J Vis (2021) 24:251–259
https://doi.org/10.1007/s12650-020-00710-6

REGULAR PAPER

Kang Yang · Xin Cao · Guohua Geng · Kang Li · Mingquan Zhou

Classification of 3D terracotta warriors fragments based on geospatial and texture information

Received: 18 June 2020 / Revised: 20 August 2020 / Accepted: 16 September 2020 / Published online: 3 January 2021

Abstract The accurate classification of the fragments is a critical step in the restoration of the Terracotta Warriors. However, the traditional manual-based method is time-consuming and labor-intensive, and the accuracy mainly depends on the archeologist's experience. In this paper, we present a novel classification framework for the 3D Terracotta Warriors fragments. The core of our framework is a dual-modal based neural network, which can incorporate geospatial and texture information of the fragments and output the category of each fragment. The geospatial information is extracted from the point cloud directly. At the same time, a method based on the 3D mesh model and improved Canny edge detection algorithm is proposed to extract the texture information. As to the real-world data experiments, the dataset includes 800 pieces of the arm, 810 pieces of the body, 810 pieces of head and 830 pieces of leg, and the mean accuracy rate is 91.41%, which is better than other existing methods, which only based on geospatial information or texture information. We hope our framework can provide a useful tool for the virtual restoration of the Terracotta Warriors.

Keywords Terracotta warriors fragments · Virtual restoration · Deep learning · Point cloud · Texture information

1 Introduction

As one of the Four Ancient Civilizations, China has a variety of cultural relics. One well-known cultural relic is the Terracotta Warriors, which is the first batch of Chinese World Heritage and known as ''the eighth wonder of the world.'' Therefore, the research on the Terracotta Warriors has highly valuable for their historical, cultural, and scientific significance. However, after thousands of years of weathering erosion, a significant number of the excavated relics have been damaged and scattered randomly. Many archeologists

K. Yang · X. Cao (✉) · G. Geng · K. Li (✉)
School of Information Science and Technology, Northwest University, Xi'an, Shaanxi, China
E-mail: xin_cao@163.com

K. Li
E-mail: likang@nwu.edu.cn

K. Yang
E-mail: 2017117334@stumail.nwu.edu.cn

G. Geng
E-mail: ghgeng@nwu.edu.cn

M. Zhou
School of Information Science and Technology, Beijing Normal University, Beijing, China
E-mail: mqzhou@bnu.edu.cn

and researchers are dedicated to the restoration of the Terracotta Warriors by resembling the relic fragments manually. With the computer-aid technology, the restoration can be done in a virtual space, which can guide the restoration of real-world Terracotta Warriors. Typically, virtual restoration can be divided into two steps. The first step is the classification of these fragments, while the second step is the reassembly of the fragments belong to one category. In this work, we focus on the first step, which is an essential step in the reconstruction of the Terracotta Warriors.

For the classification of relic fragments, traditional methods are mainly based on the information of the shape, texture, color, material characteristics, decoration, and technological elements, etc. Kampel et al. proposed a classification method based on color information (Kampel and Sablatnig 2000). The key step is the color calibration under the illustration of visible light. Similarly, Zhou et al. utilized the color properties of the porcelain image and developed a specific system for Yao Zhou's porcelain fragments classification (Zhou et al. 2011). Kampel et al. divided the complete profile into relevant sub-parts. These sub-parts, named as two-dimensional (2D) profiles, were used to classify the archeological pottery fragments (Kampel et al. 2001). Karasik et al. recognized that if their profiles can completely characterize the shape of the fragments, the ceramic would be classified well based on their profile morphology (Karasik and Smilansky 2011). Smith et al. attempted to classify the ceramic fragments based on texture and color descriptors (Smith et al. 2010). The descriptor vector of a fragment image is composed of the color histogram and total variation geometry. The proposed descriptor accurately represents texture. However, the result of classification was unsatisfactory for those images lack adequate texture and minor color and intensity variations. Qi et al. used wavelet transformations to extract surface texture characteristic (Li-Ying and Ke-Gang 2010). Then, the fuzzy C-means algorithm was applied to classify ceramic fragments. Based on the empirical mode decomposition (EMD) technique, Kang et al. proposed a method to identify the salient geometric features on the surface of relic fragments and then classified the fragments (Kang et al. 2015). Their experimental results corroborated that this method obtains a high accuracy for the classification of Terracotta Warrior fragments. Zhao et al. further improved the previous method by using Hu invariant moment to extract the shape features of fragments, which are not apparent in surface features (Zhao and Geng 2018). The support vector machine (SVM) classifier is used to classify relic fragments according to extracted features.

Although the traditional methods can work well for the classification of relic fragments, they usually require human experts to design the accurate descriptor to extract features laboriously. Nowadays, with the development of computer science and big data, the classification enters the era of data-driven methods, especially the deep learning-based methods have shown great potential in this field. Liu et al. advanced a 3D residual neural network algorithm to classify the Terracotta Warriors, which complies with the requirements of classification (Liu 2019). Wang et al. delivered 2D images of the Terracotta Warriors to a convolutional neural network for data training; the features for classification are extracted automatically instead of manually in this method (Wang 2019). In Su et al. 2015, multi-view CNNs are used for classification. Two-dimensional images from 3D shapes rendered views were fed into 2D CNN and output the classification result. Besides, some other studies have been carried out in voxelized shapes based on 3D CNN (Maturana and Scherer 2015; Wu et al. 2015; Qi et al. 2016). However, volumetric representation is constrained by its resolution as a result of data sparsity and computational complexity. Yet, this method has two main limitations. First, the point cloud data obtained by 3D scanner should be converted to voxel or mesh model, which increases the time cost; Second, compared to point cloud, the converted voxel or mesh model may lose some essential geometric details provided by the point cloud. The architecture, PointNet, proposed by Charles et al., has empowered the deep learning to consume the point cloud directly (Qi et al. 2017a, b). A single symmetric function is applied to ensure the input order invariance of a point set, which is the key step in PointNet. Later, some other methods are also proposed, such as PointNet ++, PointCNN and so on (Qi et al. 2017a, b; Li et al. 2018; Griffiths and Boehm 2019; Wang and Kim 2019; Xie et al. 2019).

Depending on the geospatial and texture information of 3D Terracotta Warrior fragments, in this paper, we propose a novel framework to classify these fragments. Specifically, to extract the geospatial features, a PointNet-based network is used. As to the texture features, the Inception-V4, one famous CNN architecture, is adopted (Szegedy et al. 2017). Details are described in Sect. 2. A series of 3D Terracotta Warrior fragments-based experiments are designed and conducted, and the results demonstrate that compared with the methods only based on geospatial or texture features, our proposed method can improve the classification accuracy significantly.

2 Materials and methods

In this paper, the fragments of 3D Terracotta Warriors were selected as the research subjects. It should be noted that these data were acquired by using a Creaform VIU handy scanner. The scan resolution was 3.91 mm, which favors speed but results in relatively low precision. Finally, the data are converted into the mesh model with a texture image by using engineering software (Geomagic Studio 2012, Raindrop Geomagic, USA).

The proposed framework can be divided into two parts. The first part is the data preparation, which can output the point cloud and 2D texture images. The second part is the training of the dual-modal-based neural network, which outputs the category of the 3D Terracotta Warrior fragment.

In the data preprocessing stage, all images and point cloud data were uniformly adjusted to the prescribed format. After that, they are put into a dual-modal-based neural network for training and classification. The overall flowchart of the proposed framework is shown in Fig. 1.

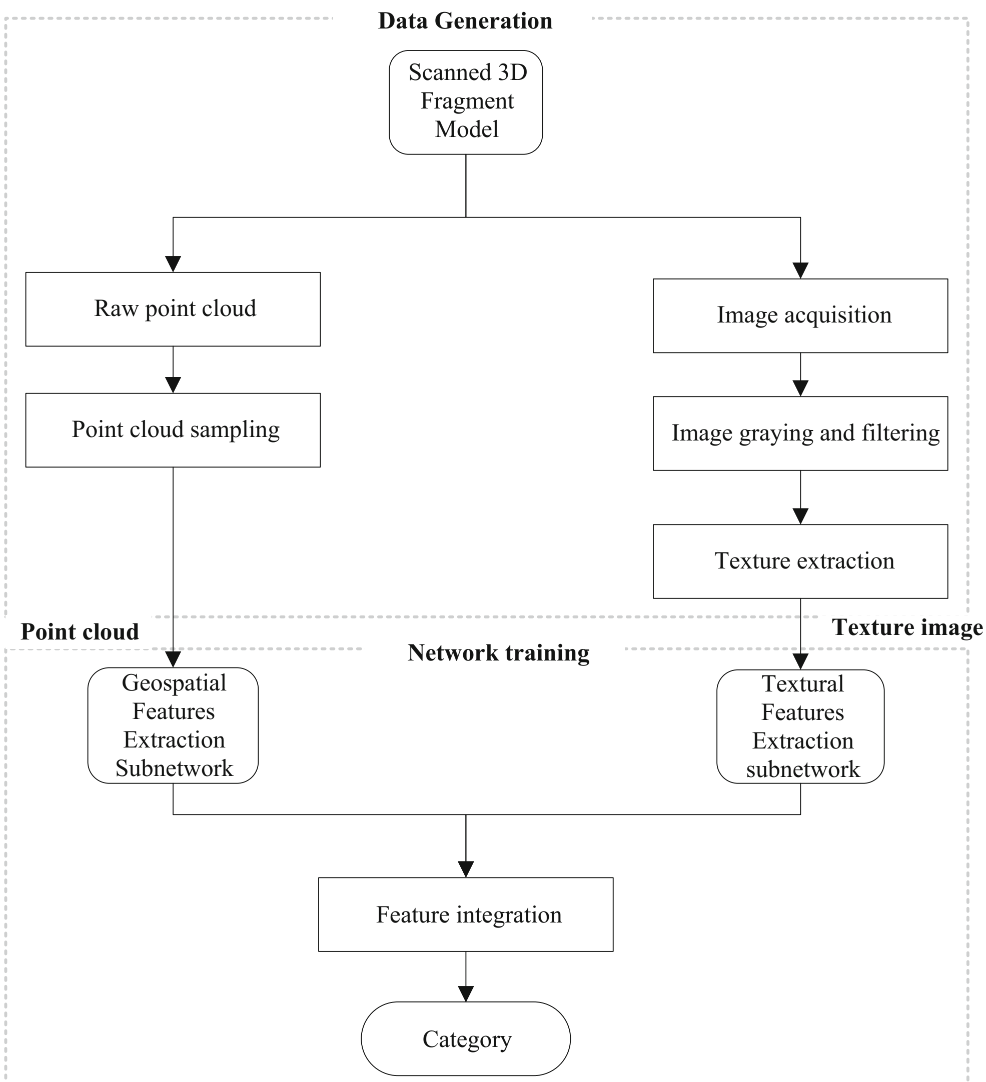

Fig. 1 Flow chart of the overall framework

2.1 Data generation

2.1.1 Point cloud generation

The raw point cloud can be extracted by Geomagic software. However, due to the high resolution of the 3D scanner, the raw point cloud usually consists of large numbers of data points with a dense distribution. As the neural network designed in this paper limits the size of the point cloud, the raw point cloud cannot be directly consumed. To this end, farthest point sampling (FPS) (Moenning and Dodgson 2003) is used to downsample the raw point cloud to generate a new point cloud with the point number of 2048. Figure 2 displays the result of downsampling on the head fragments of one of the Terracotta Warriors.

2.1.2 Texture image generation

(A) Image acquisition.

To ensure that the texture of fragment is consistently located at the centre of the picture, and the size of the image is uniform, the mesh model of the fragment is manually rotated at an appropriate angle to generate an image of 512 $\times$ 512 in the Geomagic software, as shown in Fig. 3.

(B) Image graying and filtering.

Image graying refers to the conversion of a color image to a grayscale image. Three components determine the color of each pixel in the color image: R, G, and B, and each component vary from 0 to 255. Therefore, one pixel in a color image has more than 16 million (256 $\times$ 256 $\times$ 256) colors. The huge amount of information reduces the operation rate. Moreover, the color itself is challenging to provide key information. Compared with a general color image, the grayscale image has the same components of R, G, and B, so that one pixel has a varied range of 256 colors. The gray image can greatly improve the rate of calculation for subsequent works by convert color images to grayscale images.

Specifically, the relationship between the gray value Y of the pixel in the color image and the R, G, and B values of the pixel can be determined by (1).

$$Y = 0.3R + 0.59G + 0.11B \tag{1}$$

To reduce the impact of noise, this paper applies Gaussian filter technology to image filtering. Firstly, a two-dimensional Gaussian matrix is determined by each pixel and its domain pixels in the image, as shown in (2). Then, calculating the inner product of the template and each pixel in the image. Finally, the central pixel value is replaced by the weighted mean value, which is computed as (3).

$$G(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \tag{2}$$

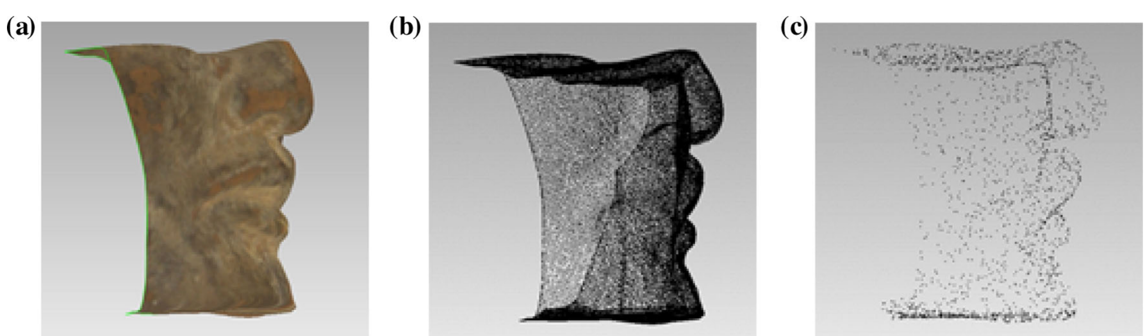

Fig. 2 Point cloud generation. **a** the mesh model of head fragment of a Terracotta Warrior; **b** the raw point cloud of (**a**), with the point number of 42,867; (**c**) the downsampled point cloud of (**b**)

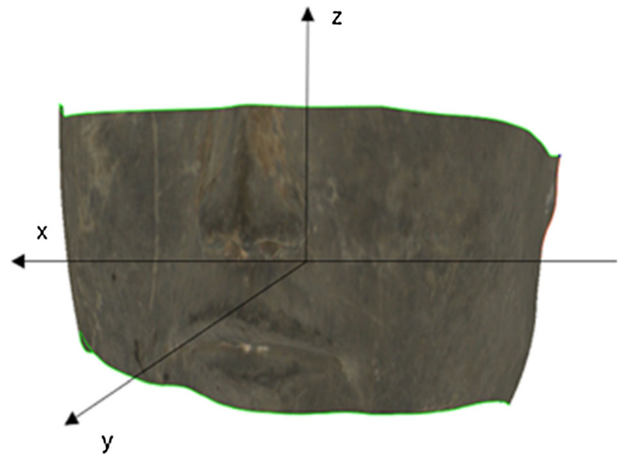

Fig. 3 Rotated image

$$I(x, y) = G(x, y) \times f(x, y) \tag{3}$$

(C) Texture image generation.

In our framework, the texture image of fragments is regarded as one of the most important elements for classification. Due to a series of reasons, the surface of the terracotta debris has lost color, resulting in only grayscale information in the collected image. To avoid the interference caused by the uneven gray value, we only use the edge information of the image to generate the texture images. An improved canny edge detection algorithm is proposed to extract the texture image of the fragments, and the details of the texture image extraction strategy are described as follows:

Algorithm: Texture image extraction strategy

Initialize. The gray image of the 3D Terracotta Warriors fragment

Step 1: Using the Gaussian filter to smooth the image and eliminate the noise.

Step 2: Estimate the edge intensity and direction at each point, calculating the gradient amplitude and direction of each pixel in the image.

Step 3: Detecting the contour by using on double threshold algorithm. Initially, selecting the appropriate high (TH) and low (TL) thresholds. Then, points with pixel values lower than TL are discarded, and points with pixel values higher than TH are marked (these are defined contour points). Finally, the point is retained if the value is between two thresholds and connected to a point higher than TH.

Step 4: Using the expansion operator to close the non-closed contour lines.

Output. The contour image of the 3D Terracotta Warriors fragment.

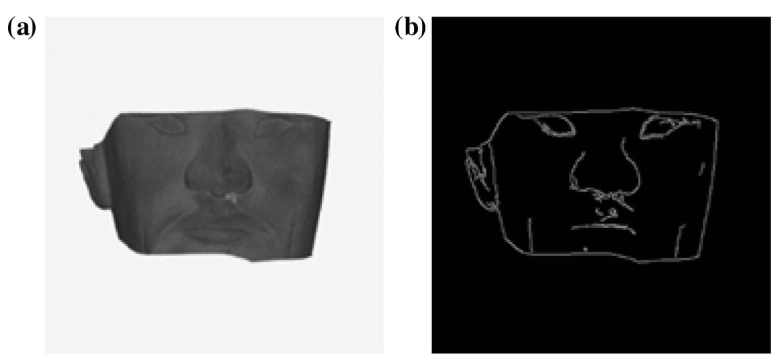

Fig. 4 Texture image generation. **a** Grayscale image of one piece of the head fragment; **b** Texture image of (a) based on the algorithm in this paper

The result of the texture extraction of one piece of the head fragment is shown in Fig. 4.

2.2 Neural network architecture

The network consumes two modalities of the 3D Terracotta Warriors fragment: Geospatial and textural information. Based on the two modalities, the network can extract the hidden features of each modality and output the category of the fragment. The basic structure of our network is shown in Fig. 5.

The subnetwork for extracting geospatial features from the point cloud is designed based on PointNet and depicted in the first part of Fig. 5. The most important component of the subnetwork is the transformation module, which is used to make the model invariant to input permutation. Details can be found in (Qi et al. 2017a, b). Another subnetwork is constructed to extract the textural features from the textural image generated in Sect. 2.1.2 and depicted in the second part of Fig. 5. This subnetwork is modified from Inception-V4, and two main components are the Stem block and Inception block, more details should be referenced to (Szegedy et al. 2017). After extracting the features from both two subnetworks, the features are then concatenated to generate the global features. The global features are then fed into an MLPs and a softmax layer, and the output is its category. The loss function used in this paper is the cross-entropy, while the Adam algorithm (Kingma and Ba 2014) (learning rate: 0.001, β_1: 0.90, β_2: 0.99) is adopted as the optimization function. It should be noted that during the training process, the batch size is set to 32 and the epochs of 300. The network was implemented using PyTorch and Python 3.7, and all computer operations were performed on a personal computer with an RTX 2080Ti GPU and a 3.40 GHz Intel Core i7 CPU.

3 Results and analysis

3.1 Databases of the terracotta warrior fragments

In this work, the 3D models of the Terracotta Warriors fragments were obtained by a 3D scanner, and a database which contained 3250 fragments of the Terracotta Warriors were established. In our database, the categories of fragments were labeled according to the parts of the Terracotta Warriors, which are arm, head, leg and body, as shown in Fig. 6. Furthermore, the database contains 800 pieces of the arm, 810 pieces of the body, 810 pieces of head and 830 pieces of the leg. The data are split into 80% for training, 10% for testing and 10% for validation. It should be noted that the texture, size and thickness of each fragment vary

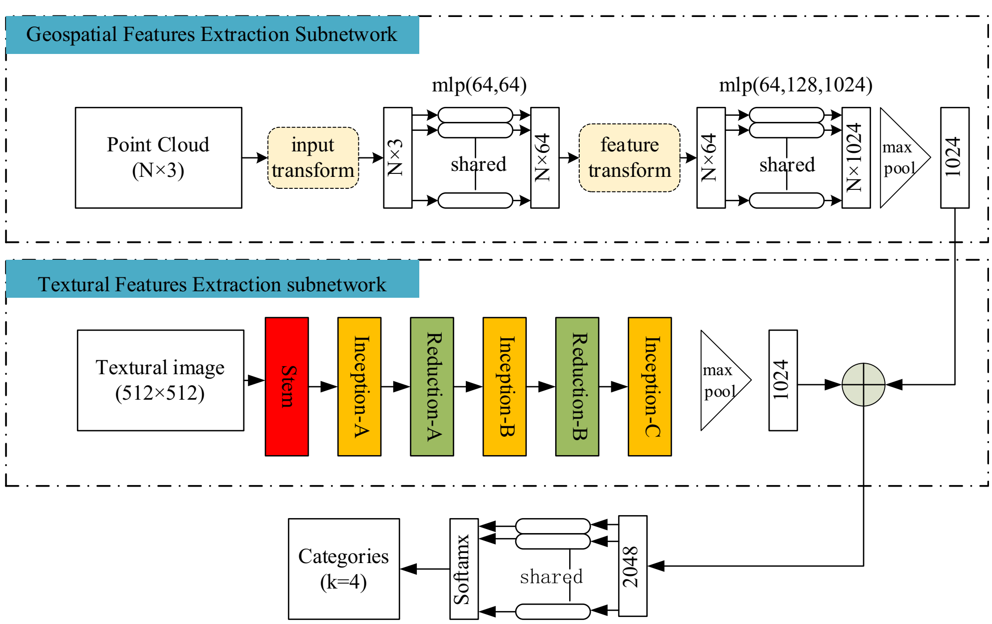

Fig. 5 Network architecture proposed in this paper

with different parts of the Terracotta Warriors. For example, the fragments of the head and body are pretty textured and characteristic. On the contrary, the pieces of legs have few noticeable texture features, and the side sections of the arm are curvy.

After the data generation procedure proposed in this paper, 13,000 fragment samples were generated in total, and the number of samples included in the different partitions is shown in Table 1.

3.2 Experimental results and comparison

Typically, a total of 10,400 fragments were assigned to the training set, while the 1300 fragments were allocated for the testing set, and 1300 fragments were used as a validation set. Implementation details are given in Sect. 2.2. Table 2 illustrates the results based on our framework.

To demonstrate that the usage of two modalities can improve the classification accuracy, we conducted two more experiments. The first one is based on the point cloud dataset, and the method used is the PointNet. While the second one is based on the textural images, and the method is the Inception-V4. Results are shown in Tables 3 and 4, respectively.

To further verify our framework can obtain a state-of-the-art classification accuracy for the Terracotta Warriors fragments, more methods are chosen to be the baseline. Results are shown in Table 5, and our framework can achieve the best mean accuracy.

Fig. 6 Selected fragments displayed by category

Table 1 The number of fragments included in particular categories

Category	Head	Body	Arm	Leg
Number	3104	3524	3232	3140

Table 2 Result of the classification based on our framework

Category	Head	Body	Arm	Leg
Accuracy(%)	94.37	95.32	87.55	88.41

Table 3 Result of the classification based on geospatial information only

Category	Head	Body	Arm	Leg
Accuracy(%)	92.36	96.45	82.51	84.41

Table 4 Result of the classification based on texture information only

Category	Head	Body	Arm	Leg
Accuracy(%)	91.50	92.75	77.75	76.25

Table 5 Comparison with the methods proposed in the references

Method	Input date type	Mean accuracy (%)
PointNet	Point cloud	88.93
Inception-V4	Image	84.56
Yang et al. (2017)	Image	71.27
Wang (2019)	Image	89.54
Liu (2019)	Volume	83.59
Gao and Geng (2019)	Point cloud	90.94
Our framework	Point cloud + textural image	**91.41**

4 Discussion and conclusion

Fragment classification is a key step in the restoration of the Terracotta Warriors; the accurate classification results can provide a great convenience for subsequent reassembly tasks. Recently, the traditional artificial approach is still widely applied to the fragment classification of the Terracotta Warriors. However, this kind of artificial classification approach is not only time-consuming and labor-intensive but also has the problem of low consistency of classification results. Therefore, many researchers have explored effective schemes to obtain a better performance in fragment classification tasks. Nevertheless, various limitations exist in these methods.

In this paper, we present a novel neural network-based framework, both geospatial and texture information of the fragments are considered. The main contributions of our work are: (1) the proposed method can directly consume point cloud and texture image of the fragment and outputs its category; (2) an improved canny edge detection algorithm is proposed to extract texture image of the fragments, compared with the original image, the edge-based image is more suitable for the classification; (3) the framework is flexible, which can incorporate more effective textural image extraction algorithm in the future. Experimental results demonstrate our framework performs better than previous methods, and the fusion model wins higher classification accuracy of the Terracotta Warriors fragments. Moreover, it is much more efficient than artificial approaches. The reason is that the proposed method uses more features for training the network than traditional single modal based methods. It is worth noting that for most parts of the Terracotta Warriors, the integration of two modalities can gain more accurate classification results, compared with the usage of the single modal. However, the result based on geospatial information only in body classification is better than that of our framework. One possible reason is that compared to the geospatial extraction subnetwork, the two-modality-based network cannot learn deep features well.

This article still has some limitations. First, the number of categories set is relatively small. Only four parts can be recognized by our framework, which does not meet the actual needs. Second, the fragments at the junction of different parts cannot be accurately classified. Third, the input size of the point cloud is limited to 2048, if the number of points of a fragment is much larger than 2048, it must be downsampled to a new sparse point cloud, which will lose the geometry information and not conducive to the geospatial extraction subnetwork to learn the local features. To address the above limitations is our future research direction.

As a summary, the framework proposed in this paper can classify the 3D Terracotta Warrior fragments accurately, and the results also demonstrate the effectiveness and practicability. We hope our work can provide a powerful tool for the restoration of the Terracotta Warriors.

Acknowledgements This work was supported in part by the National Natural Science Foundation of China under Grant 61701403, the Project funded by China Post-doctoral Science Foundation under Grant No. 2018M643719, the Young Talent Support Program of the Shaanxi Association for Science and Technology under Grant No.20190107, the National Key Research and Development Program of China under Grant 2017YFB1402103, the Scientific Research Program Funded by Shaanxi Provincial Education Department under Grant No. 18JK0767, and the Natural Science Research Plan Program in Shaanxi Province of China 2017JQ6006.

References

Gao H, Geng G (2019) Classification of 3D terracotta warrior fragments based on deep learning and template guidance. IEEE Access 8:4086

Griffiths D, Boehm J (2019) A review on deep learning techniques for 3D sensed data classification. Remote Sens 11(12):1499

Kampel M et al. (2001) Classification of archaeological fragments using profile primitives. In: Proceedings of the 25th workshop of the austrian association for pattern recognition, na

Kampel M, Sablatnig R (2000) Color classification of archaeological fragments. In: Proceedings 15th international conference on pattern recognition. ICPR-2000, IEEE

Kang X et al (2015) Classification of cultural relic fragments based on salient geometric features. J Graphics 36(4):551–556

Karasik A, Smilansky U (2011) Computerized morphological classification of ceramics. J Archaeol Sci 38(10):2644–2657

Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980

Li Y et al. (2018) Pointcnn: Convolution on x-transformed points. In: Advances in neural information processing systems

Liu X (2019) Research on residual network identification and multi-feature mosaic technology for cultural relics restoration School of Information Science and Technology Northwest University. Master Thesis

Li-Ying Q, Ke-Gang W (2010) Kernel fuzzy clustering based classification of ancient-ceramic fragments. In: 2010 2nd IEEE international conference on information management and engineering, IEEE

Maturana D, Scherer S (2015) Voxnet: a 3d convolutional neural network for real-time object recognition. In: 2015 IEEE/RSJ international conference on intelligent robots and systems (IROS), IEEE

Moenning C, Dodgson NA (2003) A new point cloud simplification algorithm. In: Proc. int. conf. on visualization, imaging and image processing

Qi CR et al. (2016) Volumetric and multi-view cnns for object classification on 3d data. In: Proceedings of the IEEE conference on computer vision and pattern recognition

Qi CR et al. (2017) Pointnet: deep learning on point sets for 3d classification and segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition

Qi CR et al. (2017) Pointnet ++: deep hierarchical feature learning on point sets in a metric space. In: Advances in neural information processing systems

Smith P et al. (2010) Classification of archaeological ceramic fragments using texture and color descriptors. In: 2010 IEEE computer society conference on computer vision and pattern recognition-workshops, IEEE

Su H et al. (2015) Multi-view convolutional neural networks for 3d shape recognition. In: Proceedings of the IEEE international conference on computer vision

Szegedy C et al. (2017) Inception-v4, inception-resnet and the impact of residual connections on learning. In: Thirty-first AAAI conference on artificial intelligence

Wang, Y. (2019). Research on the classification algorithm of terracotta warrior fragments based on the optimization model of convolutional neural network. School of Information Science and Technology, Northwest University. Master Thesis

Wang Q, Kim M-K (2019) Applications of 3D point cloud data in the construction industry: a fifteen-year review from 2004 to 2018. Adv Eng Inform 39:306–319

Wu Z et al. (2015) 3d shapenets: a deep representation for volumetric shapes. In: Proceedings of the IEEE conference on computer vision and pattern recognition

Xie Y et al. (2019) A review of point cloud semantic segmentation. arXiv preprint arXiv:1908.08854

Yang W et al. (2017) Classification of terra-cotta warriors fragments based on multi-feature and SVM. J Northwest Univ,

Zhao F, Geng G (2018) Fragments classification method of terracotta warriors based on region and shape features. J Geomat Sci Technol 35(6):584–588

Zhou P et al. (2011) Ancient porcelain shards classifications based on color features. In: 2011 Sixth international conference on image and graphics, IEEE