1. Supplementary

In this supplementary material, we include additional details for our Floorplan-Jigsaw paper. First, we show the detailed formulation of the boundary similarity analysis. Second, we describe the modifications to accommodate our method to RGB-D input. Furthermore, we show more quantitative results on our testing data.

1.1. Boundary similarity analysis

For a boundary point $\mathcal{P}$, given the closest point $\mathcal{P}'$ in the adjacent boundary of the next fragment, the probability of $\mathcal{P}$ and $\mathcal{P}'$ not belonging to a same object is computed by

$$P\left(\mathcal{P},\mathcal{P}'\right) = \Phi\left(\frac{\Delta d}{\sigma}\right) - \Phi\left(\frac{-\Delta d}{\sigma}\right), \quad (1)$$

where Δd is the distance between $\mathcal{P}$ and $\mathcal{P}'$. We consider a Gaussian measurement noise with standard deviation σ and use its cumulative distribution function (CDF) denoted as Φ to compute the target area as the probability. Then the boundary similarity energy $E_b(f_l, f_k)$ when fragment f_l is placed next to f_k is defined as

$$\frac{1}{M+N}\left(\sum_{\mathcal{P}_i \in B^r_{f_k}} P\left(\mathcal{P}_i,\mathcal{P}'_i\right) + \sum_{\mathcal{P}_j \in B^h_{f_l}} P\left(\mathcal{P}_j,\mathcal{P}'_j\right)\right), \quad (2)$$

where M and N are the number of the boundary points in f_k's rear (denote as $B^r_{f_k}$) and f_l's head (denote as $B^h_{f_l}$) respectively. This formulation gives a mismatch score $E_b(f_k, f_l)$ ranging from 0 to 1, while it is set to 0.5 if either of the two fragments does not have enough boundary points (M or $N < 50$). We sum up the mismatch scores of all adjacent pairs to derive the final boundary similarity E_b.

1.2. RGB-D images

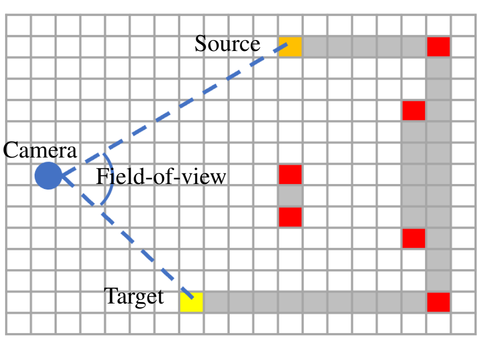

Figure 1: We project the camera frustrum onto the floorplan plane and show the relationship between the source/target points and the field-of-view.

When the input are RGB-D frames, we modify the algorithm to make full use of the properties of image input in order to improve robustness. First, the source and the target of the floorplan path can be efficiently determined by the image boundary, while we do not need to solve the *ST-graph*. Fig. 1 shows the projection of the camera frustum onto the floorplan plane, where all the projected points in the scene will fall within the 2D FoV of the camera. The boundary of the scene (i.e., layout) should intersect with the 2D camera frustum. Therefore, the source and target points p_i and p_j of a floorplan path can be inferred by

$$\arg\max_{p_i, p_j \in \overline{P}} \angle(p_i O_c p_j), \quad (3)$$

where O_c is the camera position, $\overline{P}$ the set of candidate wall keypoints.

To filter out the frames that have sufficient overlap, we detect ORB feature points for each image before running our algorithm. Any two images that have enough correct matches will be merged first and then the remaining frames can be considered as insufficiently overlapping. With these changes, the algorithm would be more robust to the input of RGB-D images.

1.3. Additional qualitative results

The input to our method can be either partial point clouds or RGB-D images. Our testing data contains 101 scenes, among which 28 scenes are captured in the real-world and 73 scenes are synthesized from the SUNCG dataset. There are 22 scenes given as point clouds and 79 scenes given in the form of RGB-D images.

We show more detailed results for both point cloud input in Fig. 2 and RGB-D input in Fig. 3. Our method is able to handle scenes with various sizes and layout complexities.

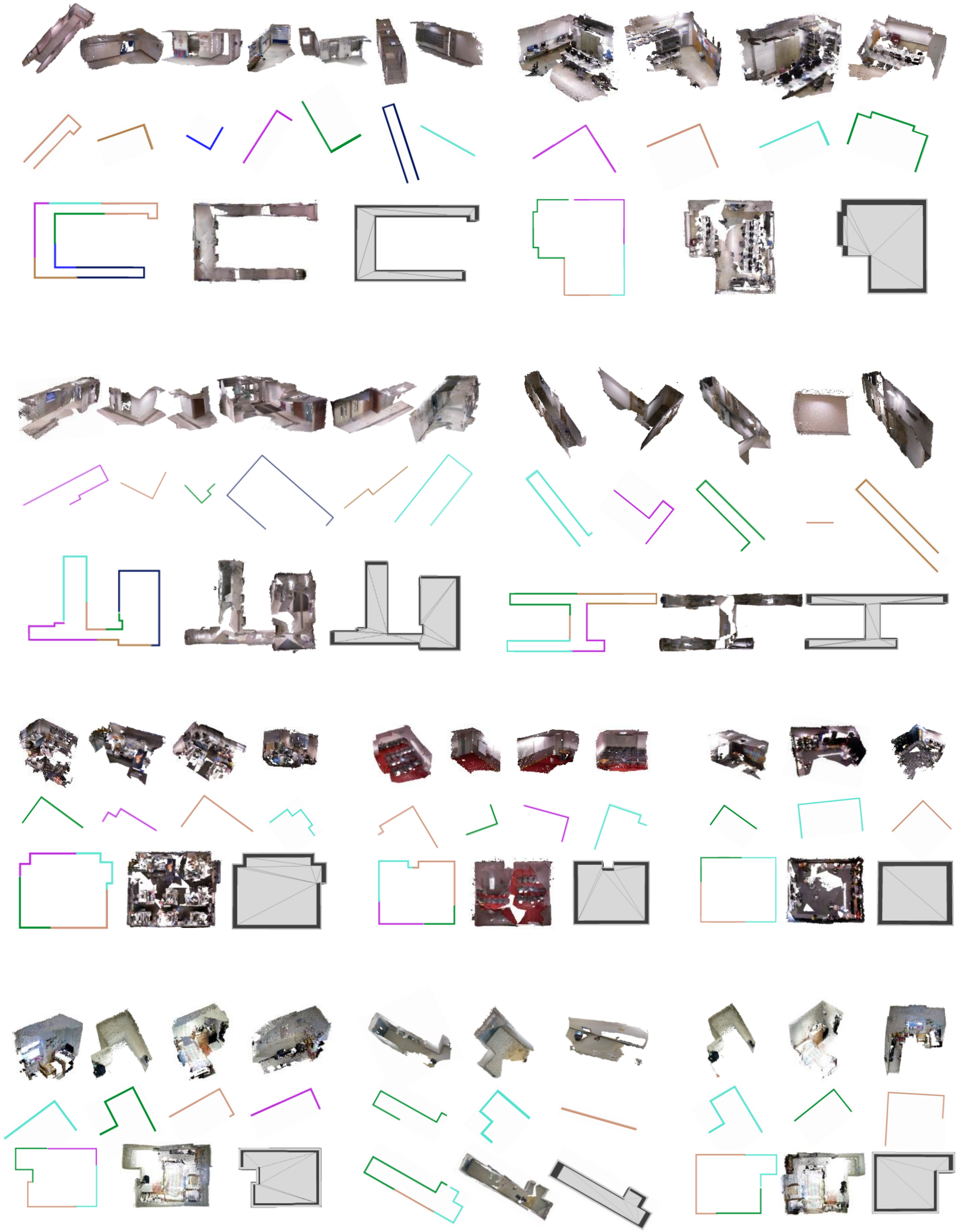

Figure 2: Additional qualitative results of partial scan alignment. We show each partial scan (first row of each sub-figure), the estimated local layout (second row of each sub-figure), the aligned global layout (first column of the last row of each sub-figure), the aligned point cloud (second column of the last row of each sub-figure) and the reconstructed layout model (third column of the last row of each sub-figure).

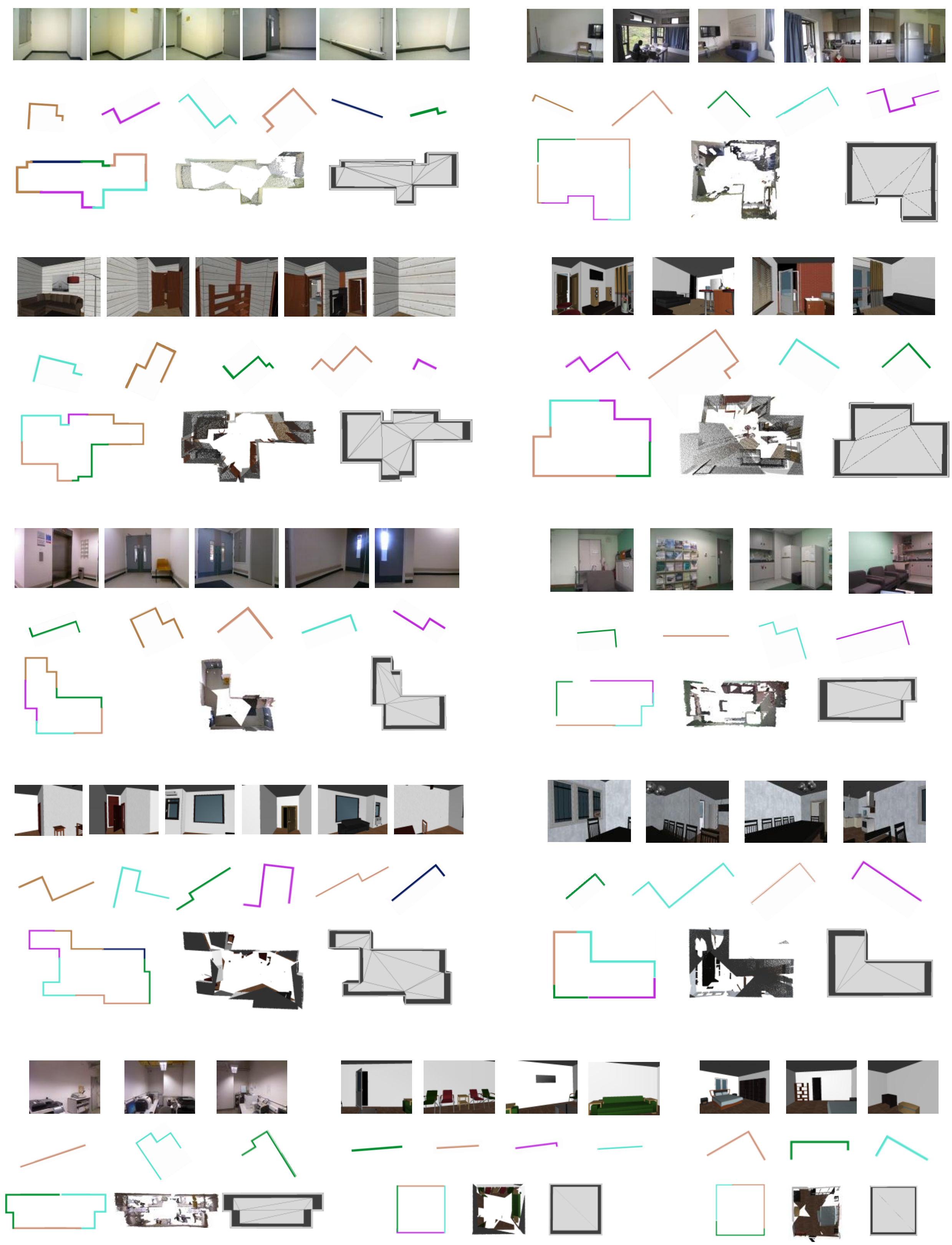

Figure 3: Additional qualitative results given RGB-D images as input. We show each RGB-D image (first row of each sub-figure), the estimated local layout (second row of each sub-figure), the aligned global layout (first column of the last row of each sub-figure), the aligned point cloud (second column of the last row of each sub-figure) and the reconstructed layout model (third column of the last row of each sub-figure).