CAOA - Completion-Assisted Object-CAD Alignment

Hiranya Garbha Kumar
University at Albany
Albany, NY, USA
hgkumar@albany.edu

Minhas Kamal
University at Albany
Albany, NY, USA
mxkamal@albany.edu

Balakrishnan Prabhakaran
University at Albany
Albany, NY, USA
bprabhakaran@albany.edu

1. Appendix Section

1.1. Qualitative Analysis

Figure 1 presents qualitative CAD alignment results, comparing CAOA (c) and CIS2VR [2] (a) with ground truth annotations (b). The input object point clouds are instance predictions generated by SoftGroup [3] for indoor scenes from the ScanNet dataset [1]. For each set of visualizations, CIS2VR results are shown on the left (a), ground truth (GT) in the middle (b), and CAOA results on the right (c). Object point clouds are depicted in black, completed point clouds (for CAOA) are in red, and CAD objects are aligned using the predicted pose in gray. The figures demonstrate a significant improvement in alignment performance with CAOA, particularly in scenarios involving incomplete instance point clouds. These results underscore the advancements achieved through the proposed approach.

1.2. S2C-Completion Dataset

We provide further examples from the S2C-Completion dataset in Figure 2. The figures show scan object color point clouds with aligned CAD models in gray. The single object pose of each object instance is visualized using a green bounding box surrounding each object.

1.3. Alignment Performance on Synthetic Datasets

CPCM Train	Avg↑	Weighted Avg↑
PCN	55.68	47.69
ShapeNet-55/34	70.82	63.77
SN-Indoor	70.78	63.89
S2C-Completion	71.05	64.38
SN-Indoor + S2C-C + Ctxt 100 cm	77.51	69.17

Table 1. Alignment performance of CAOA with CPCM trained on different datasets. We observe that the performance of CPCM on point cloud completion has significant impact on the final alignment performance.

1.4. SN-Indoor

Figure 3 presents a comparative analysis of our dataset alongside widely used synthetic datasets and a real-world dataset. Real-world scans are captured using an RGB-D sensor, with the intrinsic and extrinsic parameters of the camera utilized to transform image coordinates into world coordinates. Multiple camera frames capture distinct sets of world coordinates, which are integrated to generate the final point cloud.

1.4.1 Perspective Point cloud Generation

In our approach, we aim to replicate these real-world conditions while generating incomplete point clouds from synthetic mesh data. The following equations illustrate the transformation technique utilized in our approach:

$$PCD_o = \sum_{n=1}^{N} R_n^{-1}(d.K^{-1}(x,y) - t_n) \tag{1}$$

Here, PCD_o denotes the occluded point cloud obtained after ray casting. n represents the number of frames captured from a single camera perspective. K and $[R|t]$ correspond to the camera model parameters. d is depth of the (x,y) image coordinate.

1.4.2 Non-Linear Cropping

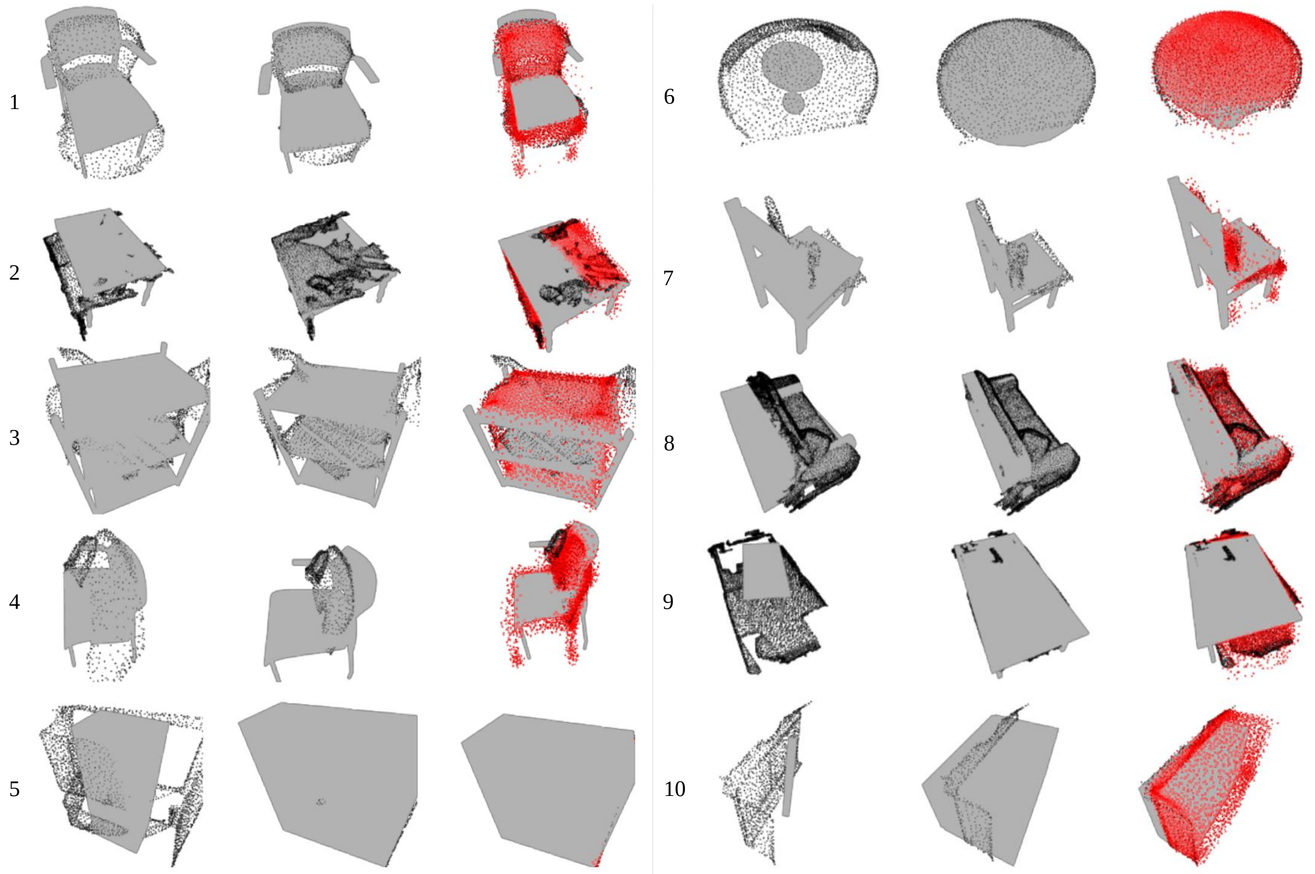
We use the following equations to define the non-linear surface (f) used for cropping. $r_1...r_9$ are randomly generated values used to define the shape of the plane. G_σ is the added gaussian noise, where σ is set to 0.005.

$$fs(x) = r_1 * \sin(r_2 * x + r_3) \tag{2}$$

$$fc(y) = r_4 * \cos(r_5 * y + r_6) \tag{3}$$

$$f(x,y,z) = fs(x) + fc(y) + r_7 * x + r_8 * y + r_9 \tag{4}$$

$$PCD_i = PCD_o[f(PCD_o)] + G_\sigma \tag{5}$$



(a) CIR2VR (b) Ground Truth (c) CAOA (a) CIR2VR (b) Ground Truth (c) CAOA

Figure 1. A qualitative comparison of alignment results using CIS2VR[2], ground truth and CAOA. Incomplete instance point clouds are shown in black, and completed point clouds in red. Predicted poses are in gray.

Here, PCD_i represents the final incomplete point cloud after cropping and adding Gaussian noise.

1.5. Code Availability

We release the code corresponding to CAOA to the public via Github.

References

[1] Angela Dai, Angel X Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5828–5839, 2017. 1

[2] Hiranya Kumar, Ninad Khargonkar, and Balakrishnan Prabhakaran. Cis2vr: Cnn-based indoor scan to vr environment authoring framework. In *2024 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality (AIxVR)*, pages 128–137. IEEE, 2024. 1, 2

[3] Thang Vu, Kookhoi Kim, Tung M Luu, Thanh Nguyen, and Chang D Yoo. Softgroup for 3d instance segmentation on point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2708–2717, 2022. 1

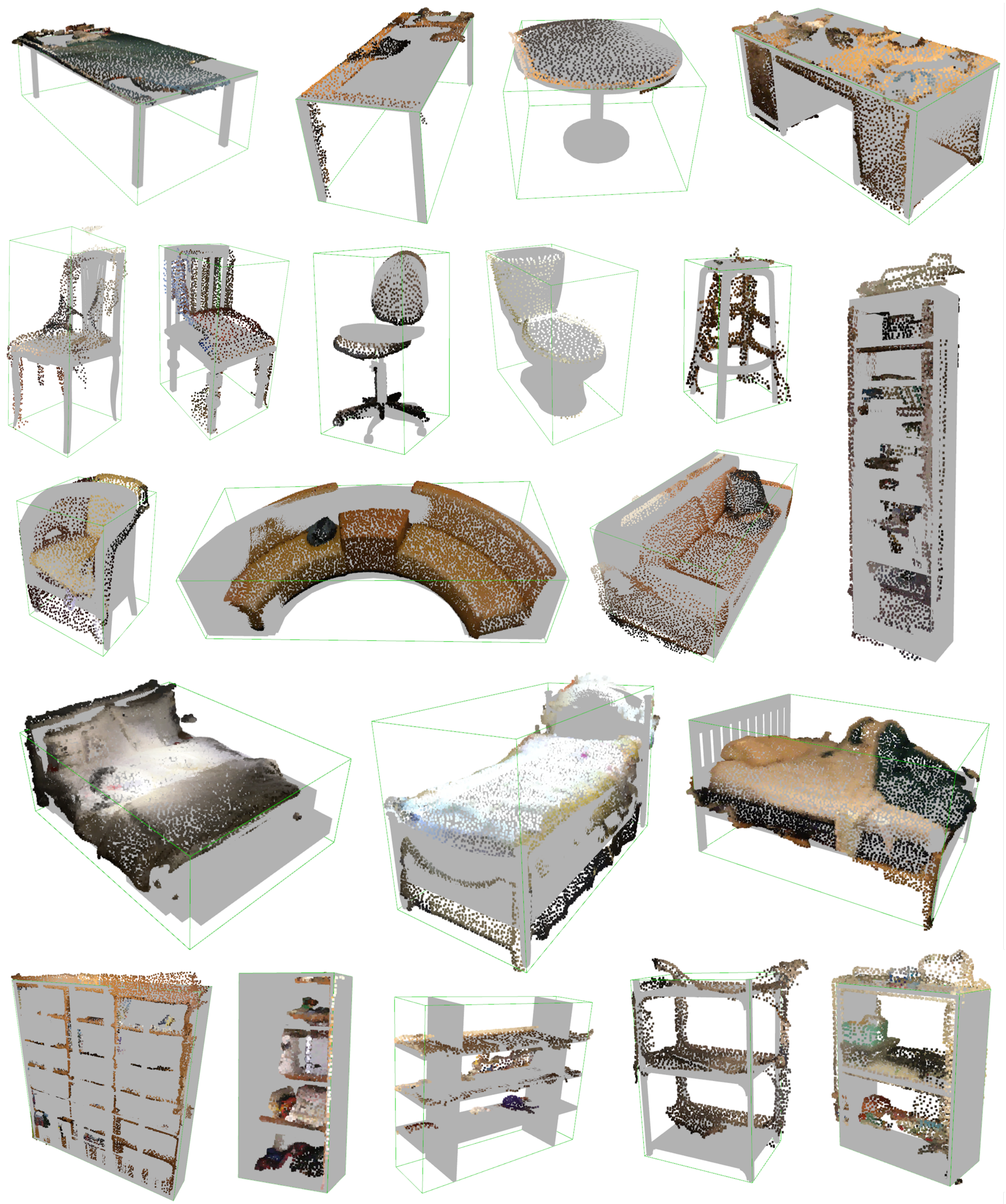
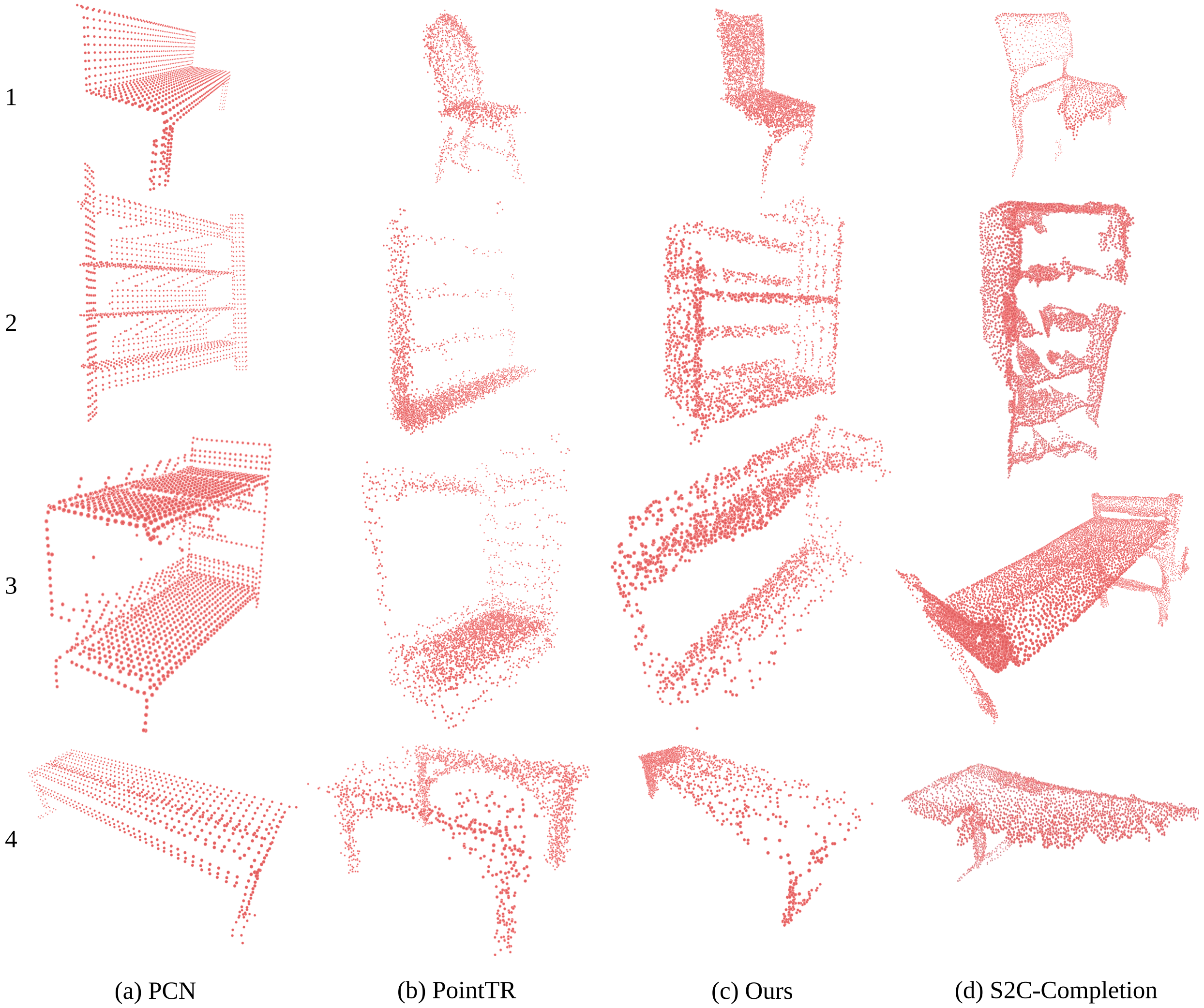


Figure 2. Examples from S2C-Completion dataset showing colored scan instance point clouds and aligned CAD models (gray). Each object's 9-DoF pose is visualized as a green bounding box.



(a) PCN (b) PointTR (c) Ours (d) S2C-Completion

Figure 3. A qualitative comparison of synthetic point cloud completion datasets. The last column is from real-world scans.