International Journal of Machine Learning and Cybernetics (2023) 14:1725–1738
https://doi.org/10.1007/s13042-022-01723-3

ORIGINAL ARTICLE

Generative face inpainting hashing for occluded face retrieval

Yuxiang Yang[1] · Xing Tian[1] · Wing W. Y. Ng[1] · Ran Wang[2] · Ying Gao[1] · Sam Kwong[3]

Received: 3 May 2022 / Accepted: 9 November 2022 / Published online: 2 December 2022

Abstract
COVID-19 has resulted in a significant impact on individual lives, bringing a unique challenge for face retrieval under occlusion. In this paper, an occluded face retrieval method which consists of generator, discriminator, and deep hashing retrieval network is proposed for face retrieval in a large-scale face image dataset under variety of occlusion situations. In the proposed method, occluded face images are firstly reconstructed using a face inpainting model, in which the adversarial loss, reconstruction loss and hash bits loss are combined for training. With the trained model, hash codes of real face images and corresponding reconstructed face images are aimed to be as similar as possible. Then, a deep hashing retrieval network is used to generate compact similarity-preserving hashing codes using reconstructed face images for a better retrieval performance. Experimental results show that the proposed method can successfully generate the reconstructed face images under occlusion. Meanwhile, the proposed deep hashing retrieval network achieves better retrieval performance for occluded face retrieval than existing state-of-the-art deep hashing retrieval methods.

Keywords Occlusion · Face retrieval · Inpainting · Generative adversarial

1 Introduction

As of this article date, because of the communicable disease (such as COVID-19) or other reasons, people always wear a mask, hat, or glasses outside. These items block most of the face information including eyes, nose, and mouth. Existing face retrieval and face recognition systems cannot perform well when encountering challenges such as large-pose variation, varying illumination, low resolution, different facial expressions, and occlusion [1]. Therefore, how to retrieve large-scale face images efficiently and accurately under occlusion has become a key problem in current human life and scientific research.

The previous work to improve the performance of face recognition under occlusion can be generally partitioned into three categories, i.e., occlusion robust feature extraction, occlusion aware face recognition, and occlusion reconstruct based face recognition. The occlusion robust feature extraction methods [2, 3] adopt the data augmentation method to expand the datasets, which alleviate the effect of face occlusion. However, these methods are limited to some special occlusion situations in recognition, which means that they may not perform well under different occlusions. Another possible approach is to add a MaskNet [4] branch in the deep networks to better learn the facial feature representation of the unoccluded region. The MaskNet is used to assign higher weights to hidden units activated by the unoccluded regions. However, there is not enough supervision information to train the MaskNet and the outputs discriminability of middle convolutional layer is not enough. Similarly, a mask learning strategy [5] is proposed to build a mask dictionary that

✉ Xing Tian
shawntian123@gmail.com

✉ Wing W. Y. Ng
wingng@ieee.org

Yuxiang Yang
fotonyoung@gmail.com

Ran Wang
wangran@szu.edu.cn

Ying Gao
gaoying@scut.edu.cn

Sam Kwong
cssamk@cityu.edu.hk

[1] School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China

[2] College of Mathematics and Statistics, Shenzhen University, Shenzhen 518060, China

[3] Department of Computer Science, City University of Hong Kong, Hongkong 999077, China

corresponds between occluded regions and missing feature representation. The occlusion reconstruct based face recognition method intends to recover a face without occlusion [6, 7] to improve the performance of face recognition systems.

Face retrieval has been widely used in many application areas, such as surveillance, forensics and security. Given a query face image, the target of face retrieval task is to retrieve face images that are similar to it from a large-scale face image dataset. Hashing technique, as an advanced indexing technique, has been widely researched to handle this task due to its high retrieval efficiency and low space cost [8–13]. With compact hash codes generated for face images, similarities between face images can be evaluated efficiently based on the Hamming distances, which can be computed quickly by computers. Generally, feature extraction plays an importance role in the performance of most existing hashing methods. Traditionally, hand-craft visual features are employed for face images, such as HOG [14], LBP [15], GIST [16] and SIFT [17]. With the development of deep learning techniques for feature learning [18–20], some deep hashing methods are proposed to improve the efficiency and retrieval accuracy of hash learning, such as CNNH [21], DH [22], DSH [23], DSHSD [24], DPSH [25], Hashnet [26], and CSQ [27]. Because the retrieval performance could be greatly improved by facial feature learning with deep neural network, many deep hashing networks have been proposed for face retrieval such as DDQH [28], DHCQ [29], DCBH [30], and DFH-GAN [31]. These methods achieve encouraging performance for face image retrieval. However, the retrieval performance has been greatly reduced because some of the face components are blocked under occlusions, which makes them fail to adapt to face retrieval problems in occlusion environments.

In this paper, we propose an effective occluded face retrieval method based on a deep generative model and hashing retrieval network, which decomposes the problem of face retrieval under partial occlusion into two stages: face inpainting and generated face retrieval. The occluded face images are reconstructed using a face inpainting model trained with a combination of adversarial loss, reconstruction loss and hash bits loss, which encourages the hash codes of the real face image and the reconstructed face image to be as close as possible. Major contributions of this work are summarized as follows:

1. An occluded face retrieval framework is proposed for face retrieval under several occlusion situations, named Generative Face Inpainting Hashing (GFIH). To the best of our knowledge, GFIH is the first approach combining generative adversarial network and deep hashing network to learn the hash codes for occluded face retrieval.
2. A joint loss function consisting of adversarial loss, reconstruction loss, and hash bits loss is proposed to encourage the generative model to reconstruct a similarity-preserving face image without occlusion. This facilitates the hashing retrieval network to generate compact similarity-preserving hashing codes.
3. Six face occlusion image datasets are created to simulate six different face occlusion situations with different occlusion regions for face retrieval performance evaluation. Quantitative experimental results show that GFIH obtains outstanding occluded face retrieval performance than other comparative methods.

The rest of this paper is organized as follows: Sect. 2 introduces related works on existing face inpainting models and hashing-based face retrieval models. The proposed GFIH is introduced in Sect. 3. Experimental results are discussed in Sect. 4. Section 5 concludes our work in this paper.

2 Related work

In this section, two most related works to the proposed method are described briefly. Section 2.1 introduces the existing face inpainting models. The hashing-based retrieval models are introduced in Sect. 2.2.

2.1 Existing face inpainting models

Previous face inpainting models can generally be divided into two categories: Non-learning inpainting methods and Learning inpainting methods. The non-learning inpainting methods [32, 33] is traditional diffusion-based or patch-based models with low-level features. The learning inpainting methods [34–36] reconstruct a face without occlusion by using deep learning and generative adversarial networks. Some previous face inpainting models use an autoencoder architecture to generate the occluded face region [6, 7]. Context Encoders [6] firstly propose a deep learning method for image inpainting tasks, which employs a generative adversarial network. The input occluded images are created by adding some masked region on the original normal images. This method can learn the feature representation of the occluded image and generate the coherent contents by optimizing the adversarial loss. However, this method focuses more on unsupervised feature learning rather than image inpainting. It is no clear if the generated content can help improve the image retrieval network to learn the compact similarity-preserving hashing codes sufficiently.

An effective object completion algorithm is proposed in [7] using a deep generative model and a face parsing network. Two adversarial loss functions are used to jointly train the autoencoder and discriminator. The first adversarial loss tried to help improve the generated content of occluded region more realistic. The second adversarial loss tries to

help improve the entire reconstructed image which consist of the generated content and unoccluded region of the original image more realistic. A face parsing network is proposed as an additional loss to regularize the generation procedure, which facilitates the generator to generate more reasonable and consistent face inpainting images. However, the performance to generate a fine-detailed content is not well enough.

A high-resolution image inpainting method is proposed in [34] using a multi-scale neural patch synthesis approach, which jointly optimize the image contents and texture constraints. The output of context encoder is employed to generate a high-resolution image by gradually increasing texture details. However, the optimization in this method significantly increases computational costs. Partial convolutions are used in [35] to help the convolution filters focus on the unoccluded regions. This approach renormalizes the convolution filter to be conditioned on only valid pixels by assigning the convolution weights with mask value. EdgeConnect [36] decomposes the image inpainting problem into two stages: structure prediction and image completion. The image structure of the occluded regions is predicted to guide the image inpainting process. However, the EdgeConnect method tend to perform unsatisfactorily in generating contents from highly textured areas and large occluded images. MAT [40] proposed a novel transformer-based model for large hole inpainting to efficiently process high-resolution images.

With the goal of achieving higher retrieval performance for occluded face retrieval by employing the reconstructed face images, an additional hash bits loss is proposed to encourage the hashing retrieval network to generate compact similarity-preserving hashing codes.

2.2 Hashing-based retrieval models

Generally, deep hashing retrieval methods construct a hash function by incorporating a convolution neural network (CNN) model to learn the similarity-preserving feature representation. Deep Supervised Hashing (DSH) [23] learns compact similarity-preserving hashing codes by using the deep feature representation extracted by convolution network of image pairs (similar/dissimilar) and the pairwise similarity. Deep supervised hashing based on stable distribution (DSHSD) [24] is proposed to solve the problem of feature distribution changes caused by the quantization regularizer. A smooth projection is used to help improve the efficiency of the training convergence and make the output binary code preserve more similarity. Deep pairwise-supervised hashing (DPSH) [25] proposes an end-to-end architecture which performs feature learning and hash-code learning simultaneously based on pairwise labels. Hashnet [26] proposes a novel deep architecture for hash code learning by continuation method with convergence guaranteed. It can learn exactly binary hash codes from imbalanced similarity data. The ill-posed gradient problem is solved by optimizing deep networks with non-smooth binary activations. In addition, a new global similarity metric, named as central similarity, is proposed in Central Similarity Quantization (CSQ) [27]. This metric is used to encourage hash codes of similar image pairs to approach a common center and encourage the dissimilar image pairs to converge to different centers.

Many deep hashing networks have been proposed for face retrieval. DDQH [28] is proposed to capture the multiscale feature of face images for hashing codes learning. The feature representation is learned by fusing the output of the last convolutional layer and the last pooling layer. Another deep hashing face retrieval method, DHCQ [29] is proposed to retrieve scalable face images. A loss function consists of quantization error and prediction error is used to optimize the by capturing discriminative facial representations retrieve the discriminative facial feature learning. To solve the problems of inter-class similarities and intra-class variations, DCBH [30] is proposed to learn the robust and multi-scale feature representations. The center-clustering loss is used to encourage the face images of intra-class to approach a common center. Besides, a block hashing layer is used to reduce the number of parameters but also can generate the compact similarity-preserving hashing codes. DFH-GAN [31] proposes a deep face hashing retrieval method combined with generative adversarial network. GAN is employed to generate fake images to augment the training dataset, so the hashing network can be trained from both real images and diverse synthesized images to learns compact binary hash codes. However, these hashing methods focus on normal images, which are not effective to handle the occluded face retrieval problem. Hence, we are motivated to propose the GFIH to combining generative adversarial network and deep hashing network to learn the hash codes for occluded face retrieval.

3 Generative face inpainting hashing

The proposed Generative Face Inpainting Hashing (GFIH) decomposes the problem of face retrieval under partial occlusion into two stages: face inpainting and hashing retrieval stages. The occluded face images are reconstructed using a face inpainting model firstly which consists of a loss function and two inpainting networks: generator and discriminator. Then, a deep hashing retrieval network is used to perform the face retrieval using reconstructed face images from the previous stage for a better retrieval performance. Figure 1 shows an overview of the GFIH. The face inpainting network, the loss function, and the hashing retrieval network of the GFIH will be described in Sects. 3.1, 3.2, and 3.3 respectively.

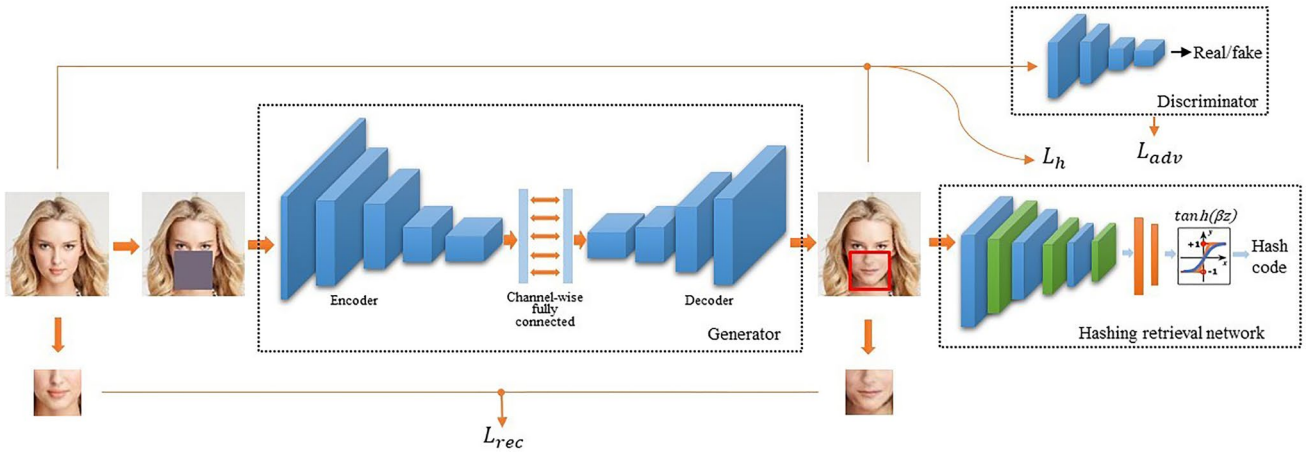

Fig. 1 Overview of the GFIH

3.1 Face inpainting networks of GFIH

The generator in the face inpainting model is designed to inpaint the masked region in the occluded face image. The overall structure of the generator is an encoder-decoder pipeline, as shown in Fig. 1. Different from the original GAN model [37], the latent feature representation is used to generate new content instead of a random noise vector. The latent feature representation is extracted by the encoder with the input occluded face images. Then, the decoder reconstructs the masked region using aforementioned feature representation.

Architectures of the discriminator and the encoder in the generator are similar to the architecture of discriminator in [6, 38], which is a series of four fractionally-stride convolutional layers. Stride convolutional layers allow the network to learn its spatial upsampling by replacing the deterministic spatial pooling functions (such as maxpooling). The encoder features are projected to a small spatial extent convolution representation with many feature maps. Then, five upconvolution layers are employed to reconstruct the occluded region of face image from aforementioned high-level feature representation. The upconvolution layers is a series of transposed convolution which can be consider as upsampling followed by fractionally strided convolutions to reconstruct a higher resolution image. The rectified linear unit (ReLU) activation function is employed in the decoder, while leaky ReLU is employed in the encoder and discriminator. Batch normalization is employed to normalize the input of each unit to zero mean and variance.

Generally, there is an explosion problem of the number of network parameters when using the fully connected layers to connect the high-level feature representation and decoder. To solve this problem, channel-wise fully connected layers [6] are employed to propagate the information across feature maps by replacing the fully connected layers. Unlike fully connected layers, there are no parameters to connecting each feature map in the channel-wise fully connected layers, which is followed by a stride 1 convolution. Therefore, the number of parameters in a channel-wise fully connected layer is mn^4 form feature maps of size $n \times n$. Because of all the activations are directly connected to each other in the fully connected layers, the number of parameters in a fully connected layer is m^2n^4 form feature maps of size $n \times n$. The number of parameters is significantly reduced, which help improve the efficiency of training model.

The occluded region of face image can be filled using the generator by minimizing the reconstruction errors, but the generator may only learn the rough shape of the unoccluded region of face image, which will result in a fuzzy and rough generated content. To encourage the reconstructed face images to look realistic and coherent, a discriminator is employed to help improve the quality of generated details. The discriminator can be trained to distinguish the real face images and inpainting face images, while help improve the ability of generator to generate a face images that can fool the discriminator. This facilitates the face inpainting model to generate a more realistic face image without occlusion.

3.2 Loss function of inpainting networks

A joint loss consisting of adversarial loss, reconstruction loss and hash bits loss is proposed to learn parameters of both the generator and the discriminator in the inpainting model. The adversarial loss tries to make the reconstructed image more realistic and has the effect of matching the distribution of the reconstructed image with the distribution of the original face image. The reconstruction loss tries to help the generator learn the knowledge of the overall structure of the unoccluded region and keep the generated

content consistent. The hash bits loss is a reflection of the Hamming distance between the hash codes of the reconstructed face image and real face image. By optimizing the hash bits loss, it facilitates the network to generate a reconstructed face image whose hash code is similar to the real face image.

By using the discriminator, the adversarial loss is employed to measure the ability of the generator to fool the discriminator, and the ability of the discriminator to distinguish the real and fake face images. The adversarial loss of the proposed model is based on generative adversarial networks. A generator G and a discriminator D are jointly trained in the GAN model. The discriminator D can provides loss gradients to generator G. The training process is a two-player game. The discriminator can be trained to distinguish the ground truth samples and the generated samples of generator G, while help improve the ability of generator G to generate the data pixels that can fool the discriminator D. The objective function for discriminator is logistic likelihood, which indicates whether the input face image is real face or generated one:

$$\min_G \max_D \mathrm{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathrm{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))] \tag{1}$$

where $p_{\text{data}}(x)$ and $p_z(z)$ denote distributions of real image and noise, respectively.

By employing this method, this face inpainting model is adopted for face inpainting by modeling generator. Let M be a binary mask corresponding to the occluded region of face image with a value of 1 for the occluded region and 0 for unoccluded region. For each face image, the occluded region $M \odot x$ is automatically generated for each face image to simulate face occlusion situations. The adversarial loss L_{adv} is defined as:

$$L_{adv} = \min_G \max_D \mathrm{E}_{x\sim p_{\text{data}}(x)}[\log D(x) + \log(1 - D(G((1 - M) \odot x)))] \tag{2}$$

where $\odot$ and $(1 - M) \odot x$ denote the element-wise product operation and the distribution of face image with masked regions, respectively. During training, generator G and discriminator D are optimized jointly using alternating SGD. This objective encourages the reconstructed face images to look realistic and coherent.

A reconstruction loss L_{rec} is another component of the joint loss function for generator, which is the L_2 distance between the reconstructed face images and real face images:

$$L_{rec} = \|M \odot (x - G((1 - M) \odot x))\|_2^2 \tag{3}$$

Hash bits loss is employed to encourage the generator to generate a reconstructed face image $x_g = G((1 - M) \odot x)$ whose hash code is close to the real face image. It is defined as:

$$L_h = \|h(G((1 - M) \odot x)) - h(x)\|_2^2 + \alpha\left(\|h(G((1 - M) \odot x)) - 1\|_1\right) \tag{4}$$

where $h(\cdot)$ and α denote the hash binary code of the images, and a weighting parameter that controls the strength of the regularizer, respectively. $h(\cdot)$ is generated by using a scaled tanh function $\tanh \beta z$ to binarized the feature representation into a K-bit binary hash code, which will be described in detail in Sect. 3.3. The first term encourages the hash codes of the reconstructed face image and corresponding original face image to be as close as possible. The second term is an additional regularizer to replace the binary constraints. Generally, a sigmoid or tanh function is used as a relaxation method to approximate the thresholding procedure. However, optimizing the generative network with these non-linear functions would cause the convergence of the network become difficult and slow [23]. To alleviate this problem, an additional regularizer is adopted to encourage the output values to approach the binary code in the hash bit loss.

The overall loss function, a combination of adversarial loss, reconstruction loss and hash bits loss, is defined as:

$$L = \lambda_1 L_{adv} + \lambda_2 L_{rec} + \lambda_3 L_h \tag{5}$$

where λ_1, λ_2 and λ_3 are weights to balance the effects of different loss.

3.3 Hashing retrieval network

The proposed occluded face retrieval method focuses on the occluded face inpainting learning. By employing the reconstructed face images, the deep hashing retrieval network is expected to achieve higher precision for face retrieval under occlusion. So, the original hash codes of the real face images directly influence the performance of the proposed occluded face retrieval method. Note that there are no limits on the method for learning the original hash codes of the real face images, which means that all existing hashing methods can be used. In this paper, Hashnet [26] is selected for binary hash codes learning in GFIH. Our proposed method first trains the Hashnet using real face image pairs and pairwise similarity to learn the similarity-preserving hash codes of real face images. Then, the trained Hashnet generates binary hash codes for newly coming inpainting face query images. By employing the computation of Hamming distances among hash codes of query images and real face images, GFIH return top k face images yielding the smallest Hamming distance.

The architecture of Hashnet consists of a convolutional neural network and a fully connected hash layer, which accepts the pairwise input faces $\left\{\left(x_i, x_j, s_{ij}\right)\right\}$. The convolutional neural network is used to learn discriminative feature representations of each face image x_i, and the feature representations are transformed into K dimensional

representation $z_i \in \mathbb{R}^k$ by the fully connected hash layer. Then the K dimensional representation z_i is binarized into a K-bits binary hash code $h_i \in \{-1, 1\}^k$ by an activation function $h_i = \text{sign}(z_i)$. For a pair of face images x_i and x_j, let dist $_H(h_i, h_j)$ be the Hamming distance between a pair of binary hash codes h_i and h_j, which is expected to preserve the similarity among image pairs. Note that, there exists a relationship between $\text{dist}_H(h_i, h_j)$ and inner product, $\text{dist}_H(h_i, h_j) = \frac{1}{2}(K - \langle h_i, h_j \rangle)$. Therefore, inner product can be adopted to represent the similarity. Given the set of pairwise similarity $S = \{s_{ij}\}$, the Weighted Maximum Likelihood (WML) estimation of the hash codes $H = [h_1, \ldots, h_N]$ for all training points is defined as:

$$\log P(S \mid H) = \sum_{s_{ij} \in S} w_{ij} \log P(s_{ij} \mid h_i, h_j) \tag{6}$$

where $P(S \mid H)$ and w_{ij} denote the weighted likelihood function and the weight for each training pair (x_i, x_j, s_{ij}), respectively. To solve the data imbalance problem caused by the difference in the number of similar image pairs and dissimilar image pairs, the w_{ij} is employed to assign the weights of the image pairs according to the importance of misclassifying that pair. For a pair of hash codes h_i and h_j, $P(s_{ij} \mid h_i, h_j)$ is the conditional probability of pairwise similarity s_{ij}. The optimization problem of Hashnet is derived as:

$$\min_{\theta} \sum_{s_{ij} \in S} w_{ij} \left(\log \left(1 + \exp \left(\gamma \langle h_i, h_j \rangle \right) \right) - \gamma s_{ij} \langle h_i, h_j \rangle \right) \tag{7}$$

where θ and γ denote the parameter of the feature learning model and the hyper-parameter of adaptive sigmoid function to control its bandwidth, respectively. By optimizing the WML estimation, the hashing network can learn exactly binary hash codes from imbalanced image pairs.

However, optimizing deep networks with sign activation may cause the gradient vanishing problem, which makes it difficult to optimize the network using the standard back-propagation. There exists a relationship between the sign function and the scaled tanh function in the concept of limit in mathematics: $\lim_{\beta \to \infty} \tanh \beta z = \text{sign}(z)$. According to this relationship, the scaled tanh function can be used to replace sign function to optimize the Hashnet. The learning procedure starts with a smoothed activation function $y = \tanh \beta z$. Then, increase the value of β to make the scaled tanh function approach to the original sign function. For face retrieval, there is a big difference in the number of intra-class face pairs and inter-class face pairs, so Hashnet is adopted to generate the compact similarity-preserving hashing codes using reconstructed face images for a better retrieval performance in the proposed method.

4 Experiments

4.1 Dataset and performance metric

Experiments are conducted on two datasets to evaluate the retrieval performance of the proposed method. The input to the face inpainting generator is an image with one or more masked regions. A masked region in the input occluded face image is filled with constant mean value. The masked region could be of any shape. Six different strategies are proposed here to simulate six different face occlusion situations in the actual environment, which consist of people wearing *hat*, *glasses*, *mask*, *hat* + *glasses*, *glasses* + *mask* and *hat* + *mask*. Among these situations, different key components (e.g., eyes, nose, and mouth) that play an important role in retrieval performance are masked. The samples of six different occlusion situations in two datasets are shown in the second row of Fig. 2.

The CelebA [39] dataset is used to generate two datasets CelebA-1H and CelebA-1K. Each face image in CelebA dataset is cropped, roughly aligned by the position of two eyes, and rescaled to $128 \times 128 \times 3$ pixels. The CelebA-1H dataset contains 2752 face images of 97 people, 1541 face images for training, 602 for validation and 609 for testing.

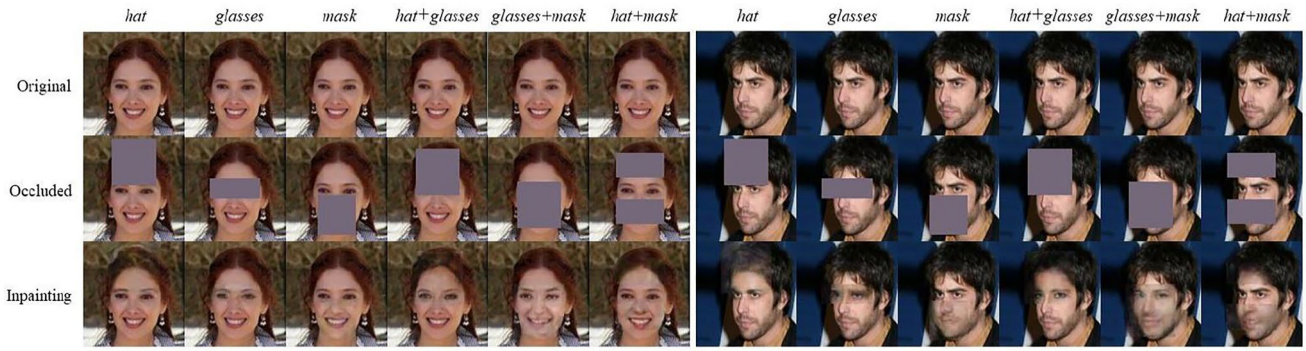

Fig. 2 The face inpainting result of six different face occlusion situations on two datasets

The CelebA-1K dataset contains 26963 face images of 954 people, 15104 face images for training, 5921 for validation and 5938 for testing. The MFRD dataset contains 90,000 face images without masks, 2203 face images with masks of 525 people. Different from CelebA-1H and CelebA-1K datasets, most of occluded face images in MFRD for test are real images in the wild.

The quality of generated face images plays an important role in improving retrieval performance. Peak Signal to Noise Ratio (PSNR), Structural Similarity Index (SSIM) are used in this work to evaluate the quality of generated face images. We can well measure the similarity between the generated face and the original face at different levels using these quality metrics, the PSNR measure the pixel-level difference well, and SSIM measures the structure similarity of generated images to original ones.

The performance metric of face retrieval employed in this work is mean average precision (mAP). mAP is the mean of average precision (AP) of all query data, which amounts to the area under the precision-recall (PR) curve. An image retrieval network with higher mAP corresponds to a larger area under its PR curve, which means that it has a better retrieval performance. For an image dataset, the AP of every images are average to output a value to evaluate the retrieval performance of the entire dataset. Therefore, mAP is used to evaluate occluded face retrieval performance on the CelebA-1h and CelebA-1k datasets.

The AP of n query data is defined as:

$$\mathrm{AP} = \frac{1}{F}\sum_{k=1}^{n}\frac{T_k}{k}\Delta T_k \tag{8}$$

where F, k, T_k, and ΔT_k denote the total number of samples in the database that have the same label with the query, the total number of returned samples, the number of returned samples which have the same label with the query among k returned samples, and the change in recall from item $k-1$ to k, respectively.

4.2 Performance for face inpainting

The generator of the proposed method is trained with the joint loss function defined in Equation 5 for the task of face inpainting under partial occlusion. By employing the reconstructed face images, the deep hashing retrieval network is expected to achieve higher precision for occluded face retrieval. The deep hashing retrieval network is learned by exploiting a deep CNN model to extract appropriate feature representation for the real face images and then generate the hashing codes using the supervised information (similarity and dissimilarity) of image pairs. While the parameters of the hashing retrieval network are already learned and the binary hash codes for the real face images are already given, they remain unchanged during the whole generator learning process. The default solver hyper-parameters λ_1, λ_2 and λ_3 are set to be 0.8325, 0.1665 and 0,001 , respectively. And a higher learning rate is used for generator (10 times) to that of adversarial discriminator.

The face inpainting result of six different face occlusion situations on two datasets are shown in Fig. 2. The first row is the real face image in two datasets. In the second row, six different face occlusion situations are presented, which can simulate most of the face occlusion situations in the actual environment. The second row of each panel shows some results of the proposed method which are visually realistic and pleasing. The third row presents the corresponding reconstructed face images using the proposed generative model. It can be seen that the inpainting face images look realistic and coherent. The occluded region is filled with generated content that fit well within the context. It shows that the inpainting results of the proposed method are encouraging regardless of the mask locations.

The quantitative evaluation results of the generated images are given in Table 1. PNSR and SSIM are calculated to evaluate the quality of generated face images. To evaluate the effectiveness of our proposed inpainting method, the inpainting method Context Encoder (CE) [6] and MAT [40] are employed as a comparison method. As shown in Table 1, the proposed generative model in GFIH achieves a better

Table 1 PSNR and SSIM comparisons of inpaining methods for CelebA-1H and CelebA-1K dataset

	Methods	CELEBA-1H		CELEBA-1K	
		PNSR	SSIM	PNSR	SSIM
Hat	CE [6]	24.056	0.851	24.207	0.858
	MAT [40]	22.785	**0.889**	24.719	**0.905**
	GFIH	**24.285**	0.863	**24.520**	0.866
Glasses	CE [6]	29.581	0.945	29.719	0.949
	MAT [40]	26.045	0.926	27.320	**0.955**
	GFIH	**29.675**	**0.950**	**29.789**	0.950
Mask	CE [6]	27.552	0.916	27.692	0.919
	MAT [40]	25.102	0.905	27.301	**0.933**
	GFIH	**27.700**	**0.919**	**28.419**	0.927
Hat + Glasses	CE [6]	24.059	0.849	23.938	0.847
	MAT [40]	23.287	**0.902**	23.09	**0.901**
	GFIH	**24.369**	0.856	**24.353**	0.858
Glasses + Mask	CE [6]	24.778	0.859	25.186	0.866
	MAT [40]	24.503	**0.903**	24.787	**0.913**
	GFIH	**25.133**	0.862	**25.356**	0.870
Hat + Mask	CE [6]	24.052	0.838	24.535	0.845
	MAT [40]	23.780	**0.863**	23.926	**0.876**
	GFIH	**24.993**	0.849	**25.273**	0.855

The bold font indicate the largest values in the corresponding column

PNSR and SSIM results against CE in most of the compared experiments on both two datasets. However, compared to MAT, the SSIM results of MAT are better than GFIH, but GFIH achieves a better PNSR. In this paper, our goal is to retrieve face images under several occlusion situations, so further retrieval experimental result are given in Sect. 4.4 to compare the performance of these two inpainting methods for occluded face retrieval.

4.3 Comparison with state-of-the-art hashing retrieval methods

To evaluate the retrieval performance of the proposed method, several deep hashing methods are compared including DSH [23], DSHSD [24], DPSH [25], Hashnet [26] and CSQ [27]. The mAP is calculated to evaluate the retrieval accuracy of the proposed method and these several

Table 2 mAP comparisons of deep hashing methods for CelebA-1H and CelebA-1K dataset

	Methods	CELEBA-1H			CELEBA-1K			
		16bits	32bits	64bits	16bits	32bits	64bits	128bits
Hat	DSH [23]	0.138	0.168	0.294	0.011	0.014	0.028	0.033
	DSHSD [24]	0.152	0.288	0.352	0.005	0.013	0.013	–
	DPSH [25]	0.143	0.150	0.172	0.013	0.021	0.040	0.045
	CSQ [27]	0.076	0.163	0.185	–	–	0.026	0.030
	Hashnet [26]	0.087	0.138	0.210	0.023	0.040	0.071	0.090
	GFIH	**0.242**	**0.351**	**0.431**	**0.042**	**0.104**	**0.182**	**0.237**
Glasses	DSH [23]	0.110	0.201	0.294	0.014	0.012	0.020	0.025
	DSHSD [24]	0.173	0.198	0.209	0.004	0.010	0.007	–
	DPSH [25]	0.141	0.290	0.311	0.028	0.045	0.103	0.105
	CSQ [27]	0.119	0.286	0.358	–	–	0.015	0.020
	Hashnet [26]	0.195	0.257	0.350	0.027	0.060	0.106	0.149
	GFIH	**0.309**	**0.475**	**0.530**	**0.046**	**0.136**	**0.197**	**0.270**
Mask	DSH [23]	0.117	0.142	0.409	0.026	0.020	0.029	0.035
	DSHSD [24]	0.254	0.350	0.435	0.011	0.023	0.024	–
	DPSH [25]	0.147	0.280	0.327	0.027	0.040	0.109	0.112
	CSQ [27]	0.122	0.293	0.415	–	–	0.017	0.160
	Hashnet [26]	0.183	0.311	0.435	0.041	0.081	0.141	0.189
	GFIH	**0.276**	**0.489**	**0.579**	**0.057**	**0.166**	**0.285**	**0.385**
Hat + Glasses	DSH [23]	0.0782	0.139	0.157	0.006	0.008	0.011	0.012
	DSHSD [24]	0.086	0.135	0.150	0.003	0.005	0.006	–
	DPSH [25]	0.083	0.141	0.160	0.015	0.021	0.053	0.055
	CSQ [27]	0.067	0.133	0.166	–	–	0.012	0.015
	Hashnet [26]	0.084	0.127	0.162	0.017	0.026	0.046	0.056
	GFIH	**0.201**	**0.288**	**0.345**	**0.035**	**0.075**	**0.117**	**0.147**
Glasses + Mask	DSH [23]	0.099	0.143	0.240	0.006	0.007	0.008	0.007
	DSHSD [24]	0.136	0.169	0.179	0.005	0.009	0.009	–
	DPSH [25]	0.121	0.210	0.254	0.022	0.031	0.079	0.071
	CSQ [27]	0.083	0.252	0.317	–	–	0.017	0.033
	Hashnet [26]	0.171	0.214	0.281	0.025	0.043	0.065	0.079
	GFIH	**0.263**	**0.357**	**0.436**	**0.039**	**0.102**	**0.149**	**0.192**
Hat + Mask	DSH [23]	0.0762	0.153	0.218	0.011	0.007	0.012	0.013
	DSHSD [24]	0.085	0.139	0.150	0.002	0.004	0.006	–
	DPSH [25]	0.065	0.082	0.158	0.011	0.012	0.031	0.036
	CSQ [27]	0.097	0.151	0.163	–	–	0.019	0.022
	Hashnet [26]	0.080	0.128	0.164	0.019	0.020	0.032	0.042
	GFIH	**0.198**	**0.337**	**0.433**	**0.043**	**0.114**	**0.198**	**0.255**

The bold font indicate the largest values in the corresponding column

competitors. A set of experiments is conducted to confirm the effectiveness of the proposed method in six different occlusion situations. The results of the proposed method compared with existing hashing methods on two datasets are given in Table 2.

The third column of Table 2 reports the mAP results of the proposed method (labeled as GFIH) and existing hashing methods using different hash bits length for CelebA-1H dataset, while the fourth column of Table 2 is for CelebaA-1K dataset. It can be seen from Table 2 that the proposed method yields the best results against the state-of-the-art hashing retrieval methods in all occlusion situations, which means that the proposed method can achieve higher precision for occluded face retrieval by employing the reconstructed face images. A combination of inpainting model and the Hashnet retrieval model enables the face retrieval model to perform better under different occlusion situations. Moreover, it can be seen that the mAP of the proposed method remains in a relatively stable and acceptable range under six different occlusion situations. This proves the effectiveness of the proposed method under different occlusion situations, which can simulate most of the face occlusion situations in the actual environment. It also should be noted that the proposed method achieves a best retrieval performance under the occlusion situation that people wearing a mask among these six different occlusion situations. This may imply that the eyes and forehead region play an important role in face inpainting than other regions.

4.4 Comparison with other inpainting methods

For the same occlusion situation and deep hashing retrieval network, the occluded face and face reconstructed by the Context Encoder (CE) [6] and MAT [40] are compared to confirm the effectiveness of the proposed generative model. The proposed generative model architecture is similar to the CE model except the loss functions. Thus, the effectiveness of the joint loss can be evaluated by using the same deep hashing retrieval network.

The results of the proposed method compared with CE and MAT inpainting model on two datasets are given in Table 3. The third column of Table 3 reports the mAP results of combination of Hashnet and non-inpainting model, CE inpainting model, MAT inpainting model or GFIH inpainting model (labeled as Hashnet, CE+Hashnet, MAT+Hashnet and GFIH) using different hash bits length

Table 3 mAP comparisons of inpainting methods for CelebA-1H and CelebA-1K dataset

	Methods	CELEBA-1H			CELEBA-1K			
		16bits	32bits	64bits	16bits	32bits	64bits	128bits
Hat	Hashnet [26]	0.087	0.138	0.210	0.023	0.040	0.071	0.090
	CE+Hashnet	0.219	0.334	0.415	0.036	0.093	0.162	0.226
	MAT+Hashnet	0.222	0.348	0.409	0.033	0.086	0.159	0.225
	GFIH	**0.242**	**0.351**	**0.431**	**0.042**	**0.104**	**0.182**	**0.237**
Glasses	Hashnet [26]	0.195	0.257	0.350	0.027	0.060	0.106	0.149
	CE+Hashnet	0.267	0.414	0.519	0.043	0.115	0.184	0.258
	MAT+Hashnet	0.239	0.395	0.500	0.039	0.108	0.173	0.252
	GFIH	**0.309**	**0.475**	**0.530**	**0.046**	**0.136**	**0.197**	**0.270**
Mask	Hashnet [26]	0.183	0.311	0.435	0.041	0.081	0.141	0.189
	CE+Hashnet	0.251	0.442	0.570	0.054	0.155	0.257	0.368
	MAT+Hashnet	0.256	0.458	0.523	0.040	0.123	0.251	0.356
	GFIH	**0.276**	**0.489**	**0.579**	**0.057**	**0.166**	**0.285**	**0.385**
Hat + Glasses	Hashnet [26]	0.084	0.127	0.162	0.017	0.026	0.046	0.056
	CE+Hashnet	0.196	0.267	0.332	0.033	0.063	0.093	0.132
	MAT+Hashnet	0.191	0.331	0.402	0.028	0.065	0.107	0.135
	GFIH	**0.201**	**0.288**	**0.345**	**0.035**	**0.075**	**0.117**	**0.147**
Glasses + Mask	Hashnet [26]	0.171	0.214	0.281	0.025	0.043	0.065	0.079
	CE+Hashnet	0.212	0.315	0.420	0.037	0.098	0.129	0.185
	MAT+Hashnet	0.228	0.337	0.471	0.032	0.086	0.135	0.181
	GFIH	**0.263**	**0.357**	**0.436**	**0.039**	**0.102**	**0.149**	**0.192**
Hat + Mask	Hashnet [26]	0.080	0.128	0.164	0.019	0.020	0.032	0.042
	CE+Hashnet	0.179	0.283	0.383	0.038	0.097	0.172	0.248
	MAT+Hashnet	0.213	0.351	0.430	0.030	0.083	0.163	0.232
	GFIH	**0.198**	**0.337**	**0.433**	**0.043**	**0.114**	**0.198**	**0.255**

The bold font indicate the largest values in the corresponding column

for Celeba-1h dataset, while the fourth column of Table 3 is for Celeba-1k dataset. It can be seen from Table 3 that the proposed generative model achieves a better mAP in most of the compared experiments on both two datasets, which demonstrate the good performance of the proposed generative model in GFIH. This proves that the combination of adversarial loss, reconstruction loss and hash bits loss enable the GFIH to perform better than other inpainting model. The possible reason for this superior performance is that the proposed generative model help generate the reconstructed face images whose hash code is closer to the real face image against other inpainting methods, so the deep hashing retrieval network can generate the compact similarity-preserving hashing codes and achieve higher mAP by employing these reconstructed face images.

4.5 Comparison with other hashing methods adopted in GFIH

Note that there are no limits on the method for learning the original hash codes of the real face images, which means that existing hashing methods can also be employed. Therefore, several deep hashing methods including DSH, DSHSD, DPSH, CSQ and Hashnet are adopted for illustration. The results of the models combining the proposed inpainting model and several hashing methods (labeled as GFIH-DSH, GFIH-DSHSD, GFIH-DPSH, GFIH-CSQ, and GFIH) on two datasets are given in Table 4. It can be seen from Table 4 that the mAP of all combined models remains in a relatively stable and acceptable range under six different occlusion situations, which proves that the proposed framework can achieve a better retrieval performance against occluded image using different hash retrieval models

Table 4 mAP comparisons of combination of generator and deep hashing methods for CelebA-1H and CelebA-1K dataset

	Methods	CELEBA-1H			CELEBA-1K			
		16bits	32bits	64bits	16bits	32bits	64bits	128bits
Hat	GFIH-DSH	0.194	0.356	**0.443**	0.023	0.031	0.043	0.050
	GFIH-DSHSD	0.262	**0.373**	0.408	0.015	0.025	0.034	–
	GFIH-DPSH	**0.275**	0.280	0.306	0.031	0.06	0.101	0.107
	GFIH-CSQ	0.160	0.286	0.416	–	–	0.041	0.131
	GFIH	0.242	0.351	0.431	**0.042**	**0.104**	**0.182**	**0.237**
Glasses	GFIH-DSH	0.229	0.385	0.443	0.031	0.049	0.032	0.081
	GFIH-DSHSD	0.302	0.385	0.457	0.014	0.028	0.035	–
	GFIH-DPSH	0.211	0.349	0.458	0.039	0.072	0.130	0.142
	GFIH-CSQ	0.232	0.355	0.529	–	–	0.040	0.046
	GFIH	**0.309**	**0.475**	**0.530**	**0.046**	**0.136**	**0.197**	**0.270**
Mask	GFIH-DSH	0.267	0.358	0.455	0.038	0.051	0.093	0.132
	GFIH-DSHSD	**0.371**	**0.501**	**0.585**	0.023	0.052	0.060	–
	GFIH-DPSH	0.213	0.361	0.472	0.037	0.069	0.137	0.151
	GFIH-CSQ	0.228	0.396	0.551	–	–	0.054	0.216
	GFIH	0.276	0.489	0.579	**0.057**	**0.166**	**0.285**	**0.385**
Hat + Glasses	GFIH-DSH	0.177	0.283	0.377	0.017	0.018	0.025	0.030
	GFIH-DSHSD	0.187	0.243	0.270	0.007	0.011	0.017	–
	GFIH-DPSH	0.189	0.260	0.316	0.033	0.057	0.105	0.103
	GFIH-CSQ	0.152	0.263	**0.381**		–	0.031	0.038
	GFIH	**0.201**	**0.288**	0.345	**0.035**	**0.075**	**0.117**	**0.147**
Glasses + Mask	GFIH-DSH	0.210	0.325	0.410	0.019	0.020	0.024	0.029
	GFIH-DSHSD	0.245	0.288	0.331	0.011	0.013	0.017	–
	GFIH-DPSH	0.201	0.315	0.394	0.039	0.070	0.122	0.126
	GFIH-CSQ	0.198	0.323	**0.455**	–	–	0.031	0.039
	GFIH	**0.263**	**0.357**	0.436	**0.039**	**0.102**	**0.149**	**0.192**
Hat + Mask	GFIH-DSH	**0.210**	0.306	0.431	0.024	0.026	0.041	0.045
	GFIH-DSHSD	0.193	0.302	0.321	0.009	0.017	0.022	–
	GFIH-DPSH	0.168	0.268	0.361	0.033	0.058	0.118	0.118
	GFIH-CSQ	0.161	0.303	0.415	–	–	0.033	0.090
	GFIH	0.198	**0.337**	**0.433**	**0.043**	**0.114**	**0.198**	**0.255**

The bold font indicate the largest values in the corresponding column

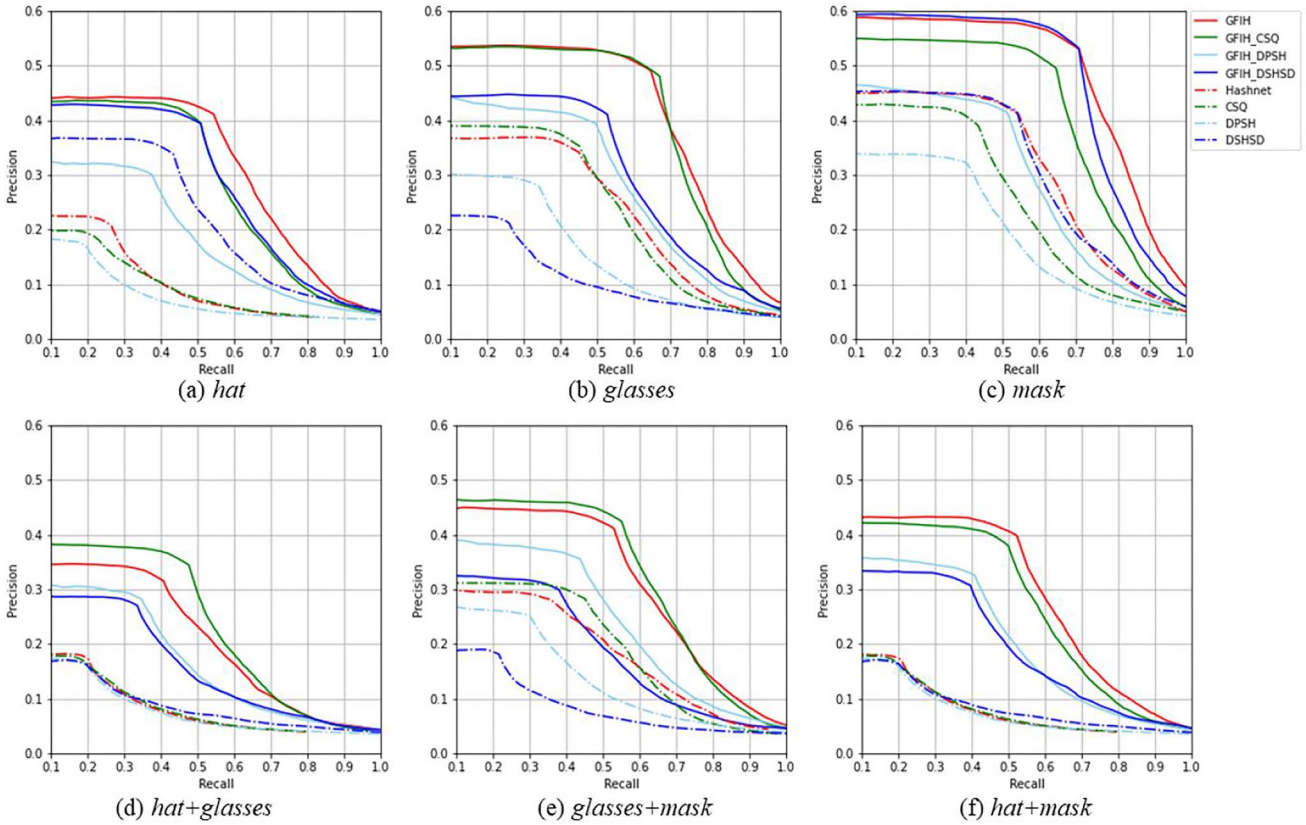

Fig. 3 The Recall-Precision curve under six different face occlusion situations on Celeba-1H dataset

and is general enough to adopt other deep hashing retrieval model to replace the Hashnet for binary hash codes learning. Figure 3 presents the recall-precision curve of the proposed method, existing hashing methods, the combinations of the proposed inpainting model and several hashing methods with 64-bit length hash codes under six different face occlusion situations on CelebA-1H dataset and Fig. 4 presents the recall-precision curve of the above methods with 128-bit length hash codes under six different face occlusion situations on CelebA-1K dataset. Compared to other methods, GFIH is proved to be efficient to balance the recall and precision while achieving higher recall and precision with the same code length under most of the occlusion situations in two datasets.

It can be concluded from Table 4 that the proposed method yields the best results against other combined retrieval methods in most occlusion situations of CelebA-1H dataset and yields the best results against other combined retrieval methods in all occlusion situations of CelebA-1K dataset. This proves the effectiveness of the hashing retrieval methods selected in this paper.

The results in Table 4 also imply that the other combined method such as the combination of GFIH and DSH, DSHSD or CSQ achieves a much lower retrieval performance than the proposed method under most occlusion situations on CelebA-1H dataset. The possible reason for this may be due to the increase of face categories, the amount of similar image pairs is much smaller than the amount of dissimilar image pairs in CelebA-1K dataset, which brings an imbalance problem between similar and dissimilar pairs in the face dataset. The data imbalance problem makes the similarity-preserving learning ineffective in these hashing methods. Because Hashnet is designed to learn similarity from imbalanced similarity relationships with a weighted pairwise cross-entropy loss function, GFIH can generate exactly binary hash codes and yield best retrieval performance on CelebA-1K dataset.

In summary, the proposed occluded face retrieval method achieves a superior performance comparing to other face inpainting models and non-inpainting models for face retrieval under occlusion. The proposed method employs a novel joint loss, which consists of adversarial loss, reconstruction loss and hash bits loss. It encourages the generative model to reconstruct a reconstructed face image, which helps the hashing retrieval network generate the compact similarity-preserving hashing codes.

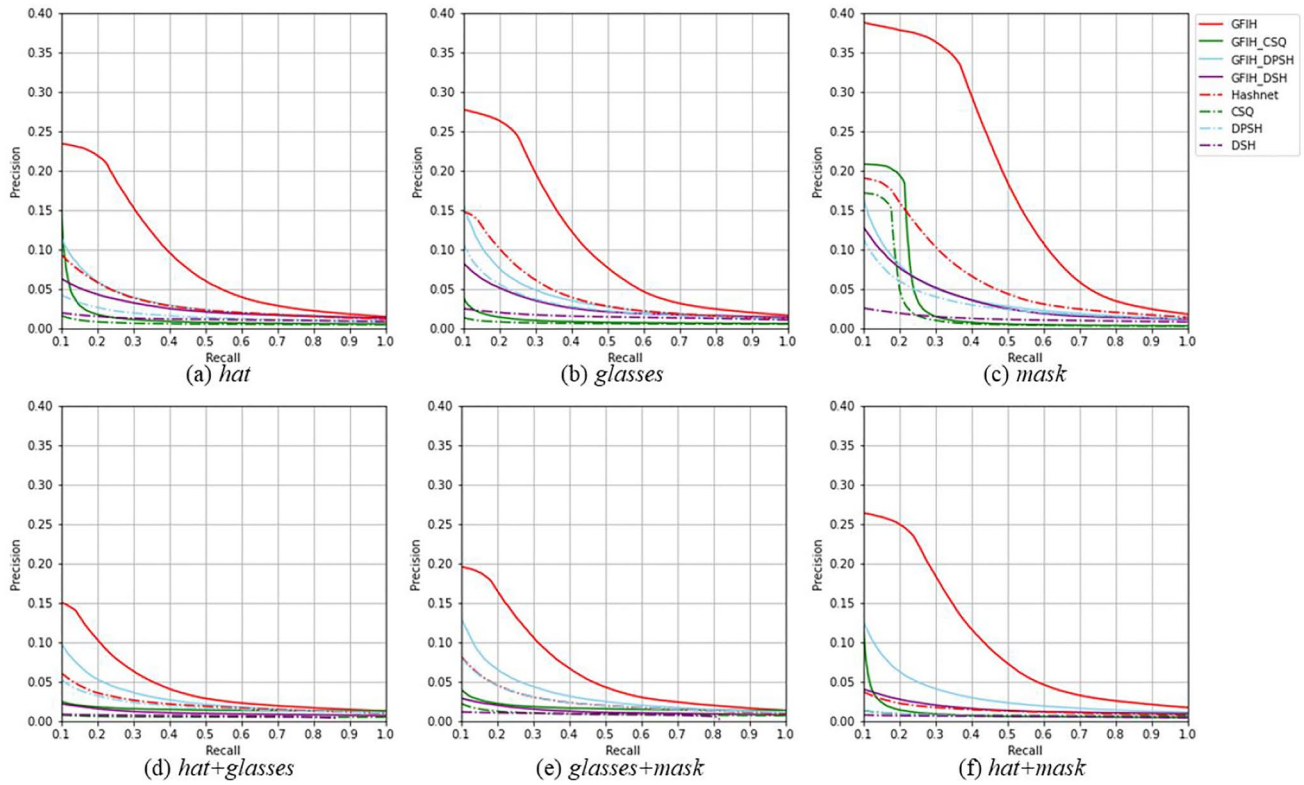

Fig. 4 The Recall-Precision curve under six different face occlusion situations on Celeba-1K dataset

Table 5 mAP comparison on MFRD dataset

Methods	MFRD		
	32bits	64bits	128bits
DSH [23]	0.035	0.052	0.057
DSHSD [24]	0.008	0.010	0.010
DPSH [25]	0.003	0.003	0.004
CSQ [27]	0.007	0.008	0.015
Hashnet [26]	0.036	0.061	0.083
CE+Hashnet	0.115	0.259	0.351
MAT+Hashnet	0.135	0.226	0.359
GFIH-DSH	0.109	0.228	0.227
GFIH-CSQ	0.023	0.105	0.157
GFIH	**0.123**	**0.276**	**0.363**

The bold font indicate the largest values in the corresponding column

4.6 Comparison with other methods on MFRD dataset

To validate the effectiveness of knowledge distillation under real occluded face situation, another dataset MFRD is conducted to validate the effectiveness of GFIH against real mask face images. Most of occluded face images in MFRD for test are real images in the wild. Experimental results of the proposed GFIH in comparison to other methods on MFRD are given in Table 5. This proves the effectiveness of the proposed KDH for real occluded face retrieval.

The MAP comparison in Table 5 shows that face inpainting methods is effective to improves retrieval performance for real occluded faces retrieval. The occlusion reconstructed based face learning methods GFIH-DSH, GFIH-CSQ and GFIH yields superior results against the original DSH, CSQ, Hashnet models. Moreover, our proposed method outperforms other inpaing methods, which proves the effectiveness of the proposed GFIH for real occluded face retrieval.

5 Conclusion

In this paper, an occluded face hash retrieval method is proposed for face retrieval under several occlusion situations. The proposed model consists of generator, discriminator, and deep hashing retrieval network. By optimizing the objective function defined over adversarial loss, reconstruction loss and hash bits loss, the learned generator model can generate the face images under occlusion and enhance the retrieval performance of occluded face retrieval by employing the reconstructed face images.

Because the dataset of six different face occlusion situations is established by artificially adding masked region, it cannot fully simulate all the possible situations of face occlusion in the natural environment. Moreover, the position of occluded region in the nature is unknown in advance. In future work, the GFIH may be extended to an effective retrieval method for different occluded face situations in the wild, so the GFIH is capable of improving the occluded face retrieval performance in the practical environment.

Acknowledgements This work was supported in part by the National Natural Science Foundation of China under Grants 62202175, 61876066, 62176160, and 61672443, the 67th Chinese Postdoctoral Science Foundation (2020M672631), the Hong Kong RGC General Research Funds under Grant 9042489 (CityU 11206317), Grant 9042816 (CityU 11209819) and Grant 9042322 (CityU 11200116), Natural Science Foundation of Guangdong Province of China (2022A1515010791), Hong Kong Innovation and Technology Commission (InnoHK Project CIMDA), and Natural Science Foundation of Shenzhen (20200804193857002).

References

1. Zeng D, Veldhuis R, Spreeuwers L (2021) A survey of face recognition techniques under occlusion. IET Biometr. 10(6):581–606
2. Lv J-J, Shao X-H, Huang J-S, Zhou X-D, Zhou X (2017) Data augmentation for face recognition. Neurocomputing 230:184–196
3. Trigueros DS, Meng L, Hartnett M (2018) Enhancing convolutional neural networks for face recognition with occlusion maps and batch triplet loss. Image Vis Comput 79:99–108
4. Wan W, Chen J (2017) Occlusion robust face recognition based on mask learning. In: 2017 IEEE international conference on image processing (ICIP). IEEE, pp 3795–3799
5. Song L, Gong D, Li Z, Liu C, Liu W (2019) Occlusion robust face recognition based on mask learning with pairwise differential siamese network. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 773–782
6. Pathak D, Krahenbuhl P, Donahue J, Darrell T, Efros AA (2016) Context encoders: feature learning by inpainting. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2536–2544
7. Li Y, Liu S, Yang J, Yang M-H (2017) Generative face completion. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3911–3919
8. Gong Y, Lazebnik S, Gordo A, Perronnin F (2012) Iterative quantization: a procrustean approach to learning binary codes for large-scale image retrieval. IEEE Trans Pattern Anal Mach Intell 35(12):2916–2929
9. Li J, Ng WW, Tian X, Kwong S, Wang H (2020) Weighted multi-deep ranking supervised hashing for efficient image retrieval. Int J Mach Learn Cybern 11(4):883–897
10. Ng WW, Tian X, Lv Y, Yeung DS, Pedrycz W (2016) Incremental hashing for semantic image retrieval in nonstationary environments. IEEE Trans Cybern 47(11):3814–3826
11. Zhu J, Shu Y, Zhang J, Wang X, Wu S (2022) Triplet-object loss for large scale deep image retrieval. Int J Mach Learn Cybern 13(1):1–9
12. Heo J-P, Lee Y, He J, Chang S-F, Yoon S-E (2015) Spherical hashing: binary code embedding with hyperspheres. IEEE Trans Pattern Anal Mach Intell 37(11):2304–2316
13. Ng WW, Jiang X, Tian X, Pelillo M, Wang H, Kwong S (2020) Incremental hashing with sample selection using dominant sets. Int J Mach Learn Cybern 11(12):2689–2702
14. Déniz O, Bueno G, Salido J, De la Torre F (2011) Face recognition using histograms of oriented gradients. Pattern Recogn Lett 32(12):1598–1603
15. Huang D, Shan C, Ardabilian M, Wang Y, Chen L (2011) Local binary patterns and its application to facial image analysis: a survey. IEEE Trans Syst Man Cybern Part C (Appl Rev) 41(6):765–781
16. Purandare V, Talele K (2014) Efficient heterogeneous face recognition using scale invariant feature transform. In: 2014 International conference on circuits, systems, communication and information technology applications (CSCITA), pp 305–310. IEEE
17. Oliva A, Torralba A (2001) Modeling the shape of the scene: a holistic representation of the spatial envelope. Int J Comput Vision 42(3):145–175
18. Li Z, Liu J, Tang J, Lu H (2015) Robust structured subspace learning for data representation. IEEE Trans Pattern Anal Mach Intell 37(10):2085–2098
19. Sun Y, Chen Y, Wang X, Tang X (2014) Deep learning face representation by joint identification-verification. In: Proceedings of the 27th International Conference on Neural Information Processing Systems vol 2 pp 1988–1996
20. Wang H, Wang Y, Zhou Z, Ji X, Gong D, Zhou J, Li Z, Liu W (2018) Cosface: large margin cosine loss for deep face recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 5265–5274
21. Xia R, Pan Y, Lai H, Liu C, Yan S (2014) Supervised hashing for image retrieval via image representation learning. In: Twenty-eighth AAAI conference on artificial intelligence
22. Erin Liong V, Lu J, Wang G, Moulin P, Zhou J (2015) Deep hashing for compact binary codes learning. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2475–2483
23. Liu H, Wang R, Shan S, Chen X (2016) Deep supervised hashing for fast image retrieval. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2064–2072
24. Wu L, Ling H, Li P, Chen J, Fang Y, Zhou F (2019) Deep supervised hashing based on stable distribution. IEEE Access 7:36489–36499
25. Li W-J, Wang S, Kang W-C (2015) Feature learning based deep supervised hashing with pairwise labels. arXiv preprint arXiv:1511.03855
26. Cao Z, Long M, Wang J, Yu PS (2017) Hashnet: deep learning to hash by continuation. In: Proceedings of the IEEE international conference on computer vision, pp 5608–5617
27. Yuan L, Wang T, Zhang X, Tay FE, Jie Z, Liu W, Feng J (2020) Central similarity quantization for efficient image and video retrieval. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 3083–3092
28. Tang J, Lin J, Li Z, Yang J (2018) Discriminative deep quantization hashing for face image retrieval. IEEE Trans Neural Netw Learn Syst 29(12):6154–6162
29. Tang J, Li Z, Zhu X (2018) Supervised deep hashing for scalable face image retrieval. Pattern Recogn 75:25–32
30. Jang YK, Jeong D-j, Lee SH, Cho NI (2018) Deep clustering and block hashing network for face image retrieval. In: Asian conference on computer vision. Springer, pp 325–339
31. Zhou L, Wang Y, Xiao B, Xu Q (2021) Dfh-gan: a deep face hashing with generative adversarial network. In: 2020 25th

international conference on pattern recognition (ICPR). IEEE, pp 7012–7019
32. Criminisi A, Pérez P, Toyama K (2004) Region filling and object removal by exemplar-based image inpainting. IEEE Trans Image Process 13(9):1200–1212
33. Xu Z, Sun J (2010) Image inpainting by patch propagation using patch sparsity. IEEE Trans Image Process 19(5):1153–1165
34. Yang C, Lu X, Lin Z, Shechtman E, Wang O, Li H (2017) High-resolution image inpainting using multi-scale neural patch synthesis. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 6721–6729
35. Liu G, Reda FA, Shih KJ, Wang T-C, Tao A, Catanzaro B (2018) Image inpainting for irregular holes using partial convolutions. In: Proceedings of the European conference on computer vision (ECCV), pp 85–100
36. Nazeri K, Ng E, Joseph T, Qureshi F, Ebrahimi M (2019) Edgeconnect: structure guided image inpainting using edge prediction. In: Proceedings of the IEEE/CVF international conference on computer vision workshops, pp 0–0
37. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. arxiv:1511.06434
38. Radford A, Metz L, Chintala S (2015) Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434
39. Liu Z, Luo P, Wang X, Tang X (2015) Deep learning face attributes in the wild. In: Proceedings of the IEEE international conference on computer vision, pp 3730–3738
40. Li W, Lin Z, Zhou K, Qi L, Wang Y, Jia J (2022) Mat: mask-aware transformer for large hole image inpainting. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition