AI-Powered Pencil Strokes: Face Sketch Synthesis Using CycleGAN and Style Transfer

Sighakolli Dheeraj Venkata Sai*, Simma Sathwik*, Solleti Venkata Dhiraj*, Peddi Deekshith*, Lekshmi C. R.*
*School of Artificial Intelligence, Amrita Vishwa Vidyapeetham, Coimbatore, India
Emails: {cb.sc.u4aie24050, cb.sc.u4aie24051, cb.sc.u4aie24052, cb.sc.u4aie24039}@cb.students.amrita.edu,
cr_lekshmi@cb.amrita.edu

***Abstract*—This paper introduces an AI-driven framework for face sketch synthesis using Cycle Generative Adversarial Networks (CycleGAN), enhanced with a style transfer module to improve the perceptual quality of generated sketches. Unlike conventional edge-detection or handcrafted filter-based techniques, our method leverages adversarial learning to capture intricate details, shading, and texture while maintaining structural coherence. By incorporating perceptual and edge-aware constraints into the objective function, the model enhances sketch quality. Qualitative analysis reveals progressive refinement over training epochs, while quantitative evaluation using the Structural Similarity Index (SSIM) confirms improved structural and perceptual similarity. The stable convergence of training loss further validates the model's effectiveness. Experimental results demonstrate that our approach generates highly realistic and visually appealing sketches, surpassing traditional methods in both fidelity and perceptual quality. Future work will focus on enhancing fine details and improving generalization across diverse datasets.**

***Index Terms*—face-to-sketch, generative adversarial networks, CycleGAN, adversarial learning, SSIM**

I. Introduction

The transformation of facial photographs into artistic pencil sketches has numerous applications in digital art, forensics, animation, and entertainment. Traditional methods for this task, such as edge-detection algorithms and handcrafted filters, often fail to capture the fine-grained textures, shading, and expressive details found in hand-drawn sketches. As a result, the generated sketches usually lack realism and aesthetic appeal.

With advancements in deep learning, Generative Adversarial Networks (GANs) have emerged as a powerful tool for image-to-image translation, enabling realistic transformations between visual domains. Among these, CycleGAN has demonstrated remarkable success in tasks such as style transfer, artistic rendering, and domain adaptation. However, direct applications of CycleGAN to photo-to-sketch transformation face challenges in preserving facial structures, delicate textures, and shading intricacies essential for high-quality sketch generation.

To address these challenges, we propose an AI-powered framework for face sketch synthesis using CycleGAN, integrated with a style transfer module to enhance the perceptual quality of generated sketches. CycleGAN is particularly effective in unpaired image translation tasks, allowing the model to learn complex mappings between face images and their corresponding sketches without requiring pixel-wise alignment. By leveraging adversarial training and cycle consistency loss, the model preserves essential structural details while generating realistic sketches.

To further refine the generated outputs, we incorporate a style transfer module that enhances sketch clarity and artistic consistency. The proposed approach is evaluated from both qualitative and quantitative perspectives, using the Structural Similarity Index (SSIM) to assess the structural integrity and perceptual quality of the generated sketches. Experimental results demonstrate that our method produces visually appealing and structurally coherent sketches, outperforming conventional approaches in terms of realism and fidelity.

II. Related work

Face photo-to-sketch transformation has gained significant attention in recent years due to its applications in forensics, digital art, and identity recognition [12].

Traditional image-to-sketch techniques have long been utilized to convert photographs into sketch-like representations. A notable method involves edge detection algorithms, such as the Canny Edge Detector [1], which identifies significant edges within an image to produce a sketch outline. However, this approach often results in sketches that lack the nuanced textures and shading found in hand-drawn art. Filter-based techniques, like the Difference of Gaussians (DoG) [9], apply Gaussian blurring to enhance edges, creating a sketch effect. Despite their simplicity, these methods can lead to oversimplified images that miss intricate details. Also, Wang et al [12] handled the face photo sketch synthesis problem to learn the relations among neighboring image patches. Handcrafted feature extraction methods, including the use of Gabor filters [15] and stroke-based rendering [3], have also been explored to capture specific orientations and frequencies in an image, aiming to simulate pencil strokes. Yet, these techniques may struggle with complex backgrounds and often require manual tuning to achieve desired results.

Several studies have employed GAN-based models for sketch synthesis. Isola et al [4] introduced conditional generative adversarial networks for image-to-image translation. Zhu

*Corresponding author: Lekshmi C. R. (Email: cr_lekshmi@cb.amrita.edu)

et al [16] introduced cycle consistency loss for perfect reconstruction. Zhang et al. [14] propose an end-to-end photo-sketch generation model based on a fully convolutional network. Similarly, Wang et al. [11] introduced an adversarial sketch-photo transformation network, improving identity preservation but still facing issues with sketch blurriness and loss of high-frequency details. To address these limitations, researchers have attempted to refine CycleGAN architectures. Fang et al. [2] proposed an Identity-Aware CycleGAN, incorporating a perceptual loss function to enhance facial details. However, the model still suffered from over-smoothing and loss of subtle textures. Another approach, HE-CycleGAN by Li et al., introduced high-frequency feature enhancement to reduce blurriness and preserve facial features, yet it remained computationally expensive and sensitive to training instability [7].

While these advancements have improved the quality of sketch synthesis, existing methods still face critical challenges such as loss of fine details, inconsistent line quality, and difficulty in handling diverse facial variations. The proposed work aims to address these gaps by enhancing CycleGAN with improved structural preservation, perceptual constraints, and fine-grained texture retention.

The outline of the rest of the paper is as follows. Section 3 explains the system description. The performance evaluation is explained in Section 4 followed by the analysis of results in Section 5. Finally, the paper is concluded in Section 6.

III. SYSTEM DESCRIPTION

The proposed system utilizes CycleGAN for face sketch synthesis, incorporating a style transfer module to enhance realism. With two generators and two discriminators, the model ensures unpaired image-to-image translation while maintaining structural consistency via cycle loss. Instance normalization stabilizes training, and the Adam optimizer enhances performance. The style transfer module further refines sketches, improving texture and stroke quality, resulting in high-fidelity, realistic outputs.

A. Cycle Generative Adversarial Network (CycleGAN)

Generative Adversarial Networks (GANs) consist of two competing networks: a generator G, which synthesizes images, and a discriminator D, which differentiates between real and generated images. Through adversarial training, G improves by generating more realistic images, while D learns to distinguish real from fake samples. However, conventional GANs often lack structured constraints, leading to unpredictable transformations.

CycleGAN [16] introduces cycle consistency loss, ensuring that an image mapped from one domain to another and back retains its original structure. The training process optimizes three key losses: adversarial loss, cycle consistency loss, and identity loss.

1) Adversarial Loss: The adversarial loss ensures that the generated images are indistinguishable from real images. It is formulated as:

$$\mathcal{L}_{\text{GAN}}(G, D) = \mathbb{E}_{x\sim P_{\text{data}}}[\log D(x)] + \mathbb{E}_{z\sim P_z}[\log(1 - D(G(z)))] \tag{1}$$

where D aims to correctly classify real and generated images, while G attempts to generate realistic outputs that deceive D.

2) Cycle Consistency Loss: To maintain structural integrity, an image x mapped from domain A to domain B and then back to A should closely resemble its original version: [16]

$$\begin{aligned} \mathcal{L}_{\text{cycle}}(G_A, G_B) = \mathbb{E}_{x\sim P_A}[\|G_B(G_A(x)) - x\|_1] \\ + \mathbb{E}_{y\sim P_B}[\|G_A(G_B(y)) - y\|_1] \end{aligned} \tag{2}$$

where G_A and G_B are the generators responsible for transformations between the two domains.

3) Identity Loss: [16] To prevent unnecessary changes when translating images that already belong to the target domain, identity loss is applied:

$$\begin{aligned} \mathcal{L}_{\text{identity}}(G_A, G_B) = \mathbb{E}_{y\sim P_B}[\|G_A(y) - y\|_1] \\ + \mathbb{E}_{x\sim P_A}[\|G_B(x) - x\|_1] \end{aligned} \tag{3}$$

The overall objective function combines these three losses: [16]

$$\begin{aligned} \mathcal{L}(G_A, G_B, D_A, D_B) = \mathcal{L}_{\text{GAN}}(G_A, D_B) + \mathcal{L}_{\text{GAN}}(G_B, D_A) \\ + \lambda\mathcal{L}_{\text{cycle}} + \gamma\mathcal{L}_{\text{identity}} \end{aligned} \tag{4}$$

where λ (typically 10) controls the importance of cycle consistency, and γ (typically 0.5λ) regulates identity preservation.

These loss functions ensure that CycleGAN produces realistic sketches while preserving the structural consistency of the input images.

B. CycleGAN Architecture

The proposed model consists of four key components: two generators and two discriminators. It is designed to process face images of size 256 × 256 x 1. The discriminators are responsible for distinguishing real face images from generated sketches and follow the PatchGAN architecture, where each prediction corresponds to a 70×70 patch of the input. The discriminator network consists of Convolutional-InstanceNorm-LeakyReLU layers, with Instance Normalization applied to standardize feature maps independently for each image rather than across a batch, unlike Batch Normalization. The discriminator processes 256 × 256 input face images and produces localized patch-based predictions. The least-squares loss (L2 loss) is used for optimization, with a weighting factor of 0.5.

The internal architecture of cycleGAN follows an encoder-decoder architecture with residual learning, utilizing nine ResNet blocks to refine image transformations. It takes a 256 × 256 face image as input and produces a generated sketch of the same size, with pixel values in the range [-1,1]. The network first downsamples the input to a bottleneck layer, processes it with ResNet blocks using skip connections,

Fig. 1. Sample face and corresponding sketch images in the dataset

TABLE I
ARCHITECTURE OF THE DISCRIMINATOR

Layer Output	Description
$1 \times 256 \times 256$	Input image
$64 \times 128 \times 128$	4×4 Convolution, 64 filters, stride = 2
$64 \times 128 \times 128$	Leaky ReLU (negative slope = 0.2)
$128 \times 64 \times 64$	4×4 Convolution, 128 filters, stride = 2
$128 \times 64 \times 64$	Instance normalization
$128 \times 64 \times 64$	Leaky ReLU (negative slope = 0.2)
$256 \times 32 \times 32$	4×4 Convolution, 256 filters, stride = 2
$256 \times 32 \times 32$	Instance normalization
$256 \times 32 \times 32$	Leaky ReLU (negative slope = 0.2)
$512 \times 31 \times 31$	4×4 Convolution, 512 filters, stride = 1
$1 \times 4 \times 4$	Final convolution layer with stride = 1

and then upsamples back to the original resolution. 3×3 convolution filters with a stride of 1×1 are used within ResNet blocks. The output from each block is concatenated channel-wise with the input to maintain spatial consistency. The architectural details of the generator and discriminator are summarized in Tables I and II.

TABLE II
ARCHITECTURE OF THE GENERATOR

Layer Output	Description
$1 \times 256 \times 256$	Input image
$64 \times 256 \times 256$	7×7 Convolution, 64 filters, stride = 1
$64 \times 256 \times 256$	Instance normalization
$64 \times 256 \times 256$	ReLU activation
$128 \times 128 \times 128$	3×3 Convolution, 128 filters, stride = 2
$128 \times 128 \times 128$	Instance normalization
$128 \times 128 \times 128$	ReLU activation
$256 \times 64 \times 64$	3×3 Convolution, 256 filters, stride = 2
$256 \times 64 \times 64$	Instance normalization
$256 \times 64 \times 64$	ReLU activation
$256 \times 64 \times 64$	9 ResNet blocks, each with 256 filters
$128 \times 128 \times 128$	3×3 Convolution, 128 filters, stride = 2
$128 \times 128 \times 128$	Instance normalization
$128 \times 128 \times 128$	ReLU activation
$64 \times 256 \times 256$	3×3 Convolution, 64 filters, stride = 1
$64 \times 256 \times 256$	Instance normalization
$64 \times 256 \times 256$	ReLU activation
$1 \times 256 \times 256$	7×7 Convolution, stride = 1
$1 \times 256 \times 256$	Instance normalization
$1 \times 256 \times 256$	Tanh activation

C. Post Processing using Style Transfer Module

In the proposed enhancement process, the CycleGAN-generated sketch undergoes further refinement using a style transfer [10] module, where the ground truth sketch serves as the reference style. This approach improves the intelligibility and artistic quality of the synthesized sketches. The process begins by extracting feature representations from both the generated sketch and the ground truth using a pre-trained deep neural network, such as VGG-19 [5]. The style transfer algorithm then optimizes a loss function that balances content preservation and style adaptation. The content loss ensures that the structural details of the generated sketch remain intact, while the style loss encourages the incorporation of textures and patterns from the ground truth. By iteratively refining the sketch, the model enhances clarity, improves stroke consistency, and produces visually appealing results that closely resemble real hand-drawn sketches. In the proposed enhancement process, the CycleGAN-generated sketch undergoes further refinement using a style transfer module, where the ground truth sketch serves as the reference style. This approach improves the intelligibility and artistic quality of the synthesized sketches. The process begins by extracting feature representations from both the generated sketch and the ground truth using a pre-trained deep neural network, such as VGG-19. The style transfer algorithm then optimizes a loss function that balances content preservation and style adaptation. The content loss ensures that the structural details of the generated sketch remain intact, while the style loss encourages the incorporation of textures and patterns from the ground truth. By iteratively refining the sketch, the model enhances clarity, improves stroke consistency, and produces visually appealing results that closely resemble

D. Experimental setup

IV. PERFORMANCE EVALUATION

A. Data Collection and Preprocessing

A diverse and high-quality dataset is essential for training a CycleGAN model for image-to-sketch conversion. Since CycleGAN does not require paired data, we collect separate datasets for real images and sketches. The real images come from CelebA (faces) [8] and the sketch dataset includes hand-drawn sketches (Chinese University of Hong Kong (CUHK), Face Sketch Database (CUFS) [12] and synthetic sketches generated using Canny, Sobel, and Laplacian edge detectors. A sample of the face and corresponding sketch images from the dataset are shown in Figure1

Before training, we preprocess the images by resizing them to 256×256 and normalizing pixel values between -1 and 1 using:

$$X_{\text{normalized}} = \frac{X - 127.5}{127.5} \tag{5}$$

We remove duplicate and corrupted images using image hashing and OpenCV [1]. Additionally, data augmentation techniques such as random rotations, flipping, brightness adjustments, and jittering are applied to increase dataset diversity. These preprocessing steps ensure a clean and varied dataset, enhancing the CycleGAN model's ability to generate realistic and detailed pencil sketches.

B. Experimental setup

The CycleGAN model is trained using a combination of adversarial loss, identity loss, and cycle consistency losses [16]. The cycle consistency losses are weighted by a parameter $\lambda = 10$, while the identity loss is assigned a weight of 0.5λ to preserve content structure during translation.

The generator and discriminator are optimized using the Adam optimizer with an initial learning rate of $2e^{-4}$. The model is trained for 1000 iterations with a batch size of 1. The learning rate remains constant for the first 100 iterations and then linearly decays to $2e^{-6}$ over the remaining iterations. Table III refers to the various hyperparameters used for our model.

TABLE III
HYPERPARAMETERS USED IN THE CYCLEGAN MODEL

Hyperparameter	**Value**
Optimizer	Adam
Learning Rate	2×10^{-4}
Batch Size	1
Weight for Cycle Consistency Loss (λ)	10
Weight for Identity Loss	0.5λ
Beta1 (Adam)	0.5
Beta2 (Adam)	0.999
Image Size	256×256

V. RESULTS AND ANALYSIS

This section evaluates the performance of the proposed model from both qualitative and quantitative perspectives, focusing on the quality and accuracy of the synthesized sketches.

A. Qualitative Analysis

The qualitative evaluation of the CycleGAN model is performed by visually analyzing the progression of generated sketches over different training epochs. Figure 2 illustrates how the model refines its output over time, showing sample sketches generated at 100, 500, and 1000 iterations.

At early training stages by 100 iterations, the sketches lack structure, often appearing incomplete with distorted edges. By 500 iterations, the model starts capturing finer details, improving overall structure, and reducing artifacts. At 1000 iterations, the generated sketches exhibit enhanced clarity, well-defined edges, and improved realism, closely resembling hand-drawn pencil sketches. This progression demonstrates the model's learning process and its ability to refine features over time. The CycleGAN-generated outputs are passed through our post-processing style transfer module to obtain the final sketch output.

B. Quantitative Analysis

To objectively assess the quality of the generated sketches, we employ the standard evaluation metrics: Structural Similarity Index (SSIM) [13]. These metrics quantify the structural preservation and overall quality of the generated images compared to ground truth sketches.

1) Structural Similarity Index (SSIM): [13] SSIM measures the perceptual similarity between the generated sketch and the reference image by evaluating luminance, contrast, and structural integrity. The SSIM formula is given as:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{6}$$

[6]

where μ_x and μ_y represent the mean intensities, σ_x^2 and σ_y^2 denote the variances, and σ_{xy} indicates the covariance between images x and y. Constants C_1 and C_2 prevent numerical instability.

TABLE IV
COMPARISON OF SSIM VALUE BEFORE AND AFTER APPLYING THE STYLE TRANSFER ENHANCEMENT.

Metric	**Without Enhancement**	**After Enhancement**
SSIM	0.52	0.56

Our model achieved an SSIM of 0.56 indicating a fair structural similarity between the generated and reference sketches and moderate feature retention. While this value suggests that the the model preserves essential image structures, there is room for improvement in capturing finer details and enhancing edge definition

C. Training Loss Analysis

The training loss curve of CycleGAN, illustrated in Figure 3, provides insights into the learning process of the model. It presents the loss values of the generator (G) and the two discriminators (dA and dB) over multiple iterations.

[1] https://opencv.org/

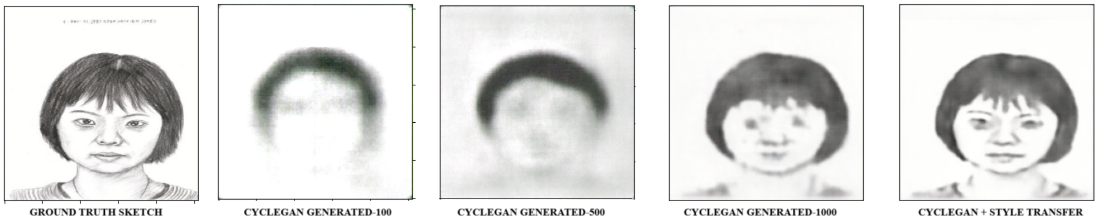

Fig. 2. Progressive refinement of generated sketches from 100 to 1000 iterations. The model gradually enhances structural details and improves sketch quality.

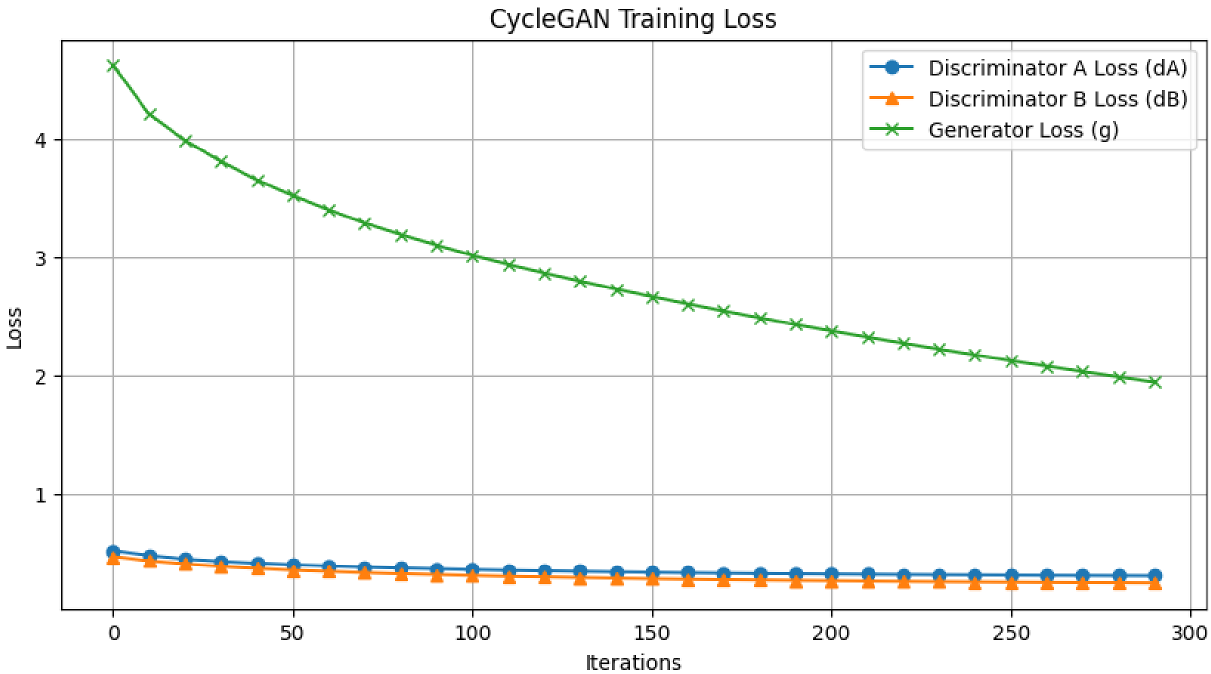

Fig. 3. CycleGAN training loss curves: Discriminator A Loss (blue), Discriminator B Loss (orange), and Generator Loss (green). The generator loss steadily decreases, indicating progressive learning.

1) Generator Loss (G): The generator loss (green curve) is initially high, indicating that the model struggles to generate convincing sketches early in training. As training progresses, the loss steadily declines, suggesting that the generator improves its ability to create sketches that closely resemble real ones. The loss does not reach zero due to the adversarial nature of the training, where the generator continuously refines its outputs to deceive the discriminators.

2) Discriminator Loss (d_A and d_B): The discriminator losses for dA (blue) and dB (orange) remain relatively low and stable, indicating that the discriminators effectively distinguish between real and generated sketches. The absence of sharp fluctuations suggests that the model maintains a balanced adversarial training dynamic, preventing either the generator or the discriminators from dominating the learning process.

3) Convergence and Model Stability: The observed loss trends indicate stable training without significant oscillations, which is crucial for ensuring convergence in adversarial models. The declining generator loss and the steady behavior of the discriminator losses confirm that the model effectively learns to transform images into sketches while maintaining structural consistency. This demonstrates the successful adaptation of CycleGAN in synthesizing high-quality sketches.

VI. Conclusion

This work presented an AI-driven approach for generating high-quality pencil sketches using CycleGAN with a style transfer enhancement module. Qualitative analysis showed progressive sketch refinement across training epochs, while quantitative evaluation demonstrated improvements in SSIM, confirming enhanced structural and perceptual quality. The stable training loss curves validated the model's robustness. Future work can explore advanced architectures for improved realism and generalization.

References

[1] John Canny. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6):679–698, 1986.

[2] Yuke Fang, Weihong Deng, Junping Du, and Jiani Hu. Identity-aware cyclegan for face photo-sketch synthesis and recognition. *Pattern Recognition*, 102:107249, 2020.

[3] Aaron Hertzmann. Stroke-based rendering. *Recent Advances in Npr for Art & Visualization*, 1, 2009.

[4] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.

[5] S Kavitha, B Dhanapriya, G Naveen Vignesh, and KR Baskaran. Neural style transfer using vgg19 and alexnet. In *2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA)*, pages 1–6. IEEE, 2021.

[6] Seungyeol Lee, Seongwoo Hong, Gwangyeol Kim, and Jaecheol Ha. Ssim-based autoencoder modeling to defeat adversarial patch attacks. *Sensors (Basel, Switzerland)*, 24(19):6461, 2024.

[7] Bin Li, Ruiqi Du, Jie Li, and Yuekai Tang. He-cyclegan: A symmetric network based on high-frequency features and edge constraints used to convert facial sketches to images. *Symmetry*, 16(8):1015, 2024.

[8] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of the IEEE international conference on computer vision*, pages 3730–3738, 2015.

[9] Ph Marthon, B Thiesse, and A Bruel. Edge detection by differences of gaussians. In *Computer Vision for Robots*, volume 595, pages 318–327. SPIE, 1986.

[10] Sihan Ren and Yiwei Sheng. Image style transfer using deep learning methods. In *2022 IEEE International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA)*, pages 1190–1195. IEEE, 2022.

[11] Lidan Wang, Vishwanath Sindagi, and Vishal Patel. High-quality facial photo-sketch synthesis using multi-adversarial networks. In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*, pages 83–90. IEEE, 2018.

[12] Xiaogang Wang and Xiaoou Tang. Face photo-sketch synthesis and recognition. *IEEE transactions on pattern analysis and machine intelligence*, 31(11):1955–1967, 2008.

[13] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[14] Liliang Zhang, Liang Lin, Xian Wu, Shengyong Ding, and Lei Zhang. End-to-end photo-sketch generation via fully convolutional representation learning. In *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, pages 627–634, 2015.

[15] Xianlin Zhang, Xueming Li, Shuxin Ouyang, and Yang Liu. Photo-to-sketch transformation in a complex background. *IEEE Access*, 5:8727–8735, 2017.

[16] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.