PolyRoom: Room-aware Transformer for Floorplan Reconstruction (Supplementary Materials)

In the supplementary material, we first provide implementation details about the loss function. Then, we provide more ablation studies to further validate our technical choices. Finally, we provide more qualitative results to show the reconstruction results.

A Loss function

The overall loss function is defined as:

$$L = \sum_{i=1}^{M} (\lambda_{cls} L_{cls}^{i} + \lambda_{coord} L_{coord}^{i} + \lambda_{ang} L_{ang}^{i} + \lambda_{ras} L_{ras}^{i}), \tag{1}$$

where L_{cls}, L_{coord}, L_{ang}, and L_{ras} are classification loss, coordinate loss, angle loss, and rasterization loss respectively. And λ_{cls}, λ_{coord}, λ_{ang}, and λ_{ras} are the corresponding loss weights.

The vertex label classification loss is defined as:

$$L_{cls}^{i} = -\frac{1}{N} \sum_{n=1}^{N} L_{ce}(l_n^i, l_n^{\sigma(i)}), \tag{2}$$

where L_{ce} indicates a binary cross-entropy.

The coordinate loss for coordinate regression is defined as:

$$L_{coord}^{i} = \frac{1}{N} \sum_{m=1}^{N} d(v_m^i, v_m^{\sigma(i)}), i \leq M_{gt}. \tag{3}$$

The angle loss calculates the angle difference between two vertices. Since the vertices are densely sampled and angles show the spatial position relationship between adjacent vertices, the introduction of angle loss is crucial for better supervision of the vertex spatial position:

$$L_{ang}^{i} = \frac{1}{N} \sum_{n=1}^{N} |ang_n^i - ang_n^{\sigma(i)}|, i \leq M_{gt}, \tag{4}$$

where ang_n^i and $ang_n^{\sigma(i)}$ are the cosine values of the vertex angle in the ground truth and matching polygon.

The rasterization loss computes the Dice loss [3] between rasterized images of ground truth and our predictions, supervising the polygon sequences from an image perspective:

$$L_{ras} = Dice(ras(R_i, R_{\sigma(i)})), i \leq M_{gt}, \tag{5}$$

where $ras(\cdot)$ means the rasterized image from room polygon sequences with a differentiable rasterizer proposed in [2].

B Additional Ablation Studies

In this section, we conduct additional ablation studies to further validate the effectiveness of PolyRoom. All experiments are performed on Structure3D [4].

B.1 Loss Function

In this section, we provide a detailed analysis of the loss function. Four types of loss functions are used to supervise the training, among which L_{cls} and L_{coord} are essential for low-level geometrical supervision of the vertex label and coordinates. L_{ras} and L_{angle} serve as higher-level topological and pair-wise supervision. Table 1 shows that the introduction of L_{ras} and L_{angle} both improve reconstruction effects and L_{angle} plays an important role in improving the angle prediction.

Table 1: Loss function analysis on Structure3D [4]. Both angle and rasterization loss improve the experimental results and using all four types of loss achieves the best reconstruction results. The best results are in bold font.

Settings				Room		Corner		Angle	
L_{cls}	L_{coord}	L_{ras}	L_{ang}	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.
✓	✓	-	-	97.7	96.3	90.8	85.0	82.0	77.0
✓	✓	✓	-	98.1	96.5	92.6	85.2	85.0	78.3
✓	✓	-	✓	97.8	96.4	91.7	85.7	85.3	79.8
✓	✓	✓	✓	**98.9**	**97.7**	**94.6**	**86.1**	**89.3**	**81.4**

Table 2: Analysis on the maximum number of polygons on Structure3D [4]. The best results are in bold font.

	Room		Corner		Angle	
Polygon	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.
15	98.8	97.3	93.9	85.7	88.4	80.9
20	**98.9**	**97.7**	**94.6**	86.1	**89.3**	81.4
25	98.7	97.4	94.3	**86.3**	89.1	**81.7**

B.2 Maximum number of polygons

In this experiment, we investigate the effects of the maximum number of polygons on the final results. To fully represent the floorplan, we have set the maximum number of polygons to exceed the maximum number of rooms in the dataset. As depicted in Table 2, the maximum number of polygons has a small impact on the reconstruction results. To achieve a balance between precision and training time, we select a polygon number of 20 as our parameter.

C Additional results

In this section, we provide more qualitative evaluations on Structured3D [4] and SceneCAD [1]. Details are shown in Figure. 1 and Figure. 2. The reconstruction results of PolyRoom outperform other competing methods in both completeness and accuracy of reconstruction. Besides, qualitative evaluations on cross-data generalization are shown in Figure. 3.

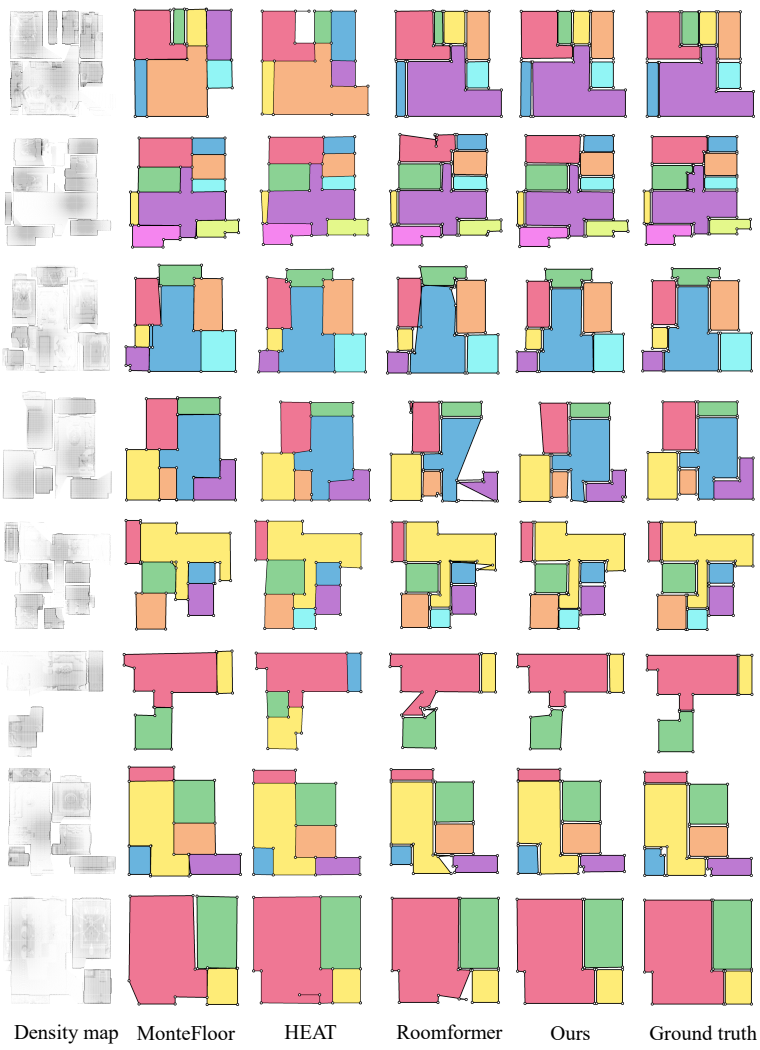

Fig. 1: Qualitative evaluations on Structured3D [4]. While competing methods suffer from incomplete rooms and missing corners, the reconstruction results of PolyRoom are more complete and accurate.

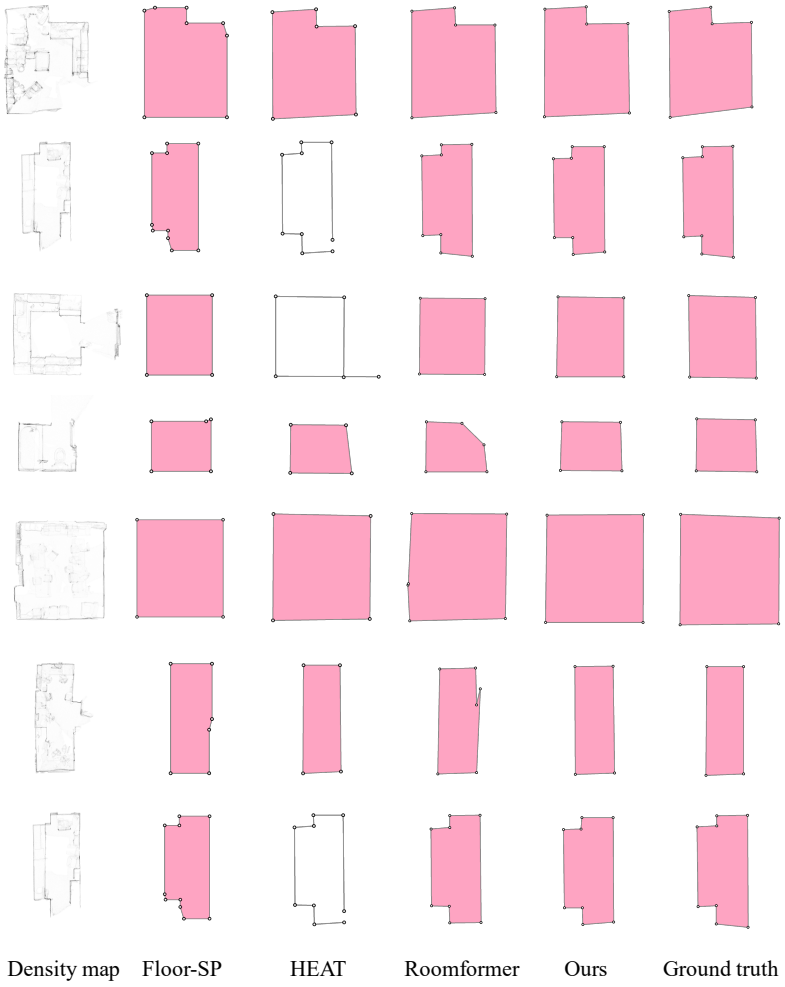

Fig. 2: Qualitative evaluations on SceneCAD [1].

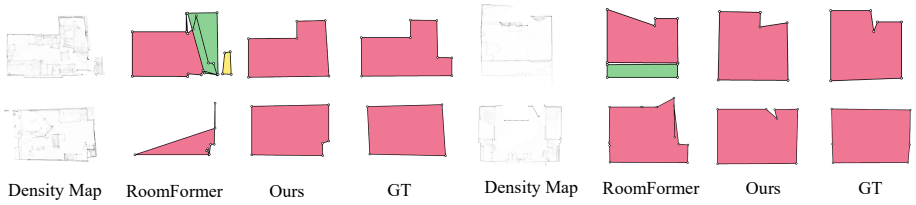

Fig. 3: Qualitative evaluations on cross-data generalization.

References

1. Avetisyan, A., Khanova, T., Choy, C., Dash, D., Dai, A., Nießner, M.: SceneCAD: Predicting object alignments and layouts in RGB-D scans. In: European Conference on Computer Vision (ECCV). pp. 596–612 (2020) 3, 4
2. Lazarow, J., Xu, W., Tu, Z.: Instance segmentation with mask-supervised polygonal boundary transformers. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4372–4381 (2022) 1
3. Milletari, F., Navab, N., Ahmadi, S.A.: V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In: International Conference on 3D Vision (3DV). pp. 565–571 (2016) 1
4. Zheng, J., Zhang, J., Li, J., Tang, R., Gao, S., Zhou, Z.: Structured3D: A large photo-realistic dataset for structured 3d modeling. In: European Conference on Computer Vision (ECCV). pp. 519–535 (2020) 2, 3