# Learning 3D Part Assembly from a Single Image Appendix

Yichen Li*[1], Kaichun Mo*[1], Lin Shao[1], Minhyuk Sung[2], and Leonidas Guibas[1]

[1] Stanford University
[2] Adobe Research

## 1 Appendix

This document provides supplementary materials accompanying the main paper, including

- Ablation Experiments
- Discussion of failure cases and future works;
- More Architecture Details;
- More Qualitative Examples.

### A. Ablation Experiments

Table 1: Ablation Experiment Results

| Ablated Module | Part Accuracies ↑ Total | Visible | Invisible | Assembly CD ↓ |
|---|---|---|---|---|
| w/o L2 Rotation loss | 0.426 | 0.445 | 0.207 | 0.070 |
| w/o Segmentation | 0.363 | 0.378 | 0.164 | 0.084 |
| w/o Graph Conv 1, 2 | 0.403 | 0.423 | 0.178 | 0.073 |
| w/o Graph Conv 2 | 0.434 | 0.456 | 0.239 | 0.073 |
| w/o Image Feature | 0.403 | 0.419 | 0.208 | 0.077 |
| w/o Global Feature | 0.418 | 0.437 | 0.202 | 0.072 |
| Ours - Full | **0.454** | **0.470** | **0.270** | **0.067** |

### B. More Failure Cases and Discussion

**Disconnected Parts** We notice that our prediction on very fine-grained instances sometimes results in unconnected parts. The assembly setting requires the physical constraint that each part must be in contact with another part. However, the implicit soft constraint enforced using the second stage graph graph convolution is not sufficient enough for this task. Ideally, the translation and

* :indicates equal contributions.

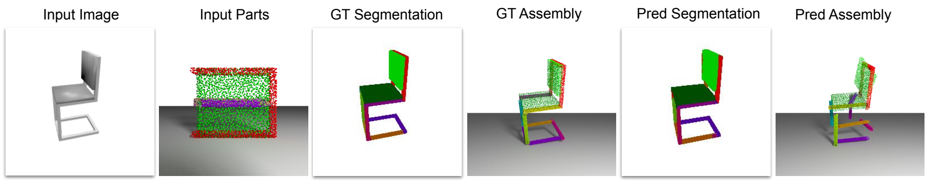


Fig. 1: **Failure Case** This figure shows that our proposed method does not well handle disconnected parts, and needs to leverage more geometric reasoning.

rotation predicted for each part is only valid if they can transform the part to be in contact at the joints between relevant parts. For example, in Figure 1 we can see that the back of the chair base bars does not connect. We plan to address this problem in future works by explicitly enforcing contact between parts in a range of contact neighborhood.

**Geometric Reasoning** Additionally, though our current proposed method makes many design choices geared for geometric reasoning between fitting of parts, however, we still see some cases that the fitting between parts is not yet perfect. For example, in Figure 1, We can see that the back pad does not fit perfectly into the back frame bar. This problem need to be addressed in future work where the method design should discover some pairwise or triplet-level geometric properties that allow fitting between parts.

## C. Architecture Details

Table 2: Part-instance Segmentation Architecture.

| layer | configuration |
|---|---|
| | UNet Encoding |
| 1 | Conv2D (3, 32, 3, 1, 1), ReLU, BN,<br>Conv2D (32, 32, 3, 1, 1), ReLU, BN, |
| 2 | Conv2D (32, 64, 3, 1, 1), ReLU, BN,<br>Conv2D (64, 64, 3, 1, 1), ReLU, BN, |
| 3 | Conv2D (64, 128, 3, 1, 1), ReLU, BN,<br>Conv2D (128, 128, 3, 1, 1), ReLU, BN, |
| 4 | Conv2D (128, 256, 3, 1, 1), ReLU, BN,<br>Conv2D (256, 256, 3, 1, 1), ReLU, BN, |
| 5 | Conv2D (256, 512, 3, 1, 1), ReLU, BN,<br>Conv2D (512, 512, 3, 1, 1), ReLU, BN, |
| | UNet Decoding |
| 1 | ConvTranspose2D(1301, 256, 2, 2) |
| 2 | ConvTranspose2D(256, 128, 2, 2) |
| 3 | ConvTranspose2D(128, 64 , 2, 2) |
| 4 | ConvTranspose2D(64, 32, 2, 2) |
| 5 | ConvTranspose2D(32, 1, 1, 1) |
| | PointNet |
| 1 | Conv1D (3, 64, 1, 1), BN, ReLU |
| 2 | Conv1D (64, 64, 1, 1), BN, ReLU |
| 3 | Conv1D (64, 64, 1, 1), BN, ReLU |
| 4 | Conv1D (64, 128, 1, 1), BN, ReLU |
| 5 | Conv1D (128, 512, 1, 1), BN, ReLU |
| | SLP1 |
| 1 | FC (512, 256), ReLU, MaxPool1D |
| | SLP2 |
| 1 | FC (256, 256), ReLU |

Table 3: Pose Prediction Architecture.

| layer | configuration |
|---|---|
| SLP 3 | |
| 1 | FC(1301, 256), ReLU |
| Pose Decoder 2 | |
| 1 | FC (1301, 256), ReLU |
| 2 | FC(256, 3) |
| 3 | FC(256, 4) |
| SLP 4 | |
| 1 | FC(1031, 256), ReLU |
| Pose Decoder 2 | |
| 1 | FC (1031, 256), ReLU |
| 2 | FC(256, 3) |
| 3 | FC(256, 4) |

## D. More Qualitative Results

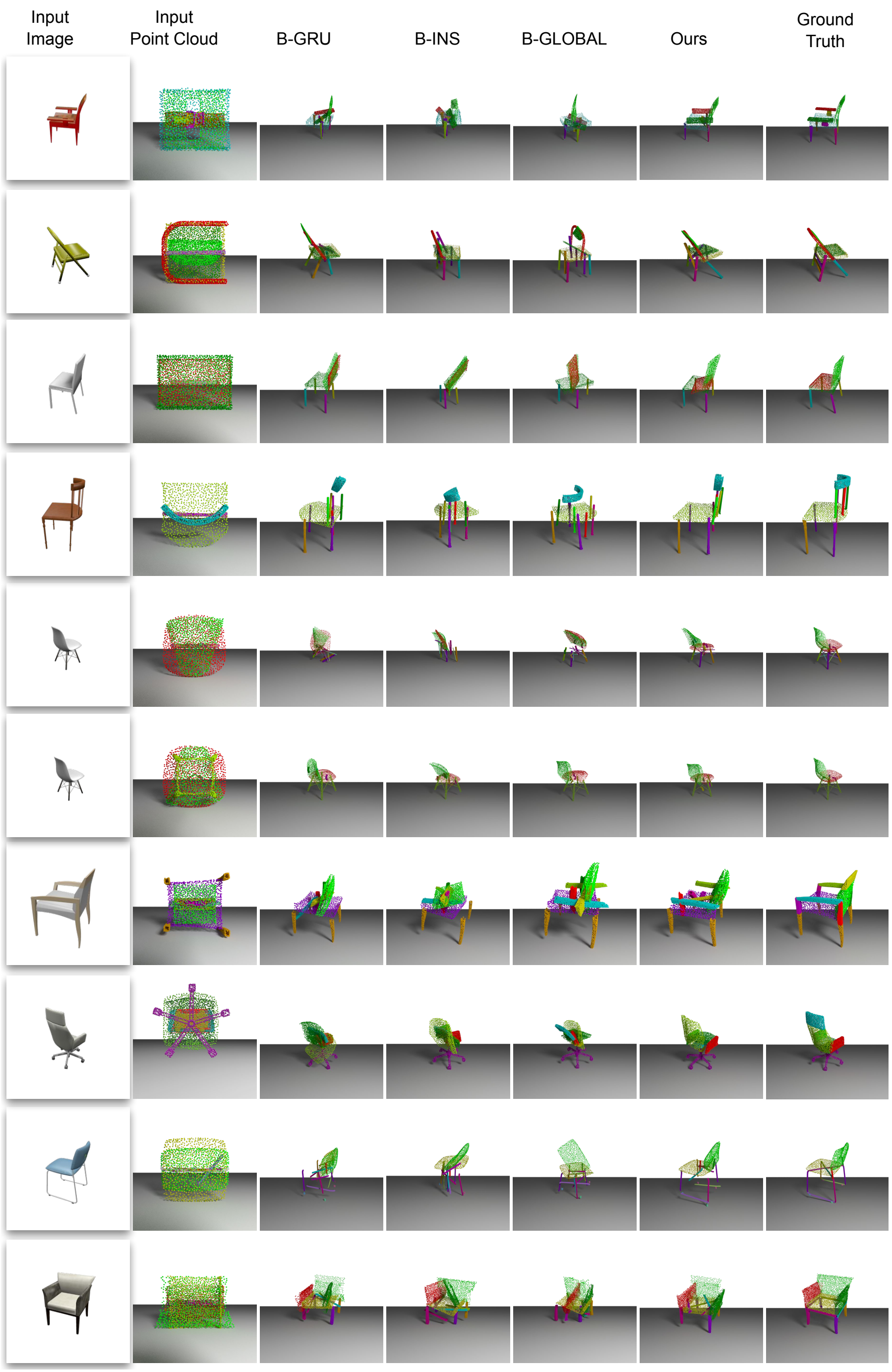


Fig. 2: **Qualitative Results for the Chair Category.** The top 5 rows show the results of Chair Level-3, and the bottom 5 rows contains the results of Chair Level-mixed.

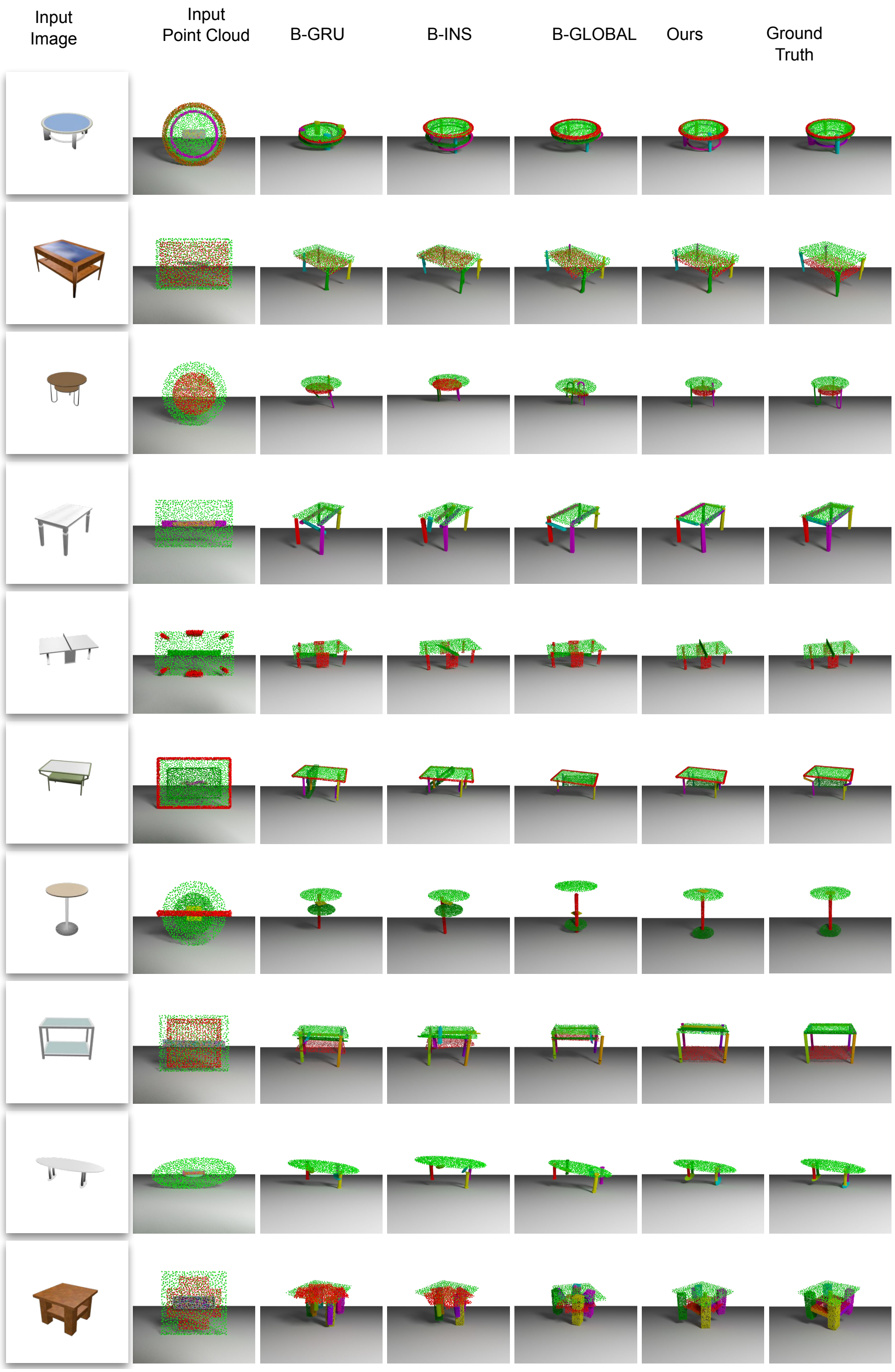


Fig. 3: **Qualitative Results for the Table Category.** The top 5 rows show the results of Table Level-3, and the bottom 5 rows contains the results of Table Level-mixed.

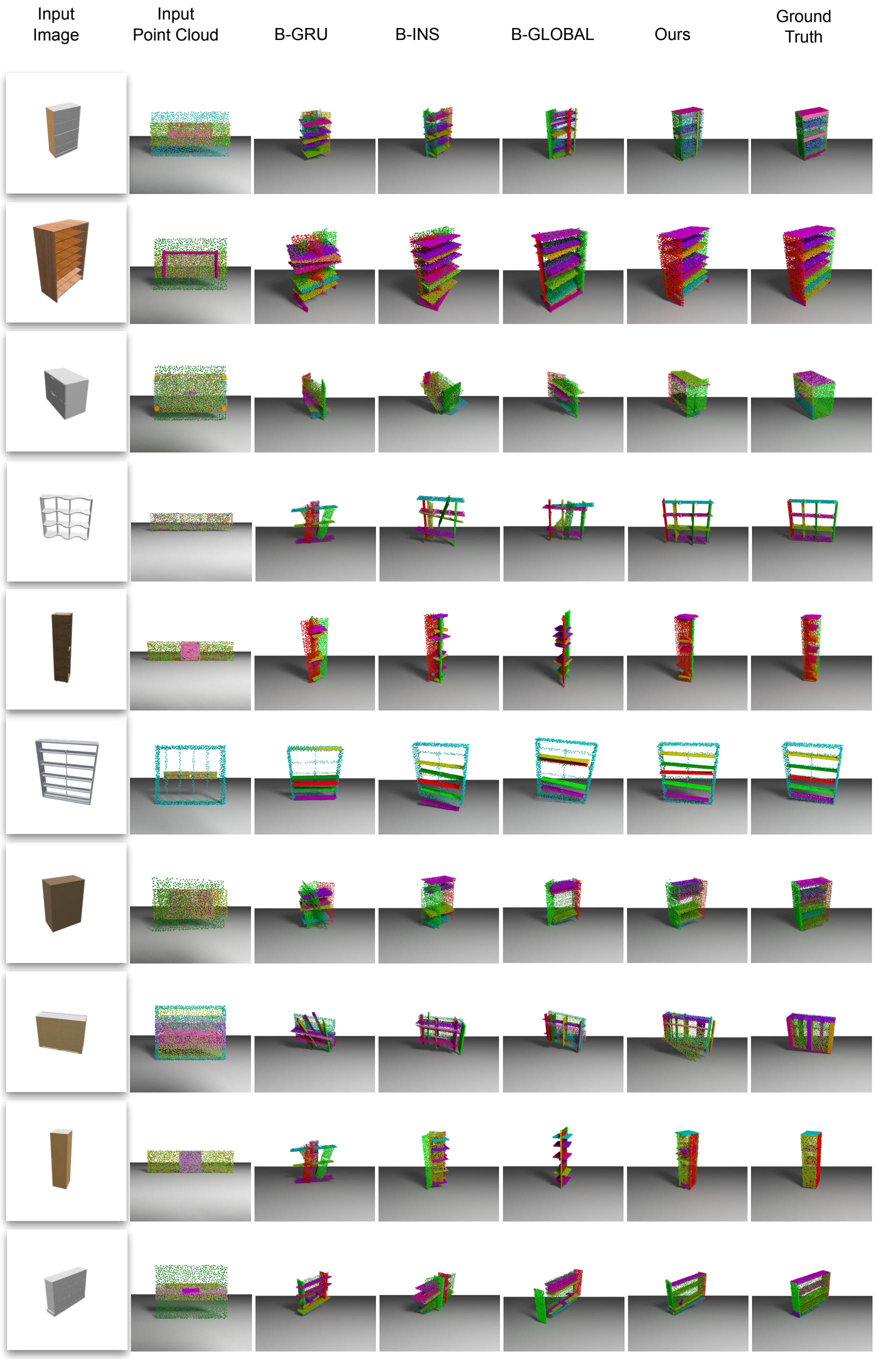


Fig. 4: **Qualitative Results for the Cabinet Category.** The top 5 rows show the results of Cabinet Level-3, and the bottom 5 rows contains the results of Cabinet Level-mixed.