Extreme Structure from Motion for Indoor Panoramas without Visual Overlaps [Supplementary Material]

Mohammad Amin Shabani
Simon Fraser University
mshabani@sfu.ca

Weilian Song
Simon Fraser University
weilians@sfu.ca

Makoto Odamaki
Ricoh Company, Ltd.
makoto.odamaki@jp.ricoh.com

Hirochika Fujiki
Ricoh Company, Ltd.
hirochika.fujiki@jp.ricoh.com

Yasutaka Furukawa
Simon Fraser University
furukawa@sfu.ca

The supplementary document provides 1) details of our annotation tools; 2) system implementation details; 3) intermediate results by pre-processing networks; 4) visualization of what our arrangement classifier learned; and 5) more experimental results.

1. Annotation tools

We use two annotation tools to create annotations at the level of panoramas and at the level of houses/apartments. Both tools are implemented with PyQT5 and Python.

Panorama-level Annotator: We used a modified version of PanoAnnotator [1] to annotate room type, room layout, and door/window bounding-boxes/segmentations. We fix the layout height to 3.2 and allow the specification of different object types. Figure 1 shows its screenshot.

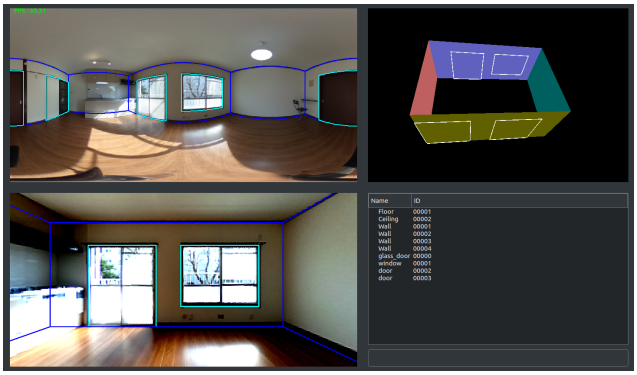

Figure 1. Panorama-level Annotator

House-level annotator: We implemented an annotation toolbox to add, rotate, and move the nadir semantic images on the floorplan image. The toolbox also allows the rescaling of the floorplan image. The 2D positions and the heading angles are annotated with respect to the floorplan. Figure 2 shows its screenshot.

Figure 2. House-level Annotator

2. System details

Table 1 provides the architecture specifications of our arrangement evaluator. The network takes an arrangement of size 256×256 × 16 as the input. Each convolutional layer has a kernel of size 3×3 with padding and stride of 1. The GroupNorm layers [2] have 4 groups except the last GroupNorm layer which has only 1 group.

3. Intermediate results

Figure 5 shows the intermediate results by pre-processing networks, namely, room layouts, door/window type/detections/segmentations, and the Nadir semantic images. Estimated layouts are visualized by the implementation of HorizonNet [1]. In addition, Figure 3 shows the confusion matrix obtained from the room classifier network over the test set. The largest confusion happens between the dining room and Western-style room when the room is large and connected to the kitchen. In these cases, it could be considered as both dining room or western-style room

Table 1. Our arrangement evaluator architecture, which is shown in the top table as the list of modules. The bottom table shows that each module consists of a list of layers.

Modules	Resolutions
ConvSet 32	256×256×32
MSP 32	256×256×32
DonwSampling 64	128×128×64
MSP 64	128×128×64
DonwSampling 128	64×64×128
MSP 128	64×64×128
DonwSampling 256	32×32×256
MSP 256	32×32×256
DonwSampling 128	16×16×128
MSP 128	16×16×128
DonwSampling 1	8×8×1
Flatten	1×64
Linear [64×1]	1

Module	Layers
ConvBlock C	Convolution C GroupNorm ReLu
ConvSet C	ConvBlock C ConvBlock C
MSP C	MessagePassing ConvSet C
DownSampling C	MaxPooling ConvSet C

based on the main overview of the house.

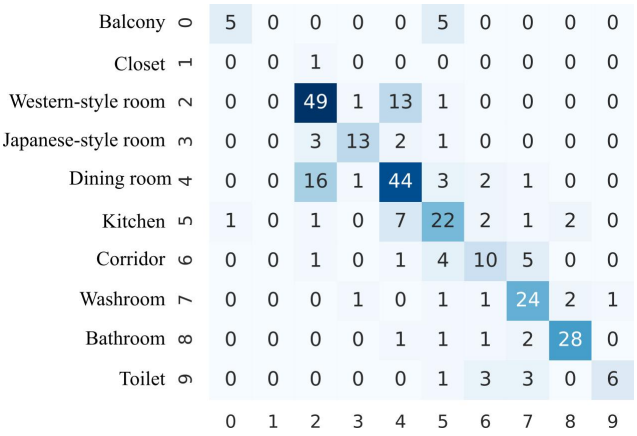

Figure 3. Confusion matrix obtained from our panorama classifier.

4. What the arrangement classifier learned

Figure 4 visualizes what the arrangement classifier learned with a simple experiment. We take a panorama arrangement and moves around a particular panorama/room

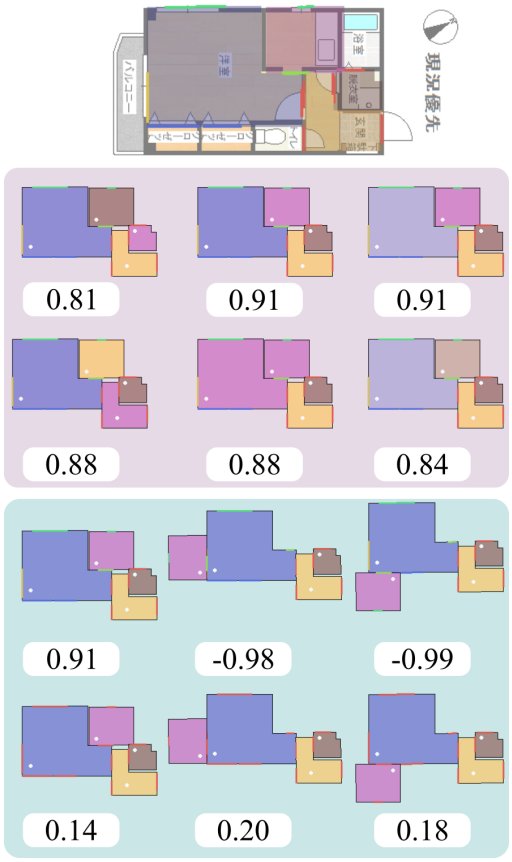

Figure 4. Visualization of what the arrangement classifier learned based on the room (pink box) and door types (green box). Each number shows the predicted score of the evaluator network given the corresponding input.

while observing the evaluation scores. This simple experiment reveals that the scores goes down when the room types in the arrangements are not a standard way or the door types are not provided properly. As it is shown in the second row of the green box in Figure 4, when we consider all of the doors as a same category, the network can not distinguish the positive arrangement (first column) among the three arrangements. However, this is significantly different when we have the corrosponding class of each door (first row).

5. More experimental results

Figure 6 shows comparisons against the three competing methods for more test samples in the same format as Fig. 8 in the main paper. Figure 7 shows qualitative evaluations for more test samples in the same format as Fig. 9 in the main paper.

References

[1] Cheng Sun, Chi-Wei Hsiao, Min Sun, and Hwann-Tzong Chen. Horizonnet: Learning room layout with 1d representa-

Figure 5. Intermediate results of converting the input panorama to the semantic Nadir image.

tion and pano stretch data augmentation. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, pages 1047–1056, 2019. 1

[2] Yuxin Wu and Kaiming He. Group normalization. In *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018. 1

Figure 6. Qualitative comparisons against the three competing methods. We show the top-2 reconstructions from each method based on their scoring functions. Room colors indicate their types

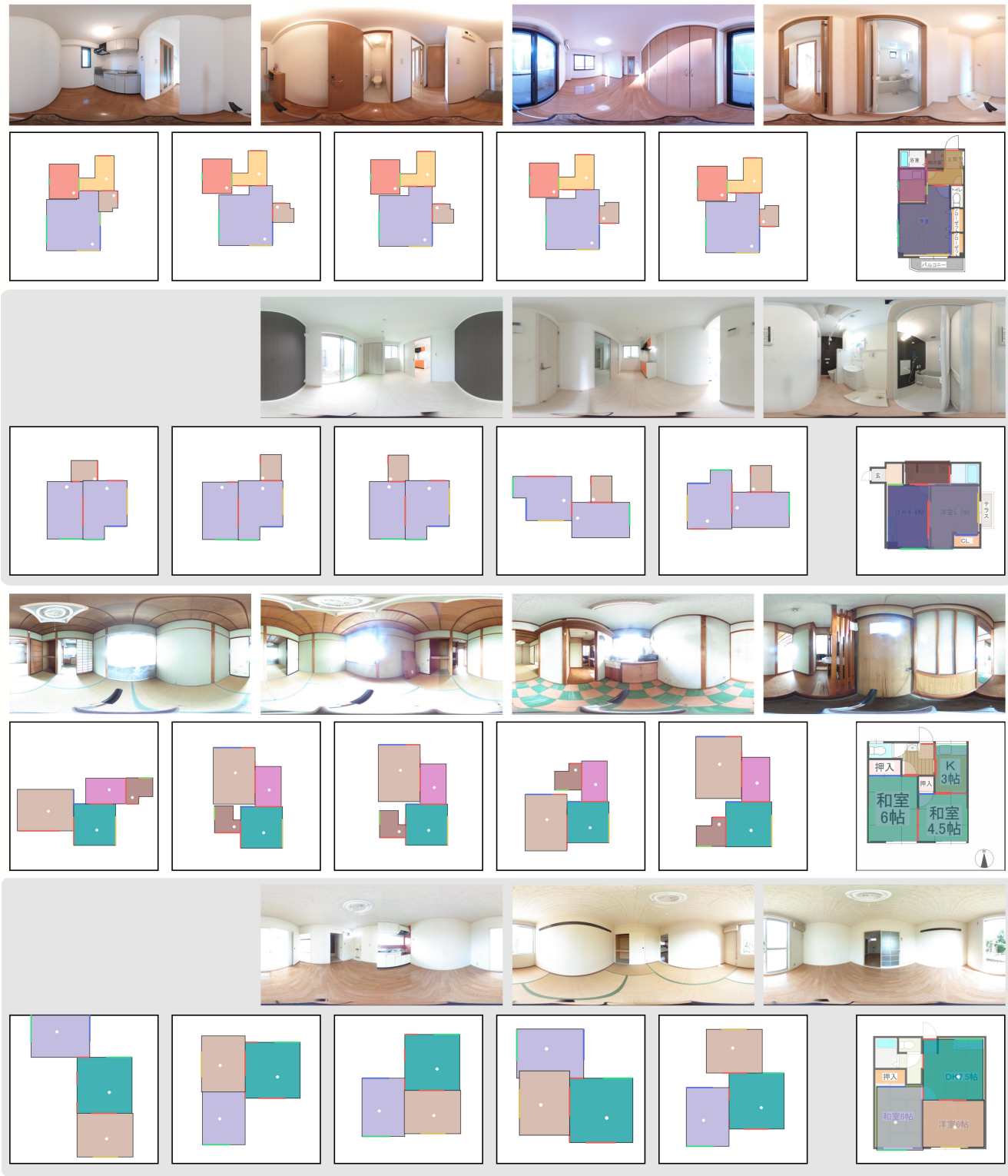

Figure 7. Qualitative evaluations. Top-5 reconstructions by our method against the ground-truth arrangement.

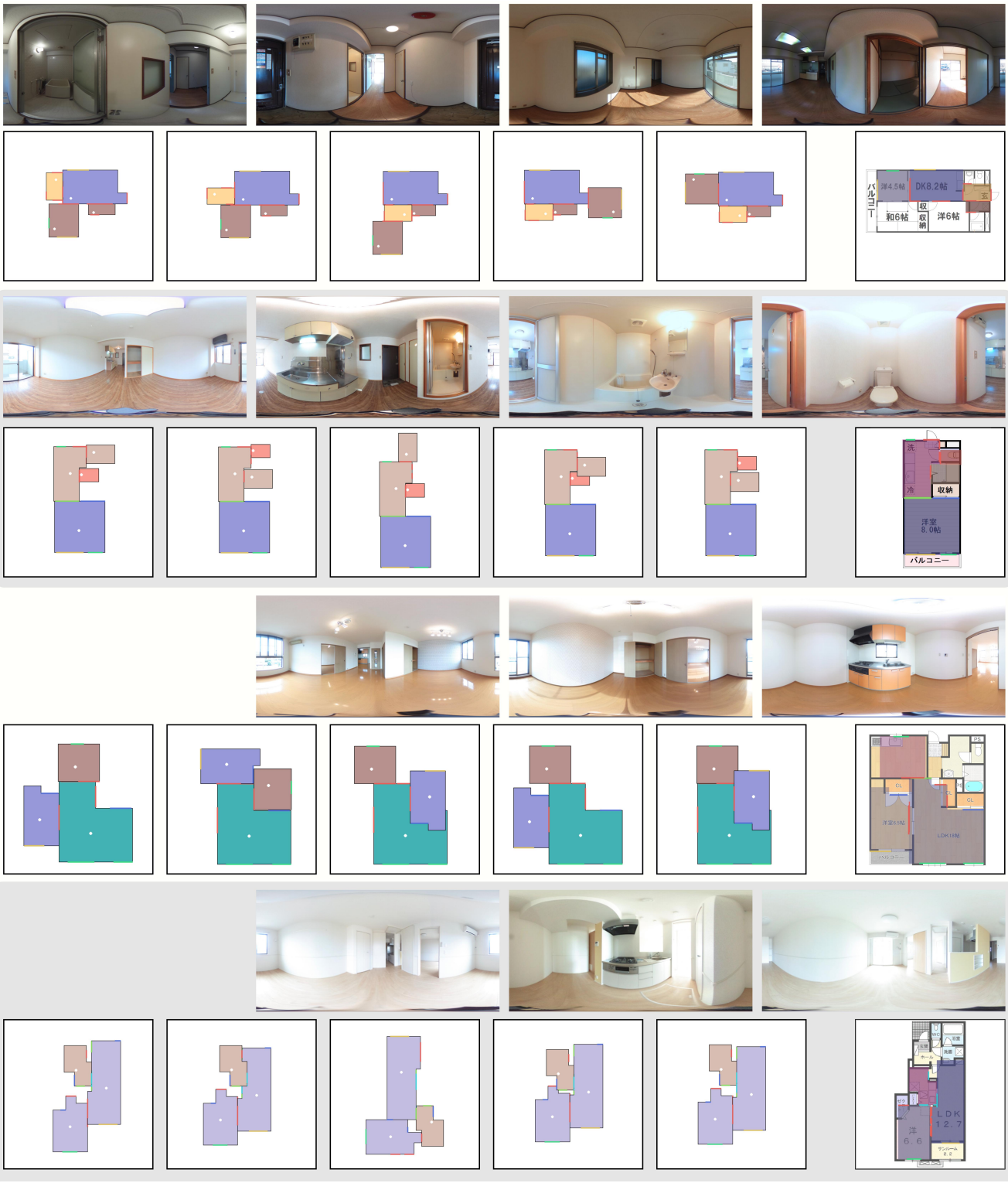

Figure 8. Continued.