International Journal of Computer (IJC)

ISSN 2307-4523 (Print & Online)

https://ijcjournal.org/index.php/InternationalJournalOfComputer/index

Realistic Sketch-based Face Photo Synthesis using Generative Adversarial Networks (GANs)

Hnin Ei Ei Cho[a]*, Aye Min Myat[b]

[a,b]*Faculty of Information and Communication Technology, University of Technology (Yadanarpon Cyber City), PyinOoLwin, Myanmar*

[a]*Email: hnineieicho87@gmail.com*

[b]*Email: awba07@gmail.com*

Abstract

Facial photo-image synthesis and sketch-based face recognition are highly advantageous, particularly in the fields of security forces and forensics. Furthermore, it makes it more feasible for law enforcement to reduce the number of possible suspects in criminal identification operations. However, since pencil drawings and photographs have different properties by nature, creating a synthesis of photographs based on sketches presents a difficult topic. In the last few decades, generative adversarial network-based systems have achieved enormous advances towards improving the performance of image synthesis. It can speed up identification times while improving matching outcomes by reducing gaps among sketch and photo representations. We perform investigations on the well-known photo-sketch pair database CUHK. First, we demonstrate how a generative adversarial network transforms hand-drawn sketches into realistic photos. Secondly, we employ suspect identification by using the pre-trained VGG16-based feature extractor network and KNN classifier. Our technique focuses on the use of deep learning-based networks, which are well-known for their capacity to process data and extract hierarchical features. The presented image-to-image translation framework minimizes the modality differences between hand-drawn face sketches and color images while improving visual quality. Tests on sketch-photo matching demonstrate significant improvements over current state-of-the-art methods on the challenging task of matching sketches with corresponding photos.

Keywords: hand-drawn sketch**;** image to image translation; sketch to photo synthesis; generative adversarial network; sketch-based face recognition.

Received: 4/10/2024
Accepted: 6/2/2024
Published: 6/15/2024

----

* Corresponding author.

1. Introduction

Today, the demand for identification of faces grows with the developments of numerous security-related industries. Fingerprints, faces, and irises are the three biometric identifiers that are most frequently utilized. Facial characteristics are much more beneficial than fingertips and retina scans. Nowadays, among the most important domains of computer vision research is face recognition, which has numerous potential uses in image database analysis, security, and other sectors. The use of facial recognition technology is crucial for helping law enforcement organizations identify an individual photo as a sketch of that person.

It often proves impossible to get a clear image of the suspect's face because any other photos that are accessible are grainy ones that were shot with partial occlusion. People have learned that a facial sketch can be a useful replacement in the search for a suspect. Due to its fundamentally imagined nature, sketching exhibits a notable contrast in expression when compared to photography. Standard facial identification systems have challenges detecting sketches from photos since drawings typically contain simple contours and little grayscale. The typical solution to this challenge involves performing synthesis tasks, followed by implementing sketch-based face recognition using a similar approach. In this regard, studying the tasks of synthesis and recognition concurrently can be beneficial. Two approaches can be used to implement synthesis-based techniques for photo-sketch recognition [1]. The first involves matching with each other in the sketch domains by converting a photo into a line drawing, and the second necessitates transforming a sketch into a photo and then comparing it in the photo domain.

Data-driven and model-driven approaches are the two categories into which conventional face synthesis techniques fall [2]. In the data-driven approach, a sketch is created by linearly integrating identical training sketch patches. Indeed, precisely. The data-driven approach involves two main steps: first, searching for similar sketch patches; and second, determining the weights for a linear combination. The experimental sketch is instantly translated into a photo in model-driven approaches through mapping-based learning. Consequently, model-driven methods allow for the quicker creation of new sketches. In contrast, data-driven approaches tend to increase computation times and expenses. Data-driven methods can lose characteristics that are present only in testing data but not in the training dataset. Model-driven techniques may eliminate some of the facial features' subtle characteristics.

Recently, methods based on deep generative modelling have led to encouraging results in the area of reliable image synthesis research. When given enough labelled data to train on, convolutional neural networks possess the capability to extract intricate visual information at a high level and generate generic feature representations that are transferrable across various datasets. These feature representations can be applied to modify both the style and content of the image, resulting in the production of images that are visually pleasing and closely resemble reality.

Among the numerous generative frameworks currently introduced, deep convolution-based generative adversarial networks (GANs) have been receiving significant interest from the scientific community. Various versions of the classical adversarial network have been used to solve a wide range of computer vision and graphics problems.

In this study, the synthesis of faces is investigated using the adversarial framework. Our goal is to take the grayscale drawings and turn them into corresponding high-quality, color-realistic images of human faces. A hand-drawn sketch image serves as the input for our system. To create an image of the face that the sketch represents, we employ a conditional Generative Adversarial Network (GAN) that has been extensively tuned. The proposed method outperforms the most current techniques for generating synthetic images in terms of the Structural Similarity Index (SSIM).

The main contribution of the proposed system is as follows:

- We propose a generative adversarial network that can handle sketch to photo synthesis challenge.
- While [3] presents a method to perform sketch to photo transformation and shows some preliminary results, this work is to study sketch-based face recognition by synthesizing face photos on given face sketches.
- We perform face recognition tasks on the generated photos by using the pretrained deep learning model. Detailed experiments are conducted to demonstrate improvements in the synthesis results.

This paper includes various sections that encompass the introduction, literature review, the proposed network and its detailed architecture, and datasets with related experimental findings and conclusions.

2. Literature Review

In this section, previous works on face photo-sketch synthesis and recognition techniques are explained. [4] proposed some methods. One is that it synthesized a pseudo-sketch by applying Eigen-transformation to the entire face photo, which is used for matching in the sketch modality. They improved the synthesis framework by applying Eigen-transformation to local patches without considering absolute recognition accuracy. The relative superior performance compared to human performance and the conventional photo-based methods demonstrate the advantage. Then, in [5], a nonlinear method for face sketch synthesis and recognition is presented. In this method, the synthesized pseudo-sketches are generated using local linear preservation of geometry between the sketches and photo images, and the probe sketch is identified using nonlinear discriminate analysis.

Reference [1] synthesized and recognized sketch pictures using a multi-scale MRF model based on the previously suggested references. Extensive experiments were carried out on a face sketch database of 606 faces. [6] improves the basic MRF method by proposing a weighted MRF algorithm called Markov weight fields (MWF). Reference [7] proposes a novel face sketch synthesis method (TFSPS) to improve the robustness of MWF based on transductive learning.

The deep Convolutional Neural Network (CNN) is employed as a complete platform for image-to-image translation challenges. CNN is composed of layers that are both convolutional and de-convolutional. During the training phase, it attempts to minimize the loss while learning the network weights. During the testing stage, the learned weights are applied to the input image to create a new representation of the input. Firstly, [8] studied pixel-wise prediction in semantic segmentation challenges. In that study, convolutional network models can be tuned end-to-end for pixel-wise prediction. A fully automatic colorization method using a deep convolutional architecture was created by [9] for converting grayscale images into color photographs. An FCN (fully convolutional network) approach is utilised by [10] in their innovative approach to handling photo-to-sketch synthesis. The goal of these techniques is to create architecture by treating the visual translation problem as an active research area.

To address the problem of image generation, Reference [11] introduced generative adversarial networks (GAN). Other computer vision applications that use GANs include text-to-image synthesis, driver drowsiness detection or recognition, image super-resolution, domain adaptation, and sketch-to-photo synthesis. Reference [12] established conditional GANs (cGAN). The authors used class labels as additional data and imposed a condition on the generator and discriminator networks of a conventional GAN.

Reference [13] was developed as a general-purpose approach based on an infrastructure of conditional GANs designed to handle image-to-image conversion tasks. Zhu and his colleagues explored CycleGAN [14], an image-to-image modification strategy that also operates on unpaired images. Reference [15] designed DualGAN, a dual-learning system that allows unsupervised translation from one image to another.

Reference [16] is a superior photo-to-sketch synthesis system made up of several adversarial networks at varied image resolutions, as suggested by Wang and his colleagues. In this study, we employed the GANs model for transforming face sketches into photo images that were validated on the CUHK benchmark dataset.

Reference [17] develop a simple face-sketch and face-photo synthesis and recognition system. The model for face to photo synthesis is designed using convolutional neural networks (CNN). The design approach for facial recognition is based on Fisherface Linear Discriminant Analysis (LDA). An accuracy of 70.731% was obtained after testing the facial recognition model.

Reference [18] performed face sketch synthesis using a new-found architecture called spiral-Net, which is a modified U-Net. The results of that network and the NLDA scheme of face recognition demonstrate an improvement in the domain. The literature study emphasizes how facial recognition systems are continually being developed, with researchers continuing to seek ideas from both new technical advancements and traditional methods.

3. Synthesis Network

This section provides an in-depth explanation of the suggested synthesis framework. Additionally, the network architectures of the discriminator and generator are described. Then an illustration of the model's training objective function is provided. We're trying to learn how to map between a photo of your face and the sketch. It is necessary to have a dataset with several photo-sketch pairs. We convert the face sketches into photos given as face images. In Figure. 1, the entire framework is displayed.

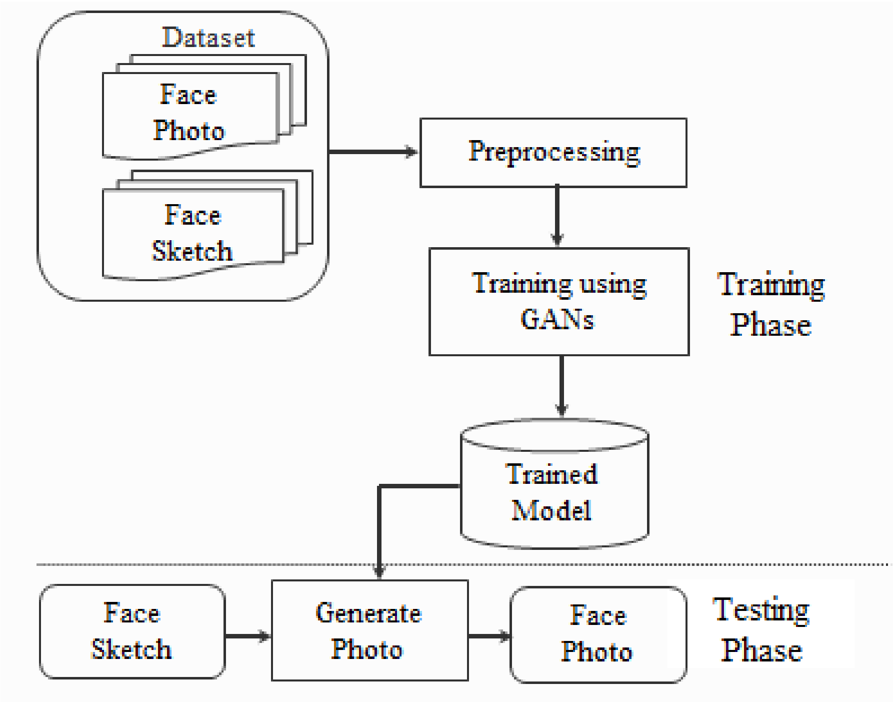

Figure 1: System design for the synthesis system

3.1. Preprocessing

A dataset comprising a number of photo-sketch pairs is required. Before we start training, we split them into two folders: training and testing. To increase quality and speed up convergence, the images being entered must have a certain dimension. To do this, the images shrink from their original size of any size to 256 × 256. Then, the sketch and real images were normalized to have pixel values within -1 and 1, respectively. This kind of approach can accelerate model convergence and lower the influence of lighting variations.

3.2. Training using GANs

For the training of the GAN network, it looks like a zero-sum game. The generator and discriminator are the individuals who participate in this game. The generators' purpose is to confuse the discriminator by producing the images. The discriminator identifies whether the input images are genuine or the generator's products. After which, it provides a classification number. Figure 2 depicts the GAN network's training process. Mode collapse is one of the most frequent GAN failure types. There are many kinds of GANs for different types of tasks. Among them, for image-to-image translation tasks, conditional GANs are less vulnerable to the mode collapse problem, which can be minimized by combining loss functions.

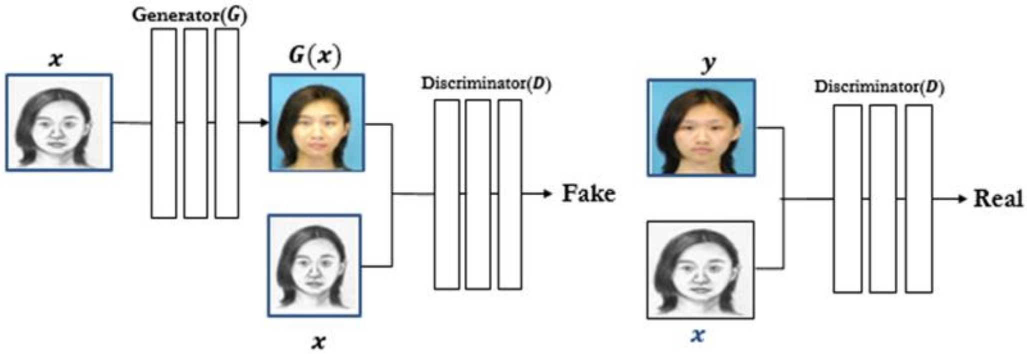

Figure 2: GAN's network training process

3.3. Network Architecture

In this proposed work, we focus on improving the images' quality by training conditional GANs. Conditional GANs provides auxiliary information to the generator and discriminator to enhance network training. Our network's discriminators also accept a synthesized face image and an input sketch image, with the real image serving as auxiliary information, as shown in Figure 3.

Figure 3: An example of an input sketch image, its corresponding photo, and the synthesized image.

3.3.1. Generator

Our generator architecture is based on the "U-Net" architecture, with skip connections added between the encoder and decoder to improve the representation of the layer connection. We adjusted the encoder architecture with the combination of the convolution, batch normalization, and Leaky ReLu layers. The decoder is a combination of deconv+bn+relu except that the activation of the last layer is a tanh function. The convolutions used in the experiments are 4 × 4. Spatial filters with a stride of 2. The architecture of the generator used in our system is shown in Figure. 4.

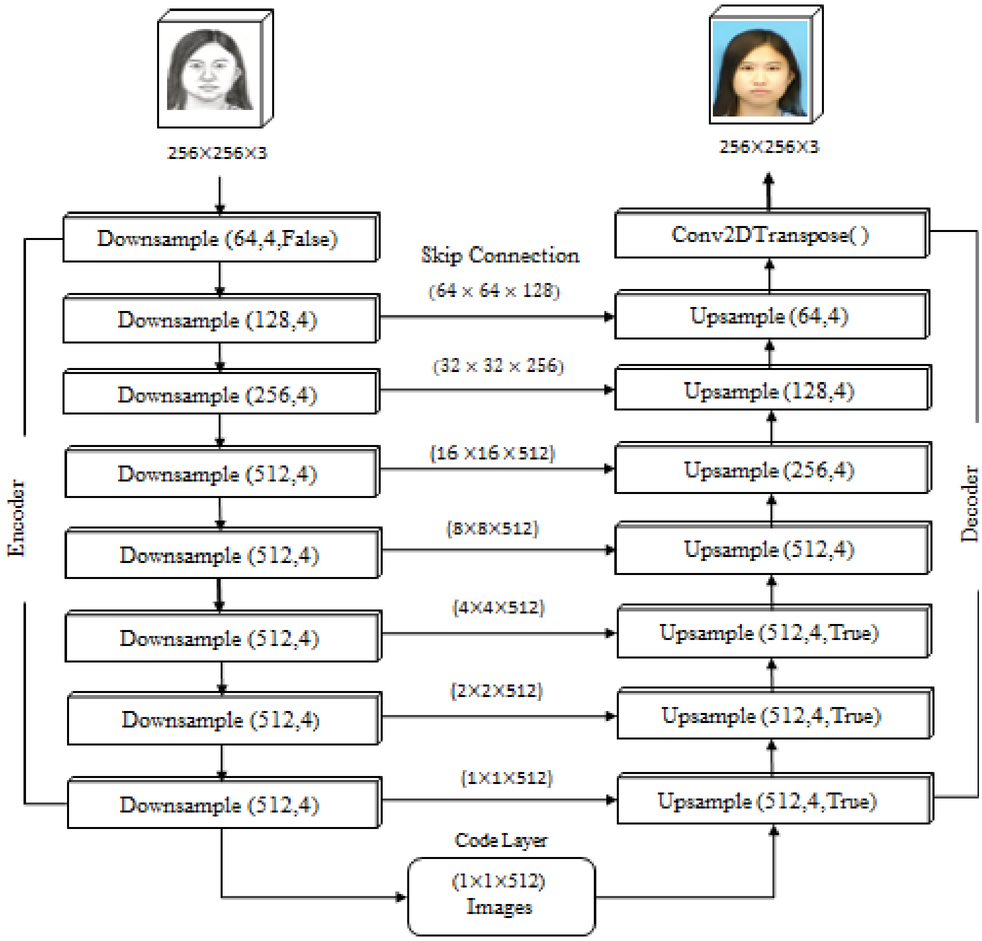

Figure 4: Illustration of the generator architecture.

3.3.2. Discriminator

We apply the PatchGAN network to classify the input image, whether it is real or fake. It distinguishes images by dividing patches instead of using the whole image.

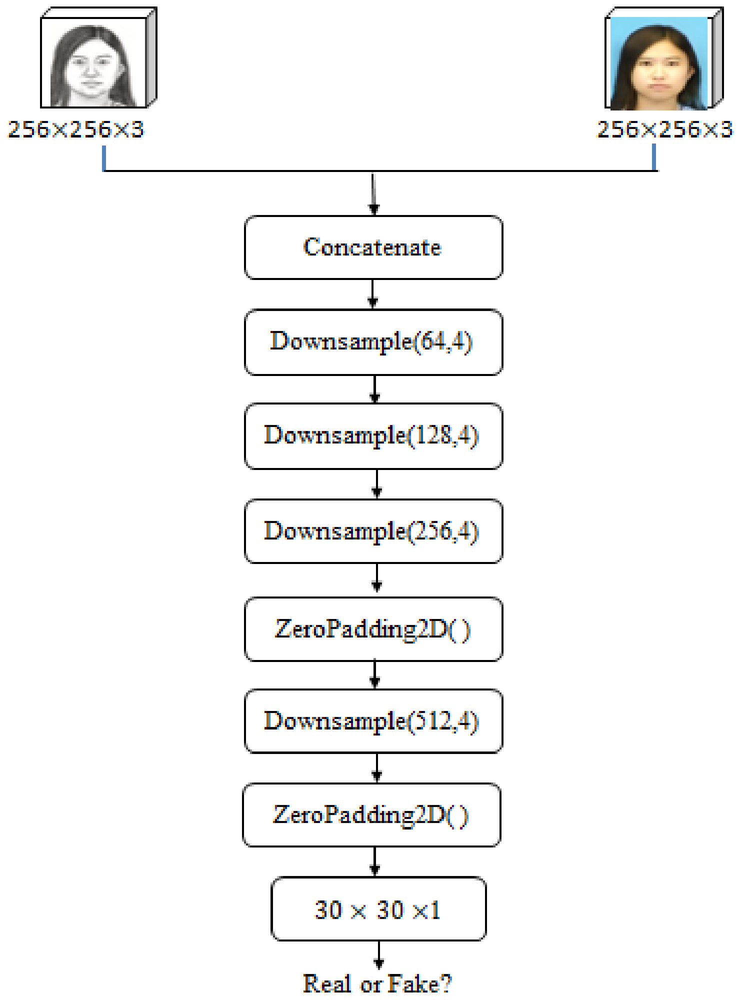

Figure 5: Illustration of the discriminator architecture.

3.3.3. Objective Function

The objective of the conditional GANs that minimizes the generator and maximizes the discriminator is computed as Equation (1):

$$L\_cGAN\ (G,D) = E\_(x,y)\ [logD(x,y)] + E\_(x,z)\ [log\big(1 - D\big(x,G(x,z)\big)\big] \tag{1}$$

$E_{x,y}[log\ D\ (x,y)]$ represents the discriminator's loss. It determines whether the real image is real or fake. The second term denotes the discriminator's loss in predicting whether a synthesized image $G(x,z)$is real or fake.

The discriminator wants to obtain the lowest feasible loss when determining actual and false images. At that time, the generator maximizes this loss. As a result of this, the ideal generator model is the most efficient solution to the min-max problem.

$$G^* = argmin_G\ max_D\ L_{cGAN}\ (G,D) \tag{2}$$

The conditional GANs also employ a new loss that improves image quality by utilising the paired dataset. It is advantageous to combine the cGANs objective with the $L1$ distance. The generator is tasked with not just tricking the discriminator but also getting close to the true result while the discriminator's work remains unaltered.

The $L1$ distance loss for the target y and generated image $G(x,z)$ is:

$$L_{L1}(G) = E_{x,y,z}[\|y - G(x,z)\|1] \tag{3}$$

The final objective function with L1 loss weight is:

$$G^* = \arg min_G max_D L_{cGAN}(G,D) + \alpha\, L_{L1}(G) \tag{4}$$

4. Experimental Details

In this section, we'll use comparison analysis to verify that our proposed approach is working well for tasks involving synthesis and recognition. Output images from the initial synthesis job are used as experimental data in our recognition study.

4.1. Photo Synthesis

4.2. Dataset

The number of datasets that may be used for this kind of work is limited by the difficulty and time required to assemble human-drawn image pairs together with matching face photos. The popular dataset, the CUHK student database, is used in this work to gather photo-sketch pairs. There are 188 pairs of student photo-sketch photos in the CUHK face sketch dataset. Each face to a single photo-sketch pair where the artist's sketch is derived from a photograph shot with neutral expressions and standard lighting. The entire set consists of 188 photos in total, of which 88 have been allocated for training and the remaining 100 for testing.

4.2.1. Implementation Setup

We set the value of batch size to 1 to train both the generator and discriminator networks. We horizontally flip the images with a probability of 0.5 to augment the data. A Gaussian distribution with a mean of 0 and a standard deviation of 0.02 is used to initialize the network's weights. The learning rate is set to 0.0002. The network is optimized using the Adam solver, with the momentum term $\beta 1$ set to 0.5. The entire model is trained on a computer equipped with an Nvidia TITAN RTX GPU for 1000 iterations, which takes around five hours.

4.2.2. Result Analysis

We explored both quantitative and qualitative evaluations to better represent the increased performance of the suggested method. The proposed method is compared against seven benchmark strategies: GAN, Pix2Pix, DualGAN, CycleGAN, PS2GAN, CSGAN, and CDGAN.

- Quantitative Evaluation: The structural similarity index metric (SSIM) is used for the quantitative aspect. It is a metric for determining the visual distinction between a synthesized image and its equivalent ground-truth image. In terms of image composition, it defines structural information as a scene attribute unaffected by brightness and contrast. The value of the metric between window x and window y of size $\mathbf{N} \times \mathbf{N}$ can be formulated as:

$$\mathrm{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x{}^2 + \mu_{xy}{}^2 + c_1)(\theta_x{}^2 + \theta_x{}^2 + c_2)} \tag{5}$$

where μ_x and μ_y refer to the average of x and y.

$\theta_x{}^2$ and $\theta_y{}^2$ refer to the variance of x and y,

σ_{xy} is the covariance of x and y

$c_1 = K_1 L$, $c_2 = K_2 L$

L is the dynamic range of pixel values

K_1=0.01 and $K_2 = 0.03$ by default

The results for CUHK using the SSIM metric are shown in Table 1. The larger SSIM scores indicate that the generated images have better quality. We can see that our suggested approach performs better than all other competing models. When compared to GAN, Pix2Pix, DualGAN, CycleGAN, PS2GAN, CSGAN, and CDGAN, the proposed method improves on each by 40.53%, 25.26%, 19.29%, 16.03%, 18.35, 14.66%, and 10.71, respectively.

Table 1: Performance comparison with state-of-the-art methods on the CUHK dataset

Methods	SSIM
GAN	0.5398
Pix2Pix	0.6056
DualGAN	0.6359
CycleGAN	0.6537
PS2GAN	0.6409
CSGAN	0.6616
CDGAN	0.6852
Proposed	0.7586

- Qualitative Evaluation: To demonstrate the improvement in image quality, we compare and display a few example image results obtained by seven state-of-the-art approaches. In the images created by the CycleGAN, PS2GAN, and CSGAN algorithms, reflections are visible on the sample faces, as shown in Figure. 6. The proposed strategy, on the other hand, is capable of eliminating this effect.

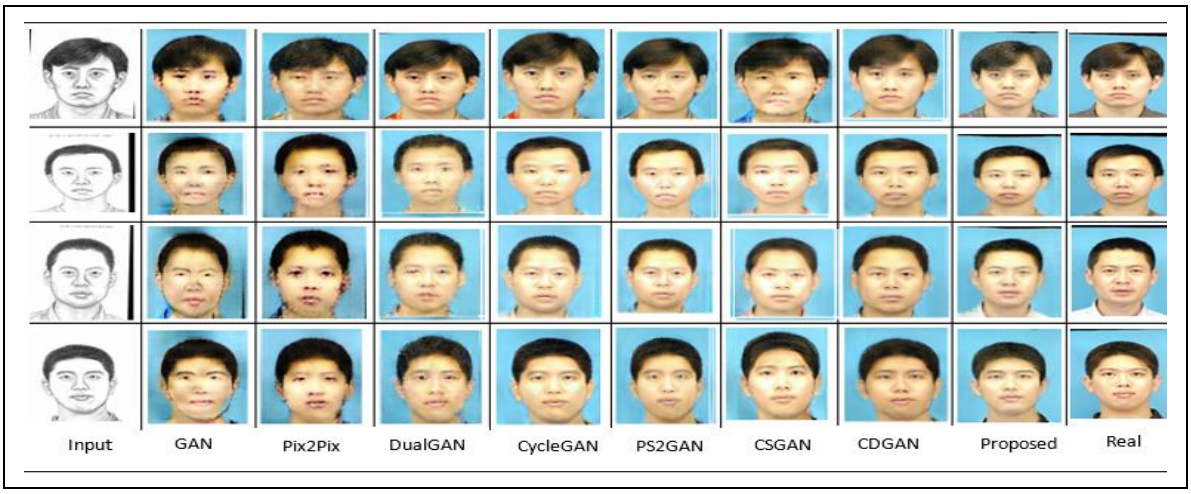

Figure 6: Comparison of the synthesized photo quality.

- Results on Unseen Dataset: In order to evaluate the performance of this approach, we test our trained network on additional, unknown datasets without knowledge of the testing data. We utilize the network trained on CUHK. For sketch-to-photo synthesis, we tested the photos from the AR dataset. Figure. 7 shows some example results, where the synthesized images look quite acceptable, though there is no training on the input datasets.

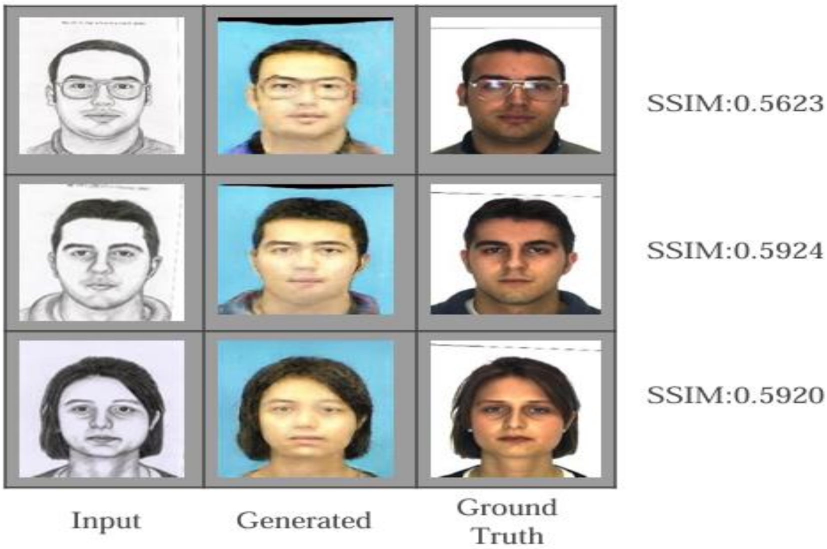

Figure 7: Sample synthesis results on the AR dataset as the training set.

4.3. Sketch-based Face Recognition

Since the primary reason for sketch-photo synthesis is distributing the realistic synthesized images to the public for suspect identification, we conducted face recognition using a pre-trained feature extraction network. Sketch-based face recognition is frequently utilized to quantitatively evaluate the quality of the synthesized photos. To identify the person, we first transform the query sketch into a photo and match the synthesized photo to the gallery photo. The accuracy of the recognition depends on various factors, such as the visual quality of the synthesized sketches as well as the recognition models we apply. In this paper, we utilized a VGG16 pre-trained network to extract the features of the image and a KNN classifier for the recognition process. The block diagram of the sketch-based face recognition process is shown in Figure 8. which contains the feature extraction phase and recognition phase.

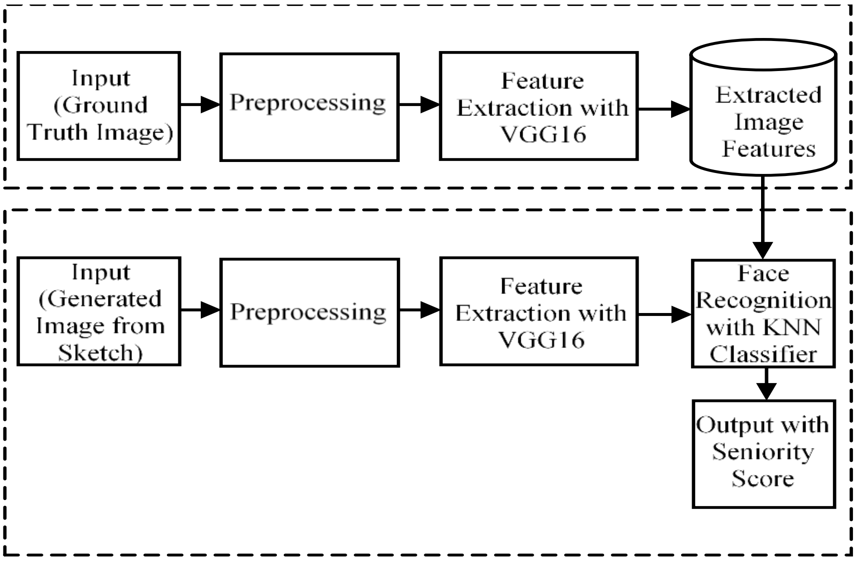

Figure 8: Face Sketch Recognition Process.

In the feature extraction phase, firstly, the ground-truth image is applied to the preprocessing step. The features of ground-truth images are extracted by using a feature extractor based on a VGG16 pre-trained network without a fully connected layer. We only need to extract features of the image. In this extractor, there are thirteen convolutional layers, five Max Pooling layers, and a flattening layer, as shown in Figure. 9. It takes the same input tensor size as 224×224 with three channels. Convolution layers use the filter size 3×3 with stride 1 and the same padding. Maxpool layers use a 2×2 filter with stride 2. Conv-1 Layer has 64 filters, Conv-2 has 128 filters, Conv-3 has 256 filters, and Conv 4 and Conv 5 have 512 filters. The value of the flatten layer's output shape is 25088.

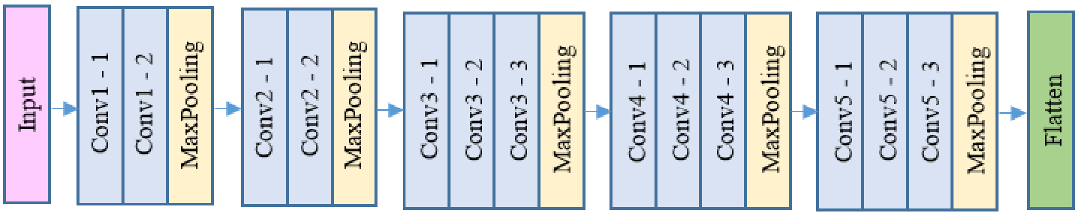

Figure 9: VGG16-based Feature Extractor.

After extracting the features, the image features are added to the feature database file. For the recognition phase, the input is the generated image from the sketch. The preprocessing step is also applied to the input image and extracts the image feature using a VGG16 pre-trained network. The K-Nearest Neighbour (KNN) classifier is used to recognize the face feature from the input based on the feature database. Finally, the results of face recognition are shown with score seniority. Firstly, the face sketch is applied to the photo synthesis task. Then, the result of photo synthesis is recognized using VGG16-based feature extractor networks. We chose to show the recognition results with six face images. These images are calculated with similarity scores. The accuracy scores are classified from rank 0 to rank 6. Rank 1 is the best similarity score, and rank 0 represents the results that do not include the ground truth photo. The proposed recognition method is tested with two types of datasets, CUHK and AR. We took 100 images from each dataset for the recognition test. All the photos from these datasets are used as the galley set. Then, the other four models, VGG19, Inception V3, ResNet50 and MobileNet are performed for the task of recognition to compare with the VGG16 network. The results of face recognition with feature extraction networks on the two datasets are shown in Table 3. According to the results of testing, the VGG16-based feature extraction network attained the highest accuracy of any other network-based face recognition method. Figure 10 illustrates the generated results that include the appropriate ground truth.

Table 3: The recognition results on the CUHK dataset.

Rank	VGG16	VGG19	InceptionV3	Resnet50	MobileNet
1	69	67	0	18	15
2	15	15	1	9	10
3	7	5	3	7	2
4	3	5	2	3	2
5	2	3	0	4	5
6	4	2	0	2	7
0	0	3	94	57	59

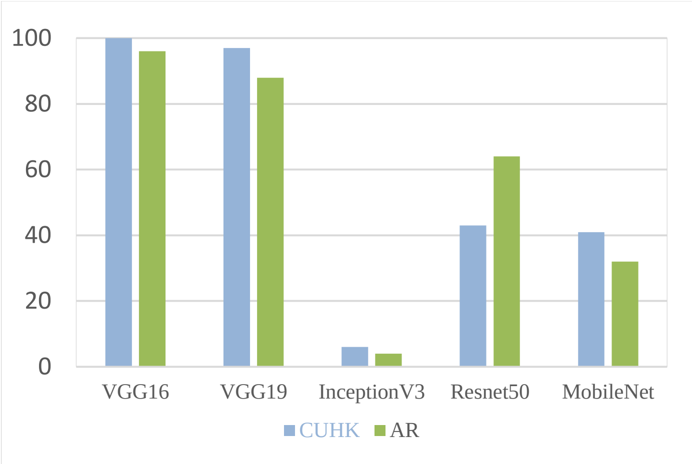

Figure 10: Comparison of the recognition results on the CUHK and AR datasets.

5. Conclusion

A facial sketch image can be turned into a realistic photo by using the suggested system, which has been studied. GAN was used as the baseline network in the development of a generative adversarial network. A novel supervision method is adopted for person identification using deep facial features from a trained recognition network. The proposed recognition method doesn't need to be retrained and can also add a new feature. This fact improves performance and shortens the deep learning training state's period. All results are evaluated on popular datasets (CUHK). The proposed method achieved considerable improvements in terms of both image quality and photo-sketch matching accuracy. The result of recognition shows that the proposed recognition method can recognize the generated output of sketch images precisely and accurately. The dataset utilized for sketch-to-photo conversion is relatively small. Given the limited data available, we plan to explore a variety of training techniques that allow our sketch to photo transformation model to converge more effectively. We plan to improve the face sketch to photo synthesis on a more common dataset by strengthening and stabilizing the connection.

Acknowledgements

I wish to express my profound gratitude to Dr. Aye Min Myat, Professor of the University of Technology in Yatanarpon Cyber City, who is my supervisor, for his excellent suggestions, guidance, consideration, and constant support throughout the preparation of this paper. I'm very grateful to all the compassionate individuals who helped directly and indirectly during the work.

References

[1] Xiaogang Wang and Xiaoou Tang, "Face Photo-Sketch Synthesis and Recognition". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(11), pp.1955-1967.

[2] Wang, Mingrui Zhu, Jie Li, Bin Song, Zan Li, "Data-driven vs. model-driven: Fast face sketch

synthesis”, Neurocomputing,257, pp. 214-221. https://doi.org/10.1016/j.neucom.2016.07.071.

[3] H. E. E. Cho and A. M. Myat, "Colorizing Face Sketch Images for Face Photo Synthesis", *2023 IEEE Conference on Computer Applications (ICCA)*, Yangon, Myanmar, 2023, pp. 119-123, doi: 10.1109/ICCA51723.2023.10182157.

[4] Xiaoou Tang and Xiaogang Wang (2004), “Face Sketch Recognition”, *IEEE Trans. CSVT*, 14(1), 50–57.

[5] Qingshan Liu, Xiaoou Tang, Hongliang Jin, Hanqing Lu andSongde Ma, “A nonlinear approach for face sketch synthesis and recognition”, IEEE *CVPR*, pp.1005-1010.

[6] Hao Zhou, Zhanghui Kuang and Kwan-Yee K. Wong, “Markov Weight Fields for face Sketch Synthesis”, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Providence, Rhode Island, June 16-22,2012, pp. 1091-1097.

[7] Nannan Wang, Dacheng Tao, Xinbo Gao, Xuelong Li, and Jie Li, “Transductive Face Sketch-photo Synthesis”, *IEEE Transactions on Neural Networks and Learning Systems*,24(9), pp. 1364-1376.

[8] Jonathan Long, Evan Shelhamer and Trevor Darrell, “Fully Convolutional Networks for Semantic Segmentation”, *IEEE Conference on Computer Vision and Pattern Segmentation*, Boston, Massachusetts, June 7-12, 2015, pp. 3431-3440.

[9] Gustav Larsson, Michael Maire, and Gregory Shakhnarovichm “. Learning representations for automatic colorization”, In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds) Computer Vision – ECCV 2016. Lecture Notes in Computer Science, vol 9908. Springer, Cham, pp. 577–593.

[10] Liliang Zhang, Liang Lin, Xian Wu, Shengyong Ding and Lei Zhang, “End-to-end photo-sketch generation via fully convolutional representation Learning”, Proceedings of the 5th ACM on International Conference on Multimedia Retrieval, Shanghal, China, June 23-26, 2015, pp. 627-634.

[11] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozairy, Aaron Courville and Yoshua Bengio. “Generative adversarial networks”, In Advances in neural information processing systems, pp. 2672-2680.

[12] Mehdi Mirza and Simon Osindero, “Conditional Generative Adversarial Nets”, arXiv preprint arXiv:1411.1784, 2014.

[13] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou and Alexei A. Efros, “Image-to-image translation with conditional adversarial networks”, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5967-5976.

[14] Jun-Yan Zhu, Taesung Park, Phillip Isola and Alexei A. Efro, “Unpaired image-to-image translation using cycle-consistent adversarial networks”, *IEEE International Conference on Computer Vision,*

Venice, Italy, October 21-29, 2017, pp. 2242-2251.

[15] Zili Yi1, Hao Zhang, Ping Tan, and Minglun Gong, “Dualgan: Unsupervised dual leaning for image-to-image translation”, *IEEE International Conference on Computer Vision*, Venice, Italy, October 21-29, 2017, pp. 2868-2876.

[16] Lidan Wang, Vishwanath A. Sindagi, Vishal M. Pate, “High-quality facial photo-sketch synthesis using multi-adversarial networks”, *IEEE International Conference on Automatic Face & Gesture Recognition*, Xi’an, China, May 15-19, 2018, pp. 83-90.

[17] Xiaogang Wang and Xiaoou Tang, “Face Sketch-Photo Synthesis and Recognition”. *Third International Conference on Image Processing and Capsule Networks*, Bangkok, Thailand, May 20-21, 2022, https://doi.org/10.1007/978-3-031-12413-6_7.

[18] Irfan Azhar, MuhammadSharif, Mudassar Raza, Muhammad Attique Khan and Hwan-SeungYong, “A decision support system for Face Sketch Synthesis using deep learning and artificial intelligence”, *Sensors* 2021, *21*(24),