SceneNN: a Scene Meshes Dataset with aNNotations

Supplementary document

Binh-Son Hua[1] Quang-Hieu Pham[1] Duc Thanh Nguyen[2] Minh-Khoi Tran[1]
Lap-Fai Yu[3] Sai-Kit Yeung[1]

[1]Singapore University of Technology and Design
[2]Deakin University [3]University of Massachusetts Boston

Abstract

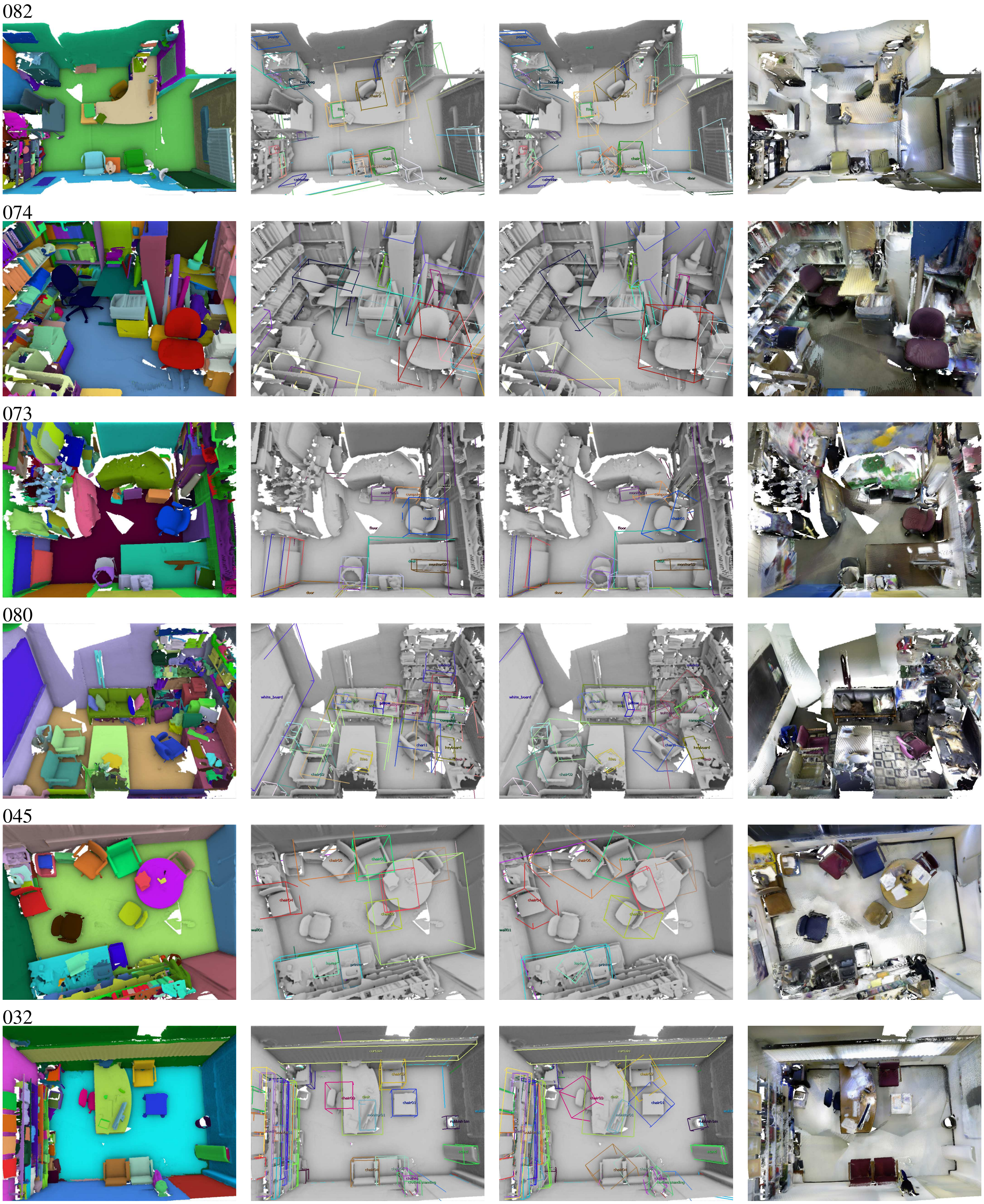
In this document, we provide screenshots of all scenes in our dataset. Our dataset has a total of 101 indoor scenes. 95 scenes are classified into seven main categories and a miscellaneous categories. 6 scenes are classified for sensor analysis and comparison. We also provided more details about the calibration experiments, and enlarged images for Figure 7, 8, and 12 in the paper. For details about the design and implementation of our annotation tool, please see our video and the technical report [3].

Contents

1. Dataset (101)

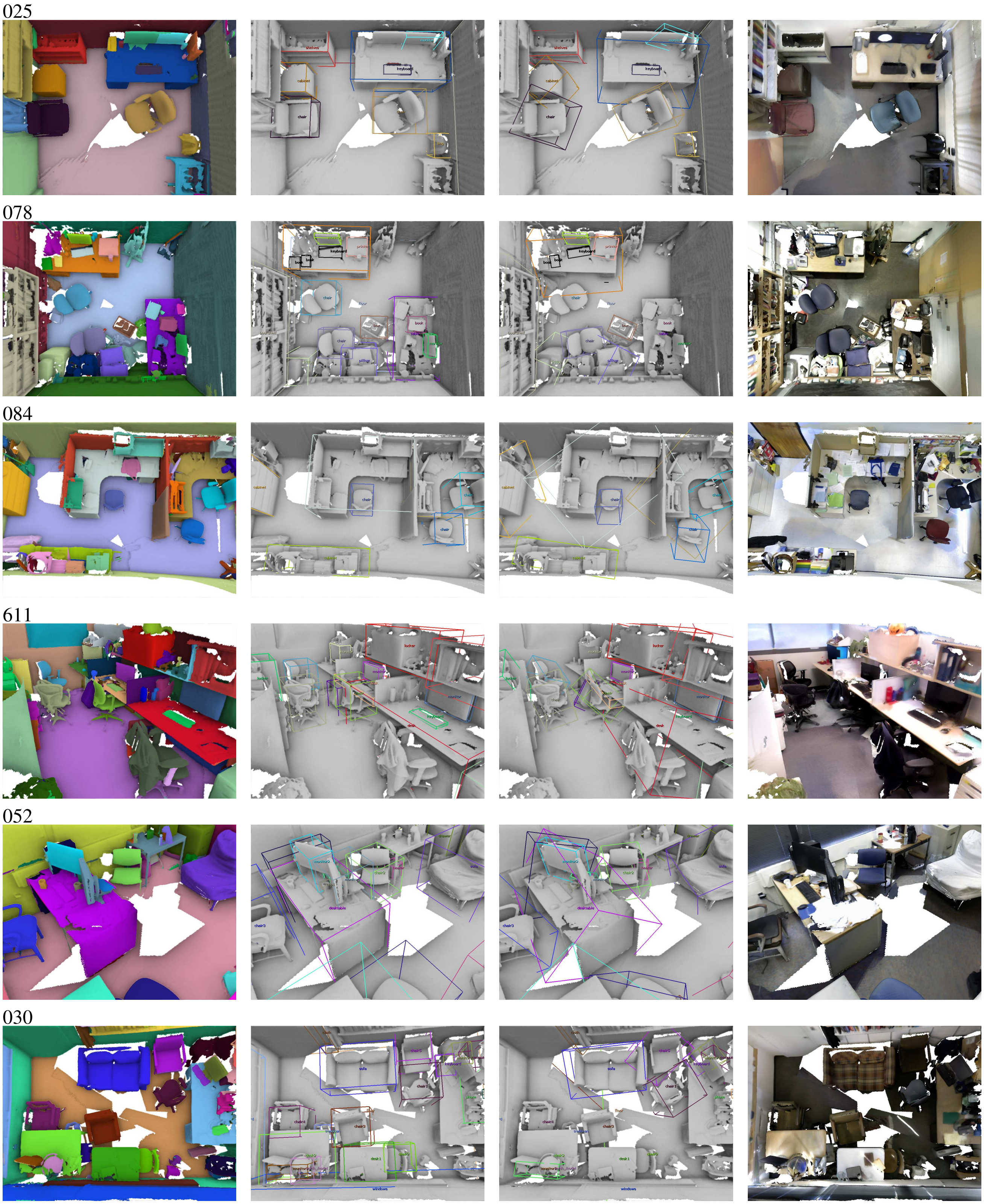
In this section, each row includes four rendered images of a scene. The labels, axis-aligned bounding boxes (AABB), oriented-bounding boxes (OBB), and color texture of each scene are displayed, respectively. To facilitate reference, the scene ID is displayed at the beginning of each row.

1.1. Workplace (27)



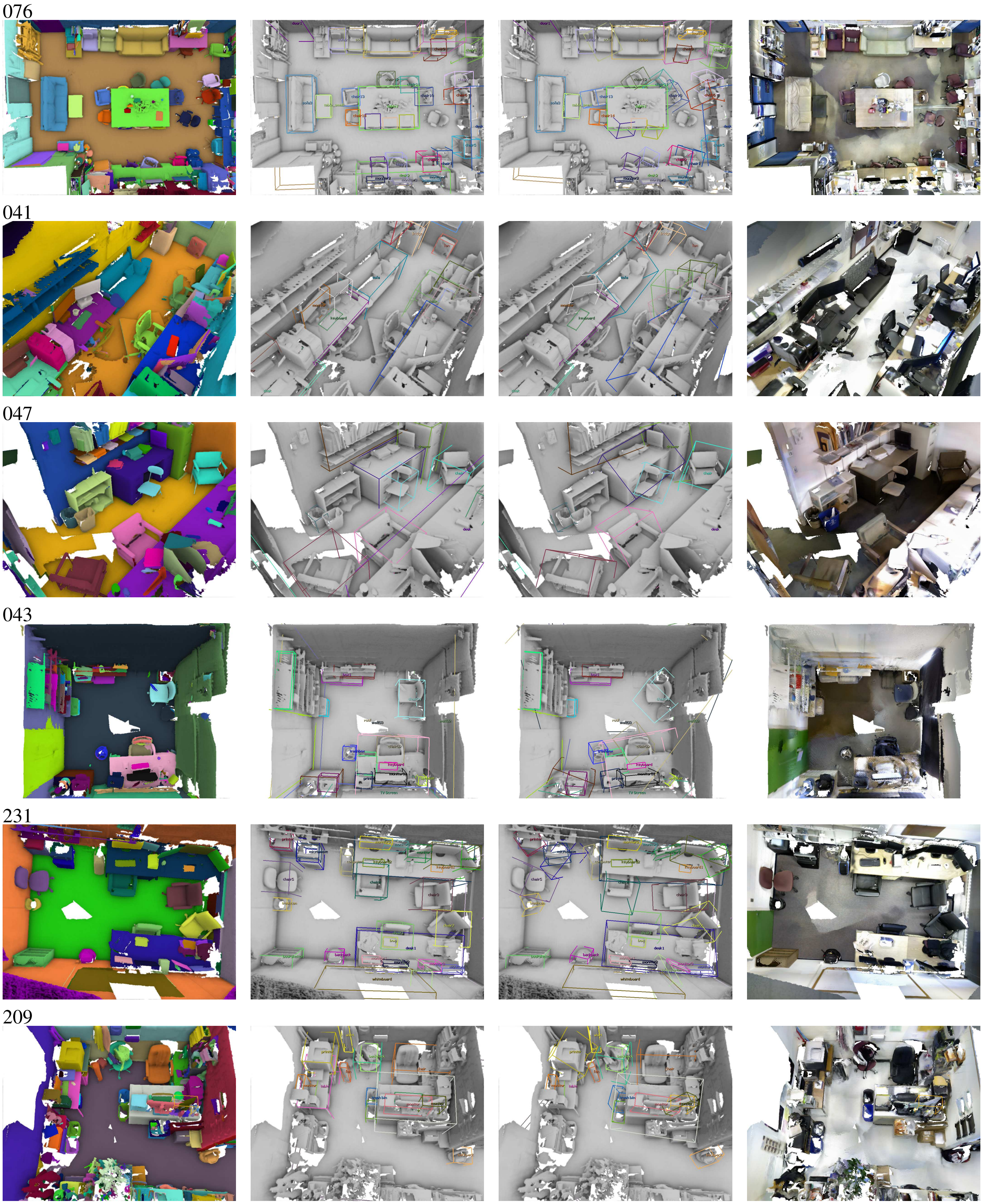
(a) Cluttered office

Figure 1: Workplace



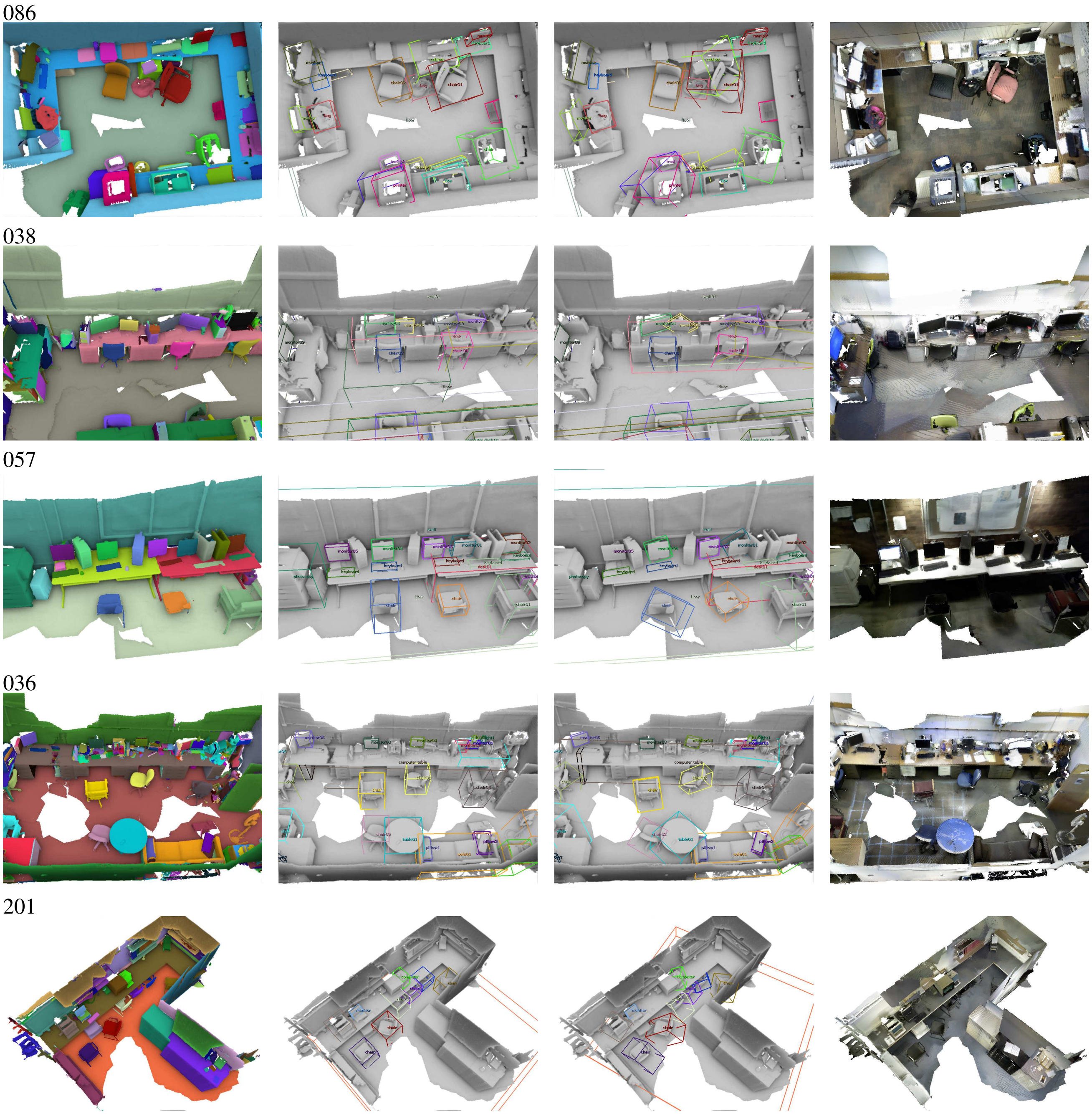
(b) Cluttered office

Figure 1: Workplace



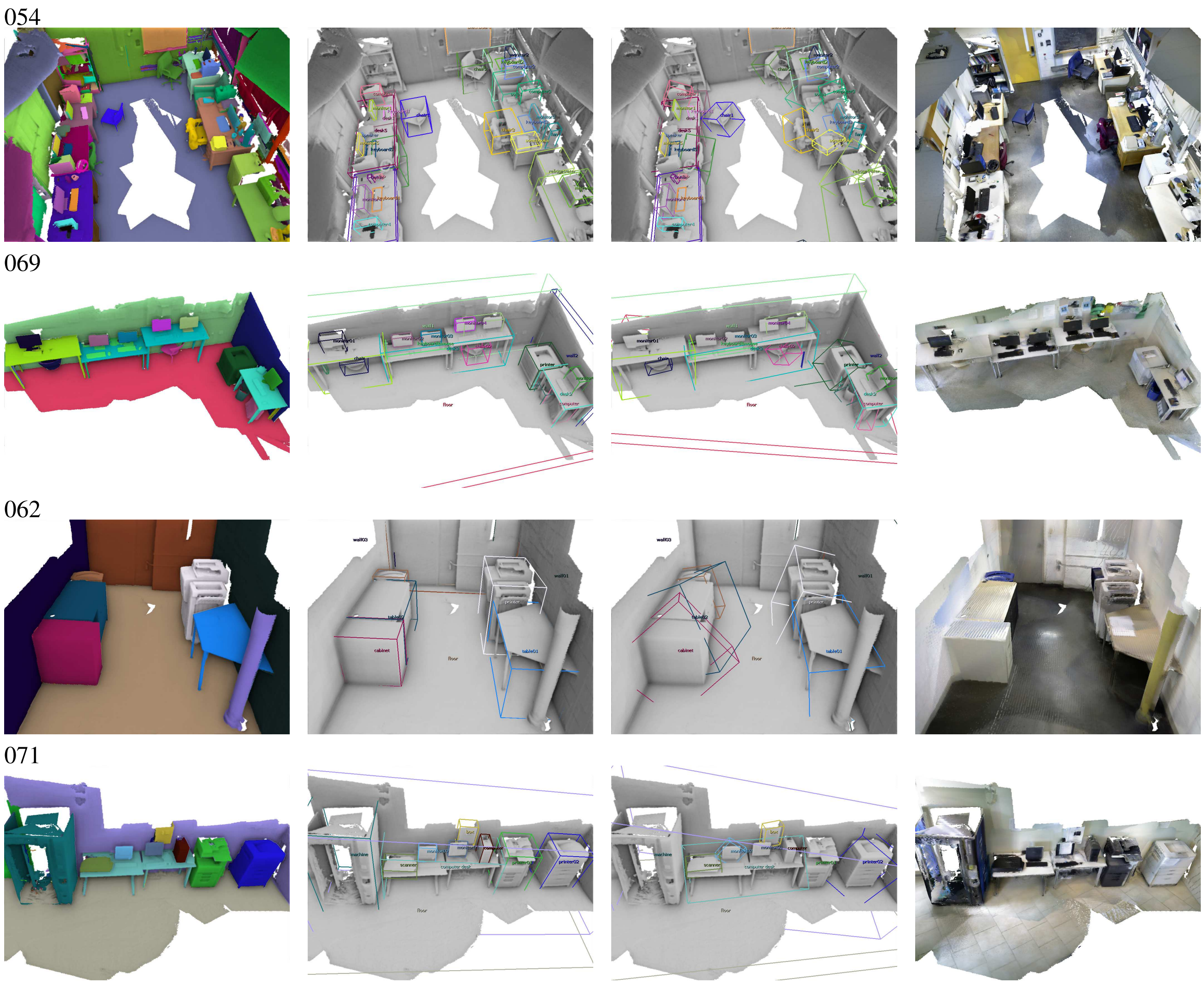
(c) Cluttered office

Figure 1: Workplace



(d) Computer lab

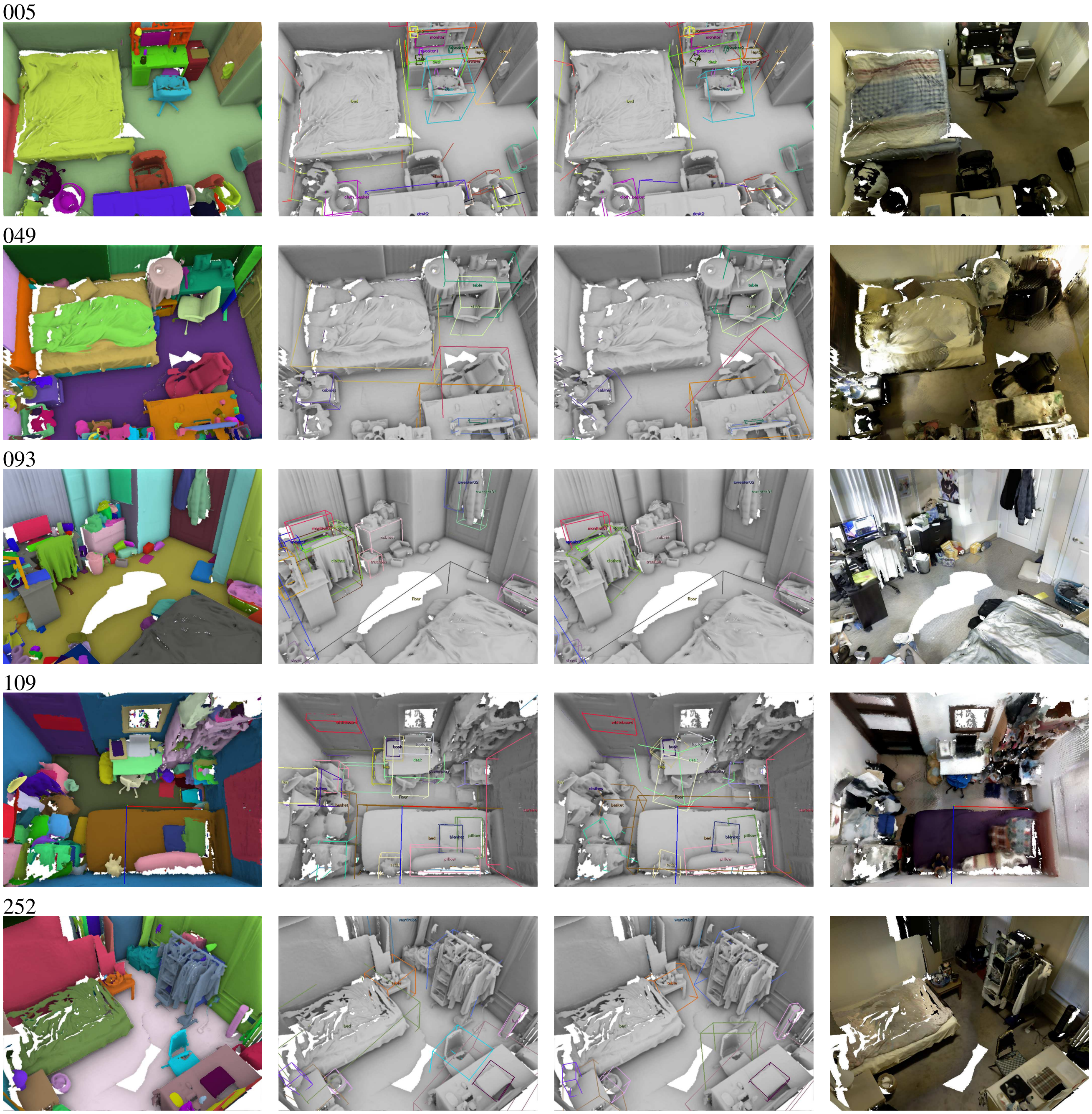
Figure 1: Workplace



(e) Computer lab and facility room

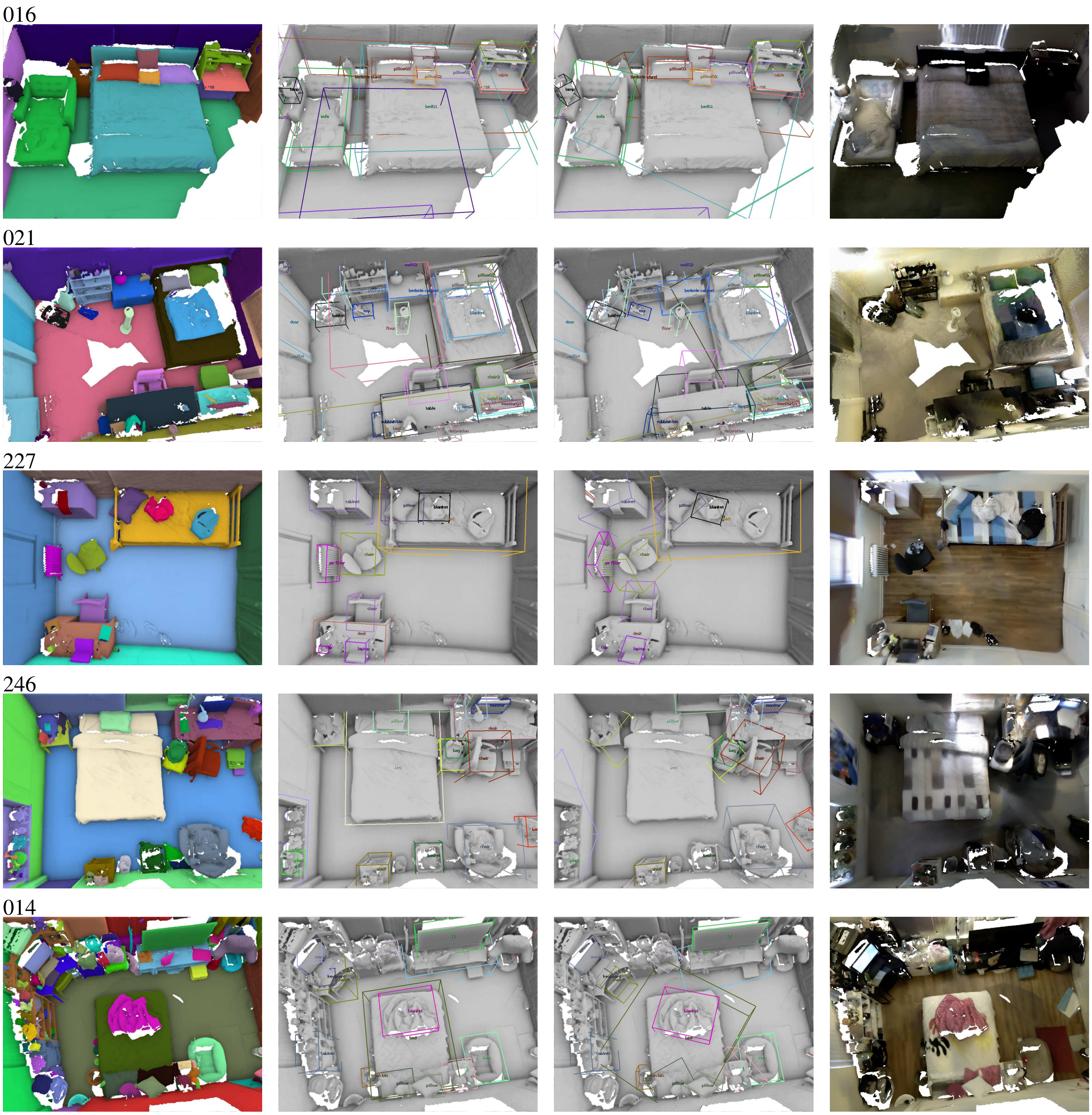
Figure 1: Workplace

1.2. Bedroom (19)



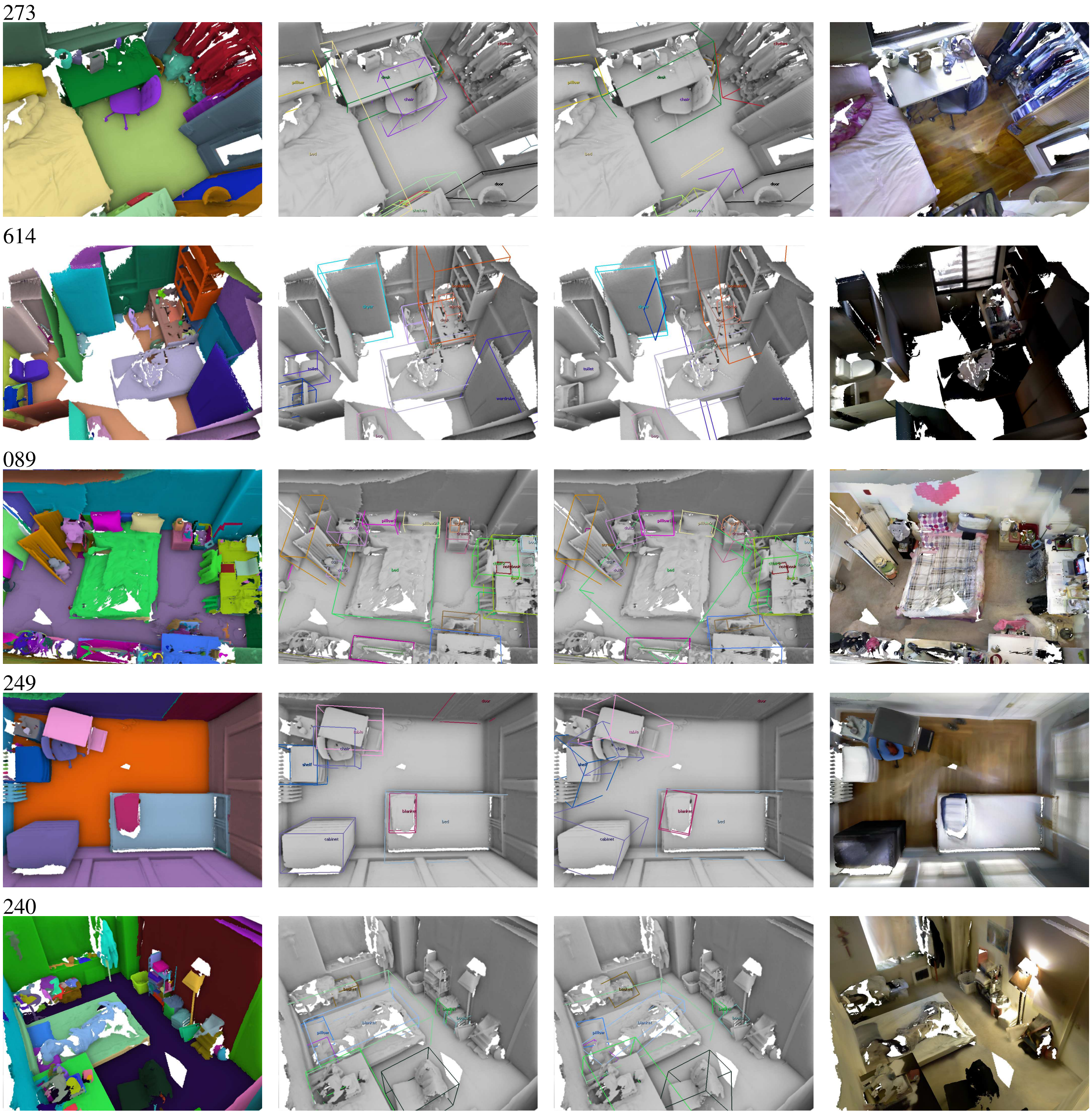
(a)

Figure 2: Bedroom



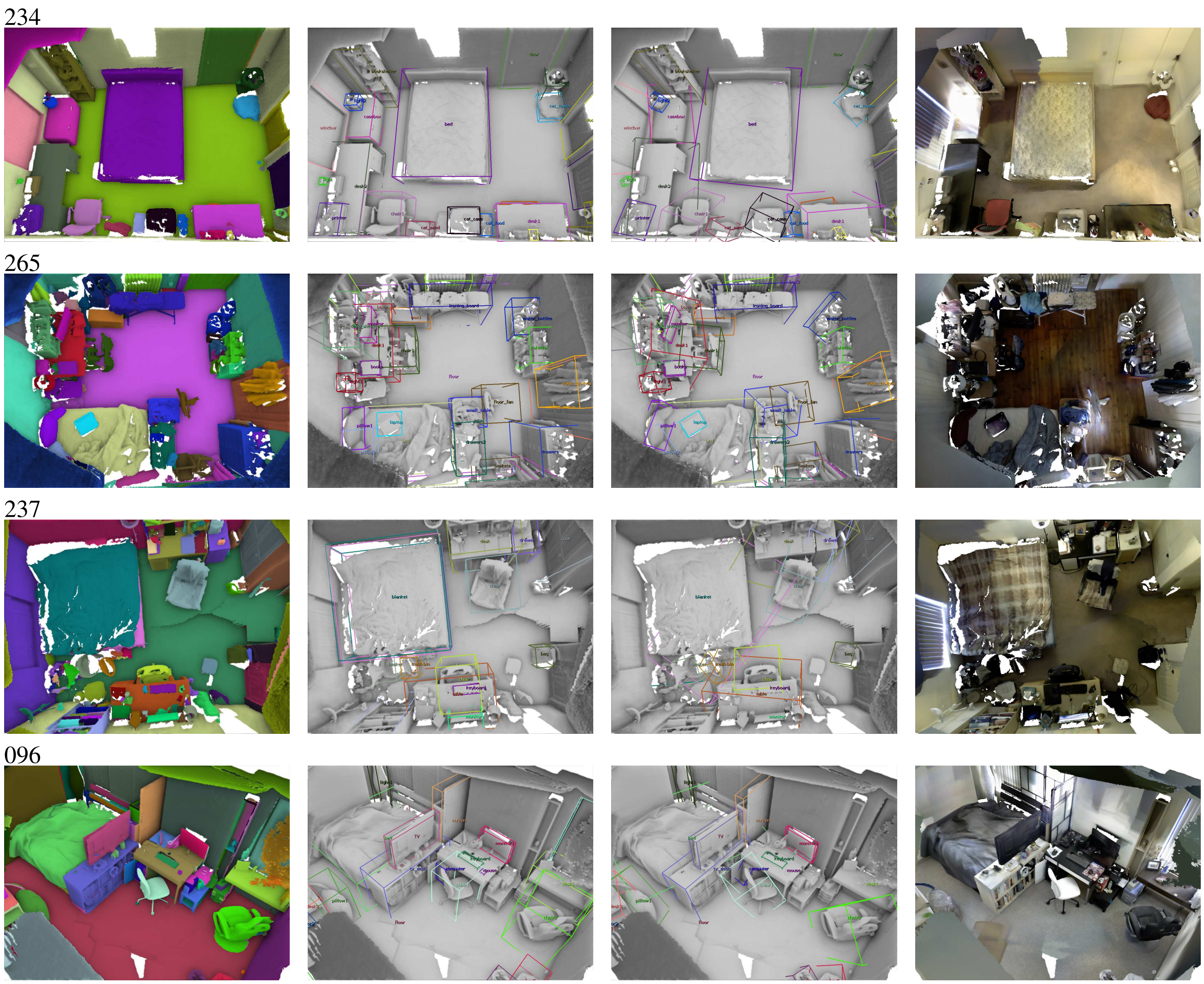
(b)

Figure 2: Bedroom



(c)

Figure 2: Bedroom



(d)

Figure 2: Bedroom

1.3. Living room (8)

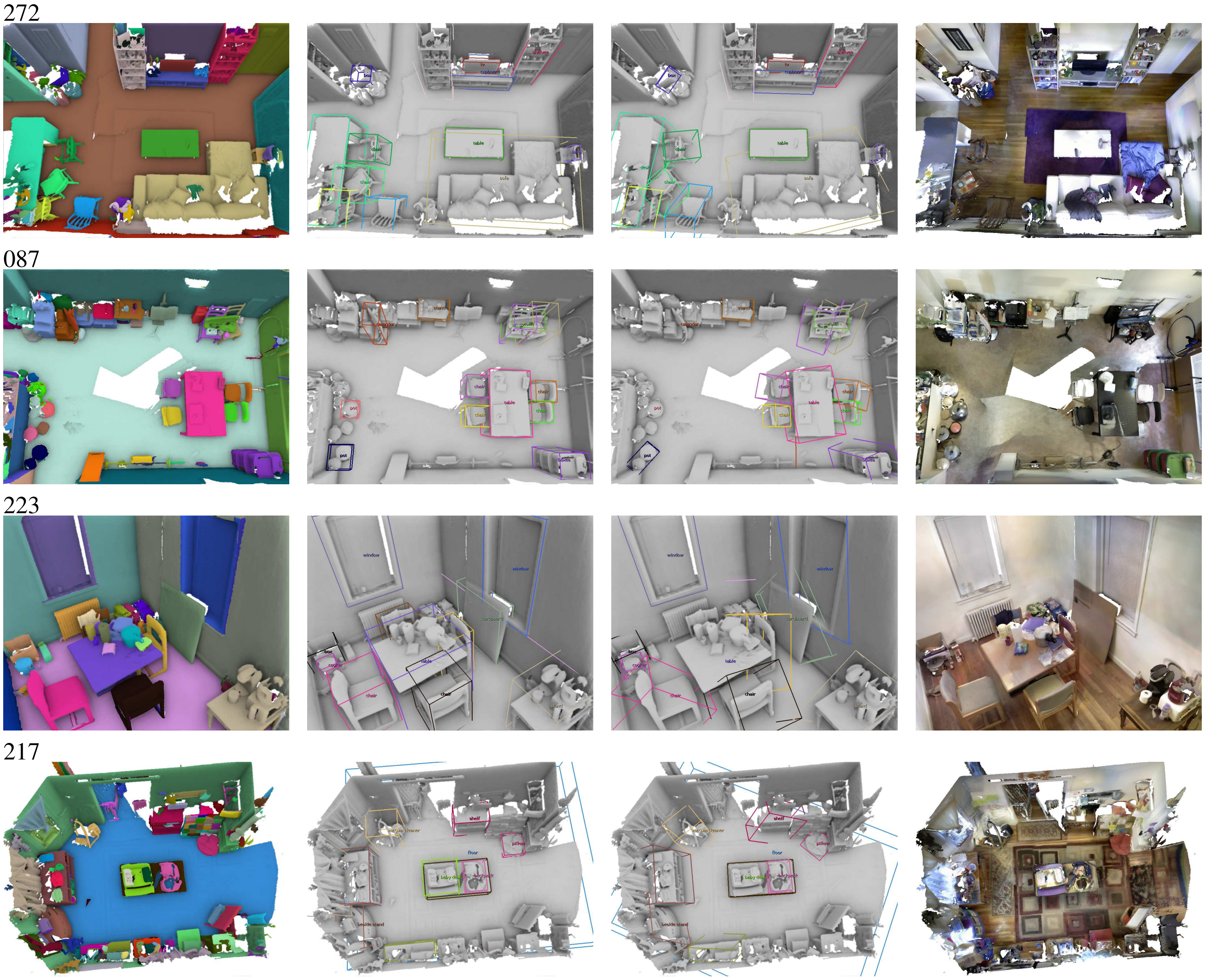
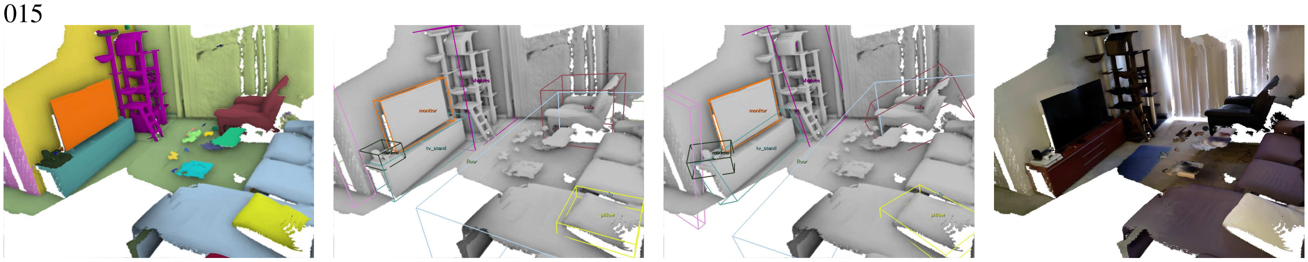
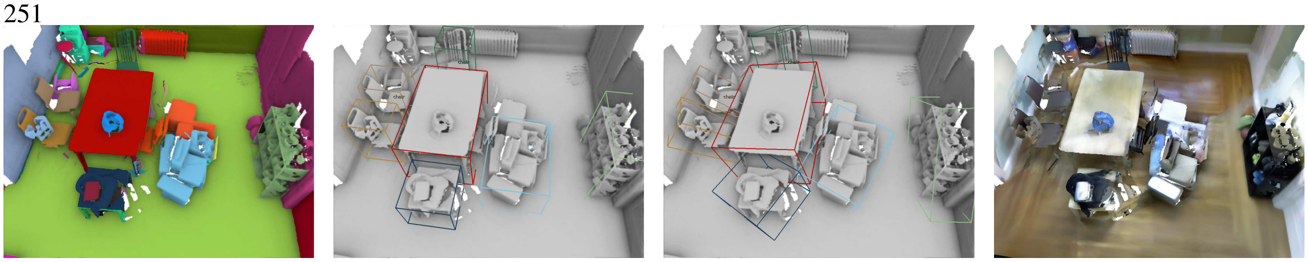


Figure 3: Living room

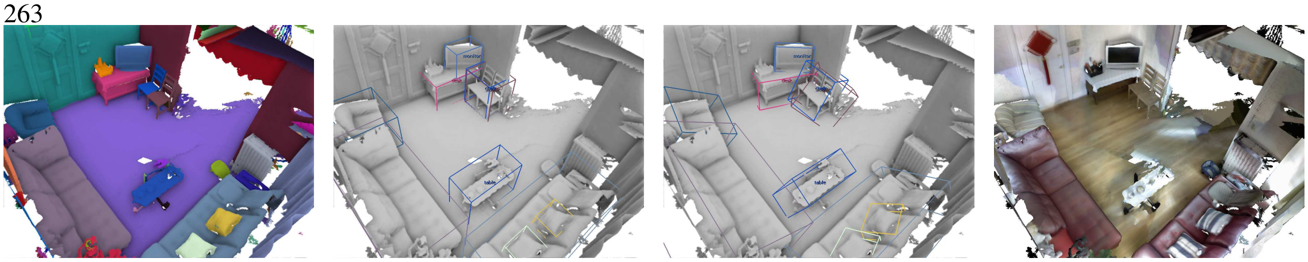
015



251



263



279

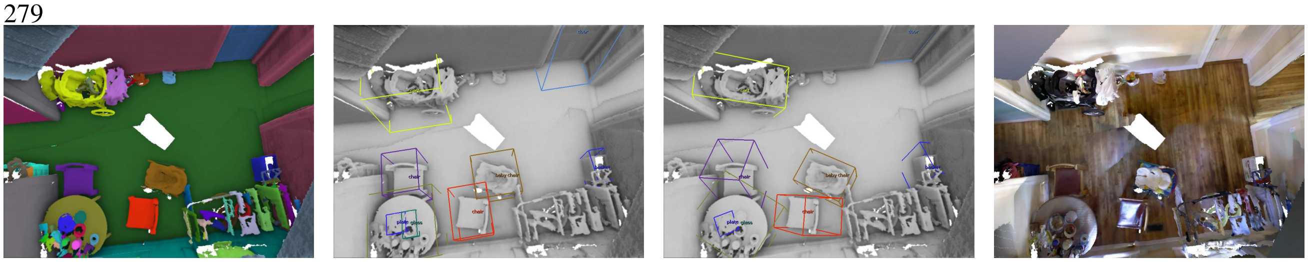
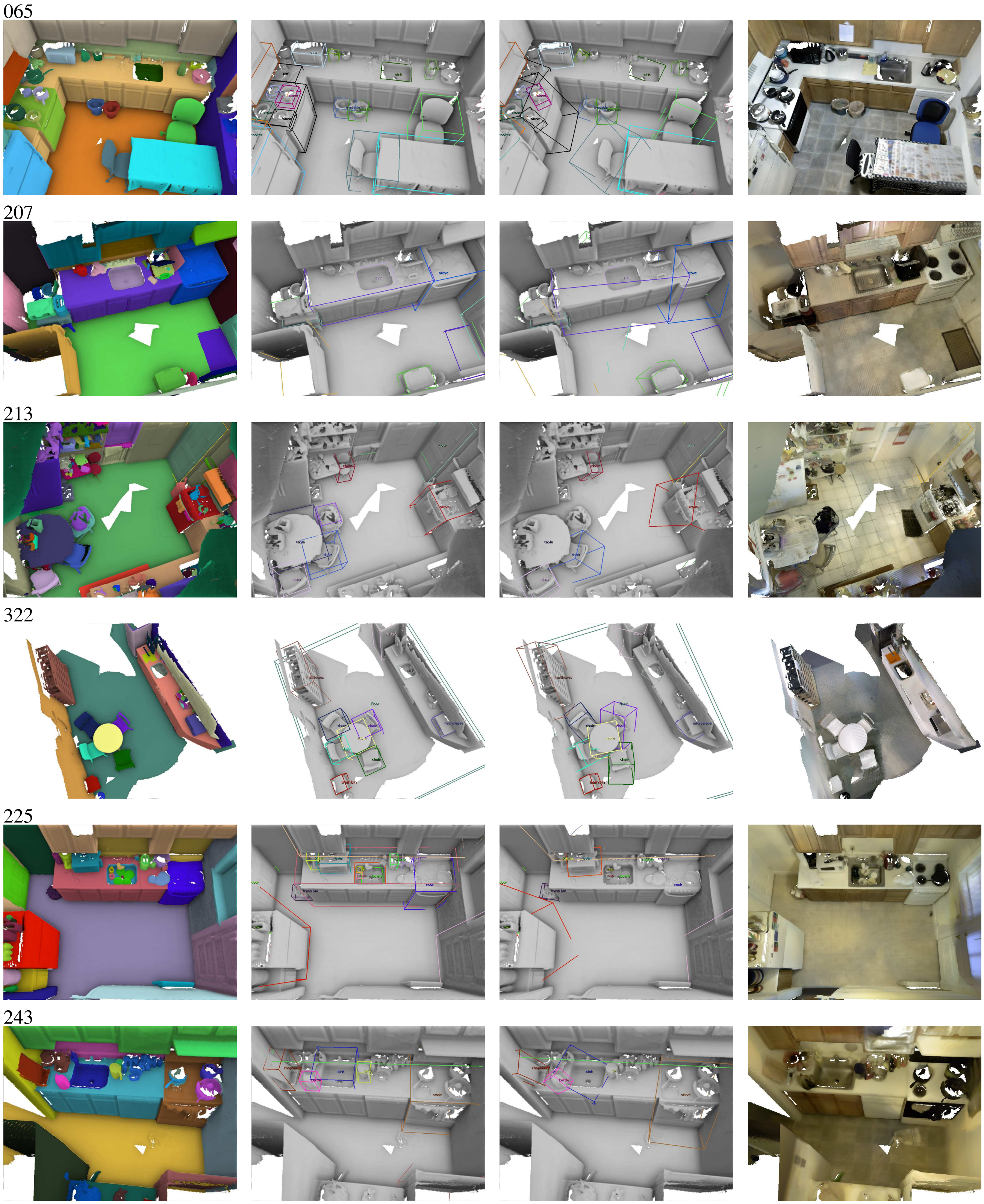


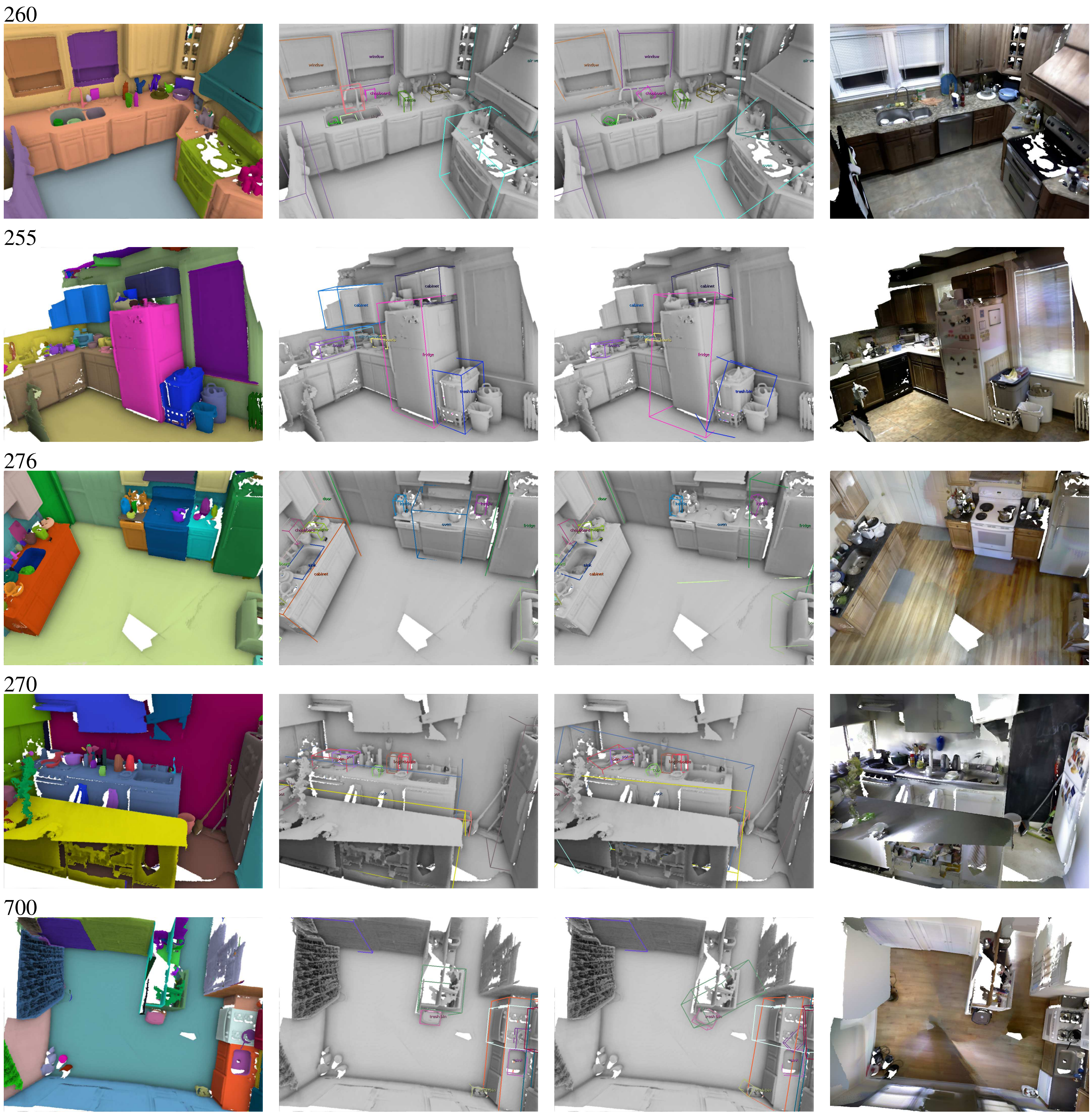
Figure 3: Living room

1.4. Kitchen (11)



(a)

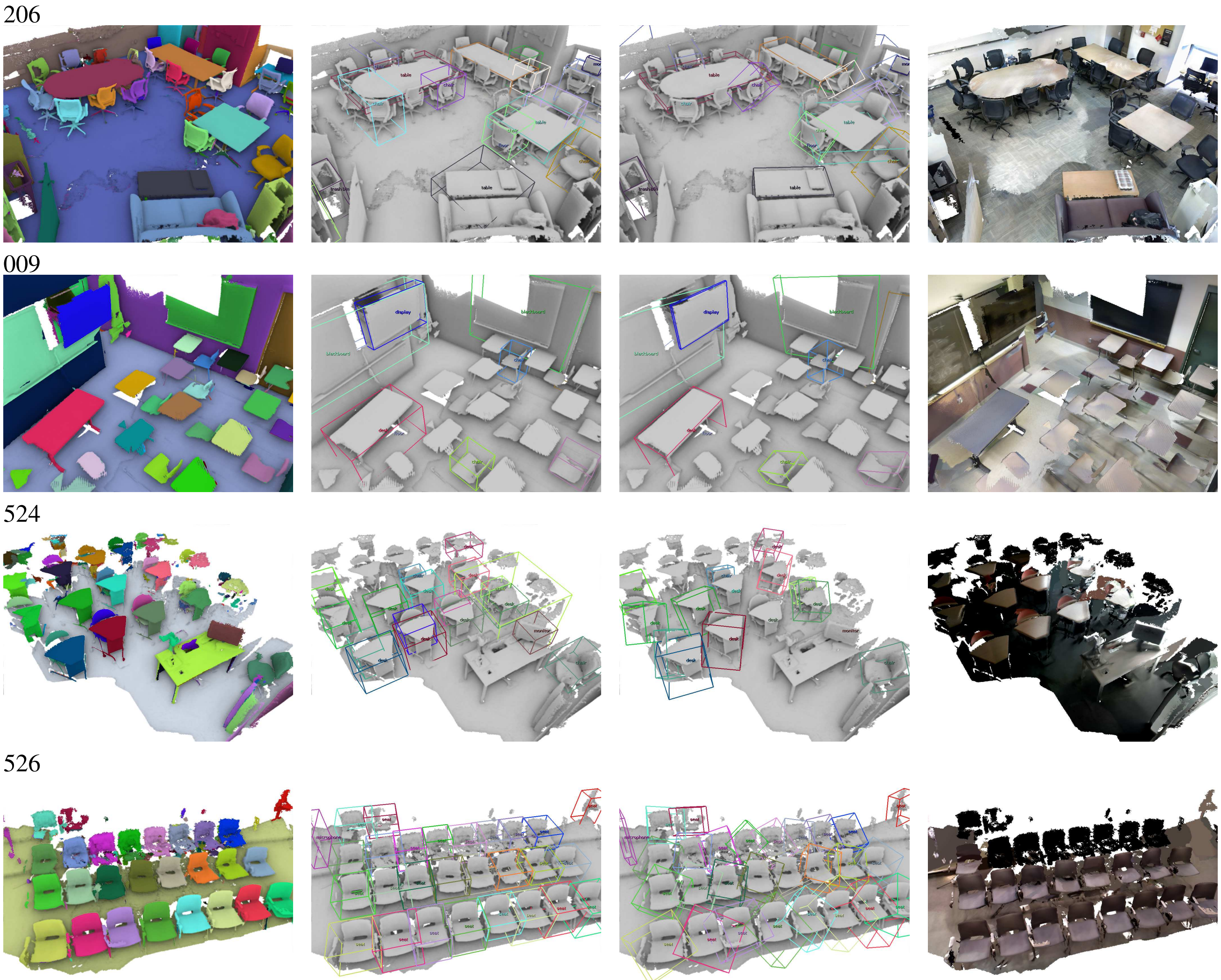
Figure 4: Kitchen



(b)

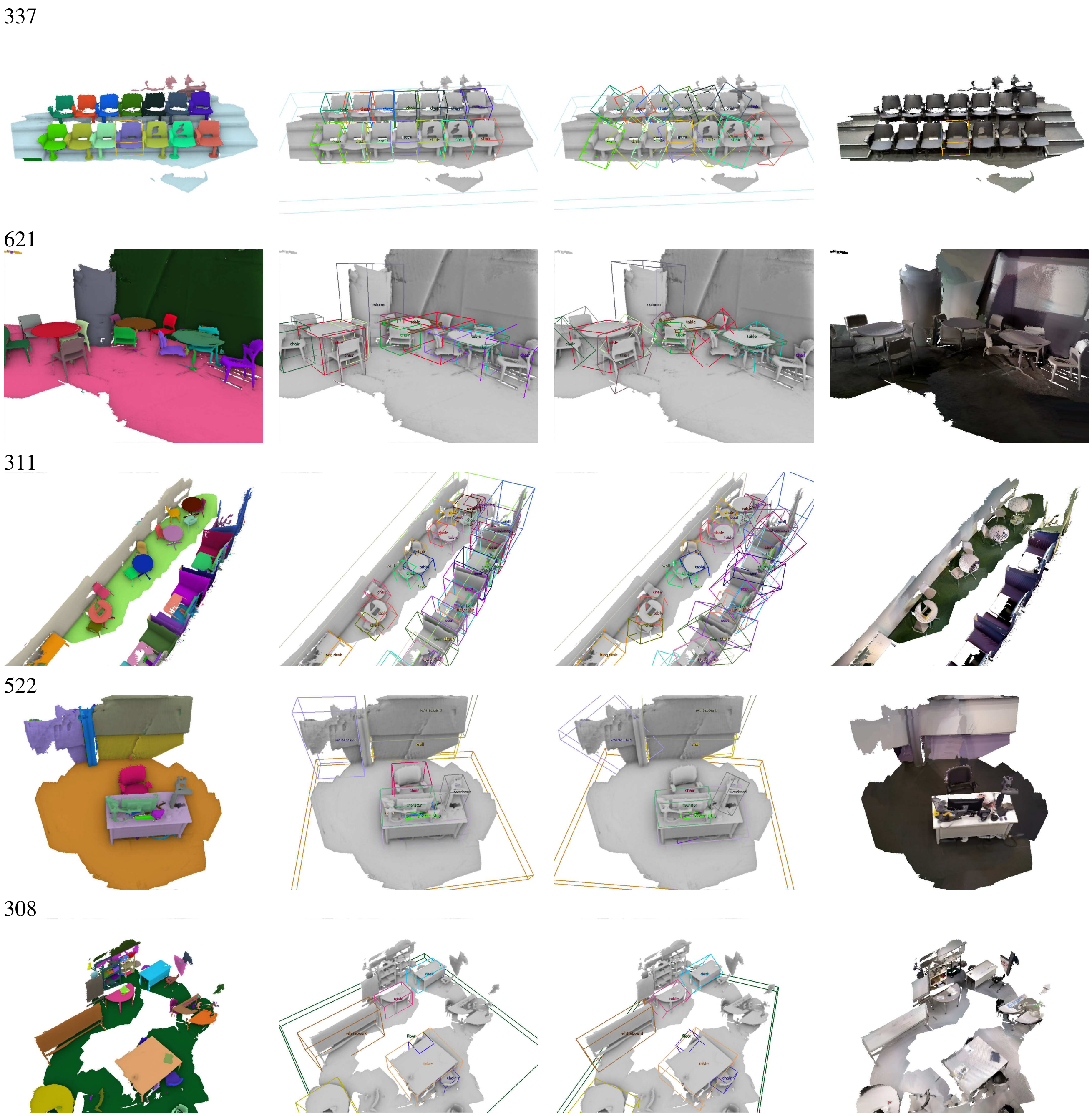
Figure 4: Kitchen

1.5. Study space (9)



(a)

Figure 5: Study space

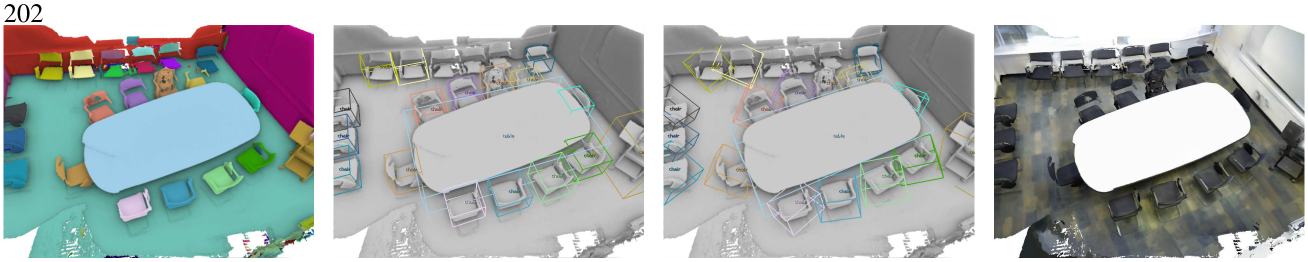


(b)

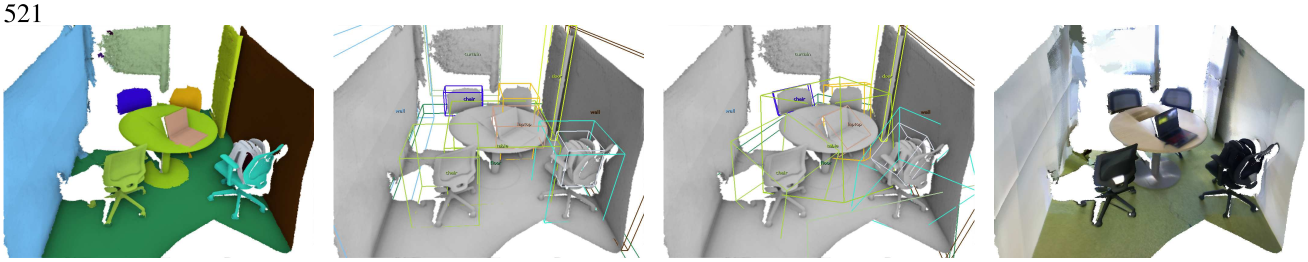
Figure 5: Study space

1.6. Meeting space (3)

202



521



613

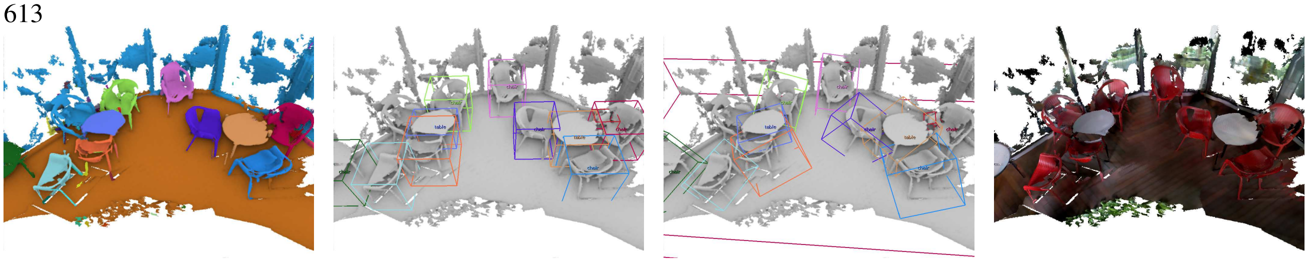
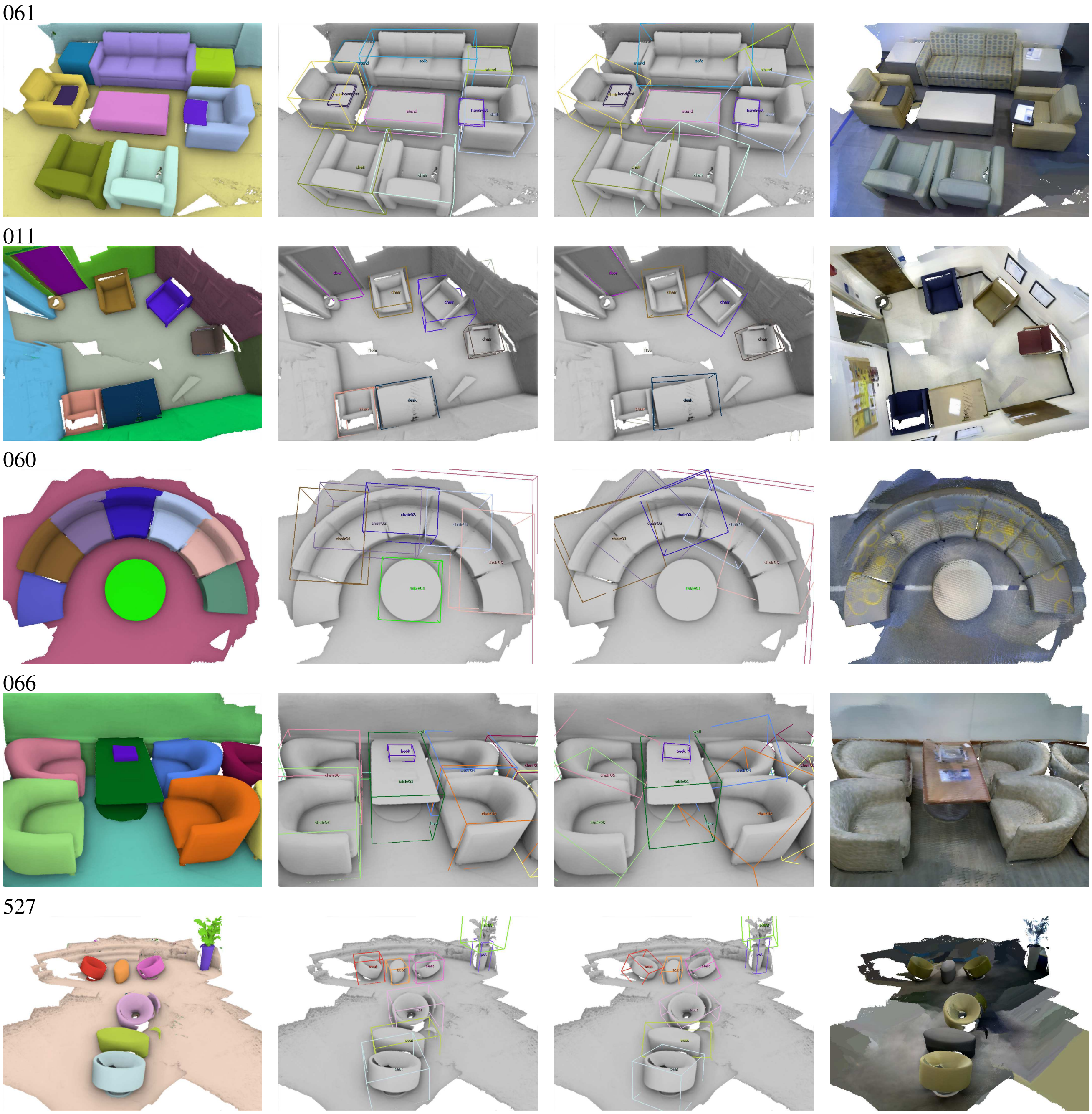


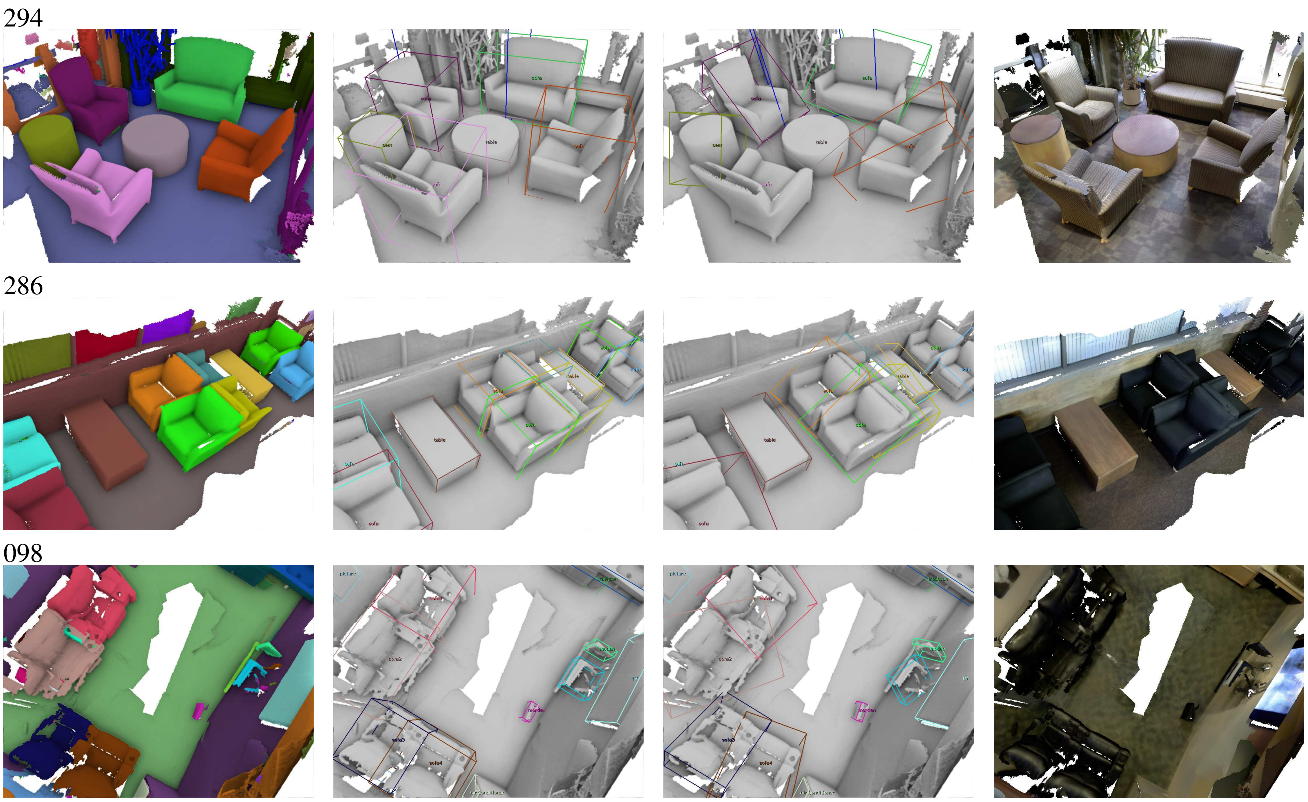
Figure 6: Meeting space

1.7. Lounge (8)



(a)

Figure 7: Lounge

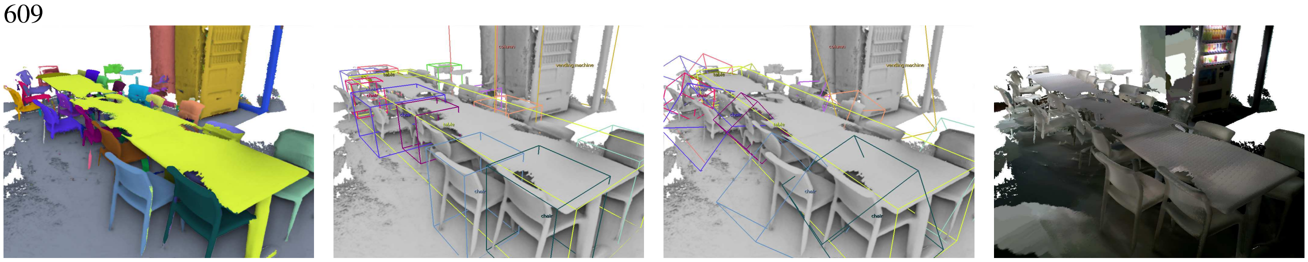


(b)

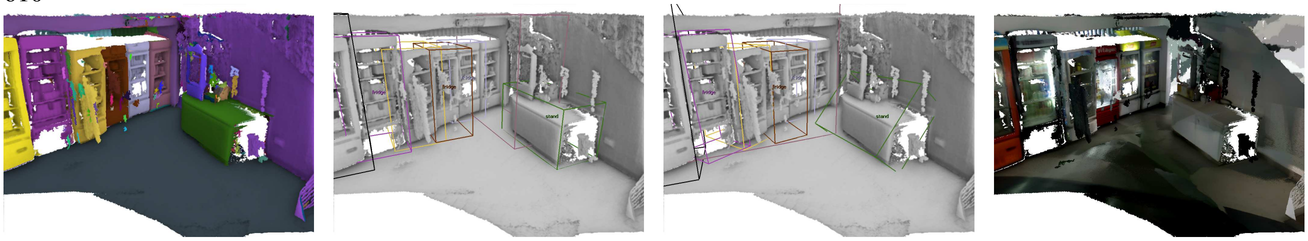
Figure 7: Lounge

1.8. Other scenes (10)

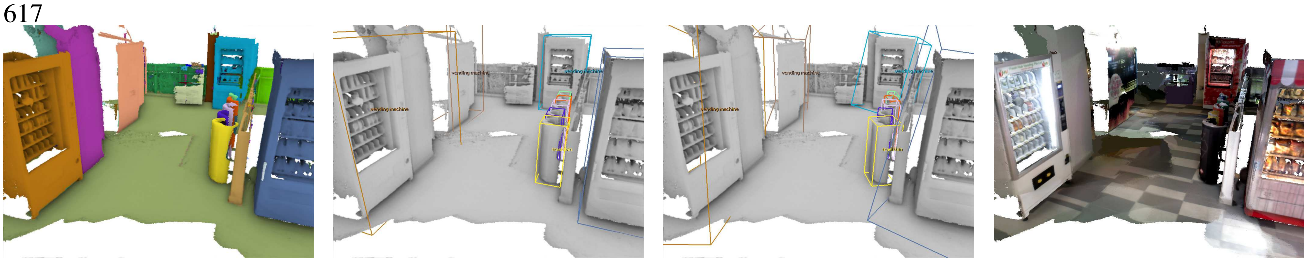
609



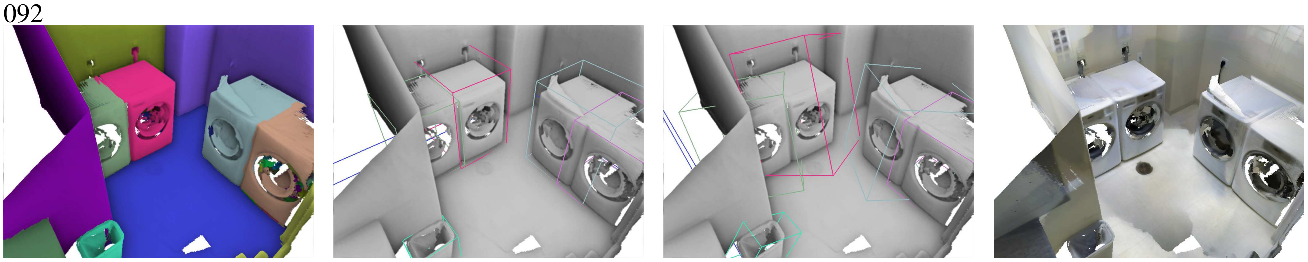
610



617

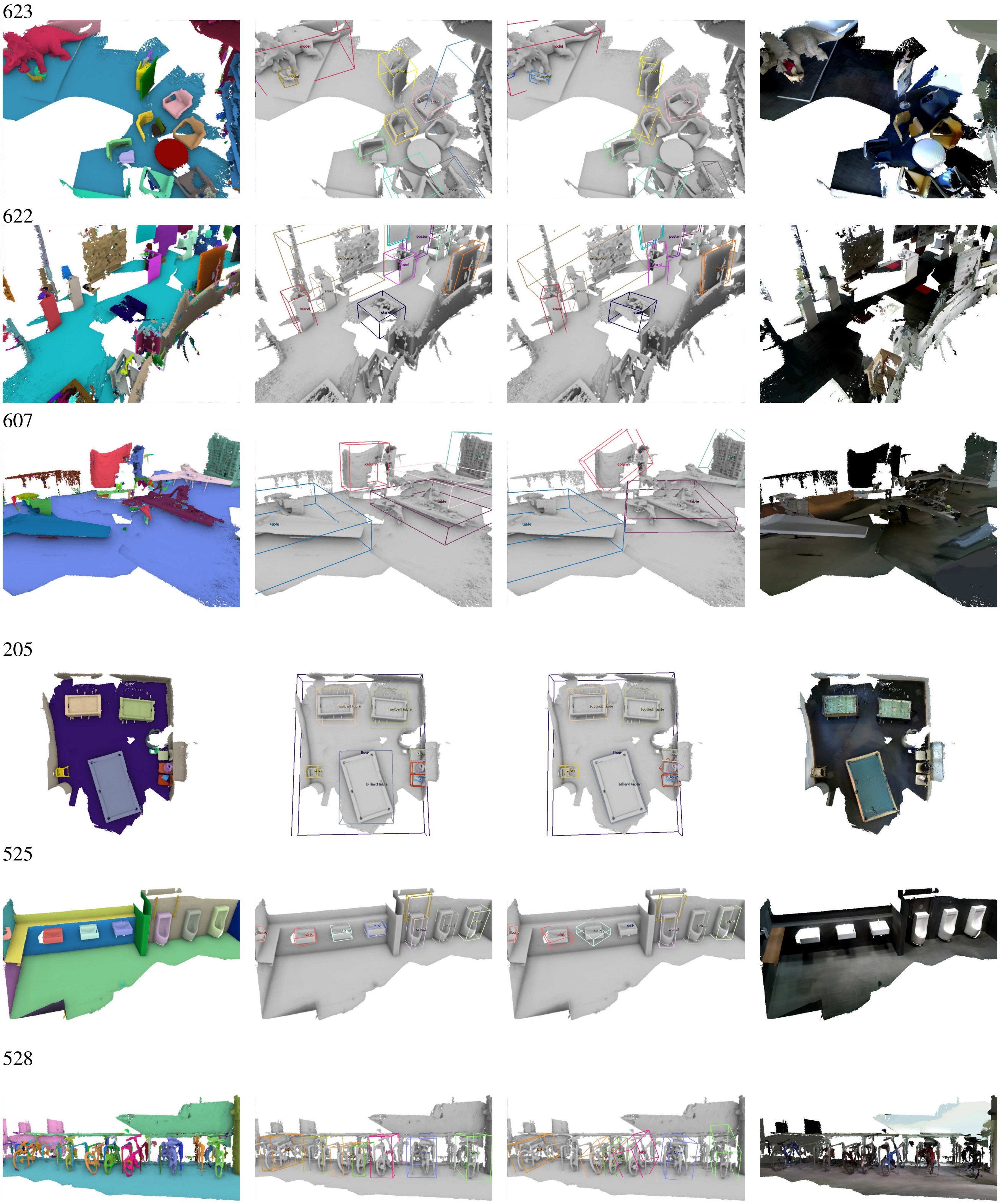


092



(a) Canteen, vending machines, laundry.

Figure 8: Other scenes



(b) Gallery, billiards, toilet, bike park.

Figure 8: Other scenes

1.9. Scenes for sensor quality comparison (6)

For studying sensor quality, our dataset has a few scenes captured by both Asus Xtion PRO (structured light sensor) and Microsoft Kinect 2 (time-of-flight sensor). In this section, scenes with ID 5xx are captured by Asus Xtion PRO. Those with ID 4xx are captured by Microsoft Kinect 2.

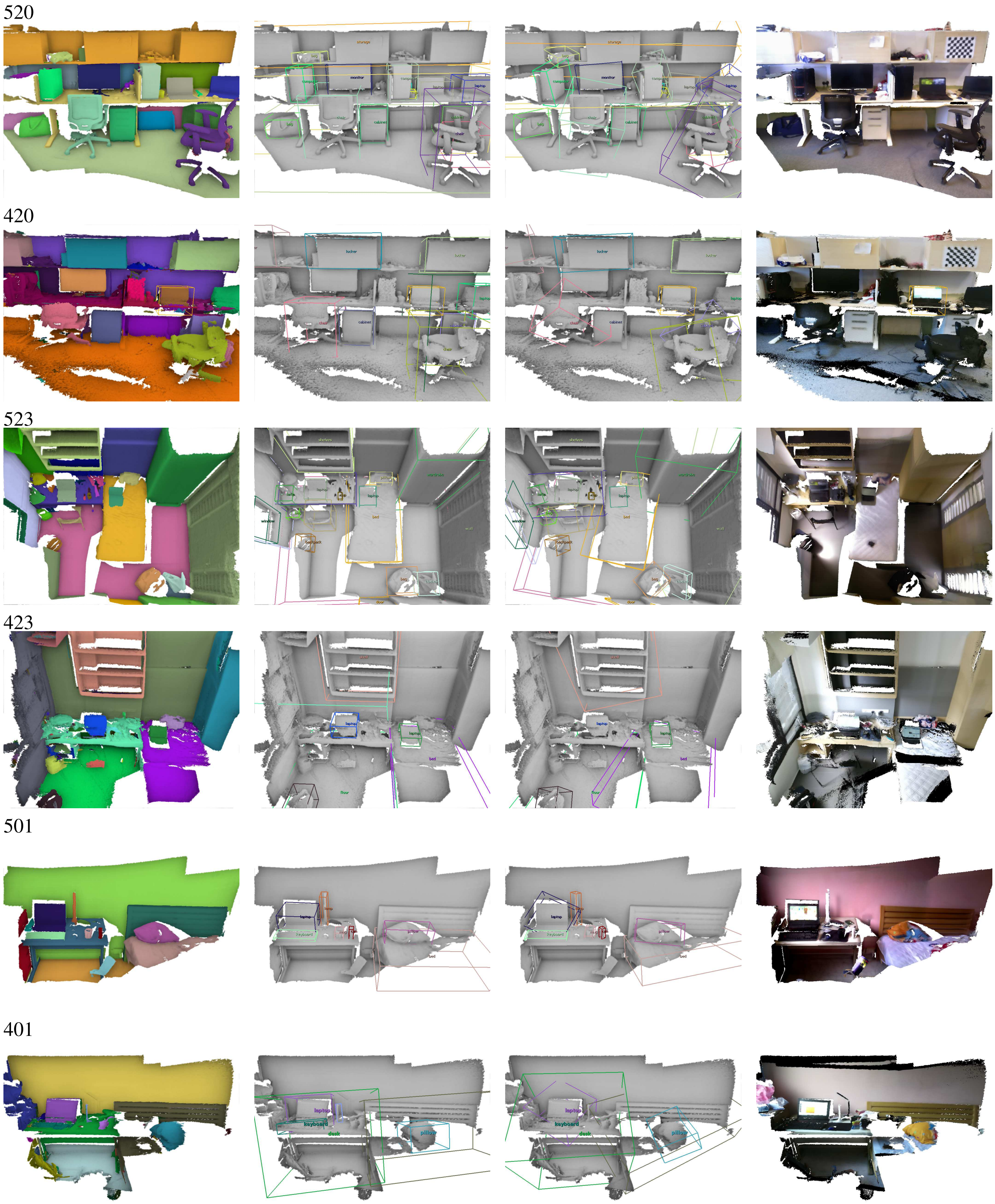


Figure 9: Scenes for sensor quality comparison.

2. Sensor Calibration

By default, the depth stream is aligned automatically to the color stream by internal processing in the RGBD camera. Therefore, we only need to calibrate the color sensor for our scene reconstruction.

We calibrate the intrinsic parameters of the color sensor using a checkerboard. Since calibration is a tedious process, we aim to seek a balance between accuracy and the number of parameters to calibrate. We measure the accuracy by depth reprojection error, which is the root mean square of the difference between reconstructed depth and captured depth for all points in all frames in a scene.

We compare the reprojection errors in four cases: (a) focal length by factory calibration, (b) focal length by checkerboard calibration, (c) focal length and principal point by checkerboard calibration, and (d) focal length, principal point, and radial distortion by checkerboard calibration.

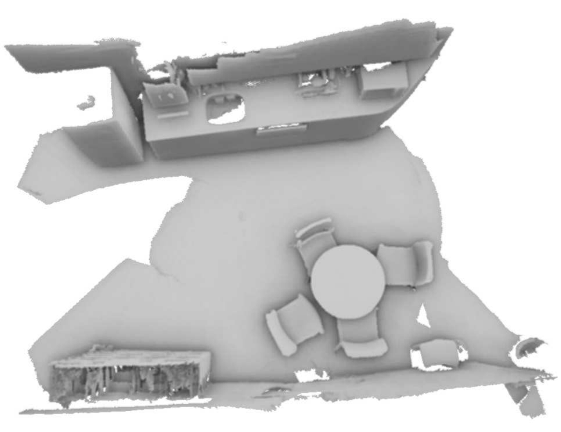
We select two scenes for this experiment. We choose a scene with low drift, and a scene with some ambiguous tracking to verify if the calibration affects the reconstruction. Table 1 lists the reprojection errors of the four cases for two example scenes. The unit is meter.

As can be seen, scene 322 has accurate surface alignment and low drift, and thus the magnitude of the reprojection error is low. Scene 311 has some drifts due to smooth walls at the right U-turn, and therefore has significantly larger reprojection errors. Despite that, it can be seen that case (c) has the lowest reprojection error; case (b) has the second lowest reprojection error for both scenes.

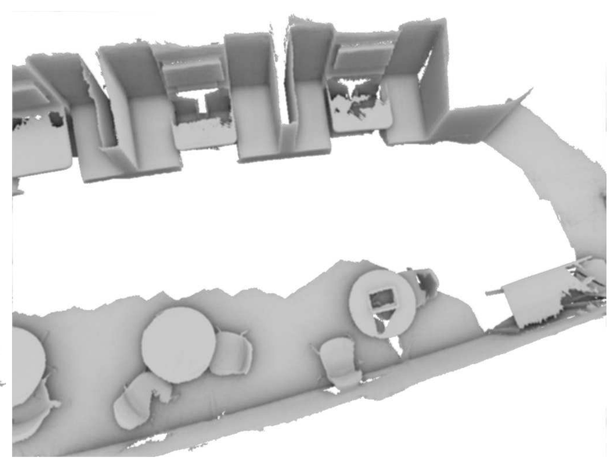
Finally, it is worth mentioning that in all cases, the reconstruction succeeds. This shows that while a good calibration can lead to a lower reprojection error, the reconstruction can tolerate a mild degree of inaccurate calibration parameters.

Scene ID	322	311
(a) Factory focal length	0.187129	0.403874
(b) Calibrated focal length	*0.159432*	*0.316245*
(c) With principal points	**0.155992**	**0.247517**
(d) With radial distortion	0.231416	0.329352

Table 1: Reconstruction based on calibrated intrinsic parameters with principal points results in the smallest error and hence a good quality mesh.



(a) Scene 322



(b) Scene 311

Figure 10: Example scenes for camera calibration.

From the experiments, we opt to reconstruct scenes with camera intrinsic parameters calibrated by focal length only (case (b)), as it yields relatively low reprojection error and is easy to use in practice. While calibrated principal points is slightly more accurate, off-center principal points are not convenient to use, especially for perspective projection using existing graphics API like OpenGL.

3. Data Capturing Statistics

Figure 11 provides more detailed statistics of the data capturing process discussed in the paper.

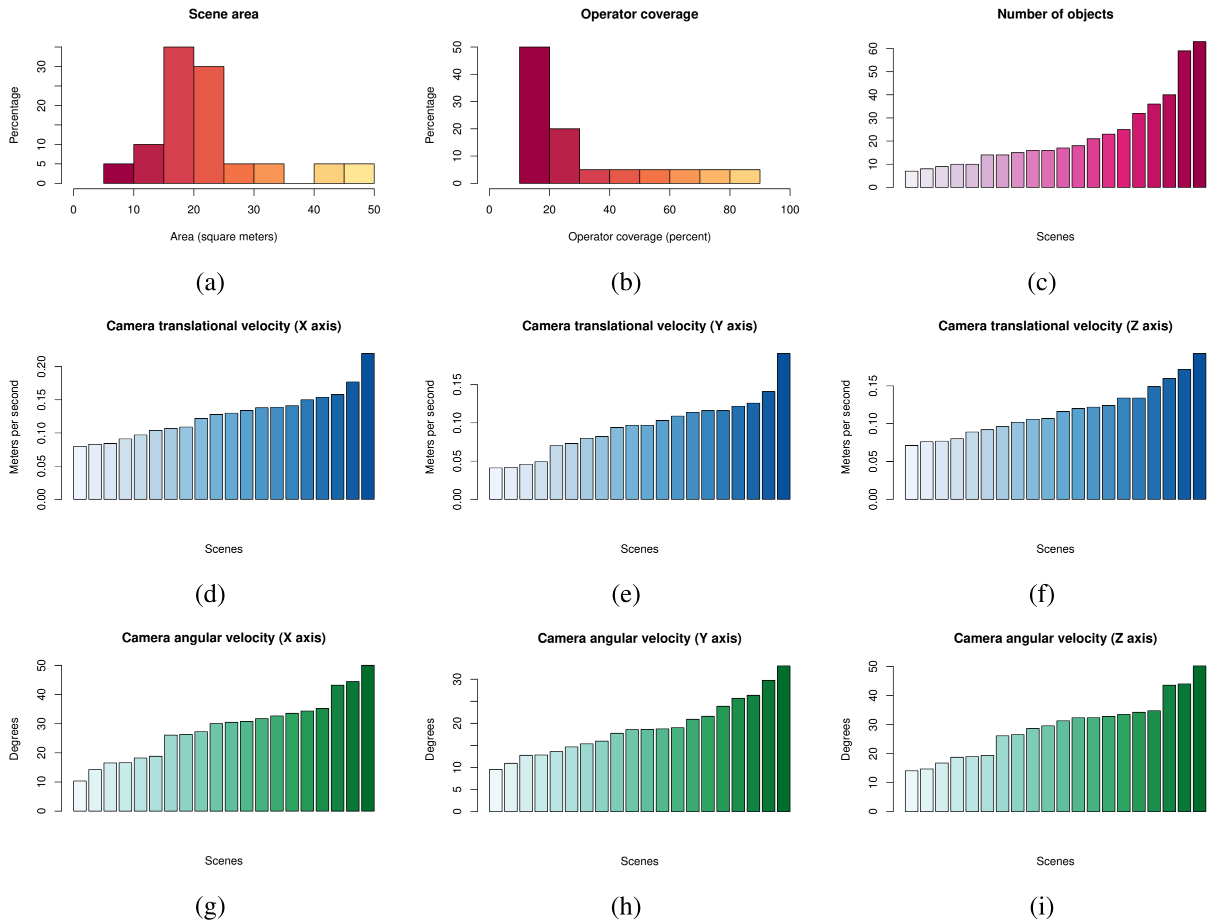


Figure 11: Statistics of data capturing. (a–b) Distribution of the scenes by area and operator coverage. (c) The number of objects per scene. (d–f) and (g-i): Camera translational and angular velocity during capturing.

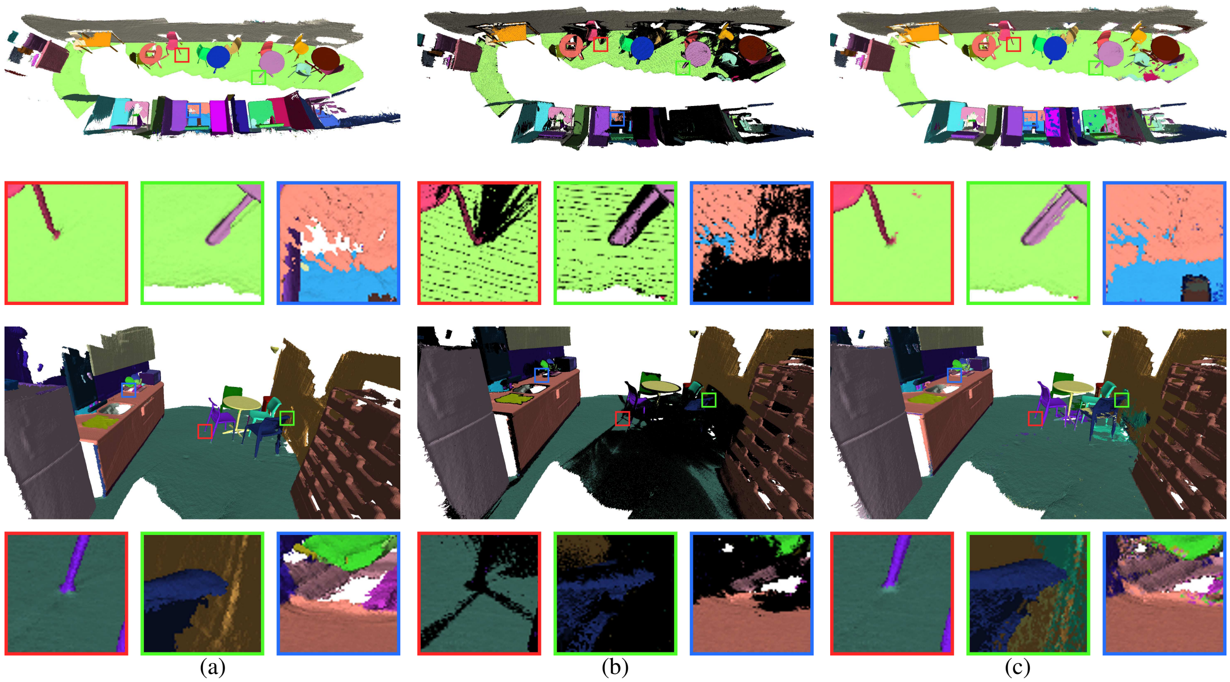
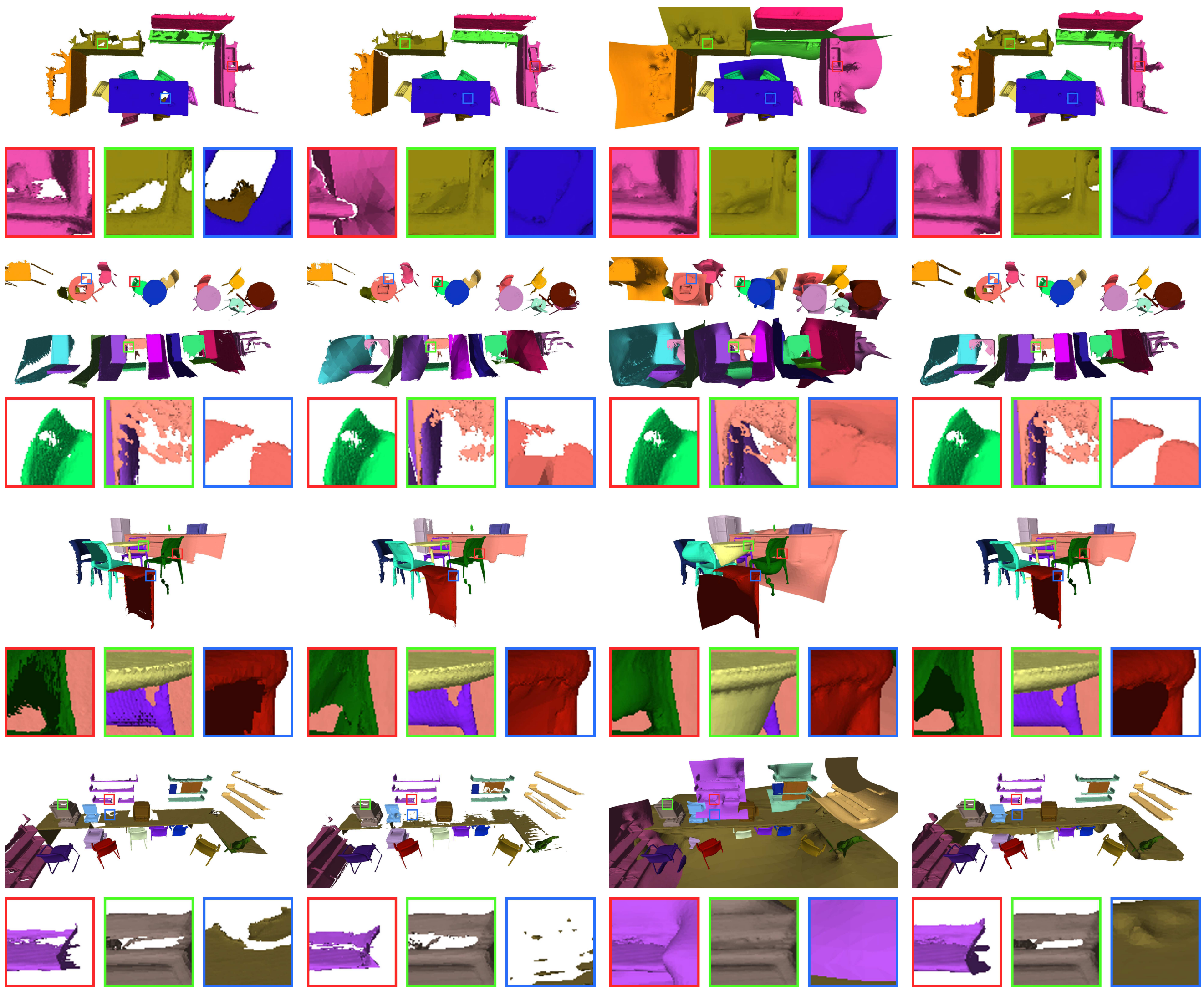


Figure 12: Annotation transfer example. (a) Source mesh with annotations. (b) Target mesh with annotations transferred from the source mesh. Note the subtle differences in geometry. Black regions correspond to unreliable annotation transfer. (c) Target mesh after propagating annotations using kd-tree search. Please refer to supplementary document for enlarged images.

4. Enlarged Figures

This section lists enlarge figures in the section of annotation transfer, shape completion, and scene synthesis.



Input PolyMender [1] Poisson [2] (before and after trimming)

Figure 13: Visual comparison of shape completion at scene level. Both methods can only complete small holes. Poisson reconstruction has less cracks and holes in general. Please refer to supplementary document for enlarged images.

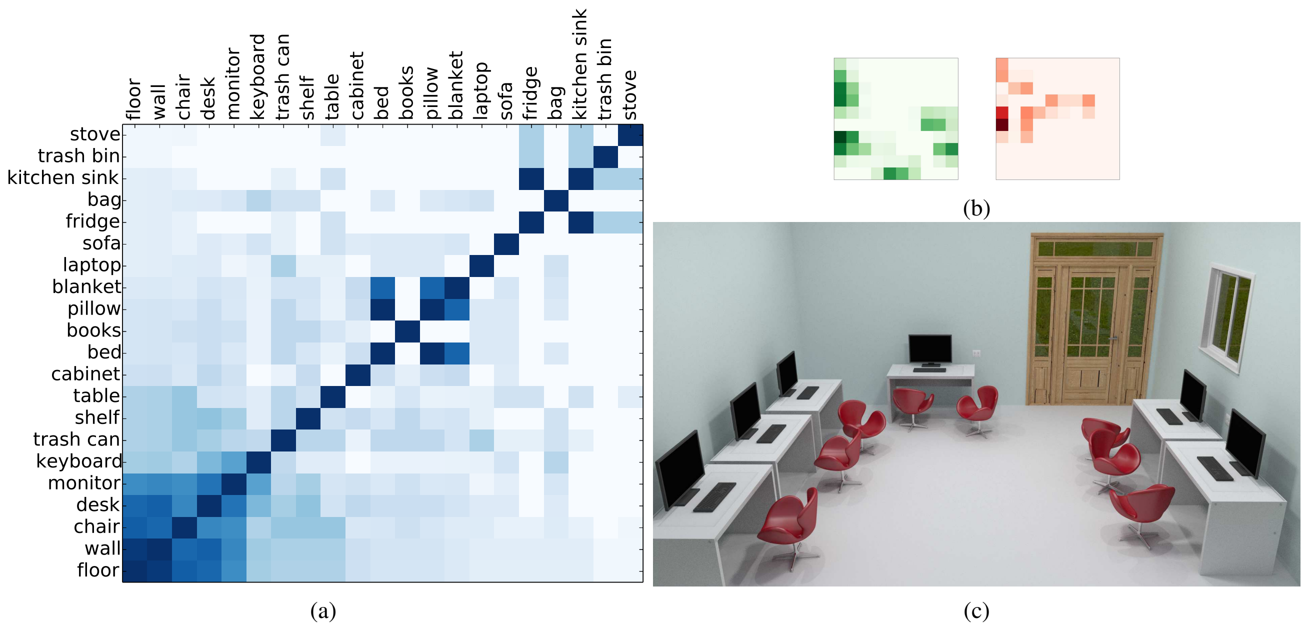


Figure 14: (a) Object co-occurrence of the 22 most common object classes in our dataset. (b) Object placement probabilities of desks (green) and chairs (red) computed from our scenes. Darker colors correspond to higher probability. The horizontal and vertical axis correspond to the X-axis and Z-axis in the world space. (c) A synthesized scene based on the statistics.

References

[1] T. Ju. Robust repair of polygonal models. In *ACM Transactions on Graphics (TOG)*, pages 888–895, New York, NY, USA, 2004. ACM. 29

[2] M. Kazhdan and H. Hoppe. Screened poisson surface reconstruction. *ACM Transactions on Graphics (TOG)*, 32(3):29:1–29:13, July 2013. 29

[3] D. Thanh Nguyen, B.-S. Hua, L.-F. Yu, and S.-K. Yeung. A robust 3d-2d interactive tool for scene segmentation and annotation. *Computing Research Repository (CoRR)*, 2016. 1