Carnegie Mellon University
The Robotics Institute

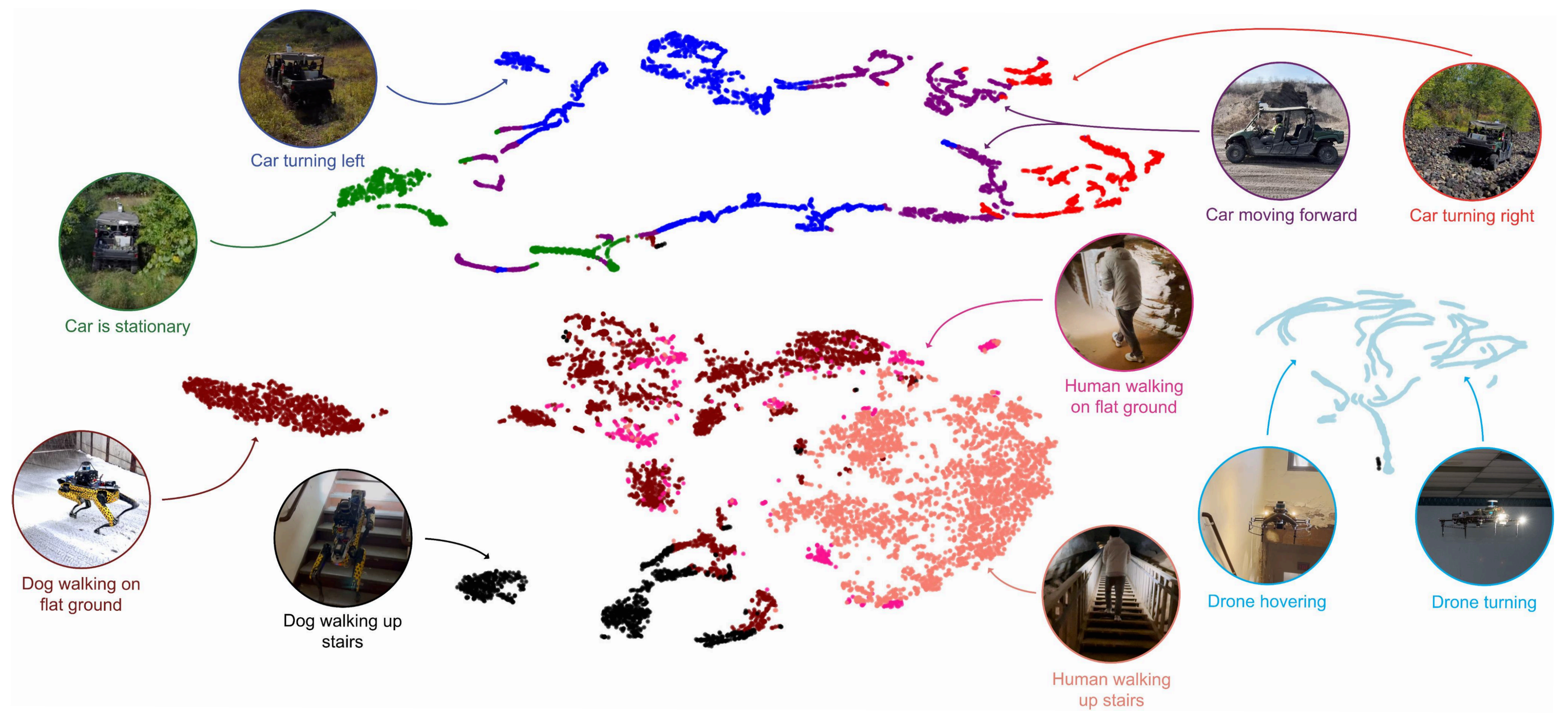
Tartan IMU: A Light Foundation Model for Inertial Positioning in Robotics

Shibo Zhao†, Sifan Zhou†, Raphael Blanchard, Yuheng Qiu, Wenshan Wang, Sebastian Scherer
{shiboz,sifanz,basti}@andrew.cmu.edu

CVPR Nashville JUNE 11-15, 2025

Motivation

How to achieve generalizable learning IMU odometry across different robotic platforms?



t-SNE Visualization of TartanIMU: Trained on 100+ hours of IMU data from ground vehicles, drones, legged robots, and humans, our foundation model embeds IMU signals into a unified high-dimensional space, capturing generalizable motion patterns.

Limitations of Existing Work:

- Rely on double integration, leading to drift.
- Often locomotion- or device-specific; fail to generalize out-of-distribution.
- Struggle to adapt to new motion platforms.
- Generalization on IMU data remains largely unexplored.

Contribution:

- **IMU Foundation Model:** scalable, cross-platform motion estimation with a shared backbone.
- **Efficient Fine-tuning:** LoRA-enabled rapid adaptation to new platforms.
- **Online Adaptation:** dynamic motion pattern selection for real-time training.

Method

Stage 1 : Pre-trained IMU Model

Acceleration in body frame: $^{B}\mathbf{a}_n = ^{B}_{W}\mathbf{R}_n(^{W}_{B}\mathbf{R}_n(\mathbf{a}-\mathbf{b}_a) - ^{W}\mathbf{g})$

Gyroscope in body frame: $^{B}\boldsymbol{\omega}_n = \boldsymbol{\omega} - \mathbf{b}_g$

Loss Function: $$L_{RL}^{MSE}(\mathbf{v},\hat{\mathbf{v}}) = \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=m}^{m+M}(v_{j\to j+1} - \hat{v}_{j\to j+1})\right)$$

Covariance Function: $$L^{\mathrm{NLL}} = \frac{1}{2}(v-\hat{v})^T\Sigma^{-1}(v-\hat{v}) + \frac{1}{2}\ln|\Sigma|$$

Body Velocity Prediction: $$(\hat{\mathbf{v}},\hat{\mathbf{u}}) = f\left(\left(^{B}\mathbf{a}_{n-N}, ^{B}\boldsymbol{\omega}_{n-N}\right),\ldots,\left(^{B}\mathbf{a}_n, ^{B}\boldsymbol{\omega}_n\right), \mathbf{h}_{n-N}\right)$$

Stage 2 : Fine-tune for Unseen Environments

LORA Adapter Network: $h = W_0x + \Delta Wx = W_0x + BAx$

Stage 3 : Online Adaptation

Gaussian Mixture Model: $$\gamma_k(x) = \frac{\pi_k\mathcal{N}(x \mid \mu_k, \Sigma_k)}{\sum_{j=1}^{K}\pi_j\mathcal{N}(x \mid \mu_j, \Sigma_j)}$$

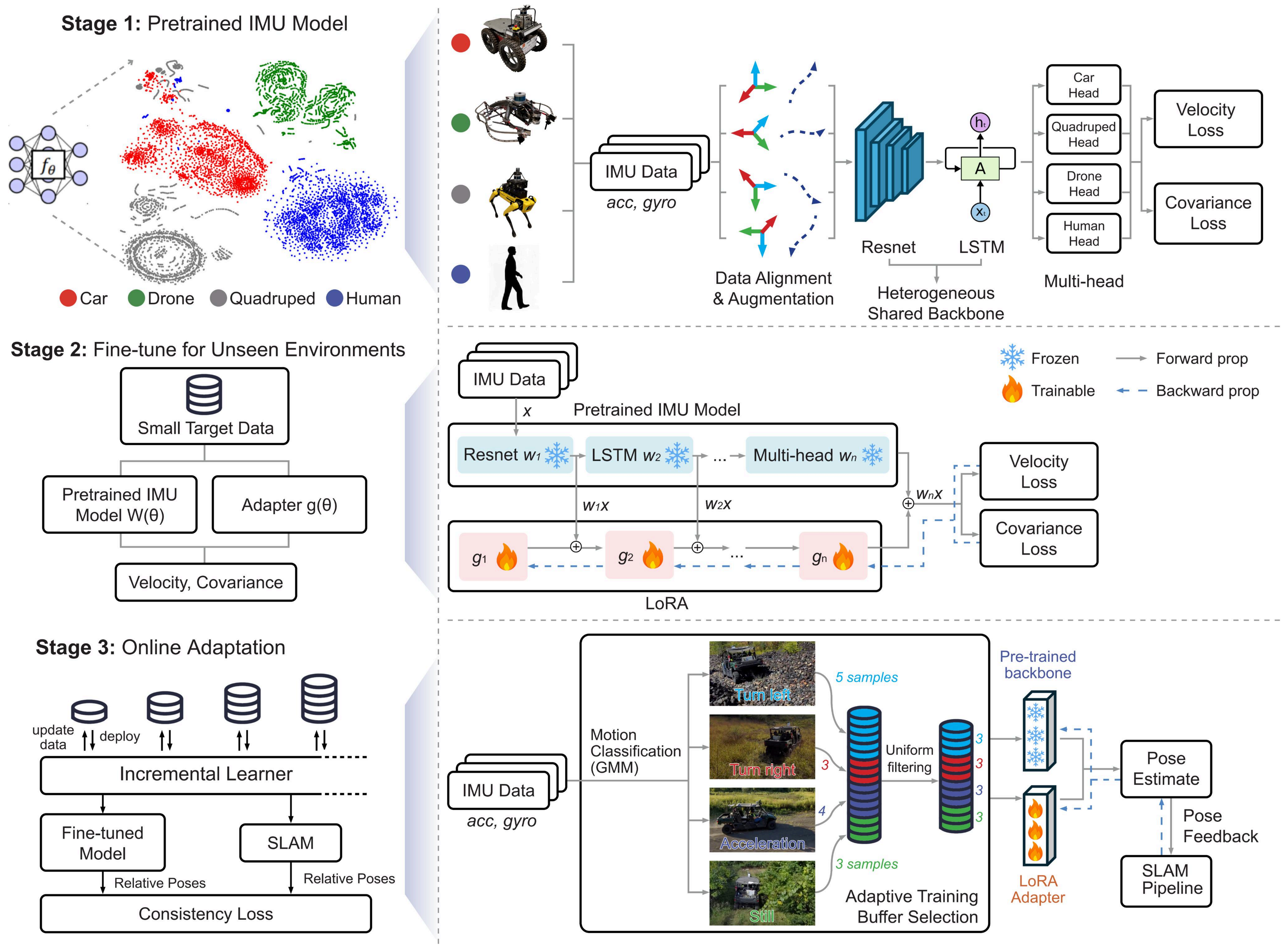
Future Work & Limitation

- Support more arbitrary robotic platforms (Mixture of Expert Model)
- Integrate sim dataset for more generalizable model
- GMM not working well for complex motion pattern

Notations

- $f(\cdot)$ represents the neural network function that processes inputs from IMU sensors.
- $\mathbf{a}$ denotes acceleration from the IMU.
- $\boldsymbol{\omega}$ is angular velocity from the IMU.
- $\mathbf{h}_{n-N}$ refers to the hidden state produced by the LSTM at the previous time step.
- $^{W}\mathbf{g} = [0, 0, 9.8]$ refers to gravity vector.
- $\mathbf{B}$ refers to the body frame.
- $\hat{\mathbf{v}}$ refers to estimated relative velocity.
- $\hat{\mathbf{u}}$ refers to uncertainty of relative velocity.
- $\mathbf{b}_a$ refers to acceleration bias.
- $\mathbf{b}_g$ refers to gyro bias.
- $W_0 \in \mathbb{R}^{d\times k}$ refers to pretrained weight matrix
- $\gamma_k(x)$ refers to posterior probability that x belongs to cluster k
- π_k refers to mixing coefficient.
- μ_k refers to mean of gaussian for cluster k
- Σ_k refers to cov matrix of gaussian for cluster k
- $\mathcal{N}(x \mid \mu_k, \Sigma_k)$ refers to multivariate Gaussian PDF
- K refers to number of clusters

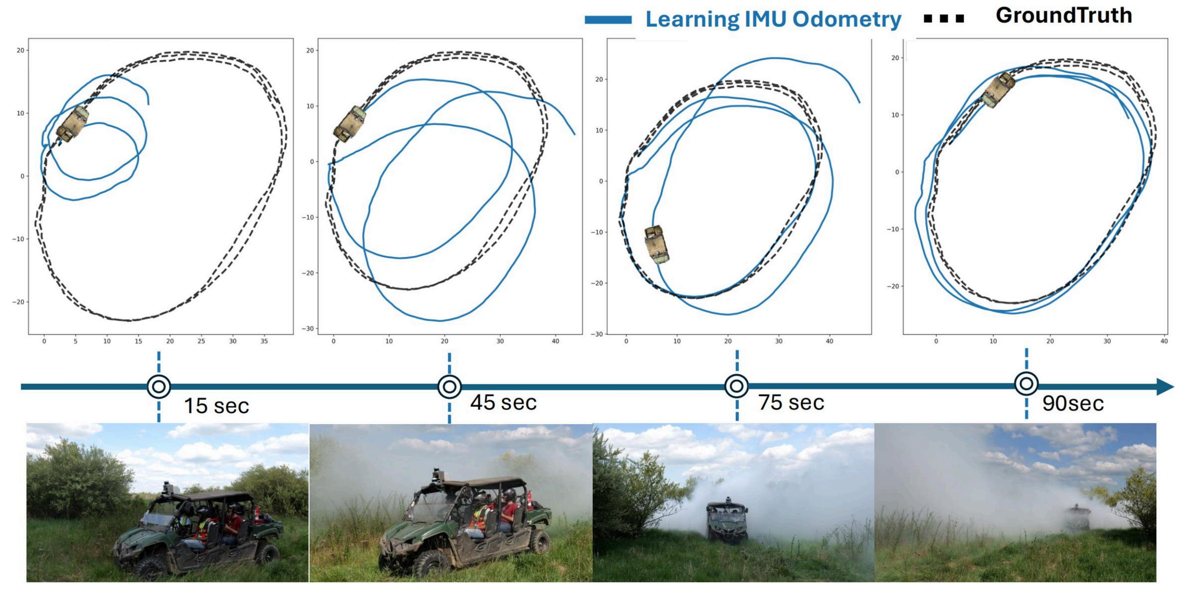
Pipeline



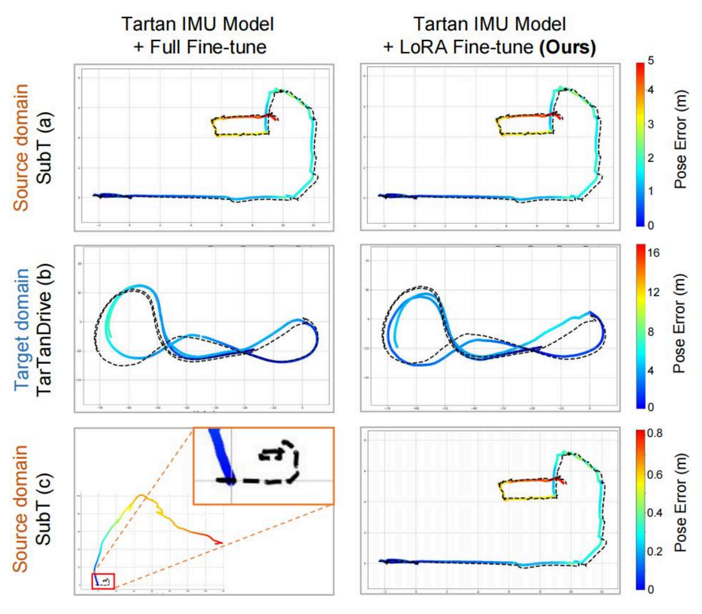
Three Learning Stages of Tartan IMU

a) Pre-trained IMU Model: features a shared backbone to capture generalizable IMU knowledge.
b) Efficient Fine-Tuning: utilizes an adapter to enable positive transfer for new tasks.
c) Online Adaptation: employs an adaptive memory buffer to support on-the-fly model updates during deployment.

Performance of Online Adaptation on Unseen Trajectory

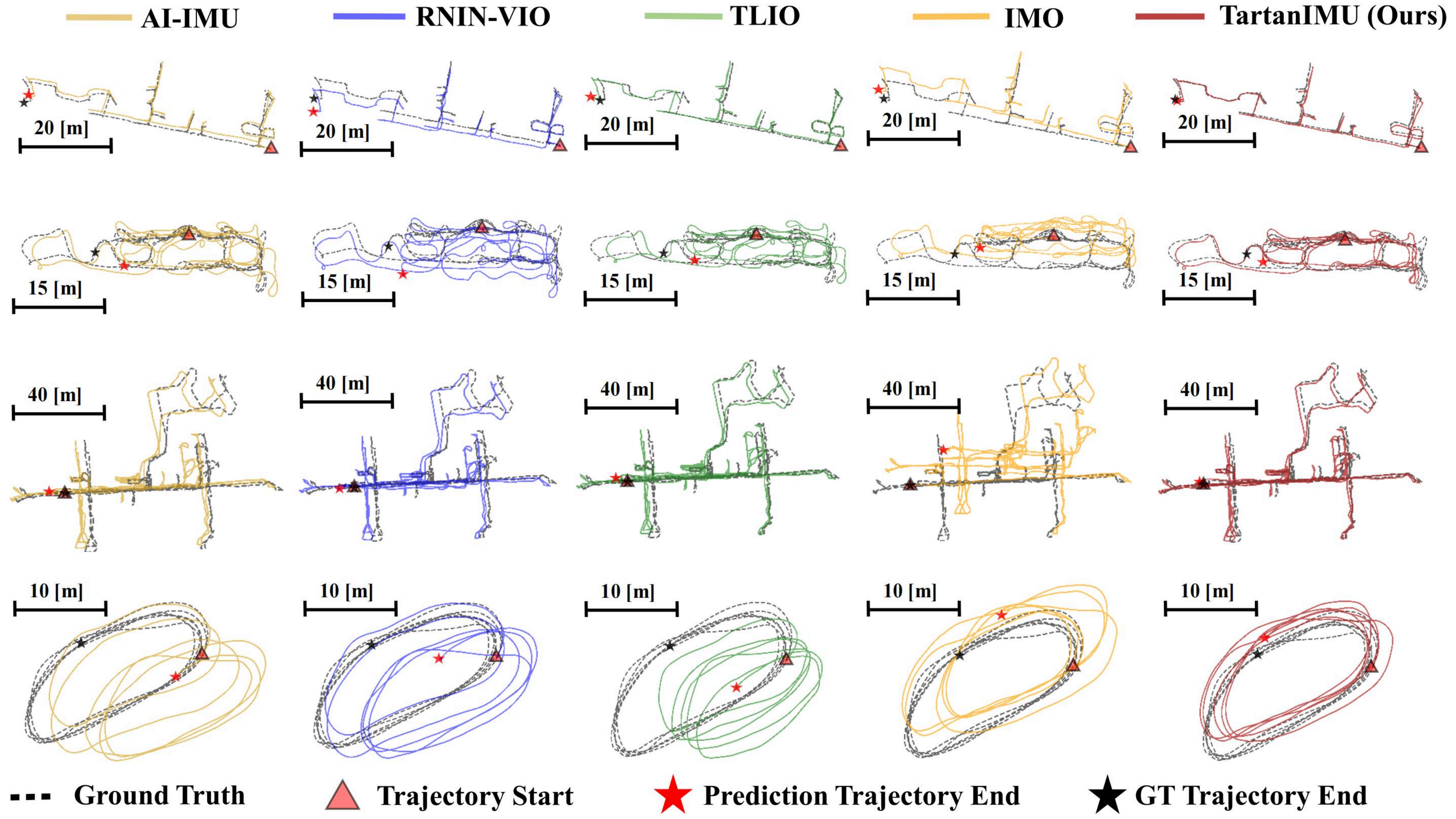


Non-forgetting of Our Model



Experiments

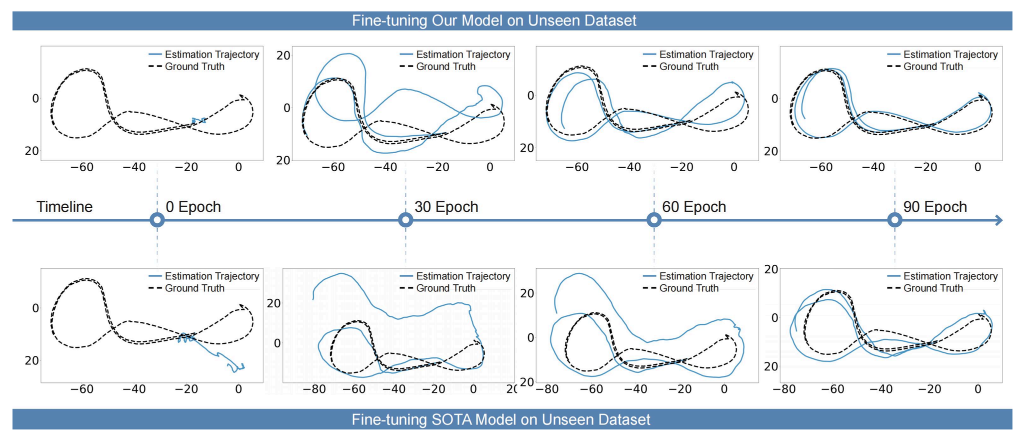
Qualitative Analysis of Our Proposed TartanIMU Compared to Other Specialized IMU Models



Quantitative Analysis of Our Proposed TartanIMU Compared to Other Specialized IMU Models

Robot Platform	AI-IMU [6]		RNIN-VIO [13]		TLIO [25]		IMO [14]		TartanIMU		Improvement	
	ATE ↓	T-RTE ↓	ATE ↓	T-RTE ↓	ATE ↓	T-RTE ↓	ATE ↓	T-RTE ↓	ATE ↓	T-RTE ↓	ATE	T-RTE
Wheeled (Car)[49]	7.68	3.33	7.82	5.06	8.12	3.73	8.12	3.73	**6.17**	**2.52**	↑ 19.63 %	↑ 24.36%
Handled (Human)[35]	8.26	4.89	7.62	5.61	6.96	4.82	10.19	5.67	**4.32**	**1.95**	↑ 47.69%	↑ 60.05%
Legged (Dog)[49]	3.23	1.60	3.10	1.58	3.61	1.73	3.35	1.64	**1.46**	**0.79**	↑ 54.83%	↑ 50.63%
Aerial (Drone)[2]	4.14	1.45	4.32	1.51	3.93	1.40	3.72	1.34	**3.32**	**1.04**	↑ 19.81%	↑ 28.97%

Fine-tuning Comparison between Ours and SOTA



Accuracy Comparison of Model Trained w/ Data from Single vs. All-robots

Robot Platform	Tartan IMU					
	ATE ↓			T-RTE ↓		
	Single	All	Improve	Single	All	Improve
Wheeled[49]	7.51	6.17	↑22%	3.17	2.52	↑26%
Handled[35]	6.64	4.32	↑54%	4.24	1.95	↑117%
Legged[49]	3.64	1.46	↑150%	1.94	0.79	↑146%
Aerial[2]	3.84	3.32	↑16%	1.36	1.04	↑30%
Average	5.41	3.82	↑42%	2.68	1.57	↑71%

Performance of Online Adaptation on Memory Buffer