Learning Regional Purity for Instance Segmentation on 3D Point Clouds (Supplementary Material)

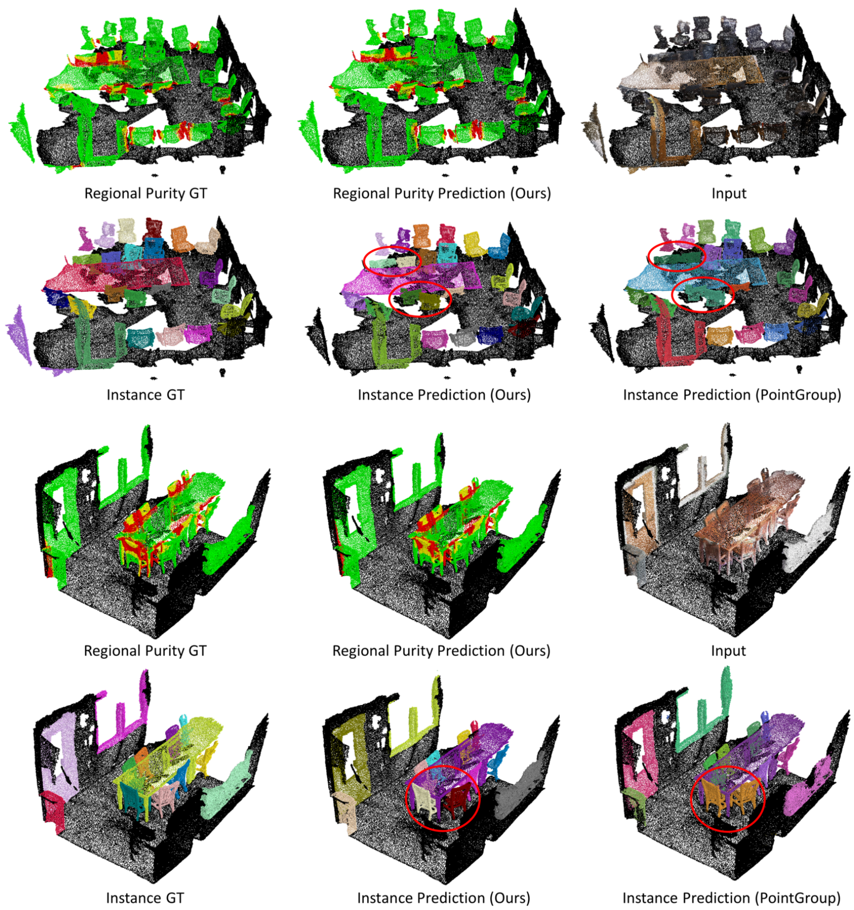

Fig. 1: More Qualitative Comparison on ScanNet v2 [2] validation set.

1 Additional Qualitative Comparisons

In this section, we show some more visualized results on ScanNet dataset [2]. As shown in Figure 1, the prediction of instance segmentation are compared with previous state-of-the-art method of PointGroup [4]. Our method shows a much better capacity of discriminating connected objects (highlighted in red circles). For regional purity signal, the predictions also have very similar patterns as ground-truth.

2 Clustering Algorithm Details

In our clustering algorithm, we introduce additional radius Δr_1, Δr_2 and cosine similarity constraint ϕ to cooperate with regional purity signal. The performance impact of using different values are shown in Table 1. In each experiment, all other parameters are fixed except the mentioned one. Parameters used for ScanNet [2] and S3DIS [1] are exactly the same. We argue that our parameter setting is generalizable and not sensitive to data distribution.

Δr_1	mAP
0.01	35.7
0.02	**35.9**
0.03	35.7
0.04	34.9

(a)

Δr_2	mAP
0.05	35.2
0.10	35.8
0.15	**35.9**
0.20	35.8

(b)

ϕ	mAP
0.6	35.8
0.7	35.8
0.8	**35.9**
0.9	35.6

(c)

Table 1: Ablation on clustering parameters on the validation set of ScanNet v2 [2]

3 Regional Purity at Different Scale

We create regional purity ground-truth on $0.2m$ radius and $0.3m$ radius respectively. To make a fair comparison, all settings are kept same except Δr_2, which is proportional to their receptive radius. The result shows that regional purity with $0.2m$ radius can be better utilized (2.5% higher than $0.3m$ scale in mAP). We argue that setting appropriate radius for regional purity can bring more value for clustering algorithm. Overlarge radius may cause regional purity to be less informative. Meanwhile, regional purity with too small radius suffers more serious class imbalance problem and the low purity area generated may not be wide enough to separate different instances.

4 Illustration of Clustering Algorithm

To better illustrate our clustering algorithm, we break the process into two steps and visualize the intermediate state. In practise, they are performed at the same time. As shown in Figure 2, the clustering algorithm takes semantic prediction and regional purity prediction as input. In step (a), it shows the shifted coordinates of foreground points. Note that, the predicted centers from points are very inaccurate in this scene. If we simply use shifted coordinates of points, it is impossible to separate objects since points from each row are almost connected as one piece. Thus, we drop out low purity points and only group high purity points and medium purity points into clusters. As visualized, high purity points are mostly concentrated and easy to be grouped. In step (b), low purity points are further grouped into clusters based on their original coordinates and direction constraint. It is important not to group low purity points on shifted coordinates, because their offset prediction is untrustworthy.

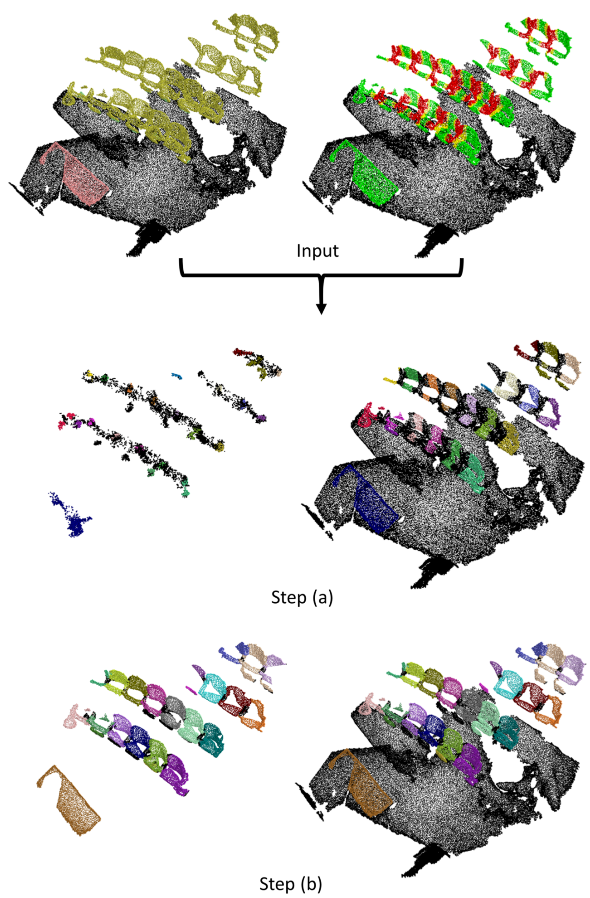

Fig. 2: Illustration of clustering algorithm. Step (a) represents the intermediate state if we only use high confidence grouping strategy for high regional purity points. Step (b) represents the final prediction.

For Algorithm 1, the descriptions in original paper are in pseudo code format, which might not be straightforward enough to understand. In this section, we further use high level explanation for better illustration.

Algorithm 1 Clustering Algorithm Explanation

```
Filter out background points
repeat
    choose an ungrouped point as initial seed point
    search all ungrouped points as target points
    for target points under same category do
        check the regional purity labels of each pair
        choose grouping strategy accordingly
        if high confidence grouping strategy then
            grouping based on shifted coordinates
            adjust reachable distance
            grouped into cluster if satisfied
        if low confidence grouping strategy then
            grouping based on original coordinates
            adjust reachable distance
            check cosine similarity
            grouped into cluster if satisfied
until No more point can be grouped
```

5 Ablation on Regional Purity Loss Function

We propose joint loss functions for dealing with unbalanced data. To analysis the effectiveness of each term in regional purity loss function, we evaluate different combinations on ScanNet v2 [2] validation set in Table 2. The precision, recall and F1 score of low purity prediction are compared. For weighted dice loss, we set $\alpha = 0.8$ and $\beta = 0.2$. Our proposed loss function achieves better trade-off between precision and recall than basic Cross-Entropy loss.

Loss Function	Precision	Recall	F1 score
CE loss	64.4	60.7	62.4
weighted dice loss	74.3	50.7	60.3
CE loss + balanced dice loss	66.6	61.7	64.1
CE loss + weighted dice loss	68.6	60.8	64.5
CE loss + weighted dice loss + distance loss	72.2	59.2	65.1

Table 2: Evaluation of Loss Functions

6 Time Complexity Analysis

In the original paper, we have reported the processing time on the full validation set of ScanNet [2]. Same as previous evaluation [3, 5], data loading time is not counted. We use a trained network for inference without label smoothing. The recorded total processing time is 83 seconds on RTX 3090 GPU. We also evaluate on RTX 2080 Ti GPU, which costs 89 seconds. Time reduction is mostly caused by following reasons: (1) our clustering algorithm only takes a single set of points as input. (2) no NMS post-processing needed (3) regional purity encourage more efficient grouping, because high purity regions can be grouped with less iterations (4) all procedures are accelerated with GPU.

References

1. Armeni, I., Sener, O., Zamir, A.R., Jiang, H., Brilakis, I., Fischer, M., Savarese, S.: 3d semantic parsing of large-scale indoor spaces. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (2016)
2. Dai, A., Chang, A.X., Savva, M., Halber, M., Funkhouser, T., Nießner, M.: Scannet: Richly-annotated 3d reconstructions of indoor scenes. In: Proc. Computer Vision and Pattern Recognition (CVPR), IEEE (2017)
3. Han, L., Zheng, T., Xu, L., Fang, L.: Occuseg: Occupancy-aware 3d instance segmentation (2020)
4. Jiang, L., Zhao, H., Shi, S., Liu, S., Fu, C.W., Jia, J.: Pointgroup: Dual-set point grouping for 3d instance segmentation (2020)
5. Yang, B., Wang, J., Clark, R., Hu, Q., Wang, S., Markham, A., Trigoni, N.: Learning object bounding boxes for 3d instance segmentation on point clouds (2019)