Ontology Based Semantic Understanding for 3D Indoor Scenes

X H Wang1*, X Gao1, K Zang1 and H J Meng1

1 Department of Information Science and Engineering, Hebei University of Science and Technology, Shijiazhuang, Hebei, 050018, China

*Corresponding author’s e-mail: feifeiwangxiaohong@126.com

Abstract. A scene understanding method of combining semantic segmentation and ontology description is proposed in this paper. The method not only obtains the object of each part of the scene but also gets the relationship between the parts of the scene. Firstly, to realize the semantic segmentation of indoor scenes for 3D point clouds, PointNet is used in this paper as a tool for processing point clouds, and the S3DIS as PointNet datasets for training and testing. Secondly, in order to combine semantic segmentation and ontology, an ontology to describe the relationship between objects and scenes of the semantic segmentation result is used in this paper. Finally, we build indoor scenes ontology based on IndoorGML and use Protégé to display the spatial position ontology of the indoor scene.

1. Introduction

Many approaches have been proposed to solve semantic segmentation, which usually focuses on traditional point cloud segmentation methods and deep learning methods. The traditional point cloud segmentation methods have certain limitations. For example [1-3] mainly concentrate on specific areas, and [4] shows that the edges are particularly sensitive to noise. Convolution of the 3D point cloud is a difficult task due to its unordered and rotational representation. In order to solve the above problem, the point cloud is converted to voxel[5-6]. The convolution of point cloud voxelization can cause a loss of resolution, so the semantic segmentation result is not very well. In the true sense, the use of point cloud for deep learning is the PointNet proposed by Charles's team [7] in 2017. This is the first network that can directly use point cloud as input. The T-NET network is used in PointNet to solve the problem of rotation, and the idea of symmetric function is used to solve the problem of disorder. In summary, this paper uses the PointNet to achieve the semantic segmentation of indoor scenes.

In this paper, we take inspiration from [8]. That reference combines the detected objects and common sense knowledge in the image to build a knowledge graph. so, we propose a semantic understanding method to describe indoor space with ontology[9-10]. This paper focuses on the combination of 3D scenes point clouds and ontology. The main work is as follows:

- The semantic segmentation of indoor space is the foundation of scene understanding. PointNet is used to do segmentation by semantic classification of each point, then points belonging to semantic classification are obtained.
- In order to construct the indoor scene ontology, the classes and levels should be determined. So, we refer to the indoor space standard IndoorGML[11] and define a semantic partition model for indoor space.
- Entity extraction from point cloud datasets is matched for a semantic partition model. Then, according to the result of partitioning, the relationship of the 3D scene is extracted combining with the indoor structural composition.
- Structuring the ontology of the indoor scenes is the last step but the key for scene understanding. Ontology is structured with a seven-step method[12] and is visualized by Protégé.

2. Process of design
Firstly, we sort out the point cloud dataset of indoor scenes. And PointNet receives the dataset as input. The stage of semantic segmentation is aimed at indoor objects including architectural structures such as walls and ceilings. After outputting the segmentation result, we use MeshLab to view the result.

Secondly, we structure the indoor scene ontology. The entity is the base element. And the classification of the entity is defined by the indoor space standard. By the analysis of the segmentation result, entity relation is extracted from the indoor space and the architecture.

Finally, we define the attributes of entities and relations, then build the ontology based on the positional relationship of indoor scenes. The overall design is shown in Figure 1.

3. Method
3.1. Semantic segmentation of 3D point cloud
PointNet is a deep neural network and is used for semantic segmentation. We take the point cloud as input, and then PointNet outputs the per point semantic class labels. Figure 2 shows the simplified PointNet architecture. Different matrices may represent the same object or scene because the point cloud can be rotated in 3D space. In order to solve this problem, PointNet uses a T-Net that a rotation matrix is an output. Then it takes N points inside a block and after a series of Multi-Layer-Perceptrons (MLP) per point, the points are mapped into a higher-dimensional space, these are called local point-features.

In order to align the local features with the original model, we input the local features to the T-Net. For the global-feature, PointNet uses MLP to upgrade 64 dimensions to 128 dimensions, and then to 1024 dimensions. At last, the output is n × 1024 feature matrix. Max pooling is applied to aggregate information from all the points, which reflects the idea of symmetric function. The global-feature is concatenated with all the point-features.

In the stage of semantic segmentation, it is important to fuse local features and global features. With the splicing operation, the input matrix size is n×1088. After another series of MLPs, these combined features are used to predict the n×m output score called the result of semantic segmentation.
3.2. semantic partition model for indoor space

IndoorGML is an open data model for describing the indoor space and is determined as the OGC international standard. According to the indoorGML standard, it is mainly divided into indoor components and indoor space composition. We defined four semantic classes for indoor spaces. One of the classes is obstacles that belong to indoor components in indoorGML, the other three are entrances, connections, and containers. The specific semantic partition model is shown in Figure 3.

3.3. Entity extraction and relation extraction

Information extraction is required to construct ontology. The most important for information extraction is entity extraction and relation extraction. The extracted entity is defined as the class name of the ontology. The relation extraction is to find the relationship between the entities.

By Figure 2, there are four categories as the first class of ontology. Through the analysis of point cloud and the class labels, we get the second class of ontology that are office, conference room, hallway, pantry, WC, auditorium, etc. They all called indoor space. And with the result of semantic segmentation, we take the semantic labels as the third class of ontology such as wall, window, door, table, chair, and so on called indoor objects.

In the part of relationship extraction, we need to visualize the indoor scene segmentation result before. And then the entities are found in the segmentation map by comparing color labels and category labels. As shown in Table 1, it is divided into two categories: inclusion relationship and position relationship.
Table 1. Relationship table of indoor space

| Attribute | Domains       | Ranges        | Description                      |
|-----------|---------------|---------------|----------------------------------|
| Have      | office        | chair, table  | Show what is in the space        |
| Top_of    | ceiling       | beam, table   | Show indoor objects position relationship |
| Left_of   | door          | window        | Show indoor objects position relationship |
| Near_by   | table         | chair         | Show the proximity of indoor objects |

3.4. Constructing ontology
The construction of ontology is divided into seven steps.

Above all, the ontology description domain and scope are determined. Ontology is described by the class or concept, relationship, function, axiom, and instance of a given domain, and so we collected and organized the concepts in the field of indoor scenes.

Then we take indoor space information standard indoorGML as the overall planning benchmark in the field of indoor scenes and carry out the division and relationship analysis of indoor space. In the part of indoor space division, the hierarchy between classes is defined for the construction of ontology. The superclasses and subclasses are also determined, and the attributes of the given class are assigned value. We used Protégé to create classes, attributes, and instances.

Finally, we use the Web Ontology Language (OWL) to describe ontology. The steps to the construction of the ontology are shown in Figure 4.

4. Implementation and result display

4.1. The datasets
The S3DIS dataset of Stanford University is used in this experiment. There are six scenes in total. The first five scenes are used as the training set and the sixth scene is used as the test set. Each scene is composed of conference room, printing room, hallway, office, food storage room, and auditorium, etc. Each indoor space is composed of several types of objects, namely ceiling, floor, beam, chair, sofa, etc.

4.2. Semantic segmentation
A Dell laptop and Centos operating system are used in the experiment. The programming environment of PointNet is python2.7. And the GPU uses the NVIDIA TitanX GPU of the Supercomputing Center of Yanshan University. This experiment mainly focuses on the semantic segmentation of indoor scenes. The PointNet has trained 50 epochs, respectively, with batch size 64, learning rate 0.001. The learning rate decay is set to 0.5 in the network. And the Adam solver is adopted to optimize the network on GPU. The test result is the segmentation accuracy of each type of object and generates a segmentation result file in obj format. Different colors are used to represent different objects in the segmentation scene.
4.3. Construction of 3D point cloud indoor scene ontology

The Protégé is used to model the ontology of the indoor space. And four major categories were created: entrance, obstacle, connect, and container. After all entities, classes and relationships are defined, they need to be instantiated. The schematic diagram of the ontology instantiated is shown in Figure 5, which is shown in the conference room.

![Figure 5. Ontology instantiation](image)

In Figure 5, the rectangular box represents the superclass, and the round rectangular box represents the subclass. It's the result of PointNet in the center of Figure 5, and the thin arrows are the object of indoor. The attributes above the thin arrows are described in Table 1. For example, at the top of the figure, the conference room has a ceiling, and the ceiling is top of the wall, door, and floor, etc. The thick arrows are the entity instance. Obviously there are some entities belong two classes.

Through the analysis of Figure 5, we can quickly build an ontology. In Figure 6, we can clearly see the distribution of hierarchy, but in the result of semantic segmentation like the center of Figure 5 emphasizes a whole and partial result and can not see the hierarchical structure well.

![Figure 6. Ontology visualizations](image)
5. Conclusion

In this paper, the deep learning is used to segment the 3D point cloud indoor scene. Then the segmentation result information is extracted, and an indoor space ontology based on the relative position relationship is established and visualized with Protégé.

The form of expression transforms the model of semantic understanding in 3D space from graphical information to textual information. On this basis, property values of indoor space and objects can be assigned, and ontology in other fields can be formulated such as the ontology in the field of architectural design industry or 3D space navigation. Eventually, a knowledge graph will be constructed. More data will be collected to build a knowledge graph in future research.

Acknowledgments

This work was supported by the university level funding project of Hebei University of science and technology. The numerical calculations in this paper have been done on the supercomputing system in the High-Performance Computing Center of Yanshan University.

References

[1] W. Yue, J. Lu, W. Zhou and Y. Miao. (2018) A new plane segmentation method of point cloud based on mean shift and RANSAC. In: Chinese Control And Decision Conference (CCDC). Shenyang. pp. 1658-1663.
[2] N. Luo, Q. Wang, Q. Wei and C. Jing, (2019) Object-Level Segmentation of Indoor Point Clouds by the Convexity of Adjacent Object Regions. IEEE Access, vol, 7. pp. 171934-171949, 2019.
[3] S. Yan, Y. Peng, G. Wang, S. Lai and M. Zhang. (2019) Weakly Supported Plane Surface Reconstruction via Plane Segmentation Guided Point Cloud Enhancement. In: IEEE Access, vol. 8. pp. 60491-60504.
[4] Z. Kun, SQ. Qiao, WZ . Zhou. (2018) Point cloud segmentation based on 3D shape matching. In: Laser & Optoelectronics Progress. 55(12): 121011.
[5] D. Maturana and S. Scherer, (2015) VoxNet: A 3D Convolutional Neural Network for real-time object recognition. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg. pp. 922-928.
[6] L. Tchapmi, C. Choy, I. Armeni, J. Gwak and S. Savarese, (2017) SEGCloud: Semantic Segmentation of 3D Point Clouds. In: International Conference on 3D Vision (3DV). Qingdao. pp. 537-547.
[7] R. Q. Charles, H. Su, M. Kaichun and L. J. Guibas, (2019) PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI. pp. 77-85.
[8] Yang J., Lu J., Lee S., Batra D., Parikh D. (2018) Graph R-CNN for Scene Graph Generation. In: Ferrari V., Hebert M., Sminchisescu C., Weiss Y. (eds) Computer Vision – ECCV 2018. Lecture Notes in Computer Science, vol 11205. Springer, Cham.
[9] L. Prananingrum, N. Aziz, L. Wulandari, L. Y. Banowosari and A. Suhendra. (2019) Web Based Ontology Implementation for Information Search System. In: 2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI). Yogyakarta, Indonesia. pp. 499-504.
[10] B. Merin and W. A. Banu. (2020) Discovering Web Services By Matching Semantic Relationships Through Ontology. In: 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India. pp. 998-1002.
[11] K. Ilku and J. Tamas. (2018) IndoorGML modeling: A case study. In: 2018 19th International Carpathian Control Conference (ICCC). Szentivanyvarad. pp. 633-638.
[12] Natalya F. Noy, and Deborah L. Mc Guinness, (2002) Ontology Development101: A Guide to Creating Your First Ontology [EB/OL]. pp. 1-24.