CS230

NeRF-to-BIM : Semantic Segmentation for Construction Project with Neural Radiance Fields

Shun Hachisuka
Department of Civil and Environmental Engineering
Stanford University
shnhchsk@stanford.edu

Abstract

Scanning buildings application for creating Building Information Modeling (BIM) relies on point clouds. Unfortunately, the application requires expensive hardware and laborious tasks. Therefore, we introduce an image-based framework, exploiting recent advancements in computer vision with Neural Radiance Fields (NeRF). NeRF is state-of-the-art (SOTA) for synthesizing novel images and reconstruction tasks but lacks specific applications in the architecture, engineering and construction (AEC) domain. We propose a 2-step approach: 1) 3D reconstruction of a construction site using NeRF. 2) Semantic segmentation with pre-trained and fine-tune deep learning algorithms. Finally, we perform qualitative and quantitative analysis of our approach in the context of facade scaffolding design.

1 Introduction

In the construction industry, collecting accurate as-built models of current conditions on a project site is becoming a growing need for efficient project management. 3D laser scanning allows the contractors to obtain accurate and up-to-date point clouds information, which can be exported for Building Information Modeling (BIM). The whole process from capturing the space as point clouds data to turning it into BIM models is generally called Scan-to-BIM. While the Scan-to-BIM idea is essential to enhance collaborative management in the construction industry, it needs to solve time-consuming and laborious tasks problems. Although Scan-to-BIM allows contractors to obtain a highly accurate 3D digital representation of the construction project, the automated process is necessary in order to deliver up-to-date BIM models quickly without manual efforts.
To address these challenges, we propose a new approach for the 3D reconstruction workflow by applying semantic segmentation algorithms to the point clouds data created with NeRF. We call this new approach NeRF-to-BIM. After the implementation of this approach, we evaluate the result and discuss the further potential to obtain enough quality segmented point cloud while reducing manual efforts.

2 Related works

Based-NeRF development study

NeRF was introduced by Mildenhall et al. in 2020 first [9], a ground-breaking method that renders realistic 3D scenes from a sparse set of input collection of 2D images. This original NeRF algorithm has several disadvantages such as slow training and rendering by optimizing. Several new papers derived from the NeRF aim at improving the rather slow training and rendering time of the original

CS230: Deep Learning, Winter 2018, Stanford University, CA. (LateX template borrowed from NIPS 2017.)

NeRF paper.
Instant NeRF[4] reduces the rendering time by a technique developed by NVIDIA called multi-resolution hash grid encoding. This new input encoding method enable to gain high-quality output with a tiny neural network that runs rapidly.

Semantic segmentation algorithms on S3DIS

3D point cloud semantic segmentation is the process of classifying point clouds into multiple homogeneous regions. The classified points in the same region can be derived the same properties, which is essential to convert to BIM. While the semantic segmentation in 2D image analysis has developed, the segmentation in point clouds is challenging because of high redundancy, uneven sampling density, and lack of labeled point clouds data.
PointNet[6] is the first promising algorithm that feeds point clouds directly into the DL architecture. This algorithm is a ground-breaking solution because it solves the problem that the point cloud format is often transformed into the data large.
PointNet++[7] is built upon PointNet, which fails to capture local structure and generalize to complex scenes. PointNet++ constructs a sampling and grouping scheme to learn hierarchical features from multiple scales and from varying densities, which usually results in greatly decreased performance. The result is significantly better than PointNet benchmarks on 3D point clouds.
PointNeXt[8] improved training strategies based on the classical PointNet++ through a systematic study of model training and scaling strategies. The inverted residual bottleneck design and separable MLPs into PointNet++ enables efficient and effective model scaling (Figure 3). PointNeXt establishes a new state-of-the-art performance with 74.9% mean IoU on S3DIS.

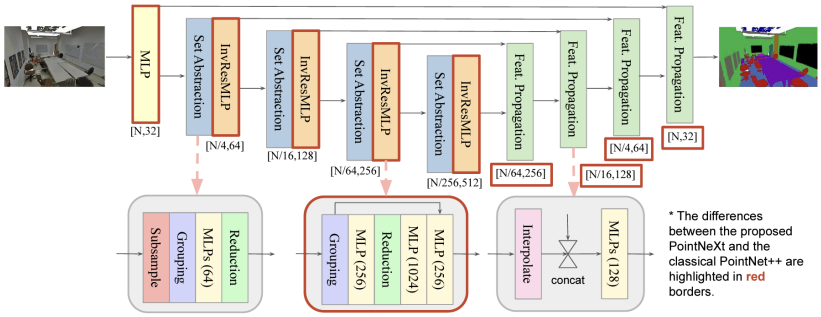

Figure 1: PointNeXt Architecture

Building-related semantic segmentation point clouds implementation

Grilli et al. (2017) [2] explains the methodologies to segment and classify 3D point clouds. The first implementation of semantic segmentation for buildings was done by Liu et al.(2017) [3]. In this study, a deep learning method for segmenting a facade into semantic categories was implemented on a full image scale in the task of building facade parsing.
Murtiyoso et al. (2022) [5] developed the implementation of deep learning-based semantic image segmentation on the photogrammetric 3D reconstruction and classification workflow. The semantically classified point clouds are automatically created as the final output.
Cao and Scaioni (2022)[1] proposes a pre-training method for 3D building point clouds that learns from a large source dataset and evaluates the proposed method by employing four fully supervised networks as backbones. The results illustrate that pre-training on the source dataset improves the performance of the target dataset with an average gain of 3.9%.

3 Dataset

At this time, while there is a lack finely-annotated of data, the S3DIS dataset[10] is the most trained in various studies among the datasets related to building. S3DIS dataset consists of a large-scale indoor environment including six indoor areas with 271 rooms for a total of 695 million points.

Each point in the scene point cloud is annotated with one of the 13 semantic categories, which are structural elements (ceiling, floor, wall, beam, column, window, and door), furniture (table, chair, sofa, bookcase, and board) and clutter for all other elements.
The S3DIS dataset used in this study is insufficient to return enough quality results because the features of our point cloud data in construction scenes are different from the S3DIS, which is a 3D interior dataset. In this scarce dataset situation, we train on 3D architectural models initializing with the pre-trained on S3DIS dataset. We create the 3D architectural model that is transformed into labeled point clouds to be used for training. If the semantic segmentation models identify some of the correct features, it is worth fine-tuning the existing model because we can redefine the required further work.

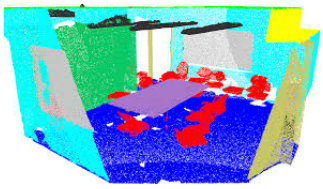

Figure 2: A S3DIS example

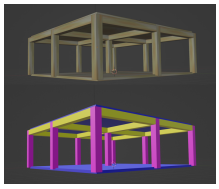

Figure 3: The 3D model for training

4 Method

In this paper, we propose 2-step-approach (Figure 4). First, we reconstruct 3D scenes of a construction site using NeRF. Second, we create semantic segmentation with pre-trained DL algorithms. For the first step, we use Instant NeRF which takes a few minutes to train a great-looking visual which is enough quality for this experimental study. In a further study, we aim to obtain large-scale 3D environments with other NeRFs implemented for large-scale scenes.
For the second step, we apply PointNeXt, which is SOTA on the S3DIS dataset, for semantic segmentation of point clouds data that is created in the first step.

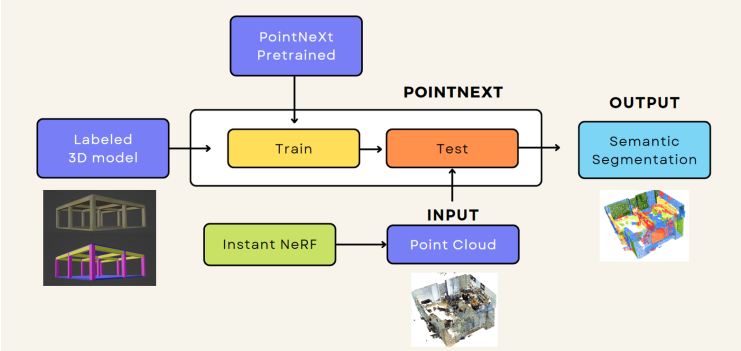

Figure 4: Diagram of NeRF-to-Semantic Segmentation

Although the classification of the point cloud elements can be the same as S3DIS dataset, the features of the point cloud data is different from the S3DIS, which is 3D interior dataset. Aiming to obtain more accurate classified output in this scarce dataset situation, initialization with the pre-trained weight is one of the options to get the best result for the semantic segmentation network. Therefore, we test based on the three training methods and evaluate the performance(Figure 5). The first way is training on only 3D model which is an annotated 3D model we prepared. The second way is using only the weights pretrained by PointNeXt. The third way is using pre-trained weight and the 3D model for training.

5 Experiments

We created a very simple structural object from 2D pictures by applying Instant-NeRF. We prepared 127 image files which are resized into 2,016 by 1,512 pixels(Figure 6). The larger resolution goes out

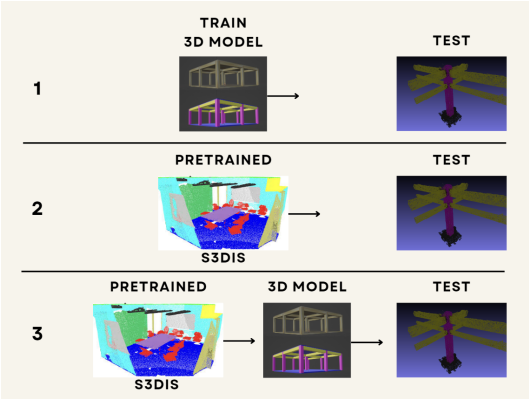

Figure 5: Three experimental methods of training

of GPU memory with NVIDIA GeForce RTX 3070 Laptop GPU. The output from these pictures includes a lot of noise as Figure 7 shows, especially in the area where photos were not taken enough. In terms of beams and columns which are taken from multiple angles, NeRF generates enough quality.

Figure 6: Partial set of pictures for NeRF

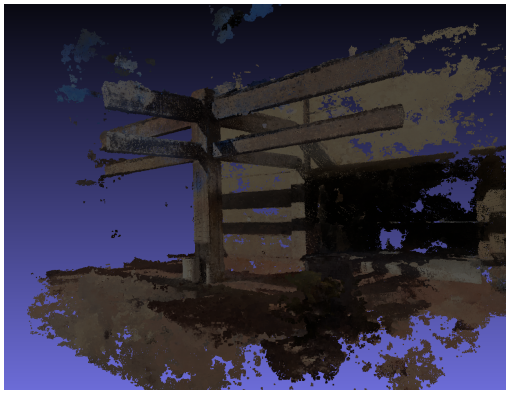

Figure 7: Execution of NeRF from the pictures

To proceed to semantic segmentation training, we removed some noise and excerpted a partial area that consists of only the columns and beams. We classified the data into beams and columns for evaluation of the semantic segmentation performance (Figure 8).

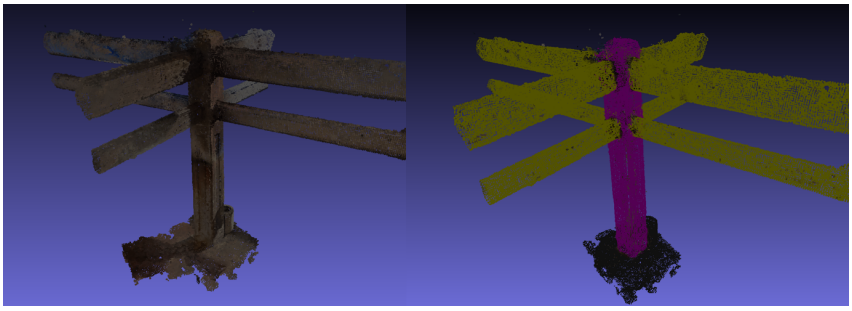

Figure 8: Left: Input point cloud, Right: Ground truth

6 Results and Discussion

Table 1 shows the result of accuracy on semantic segmentation and Figure 9 illustrates the outputs tested on each training method. While training only the 3D architectural model performed relatively lower than others, the other two methods seem to recognize the column, which is represented in pink. Looking at beams, the third method shows slightly higher performance than the performance of the pre-trained method as the yellow beam area increases. Actually, comparing the accuracy rate, after applying the new 3D architectural model for training, the algorism performs higher than the pre-trained performance (Table 1). The reason of this difference is supposed to be that there is no feature of beams in the S3DIS interior dataset.
In terms of the entire process of NeRF-to-BIM, if we can obtain the same quality as the data we used for semantic segmentation, there is enough potential to generate a semantic segmentation point cloud from 2D pictures. However, looking at the detail of the NeRF output, there are still some problems. NeRF creates blue point clouds of the sky in the middle of space and a lot of points behind the surface of objects, which are never created by the use of laser scanning devices. This noise requires manual work to remove it. In addition, the shadow affects the shape. NeRF creates more accurate points in the darker color area, whereas it creates more sparse points in the lighter color area. We need further study on generating more accurate point clouds with NeRF for large-scale scenes. This experimental study shows the potential to improve semantic segmentation accuracy in construction scenes with labeled architectural data.

Table 1: Semantic segmentation results on S3DIS dataset
(mAcc: average class accuracy OA: overall accuracy mIoU: mean Intersection-over-Union)

No	OA	mAcc	mIoU
1	11.03	80.63	18.06
2	25.90	86.99	24.08
3	27.82	87.26	47.27

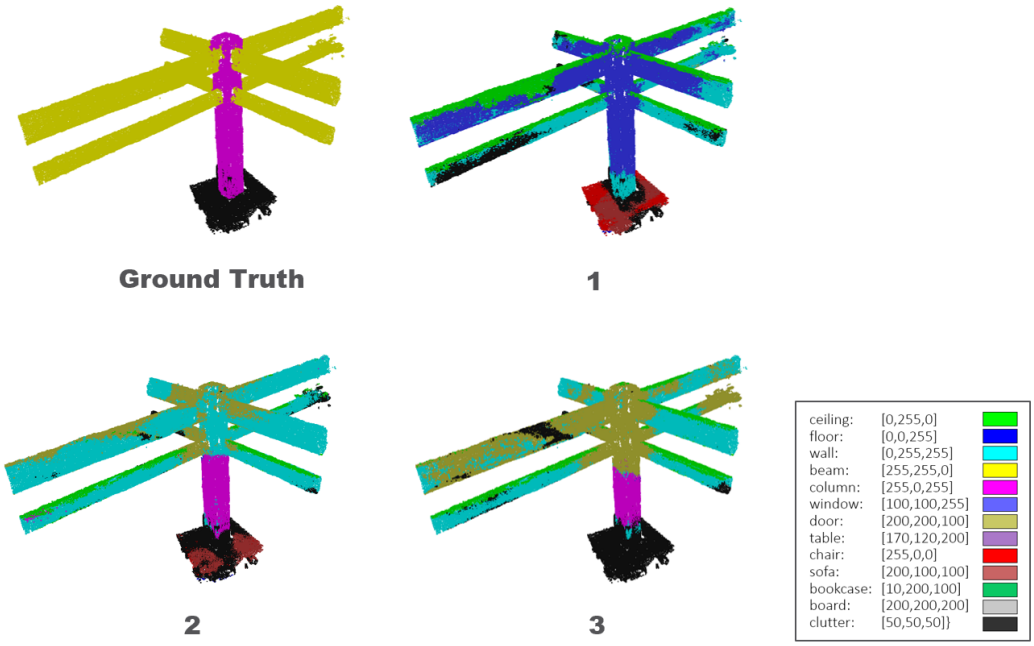

Figure 9: The results of semantic segmentation on three different training methods

7 Team Member Contributions

No collaboration with anyone else.

8 Acknowledgements

I would like to express my very great appreciation to Alberto Tono in Department of Civil and Environmental Engineering for his valuable and constructive suggestions during the planning and development of this research work. His willingness to give his time so generously has been very much appreciated.

References

[1] Yuwei Cao and Marco Scaioni. A pre-training method for 3d building point cloud semantic segmentation. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2:219–226, 2022.

[2] Eleonora Grilli, Fabio Menna, and Fabio Remondino. A review of point clouds segmentation and classification algorithms. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42:339, 2017.

[3] Hantang Liu, Jialiang Zhang, Jianke Zhu, and Steven CH Hoi. Deepfacade: A deep learning approach to facade parsing. IJCAI, 2017.

[4] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *arXiv preprint arXiv:2201.05989*, 2022.

[5] Arnadi Murtiyoso, Eugenio Pellis, Pierre Grussenmeyer, Tania Landes, and Andrea Masiero. Towards semantic photogrammetry: Generating semantically rich point clouds from architectural close-range photogrammetry. *Sensors*, 22(3):966, 2022.

[6] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660, 2017.

[7] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30, 2017.

[8] Guocheng Qian, Yuchen Li, Houwen Peng, Jinjie Mai, Hasan Abed Al Kader Hammoud, Mohamed Elhoseiny, and Bernard Ghanem. Pointnext: Revisiting pointnet++ with improved training and scaling strategies. *arXiv preprint arXiv:2206.04670*, 2022.

[9] Suhani Vora, Noha Radwan, Klaus Greff, Henning Meyer, Kyle Genova, Mehdi SM Sajjadi, Etienne Pot, Andrea Tagliasacchi, and Daniel Duckworth. Nesf: Neural semantic fields for generalizable semantic segmentation of 3d scenes. *arXiv preprint arXiv:2111.13260*, 2021.

[10] Qiangeng Xu, Xudong Sun, Cho-Ying Wu, Panqu Wang, and Ulrich Neumann. Grid-gcn for fast and scalable point cloud learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5661–5670, 2020.