Exploiting Edge-Oriented Reasoning for 3D Point-based Scene Graph Analysis (Supplementary Materials)

Chaoyi Zhang
University of Sydney
chaoyi.zhang@sydney.edu.au

Jianhui Yu
University of Sydney
jianhui.yu@sydney.edu.au

Yang Song
University of New South Wales
yang.song1@unsw.edu.au

Weidong Cai
University of Sydney
tom.cai@sydney.edu.au

The supplementary materials for [8] contain implementation and training details, as well as other additional specifications for the following studies:

A. 3D $\mathbf{SGG_{point}}$ on Real-World 3D Scans.

B. 3D $\mathbf{SGG_{point}}$ on Synthetic 3D Scenes.

C. Traditional Graph Representation Learning.

A. 3D $\mathbf{SGG_{point}}$ on Real-World 3D Scans

We adopted the same dataset split [3] for method comparisons. To alleviate the serious object class imbalance issues that appeared within the $\mathbf{SG}$ node recognition process, we selected their so-called *RIO27* annotation set (27 object classes[1]) for our $\mathbf{SGG_{point}}$ studies, rather than their initially published annotations (160 object classes) [4]. *RIO27* annotation set was a subset mapping to the raw 160-class one and it was later officially released in their repository (here). Similarly, we firstly filtered out their annotated comparative relationships (e.g., *bigger than* and *darker than*) and following [4] we considered only a subset of the relationships (16 structural relationships[2]) to formulate the $\mathbf{SG}$ edge recognition as multi-class classification problems. All irrelevant objects and inter-object structural relationships were removed to obtain our cleared $\mathbf{SG}$ node and edge annotations. Our densely sampled point cloud representations of 3D real-world scenes, together with these cleared $\mathbf{SG}$ annotations, will be published online for reproducibility, as well as fostering any further $\mathbf{SGG_{point}}$ research.

We chose Adam as the optimizer with learning rate and weight decay set to 1e-3 and 1e-4, respectively. The $\mathbf{SGG_{point}}$ framework was trained for 50 epochs with early stopping techniques applied on held-out validation set, and batch size was set to 4. We randomly cropped 4096 points on-the-fly for each scene by maintaining a same sampling ratio to be shared in between all object instances within any given scenes. Such design was insensitive to the varying object sizes and could thus ensure a balanced point sampling achieved at instance-level. The SGG_{point} framework proposed for real-world 3D scans was established and trained with four 11GB NVIDIA GeForce GTX 1080Ti GPUs. More qualitative results can be found as Fig. 1.

B. 3D $\mathbf{SGG_{point}}$ on Synthetic 3D Scenes

We followed the released dataset split and three-class structural relationship annotations [9] to establish our training procedures. Moreover, the optimizer is selected as Adam with learning rate and weight decay set to 1e-3 and 1e-4, respectively. The total training epochs were set to 100, while batch size was set to 4. Since each room category may own unique object classes, the number of object classes for each room category was listed as follows: $C_{bedroom} = C_{living} = 51$, $C_{office} = 42$, and $C_{bathroom} = 31$. Any scenes containing more than 60 object nodes were divided into sub-graphs for training. Two 11GB NVIDIA GeForce GTX 1080Ti GPUs were employed for this group of experiments.

C. Traditional Graph Representation Learning

Our contributions could also be validated on conventional graph representation learning tasks, such as node-wise classification and whole-graph recognition problems. More specifically, our method was evaluated on three

[1] $C_{object} :=$ {*wall*, *floor*, *cabinet*, *bed*, *chair*, *sofa*, *table*, *door*, *window*, *counter*, *shelf*, *curtain*, *pillow*, *clothes*, *ceiling*, *fridge*, *tv*, *towel*, *plant*, *box*, *nightstand*, *toilet*, *sink*, *lamp*, *bathtub*, *object*, *blanket*}.

[2] $C_{relationship} :=$ {*supported by*, *attached to*, *standing on*, *lying on*, *hanging on*, *connected to*, *leaning against*, *part of*, *belonging to*, *build in*, *standing in*, *cover*, *lying in*, *hanging in*, *spatial proximity*, *close by*}.

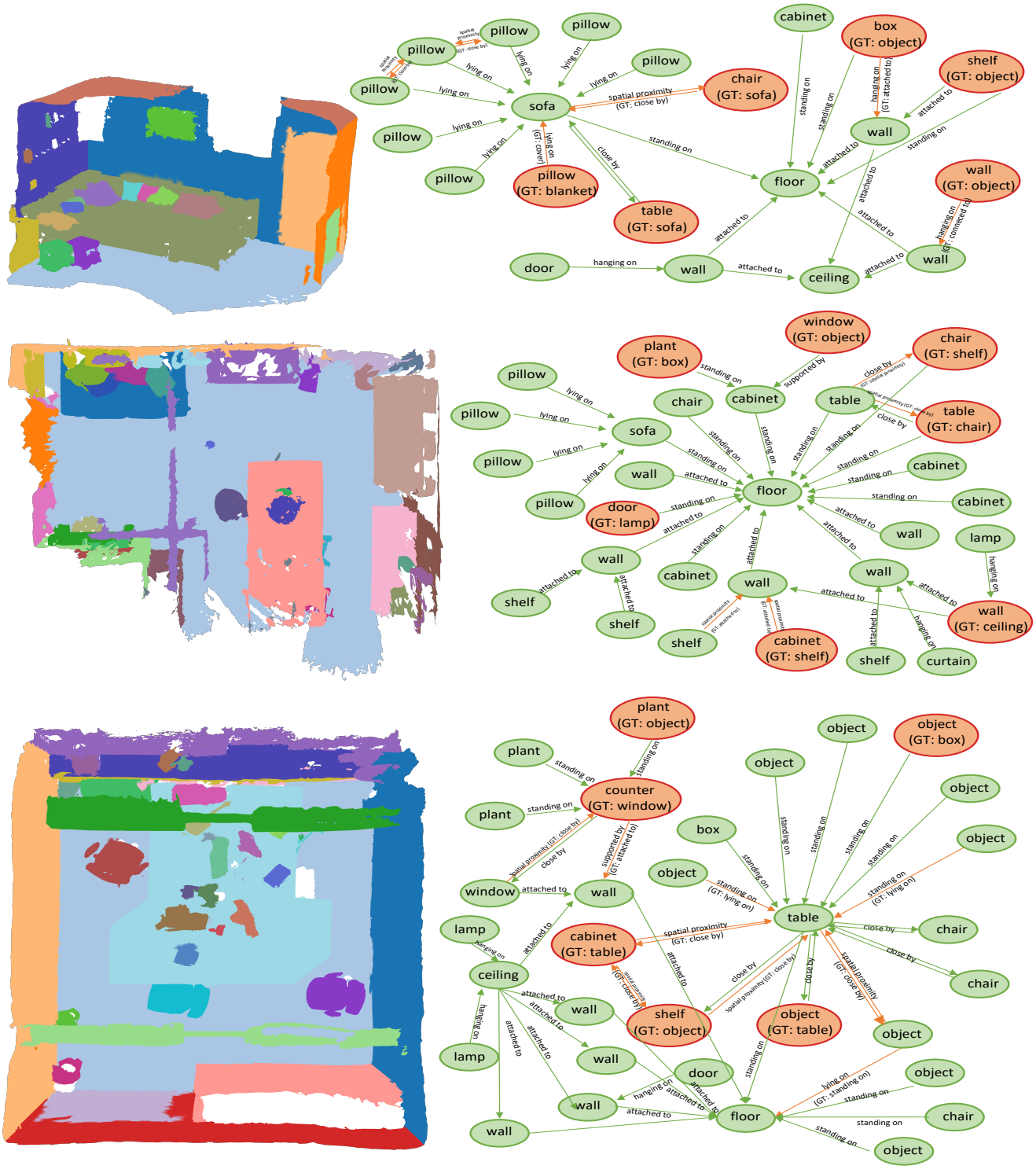

Figure 1. Qualitative visualization of the $\mathbf{SGG_{point}}$ framework, where misclassified object and structural relationship samples are marked with ground truth values in red, while the correct ones are shown in green with ground truth values omitted.

popular citation network datasets (Cora, CiteSeer, and Pubmed) [7] and two molecular datasets (Tox21 and BBBP) [6]. The evaluations were completed through two universal benchmark scripts available for graph representation learning studies, with all specific training settings unchanged for all method evaluation, except for repeating their procedure 50 times for each approach. The following experiments were conducted on one single 8GB NVIDIA GeForce GTX 1070Ti GPU.

C.1. Node-wise classification on citation datasets

We applied a Pytorch Geometric [1] script (here) to replicate the experiments on citation network datasets for evaluations among node-wise classification approaches. More specifically, we adopted Adam as the optimizer with learning rate and weight decay set to 1e-2 and 5e-4, respectively. All methods being investigated were trained over 200 epochs for each run and 50 runs in total to reach a steadily averaged accuracy for performance comparisons. All GNNs were instanced as two-layer networks with ReLU as intermediate non-linearity between, except for EGNN which was reproduced following their settings reported in [2]. Their inner channels were set to 16 by default, unless otherwise specified.

C.2. Whole-graph recognition on molecular datasets

We adopted a DGL [5] script (here) to evaluate whole-graph recognition approaches for molecular analysis. More specifically, Adam was utilized for parameter optimization with early stopping techniques applied over maximum 1000 training epochs. Scaffold splitting policy was employed to divide all datasets into 80% training, 10% validation, and 10% testing sets, where hyper-parameter searches were conducted with Bayesian Optimization for 32 trials, i.e., a randomly initialized model would be trained for each trial, and the best model achieving highest validation performance could then be selected across trials for final evaluation on testing set. We constructed GNNs with their default architectures whose configuration details, as well as their fine-tuned hyper-parameters such as learning rate and batch size, can be found available in the online DGL repository.

References

[1] Matthias Fey and Jan E. Lenssen. Fast graph representation learning with PyTorch Geometric. In *Proceedings of ICLR Workshop on Representation Learning on Graphs and Manifolds (ICLRW)*, 2019. 3

[2] Liyu Gong and Qiang Cheng. Exploiting edge features for graph neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 3

[3] Johanna Wald, Armen Avetisyan, Nassir Navab, Federico Tombari, and Matthias Niessner. RIO: 3D object instance re-localization in changing indoor environments. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2019. 1

[4] Johanna Wald, Helisa Dhamo, Nassir Navab, and Federico Tombari. Learning 3D semantic scene graphs from 3D indoor reconstructions. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 1

[5] Minjie Wang, Da Zheng, Zihao Ye, Quan Gan, Mufei Li, Xiang Song, Jinjing Zhou, Chao Ma, Lingfan Yu, Yu Gai, Tianjun Xiao, Tong He, George Karypis, Jinyang Li, and Zheng Zhang. Deep Graph Library: A graph-centric, highly-performant package for graph neural networks. *arXiv preprint arXiv:1909.01315*, 2019. 3

[6] Zhenqin Wu, Bharath Ramsundar, Evan N. Feinberg, Joseph Gomes, Caleb Geniesse, Aneesh S. Pappu, Karl Leswing, and Vijay Pande. MoleculeNet: a benchmark for molecular machine learning. *Chem. Sci.*, 9:513–530, 2018. 3

[7] Zhilin Yang, William Cohen, and Ruslan Salakhudinov. Revisiting semi-supervised learning with graph embeddings. In *Proceedings of the International Conference on Machine Learning (ICML)*, volume 48, pages 40–48, New York, New York, USA, 2016. PMLR. 3

[8] Chaoyi Zhang, Jianhui Yu, Yang Song, and Weidong Cai. Exploiting edge-oriented reasoning for 3D point-based scene graph analysis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 1

[9] Yang Zhou, Zachary While, and Evangelos Kalogerakis. SceneGraphNet: Neural message passing for 3D indoor scene augmentation. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2019. 1