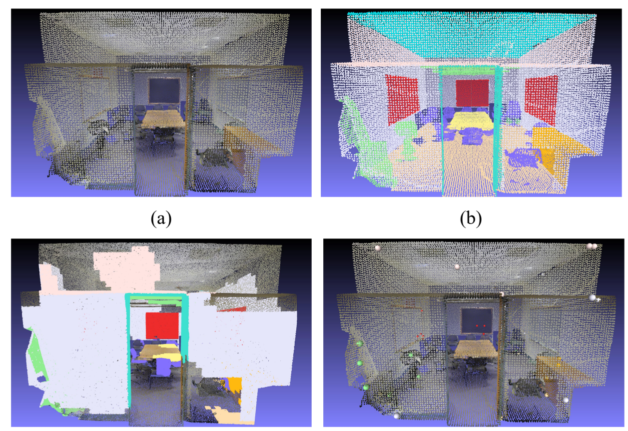
Hierarchical Point-based Active Learning for Semi-supervised Point Cloud Semantic Segmentation
– Supplementary Material –

1. Introduction

In this supplementary material, we provide the following contents:
1. The comparison of manual labelling efforts for different learning strategies is given in Sec. 2;
2. Quantitative segmentation results for each category are shown in Sec. 3;
3. More ablation studies are shown in Sec. 4;
4. Visualization of selected points in each iteration is presented in Sec. 5;
5. More qualitative segmentation results are given in Sec. 6;
6. The pseudocode of our framework is illustrated in Sec. 7.

2. The Comparison of Manual Annotation Efforts for Different Learning Strategies

As shown in the comparison in Figure 1, the full-supervised methods and region-based active learning strategy require a large amount of annotation labour, while the proposed hierarchical point-based active learning approach can significantly decrease required annotations by adopting HMMU and FDS modules to select scarce but important points.



(a) (b) (c) (d)

Figure 1. The comparisons of manual annotation efforts (coloured regions or points) by different learning strategies on the S3DIS dataset. **(a)** the original unlabelled point cloud. **(b)** the traditional fully supervised methods (e. g. MinkowskiNet) require labelling all the 3D points for the entire dataset, which is extremely expensive. **(c)** the region-based active learning method (e. g. ReDAL) needs to label a large portion of regions, inevitably causing redundant annotations. **(d)** the proposed point-based active learning strategy only requires very sparse labelling of the selected points.

3. Quantitative Segmentation Results for Each Category

We show the segmentation results of each category under different labelling budgets in Table 1 and Table 2. When our method achieves comparable segmentation results with the fully supervised baseline based on limited labelled data, we only use 0.43% and 0.1% labelled data for S3DIS and ScanNetV2 datasets respectively. Our method substantially outperforms the fully supervised baseline on categories such as column, table, bookcase, board and sofa in S3DIS and counter, curtain, sofa, table and other furniture in ScanNetV2. Moreover, our method can keep the segmentation performance without introducing noise for the other categories which can be correctly segmented even with the least labelling budget (0.02% and 20pts in S3DIS and ScanNet respectively).

4. More Ablation Studies

In this section, we show more ablation studies on HMMU and FDS modules. As shown in Table 3, we show the segmentation performance on each category with different combinations of modules. Compared with random labelling, the HMMU module pays more attention to those categories with poor segmentation results, such as column and sofa, and assigns more labelling efforts to them, resulting in great improvements in these two categories without compromising the segmentation quality of other categories.

For some challenging classes, even though more points are selected, it is hard to achieve further improvement. If we only rely on the HMMU module, we would select excessive points for these categories without further gain in

Methods	settings	mIou(%)	ceiling	floor	wall	beam	column	window	door	chair	table	bookcase	sofa	board	clutter
MinkowskiNet†	100%	64.5	92.6	96.8	81.4	0.0	23.0	44.1	79.2	88.6	76.7	71.5	52.3	77.4	56.3
Ours	0.02%	55.9	89.1	94.2	77.4	0.0	0.6	39.6	65.7	82.1	66.8	65.0	40.3	61.8	45.3
Ours	0.07%	62.3	90.6	94.6	79.1	0.0	23.8	36.7	74.1	85.1	73.6	69.8	55.6	75.1	51.4
Ours	0.43%	**65.7**	91.3	96.6	80.8	0.0	**34.5**	37.2	77.0	86.9	**78.0**	**72.6**	**64.1**	**81.0**	54.3

Table 1. Quantitative results on Area-5 of S3DIS for each category with a different number of labelling points.

Methods	settings	mIou(%)	bathtub	bed	bookshelf	cabinet	chair	counter	curtain	desk	door	floor	other furniture	picture	refrigerator	shower curtain	sink	sofa	table	toilet	wall	window
MinkowskiNet†	100%	68.0	85.3	77.3	80.7	67.5	84.7	38.8	78.1	54.7	54.5	95.2	46.8	17.4	66.5	82.7	65.6	75.1	57.5	88.1	81.9	61.9
Ours	20pts	62.5	81.3	68.6	70.2	56.4	81.6	46.2	69.7	53.9	41.8	94.2	40.3	0.6	57.7	75.0	65.3	69.0	55.9	88.4	78.3	55.4
Ours	0.1%	68.2	**91.0**	75.2	69.1	67.3	85.5	48.0	73.6	52.3	54.5	94.9	48.4	**23.6**	56.1	81.8	**73.9**	78.1	58.9	**90.9**	82.0	58.4
Ours	120pts	**69.4**	87.2	75.7	71.7	66.8	84.4	**49.1**	82.1	55.3	54.1	95.0	**49.5**	22.5	60.6	84.2	73.2	**80.6**	**63.0**	90.3	81.9	61.1

Table 2. Quantitative results on the test set of ScanNetV2 for each category with a different number of labelling points.

Algorithm 1: Pseudocode of Our Framework

```
Input: Labelled Dataset {X^l, Y^l}, Unlabelled
       Dataset {X^u}, Student Network M_s,
       Teacher Network M_t, Maximum Active
       Iteration Number I, Annotation Budget B,
       Training Steps K
Output: Student Model M_s
for all k = 1,...,K do
    Update the parameters of Student M_s;
    Update the parameters of Teacher M_t
end
for all i = 1,...,I do
    for x^u ∈ X^u do
        Calculate the uncertainty score v^u with
          HMMU module using M_s;
    end
    Sort points based on uncertainty scores.
    Select B/I points { X^p } by FDS module and
      labelling them with label {Y^p};
    Update labelled set {X^l, Y^l} = {X^l, Y^l} ∪ {
      X^p, Y^p};
    Update unlabelled set {X^u} = {X^u} / { X^p}
    for all k = 1,...,K do
        Update the parameters of Student M_s;
        Update the parameters of Teacher M_t
    end
end
```

performance. By introducing FDS, we can suppress redundant points for these categories and replace them with points representing other categories that can be further promoted, such as column, chair, sofa and bookcase. In this way, the overall performance is improved. Furthermore, by exploiting the TS module, large amounts of unlabelled data can also provide supervision signals for the model training and thus bring further improvements. However, the simple pseudo-label generation strategy does not address the class imbalance issue which might cause performance degradation in some categories, like column and sofa. In the future, we can utilize more advanced semi-supervised methods to better mine the useful information from the unlabelled data.

5. Visualization of Selected Points in Each Iteration

In Figure 2, we show the points selected with our proposed HMMU and FDS modules in each iteration. Please note that we only show the selected points from iterations 2 to 5, as points in the first iteration are chosen randomly. We can see that our selected points are distributed across the space and all categories with a decent proportion (points with different colours). This demonstrates that the FDS module plays a strong role in selecting those representative points and optimising the cost ratio between annotation quality and manual efforts.

6. Visualization of More Segmentation Results

More segmentation results of the proposed method on S3DIS and ScanNetV2 datasets are shown in Figure 3 and Figure 4. Compared to the fully supervised baseline, we can achieve comparable segmentation results and for some categories, the proposed method is able to segment point clouds more accurately.

7. Pseudocode of Our Framework

In Algorithm 1, we show the pseudocode of our framework to better illustrate the whole process of our method.

Methods	settings	mIou(%)	ceiling	floor	wall	beam	column	window	door	chair	table	bookcase	sofa	board	clutter
MinkowskiNet†	100%	64.5	92.6	96.8	81.4	0.0	23.0	44.1	79.2	88.6	76.7	71.5	52.3	77.4	56.3
base.	0.1%	54.7	87.9	95.1	76.0	0.0	13.0	35.3	59.6	78.9	66.9	65.1	29.0	61.4	43.1
HMMU	0.1%	57.7	87.3	94.7	75.8	0.1	20.7	34.3	63.4	80.7	68.2	63.7	47.7	68.0	44.7
HMMU + DFS	0.1%	59.3	87.4	94.6	77.0	0.0	33.4	32.9	64.6	82.4	69.8	65.4	56.1	61.1	46.0
HMMU + DFS + TS	0.1%	62.3	90.6	94.6	79.1	0.0	23.8	36.7	74.1	85.1	73.6	69.8	55.6	75.1	51.4

Table 3. Quantitative results on Area-5 of S3DIS with 0.1% labelled data by using different modules.

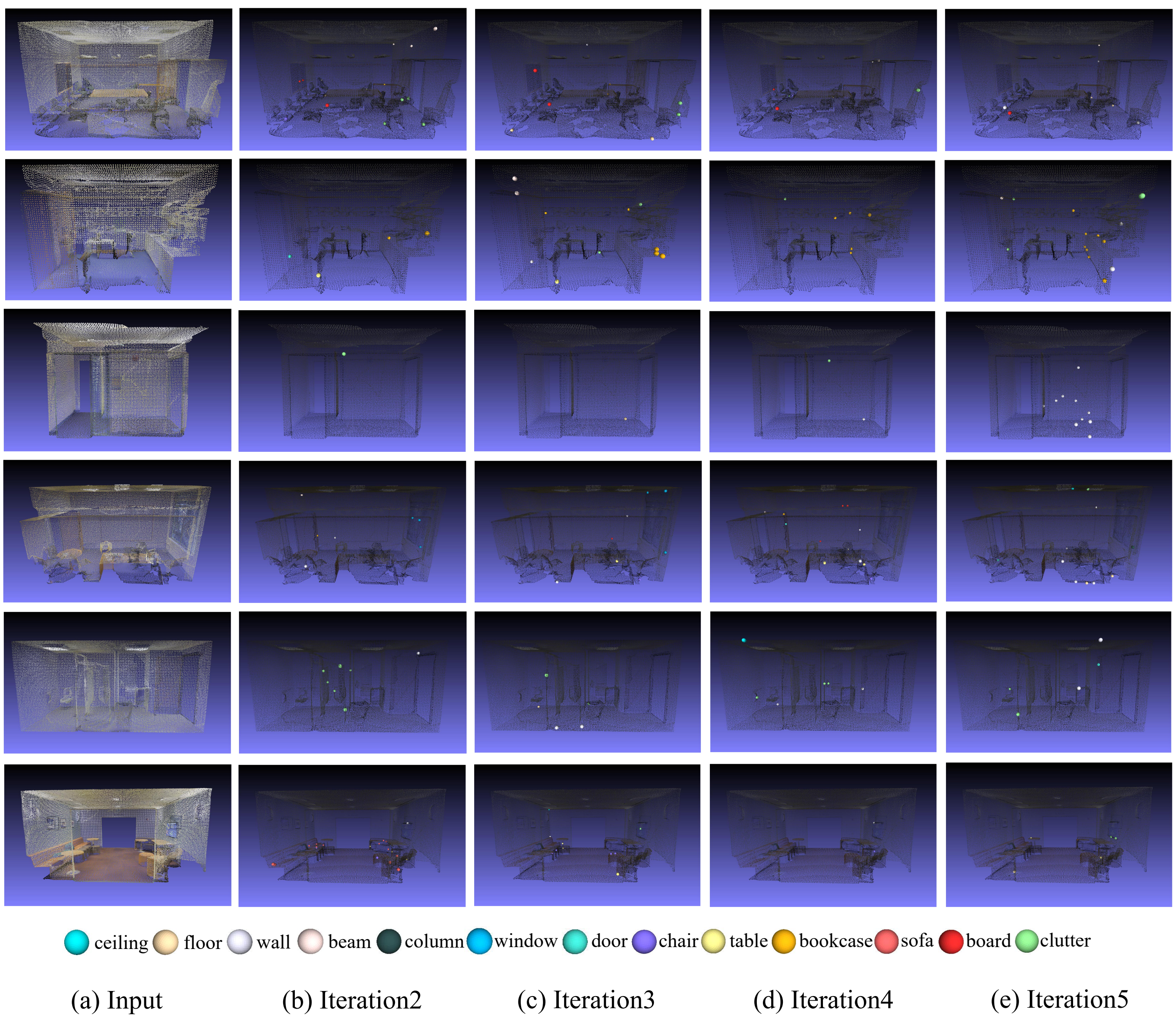


Figure 2. Visualization of labelled points selected by our method in each iteration.

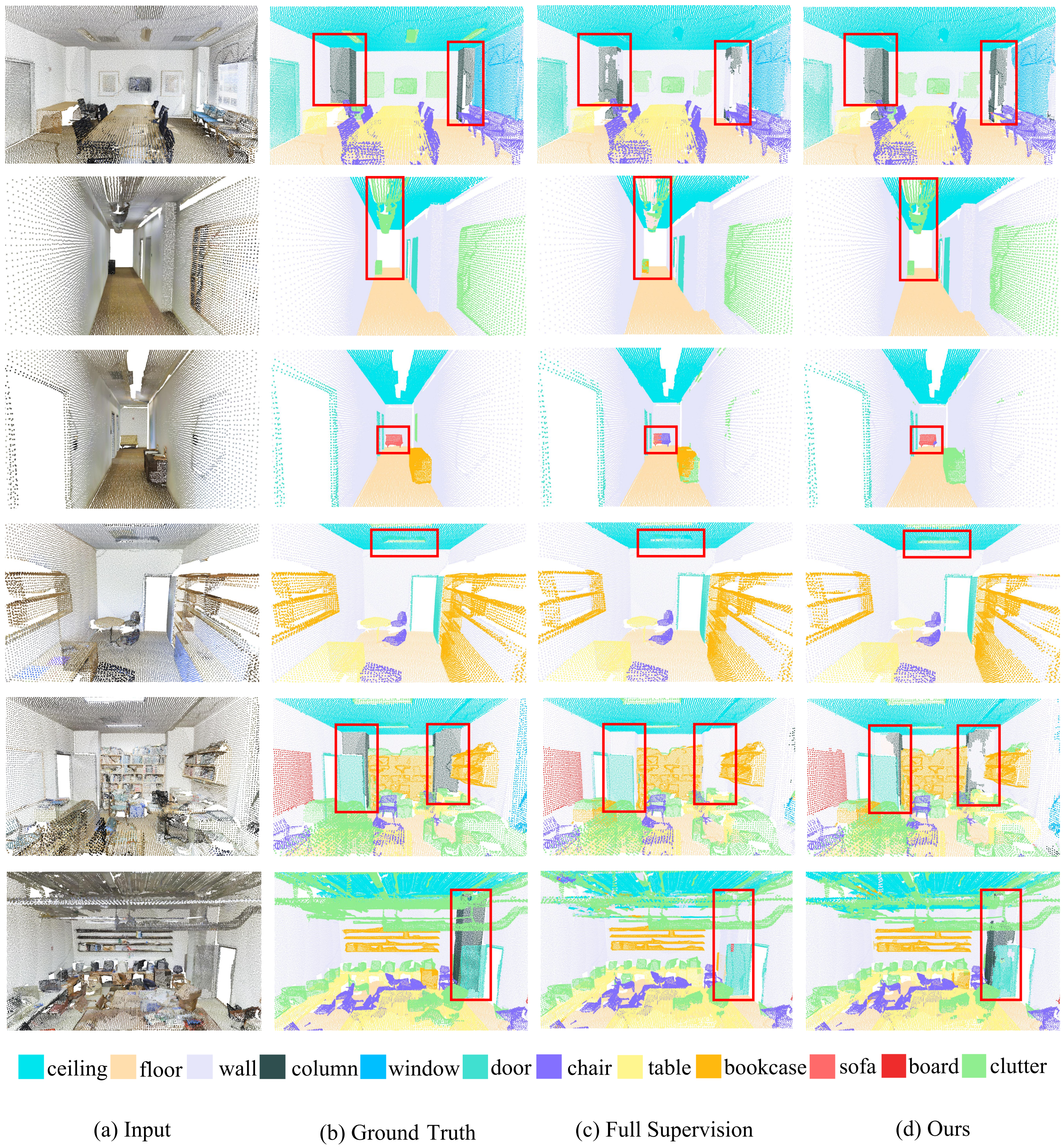


Figure 3. Visualization of segmentation results on the test set of S3DIS Area-5. Our method achieves comparable or even better results than our full-supervised baseline (MinkowskiNet).

Figure 4. Visualization of segmentation results on the validation set of ScanNetV2. Our method achieves comparable or even better results than our full-supervised baseline (MinkowskiNet).