BASED ON IMPROVED POINTNET AUTOMATIC CLASSIFICATION METHOD OF GROTTO TEMPLE STATUES

Q. Fu[1,2], M. Hou[1,2], W. Hua[3]

[1] School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture
[2] Beijing Key Laboratory for Architectural Heritage Fine Reconstruction & Health Monitoring
[3] College of Geosciences and Surveying Engineering, China University of Mining and Technology

KEY WORDS: Grotto Temple,Point Cloud,Statue Classification, Improved PointNet,Automation

ABSTRACT:

As a carrier of history, culture, religion and art, grotto temples are an important part of China's splendid cultural heritage. Among them, there are many grotto statues with similar shapes and scattered on the grotto temples, which have high artistic value. It is of great significance to classify them.The existing classification methods of grotto temple statues are mainly based on traditional manual classification and machine learning classification, which often consumes a lot of labor costs and time costs. To solve these problems, this paper improves the PointNet network and applies it to the classification of statues in grottoes, which greatly improves the automation of the classification of statues. And the point cloud data set of the grotto temple is made for experiment. The results show that the overall accuracy of the method in this paper reaches 89.73%, the average intersection and combination ratio reaches 68.9%, and the accuracy is increased by 5.47% and 4.3% respectively compared with the random forest classification method. It is of great significance to the value research, status assessment and virtual restoration of the subsequent grotto temples.

1. INTRODUCTION

Grotto temples were originally a form of Buddhist architecture in India. Most of the temple buildings built on the cliff, including caves and statues, are Buddhist statues or murals and stone carvings of Buddhist stories. Yungang Grottoes are the classic works of Buddhist art and an important part of the immovable cultural heritage (Fang Guo & Guanghui Jiang,2015). However, as an immovable cultural relic, the Grotto Temple is more vulnerable to damage from various factors after thousands of years of exposure to the wind, sun and human activities. The traditional inspection and repair work is often carried out on the entity, which is easy to cause secondary damage to the grotto temple statues. With the development of 3D laser technology, this efficient, high-precision, non-contact method has been fully applied in the digital protection of grottoes. The massive laser ranging point cloud can record the three-dimensional coordinates, reflectivity, texture and other information of the object surface in the grotto temple. On this basis, a high-precision three-dimensional model of the grotto temple can be built, which can efficiently store, analyze, display and use the three-dimensional information of the grotto temple (Bit again,2016). If all kinds of statues can be classified in the 3D data of the grotto temple, it will have scientific reference significance for the evaluation of the status quo of subsequent statues, virtual restoration and even guiding the actual restoration. However, the number of statues in grotto temples is often large, the shape is similar, the arrangement is dense and irregular, and scattered throughout the grotto. At present, the existing classification methods of statues are mainly based on traditional manual classification and machine learning classification.

The manual classification method needs to consume a lot of labor costs and time costs; Machine learning classification methods also need to construct a large number of features manually. In order to solve such problems, this paper applies the improved PointNet network to the classification of statues in grotto temples, which greatly improves the automation of the classification of statues and saves a lot of labor and time costs.

2. RESEARCH METHODS

2.1 About deep learning

In 2006, Hinton first proposed the concept of deep learning. Deep learning can be said to be a part of machine learning. It is relative to shallow learning. What is depth, simply speaking, is to learn and extract object signs through continuous convolution, so as to recognize and classify objects with the same characteristics in a neural network way like human brain. Deep learning mainly includes supervised learning and unsupervised learning. Inspired by neuroscience, early deep learning is designed by imitating the working mode of human brain. We know that neurons are composed of three parts: cell body, dendrites and axons. As shown in Figure 2.1, dendrites, as the entrance of information transmission, are the bridge connecting one neuron and another neuron. It is this abstract feature extraction ability that brings inspiration to the development of deep learning, The resulting network structures include deep feedforward neural network, deep convolution neural network, deep stack self-coding network, sparse deep neural network, deep fusion network, deep generation network, deep complex convolution neural network and deep binary neural network, deep cycle and recursive neural network, and deep reinforcement learning (Charles Ruizhongtai Qi,Hao Su,Kaichun Mo & Leonidas J. Guibas,2016). In computer vision processing, the deep learning network has played its inherent advantages, especially in the field of object classification and segmentation, and has made great progress. Scholars from all walks of life have paid more and more attention to the deep learning network .

2.2 Improved PointNet

Point clouds have three characteristics: disorder, rotatability and interaction between point clouds. (1) Disorder. Point cloud data

can be regarded as a collection of points, which is essentially a series of points without fixed order. The change of the order of points does not affect the expression of their overall shape in space, as shown in Figure 1. Although the arrangement order of the five points in the point set has changed, they represent the same point cloud; (2) Rotability. The coordinates of point clouds will also change after rigid transformation (such as rotation or translation) in space; (3) Interaction. Point cloud individuals do not exist independently, but are associated with neighborhood points and contain some common characteristics.

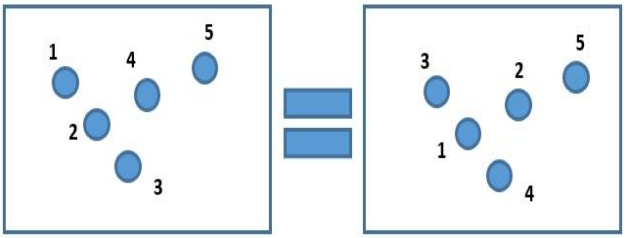

Figure 1. Schematic diagram of point cloud characteristics

Based on these characteristics of point clouds, in 2017, Charles R. Qi, a scholar at Stanford University, proposed the pioneering PointNet network that can directly process 3D point cloud data (Charles Ruizhongtai Qi,Hao Su,Kaichun Mo & Leonidas J. Guibas,2016). As the first deep neural network that directly input point cloud data into the cumulus, PointNet represents a new method for processing point cloud semantic segmentation. The network structure has two highlights, one is to solve the rotation invariance problem of point clouds, and the other is to solve the disorder problem of point clouds. On the problem of point cloud rotation invariance, the PointNet network uses a T-net matrix to ensure that the point cloud data set is a positive point cloud regardless of the angle of input, and to ensure its segmentation accuracy (Mohammadreza Tabatabaei,Roohollah Kimiaefar,Alireza Hajian & Alireza Akbari,2021); On the problem of point cloud disorder, PointNet uses maximum pooling to solve the problem. After a certain degree of feature extraction for each point, a symmetric function is used to integrate the previously extracted features to achieve point cloud processing independent of the input order (Jin Zhongxiao ,2019). In the PointNet network, the input point cloud data mainly contains spatial information, color information and label information, namely {X, Y, Z, R, G, B, L}. In the original PointNet network model, in view of the huge amount of point cloud data, coordinate normalization was also performed on the input data set. During the normalization process, the points in the data set were divided into 1 according to different object types × one × The cube of height, which is the height of the object, constitutes the data set input to the neural network.

PointNet network can be seen as composed of two parts, namely point cloud classification and point cloud segmentation. Point cloud segmentation is divided into local segmentation and whole scene semantic segmentation. Point cloud semantic segmentation is a large-scale semantic segmentation of the whole scene. Therefore, the point cloud segmentation part of PointNet network is mainly introduced. Its key modules are composed of alignment network T-net, multi-layer perceptron (MLP), and fusion layer of local and global features. In this paper, the improvement of this network structure is to use KNN algorithm to extract the local neighborhood of each point in the point cloud. The points in the neighborhood contain the local interconnection between the point clouds, and the local features obtained are more abundant. The network structure is shown in the figure2.

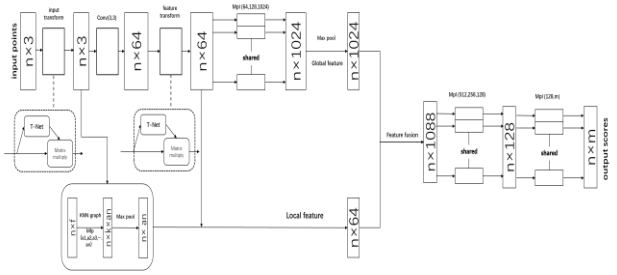

Figure 2. Improved PointNet network structure

3. RESEARCH DATA AND DATA SET CONSTRUCTION

3.1 Research object

The 18th Cave of Yungang Grottoes is one of the five caves with unique significance in the early period. It was built with the template of Tuoba Tao, the Emperor of Wei Taiwu. The overall momentum of the cave temple is magnificent. The ground inside the cave is oval, and the top is dome-shaped. The front of the cave looks like a horseshoe, which gradually shrinks from the bottom to the top, showing the embodiment of the royal weather in the Northern Wei Dynasty (Wang Yanqing ,2021).

The point cloud data used in this study comes from the overall cave replication project of the 18th Yungang Grottoes. In 2016, the research team used the ground three-dimensional laser scanner (TLS) to carry out multi-stop scanning of the 18th Cave of Yungang Grottoes through the construction of scaffolding, and the hand-held three-dimensional laser scanner was used to scan the details and areas with shelter. After point cloud pretreatment, other areas except the main image, the assistant Buddha and the disciple image were selected as the study area. Reasons for removal: the number of this part is small, training samples are small, and manual segmentation is faster and more accurate than other methods.

In order to facilitate the study of statue segmentation, the contents of the study area are divided into five categories according to the characteristics of the statue: large Buddha niches, small Buddha niches, small Buddha statues, other statues, and stone walls.

The small Buddha statue refers to the smaller Buddha statue, which is close to the size of the small Buddha niche, but not the Buddha statue in the niche, which is directly carved on the wall, as shown in the green box in Figure 5. Other statues refer to non-niche statues, which are mainly Buddha statues. There are also common Buddhist statues such as Feitian, Luxi and Buddha Treasure, as shown in the yellow box in Figure 5. The stone wall refers to the area in the study area where no statue is carved. It is generally the smooth stone wall in the cave temple, but also includes some areas with severe weathering.

3.2 Data set construction

To achieve the perfect operation of the network, it is necessary to make a high-quality dataset first, so that the subsequent segmentation work can become more efficient and accurate. During the production of the data set of the grotto temple, the

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-M-1-2023-87-2023 | © Author(s) 2023. CC BY 4.0 License.

Figure 3. Orthophoto of the north wall of Cave 18 of Yungang Grottoes

Figure 4. Orthophoto of the south wall of Cave 18 of Yungang Grottoes

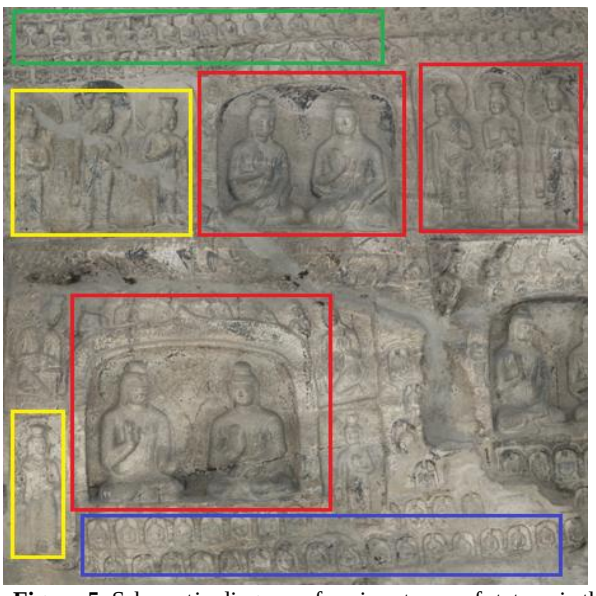

Figure 5. Schematic diagram of various types of statues in the study area

production of the point cloud data set of the grotto temple is mainly completed according to the production requirements of the S3DIS large data set. In the production process of S3DIS large data set, six indoor office areas are divided into 271 rooms. The Matterport camera is used in combination with three structured light sensors with different spacing. After scanning, the reconstructed 3D texture grid, RGB - D image and other data are generated, and the point cloud is produced by sampling the grid. A semantic tag is added to each point in the point cloud, such as 13 objects in total, such as chair, table, floor, wall, etc. The point cloud data used for training will be divided into point sets according to the room, and the point cloud data of the room will be divided into 1m × 1m blocks, and then predict the semantic label of each point for each block. According to the S3DIS large-scale data set production standard, the point cloud data of the grotto temple is divided into six different regions, as shown in the figure, and then the data set is produced for each region. The production process mainly includes point cloud labeling, point cloud fragmentation (blocking and downsampling), and coordinate normalization.

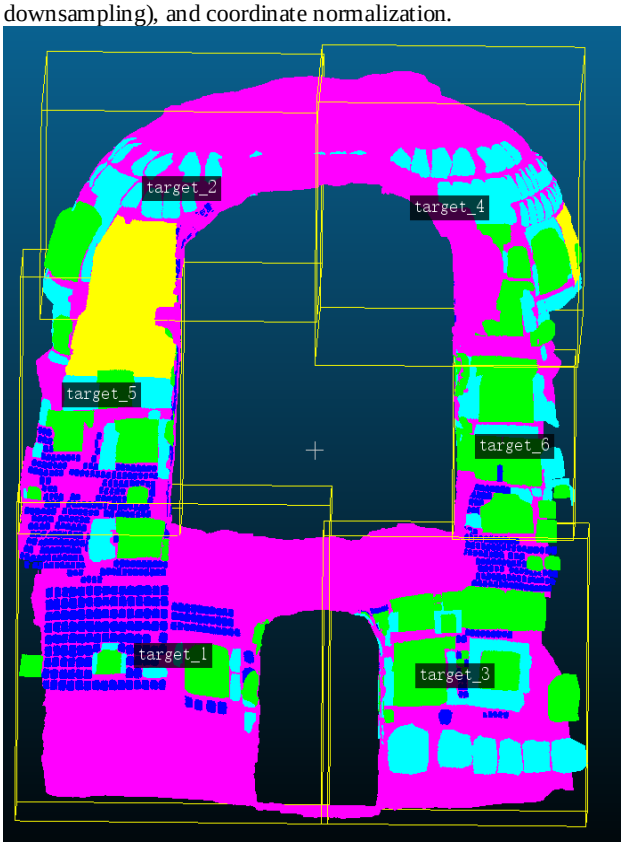

Figure 6. Six regions of grotto data

3.2.1 Point cloud annotation: The primary task of data set production is to label the pre-processed point cloud data, and each point will form corresponding label data after labeling. In deep learning, the most complicated and less technical thing should be to label the collected data. To do good work, we must use the tools first. A good labeling method is bound to affect the final classification results. The label labeling tool used in the experiment is Cloud Compare software. Compared with other point cloud labeling tools, Cloud Compare has the greatest advantage that it can retain color information for the labeled point cloud and save it as ASCII text files, which is convenient for later data conversion.

3.2.2 Point cloud fragmentation: To ensure uniform data distribution, reduce memory consumption and prevent over-fitting. In this study, according to the distribution of the grotto statues, the grotto is divided into blocks as small as possible and containing more semantic information. The number of point clouds after the block is different. In order to normalize the collected sample data, it is also necessary to reduce the sampling processing of the data. Point cloud down-sampling generally includes random sampling, uniform sampling and farthest point sampling. This paper uses the farthest point sampling method to sample the point cloud of each block of the grotto temple through batch processing. Each sample is 4096 points, and each point is represented by a 7-dimensional vector. The vector information is divided into X, Y, Z, R, G, B, and L

3.2.3 Coordinate normalization: I The coordinate value of the original point cloud data is often very large. Direct input into the network for training will cause slow operation and even collapse. Therefore, it is necessary to use the normalization method to uniformly transform the coordinate value, and limit the size of the converted coordinate value to - 1~1.

4. EXPERIMENTAL ANALYSIS

4.1 Experimental environment configuration

Before the training of deep learning network, the experimental environment should be configured first, mainly including hardware and software. Under the same network, the higher the hardware equipment is, the faster the training speed is theoretically, and the current convolutional neural network generally needs GPU to accelerate. The configuration table of the experimental environment used in this experiment is shown in Table 1. The main environment component for programming is PyCharm.

Project	Type
CPU	Intel（R）i7-11800H
GPU	NVIDIA GeForce RTX 3060
RAM	16GB
PyTorch	1.11.0+cu113
Python	3.9
system	Windows 11

Table 1. Experimental environment

4.2 Evaluation indicators

At present, the commonly used evaluation indicators are the overall accuracy (OA), classification intersection ratio (IoU), and average category intersection ratio (mIoU). OA represents the proportion of correct output points to total points; The classification intersection and union ratio represents the ratio of intersection and union between the result area and the real calibration value area; MIoU represents the average of all classes. The formulas for the three precision indicators are as follows: 1, 2, 3.

$$oA = \frac{\sum_{i=0}^{k} p_{ii}}{\sum_{i=0}^{k} \sum_{j=0}^{k} p_{ij}} \quad (1)$$

$$IoU = \frac{TP}{TP + FP + FN} \quad (2)$$

$$mIoU = \sum_{i=0}^{k} IoU_i \quad (3)$$

There are (k+1) points in total, and Pij means that it belongs to category i and is divided into category j; In the same way, Pji indicates that it belongs to category j and is divided into category i; Pii indicates that it is divided into correct categories. TP stands for positive class and is determined as positive class, that is, correct classification; FP represents negative class and is judged as positive class, that is, false positive; FN stands for positive category and is judged as negative category, that is, false negative. The purpose of our experiment can be vividly expressed. TP said that the correct result was a small niche, and it was indeed found to be a small niche; FP indicates that the correct result is a small Buddhist niche, but it is not a small Buddhist niche; FN indicates that the correct result is not a small niche, but a small niche.

Before network training, parameters need to be set. The initial learning rate is set to 0.001, the weight attenuation coefficient is set to 0.0005, the optimizer momentum momentum is set to 0.9, the threshold (sigma) is set to 0.7, and the number of iterations is 251. The activation function uses ReLU activation function . In this chapter, the input quantity of the grotto temple dataset is different from that of the PointNet network. The input information of each sample point cloud is set to {X, Y, Z, R, G, B, L} 7-dimensional features, and the number of input point clouds is 4096. Then, cross validation method was used to test and validate the data from six areas of the grotto temple. Taking this paper as an example, cross validation uses five target as the training set each time, and the other one is the test set. All validation sets use target 5.

4.3 Experimental precision analysis

In this paper, the improved PointNet network is used to experiment with the point cloud data of the Eighteen Grottoes of Yungang, and the visual analysis of the local results is aimed at intuitively viewing the classification effect. Five colors are used to represent five categories, red represents the stone wall, yellow represents the small Buddha, blue represents the small Buddha niche, green represents the large Buddha niche, and light blue represents other statues. As shown in Figure 10, the PointNet network can roughly divide all kinds of statues, but the details, such as the edge of each statue, are still relatively rough, but the whole is relatively regular, and the connectivity is also strong, which shows that the in-depth learning method is feasible to apply to the classification of statues in grottoes. In order to show the advantages of this method, this paper also carried out a comparative experiment, using a random forest classification experiment. Use commonly used features to build feature vector sets, including 13 features, such as roughness, mean curvature, Gaussian curvature, surface density, volume density, linearity, planarity, divergence, total variance, feature entropy, local curvature change, anisotropy, and verticality. After that, feature screening is carried out according to the same process. After screening, training sets and test sets are made and sent to random forest learning training. The comparison of experimental accuracy is shown in Table 2.

Figure 7. Partial photos

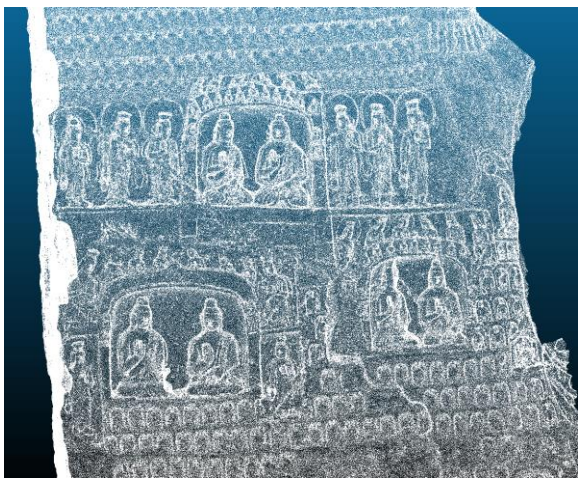

Figure 8. Point cloud data

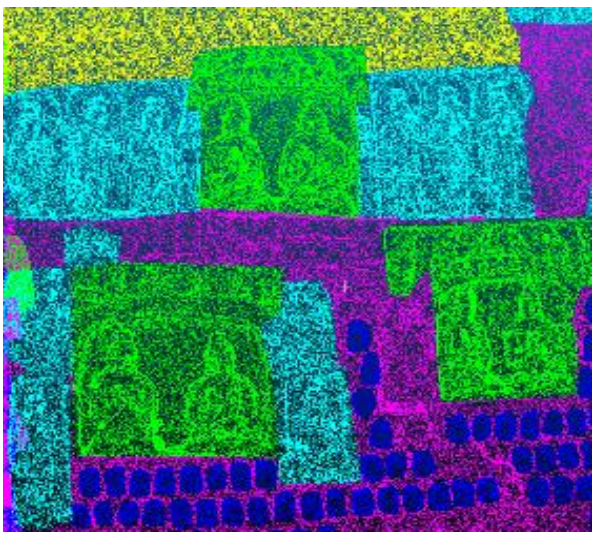

Figure 9. Manual segmentation results

Figure 10. PointNet segmentation results

method	OA	mIoU
Improved PointNet	89.73	68.9
PointNet	83.86	61.5
Random forest	84.26	64.6

Table 2. Comparison of classification accuracy

5. CONSTRUCTION

Aiming at the low degree of automation in the classification of grotto temple statues at present, this paper improves PointNet and applies the network to the classification of grotto temple statues. The self-created point cloud data set of grotto temple is used for model training. Through experiments, it is known that the overall accuracy of this method reaches 89.73%, the average intersection and combination ratio reaches 68.9%, and the accuracy is increased by 5.47% and 4.3% respectively compared with the random forest classification method, And it has greatly improved the degree of automation, which is of great significance for the subsequent value research, status assessment and virtual restoration of grotto temples.

REFERENCES

Fang Guo & Guanghui Jiang.(2015).Investigation into rock moisture and salinity regimes: implications of sandstone weathering in Yungang Grottoes, China. Carbonates and Evaporites(1). doi:10.1007/s13146-014-0191-8.

Bit again (2016). Research on the data management method of the digital project of cultural relics in the grotto temple (master's thesis, Beijing University of Architecture and Architecture)

Zhou Junzhao, Zheng Shumin, Hu Song&Zhou Jianbo (2008). Application of ground 3D laser scanning in the protection and mapping of grottoes and stone carvings Surveying and Mapping Bulletin (12), 68-69

Gao Jinhong (2018). Environmental toxicological detection technology based on high-throughput method (master's thesis, Shaanxi Normal University)

Huang Wenyi (2019). Rolling bearing fault diagnosis and performance degradation evaluation based on feature optimization and self-learning (doctoral dissertation, Hunan University)

Charles Ruizhongtai Qi,Hao Su,Kaichun Mo & Leonidas J. Guibas.(2016).PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation.. CoRR.

Jia Hongfei (2017). Research on classification and counting of white blood cells based on deep learning (master's thesis, Shenzhen University)

Mohammadreza Tabatabaei,Roohollah Kimiaefar,Alireza Hajian & Alireza Akbari.(2021).Robust outlier detection in geo-spatial data based on LOLIMOT and KNN search. Earth Science Informatics(prepublish). doi:10.1007/S12145-021-00610-9.

Jin Zhongxiao (2019). Object recognition and attitude estimation method based on 3D multi-view (master's thesis, China University of Science and Technology)

Zhang Jing (2019). Research on hole repair of cultural relic fragments and splicing method based on fracture surface (doctoral dissertation, Northwestern University).

Wang Yanqing (2021). Discussion on the types of early round arched niches in Yungang Grottoes - the second part of the discussion on the early round arched niches in Tanyao Five Grottoes Yungang Research (02), 24-39. doi: 10.19970/j.cnki.issn2096-9708.2021.02.003

Li Yi (2019). Talking about the construction of digital cultural relics in grottoes -- taking Longmen Grottoes as an example Cultural Relics Identification and Appreciation (16), 101

Li Min, Diao Changyu, Ge Yunfei, Qiu Linshan&Li Li (2021). Digital protection and utilization of cultural relics of grottoes Journal of Remote Sensing (12), 2351-2364

Li Xinglong (2021). The application of three-dimensional digital technology in Longmen Grottoes archaeology Luoyang Archaeology (01), 88-95

Lu Wenxiang, Xiong Ruiping, Xu Yisong, Yang Kang and Li Hua (2022). Point cloud registration based on feature segmentation recognition Modular machine tools and automatic processing technology (04), 32-35 doi:10.13462/j.cnki.mmtamt.2022.04.008.

Jia Shu (2018). The application and challenges of 3D laser scanning technology in the protection and restoration of contemporary cultural relics Research on Cultural Relics Restoration (00), 739-743

Dang Yuechen, Li Wan&Zhou Qiang (2022). Fragment reorganization based on multi-feature information fusion and evolutionary computation Journal of Shaanxi University of Science and Technology (02), 195-200+2006 doi:10.19481/j.cnki.issn2096-398x.2022.02.025.

He Xiaomei, Zhang Lianjun, Chen Fen, Cai Zhenzhen and Wang Xiaodong (2022). Four-reference view fusion algorithm guided by depth and structure similarity Journal of Ningbo University (Science and Engineering Edition) (02), 96-104

Li Xinglong (2021). The application of three-dimensional digital technology in Longmen Grottoes archaeology Luoyang Archaeology (01), 88-95

Jiang Yifeng, Hu Linna&Liu Ranran (2021). Research on saliency region detection method based on similarity fusion Journal of Jiangsu Institute of Technology (06), 50-58 doi:10.19831/j.cnki.2095-7394.2021.06.015.