Article

Virtual Hairstyle Service Using GANs & Segmentation Mask (Hairstyle Transfer System)

Mohamed S. Abdallah [1,2,*] and **Young-Im Cho** [1,*]

1   Department of Computer Engineering, Gachon University, Seongnam 1342, Korea
2   Informatics Department, Electronics Research Institute (ERI), Cairo 11843, Egypt
*   Correspondence: sameer@gachon.ac.kr or sameer@eri.sci.eg (M.S.A.); yicho@gachon.ac.kr (Y.-I.C.)

Abstract: The virtual hair styling service, which now is necessary for cosmetics companies and beauty centers, requires significant improvement efforts. In the existing technologies, the result is unnatural as the hairstyle image is serviced in the form of a 'composite' on the face image, image, extracts and synthesizing simple hair images. Because of complicated interactions in illumination, geometrical, and occlusions, that generate pairing among distinct areas of an image, blending features from numerous photos is extremely difficult. To compensate for the shortcomings of the current state of the art, based on GAN-Style, we address and propose an approach to image blending, specifically for the issue of visual hairstyling to increase accuracy and reproducibility, increase user convenience, increase accessibility, and minimize unnaturalness. Based on the extracted real customer image, we provide a virtual hairstyling service (Live Try-On service) that presents a new approach for image blending with maintaining details and mixing spatial features, as well as a new embedding approach-based GAN that can gradually adjust images to fit a segmentation mask, thereby proposing optimal styling and differentiated beauty tech service to users. The visual features from many images, including precise details, can be extracted using our system representation, which also enables image blending and the creation of consistent images. The Flickr-Faces-HQ Dataset (FFHQ) and the CelebA-HQ datasets, which are highly diversified, high quality datasets of human faces images, are both used by our system. In terms of the image evaluation metrics FID, PSNR, and SSIM, our system significantly outperforms the existing state of the art.

Keywords: hairstyle; StyleGAN; blending features; generative adversarial networks (GANs); segmentation mask

1. Introduction

Generative adversarial networks (GANs) [1–4] have been shown to be more popular and able of modeling finer details. The advancements of GANs has led to a recent increase in the application of GAN-based image manipulation. The majority of face editing images created by GANs alter an image's high level attributes, including age, gender, expression, position and pose [5,6]. Even while the latest progress in GAN technology have made it possible to synthesis realistic hairstyles and faces, it is still difficult to combine these elements into a unified, convincing image as opposed to a disjointed collection of image regions.

Blending issues with the face and hair in generated face images are very challenging. The visual features of the face parts are not independent of each other. For example, the hair features are affected by surrounding and reflected light, camera parameters, colors of clothes, and background, the pose and shape of the head, and the style of hair. Consequently, the pose may need to modify to accommodate the hairstyle.

Without the general consistency of the blended image, the combined parts would look disjointed, even though each component is generated with an extremely realistic level of realism. A crucial idea we put forth is a semantic-segmentation alignment based GAN that

creates a coherent image with overall realism while balancing the fitting of each part to the associated input images. The key advantage of semantic-segmentation alignment is the presence of semantic accurate image pixels in the hair-relevant areas of the image.

Refs. [7,8] coined the terms appearance, structure, shape, and identity to describe various characteristics of hair. The head and face image comprises all the features required to recognize someone. Although the two methods [7,8] show promising results, they both rely on pretrained inpainting model to repair gaps left over by mismatched hair mask, which could result in blurred artefacts and artificial borders. These previous methods do not use a semantic-segmentation alignment to combine hair and facial areas from various input images in latent space. Therefore, we discovered that they could be much better.

In this work, we propose a virtual hairstyling service (Live Try-On service) for image-realistic hairstyles by image blending elements from various images in order to create a new generated image with different aspects of the hairstyles. We approach face image editing by selecting features from multiple images (e.g., hair, face, freckles) and combining them to form a composite image. Our method can combine these four characteristics (appearance, structure, shape, and identity) to create a variety of realistic hairstyles.

StyleGAN2 [9] is used in our method to produce high reconstructions of reference images. Our method proposes a basic control of both feature spatial positions and overall appearance style attributes. We use a semantic alignment with existing GAN-embedding methods to align and embed reference images to feature locations by significantly changing reference images to comply with a new segmentation mask. Then, to eliminate the defects of existing picture compositing methods, we blend reference images in a new spatial dimension.

The following are our major contributions:

- A new space for blending images that is more capable of encoding feature spatial positions and maintaining feature details.
- A new algorithm for aligned embedding reference images that uses a semantic segmentation mask and alignment with existing GAN-embedding methods. The proposed algorithm can align and embed images with a high-quality result.
- A new method for image blending that can blend multiple images with a considerable enhancement in virtual hairstyle transfer.

The remaining parts of the paper are structured as follows: Related studies are included in Section 2. The suggested system is explained in Section 3. The experimental findings and datasets used in this study are presented and discussed in Section 4. The paper's conclusion and a plan for further study are provided in Section 5.

2. Related Research

To synthesize high quality images in an unconditional setting, various generative models have been proposed. GANs (generative adversarial networks) [1–4] have greatly assisted a massive increase in the research of high-resolution images generation. Innovative GAN networks demonstrate notable enhancements in sample variety and image quality.

Starting with a random latent code, recent GANs such as [5,10–12] can generate highly finer details images that are nearly as natural in the face domain. StyleGAN and StyleGAN2 [9,10] significantly improved image generation quality by merging the latent and noisy details into either the deep and shallow layers. As demonstrated by Karras et al. [11], a GAN may be trained using limited datasets without sacrificing its ability to generate new images.

The GAN research groups has recently become focused in GAN image manipulation and GAN understandability. Many recently published works [13–15] utilize GAN architectures and their pre-trained models to achieve a high-quality result in the image manipulation domain by editing the embedded characteristics in these networks' latent space.

The availability of datasets like FFHQ [9], CelebA-HQ [12], AFHQ [16], and LSUN [17] gives high-resolution images and variability for training GANs, as well as leading to the

growth of realistic applications and high-quality image generation. Brock et al. [18] developed large scale GAN by using ImageNet [19] dataset to generate high-quality samples.

Image semantic manipulation is conceivable using latent space or activation space. The researchers attempt to comprehend the essence of the GAN's latent space in order to derive meaningful edit directives in the domain of latent space manipulation. Principal Component Analysis (PCA) can be used as an unsupervised manner to identify the linearity properties of StyleGAN latent space [20]. Theobalt et al. [21] discover a relationship between the latent space in StyleGAN and a manipulated face model. Wonka et al. [22] investigate the nonlinearity properties of StyleGAN latent space by producing various sequential edits using normalizing flows. Lischinski et al. [23] manipulates the latent space with another strategy using text information.

On the other hand, in order to achieve the appropriate fine-grained manipulating of an image, refs. [5,24,25] edit the activation maps. Ref. [26] looks into the channel's style attributes to generate fine-grained adjustments. Ref. [27] creates spatial maps using latent codes that can be used for image manipulation.

Several hair editing methods for hair transferring are proposed using the rough hair geometry estimated from one view image [28–30]. However, the lack of interpretation and manipulation approaches to several visualizing factors severely limits the precision of the results and editing adaptability. Latest GAN-based image generated methods with high realistic image synthesis result have shown great promise in closing this gap in quality [9].

The generation of conditional images is considered to be significantly more challenging than the generation of random codes. Although recent research in conditioned hair generation [31–33] have made significant advancements by specific varieties of inputs, such approaches are still not naturally adjustable or broadly applicable.

To be manipulated, a given image must be embedded in the latent spaces of the GAN network. There are two methods for embedding images in latent spaces of GAN: (1) methods based on optimization and (2) methods based on encoders. StyleGAN [9] and Image2stylegan [13] achieved excellent embeddings into the extended space, known as latent space W+, these optimization-based approaches result in commercial software. Refs. [34,35] demonstrated that by incorporating new regularizers for optimization, embeddings can be improved. Image2StyleGAN (II2S) has shown that regularizing norm space P can lead to improved editing quality and embeddings [36]. Encoding in style [37] achieved excellent embeddings that are able to manipulate by training an encoder on the latent spaces.

Several novel studies use conditional-GAN for image generation. Making the image generation dependent on another input image is one strategy for incorporating several images for manipulation. Given the conditions, methods such as [32,38,39] are capable of producing controllable person image synthesis high-quality images. Refs. [16,40] can tweak multiple attributes, especially on face images.

We outlined two significant relating works for editing the hairstyle, hair transfer, and visual appearance based on GANs. The first work [8] uses optimization of latent spaces and orthogonal gradient to extricate hair features consciously. Ref. [8] achieved excellent hair editing by drawing new consistent hair with appearance, shape & structure, and background condition modules. The second work [7] uses the StyleGAN2 generator to break down hair into hair appearance, hair structure, and hairstyle characteristics, and then optimizes latent space to fill in absent hair features.

While both of these approaches yielded convincing preliminary achievements, we discovered that both could be significantly enhanced. For instance, the two methods require an inpainting network that fills in absent hair gaps left by an unaligned hair mask, which can result in abnormal borders and blur artifacts. Without utilizing inpainting networks, enhanced outcomes can be obtained, because turnarounds between regions are all of superior quality when synthesized by a GAN network. Additionally, using a semantic alignment to merge hair and face areas from multiple reference images in latent space can improve results.

Based on StyleGAN [9] and StyleGAN2 [10] trained on the FFHQ dataset, our research intends at the conditional face image generation for hairstyle, which is significantly complicated because of the numerous precise circumstances for dominating required hair formulation. Our system is able to generate images that are coherent with manipulations performed in semantic segmentation masks extracted from reference images.

3. Proposed System

We propose a virtual hairstyling service (Live Try-On service) for image-realistic hairstyles that use image blending to generate an entirely newly generated image with several aspects of the hairstyles. Our Face image blending approach is accomplished by selecting features from multiple images (e.g., hair appearance, hair shape, hair structure, face, and background) and combining them to create a realistic hairstyle composite image.

Figure 1 depicts the framework of our proposed model. This framework is divided into two parts: Face Segmentation Mask and Generative Hairstyle module.
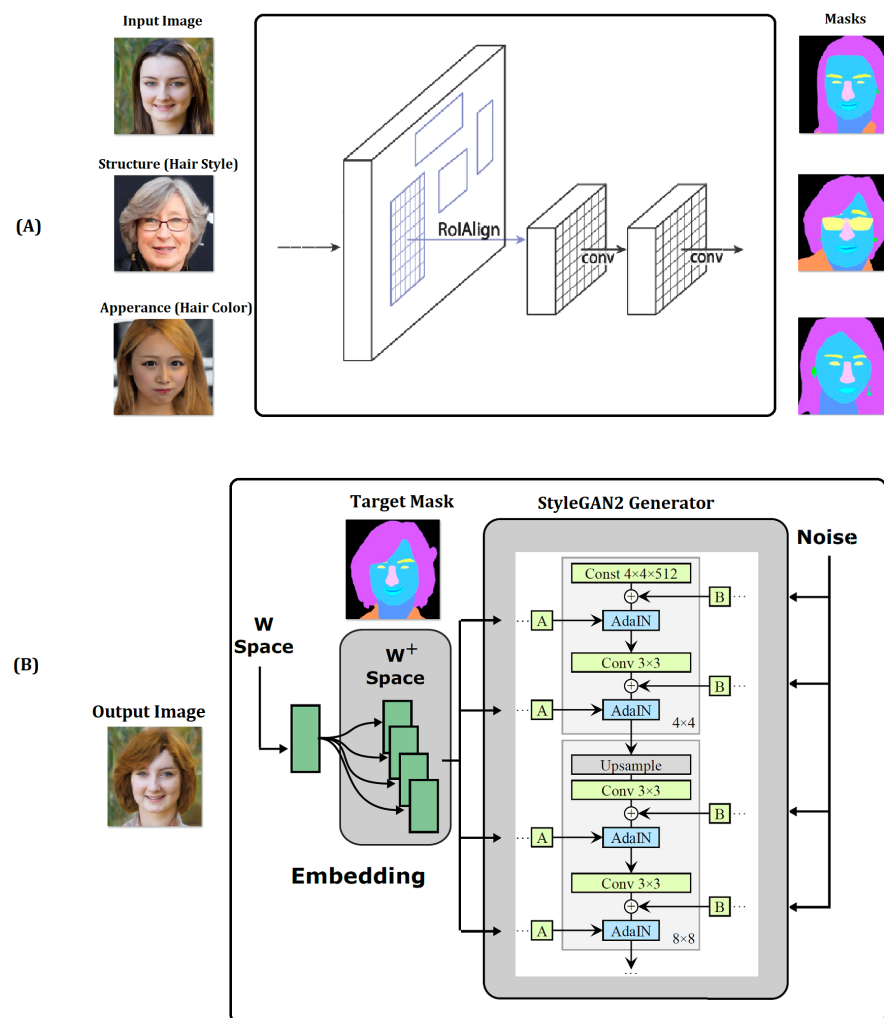

Figure 1. Architecture of our system (**A**) Face Segmentation Mask (**B**) Generative Hairstyle module.

First, in order to improve feature hair extraction and give the generation model better precise feature information, we employ a target face semantic segmentation mask image prior to generating the hairstyle. The task of face semantic segmentation is in charge of identify facial attributes and extract feature map for the face blob. Our system can use one image to transfer the hairstyle and another image to utilize for these other semantic features. Each of these reference images is fitted (aligned) to the segmentation mask before being blending to produce a new hairstyle.

As shown in Figure 1A, our face segmentation mask module extracts masks and facial characteristics relying on Mask RCNN [41]. Faster RCNN is enhanced by Mask RCNN, which includes a branch to estimating segmentation masks on every Region of Interest (RoI) [42,43]. To segment and extract facial features from the adjacent pixel spaces, we employ the region of interest (RoI) pooling layer of the RoI alignment. The quantization operation discrepancies between the acquired features and the RoI are bigger in the traditional RoI pooling layer, and this discrepancy can negatively affect the anticipated pixel mask.

The RoI alignment, on the other hand, keeps the original feature data without quantification, and the RoI features that are produced are more genuine. The RoI alignment contributes to the consistency and authenticity of the generated images by facilitating semantic identification, comprehensive mask segmentation of facial and hair areas, and image generation.

The following equation is used to express the loss of Mask R-CNN:

$$L = L_b + L_c + L_m \tag{1}$$

where L_b, L_c, and L_m are the bounding-box loss, classification loss, and binary cross entropy loss, respectively.

Mask RCNN offers a methodology for object instance segmentation and moreover predicts human poses within the same framework (See Figure 2).

Figure 2. Mask segmentation samples.

We revamped and improved the Mask R-CNN in our face segmentation mask module to retrieve masks and facial characteristics using the Res-Net50 network, which is used as the basic network for extracting the features. Our system accepts a segmentation mask as input and generates images that are consistent with the segmentation mask manipulations.

Second, rather than inventing and generating new hairstyle features from the ground up, our generative hairstyle module treats hairstyle as a transfer problem, extracting hairstyle from the reference images and transferring it to the composite one.

Our Generative Hairstyle module relies on the StyleGAN version 2 architecture and enhanced StyleGAN Embedding algorithm [9,44]. There are numerous latent spaces available for embedding. The input latent spaces Z and the intermediate latent spaces W are two

popular choices. The mapping model in StyleGAN version 2 converts the Z latent spaces to W latent spaces by passing them through a fully connected neural network. Even though the W spaces are less conflated than the Z spaces, both use 512-dimensional vectors. These latent spaces do not offer sufficient expressiveness to accurately represent all real faces.

In Our Generative Hairstyle module, we propose to embed through an expanded latent spaces called W^+, which is a combination of eighteen distinct 512-dimension vectors, one vector for every block of the StyleGAN2 model,

$$W^+ = \{W_i\}_{i=1}^{18} \tag{2}$$

This demonstrated that W^+ space, with degrees of freedom 18 times larger than W space, are proficient of regenerating images. The generative hairstyle module identifies W^+ space for the composite image that offers the advantage of removing many artifacts, especially at the boundary lines of the mixed areas as shown in Figures 3 and 4.
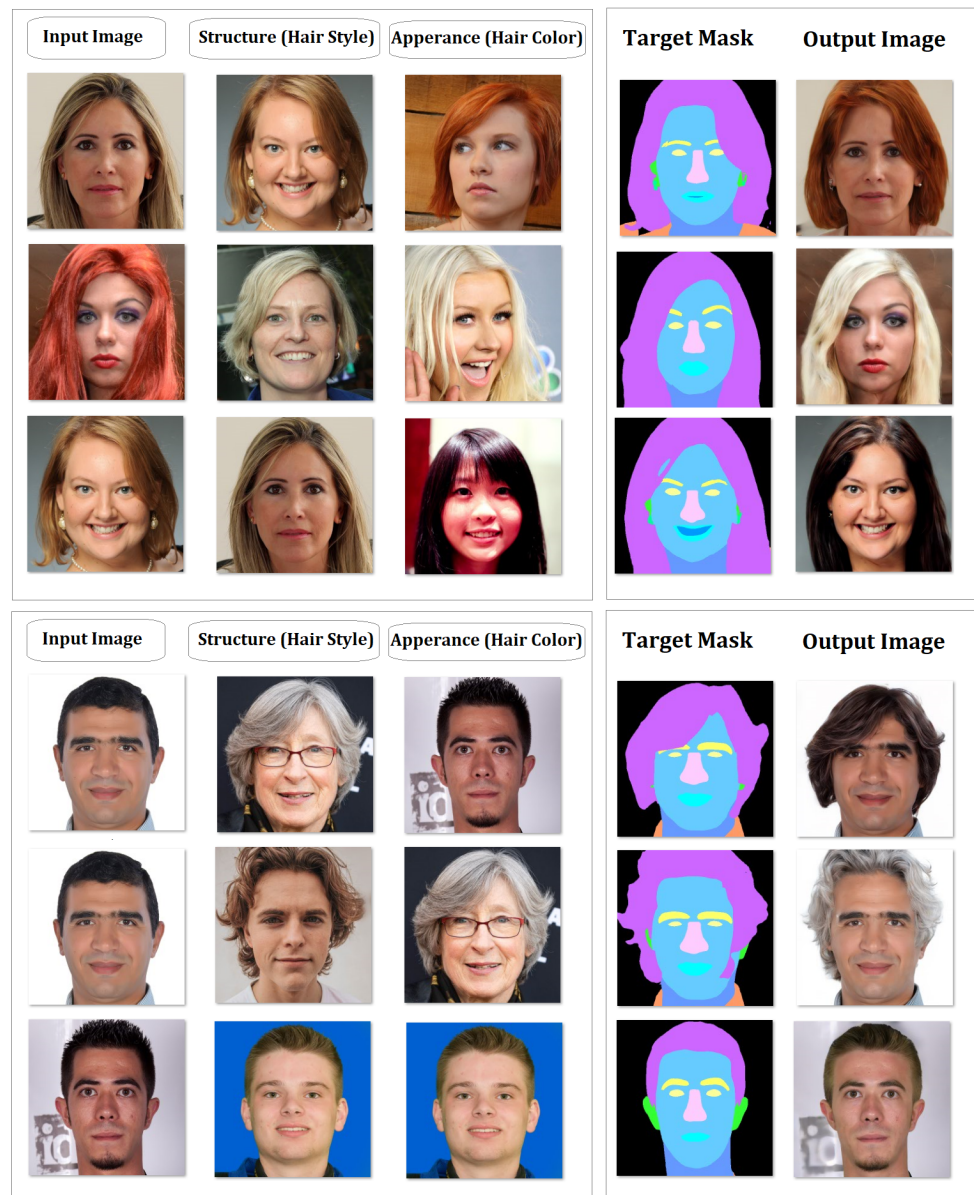

Figure 3. Hairstyle transfer is accomplished by transferring appearance features and structural details from input images.

Figure 4. Samples of generated new hairstyles.

4. Experimental Results

The experimental findings of the suggested hairstyle system are described in this section. Using the PyTorch framework, we developed, trained, and improved the suggested generative neural network models and methods. To verify the experimental findings, a number of high-resolution datasets that were produced in varied circumstances have been employed. The datasets that we used in our research are mentioned in the next subsection.

4.1. High Resolution Datasets

The availability of high-resolution datasets such as FFHQ [9], CelebA-HQ [12], AFHQ [16], and LSUN [17] participated in realistic high-resolution image production. These datasets provide enough variation and high resolution to train GANs, as well as assist in the development of realistic systems.

In our system, we trained and validated the proposed generative network model which emphasizes StyleGAN2 using the cutting-edge CelebA-HQ and FFHQ datasets (See Figure 5) for facial images generation.

CelebA-HQ is the high definition copy of the CelebA dataset, which includes 30,000 faces with a 1024 × 1024 resolution. A popular and very well dataset for producing high-quality images is FFHQ, which includes 70,000 PNG faces with a 1024 × 1024 resolution, and a wide range of age, ethnic origin, and image backgrounds. Additionally, it includes a wide variety of accessory items like caps, earrings, and glasses.

Figure 5. CelebA-HQ and FFHQ datasets.

4.2. Evaluation of Hairstyle System

We employ our innovative method to transfer hairstyles onto face images, then we evaluate our outcomes to the cutting-edge using the following established metrics:

We apply our novel approach to perform hairstyle transfer on portrait images and to evaluate our results to the state of the art, we use the following established metrics: (1) FID [45] (Fréchet inception distance), (2) PSNR [46] (peak-signal-to-noise-ratio), and (3) SSIM [46] (structural similarity index map) to figure out how much the original image differs from the one generated by our suggested model.

By evaluating the distance between Inception features for original and artificial images, FID is utilized to assess image generative neural networks. The computed FID score demonstrates that our method surpasses the latest cutting-edge hairstyle transfer outcomes. With FID, we extract features using the inception network and model the feature space using a Gaussian model. The distribution of generated images is then calculated and compared to the distribution of a set of genuine images [45]. The FID is calculated using the following formula:

$$FID(r,g) = \left\| \mu_r - \mu_g \right\|^2 + Tr(\sum_r + \sum_g - 2(\sum_r \sum_g)^{\frac{1}{2}}) \qquad (3)$$

where the mean and covariance of the distributions of the genuine and generated images, respectively, are $(\mu_r, \sum_r)$ and $(\mu_g, \sum_g)$. Lower FID values indicate smaller differences between synthetic and real data distributions [45].

The ratio of highest permissible signal power to noise power, which has an impact on the representational quality of the signal, is computed using the PSNR formula. Decibels are used to measure the ratio between two images. PSNR is frequently utilized to assess the

quality of images. Better image generation is indicated by a higher PSNR ratio [46]. Most frequently, lossy compression codecs' quality of reconstruction is assessed using PSNR. In this instance, the original data is the signal, while the error brought on by compression is the noise. The following equation is used to determine PSNR:

$$PSNR = 10\log_{10}(\frac{Max^2}{MSE}) \tag{4}$$

where *Max* is the maximum possible pixel value of the image (Peak Value), and *MSE* stands for mean square error.

SSIM is a model that is based on perception. In this strategy, image degradation is seen as a shift in how structured information is perceived. Additionally, it is compatible with a number of other crucial perception-based aspects like brightness masking, contrasting mask, and more. The phrase "structured information" refers to pixels that are strongly interdependent or spatially closed. The lower the image distortion, the higher the value of SSIM [46]. A quality measuring metric is computed using the SSIM index approach based on the computation of three key factors: brightness, contrast, and structural or correlation term. These three factors are multiplied together to get this index as shown in the following equation:

$$SSIM(r,g) = [l(r,g)]^\alpha.[c(r,g)]^\beta.[l(r,g)]^\gamma \tag{5}$$

where l, c, and s are the luminance, contrast, and structure respectively. While α, β and γ are the positive constants.

The generative network model's performance cannot be accurately assessed by a single criterion. A single evaluation factor cannot accurately evaluate the performance of the network model; thus, Table 1 demonstrates the comparison of suggested work against various techniques using FID, PSNR, and SSIM metrics.

Table 1. The comparison of proposed work against various techniques using FID, PSNR, and SSIM metrics.

Method	FID↓	PSNR↑	SSIM↑
STGAN [47]	47.54	17.92	0.72
MagGAN [48]	41.32	20.43	0.78
MichiGAN [8]	26.85	26.51	0.90
LOHO [7]	46.53	22.28	0.83
Proposed system	**22.14**	**29.98**	**0.91**

Figure 4 illustrates how our suggested system led to new, more natural-looking haircuts than other methods by preserving the original facial features.

A comparison of the outcomes produced by our approach and other generative networks is shown in Figure 6.

Despite the fact that some techniques can provide pretty clear images, their performance is limited in terms of unnatural hair and face features, and blurred edges. Our system struggles in several circumstances, like the ones when the face is concealed by a hand or by sunglasses. The generated image lacks authenticity.
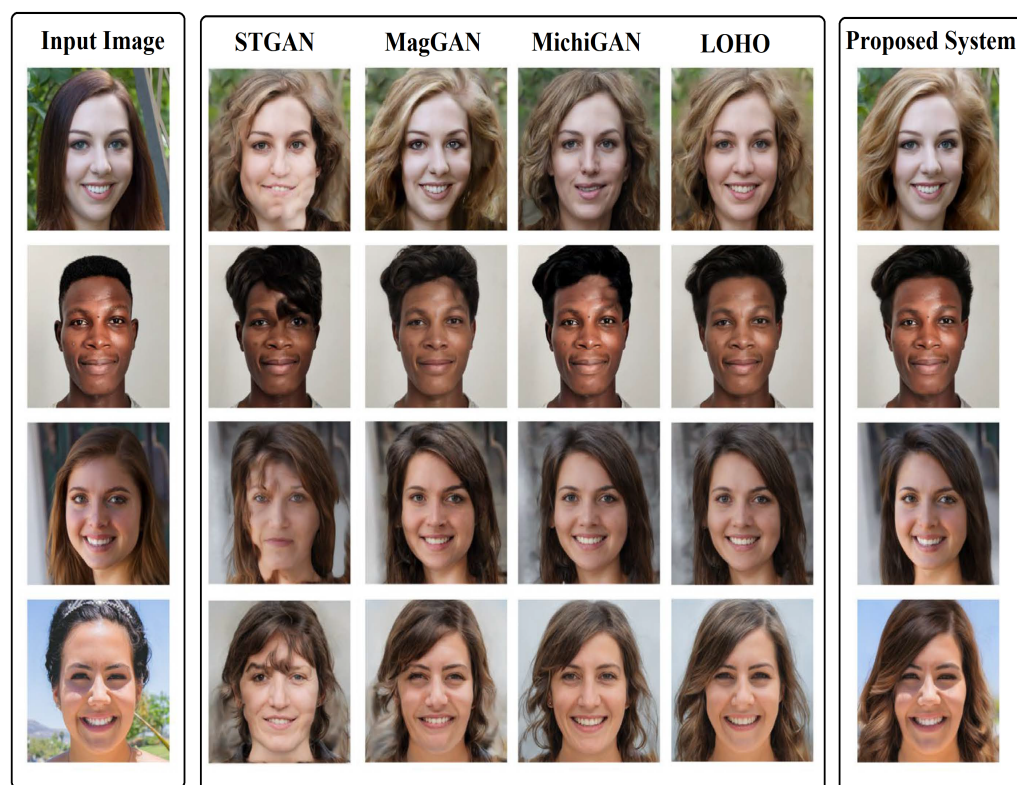
Figure 6. A comparison of the outcomes produced by our approach and other generative networks.

5. Conclusions and Future Research

By manipulating the semantic segmentation mask and blending selected features from multiple images (e.g., hair appearance, hair shape, hair structure, face, and background), we developed a novel realistic hairstyle model for GAN-based image manipulation in this study. We propose a new latent space rather than pixel space for image blending is more effective at maintaining facial characteristics and conveying spatial features.

The blending of the images in the proposed latent spaces allows us to synthesize consistent images. We suggest a GAN-based embedding approach that can embed an image and force it to adhere to our improved semantic segmentation mask. The findings show greater improvements than those of other cutting-edge techniques. Our approach generates a consistent image while avoiding the artifacts present in previous approaches.

In further studies, we will keep enhancing our model to generate facial images that are more accurately and naturally in a variety of difficult situations, such as when the face is obscured by earrings, jewelry, glasses, hand etc. Future research could resolve these limitations.

References

1. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *Commun. ACM* **2020**, *63*, 139–144. [CrossRef]
2. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv* **2015**, arXiv:1511.06434.
3. Metz, L.; Poole, B.; Pfau, D.; Sohl-Dickstein, J. Unrolled generative adversarial networks. *arXiv* **2016**, arXiv:1611.02163.
4. Mao, X.; Li, Q.; Xie, H.; Lau, R.Y.; Wang, Z.; Paul Smolley, S. Least squares generative adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2794–2802.
5. Abdal, R.; Qin, Y.; Wonka, P. Image2stylegan++: How to edit the embedded images? In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 8296–8305.
6. Huang, X.; Belongie, S. Arbitrary style transfer in real-time with adaptive instance normalization. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 1501–1510.
7. 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, Virtual, 19–25 June 2021; pp. 1984–1993.
8. 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. *ACM Trans. Graph.* **2020**, *39*, 95. [CrossRef]
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; pp. 4401–4410.
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, 14–19 June 2020; pp. 8110–8119.
11. Karras, T.; Aittala, M.; Hellsten, J.; Laine, S.; Lehtinen, J.; Aila, T. Training generative adversarial networks with limited data. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 12104–12114.
12. Karras, T.; Aila, T.; Laine, S.; Lehtinen, J. Progressive Growing of GANs for Improved Quality, Stability, and Variation. In Proceedings of the International Conference on Learning Representations, Vancouver, BC, Canada, 30 April–3 May 2018.
13. Abdal, R.; Qin, Y.; Wonka, P. Image2stylegan: How to embed images into the stylegan latent space? In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 4432–4441.
14. Shen, Y.; Yang, C.; Tang, X.; Zhou, B. Interfacegan: Interpreting the disentangled face representation learned by gans. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *44*, 2004–2018. [CrossRef] [PubMed]
15. Bau, D.; Strobelt, H.; Peebles, W.; Wulff, J.; Zhou, B.; Zhu, J.Y.; Torralba, A. Semantic photo manipulation with a generative image prior. *arXiv* **2020**, arXiv:2005.07727.
16. Choi, Y.; Uh, Y.; Yoo, J.; Ha, J.W. Stargan v2: Diverse image synthesis for multiple domains. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 8188–8197.
17. Yu, F.; Seff, A.; Zhang, Y.; Song, S.; Funkhouser, T.; Xiao, J. Lsun: Construction of a large-scale image dataset using deep learning with humans in the loop. *arXiv* **2015**, arXiv:1506.03365.
18. Brock, A.; Donahue, J.; Simonyan, K. Large scale GAN training for high fidelity natural image synthesis. *arXiv* **2018**, arXiv:1809.11096.
19. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Li, F.F. 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.
20. Härkönen, E.; Hertzmann, A.; Lehtinen, J.; Paris, S. Ganspace: Discovering interpretable gan controls. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 9841–9850.
21. Tewari, A.; Elgharib, M.; Bharaj, G.; Bernard, F.; Seidel, H.P.; Pérez, P.; Zollhofer, M.; Theobalt, C. Stylerig: Rigging stylegan for 3d control over portrait images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 6142–6151.
22. Abdal, R.; Zhu, P.; Mitra, N.J.; Wonka, P. Styleflow: Attribute-conditioned exploration of stylegan-generated images using conditional continuous normalizing flows. *ACM Trans. Graph. (ToG)* **2021**, *40*, 1–21. [CrossRef]
23. Patashnik, O.; Wu, Z.; Shechtman, E.; Cohen-Or, D.; Lischinski, D. Styleclip: Text-driven manipulation of stylegan imagery. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 2085–2094.
24. Frühstück, A.; Alhashim, I.; Wonka, P. Tilegan: Synthesis of large-scale non-homogeneous textures. *ACM Trans. Graph. (ToG)* **2019**, *38*, 1–11. [CrossRef]
25. Collins, E.; Bala, R.; Price, B.; Susstrunk, S. Editing in style: Uncovering the local semantics of gans. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 5771–5780.
26. Wu, Z.; Lischinski, D.; Shechtman, E. Stylespace analysis: Disentangled controls for stylegan image generation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, 19–25 June 2021; pp. 12863–12872.
27. Kim, H.; Choi, Y.; Kim, J.; Yoo, S.; Uh, Y. Exploiting spatial dimensions of latent in gan for real-time image editing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, 19–25 June 2021; pp. 852–861.
28. Chai, M.; Luo, L.; Sunkavalli, K.; Carr, N.; Hadap, S.; Zhou, K. High-quality hair modeling from a single portrait photo. *ACM Trans. Graph. (TOG)* **2015**, *34*, 1–10. [CrossRef]

29. Chai, M.; Wang, L.; Weng, Y.; Jin, X.; Zhou, K. Dynamic hair manipulation in images and videos. *ACM Trans. Graph. (TOG)* **2013**, *32*, 1–8. [CrossRef]

30. Weng, Y.; Wang, L.; Li, X.; Chai, M.; Zhou, K. Hair interpolation for portrait morphing. In *Computer Graphics Forum*; Wiley Online Library: Hoboken, NJ, USA, 2013; Volume 32, pp. 79–84.

31. Wei, L.; Hu, L.; Kim, V.; Yumer, E.; Li, H. Real-time hair rendering using sequential adversarial networks. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 99–116.

32. Lee, C.H.; Liu, Z.; Wu, L.; Luo, P. Maskgan: Towards diverse and interactive facial image manipulation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 5549–5558.

33. 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.

34. Tewari, A.; Elgharib, M.; Bernard, F.; Seidel, H.P.; Pérez, P.; Zollhöfer, M.; Theobalt, C. Pie: Portrait image embedding for semantic control. *ACM Trans. Graph. (TOG)* **2020**, *39*, 1–14. [CrossRef]

35. Zhu, J.; Shen, Y.; Zhao, D.; Zhou, B. In-domain gan inversion for real image editing. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; Springer: Cham, Switzerland, 2020; pp. 592–608.

36. Zhu, P.; Abdal, R.; Qin, Y.; Femiani, J.; Wonka, P. Improved stylegan embedding: Where are the good latents? *arXiv* **2020**, arXiv:2012.09036.

37. Richardson, E.; Alaluf, Y.; Patashnik, O.; Nitzan, Y.; Azar, Y.; Shapiro, S.; Cohen-Or, D. Encoding in style: A stylegan encoder for image-to-image translation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, 19–25 June 2021; pp. 2287–2296.

38. Fedus, W.; Goodfellow, I.; Dai, A.M. MaskGAN: Better Text Generation via Filling in the _. In Proceedings of the International Conference on Learning Representations, Vancouver, BC, Canada, 30 April–3 May 2018.

39. Park, T.; Liu, M.Y.; Wang, T.C.; Zhu, J.Y. Semantic image synthesis with spatially-adaptive normalization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 2337–2346.

40. 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, 18–22 June 2018; pp. 8789–8797.

41. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.

42. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems 28 (NIPS 2015)*; Curran Associates, Inc.: Red Hook, NY, USA, 2015.

43. Moussa, M.M.; Shoitan, R.; Abdallah, M.S. Efficient common objects localization based on deep hybrid Siamese network. *J. Intell. Fuzzy Syst.* **2021**, *41*, 3499–3508. [CrossRef]

44. Kynkäänniemi, T.; Karras, T.; Laine, S.; Lehtinen, J.; Aila, T. Improved precision and recall metric for assessing generative models. In *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*; Curran Associates, Inc.: Red Hook, NY, USA, 2019.

45. 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. In *Advances in Neural Information Processing Systems 30 (NIPS 2017)*; Curran Associates, Inc.: Red Hook, NY, USA, 2017.

46. Sara, U.; Akter, M.; Uddin, M.S. Image quality assessment through FSIM, SSIM, MSE and PSNR—A comparative study. *J. Comput. Commun.* **2019**, *7*, 8–18. [CrossRef]

47. Liu, M.; Ding, Y.; Xia, M.; Liu, X.; Ding, E.; Zuo, W.; Wen, S. Stgan: A unified selective transfer network for arbitrary image attribute editing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3673–3682.

48. Wei, Y.; Gan, Z.; Li, W.; Lyu, S.; Chang, M.C.; Zhang, L.; Gao, J.; Zhang, P. Maggan: High-resolution face attribute editing with mask-guided generative adversarial network. In Proceedings of the Asian Conference on Computer Vision, Kyoto, Japan, 30 November–4 December 2020.