

Article

A GAN-Based Face Rotation for Artistic Portraits

Handong Kim ^{1,†}, Junho Kim ^{1,†} and Heekyung Yang ^{2,*} 

¹ Department of Computer Science, Sangmyung University, Seoul 03016, Korea

² Division of SW Convergence, Sangmyung University, Seoul 03016, Korea

* Correspondence: yanghk@smu.ac.kr; Tel.: +82-2-2287-6156

† These authors contributed equally to this work.

Abstract: We present a GAN-based model that rotates the faces in artistic portraits to various angles. We build a dataset of artistic portraits for training our GAN-based model by applying a 3D face model to the artistic portraits. We also devise proper loss functions to preserve the styles in the artistic portraits as well as to rotate the faces in the portraits to proper angles. These approaches enable us to construct a GAN-based face rotation model. We apply this model to various artistic portraits, including photorealistic oil paint portraits, watercolor portraits, well-known portrait artworks and banknote portraits, and produce convincing rotated faces in the artistic portraits. Finally, we prove that our model can produce improved results compared with the existing models by evaluating the similarity and the angles of the rotated faces through evaluation schemes including FID estimation, recognition ratio estimation, pose estimation and user study.

Keywords: face rotation; artistic portraits; hedcut portrait; GAN; 3D face model; loss function

MSC: 94A08



Citation: Kim, H.; Kim, J.; Yang, H. A GAN-Based Face Rotation for Artistic Portraits. *Mathematics* **2022**, *10*, 3860. <https://doi.org/10.3390/math10203860>

Academic Editors: Andrea Prati, Luis Javier García Villalba and Vincent A. Cicirello

Received: 3 September 2022

Accepted: 10 October 2022

Published: 18 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The artistic portrait, one of the fine art techniques for expressing the face of a person, has been greatly beloved since the early stages of fine art. Among the various categories of artistic portraits, we focus on the hedcut portrait, where the head of the target person covers most of the region of the canvas. Before the hedcut portrait became public when *The Wall Street Journal* employed this type of portrait in their half-column portrait illustration [1], and many artists have preferred this type of portrait since the early days of the fine arts. In computer graphics and computer vision, many studies including non-photorealistic rendering (NPR) [2], face rotation [3] and face generation [4] target the hedcut portrait in their frameworks. A hedcut portrait depicts faces with various angles and expressions in order to depict the person's face in a most attractive way.

Portraits from various angles have the advantage of depicting the face attractively but have a limitation in that they cannot present complete face information. For example, a left-headed portrait cannot deliver the right aspects of the face. In this study, we propose a framework to synthesize faces of various angles from the faces in various artistic portraits, including photorealistic oil paint portraits, watercolor portraits, well-known artistic portraits and banknote portraits (see Figure 1).

Face rotation technology in the field of computer vision is technology that synthesizes face images of various angles from a single input face image by learning the characteristics of the input face image. At the early stage, this technology was mainly used to improve the accuracy of face recognition by synthesizing a human face from the front angles by rotating the face to various angles. Recently, this technique was improved to synthesize face images of various angles from a face image at a given angle. This technique is known as face frontalization. By applying this technique to artistic portraits, we aim to synthesize artistic portraits at a desired angle from input portraits.

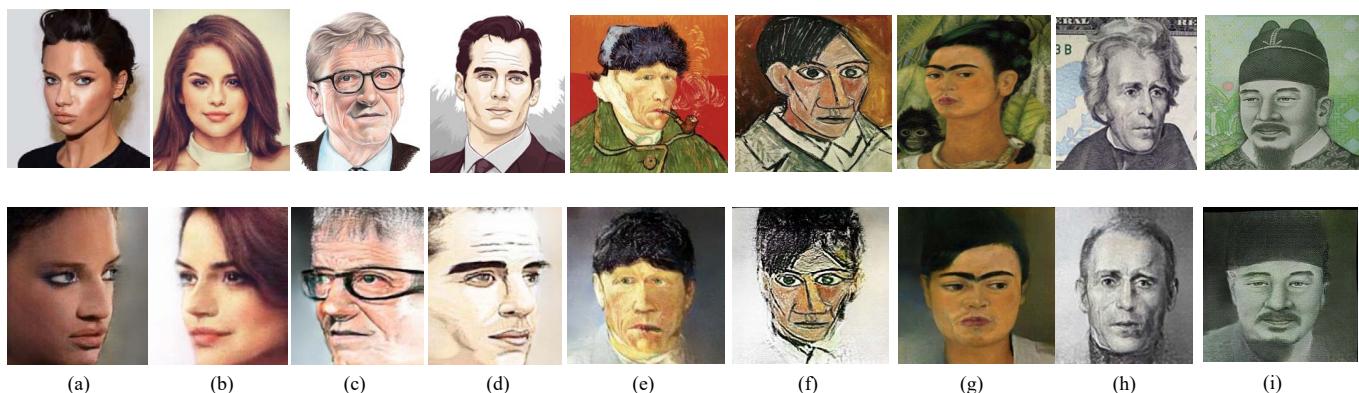


Figure 1. Teaser images. Portraits in upper row are rotated in lower row. (a,b) Photorealistic oil paint portraits. (c,d) Watercolor portraits. (e–g) Famous artistic portraits from Gogh, Picasso and Kahlo, respectively. (h,i) Banknote portraits. We rotated them to the right for (a,d), the left for (b,c) and the front angle for (e–i).

Recently, many researchers have presented a generative adversarial networks (GANs)-based approach to synthesizing rotated face images from a single face image. Training these GAN-based facial rotation techniques requires a dataset consisting of a series of face images from various angles for an identical face. Therefore, most recent facial rotation techniques focus on synthesizing photorealistic rotated faces. These techniques have a limitation for rotating faces in artistic portraits, since it is very hard to collect portraits of an identical face from various angles.

We overcome the limitations discussed above through the following process. First, a 3D face model is employed to overcome the limitation of the lack of datasets made up of art portraits. From the input artistic portrait image, we extract the texture of the data and apply it to a generic 3D face model. Through this process, a 3D face model with the facial characteristics of the input artistic portrait is created. Through the rotation of the generated 3D face model, faces of artistic portraits from various angles can be obtained. The face of the artistic portrait rotated through the 3D face model is rendered into a 2D image again. Using this approach, we can construct many artistic portraits of various angles, which enriches our artistic portrait dataset for training our model.

Second, we devise a GAN-based model for rotating faces in artistic portraits. Our GAN-based model is pretrained with a well-formed human face dataset so that the model can learn various facial features effectively. We further train the pretrained model using the dataset of artistic portraits obtained in the first step so that the model can learn the characteristics of the faces in artistic portraits. We also devise a loss function so that the styles of artistic portraits are preserved in the result images. The result images preserve the style and characteristics of the face in the input artistic portrait.

We prove our framework in the following aspects. First, we apply our framework to various artistic portraits including photorealistic oil paint portraits, watercolor portraits, well-known artistic portraits and banknote portraits. Second, we show that the similarities between the rotated face and the original face are preserved through metrics, including the Frechet inception distance (FID), recognition ratio and a user study of 50 participants. Third, we show that the faces are rotated according to the specified angles through a pose estimation model that estimates the angle of a rotated face.

The remainder of this paper is organized as follows. Section 2 suggests existing works on face rotation techniques, and Section 3 presents the outline of our framework. We explain our framework and its components as well as the loss functions in Section 4. We present the implementation details and results in Section 5 and analyze our results by comparing existing schemes in Section 6. Finally, we draw our conclusions and present future work in Section 7.

2. Related Work

In this section, we introduce various previous facial rotation techniques. These face rotation techniques can be divided into two categories: those before deep learning and those after deep learning.

2.1. Face Rotation Techniques before Deep Learning

Many facial rotation techniques conducted before deep learning employed a 3D face model as their technical background (Hassner [5], Zhu [6]). The characteristics of the input face image are applied to the 3D face model. Afterward, the input face image is rotated by rotating the 3D face model with the texture of the input face image. Moniz et al. [7] presented another face rotation framework by building a 3D transform matrix that maps each point in a 2D face image to a 3D face model. Although these studies were able to produce face rotation techniques, they revealed a limitation in the distortion and blurry effects in the resulting images caused by the process of converting 2D images into 3D models.

2.2. Latent Space-Based Attribute Control

In recent approaches, many researchers presented a latent space-based approach to control the attributes of a generated face [8–11] or edit the attributes of an input face [12–19].

2.2.1. Face Generation with Pose Control

From the works that produce realistic face images [4,20–22], various schemes are developed to control the attributes of the generated faces, including poses. Harkonen et al. [8] applied principal component analysis (PCA) on the latent space and extracted important latent directions for interpretable GAN control. The attributes they control cover a various range, including wrinkles, hair color, expression and rotation. Various objects including faces are their domain. Shen et al. [11] analyzed the semantics of latent space to control the attributes, including age, eyeglasses, gender and the poses of the generated faces. Shen and Zhou [9] improved their previous work to present a closed-form factorization of the semantics in latent space to control the attributes. They extended their domain to various objects including cars, cats, birds and bedroom scenes. Abdal et al. [10] presented StyleFlow, which extracts attributes from various target faces and composes a new latent vector that controls various attributes including gender, pose, expression and lighting.

These works present attribute-controllable face generation schemes based on StyleGAN frameworks [4,21,22]. These schemes, however, do not concentrate on face rotation or pose control. Therefore, they have a limitation in producing faces of various poses and angles.

2.2.2. Face Editing with Pose Control

From the progress of realistic face image generation, many researchers devised face editing techniques for existing face images based on the schemes that produce latent vectors from existing faces [23,24]. Deng et al. [13] presented a GAN-based approach to control the expression, lighting and pose of a generated face. They constructed a proper latent vector that produced their target attributes and employed a GAN structure for generation. Kowalski et al. [15] presented a simple pose editing method for an input face image that controls the direction of the input face at a slight angle. Yin et al. [12] presented a face editing scheme that rotates the face in two degrees of freedom (DoFs): yaw and pitch. For this purpose, they masked the components of a face using a self-attention mechanism. These facial attentions are included in a loss function for rotating the input face in two DoFs.

Zhu et al. [16] trained a domain-guided encoder that maps the input face image into a proper vector in latent space, which is further processed to a fine-tuned target image through a domain-regularized optimization scheme. Shoshan et al. [17] presented a training scheme for a GAN in a disentangled manner. Therefore, they could edit the input face for the attributes, including the expression, illumination, style and pose, in an explicit way.

Tov et al. [18] optimized an encoder for StyleGAN that pursues various input face editing, including for the pose.

Ju et al. [14] presented an obstacle-robust rotation technique for faces. The various obstacles of a face, including eyeglasses, hands and other objects, are inpainted for the rotated result. Wang et al. [19] presented a face editing scheme using high-fidelity GAN inversion. They applied the distortion consultation approach for their approach. Wu et al. [25] presented a GAN-based retrieval scheme that employs a cross-modal approach. Texts with images are processed through a GAN model that learns the modality-shared features. This model shows a similar approach with ours in that it processes faces with pose codes.

These schemes edit the pose of an input face image instead of generating a new face image from various poses. Similar to the generation schemes, they have a limitation in producing various poses from an input face image.

2.3. Deep Learning-Based Face Rotation Techniques

The progress of deep learning accelerates face rotation techniques at a great scale. Among various deep learning models, generative adversarial network (GAN)-based facial rotation techniques have been actively studied. Research on face rotation technology using GANs was largely divided into reconstruction methods that synthesize the face image at a given angle from the input image and the methods that approximates 3D geometric information from the input image.

2.3.1. Reconstruction-Based Approaches

The reconstruction-based methods employ GANs to create a composite face image at a specified angle [26–29]. Face frontalization, which is one of the most representative reconstruction-based face rotation techniques, synthesizes the frontal face image from the side view of face images in order to improve the accuracy of face recognition techniques.

Tran et al. [26] proposed a DR-GAN that separates the input image's features and the input image's angle to create a frontal image regardless of the input image's angle. Huang et al. [27] proposed a TP-GAN that synthesizes the frontal face image by separately learning the overall outline features of the input image and detailed features such as the eyes, nose and mouth. Hu et al. [28] presented a CAPG-GAN that frontalizes an input face using the heat map. Qian et al. [29] presented an FNM that improves the efficiency of learning by combining labeled and unlabeled data.

Many reconstruction-based methods produce convincing frontalized face images from input images with angles close to the front but did not produce convincing results for the input images with angles close to the side. They also have difficulties in synthesizing results for angles other than from the front.

2.3.2. 3D Geometry-Based Approaches

The 3D geometry-based approaches present face rotation techniques by combining the conventional 3D geometry-based methods and GANs (Yin [30], Deng [31], Cao [32], and Zhou [3]).

Yin et al. [30] proposed an FF-GAN, which employs the coefficients of a 3D deformable model (3DDF) with the existing 3D model and uses them to synthesize rotated face images. Deng et al. [31] proposed a UV-GAN that rotates face images using a UV map. Cao et al. [32] proposed an HF-PIM applied to the result of synthesizing the characteristic information of the input image. Zhou et al. [3] applied the texture feature information of the input image on a 3D model and rerendered the 3D model with the texture feature applied as a 2D image to construct a dataset. They presented Rotate-and-Render, which synthesizes rotated images from multiple angles from the dataset.

The 3D geometry-based approaches were able to produce convincing results for input face images at angles closer to the side than the reconstruction-based methods discussed above. However, additional processing for the 3D model is required, which results in consuming more computational resources.

2.4. The Limitations and Our Approach

Facial rotation techniques using the reconstruction-based approach and 3D geometry-based approach have their respective advantages and limitations. A common limitation of both approaches is that the ground truth (GT) is required to confirm whether the synthesized result is desired by a user or not. For this reason, existing facial rotation techniques have used datasets that have a pair of images from different angles of a specific person along with a frontal image of that person. In order to apply these approaches for artistic portraits, a vast amount of paired artistic portrait datasets is required. Since these datasets are not available, face rotation techniques on artistic portraits have not been actively conducted. Inspired by Zhou et al.'s work [3], which employed a dataset of single face images, we present a scheme for rotating the face of an artistic portrait without additional artistic portraits.

3. Overview

Figure 2 presents the overall structure of our model that synthesizes the faces in artistic portraits to a desired angle. I_n denotes an input face image of an angle n , which is used for training our model, n denotes the angle of the face in an input image and has a value between -90° and $+90^\circ$, I'_m denotes an image rendered in 2D again after applying the input image I_n to a 3D model and rotating it at an angle m , and \hat{I}_p denotes the result image synthesized from the input and angle using the model trained by the input image and the rotated image. Our model can synthesize a rotated face image from an input face image and a desired angle p .

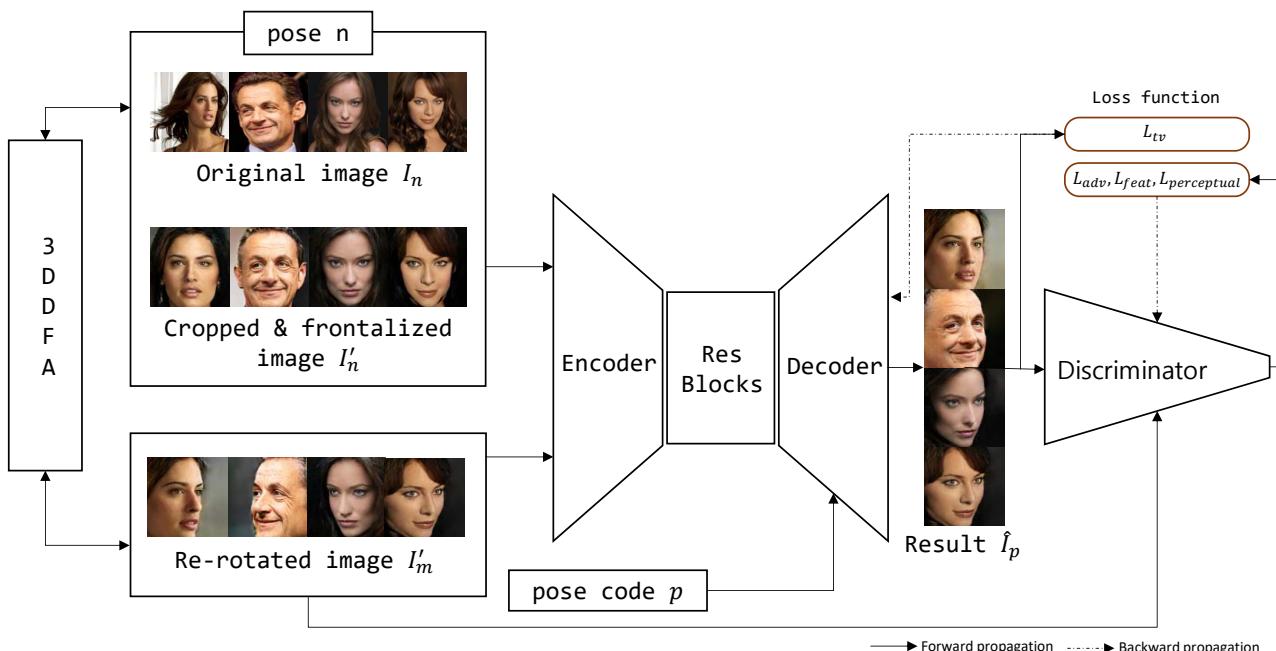


Figure 2. The process of our framework.

Our model is divided into two stages. The first stage is to construct a dataset for training our deep face rotation model. The existing deep facial rotation techniques require datasets of paired images including GT images to improve the quality of the synthesized result. Therefore, it is necessary for us to construct a dataset of a paired images including GT images.

We employ 3DDFA [33], a widely-used open source architecture, to convert a input 2D face into a 3D face model. 3DDFA processes a face image I_n whose angle is n and generates a 3D face model of an angle n where the texture of I_n is applied. The 3D model with the texture of the input image can be rotated at an arbitrary angle m . By rendering the rotated

3D model, a 2D image I'_m of an angle m is produced. To examine the verification of the approach, we reconstruct I'_n by rotating the image of an angle m to an angle n using the above approach. This strategy mimics that of CycleGAN [34].

3DDFA, which builds a 3D face model from a photograph, is employed in our study to build a 3D face model without texture. Generating a rotated face into an arbitrary angle requires both geometry and texture. We reconstruct the geometry of a rotated face using 3DDFA, which was also used in [3]. For rotating faces in artistic portraits, the stroke textures that embed artistic style are still unresolved. We attack this problem using a loss function including the total variation loss term presented in Equation (4).

The images synthesized by this approach are employed to train our model f , which learns the features of faces of different angles. The input image plays the role of the GT to verify that the model f synthesizes the desired result. The left module in Figure 2 illustrates this process. The images synthesized through the first stage are then used as a training and validation dataset for the face rotation model in the second stage.

In the second step, we train and validate our face rotation model that synthesizes a rotated face image from an input image. For this purpose, the dataset synthesized in the first step is used. This dataset consists of I_n , the input face image of angle n , I'_n , the reconstructed face image of an angle n , and I'_m , the rotated face image of an angle m . From this dataset, our model learns the features from the randomly rotated faces to the synthesized, visually pleasing rotated face images. Our model mimics the GAN structure, consisting of a generator with an encoder-decoder structure and a discriminator with an encoder structure. The details of our model are illustrated in Section 4. Our model f synthesizes a face image rotated at an angle p from the images I'_n and I'_m synthesized by applying two 3DDFAs in the first step.

The image \hat{I}_p synthesized in the generator G is processed through the discriminator D with I_n , the GT image. After training, our model synthesizes a rotated face image from the input face images and user-desired angle p . Our model is further described in Section 4, and the synthesized images are presented in Section 5.

4. Face Rotation Model

The input of our model is an artistic portrait image and the desired angle for the portrait. Our model synthesizes an artistic portrait whose face is rotated to the corresponding angle.

4.1. Architecture of the Model

The structure of our model is explained in this section. The generator G of our model consists of an encoder that learns the features from the input image and a decoder that synthesizes a rotated face image through the learned features. We assign a ResNet block between the encoder and decoder to reduce any missing features. The discriminator D , whose structure is similar to that of an encoder, determines whether the synthesized image is synthesized in a similar way to the GT. The discriminator and an encoder are distinguished from the parameters to each layer and the activation function. Figures 3 and 4 illustrate the details of our model.

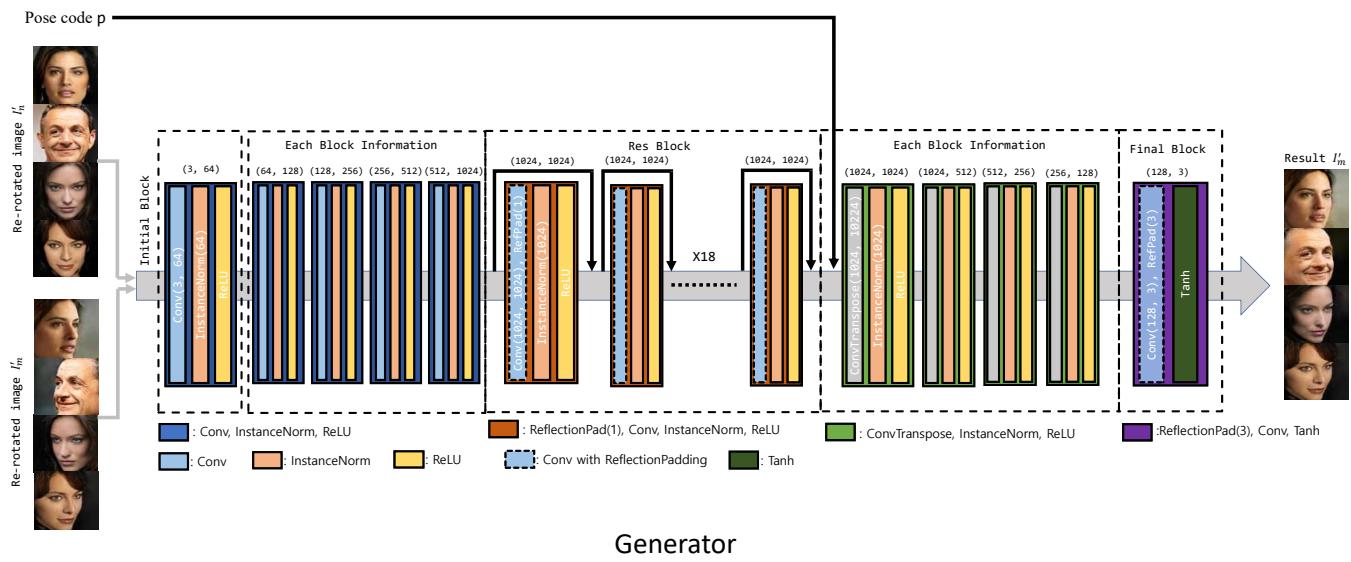


Figure 3. The generator of our face rotation framework.



Figure 4. The discriminator of our face rotation framework.

4.2. Loss Function

We present the loss function employed by our model to synthesize the rotated face image close to the GT. In our model, we devise four loss terms to accomplish our purpose. The total loss function is expressed as a weighted sum of the individual loss terms.

The first loss term in Equation (1) is an adversarial loss. This loss function used in the GAN measures how close the synthesized image is to the GT through the discriminator. Through the adversarial loss, the generator synthesizes an image so close to the GT that the discriminator hardly distinguishes the images from the GT. We formalize the first loss term in Equation (1):

$$L_{adv}(G, D) = \mathbb{E}[\log D(I_p)] + \mathbb{E}[\log(1 - D(\hat{I}_p))] \quad (1)$$

The second loss term in Equation (2) is a feature-matching loss. This loss term compares the features of the input image and the synthesized image through each layer of the discriminator. This loss term enables the synthesized images to preserve the features in the GT properly by minimizing the differences in the features extracted in the discriminators between the input image and the synthesized image. The second loss term is expressed as in Equation (2):

$$L_{feat}(G, D) = \frac{1}{N_D} \sum_{i=1}^{N_D} \|D^i(I_a) - D^i(\hat{I}_a)\|_1 \quad (2)$$

The third loss term in Equation (3) is the perceptual loss. This loss term employs VGGNet pretrained with ImageNet to extract features. Since the features extracted by the pretrained VGGNet preserve the location- and orientation-invariant features of the objects embedded in an image, the comparison between the features extracted from the input and the synthesized image plays an important role in preserving the features in the input image. The third loss term is expressed in Equation (3):

$$L_{perceptual}(G, D) = \frac{1}{N_{vgg}} \sum_{i=1}^{N_{vgg}} \|VGG^i(I_a) - VGG^i(\hat{I}_a)\|_1 \quad (3)$$

The fourth loss term in Equation (4) is the total variation regularization. This term improves the quality of the synthesized image by lowering the difference in brightness between the adjacent pixels of the synthesized image. The existing studies that employ the total variation regularization term for synthesizing photorealistic faces compare the adjacent pixels. However, our approach that synthesizes artistic portraits has a lower chance of brightness variations in the pixels. Therefore, comparing pixels of further distances guarantees improving the quality of the synthesized images. We varied the distance and concluded that the comparison of pixels at a distance of five synthesized the most visually pleasing results. The forth loss term is suggested in Equation (4):

$$L_{tv} = \sum_{c=1}^{C=3} \sum_{w,h=1}^{W,H} \sum_{k=1}^{K=5} |\hat{I}_{w+k,h,c}^f - \hat{I}_{w,h,c}^f| + |\hat{I}_{w,h+k,c}^f - \hat{I}_{w,h,c}^f|. \quad (4)$$

These four loss terms are integrated with appropriate weights obtained through the experiment, and the loss function in our study is suggested in Equation (5):

$$L_{total} = L_{adv} + \lambda_1 L_{feat} + \lambda_2 L_{perceptual} + \lambda_3 L_{tv}. \quad (5)$$

5. Implementation and Results

5.1. Implementation Environments

Our model was implemented on a personal computer with a 4.20-GHz Intel™ Core® i7-7700K CPU, 64 GB of RAM and double NVIDIA™ Titan XP® GPUs. Our model was trained and evaluated using Python and the Pytorch library under the Ubuntu® operating system.

5.2. Training Dataset

Our model was trained using two open face datasets: CASIA-WebFace [35] and MS-Celeb-1M [36]. This model was further fine-tuned using portrait images collected from various web sites including WikiArt (16,000 images from WikiArt and 1000 images from other web

sites). We cropped the collected artistic portrait images so that the faces of the portraits were in the center of the input images. Using these images, we created corresponding 3D face models with the facial characteristics of an artistic portrait. Then, we rotated the 3D face model at narrow angles (-15° , 0° and 15°) to increase the artistic portrait dataset.

5.3. Results

We present our results in the following order.

5.3.1. Results from CelebA Portraits

We applied various portrait images to our model and produced face rotation results.

Our first test images were portraits of the characters in the CelebA dataset. We could find several hand-drawn artistic portraits of the characters contained in CelebA dataset. Most of them are drawn in photorealistic watercolor styles, but some of them are in very distinguishable styles (e.g., the top row and bottom row of Figure 5). We first frontalized the faces in the input portraits (see the 0° column of Figures 5–7). We further rotated the faces at various angles, including -45° , -30° , -15° , 15° , 30° and 45° . We tested six male portraits in Figure 5 and eight female portraits in Figure 6. Furthermore, we tested four pairs of two portraits from the same person in Figure 7, where portraits of the same person drawn in different styles and poses are presented to compare the results. As illustrated in Figure 7, we could preserve the visual personal identities in the rotated portraits. We will examine these further in Section 6.

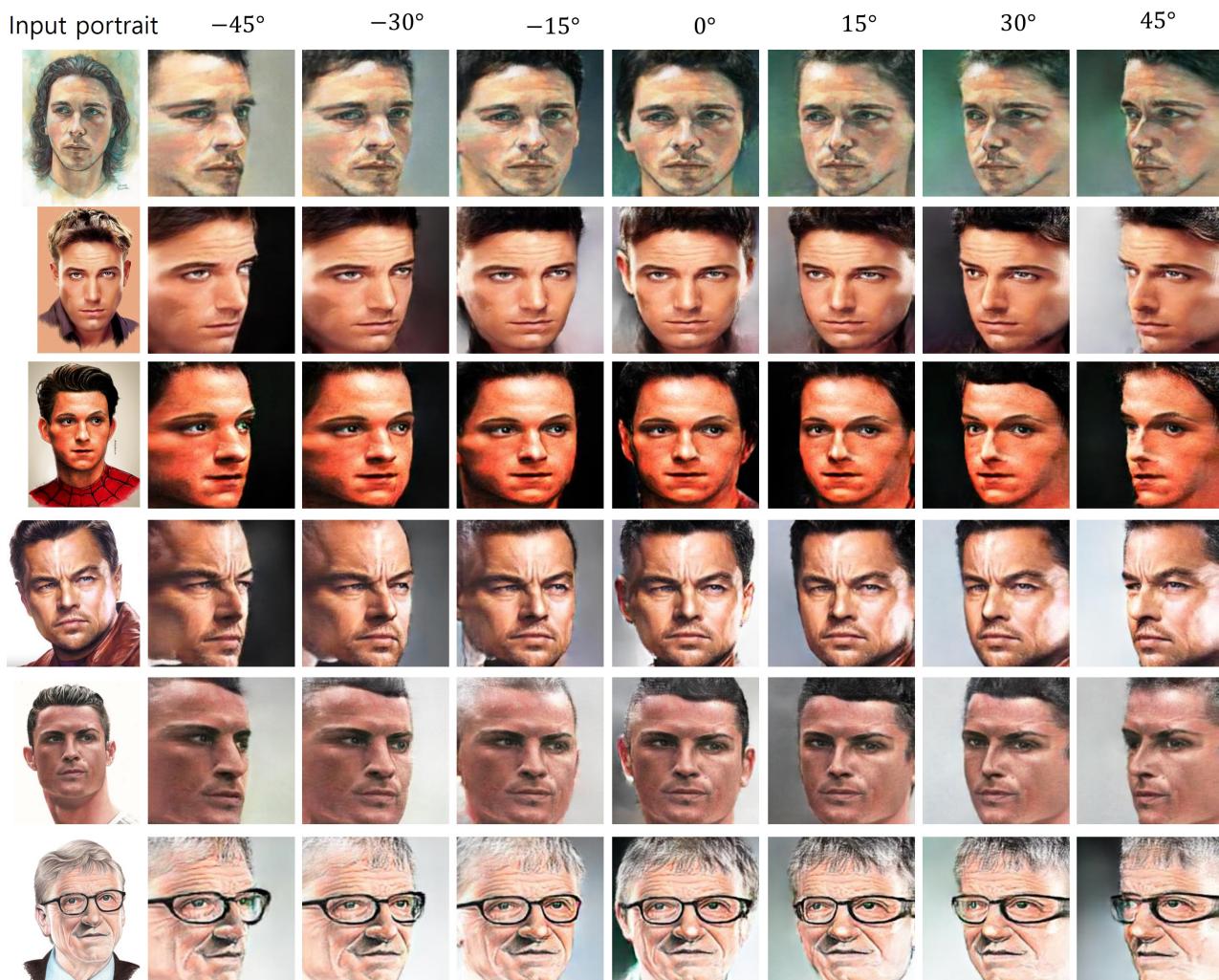


Figure 5. Male portraits from CelebA.

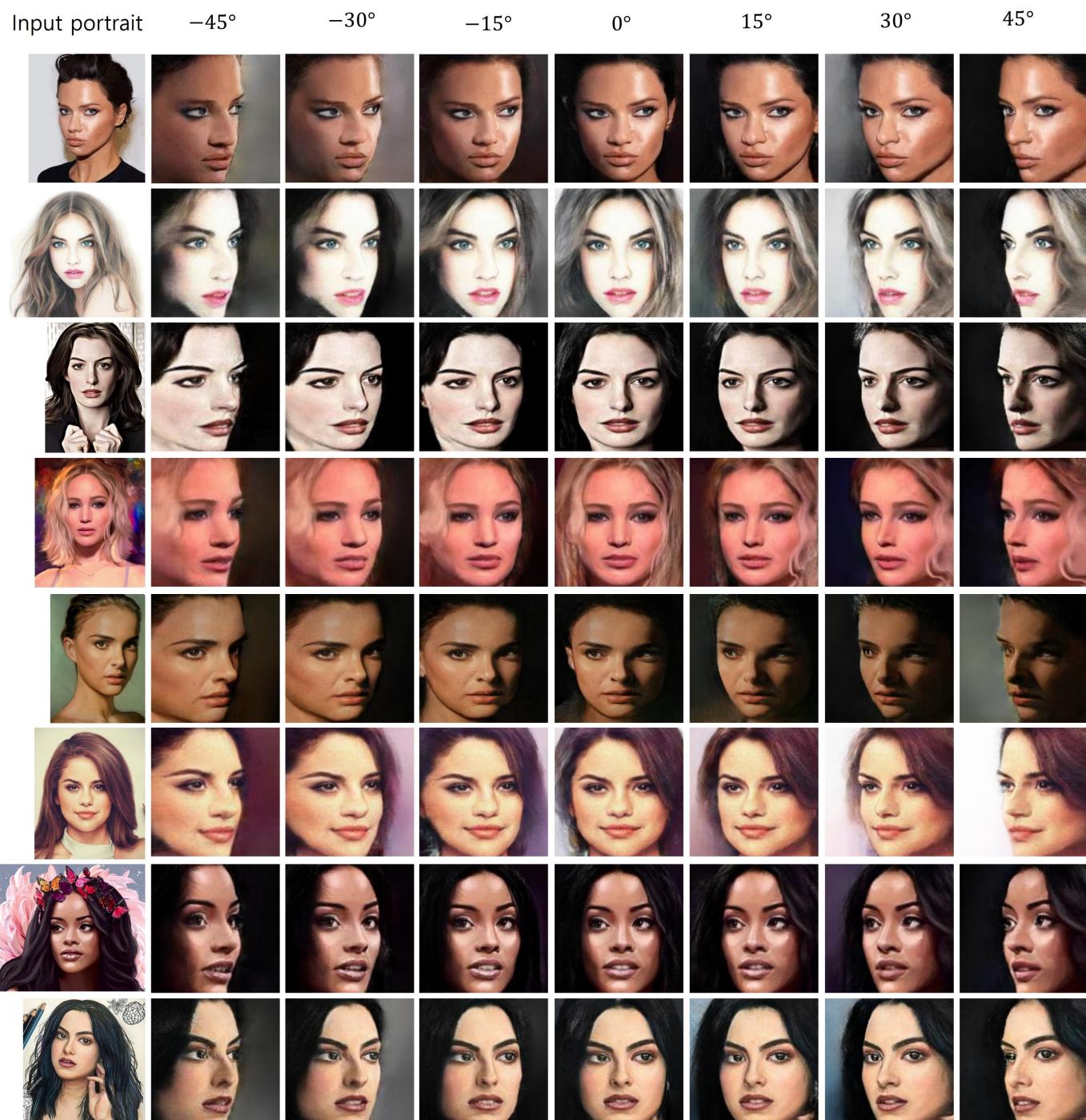


Figure 6. Female portraits from CelebA.



Figure 7. Portraits from CelebA, comparing same person with two portraits.

5.3.2. Results from Famous Artistic Portraits

We test famous artistic portraits, including those from Vincent van Gogh, Rembrandt van Rijn, Pablo Picasso and Frida Kahlo. One interesting characteristic of these portraits is that most of them are drawn with a side view of the target face. Therefore, we focused on the visible side of the face. Therefore, we present six portraits in Figure 8 and four in Figure 9. In Figure 8, we present five rotated faces at 0° , -15° , -30° , -45° and -60° . In the figure, we cover six portraits from Rembrandt's works (Baroque era) to van Gogh's (Impressionism) and Picasso's (Cubism). In Figure 9, we present five rotated faces at

0° , 15° , 30° , 45° and 60° . In the figure, we cover four portraits from van Gogh's work to van Picasso's and Khalo's.

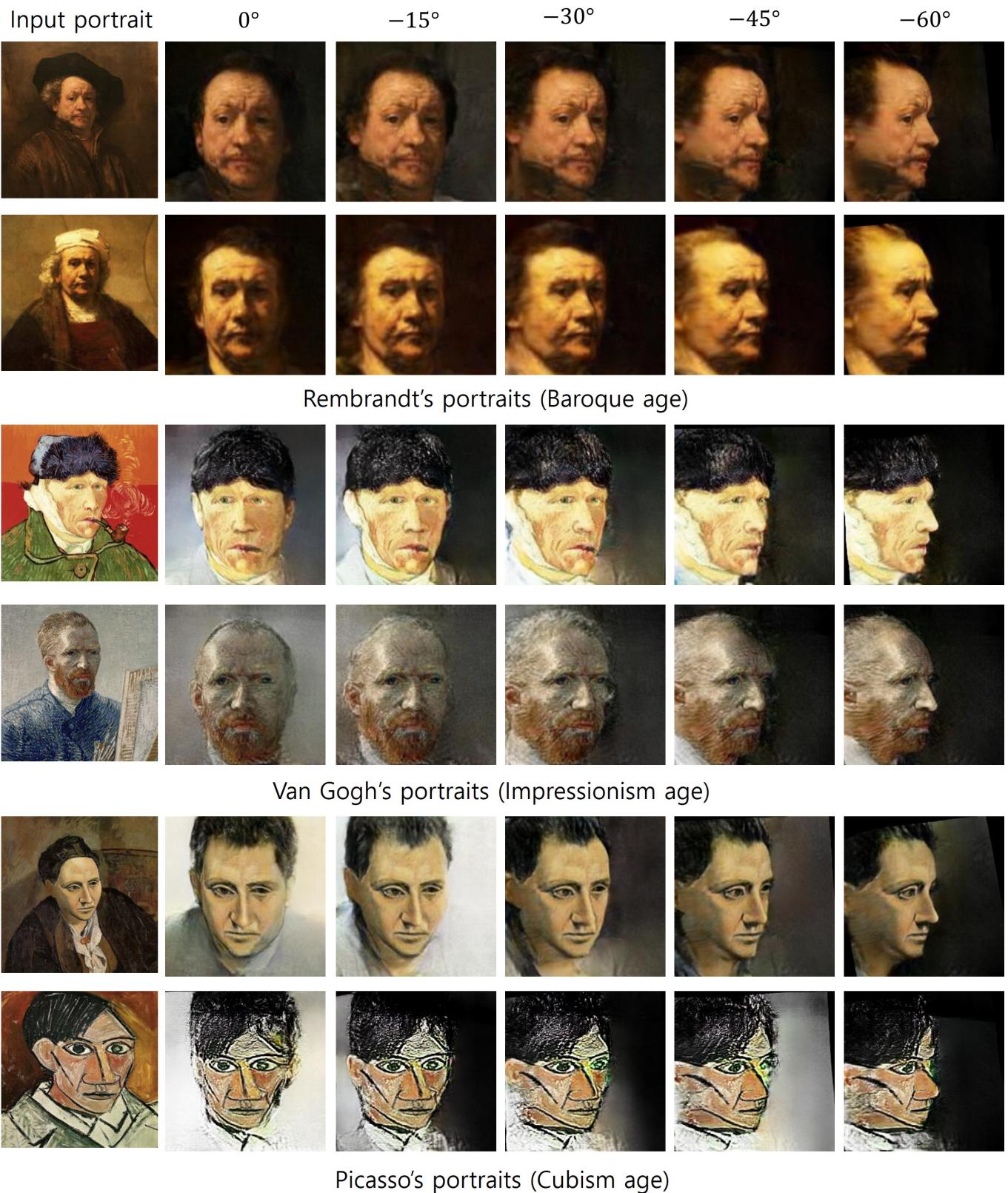


Figure 8. Right-headed artistic portraits which are rotated to the right.

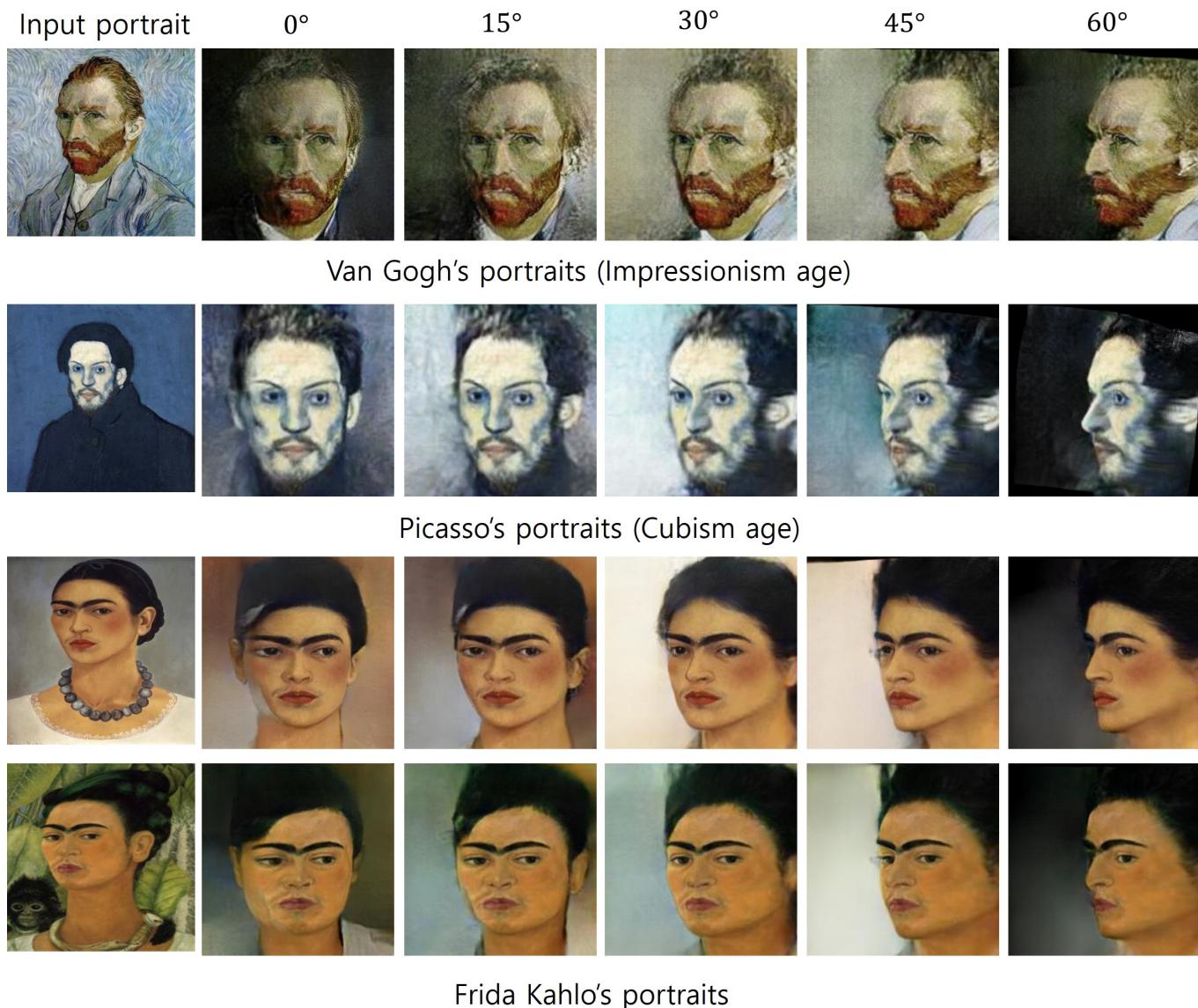


Figure 9. Left-headed artistic portraits which are rotated to the left.

5.3.3. Results from Banknote Portraits

Some of the most frequently seen portraits are on banknotes, which contain portraits of historically significant figures. We applied our face rotation model to the portraits on banknotes. Since the faces in banknote portraits are drawn with one side of a face, we rotated them in one direction. The rotated faces are presented in Figure 10. The faces were rotated at angles of 0° , -15° , -30° , -45° and -60° .

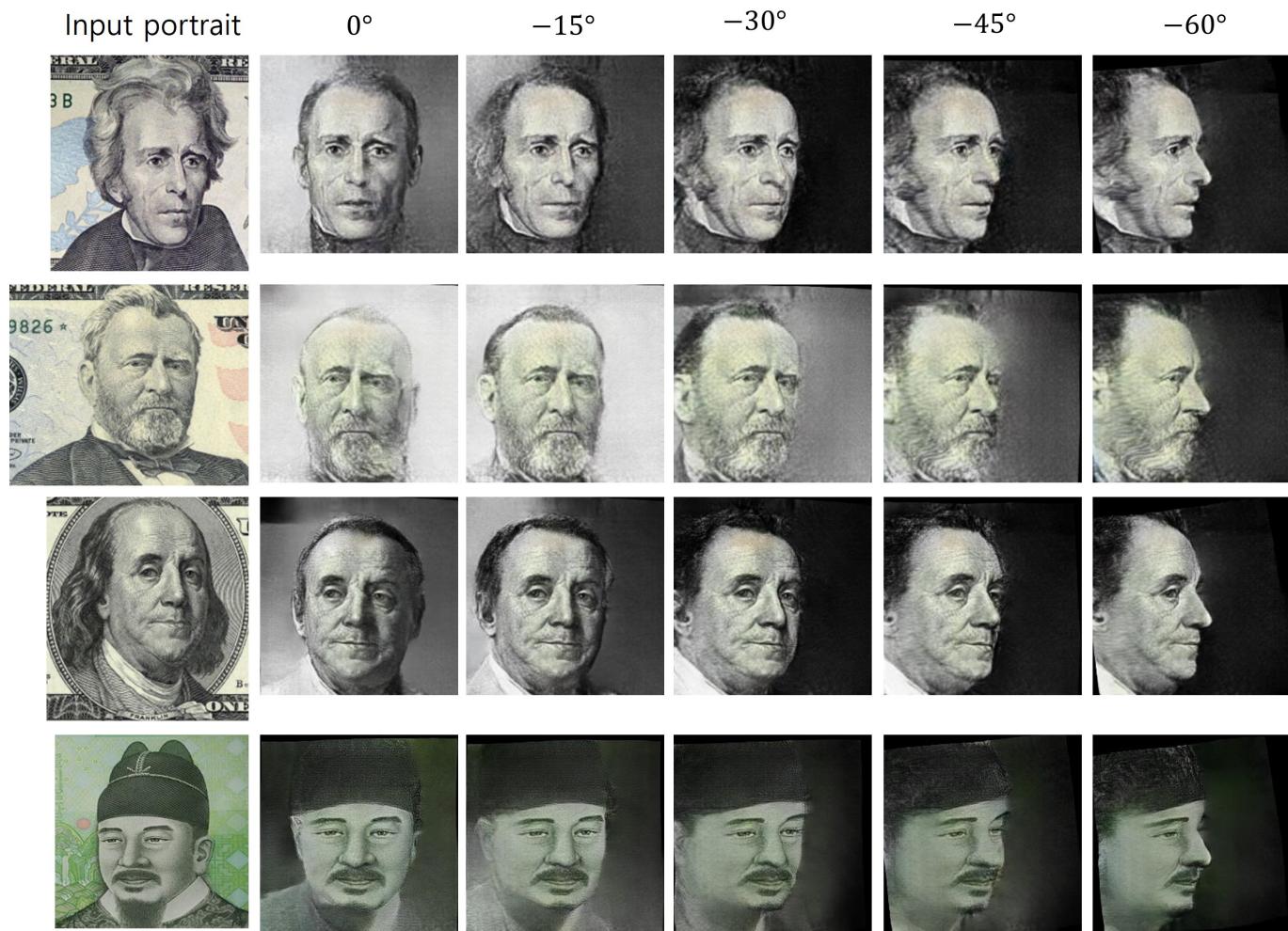


Figure 10. Portraits on banknotes.

5.3.4. Results from Illustrations

We applied our scheme for a series of illustrations of heavily rotated poses in Figure 11. Since artists express their target persons in their favorite poses, rotating these faces in the illustrations is challenging. Our scheme successfully produced rotations of their faces at various angles. The faces were rotated at angles of 60° , 45° , 30° , 15° , 0° , -15° , -30° , -45° and -60° .

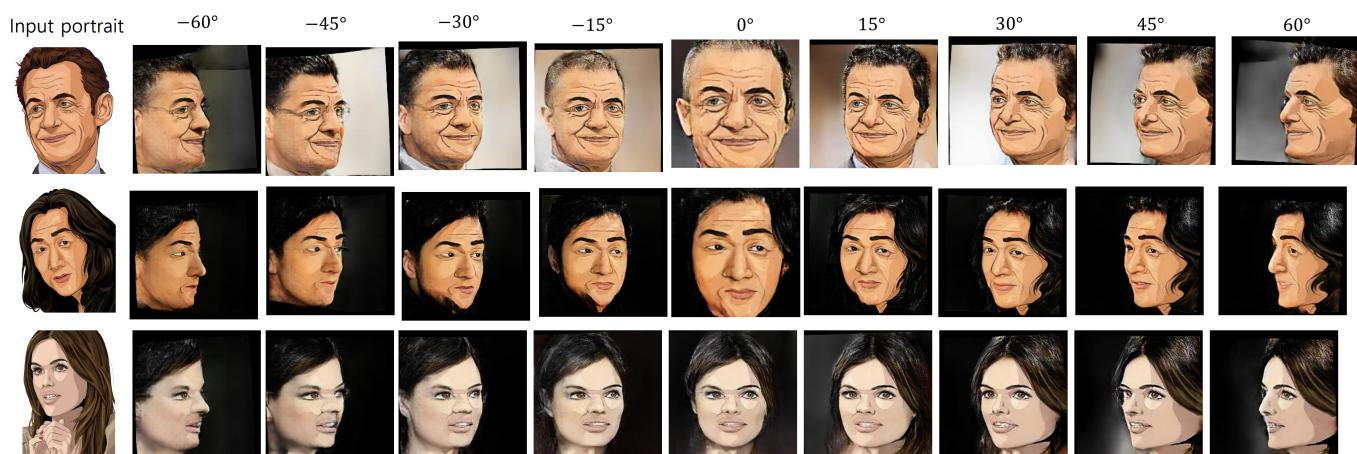


Figure 11. Rotation of illustrations.

6. Analysis

6.1. Comparison

In analyzing our results, we compared our results with those from existing works [3,26,29]. However, some of them covered only frontalization of the face images, while we frontalized the input faces of various poses and compared the results in Figure 12. The FNM [29] and DR-GAN [26], which rotate the input face images to a front angle, produced unfortunately unnatural results. Rotate-and-Render [3], which produces very convincing rotational results, has a limitation in processing the styles observed in artistic portraits. Our results, which preserved the styles in the rotated faces, successfully produced both convincing rotated faces and styles.

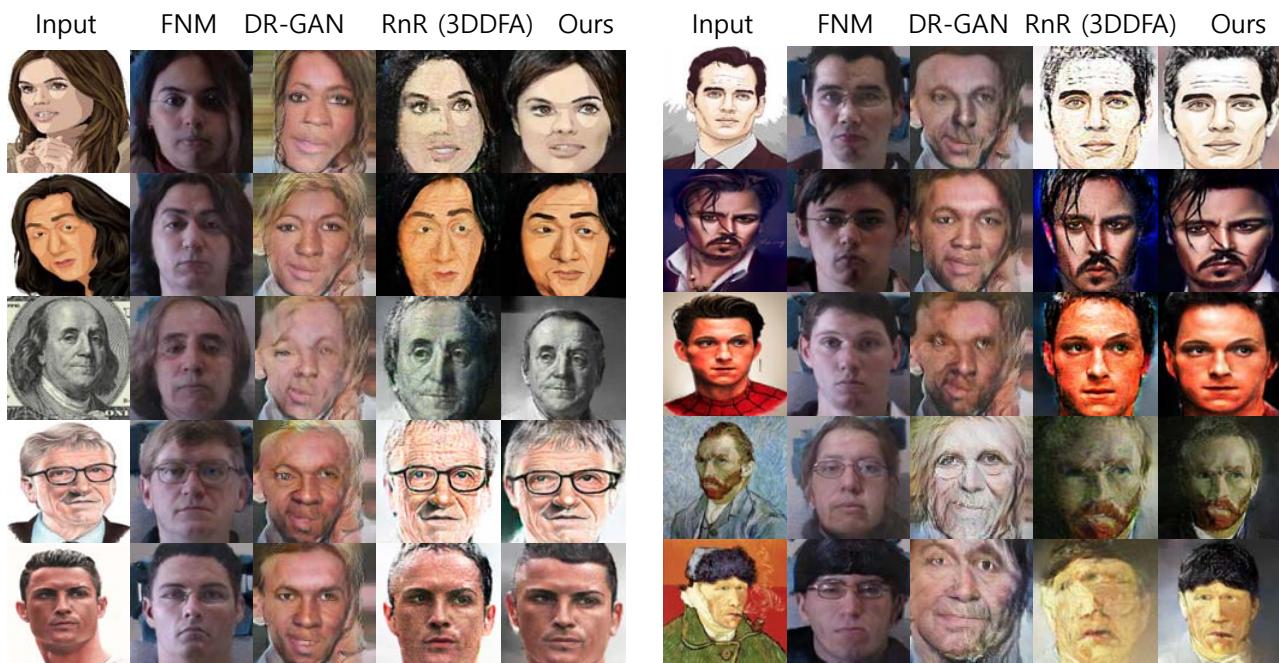


Figure 12. Comparison of our results with existing works: FNM [29], DR-GAN [26] and Rotate-and-Render (RnR) [3]. Rotate-and-Render is implemented based on 3DDFA [33].

6.2. Evaluation

It is very difficult to evaluate the results of face rotation schemes in a quantitative way. We evaluated the results of face rotation schemes from two aspects: (1) the similarity between the original faces and the rotated faces and (2) the angles of the rotated faces. In order to determine the metric that measured the similarity, we examined which metrics were used to evaluate the similarity of faces generated through existing face generation, manipulation and rotation studies and chose the most relevant one. Many face manipulation and style-mixing studies [4,13,15–18,21,22] employ the Frechet inception distance (FID) to measure the similarity between the compared pairs. Zhou et al. [3] also employed the FID for their face rotation results. Therefore, we chose the FID for measuring the similarity between a face and its rotated one. We also employed the recognition ratio with the widely used FGG-Face model pretrained with the CelebA dataset to estimate the similarity.

The angle of the rotated faces was estimated through a pose estimation framework. For this purpose, we employed the OpenFace library [37], which is one of the most widely used head pose estimation schemes. Even though the FID and pose estimation are widely used metrics, we believe these metrics are not sufficient to evaluate the rotated faces. Therefore, we executed a human study that will supplement the limitations of the FID and pose estimation. To increase the confidence of our human study, we hired 50 human participants.

6.2.1. Evaluation 1: Similarity

We evaluated the similarity of the original face and the rotated face using two metrics: the Frechet inception distance (FID) and the recognition ratio:

(1) FID

The FID is defined as in the following equation:

$$FID(X, Y) = ||\mu_X - \mu_Y||^2 - Tr(\Sigma_X + \Sigma_Y - 2\Sigma_X\Sigma_Y),$$

where X is a set of input images and Y is a set of result images. We estimated the FIDs for 39 result images, some of which are presented in Figure 12, and the suggested statistics are in Table 1, where our results in the rightmost column showed the smallest average, maximum and minimum FID values. In the bottom row of Table 1, our results show a minimum FID value for 29 images among 39 images. From these results, we proved that our framework produced better rotated face images than the existing works.

Table 1. Comparison of FID values between the existing models.

FID	FNM ([29])	DR-GAN ([26])	Rotate-and-Render ([3])	Ours
Average	461.7	507.9	459.7	382.2
Std.	183.4	193.2	155.8	156.0
Maximum	1041.4	1095.2	805.1	714.3
Minimum	225.9	248.4	180.8	131.0
No. of minimum	5	3	2	29

(2) Recognition ratio

We employed VGG-Face [38] pretrained with the CelebA dataset. Therefore, this model can recognize faces in the CelebA dataset. The recorded baseline top-five recognition ratio of this model was 93.2%. We applied the 22 images in Figures 5–7 to this model and estimated the top-5 recognition ratio. We compared our results with those from Rotate-and-Render (RnR) [3], which was implemented using 3DDFA [33]. We compared the faces rotated at -45° , -30° , -15° , 0° , 15° , 30° and 45° in Table 2, where the recognition ratio decreased according to the increase in the rotation angle. Our results showed a higher recognition ratio than that from Rotate-and-Render.

Table 2. Comparison of top-5 recognition ratios between our model and Rotate-and-Render [3].

Model	-45°	-30°	-15°	0°	15°	30°	45°
Rotate-and-Render	58.2%	67.5%	78.9%	83.3%	77.9%	68.8%	59.7%
Ours	68.4%	75.3%	86.4%	89.2%	83.9%	78.3%	67.9%

6.2.2. Evaluation 2: Pose Estimation

We employed the OpenFace library [37] for pose estimation. We test 39 result images at $-45^\circ \sim 45^\circ$ and 20 images at -60° and 60° . The statistics on this estimation are presented in Table 3. For $-30^\circ \sim 30^\circ$, our results showed differences of less than 2° , which is a very reliable value. Our method showed differences of about 5° for 45° and -45° and about 20° for 60° and -60° . From these values, we concluded that our model produced very concise results for $-30^\circ \sim 30^\circ$. However, our method became unreliable with the increase in the rotation angle.

Table 3. The yaw values estimated from our rotated face images.

	60°	45°	30°	15°	0°	-15°	-30°	-45°	-60°
Avg.	41.1°	39.3°	29.1°	16.8°	0.7°	-16.3°	-29.4°	-38.2°	-41.5°
Std.	7.6	5.4	4.3	3.2	2.7	3.6	5.6	7.9	10.7
Min.	26°	26°	20°	10°	-5°	-6°	-11°	-7°	-15°
Max.	55°	52°	40°	23°	6°	-22°	-37°	-49°	-57°
Med.	41°	40°	29°	16.5°	0.5°	-18°	-30°	-39°	-45°

6.2.3. Evaluation 3: User Study

We hired 50 human participants to estimate the quality of our results, where 23 of them were in their twenties and 27 were in their thirties. Of these participants, 26 of them were female, and 24 were male. We presented them the images in Figure 12 and asked the following three questions:

- Q1. Front: Are these faces facing front? Answer one if you agree or zero otherwise.
- Q2. Quality: Evaluate the quality of the rotated faces with a 10-point metric. Mark one for the worst and 10 for the best.
- Q3. Identity: Do the rotated faces resemble the original faces? Evaluate them with a 1-point metric. Mark one for the worst and 10 for the best.

We collected the answers from 50 participants and summarized them in Table 4.

Table 4. The results of our user study. The answers are averaged from 10 images in Figure 12 from 50 participants. The blue figures in each row are the best scores.

	FNM ([29])	DR-GAN ([26])	Rotate-and-Render ([3])	Ours
Q1. Front	8.98	9.02	9.54	9.58
Q2. Quality	2.48	1.90	8.60	8.98
Q3. Identity	2.70	1.50	8.88	9.03

We analyzed the results of the user study as follows:

- Q1. The answer to Q1 denotes the number of faces that the participants agreed that the rotated face was facing front. The maximum value was 10, and the minimum value was 0. The average answer in the models in Figure 12 varied in the range of 8.97~9.60. We can conclude that most of the participants agreed that the result images were facing front.
- Q2. The answer to Q2 denotes the average answers on the quality of the rotated faces. The figures in the middle row of Table 4 are the average value for 10 rotated faces from 30 participants. The maximum value was 10, and the minimum value was 1. For Q2, we can conclude that our model showed improved quality compared with the existing schemes, especially Rotate-and-Render [3].
- Q3. The answer to Q3 denotes the average answers about matching the rotated face and the input face. The figures in the bottom row of Table 4 are the average values for 10 rotated faces from 30 participants. The maximum value was 10, and the minimum value was 1. For Q3, we can conclude that the Rotate-and-Render model [3] and our model could preserve the identity of the input face after rotation.

6.3. Limitation

Our approach that rotates the faces in the input portrait at arbitrary angles has several limitations:

Limitation 1: Input with highly rotated poses. We had several input images with highly rotated poses (see Figure 13a). The hardly seen side of the highly rotated poses was not reconstructed in a reasonable way, since reconstruction of the

information on that side was hardly achievable. The rotation of such input images would require a constraint for the symmetry of a human face. This can be a future research direction for our studies.

- Limitation 2:** Rotation greater than 60° . The rotation of a face by more than 60° produced unacceptable results, where the identity of the face could be confused or some unwanted artifacts appeared. We believe the information for the highly rotated face was not sufficient for reconstructing the rotated face.
- Limitation 3:** Losing attachments. Some attachments on the head of a rotated face, including long hair, a pipe in a mouth or flower decorations on the head, were lost. Since our framework concentrated on face rotation rather than head rotation, it lost the attachments on the head. This limitation can be resolved through a head rotation framework that reconstructs the information for the head.

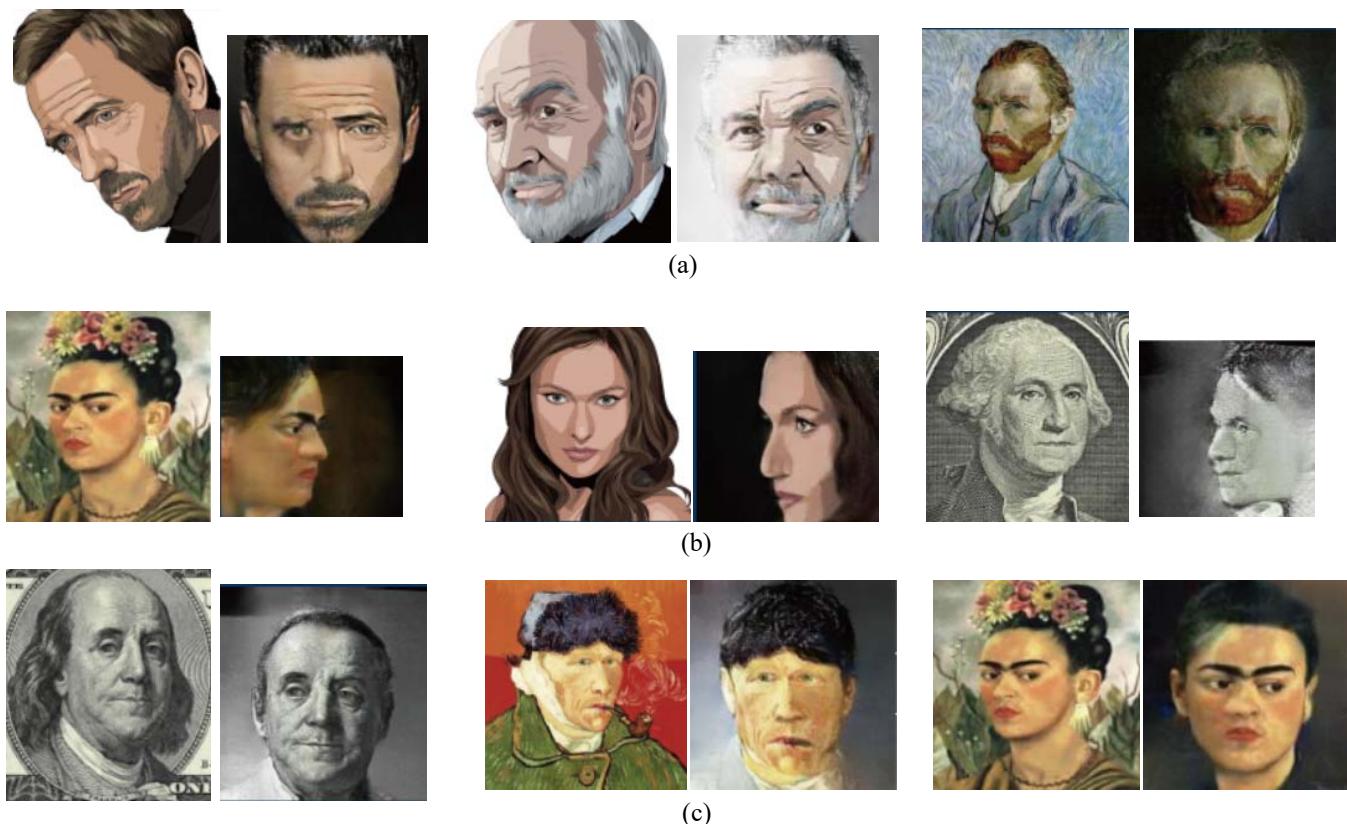


Figure 13. The limitations of our approach. (a) The input faces with highly rotated poses had problems after frontalization. (b) Rotating faces more than 60° produced poor results. (c) The attachments on the face were lost after rotation (Franklin lost his hair, Gogh lost his pipe, and Kahlo lost her flowers).

7. Conclusions and Future Work

We have presented a face rotation technique for artistic portraits using a GAN-based approach. For successful rotation, we adopted a 3D face model in order to construct a training dataset for artistic portraits. In order to preserve the styles on the artistic portraits, we devised various loss functions, including adversarial loss, feature matching loss, perceptual loss and total variation regularization. Using these terms, our model successfully rotated the faces in various portraits to diverse angles. We applied our model to photorealistic oil paint portraits, watercolor portraits, well-known portrait artworks and banknote portraits and rotated their faces.

Our primary future direction is to extend our approach to head rotation, which rotates attachments including hats, hair and decorations, as well as the faces. This rotation can help produce various visual contents from a face in a portrait. We also plan to extend our

model by considering other constraints such as the symmetry to rotate faces whose initial poses are rotated at a great angle.

Author Contributions: Conceptualization, H.K.; Data curation, J.K.; Software, H.K.; Supervision, H.Y.; Visualization, J.K.; Writing—original draft, H.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Sangmyung Univ. at 2020.

Data Availability Statement: Not applicable.

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

References

1. Available online: <https://en.wikipedia.org/wiki/Hedcut> (accessed on 1 September 2002).
2. Kim, D.; Son, M.; Lee, Y.; Kang, H.; Lee, S. Feature-guided image stippling. In *Computer Graphics Forum*; Blackwell Publishing Ltd.: Oxford, UK, 2008; pp. 1209–1216.
3. Zhou, H.; Liu, J.; Liu, Z.; Liu, Y.; Wang, X. Rotate-and-render: Unsupervised photo-realistic face rotation from single-view images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 5911–5920.
4. Karras, T.; Laine, S.; Aila, T. A Style-Based Generator Architecture for Generative Adversarial Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 4401–4410.
5. 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, MA, USA, 7–12 June 2015; pp. 4295–4303.
6. Zhu, X.; Lei, Z.; Yan, J.; Yi, D.; Li, S. High-fidelity pose and expression normalization for face recognition in the wild. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 787–796.
7. Moniz, J.; Beckham, C.; Rajotte, S.; Homari, S.; Pal, C. Unsupervised depth estimation, 3d face rotation and replacement. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 9759–9769.
8. Harkonen, E.; Hertzmann, A.; Lehtinen, J.; Paris, S. GANSpace: Discovering Interpretable GAN Controls. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 9841–9850.
9. Shen, Y.; Zhou, B. Closed-Form Factorization of Latent Semantics in GANs. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Virtual, 19–25 June 2021; pp. 1532–1540.
10. Abdal, R.; Zhu, P.; Mitra, N.; Wonka, P. StyleFlow: Attribute-conditioned Exploration of StyleGAN-Generated Images using Conditional Continuous Normalizing Flows. *ACM Trans. Graph.* **2021**, *40*, 1–21. [[CrossRef](#)]
11. Shen, Y.; Gu, J.; Tang, X.; Zhou, B. Interpreting the Latent Space of GANs for Semantic Face Editing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 9243–9252.
12. Yin, Y.; Jiang, S.; Robinson, J.; Fu, Y. Dual-Attention GAN for Large-Pose Face Frontalization. In Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020), Buenos Aires, Argentina, 16–20 November 2020; pp. 249–256.
13. Deng, Y.; Yang, J.; Chen, D.; Wen, F.; Tong, X. Disentangled and Controllable Face Image Generation via 3D Imitative-Contrastive Learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2020, Seattle, WA, USA, 14–19 June 2020; pp. 5154–5163.
14. Ju, Y.-J.; Lee, G.-H.; Hong, J.-H.; Lee, S.-W. Complete Face Recovery GAN: Unsupervised Joint Face Rotation and De-Occlusion from a Single-View Image. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 4–8 January 2022; pp. 1173–1183.
15. Kowalski, M.; Garbin, S.; Estellers, V.; Baltrusaitis, T.; Johnson, M.; Shotton, J. Config: Controllable neural face image generation. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2020; pp. 299–315.
16. Zhu, J.; Shen, Y.; Zhao, D.; Zhou, B. In-Domain GAN Inversion for Real Image Editing. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2020; pp. 592–608.
17. Shoshan, A.; Bhonker, N.; Kviatkovsky, I.; Medioni, G. GAN-Control: Explicitly Controllable GANs. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Virtual, 11–17 October 2021; pp. 14083–14093.
18. Tov, O.; Alaluf, Y.; Nitzan, Y.; Patashnik, O.; Cohen-Or, D. Designing an Encoder for Stylegan Image Manipulation. *ACM Trans. Graph.* **2021**, *40*, 1–40. [[CrossRef](#)]
19. Wang, T.; Zhang, Y.; Fan, Y.; Wang, J.; Chen, Q. High-Fidelity GAN Inversion for Image Attribute Editing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 21–24 June 2022; pp. 11379–11388.
20. Karras, T.; Aila, T.; Laine, S.; Lehtinen, J. Progressive Growing of GANs for Improved Quality, Stability, and Variation. In Proceedings of the International Conference on Learning and Representation (ICLR) 2018, Vancouver, BC, Canada, 30 April–3 May 2018.

21. Karras, T.; Laine, S.; Aittala, M.; Hellsten, J.; Lehtinen, J.; Aila, T. Analyzing and Improving the Image Quality of StyleGAN. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; pp. 8110–8119.
22. Karras, T.; Aittala, M.; Laine, S.; Harkonen, E.; Hellsten, J.; Lehtinen, J.; Aila, T. Alias-Free Generative Adversarial Networks. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 852–863.
23. Abdal, R.; Qin, Y.; Wonka, P. Image2StyleGAN: How to Embed Images Into the StyleGAN Latent Space? In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 4432–4441.
24. Abdal, R.; Qin, Y.; Wonka, P. Image2StyleGAN++: How to Edit the Embedded Images? In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, 14–19 June 2020; pp. 8296–8305.
25. Wu, F.; Jing, X.-Y.; Wu, Z.; Dong, X.; Luo, X.; Huang, Q.; Wang, R. Modality-specific and shared generative adversarial network for cross-modal retrieval. *Pattern Recognit.* **2020**, *104*, 107335. [[CrossRef](#)]
26. Tran, L.; Yin, X.; Liu, X. Disentangled representation learning GAN for pose-invariant face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1415–1424.
27. Huang, R.; Zhang, S.; Li, T.; He, R. Beyond face rotation: Global and local perception gan for photorealistic and identity preserving frontal view synthesis. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2439–2448.
28. Hu, Y.; Wu, X.; Yu, B.; He, R.; Sun, Z. Pose-guided photorealistic face rotation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8398–8406.
29. Qian, Y.; Deng, W.; Hu, J. Unsupervised face normalization with extreme pose and expression in the wild. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 9851–9858.
30. Yin, X.; Yu, X.; Sohn, K.; Liu, X.; Chandraker, M. Towards large-pose face frontalization in the wild. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 3990–3999.
31. 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 on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7093–7102.
32. Cao, J.; Hu, Y.; Zhang, H.; He, R.; Sun, Z. Learning a high fidelity pose invariant model for high-resolution face frontalization. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 2872–2882.
33. Zhu, X.; Liu, X.; Lei, Z.; Li, S. Face alignment in full pose range: A 3d total solution. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *41*, 78–92. [[CrossRef](#)] [[PubMed](#)]
34. Zhu, J.; Park, T.; Isola, P.; Efros, A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
35. Yi, D.; Lei, Z.; Liao, S.; Li, S. Learning face representation from scratch. *arXiv* **2014**, arXiv:1411.7923.
36. Guo, Y.; Zhang, L.; Hu, Y.; He, X.; Gao, J. Ms-celeb- 1m: A dataset and benchmark for large-scale face recognition. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2016; pp. 87–102.
37. Amos, B.; Ludwiczuk, B.; Satyanarayanan, M. Openface: A general-purpose face recognition library with mobile applications. *CMU Sch. Comput. Sci.* **2016**, *6*, 20.
38. Parkhi, O.M.; Vedaldi, A.; Zisserman, A. Deep Face Recognition. In Proceedings of the British Machine Vision Conference, Swansea, UK, 7–10 September 2015; pp. 41.1–41.12.