Geometry-Aware Distillation for Indoor Semantic Segmentation (Supplementary Material)

Anonymous CVPR submission

Paper ID 1440

Overview

In this supplementary material, we provide more experimental results in addition to the main paper and present more detailed analysis.

1. More Qualitative Results

In this section, we show more qualitative results of the proposed method for semantic segmentation, on the datasets of both NYU-Dv2 [1] and SUN RGBD [2]. The results on NYU-Dv2 are shown in Figure. 1 – Figure. 2, while results on SUN RGBD are shown in Figure. 3 – Figure. 4. As the categories in NYU-Dv2 and SUN RGBD are slightly different, the color coding of categories in the segmentation qualitative results are not exactly the same between these two datasets, but are consistent inside each dataset. From the results we can see that our method can well classify most of the categories and recover the shape of the objects as much as possible. Note that the category annotation in the SUN RGBD dataset is not as dense/precise as that in the NYU-Dv2 dataset. Some categories are mislabeled or omitted. A "failure" case of our method is shown in the last example (the right one in the last row of Figure. 4). In the RGB image we can observe that books are placed on the bookshelf. However in the ground-truth labeled map only the bookshelf is accurately labeled, without labeling the books. Whereas in the left example (same row) the books are well-labeled without labeling the bookshelf. Such labeling inconsistency leads our performance to be relatively lower to some extent.

2. Analysis on the Learned Depth-Aware Embedding

In this section we present more analysis on the learned depth-aware embeddings. We visualize the embeddings together with the corresponding RGB images and semantic segmentation maps, as shown in Figure. 5. We can observe from the results that the depth-aware embeddings learned by our model implicitly embed 3D depth information. With such embeddings, the semantic segmentation can be improved by taking 3D geometry into consideration, in addition to the 2D appearance. For example, in the second row the *table* and *refrigerator* share similar 2D appearances, which makes them difficult to be discriminated from each other. In the learned embedding space we can see these two objects are distant from each other, which can be leveraged to assist the semantic segmentation. Similar examples can also be noticed like the *pillows* on the *bed*, *refrigerator* next to *cabinets*, etc. As a result, the learned embeddings enhance both 2D and 3D consistencies.

References

[1] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus. Indoor segmentation and support inference from rgbd images. In *ECCV*, 2012. 1

[2] S. Song, S. P. Lichtenberg, and J. Xiao. Sun rgb-d: A rgb-d scene understanding benchmark suite. In *CVPR*, 2015. 1

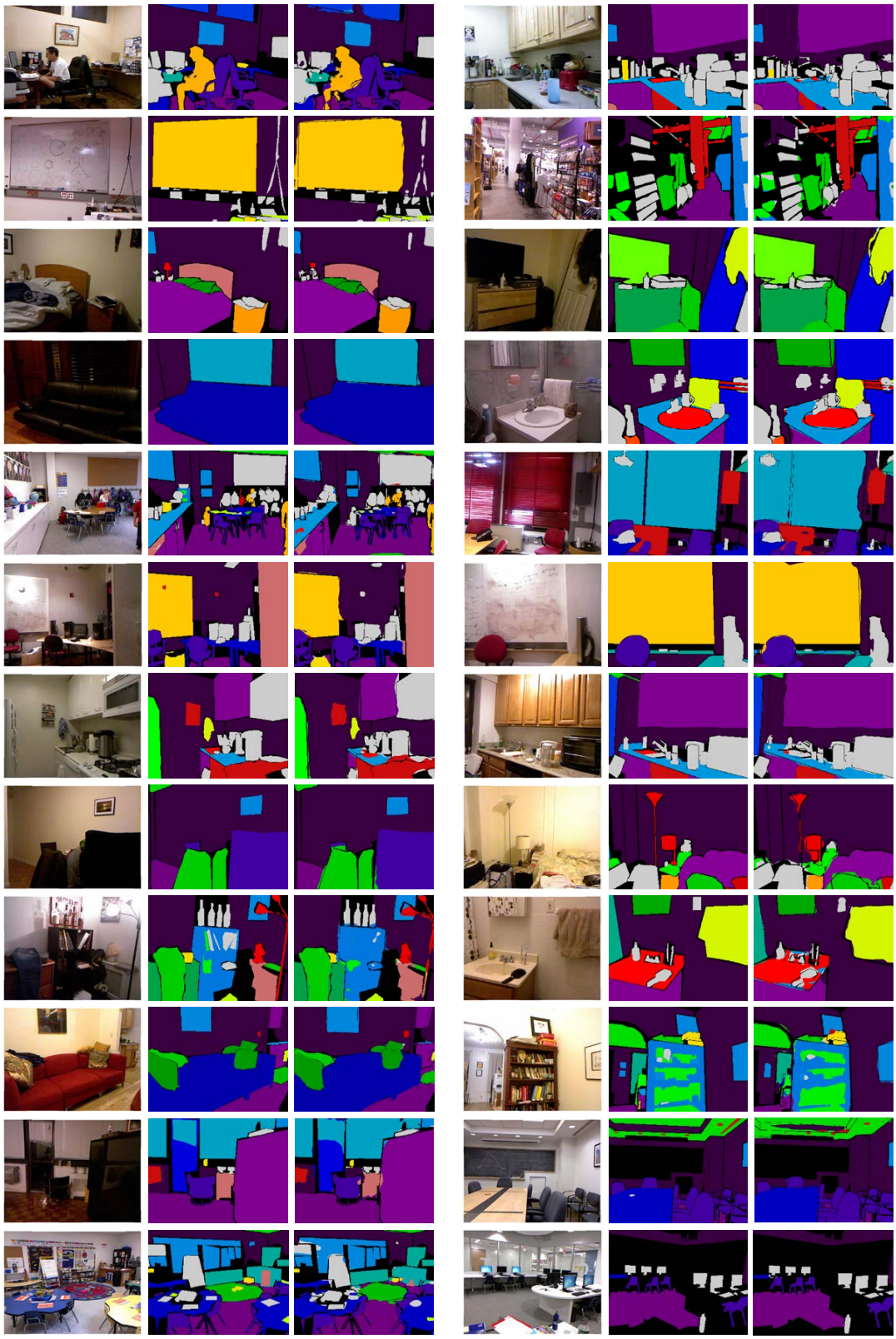

Figure 1. More qualitative results on NYU-Dv2 dataset. Each example from left to right: input RGB image, ground-truth semantic segmentation, and our result.

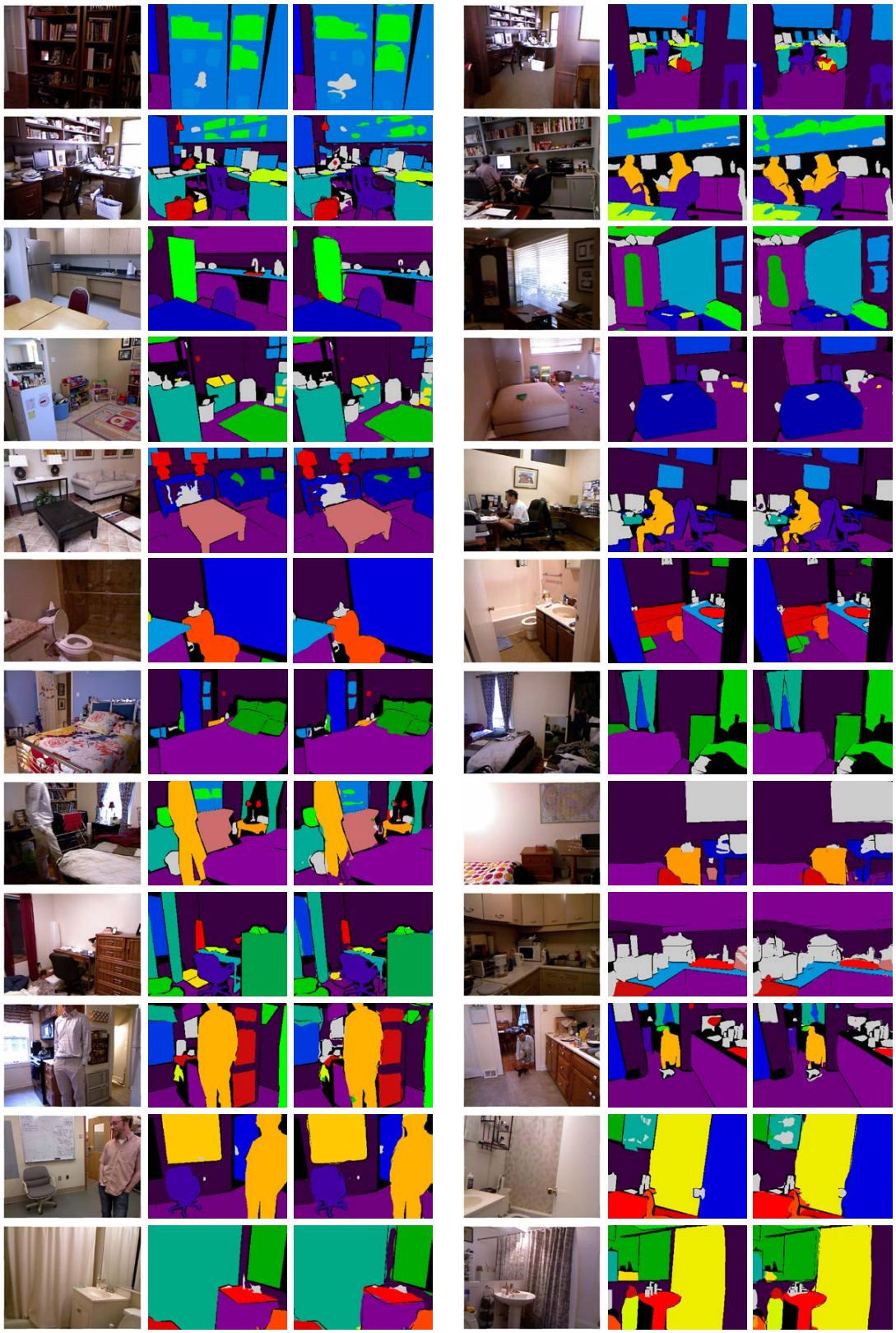

Figure 2. More qualitative results on NYU-Dv2 dataset. Each example from left to right: input RGB image, ground-truth semantic segmentation, and our result.

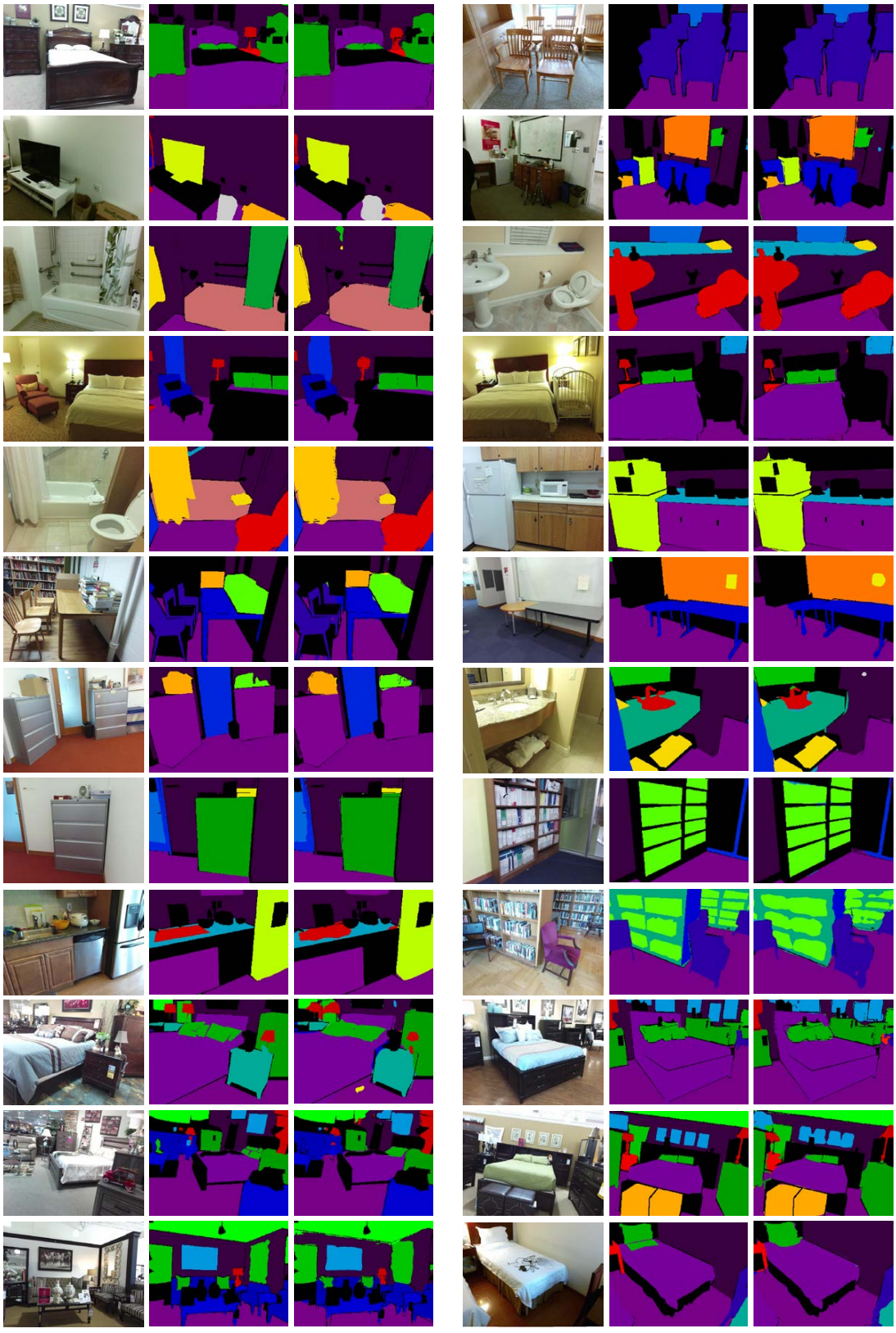

Figure 3. Qualitative results on SUN RGBD dataset. Each example from left to right: input RGB image, ground-truth semantic segmentation, and our result.

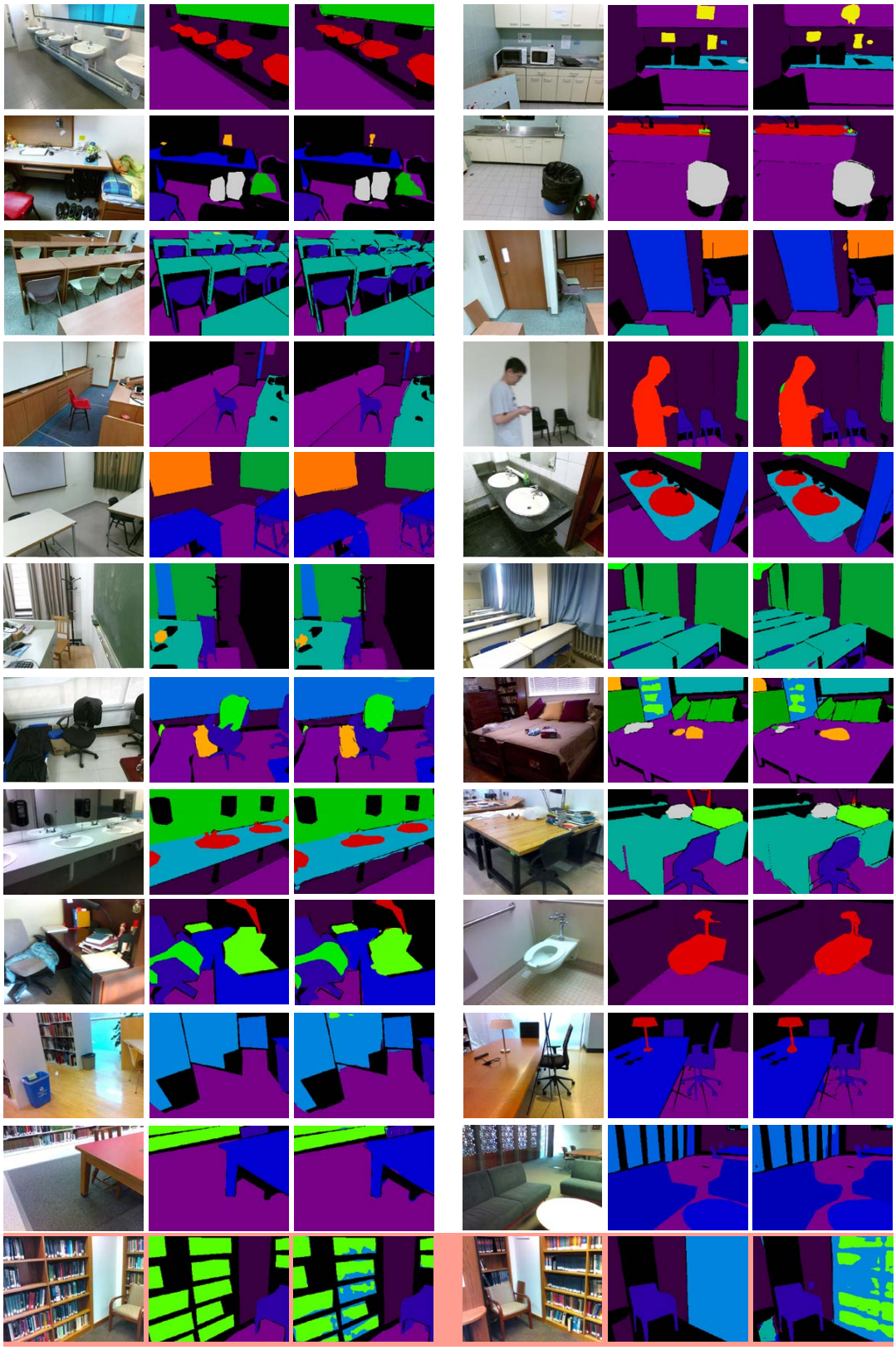

Figure 4. Qualitative results on SUN RGBD dataset. Each example from left to right: input RGB image, ground-truth semantic segmentation, and our result.

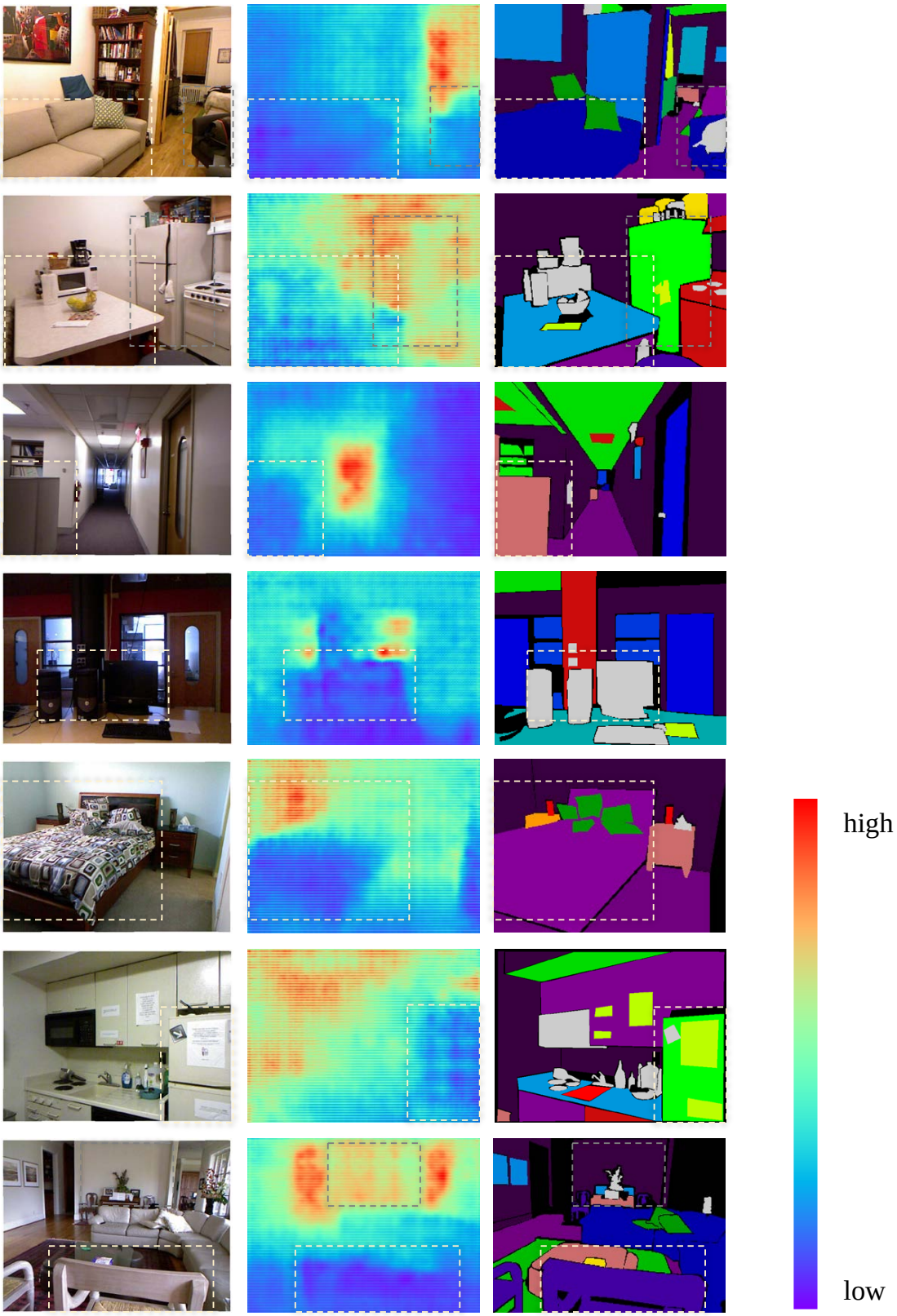

Figure 5. Visualization of the learned depth-aware embeddings. Each row from left to right: input RGB image, learned embedding, and the corresponding semantic segmentation.