Leveraging Symmetry for Learning in Physical Systems

Robin Walters

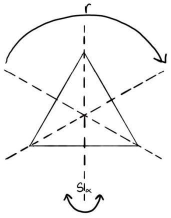
Khoury College of Computer Sciences, Northeastern University

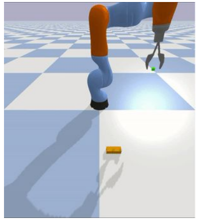
IAIFI Senior Investigator

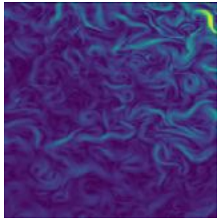
Visiting Faculty Fellow, Robotics and AI Institute

Guest Lecture

November 18, 2025



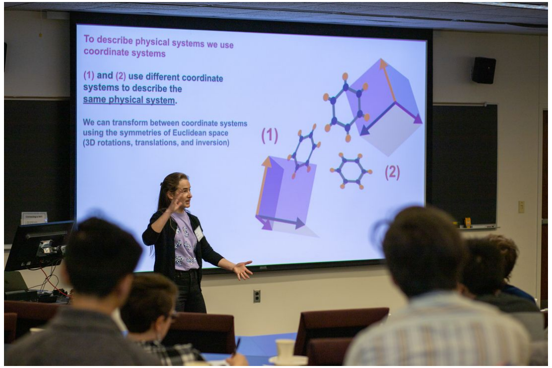




Boston Symmetry Day

Events every semester

Sign up for our mailing list at https://bostonsymmetry.github.io/.



TILOS
The Institute for Learning-Enabled Optimization at Scale (TILOS)

Boston Dynamics AI INSTITUTE
The AI Institute

Liquid
Liquid AI

Genentech
Genentech

Achira_
Achira

FODSI
Foundations of Data Science Institute (FODSI)

MITSUBISHI ELECTRIC Changes for the Better
MITSUBISHI ELECTRIC RESEARCH LABORATORIES
Mitsubishi Electric Research Laboratories (MERL)

MERCK
Merck

BOSTON Symmetry Group

Stefanie Jegelka
Associate Professor
MIT / TU Munich

Hannah Lawrence
PhD Student
MIT

Derek Lim
PhD Student
MIT

Jung Yeon Park
PhD Student
Northeastern

Robin Walters
Assistant Professor
Northeastern

Jigyasa Nigam
Postdoc
MIT

Geometric Learning Lab @Northeastern

Geometric Learning Lab
Khoury College of Computer Science
Northeastern University

Students


Lucas Laird
PhD Student


Haojie Huang
PhD Student


Neel Sortur
Undergrad Student


Edward Berman
Undergrad Student


Parnian Zameni
PhD Student


Claire Schlesinger
PhD Student


Luisa Li
Undergrad Student


Purvik Patel
Grad Student


Yuanyuan Yang
Grad Student


Yuxuan Chen
PhD Student


Marco Pacini
Visiting PhD Student


Vedanth M. Nilabh
Masters Student

Recent Grads


Dian Wang
PhD
Postdoctoral Researcher at Stanford University

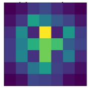

John Park
PhD


David Klee
Postdoc


Circe Hsu
Undergrad
Machine Learning Scientist at Matterworks

Research Overview

Approximate and Discovered Symmetry

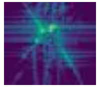


Approximate Equivariance in RL **AISTATS '25**
Topological obstructions and how to avoid them **NeurIPS '23**
Equivariant pose prediction via induced representations **NeurIPS '23**
Image to sphere: learning equivariant features for pose **Oral📣** **ICLR '23**
Approximately equivariant networks for dynamics **ICML '22**
Learning symmetric embeddings for equivariant world models **ICML '22**
Compositional generalization in object-oriented world modeling **ICML '22**
Generative adversarial symmetry discovery **ICML '23**
Symmetry discovery with lie algebra convolutional network **NeurIPS '21**

Dynamics and PDEs

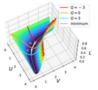
Dynamics over meshes with gauge equivariant messages **NeurIPS '23**
Meta-learning dynamics forecasting using task inference **NeurIPS '22**
Trajectory prediction using equivariant continuous convolution **ICLR '21**
Incorporating symmetry into dynamics models for generalization **ICLR '21**
Physics-informed ML for weather and climate **J Phil Trans Royal Soc A '20**

Applications



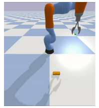
Probabilistic symmetry for multi-agent dynamics **L4DC '23**
Symmetric Models for Radar Response Modeling **NeurReps '23**
Integrating symmetry into differentiable planning **ICLR '23**
Physics-guided DL for traffic forecasting **L4DC '21**

Symmetry and Theory



The Empirical Impact of Parameter Symmetry **Best Paper🏆** **HiLD '24**
A general theory of correct, incorrect, and extrinsic equivariance **NeurIPS '23**
Physics-guided DL for spatiotemporal forecasting. **Knowledge-Guided ML '23**
Equivariant models in domains with latent symmetry **Spotlight💡** **ICLR '23**
Convergence and generalization using parameter symmetries **Oral📣** **ICLR '24**
Symmetries, flat minima, and the conserved quantities of gradient flow **ICLR '23**
Symmetry teleportation for accelerated optimization **NeurIPS '22**

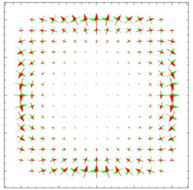
Robotics



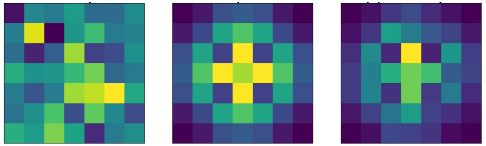
Equivariant Diffusion Policy **Outstanding Finalist🏆** **CoRL '24**
Fourier transporter: bi-equivariant robotic manipulation in 3D **ICLR '24**
On robot grasp learning using equivariant models **Autonomous Robots '23**
One-shot imitation learning via interaction warping **CoRL '23**
SEIL: Simulation-augmented equivariant imitation learning **ICRA '23**
A graph-based SE(3)-invariant approach to grasp detection **ICRA '23**
Leveraging symmetries in pick and place **IJRR '23**
On-robot learning with equivariant models **CoRL '23**
Equivariant transporter network **RSS '22**
Sample efficient grasp learning using equivariant models **RSS '22**
SO(2) equivariant reinforcement learning **ICLR '22**
Equivariant Q-learning in spatial action spaces **CoRL'21**

Talk Outline

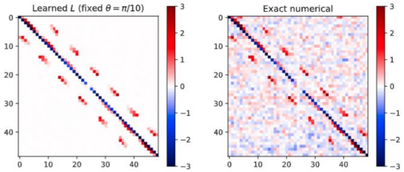
1. **Equivariant Neural Networks**



2. **Approximately Equivariant Networks**



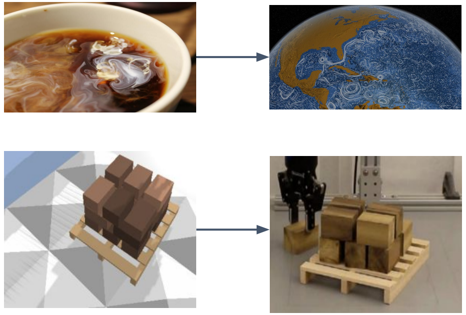
3. **Symmetry Discovery**



I. Equivariant Neural Networks for Physical Applications

Problems with Current ML

Poor Generalization

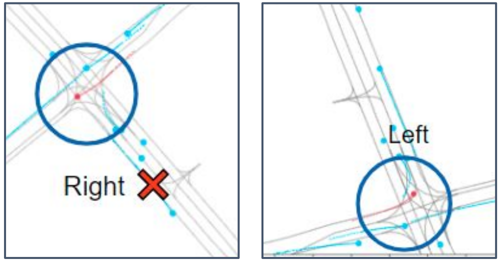


Models cannot generalize across settings

Data Hungry

Domains where data is expensive to collect

Physical Inconsistency



Physical inaccuracy undermines trust

Problems with Current ML

Poor Generalization

Data Hungry

Physical Inconsistency

Solutions?

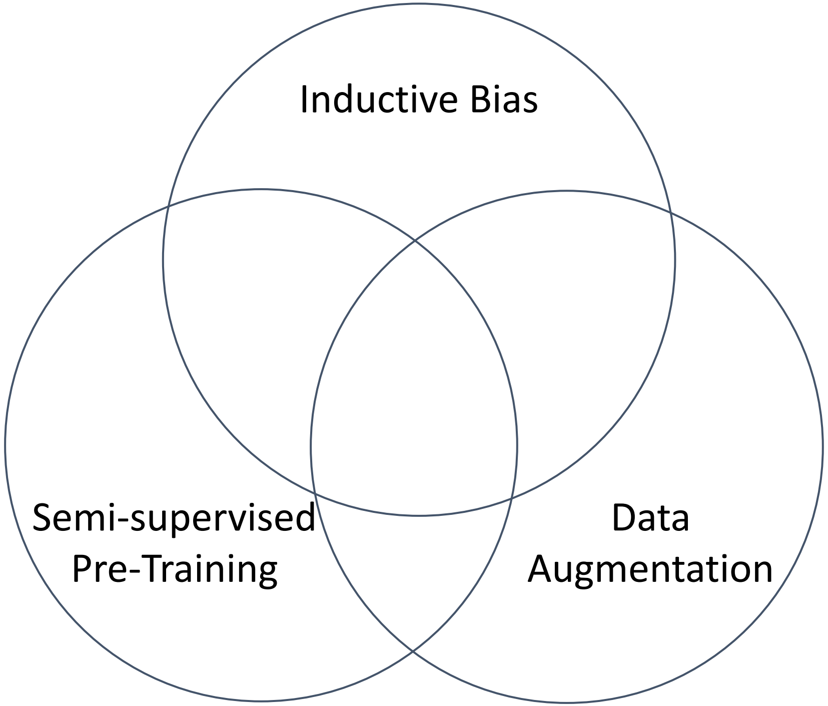
Problems with Current ML

Poor Generalization

Data Hungry

Physical Inconsistency

Solutions?



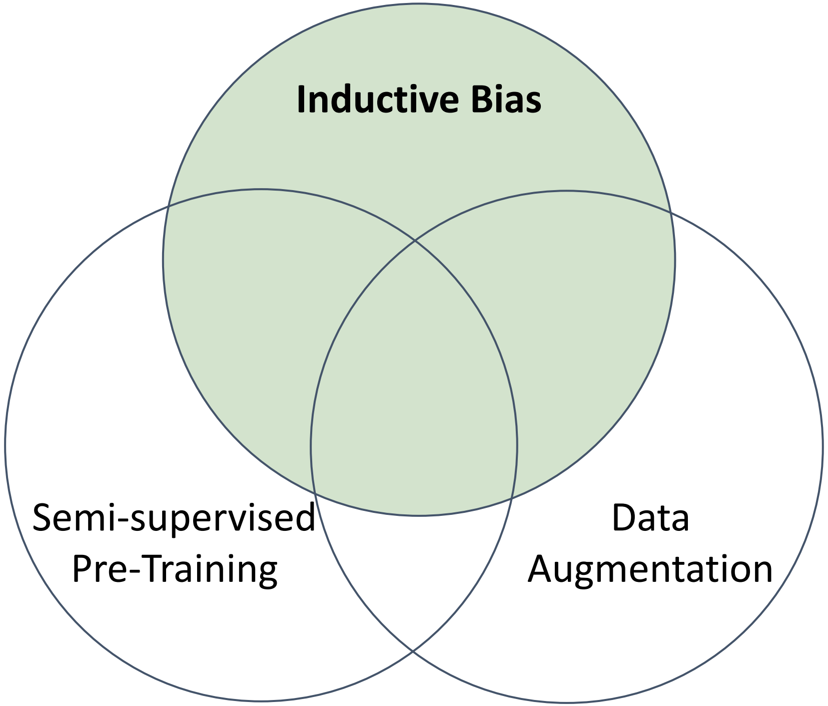
Problems with Current ML

Poor Generalization

Data Hungry

Physical Inconsistency

Solutions?



- We focus on **inductive bias.**
- But solutions are not mutually exclusive.

Symmetry in Deep Learning

Poor Generalization

Data Hungry

Physical Inconsistency

Incorporate symmetry as inductive bias into models

Generalize across symmetry

Automatic Data Augmentation

Symmetry and Conservation

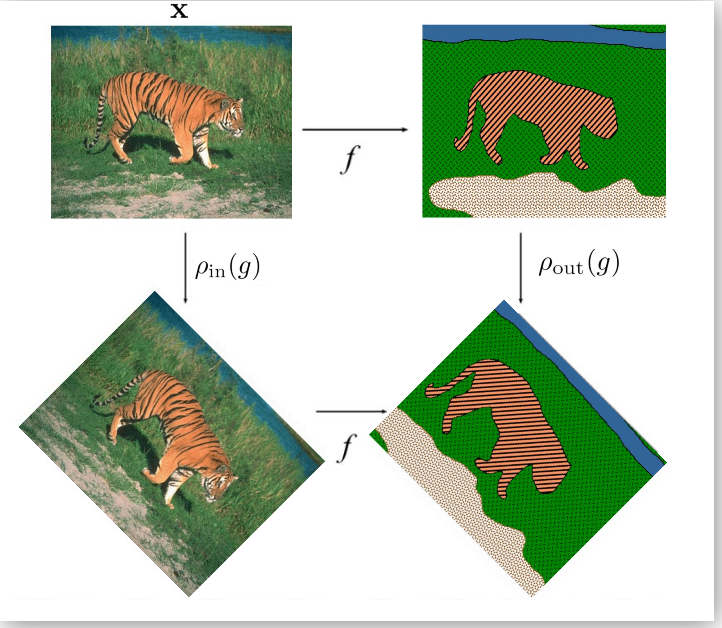
Equivariant Functions

Model functions which are **equivariant** with respect to symmetry group.

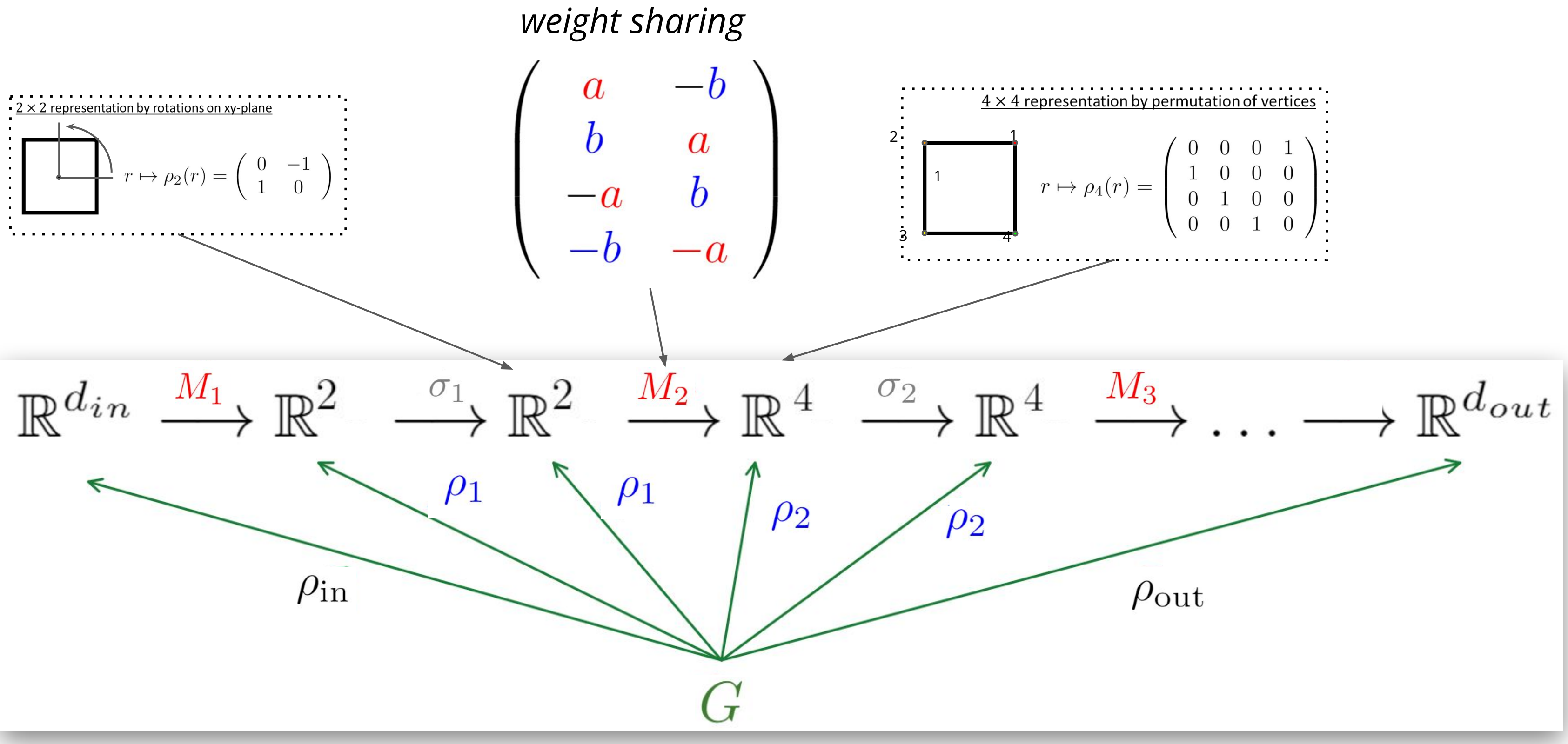
$$f : \mathbb{R}^{d_{\text{in}}} \to \mathbb{R}^{d_{\text{out}}}$$

f is G-**equivariant** if

$$f(\rho_{\text{in}}(g)\mathbf{x}) = \rho_{\text{out}}(g) f(\mathbf{x})$$

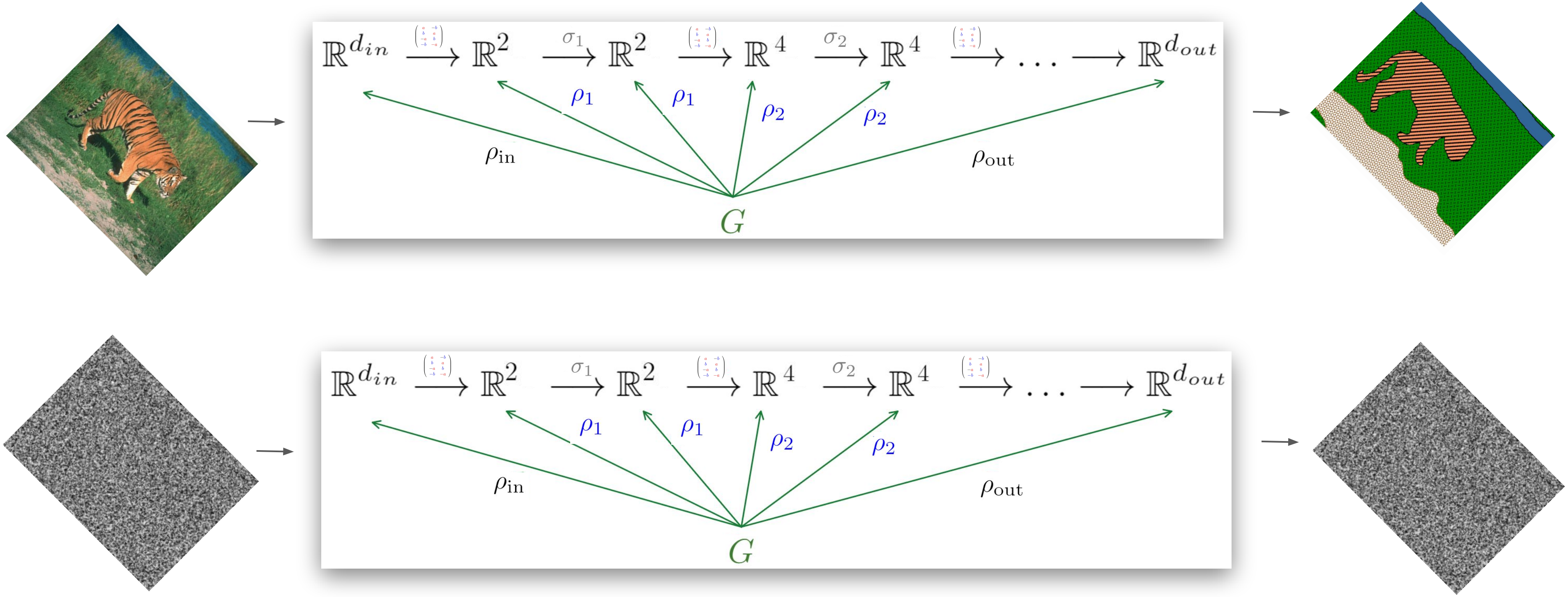


Equivariant Neural Networks



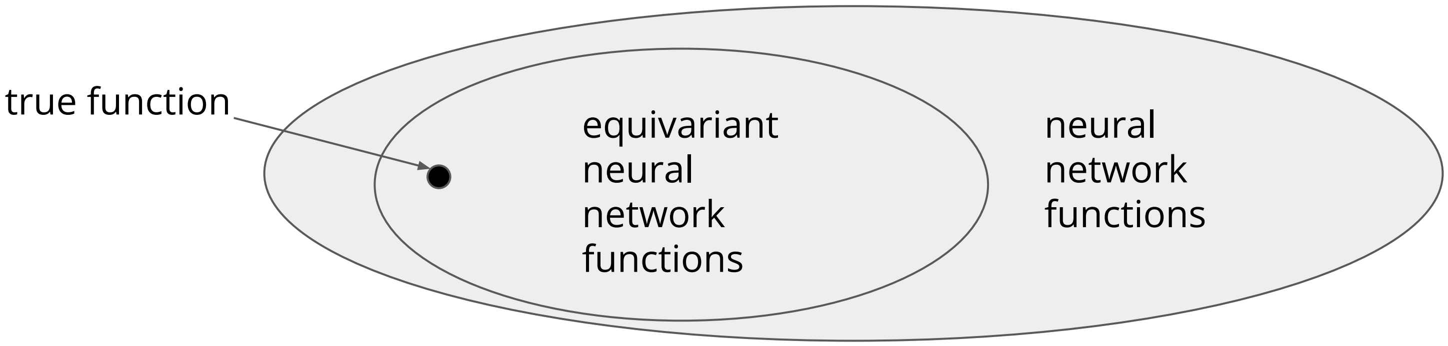
Equivariant Linear Layers and Equivariant Activations ⇒ Equivariant Network

Equivariant Neural Networks



Constraint applies at all values of parameters and all inputs, not just trained model

Equivariant Neural Networks

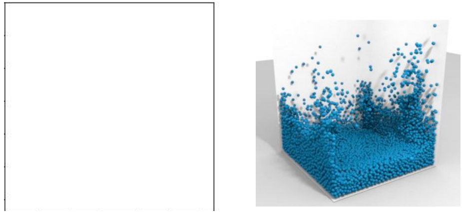


$$f \in \{f_\theta \text{ equivariant}\} \subset \{f_\theta\}$$

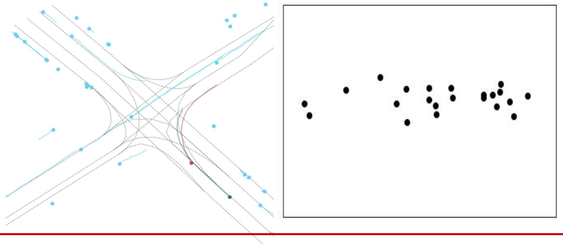
Equivariant neural networks narrow the search for the optimal model to smaller class of functions

Applications of Equivariant Neural Networks

Fluid Dynamics

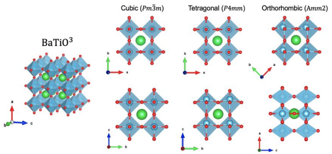


Trajectory Predictions

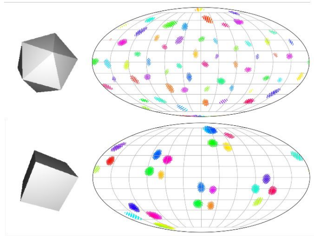


Materials

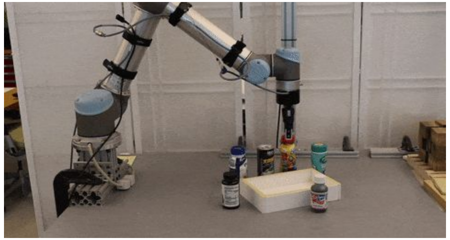
Crystal Structure



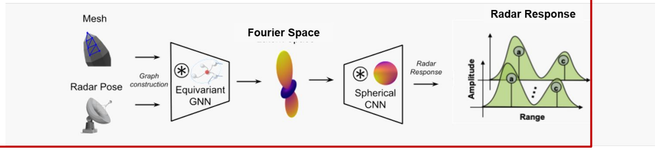
Computer Vision



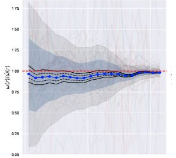
Robotics



RADAR

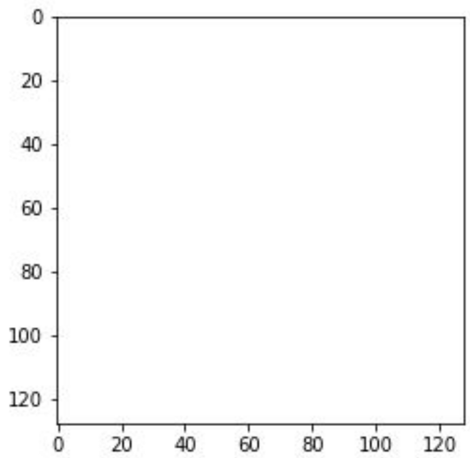


Astrophysics



1. Robin Walters*, Jinxi Li*, and Rose Yu. "Trajectory Prediction using Equivariant Continuous Convolution." *ICLR*, 2021.
2. Klee, David M., Ondrej Biza, Robert Platt, and Robin Walters. "Image to sphere: Learning equivariant features for efficient pose prediction." *ICLR*, 2023.
3. Wang, Dian, Robin Walters, and Robert Platt. SO(2)-Equivariant Reinforcement Learning. *ICLR*, 2022.
4. Robin Walters*, Rui Wang*, and Rose Yu. "Incorporating Symmetry into Deep Dynamics Models for Improved Generalization." *ICLR*, 2021.
5. Wang, Rui, Robin Walters, and Tess E. Smidt. "Relaxed Octahedral Group Convolution for Learning Symmetry Breaking in 3D Physical Systems." *arXiv preprint arXiv:2310.02299* (2023).

Turbulent Flow Prediction



Applications

- Climate Science
- Mechanical Engineering
- Medical Devices (blood)

Robin Walters*, **Rui Wang***,, and **Rose Yu**. "Incorporating Symmetry into Deep Dynamics Models for Improved Generalization." International Conference on Learning Representations (ICLR), 2021.

Equivariance of Forward Prediction Function

The forward prediction function

$$f_\theta \colon (V_{t-k}, \dots, V_t) \mapsto V_{t+1}$$

is equivariant wrt these sym.

i.e. if g is a transformation (rotation, translation, scaling,...) then

$$f_\theta \colon (gV_{t-k}, \dots, gV_t) \mapsto gV_{t+1}$$

Goal: *Incorporate this equivariance property of* D_{NS} *into model* f_Θ .

Navier Stokes Equation

Nonlinear PDE governing fluid flow.

$$\frac{\partial \boldsymbol{v}}{\partial t} = -(\boldsymbol{v} \cdot \nabla)\boldsymbol{v} - \frac{1}{\rho_0}\nabla p + \nu \nabla^2 \boldsymbol{v}$$

$$\nabla \cdot \boldsymbol{v} = 0$$

Invariant to scaling:

$$T_\lambda^{sc}\boldsymbol{v}(\boldsymbol{x}, t) = \lambda \boldsymbol{v}(\lambda \boldsymbol{x}, \lambda^2 t),\ \lambda \in \mathbb{R}_{>0}$$

Example: Deriving Scaling Symmetry

$$\left.\begin{aligned} x &\mapsto \lambda \tilde{x} \\ t &\mapsto \lambda^2 \tilde{t} \end{aligned}\right\}$$

$$\frac{\partial \mathbf{v}}{\partial t} = -(\mathbf{v} \cdot \nabla)\mathbf{v} - \frac{1}{\rho_0}\nabla p + \nu \nabla^2 \mathbf{v} + f \qquad (\mathcal{D}_{\mathrm{NS}})$$

Example: Deriving Scaling Symmetry

$$\left.\begin{aligned} x &\mapsto \lambda \tilde{x} \\ t &\mapsto \lambda^2 \tilde{t} \end{aligned}\right\} \qquad \begin{aligned} v &\mapsto \frac{1}{\lambda} \tilde{v} \\ p &\mapsto \frac{1}{\lambda^2} p \end{aligned}$$

$$\frac{\partial \mathbf{v}}{\partial t} = -(\mathbf{v} \cdot \nabla)\mathbf{v} - \frac{1}{\rho_0}\nabla p + \nu \nabla^2 \mathbf{v} + f \qquad (\mathcal{D}_{\mathrm{NS}})$$

Example: Deriving Scaling Symmetry

$$\left.\begin{aligned} x &\mapsto \lambda \tilde{x} \\ t &\mapsto \lambda^2 \tilde{t} \end{aligned}\right\} \qquad \begin{aligned} v &\mapsto \frac{1}{\lambda} \tilde{v} \\ p &\mapsto \frac{1}{\lambda^2} \tilde{p} \end{aligned}$$

$$\frac{\partial \mathbf{v}}{\partial t} = -(\mathbf{v} \cdot \nabla)\mathbf{v} - \frac{1}{\rho_0}\nabla p + \nu \nabla^2 \mathbf{v} + \cancel{f} \qquad (\mathcal{D}_{\mathrm{NS}})$$

$$\frac{\left(\frac{1}{\lambda}\right)\partial \tilde{v}}{\lambda^2 \partial t} = -\left(\frac{1}{\lambda}\tilde{v} \cdot \frac{1}{\lambda}\tilde{\nabla}\right)\frac{1}{\lambda}\tilde{v} - \frac{1}{\rho_0}\frac{1}{\lambda}\tilde{\nabla}\frac{1}{\lambda^2}\tilde{p} + \nu\frac{1}{\lambda^2}\tilde{\nabla}^2\tilde{v}$$

Example: Deriving Scaling Symmetry

$$\left.\begin{aligned} x &\mapsto \lambda \tilde{x} \\ t &\mapsto \lambda^2 \tilde{t} \end{aligned}\right\} \qquad \begin{aligned} v &\mapsto \frac{1}{\lambda} \tilde{v} \\ p &\mapsto \frac{1}{\lambda^2} \tilde{p} \end{aligned}$$

$$\frac{\partial \mathbf{v}}{\partial t} = -(\mathbf{v}\cdot\nabla)\mathbf{v} - \frac{1}{\rho_0}\nabla p + \nu \nabla^2 \mathbf{v} + \cancel{f} \qquad (\mathcal{D}_{\mathrm{NS}})$$

$$\frac{\left(\frac{1}{\lambda}\right)\partial \tilde{v}}{\lambda^2\, \partial t} = -\left(\frac{1}{\lambda}\tilde{v} \cdot \frac{1}{\lambda}\tilde{\nabla}\right)\frac{1}{\lambda}\tilde{v} - \frac{1}{\rho_0}\frac{1}{\lambda}\tilde{\nabla}\frac{1}{\lambda^2}\tilde{p} + \nu \frac{1}{\lambda^2}\tilde{\nabla}^2 \frac{1}{\lambda}\tilde{v}$$

λ^{-3} $\qquad$ λ^{-3}

Example: Deriving Scaling Symmetry

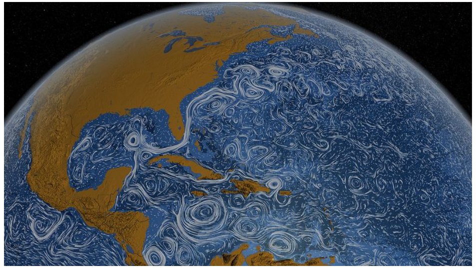
$$\left.\begin{aligned} x &\longmapsto \lambda \tilde{x} \\ t &\longmapsto \lambda^2 \tilde{t} \end{aligned}\right\} \qquad \begin{aligned} v &\longmapsto \frac{1}{\lambda} \tilde{v} \\ p &\longmapsto \frac{1}{\lambda^2} \tilde{p} \end{aligned}$$

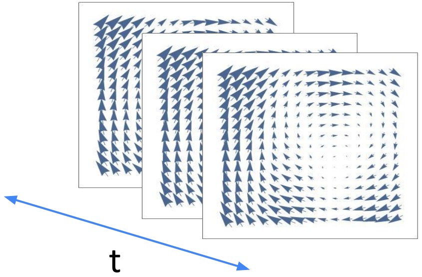
$$\frac{\partial \mathbf{v}}{\partial t} = -(\mathbf{v} \cdot \nabla)\mathbf{v} - \frac{1}{\rho_0}\nabla p + \nu \nabla^2 \mathbf{v} + \cancel{f} \qquad (\mathcal{D}_{\mathrm{NS}})$$

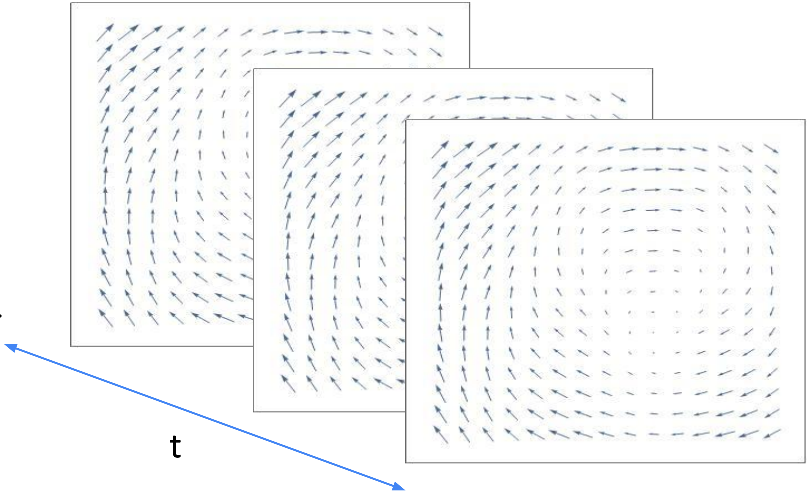
$$\frac{\left(\frac{1}{\lambda}\right)\partial \tilde{v}}{\lambda^2 \, \partial \tilde{t}} = -\left(\frac{1}{\lambda}\tilde{v} \cdot \frac{1}{\lambda}\tilde{\nabla}\right)\frac{1}{\lambda}\tilde{v} - \frac{1}{\rho_0}\frac{1}{\lambda}\tilde{\nabla}\frac{1}{\lambda^2}\tilde{p} + \nu \frac{1}{\lambda^2}\tilde{\nabla}^2 \frac{1}{\lambda}\tilde{v}$$

$$\cancel{\lambda^3} \qquad \qquad \cancel{\lambda^3}$$

Symmetry: Scaling



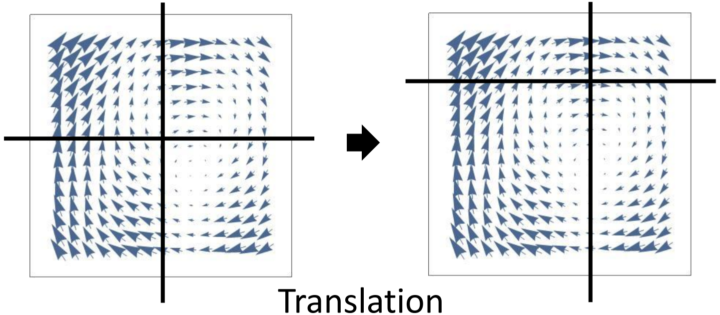




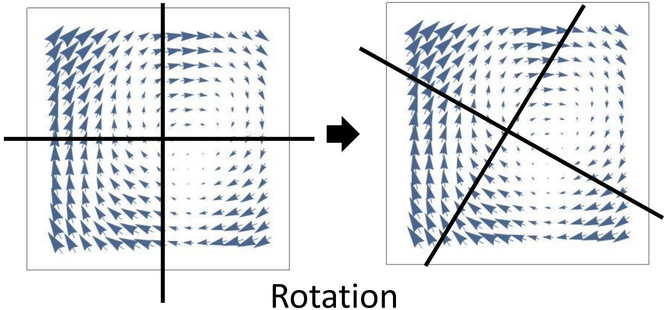
$$T^{sc}_{\lambda}\boldsymbol{v}(\boldsymbol{x},t) = \lambda\boldsymbol{v}(\lambda\boldsymbol{x},\lambda^2 t),\ \lambda \in \mathbb{R}_{>0}$$

Symmetry: Galilean Invariance

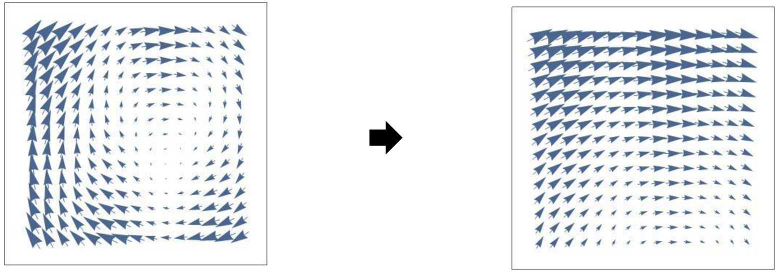
independence of physics from *arbitrary choice* of coordinate system □ Galilean invariance



Translation



Rotation



Uniform Motion

To a NN, these data all look very different

Scaling: Truncated G-Convolution

Standard convolution shares weights by translating a kernel across the input.

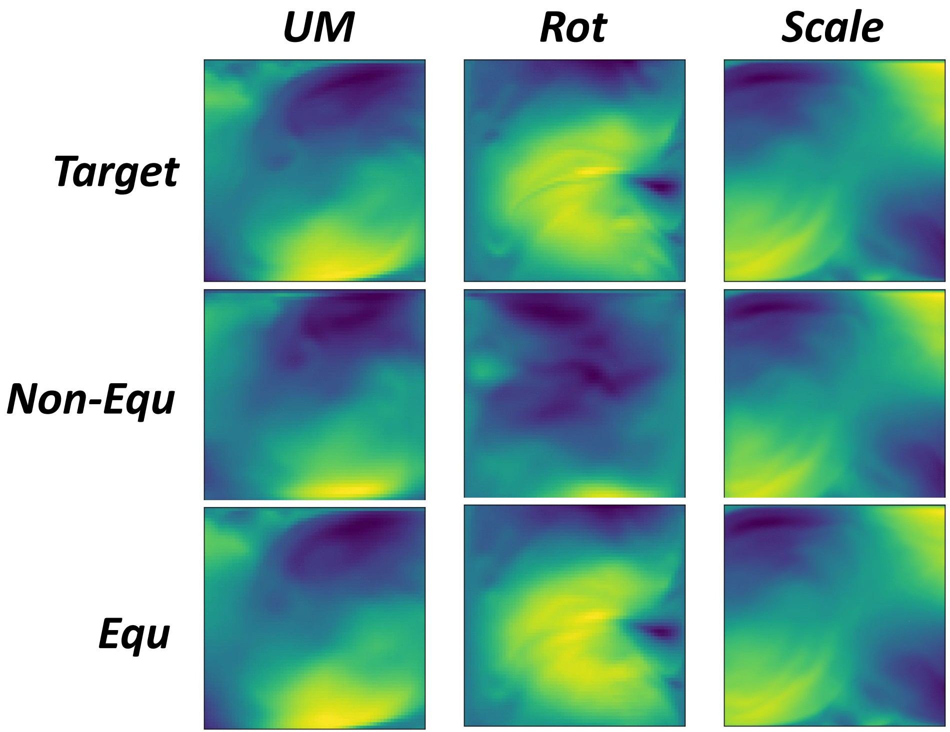
$$\boldsymbol{v}_2(\boldsymbol{p}) = \sum_{\boldsymbol{q}\in\mathbb{Z}^2} \boldsymbol{v}_1(\boldsymbol{p}+\boldsymbol{q})K(\boldsymbol{q})$$

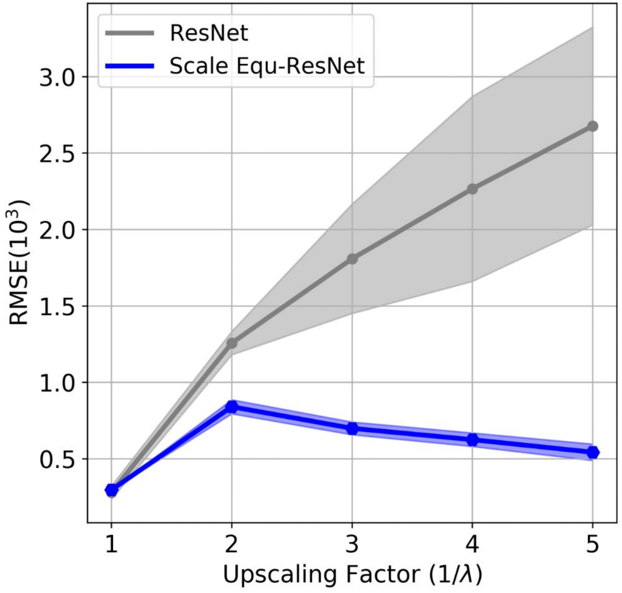
For **scale-equivariant convolution**, we translate *and scale* a kernel across the input

$$\boldsymbol{v}_2(\boldsymbol{p}, s, \mu) = \sum_{\underset{\text{scale}}{\lambda\in\mathbb{R}_{>0}},\, \underset{\text{time}}{t\in\mathbb{R}},\, \underset{\text{space}}{\boldsymbol{q}\in\mathbb{Z}^2}} \lambda \boldsymbol{v}_1(\lambda\boldsymbol{p}+\boldsymbol{q}, \lambda^2 t, \lambda\mu)K(\boldsymbol{q}, s, t, \lambda)$$

$$T^{sc}_{\lambda}\boldsymbol{v}(\boldsymbol{x}, t) = \lambda\boldsymbol{v}(\lambda\boldsymbol{x}, \lambda^2 t),\ \lambda\in\mathbb{R}_{>0}$$

Results: *Improved Generalization on Rayleigh-Bénard Convection*

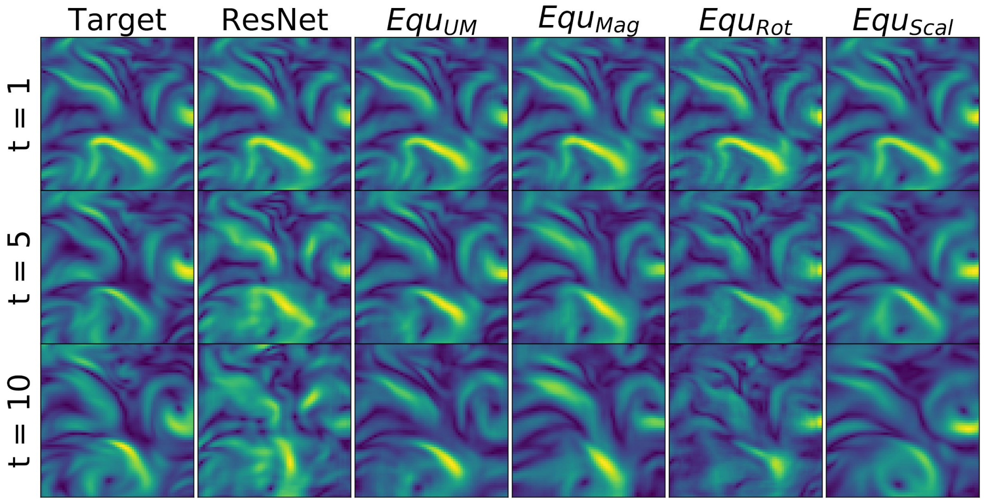




- EquNets Robust to Distributional Transformation.

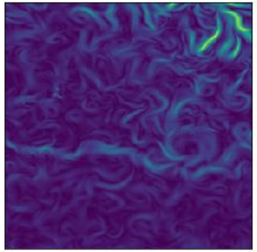
Walters*, Wang*, and Yu. "Incorporating Symmetry into Deep Dynamics Models for Improved Generalization." **ICLR**, 2021.

Results: *Real-world Ocean Currents Data*



	RMSE		ESE	
	$Test_{time}$	$Test_{domain}$	$Test_{time}$	$Test_{domain}$
ResNet	0.71±0.07	0.72±0.04	0.83±0.06	0.75±0.11
$\mathtt{Augm}_{\mathtt{UM}}$	0.70±0.01	0.70±0.07	1.06±0.06	1.06±0.04
$\mathtt{Augm}_{\mathtt{Mag}}$	0.76±0.02	0.71±0.01	1.08±0.08	1.05±0.8
$\mathtt{Augm}_{\mathtt{Rot}}$	0.73±0.01	0.69±0.01	0.94±0.01	0.86±0.01
$\mathtt{Augm}_{\mathtt{Scal}}$	0.97±0.06	0.92±0.04	0.85±0.03	0.95±0.11
$\mathtt{Equ}_{\mathtt{UM}}$	0.68±0.06	0.68±0.16	0.75±0.06	0.73±0.08
$\mathtt{Equ}_{\mathtt{Mag}}$	0.66±0.14	**0.68±0.11**	0.84±0.04	0.85±0.14
$\mathtt{Equ}_{\mathtt{Rot}}$	0.69±0.01	0.70±0.08	**0.43±0.15**	**0.28±0.20**
$\mathtt{Equ}_{\mathtt{Scal}}$	**0.63±0.02**	0.68±0.21	0.44±0.05	0.42±0.12

- Reanalysis ocean current data from NEMO engine.
- Approx 20 deg lat/long area in Indian, Pacific, Atlantic.
- Lower RMSE. Much lower ESE.



Equivariant Neural Networks for Fluid Dynamics

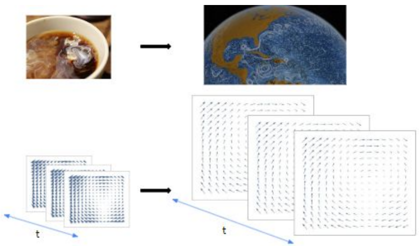
First principles model → Extract symmetries → Build network constrained by symmetry → Train to model system

Nonlinear PDE: Navier-Stokes Equations

$$\frac{\partial \boldsymbol{v}}{\partial t} = -(\boldsymbol{v} \cdot \nabla)\boldsymbol{v} - \frac{1}{\rho_0}\nabla p + \nu \nabla^2 \boldsymbol{v}$$

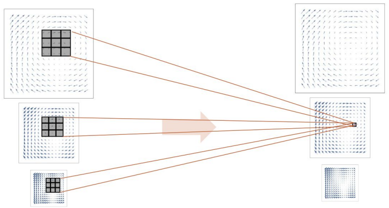
$$\nabla \cdot \boldsymbol{v} = 0$$

Galilean Group and Anisotropic Scaling Laws



$$T_\lambda^{sc}\boldsymbol{v}(\boldsymbol{x},t) = \lambda \boldsymbol{v}(\lambda \boldsymbol{x}, \lambda^2 t),\ \lambda \in \mathbb{R}_{>0}$$

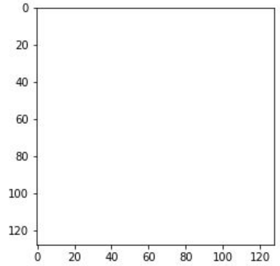
Group Convolution Neural Networks



$$\boldsymbol{v}_2(\boldsymbol{p}, s, \mu) = \sum_{\underset{\text{scale}}{\lambda\in\mathbb{R}_{>0}},\underset{\text{time}}{t\in\mathbb{R}},\underset{\text{space}}{\boldsymbol{q}\in\mathbb{Z}^2}} \lambda \boldsymbol{v}_1(\lambda \boldsymbol{p} + \boldsymbol{q}, \lambda^2 t, \lambda\mu) K(\boldsymbol{q}, s, t, \lambda)$$

$$\boxed{T_\lambda^{sc}\boldsymbol{v}(\boldsymbol{x},t) = \lambda \boldsymbol{v}(\lambda \boldsymbol{x}, \lambda^2 t),\ \lambda \in \mathbb{R}_{>0}}$$

Improved Physical Fidelity in Neural PDE Solvers

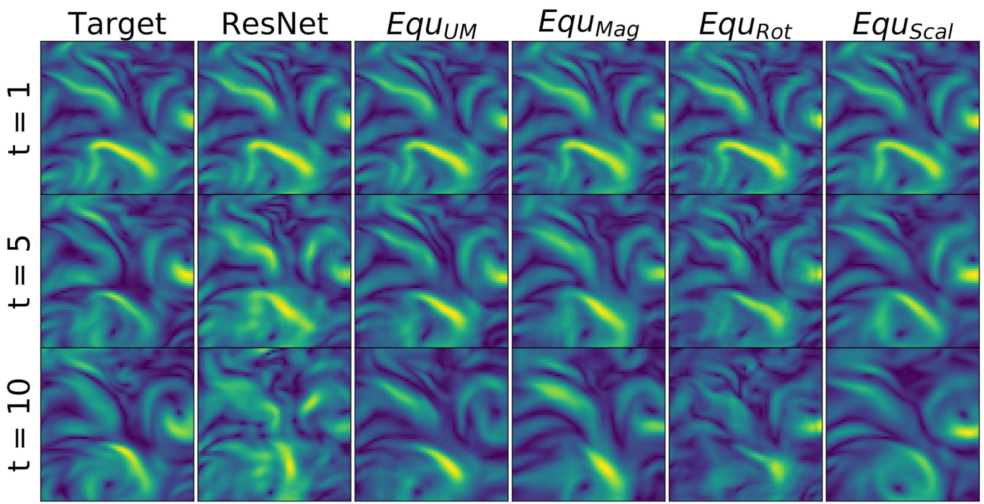


Provides a program for incorporating prior physical knowledge into deep learning.

Rui Wang*, Robin Walters*,, and Rose Yu. "Incorporating Symmetry into Deep Dynamics Models for Improved Generalization." *ICLR*, 2021.

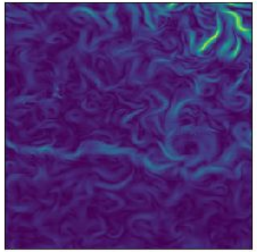
Equivariant Neural Networks for Fluid Dynamics

Real-world Ocean Currents Data

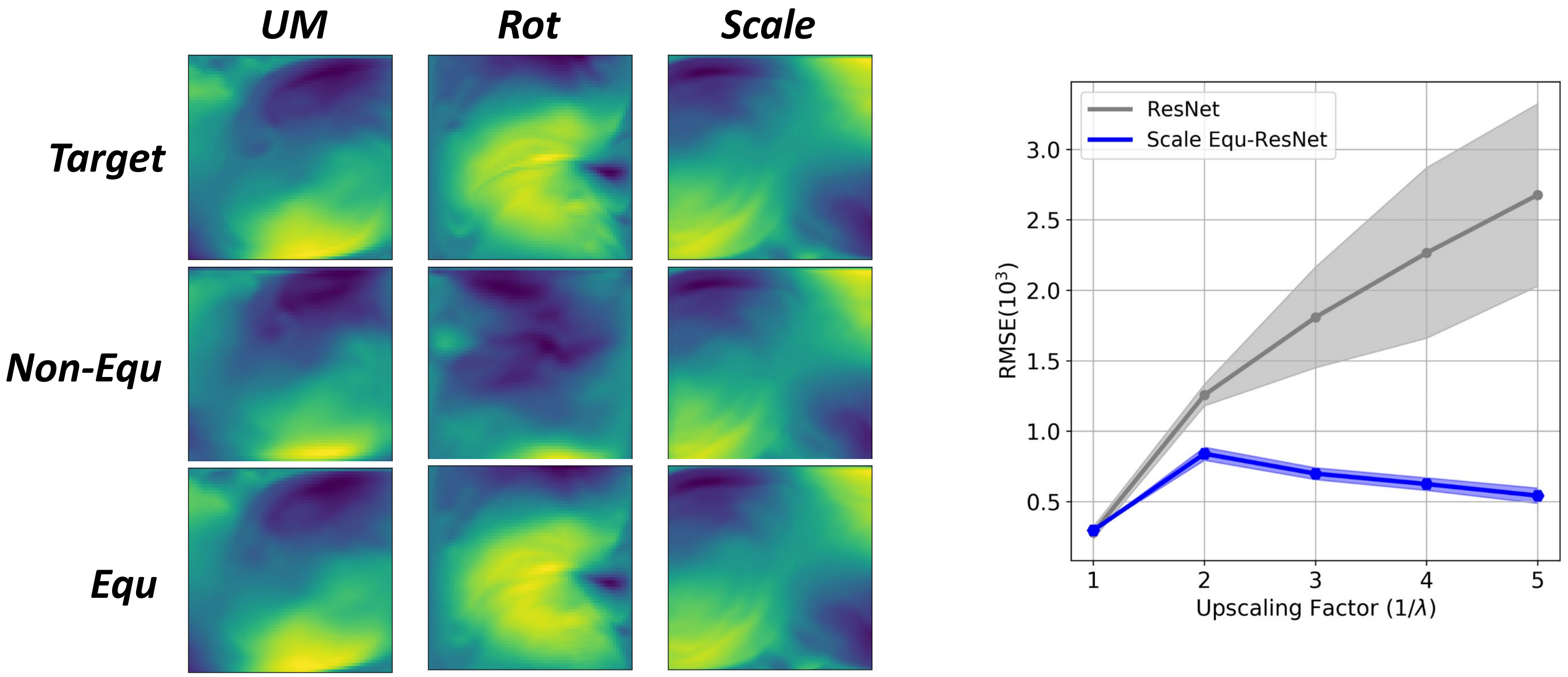


	RMSE		ESE	
	$Test_{time}$	$Test_{domain}$	$Test_{time}$	$Test_{domain}$
ResNet	0.71±0.07	0.72±0.04	0.83±0.06	0.75±0.11
$Augm_{UM}$	0.70±0.01	0.70±0.07	1.06±0.06	1.06±0.04
$Augm_{Mag}$	0.76±0.02	0.71±0.01	1.08±0.08	1.05±0.8
$Augm_{Rot}$	0.73±0.01	0.69±0.01	0.94±0.01	0.86±0.01
$Augm_{Scal}$	0.97±0.06	0.92±0.04	0.85±0.03	0.95±0.11
Equ_{UM}	0.68±0.06	0.68±0.16	0.75±0.06	0.73±0.08
Equ_{Mag}	0.66±0.14	**0.68±0.11**	0.84±0.04	0.85±0.14
Equ_{Rot}	0.69±0.01	0.70±0.08	**0.43±0.15**	**0.28±0.20**
Equ_{Scal}	**0.63±0.02**	0.68±0.21	0.44±0.05	0.42±0.12

- Reanalysis ocean current data from NEMO engine.
- Approx 20 deg lat/long area in Indian, Pacific, Atlantic.
- Lower RMSE. Much lower ESE.

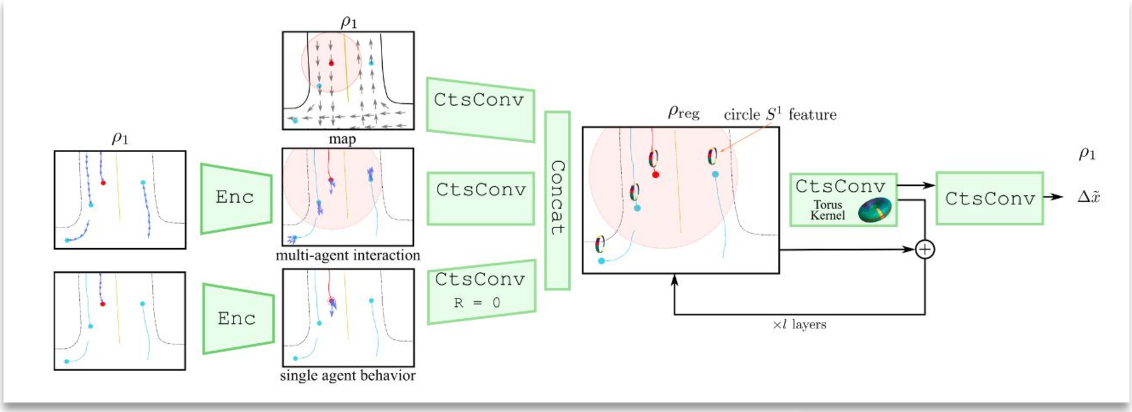


Equivariant Neural Networks for Fluid Dynamics

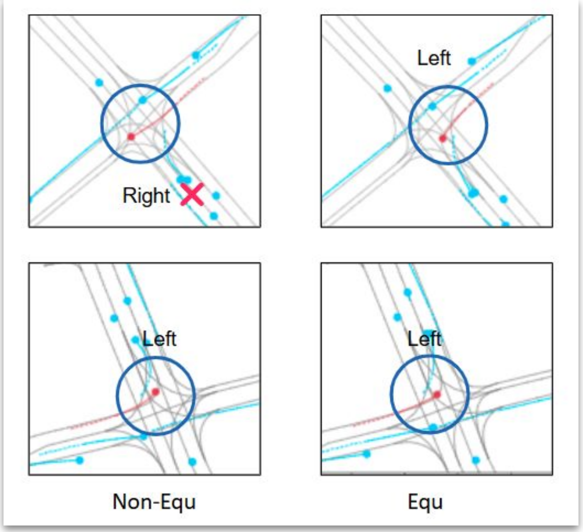


EquNets Robust to Distributional Transformation on *Rayleigh-Bénard Convection*

Equivariant Neural Networks for Trajectory Prediction

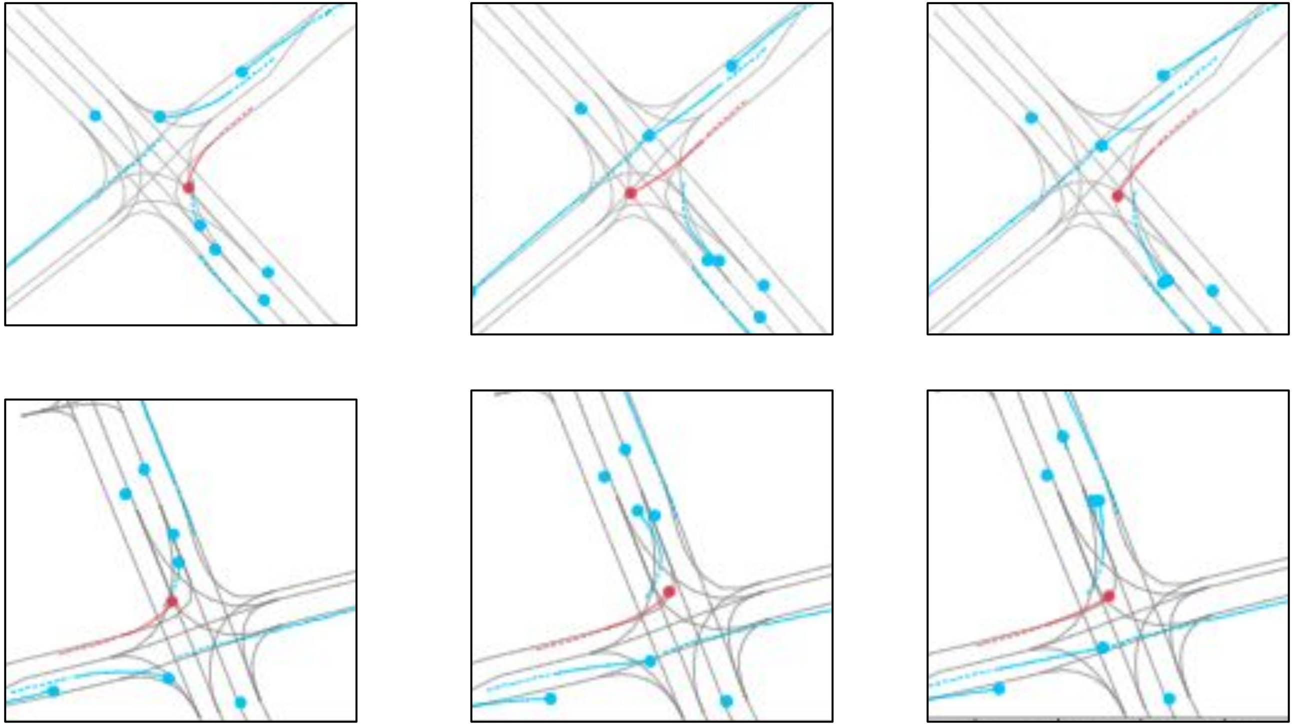


- ECCO: Equivariant Continuous Convolution
- Local Rotational Equivariance
- Consistent Outputs



Walters*, Li*, and Yu. "Trajectory Prediction using Equivariant Continuous Convolution." **ICLR**, 2021.

Equivariant Neural Networks for Trajectory Prediction

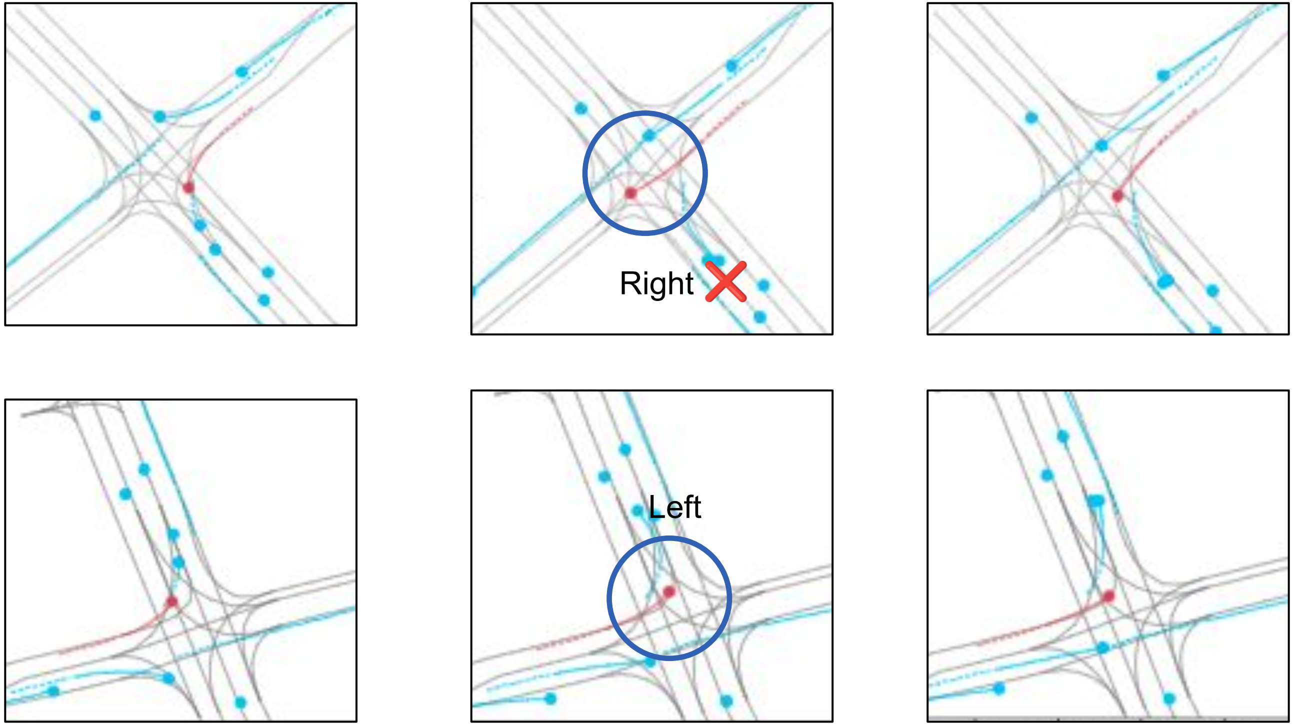


Truth Non-Equ Equ

Consistent Predictions

Walters*, Li*, and Yu. "Trajectory Prediction using Equivariant Continuous Convolution." **ICLR**, 2021.

Equivariant Neural Networks for Trajectory Prediction

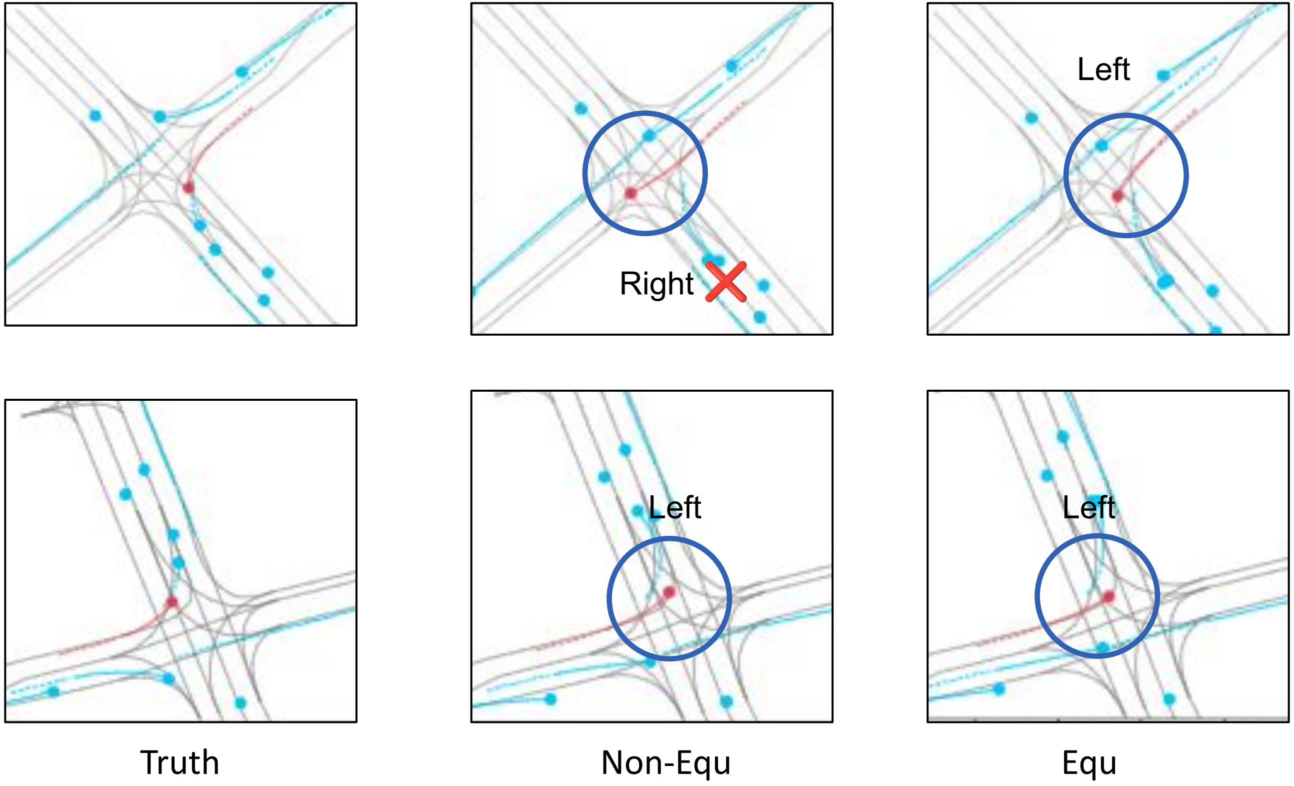


Truth   Non-Equ   Equ

Consistent Predictions

Walters*, Li*, and Yu. "Trajectory Prediction using Equivariant Continuous Convolution." **ICLR**, 2021.

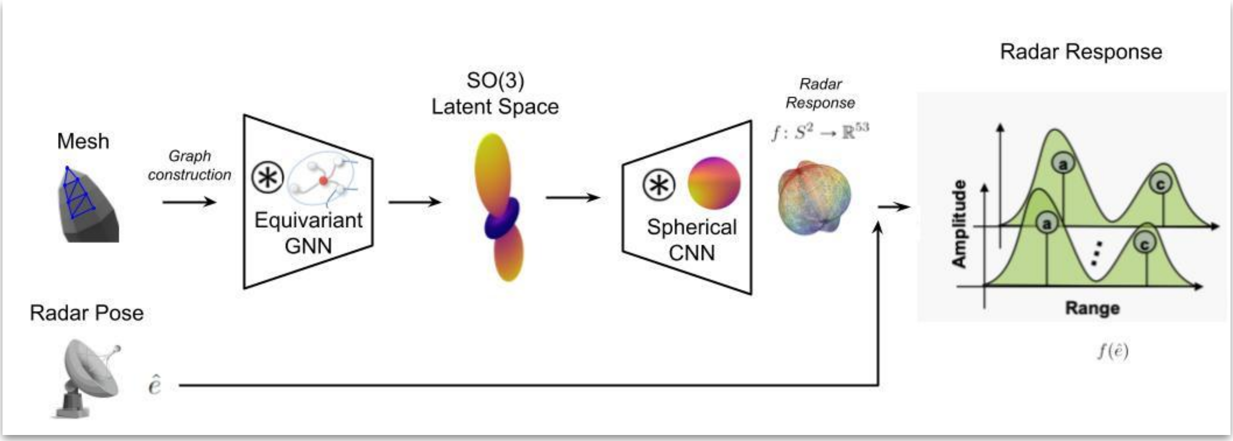
Equivariant Neural Networks for Trajectory Prediction



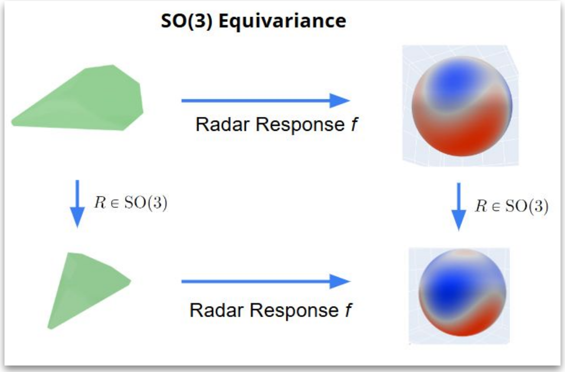
Consistent Predictions

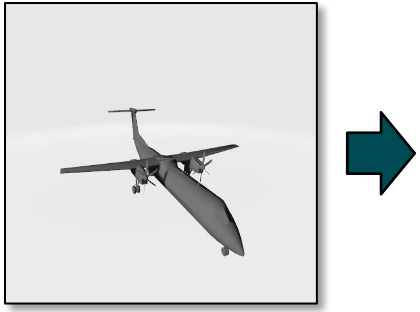
Walters*, Li*, and Yu. "Trajectory Prediction using Equivariant Continuous Convolution." **ICLR**, 2021.

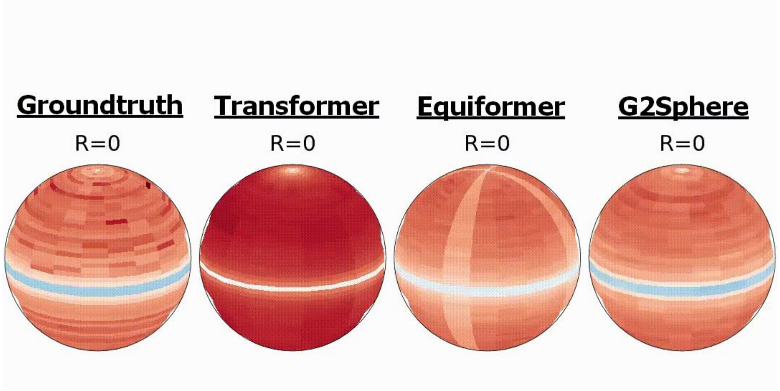
Equivariant Neural Networks for Radar Simulation



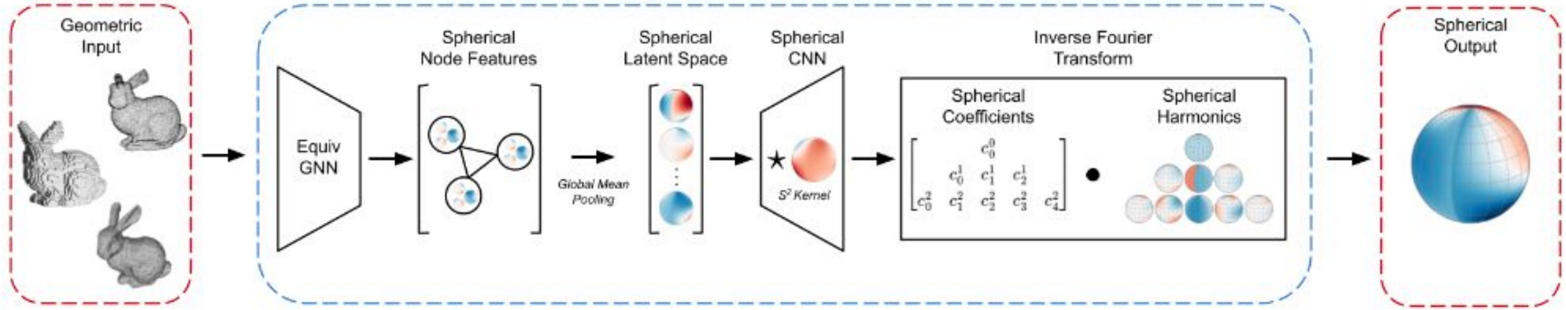
G2Sphere



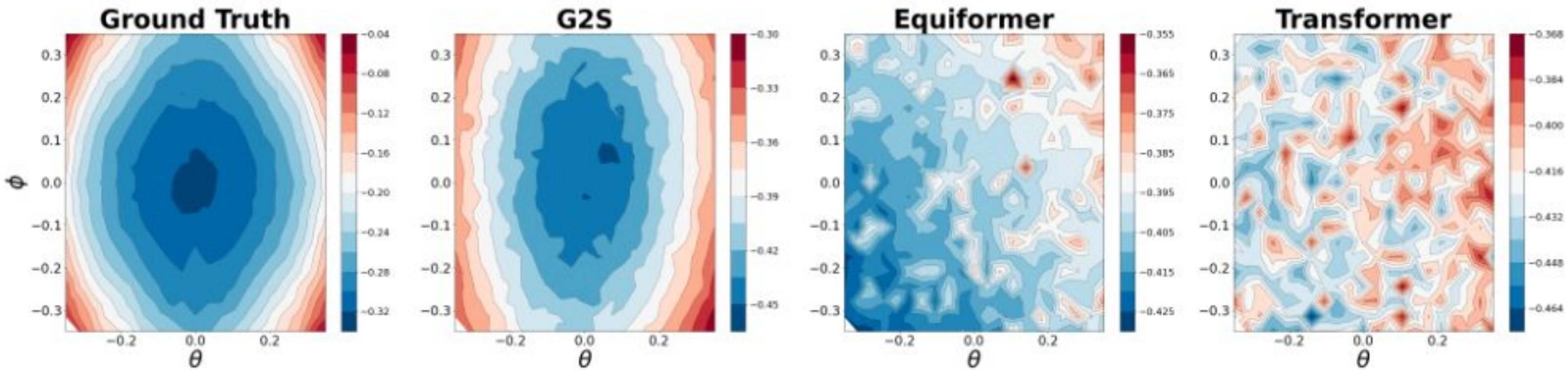




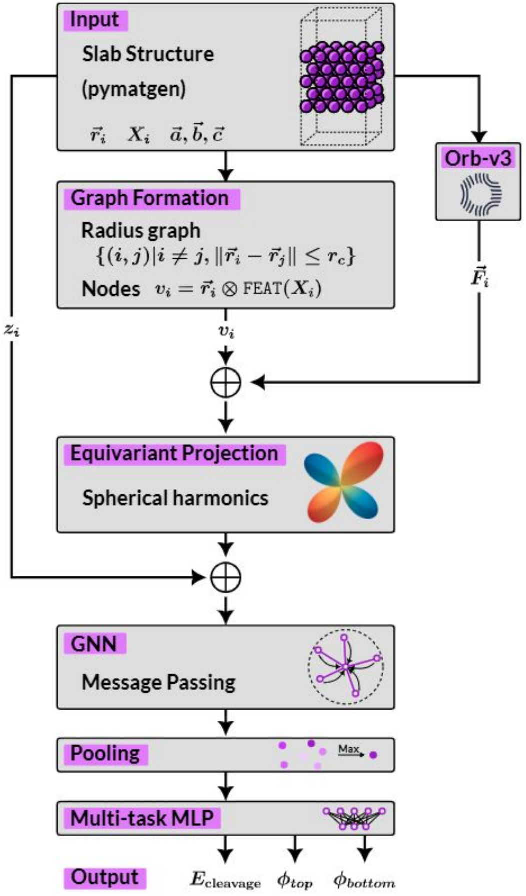
Equivariant Neural Networks for Aerodynamic Drag

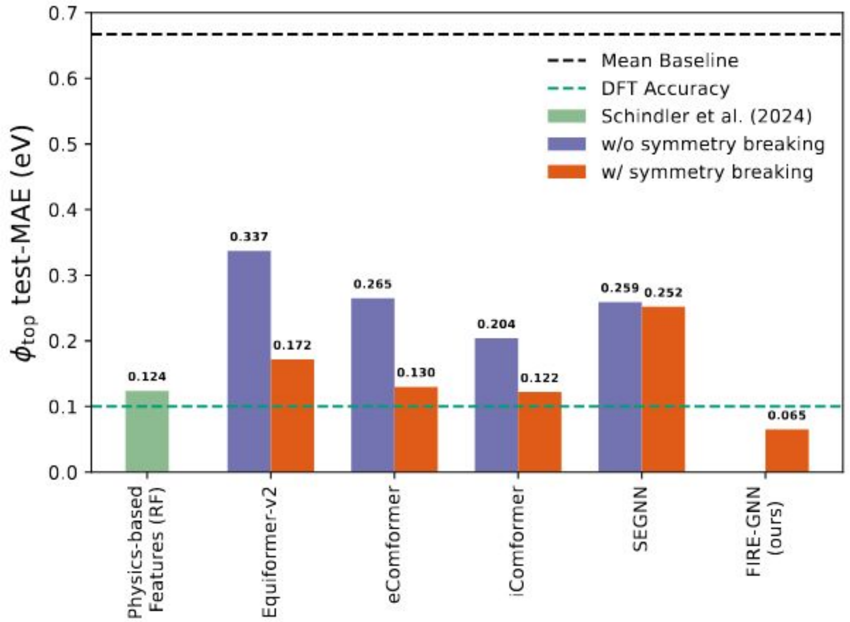


G2Sphere: Same architecture as for radar prediction



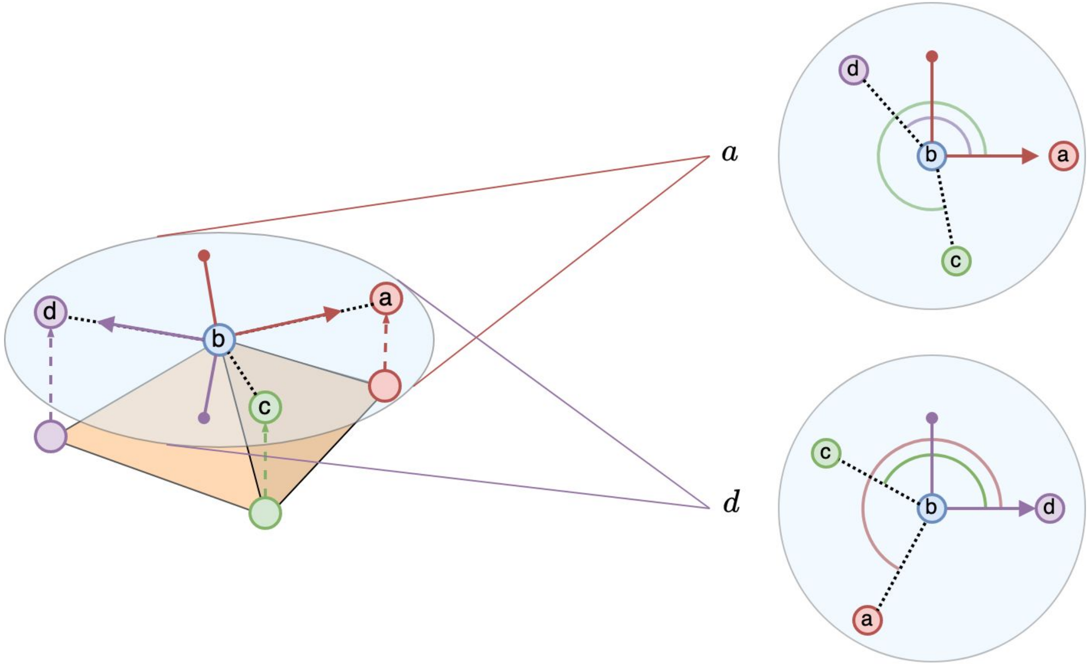
Equivariant Neural Networks for Surface Property Prediction

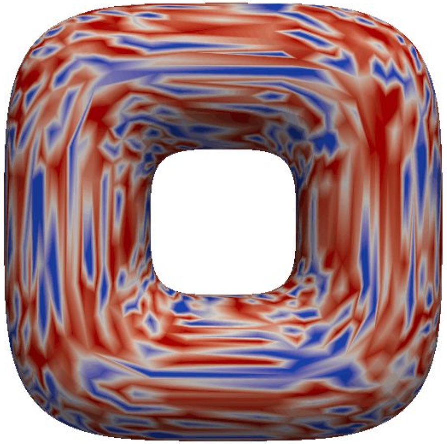




- Predict Work Function and cleavage energy for crystal slab.
- SE(3)-equiv MPNN with symmetry breaking features and force potentials

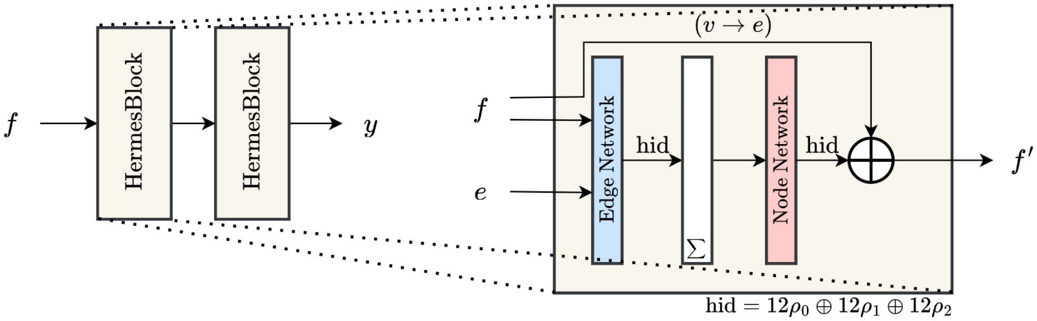
Gauge Equivariant Neural Networks for Learning PDEs over Meshes

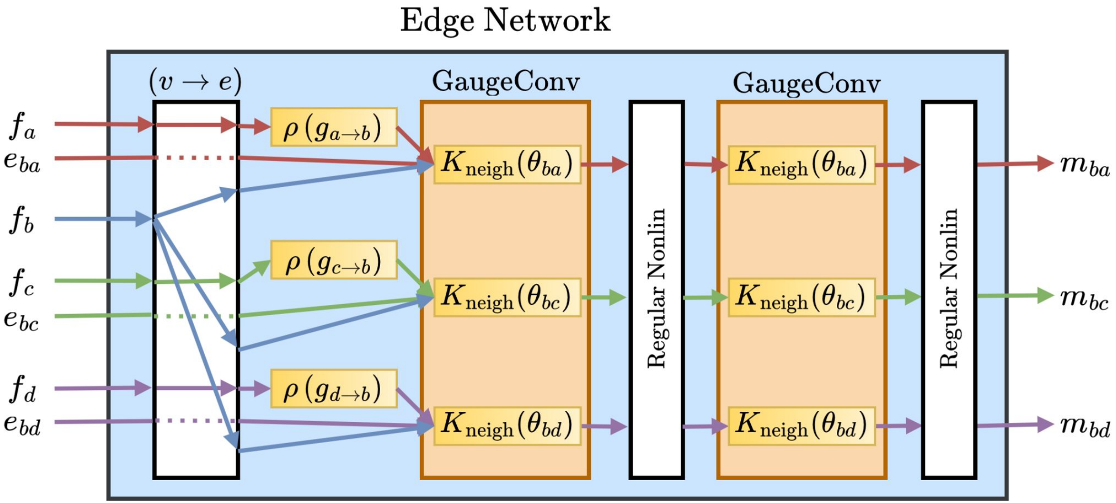


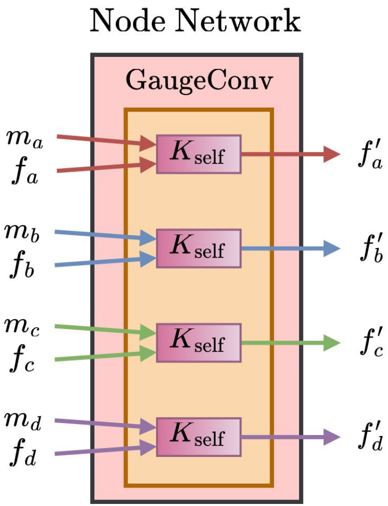


Jung Yeon Park, Lawson L.S. Wong, and Robin Walters. "**Modeling Dynamics over Meshes with Gauge Equivariant Nonlinear Message Passing" NeurIPS 2023**.

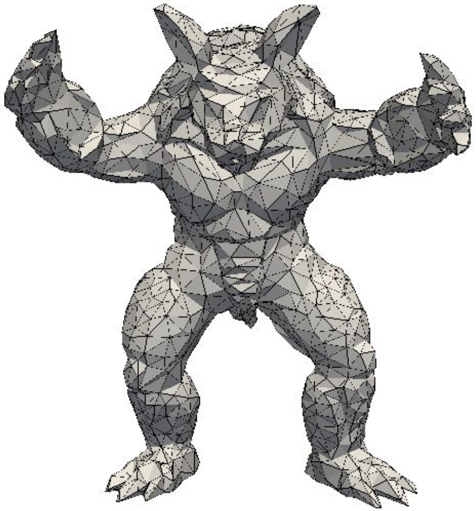
Architecture

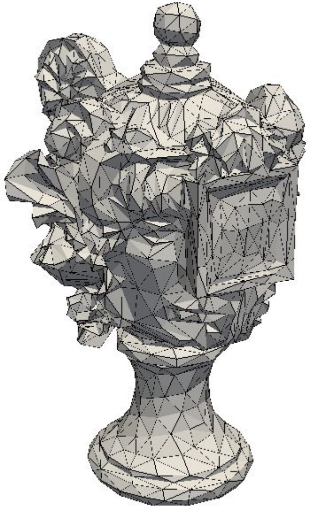


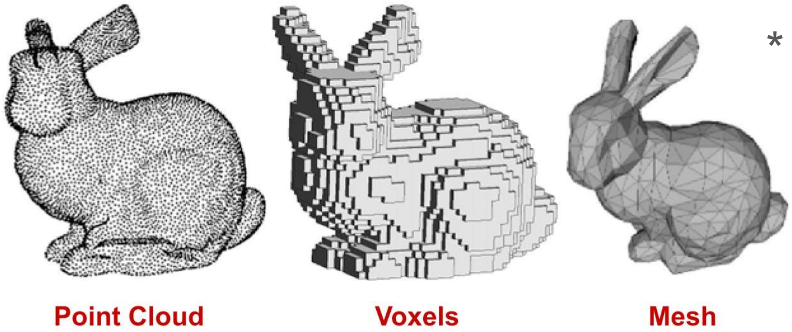




Mesh Representation



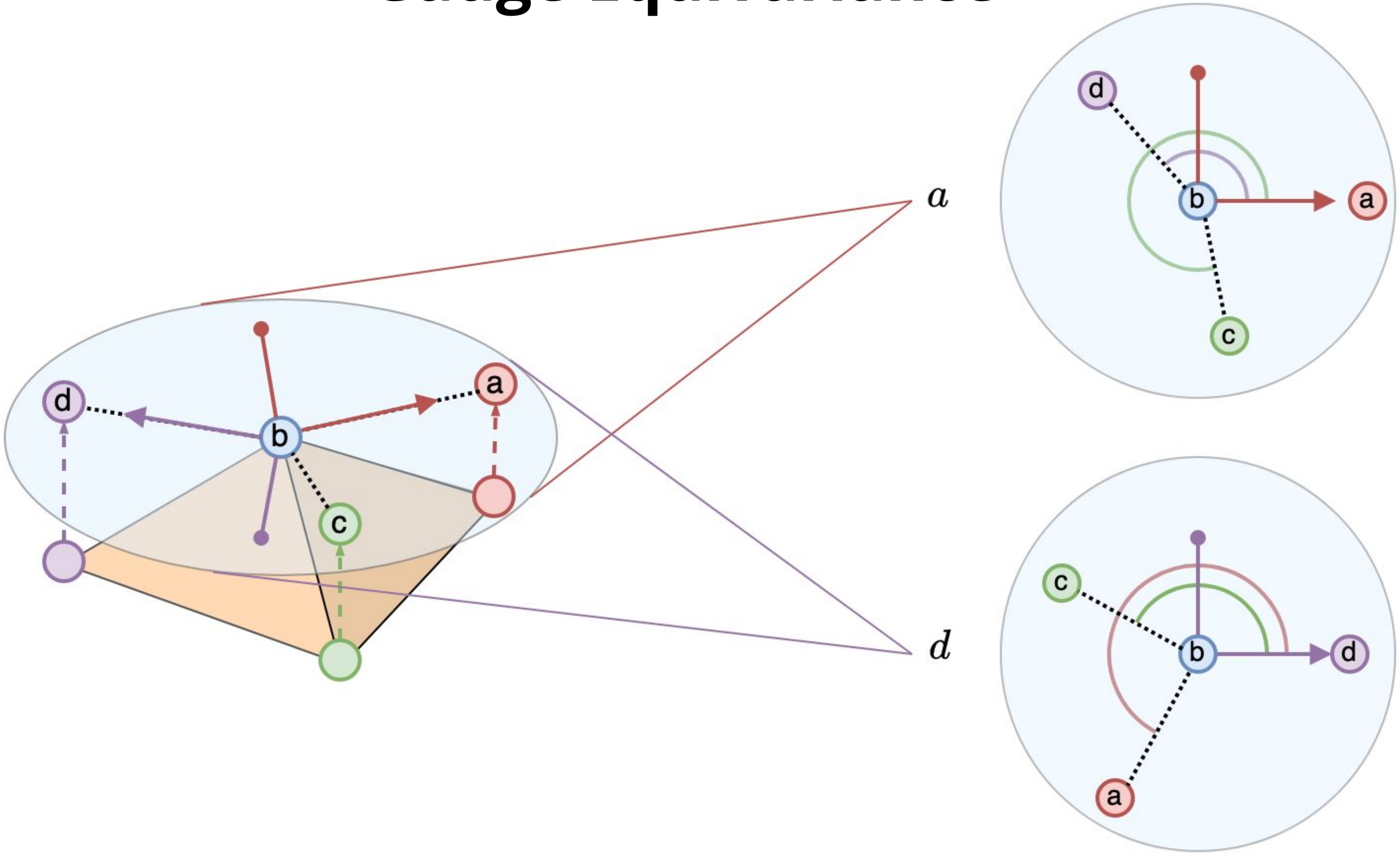




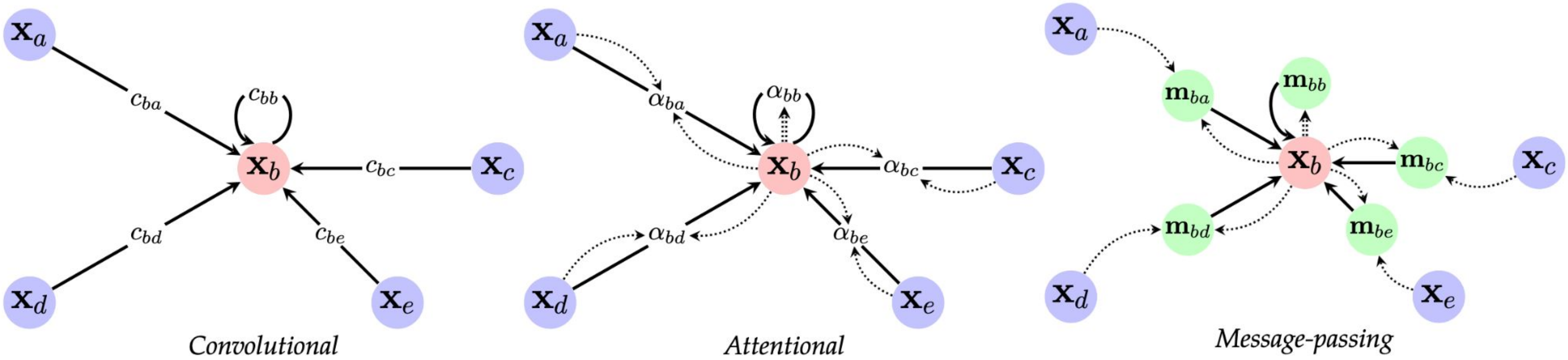
*

*https://towardsdatascience.com/shape-reconstruction-with-onets-1c1afe89c50

Gauge Equivariance

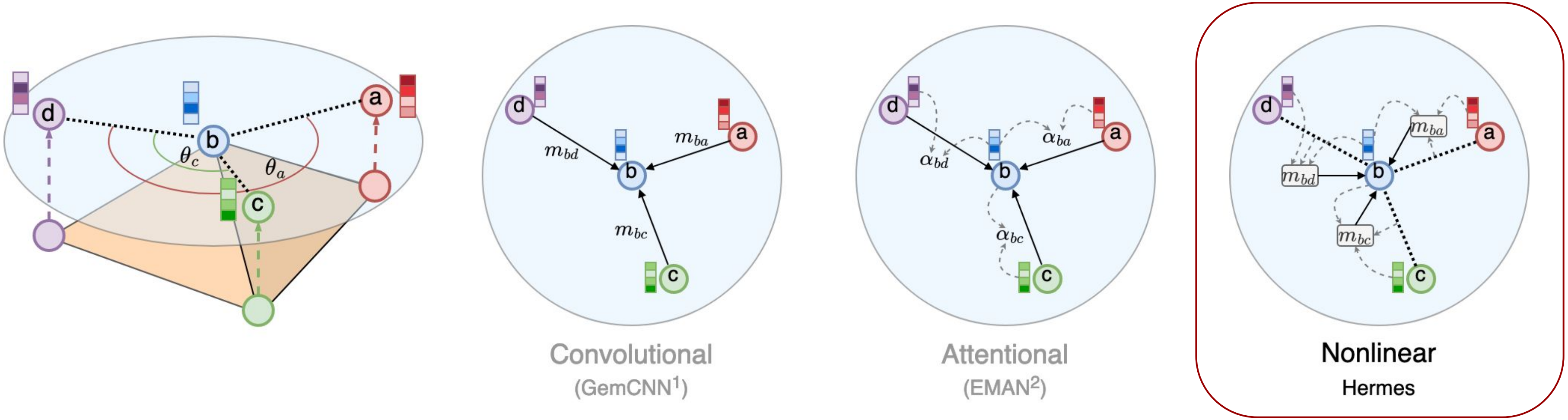


GNNs: Message Passing Flavors



Bronstein, M. M., Bruna, J., Cohen, T., & Veličković, P. (2021). Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*.

Message Passing Flavors



Combine gauge equivariance and nonlinear message passing

Better performance for modeling complex dynamics such as surface PDEs

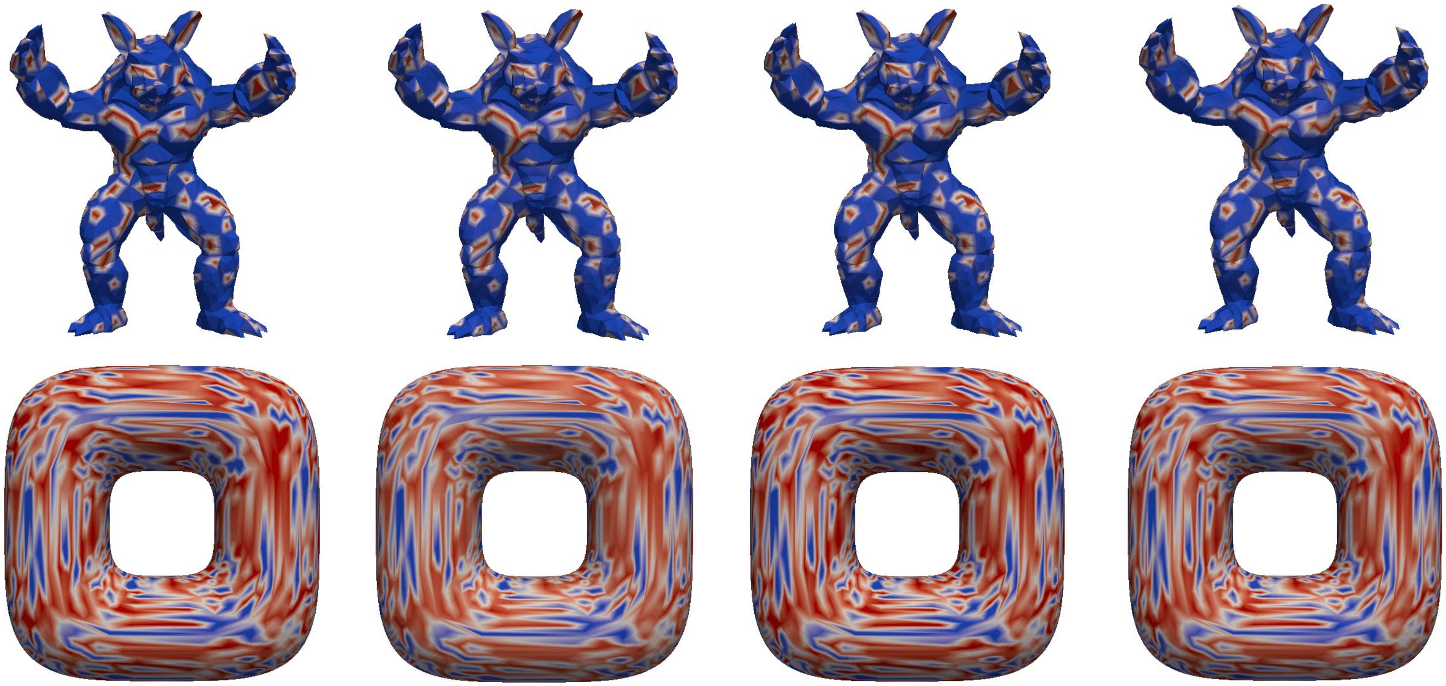
1. De Haan, P., Weiler, M., Cohen, T., & Welling, M. (2020, October). Gauge Equivariant Mesh CNNs: Anisotropic convolutions on geometric graphs. In *International Conference on Learning Representations*.
2. Basu, S., Gallego-Posada, J., Viganò, F., Rowbottom, J., & Cohen, T. (2022). Equivariant Mesh Attention Networks. *Transactions on Machine Learning Research*.

Qualitative Results

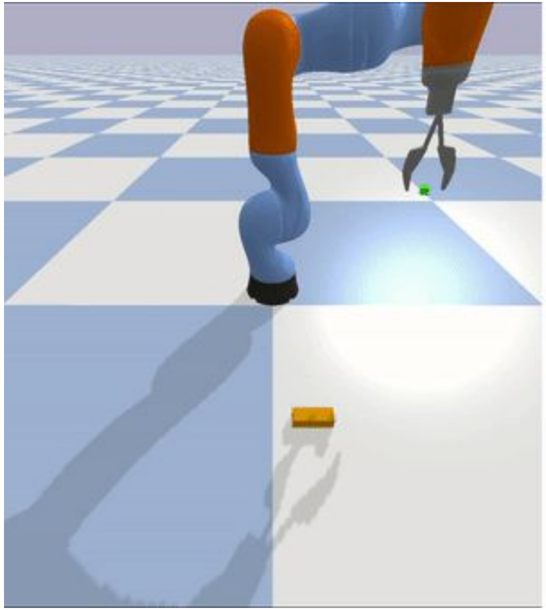
GemCNN | EMAN | Hermes | Ground Truth

Heat

Cahn-Hilliard



Equivariant Neural Networks in Robot Learning



1. Dian Wang, Robin Walters, Xupeng Zhu, and Robert Platt. "Equivariant Q-Learning in Spatial Action Spaces." *Conference on Robotics Learning (CoRL)*, 2021.
2. Dian Wang, Robin Walters, Robert Platt, "SO(2)-Equivariant Reinforcement Learning." *ICLR*, 2022.
3. Xupeng Zhu, Dian Wang, Guanang Su, Ondrej Biza, Robin Walters, Robert Platt. "On Robot Grasp Learning Using Equivariant Models." *Autonomous Robots. 2023.*
4. Xupeng Zhu, Dian Wang, Ondrej Biza, Guanang Su, Robin Walters, Robert Platt, "Sample Efficient Grasp Learning Using Equivariant Models," *RSS 2022.*
5. Haojie Huang, Dian Wang, Xupeng Zhu, Robin Walters, Robert Platt, "Edge Grasp Network: Graph-Based SE(3)-invariant Approach to Grasp Detection, *ICRA, 2023.*
6. Dian Wang, Mingxi Jia, Xupeng Zhu, Robin Walters, Robert Platt, "On-Robot Learning With Equivariant Models," *CoRL, 2022.*

Formalizing **Learning**: Markov Decision Process (MDP)

A Markov decision process is a 4-tuple $(S, A, T, \mathcal{R})$

- S: a set of all possible states
- A: a set of all possible actions

Formalizing **Learning**: Markov Decision Process (MDP)

A Markov decision process is a 4-tuple $(S, A, T, \mathcal{R})$

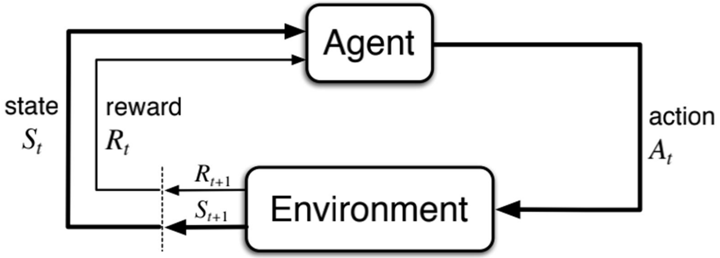
- S: a set of all possible states
- A: a set of all possible actions
- $T(s, a, s') = \Pr(s'|s, a)$: probability of reaching state s' after taking action a at state s
- $\mathcal{R}(s, a)$: the reward of taking action a at state s

Formalizing Learning: Markov Decision Process (MDP)

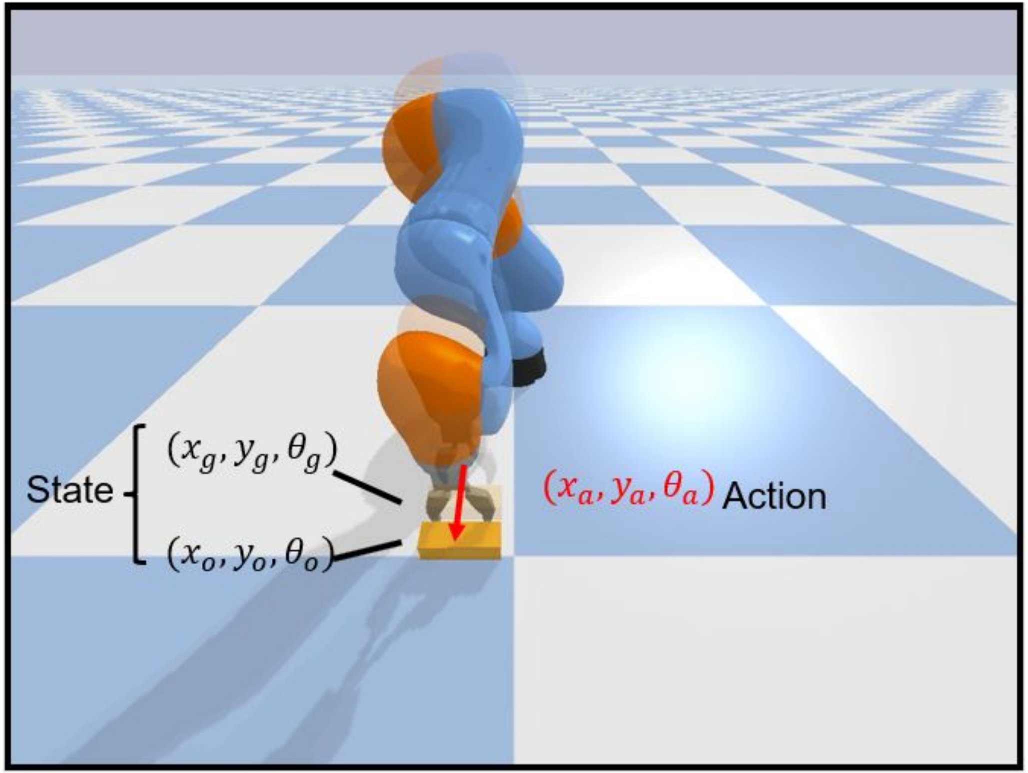
A Markov decision process is a 4-tuple $(S, A, T, \mathcal{R})$

- S: a set of all possible states
- A: a set of all possible actions
- $T(s, a, s') = \Pr(s'|s, a)$: probability of reaching state s' after taking action a at state s
- $\mathcal{R}(s, a)$: the reward of taking action a at state s

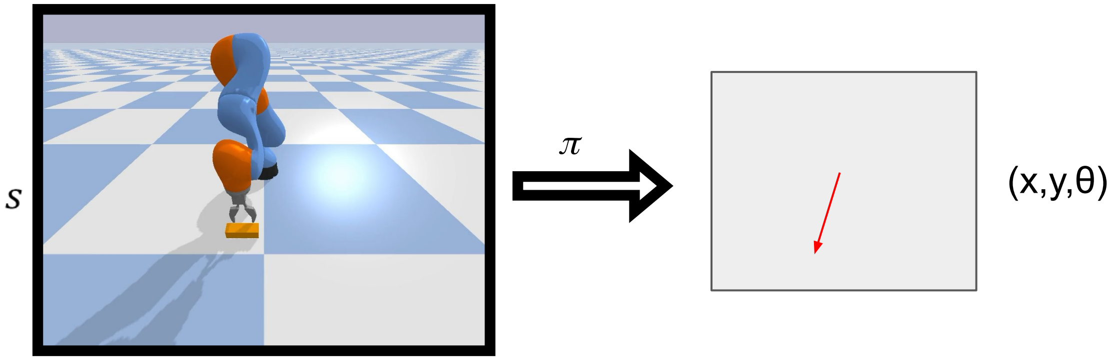
We can model robotic task as an MDP and solve it using RL



Example: Robot Manipulation



Policy Learning



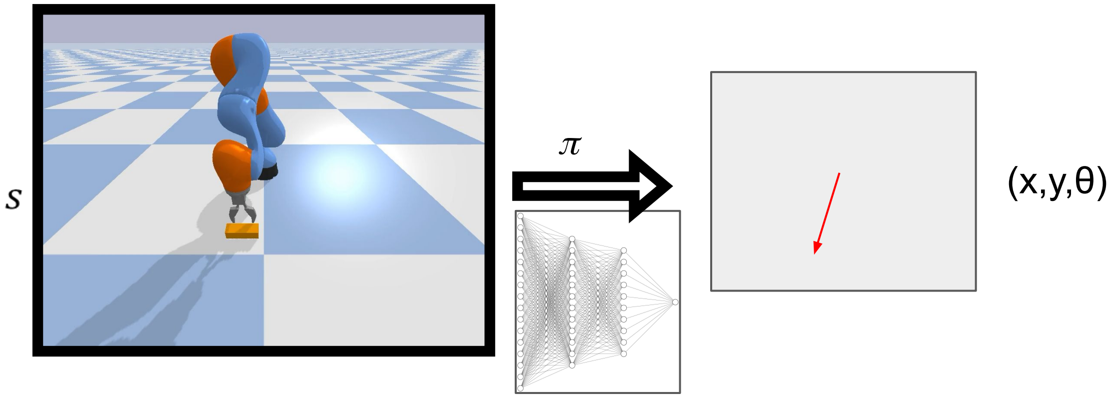
Learn Control Policy: $a = \pi(s)$

Control signal

Image, lidar,
force, tactile, etc.

- Reinforcement Learning: Algorithms to learn π by trial and error

Deep Policy Learning



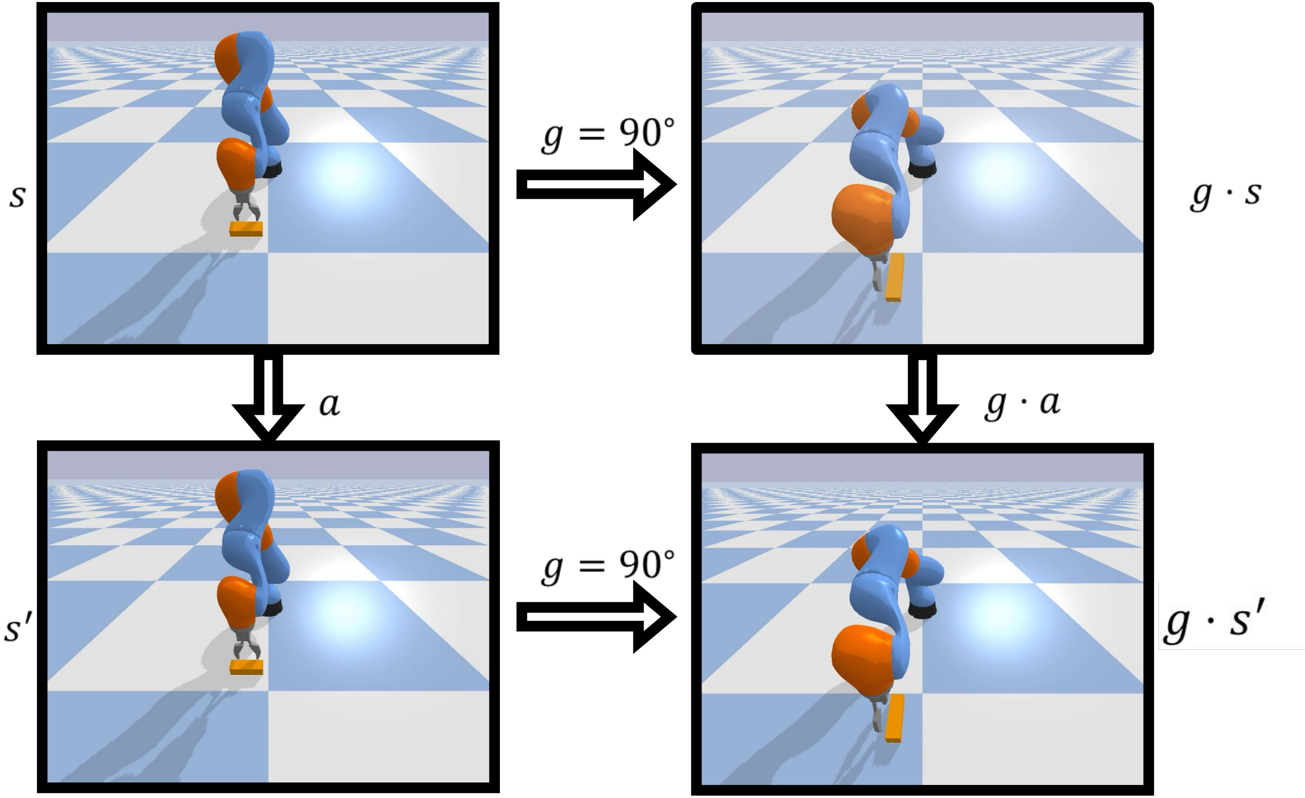
Learn Control Policy: $a = \pi(s)$

Control signal

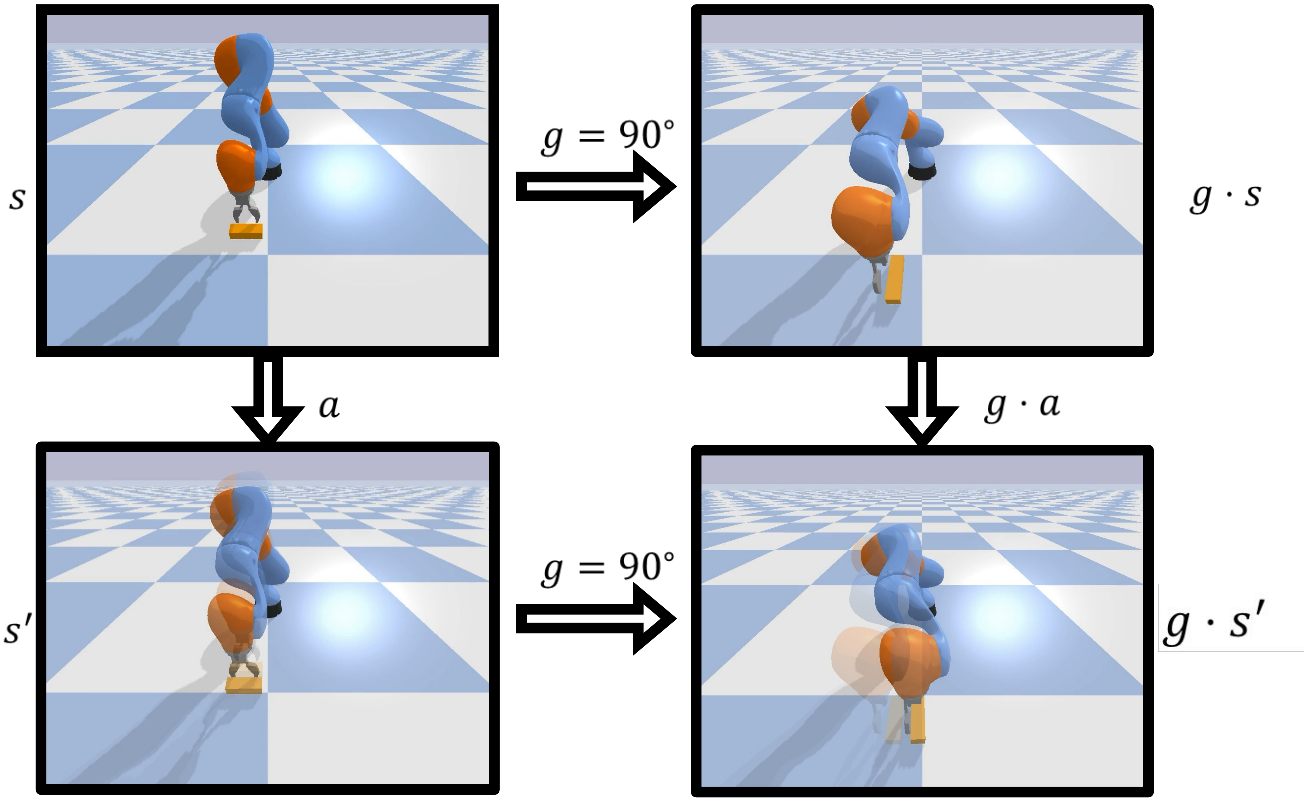
Image, lidar, force, tactile, etc.

- Reinforcement Learning: Algorithms to learn π by trial and error
- Deep Reinforcement Learning: π (or related function) is given by a neural network

SO(2) Equivariance in Manipulation



SO(2) Equivariance in Manipulation



Invariant MDPs have invariant optimal Q functions and equivariant optimal policies

Definition 4.1 (G-invariant MDP). *A G-invariant MDP $\mathcal{M}_G = (S, A, T, R, G)$ is an MDP $\mathcal{M} = (S, A, T, R)$ that satisfies the following conditions:*

1. Reward Invariance: *The reward function is invariant to the action of the group element $g \in G$, $\boxed{R(s, a) = R(gs, ga)}$.*

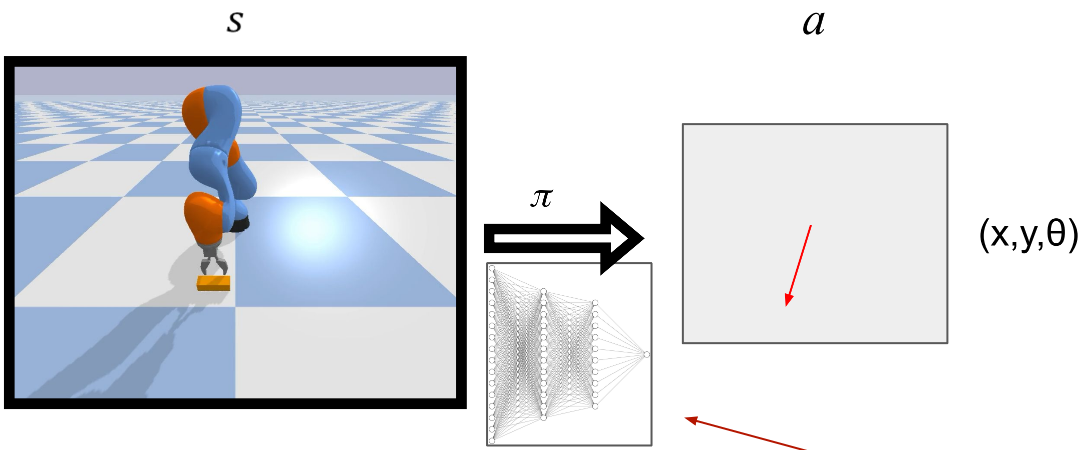
2. Transition Invariance: *The transition function is invariant to the action of the group element $g \in G$, $\boxed{T(s, a, s') = T(gs, ga, gs')}$.*

Proposition 4.1. *Let $\mathcal{M}_G$ be a group-invariant MDP. Then its optimal Q-function is group invariant, $Q^*(s, a) = Q^*(gs, ga)$, and its optimal policy is group-equivariant, $\pi^*(gs) = g\pi^*(s)$, for any $g \in G$.*

$$\boxed{\pi^*(g \cdot s) = g \cdot \pi^*(s)}$$

$$\boxed{Q^*(g \cdot s, g \cdot a) = Q^*(s, a)}$$

Equivariant Policy Learning

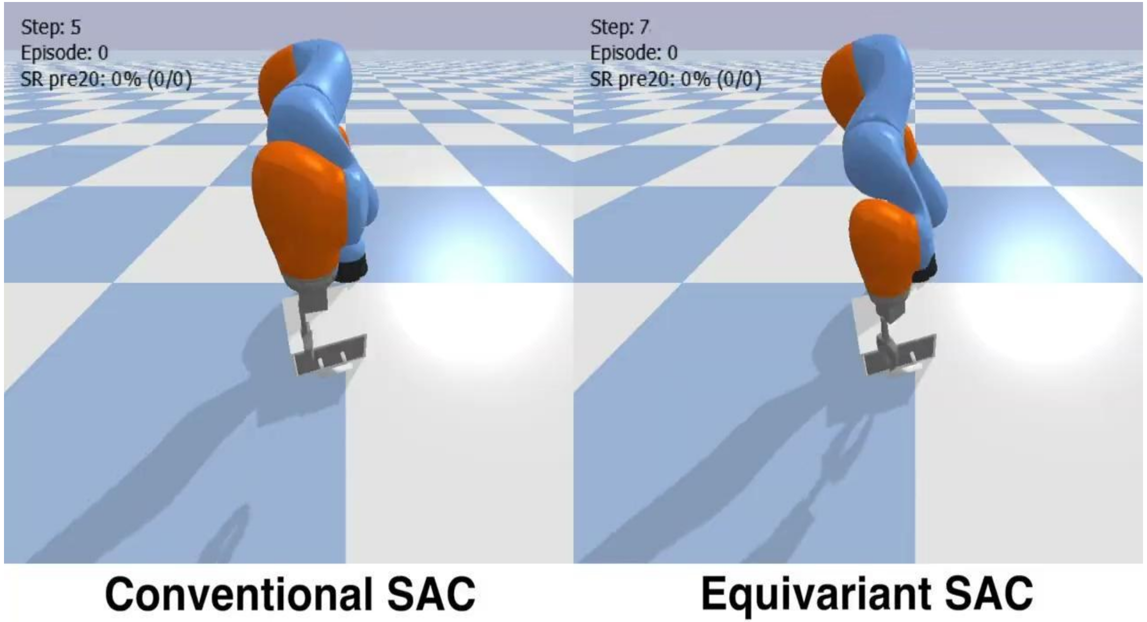


Learn Control Policy: $a = \pi(s)$

Control signal

Image, lidar,
force, tactile, etc.

Comparison - Efficiency



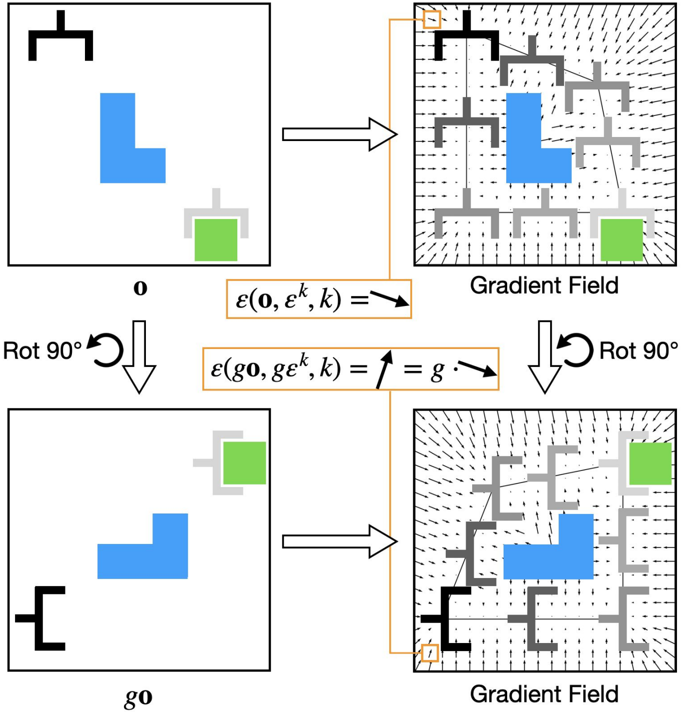
Dian Wang, Robin Walters, Robert Platt. “SO(2)-Equivariant Reinforcement Learning”. *ICLR* 2022

Grasp Learning in SE(2)

- Learn with **600** grasping attempts (**1.5 hour**)!

1. Xupeng Zhu, Dian Wang, Guanang Su, Ondrej Biza, Robin Walters, Robert Platt. “On Robot Grasp Learning Using Equivariant Models.” *Autonomous Robots. 2023.*
2. Xupeng Zhu, Dian Wang, Ondrej Biza, Guanang Su, Robin Walters, Robert Platt, “Sample Efficient Grasp Learning Using Equivariant Models,” *RSS 2022.*

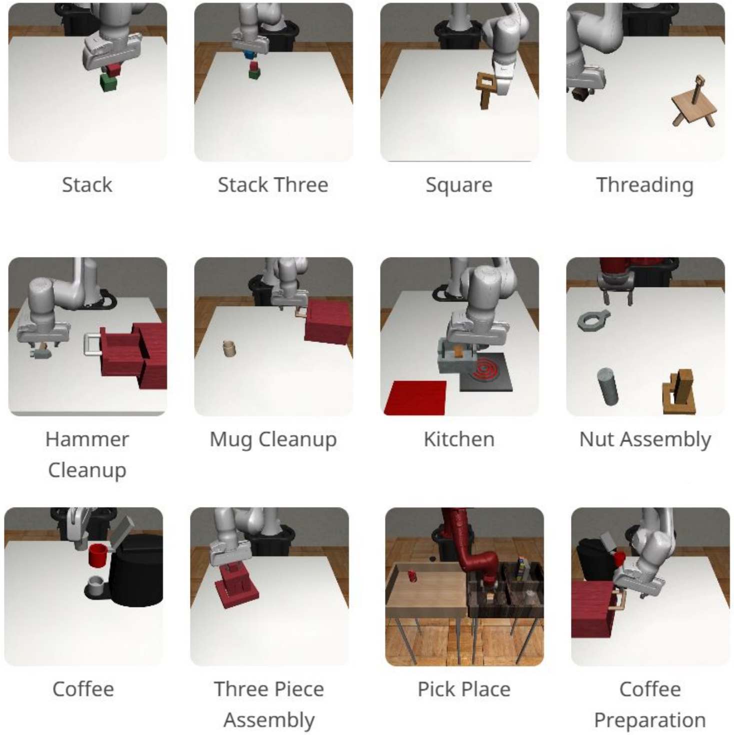
Equivariant Diffusion Policy



- Model policy as diffusion process over actions
- Equivariant conditioning on state

Dian Wang, Stephen Hart, David Surovik, Tarik Kelestemur, Haojie Huang, Haibo Zhao, Jiuguang Wang, Robin Walters, Robert Platt. Equivariant Diffusion Policy

Simulation Results on MimicGen



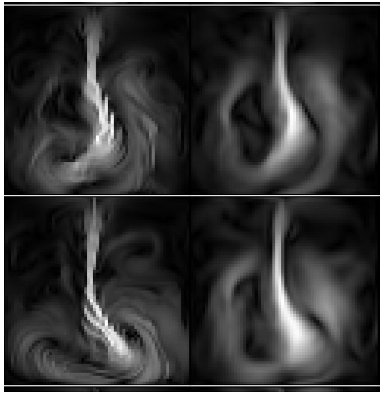
		Average over 12 Environments		
Method	Ctrl	100	200	1000
EquiDiff (PC)		76.5 (+34.5)	81.6 (+23.8)	82.3 (+10.9)
EquiDiff (Vo)		63.9 (+21.9)	72.6 (+14.8)	77.9 (+6.5)
EquiDiff (Im)		53.7 (+11.7)	68.5 (+10.7)	79.7 (+8.3)
DiffPo-C	Abs	42.0	57.8	71.4
DiffPo-T		29.0	43.0	64.9
DP3		23.9	35.1	56.8
ACT		21.3	38.2	63.3

Dramatically better performance than Diffusion Policy with 100 or 200 demos.

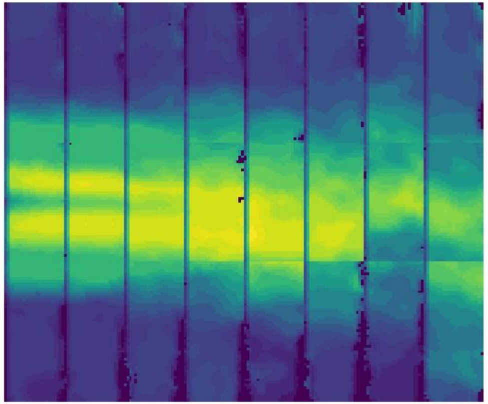
Equivariant Diffusion Policy

Can learn long horizon tasks from few demonstrations (58).

What do we do about approximate symmetry?

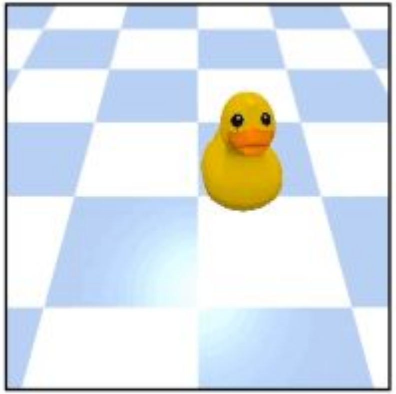


Unknown external forces and boundary conditions



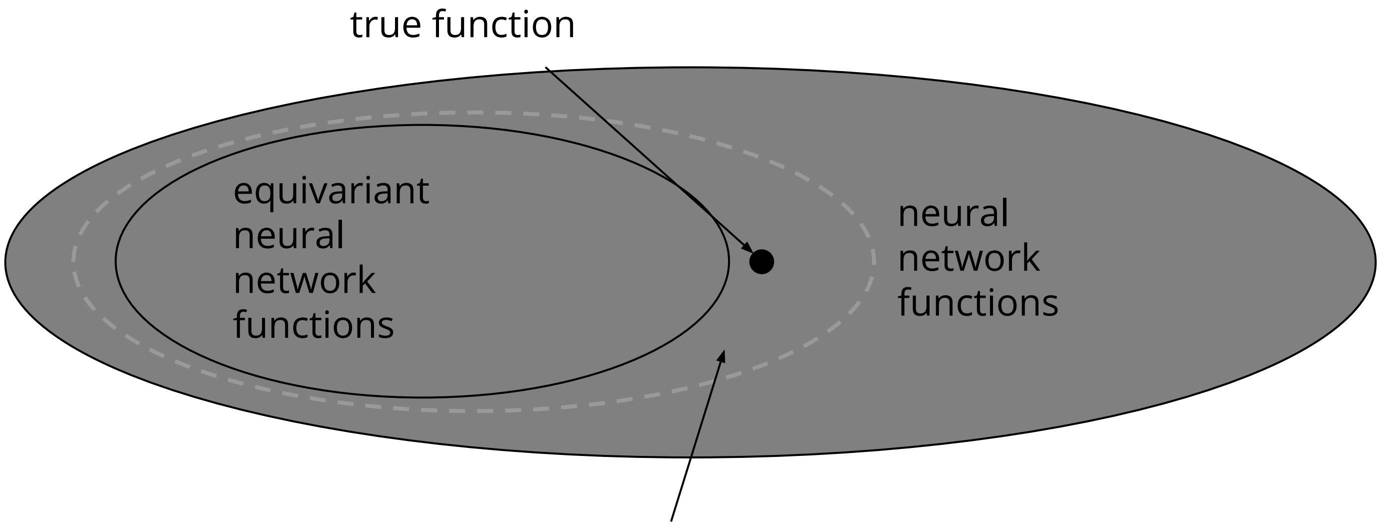
NASA LARC - Supersonic Jet flow

Limited or noisy observations



Observations not aligned with symmetry or symmetry breaking background

Option 2: Relax Equivariance Constraints



Relaxed Equivariant Neural Network Functions

Robin Walters*, **Rui Wang***,, and **Rose Yu**. "Approximately Equivariant Networks for Imperfectly Symmetric Dynamics." International Conference on Machine Learning (ICML), 2022.

What are the assumptions of an equivariant model?

1. Equivariant Task Function

$$f(gx) = gf(x)$$

2. Invariant Input Distribution

$$x \sim \mathcal{D},\ p_{\mathcal{D}}(x) = p_{\mathcal{D}}(gx)$$

Now we consider relaxing the 1st assumption

What are the assumptions of an equivariant model?

1. Equivariant Task Function

$$\cancel{f(gx) = gf(x)} \quad f(gx) \approx gf(x)$$

2. Invariant Input Distribution

$$x \sim \mathcal{D},\ p_{\mathcal{D}}(x) = p_{\mathcal{D}}(gx)$$

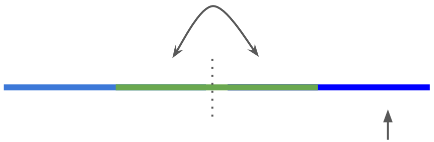
Now we consider relaxing the 1st assumption

What are the assumptions of an equivariant model?

Equivariant Task Function

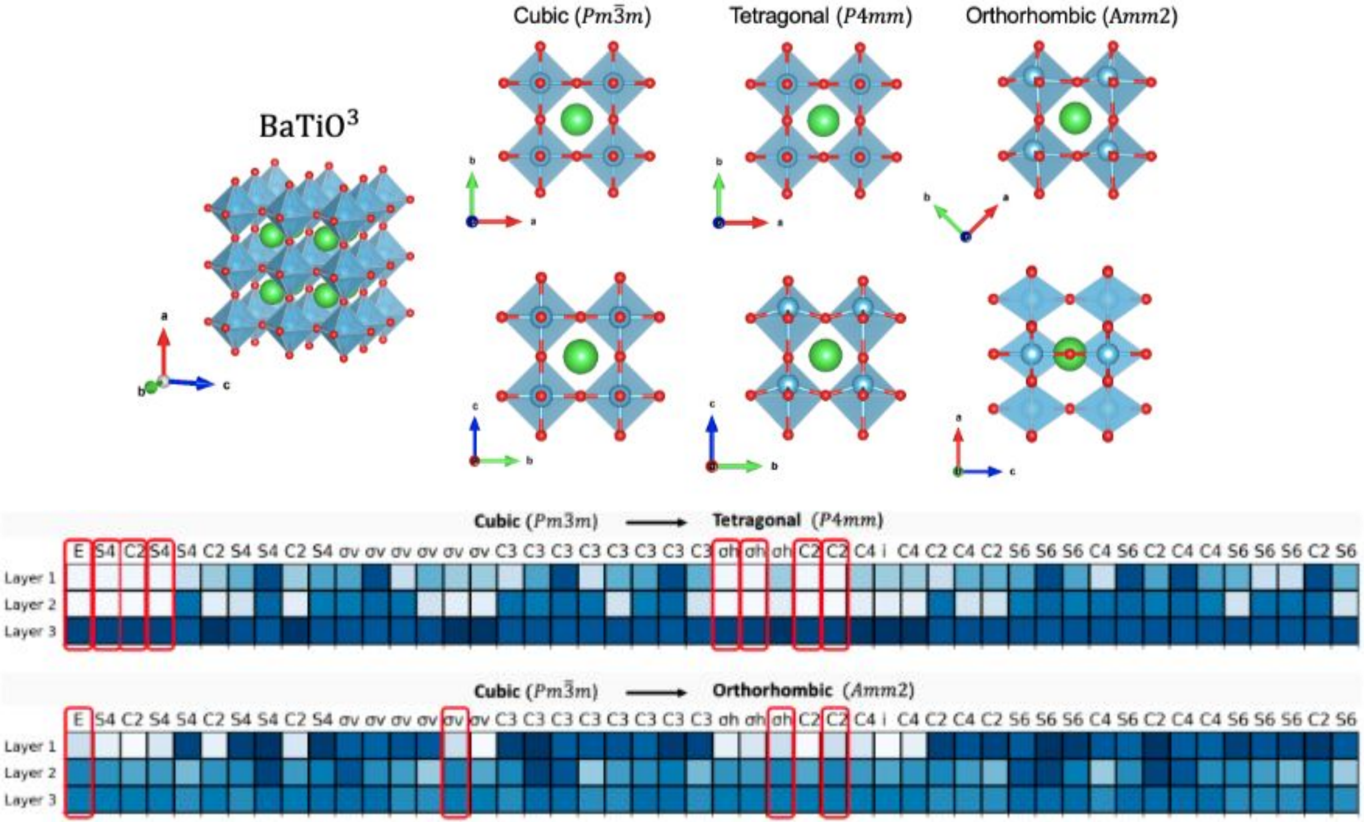
$$\cancel{f(gx) = gf(x)} \quad f(gx) \approx gf(x)$$

G**-approx-equiv**: $|f(\rho_{in}(g)x) - \rho_{out}(g)f(x)| < \varepsilon$



Incorrect
(in output,ε approximate)

Relaxed Weights Describe Symmetry Subgroup



Discovering Symmetry Breaking in Physical Systems with Relaxed Group Convolution *Rui Wang, Elyssa Hofgard, Han Gao, Robin Walters, Tess E.Smidt.*
International Conference on Machine Learning (ICML) 2024

Group Convolution

Relaxing weight-sharing constraints in Equivariant Networks by

introducing group element dependent parameters

- **Group Convolution (Cross-Correlation)**

$$[f \star_G \Psi](g) = \sum_{h \in G} f(h)\Psi(g^{-1}h)$$

Relaxed Group Convolution

Relaxing weight-sharing constraints in Equivariant Networks by

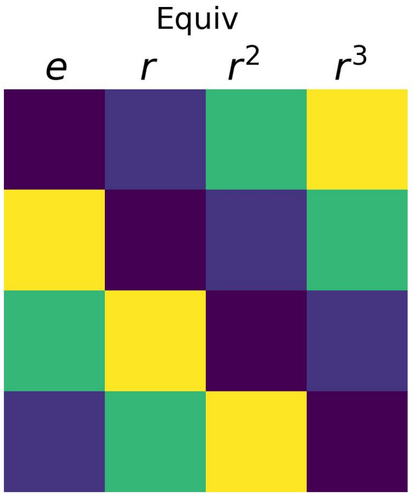
introducing group element dependent parameters

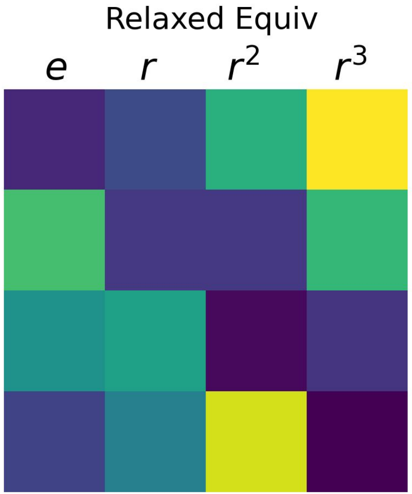
❖ **Relaxed Group Convolution (Cross-Correlation)**

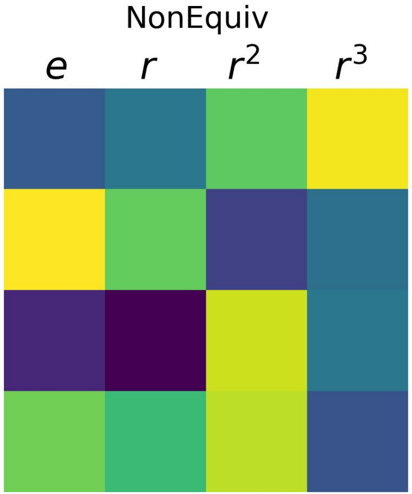
$$[f \star_G \Psi](g) = \sum_{h \in G} f(h)\Psi(g^{-1}h) \longrightarrow [f \tilde{\star}_G \Psi](g) = \sum_{h \in G} \sum_{l=1}^{L} f(h) \boxed{w_l(h)} \Psi_l(g^{-1}h)$$

Relaxed Group Convolution

$$\begin{pmatrix} a & b & c & d \\ d & a & b & c \\ c & d & a & b \\ b & c & d & a \end{pmatrix} \qquad \begin{pmatrix} xa & yb & zc & wd \\ xd & ya & zb & wc \\ xc & yd & za & wb \\ xb & yc & zd & wa \end{pmatrix} \qquad \begin{pmatrix} a & b & c & d \\ e & f & g & h \\ i & j & k & l \\ m & n & o & p \end{pmatrix}$$







Relaxed Equivariant Neural Networks

- **Steerable Convolution** $\Phi(hx) = \rho_{\text{out}}(h)\Phi(x)\rho_{\text{in}}(h^{-1}),\ \forall h \in H$

$$\sum_{\mathbf{y}\in\mathbb{Z}^2} \sum_{l=1}^{L} (w_l \odot \Phi_l(\mathbf{y})) f_{\text{in}}(\mathbf{x}+\mathbf{y})$$

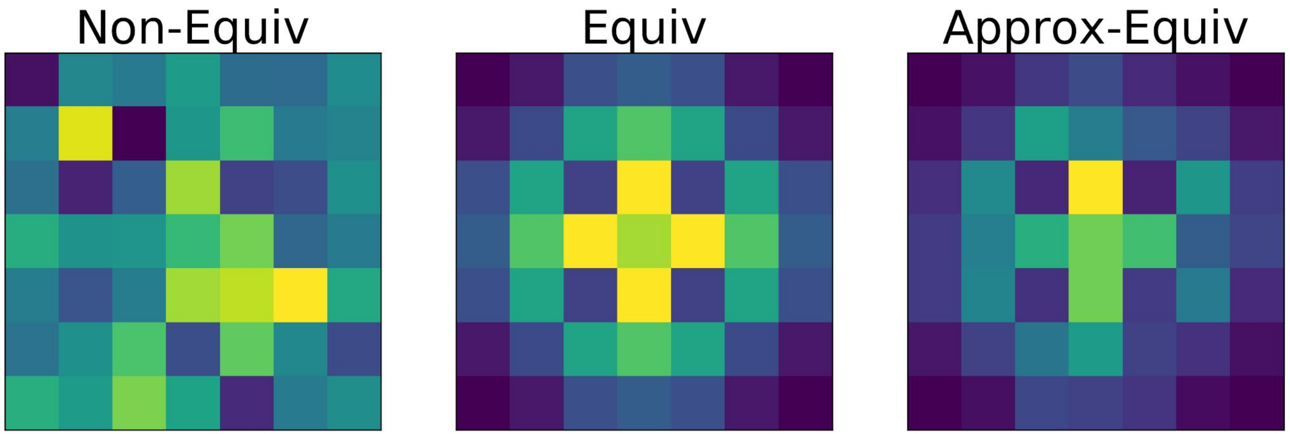
$$w \in \mathbb{R}^{c_{\text{out}} \times c_{\text{in}} \times L}$$

Relaxed Equivariant Neural Networks

- **Relaxed Steerable Convolution** $\Phi(hx) = \rho_{\text{out}}(h)\Phi(x)\rho_{\text{in}}(h^{-1}),\ \forall h \in H$

$$\sum_{\mathbf{y}\in\mathbb{Z}^2}\sum_{l=1}^{L}(w_l \odot \Phi_l(\mathbf{y}))f_{\text{in}}(\mathbf{x}+\mathbf{y}) \longrightarrow \sum_{\mathbf{y}\in\mathbb{Z}^2}\sum_{l=1}^{L}(\boxed{w_l(\mathbf{y})} \odot \Phi_l(\mathbf{y}))f_{\text{in}}(\mathbf{x}+\mathbf{y})$$

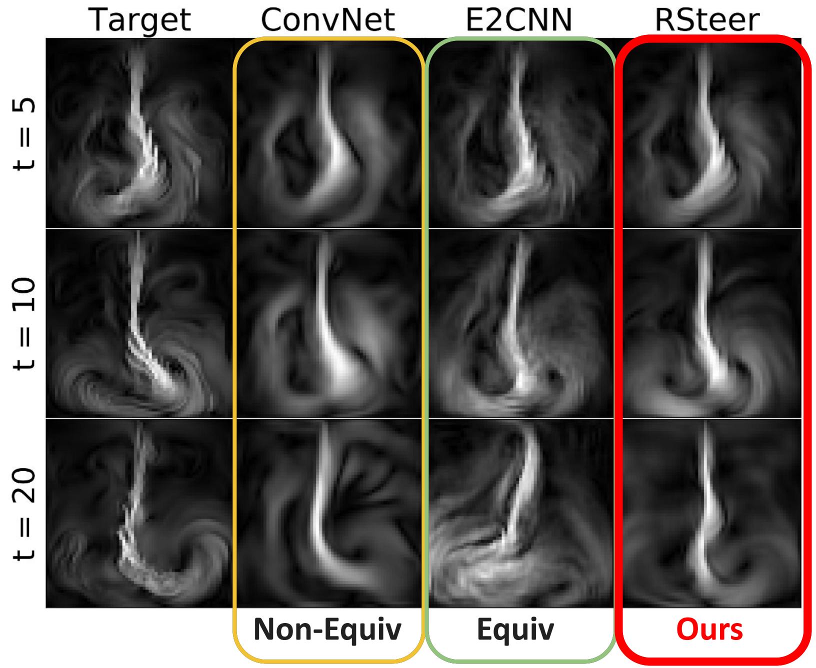
$$w \in \mathbb{R}^{c_{\text{out}}\times c_{\text{in}}\times L} \qquad w: \boxed{K^2} \to \mathbb{R}^{c_{\text{out}}\times c_{\text{in}}\times L}$$



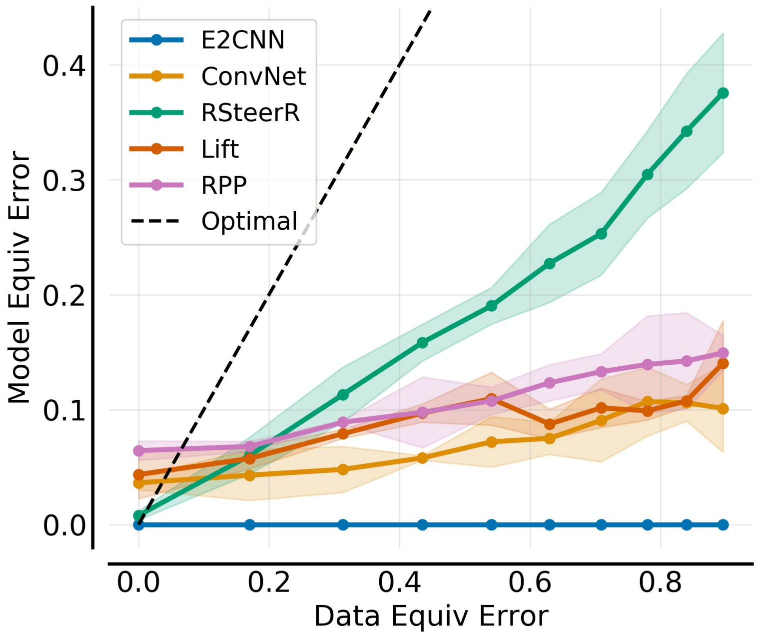
Robin Walters*, **Rui Wang***,, and **Rose Yu**. "Approximately Equivariant Networks for Imperfectly Symmetric Dynamics." International Conference on Machine Learning (ICML), 2022.

Improved Prediction on Fluid Dynamics

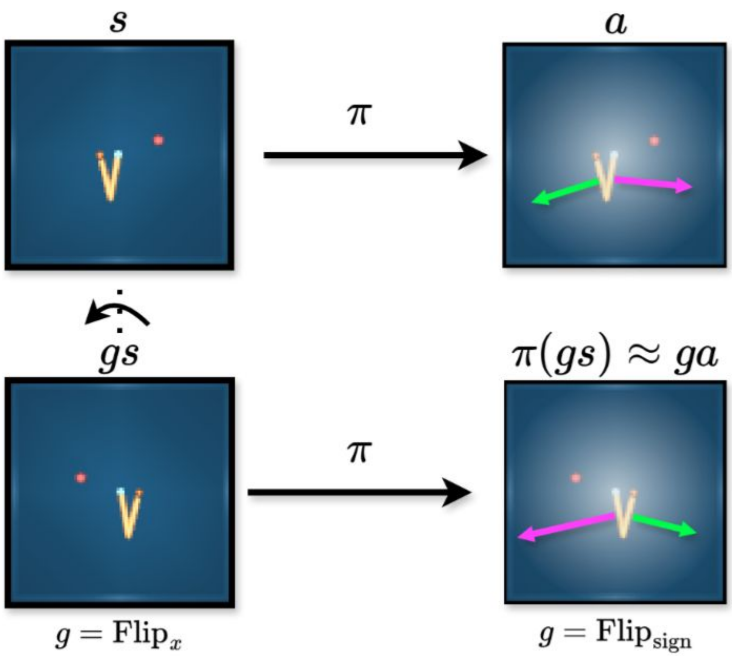
The buoyant force varies with the inflow positions to break the **rotation** symmetry



Learning Different Levels of Equivariance

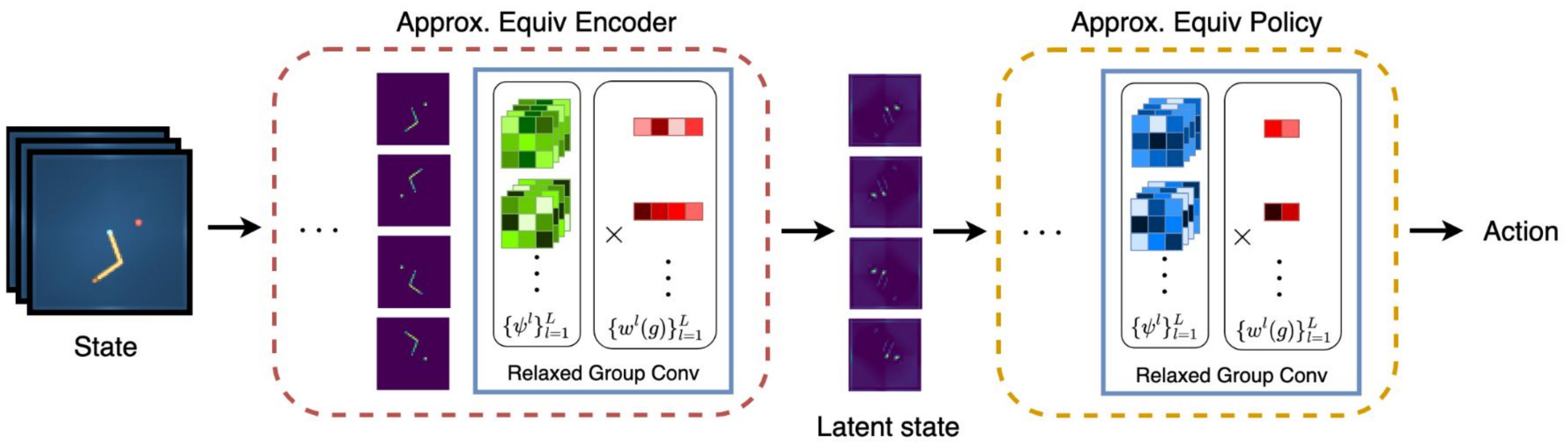


Relaxed Equivariance in Reinforcement Learning



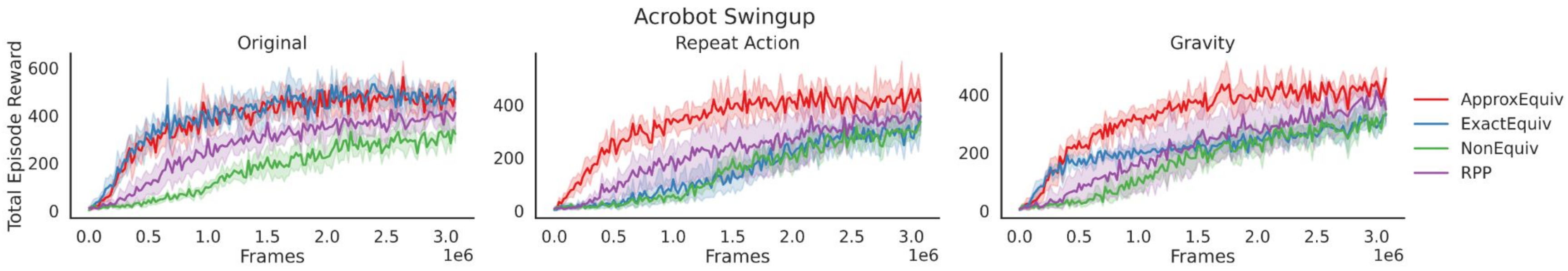
Due to wear the joint is more responsive to positive torque than negative, making the environment approximately symmetric to flips

Relaxed Equivariant Reinforcement Learning



Idea: Use relaxed equivariant models to learn the policy

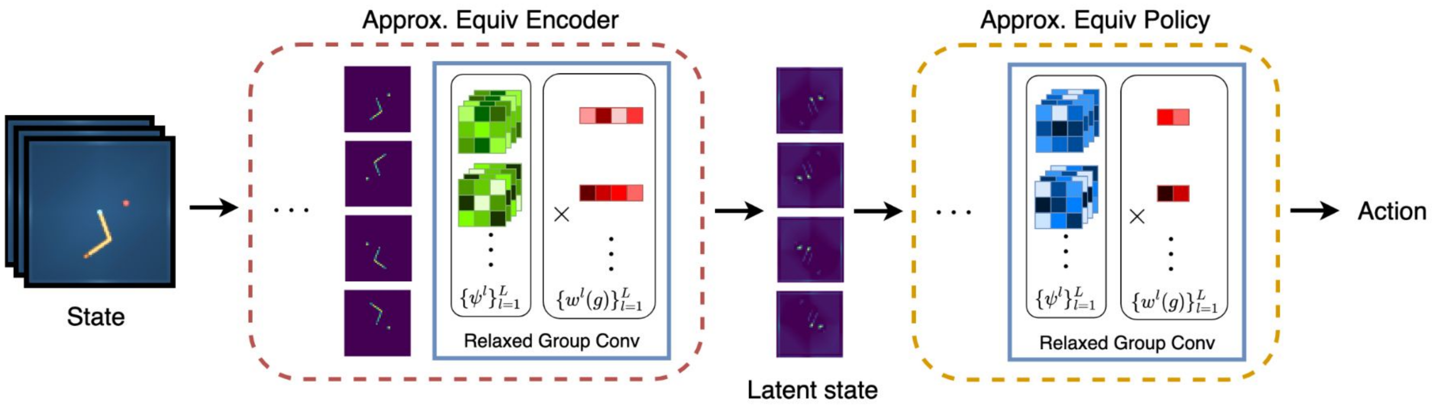
Relaxed Equivariant Reinforcement Learning

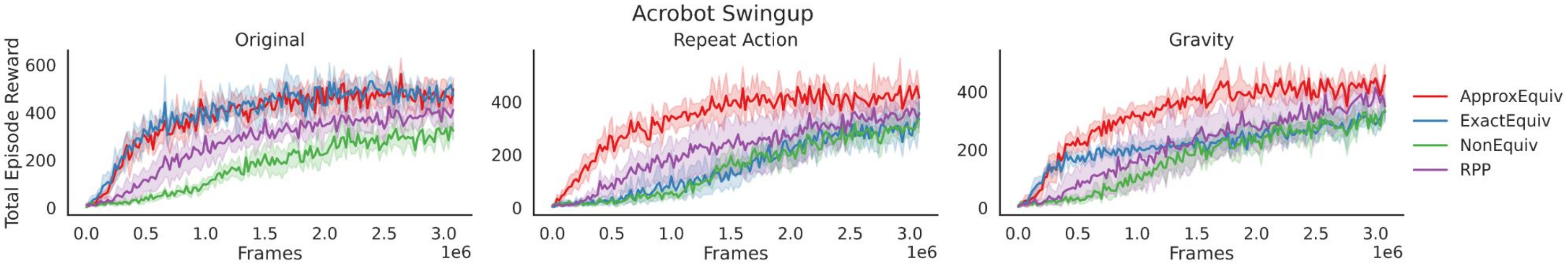


In domains with exact symmetry, the equivariant and relaxed equivariant model perform similarly, but in domains with symmetry breaking relaxed equivariant models outperform equivariant and unconstrained models.

Jung Yeon Park, Sujay Bhatt, Sihan Zeng, Lawson LS Wong, Alec Koppel, Sumitra Ganesh, Robin Walters, "Approximately Equivariance in Reinforcement Learning" AISTATS, 2025.

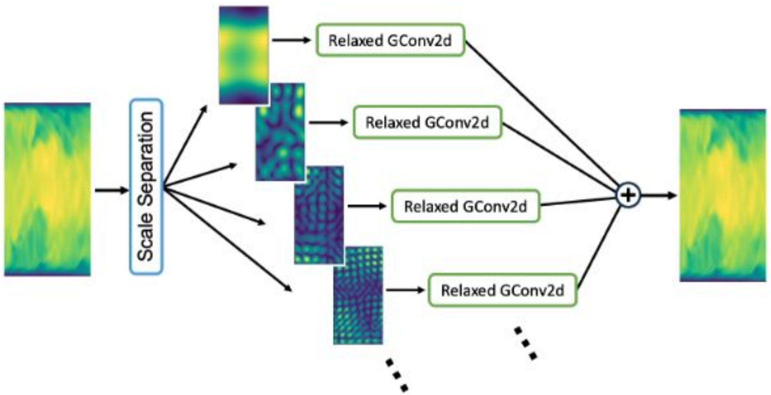
Relaxed Equivariant Reinforcement Learning

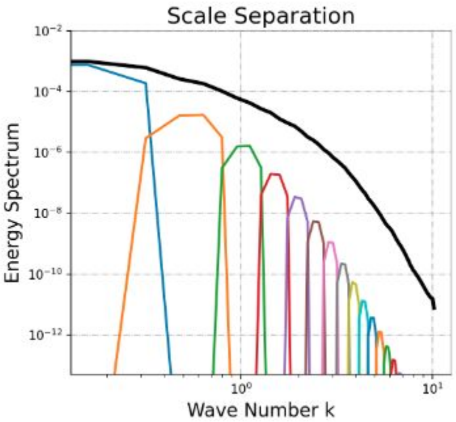


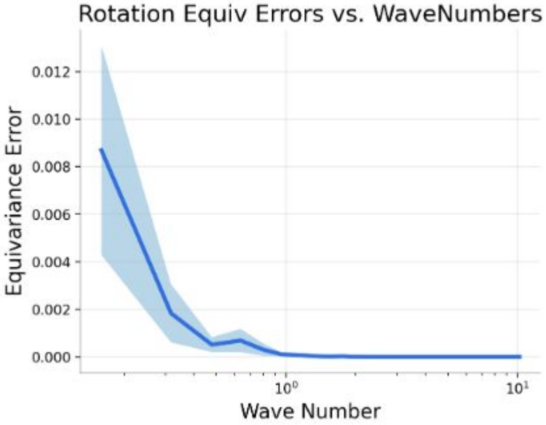


Jung Yeon Park, Sujay Bhatt, Sihan Zeng, Lawson LS Wong, Alec Koppel, Sumitra Ganesh, Robin Walters, "Approximately Equivariance in Reinforcement Learning" AISTATS, 2025.

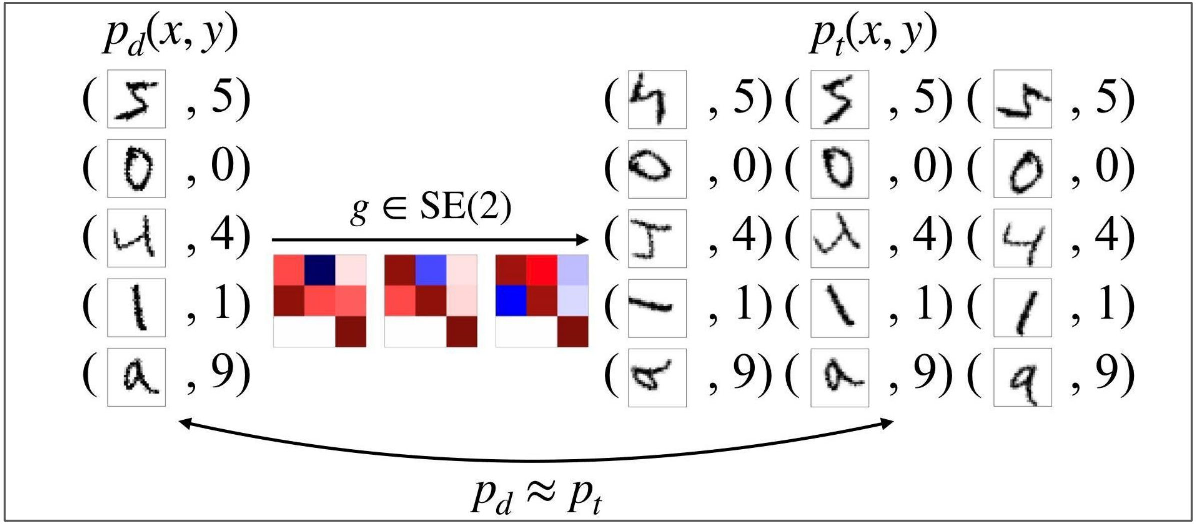
Relaxed Equivariant Models for Fluids





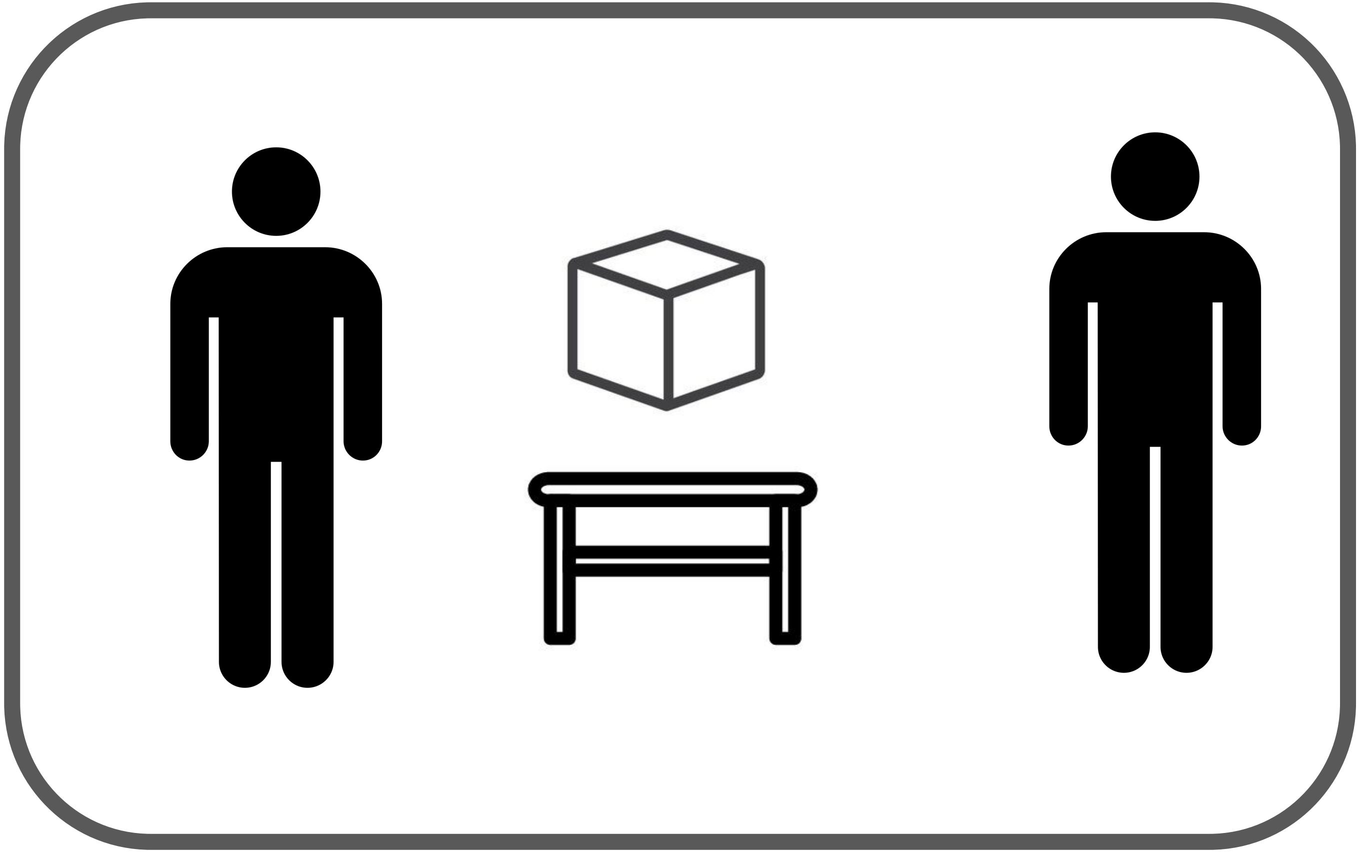


III. Symmetry Discovery for Domains with Unknown Symmetry

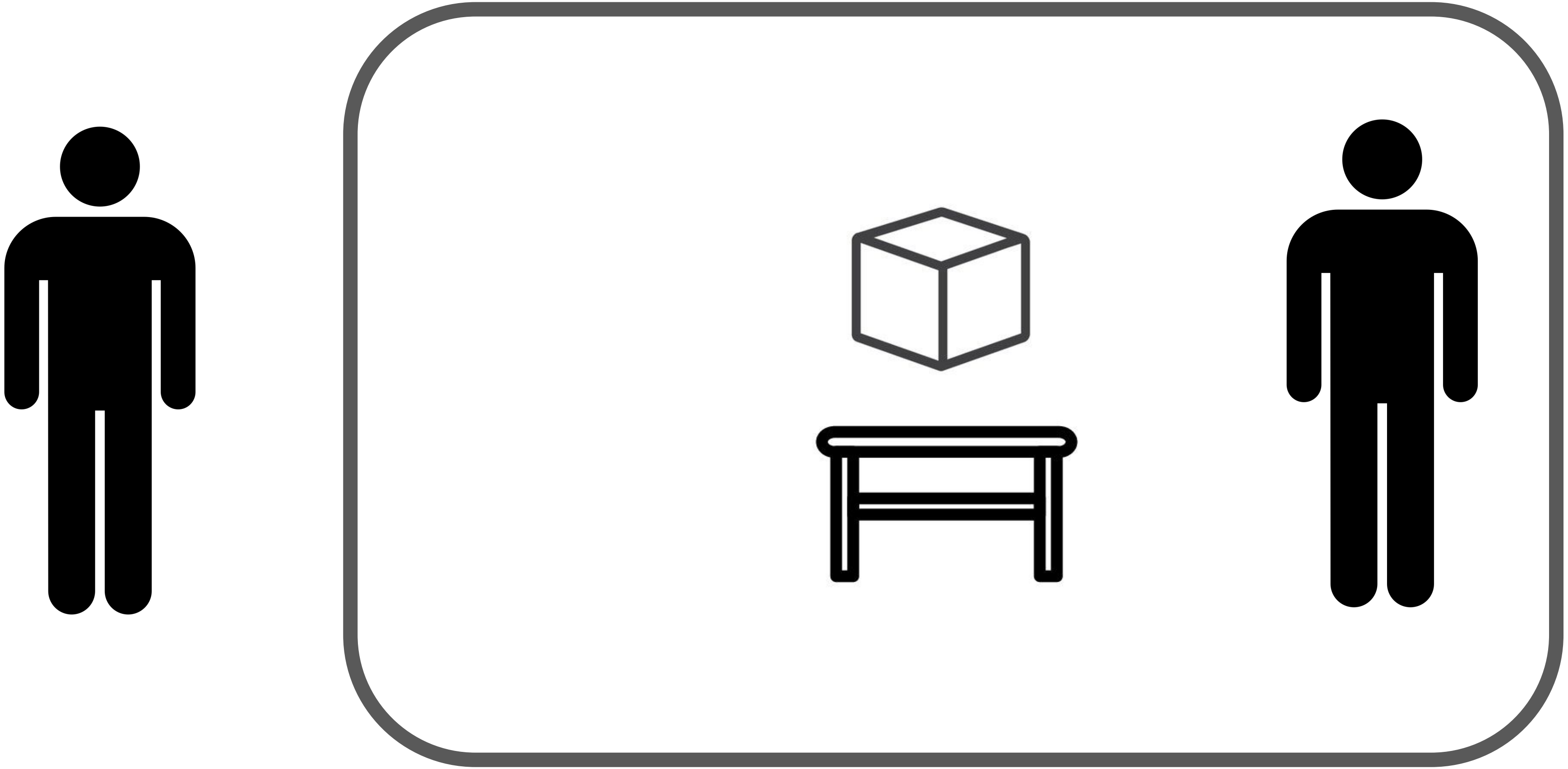


Jianke Yang, Robin Walters*, Nima Dehmamy*, and Rose Yu. "Generative Adversarial Symmetry Discovery" International Conference on Machine Learning (ICML), 2023.

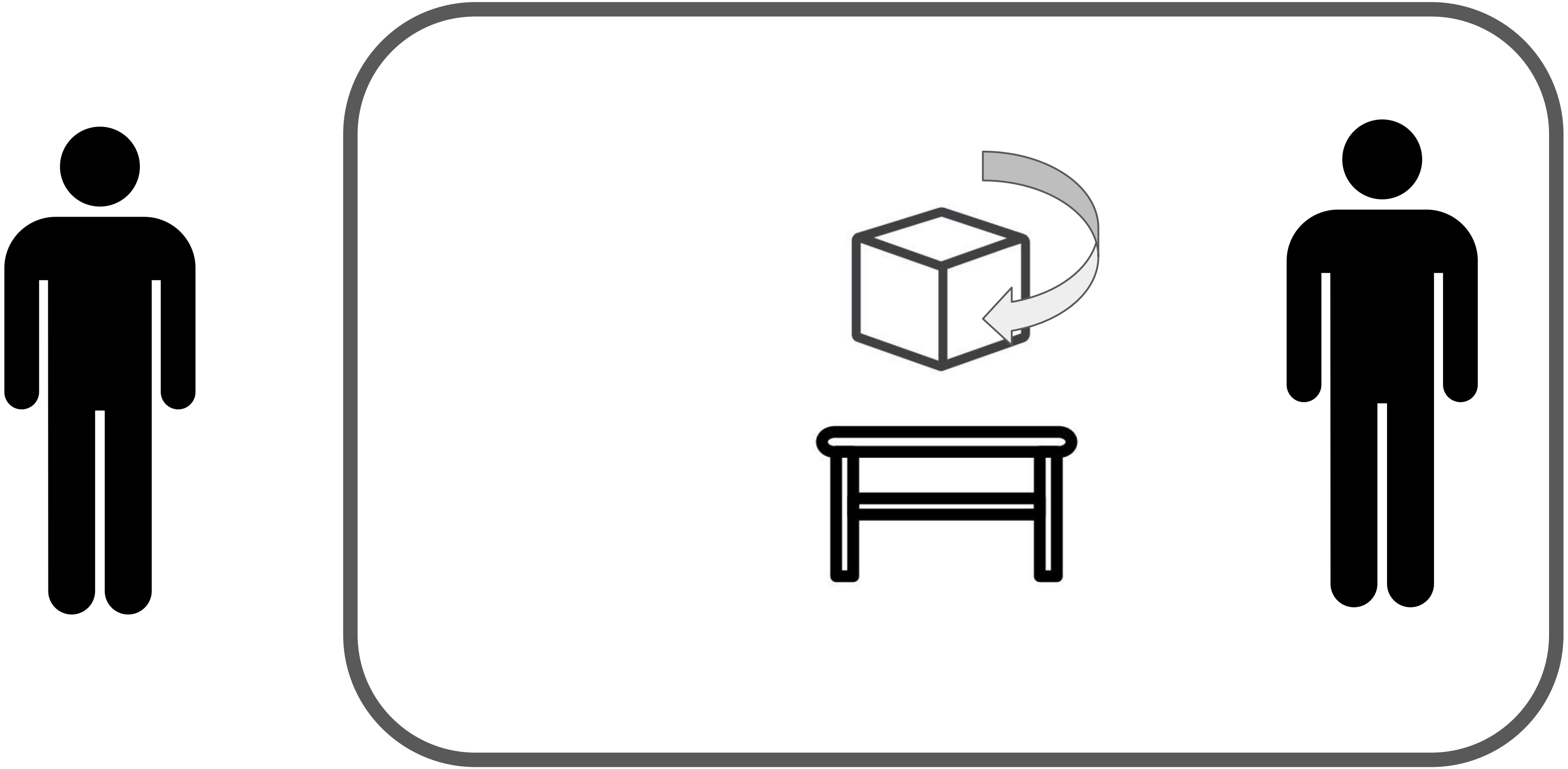
Adversarial Symmetry Discovery



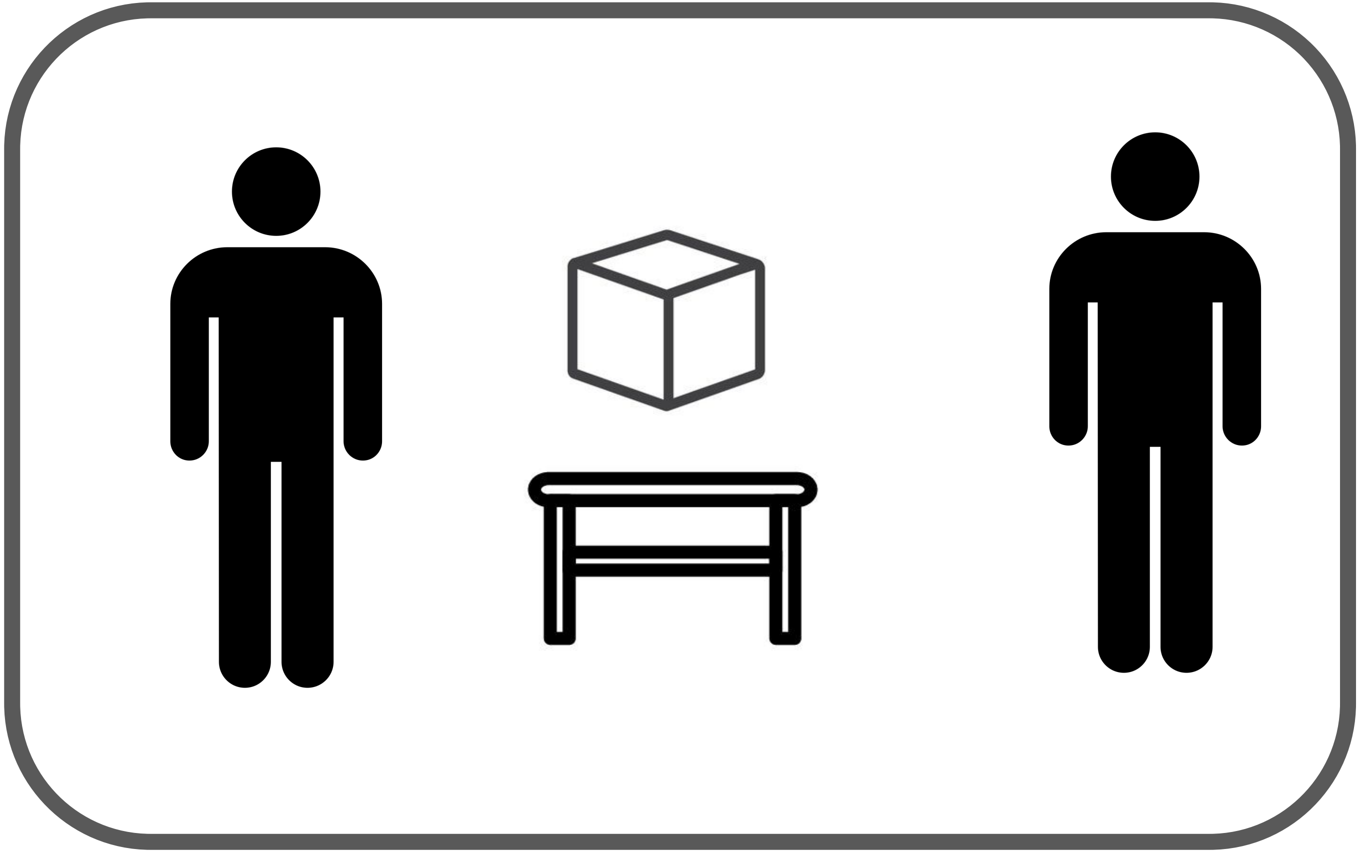
Adversarial Symmetry Discovery



Adversarial Symmetry Discovery

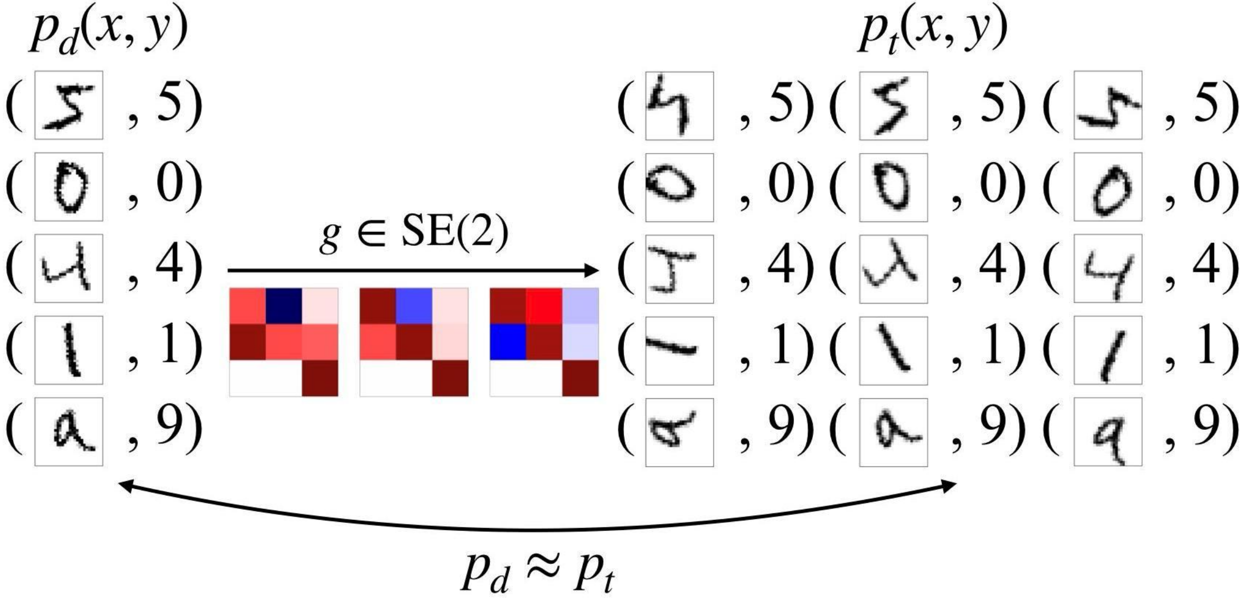


Adversarial Symmetry Discovery

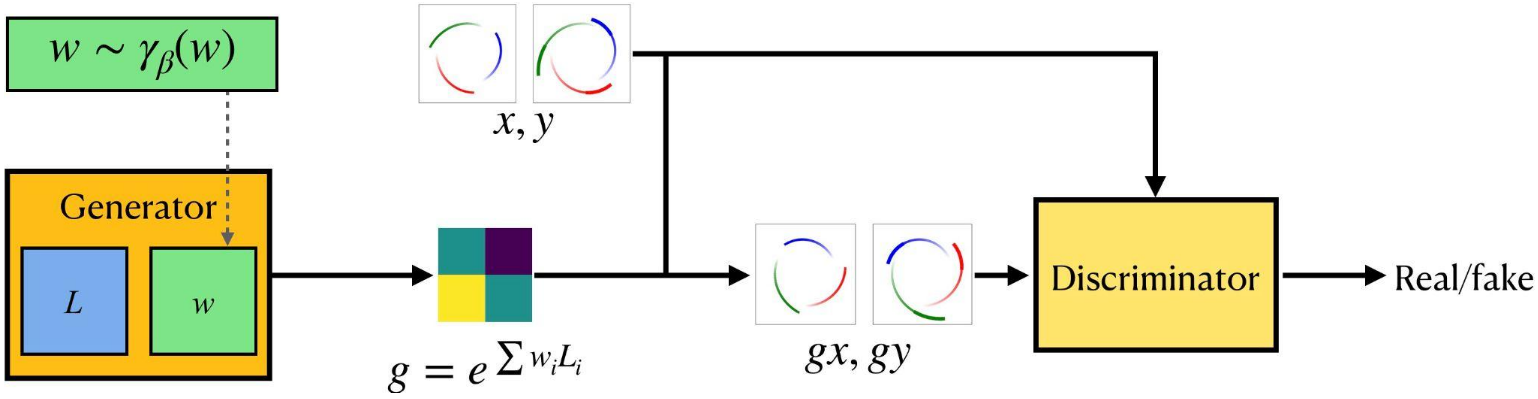


Symmetry Discovery

- Symmetry & Data distribution



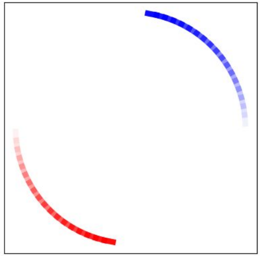
LieGAN: Generative Model for Symmetry Discovery

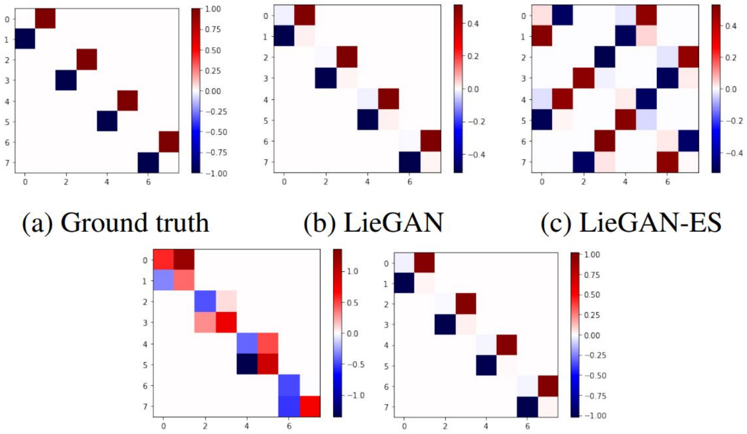


The transformation generator learns a continuous Lie group acting on the data that preserves the original joint distribution. This is an example task of predicting future 3-body movement based on past observations.

Discovering Symmetry in 2-Body Trajectory

- Task: predict future dynamics given the past observations
- Input / output: planar positions and momentums of two masses
- Rotation equivariance (SO(2))





(a) Ground truth (b) LieGAN (c) LieGAN-ES

(d) Augerino+ (e) SymmetryGAN

LieGAN discovers correct rotation symmetry with different parameterizations.

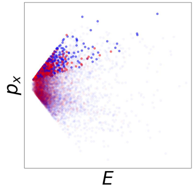
Predicting 2-Body Trajectory

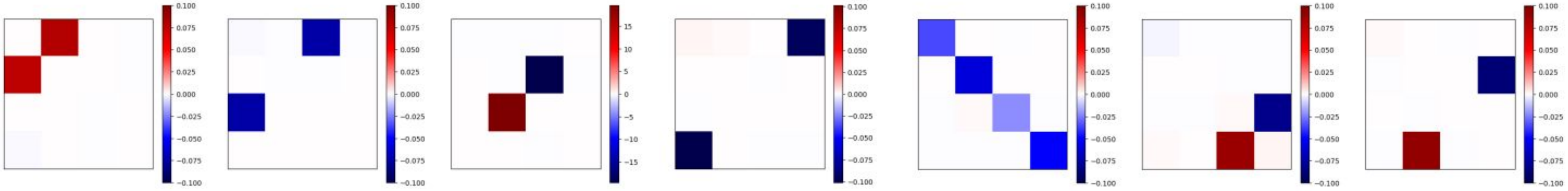
- Test MSE loss for 2-body trajectory prediction
- Symmetries from different discovery models and ground truth are inserted into EMLP or used to perform data augmentation

Model	EMLP	Data Aug.
LieGAN	6.43e-5	3.79e-5
LieGAN-ES	2.41e-4	6.17e-5
Augerino+	9.41e-4	1.47e0
SymmetryGAN	-	6.79e-4
Ground truth	**9.45e-6**	**1.39e-5**
HNN	3.63e-4	
MLP	8.49e-2	

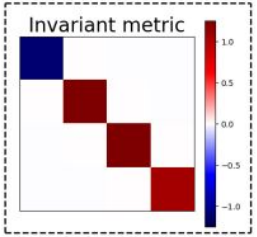
Discovering Lorentz Symmetry in Top Quark Tagging

- Task: binary classification between top quark jets and background
- Input: 4-momenta of the particle jets
- Lorentz transformation invariance (O(1,3))



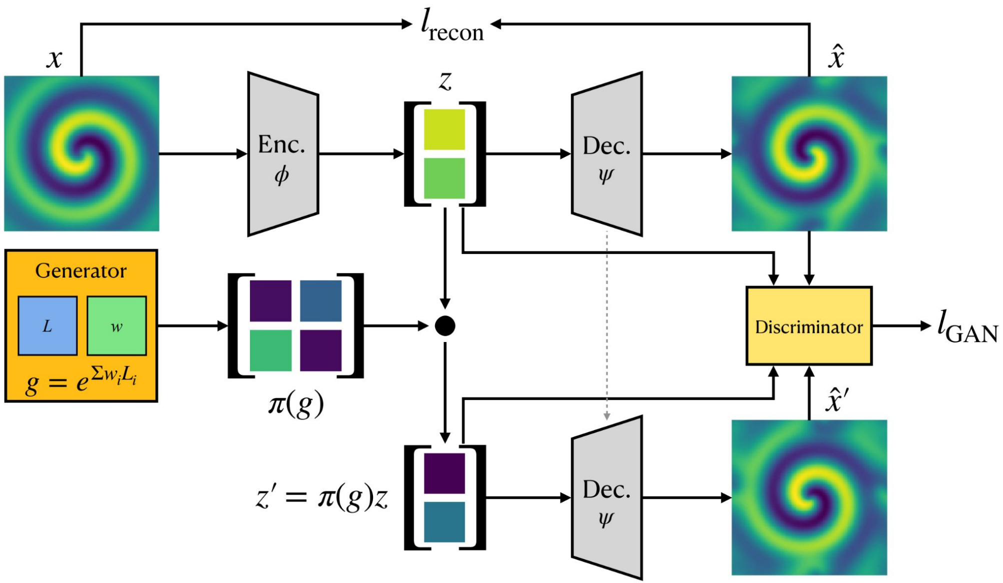


- LieGAN discovers an approximate restricted Lorentz group symmetry
- Computed invariant metric of the discovered symmetry



Latent Space Symmetry Discovery

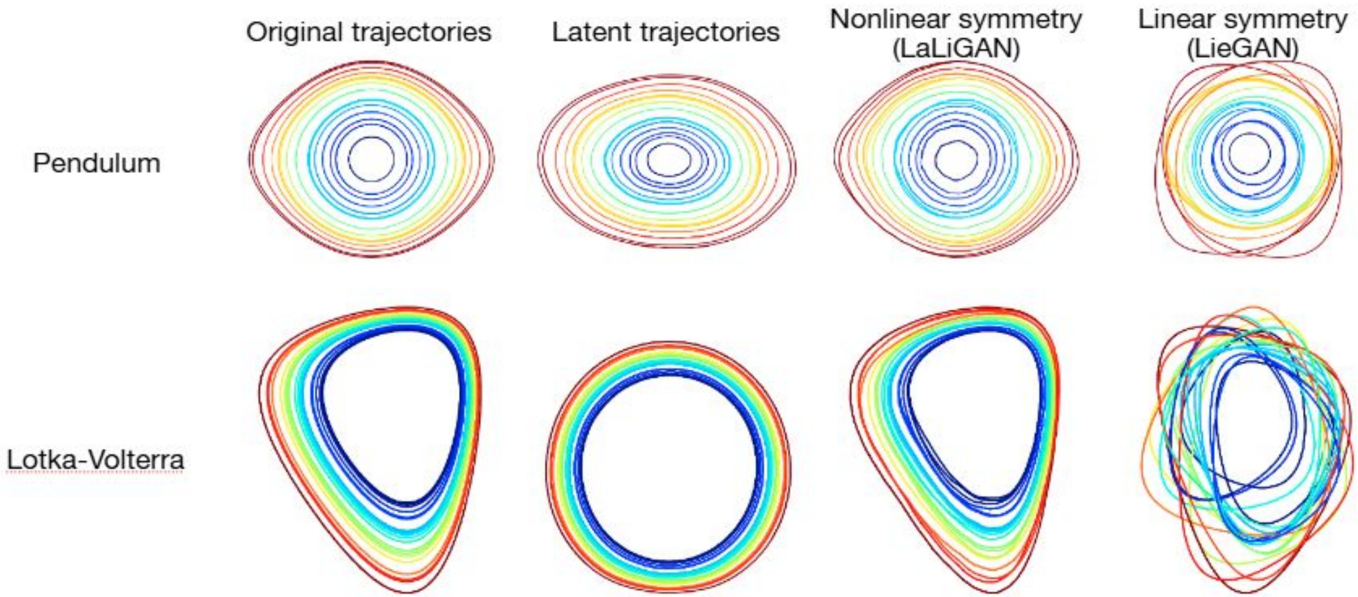
- Nonlinear group actions on data: $\pi' : G \times V \to V$



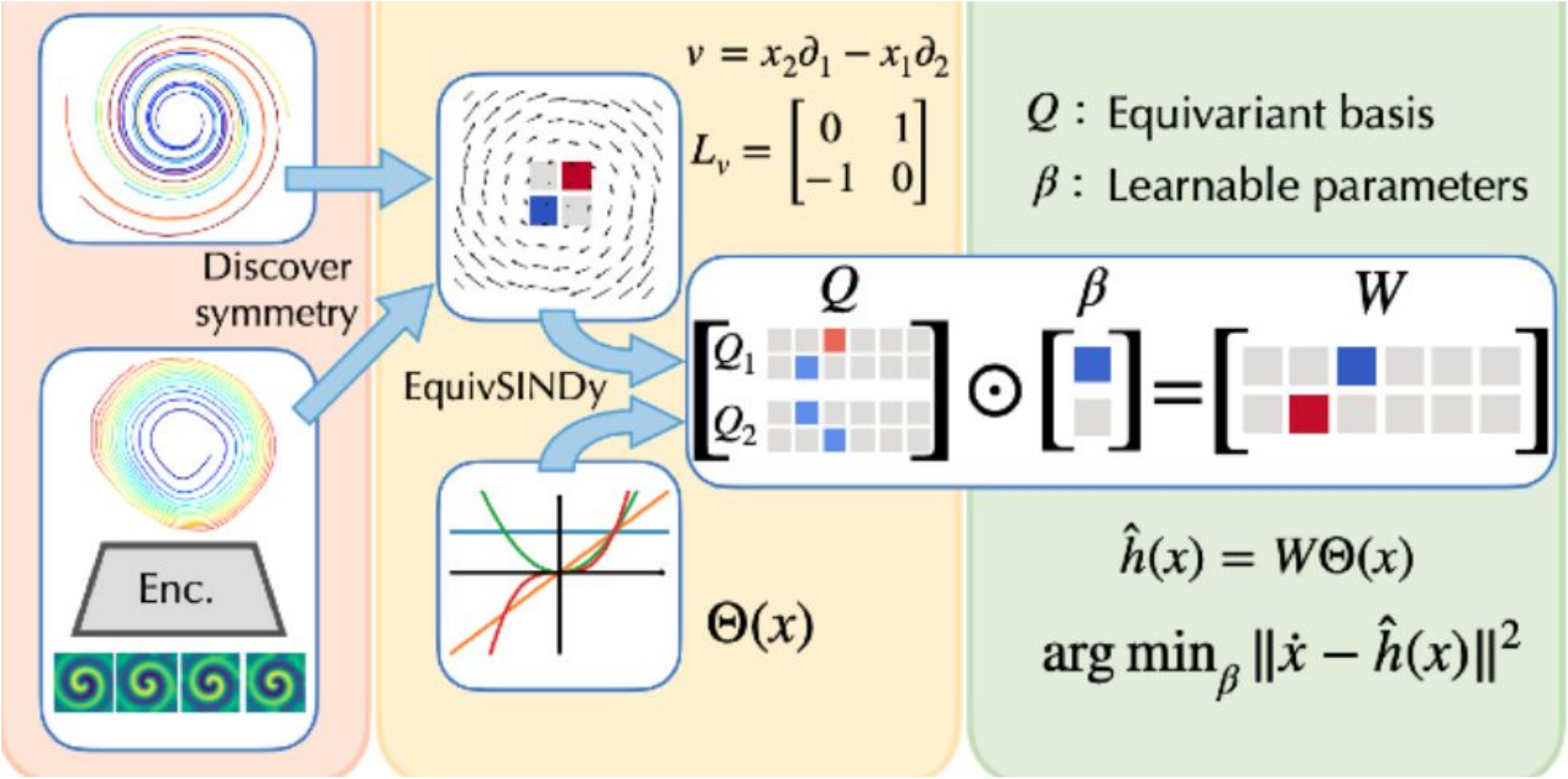
Discovering Latent Symmetry

Example: nonlinear dynamics

- Nonlinear pendulum (top) & Lotka-Volterra equations (bottom)



Joint Discovery of Symmetry and Governing Equation



Equivariant SINDy

Visualizing the equivariant parameter space

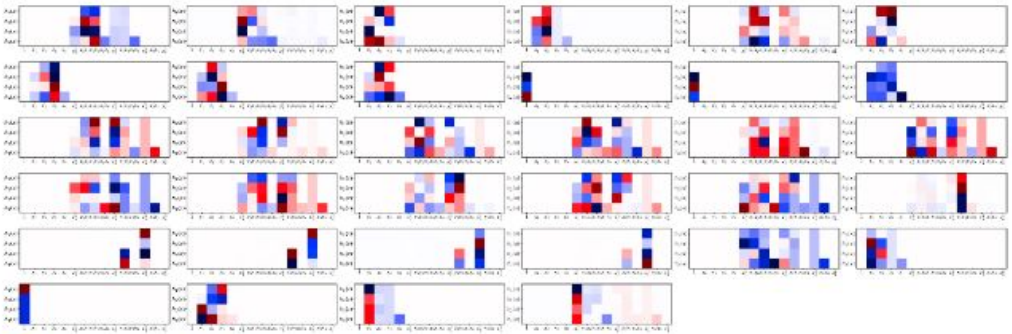
Equation

$$\begin{cases} \dot{S} = 0.15 - 0.6SI \\ \dot{E} = 0.6SI - E \\ \dot{I} = E - 0.5I \\ \dot{R} = -0.15 + 0.5I \end{cases}$$

Symmetry

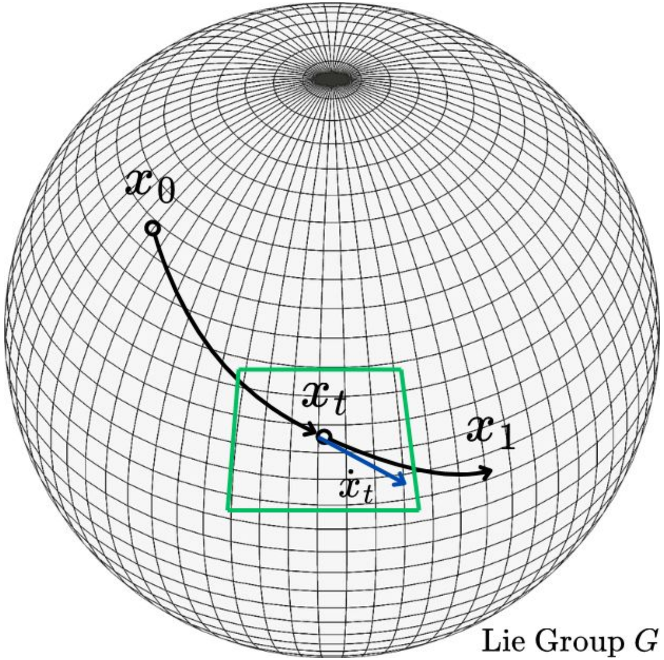
$$v = (S + E + I + R)\partial_R$$

Equivariant basis



60D parameter space reduced to 34D

Discovering Symmetries with Flow Matching



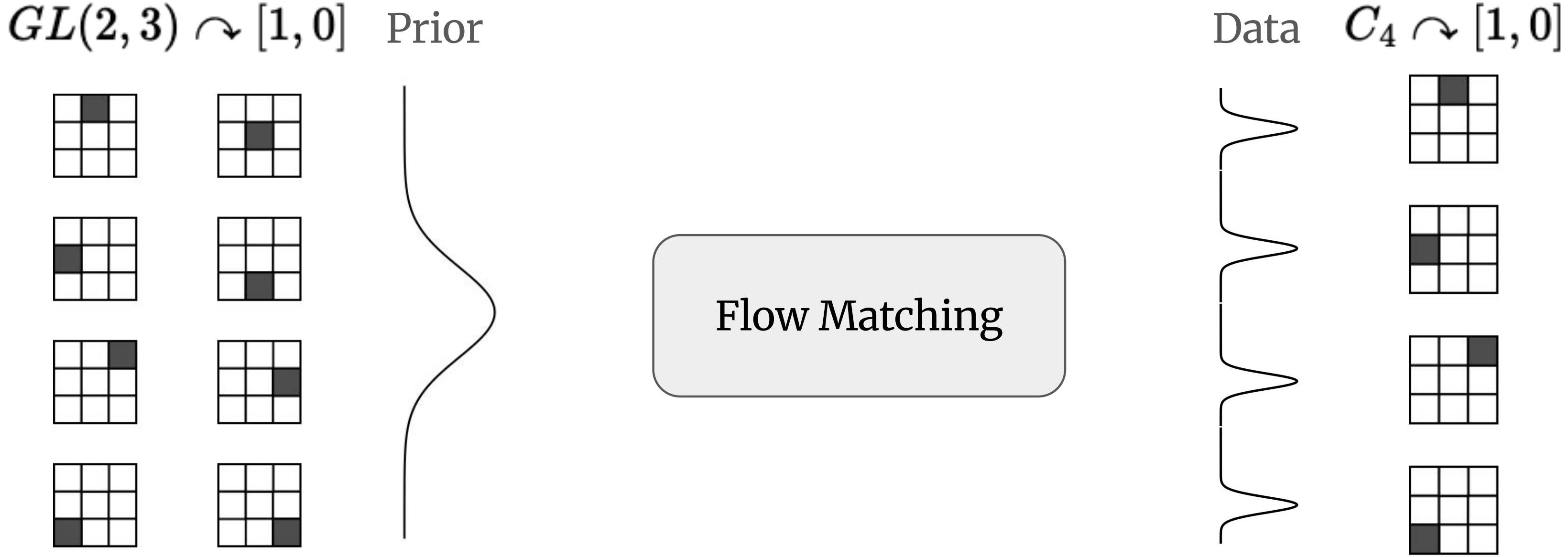
Jung Yeon Park, Yuxuan Chen, Floor Eijkelboom, Jan-Willem van de Meent, Lawson L.S. Wong, Robin Walters [Discovering Lie Groups with Flow Matching](). Under Review.

Symmetry Learning as Distribution Learning

Generative model to learn distribution of data symmetries

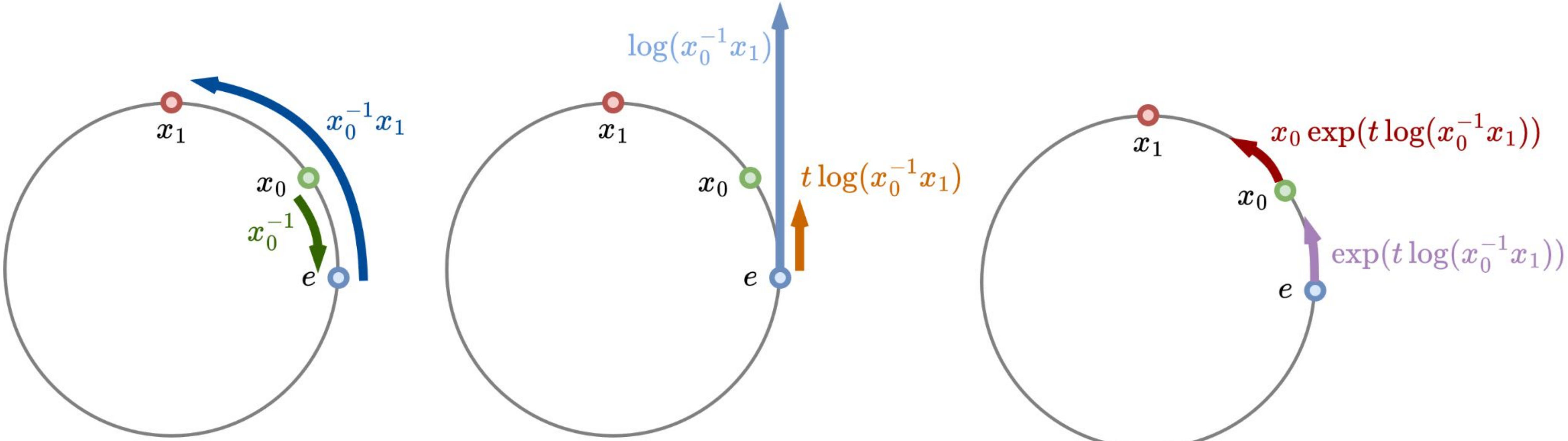
Map from large prior group to transformations observed in data

Address Limits of LieGAN: Hard to optimize, Dist over transforms, No Discrete groups



Flow Matching on Lie Groups

$$\gamma : [0,1] \to G; t \mapsto x_0 \exp(t \log(x_0^{-1} x_1)).$$



1. Sherry, F. M., & Smets, B. (2025). Flow Matching on Lie Groups. arXiv preprint arXiv:2504.00494.

Method

Fit distribution p over G. Flow Matching over the *group* conditioned on data.

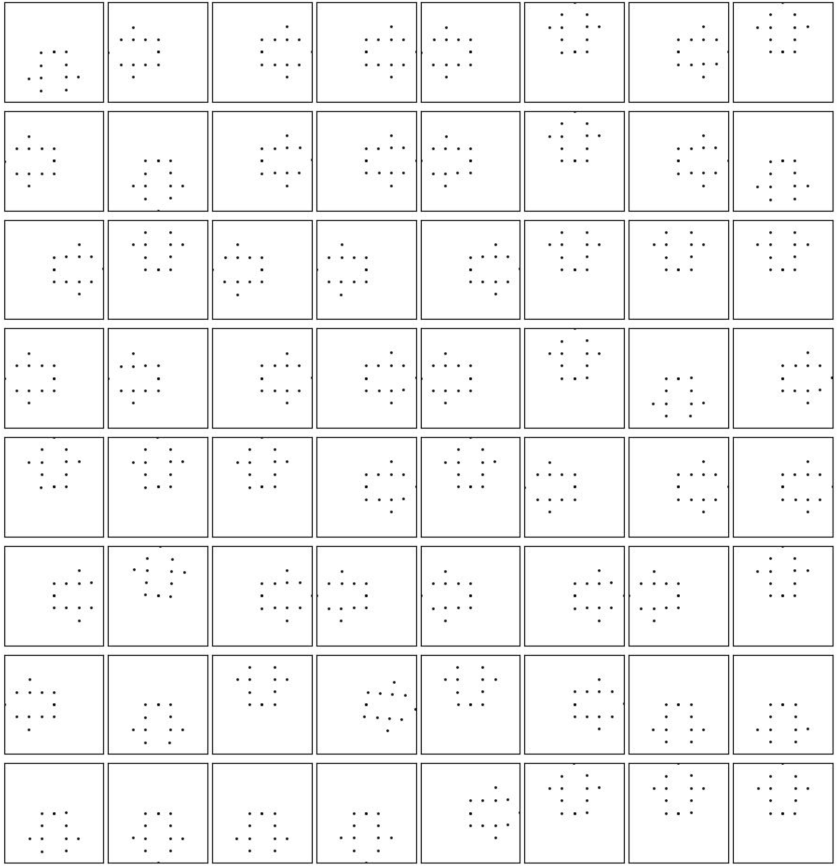
Algorithm 1 Training

repeat
  $x_1 \sim q$ ▷ q is data distribution
  $t \sim \mathcal{U}(0, 1)$
  $g \sim p(G)$ ▷ restrict domain if noncompact G
  $x_0 = gx_1$
  $A = \log(g^{-1})$
  $x_t = x_0 \exp(tA)$
  Take gradient descent step on
    $\nabla_\theta \left\| v_\theta(x_t, t) - A \right\|^2$
until converged

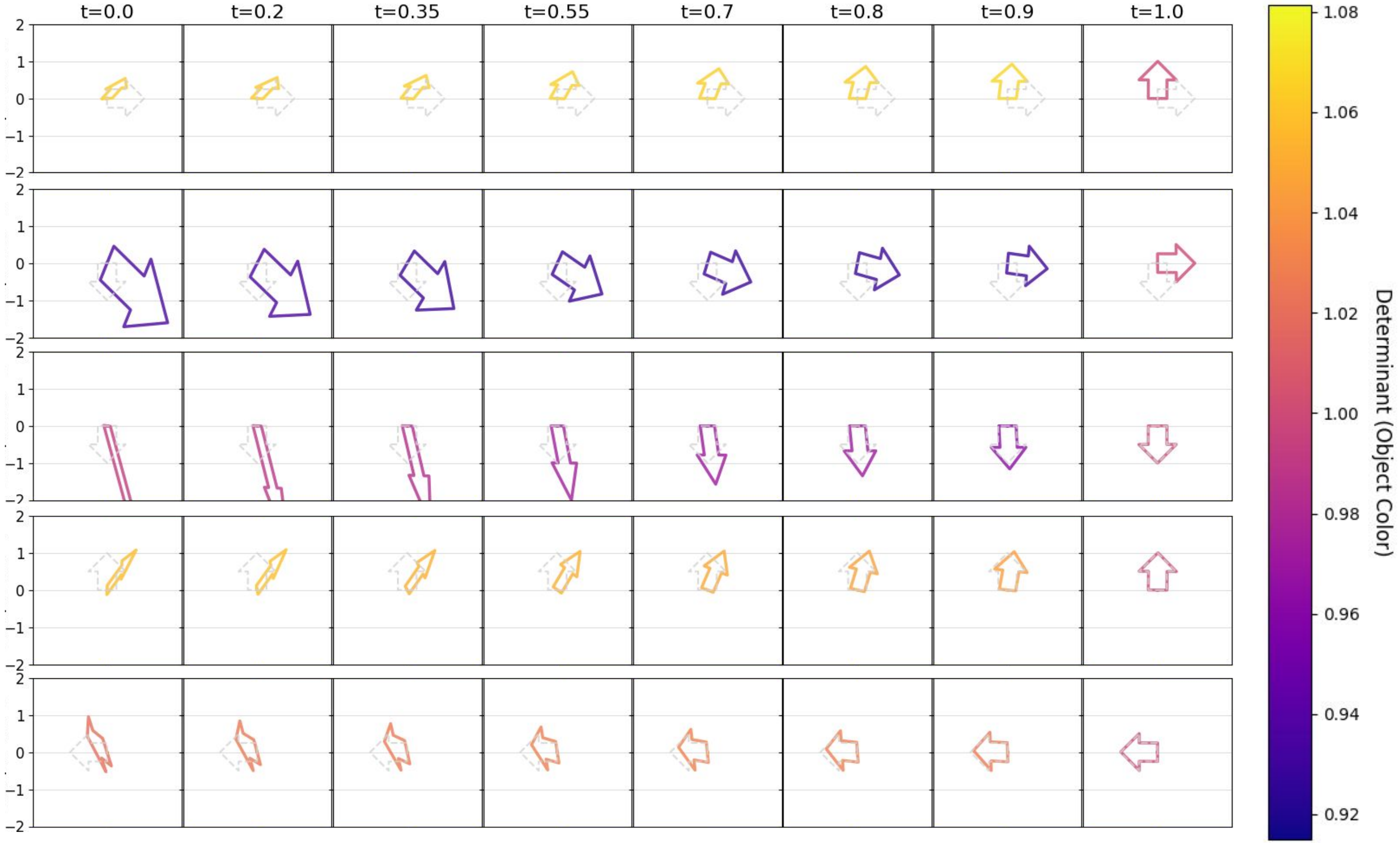
Algorithm 2 Sampling x_1

$x_1 \sim q$ ▷ q is data distribution
$g \sim p(G)$ ▷ restrict domain if noncompact G
$x_0 = gx_1$
$\Delta_t = 1/N$
for $n = 0, \ldots, N-1$ **do**
  $t = n\Delta_t$ ▷ Use Euler's method
  $A_t = v^\theta(x_t, t)$
  $x_{t+\Delta_t} = x_t \exp(\Delta_t A_t)$
end for
return x_1'

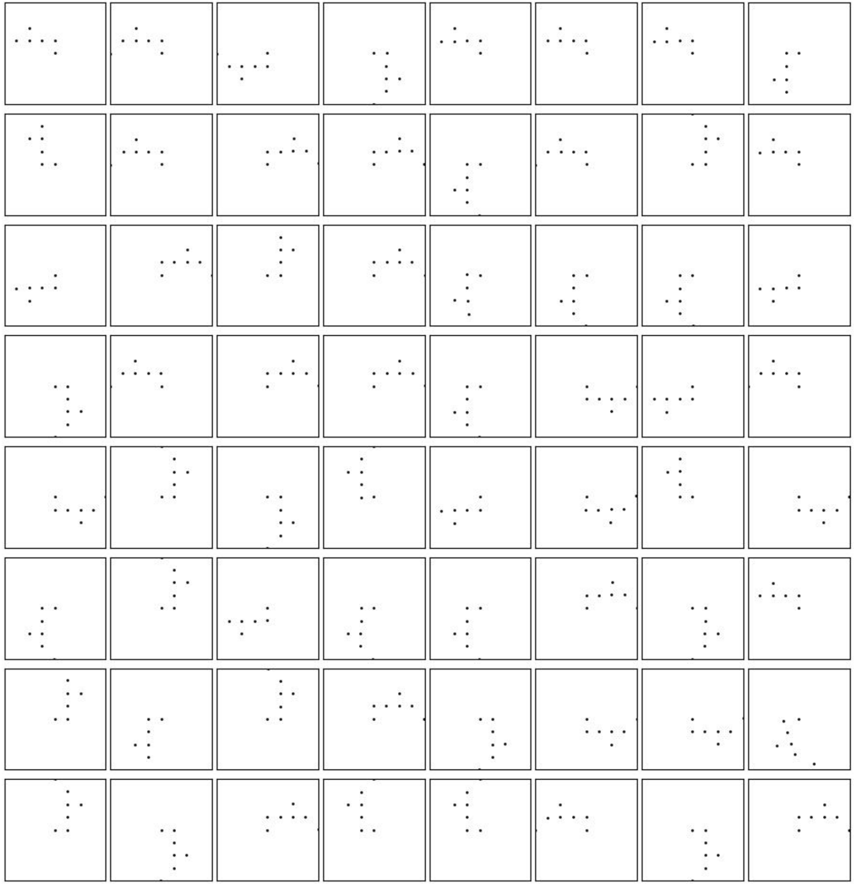
Part 3: Results on GL(2) to C_4



Results on GL(2) to C_4



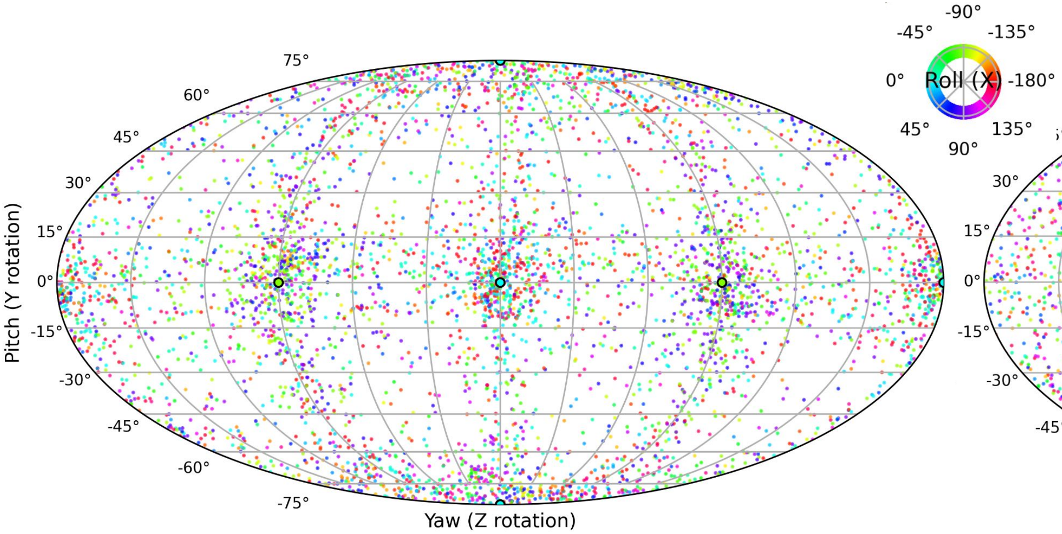
Results on GL(2) to D_4

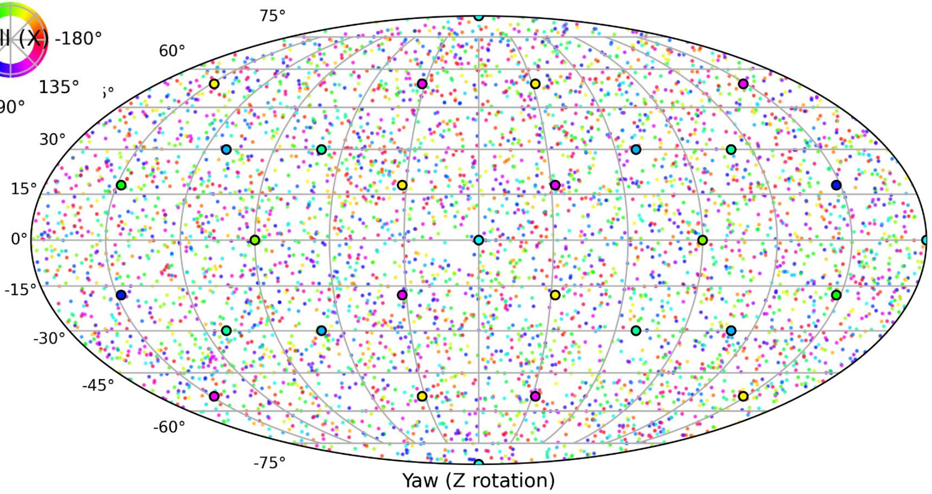


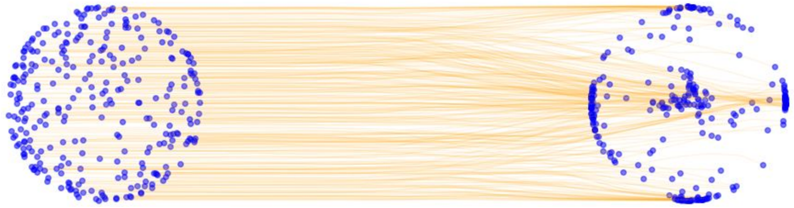
Results on SO(3)

SO(3) to Tet

SO(3) to Ico







Collaborators and support.

Thanks you for your attention!

Questions?

Rui Wang
Amazon

Rajmonda Caceres
MITLL

Peter Schindler
Northeastern

Rose Yu
UC San Diego

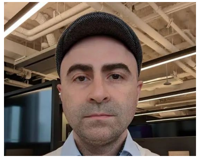
Linfeng Zhao
Northeastern

Robert Platt
Northeastern

Lawson Wong
Northeastern

Colin Kohler
RAI

Xupeng Zhu
Amazon



Nima Dehmamy
IBM

Tess Smidt
MIT

MIT LINCOLN LABORATORY

THE AI INSTITUTE

Ondrej Biza
RAI

Jan Willem van de Meent
UvAmsterdam

Elyssa Hofgard
MIT

Jianke Yang
UCSD

Floor Eijkelboom
UvAmsterdam