



Text-to-Sketch Synthesis via Adversarial Network

Jason Elroy Martis¹, Sannidhan Manjaya Shetty^{2,*}, Manas Ranjan Pradhan³, Usha Desai⁴ and
Biswaranjan Acharya^{5,*}

¹Department of Information Science and Engineering, NMAM Institute of Technology (Nitte Deemed to be University), Nitte, India

²Department of Computer Science and Engineering, NMAM Institute of Technology (Nitte Deemed to be University), Nitte, India

³School of Computing, Skyline University College, University City, Sharjah, UAE

⁴Department of Electronics and Communication Engineering, SR University, Warangal, Telangana, India

⁵Department of Computer Engineering–AI & BDA, Marwadi University, Rajkot, Gujarat, India

*Corresponding Authors: Sannidhan Manjaya Shetty. Email: sannidhan@nitte.edu.in; Biswaranjan Acharya. Email: biswaacharya@ieee.org

Received: 31 December 2022; Accepted: 18 April 2023; Published: 09 June 2023

Abstract: In the past, sketches were a standard technique used for recognizing offenders and have remained a valuable tool for law enforcement and social security purposes. However, relying on eyewitness observations can lead to discrepancies in the depictions of the sketch, depending on the experience and skills of the sketch artist. With the emergence of modern technologies such as Generative Adversarial Networks (GANs), generating images using verbal and textual cues is now possible, resulting in more accurate sketch depictions. In this study, we propose an adversarial network that generates human facial sketches using such cues provided by an observer. Additionally, we have introduced an Inverse Gamma Correction Technique to improve the training and enhance the quality of the generated sketches. To evaluate the effectiveness of our proposed method, we conducted experiments and analyzed the results using the inception score and Frechet Inception Distance metrics. Our proposed method achieved an overall inception score of 1.438 ± 0.049 and a Frechet Inception Distance of 65.29, outperforming other state-of-the-art techniques.

Keywords: Generative adversarial networks; inverse gamma correction; sketch attributes; text-to-sketch synthesis; deep learning techniques

1 Introduction

Sketch depictions have an enormous application covering larger scope that includes painting and poetry to forensic sciences. In the olden days, criminals were identified by producing sketches using sketch experts based on the oral descriptions given by the eyewitness on the scene. With the advancements in artificial intelligence that have evolved to another level, the application of sketches by law enforcement agencies have drastically changed for identifying convicts through sketch matching



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

[1–3]. The produced sketch is matched against the set of mugshots available in the dataset using standard classification techniques. However, through the abundant research studies, the technique proved inefficient for two reasons: (1) discrepancies and misinterpretation during oral communication between the user and the artist. (2) Inaccurate classifications due to the involvement of cross-domain images involving sketch and photographic mugshots.

Generative Adversarial Networks (GANs) are a type of deep learning algorithm that can be utilized to generate new data that resembles the training data. GANs consist of two neural networks, namely, the generator and the discriminator, who are trained in an adversarial way. The generator network is responsible for creating fresh data samples that match the training data, while the discriminator network's role is to differentiate between the real and generated data. Both networks work together during training, where the generator network attempts to produce data that can deceive the discriminator network into believing it is genuine data, while the discriminator network strives to identify which data is authentic and which is artificial accurately. With the introduction and massive success of GANs, sketch identification took a different dimension that overcame the fallacies involved in the early stages of its implementation. Many research exertions have implemented sketch-to-photo-matching techniques using GANs to generate color photo images from sketches supplied as input. This technique paved the way for classifying images in the same domain, avoiding cross-domain involvement. Even then, there is a challenge in generating the sketches required as input to the GANs. Generation of sketches primarily suffer from discrepancies due to inaccurate depictions and artist's misinterpretation [4–6].

Research advancements in GANs exposed the implementation of text-to-image GANs to produce images based on the textual description supplied as input. These GANs also considerably produced facial images based on the textual description defining the facial attribute sets. Few standard works proposed different techniques of text-to-facial photograph synthesis that involved generating color photographs of human faces based on facial feature descriptions [7,8]. This technique of generating facial photographs from the visual description of the face can be extended to the forensic sciences to identify convicts. However, direct generation from text-to-facial photographs involves difficulty in mapping the descriptions related to color, race, and ethnicity information. Since eyewitnesses only provide higher-level abstract definitions of facial cues, the ability to develop GANs from textual cues to sketches paves a path for creating a new variant of GANs to produce human sketches. One can later utilize these sketches with the help of various sketch-to-photo image GANs to generate color photo images.

With the challenges in generating facial photographs, there is a scope for implementing text-to-sketch GAN to generate facial sketches from the abstract textual descriptions supplied as input. The abstract textual description involves the description of facial attributes such as “long nose”, “wide chin”, “small eyes”, etc. Fig. 1 shows a typical sample of a generated sketch corresponding to its textual description.



This woman has straight hair with an expanded nose, medium eyes and a plain forehead.

Figure 1: Sample of sketch generated for typical textual description utilized from the Chinese University of Hong Kong (CUHK) dataset

The complexities encountered in text-to-facial photograph GANs [9] have motivated the proposal of a text-to-sketch GAN in this research article. This approach aims to generate facial sketches based on textual descriptions, as demonstrated in Fig. 1. To achieve this goal, a structured GAN with the employment of a generator and comparator module. The generator module creates sketches from contextual text descriptions, and the comparator module compares them to enhance the training process. The textual descriptions are based on the abstract representation of the facial attributes defining the structure of the face covering primary features [10–12]. Since the ability of GAN depends on the quality of its training, it is essential to train the GAN with improved quality sketches. Hence, to improve the training process, this work adopts the utilization of a color enhancement procedure based on an Inverse Gamma Correction Technique (Inverse GCT) that generates enhanced sketches from the color photo images. In order to validate the capacity of the proposed system, sufficient experimental evaluations were conducted on the benchmark datasets open to the research community.

Concerning the design of the proposed system, this article has succeeding contributions:

1. Design of a structured GAN that can generate facial sketches via abstract textual description of the face.
2. Training of the structured GAN effectively using the enhanced sketches generated via Inverse GCT.
3. To validate the ability of our GAN using standard evaluation metrics.

2 Related Works

GANs are neural network-based translators that translate (generate) data from one form to another with a minimal amount of training samples. An in-depth overview of the working concept of GANs in articles [13–15] pointed out its applications in different areas, from image synthesis to textual representations. Researchers in the article provided a detailed understanding of two important modules: generator and discriminator functionalities concerning the training data samples. Hence, it motivated us to explore GANs further to identify the areas of applications that utilize varieties of adversarial networks. Researchers of the articles [16–18] proposed the implementation of a conditional GAN that would translate an image from one form to another based on the training criteria. The Pix2pix model was the main component involved that would generate a colored image for any other image given as output. The experimental outcomes presented in their article were encouraging, with minimal training loss. Article [19] conducted a work for evaluating the pix2pix model for generating facial photographs from sketches. Their investigations concluded that training samples with enhanced quality improved the performance of GAN. Researchers have conducted numerous works implementing GANs with various techniques to generate facial photo images from sketches. Cycle GAN [20] is one such variant that is capable of generating a photograph from a facial sketch and vice versa. Composition Aided-GAN (CA-GAN) and Spatial Content Alignment-GAN (SCA-GAN) are utilized in combination in the article [21] to generate color photo images and to minimize loss appearing based on pixel-wise labelling of sketch images. Sketch-GAN [22] and style transfer GAN [23] are other notable works that generate photo images from facial sketches.

Advancements in GANs further spurred the discovery of a new type of translator to transform textual description into an image form. Different areas have covered various applications that use these types of GANs. Dynamic Memory-GAN (DM-GAN) [24] is one of the popular works that adopted text-to-image synthesis to improve the visual quality of the generated image based on textual descriptions. Further investigation led to exploring Attentional GAN (AttnGAN) [25], also referred to as an attentional generative adversarial network. AttnGAN focused on producing facial images

through attentional descriptions in naturally spoken words. Research papers [26–29] proposed and discussed the ability of Deep Fusion (DFGAN), a variant of GAN composed of a deep fusion network. DFGAN performed the text-to-image generation via abstract structural representation of an image. The technique engendered images of different types ranging from birds to human faces. The outcomes presented in the articles justified the ability of DFGAN to perform better with minimal description than any state-of-art techniques.

A GAN's performance equally depends on the quality of the training samples. A higher quality of training images also leads to the generation of better-quality images. Article [30] proposed a technique to generate facial photos from the sketch input using the conditional GAN (cGAN) model. In their work, to enhance the performance of their model, authors have incorporated a gamma correction-based preprocessor to generate enhanced quality sketches from color photo images. Another work carried out in research [31] proposed a novel technique of face sketch synthesis using the Internet of Things for the digital forensic system. The authors of this research have utilized gamma correction as a preprocessor enhancing the contrast to generate better sketches. A cGAN model [32] engendered Indonesian photo images from the sketches. To improve the performance of their model in the training stage, the authors have pre-processed the sketches by utilizing gamma correction inversely to stress the edges and make them appear darker.

3 Proposed System

As portrayed in Fig. 2, our proposed system has four basic modules, namely (1) the Image retrieval process, (2) the Sketch creation stage, (3) the Attribute description, and (4) the sketch generation and comparison stage. The following sections explain in detail the process of each module.

3.1 Image Retrieval Process

This is the first stage of our process responsible for grabbing a color photo of a human face in red-green-blue (RGB) format. To ensure the quality of the images, the stage ensures that they maintain a minimum resolution of 600 dpi and are in a 24-bit bitmap image format. The image contains all features of the human face, from hair to chin, ignoring the neck region. It is then passed on to the next stage.

3.2 Sketch Generation Stage

The ultimate purpose of this stage is to generate sketches from color photographs. The sketch generation process drives through a series of steps that include mathematical image morphological operations and spatial image transformations. Fig. 3 explains the process of sketch generation.

The process as depicted in Fig. 3 has four stages named as 1. Color elimination, 2. Transformer 3. Merging, and 4. Gamma Adjuster. The subsequent subsections explain how different stages work.

3.2.1 Color Elimination Stage

This stage is a preprocessor filter for the subsequent stages [33]. The output of this stage eliminates any known color to produce a grayscale output having a maximum of 255 levels. Every pixel in this stage is normalized using the color averaging technique as presented in Eq. (1).

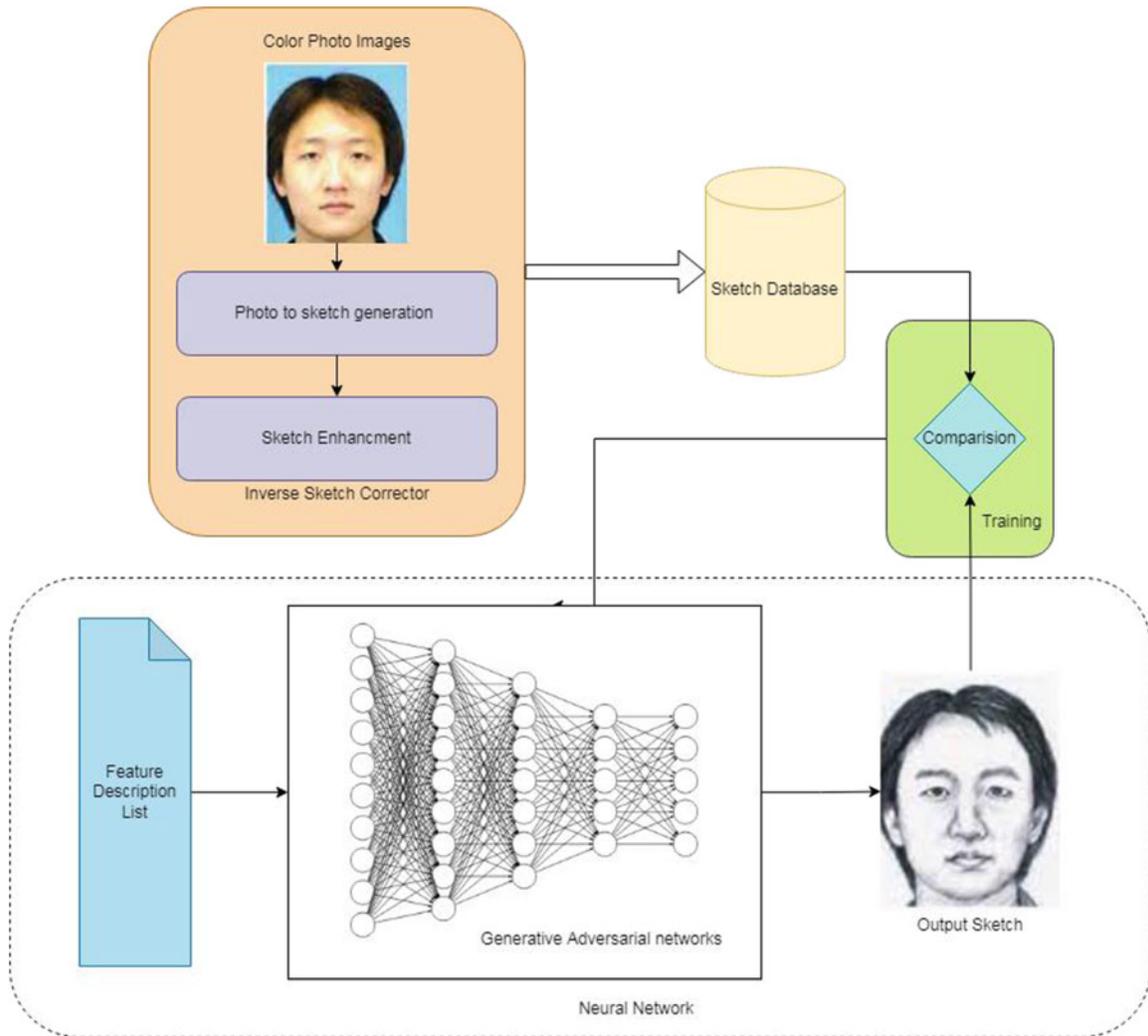


Figure 2: Architecture of our proposed system with functional modules

$$Grayscale_i = \frac{Red_i + Blue_i + Green_i}{3} \tag{1}$$

where $Grayscale_i$ represents the grayscaled averaged image pixel Red_i , $Blue_i$, and $Green_i$ are the pixel values of the red, blue and green channels, respectively.

3.2.2 Transformer Stage

The sketch generator phase relies heavily on the transformer stage, which plays a crucial role. During this stage, the inverse transformation of each known pixel is computed to obtain its corresponding inverse value. It “s” worth noting that pixels cannot have negative values, so a value equivalent to the largest identified bit precision is added. To better understand this process, one can refer to Eqs. (2) and (3), which demonstrate the steps involved in inverse transformation.

$$g_i^{-1} = 2^{bitprecision} - g_i \tag{2}$$

$$Pix_{new} = 255 - Pix_{old} \quad (3)$$

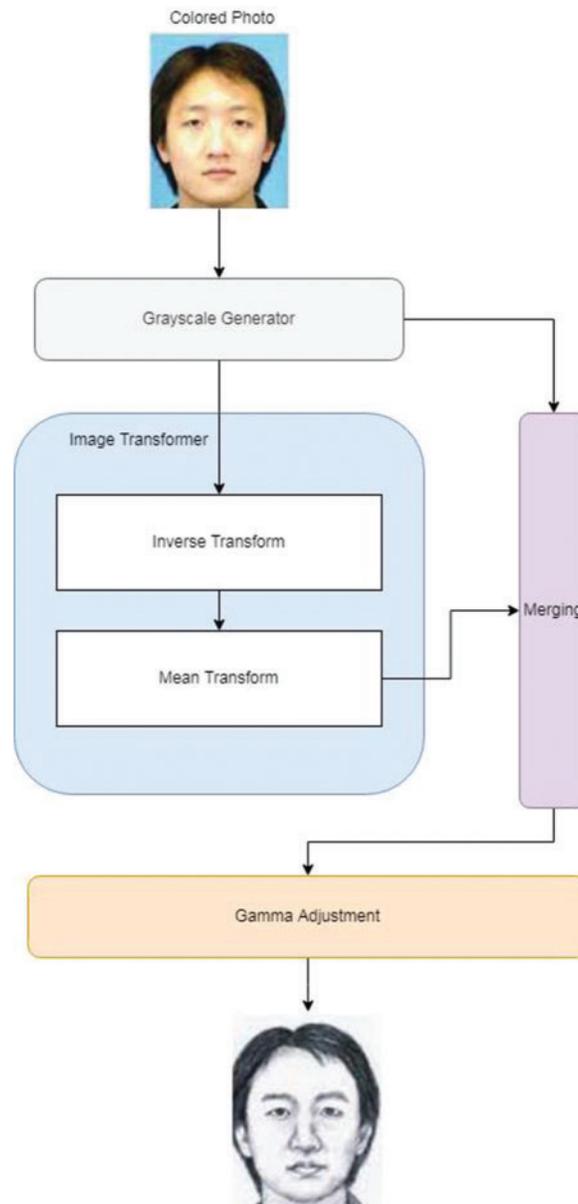


Figure 3: Workflow of Inverse GCT sketch generator

g_i^{-1} is the inverse value of each bit in the entire image., Pix_{new} and Pix_{old} denote newer and older values of the pixels. Value 255 is chosen because we use a 24-bit RGB color image that gives eight bits when divided among all three channels. The total number of gray levels for eight-bit systems is 255, corresponding to *bitprecision* in the equation. Hence, in our case, the value of *bitprecision* is 8. The next step involves a convoluting mean applied filter throughout the image. The major purpose of this step is to enhance edge highlighting to boost the sketch quality in later stages. The application of the mean

filter includes neighbouring values applied across the image. We have chosen a range of nine pixels for creating a 3×3 filter called a kernel area. The abstract mathematical transformation is shown in Eqs. (4) and (5), respectively.

$$\forall g^{-1}(x, y) \rightarrow (\bar{x}, \bar{y}) \rightarrow P_{\mu}(x, y) \quad (4)$$

$$P_{\mu}(x, y) = \frac{\sum_{R(3 \times 3)}(x, y)}{9} \quad (5)$$

Here $P_{\mu}(x, y)$ stands for the mean value of the pixels amongst the 3×3 neighbouring region of the target. $g^{-1}(x, y)$ represents the values of each pixel after applying g_i^{-1} in Eqs. (2) and (3).

3.2.3 Merging of Grayscale and Transformed Image

As the name suggests, this step accepts input from both the original grayscale image as well the transformed image to create a merged blend for the corresponding pixel. Eq. (6) shows this effect

$$Pixel_{merged} = \tilde{f}(g_i^{-1}(x, y), P_{\mu}(x, y)) \quad (6)$$

In Eq. (6), $Pixel_{merged}$ represents the concatenation operation after applying the function \tilde{f} . There are several ways to perform the blend operation, denoted as $Pixel_{merged}$. Many research exertions focus on using the $*$ operation and the addition modulus operation. In our case, we must superimpose the edges to create a vibrant sketch. Achieving this requires the implementation of a division operation. Hence the function \tilde{f} will be chosen as a division function reflected in Eq. (7).

$$\tilde{f} = \frac{g_i^{-1}(x, y)}{P_{\mu}(x, y)} \quad (7)$$

The output of this $Pixel_{merged}$ highlights streaks of hair and certain thick lines found only in sketches. Overall, the blend operation scrapes off thin lines and keeps only thick lines. The output of this step includes a dull sketched image. The highlight of the entire process is shown pictorially in Fig. 4.

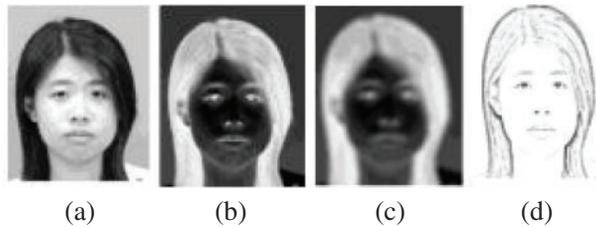


Figure 4: Pictorial representation showing the outcomes of the merging process

The merging process outcomes are illustrated in Fig. 4 through a pictorial representation. The figure shows that the process involves four steps: Firstly, the grayscale version of the image is obtained and represented in (a). Secondly, the negative version of the grayscale image is obtained from (a) and represented in (b). Thirdly, (b) is blurred to produce a smoothed version, represented in (c). Finally, the division operation is performed between (b) and (d) to obtain the final sketch, which is represented in (d).

3.2.4 Gamma Adjustment Stage

It is commonly understood that an artist's initial sketch often consists of just highlights and rough lines. Therefore, it is important for the final image to maintain a hand-drawn appearance while minimizing the use of image processing techniques. To achieve this effect, we avoid utilizing standard image processing techniques that may wash out the image and instead brighten only the necessary highlights and streaks. This can be accomplished by utilizing inverse gamma values, which darken the image instead of brightening it like gamma correction. Eq. (8) outlines the process for gamma correction [34,35].

$$Sketch_{(x,y)} = (Pixel_{merged})^{\gamma^{-1}} \quad (8)$$

Here γ^{-1} is the inverse gamma value applied to the blended image denoted as $Pixel_{merged}$. By conducting excessive trial and error, we have affixed the gamma value to 2.2.

3.3 Attribute Description

An eyewitness usually gives a verbal description to describe the victim's details. The description includes the entire details of the face, like eyes features, nose, and mouth features. The features described by the witness are abstract. Let us consider attribute descriptions taken from an eyewitness "The person had a Mouth Open with an upward curve on the face, eyes are small and partially closed, eyebrows are standard, and he had a wrinkled forehead". These features mentioned above are subjective but portray much information. Similar statements like these create the following categorizations [36–39].

3.3.1 Mouth

- The mouth can be divided into two primary components: the inner portion that displays the tongue and teeth and the outer portion that displays the lips. In our case, the tongue is disregarded to prevent any complexities and solely focus on the teeth, resulting in the following outcome. Mouth State: This category shows whether the teeth are seen, or it is invisible. If seen, it is classified as OPEN else; it is shown as CLOSED.
- Lip curve: The mouth consists of the lips divided into two parts: the upper and the lower labia. The categorization is done based on the shape of the labia form with each other. The classifications are

1. NEUTRAL: If the lines between the labia form a straight line
2. UPPER CURVE: If the line formed between the labia forms an upward curve like a smiley.
3. DOWNWARD CURVE: If the line formed between the labia falls towards the chin.

3.3.2 Eyes and Eyebrows

The right eye is separated between the nasal ridge and is independent of each other except in viewing. We perform this to aid binocular precision viewing. We can categorize the left and right eye separately, but we perform them in unison for simplicity. Eyebrows are hair-like structures that follow the eye contour. The eyebrow cannot move as the eye but signifies movements that are in association.

- Eye size: An open eye has a pre-determined size and is classified as NEUTRAL. If the eye size exceeds a pre-determined range, then the neutral is classified as LARGE. If the eye size falls behind the NEUTRAL, they are classified as SMALL.

- **Eye Brow position:** A standard brow shape has a characteristic droop shape classified as **NEUTRAL**. If the position crosses the relative position above the normal, it is classified as **HIGH**, and the below normal is **LOW**.
- **Forehead:** The forehead is the area of the face that lies between the scalp and the eyebrows. It consists of three sets of muscles controlled by the facial nerve. The muscles lie beneath a layer of fat, which creates wrinkles when compressed. The classification is based on the creation of the Wrinkles, i.e., **WRINKLED** or **PLAIN**.

3.3.3 *Nose and Nasal Ridges*

The nose is present between the eyes and the mouth extending from a range of 3.5 cm from the eyebrows. The nose muscles cannot be controlled consciously and move in unison together. Nasal ridges are the lines that lie between the nose and mouth, occurring tangentially. These are microexpressions that move in synchrony along with the nose.

- **Nose Length:** Here, the nose muscles are pushed above, creating a plethora of ridges between the eyes and the nose. These are distinct features. Whereas the nasal ridges are diminishing visible, they are absent in some individuals. The categories of Nose length are **SHRUNK** and **NORMAL**.
- **Nasal width:** The nose width determines the nostril size of the person. The nostrils can expand and contract and varies between individuals. Note that the nostrils cannot shrink in size unless there is an infection. The nostrils width is categorized into two categories, namely **EXPANDED** and **NORMAL**.

3.3.4 *Chin*

The chin is the most prominent area of the face. It is defined biologically as the protrusion part of the face below the mouth. The chin varies for different races of people, having mixed variances. Following are different categories of chin amongst the people.

- **SQUARE:** The majority of people have a square-shaped chin in which the jawline creates a parallel majorly line against the mouth.
- **POINTED:** This chin type construes with a second majority of people. Here the chin makes a pointed curve giving a sleeker finish.
- **LONG:** A rarity chin type among people creates a longer distance between the mouth and tip of the face. People having this possess a stronger lower jawline.
- **PROTUDING:** The chin protrudes until it touches the lips' virtual level. This type is a rarity amongst the Indian populace.
- **DIMPLED:** This a unique characteristic where the chin creates a fine line between the symmetry of the face.

3.3.5 *Gender*

The gender description determines the sexual orientation of the person. The options are either Male or Female. It is the most prominent feature of the human body, but it is kept at last due to the appealing effects added to the face in terms of human face generation. As discussed earlier, [Table 1](#) summarizes the complete set of feature-wise classifications.

Table 1: Classification of distinctive features

Feature	Division	Category
Mouth	Mouth state	OPEN and CLOSED
	Lip curve	Neutral, Uppercurve, Lowercurve
Eyes	Eye size	Large, Small, Neutral
	Eyebrow position	Neutral, High, Low
Forehead		Wrinkled, Plain
Nose and nasal ridges	Nose length	Shrunk, Normal
	Nasal width	Expanded, Normal
Chin		Square, Pointed, Long, Protruding, Dimpled
Gender		Male, Female

3.4 Sketch Generation and Comparison Stage

Our sketch generator consists of a structured GAN, which accepts contextual valid text input from the classes mentioned in Section 3.3 and fuses these characteristics together to form an image. Since these images are related to faces, it becomes predominant for the system to train itself with a large container of faces. Our GAN utilizes a deep fusion architecture containing two major neural networks that work together, contradicting each other. The generator network creates a raw image to dupe the discriminator, which later finds out a difference which is confirmed with a score. The process continues until the discriminator is entirely deceived. The subsequent sections explain the detailed structure of the generator and the discriminator.

3.4.1 Generator

The generator stage is a cross-domain neural network that maps the textual information domain to the image domain. This network is unique because it works based on additive noise generation, creating text to images. [Table 2](#) portrays a brief layout of the Generator architecture.

Table 2: Complexity level of proposed generator

Layer ID	Layer type	Number of hyper parameters	Description
Input	Text encoder	300976	The sequence of encoded texts
	Noise vector	0	Noise generator based on Gaussian samples
1	Fully connected	54420	Completely connected layer reshaped to 4×4 Blocks

(Continued)

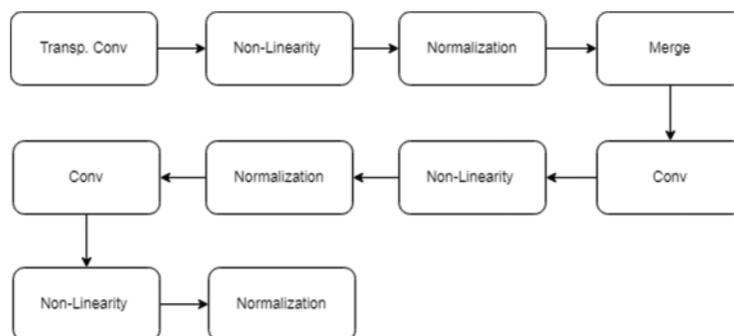
Table 2: Continued

Layer ID	Layer type	Number of hyper parameters	Description
2	UP layer		A Layer consisting of Up-sample
3	UP layer		residual and Deep Fusion
...	...	$71296 * 4 = 285184$	(DF) Blocks.
6	UP layer		
7	Reverse convolution layer	481020	Feature Conversion to Images
8	Reverse convolution layer	481020	

From [Table 2](#), we can infer that four distinct layers need explanation. They are UP-sample residual Block, DF Block and Reverse Convolved layer. Explanations of the layers are as follows.

- **Up-sample Residual Block:**

The Up-sample block consists of three main modules: The Up sampled net and the two fusion nets. The primary function of the Up sampled net is to increase the number of dimensions. The Up sampled net consists of ten specifically designed neural layers that perform a specific job which is self-explanatory. [Fig. 5](#) explains the structure of the Upsample block.

**Figure 5:** Structure of the Up-Sample Block

The Fusion net is a simple concatenation function combining various feature dimensions into smaller dimensions. In short, it is a dimensionality reducer function. Our block consists of two fusion blocks because it is always necessary to follow encoded sequences. [Fig. 6](#) shows the architecture of the Up-Sample residual block.

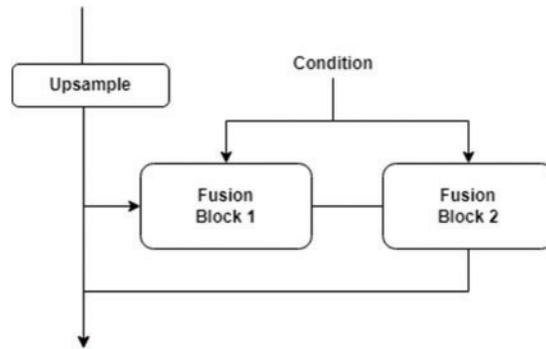


Figure 6: Structure of Up-Sample Residual Block

• Deep Fusion Block

The deep fusion block is the reason why the DF-GAN got its name. The main function of this block is to fuse all text-encoded features in the image. It accomplishes this task by increasing the number of affine blocks in the network. Fig. 7 explains the structure of the DF Block. As depicted in the figure, the architecture of the DF block consists of three unique blocks: The Affine block, the ReLU block and the Convolutional (Conv) block.

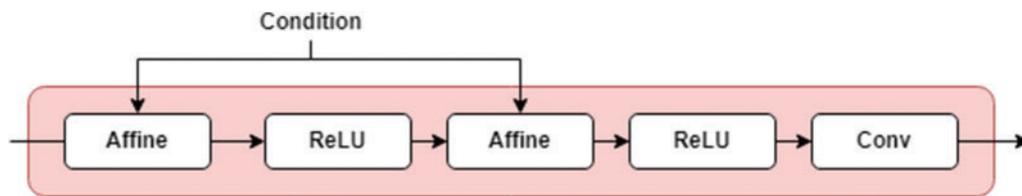


Figure 7: Structure of deep fusion block

The primary function of the affine block in deep fusion is to align and combine multiple input features at different scales or resolutions. This enables effective fusion and creates a more powerful input data representation for improved performance. ReLU (Rectified Linear Unit) block refers to a unit that applies the ReLU activation function to the input. The primary function of the ReLU block in deep fusion is to introduce non-linearity and increase the representational power of the neural network. The ReLU function is defined as $\max(0, x)$, where x is the input value. This function sets all negative input values to zero and keeps all positive values unchanged. The function of the conv layer in deep fusion blocks is to extract spatial features from the input data by applying a convolution operation. This helps to create a more informative and robust representation of the input data that can improve performance.

• Reverse Convolved Layer

The reverse convolved aliased as a convolved transposed layer is an upsampled convolved layer of higher dimensions for lowering computational complexity. The main aim of this up-sampled convolution is to map features into numbers in higher dimensions. Hence, these layers are called reverse convolved layers. If we combine these layers, they form the complete generator model of GAN. We can represent the entire structure of GAN mathematically using Eq. (9).

$$G: \{O_s, N_g\} \rightarrow O_p \quad (9)$$

where O_s is the observed features s , N_g is random noise generated from mapping sketch O_s with photo O_p . Fig. 8 depicts the entire generator procedure.

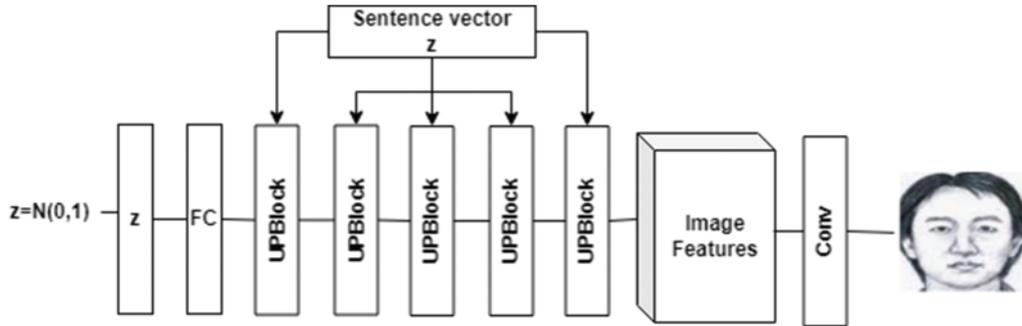


Figure 8: Different layers and their connection in DF-GAN

3.4.2 Discriminator

The discriminator is a network that contests along with the generator to disprove the generated image. Hence, the discriminator converts images into feature maps and replicates them into a concatenated sequence. The sequence is then verified to match the semantic consistency. Inertly the discriminator promotes the generator to create improved images and better consistency.

The Discriminator architecture composes of various Down blocks coupled with a convoluted layer to verify the consistency of the adversarial network. Table 3 briefs out the summary of the entire architecture.

Table 3: Complexity level of proposed discriminator

Layer ID	Layer type	Number of hyper parameters	Description
Input	Convolution	481020	Conversion of low computed image features.
1	Down layer	$71296 * 4 = 285184$	A Layer consisting of a Downsample to reduce dimensionality.
2	Down layer		
3	Down layer		
...	...		
6	Down layer		
7	Convolution layer	481020	Feature Mapping.
8	Convolution layer	481020	
9	Loss calculator	0	Matching Text aware penalty.

Table 3 presents the components of our discriminator, which include a Convolution layer, Down-sample layers, and Loss calculators. Further elaboration on these layers are subsequently discussed.

- Down Layer

As the name implies, the down sampler reduces the number of samples made by the convolution. Since the generator increases the samples by upsampling, it is evidence that the discriminator must perform the opposite.

- Convolutional Layer

This layer applies an encoded sequence repeatedly to generate feature maps as output. The main reason for applying convolution is to reduce the computation in dimensionality as seen in wavelengths. Our convolutional layers create a filter pool of $5 \times 5 \times 4$.

- Loss Calculator

The discriminator associates a loss with the help of a score indicating contextual awareness concerning the sketch. Our function calculates a score using Eq. (10).

$$L = \frac{1}{n} \sum_{k=1}^n |C_k - R_k| \quad (10)$$

Here C_k represents the contextual features generated and R_k represents the real features mentioned by the training sample. L represent the loss values obtained after applying the summation operation. n indicates the number of features by the convolution layer. We use this as a basis function to calculate the loss at the discriminator and the generator, which we explain in the next section. The discriminator network is displayed visually in Fig. 9.

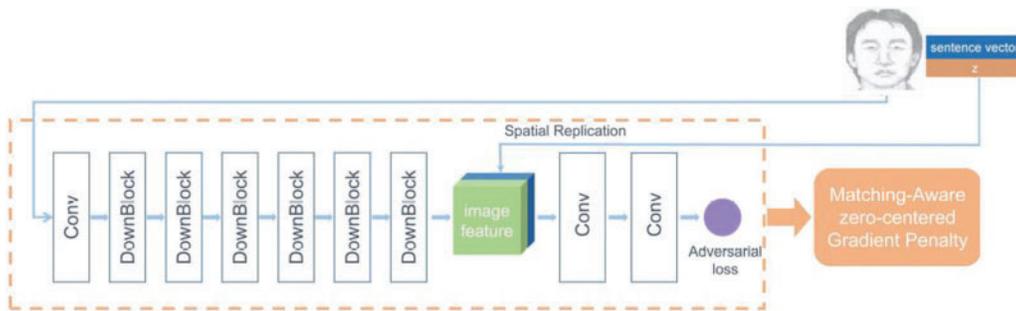


Figure 9: Different layers involved in the discriminator of our proposed DF-GAN

3.5 Generator and Discriminator Loss Function

As previously stated, the generator and discriminator engage in a competitive process to produce better quality images. This improvement occurs when the loss is minimized. As a result, it is crucial to calculate the loss of each neural network. Since the networks are interdependent, the loss functions of the individual networks rely on each other. The loss functions of the discriminator and generator are described by Eqs. (11) and (12), respectively.

$$L_D = -E_x \{ \min(0, D(x, e) - 1) \} - \frac{1}{2} [E_G \{ \min(0, -D(G(z), e) - 1) \}] \quad (11)$$

$$- \frac{1}{2} [E_x \{ \min(0, -D(x, e) - 1) \}]$$

$$L_G = -E_{G(z)} [D(G(z), e)] \quad (12)$$

From Eqs. (11) and (12), L_D and L_G represents the loss values obtained from discriminator D and generator G, respectively. Also, E determines the expected value function. Evidently, our generator network loss accepts the discriminator's score with high priority for assessment with a factor $G(z)$, indicating that the generator improvises the existing features to minimize loss. Similarly, the discriminator will enhance its precision to find minute details in the generated image. The operation continues until loss reaches a constant value. Algorithm 1 summarizes the working concept of the entire technique.

Algorithm 1: Inverse GCT Enhanced text-to-Sketch GAN

```

1: for n training samples, do
2:   for each pixel i in an image, do
3:      $G_i = \frac{Red_i + Blue_i + Green_i}{3}$ 
4:      $G_i \leftarrow 2^{bitprecision} - G_i$ 
5:     Replace each pixel in the set by transforming  $g_i^{-1}$ 
6:   end for
7:   for the set of m pixels in the entire image do
8:      $P_\mu(x, y) \leftarrow \frac{1}{9} \sum_{R(3 \times 3)}(x, y)$ 
9:     Reset every pixel in the batch by  $P_\mu(x, y)$ 
10:  end for
11:  for the entire group b in the set, do
12:    for every pixel j in batch b, do
13:       $Pixel_{merged} = \frac{g_j^{-1}(x, y)}{P_\mu(x, y)}$ 
14:      Replace each pixel in the batch by  $Pixel_{merged}$ 
15:    end for
16:  end for
17:  for every pixel j in the photo, do
18:     $p_i \leftarrow image_i^{G^{-1}}$ 
19:  end for
20:  perform a probability sample of m text samples from prior noise  $p_g(noise)$  a probability sample from the data  $p_i$ 
21:  Perform discriminator update functions:
22:   $\forall_{\theta_d} \left\{ \frac{1}{m} \sum_{i=1}^m [\log D(p_x^{(i)}) + \log(1 - D(G(noise^{(i)})))] \right\}$ 

```

(Continued)

Algorithm 1: Continued

23: perform a probability sample on the set of m noise samples for image generator updation using gradient variations

24: $\forall_{\theta_g} \left\{ \frac{1}{m} \sum_{i=1}^m \log(1 - D(G(\text{noise}^{(i)}))) \right\}$

25: end for

4 Results and Discussion

To uphold our proposed system's capabilities as well as prove the effectiveness of our system, we have deduced a series of test cases. The test cases proceed in two phases. The primary phase extrudes the ability of Inverse GCT using standard image quality assessors. The secondary phase excoriates the ability of our DF-GAN to generate better-quality images. The subsequent sections provide a detailed analysis of these two phases.

4.1 Inverse GCT

We have used two types of quality assessment to evaluate the strength of our generated sketches. The primary assessment tool uses referential quality assessors, namely Peak-Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), Mean Squared Error (MSE), Normalized Absolute Error (NAE), Naturalness Image Quality Evaluator (NIQUE), Frechet Inception Distance (FID), Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE), Deep-Convolutional Neural Network Image Quality Assessment (DIQA) [40–43]. The process involves comparing our transcribed images with our standard expected output. To evaluate our study, we have used the CUHK dataset that contains a total of 606 faces along with their corresponding sketches. We have also used the AR dataset, which contains 126 image/sketch pairs. Fig. 10 shows a sample from the CUHK dataset.



Figure 10: Three photo-sketch samples from the CUHK dataset

4.1.1 Analysis of Gamma Value

Table 4 shows the trial and error experiments performed to select a suitable γ value to generate quality sketches. To evaluate the quality of the sketch for a selected γ value, we have adopted no reference evaluator.

By analyzing the data presented in the table, it is conclusively evident that $\gamma = 2$ generates sketches of superior quality.

4.1.2 Comparison of Non-Referential Indicators

To uphold the quality of sketches generated via the Inverse GCT approach, we have compared the generated sketches with the original ones using standard no-reference evaluation metrics. Table 5

shows the values obtained for non-referential parameters. We have chosen non-referential parameters as depicted in table [44].

Table 4: Evaluation of NIQUE for different Gamma values

γ	NIQUE↓	
	CUHK	AR
1	3.352	3.125
2	2.972	2.654
2.5	2.7525	2.457
2.2	2.65133	2.34217
2.3	2.6942	2.412

Table 5: Analysis of non-referential image metrics of our proposed Inverse GCT

Quality metrics in average		Original sketch		Inverse GCT sketches	
		Image dataset		Image dataset	
		CUHK	AR	CUHK	AR
NIQUE↓		3.157	2.954	2.65133	2.34217
FID↓		82.21	75.23	77.44746	65.29599
BRISQUE↓		0.678	0.654	0.567	0.542
DIQA↑	Spearman's Rank-Order Correlation Coefficient (SRCC)	0.652	0.721	0.882	0.903
	Pearson's Linear Correlation Coefficient (PLCC)	0.774	0.727	0.891	0.917

Table 5 shows only subtle differences between the two datasets, with a differential rate of about 12%, indicating that our Inverse GCT has indeed passed all tests in these datasets.

4.2 Visual Inspection of Generated Images from Textual Cues

In this section, we have briefly described various results obtained from our proposed GAN. The left side of the table contains textual descriptions of human faces mentioned by individuals. The right side contains the generated image from the same. Table 6 shows the results obtained from our proposed system.

Table 6: Visual results obtained from our proposed system

CUHK		AR	
Description	Image	Description	Image
Raised eyebrows flat mouth big eyes, and long nose		The man has smaller eyes normal eyebrows long nose and a small mouth	
The woman has raised eyebrows broad nose plain forehead with a wide mouth		The man has a square chin, large eyes, no nasal ridges and a long nose	
The woman has a square chin with small eyes and normal eyebrows		The woman has nasal ridges with large eyes normal eyebrows, and a straight curve mouth	
The man has a raised eyebrow with a broad nose visible nasal ridge and upward curved mouth, the chin is also square shaped.		This woman has a rounded chin, normal eyes, a plain forehead with visible ridges, and a flat curved lip.	
This man has raised eyebrows with partial nasal ridges, a broad nose, an upward lip curve, a square chin.		The woman has a rounded chin with a large eye, noticeable nasal ridges and a straight lip curve.	

4.3 Training Cost Analysis

In this section, we have given a detailed performance analysis of our proposed system during training. To enhance our network performance, we utilize Adam optimization algorithm with $\beta_1 = 0.0$ and $\beta_2 = 0.9$. We assign a learning rate of 0.0001 for the generator and 0.0004 for the discriminator.

Table 7 elucidates the training performance of our proposed system in terms of loss analysis. The evaluation parameters used are Generator loss, termed as G_Loss and discriminator loss, termed as D_Loss . These loss factors are paired with every time and epoch, respectively.

It is distinct from Table 7 that the loss factors decrease gradually with every epoch. The gradual decrease indicates that the generated image's quality improves with every epoch, as expected. Fig. 11 reveals a graphical analysis of Table 7.

Table 7: Epoch wise loss analysis

Image dataset	Epoch	G_Loss↓	D_Loss↓	
			D_R	D_F
CUHK	1	13.793	0.648	0.424
	50	6.756	0.165	0.054
	100	4.732	0.152	0.173
	150	2.523	0.132	0.233
	200	2.305	0.132	0.174
AR	1	13.16	1.117	1.09
	50	6.154	0.447	0.301
	100	5.572	0.283	0.52
	150	1.667	0.721	0.186
	200	2.105	0.112	0.142

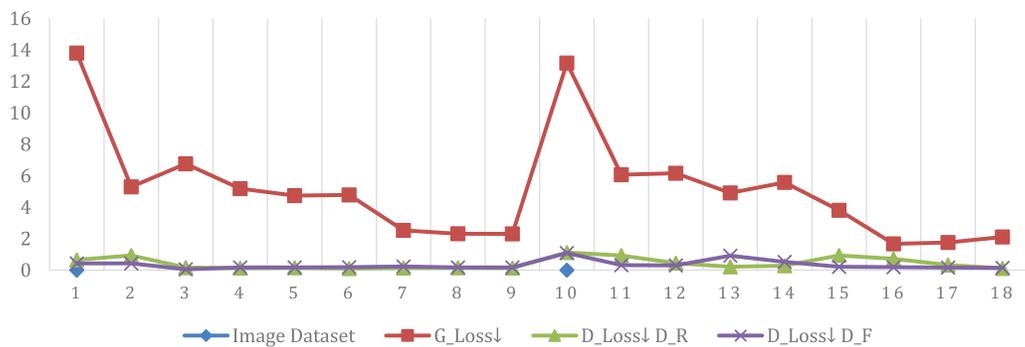


Figure 11: Comparative analysis of loss values

As part of our evaluation process, we thoroughly assessed the computational cost of our system. We considered various factors such as Random-Access Memory (RAM) utilization, model parameters, and training time. Specifically, we analyzed the RAM utilization of our system during its operation, which allowed us to identify potential memory bottlenecks. We also assessed the number of model parameters used, which can affect the accuracy and efficiency of the system. Additionally, we evaluated the time taken to train the system, which is crucial for determining its scalability and suitability. [Table 8](#) presents consolidated data revealing computational cost analysis.

Table 8: Computational cost analysis

Dataset name	Training time (Hours)	RAM utilized (Giga Bytes)	Model parameters
CUHK	1.5	6.85	3330864
AR	1.02	4.48	3330864

4.4 Comparative Examination of the Proposed System with Standard State-of-the-art Techniques

To prove that our system performs better, we rigorously tested it against state-of-the-art GANs. Table 9 gives a detailed analysis of the same.

Table 9: Examination of quality metrics with different GANs [45]

Technique	Dataset	Evaluation metrics						
		PSNR \uparrow	SSIM \uparrow	MSE \downarrow	NAE \downarrow	NIQUE \downarrow	FID \downarrow	BRISQUE \downarrow
Text2img	CUHK	18.56	0.65	4197.50	0.29	3.102	149.37	0.68
	AR	17.61	0.69	3596.72	0.062	3.031	122.27	0.70
Stack GAN	CUHK	20.39	0.78	1028.18	0.257	2.64	79.37	0.62
	AR	22.73	0.82	772.46	0.083	2.54	68.37	0.59
CVAE	CUHK	19.96	0.71	2504.72	0.352	3.23	154.3	0.78
	AR	21.81	0.79	1891.29	0.143	3.12	133.7	0.74
Attribute2 Sketch2Face	CUHK	22.69	0.83	1079.42	0.037	3.23	92.41	0.61
	AR	24.08	0.89	1016.67	0.018	3.04	75.47	0.57
Attr2face	CUHK	19.72	0.63	2345.41	0.27	3.66	188.7	0.93
	AR	19.63	0.68	1586.22	0.12	3.48	186.1	0.71
Proposed System	CUHK	25.68	0.91	580.39	0.017	2.65	77.44	0.56
	AR	24.45	0.91	648.78	0.012	2.34	65.29	0.54

From Table 9, it is evident that our system did indeed perform better with all quality metrics. The performance rate for CUHK is much better than AR because of the dataset's quality in these photos. We have also exploited GAN parameters to identify the ability of our system. The parameters chosen for this are Inception Score. Table 10 shows the inception score gathered by different GANs concerning different datasets [46].

Table 10: Comparative analysis of Inception score with different existing techniques [47]

Technique	Inception score \uparrow	
	CUHK	AR
Stack GAN	1.348 \pm 0.113	1.513 \pm 0.017
Text2img	1.268 \pm 0.044	1.481 \pm 0.018
Attr2face	1.343 \pm 0.132	1.525 \pm 0.011
CVAE	1.118 \pm 0.023	1.272 \pm 0.006
Attribute2Sketch2Face	1.432 \pm 0.041	1.557 \pm 0.010
Proposed System	1.437 \pm 0.048	1.551 \pm 0.012

Table 10 shows that our proposed system performed better when compared to other similar sets of GANs. Higher scores indicate a better ability of GAN. Our proposed system score was close to Attribute2Sketch2Score, whose internal design is remarkably similar to our DF-GAN. This change

is due to an additional module of Inverse GCT, which enhances the quality of the sketch used for generation.

5 Conclusion

In this research project, our primary goal was to use GAN to create sketches from textual descriptions, and we were successful in achieving this objective. To improve the quality of the sketch samples, we also repurposed our GAN with an inverse gamma preprocessor during the training process. We evaluated the performance of our proposed system against other state-of-the-art techniques using standard procedures to assess the visual and structural quality of the generated sketches. The results showed that our system performed better than existing methods. We also used standard metrics to evaluate the performance of our GAN compared to other state-of-the-art GANs. Our analysis revealed that our proposed GAN performed significantly better due to the improved training process resulting from the enhanced quality of sketches. We used well-known datasets, CUHK and AR, to conduct all the tests.

Our research has practical applications in health organizations and forensic sciences, where it can be used to compare the remains of a person during postmortem analysis. The generated images from sketches or actual photos can also be legalized into photos to ensure proper payment of healthcare claims. Additionally, our proposed system can be utilized as a tool to detect and prevent false healthcare claims payment fraud by stolen healthcare provider identity methods. This scam detection method can help reduce fraudulent and unnecessary healthcare expenses [48–52].

6 Future Work

As a next step, this research can be extended to include a translator component that can retrieve facial sketches from verbal descriptions in regional languages. This would greatly enhance the usability and applicability of the proposed system in diverse regions where multiple languages are spoken. Another potential area of extension for this research is the addition of a sketch-to-photo GAN module. This module would allow for the generation of photo-realistic images, providing a more accurate representation of the person being described in the textual input. This would be particularly useful in fields such as law enforcement and security, where the accuracy of visual identification is critical.

Moreover, there is a scope for extending this work to generate images with expressions based on verbal descriptions. This could be particularly beneficial for healthcare systems that use facial expressions to recognize autistic patients. By incorporating this functionality, the proposed system could help healthcare providers more accurately diagnose and treat patients with autism. Overall, the potential applications of this research are vast and varied, and there is ample room for further exploration and development. By continuing to improve and expand upon the proposed system, we can better support a wide range of industries and fields in which visual identification is essential.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] S. Nagpal, M. Singh, R. Singh, M. Vatsa, A. Noore *et al.*, “Face sketch matching via coupled deep transform learning,” in *Proc. ICCV*, Venice, Italy, pp. 5419–5428, 2017.
- [2] S. Ouyang, T. M. Hospedales, Y. Z. Song and X. Li, “ForgetMeNot: Memory-aware forensic facial sketch matching,” in *Proc. CVPR*, Las Vegas, Nevada, pp. 5571–5579, 2016.
- [3] K. Ounachad, M. Oualla, A. Souhar and A. Sadiq, “Face sketch recognition-an overview,” in *Proc. NISS*, Marrakech, Morocco, pp. 1–8, 2020.
- [4] T. Karras, S. Laine and T. Aila, “A style-based generator architecture for generative adversarial networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 12, pp. 4217–4228, 2021.
- [5] Y. Li, Q. Wang, J. Zhang, L. Hu and W. Ouyang, “The theoretical research of generative adversarial networks: An overview,” *Neurocomputing*, vol. 435, pp. 26–41, 2021.
- [6] L. Gonog and Y. Zhou, “A review: Generative adversarial networks,” in *Proc. ICIEA*, Xi’an, China, pp. 505–510, 2019.
- [7] S. Ruan, Y. Zhang, K. Zhang, Y. Fan, F. Tang *et al.*, “DAE-GAN: Dynamic aspect-aware GAN for text-to-image synthesis,” in *Proc. ICCV*, Montreal, Canada, pp. 13960–13969, 2021.
- [8] H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang *et al.*, “StackGAN: Text to photo-realistic image synthesis with stacked generative adversarial networks,” in *Proc. ICCV*, Venice, Italy, pp. 5907–5915, 2017.
- [9] M. Z. Khan, S. Jabeen, M. U. Khan, T. Saba, A. Rehmat *et al.*, “A realistic image generation of face from text description using the fully trained generative adversarial networks,” *IEEE Access*, vol. 9, pp. 1250–1260, 2021.
- [10] H. Tang, X. Chen, W. Wang, D. Xu, J. J. Corso *et al.*, “Attribute-guided sketch generation,” in *Proc. FG 2019*, Lille, France, pp. 1–7, 2019.
- [11] H. Kazemi, M. Iranmanesh, A. Dabouei, S. Soleymani and N. M. Nasrabadi, “Facial attributes guided deep sketch-to-photo synthesis,” in *Proc. WACVW*, NV, USA, pp. 1–8, 2018.
- [12] S. M. Iranmanesh, H. Kazemi, S. Soleymani, A. Dabouei and N. M. Nasrabadi, “Deep sketch-photo face recognition assisted by facial attributes,” in *Proc. BTAS*, CA, USA, pp. 1–10, 2018.
- [13] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta *et al.*, “Generative adversarial networks: An overview,” *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 53–65, 2018.
- [14] K. Wang, C. Gou, Y. Duan, Y. Lin, X. Zheng *et al.*, “Generative adversarial networks: Introduction and outlook,” *IEEE/CAA Journal of Automatica Sinica*, vol. 4, no. 4, pp. 588–598, 2017.
- [15] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley *et al.*, “Generative adversarial networks,” *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.
- [16] P. Isola, J. Y. Zhu, T. Zhou and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” in *Proc. CVPR*, Honolulu, Hawaii, pp. 1125–1134, 2017.
- [17] B. Cao, H. Zhang, N. Wang, X. Gao and D. Shen, “Auto-GAN: Self-supervised collaborative learning for medical image synthesis,” in *Proc. AAAI*, New York, USA, vol. 34, pp. 10486–10493, 2020.
- [18] A. Gooya, O. Goksel, I. Oguz and N. Burgos, “Medical image synthesis for data augmentation and anonymization using generative adversarial networks,” in *Proc. SASHIMI*, Granada, Spain, pp. 1–11, 2018.
- [19] Y. Fang, W. Deng, J. Du and J. Hu, “Identity-aware cyclegan for face photo-sketch synthesis and recognition,” *Pattern Recognition*, vol. 102, no. 11, pp. 107249, 2020.
- [20] M. S. Sannidhan, G. Ananth Prabhu, D. E. Robbins and C. Shasky, “Evaluating the performance of face sketch generation using generative adversarial networks,” *Pattern Recognition Letters*, vol. 128, pp. 452–458, 2019.
- [21] J. Yu, X. Xu, F. Gao, S. Shi, M. Wang *et al.*, “Toward realistic face photo-sketch synthesis via composition-aided GANs,” *IEEE Transactions on Cybernetics*, vol. 51, no. 9, pp. 4350–4362, 2021.
- [22] F. Liu, X. Deng, Y. -K. Lai, Y. -J. Liu, C. Ma *et al.*, “Sketchgan: Joint sketch completion and recognition with generative adversarial network,” in *Proc. CVPR*, CA, USA, pp. 5830–5839, 2019.
- [23] L. Zhang, Y. Ji, X. Lin and C. Liu, “Style transfer for anime sketches with enhanced residual U-net and auxiliary classifier GAN,” in *Proc. ACPR*, Nanjing, China, pp. 506–511, 2017.

- [24] M. Zhu, P. Pan, W. Chen and Y. Yang, "DM-GAN: Dynamic memory generative adversarial networks for text-to-image synthesis," in *Proc. CVPR*, CA, USA, pp. 5802–5810, 2019.
- [25] T. Xu, P. Zhang, Q. Huang, H. Zhang, Z. Gan *et al.*, "AttnGAN: Fine-grained text to image generation with attentional generative adversarial networks," in *Proc. CVPR*, Salt Lake City, USA, pp. 1316–1324, 2018.
- [26] G. Li, L. Li, Y. Pu, N. Wang and X. Zhang, "Semantic image inpainting with multi-stage feature reasoning generative adversarial network," *Sensors*, vol. 22, no. 8, pp. 2854, 2022.
- [27] P. Sumi, S. Sindhuja and S. Sureshkumar, "A comparison between AttnGAN and DF GAN: Text to image synthesis," in *Proc. ICPSC*, Coimbatore, India, pp. 615–619, 2021.
- [28] M. Tao, H. Tang, F. Wu, X. Jing, B. -K. Bao *et al.*, "DF-GAN: A simple and effective baseline for text-to-image synthesis," in *CVPR*, New Orleans, USA, pp. 16515–16525, 2022.
- [29] R. Mehmood, R. Bashir and K. J. Giri, "Comparative analysis of AttnGAN, DF-GAN and SSA-GAN," in *Proc. ICAC3N*, Greater Noida, India, pp. 370–375, 2021.
- [30] M. S. Sannidhan, G. A. Prabhu, K. M. Chaitra and J. R. Mohanty, "Performance enhancement of generative adversarial network for photograph-sketch identification," *Soft Computing*, vol. 27, no. 1, pp. 435–452, 2021.
- [31] M. Elhoseny, M. M. Selim and K. Shankar, "Optimal deep learning based convolution neural network for digital forensics face sketch synthesis in internet of things (IoT)," *International Journal of Machine Learning and Cybernetics*, vol. 12, no. 11, pp. 3249–3260, 2020.
- [32] M. Rizkinia, N. Faustine and M. Okuda, "Conditional generative adversarial networks with total variation and color correction for generating indonesian face photo from sketch," *Applied Sciences*, vol. 12, no. 19, pp. 10006, 2022.
- [33] M. S. Sannidhan, K. M. Chaitra and G. A. Prabhu, "Assessment of image enhancement procedures for matching sketches to photos," in *Proc. DISCOVER*, Manipal, India, pp. 1–5, 2019.
- [34] G. Cao, Y. Zhao and R. Ni, "Forensic estimation of gamma correction in digital images," in *Proc. 2010 IEEE Int. Conf. on Image Processing*, Hong Kong, China, pp. 2097–2100, 2010.
- [35] F. Kallel, M. Sahnoun, A. Ben Hamida and K. Chtourou, "CT scan contrast enhancement using singular value decomposition and adaptive gamma correction," *Signal Image and Video Processing*, vol. 12, no. 5, pp. 905–913, 2018.
- [36] S. Pallavi, M. S. Sannidhan and A. Bhandary, Retrieval of facial sketches using linguistic descriptors: An approach based on hierarchical classification of facial attributes. In: *Advances in Intelligent Systems and Computing*, Singapore, Springer, pp. 1131–1149, 2020.
- [37] M. A. Khan and A. S. Jalal, "A framework for suspect face retrieval using linguistic descriptions," *Expert Systems with Applications*, vol. 141, pp. 112925, 2020.
- [38] M. A. Khan and A. Singh Jalal, "Suspect identification using local facial attributed by fusing facial landmarks on the forensic sketch," in *Proc. IC3A*, Lucknow, India, pp. 181–186, 2020.
- [39] A. S. Jalal, D. K. Sharma and B. Sikander, Suspect face retrieval using visual and linguistic information. In: *The Visual Computer*, Germany, Springer, 2022.
- [40] S. Pallavi, M. S. Sannidhan, K. B. Sudeepa and A. Bhandary, "A novel approach for generating composite sketches from mugshot photographs," in *Proc. ICACCI*, Bangalore, India, pp. 460–465, 2018.
- [41] S. L. Fernandes and G. Josemin Bala, "Image quality assessment-based approach to estimate the age of Pencil Sketch," in *Proc. SocProS*, Roorkee, India, pp. 633–642, 2016.
- [42] S. L. Fernandes and G. Josemin Bala, "Self-similarity descriptor and local descriptor-based composite sketch matching," in *Proc. SocProS*, Roorkee, India, pp. 643–649, 2016.
- [43] S. L. Fernandes and G. J. Bala, "Odroid XU4 based implementation of decision level fusion approach for matching computer generated sketches," *Journal of Computational Science*, vol. 16, pp. 217–224, 2016.
- [44] X. Liu, M. Pedersen and C. Charrier, "Performance evaluation of no-reference image quality metrics for face biometric images," *Journal of Electronic Imaging*, vol. 27, no. 2, pp. 1, 2018.
- [45] U. Osahor, H. Kazemi, A. Dabouei and N. Nasrabadi, "Quality guided sketch-to-photo image synthesis," in *Proc. CVPRW*, WA, USA, pp. 820–821, 2020.

- [46] J. Bao, D. Chen, F. Wen, H. Li and G. Hua, "CVAE-GAN: Fine-grained image generation through asymmetric training," in *Proc. ICCV*, Venice, Italy, pp. 2745–2754, 2017.
- [47] X. Di and V. M. Patel, "Facial synthesis from visual attributes via sketch using multiscale generators," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 2, no. 1, pp. 55–67, 2020.
- [48] S. L. Fernandes and S. K. Jha, "Adversarial attack on deepfake detection using RL based texture patches," in *Proc. ECCV*, Glasgow, UK, pp. 220–235, 2020.
- [49] S. Fernandes, S. Raj, E. Ortiz, I. Vintila and S. K. Jha, "Directed adversarial attacks on fingerprints using attributions," in *Proc. ICB*, Crete, Greece, pp. 1–8, 2019.
- [50] V. P. Gurupur, S. A. Kulkarni, X. Liu, U. Desai and A. Nasir, "Analyzing the power of deep learning techniques over the traditional methods using medicare utilization and provider data," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 31, no. 1, pp. 99–115, 2018.
- [51] K. Santhi, "A survey on medical imaging techniques and applications," *Journal of Innovative Image Processing*, vol. 4, no. 3, pp. 173–182, 2022.
- [52] S. Iwin Thanakumar Joseph, "Advanced digital image processing technique based optical character recognition of scanned document," *Journal of Innovative Image Processing*, vol. 4, no. 3, pp. 195–205, 2022.