PlaneRecNet: Multi-Task Learning with Cross-Task Consistency for Piece-Wise Plane Detection and Reconstruction from a Single RGB Image (Supplementary Material)

Yaxu Xie
yaxu.xie@dkfi.de
Fangwen Shu
fangwen.shu@dkfi.de
Jason Rambach
jason_raphael.rambach@dfki.de
Alain Pagani
alain.pagani@dkfi.de
Didier Stricker
didier.stricker@dkfi.de

German Research Centre for Artificial Intelligence (DFKI)
Kaiserslautern, Germany

Overview of the Supplementary Material:

- In Section 1: we provide more qualitative results of plane segmentation in comparison with other state-of-the-art as well as the run-time comparison.
- In Section 2: we explain the different designs of surface normal constraint, and give the qualitative comparison correspondingly.

1 Comparison against State-of-the-Art

More qualitative results of plane segmentation on ScanNet dataset [1] are given in Figure 1. The run-time comparison is presented in Table 1. All timings are measured on the same computing platform with AMD Ryzen 9 3900C CPU (12-cores) and a single NVIDIA RTX 2070 GPU. Our network outperforms the others, and about 3.8x times faster than PlaneRCNN [2].

As shown in the qualitative comparison, our network provide generally more accurate boundaries for major plane instances. Trained with the same dataset, our network shows a better geometric interpretation about plane, thanks to the cross-task consistency training strategy. As instance, in the first row of Figure 1, our network does not predict the window's curtain as planar regions, which is superior than PlaneAE [6] and PlaneRCNN [2].

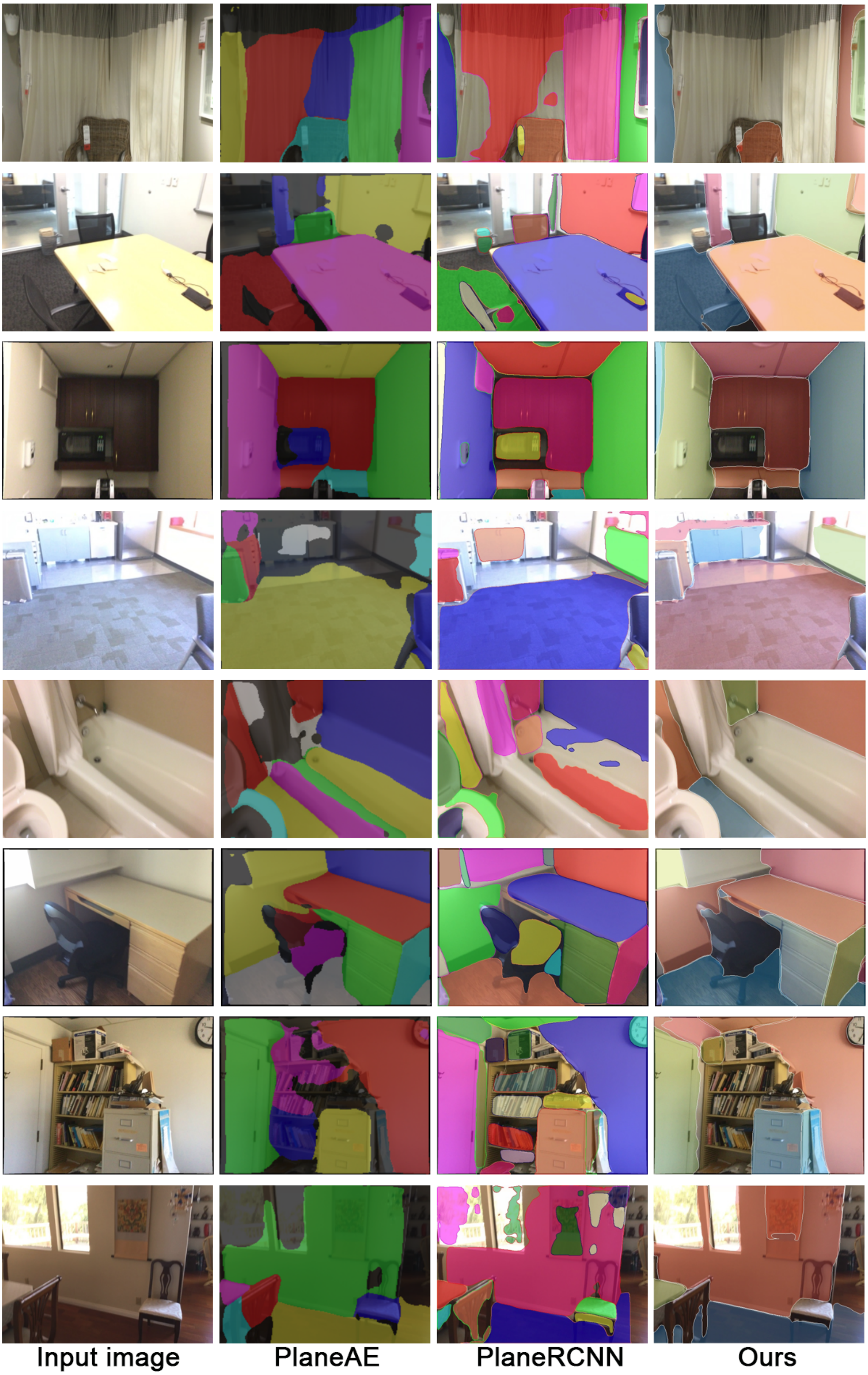

Figure 1: qualitative comparison of PlaneAE [6], PlaneRCNN [7] and ours.

Method	Output Size	Run time (FPS)
PlaneAE [6]	192×256	9.7
PlaneRCNN [2]	480×640	3.0
Ours	480×640	14.4

Table 1: Run-time comparison against state-of-the-art.

2 Ablation Study about Surface Normal Constraint

Here we trained our network with three different designs of surface normal constraint: (1) Combined Normal Map [3] with differentiable least square method [4], (2) original Virtual Normal Loss [5] and (3) our hybird method. As shown in the qualitative comparison in Figure 2, our method shows less artificial effect on planar surface with complex texture, like the poster and paint hanging on the walls.

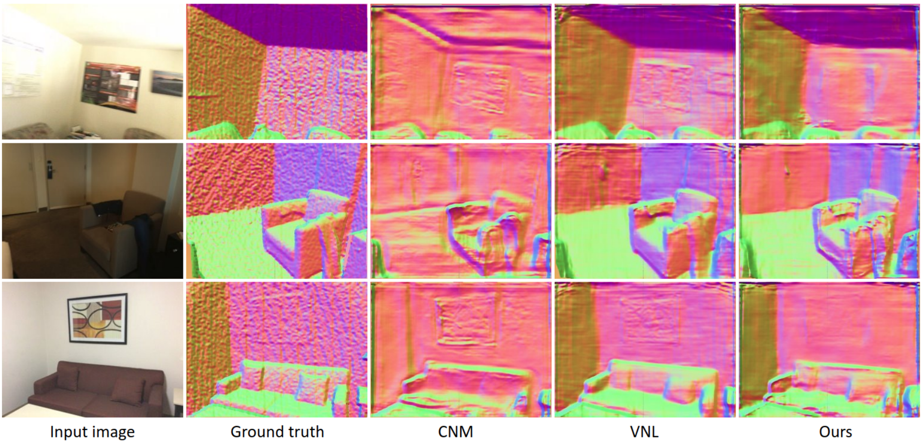

Figure 2: qualitative comparison of different surface normal constraints.

References

[1] Angela Dai, Angel X Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5828–5839, 2017.

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

[3] Xiaoxiao Long, Lingjie Liu, Christian Theobalt, and Wenping Wang. Occlusion-aware depth estimation with adaptive normal constraints. *arXiv e-prints*, pages arXiv–2004, 2020.

[4] Xiaojuan Qi, Renjie Liao, Zhengzhe Liu, Raquel Urtasun, and Jiaya Jia. Geonet: Geometric neural network for joint depth and surface normal estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 283–291, 2018.

[5] Wei Yin, Yifan Liu, Chunhua Shen, and Youliang Yan. Enforcing geometric constraints of virtual normal for depth prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.

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