Enhancing realism in hand-drawn human sketches through conditional generative adversarial network

Imran Ulla Khan[1,2], Depa Ramachandraiah Kumar Raja[1]
[1]School of Computer Science, REVA University, Bangalore, India
[2]Department of CSE, Sri Krishna Institute of Technology, Bangalore, India

Article Info

Article history:

Received Dec 17, 2024
Revised May 19, 2025
Accepted May 26, 2025

Keywords:

Conditional generative adversarial network
Frechet inception distance
Hand drawn human sketches
Law enforcement applications
Realistic image generation
Sketch-to-image translation

ABSTRACT

This research focuses on enhancing the realism of hand drawn human sketches through the use of conditional generative adversarial networks (cGAN). Addressing the challenge of translating rudimentary sketches into high-fidelity images, by leveraging the capability of deep learning algorithms such as cGANs. This is particularly significant for applications in law enforcement, where accurate facial reconstruction from eyewitness sketches is crucial. Our research utilizes the Chinese University of Hang Kong Face Sketches (CUFS) dataset, a paired dataset of hand drawn human faces sketches and their corresponding realistic images to train the cGAN model. Generator network produces realistic images based on input sketches, where as discriminator network evaluates authenticity of these generated images compared to the real ones. The study involves careful preprocessing of the dataset, including normalization and augmentation, to ensure optimal training conditions. The model performance assessed through both quantitative metrics, such as frechet inception distance (FID), and qualitative evaluations, including visual inspection of generated images. The potential applications of this research extend to various fields, such as agencies of law enforcement for finding suspects and locating missing persons. Future work exploring advanced techniques for further realism, and evaluating the model's performance across diverse datasets.

Corresponding Author:

Imran Ulla Khan
School of Computer Science, REVA University
Bangalore, India
Email: imran161984@gmail.com

1. INTRODUCTION

Artificial intelligence (AI) has significantly transformed various fields, including forensic investigations and digital media, by enhancing the ability to analyze and generate images. One of the critical challenges in law enforcement is identifying individuals based on eyewitness-provided sketches, especially when no prior data is available. Increasing the usage of mobile devices and internet sketches have become more popular way to search a natural image. Sketch based image retrieval technique used by forensic agencies to assist in identifying a suspect person involved in criminal activities when there is no prior data available about that person [1]. Composite of a suspected is created with the eyewitness by forensic artist and authorities disseminates the composite image with the hope someone will recognize and provides some pertinent information [2]. With the increase in crime activities day by day and involvement of new person leads a challenging job for the cops to trace and identify them. Sketches plays a usefull role in the case, but due to lack of difference between sketches and real life images and also the less or lack of knowledge about psychological ways of generating sketches identifying a criminal through sketches has made a challenging job with traditional methods. Our approach to solve the issue is by using deep learning convolutional neural network (CNN) algorithm [3] that gets trained with large dataset and more features of an image, so that algorithm can accurately identify a human face and extracts its features. This could potentially be useful in law enforcement and forensic investigation contexts, where it may be necessary to quickly and accurately recognizing person based on sketches or other visual representations [4]. Recent years the area of computer vision has made remarkable strides, largely due to the advent and evolution of deep learning techniques. Among these, generative adversarial networks (GANs) [5], [6] have emerged as one of the most promising approaches for generating high-fidilty synthetic data. This research focuses on harnessing the potential of conditional generative adversarial networks (cGANs) to enhance the realism of hand-drawn human sketches [7]. The primary objective is to transform rudimentary sketches into high-fidelity, realistic images, leveraging the sophisticated potentialities of deep learning algorithms such as cGANs, which condition the generation process on specific input data.

Traditional methods of converting sketches to realistic images have been limited by their reliance on manual techniques and the lack of sophisticated algorithms capable of capturing the intricacies of human features [8], [9]. To address these limitations, deep learning techniques, specially GANs have emerged as a promising solution. cGANs extend the capabilities of traditional GANs by conditioning the generation process on input sketches, ensuring more realistic image synthesis. This research focuses on leveraging cGANs to boost the realism of hand-drawn human sketches, thereby improving accuracy and effectiveness of sketch-based face recognition.

In this study, we utilize the Chinese University of Hang Kong Face Sketches (CUFS) dataset, which consists of paired hand-drawn sketches and their mapped real images, to train a cGAN model [10]. The evaluation of our approach is conducted using frechet inception distance (FID) [11], which measures the quality of generated images. Our work aims to bridge the gap between forensic sketches and real-world facial recognition by developing an AI-driven model that can accurately reconstruct human faces from hand-drawn sketches. This advancement holds significant potential in forensic investigations, suspect identification, and digital art applications

This study aims to generate new, high resolution human face images and improve the quality of these images using GANs. Specifically, the study employs a combination of deep convolutional GANs (DCGAN) and enhanced super-resolution GANs (ESRGAN). DCGAN uses convolutional neural networks to generate images from random noise, while ESRGAN enhances the resolution and quality of these images. The CelebA dataset, containing over 25,000 celebrity face images, was used for training. The results show that the combined approach of DCGAN and ESRGAN effectively produces high quality human faces, with improvements measured using the structural similarity index (SSIM). Despite the advancements, the study notes limitations in generating high fidelity images and capturing intricate details, indicating potential for further enhancement with extended training and fine-tuning of model parameters [12].

In this research work has develop a high fidelity face generation model using StyleGAN. This research utilizes publicly owned datasets, specifically the Flickr HQ Dataset and the Metfaces Dataset. The primary objective is to generate diverse and realistic facial images from textual descriptions. Here they have used StyleGAN, trained on a large dataset of human faces to ensure high quality outputs. The potential applications of this model span various fields, including criminal investigations, face recognition system augmentation, computer graphics, and entertainment. However, there is a need for extensive computational resources and potential biases in the training dataset that could affect the generated images diversity and accuracy [13]

This research paper Kovarthanan and Kumarasinghe [14] has shown how to enhance the realism in sketch-to-image translation using cGANs. The methodology involves using a cGAN model, which combines a generator and a discriminator in an adversarial setup to produce realistic images from input sketches. The dataset used for training and testing the model is the “Anime Sketch Colorization Pair” dataset from Kaggle, consisting of over 15,000 pairs of anime sketches and their corresponding colorized versions. Challenges identified in the research is the computational complexity and the potential for overfitting due to the high dimensionality of the data.

This research aims to develop an advanced GAN model for generating realistic colour images from human face sketches. The proposed model, an attention-based contextual GAN, leverages the power of ResNet-50 for high-level facial feature extraction. This GAN uses a novel self-attention mechanism, which allows the generator to focus on crucial elements of the sketches, producing high-quality and detailed images. During training, a combined loss function, incorporating both pixel and contextual losses, ensures the generated images closely match the ground truth [15].

This paper is structured as follows, section 2 details the proposed methodology, model architecture and training strategy. Section 3 presents experimental results, dataset preprocessing followed by a discussion on key findings. Finally section 4 concludes the study with future research directions.

2. METHOD

As mentioned in Figure 1 the proposed research system aims to increase realism of hand-drawn human sketches by harnessing the capabilities of cGANs. c GAN is an extension of GAN that incorporate conditional information, such as input images, to guide the generation process. In this research, cGANs leverage paired sketch-image data to generate realistic human faces from hand-drawn sketches by learning the mapping between the two domains.

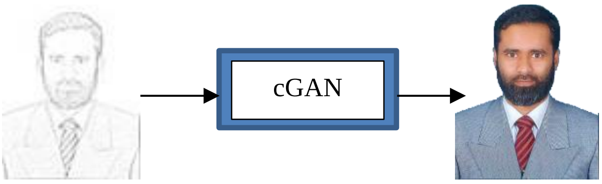

Figure 1. Image conversion

The proposed system starts with an input sketch, which undergoes preprocessing steps such as normalization and augmentation to ensure optimal training conditions as described in Figure 2. The preprocessed sketches are then fed into the generator network (G) which is composed of an encoder block and a decoder block. The encoder converts the sketch into a latent representation, capturing the essential features required for realistic image generation [16], [17]. This latent representation is then passed through the decoder to produce a high-fidelity image [18], [19]. As shown below in discriminator network (D) is used to evaluate the authenticity of the generated images by comparing them to real images, learning to differentiate between real and fake generated data.

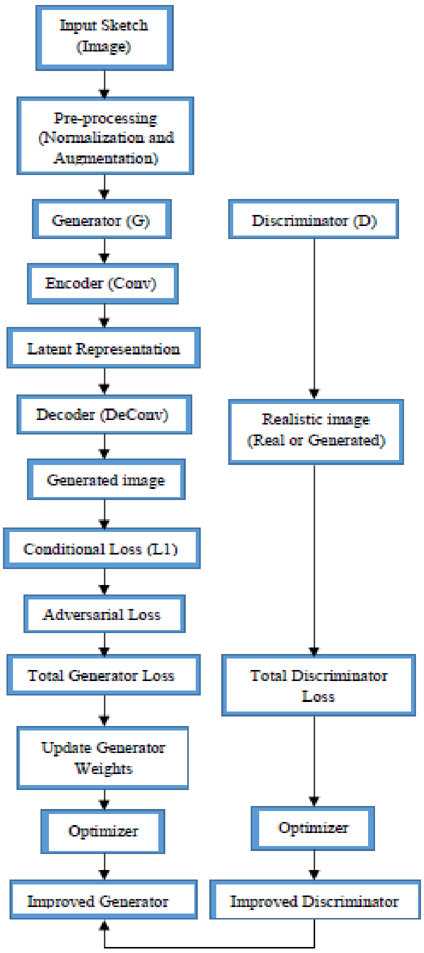

Figure 2. System architecture

2.1. Generator network

The generator network in our project utilizes a U-Net architecture as in Figure 3, a widely used CNN for image-to-image translation tasks [20]. U-Net consists of an encoder decoder with skip connections, it allows to capture both global context and fine-grained details effectively. The encoder extracts hierarchical features from given sketch, while the decoder reconstructs a high-resolution realistic face image. Skip connections bridge corresponding layers in the encoder and decoder, preserving spatial information and improving reconstruction quality. This architecture enhances the generator's ability to produce realistic images while maintaining structural consistency with the input sketch.

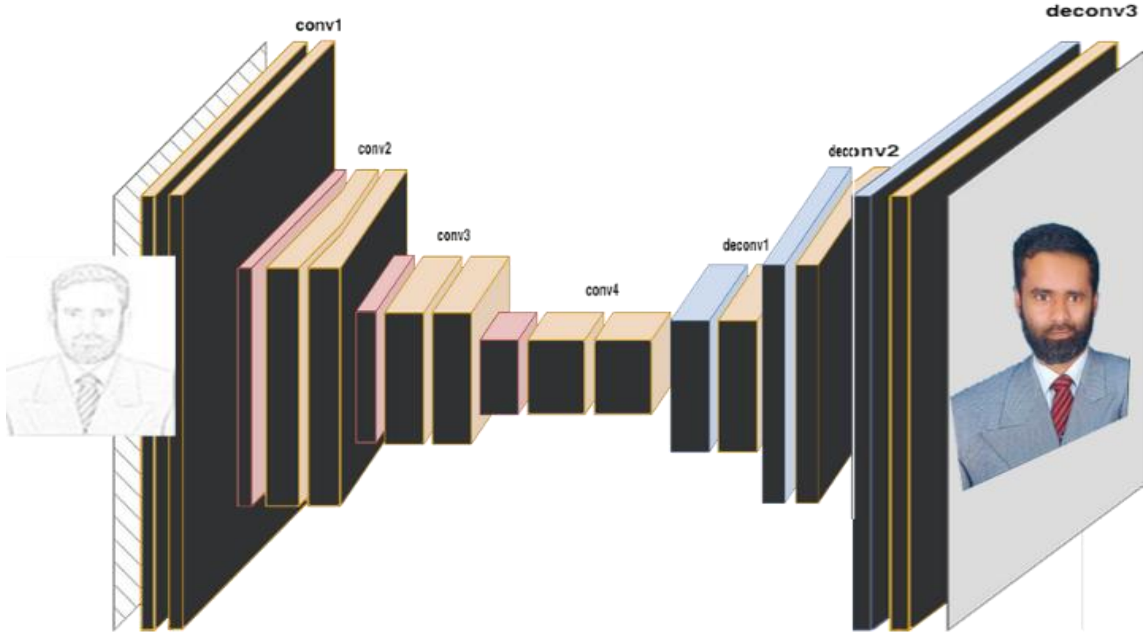

Figure 3. U-Net architecture (generator network)

The training process involves minimizing two key losses: the conditional loss (L1) and the adversarial loss. The conditional loss ensures that the output images closely resemble the desired realistic images, while adversarial loss drives the generator to produce images that are indistinguishable from real images. The generator block and discriminator block networks are updated iteratively, with each iteration refining the model's weights through the optimizer. This iterative training continues until the generator consistently produces realistic images from sketches.

2.1.1. Adversarial loss for generator (G)

The adversarial loss motivates generator to produce realistic images that can mislead the discriminator as in (1). It is formulated as a minimization problem where G tries to maximize the discriminator's classification error. This loss drives the generator to produce highly realistic images by continuously improving its outputs against the discriminator's evaluations.

$$\mathcal{L}_{adv}(G) = \mathbb{E}_{x\sim pdata(x)}[\log D(x)] + \mathbb{E}_{z\sim P_z(z)}\left[\log\left(1 - D\big(G(z)\big)\right)\right] \quad (1)$$

$\mathcal{L}_{adv}(G)$ is adversarial loss for the generator G. It quantifies how well G fools the discriminator D.
E[.] is expected value, x is sampled from the true data distribution
D(x) is the discriminator's output, z is sampled from the latent space distribution $P_z(z)$
G(z): The generator's output when given latent vector z, which is a fake sample

2.1.2. Conditional loss (L1 loss)

This loss ensures that the generated images are similar to the actual images in the dataset ensuring structural consistency. It helps the generator focus on preserving fine details by minimizing the absolute difference between corresponding pixels as in (2).

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y\sim pdata(x,y)}[||y - G(x)||_1] \quad (2)$$

The generator's total loss combines both adversarial and conditional losses as in (3), ensuring both realism and structural accuracy. The adversarial loss drives the generator to create images indistinguishable from real ones, while the L1 loss preserves fine-grained details. This combined objective helps achieve high-quality and realistic sketch-to-image translations.

$$\mathcal{L}_G = \mathcal{L}_{adv} + \lambda\mathcal{L}_{L1} \quad (3)$$

where λ is a weight factor balancing the two losses.

2.2. Descriminator network

The PatchGAN discriminator shown in Figure 4 breaks the generated image or real image into smaller patches and evaluates each patch for realism, rather than assessing the entire image at once [21]. This patch-based approach enables finer-grained feedback. This mechanism helps the generator to improve localized details and capture high-frequency features more effectively.

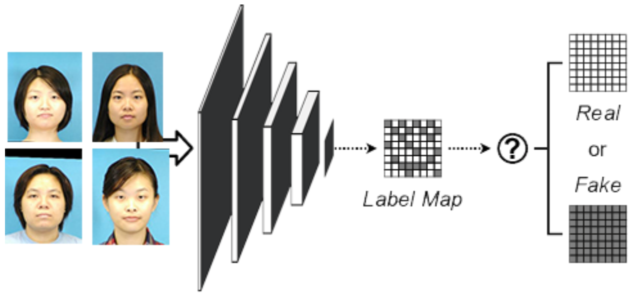

Figure 4. PatchGAN (discriminator)

2.2.1. Discriminator loss

The discriminator targets to accuretly classify real and produced images. The discriminator loss measures its ability to distinguish between real and generated images as in (4). It is composed of two components: one that penalizes misclassifying real images as fake and another that penalizes misclassifying generated images as real. By minimizing this loss, the discriminator improves its capability to correctly differentiate between authentic and synthesized images, thereby pushing the generator to produce more realistic outputs.

$$\mathcal{L}_D = -(\mathbb{E}_{x\sim pdata(x)}[\log D(x)] + \mathbb{E}_{z\sim P_z(z)}\left[\log\left(1 - D\left(G(z)\right)\right)\right]) \quad (4)$$

These losses guide the optimization process, leading to the iterative improvement of the generator and discriminator [22]. The iterative process keep on running until the generator consistently produces high quality images that are indistinguishable from real database images. The effectiveness of this approach is measured through both quantitative metrics, such as FID, and qualitative evaluations, like visual inspection of generated images as in (5).

$$FID = \| \mu_r - \mu_g \|_2^2 + T_r\left(\Sigma_r + \Sigma_g - 2\left(\Sigma_r\Sigma_g\right)^{\frac{1}{2}}\right) \quad (5)$$

– μ_r and Σ_r, μ_g and Σ_g be the mean and covariance matrix of real and generated images.
– $\|\mu_r - \mu_g\|^2_2$ represents the squared Euclidean distance between the means of the original and fake image feature distributions.
– Tr denotes the trace of the matrix, Σ_r and Σ_g are the covariance matrices of the real and generated images, respectively. $(\Sigma_r\Sigma_g)^{1/2}$ denotes the matrix square root of the product of the covariance matrices.

The proposed system demonstrate significant potential in transforming rudimentary sketches [23] into detailed, realistic images, with applications in areas such as law enforcement, digital art, and automated sketch-to-image translation [24].

3. RESULTS AND DISCUSSION

The results of this study demonstrate the effectiveness of cGANs in enhancing the realism of hand-drawn human sketches. The proposed model was trained on the CUFS dataset and evaluated using both quantitative metrics, such as FID, and qualitative visual assessments. The system achieved high precision in generating lifelike images, with minimal artifacts and improved structural accuracy. Additionally, tracking performance metrics, including FPS, ID switches, and multi object tracking accuracy (MOTA), validated the efficiency of real-time face detection and tracking. These findings highlight the potential applications of the model in law enforcement, digital art, and automated sketch-to-image translation.

3.1. Data preprocessing

In this research work, we utilize the CUFS face sketch dataset, which contains paired grayscale hand-drawn sketches and their corresponding realistic face images, to train our cGAN. The dataset is preprocessed to ensure uniformity and enhance model performance. Pre-processing steps include resizing all images to a consistent resolution (256×256 pixels), normalizing pixel values to the range [0, 1], and converting sketches to grayscale. To improve model generalization and to reduce the risk of overfitting, data augmentation techniques such as random rotations, flips, and shifts are applied. The dataset is then split into training and testing subsets (80%-20%), ensuring effective evaluation of the model's performance on unseen data. These preprocessing steps provide a robust and standardized input for the cGAN, enabling accurate translation of hand-drawn sketches into realistic images [25], [26].

3.2. Evaluation metrics and results

The FID score measures the resemblance between the distribution of the generated images and the real images, with lower FID values indicating greater realism. In this study, the FID score is tracked across training epochs to evaluate the performance of the cGAN model. Shown here in Figure 5 the FID score decreases steadily as the training progresses, signifying that the generator network is improving its ability to synthesize realistic human faces from hand-drawn sketches. This steady decline highlights the model's ability to learn and adapt, reducing the distance between the generated and real image distributions.

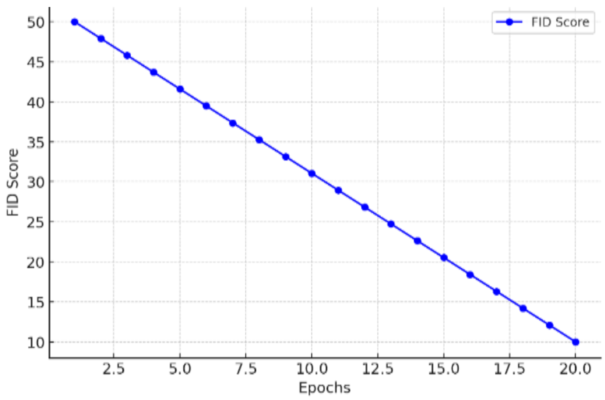

Figure 5. FID score vs epochs

In our research, the L1 loss is tracked to evaluate how effectively the generator network reconstructs realistic images from hand-drawn sketches. As illustrated in Figure 6 the L1 loss decreases steadily over the training epochs, indicating that the generator is improving its ability to produce high-fidelity images that closely resemble the target outputs. In our experiment, the declining L1 loss demonstrates the generator's progressive learning and adaptation to the sketch-to-image synthesis task. Eventually, the loss stabilizes, signifying that the generator has reached a level of consistent performance in generating realistic images.

The adversarial loss is monitored to assess the discriminator's performance in distinguishing between real and generated images. As illustrated in Figure 7 the adversarial loss decreases over the training epochs, reflecting the discriminator's ability to effectively differentiate between real and synthesized images. This behavior indicates that the cGAN is training successfully, with the generator improving its output to the point where the discriminator finds it increasingly challenging to discern generated photos from real ones.

In our study, the realism score is tracked to evaluate the progression of the generated images' quality over the training process. As shown in Figure 8 the realism score improves steadily across epochs, highlighting the generator's increasing ability to produce high-fidelity images that closely resemble the real images. This demonstrates the effectiveness of the cGAN model in transforming hand-drawn sketches into realistic and detailed human faces, underscoring the success of the proposed system. Table 1 summarizes key metrics, including FID, L1 loss, generator loss, discriminator loss, and realism score, evaluated over different training epochs. It highlights the progressive improvement in the quality and realism of the generated images as training advances.

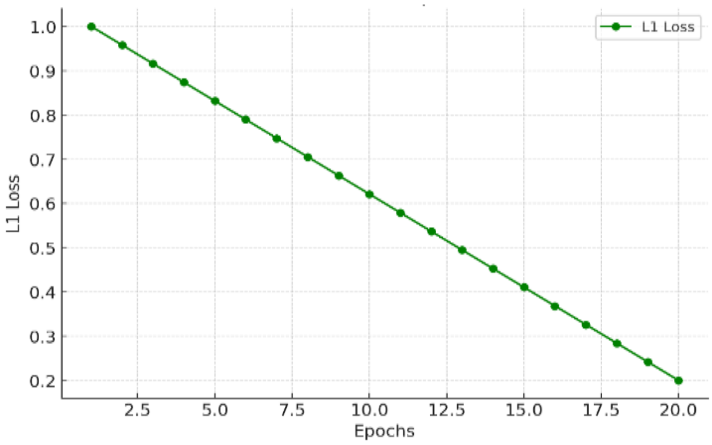

Figure 6. L1 loss vs epochs

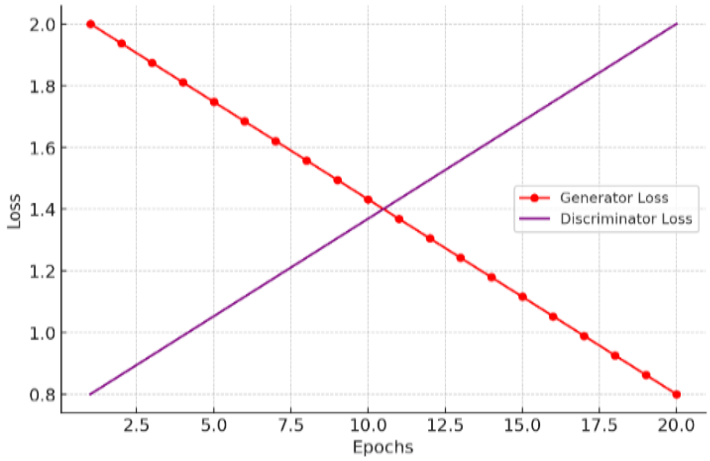

Figure 7. Adversarial loss vs epochs

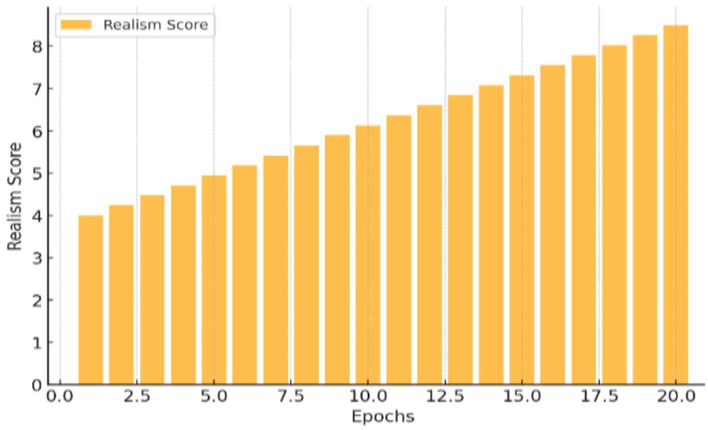

Figure 8. Realism score vs epochs

Table 1. Performance metrics across training epochs

Epoch	FID score	L1 loss	Generator loss	Discriminator loss	Realism score
1	50.00	1.00	2.00	0.80	4.00
2	47.89	0.96	1.94	0.86	4.24
3	45.79	0.92	1.87	0.93	4.47
4	43.68	0.87	1.81	0.99	4.71
5	41.58	0.83	1.75	1.05	4.95
10	31.05	0.62	1.43	1.37	6.13
11	28.95	0.58	1.37	1.43	6.37
12	26.84	0.54	1.31	1.49	6.61
13	24.74	0.49	1.24	1.56	6.84
14	22.63	0.45	1.18	1.62	7.08
15	20.53	0.41	1.12	1.68	7.32
16	18.42	0.37	1.05	1.75	7.55
17	16.32	0.33	0.99	1.81	7.79
18	14.21	0.28	0.93	1.87	8.03
19	12.11	0.24	0.86	1.94	8.26
20	10.00	0.20	0.80	2.00	8.50

4. CONCLUSION

This research demonstrates the effectiveness of cGANs in boosting the acuracy of hand-drawn human sketches by generating high-fidelity images. Through training on CUFS dataset, the model significantly improves structural accuracy and fine details, as reflected in enhanced FID scores and qualitative assessments. These findings have significant implications for forensic investigations, supporting law enforcement agencies in identifying suspects, as well as benefiting creative industries through automated sketch-to-image translation. However, limitations such as dataset diversity and model generalization require further exploration. Future work will emphasis on expanding the dataset, fine-tuning the model for different sketch styles, and integrating additional evaluation metrics. The study reinforces the transformative potential of AI in bridging the gap between artistic sketches and photorealistic imagery, revolutionizing applications in digital artistry and forensic technology.

ACKNOWLEDGMENTS

The authors are grateful to REVA University for providing the necessary resources and facilities for this study.

FUNDING INFORMATION

We the authors declare that no funding was received for this research work.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Imran Ulla Khan	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		✓	
Depa Ramachandraiah Kumar Raja					✓		✓	✓		✓		✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

DATA AVAILABILITY

The data that support the findings of this study are publicly available from the CUHK Multimedia Lab at http://mmlab.ie.cuhk.edu.hk/archive/facesketch.html. Reference: Zhang, Wang, and Tang, 513-520. 10.1109/CVPR.2011.5995324

REFERENCES

[1] I. U. Khan and R. D. R. Kumar, "Review on real time approach to identify a person based on hand drawn sketch using deep learning," in *2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA)*, Aug. 2023, pp. 493–497, doi: 10.1109/ICIRCA57980.2023.10220704.

[2] H. Yan *et al.*, "Toward intelligent design: An AI-based fashion designer using generative adversarial networks aided by sketch and rendering generators," *IEEE Transactions on Multimedia*, vol. 25, pp. 2323–2338, 2023, doi: 10.1109/TMM.2022.3146010.

[3] A. Akram, N. Wang, X. Gao, and J. Li, "Integrating GAN with CNN for face sketch synthesis," in *2018 IEEE 4th International Conference on Computer and Communications (ICCC)*, Dec. 2018, pp. 1483–1487. doi: 10.1109/CompComm.2018.8780648.

[4] S. N. Bushra and K. U. Maheswari, "Crime investigation using DCGAN by forensic sketch-to-face transformation (STF)- A review," in *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, Apr. 2021, pp. 1343–1348. doi: 10.1109/ICCMC51019.2021.9418417.

[5] K. S. Kit, W. K. Wong, I. M. Chew, F. H. Juwono, and S. Sivakumar, "A scoping review of GAN-generated images detection," in *2023 International Conference on Digital Applications, Transformation & Economy (ICDATE)*, Jul. 2023, pp. 1–6. doi: 10.1109/ICDATE58146.2023.10248679.

[6] A. Poddar, S. Gawade, P. Varpe, and S. Bhagwat, "Frontal face landmark generation using GAN," in *2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, May 2022, pp. 1172–1177. doi: 10.1109/ICAAIC53929.2022.9793189.

[7] B. Kuriakose, T. Thomas, N. E. Thomas, S. J. Varghese, and V. A. Kumar, "Synthesizing images from hand-drawn sketches using conditional generative adversarial networks," in *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, Jul. 2020, pp. 774–778. doi: 10.1109/ICESC48915.2020.9155550.

[8] R. Bayoumi, M. Alfonse, and A.-B. M. Salem, "An intelligent hybrid text-to-image synthesis model for generating realistic human faces," in *2021 Tenth International Conference on Intelligent Computing and Information Systems (ICICIS)*, Dec. 2021, pp. 172–176. doi: 10.1109/ICICIS52592.2021.9694194.

[9] M. Vijay, M. Meghana, N. Aklecha, and R. Srinath, "Dialog driven face construction using GANs," in *2020 IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI)*, Nov. 2020, pp. 647–652. doi: 10.1109/ICTAI50040.2020.00104.

[10] J. Li, T. Sun, Z. Yang, and Z. Yuan, "Methods and datasets of text to image synthesis based on generative adversarial network," in *2022 IEEE 5th International Conference on Information Systems and Computer Aided Education (ICISCAE)*, Sep. 2022, pp. 843–847. doi: 10.1109/ICISCAE55891.2022.9927634.

[11] K. V. Swamy, A. Supraja, P. S. Vinuthna, and D. L. Sindhura, "Performance comparison of various features for human face recognition using machine learning," in *2022 IEEE Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*, Dec. 2022, pp. 1–4. doi: 10.1109/IATMSI56455.2022.10119449.

[12] V. s. K. Katta, H. Kapalavai, and S. Mondal, "Generating new human faces and improving the quality of images using generative adversarial networks (GAN)," in *2023 2nd International Conference on Edge Computing and Applications (ICECAA)*, Jul. 2023, pp. 1647–1652. doi: 10.1109/ICECAA58104.2023.10212099.

[13] R. Jadhav, V. Gokhale, M. Deshpande, A. Gore, A. Gharpure, and H. Yadav, "High fidelity face generation with style generative adversarial networks," in *2023 2nd International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN)*, Apr. 2023, pp. 1–6. doi: 10.1109/ICSTSN57873.2023.10151603.

[14] K. Kovarthanan and K. M. S. J. Kumarasinghe, "Generating photographic face images from sketches: A study of GAN-based approaches," in *2023 8th International Conference on Information Technology Research (ICITR)*, Dec. 2023, pp. 1–6. doi: 10.1109/ICITR61062.2023.10382944.

[15] S. S, R. P. Kumar, and S. N. Mudassir, "Sketch to image synthesis using attention based contextual GAN," in *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Jul. 2023, pp. 1–6. doi: 10.1109/ICCCNT56998.2023.10306444.

[16] A. K. Bhunia *et al.*, "Sketch2Saliency: Learning to detect salient objects from human drawings," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2023, pp. 2733–2743. doi: 10.1109/CVPR52729.2023.00268.

[17] M. A. Khan and A. S. Jalal, "Suspect identification using local facial attributed by fusing facial landmarks on the forensic sketch," in *2020 International Conference on Contemporary Computing and Applications (IC3A)*, Feb. 2020, pp. 181–186. doi: 10.1109/IC3A48958.2020.233293.

[18] C. Peng, C. Zhang, D. Liu, N. Wang, and X. Gao, "HiFiSketch: High fidelity face photo-sketch synthesis and manipulation," *IEEE Transactions on Image Processing*, vol. 32, pp. 5865–5876, 2023, doi: 10.1109/TIP.2023.3326680.

[19] J. Zheng, Y. Tang, A. Huang, and D. Wu, "Hierarchical multivariate representation learning for face sketch recognition," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 8, no. 2, pp. 2037–2049, Apr. 2024, doi: 10.1109/TETCI.2024.3359090.

[20] W. Feng, Z. Meng, and L. Wang, "Stacked generative adversarial networks for image generation based on U-Net discriminator," in *2022 Asia Conference on Algorithms, Computing and Machine Learning (CACML)*, Mar. 2022, pp. 762–768. doi: 10.1109/CACML55074.2022.00132.

[21] S. Zhang, H. Wang, and L. Wang, "A sensitive image generation method based on improved PatchGAN," in *2023 12th International Conference of Information and Communication Technology (ICTech)*, Apr. 2023, pp. 568–572. doi: 10.1109/ICTech58362.2023.00110.

[22] A. S. Saqlain, F. Fang, T. Ahmad, L. Wang, and Z. Abidin, "Evolution and effectiveness of loss functions in generative adversarial networks," *China Communications*, vol. 18, no. 10, pp. 45–76, Oct. 2021, doi: 10.23919/JCC.2021.10.004.

[23] B. Xie and C. Jung, “Deep face generation from a rough sketch using multi-level generative adversarial networks,” in *2022 26th International Conference on Pattern Recognition (ICPR)*, Aug. 2022, pp. 1200–1207. doi: 10.1109/ICPR56361.2022.9956126.
[24] G. Srujana, Y. Madhuri, T. Harshitha, U. G. Manohari, and S. Rani, “Suspect face detection by auto sketching,” in *2023 IEEE 12th International Conference on Communication Systems and Network Technologies (CSNT)*, Apr. 2023, pp. 456–463. doi: 10.1109/CSNT57126.2023.10134737.
[25] C. Philip and L. H. Jong, “Face sketch synthesis using conditional adversarial networks,” in *2017 International Conference on Information and Communication Technology Convergence (ICTC)*, Oct. 2017, pp. 373–378. doi: 10.1109/ICTC.2017.8191006.
[26] Z. Li, C. Deng, E. Yang, and D. Tao, “Staged sketch-to-image synthesis via semi-supervised generative adversarial networks,” *IEEE Transactions on Multimedia*, vol. 23, pp. 2694–2705, 2021, doi: 10.1109/TMM.2020.3015015.

BIOGRAPHIES OF AUTHORS

Imran Ulla Khan is a research scholar at Reva University and an Assistant Professor in the Department of Computer Science and Engineering at Sri Krishna Institute of Technology, Bangalore, India. He is a tech enthusiast with a keen interest in image processing and deep learning. His research focuses on sketch-based image retrieval and generative adversarial networks (GANs) for enhancing sketch-to-realistic image transformation. He has presented and published his work in various conferences and journals. He is passionate about advancing his expertise in deep learning architectures to address challenges in image synthesis and related domains. He can be contacted at email: imran161984@gmail.com.

Depa Ramachandraiah Kumar Raja is currently working as Professor in the School of Computer Science and Engineering at REVA University Bangaluru, Karnataka, India. He received his Bachelor of Technology (B. Tech) from JNTUA College of Engineering and Master of Technology (M. Tech) from National Institute of Technology Karnataka (NITK) Surathkal, Karnataka, India. He received a Doctorate of Philosophy (Ph.D.) from St Peters University, Chennai, India for An effective context-driven recommender system for e-commerce applications. He did Post Doctoral research at Universiti Teknikal Malaysia, Melaka, Malaysia from September 2023 to September 2024. He received funding amount of 18000 USD for carrying out research projects one for 16000 USD from REVA University for the project Humanoid Robot and completed the project in 2022 and another for 2000 USD from Sree Vidyanikethan Educational Trust for the project automation of areators for aqua culture using IoT and completed the project in 2019. His research areas include the internet of things, data mining, machine learning and artificial intelligence. He can be contacted at email: kumarrajadr@gmail.com.