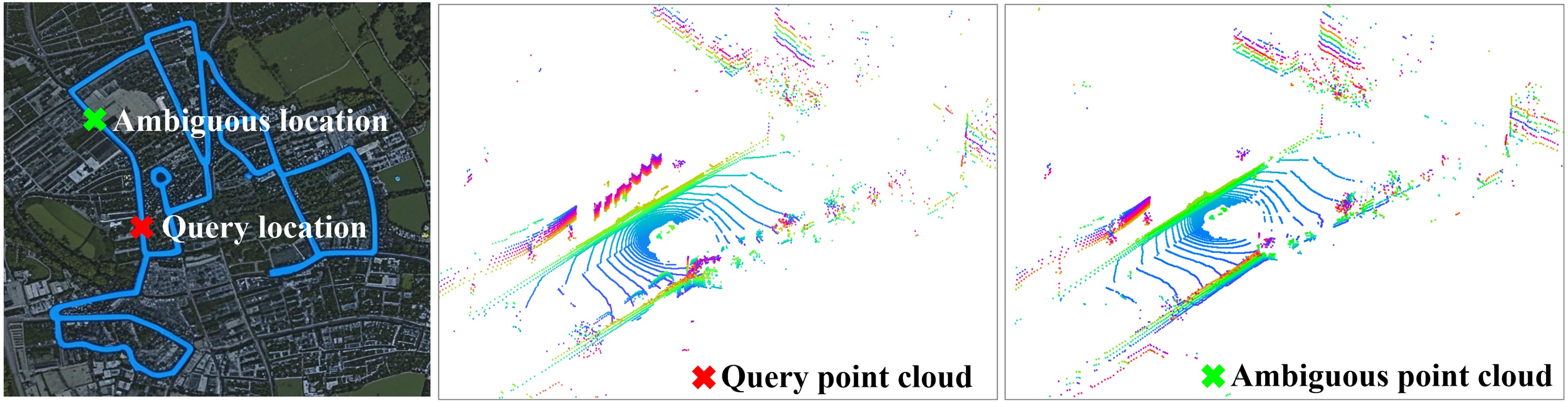
GTR-Loc: Geospatial Text Regularization Assisted Outdoor LiDAR Localization

Shangshu Yu[1], **Wen Li**[2,3], **Xiaotian Sun**[2,3], **Zhimin Yuan**[4],
Xin Wang[1], **Sijie Wang**[5], **Rui She**[6], **Cheng Wang**[2,3]*
[1]School of Computer Science and Engineering, Northeastern University, Shenyang 110819, China
[2]Fujian Key Laboratory of Sensing and Computing for Smart Cities, Xiamen University, China
[3]Key Laboratory of Multimedia Trusted Perception and Efficient Computing,
Ministry of Education of China, Xiamen University, China
[4]School of Artificial Intelligence and Software Engineering, Nanyang Normal University, China
[5]Nanyang Technological University, Singapore
[6]Beihang University, China
yushangshu@cse.neu.edu.cn cwang@xmu.edu.cn

NEURAL INFORMATION PROCESSING SYSTEMS
2025

Challenges & Motivation

NEURAL INFORMATION PROCESSING SYSTEMS 2025



(a) (b) (c)

- Prevailing **scene coordinate regression** methods for LiDAR localization suffer from **localization ambiguities**, as distinct locations can exhibit similar geometric signatures.

- Conventional text descriptions can be **subjective, inconsistent across different times**, or **ambiguous for continuous observations**, making text-enhanced localization challenging.

Contribution

- We propose GTR-Loc, a novel text-assisted LiDAR localization framework. GTR-Loc is the first work to effectively design and integrate **geospatial text descriptions as regularization** to improve LiDAR SCR, leading to promising localization performance.

- Extensive experiments on QEOxford, Oxford, and NCLT datasets demonstrate the great effectiveness of GTR-Loc, particularly outperforming state-of-the-art methods by **9.64%/8.04% on QEOxford**.

Framework

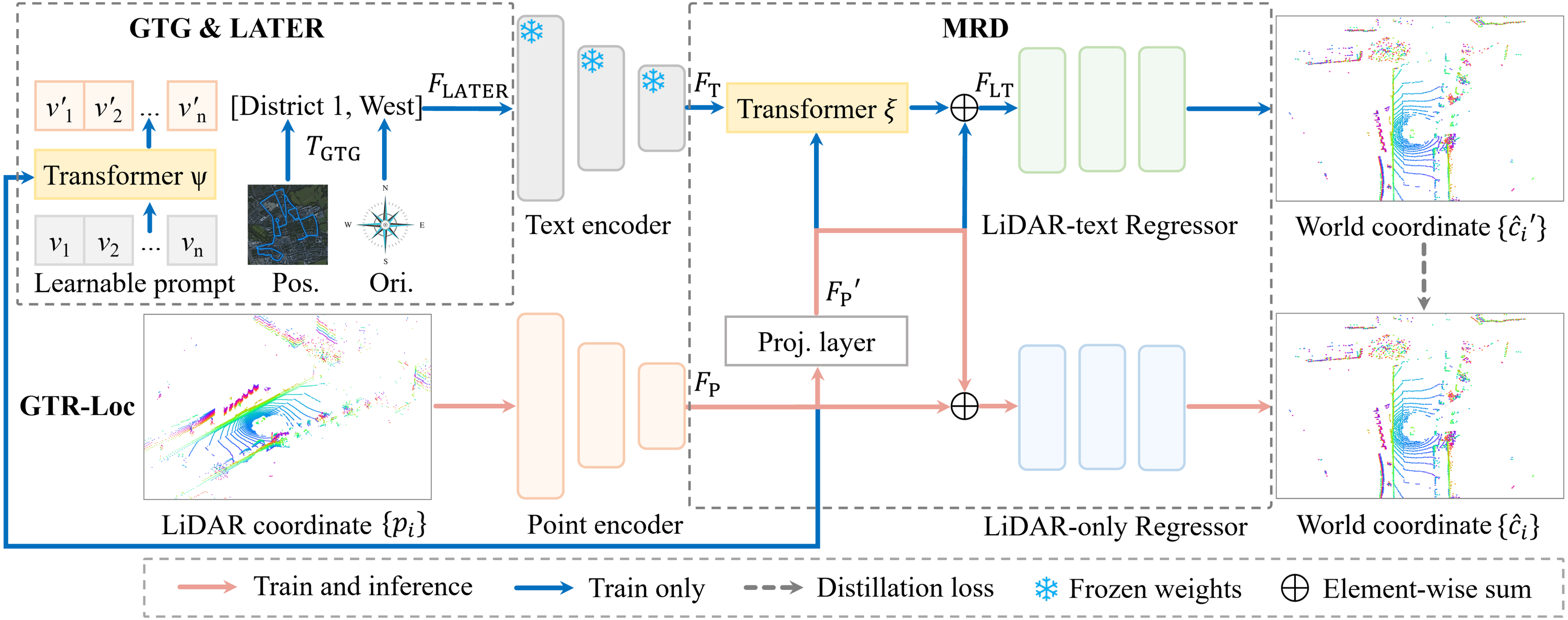
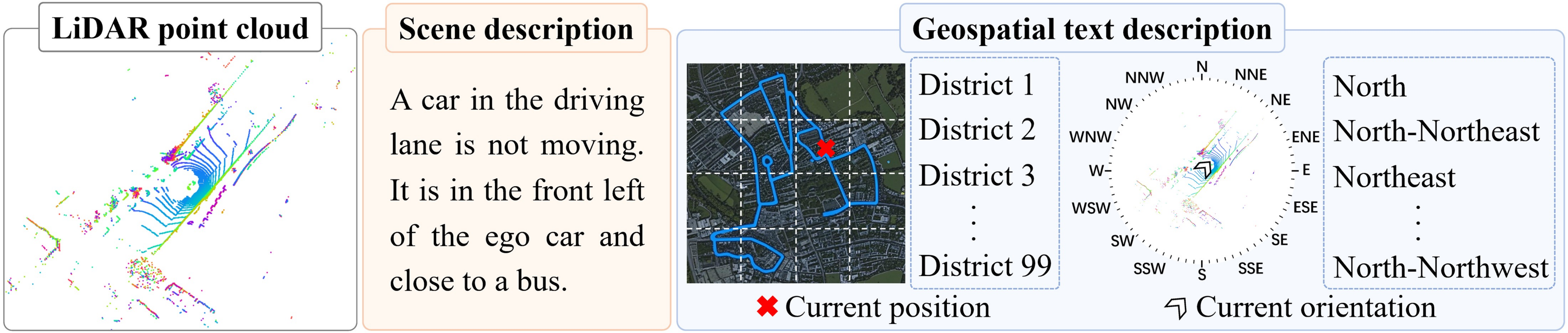


Figure 2: Overview of our method. GTR-Loc enhances LiDAR localization with text assistance to regularize SCR: a Geospatial Text Generator (**GTG**) provides discrete pose-aware text descriptions T_{GTG}, and a LiDAR-Anchored Text Embedding Refinement (**LATER**) module dynamically constructs view-specific text embeddings F_{LATER} conditioned on point features F_P. A Transformer ξ is employed to fuse multimodal features for LiDAR-text regression. Furthermore, a Modality Reduction Distillation (**MRD**) strategy enables LiDAR-only inference by distilling textural regularization.

GTG & LATER & MRD



(a) (b) "District 99, West-Southwest."

- We propose a **Geospatial Text Generator** and a **LiDAR-Anchored Text Embedding Refinement** module to dynamically create **view-specific text descriptions** focused on **discrete pose information**, providing enhanced disambiguation capabilities for LiDAR localization.

$$\{v_1^{'}, v_2^{'}, ..., v_n^{'}\} = Transformer_{\psi}(Q = \{v_1, v_2, ..., v_n\}, K, V = F_{\mathrm{P}}),$$

$$F_{\mathrm{LATER}} = \{v_1^{'}, v_2^{'}, ..., v_n^{'}, W_{\mathrm{GTG}}\},$$

- We devise a **Modality Reduction Distillation** strategy to enable LiDAR-only localization during inference.

$$\mathcal{L}_{\mathrm{LT}} = \sum_{i=1}^{N} \|\hat{c}_i^{'} - c_i\|_1, \mathcal{L}_{\mathrm{LO}} = \sum_{i=1}^{N} \|\hat{c}_i - c_i\|_1, \ \mathcal{L}_{\mathrm{D}} = \sum_{i=1}^{N} \|\hat{c}_i - \hat{c}_i^{'}\|_1, \ \mathcal{L}_{\mathrm{SCR}} = \beta_1 \mathcal{L}_{\mathrm{LT}} + \beta_2 \mathcal{L}_{\mathrm{LO}} + \beta_3 \mathcal{L}_{\mathrm{D}}.$$

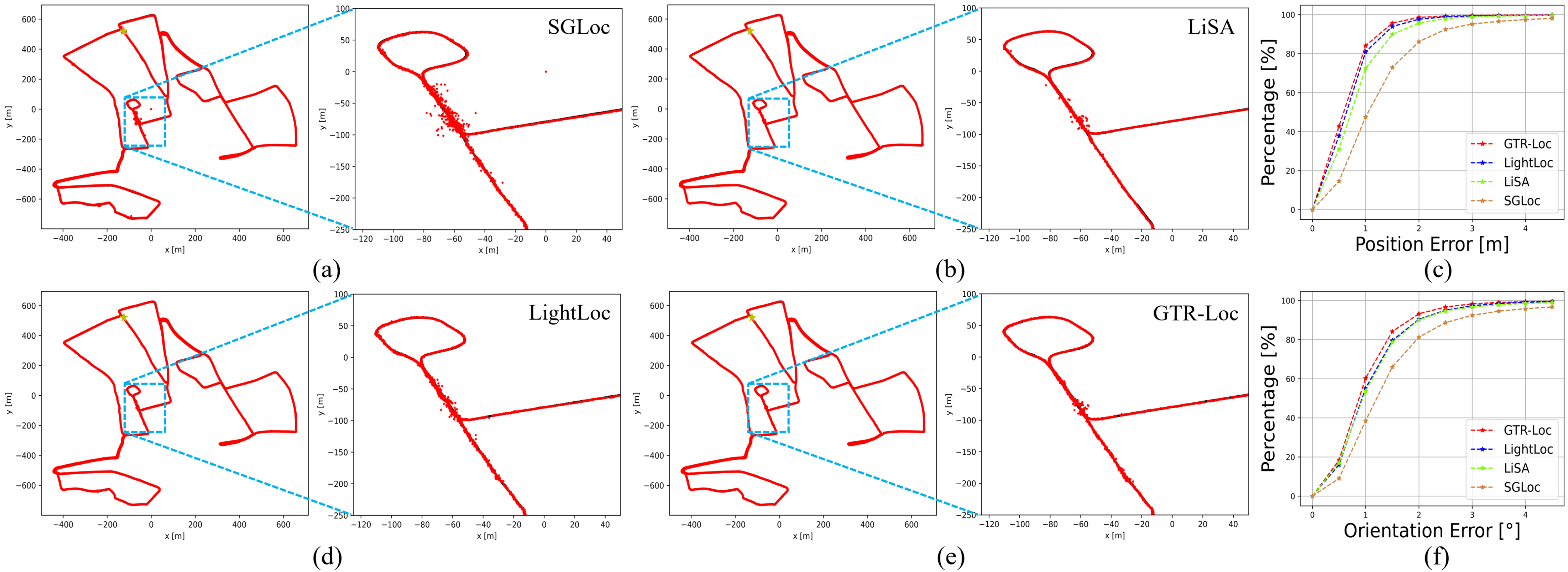
Experiment

NEURAL INFORMATION PROCESSING SYSTEMS 2025

Methods	Mech.	TFs	15-13-06-37	17-13-26-39	17-14-03-00	18-14-14-42	Avg. [m/°]
STCLoc [48]	MA	3	5.14/1.27	6.12/1.21	5.32/1.08	4.76/1.19	5.34/1.19
NIDALoc [47]	MA	5	3.71/1.50	5.40/1.40	3.94/1.30	4.08/1.30	4.28/1.38
DiffLoc [22]	MA	3	**2.03/1.04**	**1.78/0.79**	**2.05/0.83**	**1.56/0.83**	**1.86/0.87**
PointLoc [41]	SA	1	10.75/2.36	11.07/2.21	11.53/1.92	9.82/2.07	10.79/2.14
PosePN [46]	SA	1	9.47/2.80	12.98/2.35	8.64/2.19	6.26/1.64	9.34/2.25
PosePN++ [46]	SA	1	4.54/1.83	6.44/1.78	4.89/1.55	4.64/1.61	5.13/1.69
PoseMinkLoc [46]	SA	1	6.77/1.84	8.84/1.84	8.08/1.69	6.56/2.06	7.56/1.86
PoseSOE [46]	SA	1	4.17/1.76	6.16/1.81	5.42/1.87	4.16/1.70	4.98/1.79
HypLiLoc [40]	SA	1	5.03/1.46	4.31/1.43	3.61/1.11	2.61/1.09	3.89/1.27
FlashMix [12]	SA	1	2.04/1.95	1.95/1.83	2.44/2.18	2.81/2.14	2.31/2.03
SGLoc [23]	SS	1	1.79/1.67	1.81/1.76	1.33/1.59	1.19/1.39	1.53/1.60
LiSA [45]	SS	1	0.94/1.10	1.17/1.21	0.84/1.15	0.85/1.11	0.95/1.14
LightLoc [21]	SS	1	0.82/1.12	0.85/1.07	0.81/1.11	0.82/1.16	0.83/1.12
GTR-Loc	SS	1	**0.77/1.02**	**0.77/1.01**	**0.67/1.01**	**0.80/1.07**	**0.75/1.03**

- Quantitative results on the QEOxford dataset.

Experiment



- Visual comparisons on the QEOxford dataset.

Thanks!

- **Homepage：** https://psyz1234.github.io/