US DOT National
University Transportation Center for Safety

Project 564: Realtime Robot Localization and Pose Regression with Invertible Neural Networks

Rahul Mangharam (0000-0002-3388-8283)

Final Report – July 2025

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
564		

4. Title and Subtitle	5. Report Date
Project 564: Realtime Robot Localization and Pose Regression with Invertible Neural Networks	July 31, 2025
	6. Performing Organization Code University of Pennsylvania
7. Author(s) Rahul Mangharam, Ph.D. (0000-0002-3388-8283)	**8. Performing Organization Report No.** Enter any/all unique alphanumeric report numbers assigned by the performing organization, if applicable.
9. Performing Organization Name and Address University of Pennsylvania, Philadelphia, PA.	**10. Work Unit No.**
	11. Contract or Grant No. Federal Grant # 69A3552344811 / 69A3552348316
12. Sponsoring Agency Name and Address DoT UTC	**13. Type of Report and Period Covered** Final Report (July 1, 2024 – June 30, 2025)
	14. Sponsoring Agency Code USDOT

15. Supplementary Notes

Conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration.

Enter information not included elsewhere, such as translation of (or by), report supersedes, old edition number, alternate title (e.g. project name), hypertext links to documents or related information in the form of URLs, PURLs (preferred over URLs - https://archive.org/services/purl/help), DOIs (https://www.doi.org/), insertion of QR codes, copyright or disclaimer statements, etc. Edit boilerplate FHWA statement above if needed.

16. Abstract

This project explores novel approaches to robot localization and visual pose regression using Invertible Neural Networks (INNs). Addressing the critical need for efficient and accurate pose estimation in robotics, we propose two frameworks: Local_INN and PoseINN. Local_INN tackles the inverse problem of robot localization by providing an implicit map representation in its forward path and performing localization in the inverse path. It uniquely offers uncertainty estimation through latent space sampling and addresses the kidnapping problem with a global localization algorithm. PoseINN extends this work to real-time visual-based pose regression from camera data. By leveraging INNs and normalizing flows, PoseINN achieves state-of-the-art performance with significantly reduced computational costs, enabling faster training with low-resolution synthetic data and real-time deployment on mobile robots. Both frameworks demonstrate that INNs can effectively solve ambiguous inverse problems in robotics, providing robust and efficient solutions with inherent uncertainty quantification.

17. Key Words	18. Distribution Statement
Autonomous vehicles, robot localization, neural networks, normalizing flows	No restrictions. This document is available through the National Technical Information Service, Springfield, VA 22161. Enter any other agency mandated distribution statements. Remove NTIS statement if it does not apply.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	11	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Table of Contents

1. Project Abstract

This project explores novel approaches to robot localization and visual pose regression using Invertible Neural Networks (INNs). Addressing the critical need for efficient and accurate pose estimation in robotics, we propose two frameworks: Local_INN and PoseINN. Local_INN tackles the inverse problem of robot localization by providing an implicit map representation in its forward path and performing localization in the inverse path. It uniquely offers uncertainty estimation through latent space sampling and addresses the kidnapping problem with a global localization algorithm. PoseINN extends this work to real-time visual-based pose regression from camera data. By leveraging INNs and normalizing flows, PoseINN achieves state-of-the-art performance with significantly reduced computational costs, enabling faster training with low-resolution synthetic data and real-time deployment on mobile robots. Both frameworks demonstrate that INNs can effectively solve ambiguous inverse problems in robotics, providing robust and efficient solutions with inherent uncertainty quantification.

2. Project Overview

Robot localization, the process of determining a robot's pose (position and orientation) using sensor measurements and a map, is fundamental for autonomous navigation and interaction with the physical world. Similarly, visual pose regression, finding camera poses from images, is crucial for applications ranging from mobile robotics to augmented reality. Traditional geometric-based methods often incur high computational costs and latency, while many learning-based approaches suffer from low accuracy or long training times. This project investigates the application of Invertible Neural Networks (INNs) to overcome these limitations. INNs offer a unique advantage by providing bijective mappings between different data spaces, making them well-suited for inverse problems like localization and pose regression, and inherently allowing for uncertainty estimation through their probabilistic nature. The project aims to develop efficient, accurate, and robust solutions for these challenges, demonstrating their practical applicability in real-world robotic systems.

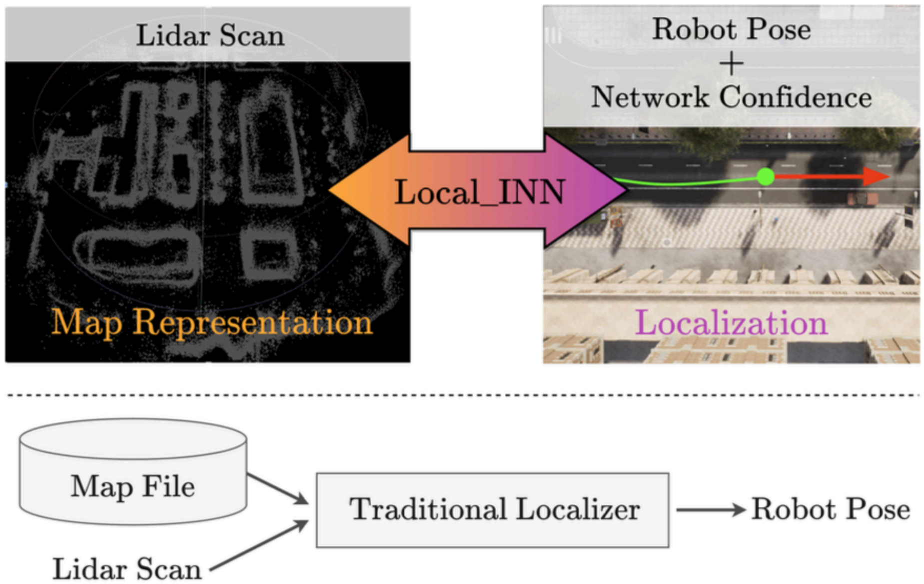


Fig. 1. Local_INN is a framework of localization with invertible neural networks. Compared to current localization methods, Local_INN stores map information within the neural network. Evaluation of Local_INN in forward direction gives compressed map information, and in the reverse direction gives accurate localization with fast runtime and uncertainty estimation.

3. Main Contributions

This project's core contributions are encapsulated in two distinct, yet complementary, frameworks: Local_INN and PoseINN, both leveraging the power of Invertible Neural Networks.

3.1. Local_INN: Implicit Map Representation and Localization

Local_INN introduces a novel framework for robot localization by formulating it as an inverse problem solved with Invertible Neural Networks.

3.1.1. Inverse Problem Formulation for Localization:
Local_INN frames robot localization as an inverse problem, where the INN's forward path learns an implicit map representation from robot poses, and its inverse path performs localization, mapping sensor measurements (e.g., LiDAR scans) back to robot poses. This contrasts with traditional methods that often rely on explicit map representations and complex probabilistic filters.

3.1.2. Implicit Map Representation:
A key innovation of Local_INN is its ability to learn and represent a map implicitly within the forward pass of the INN. This allows for a compact and flexible map representation that can be reconstructed in detail, even for poses exterior to the training set, demonstrating its generalization capabilities. The implicit nature avoids the need for explicit storage of large map structures.

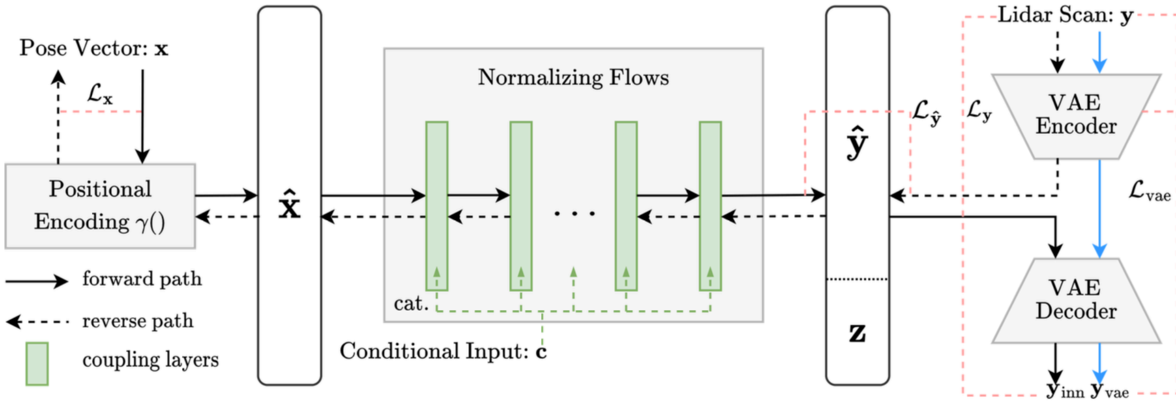


Fig. 2. Network Structure of the Local_INN. The forward path (solid arrows) is from pose to LiDAR scan. The reverse path (dashed arrows) is from LiDAR scan to robot pose. Conditional input is calculated from the robot's previous pose. The INN used in this paper has 6 coupling layers and the VAE encoder and decoder have 2 layers of MLPs for 2D LiDARs and plus 6 layers of 2D convolutions for 3D LiDARs.

3.1.3. Uncertainty Estimation with Latent Space Sampling:

By sampling the latent space during evaluation, Local_INN provides not just a single pose estimate, but also an associated covariance. This enables a robust estimation of the localization uncertainty, a crucial aspect for reliable autonomous navigation, which is often difficult to obtain directly from many learning-based methods.

3.1.4. Global Localization for Kidnapping Problem:

The framework includes a global localization algorithm designed to address the "kidnapping problem," where a robot loses its sense of location. By leveraging the INN's ability to map diverse inputs to corresponding poses, Local_INN can re-localize the robot effectively in previously unseen or ambiguous situations.

3.2. PoseINN: Realtime Visual-based Pose Regression and Localization

PoseINN extends the application of INNs to real-time visual-based pose regression, focusing on efficiency and practical deployment.

3.2.1. Visual Pose Regression using INNs:

PoseINN utilizes INNs to establish a mapping between the latent space of images and corresponding camera poses for a given scene. This allows for direct regression of ego-pose from camera inputs, bypassing the need for computationally expensive geometric pipelines.

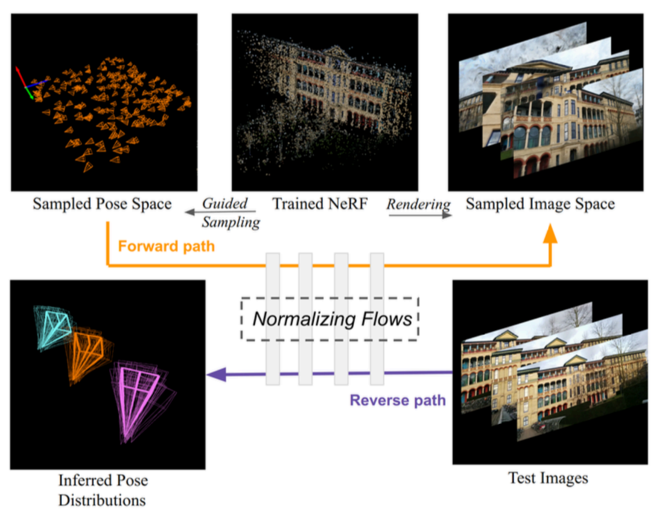


Fig. 1. We propose to learn a mapping between the latent space of the images and camera poses in an environment with an invertible neural network. We use NeRF to guide camera pose sampling and render synthetic images. Evaluating the reverse path of the INN outputs the full posterior distribution of camera poses given a test image.

3.2.2. Efficiency through Low-Resolution Synthetic Data Training:
A significant contribution is the model's ability to achieve high performance while being trained on offline rendered low-resolution synthetic data. This drastically reduces training time and computational resources, making the development and deployment of visual pose regression models more accessible and faster.

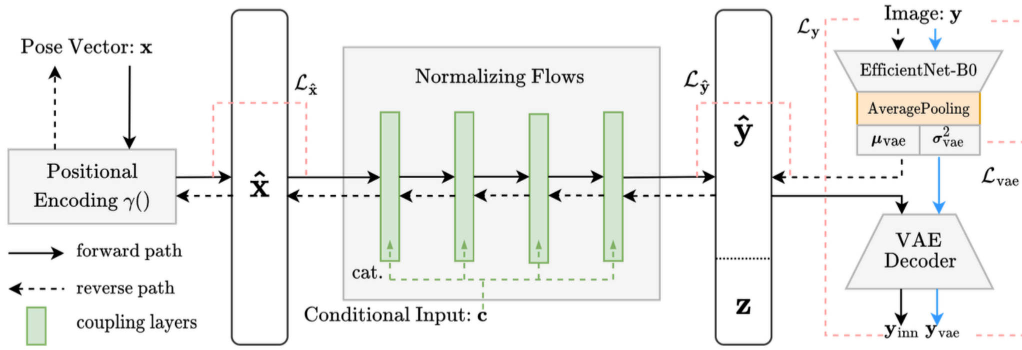


Fig. 3. Network Structure of the PoseINN. The forward path (solid) is from pose to image. The reverse path (dashed) is from image to pose.

3.2.3. Uncertainty Estimation via Normalizing Flows:
Similar to Local_INN, PoseINN leverages normalizing flows, an integral component of INNs, to inherently provide uncertainty estimation for the output poses. This allows the system to quantify its confidence in the pose predictions, which is vital for safety-critical applications.

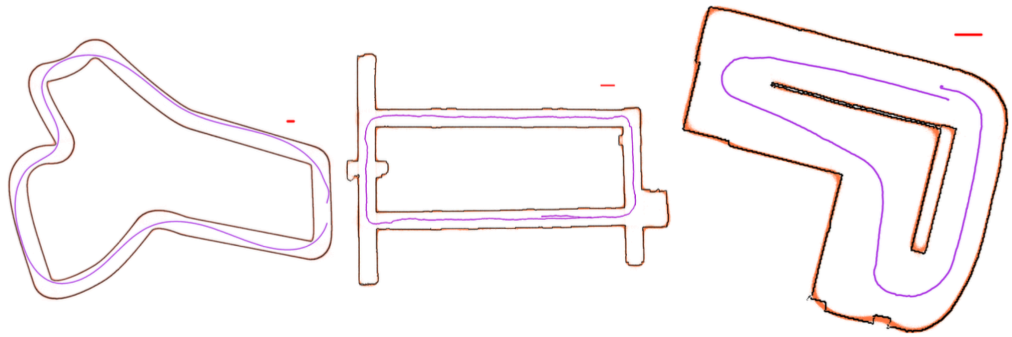
3.2.4. Mobile Robot Deployment and Efficiency:
The efficacy and efficiency of PoseINN are demonstrated through its successful deployment on a mobile robot. The model's low latency and computational requirements make it suitable for real-time applications on edge devices, paving the way for practical integration into robotic systems.

4. Results

The performance of both Local_INN and PoseINN frameworks was rigorously evaluated, demonstrating their effectiveness and efficiency.

4.1. Local_INN Results
Local_INN demonstrated localization performance on par with current state-of-the-art methods, but with significantly lower latency. This highlights its computational efficiency without sacrificing accuracy. Furthermore, the framework's ability to provide detailed 2D and 3D map reconstruction from learned implicit representations, even for poses outside the training set, showcased its robust generalization capabilities. The successful implementation of a global localization algorithm further validated its ability to handle challenging scenarios like the kidnapping problem.

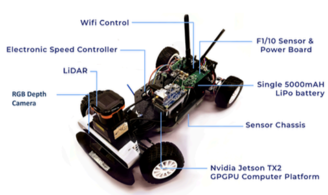
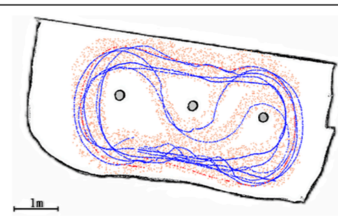
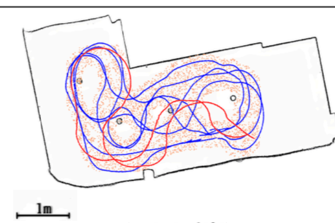
TABLE I
MAP RECONSTRUCTION AND LOCALIZATION ERRORS WITH 2D LIDAR

	Race Track (Simulation)		Hallway (Real)		Outdoor (Real)	
Original Map Reconstruction Test Trajectory						
	xy(m)	$\theta(^\circ)$	xy(m)	$\theta(^\circ)$	xy(m)	$\theta(^\circ)$
Online PF (1m/s)	$0.045 \pm \mathbf{0.058}$	0.400 ± 0.512	$\mathbf{0.039} \pm \mathbf{0.066}$	$\mathbf{0.482} \pm 0.808$	$\mathbf{0.013} \pm \mathbf{0.018}$	$\mathbf{0.358} \pm \mathbf{0.456}$
Local_INN (1m/s)	0.050 ± 0.102	0.201 ± 0.532	0.196 ± 0.433	$0.528 \pm \mathbf{0.792}$	0.034 ± 0.047	0.924 ± 1.130
↑ + EKF	$\mathbf{0.039} \pm 0.077$	0.182 ± 0.464	0.093 ± 0.139	0.536 ± 0.797	0.034 ± 0.047	0.917 ± 1.129
↑ + TensorRT	0.039 ± 0.076	$\mathbf{0.177} \pm \mathbf{0.443}$	0.104 ± 0.159	0.547 ± 0.802	0.033 ± 0.046	0.930 ± 1.142
Online PF (5m/s)	0.139 ± 0.168	1.463 ± 2.107	$\mathbf{0.071} \pm \mathbf{0.117}$	0.943 ± 1.738	0.033 ± 0.047	0.940 ± 1.371
Local_INN+EKF (5m/s)	$\mathbf{0.034} \pm \mathbf{0.056}$	$\mathbf{0.133} \pm \mathbf{0.284}$	0.100 ± 0.147	$\mathbf{0.565} \pm \mathbf{0.900}$	$\mathbf{0.032} \pm \mathbf{0.046}$	$\mathbf{0.915} \pm \mathbf{1.130}$

4.2. PoseINN Results

PoseINN achieved performance comparable to state-of-the-art visual pose regression models. Crucially, it accomplished this with considerably faster training times, primarily due to its ability to leverage low-resolution synthetic data. The real-time performance of PoseINN, as evidenced by its successful deployment on a mobile robot, underscored its efficiency and practical utility for on-device applications. The inherent uncertainty estimation provided by the normalizing flows also proved valuable, offering a measure of confidence in the regressed poses.

TABLE II
MEDIAN LOCALIZATION ERRORS WITH 2D LIDAR VS. CAMERA

Experiment Platform		Indoor	Outdoor
	train trajectory test trajectory sampled points		
		(xy[m], $\theta[^\circ]$)	(xy[m], $\theta[^\circ]$)
	Online PF (45Hz)	$\mathbf{0.01}, 0.23$	$\mathbf{0.02}, \mathbf{0.36}$
	PoseINN (**154Hz**)	$0.02, 0.31$	$0.12, 0.72$
	PoseINN + EKF	$0.02, \mathbf{0.22}$	$0.10, 0.65$

5. Authors

The research presented in this report was conducted by the following individuals:

Zirui Zang, Hongrui Zheng, Johannes Betz, Ahmad Amine and Rahul Mangharam

6. Bibliography

[1] Z. Zang, H. Zheng, J. Betz, and R. Mangharam, "Local_INN: Implicit Map Representation and Localization with Invertible Neural Networks," *arXiv preprint arXiv:2209.11925v1*, 2022.

[2] Z. Zang, A. Amine, and R. Mangharam, "PoseINN: Realtime Visual-based Pose Regression and Localization with Invertible Neural Networks," *arXiv preprint arXiv:2404.13288v3*, 2024.

[3] R. G. D. C. Souza, C. R. de Oliveira, and L. R. A. C. B. da Silva, "Learning likelihoods with conditional normalizing flows," *arXiv preprint arXiv:1912.00042*, 2019.

[4] M. O'Kelly, H. Zheng, D. Karthik, and R. Mangharam, "F1Tenth: An open-source evaluation environment for continuous control and reinforcement learning," in *Proceedings of the NeurIPS 2019 Competition and Demonstration Track, ser. Proceedings of Machine Learning Research, H. J. Escalante and R. Hadsell, Eds., vol. 123. PMLR*, 08-14 Dec 2020, pp. 77-89.

[5] C. Walsh and S. Karaman, "Cddt: Fast approximate 2d ray casting for accelerated localization," vol. abs/1705.01167, 2017. [Online]. Available: http://arxiv.org/abs/1705.01167

[6] J. Betz, H. Zheng, A. Liniger, U. Rosolia, P. Karle, M. Behl, V. Krovi, and R. Mangharam, "Autonomous Vehicles on the Edge: A Survey on Autonomous Vehicle Racing." *IEEE Open Journal of Intelligent Transportation Systems*, vol. 3, pp. 458-488, 2022.

[7] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "Carla: An open urban driving simulator," in *Conference on robot learning. PMLR*, 20...

[8] X. Wang and R. Kumar, "End-to-end learning of visual odometry from optical flow," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 10696-10705.

[9] B. Glocker. S. Izadi, J. Shotton, and A. Criminisi, "Real-time rgb-d camera relocalization," in *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). IEEE*, 2013, pp. 173-179.

[10] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning. PMLR*, 2019, pp. 6105-6114.

[11] Y. Zhou, C. Barnes, J. Lu, J. Yang, and H. Li, "On the continuity of rotation representations in neural networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 5745-5753.

[12] F. Dellaert, D. Fox, W. Burgard, and S. Thrun, "Monte carlo localization for mobile robots," in *Proceedings 1999 IEEE international conference on robotics...*