2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 14-18, 2024. Abu Dhabi, UAE

Augmenting Vision with Radar for All-weather Geo-localization without a Prior HD Map

Can Dong[1*], Ziyang Hong[1*], Siru Li[1], Liang Hu[1] and Huijun Gao[2]

***Abstract*— Accurate and robust geo-localization in all-weather conditions is essential for enabling autonomous vehicles and delivery robots to offer uninterrupted mobility services in the real world. In this paper, we propose the first camera and radar fusion based geo-localisation method that is robust to all-weather conditions. The core of the proposed method is to leverage the rich semantics information in images and sensing consistency in radars across all-weather. Our proposed method surpasses the state of the art camera-based and LiDAR-camera based methods in inclement weather conditions, shown by extensive comparative experiments. Notably, our approach requires only an open accessible map, eliminating the need for high-definition maps and offering a cost-effective solution for geo-localizing or globally localizing autonomous vehicles in any weather condition. Our code and trained model will be released publicly.**

I. INTRODUCTION

Accurate localization is fundamental for autonomous vehicles. Over the past decade, there has been significant exploration into localization with prior high definition (HD) maps, as a promising solution for localization in autonomous driving. Unfortunately the endeavor to build, update, and maintain HD maps, especially in expansive environments spanning multiple cities or even countries, has proven to be a monumental challenge. This difficulty underscores the necessity for alternative localization solutions that do not rely on HD maps.

Recently, a few novel HD map-free geo-localization methods have been proposed [1], [2], [3]. Those methods geo-localize autonomous vehicles globally by registering and associating on-board sensory data from camera, lidar with free-accessible maps such as OpenStreetMap (OSM) or satellite imagery. However, none of the above methods considers scenarios under adverse weather conditions (e.g., rain, snow and fog), in which radars demonstrate superior perception capability compare with cameras and LiDARs. Furthermore, nowadays vision-based systems is increasingly favored to complement sensors of other modalities, as even a low-cost camera can capture images with rich visual information of the surroundings, thus enhancing perception greatly.

To overcome the limitations faced by single-modality geo-localization methods, particularly in challenging weather conditions, we propose a novel all-weather geo-localization method using camera and radar fusion. As illustrated in Fig. 1, significant increase in localization accuracy is achieved by fully exploiting complementary visual and radar perceptions. Our contributions can be summarized as follows:

1) To the best of our knowledge, we propose the first HD map-free geo-localization system under all weather conditions. Building in an end-to-end learning framework, the system can be easily deployed in cities of different continents and in various weather conditions.
2) Our introduced multi-modal fusion of vision and radar enables a robust and accurate geo-localization system under all-weather conditions, showing superior localization performance compared to single-modal (camera, radar) and LiDAR-camera geo-localization methods using OSM.
3) We release our open-source codes for the community.

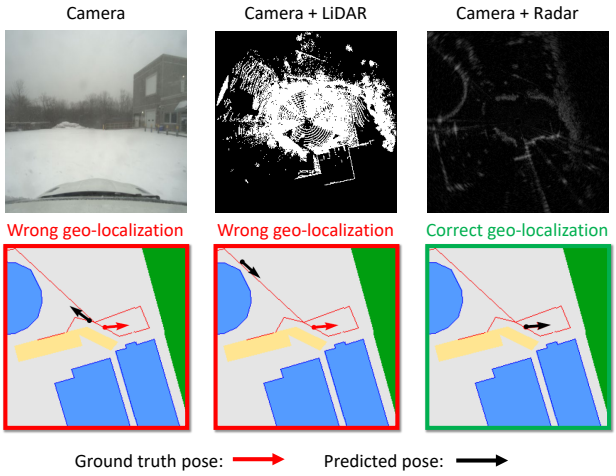

Fig. 1: Vision-based geo-localiztion method suffers from visual domain shift problem so it degrades on unseen adverse weather conditions while LiDAR is vulnerable in heavy snow condition since it cannot penetrate the snow particle. Radar is highly consistent in all-weather conditions so that it is suitable for long-term navigational task.

II. RELATED WORK

Research on robot localization has been carried out extensively. According to the data sources for localization, common methods used in autonomous driving include LiDAR, GPS, radar, and vision-based methods. When classified by map type, we can localize the autonomous vehicle in the world

This work is partially supported by Shenzhen Science and Technology Program (Project No. JCYJ20220818103000001) and Guangdong Provincial Key Laboratory of Intelligent Morphing Mechanisms and Adaptive Robotics (Project No. 2023B1212010005).

*Equal Contribution.

[1]C. Dong, Z. Hong, S. Li and L. Hu are with the Department of Automation, School of Mechanical Engineering and Automation, Harbin Institute of Technology, Shenzhen, China. For correspondence: l.hu@hit.edu.cn.

[2] H. Gao is with the Department of Control Science and Engineering, Harbin Institute of Technology, Harbin, China.

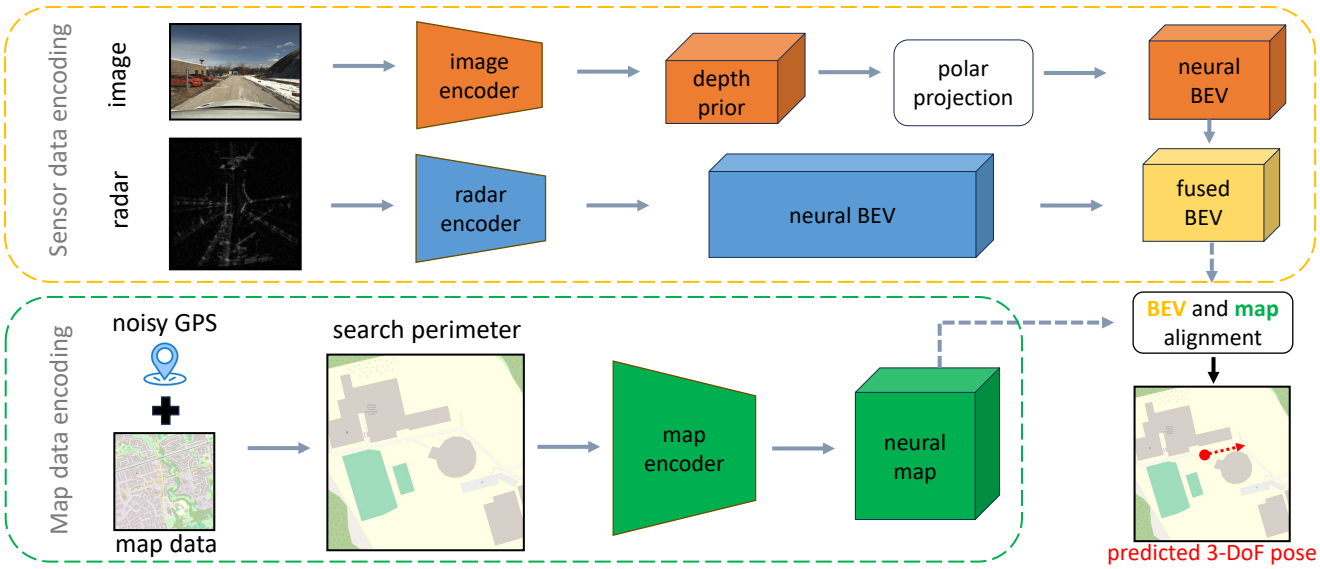

Fig. 2: Overview of framework: The model comprises two distinct pathways. The camera images undergo feature extraction and projection to the Bird's Eye View (BEV) perspective, resulting in neural BEV features. The Cartesian radar image, after feature extraction are integrated with the camera's BEV features in the fusion module. We narrow the search perimeter from the global map based on the noisy GPS information. The map data is fed into a encoder to build a neural representation. Finally, the fused BEV feature is engaged with the neural map in a probabilistic matching model for positional estimation.

using various map representations, including 3D maps built from sensor data, 2D overhead satellite images, or simpler planimetric maps from OSM.

A. Localization with Pre-built Sensor Maps

Visual localization: The most common way to perform visual re-localization or global localization is to build a map or a 3D model using Structure-from-Motion (SfM). The query image is used to perform 6-DoF pose estimation against the pre-built model. Traditionally, the problem is solved with two stage: firstly as an image retrieval problem [4] and following with pose estimation. For example, [5] propose to extract SIFT descriptor as visual word for fast searching and use RANSAC and PnP for pose estimation. As deep learning becomes popular, the community seeks to represent the image with high level CNN feature. PoseNet [6] uses a deep learning framework for regressing the camera's pose, training a CNN to perform six degrees of freedom (6-DoF) camera pose regression directly on input RGB images. Instead of performing the visual localization entirely end-to-end, we can also use trained point feature like SuperPoint [7] so that we can use classical method to compute pose estimation or trained graph neural network matcher like [8].

Range-based localization: In contrast to visual localization, range-based methods employ rich geometry information and perceive in a larger field-of-view. They are naturally invariant to lighting conditions and thus, being considered more reliable for autonomous driving application. Deep learning and classical filters are often combined to perform global localization. Deep features are first used to solve the place recognition problem while a particle filter is used to estimate the pose [9], [10]. For autonomous driving application, range-based global localization can be reduced to 3 DoF estimation since the vehicle movement is mostly perpendicular to the ground. Under this assumption, OverlapNet [11] proposes to use depth, normal, and intensity values as input channels for the deep network. Another interesting line of works involves cross-modality localization, of which the general setting is to build a map using LiDAR data while localizing it using radar data [12], [13], [14].

B. Localization using Public Maps

Localization with overhead imagery: One of the localization systems without HD maps is to perform localization using satellite imagery. Approaches relying on satellite maps are evolving by using ground-level photographs with aerial or satellite imagery. Tian et al. proposes a method based on building matching, which involves incorporating the appearance and location of many iconic buildings in [15]. To enhance the data association capability, [16] proposes a Transformer-based 2D-3D matching network, registering LiDAR point clouds and satellite images. While Tang et al. [2] proposes to use generative models to transform different range sensor modalities and satellite image into similar synthetic embedding, and then to estimate the pose via the synthetic embedding.

Localization with geographic database: Localization utilizing geographic database like OSM have been studied for more than a decade [17], [18], [19], [20], [21], [22]. More recently, [1] creates OSM descriptors by assessing the distances between a given location and buildings within OSM, and generates LiDAR descriptors by computing the nearest distances to building points from the current position at set

angles. Instead of performing geo-localization with OSM, a pose tracking system using radar sensing is proposed in [23], assuming the initial position is known.

The latest research OrienterNet [3] aims to estimate the pose of query images by creating a neural bird's-eye view (BEV) mental map of the scene. This approach demonstrates powerful generalization ability, proving that vision is a critical component in the geo-localization system.

III. METHODOLOGY

A. Problem Formulation

Consider the scenario that an autonomous ground vehicle (AGV) equipped with monocular camera, a radar and a low-cost GPS device undertakes outdoor navigation tasks. Our primary objective is to geo-localize the AGV accurately and reliably under all-weather conditions using its on-board sensors and openly-accessible maps. Formally, we aim to estimate the 3-DoF pose of the AGV $\xi = (x, y, \theta)$, where x and y signifies the spatial coordinates and θ denotes the orientation of the AGV with respect to the global frame.

B. System Overview

Inspired by the pioneering work on visual location OrienterNet [3], we propose a multi-modal fusion based geo-relocation system against all-weather conditions in this paper. Multi-modal inputs from three different types of sensor are exploited to estimate the AGV's pose: the noisy location for GPS provides an inaccurate guess of the AGV's pose $\bar{\xi}$ with typical estimate error over 20m, images from the monocular camera and data from the radar provide complementary sensory information of the surroundings.

As depicted in Fig. 2, the system first extracts features of radar data and visual images respectively and then fuse them into an integrated bird's eye view (BEV) representation, and subsequently infer the AGV's absolute pose through matching the BEV and the neural map generated by the OSM in a probabilistic approach. In this paper, we assume that we will have synchronized camera and radar data for pose estimation.

C. Camera Encoder

Images full of geometric and semantic information are useful for localization in urban environment and hence cameras are adopted as a critical sensor modality of our geo-localization system. Like OrienterNet, the raw images from RGB channels are processed by the image encoder, where a convolutional neural network (CNN) is employed to extract BEV features F_{camera} from the RGB images.

D. Radar Data & BEV Feature Extraction

Since we are interested in estimating the 3 DoF poses, a 2D scanning radar providing X-Y plane information is suffice here. The 2D scanning radar sensor rotates its antenna and reports targets across 360° at every sampling cycle so it naturally captures the scene information in a BEV manner.

Radar data: The 2D raw data comes in polar form and we transform it into a Cartesian format before entering the radar channel for feature extraction, see Fig. 3. The building layout and the curbs are visible on the Cartesian image.

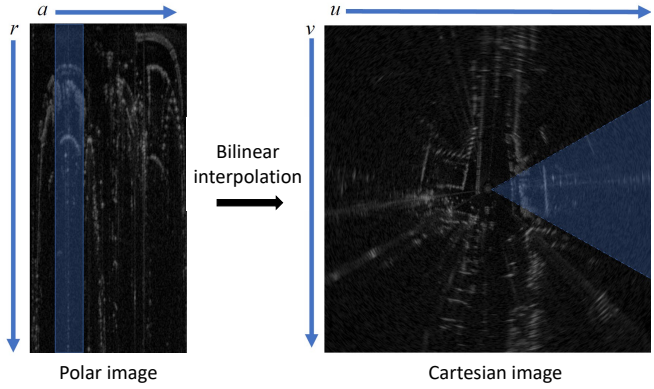

Fig. 3: Radar data

Radar encoder: To extract feature from radar Cartesian image, we employ ResNet50 as the backbone network, followed by a Feature Pyramid Network (FPN), to generate the neural BEV representation F_{radar} from radar. In contrast to the camera module, the depth prior and polar projection parts are not included since the radar Cartesian image is already in BEV form.

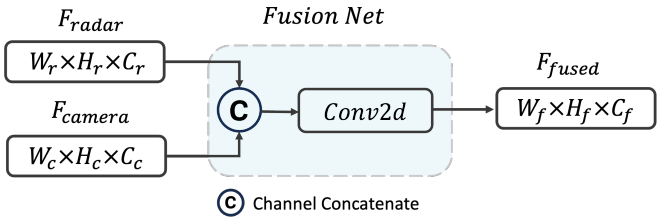

Fig. 4: Fusion Module

E. BEV Fusion

To associate image and radar features, a late fusion scheme is used here. Our fusion model is shown in Fig. 4. This module aims at fusing the radar and image BEV features and sending the fused feature to the subsequent matching module. First, the cross-modal features are concatenated for fusion. Then, in order to get more essential features, we use a multi-layer convolutional network to output a radar and image fused BEV feature grid. This fused grid is then processed by a CNN, the final fused neural BEV denoted as F_{fused}. The integration of both BEV features ensures that the fused feature maps are rich in semantic content while maintaining spatial precision in the BEV domain.

F. Neural Map and Alignment

OpenStreetMap[1]: In response to a given query region, OSM describes this region with semantic precision, organizing the geographical data into polygons for area representations, polylines for linear features, and singular points for discrete

[1]www.openstreetmap.org

TABLE I: Localization Recall on Day Sequence

	Approach	Position R@Xm			Position R@Ym			Orientation R@X°		
		1m	3m	5m	1m	3m	5m	1°	3°	5°
Boreas	Camera-only[3]	87.45	95.66	97.16	49.64	76.56	85.64	30.78	75.61	87.77
	Radar only	44.36	55.56	68.75	9.55	17.28	24.55	12.87	29.83	37.65
	LiDAR-camera	90.69	95.58	96.92	52.57	80.66	88.16	33.15	74.03	86.35
	Ours	**96.13**	**97.87**	**98.18**	**63.35**	**91.16**	**94.63**	**37.1**	**81.61**	**93.13**
Oxford	Camera-only[3]	20.54	43.91	54.63	6.13	10.72	14.55	4.74	13.71	22.27
	Radar only	10.49	31.49	44.92	1.94	6.47	11.21	0.84	2.2	3.49
	LiDAR-camera	28.81	53.12	62.65	11.85	21.72	28.31	**10.9**	**27.53**	37.53
	Ours	**34.33**	**55.5**	**63.49**	**19.16**	**30.64**	**38.36**	9.77	26.86	**38.58**

Boreas sequences: 01: 2021-01-15-12-17, 02: 2021-03-02-13-38, 03: 2021-03-30-14-23.
Oxford sequences: 01: 2019-01-10-11-46-21, 02: 2019-01-17-12-48-25.

locations. In the context of our system, the map consists of totaling 18 semantics extracted from the OSM, like "building", "parking", "road", "playground" and "curb".

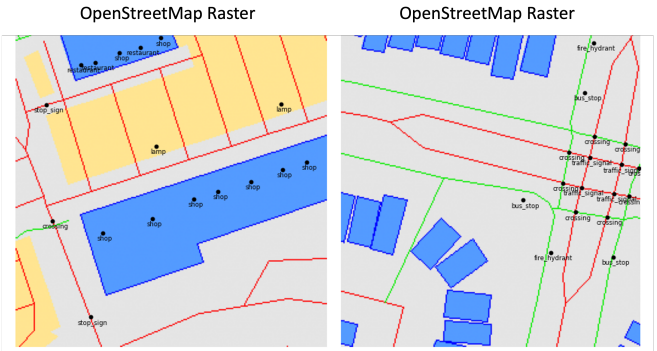

Fig. 5: The map legend uses different symbols and colors to represent various features in the mapping system.

Map encoder: In the map encoder module, we also adopted the framework from OrienterNet [3] to encode the map data (as shown in Fig.5). The encoding process involves associating each class with an N-dimensional embedding to create a feature map F_{map}, which provides the neural representation of the map.

Image-map matching: Finally, we take the fused BEV features and create multiple templates by rotating N times. Then we do *Template Matching* to exhaustively match the neural map F_{map} with the fused BEV F_{BEV}, resulting in the generation of a $W_p \times H_p \times N$ probability volume P.

G. Loss Functions

The discrete probability volume represents the likelihood of each spatial point and its orientation, enabling the estimation of the final 3-DoF pose. Thus, let $\mathbf{x}_{\text{gt}}$ be the ground truth positions, and θ_{gt} be the ground truth angles. We can define a Negative Log Likelihood (NLL) loss function as:

$$\text{NLL} = -\log P(\mathbf{x}_{\text{gt}}, \theta_{\text{gt}}; \mathbf{\Theta})$$

Accordingly, the learned parameters Θ of the deep model maximizes the likelihood function of pose estimate given the ground truth.

IV. EXPERIMENT

We follow the same evaluation protocol as detailed in OrienterNet [3] to asses the geo-localization accuracy. The recall performance is investigated which is *the higher the better*. We conducted experiments on two public datasets: Oxford Radar RobotCar Dataset [24] and the Boreas Dataset [25].

A. Datasets

Training data: For training, we selected three daytime sequences (01, 02 and 03) from the Boreas dataset, dividing them into training, validation, and test sets with an 8:1:1 ratio. From the Oxford dataset, we choose two sequences (01 and 02), ensuring that the routes selected do not overlap at all, which are used for finetuning the model.

Testing data: We assess the system's performance under varying conditions by using the Boreas dataset for daytime evaluation, and other weather conditions evaluation on sequences not included in the training set. Regarding the Oxford dataset, we meticulously split the route to ensure that there is *no overlap between the test set and the training set*.

Figure 6 illustrates the effectiveness of our multimodal fusion model by displaying comparative prediction examples from the same scenes under four varied weather conditions.

B. Competing Methods and Their Settings

Monocular: The SOTA visual geo-localization method on planimetric map is OrienterNet [3]. It possess powerful generalization capability on data from unseen cities. Specifically OrienterNet is trained on Mapillary Geo-Localization (MGL) Dataset and fine-tuned on Oxford and Boreas.

Radar baseline: To understand the necessity fusing visual data, we trained a radar baseline. The radar BEV feature is matched with the neural map directly.

LiDAR-camera fusion: LiDAR data inherently contains many reflection from the ground so it is necessary to remove them [26] before projecting the pointcloud onto X-Y plane. The LiDAR-camera fusion baseline shares the same amount of data with the Radar-Camera setting for fair comparison.

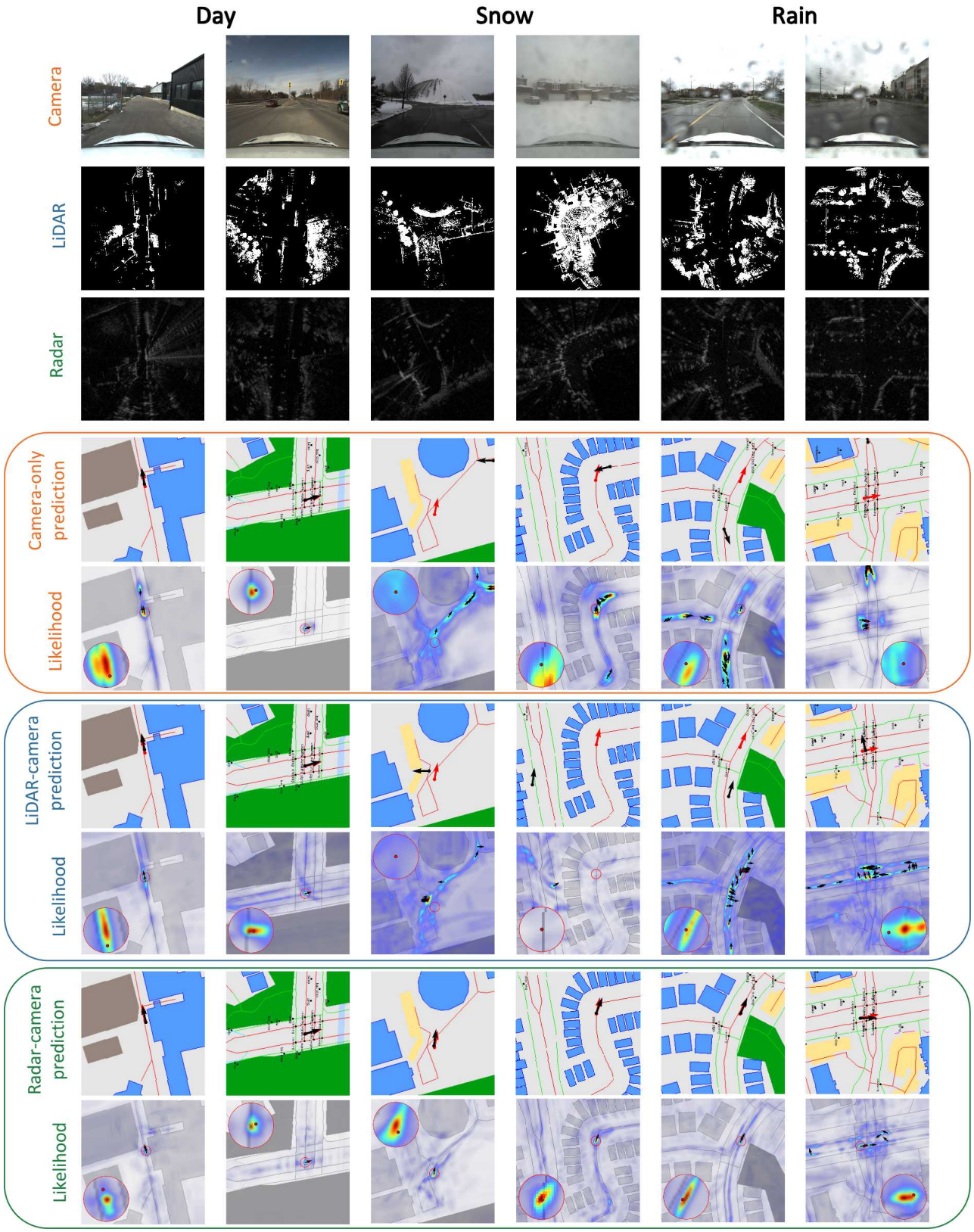

Fig. 6: **Our model is robust in all-weather conditions.** There are three different approaches mentioned in Table I: Camera-only (orange), LiDAR-camera (blue), and our Radar-camera approach (green). The visualization examples show the performance of these methods under different weather conditions. The input map is overlaid with single-image predictions (**black** arrows) and ground truth indicated (red arrows).

TABLE II: Localization Recall on Adverse Weather Sequence

Sequence	Approach	Position R@Xm			Position R@Ym			Orientation R@X°		
		1m	3m	5m	1m	3m	5m	1°	3°	5°
Light Snow-B04	Camera-only[3]	24.45	43.14	56.21	6.57	16.92	24.25	8.45	17.98	25.0
	Radar only	42.72	57.12	71.08	3.69	11.95	21.31	10.99	29.94	36.88
	LiDAR-camera	22.82	37.71	49.95	2.86	6.78	10.65	3.66	9.36	12.79
	Ours	**83.19**	**90.02**	**93.92**	**53.2**	**79.29**	**83.08**	**28.12**	**64.58**	**80.04**
Heavy Snow-B05	Camera-only[3]	9.24	21.2	35.92	2.39	6.02	10.67	3.1	7.17	10.67
	Radar only	38.77	56.25	65.96	4.3	11.82	17.13	12.47	29.94	37.07
	LiDAR-camera	20.17	32.27	42.96	3.2	8.07	11.42	2.12	5.47	7.89
	Ours	**70.21**	**83.85**	**86.45**	**43.88**	**70.01**	**74.86**	**23.73**	**56.94**	**71.04**
Rain-B06	Camera-only[3]	25.54	43.2	53.79	11.94	26.85	34.92	10.19	22.16	30.02
	Radar only	29.51	52.65	68.18	2.92	6.81	11.57	2.15	5.95	11.57
	LiDAR-camera	46.65	59.69	70.52	18.22	30.86	36.76	12.9	28.2	34.62
	Ours	**77.51**	**85.51**	**89.62**	**40.91**	**64.47**	**70.8**	**29.58**	**59.78**	**70.68**

Boreas test sequences: 04: 2020-12-01-13-26, 05: 2021-01-26-11-22, 06: 2021-04-29-15-55.

C. Implementation

We trained our network on 4 Nvidia RTX 4090 GPUs with a batch size of 4, setting the learning rate to 0.00001 for over 40 epochs to ensure thorough learning. Pre-generated Map tiles extract geographic information from GPS data and define bounding boxes for fetching OpenStreetMap tiles. We adopted a pixel density of 2 ppm for map tiles with a size of 128×128. For rotation during templating, we used 64 rotations and increased to 256 for testing. The training dataset includes over 10,000 frames from various scenes.

D. Results and Discussion

Evaluation of model performance: Table I shows the test results across two datasets employing four different methodologies in the daytime condition without snow or rain or fog. Camera-only [3], as SOTA, achieves high accuracy while the radar-only approach yields less accurate localization due to its higher noise and sparser information. The fusion of LiDAR and camera data slightly improves performance compared to the Camera-only approach as expected. However, our radar and vision fusion model significantly outperforms both the vision-only and LiDAR-camera approaches. The superior performance of our radar-camera system over the LiDAR-camera counterpart is attributed to broader sensing range of radars which mitigates ambiguities in perception.

The location accuracy obtained on the Oxford benchmark is not as good as that obtained on the Boreas dataset. There are two reasons of that. First, the distinct difference of geography and architecture features in Oxford dataset from that in the Boreas dataset, constituting shifts in domain distribution, leads to performance degradation in the camera-based deep learing models. Second, *big errors sometimes occur in GPS data (used as ground truth)* while the GPS trajectory data from Boreas is highly optimized with other sensors like wheel odometry.

The results indicate that our method comparatively outperforms the other two. Despite the substantial differences between two datasets, the model of our method trained on the Boreas dataset shows commendable generalization to the Oxford dataset, highlighting the model adaptability.

Performance in challenging conditions: Training is performed on day sequences from the Boreas dataset, with testing sequences obtained under three adverse weather conditions: light snow, heavy snow, and rain respectively.

Table II details the recall rates for each method, showing our radar-camera fusion based method outperforms sharply over all other three alternatives under various adverse weather conditions.

Fig. 7 charted the max localization errors of all four methods under both adverse and normal weather conditions. The radar-only and our radar-camera fusion based methods, compared with other two methods using cameras and LiDARs, have the least drop in localization accuracy under severe weather conditions, which proves that radars are less susceptible to adverse weather impacts. Our radar-camera fusion method further distinguishes itself by the smallest latitude and longitude errors in all four different weather conditions, compared with all other alternatives.

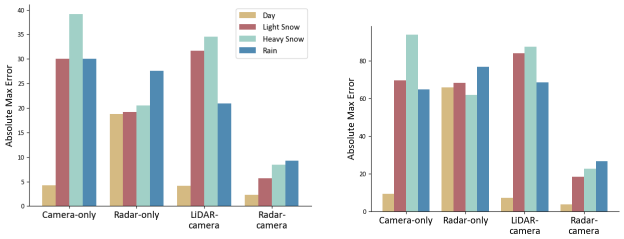

Fig. 7: The bar chart displays the maximum positional XY errors (left) and maximum yaw errors (right) of four different methods under daytime, snowy, and rainy conditions.

Reasons of visual failures in adverse weather conditions: In adverse weather conditions, visual localization often encounters challenges due to obstructions from rain, snow, or fog, as well as significant environmental variations. These factors can reduce the accuracy of depth estimation and

the extraction of BEV features. However, as experiments have demonstrated, the robustness of radar compensates for these shortcomings. Furthermore, radar sensors provide highly robust information from the BEV perspective. In an end-to-end training process, the map encoder can extract richer and more robust information from the map for localization.

V. Conclusion

This research presents an innovative HD map-free geo-localization system that integrates data from two complementary sensors of radars and cameras and use a freely-accessible OSM as the prior map. Besides three distinct feature extraction modules of the camera, the radar, and the OSM, a compact convolutional neural network is designed for feature-level fusion within the BEV domain. The proposed method enhances robustness of the geo-localization system to adverse weathers and achieves high localization accuracy even with only a OSM.

References

[1] Y. Cho, G. Kim, S. Lee, and J.-H. Ryu, "Openstreetmap-based lidar global localization in urban environment without a prior lidar map," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 4999–5006, 2022.

[2] T. Y. Tang, D. De Martini, S. Wu, and P. Newman, "Self-supervised learning for using overhead imagery as maps in outdoor range sensor localization," *The International Journal of Robotics Research*, vol. 40, no. 12-14, pp. 1488–1509, 2021.

[3] P.-E. Sarlin, D. DeTone, T.-Y. Yang, A. Avetisyan, J. Straub, T. Malisiewicz, S. R. Bulò, R. Newcombe, P. Kontschieder, and V. Balntas, "Orienternet: Visual localization in 2d public maps with neural matching," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2023, pp. 21 632–21 642.

[4] R. Arandjelovic, P. Gronat, A. Torii, T. Pajdla, and J. Sivic, "Netvlad: Cnn architecture for weakly supervised place recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 5297–5307.

[5] T. Sattler, B. Leibe, and L. Kobbelt, "Fast image-based localization using direct 2d-to-3d matching," in *2011 International Conference on Computer Vision*. IEEE, 2011, pp. 667–674.

[6] A. Kendall, M. Grimes, and R. Cipolla, "Posenet: A convolutional network for real-time 6-dof camera relocalization," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, December 2015.

[7] D. DeTone, T. Malisiewicz, and A. Rabinovich, "Superpoint: Self-supervised interest point detection and description," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2018, pp. 224–236.

[8] P.-E. Sarlin, D. DeTone, T. Malisiewicz, and A. Rabinovich, "SuperGlue: Learning feature matching with graph neural networks," in *CVPR*, 2020.

[9] H. Yin, Y. Wang, X. Ding, L. Tang, S. Huang, and R. Xiong, "3d lidar-based global localization using siamese neural network," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 4, pp. 1380–1392, 2019.

[10] L. Sun, D. Adolfsson, M. Magnusson, H. Andreasson, I. Posner, and T. Duckett, "Localising faster: Efficient and precise lidar-based robot localisation in large-scale environments," in *2020 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2020, pp. 4386–4392.

[11] X. Chen, T. Läbe, A. Milioto, T. Röhling, O. Vysotska, A. Haag, J. Behley, and C. Stachniss, "OverlapNet: Loop Closing for LiDAR-based SLAM," in *Proceedings of Robotics: Science and Systems (RSS)*, 2020.

[12] H. Yin, Y. Wang, L. Tang, and R. Xiong, "Radar-on-lidar: metric radar localization on prior lidar maps," in *2020 IEEE International Conference on Real-time Computing and Robotics (RCAR)*. IEEE, 2020, pp. 1–7.

[13] H. Yin, R. Chen, Y. Wang, and R. Xiong, "Rall: end-to-end radar localization on lidar map using differentiable measurement model," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 6737–6750, 2021.

[14] K. Burnett, Y. Wu, D. J. Yoon, A. P. Schoellig, and T. D. Barfoot, "Are we ready for radar to replace lidar in all-weather mapping and localization?" *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 10 328–10 335, 2022.

[15] Y. Tian, C. Chen, and M. Shah, "Cross-view image matching for geo-localization in urban environments," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[16] L. Li, Y. Ma, K. Tang, X. Zhao, C. Chen, J. Huang, J. Mei, and Y. Liu, "Geo-localization with transformer-based 2d-3d match network," *IEEE Robotics and Automation Letters*, vol. 8, no. 8, pp. 4855–4862, 2023.

[17] G. Floros, B. Van Der Zander, and B. Leibe, "Openstreetslam: Global vehicle localization using openstreetmaps," in *2013 IEEE International Conference on Robotics and Automation*. IEEE, 2013, pp. 1054–1059.

[18] P. Ruchti, B. Steder, M. Ruhnke, and W. Burgard, "Localization on openstreetmap data using a 3d laser scanner," in *2015 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2015, pp. 5260–5265.

[19] M. A. Brubaker, A. Geiger, and R. Urtasun, "Map-based probabilistic visual self-localization," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 4, pp. 652–665, 2015.

[20] O. Vysotska and C. Stachniss, "Exploiting building information from publicly available maps in graph-based slam," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 4511–4516.

[21] F. Yan, O. Vysotska, and C. Stachniss, "Global localization on openstreetmap using 4-bit semantic descriptors," in *2019 European Conference on Mobile Robots (ECMR)*. IEEE, 2019, pp. 1–7.

[22] N. Samano, M. Zhou, and A. Calway, "You are here: Geolocation by embedding maps and images," 2020.

[23] Z. Hong, Y. Petillot, K. Zhang, S. Xu, and S. Wang, "Large-scale radar localization using online public maps," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 3990–3996.

[24] D. Barnes, M. Gadd, P. Murcutt, P. Newman, and I. Posner, "The oxford radar robotcar dataset: A radar extension to the oxford robotcar dataset," in *IEEE International Conference on Robotics and Automation (ICRA)*. Paris: IEEE, 2020, pp. 6433–6438.

[25] K. Burnett, D. J. Yoon, Y. Wu, A. Z. Li, H. Zhang, S. Lu, J. Qian, W.-K. Tseng, A. Lambert, K. Y. Leung, A. P. Schoellig, and T. D. Barfoot, "Boreas: A multi-season autonomous driving dataset," *arXiv preprint arXiv:2203.10168*, 2022.

[26] S. Lee, H. Lim, and H. Myung, "Patchwork++: Fast and robust ground segmentation solving partial under-segmentation using 3d point cloud," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 13 276–13 283.