Supplementary Material For PointDC

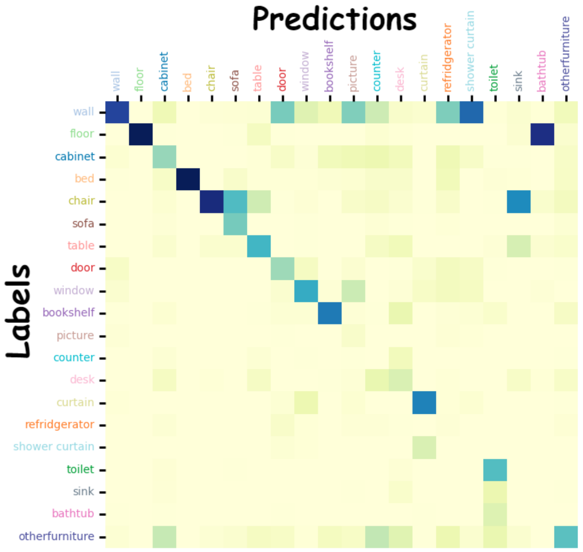
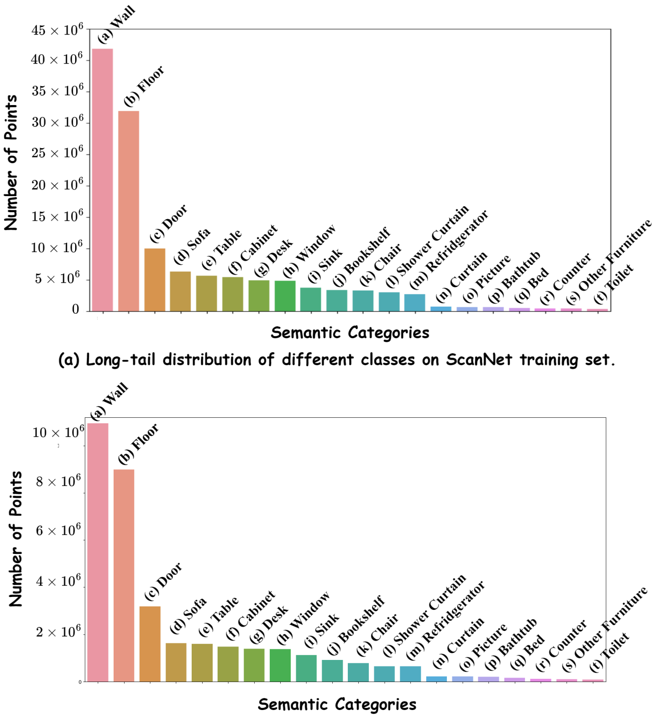


Figure 1. Confusion matrix of the unsupervised clustering predictions on ScanNet-v2 validation set.



(a) Long-tail distribution of different classes on ScanNet training set.

(b) Long-tail distribution of different classes on ScanNet validation set.

Figure 2. Long-tail distribution of all semantic categories in ScanNet-v2.

1. Failure Analysis with Confusion Matrix

To better understand the predictions and failures of our PointDC framework, we visualize the confusion matrices of ScanNet-v2 validation set in Fig. 1. We also provide a visualization of the number of points on each semantic categories to show the long-tail distribution of ScanNet-v2 for further discussion in Fig. 2. As shown in the figure, some salient failures of PointDC are listed as follows:

- The classes of ‘Door‘, ‘Picture‘, ‘Refridgerator‘ and ‘Shower curtain‘ are clustered to the class of ‘wall‘. These classes are usually placed around or inside the class of ‘wall‘. Furthermore, the imbalanced distribution of occupied points of these classes further confuse the clustering process, as shown in Fig. 2. Because the points of these classes are much less than the class of ‘wall‘, which means that their contributions to the cluster are much more likely to be ignored during clustering.
- The class of ‘Bathtub‘ is assigned to the class of ‘Floor‘. Similar to previous case, the imbalanced distribution of occupied points between these classes lead the clustering algorithm to ignore the class of ‘bathtub‘ that only has very few points compared with ‘floor‘.
- The classes of ‘Chair‘(ignoring the back of chair), ‘Sofa‘, ‘Table‘ and ‘Sink‘(watching it upside down) are sometimes confused by the framework because they share a similar 3D structure and are usually placed nearby in the indoor scene. To distinguish the difference among these clusters requires further fine-grained distance metrics to separate these classes indenpendently from the remaining classes in the dataset.

In these circumstances, due to fact that the metric of unsupervised clustering suffers from the long-tailed distribution of different semantic classes, it is hard to imagine a system that can segment these minor or relatively fine-grained categories without any additional information. This expounds the failure cases of our PointDC framework on these minor classes.

Though we firstly investigate the fully-unsupervised 3D

Models	mIoU	Acc
[4] PC-HC	9.80	42.95
[4] PC-NCE	9.08	36.74
[3] CSC	13.13	57.29
[2] PiCIE	7.05	38.03
[1] Crosspoint	3.78	25.34
PointDC	**23.85**	**65.57**

Table 1. Comparison of unsupervised segmentation results based on linear probe on the S3DIS validation set (Area 5).

semantic segmentation problem and achieve significant improvements compared with existing unsupervised systems in this manuscript, there is still a measurable gap of performance compared with fully-supervised systems. We hope that this work can provide a startup for future progresses in unsupervised 3D segmentation.

2. Linear Probe on S3DIS

In Tab. 1, we present the linear probe results of unsupervised 3D semantic segmentations methods on S3DIS validation set. The comparison includes PiCIE [2], PointContrast [4], Crosspoint [1] and CSC [3]. As shown in the table, PointDC achieves an improvement of **+10.72** linear probe mIoU and **+8.23** linear probe Accuracy.

3. Limitations

In this first attempt for learning 3D sementatic segmentation without human annotations, there are still a measurable gap between unsupervised and fully-supervised 3D semantic segmentation methods in spite of the current improvement of PointDC compared with previous unsupervised methods. As the failure analysis with confusion matrix shows, the extremely unbalanced long-tail distribution of the number of points in different categories dramatically confuse the clustering metric during training. How to handle these minor classes with few points during clustering might be a future direction for improving the performance.

References

[1] Mohamed Afham, Isuru Dissanayake, Dinithi Dissanayake, Amaya Dharmasiri, Kanchana Thilakarathna, and Ranga Rodrigo. Crosspoint: Self-supervised cross-modal contrastive learning for 3d point cloud understanding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9902–9912, 2022.

[2] Jang Hyun Cho, Utkarsh Mall, Kavita Bala, and Bharath Hariharan. Picie: Unsupervised semantic segmentation using invariance and equivariance in clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16794–16804, 2021.

[3] Ji Hou, Benjamin Graham, Matthias Nießner, and Saining Xie. Exploring data-efficient 3d scene understanding with contrastive scene contexts. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15587–15597, 2021.

[4] Saining Xie, Jiatao Gu, Demi Guo, Charles R Qi, Leonidas Guibas, and Or Litany. Pointcontrast: Unsupervised pretraining for 3d point cloud understanding. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*, pages 574–591. Springer, 2020.