RiEMann: Near Real-Time SE(3)-Equivariant Robot Manipulation without Point Cloud Segmentation

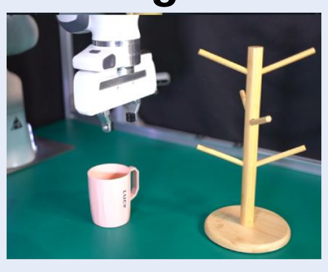
Chongkai Gao[1], Zhengrong Xue[2], Shuying Deng[2], Tianhai Liang[2], Siqi Yang[2], Lin Shao[1], and Huazhe Xu[234]

[1]National University of Singapore, [2]Tsinghua University, [3]Shanghai AI Lab, [4]Shanghai Qizhi Institute

Motivation

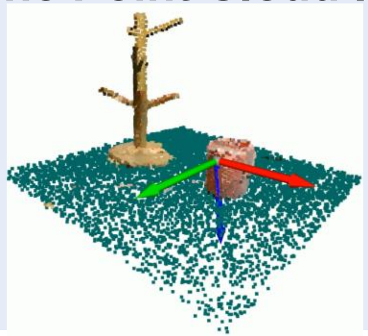
How to generalize the trained policy to new SE(3) poses, new instance, with distracting objects, without point cloud segmentation?

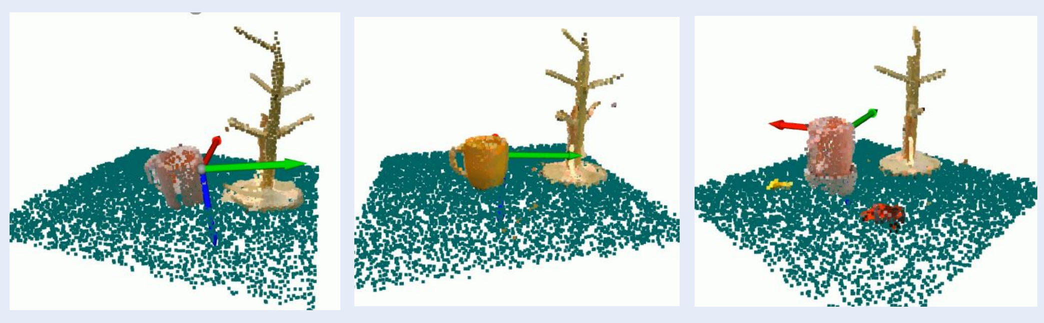
Training Cases



Testing Cases

Scene Point Cloud Input

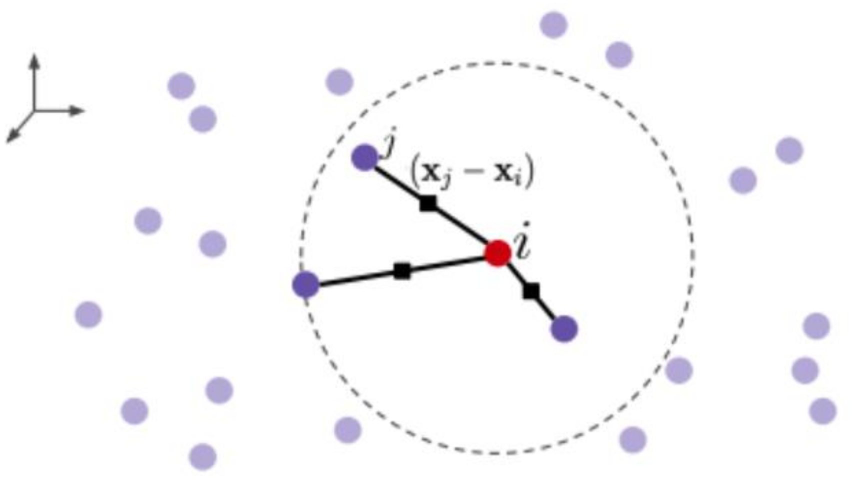




Equivariant Networks

$$S_g[f(x)] = f(T_g x), \quad \forall g \in SE(3), x \in \mathcal{X}$$

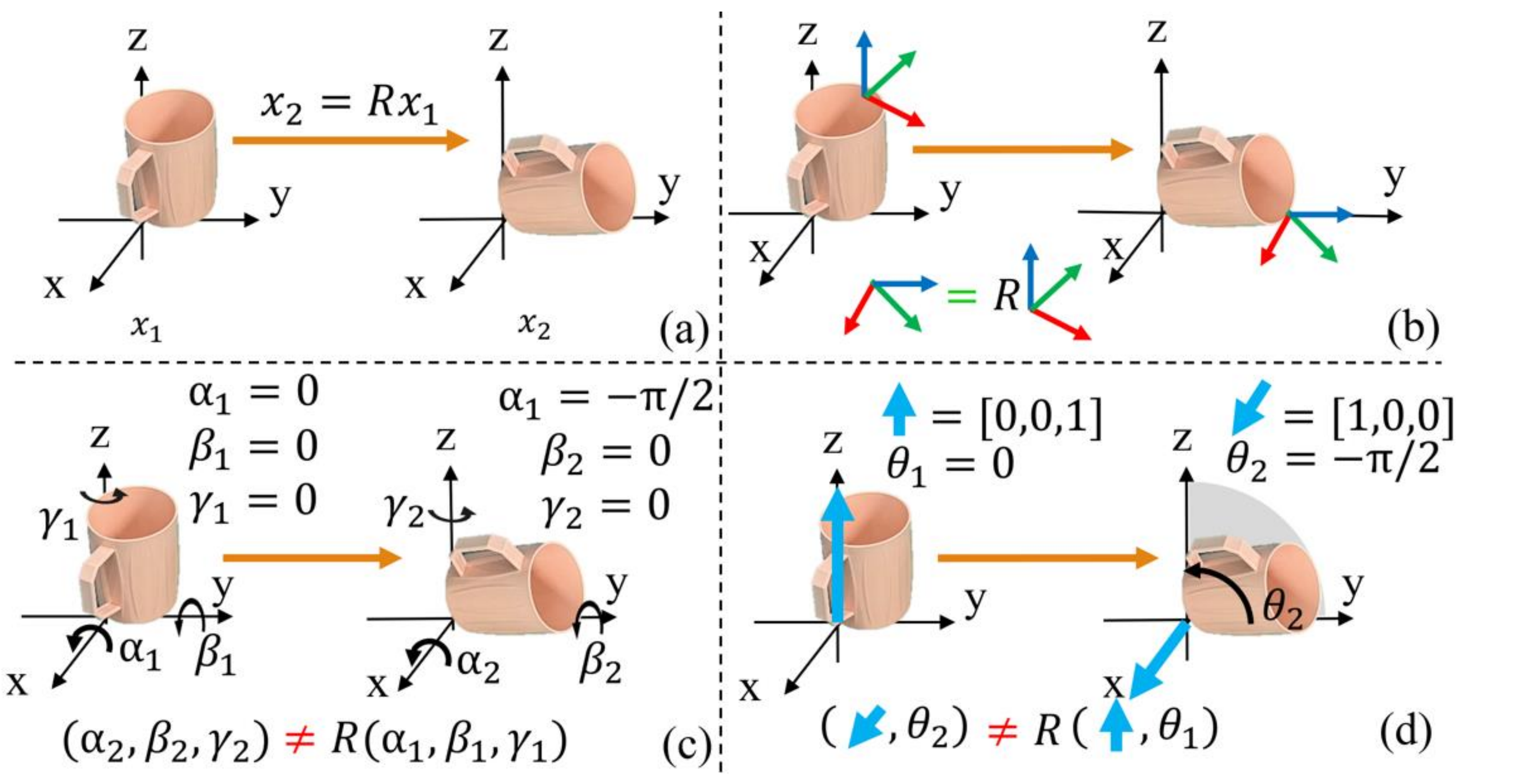
Local Mechanism



Method

Theorem 1 Rotation matrices, represented by three type-1 vectors, are SE(3)-equivariant vector field parameterization.

Theorem 2 There is no SE(3)-equivariant vector field parameterization for Euler angle, quaternion, and axis-angle.



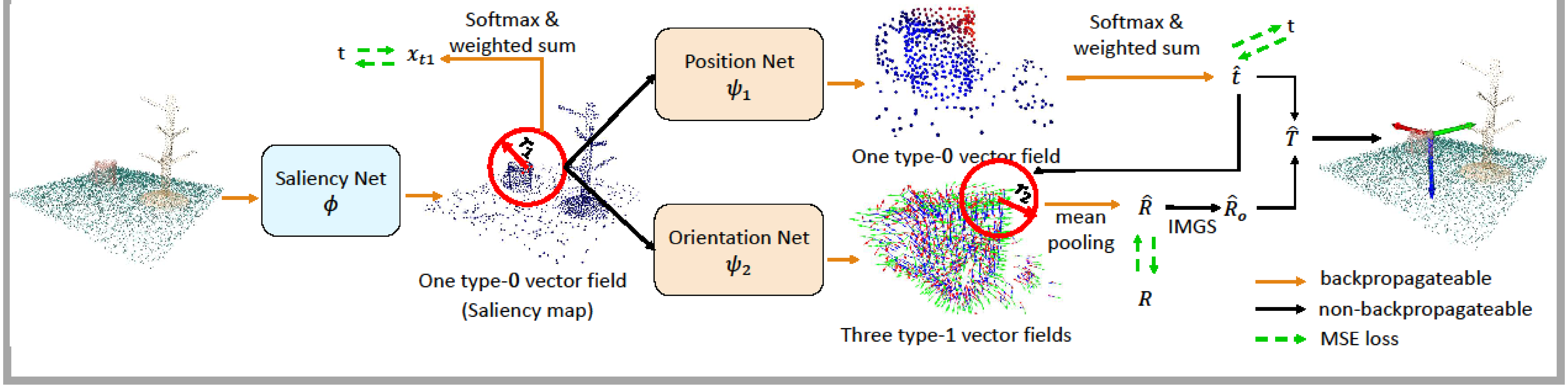
Algorithm 1 RiEMann Training

Input: Demonstrations $\{(\mathbf{P}_i, \mathbf{T}_i)\}_{i=1}^{M}$, initialized models ϕ, ψ_1, ψ_2, hyperparameters r_1 and r_2, epochs n.

for $iter = 0$ to $n-1$ **do**
- Sample a batch of m demonstrations $\{(\mathbf{P}_i, \mathbf{T}_i)\}_{i=1}^{m}$, where $\mathbf{T}_i = (\mathbf{R}_i, \mathbf{t}_i)$
- Predict the saliency map $\mathbf{f}_s(x) = \phi(x), x \in \mathbf{P}_i$
- Get x_{t1} by doing weighted sum on $\mathbf{P}$ with the soft-max weight from $\mathbf{f}_s(x)$
- Get $\mathbf{B}_{ROI}$ centered on x_{t1} with radius r_1
- Predict $\mathbf{f}_t(x) = \psi_1(x)$, $\mathbf{f}_R(x) = \psi_2(x)$, $\forall x \in \mathbf{B}_{ROI}$
- Get $\hat{\mathbf{t}}$ as the weighted position of $\mathbf{f}_t(x)$ and get $\widehat{\mathbf{R}}$ by mean pooing on $\mathbf{f}_R(x)$ on points centered at $\hat{\mathbf{t}}$ with the radius r_2
- Normalize each type-1 vector of $\widehat{\mathbf{R}}$
- Update ϕ, ψ_1, and ψ_2 with $\mathcal{L} = \sum_{i=0}^{m}[\sum_{j=1}^{N}(\mathbf{t}_i - \hat{\mathbf{t}}_i)^2 + \sum_{k=1}^{N_B}((\mathbf{t}_i - \hat{\mathbf{t}}_i)^2 + (\mathbf{R}_i - \widehat{\mathbf{R}}_i)^2)]$

end for

Output: Trained models ϕ, ψ_1, and ψ_2



Experiments

Main Results

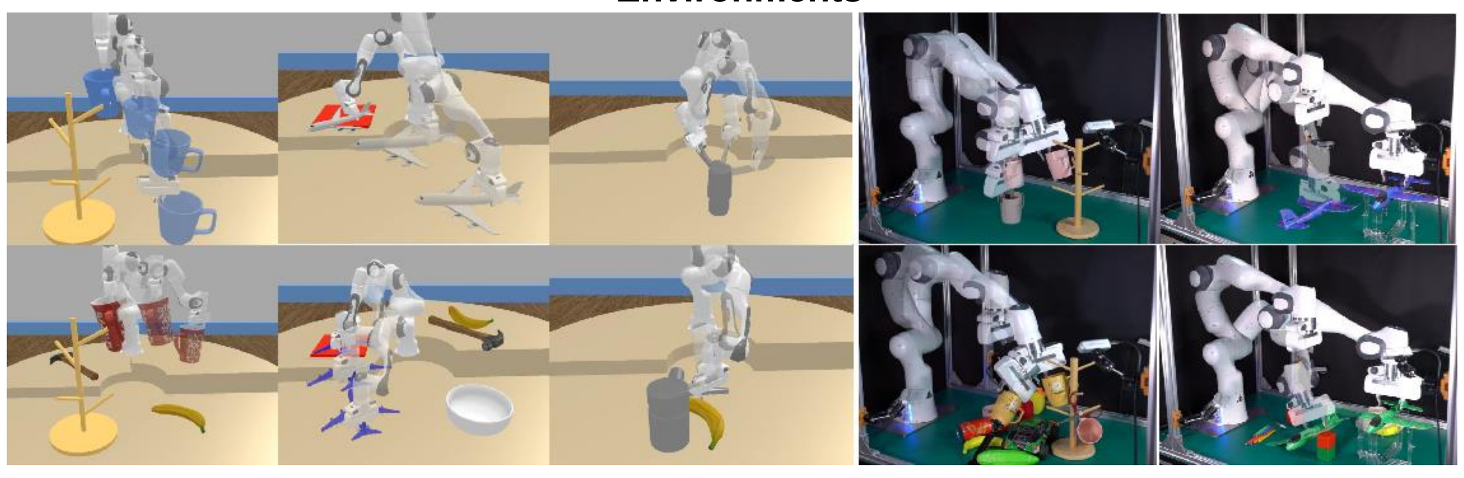
Table 1. Success rates of different tasks in simulation. Evaluated under 20 random seeds.

	Mug on Rack					Plane on Shelf					Turn Faucet				
Method	T	NI	NP	DO	ALL	T	NI	NP	DO	ALL	T	NI	NP	DO	ALL
PerAct [37]	0.85	0.00	0.70	0.00	0.00	0.90	0.00	0.80	0.00	0.00	0.45	0.00	0.50	0.00	0.00
R-NDF [39]	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	n/a	n/a	n/a	n/a	n/a
EDF [33]	1.00	0.85	1.00	0.95	0.80	0.90	0.75	0.80	0.85	0.70	n/a	n/a	n/a	n/a	n/a
D-EDF [34]	1.00	0.85	0.95	0.95	0.75	1.00	0.80	0.95	0.95	0.75	n/a	n/a	n/a	n/a	n/a
RiEMann (Ours)	1.00	0.90	0.95	1.00	0.85	1.00	0.90	1.00	1.00	0.90	1.00	0.75	1.00	1.00	0.65

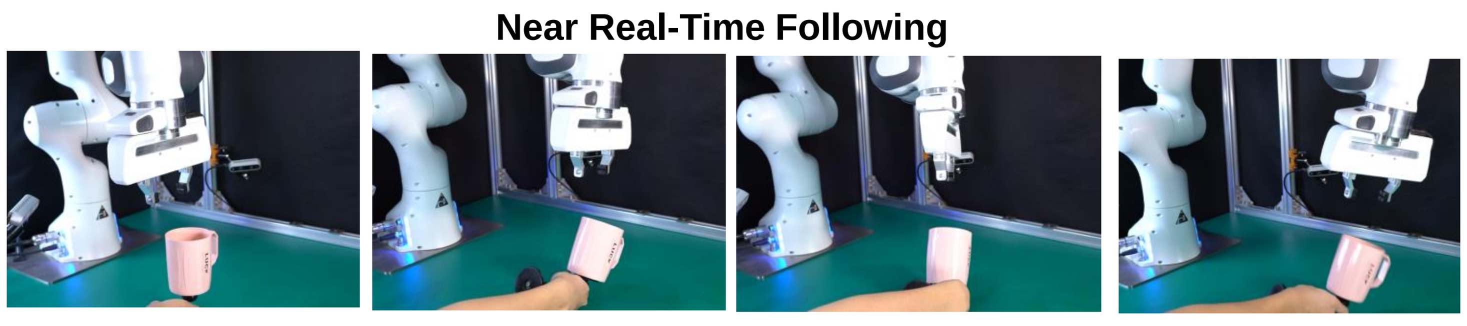
Table 2. SE(3) Geodesic distances of tasks in simulation. Evaluated under 20 random seeds.

	Mug on Rack					Plane on Shelf					Turn Faucet				
Method	T	NI	NP	DO	ALL	T	NI	NP	DO	ALL	T	NI	NP	DO	ALL
PerAct [37]	0.393	4.086	0.698	4.166	4.375	0.431	4.806	0.469	4.752	4.993	0.457	4.365	0.382	4.218	4.039
R-NDF [39]	4.855	4.298	4.178	4.509	4.662	4.277	4.361	4.179	4.466	4.989	4.996	4.374	4.278	4.229	4.560
EDF [33]	0.249	0.429	0.347	0.252	0.501	0.333	0.872	0.461	0.337	0.985	0.188	1.473	0.448	0.242	2.049
D-EDF [34]	0.312	0.545	0.425	0.337	0.682	0.328	0.966	0.417	0.345	1.024	0.304	2.047	0.567	0.488	2.249
RiEMann (Ours)	0.053	0.066	0.069	0.056	0.068	0.101	0.120	0.117	0.099	0.122	0.079	0.159	0.098	0.082	0.197

Environments



Near Real-Time Following



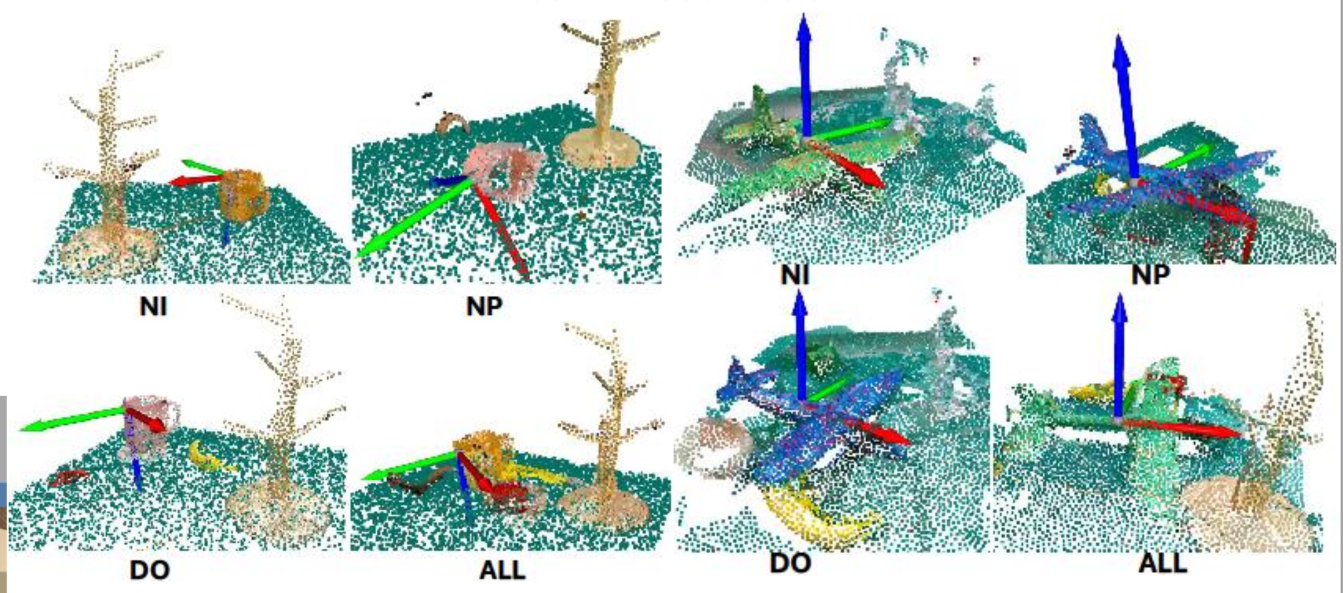
Cost and Speed

	Training Time	Inference Time	Memory Usage
RiEMann (Ours)	47mins	0.19s	11GB
D-EDFs [34]	40 mins	15.2s	42GB

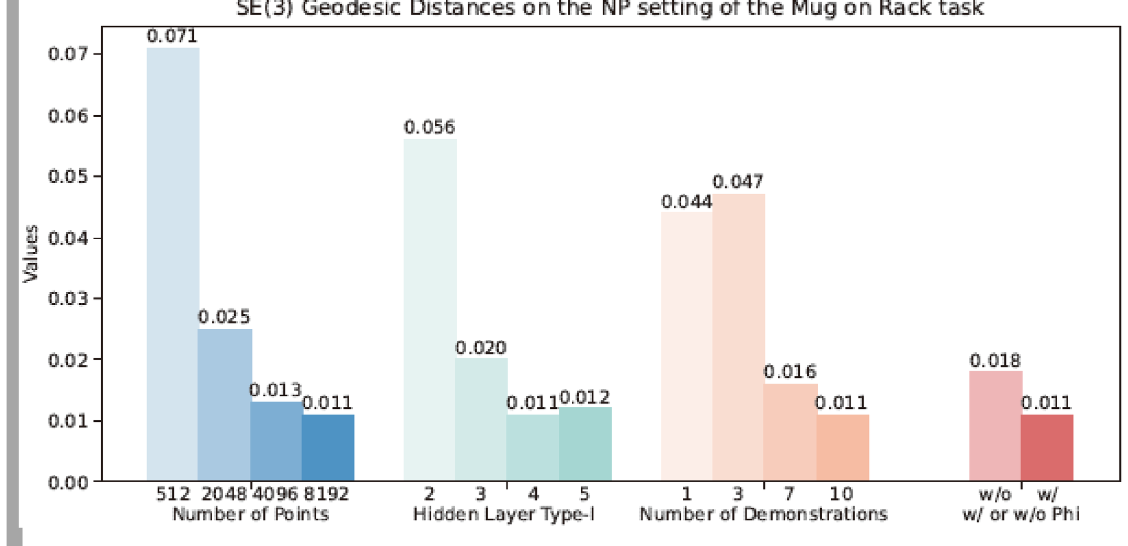
Real-World Results

	T		NI		NP		DO		ALL	
Task	G	A	G	A	G	A	G	A	G	A
Mug on Rack	1.0	1.0	1.0	1.0	0.75	0.75	0.92	0.83	0.75	0.58
Plane on Shelf	1.0	1.0	1.0	1.0	0.58	0.50	1.0	1.0	0.55	0.50

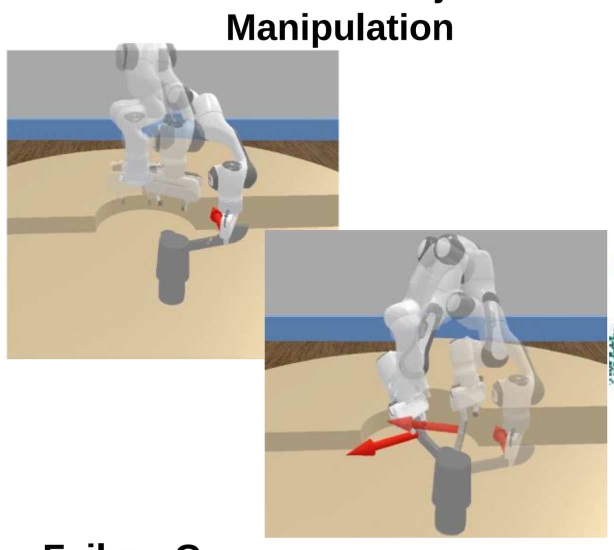
Result Visualization



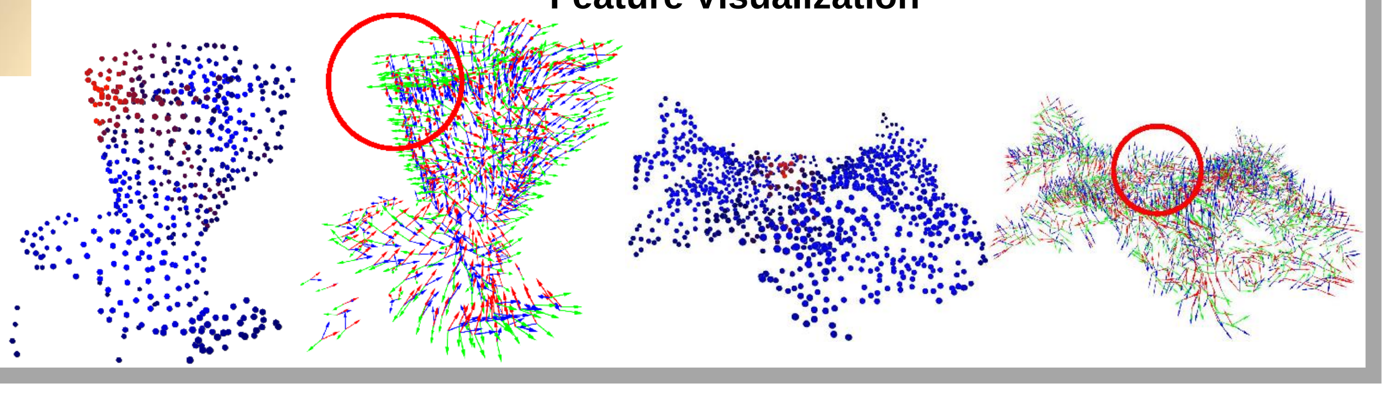
Ablation Studies



Articulated Object Manipulation



Feature Visualization



Geometry Generalization

Picture Description	original	flat	tall	fat	new
$\mathcal{D}_{geo}$	0.055	0.187	0.127	0.094	0.395

Failure Cases