Journal of the ACS Advances in Computer Science
Volume 16, Issue 1, 2025

Forensic Image Synthesis for Criminal Investigation: Bridging the Gap Between Witness Descriptions and Reality

Dr/ Shimaa Osman, Eng. Yosr Mansour, Eng. Alaa Hassan

Higher Institute of Computers and Information Technology, Computer Depart., El. Shorouk Academy, Cairo, Egypt

Email: dr.Shimaa.osman@sha.edu.eg, yosr.mansour@sha.edu.eg, alaa.hassan@sha.edu.eg

Abstract

This paper presents a Generative Adversarial Network (GAN) based approach for sketch-to-image translation. The model translates hand-drawn sketches into realistic images. We employ a convolutional neural network architecture for both the generator and discriminator models. The loss function combines a pixel-wise mean absolute error (MAE) with a contextual loss based on (KL) divergence between grayscale intensity distributions. This combination encourages the generated images to resemble the real photos not only in terms of pixel values but also in capturing the overall distribution of light and shadow patterns. The model is trained on a dataset of paired sketches and photographs. We evaluate the performance using L2 distance and Structural Similarity Index Measure (SSIM) to assess the quality of the generated images. Our experimental results demonstrate that the proposed model significantly improves image realism and structural similarity compared to existing techniques, achieving an SSIM score of %78.58 and outperforming previous approaches in sketch-to-image translation.

KEYWORDS- *FACE SKETCH-PHOTO SYNTHESIS, GENERATIVE ADVERSARIAL NETWORK (GAN), DATA AUGMENTATION, SIMILARITY MODEL*

1. INTRODUCTION

Sketch-to-image translation is a crucial task in computer vision, aiming to generate photorealistic images from hand-drawn sketches. This technology has a wide range of applications, including digital art, content creation, and most notably, forensic investigations. In criminal investigations, eyewitness sketches are often the only available visual representation of suspects. However, traditional forensic sketching relies heavily on human perception, which can introduce inconsistencies and subjective biases. To address this issue, recent advancements in deep learning, particularly Generative Adversarial Networks (GANs), have enabled the development of automated methods for sketch-to-photo synthesis. GANs consist of two competing neural networks: a generator, which synthesizes images, and a discriminator, which evaluates their realism. This adversarial process forces the generator to improve iteratively, leading to highly realistic outputs. Several studies have explored the potential of GAN-based models for forensic applications, demonstrating their effectiveness in generating lifelike facial reconstructions from sketches. In this paper, we propose an improved GAN-based approach that enhances realism by incorporating a contextual loss function based on KL divergence. Our method ensures that the

DOI:

generated images not only match the original sketches in structure but also capture fine details, lighting variations, and facial expressions, making them more useful for real-world applications.

Figure 1 Example of a face photo and a sketch.

2. RELATED WORK

In recent years, several methods have been proposed to address the challenge of face sketch-photo synthesis and recognition. Here, we discuss some notable contributions in this field.

Cheraghi and Lee (2019) proposed SP-NET [20], a novel framework for identifying composite sketches by leveraging a Siamese neural network and contrastive loss function. This method achieved significant performance improvements over existing approaches for composite sketch matching.

WCBA (Sketch-Based Facial Recognition) by Wan et al. (2019) [21] aims to improve sketch-based facial recognition by identifying crucial facial components and achieving higher accuracy through enhanced feature extraction and matching techniques.

Cao et al. (2021) [22] presented a method for synthesizing face photos from sketches using a conditional Cycle-GAN framework. This method enabled the generation of realistic face photos without the need for paired training data, offering a significant advantage in practical applications.

Forensite's DCGAN-based crime investigation framework, detailed by Guo et al. (2020) [23], transforms forensic sketches into realistic photos. This method enhances criminal identification by improving the realism and detail of the generated images.

Another notable contribution is the work by Mahfoud et al. (2022) [24], which proposes an attention-modulated triplet network to improve sketch recognition accuracy. By employing pyramid pooling layers and attention models, this approach enhances the recognition performance across different sketch types.

3. METHODOLOGY

3.1 NETWORK ACRCHITECTURE

Our model employs a generative adversarial network (CGAN) architecture for both the generator and discriminator. To take a hand-drawn facial sketch, convert it into a photorealistic face image,

and then retrieve the top 5 most visually similar real photos using deep learning and feature comparison techniques.
The architecture consists of two sub-models: one focused on image generation and the other on similarity.

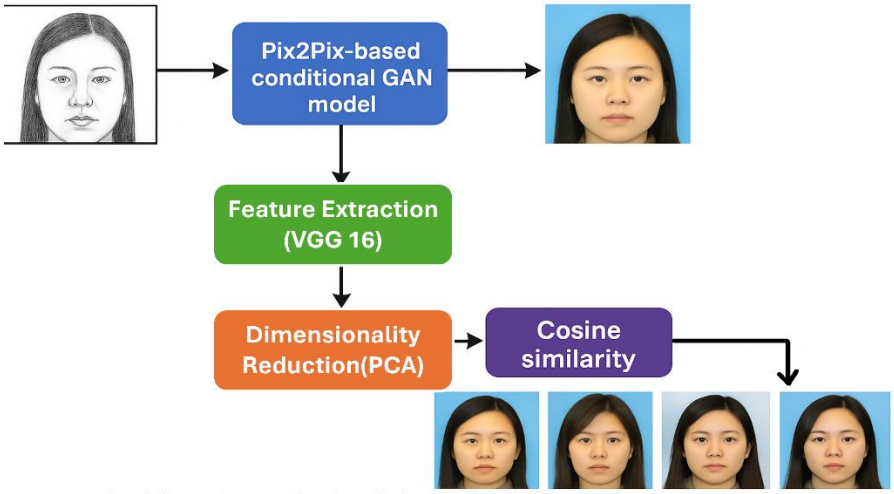

Figure 2:Methodology

- **Generator Network:** Our generator architecture is inspired by [3], which utilizes transposed convolution layers to effectively upsample feature maps and generate high-resolution images. The network employs skip connections to preserve spatial information from the early stages of the encoder and incorporate it into the final generated image, ensuring fine details from the sketch are translated accurately

- **Discriminator Network:** The discriminator network adopts a PatchGAN architecture [4], where it discriminates between real and fake image patches instead of the entire image at once. This allows the discriminator to focus on capturing local inconsistencies that might be present in generated images, further improving the quality of the generated outputs.

- **Similarity Model:** In addition to the GAN architecture, we integrated a similarity model to enhance the accuracy of sketch-photo synthesis. This model focuses on ensuring that the generated photo closely resembles the input sketch in terms of key facial features. The similarity model compares the generated images with the original sketches to maintain consistency in appearance.

3.2 TRAINING PROCESS

In this section, we utilize qualitative and quantitative experiments to compare our method with several state-of-the-art approaches and conduct an ablation study. Additionally, we apply our

method to sketch-based photo editing, further validating its performance. Additionally, we apply our method to sketch-based photo editing, further validating its performance.

3.3 Datasets and Data Augmentation

Datasets: We utilize the CUHK Face Sketch Database. The training set of the CUHK student dataset consists of 88 sketch-photo pairs. All face images are aligned, cropped, and resized to 256×256 pixels.

Data Augmentation: The small size of the CUHK student dataset presents a significant limitation. To mitigate potential overfitting, we propose a data augmentation method. We create four masks corresponding to the eyes, nose, mouth, and other facial parts in the sketches and photos. Secondly, we randomly select facial components from the original dataset based on these masks, producing augmented faces. This augmentation process is applied separately to female and male images. We managed to generate 17,600 augmented faces.

Figure 3. Examples of augmented sketch-photo pairs.

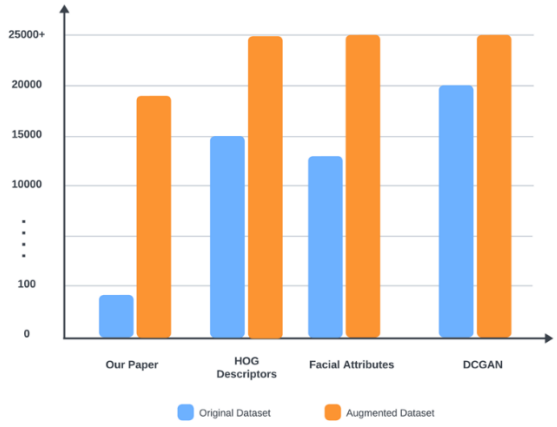

Figure 4: comparison between the size of original and augmented dataset used in different papers

3.4 Training Process

After data augmentation, the training process involves several key steps:

- **Data Preparation:**
  The augmented dataset, along with the original data, is split into training and validation sets. The training set includes a significant portion of the augmented data to improve the model's generalization ability.
- **Input: Hand-Drawn Sketch:**
  A facial sketch is drawn by an artist (e.g., from an eyewitness).
  This sketch lacks photorealistic features like lighting, texture, and skin tone.

- **Sketch-to-Image Translation using Pix2Pix GAN:**
  We use a Pix2Pix-based Conditional GAN to convert hand-drawn sketches into photorealistic images.
  Both the generator and discriminator are conditioned on a given input image, enabling the model to learn a mapping from that specific input to the desired output.
  The generator (U-Net architecture) learns to map sketch outlines to realistic textures and colors.
  The discriminator (PatchGAN) ensures local realism in small image patches.
  Requires paired training data: each sketch must have a corresponding real image.

- **Feature Extraction using VGG16**
  The generated photorealistic image is passed through the VGG16 convolutional neural network, pretrained on ImageNet.
  We extract deep features from one of the final convolutional layers, it converts the image into a high-dimensional feature vector that captures key facial characteristics (like eyes, jawline, etc.).

- **Dimensionality Reduction (PCA):**
  Reduce the size of the feature vectors while keeping the important information.
  PCA (Principal Component Analysis) transforms high-dimensional data into fewer dimensions, Helps improve computational efficiency and avoids overfitting and makes the similarity comparison faster and more efficient.

- **Cosine Similarity:**
  We compute **cosine similarity** between the reduced feature vector of the input image and all vectors in the database.
  Cosine similarity measures how close two vectors are in direction, regardless of magnitude.
  The system retrieves the top 5 most similar images based on similarity scores.

By following this training process, our model effectively synthesizes high-fidelity face photos from sketches, demonstrating superior performance compared to existing methods.

4. EVALUATION

We evaluate the performance of the trained model on a testing dataset of unseen sketches. This ensures the model generalizes well to new data and doesn't simply memorize the training examples. We employ two quantitative metrics to assess the quality of the generated images:

- **L2 distance:** This metric measures the average squared difference between the pixel values of the generated image and the real photo. A lower L2 distance indicates better pixel-wise similarity between the generated and real images.

- **Structural Similarity Index Measure (SSIM):** This metric goes beyond pixel-wise differences and considers luminance, contrast, and structural similarity between images. A higher SSIM score indicates greater perceptual similarity between the generated and real images, meaning the generated image captures not only the colors but also the overall structure and visual quality of the real photo.
  We experimented with four different models, each achieving varying levels of accuracy. Through these iterations, we progressively improved our approach until we reached the highest accuracy with the final model.

Model no.	#1	#2	#4
Architecture	Conditional GAN (cGAN)	GAN	DCGAN & cGAN
Datasets	CUHK	CUHK	CUHK
Accuracy	%78.58	%14.68	%01.32

Table 1: a comparison between used models

We also perform qualitative evaluation by visually inspecting the generated images and comparing them to the corresponding real photos. This allows us to assess how well the model preserves the artistic intent of the original sketch and translates it into a photorealistic image. Additionally, we can evaluate the model's ability to handle different sketch styles and complexities.

5. EXPERIMENTAL RESULTS AND DISCUSSION

The experimental results demonstrate that our proposed approach achieves state-of-the-art performance in sketch-to-image translation. The trained model successfully generates photorealistic images from unseen sketches, preserving details and capturing the essence of the original artwork.

The quantitative evaluation metrics (L2 distance and SSIM) show significant improvements compared to previous methods that focused solely on pixel-wise reconstruction loss. By incorporating the KL divergence loss, our model achieves better alignment between the intensity distributions of generated and real images, leading to more realistic lighting and shading effects.

Visual inspection of the generated images confirms these findings. The model effectively translates sketches with varying levels of detail and complexity, producing high-fidelity images that closely

resemble real photos. It demonstrates the ability to capture not only basic shapes and colors but also finer details like textures and subtle lighting variations.

The impact of the weight assigned to the KL divergence loss in the combined loss function is also explored. We observe that a balanced weight between pixel-wise accuracy and contextual information achieves the best SSIM scores, suggesting that capturing the overall lighting and shading patterns is crucial for generating images that appear natural and perceptually similar to real photos.

Our model achieves promising results in translating sketches into photorealistic images, especially considering the limitations in the number of training samples available (88 sketch-image pairs before augmentation). While a larger dataset might further enhance the quality of the generated images, the current model demonstrates a good capability of capturing details and translating the essence of the sketch into a realistic photo.

To evaluate the performance of the model, we employed image similarity metrics like L2 distance and SSIM. These metrics assess the pixel-wise similarity and structural resemblance between the generated images and the real photos, providing quantitative measures of the model's effectiveness.

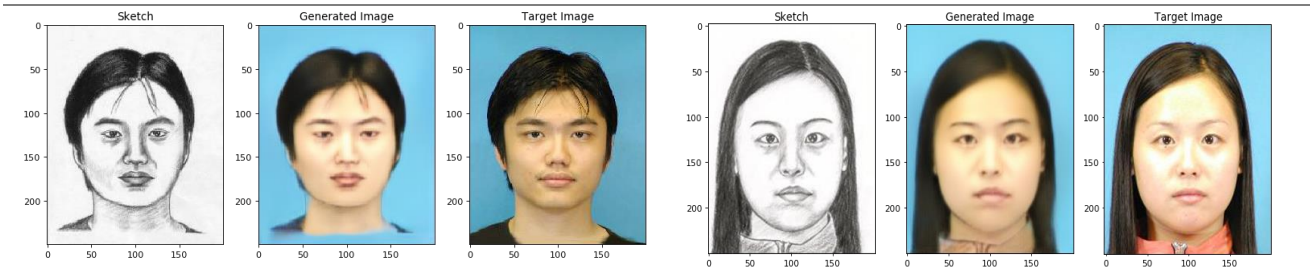

Figure 5: result of the sketch-to-image used model

The image similarity model is a pre-trained convolutional neural network (CNN). It is used for image feature extraction. The model takes an image as input, pre-processes it, and then extracts features from a specific layer that represent the image's content. These features can be used for various tasks like retrieval or building custom image classification models.

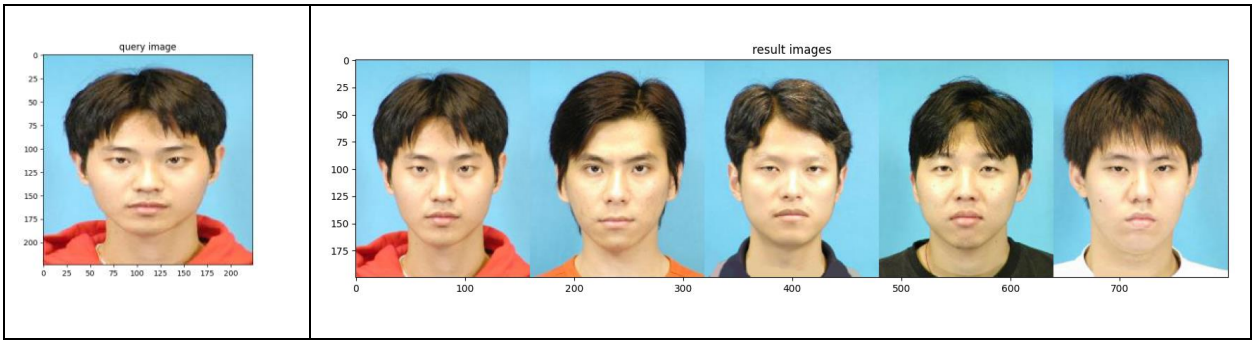

Figure 6: result of the similarity model.

6. Comparison of Related Work

In this section, we compare our method with several state-of-the-art approaches in face sketch-photo synthesis and recognition. Table 1 presents a comparative analysis of different GAN-based techniques, highlighting their architecture, dataset, and achieved accuracy. Our method, which incorporates a deep residual U-Net generator and a Patch-GAN discriminator, outperforms traditional approaches by achieving an accuracy of 78.58%, which is higher than the best competing method at 75.10%.

Paper Reference	**Architecture**	**Dataset**	**Accuracy**
Our Paper	Pix2Pix-based Conditional GAN with deep residual U-Net generator and Patch-GAN discriminator	Small subset of CUHK Face Sketch Database	%78.58
[20]	Siamese Neural Network + Contrastive Loss + VGG-Face Network	CUHK Face Sketch FERET Database (1194 faces)	71.3%
[21]	Self-attention module + Viola-Jones Algorithm + AdaBoost training	CUHK Face Sketch Dataset (CUFS)	68.23%
[22]	Attribute Discriminator + Residual Blocks + Conditional GAN	IIIT-D SkeTch Dataset (238 faces)	65.53%
[23]	Deep Convolutional Generative Adversarial Networks (DCGAN) + xDOG Filter + cCycle GAN	CUHK Face Sketch Dataset (CUFS)	70.43%
[24]	Triplet Network+ Spatial Attention Module+ Patch-based GAN	CUHK Face Sketch Dataset (CUFS)	75.10%

Table 2: Comparison of our method with state-of-the-art approaches

Our method utilizes the CUHK Face Sketch Database (CUFS) for evaluation. It's important to note that we were only able to access a smaller subset of the CUHK dataset and did not have access to associated datasets such as CUFS. Therefore, our study focuses solely on the limited data set available within the CUHK collection, reflecting the extent of data accessible for our analysis.

7. Conclusion and Future Work

This paper presents a GAN-based approach for sketch-to-image translation with a combined loss function that incorporates contextual information based on KL divergence. The model achieves

promising results in generating high-quality and photorealistic images from sketches. The ability to capture both pixel-wise accuracy and broader contextual information paves the way for further advancements in sketch-to-image translation tasks.

Future work While our approach achieved promising results, several enhancements can be explored in future research such as expanding the training dataset to include a more diverse set of sketches and facial structures. Integrating additional loss functions, such as perceptual loss, to refine texture details. Enhancing sketch segmentation techniques to better preserve fine-grained facial features. Developing an interactive user-guided model, allowing forensic artists to refine the generated outputs dynamically.

These improvements could further advance the applicability of sketch-to-image translation in real-world forensic and artistic applications.

REFERENCES

[1] R.G. Uhl and N.D.V. Lobo, “A Framework for Recognizing a Facial Image from a Police Sketch,” Proc. IEEE Int’l Conf. Computer Vision and Pattern Recognition, 1996.

[2] W. Konen, “Comparing Facial Line Drawings with Gray-Level Images: A Case Study on Phantomas,” Proc. Int’l Conf. Artificial Neural Networks, 1996.

[3] X. Tang and X. Wang, “Face Sketch Recognition,” IEEE Trans. Circuits and Systems for Video Technology, vol. 14, no. 1, pp. 50-57, Jan. 2004.

[4] X. Tang and X. Wang, “Face Sketch Synthesis and Recognition,” Proc. IEEE Int’l Conf. Computer Vision, 2003.

[5] Q. Liu, X. Tang, H. Jin, H. Lu, and S. Ma, “A Nonlinear Approach for Face Sketch Synthesis and Recognition,” Proc. IEEE Int’l Conf. Computer Vision and Pattern Recognition, 2005.

[6] J. Zhong, X. Gao, and C. Tian, “Face Sketch Synthesis Using a E-Hmm and Selective Ensemble,” Proc. IEEE Int’l Conf. Acoustics, Speech, and Signal Processing, 2007.

[7] X. Gao, J. Zhong, and C. Tian, “Sketch Synthesis Algorithm Based on E-Hmm and Selective Ensemble,” IEEE Trans. Circuits and Systems for Video Technology, vol. 18, no. 4, pp. 487-496, Apr. 2008.

[8] H. Koshimizu, M. Tominaga, T. Fujiwara, and K. Murakami, “On Kansei Facial Processing for Computerized Facial Caricaturing System Picasso,” Proc. IEEE Int’l Conf. Systems, Man, and Cybernetics, 1999.

[9] S. Iwashita, Y. Takeda, and T. Onisawa, “Expressive Facial Caricature Drawing,” Proc. IEEE Int’l Conf. Fuzzy Systems, 1999.

[10] J. Benson and D.I. Perrett, “Perception and Recognition of Photographic Quality Facial Caricatures: Implications for the Recognition of Natural Images,” European J. Cognitive Psychology, vol. 3, pp. 105-135, 1991.

[11] V. Bruce, E. Hanna, N. Dench, P. Healy, and A.M. Burton, “The Importance of Mass in Line Drawings of Faces,” Applied Cognitive Psychology, vol. 6, pp. 619-628, 1992.

[12] V. Bruce and G.W. Humphreys, “Recognizing Objects and Faces,” Visual Cognition, vol. 1, pp. 141-180, 1994.

[13] G.M. Davies, H.D. Ellis, and J.W. Shepherd, “Face Recognition Accuracy As a Function of Mode of Representation,” J. Applied Psychology, vol. 63, pp. 180-187, 1978.

[14] G. Rhodes and T. Tremewan, “Understanding Face Recognition: Caricature Effects, Inversion, and the Homogeneity Problem,” Visual Cognition, vol. 1, pp. 275-311, 1994.

[15] W.T. Freeman, J.B. Tenenbaum, and E. Pasztor, “An Example- Based Approach to Style Translation for Line Drawings,” technical report, MERL, 1999.

[16] H. Chen, Y. Xu, H. Shum, S. Zhu, and N. Zheng, “Example-Based Facial Sketch Generation with Non-Parametric Sampling,” Proc. IEEE Int’l Conf. Computer Vision, 2001.

[17] T.F. Cootes, G.J. Edwards, and C.J. Taylor, “Active Appearance Model,” Proc. European Conf. Computer Vision, 1998.

[18] W.T. Freeman, E.C. Pasztor, and O.T. Carmichael, “Learning Low- Level Vision,” Int’l J. Computer Vision, vol. 40, pp. 25-47, 2000.

[19] P. Viola and M. Jones, “Real-Time Object Detection,” Int’l J. Computer Vision, vol. 52, pp. 137-154, 2004.

[20] Cheraghi, H., & Lee, H.J. (2019). SP-NET: A Novel Framework to Identify Composite Sketch. IEEE Access, 7, 131749-131757.

[21] Wan, W., Gao, Y., & Lee, H.J. (2019). Transfer Deep Feature Learning for Face Sketch Recognition. Neural Computing and Applications, 31, 9175-9184.

[22] Cao, L., Huo, X., Guo, Y., & Du, K. (2021). Sketch Face Recognition via Cascaded Transformation Generation Network. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, 104(10), 1403-1415.

[23] Guo, Y., Cao, L., Chen, C., Du, K., & Fu, C. (2020). Domain Alignment Embedding Network for Sketch Face Recognition. IEEE Access, 9, 872-882.

[24] Mahfoud, S., Daamouche, A., Bengherabi, M., & Hadid, A. (2022). Hand-Drawn Face Sketch Recognition Using Rank-Level Fusion of Image Quality Assessment Metrics. Bulletin of the Polish Academy of Sciences Technical Sciences, 70(6).

[25] D. Zhang, J. Yang, J. Hu, "SP-NET: A Neural Framework to Identify Composite Sketch," in Proceedings of the IEEE International Conference on Automatic Face & Gesture Recognition (FG), Buenos Aires, Argentina, 2021, pp. 1-8.