September 22, 2022

Deep Closest Point: Learning Representations for Point Cloud Registration

Yue Wang & Justin M. Solomon, 2019

Mia Fornes
Norwegian University of Science and Technology

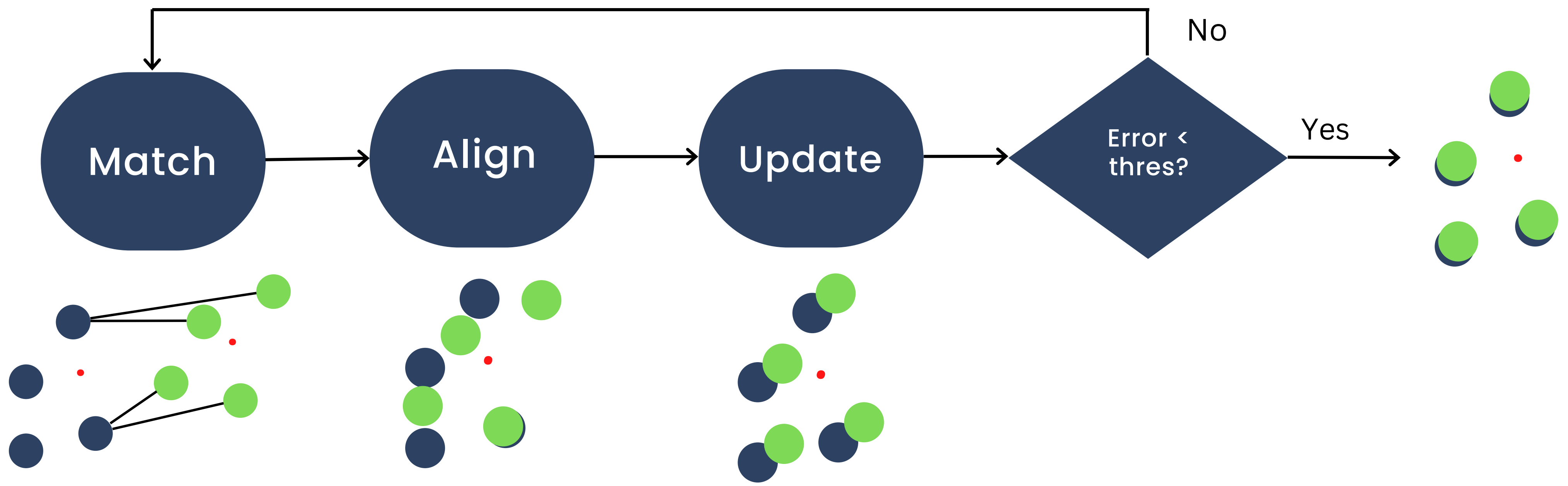
Agenda

1. Introduction
2. DCP
3. Experiments and Results
4. Ablation Study
5. Conclusion

Introduction

Motivation

- ICP and its variants
  - Simple, easily-implemented **iterative** methods
  - Local minima problem (classic ICP)



$$\mathcal{X} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_i, \dots, \boldsymbol{x}_N\} \subset \mathbb{R}^3$$

$$\mathcal{Y} = \{\boldsymbol{y}_1, \dots, \boldsymbol{y}_j, \dots, \boldsymbol{y}_M\} \subset \mathbb{R}^3$$

$$[\boldsymbol{R}_{\mathcal{XY}}, \boldsymbol{t}_{\mathcal{XY}}]$$

$$E(\boldsymbol{R}_{\mathcal{XY}}, \boldsymbol{t}_{\mathcal{XY}}) = \frac{1}{N}\sum_i^N \|\boldsymbol{R}_{\mathcal{XY}}\boldsymbol{x}_i + \boldsymbol{t}_{\mathcal{XY}} - \boldsymbol{y}_i\|^2.$$

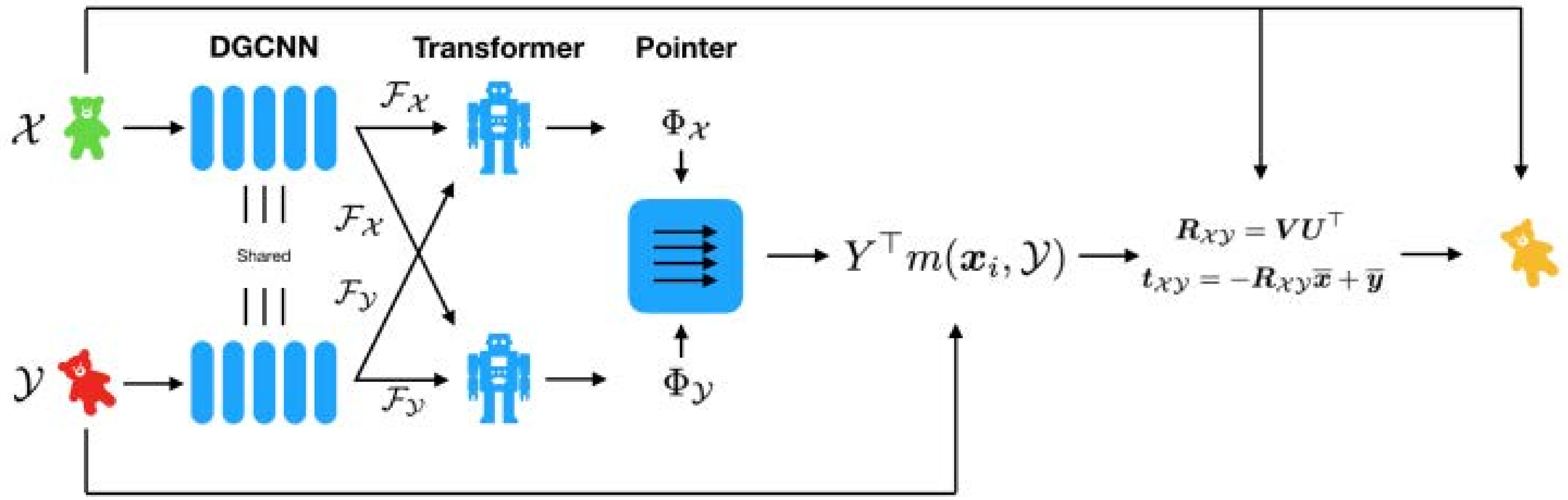
Motivation

- Wang and Solomon
  - Deep Learn perspective
  - Inspiration from computer vision, natural language techniques
  - Deep Closest Point (DCP)
    - Address ICP issues

DCP

Pipeline

- Three modules
  - Point cloud embedding network
    - Matching pairs of points
  - Attention-based module combining pointer network
    - Correspondences, two point clouds
  - Singular value decomposition (SVD) layer
    - Transformation

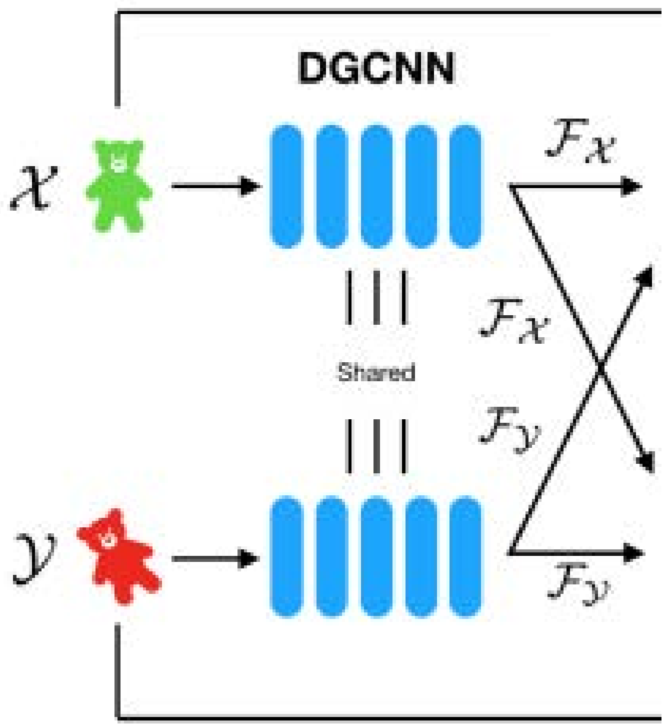


Pipeline - Point Cloud Embedding Network

- Unaligned input clouds -> find matching pairs of points
- PointNet
  - Independently
- Dynamic Graph CNN (DGCNN)
  - Local neighborhood

$$\boldsymbol{x}_i^l = f(\{h_\theta^l(\boldsymbol{x}_i^{l-1}, \boldsymbol{x}_j^{l-1})\ \forall j \in \mathcal{N}_i\}),$$

$$N_i = neighbor\ of\ vertex\ i$$



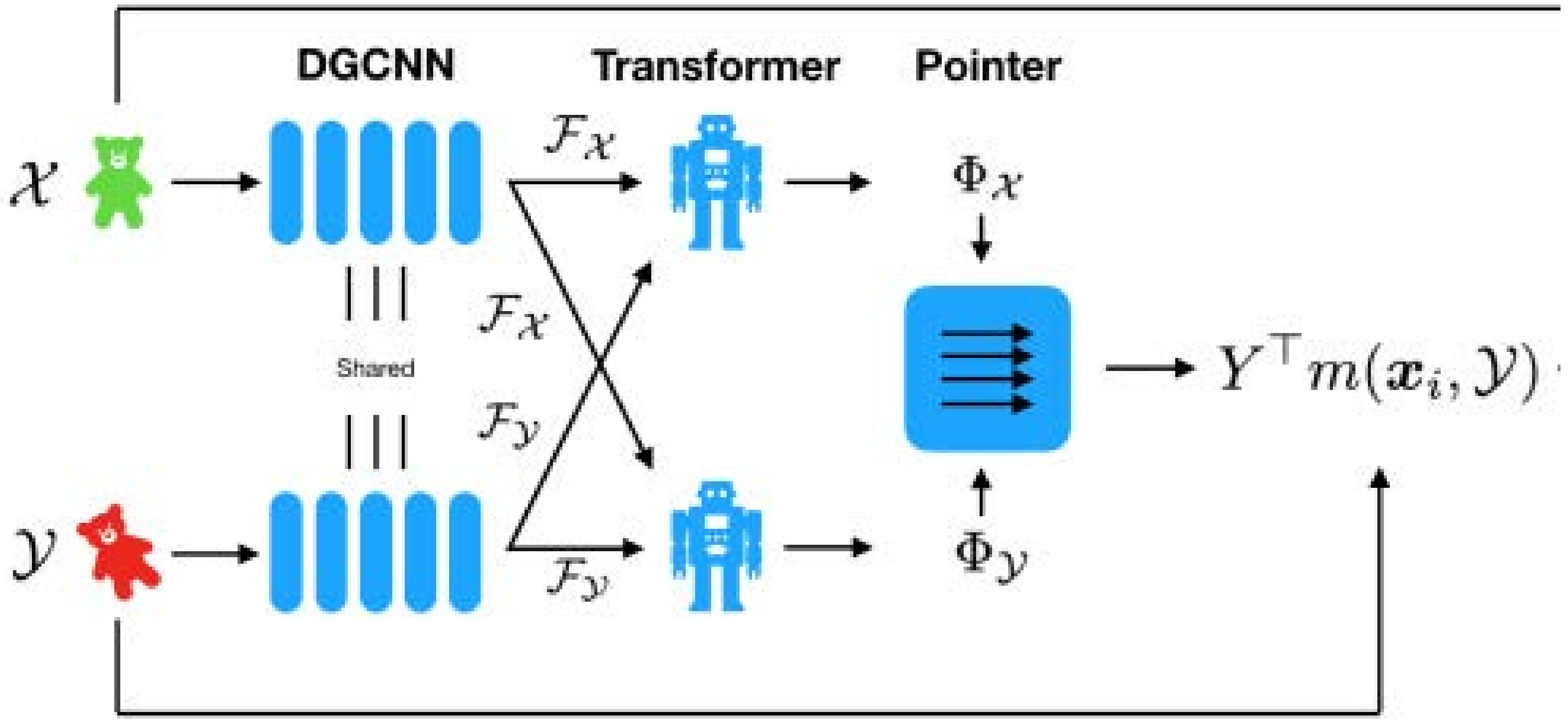
Pipeline - Transformer & Pointer

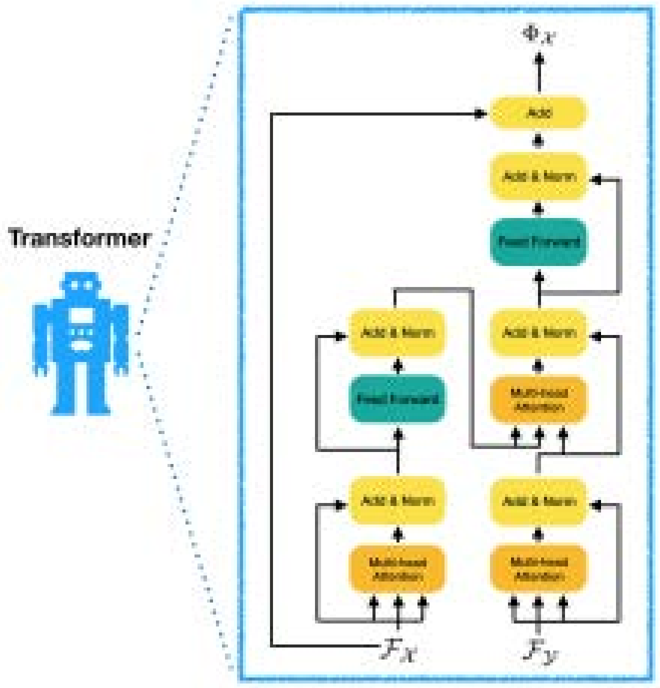
- Attention (Transformer)
  - Improve matching features
  - Learn co-contextual information
  - Encoder
    - Self-attention, shared MLP
  - Decoder
    - Encoder, co-attention
- Pointer Generation
  - Probabilistic approach

$$\Phi_{\mathcal{X}} = \mathcal{F}_{\mathcal{X}} + \phi(\mathcal{F}_{\mathcal{X}}, \mathcal{F}_{\mathcal{Y}})$$
$$\Phi_{\mathcal{Y}} = \mathcal{F}_{\mathcal{Y}} + \phi(\mathcal{F}_{\mathcal{Y}}, \mathcal{F}_{\mathcal{X}})$$

$$\Phi_{\mathcal{X}} = \mathcal{F}_{\mathcal{X}}$$
$$\Phi_{\mathcal{Y}} = \mathcal{F}_{\mathcal{Y}}$$

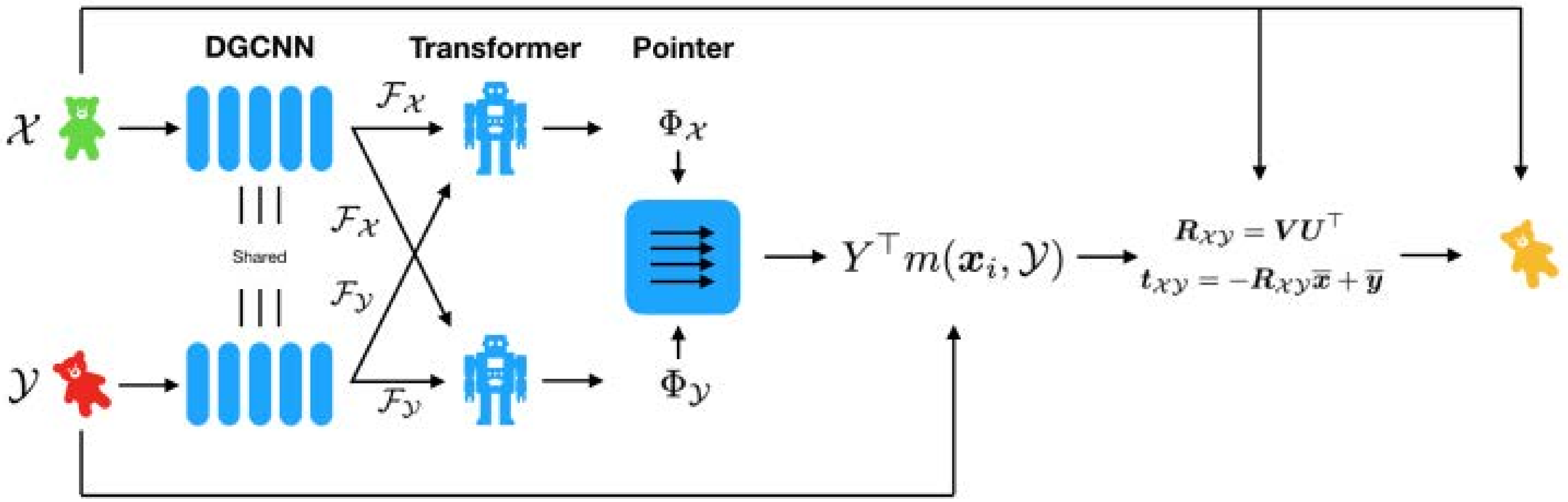
$$m(\boldsymbol{x}_i, \mathcal{Y}) = \mathrm{softmax}(\Phi_{\mathcal{Y}} \Phi_{\boldsymbol{x}_i}^{\top}).$$





Pipeline - SVD

- From soft match to rigid motion
  - $[R_{xy}, t_{xy}]$



Pipeline - Loss

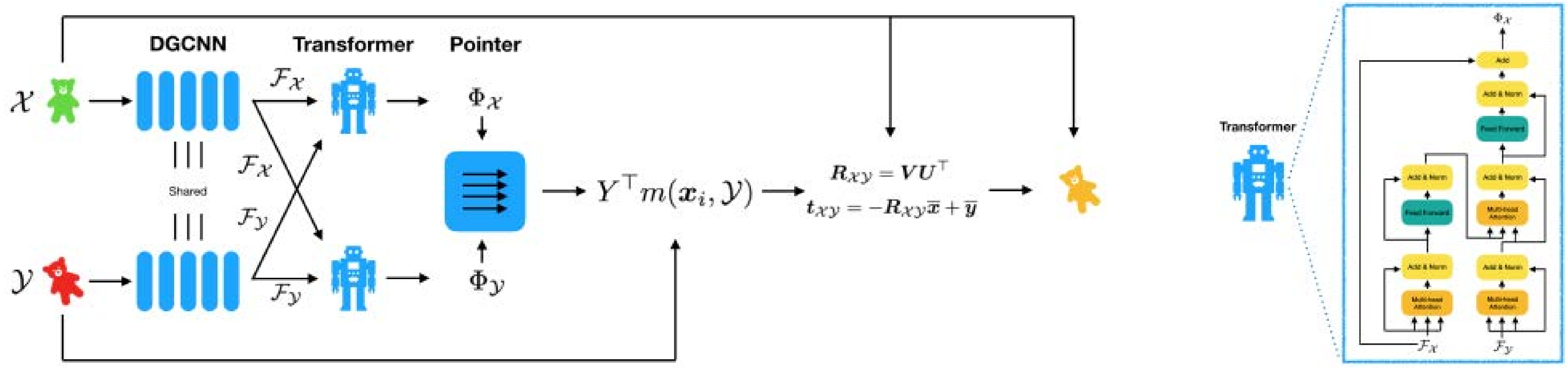
- DCGNN and attention module
  - Parameterized by NN weights

$$\text{Loss} = \underbrace{\|\boldsymbol{R}_{\mathcal{XY}}^{\top}\boldsymbol{R}_{\mathcal{XY}}^{g} - I\|^2 + \|\boldsymbol{t}_{\mathcal{XY}} - \boldsymbol{t}_{\mathcal{XY}}^{g}\|^2}_{\text{Deviation of } [R_{xy},\, t_{xy}] \text{ from GT}} + \underbrace{\lambda\|\theta\|^2}_{\text{Tikhonov regularization}}$$

Experiments and Results

Background

- Two models
  - DCP-v1 & DCP-v2
- ICP, Go-ICP, Fast Global Registration (FGR) and PointNetLK
- ModelNet40
- MSE, RMSE, MAE



Experiment 1

- Full training and test set
- No knowledge of the category label
- Test on unseen point clouds

Model	MSE($\boldsymbol{R}$)	RMSE($\boldsymbol{R}$)	MAE($\boldsymbol{R}$)	MSE($\boldsymbol{t}$)	RMSE($\boldsymbol{t}$)	MAE($\boldsymbol{t}$)
ICP	894.897339	29.914835	23.544817	0.084643	0.290935	0.248755
Go-ICP [53]	140.477325	11.852313	2.588463	0.000659	0.025665	0.007092
FGR [57]	87.661491	9.362772	1.999290	0.000194	0.013939	0.002839
PointNetLK [16]	227.870331	15.095374	4.225304	0.000487	0.022065	0.005404
DCP-v1 (ours)	6.480572	2.545697	1.505548	**0.000003**	**0.001763**	**0.001451**
DCP-v2 (ours)	**1.307329**	**1.143385**	**0.770573**	**0.000003**	0.001786	**0.001195**

Experiment 2

- Generalizability
- Dataset split evenly by category
- Test on unseen categories

Model	MSE($\boldsymbol{R}$)	RMSE($\boldsymbol{R}$)	MAE($\boldsymbol{R}$)	MSE($\boldsymbol{t}$)	RMSE($\boldsymbol{t}$)	MAE($\boldsymbol{t}$)
ICP	892.601135	29.876431	23.626110	0.086005	0.293266	0.251916
Go-ICP [53]	192.258636	13.865736	2.914169	0.000491	0.022154	0.006219
FGR [57]	97.002747	9.848997	**1.445460**	0.000182	0.013503	0.002231
PointNetLK [16]	306.323975	17.502113	5.280545	0.000784	0.028007	0.007203
DCP-v1 (ours)	19.201385	4.381938	2.680408	**0.000025**	**0.004950**	**0.003597**
DCP-v2 (ours)	**9.923701**	**3.150191**	2.007210	**0.000025**	0.005039	0.003703

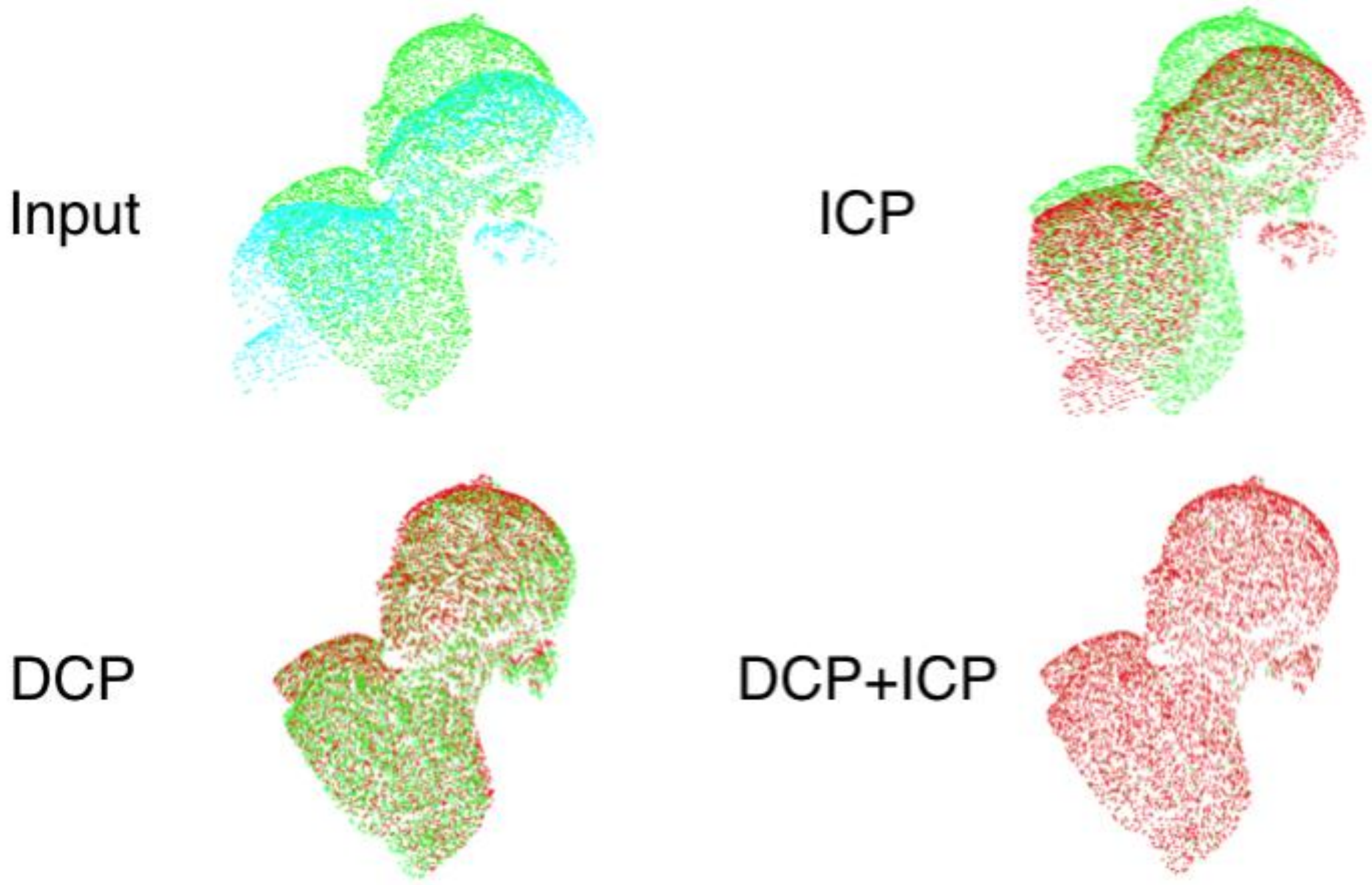
Experiment 3

- DCP-v1
- Added noise on test data

Model	MSE($\boldsymbol{R}$)	RMSE($\boldsymbol{R}$)	MAE($\boldsymbol{R}$)	MSE($\boldsymbol{t}$)	RMSE($\boldsymbol{t}$)	MAE($\boldsymbol{t}$)
ICP	882.564209	29.707983	23.557217	0.084537	0.290752	0.249092
Go-ICP [53]	131.182495	11.453493	2.534873	0.000531	0.023051	0.004192
FGR [57]	607.694885	24.651468	10.055918	0.011876	0.108977	0.027393
PointNetLK [16]	256.155548	16.004860	4.595617	0.000465	0.021558	0.005652
DCP-v1 (ours)	6.926589	2.631841	1.515879	0.000003	0.001801	0.001697
DCP-v2 (ours)	**1.169384**	**1.081380**	**0.737479**	**0.000002**	**0.001500**	**0.001053**

Experiment 4

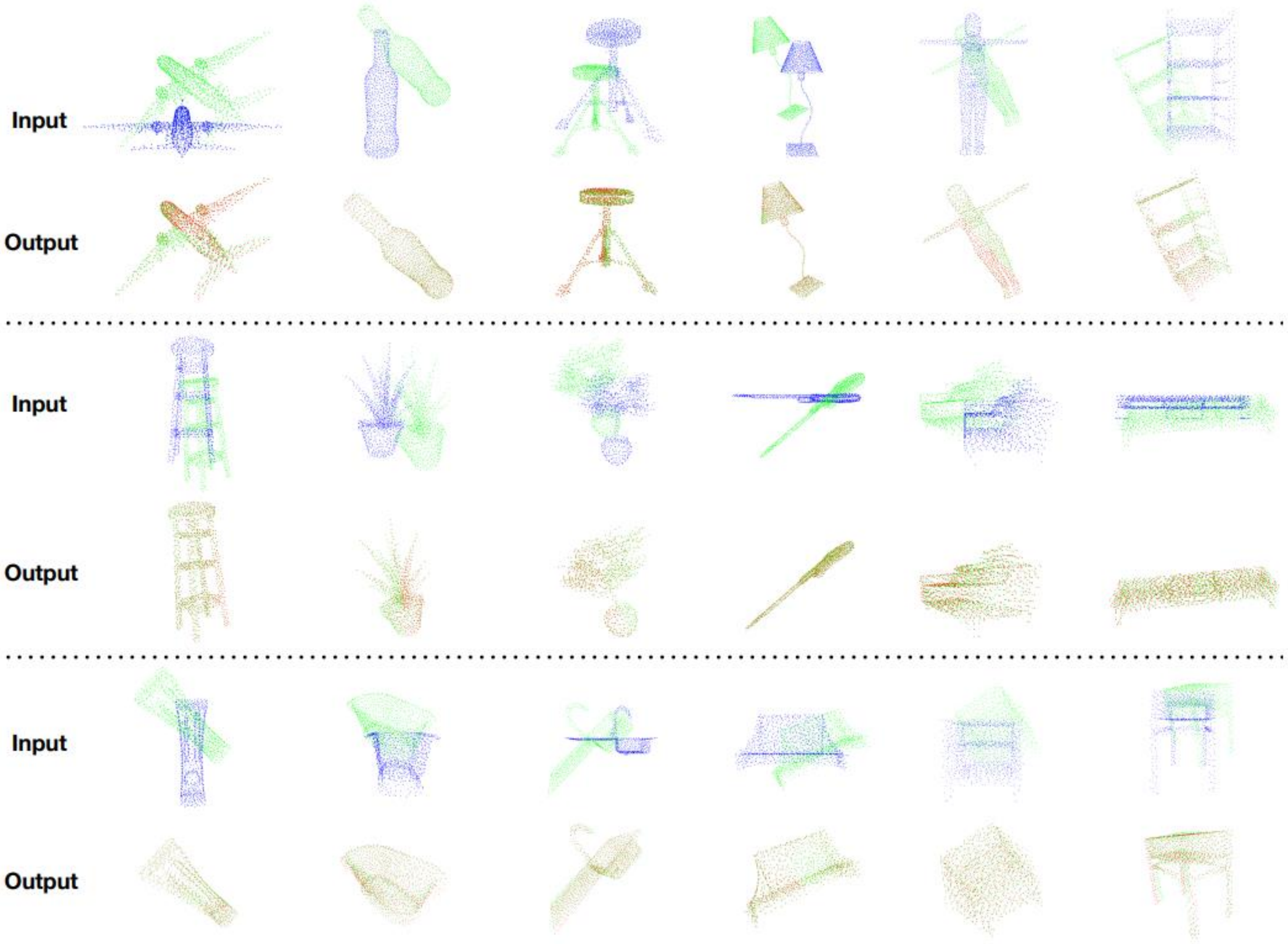
- DCP followed by ICP
  - Good initial guess for ICP



Experiment 5

- Inference time (seconds)
- Average of 100 results

# points	ICP	Go-ICP	FGR	PointNetLK	DCP-v1	DCP-v2
512	0.003972	15.012375	0.033297	0.043228	0.003197	0.007932
1024	0.004683	15.405995	0.088199	0.055630	0.003300	0.008295
2048	0.044634	15.766001	0.138076	0.146121	0.040397	0.073697
4096	0.044585	15.984596	0.157124	0.162007	0.039984	0.74263



https://arxiv.org/pdf/1905.03304.pdf

Ablation Study

PointNet vs. DGCNN

Metrics	PN+DCP-v1,	DGCNN+DCP-v1	PN+DCP-v2	DGCNN+DCP-v2
MSE($\boldsymbol{R}$)	17.008427	6.480572	49.863022	1.307329
RMSE($\boldsymbol{R}$)	4.124127	2.545697	7.061375	1.143385
MAE($\boldsymbol{R}$)	2.800184	1.505548	4.485052	0.770573
MSE($\boldsymbol{t}$)	0.000697	0.000003	0.000258	0.000003
RMSE($\boldsymbol{t}$)	0.026409	0.001763	0.016051	0.001786
MAE($\boldsymbol{t}$)	0.01327	0.001451	0.010546	0.001195

MLP vs. SVD

Metrics	DCP-v1+MLP	DCP-v1+SVD	DCP-v2+MLP	DCP-v2+SVD
MSE($\boldsymbol{R}$)	21.115917	6.480572	9.923701	1.307329
RMSE($\boldsymbol{R}$)	4.595206	2.545697	3.150191	1.143385
MAE($\boldsymbol{R}$)	3.291298	1.505548	2.007210	0.770573
MSE($\boldsymbol{t}$)	0.000861	0.000003	0.000025	0.000003
RMSE($\boldsymbol{t}$)	0.029343	0.001763	0.005039	0.001786
MAE($\boldsymbol{t}$)	0.022501	0.001451	0.003703	0.001195

Embedding Dimension

- Important parameter
- Spaces of different dimensions

Metrics	DCP-v1 (512)	DCP-v1 (1024)	DCP-v2 (512)	DCP-v2 (1024)
MSE($\boldsymbol{R}$)	6.480572	7.291216	1.307329	1.217545
RMSE($\boldsymbol{R}$)	2.545697	2.700225	1.143385	1.103424
MAE($\boldsymbol{R}$)	1.505548	1.616465	0.770573	0.750242
MSE($\boldsymbol{t}$)	0.000003	0.000001	0.000003	0.000003
RMSE($\boldsymbol{t}$)	0.001763	0.001150	0.001786	0.001696
MAE($\boldsymbol{t}$)	0.001451	0.000677	0.001195	0.001170

Conclusion

Conclusion

- High-quality alignment
- Further work
  - Iteratively
    - Reinforcement learning
  - Scenes
  - Larger pipelines
    - SLAM or SFM