Recover Realistic Faces from Sketches

Khoa Tan Truong[1,2,3], Khai Dinh Lai[1,2,3], Sang Thanh Nguyen[3], and Thai Hoang Le[1,2(✉)]

[1] Faculty of Information Technology, University of Science, Ho Chi Minh City, Vietnam
lhthai@fit.hcmus.edu.vn
[2] Vietnam National University, Ho Chi Minh City, Vietnam
[3] Faculty of Information Technology, Saigon University, Ho Chi Minh City, Vietnam
{truongtankhoa,laidinhkhai,thanhsang}@sgu.edu.vn

Abstract. Currently, Generative Adversarial Networks (GANs) is considered as the best method to solve the challenge of synthesizing realistic images from sketch images. However, the effectiveness of this method depends mainly on setting up a loss function to learn the mapping between sketches and realistic images. This leads to how to choose an optimal loss function to map them. In this paper, we investigate and propose a loss function that combines pixel-based error and context-based error on a proper ratio to obtain the best training result. The proposed loss function will be utilized to train the generator's U-Net architecture in greater detail. To convert a drawing to an actual image, the trained architecture will be applied. Based on two metrics that are the Structural Similarity Index (SSIM) and visual observations, the assessment results on the CUHK Face Sketch Database (CUFS), AR database (AR), and the CUHK ColorFERET Sketch Database (CUFSF) prove that the suggested method is feasible.

Keywords: Face sketch to image translation · Generative adversarial networks (GANs) · Sketch-based synthesis · Face image generation · Spatial attention · Dual generator · Conditional generative adversarial networks

1 Introduction

The application of image processing in criminal tracking is a much concerning issue today. By applying computer vision, we can 'translate' an input image to a corresponding output one. This is quite useful when we have an input sketch of a criminal through the description of the witness, then we can reproduce a photo-realistic face of that criminal [1]. Traditionally, current methods suggest different techniques to solve the problem, but the general idea is mostly the same: predict pixels from pixels [2–5]. Our goal in this paper is to improve a loss function in GANs [6] to enhance the realistic image prediction from the sketch.

P. Cong Vinh and A. Rakib (Eds.): ICCASA 2021, LNICST 409, pp. 105–118, 2021.
https://doi.org/10.1007/978-3-030-93179-7_9

GANs have been greatly evolved over the past decade and many of the techniques we explore in this article have been suggested previously [7–9]. Within a deep-learning network, like GANs, the most prevalent method is to minimize a loss function which is the main target for scoring the quality of results. Though the learning process is automatic, we must still spend much manual effort on designing effective loss functions. Besides, previous articles have focused on specific applications and it has remained limitations that we can optimize more for translating images. Our contribution is to customize the loss function based on pixels error and contextual error and suggest a new evaluation for synthesized images based on the Structural Similarity Index Measure (SSIM) [10] and visual inspection.

In this paper, the authors suggest a solution to enhance the quality of synthesizing photo/sketch in GANs based on improving the loss function by combining pixel and context loss evaluations. Concurrently, the paper also suggests a method to evaluate experimental results via SSIM and visual inspection.

2 Background and Related Work

2.1 Background

Currently, there are many research groups in the world that perform facial image synthesis from sketch images in many different ways. In general, these methods can be classified into two categories: data-oriented methods and model-oriented methods. Previously, to synthesize a photo/sketch, researchers used to rely on linear matching of similar training photo/sketch patches [11–16]. These approaches include two major sections: search similar photo/sketch and calculate linear association weight. Such processes consume a lot of computational time. For model-oriented method, it is essential to learn an offline mathematical function to map the image to sketch or inverse [17–20]. Usually, researchers must figure out handmade features, neighbor searching strategies, and learning techniques. However, these methods often cause image blur and large distortion in results.

Recently, the appearance of deep convolutional neural networks [13, 21, 22] has come up with great solutions for photo synthesis. Among them, Generative Adversarial Networks (GANs) [6] is the most effective method. The GAN training model behaves like a zero-sum game between the generator and the discriminator. The aim of discriminator is to decide whether a specific image is fake or real, while the generator tries to create realistic images much sophisticated that the discriminator cannot distinguish they are real or fake. Sketch-based image synthesis can be constructed as a conditional image translation problem based on an input sketch (see Fig. 1). There are several methods of using GAN to translate images from one domain to another [20, 23]. However, they are not specifically designed for photo synthesis from facial sketches.

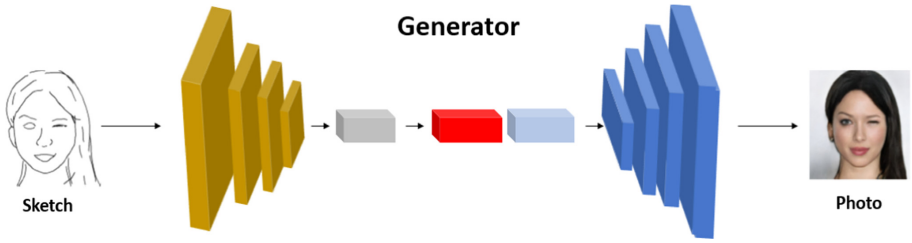

Fig. 1. General model for image translation in GANs

2.2 Related Work

Sketch to Photo Conversion in GANs. Through many studies, GANs have proven effective in creating realistic and natural images [24, 25]. Instead of directly optimizing and correcting per-pixel errors, which often affect the quality of the synthesized image, the GAN uses a discriminator to distinguish the unrealistic image among real images, then it forces the generator to produce sharper images. In the paper "pix2pix" of Isola et al. [26], he illustrated a direct approach to translate an image into another using conditional GANs. This motivates many researches on image-to-image translation methods afterward like CoupledGAN [27], CycleGAN [28], etc. These methods all yield good and promising results.

Sketch-Based Datasets. There are a few datasets on hand-drawing sketches, and they are usually small due to the effort needed to make hand-drawing images. One of the most common sketch datasets is CUHK Face Sketch dataset (CUFS) [29] which contains 188 faces from the Chinese University of Hong Kong's students. Besides, we have experienced on AR sketch dataset (AR) [30] and the CUHK ColorFERET Sketch Database (CUFSF) [31, 32].

3 Proposal Methods for Sketch-Image Translation

3.1 GANs Model Architecture

In this section, we present a Generative Adversarial Network framework that convert input sketches to images in Fig. 2. Our GANs model uses a U-Net [33] based generator with seven convolutional layers and a discriminator with four convolution layers. Here, the special improved point is the loss function which combines two evaluation methods: pixel loss (Sect. 3.2) and context loss (Sect. 3.3).

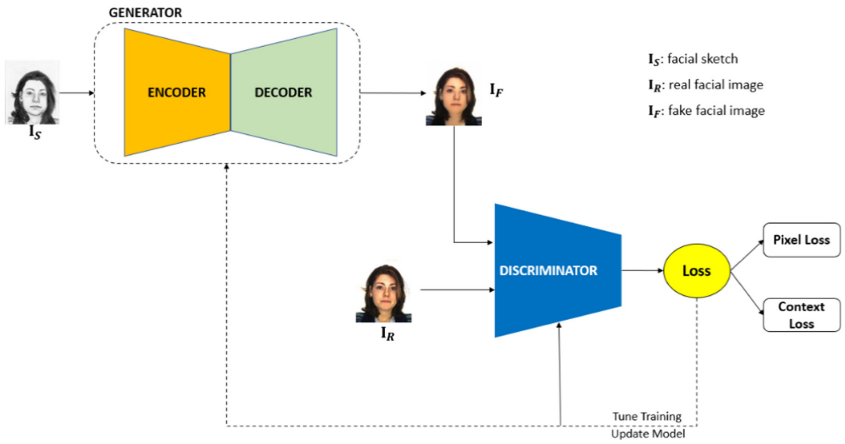

Fig. 2. Training model using GANs in proposed method

Figure 3 shows the model of the generator network based on the U-Net symmetric architecture, with n encoding units and n decoding units. The model uses a 4 × 4 convolutional filter with stride 2 and a Leaky-ReLU [34] activation function with a slope of 0.2 for downsampling. The number of channels will be doubled with each successive layer during this process. For upsampling, the preceding step's output will be doubled in size. Next, the model also uses a 4 × 4 filter with stride 1 and the ReLU [34] activation function to produce the convolution. Following batch normalization and the ReLU activation function, the output will be normalized and concatenated with the activation map of the mirrored layer in the contracting path. The network's last layer is a 1 × 1 convolution, which is the same as a cross-channel parametric pooling layer. For the last layer, we use the "*tanh*" function [35].

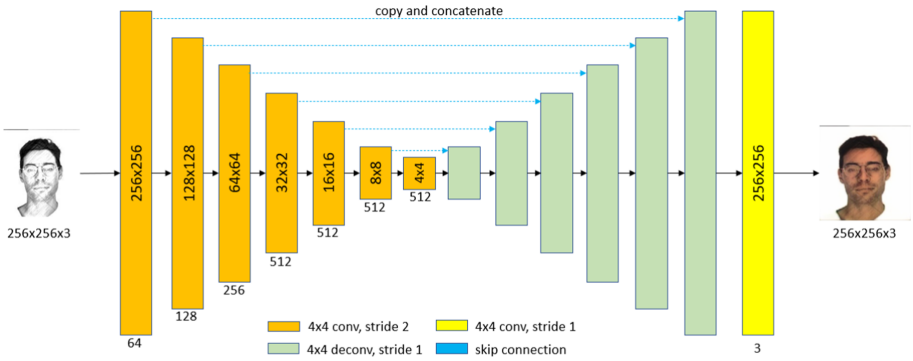

Fig. 3. The "U-Net" Generator network's architecture

Furthermore, we utilize the PatchGAN architecture [36] for the discriminator: 4×4 convolutional layers with stride 2 and the number of channels doubled after each downsampling. As illustrated in Fig. 4, all convolutional layers are followed by batch normalization and a Leaky-ReLU activation function with a slope of 0.2. For the last layer, we do a convolution operator with a filter size $4 \times 4 \times 1$ and stride 1.

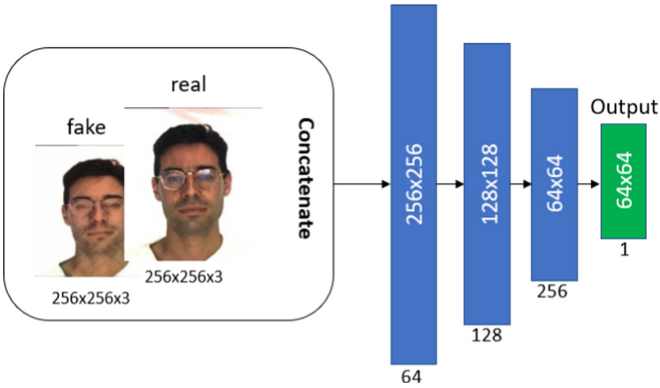

Fig. 4. Architecture of the Discriminator network

3.2 Pixel Loss

To evaluate the pixel difference between a true image and a predicted image in our model, we suggest to use the Mean Absolute Error (MAE) [37] to perform. The formula of this regression loss function is as (1).

$$\text{Pixel loss} = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \tag{1}$$

Where:

y_i is ground truth image which is used for training.
x_i is faked image generated by the generator.

3.3 Context Loss

The main idea behind this function is that it assumes the image as a collection of features, and then determines the similarity between the images by measuring the similarity between the features. This loss function allows the local deformation of the image to a certain extent, therefore the requirement for the data to be aligned at the pixel level is moderate. In addition, this function also aims to constrains the local features, which enables it to operate on the region with similar semantics. Specifically, it first finds similar features in these regions with similar semantic meanings and forms a match between these features.

The loss function is operated based on following coding steps: 1) convert two images (real and fake image) to grayscale; 2) normalize them to range [0, 1]; 3) divide non-zero pixel each other in two pixel-matrix; 4) reduce the same pixels (e.g., pixels in boundary) because their ratios are similar (Fig. 5). Mathematically, the context similarity between the uncorrupted portions, e.g., portions in ground truth and translated image, are measured by a contextual loss [38], which is defined in formula (2).

$$\mathcal{L}_{\text{contextual}}(z) = D_{\text{KL}}(M \odot y, M \odot G(z)) \tag{2}$$

where M is the binary mask of corrupted ground truth image, z is the input sketch, and $\odot$ denotes the Hadamard production [39]. We use KL-divergence [40] to compare two images: a generator-generated image G(z) and a ground truth image y. All of these photos have been binary-normalized previously. If two images y and G(z) are exactly the same ideally, then $\mathcal{L}_{\text{contextual}}(z) = 0$, but if they aren't, we punish G(z) for not producing the same image as the ground truth y.

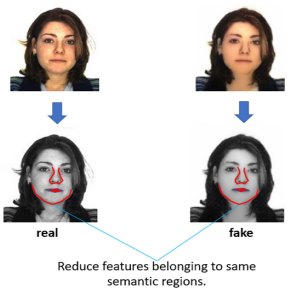

Fig. 5. Illustration for context loss evaluation

3.4 Total Loss

To increase sharpness and ensure similar semantic structure of images synthesized by the generator, we suggest to unify both previous loss function based on formula (3).

$$\text{Total loss} = \lambda * \text{Pixel loss} + \beta * \text{Context loss} \tag{3}$$

where λ is the important coefficient of pixel loss, β is the important coefficient of context loss, and $\lambda + \beta = 1$.

4 Experiment

4.1 Experiment Settings

Dataset Splitting. We experiment the proposed approach on three sketch-photo databases: the CUHK Face Sketch Database (CUFS), the AR database (AR) and the

CUHK ColorFERET Sketch Database (CUFSF). In detail, CUHK includes 188 faces, AR includes 123 face in various ages and CUFSF includes 1194 persons from the FERET database [41]. For each person, there are a face photo with lighting variation and a sketch with shape exaggeration drawn by an artist when viewing this photo. To experiment, we divide to two rounds. For the first round, in Sect. 4.2 and 4.3, we train on a small dataset mixed from CUFS and AR based on the following rule: choose randomly 90 sketch-photo pairs which compose about 50% pairs from CUFS and 50% remains from AR. In order to testing, we apply the same strategy to form 5 subsets with size: 20, 40, 60, 80, 100. The aim of the first round that combine SSIM and visual observation to find the most proper total loss from which the generator in Fig. 2 is trained. To prepare for the second round suggested in Sect. 4.4, we build a larger dataset mixed from CUFSF and CUFS based on database splitting strategy in [42]. All these images are resized in dimension 256×256. The aim of second round is to compare the generator trained from the total loss in the first round to state-of-art methods.

Implementation Details. In all experiments, we use a batch size of 16 and epoch is 100. Besides, we use the Adam optimizer [43], and set the initial learning rate 0.0001 to both generator and discriminator.

Experiment Environment. All experiments are implemented on a server PC with the configuration: Intel(R) Xeon(R) CPU E5-2609 v4 @1.70 GHz, 16 cores, 32 GB RAM, GPU NVIDIA TITAN Xp 12 GB.

Evaluation Metrics. To evaluate our task of image synthesis, we propose to use the Structural Similarity Index Measure metric to build a simple recognition framework in Fig. 6. The principle of this framework is quite simple. It is used to check if our synthesized image from the sketch is realistic enough for the facial recognition system to successfully identify the subject. The synthesis process is judged successful if the translated image is acknowledged as being comparable to the true image of the subject itself. Extensively, for each sketch-photo pair in testing subset, we pass the sketch to the generator obtained in Fig. 2 to translate it to a 2D realistic photo. Then, we compare this realistic one to the photo in the same pair with the sketch translated and to other photos in other pairs using the SSIM metric. Since all images are labeled, so we can easily calculate the accuracy of the method. In addition, we also carry out a perceptual study to consider how realistic the generated images are and how faithful they are to the input sketches. Because the suggested method is tested mostly based on the structure, so we concern much on color and visual quality on synthesized images. To resolve the problem, we manually take randomly some output samples and observe them by eyes. Fortunately, all results are as expected though some cases are abnormal.

4.2 Experimental Results

In the first testing round, we test the effect of training model based on evaluating total loss function mentioned in Sect. 3.4. This function includes one parameter tuple (λ, β) that describes combination between pixel loss and context loss. First, we initialize (λ, β) to (0, 1) and combine them 11 times with shift step 0.1 as following Table 1.

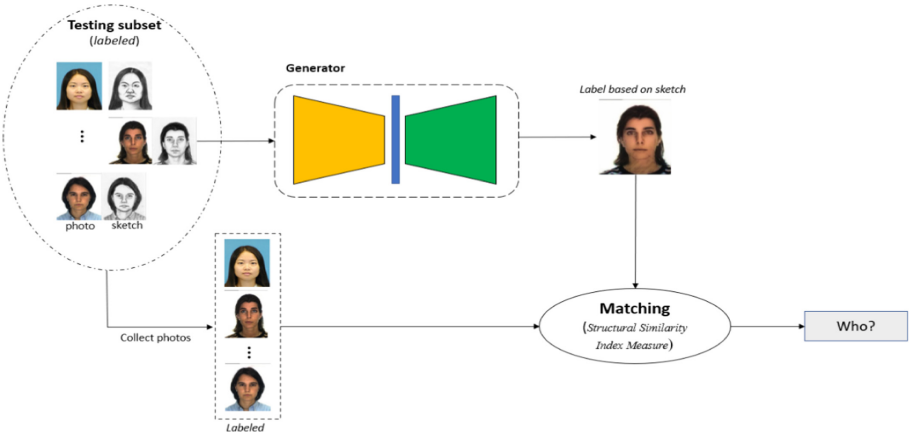

Fig. 6. Proposed recognition framework for testing

Table 1. Splitting training configurations

Total loss configuration	λ	β
0_10	0	1
1_9	0.1	0.9
2_8	0.2	0.8
3_7	0.3	0.7
4_6	0.4	0.6
5_5	0.5	0.5
6_4	0.6	0.4
7_3	0.7	0.3
8_2	0.8	0.2
9_1	0.9	0.1

For each setting, we retrain the generator model and use the testing framework in Fig. 6 to calculate accuracy scores through 5 data subsets (20, 40, 60, 80, 100).

According to Table 2, some configurations such as 0_10, 2_8, 3_7, 5_5, 7_3, 10_0 get high accuracy. Therefore, we use them to compare each other (see Fig. 7). The results in Table 2 demonstrate the effectiveness of this approach when we alter the combined ratio of pixel loss and context loss. Thereby, we learn that when the context loss is scaled down, the matching based on Structural Similarity Index Measure metric is affected negatively because the metric is mainly calculated based on context and semantics in images.

Table 2. Accuracy statistics of generator model in testing round 1

Configuration	Subset				
	20	40	60	80	100
0_10	95.00	87.50	88.33	78.75	82.00
1_9	95.00	85.00	86.67	78.75	78.00
2_8	95.00	85.00	86.67	78.75	80.00
3_7	95.00	85.00	85.00	78.75	81.00
4_6	95.00	85.00	85.00	78.75	80.00
5_5	95.00	85.00	85.00	77.50	80.00
6_4	95.00	87.50	86.67	78.75	80.00
7_3	95.00	85.00	86.67	78.75	80.00
8_2	95.00	82.50	85.00	77.50	79.00
9_1	90.00	87.50	88.33	78.75	78.00
10_0	90.00	85.00	86.67	78.75	83.00

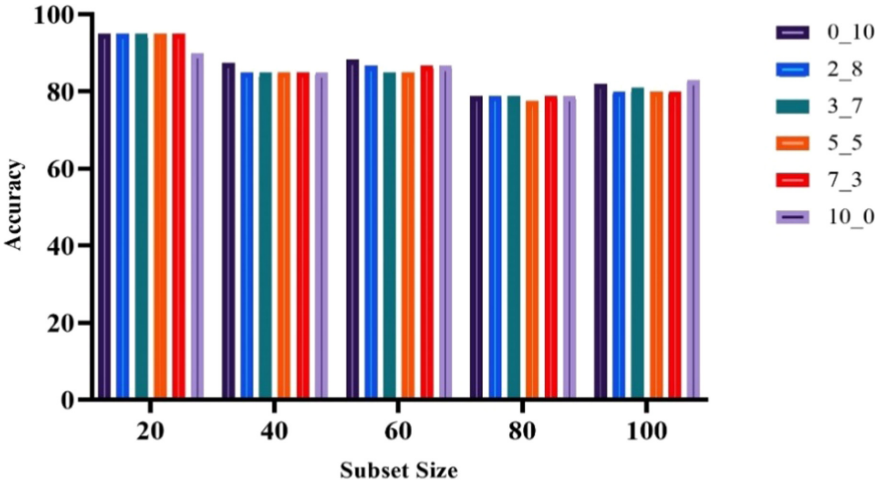

Fig. 7. Accuracy comparison for training configurations

4.3 Perceptual Validation

In addition, as shown in Fig. 7, we use the SSIM metric to evaluate the generator model, and we also use human visual perception to validate the results. Evaluation in Sect. 4.2 mainly depends on similar contextual structures, not rely on the color. We took effort to look over generated results when applying each different training configuration. After observing whole outputs over 11 settings of training configurations, we recognize that: when optimizing the generator based on high ratio of the pixel loss function, the color quality of outputs is realistic and more similar to the ground truth though they are a little blurry and unsharp; in contrast, if we decrease the ratio of the pixel loss and increase the ratio of the context loss, the output is sharp, but their color is distorted unrealistically (Fig. 8).

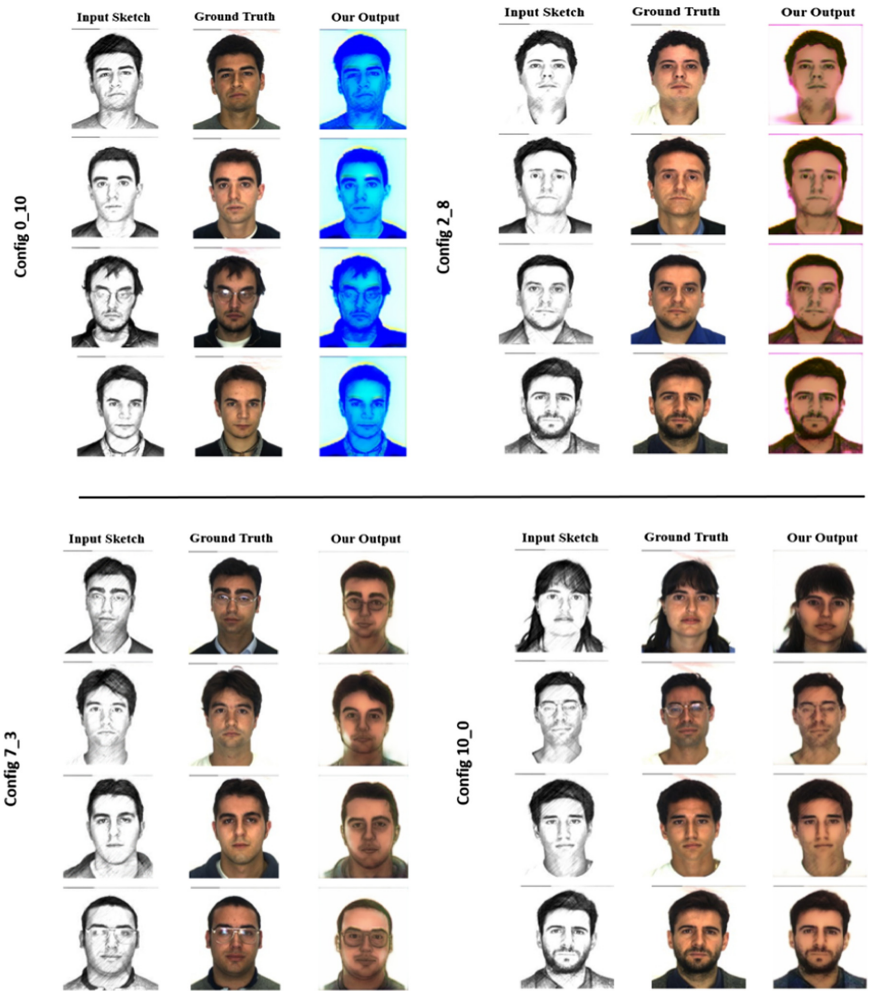

Fig. 8. Visual observation on results of proposed method over configurations

4.4 Comparison with Other Methods

Figure 7 shows an analysis of the experimental results in Sect. 4.2, which shows that the results are nearly identical. However, configurations 0_10, 2_8, 3_7 favor contextual factors that are not close to human vision (evaluating images through color); the configuration 5_5 balanced between pixel color and contextual elements also does not properly reflect human visual judgment of the image (in favor of pixel color values rather than contextual factors); configuration 10_0 does not care about the context factor, only cares about the color of the pixel, will result in a not sharp (blurred). More over, based on our visual observations in Sect. 4.3, we discover that configuration 7_3 is the best choice. Therefore, the generator trained from configuration 7_3 will be used

for image synthesis. Table 3 compares the proposed method's accuracy to that of some other modern approaches mentioned in [42]. In summary, the proposed strategy is feasible and produces positive outcomes on both huge data sets, according to statistics.

Table 3. A comparison of our method to others in testing round 2

Methods	SSIM
Baseline	0.487
cGAN-Final generator 1	0.579
cGAN-Final generator 2	0.524
cGAN-Initial generator 1	0.581
cGAN-Initial generator 2	0.608
cGAN-Both generator 1	0.564
cGAN-Both generator 2	0.529
cGAN-Final gray	0.526
cGAN-Final w/ spectral norm & self-attn, gen. 1	0.571
cGAN-Final w/ spectral norm & self-attn, gen. 2	0.530
Ours	**0.690**

5 Conclusions

With the aim of building a model with the ability to translate sketch to realistic image, our paper suggests some modifications to traditional GANs and gain some results promising. Based on analysis in Sect. 4.4, we suggest training configure 7_3 where the total loss defined as the following: 0.7 * Pixel loss + 0.3 * Context loss is the best one used for training model. Summarily, in the paper, we have three main contributions that are: 1) Improve the GAN training model to select an effective generator by the combination of pixel loss and context loss according to the set of scale factors (λ, β), of which the prominent is the proposal to use the context loss function to help ensure the basic structure of the face; 2) Suggest a new recognition framework using SSIM metric to test accuracy of realistic images translated by generator; 3) Finally, evaluate results by visual perception.

Moreover, the quality of synthesized images are still a challenge, we will be going to research and innovate the structure of the GANs to win better. The paper contributes a new approach to customize the existing GANs model to solve an interesting problem: translate sketch to image. It's a meaningful way to help police to detect criminals quickly and effectively.

References

1. Wang, N., Tao, D., Gao, X., Li, X., Li, J.: A comprehensive survey to face hallucination. Int. J. Comput. Vis. **106**(1), 9–30 (2014)
2. Efros, A.A., Freeman, W.T.: Image quilting for texture synthesis and transfer. In: SIGGRAPH (2001)
3. Hertzmann, A., Jacobs, C.E., Oliver, N., Curless, B., Salesin, D.H.: Image analogies. In: SIGGRAPH (2001)
4. Chen, T., Cheng, M.-M., Tan, P., Shamir, A., Hu, S.-M.: Sketch2Photo: internet image montage. ACM Trans. Graph. (TOG) **28**(5), 124 (2009)
5. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: CVPR (2015)
6. Goodfellow, I.J., et al.: Generative adversarial nets. In: International Conference on Neural Information Processing Systems, pp. 2672–2680 (2014)
7. Wang, N., Zha, W., Li, J., Gao, X.: Back projection: an effective postprocessing method for GAN-based face sketch synthesis. Pattern Recogn. Lett. **107**, 59–65 (2018)
8. Di, X., Patel, V.M.: Face synthesis from visual attributes via sketch using conditional VAEs and GANs. arXiv:1801.00077 (2017)
9. Zhang, S., Ji, R., Hu, J., Gao, Y., Chia-Wen, L.: Robust face sketch synthesis via generative adversarial fusion of priors and parametric sigmoid. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18), pp. 1163–1169 (2018)
10. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)
11. Song, Y., Zhang, J., Bao, L., Yang, Q.: Fast preprocessing for robust face sketch synthesis. In: Proceedings of International Joint Conference on Artificial Intelligence, pp. 4530–4536 (2017)
12. Song, Y., Bao, L., He, S., Yang, Q., Yang, M.-H.: Stylizing face images via multiple exemplars. Comput. Vis. Image Underst. **162**, 135–145 (2017)
13. Gao, X., Wang, N., Tao, D., Li, X.: Face sketchphoto synthesis and retrieval using sparse representation. IEEE Trans. Circuits Syst. Video Technol. **22**(8), 1213–1226 (2012)
14. Song, Y., Bao, L., Yang, Q., Yang, M.-H.: Real-time exemplar-based face sketch synthesis. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8694, pp. 800–813. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10599-4_51
15. Pan, Q., Liang, Y., Zhang, L., Wang, S.: Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis. In: Computer Vision and Pattern Recognition, pp. 2216–2223 (2012)
16. Wang, X., Tang, X.: Face photo-sketch synthesis and recognition. IEEE Trans. Pattern Anal. Mach. Intell. **31**(11), 1955–1967 (2009)
17. Peng, C., Gao, X., Wang, N., Tao, D., Li, X., Li, J.: Multiple representations-based face sketch-photo synthesis. IEEE Trans. Neural Netw. Learn. Syst. **27**(11), 2201–2215 (2016)
18. Zhang, S., Gao, X., Wang, N., Li, J., Zhang, M.: Face sketch synthesis via sparse representation-based greedy search. IEEE Trans. Image Process. **24**(8), 2466–2477 (2015)
19. Zhang, S., Gao, X., Wang, N., Li, J.: Robust face sketch style synthesis. IEEE Trans. Image Process. **25**(1), 220 (2016)
20. Wang, N., Tao, D., Gao, X., Li, X., Li, J.: Transductive face sketchphoto synthesis. IEEE Trans. Neural Netw. Learn. Syst. **24**(9), 1364–1376 (2013)
21. Peng, C., Wang, N., Li, J., Gao, X.: Face sketch synthesis in the wild via deep patch representation-based probabilistic graphical model. IEEE Trans. Inf. Forensics Secur. **15**, 172–183 (2020)

22. Wang, L., Sindagi, V., Patel, V.: High-quality facial photo-sketch synthesis using multi-adversarial networks. In: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), pp. 83–90. IEEE (2018)
23. Zhang, L., Zhang, L., Mou, X., Zhang, D.: FSIM: a feature similarity index for image quality assessment. IEEE Trans. Image Process. **20**(8), 2378 (2011)
24. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.C.: Improved training of Wasserstein GANs. In: Advances in Neural Information Processing Systems, pp. 5769–5779 (2017)
25. Nguyen, A., Clune, J., Bengio, Y., Dosovitskiy, A., Yosinski, J.: Plug & play generative networks: conditional iterative generation of images in latent space. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017
26. Isola, P., Zhu, J.-Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017
27. Liu, M.-Y., Tuzel, O.: Coupled generative adversarial networks. In: Advances in Neural Information Processing Systems, pp. 469–477 (2016)
28. Zhu, J.-Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: The IEEE International Conference on Computer Vision (ICCV), October 2017
29. Zhang, W., Wang, X., Tang, X.: Coupled information-theoretic encoding for face photo-sketch recognition. In: CVPR 2011, pp. 513–520 (2011). https://doi.org/10.1109/CVPR.2011.5995324
30. Martinez, A.M., Benavente, R.: The AR Face Database. CVC Technical Report #24, June 1998
31. Wang, X., Tang, X.: Face photo-sketch synthesis and recognition. IEEE Trans. Pattern Anal. Mach. Intell. (PAMI) **31**(11), 1955–1967 (2009)
32. Zhang, W., Wang, X., Tang, X.: Coupled information-theoretic encoding for face photo-sketch recognition. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2011)
33. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **39**(4), 640–651 (2014)
34. Maas, A.L., Hannun, A.Y., Ng, A.Y.: Rectifier nonlinearities improve neural network acoustic models. In: Proceedings of the ICML, vol. 30 (2013)
35. Abramowitz, M., Stegun, C.A. (eds.): Hyperbolic Functions. §4.5 in Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing, pp. 83–86. Dover, New York (1972)
36. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5967–5976 (2016)
37. Willmott, C.J., Matsuura, K.: Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Res. **30**, 79–82 (2005)
38. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. CoRRabs/1511.06434 (2015)
39. Davis, C.: The norm of the Schur product operation. Numer. Math. **4**(1), 343–344 (1962). https://doi.org/10.1007/bf01386329
40. Joyce, J.M.: Kullback-Leibler divergence. In: Lovric, M. (ed.) International Encyclopedia of Statistical Science, pp. 720–722. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-04898-2_327

41. Phillips, P.J., Moon, H., Rizvi, S.A., Rauss, P.J.: The FERET evaluation methodology for face-recognition algorithms. In: NISTIR 6264, 7 January 1999 and IEEE Trans. Pattern Anal. Mach. Intell. **22**(10) (2000)
42. Gong, J., Mistele, M.: sketch2face: Conditional Generative Adversarial Networks for Transforming Face Sketches into Photorealistic Images (2018). https://web.stanford.edu/~jxgong/docs/sketch2face.pdf
43. Kingma, D., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)