Deep Multi-Modal Image Correspondence Learning Supplementary Material

Anonymous CVPR submission

Paper ID 138

In the supplementary material, we show more comprehensive experimental results and evaluations for 1) receptive field visualization (Figure 1 and Figure 2), 2) object discovery visualization (Figure 3), 3) image placement application (Figure 4), 4) image retrieval application (Figure 5), and 5) localization application (Figure 6 and Figure 7).

References

[1] C. Barnes, E. Shechtman, A. Finkelstein, and D. Goldman. Patchmatch: a randomized correspondence algorithm for structural image editing. *ACM TOG*, 28(3):24, 2009. 3

[2] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba. Object detectors emerge in deep scene cnns. In *ICLR*, 2015. 1, 6

(a) Floorplan (b) Photograph (c) Receptive Field

Figure 1: More receptive field visualization results. We adapt the idea in [2] to visualize the learned Receptive Fields (RFs). The network learns to attend to the room in a floorplan corresponding to the photograph. These results are for bathrooms.

Figure 2: More receptive field visualization results. Rows 1-2 are for kitchens and 3-4 are for living-rooms. Row 5-7 show failure cases (bathrooms, kitchens, and living-rooms, respectively).

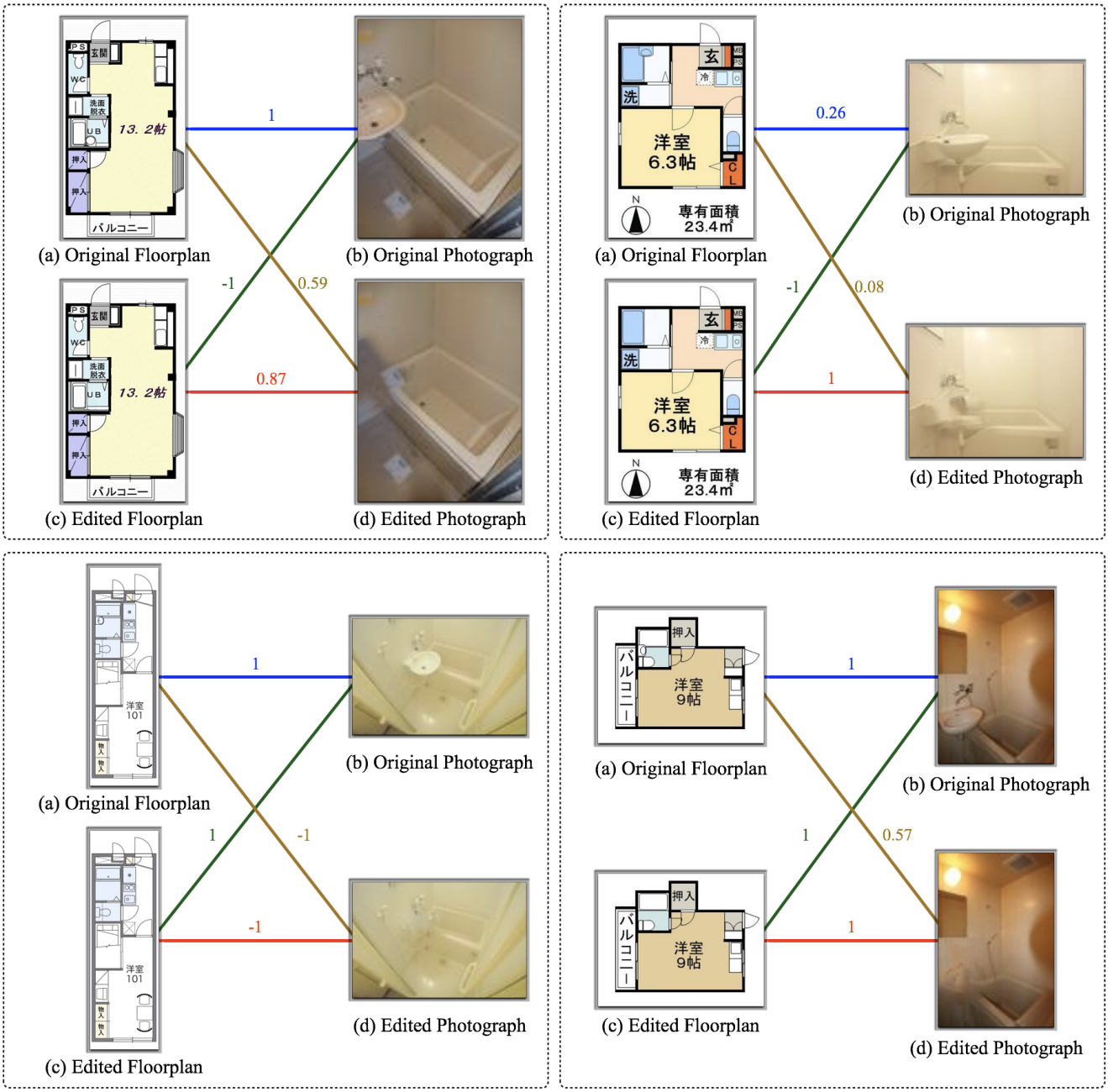

Figure 3: More object discovery visualization results. We remove an object from a photograph by PatchMatch [1] and the corresponding object icon from a floorplan manually. The similarity scores often suggest that the network has learned to detect objects and match their detections to overcome the modality differences. Two examples in the bottom row are failure cases where the similarity scores fail to follow the expected pattern.

Figure 4: More image placement results. The success of RF visualization allows us to visualize photographs in the context of a floorplan image, an effective application for real estate websites.

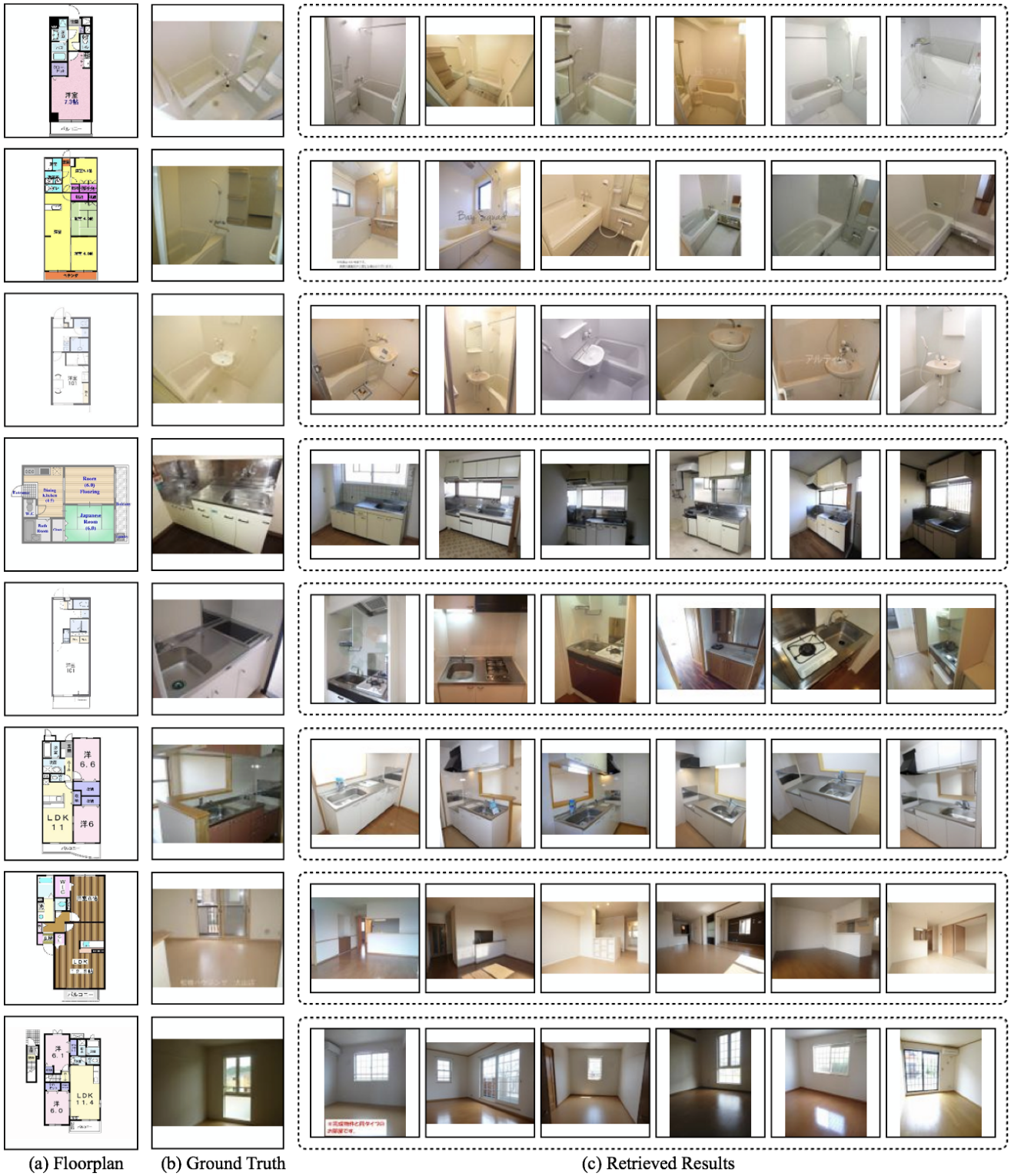

Figure 5: More image retrieval results. Our network is able to show likely room apperances only from a floorplan. Rows 1-3, 4-6, and 7-8 are for bathrooms, kitchens, and living-rooms, respectively.

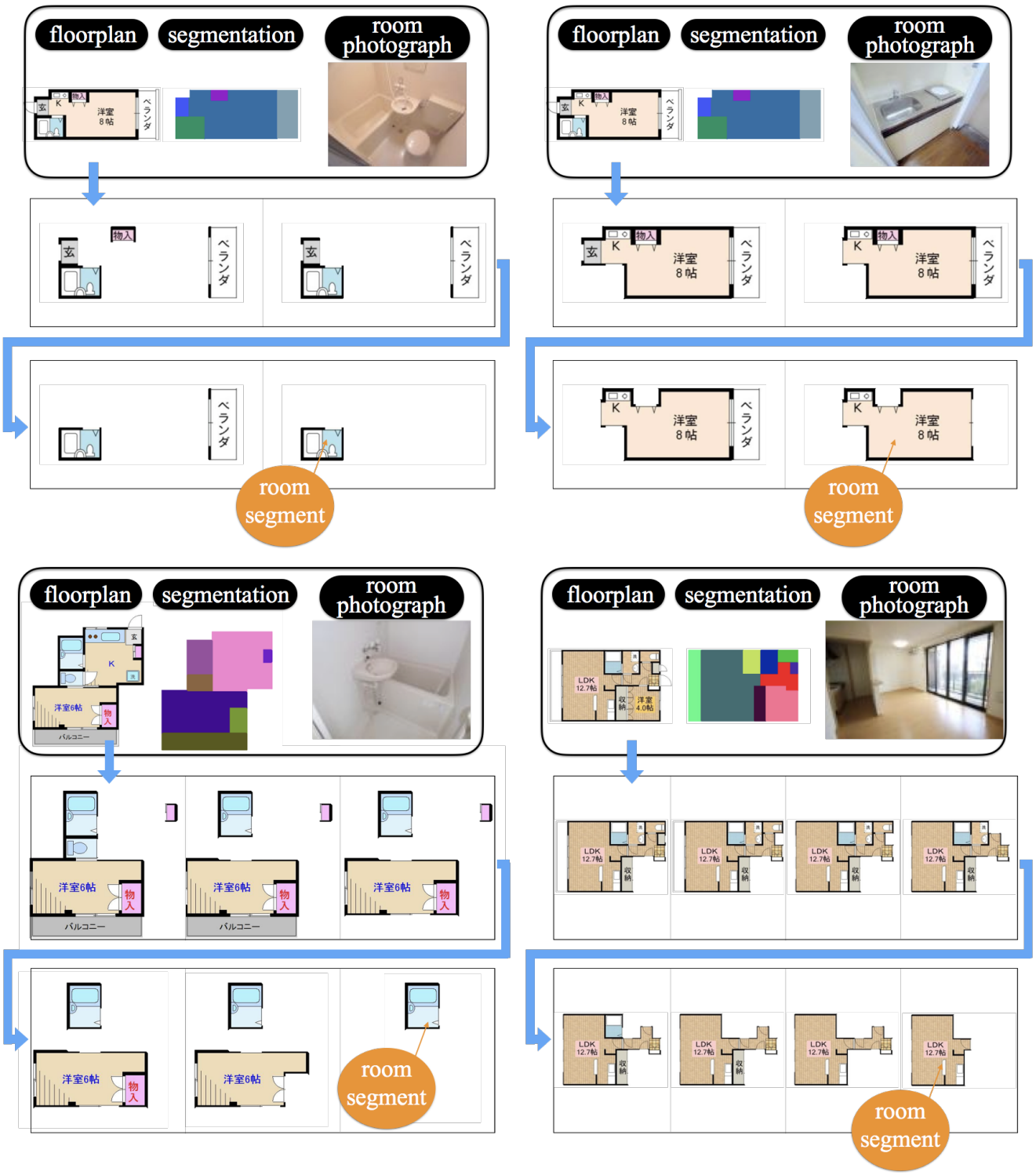

Figure 6: More localization results. The simplification technique [2] is particularly effective for floorplan images, enabling the localization of an exact image region corresponding to the photograph.

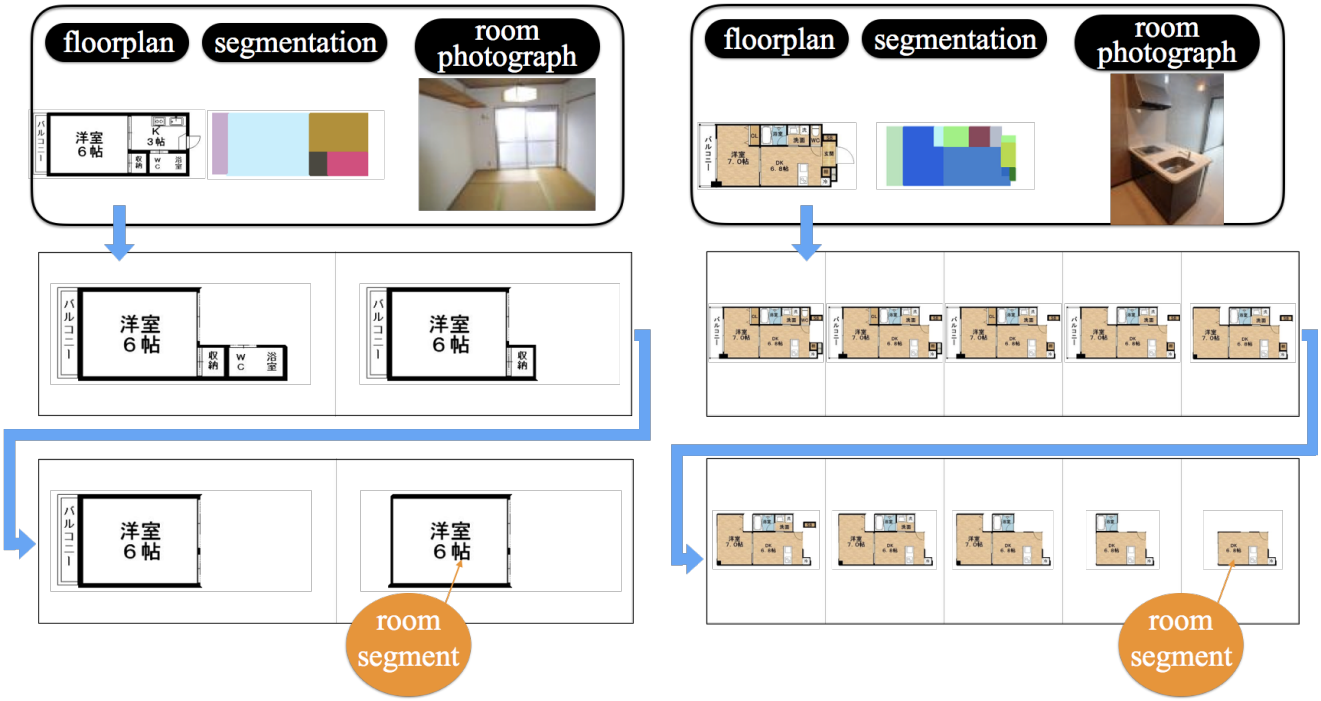

Figure 7: More localization results.