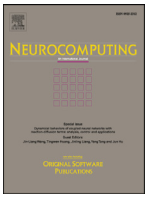
Neurocomputing 386 (2020) 54–62

Contents lists available at ScienceDirect

Neurocomputing

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



Face inpainting network for large missing regions based on weighted facial similarity

Jia Qin[a,b], Huihui Bai[a,b,*], Yao Zhao[a,b]

[a] Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China
[b] Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing 100044, China

ARTICLE INFO

Article history:
Received 6 January 2019
Revised 21 November 2019
Accepted 20 December 2019
Available online 24 December 2019

Communicated by Dr Ronghua Liang

Keywords:
Image inpainting
Face image
Face inpainting
Face similarity
Face inpainting network

ABSTRACT

Recently, deep learning has made great achievements in the field of image inpainting, especially filling the large missing regions based on generative adversarial net (GAN). However, GAN is the model to capture the data distribution rather than image content. Therefore, for the corrupted face image with large holes, it may generate a new image which is greatly different from the original one. To solve above problem, we present a face inpainting network based on weighted face similarity (WFS-Net) to generate a better restoration. Firstly, according to the structural similarity index (SSIM), a weighted similar face set (WSFS) can be built to provide more reference information for the recovery of missing regions. And then, WFS-Net is designed to fill the damaged face image by exploring the relationship between the missing region and the available information, which include the known parts of the damaged image and its WSFS. Furthermore, a new loss function is presented in pixel level and texture level. The experimental results show that our proposed approach outperforms other state-of-the-arts.

1. Introduction

Image inpainting is the task to restore missing or damaged areas in a corrupted image, which has many applications in image editing [1,2], occlusion removing [3–5] and the restoration of old photos [6]. For restoring images more realistically, many researchers focus on this ill-posed inverse problem.

Early inpainting methods are mainly defined as diffusion-based inpainting and examplar-based inpainting. Diffusion-based inpainting [7,8] tends to diffuse the information around the boundary of inpainting region into missing region by partial differential equation (PDE). However they may produce blurring artifacts when inpainting the large missing regions. And in the exemplar-based inpainting methods [5,9,10], they try to copy the suitable image patches into the missing regions, which may be not visually continuous in complex structure. Moreover, both these two kinds of methods are not able to capture high-level semantics.

In recent years, considering that the generative adversarial network (GAN) [11] can predict the missing areas with better visual quality, it is widely applied in the image inpainting, which consists of the global GAN based methods [12–14], local GAN based methods [15,16] and the GANs with both global data distribution and local details [17–19]. For example, Pathak et al. [12] and Yeh et al. [13] complete the corrupted image by global data distribution, Guo et al. [16] fill the missing regions by patch details, while Yu et al. [19] propose the inpainting network by both global image context and small region around the completed area. Nevertheless, since GAN is the method that captures the data distribution [11], the generated face images may have greatly different contents with the original ones. Furthermore, the exist inpainting networks usually fill the missing areas by the known information of damaged image, it is difficult to predict the large missing holes accurately.

Considering that the face is one of the most important human biological features, it is better to obtain the face inpainting results that similar with the original ones. In order to restore the damaged face image accurately, the face inpainting can be considered as a filling of pixels for the damaged image, rather than a capture of data distribution. Therefore, the structure of fully convolutional network rather than GAN is chosen to fill the damaged face images, for it can directly work on the pixel level. In this paper, we propose a face inpainting network based on weighted face similarity (WFS-Net), in which both the known parts of the damaged face images and other similar faces are adopted as the available information. Firstly, according to the image structure, a weighted similar face set (WSFS) is generated to provide more reference information for the recovery of missing regions. And then, WFS-Net is used to fill the big holes of damaged face images. Furthermore,

[*] Corresponding author at: Institute of Information Science, Beijing Jiaotong Univercity, Beijing 100044, China.

E-mail addresses: qinjia@bjtu.edu.cn.com (J. Qin), hhbai@bjtu.edu.cn (H. Bai), yzhao@bjtu.edu.cn (Y. Zhao).

https://doi.org/10.1016/j.neucom.2019.12.079
0925-2312/© 2019 Elsevier B.V. All rights reserved.

for better training, a new loss function is designed to focus on the differences between the recovered images and the original ones in pixel level and texture level.

In summary, the contributions of our work can be described as follows:

- In order to restore the damaged face image accurately, we design a weighted facial similarity based inpainting network (WFS-Net) to fill the large missing regions of face image in pixel level and texture level, in which the weighted similar face set (WSFS) is presented to provide more reference information.
- In this paper, a WSFS is generated to provide more prior similar faces for the recovery of missing regions. Furthermore, these similar faces are weighted as appropriate selections of reference information. Then, WFS-Net is used to explore the relationship between the missing region and more available information in WSFS to fill the large missing regions of face images accurately.
- To restore the face images realistically and accurately, a new loss function is presented in view of pixel level and texture level. Here, L2 loss is used for the restoration of image content in pixel level, while local binary pattern (LBP) is adopted to constrain the training of inpainting model in texture level.

This paper is organized as follows. In Section 2, some works about image inpainting and face similarity measurements is introduces. In Section 3, the details of proposed algorithm will be presented. In Section 4, the experimental results are shown and analyzed. And finally, the conclusion is summarized in Section 5.

2. Related work

2.1. Image inpainting

Early image inpainting algorithms are divided into diffusion-based inpainting [7,8,20–22] and examplar-based inpainting [5,9,10,23]. As one of the pioneer diffusion-base works, Shen et al. [7] propose a variational model to fill the areas involving the recovery of edges, which is closely connected to the total variation (TV) restoration model. And then, Chan et al. [8] propose a curvature-driven diffusions (CDD) based inpainting model to improve TV inpainting on the connectivity principle. Although, the diffusion-based inpainting methods can ensure local intensity smoothness, they may tend to produce blurring artifacts when filling the large missing regions. For the examplar-based inpainting methods, they can synthesize plausible stationary textures. Criminisi et al. [5] propose an examplar-based inpainting algorithm for removing large objects in a visually plausible way. Meur et al. [9] introduce a examplar-based inpainting framework, in which a coarse version of the input image is improved by the hierarchical super-resolution algorithm. Kumar et al. [10] present a formulation of exemplar-based image inpainting, which maintains better visual consistency in the inpainted region by the simulated annealing algorithm. However, these examplar-based methods may make critical failures in visually continuity in the region with complex texture or image structure.

Over the past few years, due to the rapid development of GANs, GAN-based methods have shown encouraging inpainting results. Some methods [12–14,24,25] tend to pay more attention to global data distribution of the filled image, in which the data distribution are considered as an important inpainting constrain to recognize global consistency of the scene. Here, Pathak et al. [12] present an unsupervised visual feature learning algorithm driven by context-based pixel prediction, which can generate the contents of an arbitrary image region conditioned on its surroundings. Subsequently, Yeh et al. [13] consider the semantic inpainting as a constrained image generation problem, in which the closest encoding of a corrupted image is searched by the context and prior losses. Since these global-based inpainting methods may ignore the local details of the corrupted image, many local-based inpainting networks [15,16,26,27] are proposed to focus on the patch data distributions or the small regions around the completed area. Here, Guo et al. [16] embed the progressive inpainting policy into the image inpainting to complete missing regions naturally. Yu et al. [26] present a generative image inpainting system to complete images with free-form mask and guidance. And in order to combine both the global semantics and the local context, the joint local and global networks [17–19,28–30] is presented. Here, Yu et al. [19] present a deep generative model-based approach is proposed, which can not only synthesize image structures but also utilize surrounding image features to make better predictions. Yang et al. [30] propose a multi-scale neural patch synthesis approach based on image content and texture constraints to preserve contextual structures. Although these GAN based methods are able to improve face inpainting results by capturing same data distribution, it may be greatly different in contents with the original faces. As one of the most important human biological features, it is crucial to obtain similar inpainting results.

2.2. Face similarity measurement

Many existing methods are used to explore the facial similarity by the features [31–34] and the nonlinear genetic traits [35–37]. Rahim et al. [31] test the resemblance of faces with fisher linear discriminant algorithm. Luo [32] uses the person-specific scale-invariant feature transform (person-specific SIFT) and a simple nonstatistical matching strategy to solve face recognition problems. Deng et al. [33] develop a transform-invariant principal components analysis (TIPCA) technique which aims to accurately characterize the intrinsic structures of the human face. Although, these facial feature based methods have been extensively investigated in face similarity, they still exist some limitations. As for image inpainting, it is hard to capture the complete facial features for the large holes in the corrupted face image.

Some face similarities are explored according to the nonlinear genetic features. Zhou et al. [35] propose a new kinship metric learning (KML) method to learn a coupled deep similarity metric, in which the images with kinship relation are pulled close. Mahpod et al. [36] propose a multiview hybrid combined symmetric and asymmetric distance learning network for facial similarity. Liu et al. [37] propose a status-aware projection learning (SaPL) method for facial image based parent-child kinship verification. However, there exists instability in the similarity of kin in appearance, especially in the face image with large missing regions.

In order to put more attention on image structure, SSIM is introduced in our work for more robust and effective selection of similar faces. Furthermore, because of the superiority of local binary patterns (LBP) in human face image processing [38–42], it is also adopted as the texture feature to constrain the training of inpainting model in our work.

3. The proposed method

In this section, the proposed WFS-Net will be elaborated. Firstly, an overview of WFS-Net is provided, in which the synthesis of the missing contents is introduced briefly. Then, the generation of WSFS is analyzed, which are considered as the prior information of damaged images. Furthermore, the network structure of WFS-Net will be described in detail, in which the WSFS and damaged images are used as the inputs. Finally, the loss function of WFS-Net and the filling of damaged face images will be discussed.

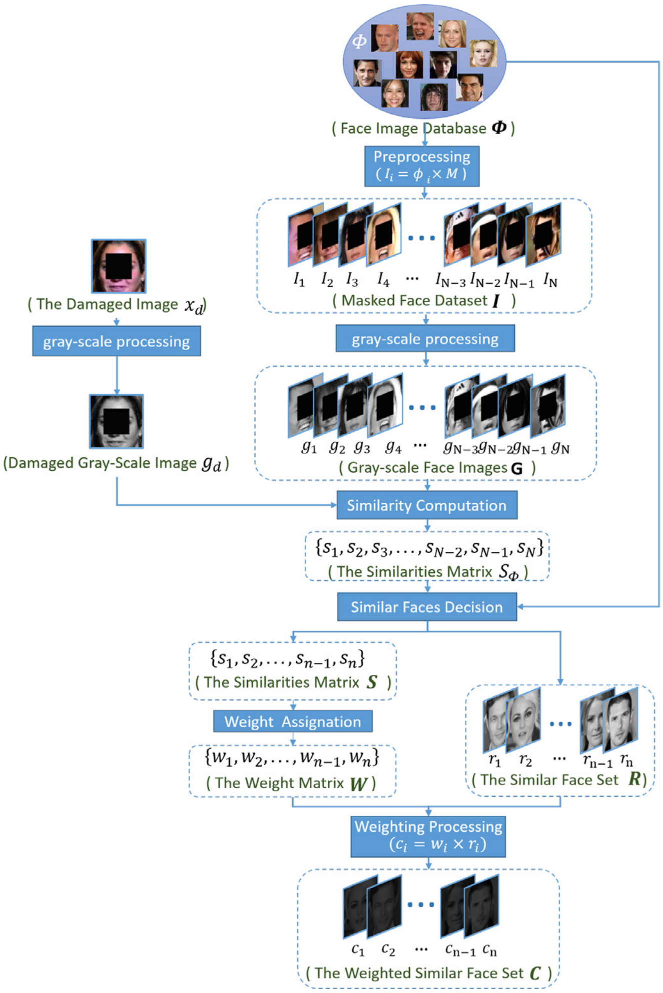
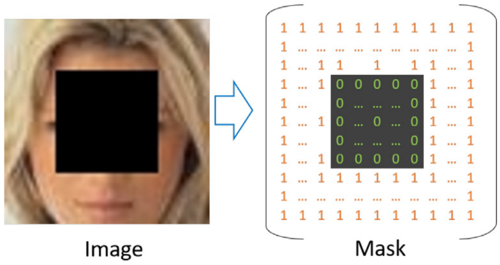

(Face Image Database Φ)
Preprocessing
($I_i = \phi_i \times M$)
I_1 I_2 I_3 I_4 ··· I_{N-3} I_{N-2} I_{N-1} I_N
(Masked Face Dataset I)
(The Damaged Image x_d)
gray-scale processing
gray-scale processing
g_1 g_2 g_3 g_4 ··· g_{N-3} g_{N-2} g_{N-1} g_N
(Gray-scale Face Images G)
(Damaged Gray-Scale Image g_d)
Similarity Computation
$\{s_1, s_2, s_3, \ldots, s_{N-2}, s_{N-1}, s_N\}$
(The Similarities Matrix S_Φ)
Similar Faces Decision
$\{s_1, s_2, \ldots, s_{n-1}, s_n\}$
(The Similarities Matrix S)
Weight Assignation
$\{w_1, w_2, \ldots, w_{n-1}, w_n\}$
(The Weight Matrix W)
r_1 r_2 ··· r_{n-1} r_n
(The Similar Face Set R)
Weighting Processing
$(c_i = w_i \times r_i)$
c_1 c_2 ··· c_{n-1} c_n
(The Weighted Similar Face Set C)

Fig. 1. The architecture of MFS-Net.

3.1. WFS-Net

To restore the face images accurately, the convolutional network is considered in our work as the inpainting method for it can directly work on the image content rather than data distribution. However, because of the local convolutional kernel, it is not effective to restore the corrupted image with large missing regions [19]. Motivated by this reason, a face inpainting network based on the weighted face similarity is proposed, in which the similar faces are introduced as the reference information. The architecture of WFS-Net shows in Fig. 1, in which the proposed algorithm is divided into two stages. The first stage is the generation of WSFS $C = \{c_1, \ldots, c_i, \ldots, c_n\}$, in which n is the size of the Similar Face Set (SFS). In this stage, the similar faces are firstly selected from a face dataset. Considering that there are different similarities between each damaged face and the face images in C, the weight assignment for each face image in WSFS is necessary to reduce the negative influences caused by these differences. The second stage is the design of WFS-Net. In this process, for providing the reference information, both the WSFS and the damaged image are considered as the inputs of WFS-Net. When the WFS-Net is trained, the features of both the pixel level and the texture level are considered in the loss function on this model. Finally, the missing regions are filled with the co-located regions of the generated image from the WFS-Net.



$$\begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1\\
1 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & 1\\
1 & \ldots & 1 & 1 & & 1 & & 1 & 1 & \ldots & 1\\
1 & \ldots & 1 & 0 & 0 & 0 & 0 & 0 & 1 & \ldots & 1\\
1 & \ldots & & 0 & \ldots & \ldots & \ldots & 0 & 1 & \ldots & 1\\
1 & \ldots & 1 & 0 & \ldots & 0 & \ldots & 0 & 1 & \ldots & 1\\
1 & \ldots & & 0 & \ldots & \ldots & \ldots & 0 & 1 & \ldots & 1\\
1 & \ldots & 1 & 0 & 0 & 0 & 0 & 0 & 1 & \ldots & 1\\
1 & \ldots & 1 & 1 & 1 & 1 & 1 & 1 & 1 & \ldots & 1\\
1 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & 1\\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{pmatrix}$$

Fig. 3. Mask matrix.

3.2. Generation of weighted similar face set

For providing the prior information for inpainting network, the WSFS is collected for each damaged face image. Here, in order to avoid an extra face dataset, the training samples are divided into two parts: the face dataset Φ for the generation of WSFS and the training sample for WFS-Net. The selection of the SFS shows in Fig. 2, in which $\Phi = \{\varphi_1, \ldots, \varphi_i, \ldots, \varphi_N\}$ is the face dataset to provide the reference information for a damaged image.

It's remarkable that before the similarity computation, a preprocessing for the face images in face dataset is crucial, which makes sure that the comparison of the similarity is only between the available regions of the damaged image and the co-located regions of face images in face dataset Φ. Firstly, a mask matrix is defined to indicate the missing parts, whose size equals the damaged image. An example of mask matrix is shown in Fig. 3, in which the missing regions are equal to 0, and the available regions are equal to 1. The equation of this preprocess shows as follows:

$$I_i = \varphi_i \odot M \tag{1}$$

where $\odot$ denotes the element-wise multiplication. M is the mask matrix, φ_i is the face image in Φ. And I_i is the masked image of the masked face dataset **I**, in which the missing regions of the damaged image and the corresponding regions of the face images in Φ are valued 0.

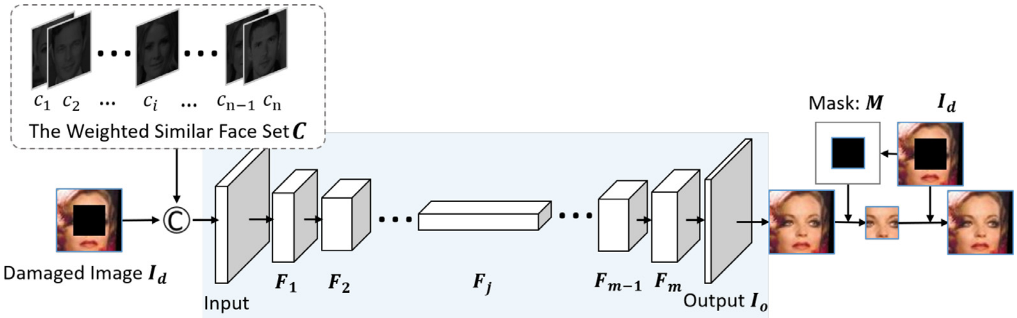


Fig. 2. The generation of weighted similar face set (WSFS).

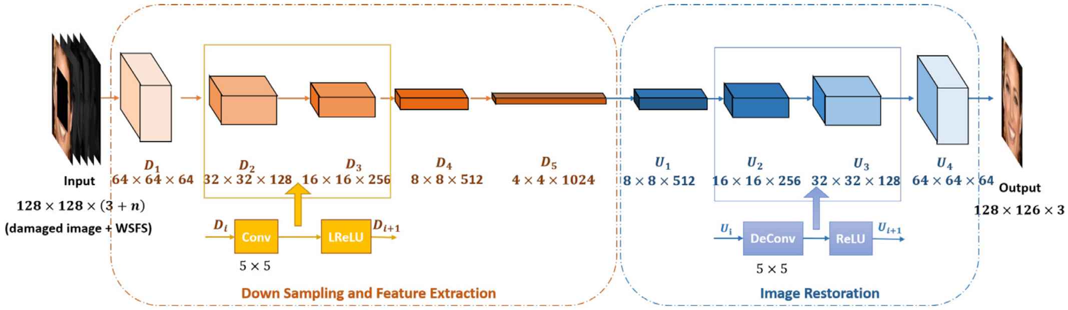


Fig. 4. The structure of weighted facial similarity based inpainting network (WFS-Net).

After the pre-processing, similarities between the damaged image and the images in Φ can be calculated for the generation of a WSFS. Here, the similarities of the damaged image and the masked images in the masked face dataset **I** are guided by the measure of SSIM for its effectiveness on similarity estimation, which can compare two face images from luminance, contrast and structure. Additionally, it is worthy of note that the color information of these similar faces is unnecessary, which may have a negative impact on the color consistency of filled image. For avoiding the negative impact caused by the color information of SFS, only the similarities of gray-scale map between the damaged image g_d and the images in G are calculated as follows:

$$s(g_d, g_i) = \frac{(2u_{g_d} + V_1)(2\sigma_{g_d g_i} + V_2)}{(u_{g_d}^2 + u_{g_i}^2 + V_1)(\sigma_{g_d}^2 + \sigma_{g_i}^2 + V_2)} \tag{2}$$

where

$$g_i = Gray(\varphi_i \odot M) \tag{3}$$

Here, u_{g_d} and u_{g_i} are the mean intensity of g_d and g_i. σ_{g_d} and σ_{g_i} are the standard deviations of these two faces g_d and g_i. V_1 and V_2 are constants to avoid instability when $(u_{g_d}^2 + u_{g_i}^2)$ or $(\sigma_{g_d}^2 + \sigma_{g_i}^2)$ are very close to zero. After the calculation of similarity, n most similar faces $R = \{r_1, \ldots, r_i, \ldots, r_n\}$ are selected as the SFS of the damaged image and the corresponding similarity matrix is $S = \{s_1, \ldots, s_i, \ldots, s_n\}$. Furthermore, considering the different similarities between the damaged face and its SFS, normalized weights of the similarity matrix S should be assigned for each similar face image, which can be calculated as follows:

$$w_i = \frac{s_i}{\sum_{a=1}^{n}(s_a)} \tag{4}$$

where s_i is the similarity of the ith similar face in SFS and $W = \{w_1, \ldots, w_i, \ldots, w_n\}$ is the weight matrix. Finally, the WSFS for a training sample $C = \{c_1, \ldots, c_i, \ldots, c_n\}$ is achieved:

$$c_i = w_i \times r_i \tag{5}$$

where c_i is the ith weighted similar face in WSFS. For each training sample, a WSFS can be found. Assuming that the size of the training samples is Num, the WSFS for all training sample is $\mathbb{C} = \{C_1, \ldots, C_i, \ldots, C_{Num}\}^T$, which size is $Num \times n$.

3.3. WSFS based face inpainting network

The design of WFS-Net is inspired by the fully convolutional network, which takes the input of arbitrary size and produces correspondingly-sized output with efficient inference and learning [43]. However, because of the local convolutional kernels, convolutional layers are not an effective method for filling the large missing regions [19]. Aiming at this problem, the WSFS is introduced as the reference information for restoring the damaged face images realistically. Here, the structure of inpainting network designed for a 128 × 128 image is shown as Fig. 4, in which the successive convolution layers and deconvolution layers are used to replace the operator of upsampling and downsampling for the more precise outputs [43]. In the network, the downsampling layers are used to extract the feature of the available information which consist of damaged image and its WSFS. And the upsampling is used to predict the content of missing region according to the extracted features. In this processing, the relationship between the large hole and the available parts is obtained. Here, we use five convolutional layers with kernel size of 5 × 5 for downsampling and feature extraction, which can be represented as:

$$D_i = \tau(f_i(D_{i-1})), (0 < i \le 5) \tag{6}$$

where f_i is the ith function of feature extraction. And D_i is the extracted features of f_i. The input is represented with D_0, which includes the damaged image and its WSFS. Furthermore, in the part of downsampling and feature extraction, the Leaky Rectified Linear Unit (Leaky ReLU) [44] is used as activation function for keeping the negative input by assigning none zero output, which is represented by τ. In the part of image restoration, five deconvolutional layers with kernel size of 5 × 5 are used for upsampling, which are denoted as follows:

$$U_i = \mu(h_i(U_{i-1})), (0 < i \le 5) \tag{7}$$

where h_i is the ith deconvolutional layer for image restoration. U_i is the feature maps of h_i. In Fig. 4, D_5 can be considered as the input U_0 in the network of image restoration. These deconvolutional layers in part of image restoration are activated by ReLU for the sparse representations and the efficiency [45], which is represented by μ.

Benefiting from the structure of feature extraction and the image restoration in WFS-Net, our network can explore the relationship between the damaged images and its WSFS effectively.

3.4. Loss function

To restore the face images realistically and accurately, the WFS-Net is trained by the loss function, which is combined with the effective reconstruction (L_2) loss and the texture information. The L_2 loss is defined as follows:

$$L_2 = \frac{\|x^* - x\|_2}{W \times H \times C} \tag{8}$$

where x^* is the output of WFS-Net and x is the original face image. C, H and W are the height, width and channel size of face image. L_2 loss is widely used in the image inpainting and it can capture the overall structure of the missing region in relation to the context [12]. However, since L_2 tends to average together the difference between the output and the ground truth, it may ignore some texture information. Therefore, for the sharper texture, we introduce the texture feature in the loss function. Here, LBP is considered as the texture feature for its superiority in face recognition

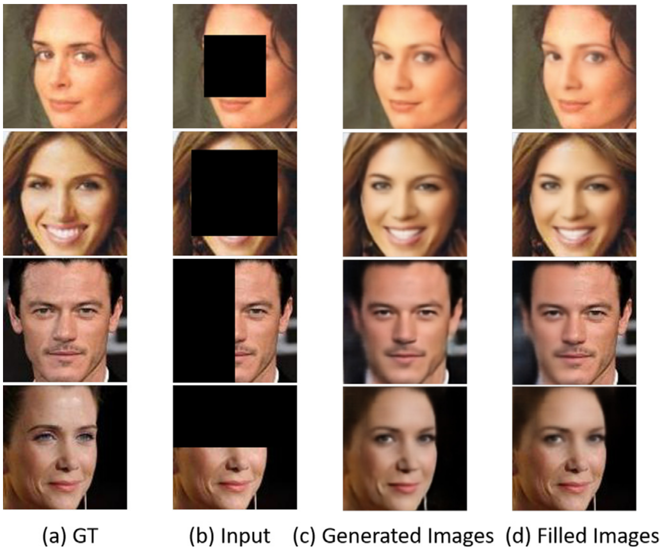


Fig. 5. The comparison of the generative images of WFS-Net and the filled images: (a) the original images, (b)the masked images, (c) the generated images and (d) the filled images.

[42]. After the calculation of LBP, the results T_{x^*} and T_x are obtained to represent the texture information of generative face image and original ones, which have the same size with the damaged image. For getting the same data range with the L_2 loss, the normalization is essential, which is shown as follows:

$$N(i,j) = \frac{T(i,j) - min(T)}{max(T) - min(T)} \tag{9}$$

where T is LBP matrix and N is the result of normalization. Finally, the normalized LBP feature is introduced in the loss function, which is calculated as follows:

$$L = L_2 + \sum_{i=1}^{H}\sum_{j=1}^{W}(N_{x^*}(i,j) - N_x(i,j))^2 \tag{10}$$

where W is the width of face image and H is the hight of the image. N_{x^*} and N_x are the normalized LBP feature of generated face image and its corresponding face respectively.

3.5. Filling of damaged face image

After achieving the generated images from the WFS-Net, a post-processing is necessary to restore the damaged image. And the filled face image can be obtained by:

$$\hat{x} = M \odot x_d + (1 - M) \odot x^* \tag{11}$$

where x_d is the damaged image. x^* is the image generated by the proposed WFS-Net. Finally, $\hat{x}$ is the result of filled image. $\odot$ is the element-wise multiplication. Fig. 5 is the comparison of the generative images of WFS-Net and the filled images.

4. Experimental results

In order to show the validity of WFS-Net, we evaluate it on the dataset Celeb Faces Attributes Dataset (CelebA) [46] and Labeled Faces in the Wild (LFW) [47]. Then, an appropriate size of face image datasets Φ is selected to balance the computational complexity and the similarity of face images in WSFS. Additionally, we carry out extensive experiments to discuss the importance of reference information and the texture feature in loss function. Furthermore, the proposed face inpainting method is compared with several state-of-the-art algorithms, which are also based on deep learning. Finally, we display some results about occlusion removal.

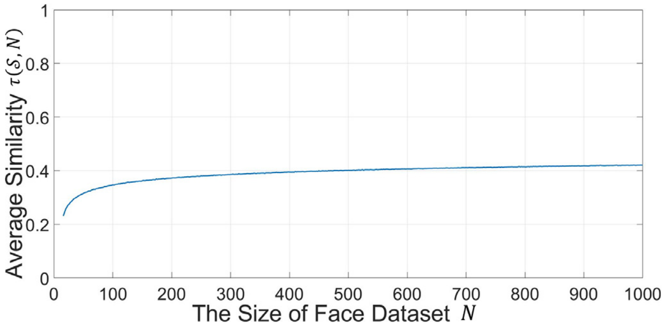


Fig. 6. The relationship between the size of face dataset N and $\tau(N)$.

4.1. Dataset and missing regions

CelebA is a large-scale face attributes dataset with 202,599 face images. In this paper, 2000 images are used for testing the trained model. 400 face images are used as face image database Φ for building the WSFS and providing the reference information, in which the size selection of Φ is discussed in the following section. Other 200,199 face images in CelebA are used as training samples. Furthermore, in order to show the effectiveness of our model, we also evaluate it on LFW dataset, which consists of 5749 face images. It illustrates that our trained model has also achieved the improvements on other face dataset. In this paper, the face images are cropped to 128 × 128.

In addition, the proposed method is tested in the different missing regions as shown in Fig. 5, which is 25% and 50% in images. As for the 25% missing regions, the missing parts are in the center of the image (25% center mode). And as for the 50% missing region, the missing parts are on the center of images (50% center mode), the left of images (vertical mode) and the up of images (horizontal mode), respectively. Additionally, we also test the proposed WFS-Net by removing the irregular occlusion.

For the experimental results, peak signal-to-noise ratio (PSNR), SSIM and semantic similarity [29] are used as the metrics of filling results. PSNR represents the difference of pixel level. SSIM is the estimates of holistic similarity between the original image and the filled image. Furthermore, identity distance is calculated by the OpenFace [48], which is the similarity of two faces in semantic. A smaller identity distance means a higher similarity of two faces, while a higher identity distance means that these two faces are not similar.

4.2. The size selection of face image datasets Φ

In Fig. 2, $\Phi = \{\varphi_1, \ldots, \varphi_i, \ldots, \varphi_N\}$ is the dataset to provide the similar faces for a damaged image. The size of face dataset is important for the selection of similar faces. Here, we assumed the face image dataset Φ with the size of N. If N is too large, it will make the inpainting difficult to realize because of the enormous computational complexity of similar face selection. And if the size of the face dataset is too small, it is hard to find the similar faces. In order to find an appropriate size for face dataset, we selected 500 damaged faces randomly from the training sample as a Training Subset (TS), which are used to observe the relationship between the similarities of SFS and the size of Φ. Here, the size of Φ is N while the size of SFS is n and $n = 4$. We calculate the similarities of each damaged face image in TS and the images in Φ, and the similarity matrix $\mathbb{S} = \{S_1, \ldots, S_p, \ldots, S_{500}\}^T$ is generated with size of 500 × n, in which $S_p = \{s_{(p,1)}, \ldots, s_{(p,q)}, \ldots, s_{(p,n)}\}$ is the SFS of p^{th} image in TS. And the equation shows as follows:

$$\tau(\mathbb{S}, N) = \frac{\sum_{p=1}^{500}(\sum_{q=1}^{n} s_{(p,q)})}{500 \times n}, n \leq N \tag{12}$$

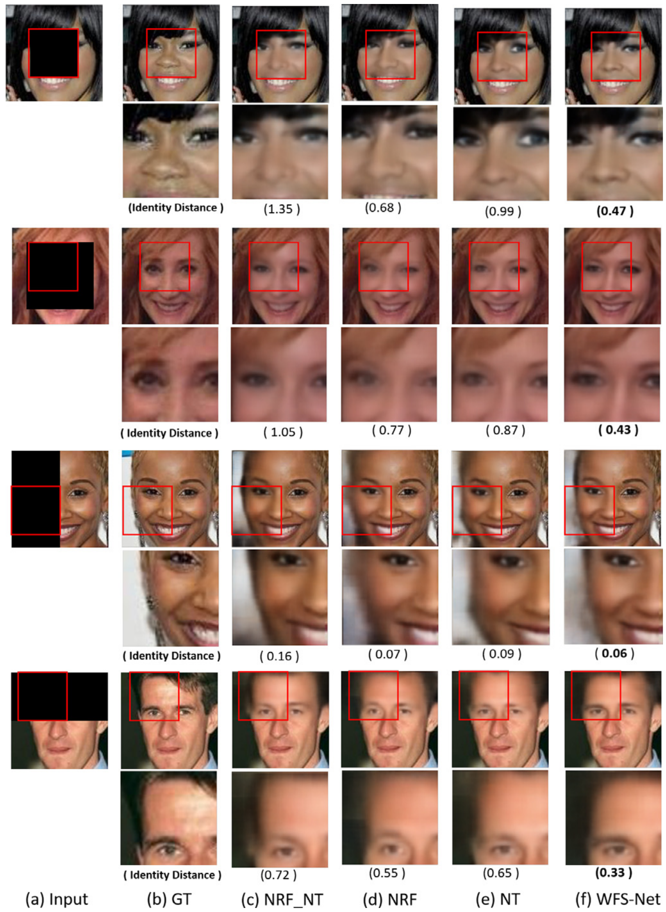


Fig. 7. The comparison of the results of the NRF_NT, the NRF, the NT and the proposed MFS-Net: (a) Input: the masked image; (b) GT: the ground truth image; (c) NRF_NT: the results with neither reference information nor texture feature; (d) NRF: the results without reference information WSFS; (e) NT: the results without texture information LBP; (f) WFS-Net: the results of proposed method, in which the WSFS is used as the reference information and the LBP is introduced in the loss function as the texture information.

Table 1
Comparison between CE [12], SemGAN [13], CA [19] and FRRN [16] and the proposed algorithm on CelebA dataset.

Missing Regions			PSNR (dB)	SSIM	Identity Distance
Center Mode	25%	CE	26.01	0.87	0.88
		SemGAN	21.85	0.80	1.36
		CA	21.51	0.83	1.02
		FRRN	27.24	0.90	0.67
		Ours	**28.59**	**0.92**	**0.65**
	50%	CE	21.47	0.71	1.16
		SemGAN	17.95	0.62	1.63
		CA	20.94	0.69	1.42
		FRRN	20.78	0.70	1.14
		Ours	**24.26**	**0.81**	**0.93**
Vertical Mode	50%	CE	18.94	0.71	0.35
		SemGAN	13.54	0.62	0.80
		CA	17.99	0.71	0.32
		FRRN	18.80	0.72	0.42
		Ours	**19.54**	**0.76**	**0.26**
Horizontal Mode	50%	CE	18.30	0.68	0.70
		SemGAN	16.15	0.65	0.88
		CA	17.73	0.69	0.63
		FRRN	18.66	0.72	0.61
		Ours	**19.80**	**0.76**	**0.48**

in which,

$$\mathbb{S} = \{S_1, \ldots, S_p, \ldots, S_{500}\}^T$$
$$= \begin{bmatrix} s_{(1,1)} & \cdots & s_{(1,q)} & \cdots & s_{(1,n)} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ s_{(p,1)} & \cdots & s_{(p,q)} & \cdots & s_{(p,n)} \\ s_{(500,1)} & \cdots & s_{(500,q)} & \cdots & s_{(500,n)} \end{bmatrix} \tag{13}$$

Here, τ is the mean of the similarity matrix $\mathbb{S}_{500\times n}$. Fig. 6 shows the relationship between N and $\tau(N)$, and it shows that 400 is a good choice for N.

4.3. Ablation study

In this paper, we introduce the similar faces as the reference information for recovering the faces accurately. In order to show the efficiency of WSFS, we compare the results of WFS-Net with WSFS and the results without reference information. In addition, considering that L_2 loss is the average together the difference and it is

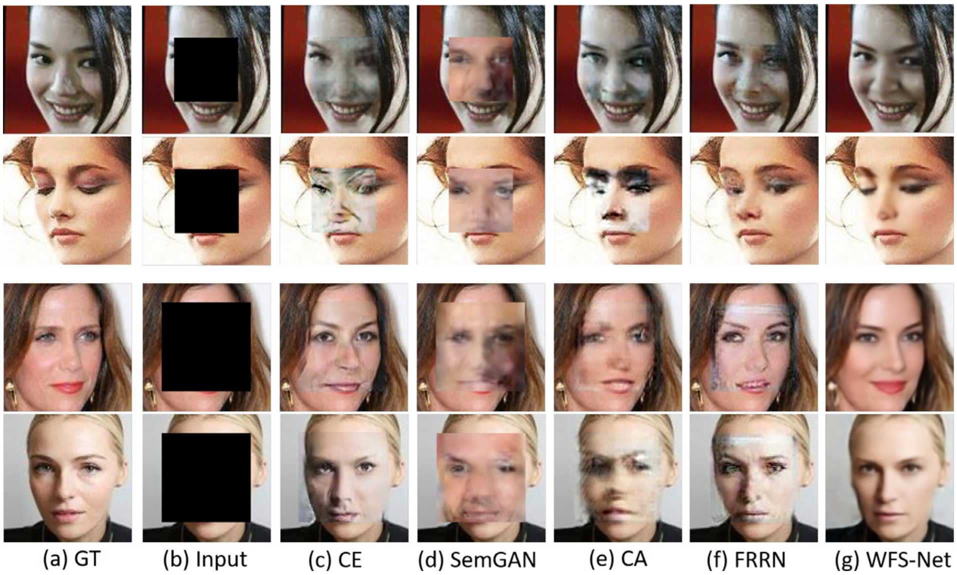


Fig. 8. Center-mode comparisons with the state-of-the-arts on CelebA: CE [12], SemGAN [13], CA [19] and FRRN [16].

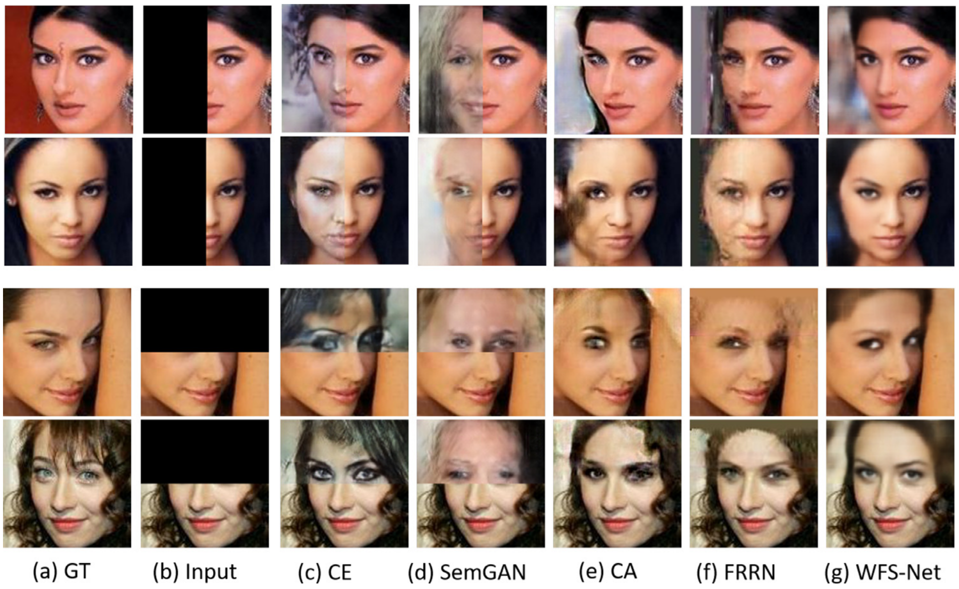


Fig. 9. Comparisons of vertical-mode and horizontal-mode results with the state-of-the-arts on CelebA: CE [12], SemGAN [13], CA [19] and FRRN [16].

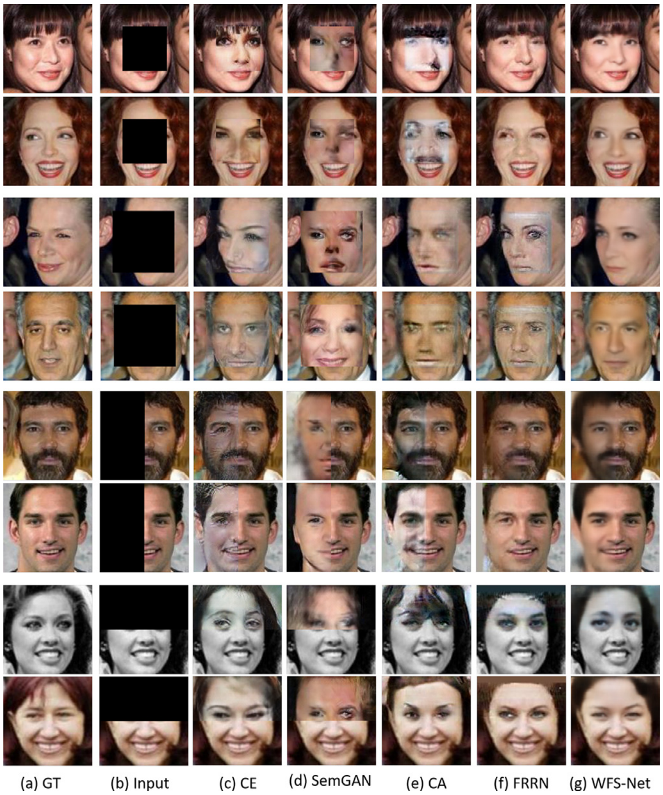


Fig. 10. Comparisons with the state-of-the-arts on LFW: CE [12], SemGAN [13], CA [19] and FRRN [16].

easy to ignore the texture information, a feature LBP is used for a sharper texture of the face images filled by WFS-Net. Therefore, we also compared the results only using the L_2 loss with the results using not only L_2 loss but also texture information. And Fig. 7 show part of the results, in which GT is the ground truth image, NRF_NT means the results without reference information and without texture feature, NRF is the results predicted without reference information WSFS, NT is the results that do not consider texture information and the last column is the results filled by the proposed WFS-Net. From this figure, we can obviously see that the results of the proposed method are much realistic and have more shaper texture. Especially in the magnified areas, the proposed method can be found to recover more details of filled face images.

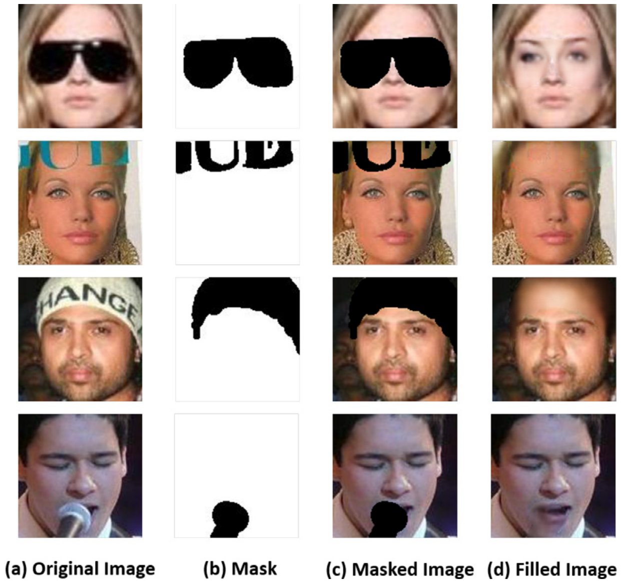


Fig. 11. The results of removing occlusions by our proposed method.

4.4. Comparisons with the state-of-the-arts

In order to evaluate our proposed WFS-Net, the model is compared with context-encoder (CE) [12], semantic image inpainting network (SemGAN) [13], generative inpainting with contextual attention (CA) [19] and full-resolution residual network (FRRN) [16], in which CE and SemGAN are the global GAN based methods, FRRN is the local GAN based method and CA is the method combined global image context and local details. For the datasets, the same pre-processing steps are used before the training of our model and the comparisons. Table 1 lists the comparison results of face inpainting, in which 2000 test images with the size of 128 × 128 are tested. From this table, the proposed WFS-Net outperforms than the other algorithms. As shown in Figs. 8 and 9, the proposed algorithm is more realistic and more accurate in the different mask mode. Furthermore, as shown in Fig. 10, we also test our trained model on LFW dataset, which illustrate that it is also effective on other face dataset.

4.5. Occlusion removal

In a face image, it may be occluded by the sunglasses, hats or other objects, which will prevent us fully observing the details of this face. Therefore, one of the main task of image inpainting is to remove unwanted occlusion. Here, we also show some examples of occlusion removal in Fig. 11, in which it can be seen that our method can fill the obscured face image naturally.

5. Conclusion

In this paper, a face inpainting network WFS-Net based on the face similarity is proposed for better restoration of the corrupted image with large holes, which focuses on filling a damaged face image that similar with the original one. And in order to restore the face realistically and accurately, a weighted similar face set is introduced into the inpainting network as prior information. And considering that L_2 loss tends to average together the difference between the output and the ground truth, which may ignore some texture information, we also introduce a texture feature LBP into the loss function for shaper texture. Finally, we carry out extensive experiments and compared the proposed WFS-Net with other state-of-the-art algorithms on pixel-level, holistic similarity and semantic similarity. And the experimental results demonstrated the superior performance on face inpainting.

Declaration of Competing Interest

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

CRediT authorship contribution statement

Jia Qin: Writing - review & editing, Software, Validation, Data curation, Writing - original draft, Conceptualization, Methodology. **Huihui Bai:** Writing - review & editing, Supervision, Conceptualization. **Yao Zhao:** Supervision, Project administration.

Acknowledgements

This work was supported in part by National Natural Science Foundation of China (No. 61972023) and Fundamental Research Funds for the Central Universities (2019JBZ102 and 2018YJS027).

References

[1] C. Barnes, E. Shechtman, A. Finkelstein, D.B. Goldman, Patchmatch: a randomized correspondence algorithm for structural image editing, ACM Trans. Graph. (TOG) (2009).
[2] A. Levin, A. Zomet, S. Peleg, Y. Weiss, Seamless image stitching in the gradient domain, in: Proceedings of the European Conference on Computer Vision (ECCV), 2004, pp. 377–389.
[3] A. Criminisi, P. Perez, K. Toyama, Object removal by examplar-based inpainting, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2003, pp. 721–728.
[4] H. Zhang, Y. Dong, Q. Fan, Wavelet frame based poisson noise removal and image deblurring, Signal Process. (2005) 363–372.
[5] A. Criminisi, P.P. erez, K. Toyama, Region filling and object removal by exemplar-based image inpainting, IEEE Trans. Image Process. 13 (9) (2004) 1200–1212.
[6] R. Chang, Y. Sie, S. Chou, T.K. Shih, Photo defect detection for image inpainting, in: Proceedings of the IEEE International Symposium on Multimedia (ISM), 2005.
[7] J. Shen, T.F. Chan, Mathematical models for local nontexture inpaintings, SIAM J. Appl. Math. (2002) 1019–1043.
[8] T.F. Chan, J. Shen, Nontexture inpainting by curvature-driven diffusions, J. Vis. Commun. Image Represent 12 (4) (2001) 436–449.
[9] O.L. Meur, M. Ebdelli, C. Guillemot, Hierarchical super-resolution-based inpainting, IEEE Trans. Image Process. 22 (10) (2013) 3779–3790.
[10] V. Kumar, J. Mukherjee, S.K.D. Mandal, Image inpainting through metric labeling via guided patch mixing, IEEE Trans. Image Process. 25 (11) (2016) 5212–5226.
[11] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, Generative adversarial nets, in: Proceedings of the International Conference on Neural Information Processing Systems, 2014, pp. 2672–2680.
[12] D. Pathak, P. Krahenbuhl, J. Donahue, Context Encoders: Feature learning by inpainting, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2536–2544.
[13] R.A. Yeh, C. Chen, T.Y. Lim, Semantic image inpainting with deep generative models, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, 2017, pp. 6882–6890.
[14] S. Xu, D. Liu, Z. Xiong, Edge-guided generative adversarial network for image inpainting, in: Proceedings of the IEEE Visual Communications and Image Processing (VCIP), 2017, pp. 1–4.
[15] K. Nazeri, E.N. E, T. Joseph, EdgeConnect: Generative image inpainting with adversarial edge learning, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2019.
[16] Z. Guo, Z. Chen, T. Yu, J. Chen, S. Liu, Progressive image inpainting with full-resolution residual network, in: Proceedings of the ACM International Conference on Multimedia (ACM MM), 2019, pp. 2496–2504.
[17] T. Ojala, M. Pietikainen, T. Maenpaa, Gray scale and rotation invariant texture classification with local binary patterns, IEEE Trans. Pattern Anal. Mach. Intell. 39 (4) (2014) 640–651.
[18] Y. Wang, X. Tao, X. Qi, X. Shen, J. Jia, Image inpainting via generative multi-column convolutional neural networks, in: Proceedings of the Advances in Neural Information Processing Systems, 2018, pp. 331–340.
[19] J. Yu, Z. Lin, J. Yang, Generative image inpainting with contextual attention, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018.
[20] N. Amrani, J. Serra-Sagrista, P. Peter, Diffusion-based inpainting for coding remote-sensing data, IEEE Geosci. Remote Sens. Lett. (2017).
[21] J. Zhang, D. Zhao, R. Xiong, Image restoration using joint statistical modeling in a space-transform domain, IEEE Trans. Circuits Syst. Video Technol. 24 (6) (2014) 915–928.
[22] U.V. Boscain, R. Chertovskih, J.P. Gauthier, Highly corrupted image inpainting through hypoelliptic diffusion, J. Math. Imaging Vis. (2018).
[23] T. Ružić, A. Pižurica, Context-aware patch-based image inpainting using markov random field modeling, IEEE Trans. Image Process. 24 (1) (2015) 444–456.
[24] Z. Yan, X. Li, M. Li, W. Zuo, S. Shan, Shift-Net: Image inpainting via deep feature rearrangement, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018.
[25] Y. Ren, X. Yu, R. Zhang, T.H. Li, S. Liu, G. Li, StructureFlow: Image inpainting via structure-aware appearance flow, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2019.
[26] J. Yu, Z. Lin, J. Yang, Free-form image inpainting with gated convolution, Proceedings of the IEEE International Conference on Computer Vision (ICCV) (2019).
[27] H. Liu, B. Jiang, Y. Xiao, C. Yang, Coherent semantic attention for image inpainting, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2019.
[28] I. Satoshi, S. Edgar, I. Hiroshi, Globally and locally consistent image completion(2017).
[29] Y. Li, S. Liu, J. Yang, Generative face completion, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, 2017, pp. 5892–5900.
[30] C. Yang, X. Lu, Z. Lin, High-resolution image inpainting using multi-scale neural patch synthesis, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1, 2017.
[31] R. Rahim, T. Afriliansyah, H. Winata, Research of face recognition with fisher linear discriminant 300 (2018) 1532–1546.
[32] J. Luo, Person-specific sift features for face recognition, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2007.
[33] W. Deng, J. Hu, J. Lu, Transform-Invariant PCA: a unified approach to fully automatic face alignment, representation, and recognition, IEEE Trans. Pattern Anal. Mach. Intell. 36 (6) (2014) 1275–1284.
[34] V. Blanz, T. Vetter, Face recognition based on fitting a 3D morphable model, IEEE Trans. Pattern Anal. Mach. Intell. 25 (9) (2003) 1063–1074.
[35] X. Zhou, K. Jin, M. Xu, G. Guo, Learning deep compact similarity metric for kinship verification from face images, Inf. Fus. (2019) 84–94.
[36] S. Mahpod, Y. Keller, Kinship verification using multiview hybrid distance learning, Comput. Vis. Image Understand. (2018) 28–36.
[37] H. Liu, J. Cheng, F. Wang, Kinship verification based on status-aware projection learning, Proceedings of the IEEE International Conference on Image Processing (ICIP) (2017) 1072–1076.
[38] L. Zhang, R. Chu, S. Xiang, Face detection based on multi-block LBP representation, Lect. Notes Comput. Sci. (2007) 11–18.
[39] C. Li, W. Wei, J. Li, A cloud-based monitoring system via face recognition using gabor and CS-LBP features, J. Supercomput. 73 (4) (2017) 1532–1546.
[40] A. Nazari, S.B. Shouraki, A constructive genetic algorithm for LBP in face recognition, in: Proceedings of the International Conference on Pattern Recognition & Image Analysis (IPRIA), 2017.

[41] J. Olivares-Mercado, K. Toscano-Medina, G. Sanchez-Perez, Face recognition system for smartphone based on LBP, in: Proceedings of the International Workshop on Biometrics & Forensics, 2017.
[42] Z. Guo, L. Zhang, D. Zhang, Hierarchical multiscale LBP for face and palmprint recognition, in: Proceedings of the IEEE International Conference on Image Processing, IEEE, 2010, pp. 4521–4524.
[43] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 39 (4) (2014) 640–651.
[44] B. Xu, N. Wang, T. Chen, M. Li, Empirical evaluation of rectified activations in convolutional network, Comput. Sci. (2015). 1–1
[45] X. Glorot, A. Bordes, Y. Bengio, Deep sparse rectifier neural networks, Journal of Machine Learning Research 15 (2010) 315–323.
[46] Z. Liu, P. Luo, X. Wang, X. Tang, Deep learning face attributes in the wild, in: Proceedings of International Conference on Computer Vision (ICCV), 2015.
[47] G. Huang, M. Mattar, L. Honglak, E.G. Learned-miller, Learning to align from scratch, in: Proceedings of the Neural Information Processing Systems (NIPS), 2012.
[48] B. Amos, B. Ludwiczuk, M. Satyanarayanan, OpenFace: A general-purpose face recognition library with mobile applications, Technical Report, CMU-CS-16-118, CMU School of Computer Science, 2016.

Jia Qin received the B.S. degree from Shanxi University, China, in 2015. Now, she is pursuing her Ph.D. degree in Institute of Information Science, Beijing Jiaotong University, Beijing, China. Her research interests are image inpainting and image/video compression, such as HEVC and rate control algorithm.

Huihui Bai received her B.S. degree and her PhD degree from Beijing Jiaotong University (BJTU), China respectively in 2001 and in 2008. She is currently a professor in Beijing Jiaotong University. Her research interests include video coding technologies and standards, such as HEVC, 3D video compression, multiple description video coding (MVC) and distributed video coding (DVC).

Yao Zhao received the B.S. degree from Fuzhou University, China, in 1989, and the M.S. degree from Southeast University, Nanjing, China, in 1992, both from the Radio Engineering Department, and the PhD degree from the Institute of Information Science, Beijing Jiaotong University (BJTU), China, in 1996. He is currently the director of the Institute of Information Science, BJTU. His current research interests include image/video coding, digital watermarking and forensics, and video analysis and understanding. He serves on the editorial boards of several international journals, including as associate editors of IEEE Transactions on Cybernetics, IEEE Signal Processing Letters, and an area editor of Signal Processing: Image Communication, etc.