Privacy-Preserving Representations are not Enough: Recovering Scene Content from Camera Poses

-

Supplementary Material

Kunal Chelani[1] Torsten Sattler[2] Fredrik Kahl[1] Zuzana Kukelova[3]
[1] Chalmers University of Technology
[2] Czech Institute of Informatics, Robotics and Cybernetics, Czech Technical University in Prague
[3] VRG, Faculty of Electrical Engineering, Czech Technical University in Prague
{chelani,fredrik.kahl}@chalmers.se torsten.sattler@cvut.cz kukelzuz@fel.cvut.cz

In Section 1, we present additional qualitative results for alignments of objects from the *IKEA-Objects* set in all scenes from the *IKEA-Scenes* dataset. Section 2 presents qualitative results for object alignments in some of the RIO10 scenes, corresponding to the quantitative results shown in Figure 6 of the main paper.

1. Qualitative results - *Ikea-Scenes* and *Ikea-Objects*

Qualitative alignment results. In Figures 1-21, we present alignment results for four selected objects present in each scene from the the *IKEA-Scenes* dataset. We include results for poses obtained by using 1) Superpoint [1] features with Superglue [4]-based matching and 2) R2D2 [2] features with Nearest Neighbor matching within the Hloc [3,4] pipeline. We selected objects of varying sizes, shapes, categories, textures *etc*. as to show the feasibility of the attack. As can be seen, it is often possible to quite accurately place objects in a scene based on camera poses estimated by a visual localization system.

Camera poses. The same figures also show the set of camera poses returned by the server (blue) and the subset of poses (green) selected as inliers by our alignment method (see Algorithm 1 and Section 3.1 in the main paper). Note that these poses are obtained by localizing a sequence of query images sampled from a video. Hence, temporally close frames can be expected to show spatial coherence in their pose estimates. However, due to the difference in appearance between objects that are present in the scenes and the objects that are used for the attack, as well as due to viewpoint changes, there can be many outlier poses. Still, Algorithm 1 (of the main paper) is able to identify subsets of camera poses that allow to appropriately position the 3D models of the objects, demonstrating the robustness of the approach.

Failure cases. For each scene, failure cases of the alignment step are highlighted using red colored boxes. Alignments that result in positioning the attacking object such that it is either too far from the corresponding object in the scene or its "up-direction" is very different from that of the corresponding object are considered as failure cases. Only visual inspection has been used to decide whether a case is considered a success or failure. As can be seen from visualizations, the 3D models of the attacking objects that result in failures are often quite different (in terms of appearance) from the corresponding objects in the scene. Some failure cases such as the *Sofa Linanas* in Figure 15 can be attributed to this difference while many other cases such as *Sofa Soderhamn* in Figure 12 and *Stool Kyrre* in Figure 15 show the success in such difficult cases. Other reasons for failure are a low number of matches from texture-less or weakly textured objects. *Chair Odger* in Figure 12 and *Chair Froset* in Figure 3 (using Superpoint features [1] with the Superglue matcher [4]), are examples of such cases. Scene02, Scene04, and Scene06, shown in Figures 4-6, 10-12, 16-18 respectively, are complex rooms (*e.g.*, an open concept kitchen in *Scene04*) composed of very similar looking objects, and hence challenging cases for such an attack. This results in more failure cases. Still, as shown in the figures, the attack succeeds in many cases, which shows its feasibility.

2. RIO10 example alignment results

In Section 6 of the main paper, we discuss a potential strategy to defend against the attack introduced in our work. Yet, Figure 6 of the main paper shows that this strategy causes the localization process to not only reject malicious

queries, but also to reject genuine query images. In Figure 6 of the main paper, we consider 3 different objects attacking 3 different scenes of the RIO10 dataset. To visualize these scenarios, Figure 22 shows qualitative results for object alignments in scenes corresponding to those plots. These results further emphasize that the attack does not require images of the exact same objects as present in the scene, but can also be carried out using images of similar instances from the same class of objects. Furthermore, preliminary results have indicated that these alignments also do not change significantly when the HLoc [3] based server uses stricter inlier thresholds (5,2 or 1 pixel as compared to the default 12 pixels) for the RANSAC based localization process.

3. Additional results for object presence classification

For the task object presence classification, we present precision and recall results for each of the objects in *IKEA-Objects* in Table 1.

References

[1] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superpoint: Self-supervised interest point detection and description, 2018. 1

[2] Jerome Revaud, Philippe Weinzaepfel, César Roberto de Souza, and Martin Humenberger. R2D2: repeatable and reliable detector and descriptor. In *NeurIPS*, 2019. 1

[3] Paul-Edouard Sarlin, Cesar Cadena, Roland Siegwart, and Marcin Dymczyk. From coarse to fine: Robust hierarchical localization at large scale. In *CVPR*, 2019. 1, 2

[4] Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. SuperGlue: Learning feature matching with graph neural networks. In *CVPR*, 2020. 1

[5] Johanna Wald, Torsten Sattler, Stuart Golodetz, Tommaso Cavallari, and Federico Tombari. Beyond controlled environments: 3d camera re-localization in changing indoor scenes. In *European Conference on Computer Vision (ECCV)*, 2020. 24

Figure 1. Scene01 of *IKEA-Scenes* with selected objects in focus.

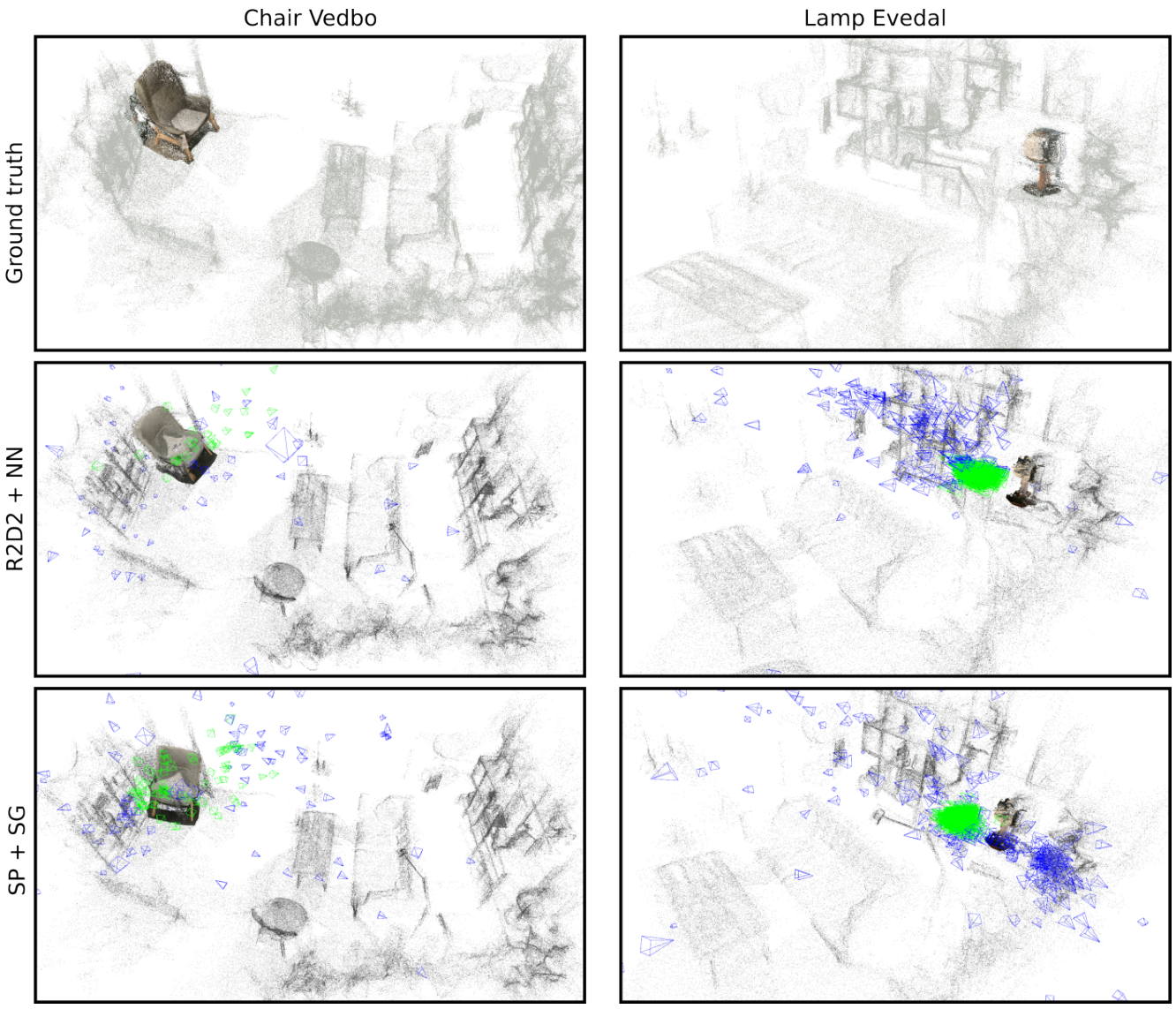

Figure 2. Qualitative results for aligning corresponding objects from *IKEA-Objects* using poses for localizing them in Scene01 of *IKEA-Scenes*.

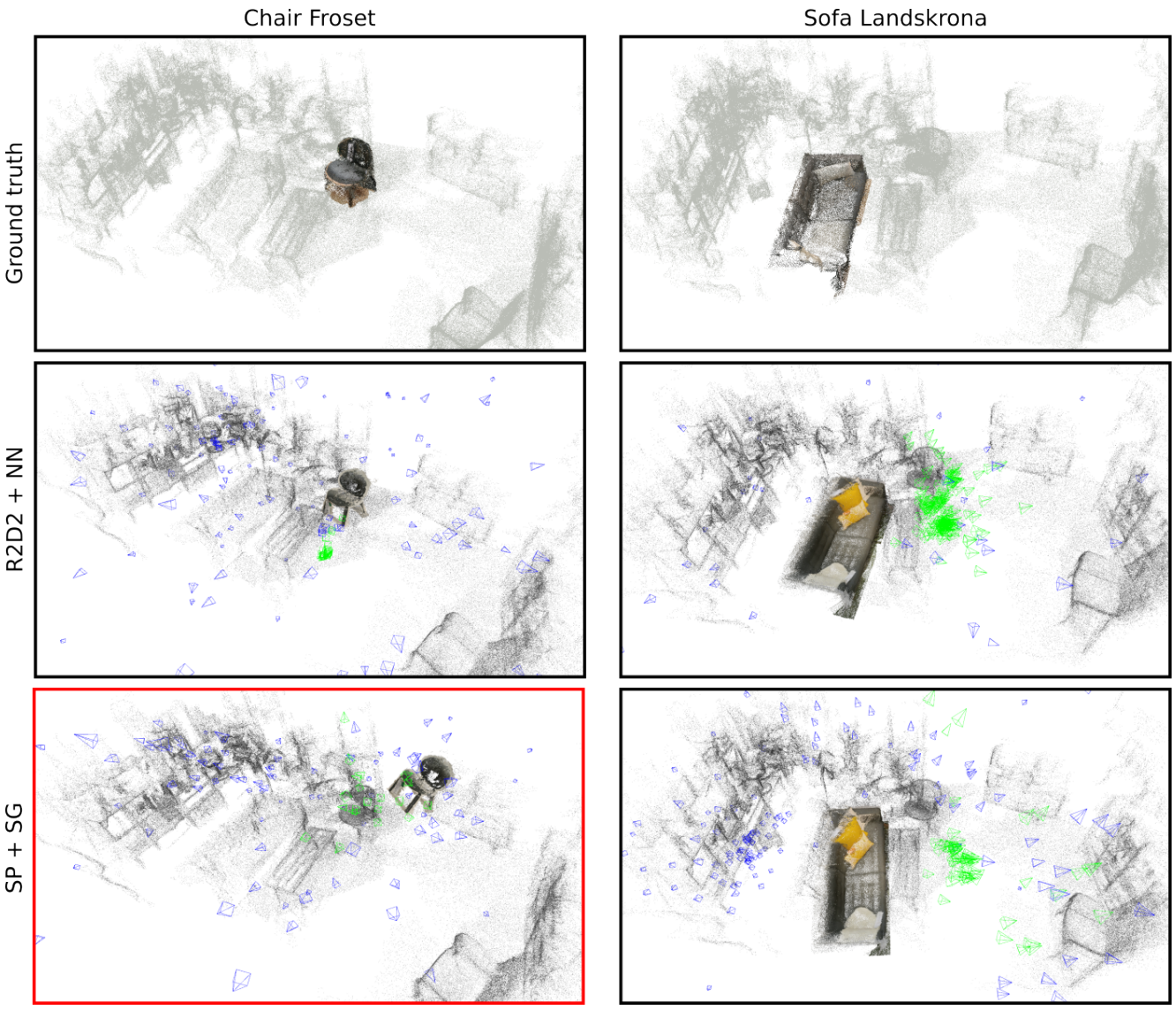

Figure 3. Qualitative results for aligning corresponding objects from *IKEA-Objects* using poses for localizing them in Scene01 of *IKEA-Scenes*. **Failure case**: The aligned model for Chair Froset when using Superpoint+Superglue is a bit far from the actual object and also oriented incorrectly.

Figure 4. Scene02 of *IKEA-Scenes* with selected objects in focus.

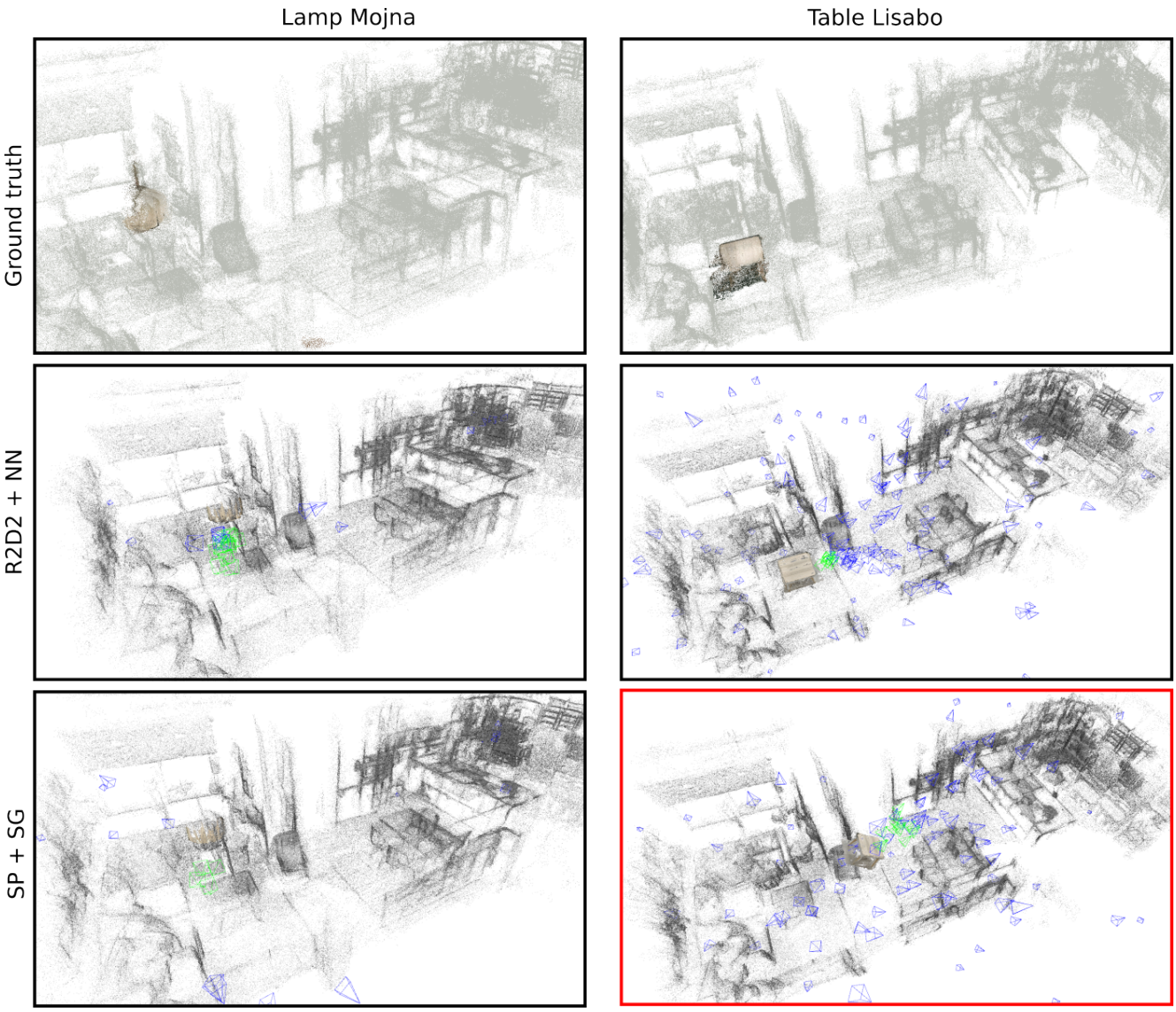

Figure 5. Qualitative results for aligning corresponding objects from *IKEA-Objects* using poses for localizing them in Scene02 of *IKEA-Scenes*. **Failure case**: The aligned model for Table Lisabo when using Superpoint+Superglue is a bit far from the actual object and also oriented incorrectly.

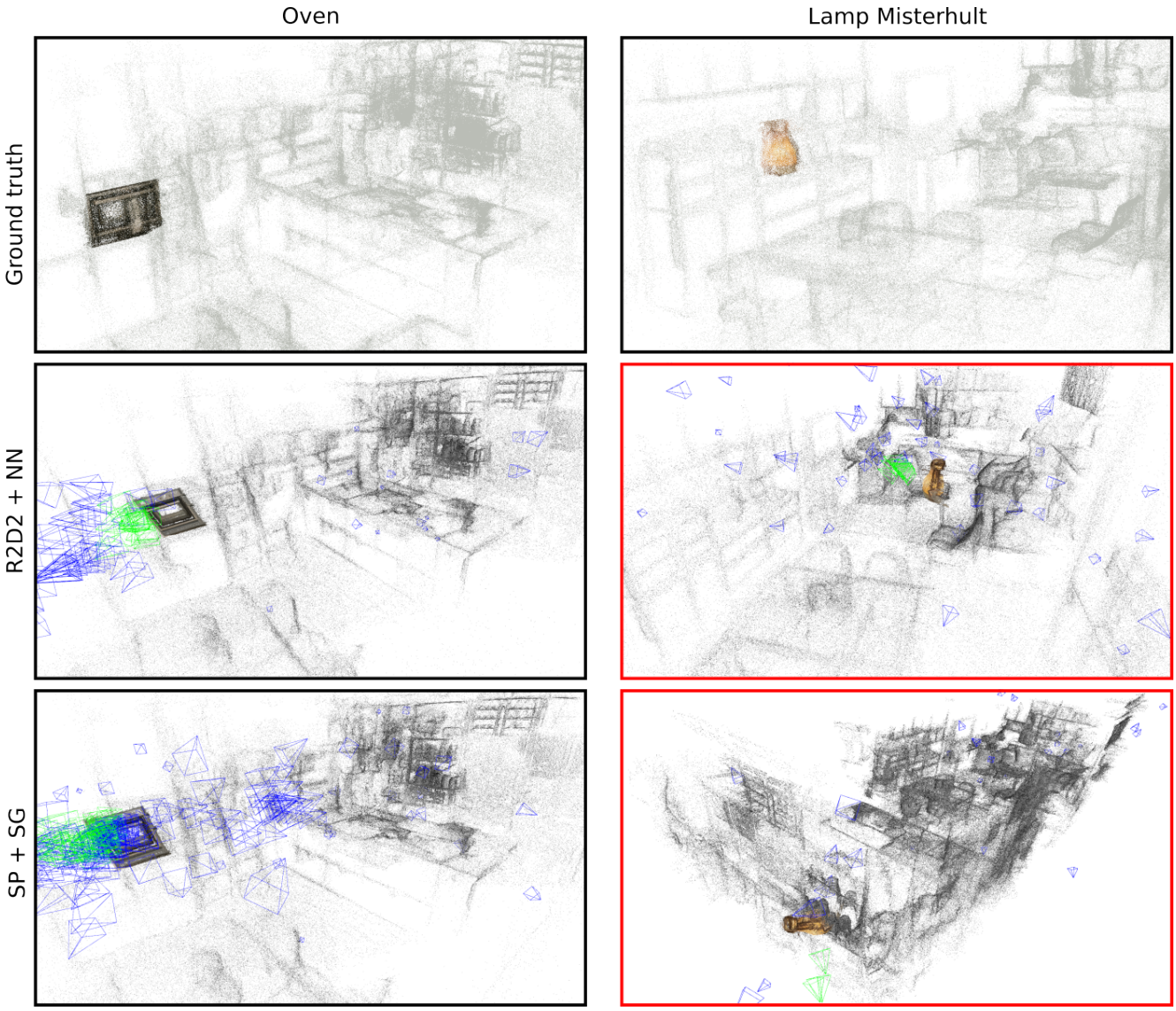

Figure 6. Qualitative results for aligning corresponding objects from *IKEA-Objects* using poses for localizing them in Scene02 of *IKEA-Scenes*. **Failure cases**: The aligned models of Lamp Misterhult when using R2D2+NN or Superpoint+Superglue are very far from the actual object. This can be attributed to the similar wooden appearance of the lamp and several wooden objects in the scene, which leads to incorrect matches.

Figure 7. Scene03 of *IKEA-Scenes* with selected objects in focus.

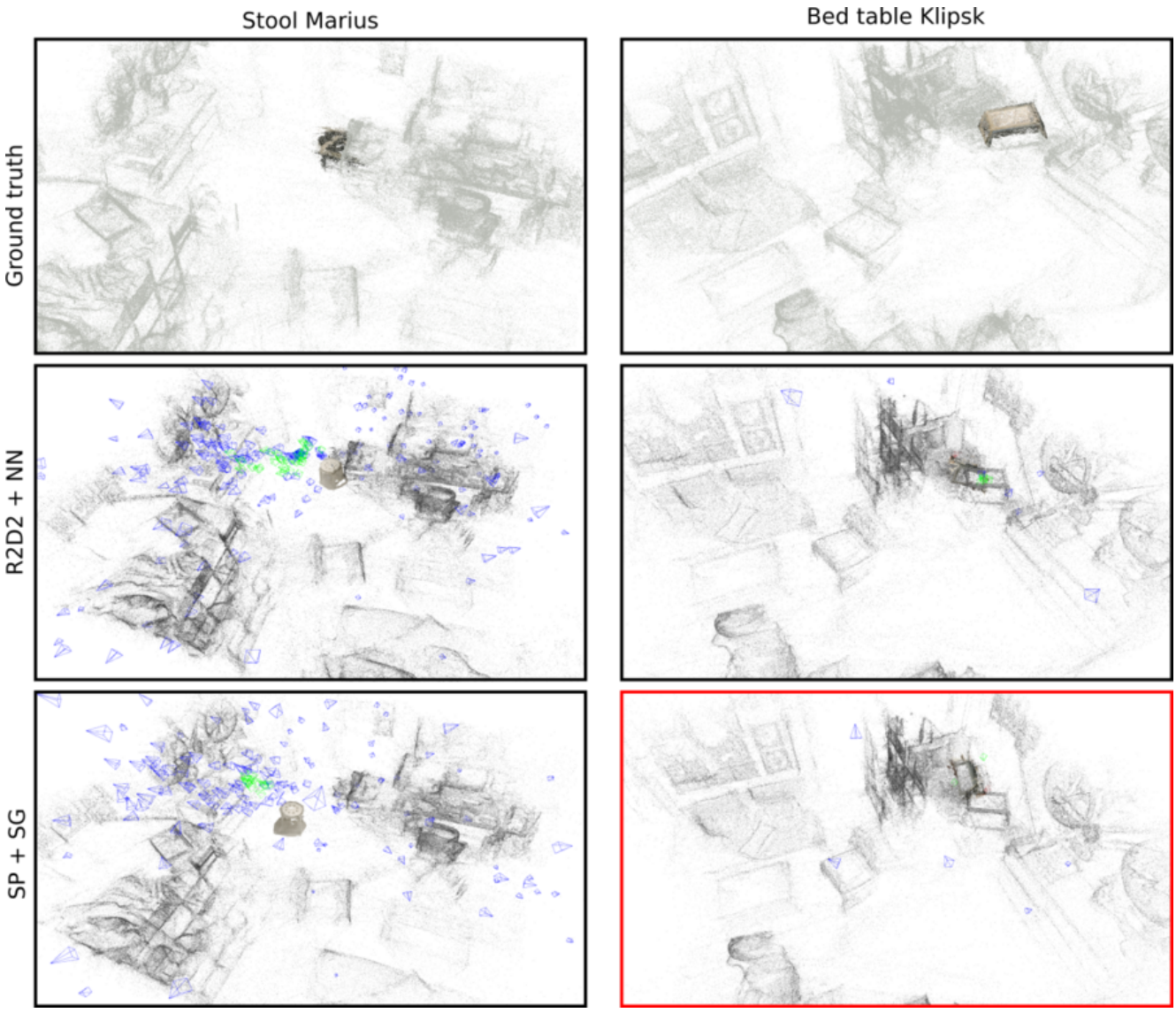

Figure 8. Qualitative results for aligning corresponding objects from *IKEA-Objects* using poses for localizing them in Scene03 of *IKEA-Scenes*. **Failure case**: The aligned model for Bed table Klipsk when using Superpoint+Superglue is close to the actual object, but incorrectly oriented.

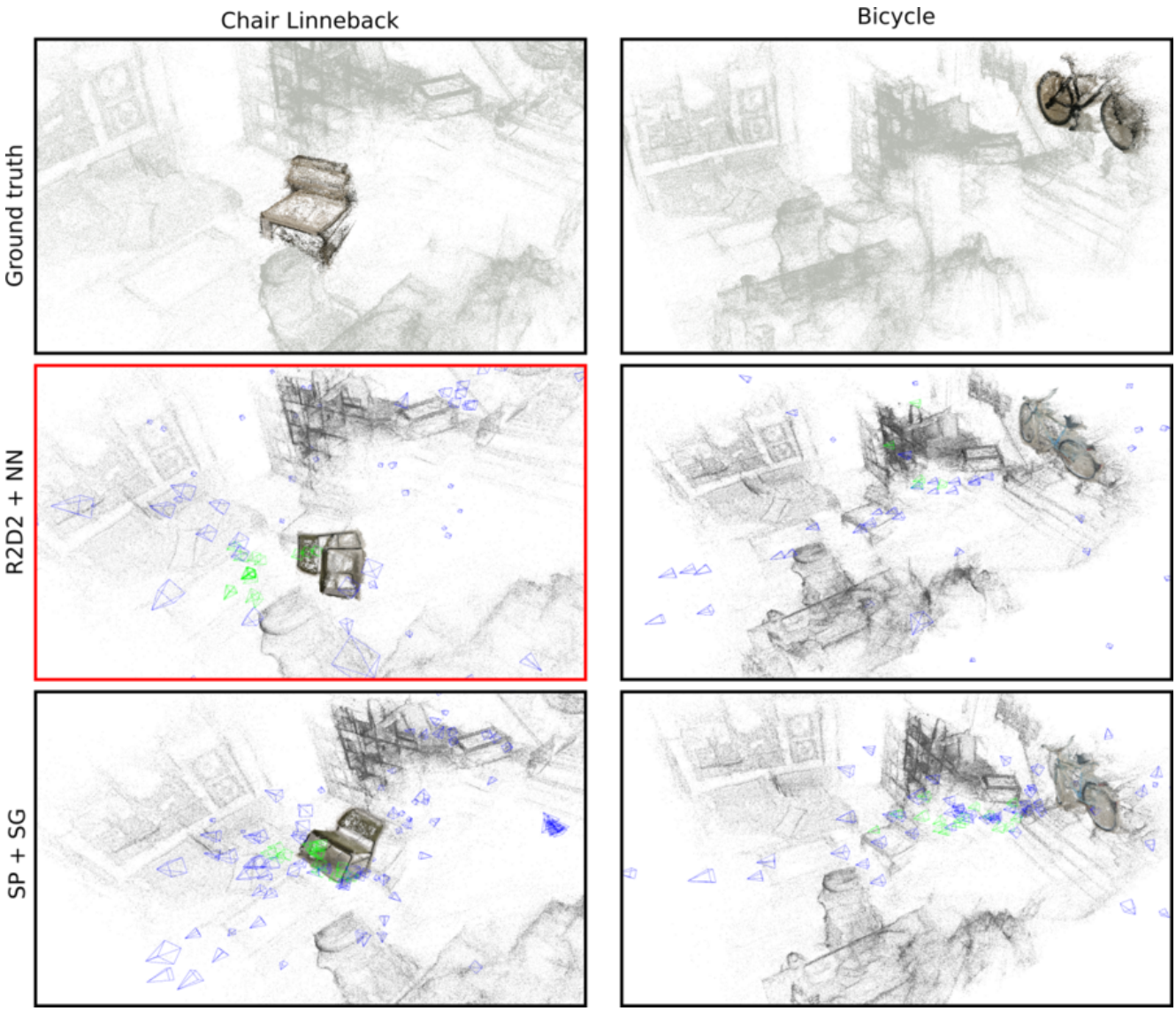

Figure 9. Qualitative results for aligning corresponding objects from *IKEA-Objects* using poses for localizing them in Scene03 of *IKEA-Scenes*. **Failure case**: The aligned model for Chair Linneback when using R2D2+NN is close to the actual object, but incorrectly oriented.

Figure 10. Scene04 of *IKEA-Scenes* with selected objects in focus.

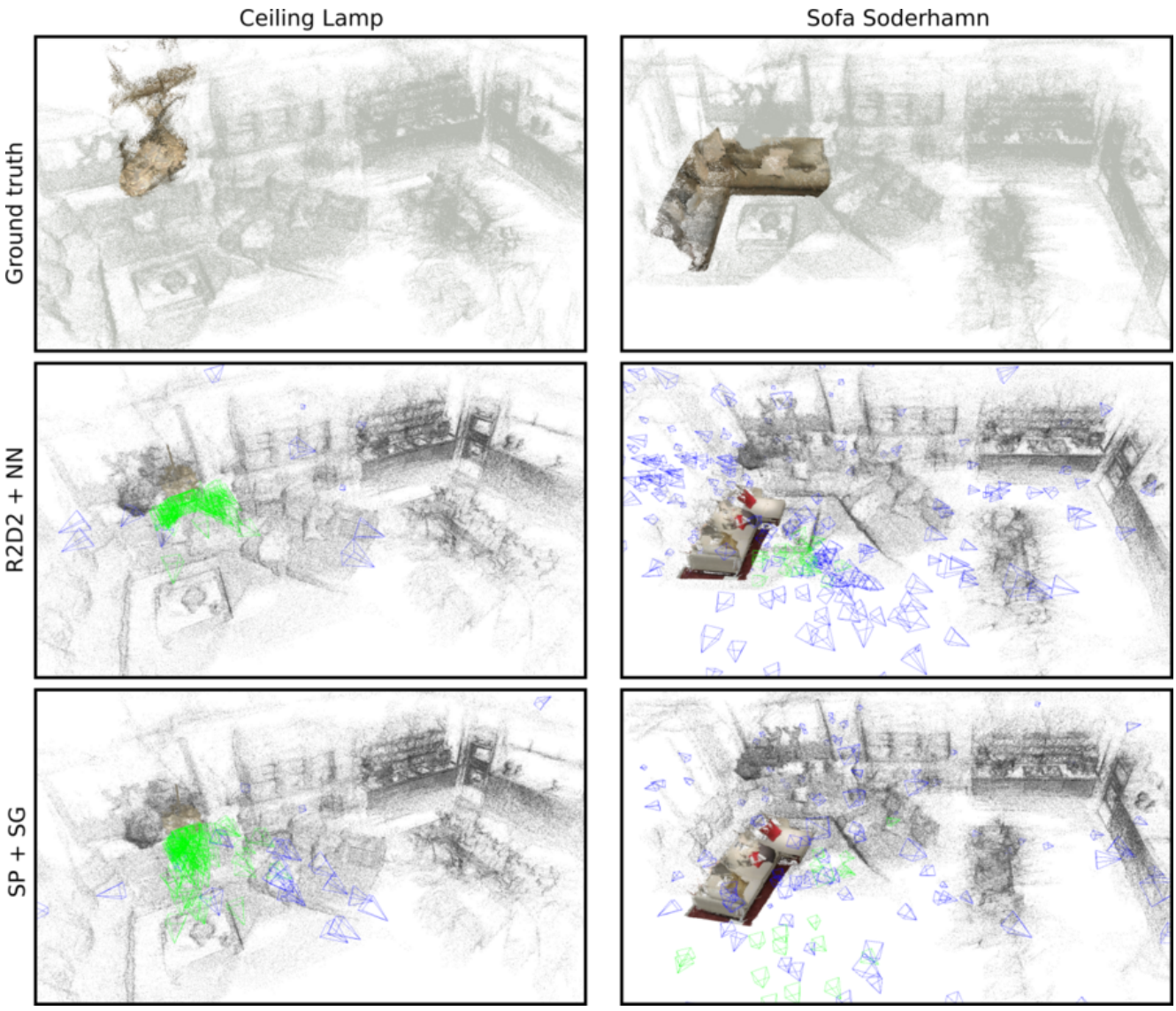

Figure 11. Qualitative results for aligning corresponding objects from *IKEA-Objects* using poses for localizing them in Scene04 of *IKEA-Scenes*.

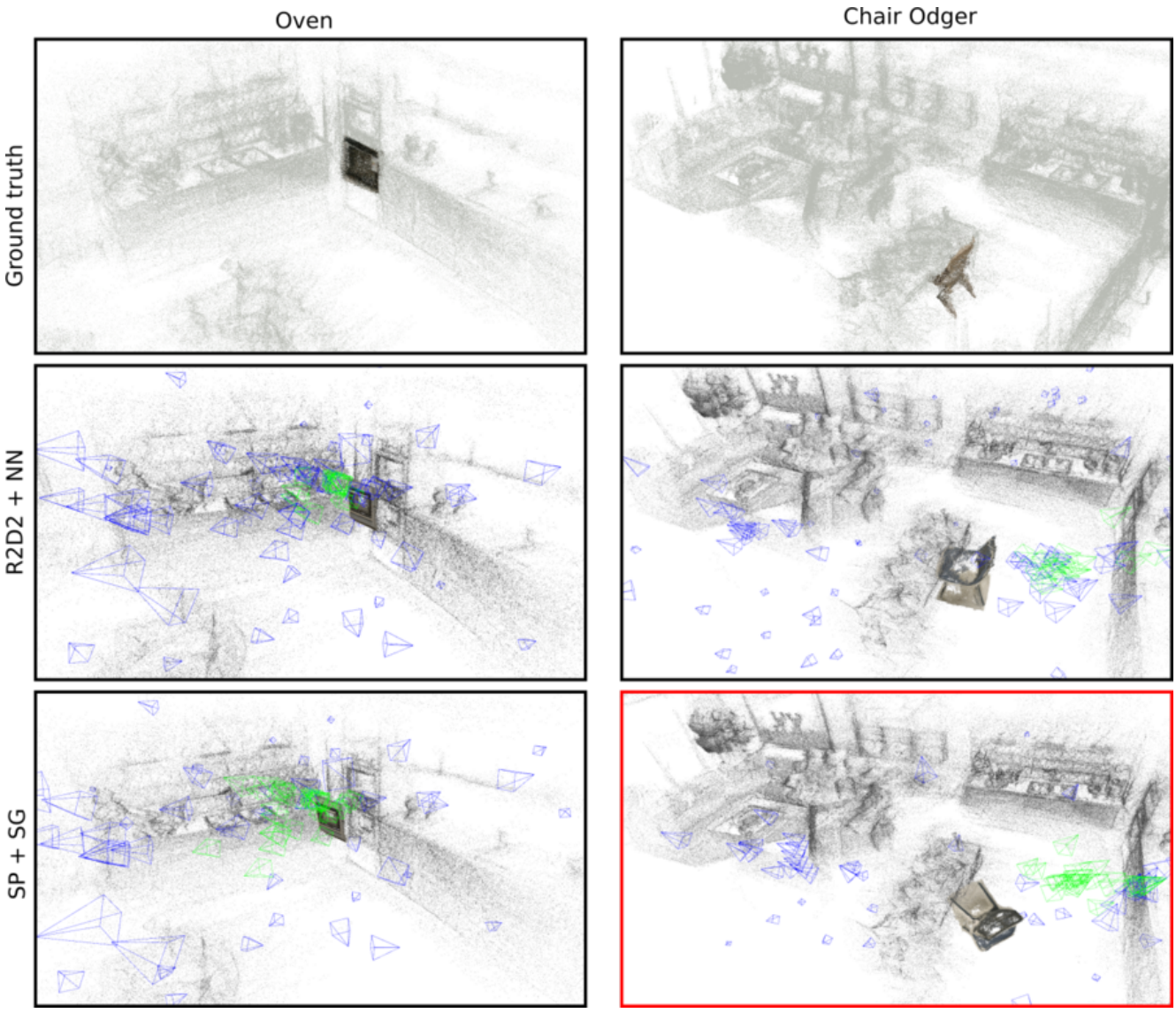

Figure 12. Qualitative results for aligning corresponding objects from *IKEA-Objects* using poses for localizing them in Scene04 of *IKEA-Scenes*. **Failure case**: The aligned model for Chair Odger when using Superpoint+Superglue is close to the actual object, but incorrectly oriented.

Figure 13. Scene05 of *IKEA-Scenes* with selected objects in focus.

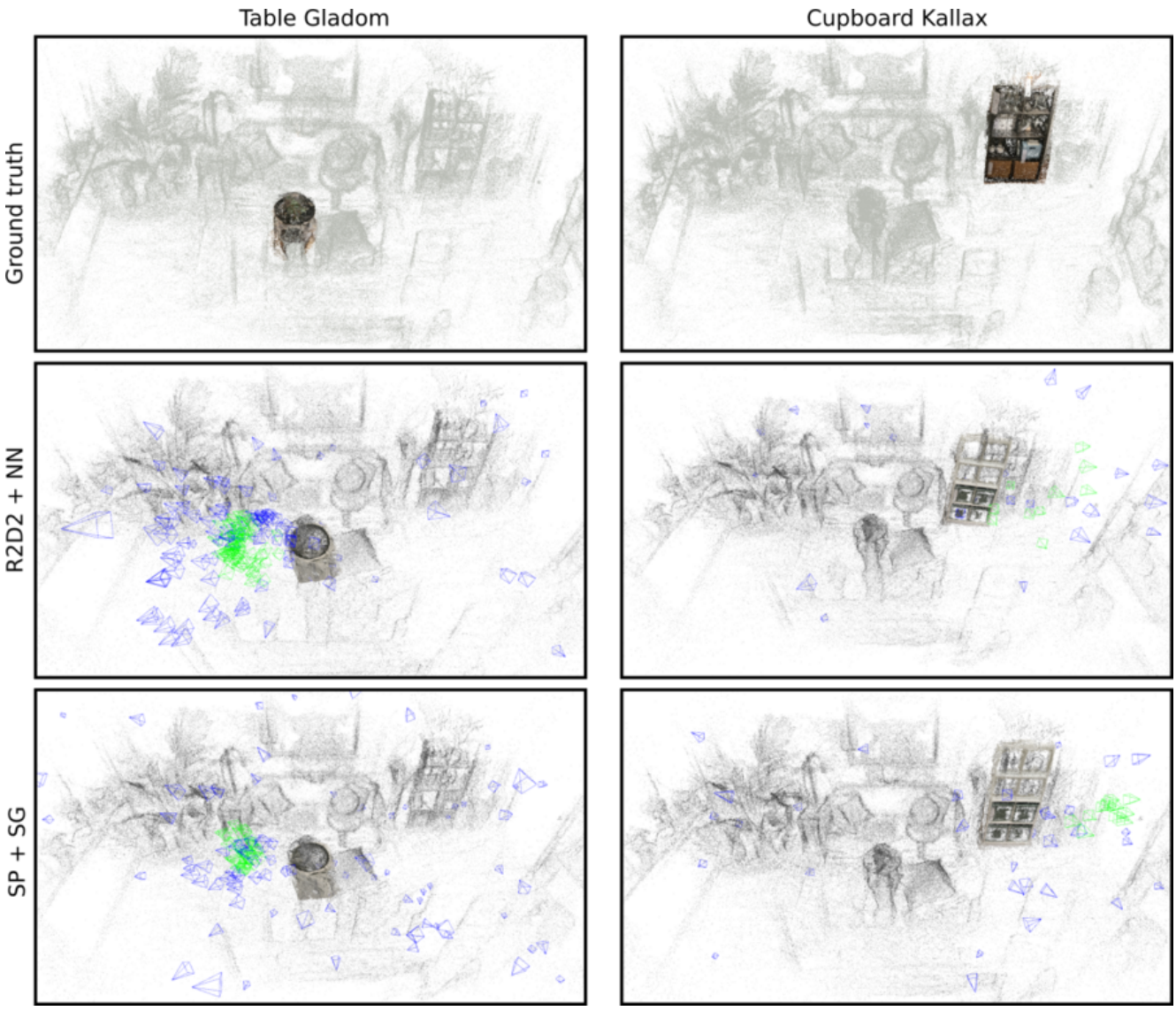

Figure 14. Qualitative results for aligning corresponding objects from *IKEA-Objects* using poses for localizing them in Scene05 of *IKEA-Scenes*.

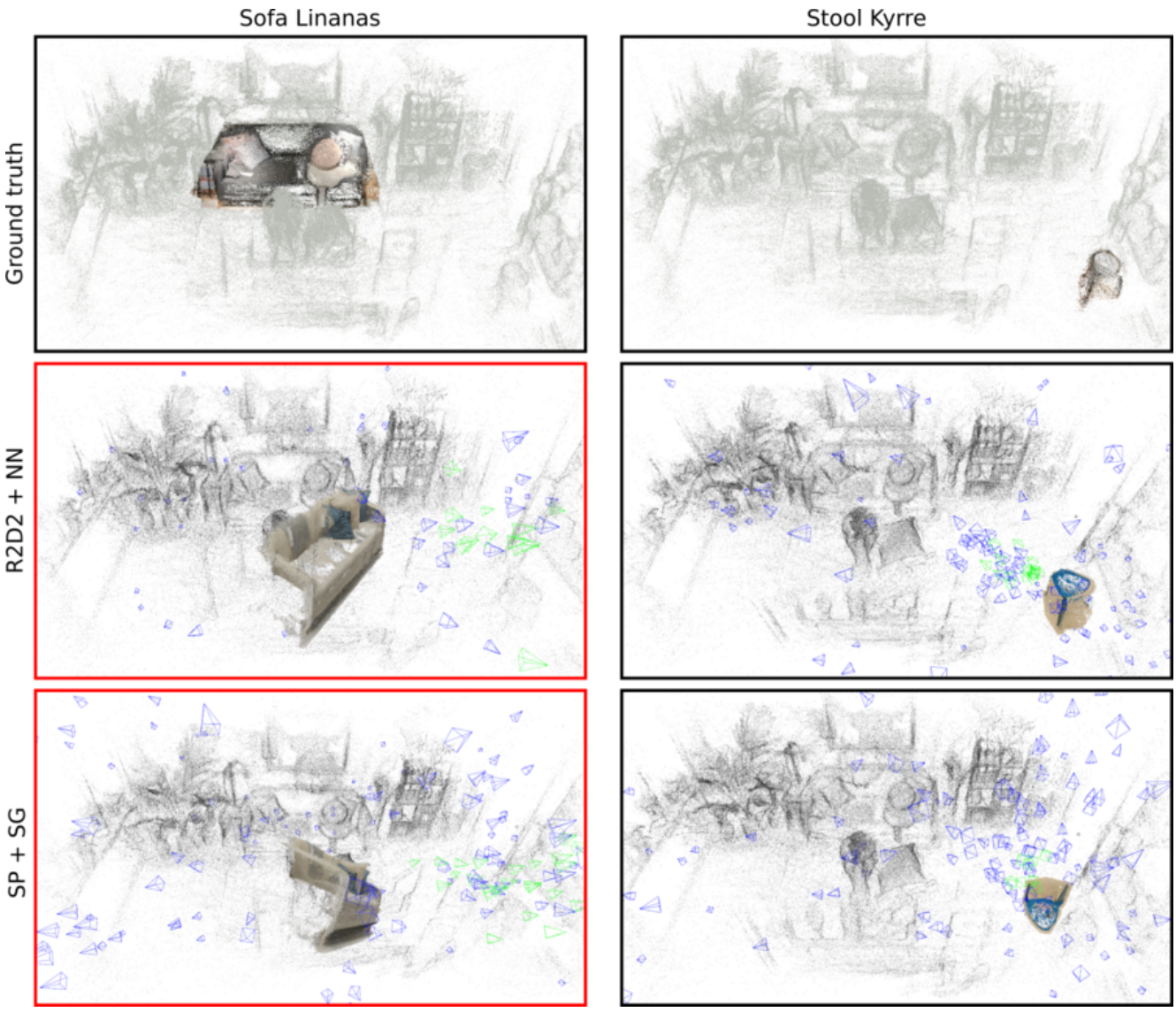

Figure 15. Qualitative results for aligning corresponding objects from *IKEA-Objects* using poses for localizing them in Scene05 of *IKEA-Scenes*. **Failure cases**: The aligned model for Sofa Linanas when using Superpoint+Superglue or R2D2+NN is incorrectly oriented and also far from the actual object.

Figure 16. Scene06 of *IKEA-Scenes* with selected objects in focus.

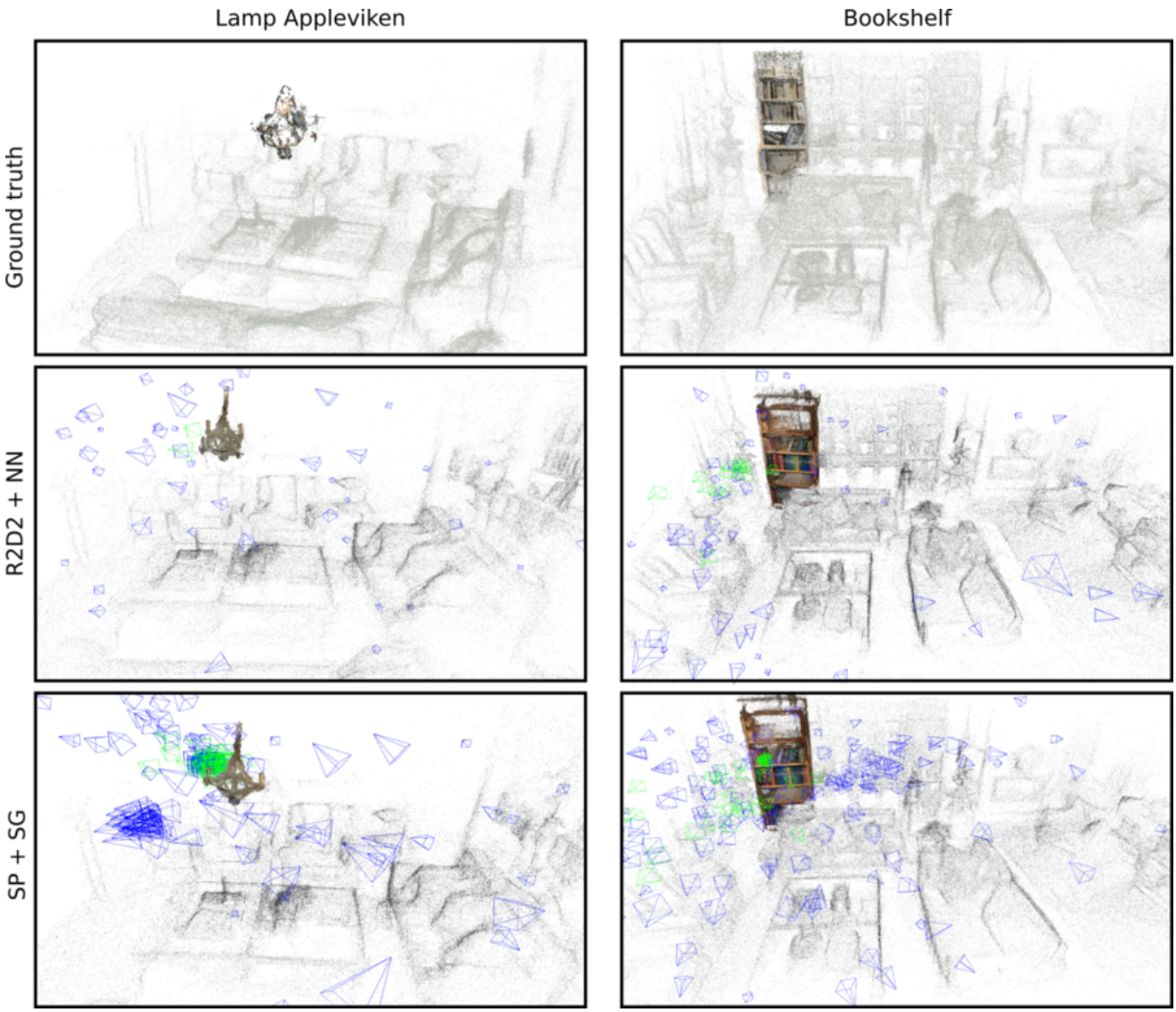

Figure 17. Qualitative results for aligning corresponding objects from *IKEA-Objects* using poses for localizing them in Scene06 of *IKEA-Scenes*.

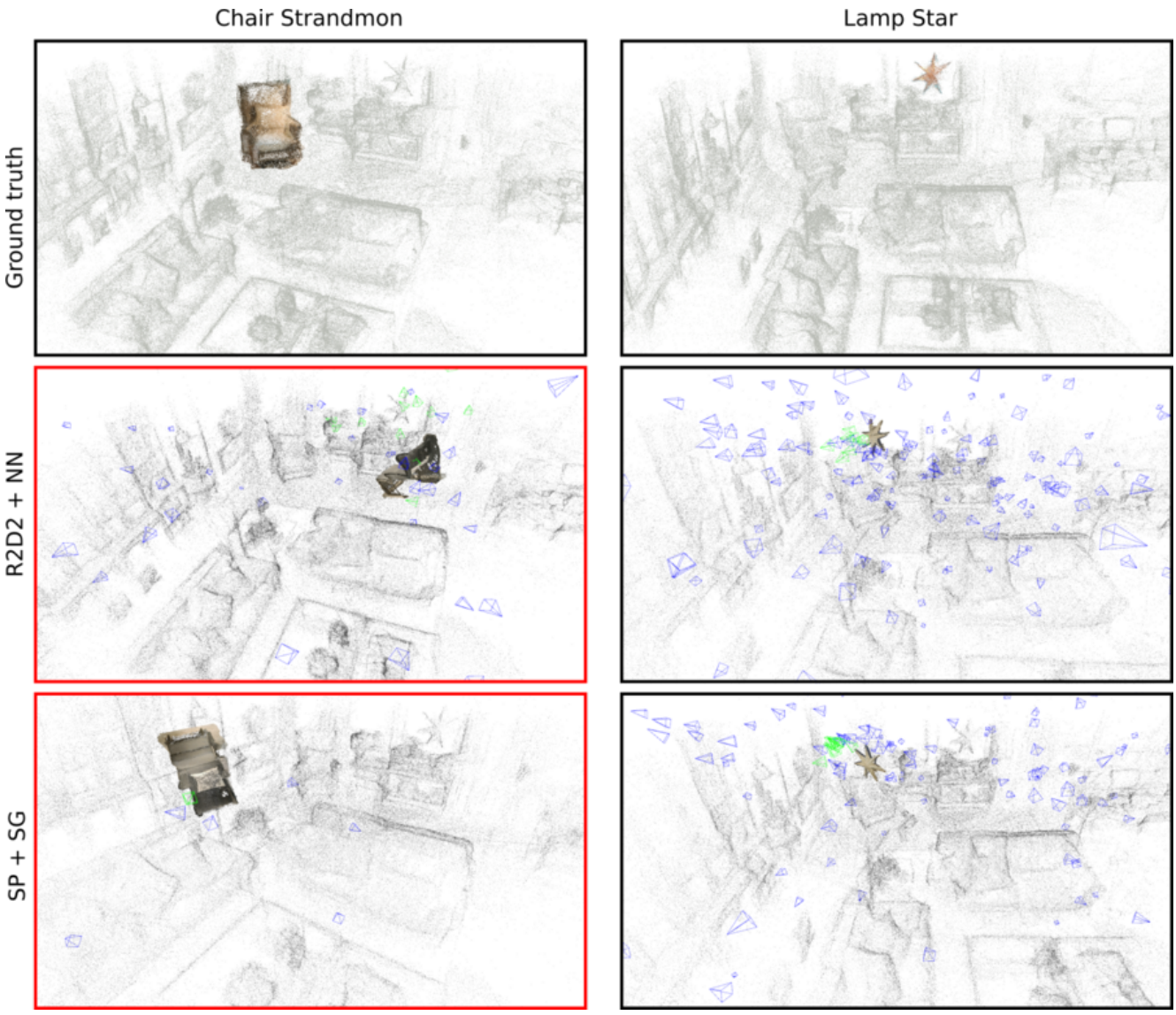

Figure 18. Qualitative results for aligning corresponding objects from *IKEA-Objects* using poses for localizing them in Scene06 of *IKEA-Scenes*. **Failure cases**: The aligned model for Chair Strandmon when using Superpoint+Superglue or R2D2+NN is incorrectly aligned and also far from the actual object.

Figure 19. Scene07 of *IKEA-Scenes* with selected objects in focus.

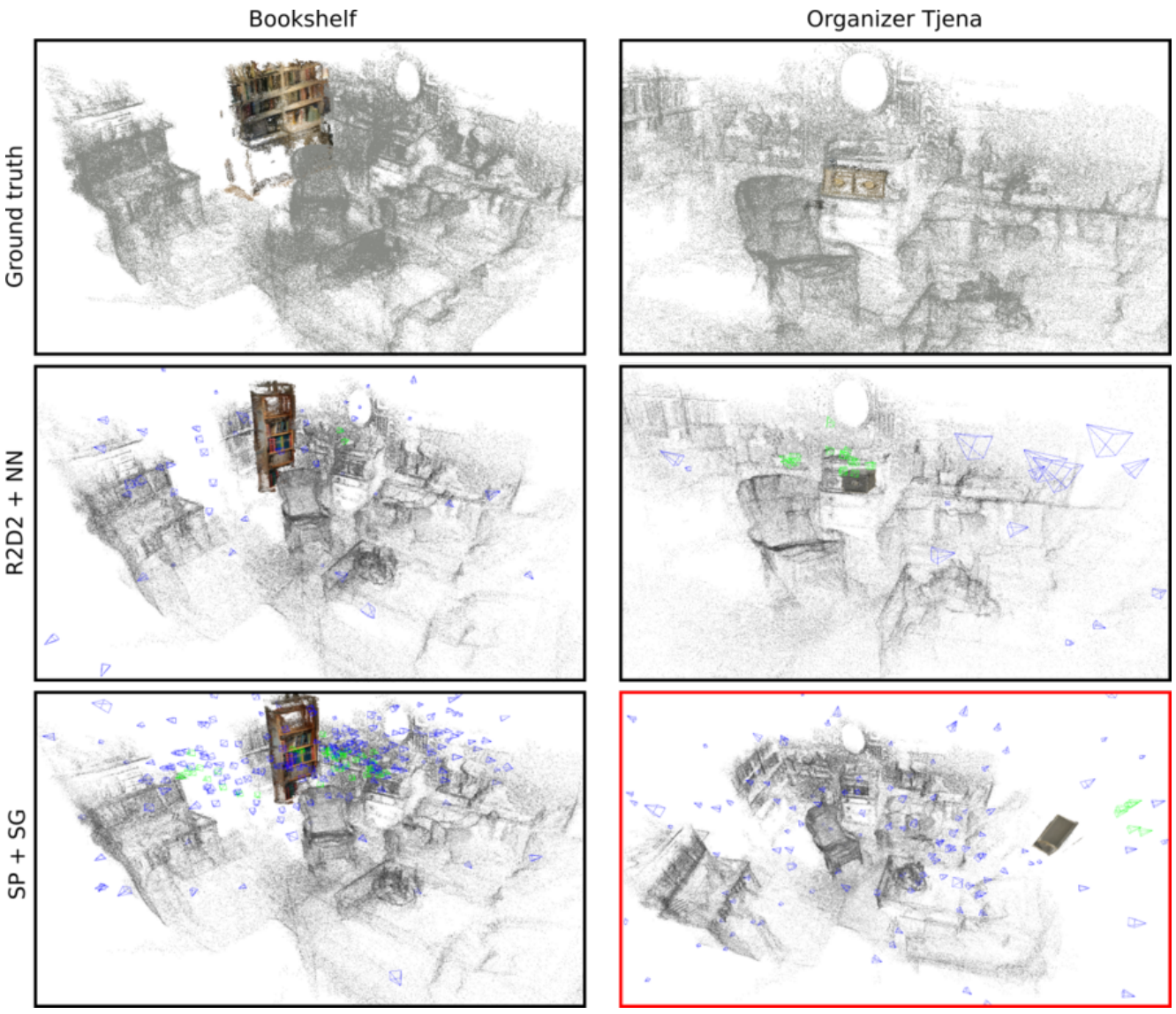

Figure 20. Qualitative results for aligning corresponding objects from *IKEA-Objects* using poses for localizing them in Scene07 of *IKEA-Scenes*. **Failure case**: The aligned model for Organizer Tjena when using Superpoint+Superglue is far from the actual object and also incorrectly oriented.

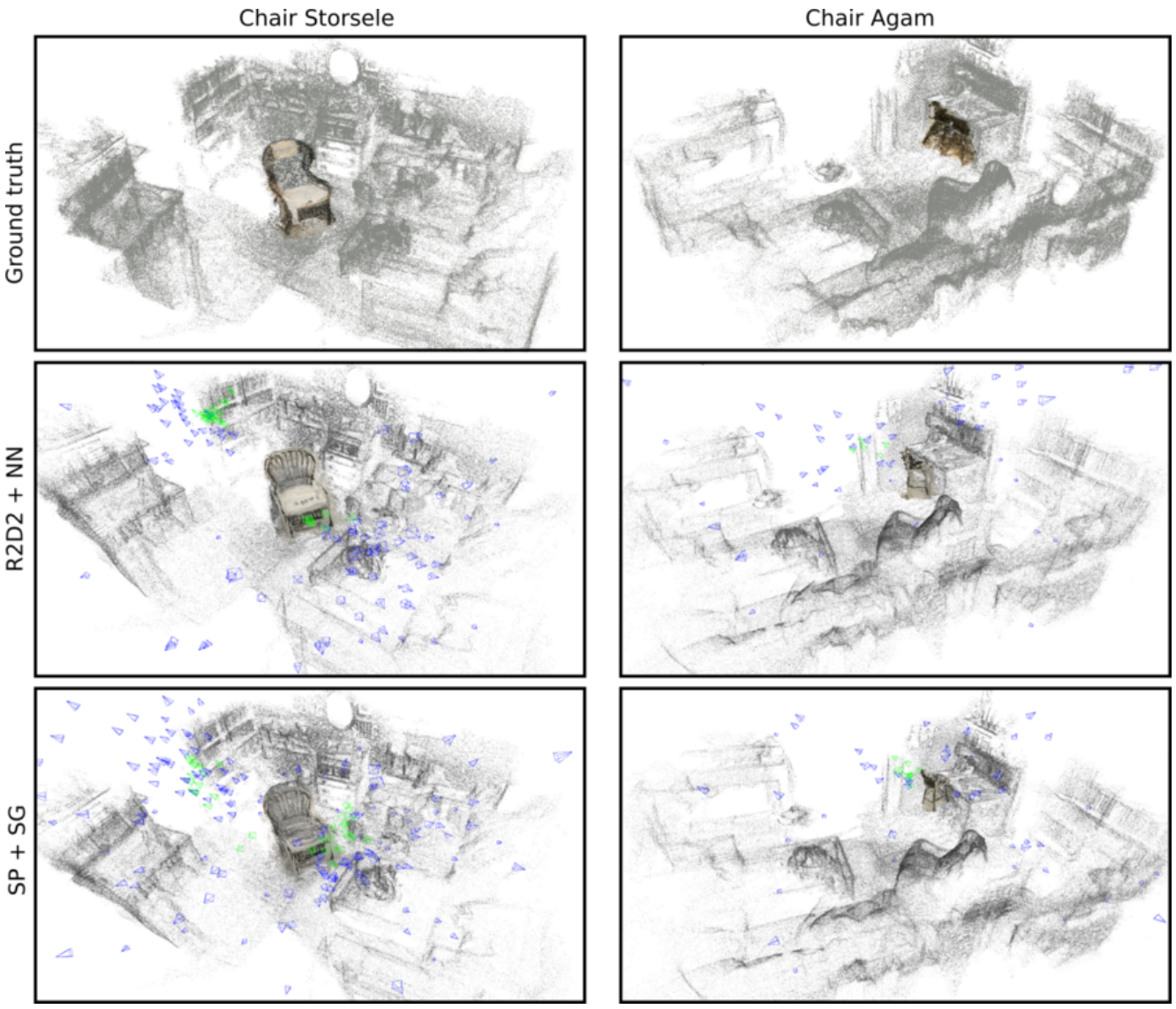

Figure 21. Qualitative results for aligning corresponding objects from *IKEA-Objects* using poses for localizing them in Scene07 of *IKEA-Scenes*.

RIO10 Scene04 + Office-Objects Chair

RIO10 Scene06 + IKEA-Objects oven

RIO10 Scene08 + IKEA-Objects Sofa Landskrona

Figure 22. Alignment results for querying scenes from the RIO10 dataset [5] with objects from the Office-Objects and IKEA-Objects datasets (*cf*. Fig. 6 in the main paper for quantitative results). The textured mesh corresponds to the scene and the point cloud corresponds to the 3D model of the object used for the attack. The alignment was produced using the poses obtained from the server, using Superpoint+Superglue for matching. As can be seen, it is possible to position the objects with reasonable accuracy, despite differences in appearance and geometry.

Object Name	Superpoint + Superglue						R2D2 + NN					
	10°,0.25m		30°, 0.5m		60°, 2m		10°,0.25m		30°, 0.5m		60°, 2m	
	P	R	P	R	P	R	P	R	P	R	P	R
bookshelf	1	0.5	0.5	0.5	1	0.5	0.5	0.5	1	0.5	0.5	0.5
chair_agam	0	0	0	0	0	0	0	0	0	0	0	0
chair_froset	0	0	0	0	0	0	0	0	0	0	0	0
chair_gaming	0.33	1	0.5	1	0	0	1	1	1	1	0.5	1
chair_linneback	0.5	1	0.5	1	0.5	1	0.5	1	0.5	1	0.5	1
chair_odger	0.33	0.5	0.33	0.5	0.5	0.5	1	0.5	1	0.5	0.5	0.5
chair_poang_small	1	1	1	1	1	1	1	1	1	1	1	1
chair_storsele	0.33	1	0.33	1	1	1	0.5	1	0.25	1	0.5	1
chair_strandmon	0	0	0	0	0	0	0.5	1	0.5	1	0.5	1
chair_vedbo	1	1	1	1	0.5	1	0.5	1	0.5	1	0.5	1
cupboard_hauga	1	1	1	1	1	1	1	1	1	1	1	1
cupboard_kallax	1	1	1	1	1	1	0	0	0	0	0.25	1
cycle	0	0	0	0	0	0	0.5	1	0.5	1	0.5	1
klipsk_bed_table	1	1	1	1	1	1	1	1	1	1	0.5	1
lamp_ceiling_agunarryd	1	1	1	1	1	1	0.33	1	0	0	0.33	1
lamp_ceiling_appleviken	1	1	1	1	1	1	0	0	0	0	0	0
lamp_ceiling_mojna	0	0	0	0	0	0	0	0	0	0	0	0
lamp_ceiling_nymane	0	0	1	0.5	0.5	0.5	1	0.5	0	0	0.25	0.5
lamp_ceiling_ranarp	0	0	0	0	0	0	0	0	0	0	0	0
lamp_evedal	1	1	1	1	1	1	1	1	1	1	1	1
lamp_fancy	1	1	1	1	1	1	1	1	1	1	1	1
lamp_navlinge	0.33	0.5	0.5	0.5	0.5	0.5	0.25	0.5	0.33	0.5	0.25	0.5
lamp_star	0	0	0	0	0	0	1	1	1	1	1	1
lamp_table_misterhult	1	1	1	1	1	1	0	0	0	0	0	0
lamp_table_nymane	0.5	0.5	1	0.5	0.33	0.5	1	0.5	1	0.5	1	0.5
lamp_table_tertial	0	0	0.5	1	1	1	0.5	1	0	0	0.33	1
organizer_kvarnik	0.5	1	0	0	0	0	1	1	1	1	0.5	1
oven	1	0.5	1	0.5	1	0.5	1	0.5	1	0.5	0.5	0.5
sofa_landskrona	1	1	1	1	0.5	1	1	1	1	1	1	1
sofa_linanas	0	0	0	0	0	0	0	0	0	0	0	0
sofa_soderhamn	0	0	0	0	0	0	0	0	0	0	0	0
stool_kyrre	1	1	1	1	1	1	1	1	1	1	1	1
stool_marius	1	1	1	1	1	1	1	1	1	1	1	1
strainer	0	0	0	0	0	0	0	0	0	0	0	0
table_corner_gladom	1	1	1	1	1	1	0.5	1	0.5	1	0.5	1
table_lisabo_square	1	0.5	0.5	0.5	0.25	0.5	1	0.5	1	0.5	1	0.5
vas_gradvis	0.25	1	1	1	0.5	1	0	0	0	0	0.5	1
wall_hanging_crescent	0.33	0.5	0.25	0.5	0.33	0.5	0	0	1	0.5	0.5	0.5

Table 1. Object wise precision and recall results for *IKEA-Objects* when attacking *IKEA-Scenes*