



# Article A Joint Training Model for Face Sketch Synthesis

# Weiguo Wan<sup>1</sup> and Hyo Jong Lee<sup>1,2,\*</sup>

- <sup>1</sup> Division of Computer Science and Engineering, Chonbuk National University, Jeonju 54896, Korea; wanwgplus@gmail.com
- <sup>2</sup> Center for Advanced Image and Information Technology, Chonbuk National University, Jeonju 54896, Korea
- \* Correspondence: hlee@chonbuk.ac.kr

Received: 18 March 2019; Accepted: 24 April 2019; Published: 26 April 2019



# Featured Application: The proposed face sketch synthesis method can be applied for various applications, such as law enforcement and digital entertainment.

Abstract: The exemplar-based method is most frequently used in face sketch synthesis because of its efficiency in representing the nonlinear mapping between face photos and sketches. However, the sketches synthesized by existing exemplar-based methods suffer from block artifacts and blur effects. In addition, most exemplar-based methods ignore the training sketches in the weight representation process. To improve synthesis performance, a novel joint training model is proposed in this paper, taking sketches into consideration. First, we construct the joint training photo and sketch by concatenating the original photo and its sketch with a high-pass filtered image of their corresponding sketch. Then, an offline random sampling strategy is adopted for each test photo patch to select the joint training photo and sketch patches in the neighboring region. Finally, a novel locality constraint is designed to calculate the reconstruction weight, allowing the synthesized sketches to have more detailed information. Extensive experimental results on public datasets show the superiority of the proposed joint training model, both from subjective perceptual and the FaceNet-based face recognition objective evaluation, compared to existing state-of-the-art sketch synthesis methods.

Keywords: face sketch synthesis; face sketch recognition; joint training model

# 1. Introduction

Face sketch synthesis is a key branch of face style transformation, which generates face sketches for given input photos with the help of face photo-sketch pairs as the training dataset [1]. It has achieved wide applications in both law enforcement and digital entertainment. For example, sketches drawn according to the description of victims or witnesses can help identify a suspect by matching the sketch against a mugshot dataset from a police department. Face sketch synthesis reduces the texture discrepancy between photos and sketches for the face recognition procedure [2] and thus increases the recognition accuracy [3]. In digital entertainment, people are increasingly preferring to use face sketches as their portrait in social media; the sketch synthesis technique can also simplify animation production [4].

During the past two decades, various sketch synthesis methods have been proposed. The exemplarbased method is an important category of existing synthesis approaches. It synthesizes sketches for test photos by utilizing photo-sketch pairs as training data. The exemplar-based method mainly consists of neighbor selection and reconstruction weight representation [5]. In the neighbor selection process, K nearest training photo patches are selected for a test photo patch. In the reconstruction weight representation, a weight vector between the test photo patch and the selected photo patches is calculated. The target sketch patch can be obtained using weighted averaging of the K training sketch patches corresponding to the selected photo patches with the calculated weight vector. The final sketch is obtained by averaging all the generated sketch patches.

Exemplar-based face sketch synthesis originates from the Eigen-transformation research by Tang et al. [6]. In their work, all training photo-sketch pairs were used to generalize the target sketch. Principal component analysis was adopted to learn the weight coefficients by projecting the input test photo onto the training photos.

It is difficult to represent nonlinear relationships between face photos and sketches by only learning one holistic reconstruction model. Thus, Liu et al. [7] proposed a locally linear embedding (LLE)-based sketch synthesis method to estimate the nonlinear mapping with piecewise linear mappings. The LLE method works at the image patch level, in which K nearest training photo patches are searched in terms of Euclidean distance for each test photo patch. However, the LLE method suffers from a serious noise problem. To resolve this problem, Song et al. [8] formulated face sketch synthesis into a spatial sketch denoising problem and calculated the reconstruction weight using the conjugate gradient solver.

To describe the dependency relationship between neighboring sketches, Wang et al. [9] introduced a multi-scale Markov random fields (MRF) model to represent the neighboring constraint between adjacent sketch patches using a compatibility function. However, this method only chose the best single sketch patch from the training data for the test photo patch, meaning it could not synthesize new sketch patches. Additionally, the optimization process in the MRF model is an NP-hard problem. To overcome these limitations, Zhou et al. [10] extended the MRF model by introducing the linear combination of nearest neighbors to structure a Markov weight fields (MWF) model, which is capable of synthesizing new sketch patches that do not exist in the training dataset. In reference [11], a sparse representation-based face sketch synthesis method was proposed by Gao et al. Peng et al. [12] proposed a multiple representation-based face sketch synthesis method, which is able to obtain high-quality sketch images. However, this method is time-consuming because of the online neighbor selection.

Recently, Wang et al. [13] proposed a state-of-the-art face sketch synthesis method, based on random sampling and locality constraint (RSLCR). They randomly sampled the training photo and sketch patches in place of a neighbor search, and then employed the locality constraint to model the distinct correlations between the test photo patch and sampled photo patches while calculating the reconstruction weight coefficients. However, the target sketch patch reconstruction obtained by weighted-averaging the hundreds of sampled sketches can be regarded as a low-pass filter process, which results in blurred synthesized sketches. In addition, the Bayesian inference was utilized in reference [14] to incorporate the neighboring constraint in both the neighbor selection and reconstruction weight representation, which can obtain impressive performance.

Apart from the exemplar-based method, deep learning techniques were also applied to face sketch synthesis, generating new trends. Zhang et al. [15] first proposed a fully convolutional network (FCN) to learn end-to-end mapping from photos to sketches. It consisted of six convolutional layers with rectified linear units as activation functions. Zhang et al. [16] utilized the branched FCN to structure a decomposition representation learning framework for sketch synthesis. Additionally, the generative adversarial network (GAN) [17] was developed for image style transformation (e.g., photo-to-sketch generation or vice versa). The deep learning-based sketch synthesis methods can preserve textural structure well; however, serious noise effects occur in the synthesized results. This is mainly because of the limited available of training data, which is insufficient to train large networks robustly [18].

A common problem with these exemplar-based approaches is that they ignore the role of training sketches when calculating reconstruction weights. This is because the basic assumption of these exemplar-based methods is that a photo patch and its corresponding sketch patch have a similar geometric manifold structure. If two photo patches are similar, then their sketch patch counterparts are also similar. However, owing to potential misalignment, the reconstruction weight obtained from the test photo patch and selected training photo patches may not be suitable for sketch patches reconstruction [19].

We propose a new exemplar-based method, the joint training model, to solve the problem. In our method, the training photo and the sketch patches are concatenated. Instead of directly using the sketch patches, the high-pass filtered component of the training sketches are adopted to reduce the effect of the modality difference between a photo and a sketch. Then, we employ the offline random sampling method to select the joint training photo and the sketch patches for the test photo patch. Moreover, a modified locality constraint is designed to calculate the reconstruction weight. With the obtained reconstruction weight, the target sketch patch can be synthesized. Experimental results indicate that the proposed joint model significantly eliminates noise and improves the synthesized sketch quality. It also preserves the detail information of the test photo, which other methods cannot do. Therefore, synthesized face sketch images using our method achieve higher accuracy in face sketch recognition.

The contributions of this paper are summarized as below.

- (1) To consider the training sketches during the reconstruction weight representation process, a joint training model is proposed to integrate the training photo and sketch information.
- (2) We design a modified locality constraint that modulates the reconstruction weight through the distance between the high-pass filtered images of test patches and the sampled training sketch patches.
- (3) The proposed method yields high quality sketches with more detail information and less noise over the wide range of datasets, promoting the accuracy of the sketch-based suspect identification.

The organization of the rest of the paper is as follows. Section 2 introduces the relevant works, including some example-based sketch synthesis methods, which are the basic works of the proposed method. The proposed model is described in detail in Section 3. Section 4 provides a comparison of experiments and their results. Conclusions are then given in Section 5.

#### 2. Technical Backgrounds

In this paper, excepted as noted, a bold uppercase letter and a bold lowercase letter represent a matrix and a column vector, respectively; regular uppercase and lowercase letters denote scalars. Given a test photo **T**, it is divided into patches  $\mathbf{t}^{(i,j)}$  with *r* pixels overlapping between neighboring patches. (i, j) denotes the location of the patch at the *i*-th row and the *j*-th column,  $i \in \{1, \dots, m\}$ ,  $j \in \{1, \dots, n\}$ . Notice that each patch is represented as a *q*-dimensional column vector, where  $q = p^2$ , and *p* is the size of the patch. Similarly, the target sketch is denoted as **S**.  $\mathbf{s}^{(i,j)}$  denotes the target sketch patch corresponding to the testing patch  $\mathbf{t}^{(i,j)}$ . The training dataset, which consists of *M* photo-sketch pairs, are similarly divided into patches. Let  $\mathbf{X}^{(i,j)} = {\{\mathbf{x}_k^{(i,j)}\}}_{k=1}^K$  and  $\mathbf{Y}^{(i,j)} = {\{\mathbf{y}_k^{(i,j)}\}}_{k=1}^K$  denote the set of *K* selected training photo patches and the corresponding sketch patches of the test photo patch  $\mathbf{t}^{(i,j)}$ , respectively. The weight coefficients  $\mathbf{w}^{(i,j)} = {\{\mathbf{w}_1^{(i,j)}, \dots, \mathbf{w}_K^{(i,j)}\}}^T$  are calculated to linearly combine the candidate sketch patches.

# 2.1. The LLE Method

In the LLE method [6], *K* nearest patches are first obtained for each test photo patch  $t^{(i,j)}$ . Then, the linear reconstruction coefficients  $w^{(i,j)}$  can be calculated by resolving the following minimization problem:

$$\min_{\mathbf{w}^{(i,j)}} \|\mathbf{t}^{(i,j)} - \mathbf{X}^{(i,j)} \mathbf{w}^{(i,j)}\|_2^2, \, s.t. \, \mathbf{1}^T \mathbf{w}^{(i,j)} = 1,$$
(1)

Then the target sketch patch  $\mathbf{s}^{(i,j)}$  can then be synthesized.

$$\mathbf{s}^{(i,j)} = \sum_{k=1}^{K} w_k^{(i,j)} y_k^{(i,j)} = \mathbf{Y}^{(i,j)} \cdot \mathbf{w}^{(i,j)},$$
(2)

After all target sketch patches are generated, the final sketch can be achieved by averaging the overlapped pixel intensities.

#### 2.2. The RSLCR Method

Instead of searching the nearest neighbor photo patches, the RSLCR method [13] proposed to randomly sample *K* photo-sketch patch pairs from training data in a predicted neighbor region for the test patch  $t^{(i,j)}$ . Additionally, to consider the correlation between different sampled patches, a locality constraint [20] was introduced to impose a weight to the distances of the test photo patch and random sampled photo patches. The reconstruction weight representation model of the RSLCR method can be written as follows.

$$\min_{\mathbf{w}^{(i,j)}} \|\mathbf{t}^{(i,j)} - \mathbf{X}^{(i,j)} \mathbf{w}^{(i,j)}\|_2^2 + \lambda \|\mathbf{d}^{(i,j)} \odot \mathbf{w}^{(i,j)}\|, s.t. \mathbf{1}^T \mathbf{w}^{(i,j)} = 1,$$
(3)

where  $\odot$  denotes element-wise multiplication,  $\lambda$  balances the reconstruction error and the locality constraint, and  $\mathbf{d}^{(i,j)}$  is the Euclidean distance vector between the test photo patch  $\mathbf{t}^{(i,j)}$  and sampled training photo patches  $\mathbf{X}^{(i,j)}$ .

## 3. Joint Training Model for Face Sketch Synthesis

In most exemplar-based face sketch synthesis methods, only the test photo and training photos are considered for selecting the candidate photo patches and calculating the reconstruction weight. This strategy cannot achieve an optimal result when the training photo and sketch are misaligned. In this paper, we put forward a novel exemplar-based face sketch synthesis approach, which takes the training sketch into account by joining it with its corresponding training photo for reconstruction weight representation.

Figure 1 shows the illustration of the proposed face sketch synthesis method. First, the high-pass filtered components of the sketch images in the training data are extracted by using the Laplacian of Gaussian (LoG) filter. Then, the extracted high-pass filtered components are attached to their corresponding photo and sketch images to form the joint training data. After that, the candidate joint patch pairs are sampled with the offline random sampling strategy. In the test phase, the high-pass filtered component of the input photo is extracted with the same filter and attached to the input photo to obtain the joint test photo. For each joint test patch, a reconstruction weight can be calculated by approximating the joint test patch and joint training sketches with the reconstruction weight and the final sketch image can be synthesized by averaging the overlapping sketch patches. The strategy to structure the joint training model and the way to calculate the reconstruction weight coefficients of the proposed method will be explained next in detail.



Figure 1. Illustration of the proposed joint model for face sketch synthesis.

# 3.1. Joint Training Model

For each training sketch, the corresponding high-pass filtered component is first obtained with a LoG filter. Let  $\mathbf{Z}^{(i,j)} = {\{\mathbf{z}_{k}^{(i,j)}\}}_{k=1}^{K}$  denote the set of *K* randomly selected high-pass filtered image patches of the training sketches corresponding to a test patch  $\mathbf{t}^{(i,j)}$ . Then, the sampled *K* joint training photo and sketch patches for  $\mathbf{t}^{(i,j)}$  can be denoted as Equations (4) and (5), respectively.

$$\mathbf{U}^{(i,j)} = \begin{pmatrix} \mathbf{X}^{(i,j)} \\ \mathbf{Z}^{(i,j)} \end{pmatrix} = \left\{ \begin{pmatrix} \mathbf{x}_{k}^{(i,j)} \\ \mathbf{z}_{k}^{(i,j)} \end{pmatrix} \right\}_{k=1}^{K},$$
(4)

$$\mathbf{V}^{(i,j)} = \begin{pmatrix} \mathbf{Y}^{(i,j)} \\ \mathbf{Z}^{(i,j)} \end{pmatrix} = \left\{ \begin{pmatrix} \mathbf{y}_{k}^{(i,j)} \\ \mathbf{z}_{k}^{(i,j)} \end{pmatrix} \right\}_{k=1}^{K},$$
(5)

For the test photo **T**, the high-pass image **H** is obtained with a LoG filter. Let  $\mathbf{h}^{(i,j)}$  denotes the high-pass image patch corresponding to the test photo patch  $\mathbf{t}^{(i,j)}$ . We concatenate these two patches as the joint test patch  $\mathbf{t}_1^{(i,j)}$ :

$$\mathbf{t}_{1}^{(i,j)} = \begin{pmatrix} \mathbf{t}^{(i,j)} \\ \mathbf{h}^{(i,j)} \end{pmatrix}.$$
 (6)

After modeling the joint training photos and sketches, the reconstruction weight can be calculated by approximating the joint test patch and joint training photos, which will be discussed next.

# 3.2. Face Sketch Synthesis

Assuming there are M pairs of joint training photos and sketches that are geometrically aligned, they are divided into patches of fixed size  $(p \times p \times 2)$ . Each joint patch is reshaped to a 2*q*-dimensional column vector. For each joint test patch location, we extend the search region around the patch with *c* pixels. Thus, there are  $(2c + 1)^2$  patches in the search region for one patch location, and there are  $(2c + 1)^2M$  joint training photo/sketch patch-pairs. Among these patch-pairs, *K* joint training photo patches  $\mathbf{U}^{(i,j)} \in \mathbb{R}^{2p^2 \times K}$  and joint training sketch patches  $\mathbf{V}^{(i,j)} \in \mathbb{R}^{2p^2 \times K}$  are randomly and simultaneously selected.

For each joint test patch  $t_1^{(i,j)}$ , the reconstruction weight is calculated as follow:

$$\min_{\mathbf{w}^{(i,j)}} \|\mathbf{t}_{1}^{(i,j)} - \mathbf{U}^{(i,j)}\mathbf{w}^{(i,j)}\|_{2}^{2} + \lambda \|\mathbf{d}_{1}^{(i,j)} \odot \mathbf{w}^{(i,j)}\| + \mu \|\mathbf{d}_{2}^{(i,j)} \odot \mathbf{w}^{(i,j)}\|, \quad s.t. \ 1^{T} \mathbf{w}^{(i,j)} = 1,$$
(7)

where  $\mathbf{w}^{(i,j)} \in \mathbb{R}^{K \times 1}$  is the weight coefficient for the joint test patch  $\mathbf{t}_1^{(i,j)}$ .  $\mathbf{d}_1^{(i,j)} \in \mathbb{R}^{K \times 1}$  is the Euclidean distance vector between the joint test patch  $\mathbf{t}_1^{(i,j)}$  and sampled joint training photo patches  $\mathbf{U}^{(i,j)}$ , and  $\mathbf{d}_2^{(i,j)} \in \mathbb{R}^{K \times 1}$  is the Euclidean distance vector between the LoG test patch image  $\mathbf{h}^{(i,j)}$  and *K* sampled LoG training sketch patches  $\mathbf{Z}^{(i,j)}$ .

Equation (7) has the closed-form solution:

$$\mathbf{aw}\boldsymbol{\nu}^{(i,j)} = (\mathbf{C}^{i,j} + \lambda \operatorname{diag}(\mathbf{d}_1^{(i,j)}) + \mu \operatorname{diag}(\mathbf{d}_2^{(i,j)})) \setminus 1, \tag{8}$$

$$\mathbf{w}^{(i,j)} = \mathbf{w}\boldsymbol{\prime}^{(i,j)} / \mathbf{1}^T \mathbf{w}\boldsymbol{\prime}^{(i,j)},\tag{9}$$

where **1** is a column vector in which all elements are 1.  $C^{i,j} = (\mathbf{U}^{(i,j)} - \mathbf{1t}_1^{(i,j)^T})(\mathbf{U}^{(i,j)} - \mathbf{1t}_1^{(i,j)^T})^T$  denotes the data covariance matrix, and diag(·) extends the vector into a diagonal matrix.

The target joint sketch patch  $\mathbf{s}^{(i,j)}$  can be synthesized by linearly combining the sampled joint training sketches with the reconstruction weight coefficients  $\mathbf{w}^{(i,j)}$ .

$$\mathbf{s}^{(i,j)} = \mathbf{V}^{(i,j)} \mathbf{w}^{(i,j)}.$$
(10)

The obtained joint sketch patch is a 2q-dimensional vector. Thus, we extract the first half and reshape it to a  $p \times p$  patch, which is the target sketch patch. After obtaining all target sketch patches, the final target sketch can be achieved with an averaged overlapping area.

# 4. Evaluation Experiments

## 4.1. Datasets

We evaluated the performance of the proposed method on two publicly available datasets: The Chinese University of Hong Kong (CUHK) face sketch (CUFS) dataset [9] and the CUHK face sketch FERET (CUFSF) dataset [21]. The CUFS dataset includes three sub-datasets: The CUHK student dataset (188 subjects) [22], the AR dataset (123 subjects) [23], and the XM2VTS dataset (295 subjects) [24]. The CUFSF dataset includes 1194 subjects from the FERET dataset [25]. Artist drew sketches corresponding to each face photo in these datasets. In our experiments, all face images were normalized into the size of 200×250 by centering the coordinates on two eyes and the mouth. Some face photo-sketch pairs from these two datasets are shown in Figure 2.



**Figure 2.** Example of face sketch-photo pairs in the CUFS dataset (first two rows) and the CUFSF dataset (last two rows). The first and the third row are face photos and the second and the last rows are corresponding face sketches drawn by the artist.

#### 4.2. Experimental Setting

In this section, the data distribution and parameter settings are introduced. For fair comparison, we employed the same cross validation technique used in most exemplar-based sketch synthesis works. For the CUFS dataset, 88 face photo-sketch pairs were taken as the training dataset and the remaining 100 pairs were taken for test in the CUHK student dataset; 80 pairs were chosen for training and the remaining 43 pairs for test in AR dataset; 100 pairs for training and the remaining 195 pairs for test in XM2VTS dataset. For the CUFSF dataset, 250 face photo-sketch pairs were randomly chosen for training and the rest of the 944 pairs for test.

Parameters were set as follow. Patch size was p = 20, overlap size was o = 14, search length was c = 5, the number for random sampling was K = 800, and the regularization parameter  $\lambda$  and  $\mu$  were both set to 0.5. The window size and standard deviation of LoG filter were 5 and 0.5, respectively.

To evaluate performance, four traditional exemplar-based methods (LLE [7], MRF [9], MWF [10], and RSLCR [13]) and two deep learning-based methods (FCN [15] and GAN [17]) were compared. These six state-of-the-art face sketch synthesis results are released by Wang et al. [13].

# 4.3. Synthesizesd Sketch Results Comparison

Figure 3 shows some synthesized face sketches from different methods on the CUFS dataset. The first two rows are from the CUHK student dataset, the middle two rows are from the AR dataset, and the last two rows are from the XM2VTS dataset. Moreover, the corresponding local blocks of the synthesized sketches are displayed in Figure 4. From Figure 3, it can be seen that the sketches synthesized by the LLE and MRF methods suffer serious block effects. The MWF method obtains better performance than the LLE and MRF methods in the CUHK student and AR datasets. However, the results are unsatisfying in the XM2VTS dataset, because it contains more face variations, such as aging, race, and hair styles. The RSLCR method generates fine textures and structures, because more candidate patches are incorporated via random sampling and the locality constraint. However, this method results in blurred outputs. The FCN and GAN methods overcome the blurring effect by using pixel-to-pixel mapping from a photo to a sketch. However, they tend to have undesirable artifacts because of instabilities in training, while generating high-resolution images. Our proposed method

achieves much better performance than these six benchmarked methods in all the three datasets. More detailed information is preserved in the synthesized sketches, such as the double-fold eyelids and hair grain. This illustrates that the proposed method is capable of generating identity-preserved sketches.



**Figure 3.** Synthesized sketches on the CUFS dataset by locally linear embedding (LLE) [7], Markov random fields (MRF) [9], Markov weight fields (MWF) [10], random sampling and locality constraint (RSLCR) [13], fully convolutional network (FCN) [15], generative adversarial network (GAN) [17], and our proposed method, respectively. Face photos in the first two rows are from the CUHK student dataset; second two rows are from the AR dataset; and the last two rows are from the XM2VTS dataset, respectively.



Figure 4. Local cut-out effect on Figure 3; same marshalling sequence as Figure 3.

We also investigated the robustness of the proposed method against shape exaggeration and illumination variations on the CUFSF dataset. Figure 5 shows some synthesized face sketches from different methods used on the CUFSF dataset. The block effect still exists in the LLE and MRF results. The MWF and RSLCR methods obtain similar performance on the CUFSF dataset. The FCN method suffers from serious noise and artifacts. The GAN results show much improvement, but distortion occurs in the synthesized sketches by this method. By comparison, the proposed method achieves the most vivid and clear sketches, reflecting the robustness of our proposed method.



**Figure 5.** Synthesized sketches on the CUFSF dataset by LLE [7], MRF [9], MWF [10], RSLCR [13], FCN [15], GAN [17], and the proposed method, respectively.

Overall, the synthesized face sketch images on the CUFS and CUFSF datasets by our method have the following superiorities compared to the benchmarked methods: (1) Less block artifacts and noise, because we calculated the target sketch patch by weighted averaging hundreds randomly sampled training sketch patches but not few nearest patches; (2) rich facial detail, due to the high-pass filtered components were adopted to build the joint training model, which is able to generate high-quality face sketch patches; (3) complete facial structures, as a result of the training sketch images were took into consideration when computing the reconstruction weights, which weakens the influence by the misalignment in training images.

Table 1 shows the average time consumption of different methods on different datasets. Here, only exemplar-based methods are compared, because it takes a very long time to train the neural network for deep learning-based methods, though the test time is quite fast once the model is trained. From Table 1, it can be seen that the LLE, MRF, and MWF methods have no scalability of training data. With the amplification of training data, the running time increased radically, such as the CUFSF dataset. The RSLCR and our proposed method were less susceptible to the size of training data, owing to the random sampling strategy. Although the proposed method is not the fastest, it still has comparable time consumption as the other methods.

Methods	LLE	MRF	MWF	RSLCR	Proposed
CUHK	536.34	8.60	16.10	18.79	20.80
AR	496.47	8.40	15.33	19.10	19.75
XM2VTS	642.50	10.40	18.80	18.14	20.78
CUFSF	1591.95	24.25	45.20	17.66	20.17

**Table 1.** Average running time (s) to generate one sketch by different methods.

# 4.4. Face Sketch Recognition

Face sketch recognition is commonly used to quantitatively evaluate the face sketch synthesis methods and to collectively compare the synthesized sketch images [8,15,26]. A higher face sketch recognition rate means that the corresponding sketch synthesis method is more effective and the synthesized sketch images are better. In this study, FaceNet [27] was employed to conduct the face sketch recognition experiments. To demonstrate the recognition performance of the synthesized sketches using our proposed method, we used the sketches synthesized using different methods as probe images to match the gallery images, consisting of the corresponding artist-drawn sketches. The 338 synthesized sketches in the CUFS dataset were taken as the probe set, and the corresponding ground-truth sketches drawn by the artist were taken as the gallery set. For the CUFS dataset, 944 synthesized sketches were taken as the probe set and the corresponding sketches drawn by the artist were taken as the gallery set.

Figure 6 shows the face sketch recognition accuracies of FaceNet on the CUFS and the CUFSF dataset, respectively. The proposed method achieved the best accuracy on both datasets, 97.04% in CUFS and 87.18% in CUFS, at rank-100, respectively. Table 2 shows the rank-1, rank-5, and rank-10 recognition rates, where rank-*n* measures the accuracy of the top-*n* best matches. The FCN and GAN methods got higher recognition accuracy on CUFS dataset, but similar accuracy on CUFSF dataset compared with the traditional exemplar-based methods. It indicates that the performance of deep learning-based methods degrades when the dataset has challenging variations. The synthesized sketches of our method obtained the highest rate in rank-1, rank-5, and rank-10. The recognition results indicated that the better generated texture features and more detailed information preserved by the proposed method.



Figure 6. Face sketch recognition accuracies of FaceNet on the CUFS (a) and the CUFSF (b) datasets.

Methods	CUFS			CUFSF			
	rank-1	rank-5	rank-10	rank-1	rank-5	rank-10	
LLE	38.7	63.6	72.8	11.3	26.3	36.2	
MRF	39.6	65.1	76.3	7.2	18.0	25.3	
MWF	47.3	68.0	77.5	10.6	23.6	31.5	
RSLCR	51.7	74.2	83.4	14.3	32.4	43.1	
FCN	51.4	76.3	84.0	11.4	26.4	35.2	
GAN	52.6	78.7	84.9	9.2	24.9	35.6	
Proposed	65.4	85.2	90.5	21.2	43.6	52.1	

Table 2. Recognition accuracies (%) of FaceNet on the CUFS and CUFSF datasets.

#### 5. Conclusions

Without considering the training sketches in the reconstruction weight representation process, the exemplar-based face sketch synthesis method had difficulty generating ideal results. This paper proposed a joint training model by concatenating the original training photos and sketches with high-pass filtered image patches of the training sketches. Additionally, we constructed a new locality constraint in the reconstruction weight process. With these improvements, more detailed information was preserved in the synthesized sketches. Experimental results demonstrated that the proposed method not only reduced the noise, but it also increased the definition of the synthesized sketches. Thus, the proposed joint training model is a practical and effective technique for face sketch synthesis. As analyzed and discussed previously, the deep learning-based face sketch synthesis methods are immature. In the future, a deep learning-based synthesis approach will be explored by employing more training data and designing optimized networks.

**Author Contributions:** Conception and design of the proposed method: H.J.L and W.W.; performance of the experiments: W.W.; writing of the paper: W.W.; paper review and editing: H.J.L.

**Funding:** This research was supported by "Research Base Construction Fund Support Program" funded by Chonbuk National University in 2018. This research was also supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (GR 2016R1D1A3B03931911). This study was also financially supported by the grants of China Scholarship Council (CSC No.2017 08260057).

Conflicts of Interest: The authors declare no conflict of interest

# References

- Wang, N.; Zhang, S.; Gao, X.; Li, J.; Song, B.; Li, Z. Unified framework for face sketch synthesis. *Signal Process*. 2017, 130, 1–11. [CrossRef]
- Li, J.; Yu, X.; Peng, C.; Wang, N. Adaptive representation-based face sketch-photo synthesis. *Neurocomputing* 2017, 269, 152–159. [CrossRef]
- Wan, W.; Lee, H.J. FaceNet Based Face Sketch Recognition. In Proceedings of the 2017 International Conference on Computational Science and Computational Intelligence, Las Vegas, NV, USA, 14–16 December 2017; pp. 432–436.
- Zhang, Y.; Wang, N.; Zhang, S.; Li, J.; Gao, X. Fast face sketch synthesis via KD-tree search. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherland, 8–16 October 2016; Springer: Cham, Switzerland, 2016; pp. 64–77.
- 5. Wang, N.; Zhu, M.; Li, J.; Song, B.; Li, Z. Data-driven vs. model-driven: Fast face sketch synthesis. *Neurocomputing* **2017**, 257, 214–221. [CrossRef]
- 6. Tang, X.; Wang, X. Face sketch recognition. IEEE Trans. Circuits Syst. Video Technol. 2004, 14, 1–7. [CrossRef]
- Liu, Q.; Tang, X.; Jin, H.; Lu, H.; Ma, S. A nonlinear approach for face sketch synthesis and recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, 20–25 June 2005; pp. 1005–1010.

- Song, Y.; Bao, L.; Yang, Q.; Yang, M. Real-time exemplar-based face sketch synthesis. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 5–12 September 2014; Springer: Cham, Switzerland, 2014; pp. 800–813.
- 9. Wang, X.; Tang, X. Face photo-sketch synthesis and recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 2009, 31, 1955–1967. [CrossRef] [PubMed]
- 10. Zhou, H.; Kuang, Z.; Wong, K. Markov weight fields for face sketch synthesis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Rhode Island, USA, 18–20 June 2012; pp. 1091–1097.
- 11. Gao, X.; Wang, N.; Tao, D.; Li, X. Face sketch–photo synthesis and retrieval using sparse representation. *IEEE Trans. Circuits Syst. Video Technol.* **2012**, *22*, 1213–1226. [CrossRef]
- 12. Peng, C.; Gao, X.; Wang, N.; Tao, D.; Li, X.; Li, J. Multiple representations-based face sketch–photo synthesis. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, *27*, 2201–2215. [CrossRef] [PubMed]
- 13. Wang, N.; Gao, X.; Li, J. Random sampling for fast face sketch synthesis. *Pattern Recognit.* **2018**, *76*, 215–227. [CrossRef]
- 14. Wang, N.; Gao, X.; Sun, L.; Li, J. Bayesian face sketch synthesis. *IEEE Trans. Image Process.* **2017**, *26*, 1264–1274. [CrossRef] [PubMed]
- Zhang, L.; Lin, L.; Wu, X.; Ding, S.; Zhang, L. End-to-end photo-sketch generation via fully convolutional representation learning. In Proceedings of the 5th ACM on International Conference on Multimedia Retrieval, Shanghai, China, 23–26 June 2015; pp. 627–634.
- 16. Zhang, D.; Lin, L.; Chen, T.; Wu, X.; Tan, W.; Izquierdo, E. Content-adaptive sketch portrait generation by decompositional representation learning. *IEEE Trans. Image Process.* **2017**, *26*, 328–339. [CrossRef] [PubMed]
- Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Proceedings of the International Conference on Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; MIT: Cambridge, MA, USA; pp. 2672–2680.
- Jiang, J.; Yu, Y.; Wang, Z.; Liu, X.; Ma, J. Graph-Regularized Locality-Constrained Joint Dictionary and Residual Learning for Face Sketch Synthesis. *IEEE Trans. Image Process.* 2019, 28, 628–641. [CrossRef] [PubMed]
- 19. Wang, N.; Gao, X.; Sun, L.; Li, J. Anchored neighborhood index for face sketch synthesis. *IEEE Trans. Circuits Syst. Video Technol.* **2018**, *28*, 2154–2163. [CrossRef]
- Wang, J.; Yang, J.; Yu, K.; Lv, F.; Huang, T.; Gong, Y. Locality-constrained linear coding for image classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 3360–3367.
- Zhang, W.; Wang, X.; Tang, X. Coupled information-theoretic encoding for face photo-sketch recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs, CO, USA, 21–25 June 2011; pp. 513–520.
- 22. Tang, X.; Wang, X. Face photo recognition using sketch. In Proceedings of the IEEE International Conference on Image Processing, New York, NY, USA, 22–25 September 2002; pp. 257–260.
- 23. Martinez, A.; Benavente, R. The AR Face Database; Technical Report; CVC: Barcelona, Spain, 1998.
- Messer, K.; Matas, J.; Kittler, J.; Luettin, J.; Maitre, G. XM2VTSDB: The extended M2VTS database. In Proceedings of the International Conference on Audio- and Video-Based Biometric Person Authentication, Washington, DC, USA, 22–24 March 1999; pp. 965–966.
- 25. Phillips, P.; Moon, H.; Rauss, P.; Rizvi, S. The FERET evaluation methodology for face recognition algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1090–1104. [CrossRef]
- Chen, C.; Tan, X.; Wong, K.K. Face sketch synthesis with style transfer using pyramid column feature. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision, Lake Tahoe, NV, USA, 12–15 March 2018; pp. 485–493.
- Schroff, F.; Kalenichenko, D.; Philbin, J. Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 8–12 June 2015; pp. 815–823.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).