

RESEARCH

Open Access



# Gender recognition of Guanyin in China based on VGGNet

Yongwen Huang<sup>1\*</sup>, Dingding Chen<sup>2\*</sup>, Haiyan Wang<sup>3</sup> and Lulu Wang<sup>2</sup>

## Abstract

Gender transformation of Guanyin (Avalokitesvara in India) in China is an intrinsically fascinating research topic. Besides the inner source from the scriptures, literatures and epigraphs, iconological analysis is usually as the external evidence of Guanyin's gender recognition. However, the ambiguous gender of the Guanyin image is often intentional and can be objectively assessed. Can computer vision be applied to the recognition objectively and quantitatively? In this paper, VGGNet (VGGNet is a very deep convolutional network for large-scale image recognition proposed by Visual Geometry Group of Oxford University) is applied to propose an automatic gender recognition system. To validate its efficiency, abundant experiments are implemented on the images of Dazu Rock Carvings, Dunhuang Mogao Caves, and Yungang Grottoes. The following conclusions can be made according to the quantitative results. Firstly, VGG-based method can be effectively applied to the gender recognition on non-Buddhist and Buddhist images. Compared with five classical feature extraction methods, VGG-based method performs not much better on non-Buddhist images, but superior on Buddhist images. Furthermore, the experiments are also carried out on three different training datasets, real-world facial datasets, including CUHK (CUHK is a student face database of Chinese University of Hong Kong), IMFDB (IMFDB is an Indian movie face database.) and CAS-PEAL (CAS-PEAL is a Chinese face database created by Chinese Academy of Sciences (CAS) with varying pose, expression, accessory, and lighting (PEAL)). The unsatisfactory results based on IMFDB indicate that it is not valid to apply Indian facial images as a training set to the gender recognition on Buddhist image in China. With the sinicization of Buddhism, there were more Chinese rather than Indian characteristics on Buddhist images in ancient China. The results based on CAS-PEAL are more robust than those based on CUHK, as the former is mainly composed of mature adult faces, and the latter consists of young student faces. It gives the evidence that Buddha and Bodhisattva (Guanyin included) were as ideally mature men in original Buddhist art. The last but the most meaningful is that besides the time factor, the relationship between image making and the scriptures, or the intentional combination of male and female features, the geographical impact should not be ignored when we talk about the gender transformation of Guanyin in ancient China. The gender of Guanyin frescoes in Dunhuang Mogao Caves painted in the Sui, Tang, Five, Song and Yuan dynasties were always with prominent male characteristics (with tadpole-like moustache), while bodhisattvas in Yungang Grottoes engraved in the Northern Wei Dynasty were feminine even though they were made earlier than those in Dunhuang Mogao Caves. It is quite different from the common idea that the feminization of Guanyin occurred in the early Tang Dynasties and completely feminized in the late Tang Dynasty. Both the quantitative results and image analysis indicate that there might be a common model in a specific region, so the image-making of Guanyin was affected much more by

\*Correspondence: hyw@cqust.edu.cn; dingding@cqu.edu.cn

<sup>1</sup> School of Intelligent Technology and Engineering, Chongqing University of Science and Technology, Chongqing 401331, China

<sup>2</sup> College of Computer Science, Chongqing University, Chongqing 400044, China

Full list of author information is available at the end of the article

geographical rather than temporal factor. In a word, it is quite a complicated issue for the feminization of Guanyin in China.

**Keywords:** Guanyin, Gender recognition, Deep convolutional neural network, Buddhist art

## Introduction and research aim

Although asexual Buddha images and sexually ambiguous bodhisattva images are symbolic representations which reflect concepts of Buddhist doctrine, Buddha usually represents an ideally mature man, and Avalokitesvara (Guanyin in Chinese) is mainly depicted as a handsome prince in India. During the Han Dynasties, Buddhism came eastward from India to China. Guanyin was also perceived as predominantly masculine in the early Chinese Buddhist art. However, with the wide spread of the Guanyin belief from the Wei and Jin Dynasties (220–420), the Sui and Tang Dynasties (581–907), to the Song Dynasty (960–1279), its gender had transformed to be female, although the *Lotus*, *Surangama* and other sutras speak of Avalokitesvara appearing in many forms, including male and female. When, how and why a male bodhisattva became a female deity in ancient China? It is an intrinsically fascinating research topic.

Generally, the academic consensus is that Guanyin has been gradually represented as a “Goddess of Mercy” with the sinicization of Buddhism. However, there are different viewpoints for the exact time of the feminization’s occurrence and completion. Although there were some signs for the transformation when the Guanyin belief began to spread in the Wei and Jin Dynasties, and some Guanyin carvings in the Sui Dynasty were with female dresses, most scholars persist that Guanyin was predominantly masculine in early Chinese Buddhist art until the early Tang Dynasty. According to the facsimiles in the Song Dynasty, Guanyin was with typical female decoration in Wu Daozi<sup>1</sup>’s painting in the early Tang Dynasty. Shuiyue (water and moon) Guanyin, created by Zhou Fang<sup>2</sup> in the middle Tang Dynasty was completely female according to Zhang Yanyuan’s *Li Dai Ming Hua Ji (Notes of Past Famous Paintings)* written in the Tang Dynasty and Shuiyue Guanyin made in the Five Dynasties and Song Dynasty. Wu Zetian<sup>3</sup> even claimed that she was the reincarnation of Maitreya Buddha. Therefore, some scholars agree that the feminine transformation of Guanyin occurred in the early period of the Tang

Dynasty and it had been basically settled by the time of Emperor Gaozong and Wu Zetian. However, it was still not popular for Guanyin as a goddess during that period. The gender of Guanyin in Dunhuang Mogao Grottoes were still ambiguous, although they were with high hair buns or wearing flower crowns as noble women. Many scholars agree that the feminization of Guanyin occurred in the middle Tang Dynasty, and increasingly feminized during the late Tang Dynasty. Jiao [1] found that during the period between the early Tang Dynasty and the period of Emperor Gaozong and Wu Zetian, both male and female Guanyin sculptures could be found. Till the late Tang Dynasty, Guanyin had shown prominent female features. Yu [2] held the point that starting from the late Tang, Guanyin was completely feminine, and tried to explain how and why the transformation happened [3] according to different Buddhist sutras. Quite a number of scholars claim that the sinicization and feminization of Guanyin completed during the Song Dynasty. From the Five Dynasties (around the tenth century) to the Song Dynasty, Guanyin statues were increasingly feminized. The folk stories of Princess Miaoshan, Yulan Guanyin, White-robed Guanyin, Nanhai Guanyin in the Song Dynasty show that Guanyin was well known as a “Goddess of Mercy”, appearing in different female forms. It indicates the completion and further consolidation of Guanyin’s feminization.

From this perspective, gender identification can help the dating study of Guanyin images in China. Sutras, literatures and epigraphs have been seen as the inner source and evidence for the cult of Guanyin developed in China. Moreover, iconological analysis is also applied to the study of the Guanyin belief because the idol worship of Guanyin is the external manifestation of the belief in Guanyin. However, the ambiguous gender of the Guanyin image is often intentional and can be objectively assessed. Can computer vision be applied to gender identification on Guanyin images objectively and efficiently? Based on deep convolutional neural networks and gender identification algorithms, an automatic gender recognition system is proposed to quantitatively study the gender of the Buddhist images. The aim is to provide some objective information for the research on the gender transformation of Guanyin.

Bodhisattva statues in Yungang Grottoes, Guanyin frescoes in Dunhuang Mogao Caves and Guanyin statues in Dazu Rock Carvings are chosen to be discussed in this

<sup>1</sup> Wu Daozi (680–759) was a famous painter known as Painting Saint in Chinese art history. His representative work is *Born of Gautama Buddha*.

<sup>2</sup> Zhou Fang was born in late eighth century, and died in early ninth century. He was famous for figure paintings in the middle Tang Dynasty. His representative work is *Portrait of Beauties Wearing Flowers*.

<sup>3</sup> Wu Zetian (624–705) was the only Queen in Chinese history.

paper, as they are the representatives of Guanyin image making during the time of the Wei and Jin Dynasties, the Sui and Tang Dynasties and the Song Dynasty, respectively, which overlaps the time clue of the spread of the Guanyin belief in ancient China.

The remains of the paper are organized as follows: Sect. 2 is the review of the related work. The gender classification system is proposed and described in Sect. 3. Experiments and analysis are shown in Sect. 4. Finally, the conclusion and the future work are discussed in Sect. 5.

## Related work

### Gender recognition

In the past few decades, much attention has been paid to gender recognition in the field of computer vision. “SexNet” [4] was one of the early methods, in which a fully connected back-propagation network with a hidden layer was trained on a small set of near-frontal facial images. In order to improve the robustness, Baback et al. [5] proposed a support vector machine (SVM) [6] classifier trained on “thumbnail” faces processed from 1755 images in the FERET face database [7] with low resolution. However, since the classifying models of these methods were trained using natural facial images, and the image preprocessing, including face alignment, image cropping and resizing, was the key step in the gender recognition system. Thus, image preprocessing has been an obstacle to gender recognition.

Instead of improving image preprocessing algorithms, many researchers paid more attention to feature-based methods. A trained active appearance models (AAM) [8] was used as a feature extractor to recognize the gender and the expression. As AAM-method was invariable to the shape, illumination and pose, the extracted features contain more related facial information for building better classifying models. Another famous feature extractor, Local Binary Patterns (LBP), encoded the local patterns in a small neighborhood, and represented the facial images with histograms of oriented gradient (HOG) [9]. Ihsan et al. [10] proposed to use Weber Local Descriptor (WLD), an attempt to simulate the well-known Weber’s law [11], as a feature extractor for gender recognition. The popular technique in machine learning, Principal Component Analysis (PCA), was also used to represent the facial image in a low dimensional space with the help of a feature vector [12].

Unfortunately, the above methods almost all focused on facial images acquired under controlled condition, so they often failed to the application in the faces captured in the natural condition. To improve the robustness, Caifeng [13] proposed a method of local binary patterns histograms (LBPH) to select boosted features with LBP

and Adaboost [14]. It was demonstrated a near-perfect performance on the labeled faces in the wild (LFW) [15] when the SVM classifier is applied.

### Deep convolutional neural networks and VGGNet

With the improvement of computational capacity, deep convolutional neural network (CNN) has become a dominant approach in many computer vision tasks, such as face classification [16–18], object detection [19], semantic segmentation [20] and many other recognition tasks [21, 22]. Deep CNN is a type of feed forward artificial neural network, in which the output of the previous layer is the input of the next layer, so as to obtain high-level features by combining the low-level features and learn the feature presentation of the input data.

Since Yann et al. [23] established the first CNN architecture named “LeNet-5” for handwritten digit classification in the 1990s, significant development has been made towards deep learning model optimization. Thanks to the huge image data available online, fast optimization achieved on modern hardware and accelerated programs on GPU,<sup>4</sup> significant breakthroughs in image recognition had been made in 2012. Alex et al. [17] proposed a classical CNN model named “AlexNet”, which was much deeper and wider than “LeNet-5” and the championship on the ILSVRC<sup>5</sup> in 2012. Then, more CNN architectures had been proposed, such as ZFNet [21], VGGNet [24], GoogLeNet [20], ResNet [22] and DensNet [25].

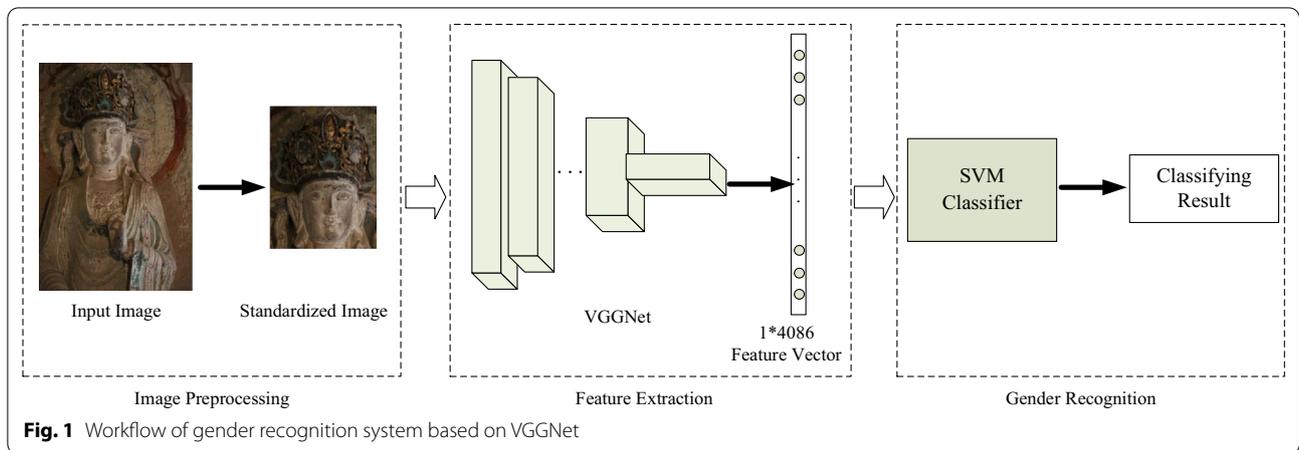
As one of the most famous deep CNN architectures, VGGNet has been successfully applied to image classification because it is easily fine-tuned. There are five different architectures, A, B, C, D and E. In which, D and E architectures can get much better recognition accuracy than others. Although E architecture slightly outperforms D architecture, we choose D architecture, called VGGNet-16 in our work, due to its lower time complexity. Its efficiency has been verified in [26] and our previous work [27].

### Gender recognition system based on VGGNet

As shown in Fig. 1, the proposed gender recognition system based on VGGNet (aliased as ‘VGG-based system/method’ for convenience) includes the following three parts: image preprocessing, feature extraction and gender recognition. The input images are firstly normalized to be standard images in the image preprocessing step. Image

<sup>4</sup> GPU is Short for Graphics Processing Unit.

<sup>5</sup> ImageNet Large Scale Visual Recognition Challenge (ILSVRC) evaluates algorithms for object detection and image classification at large scale. It is one of the most popular and authoritative academic competitions in the field of computer vision in recent years, representing the highest level in the field of graphics.



features are then extracted in VGGNet. Finally, gender recognition is achieved by applying a SVM classifier.

**Face image preprocessing**

The face image preprocessing consists of the following steps:

- Image cropping: Irrelevant information is eliminated from the image. In order to get the facial image, Photoshop is applied to crop the original full-length image.
- Image scaling: The images are resized in a consistent size. Bilinear algorithm [28] is chosen for image scaling, as it can achieve a good balance between the computational complexity and the performance.
- Global contrast normalization: Global contrast normalization (GCN) [29] is used to prevent images from various amounts of contrast by subtracting the mean from each image. The image is rescaled so that the standard deviation across its pixels is equal to some constants.

In the GCN procedure, contrast simply refers to the magnitude of the differences between the bright and the dark pixels in an image. In the context of deep CNN, contrast usually refers to the standard deviation of the pixels in an image. Suppose an image is represented by a tensor  $\in \mathbb{R}^{h \times w \times 3}$ , with  $X_{i,j,1}$  being the red intensity at row  $i$  and column  $j$ ,  $X_{i,j,2}$  giving the green intensity and  $X_{i,j,3}$  giving the blue intensity, the contrast of the entire image is given by

$$\sqrt{\frac{1}{3hw} \sum_{i=1}^h \sum_{j=1}^w \sum_{k=1}^3 (X_{i,j,k} - \bar{X})^2}, \tag{1}$$

where  $h$  and  $w$  are the height and the width of the image, respectively, and  $\bar{X}$  is the mean intensity of the entire image. The formal definition to  $\bar{X}$  is given by

$$\bar{X} = \frac{1}{3hw} \sum_{i=1}^h \sum_{j=1}^w \sum_{k=1}^3 X_{i,j,k}. \tag{2}$$

Given an input image  $X$ , an output image  $X'$  is obtained with GCN, such as in

$$X'_{i,j,k} = s \frac{X_{i,j,k} - \bar{X}}{\max \left\{ \epsilon, \sqrt{\lambda + \frac{1}{3hw} \sum_{i=1}^h \sum_{j=1}^w \sum_{k=1}^3 (X_{i,j,k} - \bar{X})^2} \right\}}, \tag{3}$$

where  $s$  is the scale parameter and set as 1 in our work.  $\epsilon$  is a constant and set as an extremely low value like  $10^{-8}$  to avoid division by 0.  $\lambda$  is the positive regularization parameter to bias the estimate of the standard deviation.

**Facial feature extraction based on VGGNet**

Feature extraction is the core step of the VGG-based system, for it cannot correctly classify the gender with poor features. In this paper, the goal of feature extraction is to obtain the most gender-related information and represent it in a lower dimensional space. Effective characterization can be directly extracted from the original face image with a few preprocessing by VGGNet model. It has been successfully applied to image recognition [27] and virtual inpainting [30, 31] in our previous work. The pseudo code of the procedure of feature extraction is described as Algorithm 1.

At the beginning of the algorithm, the image set, VGG-Net model and net parameter are loaded (line 1). After being preprocessed (line 4), every input image passes through 13 convolutional layers (line 5–6). In each convolutional layer, feature maps are extracted based on the

previous layer, and high-level features therefore can be obtained by combining low-level features. At last, the final feature vectors can be acquired through the two fully-connected layers (line 7–8).

where

$$t_n = \begin{cases} 1 & \text{if class } n \text{ is the ground truth} \\ -1 & \text{otherwise,} \end{cases} \quad (5)$$

---

**Algorithm 1** VGGNet model-based feature extraction

---

**Input:**

Image set:  $I = \{p_1, p_2, \dots, p_n\}$   
 VGGNet model  
 VGGNet net parameter

**Output:**

Feature vector set of  $I$ :  $F = \{f_1, f_2, \dots, f_n\}$

- 1: load  $I = \{p_1, p_2, \dots, p_n\}$ , VGGNet model and VGGNet parameter
  - 2: initialize  $F = \Phi$
  - 3: **for**  $i = 1$ :  $n$  **do**
  - 4:     preprocess  $i$ -th image  $p_i$
  - 5:     load  $p_i$  into the input layer of VGGNet
  - 6:     propagate forward through the network
  - 7:     extract feature vector  $f_i$  of  $p_i$  via VGGNet
  - 8:     pushback  $f_i$  into  $F$ :  $F = F \cup \{f_i\}$
  - 9: **end**
  - 10: output  $F = \{f_1, f_2, \dots, f_n\}$
- 

**Gender classification based on the linear SVM**

The linear SVM [6] is applied to gender classification based on the acquired feature vectors. The main idea of the linear SVM is to find the best hyperplane to separate the different classes (e.g., male and female). The best hyperplane for a linear SVM means the one with the largest margin between the two classes.

The linear SVM classifier with a linear kernel is trained to minimize the squared hinge loss with regularization term by using LIBLINEAR<sup>6</sup> library [32]. For a linear SVM classifier model with the weight of  $w$ , the loss function is defined as

$$\min \frac{1}{2} w^T w + \lambda \sum_{n=1}^N \max(1 - w^T f_n t_n, 0)^2, \quad (4)$$

$N$  is the number of the training images,  $f_n$  is the extracted feature vector of the  $n$ -th image.  $T$  denotes the transpose of a matrix. The loss function also involves a hyper-parameter  $\lambda$  which determines the penalty for data misclassification.

The pseudo code of gender classification applied the linear SVM is shown in Algorithm 2. Firstly, the training dataset and the test dataset are loaded (line 1). Then the training dataset is used to train the classifying model (line 2–5). Specifically, the model weights  $W$  is updated according to optimizing formula (4). For each test image, its confidence is computed by multiplying its feature vector by the model weights  $W$  (line 7). Finally, the predicted label is obtained by comparing the confidence with thresholds (line 8–9).

---

<sup>6</sup> LIBLINEAR is a Library for Large Linear Classification.

**Table 1** Dataset of different subjects

Sites	Image categories	Number of the cave\niche	Amount
Yungang Grottoes	Bodhisattva	Bodhisattva images from different niches in Yungang Grottoes ( <b>YG-B</b> ), A.D. 460–494	348
Dunhuang Mogao Caves	Guanyin	Guanyin images of different caves in Mogao Caves ( <b>MG-G</b> ), A.D. 581–907	106
Dazu Rock Carvings	Male	Shimenshan No. 10 ( <b>SMS-10</b> ), A.D. 1136–1141	336
	Male	Shizhuanshan No. 6 ( <b>SZS-6</b> ), A.D.1082–1096	288
	Female	Baodingshan civilian women images ( <b>BDS-F</b> ), A.D. 1174–1252	308
	Guanyin	Shimenshan No. 6 ( <b>SMS-6</b> ), A.D. 1136–1141	304
	Buddha	Baodingshan No. 18 ( <b>BDS-18</b> ), A.D. 1174–1252	186

The bold values are short for the image data from different caves or niches for the convenience of the description in the following sections.

---

**Algorithm 2** Gender classification based on the linear SVM

---

**Input:**

Training dataset:  $D_{tr} = \{ \langle f_1^{tr}, l_1^{tr} \rangle, \langle f_2^{tr}, l_2^{tr} \rangle, \dots, \langle f_m^{tr}, l_m^{tr} \rangle \}$

Test dataset:  $D_{te} = \{ f_1^{te}, f_2^{te}, \dots, f_n^{te} \}$

**Output:**

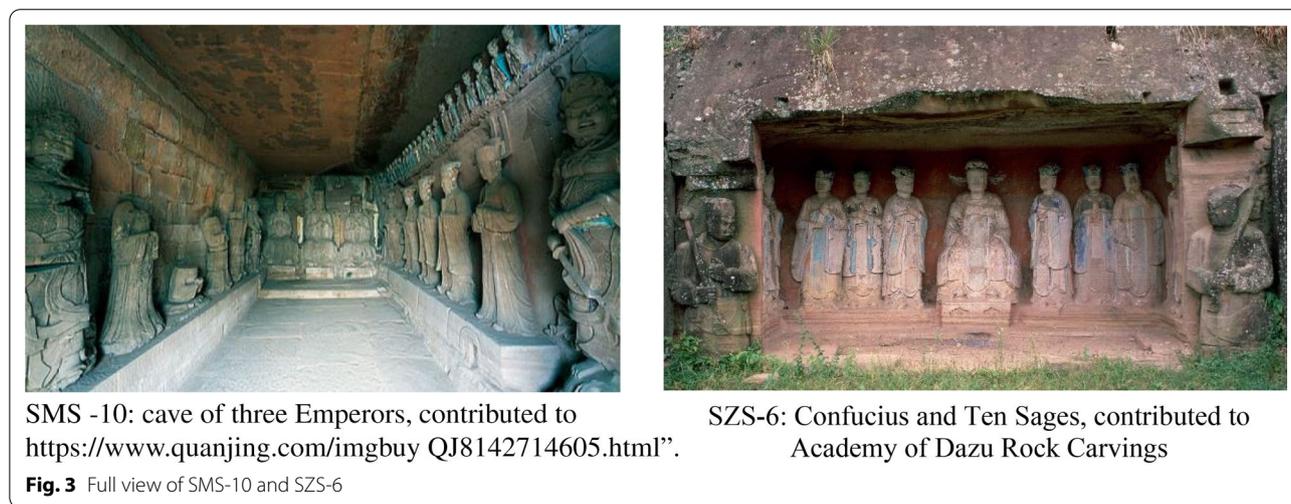
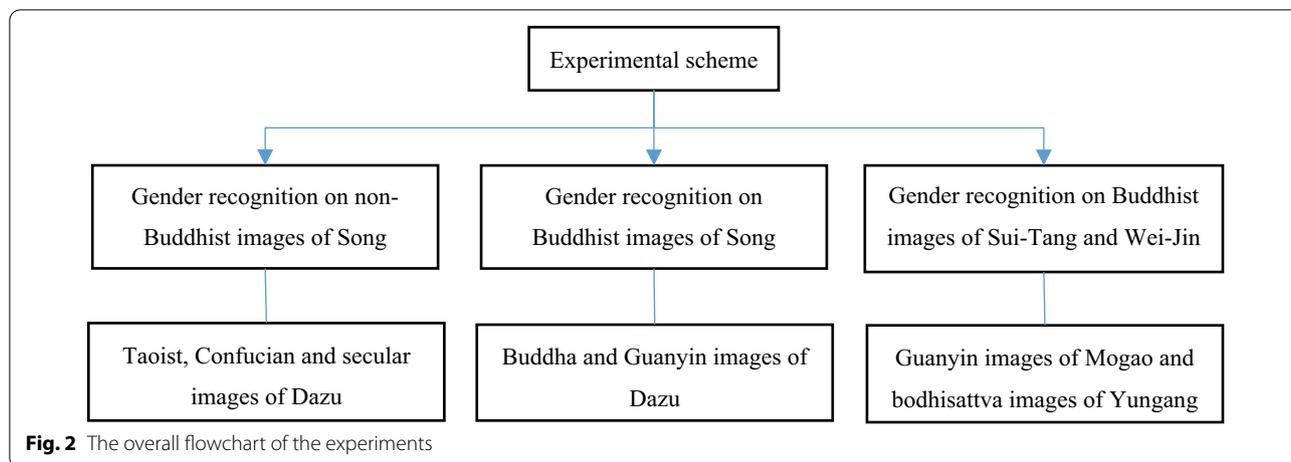
Predicted labels for  $D_{te}$ :  $L_{te} = \{ l_i^{te} | i=1, 2, \dots, n \}$

- 1: load  $D_{tr}$  and  $D_{te}$
  - 2: set weight vector  $\mathbf{W} = 0$
  - 3: for each  $\langle f_i^{tr}, l_i^{tr} \rangle \in D_{tr}$
  - 4:     update  $\mathbf{W}$  with optimization formula (4)
  - 5: end
  - 6: for each  $f_i^{te} \in D_{te}$  do
  - 7:     calculate  $c_i^{te} = \mathbf{W} * f_i^{te}$
  - 8:     if  $c_i^{te} > 0.5$ ,  $f_i^{te} = 1$
  - 9:     then  $f_i^{te} = 0$
  - 10: end
  - 11: output  $L_{te} = \{ l_i^{te} | i=1, 2, \dots, n \}$
- 

**Experiments and analysis**

Yungang Grottoes, Longmen Grottoes, Dunhuang Mogao Caves and Dazu Rock Carvings are recognized as four monuments of Buddhist image making both in China and worldwide. To validate the efficiency of VGG-based method, it is implemented on abundant image data of Yungang Grottoes, Dunhuang Mogao Caves and Dazu Rock Carvings. For they are the representatives of Buddhist images in the Northern Wei Dynasty, the Sui and

Tang Dynasties and the Song Dynasty, respectively, these images can be regarded as the clue of the transformation in ancient China. The dataset on different subjects in our experiments mainly consists of 348 color head images from Yungang Grottoes, 106 color head images from Dunhuang Mogao Caves, and 1422 color head images from three hills (Shimenshan, Shizhuanshan and Baodingshan) in Dazu, shown in Table 1.



The involved images were mainly painted or engraved during three periods: The Northern Wei Dynasty (A.D. 386–534), the Sui and Tang Dynasties (A.D. 581–907) and the Song Dynasty (A.D. 960–1279). They are divided into five categories: bodhisattva, Guanyin, male, female and Buddha images. The male images are from Shimenshan No. 10 (denoted as SMS-10 for convenience), the Cave of Three Emperor, and Shizhuanshan No. 6 (denoted as SZS-6), a shrine for Confucius and Ten Sages. The female images are secular woman images collected from the niches of No. 15, No. 17 and No. 20 at Baodingshan (denoted as BDS-F), where are famous for the secular subjects. The Buddha images are chosen from Baodingshan No. 18 (denoted as BDS-18), a niche of religious subject of Sutra of Amitabha and His Pure Land. Shimenshan No. 6 (denoted as SMS-6), the Cave of Three Sages and Ten Avalokitesvara, is naturally as the image source of Guanyin. As the image making of Guanyin in Dunhuang Mogao Caves (denoted as MG-G) was not so

integral and holistic in one cave as it was in Dazu, the images are collected from different caves. Bodhisattva images of Yungang Grottoes (denoted as YG-B) are chosen from different niches for the experiments, because the belief in Guanyin was not so popular during this period and there were only a few Guanyin images. The images of the Dazu Rock Carvings are partly provided by Academy of Dazu Rock Carvings or scanned from *the complete works of Dazu Rock Carvings* [33], and mainly photographed by the authors. The images of Dunhuang Mogao Caves are main screenshots from Digital Dunhuang (<https://www.e-dunhuang.com/>) and Chinese Treasure Museum (<http://www.lfsc.net/main.html>), and some are scanned from *Dunhuang Mogao Caves* [34]. The images of Yungang Grottoes are mainly re-shot from *Yungang Grottoes Buddha Statues* [35], and some are downloaded from the Internet, others are photographed by the authors.



Fig. 4 Samples of data 1

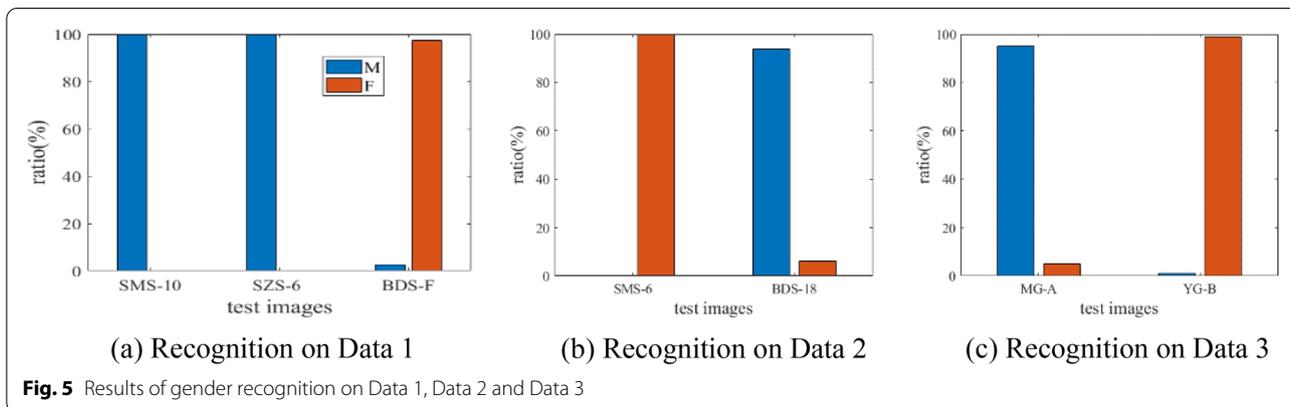


Fig. 5 Results of gender recognition on Data 1, Data 2 and Data 3

The image preprocessing, feature extraction and gender recognition are implemented on Matlab 2021a, and the VGGNet is implemented based on the Deep Learning Toolbox in Matlab.

As the gender of Guanyin is mainly ambiguous in vision before the Song Dynasty, it is difficult to validate the efficiency of the gender recognition algorithm directly on

Guanyin images. However, the gender is evident for the non-Buddhist images and Buddha images, we therefore firstly design gender recognition experiments on the two kinds of images of the Dazu Rock Carvings. Then the results on Guanyin could be more convincing. It is worth emphasizing that the experiments on Guanyin images are designed not in chronological order, but in reverse.

**Table 2** Average recognition ratios and standard deviations by different methods on Data 1

Division ratio Methods	20/80%	40/60%	60/40%	80/20%
HOG	0.9691 ±0.0339	<b>0.9939</b> ±0.0043	<b>0.9945</b> ± <b>0.0050</b>	0.9891 ±0.0163
LGBP-M	0.9455 ±0.0598	0.9915 ± <b>0.0033</b>	0.9891 ±0.0076	0.9745 ±0.0163
LGBP-P	0.9791 ± <b>0.0215</b>	<b>0.9939</b> ±0.0043	0.9891 ± <b>0.0041</b>	0.9891 ±0.0100
WLD	0.9600 ±0.0266	0.9867 ±0.0066	0.9927 ±0.0076	<b>0.9964</b> ± <b>0.0081</b>
LBPH	<b>0.9827</b> ±0.0217	<b>0.9952</b> ± <b>0.0027</b>	<b>0.9945</b> ± <b>0.0050</b>	0.9927 ±0.0100
VGG-based	<b>0.9927</b> ± <b>0.0105</b>	<b>0.9939</b> ±0.0086	<b>0.9945</b> ±0.0122	<b>1.0000</b> ± <b>0.0000</b>

It makes the experiments be from the easier to the more difficult. The overall flowchart of the experiments is shown in Fig. 2.

#### Gender recognition on non-Buddhist images

The proposed system is firstly applied to the gender recognition on non-Buddhist statues of Dazu Rock Carvings, mainly engraved during the Southern Song Dynasty. The statues of SMS-10, SZS-6, and BDS-F are used as Data 1 in gender classification.

#### Gender recognition on Taoist, Confucian and secular images of Dazu

As described by UNESCO, Dazu Rock Carvings are remarkable for their aesthetic quality, their rich diversity of subject matter, both secular and religious, and the light that they shed on everyday life in China during this period. They provide outstanding evidence of the harmonious synthesis of Buddhism, Taoism and Confucianism (<https://whc.unesco.org/en/list/912>). SMS-10, the Cave of Three Emperors, is the subject of Taoism, while SZS-6, Confucius and Ten Sages, is about the subject of Confucianism, shown in Fig. 3. The former is praised as the peak of Taoist statues in the Song Dynasty. The latter is a representative of Confucian image making in the Dazu Rock Carvings. The statues of BDS-F are secular woman images collected from some niches on Baodingshan. Obviously, three image categories are all non-Buddhist images.

As shown in Fig. 4, the statues of SMS-10 and SZS-6 are completely male images, while the statues of BDS-F are female images. These images with significant gender characteristics are taken as Data 1, and the experiments on gender identification are carried out with VGG-based method.

The feature vector of every image in Data 1 is firstly extracted based on VGGNet-16. Then the dataset is divided into two parts: training set (20% of the total) and test set (80% of the total). Finally, the linear SVM is

applied to classifying experiments. The results of the gender recognition on Data 1 can be seen in Fig. 5a.

In Fig. 5, the abscissa is the test images, and the ordinate is the recognition accuracy. The blue bar graph represents male and the orange one stands for female. It is shown in Fig. 5a that the accuracy of male recognition on SMS-10 and SZS-6 are 100%, and 97.3% images in BDS-F are classified to be female. For the gender of SMS-10 and SZS-6 are certainly male, and BDS-F is female, it can easily testify and validate the effectiveness of the VGG-based method in gender recognition on those non-Buddhist images.

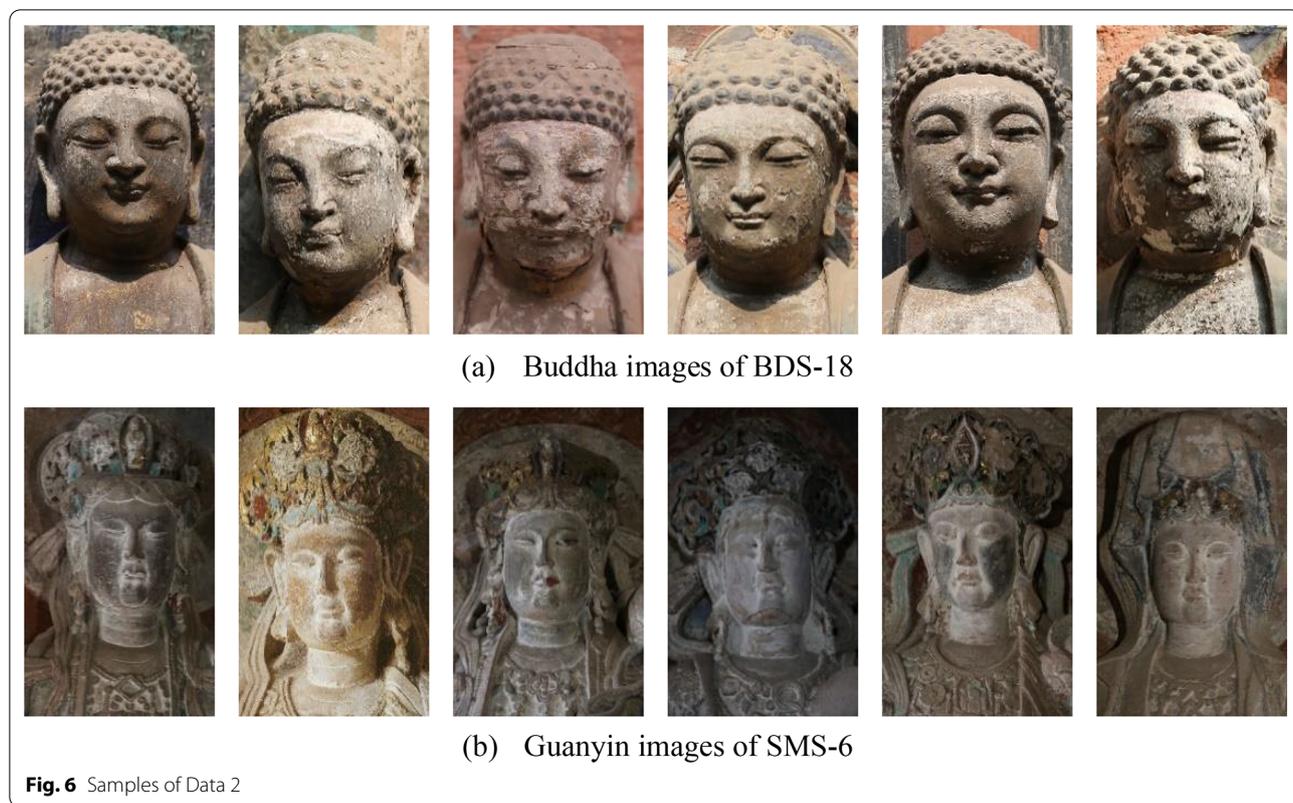
#### Efficiency comparison

In this section, the VGG-based method is compared with 5 traditional hand-crafted feature algorithms, including HOG [9], WLD [10], LBPH [13], phase-based local Gabor binary patterns (LGBP-P) [36], and magnitude-based local Gabor binary patterns (LGBP-M) [37]. The former three famous methods have been introduced in Sect. 2. The latter two are classical face representation approaches based on non-statistics, and they are not only with strong robustness to the changes in imaging conditions, but also with strong discriminative ability.

Various classifiers have been trained on Data 1 with multiple splits of the training data and test data: 20/80%, 40/60%, 60/40% and 80/20%. Every experiment is repeated 10 times. The average recognition ratios and the standard deviations<sup>7</sup> of every method are presented in Table 2.

The recognition ration is denoted as 0–1, and the higher it is, the better the method performs. The Entry in bold and red color is the highest recognition ratio for each experiment, and that in bold and black color

<sup>7</sup> A standard deviation is a statistic measure of the dispersion of a dataset relative to its mean. It is calculated as the square root of variance by determining the deviation of each data point from the mean. A low standard deviation indicates that the values of the data points tend to the mean, while a high standard deviation indicates that the value of the data points are distributed in a larger range.



is the second highest result. The standard deviation is also denoted as 0–1. A large standard deviation indicates a large difference between most values and their mean; a small one means that these values are closer to the mean and therefore the method has better performance. The lowest is in bold and blue color for each experiment, and the second lowest is in bold and black.

From the results in Table 2, it can be seen that VGG-based method outperforms a little better than five classical feature-based methods. The recognition ratios are 99.27%, 99.39%, 99.45% and 100% based on four different classifiers, respectively. In which, it performs the best twice but surpasses slightly than the second, performs the second once, and ties with two others to be the best once. The standard deviations are 0.0105, 0.0086, 0.0122, 0.0000, respectively, in which it performs the best twice. In a word, VGG-based method performs not much better than the other methods when it applied to gender recognition on non-Buddhist images.

At the same time, the results of different splits of the training set show that VGG-based method is not much dependent on the amount of training data as other image recognition based on deep CNN. Therefore, for all the experiments in the following sections, the training set is fixed to be 20% of the data.

**Gender recognition on Buddhist images**

To further prove its effectiveness of the gender recognition on Buddhist image, the VGG-based method is then applied to predict the gender of Buddhist images of Dazu Rock Carvings, Dunhuang Mogao Caves and Yungang Grottoes, which dating from the Southern Song Dynasty back to the Sui and Tang Dynasties and then the Northern Wei Dynasty.

**Gender recognition on Guanyin and Buddha in Dazu**

The Guanyin images of SMS-6 and the Buddha images of BDS-18 are collected as Data 2, shown in Fig. 6. They were engraved during the Southern Song Dynasty and the gender was certain for both of them, so the experiments can help us to validate the efficiency of VGG-based method applied to gender recognition on Buddhist images during this period.

Same as the experimental processes in “Gender recognition on non-Buddhist images” section, the feature vector of every image in Data 2 is firstly extracted; and then the dataset is divided into two parts: training set (20% of the total) and test set (80% of the total); the linear SVM is finally applied to gender classification. From the results in Fig. 5b, we can see that 100% of the Guanyin images in SMS-6 are classified as female, and 92.59% of the Buddha images in BDS-18 are recognized as male. From the

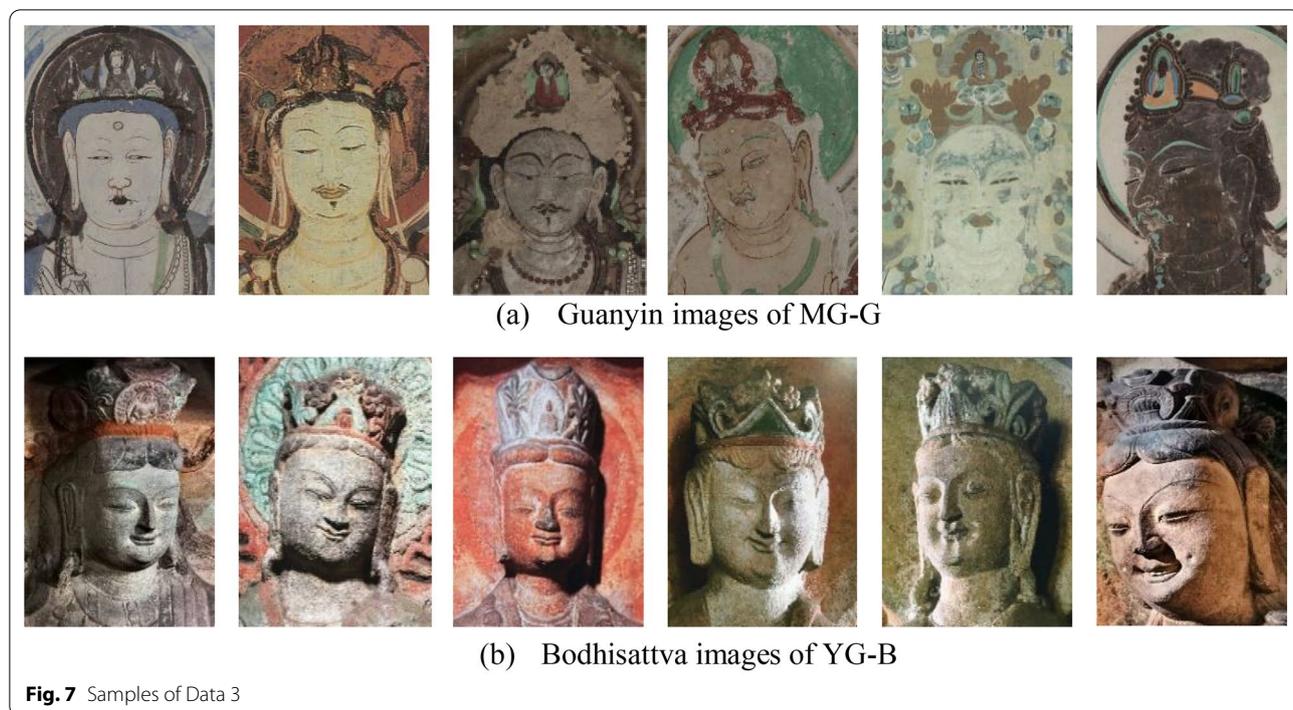


Fig. 7 Samples of Data 3

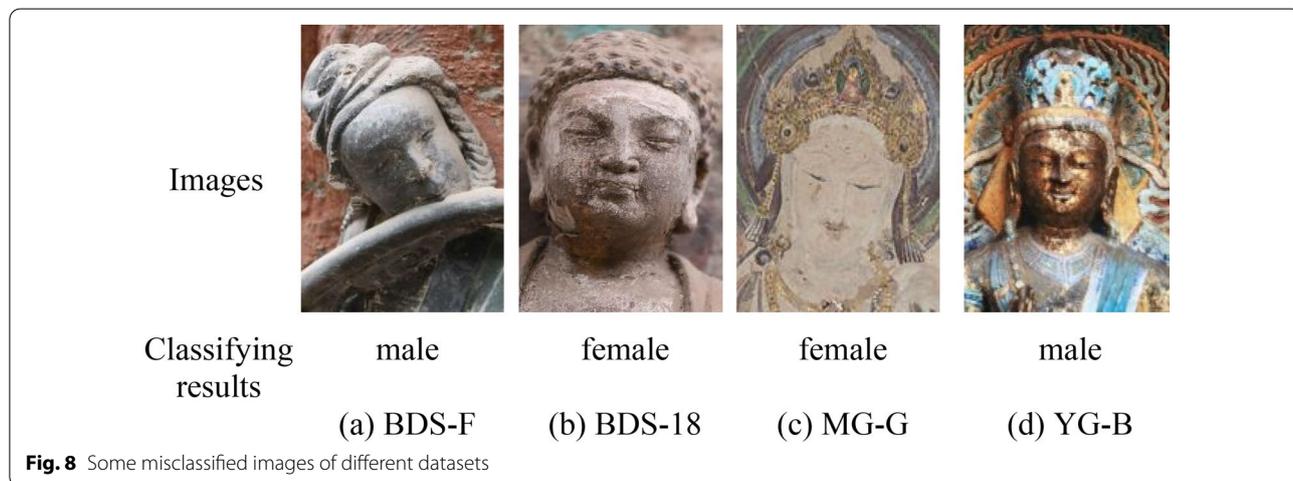


Fig. 8 Some misclassified images of different datasets

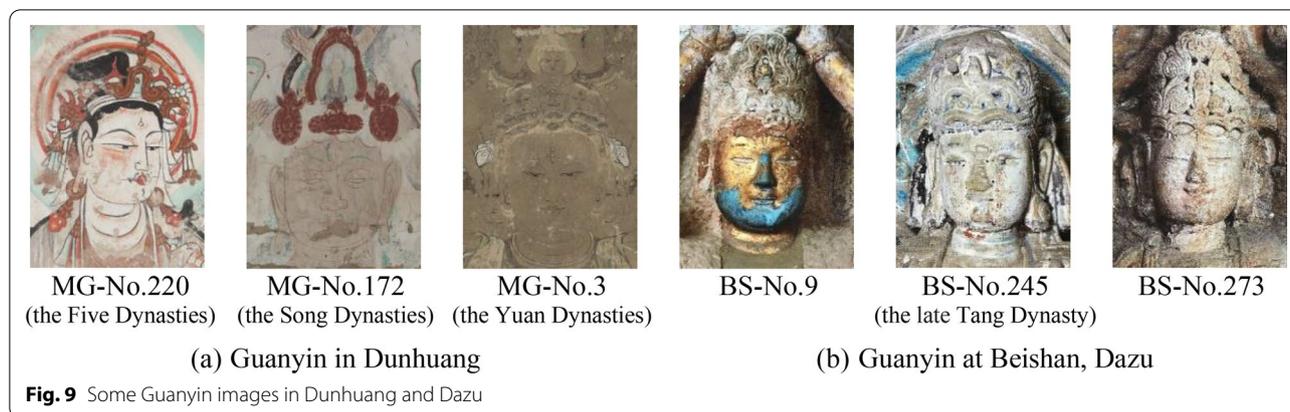
images shown in Fig. 6, we can find that there are more male characteristics for Buddha images in BDS-18, and more feminine features for Guanyin images in SMS-6. Although the misclassification of BDS-18 is 6.41%, it is shown in Fig. 8b that there is some damage on the misclassified Buddha's face. It might be the reason for the misrecognition.

Combined with the image itself, the quantitative results of this experiment not only prove that VGG-based method can be efficiently employed to gender recognition on Buddhist images, but also imply that Guanyin was

predominantly feminine and Buddha was mainly masculine in the Buddhist image making during the Southern Song Dynasty.

**Gender recognition on Guanyin in Dunhuang and bodhisattva in Yungang**

The further gender identification experiments on Buddhist images are implemented on Data 3, Guanyin frescos in Dunhuang Mogao Caves (MG-G) and bodhisattva statues in Yungang Grottoes (YG-B), shown in Fig. 7. Yungang Grottoes and Dunhuang Mogao Caves represent the



highest achievement of Buddhist art in the Northern Wei Dynasty and the Sui and Tang Dynasties, respectively.

As the belief in Guanyin was not popular during the Northern Wei Dynasty, there were only a few Guanyin images in Yungang Grottoes. However, considering that Guanyin also belongs to the series of bodhisattvas, and bodhisattvas are usually share the model during a specific period, except of the instruments in the hand or the decorations on the crown, so the bodhisattva statues of different caves in Yungang are chosen for gender identification in this paper. The experimental result is still valid for Guanyin images during that period. While the image making of Guanyin in Dunhuang was not as holistic in one whole cave as in Dazu, but scattered in different caves, so the Guanyin fresco images are also collected from different caves in Mogao Caves.

The experimental procedures applied to Data 3 are all the same as those applied to Data 1 and Data 2. The results of gender recognition are shown in Fig. 5c. It is shown from the bar graphs that 95.08% of Guanyin images in MG-G are recognized as male and 98.94% of the bodhisattva images in YG-B are classified as female. From the sample images shown in Fig. 7, it can be seen that there are tadpole-like moustache, a prominently masculine characteristic, on the faces for almost all the Guanyin images in MG-G, except of a few images as shown in Fig. 8c. While almost all the bodhisattva images in YG-B are with prominently female features. However, the misclassified image in YG-B shows some similar features to the Buddha image in BDS-18, shown in Fig. 8b, d.

Obviously, both image analysis and quantitatively experimental results can prove that the gender of Guanyin during the Sui and Tang Dynasties was mainly tend to be male, and the bodhisattvas during the Northern Wei Dynasty were with prominently feminine characteristics. As mentioned above, the results of recognition on the bodhisattvas in YG-B are valid for Guanyin during this period, so Guanyin might tend to be female during the Northern Wei Dynasty.

### **Gender analysis of Guanyin image in China**

The transformation of Indic Avalokitesvara into Chinese Guanyin is a fascinating case study of how Buddhism became indigenized in China [3]. The experimental results of gender identification carried out on the images of Guanyin and bodhisattvas in Sects. 4.2.1 and 4.2.2 can partly support the gender discussion of Guanyin in China.

As discussed in Sect. 1, with the spread of the Guanyin belief in China, Guanyin gradually feminized as time went on. Thus, there is a common idea that the earlier the bodhisattva was made, the more masculine it was. Nevertheless, according to the image analysis and quantitative results, a clue of gender transformation of Guanyin in China could be drawn as: it was mainly to be feminine in the Northern Wei Dynasty (Yungang Grottoes represented), then to be masculine tendency in the Sui and Tang Dynasties (Dunhuang frescoes represented), and to be feminine again in the Southern Song Dynasty (Dazu Rock Carvings represented).

Obviously, it is not in accord with the dominant viewpoint of the bodhisattva's feminization. In that logic, the characteristics of bodhisattvas in Yungang, Dunhuang, and Dazu should be predominantly masculine, a little feminine, predominantly feminine, respectively. But the experimental results imply that the bodhisattva's gender transformation was as early as the Northern Wei Dynasties (Yungang), and it became masculine in the Sui and Tang Dynasties (Dunhuang). It may due to the change of the relationship between image making and the Buddhist scriptures. Song Li, a distinguished Buddhist art researcher in China, sorted out the main clues of the development of bodhisattva images in China [38]. He pointed out that bodhisattva images in early China (i.e., the Wei and Jin Dynasties) is generally ambiguous due to the lack of strict adherence to the Buddhist scriptures and rituals. During the Sui and Tang dynasties, with Chinese monks' further understanding of the Buddhist



**Fig. 10** The location of Yungang Grottoes, Dunhuang Mogao Caves and Dazu Rock Carvings

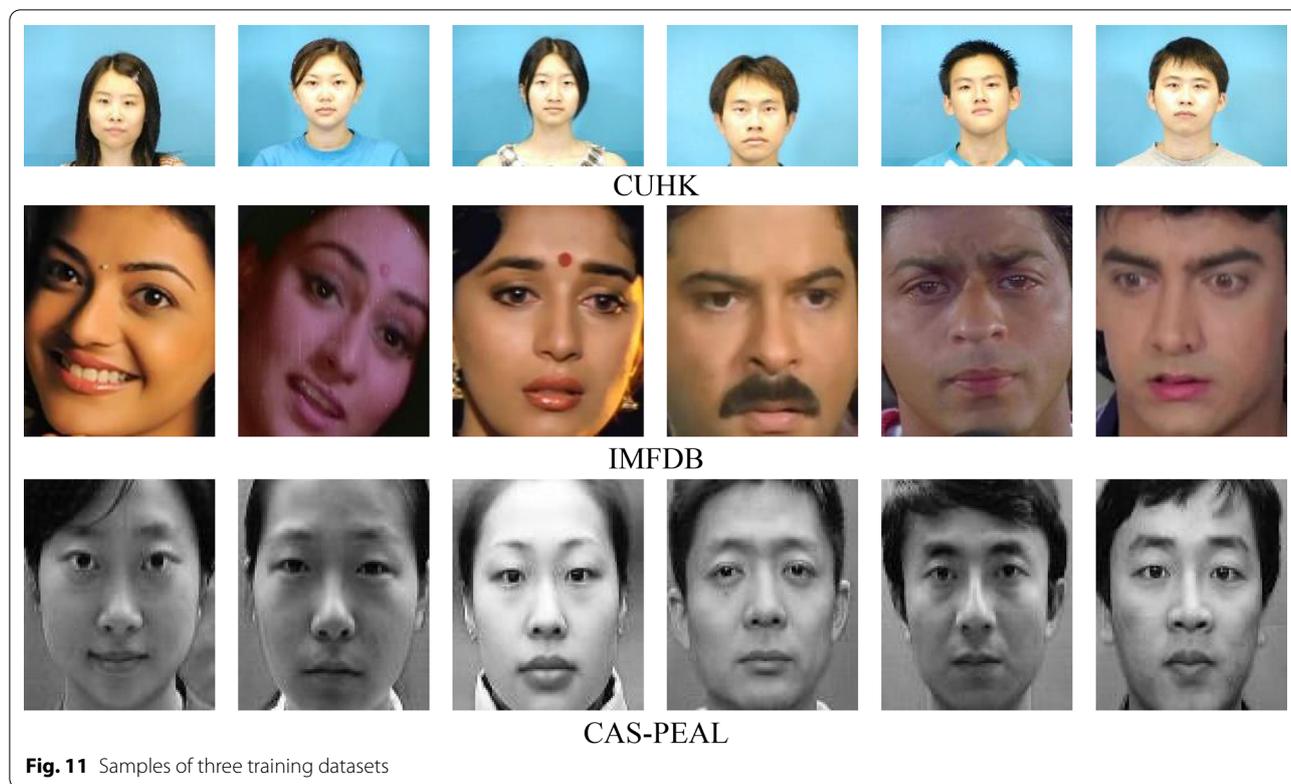
**Table 3** Average recognition ratios and standard deviations via different methods on Data 2 and 3.

Methods	SMS-6	BDS-18	MG-G	YG-B
HOG	0.9836 ±0.0120	0.8917 ± <b>0.0445</b>	0.6356 ±0.0599	0.9400 ±0.0358
LGBP-M	<b>0.9984</b> ±0.0034	0.7931 ±0.1991	0.5936 ±0.1451	0.9576 ±0.0186
LGBP-P	<b>0.9984</b> ± <b>0.0028</b>	0.8772 ±0.0511	0.7069 ±0.0946	0.9673 ± <b>0.0145</b>
WLD	0.9922 ±0.0087	0.8703 ±0.1024	0.5614 ±0.1484	0.9547 ±0.0274
LBPH	0.9914 ±0.0085	0.8864 ±0.0652	0.7374 ±0.0885	0.9234 ±0.0271
<b>VGG-based</b>	0.9975 ±0.0035	<b>0.9323</b> ±0.0708	<b>0.8767</b> ± <b>0.0509</b>	<b>0.9715</b> ±0.0239

scriptures and the direct influence of Indian Buddhist images, bodhisattva images were gradually standardized and systematized. Therefore, Guanyin images in Dunhuang were more masculine than those in Yungang due to the fact that the relationship between images and Buddhist scriptures is closer in the Sui and Tang Dynasties. Thus, the Buddhist image making in Yungang reflected much more aesthetic characteristic of scholars in the Wei and Jin Dynasties: Xiu Gu Qing Xiang (Thin and Elongated), while that in Dunhuang was more dependent on the classics. However, the indigenous scriptures, folk literatures played more important roles in Buddhist image making during the Song Dynasty, and completely broke

it away from the track of Indian Bodhisattva images and formed a Chinese system of bodhisattva images.

The geographical factor also need to be taken into account to further solve the confusion. From Fig. 10, we can see that Dunhuang Mogao Caves (in Dunhuang, Gansu Province), Yungang Grottoes (in Datong, Shanxi Province), and Dazu Rock Carvings (in Dazu, Chongqing Municipality) are in the northwestern, northern, southwestern China, respectively. Dunhuang is obviously located further west than Datong and Chongqing. Just as Yu [3] noted, in the introduction and dissemination of the faith in Guanyin, scriptures definitely played an important role. But Chinese did not simply adhere to the



**Fig. 11** Samples of three training datasets

scriptural depictions and definitions of Guanyin, nor did they strictly follow the scriptural stipulations and directions to worship Guanyin, otherwise there would not be any Chinese transformation. The craftsmen in a specific region in ancient China might have their own understanding of the Buddhist image making, and there might be a common model in a specific region. Without mass migration due to great economic or political reasons, the craftsmen in three regions hardly had the opportunity to communicate, as three sites were far from each other in ancient China. Thus, the Guanyin frescoes in the Five, Song and Yuan Dynasties in Dunhuang were always predominantly masculine for there was still moustache on the face as it was during the Sui and Tang Dynasties, shown in Figs. 9a and 7a. While the Guanyin statues in the late Tang Dynasty in Dazu were with obvious similarly female characteristics as it was later in the Song Dynasty, shown in Figs. 9b and 6b.

Therefore, an interesting and meaningful conclusion can be made that the geographical impact should not be ignored when we talk about the gender transformation of Guanyin in China, besides the effect of time, the relationship between image making and scriptures, or the intentional combination of male and female features. Archaeological and art history scholars reach an agreement that Buddhist images made in a specific location

share some universal characteristics. Moreover, from the distribution of Dunhuang Mogao Caves, Yungang Grottoes and Dazu Rock Carvings shown in Fig. 10, it may indicate that the closer to the western China it is, the more masculine the bodhisattva is, while the more inner in China, the more feminine it is. It might be attributed to the fact that the spread of Buddhism in China was from west to east.

**Efficiency comparison**

The experimental comparison on Buddhist images are also carried out to prove the superiority of the VGG-based method. The compared methods are still HOG, LBPH, WLD and LGBP-PAM. All the processes are the same as those on Data 1. Moreover, Data 2 and Data 3 are taken as a whole dataset, and it is randomly divided into two parts: 20% as a training set and 80% as a test set. The results of the comparison are shown in Table 3. In each column, the highest recognition ratio is labeled in red, and the lowest standard deviation is marked in blue.

According to the results of the average recognition ratios and the standard deviations, the superiority of VGG-based method is obvious in gender recognition on Buddhist images. The recognition ratios are 93.23%, 87.67%, 97.15% for the tests on DBS-18, MG-G, and YG-B, respectively, surpassing the second best method

**Table 4.** The accuracy of gender recognition based on different training datasets

Training	Non-Buddhist images			Buddhist images							
	SMS-10	SZS-6	BDS-F	SMS-6		BDS-18		MG-G		YG-B	
	M	M	F	M	F	M	F	M	F	M	F
CUHK	<b>0</b>	<b>0</b>	<b>1.0000</b>	0	<b>1.0000</b>	<b>0.0108</b>	0.9892	<b>0</b>	1.0000	0.0382	<b>0.9618</b>
IMFDB	<b>0.0568</b>	<b>0.2542</b>	<b>0.9750</b>	0.0559	<b>0.9441</b>	<b>0.3333</b>	0.6667	<b>0.2222</b>	0.7778	0.0828	<b>0.9172</b>
CAS-PEAL	<b>0.7955</b>	<b>0.7458</b>	<b>0.7000</b>	0.4803	<b>0.5197</b>	<b>0.9570</b>	0.0430	<b>0.5278</b>	0.4722	0.3854	<b>0.6146</b>
(a) Results based on training datasets without processing											
CUHK-gray	<b>0.8352</b>	<b>0.9661</b>	<b>0.1250</b>	0.8684	<b>0.1316</b>	<b>0.9247</b>	0.0753	<b>0.8333</b>	0.1667	0.9809	<b>0.0191</b>
IMFDB-gray	<b>0.2159</b>	<b>0.2034</b>	<b>0.3500</b>	0.2303	<b>0.7697</b>	<b>0.8817</b>	0.1183	<b>0.2389</b>	0.7611	0.5318	<b>0.4682</b>
(b) Results based on gray-processed training datasets											

by 4.06%, 13.93% and 0.42%. It implies that VGG-based method significantly outperforms the other five classical feature-based methods for the recognition on BDS-18 and MG-G, and slightly superior to the others for the classification on YG-B. Although the accuracy of LGBP-M and LGBP-P surpasses the VGG-based method for the recognition on SMS-6, it is not significant for the difference is only 0.09%. On the whole, VGGNet-based method is more powerful when it is applied to gender recognition on Buddhist images. It may be due to the fact that the most representative gender information of Buddhist images could be captured through the deep convolutional layers in VGGNet, compared to the other methods. However, the standard deviations of VGG-based method are 0.0035, 0.0708, 0.0509 and 0.0239, respectively, in which it performs the best only once. It indicates that the stability and robustness of VGG-based method are still our further work in the future. Anyway, its efficiency applied to Buddhist images is significant.

**Comparisons based on different training datasets**

To further validate the efficiency of VGG-based method, it is implemented on three different training datasets consisting of real-world facial images: CUHK [39], IMFDB [40] and CAS-PEAL [41].

Including 188 Chinese faces, the CUHK (Chinese University of Hong Kong student database) is a subset of CUHK Face Sketch database, originally for the research on face sketch synthesis and recognition. The IMFDB (Indian movie face database) is a large unconstrained face database consisting of 34,512 images of 100 Indian actors collected from more than 100 videos. It provides a detailed annotation of every image in terms of gender, age, pose and expression. The CAS-PEAL is a large-scale

Chinese face database created by Chinese Academy of Sciences (CAS), containing 99,594 images of 1040 individuals (595 males and 445 females) with varying pose, expression, accessory, and lighting (PEAL). Samples of three different datasets are shown in Fig. 11.

Three real-world facial image sets are used as training set respectively, and the images in SMS-10, SZS-6, BDS-F, SMS-6, BDS-18, MG-G, YG-B are all taken as the test set. The results of the comparison based on different training sets are shown in Table 4 (a). The gender is sufficiently definite for the non-Buddhist image, so only the results of the according gender are shown. While the gender of the Buddhist image is generally ambiguous, so the results include both female and male, denoted as F and M, respectively. According to the experimental results in previous sections and the samples shown in Data 2 and Data 3, the facial images of SMS-6, BDS-18, MG-G and YG-B are predicted as female, male, male and female, respectively. The results of the according gender are in bold. The recognition accuracy is denoted as 0–1. For the results of both non-Buddhist and Buddhist images, the highest accuracy is in bold and in purple in each column.

From the results in Table 4(a), it can be seen that the images which are predicted as female can be easily classified based on the training datasets of CUHK and IMFDB. The recognition accuracies are both 1 for BDS-F and SMS-6, and 96.18% for YG-B based on CUHK; while based on IMFDB, the accuracies are 97.5%, 94.41%, 91.72% for BDS-F, SMS-6 and YG-B, respectively. However, both of the two training sets fail in the recognition on images which are predicted as male, as the accuracies are 0, 0, 0.0108, 0 based on CUHK, and 0.0568, 0.2542, 0.3333, 0.2222 based on IMFDB for SMS-10, SZS-6, BDS-18 and MG-G, respectively. To some extent,

the recognition results based on CAS-PEAL performs much more stably and robustly for both female and male images, as the recognition accuracies are 0.7955, 0.7458, 0.9570 and 0.5278 on the images predicted as male (SMS-10, SZS-6, BDS-18, MG-G), and 0.7000, 0.5197, 0.6146 on the images predicted as female (BDS-F, SMS-6, YG-B).

Considering the facial images in CUHK and IMFDB are both color images, while those in CAS-PEAL are gray, the former two datasets are then processed to be gray to obtain more objective comparison. The experimental results are shown in Table 4(b). Obviously, it is a big reversal for the results based on CUHK before and after image-gray processing. The accuracies are improved much better on male images, 0.8352, 0.9661, 0.9247, 0.8333 on SMS-10, SZS-6, BDS-18, MG-G, respectively, versus 0, 0, 0, 0.0382 based on color images. But it is getting much worse on female images of BDS-F, MG-G, YG-B. The results are 0.1250, 0.1316, 0.0191 based on gray images, versus 1, 1, 0.9618 based on color images. While based on IMFDB, the accuracies on male images change little after and before gray-processing, except of 0.8817 versus 0.3333 on BDS-18. However, it is at the cost of worse results for the recognition on female images.

It indicates that it is complicated in gender recognition on Buddhist images when the real-world facial images are taken as the training dataset. Even though, the results based on CAS-PEAL performs not bad compared to those based on CUHK. It may be due to the fact that facial images in CAS-PEAL are mainly mature adult faces, while the images in CUHK are mainly young student faces. It implies that mature adult faces are more efficient than the student faces when they are applied as training datasets in gender recognition on Buddhist images. According to the Buddhist sutras, the Buddha gave up his royal life and became a monk to feel the pain of birth, old age, illness and death at the age of 29, and at the age of 35, he gained enlightenment under a linden tree and began to practice Buddhism. Therefore, both Buddha and bodhisattva images are depicted as an ideal, mature and virtuous man in originally Indic Buddhist art.

Hence, for the experiments of gender recognition on Buddhist images based on the real-world facial images, the training set should include sufficient mature faces. Furthermore, the recognition results based on IMFDB also indicate that the training data with Indian faces performs worse than Chinese faces, it may be the fact that both the Buddhism and its image making gradually sinicized with its spread eastward to China, then there were more Chinese rather than Indian characteristics on the faces of Buddhist images in China.

## Conclusion and future work

In order to explore gender recognition on Guanyin in China, VGGNet is applied to develop an automatic gender recognition system for Dazu Rock Carvings, Dunhuang Mogao Caves and Yungang Grottoes. The following conclusions can be made according to the quantitative results of abundant experiments.

Firstly, VGG-based system can be effectively applied to gender recognition on the Buddhist images with a 20–80% split of the training set and test set. Compared with five classical feature-based methods, VGG-based method performs better on Buddhist images, but not much better on non-Buddhist images. Anyway, the gender recognition scheme based on VGGNet is efficient when it is applied to Buddhist images.

Furthermore, when the real-world facial images are used as the training data, the recognition results are not very good. It may be due to the fact that it is quite different between the real-world faces and the facial images of carvings or frescos. Anyway, such experiments are also necessary and interesting. From the comparison based on three different kinds of real-world facial images, it can be proved that the images of Buddha and bodhisattva were both ideal and mature adult in Buddhist art, and the Buddhist image making gradually sinicized for there were more Chinese rather than Indian characteristics.

Last but the most important is that it can provide a quantitative perspective for art history. There is a common idea that the image of bodhisattvas, Guanyin as a representative, were gradually feminized with the alternation of dynasties, and then the earlier Guanyin was made, the more masculine it was. However, such stereotype is challenged by both the experimental results and the image analysis in this paper. Obviously, the feminization of Guanyin is quite a very complicated problem. Besides the time factor, the relationship between images and scriptures, or the intentional combination of male and female features, the geographical impact should not be ignored. That is because the craftsmen in different regions might have different understanding of the Buddhist images. The same masculine feature of Guanyin in different dynasties in Dunhuang Mogao Caves can be seen as the evidence. Therefore, geographical location should be considered when we talk about the clue of Guanyin's gender transformation in ancient China.

Although VGG-based system can be efficiently applied to gender recognition on Buddhist images of carvings and frescoes, it works not well when the real-world facial images are used as the training dataset. Therefore, improving its robustness based on different training datasets is becoming our future work.

## Acknowledgements

This work would not have been possible without the valuable image materials that I received from the Academy of Dazu Rock Carvings.

**Author contributions**

YH, HW and DC wrote the main manuscript text. YH, HW prepared figures. All authors reviewed the manuscript. All authors read and approved the final manuscript.

**Funding**

Chongqing Federation of Social Science, China (Grant No.2019ZCYS12).

**Availability of data and materials**

The datasets used during the current study are available from the corresponding author on reasonable request.

**Declarations****Competing interests**

Not applicable.

**Author details**

<sup>1</sup>School of Intelligent Technology and Engineering, Chongqing University of Science and Technology, Chongqing 401331, China. <sup>2</sup>College of Computer Science, Chongqing University, Chongqing 400044, China. <sup>3</sup>School of Arts and Humanity, Sichuan Fine Arts Institute, Chongqing 401331, China.

Received: 12 February 2022 Accepted: 30 May 2022

Published online: 21 June 2022

**References**

- Jie J. Change of gender: study on the feminization process of Kwan-yin in the Tang Dynasty in middle area (in Chinese). *J Guangdong Polytech Normal Univ.* 2015;4:1–9.
- Yü CF. Feminine images of Kuan-Yin in Post-T'ang China. *J Chin Religions.* 1990;18(1):61–89.
- Yü CF. Ambiguity of avalokitesvara and the scriptural sources for the cult of Kuan-yin in China. *Chung-Hwa Buddhist J.* 1997;10:409–64.
- Golomb BA, Lawrence DT, Sejnowski TJ. SEXNET: a neural network identifies sex from human faces. *Adv Neural Inform Process Syst.* 1991; 572–577.
- Moghaddam B, Yang MH. Gender classification with support vector machines//Automatic face and gesture recognition. *Proc Fourth IEEE Int Conf IEEE.* 2000;2000:306–11.
- Cortes C, Vapnik V. Support-vector networks. *Mach Learn.* 1995;20(3):273–97.
- Phillips PJ, Moon H, Rizvi SA, et al. The FERET evaluation methodology for face-recognition algorithms. *IEEE Trans Pattern Anal Mach Intell.* 2000;22(10):1090–104.
- Saatci Y, Town C. Cascaded classification of gender and facial expression using active appearance models//Automatic Face and Gesture Recognition, 2006. 7th International Conference on. IEEE, 2006, 393–398.
- Shu C, Ding X, Fang C. Histogram of the oriented gradient for face recognition. *Tsinghua Sci Technol.* 2011;16(2):216–24.
- Ullah I, Hussain M, Muhammad G, et al. Gender recognition from face images with local WLD descriptor//Systems, Signals and Image Processing (IWSSIP), 2012 19th International Conference on. IEEE, 2012, 417–420.
- Lanzara RG. Weber's law modeled by the mathematical description of a beam balance. *Math Biosci.* 1994;122(1):89–94.
- Balyan A, Suman S, Naqvi NZ, et al. Gender recognition from real-life images//Intelligent Computing in Engineering. Singapore: Springer; 2020. p. 127–34.
- Shan C. Learning local binary patterns for gender classification on real-world face images. *Pattern Recogn Lett.* 2012;33(4):431–7.
- Verschae R, Ruiz-del-Solar J, Correa M. Gender classification of faces using Adaboost. *Progr Pattern Recogn Image Anal Appl.* 2006; 68–78.
- Huang G B, Ramesh M, Berg T, et al. Labeled faces in the wild: a database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, 2007.
- Levi G, Hassner T. Age and gender classification using convolutional neural networks//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2015, 34–42.
- Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks//Advances in neural Information Processing Systems, 2012, 1097–1105.
- Ranjan R, Patel VM, Chellappa R. Hyperface: a deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. *IEEE Trans Pattern Anal Mach Intell.* 2019;41(1):121–35.
- Tompson J, Goroshin R, Jain A, et al. Efficient object localization using convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, p. 648–656.
- Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, p. 1–9.
- Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. In: European conference on computer vision. Cham: Springer; 2014. p. 818–33.
- He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016, p. 770–778.
- LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. *Proc IEEE.* 1998;86(11):2278–324.
- Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*; 2014.
- Huang G, Liu Z, Weinberger KQ, et al. Densely connected convolutional networks. *arXiv preprint arXiv:1608.06993*, 2016.
- Mittal S, Mittal S. Gender Recognition from facial images using convolutional neural network. In: 2019 Fifth International Conference on Image Information Processing (ICIIP). IEEE, 2019, p. 347–352.
- Wang H, He Z, Huang Y, et al. Bodhisattva head images modeling style recognition of Dazu Rock Carvings based on deep convolutional network. *J Cult Herit.* 2017;27:60–71.
- Press WH. Numerical recipes 3rd edition: The Art of Scientific Computing. Cambridge University Press, 2007.
- Ian Goodfellow, Yoshua Bengio and Aaron Courville. Deep Learning. MIT Press, 2016.
- Wang H, He Z, He Y, et al. Average-face-based virtual inpainting for severely damaged statues of Dazu Rock Carvings. *J Cult Herit.* 2019;36:40–50.
- Wang H, He Z, Chen D, et al. Virtual inpainting for Dazu Rock Carvings based on a sample dataset. *Journal on Computing and Cultural Heritage.* 2019;12(3):1–17.
- Fan R-E, Chang K-W, Hsieh C-J, Wang X-R, Lin C-J. LIBLINEAR: A library for large linear classification. *J Mach Learn Res.* 2008;9:1871–4.
- Li F (ed.). The Complete Works of Dazu Rock Carvings, Chongqing Press, 2017.
- Dunhuang Academy China (ed.). Dunhuang Mogao Caves, Cultural Relics Publishing House, 2013.
- Zhang Z (ed) Yungang Grottoes Buddha Statues (collector's edition). Qingdao Press, 2017.
- Zhang W, Shan S, Qing L. Are Gabor Phases really useless for face recognition? *Pattern Anal Appl.* 2009;12(3):301–7.
- Zhang W, Shan S, Gao W, Chen X. Local Gabor Binary Pattern Histogram Sequence (LGBPHS): a novel non-statistical model for face representation and recognition. In: Tenth IEEE International Conference on Computer Vision, 1, 2005, p. 786–791.
- Li S. Chinese Bodhisattva Images, Arts and Religious Civilization of Chang'an (in Chinese). In: Zhonghua Book Company, Beijing, 2002, p. 143–220.
- Wang X, Tang X. Face photo-sketch synthesis and recognition. *IEEE Trans Pattern Anal Mach Intell.* 2009;31(11):1955–67.
- Setty S, Husain M, Beham P, et al. Indian movie face database: a benchmark for face recognition under wide variations. In: Computer Vision, Pattern Recognition, Image Processing and Graphics. IEEE. 2013; 1–5.
- Gao W, Cao Bo, Shan S, et al. The CAS-PEAL large-scale Chinese face database and baseline evaluations. *IEEE Trans Syst Man Cybernet Part A.* 2008;38(1):149–61.

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.