Supplementary Materials for PlaneTR: Structure-Guided Transformers for 3D Plane Recovery

Bin Tan*[1] Nan Xue*[1] Song Bai[2] Tianfu Wu[3] Gui-Song Xia†[1]
[1] School of Computer Science, Wuhan University
[2] ByteDance AI Lab [3] Department of ECE, NC State University

In the supplementary materials, we first illustrate some detailed architectures of our proposed PlaneTR. Then, we show more qualitative results on the ScanNet [1] and NYUv2-Plane [4] datasets.

1. Detailed Architecture

Our network mainly consists of a feature extraction backbone, a Transformer branch and a convolution branch. The feature extraction backbone is built upon the HRNet-w32 [5]. Consider that the input image size is 192×256, we modify the stride in the first convolution layer of HRNet-w32 to 1 to keep enough space resolution of the output feature maps. The convolution branch contains a Plane Embedding Decoder for plane instance segmentation and a pixel-wise Depth Decoder for depth estimation in non-plane regions. Both these two decoders have the same architecture except the last convolution layer for the output prediction as shown in Fig. 1.

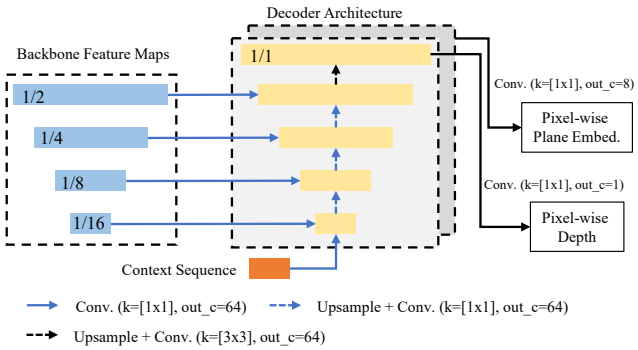


Figure 1. Architecture of the convolution decoder.

2. More Qualitative Results

Plane Instance Segmentation. In Fig. 2, we show the plane instance segmentation results of our proposed PlaneTR against PlaneNet [3], PlaneRCNN [2] and PlaneAE [7] on the ScanNet dataset [1]. Under each predicted segmentation result, we further list its segmentation metrics, including RI, VI and SC. As shown in Fig. 2, our method outperforms all other methods. Besides, more comparisons of the segmentation results on the NYUv2-Plane dataset [4] are shown in Fig. 3.

Plane Recovery. In Fig. 4, we show more piece-wise planar reconstruction results of our method on the ScanNet dataset. The input line segments are detected by recent state-of-the-art line segment detection algorithm HAWP [6] with pretrained model. Besides, the depth maps are achieved by first calculating the depth values of plane regions via the predicted 3D planar parameters and then filling out the non-plane region with predicted pixel-wise depth from the Depth Decoder. As we can see, our method can reasonably reconstruct the detected planes in the scene.

Guidance of Line Segments. In Fig. 5, we show more results about how the input line segments guide the detection of planes in our proposed PlaneTR. The setting 'w/o line segments' means that we input an empty line segment sequence into the network in inference stage. As shown in Fig. 5, with the guidance of line segments, PlaneTR can successfully detect the planes which are missed in the setting of 'w/o line segments'. It demonstrates that PlaneTR can effectively utilizes the holistic structures cues from the line segments.

*Equal Contribution.
†Correspondence Author.

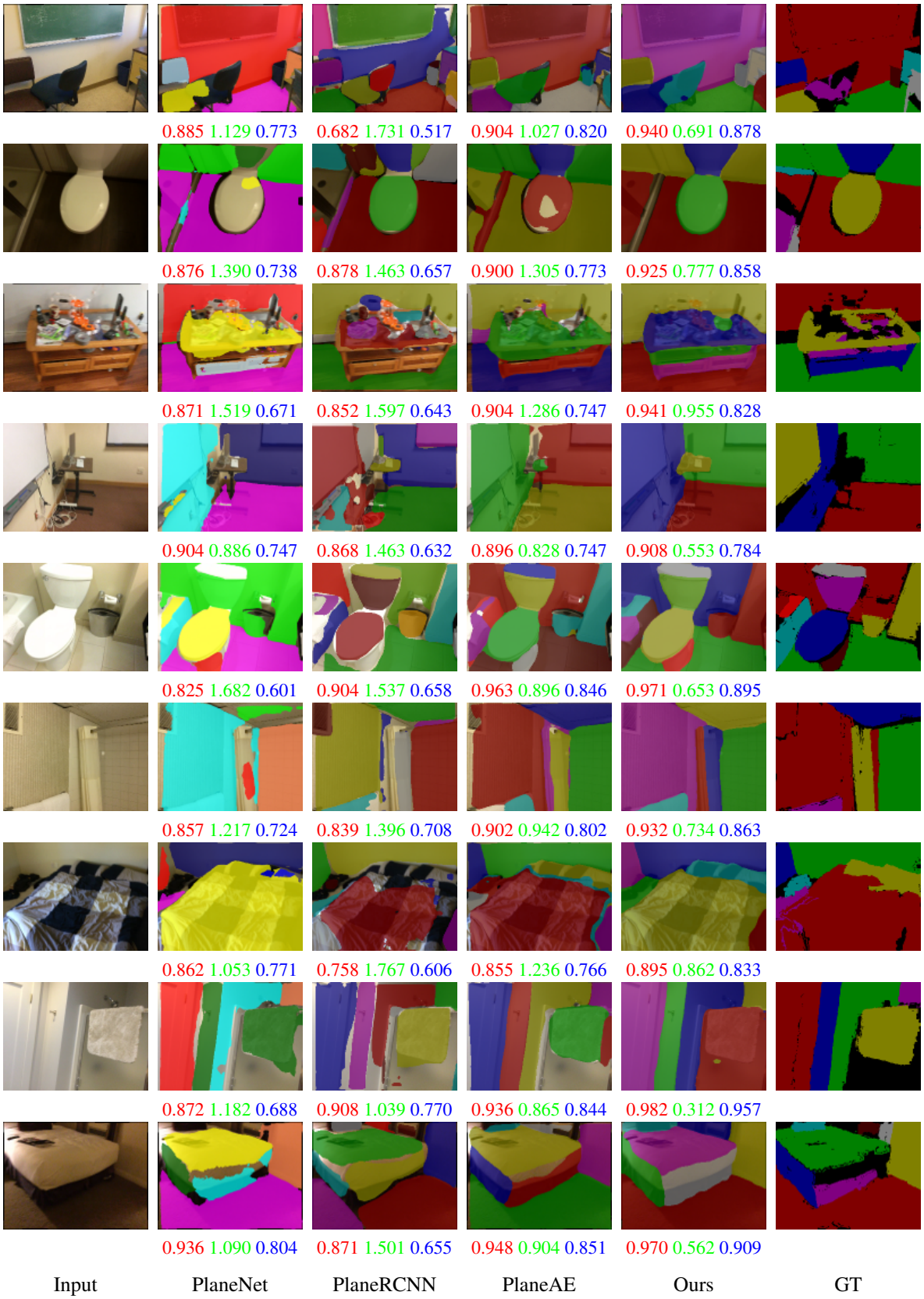
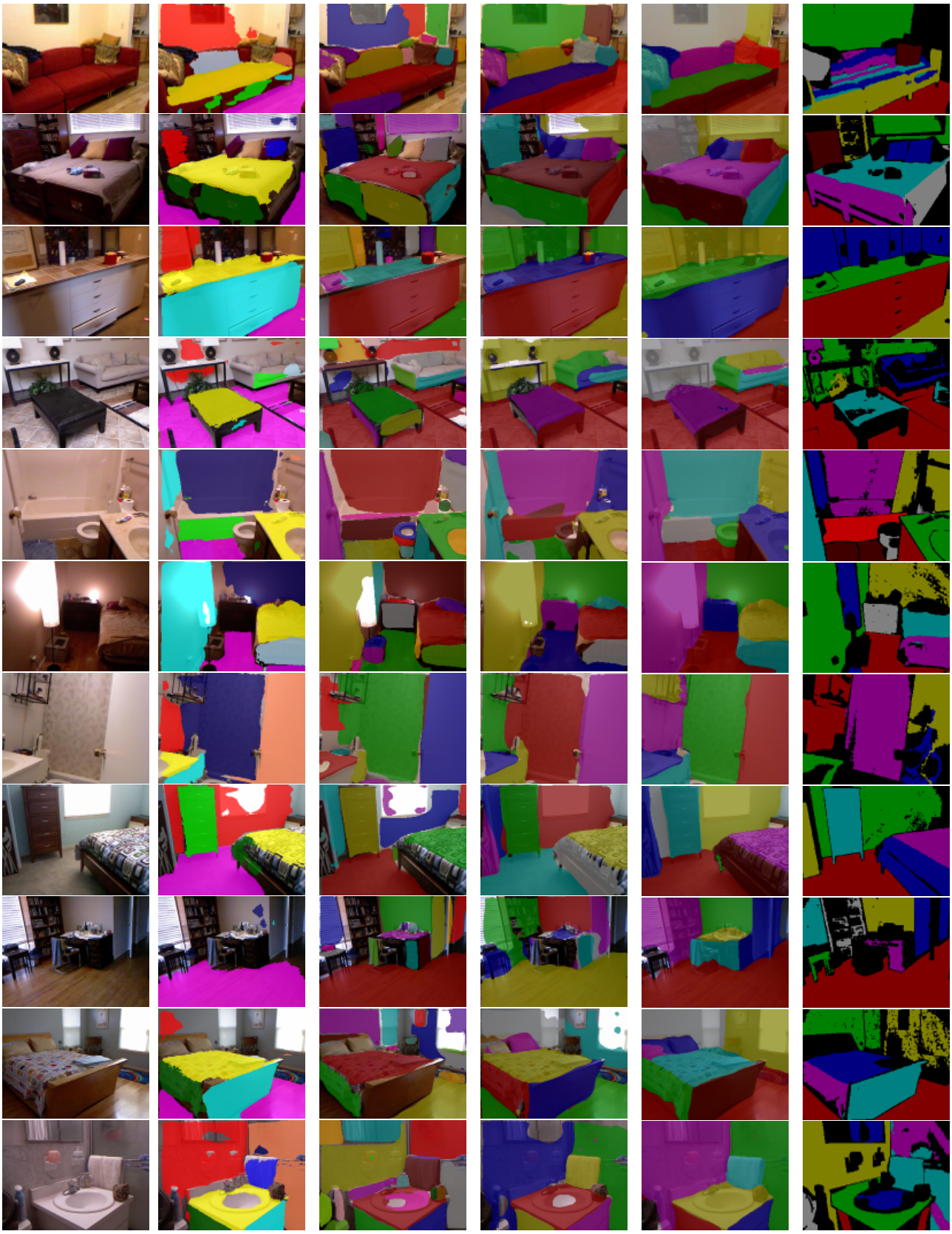
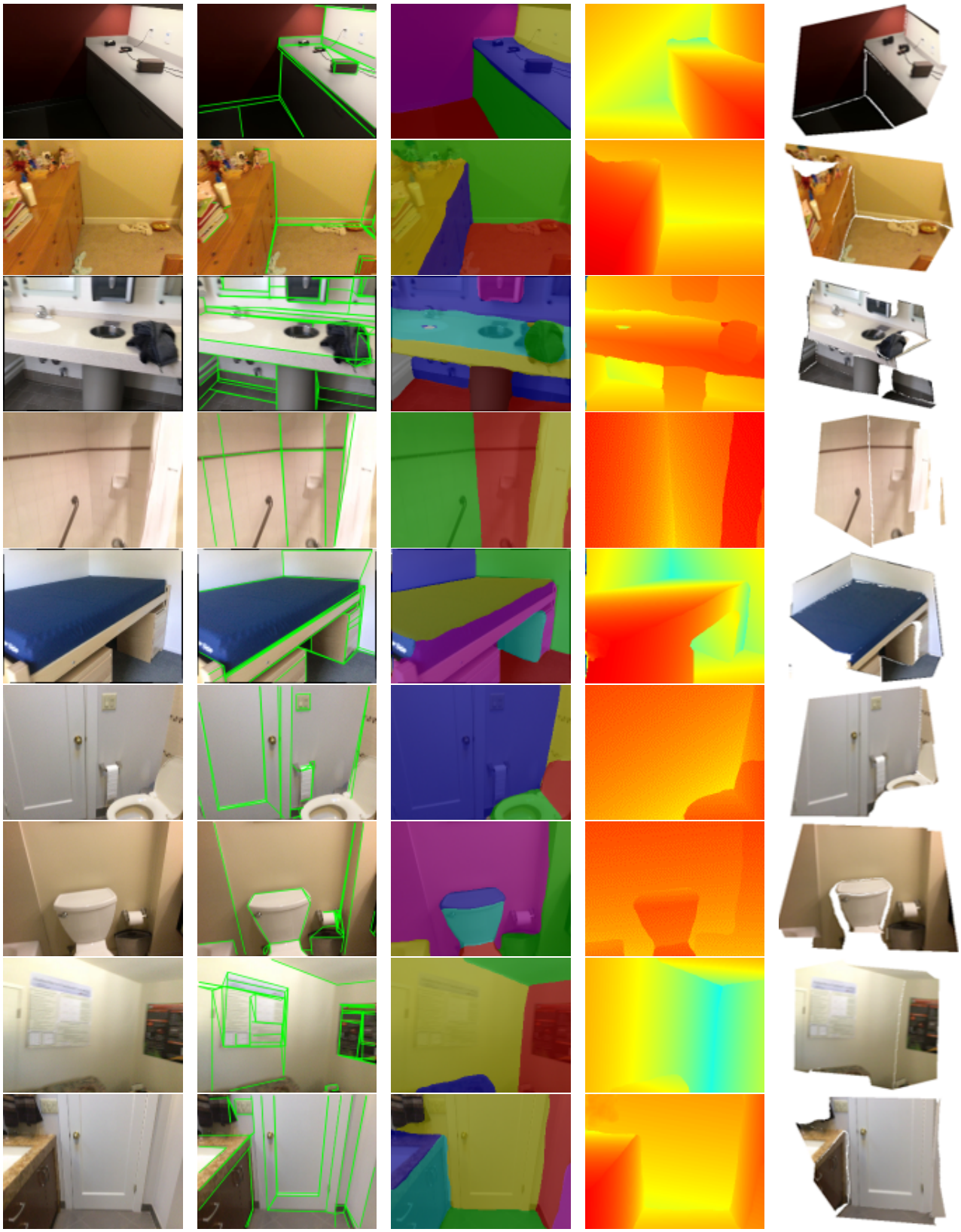


Figure 2. Comparison of plane instance segmentation results on the ScanNet dataset [1]. Red, green and blue numbers indicate RI, VI and SC, respectively.



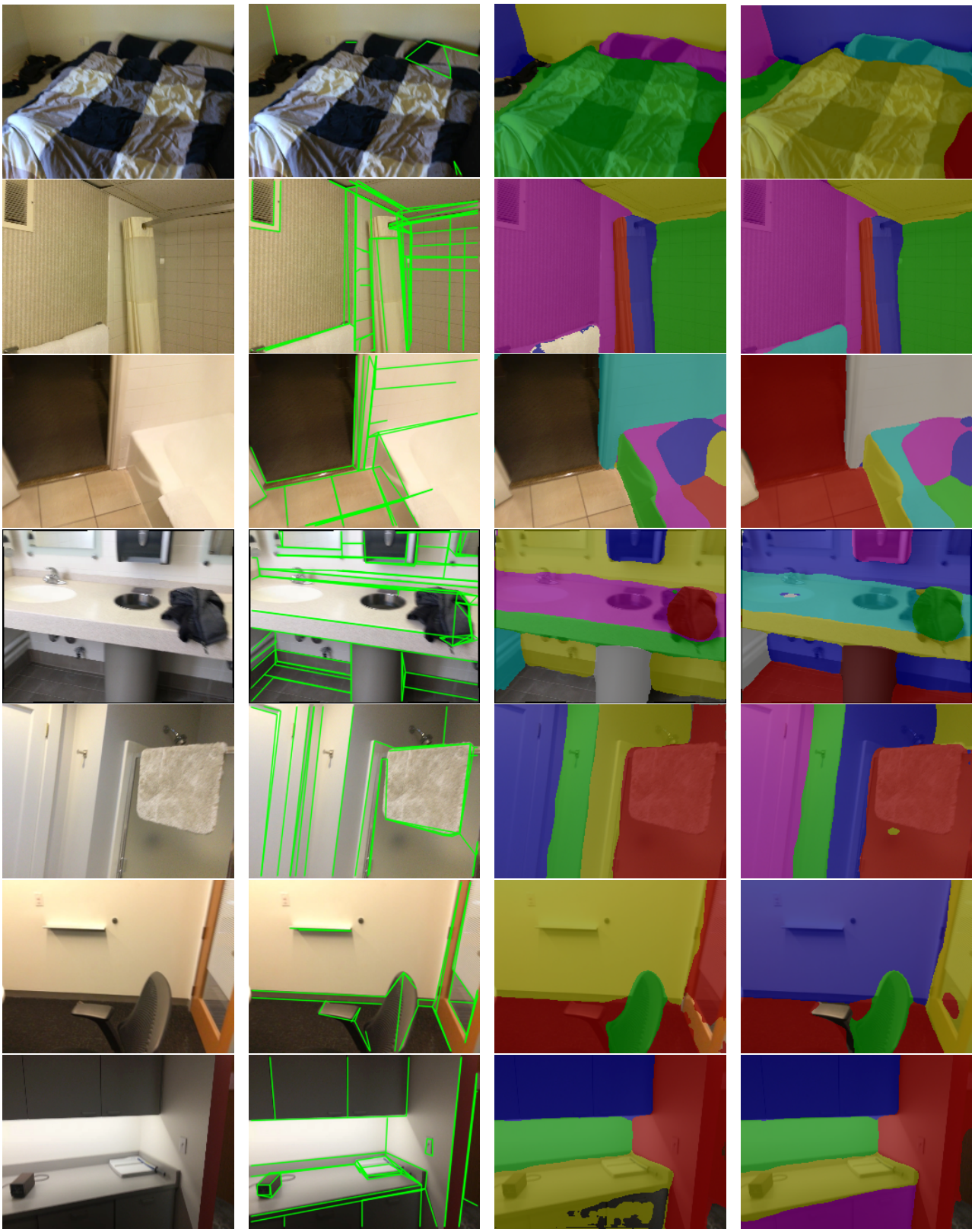
(a) Input (b) PlaneNet (c) PlaneRCNN (d) PlaneAE (e) Ours (f) Ground Truth

Figure 3. Comparison of plane instance segmentation results on the NYUv2-Plane dataset [4]



(a) Input (b) Line Segments (c) Planes (d) Depth Map (e) 3D Model

Figure 4. Piece-wise planar reconstruction results on the ScanNet dataset [1]



(a) Input (b) Line Segments (c) w/o Line Segments (d) w/ Line Segments

Figure 5. Illustration of plane detection guided by line segments in the proposed PlaneTR. Here, 'w/o' and 'w/' mean 'without' and 'with', respectively

References

[1] Angela Dai, Angel X. Chang, Manolis Savva, Maciej Halber, Thomas A. Funkhouser, and Matthias Nießner. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, pages 2432–2443, 2017.

[2] Chen Liu, Kihwan Kim, Jinwei Gu, Yasutaka Furukawa, and Jan Kautz. Planercnn: 3d plane detection and reconstruction from a single image. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, pages 4450–4459, 2019.

[3] Chen Liu, Jimei Yang, Duygu Ceylan, Ersin Yumer, and Yasutaka Furukawa. Planenet: Piece-wise planar reconstruction from a single RGB image. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, pages 2579–2588, 2018.

[4] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. Indoor segmentation and support inference from RGBD images. In *European Conference on Computer Vision, ECCV*, volume 7576, pages 746–760, 2012.

[5] Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, Wenyu Liu, and Bin Xiao. Deep high-resolution representation learning for visual recognition. *CoRR*, abs/1908.07919, 2019.

[6] Nan Xue, Tianfu Wu, Song Bai, Fudong Wang, Gui-Song Xia, Liangpei Zhang, and Philip H. S. Torr. Holistically-attracted wireframe parsing. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, pages 2785–2794, 2020.

[7] Zehao Yu, Jia Zheng, Dongze Lian, Zihan Zhou, and Shenghua Gao. Single-image piece-wise planar 3d reconstruction via associative embedding. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, pages 1029–1037, 2019.