DCT-SwinGAN: Leveraging DCT and Swin Transformer for Face Synthesis from Sketch and Thermal Domains

Haresh Kumar Kotadiya, Satish Kumar Singh, Shiv Ram Dubey, and
Nand Kumar Yadav
info.hareshkotadiya@gmail.com, sk.singh@iiita.ac.in,
srdubey@iiita.ac.in, pis2016004@iiita.ac.in

Computer Vision and Biometrics Lab (CVBL),
Indian Institute of Information Technology, Allahabad

Abstract. Face generation remains a crucial task owing to its applications in crime investigation, entertainment, etc. Sketch to face synthesis is an important task accomplished using Generative Adversarial Network (GAN) models. The generator of the GAN is usually designed using Convolutional Neural Networks (CNNs). GANs are able to deliver promising outcomes in some cases, but fail in others. Owing to the local nature of CNN, it is unable to keep track of long range dependencies, limiting the model to local information and delivering underwhelming results. However, recently developed Transformers encode the global context with long range dependencies. Swin Transformer is designed to be suitable for the images with promising performance on several vision tasks. We utilize the Swin Transformer along with Discrete Cosine Transform (DCT) in GAN framework and propose DCT-SwinGAN for face synthesis from sketch and thermal domains. DCT-SwinGAN comprises a multi scale discriminator paired with a generator comprising an attention module, DCT ResNet Convolution, Deconv decoder, and Swin Transformer. The proposed model captures not only local but also global information to generate realistic face images. The generated model outperforms the existing state-of-the-art models on CUHK sketch-to-face and the WHU-IIP thermal-to-face datasets.

Keywords: Face Synthesis · Sketch · Thermal · GAN · DCT · Swin Transformer · Multi Scale Discriminator · Image Translation.

1 Introduction

In the area of image processing and computer vision generating new images brings in a lot of new challenges. It is a cross-domain task that has multiple use cases including image-to-image translation, facial recognition, image synthesis, etc. The image to image translation is one such use case of image generation tasks. The problem at hand is the sketch and thermal to visible image translation. The solution to this problem is inspired by language translation and

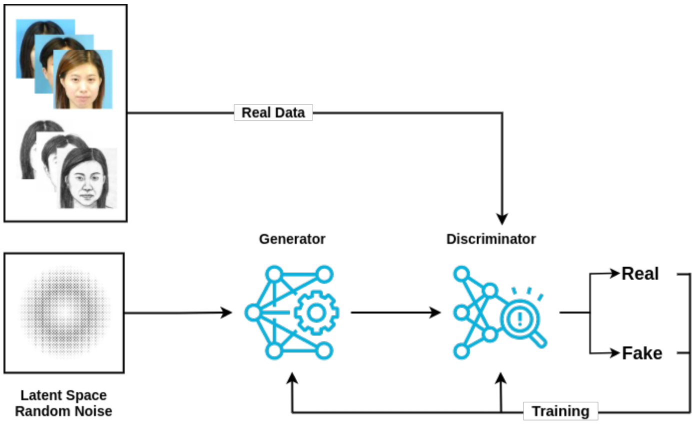

Fig. 1. Generative Adversarial Network. Facial sketches and facial images are taken from CUHK Dataset [17].

presents deep learning models. Obtaining a labeled data-set is not always possible as data labeling is quite expensive ruling out supervised learning methods. In unsupervised learning, a Generative Adversarial Network (GAN) [1] is used to generate new images using Convolutional Neural Network (CNN) [30]. GAN comprises two neural networks as shown in Fig. 1. The first neural network is the generator, which takes input from the latent space of training data distribution. The generator picks random points to generate fake images. The generator and discriminator network plays the mini-max game. The discriminator tries to label the output as fake and gives the feedback through back-propagation. By following adversarial training, the generator generates the realistic images and the discriminator accepts the fake image as a real one.

CNN fails to generate the good-quality images owing to the fact that it is not able to catch long-term dependencies in the images. Self-attention [7] modules can be used to overcome the problem of CNN. The main purpose of the attention module in the generator model for the image synthesis task is to concentrate on important features and details of the input images. A sketch image has fewer details as compared to colored images owing to the change in focus on the edge of the sketch for generating a high-quality result. Self-attention module is capable of capturing long-range dependencies but for higher quality images and to capture global dependency Transformer models can be used.

The Transformer model [3] is inspired by natural language processing and suitable for capturing sequential details hence ideal for capturing global dependencies in images. In recent years, Transformers have been very actively utilized for vision tasks [31]. Transformer [7] consists of an encoder and a decoder. Transformer leverages a self-attention module to capture long range dependencies as well as global dependencies in images. With all the advantages that the Transformer brings in, the computation complexity is also quite significant in comparison to the GAN model. To decrease the computational complexity Shifted Window (Swin) Transformer [16] is used. It works by hierarchically merging patches neighboring to each other as one goes deeper into a Transformer model. Thus. the complexity becomes linear corresponding to the size of feature maps.

The contributions of the proposed work are:

- In this paper, a novel combination of DCT Convolution and Swin Transformer is exploited in the generator architecture of the proposed DCT-SwinGAN. A multi-scale discriminator is utilized to train the generator model.
- The proposed model better encodes the global relationship in images as compared to the existing GAN models and leads to improved visual quality, accurately representing the input image with enhanced details and texture.
- The proposed model is validated using two datasets, namely the CUHK sketch-to-face and WHU-IIP thermal-to-face dataset with improved results. The analysis of loss functions is also performed to justify the used losses.

2 Related Work

GAN [1] uses mini-max optimization between the generator and discriminator networks for image synthesis from random noise in a given distribution. The GAN is also utilized for image-to-image translation in Pix2Pix [20] by considering the generator as an encoder-decoder structure which takes an image in source modality as the input and produces an image in target domain as the output. The ConditionalGAN [8] explicitly provides a condition with the standard input to generate output with a given input condition. The CycleGAN [9] utilizes a cyclic consistency adversarial network model for unpaired image-to-image translation as it is not always possible to have paired data input. CycleGAN uses a cycle consistency loss to train the model. The Identity-aware CycleGAN [11] utilizes the CycleGAN model face photo-sketch synthesis and recognition. Similar to CycleGAN, a DualGAN model [13] also utilizes the image translation in both directions, i.e., source domain to target domain and vice-versa have proposed a novel architecture that is the DualGAN model. The DualGAN is different from the CycleGAN in terms of the construction loss.

Kancharagunta et al. [5] developed a cyclic-synthesized generative adversarial network (CSGAN) using the CycleGAN framework [9]. CSGAN uses the cyclic synthesized loss to generate more realistic images. Kancharagunta et al. have also exploited the perceptual loss in PCSGAN [24]. A composition-aided GAN architecture is proposed in [12] for face photo–sketch synthesis by exploiting the compositional reconstruction loss. Coupled generative adversarial networks [15] have the ability to understand the pattern that humans can not understand or learn directly. In image synthesis using GANs, the generator using CNNs only captures the local information of the images leaving out long-range dependencies leading to the generation of poor-quality images.

The self-attention [3] mechanism has been extensively exploited for several tasks, including vision applications. Several attention-guided GANs have been proposed for image synthesis [26], [27], [14], [22]. The GAN with cycle-synthesized attention is exploited in CSA-GAN model for image synthesis [26]. CSA-GAN is trained in a cyclic manner with the cycle consistency loss. To narrow the learning in the attention direction cycle synthesis objectives are used. In TVA-GAN [27], attention-guided synthesis is utilized to convert thermal pictures into visible images. TVA-GAN exploits the recurrent inception block with an attention mechanism to learn the local and focused sparse structure. The attention module utilizes the spatial attention maps and important regions' information from the image in [14]. A self-attention GAN is proposed in [22] by preserving the global dependencies through the dissimilarity among different pixels in an image, i.e., spatial locations. Though the attention mechanism is exploited with GANs for image synthesis, the performance is still limited due to the CNN based networks.

Recent methods try to explore the Transformer based GANs for image synthesis. CNN fails in capturing the structural and global details of the input. The Vision Transformer (ViT) model [10] works on images by dividing the image into patches which are considered as the tokens. ViT captures the long-range dependencies with outstanding performance, but fails to handle the low-level vision tasks. Global and local self-attention is exploited in [7] for face photo-sketch synthesis. The ViT model is improved to Swin Transformer in [16] with the help of shifted windows. The Swin Transformer is utilized with Fast Fourier Transform (FFT) in GAN framework [25] to capture the local and global dependencies in the images for face photo-sketch synthesis. In this paper, we utilize the Swin Transformer with Discrete Cosine Transform in GAN framework for sketch/thermal face to visible face synthesis.

3 Proposed DCT-SwinGAN Model

In this paper, we propose a Swin Transformer and DCT based GAN model which performs sketch/thermal face to visible face synthesis as depicted in Fig. 2. From the given input sketch image let's say x, and we have to generate $G(x)$ using the generator. Here, we use the paired dataset which contains the ground truth images and the corresponding input sketch/thermal images. Discriminator network helps in generating more realistic images by differentiating between ground truth and generated images. The generator is a combination of an improved attention block, Discrete Cosine Transform (DCT) ResNet convolution network, Swin Transformer block in the deconvolution layer, and a deconvolutional decoder block.

If the input sketch/thermal image has insufficient information, then the resultant image will be blurry or noisy and we will get artifacts in our resultant image. So, we add the attention block to the generator, which resolves this issue and generates the spatial attention map containing the information of input features. In image synthesis, capturing the global contexts of the input plays a very important role. However, CNN fails in capturing long-range dependencies in images and hence generates poor-quality images. The input image, i.e., the

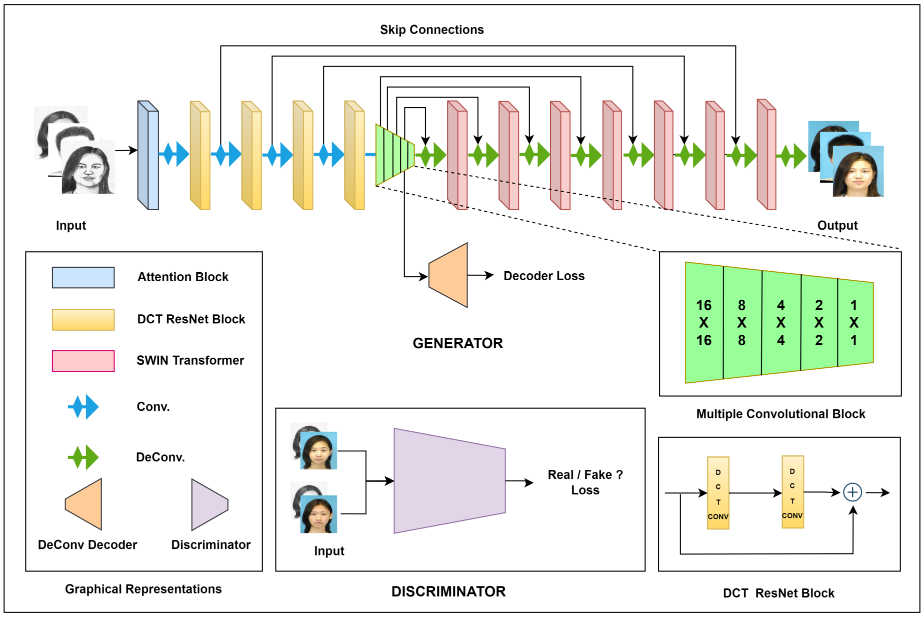

Fig. 2. Proposed Network for Image synthesis. The dimension of the image and sketch image is 256 × 256. Main blocks of architecture are attention module, DCT ResNet module for encoder and Swin Transformer in decoder. To obtain an optimal SSIM Score dimension of output image has been converted to the power of two.

sketch/thermal data, holds global structure and information that are essential for generating the resulting image. DCT convolution is crucial for capturing long-range context. DCT block can capture the global and local context of the image by converting the input into the frequency domain from the spatial domain. The spatial domain is converted into the frequency domain, as the frequency domain is able to work with more global domains in comparison to the spatial domain to get optimal results. Also, the frequency domain removes the noisy frequency elements from the data.

Due to the failure of CNN to capture long-term dependencies, one solution for that to exploit the Transformer network. However, the computational complexity of traditional Transformers is too high. Instead, we use the Swin Transformer [16], which splits the image into non overlapping windows, restricting the self attention computation to that part only, unlike a traditional Transformer. In the network, sometimes there is a chance of vanishing gradient. Introducing the deconv decoder, it is fed with the input by the preceding convolution part and generates the synthesized image. It further calculates the loss for intermediate feature maps. In the network, we use skip connections for preserving the feature information and stabilize the training network. We use a multi-scale discriminator for distinguishing between ground truth and generated images. Also, discriminator's output influence the SSIM score. When we make the discriminator's output power of two then it aligns with the image resolution and leads to better SSIM score as the generated and ground truth images match the resolution.

The optimization of a GAN model is driven by the loss function used in the training of the model and for the generation of efficient results. A GAN model comprises of a generator and discriminator networks entwined with each other to minimize their respective loss functions for optimal results. Following are the different losses used to train the proposed model:

Adversarial Loss (AL): Adversarial loss is one of the foundational components of the GAN model [1]. GAN follows the adversarial training in which both the components of the GAN model, i.e., generator (G) and discriminator (D) work like competitors. Adversarial loss is calculated between the generator and discriminator network as,

$$\mathcal{L}_{minmax} \min_{G} \max_{D} = E_{x \sim p_{data}(x)}[\log D(x)] + E_{z \sim p_z(z)}[1 - \log D(G(z))]. \tag{1}$$

Feature Matching Loss (FML): The generator network is optimized to generate the output images aligned with the target distribution and fool the discriminator simultaneously using the feature matching loss [18]. This results in generating images of high perceptible quality and leads to the stabilization of GAN. The goal of feature matching loss is to diminish the loss in terms of average feature representation of the ground truth and generated images. Feature matching loss is given as,

$$\mathcal{L}_feature = \left\| E_{x \sim p_{data}} \mathbf{f}(x) - E_{z \sim p_z(z)} \mathbf{f}(G(z)) \right\|_2^2 \tag{2}$$

where, $\mathbf{f}(x)$ represents the feature vector obtained from a specific layer of the discriminator network for an input x and $\mathbf{f}(G(z))$ represents the feature vector obtained from the same layer for a generated sample $G(z)$.

Perceptual Loss (VPL): Perceptual loss [19] aids in improving the visual quality of the images and calculated using the feature maps of intermediate layers of the network. The loss is calculated by comparing the feature map of input and output images. The generator tries to generate images that match the high level perceptual features of the input images. The perceptual loss is given as,

$$\mathcal{L}_{Perceptual} = \sum_{i} w_{\mathrm{i}}{}^{*} d(F(r), F(g)) \tag{3}$$

where $F(r)$ and $F(g)$ are the feature maps of VGG model for reference and generated images and w_{i} is the weight for i^{th} layer.

Decoder Loss (DL): Decoder loss is calculated between the real image and intermediate generated image from the deconv decoder. Mathematically, decoder loss is represented as a reconstruction loss [21] which is often represented using a distance or dissimilarity metric as,

$$\mathcal{L}_{decoder} = \frac{1}{n} \sum_{i=1}^{n} (x_i - G(z)_i)^2. \tag{4}$$

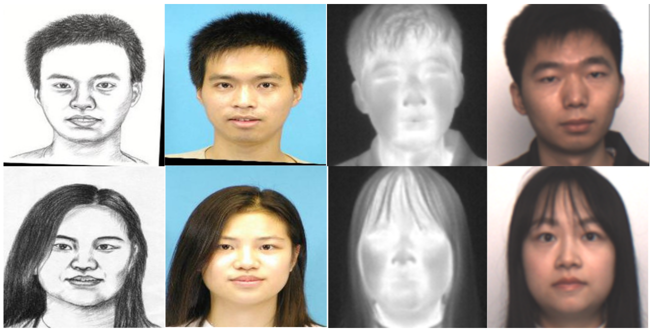

Fig. 3. Sample paired images of 256×256 dimension. From left to right: CUHK Sketch Face, CUHK Visible Face, WHU-IIP Thermal Face and WHU-IIP Visible Face images.

The discriminator and the generator are the two primary parts of GANs. While the discriminator tries to tell the difference between genuine and created images, the generator seeks to produce realistic images. The training procedure becomes more stable by including the relevant losses, such as adversarial loss, perceptual loss, feature matching loss, and decoder loss. The various losses provide the generator supplementary information that aids in directing it to create high-quality images.

4 Experimental Results

4.1 Experimental Settings

For the validation of the generated model, the CUHK Sketch-to-Face synthesis dataset is exploited [17]. It consists of 188 image pairs. This dataset is widely used amongst researchers for image translation and synthesis. The size of each of the images is 200×250 pixels which have been resized to 256×256 dimensions images. The proposed model is also validated using a Thermal-to-Visible dataset, i.e., WHU-IIP [28]. The sample images of the dataset are shown in Fig. 3.

The proposed model is trained by Adam optimizer with $\beta_1 = 0.5$ and $\beta_2 = 0.999$, respectively in the PyTorch framework. The batch size is used as 2. The training is performed using a Nvidia GTX 1080TI GPU. The proposed DCT-SwinGAN model is compared with various state-of-the-art (SOTA) image-to-image translation models, including Pix2pix [20], CycleGAN [9], DualGAN [13], CSGAN [5], AGGAN [14], Sketch2PhotoGAN [25], AttentionGAN [22], CT-GAN [23], CTIAGAN [23], and PCSGAN [24]. The structural similarity index (SSIM) score is used to evaluate the model. SSIM calculates the similarity gap between the ground truth and generated images. SSIM score takes into consideration a few parameters while comparing the images, i.e., contrast variance, luminance of parts of the data, and most importantly, the structure of the image.

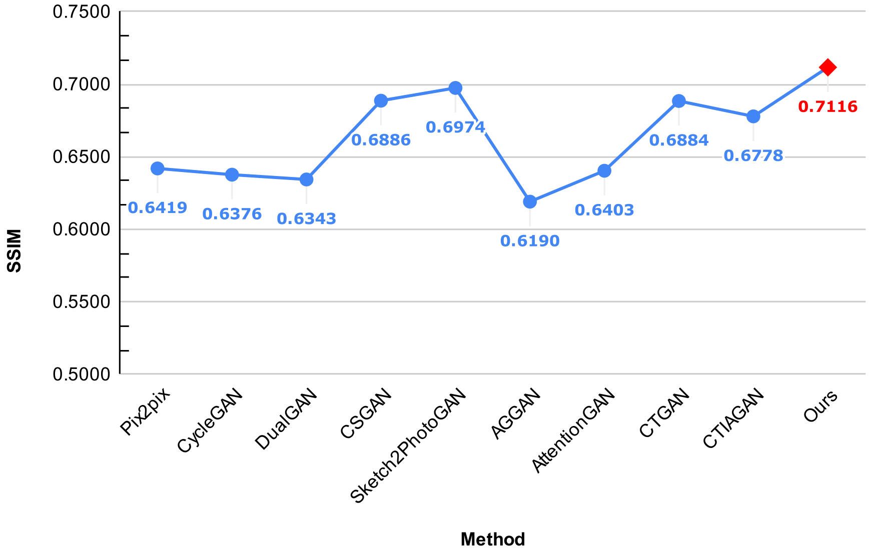

Fig. 4. Comparisons of SSIM score of various SOTA methods on CUHK face dataset. A higher value of the SSIM score is nearer to the ground truth images

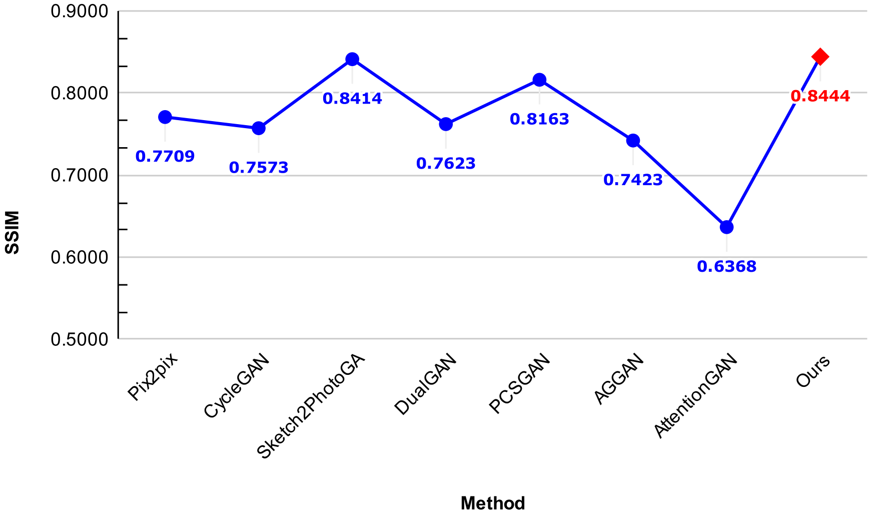

Fig. 5. Comparisons of SSIM score of various SOTA methods on WHU-IIP dataset. A higher value of the SSIM score is nearer to the ground truth images

4.2 Result Analysis

In Fig. 4, a comparison chart corresponding to the CUHK Face dataset is presented. It contains the SSIM scores of the various aforementioned SOTA methods in comparison to the proposed model. In comparison to our model, the percentage change in SSIM scores is 10.85%, 11.60%, 12.18%, 3.34%, 2.03%, 14.95%, 11.13%, 3.37%, and 4.98% less for Pix2pix, CycleGAN, DualGAN, CSGAN, Sketch2PhotoGAN, AGGAN, AttentionGAN, CTGAN and CTIAGAN, respectively. It is clear that DCT-SwinGAN model notably outperforms other GAN approaches for sketch-to-face synthesis.

In order to test the dataset generalization ability of the DCT-SwinGAN, we also validate the results for thermal-to-face synthesis. Fig. 5 presents the comparison chart of SSIM scores in correspondence to the WHU-IIP dataset. In comparison to our model, the percentage change in SSIM scores is 9.53%, 11.50%, 0.35%, 10.77%, 3.44%, 13.75%, 32.60% less for Pix2pix, CycleGAN, Sketch2PhotoGAN, DualGAN, PCSGAN, AGGAN, AttentionGAN, respectively. It noticed that the proposed model is also able to outperform the existing GAN models for thermal-to-face image synthesis.

Fig. 6 depicts the qualitative results in terms of the generated face images using the proposed DCT-SwinGAN model. The left half of the figure contains the sketch image, synthesized image, and ground truth image corresponding for the samples taken from the CUHK dataset. It can be stated that the generated image is quite similar to the ground truth image with a very minute difference in color. The right half of the figure contains the thermal input image, the synthesized image, and the ground truth image for the sample images taken from WHU-IIP dataset. The ground truth image and the synthesized image look structurally similar, leading to the good SSIM score. The generated image is softer and missing details like the shape of eyebrows and facial wrinkles, which don't create much difference optically in thermal images.

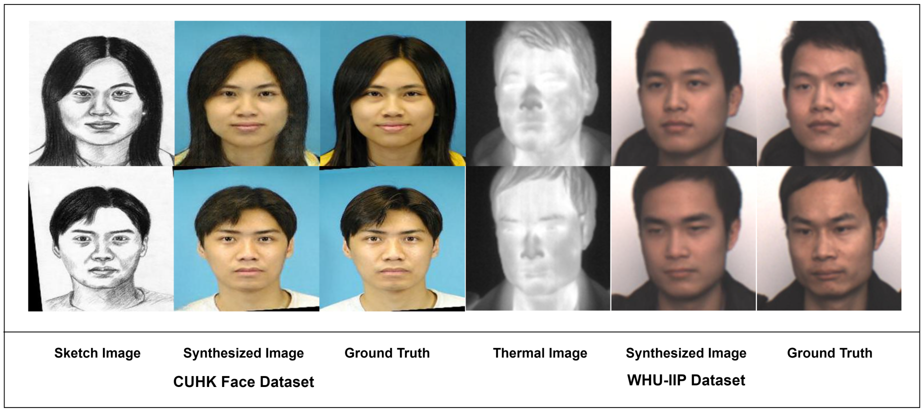

Fig. 6. From left to right: CUHK Face dataset result and WHU-IIP dataset result.

Table 1. Comparison of various losses on CUHK and WHU-IIP datasets.

Losses	SSIM on CUHK	SSIM on WHU-IIP
AL	0.6198	0.7179
AL + FML	0.6843	0.8299
AL + FML + VPL	0.6961	0.8391
AL + FML + VPL + DL	0.7116	0.8444

4.3 Loss Ablation Study

We have used the adversarial loss (AL), feature matching loss (FML), VGG perceptual loss (VPL), and decoder loss (DL) to create the best quality images for the corresponding input sketch or thermal face images. Table 1 shows the effect of different combination of losses used to train the proposed DCT-SwinGAN model in terms of the SSIM scores generated for the CUHK and WHU-IIP datasets. Initially, we use only adversarial loss, which gives an SSIM score of 61.98% on CUHK dataset, which is much lower than the SSIM score of the proposed model. Then feature matching loss is added to the objective function, increasing the SSIM score by 6.45%. To further improve the training, VGG perceptual loss is added leading to an increase in the SSIM score by 1.18%. Lastly, when decoder loss is added to the objective function, the SSIM score is increased by 1.55%, making it to 71.16% on CUHK dataset. Similarily on WHU-IIP dataset, we use only adversarial loss, which gives an SSIM score of 71.79%, which is much lower than the SSIM score of the proposed model. Then feature matching loss is added to the objective function, increasing the SSIM score by 11.2%. Next, VGG perceptual loss is added which leads to increase in the SSIM score by another 0.92%. Lastly, decoder loss increases the SSIM score by another 0.53%, making it to 84.44% on WHU-IIP dataset. It is noticed that the best performance is achieved on both the datasets when all four losses are combined in the final objective function. Thus, it can be stated that due to the use of multiple losses, DCT and Transformer model, the proposed DCT-SwinGAN model outperforms the SOTA GAN models for sketch/thermal face to visible face synthesis.

5 Conclusion

In this paper, the Swin Transformer is utilized with DCT convolution in GAN framework (i.e., DCT-SwinGAN) for face synthesis from sketch and thermal domains. The proposed model integrates attention block, DCT ResNet convolution block, deconv decoder, and Swin Transformer based decoder block in generator for capturing local as well as global information and a multi-scale discriminator. The DCT-SwinGAN is validated on two benchmark face synthesis datasets, namely the CUHK sketch-to-face dataset and the WHU-IIP thermal-to-face dataset. The proposed model outperforms the existing models in terms of the SSIM score on both the datasets. The analysis on different losses is also presented which justifies the use of adversarial, feature matching, VGG perceptual and decoder losses in the objective function. Overall, the proposed model is able to generate high-quality and realistic faces from sketch and thermal faces.

Acknowledgement

The authors acknowledge the funding and facilities provided by IIIT Allahabad to facilitate this research work.

References

1. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2020). Generative adversarial networks. Communications of the ACM, 63(11), 139-144.
2. Lei, Y., Du, W., & Hu, Q. (2020). Face sketch-to-photo transformation with multi-scale self-attention GAN. Neurocomputing, 396, 13-23.
3. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.
4. Tang, H, Xu, D, Sebe, N, Yan, Y: Attention-guided generative adversarial networks for unsupervised image-to-image translation. In: 2019 International Joint Conference on Neural Networks (IJCNN), pp 1–8. IEEE (2019)
5. Kancharagunta, K. B., & Dubey, S. R. (2019). Csgan: Cyclic-synthesized generative adversarial networks for image-to-image transformation. arXiv preprint arXiv:1901.03554.
6. Cao, B., Wang, N., Li, J., Hu, Q., & Gao, X. (2022). Face photo-sketch synthesis via full-scale identity supervision. Pattern Recognition, 124, 108446.
7. Yu, W., Zhu, M., Wang, N., Wang, X., & Gao, X. (2022). An Efficient Transformer Based on Global and Local Self-Attention for Face Photo-Sketch Synthesis. IEEE Transactions on Image Processing, 32, 483-495.
8. Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784.
9. Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision (pp. 2223-2232).
10. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. International Conference on Learning Representations.
11. Fang, Y., Deng, W., Du, J., & Hu, J. (2020). Identity-aware CycleGAN for face photo-sketch synthesis and recognition. Pattern Recognition, 102, 107249.
12. Yu, J., Xu, X., Gao, F., Shi, S., Wang, M., Tao, D., & Huang, Q. (2020). Toward realistic face photo–sketch synthesis via composition-aided GANs. IEEE transactions on cybernetics, 51(9), 4350-4362.
13. Yi, Z., Zhang, H., Tan, P., & Gong, M. (2017, October). DualGAN: Unsupervised Dual Learning for Image-To-Image Translation. Proceedings of the IEEE International Conference on Computer Vision (ICCV).
14. Alami Mejjati, Y., Richardt, C., Tompkin, J., Cosker, D., & Kim, K. I. (2018). Unsupervised attention-guided image-to-image translation. Advances in neural information processing systems, 31.

15. Liu, M. Y., & Tuzel, O. (2016). Coupled generative adversarial networks. Advances in neural information processing systems, 29.
16. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... & Guo, B. (2021). Swin Transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 10012-10022).
17. Wang, X., & Tang, X. (2008). Face photo-sketch synthesis and recognition. IEEE transactions on pattern analysis and machine intelligence, 31(11), 1955-1967.
18. Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., & Chen, X. (2016). Improved Techniques for Training GANs. arXiv [Cs.LG]. Retrieved from http://arxiv.org/abs/1606.03498
19. Johnson, J., Alahi, A., & Fei-Fei, L. (2016). Perceptual losses for real-time style transfer and super-resolution. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14 (pp. 694-711). Springer International Publishing.
20. Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1125-1134).
21. Ganguli, S., Garzon, P., & Glaser, N. (2019). GeoGAN: A conditional GAN with reconstruction and style loss to generate standard layer of maps from satellite images. arXiv preprint arXiv:1902.05611.
22. Tang, H., Xu, D., Sebe, N., & Yan, Y. (2019, July). Attention-guided generative adversarial networks for unsupervised image-to-image translation. In 2019 International Joint Conference on Neural Networks (IJCNN) (pp. 1-8). IEEE.
23. Cao, B., Wang, N., Li, J., Hu, Q., & Gao, X. (2022). Face photo-sketch synthesis via full-scale identity supervision. Pattern Recognition, 124, 108446.
24. Babu, K. K., & Dubey, S. R. (2020). PCSGAN: Perceptual cyclic-synthesized generative adversarial networks for thermal and NIR to visible image transformation. Neurocomputing, 413, 41-50.
25. Liu, H., Xu, Y., & Chen, F. (2023). Sketch2Photo: Synthesizing photo-realistic images from sketches via global contexts. Engineering Applications of Artificial Intelligence, 117, 105608.
26. Yadav, N. K., Singh, S. K., & Dubey, S. R. (2022). CSA-GAN: Cyclic synthesized attention guided generative adversarial network for face synthesis. Applied Intelligence, 52(11), 12704-12723.
27. Yadav, N. K., Singh, S. K., & Dubey, S. R. (2023). TVA-GAN: attention guided generative adversarial network for thermal to visible image transformations. Neural Computing and Applications, 1-21.
28. Wang, Z., Chen, Z., & Wu, F. (2018). Thermal to visible facial image translation using generative adversarial networks. IEEE Signal Processing Letters, 25(8), 1161-1165.
29. Owens, J. D., Houston, M., Luebke, D., Green, S., Stone, J. E., & Phillips, J. C. (2008). GPU computing. Proceedings of the IEEE, 96(5), 879-899.
30. Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). Image style transfer using convolutional neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2414-2423).
31. Dubey, S. R., & Singh, S. K. (2023). Transformer-based Generative Adversarial Networks in Computer Vision: A Comprehensive Survey. arXiv preprint arXiv:2302.08641.