Mining and Transferring Feature-Geometry Coherence for Unsupervised Point Cloud Registration

KeZheng Xiong[1], Haoen Xiang[1], Qingshan Xu[2], Chenglu Wen[1], Siqi Shen[1], Jonathan Li[3], Cheng Wang[1]

[1]Xiamen University, China.
[2]Nanyang Technological University, Singapore.
[3]University of Waterloo, Waterloo, Canada

NEURAL INFORMATION PROCESSING SYSTEMS

Preliminaries

➢ What is point cloud registration?

- Point cloud registration estimates a rigid transformation $\mathbf{T} = \{\mathbf{R}, \mathbf{t}\}$ that aligns two point clouds.

$$\min_{\mathbf{R},\mathbf{t}} \sum_{(\mathbf{p}_{x_i},\mathbf{q}_{y_i})\in\mathcal{C}^\star} \|\mathbf{R}\mathbf{p}_{x_i} + \mathbf{t} - \mathbf{q}_{y_i}\|^2,$$

- When the two point clouds are acquired at **a large distance** d such as when d ∈ [5m, 50m], the registration task faces the challenges of low overlap and density variation. Therefore, it is crucial to **learn density-invariant features.**

Challenges

➢ Challenges:

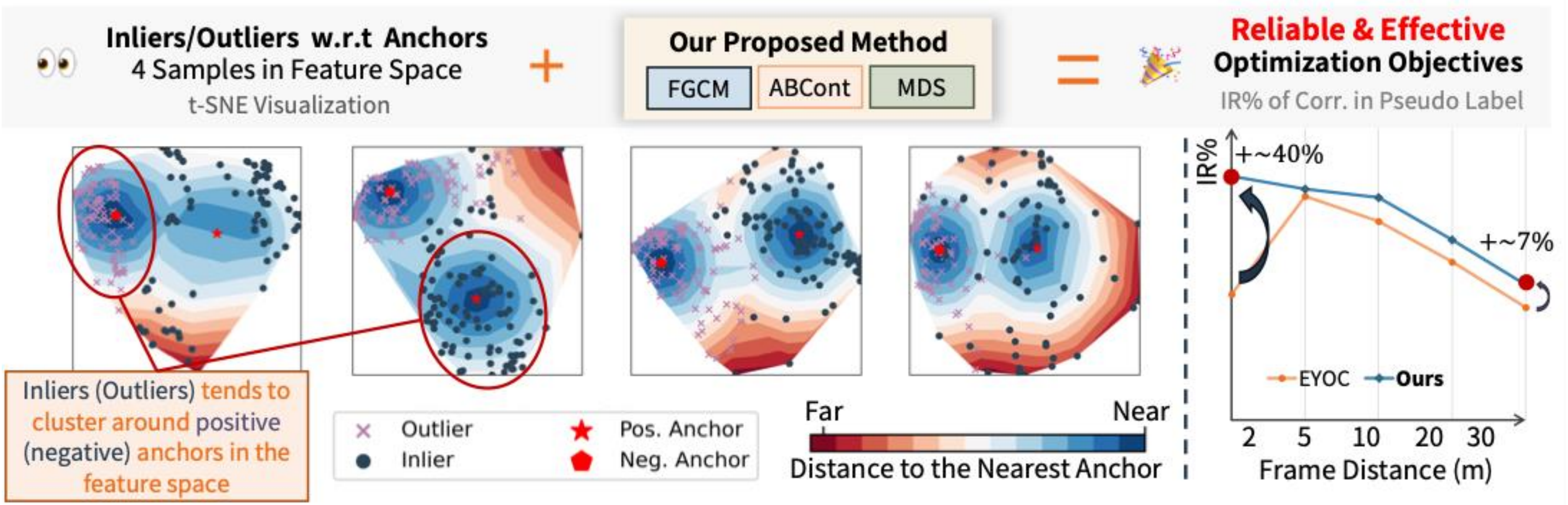
- **Costly** pose annotations
- **Poor generalizability** of supervised methods
- Large-scale and complexly-distributed outdoor LiDAR point.

➢ Existing Efforts:

- Relying on **overly-strong** geometric assumptions
- **Poor quality of pseudo-labels** due to inadequate integration of low-level geometric and high-level contextual information.

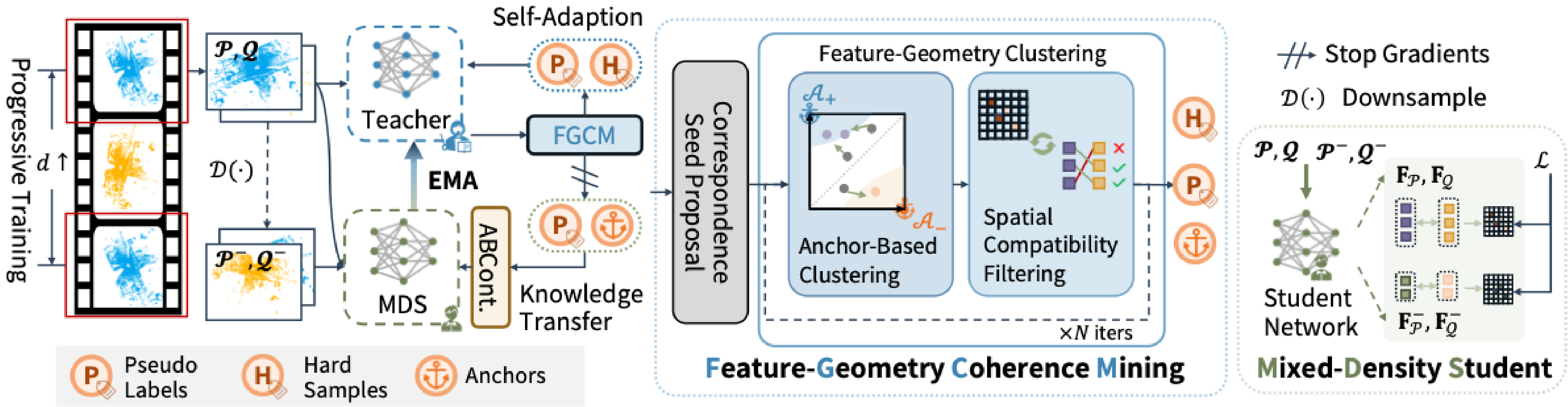
Suboptimal Performance of Existing Methods

Motivation



- Our **observation** in the feature space
  - High-level contextual information **is adept at discovering inliers** from a global perspective of the scene.
  - Low-level geometric cues have proven **effective in rejecting outliers.**

Methodology



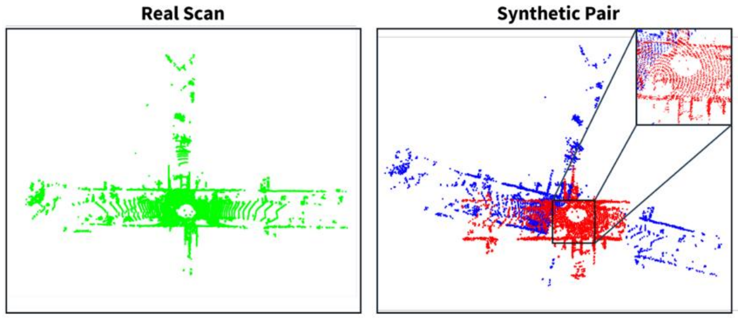
Main Contributions:

- **INTEGER**, a novel teacher-student framework that exploits low-level and high-level information for unsupervised point cloud registration, demonstrating its superior registration performance in complex outdoor environments.
- **FCEM and MDS** for the teacher and student, respectively, to mine reliable pseudo-labels and learn density-invariant features.
- **ABCont** to mitigate pseudo-label noise and facilitate contrastive learning with anchors for a robust feature space.

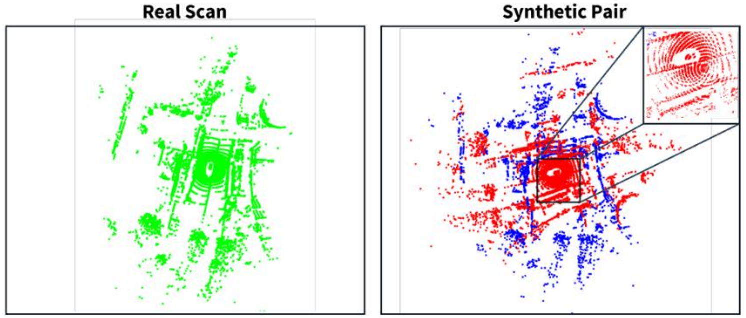
Methodology

➢ Synthetic Teacher Initialization

- Initialize the teacher by training with synthetic pairs.
- Following **PointContrast**[1] to generate two partially-overlap fragments for each scan.
- Additionally apply **periodic sampling**[2] to remove points periodically with respect to a random center, simulating the **irregular sampling of LiDAR**.



(a) A Sample from KITTI



(b) A Sample from nuScenes

1. Saining Xie, Jiatao Gu, Demi Guo, Charles R Qi, Leonidas Guibas, and Or Litany. Pointcontrast: Unsupervised pre-training for 3d point cloud understanding. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16, pages 574–591. Springer, 2020.
2. Sofiane Horache, Jean-Emmanuel Deschaud, and François Goulette. 3d point cloud registration with multi-scale architecture and unsupervised transfer learning. In 2021 international conference on 3D vision (3DV), pages 1351–1361. IEEE, 2021.

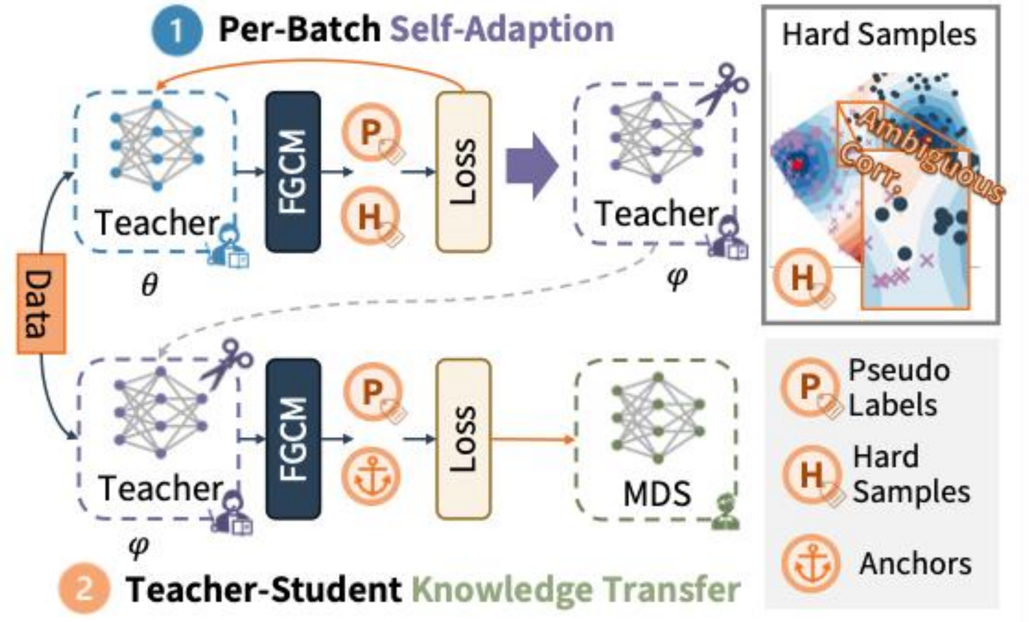
空间感知与计算实验室
spAtial Sensing & Computing Lab

Methodology

➢ Feature-Geometry Coherence Mining

Algorithm 1: Feature-Geometry Clustering

Input: Initial correspondence seed proposals $\mathcal{C}^0$
Compute initial $\mathbf{U}^{\mathcal{P}}$, $\mathbf{U}^{\mathcal{Q}}$ and anchors $\mathcal{A}_+$, $\mathcal{A}_-$ with Eq. 1
for $i = 1$ *to* max_iters **do**
  Generate unclassified correspondences $\mathcal{C}_{\mathrm{U}} \leftarrow \mathrm{FeatureMatching}\left(\mathbf{U}^{\mathcal{P}}, \mathbf{U}^{\mathcal{Q}}\right)$
  Select $\mathcal{C}_{\mathrm{U}}^{\text{top-}k}$ with top-k S_c^+ satisfying $S_c^+ > S_c^-$ based on Eq. 2
  Update $\mathcal{C}^i \leftarrow \mathcal{C}^{i-1} \cup \mathcal{C}_{\mathrm{U}}^{\text{top-}k}$ `// Anchor-Based Clustering`
  Filter $\mathcal{C}^i$ with spatial compatibility to produce $\mathcal{C}_+^i$, $\mathcal{C}_-^i$ `// Spatial Compatibility Filtering`
  Update $\mathbf{U}^{\mathcal{P}}$, $\mathbf{U}^{\mathcal{Q}}$ and $\mathcal{A}_+$, $\mathcal{A}_-$ with Eq. 1
  if $|\mathcal{C}_+^i| = |\mathcal{C}_+^{i-1}|$ *or* $|\mathcal{C}_-^i| = |\mathcal{C}_-^{i-1}|$ **then**
    $\mathcal{C} \leftarrow \mathcal{C}_+^i$
    break
return $\mathcal{C}$, $\mathcal{A}_+$, $\mathcal{A}_-$



- **In the first forward pass**, we perform Per-Batch Self-Adaption on the teacher model θ to establish a denoised feature space, yielding a data-specific teacher φ.
- **In the second forward pass**, the adapted teacher φ and FGCM are used to mine reliable pseudo-labels l, which are then used to train the student, achieving Teacher-Student Knowledge Transfer

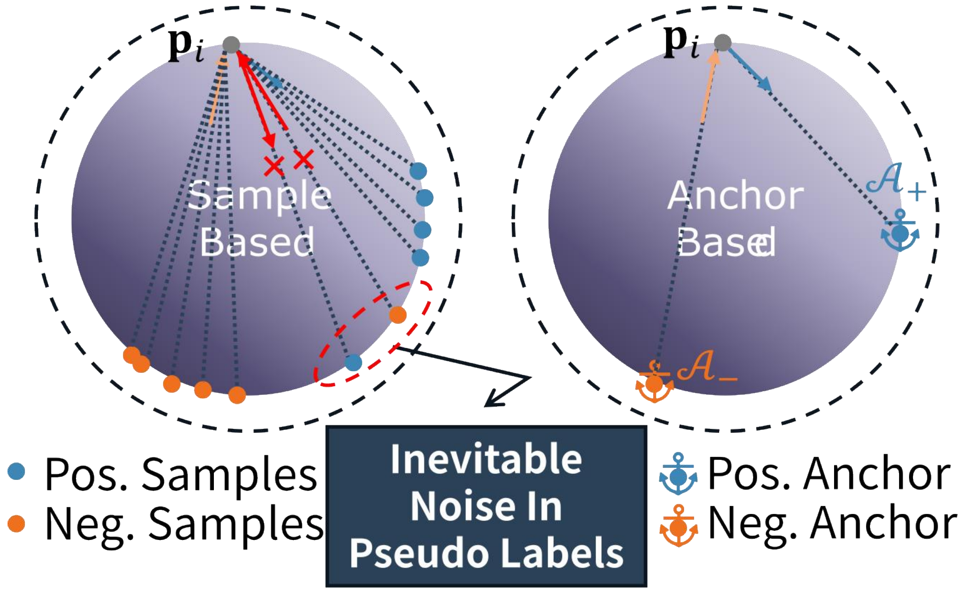
Methodology

➢ Anchor-Based Contrastive Learning (ABCont)

$$\mathcal{L}_{\text{ABCont}} = \mathcal{L}_{\text{reg}} + \lambda_{\text{corr}} \mathcal{L}_{\text{corr}},$$

$$\mathcal{C}_+^\star = \mathcal{C}_+ \cup \text{sg}(\mathcal{A}_+),\ \mathcal{C}_-^\star = \mathcal{C}_- \cup \text{sg}(\mathcal{A}_-),$$

$$\mathcal{L}_{\text{corr}} = -\frac{1}{n_p} \sum_{i=1}^{n_p} \log \frac{\exp(\beta_p^i)}{\exp(\beta_p^i) + \sum_{j=1}^{n_n} \exp(\beta_n^j)},$$



Toy Example for ABCont. Anchor-based methods introduce fewer pairwise relationships and are robust against inevitable label noise.

Methodology

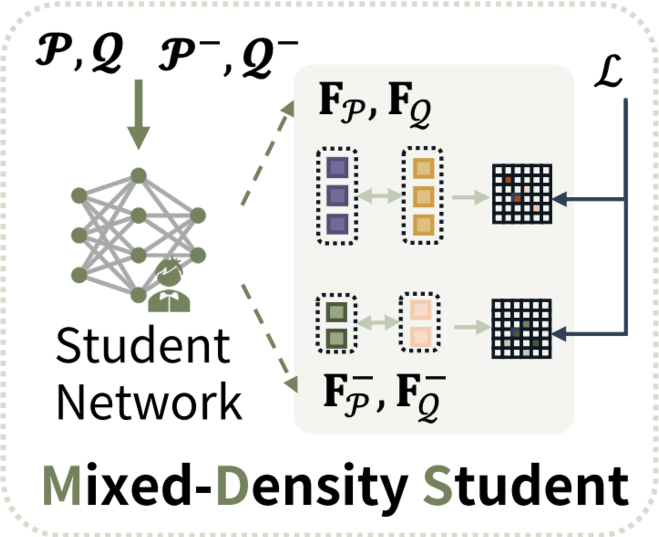
➢ Mixed-Density Student(MDS)

Challenges:

- The density of LiDAR point clouds varies greatly with the distance to the sensor, posing challenges for matching distance point clouds effectively

Using features from downsampled views for **density-invariant** training

$$\mathcal{L} = \mathcal{L}_{\text{ABCont}}^{(\mathcal{P},\mathcal{Q})} + \lambda_1 \mathcal{L}_{\text{ABCont}}^{(\mathcal{P}^-,\mathcal{Q}^-)}$$



Experiments

- Comparison with SOTAs

Dataset	Method	U	mRR	RR@$d \in$				
				$[5,10)$	$[10,20)$	$[20,30)$	$[30,40)$	$[40,50)$
KITTI	FCGF	–	77.4	98.4	95.3	86.8	69.7	36.9
	FCGF+C	–	84.6	100.0	97.5	90.1	79.1	56.3
	Predator	–	87.9	100.0	97.5	90.1	79.1	56.3
	SpinNet	–	39.1	99.1	82.5	13.7	0.0	0.0
	D3Feat	–	66.4	99.8	98.2	90.7	38.6	4.5
	CoFiNet	–	82.1	99.9	99.1	94.1	78.6	38.7
	GeoTrans.	–	42.2	100.0	93.9	16.6	0.7	0.0
	EYOC	✓	83.2	**99.5**	96.6	89.1	78.6	52.3
	RIENet	✓	50.7	96.3	72.1	38.2	24.4	22.6
	Ours	✓	**84.0**	**99.5**	**97.1**	**89.6**	**79.6**	**54.2**
nuScenes	FCGF	–	39.5	87.9	63.9	23.6	11.8	10.2
	FCGF+C	–	59.3	96.2	85.1	59.6	35.8	20.0
	Predator	–	51.0	99.7	72.2	52.8	16.2	14.3
	EYOC	✓	61.7	96.7	85.6	61.8	37.5	26.9
	RIENet	✓	47.1	96.5	57.9	36.6	25.8	18.9
	Ours	✓	**63.1**	**97.1**	**86.9**	**62.9**	**39.6**	**29.4**
KITTI ↓ nuScenes	EYOC	✓	55.3	96.2	75.6	58.7	26.6	19.7
	RIENet	✓	46.2	83.3	73.2	43.5	19.8	11.1
	Ours	✓	**62.6**	**97.5**	**84.6**	**62.6**	**37.8**	**30.2**

- Ablation Studies

Methods	tIR@1st Epoch	mRR	$d \in [40,50)$		
			RR	RRE	RTE
Full	**81.2**	**84.0**	**54.2**	**1.1**	**0.54**
w/o ABCont	80.3	83.5	53.7	1.3	0.58
w/o GSA	43.3	80.9	50.2	1.7	0.79
w/o FGC	67.6	82.8	52.7	1.4	0.61
w/o MDS	81.2	82.7	52.3	1.3	0.71
w/o S.T.I	71.9	83.7	53.7	1.2	0.55

a) Ablation Study of Proposed Components

Pose Estimators	tIR@1st Epoch	Time (s)
PointDSC	**81.3**	1.13
MAC	80.1	28.2
FastMAC	79.3	**0.67**
SC^2-PCR	<u>81.2</u>	<u>0.75</u>

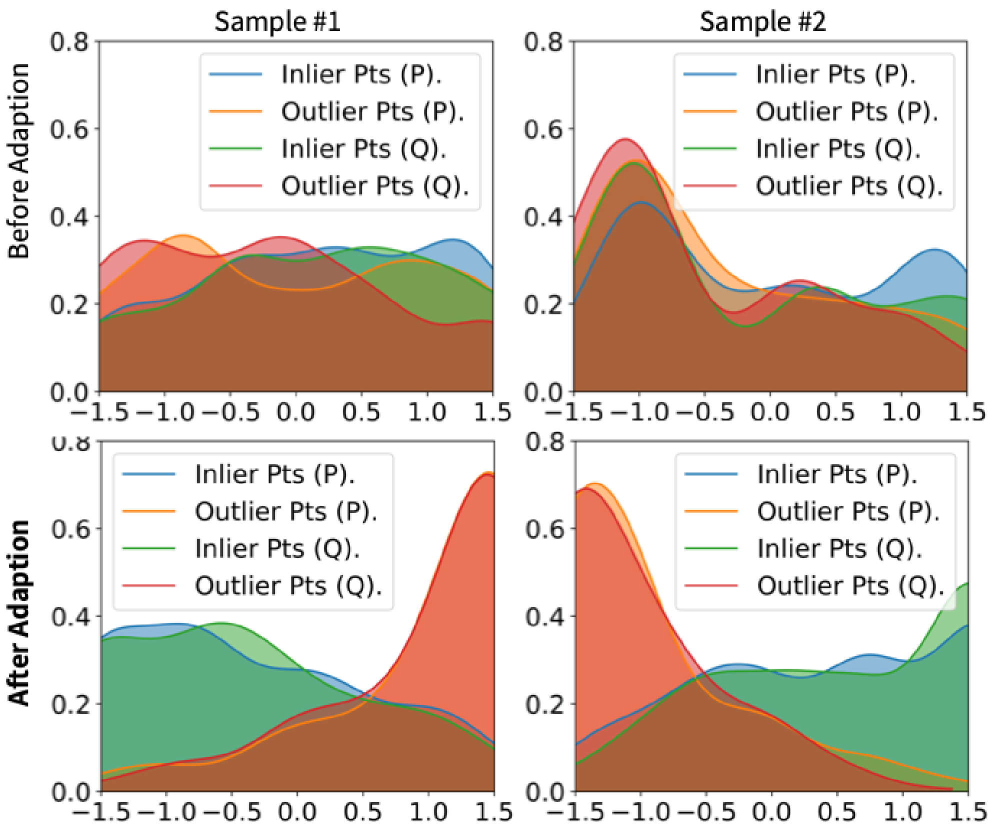
b) Different Pose Estimators for FGC-Branch in FCEM

空间感知与计算实验室
spAtial Sensing & Computing Lab

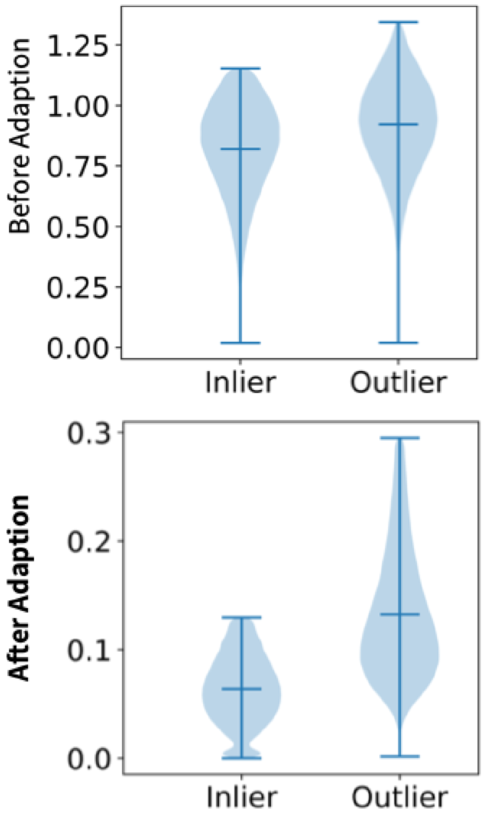
Experiments

- Effectiveness of GSA-Branch for Discriminative Features

Feature Distribution of Samples



Feature Similarity Distribution of Inlier/Outlier Correspondences



Before v.s. After Self-Adaption in GSA-Branch: Point-wise Feature & Correspondence-wise Similarity Distribution indicate that the **self-adaption results in more discriminative features.**

空间感知与计算实验室
spAtial Sensing & Computing Lab

Thanks

Project Page

github.com/kezheng1204/INTEGER

NEURAL INFORMATION PROCESSING SYSTEMS