# Supplementary Material for Learning with Noisy Labels for Robust Point Cloud Segmentation

Anonymous ICCV submission

Paper ID 6600

## 1. Overview

In this supplemental material, we first show the performance comparison of our proposed PNAL with different methods on artificially created asymmetric noisy S3DIS, by following the setting in image noisy label learning works. Then, we show our statistics for ScanNetV2 **validation set** to get a deeper understanding of the noise level of real-world noisy datasets. Finally, we show more visualization results on the ScanNetV2 **validation set** and provide insights for future studies.

## 2. Performance Comparison on Asymmetric Noise

Consistent with our observations on real-world noisy dataset ScanNetV2, we designed the "*asymmetric noise*" as a combination of asymmetric and symmetric noise. The presence of asymmetric noise is due to the similarity among the confusing categories. And symmetric noise exists because of the wrong category selection during the labeling process, i. e. Fig. 1 in the paper, where the floor is labeled as a chair. Thus, the annotator introduces noise not only in the confusing classes, but also in the other classes, with random annotation errors.

Different from the "*asymmetric noise*" ($\tau_{pair} = 40\%, \tau = 60\%$) setting in the main manuscript, which is a combination of asymmetric and symmetric noise, here we use **only** asymmetric noise following the similar setting in the image noisy label learning works. In other words, we only randomly flip the label with a probability $\tau$ between our identified easily misclassified label pairs, including door-wall, board-window, and sofa-chair, while the labels of the other classes remain unchanged.

| Methods | Asymmetric Noise ($\tau$) | |
|---|---|---|
| | 20% | 40% |
| DGCNN[5]+CE | 0.8288 | 0.7513 |
| DGCNN[5]+SCE[4] | 0.7593 | 0.7308 |
| DGCNN[5]+GCE[6] | 0.7024 | 0.7006 |
| DGCNN[5]+SELFIE[3]* | 0.8565 | 0.8095 |
| DGCNN[5]+PNAL | **0.8635** | **0.8498** |
| PointNet2[2]+CE | 0.8301 | 0.7442 |
| PointNet2[2]+PNAL | **0.8670** | **0.8321** |

Table 1. OA comparison of different methods on artificially created noisy S3DIS with asymmetric noise following the setting in previous works on image.

As shown in Table 1, we report the results in terms of overall accuracy (OA). The first five rows show the comparison with DGCNN backbone. The performance of DGCNN+CE drops by $0.0404$ at $20\%$ asymmetric noise rate and $0.1179$ at $40\%$ compared to the baseline that trains on purely clean samples ($0.8692$). The decline is not as dramatic as with symmetric noise of Table 1 in the main manuscript, because noises in this traditional asymmetric setting are only present in the easily confusing classes, which are a small percentage of the whole data. As shown in the second and third rows, the previous noise robust loss methods SCE and GCE fail to work and even produce worse results. This is expected as they cannot cope with potentially large noise rates. Though "DGCNN+SELFIE*" shows a significant improvement over the CE results, the proposed PNAL can produce much better results than SELFIE. We attribute this to the noise-blind design and local region correlation modeling. When using PointNet2 as backbone as shown in the last two rows, our PNAL also improves the performance significantly.

## 3. Statistics of Noisy Label on the ScanNetV2 dataset

| Dataset | # Scenes | # Noisy Labeled Scenes | Ratio |
|---|---|---|---|
| ScanNetV2 | 312 | 228 | 73.3% |

Table 2. Statistics of Noisy Label on Real-World Noisy ScanNetV2 validation set.

For a better understanding of the noise issue on the ScanNetv2 dataset, we leverage manual inspection to check the whole ScanNetV2 validation set. For each scene, there were at least two human inspectors and the scene is counted as mislabeled when they both agreed that noisy class labels exist in the scene. As shown in Table 2, 228 scenes out of 312 (73.3%) suffer from label noise issue. Therefore, even though ScanNetV2 is already a re-labeled version of ScanNet, the label noise issue is widely present in its validation set. This further proves the existence and significance of the noisy labeling problem and the urgency of the need for a noise-robust framework.

## 4. Performance on ScanNetV2

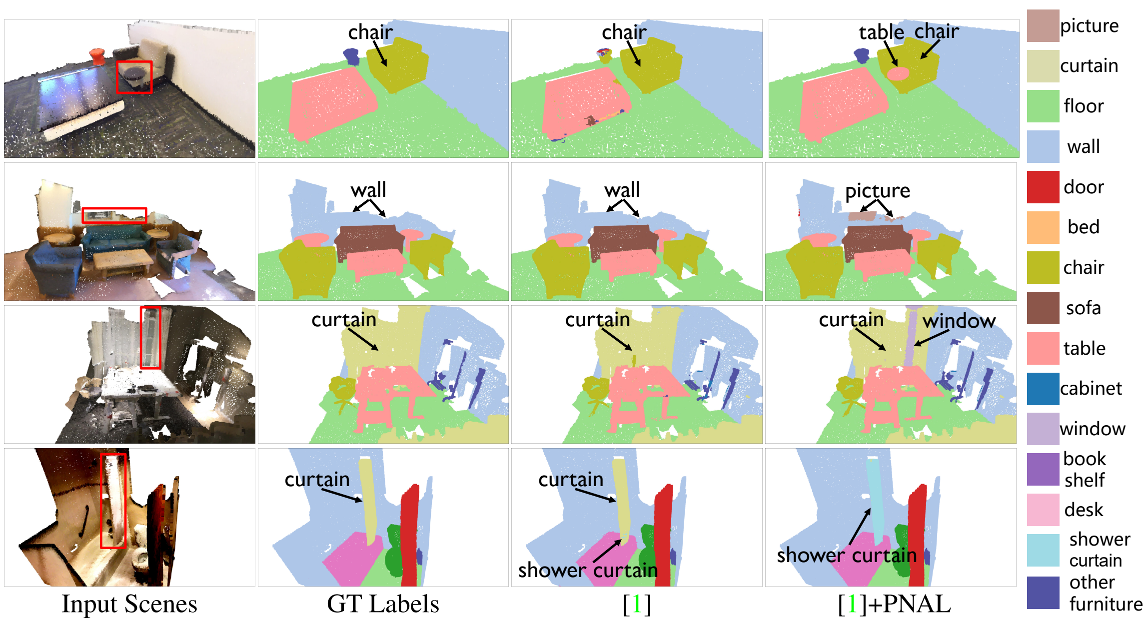


Figure 1. From left to right: Scenes in ScanNetV2 **validation set**, GT labels given by ScanNetV2, predictions of SparseConvNet [1], and predictions of [1]+PNAL. Ours get more reasonable labels than GT labels.

We show more scene visualization examples from the **validation set** of ScanNetV2 in the Fig. 1, where our method gets more reasonable results than the baseline method and even than GT. These visualization results further illustrate the existence and complexity of real-world label noises. Specifically, the GT labeling error in the second and fourth rows are from confusing category pairs "photos *vs* walls" and "shower curtains *vs* curtains",which fit the asymmetric noise pattern. For the first and third row, unlike the aforementioned noise types, their labeling error pattern are sample-related, called instance- and label-dependent noises. Owing to its complexity, modeling this kind of noise has not been extensively investigated yet. In the future, studying more diverse label noises in point cloud segmentation would be a great research direction.

## References

[1] Benjamin Graham, Martin Engelcke, and Laurens van der Maaten. 3d semantic segmentation with submanifold sparse convolutional networks. *CVPR*, 2018. 2

[2] Charles R Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *arXiv preprint arXiv:1706.02413*, 2017. 1

[3] Hwanjun Song, Minseok Kim, and Jae-Gil Lee. Selfie: Refurbishing unclean samples for robust deep learning. In *International Conference on Machine Learning*, pages 5907–5915. PMLR, 2019. 1

[4] Yisen Wang, Xingjun Ma, Zaiyi Chen, Yuan Luo, Jinfeng Yi, and James Bailey. Symmetric cross entropy for robust learning with noisy labels. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019. 1

[5] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E. Sarma, Michael M. Bronstein, and Justin M. Solomon. Dynamic graph cnn for learning on point clouds. *ACM Transactions on Graphics (TOG)*, 2019. 1

[6] Zhilu Zhang and Mert R Sabuncu. Generalized cross entropy loss for training deep neural networks with noisy labels. In *NeurIPS*, 2018. 1