Supplementary Material for
CTRL-C: Camera calibration TRansformer with Line-Classification

Jinwoo Lee, Kakao Brain
Hyunsung Go, Kookmin University
Hyunjoon Lee, Kakao Brain
Sunghyun Cho, POSTECH
Minhyuk Sung, KAIST
Junho Kim, Kookmin University

https://github.com/jwlee-vcl/CTRL-C

S.1. Generalization Tests

We compare the generalization capacity of our network with those of three network-based approaches, DeepHorizon [8], Perceptual [2] and GPNet [4]. Specifically, we train our CTRL-C and other networks with either the Google Street View (GSV) [1] or SUN360 [9] datasets and then test them with the other datasets, including HoliCity [10] dataset, Horizon Lines in the Wild (HLW) [8] dataset, and Eurasian Cities (ECD) [6] dataset. For the HoliCity [10] dataset, we evaluate networks trained with the SUN360 [9] dataset for comparison, since the sampling range of FoV in the SUN360 [9] dataset covers 90°, which is the FoV used in the entire HoliCity [10] dataset. For HLW [8] and ECD [6] datasets, we only measure the horizon line prediction accuracy since these datasets do not provide GT for the other camera parameters.

HoliCity [10] dataset. Table S1 shows the comparison results evaluated on the HoliCity [10] test set. Compared with the other methods, our CTRL-C provides better accuracy overall except for pitch. The AUC of the horizon line errors is improved from 81.72% of the previous SotA to 84.16% of ours, which has a 2.44% gap. Fig. S1 shows experimental results on the HoliCity [10] test set, as in Fig. 5 in the paper, visualizing qualitative evaluations on horizon line predictions. Fig. S2 shows further examples illustrating the inputs and outputs of our network, as in Fig. 6 in the paper.

HLW [8] dataset. Table S2 shows the comparison results evaluated on the HLW [8] test set. Our CTRL-C provides better overall accuracy and less sensitivity to the choice of the training datasets compared to the other network-based approaches. For instance, the difference of AUCs with the different training datasets (GSV [1] and SUN360 [9]) is 5.15% for our CTRL-C while 8.37% for GPNet [4].

ECD [6] dataset. Table S3 shows the comparison results evaluated on the ECD [6] dataset. Our CTRL-C shows better generalization performance than those of the other neural approaches, in both cases of training with the GSV [1] and SUN360 [9] datasets.

Table S1. Quantitative evaluation results on the HoliCity [10] dataset. All the networks are trained with the SUN360 [9] dataset.

Method	Up Dir (°) ↓		Pitch (°) ↓		Roll (°) ↓		FoV (°) ↓		AUC (%) ↑
	Mean	Med.	Mean	Med.	Mean	Med.	Mean	Med.	
DeepHorizon [8]	7.82	3.99	6.10	2.73	3.97	2.67	-	-	70.13
Perceptual [2]	7.37	3.29	6.32	2.86	3.10	1.82	5.48	2.80	70.80
GPNet [4]	4.17	**1.73**	**1.46**	**0.74**	3.65	1.36	10.03	4.29	81.72
CTRL-C	**2.90**	1.99	2.43	1.50	**1.36**	**0.95**	**2.68**	**1.56**	**84.16**

Table S2. Horizon line prediction results on the HLW [8] dataset.

Method	AUC (%) ↑	
Training Sets	GSV	SUN360
DeepHorizon [8]	45.63	40.63
Perceptual [2]	38.29	46.70
GPNet [4]	**48.90**	40.53
CTRL-C	44.93	**50.08**

Table S3. Horizon line prediction results on the ECD dataset [6].

Method	AUC (%) ↑	
Training Sets	GSV	SUN360
DeepHorizon [8]	74.26	74.60
Perceptual [2]	67.97	76.53
GPNet [4]	77.61	75.04
CTRL-C	**77.66**	**79.83**

S.2. Additional Results

Google Street View [1] dataset. Fig. S3 shows additional results on the Google Street View [1] test set, as in Fig. 5 in the paper, visualizing qualitative evaluations on horizon line predictions. Fig. S4 shows additional results with the Google Street View [1] test set, as in Fig. 6 in the paper, illustrating the inputs and outputs of our network.

SUN360 [9] dataset. Fig. S5 shows additional results on the SUN360 [9] test set, as in Fig. 5 in the paper, visualizing qualitative evaluations on horizon line predictions. Fig. S6 shows additional results with the SUN360 [9] test set, as in Fig. 6 in the paper, illustrating the inputs and outputs of our network.

Figure S1. Examples of horizon line prediction on the HoliCity [10] test set.

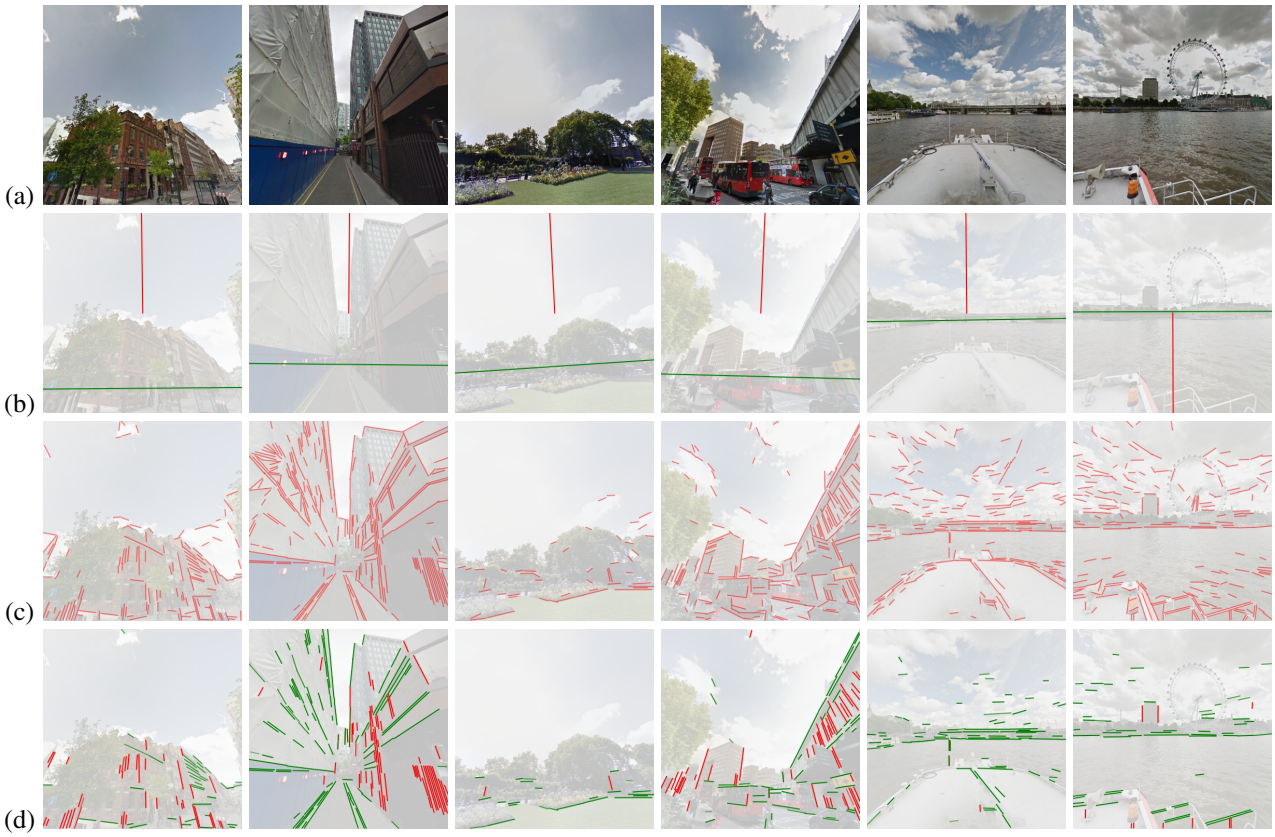

Figure S2. More results with HoliCity [10] test set: (a) input image, (b) estimated horizon line (green) and vertical direction along with the zenith VP (red), (c) detected lines with LSD [7], (d) estimated vertical (red) and horizontal (green) convergence line segments of (c).

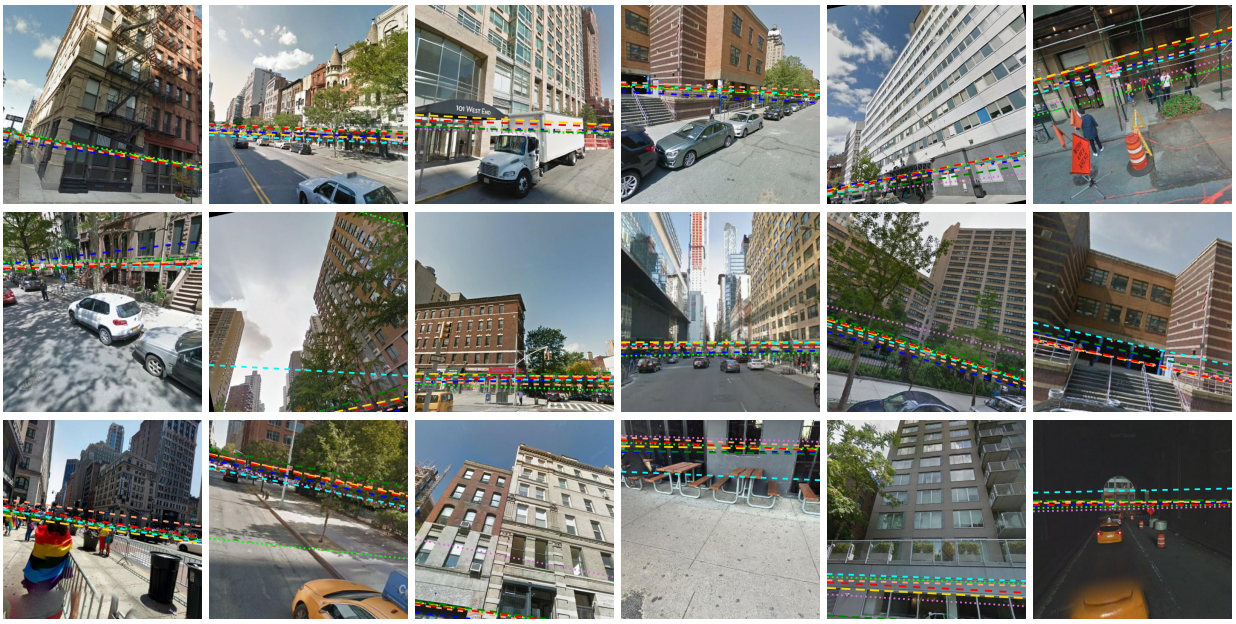

Figure S3. Examples of horizon line prediction on the Google Street View [1] test set.

Figure S4. More results with Google Street View [1] test set: (a) input image, (b) estimated horizon line (green) and vertical direction along with the zenith VP (red), (c) detected lines with LSD [7], (d) estimated vertical (red) and horizontal (green) convergence line segments of (c).

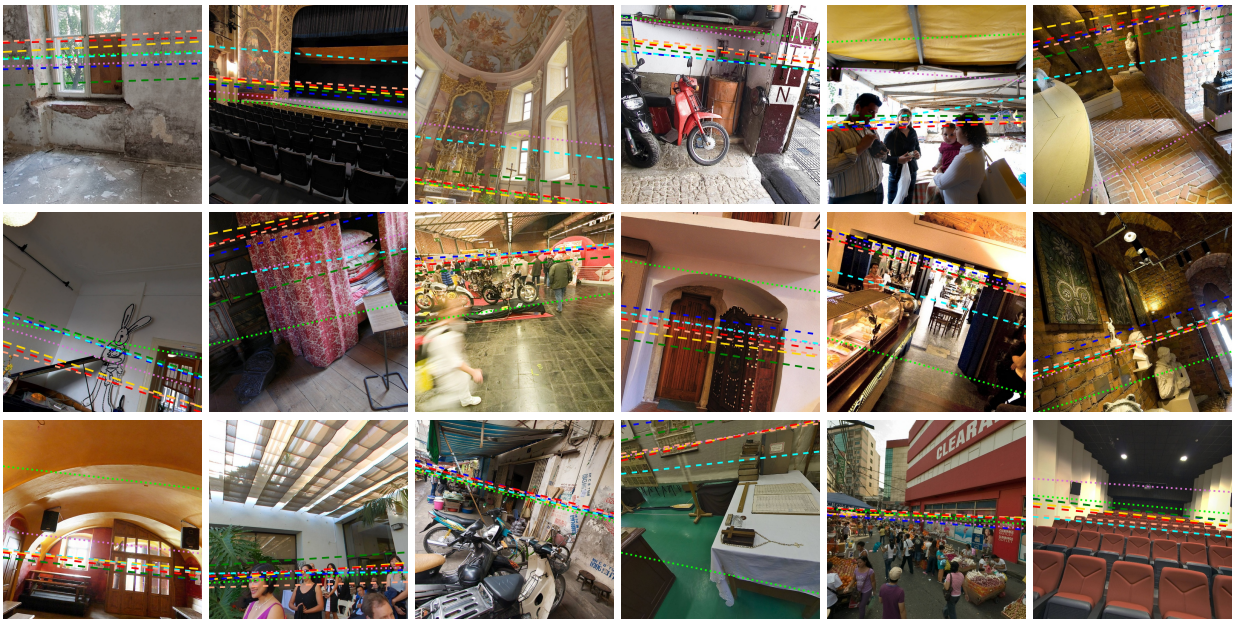

Figure S5. Examples of horizon line prediction on the SUN360 [9] test set.

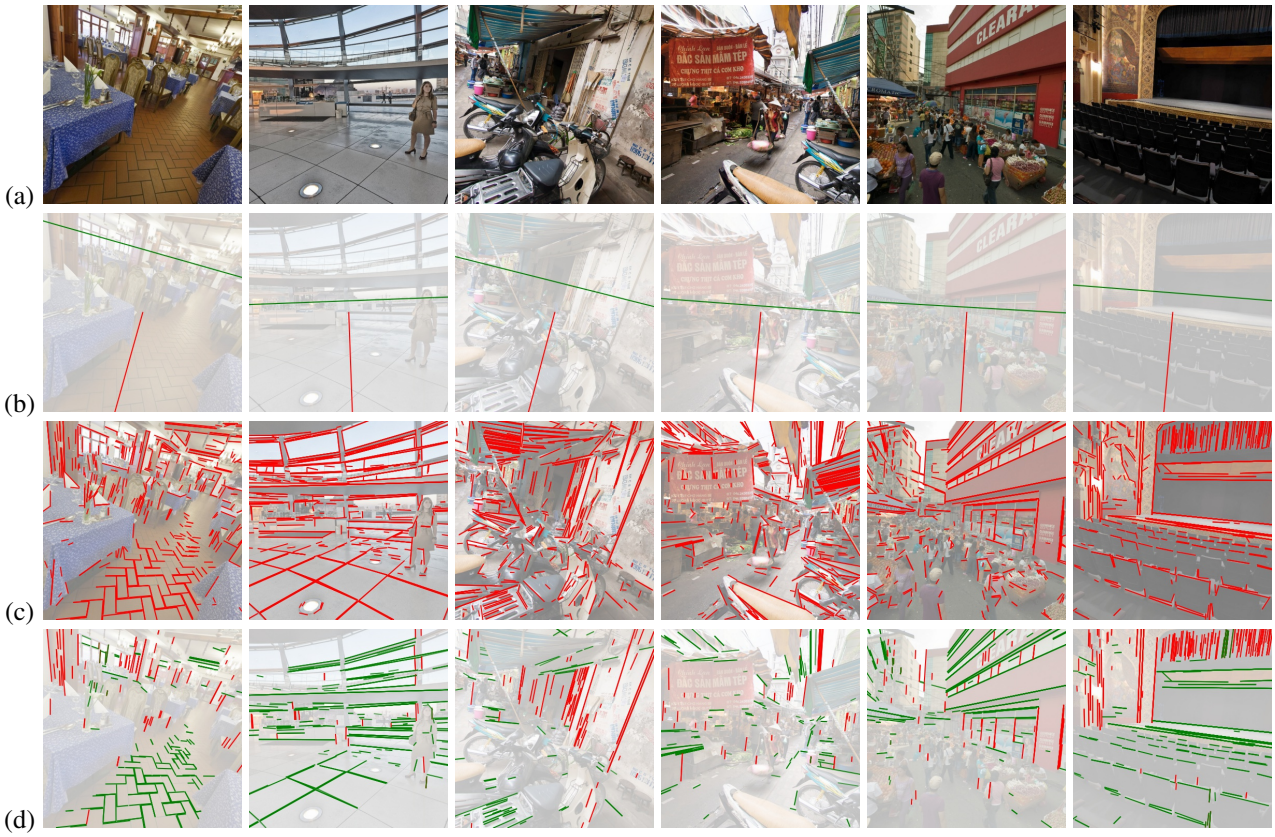

Figure S6. More results with SUN360 [9] test set: (a) input image, (b) estimated horizon line (green) and vertical direction along with the zenith VP (red), (c) detected lines with LSD [7], (d) estimated vertical (red) and horizontal (green) convergence line segments of (c).

References

[1] Google Street View Images API. https://developers.google.com/maps/.

[2] Yannick Hold-Geoffroy, Kalyan Sunkavalli, Jonathan Eisenmann, Matthew Fisher, Emiliano Gambaretto, Sunil Hadap, and Jean-François Lalonde. A Perceptual Measure for Deep Single Image Camera Calibration. In *Proc. CVPR*, pages 2354–2363, 2018.

[3] Hyunjoon Lee, Eli Shechtman, Jue Wang, and Seungyong Lee. Automatic Upright Adjustment of Photographs with Robust Camera Calibration. *IEEE Trans. Pattern Analysis Machine Intelligence*, 36(5):833–844, 2014.

[4] Jinwoo Lee, Minhyuk Sung, Hyunjoon Lee, and Junho Kim. Neural Geometric Parser for Single Image Camera Calibration. In *Proc. ECCV*, pages 541–557, 2020.

[5] Gilles Simon, Antoine Fond, and Marie-Odile Berger. A-Contrario Horizon-First Vanishing Point Detection Using Second-Order Grouping Laws. In *Proc. ECCV*, pages 318–333, 2018.

[6] Elena Tretyak, Olga Barinova, Pushmeet Kohli, and Victor Lempitsky. Geometric Image Parsing in Man-Made Environments. *International Journal of Computer Vision*, 97:305–321, 2012.

[7] Rafael Grompone von Gioi, Jérémie Jakubowicz, Jean-Michel Morel, and Gregory Randall. LSD: A Fast Line Segment Detector with a False Detection Control. *IEEE Trans. Pattern Analysis Machine Intelligence*, 32(4):722–732, 2010.

[8] Scott Workman, Menghua Zhai, and Nathan Jacobs. Horizon Lines in the Wild. In *Proc. BMVC*, pages 20.1–20.12, 2016.

[9] Jianxiong Xiao, Krista A. Ehinger, Aude Oliva, and Antonio Torralba. Recognizing Scene Viewpoint using Panoramic Place Representation. In *Proc. CVPR*, pages 2695–2702, 2012.

[10] Yichao Zhou, Jingwei Huang, Xili Dai, Linjie Luo, Zhili Chen, and Yi Ma. HoliCity: A City-Scale Data Platform for Learning Holistic 3D Structures. *arXiv*, 2020.