Face Sketch Synthesis Using Regularized Broad Learning System

Ping Li, *Member, IEEE*, Bin Sheng, *Member, IEEE*, and C. L. Philip Chen, *Fellow, IEEE*

***Abstract*— There are two main categories of face sketch synthesis: data- and model-driven. The data-driven method synthesizes sketches from training photograph–sketch patches at the cost of detail loss. The model-driven method can preserve more details, but the mapping from photographs to sketches is a time-consuming training process, especially when the deep structures require to be refined. We propose a face sketch synthesis method via regularized broad learning system (RBLS). The broad learning-based system directly transforms photographs into sketches with rich details preserved. Also, the incremental learning scheme of broad learning system (BLS) ensures that our method easily increases feature mappings and remodels the network without retraining when the extracted feature mapping nodes are not sufficient. Besides, a Bayesian estimation-based regularization is introduced with the BLS to aid further feature selection and improve the generalization ability and robustness. Various experiments on the CUHK student data set and Aleix Robert (AR) data set demonstrated the effectiveness and efficiency of our RBLS method. Unlike existing methods, our method synthesizes high-quality face sketches much efficiently and greatly reduces computational complexity both in the training and test processes.**

***Index Terms*— Bayesian estimation, face sketch synthesis, incremental learning, regularized broad learning system (RBLS).**

I. Introduction

FACE sketch synthesis has been widely applied in digital entertainment and law enforcement [1]–[3]. In terms of digital entertainment, everyone can become a painter and get easy access to sketches that are originally drawn by skilled artists. Another application is for assisting law enforcement. As the photographs of a criminal suspect are not available in most cases, the best way to find the suspect in a mug-shot database is to match sketches, which are drawn by skilled artists with the aid of eyewitnesses, with photographs. However, due to the different modalities between photographs and sketches as shown in Fig. 1, face sketch recognition encounters great difficulty. Thus, the solution is to transform them into the same modality.

(a) (b) (c) (d)

Fig. 1. Examples of face photographs and sketches. (a) and (b) Photographs. (c) and (d) Corresponding sketches drawn by artists.

Existing face sketch synthesis methods can be divided into two categories: data-driven methods and model-driven methods [4]. Data-driven methods search similar sketch patches from the training photograph–sketch pairs and combine them via weights to get a new sketch of a test photograph. Tang and Wang [5], [6] employed the principal component analysis (PCA) to reconstruct sketches through a linear combination of training sketches. Liu *et al.* [7] took the nonlinear mapping from photographs to sketches to synthesize face sketches based on locally linear embedding (LLE). Wang *et al.* [8] proposed a Bayesian framework for face sketch synthesis and divided the synthesis process into two parts: the neighbor selection model and the weight computation model. However, their method still suffered from time-consuming computation as complex optimization problems. Besides, some necessary preprocessing/postprocessing, such as image blocking and patch aggregation, have to be employed in the synthesis [9]–[11]. Moreover, some details in the test photograph, which do not exist in the training set may lost. All these factors limit the practical efficiency of data-driven methods.

Model-driven methods learn a mapping function from the training photograph–sketch pairs. They learn the conversion model between two different modalities and can directly transform photographs to sketches in batch with the trained model. Wang *et al.* [12] proposed to learn some linear regressors from training photograph–sketch pairs. Zhang *et al.* [13] used a fully convolutional network (FCN) to learn the end-to-end nonlinear photograph–sketch mapping. Wang *et al.* [4] utilized a backprojection generative adversarial

Manuscript received January 23, 2020; revised September 3, 2020 and January 14, 2021; accepted March 30, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61872241 and Grant 61572316; in part by the National Key Research and Development Program of China under Grant 2019YFB1703600; and in part by The Hong Kong Polytechnic University under Grant P0030419, Grant P0030929, and Grant P0035358. *(Ping Li and Bin Sheng contributed equally to this work.) (Corresponding author: Bin Sheng.)*

Ping Li is with the Department of Computing, The Hong Kong Polytechnic University, Hong Kong (e-mail: p.li@polyu.edu.hk).

Bin Sheng is with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: shengbin@sjtu.edu.cn).

C. L. Philip Chen is with the School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China, also with the Navigation College, Dalian Maritime University, Dalian 116026, China, and also with the Faculty of Science and Technology, University of Macau, Macau 999078, China (e-mail: philip.chen@ieee.org).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TNNLS.2021.3070463.

Digital Object Identifier 10.1109/TNNLS.2021.3070463

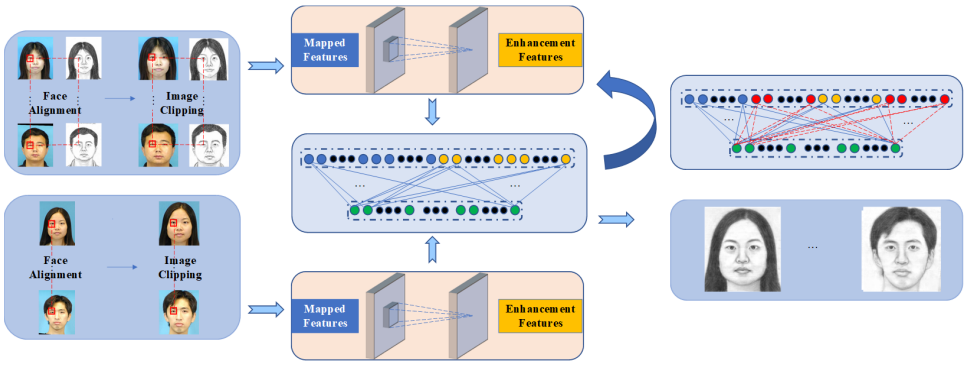


Fig. 2. Overall framework of the proposed method. Face alignment and image clipping are necessary preprocess operations to fit for the system. Random convolution is employed to form mapped features, and features are improved to obtain enhancement features. The input layer of the flat network consists of feature nodes (blue nodes) and enhancement nodes (orange nodes), constructing by concatenation of mapped features and enhancement features, respectively. All nodes are directly connected with target sketches. Once incremental nodes (red nodes) are necessary to improve the network, incremental learning is implemented. The regularized algorithm is used to further enhance the generalization. After training, face photographs can be transformed into sketches.

network (BP-GAN) to perform the synthesis tasks to improve the results. While deep neural networks have shown great capability in feature extraction and function approximation, they suffer from a time-consuming training process using backpropagation methods due to the deep structure. Moreover, the hyperparameters of network structures, such as numbers of layers, number of filters, and size of filters, have to be selected artificially by prior knowledge before training. Thus, if the structure is insufficient for modeling, it has to be remodeled and a complete retraining is needed. Recently, new advancements are made to enhance the face sketch synthesis. Zhu *et al.* [14] proposed a face sketch synthesis using deep graphical feature learning. Chen *et al.* [15] introduced a semisupervised deep learning to deal with face photographs in the wild. Wang *et al.* [16] produced face sketches in an adversarial manner. Wang *et al.* [17] proposed an anchored neighborhood index method to utilize the training face sketch patches. Wang *et al.* [18] further introduced a random sampling with locality constraint for fast face sketch synthesis. Zhu *et al.* [19] proposed a deep collaborative framework for face photograph–sketch synthesis. Radman and Suandi [20] enhanced the facial structures in face sketch synthesis via a superpixel-wise method.

To solve the problems of long training and retraining in deep learning, inspired by the random vector functional-link neural network (RVFLNN) and the dynamic step-wise updating algorithm [21], Chen *et al.* [22], [23] recently proposed a broad learning system (BLS) to provide an alternative to deep structures. The system has been demonstrated great performance in classification and regression [22]–[24].

We propose a face sketch synthesis method via BLS [22], [23] with a flat network, and the system is refined to adapt to the face sketch synthesis task by a regularized algorithm, which we call regularized BLS (RBLS). The proposed overall face sketch synthesis framework is shown in Fig. 2. A feedforward face sketch synthesis network (see Fig. 3), which is a one-layer structure with the input layer consisting of feature nodes and enhancement nodes, is trained by the system. Face photographs and training sketches are aligned and clipped to fit the BLS. A nonlinear conversion model from photographs to sketches is learned by training photograph–sketch pairs based on RBLS. After training, photographs can be transformed into sketches by the trained model directly. The modeling process can be divided into two steps: feature extraction and weight computation. Preprocessed training photographs are extracted face features by several randomly generated convolutional filters to form feature nodes in the input layer. Then, the network is extended by enhancement nodes which we call feature enhancement. Ridge regression learning is introduced to analytically compute initial connecting weights between the input feature layer and the output sketch layer. An incremental learning algorithm for increment feature mappings is provided for fast remodeling when extra features are necessarily added to improve the performance. Finally, a regularized algorithm is applied to improve the generalization ability and robustness and get the final trained network. Instead of the long training and complex optimization in deep learning, we use the flat network of RBLS to extract features and get the mapping model from features space of photographs to sketches via a very short training process. Combining with incremental learning, the network structure is improved without a retraining process by increasing feature mappings. The regularized algorithm used is for further feature selection and makes the output weights smooth and sparse for generalization ability and robustness improvement. In sum, our approach makes the following four main contributions.

1) *A Feedforward Face Sketch Synthesis Network via RBLS:* We propose a model-driven face sketch synthesis method. A feedforward sketch synthesis network via RBLS is trained, by which photographs can be directly transformed into sketches in batch with more details preserved. Experiments show that our approach outperforms most data- and model-driven synthesis methods.
2) *Incremental Learning for Fast Remodeling:* Incremental learning algorithm for increment feature mappings is proposed. When the extracted feature mapping nodes are not sufficient for the desired quality of synthesis sketches, increasing feature mappings and remodeling

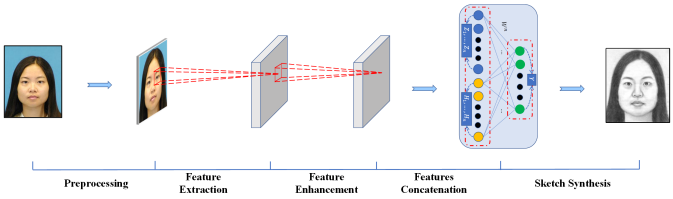


Fig. 3. Feedforward face sketch synthesis network via RBLS. The input face photograph is preprocessed and extracted features. Then, the mapped features are improved as enhancement features. Mapped features are concatenated as feature nodes and enhancement features are concatenated as enhancement nodes. All nodes are taken as input of the flat network and connected with the target sketch by connecting weights W^n.

the network are to be implemented without a retraining process.

3) *Regularized Algorithm for Generalization Ability and Robustness Improvement:* A regularized method via Bayesian estimation is proposed for further feature selection and to improve generalization ability and robustness.
4) **Time-Saving for Modeling/Remodeling:** Taking advantage of flat network, our method significantly shortens training and test time and greatly improves efficiency. Combining with incremental learning, the conversion network can be easily improved and fast remodeled.

II. Related Work

A. Data-Driven Face Sketch Synthesis

The conventional face sketch synthesis methods are data-driven, which linearly combine candidate sketch patches selected from the training sketches to synthesize a sketch of a test photograph. Tang and Wang [5], [6] first proposed to synthesize sketches by edge transformation based on the assumption of linear mapping from photographs to sketches. However, the whole linear assumption limited the ability to represent the nonlinear mapping from photographs to sketches. Liu *et al.* [7] presented a nonlinear approach via local geometry preserving to generate sketches, and a patch-based strategy is adopted. Gao *et al.* [25] proposed to adaptively select the number by sparse representation. Zhang *et al.* [26] converted photographs to arbitrarily stylistic sketches based on only one corresponding template sketch. Zhang *et al.* [27] proposed to handle nonfacial factors with a single photograph–sketch pair from coarse to fine. Zhang *et al.* [28] synthesized face sketches based on prior knowledge and similarity between different image patches. Considering the similarity between adjacent patches, Wang and Tang [29] introduced Markov random fields (MRF) into face sketch synthesis and employed belief propagation to obtain face sketches. Zhang *et al.* [30] proposed a face sketch generator via the compositional model (CM). They imitated the drawing process of artists and decomposed a face into components instead of patches. Wang *et al.* [8] presented a Bayesian face sketch synthesis framework consisting of two parts, the neighbor selection model and the weight computation model, and unified the existing data-driven methods. For the neighbor selection model, based on whether or not considering the neighboring constraint between adjacent patches, they divided the existing methods into two categories: one is neglecting the constraint, representative methods being K-nearest neighbors (K-NN) search-based methods [7] and sparse representation methods [25], [31], [32]; the other is based on the MRF selection strategy [29], which takes the spatial neighboring constraint into consideration. Similar to the neighbor selection model, they also made a division for the weight computation model. Although data-driven methods have shown good performance, they are limited in practical use. There are three main drawbacks in data-driven methods: high computation complexity, complicated preprocess/postprocess, and loss of particular detailed information. First, many complex optimization problems, such as belief propagation and least square, have to be solved in the whole synthesis process. Among them, the most time-consuming part is the neighbor selection process, especially performing neighbor searching on a large training data set. Second, necessary preprocessing/postprocessing, such as image blocking and patch aggregation, has to be employed. Finally, all of these methods are based on the assumption that a test photograph can be reconstructed from training data so that the sketch can be estimated by training data. When the training samples are insufficient to reconstruct the test photograph, loss of detailed information occurs.

B. Model-Driven Face Sketch Synthesis

Data-driven methods have some limitations that their computation complexity grows linearly with the size of the training set as they have to traverse the whole training set to find nearest neighbors. To solve this problem, Wang *et al.* [12] proposed to learn a mathematical model from training photograph–sketch pairs using a model-driven strategy that sped the synthesis process a lot compared with data-driven methods. However, the linear assumption has limitations in learning the nonlinear mapping from photographs to sketches. Recently, convolutional neural networks (CNNs) have demonstrated outstanding capability in image style transfer [33]–[36] and nonlinear mapping. Zhang *et al.* [13] developed an FCN to learn the nonlinear mapping from photographs to sketches. Jiao *et al.* [9] proposed a lightweight four-layer CNN method. Though most details are preserved, the blurring effect happens in the final results, and some textures and shadings are lost. Chen *et al.* [37] proposed a pyramid column feature to introduce sketch textures and shadings into the content image created by a fully CNN.

Apart from CNN-based methods, generative adversarial networks (GANs) have also made great contributions to image generation. Philip and Jong [38] proposed a conditional GAN face sketch synthesis framework and demonstrated the great potential of GAN for face sketch synthesis. However, the sketches are generated accompanying noise. Wang *et al.* [4] proposed a BP-GAN-based face sketch synthesis to denoise and synthesize high-quality sketches. Bi *et al.* [39] designed a three-layer pyramid model for getting multiscale information and trained the mapping relationships

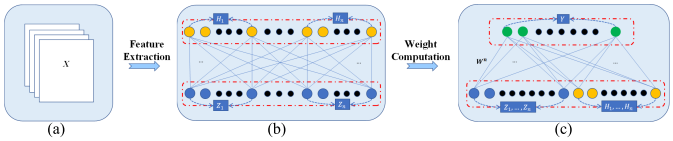


Fig. 4. Structure of BLS. The input data are extracted features to construct feature nodes $Z^n = [Z_1, \ldots, Z_n]$ by $Z_i = \phi(XW_{e_i} + \beta_{e_i})$. Then, the network is extended by enhancement nodes $H^n = [H_1, \ldots, H_n]$, where $H_i = \xi(Z_i W_{h_i} + \beta_{h_i})$. (c) Final flat network structure for BLS, whose input layer consists of feature nodes and enhancement nodes and output layer is target results Y. The input layer and output layer are connected by weight parameters W^n as $W^n = [Z^n|H^n]^\dagger Y$, where † denotes pseudoinverse. (a) Input date. (b) Nodes contribution for input layer. (c) Flat networks structure for BLS.

with a multiscale conditional GAN achieving nice Scoot score [40]. Yi *et al.* [41] proposed a novel GAN-based architecture named APDrawingGAN that builds upon hierarchical generators and discriminators. Inspired by knowledge transfer, Zhu *et al.* [42] learned knowledge of face photographs and face sketches separately by training two teacher networks on a large amount of data in related tasks simultaneously. Then, they used a student network to mimic and transform the obtained two kinds of knowledge and transfer them mutually. The proposed method can well synthesis the face sketch even with a small set of photograph–sketch pairs. Zhang *et al.* [30] introduced a CM-based face sketch generator, which nicely reduced high-frequency loss and enhanced the sketch synthesis results. Zhang *et al.* [43] further proposed dual-transfer face sketch–photograph synthesis framework, which generated more identifiable facial structures with higher recognition rate. Zhang *et al.* [44] again presented a neural probabilistic graphical model for face sketch synthesis, which could obtain specific features effectively and recover common structures in synthesized sketches. All these methods are based on deep structures, which means time-consuming training with error backpropagation. Moreover, a completely retraining process occurs when the deep structure is insufficient to learn the model.

C. Broad Learning System

Most model-driven methods are based on deep learning methods. However, due to the deep structure and abundant connecting weights, these methods suffered from time-consuming training processes. Pao and Takefuji [45] proposed an RVFLNN to solve the long training process problem in deep structure and demonstrated its generalization capability in function approximation [46]. They moved down hidden nodes in the traditional single-layer feedforward neural network (SLFN) as enhancement nodes and connected output nodes with input nodes and enhancement nodes. This simple improvement makes a direct connection between input and output and significantly improves the performance of the network. Taking inspiration from RVFLNN and adopting mapped features as its input, Chen *et al.* [22], [23], [47] proposed a BLS to offer an alternative way of deep learning. The structure of BLS is shown in Fig. 4. Instead of deep structures, the system is established as a flat network whose input layer consists of two parts, feature nodes generated from the original inputs, and enhancement nodes generated from feature nodes to expand the structure [see Fig. 4(b)]. The weights and biases to generated feature nodes and enhancement nodes are generated randomly. Then, all nodes are connected with outputs through connecting weights [see Fig. 4(c)]. Ridge regression learning algorithms are given to solve the connecting weights within a single learning step. In addition, incremental learning algorithms are designed for fast remodeling without retraining process when encountering increasing nodes and input data. Inspired by their work, we propose a feedforward face sketch synthesis network via an RBLS. Taking full advantage of the flat structure of BLS, the proposed network can be constructed at a fast speed. Meanwhile, the trained network can be demonstrated a better performance in face sketch synthesis with more detailed information preserved. Incremental learning for increment feature mapping is developed to improve the performance and for fast remodeling. Also, the regularized algorithm applied refines the system and improves the generalization ability and robustness.

III. Overview

Given a face photograph x, to estimate its corresponding sketch y, we first learn a feedforward face sketch synthesis network from training photograph–sketch pairs based on RBLS. Taking advantage of the great ability in feature extraction and function approximation of the network, the trained network can approximate a nonlinear function $F(\cdot)$, which maps photographs to sketches. Then, the target sketch y can be estimated by $y = F(x)$. The overall framework of the proposed method is shown in Fig. 2. A feedforward face sketch synthesis network (as shown in Fig. 3) based on a flat network is trained. As we can see, unlike the conventional RVFLNN taking the input photograph image directly, it takes mapped features as the input of the network. First, input face photographs X are mapped to construct n feature maps $Z^n = [Z_1, \ldots, Z_n]$ (denoted by blue nodes called feature nodes in Figs. 2 and 3). Then, the mapped features are enhanced as enhancement nodes $H^n = [H_1, \ldots, H_n]$ (denoted by orange nodes in Figs. 2 and 3). Finally, the feature nodes and enhancement nodes are concatenated together and connected with the output sketches Y through connected weights W^n. The mapping from the mapped features to the outputs can be denoted as

$$Y = [Z^n|H^n]W^n. \tag{1}$$

The proposed feedforward face sketch synthesis network is trained to learn the connecting weights W^n. To make W^n smooth and sparse, a regularized algorithm based on the Bayesian estimation is introduced to refine the system. Then, the total system approximates the nonlinear function $F(\cdot)$, which maps face photographs to sketches.

The learning process can be divided into two steps: 1) feature extraction and 2) weight computation. In this article, we employ random feature extraction and feature enhancement using convolution kernels generated randomly. After the necessary preprocessing operations such as face alignment and image clipping, all training face photographs are convoluted with randomly generated convolution kernels to obtain mapped

features. Next, the mapped features are improved by random convolution once again to form enhancement features. All the mapped features are concatenated to construct feature nodes and so are the enhancement features to construct enhancement nodes. Then, these two types of nodes are concatenated together and taken as the input of the flat network and are connected with target sketches directly. Initial connecting weights W^n can be solved by the error backpropagation algorithm. Another more direct solution without iteration is to compute pseudoinverse by the ridge regression learning algorithm. More details are shown in Section IV.

Once the network is insufficient to learn the mapping from face photographs to sketches that are usually determined by numbers of feature nodes and enhancement nodes, hyperparameters adjustment and remodel are necessary. In general, for deep structures, when the network parameters are adjusted and a remodeling process occurs, they will encounter a complete retraining process that consumes much time. However, for the proposed method, an incremental learning algorithm for increment feature mapping is given to avoid the retraining process and achieve fast remodel. More details are shown in Section V. Through the standard BLS above, we can get initial connecting weights. However, as noise exists and the small size of training data, the system has to be further refined for better generalization ability and robustness. Recent success in pattern classification and recognition has demonstrated that the Laplacian error distribution could be more realistic [48]–[50], which can fit long tails caused by small fraction outliers better [51], [52]. Therefore, we use the Laplacian assumption to model errors realistically [48]–[52], and the l_1 and l_2 regularizers are combined to form the loss function based on the Bayesian estimation [52]. Then, the system is optimized by the optimization function. More details are shown in Section VI.

IV. Face Sketch Synthesis Using BLS

In this section, details of face sketch synthesis by BLS are given. The learning process can be divided into two steps: feature extraction and weight computation. First, all of the training photograph–sketches pairs and test photographs are preprocessed to fit for the BLS. Then, random convolution feature extraction is employed for face feature extraction and feature improvement. Finally, the connecting weights are solved by ridge regression. Then, we get the initial connecting weights for the synthesis network.

A. Preprocessing

Given training photograph–sketch pairs, we first make two eye centers of all face images be at a fixed position by translating, rotating, and scaling, namely face alignment. As described in [29], this simple face alignment step arranges the same face components to the same region in different images, making it robust to map the photographs to sketches. Another preprocessing step takes its inspiration from [13]. We crop original images with a size of 200×250 into images with a size of 180×200 to throw away the blank regions. Experiments show that this step can effectively improve the quality of generated sketches.

B. Feature Extraction

Traditional RVFLNN takes original inputs directly as the input of the first layer in the network. Another feasible choice is to take feature mapping as input. As mentioned in Section III, BLS first maps inputs into feature spaces which we call feature mapping or feature extraction. Feature extraction consists of two steps: feature nodes generation and enhancement nodes generation. Assume an input face photograph x, which is preprocessed, with a size of 180×200. Features of the input photograph are extracted and enhanced by randomly generated convolution kernels to form feature nodes and enhancement nodes as the input of the network. For n feature mappings, z_i denotes the ith feature map of x, which can be represented as

$$z_i = \phi(x \otimes w_{f_i} + \beta_{f_i}), \quad i = 1, \ldots, n \tag{2}$$

where w_{f_i} and β_{f_i} are the randomly generated filter and bias corresponding to the ith feature map, respectively. $\phi(\cdot)$ is an activation function. $\otimes$ represents the convolution operation. Feature nodes are constructed by concatenating all the n mapping features. We denote all the feature nodes as

$$z^n \equiv [z_1, \ldots, z_n]. \tag{3}$$

For each feature mapping, we enhance it as an enhancement mapping by

$$h_j = \xi(z_j \otimes w_{e_j} + \beta_{e_j}), \quad j = 1, \ldots, n \tag{4}$$

where w_{e_j} and β_{e_j} are the randomly generated filter and bias similar to w_{f_i} and β_{f_i}, respectively. $\xi(\cdot)$ is a activation function like $\phi(\cdot)$. All of the n enhancement feature maps are concatenated to construct the enhancement nodes as

$$h^n \equiv [h_1, \ldots, h_n]. \tag{5}$$

Then, feature nodes and enhancement nodes are concatenated together to construct the input layer of the flat network of BLS. This step is equal to moving down the hidden layer of RVFLNN, which means connecting its input layer with the output layer, expanding the structure of BLS in the wide sense. The final broad model can be denoted as

$$y = [z^n | h^n] W^n \tag{6}$$

where W^n are connecting weights for the synthesis network and y is the target sketch of the input face photograph x.

C. Weight Computation

Giving training photograph–sketch pairs, we first map the face photographs to feature spaces and construct feature nodes and enhancement nodes. Meantime, the randomly generated filters and biases are preserved for testing. The connecting weights W^n can be solved under the constraint that the generated sketch $F(x; W^n)$ by BLS should be as close to the ground truth which is drawn by artists. Let (x_i, y_{t_i}) denote the training photograph–sketch pairs. For N pairs of training images, the weight computation can be represented as the following optimization:

$$\min_{W^n} \frac{1}{N} \sum_{i=1}^{N} \left\| F(x_i; W^n) - y_{t_i} \right\|^2 \tag{7}$$

where we adopt the mean squared error (MSE) as the loss function. Gradient descent can be applied to train the network. While iterative computation is needed when implementing the gradient of the loss function, a straightforward method is to compute pseudoinverse and solve the problem in a single step. Let $X = [x_1^T, \ldots, x_N^T]^T$ and $Y = [y_{t_1}^T, \ldots, y_{t_N}^T]^T$ denote the concatenation of the N training photographs and training sketches, respectively. We rewrite (6) as

$$Y = [Z^n | H^n] W^n \tag{8}$$

where $Z^n = [(z_1^n)^T, \ldots, (z_N^n)^T]^T$ is the concatenation of all features nodes for the N face photographs and $H^n = [(h_1^n)^T, \ldots, (h_N^n)^T]^T$ is the concatenation of all enhancement nodes. The connecting weights W^n can be solved by

$$W^n = [Z^n | H^n]^{\dagger} Y. \tag{9}$$

Denote

$$A = [Z^n | H^n]. \tag{10}$$

The pseudoinverse $A^{\dagger}$ can be solved by

$$A^{\dagger} = (A^T A)^{-1} A^T. \tag{11}$$

To get a more stable estimate for W^n, a weight penalty term is added to the loss function. We follow [52] to optimize W^n through solving the following minimization as:

$$\min_{W^n} \left\| A W^{n\text{-}Y} \right\|_2^2 + \lambda \left\| W^n \right\|_2^2 \tag{12}$$

where λ is a regularizer to control the network structure complexity and enhance the generality. Equation (12) is actually the loss function of BLS. Thus, through assigning the derivation of (12) with respect to W^n as zero, we could, hence, obtain the solution of W^n as follows:

$$W^n = (\lambda I + A^T A)^{-1} A^T Y. \tag{13}$$

For the pseudoinverse, we have

$$A^{\dagger} = \lim_{\lambda \to 0} \left(\lambda I + A^T A\right)^{-1} A^T. \tag{14}$$

Then, we get the initial connecting weights. Algorithm 1 shows the training process.

Algorithm 1 Face Sketch Synthesis Using BLS

Input: N training photo-sketch pairs (x_i, y_{t_i}), the number of feature maps n

Output: Connecting weights W^n

1: **for** $i = 1; i \le n$ **do**
2: Random w_{f_i}, β_{f_i}, w_{e_i}, β_{e_i};
3: **for** $k = 1; k \le N$ **do**
4: Calculate $z_i(x_k) = \phi(x_k \otimes w_{f_i} + \beta_{f_i})$;
5: Calculate $h_i(z_i(x_k)) = \xi(z_i(x_k) \otimes w_{e_i} + \beta_{e_i})$;
6: **end for**
7: **end for**
8: **for** $k = 1; k \le N$ **do**
9: Set the feature nodes $z_k^n = [z_1(x_k), \ldots, z_n(x_k)]$;
10: Set the enhancement nodes $h_k^n = [h_1(x_k), \ldots, h_n(x_k)]$;
11: **end for**
12: Concatenate all the feature nodes for the N face photos $Z^n = [(z_1^n)^T, \ldots, (z_N^n)^T]^T$;
13: Concatenate all the enhancement nodes for the N face photos $H^n = [(h_1^n)^T, \ldots, (h_N^n)^T]^T$;
14: Concatenate all nodes $A = [Z^n | H^n]$;
15: Obtain pseudoinverse $A^{\dagger} = \lim_{\lambda \to 0} \left(\lambda I + A^T A\right)^{-1} A^T$;
16: **return** $W^n = A^{\dagger} Y$;

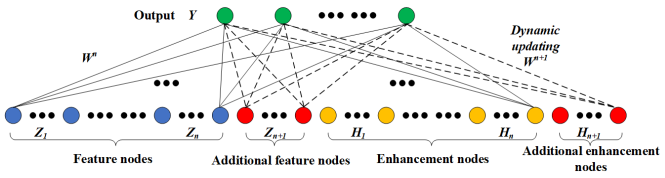


Fig. 5. Increment of feature mappings. While the selected feature mappings are insufficient for learning the photograph-to-sketch mapping, increasing the feature mappings and remodeling are adopted to improve the performance, accompanied by the increasing of enhancement nodes. Additional feature nodes and enhancement nodes are added by the proposed incremental algorithm, and the updated connecting weights can be calculated without a complete retraining process.

V. Incremental Learning Approach

While the selected feature mappings are insufficient for learning the photograph-to-sketch mapping, the solution we usually adopt is to increase the feature mapping, accompanied by the increasing of enhancement nodes (see Fig. 5). This is similar to most popular deep structures such as CNN. When the established structures cannot learn the task well, increasing the number of filters or layers may be applied in deep structures. While the remodeling process in deep structures suffers from a complete retraining process that takes up most of the time in the process of model selection, the incremental learning algorithm developed for a BLS can avoid the retraining process, which provides a better choice for fast model selection. Here, we consider the incremental learning for additional feature nodes accompanied by increasing of enhancement nodes. Assume the initial network with n feature mappings and n enhancement mappings. The $(n+1)$th feature mapping and its corresponding enhancement mapping can be denoted as

$$z_{n+1} = \phi(x \otimes w_{f_{n+1}} + \beta_{f_{n+1}}) \tag{15}$$

$$h_{n+1} = \xi(z_{n+1} \otimes w_{h_{n+1}} + \beta_{h_{n+1}}) \tag{16}$$

where $w_{f_{n+1}}$, $\beta_{f_{n+1}}$, $w_{h_{n+1}}$, and $\beta_{h_{n+1}}$ are randomly generated. Denote $Z_{n+1} = [(z_{n+1}(x_1))^T, \ldots, (z_{n+1}(x_N))^T]^T$ and $H_{n+1} = [(h_{n+1}(x_1))^T, \ldots, (h_{n+1}(x_N))^T]^T$ as the $(n+1)$th feature mapping and enhancement mapping of the N face photographs, respectively.

We perform feature mappings for the N face photographs, and we represent all the input nodes with $F_n^n = [Z_1, Z_2, \ldots, Z_n | H_1, H_2, \ldots, H_n]$, containing n groups of feature nodes and n groups of enhanced nodes. For simplicity, we denote the concatenation of the feature nodes and enhancement nodes as $A^n = F_n^n$. The final BLS is formulated as: $Y = A^n W^n$. The weight coefficient W^n could be calculated by pseudoinverse in BLS as $W^n = (A^n)^{\dagger} Y$, and the pseudoinverse of input nodes $(A^n)^{\dagger}$ could be calculated by

$(A^n)^\dagger = \lim_{\lambda\to 0}\left(\lambda I + A^n(A^n)^T\right)^{-1}(A^n)^T$. I is a unit matrix with proper dimensions, and λ is a regularization parameter for ridge regression (see the detail in [22]) and λ is set as 10^{-8}. The updated input nodes $A^{n+1} = F_{n+1}^{n+1}$ are denoted as

$$\begin{aligned} A^{n+1} &= [Z_1, Z_2, \ldots, Z_{n+1}|H_1, H_2, \ldots, H_{n+1}] \\ &\equiv [F_n^n|Z_{n+1}|H_{n+1}] = [A^n|Z_{n+1}|H_{n+1}] \end{aligned} \tag{17}$$

which is the upgrade of new mapped feature nodes and the corresponding enhancement nodes. Then, the pseudoinverse of the new input nodes $\left(A^{n+1}\right)^\dagger$ can be calculated as

$$\left(A^{n+1}\right)^\dagger = \begin{bmatrix} \left(A^n\right)^\dagger - \left(A^n\right)^\dagger [Z_{n+1}|H_{n+1}]B^T \\ B^T \end{bmatrix} \tag{18}$$

where B^T can be calculated by the additional feature node Z_{n+1}, the additional enhanced node H_{n+1}, and the original input nodes $A^n = F_n^n$. According to [22], let $D = (A^n)^\dagger [Z_{n+1}|H_{n+1}]$, and we hence have

$$B^T = \begin{cases} (C)^\dagger & \text{if } C \neq 0 \\ (1 + D^T D)^{-1} D^T (A^n)^\dagger & \text{if } C = 0 \end{cases} \tag{19}$$

where $C = [Z_{n+1}|H_{n+1}] - A^n D$.

Algorithm 2 Incremental Algorithm of the Increment Feature Mapping

Input: N training photo-sketch pairs $\left(x_i, y_{t_i}\right)$, the initial connecting weights W^n

Output: Updated connecting weights W

1: **while** *The training error threshold* is not satisfied **do**
2: Random $w_{f_{n+1}}$, $\beta_{f_{n+1}}$, $w_{h_{n+1}}$, $\beta_{h_{n+1}}$;
3: **for** $k = 1$; $k \leq N$ **do**
4: Calculate $z_{n+1}(x_k) = \phi(x_k \otimes w_{f_{n+1}} + \beta_{f_{n+1}})$;
5: Calculate $h_{n+1}(x_k) = \xi(z_{n+1}(x_k) \otimes w_{h_{n+1}} + \beta_{h_{n+1}})$;
6: **end for**
7: Concatenate all the new feature nodes for N face photos $Z_{n+1} = [(z_{n+1}(x_1))^T, \ldots, (z_{n+1}(x_N))^T]^T$;
8: Concatenate all the new enhancement nodes for N face photos $H_{n+1} = [(h_{n+1}(x_1))^T, \ldots, (h_{n+1}(x_N))^T]^T$;
9: Update A^{n+1} by Eq. (17);
10: Update $(A^{n+1})^\dagger$ and W^{n+1} by Eq. (18) and Eq. (20);
11: Set $n = n + 1$;
12: **end while**
13: **return** $W = W^n$;

Obviously, the weight W^{n+1} can be computed as

$$W^{n+1} = \begin{bmatrix} W^n - \left(A^n\right)^\dagger [Z_{n+1}|H_{n+1}]B^T Y \\ B^T Y \end{bmatrix}. \tag{20}$$

Obviously, the new weights W^{n+1} could be computed based on the pseudoinverse of all input nodes $A^{n+1} = F_{n+1}^{n+1}$, and only the pseudoinverse of additional nodes needs to be calculated without calculating the entire input nodes so that the system could be built quickly. The incremental algorithm of the increment feature mapping is shown in Algorithm 2.

VI. Regularized Method via Bayesian Estimation

What is mentioned above is the standard BLS and its incremental learning used for face sketch synthesis. However, the MSE criterion applied in the standard BLS is insufficient for the situation with unexpected input such as images with noise. Besides, as for the face sketch synthesis problem, the training data are usually too small to get a good generalization ability and a stable output by the standard BLS. Thus, to better model the errors in most real situation where noise exists and performs feature selection, a regularized algorithm for better generalization ability and robustness has to be developed particularly. In the standard BLS, MSE is used to model the errors with an assumption of Gaussian distribution. However, the Gaussian distribution has a weakness to model the real errors. To make the algorithm more robust, in this section, the widely used Laplacian distribution is used to model the errors more realistically [48]–[52] and optimize the standard BLS. We follow [52] to employ the Bayesian estimation to make our first assumption that W^n follows a Gaussian distribution, and we, hence, get the loss function as:

$$\min_{W^n} \left\| AW^{\text{n-Y}} \right\|_1 + \alpha \left\| W^n \right\|_2^2. \tag{21}$$

Comparing (21) and (12), we can find that the l_2 loss function in (12) is replaced by the l_1 loss function to improve robustness. We utilize the Laplacian error distribution to solve the problem when unknown noise/error exists and make the algorithm more robust. The l_1 loss function, which is the difference in (21) from (12), can be explained to reflect the sparsity of noise and the robustness of the improved BLS. Also, the l_2-norm regular term in (21) introduced by the Gaussian assumption of W^n can make the connecting weights smooth. For small training data problems, a feature selection strategy is employed, which gives a sparsity constraint to the connecting weights W^n.

We then follow [52] to make our second assumption that W^n follow the Laplacian distribution, and hence, we could get the l_1-norm regular term for the second term in (21). To combine the advantages of both l_1 and l_2 regularizers, we get a loss function by combining l_2 regularizer and l_1 regularizer as

$$\min_{W^n} \left\| AW^{\text{n-Y}} \right\|_1 + \alpha_1 \left\| W^n \right\|_1 + \alpha_2 \left\| W^n \right\|_2^2 \tag{22}$$

where α_1 and α_2 are adjustable parameters. Unlike (12), (22) cannot be solved analytically. We use it to further optimize iteratively the weight W^n calculated by the standard BLS. Here, an iterative optimization based on the augmented Lagrange multiplier (ALM) [53] is given to solve the problem.

Through the strategy of ALM, (22) can be rewritten as augmented Lagrangian function as

$$\begin{aligned} L(E, Q, W^n, \lambda_1, \lambda_2) = \|E\|_1 + \alpha_1 \|Q\|_1 + \alpha_2 \left\| W^n \right\|_2^2 \\ + \frac{\mu}{2} \left\| Y - AW^n - E + \frac{\lambda_1}{\mu} \right\|_2^2 \\ + \frac{\mu}{2} \left\| Q - W^n + \frac{\lambda_2}{\mu} \right\|_2^2 \end{aligned} \tag{23}$$

where λ_1 and λ_2 are Lagrange multipliers and $\mu \geq 0$ is the penalty parameter. Optimized by the block coordinate

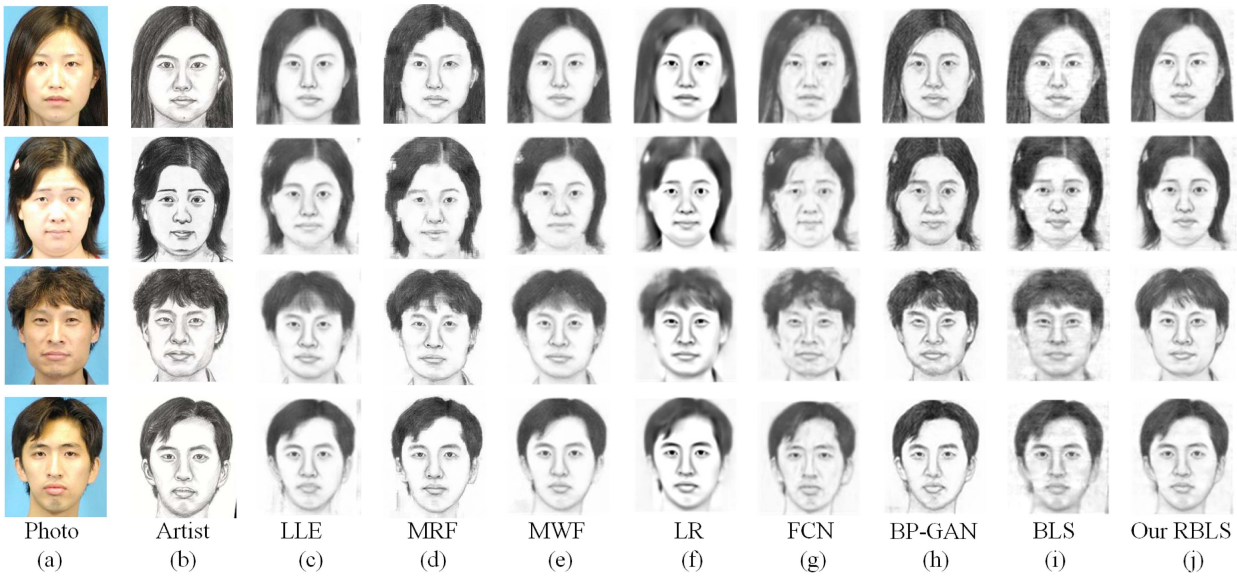


Fig. 6. Sketch synthesis results on the CUHK student data set. (a) Four test photographs from the CUHK student data set. (b) Corresponding sketches drawn by artists used as ground truth. (c) Results of the LLE-based method [7]. (d) Results of the MRF method [29]. (e) Results of the MWF method [54]. (f) Results of the LR-based method [12]. (g) Results of the FCN method [13]. (h) Results of the BP-GAN method [4]. (i) Results of the standard BLS method without the regularized algorithm. (j) Results of our proposed RBLS method.

Algorithm 3 Regularized Algorithm via Bayesian Estimation

Input: The concatenation of feature nodes and enhancement nodes A, initial connecting weights W^n, regularization parameter α_1, α_2, penalty parameter μ

Output: Connecting weights W^n

```
Initialize λ1, λ2;
while not converged do
    Optimize E by Eq. (24);
    Optimize Q by Eq. (25);
    Optimize λ1 and λ2 by Eq. (26) and Eq. (27);
    Optimize W^n by Eq. (28);
end while
return W^n;
```

descent (BCD) method, the connecting weights can be updated iteratively as follows:

$$E = \max(|S| - 1/\mu, 0) \odot \text{sign}(S) \tag{24}$$

$$Q = \max(|M| - \alpha_1/\mu, 0) \odot \text{sign}(M) \tag{25}$$

$$\lambda_1 = \lambda_1 + \mu(Y - AW^n - E) \tag{26}$$

$$\lambda_2 = \lambda_2 + \mu(Q - W^n) \tag{27}$$

$$W^n = (A^T A + (1 + 2\alpha_2/\mu)I)^{-1}(A^T T + F) \tag{28}$$

where $S = Y - AW^n + \lambda_1/\mu$, $M = W^n - \lambda_2/\mu$, $T = Y - E + \lambda_1/\mu$, $F = Q + \lambda_2/\mu$, and $\odot$ represent the element-wise multiplication. The process is iteratively optimized until reaching to convergence. The computational complexity can be reduced by using the partial singular value decomposition (SVD) [55] to simplify the computing process of the inverse in (28). Algorithm 3 shows the implementation process of the regularized algorithm.

VII. Experimental Results

We evaluate the proposed method on the CUHK student data set, which includes 188 face photograph–sketch pairs and Aleix Robert (AR) data set, which includes 123 pairs [29]. Each face photograph has its corresponding sketch drawn by the artist (see Fig. 1). We select 88 face photograph–sketch pairs for training and the remaining for testing from the CUHK student data set. For the AR data set, 80 pairs are selected randomly for training and the remaining for testing.

A. Face Sketch Synthesis

Fig. 6 shows the sketch synthesis results for four test examples by LLE [7], MRF [29], Markov weight fields (MWF) [54], linear regression (LR) [12], FCN [13], BP-GAN [4], BLS, and our RBLS method on the CUHK student data set. The first three methods are data-driven, and the other five methods are model-driven. Among these, FCN [13] and BP-GAN [4] are deep learning-based methods. Compared with data-driven methods, our method keeps more detailed information. For example, the man in row three has small eyes whose shapes are changed by the first three methods. Similarly, on the AR data set, Fig. 7 shows that our approach can preserve more face details. As we can see from Fig. 7, the man in row one wears glasses. The detailed glass structures are destroyed when using data-driven methods, while the result of our RBLS shows a more reliable result. The detailed structural information of the women's hair from row two to row three can be found destroyed by different degrees using data-driven methods, while model-driven methods can preserve more structural information. This is because most data-driven methods are based on matching algorithms, which search similar sketch patches from the training sets and synthesize the final face sketch. It is, hence, difficult for data-driven methods to

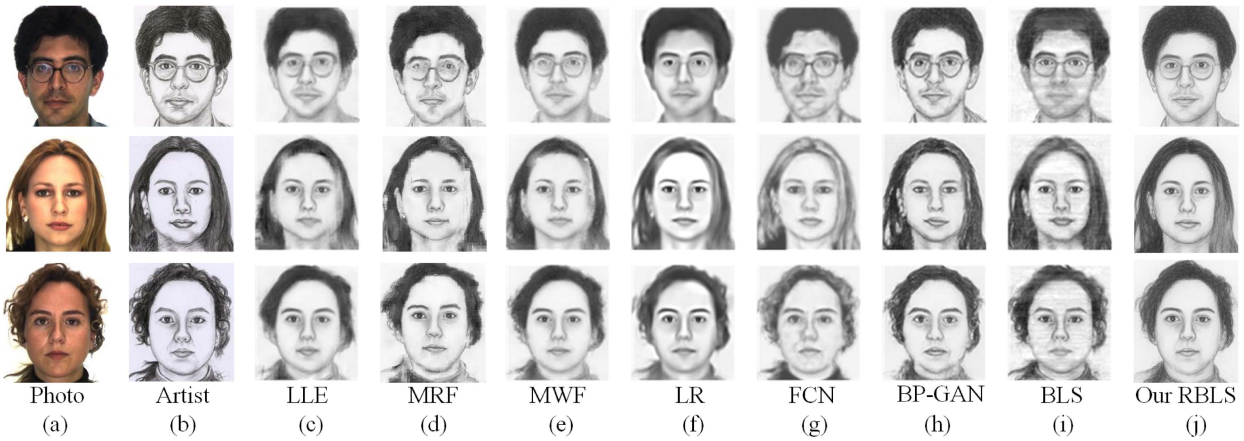


Fig. 7. Sketch synthesis results on the AR data set. (a) Four test photographs from the AR data set. (b) Corresponding sketches drawn by artists used as ground truth. (c) Results of the LLE-based method [7]. (d) Results of the MRF method [29]. (e) Results of the MWF method [54]. (f) Results of the LR-based method [12]. (g) Results of the FCN method [13]. (h) Results of the BP-GAN method [4]. (i) Results of the standard BLS method without the regularized algorithm. (j) Results of our proposed RBLS method.

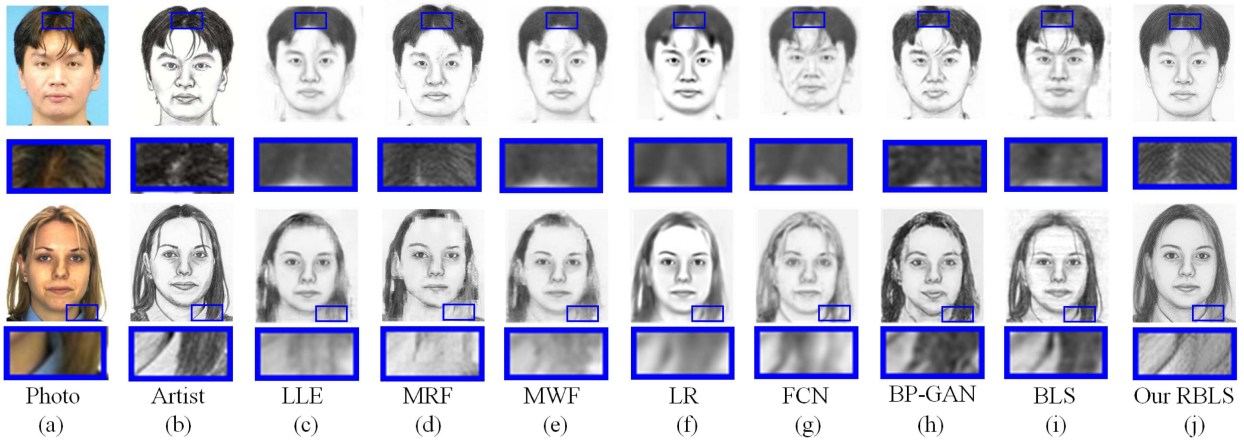


Fig. 8. Qualitative evaluation with local detail magnification for different approaches. (a) Input face photographs. (b) Corresponding sketches drawn by artists used as ground truth. (c) Results of the LLE-based method [7]. (d) Results of the MRF method [29]. (e) Results of the MWF method [54]. (f) Results of the LR-based method [12]. (g) Results of the FCN method [13]. (h) Results of the BP-GAN method [4]. (i) Results of the standard BLS method without the regularized algorithm. (j) Results of our proposed RBLS method.

reconstruct uncommon details. However, model-driven methods directly learn the modality of sketches from training sets and transform target photographs into sketches with vital details preserved through the trained model. In this article, taking the advantages of model-driven methods, we train a broad learning network to learn the sketch modality and realize direct photograph–sketch transformation. In this way, our approach is able to imitate the underlying sketch styles and preserve more semantic details from the target photographs.

In comparison with some model-driven methods as shown in Figs. 6 and 7, our method provides more reliable results. As the linear assumption has limitations in learning the model of photograph–sketch transformation, the LR method used in Figs. 6(f) and 7(f) has a limitation in performance. For the deep learning methods (i.e., FCN and BP-GAN), our results are better in visual detail quality. As can be seen from Figs. 6 and 7, the results of FCN are accompanied with noise and are blurred, and the synthesis results of BP-GAN have many detail loss [see row two in Fig. 6(h) and row two in Fig. 7(h)] and deformation [see Fig. 7(h)]. In Fig. 6 and 7, we can find that the face sketch synthesis results of our RBLS are much smoother than that of the traditional BLS. The results synthesized by BLS are corrupted by noise and some details are vague, while these problems are effectively addressed by our proposed RBLS-based approach, which demonstrates the effectiveness of our proposed sketch synthesis method.

Fig. 8 shows two detailed face sketch synthesis results with enlarged views to demonstrate the advantages of our RBLS synthesis. The first row of Fig. 8 shows a boy with two strands of delicate hair from the CUHK student data set. From the comparison of the enlarged views of the critical hair region, we can see that only our synthesis result offers clear hair structures. All the other methods could not restore the delicate hair structures well. The second row of Fig. 8 shows a young girl with a neat collar from the AR data set. All the methods except BLS and our RBLS could not preserve the neat collar structures. While BLS introduces many noises, our RBLS restore clear collar structures are the original photograph and

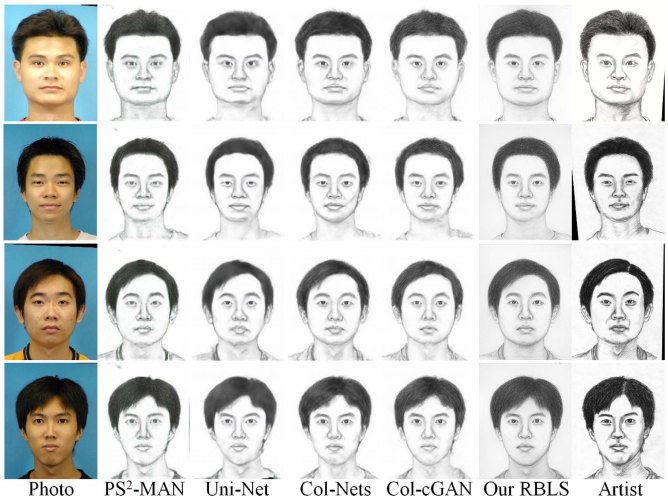


Fig. 9. Extended comparison of face sketch synthesis results on the CUHK student data set. From left to right: input face photograph, PS^2-MAN [16], Uni-Net [19], Col-Nets [19], Col-cGAN [19], our proposed RBLS, and ground truth drawn by artists.

the sketch by an artist while significantly reducing noises. At the same time, our proposed RBLS synthesis also offers clear details for the mouth, eyes, eyebrow, and hair of the girl, which shows the robustness of our RBLS-based face sketch synthesis.

B. Extended Sketch Synthesis

Since the key advantages of our proposed RBLS approach lies in the efficient high-quality face sketch synthesis, we further test the effectiveness of our RBLS method on different data sets in comparison with more state-of-the-art methods, i.e., PS^2-MAN [16], Uni-Net [19], Col-Nets [19], Col-cGAN [19], FCN [13], LR [12], BP-GAN [4], Fast-RSLCR [18], and SuperLLC [20]. Fig. 9 shows more extended comparison results of face sketch synthesis on the CUHK student data set. In Fig. 9, we compare our proposed RBLS face sketch synthesis with PS^2-MAN [16], Uni-Net [19], Col-Nets [19], and Col-cGAN [19]. The first row contains a man in relatively good lighting conditions, and the other three rows have a bit darker lighting condition. In all four cases, our RBLS synthesis provides realistic face sketch synthesis results. PS^2-MAN tends to produce unnecessary noises around the boundary areas (e.g., noses and necks), Uni-Net has some distortions in the collar regions, Col-Nets generated overly highlighted wrinkles (e.g., the face on the second row), and Col-cGAN generally has good visual effects but also produces some distortions in the collar regions (e.g., the last three rows). Our RBLS excels in the term structure details, especially good at preserving detailed subtle structural information such as hair structures.

In addition, we have utilized more data sets to test our RBLS approach, i.e., CUHK Face Sketch XM2VTS and CUHK Face Sketch FERET (CUFSF). The XM2VTS and CUFSF databases contain face sketches with relatively shape exaggeration designed by artists. The CUFSF data set contains 1194 pairs of face photograph and face sketch, and the

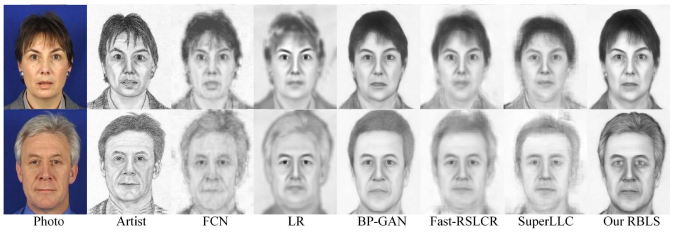


Fig. 10. Extended comparison of sketch synthesis on the XM2VTS data set. From left to right: input face photograph, ground truth drawn by artists, FCN [13], LR [12], BP-GAN [4], Fast-RSLCR [18], SuperLLC [20], and our proposed RBLS synthesis.

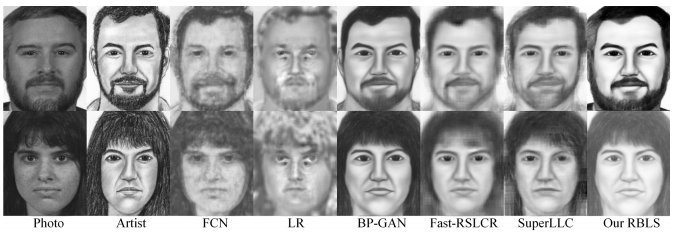


Fig. 11. Extended comparison of sketch synthesis on the CUFSF data set. From left to right: input face photograph, ground truth drawn by artists, FCN [13], LR [12], BP-GAN [4], Fast-RSLCR [18], SuperLLC [20], and our proposed RBLS synthesis.

XM2VTS data set consists of 295 pairs of face photograph and face sketch. We utilize the leave-one-out method to test our RBLS synthesis on the XM2VTS data set, while we utilize half pairs for training and half pairs for testing for applying our RBLS synthesis on the CUFSF data set. Fig. 10 shows more results of our proposed RBLS synthesis on the XM2VTS data set in comparison with FCN [13], LR [12], BP-GAN [4], Fast-RSLCR [18], and SuperLLC [20]. Fig. 11 shows more results of our proposed RBLS synthesis on the CUFSF data set in comparison with these state-of-the-art methods. From Fig. 10, we could find that our RBLS results are of high-quality overall visual effects as good as BP-GAN, but our approach outperforms the other state-of-the-art methods in the ways of better-detailed structures (e.g., the hair). From Fig. 11, we could see that our RBLS synthesis method offers comparable visual effects with the state-of-the-art methods (e.g., BP-GAN, Fast-RSLCR, and SuperLLC). Again, in terms of detailed structures like the hair regions, our RBLS demonstrates its superiority. Please note that the drawing styles of face sketches by artists for the CUFSF database are not in a similar style with the CUHK Student, AR, and XM2VTS databases. Therefore, all the generated sketch results of Fig. 11 are of quite different styles compared with results from the other three databases. We again show more experimental results in comparison with some representative methods on different data sets in Fig. 12. The extended sketch synthesis experiments in Figs. 9–12 have demonstrated the effectiveness and robustness of our RBLS face sketch synthesis approach on different data sets.

C. Objective Image Quality Evaluation

In Sections VII-A and VII-B, the proposed method is intuitively compared with some existing methods, and the

TABLE I
SIMILARITY MEASUREMENT BETWEEN THE ESTIMATED SKETCH AND THE GROUND TRUTH

Dataset	Meas.	Methods							
		LLE [7]	MRF [29]	MWF [55]	LR [12]	FCN [13]	BP-GAN [4]	BLS	RBLS
CUHK	MSE	1167	1379	1151	1200	1157	1339	963	813
	PSNR	17.75	16.99	17.86	17.62	17.79	17.20	18.42	19.20
	SSIM	0.8860	0.8728	0.8897	0.8827	0.8790	0.8718	0.9013	0.9152
	Scoot	0.4789	0.5149	0.4832	-	0.4527	-	0.5676	0.7235
AR	MSE	824	1220	809	981	969	988	724	594
	PSNR	19.39	18.04	19.53	18.48	18.77	18.43	19.53	20.40
	SSIM	0.9069	0.8728	0.9100	0.8979	0.8853	0.8930	0.9180	0.9327
	FSIM	0.7216	0.7496	0.7579	0.7083	0.6824	0.7663	0.7657	0.7935

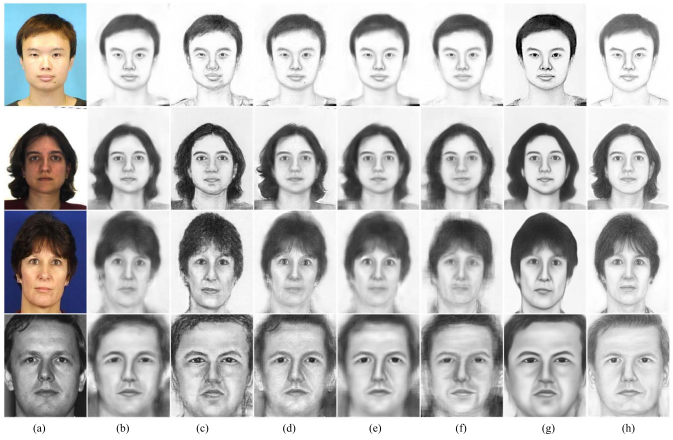


Fig. 12. Comparison with some representative methods for synthesizing sketches on different data sets. The photograph in the first row is from the CUHK data set. The photograph in the second row is from the AR data set. The photograph in the third row is from the XM2VTS data set. The photograph in the fourth row is from the CUFSF data set. (a) Input photographs. (b) Results of the spatial sketch denoising (SSD) method [56]. (c) Results of the BP-GAN method [4]. (d) Results of the *RL* method [57]. (e) Results of the Fast-RSLCR method [18]. (f) Results of the SuperLLC method [20]. (g) Results of the cascaded low-rank representation (CLRR) method [58]. (h) Results of our RBLS.

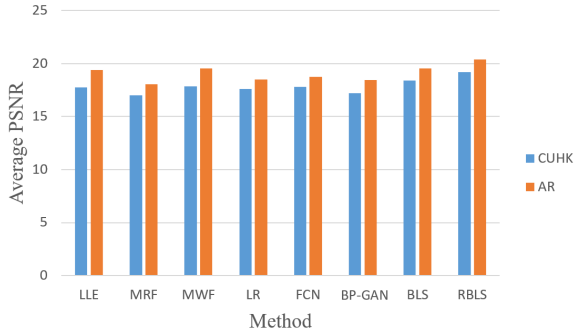


Fig. 13. Average PSNR values for 100 test images of the CUHK data set and average PSNR values for 43 test images of AR data set. The proposed RBLS method is compared with the other seven methods, i.e., LLE [7], MRF [29], MWF [54], LR [12], FCN [13], BP-GAN [4], and the standard BLS methods. Results show that our method achieves a better image quality.

conclusion that our method can preserve more details thus giving better qualities is made. To be more convincing, in this section, we use the sketches drawn by artists as reference images and make objective evaluation by three objective image quality measurement methods, namely MSE, peak signal-to-noise ratio (PSNR), and structural similarity (SSIM). Table I gives the average MSE, PSNR, and SSIM scores of the estimated sketches on CUHK and AR. We also compared the Scoot score [39], [40] on the CUHK data set and the feature similarity (FSIM) score on the AR data set in Table I. As mentioned above, the MSE is used as the loss function whose values represent the loss of synthesis methods. As we can see from Table I, the proposed method gives the minimum loss both on the CUHK student data set and the AR data set compared with the other methods. The PSNR and SSIM are similarity measurements between the generated sketches and the ground truth, which means that the greater the values are, the sketches generated are more similar to the expected results. The FSIM measures the FSIM, while the Scoot is a perceptual similarity metric. We can find in Table I that our results give the maximum similarity and thus provide better image qualities. Fig. 13 shows the average PSNR values for the 100 test images of the CUHK data set and the average PSNR values for the 43 test images of the AR data set. Fig. 14 shows the statistics of SSIM scores on the CUHK data set and AR data set. The values of the y-axis are percentages of synthesized sketches whose SSIM scores are not smaller than the corresponding values of the x-axis. As we can see clearly from Figs. 13 and 14, our approach outperforms the other methods in most cases.

D. Face Sketch Recognition

An important application of face sketch synthesis is assisting law enforcement, which matches the sketches of a suspect with the photographs in the mug-shot database by face sketch recognition. Thus, besides objective image quality evaluation, face sketch recognition accuracy also reflects the quality of synthesized sketches quantitatively. In this section, the PCA-based face recognition method is employed to conduct face recognition experiments. The face photographs in the gallery are first transformed into sketches by face sketch synthesis methods, and then, the query sketch is matched with the synthesized sketches. Table II gives the face recognition accuracy of RBLS compared with several sketch synthesis methods, including LLE [7], MRF [29], CM-based [30] (denoted by CM in Table II), FCN [13], and the standard BLS. We can find that our method achieves the best performance both on the CUHK student data set and AR data set, which

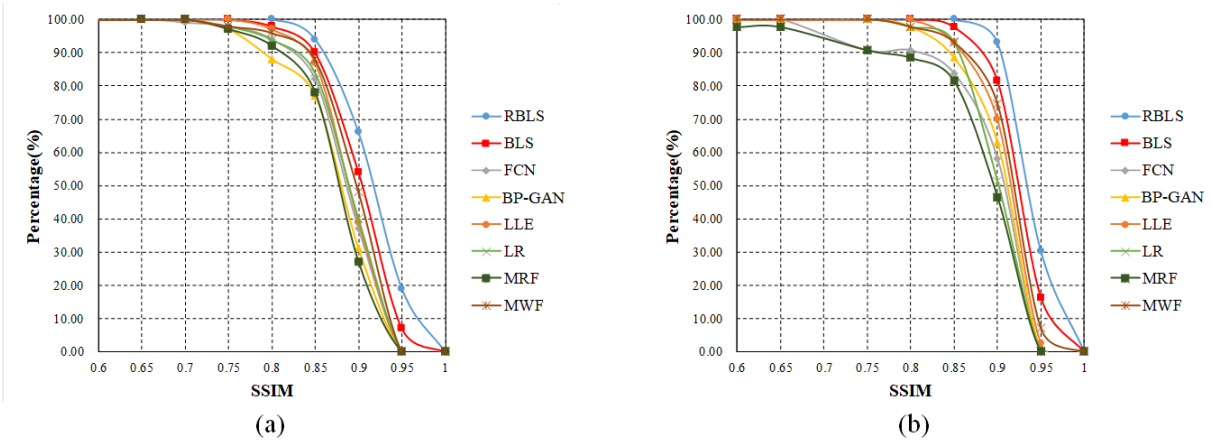


Fig. 14. SSIM scores on the CUHK data set and AR data set. (a) Results on CUHK data set. (b) Results on AR data set. The proposed RBLS method is compared with other seven methods, including the standard BLS, FCN [13], BP-GAN [4], LLE [7], LR [12], MRF [29], and MWF [54] methods. The values of the y-axis are percentages of synthesized sketches whose SSIM scores are not smaller than the corresponding values of the x-axis. Experiments show that our method provides a better image quality.

TABLE II
COMPARISON OF THE RANK ONE FACE SKETCH RECOGNITION ACCURACY ON THE CUHK AND AR DATA SETS

Dataset	LLE [7]	MRF [29]	CM [30]	FCN [13]	BLS	RBLS
CUHK	0.85	0.83	0.89	0.81	0.90	0.92
AR	0.95	0.93	0.96	0.93	0.97	0.98

demonstrates the effectiveness of our method. As for LLE and MRF methods, the image is blurred by the linear combination and average of sketch patches, which leads to image quality deterioration. The CM-based method imitates the drawing process of artists and decomposed a face into components. In this way, the synthesis quality can be improved to a certain extent. However, its performance is still limited by the drawbacks of data-driven methods. For model-driven methods, although they overcome the weakness of data-driven methods, their performance is limited by the small training samples and unknown noises. Through our method, various experiments have shown that our method well improves performance.

E. Time Complexity

As previously mentioned, the great advantages of our method are the generalization capability in function approximation and the ability for fast training and remodeling. The aforementioned experiments have demonstrated the excellent performance of RBLS in approximating the nonlinear function, which maps face photographs into sketches from both subjective and objective aspects. In this section, the time complexity is discussed to show another advantage of the proposed method. Compared with most of the data-driven methods based on neighbor selection to synthesize a sketch, the proposed method takes much less time. Even taking into account the training time, the total synthesis time is about 70 s for the 88 training pairs and the 100 test photographs on the CUHK student data set. Meanwhile, our RBLS could offer a better synthesis quality efficiently, while other methods usually take hundreds of seconds to synthesize only one sketch. This is mainly caused by the time-consuming part of data-driven methods, namely, the neighbor selection process whose complexity is related to the number of training pairs and patches.

Compared to most deep learning-based structures, such as CNN, FCN, and BP-GAN, the proposed method can save a lot of time in the training process, which only consumes 30 s for 100 training pairs in a regular PC, while the deep structures usually take several hours even on a GPU with high performance. This mainly benefits from the flat structure of RBLS. Thus, there is no need for the iterative parameter optimization process with complex error backpropagation, which saves much time. The time complexity is proportional to the numbers of filters or feature mappings. Table III gives the timing performance results on the CUHK student data set using an incremental learning algorithm for incremental feature mapping. The initial number of feature mappings is set as 3. To improve the performance and find the best network structure, the number of filters (namely feature mappings) is increased. With the proposed increment learning algorithm, the remodeling process saves a lot of time. This is mainly because, for deep structures, model selection and hyperparameters setting are usually done artificially and empirically. Thus, once the hyperparameters are changed, a complete retraining process is necessary to remodel the network, which consumes most of the time in modeling. However, in the RBLS, the hyperparameters can be adaptively set to improve the network at a fast speed by incremental learning.

F. Parameter Analysis

When applied to the sketch synthesis problem, the proposed network based on RBLS has two network parameters: filter number n and filter size s for feature extraction. In this section, we use the CUHK student data set to discuss the situation when using different sizes of filters. Then, the results of the incremental learning algorithm are given to analyze the performance of different numbers of filters. For simplicity, we use

TABLE III

TIMING PERFORMANCE OF OUR RBLS INCREMENTAL LEARNING ON THE CUHK STUDENT DATA SET

Number of Filters	Size of Filters	PSNR	Each Additional Training - Time Cost (s)	Accumulative Training - Time Cost (s)	Each Additional Testing - Time Cost (s)	Accumulative Testing - Time Cost (s)
9	9	17.94	26.2027	26.2027	27.4384	27.4384
3	9	17.35	8.4329	8.4329	9.0154	9.0154
3→4	9	18.27	3.4527	11.8856	4.0234	13.0388
4→5	9	18.95	3.2654	15.1510	3.8720	16.9108
5→6	9	19.20	3.8560	19.0070	3.4291	20.3399
6→7	9	19.04	3.6245	22.6315	3.9026	24.2425
7→8	9	18.73	3.8201	26.4516	3.7830	28.0255
8→9	9	17.89	3.5295	29.9811	3.6192	31.6447

TABLE IV

USER STUDY RESULTS

Sketch set (No.)	Approach	Short-term recognition speed (s)	Short-term retention (%)	Long-term recognition speed (s)	Long-term retention (%)	Attractiveness
1	BP-GAN [4]	2.1	58	3.8	49	Medium
	MWF [54]	2.5	49	4.3	37	Medium
	Our RBLS	1.8	87	2.6	68	High
2	BP-GAN [4]	2.3	57	3.5	51	Medium
	MWF [54]	2.7	45	4.6	43	Medium
	Our RBLS	1.9	81	2.1	75	High
3	BP-GAN [4]	2.2	46	3.3	39	Medium
	MWF [54]	2.5	43	4.2	37	Medium
	Our RBLS	1.8	72	2.3	61	High

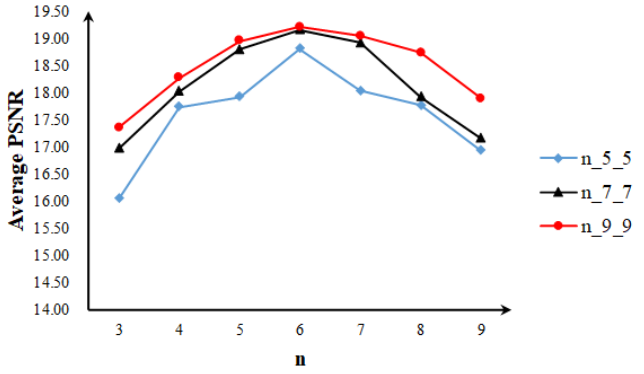


Fig. 15. Average PSNR values for different numbers of filters n when $s = 5$, $s = 7$, and $s = 9$.

the notation n_s_s to represent the proposed method based on RBLS with n filters with the size of $s \times s$ during the process of feature nodes construction and the process of enhancement nodes construction. Fig. 15 shows the performance when using different filter sizes. We can see from Fig. 15 that, as the filter size is increased, so does the average PSNR value, which means better synthetic image quality. However, when the filter size is increased close to 9, the performance does not improve a lot. Overall, we find that appropriately increasing the filter sizes can improve the performance, but the improvement is limited to a certain extent, which may be caused by the network structure that can be viewed as a two-layer perceptron. Theoretically, in the case that the filter size is equal to 1, the feature nodes and enhancement nodes are constructed by feature components independently. However, when the filter size is larger than 1, the conversion process combines the neighborhood information. The experimental results have shown that feature vectors are related to their neighbor feature components, which is consistent with the theoretical analysis.

As mentioned above, the performance of the RBLS was limited by the two-layer network structures. An intuitive approach to improve the performance is to increase the number of nodes in the network by increasing the number of filters. Fig. 15 shows the experimental results of using different numbers of filters for face sketch synthesis on the CUHK student data set. As shown in Fig. 15, the performance is improved as the number of filters increases as expected in a certain range. However, continuing to increase the number of filters leads to performance degradation. This is because when the number of filters is larger, a lot of unnecessary feature information is extracted. The information will corrupt the final synthesis sketches and make the RBLS overfitting.

G. User Study

We perform both quantitative and qualitative user study evaluations to test the performance of sketch synthesis results. One aspect is the recognition speed (s) of users. For each test photograph, users are provided with six different sketches. We record the time users took to identify the exact person. Another aspect we test is the retention percentage for users. After a fixed time, users are again provided with the sketches to check whether they can figure out the identities. We record the mean value of the two aspects. The two aspects (recognition speed and retention) are evaluated in long- and short-term manners. Short term denotes the first time we conduct the test. Long term denotes the same test making two weeks later. Again, we require the users to describe the attractiveness of the sketches as: "High," "Medium," or "Low," We studied three groups of such sketch test, each with six different sketches. The user study results are shown in Table IV.

From Table IV, we can see that our sketches surpass BP-GAN [4] and MWF [54] approaches in both retention (short term and long term) and recognition speed (short term and long term). Users also consider our sketches with high evaluation for attractiveness.

VIII. Conclusion and Future Work

In this article, we propose a new face sketch synthesis approach via the RBLS. The core of this system is a flat network with the advantages of fast modeling and remodeling. To make full use of the 2-D information of photographs, we extract image features by convolution. Besides, the incremental learning algorithm is proposed to adaptively refine the network for fast remodeling, and a regularized algorithm via Bayesian estimation is utilized to further improve the generalization ability and robustness. Experiments have shown that the synthesized sketches have good perceptual quality with more details preserved compared with traditional data-driven methods. In comparison to deep learning-based methods, the proposed RBLS approach can synthesize comparative sketches with a faster speed due to the time-saving training and remodeling. In summary, our proposed RBLS synthesis provides a better choice for face sketch synthesis in practical use.

Our RBLS synthesis still has its limitation. The results generated by our approach may not be fully satisfactory when the facial structures/poses are too complicated. We will work on to solve such complications as our future work. We will also utilize the latest face registration methods to enhance the spatial correspondence between the face photograph and the face sketch for more general applications. Some of the generated face sketch results using our RBLS approach may have horizontal bar artifacts and may not be fully satisfactory, especially when dealing with complicated scenarios. In the future, we will work on coprocessing and/or postprocessing to solve such complications. Inspired by Zhang *et al.* [59], we will also try to use scribbles locating key components of face image for aiding to get better visual quality. Besides, we attempt to utilize local edge information of facial features to guide the face sketch synthesis like [60] so that the spatial consistency can be ensured. Existing data sets for face sketch synthesis mainly contain well-positioned front faces, and however, scenes are usually complex in the wild. Thus, like [61], we will consider building a new face data set of complex scenes and evaluate existing face sketch synthesis and the proposed method on the new data set. Handling images with depth has been a new trend [62], and we will also extend our proposed method to face images with depth, which means that we will try to create a data set of face images with depth.

References

[1] M. Zhang, N. Wang, Y. Li, and X. Gao, "Deep latent low-rank representation for face sketch synthesis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 10, pp. 3109–3123, Oct. 2019.

[2] N. Wang, D. Tao, X. Gao, X. Li, and J. Li, "A comprehensive survey to face hallucination," *Int. J. Comput. Vis.*, vol. 106, no. 1, pp. 9–30, Jan. 2014.

[3] S. Agarwal and D. P. Mukherjee, "Synthesis of realistic facial expressions using expression map," *IEEE Trans. Multimedia*, vol. 21, no. 4, pp. 902–914, Apr. 2019.

[4] N. Wang, W. Zha, J. Li, and X. Gao, "Back projection: An effective postprocessing method for GAN-based face sketch synthesis," *Pattern Recognit. Lett.*, vol. 107, pp. 59–65, May 2018.

[5] X. Tang and X. Wang, "Face photo recognition using sketch," in *Proc. Int. Conf. Image Process.*, 2002, pp. 257–260.

[6] X. Tang and X. Wang, "Face sketch synthesis and recognition," in *Proc. ICCV*, 2003, pp. 687–694.

[7] Q. Liu, X. Tang, H. Jin, H. Lu, and S. Ma, "A nonlinear approach for face sketch synthesis and recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Oct. 2005, pp. 1005–1010.

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

[9] L. Jiao, S. Zhang, L. Li, F. Liu, and W. Ma, "A modified convolutional neural network for face sketch synthesis," *Pattern Recognit.*, vol. 76, pp. 125–136, Apr. 2018.

[10] X. Fu, B. Liang, Y. Huang, X. Ding, and J. Paisley, "Lightweight pyramid networks for image deraining," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 6, pp. 1794–1807, Jun. 2020.

[11] Y. Tang, W. Gong, X. Chen, and W. Li, "Deep inception-residual Laplacian pyramid networks for accurate single-image super-resolution," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 5, pp. 1514–1528, May 2020.

[12] N. Wang, M. Zhu, J. Li, B. Song, and Z. Li, "Data-driven vs. model-driven: Fast face sketch synthesis," *Neurocomputing*, vol. 257, pp. 214–221, Sep. 2017.

[13] L. Zhang, L. Lin, X. Wu, S. Ding, and L. Zhang, "End-to-end photo-sketch generation via fully convolutional representation learning," in *Proc. 5th ACM Int. Conf. Multimedia Retr.*, Jun. 2015, pp. 627–634.

[14] M. Zhu, N. Wang, X. Gao, and J. Li, "Deep graphical feature learning for face sketch synthesis," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, Aug. 2017, pp. 3574–3580.

[15] C. Chen, W. Liu, X. Tan, and K.-Y. K. Wong, "Semi-supervised learning for face sketch synthesis in the wild," in *Proc. ACCV*, 2018, pp. 216–231.

[16] L. Wang, V. Sindagi, and V. Patel, "High-quality facial photo-sketch synthesis using multi-adversarial networks," in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2018, pp. 83–90.

[17] N. Wang, X. Gao, L. Sun, and J. Li, "Anchored neighborhood index for face sketch synthesis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 9, pp. 2154–2163, Sep. 2018.

[18] N. Wang, X. Gao, and J. Li, "Random sampling for fast face sketch synthesis," *Pattern Recognit.*, vol. 76, pp. 215–227, Apr. 2018.

[19] M. Zhu, J. Li, N. Wang, and X. Gao, "A deep collaborative framework for face photo–sketch synthesis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 10, pp. 3096–3108, Oct. 2019.

[20] A. Radman and S. A. Suandi, "A superpixel-wise approach for face sketch synthesis," *IEEE Access*, vol. 7, pp. 108838–108849, 2019.

[21] C. L. P. Chen and J. Z. Wan, "A rapid learning and dynamic stepwise updating algorithm for flat neural networks and the application to time-series prediction," *IEEE Trans. Syst., Man, Cybern. B. Cybern.*, vol. 29, no. 1, pp. 62–72, Oct. 1999.

[22] C. L. P. Chen and Z. Liu, "Broad learning system: An effective and efficient incremental learning system without the need for deep architecture," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 1, pp. 10–24, Jan. 2018.

[23] C. L. P. Chen, Z. Liu, and S. Feng, "Universal approximation capability of broad learning system and its structural variations," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 4, pp. 1191–1204, Apr. 2019.

[24] S. Feng and C. L. P. Chen, "Fuzzy broad learning system: A novel neuro-fuzzy model for regression and classification," *IEEE Trans. Cybern.*, vol. 50, no. 2, pp. 414–424, Feb. 2018.

[25] X. Gao, N. Wang, D. Tao, and X. Li, "Face sketch–photo synthesis and retrieval using sparse representation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 8, pp. 1213–1226, Aug. 2012.

[26] S. Zhang, X. Gao, N. Wang, and J. Li, "Robust face sketch style synthesis," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 220–232, Jan. 2016.

[27] S. Zhang, X. Gao, N. Wang, and J. Li, "Face sketch synthesis from a single photo-sketch pair," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 2, pp. 275–287, Feb. 2017.

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

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

[30] M. Zhang, J. Li, N. Wang, and X. Gao, "Compositional model-based sketch generator in facial entertainment," *IEEE Trans. Cybern.*, vol. 48, no. 3, pp. 904–915, Mar. 2018.

[31] L. Chang, M. Zhou, Y. Han, and X. Deng, "Face sketch synthesis via sparse representation," in *Proc. ICPR*, 2010, pp. 2146–2149.

[32] N. Wang, X. Gao, D. Tao, and X. Li, "Face sketch-photo synthesis under multi-dictionary sparse representation framework," in *Proc. ICIG*, 2011, pp. 82–87.

[33] L. Gatys, A. Ecker, and M. Bethge, "A neural algorithm of artistic style," *J. Vis.*, vol. 16, no. 12, p. 326, Sep. 2016.

[34] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2414–2423.

[35] D. Ulyanov, V. Lebedev, and V. Lempitsky, "Texture networks: Feed-forward synthesis of textures and stylized images," in *Proc. ICML*, 2016, pp. 1349–1357.

[36] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 694–711.

[37] C. Chen, X. Tan, and K.-Y.-K. Wong, "Face sketch synthesis with style transfer using pyramid column feature," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2018, pp. 485–493.

[38] C. Philip and L. H. Jong, "Face sketch synthesis using conditional adversarial networks," in *Proc. ICTC*, 2017, pp. 373–378.

[39] H. Bi, N. Li, H. Guan, D. Lu, and L. Yang, "A multi-scale conditional generative adversarial network for face sketch synthesis," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 3876–3880.

[40] D.-P. Fan *et al.*, "Scoot: A perceptual metric for facial sketches," in *Proc. ICCV*, 2019, pp. 5611–5621.

[41] R. Yi, Y.-J. Liu, Y.-K. Lai, and P. L. Rosin, "APDrawingGAN: Generating artistic portrait drawings from face photos with hierarchical GANs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 10735–10744.

[42] M. Zhu, N. Wang, X. Gao, J. Li, and Z. Li, "Face photo-sketch synthesis via knowledge transfer," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, Aug. 2019, pp. 1048–1054.

[43] M. Zhang, R. Wang, X. Gao, J. Li, and D. Tao, "Dual-transfer face Sketch–Photo synthesis," *IEEE Trans. Image Process.*, vol. 28, no. 2, pp. 642–657, Feb. 2019.

[44] M. Zhang, N. Wang, Y. Li, and X. Gao, "Neural probabilistic graphical model for face sketch synthesis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 7, pp. 2623–2637, Jul. 2020.

[45] Y.-H. Pao and Y. Takefuji, "Functional-link net computing: Theory, system architecture, and functionalities," *Computer*, vol. 25, no. 5, pp. 76–79, May 1992.

[46] Y.-H. Pao, G.-H. Park, and D. J. Sobajic, "Learning and generalization characteristics of the random vector functional-link net," *Neurocomputing*, vol. 6, no. 2, pp. 163–180, Apr. 1994.

[47] C. L. P. Chen and Z. Liu, "Broad learning system: A new learning paradigm and system without going deep," in *Proc. 32nd Youth Academic Annu. Conf. Chin. Assoc. Autom. (YAC)*, May 2017, pp. 1271–1276.

[48] F. Cao, H. Ye, and D. Wang, "A probabilistic learning algorithm for robust modeling using neural networks with random weights," *Inf. Sci.*, vol. 313, pp. 62–78, Aug. 2015.

[49] S. Cai, L. Zhang, W. Zuo, and X. Feng, "A probabilistic collaborative representation based approach for pattern classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2950–2959.

[50] K. Chen, Q. Lv, Y. Lu, and Y. Dou, "Robust regularized extreme learning machine for regression using iteratively reweighted least squares," *Neurocomputing*, vol. 230, pp. 345–358, Mar. 2017.

[51] M. Yang, L. Zhang, J. Yang, and D. Zhang, "Regularized robust coding for face recognition," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1753–1766, May 2013.

[52] J.-W. Jin and C. L. Philip Chen, "Regularized robust broad learning system for uncertain data modeling," *Neurocomputing*, vol. 322, pp. 58–69, Dec. 2018.

[53] X. Fang, Y. Xu, X. Li, Z. Lai, W. K. Wong, and B. Fang, "Regularized label relaxation linear regression," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 4, pp. 1006–1018, Apr. 2018.

[54] H. Zhou, Z. Kuang, and K. K. Wong, "Markov weight fields for face sketch synthesis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 1091–1097.

[55] S. Van Huffel, "Partial singular value decomposition algorithm," *J. Comput. Appl. Math.*, vol. 33, no. 1, pp. 105–112, Dec. 1990.

[56] Y. Song, L. Bao, Q. Yang, and M.-H. Yang, "Real-time exemplar-based face sketch synthesis," in *Proc. ECCV*, 2014, pp. 800–813.

[57] J. Jiang, Y. Yu, Z. Wang, X. Liu, and J. Ma, "Graph-regularized locality-constrained joint dictionary and residual learning for face sketch synthesis," *IEEE Trans. Image Process.*, vol. 28, no. 2, pp. 628–641, Feb. 2019.

[58] M. Zhang, Y. Li, N. Wang, Y. Chi, and X. Gao, "Cascaded face sketch synthesis under various illuminations," *IEEE Trans. Image Process.*, vol. 29, pp. 1507–1521, 2020.

[59] J. Zhang, X. Yu, A. Li, P. Song, B. Liu, and Y. Dai, "Weakly-supervised salient object detection via scribble annotations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 12543–12552.

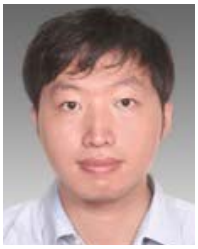
[60] J.-X. Zhao, J.-J. Liu, D.-P. Fan, J.-F. Cao, J. Yang, and M.-M. Cheng, "EGNet: Edge guidance network for salient object detection," in *Proc. ICCV*, 2019, pp. 8778–8787.

[61] D.-P. Fan, M.-M. Cheng, J.-J. Liu, S.-H. Gao, Q. Hou, and A. Borji, "Salient objects in clutter: Bringing salient object detection to the foreground," in *Proc. ECCV*, 2018, pp. 196–212.

[62] J. Zhang *et al.*, "UC-net: Uncertainty inspired RGB-D saliency detection via conditional variational autoencoders," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 8579–8588.

Ping Li (Member, IEEE) received the Ph.D. degree in computer science and engineering from The Chinese University of Hong Kong, Hong Kong, in 2013.

He is currently a Research Assistant Professor with The Hong Kong Polytechnic University, Hong Kong. He has authored or coauthored one image/video processing national invention patent and has excellent research project reported worldwide by *ACM TechNews*. His current research interests include image/video stylization, artistic rendering and synthesis, and creative media.



Bin Sheng (Member, IEEE) received the Ph.D. degree in computer science and engineering from The Chinese University of Hong Kong, Hong Kong, in 2011.

He is currently a Full Professor with the Shanghai Jiao Tong University, Shanghai, China. He is an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY. His current research interests include virtual reality and computer graphics.

C. L. Philip Chen (Fellow, IEEE) received the Ph.D. degree in electrical engineering from Purdue University, West Lafayette, IN, USA, in 1988.

He is currently a Chair Professor and the Dean of the School of Computer Science and Engineering, South China University of Technology, Guangzhou, China. Being a Program Evaluator of the Accreditation Board of Engineering and Technology Education (ABET) in the United States for computer engineering, electrical engineering, and software engineering programs, he successfully architects the University of Macau's Engineering and Computer Science programs receiving accreditations from Washington/Seoul Accord through the Hong Kong Institute of Engineers (HKIE), which is considered as his utmost contribution in engineering/computer science education for Macau as the former Dean of the Faculty of Science and Technology. His current research interests include systems, cybernetics, and computational intelligence.

Dr. Chen is a fellow of AAAS, IAPR, CAA, and HKIE; a member of Academia Europaea (AE), the European Academy of Sciences and Arts (EASA), and the International Academy of Systems and Cybernetics Science (IASCYS). He was a recipient of the IEEE Norbert Wiener Award in 2018 for his contribution to systems and cybernetics, and machine learning. He is also a highly cited researcher by Clarivate Analytics in 2018 and 2019. He was a recipient of the 2016 Outstanding Electrical and Computer Engineers Award from his alma mater, Purdue University (in 1988), after he graduated from the University of Michigan at Ann Arbor, Ann Arbor, MI, USA, in 1985. He was the IEEE Systems, Man, and Cybernetics Society President from 2012 to 2013 and the Editor-in-Chief of the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS from 2014 to 2019. He is currently the Editor-in-Chief of the IEEE TRANSACTIONS ON CYBERNETICS and an Associate Editor of the IEEE TRANSACTIONS ON FUZZY SYSTEMS. He was the Chair of TC 9.1 Economic and Business Systems of the International Federation of Automatic Control from 2015 to 2017. He is currently a Vice President of the Chinese Association of Automation (CAA).