PlaneRAS: Learning Planar Primitives for 3D Plane Recovery

Supplementary Material

A. More Quantitative Results on NYUv2-Plane

We further report additional evaluation metrics on the unseen NYUv2-Plane [4] dataset in Table 4, following the evaluation protocol of PlaneRecTR [3]. For simplicity, all comparisons are conducted using a ResNet-50 [1] image backbone. Additionally, for the pretrained model, only monocular depth from Metric3D [2] is utilized as the base geometric prior of the scene.

Method	Per-Pixel/Per-Plane Recalls ↑				Plane Parameters Estimation Errors ↓	
	Depth		Normal			
	@0.10m	@0.60m	@5°	@30°	Normal (°)	Offset (mm)
PlaneTR	7.08/5.07	41.98/27.10	20.08/11.69	52.08/32.85	17.09	615.92
PlaneRecTR	**7.72/6.48**	44.44/35.70	14.43/10.56	55.99/42.24	15.98	**611.82**
PlaneRAS	4.52/3.86	**56.25/43.53**	**46.12/29.66**	**63.53/48.61**	**15.38**	621.47

Table 4. Comparison of different methods on NYUv2-Plane dataset.

As shown in Table 4, PlaneRAS achieves higher recall under larger depth thresholds. However, its performance deteriorates under more stringent depth evaluations, consistently reflecting the suboptimal depth prediction accuracy observed in Table 2 of the main text. The relatively coarse geometric estimations also lead to increased offset errors in plane parameter predictions. This may be attributed to the depth prediction paradigm adopted in the reconstruction module, as defined in Eq. (8) of the main text. Specifically, the model is predominantly constrained to learn the residual between the Metric3D [2] depth and the ground-truth depth from the ScanNet training set, which may in turn compromise its generalization to diverse real-world scenes.

Regarding surface normal evaluation, PlaneRAS exhibits significantly higher recall and lower parameter errors compared to the baseline methods. We attribute this to the fact that the fine-grained planar primitive representation facilitates more effective supervision of surface normals. The complete recall curves are provided in Figure 7.

B. More Qualitative Results

Figure 8 presents a qualitative comparison between our method and PlaneRecTR, along with frame-wise plane recovery metrics, which further highlight the performance differences between the two approaches. The first three columns in Figure 8 depict relatively simple scenes, where both methods achieve satisfactory plane recovery. PlaneRAS, in particular, produces segmentation results that tend to better preserve the underlying 3D structure of the scene, owing to its 3D-aware architecture. The last three columns in Figure 8 represent more challenging cases, in which neither method performs ideally. Nonetheless, our approach demonstrates two key advantages: (1) clearer delineation of planar boundaries, and (2) improved suppression of background interference, leading to significantly better performance in the associated metrics.

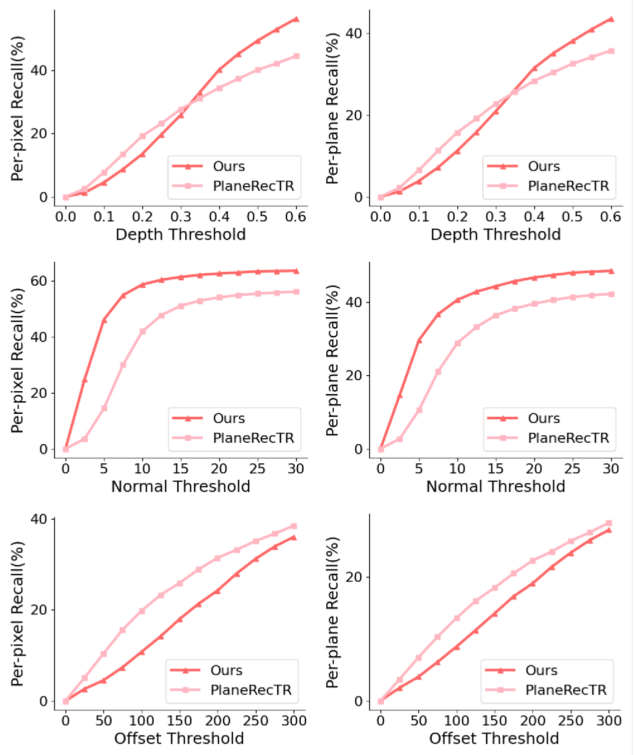

Figure 7. Per-pixel and per-plane recalls on the NYUv2-plane dataset

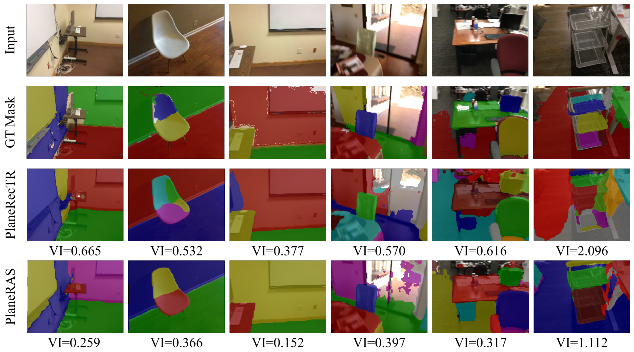

Figure 8. Additional qualitative results comparing PlaneRAS with PlaneRecTR.

However, while the model's ability to filter out background elements contributes to strong plane recovery performance on the current dataset, we observe that this behavior may negatively impact generalization. This issue will be further discussed in the following subsection.

C. Key Limitation

We present two example images from the ScanNet++ [5] dataset in Figure 9, where the depth range significantly deviates from that of typical ScanNet scenes. In these cases, PlaneRAS incorrectly interprets distant regions as background, resulting in severe failure cases. We attribute this issue primarily to the depth distribution in the ScanNet training set, which is mostly concentrated within a relatively short range of 1–2 meters or even less. Given that PlaneRAS adopts a reconstruction-aggregation prediction paradigm that models full-scene geometry, its performance is more susceptible to the limited depth statistics of the training data. Consequently, the model fails to effectively predict planar regions in distant areas. In contrast, PlaneRecTR [3], which relies on 2D pixel-level features, demonstrates greater robustness in these scenarios.

Overall, our method represents an initial attempt to extend existing frameworks toward full 3D-aware plane prediction. However, there remains substantial room for improvement. Future work will focus on enhancing generalization capability, incorporating multi-view information, and ultimately achieving more robust and broadly applicable planar reconstruction in complex 3D environments.

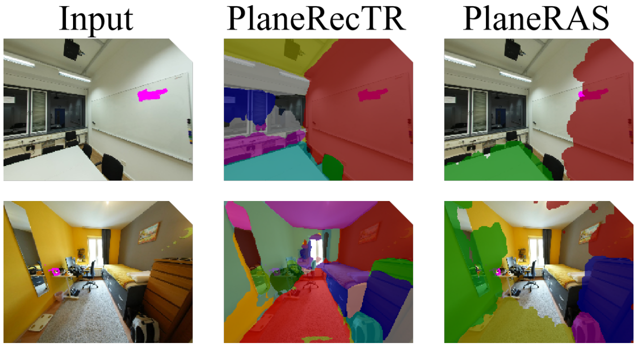

Figure 9. Catastrophic failure cases.

References

[1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016. 1

[2] Mu Hu, Wei Yin, Chi Zhang, Zhipeng Cai, Xiaoxiao Long, Hao Chen, Kaixuan Wang, Gang Yu, Chunhua Shen, and Shaojie Shen. Metric3d v2: A versatile monocular geometric foundation model for zero-shot metric depth and surface normal estimation. *TPAMI*, 2024. 1

[3] Jingjia Shi, Shuaifeng Zhi, and Kai Xu. Planerectr: Unified query learning for 3d plane recovery from a single view. In *ICCV*, pages 9377–9386, 2023. 1, 2

[4] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. Indoor segmentation and support inference from rgbd images. In *ECCV*, pages 746–760, 2012. 1

[5] Chandan Yeshwanth, Yueh-Cheng Liu, Matthias Nießner, and Angela Dai. Scannet++: A high-fidelity dataset of 3d indoor scenes. In *ICCV*, pages 12–22, 2023. 2