Neurocomputing 438 (2021) 107–121

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Generative adversarial learning for detail-preserving face sketch synthesis

Weiguo Wan [a], Yong Yang [b], Hyo Jong Lee [c,*]

[a] School of Software and Internet of Things Engineering, Jiangxi University of Finance and Economics, Nanchang 330032, China
[b] School of Information Technology, Jiangxi University of Finance and Economics, Nanchang 330032, China
[c] Division of Computer Sciences and Engineering, CAIIT, Jeonbuk National University, Jeonju 54896, South Korea

ARTICLE INFO

Article history:
Received 19 August 2020
Revised 28 November 2020
Accepted 8 January 2021
Available online 18 January 2021
Communicated by Zidong Wang

Keywords:
Face sketch synthesis
Detail-preserving
Generative adversarial learning
High-resolution network
Face sketch recognition

ABSTRACT

Face sketch synthesis aims to generate a face sketch image from a corresponding photo image and has wide applications in law enforcement and digital entertainment. Despite the remarkable achievements that have been made in face sketch synthesis, most existing works pay main attention to the facial content transfer, at the expense of facial detail information. In this paper, we present a new generative adversarial learning framework to focus on detail preservation for realistic face sketch synthesis. Specifically, the high-resolution network is modified as generator to transform a face image from photograph to sketch domain. Except for the common adversarial loss, we design a detail loss to force the synthesized face sketch images have proximate details to its corresponding photo images. In addition, the style loss is adopted to restrain the synthesized face sketch images have vivid sketch style as the hand-drawn sketch images. Experimental results demonstrate that the proposed approach achieves superior performance, compared to state-of-the-art approaches, both on visual perception and objective evaluation. Specifically, this study indicated the higher FSIM values (0.7345 and 0.7080) and Scoot values (0.5317 and 0.5091) than most comparison methods on the CUFS and CUFSF datasets, respectively.

1. Introduction

In recent years, face sketch synthesis has attracted significant attention in the computer vision and pattern recognition area, due to its important applications in law enforcement agencies and digital entertainment [1]. In many criminal cases, only limited information about the suspects is available, because of the low quality or lack of surveillance videos. In these situations, the sketches drawn by artists according to the reminiscence of eyewitnesses are usually taken as the substitute for identifying the suspects. For example, Fig. 1 shows real instances of forensic face sketches released by the FBI. Once the police obtain the sketches, they can narrow down the list of possible suspects by retrieving the face datasets in law enforcement departments or surveillance camera footage with the sketches [2]. Moreover, the sketch-style face images are also applied in animation production and used as avatars on social media[3].

For face sketch to photo recognition, due to the large modality discrepancy between face sketches and digital photos, it is more difficult to directly match face sketch images to photo images compared with homogeneous face recognition [4,5]. Various approaches have been put forward to solve the modality gap problem in face sketch recognition, such as modality-invariant feature extraction [6], common subspace projection [7], and photo to sketch synthesis [8–12]. Among them, face sketch synthesis is the most commonly used method, which transforms the face modality from photo to sketch by utilizing photo–sketch image pairs as training data, thus the face modality discrepancy is reduced. After generating face sketches from digital photos, the face sketch recognition can be performed with conventional homogeneous face recognition methods.

Existing face sketch synthesis approaches can be roughly classified into two categories: traditional exemplar-based approaches, and deep learning-based approaches. Exemplar-based approaches search several nearest photo patches from the training photos for each patch in the test photo. Then, the reconstruction weights can be calculated by approximating the nearest photo patches to the test photo patch, which are then utilized to construct the target sketch patch [13]. The exemplar-based methods are easy to understand and implement. But the searching of nearest photo patches is time-consuming, and the synthesized sketch images suffer from block artifacts. Recently, deep learning-based methods have

* Corresponding author.
E-mail address: hlee@chonbuk.ac.kr (H.J. Lee).

https://doi.org/10.1016/j.neucom.2021.01.050
0925-2312/

Fig. 1. Examples of face sketch for real law enforcement application.

become new trends in the face sketch synthesis field. These methods aim at training a deep model that is able to generate face sketches from face photos, without further training photo–sketch pairs during testing. The transformation process of deep learning-based approaches is fast; however, the synthesized sketch images usually lack facial details, and contain serious noise effects [14].

In order to address the aforementioned challenges, we present a face photo-to-sketch synthesis approach based on generative adversarial learning framework. First, we modify the high-resolution network [15], which has the ability to maintain high-resolution representations through the whole process, to generate realistic face sketch images. Then, a detail loss function, which is calculated between the Laplacian of Gaussian (LoG) filter of the input photo image and the predicted sketch image, is proposed, to restrain the generator to produce face sketch images with more detailed information. In addition, the pseudo sketch feature is utilized to calculate style loss to make the synthesized face sketch image having a similar style to the training sketch images. Experimental results indicate that the proposed face sketch synthesis method achieves superior performance, both on visual perceptual and objective evaluation.

1.1. Contributions

The main contributions of our work in this paper are summarized as follows:

1) We construct a new generative adversarial learning framework for face sketch synthesis. Our method is capable of generating unabridged facial content and preserving face details of the input photo image.
2) We modify the high-resolution network to merge the high-level feature maps gradually, and utilize it as generator to synthesize realistic face sketch images. The gradual merging strategy has been experimentally proved can preserve more detail information from source photos.
3) A LoG-based detail loss is designed to cope with the problem of lack of facial details in synthesized sketches. To the best of our knowledge, it is the first time to employ this idea to constraint the detail information in image transformation tasks.
4) The effectiveness of the proposed method is qualitatively and quantitatively validated on multiple public datasets. Compared with the state-of-the-art approaches, our results have abundant facial details and vivid sketch style.

The remaining parts of this paper are organized as follows: Section 2 introduces the related works, while Section 3 introduces our face sketch synthesis framework in detail. Section 4 describes the experiments on multiple face sketch datasets and evaluates the performance of the proposed face sketch synthesis approach. Section 5 concludes the paper and recommends future works.

2. Related works

In this section, we survey the representative works related to face sketch synthesis, and introduce the Generative Adversarial Network (GAN) briefly.

2.1. Face sketch synthesis

Traditional exemplar-based face photo-to-sketch synthesis approaches generate a face sketch image by linear combination of training sketch patches. The exemplar-based methods mainly consist of neighbor patch selection in the training photo–sketch image pairs, and linear combinations with reconstruction weights in two steps. In the first step, for each test photo patch, several closest training photo patches are chosen from contiguous positions. In the second step, a weight vector is computed between the patch of test photo and the chosen patches of training photos. After that, the image patch of target sketch can be achieved through weighted averaging of the corresponding training sketch patches to the chosen training photo patches. The exemplar-based face sketch synthesis approach was firstly put forward by Tang et al. [16], in which all of the training photo–sketch pairs were utilized to generate a target sketch. Liu et al. [17] suggested to perform face photo to sketch generation on the image patch level, where the reconstruction coefficients are computed with locally linear embedding. Gao et al. [18] proposed a face sketch synthesis method based on an embedded hidden Markov model. To consider the dependency relationship between neighboring sketches, Wang et al. [19] employed the Markov random fields (MRF) algorithm to represent the neighbor constraints among adjacent sketch patches. It chose only the best sketch patch from the training data for each test photo patch. Later, Zhou et al. [20] utilized a Markov weight fields (MWF) model to select multiple nearest patches to generate the sketch patch. To speed up the sketch generation process, Song et al. [21] transformed face sketch generation task to a spatial sketch denoising (SSD) issue, and obtained real-time performance on GPU. Peng et al. [22] presented a face photo-sketch synthesis approach based on superpixel seg-

mentation. Zhang et al. [23] developed a robust face sketch generation approach that can produce arbitrarily stylistic sketches with only a single sketch example. Wang et al. [24] conducted the neighbor sketch selection and reconstruction weight calculation with a Bayesian framework. Wang et al. [13] also proposed a random sampling and locality constraint (RSLCR) based synthesis method, in which the training photo and sketch patches are randomly sampled, and the locality constraint was utilized to calculate the reconstruction weight coefficients. Recently, Zhang et al. [25] designed a dual-transfer framework to preserve the identity-specific information with inter- and intra-domain transfer processes. While the exemplar-based methods can synthesize facial components well, most of these face sketch synthesis methods are usually time-consuming, and suffer from blocking effects around image boundaries.

Deep learning-based methods aim at learning a deep network model that has the ability to rapidly generate a face sketch image from the face photo image in the testing phase. Zhang et al. [26] firstly put forward a deep learning-based face sketch generation approach with fully convolutional networks (FCN). Reference [11] modified the convolutional neural network (CNN) with a multi-layer perceptron convolutional layer to generate sketch images. Sheng et al. [27] proposed another CNN-based method that utilized the enhanced cross-layer cost aggregation and 3D PatchMatch to extract the feature maps to generate face sketch images. Instead of generating the sketch image directly, Jiang et al. [12] put forward the learning of the residual map between the face photo and its corresponding sketch image. For the purpose of improving the quality of the synthesized sketches, some researches proposed to train deep networks with the assistance of facial components [28–30]. With the great success of GAN in image-to-image transformation, the researches have explored the employment of GAN in face sketch synthesis, and achieved impressive performance. While the deep learning-based face sketch generation approaches have an aptitude for identity-preserving, facial structures and details are usually missed during the synthesis phase.

To cope with the weaknesses of the traditional exemplar- and deep learning-based face sketch synthesis methods. The proposed method put forward a novel face sketch synthesis framework which takes the texture details and sketch style into consideration, and the synthesized face sketches not only can preserve the facial details of the input face photos, but also have vivid style as the real face sketches in training dataset.

2.2. Generative adversarial networks

GAN was first devised by Goodfellow et al. [31] to generate realistic images by learning the distribution of the training images with a game theoretic min–max optimization framework. It has made great achievements in numerous computer vision applications, such as image super-resolution [32], image style transfer [33], image restoration [34], and texture synthesis [35]. Two neural networks are trained alternately in GAN: a generative model G that generates new samples that are similar to the training data, and a discriminative model D that distinguish samples in disguise. They are trained by using adversarial loss, which compels the generated images to have similar distribution to the real images in the training data. In order to add more constraints in the generation process, various derivatives of GAN have been developed, such as conditional GAN (CGAN) [36], and auxiliary classifier GAN (ACGAN) [37]. They usually took extra information (e.g. labels, attention, and attributes) as auxiliary input to serve as specific conditions. Due to the great generation capability of GAN, numerous GAN-based face sketch synthesis methods have been proposed. Wang et al. [38] presented a back-projection procedure for the synthesized sketches with GAN (BPGAN). Gao et al. [39] took advantage of face labels and compositional loss to add facial details for sketch portrait generation. Zhang et al. [40] utilized multi-domain adversarial learning to synthesize high-quality sketch images. In [41], Zhu et al. proposed a GAN-based deep collaborative face photo-sketch synthesis framework with two opposite networks which are able to utilize the mutual interaction of two opposite mappings. The previous methods require a mass of paired photo–sketch images as training data, which are difficult to obtain. To handle this problem, Zhu et al. [42] developed the Cycle-consistent adversarial networks (CycleGAN) to deal with the unpaired image-to-image translation tasks. Based on CycleGAN, Wang et al. [43] applied the multiple adversarial networks on the feature maps with different resolution. Chen et al. [44] presented a semi-supervised learning (SSL) framework to resolve the limited training photo–sketch pairs for face sketch synthesis, which obtained impressive performance in the wild. Zhu et al. [45] employed knowledge transfer framework for training a face photo-sketch synthesis model with a small set of photo-sketch pairs.

It can be seen that GAN has become an important deep learning framework for face sketch synthesis research and attracted more and more attentions. For GAN, the design of the generator is most critical. In this study, a modified high-resolution network is proposed as the generator to produce realistic face sketch images.

3. The proposed method

This section presents a generative adversarial learning framework for face photo-to-sketch synthesis. The overall architecture of the proposed method is first introduced, and then the designed generator and discriminator structures are described in detail. Afterwards, the loss functions employed for training the proposed networks are defined.

3.1. Overall architecture

This paper aims to design a framework that is able to synthesize a realistic face sketch image with abundant facial details. Fig. 2 shows the overall architecture, which mainly consists of a GAN deep learning framework. In this work, we firstly improve the high-resolution network as generator G, which is fed with a face photo as input image, and a corresponding sketch image with rich facial details can be obtained. Then, a naive CNN is used as discriminator D to distinguish the synthesized sketches and real sketches.

Assuming (x, y) represents a photo–sketch image pair sampled from training data, $\hat{y}=G(x)$ is the target face sketch image. The objective of GAN for face sketch synthesis can be represented as:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim P_x(x)}[\log(1-D(G(x)))] + \mathbb{E}_{y\sim P_y(y)}[\log D(y)] \quad (1)$$

where $P_x(x)$ and $P_y(y)$ are the data probability distributions of training photo and sketch images. $G(x)$ is the synthesized fac sketch image by the generator G. $D(G(x))$ and $D(y)$ are the outputs of the discriminator D whose inputs are the synthesized fac sketch image $G(x)$ and the real fac sketch image y, respectively.

To train the proposed face sketch synthesis framework, beside the adversarial loss, three more loss functions are employed, including detail loss, style loss, and total variation loss. For simple depiction, Fig. 2 only shows the detail loss and the style loss. In the detail loss, the LoG feature maps of the input photo and synthesized sketch images are extracted, and their difference is calculated as loss value. In the style loss, the K nearest photo-sketch pairs are first selected based on the VGG19 feature similarity of input photo and training photos. Then, the distance between feature maps of

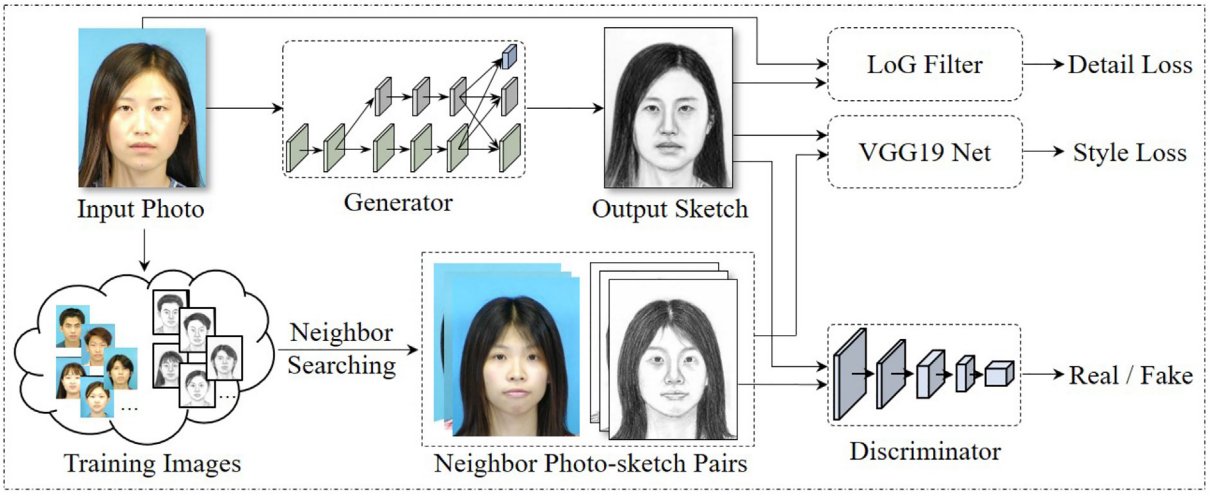


Fig. 2. Illustration of the proposed face sketch synthesis method.

synthesized sketch and pseudo feature maps constructed from the K selected neighbor sketches measures the style loss value.

3.2. Detailed generator and discriminator networks

3.2.1. Generator network

In the proposed method, the modified high-resolution network is utilized as the basic structure of the generator. The original high-resolution network was first proposed for human pose estimation [46], and further applied for other computer vision tasks, e.g. face landmark detection, semantic segmentation, object detection, and image classification [15]. The main characteristics of the high-resolution network are that it can maintain the high-resolution representation throughout the whole network, and repeatedly conducts multi-scale fusion across parallel convolutions from different stages. Motivated by its remarkable success in high-resolution representation, we employ the high-resolution network as generator for face sketch synthesis, and modify its head structure, in order to gradually aggregate the feature maps from low-resolution to high-resolution.

Fig. 3 shows the basic high-resolution network, which in the proposed approach is composed of four stages. The 1st stage is the high-resolution convolutions. From the 2nd stage, a multi-resolution block is composed of a multi-resolution group convolution, and a multi-resolution convolution is placed. In multi-resolution group convolution, the input channels are divided into several subsets of channels, and the convolutions are separately performed over each subset at different spatial resolutions. In multi-resolution convolution, the input channels are divided into several subsets with multiple resolutions, and the output channels are also divided into several subsets with multiple resolutions. The input and output subsets are connected in a fully connected fashion, and the connections include three different types: 1) regular convolution; 2) upsampling with bilinear interpolation; and 3) downsampling with 2-strided and 3×3 convolutions. The output channel maps of each subset are a summation of the outputs from each subset of the input channel maps.

In the original high-resolution network, the outputs of different resolution convolutions at the final stage are rescaled and concatenated at once, which leads to loss of detail information in lower resolution feature maps. Instead, we upsample and merge them gradually, as compared in Fig. 4(a) and (b). The gradual merging strategy helps to preserve more facial details in synthesized sketch image. As depicted in Algorithm 1, from the lowest resolution, we first interpolate the feature map up a scale, and concatenate it with the higher resolution feature map. Afterward, two convolutional layers with kernel sizes of 1×1 and 3×3 are conducted, and the channels of the output feature map are the same as those of the higher feature map. At the end, the high-resolution output can be obtained. Moreover, we replace the batch normalization

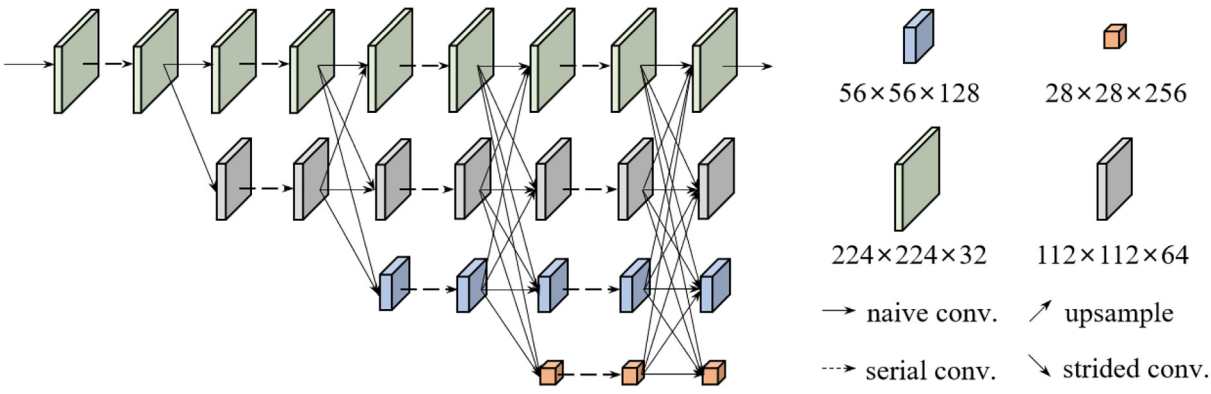


Fig. 3. The basic structure of the high-resolution network. The horizontal direction corresponds to the stage of the high-resolution network, while the vertical direction represents the resolution of the feature maps.

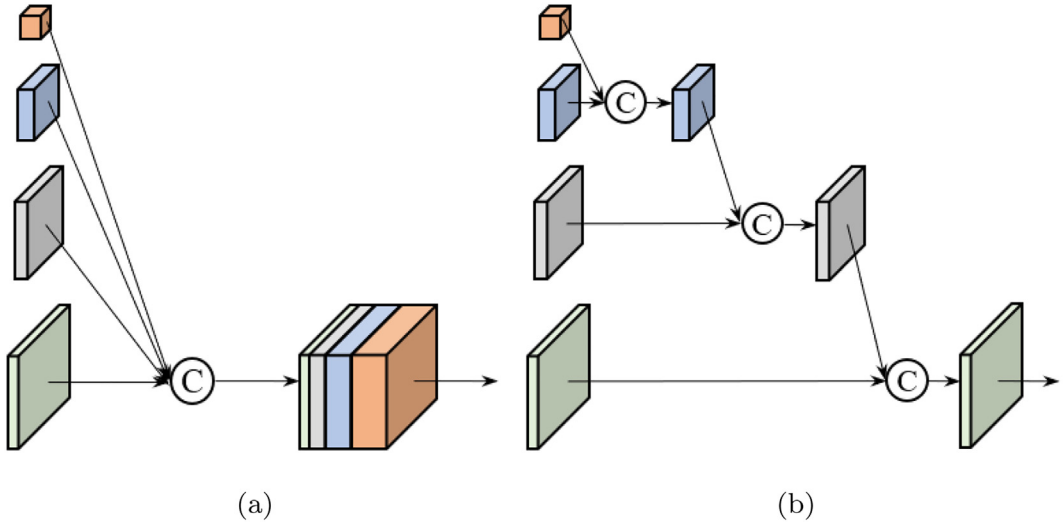


Fig. 4. The fusion strategies of high-level feature maps in (a) original high-resolution network, and (b) our method. Here, © indicates the concatenation operation.

in the original high-resolution network with the instance normalization, which has been proven to achieve better performance in image-to-image transform tasks [47].

Considering the small numbers of training photo-sketch pairs, we narrow down the size of high-resolution network model. The detailed instantiation of our generator network is described here. The 1st stage is composed of three residual units, and the feature map in each unit has the shape of 224 × 224 × 32. The size of the feature map compresses by half, and the channels of the feature map double every resolution reduction. There are 1, 3, and 2 multi-resolution blocks in the 2nd, 3rd, and 4th stages, respectively. Each branch in the multi-resolution group convolution contains three residual units, while each unit contains two convolution layers with filter size of 3 × 3 in each resolution.

Algorithm 1. Modified Feature Map Fusion Strategy.

Input: Feature maps of different resolution convolutions at the final stage: $F_1, F_2, \ldots, F_N$ (N is the number of resolutions);
for $t = 1$ to $N-1$ **do**
1. Interpolate the feature map up a scale:
$\overline{F}_t = upsample(F_t)$
2. Concatenate it with the higher resolution feature map:
$\widehat{F}_t = concatnate(\overline{F}_t, F_{t+1})$
3. Conduct convolutions on concatenated feature map:
$F_{t+1} = convolute(\widehat{F}_t)$
end for
Output: fused high-resolution feature map F_N.

3.2.2. Discriminator network

The basic structure of the discriminator network includes five convolutional layers with strides 2, kernel size 3, and padding 1. The numbers of filters are 32, 32, 64, 64, and 128, respectively. After each convolutional layer, the batch normalization and ReLU activation are stacked. Followed the basic discriminator network, a convolutional layer with output size of 7 × 7 × 1 and a Sigmoid activation layer are conducted to predict probability scores between 0 and 1, which are utilized to distinguish whether the observed sketch image is real or fake. The detailed structure and parameter setting of the discriminator network can be seen in Table 1.

3.3. Loss functions

This part introduces the loss functions employed for training the proposed GAN model. Assume that x and y are the training photo–sketch pair, z is the selected neighbor sketch image, and $\hat{y}$ is the synthesized sketch.

1) *Adversarial loss for generator* is a basic loss in GAN, which aims at guiding the generator to produce data that look real, and is able to deceive the discriminator. It can be represented as:

$$L_{adv\_G} = \mathbb{E}_{x\sim P_{photo}(x)}\left[(D(G(x)) - 1)^2\right] \tag{2}$$

where L_{adv_G} is the adversarial loss for generator, $P_{photo}(x)$ is data probability distributions of training photo images.

2) *Detail loss* is designed to measure the facial detail discrepancy. Existing face sketch synthesis methods rarely pay attention to detail preservation, and thus inevitably result in blur and artifact effect. In this work, we propose a detail loss to increase the detailed information in resultant sketch images. In order to avoid the influence by the deformation in hand-drawn sketch image, we calculate the mean square error between the LoG maps of the predicted sketch image and the input photo image, instead of the ground-truth hand-drawn sketch. That is although the face photo and its synthesized sketch have diverse modality, their facial details should be similar, as shown in Fig. 5. Therefore, the detail loss can constrain the synthesized image has proximate detail information with source image. Eq. (3) represents the definition of detail loss:

$$L_{detail}(x, \hat{y}) = \|LoG(x) - LoG(\hat{y})\|_2^2 \tag{3}$$

where L_{detail} represents the detail loss, LoG means the Laplacian of Gaussian filter process.

3) *Style loss* is utilized to enforce the generated face sketch images to have similar style to the real hand-drawn face sketch images. It also helps for face sketch recognition due to the modality-discrepancy being reduced after transforming face photo into sketch domain. The high-level feature maps extracted by the pre-trained VGG deep model are commonly used to represent image style [33]. In this paper, we adopt the pseudo-sketch feature [44] to measure the style loss. The flowchart of style loss is displayed in Fig. 6, and the formula of style loss is as follow:

Table 1
Architecture of the discriminator network, where 'conv' means the convolution layers.

Layer	Conv	Conv	Conv	Conv	Conv	Conv
Kernel Size	3 × 3	3 × 3	3 × 3	3 × 3	3 × 3	3 × 3
Stride	2	2	2	2	2	1
Output size	112 × 112 × 32	56 × 56 × 32	28 × 28 × 64	14 × 14 × 64	7 × 7×128	7 × 7×1

Fig. 5. Comparison of the facial details between face photo and sketch. (a) Photo image and its facial details extracted with LoG filter; (b) sketch image and its facial details extracted with LoG filter.

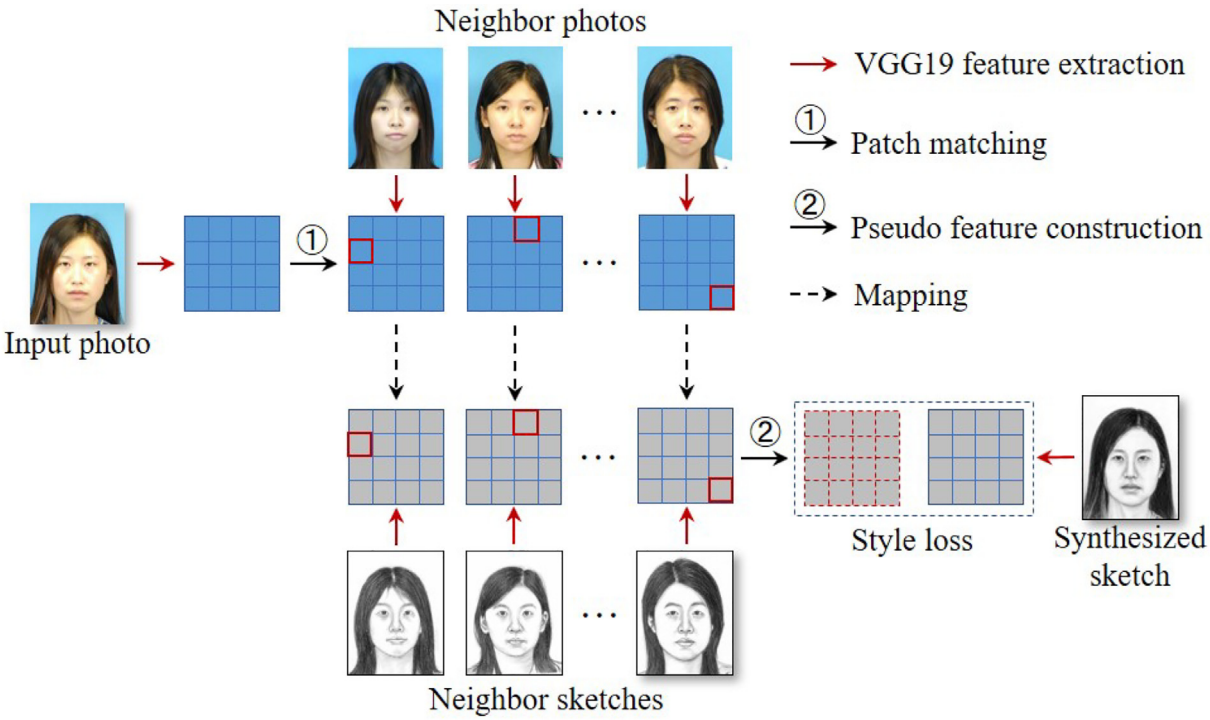


Fig. 6. Illustration of the style loss. The face feature maps are extracted with the VGG19 network. For each feature map batch of input photo, the nearest batch is selected from feature map batches of neighbor photos at the same position. Then, the corresponding feature map batches of sketches are used to construct the pseudo sketch feature map. The distance between the pseudo sketch feature map and the feature map of the synthesized sketch forms the style loss.

$$L_{style}(x,\hat{y}) = \sum_{i=1}^{5}\sum_{j=1}^{J}\left\|\Psi_j\left(\Phi^i(\hat{y})\right) - \Psi'_j\left(\Phi^i(x)\right)\right\|_2^2 \tag{4}$$

where L_{style} represents the style loss, $\Phi^i(\cdot)$ is the extracted feature map at the first channel of ith group of convolution layers in the VGG19 network, which has 5 groups of convolution layers; $\Psi_j\left(\Phi^i(\hat{y})\right)$ is a $k \times k$ patch centered at a point j of feature map $\Phi^i(\hat{y})$, J is the number of total patches; and $\Psi'_j\left(\Phi^i(x)\right)$ is the feature map patch of the neighbor sketch whose corresponding photo has the best match feature map in this batch with input photo x.

4) *Total variation loss* is commonly used in deep learning-based style transfer tasks, which is able to eliminate noise and artifact

in the target images [48–50]. To further improve the quality of the generated face sketch image, we conduct the total variation loss on the output sketch of the generator:

$$L_t(\hat{y}) = \sum_{m,n}\left((\hat{y}_{m+1,n} - \hat{y}_{m,n})^2 + (\hat{y}_{m,n+1} - \hat{y}_{m,n})^2\right) \tag{5}$$

where $\hat{y}_{m,n}$ is the pixel value at (m, n) of the synthesized face sketch image.

The total loss function for training the generator is formed by weighted combining of the aforesaid loss functions:

$$L_G = \lambda_1 L_{adv\_G} + \lambda_2 L_{style} + \lambda_3 L_{detail} + \lambda_4 L_t \tag{6}$$

where the parameters of λ are the weighting coefficients, which are set as [1] to scale the loss values at the same order of magnitude.

5) *Adversarial loss* for discriminator is used to train the discriminator network in GAN. The well-trained discriminator is able to distinguish the true sketch image from a generated pseudo sketch image. The loss function for the discriminator is calculated on the generated sketch $G(x)$ and real sketch z, which we randomly select from the neighbor sketch images:

$$L_D = \frac{1}{2}\mathbb{E}_{z\sim P_{sketch}(z)}\left[(D(z)-1)^2\right] + \frac{1}{2}\mathbb{E}_{x\sim P_{photo}(x)}\left[D(G(x))^2\right] \tag{7}$$

where L_D is the adversarial loss for discriminator, $P_{sketch}(z)$ is data probability distributions of training sketch images.

4. Experimental results

4.1. Datasets and implemental details

We evaluate the proposed method on two public face sketch datasets: CUFS [19] and CUFSF [51], which comprise 606 and 1143 photo–sketch pairs, respectively. The photo–sketch pairs in the CUFS dataset are well-aligned; however, CUFSF is a more challenging dataset, due to the sketches having serious deformation, and are not being well aligned with the photos. Moreover, the photos in the CUFSF dataset are captured under various light environments. Fig. 7 shows the examples of photo-sketch pairs in the CUFS and CUFSF datasets.

The experiment setting and implementation details of the proposed method are introduced here. For convenient comparison, we select 268 photo–sketch pairs in the CUFS dataset and 250 photo–sketch pairs in the CUFSF dataset for training, as the commonly-used data splitting rule in most face sketch synthesis literature [13][38]. The remaining photo-sketch pairs are used for testing. The photo and sketch images of CUFS and CUFSF datasets are with size of 200 × 250, and when feeding into network for training, they are resized to 224 × 224. The deep network models in this work are implemented with Pytorch and trained on a Nvidia Titan X GPU. The training epochs number is 40, and the Adam optimizer [52] with learning rate 0.002 is employed for network training.

To evaluate the performance of different face sketch synthesis methods, apart from the visual comparison, two image quality evaluation indices were employed for objectively assessing the synthesized face sketches. Feature similarity image measurement (FSIM) [53] is a frequently used index for evaluating image quality, which has high consistency with human visual perception and can capture similarity between low-level features of the synthesized image and the ground-truth image. Assume that M, N be the synthesized face sketch image and its ground-truth hand-drawn sketch image, respectively. The FSIM value of M and N can be represented as:

$$\mathrm{FSIM}(M,N) = \frac{\sum S_{PC}\cdot S_G \max(PC(M), PC(N))}{\sum \max(PC(M), PC(N))} \tag{8}$$

where PC is image phase congruency, S_{PC} and S_G are defined as follows:

$$S_{PC} = \frac{2PC(M)\cdot PC(N) + T_1}{PC^2(M) + PC^2(N) + T_1} \tag{9}$$

$$S_G = \frac{2GM(M)\cdot GM(N) + T_1}{GM^2(M) + GM^2(N) + T_1} \tag{10}$$

where GM means image gradient magnitude, T_1 and T_2 are positive constants to increase the stabilities of S_{PC} and S_G.

In addition, Fan et al. [54] proposed a new image quality evaluation index specifically for face sketch synthesis, named structure co-occurrence texture (Scoot), which has the ability to measure the perceptual similarity between face sketches by simultaneously considering the co-occurrence texture statistics and block-level spatial structure. The Scoot value of M and N can be represented as:

Fig. 7. Examples of the CUFS dataset (first three columns) and CUFSF dataset (last two columns).

$$\text{Scoot}(M,N) = \frac{1}{1+\left\|\vec{\Psi}(M') - \vec{\Psi}(N')\right\|_2} \tag{11}$$

where M' and N' are the quantized M and N, respectively. $\vec{\Psi}(\cdot)$ represents the operation of calculating the average feature at different orientation vectors.

4.2. Ablation study

The proposed face sketch synthesis method integrates the modified high-resolution network together with several elaborate loss functions. In order to verify the effectiveness of these components, two groups of experiments were conducted on the CUFS dataset, in which the effectiveness of our proposed feature map fusion strategy and detail loss were evaluated emphatically.

4.2.1. Effectiveness of the modified fusion strategy

The first experiment fixed all of the loss functions, and tested the performance of the different components in the modified high-resolution network. The images in first row of Fig. 8 illustrate the synthesized face sketches by original fusion strategy and our modified one, from which we can observe that less noise and finer detail information are achieved with the proposed feature map fusion strategy. From Table 2, it can be seen that the performance improves as more components are added in the generator network.

4.2.2. Effectiveness of the proposed detail loss

The second experiment fixed the modified high-resolution network, and tested the performance of the loss functions in our method. The images in second row of Fig. 8 depict the input photo and the synthesized results without style loss and detail loss, without detail loss, and with all aforementioned losses, respectively. From visual comparison, the style loss polishes up sketch style in the generated results; however, the detail loss preserves more facial detail from source photo significantly. Table 3 indicates that each loss function has positive effectiveness for training the proposed generator network. Please note that the first adversarial loss in Table 3 includes a simple mean squared error loss, because only the adversarial loss cannot train the GAN model well.

4.3. Comparison with state-of-the-art face sketch synthesis methods

To demonstrate the performance of the proposed face sketch synthesis method, six state-of-the-art approaches are compared,

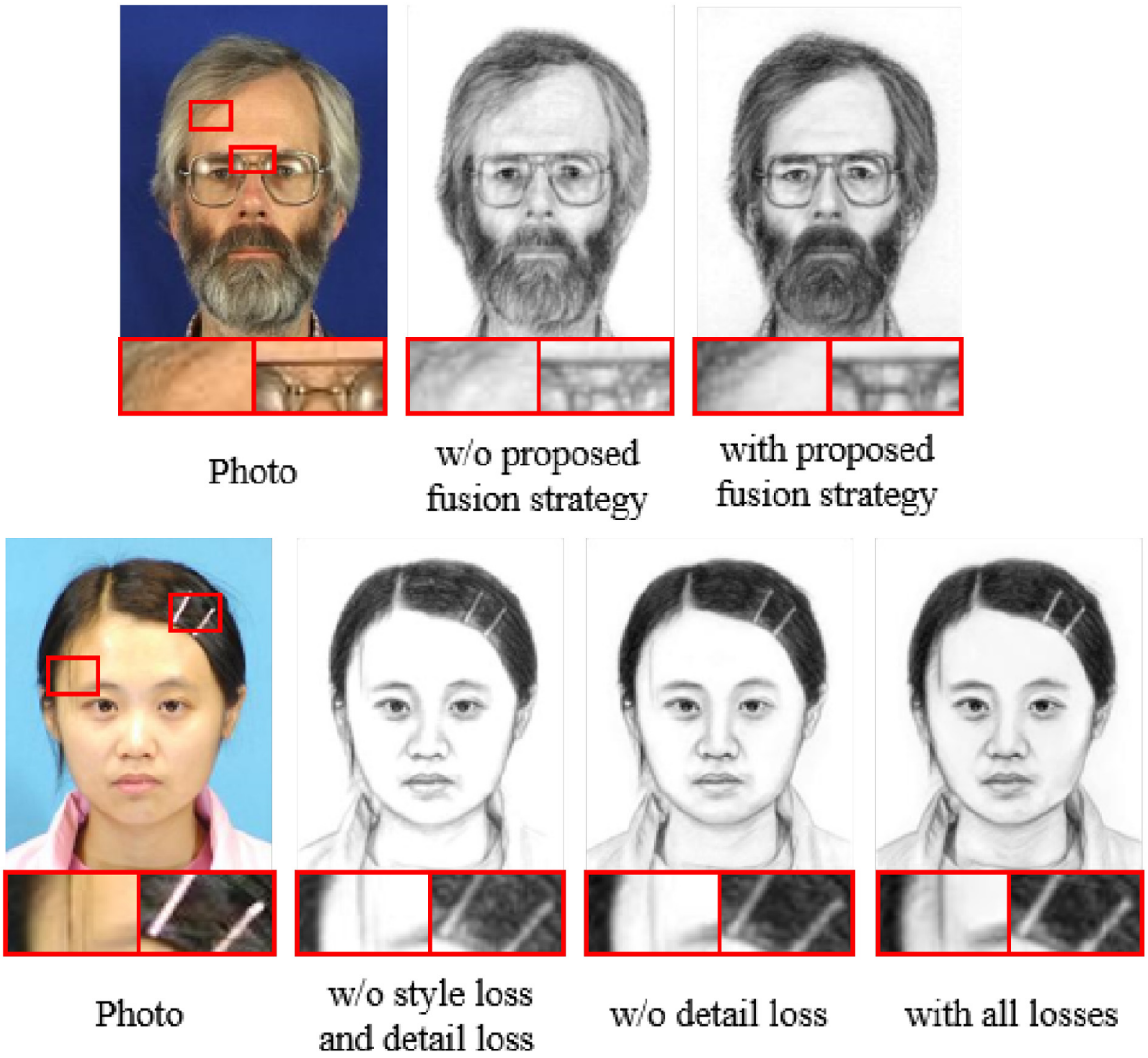


Fig. 8. Results of face sketch synthesis with different fusion strategies and different loss functions.

Table 2
The ablation study of the components in the modified high-resolution network on the CUFS dataset.

High-resolution network	Modified fusion strategy	Instance normalization	FSIM	Scoot
√			0.7273	0.5123
√	√		0.7316	0.5268
√	√	√	**0.7345**	**0.5317**

Table 3
The ablation study of the loss functions on the CUFS dataset.

Adversarial loss	Style loss	Detail loss	Total variation loss	FSIM	Scoot
√				0.7081	0.4750
√	√			0.7290	0.5236
√	√	√		0.7313	0.5258
√	√	√	√	**0.7345**	**0.5317**

including exemplar-based methods (MWF [20], SSD [21], and RSLCR [13]), and deep learning-based methods (FCN [25], BPGAN [38], and SSL [44]). Figs. 9 and 11 make visual comparisons between these methods and the proposed method on the CUFS and CUFSF datasets, respectively. From them we can observe that the MWF method leads to blurring effects around the image edge area, because of the weighted averaging of multiple neighbor sketches. Due to the denoising process in the SSD method, the detailed information of the sketch image weakens together with the noise. For RSLCR method, while better results can be obtained, the weighted averaging on randomly selected sketch patches results in blurring as well, like the MWF method. The FCN method is able to generate intact facial structure, but a serious noise problem exists in the synthesized sketch images. The BPGAN method conducts a back-projection step on the synthesized sketches by GAN model, and can generate clear and clean face sketch images. However, the sketch images by the BPGAN method have some deformation compared with the input photo images, especially on the CUFSF dataset. The SSL method can obtain excellent results, and has the closest performance to our method, except for some fine-grained facial components, like eyes and plications. In summary, the exemplar-based methods usually suffer from block and blur effects, while the deep learning-based methods show noise and deformation problems. Comparatively, the proposed method generates the most realistic face sketch images with abundant facial details, unabridged facial content,

Fig. 9. Comparisons of the synthesized face sketch images by different methods on the CUFS dataset.

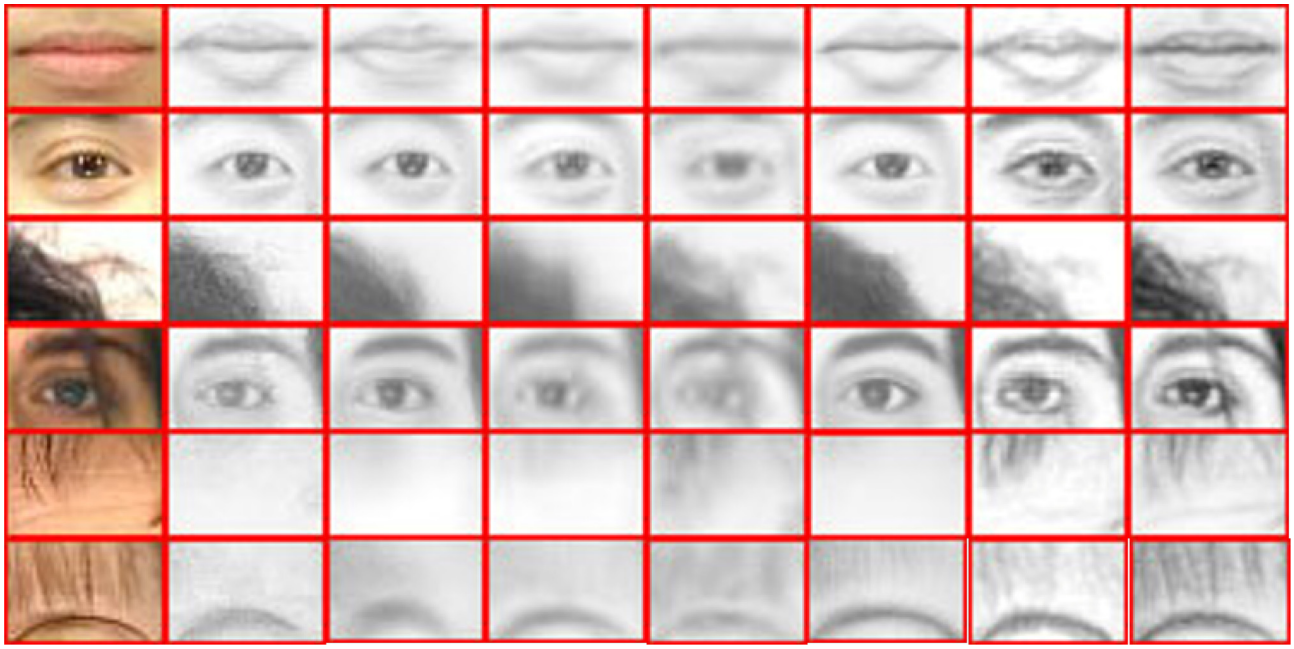


Fig. 10. Local cut-out effect on Fig. 9; same marshalling sequence as Fig. 9.

and less noise. In order to visually compare the synthesized sketch images by different methods more clearly, we cut the sub-images in Fig. 9, and display them in Fig. 10,11. It can be observed that the proposed approach can surpassingly preserve the detailed information of the face photo images, such as hair, eyelids, and plication.

Table 4 shows the average FSIM and Scoot indices that we calculated as objective evaluation of the CUFS and CUFSF datasets. It indicates that our approach achieved the highest index values on the CUFS dataset, and comparable results on the challenging CUFSF dataset. The objective evaluation results further support the effectiveness of our approach.

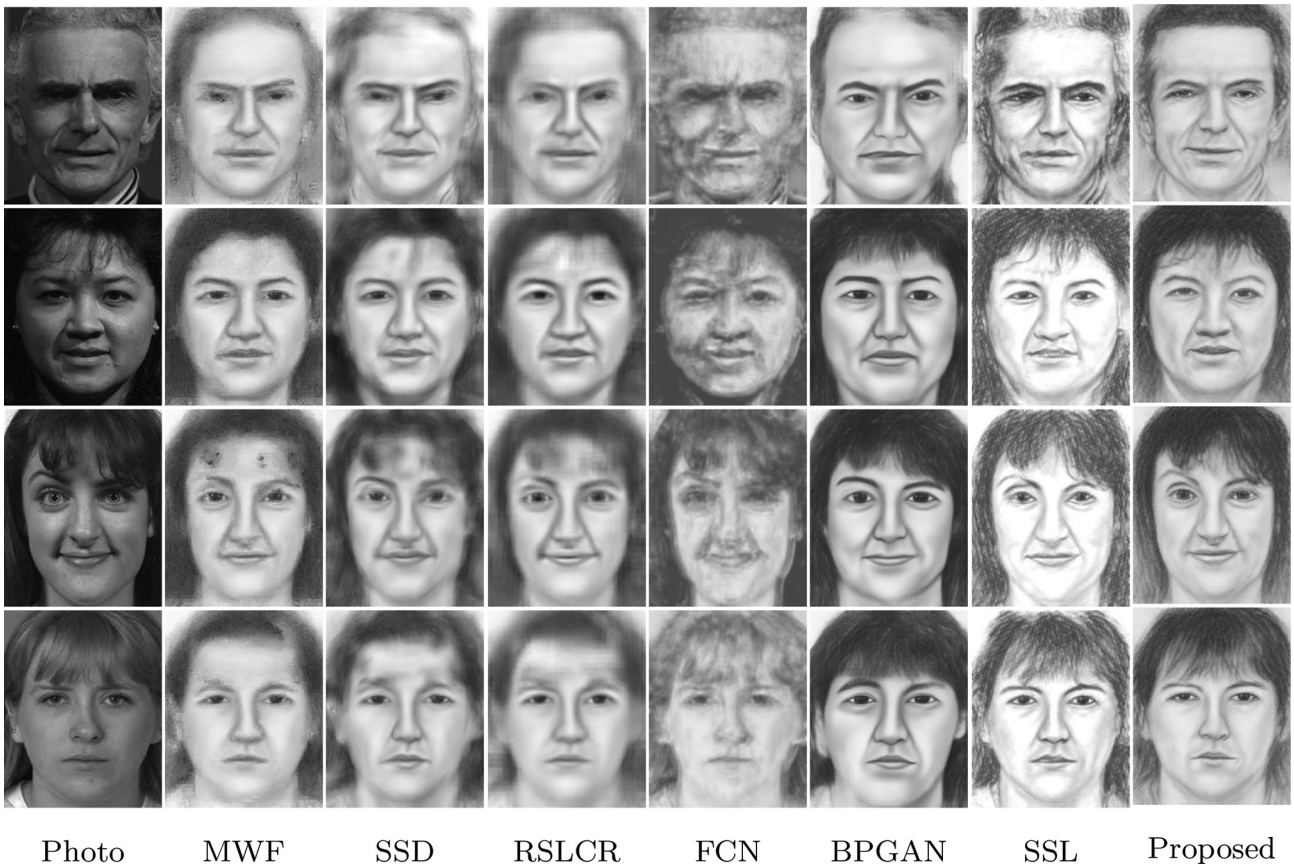


Fig. 11. Comparisons of the synthesized face sketch images by different methods on the CUFSF dataset.

Table 4
Objective evaluation of the synthesized sketch images by different methods on the CUFS and CUFSF datasets. The best values are depicted in bold font.

	Methods	MWF	SSD	RSLCR	FCN	BPGAN	SSL	Proposed
CUFS	FSIM	0.7145	0.6959	0.6966	0.6936	0.6899	0.7256	**0.7345**
	Scoot	0.4833	0.4544	0.4499	0.4527	0.4680	0.4878	**0.5317**
CUFSF	FSIM	0.7029	0.6824	0.6650	0.6624	0.6814	**0.7159**	0.7080
	Scoot	0.4882	0.4687	0.4531	0.4378	0.4936	0.5038	**0.5091**

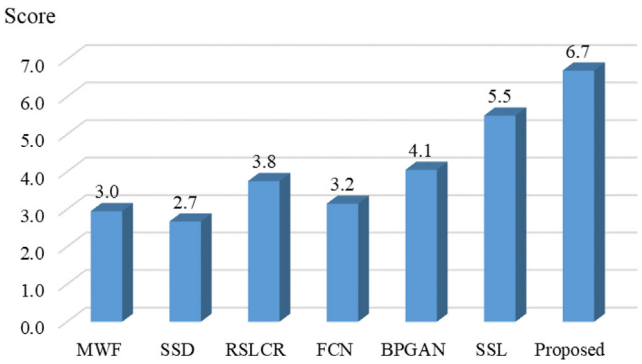


Fig. 12. User study results on the synthesized face sketches of Fig. 9.

Moreover, a user study is conducted to help with the performance measurement of the different face sketch synthesis methods. The synthesized face sketches by different methods in Fig. 9 are rearranged randomly, and the method names are hided before showing to the participants. There are 20 subjects participated in the user study, and all of them have prior experience on image processing. The subjects are asked to evaluate the synthesized face sketches from various aspects, e.g. facial details, sketch style et al., and to qualitatively mark the synthesized face sketches of the comparison methods from 1 to 7. The statistical results of the user study are shown in Fig. 12. From the figure, it can be seen that the proposed method achieves the highest mean opinion score of 6.7, which indicates the superiority of the proposed method.

4.4. Face sketch synthesis on real-world photos

In previous section, the test photos for sketch synthesis experiments were captured under the same controlled environment as the training photos, such as simple background, uniform illumination, and normal expression. However, in the real-world situation, the background, illumination, and expression may vary. To prove the robust of our approach in the real-world case, we conducted the face sketch synthesis experiment on photo images from the CelebA face dataset [55]. The test photo images are aligned according to the coordinates of the center of two eyes, and cropped with size of 100 × 125. The lower image resolution, compared to the training data, makes the synthesis task more challenging.

The RSLCR method [13] and the SSL method [44], as the representative of exemplar- and deep learning-based face sketch synthesis methods, respectively, are utilized to compare with our

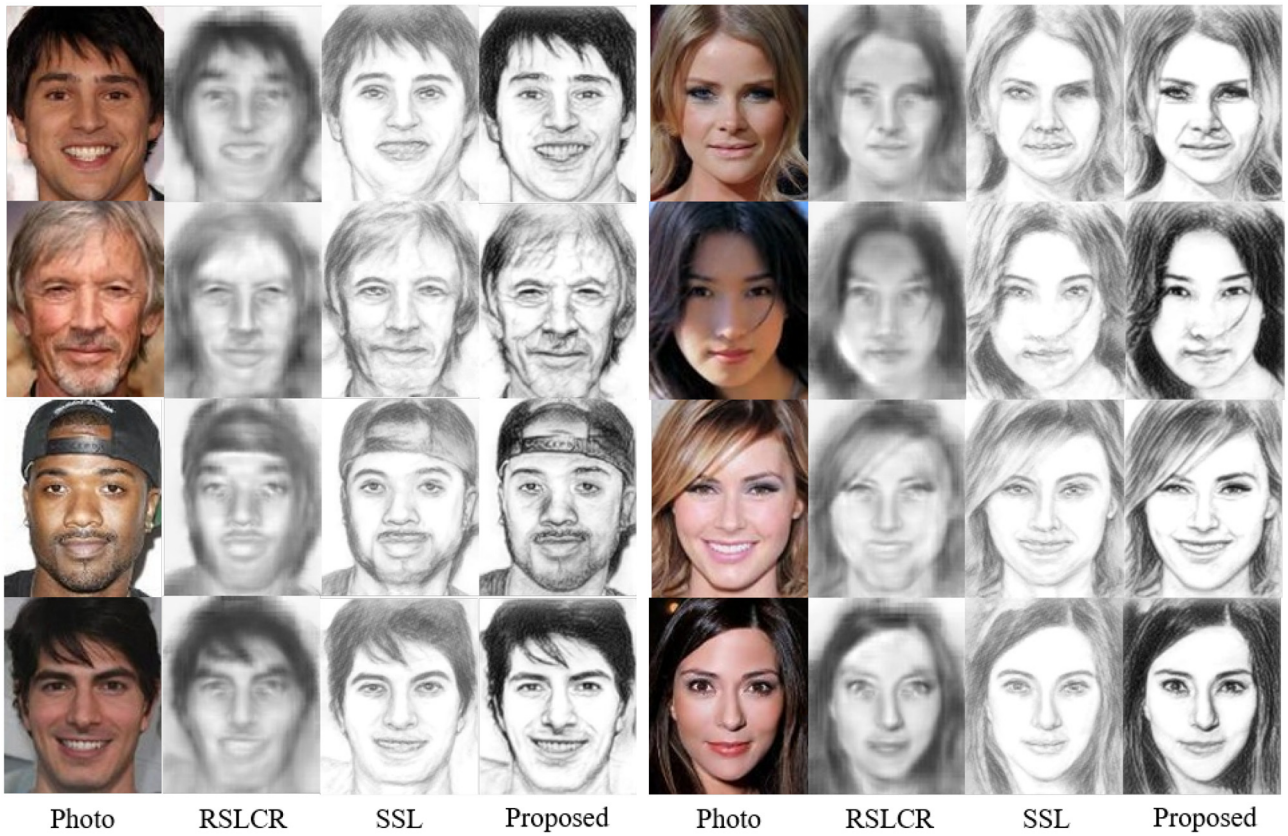


Fig. 13. Face sketch synthesis results on real-world photos from the CelebA face dataset.

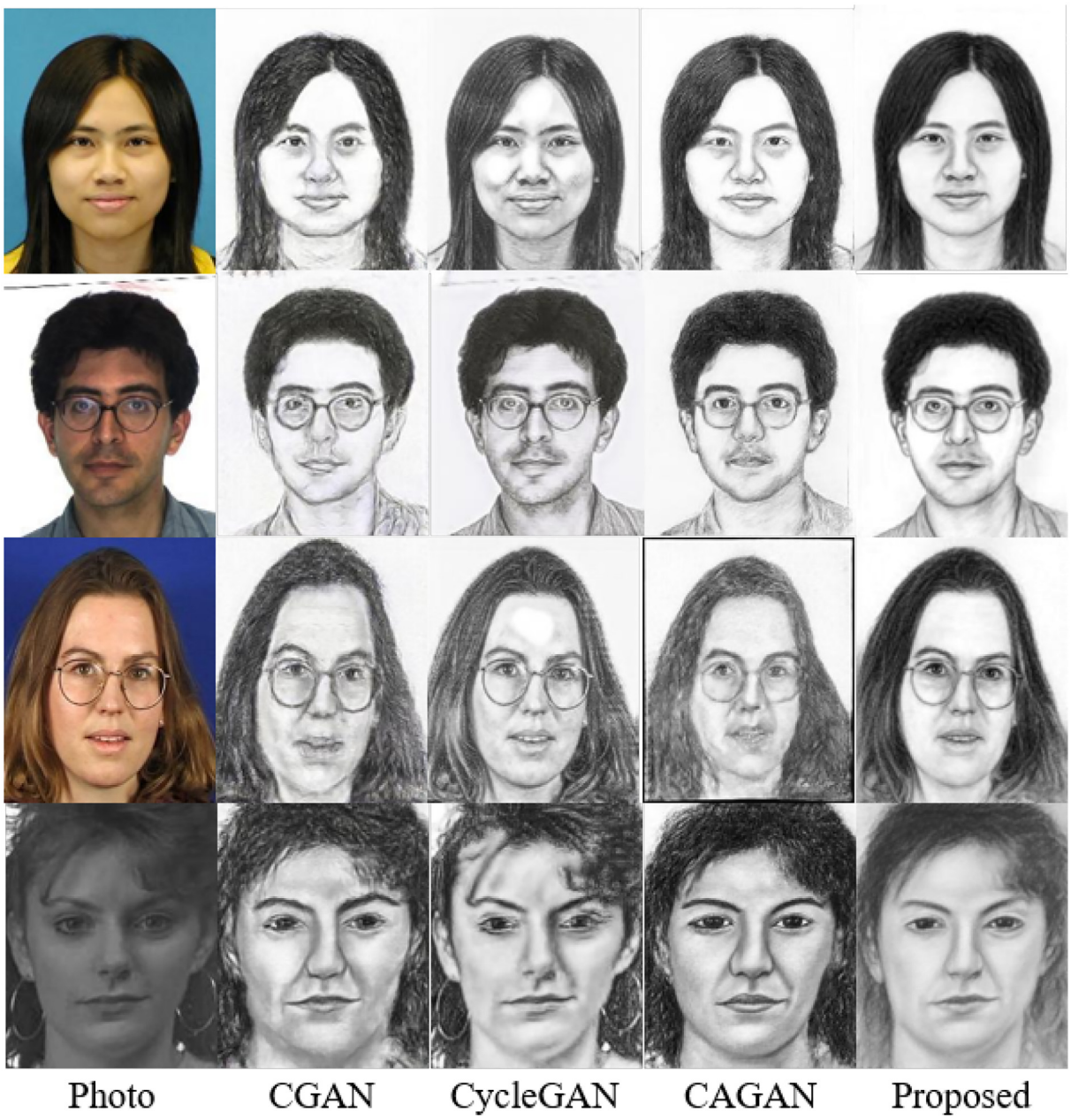


Fig. 14. Comparison results with state-of-the-art GAN-based methods. First three rows are from CUFS dataset and the last row are from CUFSF dataset.

approach. Fig. 13 displays the comparison results, from which we can see that the synthesized sketches by the RSLCR method are blurred and smooth. The main reason is the huge differences between the training photos and test photos make the test photo patch hard to approximate with exemplar-based methods. In contrast, the deep learning-based methods like SSL and our method are little influenced by image discrepancy between the training photos and test photos. From the visual comparisons, our results have better sketch style and more facial details, e.g. skin texture and teeth area. The experiment on real-world photos indicates the robustness and generalization of our proposed method.

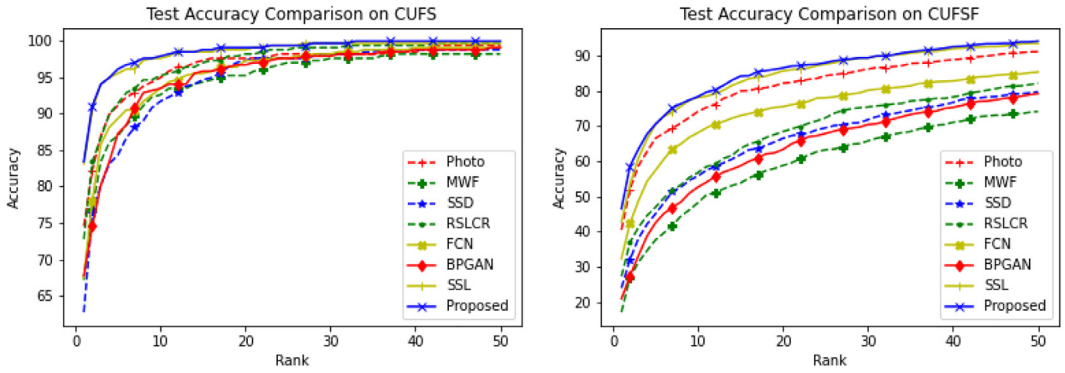
4.5. Comparison with state-of-the-art GAN-based methods

To further demonstrate the superiority of the proposed face sketch synthesis approach, more recent GAN-based methods were compared, include CGAN [36], CycleGAN [42], and CA-GAN [39]. The visual comparisons of synthesized sketches are shown in Fig. 14. It is obvious that the CGAN method causes various degrees of deformation and noise in synthesized face sketches. The results by the CycleGAN method have serious artifacts and thus look unnatural. The CA-GAN method can capture overall face structure well, however, some details are lost, especially for the photos from CUFS dataset. By contrast, our results achieve best performance in terms of style transfer and detail preserving.

4.6. Generalizability

To indicate the generalizability of the proposed method, we have added the experiments on non-frontal faces and images without human faces, the synthesized results are shown in Fig. 15. In this figure, the first row displays the non-frontal face images and their corresponding sketch images, the second row displays the images with human faces and their corresponding sketch images.

Fig. 15. Synthesized sketches on non-frontal faces and images without human faces.



(a) (b)

Fig. 16. Face sketch recognition results by different methods on the (a) CUFS and (b) CUFSF datasets.

Table 5
Face sketch recognition accuracies (%) of VGGFace2 on the CUFS and CUFSF datasets.

Methods	CUFS			CUFSF		
	Rank-1	Rank-5	Rank-10	Rank-1	Rank-5	Rank-10
Photo	74.5	91.1	95.0	40.5	66.5	74.1
MWF	67.2	87.0	92.6	17.1	37.6	48.2
SSD	62.7	84.3	91.7	23.9	45.0	55.8
RSLCR	72.8	91.4	95.3	27.2	46.9	56.8
FCN	67.5	89.3	93.5	32.3	57.4	68.1
BPGAN	67.8	87.3	93.5	20.9	42.5	52.6
SSL	83.1	95.6	97.6	42.8	70.4	78.2
Proposed	**83.4**	**96.2**	**97.9**	**46.5**	**70.6**	**78.4**

From Fig. 15, it can be observed that the proposed method can achieve impressive results, even though in the challenging scenarios.

4.7. Face sketch recognition

Face sketch recognition between synthesized face sketches and original hand-drawn sketches is an alternative way to quantitatively assess the performance of the face sketch synthesis approaches [?,13]. Higher face sketch recognition accuracy represents better synthesized sketch quality, and more effective sketch generation method. To conduct the face sketch recognition experiments, VGGFace2 [56] was adopted to extract facial features from sketch images, and Euclidean distance was utilized to measure the feature similarity. In this work, the synthesized sketches by different methods or the face photos were taken as probe images to match the gallery images consisting of the corresponding hand-drawn sketches. For the CUFS dataset, 338 synthesized sketches or face photos were used as the probe set, and the corresponding ground-truth sketches drawn by the artist were taken as

the gallery set. For the CUFSF dataset, 944 synthesized sketches or face photos were taken as the probe set, and the corresponding hand-drawn sketch images were taken as the gallery set.

Fig. 16 shows the face sketch recognition results on the CUFS and the CUFSF datasets. In comparison, our method obtained the highest recognition accuracies on both datasets, of 100% and 94.17% in the CUFS and CUFS datasets at Rank-50, respectively. Here, Rank-*n* represents the recognition accuracy of the top-*n* best matches. Table 5 displays the exact recognition accuracies at Rank-1, Rank-5, and Rank-10. It is evident that the proposed method achieved best values at all the three indices. The preeminent recognition results demonstrate that our synthesized face sketch images with more facial details and better sketch style promote the face sketch recognition performance. In addition, compared to the direct face sketch recognition without synthesis process, transferring face images from photo domain to sketch domain by the proposed method, the face sketch recognition performance is indeed improved.

4.8. Future scopes

With the development of deep learning technology, the face sketch synthesis has achieved excellent performance in generating vivid face sketch images. However, there are still shortcomings which are not addressed very well in the related research topics: 1) The existing works mainly focus on the frontal face transformation, the results on non-frontal face are still unsatisfactory. More attentions and efforts can be paid to address this problem; 2) The semi-supervised and unsupervised face sketch synthesis methods will become trends to handle the lack of training photo-sketch pair data; 3) Although the performance of face sketch recognition has been improved by transforming the face photo to sketch domain, the recognition accuracy is not high as normal face recognition. More factors, e.g. facial attributes and identity-aware, can be considered to further boost the accuracy of face sketch recognition.

5. Conclusions

In order to handle the loss of facial details in generated face sketches, we proposed a detail-preserving face photo-to-sketch synthesis approach based on GAN in this paper. Firstly, we modified the high-resolution network to gradually fuse the feature maps from different resolutions. In addition, we designed a loss function named detail loss to force the generated face sketch image to preserve more detailed information from the photo image. We utilized the style loss and total variation loss to further improve the performance of our approach. The experimental results on the CUFS, CUFSF, and CelebA face datasets indicate that the face sketch images synthesized by our approach not only have better facial detail information, but also lead to higher objective evaluation indices and face recognition accuracy, compared to the existing face sketch synthesis approaches. In future, we will explore face sketch synthesis with unpaired data, hence vast photo images from normal face datasets can be used to train face sketch synthesis networks.

CRediT authorship contribution statement

Weiguo Wan: Methodology, Writing - original draft. **Yong Yang:** Investigation. **Hyo Jong Lee:** Supervision, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study has been supported in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (GR2019R1D1A3A03103736), and by the National Natural Science Foundation of China (62072218), and by the Natural Science Foundation of Jiangxi Province (20192ACB20002 and 20192ACBL21008), and by the Talent project of Jiangxi Thousand Talents Program (jxsq2019201056), and by the Project of the Education Department of Jiangxi Province (GJJ200541), and by the Postdoctoral Research Projects of Jiangxi Province (2020KY44).

References

[1] M. Zhang, R. Wang, X. Gao, J. Li, D. Tao, Dual-transfer face sketch–photo synthesis, IEEE Trans. Image Process. 28 (2) (2019) 642–657.
[2] N. Wang, M. Zhu, J. Li, B. Song, Z. Li, Data-driven vs. model-driven: fast face sketch synthesis, Neurocomputing 257 (2017) 214–221.
[3] M. Zhang, J. Li, N. Wang, X. Gao, Compositional model-based sketch generator in facial entertainment, IEEE Trans. Cybern. 48 (3) (2018) 904–915.
[4] W. Wan, Y. Gao, H.J. Lee, Transfer deep feature learning for face sketch recognition, Neural Comput. Appl. 31 (12) (2019) 9175–9184.
[5] H. Cheraghi, H.J. Lee, Sp-net: A novel framework to identify composite sketch, IEEE Access 7 (2019) 131749–131757.
[6] Y. Jin, J. Lu, Q. Ruan, Coupled discriminative feature learning for heterogeneous face recognition, IEEE Trans. Inf. Forensics Secur. 10 (3) (2015) 640–652.
[7] J. Huo, Y. Gao, Y. Shi, W. Yang, H. Yin, Heterogeneous face recognition by margin-based cross-modality metric learning, IEEE Trans. Cybern. 48 (6) (2018) 1814–1826.
[8] M. Zhang, N. Wang, Y. Li, X. Gao, Neural probabilistic graphical model for face sketch synthesis, IEEE Trans. Neural Networks Learn. Syst. 31 (7) (2020) 2623–2637.
[9] M. Zhang, N. Wang, Y. Li, R. Wang, X. Gao, Face sketch synthesis from coarse to fine, in: Thirty-Second AAAI Conference on Artificial Intelligence, pp. 7558–7565..
[10] W. Wan, H.J. Lee, A joint training model for face sketch synthesis, Appl. Sci. 9 (9) (2019) 1731.
[11] L. Jiao, S. Zhang, L. Li, F. Liu, W. Ma, A modified convolutional neural network for face sketch synthesis, Pattern Recogn. 76 (2018) 125–136.
[12] J. Jiang, Y. Yu, Z. Wang, X. Liu, J. Ma, Graph-regularized locality-constrained joint dictionary and residual learning for face sketch synthesis, IEEE Trans. Image Process. 28 (2) (2018) 628–641.
[13] N. Wang, X. Gao, J. Li, Random sampling for fast face sketch synthesis, Pattern Recogn. 76 (2018) 215–227.
[14] M. Zhang, N. Wang, Y. Li, X. Gao, Deep latent low-rank representation for face sketch synthesis, IEEE Trans. Neural Networks Learn. Syst. 30 (10) (2019) 3109–3123.
[15] K. Sun, Y. Zhao, B. Jiang, T. Cheng, B. Xiao, D. Liu, Y. Mu, X. Wang, W. Liu, J. Wang, High-resolution representations for labeling pixels and regions, arXiv preprint arXiv:1904.04514..
[16] X. Tang, X. Wang, Face photo recognition using sketch, in: Proceedings. International Conference on Image Processing..
[17] Q. Liu, X. Tang, H. Jin, H. Lu, 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)..
[18] X. Gao, J. Zhong, J. Li, C. Tian, Face sketch synthesis algorithm based on e-hmm and selective ensemble, IEEE Trans. Circ. Syst. Video Technol. 18 (4) (2008) 487–496.
[19] X. Wang, X. Tang, Face photo-sketch synthesis and recognition, IEEE Trans. Pattern Anal. Mach. Intell. 31 (11) (2008) 1955–1967.
[20] H. Zhou, Z. Kuang, K.-Y.K. Wong, Markov weight fields for face sketch synthesis, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2012, pp. 1091–1097.
[21] Y. Song, L. Bao, Q. Yang, M.-H. Yang, Real-time exemplar-based face sketch synthesis, in: European Conference on Computer Vision, Springer, 2014, pp. 800–813.
[22] C. Peng, X. Gao, N. Wang, J. Li, Superpixel-based face sketch–photo synthesis, IEEE Trans. Circ. Syst. Video Technol. 27 (2) (2015) 288–299.
[23] S. Zhang, X. Gao, N. Wang, J. Li, Robust face sketch style synthesis, IEEE Trans. Image Process. 25 (1) (2015) 220–232.
[24] N. Wang, X. Gao, L. Sun, J. Li, Bayesian face sketch synthesis, IEEE Trans. Image Process. 26 (3) (2017) 1264–1274.

[25] M. Zhang, R. Wang, X. Gao, J. Li, D. Tao, Dual-transfer face sketch–photo synthesis, IEEE Trans. Image Process. 28 (2) (2018) 642–657.
[26] L. Zhang, L. Lin, X. Wu, S. Ding, 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, 2015, pp. 627–634.
[27] B. Sheng, P. Li, C. Gao, K.-L. Ma, Deep neural representation guided face sketch synthesis, IEEE Trans. Visualiz. Comput. Graph. 25 (12) (2018) 3216–3230.
[28] D. Zhang, L. Lin, T. Chen, X. Wu, W. Tan, E. Izquierdo, Content-adaptive sketch portrait generation by decompositional representation learning, IEEE Trans. Image Process. 26 (1) (2016) 328–339.
[29] J. Yu, S. Shi, F. Gao, D. Tao, Q. Huang, Composition-aided face photo-sketch synthesis, arXiv preprint arXiv:1712.00899..
[30] S. Zhang, R. Ji, J. Hu, Y. Gao, C.-W. Lin, Robust face sketch synthesis via generative adversarial fusion of priors and parametric sigmoid, IJCAI (2018) 1163–1169.
[31] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Advances in neural information processing systems, 2014, pp. 2672–2680..
[32] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al., Photo-realistic single image super-resolution using a generative adversarial network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4681–4690.
[33] Z. Xu, M. Wilber, C. Fang, A. Hertzmann, H. Jin, Learning from multi-domain artistic images for arbitrary style transfer, arXiv preprint arXiv:1805.09987..
[34] X. Yu, Y. Qu, M. Hong, Underwater-gan: Underwater image restoration via conditional generative adversarial network, in: International Conference on Pattern Recognition, Springer, 2018, pp. 66–75..
[35] W. Xian, P. Sangkloy, V. Agrawal, A. Raj, J. Lu, C. Fang, F. Yu, J. Hays, Texturegan: controlling deep image synthesis with texture patches, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 8456–8465.
[36] P. Isola, J.-Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1125–1134.
[37] A. Odena, C. Olah, J. Shlens, Conditional image synthesis with auxiliary classifier gans, in: International Conference on Machine Learning, 2017, pp. 2642–2651.
[38] N. Wang, W. Zha, J. Li, X. Gao, Back projection: an effective postprocessing method for gan-based face sketch synthesis, Pattern Recogn. Lett. 107 (2018) 59–65.
[39] F. Gao, S. Shi, J. Yu, Q. Huang, Composition-aided sketch-realistic portrait generation, arXiv preprint arXiv:1712.00899..
[40] S. Zhang, R. Ji, J. Hu, X. Lu, X. Li, Face sketch synthesis by multidomain adversarial learning, IEEE Trans. Neural Networks Learn. Syst. 30 (5) (2018) 1419–1428.
[41] M. Zhu, N. Wang, X. Gao, J. Li, Z. Li, Face photo-sketch synthesis via knowledge transfer, IJCAI (2019) 1048–1054.
[42] J.-Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2223–2232.
[43] L. Wang, V. Sindagi, V. Patel, High-quality facial photo-sketch synthesis using multi-adversarial networks, in: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), IEEE, 2018, pp. 83–90..
[44] C. Chen, W. Liu, X. Tan, K.-Y. K. Wong, Semi-supervised learning for face sketch synthesis in the wild, in: Asian Conference on Computer Vision, Springer, 2018, pp. 216–231..
[45] M. Zhu, J. Li, N. Wang, X. Gao, A deep collaborative framework for face photo–sketch synthesis, IEEE Trans. Neural Networks Learn. Syst. 30 (10) (2019) 3096–3108.
[46] K. Sun, B. Xiao, D. Liu, J. Wang, Deep high-resolution representation learning for human pose estimation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 5693–5703.
[47] Z. Xu, X. Yang, X. Li, X. Sun, The effectiveness of instance normalization: a strong baseline for single image dehazing, arXiv preprint arXiv:1805.03305..
[48] C. Li, M. Wand, Combining markov random fields and convolutional neural networks for image synthesis, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2479–2486.
[49] P. Kaur, H. Zhang, K. Dana, Photo-realistic facial texture transfer, in: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, 2019, pp. 2097–2105..
[50] H. Zhang, K. Dana, Multi-style generative network for real-time transfer, in: Proceedings of the European Conference on Computer Vision (ECCV), 2019, pp. 349–365.
[51] W. Zhang, X. Wang, X. Tang, Coupled information-theoretic encoding for face photo-sketch recognition, in: CVPR 2011, IEEE, 2011, pp. 513–520..
[52] D. P. Kingma, J. Ba, Adam: a method for stochastic optimization, arXiv preprint arXiv:1412.6980..
[53] L. Zhang, L. Zhang, X. Mou, D. Zhang, Fsim: a feature similarity index for image quality assessment, IEEE Trans. Image Process. 20 (8) (2011) 2378–2386.
[54] D.-P. Fan, S. Zhang, Y.-H. Wu, Y. Liu, M.-M. Cheng, B. Ren, P.L. Rosin, R. Ji, Scoot: a perceptual metric for facial sketches, in: Proceedings of the IEEE International Conference on Computer Vision, 2019, pp. 5612–5622.
[55] Z. Liu, P. Luo, X. Wang, X. Tang, Deep learning face attributes in the wild, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 3730–3738.
[56] Q. Cao, L. Shen, W. Xie, O.M. Parkhi, A. Zisserman, Vggface2: a dataset for recognising faces across pose and age, in: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), IEEE, 2018, pp. 67–74..

Weiguo Wan received the B.S. degree in mathematics and applied mathematics from Jiangxi Normal University, Nanchang, China, in 2014 and Ph.D. degree in computer science and engineering from Jeonbuk National University, Jeonju, South Korea, in 2020. He is currently a Lecturer with the School of Software and Internet of Things Engineering, Jiangxi University of Finance and Economics, Nanchang, China. His current research interests include computer vision, deep learning, face sketch synthesis and recognition, and remote sensing image fusion.

Yong Yang (M'13?SM'16) received the Ph.D. degree from Xi'an Jiaotong University, Xi'an, China, in 2005. From 2009 to 2010, he was a Post-Doctoral Research Fellow with Chonbuk National University, Jeonju, South Korea. He is currently a Full Professor and the Vice Dean with the School of Information Technology, Jiangxi University of Finance and Economics, Nanchang, China. His current research interests include image fusion and super resolution, image processing and analysis, and pattern recognition. He is a Senior Member of CCF. He received the title of Jiangxi Province Young Scientist in 2012 and was selected as the Jiangxi Province Thousand and Ten Thousand Talent in 2015.

Hyo Jong Lee (M'91) received the B.S., M.S., and Ph.D. degrees in computer science from The University of Utah, USA, where he was involved in computer graphics and parallel processing. He is currently a Professor with the Division of Computer Science and Engineering and the Director of the Center for Advanced Image and Information Technology, Chonbuk National University, Jeonju, South Korea. His research interests include image processing, medical imaging, parallel algorithms, deep learning, and brain science.