Supplemental Materials of HorizonNet

1. Pano Stretch Augmentation for Semantic Segmentation

We evaluate the potential advantage of the proposed Pano Stretch Data Augmentation on semantic segmentation task. We train and test on Stanford 2D3D [1] semantic segmentation benchmark. We train PSPNet [3] on subsampled training set and test on the whole testing set. The results are summarized in Table 1.

# of training images	20	50	100	200	500	1040
wo/ pano stretch aug.	31.5	34.9	37.1	40.7	44.2	44.8
w/ pano stretch aug.	**33.2**	**36.1**	**38.4**	**41.9**	**44.3**	**44.9**

Table 1. We evaluate the effect of Pano Stretch Augmentation on semantic segmentation task using the standard metric - mIoU (%). The result implies that the new augmentation technique has the potential to mitigate the "lack of training data" problem for other tasks like semantic segmentation.

2. More Qualitative Results of Cuboid Room Layout Reconstruction

Figure 1: Qualitative results of cuboid layout estimation on PanoContext [2] dataset. The results in the first to the fourth rows are separately sampled from four groups that comprise results with the best 0–25%, 25–50%, 50–75% and 75-100% corner errors, and the four results with the worst corner errors are displayed in the last row. The green lines are ground truth layout while the orange lines are estimated layout.

Figure 2: Qualitative results of cuboid layout estimation on Stanford 2D-3D [1] dataset. The results in the first to the fourth rows are separately sampled from four groups that comprise results with the best 0–25%, 25–50%, 50–75% and 75-100% corner errors, and the four results with the worst corner errors are displayed in the last row. The green lines are ground truth layout while the orange lines are estimated layout.

3. More Qualitative Results of Non-Cuboid Room Layout Reconstruction

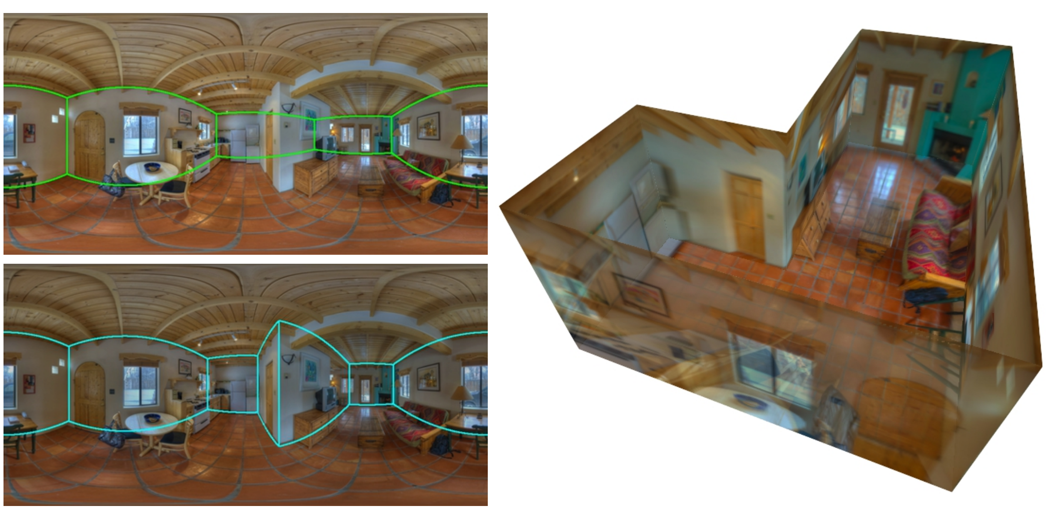

Figure 3: Green lines are original ground truth annotation. Blue lines are room layout estimated by our model.

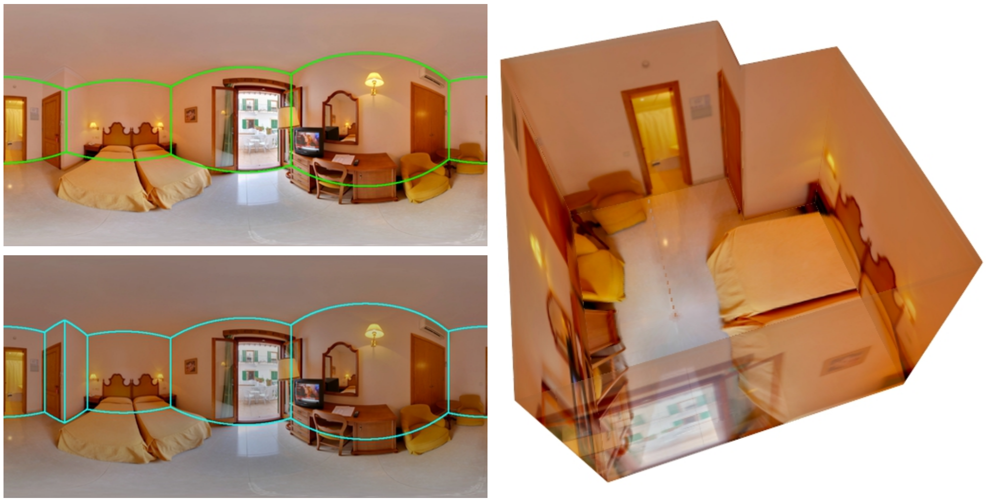

Figure 4: Green lines are original ground truth annotation. Blue lines are room layout estimated by our model.

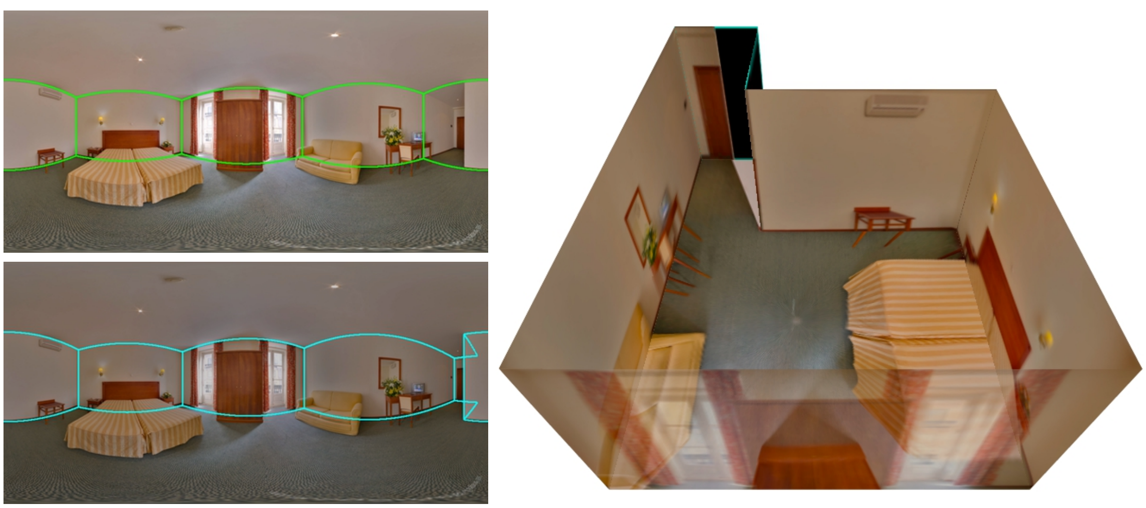

Figure 5: Green lines are original ground truth annotation. Blue lines are room layout estimated by our model. The occlusion walls are filled with black.

Figure 6: Green lines are original ground truth annotation. Blue lines are room layout estimated by our model. The occlusion walls are filled with black.

Figure 7: Green lines are original ground truth annotation. Blue lines are room layout estimated by our model. The occlusion walls are filled with black.

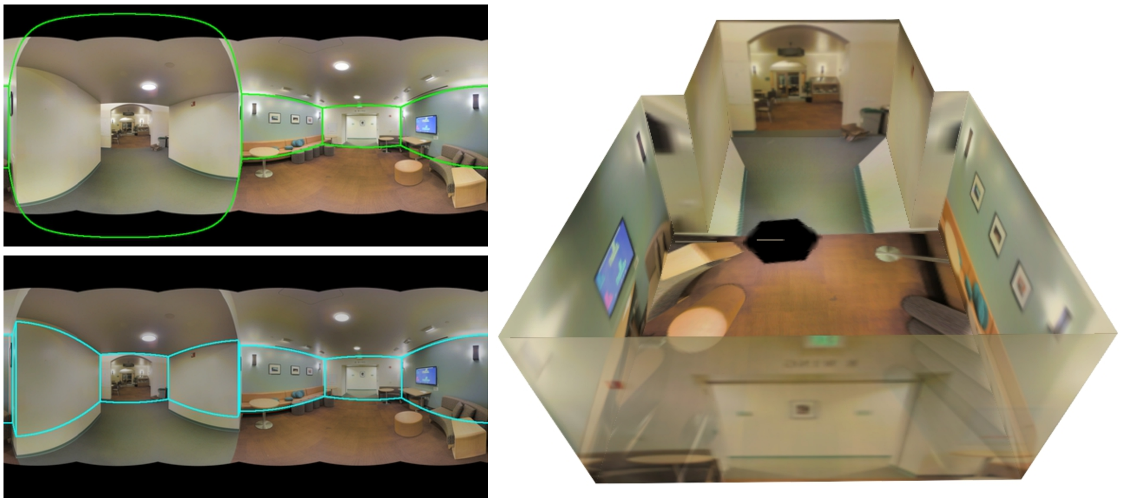

Figure 8: Green lines are original ground truth annotation. Blue lines are room layout estimated by our model.

Figure 9: Green lines are original ground truth annotation. Blue lines are room layout estimated by our model.

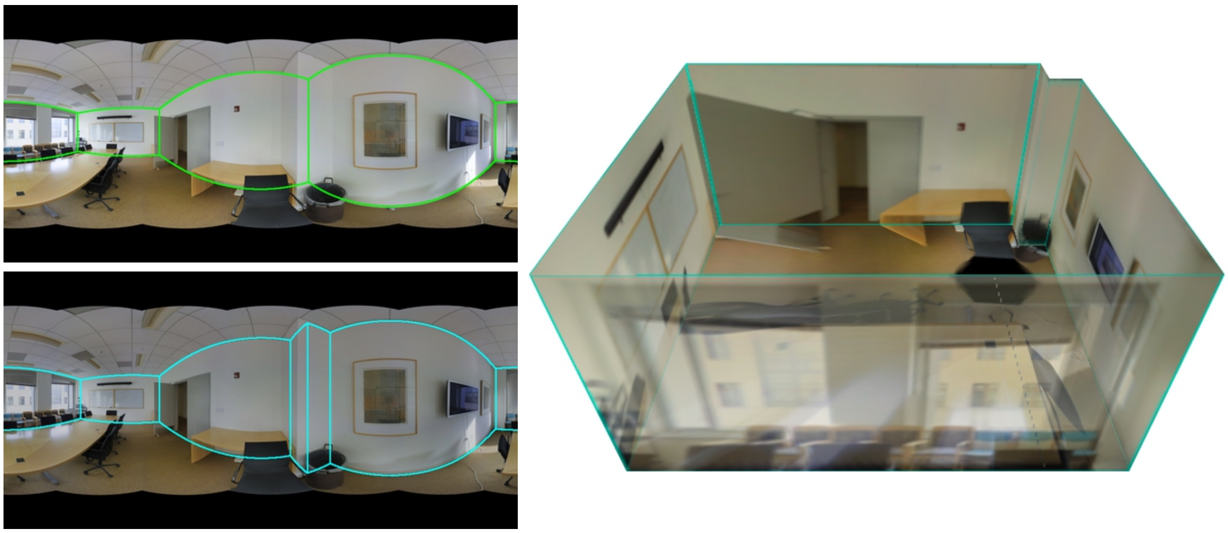

Figure 10: Green lines are original ground truth annotation. Blue lines are room layout estimated by our model.

Figure 11: Green lines are original ground truth annotation. Blue lines are room layout estimated by our model.

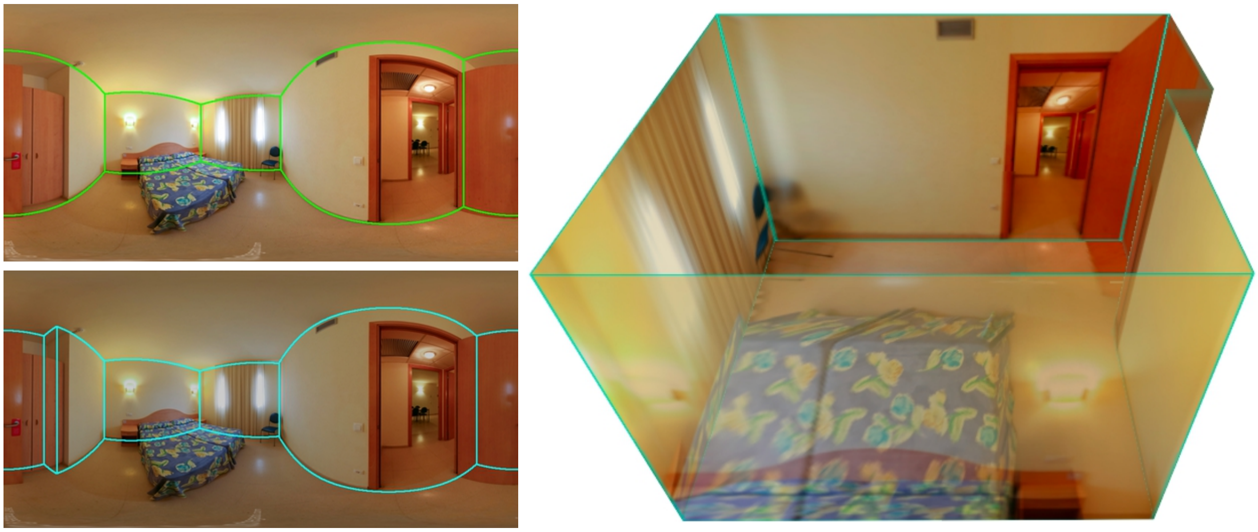

Figure 12: Green lines are original ground truth annotation. Blue lines are room layout estimated by our model.

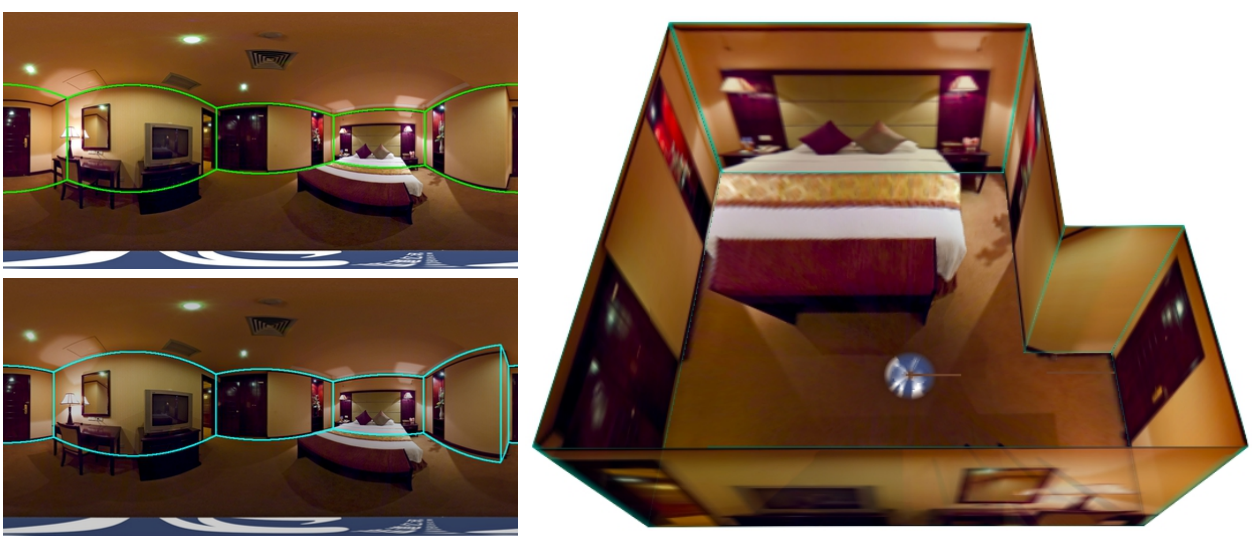

Figure 13: Green lines are original ground truth annotation. Blue lines are room layout estimated by our model.

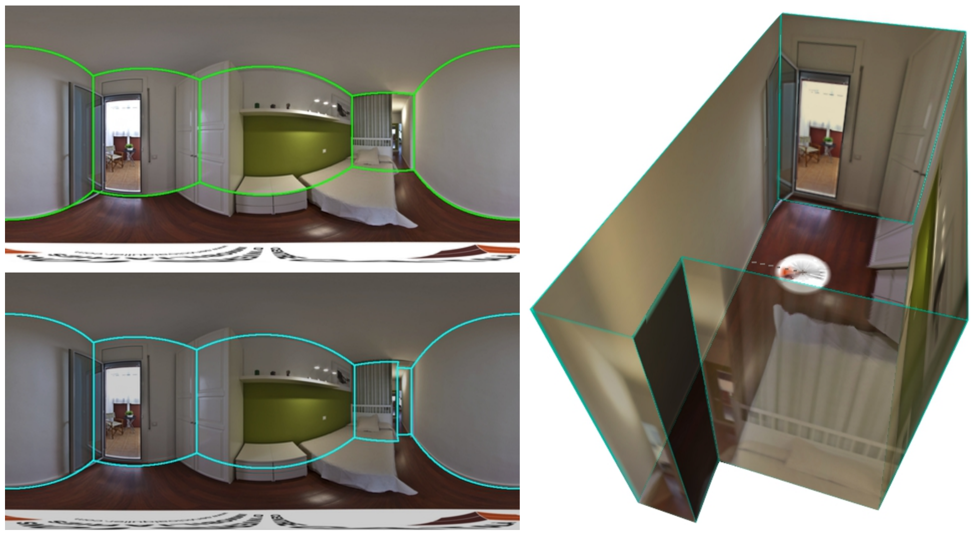

Figure 14: Green lines are original ground truth annotation. Blue lines are room layout estimated by our model. The occlusion wall is filled with black.

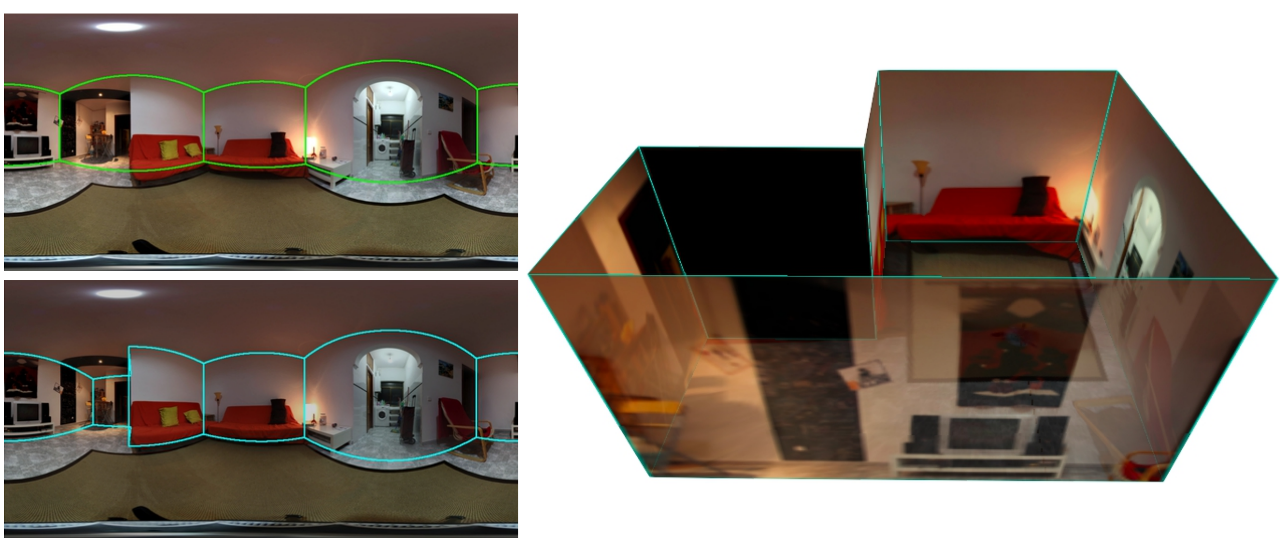

Figure 15: Green lines are original ground truth annotation. Blue lines are room layout estimated by our model. The occlusion wall is filled with black.

References

[1] I. Armeni, A. Sax, A. R. Zamir, and S. Savarese. Joint 2D-3D-Semantic Data for Indoor Scene Understanding. *ArXiv e-prints*, Feb. 2017. 1, 2

[2] Yinda Zhang, Shuran Song, Ping Tan, and Jianxiong Xiao. Panocontext: A whole-room 3d context model for panoramic scene understanding. In *European Conference on Computer Vision*, pages 668–686. Springer, 2014. 1

[3] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2881–2890, 2017. 1