SPAFormer: Sequential 3D Part Assembly with Transformers

Supplementary Material

1. Additional Quantitative Results

1.1. General-purpose Part Assembly is Beneficial

We have trained a versatile model ("ours-general") on all available object categories, which aims to perform part assembly across a variety of categories, and category-specific models ("ours-specific"), where each model is tailored to a specific category.

Influence on Assembly Sequence Length. We analyze the impact of multi-task learning on varied assembly lengths in Tab. 1. Compared to the category-specific models, the multi-task model continues to gain a notable improvement of $+2.77\%$ in long-horizon assembly, and also improve by $+1.9\%$ in short-horizon assembly.

Table 1. Average results across all categories. The numbers in brackets denote the object number in the test set.

	≤10 parts (3318)		>10 parts (1223)	
Method	PA	SR	PA	SR
ours-specific	61.74	36.20	53.25	7.59
ours-multitask	63.64	37.46	56.02	9.31

Detailed Results on Daily Object Assembly. In Tab. 2, we conduct comparisons between the versatile model and category-specific models. Our versatile model demonstrates significant improvements in categories with both abundant (e.g., tables, storage furniture) and limited (e.g., mugs with 120 samples, bowls with 130 samples) training data. Given these promising results, we will further explore this direction, such as discovering which categories benefit most from shared assembly knowledge, or how to effectively manage the output categories.

1.2. More Details and Analysis on Category-specialized Part Assembly

Unlike the less-explored general-purpose assembly mentioned above, the experiments presented in our main paper follow the traditional benchmark of category-specialized part assembly, which has been commonly adopted by previous works. In this subsection, we provide additional experimental details.

Additional Implementation Details. In our approach, each part is represented by a point cloud, consisting of 1,000 points obtained through Furthest Point Sampling [1]. We ensure these sampled point clouds are zero-centered and aligned with the principal axes determined by Principal Component Analysis (PCA) [2]. In our experimental results, the shape chamfer distance (SCD) metric is scaled up by a factor of 1000, while the part accuracy (PA), connectivity accuracy (CA), and success rate (SR) are expressed as percentages. Moreover, the assembly sequences are constructed based on the relative positions of parts in their original configuration.

Additional evaluation on Per-class Part Assembly. To offer a comprehensive comparison of different object categories, we have included per-class evaluation results in Tab. 2. Our proposed method consistently outperforms previous works in most categories. It's important to note that the success rates vary among categories, We attribute this variation more to the diversity of object structures and fundamental geometries encountered during training, rather than the quantity of training samples. For example, categories with a large number of training samples, such as faucet (with approximately 460 samples) and lamp (with approximately 1420 samples), do not necessarily yield higher success rates. Specifically, faucets and lamps demonstrate success rates of only 5.3% and 17.2%, respectively, when using our category-specific model. This is largely due to the complex assembly required by the intricate geometrical composition of their parts.

2. Additional Qualitative Examples

Additional visualizations, including both successful and unsuccessful cases, are presented in Figs. 1 to 3. Our approach not only excels in effectively assembling structural tables that have small connection areas but also shows significant improvement in assembling larger parts of storage furniture, which typically have more extensive connection areas.

3. Limitations and Directions

First, acquiring the assembly sequence is sometimes unfeasible, especially when predicted by assembly sequence planning algorithms. On the other hand, while assembly sequences can be derived from various sources such as instructional manuals, the goal for a truly autonomous agent is to independently determine assembly sequences and navigate the assembly process on its own, mirroring human capabilities. Therefore, combining the task of assembly sequence planning and 3D part assembly would significantly benefit the object assembly problem.

Second, current methods for 3D-PA primarily focus on point cloud processing without considering motion planning in the real world, which limits the application in robotics. Working towards this path and deploying the models in the real world can significantly expand the scope of this problem.

Table 2. Comparisons of per-class results across 18 daily objects, where "ours-multitask" and "ours-specific" denote the unified model for multi-task assembly and the class-specific models, respectively. The abbreviations "Dish", "Disp", "Ear", "Fauc" and "Frid" denote Dishwasher, Display, Earphone, Faucet and Refrigerator, respectively.

	Method	Avg	Bag	Bed	Bottle	Bowl	Clock	Dish	Disp	Door	Ear	Fauc	Hat	Knife	Lamp	Lap	Mug	Frid	Trash	Vase
SCD	DGL	6.57	23.7	20.2	6.0	20.2	**3.1**	7.0	**1.4**	4.5	16.6	13.8	3.6	**2.0**	**8.0**	**0.9**	5.2	5.8	9.1	2.4
	RGL	10.93	11.8	9.6	3.5	22.4	9.1	4.5	3.6	3.7	27.2	17.2	9.6	3.4	15.8	17.1	7.8	4.2	4.7	8.7
	Score	**6.38**	**1.3**	19.8	13.6	**0.6**	7.1	7.8	2.9	9.7	**13.6**	**9.0**	**1.6**	10.4	12.7	12.7	**1.4**	**2.5**	3.6	**1.2**
	ours-specific	8.61	7.6	**7.7**	**3.0**	23.3	7.7	4.3	3.3	**3.4**	24.5	11.2	9.2	3.3	11.7	5.1	8.5	2.8	4.7	8.9
	ours-multitask	8.60	8.5	12.0	3.6	10.7	8.8	**3.5**	4.4	6.5	24.4	12.4	10.8	3.3	11.6	4.0	7.0	2.9	**2.7**	8.7
PA	DGL	36.55	21.7	11.8	80.4	63.0	45.2	11.3	62.9	26.7	27.8	13.1	36.7	41.0	32.8	65.2	13.1	16.5	11.2	23.3
	RGL	51.33	36.2	13.3	86.1	46.3	52.0	45.2	82.6	**56.7**	41.0	33.7	55.1	82.0	33.3	60.0	**59.5**	48.5	49.4	47.3
	Score	32.39	**62.8**	1.5	2.2	72.2	11.6	1.8	10.2	10.2	14.0	23.5	58.2	59.8	34.9	62.8	54.8	32.5	20.6	51.4
	ours-specific	54.32	52.2	**28.0**	**87.8**	42.6	**58.4**	52.5	**83.6**	47.1	**52.0**	39.7	**62.2**	76.2	34.2	87.2	37.0	52.1	59.4	48.5
	ours-multitask	**59.93**	46.4	11.0	84.8	**77.8**	56.0	**62.0**	82.2	42.0	50.5	**42.8**	55.1	**84.0**	**36.9**	**90.2**	55.9	**55.7**	**65.6**	**52.7**
CA	DGL	40.01	29.8	21.8	76.5	64.3	**41.1**	14.7	66.3	58.1	38.0	29.7	36.5	42.5	38.7	31.3	9.4	28.7	10.8	27.4
	RGL	50.52	23.4	12.6	76.5	**71.4**	36.9	27.5	76.0	**83.8**	42.2	47.5	50.0	74.0	**47.4**	71.3	**11.3**	37.3	20.7	32.21
	Score	32.16	16.2	19.4	24.8	50	26.2	10.9	19.8	19.6	24.1	37.8	**55.8**	51.4	41.4	16.2	**11.3**	32.5	9.8	**38.5**
	ours-specific	51.60	**42.6**	**23.2**	**77.1**	57.1	37.6	29.4	**80.2**	70.4	**55.1**	44.2	51.9	**77.3**	46.9	**95**	1.9	50.4	27.6	26.4
	ours-multitask	**52.86**	**42.6**	15.0	71.2	42.8	**41.1**	**56.8**	77.8	63.7	**55.1**	**53.9**	44.2	76.2	47.2	91.2	9.4	**55.6**	**33.0**	31.2
SR	DGL	18.73	13.8	0	69.0	53.8	23.9	0	37.7	3.9	0	0	13.3	7.8	14.5	47.6	0	0	0	11.6
	RGL	32.0	17.2	0	75.0	25.6	47.7	13.7	67.5	**47.1**	1.9	1.5	24.4	59.7	14.0	41.5	**31.4**	12.9	26.3	40.0
	Score	15.05	**34.5**	0	0	66.7	3.4	0	8.4	11.8	0	1.5	26.7	19.5	10.1	12.2	25.7	3.2	2.6	43.6
	ours-specific	37.12	31.0	**6.3**	**78.6**	25.6	**53.4**	15.7	**69.7**	45.1	**17.0**	5.3	**37.8**	50.6	**17.2**	78.0	5.7	16.2	**42.1**	42.2
	ours-multitask	**39.06**	24.1	0	75.0	**71.8**	46.6	**25.5**	68.0	33.3	**17.0**	**10.6**	26.7	**62.3**	15.7	**82.9**	**31.4**	**19.4**	39.5	**47.8**
Test Number			29	16	84	39	88	51	191	51	53	132	45	77	407	82	35	31	38	232

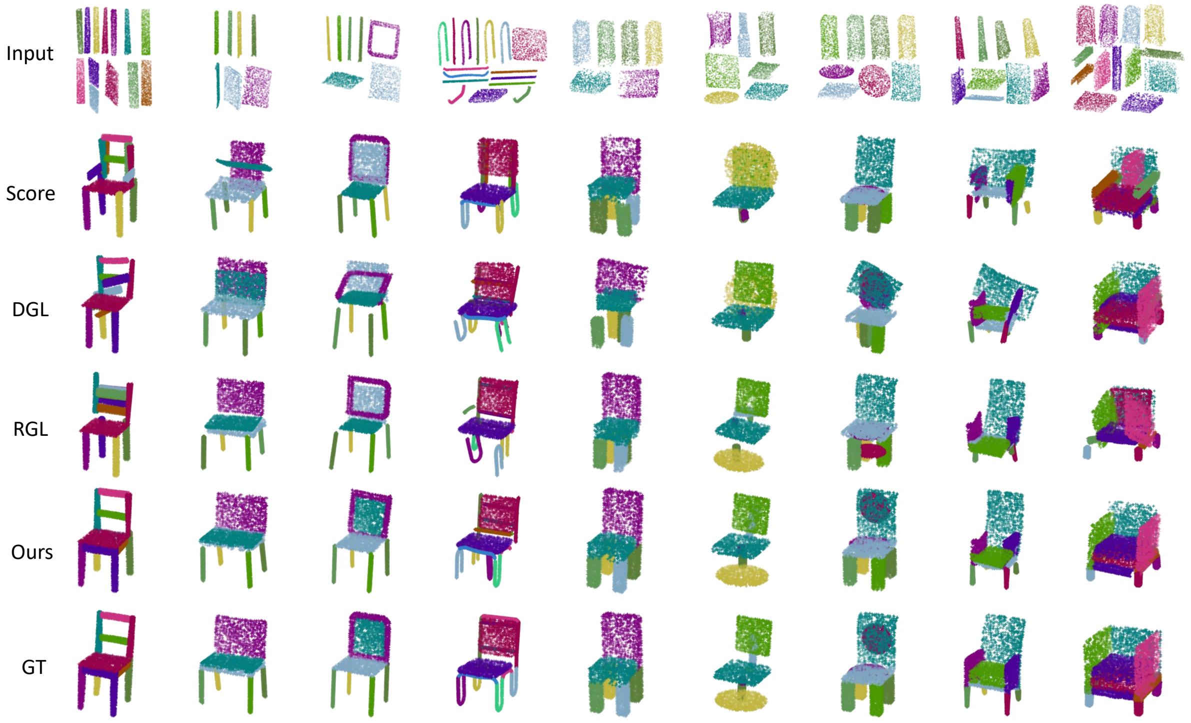

Figure 1. Qualitative results and comparisons on the chair assembly task. Distinct colors within a single shape denote various parts of the chair, whereas consistent coloring in a row signifies identical parts. Our SPAFormer is able to identify and adhere to appropriate assembly patterns to ensure accurate assembly of structured objects.

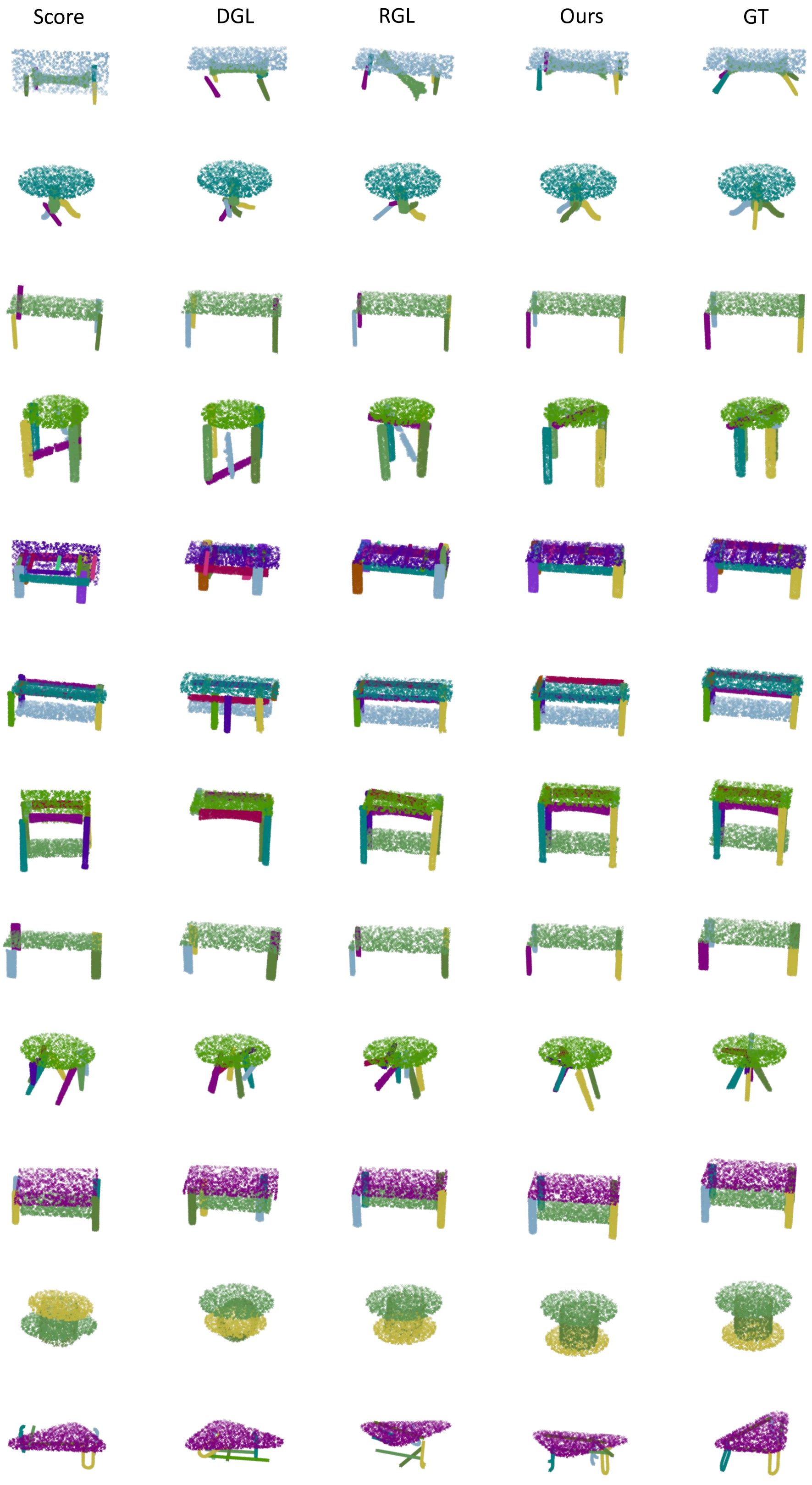

Figure 2. Qualitative results and comparisons on tables.

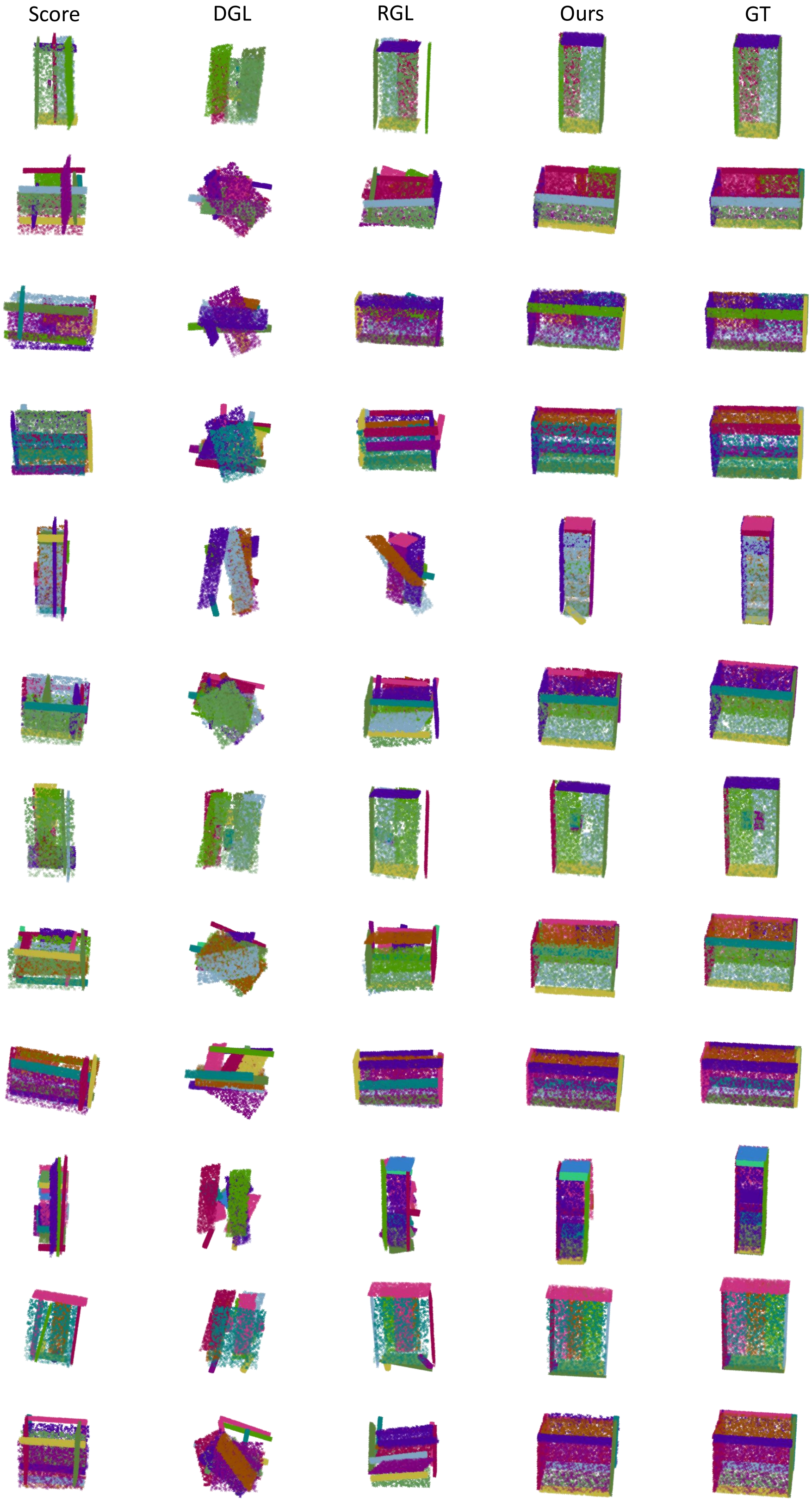

Figure 3. Qualitative results and comparisons on storage furniture.

References

[1] Eldar, Y., Lindenbaum, M., Porat, M., Zeevi, Y.Y.: The farthest point strategy for progressive image sampling. IEEE Transactions on Image Processing **6**(9), 1305–1315 (1997) 1

[2] Jolliffe, I.T.: Principal component analysis. Journal of Marketing Research **87**(4), 513 (2002) 1