# Forensic Facial Reconstruction from Sketch in Crime Investigation

Doaa M. Mohammed, Mostafa Elgendy, Mohamed Taha
Department of Computer Science-Faculty of Computers and Artificial Intelligence, Benha University, Benha, Egypt

***Abstract*—Many crimes are committed every day all over the world, and one of them is a criminal offense that includes a wide range of illegal acts such as murder, theft, assault, rape, kidnapping, fraud, and others. Criminals pose a threat to security, which harms the public interest. In this case, the police question all eyewitnesses at the crime scene, and sometimes, witnesses who were present at the crime scene can remember the face of the criminal. The witness accurately describes the person's facial features in the report, such as eyes, nose, etc. Law enforcement authorities use eyewitness information to identify the person. Criminal investigations can be accelerated by converting sketched faces into actual images, but this requires eyewitnesses to confirm the description in the report. Drawings make it very difficult to identify real human faces because they do not contain the details that help to catch criminals. In contrast, color photographs contain many details that help to identify facial features more clearly. This work proposes to generate color images using the modified modulation Sketch-to-Face CycleGAN and then pass them through Generative Facial Prior-GAN. CycleGAN consists of a generator and discriminator. The generator is used to generate colored images, and the discriminator is used to identify whether the images are real or fake. These are then passed to GFPGAN to improve the quality of the colored images. The structural similarity index measure of 0.8154 is achieved when creating photorealistic images from drawings.**

***Keywords*—Sketch-to-Face; facial features; Sketch-to-Face CycleGAN; victim's identification; criminal offense**

## I. Introduction

Crime is a pervasive issue that transcends cultural, geographical, and social boundaries, impacting individuals and communities worldwide. Crime is broadly defined as an act that violates the law and is punishable by the state. It manifests in various forms, from minor offenses to serious felonies. The consequences of criminal activities extend beyond the immediate victim, affecting families, neighborhoods, and society. As crime rates fluctuate, the need for effective crime investigation becomes increasingly critical, serving as a fundamental component of law enforcement and public safety. Crimes can be categorized into several types, including violent crimes, property crimes, white-collar crimes, and cybercrimes, each posing unique dangers to society. Murder and assault are examples of violent crimes. Murders are one of the crimes that have the greatest impact on global societies. They are not just isolated acts of violence but represent a flagrant violation of human rights and a threat to public safety [1]. In the aftermath of a murder, law enforcement's role in identifying suspects is critical to public security and criminal justice operations. This role entails using various legal and technical techniques and tools that aid in accurately and effectively identifying suspected individuals [2]. The role of law enforcement begins with gathering information from multiple sources, such as witnesses, physical evidence, and field investigations. The services of legal drawing experts are used to draw portraits of criminals based on eyewitness descriptions. It entails the process of creating a visual representation of an unidentified person or person of interest using memories and details provided by witnesses or victims [3]. Experts specializing in drawing arts sketch suspected individuals. Forensic artists interview witnesses or victims to obtain accurate details about the person they observed [4]. The information provided may include details related to the suspect's facial features, hair, clothes, tattoos, scars, and any other distinguishing features. Expert forensic artists can transform an accurate description of the human form into a visual image that can be used as evidence in criminal investigations and trials. Forensic artists use manual or computer-based techniques [5]. Drawings are used as evidence that reinforces other available evidence, such as fingerprints, certificates, or physical evidence. When the hand-drawn sketches are completed, they are distributed to the media to facilitate the perpetrator's arrest [6]. The current identification techniques, particularly those relying on forensic sketches, often suffer from inaccuracies due to the subjective nature of eyewitness accounts. These limitations hinder law enforcement's ability to identify and apprehend suspects quickly. This study aims to address these challenges by improving the Structural Similarity Index Measure and the effectiveness of suspect identification through advanced imaging techniques. This study employs modified STF CycleGAN (Sketch-to-Face CycleGAN) in conjunction with GFP-GAN (Generative Facial Prior GAN) to reconstruct images generated by the modified STF CycleGAN model, resulting in a high-fidelity representation of the colored generated image. This paper mainly uses a modified CycleGAN model with GFP-GAN to transform sketch hand-drawn face images into colored real images. Colored images are useful in forensic science and law enforcement, where suspect sketches based on eyewitness statements can be turned into realistic graphics to help law enforcement identify and apprehend criminals. Law enforcement enhances the likelihood of finding missing individuals and criminals. The rest of this work is organized in the following way: Section II is Related Work, and Section III is Methodology. Section IV is Experimental Results. Section V is Discussion. Section VI is Conclusion and Future Work. Finally, References are given.

## II. Related Work

We summarize past research on facial reconstruction, utilizing techniques from computer vision and deep learning to reconstruct the face. Shikang Yu et al. introduced a generative adversarial network (GAN) that utilizes the benefits of CycleGAN and conditional GANs to perform face sketch-to-photo transformation [7]. This study presented the development of a novel feature-level loss function, i.e., integrated with the conventional image-level adversarial loss function to enhance the quality of synthesized images. The generator and discriminator provide additional data to supplement the network's training. In addition to the synthetic facial picture, the network's discriminators also receive a genuine image from the other modality as input, with the real image serving as auxiliary information. A feature-level loss is used to assign a penalty to the disparities between a processed image and the original image within the feature space. The generator employed a U-Net architecture [8] because of its efficacy in Pix2Pix, a technique commonly used for picture processing [9]. There are three convolution layers in the discriminator. The CUHK Face Sketch FERET Database (CUFSF) was utilized in conjunction with SSIM 0.5517 [10].

Sreedev Devakumar et al. converted pencil sketches into actual photographs for forensic investigation to ascertain an individual's identity. DCGAN (Deep Convolutional Generative Adversarial Network) transformed the sketch image into a real image [11]. The proposed model comprises a single generator network (G) and two discriminator networks (D1, D2). The photo generator employs convolutional operations on the input sketch to produce images. A patch GAN serves as the initial discriminator, calculating the L1 loss or discriminator loss. As a result, the generator continues to apply this discriminator loss over subsequent epochs. The generator will activate the second discriminator after a total of 30 epochs.

In contrast to the initial scenario, the second discriminator is a conventional discriminator incorporating an additional dense layer. The second classifier establishes a correspondence between the target-generated image and the target photo. The proposed network architecture employs a patch GAN as one of its discriminators. The second option is a conventional discriminator that incorporates an additional dense layer. The conv2D generator employs batch normalization and a Relu layer. The discriminator conv2D observes the application of batch normalization and a leaky Relu layer. The discriminator uses 4x4 filters and a stride of 2 for each convolution. Simultaneously, the generator uses 3x3 filters with a stride value of 2. The G Generator uses an expanded U-Net architecture, employing DCGAN and incorporating six batches of convolutions to enhance the quality of generated images. This task utilized the CUHK Face Sketch FERET Database (CUFSF) dataset, resulting in a final average SSIM of 0.587 [12].

Lidan Wang et al. presented The Photo-Sketch Synthesis using the Multi-Adversarial Networks (PS2-MAN) approach, which involves the iterative generation of low-resolution to high-resolution images in an adversarial manner. This study implements adversarial supervision at all resolution levels. More precisely, the feature maps at each deconvolution layer use 3 x 3 convolutions to provide outputs with varying resolutions. Two generator sub-networks, namely GA and GB, respectively, perform the transformation from photo to sketch and from sketch to photo. The Genetic Algorithm (GA) uses a genuine facial photograph of RA as input and a synthetic (false) representation of FB as output. GB endeavors to convert sketches into photographs. The generator sub-networks objective is to generate images that closely resemble real ones to deceive the discriminator sub-networks. Conversely, the objective of the discriminator sub-networks is to acquire the ability to distinguish between generated and authentic samples. CUFS and CUFSF datasets are commonly used in this study [13]. The CUHK Face Sketch Database (CUFS) is a comprehensive collection of sketches. It consists of 188 faces from the Chinese University of Hong Kong (CUHK) student database, 123 from the AR database, and 295 from the XM2VTS database [14]. The SSIM (picture synthesis) resulted in 0.7915 for picture synthesis and 0.6156 for sketch synthesis [15].

Sparsh Nagpal et al. introduced a novel contextual generative adversarial network designed specifically for sketch-to-image production. The sketch was employed as a lax constraint to identify the most similar mapping and establish our objective function, which comprises a contextual loss and a conventional GAN loss. Additionally, they use a straightforward approach to enhance the start of sketches. This study posits that inpainting produces an image in and of itself. The use of a mask distorts a fraction of the input image. The image, partially obscured by a mask, is called the context. The model attempts to produce a comprehensive representation of this situation. The researchers employed seven up-convolutional layers, each having a kernel size of four and a stride of two. A batch normalizing layer is applied after each up-convolutional layer to expedite the training process and enhance the stability of the learning. Reputation linear unit (Relu) activation is employed at all levels. Ultimately, we employ the tanh function on the output layer. The resulting photos were fed into a layer of pre-trained GFP GANs that were trained using low-quality portrait enhancers. This enhanced image quality and generated more realistic outputs, even when utilizing a less sophisticated model. The dataset utilized the CUHK Face Sketch database (CUFS), the XM2VTS database, and the AR database, with a simultaneous similarity index (SSIM) of 0.78 [16].

Sumit Gunjate et al. introduced a methodology for converting sketches into images using generative adversarial networks. Converting an individual's representation into an image that encompasses the characteristics or attributes associated with the representation necessitates the utilization of many categories of machine learning algorithms. The models comprised a GAN discriminator, which functions just as a classifier. The task involves distinguishing between factual data and data produced by the generator. The generator component of a generative adversarial network (GAN) acquires the ability to generate false data. It acquires the capability to designate its output as genuine. In contrast to the process of discriminator training, generator training requires a more advanced level of integration between the generator and the discriminator. The generator model used the U-Net architecture, which consists of encoder and decoder layers. The model takes a picture (the sketch) as input and applies 7 encoding layers with filters (C64, C128, C256, C512, C512, C512, C512, C512, C512) to make the image smaller in size. The encoding layers employ batch normalization except for the first layer (C64). Their research aims to determine whether the generated facial structures if deemed credible, share the same identity label as authentic faces. The researchers specifically designed a light CNN to extract identity-preserving features and utilized the L2 norm for comparison purposes [17].

## III. Methodology

This research suggests using a modified STF CycleGAN model output as an input to the GFP-GAN model to obtain the result, as shown in Fig. 1.


Fig. 1. The architecture used in this paper.

### A. Preprocess

The technique entails analyzing and utilizing sketches and target images in RGB color. The Gaussian blur feature is achieved by applying a Gaussian function to a picture to reduce noise and generate a more refined visual effect. The filter described above can be considered a nonuniform low-pass filter [19], efficiently preserving low spatial frequency while simultaneously reducing image noise and inconsequential details inside an image. The sketches were subjected to Gaussian blur with a blur radius of 0.3, while the color images were exposed to a blur radius of 0.8. The calculation of Gaussian blur is derived from (1) [18]-[19] determines the Gaussian blur calculation.

$$G_{\sigma}(x) = \frac{1}{2\pi\delta^2} e^{-(x^2+y^2)/2\delta^2} \quad (1)$$

Following the noise reduction process, we employed rescaling augmentation on the photos from the Chuck database with various scales. As a result, 2256 images were generated, including both sketch and ground truth images.

### B. Modified STF CycleGAN Model

CycleGAN is a kind of Generative Adversarial Network (GAN) architecture built primarily for challenges involving the translation of unpaired images. Jun-Yan Zhu et al. introduced it in their 2017 publication [7]. CycleGAN enables the acquisition of mappings between distinct domains, facilitating the transformation of images from one domain to another without necessitating the inclusion of paired data samples from both domains during the training process. CycleGAN can learn image translation across different domains, even without paired data, unlike typical supervised learning approaches that require each input image to be matched with its corresponding destination image. Our objective is to convert prototype photos into their corresponding actual ones. The model comprises two fundamental components, namely a Generator and a Discriminator.

*1) Generator*: The generator is a neural network, i.e. tasked with translating images from one domain to another. The generator functions unsupervised, enabling it to execute the translation task without relying on paired data samples throughout the training process. The generator comprises convolutional neural network (CNN) layers that receive an input image from one domain and generate a corresponding output image in the target domain. The input image size is [256, 256, 3] and undergoes a sequence of downsampling and upsampling modules.

*a) Down-sampling*: Downsampling is a technique used to decrease the spatial resolution of an input image, usually achieved by employing convolutional procedures like step convolutions. One essential operation in the generator network is down-sampling, which extracts hierarchical information from the input image and aids in the translation process. The down-sampling blocks progressively decrease the resolution from 256x256 to 1x1 spatial dimensions. Convolution layers, batch normalization, and the Leaky-Relu activation function are employed in the down-sampling process.

- Convolution Layer: Convolutional layers, sometimes called conv layers, are essential components of convolutional neural networks (CNNs). Convolutional layers are of utmost importance in the process of extracting features from input data through the application of convolutional operations. In this context, the weights assigned to each filter function represent a vocabulary of feature patterns. Filters used are (64, 128, 256, 512, 512, 512, 512, 512) with kernel size 4. As calculated in (2) [20], the convolution operation is computed for the input patch X and the filter kernel K, which determines the convolution layer's calculation. Fig. 2 illustrates the convolution operation.

$$C(m, n) = \sum_{p=0}^{p} \sum_{q=0}^{q} K(p; q) * X(m+p, n+q) + b \quad (2)$$

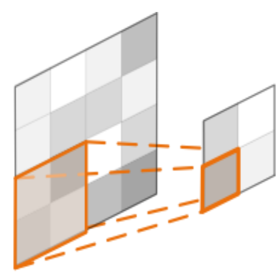

Fig. 2. Convolution operation.

- Batch Normalization: Batch normalization is a technique often used in deep neural networks, particularly in the convolutional and fully connected layers, to improve the training process and the model's overall performance. People widely recognize batch normalization as a prominent technique that enhances network training speed and improves accuracy. [21] Batch normalization changes the data distribution to have a mean of 0 and a variance of 1, expressed in units of the minibatch. This approach is effective in mitigating the possibility of overfitting [22].

- Leaky-Relu Activation Function: The Leaky Rectified Linear Unit (Leaky-Relu) is a frequently employed activation function in neural networks, specifically within deep learning architectures. The proposed approach is an expansion of the conventional Rectified Linear Unit (Relu) function, which aims to overcome certain constraints, notably the issue of "dying Relu" This difficulty arises when neurons that produce a weight of zero for all inputs cease to update their weights during the training process. The Leaky-Relu has a small gradient (usually a small positive value, like 0.01) for negative inputs. This lets them spread a gradient throughout the backpropagation process. Fig. 3 illustrates the Leaky-Relu activation function. The Leaky Relu activation function is defined in (3) [23], which specifies how the function is calculated.

$$Leakly - Relu(x) = max(kx, x) = \begin{cases} x, & \text{if } x > 0 \\ kx, & x \leq 0 \end{cases} Leakly - Relu(x) = max(kx, x) = \begin{cases} x, & \text{if } x > 0 \\ kx, & x \leq 0 \end{cases} \quad (3)$$

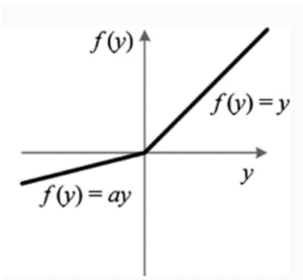


Fig. 3. Leaky-Relu activation function.

*b) Up-sampling:* Upsampling is a technique used to increase the spatial resolution of feature maps. It is usually achieved using techniques such as transposed convolutions. Upsampling plays a crucial role in the generator network by efficiently collaborating with downsampling to convert images across different domains. A 2D transposed convolutional layer, batch normalization, and the Rectified Linear Unit (Relu) activation function are used in upsampling to expand the spatial dimensions. The generator network efficiently produces images with suitable characteristics, and the ultimate output has dimensions of 256x256x3, corresponding to color images.

- Deconvolution Layer: Deconvolution, also known as transposed convolution, is a method that increases the sampling level of feature maps while preserving the connectivity pattern. The deconvolutional layers employ convolution-like procedures with several filters to expand and densify the input. The filters used are (512, 512, 512, 512, 256, 128, 64, 32). Unlike the current resizing methods, the deconvolution process includes adjustable parameters, as shown in Fig. 4 [20]. During network training, the weights of deconvolutional layers undergo continuous updates and refinements. On the input side, the process starts with adding zeros between neurons in the receptive field, and then a convolution kernel with a unit stride is used at the very top.

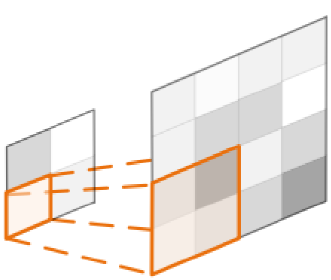

Fig. 4. Deconvolution operation.

- Relu Activation Function: The Rectified Linear Unit (Relu) is a highly prevalent activation function commonly employed in neural networks, particularly in deep learning models. Relu's linear operation results in a higher convergence rate and calculation speed. The neural network's performance is negatively impacted by the presence of dead neurons, which Relu causes. Negative input values result in a constant output of zero for the Rectified Linear Unit (Relu) function, and consequently, the first derivative is also zero. This renders the neuron incapable of updating its parameters, as shown in Fig. 5. The Relu activation function is calculated in (4) [24], which specifies how it is calculated.

$$f(x) = \max(0, x) \quad (4)$$

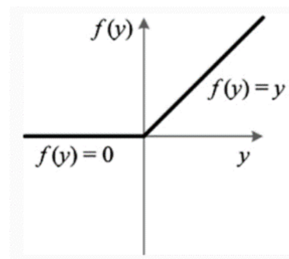


Fig. 5. Relu activation function.

*c) Skip connections*: Skip connections, often referred to as residual connections, serve a vital function in the generator, easing the transmission of information across several layers and enhancing the overall quality of the generated images. The establishment of these links serves to mitigate the issue of information loss that arises during the translation process, hence enabling the generator to retain intricate details within the input images more effectively. The generator's skip connections facilitate the preservation of low-level details and fine-grained information from the input image. Skip links enable the network to access both the original input and the higher-level information acquired during the translation process by directly transmitting the input from a certain layer to a subsequent layer.

*2) Discriminator*: The discriminator is an essential component of the adversarial training process and plays a critical role. The discriminator has a crucial function in differentiating between genuine images from the desired domain and artificial images produced by the generator. It offers input for the generator network to enhance the quality of its generated images. The following items are included: Downsampling blocks are employed to minimize spatial dimensions. We use filters (32, 128, 256, 512) with a filter size of 4, batch normalization, Leaky-Relu, zero padding, and a convolutional layer with only one output channel.

*3) Modified STF CycleGAN loss*: Modified STF CycleGan loss consists of two losses.

*a) Generator loss function*: The generator loss function consists of the following components:

- Adversarial loss (GAN loss): This causes the generator to generate images that are so similar to the actual ones that they cannot be distinguished. The formulation employs binary cross-entropy loss. We use the Adversarial loss to calculate the min-max loss. Adversarial loss is calculated in (5).

$$L_{Gan} = E_x[\log(D(x))] + E_z[\log(1-D_x(G(x)))] \quad (5)$$

- L1 loss is a loss term that ensures the generated and target images have pixel-wise similarity.

To calculate the overall generator loss, we combine these components, using weights that determine the relative significance of each loss item. The generator loss function is calculated in (6).

$$L_{Gen} = L_{GAN} + \lambda 1 \cdot L_{cyc}(G) \quad (6)$$

Where $\lambda 1$ is the weighting factor for the L1 loss, and $L_{GAN}$ represents the adversarial loss, $L_{cyc}$ represents cycle consistency loss. $L_{cyc}$ is calculated in (7).

$$L_{cyc} = E \sim p_{data}(x)\ [\|G(x)-x\|] \quad (7)$$

*b) Discriminator loss function*: The discriminator loss function consists of two distinct components:

- The discriminator's ability to accurately categorize authentic photos is referred to as a real loss.
- Generated Loss: This metric measures the discriminator's accuracy in correctly categorizing generated images.

The discriminator loss function is calculated in (8).

$$L_{disc} = L_{real} + L_{generated} \quad (8)$$

where $L_{real}$ is the real loss term, and $L_{generated}$ is the generated loss

### C. GFP-GAN Model

Generative Facial Parsing (GFP) is a specific adaptation of generative adversarial networks (GANs).

The modified STF CycleGAN model generates images that are input into the GFP-GAN model. The GFP-GAN can be used to fix images by identifying small imperfections, such as scratches and blemishes. Fig. 6 shows the inpainting techniques used to fix the damaged regions [25]. The face restoration method uses GFP with innovative channel-split spatial feature transform layers, striking an optimal balance between authenticity and accuracy. The GFP-GAN successfully upgraded colors and restored facial details with a single forward pass, owing to its sophisticated designs and robust generative facial prior. Simultaneously, GAN inversion techniques necessitate costly image-specific optimization [26].

### D. Performance Evaluation

We employed modified STF CycleGAN neural networks, which were trained and tested on a dataset consisting of human sketches paired with their corresponding images. To assess the similarity between the generated image and the original image, we utilize the following parameters to evaluate the performance.

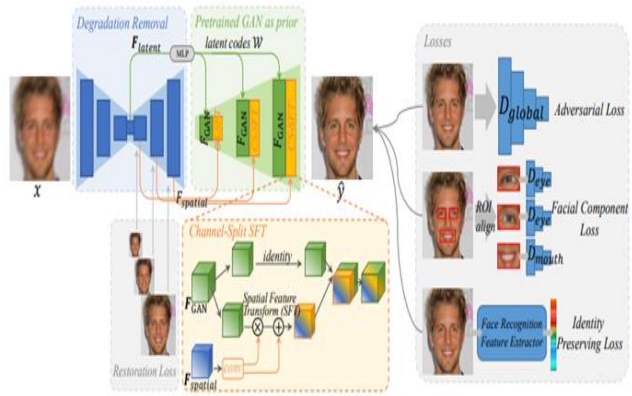


Fig. 6. GFP-GAN framework [26].

The SSIM (Structural Similarity Index Measure) is a technique used to estimate the perceived quality of photos, digital images, and videos [27]. SSIM is mostly employed to quantify the resemblance between two photographs. The SSIM values range from 0 to 1, with a value of 1 indicating a perfect match between the reconstructed image and the original image. The SSIM metric determines the degree of alignment or similarity in pixel density values between two photographs. The calculation is performed in the following manner:

$$\text{SSIM}(x, y) = \frac{(2\,\mu_x \mu_y + C1)(2\sigma_{xy} + C2)}{(\mu_x^2 + \mu_y^2 + C1)(\sigma_x^2 + \sigma_y^2 + C2)} \quad (9)$$

Where µx the mean of x

- $\mu_y$ the mean of y

- $\sigma_x^2$ the variance of x

- $\sigma_y^2$ the variance of y

- $\sigma_{xy}$ the cross-correlation of x and y

- c1, c2 Constants are added into the formula to provide numerical stability, particularly when the denominator terms (variances and covariances) are close to zero.

## IV. Experimental Result

The system operates on the Google Colab TPU runtime, which comprises an Intel Xeon CPU operating at a frequency of 2.30 GHz, 13 GB of RAM, and a cloud TPU with a computing capacity of 180 teraflops. The computer vision system operates on Python 3.10 and utilizes OpenCV.

### A. Dataset

The dataset used in this study is the CUHK Face Sketch database (CUFS) for research on face sketch synthesis and recognition. It includes 188 faces from the Chinese University of Hong Kong (CUHK) student database, 123 from the AR database, and 295 from the XM2VTS database. The images were taken during a video recording session. In total, there are 606 faces. For each face, an artist draws a sketch based on a photo taken in a frontal pose under normal lighting conditions with a neutral expression. The CUHK Face Sketch FERET Database (CUFSF) is used for face sketch synthesis and recognition research. It includes 1,194 individuals from the FERET database. The augmentation technique involves rescaling the photos in the Chuck database, resulting in 2256 images for both the sketch and target images. Combining sketches and photographs allows for more realistic training as the model learns to translate artistic representations into lifelike images, crucial for forensic applications.

### *B. Experiment*

We partitioned the dataset into two subsets, allocating 80% for training and 20% for testing the modified STF CycleGAN model. The input for the modified STF CycleGAN model is an image with a shape of (256, 256, 3). We then downsample this picture using filters of sizes (64, 128, 256, 512, 512, 512, and 512). Additionally, the image is upsampled using filters of sizes 512, 512, and 512, with dropout applied at a rate of 0.5. The size of the filter employed is 4. The modified STF CycleGAN model's discriminator compresses the input image with a shape of (256, 256, 3) using filters (32, 128, 256, 512). We train the model on the dataset for a total of 100 epochs. The model's SSIM yielded a value of SSIM is 0.637. Fig. 7 displays the outcome of the modified STF CycleGAN model.


Fig. 7. Result of modified STF CycleGAN model.

The dataset has been partitioned into two subsets, with 80% allocated for training and 20% for testing. The model is trained for 500 epochs using the dataset. The model yielded a result of SSIM is 0.8154 when the generated images from modified STF CycleGAN were used as input for GFP-GAN to recover the images and produce high-quality output images. Fig. 8 displays the outcome of the modified STF CycleGAN algorithm when applied to the GFP-GAN model.

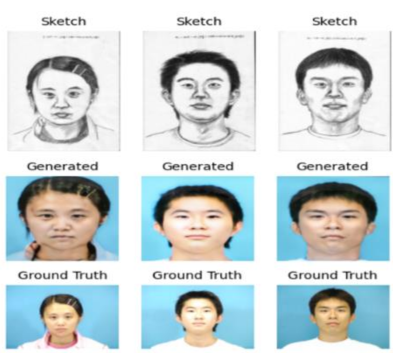


Fig. 8. Result of modified STF CycleGAN with GFP-GAN model.

## V. DISCUSSION

In this study, we developed a deep-learning model using modified STF CycleGAN with GFP-GAN to generate realistic faces from hand-drawn sketches, which helps law enforcement authorities identify and arrest criminals. The model outperformed traditional methods, achieving higher visual accuracy and preserving key facial features with an average SSIM score of 0.8154. The system uses CUFS, CUFSF, AR, and XM2VTS are public and available datasets. We use a modified STF CycleGAN model to acquire high-level facial representations and enhance the quality of the generated images. However, the model faced challenges when processing incomplete or low-quality sketches, and it struggled with extreme facial poses or expressions, highlighting the need for more diverse training data. Our system is better than other models GAN and DCGAN [10, 12]. It also outperforms PS2-MAN and contextual generative adversarial networks with GFP GANs [15, 16]. Despite the limitations, this model has practical potential in fields like forensic sketching and digital art, where accurate facial generation from sketches is essential. Developing a sketch-to-face generation system has promise for advancing techniques in forensic sketching, character design, and digital art, with potential applications in law enforcement and personalized creative tools. This technology could revolutionize how hand-drawn sketches are used in practical settings, providing more accurate visualizations for identification, design, and artistic expression.

TABLE I. PERFORMANCE COMPARISON OF DIFFERENT MODELS

| | **Method** | **Dataset** | **SSIM** |
|---|---|---|---|
| **[10]** | Generative Adversarial Network (GAN) | (CUFSF) | 0.5517 |
| **[12]** | DCGAN (Deep Convolutional Generative Adversarial Network) | (CUFSF) | 0.587 |
| **[15]** | Multi-adversarial networks (PS2 - MAN) | (CUFS), AR database, and the XM2VTS database | SSIM (Photo Synthesis) is 0.7915, SSIM (Sketch Synthesis) is 0.6156 |
| **[16]** | contextual generative adversarial network with GFP GANs | (CUFS), the XM2VTS database, and the AR database | 0.78 |
| **Ours** | modified STF CycleGAN | (CUFS), the XM2VTS database, and the AR database | 0. 637 |
| **Ours** | modified STF CycleGAN with GFP-GAN | (CUFS),(CUFSF), the XM2VTS database, and the AR database | 0. 8154 |

## VI. CONCLUSION AND FUTURE WORK

In conclusion, this research paper presents a significant advancement in forensic facial reconstruction by utilizing a modified STF CycleGAN that effectively transforms hand-drawn sketches into high-fidelity colored images, enhancing the identification process of suspects in criminal investigations. Additionally, using GFP-GAN enhances the quality of images generated from modified STF CycleGAN. The study uses the CUHK Face Sketch database (CUFS) and other datasets such as CUHK, CUFSF, FERET, XM2VTS, and AR. The study involves assembling a comprehensive collection of 2,256 images for training and testing the model. The modified STF ycleGAN achieves a Structural Similarity Index Measure (SSIM) of 0.8154, demonstrating a high level of reconstructioned image compared to previous models. Future directions include further exploring the integration of modified StyleGAN with GFP-GAN to improve the Structural Similarity Index Measure and quality of generated images. Overall, this research underscores the potential of advanced generative models in aiding law enforcement and enhancing the efficiency of criminal investigations.

## REFERENCES

[1] B. Cao, N. Wang, J. Li, Q. Hu, and X. Gao, "Face photo-sketch synthesis via full-scale identity supervision," *Pattern Recognition*, vol. 123, pp. 107654, Apr. 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320321006221

[2] G. Grana and J. Windell, *Crime and intelligence analysis*. 2021. doi: 10.4324/9781003005346.

[3] IEEE Journals & Magazine, "Face sketch recognition," *IEEE Xplore*, vol. 92, no. 1, pp. 34-45, Jan. 2004. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/1262031/

[4] N. M. Farid, M. S. Fard, and A. Nickabadi, "Face sketch to photo translation using generative adversarial networks," arXiv preprint arXiv:2110.12290, Oct. 23, 2021. [Online]. Available: https://arxiv.org/abs/2110.12290

[5] A. Aglasia, A. Adimas, S. Y. Irianto, S. Karnila, and D. Yuliawati, "Image Sketch Based Criminal Face Recognition Using Content Based Image Retrieval," *Scientific Journal of Informatics*, vol. 8, no. 2, pp. 177-188, 2021.

[6] IEEE Conference Publication, "Face photo recognition using sketch," *IEEE Xplore*, 2002. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/1038008/

[7] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-To-Image Translation Using Cycle-Consistent Adversarial Networks," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017. [Online]. Available: http://openaccess.thecvf.com/content_iccv_2017/html/Zhu_Unpaired_Image-To-Image_Translation_ICCV_2017_paper.html

[8] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Lecture Notes in Computer Science*, vol. 9351, pp. 234-241, Jan. 2015. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-24574-4_28

[9] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-To-Image Translation With Conditional Adversarial Networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. [Online]. Available: http://openaccess.thecvf.com/content_cvpr_2017/html/Isola_Image-To-Image_Translation_With_CVPR_2017_paper.html

[10] IEEE Conference Publication, "Improving Face Sketch Recognition via Adversarial Sketch-Photo Transformation," *IEEE Xplore*, May 01, 2019. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8756563/

[11] IEEE Conference Publication, "Crime Investigation using DCGAN by Forensic Sketch-to-Face Transformation (STF)- A Review," Apr. 8, 2021. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9418417/

[12] S. Devakumar and G. Sarath, "Forensic Sketch to Real Image Using DCGAN," *Procedia Computer Science*, vol. 203, pp. 101-108, Jan. 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050923001394

[13] IEEE Journals & Magazine, "Face Photo-Sketch Synthesis and Recognition," *IEEE Xplore*, vol. 97, no. 11, pp. 105-112, Nov. 2009. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/4624272/

[14] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre, "XM2VTSDB: The extended M2VTS database," in *Proc. Second Int. Conf. Audio and Video-based Biometric Person Authentication*, vol. 964, pp. 965-966, 1999.

[15] IEEE Conference Publication, "High-Quality Facial Photo-Sketch Synthesis Using Multi-Adversarial Networks," *IEEE Xplore*, May 1, 2018. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8373815/

[16] S. Nagpal, "Sketch-to-Face Image Translation and Enhancement Using a Multi-GAN Approach," *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, vol. 10, no. 12, pp. 111-119, Dec. 2022.

[17] S. Gunjate, T. Nakhate, T. Kshirsagar, Y. Sapat, and S. Guhe, "Sketch to Image Using GAN," *International Journal of Innovative Science and Research Technology*, vol. 8, no. 1, pp. 45-52, Jan. 2023.

[18] IEEE Journals & Magazine, "A Generalized Laplacian of Gaussian Filter for Blob Detection and Its Applications," *IEEE Xplore*, vol. 101, no. 12, pp. 1356-1365, Dec. 2013. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6408211/

[19] S. Misra and Y. Wu, "Machine learning assisted segmentation of scanning electron microscopy images of organic-rich shales with feature extraction and feature ranking," in *Elsevier eBooks*, vol. 35, pp. 205-221, Jan. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/B9780128177365000107

[20] Y. Yang, W. Zhang, J. Wu, W. Zhao, and A. Chen, "Deconvolution-and-Convolution Networks," arXiv preprint arXiv:2103.11887, Mar. 22, 2021. [Online]. Available: https://arxiv.org/abs/2103.11887

[21] IEEE Conference Publication, "Batch Normalization in Convolutional Neural Networks — A comparative study with CIFAR-10 data," Jan. 1, 2018. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8470438

[22] K. Yasaka, H. Akai, A. Kunimatsu, S. Kiryu, and O. Abe, "Deep learning with convolutional neural network in radiology," *Japanese Journal of Radiology (Print)*, vol. 36, no. 3, pp. 256-269, Mar. 2018. [Online]. Available: https://link.springer.com/article/10.1007/s11604-018-0726-3

[23] IEEE Conference Publication, "Reluplex made more practical: Leaky ReLU," Jul. 1, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9219587/

[24] D. Ak and V. Jain, "Comparative Study of Convolution Neural Network's Relu and Leaky-Relu Activation Functions," in *Lecture Notes in Electrical Engineering*, vol. 539, pp. 432-439, Jan. 2019. [Online]. Available: https://link.springer.com/chapter/10.1007/978-981-13-6772-4_76

[25] A. Kumari, R. K. Dubey, and S. K. Mishra, "A Cascaded Method for Real Face Image Restoration using GFP-GAN," *International Journal of Innovative Research in Technology and Management*, vol. 6, pp. 23-32, 2022.

[26] X. Wang, Y. Li, H. Zhang, and Y. Shan, "Towards Real-World Blind Face Restoration With Generative Facial Prior," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. [Online]. Available: http://openaccess.thecvf.com/content/CVPR2021/html/Wang_Towards_Real-World_Blind_Face_Restoration_With_Generative_Facial_Prior_CVPR_2021_paper.html

[27] M. Chen and A. C. Bovik, "Fast structural similarity index algorithm," *Journal of Real-Time Image Processing*, vol. 6, no. 4, pp. 243-254, Aug. 2010. [Online]. Available: https://link.springer.com/article/10.1007/s11554-010-0170-9.