Australian National University

MITSUBISHI ELECTRIC

AUSTRALIAN INSTITUTE FOR MACHINE LEARNING

ICCV OCT 19-23, 2025 HONOLULU HAWAII

Manual-PA:
Learning 3D Part Assembly from Instruction Diagrams

Jiahao Zhang[1]

Anoop Cherian[2]

Cristian Rodriguez[3]

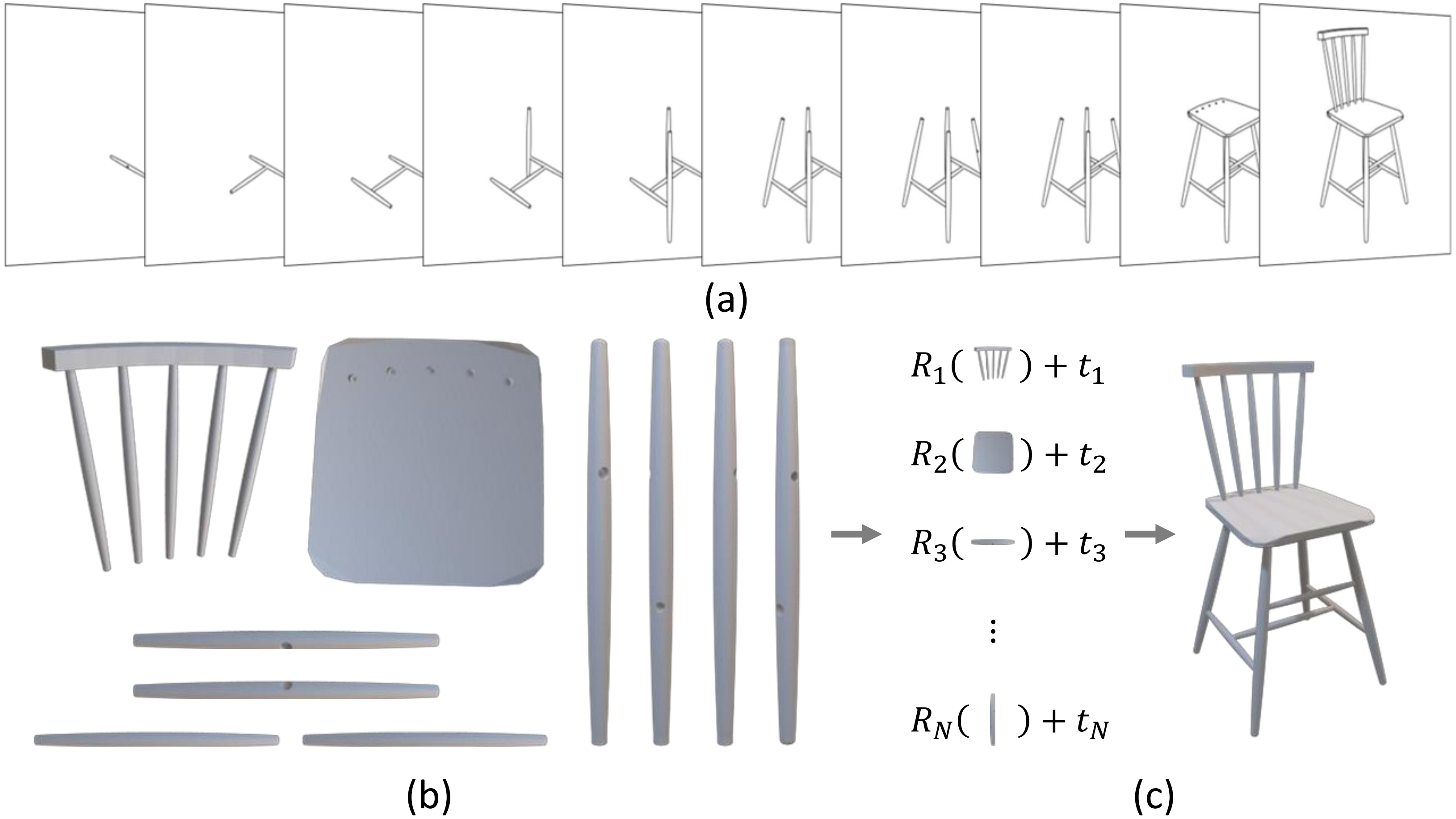
Weijian Deng[1]

Stephen Gould[1]

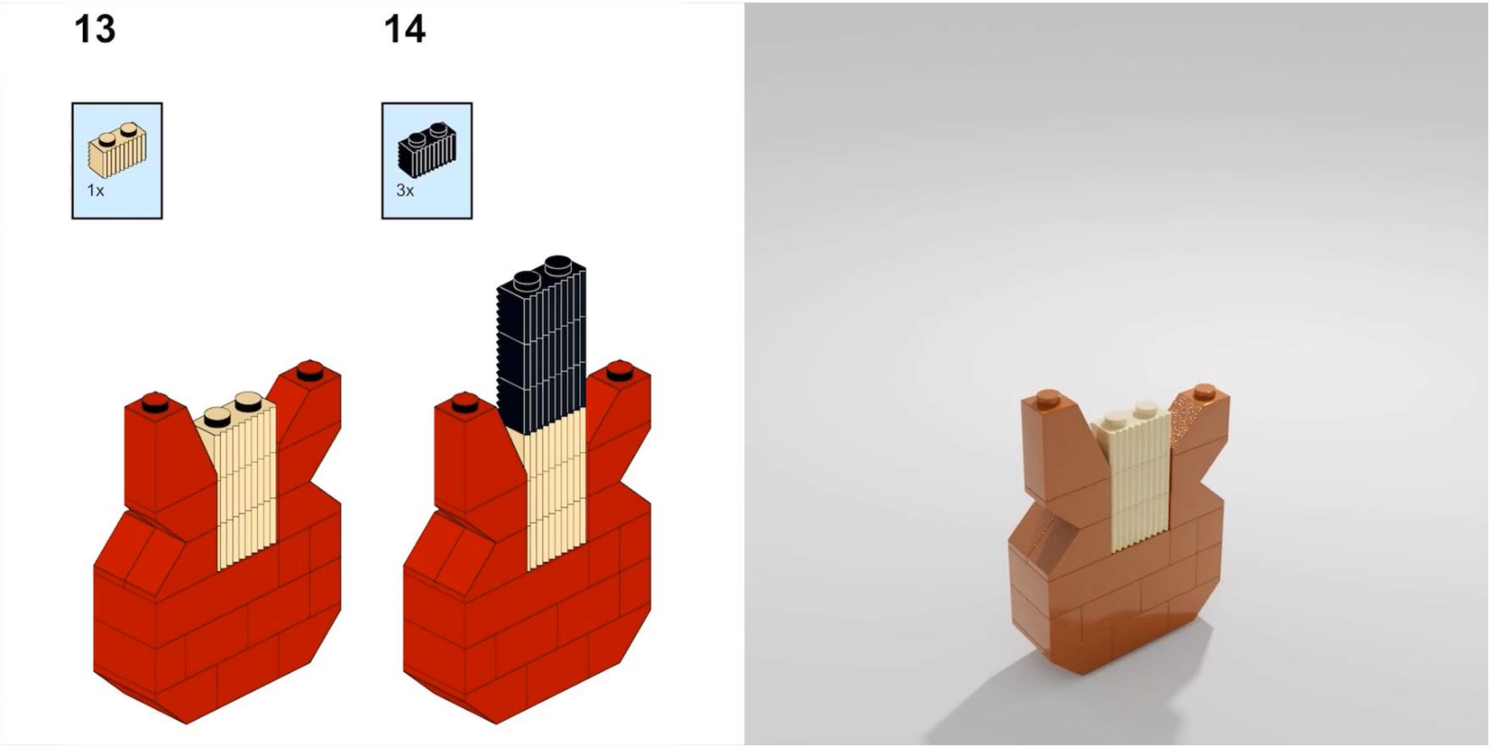
[1]The Australian National University
[2]Mitsubishi Electric Research Labs
[3]The Australian Institute for Machine Learning

Code & Dataset: https://github.com/DavidZhang73/Manual-PA
Poster: Session 2 (Oct 21 PM)

Problem Formulation



Related Works – Assembly Manual – MEPNet



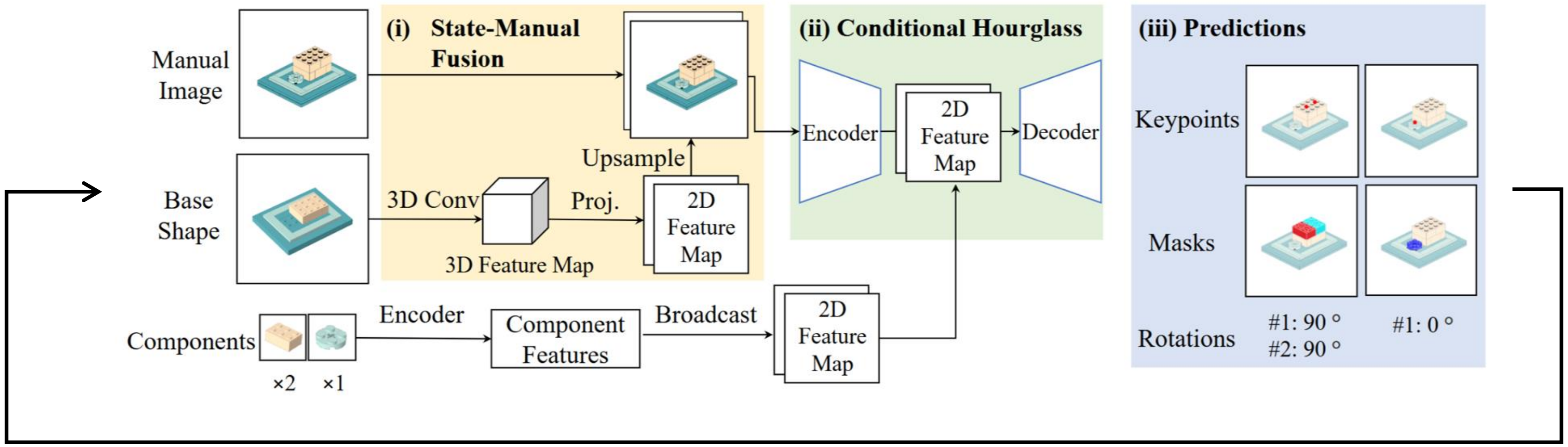
(a) Step-by-step LEGO Manual

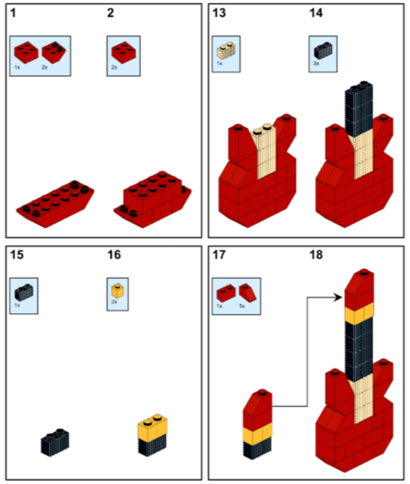
(b) Inferred Assembly Process

Our Model can assemble LEGO objects according to manuals.

Wang et al.: Translating a Visual LEGO Manual to a Machine-Executable Plan (ECCV'22)

Accumulating Errors!



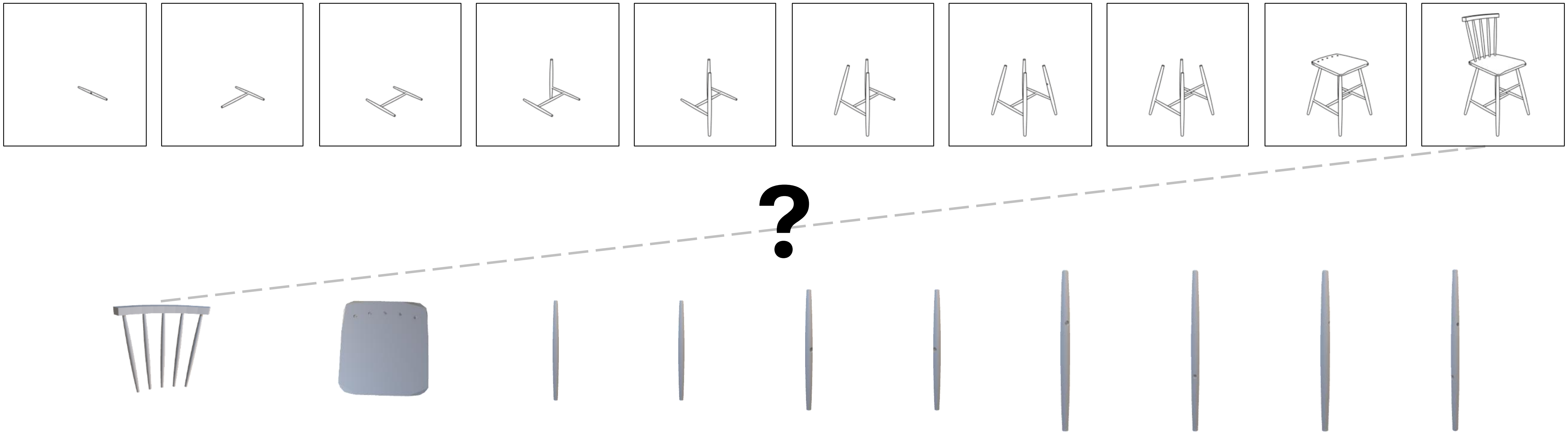


- Autoregressive arch.
  - Base shape => predicted result from pervious step
- Components => from ground truth

Wang et al.: Translating a Visual LEGO Manual to a Machine-Executable Plan (ECCV'22)

Key Issues

- How to learn correspondence between step diagrams and 3D parts?



Key Issues

- How to learn correspondence between step diagrams and 3D parts?
  - Contrastive learning

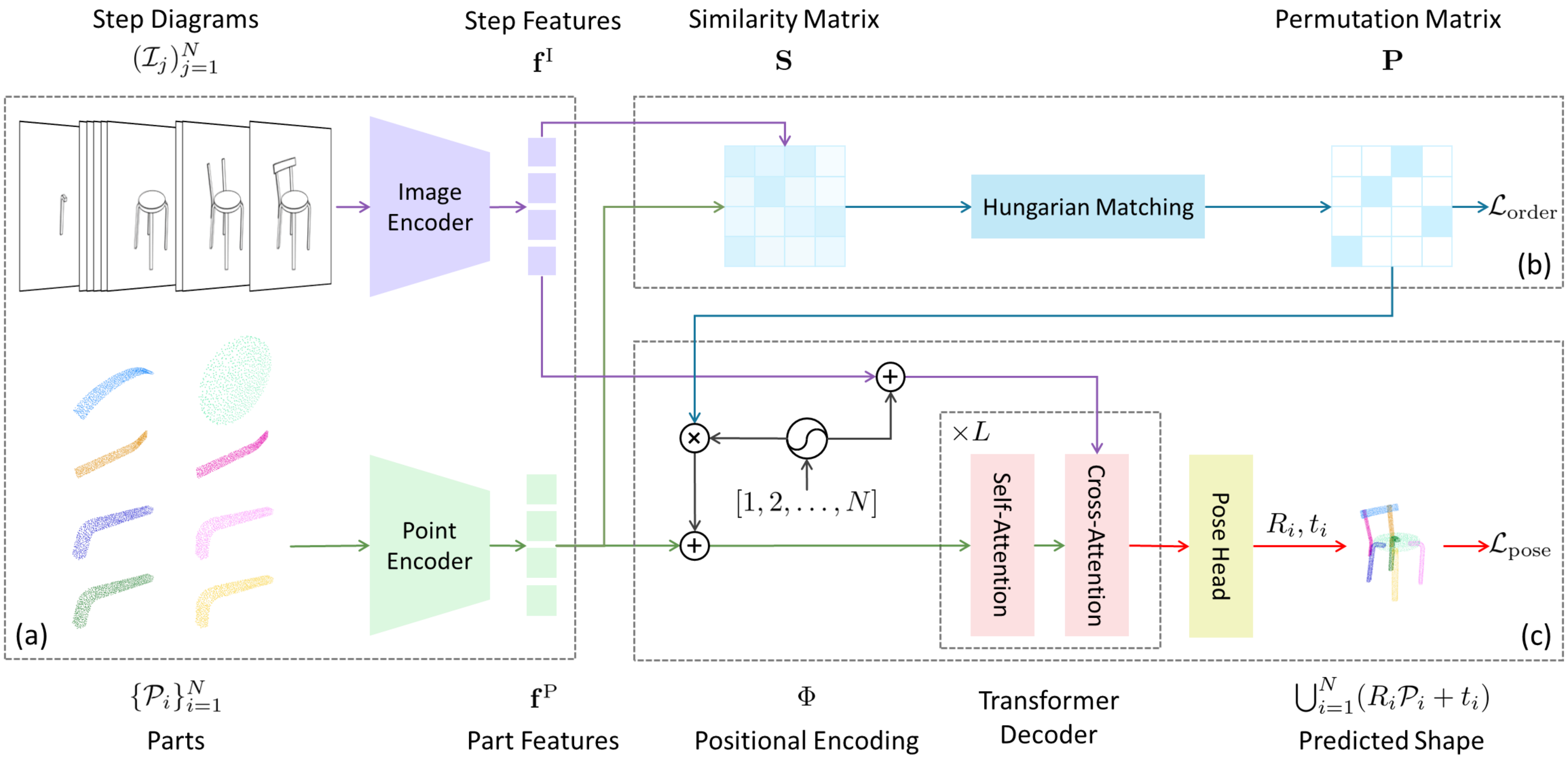
Key Issues

- How to learn correspondence between step diagrams and 3D parts?
  - Contrastive learning
- How to incorporate with the learned correspondence/order for pose estimation?

Key Issues

- How to learn correspondence between step diagrams and 3D parts?
  - Contrastive learning
- How to incorporate with the learned correspondence/order for pose estimation?
  - Via positional encoding as a soft guidance

⭐ Ours: Manual-PA

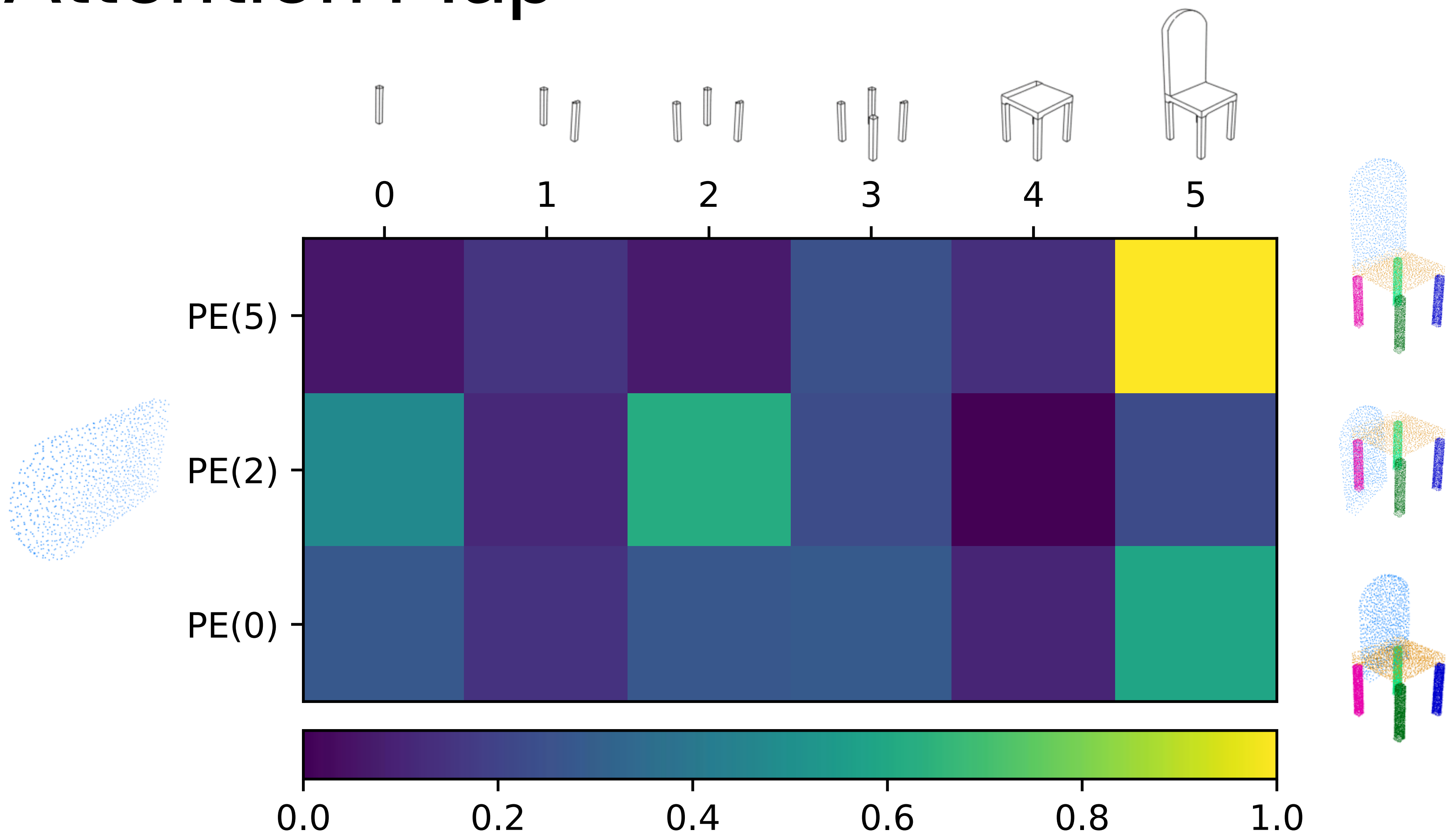


Zhang et al., ICCV 2025

Soft Guidance?!

- Hard guidance?
  - The 3D assembly model can **NOT** self-correct if the predicted correspondence is wrong.

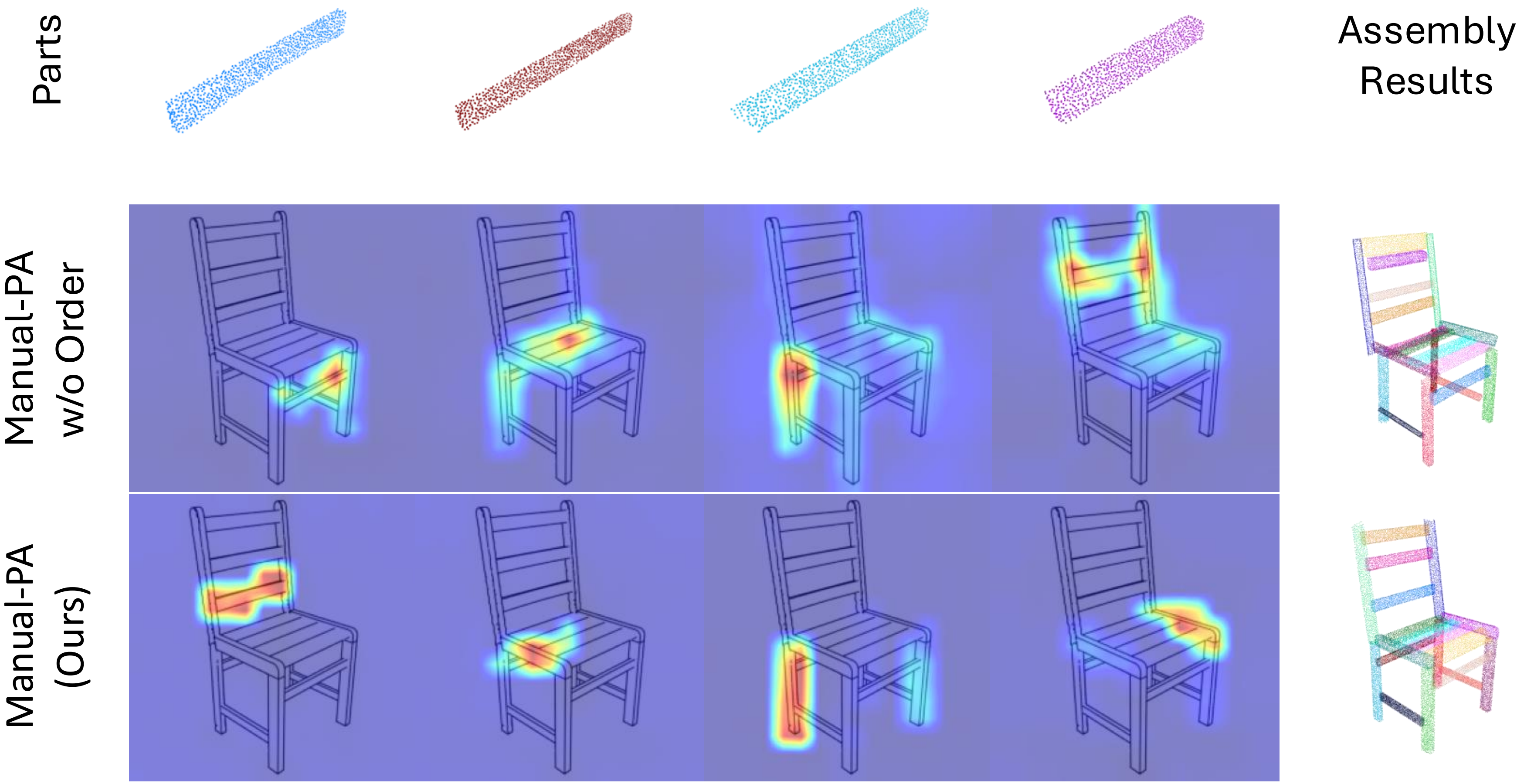
⭐Attention Map



Soft Guidance?!

- Hard guidance?
  - The 3D assembly model can **NOT** self-correct if the predicted correspondence is wrong.
- No guidance?
  - The 3D assembly model need to learn the correspondence implicitly by itself.

⭐ Attention Map



Quantitative Results

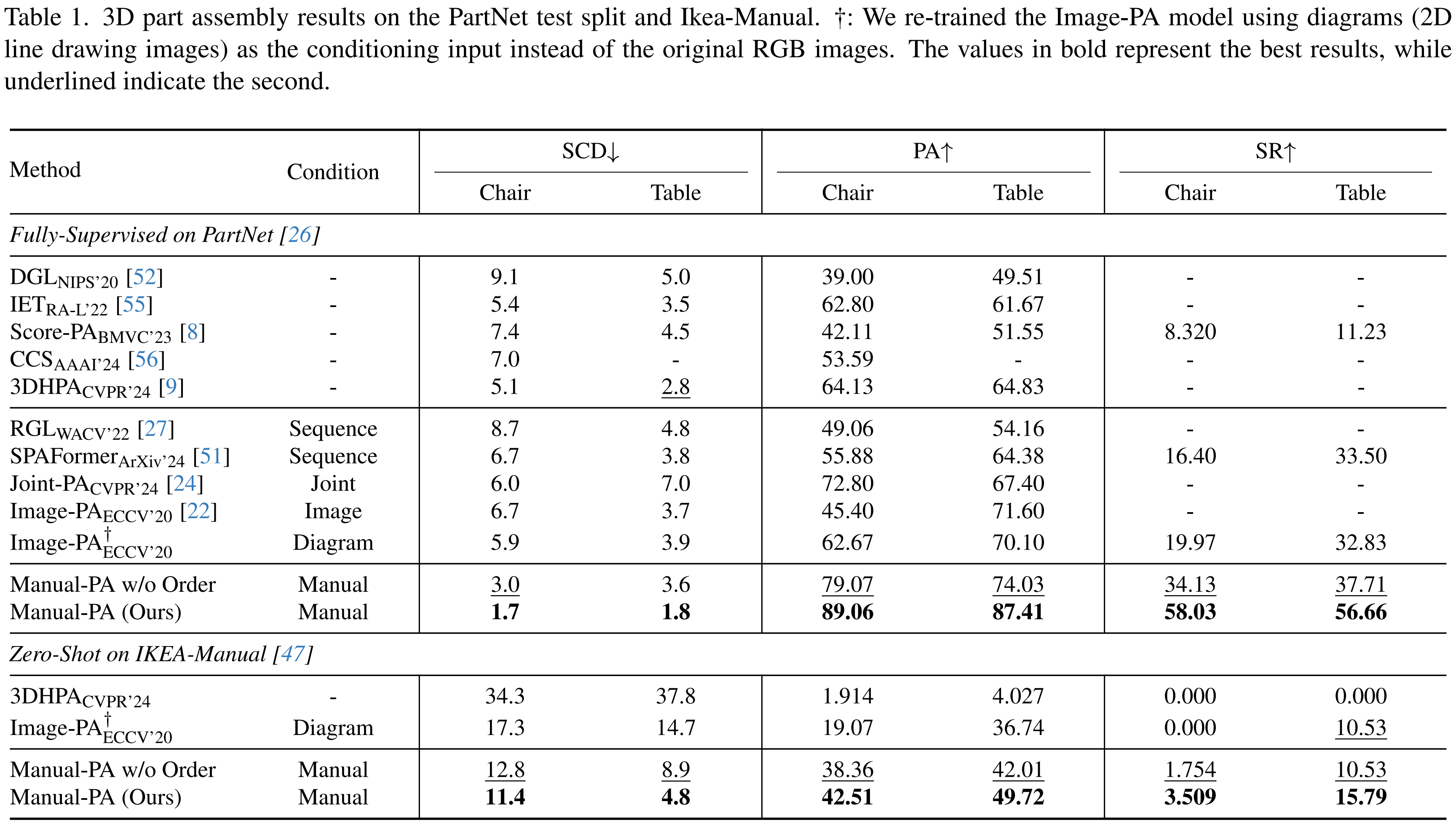
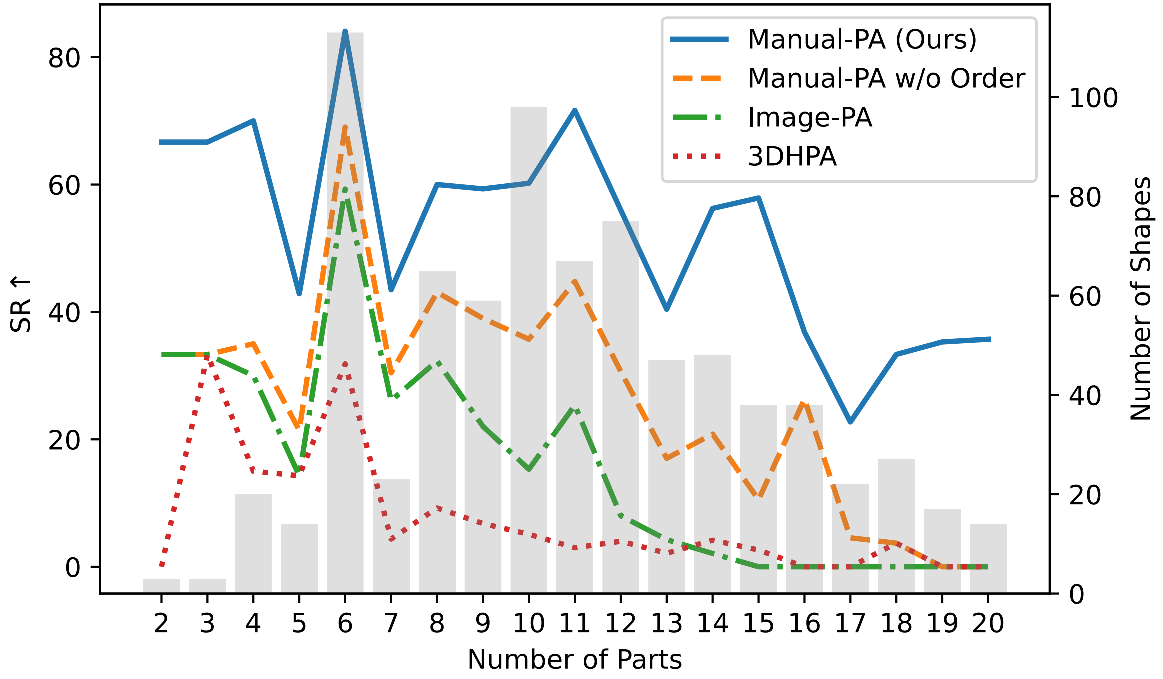


Table 1. 3D part assembly results on the PartNet test split and Ikea-Manual. †: We re-trained the Image-PA model using diagrams (2D line drawing images) as the conditioning input instead of the original RGB images. The values in bold represent the best results, while underlined indicate the second.

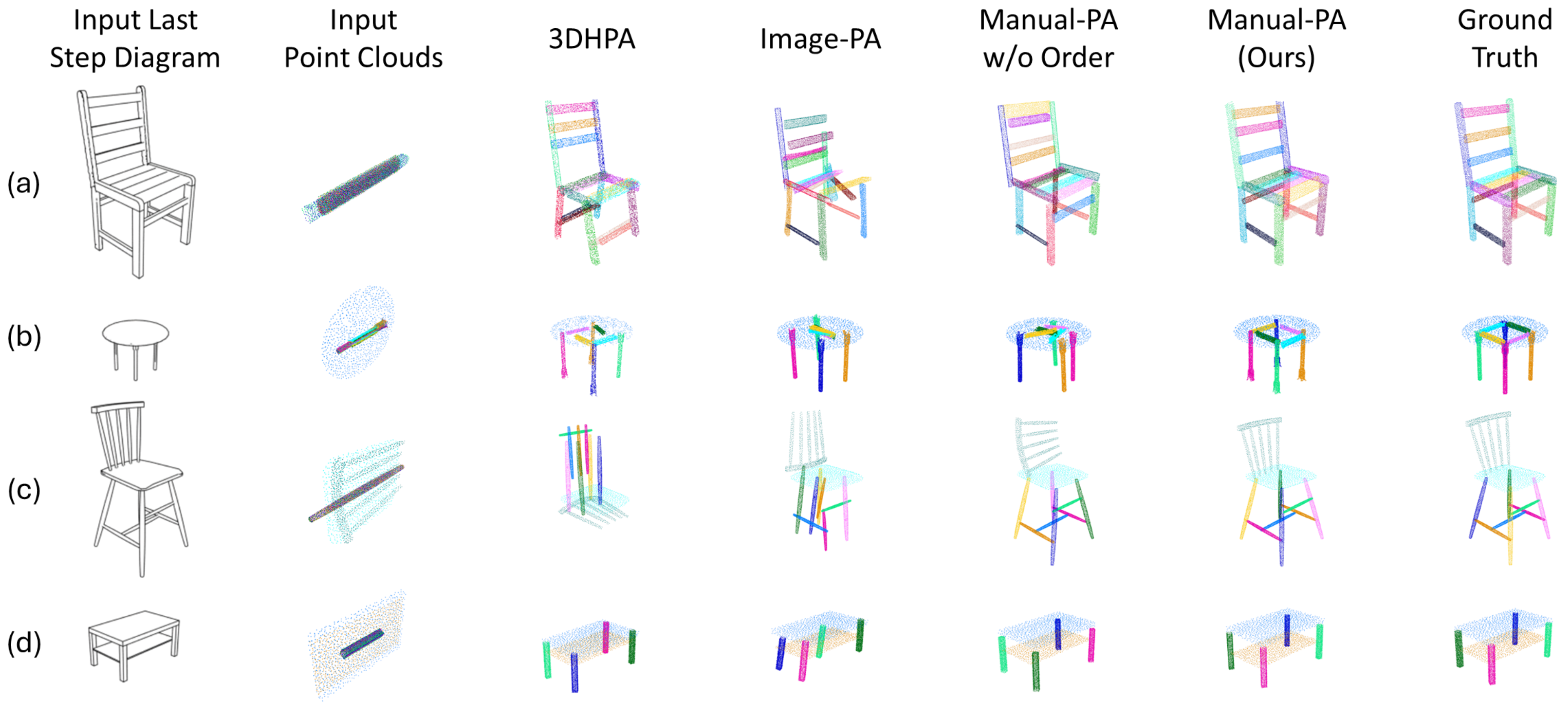
Method	Condition	SCD↓		PA↑		SR↑	
		Chair	Table	Chair	Table	Chair	Table
Fully-Supervised on PartNet [26]							
DGL$_{\text{NIPS'20}}$ [52]	-	9.1	5.0	39.00	49.51	-	-
IET$_{\text{RA-L'22}}$ [55]	-	5.4	3.5	62.80	61.67	-	-
Score-PA$_{\text{BMVC'23}}$ [8]	-	7.4	4.5	42.11	51.55	8.320	11.23
CCS$_{\text{AAAI'24}}$ [56]	-	7.0	-	53.59	-	-	-
3DHPA$_{\text{CVPR'24}}$ [9]	-	5.1	<u>2.8</u>	64.13	64.83	-	-
RGL$_{\text{WACV'22}}$ [27]	Sequence	8.7	4.8	49.06	54.16	-	-
SPAFormer$_{\text{ArXiv'24}}$ [51]	Sequence	6.7	3.8	55.88	64.38	16.40	33.50
Joint-PA$_{\text{CVPR'24}}$ [24]	Joint	6.0	7.0	72.80	67.40	-	-
Image-PA$_{\text{ECCV'20}}$ [22]	Image	6.7	3.7	45.40	71.60	-	-
Image-PA$^{\dagger}_{\text{ECCV'20}}$	Diagram	5.9	3.9	62.67	70.10	19.97	32.83
Manual-PA w/o Order	Manual	<u>3.0</u>	3.6	<u>79.07</u>	<u>74.03</u>	<u>34.13</u>	<u>37.71</u>
Manual-PA (Ours)	Manual	**1.7**	**1.8**	**89.06**	**87.41**	**58.03**	**56.66**
Zero-Shot on IKEA-Manual [47]							
3DHPA$_{\text{CVPR'24}}$	-	34.3	37.8	1.914	4.027	0.000	0.000
Image-PA$^{\dagger}_{\text{ECCV'20}}$	Diagram	17.3	14.7	19.07	36.74	0.000	<u>10.53</u>
Manual-PA w/o Order	Manual	<u>12.8</u>	<u>8.9</u>	<u>38.36</u>	<u>42.01</u>	<u>1.754</u>	<u>10.53</u>
Manual-PA (Ours)	Manual	**11.4**	**4.8**	**42.51**	**49.72**	**3.509**	**15.79**

Quantitative Results



Zhang et al., ICCV 2025

Qualitative Results

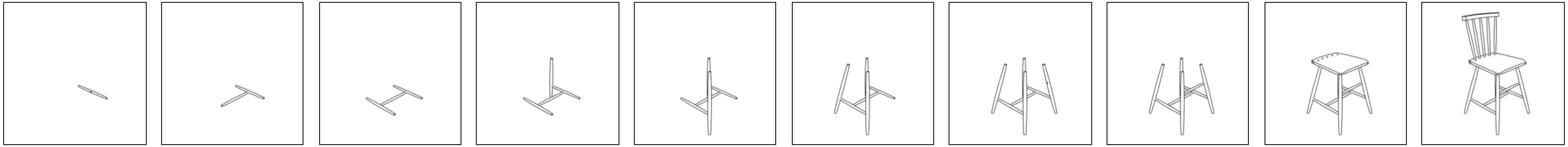


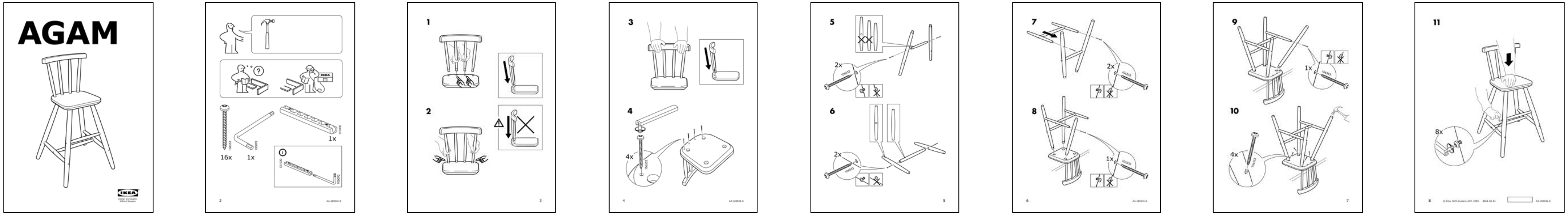
Limitations and Future Works

- Category-level vs. Universal

Limitations and Future Works

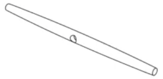
- Synthetic vs. Real World Manual





Demonstration Video

1



Manual

Image-PA

Manual-PA (Ours)

Thanks!

https://github.com/DavidZhang73/Manual-PA