BULLETIN OF THE POLISH ACADEMY OF SCIENCES
TECHNICAL SCIENCES, Vol. 72(4), 2024, Article number: e150109
DOI: 10.24425/bpasts.2024.150109

ARTIFICIAL AND COMPUTATIONAL INTELLIGENCE

High-quality synthesized face sketch using generative reference prior

Sami MAHFOUD[1,2]*, Messaoud BENGHERABI[2], Abdelhamid DAAMOUCHE[3], Elhocine BOUTELLAA[3], and Abdenour HADID[4]

[1] University of Algiers 3, Faculty of Economic, Commercial and Management Sciences, Laboratory of Governance and Modernization of Public Management, 02 Ahmed Ouaked Street Dely Ibrahim 16302, Algiers, Algeria
[2] Center for Development of Advanced Technologies, Telecom Division, P.O. Box 17 Baba-Hassen 16303, Algiers, Algeria
[3] SUniversity M'Hamed Bougara of Boumerdes, Institute of Electrical and Electronic Engineering, Laboratory of Signals and Systems, Boumerdes, 35000, Algeria
[4] Sorbonne University Abu Dhabi, Sorbonne Center for Artificial Intelligence, Abu Dhabi, UAE

Abstract. Face sketch synthesis (FSS) is considered an image-to-image translation problem, where a face sketch is generated from an input face photo. FSS plays a vital role in video/image surveillance-based law enforcement. In this paper, motivated by the recent success of generative adversarial networks (GAN), we consider conditional GAN (cGAN) to approach the problem of face sketch synthesis. However, despite the powerful cGAN model ability to generate fine textures, low-quality inputs characterized by the facial sketches drawn by artists cannot offer realistic and faithful details and have unknown degradation due to the drawing process, while high-quality references are inaccessible or even nonexistent. In this context, we propose an approach based on generative reference prior (GRP) to improve the synthesized face sketch perception. Our proposed model, which we call cGAN-GRP, leverages diverse and rich priors encapsulated in a pre-trained face GAN for generating high-quality facial sketch synthesis. Extensive experiments on publicly available face databases using facial sketch recognition rate and image quality assessment metrics as criteria demonstrate the effectiveness of our proposed model compared to several state-of-the-art methods.

Keywords: generative adversarial networks; face sketch synthesis; generative reference prior.

1. INTRODUCTION

In security and law enforcement, police agencies can quickly identify potential suspects by automatically retrieving the suspect's photos from the mugshot database [1]. Actually, the face photo of criminal suspects may not be available, and the face sketch of the likely suspect, drawn by expert artists based on the description of eyewitnesses is an alternative way to assist in face sketch matching applications [2]. In addition to security applications, face sketch synthesis (FSS) can be used in diverse digital entertainment applications. Face sketches are becoming more and more popular amongst social network users and smartphones, where face sketch is utilized as profile photos or avatars. However, face sketch synthesis and recognition may yield challenging problems due to the significant discrepancy in texture and structure between the facial photo and facial sketch.

FSS techniques have been classified into two groups, namely data-driven and model-driven strategies [3]. Data-driven strategies synthesize facial sketch patches using a linear combination of similar training photo-sketch pairs. Model-driven strategies are generally based on the trained model, which can directly synthesize facial sketches from facial photos after learning an offline mapping function between two heterogeneous modalities.

Data-driven models synthesize a facial sketch of a test photo by linearly combining candidate facial sketch patches chosen from the training facial photo-sketch pairs. For instance, Tang and Wang [4] proposed synthesizing facial sketches using the principal component analysis technique to get the coefficients used in the synthesis process. However, the linear hypothesis restricted the capability to characterize the non-linear aspect between facial photos and sketches. Liu *et al.* [5] introduced a method via locally linear embedding (LLE) to synthesize face sketches from face photos while the K-nearest neighbor (K-NN) is used to search similar neighbors at the image patch level. Zhang *et al.* [6] synthesized face sketches using similarity and prior knowledge between different facial image patches. Regarding the similarity between neighboring facial image patches, Wang and Tang [2] employed Markov random fields (MRF) in the FSS process and then utilized belief propagation to generate facial sketches. Zhou *et al.* [7] introduced a combination of K-neighbour patches into the MRF architecture named Markov weight fields (MWF), to resolve their deformation issue. Song *et al.* [8] proposed a fast synthesized method using spatial sketch denoising (SSD) problem.

Deep learning, on the other hand, has recently been gaining interest and has been extensively applied to related problems

*e-mail: mahfoud.sami@univ-alger3.dz

Manuscript submitted 2023-10-24, revised 2024-02-09, initially accepted for publication 2024-03-17, published in July 2024.

such as image style transfer, image super-resolution, image classification, and image fusion [9–12]. For the particular issue at hand, Zhang *et al.* [13] designed a model using a convolutional neural network (CNN) to learn the end-to-end facial photo-to-facial sketch mapping. More recently, inspired by their significant contributions to different image-to-image translation problems [14], GANs have proven to be the most performing among model-driven FSS methods. Different facial sketch synthesis approaches based on GANs have been proposed, demonstrating significant improvements. For instance, Zhu *et al.* [15] developed an unpaired GAN-based method named Cycle-GAN for transferring images between different modalities using unpaired data. However, the generated face sketches contain some noise. Wang *et al.* [16] introduced a back projection based on a data-driven technique to improve GAN-based FSS. Bi *et al.* [17] developed a multi-layer pyramid framework to extract the multi-scale data and train a multi-scale conditional GAN to learn the mapping function. Zhu *et al.* [18] employed knowledge transfer for learning the facial photo and sketch features separately by learning two models simultaneously. Zhang *et al.* [19] presented a dual-transfer facial photo-sketch synthesis model that synthesized more recognizable facial structures while preserving high matching accuracy.

Although the facial sketch generated by the conditional GAN-based method preserves fine texture, as shown in Fig. 1, noise and blur accompanying the fine texture usually appear in the generated sketches as a result of the pixel-to-pixel mapping. To enhance the perceptual quality of the generated facial sketches, we propose a model based on the generative reference prior, which is widely used in blind restoration and image super-resolution, to reduce the appearance of noise and blurring in the synthesized face sketch [20–22]. Our proposed cGAN-GRP model exploits extra and various prior information incorporated in a pre-trained facial GAN (such as StyleGAN2 [23]) to restore textures and structures as close to the real ones as possible, thus achieving a higher perceptual quality. Our contribution in this paper is threefold. First, we have adopted the conditional GAN for generating the facial sketch from the facial photo. Second, we have proposed improving the perceptual quality of synthesized facial sketches using the generative reference prior approach. Third, the proposed synthesis method strategy is benchmarked and compared favorably to the state-of-the-art synthesis methods, demonstrating a significant performance improvement.

Fig. 1. Facial sketches synthesized by Cycle-GAN, cGAN and the proposed cGAN-GRP method. **(Zoom in for the best view)**

2. PROPOSED METHOD

We describe below the overall architecture of our proposed cGAN-GRP approach along with a problem formulation, and details of each module of the proposed cGAN-GRP framework. Our proposed cGAN-GRP architecture combines two GAN models, namely the conditional GAN model for face sketch synthesis (cGAN) and generative reference prior (GRP), to enhance the perceptual quality of the synthesized face sketch. The overall cGAN-GRP framework is depicted in Fig. 2.

2.1. Conditional GAN model for the face sketch synthesis

We briefly describe the notations used to represent the cGAN for synthesized facial sketch. Depending on the M training facial sketch-photo pairs, the purpose is to produce the output $\mathbf{s}$, generated face sketch, from an observable facial image $\mathbf{t}$.

As shown in Fig. 2 (cGAN model part), conditional generative adversarial networks (cGAN) are a type of generative network that attempt to train a non-linear mapping function from the observable facial image $\mathbf{t}$ and a random noise vector $\mathbf{z}$ to create a sketch $\mathbf{x}$, $cGAN : \{\mathbf{t}, \mathbf{z}\} \mapsto \mathbf{x}$ instead of $\{\mathbf{z}\} \mapsto \mathbf{x}$ as generative adversarial networks (GANs) do.

The generator $\mathscr{G}$ attempts to produce $fake$ face sketches that are unable to be differentiated by the discriminator $\mathscr{D}$ against the $real$ sketches painted by professional artists. Simultaneously, the discriminator $\mathscr{D}$ attempts to differentiate the $fake$ of the generator $\mathscr{G}$ among the $real$, as depicted in Fig. 2. The conditional generative adversarial network objective [14] is expressed as follows:

$$\mathscr{G}^* = \arg\min_{\mathscr{G}} \max_{\mathscr{D}} \mathscr{L}_{\text{cGAN}}(\mathscr{G}, \mathscr{D}) + \lambda \mathscr{L}_{\text{L1}}(\mathscr{G}), \tag{1}$$

where $\mathscr{L}_{\text{cGAN}}$ is the cGAN loss, $\mathscr{L}_{\text{L}}(\mathscr{G})$ is the regularization loss, and λ is utilized to make a balance within the $\mathscr{L}_{\text{cGAN}}$ lose and the $\mathscr{L}_{\text{L1}}(\mathscr{G})$ loss. The $\mathscr{L}_{\text{cGAN}}$ has the following definition:

$$\begin{aligned} \mathscr{L}_{\text{cGAN}}(\mathscr{G}, \mathscr{D}) &= \mathbb{E}_{\mathbf{t},\mathbf{x}\sim P_{data}(\mathbf{t},\mathbf{x})}[\log \mathscr{D}(\mathbf{t}, \mathbf{x})] \\ &+ \mathbb{E}_{\mathbf{t}\sim P_{data}(\mathbf{t}),\mathbf{z}\sim P_z(\mathbf{z})}[\log(1 - \mathscr{D}(\mathbf{t}, \mathscr{G}(\mathbf{t}, \mathbf{z})))], \end{aligned} \tag{2}$$

and the $\mathscr{L}_{\text{L1}}(\mathscr{G})$ is expressed as follows:

$$\mathscr{L}_{\text{L1}}(\mathscr{G}) = \mathbb{E}_{\mathbf{t},\mathbf{x}\sim P_{data}(\mathbf{t},\mathbf{x}),\mathbf{z}\sim P_z(\mathbf{z})}[\|(\mathbf{x} - \mathscr{G}(\mathbf{t}, \mathbf{z}))\|], \tag{3}$$

The discriminator $\mathscr{D}$ in the cGAN model is exactly a classifier. Its aim is to learn to differentiate the real data from the fake data generated by $\mathscr{G}$. Any network appropriate to the facial image data could be used. For the discriminator, in our case, the PatchGAN [24] is utilized as the convolutional classifier. We adopt an architecture based on U-Net [25] as the generator $\mathscr{G}$.

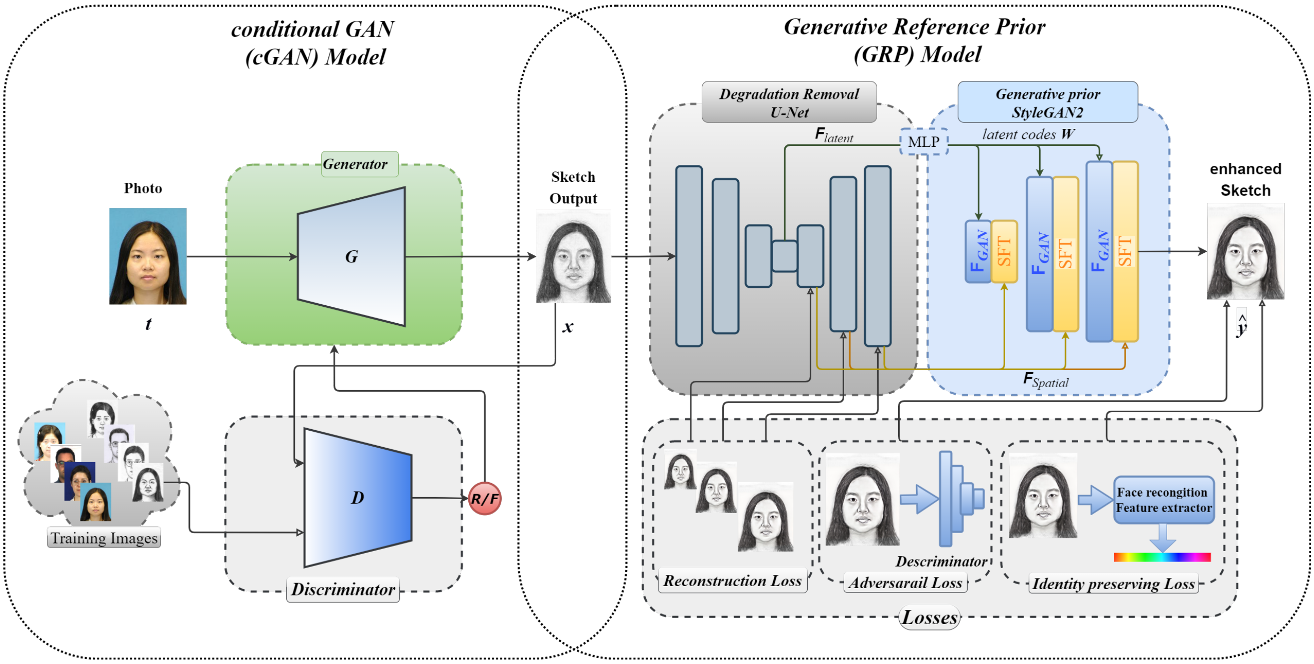

Fig. 2. Overview of cGAN-GRP framework. The framework combines two models, conditional GAN (cGAN) and generative reference prior (GRP). The GRP model composes of a blur and noise removal module (such as U-Net) and a pre-trained face GAN (such as StyleGAN2) as a generative prior module

2.2. Generative reference prior model

As shown in Fig. 1 and Fig. 6, noise and blur attached to fine texture appear in the generated sketches by cGAN due to the pixel-to-pixel mapping on the one hand and the low-quality facial sketches drawn by artists and the unknown degradation caused by the drawing process on the other hand. To solve this issue, we propose a new framework based on the generative reference prior, which significantly reduces the noise and blurring in the synthesized face sketch. Our proposed GRP model takes advantage of extra and various prior information incorporated in a pre-trained facial StyleGAN2 to reconstruct realistic textures and structures, aiming to achieve a higher perceptual quality.

Given an input face photo (here, the synthesized facial sketch $\mathbf{x}$ by the cGAN model), the generative reference prior aims to assess a high-perceptual quality of the facial photo (here, the enhanced synthesized facial sketch $\hat{\mathbf{y}}$) that resembles the high-quality ground truth photo $\hat{\mathbf{y}}$ as closely as possible in terms of texture and structure, and realness.

Our proposed GRP is composed of two sub-models: the blur and noise removal module (such as U-Net [25]) and the generative prior module (such as StyleGAN2 [23]). These two sub-modules are mapped by a code and several intermediate layers of spatial feature transform (SFT) [26]. More details of these components are provided hereafter.

Blur and noise removal module is implemented as a U-Net architecture aiming to eliminate complicated degradation and obtain two features: multi-spatial F_{Spat} features and the latent F_{lat} features. The U-Net formulation is as follows:

$$F_{lat}, F_{Spat} = U\text{-}Net(x), \tag{4}$$

F_{lat} maps the test photo into the latent code in the pre-trained StyleGAN2, whereas F_{Spat} modulates StyleGAN2 features.

Generative prior module a pre-trained GAN model, such as StyleGAN2, encapsulates a learned distribution of facial features within its convolutional weights, known as the generative prior [27, 28]. We take advantage of these pre-trained facial GANs to generate diverse and rich facial details. One common method for exploiting generative priors involves mapping the input image to its most similar latent codes, denoted as Z, and subsequently producing the corresponding output using a pre-trained GAN [27–29]. Alternatively, intermediate convolutional features (F_{GAN}) of the closest face can be generated, as they offer more detail and can be adjusted by input features to enhance fidelity [22]. In our task, using the latter method, given the latent features F_{lat} of the input photo (result of the U-Net, Eq. (4)), the first step is to map it to intermediate latent codes W, i.e., the intermediate space transformed from Z with several Multi-Layer Perceptron (MLP) [29]. Subsequently, the latent codes W generate GAN features for every resolution scale by passing through every convolution layer in the pre-trained StyleGAN2. the formulas are as follows:

$$\begin{gathered} W = MLP(F_{lat}), \\ F_{GAN} = StyleGAN(W). \end{gathered} \tag{5}$$

Multi-spatial features F_{Spat} are utilized to adjust the pre-trained facial GAN features F_{GAN} to perform realistic results and faithful details while maintaining high fidelity. Specifically, we generate two transformation parameters (α, β) from the input F_{Spat} by several intermediate convolutional layers at each resolution scale. Hereafter, the vector F_{GAN} is scaled and shifted to produce the modulation (α, β). The formulas are expressed as follows:

$$\begin{aligned} \alpha, \beta &= Convolu(F_{Spat}), \\ F_{output} &= SFT(F_{GAN} \mid \alpha, \beta) = \alpha \odot F_{GAN} + \beta. \end{aligned} \tag{6}$$

Consequently, SFT has the advantages of directly incorporating prior information and effective modulation of input images, performing the best balance between texture faithfulness and fidelity to produce the final sketch $\hat{y}$.

2.3. Objective functions

The objective functions for training our GRP consist of i) reconstruction loss, ii) adversarial loss, and iii) identity-preserving loss.

2.3.1. Reconstruction loss

To restrict the output sketch $\hat{y}$ to be as similar as possible to the ground-truth facial photo y, we use the perceptual loss [30] and the most utilized L_1 loss as the reconstruction loss function $\mathscr{L}_{reco}$, expressed below:

$$\mathscr{L}_{reco} = \lambda_{L_1} \|\hat{y} - y\|_1 + \lambda_{perc} \|\psi(\hat{y}) - \psi(y)\|_1 , \tag{7}$$

where ψ denotes the pre-trained VGG-19 model [31]. λ_{L_1} and λ_{perc} are loss weights of L_1 and $\mathscr{L}_{reco}$, respectively.

2.3.2. Adversarial loss

We use this loss function $\mathscr{L}_{adve}$ to boost our GRP to generate realistic textures and structure. As in StyleGAN2 [23], we adopt the logistic loss [32], described as follows:

$$\mathscr{L}_{reco} = -\lambda_{adve} \mathbb{E}_{\hat{y}}\ softplus(\mathscr{D}(\hat{y})), \tag{8}$$

where the loss weight is denoted by λ and the discriminator is represented by $\mathscr{D}$.

2.3.3. Identity-preserving loss

We consider this loss function in the GRP architecture to encourage preserving identity. Similar to perceptual loss [30], we utilize the pre-trained ArcFace [33] network for its ability to capture the principal features required for discrimination. The identity-preserving loss $\mathscr{L}_{ide}$ is described as follows:

$$\mathscr{L}_{ide} = \lambda_{ide} \|\upsilon(\hat{y}) - \upsilon(y)\|_1 , \tag{9}$$

where υ symbolises the pre-trained ArcFace [33] and λ_{ide} indicates the loss weight.

2.3.4. Overall objective loss

The overall objective of GRP model is a sum of the aforementioned losses:

$$\mathscr{L}_{total} = \mathscr{L}_{reco} + \mathscr{L}_{adve} + \mathscr{L}_{ide} . \tag{10}$$

The setting of loss weights are as follows: $\lambda_{l1} = 0.1$, $\lambda_{perc} = 1$, $\lambda_{adve} = 0.1$ and $\lambda_{ide} = 10$.

3. EXPERIMENTAL RESULTS AND ANALYSIS

3.1. Settings

3.1.1. Dataset

Our proposed method has been validated on the public University of Hong Kong (CUHK) Face Sketch Database (CUFS) [2]. In CUFS Database, there are 606 facial images and their associated facial sketches penciled by expert artists. These facial photos are collected from three datasets, including the CUHK student database [4], the XM2VTS database [34], and the AR database [35], which yields 188, 295, and 123 face photo-sketch pairs, respectively. Figure 3 depicts some facial photos and sketches from this database. Following the standard settings [16, 36, 37], we utilize 268 photos and their sketches to train the model and the rest (338 samples) for testing.

Fig. 3. Some facial image examples and their sketches from the CUFS database. These three photo-sketch pairs are from the XM2VTS, CUHK student, and AR datasets, respectively

3.1.2. Implementation details

The cGAN-GRP architecture combines two networks, the cGAN network (Fig. 2 part cGAN model) for synthesized facial sketch and the generative reference prior model (Fig. 2 part GRP model). For the cGAN model, we have adopted the source codes from here[1] to synthesize the face sketch in our scenario keeping the same experimental parameters and settings, and then we trained the model on the CUFS database. For the GRP model, we have used the pre-trained model from here[2], where the model has been trained on the FFHQ dataset [38].

3.1.3. Criteria

We considered three widely-used performance indices as criteria for quantitative evaluation, namely, the "Learned Perceptual Image Patch Similarity (LPIPS)" [39] metric, the "Structure Co-occurrence Texture (Scoot)" [40] metric, and the face sketch matching rate by employing the "Null-space Linear Discriminant Analysis (NLDA)" [41]. The two former criteria, LPIPS and Scoot, are full-reference image quality assessment metrics to objectively assess the facial sketch synthesis quality. The facial sketches made by the professional artist are taken as reference photos, and even the reference sketches have been enhanced using the GPR model and retaken as a reference for our approach to restoring unknown degradation due to the drawing process, as shown in Fig. 4. The generated sketches are used as distorted photos. Lower LPIPS scores demonstrate that the approach synthesizes better facial sketches in terms of perceptual quality.

[1] https://github.com/phillipi/pix2pix.

[2] https://github.com/TencentARC/GFPGAN.

Higher Scoot values generally demonstrate the best similarities between a generated face sketch and its related target facial sketch.

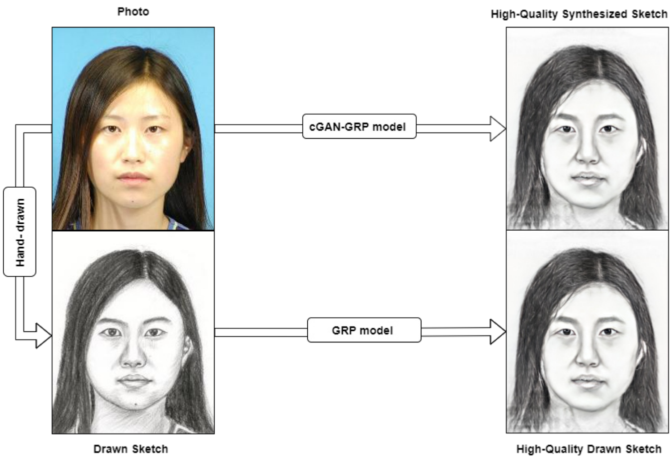

Fig. 4. High-quality synthesized sketch and the improved hand-drawn face sketch using the proposed cGAN-GRP and GRP models, respectively. The photo and its corresponding hand-drawn sketch are from the CUFS database

The latter criterion, face sketch matching rate, is an alternate way to assess the capability of synthesized facial sketch methods quantitatively [36,42,43]. Such a higher sketch recognition rate (NLDA rate) indicates better-synthesized facial sketch quality and a more effective sketch generation approach [44,45] without losing its identity. To accomplish the face sketch recognition experiments using the NLDA method, we randomly selected 150 synthesized facial sketches and their associated facial sketches painted by the professional artists for training. The remaining 188 synthesized facial sketches are used as the gallery test. By randomly partitioning the data, we repeat the process 20 times.

3.1.4. Comparison with the state-of-the arts

For evaluating the effectiveness of the proposed strategy, we considered various state of the art methods for comparison, including LLE [5], SSD [8], MRF [2], MWF [7], Cycle-GAN [15], BP-GAN [16] and GAN [14]. The LLE, MWF, MRF, and SSD techniques are considered data-driven approaches, while the BP-GAN, cGAN, Cycle-GAN, and our proposed cGAN-GRP are model-driven.

3.2. Results and discussion

3.2.1. Quantitative evaluation

Figure 5 provides the statistics of LPIPS and Scoot values on the CUFS database as boxplots, and Table 1 shows the quantitative evaluation metrics for each approach used for comparison. The best results are highlighted in bold font. The results from Table 1 and Fig. 5 demonstrate that our proposed cGAN-GRP attains the best values of LPIPS and Scoot while preserving the higher face sketch matching accuracy. Specifically, cGAN-GRP significantly increases the previous state-of-the-art Scoot by a considerable margin, 54.33%, achieved by cGAN method, to

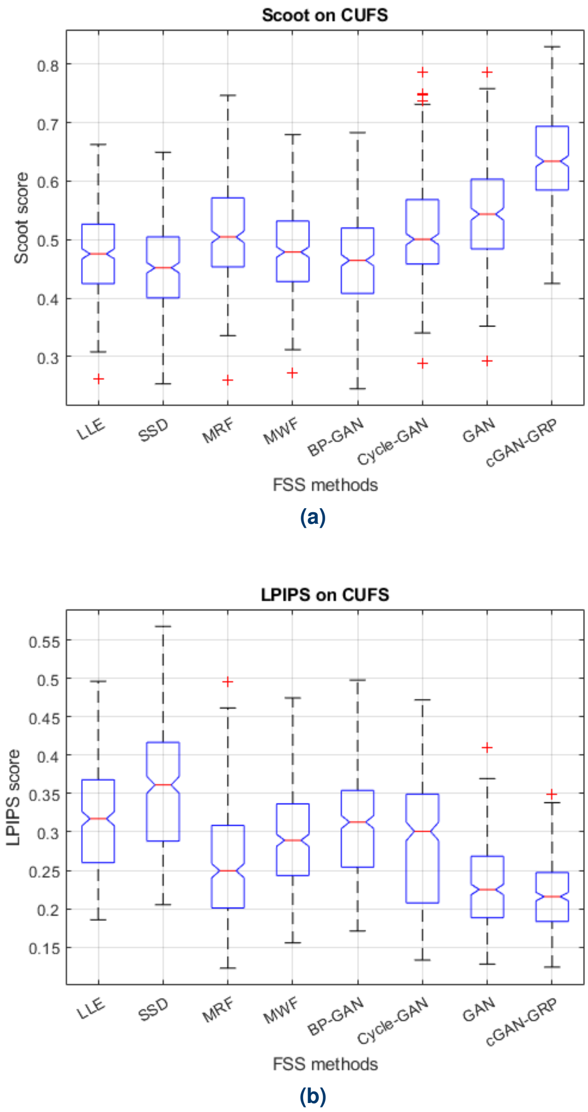

Fig. 5. Scoot and LPIPS values of the different compared face sketch synthesis methods on the CUFS database, represented as box-plot curves, respectively

63.39%. Besides, cGAN-GRP reduces the prior best state-of-the-art LPIPS from 22.49% to 21.57%. Such low LPIPS scores mean that the synthesized sketch by cGAN-GRP is more natural in terms of perceptual quality and appearance. In addition, the higher Scoot scores mean that the synthesized facial sketch by cGAN-GRP is more similar to those sketched by expert artists regarding textures and structures. In terms of face sketch matching, cGAN-GRP increases the accuracy from 95.53% to 95.66% as shown in Table 1.

According to these criteria, our cGAN-GRP demonstrates significant superiority over existing approaches and attains the best performance. Thus, our cGAN-GRP improves the perceptual quality of synthesized sketches while preserving high fidelity.

3.2.2. Qualitative evaluation

Figure 6 shows some synthesized facial sketches by the different techniques on the CUFS database. First, it can be illustrated from

Table 1
The quantitative evaluation metrics for each compared approach on the CUFS database. The best results are in bold. ↓ means that the lower is the best, while ↑ means that the higher is the best

FSS methods	Criteria	Scoot (%) ↑	LPIPS(%) ↓	NLDA(%) ↑
Data-driven methods	LLE	47.55	31.73	92.31
	SSD	45.16	36.13	92.36
	MRF	50.44	24.94	87.5
	MWF	47.85	28.91	93.35
Model-driven (GAN-based) methods	BP-GAN	46.44	31.29	95.1
	Cycle-GAN	50.04	30.07	86.06
	cGAN	54.33	22.49	95.53
	cGAN-GRP	**63.39**	**21.57**	**95.66**

Fig. 6 that model-driven (BP-GAN, cGAN, Cycle-GAN, and our proposed cGAN-GRP) strategies generate finer details better than data-driven strategies, which include LLE, MWF, MRF, and SSD techniques. These outcomes explain the advantage of an adversarial process like the GAN model used for image-to-image translation problems. The data-driven approaches fail to provide fine details due to the existing nonlinear aspects between facial photos and facial sketches penciled by professional artists, including shape exaggeration, lighting variations, and different races. This justifies the blur and noise appearing on the generated face sketches using data-driven methods. Second, although the GAN-based methods, such as BP-GAN, cGAN, and Cycle-GAN, produce fine texture details, there is still some noise present among the facial parts, and the synthesized outcomes do not exhibit a net appearance, as depicted in Fig. 6. In contrast, the synthesized facial sketches using the proposed cGAN-GRP framework demonstrate a more realistic appearance and faithful details, as well as less noise. This is particularly more perceptible around the mouth, hair, and eye regions. Such superiority confirms the effectiveness of our proposed cGAN-GRP technique.

Our both quantitative and qualitative evaluations confirm that our cGAN-GRP outperforms previous SOTA methods and is capable of synthesizing high-quality face sketches within preserving high fidelity and identity.

4. CONCLUSIONS

In this work, we adopted the conditional GAN for facial sketch synthesis. Then, we proposed enhancing the perceptual quality of synthesized facial sketches using the generative reference prior strategy. Extensive experiment showed that despite the GAN-based approaches effectively maintaining texture and structure details, they also generate some noise and blur around the facial part, and some generated outputs are unclear. Our proposed cGAN-GRP model significantly reduces the noise, enhances the quality of the facial sketches, and provides more faithful and realistic details. Both quantitative and qualitative evaluations demonstrated that our proposed cGAN-GRP outperforms the current FSS techniques. In the proposed framework, the GRF and cGAN models have been separated. In the future, we plan to integrate the GRP into the cGAN architecture and generalize the proposed framework to other heterogeneous face recognition applications.

Input Photo LLE SSD MRF MWF Cycle-GAN BP-GAN cGAN cGAN-GRP

Fig. 6. Synthesized facial sketch examples by the different compared methods on the CUFS database. The input facial photos are in the first column, and from the second to the last are their corresponding synthesized sketches by LLE, SSD, MRF, MWF, Cycle-GAN, BP-GAN, cGAN, and the proposed cGAN-GRP methods, respectively. These facial photos are selected from the CUHK Student, AR, and XM2VTS datasets, respectively

DECLARATION OF COMPETING INTERESTS

The authors certify that they have no known financial or interpersonal conflicts that could have been expected to have an impact on the research presented in this study.

ACKNOWLEDGEMENTS

This work is supported by the DGRSDT (General Directorate of Scientific Research and Technological Development), Algeria, the CDTA (Center for Development of Advanced Technologies), Algeria, and the GMPM (Laboratory of Governance and Modernization of Public Management), University of Algiers 3, Algeria.

REFERENCES

[1] H. Bi, Z. Liu, L. Yang, K. Wang, and N. Li, "Face sketch synthesis: a survey," *Multimed. Tools Appl.*, vol. 80, no. 12, pp. 18 007–18 026, Feb 2021, doi: 10.1007/s11042-020-10301-0.

[2] X. Wang and X. Tang, "Face photo-sketch synthesis and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 11, pp. 1955–1967, Nov 2009, doi: 10.1109/TPAMI.2008.222.

[3] P. Li, B. Sheng, and C.L. P. Chen, "Face sketch synthesis using regularized broad learning system," *IEEE Trans. Neural Netw. Learn. Syst.*, pp. 1–15, 2021, doi: 10.1109/TNNLS.2021.3070463.

[4] X. Tang and X. Wang, "Face photo recognition using sketch," in *Proceedings. International Conference on Image Processing*, vol. 1. IEEE, 2002, pp. I–I, doi: 10.1109/ICIP.2002.1038008.

[5] Q. Liu, X. Tang, H. Jin, H. Lu, and S. Ma, "A nonlinear approach for face sketch synthesis and recognition," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1. IEEE, 2005, pp. 1005–1010, doi: 10.1109/CVPR.2005.39.

[6] S. Zhang, X. Gao, N. Wang, J. Li, and M. Zhang, "Face sketch synthesis via sparse representation-based greedy search," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2466–2477, 2015, doi: 10.1109/tip.2015.2422578.

[7] H. Zhou, Z. Kuang, and K. K. Wong, "Markov weight fields for face sketch synthesis," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, Jun 2012, pp. 1091–1097, doi: 10.1109/cvpr.2012.6247788.

[8] Y. Song, L. Bao, Q. Yang, and M.-H. Yang, "Real-time exemplar-based face sketch synthesis," in *Computer Vision – ECCV 2014*. Springer International Publishing, 2014, pp. 800–813, doi: 10.1007/978-3-319-10599-451.

[9] A. Cichocki, T. Poggio, S. Osowski, and V. Lempitsky, "Deep learning: Theory and practice," *Bull. Pol. Acad. Sci. Tech. Sci.*, vol. 66, no. No 6 (Special Section on Deep Learning: Theory and Practice), pp. 757–759, 2018, doi: 10.24425/bpas. 2018.125923.

[10] X. Yang, Y. Zhang, and D. Zhou, "Deep networks for image super-resolution using hierarchical features," *Bull. Pol. Acad. Sci. Tech. Sci.*, vol. 70, no. 1, p. e139616, 2022, doi: 10.24425/bpasts.2021.139616.

[11] K. Hawari and I. Ismail, "The automatic focus segmentation of multi-focus image fusion," *Bull. Pol. Acad. Sci. Tech. Sci.*, vol. 70, no. 1, p. e140352, 2022, doi: 10.24425/bpasts.2022. 140352.

[12] M. Grochowski, A. Kwasigroch, and A. Mikołajczyk, "Selected technical issues of deep neural networks for image classification purposes," *Bull. Pol. Acad. Sci. Tech. Sci.*, vol. 67, no. 2, pp. 363–376, 2019, doi: 10.24425/bpas.2019.128485.

[13] 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*. ACM, Jun 2015, pp. 627–634, doi: 10.1145/2671188.2749321.

[14] 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)*. IEEE, Jul 2017, pp. 1125–1134, doi: 10.1109/cvpr.2017.632.

[15] J.-Y. Zhu, T. Park, P. Isola, and A.A. Efros, "Unpaired image- to-image translation using cycle-consistent adversarial networks," in *2017 IEEE International Conference on Computer Vision (ICCV)*. IEEE, Oct 2017, pp. 2223–2232, doi: 10.1109/iccv.2017.244.

[16] N. Wang, W. Zha, J. Li, and X. Gao, "Back projection: An effective postprocessing method for gan-based face sketch synthesis," *Pattern Recognit. Lett.*, vol. 107, pp. 59–65, 2018, doi: 10.1016/j.patrec.2017.06.012.

[17] 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)*. IEEE, 2019, pp. 3876–3880, doi: 10.1109/icip. 2019.8803629.

[18] M. Zhu, N. Wang, X. Gao, J. Li, and Z. Li, "Face photo-sketch synthesis via knowledge transfer." in *IJCAI*, 2019, pp. 1048–1054, doi: 10.24963/ijcai.2019/147.

[19] M. Zhang, R. Wang, X. Gao, J. Li, and D. Tao, "Dual-transfer face sketch-photo synthesis," *IEEE Trans. Image Process.*, vol. 28, no. 2, pp. 642–657, 2019, doi: 10.1109/tip. 2018.2869688.

[20] X. Deng, H. Zhang, and X. Li, "Hpg-gan: High-quality prior-guided blind face restoration generative adversarial network," *Electronics*, vol. 12, no. 16, p. 3418, 2023, doi: 10.3390/electronics12163418.

[21] X. Wang, Y. Li, H. Zhang, and Y. Shan, "Towards real-world blind face restoration with generative facial prior," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 9168–9178, doi: 10.1109/cvpr46437.2021.00905.

[22] X. Guo, L. Yi, H. Zou, and Y. Gao, "Generative facial prior for large-factor blind face super-resolution," in *J. Phys.-Conf. Ser.*, vol. 2078, no. 1. Publishing, 2021, p. 012045, doi: 10.1088/1742-6596/2078/1/012045.

[23] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of stylegan," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 8110–8119, doi: 10.1109/cvpr42600.2020.00813.

[24] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015, doi: 10.48550/arXiv.1511.06434.

[25] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-428.

[26] X. Wang, K. Yu, C. Dong, and C. C. Loy, "Recovering realistic texture in image super-resolution by deep spatial feature transform," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Jun 2018, pp. 606–615, doi: 10.1109/cvpr.2018.00070.

[27] J. Gu, Y. Shen, and B. Zhou, "Image processing using multi-code gan prior," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Jun. 2020, pp. 3012–3021, doi: 10.1109/cvpr42600.2020.00308.

[28] X. Pan, X. Zhan, B. Dai, D. Lin, C. C. Loy, and P. Luo, "Exploiting deep generative prior for versatile image restoration and manipulation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 11, pp. 7474–7489, Nov. 2022, doi: 10.1109/ tpami.2021.3115428.

[29] J. Zhu, Y. Shen, D. Zhao, and B. Zhou, *In-Domain GAN Inversion for Real Image Editing*. Springer International Publishing, 2020, pp. 592–608, doi: 10.1007/978-3-030-58520-4_35.

[30] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Computer Vision – ECCV 2016*, Springer. Springer International Publishing, 2016, pp. 694–711, doi: 10.1007/978-3-319-46475-6-43.

[31] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014, doi: 10.48550/arXiv.1409.1556.

[32] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020, doi: 10.1145/3422622.

[33] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Jun 2019, pp. 4690–4699, doi: 10.1109/ cvpr.2019.00482.

[34] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre, "Xm2vtsdb: The extended m2vts database," in *Second International Conference on Audio and Video-Based Biometric Person Authentication*, vol. 964, 1999, pp. 965–966, Database: XM2VTSDB.

[35] A.M. Martinez, "The ar face database," CVC Technical Report24, Tech. Rep., 1998, AR Face database.

[36] W. Wan, Y. Yang, and H. J. Lee, "Generative adversarial learning for detail-preserving face sketch synthesis," *Neurocomputing*, vol. 438, pp. 107–121, May 2021, doi: 10.1016/j.neucom. 2021.01.050.

[37] N. Wang, X. Gao, L. Sun, and J. Li, "Bayesian face sketch synthesis," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1264–1274, Mar 2017, doi: 10.1109/TIP.2017.2651375.

[38] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *2019 IEEE/ CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Jun 2019, pp. 4401–4410, doi: 10.1109/cvpr. 2019.00453.

[39] R. Zhang, P. Isola, A.A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Jun 2018, pp. 586–595, doi: 10.1109/cvpr.2018.00068.

[40] D.-P. Fan, S. Zhang, Y.-H. Wu, Y. Liu, M.-M. Cheng, B. Ren, P. Rosin, and R. Ji, "Scoot: A perceptual metric for facial sketches," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, Oct 2019, pp. 5612–5622, doi: 10.1109/iccv.2019.00571.

[41] L.-F. Chen, H.-Y. M. Liao, M.-T. Ko, J.-C. Lin, and G.-J. Yu, "A new LDA-based face recognition system which can solve the small sample size problem," *Pattern Recognit.*, vol. 33, no. 10, pp. 1713–1726, Oct 2000, doi: 10.1016/s0031-3203(99) 00139-9.

[42] X. Li, F. Gao, and F. Huang, "High-quality face sketch synthesis via geometric normalization and regularization," in *2021 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, Jul 2021, pp. 1–6, doi: 10.1109/ICME51207. 2021.9428348.

[43] N. Wang, X. Gao, and J. Li, "Random sampling for fast face sketch synthesis," *Pattern Recognit.*, vol. 76, pp. 215–227, Apr 2018, doi: 10.1016/j.patcog.2017.11.008.

[44] N. Balayesu and H. K. Kalluri, "An extensive survey on traditional and deep learning-based face sketch synthesis models," *Int. J. Inf. Technol.*, vol. 12, no. 3, pp. 995–1004, Nov 2020, doi: 10.1007/s41870-019-00386-8.

[45] S. Mahfoud, A. Daamouche, M. Bengherabi, and A. Hadid, "Hand-drawn face sketch recognition using rank-level fusion of image quality assessment metrics," *Bull. Pol. Acad. Sci. Tech. Sci.*, vol. 70, no. 6, 2022, doi: 10.24425/bpasts.2022.143554.