Supplementary Materials for "GeoLayout: Geometry Driven Room Layout Estimation Based on Depth of Planes"

Weidong Zhang[1,2], Wei Zhang[1], and Yinda Zhang[3]

[1] School of Control Science and Engineering, Shandong University
[2] School of Communications and Information Engineering, Xi'an University of Posts & Telecommunications
[3] Google Research

1 Ablation Study

1.1 Performance of Loss Functions

In order to evaluate the performance of the proposed loss functions in Eq. (9), we train the models with different losses: (1) $\mathcal{L}_p$, (2) $\mathcal{L}_p + \alpha\mathcal{L}_{var}$, (3) $\mathcal{L}_p + \beta\mathcal{L}_z$, (4) full loss $\mathcal{L}_p + \alpha\mathcal{L}_{var} + \beta\mathcal{L}_z$. The results are shown in Table 1. As can be seen, although the ground truth surface parameters are available during training, the variance loss $\mathcal{L}_{var}$ and inverse depth loss $\mathcal{L}_z$ show improvements to the layout estimation result.

Table 1: Results of our method trained with different loss functions on the Matterport3D-Layout dataset.

Method	$e_{\text{pix.}}$	$e_{\text{cor.}}$	$e_{\text{3D\_cor.}}$	rms	rel	log10	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
$\mathcal{L}_p$	10.96	7.52	19.20	0.724	0.147	0.063	0.805	0.955	0.986
$\mathcal{L}_p + \alpha\mathcal{L}_{var}$	11.19	7.74	18.47	0.745	0.154	0.066	0.787	0.951	0.985
$\mathcal{L}_p + \beta\mathcal{L}_z$	10.33	7.30	16.82	0.640	0.139	0.059	0.821	0.961	0.989
full loss	**5.24**	**4.36**	**12.82**	**0.456**	**0.111**	**0.047**	**0.892**	**0.975**	**0.994**

Table 2: Results of our method trained with different loss functions on the LSUN dataset.

Loss function	e_{pixel} (%)	e_{corner} (%)
w/o $\mathcal{L}_d$	31.40	19.37
w/o $\mathcal{L}_z$	14.78	10.16
w/o $\mathcal{L}_s$	20.18	16.37
full loss	**6.10**	**4.66**

Similarly, to evaluate the performance of the loss functions for 2D dataset as in Eq. (13), the following models trained with different losses are compared: (1) w/o $\mathcal{L}_d$, (2) w/o $\mathcal{L}_z$, (3) w/o $\mathcal{L}_s$, (4) full loss $\mathcal{L}_d + \eta\mathcal{L}_z + \theta\mathcal{L}_s$. The quantitative results are shown in Table 2. Fig. 1 visualizes several results of the models trained with different losses. It can be observed that the predicted surface parameters without the discriminative loss are inconsistent within each surface. The performance without the inverse depth loss is not very bad because the stretch loss also encourages the generated layout to be consistent with the ground truth segmentation. The impact of the stretch loss can be found in the layout generation process. The model trained with full loss shows the best performance.

1.2 Fine-tuning on LSUN

On the LSUN validation set, the performance of the model pre-trained on Matterport3D-Layout with and without fine-tuning on LSUN is compared in Table 3 of the main manuscript. As a complement to Table 3, the quantitative results are shown in Fig. 2. It can be observed that the model with fine-tuning works better than the pre-trained model, especially when the layout is ambiguous. The reason might be the domain gaps between the two datasets.

2 Additional Results

Additional results on the Matterport3D-Layout dataset are shown in Fig. 3 and Fig. 4, where Fig. 3 shows the results of cuboid rooms and Fig. 4 shows the results of non-cuboid rooms.

More results on the LSUN dataset are given in Fig. 5, with the results of Zhang et al. [38] for comparison.

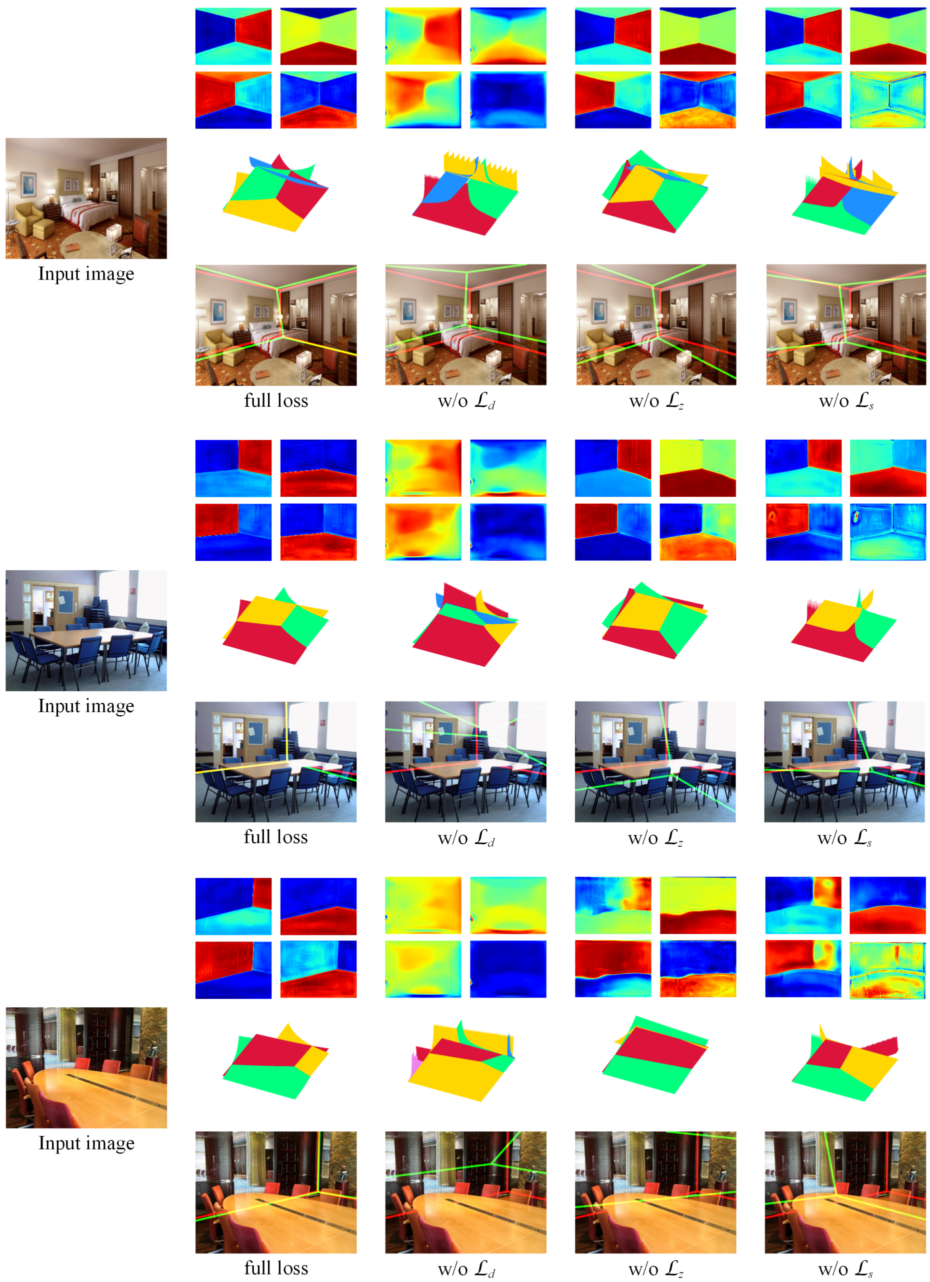


Fig. 1: Results of our method on the LSUN validation set trained with different loss functions.

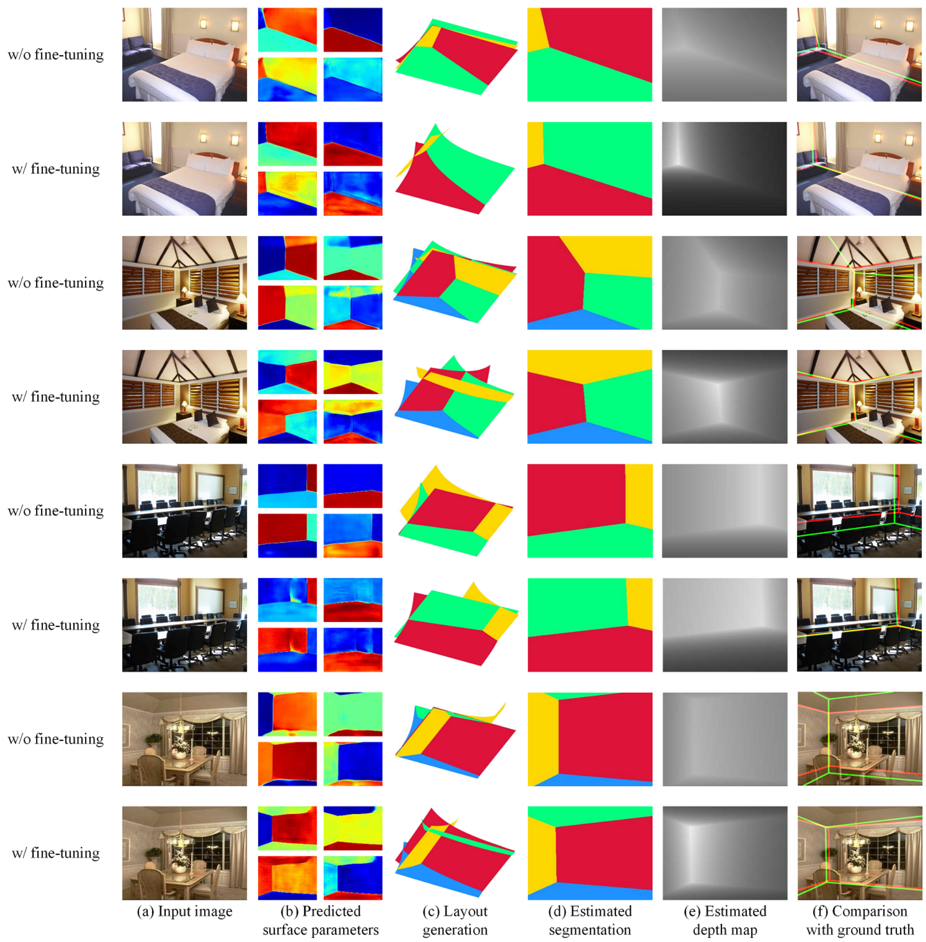


Fig. 2: Quantitative results on the LSUN validation set of the pre-trained model w/wo fine-tuning on LSUN.

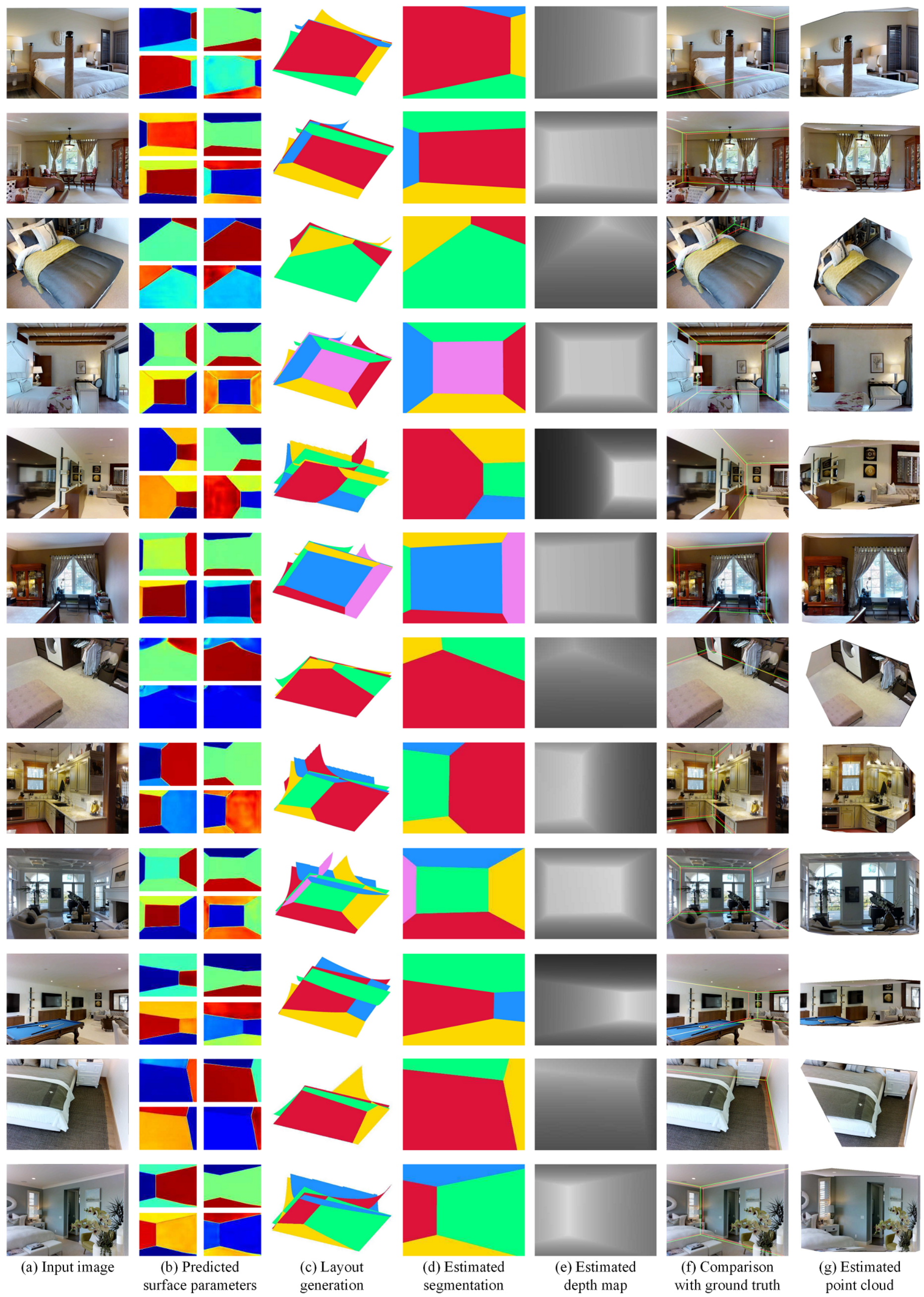


Fig. 3: More layout estimation results of **cuboid** rooms on the Matterport3D-Layout dataset.

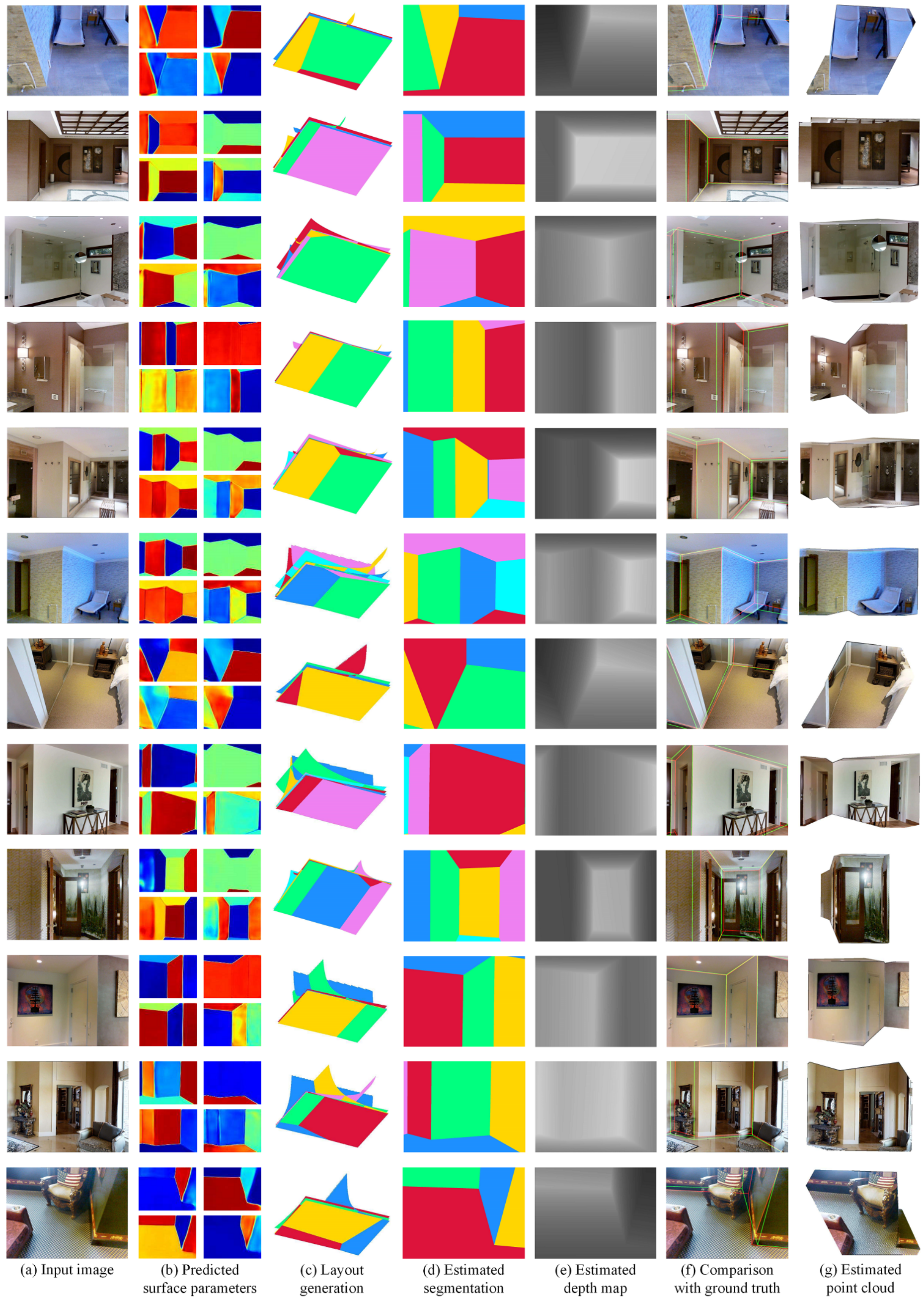
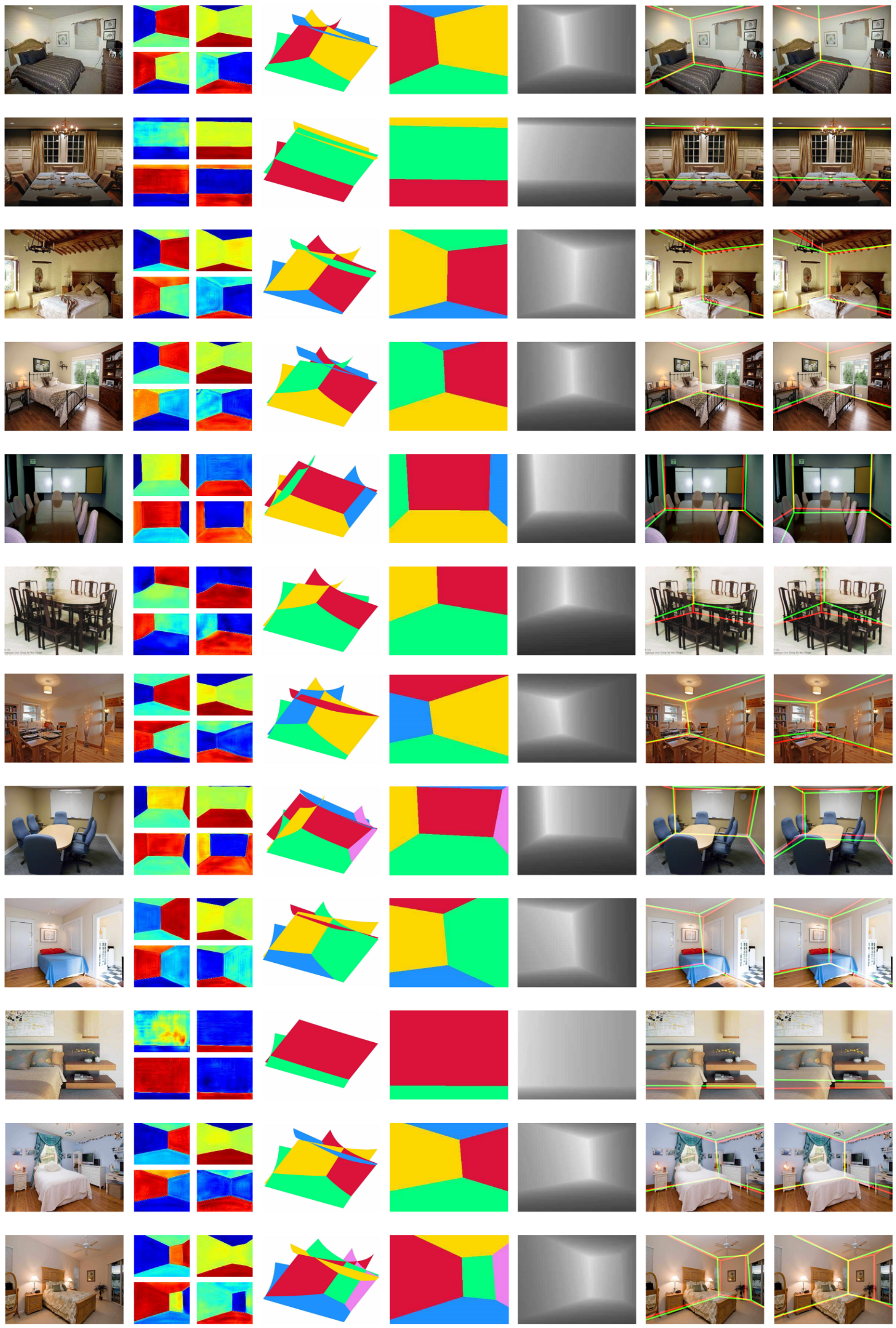
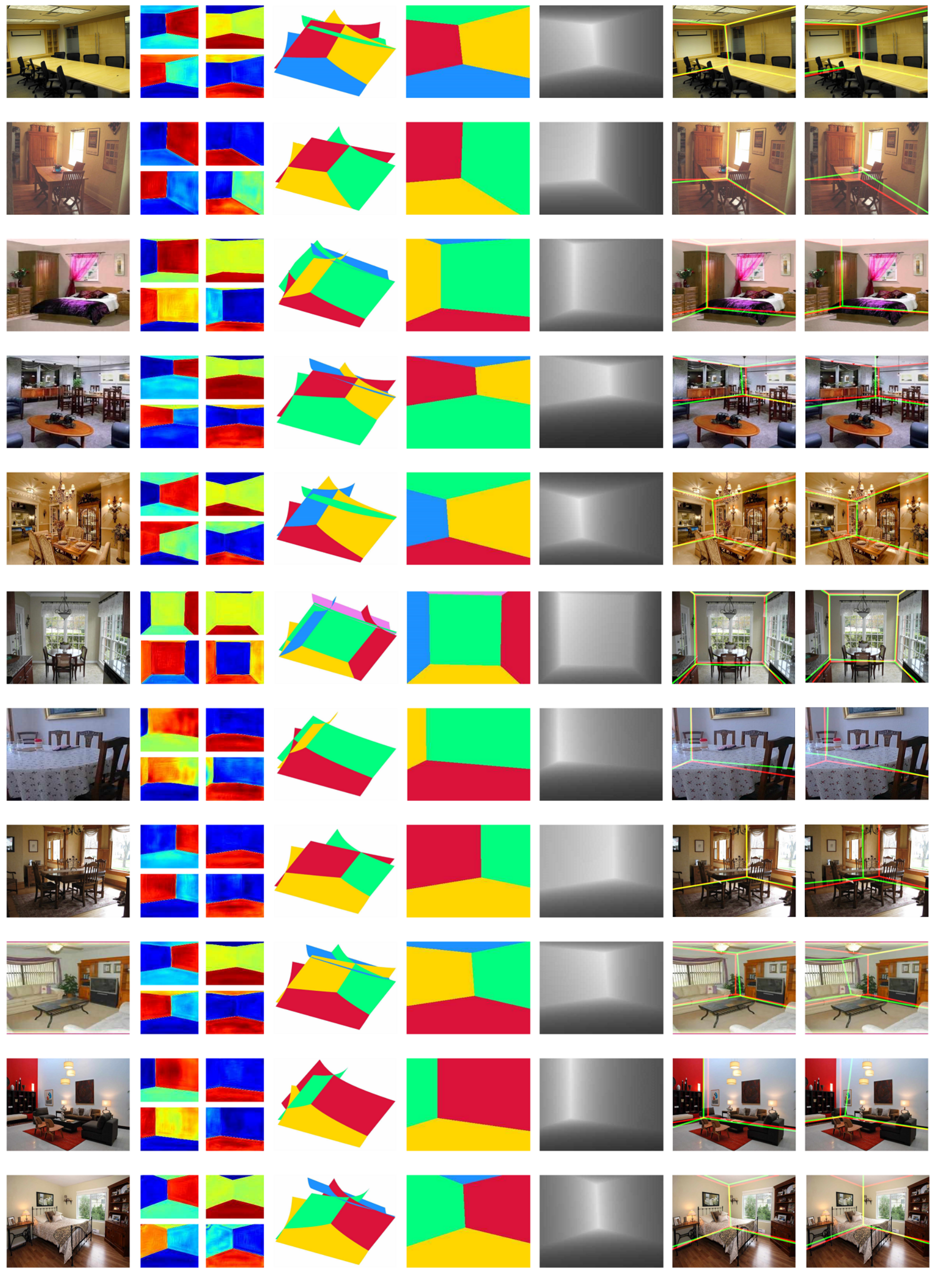


Fig. 4: More layout estimation results of **non-cuboid** rooms on the Matterport3D-Layout dataset.



(a) Input image (b) Predicted surface parameters (c) Layout generation (d) Estimated segmentation (e) Estimated depth map (f) Our result (g) Zhang et al. [38]



(a) Input image | (b) Predicted surface parameters | (c) Layout generation | (d) Estimated segmentation | (e) Estimated depth map | (f) Our result | (g) Zhang et al. [38]

Fig. 5: More layout estimation results on the LSUN validation set. The results of Zhang *et al.* [38] are also shown for comparison.