Global Bird's Eye View Semantic Mapping and Localization

Hefeng Zhou, Ding Zhong, Yiwei Gui, and Yiting Wang
Electrical and Computer Engineering Department, University of Michigan

Abstract

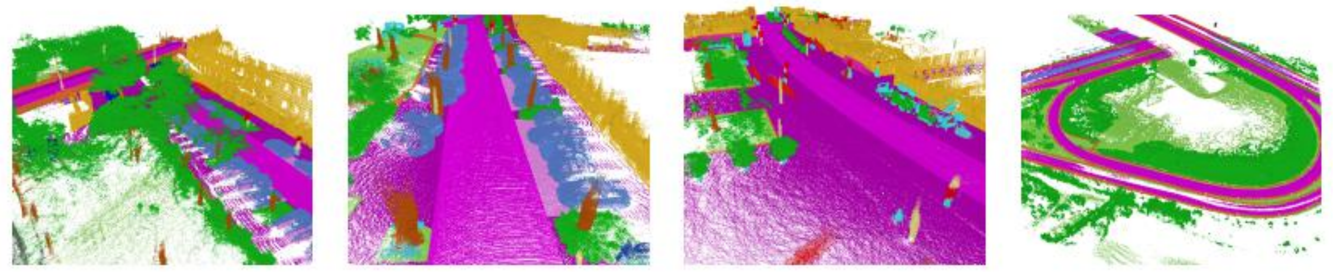
- We integrate AI-IMU[1] and MotionNet[2] to achieve robust localization and semantic scene understanding. AI-IMU dynamically adapts to varying sensor noise and motion patterns.MotionNet generates dense semantic BEV maps and effectively filters dynamic obstacles.
- Experimental validation on the SemanticKITTI dataset demonstrates strong robustness and accuracy in challenging scenarios, including dynamic objects and sensor dropout.

Motivation

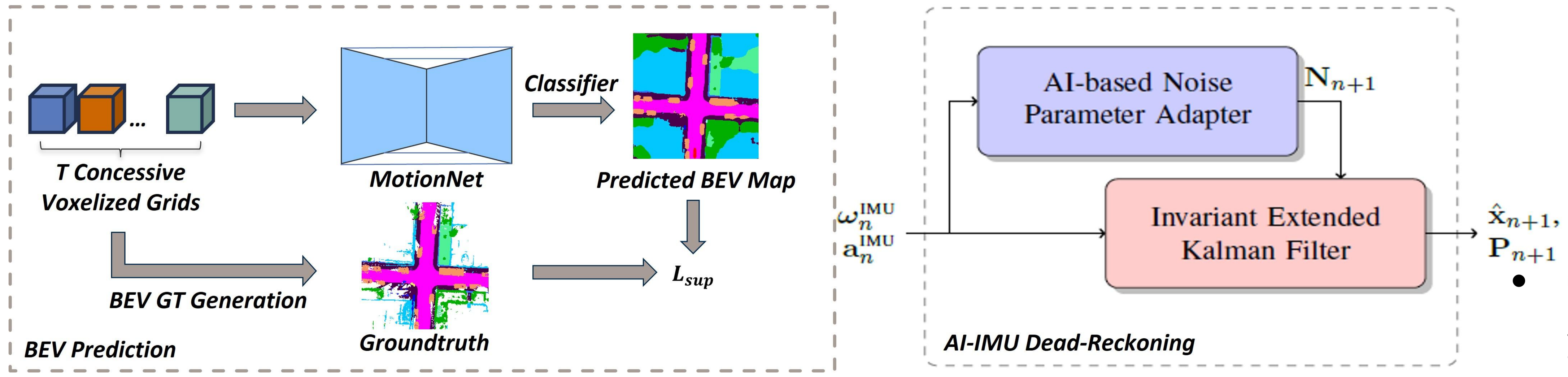
- For autonomous driving, behavior prediction and planning are typically done in the bird's eye view
- IMUs suffer from drift and error accumulation when used in isolation.

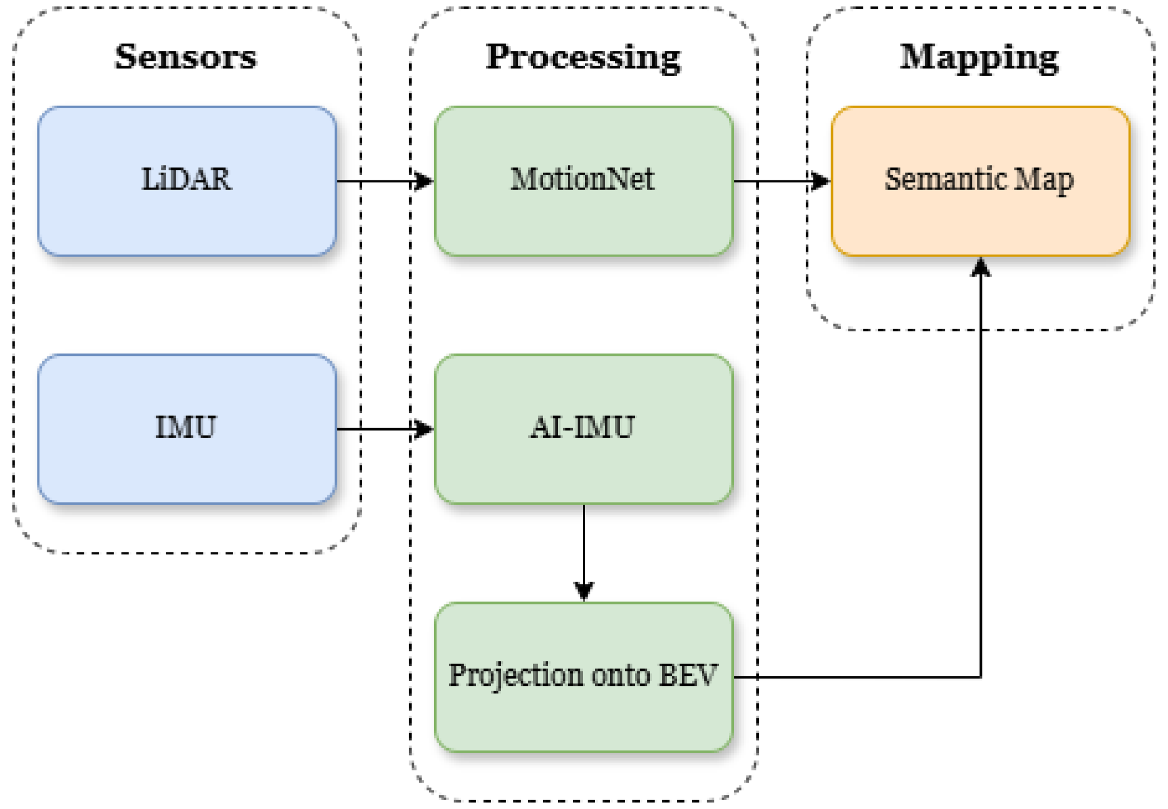
Data Sets

- We evaluate our method on the SemanticKITTI dataset



Methodology



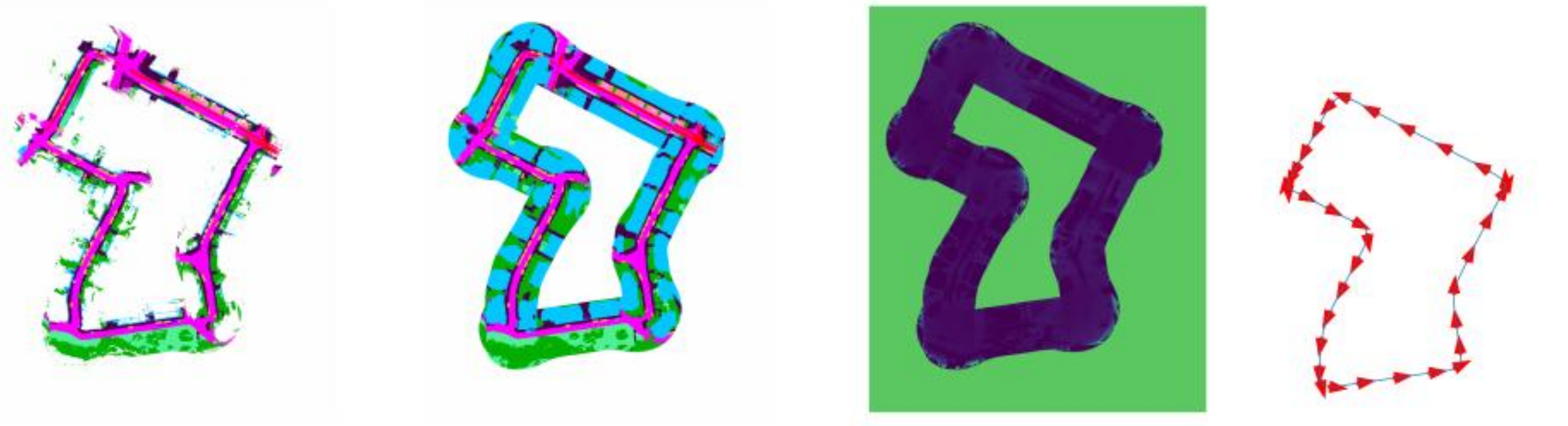


- MotionNet processes LiDAR point clouds to produce accurate semantic BEV maps using a Spatio-Temporal Pyramid Network (STPN).
- AI-IMU employs an Invariant Extended Kalman Filter (IEKF) enhanced by CNN-adaptive pseudo-measurement covariance for precise IMU-based localization.
- Semantic maps from MotionNet are spatially aligned with accurate poses estimated by AI-IMU, enabling consistent semantic mapping.
- Our integrated framework effectively captures dynamic scenes and maintains robustness even when LiDAR or visual data are compromised.

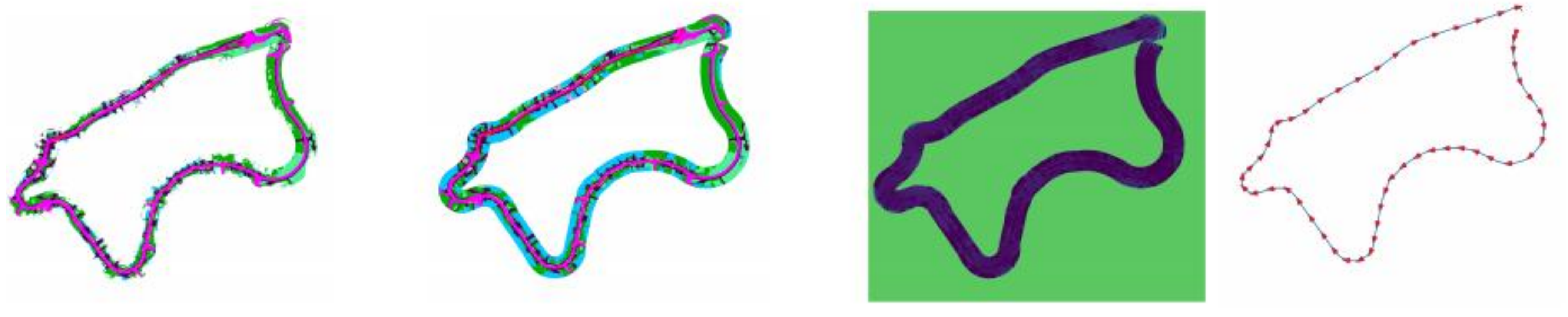
Results

Class ID	car	road	parking	sidewalk	building	fence	vegetation	terrain
Baseline IoU (%)	**74.8**	82.8	24.6	53.0	16.0	15.6	**49.3**	50.3
Ours IoU (%)	71.59	**87.09**	**40.32**	**67.07**	**19.63**	**29.63**	45.91	**64.44**
mIoU (%)			Ours: 53.21		Baseline: 45.80			

- Our method improved IoU scores on crucial urban classes, such as roads, sidewalks, and terrains.
- Compared with baseline model (single-frame prediction), our model demonstrates a superior performance (53.21% in mIoU) across 8 classes, yielding a 7.41% mIoU increase. These results highlight our model's strong capability in critical urban segmentation scenarios.



(a) Ground Truth Semantic Map (b) Predicted Semantic Map (c) Prediction Variance (Uncertainty) (d) Trajectory with Heading Angles



(a) Ground Truth Semantic Map (b) Predicted Semantic Map (c) Prediction Variance (Uncertainty) (d) Trajectory with Heading Angles

[1] M. Brossard, A. Barrau, and S. Bonnabel, "Ai-imu dead-reckoning,"IEEE Transactions on Intelligent Vehicles, vol. 5, no. 4, pp. 585–595,2020
[2] P. Wu, S. Chen, and D. N. Metaxas, "Motionnet: Joint perceptionand motion prediction for autonomous driving based on bird's eyeview maps," in Proceedings of the IEEE/CVF Conference on ComputerVision and Pattern Recognition (CVPR), June 2020

ECE ELECTRICAL & COMPUTER ENGINEERING
UNIVERSITY OF MICHIGAN