

RESEARCH

Open Access



# Face repairing based on transfer learning method with fewer training samples: application to a Terracotta Warrior with facial cracks and a Buddha with a broken nose

Jian Zhu<sup>1,2\*</sup>, Bowei Fang<sup>1</sup>, Tianning Chen<sup>1,2</sup> and Hesong Yang<sup>1\*</sup>

## Abstract

In this paper, a method based on transfer learning is proposed to recover the three-dimensional shape of cultural relics faces from a single old photo. It can simultaneously reconstruct the three-dimensional facial structure and align the texture of the cultural relics with fewer training samples. The UV position map is used to represent the three-dimensional shape in space and act as the output of the network. A convolutional neural network is used to reconstruct the UV position map from a single 2D image. In the training process, the human face data is used for pre-training, and then a small amount of artifact data is used for fine-tuning. A deep learning model with strong generalization ability is trained with fewer artifact data, and a three-dimensional model of the cultural relic face can be reconstructed from a single old photograph. The methods can train more complex deep networks without a large amount of cultural relic data, and no over-fitting phenomenon occurs, which effectively solves the problem of fewer cultural relic samples. The method is verified by restoring a Chinese Terracotta Warrior with facial cracks and a Buddha with a broken nose. Other applications can be used in the fields such as texture recovery, facial feature extraction, and three-dimensional model estimation of the damaged cultural relics or sculptures in the photos.

**Keywords** Cultural relics face repairing, Convolutional neural network, Deep learning, Facial reconstruction, Texture alignment

## Introduction

As time went by, many sculptures were eroded by wind and rain or destroyed by war, with varying degrees of damage. In the early days, many artifacts were not scanned or archived in time by structured light.

Moreover, the history of photography technology is only two or three hundred years, which is far from the existence time of cultural relics, e.g., the Chinese Terracotta Warriors, which is known as the “eighth wonder of the world” has the history of more than 2000 years [1]. Many cultural relics have been damaged when they are photographed, which increases the difficulty of cultural relics restoration. For example, Fig. 1 shows the photograph of a Chinese Terracotta Warrior with facial cracks and a Buddha with a broken nose. Although some old photographs taken when the cultural relics were not damaged can be found, the results of the restoration based on the experience of the craftsmen and the information provided by the old photographs are highly uncertain, and the restoration results are greatly affected by the

\*Correspondence:

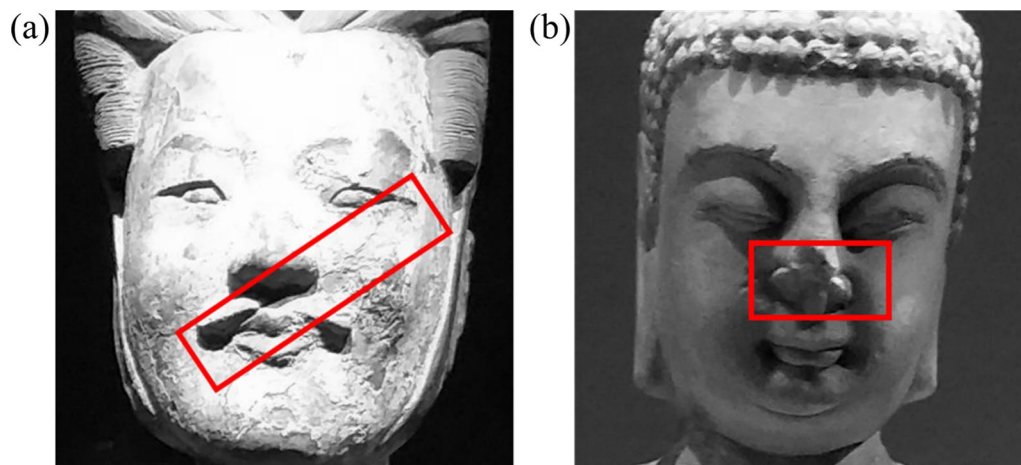
Jian Zhu  
jianzhuxj@xjtu.edu.cn  
Hesong Yang  
hesong.yang@foxmail.com

<sup>1</sup> School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, Shaanxi, People's Republic of China

<sup>2</sup> State Key Laboratory of Strength & Vibration of Mechanical Structures, Xi'an Jiaotong University, Xi'an 710049, Shaanxi, People's Republic of China



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.



**Fig. 1** Photograph of a Chinese Terracotta Warrior with facial cracks and a Buddha with a broken nose. The red panels indicate the damaged parts of the cultural relics

differences in the individual experience of the craftsmen. Additionally, in many cases, the number of photos collected is too small, and the photos are not necessarily in the same period, so it is difficult to repair them by multi-visual methods [2–4]. The information obtained from a single photo has some ambiguity, and it is difficult to determine the three-dimensional model of the artifact from a single photo.

Deep learning is an algorithm that uses artificial neural networks as a framework to characterize and learn data. It has been applied to many fields, and convolutional neural networks have solved many computer vision problems [5–10]. In 1999, Blankz and Vetter proposed the 3D deformation model (3DMM) [11], but the method relies on the accuracy of feature points and detectors [12, 13]. Recently, convolutional neural networks (CNN) have been used to predict 3DMM parameters, but they take much time [14–17]. In addition, the proposed unsupervised learning method can achieve the regression of 3DMM parameters without training data, but it is not effective in the case of occlusion and non-positive faces [18, 19]. Although the PRNet [12] has good effects, lightweight features, and can realize real-time facial reconstruction, it requires a large amount of training data, which is not suitable for the case of fewer samples. Wang et al. use the information of other similar artifacts to repair the missing artifacts, which is difficult to deal with the situation without similar data [20].

Facial texture alignment is a long-standing problem in computer vision. There are many methods for two-dimensional planes, such as the classic Active Appearance Model (AAM) and the Constrained Local Model (CLM) [21–23]. Then, the neural network-based method achieves better results but requires more data training

and cannot handle occlusion [24, 25]. Recently, some work has begun to return to the 3DMM model [14–16, 25–33], but it still needs more data, which is not suitable for the scene of cultural relics restoration. The model method can be used to achieve the goal of reconstructing the three-dimensional face alignment better, but large samples are needed for training. In the restoration of cultural relics, due to the small amount of data, an overfitting phenomenon will occur, which makes it difficult to achieve better results.

Recently, a method using convolutional neural networks has been used to estimate the depth of a photograph from a single photo [34–37], but this method cannot solve the occluded portion. As shown in previous works [20, 38, 39], the result can be obtained by using other similar types of images for alignment and averaging. This method is used to repair the sculpture that has suffered severe damage. It also relies on a lot of similar pictures of cultural relics and only achieves better results when the artifacts are not much different. Moreover, the method can only restore the image and cannot restore the three-dimensional model of the lost artifacts. Many works [26, 32, 33, 40, 41] use a lot of training data and have achieved good results in facial reconstruction, which better solves the problem of large posture and occlusion. However, the deep learning method also requires a large number of training samples, and the data of cultural relics is very limited. Even with data expansion methods such as panning, rotating, folding, scaling color channels, and adding image random noise, it is difficult to train a deep learning model with strong generalization ability.

In order to recover the three-dimensional shape of cultural relics from a single photo, this paper proposes a

training method based on transfer learning, which effectively solves the problem of fewer cultural relics training samples when using deep learning method to recover the three-dimensional shape of cultural relics and improve the credibility of cultural relics restoration. The method realizes extracting features from a single photo, reconstructing the three-dimensional model of the artifact surface, and obtaining good results. The face dataset pre-training model is used to make the model learn the basic facial features, which provides a good initial value for subsequent feature refinement. Then, a small amount of cultural relic data is used to fine-tune all the weights to make the model learn so that the model can learn the difference between the face and the face of other artifacts. The trained convolutional neural network can effectively extract the features of the facial surface, reconstruct the sculpted facial shape from a single old black-and-white photograph, and realize the texture alignment. The network trained by the method of this paper also effectively solves the problems of occlusion, side face, and shadow. In addition, it can also extract features from the face of damaged artifacts and estimate the geometry before the damage.

### The study motivation and main contributions

The purpose of this work is to train a more generalized model using less training data, to reconstruct the three-dimensional shape and texture of the cultural relic using the information provided by a single old photo, and to repair the damaged parts. The main contributions of this study are:

- (1) Use a small amount of cultural relic data to train a deep neural network with strong generalization ability based on transfer learning and realize the reconstruction of cultural relics and texture alignment from a single old photograph. Training with a small amount of data not only aligns with the current situation of limited samples in cultural relics but also offers fast training speeds and low computational costs.
- (2) Overcome the problem of side face and shadow in old photos, realize face reconstruction and texture alignment in corresponding situations.
- (3) The network trained using the method of this paper can extract features from photos of damaged artifacts and estimate the model before the damage.

### Methods

In this section, the details of the training methods proposed in this paper will be carefully described. First, the training data used in the training model is introduced. Then, the representation of the data in training

is expounded. Finally, the training methods proposed in this paper are introduced, including the specific data and training methods used in different stages, as well as the data expansion method, network structure, and loss function.

### Data used by the training model

Although the cultural relic sculpture has a certain artistic style and is different from the real person's face, it still has a lot of similarities with the real person's face in general. Therefore, this article proposes to use the face dataset for pre-training to extract the surface features of the figure sculpture roughly. The pre-training model in this article uses face data in 300W-LP. The 300W-LP is a 3D face reconstruction dataset derived from the 300W dataset and 3DMM simulation, containing annotations for 68 key points, camera parameters, and 3DMM coefficients. For the artifact data used in the training model, we scan the existing artifact model using structured light to obtain a three-dimensional model of the artifact. The data used is part of the data in the Shaanxi History Museum. For the collected cultural relics data, this paper first uses the methods of panning, rotating, folding, and zooming color channels, adding image random noise [29, 42] to expand. For the expanded data, cross-validation is used for training to improve data utilization efficiency.

### The representation of facial data

The purpose of this paper is to extract features from old photographs and to regress the three-dimensional parameters of the facial model. Therefore, we have to convert the raw data into representations of the neural network's input and output. If the three-dimensional data of the model is converted into one-dimensional data and connected by using the full connection method, the position information of the three-dimensional model in the space will be lost, and the position of the adjacent points in the space won't be reflected. At the same time, the use of the full connection layer also greatly increases the parameters that need to be trained. Moreover, the possibility of overfitting is increased because of the lack of artifact data, so this method is not used to represent data. Fan et al. [43] used a point cloud to represent the 3D model, but the maximum number of points is only 1024, which cannot effectively express the details of the face unit model. In order to save the positional relationship information between points and points in the data set, referring to previous works [12, 44–47], the UV position map is used as the alignment and reconstruction of the facial structure information, besides, as the output of the neural network.

### The method of model training

The purpose of using old photos to restore the faces of cultural relic sculptures is to extract facial features from a single old photo so as to construct the parameters of the face 3D model. If the deep learning method is used to extract features from 2D images, it can be achieved through convolutional neural networks. However, this article needs to learn a neural network model with strong generalization ability through a small amount of data, so it should make full use of the only cultural relic data, fully mine the characteristics of the image data, and introduce the information of the facial features to eliminate leaflets ambiguity in the photo. A relatively simple idea is to directly expand the cultural relics data set by translation, rotation, folding, and zooming color channels, adding image random noise, etc. Nevertheless, the combination of facial features of the expanded data is relatively single, and the training difficulty of the model is still great. In addition, the generalization ability of the model is not high. In order to solve the problems of the previous work, this paper proposes a transfer learning method. In the past, the method of transfer learning has been mainly used in the classification problem. Combined with a small amount of data in the training set, this method can better combine the features of the pre-training model, and give a more accurate classification to the target [48].

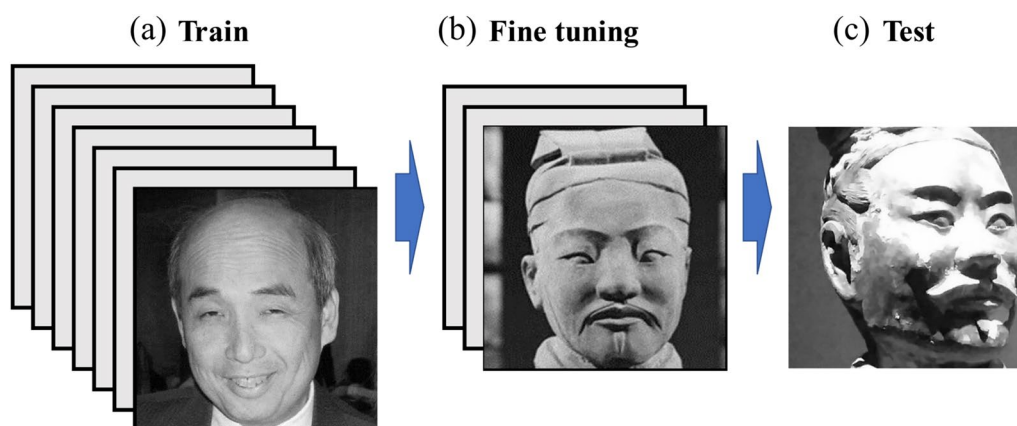
This paper first uses face data to train a large network, which can better preserve the information about human face data. The network has many different low-level and high-level kernels. The features extracted in this way are similar to the features extracted from the face of the cultural relics, which provides a good initial value for fine-tuning the three-dimensional data of the cultural relics. After training the model using face data, the model is trained using the artifact data to eliminate the difference

between the human face and the artifact face using the artifact data. Because the cultural relics data is very limited, in order to be able to use the cultural relic data, this paper uses ten-fold cross-validation to train the model using cultural relics data. The specific training process is shown in Fig. 2.

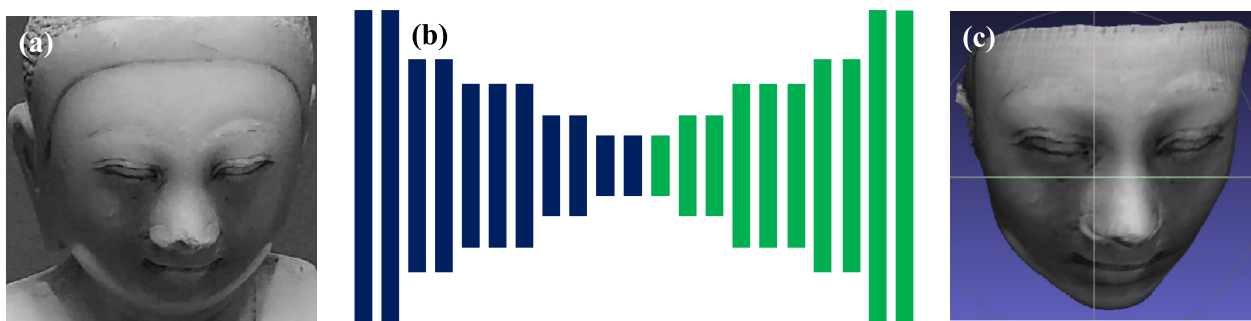
### Network structure and loss function

The network structure of this paper refers to the method of using convolutional neural networks and residual neural networks. The input of the network is a single-channel black-and-white image. This paper adopts the end-to-end learning method. A convolutional residual network was used in the previous paragraph. The first layer of the network uses a convolutional layer and then connects 10 residual network blocks to convert the input  $256 \times 256 \times 1$  single-channel image into an  $8 \times 8 \times 512$  feature map. The second half is also combined using a convolution and residual network, which contains 17 convolutional layers, and finally generates a  $256 \times 256 \times 3$  position map. For all convolutional layers, we use a filter size of  $4 \times 4$ . The activation function in the network is selected as ReLU, which can effectively solve the problem of gradient disappearance in training. The functions implemented in this article can only be implemented by one of the above network structures. The specific network structure is shown in Fig. 3.

MSE is generally chosen as the objective function for commonly used networks [30, 49]. However, the MSE loss function is equivalent to all areas on the face. But the face center has more detailed features than other parts. As a loss function, MSE cannot distinguish key points and is not suitable for learning position maps [12]. Therefore, in this paper, different weights are selected for different parts of the face. The method used to weigh the



**Fig. 2** Training process. First use the face data for pre-training, then use the artifact training data to fine-tune, and finally reconstruct the artifacts in the old photos



**Fig. 3** The network structure of cultural relics repairing. **a** Input picture. **b** The network structure. The blue part represents the residual block, and the green part represents the transposed convolution block. **c** The repaired artifact model

different parts of the face is referred to in Feng et al.'s research [12]. Since the features of the face center are more pronounced, there is a higher proportion at the center of the face. The specific ratio is 16:4:3:0.

#### Training details

The data in 300W-LP [26] was used for pre-training of the model. Since the images in the training set are all color images, the color images are first converted to black and white images to accommodate the input of the model. The corresponding 3D point cloud map is generated according to the parameters given by the dataset. By performing meshing and parameterization on the 3D point cloud and then mapping the attributes of the point cloud (such as color, coordinates, etc.) to UV space, a UV texture map is obtained. Replacing the RGB components in the UV texture map with the Cartesian coordinates of the 3D point cloud data yields the UV position map. In order to achieve better training results, the face area in the image is cut and divided, and the large-area image background is cut off, so that the main body of the image is mainly the human face.

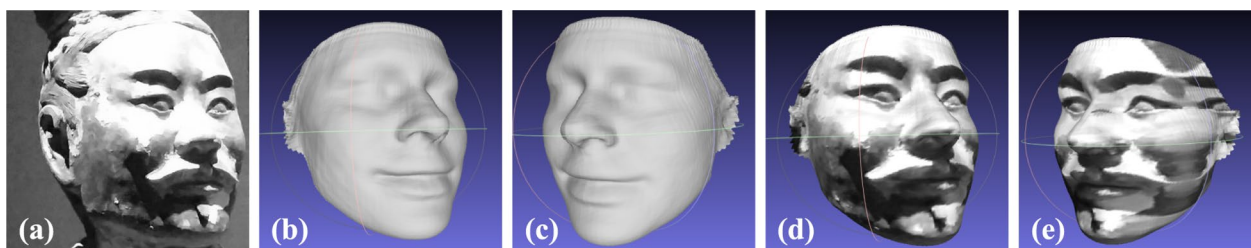
For the collected artifact data, we position, crop, and scale it to a size of  $256 \times 256$  to match the model. Similarly, the image of 300W-LP is also positioned, clipped, and scaled to a size of  $256 \times 256$ . In order to increase the

amount of data and improve the generalization ability of the model, some data transformations are made. The main transformations include panning, rotating, folding, scaling color channels, and adding image random noise. The translation transformation mainly performs random translation of the image up and down and left and right within 15%. The range of zoomed image color passes is 0.5 to 1.5. The original data is randomly rotated, and the range of random rotation is  $-45$  degrees to  $45$  degrees. The image is folded up and down and left and right, so that all possible situations can be covered by the angle range of rotation. The variance of the random noise added to the original image is 200. Adam optimizer is used during training. The learning rate starts from 0.001, and every 5 epochs become half of the original. The batch size in training is 16.

#### Results

##### Terracotta Warriors data reconstruction and mapping test

Firstly, the proposed method in this work is tested on the Terracotta Warrior. The data from the 300W-LP was first used for pre-training, then four simultaneous Terracotta Warriors were used for fine-tuning, and finally, other data was used for testing. An example of repair is shown in Fig. 4, which presents the test 3D reconstruction and texture alignment on Terracotta Warriors.



**Fig. 4** Test 3D reconstruction and texture alignment on a Terracotta Warrior. **a** The input picture of Terracotta Warrior. **b, c** Three-dimensional model of the Terracotta Warrior face. **d, e** Three-dimensional model of the Terracotta Warrior with texture

Figure 4a is the input picture of the Terracotta Warrior, and the three-dimensional model of the Terracotta Warrior face without and with texture are shown in Fig. 4b–e, respectively. The iterative closest point (ICP) method is an algorithm that iteratively seeks the optimal rigid transformation between two point clouds under certain constraints, aiming to align them as closely as possible. The corresponding nearest point will be found using the ICP of the restored 3D model and the original 3D model, and then it will be normalized using the distance outside the eye. The mean square error (MSE) was calculated and the MSE for all test samples was 4.19. As can be seen from the figure, the position corresponding to reconstruction and mapping is better.

#### Reconstruction map test under small sample

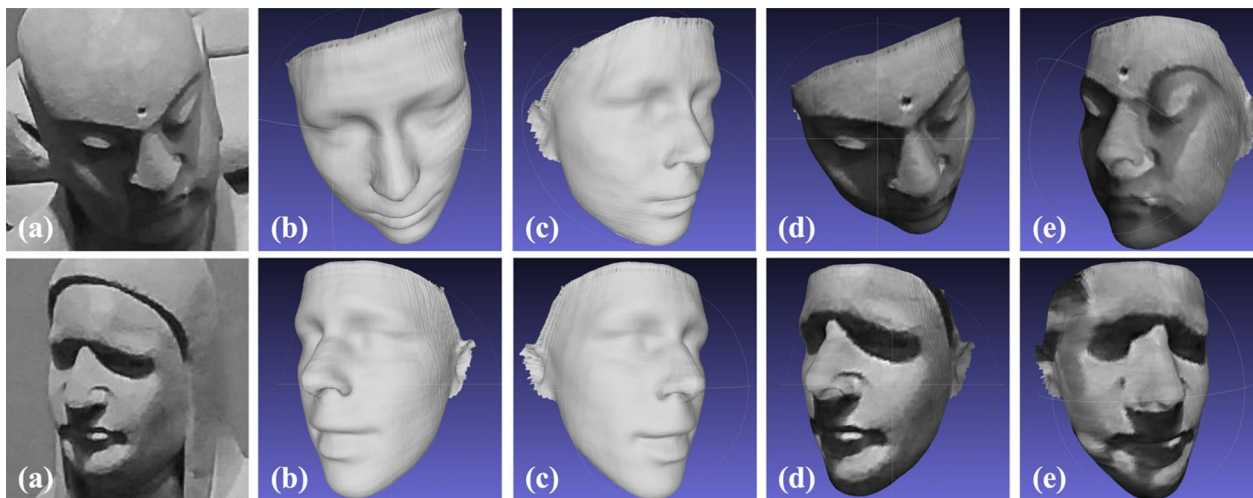
The sample was also tested in a very small case, first using the data in the 300W-LP for pre-training, then using two Luo Han from the Northern Song Dynasty to fine-tune the model, and finally using the following images with shadows and side faces to test the model. The mean square error (MSE) on all test samples was 5.73. The

result of restoring the 3D model from the old photo is shown in Fig. 5. As can be seen from the figure, our model has achieved good results in the reconstruction of the facial model with large shadows, side faces, and shadows in the old photos. However, the texture of occlusion and shadows is not well estimated.

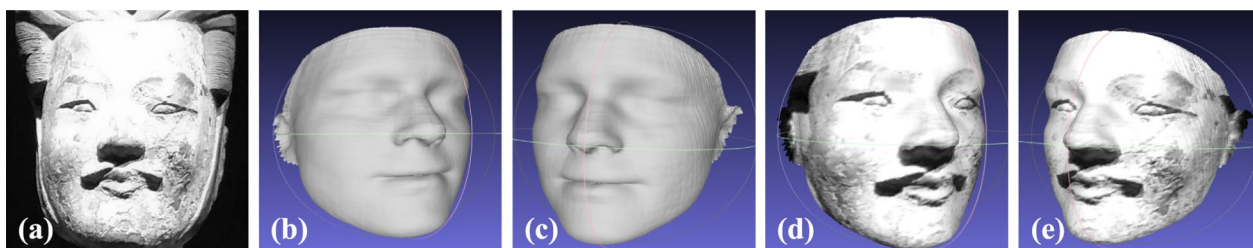
#### Restoration of damaged artifacts from old photos

As for the damaged Chinese objects in the old photos, we used the Terracotta Warrior data and the Buddha data to test. In both tests, 300W-LP was used for pre-training. Then, four samples were used for fine adjustment of the Terracotta Warrior and the Buddha images, and other data were used for testing.

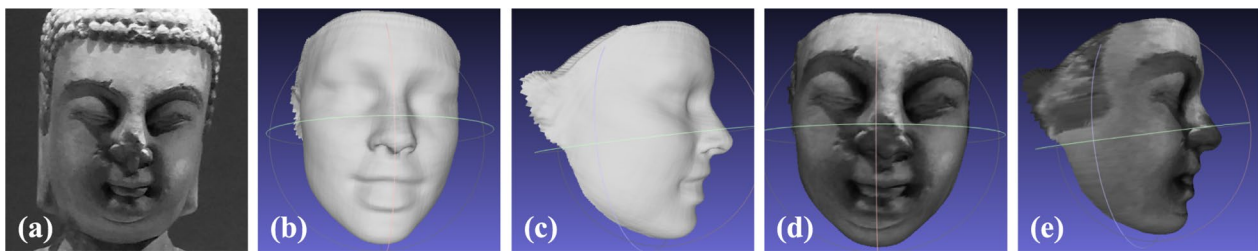
Figure 6 shows the restoration of a damaged Terracotta Warrior from an old photo. Figure 6(a) represents the input picture of the Terracotta Warrior, and Fig. 6b–e illustrate the three-dimensional model of the restored Terracotta Warrior without and with texture. Figure 7 shows the restoration of a damaged Buddha in an old photo. Figure 7a is the input picture of the Buddha, and Fig. 7b–e represent the three-dimensional model of the



**Fig. 5** Test 3D reconstruction and texture alignment on two Luo Hans. **a** The input picture of Luo Hans. **b, c** Three-dimensional model of the Luo Hans' face. **d, e** Three-dimensional model of Luo Hans with texture



**Fig. 6** Restoration of a damaged Terracotta Warrior in an old photo. **a** The input picture of the Terracotta Warrior. **b, c** The three-dimensional model of the restored Terracotta Warrior. **d, e** The three-dimensional model of the restored Terracotta Warrior with texture



**Fig. 7** Restoration of a damaged Buddha in an old photo. **a** The input picture of the Buddha. **b, c** The three-dimensional model of the restored Buddha. **d, e** The three-dimensional model of the restored Buddha with texture

restored Buddha without and with texture. It can be seen from the results that for the problems of facial cracks and a broken nose in old photos, the model in this work can extract the features in the photos and repair very well of the corresponding 3D models.

### Discussion and conclusion

Reconstructing and repairing the faces of cultural relics is of great significance for presenting the original appearance of relics, enhancing overall aesthetics and artistic value, and promoting historical and cultural research. This paper addresses the current situation of limited cultural relic samples by training a neural network with strong generalization ability based on transfer learning, achieving reconstruction and texture alignment of cultural relics from a single photo. Compared with previous methods, the method in this paper requires fewer relic samples and can address issues such as occlusion, side faces, and shadows. Additionally, both the facially damaged Terracotta Warrior and the broken-nosed Buddha have been effectively repaired, providing a model for similar relic restoration issues. However, due to the limitation of relic quantity, this paper only tested two types of relics. In future research, more relics with facial defects can be reconstructed to verify the algorithm's generalization ability. Furthermore, the damaged areas of the relics in this paper are not extensive, and subsequent studies can verify and research cases with large areas of facial damage.

In summary, this paper proposes a method of training end-to-end networks by transfer learning methods. The results show that the method makes it possible to train a deep neural network model with strong generalization ability using a small amount of data. The Terracotta Warrior and Buddha are taken as examples to demonstrate that the model can better realize the extraction of information from a single old photo, and the establishment of a face 3D model and texture alignment. The model can also solve the problem of side face, shadow, and occlusion in old photos. In addition, the model is further extended

to feature extraction and 3D model reconstruction for the old photos of damaged statues. After the restoration, the texture alignment is realized, and a better effect is achieved, which provides a reference for the restoration of cultural relics.

### Acknowledgements

Not applicable.

### Author contributions

All the authors contributed to the current work. Conceptualization: JZ, HY and TC; methodology, software, validation, formal analysis, investigation, resources, and writing—original draft preparation: JZ, BF, HY; review and editing: JZ; supervision, and funding acquisition: JZ and TC.

### Funding

This work was financially supported by the China Postdoctoral Science Foundation (No. 2022M712540) and the State Key Laboratory for Strength and Vibration of Mechanical Structures (No. SV2023-KF-08).

### Availability of data and materials

Data can be made available on request.

### Declarations

#### Competing interests

The authors declare that they have no competing interests.

Received: 22 March 2024 Accepted: 21 May 2024

Published online: 07 June 2024

### References

1. Yang K, Cao X, Geng G, Li K, Zhou M. Classification of 3D Terracotta Warriors fragments based on geospatial and texture information. *J Vis.* 2021;24(2):251–9.
2. Snavely N, Seitz S M, Szeliski R. Photo tourism: exploring photo collections in 3D. *ACM siggraph 2006 papers.* 2006; 835–846.
3. Sharma S, Kumar V. 3D face reconstruction in deep learning era: a survey. *Arch Computat Methods Eng.* 2022;29(5):3475–507.
4. Deng Z, Liang Y, Pan J, Liao J, Hao Y, Wen X. Fast 3D face reconstruction from a single image combining attention mechanism and graph convolutional network. *Vis Comput.* 2023;39(11):5547–61.
5. Di Angelo L, Di Stefano P, Guardiani E. A review of computer-based methods for classification and reconstruction of 3D high-density scanned archaeological pottery. *J Cult Herit.* 2022;56:10–24.
6. Chen M, Zang S, Ai Z, Chi J, Yang G, Chen C, et al. RFA-Net: Residual feature attention network for fine-grained image inpainting. *Eng Appl Artif Intell.* 2023;119: 105814.

7. Qin Z, Zeng Q, Zong Y, Xu F. Image inpainting based on deep learning: a review. *Displays*. 2021;69: 102028.
8. Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans Pattern Anal Mach Intell*. 2017;40(4):834–48.
9. Wang W, Shen J, Ling H. A deep network solution for attention and aesthetics aware photo cropping. *IEEE Trans Pattern Anal Mach Intell*. 2018;41(7):1531–44.
10. Bejnordi BE, Veta M, Van Diest PJ, Van Ginneken B, Karssemeijer N, Litjens G, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA*. 2017;318(22):2199–210.
11. Blanz V, Vetter T. A morphable model for the synthesis of 3D faces. *Sem Graph Papers Push Bound*. 2023;2:157–64.
12. Feng Y, Wu F, Shao X, Wang Y, Zhou X. Joint 3d face reconstruction and dense alignment with position map regression network. *Proceedings of the European conference on computer vision (ECCV)*. 2018; 534–51.
13. Zhao R, Wang Y, Benitez-Quiroz CF, Liu Y, Martinez AM. Fast and Precise Face Alignment and 3D Shape Reconstruction from a Single 2D Image. In: Hua G, Jégou H, editors. *Computer Vision—ECCV 2016 Workshops*. Cham: Springer International Publishing; 2016. p. 590–603.
14. Richardson E, Sela M, Or-El R, Kimmel R. Learning detailed face reconstruction from a single image. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017; 1259–68.
15. Richardson E, Sela M, Kimmel R. 3D face reconstruction by learning from synthetic data. *2016 fourth international conference on 3D vision (3DV)*. 2016; 460–9.
16. Jourabloo A, Liu X. Large-pose face alignment via CNN-based dense 3D model fitting. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016; 4188–96.
17. Peng X, Feris RS, Wang X, Metaxas DN. A recurrent encoder-decoder network for sequential face alignment. In: Leibe B, Matas J, Sebe N, Welling M, editors. *Computer Vision—ECCV 2016*. Cham: Springer International Publishing; 2016. p. 38–56.
18. Tewari A, Zollhofer M, Kim H, Garrido P, Bernard F, Perez P, et al. Mofa: Model-based deep convolutional face autoencoder for unsupervised monocular reconstruction. *Proceedings of the IEEE international conference on computer vision workshops*. 2017; 1274–83.
19. Bas A, Huber P, Smith WA, Awais M, Kittler J. 3D morphable models as spatial transformer networks. *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 2017; 904–12.
20. Wang H, He Z, He Y, Chen D, Huang Y. Average-face-based virtual inpainting for severely damaged statues of Dazu Rock Carvings. *J Cult Herit*. 2019;36:40–50.
21. Asthana A, Zafeiriou S, Cheng S, Pantic M. Robust discriminative response map fitting with constrained local models. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2013; 3444–51.
22. Kim J, Liu C, Sha F, Grauman K. Deformable spatial pyramid matching for fast dense correspondences. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2013; 2307–14.
23. Saragih J, Goecke R. A nonlinear discriminative approach to AAM fitting. *2007 IEEE 11th International Conference on Computer Vision*. 2007;1–8.
24. Xiong X, De la Torre F. Global supervised descent method. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015;2664–73.
25. Dollár P, Welinder P, Perona P. Cascaded pose regression. *2010 IEEE computer society conference on computer vision and pattern recognition*. 2010;1078–85.
26. Zhu X, Lei Z, Liu X, Shi H, Li SZ. Face alignment across large poses: A 3d solution. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016; 146–55.
27. Jourabloo A, Liu X. Pose-invariant 3D face alignment. *Proceedings of the IEEE international conference on computer vision*. 2015; 3694–702.
28. Tran AT, Hassner T, Masi I, Medioni G. Regressing robust and discriminative 3d morphable models with a very deep neural network. *2017 IEEE Conference on computer vision and pattern recognition (CVPR)*. 2017; 1493–502.
29. Dou P, Shah SK, Kakadiaris IA. End-to-end 3D face reconstruction with deep neural networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017; 5908–17.
30. Yu R, Saito S, Li H, Ceylan D, Li H. Learning dense facial correspondences in unconstrained images. *Proceedings of the IEEE international conference on computer vision*. 2017; 4723–32.
31. Alp Guler R, Trigeorgis G, Antonakos E, Snape P, Zafeiriou S, Kokkinos I. Densereg: Fully convolutional dense shape regression in-the-wild. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017; 6799–808.
32. Liu F, Zeng D, Zhao Q, Liu X. Joint Face Alignment and 3D Face Reconstruction. In: Leibe B, Matas J, Sebe N, Welling M, editors. *Computer Vision—ECCV 2016*. Cham: Springer International Publishing; 2016. p. 545–60.
33. Liu Y, Jourabloo A, Ren W, Liu X. Dense face alignment. *Proceedings of the IEEE international conference on computer vision workshops*. 2017;1619–28.
34. Riegler G, Liao Y, Donne S, Koltun V, Geiger A. Connecting the dots: Learning representations for active monocular depth estimation. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019;7624–33.
35. Blanchon M, Sidibé D, Morel O, Seulin R, Braun D, Meriaudeau F. P2D: a self-supervised method for depth estimation from polarimetry. *2020 25th International Conference on Pattern Recognition (ICPR)*. 2021;7357–64.
36. Song M, Kim W. Depth estimation from a single image using guided deep network. *IEEE Access*. 2019;7:142595–606.
37. Zhang Z, Xu C, Yang J, Gao J, Cui Z. Progressive hard-mining network for monocular depth estimation. *IEEE Trans Image Process*. 2018;27(8):3691–702.
38. Huang J, Nara K, Zong K, Wang J, Xue S, Peng K, et al. Ectomycorrhizal fungal communities associated with Masson pine (*Pinus massoniana*) and white oak (*Quercus fabri*) in a manganese mining region in Hunan Province. *China Fungal Ecol*. 2014;9(1):1–10.
39. Hays J, Efros AA. Scene completion using millions of photographs. *Commun ACM*. 2008;51(10):87–94.
40. Jackson AS, Bulat A, Argyriou V, Tzimiropoulos G. Large pose 3D face reconstruction from a single image via direct volumetric CNN regression. *Proceedings of the IEEE international conference on computer vision*. 2017; 1031–9.
41. Bulat A, Tzimiropoulos G. How far are we from solving the 2d & 3d face alignment problem?(and a dataset of 230,000 3d facial landmarks). *Proceedings of the IEEE international conference on computer vision*. 2017; 1021–30.
42. Saito S, Li T, Li H. Real-Time Facial Segmentation and Performance Capture from RGB Input. In: Leibe B, Matas J, Sebe N, Welling M, editors. *Computer Vision—ECCV 2016*. Cham: Springer International Publishing; 2016. p. 244–61.
43. Fan H, Su H, Guibas LJ. A point set generation network for 3d object reconstruction from a single image. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017; 605–13.
44. Xue N, Deng J, Cheng S, Panagakis Y, Zafeiriou S. Side information for face completion: a robust PCA approach. *IEEE Trans Pattern Anal Mach Intell*. 2019;41(10):2349–64.
45. Moschoglou S, Ververas E, Panagakis Y, Nicolaou MA, Zafeiriou S. Multi-attribute robust component analysis for facial uv maps. *IEEE J Sel Top Signal Process*. 2018;12(6):1324–37.
46. Deng J, Cheng S, Xue N, Zhou Y, Zafeiriou S. UV-GAN: Adversarial Facial UV Map Completion for Pose-Invariant Face Recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018; 7093–102.
47. Floater MS. Parametrization and smooth approximation of surface triangulations. *Comput Aided Geom Des*. 1997;14(3):231–50.
48. Pan SJ, Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng*. 2009;22(10):1345–59.
49. Crispell D, Bazik M. Pix2face: Direct 3d face model estimation. *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 2017; 2512–8.

## Publisher's Note

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