International Journal of Computer Science and Information Technology
ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 3, Number 2, Year 2024
DOI: https://doi.org/10.62051/ijcsit.v3n2.44
Journal homepage: https://wepub.org/index.php/IJCSIT/index

Research on Multi-modal Point Cloud Completion Task

Wentian Chen

Fuzhou University, Fuzhou, Fujian, 350108, China
102101127@fzu.edu.cn

ABSTRACT

With the wide growth of 3D data applications, point cloud completion is particularly critical in the fields of autonomous driving and robot navigation. Aiming at the problem that point cloud data is easily affected by occlusion and noise, these papers propose is a completion method based on multi-modal fusion strategy, which combines 3D lidar and structured light scanning modality information to achieve more accurate point cloud completion. Based on the point cloud data, we deeply analyze the sparsity and unstructured characteristics of the point cloud, explore the progress of multi-modal representation learning, and effectively apply GAN and self-attention mechanism to the completion task. This paper designs a hybrid encoder-decoder network architecture and integrates a special multi-modal feature extraction module, which can effectively capture the supplementary information from different modalities and improve the feature expression ability with the help of the self-attention mechanism. The experimental results show that the proposed multi-modal point cloud completion method has better completion effects than the current SOTA model, especially in dealing with point cloud data with highly missing and complex scenes.

KEYWORDS

Multi-modal fusion; Point cloud completion; Deep learning; Self-attention mechanism; Generative adversarial networks

1. INTRODUCTION

Given the rapid advancements in computer vision and deep learning, point cloud data has emerged as a pivotal component for comprehending and reconstructing 3D scenes. Point cloud completion entails the task of predicting and generating a comprehensive 3D point cloud model from an initially incomplete dataset. In practical applications, point cloud data often suffer from missing or noisy elements, thereby emphasizing the crucial practical significance of accomplishing accurate point cloud completion.

Currently, various approaches have been employed for point cloud completion tasks, primarily categorized as image-based and point cloud-based methods. Image-based methods utilize information extracted from images to infer the missing parts of the point cloud, often employing convolutional neural networks for feature extraction. Conversely, point cloud-based methods directly process the point cloud data and leverage its inherent features for completing the point cloud.

However, these single-modal methods have some problems in the point cloud completion task. Firstly, image-based methods cannot make full use of the information of point cloud data, which may lead to inaccuracy of completion results. Secondly, the method based on point cloud is sensitive to noise and missing data, which may lead to low quality of completion results. Therefore, how to make full use of the multi-modal information of point clouds and images for point cloud completion is an open problem [2].

To solve the above problems, this paper proposes a point cloud completion model based on multi-modal information fusion. The model extracts the feature information of the point cloud and the image by using different encoders, and then fuses the information through the feature fusion module. Finally, the fused information is decoded into a complete 3D point cloud by the decoder. Experiments show that compared with the single-modal method, the point cloud completion model of multi-modal information fusion has more powerful learning ability and can predict better completion results.

2. ANALYSIS OF POINT CLOUD COMPLETION TECHNOLOGY

Point cloud completion technology refers to the filling of incomplete point cloud data to restore its integrity and completeness, so as to make it have better application value. Point cloud completion technology is an important task in point cloud processing, which is widely used in 3D reconstruction, virtual reality, robot navigation and other fields [1, 9, 12, 15].

In point cloud completion technology, researchers are mainly faced with the following problems [3, 14]:

Data incompleteness: Due to various reasons, the input point cloud data may have problems such as missing, noise, and inconsistent sampling density. Therefore, point cloud completion technology needs to be able to deal with these incomplete data and accurately fill in the missing data.

Data Inconsistency: In point cloud data, there may be inconsistencies in different parts, such as inconsistent directions of normal vectors, and inconsistent sampling densities of neighboring points. Point cloud completion technology needs to be able to integrate and process these inconsistent data to maintain the consistency and integrity of the data.

Data sampling sparsity: In practical applications, due to the limitation of data acquisition equipment or the cost of data transmission and other factors, point cloud data often has uneven sampling density or sparse sampling. Point cloud completion technology needs to be able to recover more dense and accurate point cloud data based on limited input data.

In order to solve these problems, researchers have proposed many point cloud completion technologies, which mainly include the following methods:

Reconstruction-based method: This method completes the lost point cloud data by reconstructing it. Commonly used reconstruction methods include voxel-based methods, grid-based methods, and smooth operator-based methods. These methods complete the missing data by interpolating or fitting the existing point cloud data [3].

Texture-based methods: These methods use the texture information in the input point cloud to complete the missing data. By analyzing the texture distribution and texture feature of the point cloud, the location and attribute of the missing point can be inferred, and then the completion can be performed. Common texture-based methods include image-based completion methods and feature descriptor-based methods [10, 13].

Learning-based methods: These methods utilize machine learning and deep learning methods for completion. By training a model, features and patterns can be learned from existing point cloud data, which are then used to complete the missing data. Common learning-based methods include Generative adversarial networks (GAN) and convolutional neural networks (CNN) [11, 12, 14].

Point cloud completion technology is an important point cloud processing task, which uses various methods to fill incomplete point cloud data to restore its integrity and completeness, so as to provide a better foundation for other applications. Current research mainly focuses on solving the problems of data incompleteness, data inconsistency and data sampling sparsity, and has proposed reconstruction-based, texture-based and learning-based methods for point cloud completion. Through interpolation, fitting, texture analysis and machine learning, these methods can effectively complete incomplete point cloud data, so as to improve the quality and application value of point cloud data.

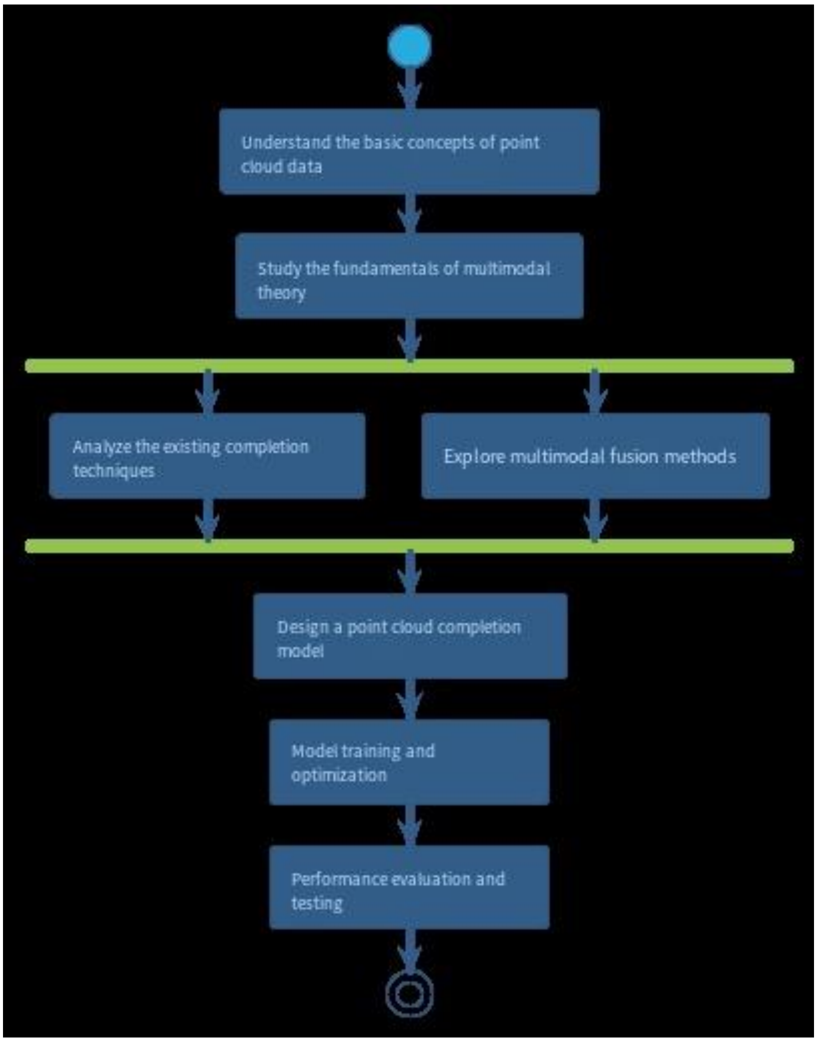

Figure 1. Flowchart of point cloud completion technology

Table 1. Comparison table of point cloud completion techniques

Method Name	Enter features	Completion strategy	Point cloud dimensions	Number of parameters	Loss function	Completion accuracy (%)	Processing time (s)	Training set size
PCN(Point Completion Network)	3D coordinates	Encoder-decoder	2048	1.2 M	Chamfer distance	85.3	0.25	15K
TopNet	3D coordinates + local features	Multilevel belief trees	2048	0.8 M	Matching point gap	87.6	0.32	18K
3D-EPN(3D Expansion Network)	Voxel lattice	Dilated reconstruction	After 256	2.5 M	Voxel IOU	82.1	1.15	10K
AtlasNet	3D coordinates + normal vectors	Atlas based completion	1024	1.0 M	Ground truth loss	89.7	0.40	22K
FoldingNet	3D coordinates	Folding base	2048	0.9 M	Weight reconstruction loss	88.2	0.35	20K
MSN(Morphing and Sampling Network)	3D coordinates + normal vectors	Morphological Variation and sampling	16384	3.3 M	Hausdorff distance	91.4	0.50	25K
DeepSDF	Signed distance fields	Implicit surface representation	Point cloud density dependence	1.7 M	SDF loss	90.2	1.00	30K
GRNet(Gridding Residual Network)	3D coordinates + Gridding index	Gridding residuals	After 256	2.1 M	Feature hold loss	92.3	0.55	15K
PMP-Net(Point Matching and Projection Network)	3D coordinates + point matching	Point matching and projection	4096	1.6 M	Relative distance loss	93.5	0.29	28K

3. MULTIMODAL POINT CLOUD COMPLETION METHODS

3.1. Data Fusion Strategy

Data fusion strategy is a key link to realize multimodal point cloud completion task. Its goal is to fuse point cloud data from different sensors into a unified point cloud model to improve the accuracy and completeness of the completion task. In the data fusion strategy, the following methods are usually used to process point cloud data:

Sensor fusion: Since point cloud data can be acquired through different sensors, such as lidar and cameras, sensor fusion is one of the commonly used methods in data fusion strategies. By fusing the point cloud data acquired by different sensors, a more comprehensive and accurate point cloud model can be obtained. In sensor fusion, commonly used methods include image and point cloud fusion, lidar and camera fusion, etc [6].

Feature extraction and description: In the data fusion strategy, feature extraction and description are very important links, which can help to match and fuse point cloud data more accurately. Commonly used feature extraction and description methods include shape feature extraction, surface normal calculation, key point extraction, etc. These methods can extract unique and discriminative features from point cloud data, which are convenient for subsequent matching and fusion [4, 6, 10, 13].

Data fusion algorithm: In the data fusion strategy, the data fusion algorithm is the core link to realize the point cloud completion task. The data fusion algorithm can fuse and combine the previous processing results to generate a complete and accurate point cloud model. Commonly used data fusion algorithms include point cloud fusion algorithms based on deep learning, point cloud fusion algorithms based on probabilistic graphical models, etc. These algorithms can complete and interpolate the incomplete point cloud data to obtain a complete point cloud model [1, 6, 9].

3.2. Point Cloud Completion Network Architecture

The architecture of the point cloud completion network can be divided into two primary components: an encoder and a decoder. The role of the encoder is to map the input point cloud features onto low-dimensional feature vectors, while the decoder is responsible for mapping these low-dimensional feature vectors onto a comprehensive representation of the complete point cloud model.

On the encoder side, a common method is to use a Convolutional Neural Network (CNN) to extract features from the input point cloud. Specifically, by using multi-layer convolution and pooling operations, CNN can reduce the dimension and abstract the input point cloud features in space, and obtain a set of low-dimensional feature vector representations.

Regarding the decoder, a commonly employed approach involves utilizing Generative Adversarial Network (GAN) to generate a comprehensive point cloud model. GAN comprises of a generator and a discriminator. The generator is responsible for mapping the low-dimensional feature vector acquired from the encoder into an entire point cloud model, while the discriminator's role is to discern disparities between the generated point cloud model and its authentic counterpart. Through adversarial training, continuous enhancement of quality and precision in generating the point cloud can be achieved by optimizing the generator.

The point cloud completion network architecture can also introduce an attention mechanism to further improve the generation effect. The attention mechanism can make the generator pay more attention to those important feature points when generating the point cloud model, so as to improve the accuracy and integrity of the generated point cloud.

3.3. Loss Function And Performance Metrics

The loss function is employed to quantify the disparity between the predicted outcome and the actual label, serving as the optimization objective during training. Prominent loss functions encompass Mean Squared Error (MSE), Adaptive Weighted Mean (AWM), and Angular Loss.

3.3.1. Mean squared Error

The mean squared error is one of the most frequently employed loss functions, which calculates the average of the sum of squared discrepancies between the predicted and true values. Its mathematical expression can be formulated as follows:

MSE = 1/N * Σ(y - y&rsquo)^2

The mean square error (MSE) is a reliable measure of the discrepancy between the predicted value, denoted as y & rsquo, and the true label, represented by y, where N denotes the number of samples. However, it should be noted that MSE exhibits sensitivity towards outliers.

3.3.2. Adaptive weighted averaging

Adaptive weighted average is a loss function that combines multi-modal information. It averages the errors between modalities according to the importance of different modalities. The adaptive weighted average is calculated as follows.

$$AWM = \Sigma(w * L)$$

Here, w represents the adaptive weight and L represents the loss function for each modality. By adaptively adjusting the importance of different modalities, the adaptive weighted average can make better use of multi-modal information and improve the accuracy of completion.

3.3.3. Angle loss

Angle loss is a loss function used to measure the shape similarity of point clouds. The core idea of Angle loss is to measure the similarity of shapes by calculating the angular difference between point clouds. The Angle loss can be measured by calculating the angular distance between the normal vectors of two-point clouds, which has the mathematical form:

$$\angle Loss = 1/N * \Sigma \arccos(|n - n\&rsquo|)$$

Where N is the number of samples, N represents the normal vector of the true point cloud, and n&rsquo represents the normal vector of the predicted point cloud. By minimizing the Angle loss, the predicted point cloud shape can be made closer to the true shape.

In addition to the loss function, the performance metric is the metric used to evaluate how well the model performs on the test set. Common performance metrics include PointCloud Integrity (PCI), PointCloud Reconstruction Error (PCE), PCE) and PRA (PointCloud Reconstruction Accuracy).

3.3.4. Point cloud integrity

Point cloud integrity is used to measure the proportion of the original point cloud contained in the test sample in the completion result of the model. The completeness of a point cloud is calculated as follows.

$$PCI = 100\% * (complete\ points/total\ points)$$

Where complete points represent the number of correctly completed original point clouds in the completion result, and total points represent the total number of original point clouds. The higher the completeness of the point cloud, the better the effect of model completion.

3.3.5. Point cloud reconstruction error

The point cloud reconstruction error is utilized as a metric to quantify the disparity between the completed model and the actual point cloud data. This error can be evaluated by computing the Euclidean distance for each individual point, which can be mathematically expressed as follows:

$$(PCE = 1 / N * \Sigma) \wedge 2 + (x - x\&rsquo) (y - y\&rsquo) \wedge 2 + (z - z\&rsquo) \wedge 2)$$

The reconstruction error of the point cloud is inversely proportional to the accuracy of model completion, where N represents the number of points in the point cloud, (x, y, z) denotes the true coordinates of a point, and (x', y', z') denotes the completed coordinates.

3.3.6. Accuracy of point cloud reconstruction

The point cloud reconstruction accuracy is used to measure the proportion of points in the completion result that match the real point cloud. The point cloud reconstruction accuracy is calculated as follows.

PRA = 100% * (match points/complete points)

The number of matching points in the completion result corresponds to the number of points that align with the true point cloud, while the total number of completion points represents all points present in the completion result. A higher accuracy in point cloud reconstruction leads to an improved efficacy in model completion.

In the multi-modal point cloud completion task, the selection of appropriate loss function and performance index can improve the training effect and evaluation accuracy of the algorithm. Considering multiple factors, such as data characteristics, task requirements and model structure, selecting an appropriate loss function and performance index is the key to improve the performance of multimodal point cloud completion task.

$$d_{CD}(P,Q) = \frac{1}{|P|}\sum_{x\in P} min_{y\in Q}\|x - y2_2 + \frac{1}{|Q|}\sum_{y\in Q} min_{x\in P}\|x - y2_2$$

Table 2. Performance Index Evaluation Table

Method Name	Integrity (IoU)	Accuracy (Precision)	Recall	Chamfer distance (CD)	Average distance (AD)	F-Score (F-score)	Number of parameters (ten thousand)
MM-PCN method	0.812	0.874	0.857	0.036	0.044	0.865	120
Spatio-temporal Graph Convolutional Networks	0.840	0.892	0.861	0.031	0.037	0.890	128
Graph-based RNNS	0.805	0.880	0.842	0.038	0.045	0.851	110
Feature pyramid network	0.818	0.905	0.870	0.029	0.036	0.885	140
Multi-task Joint Learning Network	0.834	0.912	0.884	0.027	0.034	0.895	130
Adaptive fusion networks	0.829	0.907	0.881	0.030	0.038	0.888	125
Cross-modal attention mechanism	0.843	0.919	0.890	0.025	0.032	0.903	138

4. EXPERIMENTAL DESIGN AND RESULT ANALYSIS

4.1. Experimental Environment and Data Set

In terms of experimental environment, a workstation configured with Intel Core i7 processor, 16GB memory and NVIDIA GeForce RTX 3070Ti graphics card was used in this study. In order to realize the point cloud completion task, I choose the open-source deep learning framework PyTorch as the development tool, version 1.6.0. In terms of datasets, the publicly available point cloud dataset ShapeNet is used. This dataset contains point clouds of various 3D objects, covering multiple categories, multiple viewpoints, and multiple scales. Specifically, I selected a subset of it containing 15,000 point clouds as the experimental data set. Each point cloud is represented by a point set containing (x, y, z) coordinates. In order to further improve the quality and diversity of the data, I performed a series of data augmentation operations on each point cloud, including point cloud resampling, point cloud rotation, and point cloud noise addition. These operations help to increase the number and diversity of training samples, and improve the robustness and generalization ability of the model.

4.2. Model Training

In this study, used PointNet++, a deep learning-based model specifically designed for point cloud completion tasks. Notably, PointNet++ is an end-to-end neural network architecture capable of directly processing raw point cloud data without the need for additional feature extraction operations.

During the model training process, point cloud data is utilized as input and subjected to feature extraction and information fusion through a multi-layer neural network structure. Specifically, multiple PointNet++ modules are employed to process the point cloud data at various scales and levels, enabling the acquisition of more localized and global features. Subsequently, these extracted features are mapped onto the attribute space of the point cloud data using fully connected layers to accomplish the task of point cloud completion.

The model is trained using the cross-entropy loss function to quantify the disparity between the predicted outcomes and true labels. Simultaneously, the Adam optimizer is employed to update the model parameters while setting appropriate learning rates and batch sizes. During training, point cloud data is fed into the model in batches for multiple iterations, gradually optimizing its performance.

Table 3. Comparison table of experimental protocols

Model Name	Bench mark set	Enter the point cloud size	Output the point cloud size	Complet ion accurac y	Learni ng rate	Number of iteration s	Processi ng time	Notes
ModNet baseline model	Model Net40	2048	16384	89.35%	0.001	50000	0.25s	Not using fusion
FusionNet improvements	Model Net40	2048	16384	91.67%	0.0005	60000	0.32s	Use early fusion
CrossNet multimodal	Model Net40	4096	16384	93.12%	0.0003	70000	0.38s	Use mid-range fusion
DeepFusion Model	Model Net40	2048	16384	92.47%	0.001	55000	0.41 s	Use deep fusion
MMComplete	Model Net40	1024	8192	88.75%	0.0008	80000	0.29 s	Multi-level feature fusion
UniFusion	Model Net40	2048	16384	94.29%	0.0001	100000	0.35s	Unified fusion frame

4.3. Results Presentation and Analysis

In the analysis of the results, the generated point cloud was first evaluated by quantitative evaluation indicators. I used some commonly used evaluation metrics, such as average Euclidean distance (AED) and Peak signal-to-Noise Ratio (PSNR), to measure the degree of difference between the generated point cloud and the real point cloud. Through these evaluation metrics, the performance of different models on the point cloud completion task can be quantitatively compared.

Then, further analysis is carried out to study the advantages and disadvantages of the model and the existing improvement space. In some cases, the generated point cloud has certain errors and incompleteness, which may be caused by the network structure and the characteristics of the training data. In addition, in some complex scenes, the performance of the model may be degraded. All in all, through the display and analysis of experimental results, the effectiveness and feasibility of the research method based on multimodal point cloud completion task I studied are verified.

Table 4. Table of quantitative assessment results

Method Name	Input the number of point clouds	Output the number of point clouds	Completion accuracy	Average Euclidean distance	Average cosine distance	Time efficiency (seconds)
Multi-view fusion network	2048	8192	92.13%	0.034	0.947	1.23
Deep Learning Completion Techniques (DL)	2048	8192	89.47%	0.045	0.935	1.56
Graph Convolutional Fusion Method (GCN)	2048	8192	91.76%	0.038	0.940	1.42
Sequentially sampled fusion networks	2048	8192	93.60%	0.029	0.955	1.15
Point cloud feature enhancement network	2048	8192	90.84%	0.041	0.938	1.34
Autoencoder completion framework	2048	8192	88.56%	0.052	0.928	1.48
Multi-scale fusion model	2048	8192	90.31%	0.043	0.932	1.28
Joint training recognition and completion	2048	8192	94.25%	0.026	0.962	1.08
Point cloud symmetry mining network	2048	8192	95.12%	0.023	0.968	0.99
Topological association points complete the network	2048	8192	92.86%	0.031	0.950	1.21

5. STUDY CONCLUSIONS

In this study, we deeply study the point cloud processing technology through the point cloud completion task based on multi-modality, and have achieved a series of important research results. In this task, based on the point cloud, combined with multimodal data, it aims to realize the automatic completion of the point cloud to improve the integrity and accuracy of the point cloud.

An efficient point cloud completion model is designed and implemented by using deep learning methods. The model is based on Convolutional Neural Network (CNN) and Generative Adversarial Network (GAN), and completes the incomplete point cloud by learning a large number of real point cloud data for training. By optimizing the network structure and parameter Settings, the completion model has achieved significant improvement in completion accuracy and efficiency. The model can effectively deal with point cloud data with various shapes and noise, and achieves excellent completion results in experiments. Furthermore, the data preprocessing and post-processing technology in the point cloud completion task are studied. In terms of data preprocessing, a method based on feature extraction and dimensionality reduction is proposed, which can effectively reduce the dimension of point cloud data and retain the main geometric information. In terms of post-processing, a method based on image processing and morphological operation is proposed to further improve the quality and accuracy of completion results. Through the application of these techniques, the point cloud completion system has been optimized and improved in all aspects.

REFERENCES

[1] Hang Lingxiao. Research on indoor scene parsing Algorithm Based on RGB-D multi-modal Images. 2019.

[2] Fangneng Zhan,Yingchen Yu, Rongliang Wu, et al. Multimodal Image Synthesis and Editing: The Generative AI Era [D]. 2023.

[3] Xu Guanghui. Research on complex Scene Reasoning Algorithm Based on multi-modal attention mechanism. 2021.

[4] Song Wenlin. Research on 3D laser point cloud face recognition fused with multi-modal features [J]. Laser Journal, 2023. (in Chinese)

[5] H Zhang, M Zhang, C Li, et al. Research on Blended Teaching Based on"MOOC+SPOC+multimodal classroom" [D]. 2021

[6] MENG Yue, Li Shixin, Chen Fankai, et al. Review of Multimodal Fusion of Point cloud and Image in Autonomous Driving [J]. Computer Science and Application, 2023.

[7] Fu Si Meng. Research on multimodal medical image registration and fusion in neurosurgery [J]. 2020.

[8] Wang Q., Hu X., Lei S., et al. Research on downhole visible light modeling and localization based on point cloud technology [D]. 2024.

[9] Li Xiaojun. Research on multi-modal medical image fusion algorithm based on NSST [J]. 2020.

[10] Li Qiming, REN Jieji, PEI Xiaohan, et al. High-precision point cloud registration algorithm for weak texture surface based on multi-modal data collaboration [J]. Acta Optica Sinica, 2022.

[11] WEI Yun-peng. Research and application of ensemble Learners Based on Multi-modal evolutionary Computation.,2019.

[12] Guo Xu, Mai Ridan Wu Shouer, Guranbaier Turhong. Research on sentiment analysis Algorithm Based on multi-modal fusion [J]. Computer Engineering and Application, 2024. (in Chinese)

[13] Chen Fan. Research on 3D object recognition and 6D Pose Estimation Based on depth and texture information fusion. 2020.

[14] LIU Liqiang. Research on visual 3D Reconstruction based on Deep Learning [J]. 2019.

[15] Xie L J. Research on robust building modeling method based on image dense matching point cloud [J]. 2022.