Deep Global Registration: Supplementary Materials

Christopher Choy
Stanford University

Wei Dong
Carnegie Mellon University

Vladlen Koltun
Intel Labs

1. Weighted Procrustes

In this section, we provide the full proof of Theorem 1 in the paper. The weighed mean squared error is defined as follows

$$e^2 = \sum_{(i,j)\in\mathcal{M}} \tilde{w}_{(i,j)}(\mathbf{y}_j - (R\mathbf{x}_i + \mathbf{t}))^2 \tag{1}$$

$$= \text{Tr}\left((Y - RX - \mathbf{t}\mathbf{1}^T)W(Y - RX - \mathbf{t}\mathbf{1}^T)^T\right). \tag{2}$$

Theorem 1 *: The R and $\boldsymbol{t}$ that minimize the squared error $e^2(R,\boldsymbol{t}) = \sum_{(i,j)\in\mathcal{M}} w_{(i,j)}(\boldsymbol{y}_j - R\boldsymbol{x}_i - \boldsymbol{t})^2$ are $\hat{\boldsymbol{t}} = (Y - RX)W\mathbf{1}$ and $\hat{R} = USV^T$ where $U\Sigma V^T = SVD(\Sigma_{xy})$, $\Sigma_{xy} = YKWKX^T$, $K = I - \sqrt{\tilde{\mathbf{w}}}\sqrt{\tilde{\mathbf{w}}}^T$, and $S = diag\,(1, \cdots, 1, det(U)det(V))$.*

Proof. First, we differentiate e^2 w.r.t. $\mathbf{t}$ and equates the partial derivative to 0:

$$\frac{\partial}{\partial \mathbf{t}} e^2 = \frac{\partial}{\partial \mathbf{t}} \sum_{(i,j)\in\mathcal{M}} \tilde{\mathbf{w}}_{(i,j)}(\mathbf{y}_j - R\mathbf{x}_i - \mathbf{t}) \tag{3}$$

$$= -2\left(\sum_{(i,j)} \tilde{\mathbf{w}}_{(i,j)}\mathbf{y}_j - \sum_{(i,j)} \tilde{\mathbf{w}}_{(i,j)}R\mathbf{x}_i - \sum_{(i,j)} \tilde{\mathbf{w}}_{(i,j)}\mathbf{t}\right) = 0. \tag{4}$$

Thus, $\hat{\mathbf{t}} = (Y - RX)W\mathbf{1}$. Next, we substitute $X = KX + X\sqrt{\tilde{\mathbf{w}}}\sqrt{\tilde{\mathbf{w}}}^T$ on Eq. 2 and do the same for Y:

$$e^2 = \text{Tr}\left((Y - RX - \mathbf{t}\mathbf{1}^T)W(Y - RX - \mathbf{t}\mathbf{1}^T)^T\right) \tag{5}$$

$$= \text{Tr}\left((YK + Y\sqrt{\tilde{\mathbf{w}}}\sqrt{\tilde{\mathbf{w}}}^T - RXK - RX\sqrt{\tilde{\mathbf{w}}}\sqrt{\tilde{\mathbf{w}}}^T - \mathbf{t}\mathbf{1}^T)W(YK + Y\sqrt{\tilde{\mathbf{w}}}\sqrt{\tilde{\mathbf{w}}}^T - RXK - RX\sqrt{\tilde{\mathbf{w}}}\sqrt{\tilde{\mathbf{w}}}^T - \mathbf{t}\mathbf{1}^T)^T\right)$$

$$= \text{Tr}((YK - RXK)W(YK - RXK)^T) \tag{6}$$

$$= \text{Tr}(YKWK^TY^T) + \text{Tr}(RXKWK^TX^TR^R) - 2\text{Tr}(YKWK^TX^TR^T), \tag{7}$$

where we use the fact that $W\mathbf{1}\mathbf{1}^T = \sqrt{\tilde{\mathbf{w}}}\sqrt{\tilde{\mathbf{w}}}^T$. The minimum occurs when we maximize the last negative term:

$$\max_R \text{Tr}(YKWK^TX^TR^T) = \sum_k \sigma_k(YKWK^TX^T), \tag{8}$$

where $\sigma_k(A)$ is the k-th largest singular value of the matrix A. Thus, the maximum of Eq. 8 occurs when $R = USV^T$ where $U\Sigma V^T = \text{SVD}(\Sigma_{xy})$, $\Sigma_{xy} = YKWKX^T$ and $S = \text{diag}\,(1, \cdots, 1, \det(U)\det(V))$. The last $\det(U)\det(V)$ is either +1 or -1 depending on the direction of the orthonomal basis. □

2. Additional scene-wise statistics

We report scene-wise 3DMatch registration recalls and error metrics of our method. We find that our method performs relatively worse on *Study, Lab*, which include noisy long range scans and objects with similar geometry that introduce ambiguity. One possible future extension of our model is batch optimization over multiple scans jointly to enhance robustness.

In Fig. 1, we show another visualization of Fig.4 in the paper, which emphasizes the recall with more strict bounds. We observe that our approach still significantly outperforms baselines at a finer scale.

Table 1: Scene-wise statistics for our method on 3DMatch test set.

Scene	Recall	TE (cm)	RE (deg)	Time (s)
Kitchen	94.5%	6.30	2.25	0.70
Home1	89.7%	7.78	2.34	0.66
Home2	77.9%	8.31	3.49	0.62
Hotel1	92.9%	7.22	2.38	0.74
Hotel2	85.5%	6.49	2.43	0.72
Hotel3	79.6%	7.13	2.44	0.68
Study	69.9%	11.59	3.23	0.73
Lab	72.7%	8.36	2.04	0.73

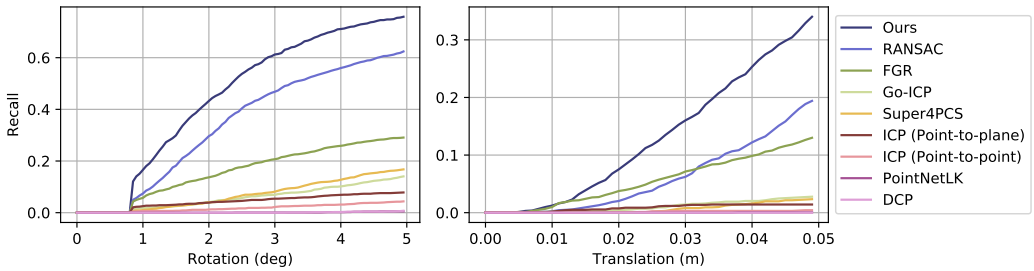


Figure 1: Overall pairwise registration recall (y-axis) on 3DMatch test set with varying rotation (left image) and translation (right image) error thresholds (x-axis).

3. Analysis of the voxel size

In the main paper, we empirically select 5cm as the voxel size for quantizing input point clouds. In Table 2 and Fig. 3, we report registration results of networks trained with 2.5cm voxel down-sampling. The name next to the voxel size indicates the robust optimization module initialization method which we will discuss in the next section. In general, 2.5cm voxel downsampling leads to lower registration recall. We suspect that high-resolution point clouds results in relatively fewer inliers. Thus, the pose optimiation stage fails more often due to lower inlier ratio. However, we observe that translation and rotation errors are lower for 2.5cm voxelization due to high-resolution registration.

Table 2: Registration results w/ and w/o Weighted Procrustes, with different voxel sizes.

	Recall	TE (cm)	RE (deg)	Time (s)
2.5cm-weighted	82.4%	7.61	2.44	2.24
2.5cm-nonweighted	81.9%	**7.21**	**2.39**	2.27
5cm-weighted	**85.2%**	7.73	2.58	0.70
5cm-nonweighted	85.0%	7.44	2.42	**0.69**

4. Analysis of Procrustes methods

In this section, we provide further analysis on the Weighted Procrustes method. First, we compare the registration recall of our Weighted Procrustes and non-weighted Procrustes method in a pair-wise registration pipeline in Fig. 3 and Table 2. For non-weighted Procrustes method, we simply weigh all correspondences above $\tau = 0.05$ to be inliers and use the computed pose to initialized the robust registration stage. As the robust registration stage can also handle outliers, the improvement from the Weighted Procrustes method is marginal.

Next, we analyze the effect of the clipping threshold on registration accuracy. The Weighted Procrustes rely on the weight vector $\mathbf{w}$ to filter out noise and outliers. We use a sigmoid function to map a logit score from a neural network to a weight value that range from 0 to 1. However, to achieve $w = 0$ for noise and outlier correspondences, a neural network must predict

a logit score of $-\infty$ which is not possible. Thus, we use a clip function $\phi(w) = I(w > \tau)w$ element-wise to filter small weights. We use $\tau = 0.05$ in all our experiments in the main paper, but we report the registration accuracy for varying τ in Fig. 2 at test time.

Note that increasing τ improves the accuracy of inlier prediction but decreases the recall of retrieving inlier correspondences. We find a marginal gain may be achieved at 0.25, giving us a sense how tolerant our Weighted Procurstes is regarding the clip threshold.

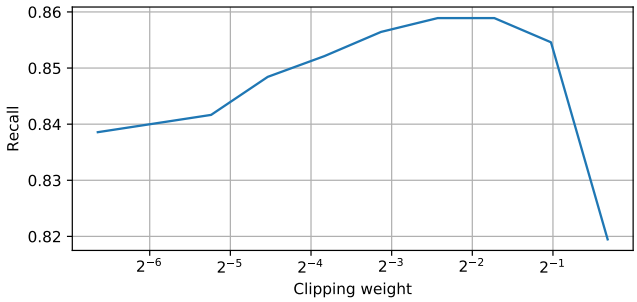


Figure 2: Ablation of clipping threshold τ. The curve is relatively stable, reaching its peak at around 0.25. X-axis is in log scale.

5. Multi-way registration

We demonstrate the qualitative full scene reconstruction on the large scale indoor LiDAR RGBD dataset [1] in Fig. 4, where our method is used for an open-source multi-way registration pipeline [6]. In Fig. 5, we visualize successful global registrations as green edges in the pose graph. Our method provides dense true-positive pairwise global registrations, which improve the global accuracy of scene reconstruction.

6. Additional qualitative results and failure cases

We show additional qualitative comparisons with baselines in Fig. 6 and successful registration of our method in Fig. 7. We also list typical failure cases of our method in Fig. 8, where small overlaps, similar object shapes, and repetitive structures result in incorrect registrations.

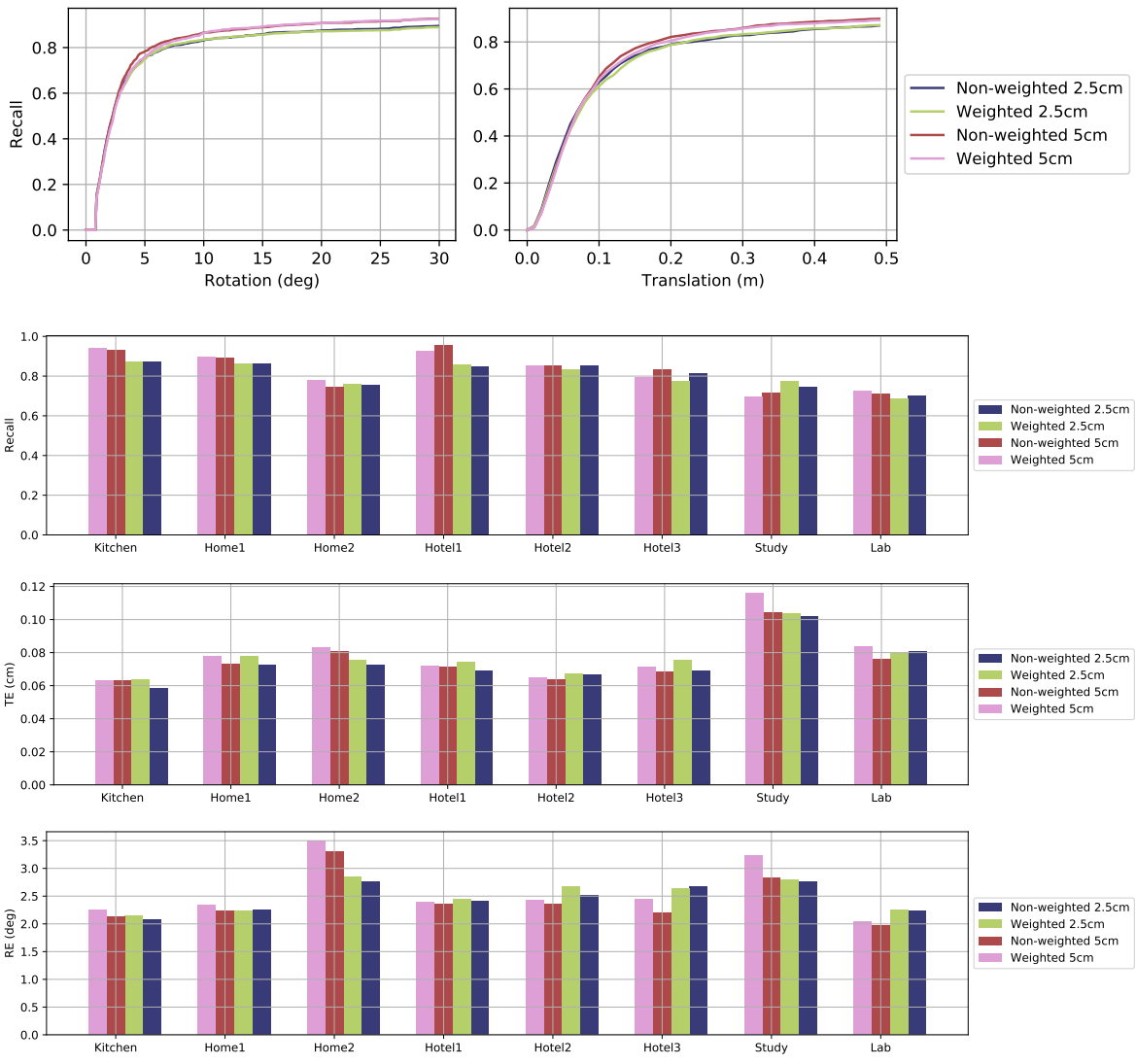


Figure 3: Scene-wise analysis of recall, TE, and RE w/ and w/o Weighted Procrustes on the 3DMatch [4] dataset.

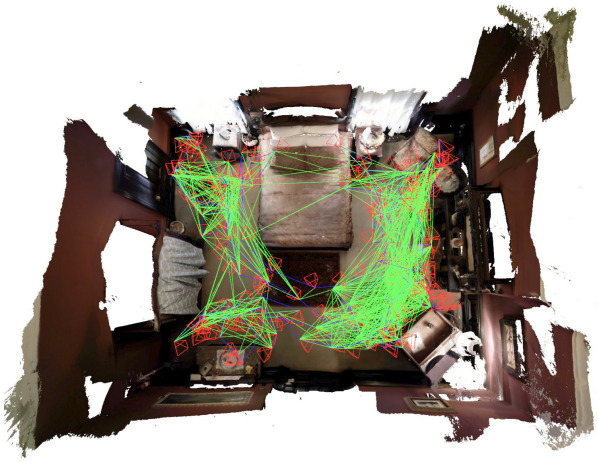
Bedroom

Boardroom

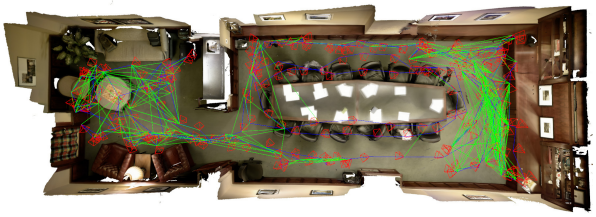
Loft

Apartment

Figure 4: Full scene reconstructions from a modified multi-way registration pipeline with deep global registration for pairwise registration.



Bedroom

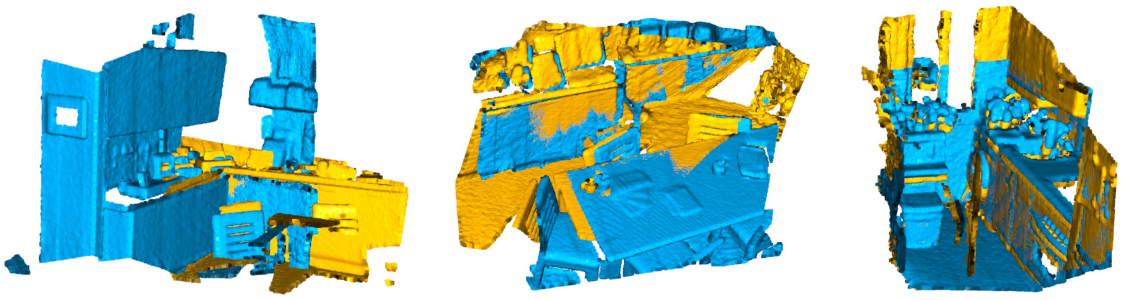


Boardroom

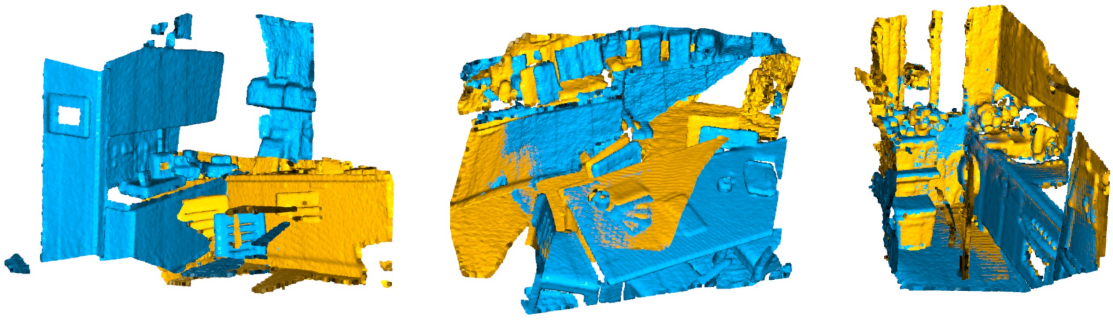
Loft

Apartment

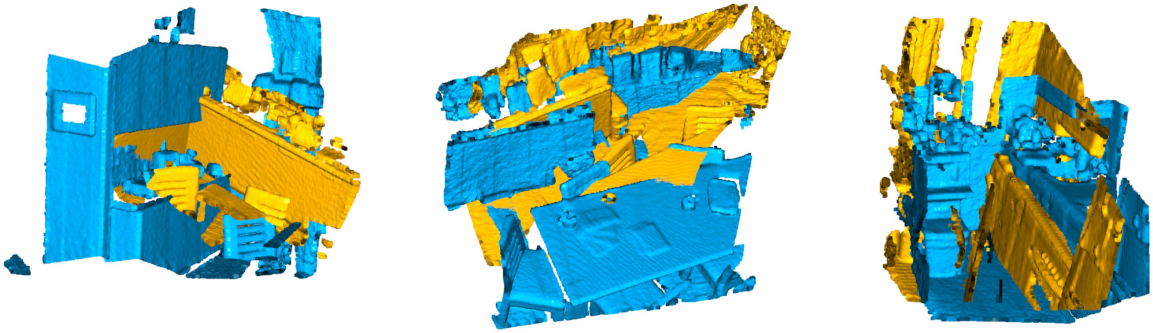
Figure 5: Visualization of full scene reconstructions with optimized pose graphs. Green edges correspond to successful global registrations generated from our method. Successful registration are used for loop closures in multi-way registration. Red pyramids indicate camera poses after multi-way registration.



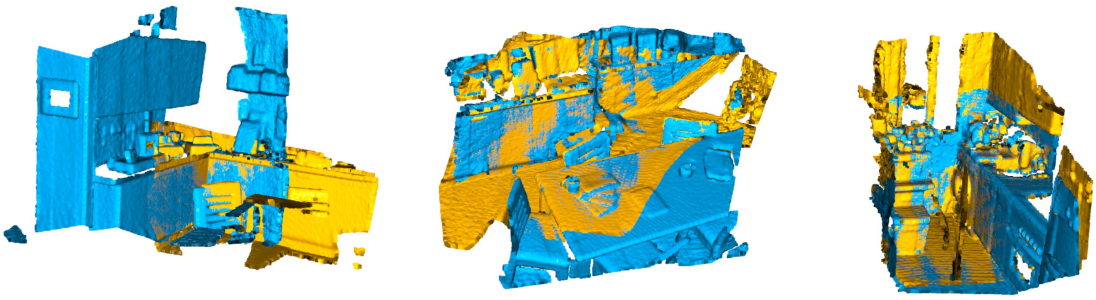
(a) RANSAC [2]



(b) FGR [5]



(c) DCP [3]



(d) Ours

Figure 6: Additional registration results on the 3DMatch benchmark [4]. Our method performs better than RANSAC [2], FGR [5], and DCP [3].

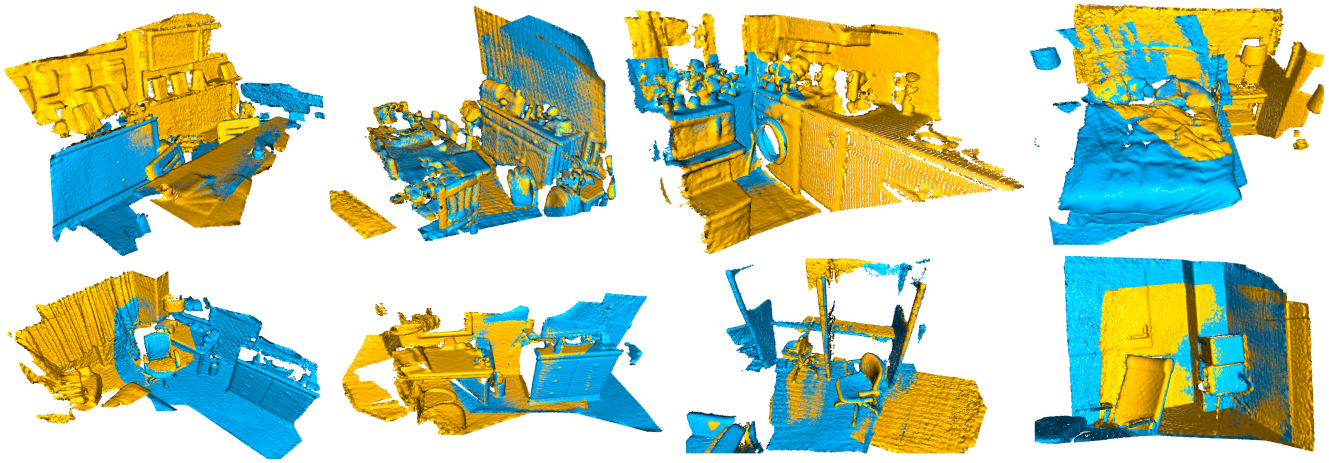


Figure 7: Visualization of registration results of our method on all 8 test scenes in the 3DMatch [4] benchmark. Best viewed in color.

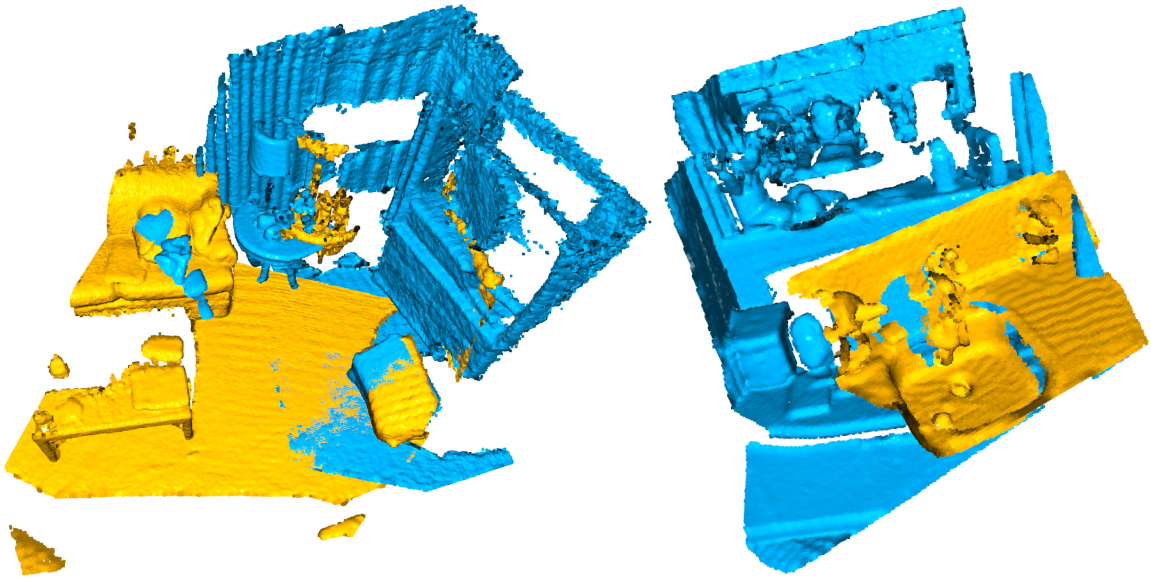


Figure 8: Failure cases of our method. Registration fails due to small overlap (left), or similar objects (right).

References

[1] Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. Colored point cloud registration revisited. In *ICCV*, 2017. 3

[2] R. B. Rusu, N. Blodow, and M. Beetz. Fast point feature histograms (FPFH) for 3D registration. In *ICRA*, 2009. 7

[3] Yue Wang and Justin M. Solomon. Deep closest point: Learning representations for point cloud registration. In *ICCV*, 2019. 7

[4] Andy Zeng, Shuran Song, Matthias Nießner, Matthew Fisher, Jianxiong Xiao, and Thomas Funkhouser. 3DMatch: Learning local geometric descriptors from RGB-D reconstructions. In *CVPR*, 2017. 4, 7, 8

[5] Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. Fast global registration. In *ECCV*, 2016. 7

[6] Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. Open3D: A modern library for 3D data processing. *arXiv*, 2018. 3