




Article

GAN-Based ROI Image Translation Method for Predicting Image after Hair Transplant Surgery

Do-Yeon Hwang ¹, Seok-Hwan Choi ¹ , Jinmyeong Shin ¹ , Moonkyu Kim ^{2,*} and Yoon-Ho Choi ^{1,*} 

¹ School of Computer Science and Engineering, Pusan National University, Busan 46242, Korea; h1d2y3@pusan.ac.kr (D.-Y.H.); danialsh@pusan.ac.kr (S.-H.C.); sinryang@pusan.ac.kr (J.S.)

² Kyungpook National University Hospital Hair Transplantation Center, Daegu 41913, Korea

* Correspondence: moonkim@knu.ac.kr (M.K.); yhchoi@pusan.ac.kr (Y.-H.C.)

Abstract: In this paper, we propose a new deep learning-based image translation method to predict and generate images after hair transplant surgery from images before hair transplant surgery. Since existing image translation models use a naive strategy that trains the whole distribution of translation, the image translation models using the original image as the input data result in converting not only the hair transplant surgery region, which is the region of interest (ROI) for image translation, but also the other image regions, which are not the ROI. To solve this problem, we proposed a novel generative adversarial network (GAN)-based ROI image translation method, which converts only the ROI and retains the image for the non-ROI. Specifically, by performing image translation and image segmentation independently, the proposed method generates predictive images from the distribution of images after hair transplant surgery and specifies the ROI to be used for generated images. In addition, by applying the ensemble method to image segmentation, we propose a more robust method through complementing the shortages of various image segmentation models. From the experimental results using a real medical image dataset, e.g., 1394 images before hair transplantation and 896 images after hair transplantation, to train the GAN model, we show that the proposed GAN-based ROI image translation method performed better than the other GAN-based image translation methods, e.g., by 23% in SSIM (Structural Similarity Index Measure), 45% in IoU (Intersection over Union), and 42% in FID (Fréchet Inception Distance), on average. Furthermore, the ensemble method that we propose not only improves ROI detection performance but also shows consistent performances in generating better predictive images from preoperative images taken from diverse angles.

Keywords: hair loss; hair transplant surgery; image translation; image segmentation



check for updates

Citation: Hwang, D.-Y.; Choi, S.-H.; Shin, J.; Kim, M.; Choi, Y.-H. GAN-Based ROI Image Translation Method for Predicting Image after Hair Transplant Surgery. *Electronics* **2021**, *10*, 3066. <https://doi.org/10.3390/electronics10243066>

Academic Editor: Prasan Kumar Sahoo

Received: 27 October 2021

Accepted: 5 December 2021

Published: 9 December 2021

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



Copyright: © 2021 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

Hair loss, i.e., no hair in areas where there should normally be hair, is not only a matter of visual appearance but also negatively affects individual self-esteem [1]. Accordingly, interest in hair loss treatment is increasing irrespective of age, race, and gender. In particular, because hair transplant surgery can treat various symptoms of hair loss and improve the hairlines of patients after surgery, choosing hair transplant surgery is becoming more common [2]. Nevertheless, most of the patients who visit a hair loss treatment hospital or a hair transplant hospital cannot easily decide on a hair transplant because they cannot anticipate the appearance of the hair after the surgery.

As the field of analyzing and diagnosing medical images using Artificial Intelligence (AI) technology has become common, various studies have been conducted recently to analyze and diagnose the hair condition of hair loss patients using AI technology [3–5]. To date, research for analyzing and diagnosing the hair status of patients using AI technology has been conducted to estimate the severity and presence of hair loss and to detect diverse scalp hair symptoms, i.e., the hair medical status. Even though these research results can

be used to measure a patient's hair loss status and to analyze and predict the possibility of hair loss in the future, they cannot help resolve concerns about artificial and unnatural hair after surgery for patients considering hair transplant surgery.

To solve this problem, we propose a new deep learning-based image translation method to predict images after hair transplantation from images before hair transplant surgery. In the task, a predictive image generated through the deep learning-based model should be a sufficient reference for patients who want hair transplant surgery, so it is important to generate a clear predictive image without blurred regions. For this reason, even though deep learning-based models applying autoencoders showed excellent performance [6,7], the models were not suitable for this task in that the models produce blurred images [8]. Conversely, a generative adversarial network (GAN) can generate clearer images by applying a learning method that can solve the blurring problem [9]. Therefore, we selected GAN as a model to achieve our goal. Using the GAN model, which is an unsupervised deep learning technique, we analyzed medical image data before and after hair transplant surgery and generated predictive images after hair transplant surgery from images before hair transplant surgery. They can be used as data to help patients decide on surgery and to help medical staff diagnose issues before hair transplant surgery.

Among the deep learning models, the GAN model, which is widely used as an image generation and translation technology, learns two models (generator and discriminator) via an adversarial process at the same time. The GAN model is composed of a generator that generates or converts images by receiving various noises or images as input, and a discriminator that receives the image created by the generator as an input and determines whether it is a real image or an image created by the generator. By preparing two datasets and putting one dataset (images before hair transplantation) as an input to the generator, it is possible to train a generator that produces an output having the characteristics of another dataset, a so-called image translation function.

However, when the GAN model is used for analyzing medical images before and after hair transplant surgery, the generator generally uses not only the region of interest (ROI) of the image but also the image regions including the non-ROI, clothes and ornaments, and so on. That is, such an image translation method using the original image as the input data results in converting not only the hair transplant surgery region, which is the ROI, but also the other image regions, which are not the ROI [10]. For example, as shown in Figure 1a, when the patient's hair image before hair transplant surgery is used to predict the patient's hair image after hair transplant surgery, the nonsurgical region, such as the shape of the patient's clothes and the state of wearing a headband, has also changed.

In this paper, to solve the problem that occurs when translating an original image before hair transplant surgery into an image after hair transplant surgery, we proposed a novel GAN-based ROI image translation method that converts only the ROI and retains the image for the non-ROI. Through the following three phases, only the hair transplant region, i.e., the ROI, is converted to generate a predictive image after hair transplant surgery as shown in Figure 1b: (1) image translation, which generates a naive prediction for the input image before hair transplant surgery; (2) image segmentation, which predicts the surgical area, i.e., the ROI, in the input image before hair transplant surgery using a segmentation model; and (3) image synthesis, which synthesizes the input image and the naive prediction for only the ROI to generate the predictive image translated only on the hair transplant region.

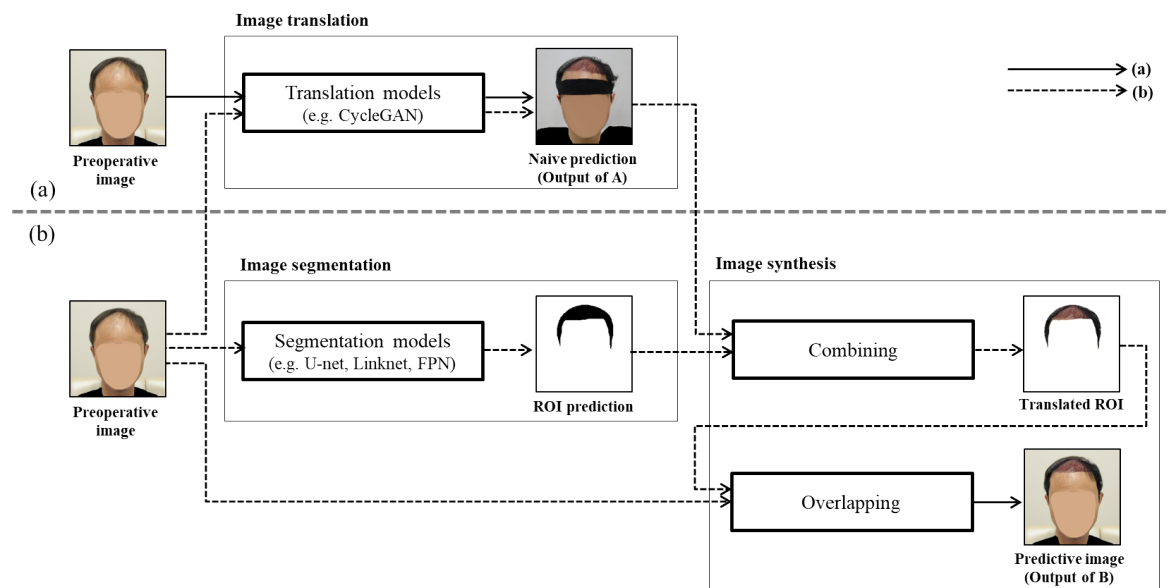


Figure 1. Comparison of the GAN-based ROI image translation method with the existing GAN-based image translation method: (a) using the original medical image before hair transplant surgery and (b) using both the original medical image before hair transplant surgery and the ROI for hair transplant surgery.

To evaluate the performance of the proposed method, we collected medical image data from various medical institutions in Korea for training and testing the proposed method, which is the most important factor in determining the performance of the GAN-based image translation model. The medical image data consisted of two categories, 1768 images before hair transplantation and 1064 images after hair transplantation. The before images were taken with the head facing forward. Furthermore, the before images included people with varying degrees of hair loss from early to late stages, and the age range varied from young to old. Images after hair transplantation were also taken with the head facing forward. The after images had a wide range of surgeries, from those who had surgery intensively on the hairline to those who had surgery all over the head. Furthermore, to protect the privacy of patients, personal identification information, e.g., face and clothes, was blurred to be de-identified as shown in Figure 1. Specifically, 1394 images before hair transplantation and 896 images after hair transplantation were collected and used to train the GAN-based image translation model; 374 images before hair surgery were used in the test process and 168 postoperative images for a standard of comparison. In addition, annotations of training/test datasets were made to train segmentation models designed for detecting ROI. The annotations of preoperative images were made by forming a polygon along the outline of the head shape and including the upper part of the forehead (slightly below the hairline). Through commonly used performance measurement metrics such as SSIM [11] (Structural Similarity Index Measure), IoU [12] (Intersection over Union), and FID [13] (Fréchet Inception Distance), we showed that the proposed GAN-based ROI image translation method had superior image translation performance compared to other GAN-based image translation methods in generating a predictive image after hair transplant surgery. In addition, we proposed an ensemble method to improve image segmentation performance and showed that the ensemble method was robust for images of various angles.

The rest of the paper is organized as follows. In Section 2, we overview the GAN-based image translation methods. In Section 3, we show the operation of the proposed GAN-based ROI image translation method in detail. In Section 4, we show evaluation results of the proposed method under various conditions. In Section 5, we analyze the limitations of the proposed method. Finally, we summarize this paper in Section 6.

2. Related Work

Recently, image translation research has rapidly grown with the advancement of GAN models. For example, Isola et al. [9] proposed a pix2pix model that applied a so-called conditional GAN for image translation. Pix2pix converts an input image with a specific subject X into an output image of a desired subject Y through a conditional GAN. However, pix2pix requires a lot of paired data. To solve this problem, Zhu et al. [14] proposed CycleGAN that can learn with unpaired data through cycle-consistency loss. Even though the CycleGAN model showed a good performance for diverse subjects [15,16], the performance was limited because it converted areas other than the ROI [10]. To resolve this problem, models using the attention technique [10,17] and using shared-space to convert images [18,19] were proposed. However, such models did not resolve the problem due to some constraints, e.g., that the attention and the ROI should coincide.

With the advancement of deep neural networks, GAN-based image translation models have been used in hair translation research. Jo et al. [20] presented a model that allowed users to convert images in free-form and showed that the model appropriately converted the hairstyle of the source image based on the sketch drawn by the user. The function that hair translation can be freely performed by users is similar to [20], but MichiGAN [21] can translate hair on a source image preserving the background through sophisticated neural network architecture. In addition, after MichiGAN, LOHO [22], which applied gradient orthogonalization, and HSEGAN [23], which used hair structure-adaptive normalization, showed good hair translation performance while preserving background information. However, such GAN-based hair image translation models have the disadvantage that they need not only an image that is a source of translation but also additional inputs (e.g., reference images) to guide hair translation. For the convenience of patients, we need a translation model that can generate a predictive image for hair transplant surgery without additional inputs. Thus, we proposed the GAN-based ROI image translation method, which converts only the ROI after the hair transplant surgery while retaining the image for the non-ROI without additional materials.

3. Proposed Method

3.1. Overall Operation

To generate a predictive image after the hair transplant surgery only on the hair transplant region, i.e., the ROI, the proposed GAN-based ROI image translation model consists of the following three phases as shown in Figure 2:

1. **Image Translation:** To generate a naive prediction for the input image before hair transplant surgery. That is, the original preoperative image is converted into the postoperative image using a GAN model, e.g., CycleGAN, where the converted image is called a naive prediction image.
2. **Image Segmentation:** To estimate the surgical area, i.e., the ROI, in the input image before hair transplant surgery using a segmentation model. That is, the ROI from the original preoperative image using an image segmentation model, e.g., U-net, is extracted after image segmentation into ROI and non-ROI.
3. **Image Synthesis:** To generate the predictive image by combining the extracted ROI from the image segmentation phase and the naive prediction. After combining the extracted ROI in the image segmentation phase with the naive prediction image, we generate the translated ROI image. Finally, we generate the predictive image, where only the ROI is translated, after hair transplant surgery through overlapping translated ROI on the preoperative image.

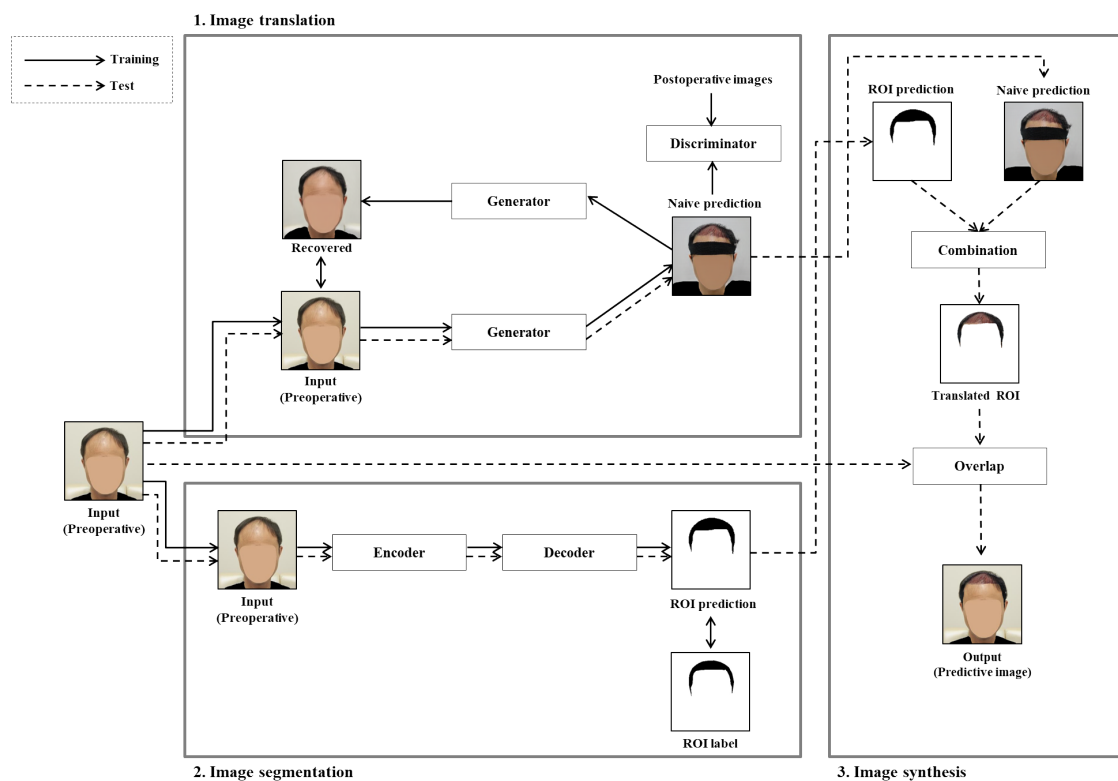


Figure 2. Overall workflow of the proposed GAN-based ROI image translation method consisting of (1) image translation, (2) image segmentation, and (3) image synthesis.

3.2. Image Translation

In the image translation phase, the proposed method generates a naive prediction of an image after hair transplant surgery using a GAN model. Hereafter, for simplicity of explanation, we call the image before hair transplant surgery the preoperative image and the image after hair transplant surgery the postoperative image. The preoperative and postoperative images are not pairwise, since it is impossible to take before and after photos at the same angle, same pose, same status. Therefore, we used unpaired data that consisted of a preoperative image set and a postoperative image set. To conduct image translation through unpaired data, CycleGAN was used as a representative GAN, which is an original model that uses cycle-consistency loss through a structure of generator–discriminator on which various image translation models are based [10,17,18,24]. The CycleGAN model consists of two pairs of generator–discriminators. Each generator converts an input image into an output image with a specific subject. While the generator learns the distribution of the subject, which is the target of the translation, the discriminator decides whether the image converted by the corresponding generator is real or fake. By doing so, the discriminator induces the generator to convert the image precisely. In Figure 3a, X represents the preoperative subject and Y represents the postoperative subject. The adversarial loss in Equation (1) for the generator G and the discriminator D_Y , while converting the image corresponding to subject X into the image corresponding to subject Y is expressed as follows:

$$L_{GAN}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{data}(y)} [\log D_Y(y)] + \mathbb{E}_{x \sim p_{data}(x)} [\log (1 - D_Y(G(x)))] \quad (1)$$

where x represents the actual preoperative image corresponding to subject X , and y represents the actual postoperative image corresponding to the subject Y . $x \sim p_{data}(x)$ represents x following the probability distribution of $data(x)$ and, $y \sim p_{data}(y)$ represents y following the probability distribution of $data(y)$. $G(x)$ represents a converted image to a postoperative image from x using G . $D_Y(y)$ represents the determination of D_Y , i.e., whether

y is real or fake, and $D_Y(G(x))$ represents the determination of D_Y , i.e., whether $G(x)$ is real or fake. When computing the adversarial loss, G aims to minimize Equation (1), and D_Y aims to maximize Equation (1). In the other pair of CycleGAN, generator F creates a converted image $F(y)$ corresponding to subject X from the actual postoperative image y corresponding to subject Y . Discriminator D_X determines whether $F(y)$ and x are the actual images. Thus, the adversarial loss of F and D_X is calculated using $L_{GAN}(F, D_X, Y, X)$.

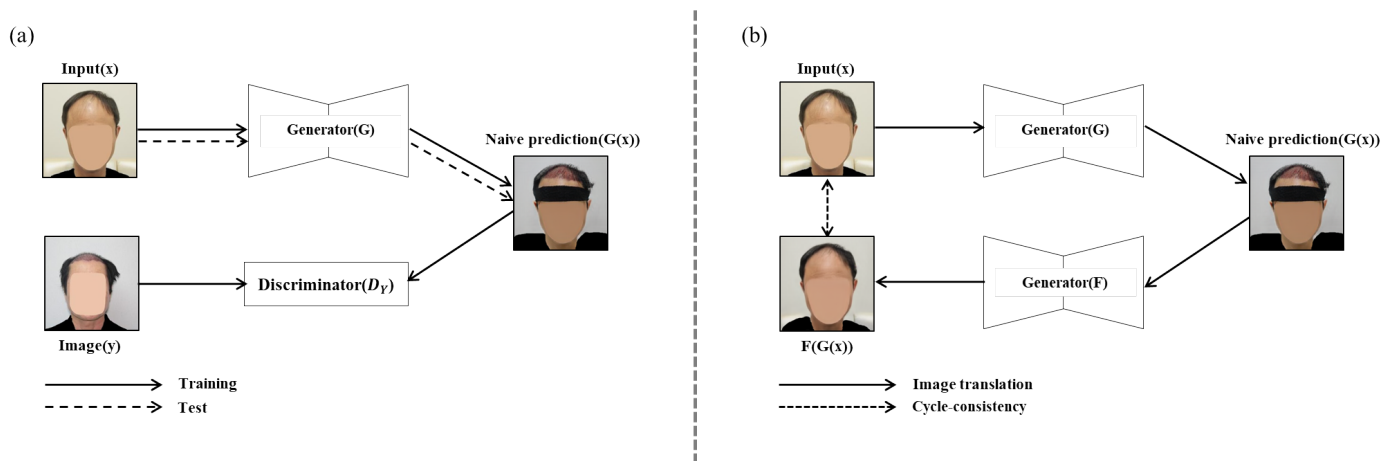


Figure 3. GAN-based image translation: (a) Workflow of G and D_Y for converting a preoperative image into a naive prediction (Workflow of F and D_X is vice versa); and (b) Cycle-consistency loss calculation for x and $F(G(x))$ (Calculation for y and $G(F(y))$ is vice versa).

Since the adversarial loss only focuses on fooling the discriminator, the adversarial loss alone cannot sufficiently train the CycleGAN network. If the adversarial loss alone is used, the generator would generate an image irrelevant to the individual characters in the input image even though the generator can deceive the discriminator [14]. As shown in Figure 3b, to solve this problem, a cycle-consistency loss proposed is as follows:

$$L_{cyc}(G, F) = \mathbb{E}_{x \sim p_{data}(x)} [\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim p_{data}(y)} [\|G(F(y)) - y\|_1], \tag{2}$$

where $F(G(x))$ represents a preoperative image recovered by F from $G(x)$ and $G(F(y))$ represents a postoperative image recovered by G from $F(y)$. In addition, $\| \cdot \|_1$ represents the L1 norm. The cycle-consistency loss aims for both generators G and F to learn in the direction of minimizing Equation (2) in the L1 norm. In other words, $F(G(x))$ and x , $G(F(y))$ and y should be as similar as possible in each pair. Through the loss function, CycleGAN can fool the discriminator and generate an output image relevant to an input image.

That is, from the generation of a naive prediction perspective, the generator G learns how to generate a naive prediction from the preoperative image. The discriminator D_Y is responsible for judging whether the naive prediction $G(x)$ and the actual postoperative image y are real or not to induce G to generate a naive prediction more precisely. In contrast, generator F restores the naive prediction $G(x)$ back to the preoperative image $F(G(x))$. The discriminator D_X is responsible for discriminating between the actual preoperative image x and the converted image $F(y)$ to improve the restoration ability of F . Thus, the final loss is expressed as follows:

$$L(G, F, D_X, D_Y) = L_{GAN}(G, D_Y, X, Y) + L_{GAN}(F, D_X, Y, X) + L_{cyc}(G, F). \tag{3}$$

Note that since the final loss is a basis of the loss functions in other image translation models [10,17,18,24], the image translation phase can be conducted by other image transla-

tion models that are able to be trained through unpaired data. For this reason, although CycleGAN was used as an example, users can select an image translation model to use freely if the model can be trained through unpaired data.

3.3. Image Segmentation

Naive prediction has a problem in that it converts not only the ROI but also regions other than the ROI. To generate more sophisticated final predictive images, the image segmentation model was used to predict the hair transplant surgery area as the ROI. Specifically, we used representative segmentation models such as U-net [25], Linknet [26], and FPN [27], which are frequently used for image segmentation tasks [28–30], and an ensemble model that combined single segmentation models. Each segmentation model has the following characteristics:

- **U-net [25]:** U-net aims to achieve good performance with fewer data by using a contracting path–expansive path structure. The contracting path captures the context information of the input image, and the expansive path performs segmentation on the input image. The contracting path and the expansive path are connected through skip connection. The expansive path performs segmentation based on the context information of the image captured by the contracting path.
- **Linknet [26]:** Linknet is a model that improves segmentation speed by removing unnecessary parameters. Similar to U-net, Linknet is divided into two modules. A module corresponding to the expansive path takes information through connection from another module of the same level. Different from U-net, Linknet uses ResNet18, which showed good performance in the image classification field when extracting features of an image. Furthermore, different from U-net, which uses concatenation for skip connection, Linknet uses the summation method to connect two modules [31].
- **FPN [27]:** FPN enables general image feature extraction by applying a pyramid structure. The structure consists of bottom-up and top-down processes based on a pyramid structure. In the bottom-up process, the FPN extracts image feature information, and the model performs localization for each layer of the input image through the top-down method. In this process, the FPN performs lateral connections to ensure accurate image feature information and combines the bottom-up pathway with the top-down pathway in the same-size space through lateral connections.
- **Ensemble model:** The ensemble method is a well-known strategy to improve performance by combining several models with each other. To apply the ensemble method to image segmentation phase, we combined the prediction results of U-net, Linknet, and FPN. The ensemble model determines whether or not a specific region is ROI based on the combined models' judgments about each pixel.

For the convenience of explanation about image segmentation, we explain the single segmentation models before explaining the ensemble model. The single segmentation models consist of an encoder–decoder, one module extracts information from the input image, and another module conducts segmentation based on the extracted information. To maximize the performance of the segmentation models, we considered ResNet34, which was pre-trained using the ImageNet dataset, as the backbone for the encoder. Such segmentation models are trained using a pre-annotated ROI, which is the groundtruth about a preoperative image given as an input. For this process, the image segmentation models were calculated through the loss function combining binary cross entropy Equation (4) and jaccard loss Equation (5) as follows:

$$L_{bce}(y, \hat{y}) = -\frac{1}{N} \sum_{i=1}^N (y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)). \quad (4)$$

$$L_{jac}(y, \hat{y}) = 1 - \frac{y \cap \hat{y}}{y \cup \hat{y}}. \quad (5)$$

In Equation (4), N is the image size, i is the arbitrary pixel index in the image, y_i represents the groundtruth value for the i_{th} arbitrary pixel, and \hat{y}_i represents a predictive value for the i_{th} arbitrary pixel. In addition, y represents a set of groundtruth values, \hat{y} represents a set of predictive values, \cap represents intersection operation, and \cup represents the sum-set operation in Equation (5). The sum of binary cross-entropy and jaccard loss is given into the final loss as follows:

$$L(y, \hat{y}) = L_{bce}(y, \hat{y}) + L_{jac}(y, \hat{y}). \quad (6)$$

Through this loss function, the single image segmentation models are trained to generate ROI prediction, that is, information about whether each pixel of the preoperative image is included in the surgical area or not. After the single segmentation models generate ROI predictions, the ensemble model combines the ROI predictions from U-net, Linknet, and FPN setting the thresholds of the ensemble model as 1, 2, 3. Hereafter, the threshold means the number of models required for a specific pixel to be predicted as ROI by the ensemble model. For example, when the threshold is 1, if any of the single segmentation models used predicts a pixel as ROI, the ensemble model predicts the pixel as ROI.

3.4. Image Synthesis

To generate a high-quality predictive image, the proposed method synthesized the naive prediction on the preoperative image at the image synthesis phase. Specifically, we changed specific pixels in the predicted region as the ROI of the preoperative image into a pixel of naive prediction at the same location using the ROI prediction.

As described in Algorithm 1, the image synthesis procedure uses a preoperative image, naive prediction, and ROI prediction as input values and accesses all pixels of the preoperative image (Lines 1–3). Subsequently, only the preoperative image pixels on the ROI are changed into the pixel values on the naive prediction image at the same position (Lines 4–5). Finally, a predictive image is generated where the preoperative image is changed into naive prediction only on the ROI (Lines 9–10).

Algorithm 1 Image synthesis procedure.

```

1: procedure IMAGE SYNTHESIS(Preoperative_image, Naive_prediction, ROI_prediction)
2:   for  $i$  in Preoperative_image.height do
3:     for  $j$  in Preoperative_image.width do
4:       if ROI_prediction[ $i$ ][ $j$ ]  $\in$  ROI then
5:         Preoperative_image[ $i$ ][ $j$ ]
            $\leftarrow$  Naive_prediction[ $i$ ][ $j$ ]
6:       end if
7:     end for
8:   end for
9:   Predictive_image  $\leftarrow$  Preoperative_image
10:  return Predictive_image
11: end procedure

```

4. Experiment

In this section, to show the proposed method can generate the predictive images while preserving the characters of individuals, we compared the performances experimentally and analyzed the results. We conducted the experiment to answer the following four questions:

- **RQ1:** How does a hyperparameter of cycle-consistency loss (λ) influence the quality of predictive images generated by the proposed method?

- **RQ2:** Does the framework show a better performance than other image translation models?
- **RQ3:** Can applying an ensemble method to segmentation models improve the performance of the proposed method?
- **RQ4:** What is the difference between segmentation and attention in predicting hair transplant surgery results?

4.1. Experimental Environment

In the image translation phase, we trained CycleGAN using 1394 preoperative images and 896 postoperative images of $256 \times 256 \times 3$ size. As the parameter values of CycleGAN, we used a learning rate of 0.0002, an Adam optimizer, a batch size of 1, and 200 epochs. When training the image segmentation model, 1394 pairs of $256 \times 256 \times 3$ preoperative images and corresponding annotation information about the surgical area were used. For the test, we used 374 preoperative images as the test data and 168 postoperative images for a standard of the FID in comparison. We set the parameter values of the segmentation model as follows: an Adam optimizer, a batch size 1, and 50 epochs. We used a Linux operating system and machine equipped with an AMD EPYC 7301 16-Core Processor at CPU, a GeForce RTX 2080 Ti GPU, and 128 GB of RAM.

4.2. Evaluation Metrics

To measure the quality of the predictive images generated by different models, we used three metrics.

- **SSIM (Structural Similarity Index Measure)** [11] is a measure to determine the structural differences of an image in consideration of the luminance and contrast. To apply this metric to the present experiment, we calculated the SSIM between the preoperative and the predictive images. SSIM indicates how well the predictive image preserves the input image information. The higher the SSIM score is, the better the performance.
- **IoU (Intersection over Union)** [12] is a measure of how accurately models detect the ROI. We calculated the IoU using the actual ROI of the preoperative image and the translated area of the predictive image. Specifically, the IoU metric is calculated as follows:

$$IoU(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}, \quad (7)$$

where A represents the actual ROI of the preoperative image, B represents the translated area of the predictive image, $A \cap B$ represents the intersection area between A and B , and $A \cup B$ represents the union area between A and B . An increase in the IoU score indicates that the predictive model accurately recognizes the ROI.

- **FID (Frechet Inception Distance)** [13] determines the similarity between the actual images and the predictive images. To calculate the FID metric, the Inception-V3 network was used to convert images into feature vectors, and then, we calculated the Wasserstein-2 distance on these feature vectors. The shorter the distance between the two images, the better the similarity between the actual postoperative images and the predictive images.

Since the dataset we used was very sensitive to privacy issues, this experiment was not possible for perceptual study based on human evaluation. If the dataset would be shown without blurring, individuals' identities could be easily recognized. For this reason, we carefully selected metrics to measure performances. There are metrics that can measure the performance of the image translation models well (e.g., Wasserstein metric [8]); however, we used SSIM, IoU, and FID which are commonly used to evaluate performances of image translation models. Through the metrics that we used, the proposed method could be not only compared with other and future image translation models but also

evaluated considering the human visual system that is taken into account in image quality assessment [32,33]. As for the metrics, SSIM evaluates how well the model preserves the structure of input while translating, IoU evaluates whether it accurately translates only the ROI, and FID evaluates how similar the distribution of the generated images and the distribution of the actual image are through feature vectors of Inception-V3 [34]. By evaluating performance using these three metrics, we enabled systematic analysis of the models in quantitative evaluation.

4.3. How Does a Hyperparameter of Cycle-Consistency Loss (λ) Influence the Quality of Predictive Images Generated by the Proposed Method? (RQ1)

To examine the influence and find the optimal weight of the cycle-consistency loss (λ), we set λ as 0, 3, 5, 10, 15, and 20, and CycleGAN was trained 10 times for each λ . Using the trained CycleGAN under such conditions in the image translation phase, we generated predictive images for the test data and compared their means of performance under each condition.

4.3.1. Quantitative Evaluation

In Table 1, we summarize the results of our experiment for RQ1. An increase in λ improved the performances in the SSIM metric. For example, when λ was 20, we observed the highest SSIM scores of 0.957, 0.956, 0.954, respectively, while the worst SSIM score appeared when λ was 0. Comparing the two cases, the average of the SSIM scores when λ was 20 was 6% better than the score when λ was 0. To show there were meaningful improvements according to an increment of λ from a statistical perspective, we conducted a *t*-test on the SSIM scores. The *t*-test for 15 and 20 in λ , which had the smallest difference in the SSIM between one another, had a *p*-value of 0.02 in U-net, Linknet, and FPN, meaning significant improvements. This result implies that the increase in the cycle-consistency loss weight better preserves the information of the input images.

Table 1. Comparison of averages and standard deviations of performances according to the cycle-consistency loss hyperparameter (λ).

Segmentation Models	λ	SSIM	IoU	FID
U-net [25]	0	0.899 \pm 0.0090	0.74025 \pm 0.000010	58.407 \pm 2.88
	3	0.938 \pm 0.0030	0.74016 \pm 0.000019	54.342 \pm 0.46
	5	0.946 \pm 0.0027	0.74014 \pm 0.000012	53.960 \pm 0.23
	10	0.950 \pm 0.0026	0.74016 \pm 0.000020	53.680 \pm 0.29
	15	0.954 \pm 0.0031	0.74016 \pm 0.000012	53.569 \pm 0.23
	20	0.957 \pm 0.0017	0.74017 \pm 0.000015	53.632 \pm 0.28
Linknet [26]	0	0.897 \pm 0.0090	0.74591 \pm 0.000010	58.999 \pm 2.87
	3	0.937 \pm 0.0030	0.74582 \pm 0.000019	54.814 \pm 0.43
	5	0.945 \pm 0.0027	0.74580 \pm 0.000013	54.442 \pm 0.33
	10	0.949 \pm 0.0027	0.74583 \pm 0.000020	54.152 \pm 0.22
	15	0.953 \pm 0.0032	0.74583 \pm 0.000012	53.863 \pm 0.35
	20	0.956 \pm 0.0017	0.74584 \pm 0.000015	53.853 \pm 0.39
FPN [27]	0	0.893 \pm 0.0094	0.76754 \pm 0.000010	59.508 \pm 3.18
	3	0.934 \pm 0.0031	0.76745 \pm 0.000020	54.749 \pm 0.47
	5	0.943 \pm 0.0028	0.76743 \pm 0.000014	54.358 \pm 0.26
	10	0.947 \pm 0.0028	0.76746 \pm 0.000021	53.885 \pm 0.28
	15	0.951 \pm 0.0032	0.76746 \pm 0.000012	53.594 \pm 0.28
	20	0.954 \pm 0.0018	0.76747 \pm 0.000016	53.547 \pm 0.28

On the other hand, the result of the IoU metric showed a similar performance under the same segmentation condition regardless of the changes in λ . The results of the IoU score show that the capacity to detect the ROI is up to the segmentation model applied.

In the FID metric, an increment of λ generally improved the scores. For instance, when λ was 0, predictive images showed the worst performances of 58.407, 58.999, 59.508, respectively, while when λ was 20, the FID scores of predictive images exhibited improved performance of 53.632, 53.853, 53.547. Comparing the two cases, the FID score at λ 20 was 9% better than the score under the condition of λ 0. These results imply that the adoption and increase in λ showed a positive effect on generating more realistic and diverse predictive images. However, we note that the comparison did not show significant improvements in performance after 10 in λ . In the *t*-test, the *p*-values between 15 and 20 in λ were 0.61, 0.95, and 0.72, respectively, showing no significant difference. Furthermore, the *p*-values between 10 and 15 in λ were 0.38, 0.05, and 0.04, showing a meaningful difference on FPN only. Unlike the previous results, in 5 and 10 in λ , the *p*-values were 0.04, 0.04, and 0.001, respectively, indicating statistically significant results in all models. This means the increase in λ improves performance in the FID until 10 and, the improvement is limited after 10 λ .

4.3.2. Qualitative Evaluation

In Figure 4, we show predictive images according to the change in λ . First, when λ was 0, the proposed method generated the artificial outputs without considering the features of the input images. For example, as shown in the fourth row of Figure 4, the proposed method generated a predictive image regardless of the hair shape of the input image under the condition of λ 0. Unlike the outputs when λ was 0, outputs under the condition of cycle-consistency showed the conversion considered the features of the input images. Comparing between λ 0 and λ 3, we observed that the introduction of λ enabled the proposed method to reflect the characteristics of the input images. Specifically, in the predictive images of λ 3, information from the input images (e.g., hair shape and location of forehead) was preserved unlike predictive images with λ 0. Furthermore, the expansion of the influence of the cycle-consistency led to the generation of more natural predictive images. For example, in the second row of Figure 4, between λ 3 and λ 5, there was a difference in the ROI of the predictive images. Specifically, the predictive image created with λ 5 was more natural than the predictive image created with λ 3. From a generalization point of view, the improvement according to this increment of λ in image quality can be clearly recognized in the third and fourth rows of Figure 4. We can see the larger λ became, the better the quality of the predictive images. In particular, the upper part of the head of the subject was more properly preserved, and the hairline was very natural in the predictive images of λ 20. However, similar to the quantitative evaluation, significant changes in image quality were not seen after λ 10. Considering such results, although we did not find the upper bound of λ , we concluded that 20 is the optimal hyperparameter of the cycle consistency loss on the experiment in RQ1.

Result 1: The cycle-consistency loss contributed to better predictive image generation, which showed the optimal result when the weight value was 20.

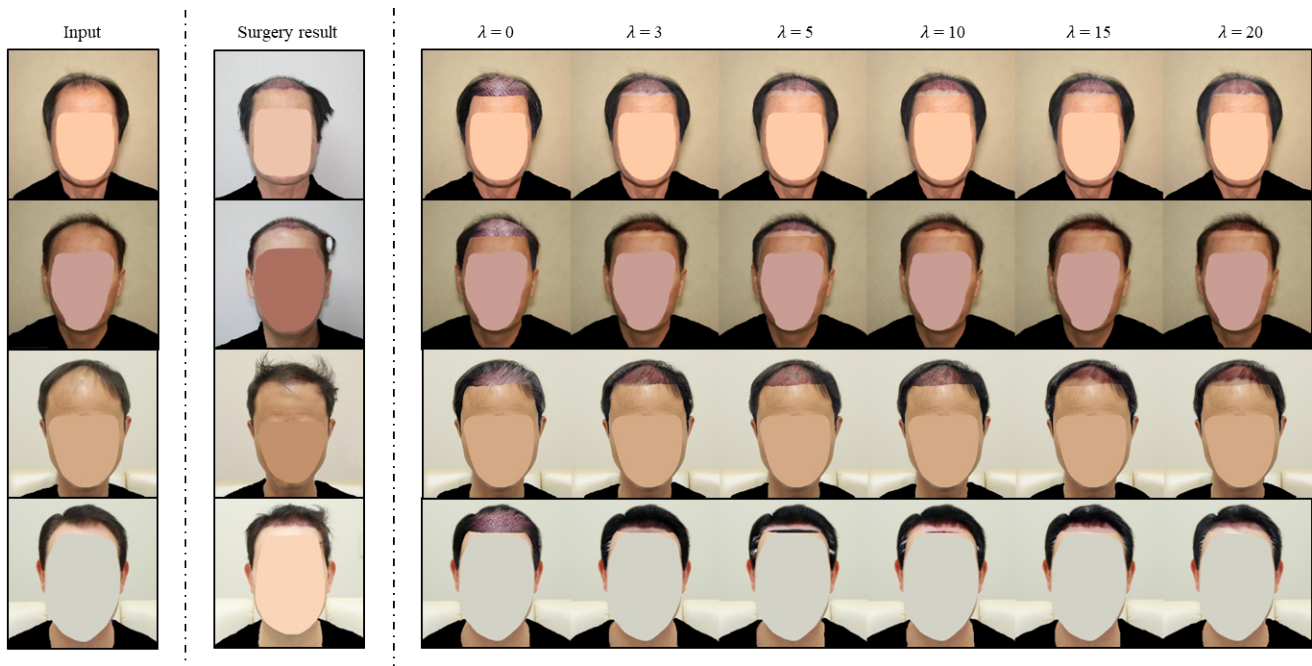


Figure 4. Predictive images generated by the proposed method trained based on different λ (the clothing was blocked for privacy issues).

4.4. Does the Proposed Method Show a Better Performance Than Other Image Translation Models? (RQ2)

To confirm whether the proposed method can achieve better performance than other image translation models, we compared the proposed method with other image translation models for which we could obtain source codes. In the experiment, we selected CycleGAN [14], UNIT [18], AGGAN [10], and CUT [34] as competitive models since the models have good characteristics for comparison. Specifically, CycleGAN is the one that first proposed cycle-consistency loss, which is the core of image translation using unpaired data. In addition, UNIT is a representative model based on shared-space for image translation applied by various models [19,35]. Furthermore, AGGAN is a model applying attention module and showed better performance than the competitive model [17] to which an attention module was applied [24]. Finally, CUT is a state-of-the-art model, which applied a contrastive learning method to image translation through a patch-based approach.

In order to know how good the quality for each region is in predictive images, we not only conducted performance measurement on whole image but also separated ROI and non-ROI in an image. In addition, we used one of the models with λ 20 as in RQ1, and we use this model in future experiments.

In the experiment for the whole image, we used SSIM, IoU, and FID. SSIM used 374 preoperative images used as test data, and IoU leveraged 374 ROI annotations as standard. FID was calculated through 168 postoperative images not included in the training data. The reason for using postoperative images not included in the training data was to prevent factors unrelated to the hair transplantation area from affecting the score.

4.4.1. Quantitative Evaluation on Whole Image

In Table 2, we summarize the evaluation results from experiment RQ2. The proposed method showed a better performance than other image translation models in the SSIM metric. Specifically, CycleGAN, UNIT, AGGAN, and CUT obtained SSIM scores of 0.768, 0.646, 0.908, and 0.764, respectively. The proposed method obtained SSIM scores of 0.958, 0.957, and 0.955, respectively. Comparing the other image translation models and the proposed method, the average score of the proposed method was 23% higher than the

average score of other image translation models in the SSIM. This result implies that the proposed method better preserves the characteristics of the individual in images.

Table 2. Performance comparison of the proposed method and other image translation models on whole image.

Models	SSIM	IoU	FID
CycleGAN [14]	0.768	0.13589	67.191
UNIT [18]	0.646	0.13590	116.798
AGGAN [10]	0.908	0.13594	56.968
CUT [34]	0.764	0.13589	63.153
Proposed method (U-net [25])	0.958	0.74015	53.498
Proposed method (Linknet [26])	0.957	0.74582	53.666
Proposed method (FPN [27])	0.955	0.76745	53.223

In the IoU metric, we observed that the proposed method resolved the limitation of other image translation models. For example, CycleGAN, UNIT, AGGAN, and CUT showed almost identical IoU scores 0.13589, 0.13590, 0.13594, and 0.13589, respectively. Note that when the values of all pixels in the test image were changed, the IoU score was 0.13590. The score of the work was similar to the scores of other image translation models in terms of the IoU. For this reason, we can conclude that other image translation models changed almost all pixels during this experiment. In contrast, the proposed method showed 0.74015, 0.74582, and 0.76745, respectively. The average score of the proposed method was 452% higher than the average score of other image translation models in the IoU. In addition, the proposed method showed a good performance with all segmentation models applied. Therefore, this result indicates that the proposed method solves the problem of other image translation models, and it is compatible with various image segmentation models.

The proposed method showed superiority over other image translation models in the FID metric. Specifically, the proposed method obtained FID scores of 53.498, 53.666, and 53.223, respectively, while other image translation models obtained FID scores of 67.191, 116.798, 56.968, and 63.153, respectively. The average score of the proposed method was 42% better than the average score of other image translation models in terms of the FID. In addition, note that UNIT showed the worst FID score, since most predictive images from UNIT shared similar features through the shared-space used for the image translation. Specifically, the tendency of UNIT made all predictive images appear similar to each other, which caused the distribution of predictive images to move far from the distribution of the actual postoperative images. From this point of view, the proposed method showed a good score by preserving individual characters. Therefore, the results indicate that the distribution of the actual postoperative images and the distribution of the predictive images generated by the proposed method were more similar than the distribution of translated images generated by other image translation models. Considering the results for all metrics comprehensively, we conclude that the proposed method cannot only detect the ROI well but also convert the area similar to the actual postoperative images.

4.4.2. Qualitative Evaluation on Whole Images

In Figure 5, we visualized the ROI detection of predictive images by the different models used in RQ2. The groundtruth of the ROI was represented by black and the true positive, true negative, false positive, and false negative were represented by yellow, white, green, red, respectively. Other image translation models changed all parts of the images. As shown in the third to the sixth columns of Figure 5, green appeared in the non-ROIs of all images, since all parts except the ROI became false positives in all images. This result means that the existing image translation models converted whole images. Unlike CycleGAN and UNIT, CUT showed true negative regions in the second row and AGGAN showed some regions of true negatives in the third and fourth rows. However, the regions

of true negatives were very small. In contrast, the proposed method generated predictive images by appropriately selecting the ROI through an image segmentation process. For instance, the predictive images of the proposed method had few false positives in the third row in Figure 5. In addition, the proposed method showed stable performance in all the segmentation models used in the experiment. These results indicate that the proposed method detects the ROI well, and diverse image segmentation models are compatible with the proposed method.

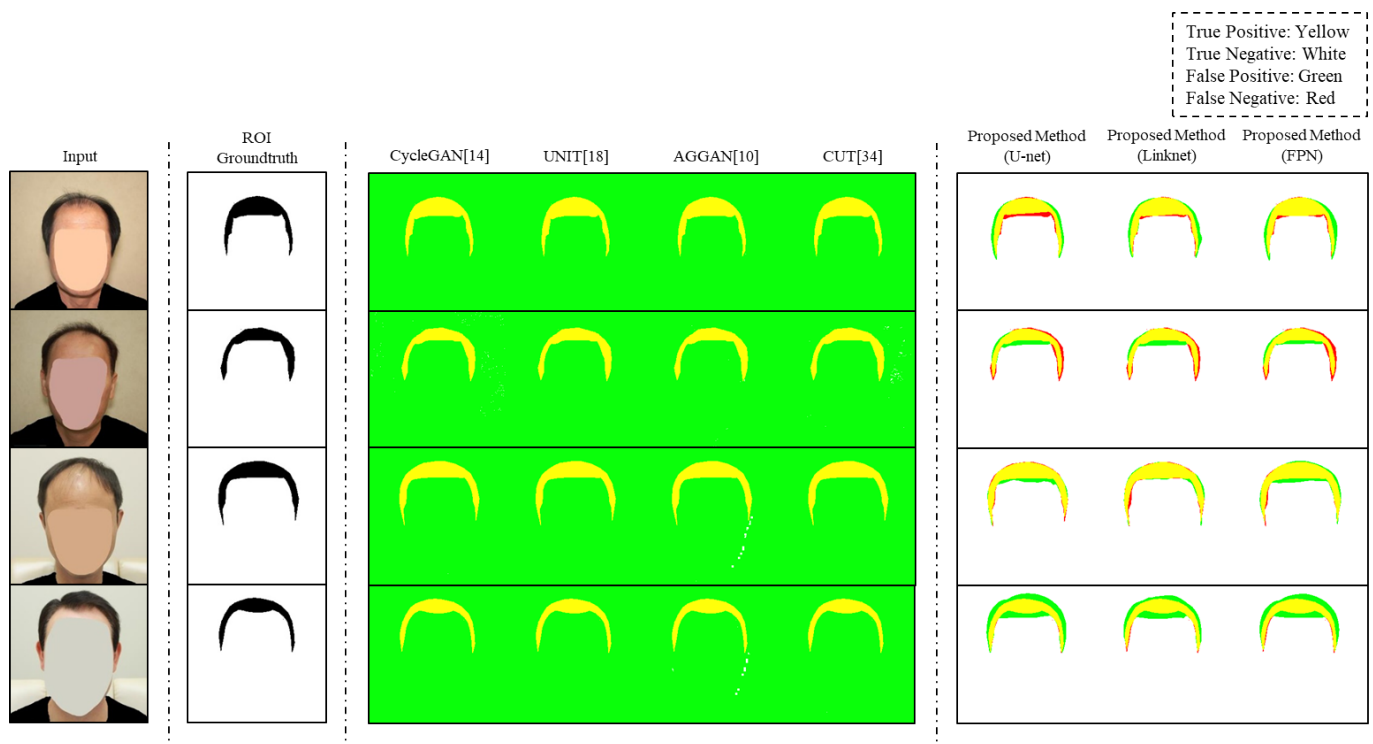


Figure 5. Visualization of the ROI detection of predictive images generated by the proposed method and other image translation models, i.e., CycleGAN [14], UNIT [18], AGGAN [10], and CUT [34].

In Figure 6, we show the comparative image translation results between other image translation models and the proposed method. CycleGAN generated the predictive images differently from input images in regions that were not included in the ROI. From the comparison between the first column and the third column in Figure 6, we could observe the hairs of subjects in the predictive images changed faintly, and unrelated factors such as a headband appeared. Similar to CycleGAN, all predictive images of UNIT showed irrelevant features (e.g., blurred color, shape of cloth, and headband). Note that the quality of the predictive images of AGGAN was also poor, even if the model applied the attention module to resolve the limitation of other image translation models. Specifically, AGGAN produced a predictive image in which the range of the hair transplant surgery was estimated incorrectly, as shown in the third row of the fifth column in Figure 6. The result indicates that applying the attention module cannot solve the problem perfectly. In addition, CUT showed the same limitation, in that irrelevant factors were generated in predictive images in the third and fourth rows of Figure 6. This means contrastive learning is not a useful method to train to preserve non-ROI. Limitations of the existing image translation models were caused by the simple learning method. Contrary to other image translation models, the proposed method generated predictive images that were modified only for the ROI. For example, the predictive images of the proposed method did not show any different elements from those of the input images, except for the predicted region of the hair transplant surgery area in Figure 6. This result implies that the proposed method overcomes the existing problem of CUT of other image translation models by using a

structure that can specify the ROI. In other words, by using other image translation and image segmentation independently, the proposed method can learn the distribution of the actual hair transplant surgery result and preserve the information of the input image simultaneously.

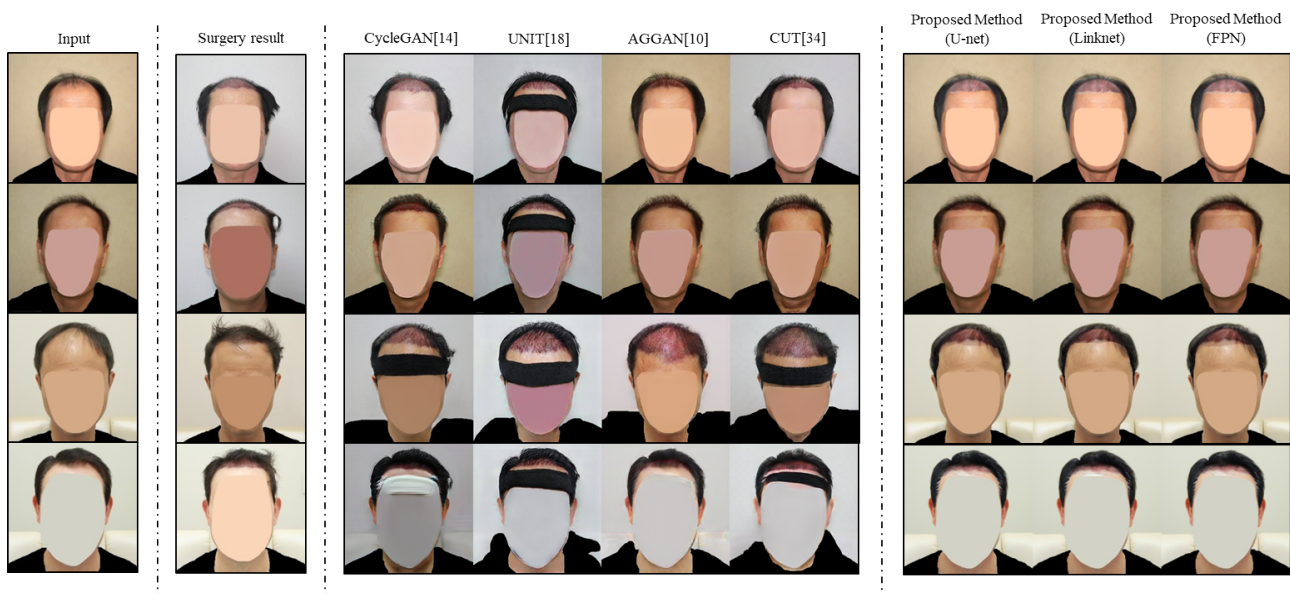


Figure 6. Final predictive images converted by the proposed method and other GAN-based image translation models.

4.4.3. Quantitative Evaluation on ROI and Non-ROI

Although the comparison was conducted on the whole image in Section 4.4.1, we performed additional comparisons for a more sophisticated analysis. Specifically, we separated the whole image into ROI and non-ROI and performed experiments, respectively. In the experiment for ROI, we measured how well the proposed method performed in the field that the existing image translation models use. Conversely, through an experiment for non-ROI, we measured how well the proposed method was in the field that the other image translation models cannot use. In the experiment for ROI, we used 374 ROI images of preoperative images as test data to calculate the SSIM scores. To calculate the FID, 896 ROI images of postoperative images from the training data were used, since the ROI images did not involve irrelevant factors from the hair transplantation region. In the experiment for non-ROI, 374 non-ROI preoperative images were used as a standard for the SSIM and FID, since the non-ROI images of preoperative images used as test data are groundtruth that translation models should preserve from the perspective of structure and distribution. In addition, since data were separated as ROI and non-ROI already, the IoU was not meaningful; hence, it was excluded from the measurements.

We summarize the evaluation results on ROI only in Table 3. AGGAN showed the best score in the SSIM metric of 0.978, and the proposed methods showed the second-best scores with an average of 0.965. CUT, CycleGAN, and UNIT were 0.952, 0.951, and 0.925, respectively, in the SSIM metric. In the FID metric, AGGAN showed the best performance with 86.167 and CycleGAN and CUT showed better performances with 91.236 and 91.595, respectively. The proposed method scored 93.313, 93.284, and 93.027, and UNIT was 108.581. Note that since the ROI images were much simpler than whole images, the FID scores in the ROI images showed a tendency to have worse scores overall compared to the FID scores of whole images.

In addition, we organized the evaluation results on non-ROI in Table 4. The proposed method showed the best performance compared to other image translation models in the SSIM metric. The proposed method had SSIM scores of 0.998, 0.998, and 0.997. CycleGAN, UNIT, AGGAN, and CUT were 0.831, 0.744, 0.936, and 0.825 in the SSIM metric, respectively.

The average score of the proposed method was 19% higher than the average score of the other image translation models in the SSIM. As in the SSIM, the proposed method showed the best scores of 1.011, 1.008, and 1.260 in the FID. CycleGAN, UNIT, AGGAN, and CUT had FID scores of 57.830, 94.178, 31.557, and 51.066, respectively. In the FID, the proposed model was 526% better than the average score of other image translation models.

Table 3. Performance comparison of the proposed method and other image translation models on ROI.

Models	SSIM	FID
CycleGAN [14]	0.951	91.237
UNIT [18]	0.925	108.580
AGGAN [10]	0.978	86.176
CUT [34]	0.952	91.595
Proposed method (U-net [25])	0.966	93.306
Proposed method (Linknet [26])	0.965	93.280
Proposed method (FPN [27])	0.964	93.026

Table 4. Performance comparison of the proposed method and other image translation models on non-ROI.

Models	SSIM	FID
CycleGAN [14]	0.831	57.830
UNIT [18]	0.744	94.178
AGGAN [10]	0.936	31.557
CUT [34]	0.825	51.066
Proposed method (U-net [25])	0.998	1.011
Proposed method (Linknet [26])	0.998	1.008
Proposed method (FPN [27])	0.997	1.260

The result on ROI implies AGGAN has the ability to translate ROI properly while preserving the structure of ROI. In addition, the proposed method preserved the structure of ROI well but did not translate ROI perfectly since if the ROI prediction was smaller than the actual ROI, the proposed method was not able to translate the region that was not predicted. Conversely, since CUT and CycleGAN translate all regions, the image structure preservation ability was worse than the proposed model. However, CUT and CycleGAN can translate preoperative ROI into predictive ROI regardless of ROI detection. Compared to the other models, UNIT was the worst in SSIM and FID. Since UNIT forces all preoperative images to be translated through one shared-space, the translated results (predictive images) became very monotonous and collapsed the structure of each image. In the result on non-ROI, the result showed the proposed method had outstanding ability to preserve non-ROI. The proposed method earned the best scores in the SSIM and FID. Specifically, we know that the method was strong in preserving distribution on non-ROI compared to the existing translation models. Although AGGAN preserved non-ROI information relatively well among the existing image translation models, a significant difference existed compared to the proposed method. CUT, CycleGAN, and UNIT converted the non-ROI more than AGGAN did, and in particular, UNIT showed the worst performance in terms of image distribution. Considering the two results from the experiments on ROI and non-ROI, we can conclude the proposed method can generate predictive images on ROI with good structure, preserving non-ROI information.

4.4.4. Qualitative Evaluation on ROI and Non-ROI

To analyze the predictive images more carefully, we show predictive images on ROI and non-ROI translated by the existing image translation models and the proposed method

in Figure 7. CycleGAN translated for the headless region well except the point where the left hair of the patient was changed into being faintly in ROI. However, in non-ROI, we can see that the right forehead was translated into the postoperative region even though the region is not the expected surgical area. UNIT showed inappropriate translation such as hair growth about the central region where there was no hair in the ROI of the preoperative image. Furthermore, in non-ROI, about half of the forehead was translated into the postoperative region. AGGAN not only preserved well the region where hair exists but also translated the expected surgical region to red in ROI. However, in non-ROI, a large region of the forehead was translated into the postoperative region, and the boundary of the hair transplant surgery was not clearly revealed. CUT preserved the existing hair well; however, the red surgical marks were not clearly visible in the surgical region. Furthermore, CUT translated more than half of the forehead into a surgical region in non-ROI and, it made the patient appear to be bowing their head. In contrast, the proposed method translated ROI properly and preserved non-ROI well in all image segmentation models. Specifically, the results of the proposed method not only preserved a region where hair exists but also clearly revealed the boundary of hair transplant surgery in ROI. Furthermore, in non-ROI, there were some differences (e.g., the boundary of the synthesis under the forehead) depending on the image segmentation model used; however, all results did not show the translated forehead of a patient into the postoperative region, unlike the existing image translation models. From the perspective of patients, judging how natural the hair transplant results will be through the predictive image is important. For this reason, preserving non-ROI is as important as translating ROI. It means that the proposed method, which predicts the surgical area more accurately through the image segmentation phase, can provide more proper results to patients.

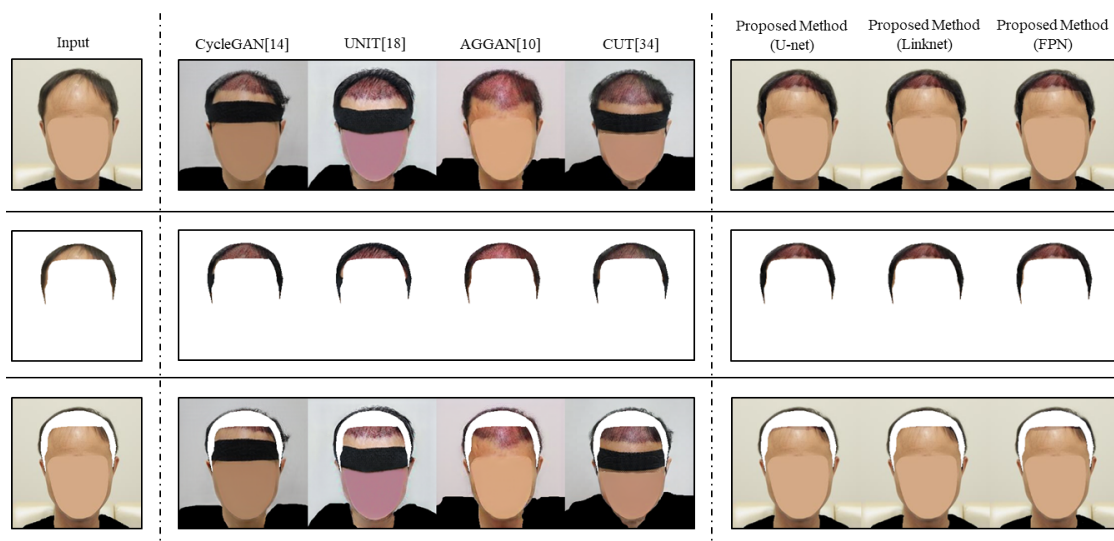


Figure 7. Comparison predictive images on ROI and non-ROI converted by the proposed method and other GAN-based image translation models.

Result 2: The proposed method generated a high-quality predictive image after hair transplant surgery while translating only the hair transplant surgery region.

4.5. Can Applying an Ensemble Method to Segmentation Models Improve the Performance of the Proposed Method? (RQ3)

Since preoperative images can be taken from various angles, segmentation models should be robust enough to detect ROI from different images from diverse angles. However, the single segmentation model did not show consistent performances from the different images. To improve the robustness of ROI prediction, we can apply the ensemble method to the proposed method. The ensemble method is a well-known strategy to improve accuracy

through combining several models with each other. In RQ3, we combined ROI predictions from U-net, Linknet, and FPN and measured the performance of the model applying the ensemble method. For the experiment of RO3, we used 374 preoperative images for SSIM and 374 ROI annotations as standard for IoU. Furthermore, we leveraged 168 postoperative images for FID, which is the same condition as RQ2 for the whole images.

4.5.1. Quantitative Evaluation

We summarize the evaluation results of the segmentation models in Table 5. The ensemble methods were specified into three types according to the threshold. In the SSIM metric, the ensemble model with threshold 3 had the best performance of 0.963. In addition, the ensemble model with threshold 2 had a mediocre performance of 0.957, and the value was similar to the results from the single segmentation models. The single segmentation models had SSIM scores of 0.958, 0.957, and 0.955, respectively. The ensemble model with threshold 1 showed the worst SSIM score of 0.951.

Table 5. Performance comparison of single segmentation models and ensemble models with different thresholds.

Models	Threshold	SSIM	IoU	FID
U-net [25]	-	0.958	0.74015	53.498
Linknet [26]	-	0.957	0.74582	53.666
FPN [27]	-	0.955	0.76745	53.223
Ensemble	1	0.951	0.80171	53.160
	2	0.957	0.76548	53.680
	3	0.963	0.68461	53.823

However, note that the more a model changes region in a preoperative image, the worse the results in the SSIM. In the IoU metric, the ensemble model with threshold 1 showed the best performance of 0.80171. The reason why the ensemble model with threshold 1 showed the worst score in the SSIM was that the model changed ROI more than other models did. Other models showed consistent results with the interpretation of the relationship between SSIM and FID. The ensemble model with threshold 2 was 0.76548 and the single segmentation models were 0.74015, 0.74582, and 0.76745, respectively, in the IoU metric. The ensemble model with threshold 3 showed the worst IoU score of 0.68461.

Likewise, the ensemble model with threshold 1 had the best FID score of 53.160. Models applying single segmentation had FID scores of 53.498, 53.666, and 53.223, and the other ensemble models had FID scores of 53.680 and 53.823. These results imply the ensemble model with threshold 1 has the ability to generate predictive images more precisely through detecting ROI more accurately. Considering the results for all metrics comprehensively, we conclude that an ensemble model with threshold 1 generates more realistic predictive images based on more accurate ROI detection performance.

4.5.2. Qualitative Evaluation

In Figure 8, we visualized ROI detection generated by single segmentation models and the ensemble models with different thresholds. In Figure 8, true positive, true negative, false positive, and false negative are indicated in the same method used in Figure 5. As shown in the second to the fourth columns of Figure 8, each single segmentation model had strong points in different images. For instance, U-net showed a good prediction result in the second row, however, in other rows, the prediction results were bad or mediocre. Linknet had strong points in the third and the fifth rows, but the model also showed bad prediction results in the fourth and the sixth rows. Likewise, FPN showed good results in the first and the sixth rows and bad results in the second and the fifth rows, simultaneously.

Unlike the results from the single image segmentation models, the ensemble model with threshold 1 showed good results consistently. For instance, in the first, second, and

fifth rows of Figure 8, the results from the ensemble model with threshold 1 indicated red regions much less than other competitive models. Another ensemble model with 2 as the threshold also showed good results. However, unlike the ensemble model with threshold 1, the model had limitations in that the results seemed similar to results from the second-best model in the single segmentation models from the third to the fifth rows. The results show that the ensemble model with threshold 2 was inefficient. Similarly, an ensemble model with threshold 3 revealed limitations showing similar results to those of the worst model in the single segmentation models in the third to the fifth rows.

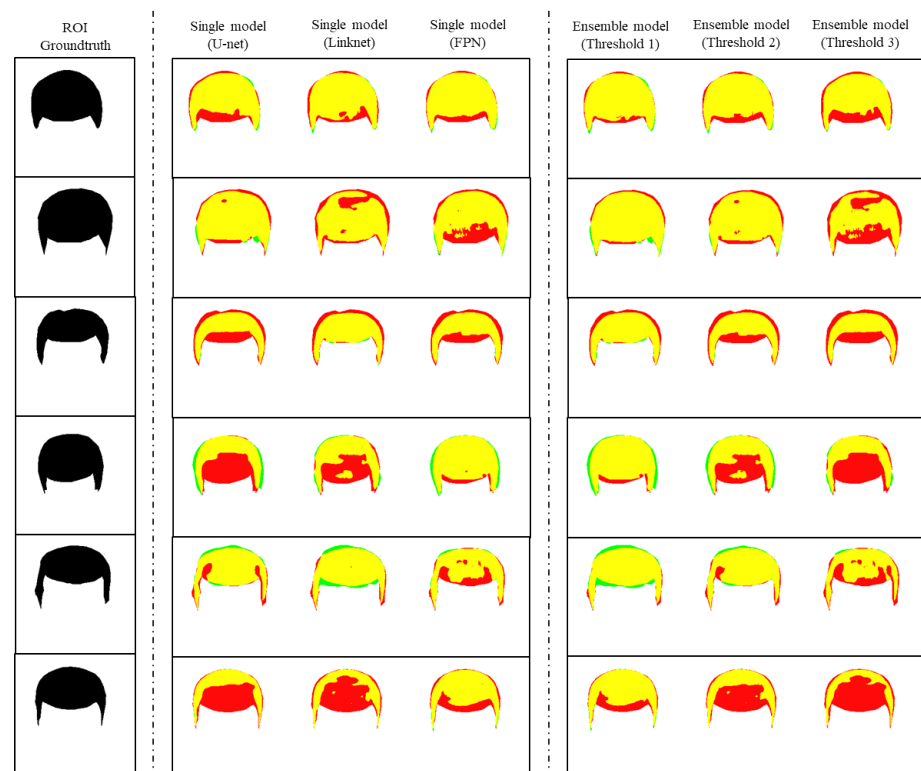


Figure 8. Visualization of the ROI detection generated by single segmentation models and ensemble models with different thresholds.

These results indicate the ensemble model with threshold 1 has the ability to detect ROI on preoperative images well.

Result 3: Applying the ensemble method with threshold 1 improved prediction of ROI on preoperative images compared to the single segmentation models.

4.6. What Is the Difference between Segmentation and Attention in Predicting Hair Transplant Surgery? (RQ4)

To translate ROI while preserving non-ROI in images, an attention module has been used conventionally, and applying the attention module achieved good results in diverse subjects. However, we used segmentation models to conduct translation on ROI, since using an attention module has limitations for predicting the result of hair transplant surgery. To show that using segmentation models was appropriate for this subject, we compared the proposed method with AGGAN [10], a well-known model applying an attention module.

Since the workflows of both models are very different from each other, comparing attention with segmentation in quantitative evaluation is difficult. To show differences between the two models through qualitative evaluation instead of quantitative evaluation, we compared the attention and predictive images of AGGAN with the segmentation and predictive images of the proposed method in Figure 9. As shown in the third column of

Figure 9, the results of attention from AGGAN showed factors (e.g., clothes, structure of humans) irrelevant to the hair transplant surgery. Unlike the attention results, the results of segmentation showed only regions involved with hair transplant surgery. Similar to the difference between attention and segmentation, the two models showed different results in the predictive images. As shown from the third to the fourth rows of Figure 9, the predictive images of AGGAN showed converted shapes of clothes. In addition, the predictive image from AGGAN in the third row showed a dramatically different prediction of hair transplantation. Conversely, the proposed method had predictive images translated on ROI only. This difference between the results from the two models occurred, since AGGAN learns the distribution over the whole image. Specifically, AGGAN would give attention to any region, even if the region is not the area where hair transplantation takes place (e.g., patient clothes). Therefore, the attention module of AGGAN cannot specify only a specific element among the hair transplant-related elements existing in the preoperative image. On the other hand, the segmentation model of the proposed method can perform more accurate ROI detection through learning only the expected surgical region for hair transplantation. In summary, through segmentation, the proposed method can specify ROI where hair transplant surgery will take place; however, AGGAN cannot because of the limitation of attention. Therefore, the results imply that applying the segmentation model was appropriate for ROI detection, since segmentation models can specify the expected hair transplant surgery region.

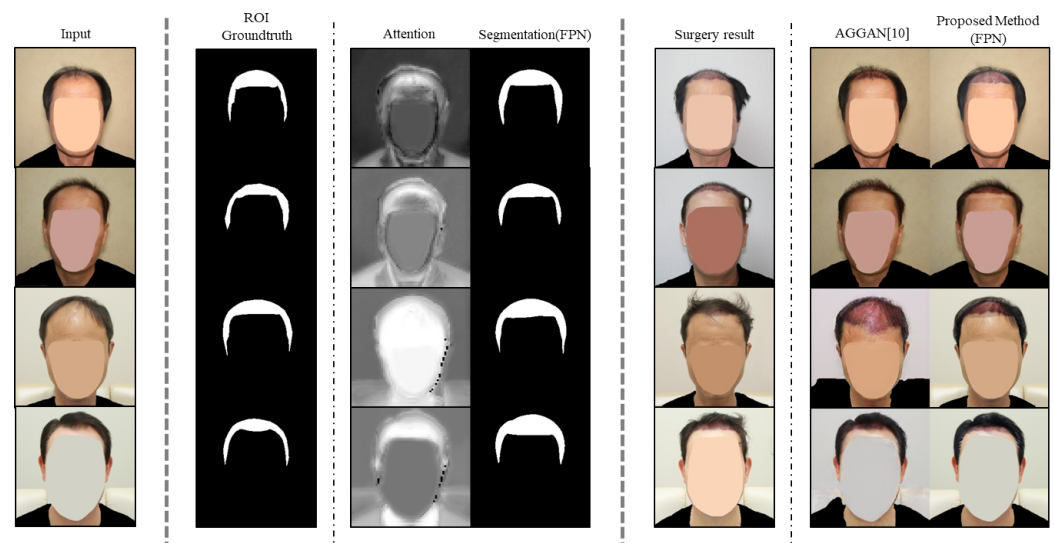


Figure 9. Comparison between AGGAN [10] and the proposed method in ROI and the actual hair transplant surgery results (ROI was represented as white to compare attention with segmentation in this figure only).

Result 4: Segmentation is better than attention for hair transplant surgery prediction, since segmentation models can specify a focus.

5. Limitations

The proposed method solves the problem of changing non-ROI in image translation. However, if the ROI prediction from the image segmentation phase is not perfect, the predictive image can be unnatural. For instance, in predictive images from our framework from the first to the second columns of Figure 6, the boundaries between non-ROI and ROI are visible. To deal with this problem, the proposed method needs a better methodology to combine preoperative images with naive predictions naturally, even if the ROI prediction is not perfect.

In addition, the proposed method is susceptible to the quality of naive prediction from an image translation model selected in the image translation phase. If a selected translation model did not translate the ROI of a preoperative image, the proposed method cannot achieve a high-quality predictive image even if ROI detection was conducted well. The proposed method needs forcing methodology to guarantee correct ROI conversion.

Comprehensively, to generate better predictive images, it is necessary to consider the distribution of non-ROI in harmony with ROI in the entire image. At the same time, we need a model which can specify the ROI to be translated and force it to be translated. In the future, we can design one overall system that takes into account distributions of ROI and non-ROI and harmonize two distributions simultaneously.

6. Conclusions

Hair loss, i.e., no hair in areas where there should normally be hair, is not only a matter of visual appearance but also negatively affects individual self-esteem. However, no practical solutions can currently predict the outcomes of hair transplant surgery. Even though we can deliberate over the GAN-based image translation models that showed good performance, such models have a problem in translating regions other than the surgical area. To overcome the limitations of other image translation models and to generate high-quality predictive images, we proposed a new GAN-based ROI image translation method that combined image translation with image segmentation. As for the proposed method, since it conducts image translation and image segmentation independently, the method generated predictive images that translated the ROI only retaining the non-ROI. Furthermore, we applied the ensemble method to the image segmentation phase to supplement a single image segmentation model's weaknesses on various images. From the experiments using various quantitative metrics, we showed that the proposed method was better than the other GAN-based image translation methods by 23%, 452%, and 42% in SSIM, IoU, and FID, respectively. In addition, the ensemble method proposed for image segmentation showed better performance than a single segmentation model in the ROI detection. Specifically, the ensemble model had more robust performances in preoperative images taken from various angles. From these experimental results, we observed that the predictive image of the proposed method preserved the information of the input image and properly predicted the hair transplant surgery output.

Author Contributions: Conceptualization, D.-Y.H., M.K. and Y.-H.C.; methodology, D.-Y.H. and Y.-H.C.; software, D.-Y.H.; validation, D.-Y.H., S.-H.C. and J.S.; formal analysis, D.-Y.H. and Y.-H.C.; investigation, D.-Y.H., S.-H.C. and J.S.; resources, D.-Y.H. and S.-H.C.; data curation, M.K.; writing—original draft preparation, D.-Y.H. and S.-H.C.; writing—review and editing, S.-H.C., M.K. and Y.-H.C.; visualization, D.-Y.H.; supervision and project administration, M.K. and Y.-H.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study protocol was approved by Yonsei University Health System, Severance Hospital, Institutional Review Board (IRB No. 4-2020-1338). All clinical investigations were conducted in accordance with the guidelines of Declaration of Helsinki and Good Clinical Practice guidelines.

Informed Consent Statement: Written informed consent was obtained from the patients to publish this paper.

Data Availability Statement: We did not report any data.

Acknowledgments: This work was supported by a Biomedical Research Institute grant, Kyungpook National University Hospital (2019).

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

References

1. Camacho, F.M.; García-Hernández, M.J. Psychological features of androgenetic alopecia 1. *J. Eur. Acad. Dermatol. Venereol.* **2002**, *16*, 476–480. [[CrossRef](#)] [[PubMed](#)]
2. Bater, K.L.; Ishii, M.; Joseph, A.; Su, P.; Nellis, J.; Ishii, L.E. Perception of hair transplant for androgenetic alopecia. *JAMA Fac. Plast. Surg.* **2016**, *18*, 413–418. [[CrossRef](#)] [[PubMed](#)]
3. Lee, S.; Lee, J.W.; Choe, S.J.; Yang, S.; Koh, S.B.; Ahn, Y.S.; Lee, W.S. Clinically Applicable Deep Learning Framework for Measurement of the Extent of Hair Loss in Patients With Alopecia Areata. *JAMA Dermatol.* **2020**, *156*, 1018–1020. [[CrossRef](#)] [[PubMed](#)]
4. Chang, W.J.; Chen, L.B.; Chen, M.C.; Chiu, Y.C.; Lin, J.Y. ScalpEye: A Deep Learning-Based Scalp Hair Inspection and Diagnosis System for Scalp Health. *IEEE Access* **2020**, *8*, 134826–134837. [[CrossRef](#)]
5. Kapoor, I.; Mishra, A. Automated classification method for early diagnosis of alopecia using machine learning. *JAMA Fac. Plast. Surg.* **2018**, *132*, 437–443. [[CrossRef](#)]
6. Kingma, D.P.; Welling, M. Auto-encoding variational bayes. *arXiv* **2013**, arXiv:1312.6114.
7. He, K.; Chen, X.; Xie, S.; Li, Y.; Dollár, P.; Girshick, R. Masked autoencoders are scalable vision learners. *arXiv* **2021**, arXiv:2111.06377.
8. Gao, J.; Tembine, H. Distributionally robust games: Wasserstein metric. In Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 8–13 July 2018; pp. 1–8.
9. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134.
10. Mejjati, Y.A.; Richardt, C.; Tompkin, J.; Cosker, D.; Kim, K.I. Unsupervised attention-guided image to image translation. *arXiv* **2018**, arXiv:1806.02311.
11. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)] [[PubMed](#)]
12. Shvets, A.A.; Rakhlin, A.; Kalinin, A.A.; Iglovikov, V.I. Automatic instrument segmentation in robot-assisted surgery using deep learning. In Proceedings of the 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, USA, 17–20 December 2018; pp. 624–628.
13. Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; Hochreiter, S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *arXiv* **2017**, arXiv:1706.08500.
14. 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, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
15. Zhao, K.; Zhou, L.; Gao, S.; Wang, X.; Wang, Y.; Zhao, X.; Wang, H.; Liu, K.; Zhu, Y.; Ye, H. Study of low-dose PET image recovery using supervised learning with CycleGAN. *PLoS ONE* **2020**, *15*, e0238455. [[CrossRef](#)] [[PubMed](#)]
16. Mathew, S.; Nadeem, S.; Kumari, S.; Kaufman, A. Augmenting Colonoscopy using Extended and Directional CycleGAN for Lossy Image Translation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 4696–4705.
17. Chen, X.; Xu, C.; Yang, X.; Tao, D. Attention-gan for object transfiguration in wild images. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 164–180.
18. Liu, M.Y.; Breuel, T.; Kautz, J. Unsupervised image-to-image translation networks. *arXiv* **2017**, arXiv:1703.00848.
19. 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; pp. 172–189.
20. Jo, Y.; Park, J. SC-FEGAN: Face editing generative adversarial network with user’s sketch and color. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 1745–1753.
21. Tan, Z.; Chai, M.; Chen, D.; Liao, J.; Chu, Q.; Yuan, L.; Tulyakov, S.; Yu, N. MichiGAN: Multi-input-conditioned hair image generation for portrait editing. *arXiv* **2020**, arXiv:2010.16417.
22. Saha, R.; Duke, B.; Shkurti, F.; Taylor, G.W.; Aarabi, P. LOHO: Latent Optimization of Hairstyles via Orthogonalization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 1984–1993.
23. Fan, W.; Fan, J.; Yu, G.; Fu, B.; Chen, T. HSEGAN: Hair Synthesis and Editing Using Structure-Adaptive Normalization on Generative Adversarial Network. In Proceedings of the 2021 IEEE International Conference on Image Processing (ICIP), Anchorage, AK, USA, 19–22 September 2021; pp. 1324–1328.
24. Emami, H.; Aliabadi, M.M.; Dong, M.; Chinnam, R.B. Spa-gan: Spatial attention gan for image-to-image translation. *IEEE Trans. Multimed.* **2020**, *23*, 391–401. [[CrossRef](#)]
25. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
26. Chaurasia, A.; Culurciello, E. Linknet: Exploiting encoder representations for efficient semantic segmentation. In Proceedings of the 2017 IEEE Visual Communications and Image Processing (VCIP), St. Petersburg, FL, USA, 10–13 December 2017; pp. 1–4.
27. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.

28. McGlinchy, J.; Johnson, B.; Muller, B.; Joseph, M.; Diaz, J. Application of UNet fully convolutional neural network to impervious surface segmentation in urban environment from high resolution satellite imagery. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 3915–3918.
29. Zhou, L.; Zhang, C.; Wu, M. D-linknet: Linknet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 182–186.
30. Kuang, H.; Wang, B.; An, J.; Zhang, M.; Zhang, Z. Voxel-FPN: Multi-scale voxel feature aggregation for 3D object detection from LIDAR point clouds. *Sensors* **2020**, *20*, 704. [[CrossRef](#)] [[PubMed](#)]
31. Kholiavchenko, M.; Sirazitdinov, I.; Kubrak, K.; Badrutdinova, R.; Kuleev, R.; Yuan, Y.; Vrtovec, T.; Ibragimov, B. Contour-aware multi-label chest X-ray organ segmentation. *Int. J. Comput. Assist. Radiol. Surg.* **2020**, *15*, 425–436. [[CrossRef](#)] [[PubMed](#)]
32. Yang, J.; Zhao, Y.; Liu, J.; Jiang, B.; Meng, Q.; Lu, W.; Gao, X. No reference quality assessment for screen content images using stacked autoencoders in pictorial and textual regions. *IEEE Trans. Cybern.* **2020**. [[CrossRef](#)] [[PubMed](#)]
33. Sim, K.; Yang, J.; Lu, W.; Gao, X. MaD-DLS: Mean and deviation of deep and local similarity for image quality assessment. *IEEE Trans. Multimed.* **2020**. [[CrossRef](#)]
34. Park, T.; Efros, A.A.; Zhang, R.; Zhu, J.Y. Contrastive learning for unpaired image-to-image translation. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 319–345.
35. Lee, H.Y.; Tseng, H.Y.; Huang, J.B.; Singh, M.; Yang, M.H. Diverse image-to-image translation via disentangled representations. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 35–51.