Combinative Matching for Geometric Shape Assembly

Nahyuk Lee[1*] Juhong Min[1,2*] Junhong Lee[1] Chunghyun Park[1] Minsu Cho[1,3]

[1] Pohang University of Science and Technology (POSTECH) [2] Samsung Research America [3] RLWRLD

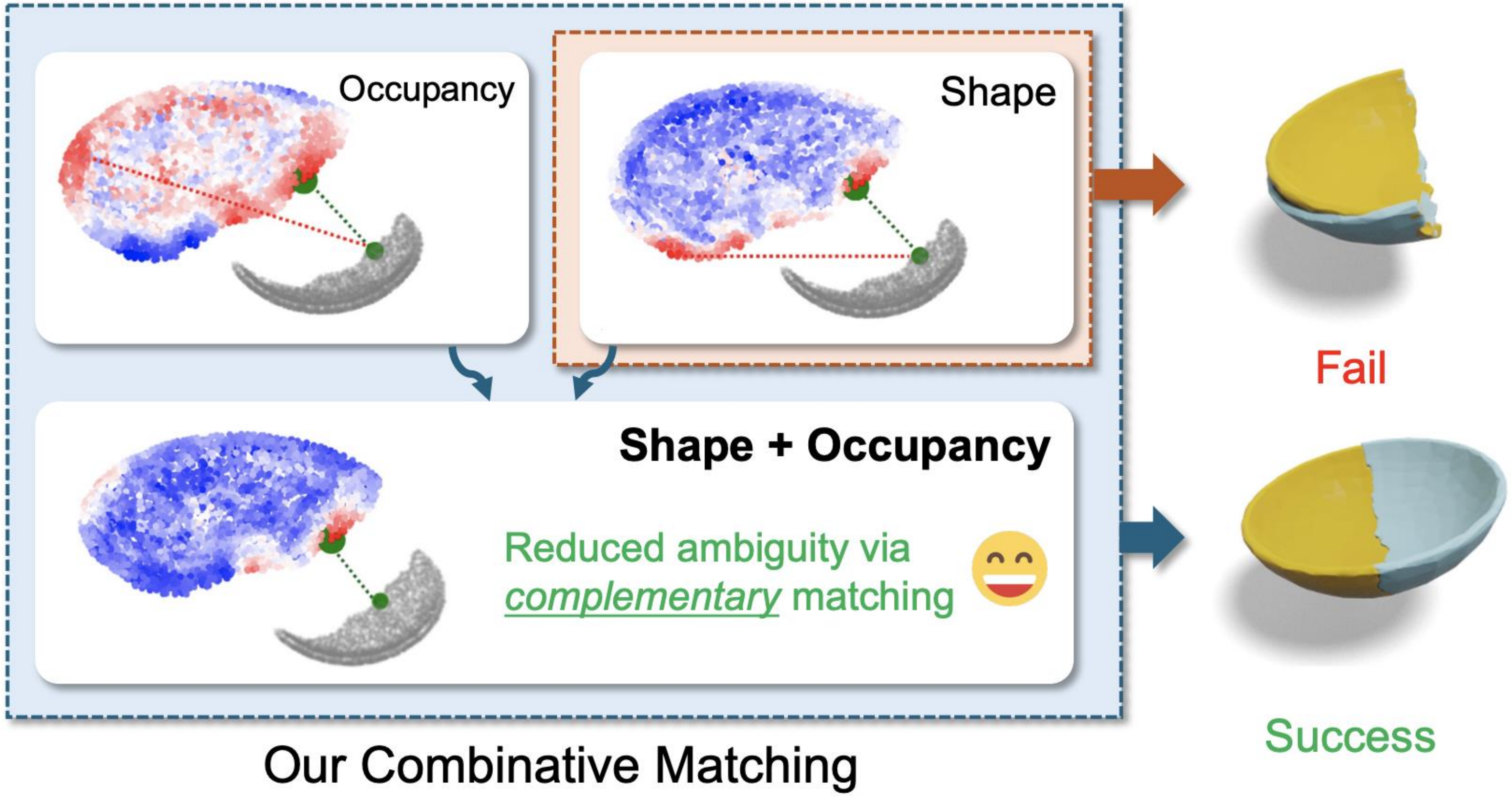
* Equal Contribution

ICCV OCT 19-23, 2025 HONOLULU HAWAII

Highlight Presentation ✨

POSTECH POHANG UNIVERSITY OF SCIENCE AND TECHNOLOGY SAMSUNG RLWRLD

Combinative Matching for Geometric Shape Assembly



3D Geometric Shape Assembly

Given a set of fractured pieces $\mathcal{P} = \{P_1, P_2, \ldots, P_N\}$, our goal is to recover the 6-DoF pose $\{T_1, T_2, \ldots, T_N\}$, in SE(3) for each piece and restore the underlying object.

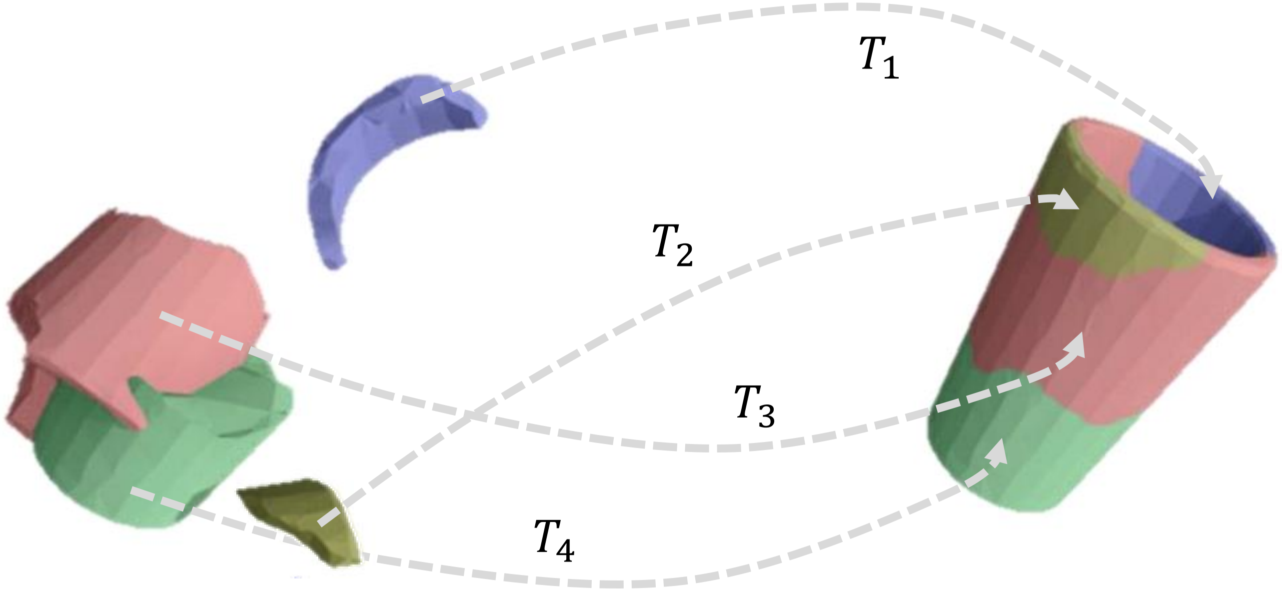
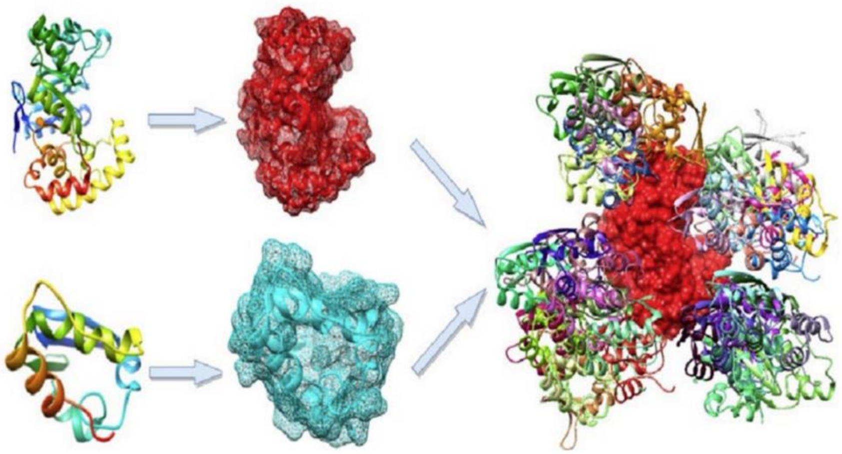


*Figure courtesy of Lu et al.

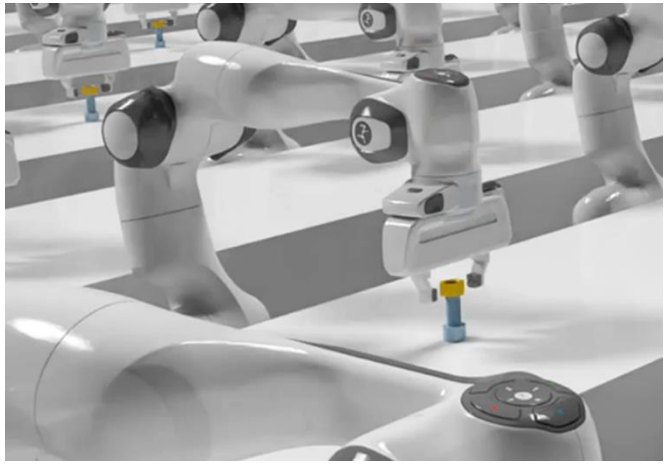
3D Geometric Shape Assembly

Archaeology



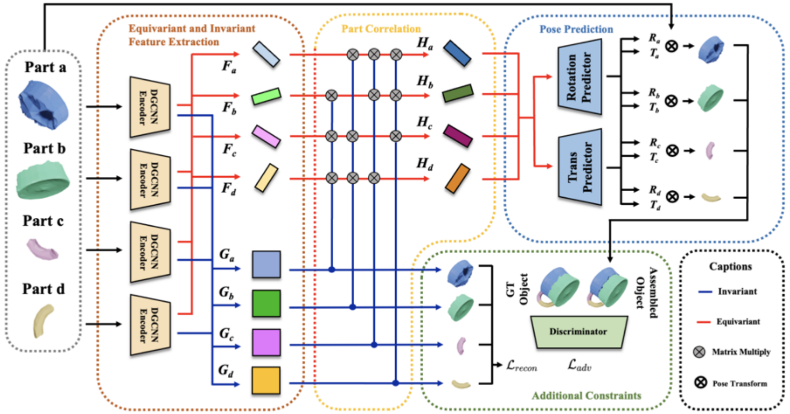
Scientific domain

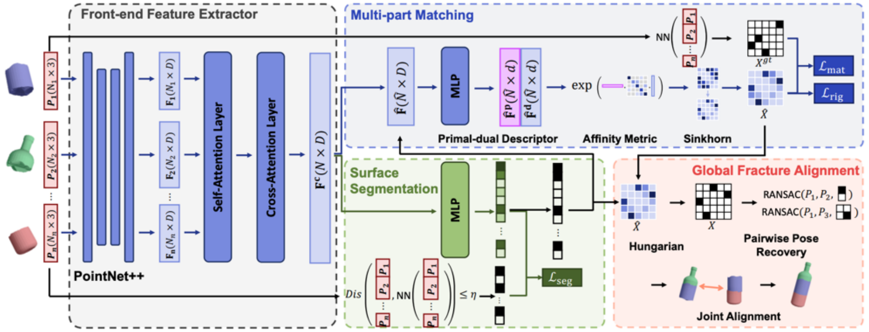
(e.g., Protein docking)

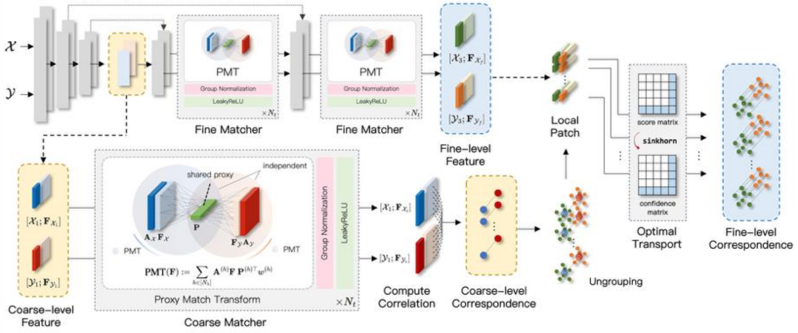


Robotics

State-of-the-art methods







Wu et al.

(Wu et al., ICCV 2023)

☺ Leveraging SE(3) equivariance for multi-part correlations

☹ Pose regression with limited spatial information

Jigsaw

(Lu et al., NeurIPS 2023)

☺ Effective matching between mating surfaces of parts

☹ Heavily rely on segmentation of mating surfaces

PMTR

(Lee et al., ICML 2024)

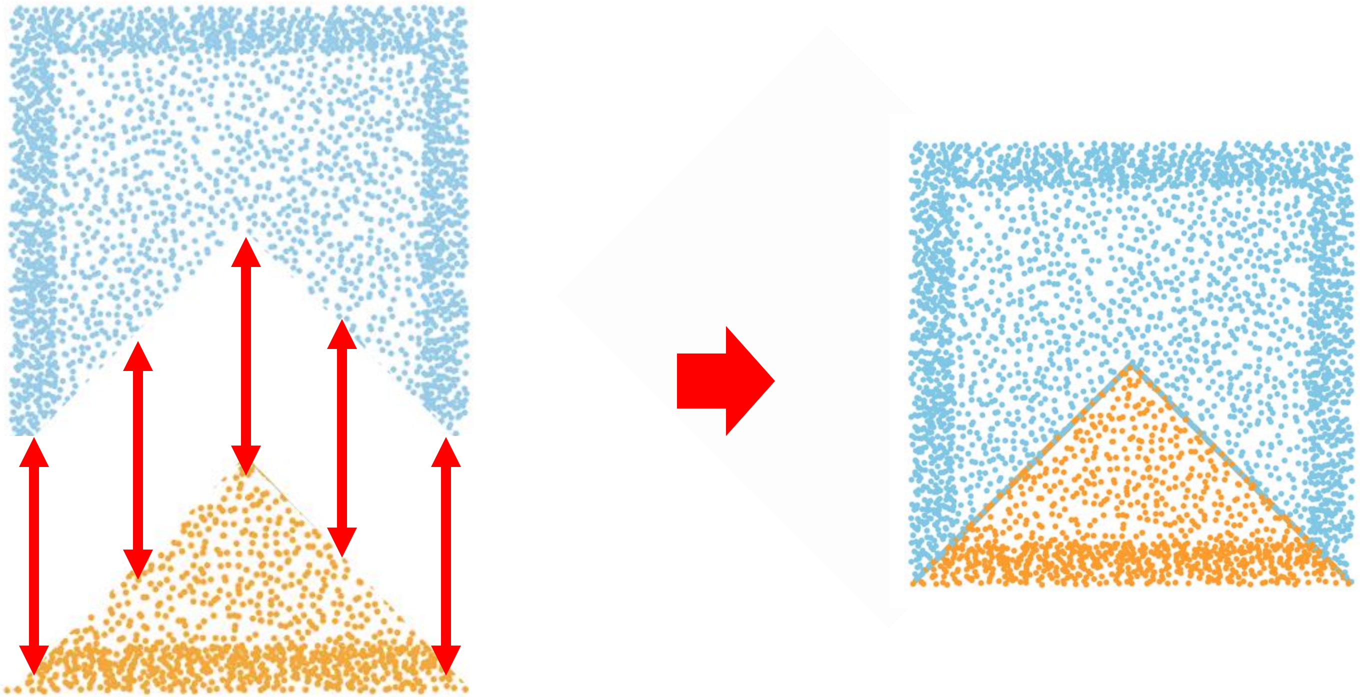
☺ Approximate high-order conv. for dense matching of parts

☹ Often fails to establish good correspondences between mating surfaces

Wu et al., "Leveraging SE(3) Equivariance for Learning 3D Geometric Shape Assembly", ICCV 2023
Lu et al., "Jigsaw: Learning to Assemble Multiple Fractured Objects", NeurIPS 2023
Lee et al., "3D Geometric Shape Assembly via Efficient Point Cloud Matching", ICML 2024

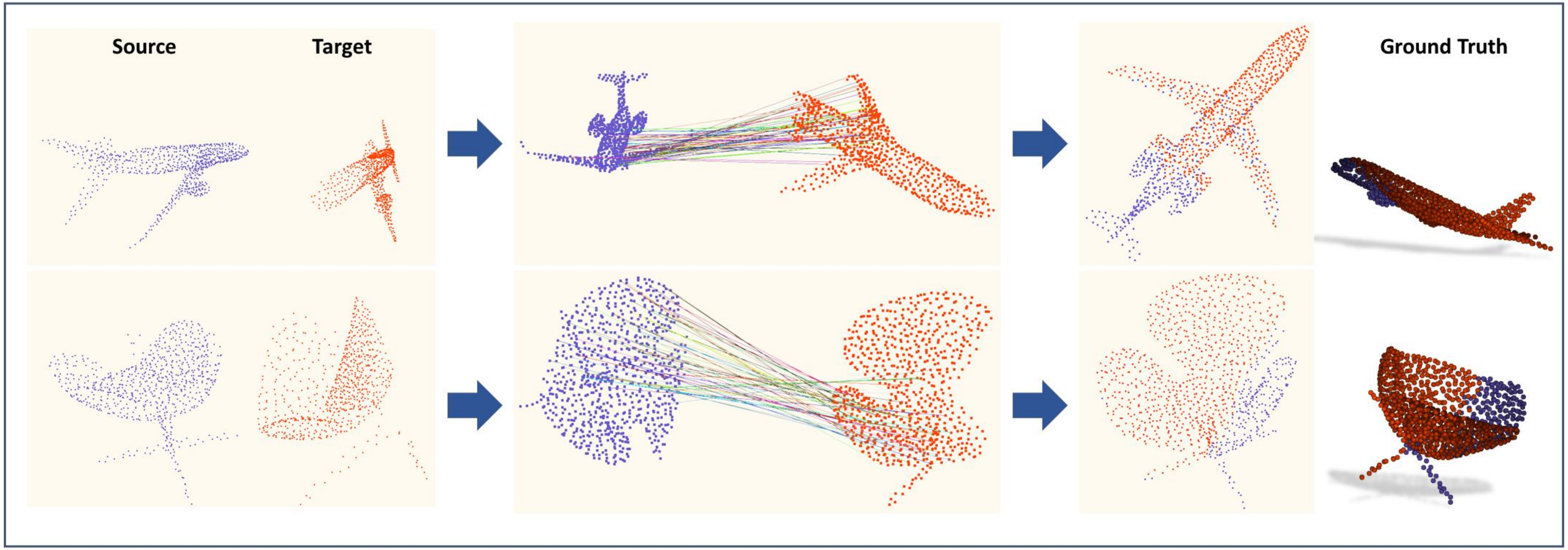
Previous Approach

Equative matching: establishing correspondences based on visual similarity alone *under the assumption of the same appearance between mating parts.*



Previous Approach

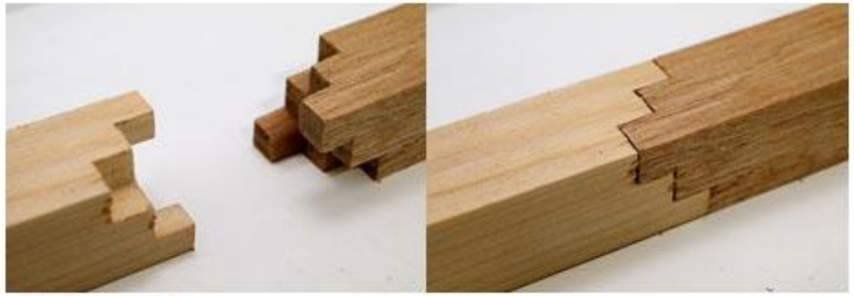
Equative matching: establishing correspondences based on visual similarity alone *under the assumption of the same appearance between mating parts.*



Point Cloud Registration

https://mvp-dataset.github.io/MVP/Registration.html

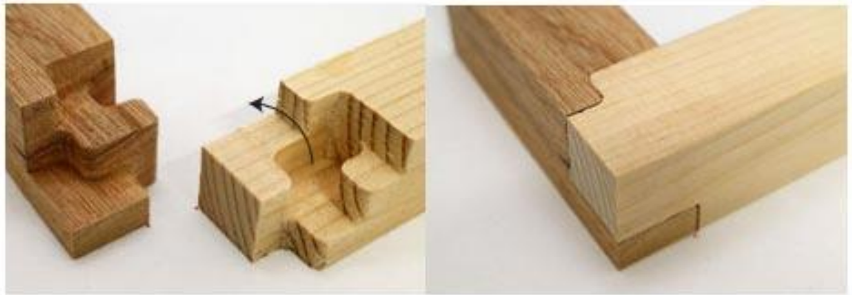
Beyond Equative Matching



a) I-axial

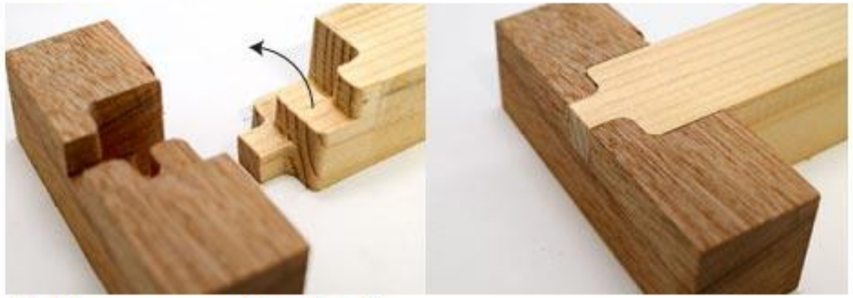
b) I-perpendicular*

c) L-axial



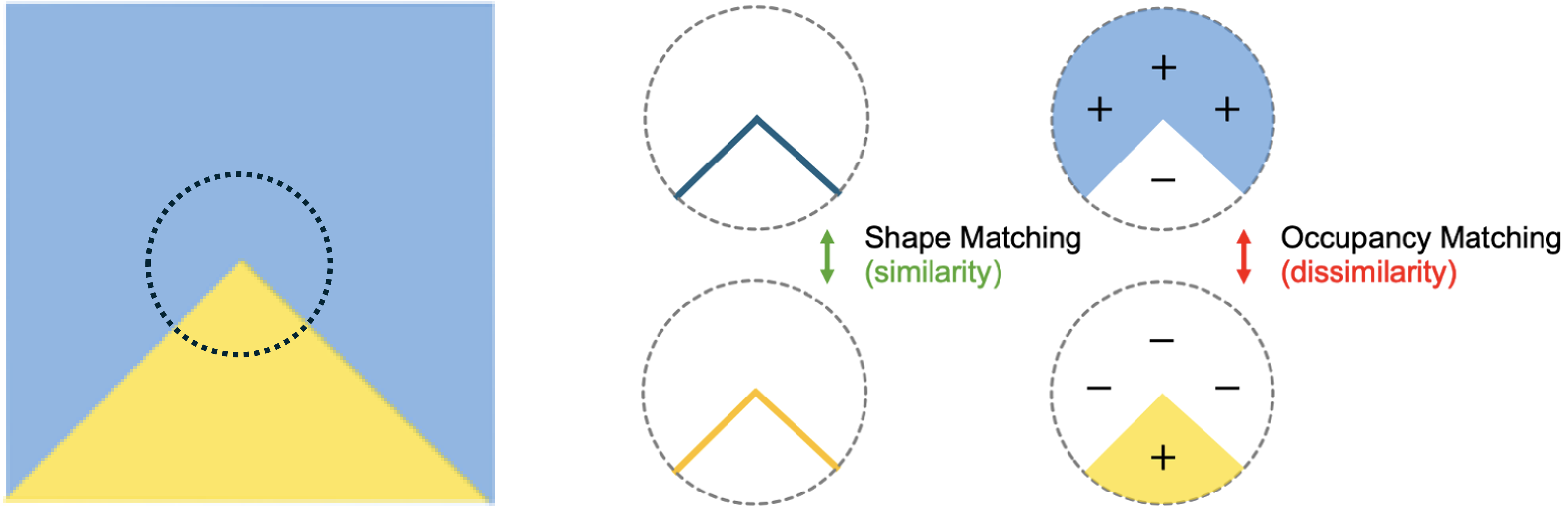
d) L-perpendicular*

e) T-axial

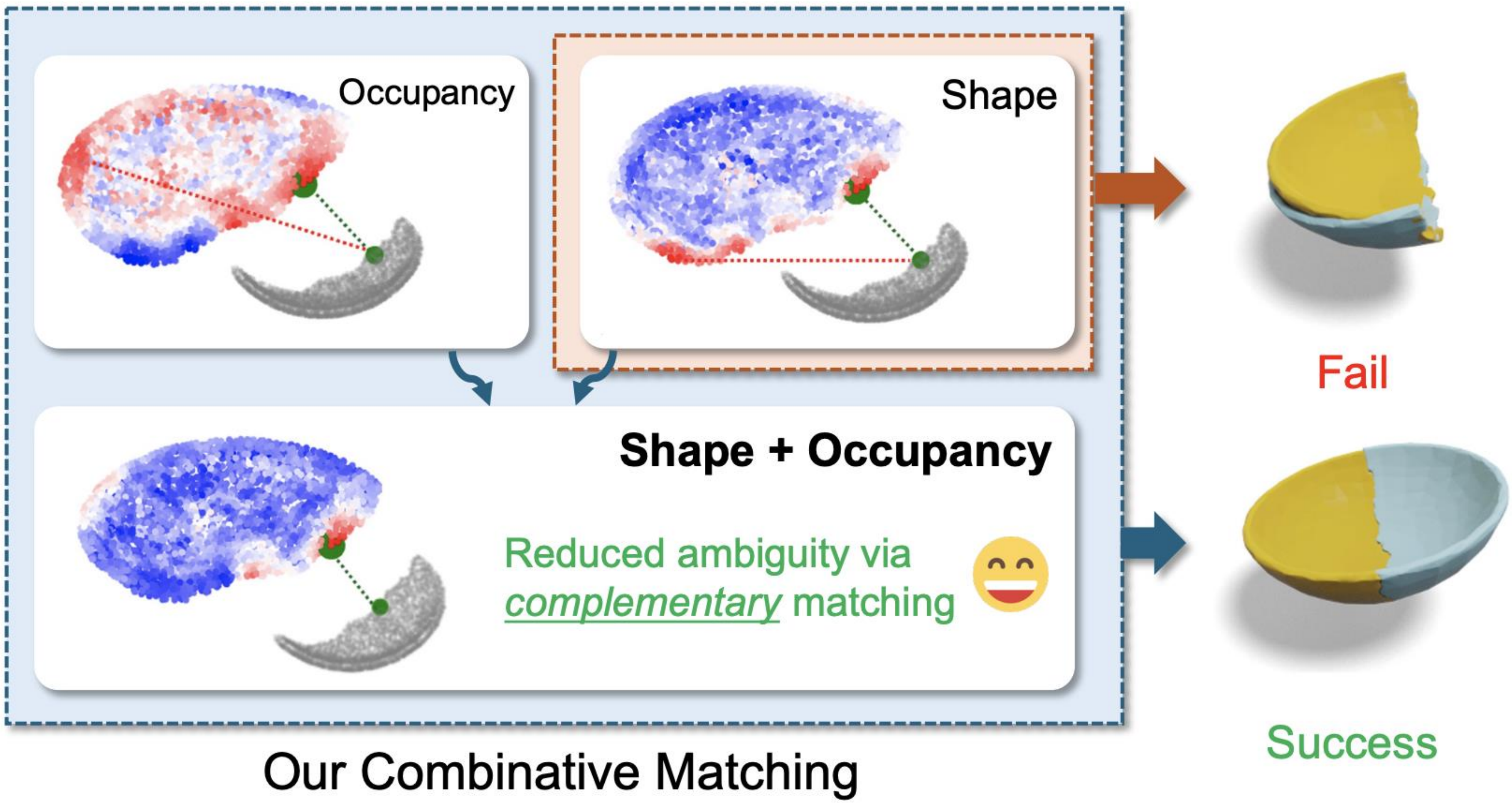


f) T-perpendicular*

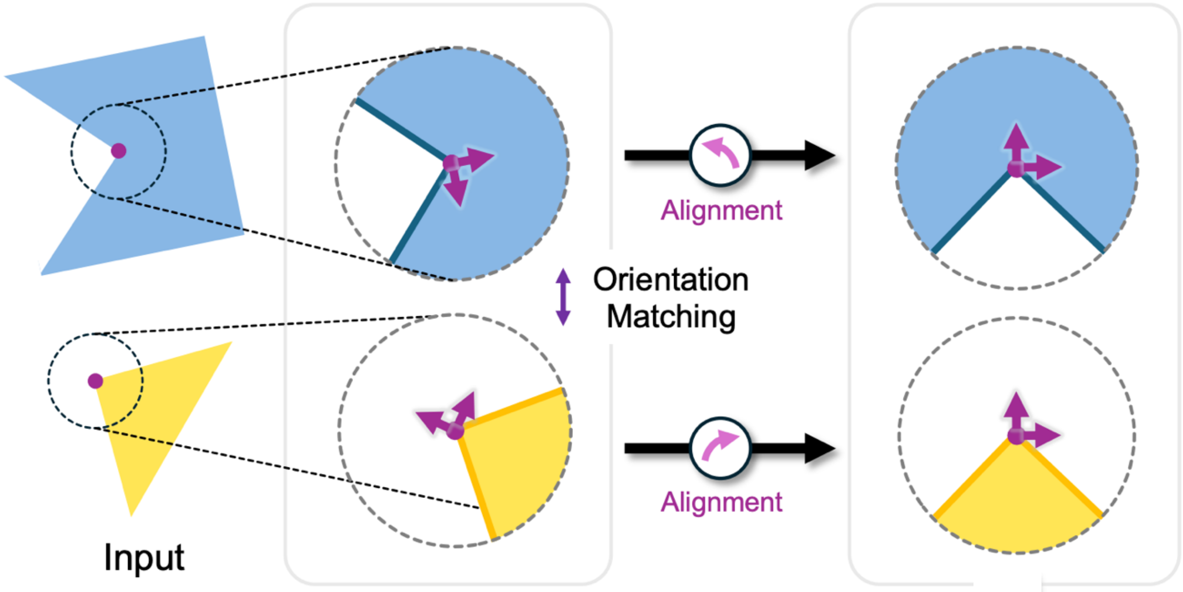
Beyond Equative Matching



Combinative Matching: 'combination' or 'joining' of elements form the basis for matching



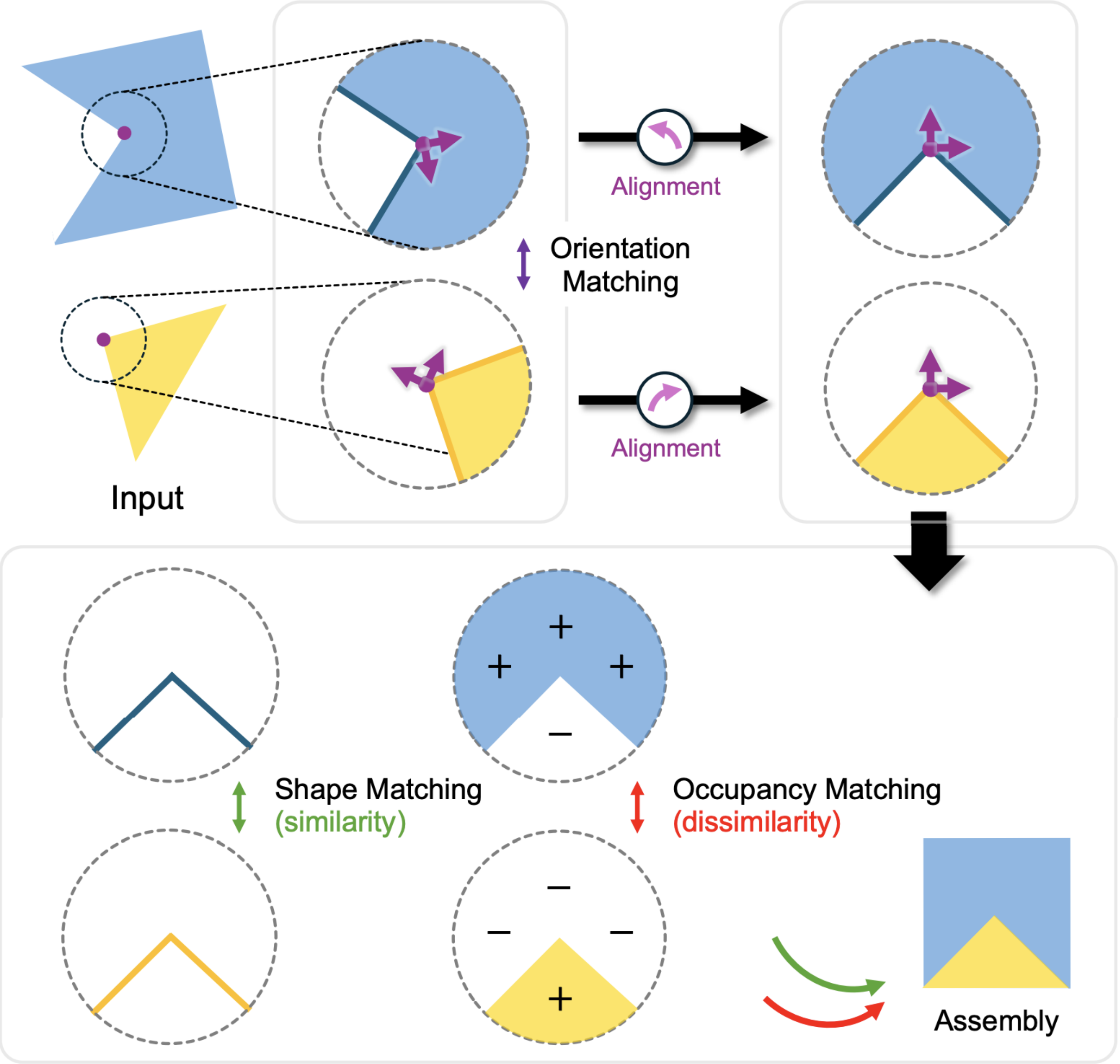
Objectives for Combinative Matching



$\mathcal{L}_{\mathrm{d}}$: maximize ‘**orientation**’ **consistency**

(in SO(3))

Objectives for Combinative Matching



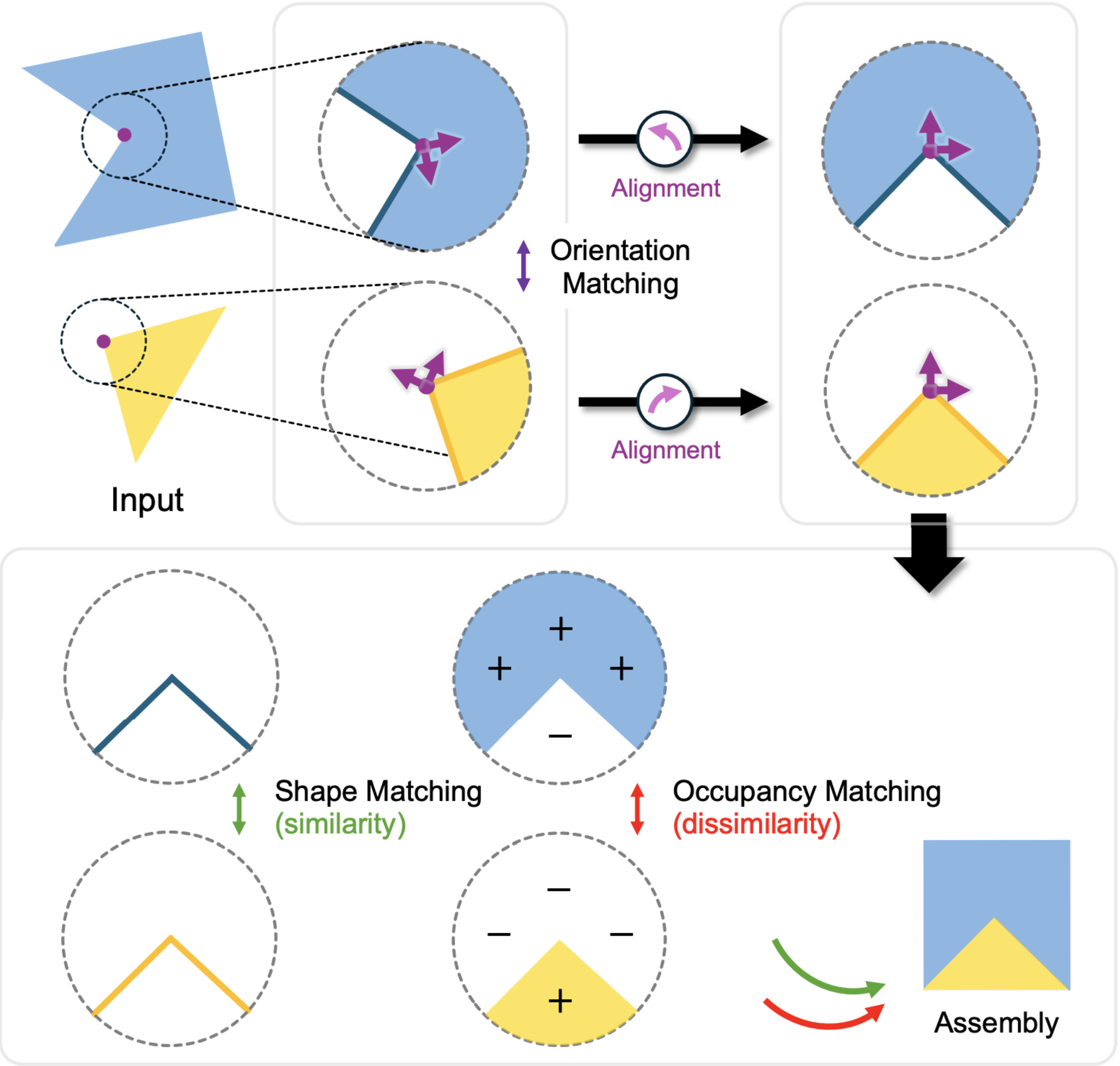
$\mathcal{L}_{\mathrm{d}}$: maximize '**orientation**' **consistency**

(in SO(3))

$\mathcal{L}_{\mathrm{s}}$: maximize '**visual**' feature **similarity**

$\mathcal{L}_{\mathrm{o}}$: maximize '**occupancy**' feature **dissimilarity**

Objectives for Combinative Matching



$\mathcal{L}_{\mathrm{d}}$: maximize '**orientation**' **consistency**

$$\mathcal{L}_{\mathrm{d}} = \frac{1}{|\mathcal{C}|} \sum_{(i,j)\in\mathcal{C}} \left\| (\mathbf{F}^{\mathrm{P}}_{\mathrm{d}})_i \, \mathbf{R}^{\mathrm{P}} - (\mathbf{F}^{\mathrm{Q}}_{\mathrm{d}})_j \, \mathbf{R}^{\mathrm{Q}} \right\|_F$$

$\mathcal{L}_{\mathrm{s}}$: maximize '**visual**' feature **similarity**

$$\mathcal{L}_{\mathrm{s}} = \mathop{\mathbb{E}}_{i\sim\mathcal{I}} \left[\log\left(1 + \sum_{j\in\mathcal{E}_{\mathrm{p}}(i)} e^{\alpha_{ij}(d^{\mathrm{p}}_{ij} - \Delta_{\mathrm{p}})} \cdot \sum_{k\in\mathcal{E}_{\mathrm{n}}(i)} e^{\beta_{ik}(\Delta_{\mathrm{n}} - d^{\mathrm{n}}_{ik})} \right) \right]$$

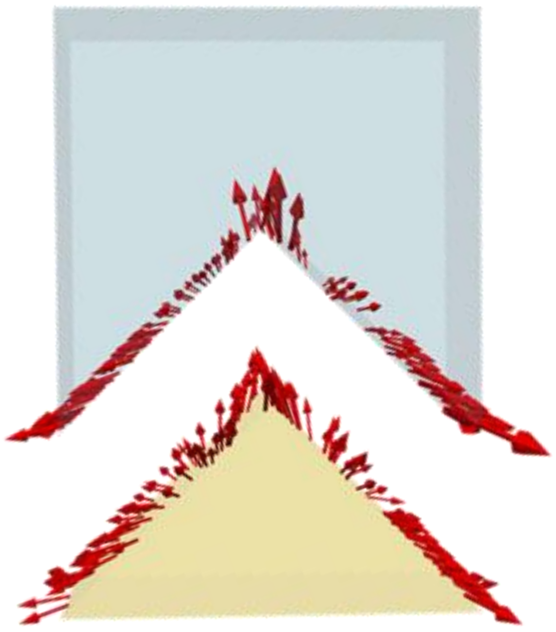
$$\alpha_{ij} = \gamma[d^{\mathrm{p}}_{ij} - \Delta_{\mathrm{p}}]_+, \quad \beta_{ik} = \gamma[\Delta_{\mathrm{n}} - d^{\mathrm{n}}_{ik}]_+ \qquad d^{\mathrm{p}}_{ij} = \|\hat{\mathbf{F}}^{\mathrm{P}}_{\mathrm{s},i} - \hat{\mathbf{F}}^{\mathrm{Q}}_{\mathrm{s},j}\|_2$$

$\mathcal{L}_{\mathrm{o}}$: maximize '**occupancy**' feature **dissimilarity**

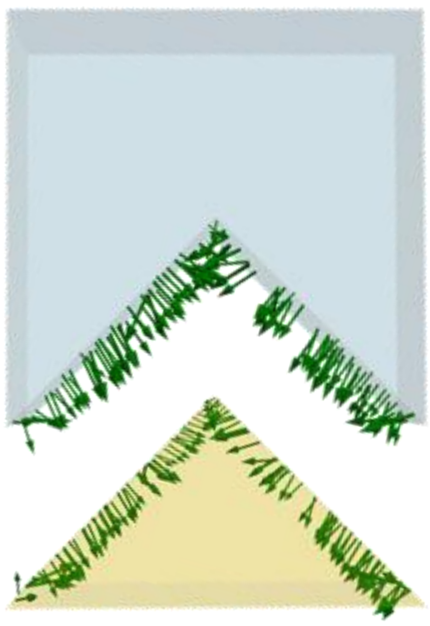
$$\mathcal{L}_{\mathrm{o}} = \mathop{\mathbb{E}}_{i\sim\mathcal{I}} \left[\log\left(1 + \sum_{j\in\mathcal{E}_{\mathrm{p}}(i)} e^{\alpha_{ij}(s^{\mathrm{p}}_{ij} - \Delta_{\mathrm{p}})} \cdot \sum_{k\in\mathcal{E}_{\mathrm{n}}(i)} e^{\beta_{ik}(\Delta_{\mathrm{n}} - s^{\mathrm{n}}_{ik})} \right) \right]$$

$$\alpha_{ij} = \gamma[s^{\mathrm{p}}_{ij} - \Delta_{\mathrm{p}}]_+, \quad \beta_{ik} = \gamma[\Delta_{\mathrm{n}} - s^{\mathrm{n}}_{ik}]_+ \qquad s^{\mathrm{p}}_{ij} \approx \cos\left(\mathbf{F}^{\mathrm{P}}_{\mathrm{o},i}, \mathbf{F}^{\mathrm{Q}}_{\mathrm{o},j}\right)$$

Learned orientation analysis



$\{\mathbf{x}_i\}_{i=1}^K$



$\{\mathbf{y}_i\}_{i=1}^K$

<u>Several notable patterns</u>

$\boldsymbol{x}_i$: directed toward the center of the surface

$\boldsymbol{x}_i$: parallel to the 2D plane of the interface

$\boldsymbol{y}_i$: pointing in/outward on concave/convex

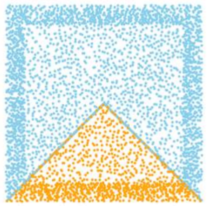
$\boldsymbol{y}_i$: magnitudes correlating with the degree of convexity/concavity

src and trg orientations are aligned in parallel (enforced by objective)

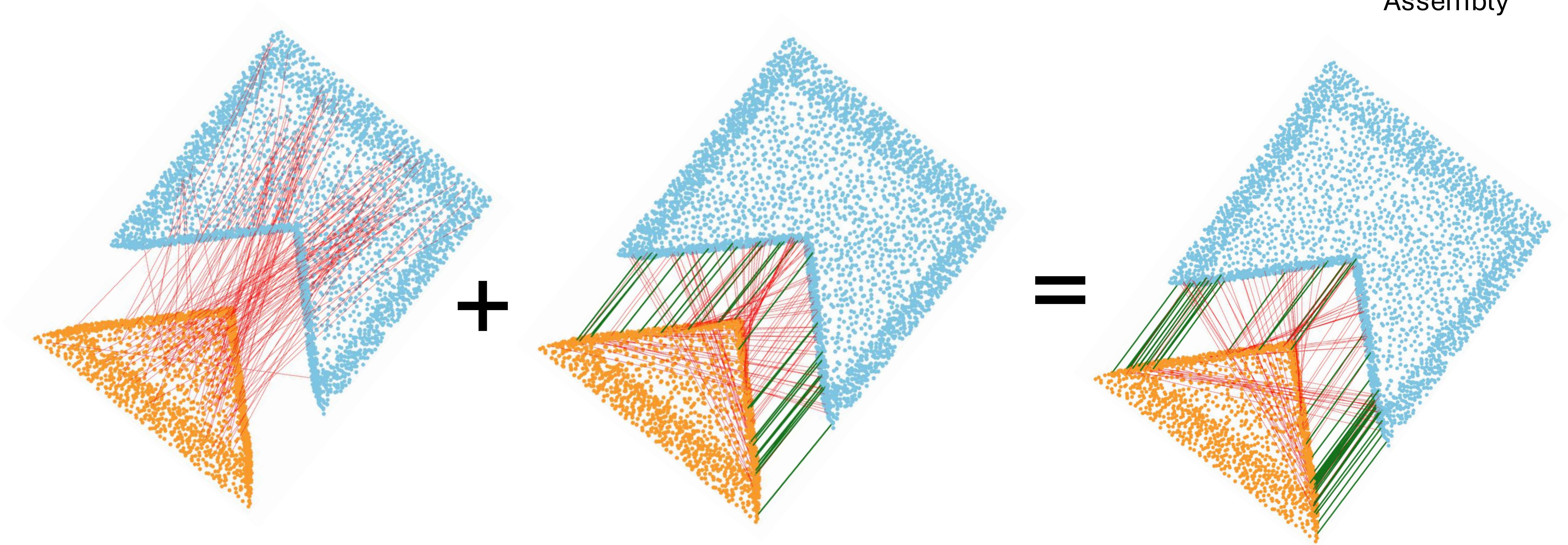
Learned orientations capture surface structures (convex, concave, curvature, 2D plane)
without any explicit supervisions dedicated to these aspects

Learned correlation analysis

Learning occupancy 'dissimilarity' effectively resolves local ambiguity

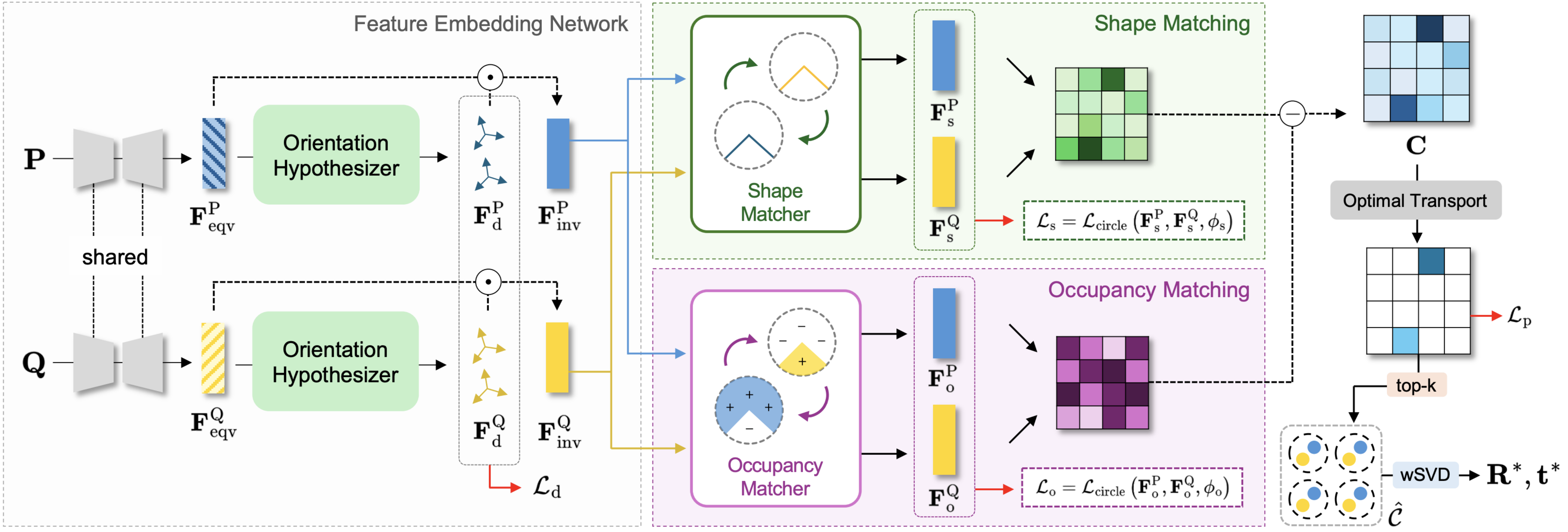


Assembly



Shape only   Occupancy only   Shape + Occupancy

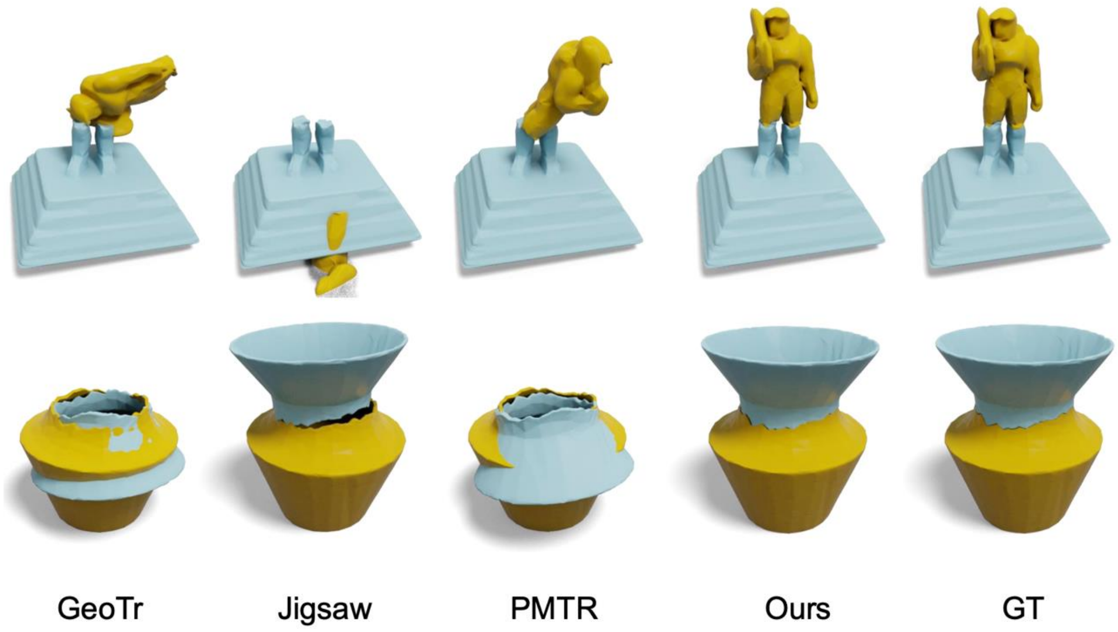
Network Architecture



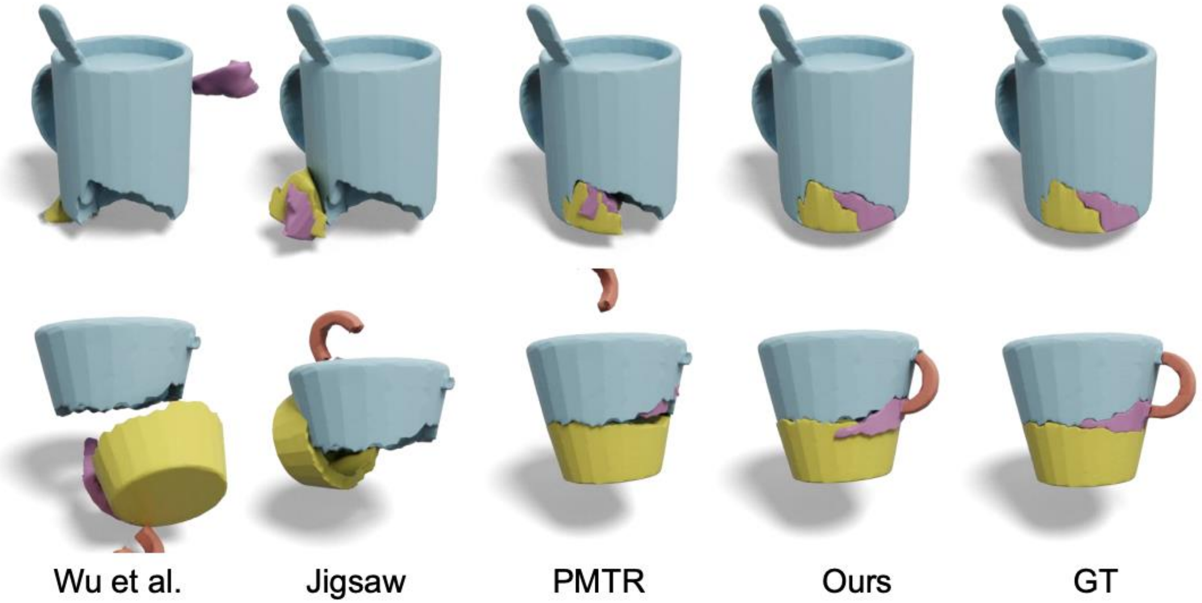
$$\mathcal{L} = \underbrace{\lambda_{\mathrm{s}}\mathcal{L}_{\mathrm{s}} + \lambda_{\mathrm{o}}\mathcal{L}_{\mathrm{o}} + \lambda_{\mathrm{d}}\mathcal{L}_{\mathrm{d}}}_{\textbf{Combinative matching objective}} + \mathcal{L}_{\mathrm{p}}$$

← Point matching loss

Qualitative Comparisons



Two-part assembly result



Multi-part assembly result

Comparison with SoTA

Method	CRD ↓ (10^{-2})	CD ↓ (10^{-3})	RMSE(R) ↓ (°)	RMSE(T) ↓ (10^{-2})
everyday				
NSM [5]	21.71	11.09	83.38	23.71
Wu et al. [48]	20.65	11.66	84.58	22.90
GeoTransformer [34]	0.61	0.51	22.81	7.28
Jigsaw [24]	5.48	1.34	38.73	**2.73**
PMTR [16]	0.39	0.25	17.14	5.53
CMNet (Ours)	**0.28**	**0.17**	**12.88**	3.78
artifact				
NSM [5]	19.44	6.33	83.22	21.41
Wu et al. [48]	19.17	7.97	85.04	20.90
GeoTransformer [34]	0.89	0.70	33.23	10.30
Jigsaw [24]	6.36	1.45	39.71	**3.02**
PMTR [16]	0.60	0.42	23.28	7.27
CMNet (Ours)	**0.49**	**0.34**	**18.77**	5.57

Pairwise Shape Assembly

Method	CRD ↓ (10^{-2})	CD ↓ (10^{-3})	RMSE(R) ↓ (°)	RMSE(T) ↓ (10^{-2})	PA_{CRD} ↑ (%)	PA_{CD} ↑ (%)
everyday						
Global [17, 37]	27.79	15.30	55.42	15.31	36.42	37.90
LSTM [47]	27.69	15.23	54.78	15.24	36.74	38.97
DGL [12]	27.90	13.23	55.76	15.33	36.99	39.70
Wu et al. [48]	28.18	19.70	54.98	15.59	35.66	36.28
Jigsaw [24]	14.13	11.82	41.12	11.74	52.48	60.26
PMTR [16]	6.51	5.56	31.57	9.95	66.95	70.56
CMNet (Ours)	**5.18**	**3.65**	**27.11**	**8.13**	**73.88**	**77.88**
artifact						
Global [17, 37]	26.42	14.92	54.41	14.48	36.67	36.97
LSTM [47]	28.15	14.61	53.59	15.49	36.67	37.25
DGL [12]	27.48	13.91	54.66	15.10	36.66	37.40
Wu et al. [48]	26.02	15.81	54.35	14.27	36.63	37.02
Jigsaw [24]	16.10	9.53	42.01	17.47	56.93	65.58
PMTR [16]	5.67	4.33	31.58	10.08	66.96	71.61
CMNet (Ours)	**4.56**	**3.04**	**29.21**	**8.99**	**71.02**	**76.32**

Multi-part Assembly (Volume-constrained ver.)

Method	RMSE(R) ↓ (°)	RMSE(T) ↓ (10^{-2})	PA_{CD} ↑ (%)	CD ↓ (10^{-3})
everyday				
Global [17, 37]	80.7	15.1	24.6	14.6
LSTM [47]	84.2	16.2	22.7	15.8
DGL [12]	79.4	15.0	31.0	14.3
Wu et al. [48]	79.3	16.9	8.41	28.5
DiffAssemble [36]	73.3	14.8	27.5	-
Jigsaw [24]	42.3	10.7	57.3	13.3
PuzzleFusion++ [46]	38.1	**8.0**	71.0	6.0
CMNet (Ours)	**32.0**	9.6	**77.3**	**3.5**
everyday → artifact				
Jigsaw [24]	52.4	22.2	45.6	14.3
PuzzleFusion++ [46]	52.1	**13.9**	49.6	14.5
CMNet (Ours)	**46.0**	14.3	**52.6**	**9.8**

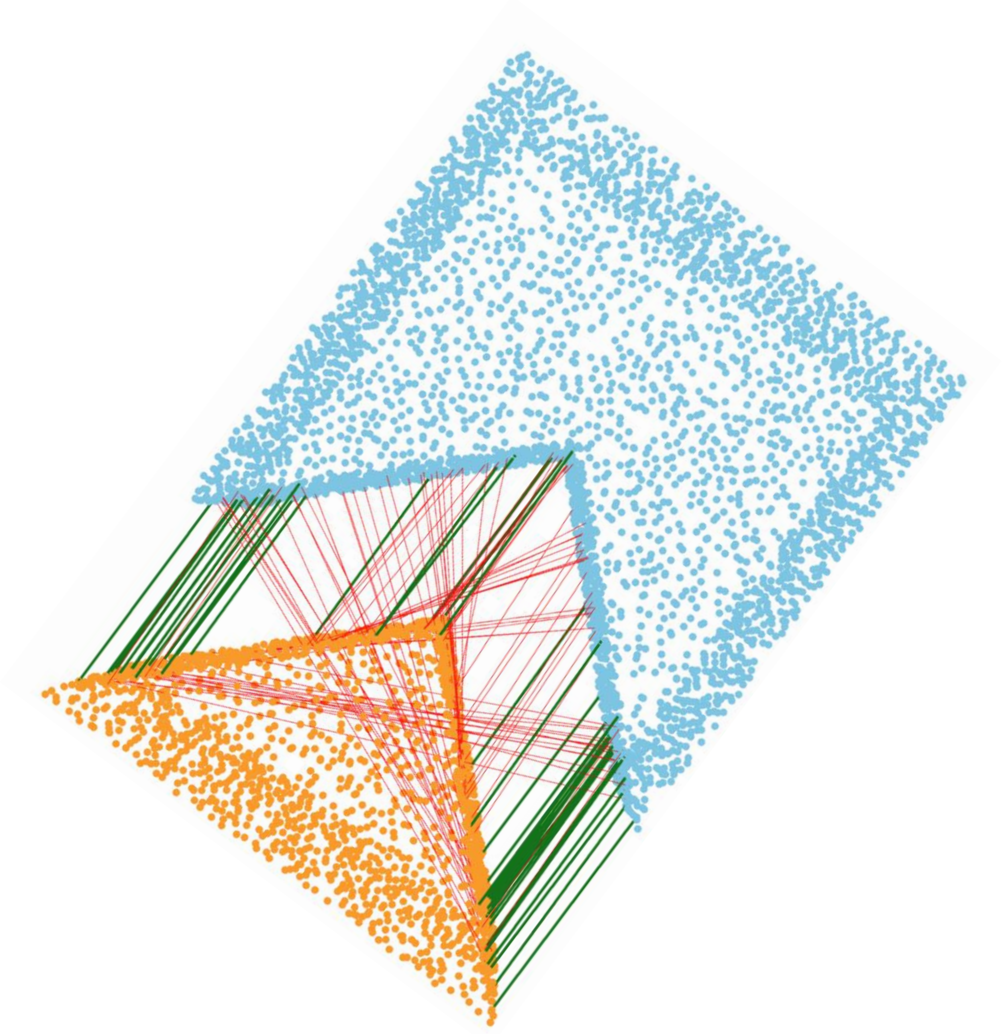
Multi-part Assembly (Vanilla ver.)

Summary

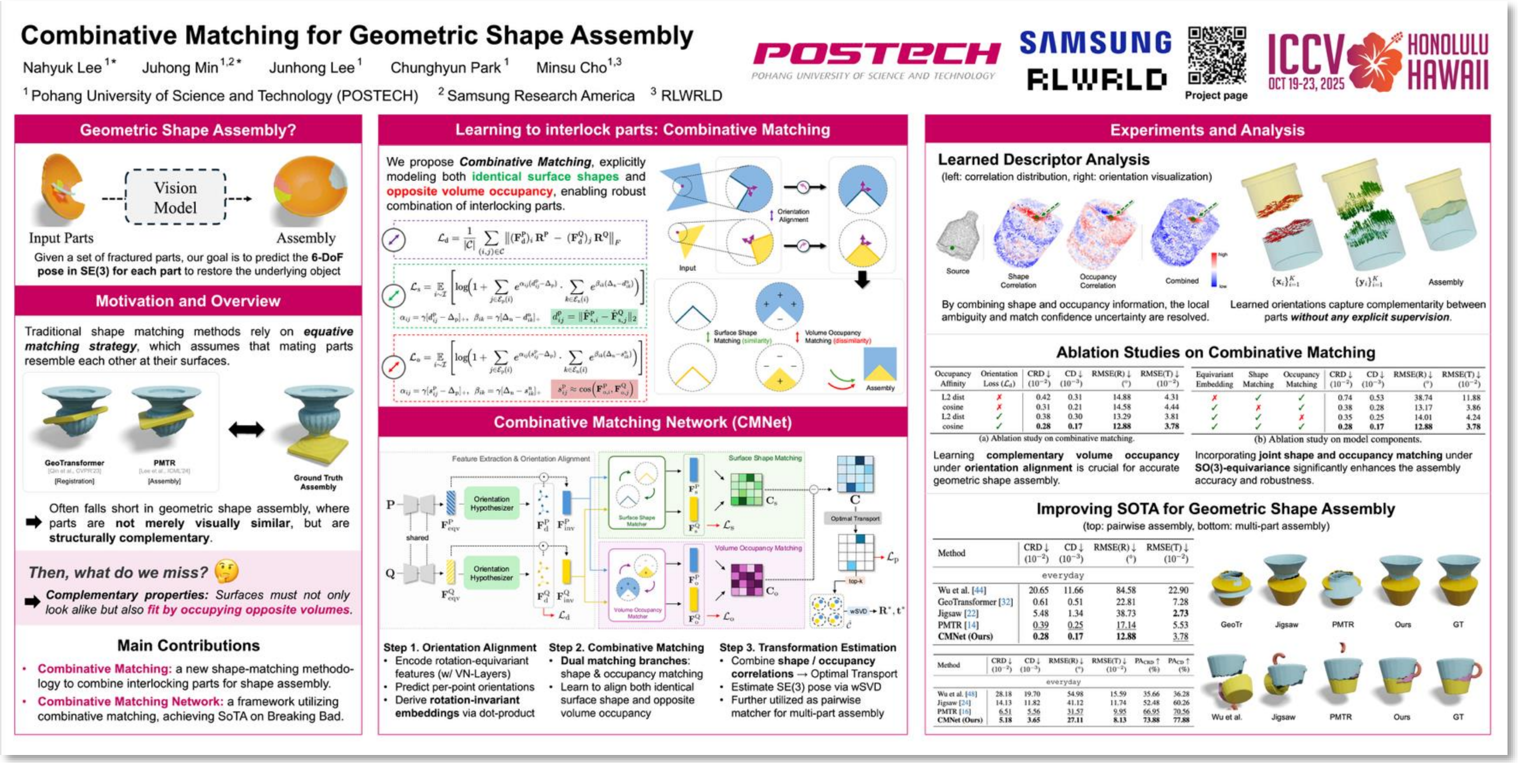
- Assembly as a more realistic and challenging generative task is key to interacting AI with reality, e.g., manufacturing and design.

- Leveraging **geometric equivariance and invariance** in a careful design is crucial for addressing the tasks in a realistic environment.

- **Combinative matching**, which considers both shape and occupancy relations in matching, significantly improves the performance of geometric assembly and potentially extends to other relational tasks.



See you in Hawaii!



OUR POSTER PRESENTATION

Tue 21 Oct, 2025
3:15 p.m. HST — 5:15 p.m. HST

@ Exhibit Hall I #885

Project page