NEURAL INFORMATION PROCESSING SYSTEMS

Coarse-to-Fine 3D Part Assembly via Semantic Super-Parts and Symmetry-Aware Pose Estimation

Xinyi Zhang*[1], Bingyang Wei*[1], Ruixuan Yu[1†], Jian Sun[2,3]

[1]Shandong University [2]Xi'an Jiaotong University [3]Pazhou Laboratory (Huangpu)

Contents

Introduction: 3D Part Assembly

3D Part Assembly

- Reconstruct a coherent 3D shape by predicting the 6-DoF poses for a set of individual parts.

Challenges:

1. **Over-reliance on Geometric Relationships**

- Limited ability to capture high-level, semantic object structure.

2. **Difficulties in Handling Symmetries**

- Real-world objects are full of symmetries (e.g., identical chair legs, symmetric chair seats).
- Leads to multiple valid assembly configurations, but most methods are designed to find only a single solution, overlooking the interchangeability of parts.

Our Contributions:

We propose **CFPA** (**Coarse-to-Fine Part Assembly**), a two-stage framework that unifies semantic abstraction, hierarchical reasoning, and symmetry awareness.

- **Semantic Super-Parts** via **Optimal Transport**: Captures **high-level semantic** structure and supports more coherent and semantically aware assembly.
- **Coarse-to-Fine Pose Estimation:** A two-stage framework incorporates with a **dual-range feature propagation** mechanism for coarse stage estimation, followed by a refinement stage with **cross-stage attention**.
- **Symmetry-Aware Loss:** A novel objective that supervises multiple consistent pose configurations by explicitly modeling both **intra-part** and **inter-part** symmetries.

Coarse Stage Estimation: Semantic Super-Parts Construction via Optimal Transport

- Construct a set of high-level semantic super-parts $\{h_j\}_{j=1}^{M}$ from basic part features $\{f_i\}_{i=1}^{N}$ ($M \leq N$).
- Each super-part is computed as a weighted aggregation of part features according to the transport matrix T:

$$h_j = \sum_{i=1}^{N} T_{ij} f_i, \quad j = 1, \ldots, M$$

- **OT Objective:**

$$T^* = \arg\min_{T} \sum_{i=1}^{N}\sum_{j=1}^{M} T_{ij}C_{ij} - \epsilon \sum_{i=1}^{N}\sum_{j=1}^{M} T_{ij}\log T_{ij}$$

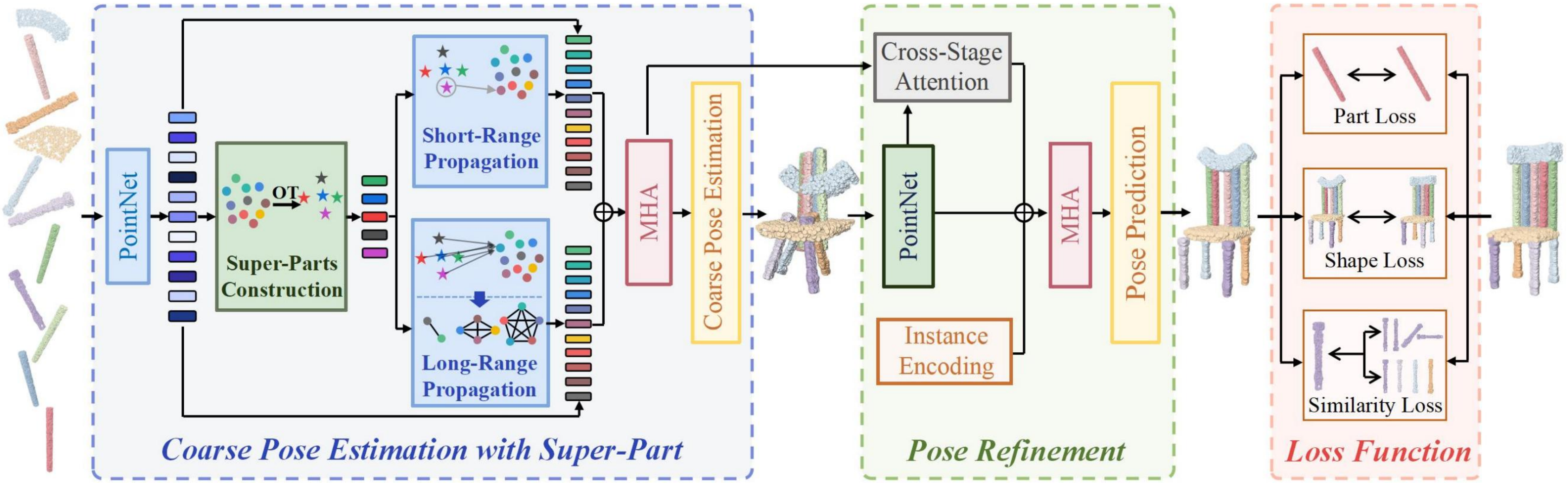


Figure1: The CFPA framework consists of two main stages: coarse pose estimation and pose refinement, supervised by a symmetry-aware loss.

Method: CFPA Pipeline

● Coarse Stage Estimation: Dual-Range Feature Propagation

➢ **Short-Range Feature Propagation**

✓ Propagate features from the nearest (Euclidean distance in the feature space) super-part.

➢ **Long-Range Feature Propagation**

✓ Integrates semantic information from all super-parts and reinforces spatial coherence through geometry-aware message passing.

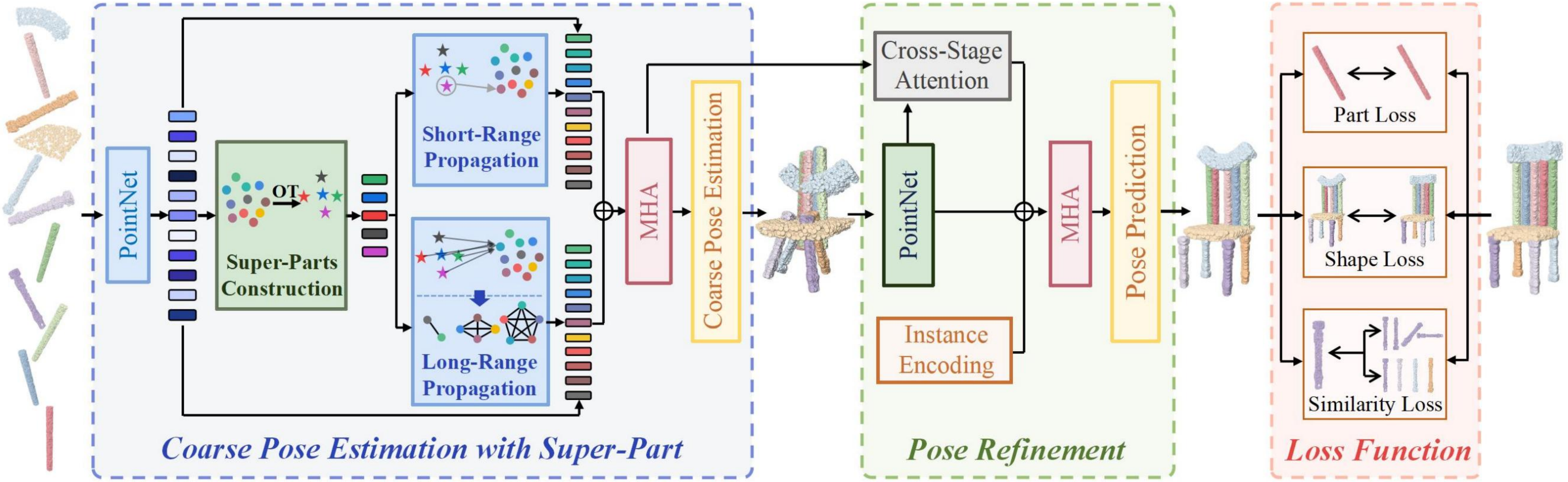


Figure1: The CFPA framework consists of two main stages: coarse pose estimation and pose refinement, supervised by a symmetry-aware loss.

● Pose Refinement

➢ **Cross-Stage Attention:**

✓ Uses coarse-stage features as guidance (Key/Value) for refining fine-stage features (Query).

➢ **Instance Encoding:**

✓ Disambiguates geometrically similar parts.

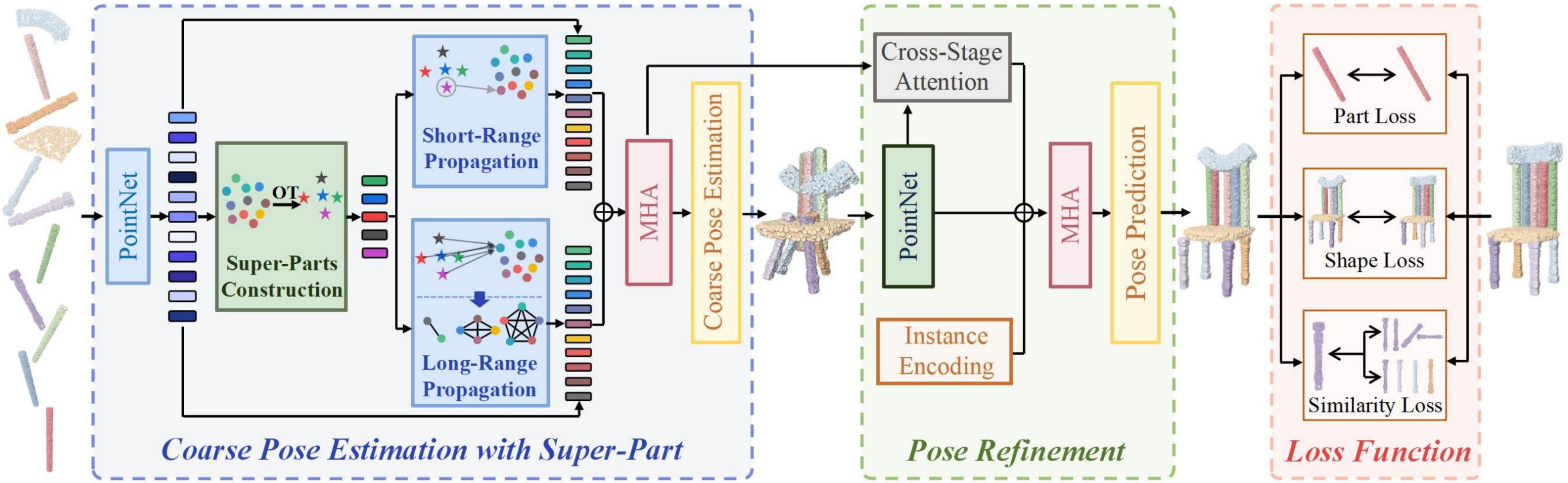


Figure1: The CFPA framework consists of two main stages: coarse pose estimation and pose refinement, supervised by a symmetry-aware loss.

Loss Function

- **Part Loss** $\mathcal{L}_{\text{Part}} = \sum_{i=1}^{N} L_{\text{pose}} \left(\mathcal{F}(P_i), \{r_i^{\text{GT}}, t_i^{\text{GT}}\}\right)$
- **Shape Loss** $\mathcal{L}_{\text{Shape}} = d_c\left(\mathcal{S}^*, \mathcal{S}^{\text{GT}}\right)$
- **Symmetry-Aware Loss**
- **Total Loss:** $\mathcal{L} = \mathcal{L}_{\text{Part}} + \mathcal{L}_{\text{Shape}} + \lambda\mathcal{L}_{\text{Sym}}$

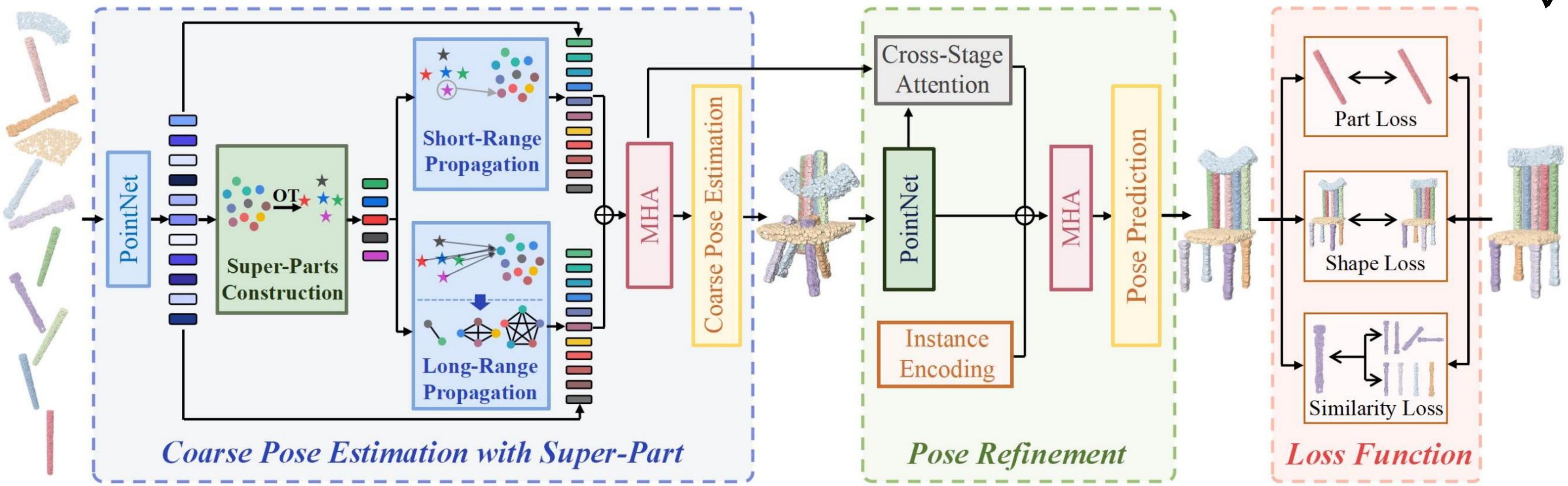
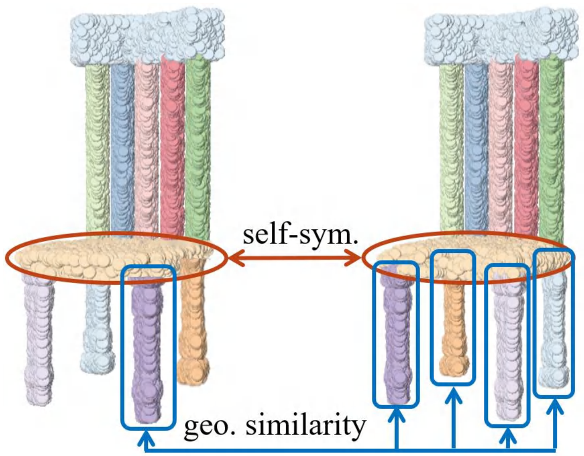


Figure1: The CFPA framework consists of two main stages: coarse pose estimation and pose refinement, supervised by a symmetry-aware loss.

Symmetry-Aware Loss

- **Intra-part self symmetry:**
  - ✓ e.g., a chair seat symmetric under vertical flipping
- **Inter-part geometric similarity:**
  - ✓ e.g., chair legs



- **Dataset:**

  PartNet, focusing on the three largest categories: Chair, Table, and Lamp.

- **Evaluation Metrics:**

  - **Accuracy:** Shape Chamfer Distance (SCD), Part Accuracy (PA), Connectivity Accuracy (CA).
  - **Diversity:** Quality-Diversity Score (QDS), Weighted QDS (WQDS).

- **Baselines:**

  We compare against state-of-the-art methods including B-Global, B-LSTM, DGL, Score-PA, IET, SPAFormer, RGL and 3DHPA.

Main Quantitative Results

● Accuracy:

Table 1: Comparison for assembly accuracy evaluated by SCD, PA, and CA.

Methods	SCD(10^{-2}) ↓			PA(%) ↑			CA(%) ↑		
	Chair	Table	Lamp	Chair	Table	Lamp	Chair	Table	Lamp
B-Global [18, 54]	1.46	1.12	0.79	15.70	15.37	22.61	9.90	33.84	18.60
B-LSTM [55]	2.35	1.71	0.90	8.08	10.55	24.68	10.05	18.28	30.23
DGL [6]	0.91	0.50	0.93	39.00	49.51	33.33	23.87	39.96	41.70
Score-PA [15]	0.71	0.42	1.11	44.51	52.78	34.32	30.32	40.59	49.07
IET [8]	1.34	0.66	0.89	37.60	48.86	32.86	25.44	40.35	52.75
SPAFormer [38]	0.67	0.38	-	55.88	64.38	-	36.39	57.60	-
RGL [7]	0.98	0.40	1.05	48.85	55.13	35.54	30.68	41.41	50.09
3DHPA [17]	0.51	**0.32**	0.82	63.01	64.58	33.49	48.28	58.00	62.01
Ours	**0.49**	0.33	**0.77**	**69.24**	**68.48**	**36.35**	**49.20**	**58.51**	**63.32**

➢ CFPA achieves the best results in **8** out of **9** cases and ranks second in the remaining one, demonstrating its effectiveness in both part assembly accuracy and structural consistency.

Main Quantitative Results

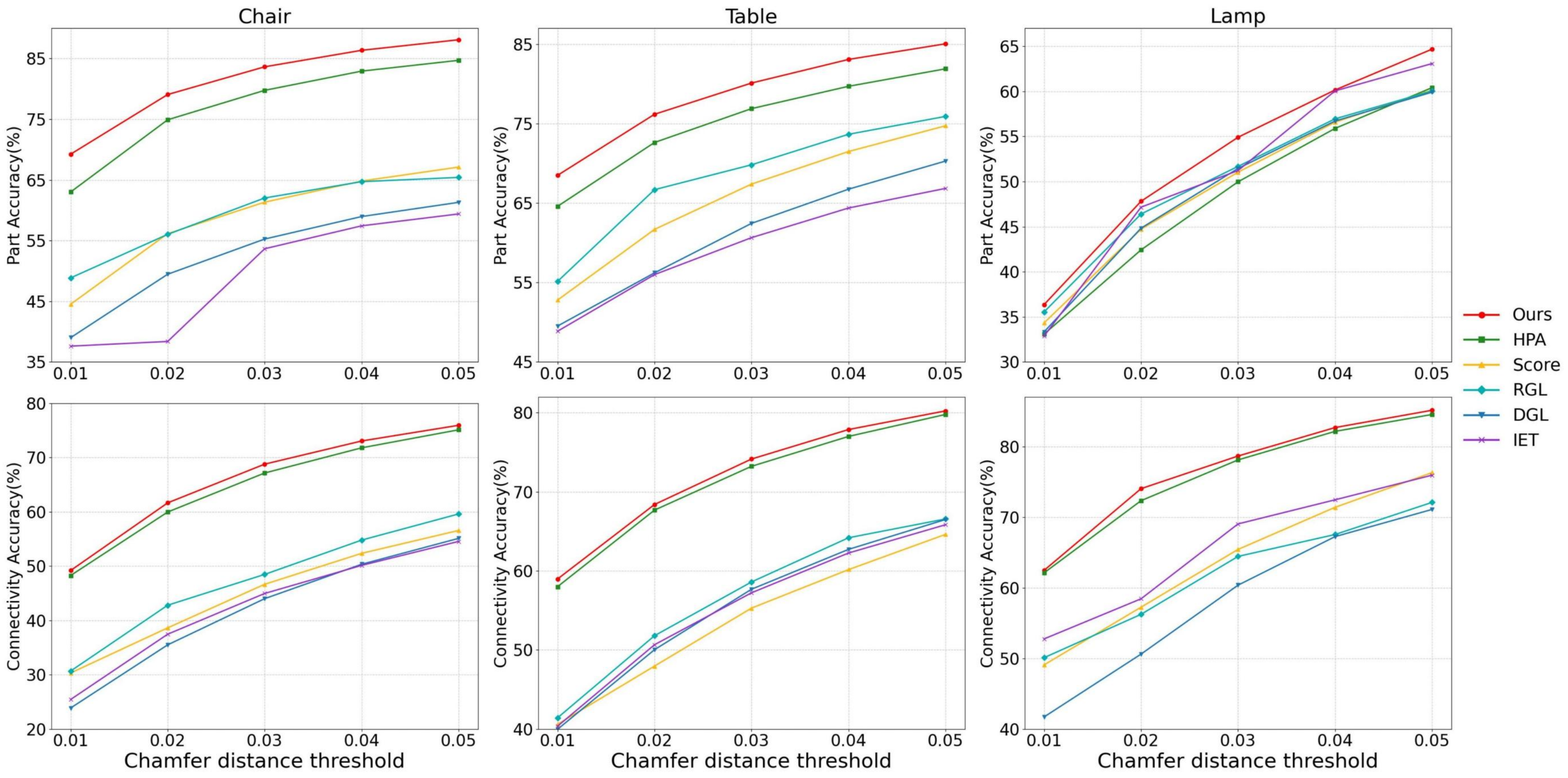
Diversity:

Table 3: Comparison of assembled shape diversity evaluated using QDS and WQDS.

Methods	QDS(10^{-5}) ↑			WQDS(10^{-5}) ↑		
	Chair	Table	Lamp	Chair	Table	Lamp
B-Global [18, 54]	0.15	0.20	0.76	1.25	1.40	0.58
B-LSTM [55]	3.92	1.33	3.05	1.26	0.55	2.01
DGL [6]	1.69	3.05	1.84	1.35	2.97	1.73
Score-PA [15]	3.36	**9.17**	**6.83**	1.70	3.81	2.82
IET [8]	3.33	6.22	4.93	1.85	2.35	3.43
RGL [7]	5.85	7.55	6.37	2.09	3.51	3.15
3DHPA [17]	4.42	7.15	4.67	1.90	3.80	3.16
Ours	**6.71**	7.28	5.65	**2.75**	**3.92**	**3.74**

- ➢ CFPA achieves the highest WQDS across all categories and the highest QDS on Chair, indicating its ability to generate shape diverse yet structurally valid assemblies.

Comparison: Performance Curves



➢ Performance curves of our CFPA and compared models on the Chair, Table and Lamp categories under Chamfer distance threshold ranging from 0.01 to 0.05. Best viewed in color.

Effectiveness of Super-Part.

1) removes super-part guidance;

2) builds super-parts based on geometric similarity;

3) uses K-means clustering to build semantic super-parts based on basic part features.

Table 4: Ablation study on super-parts.

Methods	SCD(10^{-2})↓	PA(%)↑	CA(%)↑
1) CFPA-w/o-SP	0.54	66.75	47.75
2) CFPA-GE-SP	0.53	67.60	47.97
3) CFPA-KM-SP	0.54	69.20	47.48
CFPA	**0.49**	**69.24**	**49.20**

Effectiveness of Dual-Range Feature Propagation.

4) removes short-range propagation;

5) removes long-range propagation;

6) removes message passing

7) removescross-stage attention;

8) removes instance encoding.

Table 5: Ablation study on designs in the coarse pose estimation stage and pose refinement stage.

Methods	SCD(10^{-2})↓	PA(%)↑	CA(%)↑
4) CFPA-w/o-SRFP	0.57	66.77	44.71
5) CFPA-w/o-LRFP	0.51	68.93	47.56
6) CFPA-w/o-MP	0.60	63.60	40.28
7) CFPA-w/o-CA	0.51	67.47	44.42
8) CFPA-w/o-IE	0.55	64.81	44.44
CFPA	**0.49**	**69.24**	**49.20**

Effectiveness of Symmetry-Aware Loss.

9) removes self-symmetry supervision;

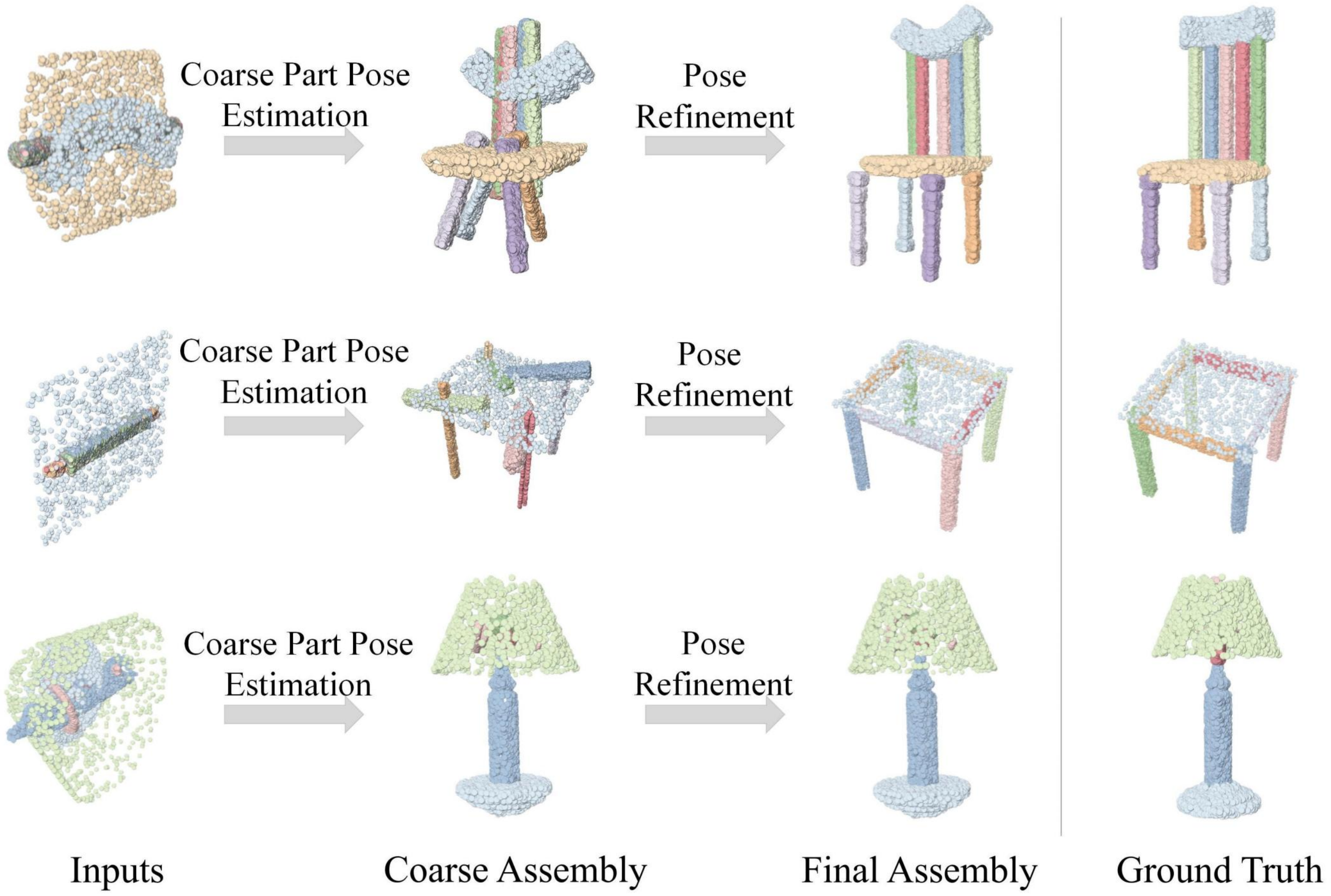
10) removes constraints on geometrically similar parts;

11) disables the symmetry-aware loss entirely.

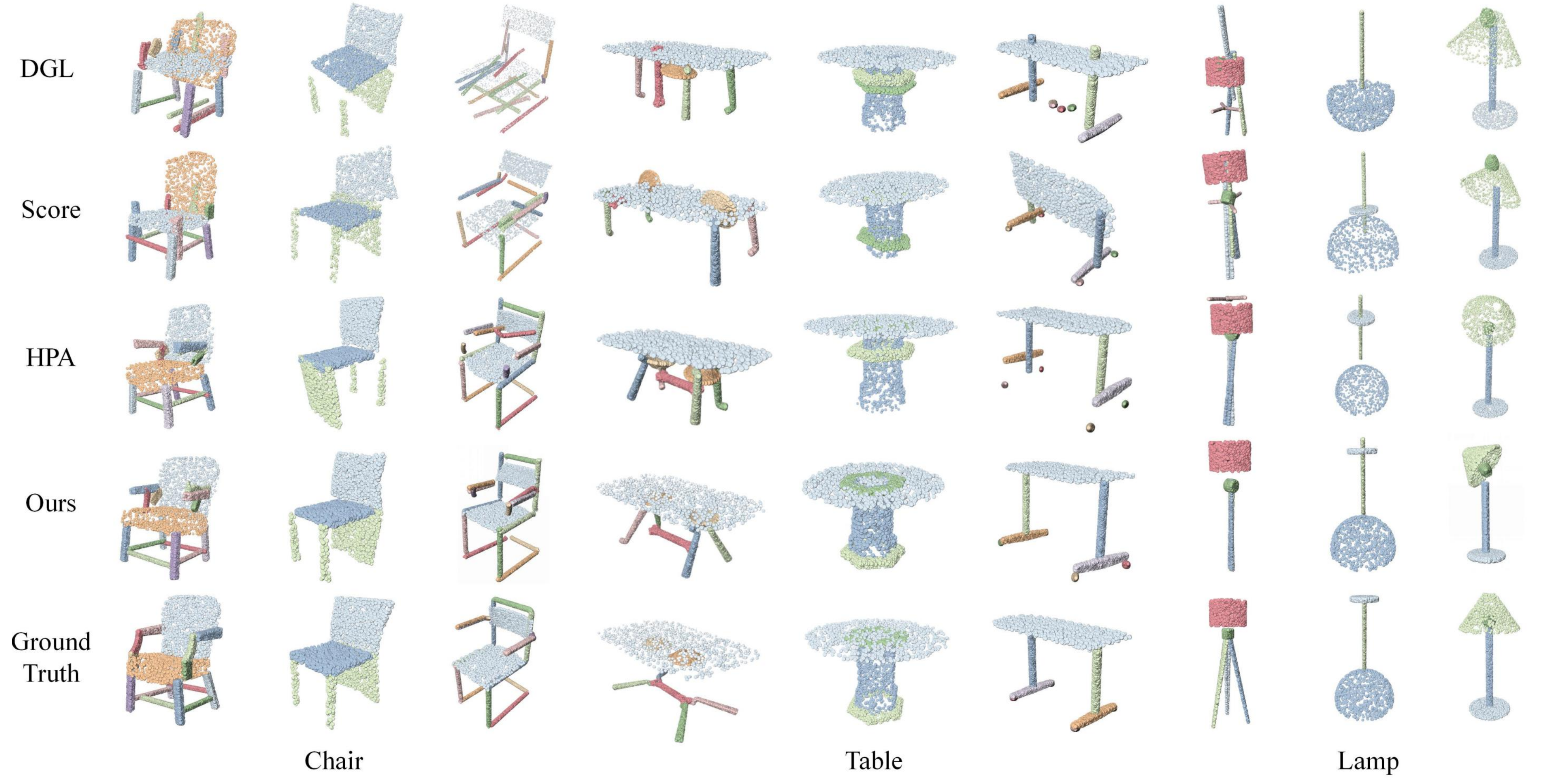
Table 6: Ablation study on symmetry-aware loss.

Methods	SCD(10^{-2})↓	PA(%)↑	CA(%)↑
9) CFPA-w/o-SS	0.51	67.86	47.24
10) CFPA-w/o-GS	0.51	68.98	47.49
11) CFPA-w/o-SL	0.52	67.51	47.17
CFPA	**0.49**	**69.24**	**49.20**

- **Visualization of coarse-to-fine progress:**



- **Qualitative results on the Chair, Table and Lamp categories:**



Conclusion

We introduced CFPA, a coarse-to-fine framework for 3D part assembly that achieves state-of-the-art performance by:

- Learning and propagating semantic structure via super-parts.
- Explicitly modeling geometric symmetries through a novel symmetry-aware loss.

Our method produces assemblies that are **not only accurate but also structurally consistent and diverse.**

Limitation

The current framework is designed for semantic parts and is not directly applicable to the reassembly of irregular fragments.

Future Work

- Extend CFPA to fragment reassembly.
- Explore unsupervised learning to improve robustness to unseen parts.