Realistic and Robust Image Transfer using Deep Learning

Rhitik Prajapati
Department of Computer Engineering,
RMD Sinhgad School of Engineering
Pune, India

Sonal Fatangare
Department of Computer Engineering,
RMD Sinhgad School of Engineering
Pune, India

Shreya Nikam
Department of ComputerEngineering,
RMD Sinhgad School of Engineering Pune, India

Devashri Suravase
Department of ComputerEngineering,
RMD Sinhgad School of Engineering
Pune,India

Tilak Raut
Department of Computer Engineering,
RMD Sinhgad School of Engineering Pune, India

ABSTRACT

In the ever-evolving world of deep learning, creating photorealistic photo editing poses is challenging, especially when modifying features such as hairstyles in photos. The system leverages advanced generative adversarial networks (GANs) to solve problems such as misalignment, texturing, and lighting conflicts. A dedicated color adjustment module controls hair color change even under different lighting conditions, while a refinement module restores fine details for highly realistic final images. Recent solutions have shown significant improvements in both speed and accuracy. These advances are paving the way for more implementation in areas like virtual experiments, interactive tournaments, and design tools. In this survey, we examine the most advanced deep learning techniques for processing real-life images, focusing on their ability to handle complex transformations like hair editing.

Keywords

Generative Adversarial Networks (GANs), Image-to-Image Translation, Encoder-Based Approach, StyleGAN, Pose Alignment, Shape and Color Alignment, Image Synthesis.

1. INTRODUCTION

In past few years, the implementation of Generative Adversarial Networks (GANs) in image generation, particularly in facial editing, has significantly advanced. One area that has drawn attention is hairstyle transfer, where the goal is to manipulate hair attributes—such as shape, color, and texture—while maintaining the identity and background of the object in the image. This task is not only challenging due to the complexity of hair structure but also because it requires careful handling of various factors like pose differences between images. These challenges are particularly relevant in fields like virtual reality, gaming, and photo editing applications.

Two primary approaches have emerged in solving this problem: optimization-based methods, which provide high-quality results but are often slow, and encoder-based methods, which are faster but tend to compromise on output quality. Despite progress, limitations remain, especially when dealing with large pose differences. To overcome these challenges, this paper Bintroduces an innovative technique called HairFast, which merges the strengths of existing methods. HairFast enables highresolution hair transfers while improving both speed and output quality, making it highly effective for real-time applications. By incorporating new techniques for pose adaptation, shape alignment, and color transfer, HairFast offers a comprehensive solution to the problem of hairstyle transfer.

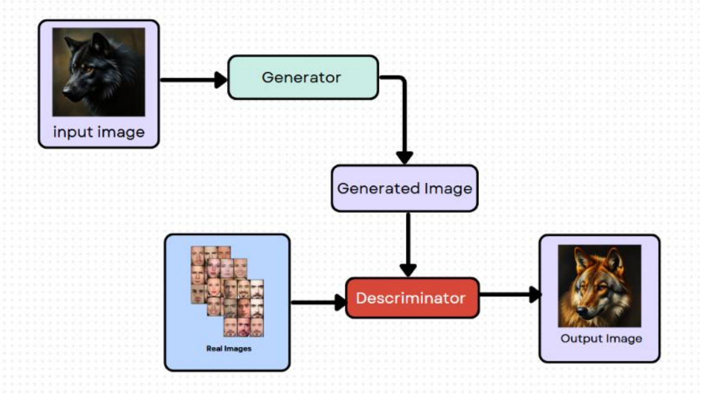

Fig.1: Outline of Image-to-image transfer system

2. LITERATURE REVIEW

1. Generative Adversarial Networks

The evolution of GANs (Generative Adversarial Networks) is a journey of deep learning that is redefining image generation. Introduced by Ian Goodfellow in 2014, GANs start with a simple yet influentialthought of pitting two neural networks (a generator and a discriminator) against each other to create real images from loud noises.

2.GAN-based Unsupervised Image Translation

Liu & Tuzel (2016) developed Coupled GANs for unsupervised image-to-image interpretation, aligning distributions between different domains.Huang & Belongie (2017) introduced a multimodal image-to-image translation method that learns mappings between domains without paired training data.Mejjati et al. (2018) proposed an attention-guided unsupervised image translation framework that focuses on salient regions for better feature adaptation.

3.Conditional GANs (Mirza & Osindero, 2014)

Conditional GANs extend the GAN framework by conditioning the generation process on additional information, such as class labels or input images. This technique is foundational for many image-to-image translation tasks, allowing for targeted transformations based on specific input conditions.

Table 1: Dataset Survey

DATASETUSED	PAPERTITLE
CelebA	● Generating Synthetic Images for Health care - 13 ● Hair Fast GAN: Realistic and Robust Hair Transfer - 1 ● A Study of State-of-the-Art GAN-Based Strategy - 16 ● High-Resolution Image Generation and Semantic Manipulation - 29 ● Geometry Structure Preserving Based GAN - 38 ● Training Transformers for High-Resolution Image Integration - 17 ● Style and Pose Control for Image Synthesis of Humans - 20 ● Spatial Fusion GAN for Image Synthesis - 14
COCO	● Multimodal Image Synthesis and Editing - 3 ● Stack GAN++: Realistic Image Synthesis - 10 ● Photo-Realistic Image Synthesis from Text Descriptions - 22 ● Scaling up GANs for Text-to-Image Synthesis - 14
FFHQ(Flicker-Faces-HQ)	● Efficient Hair Style Transfer with GANs - 4 ● Rethinking and Improving Robustness of Image Style Transfer - 6 ● Synthesis of Facial Image using Conditional GAN – 35
ImageNet	● Inverting Adversarially Robust Networks - 7 ● Generating Synthetic Images for Healthcare - 13 ● A Survey of State-of-the-Art GAN-Based Approaches - 16
CUB	● A Robust Pose Transformational GAN for Pose Guided - 21 ● Adversarial Text-to-Image Synthesis - 31 ● Photo-Realistic Image Synthesis from Text Descriptions - 22

4.CycleGAN (Zhu et al., 2017)

CycleGAN introduced a framework for unpaired image-to-image translation, allowing the transformation of images from one domain to another without requiring paired examples. The method employs two GANs to seek mappings between two fields, using cycle consistency loss to asure that the translated images can be converted back to their original form.

5. Pix2Pix (Isola et al., 2017)

Pix2Pix is a conditional GAN framework designed for paired image-to-image translation. It uses a U-Net architecture as the generator and a PatchGAN discriminator to ensure that the generated images are realistic at the patch level.

6. StyleGAN (Karras et al., 2019)

StyleGAN introduced a style-based generator architecture that allows for fine control over the generated images' styles at different levels of detail.

7. Hair Fast GAN (Nikolaev et al., 2021)

Hair Fast GAN specifically addresses the challenge of hair transfer in image synthesis. By utilizing a fast encoder-based approach, it enables efficient and realistic hair transfer between images, showcasing a specialized application of image-to-image translation techniques.

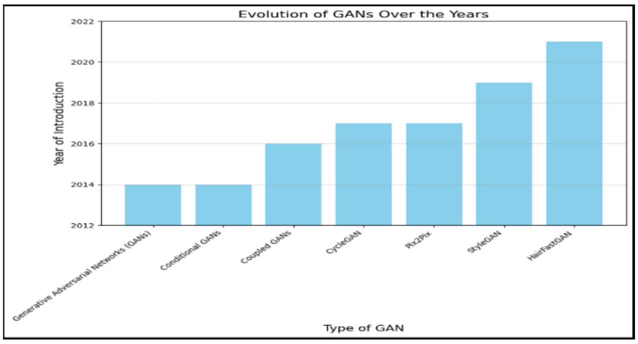

Fig.2 Evolution of GANS & Applications

A. FID Score

For cases involving generated images, the FID (Fréchet Inception Distance) score has become famous. FID evaluates the quality of a set of constructed images by comparing their distribution to that of real images. It analyzes how similar the generated image “distribution” is to the real one using high-level features extracted from a pre-trained neural network. This method is commonly used with GANs (Generative Adversarial Networks) and similar models to judge how realistic the generated images appear as a whole. Lower FID scores indicate that the generated photos are closer to real images in terms of their overall style and contentdistribution.

$$FID=|\mu_r-\mu_g|^2+Tr(\sum_r+\sum_g-2\sqrt{\sum_r\sum_g})$$

B. LPIPS Score

The LPIPS (Learned Perceptual Image Patch Similarity) score is a metric that mainly aims on perceptual similarity, aiming to measure how similar images "feel" from a human perspective. Unlike conventional pixel-based methods, LPIPS compares the high-level features of images by leveraging deep neural networks, meaning it can identify if two images have similar textures and structural characteristics. This score is particularly valued in image generation tasks (such as enhancing resolution or transforming images) where the "visual quality" perceived by humans is paramount. A lower LPIPS score indicates a closer match in perceptual terms, often corresponding well with human judgments of similarity.

$$\text{LPIPS(I,K)}=\frac{1}{L}\sum_{i=1}^{L} \mid \phi l(I)-\phi l(K) \mid^{2}$$

Table 1.Fid Score

Paper Title	FID Score	Methodology	Research Gap
HairFastGAN: Realistic and Robust Hair Transfer[1]	13.7	StyleGANFS,E4E Encoder	Slow optimization, pose misalignment, quality issues in fast methods
Multimodal Image Synthesis and Editing [3]	8.5	GANs, Diffusion Models, NeRF	Computational complexity, alignment challenges across modalities
Efficient Hair Style Transfer with GANs [4]	12.1	GAN, AdaIN	High computational costs, real-time limitations
Inverting Adversarially Robust Networks [7]	15.3	Adversarially Robust Encoder, GAN Inversion	High computational cost, complexity in feature inversion
A Survey of Image Synthesis Methods [9]	11.4	GANs, VAEs, CG-based Methods	Balancing synthetic data realism, domain fidelity
End-to-End Learning for HDR Image Synthesis [12]	7.9	Multi-ExposureHDR, Recurrent Networks	Ghosting in HDR, correlation complexity between HDR stacks and images
Generating Synthetic Images for Healthcare [13]	9.8	Pix2Pix, CycleGAN, StyleGAN	GAN instability, mode collapse, diversity issues
SpatialFusionGANfor Image Synthesis [14]	11.0	Spatial Fusion GAN(SF- GAN)	Limited exploration in complex scenes, challenges with high-resolution outputs

Table 2. LPIPS Score

Paper Title	LPIPS Score	Methodology	Research Gap
HairFastGAN: Realistic And Robust Hair Transfer[1]	0.23	StyleGANFS,E4E Encoder	Slow optimization, pose misalignment, quality issues in fast methods
Multimodal Image Synthesisand Editing [3]	0.19	GANs, Diffusion Models, NeRF	Computational complexity, alignment challenges across modalities
Efficient Hair Style Transfer with GANs [4]	0.21	GAN, AdaIN	High computational costs, real-time limitations
Inverting Adversarially Robust Networks [7]	0.28	Adversarially Robust Encoder, GAN Inversion	High computational cost, complexity in feature inversion
A Survey of Image Synthesis Methods [9]	0.25	GANs,VAEs,CG-based Methods	Balancing synthetic data realism, domain fidelity

End-to-End Learning for HDR Image Synthesis [12]	0.16	Multi-Exposure HDR, Recurrent Networks	Ghosting in HDR, correlation complexity between HDR stacks and images
Generating Synthetic Images for Healthcare [13]	0.20	Pix2Pix,CycleGAN,StyleGAN	GAN instability, mode collapse, diversity issues
Spatial Fusion GANfor Image Synthesis [14]	0.22	Spatial Fusion GAN(SF-GAN)	Limited exploration in complex scenes, challenges with high-resolution outputs

C. PSNR *Score*

The PSNR (Peak Signal-to-Noise Ratio) score provides a simple pixel-based assessment by measuring the ratio between the maximum signal strength and the strength of corrupting noise. It is commonly used in applications like image compression, where higher PSNR indicates less distortion and greater fidelity to the original image. While PSNR and SSIM both offer valuable information for structural similarity, they are not as perceptually aligned with human judgment as LPIPS or FID, which consider more nuanced and complex features in their calculations.

$$\text{PSNR}=10\cdot\log_{10}\left(\frac{R^2}{MSE}\right)$$

D. SSIM Score

The SSIM (Structural Similarity Index Measure) score assesses images based on three factors: luminance, contrast, and structure. This measure is particularly well- suited for tasks that need to preserve fine details and textures, such as image compression or transmission. SSIM values range from 0 to 1, where a score closer to 1 means the images are almost identical in structure. SSIM is therefore often used to maintain image quality when compressing files for storage or streaming because it focuses on preserving details that are critical to visual appeal, like edges and contrast.

$$\text{SSIM(I,K)}=\frac{(2\mu_I\mu_k+C_1)(2\sigma_{IK}+C_2)}{(\mu_I^2+\mu_K^2+C_1)(\sigma_I^2+\sigma_K^2+C_2)}$$

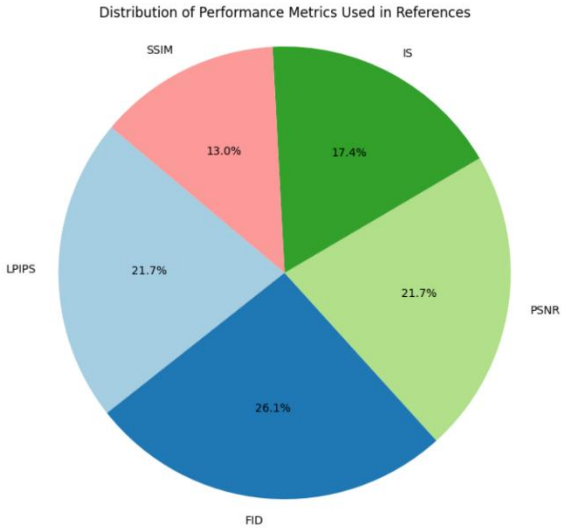

Fig 3: Distribution of performance metric used in papers

E. IS Score

The IS (Inception Score) is another important metric in image generation and evaluation, primarily used to assess the quality and diversity of images generated by AI models, especially GANs (Generative Adversarial Networks). The Inception Score measures two things: the quality of individual generated images and the variety across a set of generated images. To compute this score, each generated image is transpired through a pre-trained Inception network, which classifies the image and shows a probability distribution for various object classes. If an image has a clear, recognizable object, the probability distribution will be highly peaked around that class. If the generated images are diverse, the probability distributions for different images will vary. IS thus aims to achieve both sharp, high-quality images (indicating recognizable content) and a broad distribution over classes (indicating diversity in the generated.

$$\text{IS(G)}=\exp(E_{x\sim pg}[D_{KL}(p(y|x)||p(y))])$$

3. CONCLUSION

Techniques like HairFastGAN, StyleGAN, and StackGAN++ achieve high levels of realism, evident in favorable LPIPS and FID scores. These methods successfully create lifelike details crucial for applications in face and hair synthesis, ensuring perceptual quality and natural-looking results.

Approaches like Efficient Hair Style Transfer aim to balance image quality with reduced computational costs. While progress has been made toward real-time synthesis, these methods still grapple with the trade-off between quality and efficiency in resource-intensive tasks.

Limited dataset diversity, especially in specialized fields like healthcare, constrains the generalizability of models. While general datasets (e.g., COCO, ImageNet) are widely used, fields like medical imaging require richer, domain-specific data for more robust model performance.

Pose-transformational GANs and spatial fusion methods have advanced pose-aware synthesis, which is essential for generating adaptable human figures. However, maintaining structural consistency remains a challenge, especially in complex poses where alignment may falter.

4. REFERENCES

[1] Aljohani, A., & Alharbe, N. (2022). Generating Synthetic Images for Healthcare with Novel Deep Pix2Pix GAN. *Electronics*, *11*(21), 3470. https://doi.org/10.3390/electronics11213470

[2] Brock, A., Donahue, J., & Simonyan, K. (2018). Large scale GAN training for high fidelity natural image synthesis. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1809.11096

[3] Chen, Q., & Koltun, V. (2017). Photographic Image Synthesis with Cascaded Refinement Networks. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1707.09405

[4] Chen, X., Duan, Y., Houthooft, R., Schulman, J.,

Sutskever, I., & Abbeel, P. (2016). InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial NETS. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1606.03657

[5] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2010.11929

[6] Dosovitskiy, A., & Brox, T. (2016). Generating Images with Perceptual Similarity Metrics based on Deep Networks. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1602.02644

[7] Esser, P., Rombach, R., & Ommer, B. (2020). Taming Transformers for High-Resolution Image Synthesis. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2012.09841

[8] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial networks. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1406.2661

[9] Huang, X., & Belongie, S. J. (2017). Arbitrary Style Transfer in Real-time with Adaptive Instance Normalization. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1703.06868

[10] Isola, P., Zhu, J., Zhou, T., & Efros, A. A. (2016). Image-to-Image Translation with Conditional Adversarial Networks. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1611.07004

[11] Karmakar, A., & Mishra, D. (2020). A robust pose Transformational GAN for pose guided person image synthesis. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2001.01259

[12] Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2017). Progressive growing of GANs for improved quality, stability, and variation. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1710.10196

[13] Karras, T., Laine, S., & Aila, T. (2018). A Style-Based generator architecture for generative adversarial networks. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1812.04948

[14] Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., & Aila, T. (2019). Analyzing and improving the image quality of StyleGAN. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1912.04958

[15] Kim, J., Lee, S., & Kang, S. (2021). End-to-End differentiable learning to HDR image synthesis for multi-exposure images. *Proceedings of the AAAI Conference on Artificial Intelligence*, *35*(2), 1780–1788. https://doi.org/10.1609/aaai.v35i2.16272

[16] Kingma, D. P., & Welling, M. (2013). Auto-Encoding variational Bayes. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1312.6114

[17] Li, B., Qi, X., Lukasiewicz, T., & Torr, P. H. S. (2019). Controllable Text-to-Image generation. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1909.07083

[18] Li, C., & Wand, M. (2016). Combining Markov random fields and convolutional neural networks for image synthesis. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1601.04589

[19] Liu, M., & Tuzel, O. (2016). Coupled generative adversarial networks. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1606.07536

[20] Lu, Z., Li, Z., Cao, J., He, R., & Sun, Z. (2017). Recent progress of face image synthesis. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1706.04717

[21] Mikołajczyk, A., & Grochowski, M. (2019). Style transfer-based image synthesis as an efficient regularization technique in deep learning. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1905.10974

[22] Mirza, M., & Osindero, S. (2014). Conditional generative adversarial Nets. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1411.1784

[23] Nie, D., Trullo, R., Lian, J., Wang, L., Petitjean, C., Ruan, S., Wang, Q., & Shen, D. (2018). Medical Image Synthesis with Deep Convolutional Adversarial Networks. *IEEE Transactions on Biomedical Engineering*, *65*(12), 2720–2730. https://doi.org/10.1109/tbme.2018.2814538

[24] Nikolaev, M., Kuznetsov, M., Vetrov, D., & Alanov, A. (2024). HairFastGAN: Realistic and Robust Hair Transfer with a Fast Encoder-Based Approach. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2404.01094

[25] Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., & Efros, A. A. (2016). Context Encoders: Feature learning by inpainting. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1604.07379

[26] Pektas, M., Gecer, B., & Ugur, A. (2022). Efficient Hair Style Transfer with Generative Adversarial Networks. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2210.12524

[27] Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1511.06434

[28] Rojas-Gomez, R. A., Yeh, R. A., Do, M. N., & Nguyen, A. (2021). Inverting adversarially robust networks for image synthesis. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2106.06927

[29] Sarkar, K., Golyanik, V., Liu, L., & Theobalt, C. (2021). Style and Pose Control for Image Synthesis of Humans from a Single Monocular View. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2102.11263

[30] Tsirikoglou, A., Eilertsen, G., & Unger, J. (2020a). A survey of image Synthesis Methods for Visual Machine Learning. *Computer Graphics Forum*, *39*(6), 426–451. https://doi.org/10.1111/cgf.14047

[31] Tsirikoglou, A., Eilertsen, G., & Unger, J. (2020b). A survey of image Synthesis Methods for Visual Machine Learning. *Computer Graphics Forum*, *39*(6), 426–451. https://doi.org/10.1111/cgf.14047

[32] Van Den Oord, A., Kalchbrenner, N., & Kavukcuoglu, K. (2016). Pixel recurrent neural networks. *arXiv*

(Cornell University). https://doi.org/10.48550/arxiv.1601.06759

[33] Vondrick, C., Pirsiavash, H., & Torralba, A. (2016). Generating Videos with Scene Dynamics. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1609.02612

[34] Wang, L., Chen, W., Yang, W., Bi, F., & Yu, F. R. (2020). A State-of-the-Art Review on image synthesis with generative adversarial networks. *IEEE Access*, *8*, 63514–63537. https://doi.org/10.1109/access.2020.2982224

[35] Wang, T., Liu, M., Zhu, J., Tao, A., Kautz, J., & Catanzaro, B. (2017). High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1711.11585

[36] Wang, X., & Gupta, A. (2016). Generative Image Modeling using Style and Structure Adversarial Networks. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1603.05631

[37] Xian, W., Sangkloy, P., Agrawal, V., Raj, A., Lu, J., Fang, C., Yu, F., & Hays, J. (2017). TextureGAN: Controlling Deep Image Synthesis with Texture Patches. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1706.02823

[38] Zhan, F., Yu, Y., Wu, R., Zhang, J., Lu, S., Liu, L., Kortylewski, A., Theobalt, C., & Xing, E. (2021). Multimodal image Synthesis and Editing: the Generative AI Era. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2112.13592

[39] Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X., Huang, X., & Metaxas, D. (2017). StackGAN++: Realistic Image Synthesis with Stacked Generative Adversarial Networks. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1710.10916

[40] Zhang, M., & Zheng, Y. (2018). Hair-GANs: Recovering 3D Hair Structure from a Single Image. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1811.06229

[41] Zhang, R., Isola, P., Efros, A. A., Shechtman, E., & Wang, O. (2018). The unreasonable effectiveness of deep features as a perceptual metric. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1801.03924

[42] Zhao, J. J., Mathieu, M., & LeCun, Y. (2016). Energy-based generative adversarial network. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1609.03126.

[43] Zhu, J., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1703.10593