HypLiLoc: Towards Effective LiDAR Pose Regression with Hyperbolic Fusion

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE
UNIVERSITY OF OXFORD
C3.ai

Sijie Wang[1*] Qiyu Kang[1*] Rui She[1] Wei Wang[2] Kai Zhao[1] Yang Song[3] Wee Peng Tay[1]

[1] Nanyang Technological University [2] University of Oxford [3] C3.AI

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Abstract

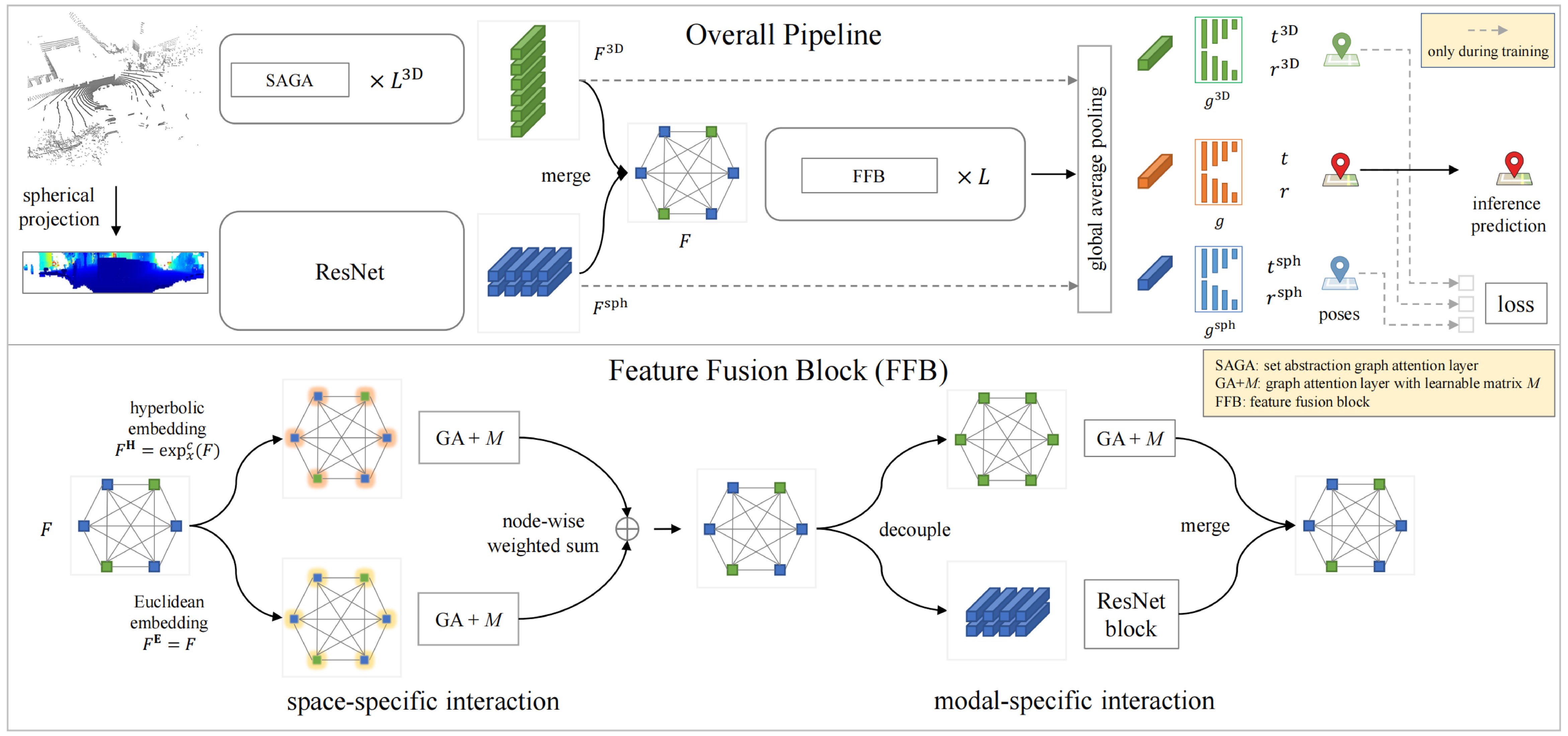
LiDAR relocalization plays a crucial role in many fields, including robotics, autonomous driving, and computer vision. LiDAR-based retrieval from a database typically incurs high computation storage costs and can lead to globally inaccurate pose estimations if the database is too sparse. On the other hand, pose regression methods take images or point clouds as inputs and directly regress global poses in an end-to-end manner. They do not perform database matching and are more computationally efficient than retrieval techniques. We propose HypLiLoc, a new model for LiDAR pose regression. We use two branched backbones to extract 3D features and 2D projection features, respectively. We consider multi-modal feature fusion in both Euclidean and hyperbolic spaces to obtain more effective feature representations. Experimental results indicate that HypLiLoc achieves state-of-the-art performance in both outdoor and indoor datasets. We also conduct extensive ablation studies on the framework design, which demonstrate the effectiveness of multi-modal feature extraction and multi-space embedding. Our code is released at: https://github.com/sijieaaa/HypLiLoc

Abstract

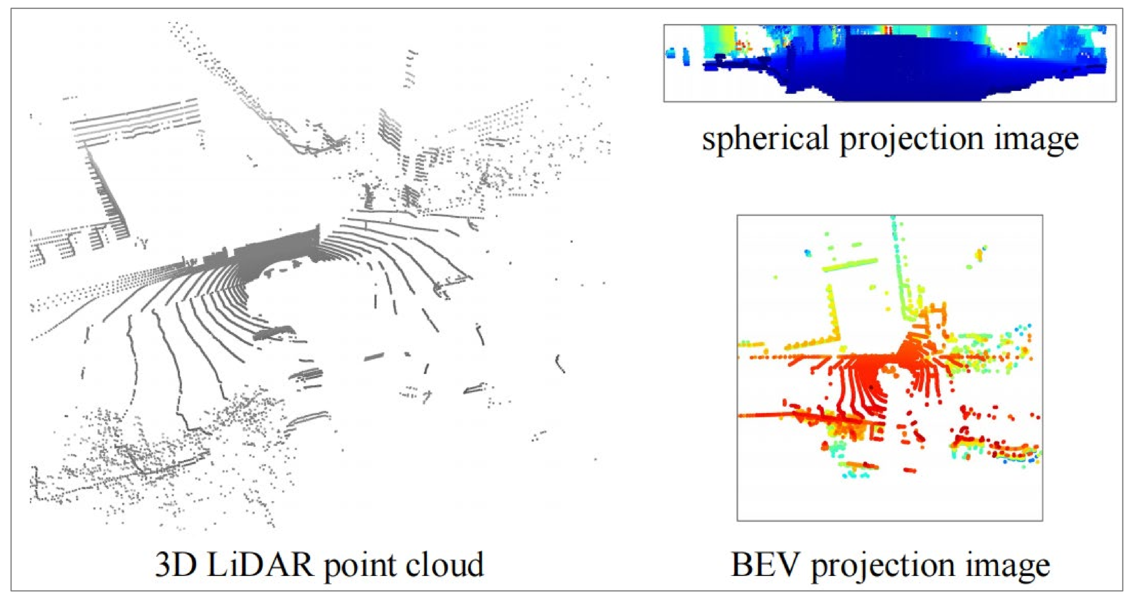
· We propose a novel LiDAR-based pose regression network HypLiLoc. It has one backbone that learns 3D features directly from the 3D point cloud and another backbone that learns features from a 2D projection of the point cloud onto a spherical surface.To achieve effective multi-modal feature interaction, the features are embedded in both Euclidean and hyperbolic spaces using multi-space learning. An attention mechanism is then used to fuse the features from different spaces together.

· We test our network in both outdoor and indoor datasets, where it outperforms current LiDAR pose regression counterparts and achieves SOTA performance. We also conduct extensive ablation studies on the effectiveness of each design component.

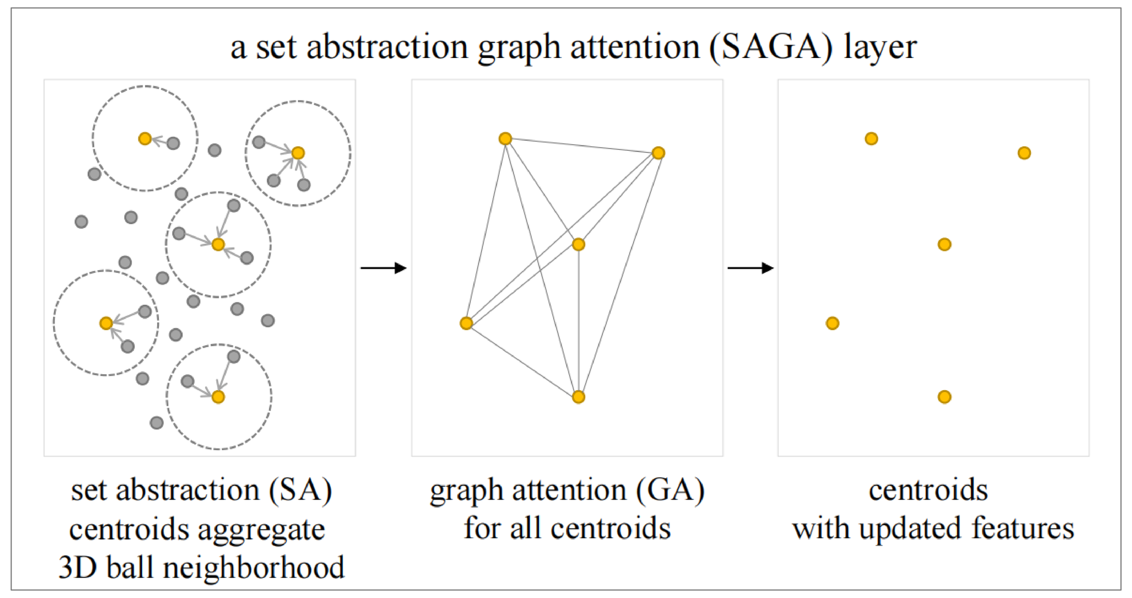
Architecture



The overall architecture of our proposed HypLiLoc. We use two backbone branches to perform feature extraction. In the 3D backbone, we consider both local set abstraction and global attention aggregation. In the feature fusion block, the extracted multi-modal features are embedded into both Euclidean and hyperbolic spaces to achieve space-specific interaction. The fusion features are then decoupled to their own modality to perform modal-specific interaction. The final training loss is applied on both the 3D/projection level and the final fusion level.

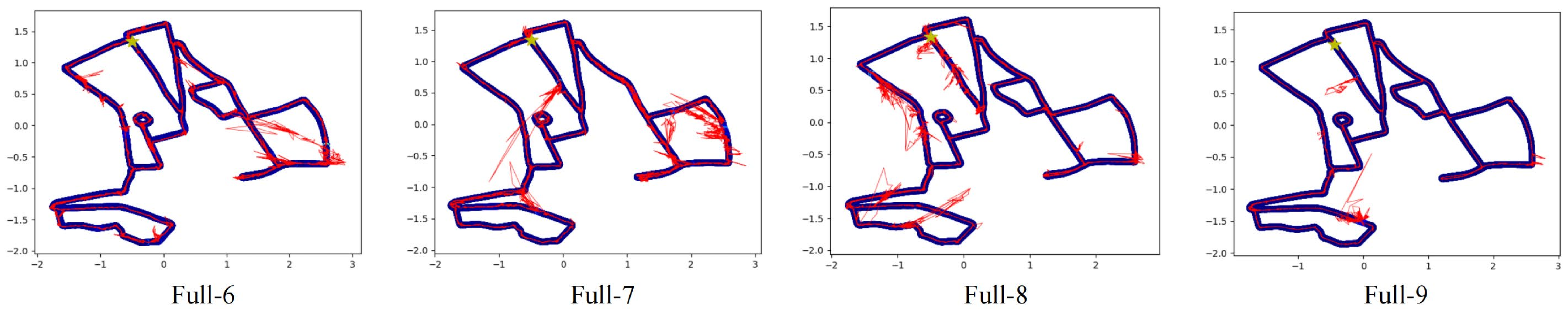


Visualization of the spherical and BEV projection methods.



The SAGA layer consists of a SA layer and a GA layer

Trajectory Visualization



Experiments

	Model	Seq-05	Seq-06	Seq-07	Seq-14
Image-based PR	PoseLSTM [35]	0.16 / 4.23	0.18 / 5.28	0.24 / 7.05	0.13 / 4.81
	MapNet [15]	0.26 / 6.67	0.28 / 6.91	0.39 / 9.17	0.25 / 6.85
	AD-MapNet [16]	0.17 / 3.33	0.21 / 3.37	0.24 / 4.38	0.14 / 4.12
	AtLoc+ [36]	0.18 / 4.32	0.24 / 5.14	0.26 / 6.04	0.16 / 4.61
	MS-Transformer [33]	0.16 / 3.98	0.15 / 3.56	0.18 / 5.32	0.13 / 4.83
LiDAR-based PR	PointLoc [38]	<u>0.12</u> / 3.00	0.10 / 2.97	**0.13** / 3.47	0.11 / 2.84
	PosePN [45]	<u>0.12</u> / 4.38	<u>0.09</u> / 3.16	0.17 / 3.94	**0.08** / 3.27
	PosePN++ [45]	0.15 / 3.12	0.10 / 3.31	<u>0.15</u> / <u>2.92</u>	0.10 / <u>2.80</u>
	PoseSOE [45]	0.14 / 3.15	0.11 / <u>2.90</u>	<u>0.15</u> / 3.06	0.11 / 3.20
	PoseMinkLoc [45]	0.16 / 5.17	0.11 / 3.74	0.21 / 5.74	0.12 / 3.64
	HypLiLoc (ours)	**0.09 / 2.52**	**0.08 / 2.58**	**0.13 / 2.55**	<u>0.09</u> / **2.34**

Mean translation and rotation error (m/°) on the Oxford Radar dataset. The best and the second-best results in each metric are highlighted in bold and <u>underlined</u>, respectively. PR stands for pose regression. HypLiLoc achieves the best performance in all metrics.

	Model	Full-6	Full-7	Full-8	Full-9
LiDAR Retrieval	PointNetVLAD [34]	28.48 / 5.19	17.62 / 3.95	23.59 / 5.87	13.71 / 2.57
LiDAR Odometry	DCP [40]	18.45 / 2.08	14.84 / 2.17	16.39 / 2.26	13.60 / 1.86
Image-based PR	PoseLSTM [35]	26.36 / 6.54	74.00 / 9.85	128.25 / 18.59	19.12 / 3.05
	MapNet [15]	48.21 / 6.06	61.01 / 5.85	75.35 / 9.67	44.34 / 4.54
	AD-MapNet [16]	18.43 / 3.28	19.18 / 3.95	66.21 / 9.42	15.10 / 1.82
	AtLoc+ [36]	17.92 / 4.73	34.03 / 4.01	71.51 / 9.91	10.53 / 1.97
	MS-Transformer [33]	11.69 / 5.66	65.38 / 9.01	88.63 / 19.80	7.62 / 2.53
LiDAR-based PR	PointLoc [38]	13.81 / <u>1.53</u>	9.81 / <u>1.27</u>	11.51 / <u>1.34</u>	9.51 / <u>1.07</u>
	PosePN [45]	16.32 / 2.43	14.32 / 3.06	13.48 / 2.60	9.14 / 1.78
	PosePN++ [45]	10.64 / 1.78	9.59 / 1.92	<u>9.01</u> / 1.51	8.44 / 1.71
	PoseSOE [45]	<u>8.81</u> / 2.04	<u>7.59</u> / 1.94	9.21 / 2.12	<u>7.27</u> / 1.87
	PoseMinkLoc [45]	11.20 / 2.62	14.69 / 2.90	12.35 / 2.46	10.06 / 2.15
	HypLiLoc (ours)	**6.00 / 1.31**	**6.88 / 1.09**	**5.82 / 0.97**	**3.45 / 0.84**

Median translation and rotation error (m/°) on the vReLoc dataset. The best and the second-best results in each metric are highlighted in bold and <u>underlined</u>, respectively. PR stands for pose regression. HypLiLoc achieves the best performance in 7 out of 8 metrics.

Dataset Visualization