electronics

MDPI

Article

Face Attribute Modification Using Fine-Tuned Attribute-Modification Network

Naeem Ul Islam [1] and Jaebyung Park [1,2,*]

[1] Core Research Institute of Intelligent Robots, Jeonbuk National University, Jeonju 54896, Korea; naeem@jbnu.ac.kr

[2] Division of Electronics and Information Engineering, Jeonbuk National University, Jeonju 54896, Korea

* Correspondence: jbpark@jbnu.ac.kr; Tel.: +82-63-270-4283

Received: 18 April 2020; Accepted: 29 April 2020; Published: 30 April 2020

Abstract: Multi-domain image-to-image translation with the desired attributes is an important approach for modifying single or multiple attributes of a face image, but is still a challenging task in the computer vision field. Previous methods were based on either attribute-independent or attribute-dependent approaches. The attribute-independent approach, in which the modification is performed in the latent representation, has performance limitations because it requires paired data for changing the desired attributes. In contrast, the attribute-dependent approach is effective because it can modify the required features while maintaining the information in the given image. However, the attribute-dependent approach is sensitive to attribute modifications performed while preserving the face identity, and requires a careful model design for generating high-quality results. To address this problem, we propose a fine-tuned attribute modification network (FTAMN). The FTAMN comprises a single generator and two discriminators. The discriminators use the modified image in two configurations with the binary attributes to fine tune the generator such that the generator can generate high-quality attribute-modification results. Experimental results obtained using the CelebA dataset verify the feasibility and effectiveness of the proposed FTAMN for editing multiple facial attributes while preserving the other details.

Keywords: generative adversarial network; convolutional neural network; fine-tuned attribute-modification network; autoencoders

1. Introduction

Facial attributes represent intuitive semantic features such as "male", "female", "person with eyeglasses" and "smiling" that describe the biological identity or expression of a person. Extensive research has been conducted on the biological identity of human faces in the field of computer vision, from face identification and detection [1–4] to face-attribute modification [5–7]. These approaches were successful based on three factors: (i) access to numerous publicly available training data with labels, (ii) the high computational capabilities of GPUs, and (iii) access to open-source libraries. The availability of the aforementioned resources made it possible for researchers to perform a large amount of work in the fields of face identification, detection, and attribute modification. The prominent facial-attribute modification task is, however, more challenging as compared to those of face recognition and detection, wherein a careful description of the semantic aspects of the face is required while modifying the required attributes, in addition to keeping the face identity intact. For example, if we want to modify a specific attribute, such as hair color, we need to have semantic information about the hair and modify only that part of the image without changing any facial details.

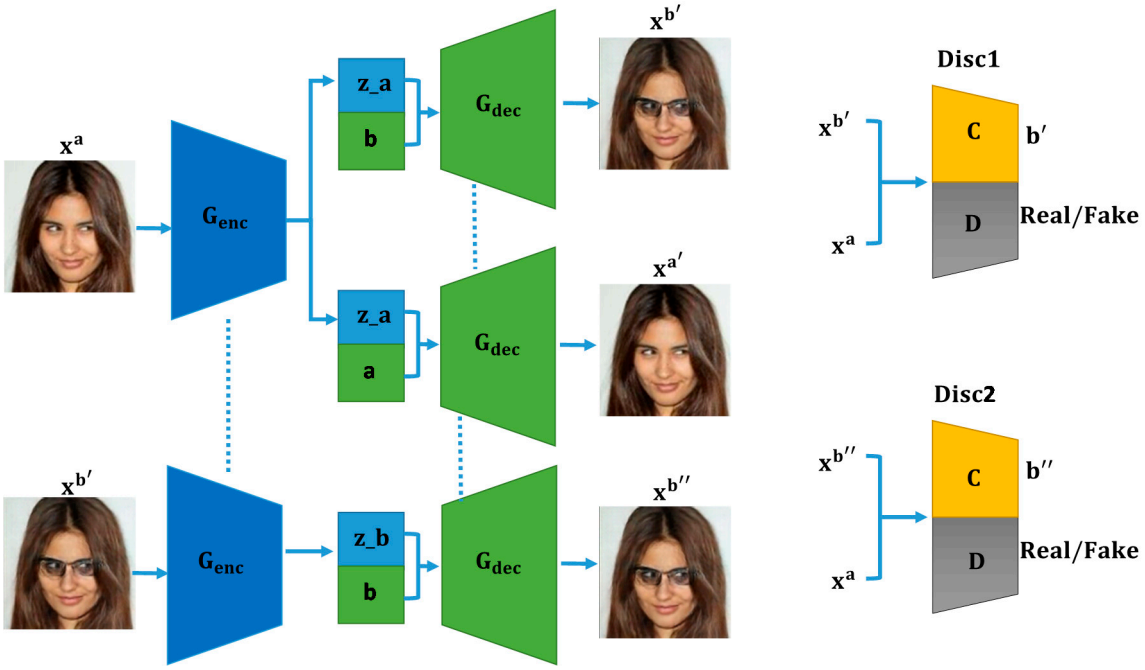
Extensive studies have been conducted on facial attribute modification and image-to-image translation in computer graphics in terms of various applications such as color modification [8],

content modification [9], image wrapping [10,11], image translation [12–17], and interpretation [18]. The image editing problem has been handled using two types of approaches: example-based [19–21] and model-based approaches [11,22,23]. In the example-based approach, the required attribute is searched for in the given reference image and transferred to the target image. This makes the image editing dependent on the available reference image in many ways [19–21]. However, the reference image must be of the same person with an appropriate face alignment and the same lighting conditions. In the model-based approach, the model of the required face is first built and the image is then modified accordingly [11,22,23]. Although these approaches have proven to be successful in the modification of particular attributes, they are task-specific, and thus, it is impossible to apply them to arbitrary attribute-modification problems.

Several face-attribute modification approaches [6,24] and distortion-removal approaches for real images [24] have been developed based on recent developments in deep neural networks, such as generative adversarial networks (GANs) [25] and variational autoencoders (VAEs) [26]. Both GANs and VAEs are powerful models and are capable of generating images. GANs generate more realistic images as compared to VAEs. However, the GAN cannot encode images as it uses random noise as an input. In contrast, the VAE is capable of encoding the image to its corresponding latent representation although its generated image is blurry as compared to that of the GAN. A combination of the GAN and autoencoder makes a powerful tool for image-attribute editing. In Invertable Conditional GAN (IcGAN) [27] and Conditional GAN (cGAN) [28], the GAN and autoencoder are combined for editing the attributes of an image. They modify the latent representation to reflect the expected attribute and then decode the modified image. The object transfiguration is learned by GeneGAN [5] from two unpaired sets of images: one set of images with specific attributes and the other without those attributes. The only constraint faced here is that the objects are located at approximately the same place. For example, the training data can comprise one set of reference images of faces with eyeglasses and another set of images of faces without eyeglasses, where both sets are spatially aligned using face landmarks. DNA-GAN [29] shares similar traits with GeneGAN and provides "crossbreed" images by swapping the latent representation of the corresponding attributes between the given pair of images. Hence, the DNA-GAN can be considered as an extension of the GeneGAN. These methods have been proved to be useful in image editing tasks, but they require different models for different attributes, which is not practical in real-world applications owing to the corresponding time complexity. The proposed work is based on the attribute-dependent approach, wherein an attribute classification constraint is applied to the generated images. Attribute GAN (AttGAN) [6] can change the required attribute while keeping the other details unchanged. The modified attributes, however, are not prominent, although the face identity is well preserved. In contrast, the proposed approach is focused on the prominent modification of the attributes while keeping the rest of the attributes unchanged. The proposed approach is different from the AttGAN because it has an additional discriminator that we call the refined discriminator. The refined discriminator takes the modified image from the generator as an input and helps the generator to refine it further to obtain a better output. Further explanation of the refined discriminator is presented in Section 2. RelGAN in [7] can effectively modify attributes simultaneously with an additional capability of interpolation. They obtain the relative attributes by determining the difference between the given and predefined target attributes. In contrast, the proposed approach does not consider the predefined attributes but uses the already-available target attributes with a further refinement, which is provided by the refined discriminator. Recently, a facial attribute modification network (FAMN) was proposed in [30], which has an architecture style that is common with that of the AttGAN and the proposed approach. The FAMN has two generators and two discriminators, and both the generators share the same encoder part. The decoder part of both the generators takes the latent representation as its input along with the binary and mean attribute vectors, whereas each generator has its own discriminator. In contrast, the proposed approach consists of a single generator and two discriminators, where the role of the second discriminator is to refine the modified output further and, hence, to generate more realistic results with a lower computational

complexity. In [31], the authors proposed the Multimodal Unsupervised Image-to-Image Translation (MUNIT) framework, where they decomposed the image representation into a content code and a style code and then recombined the content code with a random style code sampled from the style space of the target domain to generate different style output. The cycle-consistent adversarial networks (CycleGANs) [32] translate an input image from one domain to another domain in an unsupervised manner, where the corresponding target pair is not given. Both MUNIT and CycleGAN can effectively translate images from one domain to another domain. However, their focus is on the style of the whole image rather than a specific attribute. Considering the time efficiency, because the proposed network has a single generator and two discriminators while the AttGAN has a single generator and a single discriminator, the number of parameters in the proposed network are more than those of the AttGAN and thus the proposed network requires more training time than the AttGAN. However, the discriminator is only included in training as shown in Figure 1 and does not affect the time complexity during the testing. The FAMN, on the other hand, has two discriminators and an extra decoder whereas the proposed approach does not have the extra decoder. Thus, the number of parameters in the proposed network are less than those of the FAMN and the proposed network requires less training time than the FAMN. The testing time for the AttGAN, the FAMN, and the proposed approach is same. The contribution of this work can be summarized as follows:

- A novel network architecture that utilizes the latent representation of the given and modified outputs along with the given and required attributes is proposed to generate prominent modification in the given input.
- The tuning and refining discriminators ensure the prominent modifications by guiding the generators.
- A unified approach is proposed to effectively change the appearance of faces while preserving the identity.
- Multiple experiments are carried out to validate the proposed approach.



(**a**)

Figure 1. *Cont.*

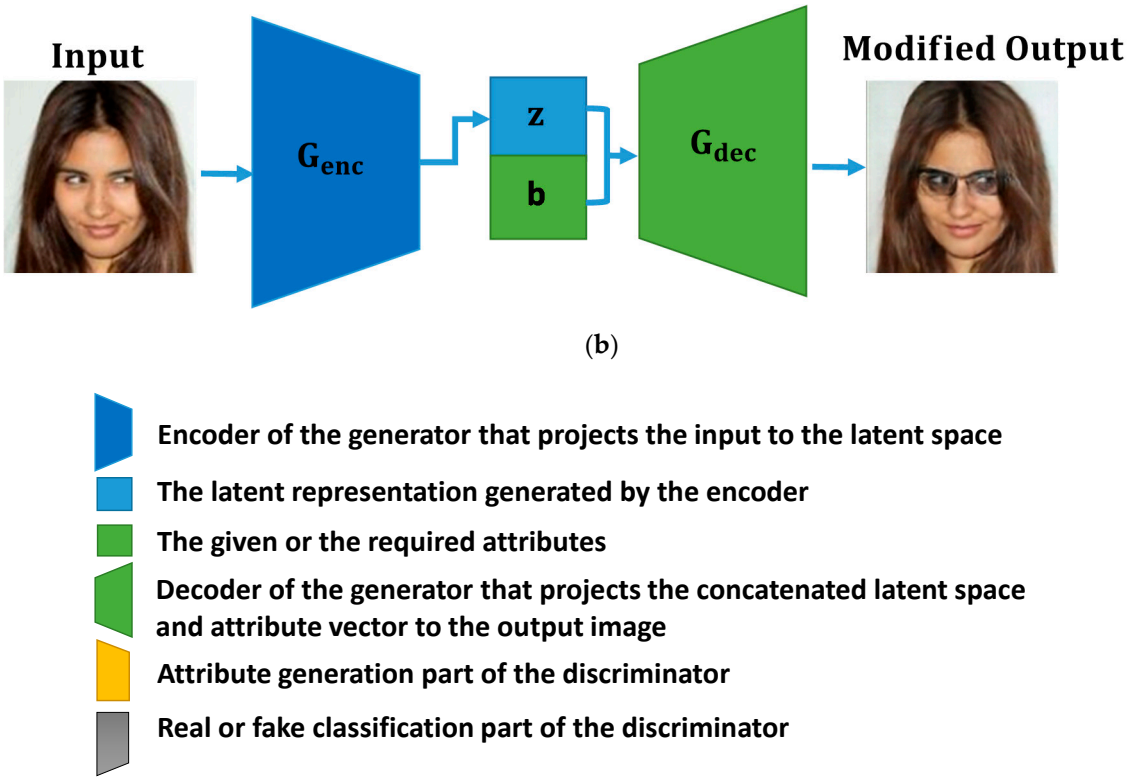


Figure 1. Fine-tuned attribute-modification network: (**a**) Training; (**b**) Testing; (**c**) Network descriptions.

2. Proposed Approach

This section introduces the fine-tuned attribute-modification approach for editing facial attributes. The facial attributes are represented by a binary vector, where "1" represents the presence of a particular attribute in the given face image and "0" represents its absence. The configuration of the proposed FTAMN is presented in Figure 1. The FTAMN consists of a single generator with an encoder G_{enc} and a decoder G_{dec}, and two discriminators Disc1 and Disc2. The generator takes the input face image in two different steps. In the first step, the encoder part G_{enc} of the generator takes the face image $\mathbf{x^a}$ that is required to be modified and projects it to the latent representation **z_a**. The decoder G_{dec} then takes the pair of the latent representation **z_a** and given attribute vector **a** and generates the image $\mathbf{x^{a'}}$. Similarly, the latent representation **z_a** and required attribute vector **b** are decoded back to the modified image $\mathbf{x^{b'}}$ using G_{dec}. In the second step, G_{enc} takes the generated image $\mathbf{x^{b'}}$ as its input and projects it to its corresponding latent representation **z_b**. Next, G_{dec} takes the pair of the latent representation **z_b** and the required attribute vector **b** and then generates the modified image $\mathbf{x^{b''}}$. In terms of the discriminators, the first discriminator Disc1 takes either the real image $\mathbf{x^a}$ or fake image $\mathbf{x^{b'}}$ generated by G_{dec} and maps it to the attribute vectors **a** or **b** along with real and fake classification labels. It thus guides the generator to generate the modified image. Similarly, the second discriminator Disc2 is used as a tuning discriminator that takes the real image $\mathbf{x^a}$ or fake image $\mathbf{x^{b''}}$ generated by G_{dec} and maps it to the corresponding attribute vectors **a** or **b** along with real and fake classification labels. It thus guides the generator to generate the required modified image with prominent attributes.

3. Training

The purpose of the proposed FTAMN is to modify multiple attributes simultaneously using the attribute information available in the input data. The generator takes the required image as an input, and we modify some of its contents by decoding a different combination of attributes. In this approach,

we selected thirteen attributes that are required to be modified. These attributes include baldness, bangs, black hair, blond hair, brown hair, bushy eyebrows, eyeglasses, gender, mouth open/closed, mustache, no beard, pale skin, and young. For example, if black hair and a mustache are the desired attributes in the input testing sample, this approach will make the hair black and put a mustache on the face in the given image. During the training, the attributes of each image are concatenated in three configurations. In the first configuration, the original attributes are concatenated with the latent representation of the input image, and the combination of the attributes and latent representation are then decoded back to the original image using G_{dec}. The original attribute objective along with the GAN objective is used to train Disc1 and Disc2. In the second configuration, the input attributes are first shuffled and then concatenated with the latent representation of the input image. Next, the concatenated pair is decoded back to the modified image using G_{dec}. In the third configuration, the generated modified image is first fed back to the encoder G_{enc} and is projected to its corresponding latent representation. The latent representation and the required attributes are concatenated. Next, the concatenated pair is decoded back to the modified image using G_{dec}. The attribute objective along with the reconstruction and the GAN objectives are used to tune the parameters of the generator of the network.

The detailed explanation of this process and stepwise training procedure are as follows. For a given face input image $\mathbf{x^a}$, its attributes are defined as a vector **a** as follows.

$$\mathbf{a} = [a_1,\ a_2,\ \ldots,\ a_n] \tag{1}$$

In the first step, the encoder part G_{enc} of the FTAMN takes $\mathbf{x^a}$ as its input and transforms it into its corresponding latent representation **z_a** as follows.

$$\mathbf{z\_a} = G_{enc}(\mathbf{x^a}) \tag{2}$$

During the first configuration, the decoder G_{dec} translates the latent representation **z_a** along with its attribute vector **a** to $\mathbf{x^{a'}}$ as follows.

$$\mathbf{x^{a'}} = G_{dec}(\mathbf{z\_a}, \mathbf{a}) \tag{3}$$

During the second configuration, the attribute vector **a** is first modified to reflect the required attributes as follows.

$$\mathbf{b} = f(\mathbf{a}) \tag{4}$$

where $f(\cdot)$ is a function used for changing the elements of a binary vector. The decoder G_{dec} then translates the latent representation **z_a** along with the required attribute vector **b** to the modified image $\mathbf{x^{b'}}$ as follows.

$$\mathbf{x^{b'}} = G_{dec}(\mathbf{z\_a}, \mathbf{b}) \tag{5}$$

During the third configuration, the latent representation **z_b** is first obtained as follows.

$$\mathbf{z\_b} = G_{enc}\left(\mathbf{x^{b'}}\right) \tag{6}$$

The decoder G_{dec} then translates the latent representation **z_b** along with the required attribute vector **b** to the refined modified image $\mathbf{x^{b''}}$.

$$\mathbf{x^{b''}} = G_{dec}(\mathbf{z\_b}, \mathbf{b}) \tag{7}$$

The GAN objective functions are defined for training the discriminators Disc1 and Disc2 as follows.

$$L_{Disc1\_GAN} = \gamma_d\left(Disc1\_D(\mathbf{x^a}) + 1 - Disc1\_D\left(\mathbf{x^{b'}}\right)\right) \tag{8}$$

$$L_{Disc2\_GAN} = \gamma_d\left(Disc2\_D(\mathbf{x^a}) + 1 - Disc2\_D\left(\mathbf{x^{b''}}\right)\right) \tag{9}$$

where L_{Disc1_GAN} and L_{Disc2_GAN}, respectively, are the GAN losses for training each of the discriminators. Disc1_D($\mathbf{x^a}$) and Disc2_D($\mathbf{x^a}$) represent the real outputs of the discriminators for the given original input that is required to be modified. Disc1_D$\left(\mathbf{x^{b'}}\right)$ and Disc2_D$\left(\mathbf{x^{b''}}\right)$ represent the fake outputs of the discriminators for the given modified face image generated by G_{dec}. The attribute objective for preserving the remaining attributes of the face image is defined as follows.

$$\begin{aligned} LDisc1_{att\_a} &= -\sum_{i=1}^{n} a_i \log Disc1\_C\left(x_i^a\right) \\ &= -a_i \log\left(Disc1\_C\left(x_i^a\right)\right) - (1-a_i)\log\left(1 - Disc1\_C\left(x_i^a\right)\right) \end{aligned} \tag{10}$$

where $LDisc1_{att\_a}$ represents the sigmoid cross-entropy loss for the given attributes, a_i represents the target attribute, and Disc1_C$\left(x_i^a\right)$ represents the generated predicted attributes.

$$\begin{aligned} LDisc2_{att\_a} &= -\sum_{i=1}^{n} a_i \log Disc2\_C\left(x_i^a\right) \\ &= -a_i \log\left(Disc2\_C\left(x_i^a\right)\right) - (1-a_i)\log\left(1 - Disc2\_C\left(x_i^a\right)\right) \end{aligned} \tag{11}$$

where $LDisc2_{att\_a}$ represents the sigmoid cross-entropy loss for the given attributes, and Disc2_C$\left(x_i^a\right)$ represents the generated predicted attributes.

The overall objective for training the discriminator Disc1 is defined as follows.

$$L_{Disc1} = L_{Disc1\_GAN} + \alpha LDisc1_{att\_a} \tag{12}$$

where α represents the control parameter for the original attribute objective. The overall objective for training the discriminator Disc2 is defined as follows.

$$L_{Disc2} = L_{Disc2\_GAN} + \alpha LDisc2_{att\_a} \tag{13}$$

Similarly, the training objective functions for the generator in the first and second steps are defined as follows.

$$L_{G1\_GAN} = \gamma_d\left(1 - Disc1\_D\left(\mathbf{x^{b'}}\right)\right) \tag{14}$$

$$L_{G2\_GAN} = \gamma_d\left(1 - Disc2\_D\left(\mathbf{x^{b''}}\right)\right) \tag{15}$$

where L_{G1_GAN} and L_{G2_GAN} are the GAN losses for training the generator in the first and second steps, respectively.

$$\begin{aligned} LDisc1_{att\_b} &= -\sum_{i=1}^{n} b_i \log Disc1\_C\left(x_i^{b'}\right) \\ &= -b_i \log\left(Disc1\_C\left(x_i^{b'}\right)\right) - (1-b_i)\log\left(1 - Disc1\_C\left(x_i^{b'}\right)\right) \end{aligned} \tag{16}$$

$$\begin{aligned} LDisc2_{att\_b} &= -\sum_{i=1}^{n} b_i \log Disc2\_C\left(x_i^{b''}\right) \\ &= -b_i \log\left(Disc2\_C\left(x_i^{b''}\right)\right) - (1-b_i)\log\left(1 - Disc2\_C\left(x_i^{b''}\right)\right) \end{aligned} \tag{17}$$

where $LDisc1_{att\_b}$ and $LDisc2_{att\_b}$ represent the sigmoid cross-entropy losses for the modified attributes b_i, where b_i represents the target binary attributes. Disc1_C$\left(x_i^{b'}\right)$ and Disc2_C$\left(x_i^{b''}\right)$ represent the corresponding generated predicted attributes.

$$L_{recons} = \zeta\left(abs\left(\mathbf{x^a} - \mathbf{x^{a'}}\right)\right) \tag{18}$$

where L_{recons} represents the absolute reconstruction objective, where the aim is to preserve the remaining attributes of the face image during the modification of the required attribute objectives. $\mathbf{x^a}$ represents the input face image, and $\mathbf{x^{a'}}$ represents the image reconstructed with the intention of preserving the

remaining features of the given face image. ζ is the control parameter for the reconstruction objective that preserves the face identity.

$$L_{refine} = \zeta\left(abs\left(\mathbf{x}^{\mathbf{b}'} - \mathbf{x}^{\mathbf{b}''}\right)\right) \quad (19)$$

where L_{refine} represents the absolute refinement objective, where the aim is to further refine the modification in the given image while considering the provided attributes. $\mathbf{x}^{\mathbf{b}'}$ represents the generated modified image in the first step, and $\mathbf{x}^{\mathbf{b}''}$ represents the generated modified image in the second step.

The overall training objective for the generator part of the network is expressed as follows.

$$L_{Gen} = L_{G1\_GAN} + L_{G2\_GAN} + \lambda LDisc1_{att\_b} + \lambda LDisc2_{att\_b} + L_{recons} + L_{refine} \quad (20)$$

where λ represents the control parameter for the required attribute objective, and L_{Gen} represents the total loss of the generator.

4. Experiments

For training the FTAMN, we used the CelebA dataset [33]. The CelebA is a large-scale face dataset comprising 202,599 face images. We divided the CelebA dataset into training and testing sets. The training set comprises 182,000 images, and the testing set comprises the remaining 20,599 images. After the training, we evaluated the network for its ability to modify the input images according to the required attributes. We analyzed the experimental results qualitatively and quantitatively by defining the structural similarity index (SSIM) from the reconstructed images and the modified images as shown in Table 1. The proposed network was implemented by using TensorFlow 1.7, an open-source deep learning framework, on the GPU-based PC, which was comprised of an Intel(R) Core i9-9940X CPU, 132.0 GB RAM, and four NVIDIA GeForce RTX 2080 Ti graphics cards.

Table 1. Structural similarity index (SSIM) between the reconstructed images and their corresponding modified images.

FTAMN (proposed approach)					
Attributes	**Bald**	**Mouth Open/Closed**	**Bangs**	**Eyeglasses**	**Mustache**
SSIM per pixel	5.3262×10^{-6}	3.9688×10^{-6}	5.4398×10^{-6}	6.0457×10^{-6}	6.3409×10^{-6}
SSIM per image	0.9474	0.9883	0.9676	0.9410	0.9869
AttGAN [6]					
Attributes	**Bald**	**Mouth Open/Closed**	**Bangs**	**Eyeglasses**	**Mustache**
SSIM per pixel	5.4324×10^{-6}	3.9795×10^{-6}	5.4689×10^{-6}	6.1952×10^{-6}	6.3807×10^{-6}
SSIM per image	0.9663	0.9910	0.9728	0.9642	0.9931
FAMN [30]					
Attributes	**Bald**	**Mouth Open/Closed**	**Bangs**	**Eyeglasses**	**Mustache**
SSIM per pixel	5.3993×10^{-6}	3.9778×10^{-6}	5.4359×10^{-6}	6.2458×10^{-6}	6.3415×10^{-6}
SSIM per image	0.9604	0.9906	0.9669	0.9721	0.9870

In the first experimental analysis, we evaluated the proposed approach in terms of the baldness of the given input image. If the given image is not bald, the proposed approach will inverse the bald attribute and generate a bald image. The comparison of the qualitative results of the proposed approach with those of the AttGAN [6] and the FAMN [30] in terms of baldness is presented in Figure 2. Figure 2a–c shows the results obtained using the proposed FTAMN, the AttGAN, and the FAMN, respectively. The green frame indicates the required successful modification, the blue frame indicates the successful modification with lost identity, and the red frame indicates the failure of the

required attribute modifications. Considering the modified images of the AttGAN, we can observe that the generated images are smooth but their smoothness affects the quality of the modified image, as indicated by the red frames in Figure 2b. In other words, the baldness effect is not distinct in the results of the AttGAN as compared to our results. In the results of the FAMN, we can observe that the generated images show prominent baldness but the identity of the face is affected as indicated by the blue frames in Figure 2c. In contrast, the FTAMN translates the attributes in a prominent manner with considerable smoothness and retains the face identity as compared to the FAMN, as indicated by the green frames in Figure 2a. This proves that, as compared to the AttGAN and the FAMN, the proposed FTAMN is more effective in translating the required attributes to the given input face images.

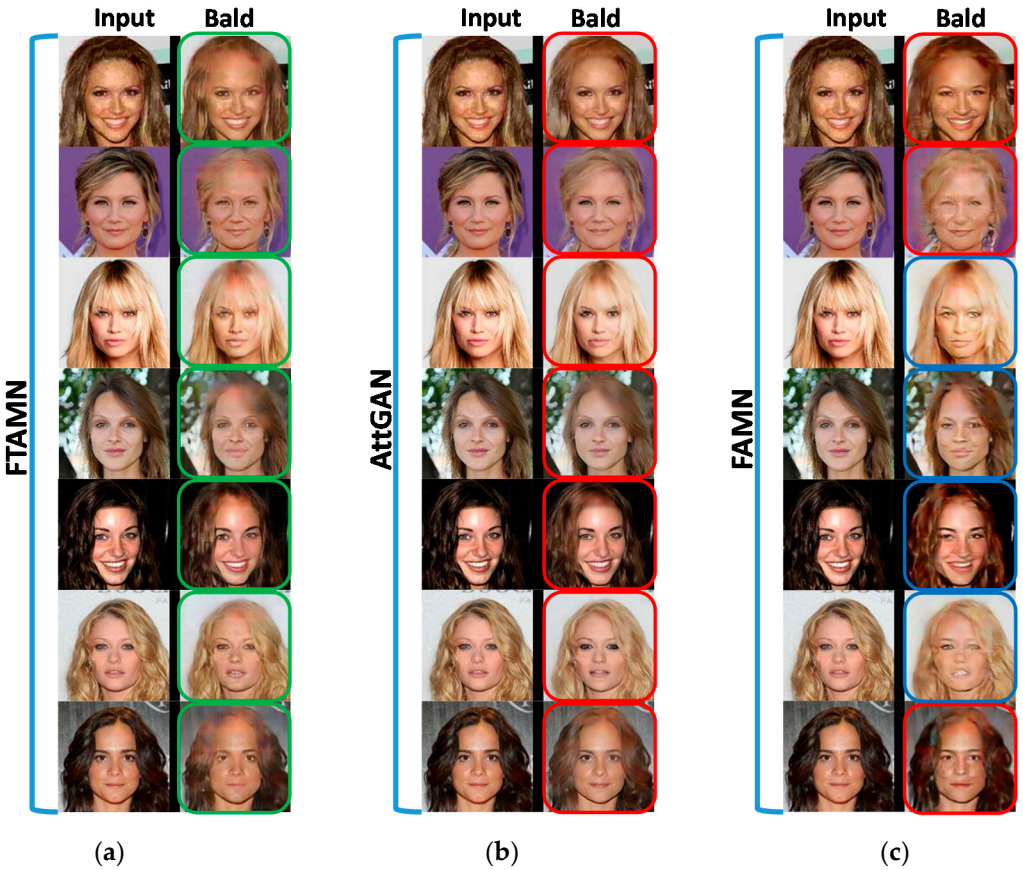


Figure 2. Modification of the bald attributes in the given face images. (**a**) Fine-tuned attribute modification network (FTAMN) (proposed approach); (**b**) Attribute GAN (AttGAN) [6]; (**c**) Facial attribute modification network (FAMN) [30]. The input images in the first column of each subfigure with hair are transformed into the output images that show bald faces.

In the second analysis, we evaluated the proposed approach in terms of the mouth open/closed attribute, where the given input image with an open mouth is modified to obtain an output image with a closed mouth. The comparative analysis of the proposed approach with the AttGAN and the FAMN is presented in Figure 3. Figure 3a–c presents the results obtained using the proposed FTAMN, the AttGAN, and the FAMN, respectively. From this analysis, we can observe that the AttGAN translates the input images with an open mouth into the required output images with a closed mouth. However, the attribute modification is not prominent, as indicated by the red frames in Figure 3b. In contrast, the FTAMN translates the required attribute in a prominent manner in the output images, as indicated by the green frames in Figure 3a. The modification performance of the FTAMN is comparable with that of the FAMN, but the FTAMN preserves the face identity in the images better than the FAMN.

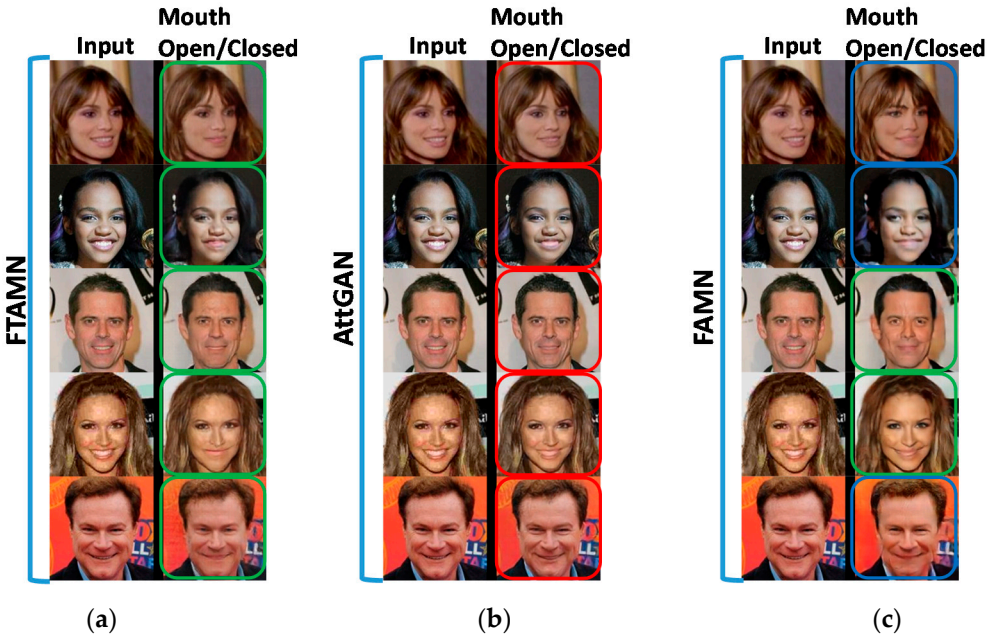


Figure 3. Modification of the mouth open or closed attributes in the given face images. (**a**) FTAMN (proposed approach); (**b**) AttGAN [6]; (**c**) FAMN [30]. The input images in the first column of each subfigure with an open mouth are transformed into the output images with a closed mouth.

In the third analysis, we evaluated the proposed approach in terms of the bangs attribute, where the given input image without bangs is modified to an output image that comprises bangs. The comparative analysis of the proposed approach with the AttGAN and the FAMN in terms of the bangs attribute translation is presented in Figure 4. From this analysis, we can observe that the bangs attribute effect is applied well to the given input images on using the proposed FTAMN, as shown in Figure 4a. In Figure 4a, the first and second columns, respectively, comprise the candidate input images and generated images with the required bangs effect. Figure 4b shows the results obtained from the AttGAN, where the given input images in the first column are translated to the output images in the second column with the required modification. The bangs effect is not prominent in the translated images with this approach as compared to the proposed FTAMN, although the face identity is preserved well. Figure 4c shows the results obtained with the FAMN, where the given input images in the first column are also translated to the output images in the second column with the required modification. The bangs effect is not prominent in the majority of the translated images of the FAMN as compared to the proposed FTAMN, especially in the red frame of Figure 4c. Furthermore, the FTAMN preserves the face identity better than the FAMN, as indicated by the green frame in Figure 4a.

In the fourth analysis, we evaluated the proposed approach in terms of the eyeglasses attribute, where the given input image without eyeglasses is modified into the output image with eyeglasses. The comparative analysis of the FTAMN with the AttGAN and the FAMN is presented in Figure 5, where Figure 5a–c presents the results of the FTAMN, the AttGAN, and the FAMN, respectively. As shown in the second columns of each subfigure in Figure 5, the FTAMN translates the eyeglasses attribute in more prominent manner in the case of a majority of the images as compared to the AttGAN and the FAMN, as indicated by the green frames in Figure 5a. Furthermore, the FTAMN preserves the face identity better than the FAMN, as indicated by the green frames in Figure 5a.

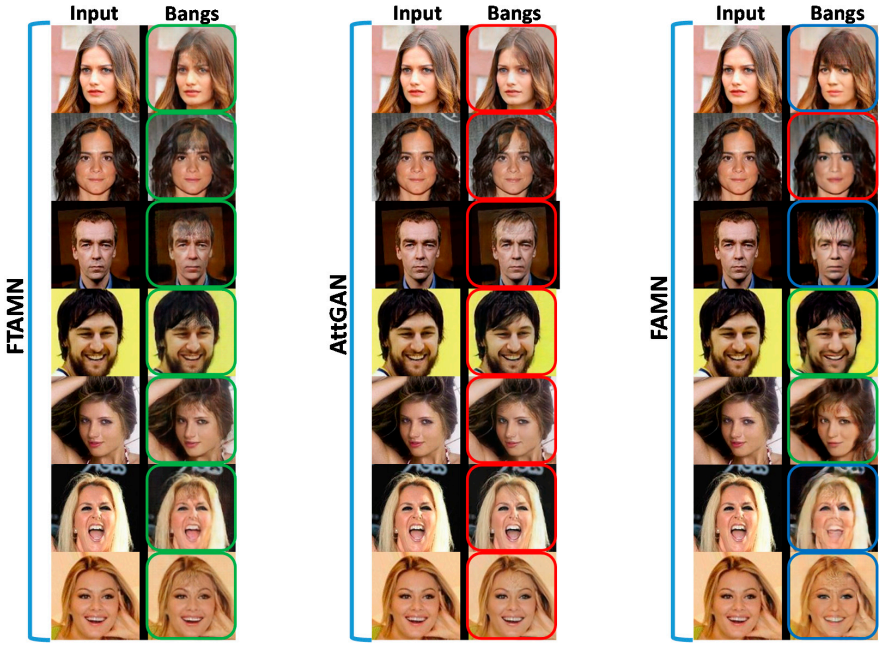


Figure 4. Modification of the bangs attribute in the given face images. (**a**) FTAMN (proposed approach); (**b**) AttGAN [6]; (**c**) FAMN [30]. The input images in the first column of each subfigure are transformed to the output images with bangs.

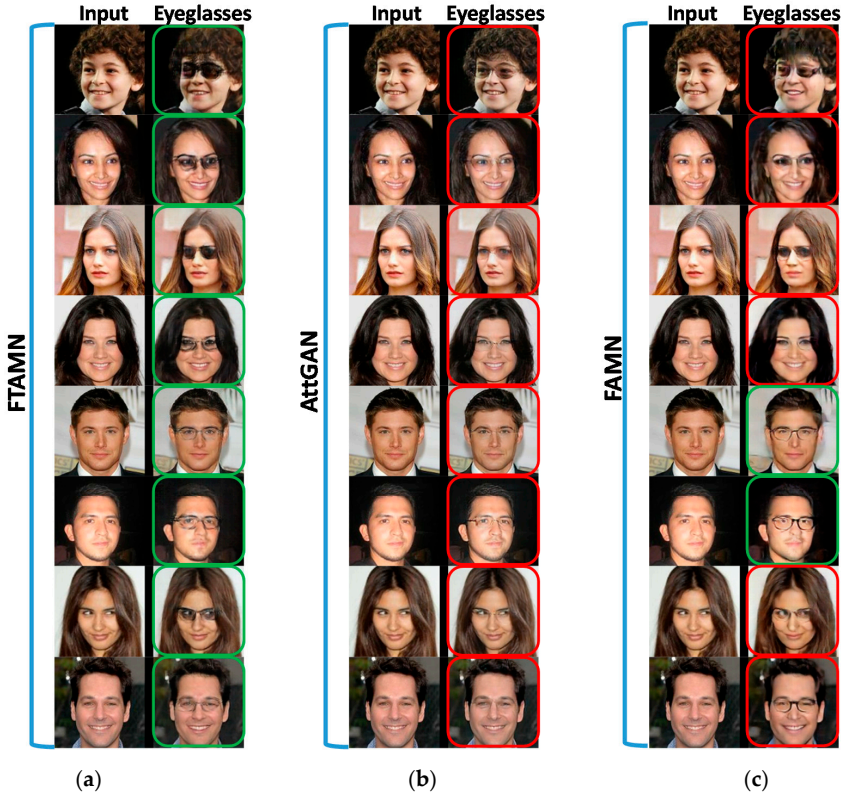


Figure 5. Modification of the eyeglasses attributes in the given face images. (**a**) FTAMN (proposed approach); (**b**) AttGAN [6]; (**c**) FAMN [30]. The input images in the first column of each subfigure without eyeglasses are transformed into the output images with eyeglasses.

In the fifth analysis, we evaluated the proposed approach in terms of the mustache attribute, where the given input image without a mustache is modified into the output image with a mustache irrespective of the gender. For example, if the given input image comprises no mustache, whether its face is male or female, we modify the input image to have a mustache. We performed a comparative analysis of the proposed approach with the AttGAN and the FAMN in terms of putting a mustache in the image, as presented in Figure 6. We can observe that the mustache attribute effect is applied well to the given input images by the FTAMN, as shown in Figure 6a. The first and second columns in Figure 6a show the candidate input images and the generated images with the required mustache on the face. Figure 6b,c presents the results obtained using the AttGAN and the FAMN, respectively, where the given input images in the first column are translated to the output images in the second column with the required modification. In the case of the AttGAN, the effect of the mustache attribute is not prominent in the translated images as compared to the FTAMN, although the face identity is preserved well, as indicated by the red frames in Figure 6b. In the case of the FAMN, as compared to the FTAMN, although the effect of the mustache attribute is prominent, the identity of the face image is affected, as indicated by the blue frames in Figure 6c.

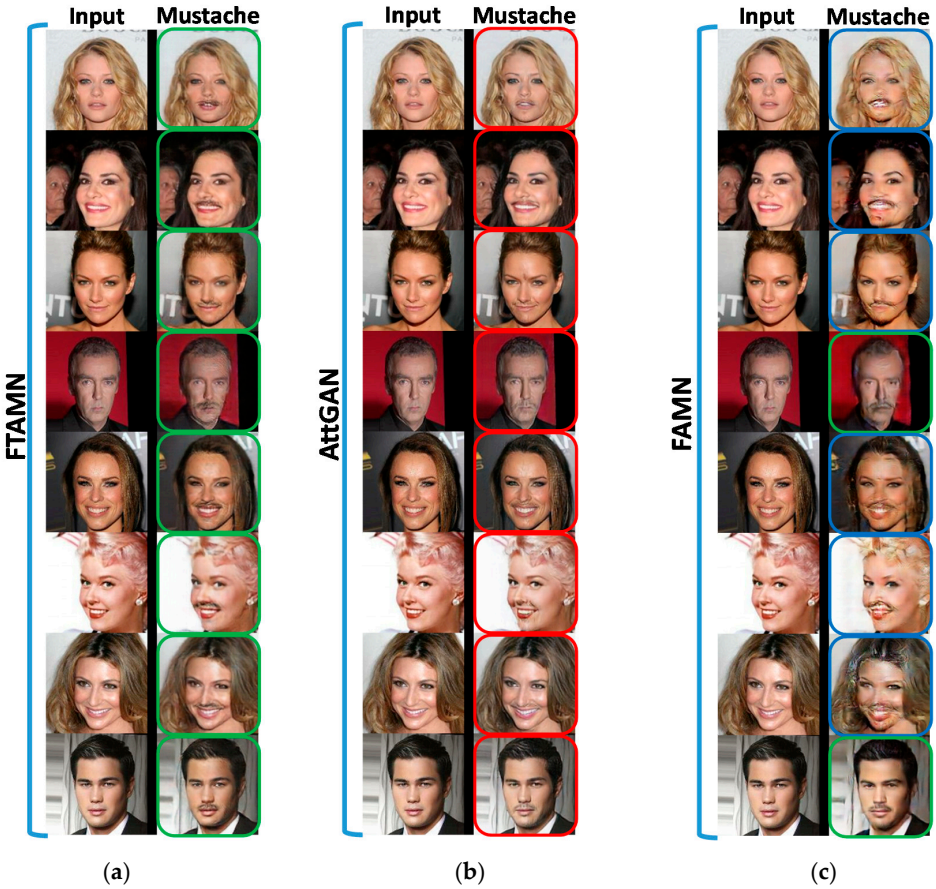


Figure 6. Modification of the mustache attribute in the given face images. (**a**) FTAMN (proposed approach); (**b**) AttGAN [6]; (**c**) FAMN [30]. The input images in the first column of each subfigure without a mustache are transformed into output images with a mustache.

In the final analysis, we performed a comparative qualitative analysis of the proposed FTAMN with the AttGAN and the FAMN for all of the thirteen attributes, as shown in Figure 7. Figure 7a–c shows the results obtained using the AttGAN, the FAMN, and the proposed FTAMN, respectively. The input images in the first column are the candidate samples that are required to be modified according to the required attributes. The second column shows the reconstructed results obtained using the AttGAN, the FAMN, and the FTAMN, and the remaining columns comprise the modified

results with the various aforementioned attributes ranging from bald to young. These results reflect that the proposed FTAMN is capable of modifying multiple attributes more efficiently as compared to the AttGAN and the FAMN.

Figure 7. Modification of all the attributes in the given face images. (**a**) AttGAN [6]; (**b**) FAMN [30]; (**c**) FTAMN (proposed approach). The input images in the first column of each subfigure are transformed into the output images with the required attributes in their corresponding columns.

The qualitative results discussed above showed the effectiveness of the proposed network for realistically modifying the given input images to their corresponding required attributes, compared with the AttGAN and the FAMN approaches. However, for insight into the analysis of the proposed approach, we performed a quantitative analysis. We selected five attributes such as bald, mouth open/closed, bangs, eyeglasses, and mustache. First, we randomly selected the testing samples and then generated the reconstructed samples as well as the modified images using the proposed FTAMN, the FAMN and the AttGAN approaches. We analyzed the structure of the images generated by the proposed FTAMN, the FAMN and the AttGAN approaches with the SSIM. To perform specific attribute dependent analysis, we first located the required target attribute region in the input image and then cropped the attribute region for comparative analysis using the SSIM as shown in Figure 8. The SSIM showed the structural similarity between the modified images and the corresponding reconstructed images. Under the rough estimation, the lower the SSIM was, the better the results were because we want prominent modification. The per-pixel and per-image SSIM results for the CelebA dataset are listed in Table 1. The per-pixel and per-image SSIM values were lower for the proposed FTAMN than the FAMN and the AttGAN, showing that the proposed approach outperformed the FAMN and the AttGAN in terms of prominent attribute modification.

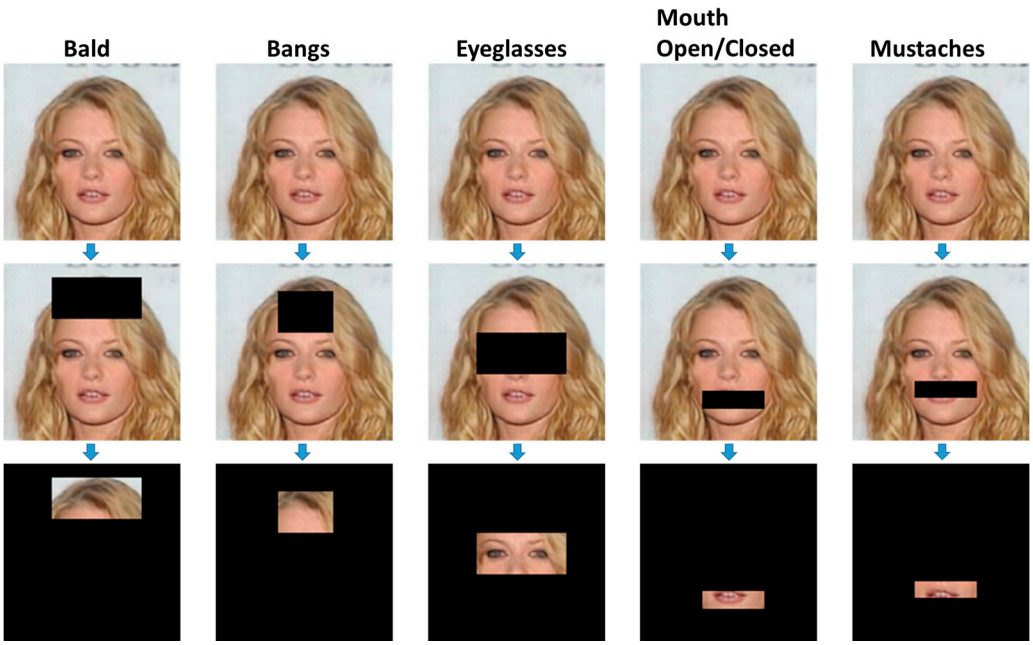


Figure 8. Data generation for quantitative analysis of the required attributes. The top row shows the reconstructed images, the middle row shows the required attribute region and the last row shows the cropped attribute region for comparative quantitative analysis.

5. Conclusions

In this paper, we proposed the FTAMN for efficiently modifying multiple attributes in the given input images. The proposed FTAMN consists of a generator and two discriminators. The first discriminator guides the generator to generate the required modified output face images, while the second refined discriminator further guides the generator to generate a realistic modification in the given face images. The proposed approach, with additional training strategy of the refined discriminator, can modify the input images effectively, as demonstrated in the experimental results. The FTAMN modifies the given attributes prominently while preserving the identity of the remaining attributes well in the given input images. Furthermore, as interpolation is a desirable feature in image modification, in our future work, we intend to extend the proposed approach to gain the

additional capability of interpolation along with attribute-independent modification for realizing a further improvement in prominent attribute modification while preserving the other features well.

Author Contributions: Conceptualization, N.U.I. and J.P.; methodology, N.U.I.; software, N.U.I.; validation, N.U.I.; formal analysis, N.U.I.; investigation, N.U.I. and J.P.; resources, N.U.I.; data curation, N.U.I.; writing—original draft preparation, N.U.I.; writing, review and editing, N.U.I. and J.P.; visualization, N.U.I.; supervision, J.P.; project administration, J.P.; funding acquisition, J.P. Both authors have read and agreed to the published version of the manuscript.

Funding: This research was partly supported by the Basic Science Research Programs (NRF-2019R1A6A1A09031717 and NRF-2018R1D1A1B07049270) through the National Research Foundation of Korea (NRF) funded by the Ministry of Education and was supported by "Research Base Construction Fund Support Program" funded by Jeonbuk National University in 2020.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Li, H.; Lin, Z.; Shen, X.; Brandt, J.; Hua, G. A convolutional neural network cascade for face detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, NJ, USA, 8–10 June 2015.
2. Bai, Y.; Zhang, Y.; Ding, M.; Ghanem, B. Finding tiny faces in the wild with generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018.
3. Deng, J.; Cheng, S.; Xue, N.; Zhou, Y.; Zafeiriou, S. UV-GAN: Adversarial facial UV map completion for pose-invariant face recognition. In Proceedings of the IEEE Conference of Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018.
4. Ahonen, T.; Hadid, A.; Pietikainen, M. Face description with local binary patterns: Application to face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 2037–2041. [CrossRef] [PubMed]
5. Zhou, S.; Xiao, T.; Yang, Y.; Feng, D.; He, Q.; He, W. GeneGAN: Learning object transfiguration and attribute subspace from unpaired data. *arXiv* **2017**, arXiv:1705.04932.
6. He, Z.; Zuo, W.; Kan, M.; Shan, S.; Chen, X. AttGAN: Facial Attribute Editing by Only Changing What You Want. *IEEE Trans. Image Process.* **2019**, *28*, 5464–5478. [CrossRef] [PubMed]
7. Wu, P.-W.; Lin, Y.-J.; Chang, C.-H.; Chang, E.Y.; Liao, S.-W. RelGAN: Multi-Domain Image-to-Image Translation via Relative Attributes. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019.
8. Reinhard, E.; Ashikhmin, M.; Gooch, B.; Shirley, P. Color transfer between images. *IEEE Comput. Graph. Appl.* **2001**, *21*, 34–41. [CrossRef]
9. Barnes, C.; Shechtman, E.; Finkelstein, A.; Goldman, D. PatchMatch: A randomized correspondence algorithm for structural image editing. *ACM Trans. Graph.* **2009**, *28*, 24:1–24:11. [CrossRef]
10. Alexa, M.; Cohen-Or, D.; Levin, D. As-rigid-as-possible shape interpolation. In Proceedings of the SIGGRAPH'00: The 27th International Conference on Computer Graphics and Interactive Techniques, New Orleans, LA, USA, 23–28 July 2000.
11. Hassner, T.; Harel, S.; Paz, E.; Enbar, R. Effective face frontalization in unconstrained images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, NJ, USA, 8–10 June 2015.
12. Islam, N.U.; Lee, S.; Park, J. Accurate and Consistent Image-to-Image Conditional Adversarial Network. *Electronics* **2020**, *9*, 395. [CrossRef]
13. Lee, S.; Islam, N.U. Robust Image Translation and Completion Based on Dual Auto-Encoder with Bidirectional Latent Space Regression. *IEEE Access* **2019**, *7*, 58695–58703. [CrossRef]
14. Islam, N.U.; Lee, S. Cross Domain Image Transformation Using Effective Latent Space Association. In Proceedings of the 15th International Conference IAS-15, Baden-Baden, Germany, 11–14 June 2018.
15. Islam, N.U.; Lee, S. Learning Typical 3D Representation from a Single 2D Correspondence using 2D-3D Transformation Network. In Proceedings of the International Conference on Ubiquitous Information Management and Communication, Phuket, Thailand, 4–6 January 2019.

16. Romera, E.; Bergasa, L.M.; Yang, K.; Alvarez, J.M.; Barea, R. Bridging the day and night domain gap for semantic segmentation. In Proceedings of the 2019 IEEE Intelligent Vehicles Symposium (IV), Paris, France, 9–12 June 2019.
17. Zhao, X.; Wang, K.; Yang, K.; Hu, W. Unconstrained face detection and recognition based on RGB-D camera for the visually impaired. In Proceedings of the 8th International Conference on Graphic and Image Processing (ICGIP 2016), Tokyo, Japan, 29–31 October 2016.
18. Islam, N.U.; Lee, S. Interpretation of deep CNN based on learning feature reconstruction with feedback weights. *IEEE Access* **2019**, *7*, 25195–25208. [CrossRef]
19. Guo, D.; Sim, T. Digital face makeup by example. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Miami Beach, FL, USA, 22–24 June 2009.
20. Liu, L.; Xu, H.; Xing, J.; Liu, S.; Zhou, X.; Yan, S. Wow! You are so beautiful today! In Proceedings of the 21st ACM International Conference on Multimedia, Barcelona, Catalunya, Spain, 21–25 October 2013.
21. Yang, F.; Wang, J.; Shechtman, E.; Bourdev, L.; Metaxas, D. Expression flow for 3D-aware face component transfer. In Proceedings of the SIGGRAPH'11: The 38th International Conference and Exhibition on Computer Graphics and Interactive Techniques, Vancouver, BC, Canada, 9–11 August 2011.
22. Kemelmacher-Shlizerman, I.; Suwajanakorn, S.; Seitz, S.M. Illumination-aware age progression. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 24–27 June 2014.
23. Kossaifi, J.; Tran, L.; Panagakis, Y.; Pantic, M. Geometry-aware generative adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018.
24. Park, D.-H.; Kakani, V.; Kim, H.-I. Automatic Radial Un-distortion using Conditional Generative Adversarial Network. *J. Inst. Control Robot. Syst.* **2019**, *25*, 1007–1013. [CrossRef]
25. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014.
26. Kingma, D.P.; Welling, M. Auto-encoding variational bayes. In Proceedings of the 2nd International Conference on Learning Representations, Banff, AB, Canada, 14–16 April 2014.
27. Perarnau, G.; Weijer, J.; Raducanu, B.; Alvarez, J.M. Invertible Conditional GANs for image editing. *arXiv* **2016**, arXiv:1611.06355.
28. Mirza, M.; Osindero, S. Conditional generative adversarial nets. *arXiv* **2014**, arXiv:1411.1784.
29. Kim, T.; Kim, B.; Cha, M.; Kim, J. Unsupervised visual attribute transfer with reconfigurable generative adversarial networks. *arXiv* **2017**, arXiv:1707.09798.
30. Islam, N.U.; Lee, S.; Park, J. Prominent Attribute Modification using Attribute Dependent Generative Adversarial Network. In Proceedings of the 17th International Conference on Ubiquitous Robots, Kyoto, Japan, 22–26 June 2020.
31. Huang, X.; Liu, M.-Y.; Belongie, S.; Kautz, J. Multimodal unsupervised image-to-image translation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
32. Zhu, J.-Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Honolulu, HI, USA, 22–25 July 2017.
33. Large-scale CelebFaces Attributes (CelebA) Dataset. Available online: http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html (accessed on 28 April 2020).