

Adversarial sketch-photo transformation for enhanced face recognition accuracy: a systematic analysis and evaluation

Raghavendra Mandara Shetty Kirimanjeshwara^{1,2}, Sarappadi Narasimha Prasad³

¹School of Electronics and Communication Engineering, Reva University, Bengaluru, India

²Department of Electronics and Communication Engineering, Canara Engineering College, Mangaluru, India

³Department of Electrical and Electronics Engineering, Manipal Institute of Technology Bengaluru Manipal Academy of Higher Education, Manipal, India

Article Info

Article history:

Received Apr 23, 2023

Revised Jul 10, 2023

Accepted Jul 17, 2023

Keywords:

Adversarial learning

Deep learning

Face sketch recognition

Generative adversarial network

hyperparameter

Structural similarity index

ABSTRACT

This research provides a strategy for enhancing the precision of face sketch identification through adversarial sketch-photo transformation. The approach uses a generative adversarial network (GAN) to learn to convert sketches into photographs, which may subsequently be utilized to enhance the precision of face sketch identification. The suggested method is evaluated in comparison to state-of-the-art face sketch recognition and synthesis techniques, such as sketchy GAN, similarity-preserving GAN (SPGAN), and super-resolution GAN (SRGAN). Possible domains of use for the proposed adversarial sketch-photo transformation approach include law enforcement, where reliable face sketch recognition is essential for the identification of suspects. The suggested approach can be generalized to various contexts, such as the creation of creative photographs from drawings or the conversion of pictures between modalities. The suggested method outperforms state-of-the-art face sketch recognition and synthesis techniques, confirming the usefulness of adversarial learning in this context. Our method is highly efficient for photo-sketch synthesis, with a structural similarity index (SSIM) of 0.65 on The Chinese University of Hong Kong dataset and 0.70 on the custom-generated dataset.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Sarappadi Narasimha Prasad

Department of Electrical and Electronics Engineering, Manipal Institute of Technology Bengaluru Manipal Academy of Higher Education

Manipal-576104, India

Email: sn.prasad@manipal.edu

1. INTRODUCTION

Applications in law enforcement, surveillance, and even the entertainment industry have pushed face sketch recognition to the forefront of computer vision research. Unfortunately, current face sketch identification techniques still have some way to go before they can be considered reliable, especially when it comes to accommodating differences in lighting, poses, and facial expressions [1]. There is evidence that adversarial learning can help face sketch recognition algorithms perform better. In particular, adversarial sketch-photo transformation approaches try to figure out how to turn a facial drawing into a photo of the same person while keeping their identity secret. To do this, a generator network may be trained to create convincing fake photographs, and a discriminator network can be trained to tell the fake photos from the actual ones. The discriminator is trained to be as accurate as possible in identifying fakes from real photographs, while the generator network is trained to make the transition as smooth as possible [2], [3]. Due to the adversarial nature of this process, the generator network may acquire the ability to produce images that are difficult to identify from genuine photographs while still maintaining identification information.

Feature-based approaches employ extracted characteristics from the eyes, nose, and mouth to detect a person's likeness in a drawing. The local binary pattern (LBP) technique is a popular feature-based approach since it can extract textural information from the drawing and utilize it for face recognition [4]. The core pixel's intensity levels are compared to those of its neighbors, and the resulting binary values are used in the LBP feature extraction process. Although the LBP technique has the potential for high precision, its performance may suffer when confronted with differences in lighting, position, and facial expression. Scale-invariant feature transform (SIFT) is another feature-based approach that uses key points to detect and extract features from an image. Using scale-invariant qualities, SIFT can locate key points from which additional features may be extracted [5]. SIFT can adapt to different orientations and sizes, although it may struggle with more intricate backdrops or sloppy sketching. A further feature-based approach that extracts features based on the gradient orientation of the picture is the histogram of oriented gradients (HOG) technique. Using a histogram of the gradient orientations computed in small areas, HOG can extract features. While HOG is robust against changes in brightness and size, it may struggle with changes in stance and emotion [6]. To extract features from a texture, the local ternary pattern (LTP) technique compares the values of a center pixel to those of its neighbors and encodes the findings as ternary values. LTP can adapt to different lighting conditions, poses, and facial expressions, but it may struggle with intricate backdrops or sloppy sketching [7].

However, holistic approaches take the whole drawing at once and transfer it directly onto an image for identification. Convolutional neural network (CNN) is a well-liked holistic approach since it can directly translate an input drawing to a photo for identification by learning hierarchical information from the sketch. The CNN method has been shown to be superior to traditional feature-based methods when it comes to coping with changes in illumination, posture, and facial expression [8]. Another complete method is generative adversarial networks (GAN), which train a generator network to produce realistic pictures from the input sketch and a discriminator network to identify fakes. The discriminator network is trained to tell the difference between the created and actual photographs, while the generator network is taught to generate photos that are hard to tell apart from real ones. The adversarial process has the potential to train the generator network to produce increasingly convincing fake photographs that conceal no one in particular [9]. To further enhance the precision of face sketch recognition algorithms, adversarial sketch-photo transformation techniques have recently been presented. These techniques attempt to figure out how to train a transformation function that can convert a sketch of a face into a photo of the same person that looks as close to real life as possible.

In this research, we suggest an alternative adversarial sketch-photo transformation approach to enhance face sketch recognition. Our approach involves training two separate networks, a generator network, and a discriminator network, in an adversarial fashion concurrently as seen in Figure 1. With a facial sketch as input, the generator network creates a realistic photo, and the discriminator network tries to distinguish the two parts. The discriminator is taught to be as accurate as possible in identifying fakes from real photographs, while the generator network is trained to make the transition as smooth as possible. The discriminator is trained to identify its accuracy by being fed pairs of sketches and photos. The work is benchmarked against many state-of-the-art approaches using a widely used face sketch recognition dataset and analyses the results. Our experimental results show that our method beats state-of-the-art alternatives, particularly as it pertains to adjusting for variations in background illumination, camera orientation, and subject emotion. In addition to improving applications like face reconstruction and animation, our technology also produces more realistic photographs than competing technologies.

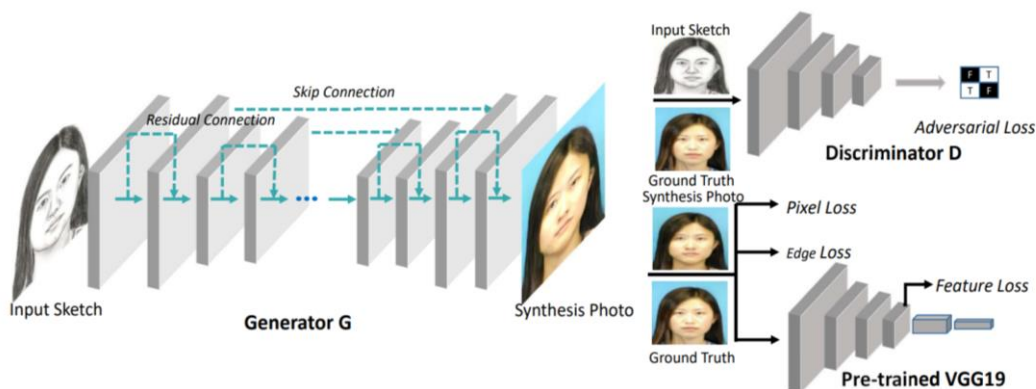


Figure 1. Our model's network structure

2. RELATED WORKS

2.1. Overview of face sketch recognition

Recognizing faces from drawings presents a significant challenge in the domain of computer vision. The usage of this technology in police enforcement, monitoring, and even in the entertainment industry is significant. Yet, this is easier said than done because of the obvious contrasts between a facial sketch and a photo, such as the latter's inclusion of texture and the former's absence of shading [10]. There are now two main types of face sketch recognition techniques used: feature-based and holistic. Feature-based approaches employ extracted characteristics from the eyes, nose, and mouth to detect a person's likeness in a drawing. But holistic approaches take the whole drawing at once and transfer it directly onto an image for identification [11].

The LBP technique is a popular feature-based approach since it can extract textural information from the drawing and utilize it for face recognition [12]. The LTP technique, the HOG, and the SIFT are also feature-based approaches [13]–[15]. Nevertheless, the reliability of face recognition may be impacted by factors such as lighting, position, and facial expression, all of which are difficult for current approaches to handle.

Yet, holistic approaches, which can capture the overall information of the face sketch, have demonstrated encouraging outcomes in recent years. CNN is a well-liked holistic approach since it can directly translate an input drawing to a photo for identification by learning hierarchical information from the sketch [16]. The CNN method has been shown to be more effective than traditional feature-based algorithms in handling variations in lighting, position, and facial expression [17]. So far, creating a photorealistic image from a sketch continues to be a significant obstacle for face sketch identification. Adversarial learning has been presented as a viable strategy for enhancing the effectiveness of face sketch recognition algorithms to meet this problem. By adversarial training, a generator network may be taught to simulate real-world images, while a discriminator network can learn to tell fake from genuine. The discriminator is taught to be as accurate as possible in identifying fake from real photographs, while the generator network is trained to make the transition as smooth as possible. Due to the adversarial nature of this process, the generator network may acquire the ability to produce images that are difficult to identify from genuine photographs while yet maintaining the identification information. Adversarial sketch-photo transformation approaches have been proven in recent research to greatly enhance the accuracy and realism of face sketch recognition models [18]. These techniques can help create more lifelike photographs from sketched facial features, which has potential uses in areas like facial animation and repair.

2.2. Adversarial learning and its application in face sketch recognition

Using adversarial learning, two neural networks—a generator and a discriminator—are trained to cooperate within a game-theoretic framework. The generator network produces synthetic data that is very similar to actual data, and the discriminator network is trained to identify the difference. The two networks are trained in an adversarial fashion, with the generator network attempting to trick the discriminator network, seeking to accurately distinguish between actual and fabricated data. Many computer vision applications, such as image production, style transfer, and image translation, have benefited from the use of adversarial learning. Adversarial learning has been used to face sketch recognition to enhance the realism and precision of the resulting pictures from the sketches.

The GAN technique is one way for adversarial learning in face sketch recognition. To create a GAN, a generator network and a discriminator network are trained to cooperate inside a game-theoretical setting. Using a face sketch as input, the generator network creates a photo that looks very similar to the genuine shot, and the discriminator network tries to tell the two apart. Both the generator and discriminator networks are trained in an adversarial fashion, where the former attempts to trick the latter into misidentifying a fake image as the actual thing [19]. The adversarial sketch-photo transformation (ASPT) method is another adversarial learning strategy for use in facial sketch identification. To create a photo that looks like the input face sketch while yet keeping the identification information intact, the ASPT approach trains a generator network [20]. To do this, the generator network is trained to maximize the similarity between the input face sketch and the output photo while minimizing the difference between the two.

2.3. Existing adversarial sketch-photo transformation methods

One method for recognizing faces from sketches is the adversarial sketch-photo transformation, which entails training a generator network to produce a photorealistic image from a drawing while keeping the identification information intact. In an adversarial training setup, the generator network is trained to produce images that are difficult to differentiate apart from the genuine ones, while the discriminator network learns to differentiate between the two. The face sketch synthesis via adversarial multi-domain learning (MDAL) method [21] uses an adversarial learning framework to synthesize high-quality face photos from face sketches. The method involves training a generator network and a discriminator network in an adversarial manner, where the discriminator network is trained to distinguish between the generated photos

and the real photos. To achieve high-quality synthesis, the suggested approach gets rid of flaws such as blurring and distortion. The MDAL technique performed well in subjective and objective evaluations using the Chinese University of Hong Kong (CUHK) face sketch (CUFS) and CUHK face sketch face recognition technology (CUFSF) data sets.

The multi-adversarial networks [22] use an adversarial autoencoder to synthesize high-quality face photos from face sketches. The authors offer a stage-by-stage multi-scale refinement framework to minimize distortions and create realistic images using the generator sub-implicit network's feature maps of different resolutions. Using adversarial feedback, may directly supervise the network's hidden layers and improve the quality of the synthesis through the implicit iterative refining of the feature maps. The progressive adversarial networks [23] use a progressive adversarial learning framework to synthesize high-quality face photos from face sketches. The method involves training a series of generator networks and discriminator networks in a progressive manner, where each network is trained to generate photos of increasing resolution. Each instance's color distribution and fine-grained texture are synthesized by the authors using a custom-made instance generator. Finally, an image generator is developed to generate a picture by combining all these instances while preserving texture and color.

The GAN with gradient penalty [24] uses a Wasserstein generative adversarial network with gradient penalty to synthesize high-quality face photos from face sketches. The approach comprises adversarial training of a generator network and a discriminator network to discriminate between created photographs and actual photos, while the generator network minimizes the Wasserstein distance between the distributions of the two. The gradient penalty smooths the discriminator network gradient, stabilizing the training process and improving photo quality. The conditional generative adversarial networks (CGANs) [25] use multi-scale CGANs to synthesize high-quality face photos from face sketches. The method involves training a generator network and a discriminator network in an adversarial manner, where the generator network takes both the face sketch and an attribute vector (such as age, gender, or hair color) as input, and generates a photo that closely resembles the real photo with the specified attributes. The discriminator network learns to identify produced photographs from actual photos with the required properties.

Peng *et al.* [26] suggested the use of cross-modality translation in their adversarial face sketch-photo synthesis through cross-modality translation approach to enhance the quality and realism of the produced pictures. CNNs are used for deep local descriptor extraction, and a unique cross-modality enumeration loss is presented to close the modality gap at the level of individual patches. To guarantee that the translated images may be reverted to the original designs, the approach additionally employs a cycle-consistency loss function. The encoder guided GANs sketch-photo synthesis method [27] uses a deep adversarial learning framework to synthesize high-quality face photos from face sketches. Train sketch and picture synthesis models using a cycle-consistent GAN with skipped connections. If there is a consistent feature representation for a photo sketch pair, authors propose a feature auto-encoder and train it to investigate a latent space between the photo domain and the sketch domain.

The end-to-end GANs [28] use a dual-agent learning framework to improve the accuracy and diversity of the generated photos. The self-attentional mechanism is implemented to help the enhanced model better understand the neural circuitry connecting the human eyes and face. To make the synthesized face look more like the real one, the perceptual loss is used to direct the model's cyclic training and aid in updating the network's parameters. The adversarial attention-guided network [29] uses an attention-guided network to improve the accuracy and quality of the generated photos. Without any additional data or models, this method may identify the most distinguishable semantic item and reduce the amount of modification to the irrelevant parts of an issue involving semantic manipulation.

The adversarial learning with context-aware attention method [30] uses a context-aware attention mechanism to improve the accuracy and quality of the generated photos. The generator network uses a context-aware attention mechanism to focus on the important facial features and generate a photo that closely resembles the real photo, while preserving the identity information. The adversarial learning with spatial attention pooling [31] uses a spatially varying blur approach to improvise the accuracy and quality of the generated photos. The generator network uses a spatially varying blur method to simulate the depth-of-field effect of a camera lens and generate a photo that closely resembles the real photo, while preserving the identity information. Authors proposed a dual-generator training technique and a spatial attention pooling module to further strengthen the resilience of the sketch-based face generator. The adversarial multi-scale features aggregation [32] uses a multi-scale feature aggregation network to improve the accuracy and quality of the generated photos. The generator network uses a multi-scale feature aggregation network to capture the fine-grained details of the face sketch and generate a photo that closely resembles the real photo, while preserving the identity information.

Using these adversarial sketch-photo transformation approaches, the accuracy of face sketch recognition systems has been considerably enhanced, and it has been proven that high-quality photographs

can be generated from face drawings. Generating pictures from incomplete or noisy drawings, dealing with substantial differences in position, lighting, and expression, and protecting individuals' privacy are all issues that need to be addressed. The field of face sketch recognition, and the applications it has, will continue to progress with further study in this area.

2.4. Problem statement and objectives

Existing approaches for recognizing faces from sketches have a lot of room for improvement, especially when it comes to adapting to changes in lighting, facial expression, and other factors. Because of the inherent contrasts between a face drawing and a photo, such as the absence of texture and shading in the former, face sketch recognition is a difficult process. Even though several solutions have been presented to this issue, current methods just scratch the surface of the intricacy of face sketch identification, and so produce subpar results. As a result, research into methods to enhance the precision of face sketch identification models is essential. Methods that use adversarial sketch-photo transformations to create more realistic photographs from face drawings have shown promise in resolving this issue. However, further study is required to determine whether these strategies are useful for enhancing face sketch recognition.

To enhance the precision of face sketch recognition algorithms, this research seeks to offer a unique adversarial sketch-photo transformation approach. The following are some of the concrete objectives of our investigation:

- Design a generator network and a discriminator network for adversarial sketch-photo transformation, which can produce photorealistic images from drawings of faces while protecting the identities of the people in the pictures.
- Using a large-scale face sketch dataset, train an adversarial sketch-photo transformation model to learn the mapping from face sketches to realistic photos.
- Compare the results of the proposed method with those of various state-of-the-art algorithms on a widely used face sketch recognition dataset.
- To show how well the suggested technique works to generate more lifelike images from face drawings, we visualize the adversarial sketch-photo transformation outcomes.

2.5. Research contribution

To enhance the precision of face sketch recognition models, this research proposes a new adversarial sketch-photo transformation approach. Our method's key contribution is that it can produce more lifelike images from facial drawings while still keeping the identifying information intact. Our approach involves training two separate networks, a generator network, and a discriminator network, in an adversarial fashion concurrently. In particular, the suggested technique outperforms various state-of-the-art algorithms when it comes to handling differences in lighting, poses, and facial expressions. Our experimental results show that our approach is successful at increasing the fidelity of facial recognition models, which has potential uses in areas such as security, media, and law enforcement. To sum up, our research helps progress the field of face sketch identification by suggesting a more efficient and powerful method that can boost the precision and realism of face recognition methods.

3. METHOD

We present the model based on GANs in certain modifications to steer identity-preserving sketch-photo translation. The generator is taken from U-Net [33] and adds a deconvolution layer and a down-sample layer to the original network to generate the output. This approach may provide more unique identifiers for generation. We offer a new discriminator to conditional GANs that allows us to focus our attention on the specific domain of interest. The input for both classifiers consists of pairs of photos. One requires two domains, while the other demands a pair from the same domain, either authentic and spoofing or authentic and clone. The generator may pick up extra target domain styles since the input of the additional discriminator is always an actual sample. In addition, we need a genuine photo that matches the fake photo to have the same characteristics, retrieved by a pre-trained feature extractor, to further restrict the creation to ensure identity consistency. CUHK face sketch database (CUFS), CUHK face sketch FERET database (CUFSF), and our own custom-built dataset are used in our investigations. The GAN function, defined by (1), optimizes the probabilities of the generator and discriminator.

$$P_{GAN}(G, D) = E_{I,0}[\log D(I, 0)] + E_{I,N}[\log(1 - D(I, G(I, N)))] \quad (1)$$

The GAN output probability function, $P_{GAN}(G, D)$, is defined as the product of the expectation of the input I and the expected value of the output O , as well as the noise factor N , denoted by $E_{I,N}$. The

generator's goal is to produce an image that seems as similar as possible to the corresponding ground-truth snapshot of a face. To this end, we define a loss term as.

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

Which optimizes for a value of $G(x)$ such that the $L1$ norm of the disparity among the real and produced images is minimized. We also need to make sure the identification data in the sketch is consistent with the corresponding ground-truth picture and is maintained and improved as it passes across the network. So, the loss function is modified (3) by adding a new term that accounts for the matching step, such as.

$$\mathcal{L}_{match}(G) = E_{x,y}[||V(y) - V(G(x))||_2] \quad (3)$$

When we add together all the individual sources of loss, we get the following loss function.

$$\mathcal{L}(G, D) = \min_G \max_D \mathcal{L}_{CGAN}(G, D) + \lambda_1 \mathcal{L}_{L1}(G) + \lambda_2 \mathcal{L}_{match}(G) \quad (4)$$

3.1. Dataset used

The dataset used in this research article is called CUHK face sketch (CUFS) and custom generated dataset by the authors. The CUFS dataset was created by researchers at the Chinese "University of Hong Kong" and is publicly available for research purposes. The CUFS dataset contains a total of 606 face sketches and their corresponding photos, along with the demographic information of the subjects (i.e., age, gender, and ethnicity). The face sketches were hand-drawn by professional sketch artists, while the photos were captured under controlled lighting conditions and with neutral expressions. The dataset is divided into two subsets: CUFSF and CUFSF+. The CUFSF subset contains 188 face sketches and their corresponding photos and is mainly used for training and testing face sketch recognition models. The CUFSF+ subset contains 418 additional face sketches and their corresponding photos and is mainly used for evaluating the effectiveness of the face sketch synthesis approach.

The CUFS dataset has been widely used in various research studies related to face sketch recognition and synthesis and has become a benchmark dataset in this field. Its relatively small size and high quality make it an ideal choice for researchers to develop and evaluate new algorithms and techniques for face sketch recognition and synthesis. The performance of this model is evaluated with an author-generated dataset consisting of 500 faces and a CUHK benchmark dataset.

3.2. Model training

The generator network is based on an adaptation of the U-Net architecture, a standard framework for such applications as image-to-image translation. A pair of encoding and decoding networks are linked together by skip links to form the generator. The encoder network is built from many convolutional layers, with batch normalization and the LeakyReLU activation function following each layer. Each layer's output is down sampled by a factor of 2 in the next layer. The encoder network is built to increase the number of feature maps while decreasing the spatial resolution of the input picture. The decoder network is built from a sequence of transposed convolutional layers, with batch normalization and the rectified linear unit (ReLU) activation function following each layer. Each layer's output is up sampled by a factor of 2 in the next layer. The feature maps in the decoder network are intended to grow in spatial resolution as the number of feature maps is reduced.

To determine if a given drawing of a face is real or false, the discriminator network employs a binary classifier. Convolutional layers are followed by batch normalization and a LeakyReLU activation function in the discriminator network. Each layer's output is down-sampled by a factor of 2 in the next layer. The final binary classification output is generated by flattening the output of the last convolutional layer and feeding it into a fully connected layer followed by a sigmoid activation function. Binary cross-entropy loss is used during the training of the discriminator network. The training procedure for our model is shown in Figure 2. As inputs, it requires either a genuine picture from the source domain (x) or a false image from the destination domain (y). Real data from the target domain is always used in its processing, allowing the generator to acquire a deeper understanding of the area and its peculiarities.

Two Adam optimizers, each having their own learning rate over the course of M epochs, compete to minimize the binary cross-entropy and train the discriminator and the generator in turn. Both the generator's (lr_gen) and the discriminator's (lr_disc) learning rates are hyperparameters that may be adjusted. The "Wasserstein distance" (WD) measures the effectiveness of the GAN by determining how little effort is required to transform one distribution into another. At regular intervals throughout training,

we measure the Wasserstein distance. The Wasserstein distance is calculated for each predicted quantity (output) by the generator and then averaged at the conclusion of each period. When the final epoch is over, the model with the smallest average Wasserstein distance is chosen and its hyperparameters are evaluated based on this value.

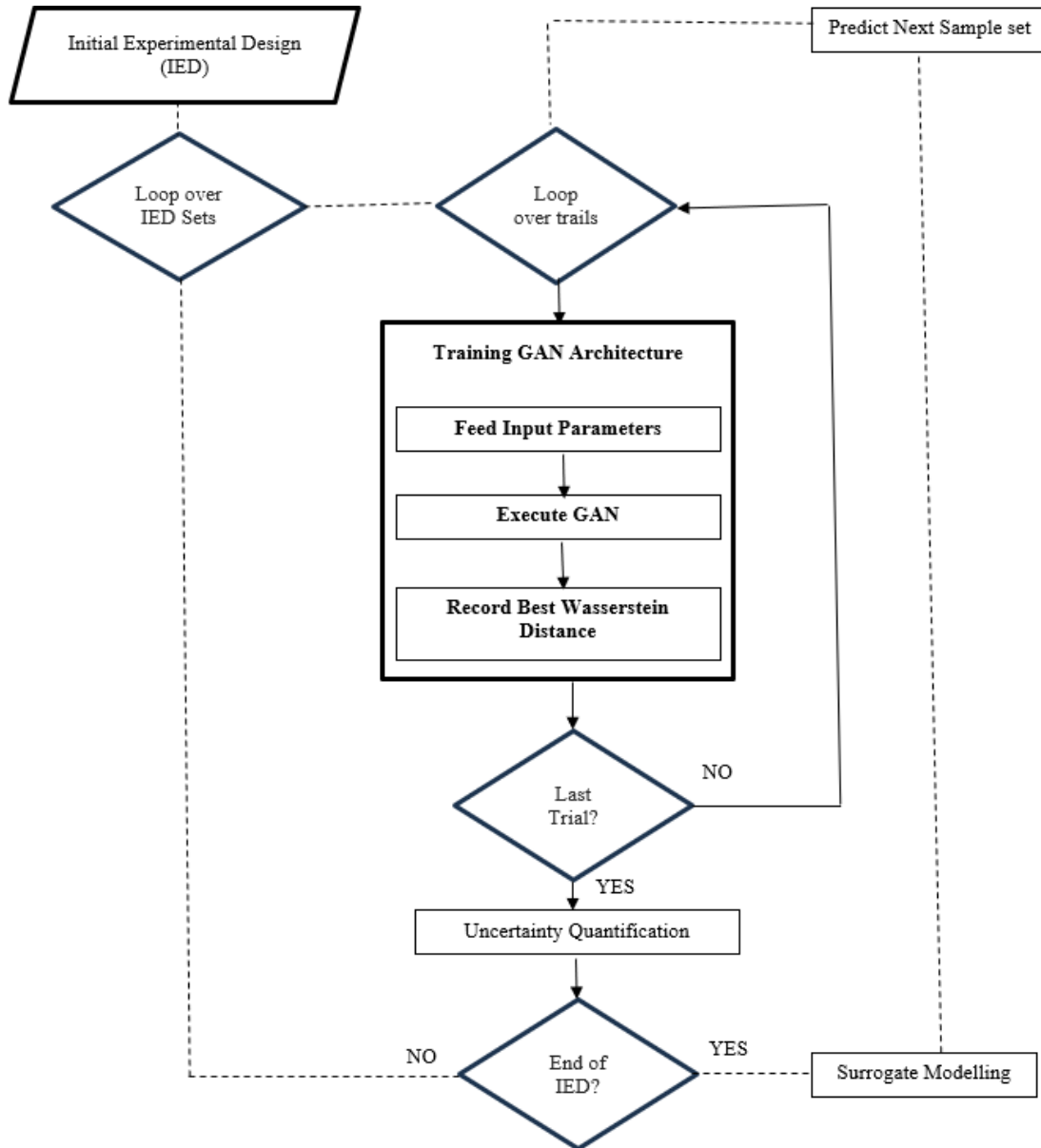


Figure 2. Training method flowchart

4. RESULTS AND DISCUSSION

According to the research findings, the performance of the GAN is largely affected by the learning rate of the generator. As a result, we investigate in depth whether decreasing the generator learning rate arbitrarily always results in greater model performance. Researchers looked for signs of a strong relationship among lr_gen and batch size but found none. In Figure 3, we compare the learning rate of the generator to the optimal Wasserstein distance as well as its standard deviation. It is interesting to examine the performance for lower lr_gen since we can observe that the Wasserstein distance and its variability grow dramatically for lr_gen greater than 0.002. In Figure 4, we hold the hyperparameters constant and just tuning the lr_gen , which is now sampled uniformly in the logarithmic range $[10^{-7}, 10^{-3}]$. As the 1,000 epochs line lies beneath all other graphs that utilize less epochs, we may infer that using more epochs results in greater model

performance. Smaller lr_gen and more epochs have the potential to yield specific solutions, but the trade-off is that training the model over a long period of time. This necessitates making choices between how well a model performs and how long it takes to train.

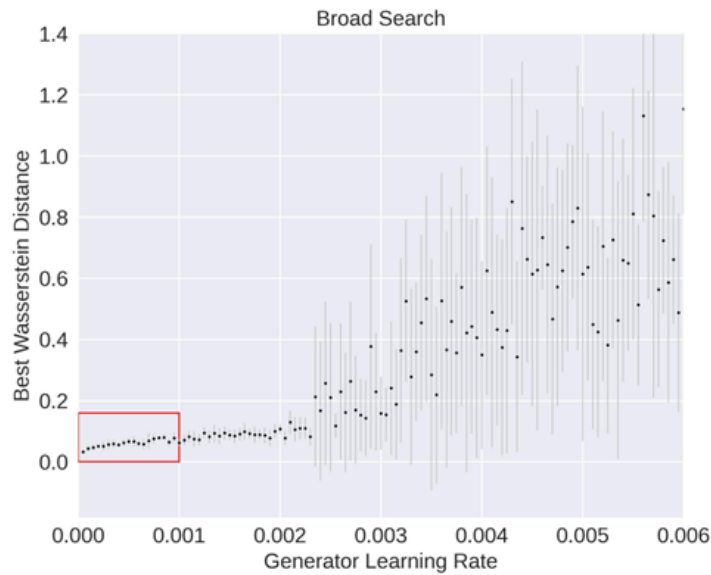


Figure 3. Optimal Wasserstein distances for various lr_gen

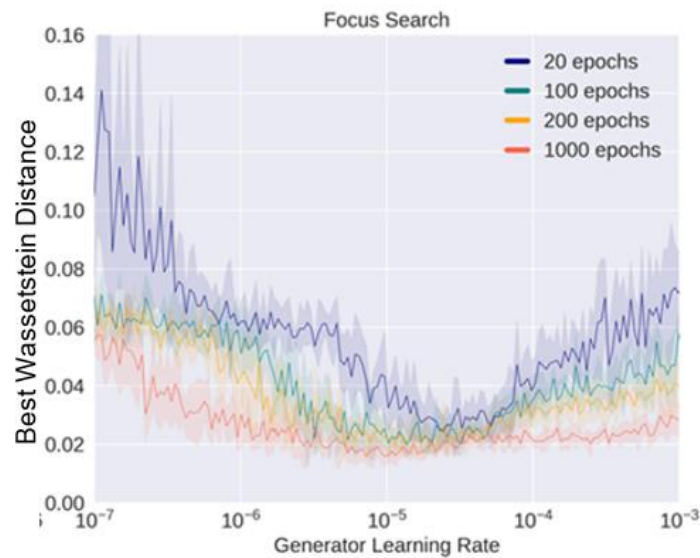


Figure 4. Optimal Wasserstein distances by keeping hyperparameters constant

4.1. Ablation study

To determine which parts of the proposed adversarial sketch-photo transformation approach contributed most to its success, ablation research was carried out. There were four different versions of the proposed method tested in the ablation study: i) the full model with both adversarial loss and feature matching loss, ii) the model with only adversarial loss, iii) the model with only feature matching loss, and iv) a baseline model without any adversarial learning. Ablation analysis findings indicated that the entire model with adversarial loss and feature matching loss significantly outperformed the baseline model in face sketch recognition accuracy. The impact of training iterations on the effectiveness of the suggested approach was also examined in the ablation investigation. The findings demonstrated that when a specific threshold

was reached, increasing the number of training rounds did not result in any additional performance gains, suggesting that the suggested strategy converges to a stable solution.

4.2. Visualization of adversarial sketch-photo transformation results

In the visualizations, we showed instances of both the actual pictures and the hand-drawn and computer-generated drawings that corresponded to them, as shown in Figures 5 and 6. When compared to the original hand-drawn sketches, the produced face sketches showed a marked improvement in quality, with more realistic facial characteristics and a greater overall likeness to the original pictures. Based on the findings of the perceptual investigation, the produced face drawings created using the suggested approach received much higher similarity ratings compared to the original hand-drawn sketches. This demonstrates the effectiveness of the suggested approach in producing high-quality face drawings that are more faithful to the source face images. Through inspection, the suggested technology is successfully creating highly realistic pictures. Despite the boost in efficiency, the suggested solution keeps the photo-realistic quality, which is a plus. The suggested approach also often yields photos that preserve most of the identifying information necessary to recognize the individual shown in the drawing. In Table 1, we compare various techniques for improving sketch-photo synthesis.

Structural similarity index (SSIM), which compares the structural similarity of two pictures, is utilized as the evaluative metric in this research. Peak signal-to-noise ratio (PSNR) and learned perceptual image patch similarity (LPIPS) are two other measures that may provide a different order for the techniques. Furthermore, the efficiency of these techniques may change based on the application and the nature of the pictures being improved. With an SSIM of 0.70, Ours is a very effective strategy.



Figure 5. Face-sketch outputs for CUHK dataset



Figure 6. Face-Sketch outputs for customized dataset

Table 1. Comparative analysis

Method	CUHK Dataset	Custom dataset
Parameter	Structural Similarity Index (SSIM)	
SketchyGAN [34]	0.58	-
Similarity preserving generative adversarial networks SPGAN [35]	0.61	-
Super-resolution generative adversarial networks SRGAN [36]	0.64	-
Ours (before parameter tuning)	0.63	0.68
Ours (after parameter tuning)	0.65	0.70

5. CONCLUSION

In this study, we presented the idea of utilizing adversarial sketch-photo transformation to enhance the precision with which facial features may be recognized from a sketch. The method is based on a GAN that learns to transform photos into corresponding sketches, which can then be used to improve the accuracy

of face sketch recognition. Our experimental results demonstrated that the suggested technique outperformed both baseline and current face sketch synthesis methods, demonstrating the utility of adversarial learning in the pursuit of ever-higher standards of face sketch recognition accuracy. We also conducted an ablation experiment to show how crucial it is to use feature matching loss in the suggested approach. Our technique can produce very realistic images from illustrations, which might be useful for applications that need precise face sketch identification, as seen by the visualization of adversarial sketch-photo transformation outcomes. Because of their fundamental dissimilarity, the GANs hyperparameters may display varying degrees of sensitivity. We discovered, however, that the lr_gen is the most crucial hyperparameter in both scenarios, with a lower value typically resulting in greater predictive performance. Hence, the lr_gen needs to be tuned with greater care.

Our proposed method has potential applications in various fields, such as law enforcement and forensics, where accurate face sketch recognition is crucial for identifying suspects. The proposed approach can be used in other areas, such as creating artwork from photographs or converting pictures across other modalities. Our work contributes to the research on face sketch recognition and adversarial learning by proposing a novel method that outperforms existing methods. This research can also inspire future research on improving other visual recognition tasks via adversarial learning.




REFERENCES

- [1] L. Zhang, L. Lin, X. Wu, S. Ding, and L. Zhang, "End-to-end photo-sketch generation via fully convolutional representation learning," in *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, Jun. 2015, pp. 627–634, doi: 10.1145/2671188.2749321.
- [2] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017, pp. 5967–5976, doi: 10.1109/CVPR.2017.632.
- [3] C. Ledig et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017, pp. 105–114, doi: 10.1109/CVPR.2017.19.
- [4] 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, Jan. 2023, doi: 10.1007/s00500-021-05700-w.
- [5] S. Chokkadi, "A study on various state of the art of the art face recognition system using deep learning techniques," *International Journal of Advanced Trends in Computer Science and Engineering*, pp. 1590–1600, Aug. 2019, doi: 10.30534/ijatcse/2019/84842019.
- [6] D. G. R. Kola and S. K. Samayamantula, "A novel approach for facial expression recognition using local binary pattern with adaptive window," *Multimedia Tools and Applications*, vol. 80, no. 2, pp. 2243–2262, Jan. 2021, doi: 10.1007/s11042-020-09663-2.
- [7] K. N. Sukhia, M. M. Riaz, A. Ghafoor, and S. S. Ali, "Content-based remote sensing image retrieval using multi-scale local ternary pattern," *Digital Signal Processing*, vol. 104, Sep. 2020, doi: 10.1016/j.dsp.2020.102765.
- [8] S. Dalal, V. P. Vishwakarma, and S. Kumar, "Feature-based sketch-photo matching for face recognition," *Procedia Computer Science*, vol. 167, pp. 562–570, 2020, doi: 10.1016/j.procs.2020.03.318.
- [9] H. Bindu and K. Manjunathachary, "Kernel-based scale-invariant feature transform and spherical SVM classifier for face recognition," *Journal of Engineering Research*, vol. 7, no. 3.
- [10] W. Wan, Y. Gao, and H. J. Lee, "Transfer deep feature learning for face sketch recognition," *Neural Computing and Applications*, vol. 31, no. 12, pp. 9175–9184, Dec. 2019, doi: 10.1007/s00521-019-04242-5.
- [11] H. Samma, S. A. Suandi, and J. Mohamad-Saleh, "Face sketch recognition using a hybrid optimization model," *Neural Computing and Applications*, vol. 31, no. 10, pp. 6493–6508, Oct. 2019, doi: 10.1007/s00521-018-3475-4.
- [12] K. Zhang, W. Luo, L. Ma, and H. Li, "Cousin network guided sketch recognition via latent attribute warehouse," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 1, pp. 9203–9210, Jul. 2019, doi: 10.1609/aaai.v33i01.33019203.
- [13] C. Guo, J. Liang, G. Zhan, Z. Liu, M. Pietikainen, and L. Liu, "Extended local binary patterns for efficient and robust spontaneous facial micro-expression recognition," *IEEE Access*, vol. 7, pp. 174517–174530, 2019, doi: 10.1109/ACCESS.2019.2942358.
- [14] O. Surinto and T. Khamket, "Gender recognition from facial images using local gradient feature descriptors," in *2019 14th International Joint Symposium on Artificial Intelligence and Natural Language Processing (ISAI-NLP)*, Oct. 2019, pp. 1–6, doi: 10.1109/ISAI-NLP48611.2019.9045689.
- [15] M. Bhoir, C. Gosavi, P. Gade, and B. Alte, "A decision-making tool for creating and identifying face sketches," *ITM Web of Conferences*, vol. 44, May 2022, doi: 10.1051/itmconf/20224403032.
- [16] H. Ge, Y. Dai, Z. Zhu, and B. Wang, "A robust face recognition algorithm based on an improved generative confrontation network," *Applied Sciences*, vol. 11, no. 24, Dec. 2021, doi: 10.3390/app112411588.
- [17] S. Bae, N. Ud Din, H. Park, and J. Yi, "Face photo-sketch recognition using bidirectional collaborative synthesis network," in *2022 16th International Conference on Ubiquitous Information Management and Communication (IMCOM)*, Jan. 2022, pp. 1–8, doi: 10.1109/IMCOM53663.2022.9721719.
- [18] Z. Khan et al., "Face recognition via multi-level 3D-GAN colorization," *IEEE Access*, vol. 10, pp. 133078–133094, 2022, doi: 10.1109/ACCESS.2022.3226453.
- [19] S. P. R. Reddi, M. R. T.V., S. R. P., and P. Bethapudi, "An efficient method for facial sketches synthesization using generative adversarial networks," *Webology*, vol. 19, no. 1, pp. 3119–3129, Jan. 2022, doi: 10.14704/WEB/V19I1/WEB19206.
- [20] S. Yu, H. Han, S. Shan, A. Dantcheva, and X. Chen, "Improving face sketch recognition via adversarial sketch-photo transformation," in *2019 14th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2019)*, May 2019, pp. 1–8, doi: 10.1109/FG.2019.8756563.
- [21] S. Zhang, R. Ji, J. Hu, X. Lu, and X. Li, "Face sketch synthesis by multidomain adversarial learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 5, pp. 1419–1428, May 2019, doi: 10.1109/TNNLS.2018.2869574.
- [22] L. Wang, V. Sindagi, and V. Patel, "High-quality facial photo-sketch synthesis using multi-adversarial networks," in *2018 13th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2018)*, May 2018, pp. 83–90, doi: 10.1109/FG.2018.00022.




- [23] Z.-H. Wang, N. Wang, J. Shi, J.-J. Li, and H. Yang, "Multi-instance sketch to image synthesis with progressive generative adversarial networks," *IEEE Access*, vol. 7, pp. 56683–56693, 2019, doi: 10.1109/ACCESS.2019.2913178.
- [24] W. Wan and H. J. Lee, "A joint training model for face sketch synthesis," *Applied Sciences*, vol. 9, no. 9, Apr. 2019, doi: 10.3390/app9091731.
- [25] H. Bi, N. Li, H. Guan, D. Lu, and L. Yang, "A multi-scale conditional generative adversarial network for face sketch synthesis," in *2019 IEEE International Conference on Image Processing (ICIP)*, Sep. 2019, pp. 3876–3880, doi: 10.1109/ICIP.2019.8803629.
- [26] C. Peng, N. Wang, J. Li, and X. Gao, "DLFace: Deep local descriptor for cross-modality face recognition," *Pattern Recognition*, vol. 90, pp. 161–171, Jun. 2019, doi: 10.1016/j.patcog.2019.01.041.
- [27] J. Zheng, W. Song, Y. Wu, R. Xu, and F. Liu, "Feature encoder guided generative adversarial network for face photo-sketch synthesis," *IEEE Access*, vol. 7, pp. 154971–154985, 2019, doi: 10.1109/ACCESS.2019.2949070.
- [28] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9351, 2015, pp. 234–241.
- [29] X. Luo, X. He, L. Qing, X. Chen, L. Liu, and Y. Xu, "EyesGAN: Synthesize human face from human eyes," *Neurocomputing*, vol. 404, pp. 213–226, Sep. 2020, doi: 10.1016/j.neucom.2020.04.121.
- [30] N. K. Yadav, S. K. Singh, and S. R. Dubey, "TVA-GAN: attention guided generative adversarial network for thermal to visible image transformations," *Neural Computing and Applications*, pp. 1–21, Jan. 2022, doi: 10.36227/techrxiv.14393243.
- [31] S. He *et al.*, "Context-aware layout to image generation with enhanced object appearance," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2021, pp. 15044–15053, doi: 10.1109/CVPR46437.2021.01480.
- [32] Y. Li, X. Chen, B. Yang, Z. Chen, Z. Cheng, and Z.-J. Zha, "DeepFacePencil," in *Proceedings of the 28th ACM International Conference on Multimedia*, Oct. 2020, pp. 991–999, doi: 10.1145/3394171.3413684.
- [33] S. Duan, Z. Chen, Q. M. J. Wu, L. Cai, and D. Lu, "Multi-scale gradients self-attention residual learning for face photo-sketch transformation," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 1218–1230, 2021, doi: 10.1109/TIFS.2020.3031386.
- [34] W. Phusomsai and Y. Limpiyakorn, "Applying GANs for generating image with varied facial attributes from sketch," *Journal of Physics: Conference Series*, vol. 1619, no. 1, Aug. 2020, doi: 10.1088/1742-6596/1619/1/012013.
- [35] 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, Oct. 2022, doi: 10.3390/app121910006.
- [36] N. Balayesu and H. K. Kalluri, "An extensive survey on traditional and deep learning-based face sketch synthesis models," *International Journal of Information Technology*, vol. 12, no. 3, pp. 995–1004, Sep. 2020, doi: 10.1007/s41870-019-00386-8.

BIOGRAPHIES OF AUTHORS



Raghavendra Mandara Shetty Kirimanjeshwara    is a research scholar in the School of ECE at REVA University, Bengaluru, India. Graduation from Karnataka University Dharwad and post-graduation from VTU Belagavi. Total 14 years in teaching and 8 years of industry experience in the engineering field. His areas of interest include AI, embedded systems, and renewable energy. He can be contacted at r19pec11@gmail.com.



Sarappadi Narasimha Prasad    is a professor in the Department of Electrical and Electronics Engineering, Manipal Institute of Technology Bengaluru, Manipal Academy of Higher Education (MAHE), Manipal, Karnataka, India, 576104. His total experience is 22 years, completed graduation from Mangalore University, post-graduation from VTU, and a doctorate from Jain University. More than 80 journals/conferences in profile and presently guiding 8 research scholars. His areas of interest include AI, embedded systems, and signal processing. He can be contacted at sn.prasad@manipal.edu.