LEARNING GNNS WITH NOISY LABELS

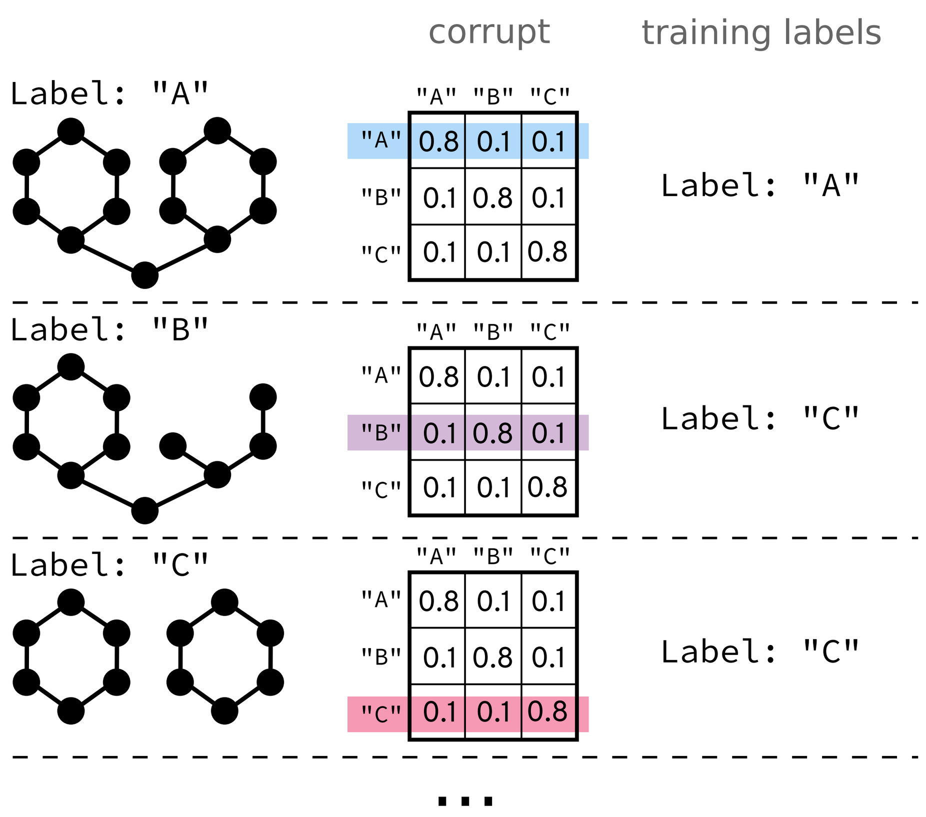
HOANG NT, JUN JIN CHOONG, TSUYOSHI MURATA
TOKYO INSTITUTE OF TECHNOLOGY

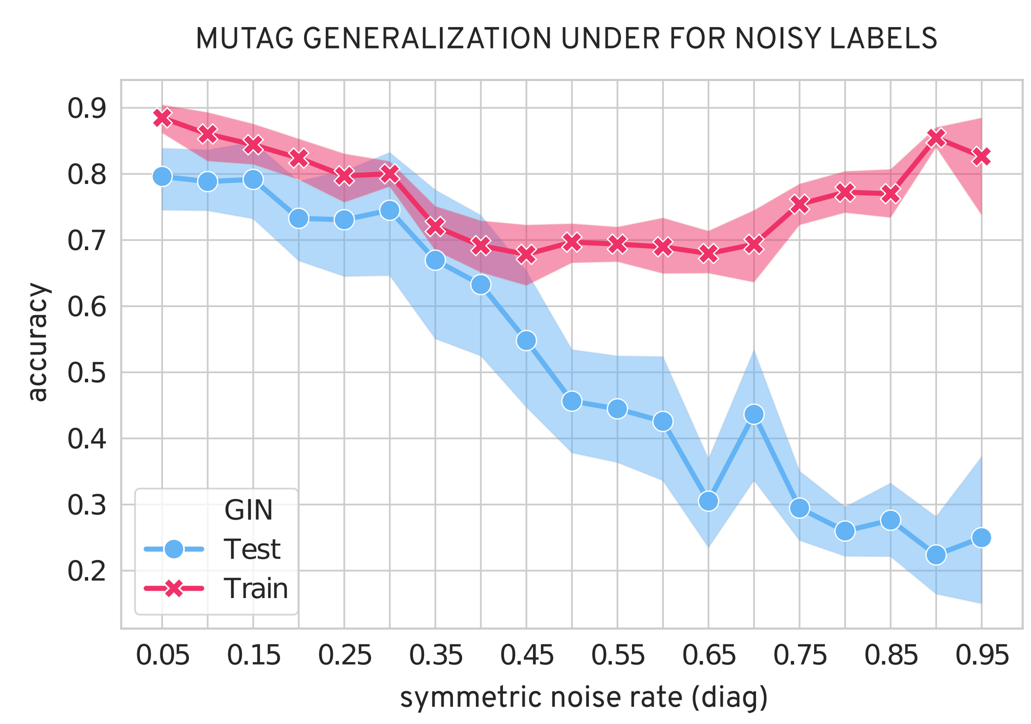
東京工業大学
Tokyo Institute of Technology

LEARNING WITH LIMITED LABELED DATA
ICLR 2019 New Orleans, USA

INTRODUCTION

We study the robustness to symmetric label noise of GNNs training procedures. By combining the nonlinear neural message-passing models (e.g. Graph Isomorphism Networks, GraphSAGE, etc.) with loss correction methods, we present a noise-tolerant approach for the graph classification task. Our experiments show that test accuracy can be improved under the artificial symmetric noisy setting.





GRAPH CLASSIFICATION MODEL

We use a nonlinear message passing model similar to *GIN* [2] and *GraphSAGE* [3]. In the supervised learning setting, such model has two main steps:

1. Accumulate neighborhood information to each vertex v and neural network layer (k):

$$\mathbf{a}_v^{(k)} = \texttt{AGGREGATE}^{(k)}(\{\mathbf{h}_u^{(k-1)} : u \in \mathcal{N}(v)\}),$$
$$\mathbf{h}_v^{(k)} = \texttt{COMBINE}^{(k)}(\mathbf{h}_v^{(k-1)}, \mathbf{a}_v^{(k)})$$

2. Employ a function to learn the overall representation of the graph, then the objective ℓ is optimized with standard backpropagation.

$$\mathbf{h}_{\mathcal{G}} = \texttt{READOUT}(\{\mathbf{h}_v^{(K)} : v \in \mathcal{G}\}),$$
$$\ell(p(y|\mathbf{h}_{\mathcal{G}}), y_{\mathcal{G}}) = \texttt{XENTROPY}(p(y|\mathbf{h}_{\mathcal{G}}), y_{\mathcal{G}})$$

SHORT REFERENCES

[1] Giorgio Patrini, Alessandro Rozza, Aditya Krishna Menon, Richard Nock, and Lizhen Qu. "Making deep neural networks robust to label noise: A loss correction approach," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1944–1952, 2017.

[2] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. "How powerful are graph neural networks?," *International Conference on Learning Representations*, ICLR 2019.

[3] Will Hamilton, Zhitao Ying, and Jure Leskovec. "Inductive representation learning on large graphs," *Advances in Neural Information Processing Systems*, pp. 1024–1034, 2017.

[4] Thomas N. Kipf and Max Welling. "Semi-supervised classification with graph convolutional networks," *In International Conference on Learning Representations*, ICLR 2017.

LOSS CORRECTION SCHEMES

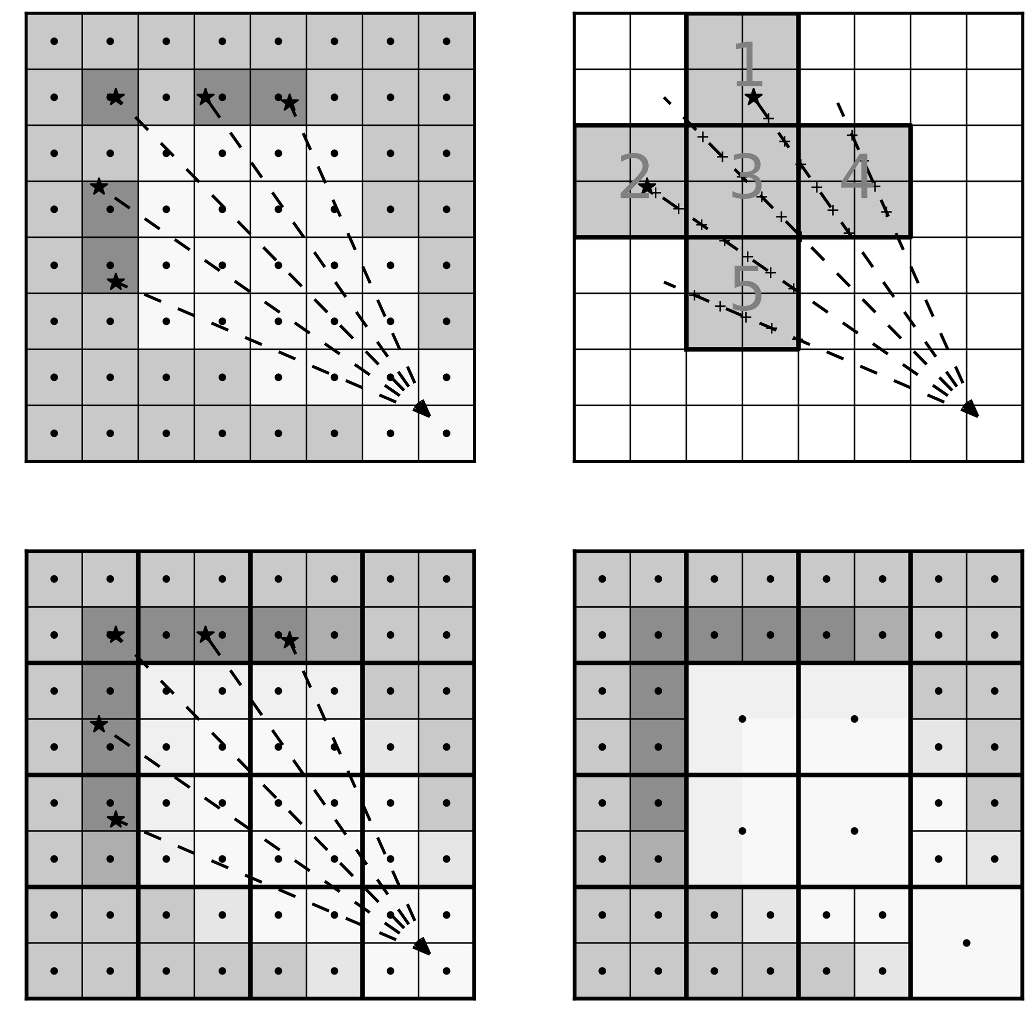
To the right we show the core data structure involved in our approach to occupancy maps: the "test data octree" from [?], which allows us to prune our environment representation to a lower resolution.

Top-left: Standard occupancy grid mapping.

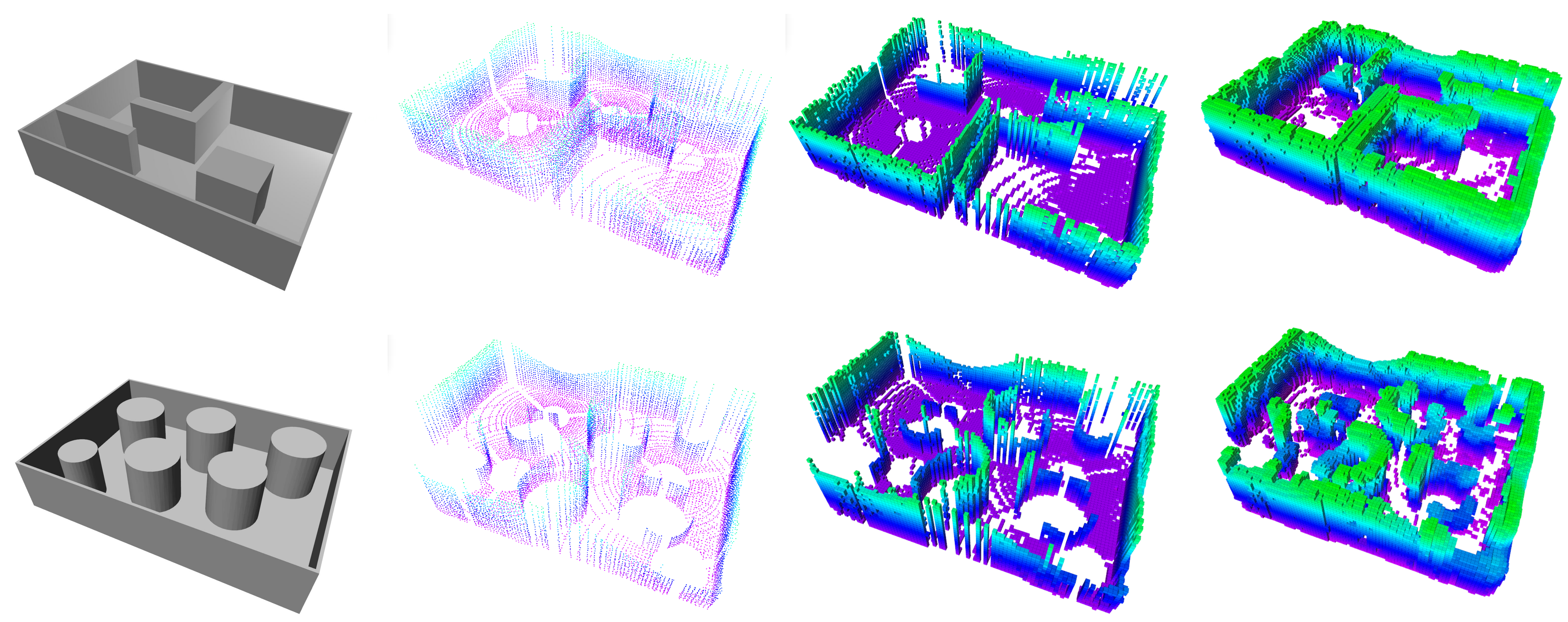
Top-right: Setup for prediction of the occupancy probability of all cells in block 3. The extended block consists of all blocks within distance l of block 3 that contain sensor data or sampled free-space points.

Bottom-left: For each block, the data from the corresponding extended block is aggregated and inference is performed.

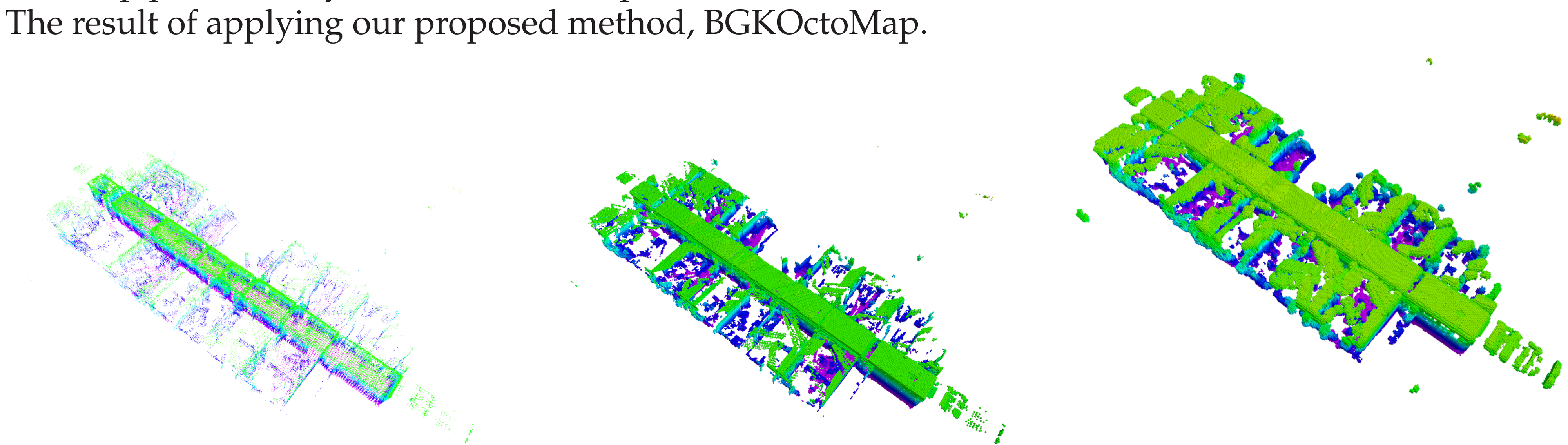
Bottom-right: Neighboring cells within a block with the same occupancy state are pruned.



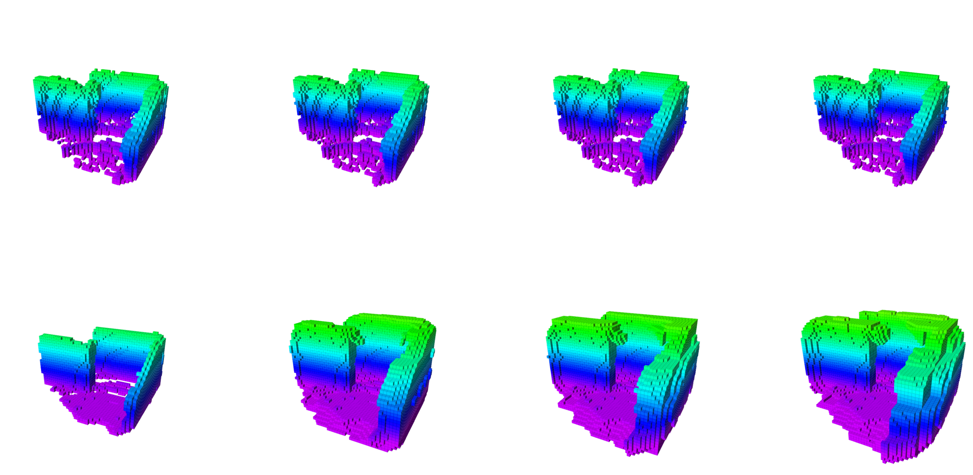
EXPERIMENTAL RESULTS



We tested our method "Bayesian Generalized Kernel OctoMap" (BGKOctoMap) in simulated *structured* (**top**) and *unstructured* (**bottom**) environments. We show (colored by height), from **left** to **right**:
1) The Gazebo simulation model for the environment
2) The simulated raw sensor data
3) The map produced by standard OctoMap
4) The result of applying our proposed method, BGKOctoMap.



We also evaluated our method qualitatively on real data from the University of Freiburg, shown above. From **left** to **right** we have:
1) The raw range-sensor data from University of Freiburg Corridor FR-079
2) The map produced by standard OctoMap
3) The map produced by BGKOctoMap



In this experiment, we simulate a robot keeping station, repeatedly scanning the same area.
Top: Our method, BGKOctoMap updated with information from 1, 15, 30, and 60 scans (**left** to **right**) containing the same data.
Bottom: The result after applying the online Gaussian process occupancy mapping method in [?] to the same data.
Qualitatively, our method (**top**) is more stable for long-term mapping scenarios with many repeat observations than the previous method (**bottom**) in which the occupied voxels tend to grow continuously with repeat observations.

We quantitatively evaluated the inference model in both simulated environments. On the **left** we show the receiver operating characteristic (ROC) curve of the classifier for the *structured* map, and for the *unstructured* map on the **right**. Each point on the curve shows the false-positive rate and true-positive rate at a particular threshold separating the positive class "occupied" and the negative class "unoccupied" (up and to the left is better). The area under the curve measures the overall accuracy of the classifier. We compare our method to GPOctomap-NBCM-P presented in [?] and standard OctoMap.