Supplemental Material for 'Patch2CAD: Patchwise Embedding Learning for In-the-Wild Shape Retrieval from a Single Image'

Weicheng Kuo[1], Anelia Angelova[1], Tsung-Yi Lin[1], Angela Dai[2]
[1] Google Research, Brain Team
[2] Technical University of Munich
{weicheng, anelia, tsungyi}@google.com, angela.dai@tum.de

1. Additional Top-K Retrieval Qualitative Results

In Figure 1, we show additional qualitative results of our Patch2CAD top-K retrieval vs Mask2CAD. We observe that ours can retrieve better and more consistent shapes in the top-K pool than Mask2CAD.

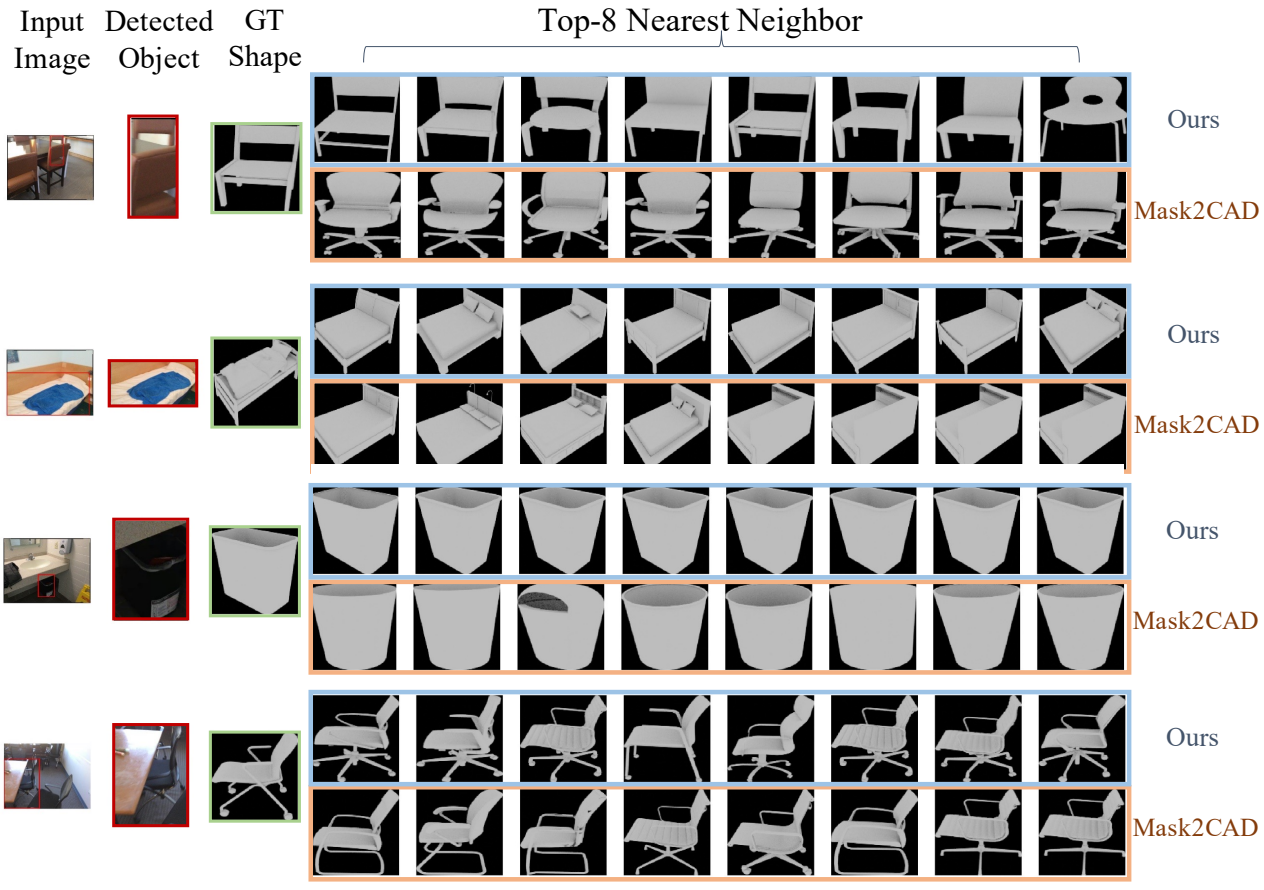

Figure 1: Additional Patch2CAD Top-K retrival qualitative results on various ScanNet [2] images in comparison with Mask2CAD.

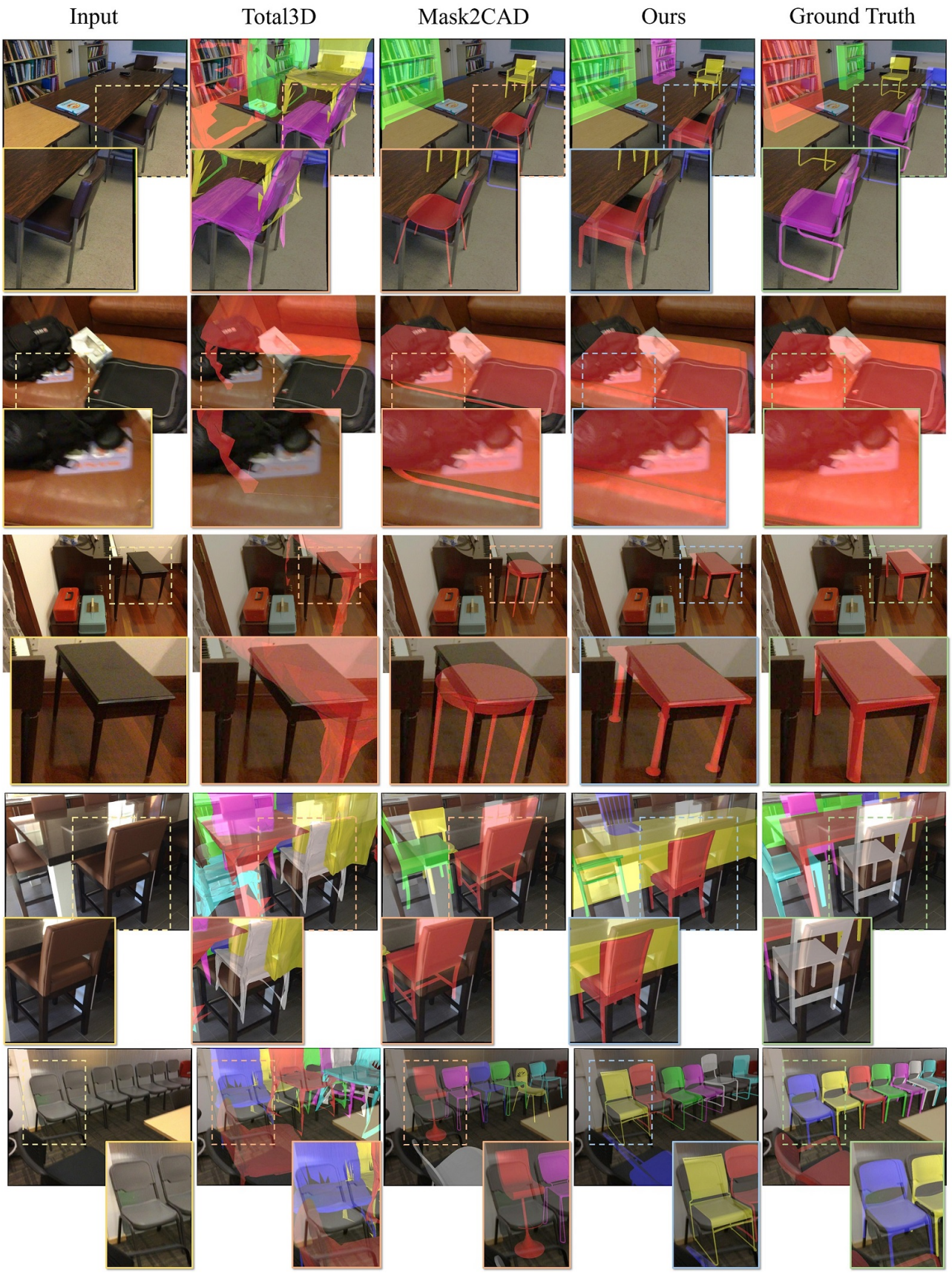

Figure 2: Additional qualitative results of Patch2CAD (ours) on various ScanNet [2] images.

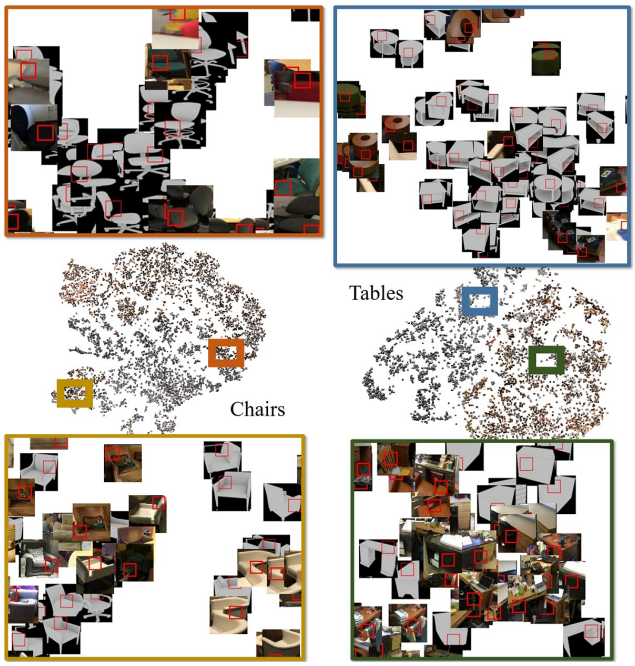

Figure 3: t-SNE embeddings of our patch-wise embedding of images and CAD shapes (patches demarcated in red) for the 'chairs' and 'tables' categories.

2. Additional Qualitative Results

In Figure 2, we show additional qualitative results of Patch2CAD on ScanNet [2] images, with Scan2CAD [1] targets. Ours is able to retrieve better matching shapes to the groundtruth than Mask2CAD [3] or Total3D [4].

3. t-SNE embedding of Patch2CAD

We visualize several t-SNE embeddings in Figure 3, where CAD patches can tend to cluster near each other (there are many locally very similar patches), but also near similar image patches (*e.g.*, chair seat corner, tabletop).

4. Effect of the number of query (K_q) and retrieved patches (K_r).

We use one model for all inference time ablation studies in this section. All parameters are the same as the main paper unless stated otherwise. The noise across independent runs are ≈ 0.1 Mesh AP.

Table 1 analyzes query K_q patches per detection at test time. We see that more patches result in better retrieval.

Table 2 shows improvement with retrieved K_r per test query, due to robustness of voting when K_r is high.

K_q	1	3	6	9	12
AP	9.2	9.8	10.2	**10.3**	10.2

Table 1: Mesh AP vs the number of query patches.

K_r	1	3	6	12	24	48	96
AP	9.3	9.4	9.8	10.0	10.3	**10.6**	10.6

Table 2: Mesh AP vs the number of retrieved patches.

References

[1] Armen Avetisyan, Manuel Dahnert, Angela Dai, Manolis Savva, Angel X. Chang, and Matthias Nießner. Scan2cad: Learning cad model alignment in rgb-d scans. *CVPR*, 2019. 3

[2] Angela Dai, Angel X. Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. ScanNet: Richly-annotated 3D reconstructions of indoor scenes. In *Proc. Computer Vision and Pattern Recognition (CVPR), IEEE*, 2017. 1, 2, 3

[3] Weicheng Kuo, Anelia Angelova, Tsung-Yi Lin, and Angela Dai. Mask2CAD: 3D shape prediction by learning to segment and retrieve. In *Eur. Conf. Comput. Vis.*, 2020. 3

[4] Yinyu Nie, Xiaoguang Han, Shihui Guo, Yujian Zheng, Jian Chang, and Jian Jun Zhang. Total3dunderstanding: Joint layout, object pose and mesh reconstruction for indoor scenes from a single image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 55–64, 2020. 3