GPR-Net: Multi-view Layout Estimation via a Geometry-aware Panorama Registration Network
Supplemental Material

Jheng-Wei Su[1] Chi-Han Peng[2] Peter Wonka[3] Hung-Kuo Chu[1]

[1]National Tsing Hua University [2]National Yang Ming Chiao Tung University
[3]KASUT

1. More baselines comparison on the panorama registration

In Table 1, we report more competing baselines on the panorama registration, including DirectionNet [1] and LoFTR [5] + OpenMVG [4] following the setting in the CoVisPose [2]. Comparing with all the other competing methods, our GPR-Net (Direct) still achieves the best performance overall. The main problem of OpenMVG is that it returns lots of unsuccessful registrations. Although using OpenMVG with the LoFTR (a robust learning-based feature matcher) improves the success rate, the translation and rotation retain low accuracy. DirectionNet can predict a registration for all the panorama pairs (100% success rate), but the registration accuracy is usually not good, especially for the Overlap-Low category. In contrast, GPR-Net can estimate robust registration results even in the Overlap-Low category, and achieve the best registration accuracy across all the metrics and different categories.

Table 1. **Quantitative evaluation on panorama registration.** We categorize the test dataset according to the spatial overlap ratio and highlight the best results in yellow.

Overlap	Method	Success (% ↑)	Rotation					Translation angle					Translation vector		
			Mn (° ↓)	Med (° ↓)	2.5° ↑	5° ↑	10° ↑	Mn (° ↓)	Med (° ↓)	2.5° ↑	5° ↑	10° ↑	Mn ($m.$ ↓)	Med ($m.$ ↓)	$0.5m.$ ↑
Overall	GPR-Net (RANSAC)	99.94	4.4552	1.6138	0.6733	0.8791	0.9464	6.3809	1.8979	0.6026	0.8107	0.9051	0.3203	0.1172	0.8907
	GPR-Net (Direct)	100	2.0687	0.7670	0.9661	0.9822	0.9858	5.3703	2.3710	0.5238	0.7953	0.9296	0.2365	0.1468	0.9176
	OpenMVG (SIFT)	22.93	72.1806	66.9710	0.2138	0.2138	0.2138	73.6517	68.4432	0.1006	0.1369	0.1627	-	-	-
	OpenMVG (LoFTR)	97.31	75.4980	81.1601	0.0852	0.1486	0.2222	90.5169	90.3012	0.0516	0.0895	0.1355	-	-	-
	DirectionNet	100	26.7651	4.3237	0.3434	0.5435	0.7172	26.3579	9.8102	0.1428	0.2876	0.5074	-	-	-
High	GPR-Net (RANSAC)	100	2.1114	1.3080	0.7806	0.9599	0.9897	5.9692	1.6772	0.6511	0.8494	0.9317	0.1260	0.0788	0.9783
	GPR-Net (Direct)	100	1.2105	0.7811	0.9794	0.9935	0.9951	6.8568	2.2832	0.5395	0.8082	0.9193	0.1437	0.1076	0.9821
	OpenMVG (SIFT)	30.50	69.4531	60.2696	0.2719	0.2719	0.2719	67.6999	59.4363	0.1381	0.1934	0.2297	-	-	-
	OpenMVG (LoFTR)	97.56	72.3996	77.2172	0.1105	0.1853	0.2692	86.9587	88.9230	0.0520	0.1002	0.1419	-	-	-
	DirectionNet	100	3.3270	1.9398	0.6163	0.8818	0.9843	7.6541	4.9998	0.2488	0.5003	0.7935	-	-	-
Medium	GPR-Net (RANSAC)	100	3.9025	1.6306	0.6761	0.8867	0.9519	5.4238	1.8655	0.6109	0.8270	0.9168	0.2964	0.1340	0.8882
	GPR-Net (Direct)	100	1.6852	0.7571	0.9696	0.9866	0.9902	4.4067	2.3046	0.5396	0.8172	0.9457	0.2411	0.1624	0.9121
	OpenMVG (SIFT)	22.04	75.3876	72.5451	0.2067	0.2067	0.2067	74.0510	69.6638	0.0974	0.1307	0.1535	-	-	-
	OpenMVG (LoFTR)	97.93	75.8301	84.0298	0.0806	0.1503	0.2244	90.0503	91.0122	0.0534	0.0915	0.1383	-	-	-
	DirectionNet	100	30.1753	5.4123	0.2868	0.4667	0.6903	27.4318	11.9780	0.1287	0.2345	0.4384	-	-	-
Low	GPR-Net (RANSAC)	99.78	7.6195	2.1033	0.5622	0.7870	0.8951	8.1646	2.2390	0.5416	0.7476	0.8611	0.5499	0.1560	0.8070
	GPR-Net (Direct)	100	3.4979	0.7700	0.9476	0.9643	0.9697	5.3262	2.6039	0.4843	0.7497	0.9157	0.3221	0.1866	0.8616
	OpenMVG (SIFT)	0.1957	74.4980	72.2465	0.1665	0.1665	0.1665	79.8191	77.3498	0.0681	0.0897	0.1097	-	-	-
	OpenMVG (LoFTR)	97.02	76.9634	82.0682	0.0736	0.1299	0.1981	92.4105	90.9182	0.0509	0.0836	0.1315	-	-	-
	DirectionNet	100	37.9143	7.2713	0.2225	0.3936	0.5901	35.4612	14.9505	0.0931	0.1984	0.3806	-	-	-

2. Additional ablation study

Joint layout and correspondence optimization architecture. In this experiment, we want to know whether jointly learning the layout and correspondence in the geometry transformer would benefit the pose and layout estimation. To this end, we substitute the layout estimation part (associated MLP heads) with the LED^2-Net [6] and directly estimate a layout for each input panorama. As shown in Table 2, we obtain better accuracy using our joint optimization architecture in the geometry transformer.

Table 2. **Ablation study on the joint layout and correspondence optimization architecture.** The best results are in yellow highlight.

Method	w/ GT pose			w/o GT pose		
	2D IoU↑	δ^i ↑	3D IoU↑	2D IoU↑	δ^i ↑	3D IoU↑
w/o joint optimization	0.8364	0.9557	0.8131	0.7920	0.9413	0.7713
w/ joint optimization	0.8449	0.9603	0.8211	0.8026	0.9452	0.7816

Table 3. **Ablation study on the effect of geometry transformer's outputs.** The best results are in yellow highlight.

Boundary coordinate	Correspondence	Co-visibility	Success (% ↑)	Rotation					Translation angle					Translation vector		
				Mn (° ↓)	Med (° ↓)	2.5° ↑	5° ↑	10° ↑	Mn (° ↓)	Med (° ↓)	2.5° ↑	5° ↑	10° ↑	Mn ($m.$ ↓)	Med ($m.$ ↓)	0.5$m.$ ↑
GT	predicted	predicted	99.94	3.8930	1.4490	0.7150	0.9001	0.9602	5.2676	1.4035	0.6843	0.8460	0.9220	0.2631	0.0682	0.9107
predicted	GT	predicted	99.94	2.3105	0.9447	0.8681	0.9371	0.9683	3.5325	0.8734	0.8083	0.8997	0.9466	0.1930	0.0640	0.9309
predicted	predicted	GT	99.95	4.4887	1.6501	0.6659	0.8763	0.9458	6.4160	1.8800	0.6024	0.8089	0.9093	0.3202	0.1172	0.8922
predicted	predicted	predicted	99.94	4.4552	1.6138	0.6733	0.8791	0.9464	6.3809	1.8979	0.6026	0.8107	0.9051	0.3203	0.1172	0.8907

The effect of geometry transformer's outputs. To further understand how the geometry transformer's outputs (ceiling/floor boundary coordinates, correspondence, co-visibility) influence the registration performance, we conduct an ablated setting by substituting the outputs of respective MLP heads with ground truth data. As shown in Table 3, using the ground truth correspondence map achieves the most significant performance boost on all the registration metrics, followed by the ground truth ceiling/floor boundary coordinates. However, improvements in using ground truth co-visibility are marginal because the co-visibility is mainly used to filter out invalid matching pairs in the RANSAC-based pose estimation.

3. Visual Comparison

We show visual comparisons with other competing methods. The first column shows two input panoramas with their estimated layouts. The remaining columns show the ground-truth layout, our layouts, LED^2-Net's layouts, and LGT-Net's [3] layouts in blue, green, yellow, red, respectively.

References

[1] Kefan Chen, Noah Snavely, and Ameesh Makadia. Wide-baseline relative camera pose estimation with directional learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3258–3268, June 2021. 1

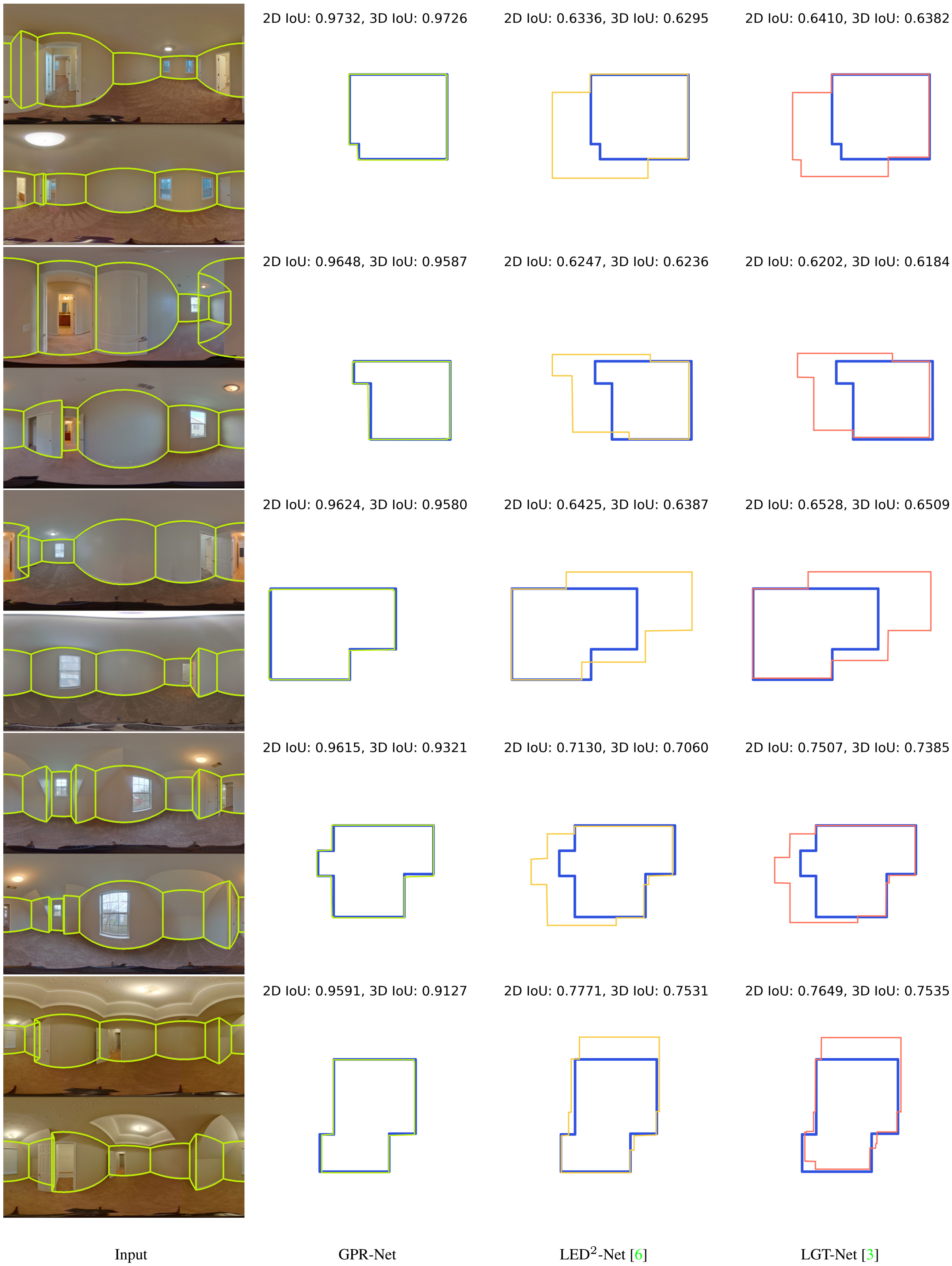
[2] Will Hutchcroft, Yuguang Li, Ivaylo Boyadzhiev, Zhiqiang Wan, Haiyan Wang, and Sing Bing Kang. Covispose: Co-visibility pose transformer for wide-baseline relative pose estimation in 360° indoor panoramas. In *ECCV*, 2022. 1

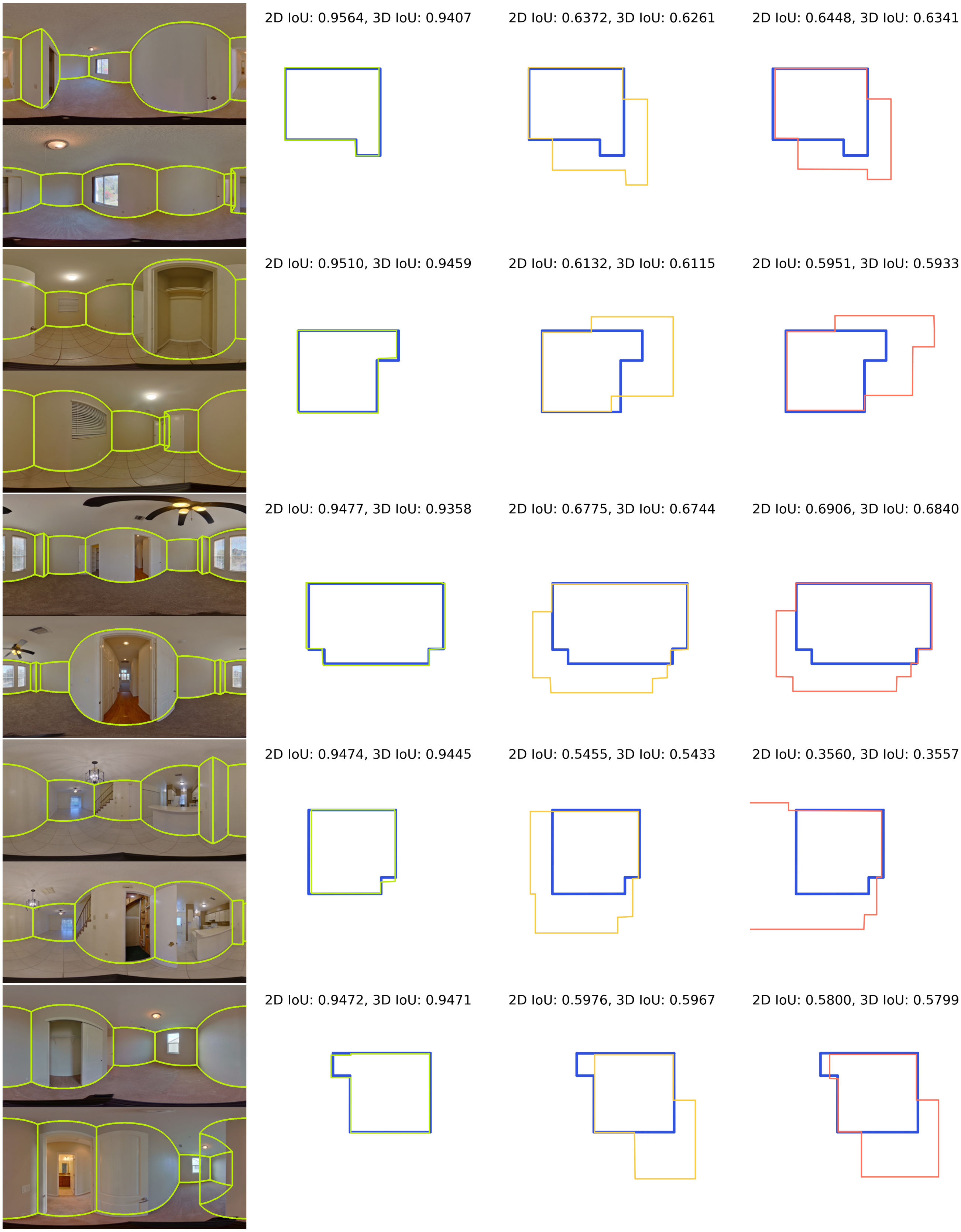
[3] Zhigang Jiang, Zhongzheng Xiang, Jinhua Xu, and Ming Zhao. Lgt-net: Indoor panoramic room layout estimation with geometry-aware transformer network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1654–1663, June 2022. 2, 3, 4, 5, 6, 7, 8

[4] Pierre Moulon, Pascal Monasse, Romuald Perrot, and Renaud Marlet. Openmvg: Open multiple view geometry. In *International Workshop on Reproducible Research in Pattern Recognition*, pages 60–74. Springer, 2016. 1

[5] Jiaming Sun, Zehong Shen, Yuang Wang, Hujun Bao, and Xiaowei Zhou. LoFTR: Detector-free local feature matching with transformers. *CVPR*, 2021. 1

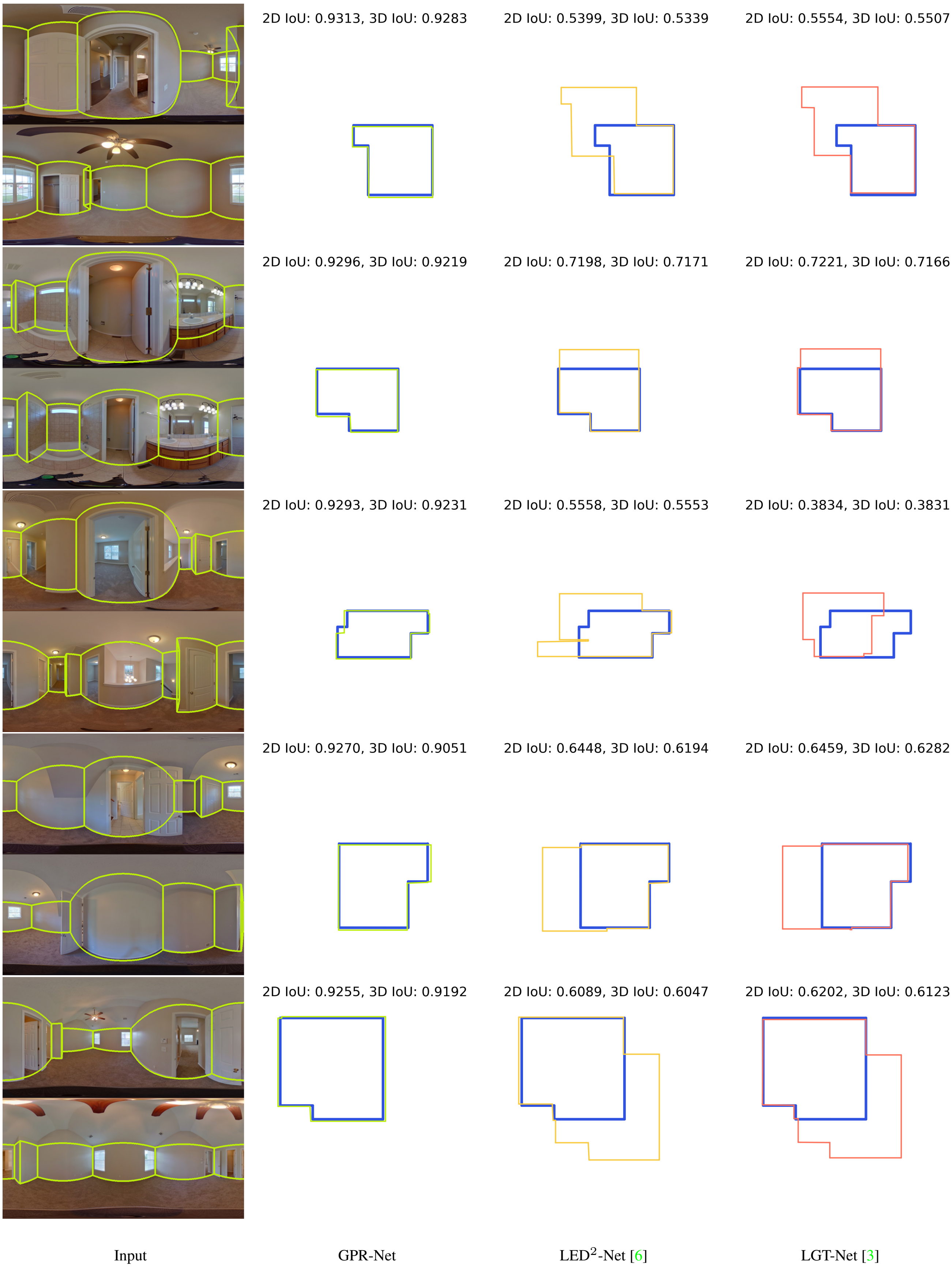
[6] Fu-En Wang, Yu-Hsuan Yeh, Min Sun, Wei-Chen Chiu, and Yi-Hsuan Tsai. Led2-net: Monocular 360deg layout estimation via differentiable depth rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12956–12965, June 2021. 1, 3, 4, 5, 6, 7, 8

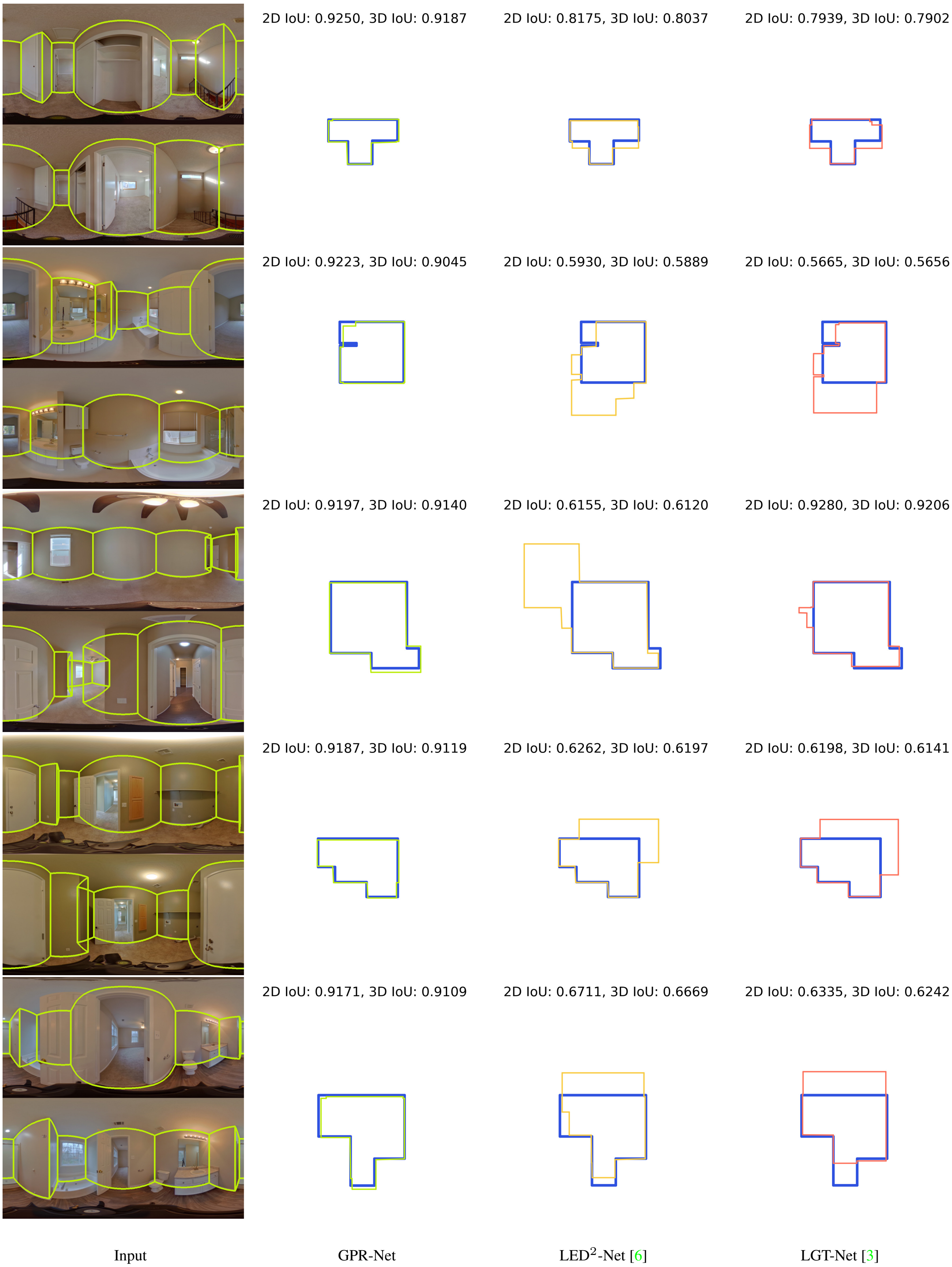

2D IoU: 0.9732, 3D IoU: 0.9726
2D IoU: 0.6336, 3D IoU: 0.6295
2D IoU: 0.6410, 3D IoU: 0.6382
2D IoU: 0.9648, 3D IoU: 0.9587
2D IoU: 0.6247, 3D IoU: 0.6236
2D IoU: 0.6202, 3D IoU: 0.6184
2D IoU: 0.9624, 3D IoU: 0.9580
2D IoU: 0.6425, 3D IoU: 0.6387
2D IoU: 0.6528, 3D IoU: 0.6509
2D IoU: 0.9615, 3D IoU: 0.9321
2D IoU: 0.7130, 3D IoU: 0.7060
2D IoU: 0.7507, 3D IoU: 0.7385
2D IoU: 0.9591, 3D IoU: 0.9127
2D IoU: 0.7771, 3D IoU: 0.7531
2D IoU: 0.7649, 3D IoU: 0.7535
Input
GPR-Net
LED^2-Net [6]
LGT-Net [3]



Input | GPR-Net | LED^2-Net [6] | LGT-Net [3]

Input GPR-Net LED^2-Net [6] LGT-Net [3]


2D IoU: 0.9313, 3D IoU: 0.9283
2D IoU: 0.5399, 3D IoU: 0.5339
2D IoU: 0.5554, 3D IoU: 0.5507
2D IoU: 0.9296, 3D IoU: 0.9219
2D IoU: 0.7198, 3D IoU: 0.7171
2D IoU: 0.7221, 3D IoU: 0.7166
2D IoU: 0.9293, 3D IoU: 0.9231
2D IoU: 0.5558, 3D IoU: 0.5553
2D IoU: 0.3834, 3D IoU: 0.3831
2D IoU: 0.9270, 3D IoU: 0.9051
2D IoU: 0.6448, 3D IoU: 0.6194
2D IoU: 0.6459, 3D IoU: 0.6282
2D IoU: 0.9255, 3D IoU: 0.9192
2D IoU: 0.6089, 3D IoU: 0.6047
2D IoU: 0.6202, 3D IoU: 0.6123
Input
GPR-Net
LED^2-Net [6]
LGT-Net [3]


2D IoU: 0.9250, 3D IoU: 0.9187
2D IoU: 0.8175, 3D IoU: 0.8037
2D IoU: 0.7939, 3D IoU: 0.7902
2D IoU: 0.9223, 3D IoU: 0.9045
2D IoU: 0.5930, 3D IoU: 0.5889
2D IoU: 0.5665, 3D IoU: 0.5656
2D IoU: 0.9197, 3D IoU: 0.9140
2D IoU: 0.6155, 3D IoU: 0.6120
2D IoU: 0.9280, 3D IoU: 0.9206
2D IoU: 0.9187, 3D IoU: 0.9119
2D IoU: 0.6262, 3D IoU: 0.6197
2D IoU: 0.6198, 3D IoU: 0.6141
2D IoU: 0.9171, 3D IoU: 0.9109
2D IoU: 0.6711, 3D IoU: 0.6669
2D IoU: 0.6335, 3D IoU: 0.6242
Input
GPR-Net
LED^2-Net [6]
LGT-Net [3]


2D IoU: 0.9157, 3D IoU: 0.9015
2D IoU: 0.8134, 3D IoU: 0.8038
2D IoU: 0.6091, 3D IoU: 0.6032
2D IoU: 0.9111, 3D IoU: 0.9054
2D IoU: 0.7181, 3D IoU: 0.7097
2D IoU: 0.6956, 3D IoU: 0.6908
2D IoU: 0.9075, 3D IoU: 0.9010
2D IoU: 0.7086, 3D IoU: 0.7074
2D IoU: 0.6981, 3D IoU: 0.6925
2D IoU: 0.8882, 3D IoU: 0.8712
2D IoU: 0.6335, 3D IoU: 0.6253
2D IoU: 0.7270, 3D IoU: 0.7096
2D IoU: 0.8464, 3D IoU: 0.8427
2D IoU: 0.2443, 3D IoU: 0.2399
2D IoU: 0.4785, 3D IoU: 0.4561
Input
GPR-Net
LED^2-Net [6]
LGT-Net [3]