# Supplementary Materials for
# PlankAssembly: Robust 3D Reconstruction from Three Orthographic Views with Learnt Shape Programs

Wentao Hu[1,2†*] Jia Zheng[3*] Zixin Zhang[4*] Xiaojun Yuan[4] Jian Yin[1,2] Zihan Zhou[3]
[1]Sun Yat-Sen University [2]Guangdong Key Laboratory of Big Data Analysis and Processing
[3]Manycore Tech Inc. [4]University of Electronic Science and Technology of China
https://manycore-research.github.io/PlankAssembly

In this document, we show some statistics of the new benchmark dataset used in the paper (Section 1), conduct an ablation study on data augmentation (Section 2), and report the detailed numbers of experiment results (Section 3).

## 1. PlankAssembly Dataset

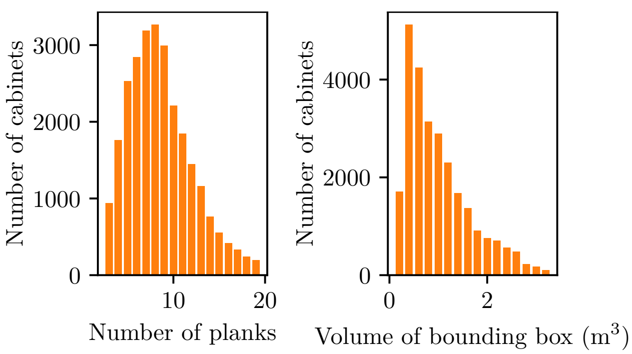


Figure 1. Statistics of the PlankAssembly dataset.

Figure 1 reports some statistics of our new benchmark dataset, including the number of planks per cabinet, and the volume of bounding boxes. Figure 3 shows more examples of the dataset.

## 2. Ablation Study on Data Augmentation

In deep learning, data augmentation is commonly used to improve the model's robustness to noises in the input. For our task, we have also experimented with data augmentation strategies by injecting noises into the inputs during training. Specifically, given the input views, we randomly select $15\%$ of the edges. These edges are either deleted or modified by perturbing the endpoints along the edge direction.

† Work done during internship at Manycore Tech Inc.
* Equal contributions.

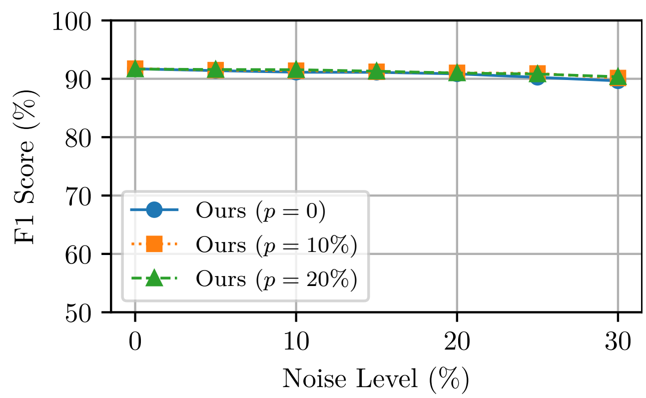


Figure 2. Ablation study on data augmentation.

In this experiment, we apply the data augmentation scheme to the training samples with probability $p$, where we set $p = \{0, 0.1, 0.2\}$. The results are shown in Figure 2. As one can see, injecting noises to the input during training appears to have only a marginal effect on the model's performance, with the model trained with $p = 0.1$ achieving slightly higher F1 scores than the other. It is worth noting that the model trained with $p = 0$ (*i.e.*, no data augmentation) performs well across a wide range of noise levels on the testing set. We hypothesize that this is partly due to the network's ability to learn flexible mappings between the input and output.

## 3. Detailed Numbers for Experiment Results

In Table 1, we report all precision, recall, and F1 score numbers for the experiment on the comparison to traditional methods (Section 5.2 and Figure 4 of the paper). We also report the number of failed cases. The first variant fails due to exceeding the time budget (5 minutes). The second variant fails if it does not produce any results.

In Table 2, we report all precision, recall, and F1 score numbers for the ablation studies (Section 5.3 and Figure 7

| Methods | Precision | Recall | F1 Score | #Failed |
|---|---|---|---|---|
| Noise level 0% | | | | |
| [1, 2], variant 1 | **95.14** | 93.56 | **94.07** | 518 |
| [1, 2], variant 2 | 83.75 | **100.00** | 90.67 | 0 |
| Ours | 92.21 | 91.51 | 91.75 | 0 |
| Noise level 5% | | | | |
| [1, 2], variant 2 | 69.80 | 25.98 | 36.05 | 113 |
| Ours | **92.13** | **91.10** | **91.49** | 0 |
| Noise level 10% | | | | |
| [1, 2], variant 2 | 59.69 | 12.91 | 20.39 | 440 |
| Ours | **92.06** | **90.99** | **91.40** | 0 |
| Noise level 15% | | | | |
| [1, 2], variant 2 | 51.56 | 8.28 | 13.86 | 777 |
| Ours | **92.03** | **90.65** | **91.21** | 0 |
| Noise level 20% | | | | |
| [1, 2], variant 2 | 45.86 | 5.99 | 10.35 | 1039 |
| Ours | **91.89** | **90.43** | **91.03** | 0 |
| Noise level 25% | | | | |
| [1, 2], variant 2 | 42.71 | 4.98 | 8.80 | 1193 |
| Ours | **92.04** | **90.11** | **90.92** | 0 |
| Noise level 30% | | | | |
| [1, 2], variant 2 | 39.82 | 4.71 | 8.20 | 1285 |
| Ours | **91.46** | **89.19** | **90.14** | 0 |

Table 1. Comparison to traditional methods on varying input noise levels (Section 5.2 and Figure 4 of the paper).

of the paper). We also report the number of failed cases. *Ours (sideface)* fails if no sidefaces are detected from input views.

## References


[1] Hiroshi Sakurai and David C. Gossard. Solid model input through orthographic views. In *ACM SIGGRAPH*, pages 243–252, 1983. 2

[2] Byeong-Seok Shin and Yeong-Gil Shin. Fast 3d solid model reconstruction from orthographic views. *Comput. Aided Des.*, 30(1):63–76, 1998. 2


| Methods | Precision | Recall | F1 Score | #Failed |
|---|---|---|---|---|
| Noise level 0% | | | | |
| Ours | 92.21 | 91.51 | 91.75 | 0 |
| Ours (image) | 83.29 | 81.66 | 82.34 | 0 |
| Ours (sideface) | **92.45** | 91.86 | **92.02** | 0 |
| PolyGen | 92.43 | **92.03** | 91.71 | 0 |
| Noise level 5% | | | | |
| Ours | **92.13** | 91.10 | **91.49** | 0 |
| Ours (image) | 82.43 | 80.19 | 81.14 | 0 |
| Ours (sideface) | 88.78 | 87.84 | 88.06 | 0 |
| PolyGen | 91.99 | **91.41** | 91.10 | 0 |
| Noise level 10% | | | | |
| Ours | **92.06** | 90.99 | **91.40** | 0 |
| Ours (image) | 81.42 | 78.52 | 79.76 | 0 |
| Ours (sideface) | 85.19 | 83.05 | 83.75 | 0 |
| PolyGen | 91.86 | **91.38** | 91.08 | 0 |
| Noise level 15% | | | | |
| Ours | **92.03** | 90.65 | **91.21** | 0 |
| Ours (image) | 80.78 | 77.43 | 78.87 | 0 |
| Ours (sideface) | 81.38 | 88.02 | 78.63 | 5 |
| PolyGen | 91.70 | **90.80** | 90.68 | 0 |
| Noise level 20% | | | | |
| Ours | **91.89** | **90.43** | **91.03** | 0 |
| Ours (image) | 79.64 | 75.44 | 77.27 | 0 |
| Ours (sideface) | 74.73 | 68.50 | 70.85 | 15 |
| PolyGen | 91.34 | 90.28 | 90.22 | 0 |
| Noise level 25% | | | | |
| Ours | **92.04** | **90.11** | **90.92** | 0 |
| Ours (image) | 77.99 | 72.64 | 74.96 | 0 |
| Ours (sideface) | 67.08 | 58.85 | 61.89 | 37 |
| PolyGen | 90.88 | 89.27 | 89.42 | 0 |
| Noise level 30% | | | | |
| Ours | **91.46** | **89.19** | **90.14** | 0 |
| Ours (image) | 77.25 | 70.67 | 73.53 | 0 |
| Ours (sideface) | 59.87 | 49.28 | 53.21 | 93 |
| PolyGen | 90.44 | 87.97 | 88.46 | 0 |

Table 2. Ablation studies on the input sequence and the output sequence (Section 5.3 and Figure 7 of the paper).

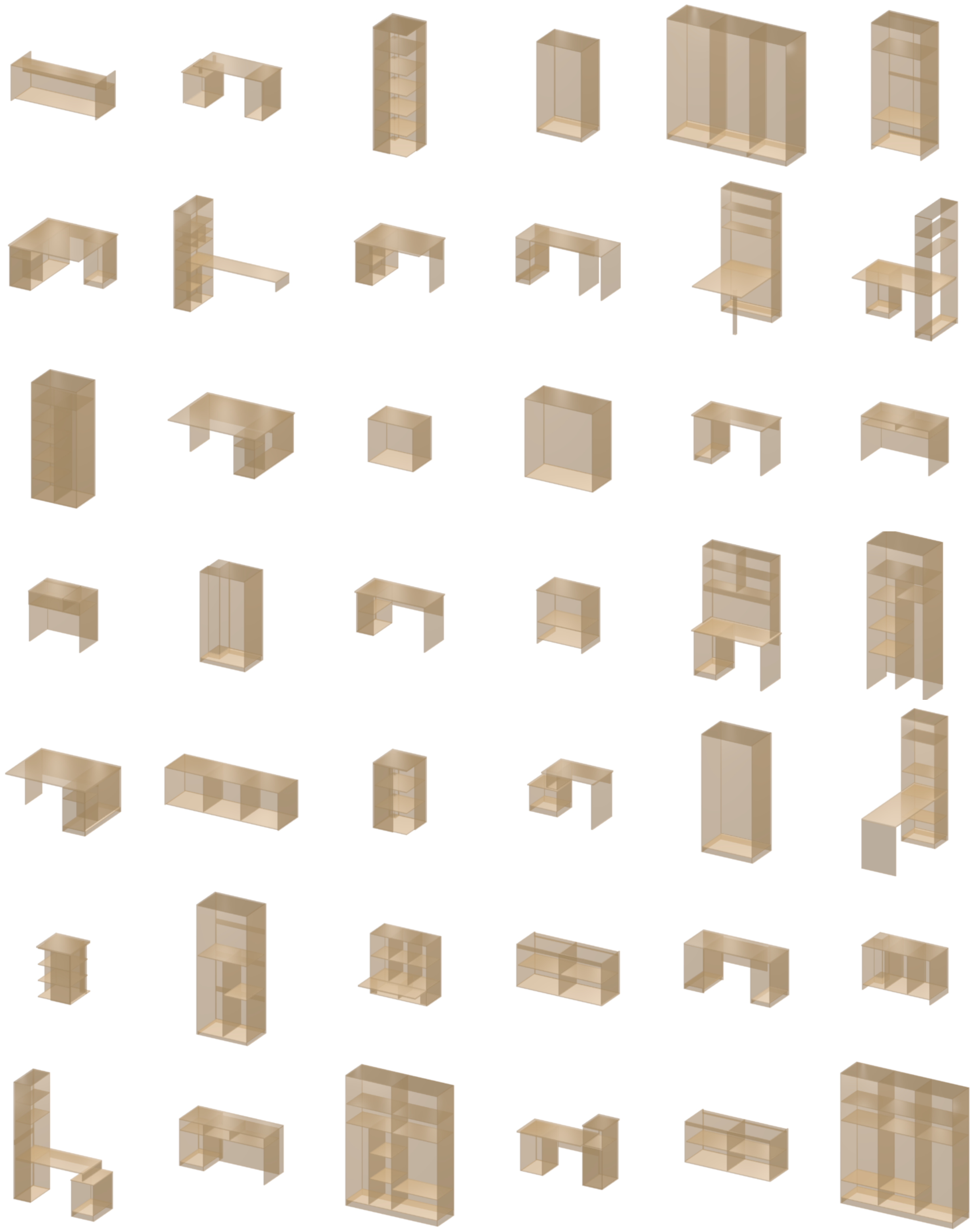

Figure 3. Example cabinet models in our PlankAssembly dataset.