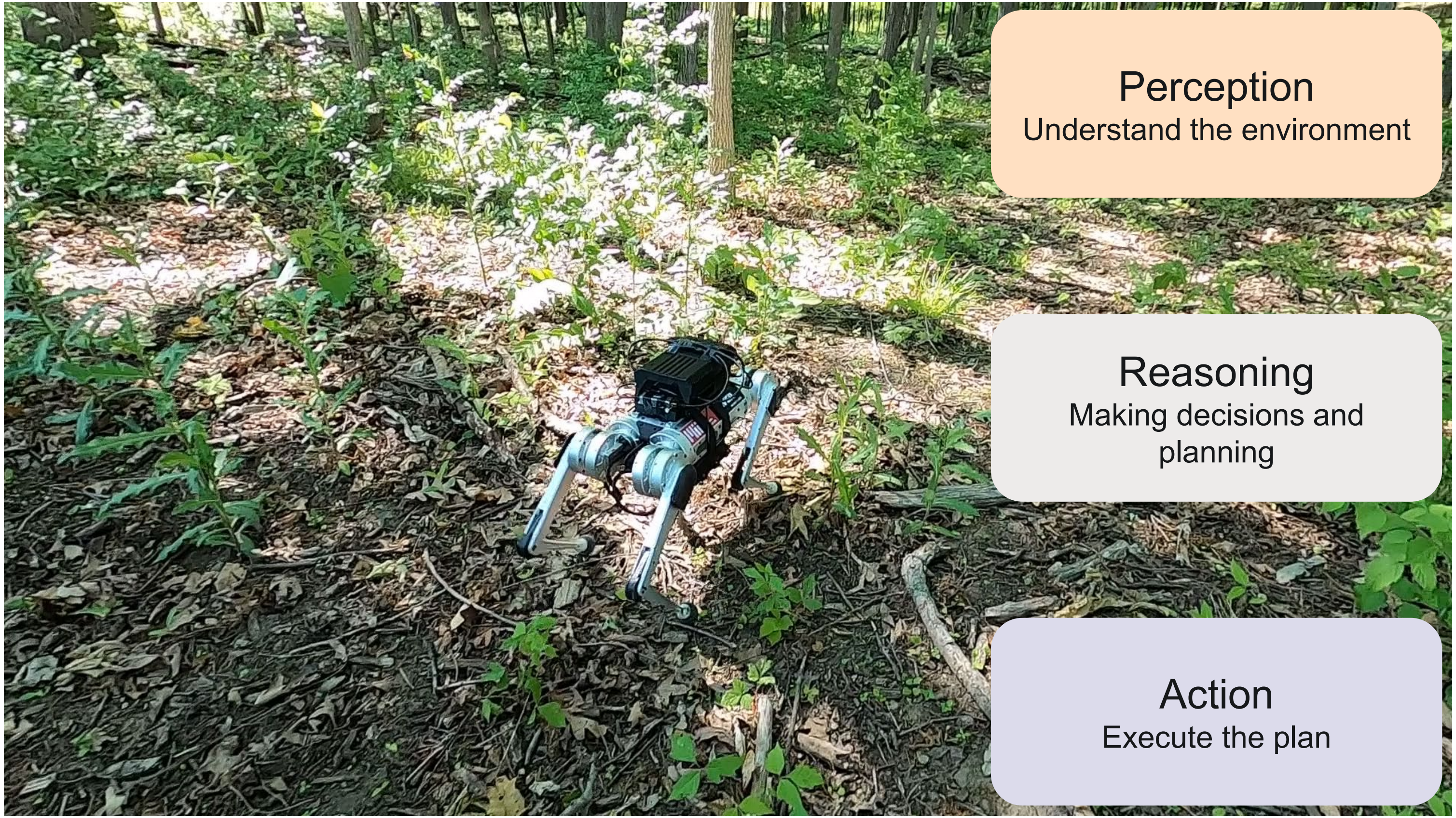
It’s the Same Everywhere: Leveraging Symmetry for Robot Perception and Localization

Chien Erh (Cynthia) Lin & Tzu-Yuan (Justin) Lin

University of Notre Dame
June 4th, 2024

ROBOTICS CURLY Explore the Unknown

Robots achieve human-level autonomy in the future



Perception
Understand the environment

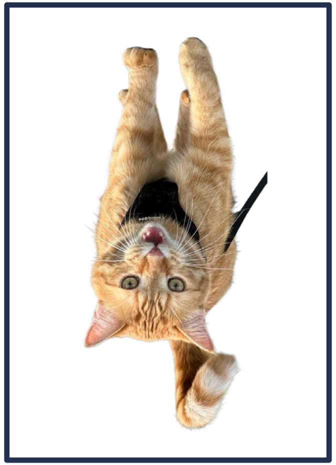
Reasoning
Making decisions and planning

Action
Execute the plan

Generalizable Robotic Systems?

- Equivalent object input
- Does robotic systems understand they are the same?





Symmetry: immunity to a possible change

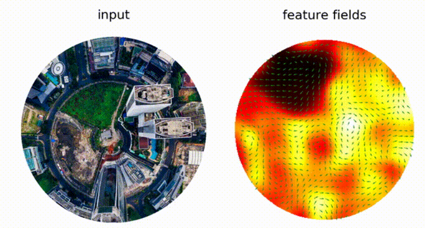
Symmetry can help designing efficient and robust algorithms

Equivariance: functions that preserve the transformation applied on the input to the output.

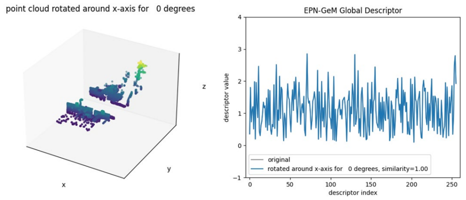
$$f(g \circ x) = g \circ f(x)$$

Invariance: output of functions is independent to the transformations applied to the input.

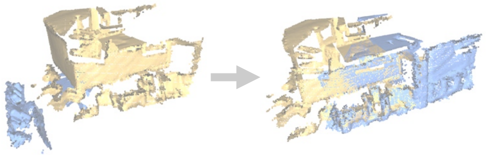
$$f(g \circ x) = f(x)$$



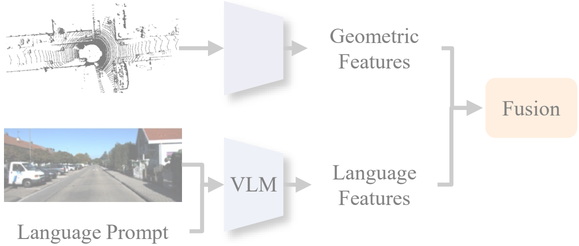
Outline



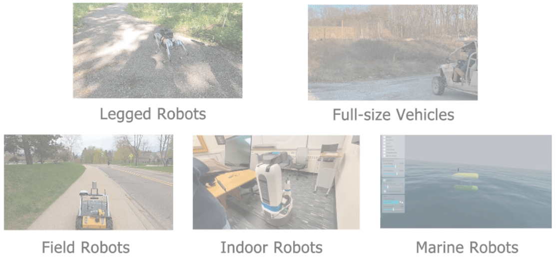
Place Recognition



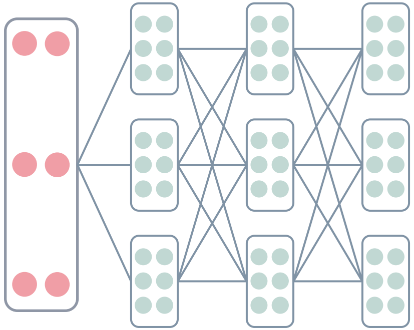
Point Cloud Registration



Foundation Models



Proprioceptive State Estimation



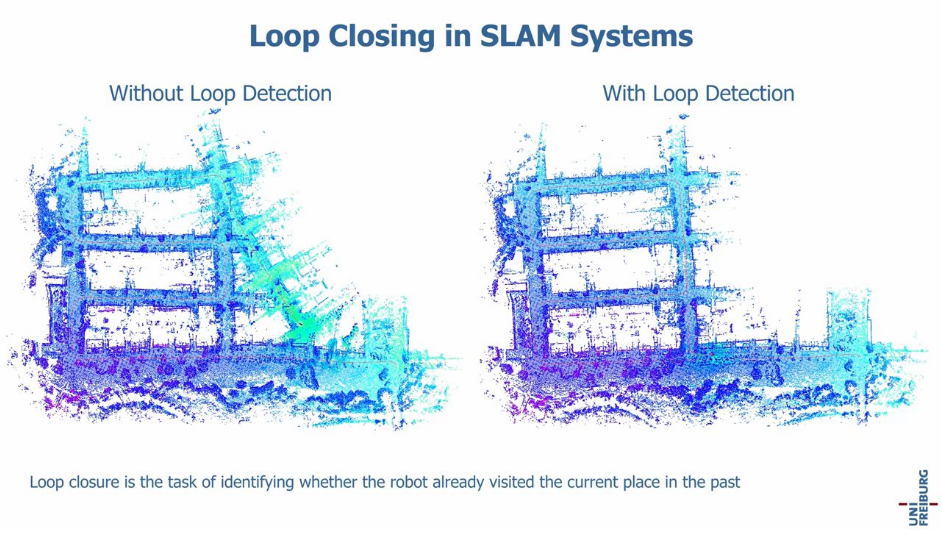
Lie Algebraic Neuron Networks

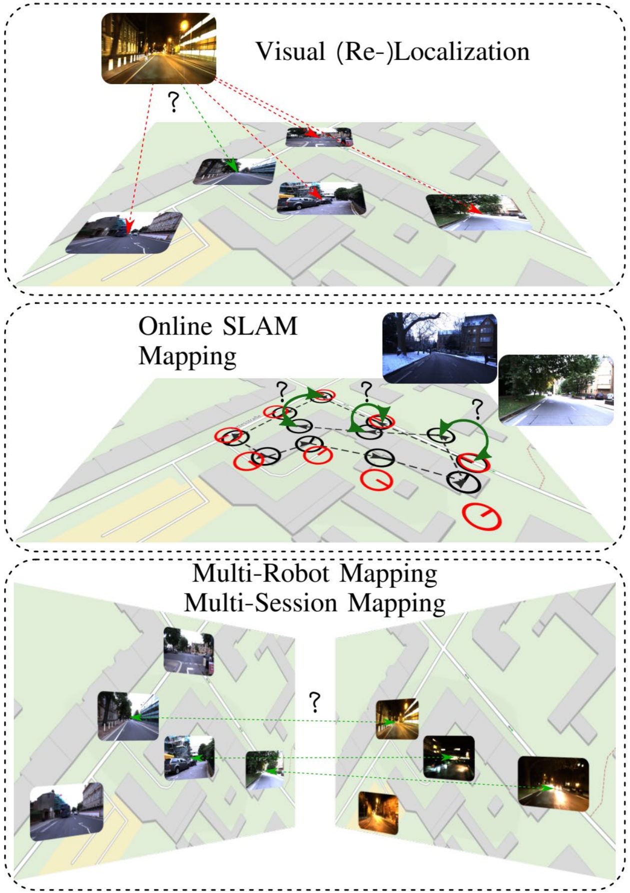
Perspective Equivariant Representation Learning

Place Recognition

Has the robot been to this place before?

Another name: Loop Closure Detection

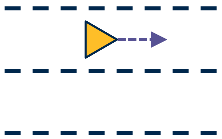




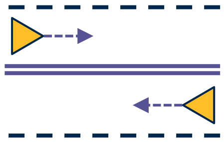
Schubert, Stefan, et al. "Visual place recognition: A tutorial." IEEE Robotics & Automation Magazine (2023).
Cattaneo, Daniele, Matteo Vaghi, and Abhinav Valada. "LCDnet: Deep loop closure detection and point cloud registration for lidar slam." T-RO, 2022. https://www.youtube.com/watch?v=nAvTdEFRh_s&ab_channel=RobotLearningFreiburg

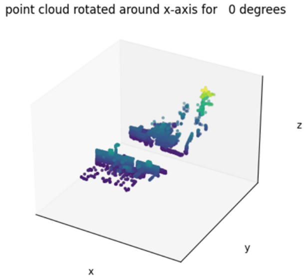
Challenges in Place Recognition

Learned features are sensitive to transformation changes in 3D data



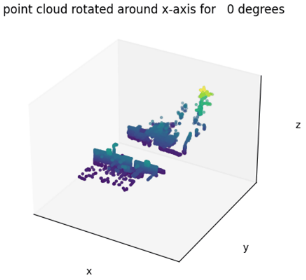
- Vehicle changes lanes
- Different orientation in a similar location
- Random rotation and drift from drones



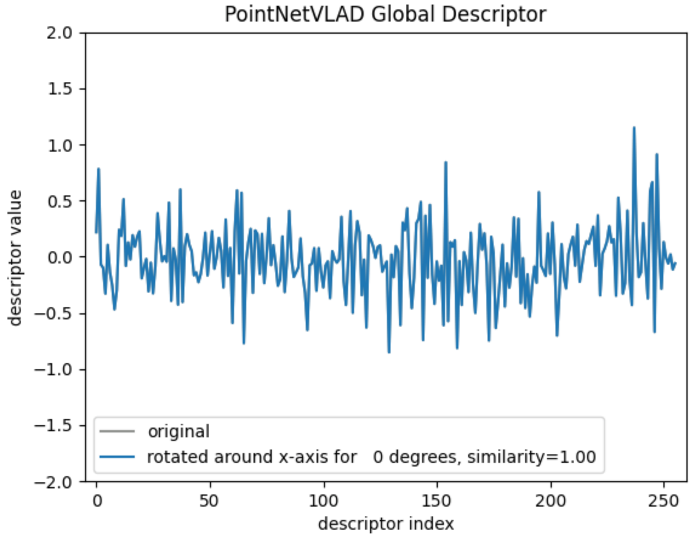


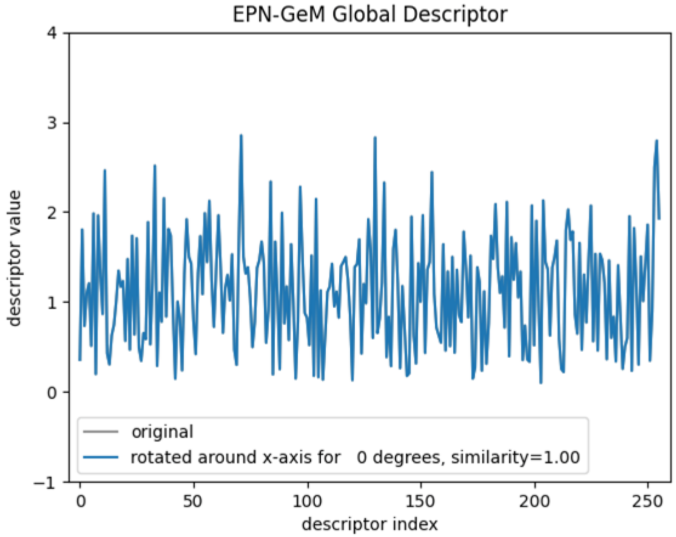
Adding Symmetry can help stabilizing the feature

Existing Method: sensitive



Ours: stable





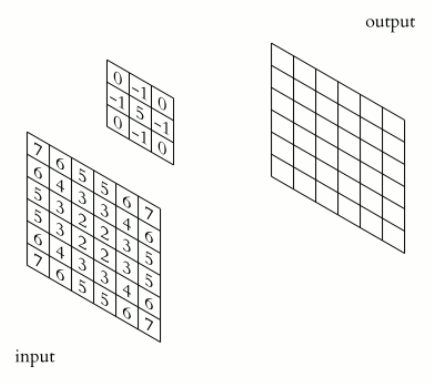
ROBOTICS CURLY Explore the Unknown

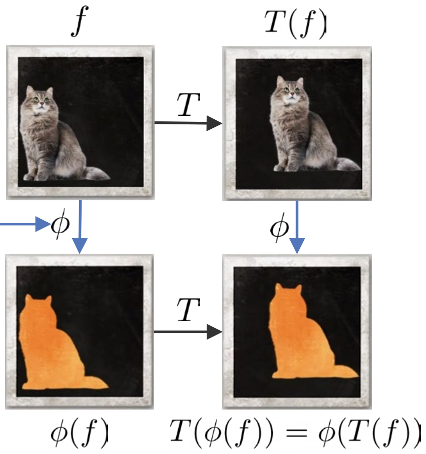
Uy, Mikaela Angelina, and Gim Hee Lee. "Pointnetvlad: Deep point cloud based retrieval for large-scale place recognition." *CVPR* 2018.
Lin, Chien Erh, et al. "SE(3)-Equivariant Point Cloud-Based Place Recognition." CoRL 2022.

We achieve it by utilizing group convolution

- Standard 2D convolutions are translation-equivariant
  - Inner product of function f and a shifted kernel k

$$(f * k)(\mathbf{x}) = \int_{\mathbb{R}^2} f(\tilde{\mathbf{x}})k(\tilde{\mathbf{x}} - \mathbf{x})\mathrm{d}\tilde{\mathbf{x}}$$

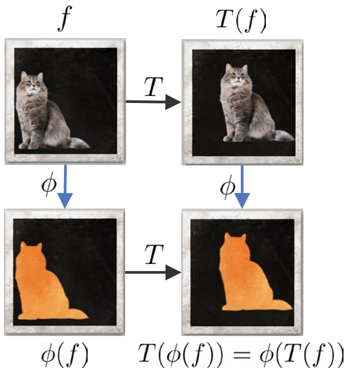




We achieve it by utilizing group convolution

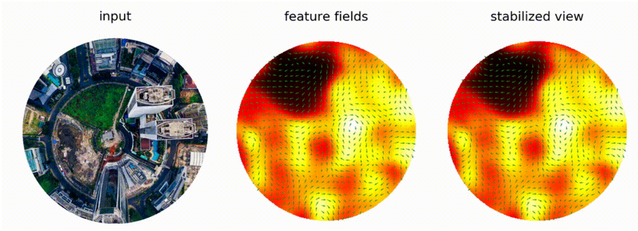
- Standard 2D convolutions are translation-equivariant
    - Inner product of function f and a shifted kernel k

$$(f * k)(\mathbf{x}) = \int_{\mathbb{R}^2} f(\tilde{\mathbf{x}}) k(\tilde{\mathbf{x}} - \mathbf{x}) d\tilde{\mathbf{x}}$$



- Group convolutions extend equivariance beyond translations

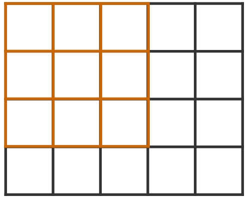
$$(f * k)(g) = \int_{G} f(\tilde{g}) k(g^{-1} \cdot \tilde{g}) d\tilde{g}$$

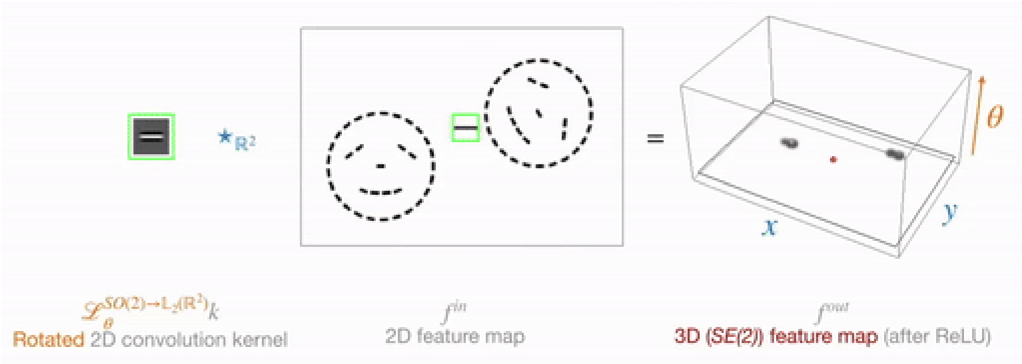


Erik Bekkers. "Group Equivariant Deep Learning - Lecture 1.3: Regular group convolutional neural networks." https://www.youtube.com/watch?v=cWG_1IzI0uI
Weiler, Maurice and Cesa, Gabriele. "General E(2)-Equivariant Steerable CNNs. " NeurIPS 2019. https://github.com/QUVA-Lab/e2cnn

Regular Group Convolution

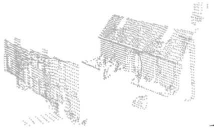
- Standard 2D convolutions convolute over pixels
- Group convolution expands additional dimensions
  - We use a SE(3)-equivariant network

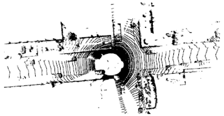




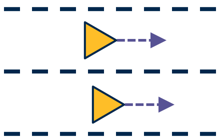
Erik Bekkers. “Group Equivariant Deep Learning - Lecture 1.3: Regular group convolutional neural networks.“ https://www.youtube.com/watch?v=cWG_1IzI0uI
Zhu, Minghan, et al. "E2PN: Efficient SE(3)-Equivariant Point Network." CVPR 2023.

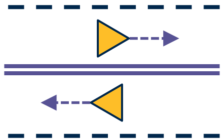
Results on Unseen and Challenging Data - KITTI

Trained on pre-processed submap 

Test on unseen sensor measurement 

Methods	Average Recall @ 1 % (%) ↑	
	Same Direction	Opposite Direction
PointNetVLAD[1]	73.18	32.47
MinkLoc3D[2]	28.07	17.30
Ours[3]	**86.22**	**71.70**



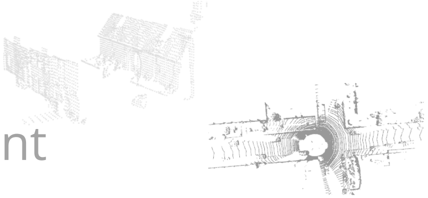


[1] Uy, Mikaela Angelina, and Gim Hee Lee. "Pointnetvlad: Deep point cloud based retrieval for large-scale place recognition." *CVPR* 2018.
[2] Komorowski, Jacek. "Minkloc3d: Point cloud based large-scale place recognition." *WACV* 2021.
[3] Lin, Chien Erh, et al. “SE(3)-Equivariant Point Cloud-Based Place Recognition.” CoRL 2022.

Place Recognition – Key Takeaway

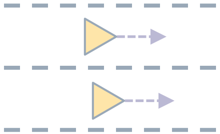
Trained on pre-processed submap

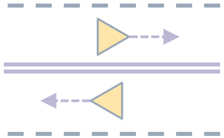
Test on unseen sensor measurement



Generalizable to unseen data

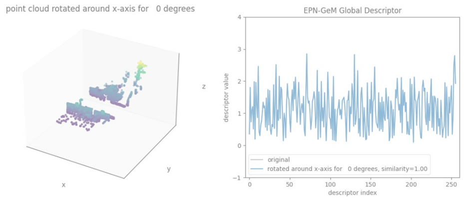
Symmetry helps in challenging scenarios



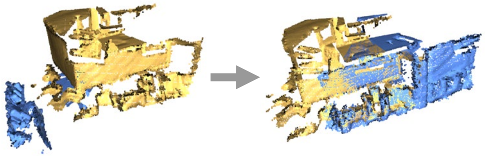


[1] Uy, Mikaela Angelina, and Gim Hee Lee. "Pointnetvlad: Deep point cloud based retrieval for large-scale place recognition." *CVPR* 2018.
[2] Komorowski, Jacek. "Minkloc3d: Point cloud based large-scale place recognition." *WACV* 2021.
[3] Lin, Chien Erh, et al. "SE(3)-Equivariant Point Cloud-Based Place Recognition." CoRL 2022.

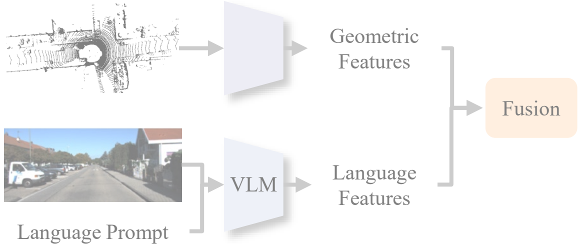
Outline



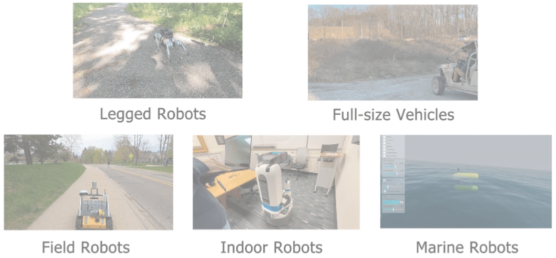
Place Recognition



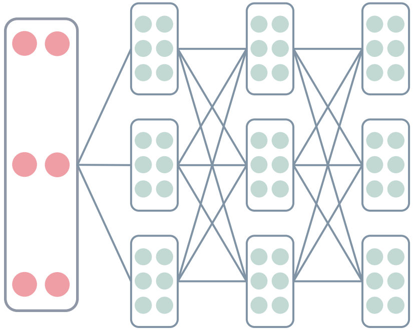
Point Cloud Registration



Foundation Models



Proprioceptive State Estimation



Lie Algebraic Neuron Networks

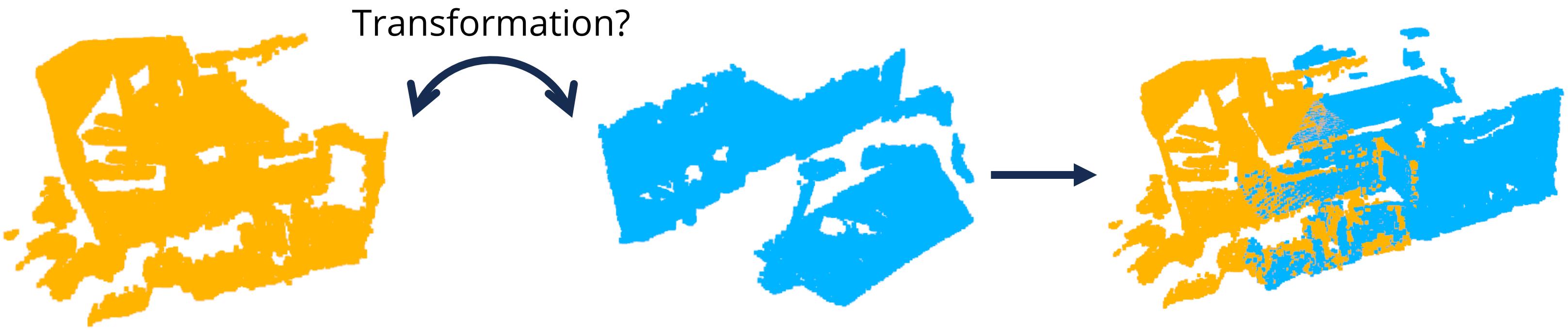
Perspective Equivariant Representation Learning

ROBOTICS

CURLY
Explore the Unknown

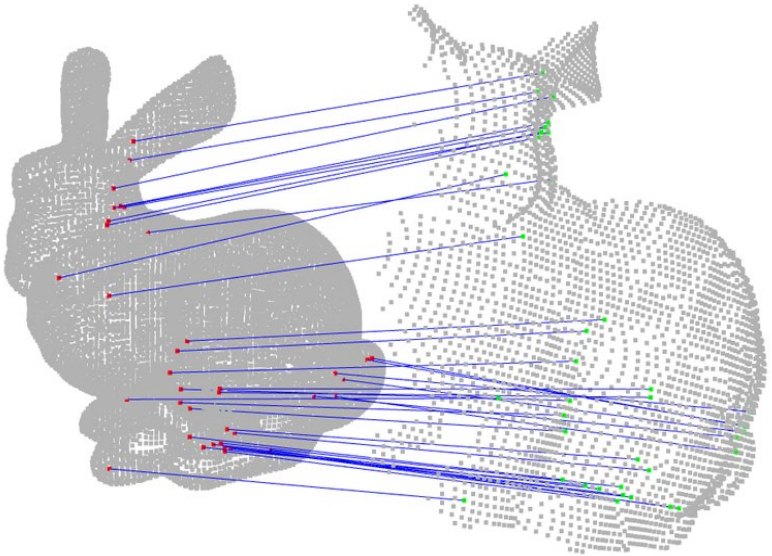
Point Cloud Registration

- Find the transformation between two point clouds
- Challenges
  - Low overlap
  - Large arbitrary transformation



Finding Correspondence

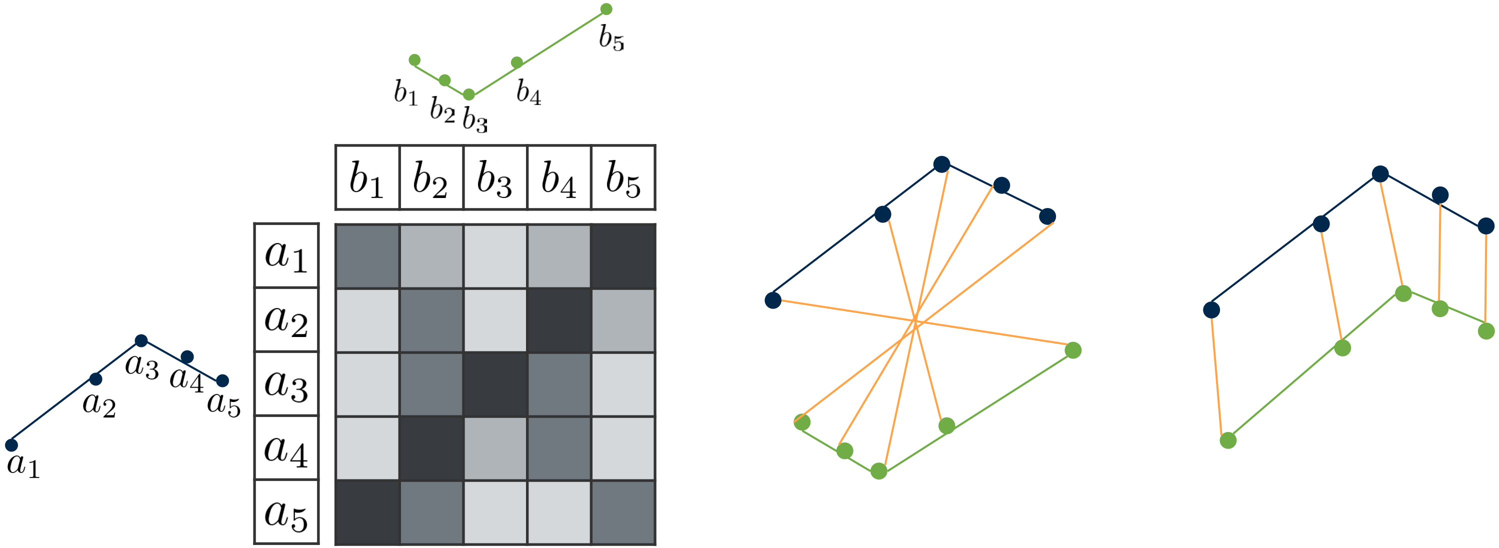
- Finding correct correspondence is essential.
- The accuracy is highly relied on the correspondence.



Laserna, Javier, Pablo San Segundo, and David Alvarez. "CliReg: Clique-based robust Point Cloud Registration." *imaging* 2024.

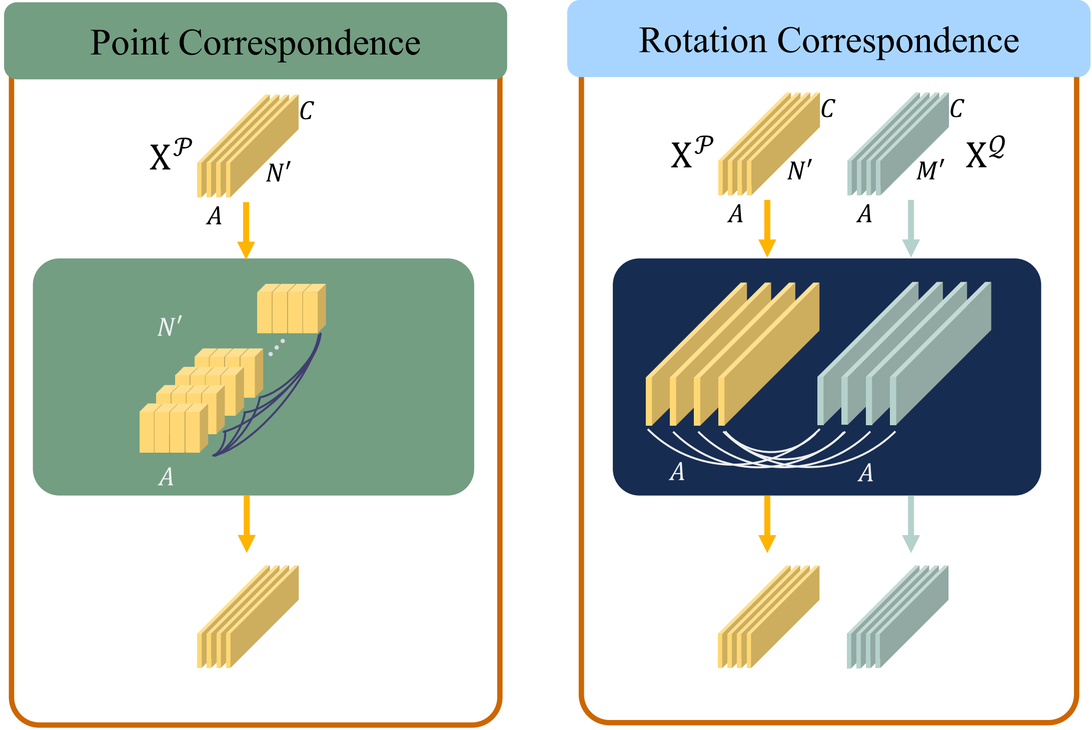
Transformer helps finding correspondence

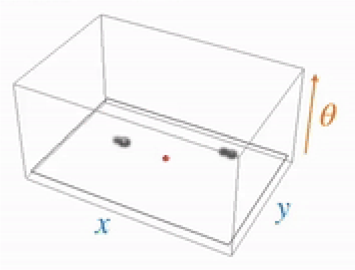
- Simplified explanation of Transformer:
    - Learn where we should pay more attention at when comparing one input with another



Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems 30 (2017).

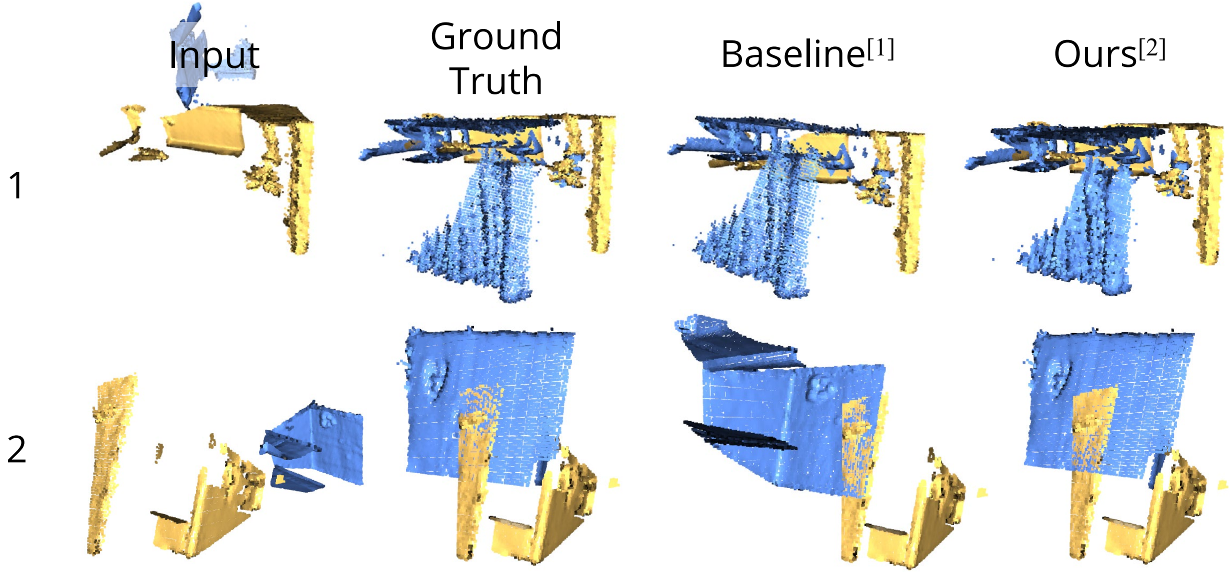
Adding Symmetry in Transformers





Erik Bekkers. "Group Equivariant Deep Learning - Lecture 1.3: Regular group convolutional neural networks." https://www.youtube.com/watch?v=cWG_1IzI0uI
Lin, Chien Erh, et al. "SE3ET: SE(3)-Equivariant Transformer for Low-Overlap Point Cloud Registration." Submitted to RA-L 2024.

Results on Rotated 3DLoMatch



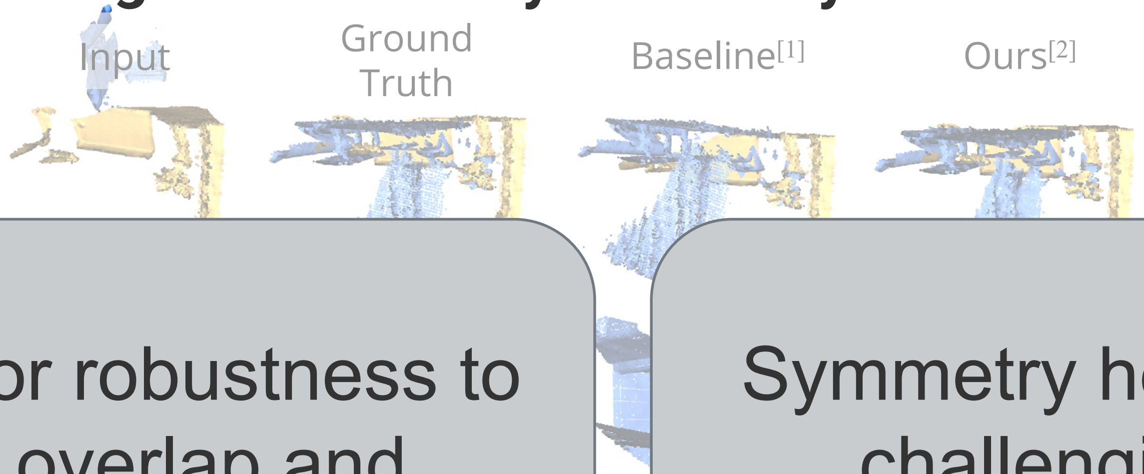
	Baseline[1]		Ours[2]	
Example #	RRE (deg)	RTE (m)	RRE (deg)	RTE (m)
1	7.953	0.137	**0.480**	**0.054**
2	176.097	4.585	**7.488**	**0.167**

[1] Qin, Zheng, et al. "Geometric transformer for fast and robust point cloud registration." *CVPR* 2022.
[2] Lin, Chien Erh, et al. "SE3ET: SE(3)-Equivariant Transformer for Low-Overlap Point Cloud Registration." Submitted to RA-L 2024.

Point Cloud Registration – Key Takeaway

Input | Ground Truth | Baseline[1] | Ours[2]

1



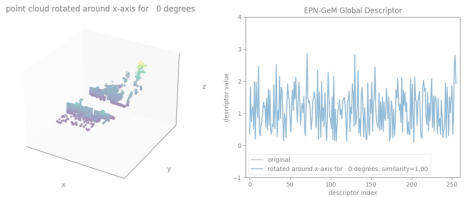
Superior robustness to low overlap and arbitrary rotations

Symmetry helps in challenging scenarios

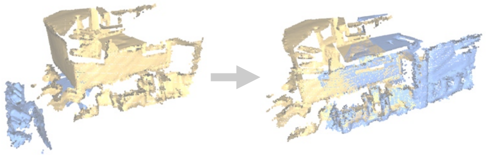
1	7.953	0.137	**0.480**	**0.054**
2	176.097	4.585	**7.488**	**0.167**

[1] Qin, Zheng, et al. "Geometric transformer for fast and robust point cloud registration." *CVPR* 2022.
[2] Lin, Chien Erh, et al. "SE3ET: SE(3)-Equivariant Transformer for Low-Overlap Point Cloud Registration." Submitted to RA-L 2024.

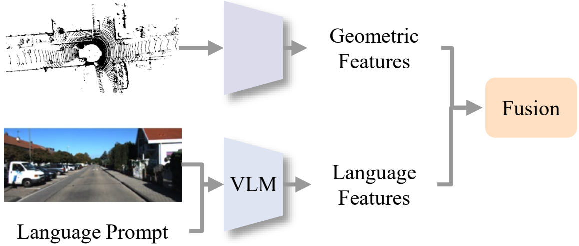
Outline



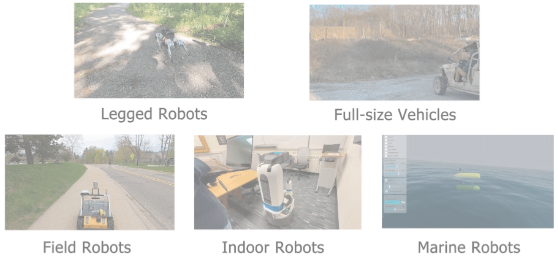
Place Recognition



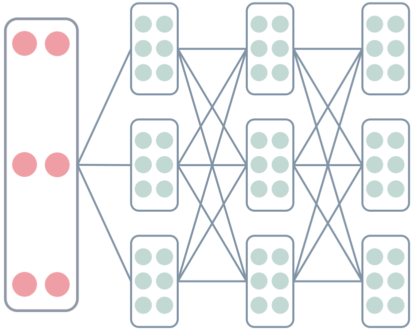
Point Cloud Registration



Foundation Models



Proprioceptive State Estimation

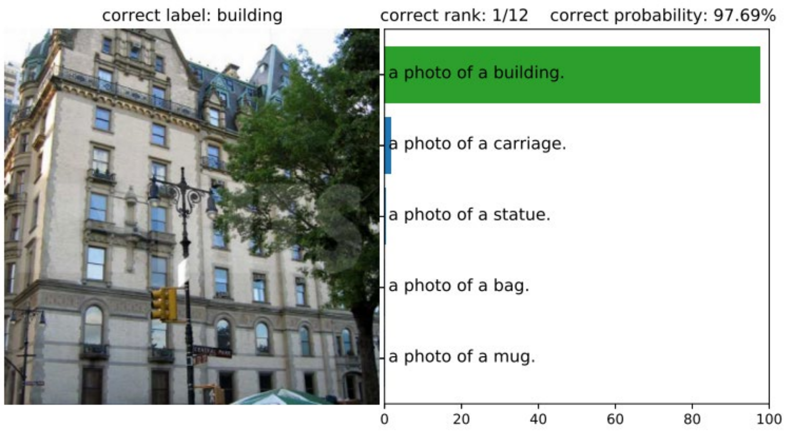


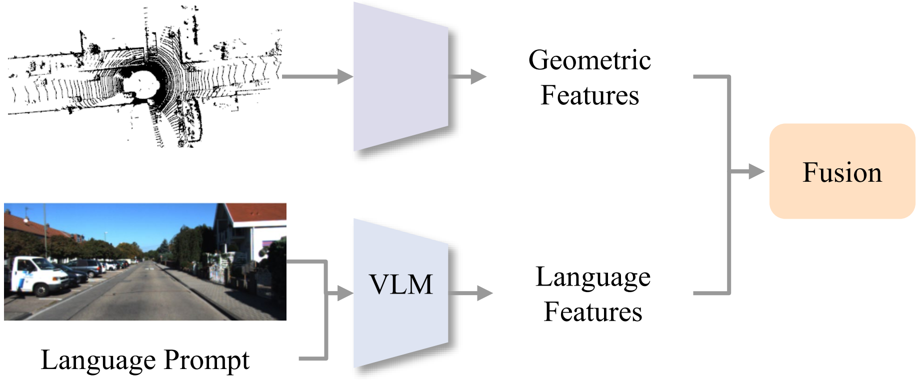
Lie Algebraic Neuron Networks

Perspective Equivariant Representation Learning

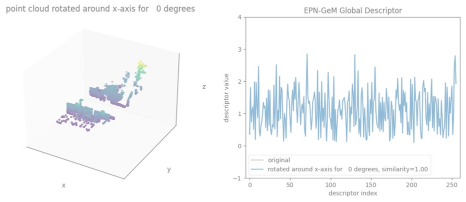
Fusing with Foundation Models

- Foundation models enable zero-shot transfer (without training on the specific data)
- How to obtain approximately invariant features from VLMs?



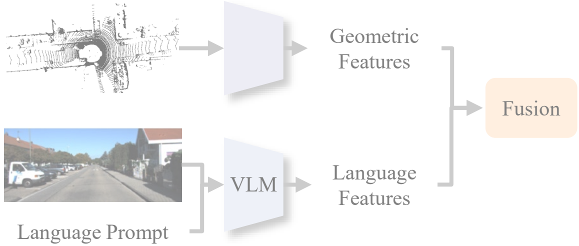


Outline

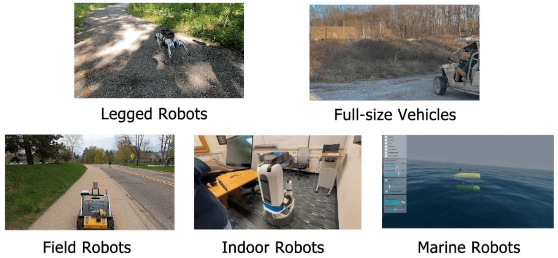


Place Recognition

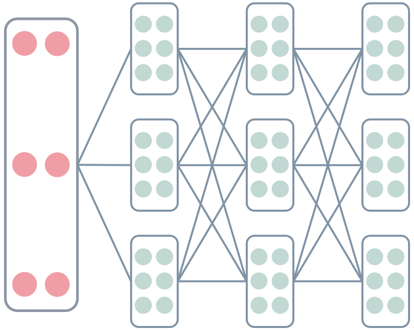
Point Cloud Registration



Foundation Models



Proprioceptive State Estimation

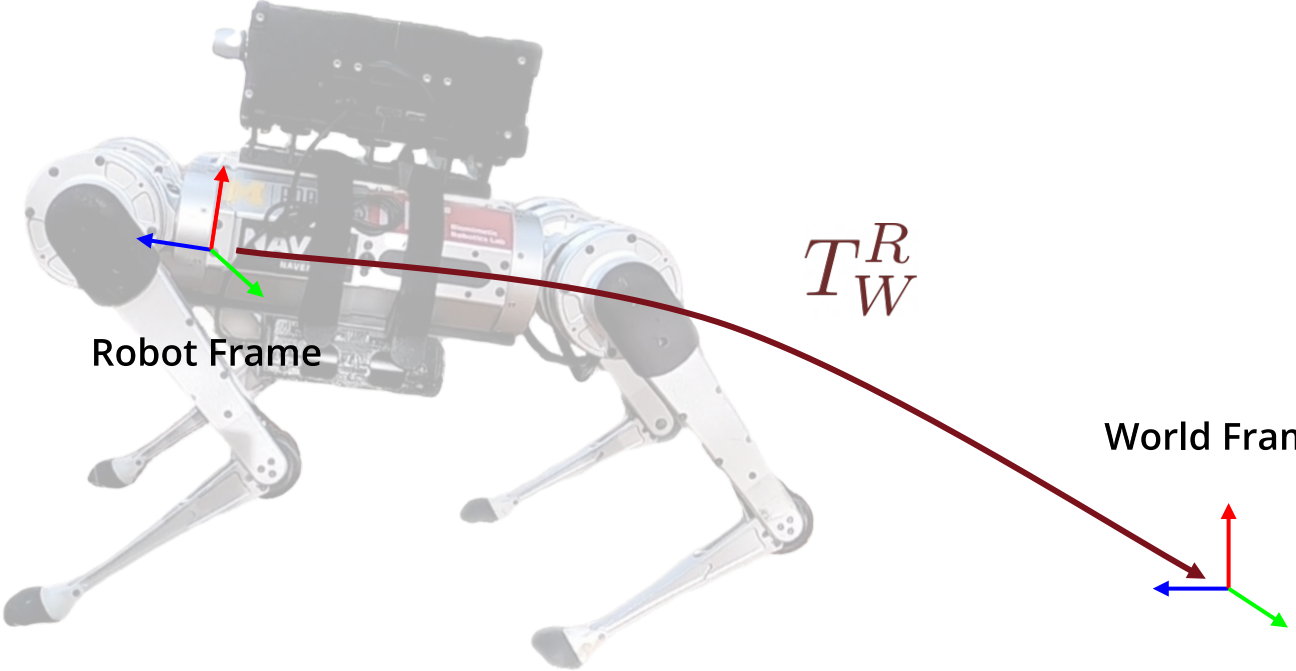


Lie Algebraic Neuron Networks

Perspective Equivariant Representation Learning

Proprioceptive State Estimation

$$X = \begin{bmatrix} R & v & p \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$



Kalman Filtering

- Kalman Filter [1]

$$\frac{d}{dt}x_t = A_t x_t + B_t u_t + w_t$$

- Extended Kalman Filter (EKF)

$$\frac{d}{dt}x_t = f(x_t, u_t, w_t) \qquad A_t = \frac{\partial f}{\partial x_t}\Big|_{x=x_t}$$

An incorrect estimation of the states can lead to a wrong linearization!!

- Error-State EKF [2]

$$e_t \triangleq x_t \boxminus \hat{x}_t \qquad \begin{aligned}\frac{d}{dt}e_t &= g(e_t, x_t, u_t, w_t) \\ &\approx A_t(x_t, u_t)e_t + w_t\end{aligned}$$

ROBOTICS CURLY Explore the Unknown

1. Kalman, Rudolph Emil. "A new approach to linear filtering and prediction problems." (1960): 35-45.
2. Trawny, Nikolas, and Stergios I. Roumeliotis. "Indirect Kalman filter for 3D attitude estimation." University of Minnesota, Dept. of Comp. Sci. & Eng., Tech. Rep 2 (2005): 2005.

Symmetry?

Can we define an error such that it respect the symmetry of the system?

$$e_t \triangleq x_t \boxminus \hat{x}_t$$

$\boxed{\text{Yes!}}$

Define our states on a matrix Lie group: $X \in \mathcal{G}$ Ex: $SO(3), SE(3)$

Right-invariant Error: $\eta_t^r = \bar{X}_t X_t^{-1} = (\bar{X}_t L)(X_t L)^{-1}$

Left-invariant Error: $\eta_t^l = X_t^{-1} \bar{X}_t = (L\bar{X}_t)^{-1}(LX_t)$

Invariant Kalman Filtering [3]

If the system dynamics satisfy the group affine property:

$$f_{u_t}(X_1 X_2) = f_{u_t}(X_1) X_2 + X_1 f_{u_t}(X_2) - X_1 f_{u_t}(I) X_2$$

The error dynamic can be exactly model as a linear system in the Lie algebra:

$$\frac{d}{dt}\eta_t = g_{u_t}(\eta_t) \quad \Rightarrow \quad \frac{d}{dt}\xi_t = A_t \xi_t$$

3. Barrau, Axel, and Silvère Bonnabel. "The invariant extended Kalman filter as a stable observer." IEEE Transactions on Automatic Control 62.4 (2016): 1797-1812.

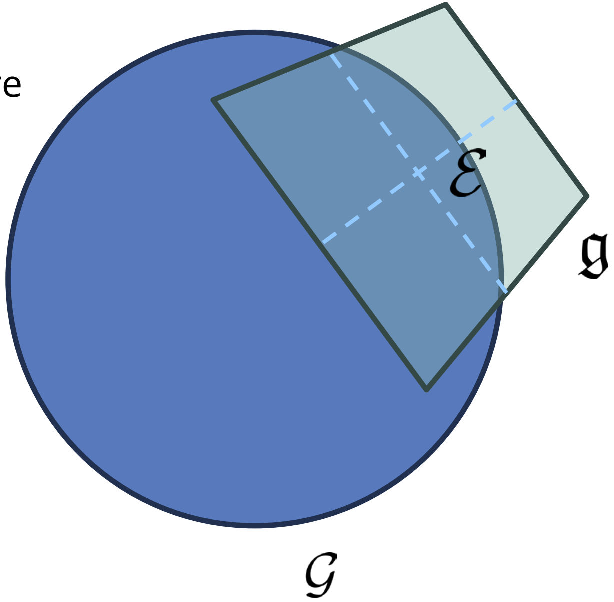
Lie Groups

- A **Lie group** $\mathcal{G}$ is a group that is also a differentiable manifold

- The **Lie algebra** $\mathfrak{g}$ is the tangent space at the identity $\mathcal{E}$
  - It is a vector space that locally captures the group structure

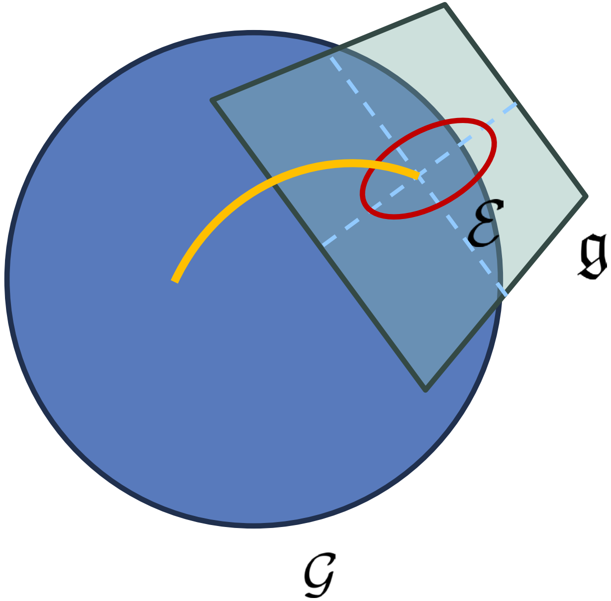
- One can move between $\mathcal{G}$ and $\mathfrak{g}$ using the **exponential** and **log** maps

$$\begin{aligned} \exp: &\quad \mathfrak{g} \mapsto \mathcal{G}, \quad X \mapsto \exp(X) \\ \log: &\quad \mathcal{G} \to \mathfrak{g}, \quad g \mapsto \log(g) \end{aligned}$$



Invariant Kalman Filtering [3]

- Means evolves on the group.
- Tracks the covariance in the Lie algebra.



3. Barrau, Axel, and Silvère Bonnabel. "The invariant extended Kalman filter as a stable observer." IEEE Transactions on Automatic Control 62.4 (2016): 1797-1812.

Invariant Kalman Filtering [3]

Propagation:

$$\frac{\mathrm{d}}{\mathrm{d}t}\bar{\mathbf{X}}_t = f_{u_t}(\bar{\mathbf{X}}_t)$$
$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{P}_t = \mathbf{A}_t\mathbf{P}_t + \mathbf{P}_t\mathbf{A}_t^\mathsf{T} + \bar{\mathbf{Q}}_t,$$

Correction:

$$\bar{\mathbf{X}}_t^+ = \mathrm{Exp}\left(\overbrace{\mathbf{K}_t\mathbf{\Pi}\left(\bar{\mathbf{X}}_t\mathbf{Y}_t\right)}^{\text{correction vector}}\right)\bar{\mathbf{X}}_t$$
$$\mathbf{P}_t^+ = (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\mathbf{P}_t$$

$$\left.\begin{aligned}\mathbf{S}_t &= \mathbf{H}_t\mathbf{P}_t\mathbf{H}_t^\mathsf{T} + \bar{\mathbf{N}}_t \\ \mathbf{K}_t &= \mathbf{P}_t\mathbf{H}_t^\mathsf{T}\mathbf{S}_t^{-1}\end{aligned}\right\}\ \text{Computing Kalman Gain}$$

Linearization are constant!

3. Barrau, Axel, and Silvère Bonnabel. "The invariant extended Kalman filter as a stable observer.“ IEEE Transactions on Automatic Control 62.4 (2016): 1797-1812.

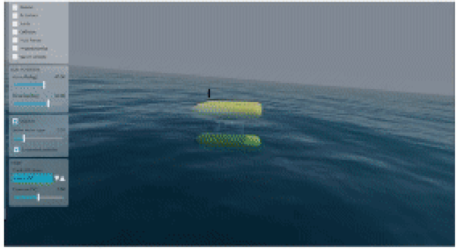
DRIFT: Dead Reckoning In Field Time [4]

Legged Robots

Full-size Vehicles

Field Robots

Indoor Robots



Marine Robots

4. Lin, Tzu-Yuan, et al. "Proprioceptive Invariant Robot State Estimation." arXiv preprint arXiv:2311.04320 (2023).

DRIFT - Estimating orientation, velocity, and position

Propagation:

IMU

Correction:

Encoder → Kinematics → Foot Position

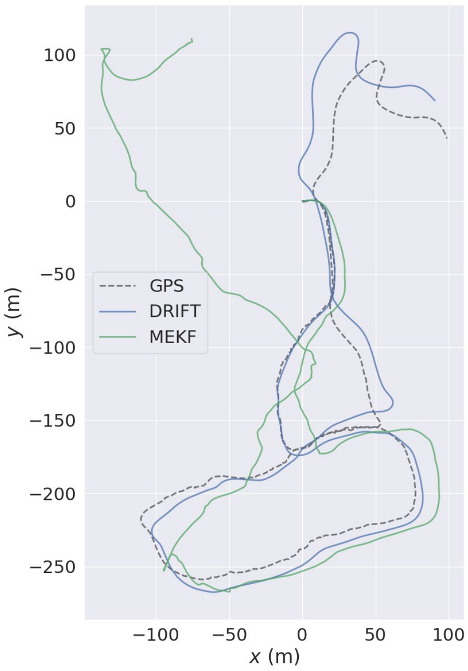
Encoder → Kinematics → Velocity

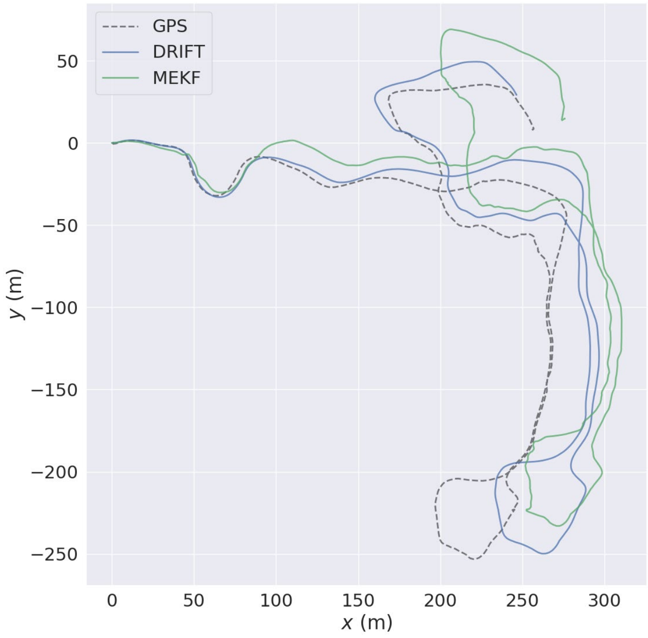
Body Velocity Sensor → Velocity

Full-size Vehicles

Full-Size Vehicle

NEYASYSTEMS
A DIVISION OF APPLIED RESEARCH ASSOCIATES, INC.





- 3 Sequences
- Avg. Distance: 1510.43 m
- Avg. Duration: 449.15 sec

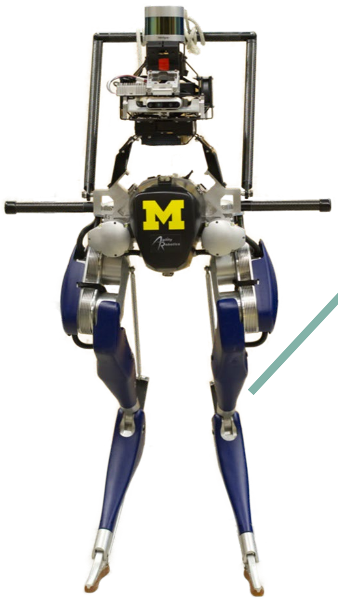
	MEKF [5]	DRIFT [4]
Final Drift (m)	203.02	**51.08**
Percentage (%)	12.32%	**3.18%**

ROBOTICS CURLY Explore the Unknown

4. Lin, Tzu-Yuan, et al. "Proprioceptive Invariant Robot State Estimation." arXiv preprint arXiv:2311.04320 (2023).
5. Sola, Joan. "Quaternion kinematics for the error-state Kalman filter." *arXiv preprint arXiv:1711.02508* (2017).

Field Robots

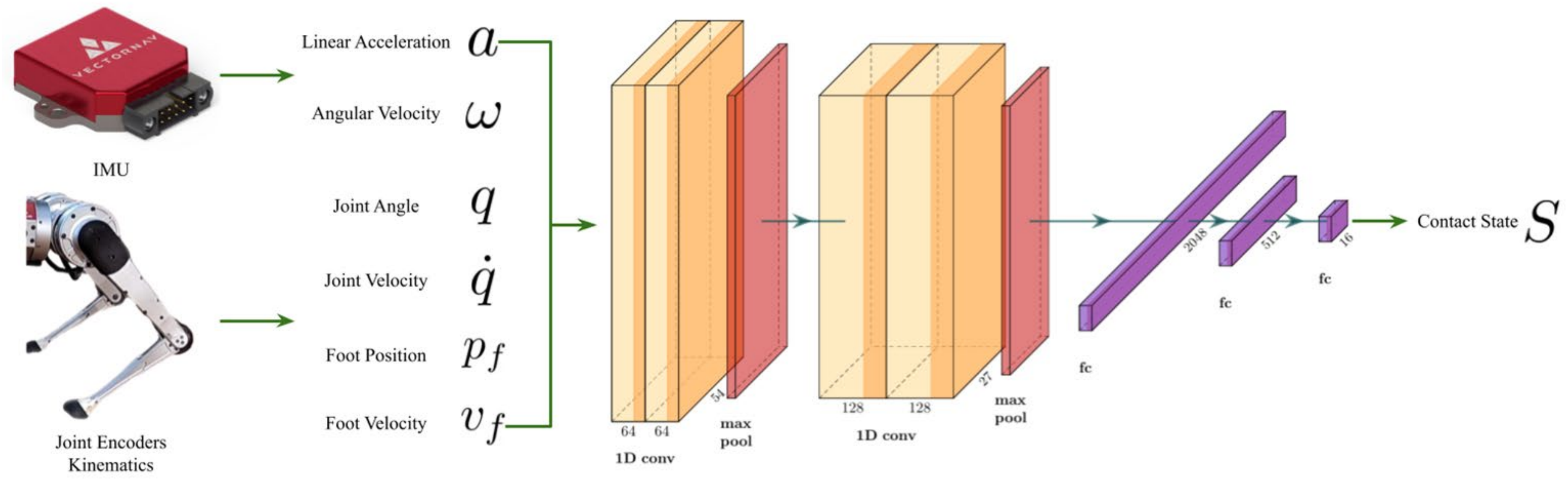
Legged Robot - Contact Detection



Spring for contact detection!

How do we accurately determine contact events?

Deep Contact Estimator [5]



Runs real-time on an NVIDIA Jetson AGX Xavier at 830 Hz!

5. Lin, Tzu-Yuan, et al. "Legged Robot State Estimation using Invariant Kalman Filtering and Learned Contact Events." 5th Annual Conference on Robot Learning. 2021.

Legged Robot

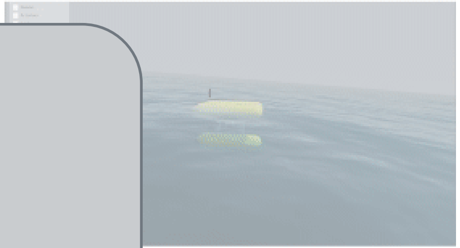
Proprioceptive State Estimation – Key Takeaway

Symmetry helps improve the consistency

Learned contacts help legged state estimation

Field Robots

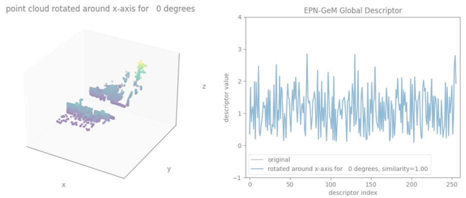
DRIFT:
open-sourced ready-to-use library



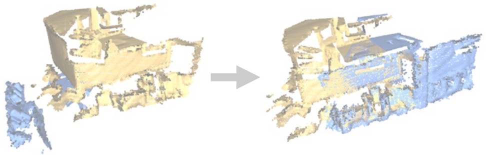
arine Robots

4. Lin, Tzu-Yuan, et al. "Proprioceptive Invariant Robot State Estimation." arXiv preprint arXiv:2311.04320 (2023).

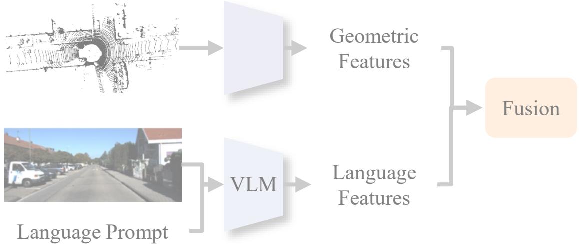
Outline



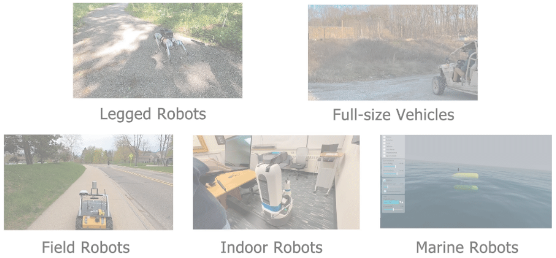
Place Recognition



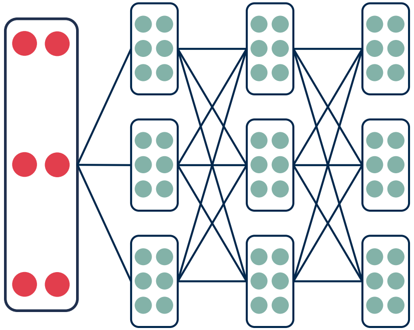
Point Cloud Registration



Foundation Models



Proprioceptive State Estimation



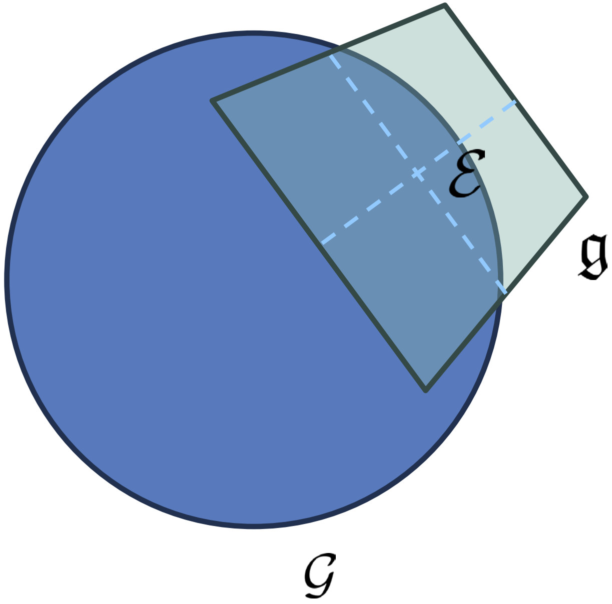
Lie Algebraic Neuron Networks

Perspective Equivariant Representation Learning

Neural Networks in a Lie Algebra?

- Takes elements in the Lie algebra as input.
- Adjoint (conjugation) equivariant by design.

$$f(gXg^{-1}) = gf(X)g^{-1}$$

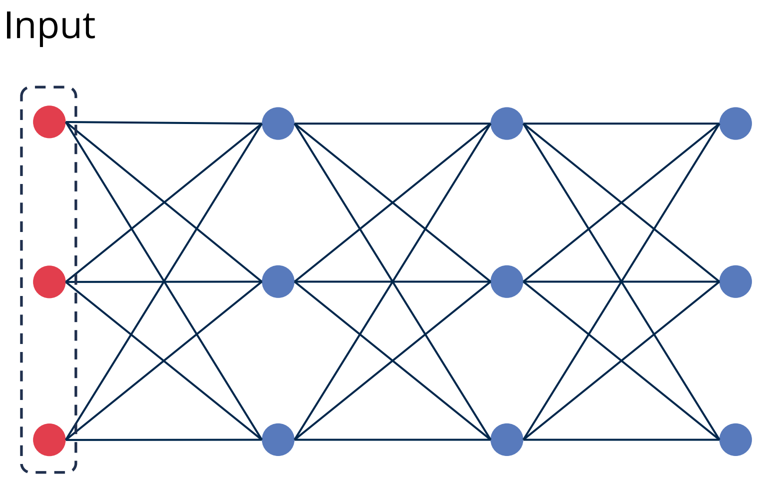


$f(\cdot)$: Lie Neuron Networks

$g \in G$: Elements in a Lie group

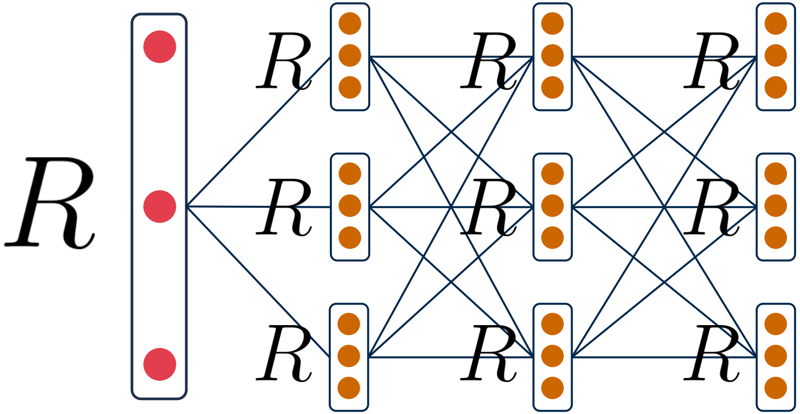
$X \in \mathfrak{g}$: Elements in a Lie algebra

Multi Layer Perceptron (MLP)



Each neuron is a scalar

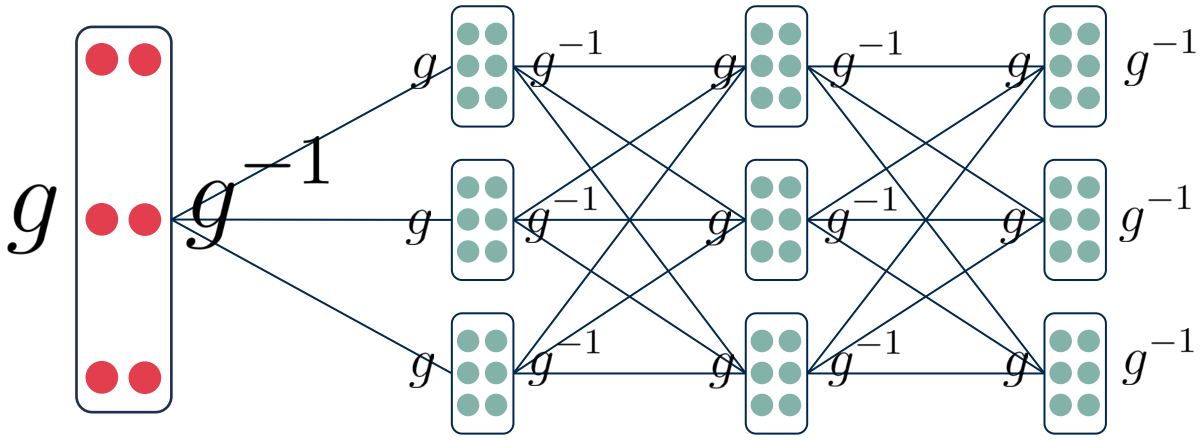
Vector Neurons [6]



Each neuron is a $\mathbb{R}^3$ vector

6. Deng, Congyue, et al. "Vector neurons: A general framework for so (3)-equivariant networks." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021.

Lie Neurons [7]



Each neuron is an element in the Lie algebra

7. Lin, Tzu-Yuan, Minghan Zhu, and Maani Ghaffari. "Lie Neurons: Adjoint-Equivariant Neural Networks for Semisimple Lie Algebras." arXiv preprint arXiv:2310.04521 (2023) (Accepted for ICML 2024).

Module Overview

Nonlinearity

Linear Mapping

Invariant Feature

Down sampling

LN-ReLU

$$\begin{cases} x, & \text{if } B(x, xU) \leq 0 \\ x + B(x, xU)xU, & \text{otherwise.} \end{cases}$$

LN-Linear

$$xW$$

LN-Pooling

$$\arg\max_n \; B(x_n[c]W, x_n[c])$$

LN-Invariant

$$B(x, x)$$

LN-Bracket

$$x + [(xU)^\wedge, (xV)^\wedge]^\vee$$

$B(X, Y)$: the Killing form

$$= Tr(ad(X) \circ ad(Y))$$

$$[X, Y] = XY - YX$$

ROBOTICS CURLY Explore the Unknown

Learning Dynamics

Free-rotating international space station

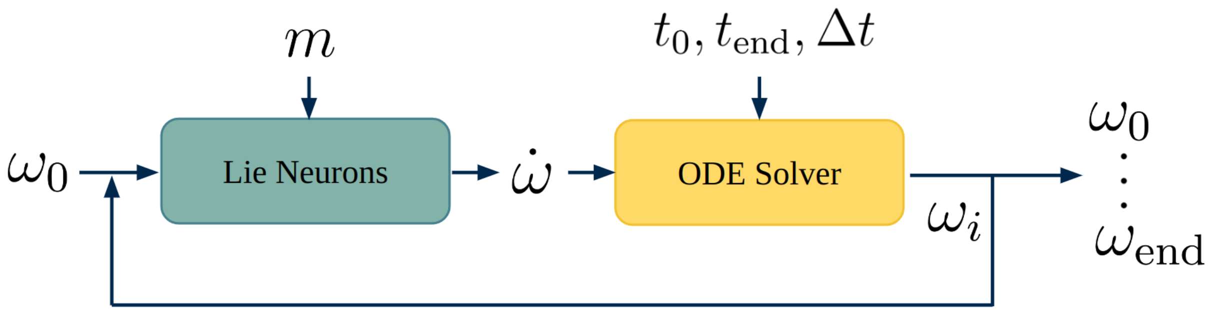
Euler equation of motion

$$I\dot{\omega}(t) + \omega(t) \times I\omega(t) = 0$$

ROBOTICS CURLY Explore the Unknown

7. National Aeronautics and Space Administration, International Space Station Program. On-Orbit Assembly, Modeling, and Mass Properties Data Book, Volume 1. June 2002.

Learning Dynamics

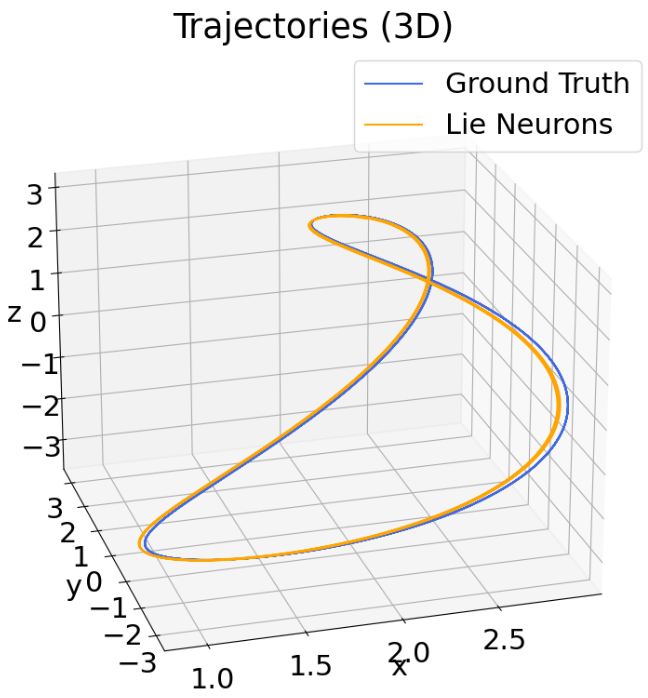


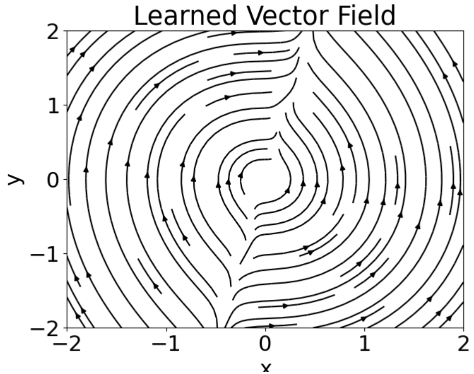
Using Neural ODE[8] framework

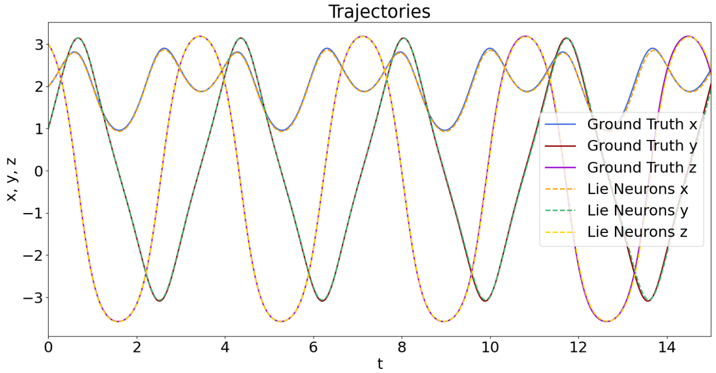
Lie Neurons learns the underlying vector field of the dynamics.

8. Chen, Ricky TQ, et al. "Neural ordinary differential equations." Advances in neural information processing systems 31 (2018).

Learning Dynamics







Learning Dynamics

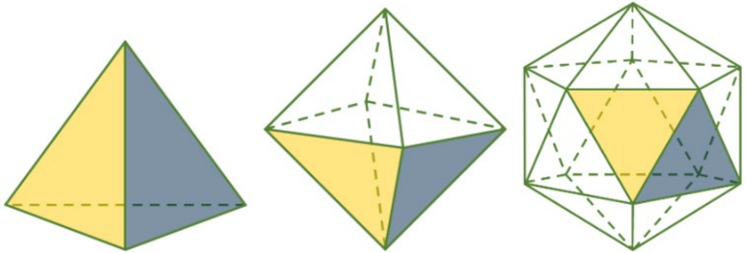
Train on multiple trajectories and evaluate on unseen data

Unit: rad/s	Error			Error (Change of Frame)		
Time (s)	5	15	25	5	15	25
MLP	0.428	0.717	0.800	0.474	0.733	0.805
Lie Neurons	**0.005**	**0.014**	**0.018**	**0.005**	**0.014**	**0.018**

Error: Norm distance error

Platonic Solid Classification

- Input: $\mathfrak{sl}(3)$ transformation between faces.
- Output: Platonic solid class.
- Randomly rotate the solids in *test set*.

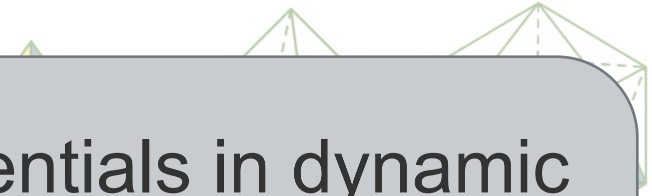


(a) Tetrahedron (b) Octahedron (c) Icosahedron

	Acc	Acc (Rotated)
MLP	95.76	36.54
Lie Neurons	**99.62**	**99.61**

Lie Algebraic Neural Network – Key Takeaway

- Input: $\mathfrak{sl}(3)$ transformation between faces.

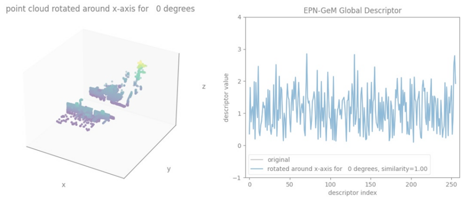


Lie algebraic network that is adjoint equivariant

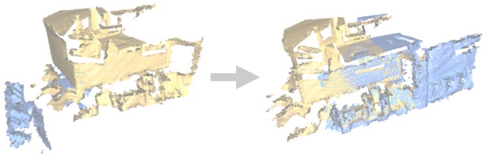
Potentials in dynamic modeling and perspective equivariant learning

	.76	
Lie Neurons	**99.62**	**99.61**

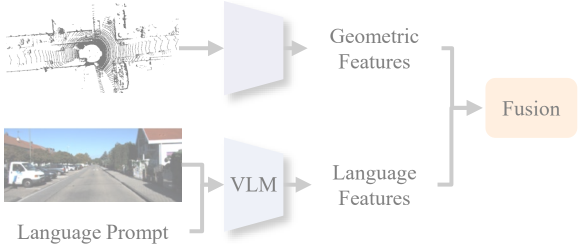
Outline



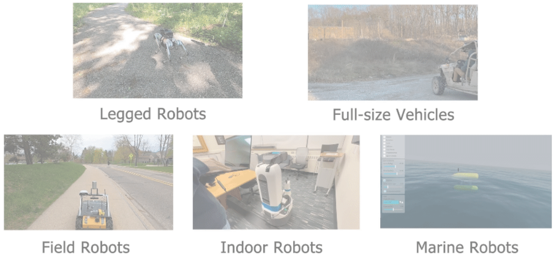
Place Recognition



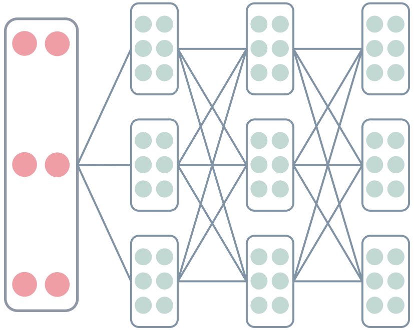
Point Cloud Registration



Foundation Models



Proprioceptive State Estimation



Lie Algebraic Neuron Networks

Perspective Equivariant Representation Learning

Perspective Changes

Homography Representation

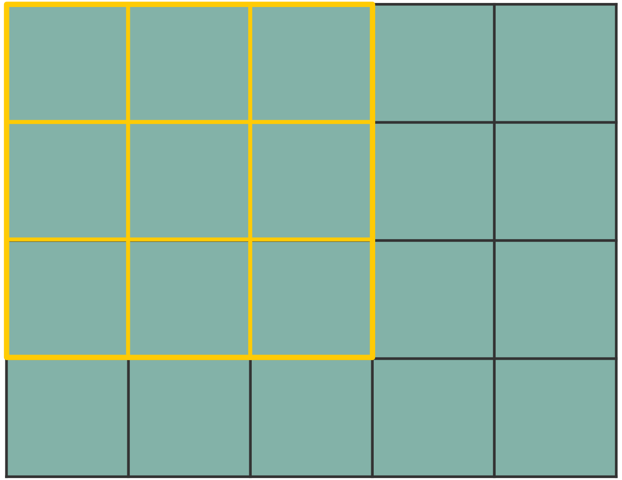
- Homography matrix
  - 8 degree of freedom
- Special linear group: $SL(3)$
  - 3 by 3 matrices with determinant 1

$$H \in SL(3)$$

Group Convolutional Neural Network

Need to discretize and define "grid" in the group.

$$H = \begin{bmatrix} c & d & e \\ f & g & h \\ i & j & k \end{bmatrix}, \quad \det(H) = 1$$



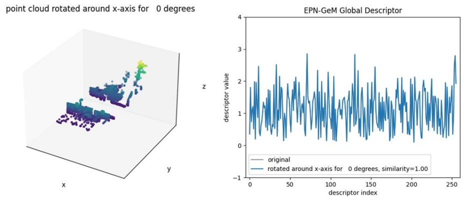
How do we discretize $SL(3)$?

Iwasawa Decomposition

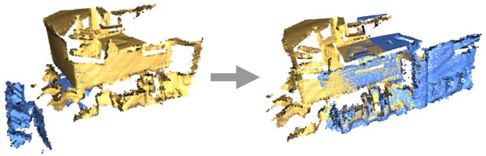
$$H = KAN, H \in SL(3)$$

$$K \in SO(3) \quad A = \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & \frac{1}{ab} \end{pmatrix} \quad N = \begin{pmatrix} 1 & z & x \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix}$$

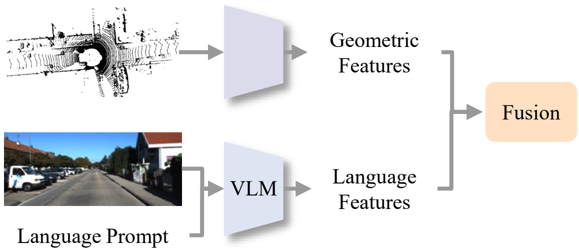
Conclusion



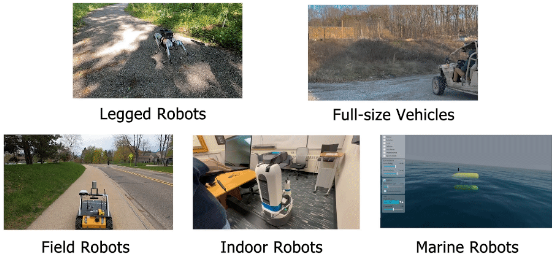
Place Recognition



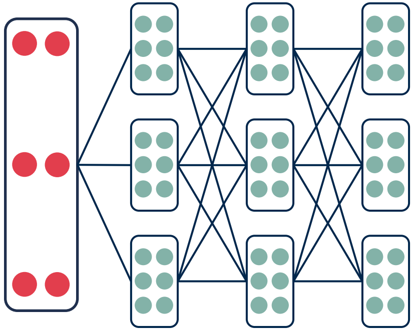
Point Cloud Registration



Foundation Models



Proprioceptive State Estimation



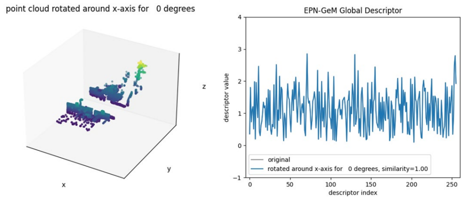
Lie Algebraic Neuron Networks

Perspective Equivariant Representation Learning

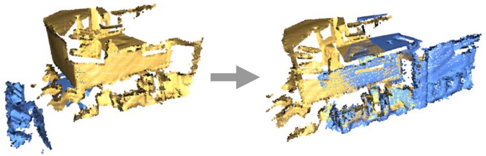
Open-sourced Software

- https://github.com/UMich-CURLY/se3_equivariant_place_recognition
- https://github.com/UMich-CURLY/SE3ET
- https://github.com/UMich-CURLY/drift
- https://github.com/UMich-CURLY/LieNeurons

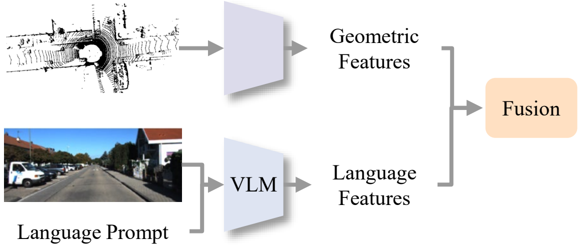
Questions?



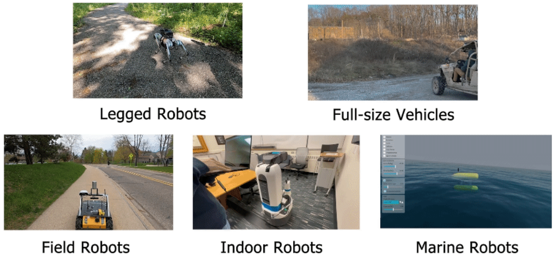
Place Recognition



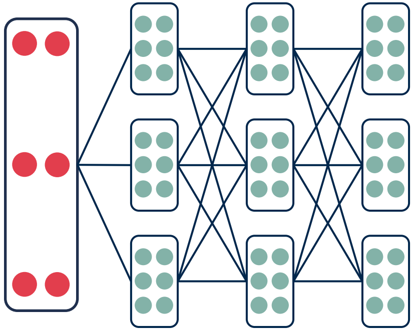
Point Cloud Registration



Foundation Models



Proprioceptive State Estimation



Lie Algebraic Neuron Networks

Perspective Equivariant Representation Learning

ROBOTICS CURLY Explore the Unknown