Learning SO(3)-Invariant Semantic Correspondence via Local Shape Transform

Chunghyun Park[1]*, Seungwook Kim[1]*, Jaesik Park[2], and Minsu Cho[1] (*equal contribution)
CVPR 2024 · [1]POSTECH · [2]Seoul National University

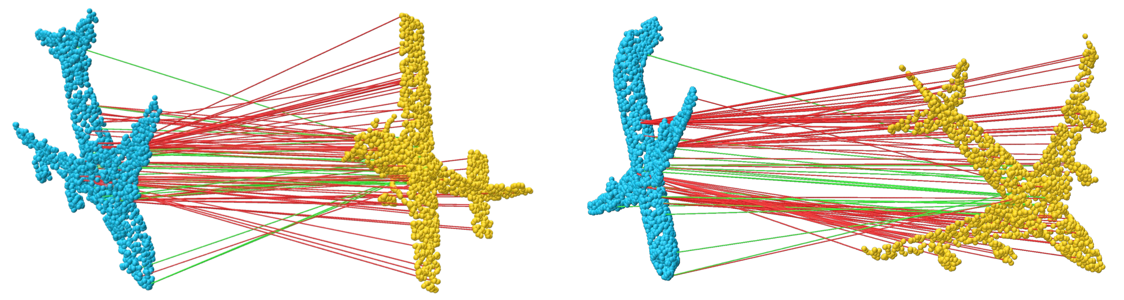
Website

3D Semantic Correspondence

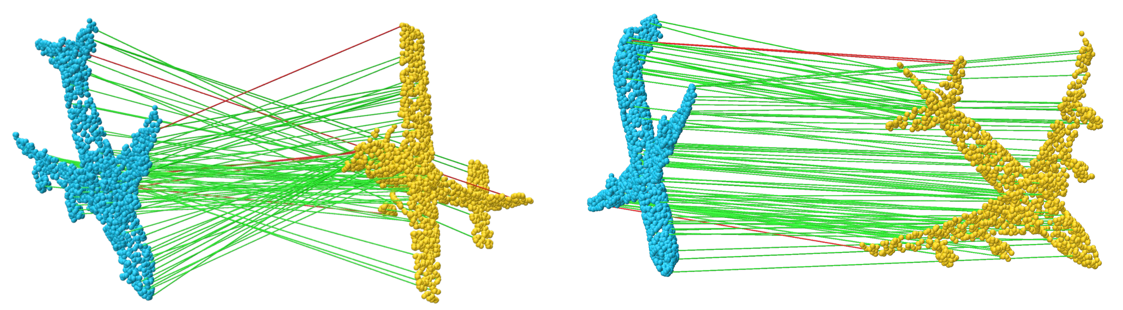
- **Problem Definition**
  Given two different shapes $\mathbf{P}_1 \in \mathbb{R}^{N\times 3}$ and $\mathbf{P}_2 \in \mathbb{R}^{N\times 3}$ of the same semantic category, find all semantically matching point pairs $\{\mathbf{p}_i, \mathbf{q}_i\}_{i=1}^{N'}$ such that $\mathbf{p}_i \in \mathbf{P}_1$ and $\mathbf{q}_i \in \mathbf{P}_2$.

  *$N' \leq N$; there could be points with no pairs.

- **Limitations of Previous Approaches**
  - Impractical assumption of aligned shapes
  - Fail to match rotated shapes even with rotation augmentation during training

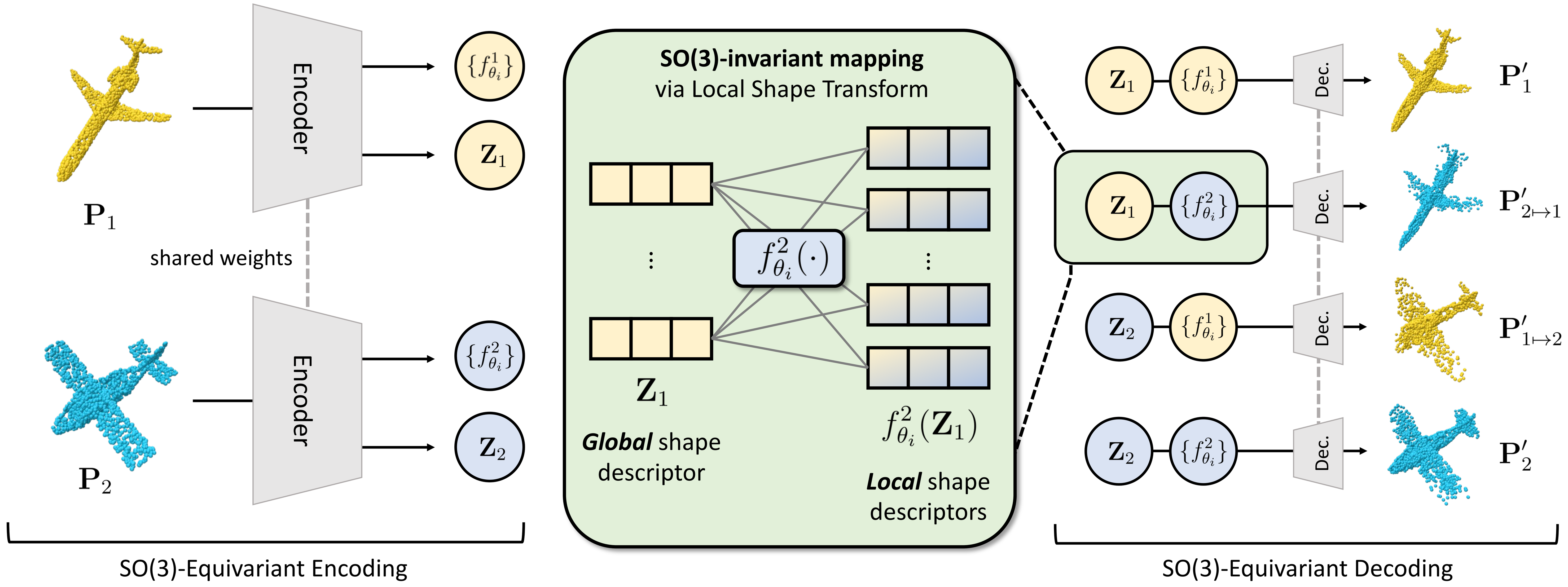


Our Contributions



- **Local Shape Transform**
  We formulate local shape information of each point as a novel function called ***local shape transform*** with dynamic input-dependent parameters.
- **Self-Supervised Learning for SO(3)-Invariant Semantic Correspondence**
  The proposed SO(3)-Invariant local shape transform enables a self-supervised approach for matching two rotated shapes.

RIST: Rotation-Invariant Local Shape Transform



- Self-reconstruction Loss: $\mathcal{L}_{\mathrm{SR}} = \lambda_{\mathrm{MSE}}\,\mathrm{MSE}(\mathbf{P},\mathbf{P}') + \lambda_{\mathrm{EMD}}\,\mathrm{EMD}(\mathbf{P},\mathbf{P}')$
- Cross-reconstruction Loss: $\mathcal{L}_{\mathrm{CR}} = \lambda_{\mathrm{CD}}\,\mathrm{CD}(\mathbf{P}_1,\mathbf{P}'_{2\mapsto 1})$

Self-supervised training of **RIST** for 3D semantic correspondence!

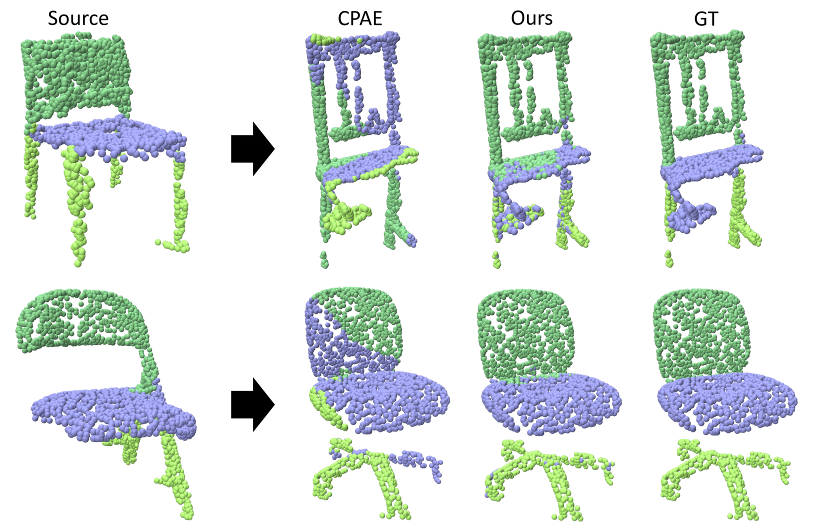
Experiments – I. Part Label Transfer

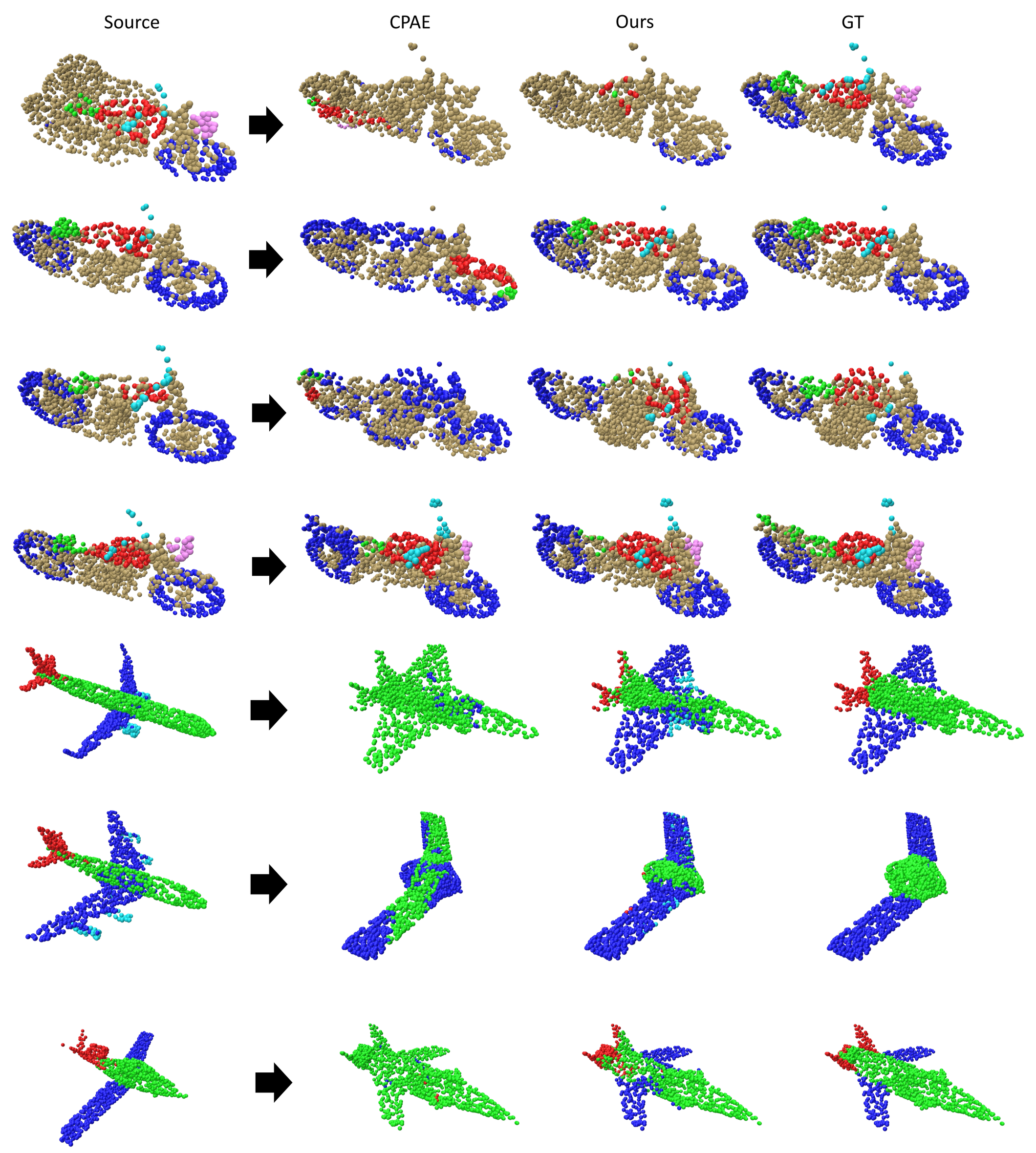
- ShapeNet

Training	Method	Airplane	Cap	Chair	Guitar	Laptop	Motorcycle	Mug	Table	Average
w/o Rotations	FoldingNet	17.8	34.7	22.5	22.1	36.2	12.6	50.0	34.6	28.8
	AtlasNetV2	19.7	31.4	23.6	22.7	36.0	13.1	49.7	35.2	28.9
	DPC	22.7	37.1	25.6	31.9	35.0	17.5	51.3	36.8	32.2
	CPAE	21.0	38.0	26.0	22.7	34.9	14.7	51.4	35.5	30.5
	RIST (ours)	**52.1**	**54.5**	**58.3**	**74.1**	**56.5**	**48.6**	**75.0**	**41.3**	**57.6**
w/ Rotations	FoldingNet	22.5	33.2	24.0	31.0	35.9	13.5	49.9	37.0	30.9
	AtlasNetV2	21.1	32.7	25.2	28.8	35.5	14.5	49.9	41.0	31.1
	DPC	24.6	38.5	25.6	40.2	34.9	19.3	51.8	37.3	34.0
	CPAE	17.0	36.6	24.5	39.4	37.4	15.8	51.9	36.7	32.4
	RIST (ours)	**51.2**	**57.0**	**55.0**	**73.5**	**60.6**	**48.5**	**72.2**	**44.4**	**57.8**

- ScanObjectNN

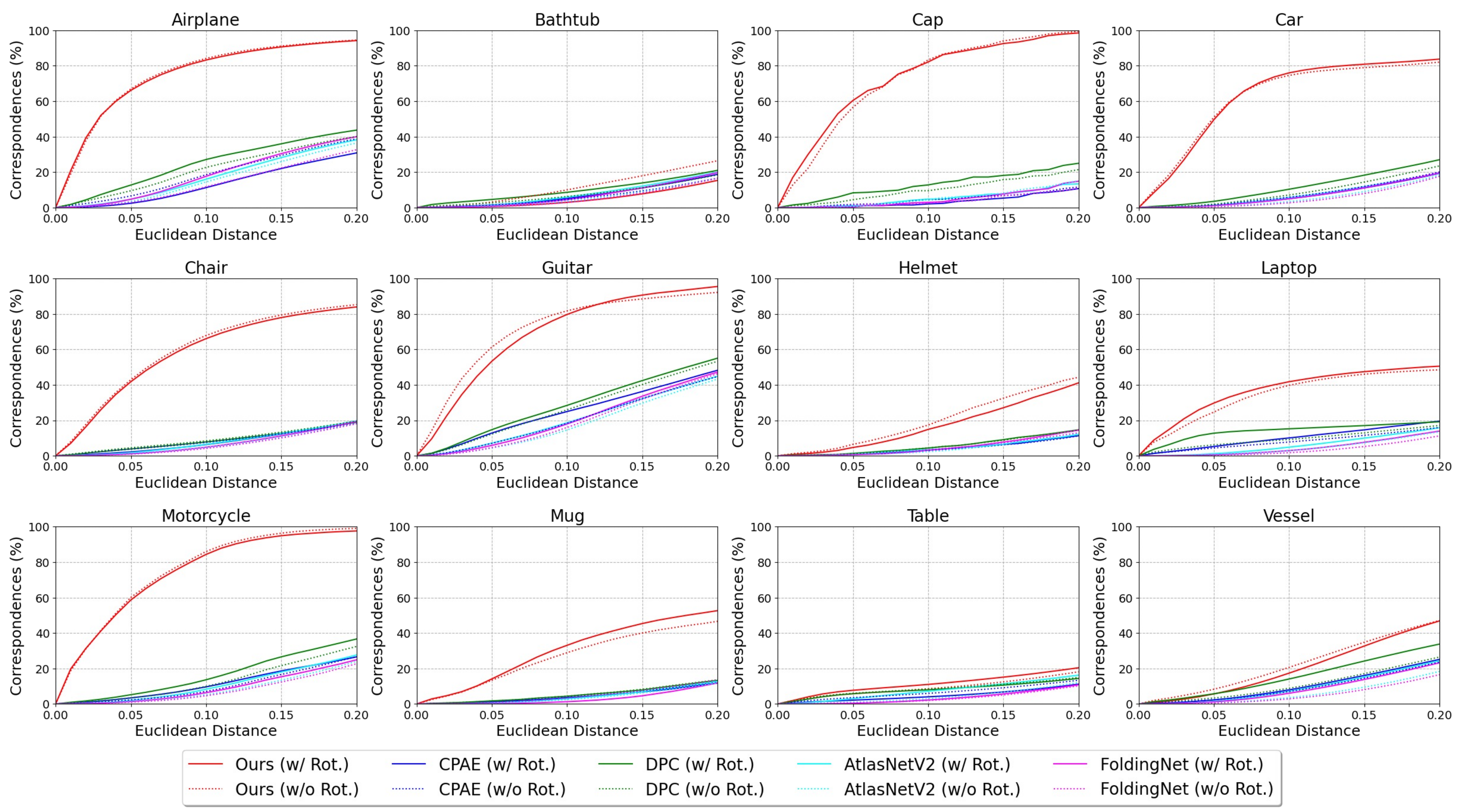
Method	w/o Rotations	w/ Rotations
FoldingNet	23.2	23.3
AtlasNetV2	23.6	24.1
DPC	23.9	23.9
CPAE	24.4	23.9
RIST (ours)	**39.6**	**37.9**

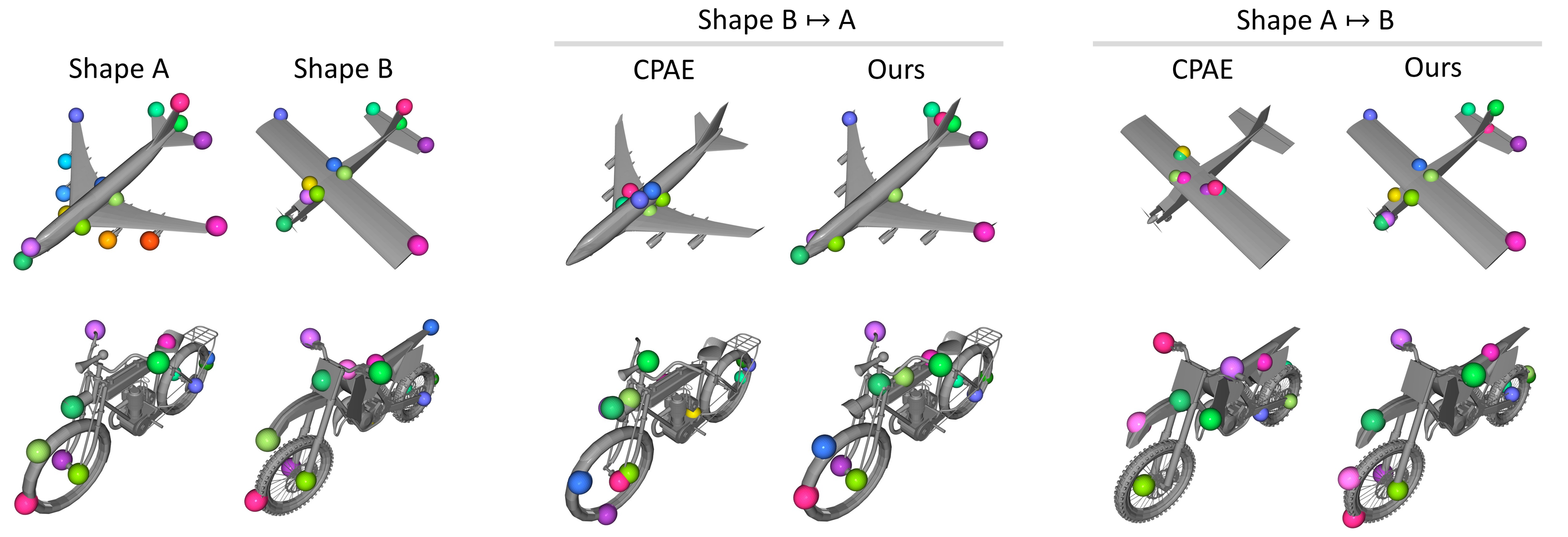




Experiments – II. Keypoint Transfer

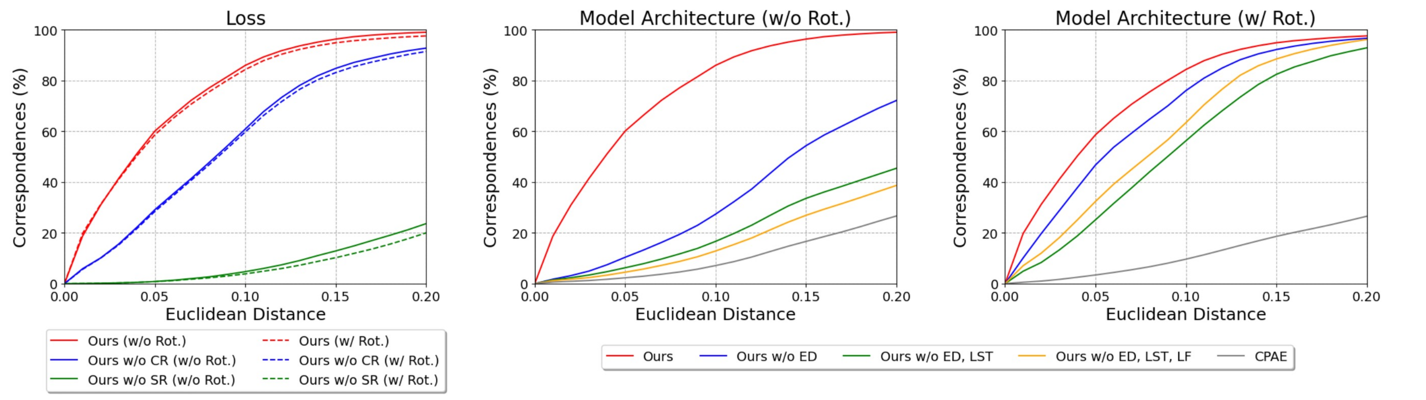
- Quantitative & Qualitative Results on KeypointNet





- Ablation Study on Losses and Model Components

SR: Self-reconstruction
CR: Cross-reconstruction
ED: Equivariant Decoder
LST: Local Shape Transform
LF: Local Feature



Acknowledgement

This work was supported by IITP grants (No.2021-0-02068: AI Innovation Hub (50%), No. 2022-0-00290: Visual Intelligence for Space-Time Understanding and Generation (40%), No.2019-0-01906: Artificial Intelligence Graduate School Program at POSTECH (5%), NO.2021-0-01343, Artificial Intelligence Graduate School Program at Seoul National University (5%)) funded by the Korea government (MSIT). Seungwook was supported by the Hyundai-Motor Chung Mong-koo Foundation.