applied sciences

MDPI

Article

Predicting Post-Liposuction Body Shape Using RGB Image-to-Image Translation

Minji Kim †, Jiseong Byeon †, Jihun Chang and Sekyoung Youm *

Department of Industrial and Systems Engineering, Dongguk University, Seoul 04620, Republic of Korea; mostar39@naver.com (M.K.); quswltjd12@naver.com (J.B.); auau0918@naver.com (J.C.)

* Correspondence: skyoum@dongguk.edu; Tel.: +82-2-2260-3377

† These authors contributed equally to this work.

Abstract: The growing interest in weight management has elevated the popularity of liposuction. Individuals deciding whether to undergo liposuction must rely on a doctor's subjective projections or surgical outcomes for other people to gauge how their own body shape will change. However, such predictions may not be accurate. Although deep learning technology has recently achieved breakthroughs in analyzing medical images and rendering diagnoses, predicting surgical outcomes based on medical images outside clinical settings remains challenging. Hence, this study aimed to develop a method for predicting body shape changes after liposuction using only images of the subject's own body. To achieve this, we utilize data augmentation based on a conditional continuous Generative Adversarial Network (CcGAN), which generates realistic synthetic data conditioned on continuous variables. Additionally, we modify the loss function of Pix2Pix—a supervised image-to-image translation technique based on Generative Adversarial Networks (GANs)—to enhance prediction quality. Our approach quantitatively and qualitatively demonstrates that accurate, intuitive predictions before liposuction are possible.

Keywords: prediction of liposuction outcome; image-to-image translation; Pix2Pix; deep learning; GAN

check for updates

Academic Editors: Antonio Fernández-Caballero, Haitao Zhao and Meng Wang

Received: 11 February 2025
Revised: 23 April 2025
Accepted: 23 April 2025
Published: 25 April 2025

Citation: Kim, M.; Byeon, J.; Chang, J.; Youm, S. Predicting Post-Liposuction Body Shape Using RGB Image-to-Image Translation. *Appl. Sci.* **2025**, *15*, 4787. https://doi.org/10.3390/app15094787

Copyright: © 2025 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

With growing interest in weight management worldwide, liposuction—a surgical obesity treatment that removes accumulated mast cells—has been garnering attention as well [1]. However, liposuction substantially alters an individual's appearance; hence, the patient cannot determine exactly how their appearance will change after the operation. Accordingly, if patients can predict their post-surgery appearance before the surgery, then they can understand the potential outcome better, which can help them develop a more positive perception of the operation [2]. Currently, to predict the changes to their own appearance post-surgery, patients must rely on subjective projections based on the doctor's surgical experience, which may not always be accurate [3]. Additionally, a certain liposuction clinic in South Korea uses physical body measurements (e.g., height and weight) and lifestyle details (e.g., drinking habits) to find the previous patients most similar to the potential new patients, who are then shown images of these previous patients before and after liposuction [4]; this allows patients to gain an intuitive idea of how their own appearance may change after liposuction. However, because such predictions are not based on the unique physical features of the actual patient, they may differ from the final surgical outcome. Therefore, to overcome these limitations, post-liposuction appearance must be predicted based on the potential patient's own features.

In this study, we constructed a paired image dataset consisting of RGB images taken before and after liposuction and developed a Pix2Pix architecture with a modified loss function to predict post-liposuction appearance based on a pre-liposuction image that can be easily captured during everyday life. Our approach does not incur any separate optimization time and reduces blur. This study marks the first attempt to predict the appearance of patients after liposuction using RGB images. The proposed approach represents an accessible tool for making informed decisions about liposuction.

2. Related Work

Babajide et al. [5] and Weber and Achananuparp [6] focused on predicting future weight and the success of diet control based on factors such as diet management, body fat, sleep time, and physical condition. However, the methods developed in these studies cannot predict visual changes in individuals. To address this limitation, Kuzmar et al. [7] developed a technique to predict 3D body shape based on height, weight, and four other body measurements using the Skinned Multi-Person Linear (SMPL) model. Similarly, Zeng et al. [8] incorporated height, weight, and 17 other more deformable body dimensions to obtain more detailed body shape predictions. However, as these studies primarily utilized databases related to Western populations, their findings may not be applicable to Koreans. Additionally, relying solely on height and weight may not be sufficient for determining 3D body shapes similar to the subject's, while incorporating additional human body measurements into a prediction model can be a complex process. Furthermore, 3D models based on the SMPL may not resemble the subject perfectly, as the SMPL model is based on a mannequin.

Image-to-image translation (I2I) is a computer vision technique that transforms input images into corresponding output images. This technique is particularly useful for predicting changes in images based on specific criteria. Advancements in deep learning have increased its utilization in the medical field as well. A prominent research focus within I2I is the prediction of changes in facial style. For instance, Karras et al. introduced the Style GAN and Style GAN2 algorithms, which enable various style transformations in facial images, such as alterations in hair color, based on human facial data [9,10]. These algorithms utilize deep learning to generate realistic and accurate transformations in facial appearance. In a related study, URSF-GAN [11] was proposed as an unsupervised framework for face frontalization that lifts single-view 2D facial images into 3D space, applies symmetric texture filling, and leverages a Pix2Pix-based network to eliminate artifacts. This approach shows that combining geometric priors such as 3D rotation with adversarial learning can produce photorealistic and identity-preserving facial outputs, even without paired training data.

Generative adversarial networks (GANs) [12] offer various capabilities, such as data augmentation and Image-to-Image(I2I), which have become increasingly important in the medical field [13,14]. Specifically, conditional GAN (cGAN) models have been used for pixel unit prediction in various I2I applications involving medical images [13,15,16]. Representative examples include the prediction of visceral and subcutaneous adipose tissue maps based on 2D body shape data obtained by projecting reconstructed 3D Computed Tomography(CT) data in one direction [15] and long-term predictions based on X-ray radiography to help doctors with segmentation and diagnosis by erasing the spine [16].

However, the aforementioned methods are not readily accessible to the general population because of the difficulty in acquiring appropriate data. GAN-based I2I can be used to predict postoperative appearance based on a preoperative image [17]. I2I has also been employed to predict patient appearance after orbital decompression using RGB images [13]. However, since the input image cannot be translated directly, latent vector

optimization must be performed, which is a notable limitation as it requires about 20 min to complete. This problem can be overcome by employing the Pix2Pix algorithm, which allows for constructing a latent space using the image itself as an input before performing the prediction, all without requiring a separate latent optimization process [18]. However, studies have reported that Pix2Pix creates blurry images [13,19,20].

In addition to the advancements in facial style prediction, Choi et al. proposed Star GAN, which is a single model that can accommodate multiple data domains, such as age, hair color, and gender [21], thereby enabling more versatile I2I across different attributes. Researchers have also explored the application of I2I techniques in the medical field. For example, Yoo et al. predicted patient appearances post-orbital decompression surgery based on pre-surgery images, assisting patients in making informed decisions considering the potential surgical outcomes [13]. Similarly, Wang et al. employed I2I to predict maps of visceral fat and subcutaneous fat using 2D body shape data derived from 3D reconstructions of CT scans [15]. These studies demonstrate the potential of I2I techniques in various medical applications. However, despite active research related to facial changes in medical applications, few studies have focused on the entire human body because of the limited availability of comprehensive body shape data.

3. Materials and Methods

Figure 1 presents the proposed framework, which integrates data augmentation using CcGAN and image-to-image translation with Pix2Pix to predict post-liposuction body shape from preoperative RGB images. Both real and augmented paired data are used to compensate for the limited sample size.

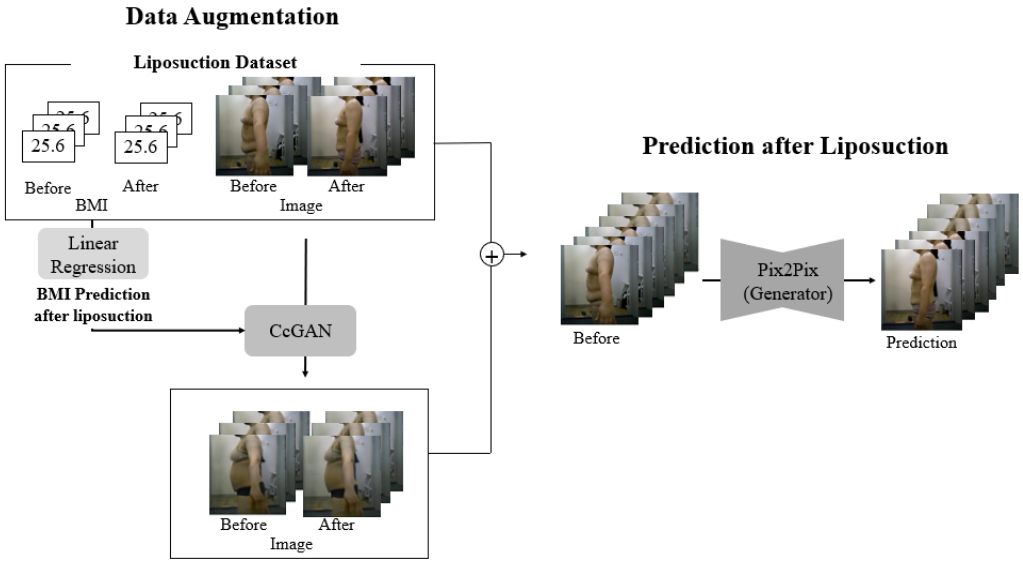


Figure 1. Overall prediction framework. This figure illustrates the complete pipeline for predicting post-liposuction body shape using RGB images. The process begins with a paired dataset containing pre- and post-liposuction images and corresponding BMI values. To address the limited sample size, data augmentation is performed using a Continuous Conditional GAN (CcGAN), which generates synthetic images conditioned on continuous BMI values. A linear regression model estimates post-liposuction BMI from pre-liposuction BMI, enabling the generation of realistic pre- and postoperative image pairs for the same individual. The original and augmented datasets are then combined to train a Pix2Pix-based image-to-image translation model, which predicts postoperative body shape from preoperative images. This framework enhances prediction accuracy and visual realism by integrating data augmentation and loss-function improvements.

(Data Collection) Since post-liposuction appearance must be predicted using images captured before liposuction, data to be collected must satisfy several conditions: First, to clearly determine the change in body shape, the images must depict the subject with only underwear on. Second, image pairs depicting the same individual before and after liposuction are required. Accordingly, images of 230 patients who underwent liposuction were collected. This study's protocol was approved by the Institutional Review Board of Dongguk University (DUIRB-202202-14), and the relevant procedures complied with the principles of the Declaration of Helsinki. On the day of liposuction, the patient captured images of their upper and lower body in a fixed position from the front, left, right, and back; two months after liposuction, they captured the same images to record the postoperative changes. All images were captured with the camera 110 cm above the ground and 100 cm away from the heel. To ensure consistency in posture during image acquisition, all participants were instructed to stand with their arms naturally at their sides, feet shoulder-width apart, and eyes facing forward. However, due to the absence of explicit guidelines regarding arm angles or spacing between the arms and torso, individual variations in arm positioning occurred. This led to ambiguous boundaries between the abdomen and the arms in some cases. All images were captured in a closed studio environment without windows to eliminate the influence of natural light, using consistent lighting conditions and fixed camera settings. For participant safety and comfort, the entire process was supervised by medical professionals. A total of 920 images were collected, including front, back, left, and right views of the upper and lower body. Among these, 230 left-side upper body images were selected for model training and evaluation, as this viewpoint most clearly captured changes in the abdominal and flank areas and provided a distinct view of the target region for prediction. The images were taken under controlled conditions in a clinical environment using a DSLR camera (Canon EOS 800D (Canon Inc., Tokyo, Japan), equipped with a 24.2 MP APS-C sensor). All subjects were photographed against a matte white background, and the lighting setup followed a three-point lighting system (left and right-side lights and a front-facing soft light) to minimize skin reflections and shadow artifacts. Continuous lighting with a color temperature of 5500 K was used, and no flash was applied. Additionally, before each session, subjects were instructed to remain under the lighting for five seconds to allow their skin tone to adjust to the environment. Lighting intensity and camera exposure settings were kept consistent across all sessions. These conditions were designed to ensure image consistency and allow accurate comparison of pre- and postoperative body shapes without distortion from lighting variations. All experiments were implemented in Python 3.9.7 using PyTorch 1.13.1 and torchvision 0.14.1. Model training was performed with CUDA 11.4 on an NVIDIA RTX 2060 GPU (NVIDIA Corporation, Santa Clara, CA, USA). Body mass index (BMI) data for the patients were also collected.

(Data Augmentation) As only 230 image pairs were available, data augmentation was necessary to obtain meaningful results for analysis [22,23]. Surgical procedures such as double-eyelid surgery and orbital decompression involve a clear and consistent pattern of alteration in appearance. By contrast, although liposuction always results in weight loss, determining whether a specific image depicts pre-liposuction or post-liposuction appearance is impossible. Specifically, for slim people undergoing liposuction [24], the amount of fat removable is relatively limited, whereas for heavier individuals, removing too much in a single session could be risky [25]. Thus, images of certain individuals after liposuction could depict more fat than images of other individuals before liposuction. Therefore, conventional GAN-based data augmentation, which relies on discrete categorical labels, could not be performed. Instead, data augmentation was performed using continuous cGAN (CcGAN) [25] based on BMI [26,27], a continuous variable that reflects overall body

characteristics and is commonly used to evaluate obesity. Figure 2 presents side-by-side examples of actual patient images and augmented images generated by the CcGAN. Each pair includes both real and synthetic pre- and post-liposuction images, demonstrating that the generated images realistically reflect changes in BMI.

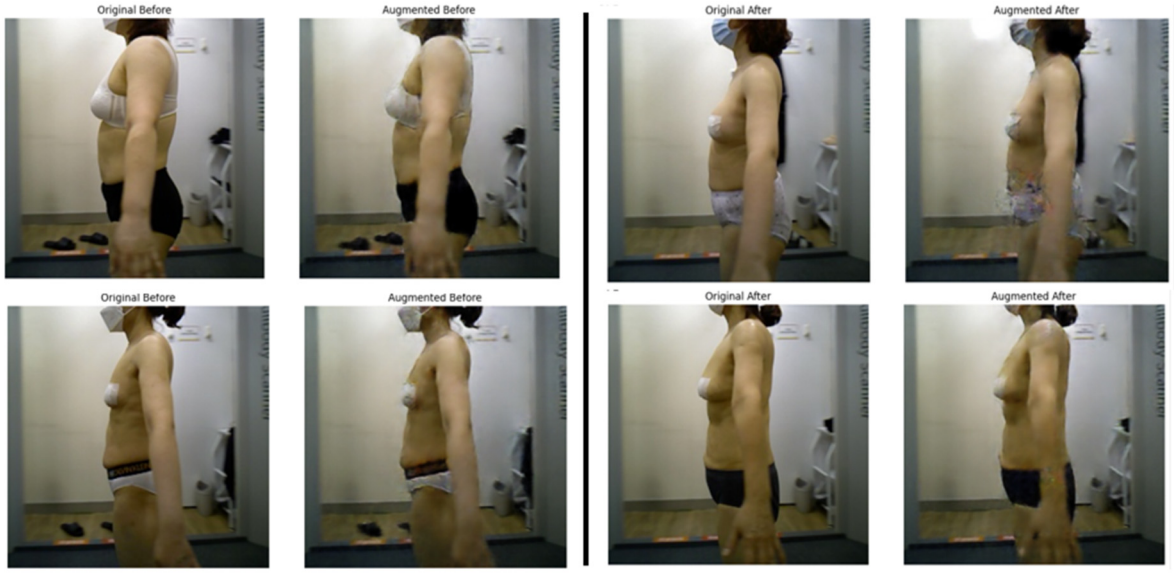


Figure 2. Examples of real and CcGAN-generated images before and after liposuction This figure presents side-by-side comparisons of actual patient images (Original) and CcGAN-generated synthetic images (Augmented) before and after liposuction. Each row corresponds to one patient, with pre- and postoperative images shown for both real and generated cases. The augmented images were produced using fixed noise vectors and BMI values estimated via linear regression to condition the CcGAN. While the generated images generally resemble the real ones and reflect BMI-associated body shape changes, some visual artifacts or soft boundaries—particularly around the abdomen or arms—may be observed. These examples are intended to provide readers with an intuitive understanding of the characteristics of both the original and augmented datasets used for training.

(Linear Regression) Based on the BMI distributions before and after liposuction in Figure 3a, liposuction appears to shift the BMI to the left (i.e., toward small values). A paired t-test on the BMI values before and after liposuction showed a p-value of 1.0210×10^{-33}, which indicates a significant difference. A Pearson correlation analysis was also performed to determine whether the BMI values before and after liposuction share a linear relationship [28]. Pearson correlation coefficients closer to 1 indicate more positive linear relationships [29]. The calculated Pearson correlation coefficient in this study was 0.9552, which indicates a highly linear relationship. Figure 3b presents a distribution graph visualizing this linear relationship. Therefore, we performed linear regression to predict the BMI after liposuction based on the BMI before liposuction. Based on these predictions, paired data samples (images before and after liposuction) were generated during data augmentation.

(CcGAN) CcGAN is a generative architecture for conditional image generation based on continuous numerical values as auxiliary information instead of discrete class labels [30]. While existing cGANs generate images mainly based on categorical labels, they cannot directly utilize continuous data as labels. This limitation is addressed by CcGAN, which comprises a generator and a discriminator. Figure 4 illustrates the CcGAN architecture used in this study. Both the generator and discriminator in CcGAN are implemented as CNNs. The generator receives a noise vector and embedded BMI value, processed through a series of convolutional layers and upsampling blocks. Specifically, the generator consists of five transposed convolutional layers with filter sizes of 512, 256, 128, 64, and 3, respectively.

Each layer uses a 4 × 4 kernel, stride of 2, and padding of 1. Batch normalization and ReLU activation are applied to all layers except the last, which uses a tanh activation function to produce a 3-channel RGB image.

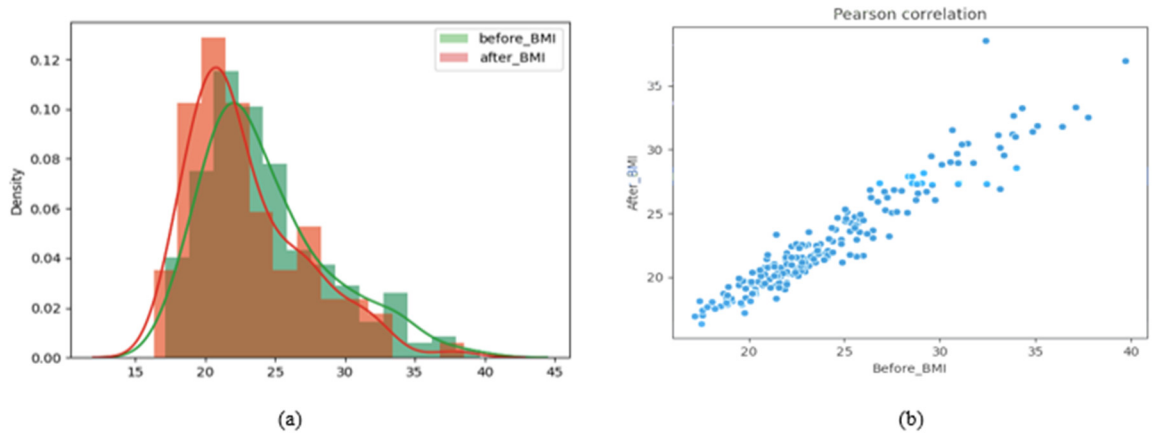


Figure 3. (**a**) Histogram showing the distribution of BMI before and after liposuction. The post-liposuction BMI values are generally lower, indicating an overall reduction in body mass index. (**b**) Scatter plot of pre- and post-liposuction BMI, revealing a strong linear correlation (Pearson r = 0.9552). This result supports the use of linear regression to predict post-liposuction BMI from pre-liposuction values for data augmentation.

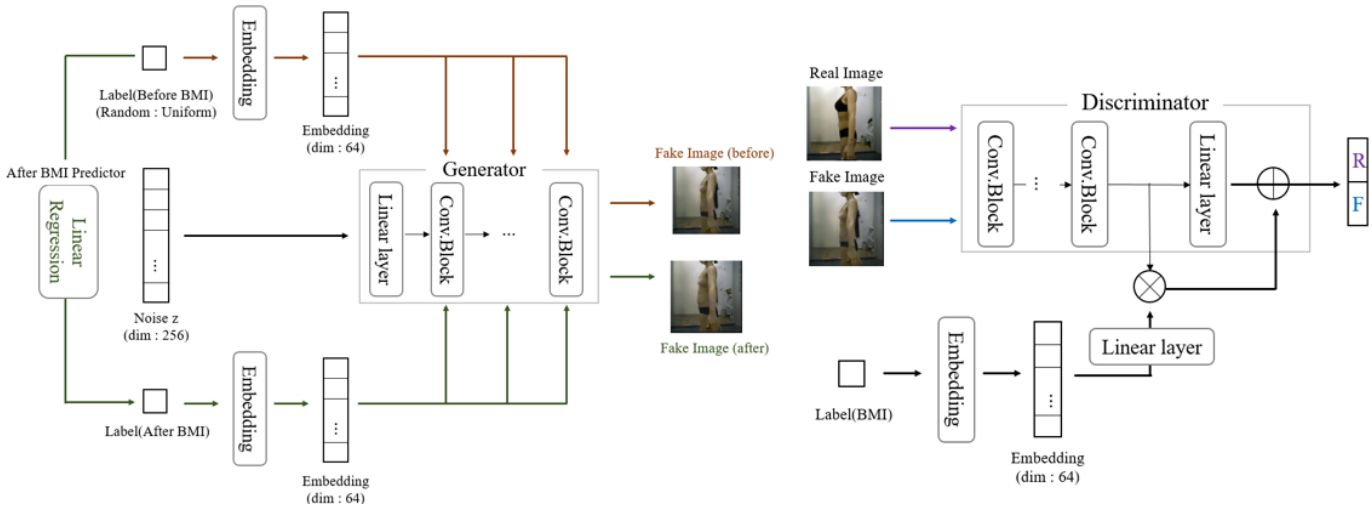


Figure 4. Architecture of the Continuous Conditional Generative Adversarial Network (CcGAN) used for data augmentation. The generator receives a noise vector z and a continuous BMI value (before or after liposuction), which is embedded into a fixed-dimensional vector and concatenated with the noise input. The generator outputs synthetic images corresponding to the specified BMI. A linear regression model is used to derive the "after" BMI from a sampled "before" BMI, allowing for paired generation of pre- and post-liposuction images using the same noise. The discriminator receives real or generated images and evaluates their realism while simultaneously verifying whether the image condition (BMI) matches the embedded label. This architecture enables realistic and condition-consistent data augmentation for both pre- and post-liposuction images.

The discriminator evaluates the realism of the image using a CNN that processes the image and embedded label through parallel streams before concatenation and final classification. The discriminator consists of five convolutional layers with filter sizes of 64, 128, 256, 512, and 1. All layers use a 4 × 4 kernel. The first three layers apply a stride of 2, and the last two apply a stride of 1. Instance normalization is applied from the second layer onward, and LeakyReLU (slope = 0.2) is used as the activation function. The final

layer uses a sigmoid function to classify real vs. fake. More detailed descriptions of the generator and discriminator structures are provided in the later section titled 'Network Architecture: Pix2Pix'.

The generator generates images based on random noise z and a continuous label. The continuous label is transformed to a specific dimension through embedding before being used. Unlike Pix2Pix, CcGAN does not require an input image, as it synthesizes images from a random noise vector and condition value. Therefore, it does not employ an encoder–decoder structure such as U-Net. Instead, it uses a stack of transposed convolutional layers to upsample the input latent representation into a full-resolution image gradually. We performed data augmentation to create pre-liposuction and post-liposuction images for the same person.

While the noise z was fixed, the BMI values before and after liposuction were utilized differently to generate images of a person before and after liposuction, respectively. Fixing the noise vector z ensures that the identity-related latent features remain constant across the generated pre- and post-liposuction images. This design allows the BMI condition to modulate only the morphological aspects relevant to fat reduction, thereby enabling identity-consistent yet BMI-dependent image generation. Such a strategy aligns with prior work—InfoGAN [31]—that explicitly separates incompressible noise from semantic latent codes to disentangle identity and attribute factors in conditional generative models. This approach is particularly useful for simulating before-and-after transformations in medical imaging, where preserving individual identity is critical. Specifically, z was randomly derived from a Gaussian function, and the pre-liposuction BMI was derived using a uniform function based on the smallest (16.3176) and largest (39.7025) BMI values in the real dataset. Accordingly, pre-liposuction images were generated by combining the derived pre-liposuction BMI with z.

Further, the BMI after liposuction was predicted via linear regression based on the BMI before liposuction, as mentioned earlier. Post-liposuction images were then generated by combining the predicted post-liposuction BMI with noise z. Thus, 2000 pairs of pre-liposuction and post-liposuction images were created in total. Finally, the complete dataset was separated into training and testing subsets, as outlined in Table 1.

Table 1. Liposuction dataset used in this study. The table presents the number of paired images (pre- and post-liposuction) used for training and testing. A total of 230 real image pairs were collected from actual patients, and an additional 2000 synthetic pairs were generated using CcGAN, conditioned on BMI before and after liposuction. For each subject, the same noise vector was used to generate both pre- and post-liposuction images based on BMI values predicted through linear regression. The full dataset was divided into training and test sets, as shown.

	Real Liposuction Data	Augmented Data	Total
Train	200 pairs	1800 pairs	2000 pairs
Test	30 pairs	200 pairs	230 pairs
Total	230 pairs	2000 pairs	2230 pairs

To confirm whether the generated images corresponded to the input BMI value, we designed a model based on a pretrained ResNet34 architecture that predicts BMI using body images as input. ResNet34 [32] offers excellent image classification performance [33] and can be used to derive feature maps reflecting the characteristics of any given input images [32]. The BMI predictor was trained using BMI data from a total of 400 pre-liposuction and post-liposuction images. For training the BMI predictor, we utilized the Adam optimizer with a learning rate of 0.001, beta values of (0.9, 0.999), and a batch size of 16. The model was trained for 160 epochs, with the mean squared error (MSE) loss between

the actual BMI and predicted BMI as the optimization criterion. To ensure robust training, we applied a weight decay of 0.0001 for regularization to prevent overfitting. The dataset was split into 80% training and 20% validation to monitor generalization performance. The learning rate was reduced by a factor of 0.1 if the validation loss did not improve for 10 consecutive epochs using a learning rate scheduler. The model with the smallest validation MSE loss, achieved at epoch 160 with an MSE loss of 0.0082, was selected as the final BMI predictor. The ResNet34 model was implemented using PyTorch 1.13.1 and torchvision 0.14.1. Training was performed on an NVIDIA RTX 2060 GPU (NVIDIA Corporation, Santa Clara, CA, USA), taking approximately 4 h. The input images were resized to 256 × 256 pixels and normalized using ImageNet mean and standard deviation values to align with the pretrained Resnet34 weights. This model was used to evaluate the validity of data generated by CcGAN.

(Network Architecture: Pix2Pix) As our goal was to generate a post-liposuction image based on a pre-liposuction image, we utilized I2I. Most studies on I2I have employed convolutional neural networks with encoder–decoder architectures [18,34–36]. Pix2Pix is a representative model that builds upon the cGAN framework [37] to create realistic images via I2I [18]. Pix2Pix comprises a generator and a discriminator, which are trained in an adversarial manner, as shown in Figure 5.

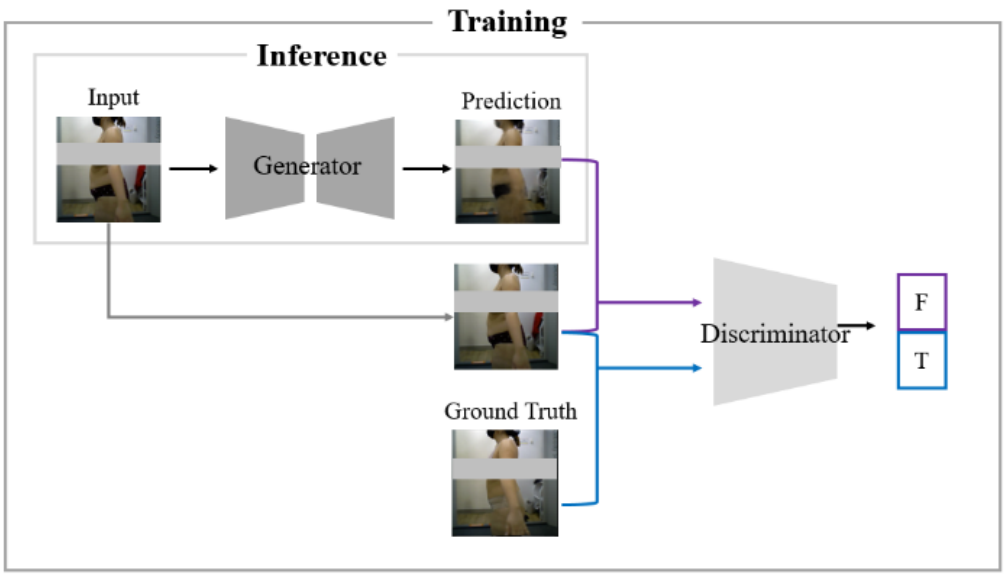


Figure 5. Overview of the Pix2Pix Architecture used for predicting post-liposuction images. During training, the generator learns to produce realistic post-liposuction images from pre-liposuction inputs, which are then compared with ground truth images. The discriminator receives both real (input + ground truth) and fake (input + generated) image pairs to distinguish between authentic and generated outputs. During inference, only the generator is used to produce predictions. This architecture is optimized using a combination of adversarial loss and reconstruction loss functions.

The generator of the Pix2Pix model is implemented as a Convolutional Neural Network (CNN) based on a U-Net architecture consisting of eight encoder layers and eight decoder layers. The encoder progressively reduces the spatial resolution of the input image (256 × 256 pixels) to 1 × 1, using convolutional layers with filter sizes of 64, 128, 256, 512, 512, 512, 512, and 512. Each encoder layer applies a convolution operation with a kernel size of 4 × 4, stride of 2, and padding of 1. Batch normalization and LeakyReLU activation (with a slope of 0.2) are applied to all encoder layers except the first. The decoder mirrors the encoder's structure and restores resolution using transposed convolutions with a kernel size of 4 × 4, stride of 2, and padding of 1. Each decoder layer uses ReLU activation and

batch normalization, except for the last layer, which applies a tanh activation function without normalization to produce a 3-channel RGB output. Specifically, the decoder consists of eight layers with filter sizes of 512, 512, 512, 512, 256, 128, 64, and 3, corresponding to each upsampling stage. In accordance with the original Pix2Pix implementation, dropout with a rate of 0.5 is applied to the first three decoder layers during training to improve generalization and prevent overfitting. Skip connections are implemented via channel-wise concatenation between each encoder layer and its corresponding decoder layer, allowing spatial information and fine-grained details to be preserved throughout the reconstruction process. This encoder–decoder CNN structure allows feature maps and spatial information to be effectively transmitted without loss [38–40]. The discriminator of the Pix2Pix model is implemented as a Convolutional Neural Network (CNN) and follows the PatchGAN architecture, which evaluates the realism of overlapping local image patches instead of assessing the entire image at once. This design is particularly effective for capturing high-frequency visual details, such as contours and textures, that are essential for generating realistic images. The discriminator receives as input a 256 × 256 × 6 tensor formed by concatenating a pre-liposuction image and either a real or generated post-liposuction image along the channel dimension. The network consists of five convolutional layers with filter sizes of 64, 128, 256, 512, and 1, respectively. All layers use a 4 × 4 kernel, with the first three layers applying a stride of 2 and the last two a stride of 1. A padding of 1 is applied to all layers to maintain spatial consistency throughout the network. Each layer uses a LeakyReLU activation function (with a slope of 0.2), and instance normalization is applied from the second layer onward to stabilize training and improve generalization. The final output layer uses a sigmoid activation function to predict the authenticity of each local patch. Unlike traditional discriminators that produce a single real/fake score for the whole image, PatchGAN outputs a probability map—for example, 30 × 30—where each value represents the likelihood that a particular image patch is real. During training, the discriminator is trained on two types of image pairs: one consisting of a pre-liposuction and a real post-liposuction image and the other consisting of a pre-liposuction and a generator-produced post-liposuction image. This patch-based discrimination encourages the generator to produce more realistic local details throughout the image [18].

The objective function of Pix2Pix comprises cGAN loss and reconstruction loss. Equation (1) defines cGAN loss, where x represents the pre-liposuction image, y denotes the post-liposuction image, and G is the generator. Therefore, $G(x)$ is a fake image produced by the generator using the input image x. In Equation (1), the error is calculated using cross-entropy. Equation (2) serves to reduce the loss in units of pixels as a reconstruction loss. The L1 loss signifies the error between the image created by the generator and the real post-liposuction image. Therefore, Equation (3) represents the final objective function based on Equations (1) and (2), where λ denotes a weight.

$$L_{cGAN}(G,\ D) = E_{x,y}[logD(x,y)] + E_x[\log(1 - D(x, G(x))], \tag{1}$$

$$L_{L1}(G) = E_{x,y}[||y - G(x)||_1], \tag{2}$$

$$\text{Object Function} = \min_G \max_D L_{cGAN}(G, D) + \lambda L_{L1}(G). \tag{3}$$

Several studies have shown that Pix2Pix produces blurry predictions [13,19,20]. As the L1 loss in Equation (2) represents the average pixel-wise difference between the predicted and real images, the model tends to produce smoothed results to minimize this average [41]. Thus, the model produces blurry predictions, specifically around high-frequency parts such as edges. To address this issue, the perceptual loss function based on VGGNet [42] has been proposed as an alternative [41,43]. Therefore, we used perceptual loss—which

considers the Euclidean distance of the activation layer of VGG16 pretrained using the ImageNet dataset [33]—together with L1 loss to represent the overall reconstruction loss.

In Equation (4), φ represents VGG16, and $\varphi_j(x)$ denotes the feature map derived from the j-th activation layer in the convolution layer. We used four ReLU Layers in total. As shown in Equation (5), the perceptual loss is the sum of the Euclidean distances of the feature maps corresponding to the real and predicted post-liposuction images. In Equation (4), W, H, and C denote the dimensions of the feature map. Equation (6) represents the final objective function used in this study, where λ signifies a weight.

$$l_j(G) = \frac{1}{C_j \times H_j \times W_j} ||\varphi_j(G(x)) - \varphi_j(y)||_2^2, \tag{4}$$

$$L_{VGG}(G) = \sum l_j(G), \tag{5}$$

$$\text{Object Function} = \min_G \max_D L_{cGAN}(G, D) + \lambda_{VGG} L_{VGG}(G) + \lambda_{L1} L_{L1}(G). \tag{6}$$

4. Results

(Data Augmentation) To determine whether the CcGAN correctly generated body shape images according to the input BMI value, we used the ResNet34-based BMI predictor, as mentioned in Section 3. The CcGAN was trained over a total of 20,000 iterations, and the generator model was saved every 1000 iterations. For training the CcGAN, we used the Adam optimizer with a learning rate of 0.0002, beta values of (0.5, 0.999), and a batch size of 32, following standard GAN training practices [12]. Continuous conditional labels (BMI values) were embedded into a 128-dimensional vector to ensure stable training. The model was trained for approximately 286 epochs, equivalent to 20,000 iterations, given the dataset size and batch size. The training was conducted on an NVIDIA RTX 2060 GPU. The images generated by each model were evaluated for accuracy by the BMI predictor. We calculated MSELoss between the predicted BMI and the actual BMI. As MSELoss approaches 0, the generated image means that the image corresponding to the target BMI Label is generated. Figure 5 presents a graph showing the MSELoss calculated every 1000 iterations, along with some of the images generated. During the initial iterations, the images were severely blurred, and the body type did not vary with BMI. The 14,000th-iteration generator model was deemed optimal as it achieved the smallest MSELoss (0.0119) and could generate appropriate body shapes for different BMI values. The 19,000th-iteration model suffered mode collapse [40] and could generate only similar images or visually abnormal images. Therefore, the 14,000th-iteration generator model was used for data augmentation.

Equation (7) presents the linear regression equation for calculating the BMI after liposuction based on the BMI before liposuction. Table 2 summarizes the metrics used to evaluate the effectiveness of the linear regression process. R^2 represents the explanatory power of linear regression; the closer its value is to 1, the more accurate the prediction. The R^2 value was determined to be 0.912 in this study, which underscores the accuracy of our predictive model.

$$\text{Post-liposuction BMI} = 0.8706 \times \text{Pre-liposuction BMI} + 1.7281. \tag{7}$$

Table 2. Performance metrics of linear regression predicting post-liposuction BMI from pre-liposuction BMI. This table presents the results of a linear regression model used to estimate BMI after liposuction based on BMI before surgery. The model achieved a high R^2 value of 0.912, indicating strong predictive power. The F-statistics and corresponding *p*-value (1.76×10^{-106}) confirm the statistical significance of the model. These results validate the use of linear regression for estimating post-surgical BMI, which was then used to condition synthetic image generation in the CcGAN framework.

Metric	Value
R^2	0.912
F-statistics	3063
Prob (F-statistics)	1.76×10^{-106}

With a fixed noise z, the CcGAN generator created images of individuals before and after liposuction. Figure 6 shows the images generated considering a total of 4 noise z. When the BMI before liposuction was low, no significant difference could be observed between the images before and after liposuction. This occurs because the BMI values before and after liposuction share a linear relationship, whereby small BMI values are not significantly altered after liposuction. Accordingly, the larger the BMI before liposuction, the more pronounced the difference in body shape between the images before and after liposuction. The pre-liposuction and post-liposuction images in Figure 7 can be confirmed as depicting the same person for every z value.

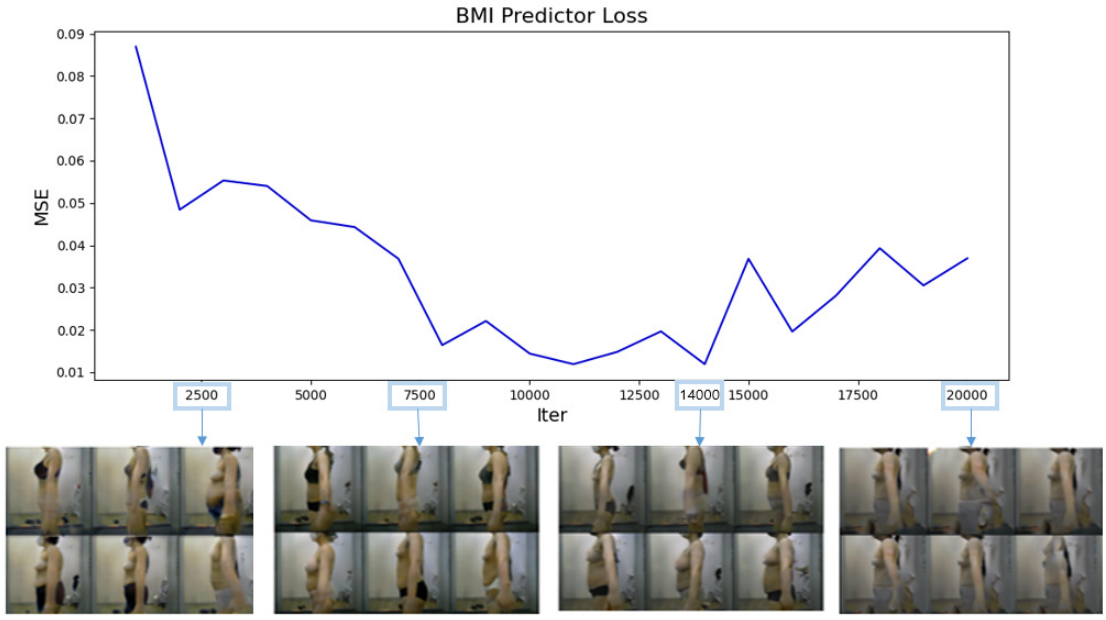


Figure 6. BMI Predictor Loss (mean squared error) and Corresponding Image Samples at Key Iterations during CcGAN training. The graph shows the change in MSE loss values predicted by the ResNet34-based BMI predictor for images generated by CcGAN at every 1000 iterations. Lower loss indicates better alignment between the generated image and the target BMI condition. Images below the graph correspond to models saved at selected iterations (2500, 7500, 14,000, and 20,000). Blurring and lack of shape variation were observed at early iterations. The model at 14,000 iterations achieved the lowest MSE (0.0119), producing images that visually align with target BMI values. The severe mode collapse occurred at 19,000 iterations. Based on this evaluation, the 14,000-iteration generator was selected for final data augmentation.

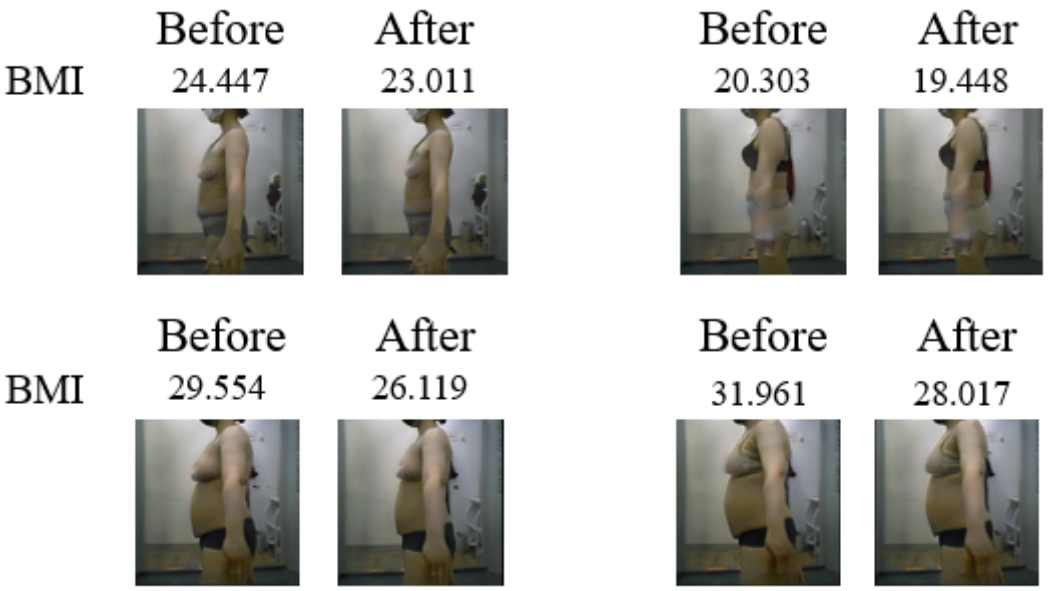


Figure 7. Examples of CcGAN-generated paired pre- and post-liposuction images using fixed noise and varying BMI values. This figure illustrates synthetic image pairs generated by CcGAN with a fixed noise vector and BMI values predicted through linear regression. Each row presents a different sample, where the left image represents the pre-liposuction body shape, and the right image depicts the post-liposuction body shape. As BMI increases, the visual difference between pre- and post-liposuction images becomes more pronounced. This result highlights that CcGAN successfully generates identity-consistent and BMI-conditioned transformations.

(Predicted Image after Liposuction) Figure 8 shows the reconstruction loss trend. Referring to the objective function in Equation (6), λ_{VGG} and λ_{L1} were set to 10 and 100, respectively, before training the model. Without data augmentation, as the epochs progressed, the model's effectiveness became increasingly dependent on training data, which ultimately resulted in overfitting. Thus, when the model was evaluated on test data, its reconstruction loss showed an increasing trend, which indicates that it did not learn appropriately. Conversely, when data augmentation was performed, the model achieved stable learning.

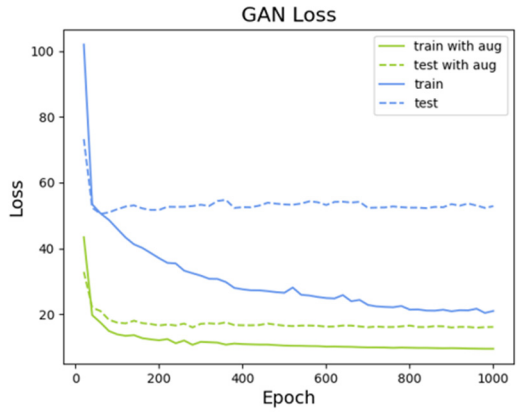


Figure 8. Comparison of GAN training and test loss trends with and without data augmentation over 1000 epochs. This figure shows the training and testing loss curves of the Pix2Pix-based model under two conditions: with data augmentation (green lines) and without augmentation (blue lines). Solid lines represent training loss, and dashed lines indicate test loss. Without augmentation, the model exhibits overfitting, as seen in the increasing test loss despite decreasing training loss. In contrast, with CcGAN-based data augmentation, both training and test losses decrease stably, indicating improved generalization and learning stability across epochs.

The images generated by the model without and with data augmentation were qualitatively compared. Figure 8 shows the reconstruction loss trend for the model trained over 860 epochs, which achieved the smallest error. Without data augmentation, although the minimum error was initially achieved, the model was not trained properly; hence, the 200-epoch model—the one with the minimum error—was trained for additional epochs. As seen in Figure 9, the images generated by the model without data augmentation appeared distorted; however, when data augmentation was performed, although some blurring was still observed around the arm, it was noticeably less than without data augmentation. Pix2Pix was trained using the Adam optimizer with a learning rate of 0.0002 and beta values of (0.5, 0.999), following standard GAN training practices. The Adam optimizer was used as implemented in PyTorch 1.13.1, and model input preprocessing followed the ImageNet statistics for compatibility with the pretrained VGG16 (torchvision 0.14.1) used in the perceptual loss calculation. The batch size was set to 16, and the model was trained for up to 1000 epochs. The objective function combined cGAN loss, L1 loss, and perceptual loss, with their weights empirically set to $\lambda_{L1} = 100$ and $\lambda_{VGG} = 10$. All training was conducted on an NVIDIA RTX 2060 GPU. Input images were resized to 256 $\times$ 256 pixels and normalized using ImageNet statistics, specifically by subtracting the channel-wise mean [0.485, 0.456, 0.406] and dividing by the standard deviation [0.229, 0.224, 0.225]. These preprocessing settings ensured consistency with the pretrained VGG16 model used in the perceptual loss calculation. These settings enabled stable training and ensured that the model captured both pixel-level and structural differences between pre- and post-liposuction images. Figure 10 provides several examples of predicted post-liposuction images generated by the Pix2Pix model trained with L1 and perceptual loss.

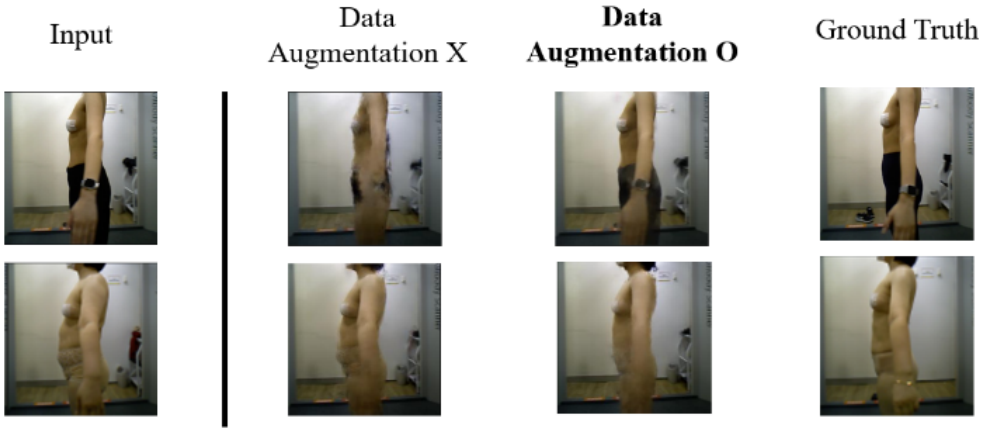


Figure 9. Qualitative comparison of predicted post-liposuction images with and without data augmentation. This figure compares the predicted results from models trained without augmentation (second column) and with CcGAN-based data augmentation (third column), given the same pre-liposuction input images (first column). Ground truth post-liposuction images are shown in the last column for reference. Without augmentation, the generated images exhibit significant artifacts and structural inconsistencies, while the augmented model produces more realistic contours despite minor blurring because of arm pose variability. These results demonstrate that data augmentation helps improve prediction fidelity and visual quality.

This blurring was primarily caused by inconsistent arm placement during image acquisition. Although the camera height and distance were fixed, no explicit instructions were given regarding arm angle or position, leading to variations between subjects. As a result, the boundary between the arms and torso became ambiguous, especially in the abdominal region—reducing the clarity of the predicted contour. This represents a failure

case where insufficient posture standardization during data collection negatively impacted prediction accuracy. The corresponding results are shown in Figure 11.

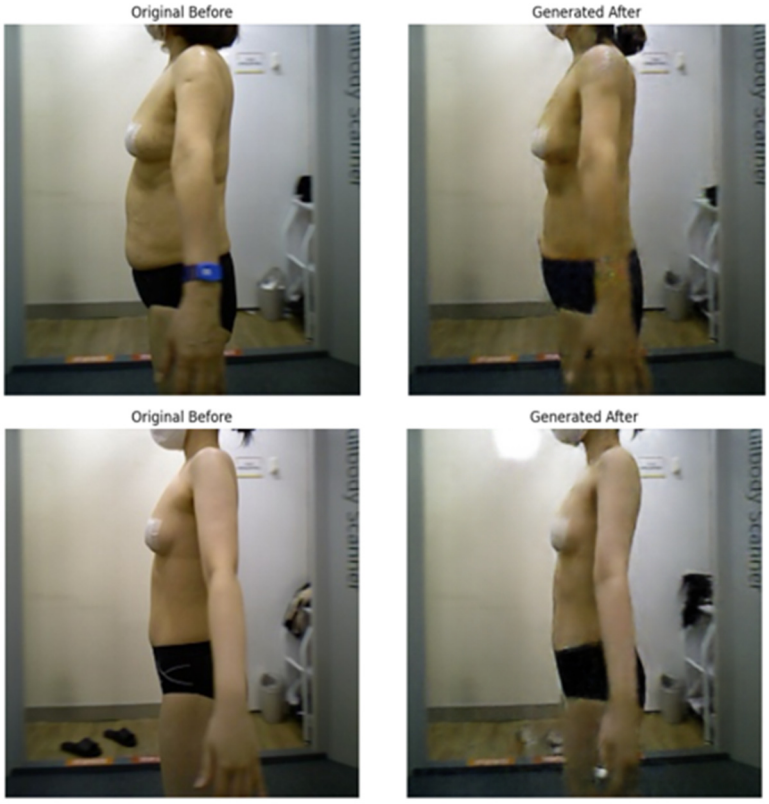


Figure 10. Examples of post-liposuction prediction results generated by the Pix2Pix model trained with L1 and perceptual loss. This figure presents qualitative examples comparing preoperative images (**left column**) and corresponding predicted postoperative images (**right column**), generated using the Pix2Pix model trained with a combination of L1 and perceptual loss. The original images were captured before liposuction, and the predicted images aim to simulate postoperative body shapes based solely on the input pre-liposuction image. The output generated reflects changes primarily in the abdominal and waist regions. Minor blurring can be observed, particularly around the arms, because of variability in arm positioning during data acquisition. These examples illustrate the model's ability to generate visually plausible postoperative predictions under realistic conditions.

Another failure case was found in patients with relatively low BMI. In such cases, the physical change after liposuction is often minimal and difficult to detect visually. As a result, the model occasionally failed to generate images with noticeable differences, producing predictions nearly identical to the preoperative image. This limitation is illustrated in Figure 12.

We compared the results of the models based on L1 loss and perceptual loss. The model with the smallest L1 loss was evaluated on the test dataset. As shown in Figure 9, when L1 loss was used, the images tended to be blurry. This occurred because edge details and other information were not considered, as the model focused solely on minimizing pixel-wise errors. The images generated by the model based on perceptual loss were less blurry. However, in the case of the first subject, when only perceptual loss is used, the body shape exhibited minimal change. By comparison, when L1 loss and perceptual loss were used together, the generated images were even less blurry, and body shape changes were accurately reflected. As shown in Figure 13, the comparison of predicted postoperative images generated by different loss functions demonstrates the differences clearly. The

combination of L1 and perceptual loss results in more realistic and sharper predictions, better reflecting the structural and contour changes after liposuction.



Figure 11. Prediction failure due to inconsistent arm position during data acquisition. This figure illustrates a failure case in which the predicted post-liposuction image (middle) fails to reflect morphological changes accurately, primarily because of inconsistent arm positioning between pre- (left) and postoperative (right) images. Although camera height and distance were fixed, the lack of explicit posture guidelines caused large variations in arm placement across individuals. This led to an ambiguous boundary between the arms and abdomen, introducing blurring artifacts and reducing the model's ability to learn precise contours in the abdominal region.



Figure 12. Prediction failure case in a low-BMI patient with minimal postoperative change. From left to right: the first image shows the ground truth (GT) before liposuction, the second is the predicted image generated by the model, and the third is the ground truth (GT) after liposuction. In this case, the patient's BMI changed only slightly (from 18.80 to 18.07), resulting in a minimal visual difference between the pre- and post-liposuction appearances. The predicted image closely resembles the preoperative image, failing to reflect the subtle morphological change. This example highlights the model's current limitation in capturing slight body contour differences, particularly in patients with low BMI. Future improvements should focus on enhancing the model's sensitivity to subtle variations and preserving fine-grained details to ensure accurate postoperative predictions in such cases. From left to right: the first image shows the ground truth (GT) before liposuction, the second is the predicted result after liposuction, and the third is the ground truth (GT) after liposuction. In this example, the model failed to capture the subtle morphological differences caused by minimal fat reduction. As a result, the predicted image closely resembles the preoperative state rather than the actual postoperative outcome. This highlights the model's limitation in handling slight body contour changes, especially common in patients with low BMI. Enhancing the model's sensitivity to such subtle variations remains an important area for future improvement.

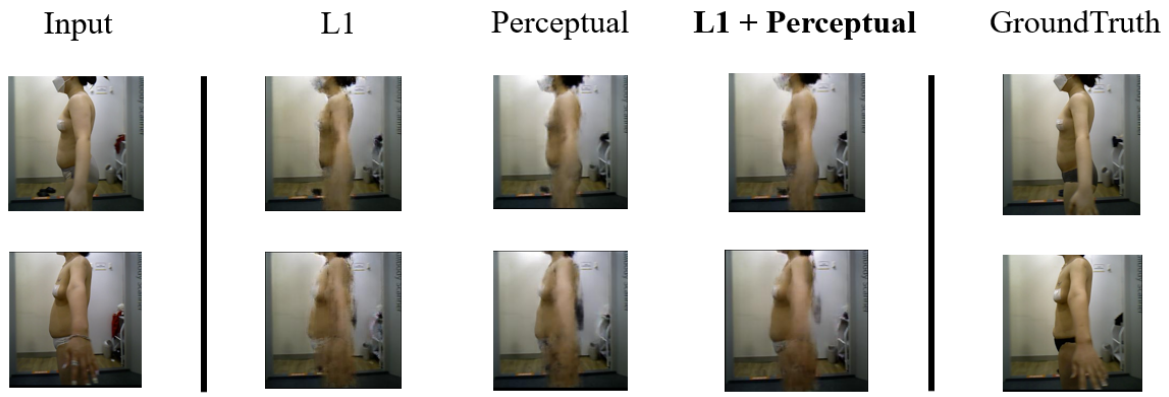


Figure 13. Comparison of predicted postoperative images generated by different loss functions: L1, Perceptual, and a combination of both (L1 + Perceptual). From left to right, the columns show the preoperative input image, images generated using only L1 loss, only perceptual loss, both L1 and perceptual losses combined, and the ground truth postoperative image. When using only L1 loss, the generated images appear overly smooth and blurry due to the averaging effect of pixel-wise loss. Perceptual loss reduces blurriness but often fails to capture accurate morphological changes. The combined L1 + perceptual loss results in more realistic and sharper predictions, better reflecting the structural and contour changes after liposuction. These findings support the use of hybrid loss functions for improving visual realism and structural fidelity in medical image translation.

(Quantitative Evaluation) We conducted four types of quantitative evaluation. First, we evaluated the accuracy of the generated images using the BMI predictor, as described earlier. This was measured by the mean squared error (MSE) between the predicted BMI and the ground truth BMI. A lower BMI MSE loss indicates higher prediction accuracy and consistency between the generated image and the actual outcome. Second, the similarity between two given images was determined based on the Learned Perceptual Image Patch Similarity(LPIPS) metric, which can mimic human evaluation. Specifically, the LPIPS metric calculates the similarity between the features of two images using AlexNet [44] pretrained with ImageNet [33]; when linear weights are incorporated, the evaluation resembles subjective judgment based on human cognition [43]. Since LPIPS measures dissimilarity, a lower LPIPS value indicates greater perceptual similarity between the generated and ground truth images and is therefore considered better. Third, we adopted the Structural Similarity Index Measure (SSIM) [45], a widely used metric in medical image translation, to evaluate the structural fidelity of the generated images. SSIM measures the similarity in structural information, luminance, and contrast between two images. A higher SSIM value (closer to 1.0) indicates better preservation of spatial structure and local textures. Fourth, we used the Peak Signal-to-Noise Ratio (PSNR) [46] to evaluate image clarity. PSNR measures the ratio between the maximum possible power of a signal and the power of corrupting noise. Higher PSNR values (typically above 30 dB) represent clearer and more accurate reconstructions with less distortion.

Table 3 lists the results of the quantitative evaluation. Among the baseline methods, the L1 + Augmentation model achieved the lowest BMI MSE loss (0.0050), indicating high consistency between predicted and actual BMI values. However, this model yielded relatively high LPIPS (0.2861) and lower SSIM and PSNR scores, suggesting that the generated images lacked visual sharpness and structural clarity—likely due to the tendency of L1 Loss to produce smooth but blurry outputs.

Table 3. Quantitative comparison of different Pix2Pix-based loss function configurations for predicting post-liposuction images. The table reports BMI MSE loss, LPIPS, SSIM, and PSNR values for models using L1 Loss, perceptual loss, and a combined loss, with and without data augmentation using CcGAN. Our proposed model (L1 + Perceptual + Augmentation) achieved the lowest BMI MSE loss (0.0057), lowest LPIPS (0.2709), highest SSIM (0.7814), and highest PSNR (40.34), demonstrating superior structural accuracy, perceptual similarity, and image clarity. These results confirm that combining L1 and perceptual loss with CcGAN-based augmentation enhances both quantitative and visual prediction performance.

Method	BMI MSELoss	LPIPS	SSIM	PSNR
L1 + Augmentation	0.0050	0.2861	0.7162	38.95
Perceptual + Augmentation	0.0118	0.2820	0.7095	37.10
L1 + Perceptual	0.0060	0.3145	0.7053	38.87
L1 + Perceptual + Augmentation (ours)	0.0057	0.2709	0.7814	40.34

In contrast, Perceptual + Augmentation achieved the lowest LPIPS (0.2820) among the baselines, reflecting improved perceptual similarity to ground truth images. However, this method showed the highest BMI MSE loss (0.0118), indicating reduced consistency in predicted BMI. Although it performed best among baseline methods in terms of LPIPS, its performance was still inferior to our proposed model, which achieved an even lower LPIPS of 0.2709 while also improving structural fidelity and BMI consistency. L1 + Perceptual, trained without augmentation, showed moderate performance but the worst LPIPS (0.3145) and relatively lower PSNR.

Our proposed model, L1 + Perceptual + Augmentation, achieved a strong balance across all metrics. It yielded a low BMI MSE loss (0.0057), the lowest LPIPS (0.2709) overall, and the highest SSIM (0.7814) and PSNR (40.34), indicating that it not only maintained BMI consistency but also improved perceptual realism and image quality. This result demonstrates that combining L1 and perceptual loss with augmentation effectively integrates the advantages of both loss types while improving generalization.

5. Discussion and Conclusions

With the continued development of deep learning, I2I has progressively gained importance in the medical field, with several excellent algorithms being developed. While some methods utilize medical data, most techniques involve using CT or X-ray data, which limits their accessibility to the general public [15,16]. This study introduced a model to predict post-liposuction appearance using RGB images, which can be easily captured by potential patients themselves. The proposed model can predict how the body shape will change after liposuction based on the subject's body measurements. This enables intuitive predictions, which would not be possible when predicting the results for one individual based on those of other people. To the best of our knowledge, this study represents the first attempt at predicting post-liposuction appearance using body shape data. Despite being designed around liposuction, the proposed technique is expected to be useful for predicting changes caused by various other types of surgery that alter appearance and body shape. However, as few studies have sought to exploit such possibilities thus far, active research should be conducted on postoperative prediction using RGB images.

Our method combines data augmentation using a Continuous Conditional GAN (CcGAN) with a Pix2Pix-based image-to-image translation model. The effectiveness of this approach was verified through quantitative evaluation using four metrics: BMI MSE loss, LPIPS, SSIM, and PSNR. Our proposed model, which integrates both L1 and perceptual loss functions with CcGAN-based augmentation, achieved the lowest BMI MSE loss (0.0057),

lowest LPIPS (0.2709), highest SSIM (0.7814), and highest PSNR (40.34) among all compared methods (Table 3). These results indicate that the proposed model generates images that are not only quantitatively consistent with BMI predictions but also visually realistic and structurally faithful.

Notwithstanding its benefits, the proposed method has certain limitations as well. For instance, the BMI predictor used to evaluate the CcGAN was based on ResNet34, and the model with the lowest error was used for data augmentation. Thus, the selection of the optimal model was entirely dependent on the BMI predictor's evaluation. As described in Section 3, blurring was frequently observed in the abdominal region, which was primarily attributed to the inconsistency in arm positioning during image acquisition. Although the images were captured from a fixed distance and camera height, no clear instructions were provided regarding the angle, position, or spacing of the arms. As a result, each subject adopted different poses—some with arms naturally hanging close to the body and others with arms slightly spread apart. This inconsistency in arm placement led to ambiguity in the boundary between the abdomen and the arms, especially causing loss of contour clarity or distortion around the abdominal area. Such variations disrupted the model's ability to learn the structural features of the abdomen consistently and ultimately resulted in blurring or smearing artifacts in the predicted images. To address this issue, future data collection protocols should include explicit pose guidelines, including the angle of the arms, the width between the shoulders, and the distance between the arms and torso. Providing standardized posture instructions to participants prior to image capture will help ensure greater consistency across samples, thereby enabling the model to more accurately learn and reproduce key body regions, particularly the abdominal area.

This study also faced a limitation from the use of only a single-view input—the left lateral image of the upper body. While this viewpoint effectively captures changes in the abdominal and flank areas, it is insufficient for ensuring anatomical consistency from other angles, such as frontal or dorsal views. This limitation may restrict the model's ability to learn comprehensive morphological changes. Future studies are expected to incorporate multi-view image inputs, including front, side, and back views, to enable more coherent and spatially consistent predictions of body contour changes. Furthermore, conducting quantitative validations for each viewpoint will help improve anatomical fidelity and enhance the generalizability of the model.

A further limitation was observed in patients with relatively low BMI; in such cases, the physical changes before and after liposuction tend to be subtle and less visually noticeable. As a result, the model sometimes failed to capture these minimal differences, generating output images with little to no perceived change. This suggests that the model's performance may be less effective in cases where the visual impact of liposuction is inherently limited, which poses a challenge for prediction based on subtle morphological cues.

Another critical factor that was not explicitly addressed in this study is the potential variation in fat distribution patterns between genders. Numerous clinical and physiological studies have shown that males and females tend to accumulate and reduce fat differently—males often in the abdominal area and females more evenly distributed across the hips, thighs, and waist. Since our model was trained using images of patients without explicitly separating by gender, these biological differences may influence both the training dynamics and prediction accuracy. Particularly, morphological changes that are subtle or anatomically different because of gender-specific fat localization may not be fully captured by a single unified model. Therefore, it is important to acknowledge that the applicability of the proposed model may vary depending on gender, and future research should consider stratifying training data or building gender-specific models to enhance accuracy and

interpretability. Incorporating such considerations will also help define the boundaries of the model's generalization and clarify its clinical utility.

In addition, the dataset used in this study does not include patients with clinically extreme BMI values (i.e., BMI $\geq$ 40 or BMI < 16); the model's generalizability to such cases could not be evaluated. Therefore, future research should involve the collection of data from patients with pathological obesity or severely underweight to more rigorously assess the clinical applicability and robustness of the proposed model.

Moreover, the training dataset used in this study was composed entirely of Korean individuals. While this allows the model to learn the body characteristics and liposuction-related changes typical of this population, it also raises concerns about the model's ability to generalize to individuals from other ethnicities and body types. Factors such as body shape distribution, fat localization, and skin tone can vary significantly across different populations, potentially reducing prediction accuracy when the model is applied outside the original data domain.

To improve the model's generalizability and applicability, future studies should aim to collect more diverse datasets that include a broader range of ethnic groups, body shapes, and demographic characteristics. In addition, applying domain adaptation techniques or style transfer-based approaches may help bridge the distribution gap between datasets and enhance model robustness across varying populations.

Furthermore, BMI—a common index for classifying obesity [27]—was used as the conditional variable for data augmentation. However, while approximate body shapes can be predicted based on BMI, its accuracy may not be sufficient because it depends solely on height and weight [26,47,48]. Therefore, more accurate data could be generated if Dual-energy X-ray Absorptiometry (DXA) [49], the gold standard for body fat mass evaluation, is used as the conditional variable instead of BMI. Finally, since I2I based on body shape has not been performed previously, we had no reference to assess whether the error of our model is within an allowable range. Therefore, we had to rely on qualitative evaluations mimicking human judgment instead.

The Pix2Pix-based I2I model proposed herein enables patients to predict their appearance after liposuction based on their own body shape data. Building on this approach, we developed a deep learning-based prediction model trained on a paired dataset of pre- and post-liposuction RGB images. This model enables personalized appearance prediction from preoperative images, making the process more intuitive and accessible. To combat the lack of available data, we performed data augmentation using CcGAN, which employs continuous numerical values as conditional variables. The model's predictive performance was improved by generating body shape data based on BMI values as inputs. Further, by leveraging Pix2Pix's reconstruction loss along with L1 loss and perceptual loss, our model outputs less blurry images as it minimizes both pixel-wise errors and deviations in the overall image structure. Because our model uses only RGB images rather than professional medical data, it can be utilized by the public to predict their post-surgery appearance without having to visit a hospital or any other specialized medical facility. Therefore, the proposed method can assist people in making informed decisions about whether to undergo liposuction.

In future work, we plan to explore the applicability of advanced generative models such as StyleGAN3, diffusion-based models, and transformer-based conditional generators, which may better accommodate continuous condition variables such as BMI. Models such as DCAN, cGAN, PConv-GAN, StyleGAN, and StyleGAN2 were not adopted in this study because of specific limitations—such as segmentation-focused design, support for only categorical conditions, specialization in inpainting, reliance on latent space optimization, or the need for large-scale datasets—we recognize the potential value of these architec-

tures in related domains. In future research, we aim to revisit these models by applying dataset expansion, customized training strategies, and architectural modifications to assess whether they can be effectively adapted to our paired RGB image setting. This may lead to improvements in image realism, training stability, and generalization across diverse populations. Ultimately, by integrating more advanced generative frameworks into clinical applications, we hope to enhance the robustness, scalability, and personalized utility of postoperative prediction systems in medical practice.

Author Contributions: M.K. was responsible for methodology, analysis, and writing. J.B. was responsible for data processing and analysis. J.C. was responsible for methodology. S.Y. was responsible for methodology, review, and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Research Foundation of Korea (grant number NRF-2023R1A2C2004252).

Institutional Review Board Statement: The data collection process was reviewed and approved by the Institutional Review Board (IRB) of Dongguk University under the approval number DUIRB-202202-14 (approved 7 March 2022). All methods were conducted in accordance with the relevant ethical guidelines and regulations.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Due to ethical considerations and privacy regulations, the dataset used in this study cannot be publicly shared. The data contains sensitive patient information, including pre- and post-operative images, which are protected under the Institutional Review Board (IRB) guidelines at Dongguk University (DUIRB-202202-14). Access to the dataset may be granted upon reasonable request, subject to ethical approval and institutional data-sharing policies.

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

References

1. Kim, Y.J. Effects of Liposuction on Body Image, Self-Esteem, and Satisfaction with Outcomes. Master's Thesis, Department of Occupational Health, Yonsei University, Seoul, Republic of Korea, 2006.
2. Parsa, S.; Basagaoglu, B.; Mackley, K.; Aitson, P.; Kenkel, J.; Amirlak, B. Current and future photography techniques in aesthetic surgery. *Aesthetic Surg. J. Open Forum* **2022**, *4*, ojab050. [CrossRef] [PubMed]
3. Shah, S.A.A.; Bennamoun, M.; Molton, M.K. Machine learning approaches for prediction of facial rejuvenation using real and synthetic data. *IEEE Access* **2019**, *7*, 23779–23787. [CrossRef]
4. 365mc.co.kr. 365MC. 2022. Available online: http://www.365mc.co.kr/diettest/pc/?src=image&kw=000022&route=WCDTEST_H (accessed on 30 March 2022).
5. Babajide, B.O.; Hissam, T.; Anna, A.P.; Anatoliy, A.G.; Astrup, A.J.; Martinez, J.A. A machine learning approach to short-term body weight prediction in a dietary intervention program. In *Computational Science–ICCS 2020: Proceedings of the 20th International Conference, Amsterdam, The Netherlands, 3–5 June 2020*; Proceedings, Part IV; Springer: Cham, Switzerland, 2020.
6. Weber, I.; Achananuparp, P. Insights from machine-learned diet success prediction. In *Biocomputing 2016: Proceedings of the Pacific Symposium, Kohala Coast, HI, USA, 4–8 January 2016*; World Scientific: Singapore, 2016.
7. Kuzmar, I.; Rizo, M.; Cortés-Castell, E. Adherence to an overweight and obesity treatment: How to motivate a patient? *PeerJ* **2014**, *2*, e495. [CrossRef]
8. Zeng, Y.; Fu, J.; Chao, H. 3D human body reshaping with anthropometric modeling. In *Internet Multimedia Computing and Service, Proceedings of the 9th International Conference, ICIMCS 2017, Qingdao, China, 23–25 August 2017*; Revised Selected Papers; Springer: Singapore, 2018.
9. 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, Long Beach, CA, USA, 15–20 June 2019.
10. 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, Seattle, WA, USA, 13–19 June 2020.
11. Dong, G.; Basu, A. Unsupervised Face Frontalization GAN Driven by 3D Rotation and Symmetric Filling. In *International Conference on Smart Multimedia*; Springer International Publishing: Cham, Switzerland, 2022; pp. 59–74.

12. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 1–9.
13. Yoo, T.K.; Choi, J.Y.; Kim, H.K. A generative adversarial network approach to predicting postoperative appearance after orbital decompression surgery for thyroid eye disease. *Comput. Biol. Med.* **2020**, *118*, 103628. [CrossRef] [PubMed]
14. Kaji, S.; Kida, S. Overview of image-to-image translation by use of deep neural networks: Denoising, super-resolution, modality conversion, and reconstruction in medical imaging. *Radiol. Phys. Technol.* **2019**, *12*, 235–248. [CrossRef]
15. Wang, Q.; Xue, W.; Zhang, X.; Jin, F.; Hahn, J. Pixel-wise body composition prediction with a multi-task conditional generative adversarial network. *J. Biomed. Inform.* **2021**, *120*, 103866. [CrossRef]
16. Eslami, M.; Tabarestani, S.; Albarqouni, S.; Adeli, E.; Navab, N.; Adjouadi, M. Image-to-image translation for multi-task organ segmentation and bone suppression in chest X-ray radiography. *IEEE Trans. Med. Imaging* **2020**, *39*, 2553–2565. [CrossRef]
17. Ryu, J.Y.; Chung, H.Y.; Choi, K.Y. Potential role of artificial intelligence in craniofacial surgery. *Arch. Craniofacial Surg.* **2021**, *22*, 223. [CrossRef]
18. 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.
19. Yang, H.; Lu, X.; Wang, S.H.; Lu, Z.; Yao, J.; Jiang, Y.; Qian, P. Synthesizing multi-contrast MR images via novel 3D conditional variational auto-encoding GAN. *Mob. Netw. Appl.* **2021**, *26*, 415–424. [CrossRef]
20. Armanious, K.; Kumar, V.; Abdulatif, S.; Hepp, T.; Gatidis, S.; Yang, B. ipA-MedGAN: Inpainting of arbitrary regions in medical imaging. In Proceedings of the 2020 IEEE International Conference on Image Processing (ICIP), Abu Dhabi, United Arab Emirates, 25–28 October 2020; pp. 3005–3009.
21. Choi, Y.; Choi, M.; Kim, M.; Ha, J.-W.; Kim, S.; Choo, J. StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.
22. Wang, J.; Perez, L. The effectiveness of data augmentation in image classification using deep learning. *Convolutional Neural Netw. Vis. Recognit.* **2017**, *11*, 1–8.
23. Shi, H.; Wang, L.; Ding, G.; Yang, F.; Li, X. Data augmentation with improved generative adversarial networks. In Proceedings of the 2018 24th International Conference on Pattern Recognition (ICPR), Beijing, China, 20–24 August 2018; pp. 73–78.
24. Frederick, D.A.; Lever, J.; Peplau, L.A. Interest in cosmetic surgery and body image: Views of men and women across the lifespan. *Plast. Reconstr. Surg.* **2007**, *120*, 1407–1415. [CrossRef] [PubMed]
25. Fulton, J.E.; Rahimi, A.D.; Helton, P. Modified tumescent liposuction. *Dermatol. Surg.* **1999**, *25*, 755–766. [CrossRef] [PubMed]
26. Nuttall, F.Q. Body mass index: Obesity, BMI, and health: A critical review. *Nutr. Today* **2015**, *50*, 117. [CrossRef]
27. World Health Organization. *Physical Status: The Use and Interpretation of Anthropometry. Report of a WHO Expert Committee*; World Health Organization: Geneva, Switzerland, 1995.
28. Altman, D.G. *Practical Statistics for Medical Research*; CRC Press: Boca Raton, FL, USA, 1990.
29. Ratner, B. The correlation coefficient: Its values range between +1/−1, or do they? *J. Target. Meas. Anal. Mark* **2009**, *17*, 139–142. [CrossRef]
30. Ding, X.; Wang, Y.; Xu, Z.; Welch, W.J.; Wang, Z.J. CcGAN: Continuous conditional generative adversarial networks for image generation. In Proceedings of the International Conference on Learning Representations, Addis Ababa, Ethiopia, 26–30 April 2020.
31. Chen, X.; Duan, Y.; Houthooft, R.; Schulman, J.; Sutskever, I.; Abbeel, P. InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2016**, *29*, 1–9.
32. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
33. Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; Fei-Fei, L. ImageNet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
34. Kodirov, E.; Xiang, T.; Gong, S. Semantic autoencoder for zero-shot learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 3174–3183.
35. Zhu, J.-Y.; Krähenbühl, P.; Shechtman, E.; Efros, A.A. Generative visual manipulation on the natural image manifold. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2016; pp. 597–613.
36. Laffont, P.-Y.; Ren, Z.; Tao, X.; Qian, C.; Hays, J. Transient attributes for high-level understanding and editing of outdoor scenes. *ACM Trans. Graph. (TOG)* **2014**, *33*, 1–11. [CrossRef]
37. Mirza, M.; Osindero, S. Conditional generative adversarial nets. *arXiv* **2014**, arXiv:1411.1784.
38. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Cham, Switzerland, 2015; pp. 234–241.
39. Latif, U.; Shahid, A.R.; Raza, B.; Ziauddin, S.; Khan, M.A. An end-to-end brain tumor segmentation system using multi-inception-UNET. *Int. J. Imaging Syst. Technol.* **2021**, *31*, 1803–1816. [CrossRef]

40. Amer, A.; Ye, X.; Zolgharni, M.; Janan, F. ResDUnet: Residual dilated UNet for left ventricle segmentation from echocardiographic images. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC), Montreal, QC, Canada, 20–24 July 2020; pp. 2019–2022.
41. Johnson, J.; Alahi, A.; Fei-Fei, L. Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2016; pp. 694–711.
42. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
43. Zhang, R.; Isola, P.; Efros, A.A.; Shechtman, E.; Wang, O. The unreasonable effectiveness of deep features as a perceptual metric. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 586–595.
44. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1–9. [CrossRef]
45. 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]
46. Huynh-Thu, Q.; Ghanbari, M. Scope of validity of PSNR in image/video quality assessment. *Electron. Lett.* **2008**, *44*, 800–801. [CrossRef]
47. Keys, A.; Fidanza, F.; Karvonen, M.J.; Kimura, N.; Taylor, H.L. Indices of relative weight and obesity. *J. Chronic Dis.* **1972**, *25*, 329–343. [CrossRef] [PubMed]
48. Wellens, R.I.; Roche, A.F.; Khamis, H.J.; Jackson, A.S.; Pollock, M.L.; Siervogel, R.M. Relationships between the body mass index and body composition. *Obes. Res.* **1996**, *4*, 35–44. [CrossRef]
49. Ball, S.D.; Altena, T.S. Comparison of the Bod Pod and dual-energy X-ray absorptiometry in men. *Physiol. Meas.* **2004**, *25*, 671. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.