GeoFlow: Real-Time Fine-Grained Cross-View Geolocalization via Iterative Flow Prediction

CVPR 2026

Ayesh Abu Lehyeh, Xiaohan Zhang, Ahmad Arrabi, Waqas Sultani, Chen Chen, Safwan Wshah

University of Vermont

CVPR JUNE 3-7, 2026 DENVER COLORADO

VaiL Vermont Artificial Intelligence Laboratory

The Core Problem (FG-CVG)

Query ground view

x , y , θ ?

Satellite overhead map

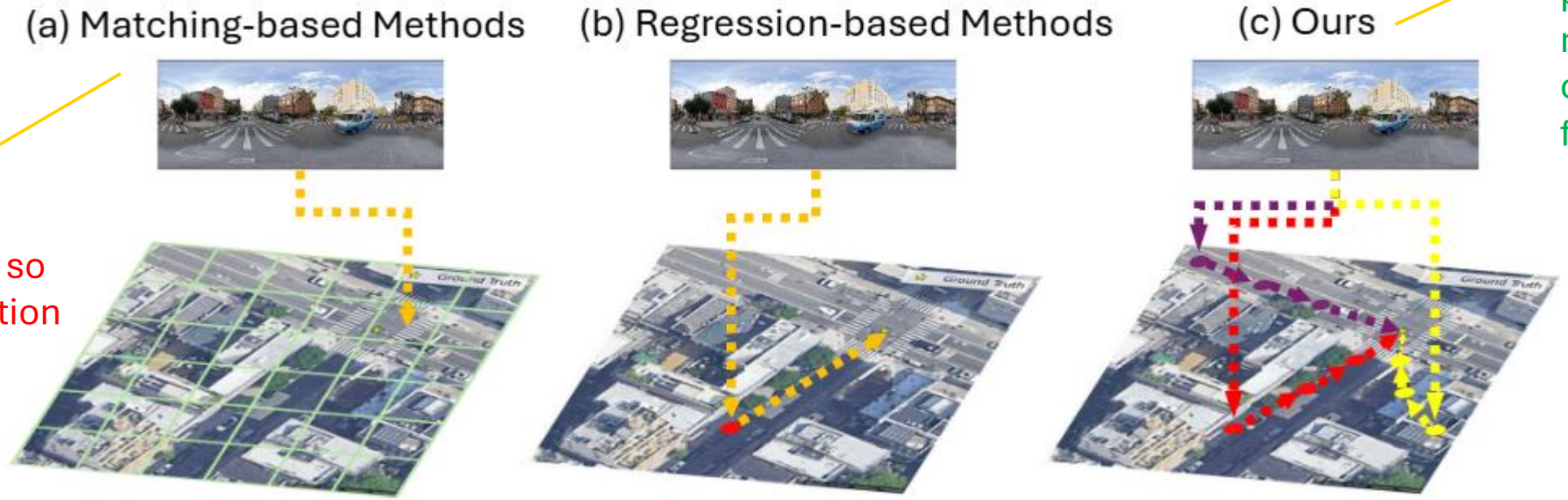
The Goal of Fine-Grained Cross View Geolocalization (FG-CVG): Estimate precise **3-DoF (location (x , y) + orientation(θ))** from ground-to-satellite images.

University of Vermont

Motivations

Continuous but computationally heavy (often needing camera intrinsics to work as well).

GeoFlow: Direct probabilistic mapping via continuous flow.



Discretizes search space so high quantization error (lower accuracy)

The problem: Current methods force a harsh **trade-off between accuracy and efficiency**. High-accuracy models are too slow for real-time use and lack real world deployment requirements.

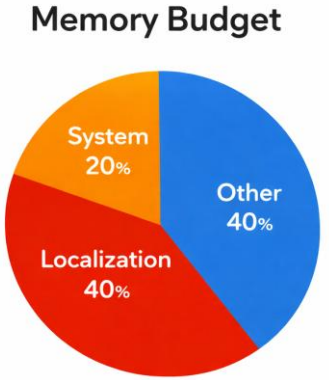
University of Vermont

The Edge Deployment Reality ("Why GeoFlow?)

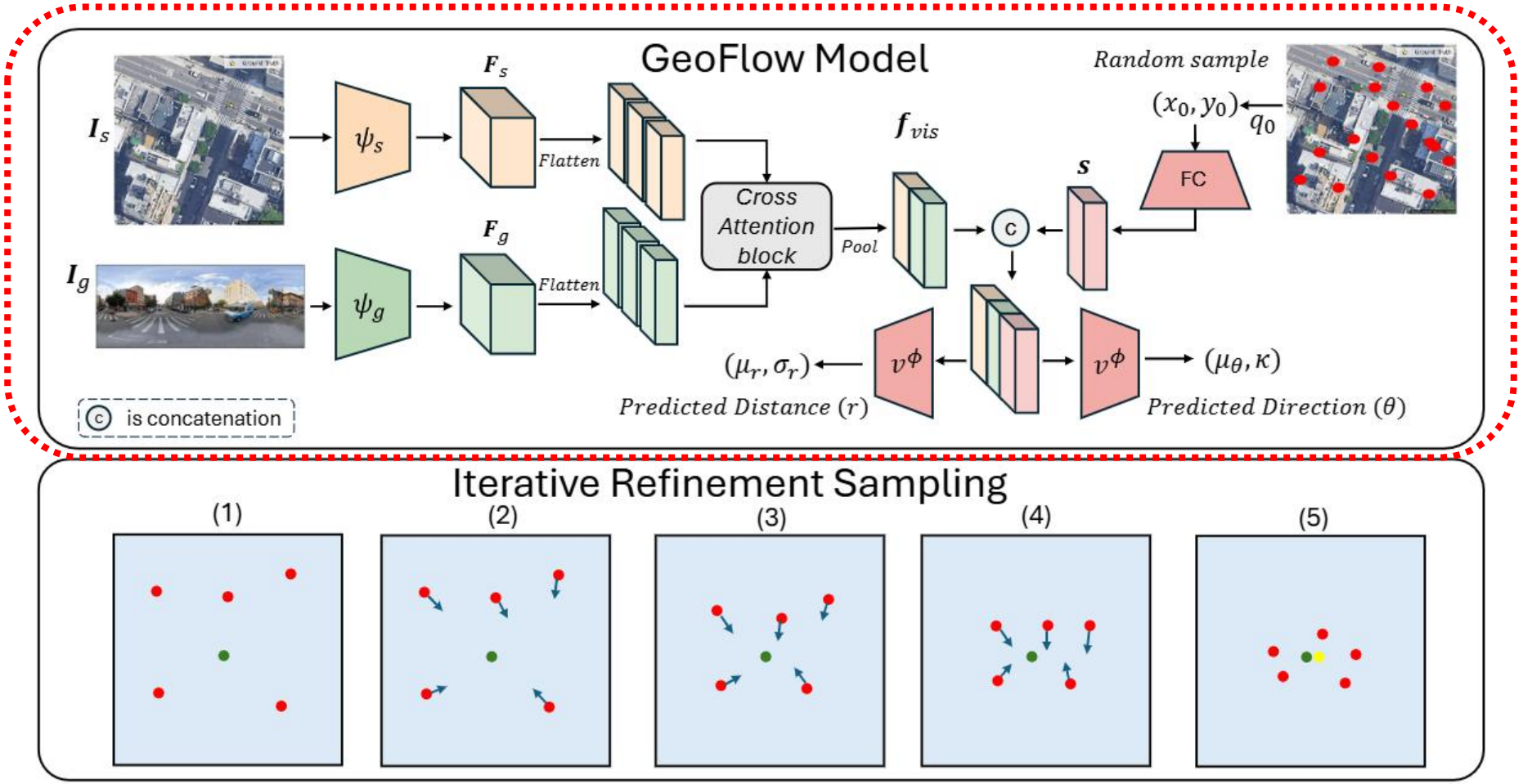
Targeting edge devices (Ex., Sub-250g Micro-UAVs) that need real-time speed.

Strict hardware limits: 2GB to 4GB RAM.

Concurrent tasks (OS, other tasks) consume >2GB.



GeoFlow Architecture

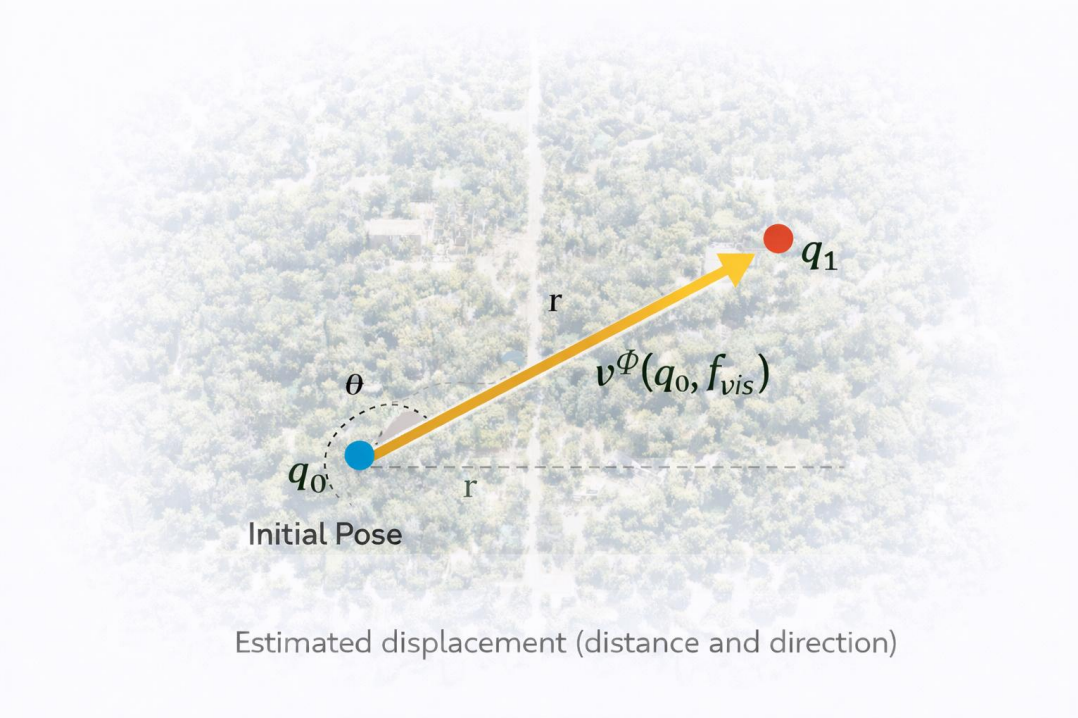


University of Vermont

Probabilistic Displacement Regression

The Goal: Learn a regression field $v^{\phi}(q_0, f_{vis})$ to estimate the displacement (distance and direction) from an initial pose q_0 to the target q_1.

The Decoder: A simple MLP processes z and branches into two distinct probabilistic prediction heads.



Objective Functions

The Distance Head:

- **Distribution:** Predicts parameters for a standard Gaussian distribution $\mathcal{N}(\mu_r, \sigma_r^2)$.
- **Loss Function (Gaussian NLL):**

$$\mathcal{L}_r = \frac{1}{2}\left(\frac{(r_{gt} - \mu_r)^2}{\sigma_r^2} + \log \sigma_r^2\right)$$

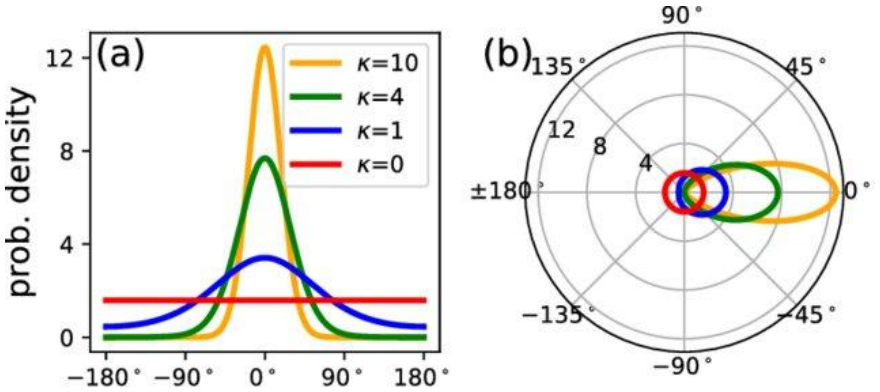
The Direction Head (von Mises-Fisher)

- **The Problem with L2:** Standard Euclidean losses and NLL fail on angles due to circular discontinuity (the jump from 359 to 0).
- **The Solution:** The von Mises-Fisher (vMF) distribution models data on the unit hypersphere S^1.
- **Probability Density:**

$$p(u|\mu_\theta, \kappa) = C_2(\kappa)\exp\left(\kappa \mu_\theta^T u\right)$$

- **Loss Function (Angular MF Loss):**

$$\mathcal{L}_\theta = -\log(\kappa^2 + 1) + \kappa \cdot \cos^{-1}\left(\mu_\theta^T \cdot \theta_{gt}\right) + \log(1 + \exp(-\kappa\pi))$$



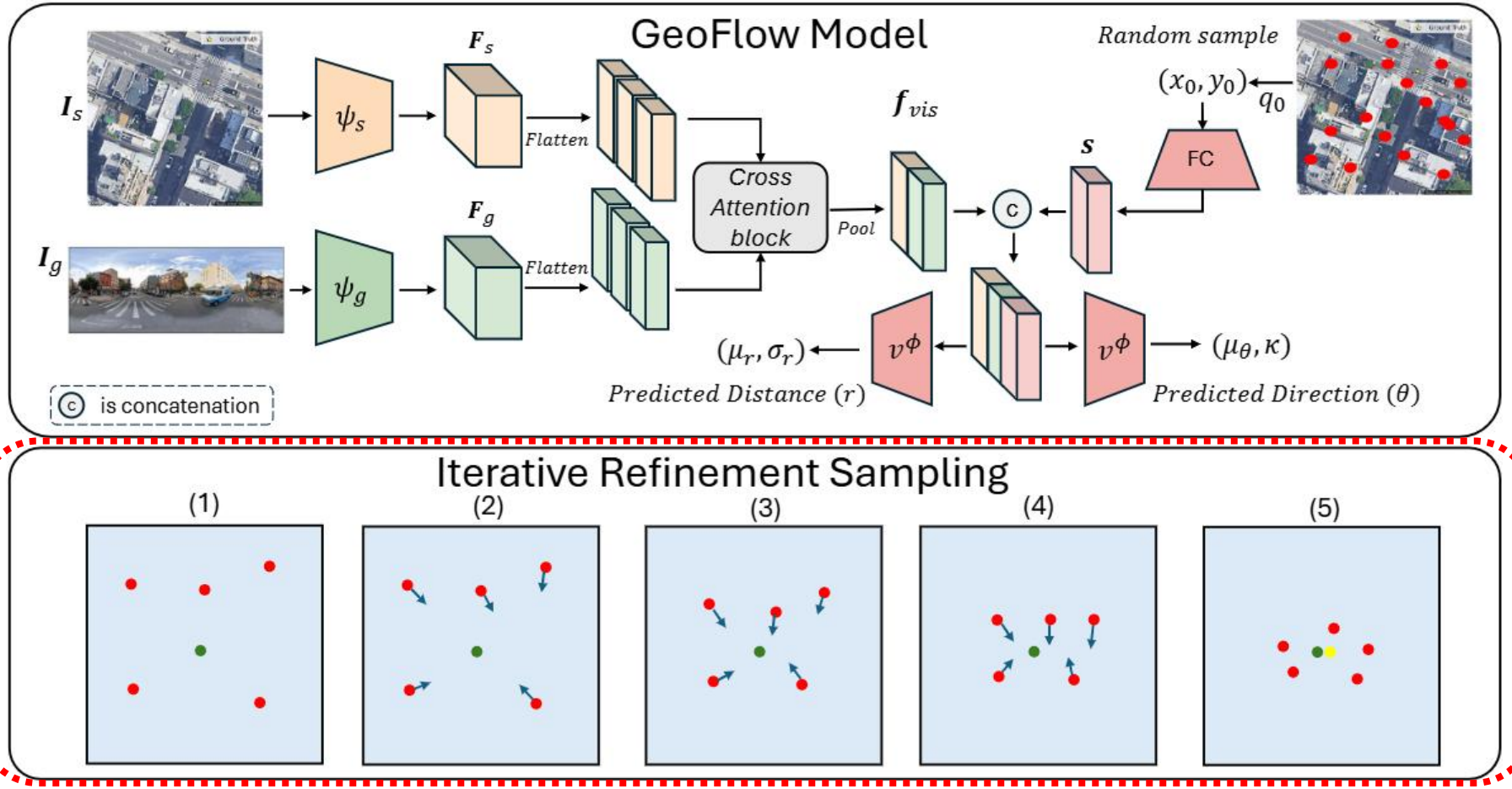
***κ* acts as the concentration parameter (confidence).**

The Output: The combined training objective is :

$$\mathcal{L} = \mathcal{L}_r + \mathcal{L}_\theta.$$

University of Vermont

Iterative Refinement Sampling (IRS)



Inference time

Quantitative Results

Table 1. **KITTI** test results comparison. We report localization error (Mean/Median, m), lateral/longitudinal recall (R@1m/R@5m, %), and inference speed (FPS) for both Same-Area and Cross-Area splits. Best results are in **bold**.

	Efficiency	Same-area						Cross-area					
	FPS↑	↓Loc. (m)		↑Lateral (%)		↑Long. (%)		↓Loc. (m)		↑Lateral (%)		↑Long. (%)	
Method		Mean	Median	R@1m	R@5m	R@1m	R@5m	Mean	Median	R@1m	R@5m	R@1m	R@5m
GGCVT [14]	4.17	–	–	76.44	98.89	23.54	62.18	–	–	57.72	91.16	14.15	45.00
CCVPE [23]	24.00	1.22	0.62	97.35	99.71	77.13	97.16	9.16	**3.33**	44.06	90.23	23.08	64.31
HC-Net [18]	25.00	0.80	0.50	99.01	99.73	**92.20**	**99.25**	8.47	4.57	75.00	97.76	**58.93**	**76.46**
DenseFlow [16]	7.30	1.48	**0.47**	95.47	99.79	87.89	94.78	7.97	3.52	54.19	91.74	23.10	61.75
FG^2 [21]	4.20	**0.75**	0.52	**99.73**	**100.00**	86.99	98.75	**7.45**	4.03	**89.46**	**99.80**	12.42	55.73
GeoFlow (Ours)	**29.49**	0.98	0.68	96.85	99.68	74.05	98.75	8.42	5.60	36.36	83.85	14.76	52.51

Outperforms CCVPE in mean accuracy (8.42m vs 9.16m) on KITTI Cross-Area while achieving real time speed.

Inference-Time Scaling

Table 3. Impact of refinement rounds (R) on KITTI Cross-Area.

Rounds (R)	↓Mean (m)	↓Median (m)	↑Speed (FPS)
1	10.69	9.95	**32.55**
3	8.47	5.88	31.23
5	8.42	5.60	29.49
10	**8.41**	**5.59**	26.23

Table 4. Impact of initial seeds (N) on KITTI Cross-Area.

Seeds (N)	↓Mean (m)	↓Median (m)	↑Speed (FPS)
1	8.58	5.75	**30.70**
5	8.49	5.66	30.05
10	8.42	5.60	29.49
20	**8.41**	**5.60**	28.08

Heavy lifting (visual context) is done once.
Dynamic trade-off between accuracy and speed.

State-of-the-Art Efficiency

VRAM: 686 MiB (7x reduction vs. CCVPE).

Speed: 26 ms (35% speedup in pure inference).

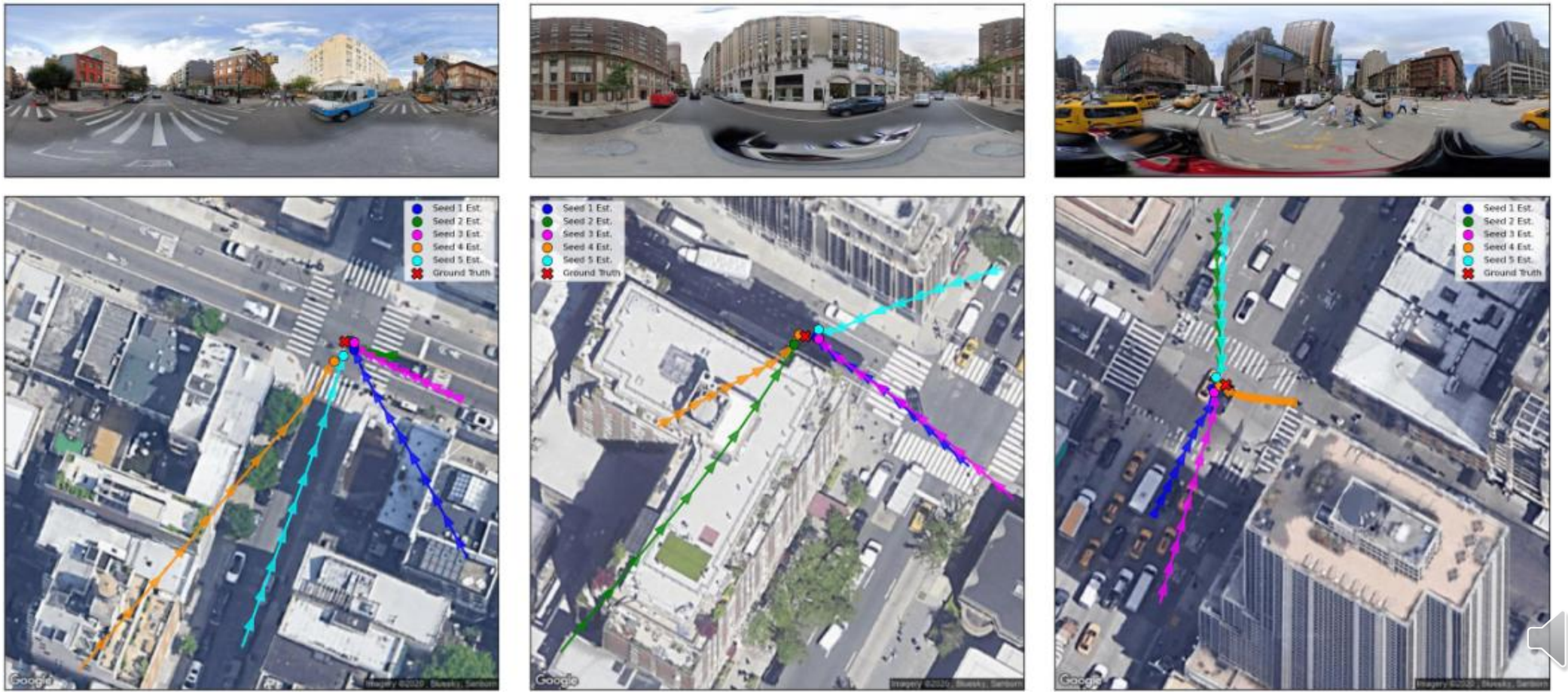
Compute: 7.65 GFLOPs (reduces thermal throttling and power consumption).

Table 1. Unified Efficiency Profile. GeoFlow balances accurac and efficiency.

Model	GFLOPs	VRAM	Inference Time	FPS
CCVPE	31.18	4730 MiB	41.7 ms	24.0
HC-Net	11.56	1900 MiB	40.0 ms	25.0
GeoFlow	**7.65**	**686 MiB**	**26.0 ms**	**29.5**
Gain (vs HC-Net)	**-34%**	**-64%**	**-35%**	**+18%**

University of Vermont

Qualitative Results



University of Vermont

Qualitative Results (Confidence)

$R = 1$ $R = 2$ $R = 3$ $R = 4$

IRS actively sculpts predictive uncertainty

Conclusion & Broader Impact

- **Summary**: GeoFlow successfully bridges the gap between SOTA benchmark accuracy and practical edge deployment constraints.

- **Innovation**: The probabilistic displacement field combined with Iterative Refinement Sampling (IRS) actively models confidence and consensus.

- **Future Work**: Extending the continuous flow principles to full 6-DoF estimation and downstream path-planning tasks for autonomous micro-robotics. We are also working now on a unified framework that gives an end-to-end geolocalization.

University of Vermont

Scan For Project Page