Can GAN-induced Attribute Manipulations Impact Face Recognition?

by

Sudipto Banerjee, Aditi Aggarwal, Arun Ross

in

*International Conference on Pattern Recognition*

: 1
-15

Report No: IIIT/TR/2022/-1

Centre for Others
International Institute of Information Technology
Hyderabad - 500 032, INDIA
August 2022
Can GAN-induced Attribute Manipulations Impact Face Recognition?

Sudipta Banerjee\textsuperscript{1}[0000−0003−1623−7697], Aditi Aggarwal\textsuperscript{1}, and Arun Ross\textsuperscript{2}[0000−0001−8850−3013]

\textsuperscript{1} International Institute of Information Technology, Hyderabad, India
sudipta.b@iiit.ac.in, aditi.aggarwal@students.iiit.ac.in
\textsuperscript{2} Michigan State University, East Lansing, USA
rossarun@cse.msu.edu

Abstract. Impact due to demographic factors such as age, sex, race, etc., has been studied extensively in automated face recognition systems. However, the impact of digitally modified demographic and facial attributes on face recognition is relatively under-explored. In this work, we study the effect of attribute manipulations induced via generative adversarial networks (GANs) on face recognition performance. We conduct experiments on the CelebA dataset by intentionally modifying thirteen attributes using AttGAN and STGAN and evaluating their impact on two deep learning-based face verification methods, ArcFace and VGGFace. Our findings indicate that some attribute manipulations involving eyeglasses and digital alteration of sex cues can significantly impair face recognition by up to 73\% and need further analysis.

Keywords: Face recognition \cdot Generative adversarial network (GAN) \cdot Attribute manipulation.

1 Introduction

Demographic attributes (race, age and sex) \textsuperscript{[14]}\textsuperscript{[14]}, face and hair accessories or attributes (glasses, makeup, hairstyle, beard and hair color) \textsuperscript{[28]}, and data acquisition factors (environment and sensors) \textsuperscript{[24],[22]} play an important role in the performance of automated face recognition systems. Demographic factors can potentially introduce biases in face recognition systems and are well studied in the literature \textsuperscript{[21],[15],[1]}. Work to mitigate biases due to demographic factors are currently being investigated \textsuperscript{[20],[18]}. Typically, some of these attributes can be modified physically, e.g., by applying hair dye or undergoing surgery. But what if these attributes are digitally modified? Individuals can alter their facial features in photos using image editors for cheekbone highlighting, forehead reduction, etc. The modified images, with revised features, may be posted in social media websites. But do such manipulations affect the biometric utility of these images? Assessing the impact of digital retouching on biometric identification accuracy

\textsuperscript{*} Corresponding author
was done in [4]. With the arrival of generative adversarial networks (GANs), the possibilities of automated attribute editing have exploded [11,17,16,18]. GANs can be used to change the direction of hair bangs, remove facial hair, change the intensity of tinted eyeglasses and even add facial expressions to face images. With GANs for facial age progression [29], a person’s appearance can seamlessly transit from looking decades younger to appearing as an elderly individual. Note that the user does not have to be a deep learning expert to edit attributes in face images. Several smartphone-based applications have such attribute modifications in the form of filters, e.g., FaceApp [7]. Open-source applications make it easy to modify an image by uploading the image, selecting the attribute to be edited, using a slider to regulate the magnitude of the change, and downloading the edited photo. The entire process can be easily accomplished in under five minutes [5]. Recently, mask-aware face editing methods have emerged [26].

Motivation: Although the intent of using image editing routines stems from personal preferences, they can be misused for obscuring an identity or impersonating another identity. The style transfer networks are typically evaluated from the perspective of visual realism, i.e., how realistic do the generated images look? However, we rarely investigate the impact of such digital manipulations on biometric face recognition. Work has been done to localize the manipulations [13], estimate the generative model responsible for producing the effects [30], and gauge the robustness of face recognition networks with respect to pose and expression [10]. But we need to investigate GAN-based attribute manipulations from the perspective of biometric recognition. Studying the influence of digital manipulations of both demographic and facial attributes on face images is pivotal because an individual can use the edited images in identification documents. Therefore, it is imperative to assess the impact of GAN-based attribute manipulations on biometric recognition to evaluate the robustness of existing open-source deep learning-based face recognition systems [23]. Our objective is to conduct an investigative study that examines the impact of attribute editing (thirteen attributes) of face images by AttGAN [11] and STGAN [17] on two popular open-source face matchers, namely, ArcFace [6] and VGGFace [27].

The remainder of the paper is organized as follows. Section 2 describes image attribute editing GANs and open-source face recognition networks analyzed in the work. Section 3 describes the experimental protocols followed in this work. Section 4 reports and analyzes the findings. Section 5 concludes the paper.

2 Proposed Study

In this work, we investigate how GAN-induced attribute manipulations affect face recognition. In the process, we will review the following research questions through our study.

1. Does GAN-based attribute editing only produce perceptual changes in face images?
Can GAN-induced Attribute Manipulations Impact Face Recognition?  

Fig. 1: Examples of attribute-manipulated images using AttGAN and STGAN for a (a) male subject and a (b) female subject. To demonstrate the ability of the GAN as a faithful autoencoder, ‘Reconstruction’ images are also displayed.

2. Are there certain attribute manipulations that are more detrimental than others on face verification performance?

3. Is the impact of GAN-based manipulations consistent across different face recognition networks?

To answer the above questions, we used two GAN-based image editing deep networks, viz., AttGAN and STGAN to modify thirteen attributes on a set of face images from the CelebA dataset. The edited images are then compared with the original face images in terms of biometric utility using two deep learning-based face recognition networks, namely, ArcFace and VGGFace. We hypothesize that digital manipulations induced using GANs can alter the visual perceptibility of the images, but more importantly, affect the biometric utility of the images. Note that these attributes are not originally present in the images: they are artificially induced. Although such manipulations may seem innocuous, they can produce unexpected changes in the face recognition performance and should be handled cautiously. We now discuss the GANs used for attribute manipulation and the deep face networks considered in this work.

AttGAN [11]: It can perform binary facial attribute manipulation by modeling the relationship between the attributes and the latent representation of the face. It consists of an encoder, a decoder, an attribute classification network and a discriminator. The attribute classifier imposes an attribute classification constraint to ensure that the generated image possesses the desired attribute.
Reconstruction learning is introduced such that the generated image mimics the original image, i.e., only the desired attribute should change without compromising the integrity of the remaining details in the image. Finally, adversarial learning maintains the overall visual realism of the generated images. The network allows high quality facial attribute editing with control over the attribute intensity and accommodates changes in the attribute style.

**STGAN** [17]: AttGAN employs an encoder-decoder structure with spatial pooling or downsampling that results in reduction in the resolution and irrecoverable loss of fine details. Therefore, the generated images are susceptible to loss of features and look blurred. Skip connections are introduced in AttGAN that directly concatenate encoder and decoder features and can improve the reconstruction image quality but at the expense of reduced attribute manipulation ability. In contrast, STGAN adopts selective transfer units that can adaptively transform the encoder features supervised by the attributes to be edited. STGAN accepts the difference between the target and source attribute vector, known as the difference attribute vector, as input unlike the AttGAN that takes the entire target attribute vector as input. Taking the difference vector results in more controlled manipulation of the attribute and simplifies the training process.

**VGGFace** [27]: “Very deep” convolutional neural networks are bootstrapped to learn a \( N \)-way face classifier to recognize \( N \) subjects. It is designed such that the network associates each training image to a score vector by using the final fully connected layer that comprises \( N \) linear predictors. Each image is mapped to one of the \( N \) identities present during training. The score vector needs fine-tuning to perform verification by comparing face descriptors learned in the Euclidean space. To that end, triplet-loss is employed during training to learn a compact face representation (projection) that is well separated across disjoint identities.

**ArcFace** [6]: Existing face recognition methods attempt to directly learn the face embedding using a softmax loss function or a triplet loss function. However, each of these embedding approaches suffer from drawbacks. Although softmax loss is effective on closed-set face recognition, the face representations are not discriminative enough for open-set face recognition. Triplet loss can cause a huge number of face triplets for large datasets, and semi-hard triplet mining is challenging for effective training. In contrast, additive angular margin loss (ArcFace) optimises the geodesic distance margin by establishing correspondence between the angle and arc in normalized hypersphere. Due to its concise geometric interpretation, ArcFace can provide highly discriminative features for face recognition, and has become one of the most widely-used open-source deep face recognition networks.

### 3 Experiments

We used 2,641 images belonging to 853 unique individuals with an average of \( \sim 3 \) images per subject from the CelebA [19] dataset for conducting the investigative study in this work. Next, we edited each of these images to alter 13
Can GAN-induced Attribute Manipulations Impact Face Recognition?

Fig. 2: DET curves obtained for all pairs of GANs (for attribute editing) and deep face recognition networks (for biometric matching). (a) STGAN-VGGFace, (b) STGAN-ArcFace, (c) AttGAN-VGGFace and (d) AttGAN-ArcFace.

attributes, one at a time, using AttGAN and STGAN to generate $2,641 \times 13 \times 2 = 68,666$ manipulated images. We used the codes and pre-trained models provided by the original authors of AttGAN [3] and STGAN [25] in our work. The list of attributes edited are as follows: Bald, Bangs, Black_Hair, Blond_Hair, Brown_Hair, Bushy_Eyebrows, Eyeglasses, Male, Mouth_Slightly_Open, Mustache, No_Beard, Pale_Skin, and Young. We selected these thirteen attributes as they were modified using the original AttGAN and STGAN models. Examples of GAN edited images are presented in Figures 1(a) and 1(b). Note that the attributes are applied in a toggle fashion. For example, the Male attribute changes the face image of the male individual by inducing makeup to impart a feminine appearance (see Figure 1(a) ninth image from the left), while the same attribute induces facial hair in the face image of the female subject to impart a masculine appearance (see Figure 1(b) ninth image from the left). We can also observe that the manipulations are not identical across the two GANs and can manifest differently across different sexes. For example, attribute Bald is more effectively induced by STGAN compared to AttGAN and has a more
pronounced effect on male images than female images (see Figures 1(a) and (b) second image from the left). To perform quantitative evaluation, we compared an original image \((I_i)\) with its respective attribute manipulated image \((O_j)\) using a biometric comparator, \(B(I_i, O_j)\), where \(B(\cdot, \cdot)\) extracts face representations from each image and computes the vector distance between them. We used cosine distance in this work. The distance value is termed as the biometric match score between a pair of face images. The subscripts indicate the subject identifier. If \(i = j\), then \((I_i, O_j)\) forms a genuine pair (images belong to the same individual). If \(i \neq j\), then \((I_i, O_j)\) constitutes an impostor pair (images belong to different individuals). We used the genuine and impostor scores to compute the detection error trade-off (DET) curve, and compared the face recognition performance between the original images and the corresponding attribute manipulated images. In the DET curve, we plot False Non-Match Rate (FNMR) vs. False Match Rate (FMR) at various thresholds. We repeated this process to obtain fourteen curves, \(viz.,\) one curve corresponding to the original images (original-original comparison) and the remaining thirteen curves corresponding to the thirteen attributes manipulated by the GAN (original-attribute edited comparison). We used an open-source library for the implementation of ArcFace and VGGFace [2].

4 Findings

We present the DET curves for four different combinations of attribute manipulating GANs and face comparators used in this work, namely, STGAN-VGGFace, STGAN-ArcFace, AttGAN-VGGFace and AttGAN-ArcFace in Figure 2. We also tabulated the results from the DET curves at FMR=0.01 in Table 1 and at FMR=0.1 in Table 2.

Additionally, we performed an experiment to determine how the number of impostor scores affect the overall biometric recognition performance for GAN-edited images. The number of scores corresponding to genuine pairs was \(\sim 10 \text{K}\), and the number of scores corresponding to impostor pairs was \(\sim 7 \text{M}\). Due to memory constraints, we selected a subset of impostor scores for plotting the DET curves. We randomly selected impostor scores, without replacement, in factors of 5X, 10X, 50X and 100X of the total number of genuine scores (10K). This resulted in 50K, 100K, 500K and 1M impostor scores, respectively. We observed that the DET curves obtained by varying the number of impostor scores from the AttGAN-ArcFace combination are almost identical. See Figure 3. Therefore, for the remaining three combinations, we used the entire set of genuine scores but restricted to utilizing one million impostor scores.

Analysis: In Table 1, we reported the results @FMR=0.01. We observed that the attribute ‘Eyeglasses’, a facial attribute produced the highest degradation in the biometric recognition performance (see the red colored cells). AttGAN-ArcFace achieved FNMR=0.98 compared to the original performance of FNMR=0.25 @FMR=0.01 (reduction in performance by 73%). Note that we reported the reduction as the difference between the FNMR obtained for
Can GAN-induced Attribute Manipulations Impact Face Recognition? 7

Fig. 3: DET curves of AttGAN-ArcFace obtained after varying the number of impostor scores (a) 50K, (b) 100K, (c) 500K and (d) 1M. We varied the number of impostor scores with respect to the number of genuine scores (~10K) in factors of 5X, 10X, 50X and 100X, respectively. Note the DET curves show minimal variations.

modified images and the original images. The DET curve indicated as Original in Figure 2 denotes the face recognition performance computed using only the unmodified images. AttGAN-VGGFace achieved FNMR=0.52 compared to the original performance of FNMR=0.32 @FMR=0.01 (reduction in performance by 20%). STGAN-ArcFace achieved FNMR=0.31 compared to the original performance of FNMR=0.25 @FMR=0.01 (reduction in performance by 6%). STGAN-VGGFace achieved FNMR=0.39 compared to the original performance of FNMR=0.32 @FMR=0.01 (reduction in performance by 7%). The second highest degradation was caused due to the change in ‘Male’, a demographic attribute. AttGAN-ArcFace achieved FNMR=0.47 compared to the original performance of FNMR=0.25 @FMR=0.01 (reduction in performance by 22%). AttGAN-VGGFace achieved FNMR=0.43 compared to the original performance of FNMR=0.32 @FMR=0.01 (reduction in performance by 11%). STGAN-ArcFace achieved FNMR=0.30 compared to the original performance of
Table 1: Face recognition performance: False Non-Match Rate (FNMR) at a False Match Rate (FMR) = 0.01 for attribute manipulated images. The red colored cells correspond to the maximum degradation (if any) in the biometric recognition performance compared to the original in each column. The green colored cells correspond to the maximum improvement (if any) in the biometric recognition performance compared to the original in each column. Identical performance results in multiple colored cells within a column. Note majority of the attributes resulted in degradation. The attribute ‘Eyeglasses’ caused the worst degradation in biometric recognition performance followed by the ‘Male’ attribute.

| Attribute          | STGAN-VGGFace | STGAN-ArcFace | AttGAN-VGGFace | AttGAN-ArcFace |
|--------------------|---------------|---------------|----------------|----------------|
| Original           | 0.32          | 0.25          | 0.32           | 0.25           |
| Bald               | 0.34          | 0.22          | 0.35           | 0.40           |
| Bangs              | 0.28          | 0.23          | 0.35           | 0.57           |
| Black_Hair         | 0.29          | 0.20          | 0.33           | 0.35           |
| Blond_Hair         | 0.30          | 0.21          | 0.33           | 0.39           |
| Brown_hair         | 0.26          | 0.19          | 0.31           | 0.34           |
| Bushy_Eyebrows     | 0.30          | 0.22          | 0.35           | 0.40           |
| Eyeglasses         | 0.39          | 0.31          | 0.52           | 0.98           |
| Male               | 0.42          | 0.30          | 0.43           | 0.47           |
| Mouth_Slightly_Open| 0.26          | 0.21          | 0.32           | 0.38           |
| Mustache           | 0.28          | 0.20          | 0.34           | 0.45           |
| No_Beard           | 0.28          | 0.20          | 0.33           | 0.41           |
| Pale_Skin          | 0.27          | 0.18          | 0.34           | 0.40           |
| Young              | 0.31          | 0.22          | 0.37           | 0.40           |

FNMR=0.25 @FMR=0.01 (reduction in performance by 5%). STGAN-VGGFace achieved FNMR=0.42 compared to the original performance of FNMR=0.32 @FMR=0.01 (reduction in performance by 10%). STGAN-ArcFace achieved FNMR=0.18 compared to the original performance of FNMR=0.25 @FMR=0.01 (improvement in performance by 7%) for the attribute ‘Pale_Skin’. In Table 2 we reported the results @FMR=0.1. We observed that the attributes seem to improve the recognition performance by up to 4% in FNMR for STGAN-ArcFace. On the contrary, for the remaining three sets, viz., STGAN-VGGFace, AttGAN-ArcFace and AttGAN-VGGFace, we continued to observe a drop in recognition performance with ‘Eyeglasses’ resulting in the worst drop in performance by 14%, followed by ‘Male’ resulting in the second highest drop in performance by 7%. Now, let us review the questions posited at the beginning of the paper.

**Question #1:** Does GAN-based attribute editing only produce perceptual changes in face images?

**Observation:** No. GAN-based attribute editing not only alters the perceptual quality of the images but also significantly impacts the biometric recognition performance. See Tables 1 and 2 where a majority of the attributes strongly degraded the face recognition performance.
Table 2: Face recognition performance: False Non-Match Rate (FNMR) at a False Match Rate (FMR) = 0.1 for attribute manipulated images. The red colored cells correspond to the maximum degradation (if any) in the biometric recognition performance compared to the original in each column. The green colored cells correspond to the maximum improvement (if any) in the biometric recognition performance compared to the original in each column. Identical performance results in multiple colored cells within a column. Note @FMR=0.1, STGAN-ArcFace (attribute editing using STGAN and face matching using ArcFace) caused the least degradation in face recognition.

| Attribute       | STGAN-VGGFace | STGAN-ArcFace | AttGAN-VGGFace | AttGAN-ArcFace |
|-----------------|---------------|---------------|----------------|----------------|
| Original        | 0.10          | 0.14          | 0.10           | 0.14           |
| Bald            | 0.12          | 0.12          | 0.11           | 0.16           |
| Bangs           | 0.09          | 0.11          | 0.11           | 0.17           |
| Black_Hair      | 0.09          | 0.10          | 0.11           | 0.14           |
| Blond_Hair      | 0.10          | 0.11          | 0.10           | 0.15           |
| Brown_hair      | 0.08          | 0.10          | 0.10           | 0.14           |
| Bushy_Eyebrows  | 0.09          | 0.11          | 0.11           | 0.16           |
| Eyeglasses      | 0.13          | 0.13          | 0.19           | 0.28           |
| Male            | 0.17          | 0.14          | 0.16           | 0.18           |
| Mouth_Slightly_Open | 0.08        | 0.10          | 0.10           | 0.15           |
| Mustache        | 0.09          | 0.10          | 0.10           | 0.16           |
| No_Beard        | 0.08          | 0.10          | 0.10           | 0.16           |
| Pale_Skin       | 0.08          | 0.10          | 0.10           | 0.16           |
| Young           | 0.10          | 0.11          | 0.13           | 0.16           |

Question #2: Are there certain attribute manipulations that are more detrimental than others on face verification performance?

Observation: Yes. Editing of 'Eyeglasses' attribute caused a significant degradation in FNMR by up to 73% @FMR=0.01 for AttGAN-ArcFace. It was followed by the the 'Male' attribute that caused the second highest degradation in FNMR by up to 22% @FMR=0.01 for AttGAN-ArcFace. Surprisingly, a facial attribute, 'Eyeglasses', and a demographic attribute, 'Male', when edited separately, were responsible for significant degradation in the biometric recognition performance in a majority of the scenarios.

Question #3: Is the impact of GAN-based manipulations consistent across different face recognition networks?

Observation: No. The impact of GAN-based attribute manipulations on face recognition depends on two factors, firstly, which GAN was used to perform the attribute editing operation, and secondly, which face recognition network was used to measure the biometric recognition performance. For example, the ‘Bald’ attribute caused degradation in the performance for AttGAN, irrespective of which face matcher was used. However, the same attribute caused improvement in the face recognition accuracy for STGAN when ArcFace was used as the face matcher, but caused reduction in the recognition performance when VGGFace
was used as the face comparator. Similar findings were observed for the ‘Bangs’
attribute.

![Fig. 4: Examples depicting ‘Male’ attribute edited using STGAN and AttGAN, respectively, in male (top row) and female (bottom row) individuals. AttGAN output images appear to contain more visual artifacts compared to STGAN outputs.](image)

To investigate the possibility of **dataset-specific bias**, we selected 495 im-
gages from the Labeled Faces in-the-Wild (LFW) dataset [12] belonging to 100
individuals, and applied attribute editing using AttGAN. AttGAN resulted in
the worst drop in performance on the CelebA dataset, so we employed it for this
experiment. This is done to examine whether the impact of attribute editing
on face recognition performance manifests across datasets. Next, we computed
the face recognition performance between the original and the attribute-edited
images using ArcFace and VGGFace comparators and observed the following

![Fig. 5: Examples depicting ‘Eyeglasses’ attribute edited using STGAN and AttGAN, respectively, in original images (a) Without Eyeglasses and (b) With Eyeglasses. Note in images containing eyeglasses, the GANs fail to remove the glasses completely, instead they lighten the shades of the eyeglasses or the glass rims. STGAN outputs appear better for images with and without eyeglasses.](image)
results. ‘Eyeglasses’ attribute editing resulted in the worst degradation in the performance by 83% while using ArcFace and by 37% while using VGGFace in terms of FNMR at FMR=0.01. It was followed by ‘Bangs’ attribute resulting in second-highest degradation in the performance by 69% while using ArcFace and by 32% while using VGGFace in terms of FNMR at FMR=0.01. See Table 3.

To investigate the possibility of **GAN-specific artifacts**, we compared the ‘Reconstructed’ images with the attribute edited images, specifically, the ‘Eyeglasses’ and ‘Male’ attribute edited images, to evaluate the face recognition performance. This is done to assess whether face recognition performance is affected solely by attribute editing or influenced by GAN-specific artifacts. Note that ‘Reconstructed’ images are faithful reconstruction of ‘Original’ images that demonstrate the fidelity of the GAN as an effective autoencoder. See Table 4 for FNMR at FMR=0.01 and 0.1 for Reconstructed-Reconstructed, Reconstructed-Eyeglasses and Reconstructed-Male comparisons. We observed similar degradation in performance when comparing attribute edited images with GAN-reconstructed images by up to 47% for ‘Eyeglasses’ attribute and by up to 12% for ‘Male’ attribute in terms of FNMR at FMR=0.01. The results indicate that AttGAN has an overall weaker reconstruction and attribute editing capability than STGAN.

The implications of our findings are as follows. Bias due to naturally prevalent demographic factors in automated face recognition systems can be further aggravated when attributes are digitally modified. Digitally altering the sex cues (de-
Table 4: Face recognition performance: False Non-Match Rate (FNMR) at a False Match Rate (FMR) = 0.01/0.1 for attribute manipulated images. We report the face recognition performance by comparing ‘Reconstructed’ images with ‘Eyeglasses’ attribute and ‘Male’ attribute edited images.

|                  | STGAN-VGGFace | STGAN-ArcFace | AttGAN-VGGFace | AttGAN-ArcFace |
|------------------|---------------|---------------|----------------|---------------|
| Reconstructed    | 0.30/0.08     | 0.21/0.10     | 0.42/0.12      | 0.49/0.20     |
| Eyeglasses       | 0.38/0.12     | 0.27/0.11     | 0.50/0.19      | 0.96/0.29     |
| Male             | 0.42/0.16     | 0.28/0.12     | 0.47/0.20      | 0.46/0.19     |

noted by ‘Male’ attribute) can be considered as manipulating a demographic attribute. It involves adding facial hair to images of female individuals and adding makeup to impart feminine appearance to images of male individuals. See Figure 4. These artificial manipulations affect face recognition performance. Surprisingly, altering a facial attribute like ‘Eyeglasses’ caused an excessive degradation in face recognition performance. Modifying the ‘Eyeglasses’ attribute involves adding glasses to individual face images where no eyeglasses are present and removing eyeglasses in the images where the individual is wearing one. See Figure 5. AttGAN struggles with addition as well as removal of eyeglasses from the images, and, instead, produces visually apparent artifacts that might be responsible for significant degradation in the biometric recognition performance. Removal of glasses is particularly hard: the GANs are only able to lighten the lens shades or the color of the eyeglass frames, but not completely remove the glasses. To check for any statistical variation between the results produced by AttGAN for the ‘Eyeglasses’ attribute, we repeated the experiment with the original images five times. Each time, we followed the same procedure and executed basic attribute editing and not attribute sliding (sliding regulates the intensity of attribute modification). We obtained exactly identical results for each of the five runs, i.e., a decrease by 73% in FNMR @FMR=0.01. On the LFW dataset, ‘Eyeglasses’ attribute resulted in a decrease by up to 83% in FNMR @FMR=0.01, while ‘Bangs’, also a type of facial attribute, reduced the biometric recognition performance by up to 69%. Therefore, we observed that ‘Eyeglasses’ attribute editing reduced the biometric recognition performance considerably on both CelebA and LFW datasets. Although the exact reason responsible for significant degradation in face recognition performance caused by ‘Eyeglasses’ attribute editing is unknown, we speculate that the attribute manipulations may produce severe changes in texture around the facial landmarks resulting in the drop in face recognition performance.

Therefore, our findings indicate that digitally modified attributes, both demographic and facial, can have a major impact on automated face recognition systems and can potentially introduce new biases that require further examination.
5 Conclusion

In this paper, we examined the impact of GAN-based attribute editing on face images in terms of face recognition performance. GAN-generated images are typically evaluated with respect to visual realism but their influence on biometric recognition is rarely analyzed. Therefore, we studied face recognition performance obtained using ArcFace and VGGFace after modifying thirteen attributes via AttGAN and STGAN on a total of \( \sim 68,000 \) images belonging to 853 individuals from the CelebA dataset. Our findings indicated some interesting aspects:

(i) Insertion or deletion of eyeglasses from a face image can significantly impair biometric recognition performance by up to 73%. Digitally modifying the sex cues caused the second highest degradation in the performance by up to 22%.

(ii) There can be an artificial boost in the recognition accuracy by up to 7% depending on the GAN used to modify the attributes and the deep learning-based face recognition network used to evaluate the biometric performance. Similar observations were reported when tested on a different dataset. Our findings indicate that attribute manipulations accomplished via GANs can significantly affect automated face recognition performance and need extensive analysis. Future work will focus on examining the effect of editing multiple attributes simultaneously, in a single face image, on face recognition.

References

1. Albiero, V., Zhang, K., King, M.C., Bowyer, K.W.: Gendered differences in face recognition accuracy explained by hairstyles, makeup, and facial morphology. IEEE Transactions on Information Forensics and Security 17, 127–137 (2022)
2. ArcFace and VGGFace implementation: https://pypi.org/project/deepface/ [Online accessed: 18th May, 2022]
3. AttGAN implementation: https://github.com/LynnHo/AttGAN-Tensorflow [Online accessed: 18th May, 2022]
4. Bharati, A., Singh, R., Vatsa, M., Bowyer, K.W.: Detecting facial retouching using supervised deep learning. IEEE Transactions on Information Forensics and Security 11(9), 1903–1913 (2016)
5. Chen, H.J., Hui, K.M., Wang, S.Y., Tsao, L.W., Shuai, H.H., Cheng, W.H.: Beauty-Glow: On-Demand Makeup Transfer Framework With Reversible Generative Network. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 10034–10042 (2019)
6. Deng, J., Guo, J., Xue, N., Zafeiriou, S.: ArcFace: Additive Angular Margin Loss for Deep Face Recognition. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4685–4694 (2019)
7. FaceApp: https://www.faceapp.com/ [Online accessed: 17th May, 2022]
8. Gong, S., Liu, X., Jain, A.K.: Jointly De-Biasing Face Recognition and Demographic Attribute Estimation. In: 16th European Conference on Computer Vision (ECCV). p. 330–347 (2020)
9. Grother, P., Ngan, M., Hanaoka, K.: Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects. NIST IR 8280 https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST_IR.8280.pdf (2019)
10. Gutierrez, N.R., Theobald, B., Ranjan, A., Abdelaziz, A.H., Apostoloff, N.: MorphGAN: Controllable One-Shot Face Synthesis. British Machine Vision Conference (BMVC) (2021)
11. He, Z., Zuo, W., Kan, M., Shan, S., Chen, X.: AttGAN: Facial Attribute Editing by Only Changing What You Want. IEEE Transactions on Image Processing 28(11), 5464–5478 (2019)
12. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Tech. Rep. 07-49, University of Massachusetts, Amherst (October 2007)
13. Huang, Y., Juefei-Xu, F., Wang, R., Xie, X., Ma, L., Li, J., Miao, W., Liu, Y., Pu, G.: FakeLocator: Robust Localization of GAN-Based Face Manipulations via Semantic Segmentation Networks with Bells and Whistles. ArXiv (2022)
14. Klare, B.F., Burge, M.J., Klontz, J.C., Vorder Bruegge, R.W., Jain, A.K.: Face recognition performance: Role of demographic information. IEEE Transactions on Information Forensics and Security 7(6), 1789–1801 (2012)
15. Krishnan, A., Almadan, A., Rattani, A.: Understanding fairness of gender classification algorithms across gender-race groups. In: 19th IEEE International Conference on Machine Learning and Applications (ICMLA). pp. 1028–1035 (2020)
16. Kwak, J., Han, D.K., Ko, H.: CAFE-GAN: Arbitrary Face Attribute Editing with Complementary Attention Feature. In: European Conference on Computer Vision (ECCV) (2020)
17. Liu, M., Ding, Y., Xia, M., Liu, X., Ding, E., Zuo, W., Wen, S.: STGAN: A Unified Selective Transfer Network for Arbitrary Image Attribute Editing. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3668–3677 (2019)
18. Liu, S., Li, D., Cao, T., Sun, Y., Hu, Y., Ji, J.: GAN-Based Face Attribute Editing. IEEE Access 8, 34854–34867 (2020)
19. Liu, Z., Luo, P., Wang, X., Tang, X.: Deep Learning Face Attributes in the Wild. In: Proceedings of International Conference on Computer Vision (December 2015)
20. Mirjalili, V., Raschka, S., Ross, A.: PrivacyNet: Semi-Adversarial Networks for Multi-Attribute Face Privacy. IEEE Transactions on Image Processing 29, 9400–9412 (2020)
21. NIST Study Evaluates Effects of Race, Age, Sex on Face Recognition Software: [https://www.nist.gov/news-events/news/2019/12/nist-study-evaluates-effects-race-age-sex-face-recognition-software] [Online accessed: 4th May, 2022]
22. Phillips, P.J., Flynn, P.J., Bowyer, K.W.: Lessons from collecting a million biometric samples. Image and Vision Computing 58, 96–107 (2017)
23. Ross, A., Banerjee, S., Chen, C., Chowdhury, A., Mirjalili, V., Sharma, R., Swearingen, T., Yadav, S.: Some research problems in biometrics: The future beckons. In: International Conference on Biometrics (ICB). pp. 1–8 (2019)
24. Sgroi, A., Garvey, H., Bowyer, K., Flynn, P.: Location matters: A study of the effects of environment on facial recognition for biometric security. In: 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition. vol. 02, pp. 1–7 (2015)
25. STGAN implementation: [https://github.com/csmliu/STGAN] [Online accessed: 18th May, 2022]
26. Sun, R., Huang, C., Zhu, H., Ma, L.: Mask-aware photorealistic facial attribute manipulation. In: Computational Visual Media. vol. 7, pp. 363–374 (2021)
27. Taigman, Y., Yang, M., Ranzato, M., Wolf, L.: DeepFace: Closing the Gap to Human-Level Performance in Face Verification. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 1701–1708 (2014)
28. Terhörst, P., Kolf, J.N., Huber, M., Kirchbuchner, F., Damer, N., Moreno, A.M., Fierrez, J., Kuijper, A.: A Comprehensive Study on Face Recognition Biases Beyond Demographics. IEEE Transactions on Technology and Society 3(1), 16–30 (2022)
29. Yang, H., Huang, D., Wang, Y., Jain, A.K.: Learning Face Age Progression: A Pyramid Architecture of GANs. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 31–39 (2018)
30. Yu, N., Davis, L., Fritz, M.: Attributing Fake Images to GANs: Learning and Analyzing GAN Fingerprints. In: IEEE/CVF International Conference on Computer Vision. pp. 7555–7565 (2019)