UNIVERSITY OF TECHNOLOGY SYDNEY
Faculty of Engineering and Information Technology

Research on Point Cloud Segmentation and 3D Scene Understanding

by

Anan Du

Supervisor: Jian Zhang
Co-Supervisor: Qiang Wu

A Thesis Submitted
in Partial Fulfillment of the
Requirements for the Degree

Doctor of Philosophy

Sydney, Australia

April, 2022

Certificate of Authorship/Originality

I, Anan Du declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Electrical and Data Engineering/Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Signature: Production Note: Signature removed prior to publication.

Date: 17/04/2022

Acknowledgements

During my doctoral study in UTS, I have received a lot of support from my supervisors, colleagues, and family. Here, I would like to express my sincere gratitude to them.

First and foremost, I would like to thank my Ph.D supervisor, Prof. Jian Zhang, for his guidance and support throughout my research and completion of this thesis. Prof. Zhang Jian has been an academic role model to me. He is rigorous and enthusiastic in work, has deep insights in academic, has rich skills on research writing and presentation and always keeps curiosity on new knowledge. All of these characters have guided me and will continue encouraging me to be a professional researcher. I also want to thank my co-supervisor, Associate Prof. Qiang Wu, for his constructive guidance regarding my research directions. Besides, his optimistic attitude to work and life has inspired me a lot.

I would like to thank my colleague, Dr. Xiaoshui Huang, who is also engaged in research on 3D point clouds. Dr. Xiaoshui has given me a lot of professional guidance during my entire Ph.D. study, from introducing me to the basic knowledge about 3D point cloud, to discussing my follow-up research works, to writing articles. He always works hard and efficiently. He also has a wide range of scientific research fields and often has novel ideas. It is an honor to work with him. I want to thank my co-author Dr. Shuchao Pang for supporting my study and life. In my Ph.D study, we often shared the articles we read, discussed our views, debugged the experiments, and polished each other's articles. In life, when I encounter setbacks, Dr. Shuchao provided me a lot of encouragement. I also want to thank my research collaborator Dr. Tianfei Zhou, for his help to my research. He has given me a lot of support in the method design, validation, and article writing. I also appreciate my colleague, Dr. Lingxiang Yao, for his help in polishing the articles and guidance in doing the

presentation with his excellent communication skills.

I want to thank all my colleagues supervised by Prof. Jian and Associate Prof. Qiang during my Ph.D study. Thanks to Dr. Jinsong Xu, Dr. Junjie Zhang, Dr. Muming Zhang, Dr. Lu Zhang, Dr. Yongshun Gong, Dr. Zhibin Li, Dr. Huaxi Huang, Dr. Litao Yu, Dr. Zongjian Zhang, Dr. Guofeng Mei, Dr. Wenbo Xu, Dr. Yan Huang, Dr. Peng Zhang, Dr. Qiang Li, Dr. Xunxiang Yao. I also thank all other colleagues I have studied with. They make my study and life in UTS beautiful.

At last, I want to thank my parents, brother and sister. They always show their patience, trust and love to me. I love them.

Anan Du

Sydney, Australia, 2022.

ABSTRACT

Recently, the amount of 3D data has been sharply increasing thanks to widely available 3D sensors like Lidar, Kinect and RealSense. Compared to 2D images, 3D data provides clear topology and geometric information, which is very important for many computer vision applications. Among many types of 3D data, the point cloud is widely used because of its easy availability and storage. This thesis summaries the works that have been conducted on understanding the high-level semantics and basic structure of 3D scenes based on point cloud data.

Firstly, a fully-supervised point cloud semantic segmentation framework is designed. Contextual information can help resolve ambiguity and improve the robustness of a recognition system. To capture long-range context, a long-short-term feature bank based framework is introduced to exploit the patch-wise relationship for point cloud semantic segmentation. This approach can capture context in an arbitrary range by costing a little extra computation than a standard segmentation model. Experiments demonstrate that the proposed approach outperforms other point cloud semantic segmentation approaches exploring long-range context.

Manually labeling point cloud datasets, especially point-wise annotations, for fully-supervised methods is expensive. To avoid manually annotating 3D keypoints, a weakly-supervised 3D keypoint extraction method for point cloud registration, called KPSNet, is proposed, which only uses the relative transformation matrices between input pairs of point clouds as weak labels to establish point-to-point correspondences on-the-fly for training. Moreover, KPSNet can simultaneously detect 3D keypoints and learn the representations. Experimental results reveal that our proposed method performs better in point cloud registration than other methods without keypoint annotations.

To cut the need for manual point-wise annotations, a weakly-supervised point

cloud semantic segmentation approach is proposed, which uses coarse labels only. The proposed approach reformulates this problem as a semi-supervised point cloud semantic segmentation problem with noisy pseudo labels. Then a three-branch network is proposed, which can reduce the impact of noises in pseudo labels and leverage the inner structure of point clouds to improve the segmentation performance. By evaluating on benchmark datasets, this method surpasses other weakly-supervised methods and fully-supervised method PointNet++, and narrows the gap with other outstanding fully-supervised approaches. Subsequently, another weakly-supervised point cloud semantic segmentation framework is introduced, which uses incomplete point-level labels. It explores contrastive learning with cross-sample contrast and low-level contrast and a pseudo label refinement module to mine more helpful supervision information from pseudo labels. This method gains state-of-the-art performance on benchmark datasets with several weakly-supervised annotation settings.

Keywords: Point cloud, semantic segmentation, 3D keypoint, weakly-supervised, 3D scene understanding

Contents

List of Figures

List of Tables

List of Publications

Conference Papers

1. **Anan Du**, Shuchao Pang, Xiaoshui Huang, Jian Zhang, Qiang Wu, "Exploring Long-Short-Term Context For Point Cloud Semantic Segmentation", In *2020 IEEE International Conference on Image Processing (ICIP)*, pp. 2755-2759. IEEE, 2020.
2. **Anan Du**, Xiaoshui Huang, Jian Zhang, Lingxiang Yao, Qiang Wu, "KPSNet: Keypoint Detection and Feature Extraction for Point Cloud Registration", In *2019 IEEE International Conference on Image Processing (ICIP)*, pp. 2755-2759. IEEE, 2019.

Journal Papers

1. **Anan Du**, Shuchao Pang, Xiaoshui Huang, Jian Zhang, Qiang Wu, "Weakly-supervised Point Cloud Semantic Segmentation with Coarse-grained Labels." In *IEEE Transactions on Multimedia*. (Submitted)
2. **Anan Du**, Tianfei Zhou, Shuchao Pang, Jian Zhang, Qiang Wu, "Point Contrast and Labelling for Weakly-supervised Point Cloud Semantic Segmentation." In *IEEE Transactions on Image Processing*. (Under submission)