LiDAR Data Segmentation using Deep Learning for Indoor Mapping

Enbo Zhou*, Keith C. Clarke, Joel Salzman, Haoyu Shi and Bryn Morgan

Department of Geography, UC Santa Barbara, Santa Barbara, CA, 93106-4060 USA
* ebzhou@geog.ucsb.edu

Keywords: Indoor Mapping, LiDAR, Segmentation, Deep Learning

Introduction

Indoor mapping is an important precondition for indoor cartography (Chen and Clarke 2019). Recent years have seen a rising demand for indoor maps in fields such as indoor navigation, building information management, 3D-GIS, robotics, etc. Traditionally, indoor maps are made manually or semi-automatically from images or manual measurements. Many emerging techniques can provide useful data for mapping indoor scenes, for example, WIFI positioning, LiDAR scanning, and Bluetooth positioning. Among these data sources, LiDAR scanners can provide both accurate and rich surface information for indoor scenes and their component objects. Portable LiDAR scanners are easy to deploy to retrieve data in different kinds of indoor spaces, making them a good choice to acquire indoor map information. However, due to the irregular and unordered structure of the scans, the difficulty in conflating overlapping and adjacent scans, varying point density, and the massive volume of LiDAR data, there is a huge gap between LiDAR point cloud data and the use of LiDAR-based 3D models for indoor mapping (Xie et al. 2019; Zhang et al. 2019).

In this study, we first analyzed the information requirements of indoor mapping and the characteristics of LiDAR point cloud data. Then two deep learning frameworks *PointNet* (Qi et al. 2017a) and *PointNet++* (Qi et al. 2017b) were implemented to recognize objects in LiDAR point clouds. PointNet can deal with unordered irregular individual points and combine global feature with pointwise features. Based on PointNet, PointNet++ considers neighborhood features and incorporates multiscale features for segmentation. We trained and evaluated these networks with the open Stanford Large-scale 3D Indoor Spaces Dataset (S3DIS, Armeni et al. 2017). Then we segmented real-world data from offices in the UCSB Geography department with the derived networks. To meet the demand of multiple levels of detail (LODs) (Chen and Clarke 2019) for indoor mapping, we designed a multiscale sampling strategy to sample the original dense LiDAR data. Multiscale samples were fed into the network to get segmented results with different LODs. Various segmentation results showed the feasibility of methods for indoor mapping information extraction.

The main contributions of this research are to: (i) analyze LiDAR data characteristics and indoor mapping information demands; (ii) implement state-of-the-art deep learning models for indoor point cloud segmentation; and (iii) design a way to get different LODs from a single dense point cloud.

Methods

Deep learning methods such as PointNet and PointNet++ consume point cloud data directly. Before training and predicting, LiDAR data are rotated and aligned in the same coordinate system (Chen 2018) and joined. In addition, LiDAR data are divided and sampled into small 3D blocks according to network configurations. When making predictions, data are sampled based on the required LOD. Farthest Point Sampling (FPS) was used to sample LiDAR point clouds at different scales, and then to process them with neural networks.

1. PointNet

The PointNet software library was a pioneer for deep learning using point sets. It considers three characteristics of point cloud data, namely unordered, interaction among points and invariance under transformations. Its network architecture is shown in Fig.1, which includes classification and segmentation networks. In PointNet, several multi-layer perceptions are used to get pointwise features. To deal with unordered input, PointNet uses a symmetry function *MaxPooling* to get global features. Then in the segmentation network, global features and pointwise features are concatenated for point cloud segmentation. Within PointNet, two transform networks are implemented to conduct geometric transformations. These transform networks can retain invariance when handling point clouds with transformations.

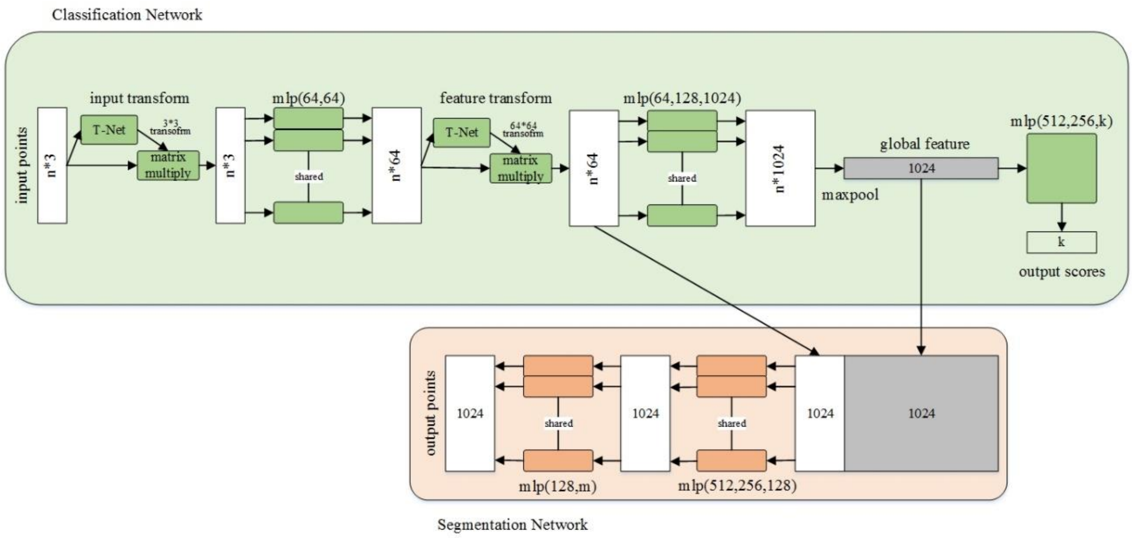

Fig. 1 Network Architecture of PointNet (Adapted from: Qi et al. 2017a)

2. PointNet++

PointNet++ is a deep neural network software library that followed from PointNet, which captures local structures and incorporates multiscale features for point cloud segmentation. The network architecture of PointNet++ is shown in Fig. 2, which also includes classification and segmentation networks. For the segmentation network, it has a downsampling part (encoder, on the left) and an upsampling part (decoder, in the upper right). The main component in the downsampling part is the set abstraction level including sampling layer, grouping layer and density adaptive PointNet layer. The sampling layer selects certain numbers of points as the centroids of local regions. The grouping layer constructs local regions by finding neighboring points. The density adaptive PointNet encodes local regions into features. The upsampling part uses interpolation and PointNet to upsample features into the original space and generate per-point predictions.

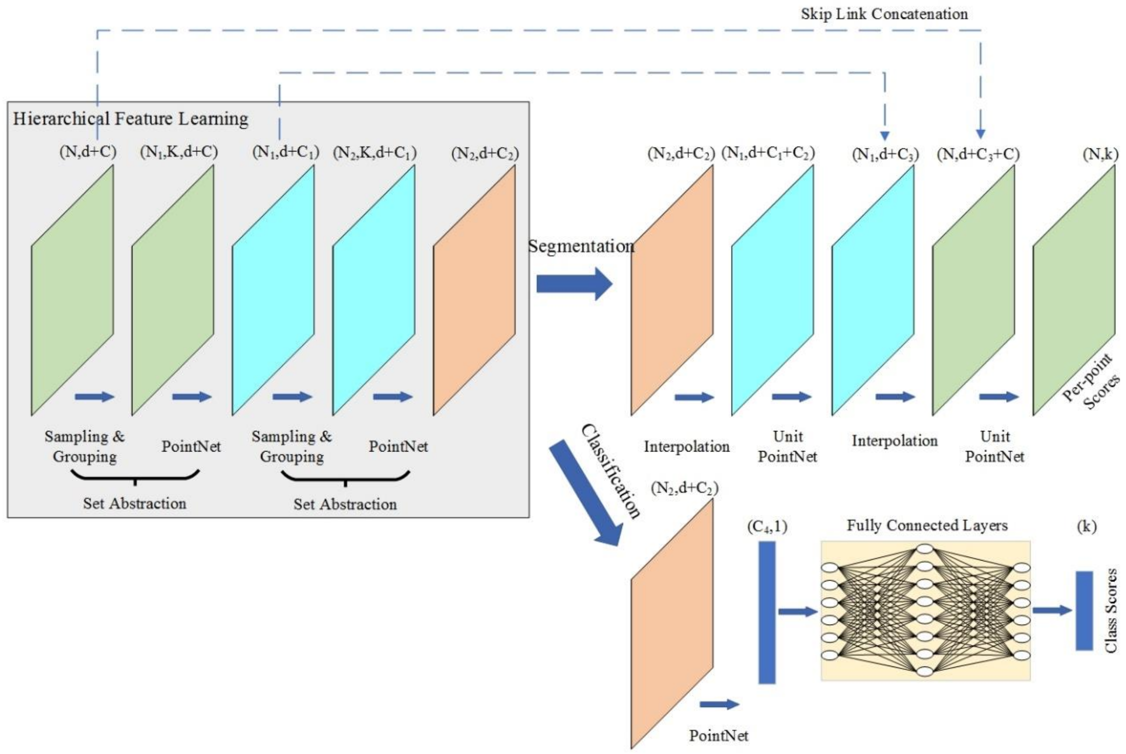

Fig. 2 Network Architecture of PointNet++ (Adapted from Qi et al. 2017b)

Results

We trained the neural networks in PointNet and PointNet++ with S3DIS dataset. The dataset contains 3D scans from Matterport scanners in 6 areas including 271 rooms at Stanford University. To prepare the training data, we first split points by room and then sampled rooms into blocks with voxels 1m x 1m x 1m. Each point is represented by a 9-dimensional vector of XYZ, RGB and normalized location as related to the room (from 0 to 1). The points are labeled as one of 13 categories (ceiling, floor, wall, beam, column, window, door, table, chair, sofa, bookcase, board, clutter). At training time, we randomly sampled 4096 points in each block from areas 1 to 5 on-the-fly. At test time, we tested only on sample points from area 6. The network was trained for 200 epochs. The batch size was 24, and the optimizer used was Adam, an adaptive learning rate optimization algorithm designed specifically for training deep neural networks (Kingma and Ba 2014).

The accuracy and loss of training and testing processes are shown in Figs. 3 and 4. As we can see from Fig.3, training accuracy reached above 95% while testing accuracy reached around 88%. The evaluation accuracy for all rooms was about 0.7379. Fig. 5 shows the distribution of the number of testing areas by accuracy and loss.

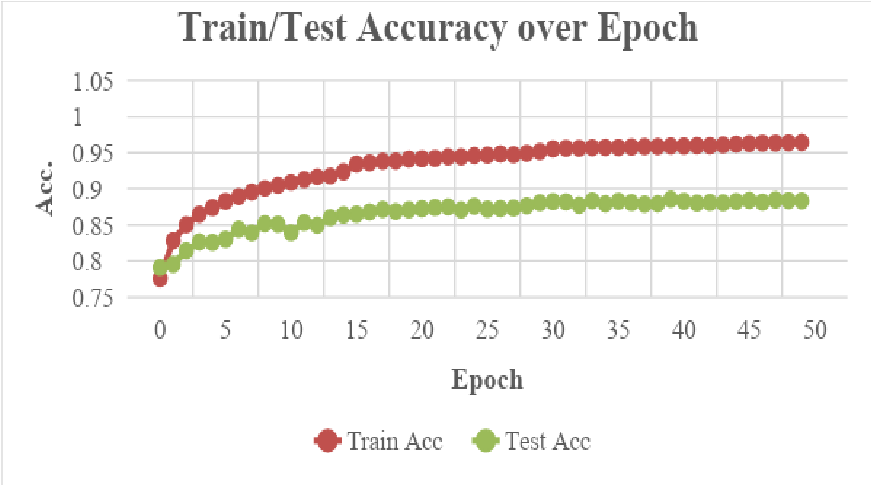

Fig. 3 Training and Testing Accuracy during Training

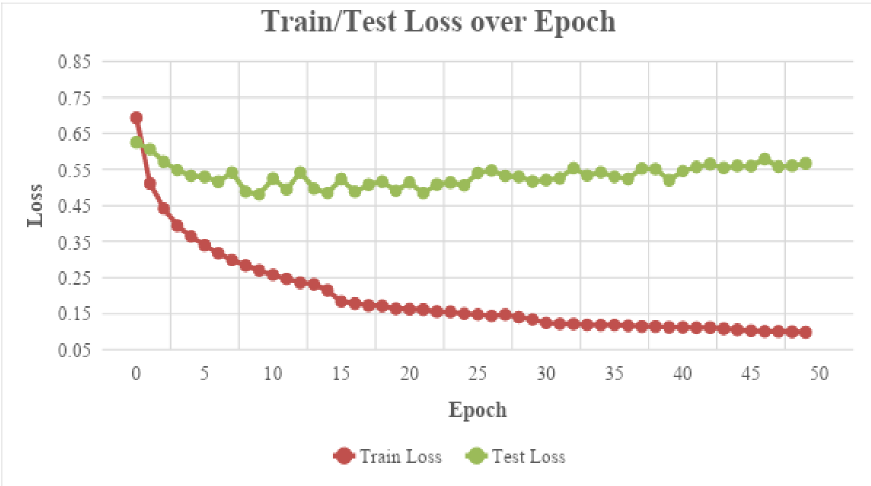

Fig. 4 Training and Testing Loss during Training

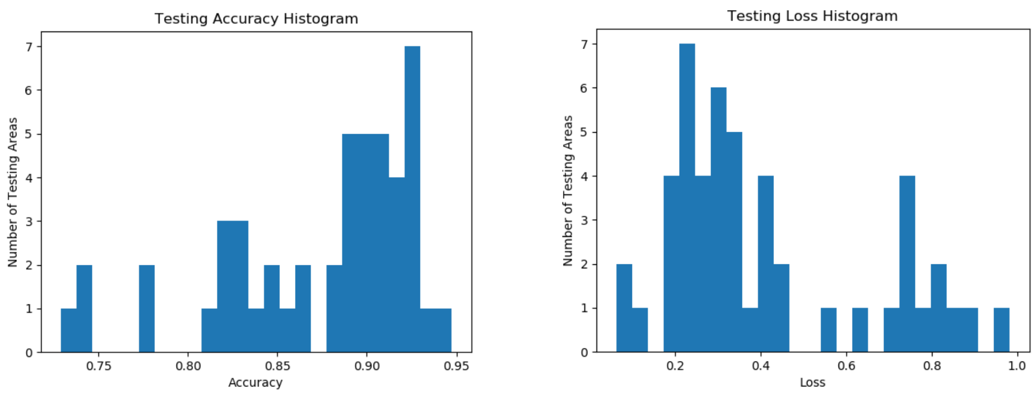

Fig. 5 Sample Accuracy and Loss Distribution

Figs. 6 and 7 show some results for the testing data, in which different colors represent different object types. We then used the trained model to segment LiDAR data acquired in offices at the UCSB Geography department.

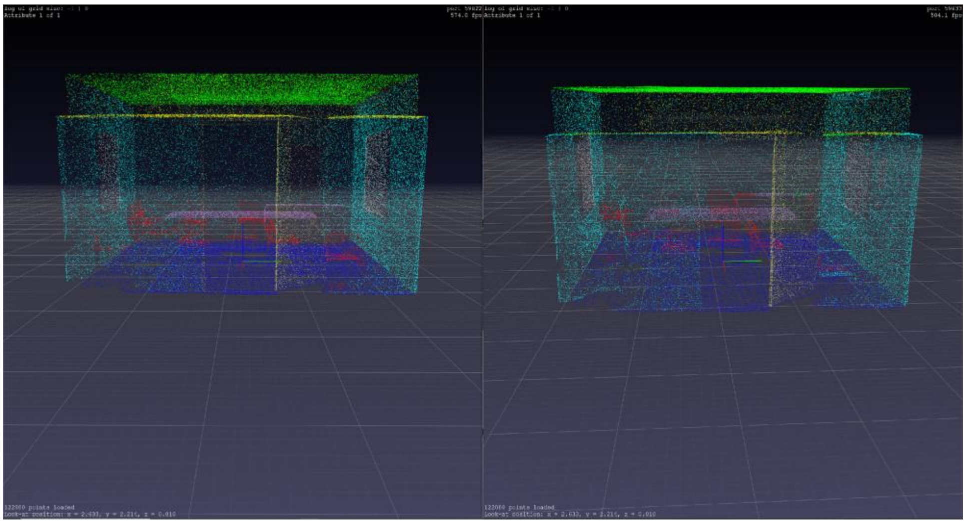

Fig. 6 Ground Truth (Left) and Prediction (Right) of an Office

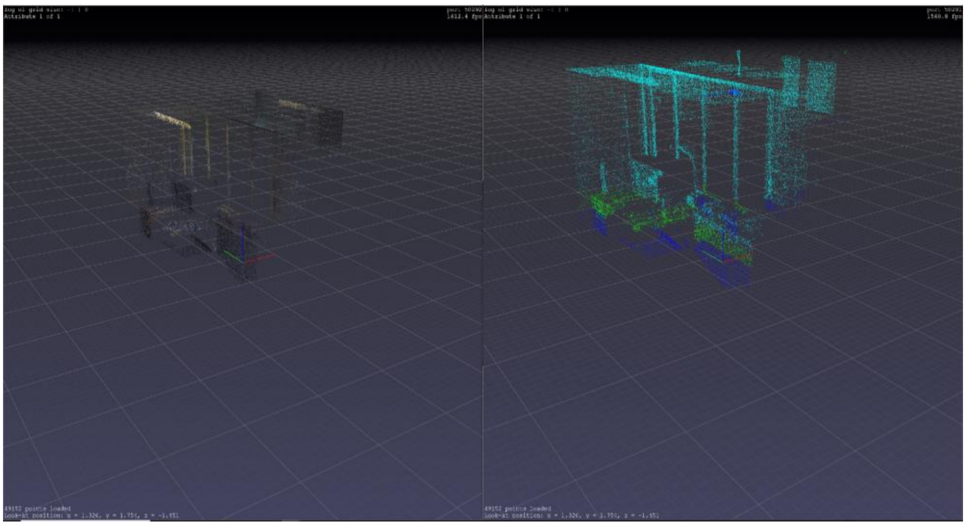

Fig. 7 Sampled True Color Point Cloud and Prediction Label

Discussion and Conclusion

Accurate indoor mapping is important for indoor cartography. However, many problems exist in collecting interior data and extracting useful structured cartographic information. Here, we implemented state-of-art deep learning models to extract indoor objects from LiDAR point cloud data. The deep learning models PointNet and PointNet++ automatically computed pointwise features by considering neighboring points and fusing multiscale features for segmentation. The results using LiDAR data from UCSB offices show promising potential for larger-scale applications of indoor mapping, say to whole floors and buildings. Nevertheless, there are several drawbacks to these methods. Trained models are only capable of identifying a limited number of indoor objects, for example the 13 categories used in this research. LiDAR data needs to be sampled for prediction, which limits the spatial accuracy of the extracted objects. In prior work, we estimated that a point accuracy of 1mm is necessary for indoor cartography (Chen 2018). Finally, large training datasets are needed to train the models. Progress with the same method we used has been achieved by Wu et al. (2019), but far more research is necessary to achieve fully automated, accurate, and reliable indoor mapping.

Acknowledgements:

This research was supported by the National Geospatial-Intelligence Agency's University Research Initiatives (NURI) NGA Grant HM0476-17-1-2002.

References

Armeni, I., Sax, S., Zamir, A. R., & Savarese, S. (2017) Joint 2d-3d-semantic data for indoor scene understanding. *arXiv preprint* arXiv:1702.01105.

Chen, J. (2018). Grid referencing of buildings. In P. Kiefer, H. Huang, N. Van de Weghe, & M. Raubal (Eds.), *Adjunct proceedings of the 14th conference on location based services* (pp. 38–43). Zurich: ETH Zurich.

Chen, J., & Clarke, K. C. (2019). Indoor cartography. *Cartography and Geographic Information Science*, 1-15.

Kingma, D.P. & Ba, J.L. (2014) Adam: A Method for Stochastic Optimization. *arXiv* 2014, arXiv:1412.6980v9.

Qi, C. R., Su, H., Mo, K., & Guibas, L. J. (2017). Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 652-660).

Qi, C. R., Yi, L., Su, H., & Guibas, L. J. (2017). Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *Advances in neural information processing systems* (pp. 5099-5108).

Wu, W., Qi, Z., & Fuxin, L. (2019). Pointconv: Deep convolutional networks on 3d point clouds. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 9621-9630).

Xie, Y., Tian, J., & Zhu, X. X. (2019). A Review of Point Cloud Semantic Segmentation. *arXiv* preprint arXiv:1908.08854.

Zhang, J., Zhao, X., Chen, Z., & Lu, Z. (2019). A Review of Deep Learning-based Semantic Segmentation for Point Cloud (November 2019). *IEEE Access.*