Generative 3D Part Assembly via Dynamic Graph Learning

Jialei Huang[1*] Guanqi Zhan[1*] Qingnan Fan[2] Kaichun Mo[2]

Lin Shao[2] Baoquan Chen[1] Leonidas J. Guibas[2] Hao Dong[1]

(*: indicates joint first authors)

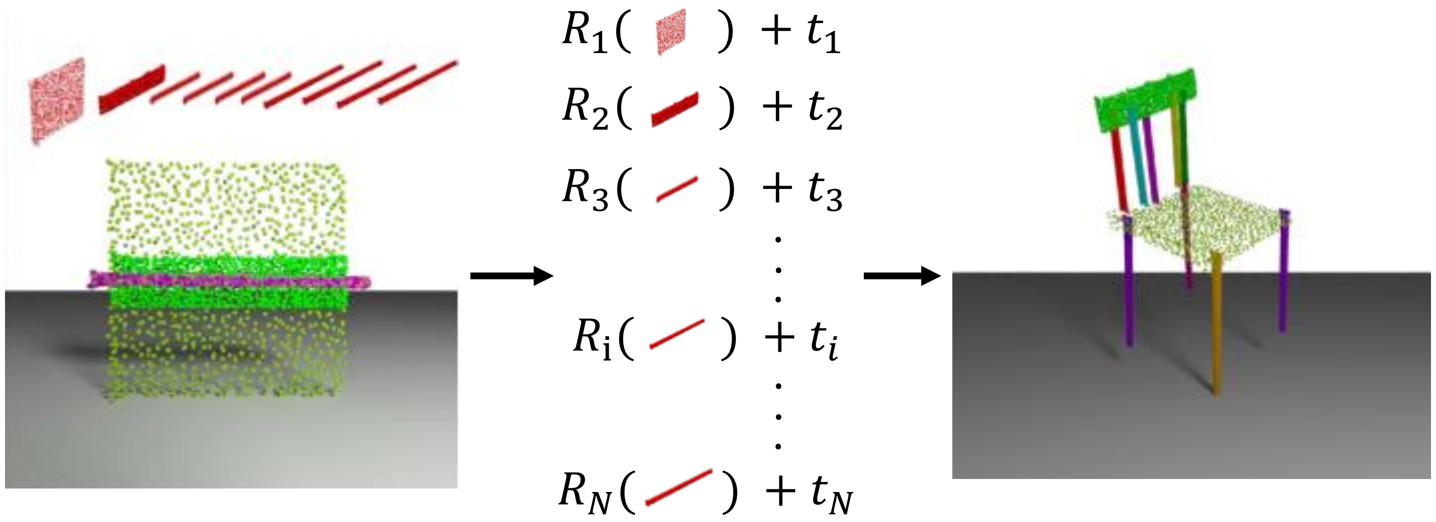
[1]Peking University [2]Stanford University

Code: https://github.com/hyperplane-lab/Generative-3D-Part-Assembly,

Contact: guanqi_zhan@pku.edu.cn

Introduction

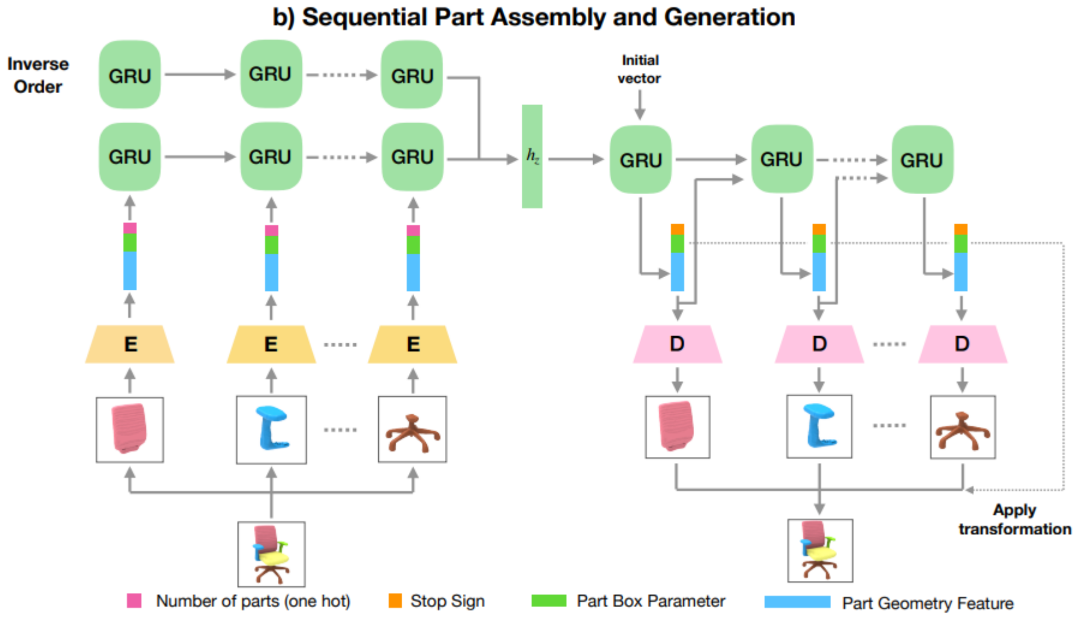
- Given parts laid out in the part canonical space, the goal is to compose a full shape.
  - formulated as the **pose estimation** problem



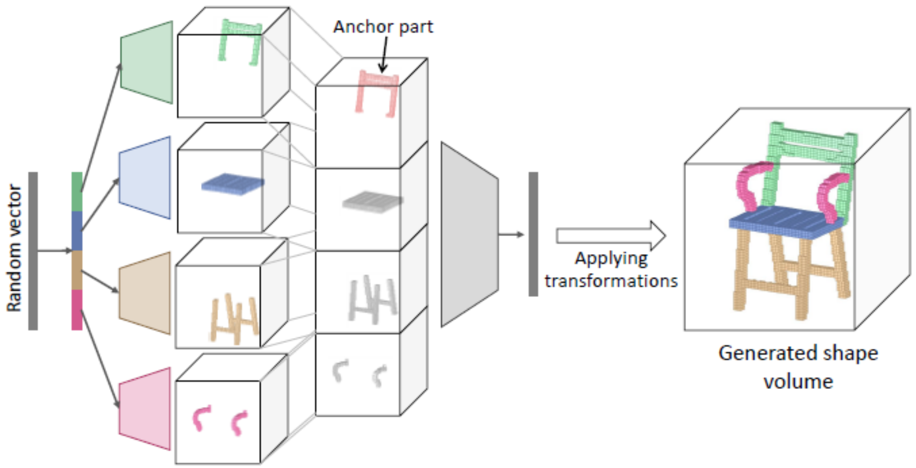
Suárez-Ruiz et al. Can robots assemble an IKEA chair? Science Robotics 2020

Literature

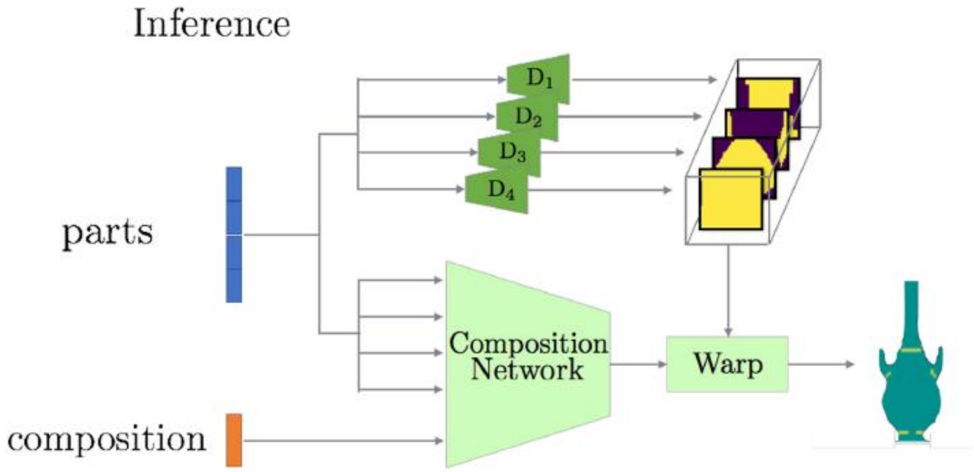
- Learn a per-part AE for the generative task.
- Transform: scale and translation.



[Wu et al. 2020]



[Li et al. 2020]

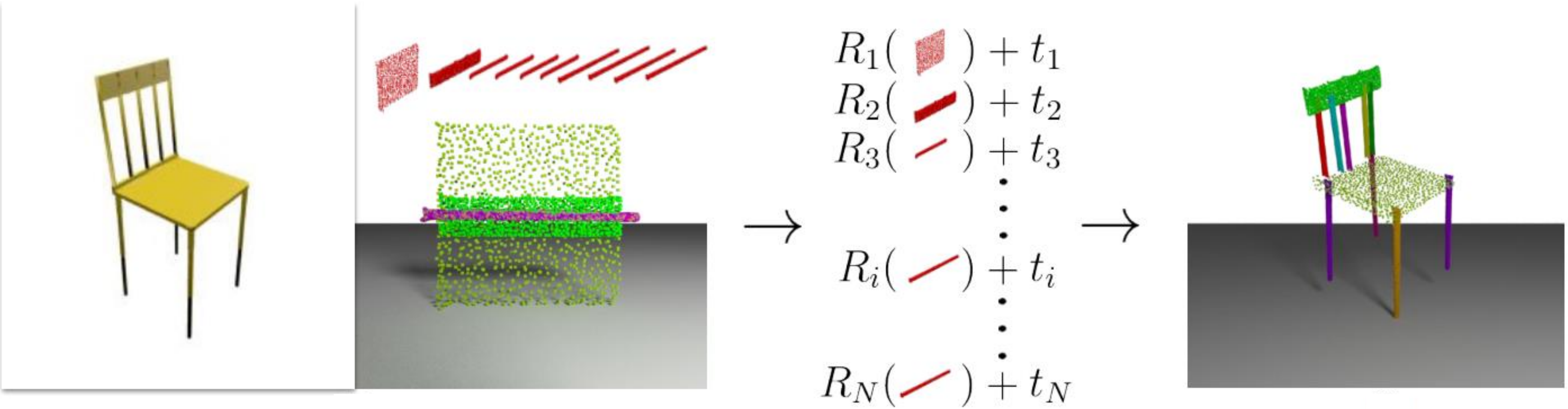


[Schor et al. 2019]

Wu et al. PQ-NET: A Generative Part Seq2Seq Network for 3D Shapes. CVPR 2020.
Li et al. Learning Part Generation and Assembly for Structure-aware Shape Synthesis. AAAI 2020.
Schor et al. CompoNet: Learning to Generate the Unseen by Part Synthesis and Composition. ICCV 2019

Literature

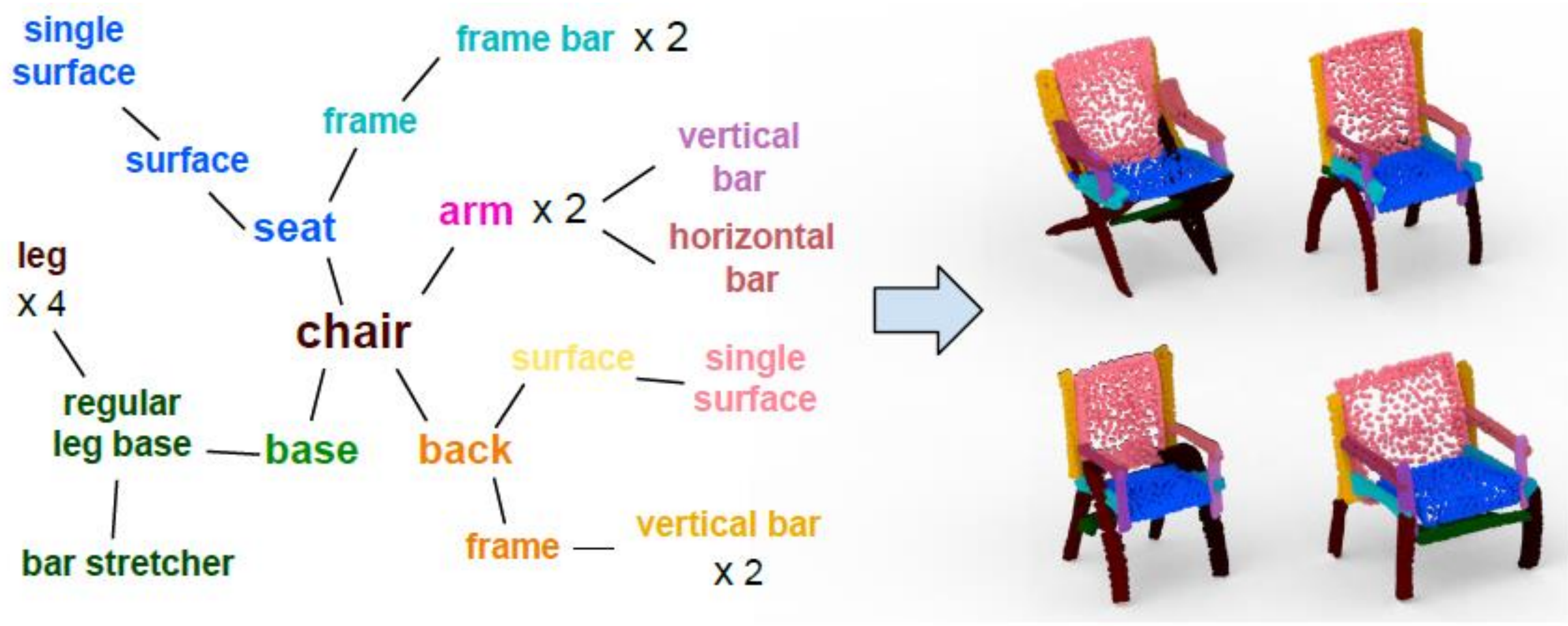
Part assembly in the camera space from a guidance image.



Li and Mo et al. Learning 3D Part Assembly from a Single Image. 2020.

Literature

Generate shapes from a fixed shape tree structure.



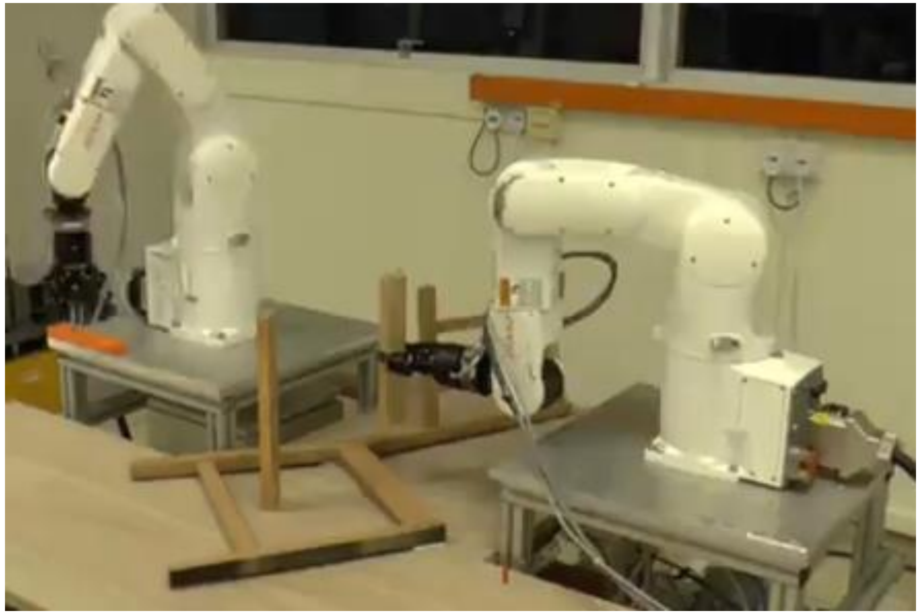
Mo et al. PT2PC: Learning to Generate 3D Point Cloud Shapes from Part Tree Conditions. 2020

北京大學 PEKING UNIVERSITY

Motivation

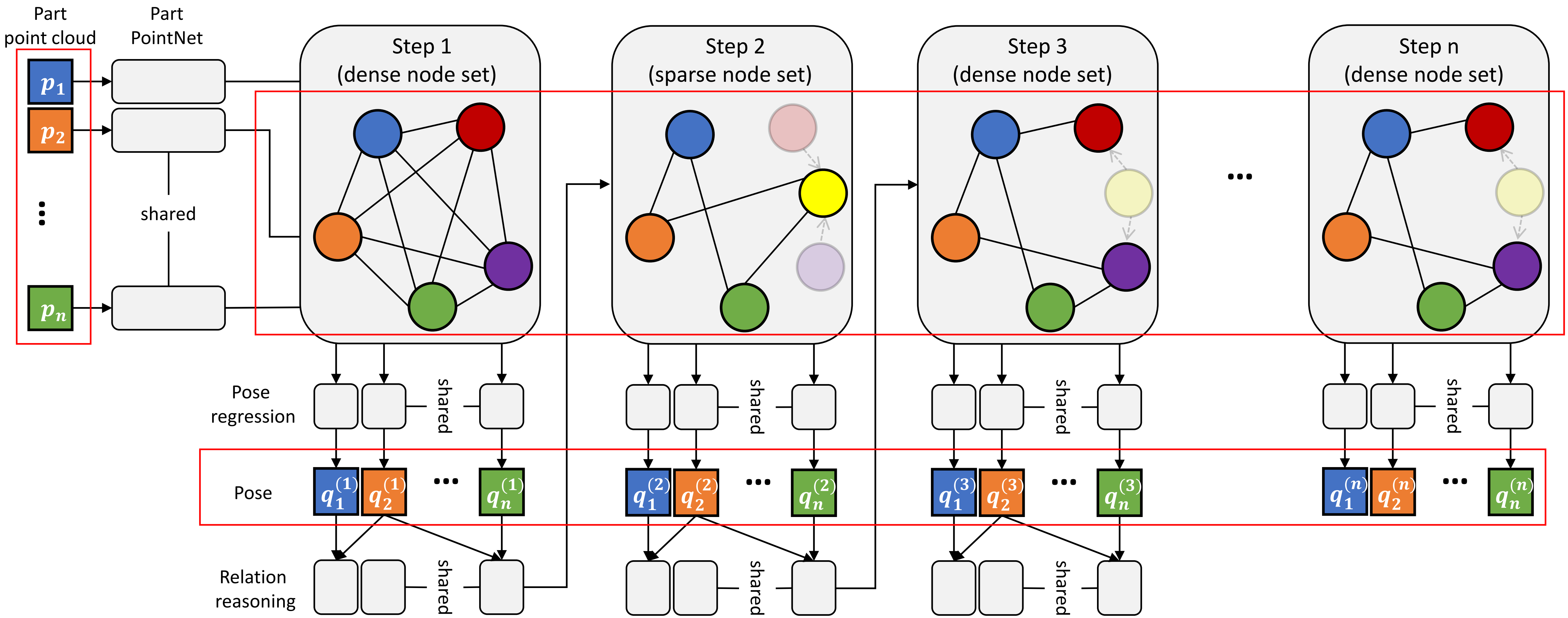
- **Important characteristics of the previous works:**
  - Allow **free-form part generation** for part geometry
  - Assume **well-rotated** part pose in the beginning
  - Assume certain **part priors**

- **A more practical problem setting in our project:**
  - **Parts are provided**, and no geometry is predicted.
  - 6D part pose of **rotation** and translation.
  - Assume **no prior knowledge** upon the input parts



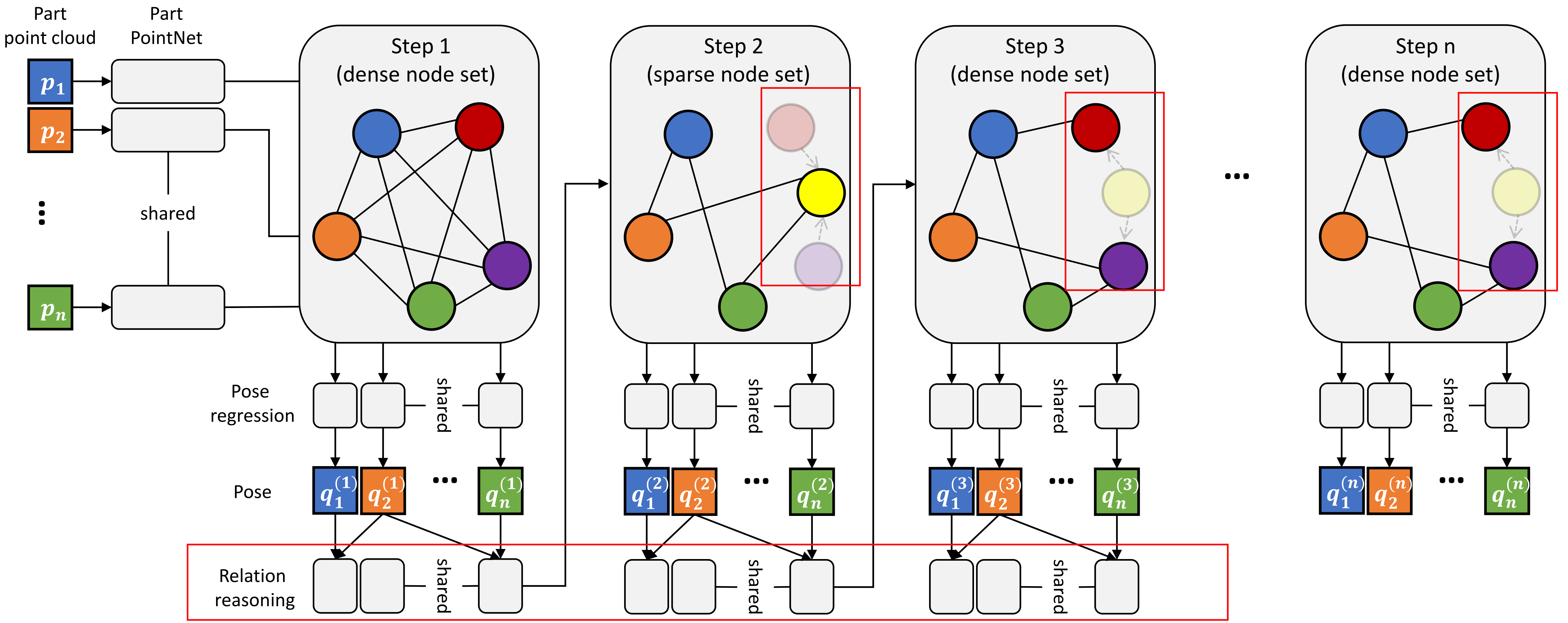
Method

- **The proposed dynamic graph learning framework**



Method

- **The proposed dynamic graph learning framework**



Method

- **Iterative GNN Backbone**
  - Edge attribute from node $v_j^{(t)}$ to $v_i^{(t)}$: $e_{ij}^{(t)} = f_{edge}(v_i^{(t)}, v_j^{(t)})$
  - Update the node attribute $v_i^{(t+1)}$: $v_i^{(t+1)} = f_{node}(v_i^{(t)}, \frac{1}{N}\sum_{j=1}^{N} e_{ij}^{(t)})$
  - Predict the 6-DoF part pose $q_i^{(t+1)}$: $q_i^{(t+1)} = f_{pose}(v_i^{(0)}, v_i^{(t+1)}, q_i^{(t)})$
- **Dynamic Relation Reasoning Module**
  - Reasoning the relation from the estimated poses: $r_{ij}^{(t)} = f_{relation}(f_{feat}\left(q_i^{(t-1)}\right), f_{feat}(q_j^{(t-1)}))$
  - Update the node attribute $v_i^{(t+1)}$ according to the implicitly learned relation weight: $v_i^{(t+1)} = f_{node}(\mathrm{v_i^{(t)}}, \frac{\sum_j e_{ij}^{(t)} r_{ij}^{(t)}}{\sum_j r_{ij}^{(t)}})$
- **Dynamic Part Aggregation Module**
  - Aggregate the node attributes among the geometrically-equivalent parts V_g into a single node $v_k^{(t)}$: $v_j^{(t)} = pooling_{k \in v_g}(v_k^{(t)})$

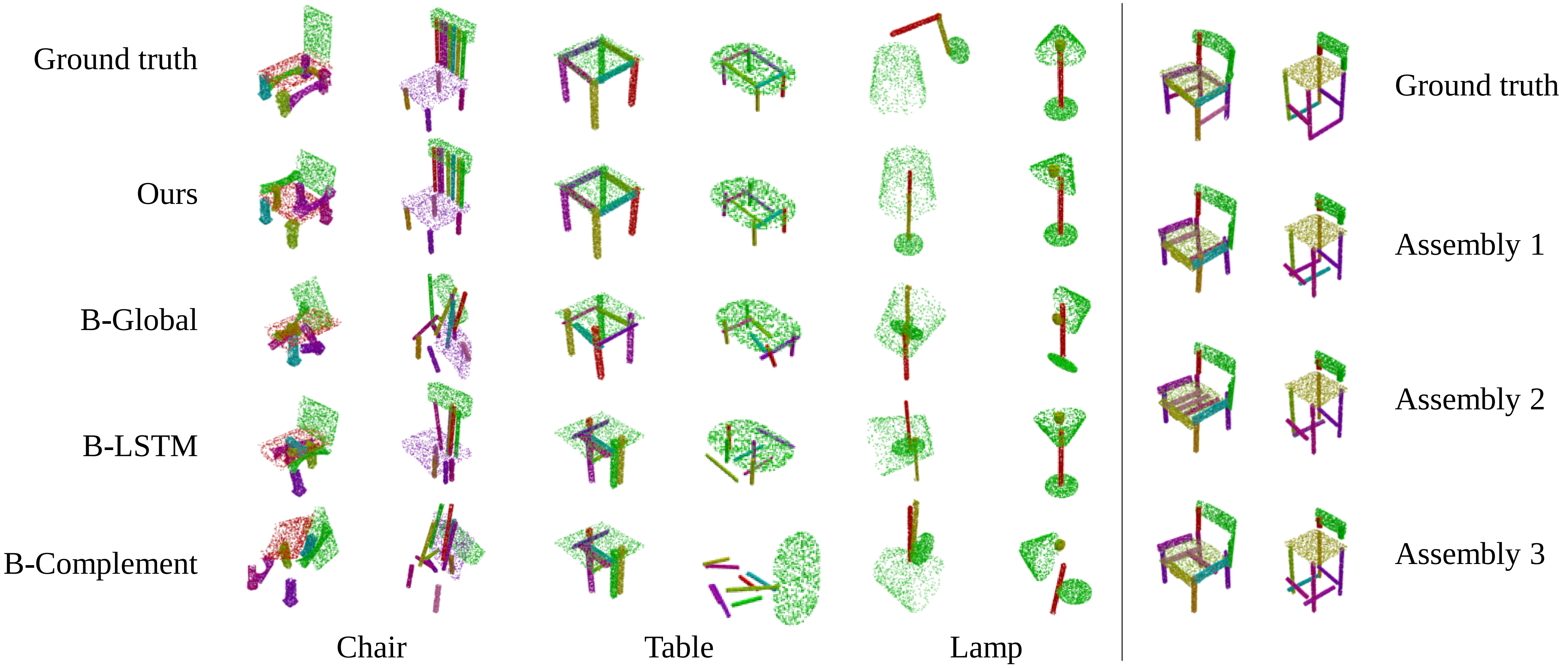
Experiments Results

- **Baselines**
  - B-Complement
  - B-LSTM
  - B-Global
- **Evaluation metrics**
  - Shape Chamfer Distance
  - Part Accuracy
  - Connectivity Accuracy

	Shape Chamfer Distance↓			**Part Accuracy↑**			**Connect Accuracy↑**		
	Chair	Table	Lamp	Chair	Table	Lamp	Chair	Table	Lamp
B-Global	0.0146	0.0112	0.0079	15.7	15.37	22.61	9.90	33.84	18.6
B-LSTM	0.0131	0.0125	**0.0077**	21.77	28.64	20.78	6.80	22.56	14.05
B-Complement	0.0241	0.0298	0.0150	8.78	2.32	12.67	9.19	15.57	26.56
Ours	**0.0091**	**0.0050**	0.0093	**39.00**	**49.51**	**33.33**	**23.87**	**33.96**	**41.70**

Quantitative Comparison between our approach and the baseline methods

Experiments Results



Qualitative Results. Left: visual comparisons between our algorithm and the baseline methods; Right: multiple plausible assembly results generated by our network.

Experiments Results

- **Ablation study**

	Shape CD↓	PA↑	CA↑
Our backbone w/o graph learning	0.0086	26.05	28.07
Our backbone	0.0055	42.09	35.87
Our backbone + relation reasoning	0.0052	46.85	38.60
Our backbone + part aggregation	0.0051	48.01	38.13
Exchange dense/sparse node set iteration	0.0052	49.19	39.62
Input GT adjacency relation	0.0053	45.43	35.66
Reasoning relation from geometry	0.0053	45.11	39.21
Our full algorithm	**0.0050**	**49.51**	**39.96**

Ablation study to demonstrate the effectiveness of each component of our algorithm. Shape CD, PA and CA are short for Shape Chamfer Distance, Part Accuracy and Connectivity Accuracy respectively.

Experiments and Analysis

- **Dynamically evolving part relation weights $r_{ij}^{(t)}$** among four common chair part types:
  - The orange cells highlight the four directed edges with the **maximal learned relation weight** in the matrix, while the yellow cells indicate the **minimal ones**.
  - The vertical axis denotes the **emitting** parts, and the horizontal axis denotes the **receiving** parts.

	back	seat	leg	arm
back	0.82	1.27	1.83	1.92
seat	0.67	0.27	2.75	2.43
leg	1.24	1.05	0.23	0.62
arm	1.14	1.21	0.92	0.58

Step 2

	back	seat	leg	arm
back	0.96	1.25	0.77	0.83
seat	0.64	0.78	0.66	0.50
leg	1.26	1.18	1.58	1.61
arm	1.00	0.67	0.43	0.85

Step 3

	back	seat	leg	arm
back	0.85	0.84	2.09	3.55
seat	3.16	0.55	1.76	0.94
leg	0.45	1.13	0.25	0.55
arm	0.95	1.40	1.25	0.71

Step 4

	back	seat	leg	arm
back	1.00	1.35	0.85	0.93
seat	0.90	0.83	0.84	0.65
leg	1.04	1.03	1.29	1.19
arm	1.09	0.68	0.65	1.02

Step 5

- **Some discoveries:**
  - **Similar pattern in both even and odd steps**
  - On average, **central parts** (eg: back, seat) emit more relation weights than **peripheral parts** (eg: leg, arm)
  - Central parts **guide the part assembly process** more

Experiments and Analysis

- **The time-varying part assembly results**
  - The poses for the central parts are firstly determined
  - And then the peripheral parts gradually adjust their poses to match the central parts

Step 1 Step 2 Step 3 Step 4 Step 5 Ground Truth

Additional Results - Analysis

- **Implicitly learned relation weights on additional parts and object categories**

	Emitted relations: Step 1	Step 2	Step 3	Step 4	Average	Received relations: Step 1	Step 2	Step 3	Step 4	Average
back	1.60	1.12	1.59	1.28	1.40	1.17	1.00	1.17	0.99	1.08
seat	1.45	0.69	1.57	0.83	1.14	1.01	1.01	0.90	1.00	0.98
head	1.60	1.54	1.01	1.21	1.34	0.66	1.02	1.43	0.76	0.97
regular_leg	0.92	1.42	0.84	1.14	0.94	1.35	0.84	1.36	0.92	1.12
arm	1.06	0.65	1.27	0.80	0.94	1.07	1.08	1.18	0.98	1.08
footrest	0.83	0.82	1.14	1.22	1.00	1.49	1.20	1.07	1.13	1.22
	0.90	1.22	0.77	1.16	1.01	0.67	1.09	1.39	1.03	1.05
star_leg	0.85	1.05	0.71	1.11	0.93	2.03	0.81	1.39	1.05	1.32

Emitted relations Received relations

(a) Chair

	Average (Emitted relations)	Average (Received relations)
base	1.58	1.09
body	1.45	0.97
power_cord	0.42	1.17
lamp_unit	0.82	1.04

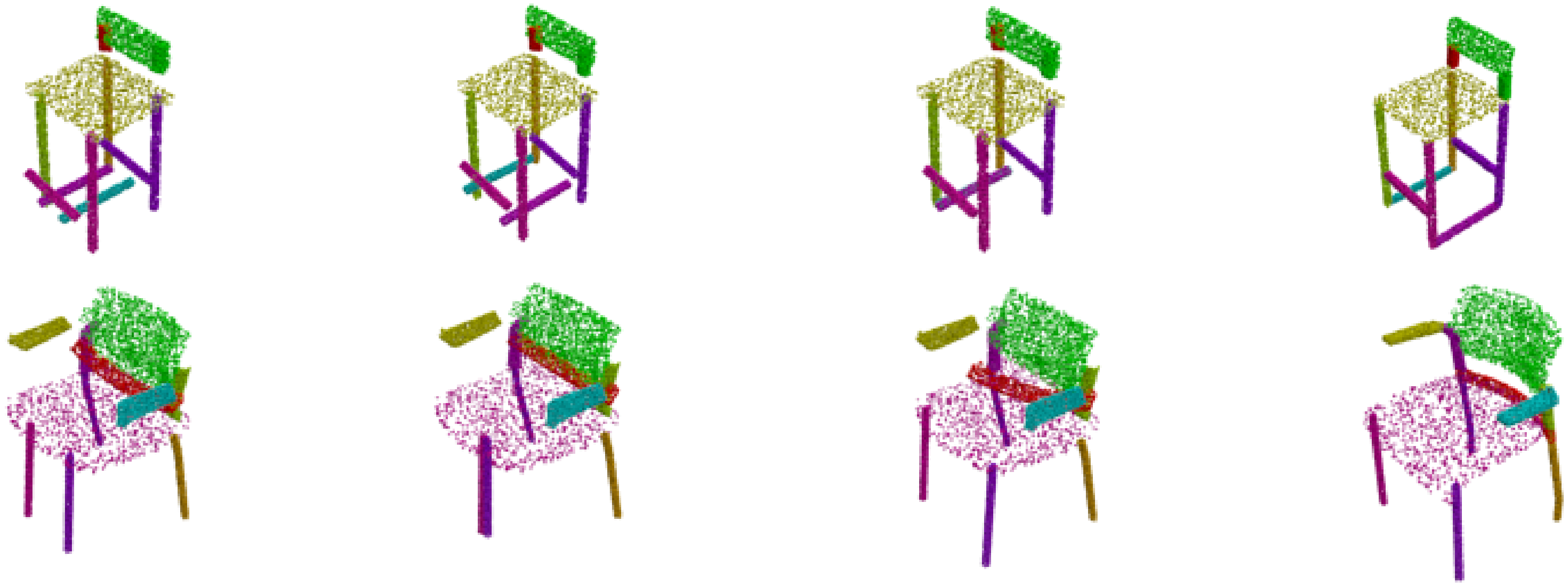
(b) Lamp

	Average (Emitted relations)	Average (Received relations)
shelf	1.28	0.95
drawer	0.68	0.63
cabinet_door	0.94	1.14
back_panel	0.97	1.15

(c) Table

Additional Results – Structural Variation

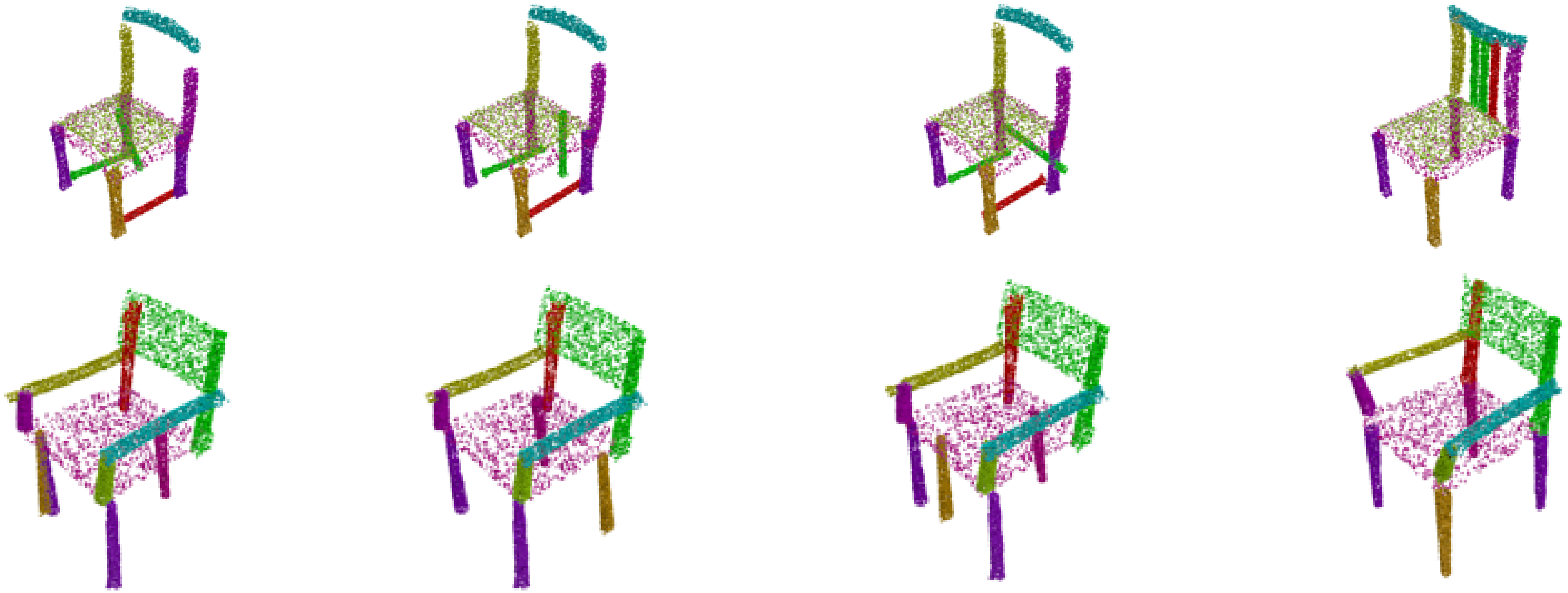
- **Structural variation demonstrated in part assembly**



Assembly1 Assembly2 Assembly3 Ground Truth

Additional Results – Structural Variation

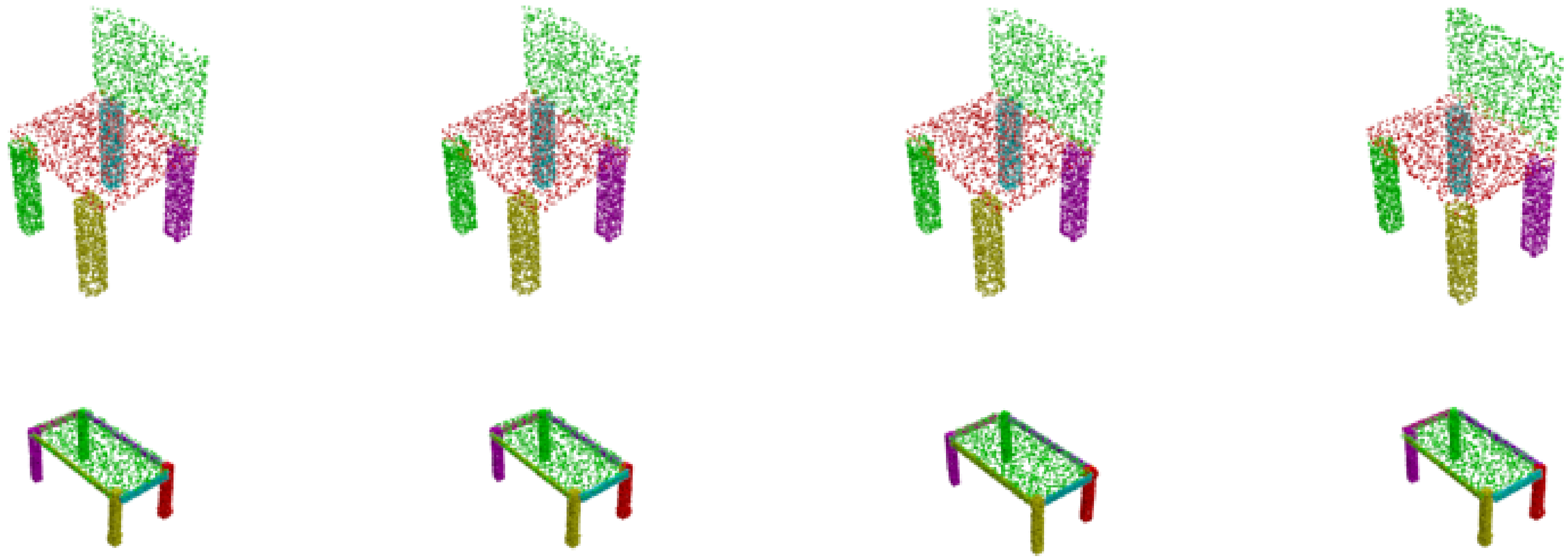
- **Structural variation demonstrated in part assembly**



Assembly1 Assembly2 Assembly3 Ground Truth

Additional Results – Structural Variation

- **No structural variation for the case with very limited parts**

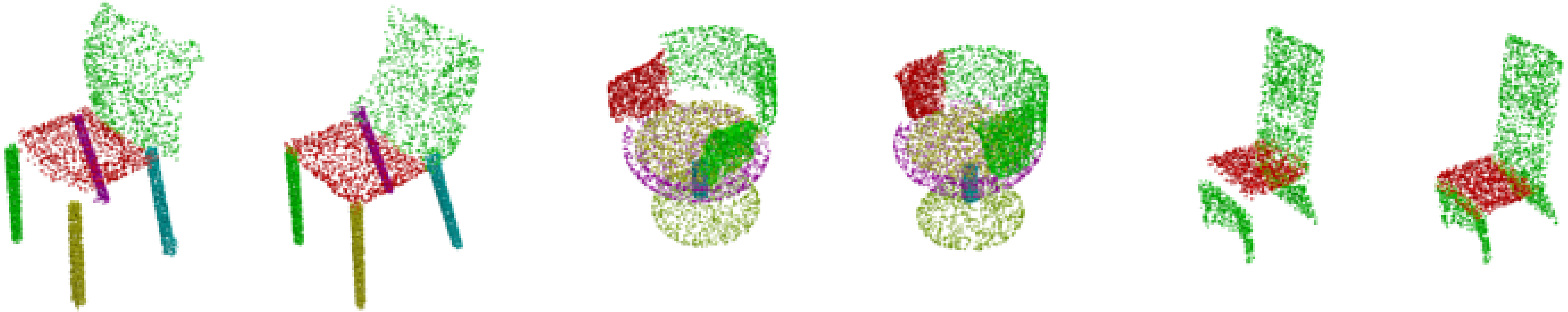


Assembly1 Assembly2 Assembly3 Ground Truth

北京大學 PEKING UNIVERSITY

Future Work

- Floating parts: sometimes fails to assemble a **well-connected** shape
  - eg: legs and arms **disconnected/misaligned** from base and back



Ours Ground Truth Ours Ground Truth Ours Ground Truth

- Floating **Implicit** soft relation --> **explicit** hard connection constraints
- Approach**: incorporate a joint-centric assembly framework**

Thanks