that the concept of single and multiple facial attributes is manipulated employing the pre-trained (PGGAN) model.

The idea is to integrate a searching mechanism and edit the latent variable to manipulate facial characteristics. Face manipulation becomes more challenging when images with face masks or one side cover face come in the count. We scrutinized the StyleFace Generative Adversarial Network (StyleFaceGAN) and performed face manipulation using single and multiple attributes. This

* Corresponding Author: Md. Abu Layek
Email: layek@cse.jnu.ac.bd

method accomplishes the GANs inversion (Xia et al., 2021; Zhu et al., 2020) for source and target images before face manipulation and interpolation. We investigated and approximated various models like PGGAN (Liu et al., 2020), StyleGAN (Richardson et al., 2021), (LB-GAN), (CT- GAN), (M-GAN) (Bi et al., 2017), VGG-FACE (Zhang et al., 2018) and tabled the correlation matrix (Bau et al., 2019), verification, validation, distance measurement, and PSNR value for single and multiple attribute manipulations to generate images. In this work, we used StyleFaceGAN for facial image generation. Image manipulation and transformation exploit StyleFace Generative adversarial networks to achieve enhanced results. We have investigated three basic operations using GANs: interpolation, manipulation, and diffusion. The image emending method is based on StyleFace Generative Adversarial Network, a generative model synthesizing human faces. We implemented the GANs models and investigated to generate data distribution from the training dataset and random uniform distribution via Generative Adversarial Networks.

Generative adversarial networks simultaneously train two models. One is Generator, and another is Discriminator that is pitting against each other, called adversarial. The Generator generates a model G to capture the distribution of sample data and generate a sample similar to actual training data, with noise z obeying a particular distribution. Discriminator model D captures the probability between randomly generated data and the actual data distribution. The image emending method performs three different approaches to generate data distribution by using a generative adversarial network GANs. We call them diffusion, semantic manipulation, and interpolation. Firstly, the diffusion approach changes targeted properties to generate new data distribution. Moreover, by utilizing Semantic Manipulation, certain properties of sample distribution are controlled (Vinker et al., 2021; Zhu et al., 2016). Finally, in terms of Interpolation (Pinkney & Adler, 2020), this system can generate various new intermediary data distributions from sample distribution. The main contributions of this disquisition are described as follows. The main contributions of our research are to scrutinize, investigate and compare single attribute manipulation, multiple attributes manipulation, correlation measurement, distance measurement, finding challenges. Theexperimental GANs systems we implemented can discriminate between actual and fake images that are diffused, manipulated, and interpolated by combining different loss functions. Single and multiple attribute manipulations are shown in (Shen et al., 2020) and correlation measurement between the reconstructed image and its actual value in (Bau et al., 2019). We measured and compared a correlation matrix, single and multiple attributes for various image manipulation attributes. In this experiment, we have used the Flickr- Faces-HQ Dataset (Karras et al., 2019) (Nvidia) Resized 256px (ffhq256) dataset. ffhq256 dataset consists of 70,000 human countenance images. Flickr-Faces-HQ FFHQ is a high-quality image dataset of human countenance, created as a benchmark for generative adversarial networks GANs. We used images of 256×256 resolution and a 1:1 aspect ratio here. It has a good quality of fittings like eyeglasses, hats, sunglasses, and others. Also, it has a good amount of different aged,posed quality images and images of all gender. This paper is organized as follows. Different strategies and theories are described in section 1.

Figure 1. Manipulating various facial properties using StyleFaceGAN

Section 2 explains the techniques of the image manipulation and transformation study. Section 3 describes the evaluation and comparison of different models for face generation. Finally, section 4 describes the conclusion and limitations of our work.

1 Background Study

In this section, we have briefly inspected the related strategy of our research work, such as the correlation metrics, validation, structural similarity, and distance. GANs can also play an important role in animation movies, where animated images can be generated quickly using GANs. Moreover, in high-resolution videos like 4k and 8k, we can use GANs to increase the resolution of a low-quality image and thus maintain the resolution. Thus, using GANs, our still nascent gaming, animation, and video-making industries can flourish rapidly. These

applications of GANs have gained global popularity and significance.

1.1 Related Work

GANs can generate highly visually appealing images,

images of human faces, combine an image with the style of another image, conversion of male faces to female faces, developed from the sample images generated from the generator, which is hardly distinguishable. Face manipulation is editing and modifying human countenance images to drive them more appropriate for personal aesthetics or medical applications. These conversion types in a Generative Adversarial Network can generate manipulated human countenance images from the sample images. GANs are significantly used to render the face, obtaining promising results in the output samples. Manipulating a portrait is accomplished by operating some semantic parameters. Face aging, eyeglasses, gender, pose, and expression are mainly used. Liu, et al presented a new face attribute editing strategy

Table 1.1. PSNR Value for different types of GANs in CELEBA and FFHQ datasets

PSNR					
Ref.	**Attributes**	**Age**	**Smile**	**Eyeglasses**	**Gender**
(Huang etal., 2022)	**STGAN**	34.72	-	33.23	-
(Huang etal., 2022)	**StyleGAN**	20.71	22.68	-	-
(Huang etal., 2022)	**StarGAN**	26.38	-	-	26.71
(Huang etal., 2022)	**AttGAN**	-	29.59	28.19	-

(Liu et al., 2020) that operates the pre-trained PGGAN model and incorporates exploring and reworking the latent variable to achieve countenance characteristic editing.

H. Zhang et al. wanted to produce high-resolution text to photorealistic images (Zhang et al., 2017), and they modeled their GANs with an architecture similar to StackGAN. Y. Choi et al. proposed multi-domain image to image translation (Choi et al., 2018) with an architecture similar to Star GANs. K Gorti et al. proposed that Text-to-image-to-text translation using cycle consistent generative networks can generate realistic images from the given text (Gorti & Ma, 2018). To learn multifarious realistic images, J.-Y. Zhu et al. presented the generative adversarial neural network (Zhu et al., 2016). To ensure the output is kept the learned multifarious, they used a neural network to constrain various image manipulation operations. They have shown several editing operations, precisely color and shape manipulations. For instance, color adjustment, image blending, or image reshuffling are shown in (Song et al., 2009). The low-level features of an image can be modified using these editing tools. Yang et al. (Yang & Lim, 2020) considered Face2Face (Thies et al., 2016), encoder-decoder DeepFake methods, and accomplished face attribute manipulation. They have compared different models like StyleGAN, proGAN, WGAN − GP and measured the classification accuracy against each other, and accuracy 99.6 and 99.4 percentages are obtained, respectively. Pinkney et al. simply applying linear interpolation (Pinkney & Adler, 2020). StyleGAN can generate intermediary images through interpolation, which has the original properties but still matches neither. Also, it can be seen that the resultant images are pretty appealing to the visuals. Tolosana Ruben et al. specifically proposed total face synthesis and attribute manipulation (Tolosana et al., 2020). Zaeemzadeh et al. introduced a new face image recovery model that uses an adjustment vector and a preference vector for input countenance (Zaeemzadeh et al., 2021). The input countenance is modified using an adjustment vector.

Various priority levels are set to various facial features using a preference vector. The preference vector and an adjustment vector operate jointly. The methodology of sample inversion is followed by StyleGAN. Richardson et al. introduced the pixel2style2pixel framework (Richardson et al., 2021) to propose a solution to handling multiple image-to-image translation tasks (Choi et al., 2018; Zhu et al., 2017) in case the StyleGAN domain does not represent the input portrait. Through resampling of styles, the model is compatible with the multimodal synthesis. The model can calculate quantitative results for image reconstructions such as similarity 0.56, (LPIPS) 0.17, (MSE) 0.03. Bi, Lei et al. proposed a new method to produce synthetic PET images (Bi et al., 2017) employing a multichannel generative adversarial network M-GAN and compared M-GAN to single-channel variants, the LB-GAN and the CT-GAN. They used mean absolute error (MAE) and peak signal-to-noise ratio PSNR to evaluate the different methods. After comparing different generative adversarial network approaches, they showed the highest MAE value of 7.98, the lowest PSNR value 24.25 for LB-GAN, MAE 4.77, and PSNR 26.65 for CT-GAN finally got the best performance for M-GAN with the lowest MAE 4.60 and highest PSNR 28.06. Phillip Isola et al. have shown the potential of conditional adversarial networks (Isola et al., 2017) to overcome the difficulty of image-to-image translation. These networks comprehend how an input portrait can be mapped to an output portrait and find a loss function to train this mapping. Liu et al. have shown conditional convolutional networks (Liu et al., 2019) as a general solution to predict layout-to-image translation problems for semantic image synthesis. These networks take a low-resolution noise map as input and generate output using conditional convolution blocks and up sampling layers to refine the intermediary feature maps gradually. Quantitative results are calculated on multiple

datasets determining (MIOU) and FID scores and finally got the best performance with the highest MIOU and lowest FID scores. Starganv2 is presented in (Choi et al., 2020) and is experimented with composite image synthesis results on different datasets. Quantitative results are calculated on latent- guided synthesis determining FID, and LPIPS and finally got the best performancewith the FID 13.8 and LPIPS 0.453 for Celeba − HQ dataset and FID 16.3 and LPIPS 0.451 for (AFHQ) dataset. (Shen et al., 2020) presented the correlation matrix of attribute boundaries with different manipulated attributes like pose, eyeglass, smile, age, and gender. They have shown different manipulation attributes correlation, with a maximum value of 0.52 when eyeglasses and gender take into account. (Rai et al.,2021) presented using StyleGAN the distance measurement of single attribute manipulation with the range of -3 to 3 and has shown correlation measurement for various attributes: beard, pointy nose, chubby, race, hairstyle, and face shape and obtained the best correlation value 0.56 for chubby face against age attribute manipulation. Table 1.1 shows different PSNR values of multifarious GANs. We have shown single attribute manipulation where CELEBA for fake images and FFHQ datasets for real images are used.

2. Methodology

We proposed the StyleFaceGAN by which various manipulated faces are generated from the original image.

Figure 2.1. *StyleFaceGAN* Architecture

Figure 2.1 illustrates the StyleFaceGAN architecture. Using StyleFaceGAN, first, face detection on the input is done in image generation. Only facial properties are captured and extracted when the input image generates fake samples. After that, image inversion is completed to manipulate, diffuse, and interpolate facial properties like age, eyeglasses, expression, gender, and pose.

2.1 Generative Adversarial Networks

In Generative Adversarial Networks, GANs consist of Generator G and Discriminator D. The Generator generates a sample from a latent space distribution to the data distribution. The Discriminator determines the samples which belong to the data distribution or not. The two-player game inspires GANs in game theory. Based on game-theoretic min-max principles (Goodfellow et al., 2014), the Generator and the Discriminator are generally learned together by alternating the training of the Discriminator and Generator. The Generator generates a model G to catch the distribution of sample data and generate a sample similar to actual training data, with latent space z regarding a particular distribution. The new distribution of images generated by the Generator is visually appealing and realistic to humans (Salimans et al., 2016).

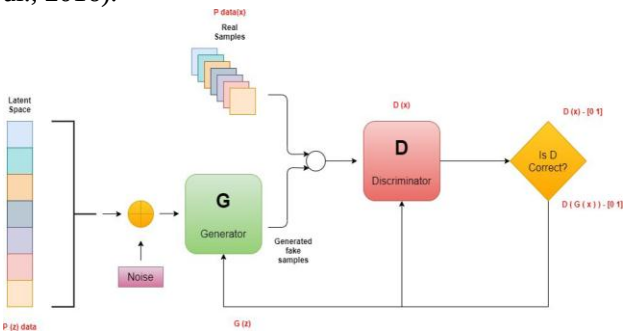

Figure 2.2. *GANs* Architecture

The Generator's objective is to create fake samples as close as possible to the actual samples. The objectives of the Discriminator are to discriminate between generated fake samples from the generator and actual samples from the datasets (Goodfellow et al., 2014). The objective is to produce fake samples close to the actual samples so that the Discriminator becomes a fool to discriminate between the actual and fake samples and produces an output that is neither 0 nor 1 for both of these samples. Each of these is held constant while training the other. Equation 2.1 describes the minimization and maximization ofGenerator and Discriminator, respectively.

$$\min_G \max_D E_{x\sim p_{data}(x)}[\log(1-D(x,G(x))] + E_{x\sim p_x(x)}[\log(D(x,y))] \quad (2.1)$$

This GANs aims to approximate this distribution as close as possible to dataset (D). D(x) is the discriminator's estimate of the probability that actual data instance x is real. G(z) is the Generator's output when given noise z. D(G(z)) is the discriminator's estimate of the probability that a fake instance is actual. The Generator is a primary simple deep neural network. It is taking a random distribution Z which is entirely random. z~Z, here z is a random variable, and Z is a probability distribution function. This distribution feeds into the Generator and produces modified random variables G(z). Then modified random variables are fed into the discriminative model. Then the discriminative model gets two images from random distribution (D) and another from G(z). The label for D is y (actual), and the label for G(z) is y (fake). It is called adversarial because G and D are fighting eachother to achieve this accuracy.

2.2 Loss Function

GANs try to imitate a probability distribution. Using loss functions GANs can measure the distance between generated and actual samples. Generator and discriminator have different loss functions. In both scenarios, however, the generator can only affect one term in the distance measure: the term that deliberates the

distribution of the fake data. So during generator training, we lower the other term, which deliberates the actual data distribution.

$$L(\hat{y}, y) = [ylog\hat{y} + (1-y)log(1-\hat{y})] \quad (2.2)$$

$\hat{y}$ is the reconstructed label and y is the original label. The label for the data coming from Pdata(x) is y=1, $\hat{y}$=D(x), so we obtain,

$$L(D(x), 1) = 1 * log(D(x)) = log(D(x)) \quad (2.3)$$

For data coming from generator the label is y=0 $\hat{y}$=D(G(z)), so in that case,

$$L(D(G(z)), 0) = [0 + (1-0)log(1 - D(G(z)))] = log(1 - D(G(z))) \quad (2.4)$$

The objective of the Discriminator is to classify fake versus real datasets correctly. For this, 2.3 to 2.4 should be maximized. The Equation for Discriminator:

$$Discriminator = \max_{D}[\log D(x) + \log(1-D(G(z)))] \quad (2.5)$$

The objective of the generator is to fool the discriminator by producing the probability one. That means, D(G(z)) is equal to one. For this, 2.3 to 2.4 should be minimized.

The Equation for Generator:

$$Generator = \min_{G}[\log D(x) + \log(1-D(G(z)))] \quad (2.6)$$

Only for one sample, the equation of GANs:

$$\min_{G}\max_{D} V(D,G) = [\log D(x) + \log(1-D(G(z)))] \quad (2.7)$$

For all samples, the equation of GANs:

$$\min_{G}\max_{D} V(D,\ G) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))] \quad (2.8)$$

3. Experimental results analysis

3.1 Semantic Diffusion

Semantic diffusion is employed to diffuse a distinct part of the planned image into the context of another image. The combined result must preserve the characteristics of the mark image, not the original one, and simultaneously adapt the context information. Figure 3.2 illustrates the diffused mark countenance. The first rows of images illustrate the sample images. The middle part of these images faces is diffused into the target image's context and produces the unique images in the final row. The images seem reasonably natural and do not contain many artifacts. It can be seen that the outcomes retain the identicalness of the mark image and incorporate it into the surroundings reasonably well. Unlike the style mixing process, the central portion of the outcome image is preserved identically as the mark image.

Figure 3.1. Before Diffusion

Figure 3.2. After Diffusion

3.2 Semantic Manipulation

Image Manipulation refers to bringing changes to a digitized image to transform it into the desired image. Semantic Image manipulation helps to test the alignment of the embedded latent codes with the semantic knowledge learned by GANs. Any number of effects is possible to achieve through image manipulation. One of the most common applications in GANs is face manipulation. A human face can be aged or made young. Gender can be changed. Facial expression can be altered. The pattern of eyeglasses or smiling faces can be modified, and so on. The outcome images are generated by applying various manipulation attributes such as age, eyeglasses, gender, pose, and expression, shown in figure 3.3. For example, in figure 3.4, a human male face has been changed to a female countenance. A previously unnoticeable ear can be seen in the resulting image by applying the pose parameter. Also, the facial expression has altered, which is evident from the broader smile. In the second row, a human male face has been changed to a broader male face with a smile and eyeglasses. Another male face has been changed to a more beard face.

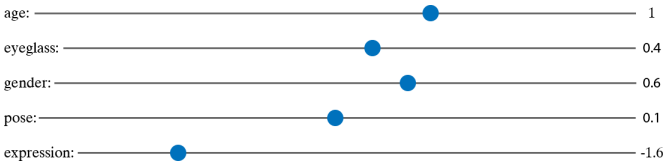

Figure 3.3. Manipulation Attributes

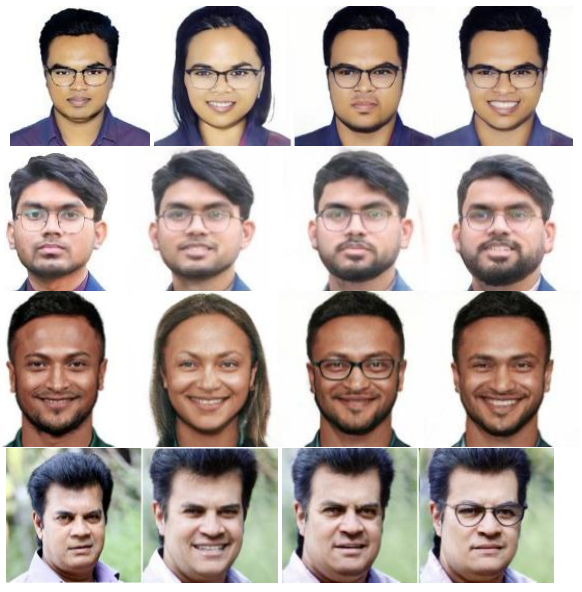

Figure 3.4. After Manipulation

Figure 3.5 shows the effect of the Interpolation of two images. The original images are on the leftmost and rightmost sides, respectively. We take some pixels from each image in each step of Interpolation and apply them to the other. Likewise, we have generated interpolated images from two actual sample images by operating five to seven interpolation steps shown in figure 3.5 respectively. Finally, a generated intermediary image is created, which has both the original's properties but still matches neither. Also, it can be seen that the resultant images are pretty appealing to the visuals.

In Figure 3.6, we experimented with image Interpolation using different characters to generate multiple intermediary images. In figure 3.6, the Red and orange lines indicate primarily distinguishable images from others. In the Red rectangle, we have shown interpolated intermediary images, which have the original properties from two actual images and aremore dissimilar than the orange rectangle intermediary images.

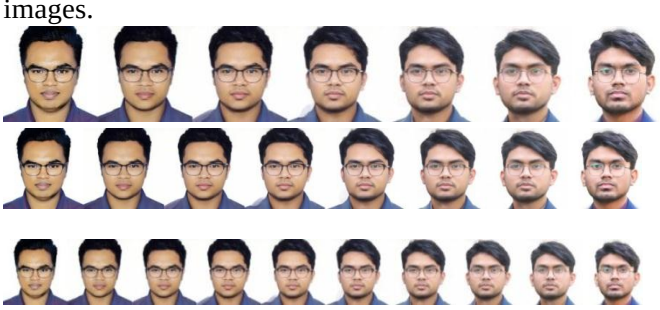

Figure 3.5: Interpolation in 5, 6, and 7 steps

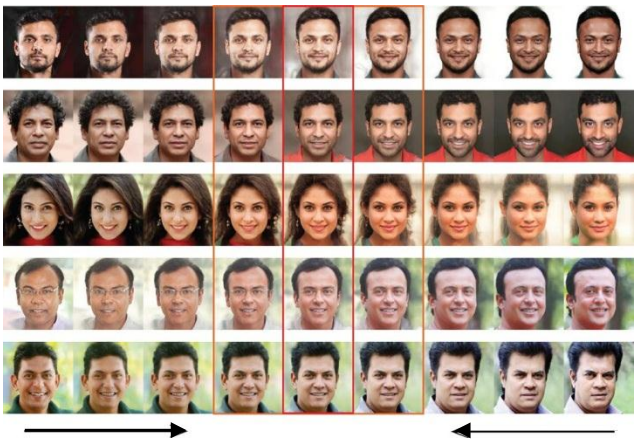

Figure 3.6. Image Interpolation

3.3 Image Interpolation

Image interpolation is the process of interpolating two images. Resizing or distortion of an image in pixel by pixel manner image interpolation is done. It is the process of finding out the unknown pixels of the image. In other words, Interpolation is the strategy of using available data to estimate values at unknown locations.

3.4 Single Attribute Manipulation

This section experimented with single attribute manipulation and compared with (Shen et al., 2020) and finally got these outcome images. We have used three iconic Bangladeshi characters to explore the

manipulation. A single attribute countenance generation using multiple characters, shown in figure 3.7. The input images are positioned at the second, fifth, and eighth. All manipulated images are arranged at both sides of theinput image. The first row depicts the age manipulation. The young and old attributes of aging are illustrated on both sides. The second row characterizes the manipulation of eyeglasses. The reduced eyeglass properties are presented on the left side, and the improved eyeglass properties of the actual nature are on the other side. Gender manipulation is positioned in the third row. In terms of gender manipulation, the left image shows the increasing male characteristics and an other- sided image indicating female characteristics. In the case of pose manipulation, the fourth row expresses as left posed on the left side and an other- sided image displaying right posed. The last row characterizes the manipulation of expression. Happy expressions and sad expressions are illustrated on both sides, respectively.

3.5 Multiple Attributes Manipulation

This section exemplifies the experiment of the multiple attributes manipulation in figure 3.8 and depicts the gradual improvement of different characteristics. The first portion shows the same person under gradually changed age, expression, and gender manipulation. Inthis way, these three manipulation properties are acclimated to the actual image, and finally, we reached the sequential illustration of multiple attributes to present a synthesized image where all other properties remain unchanged. The following image of another person corresponds to manipulating five different attributes. The last row of the second portion depicts the flow of old aging, happy expression, gender changing, left-sided pose, removing eyeglasses, and eventually, all five different attributes are acclimated to generate a synthesized image. In (Shen et al., 2020), conditional manipulation is shown to generate synthesized images where manipulating attributes like age, gender, eyeglasses are considered and got the result with remaining facial properties unchanged.

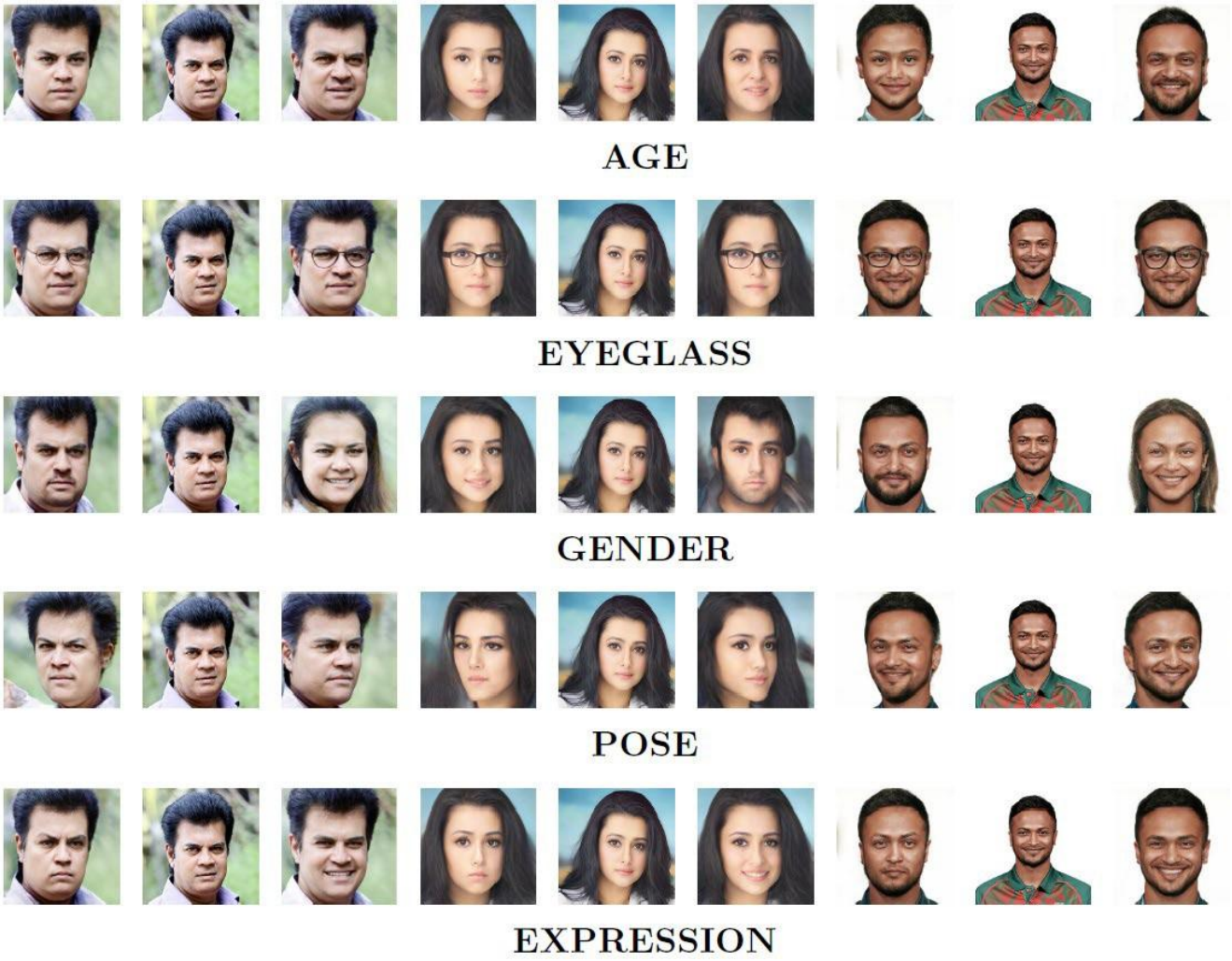

Figure 3.7. Single Attribute Manipulation

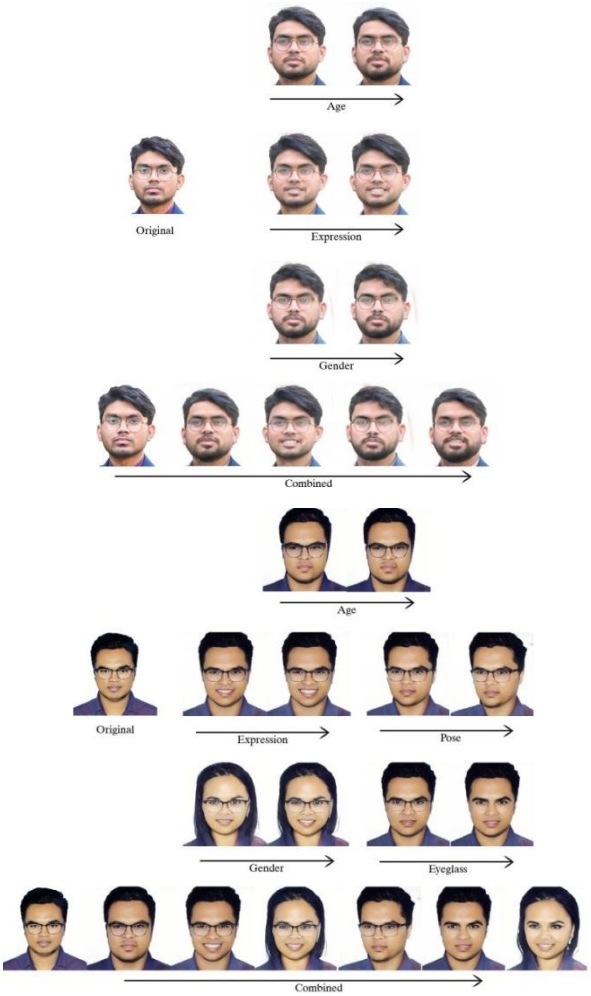

Figure 3.8. Multiple attribute manipulation

3.6 Correlation Matrix

The mixture of multiple manipulating facial attributes manipulated countenance is generated in figure 3.9. The mixture image has the combination of age, eyeglasses, and smiling attributes. Firstly, the original image is manipulated using age-decreasing attributes and looks much younger than that. Then at the second step, eyeglass is added to the newly generated image. After that, a smiling expression is added. Finally, we got a visually appealing image that combines those three attributes.

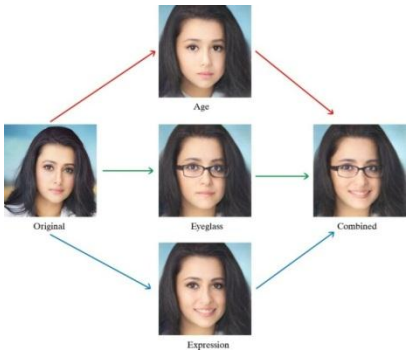

Figure 3.9. Combination of multiple manipulatedimages

Equation 3.1 derives the Pearson correlation coefficient.

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (3.1)$$

Table 3.1. Correlation between multiple manipulated face synthesis

	Original	Age	Expression	Eyeglasses	Combined
Original	**1.00**	**0.9388**	**0.9674**	**0.9291**	**0.9408**
Age	**-**	**1.00**	**0.9563**	**0.8953**	**0.9208**
Expression	**-**	**-**	**1.00**	**0.9332**	**0.9716**
Eyeglasses	**-**	**-**	**-**	**1.00**	**0.9515**
Combined	**-**	**-**	**-**	**-**	**1.00**

The correlation between the original and other manipulated images is shown in Table 3.1. Every diagonal represents the correlation between the same image, and the correlation matrix is one. Other than the diagonal cell, all manipulated images have the value of correlation matrix range 0.8953 to 0.9716, indicating the significant similarity between the original and other images. The value of the correlation matrix of the combined manipulated image is 0.9408 against the original image, as close as one and indicating a notable correlation against the original image. Figures 3.10. and 3.11. illustrate scatter plots and correlation, where x-label and y-label represent the original and combined images. In (Bau et al., 2019) have shown the scatter plots and correlation of reconstruction images of GANs generated images against its truth value and

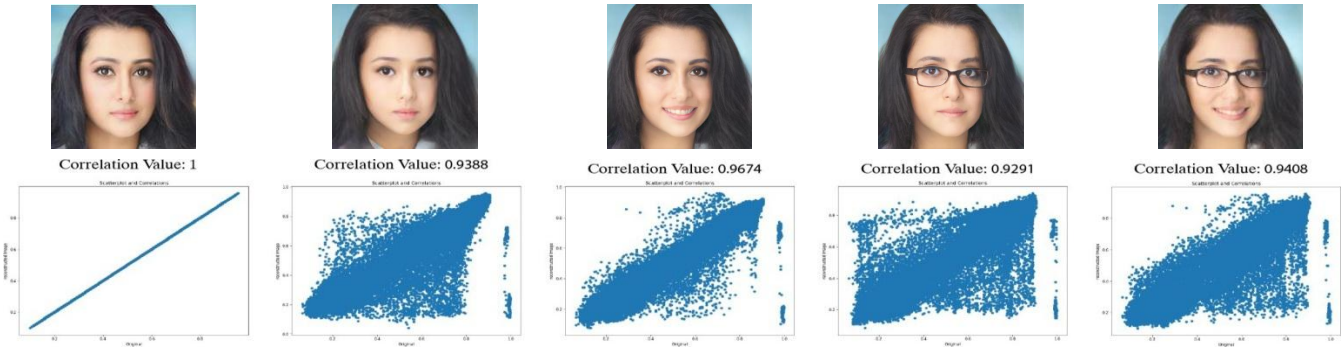

Figure 3.10. Correlation scatter plot between real and different face synthesis

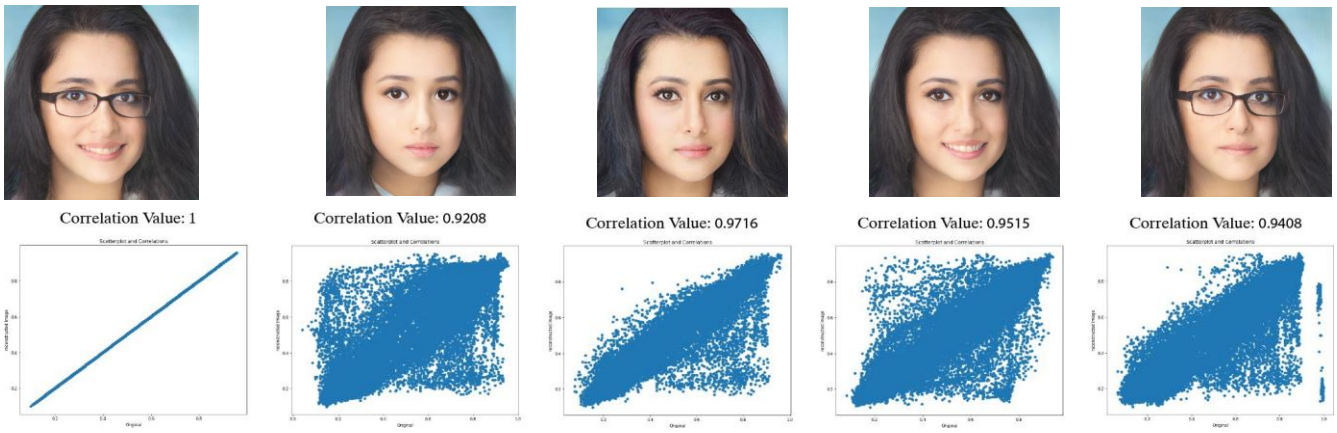

Figure 3.11. Correlation scatter plot between combined and different face synthesis

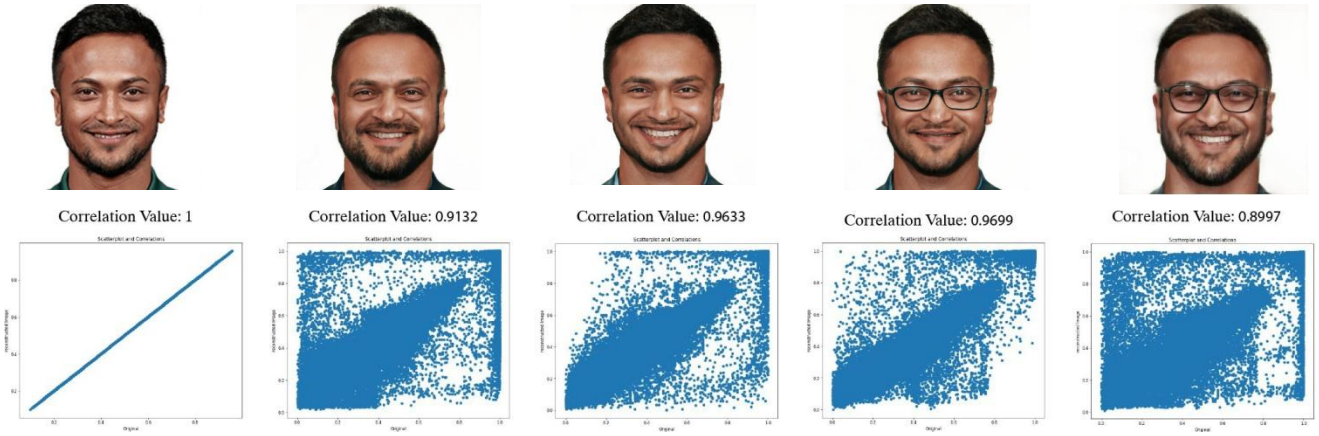

Figure 3.12. shows the correlation scatter plot between real and different attributes applying age, expression, eyeglasses, and combined image

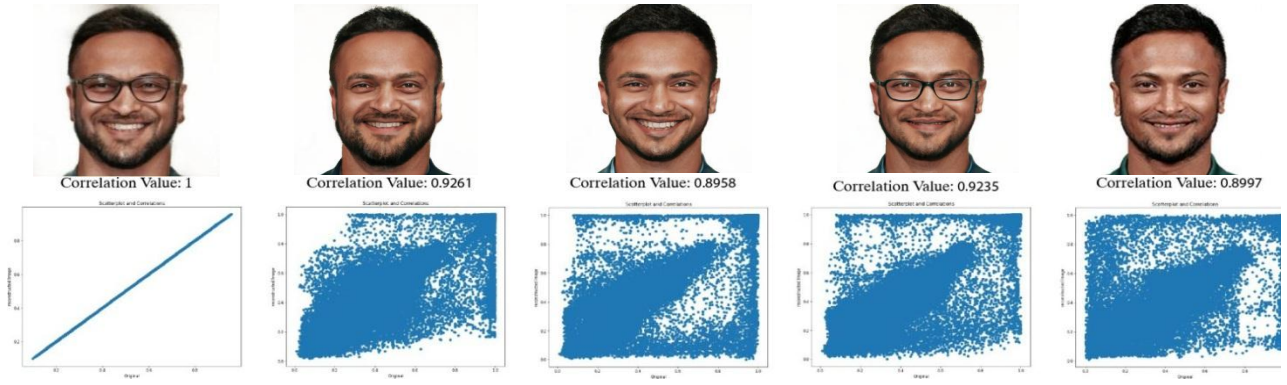

Figure 3.13. shows the correlation scatter plot between combined and different attributes applying age, expression, eyeglass, and real image

correlation value 0.9984 for the reconstructed layer4 features, 0.99939 for reconstructed pixel channels. The scatter plot and similarity of the original image against the age, happy expression, eyeglass, and the combined image is pictured in figure 3.10. In contrast, the scatter plot and similarity of the combined image against the age, happy expression, eyeglass, and the original image is pictured in figure 3.11. Eventually, we got a relative correlation value against those three manipulating attributes.

3.7 Distance Measurement

This section compares the distance, validation, and PSNR to other models and verifies the face manipulation. This section compares the distance, validation, and PSNR to other models and verifies the face manipulation. The original image is represented in the second row in figure 3.14. The first and third row of age, eyeglass, gender, pose, and expression attributes generating manipulated countenance represent the positive and negative distance.

The quantitative measurement for face manipulation using multiple manipulated attributes for the VGG-FACE model is shown in tables 3.2 and 3.3. The threshold value for distance measurement is 0.4. When the distance measurement result is more petite, indicating an excellent image. In terms of validity and peak error, for the VGG-FACE model, the average value is 76.41 and 26.41, respectively. Equation (3.2) and (3.3) (Serengil & Ozpinar, 2020) derive the cosine similarity and distance measurement.

$$cos\ \theta = \frac{\sum_{i=0}^{n} x_i y_i}{\sqrt{\sum_{i=0}^{n} x_{i^2}} \sqrt{\sum_{i=0}^{n} y_{i^2}}} \tag{3.2}$$

$$D_c(x, y) = 1 - cos\theta \tag{3.3}$$

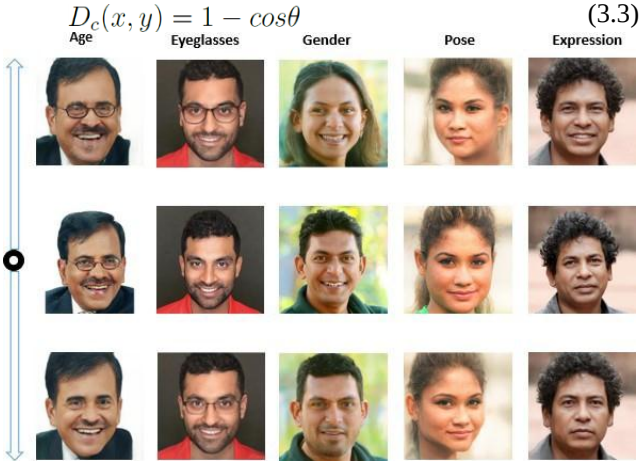

Figure 3.14. Distance measurement

More, we have measured the structural similarity and FID. SSIM indicates the structural similitude between actual and generated images using equation 3.4. To overcome the constraints of different conventional matrices, SSIM is used, where higher SSIM indicates a more appealing structural similarity index.

$$SSIM(x, y) = \frac{(2\sigma_x\sigma_y + k_1)(2\mu_{xy} + k_2)}{(\sigma_x^2 + \sigma_y^2 + k_1)(\mu_x^2 + \mu_y^2 + k_2)} \tag{3.4}$$

In the case of Frechet Inception Distance, FID measures the differentiation be- tween actual and generated images using equation 3.5, where low FID indicates the better appealing outcome.

$$FID(R,\ G) = \|\mu_r - \mu_g\|_2^2 + Tr(\Sigma_r + \Sigma_g - 2(\Sigma_r\Sigma_g)^{1/2}) \tag{3.5}$$

Table 3.2. Classification accuracy of distance, validation, PSNR, SSIM, and FID are shown for various methods with respect to different attributes. Positive (+) represents the manipulated image by counting the attributes, and negative (-) represents the manipulated image by removing the attributes.

Attrib-utes	Age (+)	Age (-)	Eyeglass-es (+)	Eyeglass -es (-)	Male	Fe-male
Distance	0.143	0.187	0.268	0.194	0.323	0.373
Valida-tion	85.7	81.3	73.2	80.6	67.7	62.7
PSNR	29.948	30.045	30.294	30.650	28.462	28.236
SSIM	0.609	0.606	0.720	0.683	0.555	0.477
FID	23.985	26.291	29.124	32.986	22.639	37.210

Table 3.3. Classification accuracy

Attributes	Pose(Left)	ose(Right)	Expres-sion(Happy)	Expres-sion(Sad)
Distance	0.224	0.302	0.197	0.148
Validation	77.6	69.8	80.3	85.2
PSNR	29.060	28.394	29.565	29.718
SSIM	0.594	0.433	0.600	0.639
FID	23.989	54.896	30.282	27.873

3.8 Comparison with different models

In the case of validity and PSNR comparison of various models is shown in Tables 3.4 and 3.5. When the similarity measurement results are significant, lower the error. In the case of PSNR results, the higher value indicates visually appealing images.

Table 3.4. Quantitative measurement for image manipulation

Method	Validation
PGGAN (Shen etal., 2020)	0.97
Karras et al. (Karras et al., 2020)	0.77
StyleFaceGAN	**0.76**

Table 3.5. Quantitative measure- ment for image manipulation (PSNR)

Model	PSNR
LB-GAN (Bi et al.,2017)	24.25
CT-GAN (Bi et al.,2017)	26.65
M-GAN (Bi et al.,2017)	28.06
StyleFaceGAN	**26.41**

Table 3.6. FID values for different types of GANs. lower FID indicates the better quality image.

FID							
Ref.	**Attrib-utes**	**Smile (+)**	**Smil e(-)**	**Eyegla ss-es(+)**	**Eye-glasses (-)**	**Mal e**	**Fe-male**
(Xiao et al., 2018)	Cy-cleGAN	23.23	22.74	36.87	48.82	60.25	46.25
Xiao et al., 2018)	StarGAN	51.36	78.87	70.40	142.35	70.14	206.21
(Xiao et al., 2018)	LEGAN T	25.71	24.88	47.35	60.71	59.37	56.80
	StyleFace GAN	30.28	27.87	**29.12**	**32.98**	**22.63**	**37.21**

Table 3.7. SSIM Value for different types of GANs in CELEBA and FFHQ datasets. The higher SSIM indicates the better quality image

		SSIM			
Ref	**Attrib-utes**	**Age**	**Expres-sion**	**Eyeglass-es**	**Gender**
(Huang et al., 2022)	STGAN	-	0.9086	0.8663	-
(Huang et al., 2022)	Style-GAN	0.6159	0.6625	-	-
(Huang et al., 2022)	StarGAN	0.7929	-	-	0.7952
(Huang et al., 2022)	AttGAN	-	0.8736	0.8207	-
	StyleFac eGAN	0.6075	0.6195	0.7015	0.5160

3.9 Challenges and limitations

This section shows some challenges in StyleFaceGAN. We investigated and found some challenges that face masks and one side-covered face cannot be manipulated by StyleFaceGAN. In Figures 3.15 and 3.16, countenance manipulation is restrained when one-side covered face and face masks are taken into account.

Figure 3.15. One side covered face

Figure 3.16. Face mask

as a noise function in image interpolations. These various noise functions are applied with the actual sample images to fool the discriminator. Analyzing single and multiple attribute manipulations, we observed that generated images are nearly close to the actual image. Individual attributes to image manipulation are applied in single attribute manipulation, and more than two or three

4. Conclusion and future work

The main contributions of our research are to scrutinize, investigate and compare single attribute manipulation, multiple attributes manipulation, correlation measurement, distance measurement, finding challenges. The experimental GANs systems we implemented can discriminate between actual and fake images that are diffused, manipulated, and interpolated by combining different loss functions. Initially, a generator can generate the fake one, and the dataset contains the actual one. The discriminator can reduce the gap between the fake and the actual one by improving the loss function and de-creasing discrimination. Then the generated loss function is applied to newly generated fake samples, the discriminator fools to discriminate. In diffusion, a particular targeted property of an actual sample is used asa noise function. In the case of manipulation, different characteristics of human faces like age, eyeglass, gender, pose, and expression are used as a noise function. Various properties of two actual sample images are used attributes are adjusted to generate manipulation inmultiple attribute manipulation. After that, we measured the correlation matrix and examined all generated imagesnear value one. Verification, validation, and peak error are also examined for the VGG-FACE model (Zhang et al., 2018) using distance measurement compared to the other model. Finally, we found some challenges like image manipulation where face masks presence, one side covered face. Generated fake sample images may be used in game development, animated movies, and many more domains where privacy preservation is required. We want to investigate another efficient encoder to manipulate the challenges we discussed.

Acknowledgements

This work was supported by the Information Technology Research and Resource Center (ITRRC, web: http://itrrc.com), and the JnU research grant (জীব/ গবেষণা/গপ্র/২০২০-২০২১/বিজ্ঞান/৩৩) Jagannath University, Dhaka, Bangladesh. Dr. Md. Abu Layek (layek@cse.jnu.ac.bd) is the corresponding author.

References

Bau, D., Zhu, J.-Y., Wulff, J., Peebles, W., Strobelt, H., Zhou, B., & Torralba, A. (2019). Seeing what a gan cannot generate. Proceedings of the IEEE/CVF International Conference on Computer Vision,

Bi, L., Kim, J., Kumar, A., Feng, D., & Fulham, M. (2017). Synthesis of positron emission tomography (PET) images via multi-channel generative adversarial networks (GANs). In *molecular imaging, reconstruction and analysis of moving body organs, and stroke imaging and treatment* (pp. 43-51). Springer.

Choi, Y., Choi, M., Kim, M., Ha, J.-W., Kim, S., & Choo, J. (2018). Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. Proceedings of the IEEE conference on computer vision and pattern recognition,

Choi, Y., Uh, Y., Yoo, J., & Ha, J.-W. (2020). Stargan v2: Diverse image synthesis for multiple domains. Proceedings of the IEEE/CVF conference on computer vision and patternrecognition,

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde- Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, *27*.

Gorti, S. K., & Ma, J. (2018). Text-to-image-to-text translation using cycle consistent adversarial networks. *arXiv preprint arXiv:1808.04538*.

Huang, Y., Juefei-Xu, F., Guo, Q., Liu, Y., & Pu, G. (2022). FakeLocator: Robust localization of GAN-based face manipulations. *IEEE Transactions on Information Forensics and Security*.

Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2017). Image-to- image translation with conditional adversarial networks. Proceedings of the IEEE conference on computer vision and pattern recognition,

Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition,

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., & Aila,

T. (2020). Analyzing and improving the image quality of stylegan. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition,

Liu, S., Li, D., Cao, T., Sun, Y., Hu, Y., & Ji, J. (2020). GAN-based face attribute editing. *IEEE Access*, *8*, 34854-34867.

Liu, X., Yin, G., Shao, J., & Wang, X. (2019). Learning to predict layout-to-image conditional convolutions for semantic image synthesis. *Advances in Neural Information Processing Systems*, *32*.

Pinkney, J. N., & Adler, D. (2020). Resolution dependent gan interpolation for controllable image synthesis between domains. *arXiv preprint arXiv:2010.05334*.

Rai, A., Ducher, C., & Cooperstock, J. R. (2021). Improved Attribute Manipulation in the Latent Space of StyleGAN for Semantic Face Editing. 2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA),

Richardson, E., Alaluf, Y., Patashnik, O., Nitzan, Y., Azar, Y., Shapiro, S., & Cohen-Or, D. (2021). Encoding in style: a stylegan encoder for image-to-image translation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,

Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., & Chen, X. (2016). Improved techniques for training gans. *Advances in Neural Information Processing Systems*, *29*.

Serengil, S. I., & Ozpinar, A. (2020). Lightface: A hybrid deep face recognition framework. 2020 Innovations in Intelligent Systems and Applications Conference (ASYU),

Shen, Y., Gu, J., Tang, X., & Zhou, B. (2020). Interpreting the latent space of gans for semantic face editing. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,

Song, Y., Tong, X., Pellacini, F., & Peers, P. (2009). Subedit: a representation for editing measured heterogeneous subsurface scattering. *ACM Transactions on Graphics (TOG)*, *28*(3), 1-10.

Thies, J., Zollhofer, M., Stamminger, M., Theobalt, C., & Nießner, M. (2016). Face2face: Real-time face capture and reenactment of rgb videos. Proceedings of the IEEE conference on computer vision and pattern recognition,

Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). Deepfakes and beyond: A survey of face manipulation and fake detection. *Information Fusion*, *64*, 131-148.

Vinker, Y., Horwitz, E., Zabari, N., & Hoshen, Y. (2021). Image Shape Manipulation from a Single Augmented Training Sample. Proceedings of the IEEE/CVF International Conference on Computer Vision,

Wang, T.-C., Liu, M.-Y., Zhu, J.-Y., Liu, G., Tao, A., Kautz, J., & Catanzaro, B. (2018). Video-to-video synthesis. *arXiv preprint arXiv:1808.06601*.

Xia, W., Zhang, Y., Yang, Y., Xue, J.-H., Zhou, B., & Yang, M.-H. (2021). GAN inversion: A survey. *arXiv preprint arXiv:2101.05278*.

Xiao, T., Hong, J., & Ma, J. (2018). Elegant: Exchanging latent encodings with gan for transferring multiple face attributes. Proceedings of the European conference on computer vision (ECCV),

Yang, C., & Lim, S.-N. (2020). One-shot domain adaptation for face generation. Proceedings of the IEEE/CVF conferenceon computer vision and pattern recognition,

Zaeemzadeh, A., Ghadar, S., Faieta, B., Lin, Z., Rahnavard, N., Shah, M., & Kalarot, R. (2021). Face Image Retrieval With Attribute Manipulation. Proceedings of the IEEE/CVF International Conference on Computer Vision,

Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X., Huang, X., & Metaxas, D. N. (2017). Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. Proceedings of the IEEE international conference on computer vision,

Zhang, T., Wiliem, A., Yang, S., & Lovell, B. (2018). TV-GAN: Generative adversarial network based thermal to visible face recognition. 2018 international conference on biometrics (ICB),

Zhu, J.-Y., Krähenbühl, P., Shechtman, E., & Efros, A. A. (2016). Generative visual manipulation on the natural image manifold. European conference on computer vision,

Zhu, J.-Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. Proceedings of the IEEE international conference on computer vision,

Zhu, J., Shen, Y., Zhao, D., & Zhou, B. (2020). In-domain gan inversion for real image editing. European conference on computer vision,