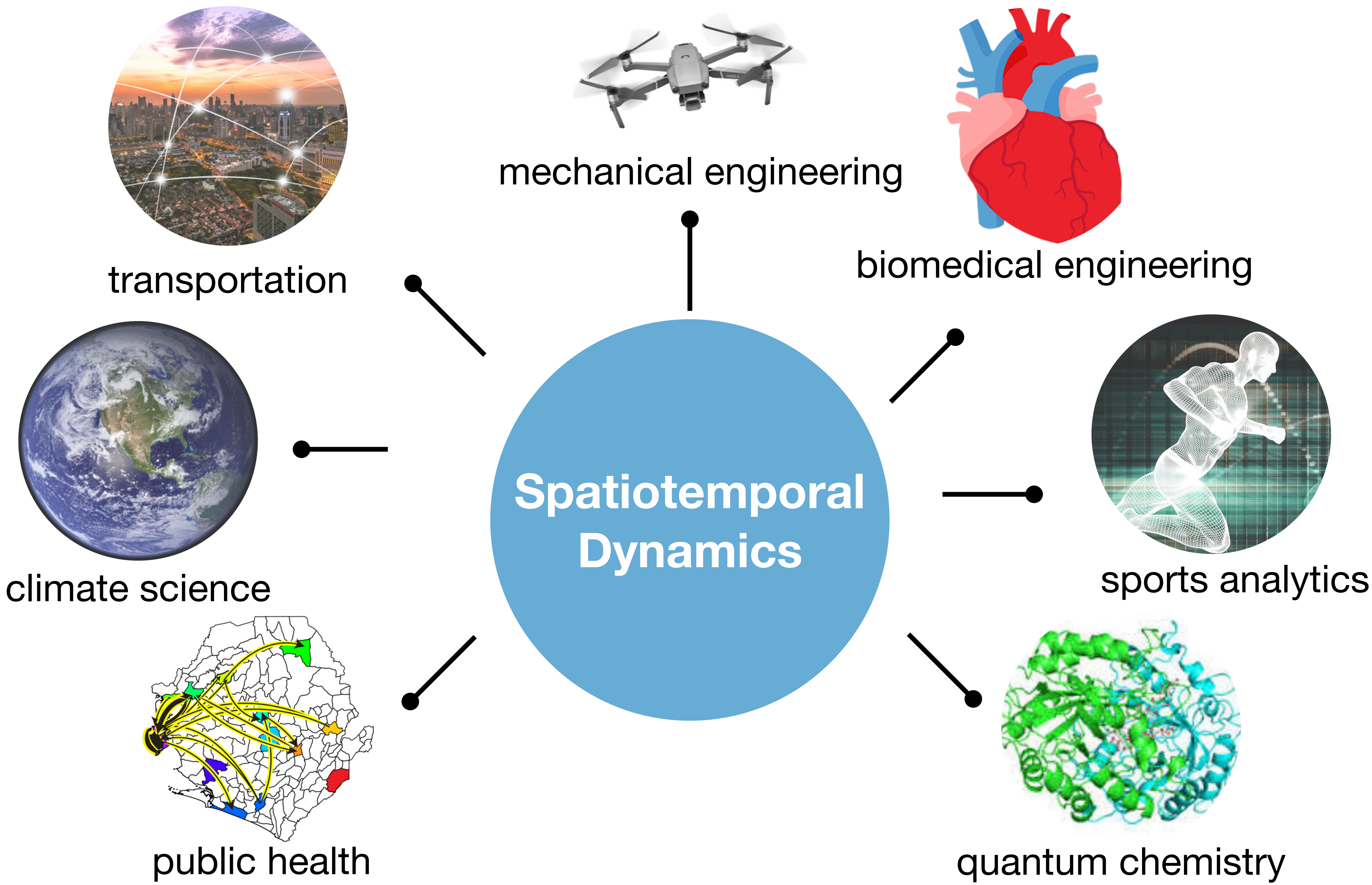
UC San Diego

Incorporating Symmetry for Learning Spatiotemporal Dynamics

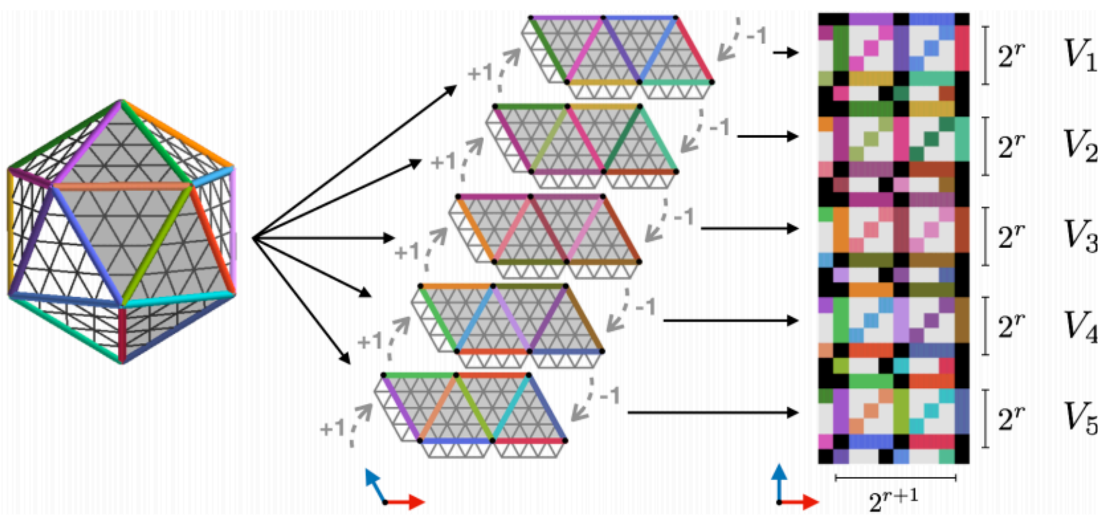
Rose Yu

Assistant Professor
University of California, San Diego

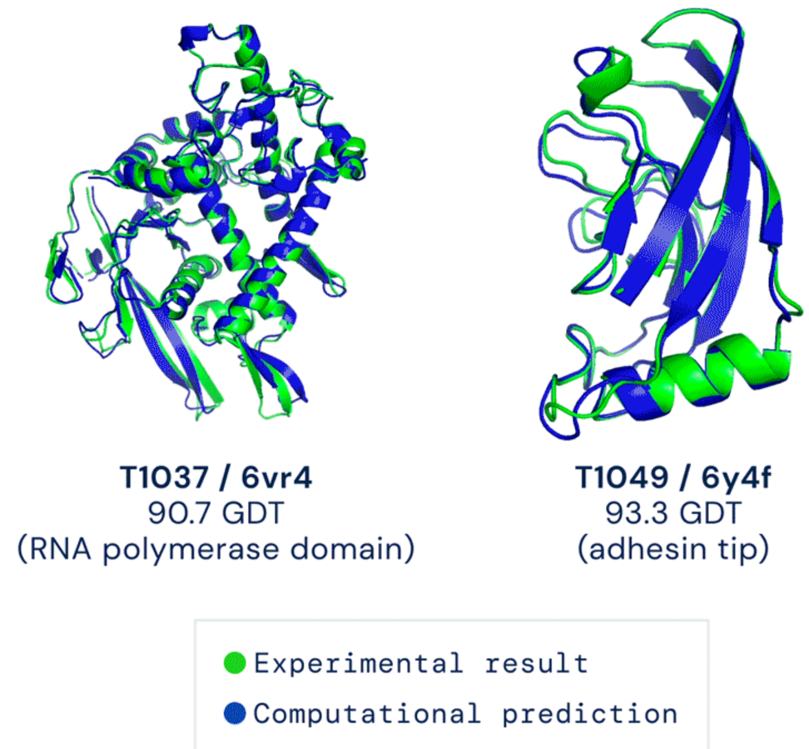
Learning Spatiotemporal Dynamics



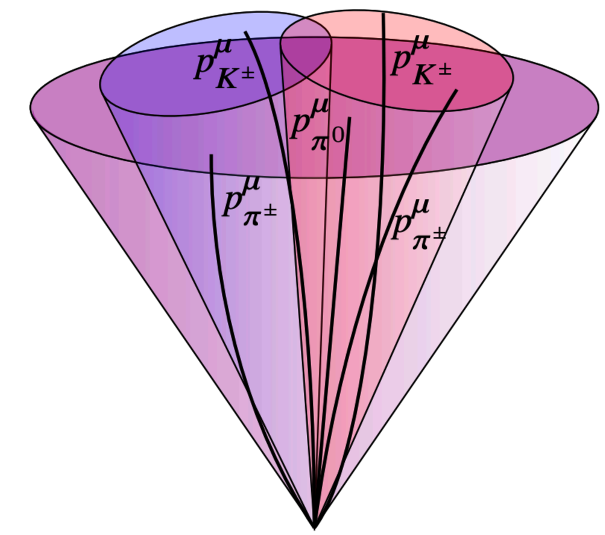
Success of Symmetry



Cohen et al 2019



Jumper et al 2021

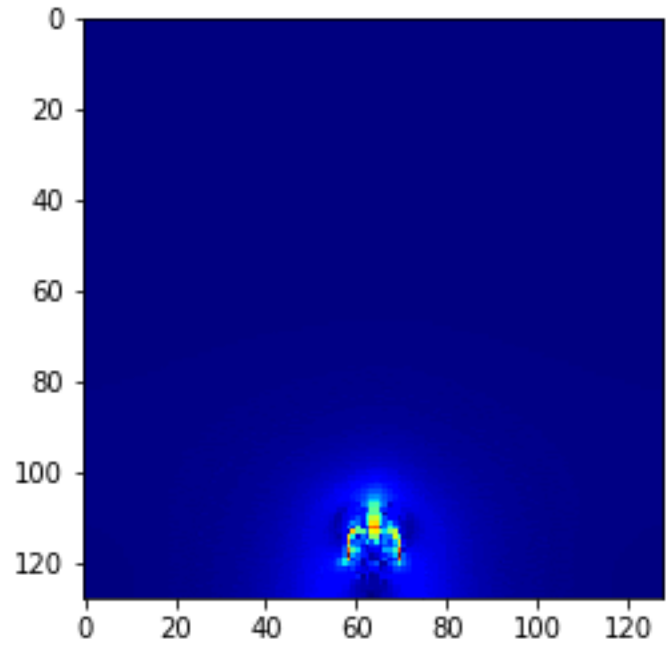


Bogatskiy et al 2020

Ravanbakhsh et al. (2017); Kondor & Trivedi (2018); Cohen & Welling (2016b); Thomas et al. (2018); Maron et al. (2020); Walters et al. (2021)…..

How about Spatiotemporal Dynamics?

Incorporating Symmetry for Generalization

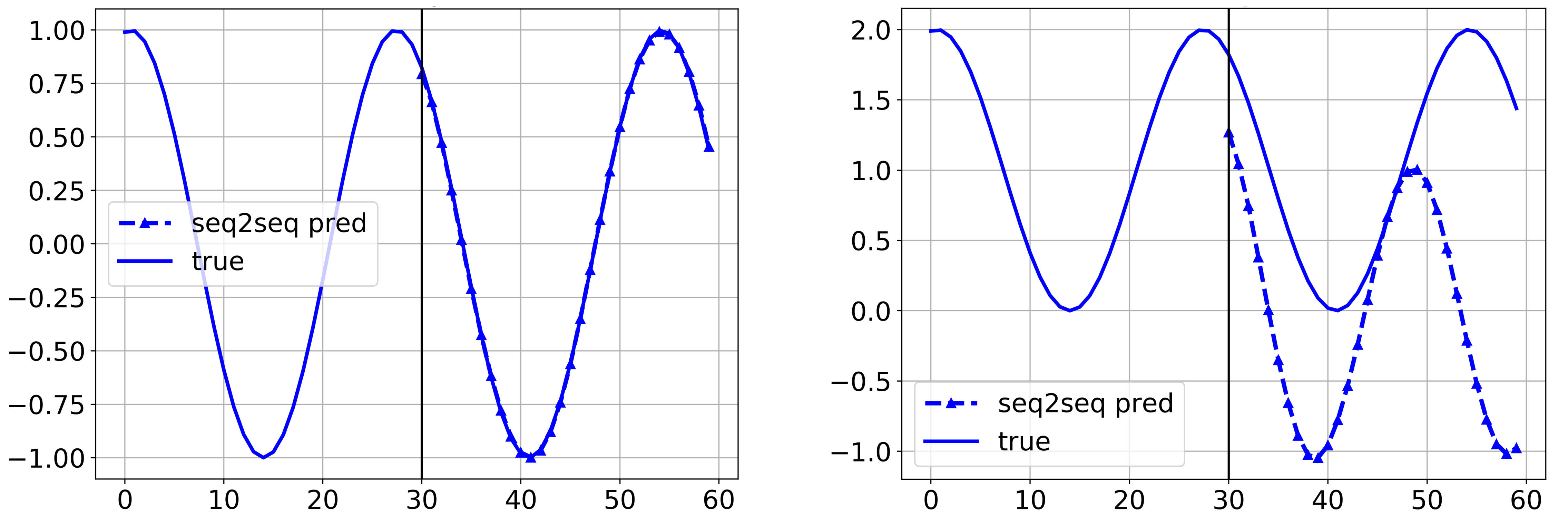


Rui Wang

Robin Walters

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

Generalization Challenge



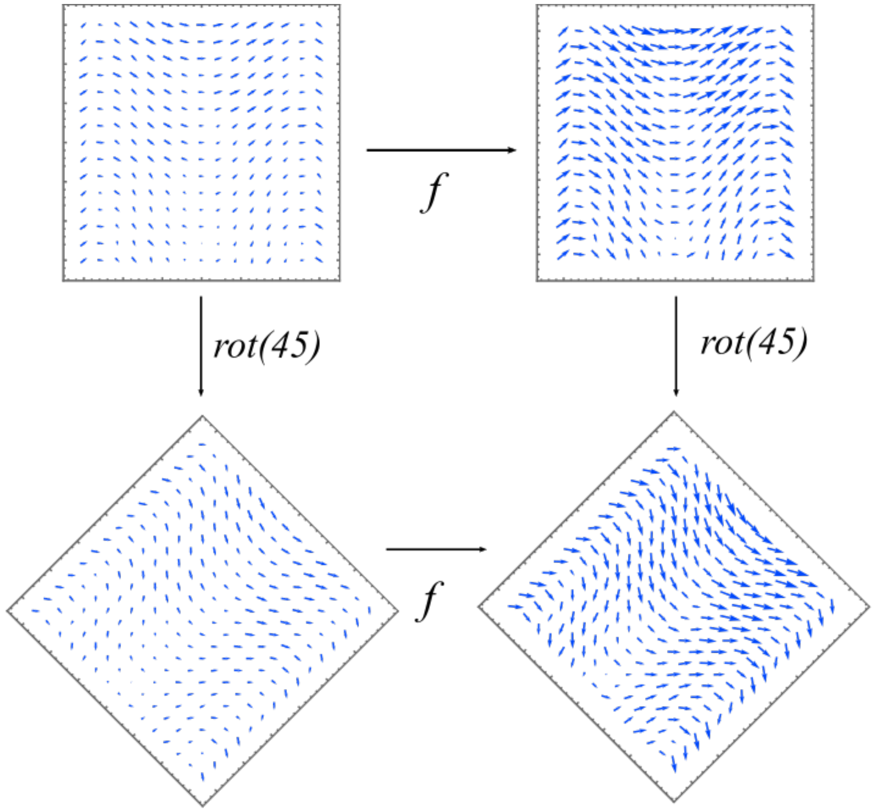
$\theta^{\textbf{tr}} \in [a, b]$ $\theta^{\textbf{te}} \in [a, b]$ $\theta^{\textbf{tr}} \in [a, b]$ $\theta^{\textbf{te}} \in [b, c]$

- Generalization: fundamental challenge in dynamics forecasting
  - Performance degrades with test *distributional shift*
  - Punchline*:* distributions change, laws of physics do not!

Bridging Physics-based and Data-driven modeling for Learning Dynamical Systems
Rui Wang, Danielle Maddix, Christos Faloutsos, Yuyang Wang, Rose Yu.
International Conference in Learning for Dynamics and Control (L4DC), 2021

Conservation Laws and Symmetry

- **Noether's theorem**: *For every symmetry, there is a corresponding conservation law*



- **Invariance, Equivariance**:
  - G-invariant: $f(g(x)) = f(x)$
  - G-equivariant: $f(gx) = gf(x)$

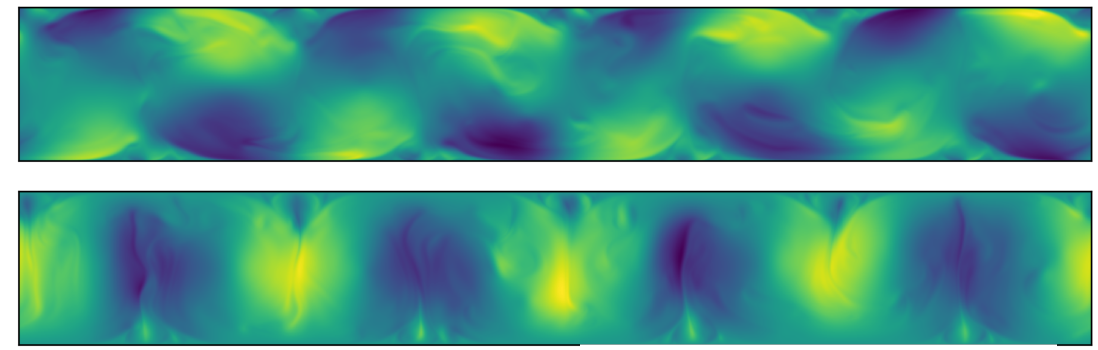
Symmetry in Dynamical Systems

- A system of differential operators $D = \{P_1, \cdots, P_r\}$
- if ϕ is a solution of D, then for all $g \in G$, $g(\phi)$ is also a solution
- **2D Navier-Stokes Equations**

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

$$\nabla \cdot \mathbf{w} = 0$$

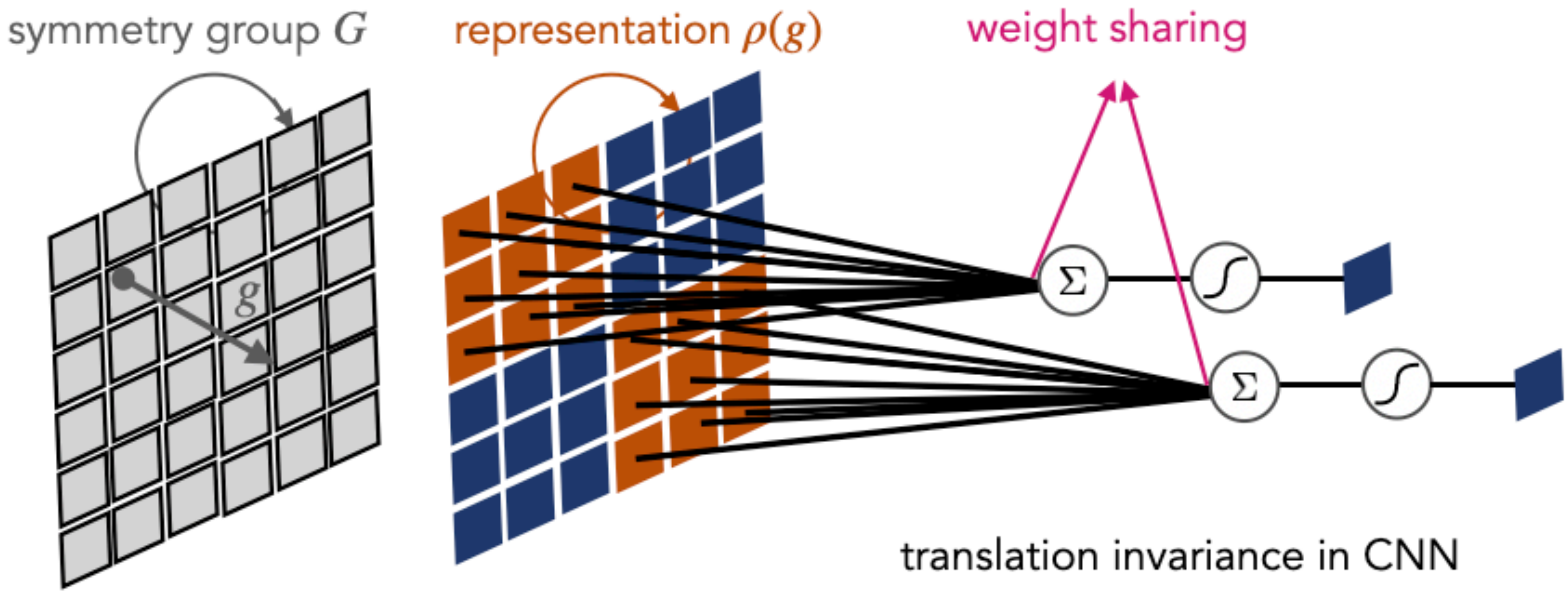
$$\frac{\partial T}{\partial t} + (\mathbf{w} \cdot \nabla)T = \kappa \nabla^2 T$$



Symmetries	*NS Equ.*
Space translation	$\boldsymbol{w}(\boldsymbol{x}-\boldsymbol{v}, t)$
Time translation	$\boldsymbol{w}(\boldsymbol{x}, t-\tau)$
Uniform Motion	$\boldsymbol{w}(\boldsymbol{x}, t)+\boldsymbol{c}$
Reflect/rotation	$R\boldsymbol{w}(R^{-1}\boldsymbol{x}, t)$
Scaling	$\lambda\boldsymbol{w}(\lambda\boldsymbol{x}, \lambda^2 t)$

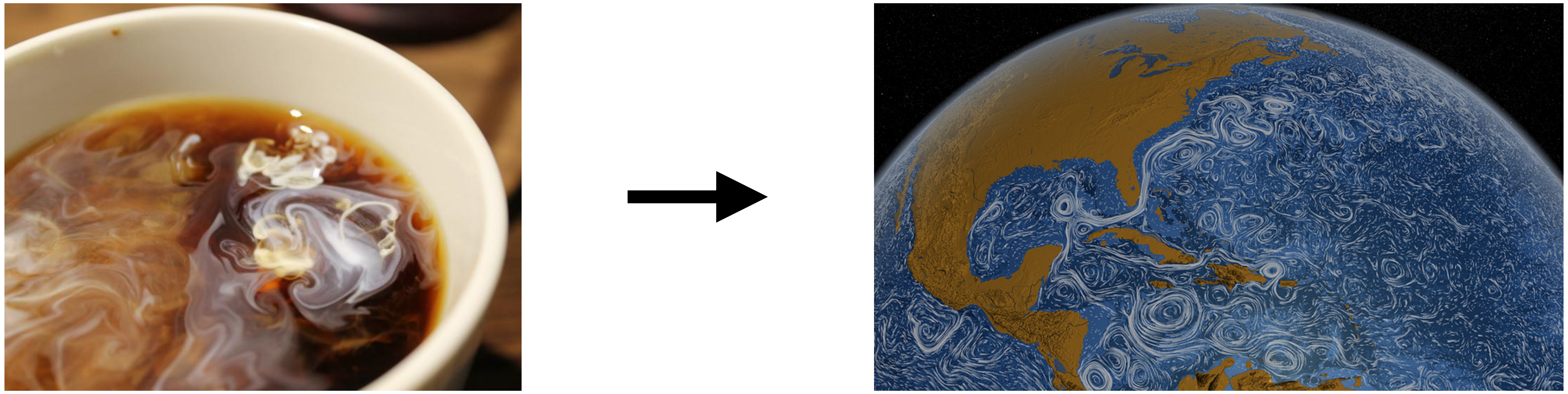
Scaling Law

Weight Symmetry



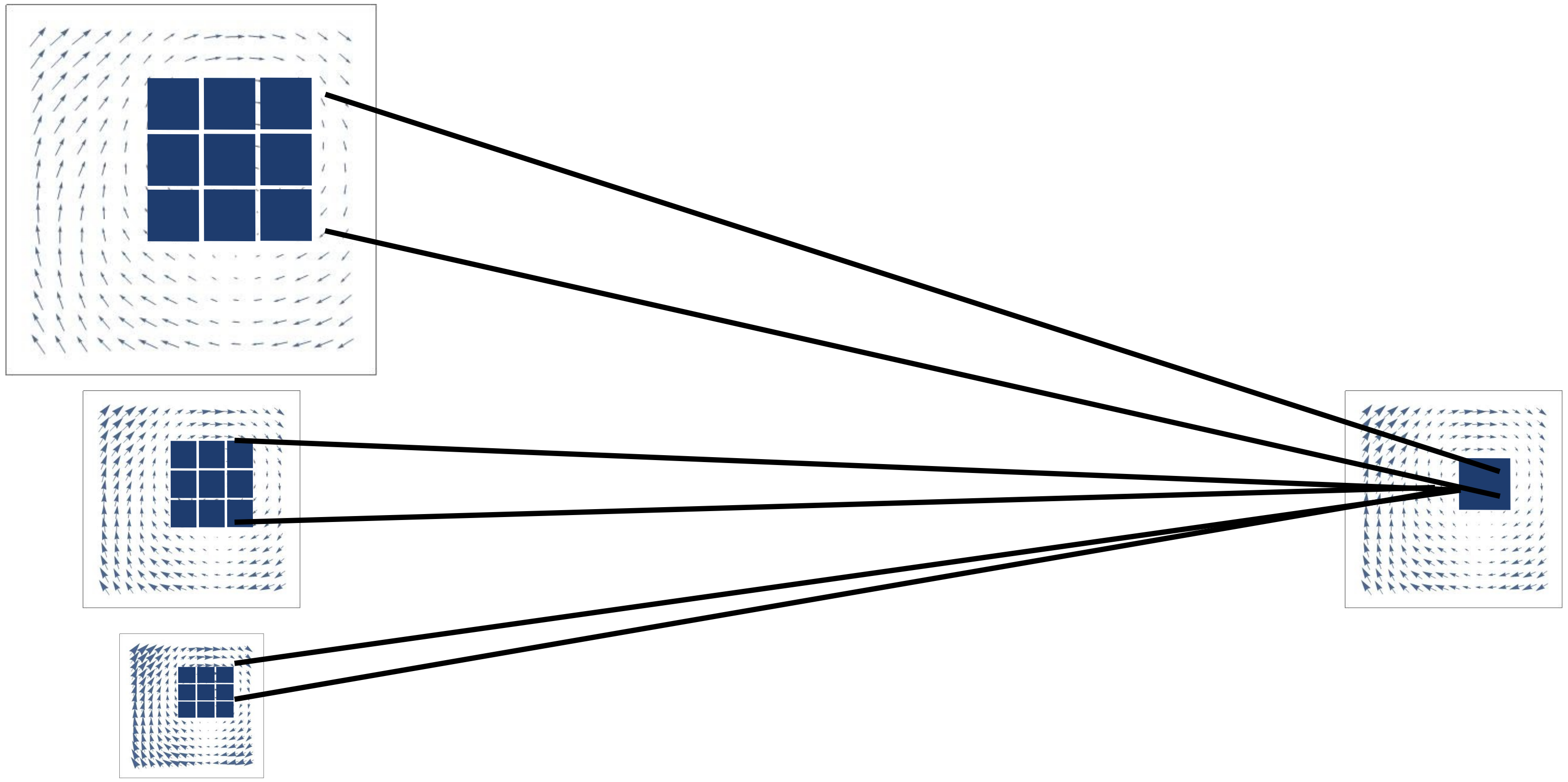
Theorem (Weiler & Cesa 2019): a convolutional layer is G-equivariant if and only if the kernel satisfies $K(gv) = \rho_{out}^{-1}(g)K(v)\rho_{in}(g)$ for all $g \in G$, with action maps ρ_{in} and ρ_{out} .

Symmetry: Scaling



- Standard convolution shares weights across the input by translating a kernel across the input.
- For scale-equivariant convolution, we must translate and scale a kernel across the input.

Symmetry: Scaling

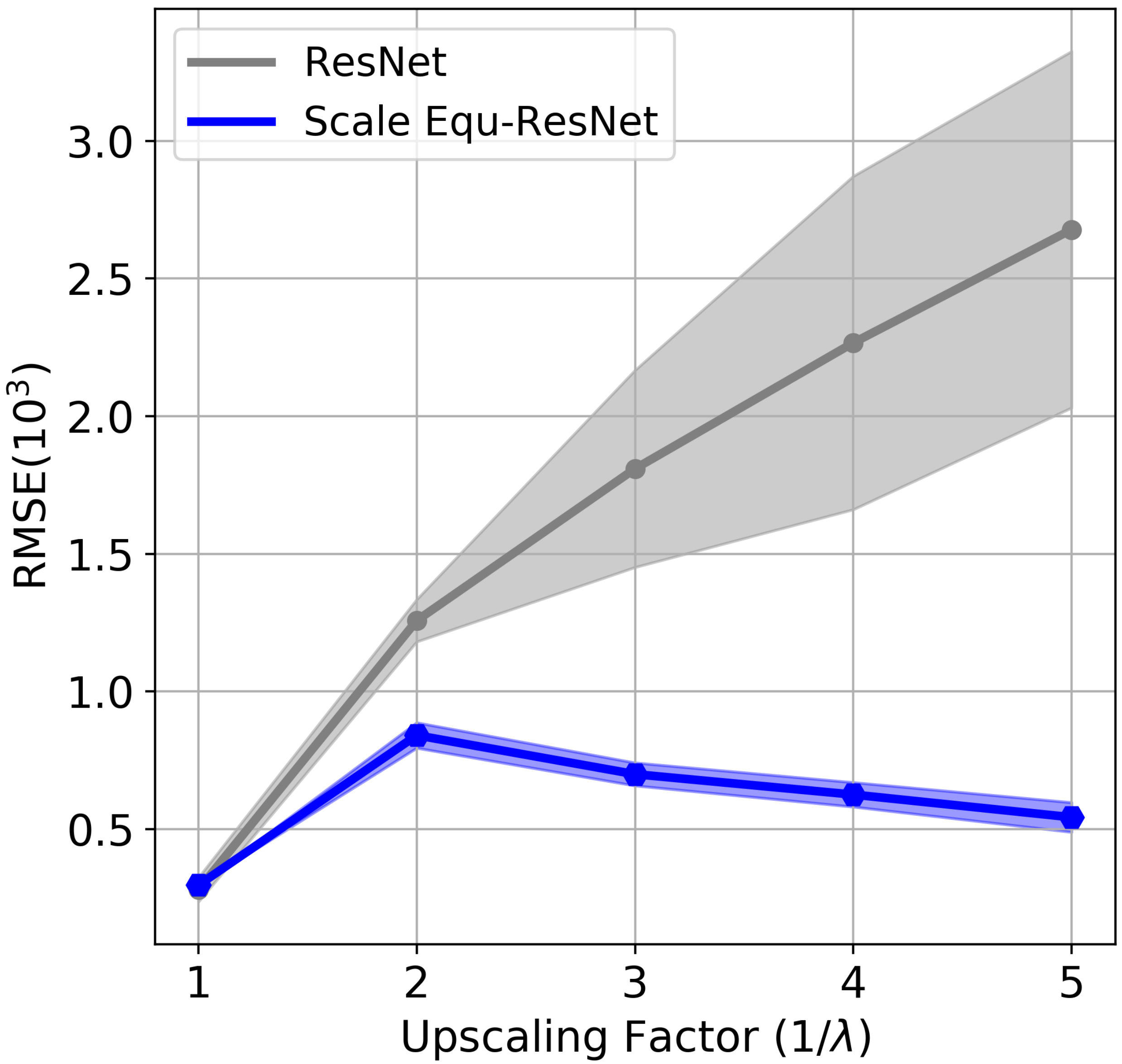


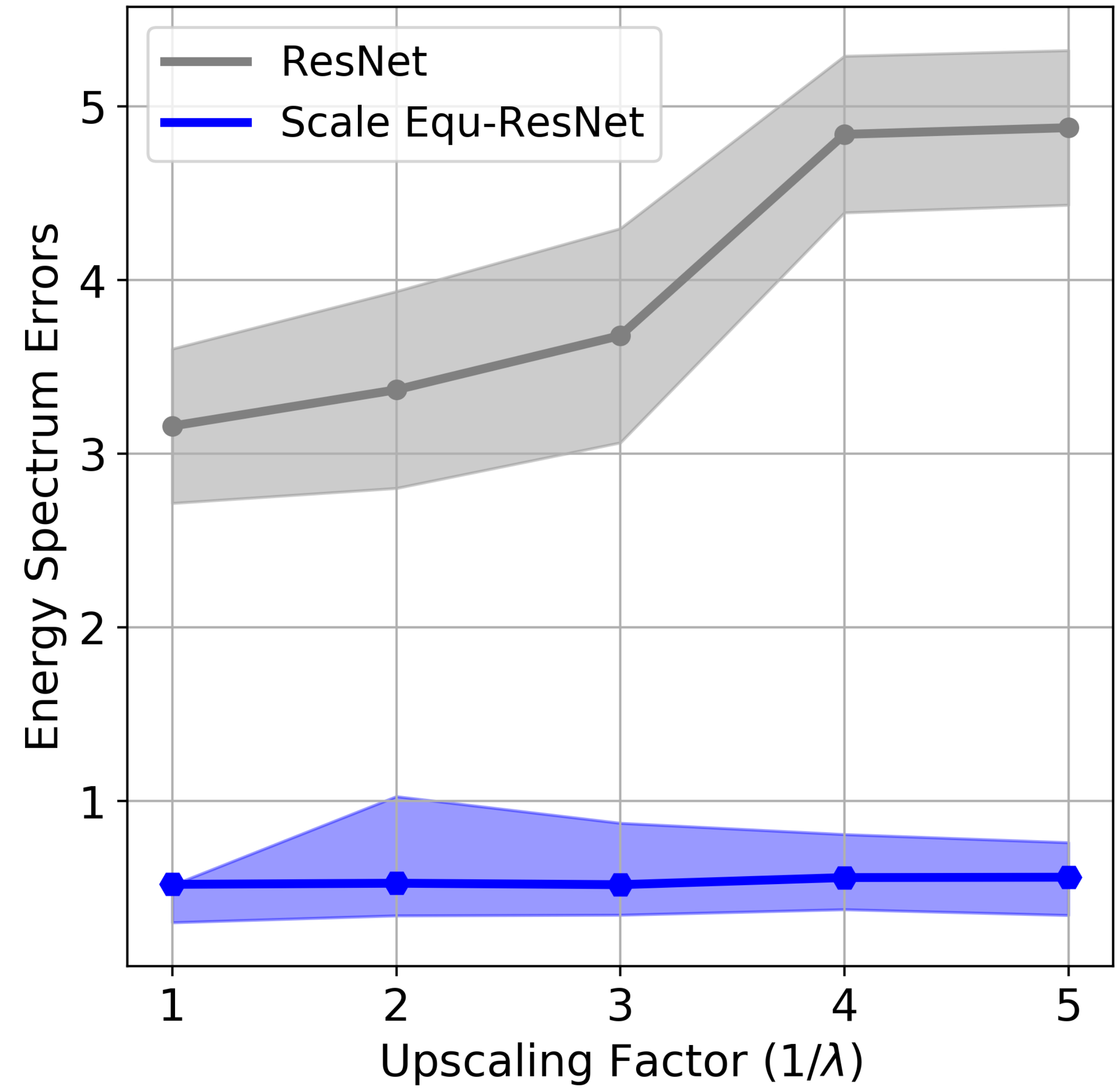
- *Scale* equivariant

$$\boldsymbol{v}(p) = \sum_{\lambda \in \mathbb{Z}_{>0}, q \in \mathbb{Z}^2} (T_\lambda \boldsymbol{w})(p+q)(T_\lambda K)(q),$$

$T_\lambda w(x, t) = \lambda w(\lambda x, \lambda^2 t)$

Ocean Currents Forecast





Physically Consistent Predictions!

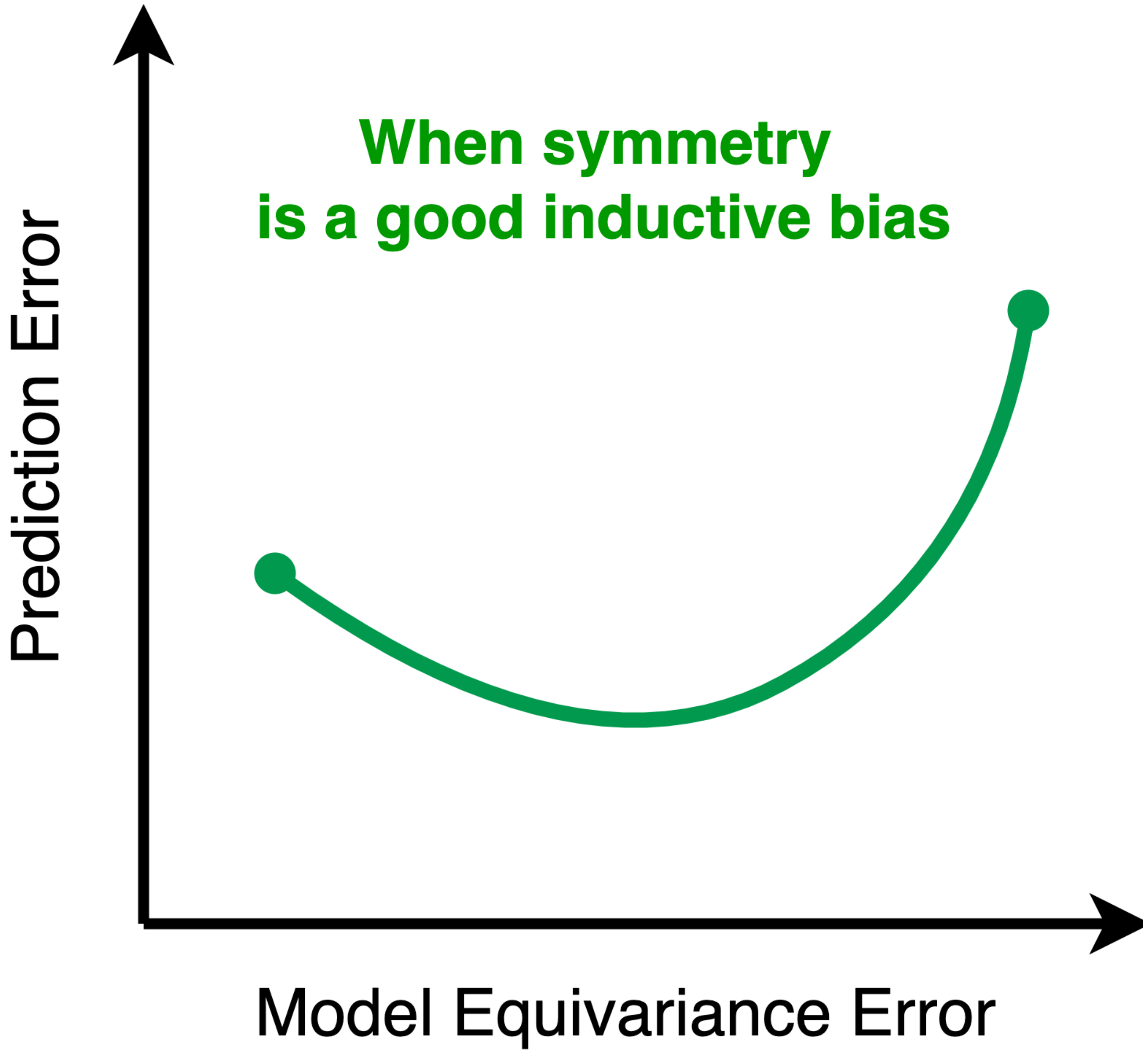
Approximately Equivariant Networks

Rui Wang

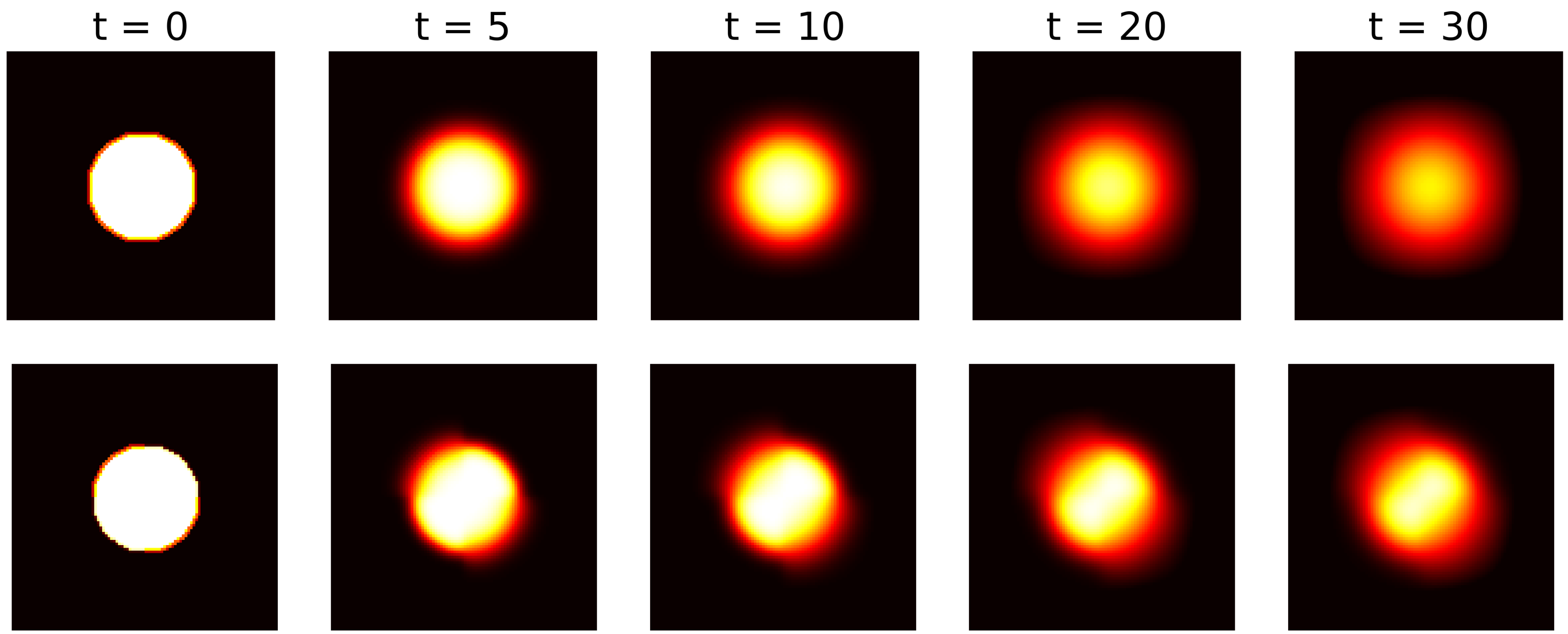
Robin Walters

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

Symmetry as Inductive Bias



Approximate Symmetry



Definition: Let $f: X \to Y$ be a function and G be a group. Assume that G acts on X and Y via representations ρ_X and ρ_Y. We say f is ϵ-approximately G-equivariant if for any $g \in G$,

$$\|f(\rho_X(g)(x)) - \rho_Y(g)f(x)\| \leq \epsilon.$$

Equivariant Convolution

- Group Convolution **(G-conv)**

$$f *_G K(g) = \sum_{h \in G} f(h) K(g^{-1}h)$$

  - G-conv does not need to precompute an equivariant kernel basis
  - But limited to discrete (compact) group, not efficient when the group order is large
- G-Steerable Convolution **(Steer)**

$$K(hx) = \rho_{out}(h) K(x) \rho_{in}(h^{-1})$$

Relaxed Equivariance

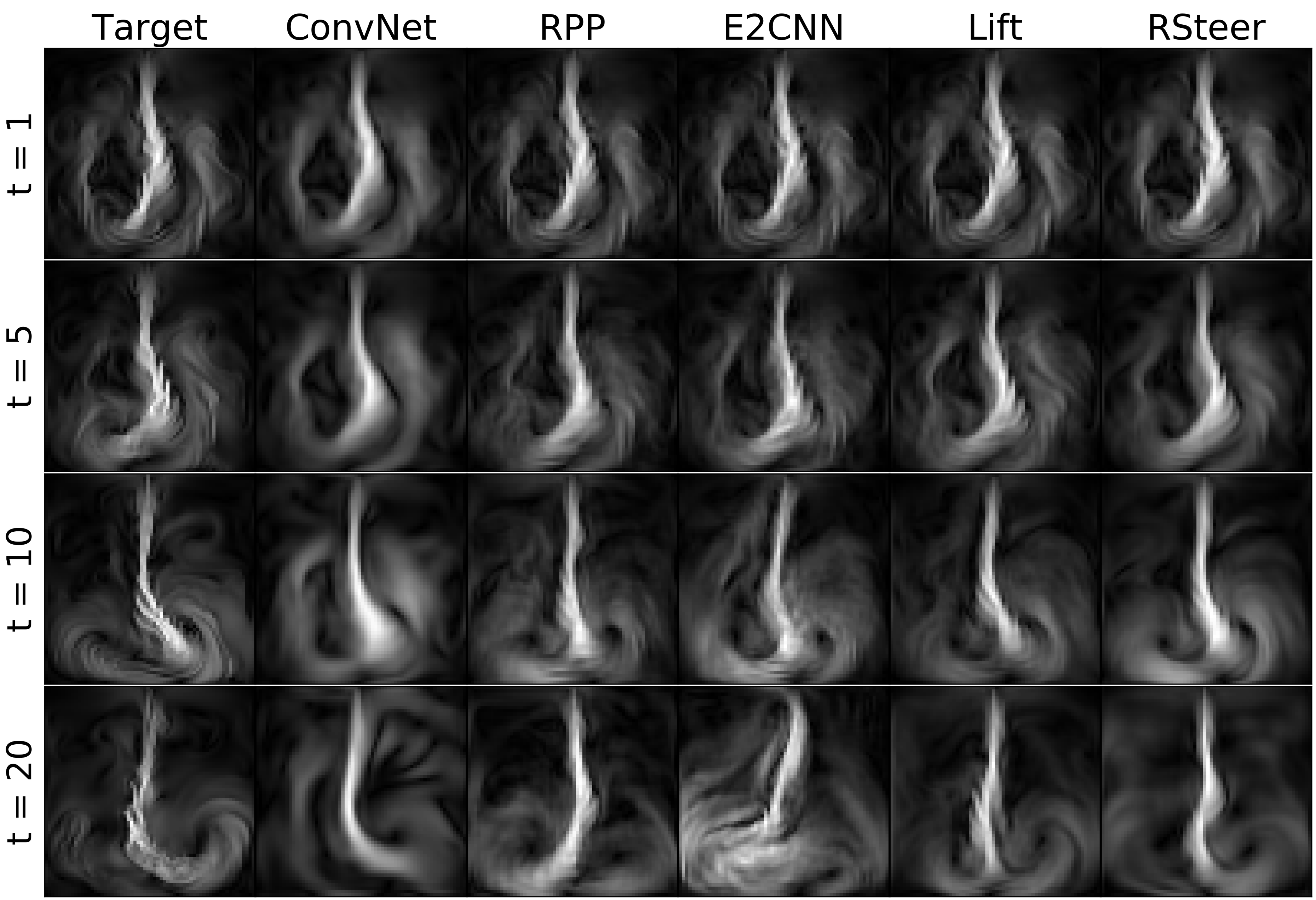
- Relaxed G-conv (**RGroup**):

$$f \tilde{*}_G K(g) = \sum_{h \in G} f(h) \sum_{l=1}^{L} w_l(h) K_l(g^{-1}h)$$

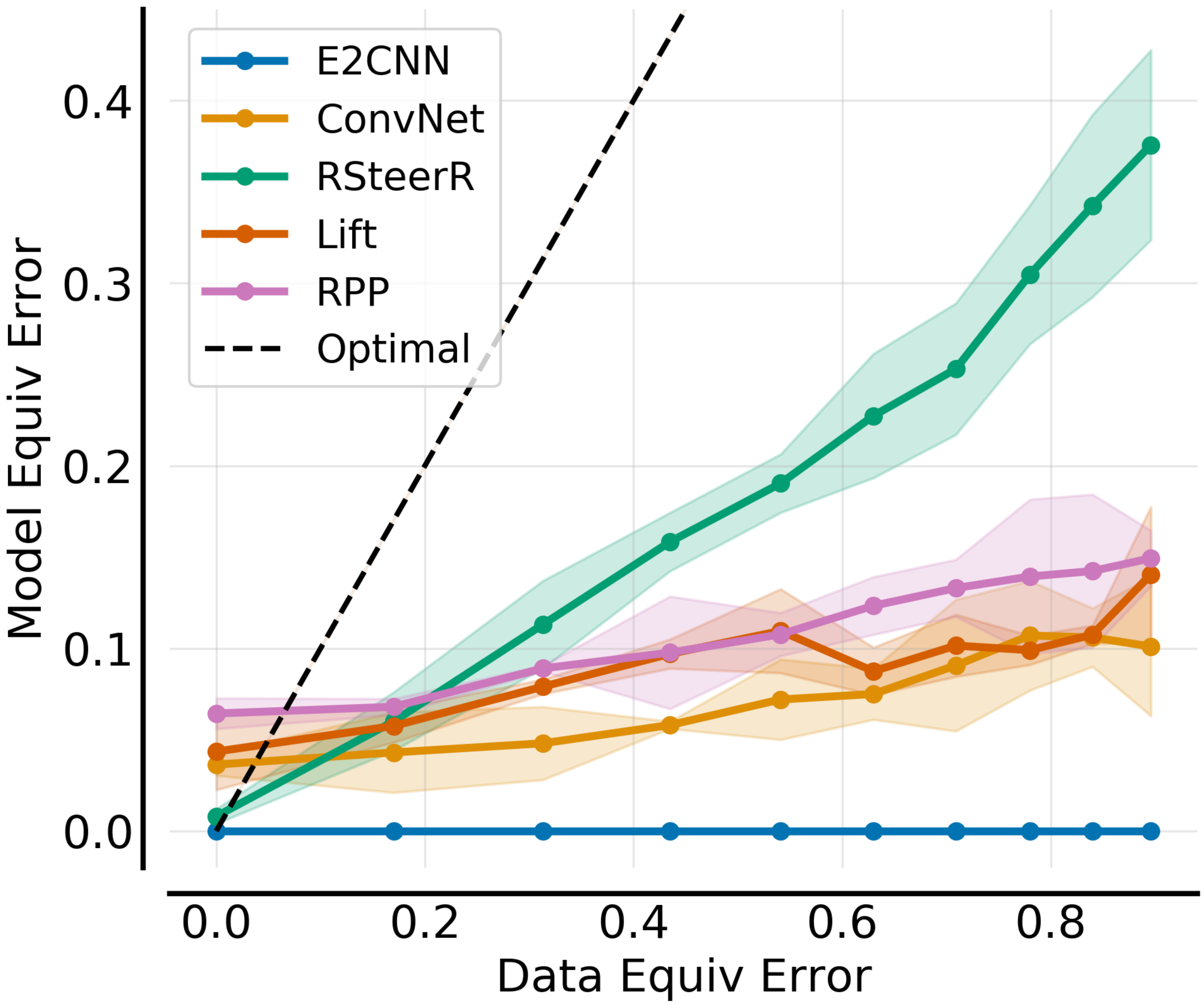
- Relaxed Steerable (**RSteer**):

$$\tilde{K}(hx) = \rho_{out}(h) \sum_{l=1}^{L} w_l(h) K_l(x) \rho_{in}(h^{-1})$$

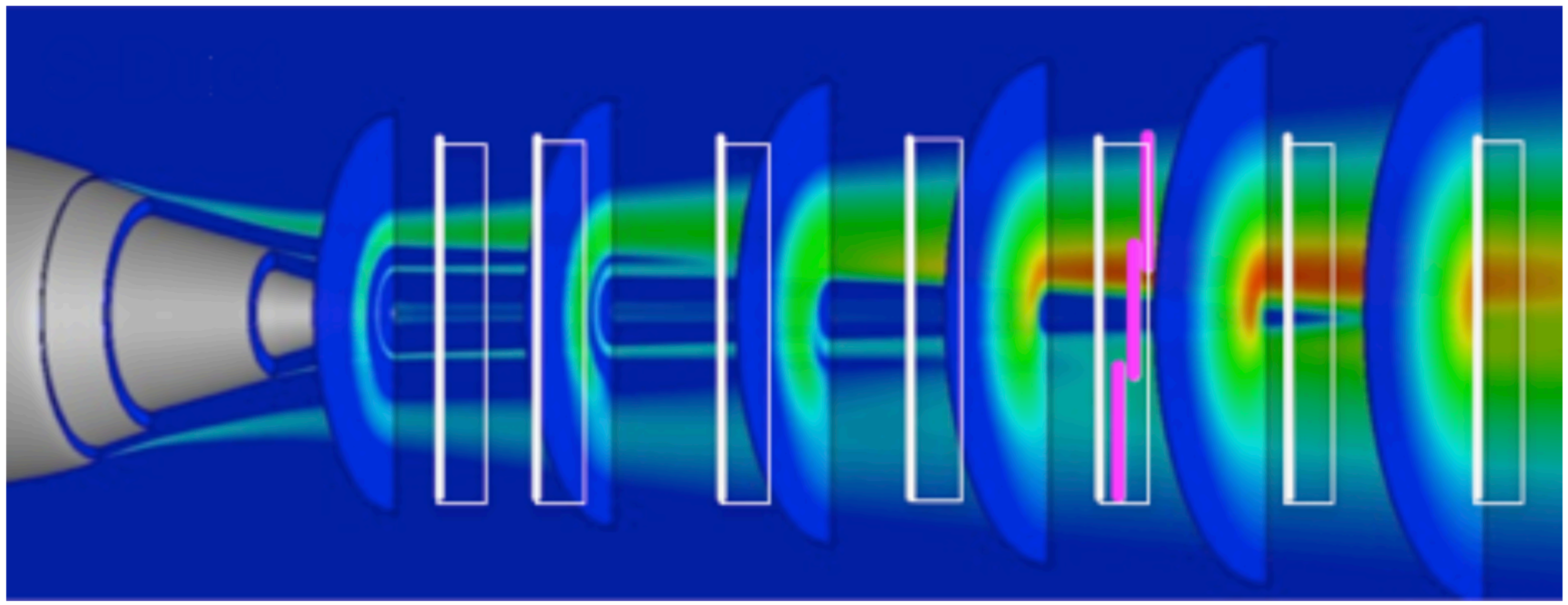
Smoke Plume



Smoke Plume Results



Supersonic Jet Flow



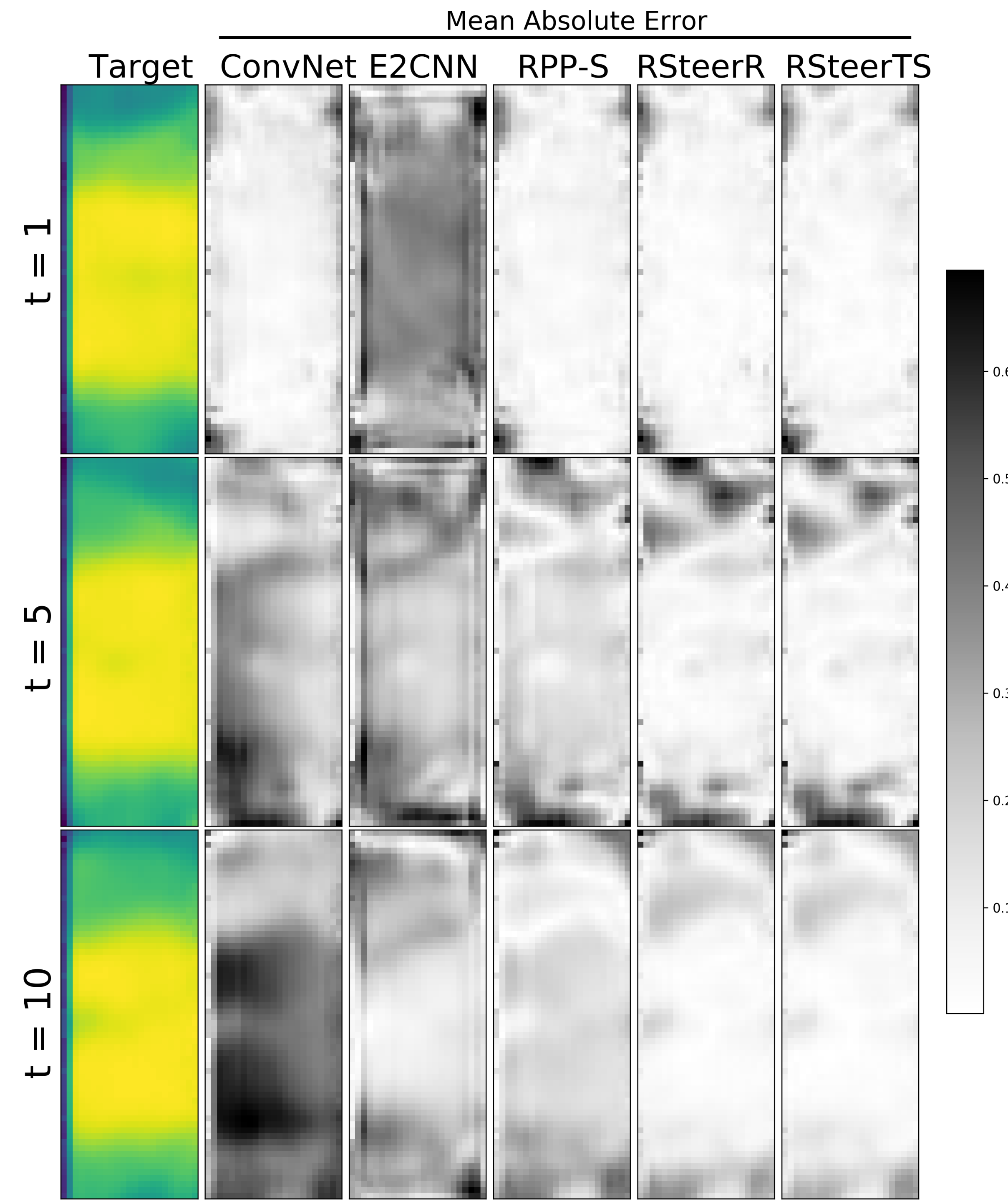
- Real experimental data of 2D turbulent velocity in multi-stream jets from NASA

- Measured using time-solved partial image velocimetry

Prediction Performance

Model	Conv	Lift	RGroup
	Translation		
Future	0.22±0.06	0.17±0.02	0.15±0.00
Domain	0.23±0.06	0.18±0.02	0.16±0.01
	E2CNN	Lift	RSteer
	Rotation		
	0.21±0.02	0.18±0.02	0.17±0.01
	0.27±0.03	0.21±0.04	0.16±0.01
	SESN	Rpp	RSteer
	Scaling		
	0.15±0.00	0.16±0.06	0.14±0.01
	0.16±0.01	0.16±0.07	0.15±0.00
	RSteerTR	RSteerTS	
	Combination		
	0.14±0.01	0.14±0.02	
	0.15±0.01	0.15±0.00	

20% better



Conclusion

- Incorporating symmetry in deep learning for learning spatiotemporal dynamics
    - **EquNet**: symmetry in differential equations
    - **Relaxed-EquNet**: approximate symmetry
- Probabilistic modeling, symmetry discovery, etc…

Acknowledgment

Open Source Code and Data: roseyu.com

@yuqirose

Equivariant Neural Networks & Symmetry Discovery

Presenter: Jianke Yang
Mar 9, 2023

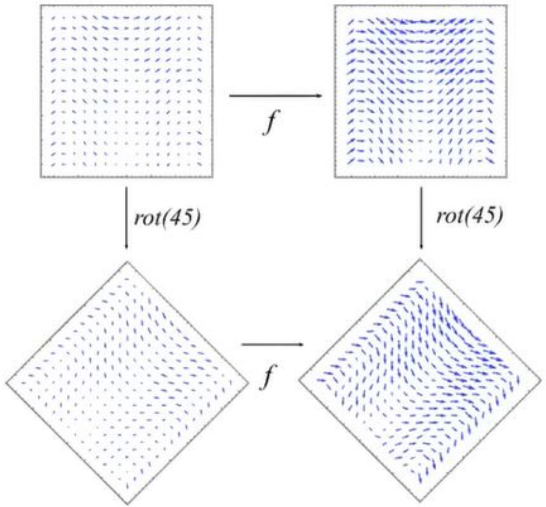
Authors: **Jianke Yang** <jiy065@ucsd.edu>, Robin Walters <r.walters@northeastern.edu>, Nima Dehmamy <Nima.Dehmamy@ibm.com>, Rose Yu <q6yu@ucsd.edu>
Paper: https://arxiv.org/abs/2302.00236

Symmetry

- **Group**: a set with an operation satisfying group axioms
  - Associativity
  - Identity element
  - Inverse elements
- Invariance & Equivariance: function and group
  - G-invariant: $f(gx) = f(x), \ \forall g \in G$
  - G-equivariant: $f(gx) = gf(x), \ \forall g \in G$

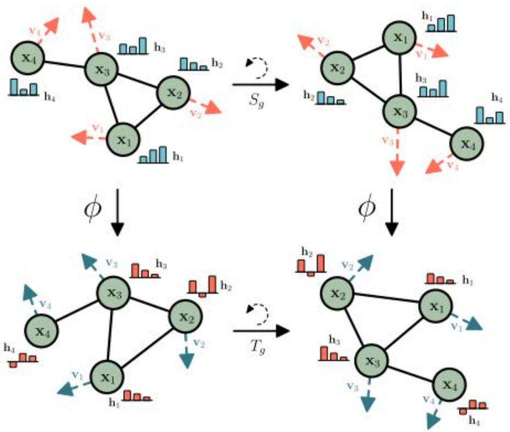
$$f(x, y) = (x, 2y)$$

$$\rho(g_\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

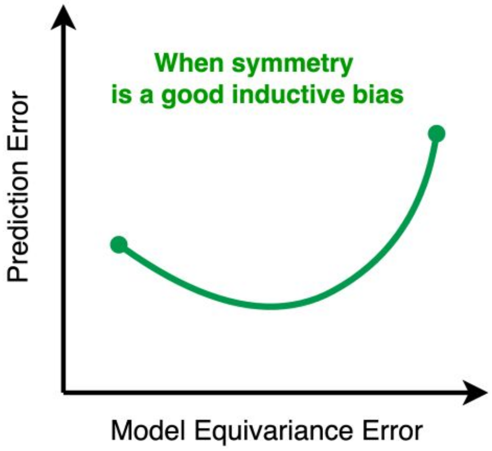


Symmetry

- Equivariance is an important inductive bias in deep learning [Bronstein et al., 2021]
- Most existing equivariant NN models require knowing the symmetry before constructing the model
    - which is sometimes unrealistic and not optimal



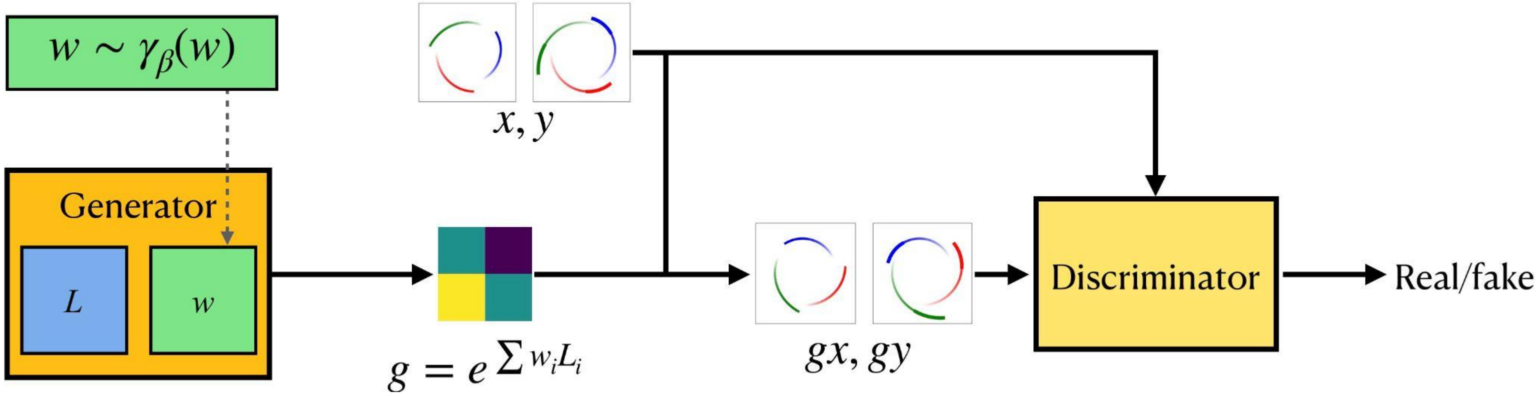
[Satorras et al., 2022]



[Wang et al., 2022]

Symmetry Discovery

- LieGAN framework



Structure of the proposed LieGAN model. 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, where the generator could learn rotation symmetry.

Symmetry Discovery

- Meta-learning Symmetries by Reparameterization [Zhou et al., 2021]

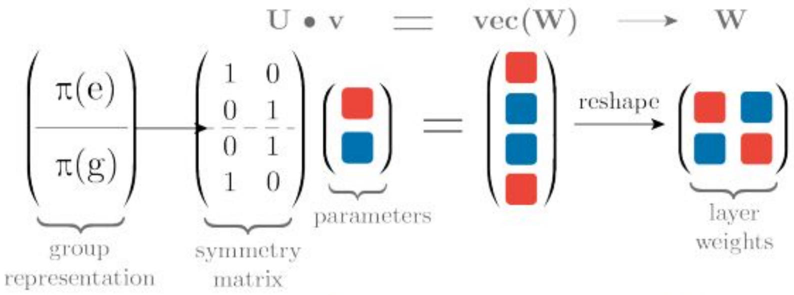


Figure 2: We reparameterize the weights of each layer in terms of a symmetry matrix U that can enforce equivariant sharing patterns of the filter parameters v. Here we show a U that enforces permutation equivariance. More technically, the layer implements group convolution on the permutation group S_2: U's block submatrices $\pi(e), \pi(g)$ define the action of each permutation on filter v. Note that U need not be binary in general.

Proposition 1 *Suppose G is a finite group $\{g_1, \ldots, g_m\}$. There exists a $U^G \in \mathbb{R}^{mn \times n}$ such that for any $v \in \mathbb{R}^n$, the layer with weights $vec(W) = U^G v$ implements G-convolution on input $x \in \mathbb{R}^n$. Moreover, with this fixed choice of U^G, any G-convolution can be represented by a weight matrix $vec(W) = U^G v$ for some $v \in \mathbb{R}^n$.*

Symmetry Discovery

- Augerino [Benton et al., 2020]: finding the extent of symmetry

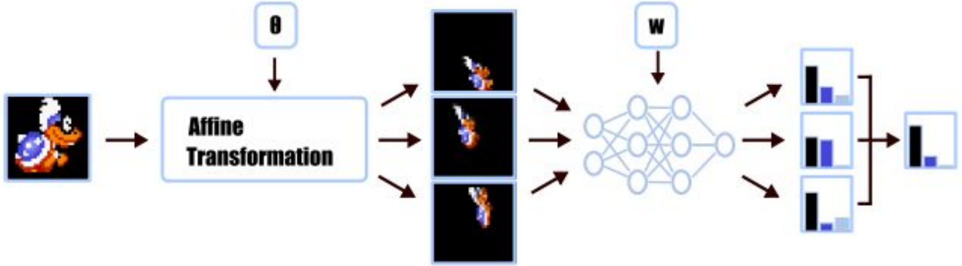
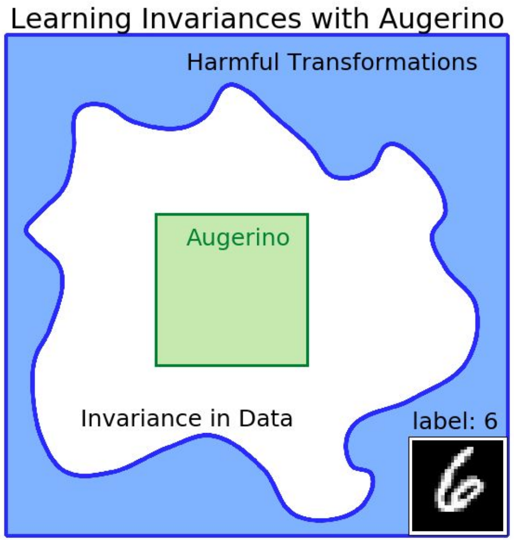


Figure 1: The Augerino framework. Augmentations are sampled from a distribution governed by parameters θ, and applied to an input to produce multiple augmented inputs. These augmented inputs are then passed to a neural network with weights w, and the final prediction is generated by averaging over the multiple outputs. Augerino discovers invariances by learning θ from training data alone.



Symmetry Discovery

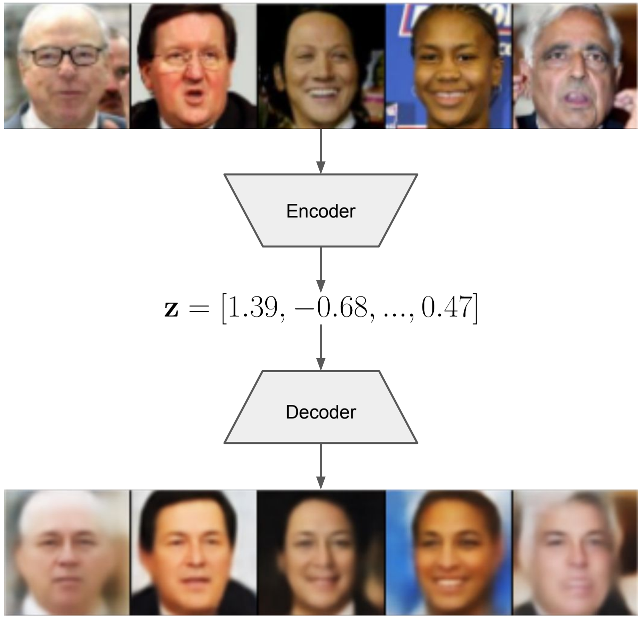
Table 1. Comparison of different models' capability of discovering different kinds of symmetries

SYMMETRY	MSR	AUGERINO	LIEGAN
DISCRETE	✓	✗	✓
CONTINUOUS	✗	✗	✓
GIVEN GROUP SUBSET	✗	✓	✓
UNKNOWN GROUP SUBSET	✗	✗	✓

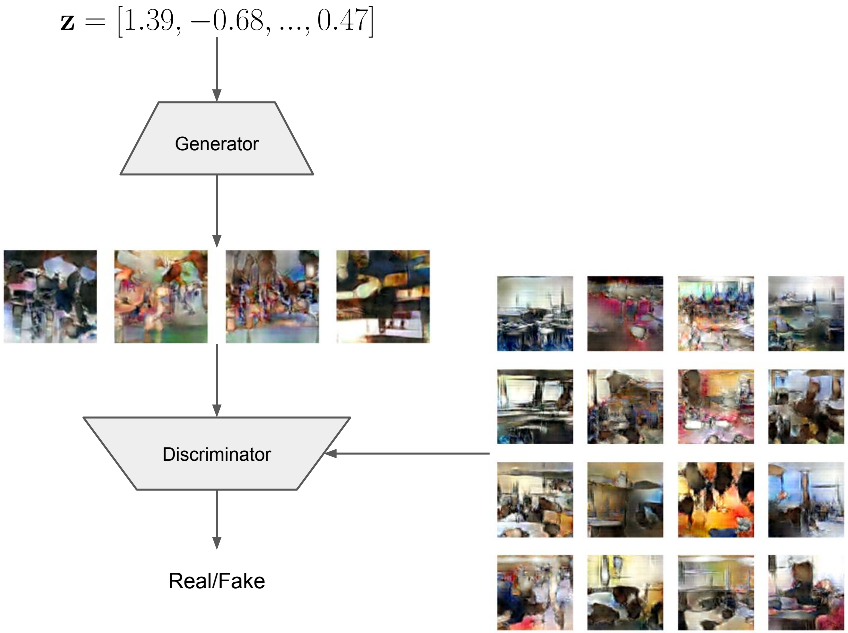
Supervised vs Unsupervised Learning

Supervised Learning	Unsupervised Learning
Input & output data	Input data
Classification Regression …	Clustering Generation …
Predictive models	Analysis & discovery

Generative Models



Variational Autoencoder



Generative Adversarial Network

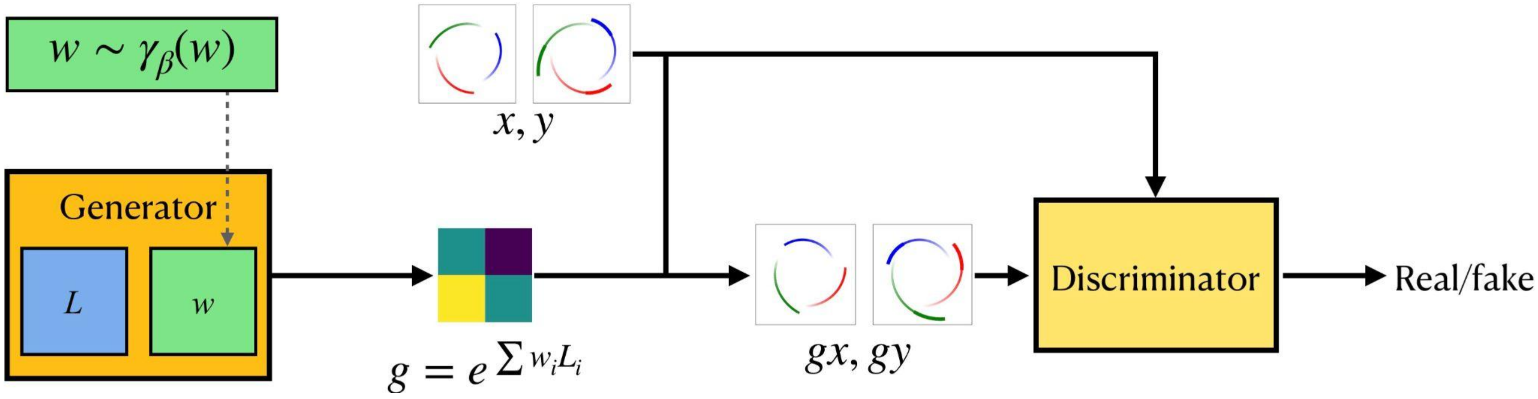
Symmetry Discovery

- Problem definition: what are we trying to discover?
    - The unknown *equivariance* property of a given dataset

Given a dataset $\mathcal{D} = \{(\mathbf{x}, \mathbf{y}) | \mathbf{x} \sim p(\mathbf{x}), \mathbf{y} = f(\mathbf{x})\} \subset \mathcal{X} \times \mathcal{Y} = \mathbb{R}^n \times \mathbb{R}^m$ with an unobserved function $f : \mathcal{X} \to \mathcal{Y}$ that maps $\mathbf{x}$ to $\mathbf{y}$, we want to discover the equivariance property of this function, that is, to find a group G acting on $\mathcal{X}$ and $\mathcal{Y}$ through linear group representations $\rho_{\mathcal{X}} : G \to GL(n)$ and $\rho_{\mathcal{Y}} : G \to GL(m)$ such that $\forall g \in G, (\mathbf{x}, \mathbf{y}) \in \mathcal{D}, \rho_{\mathcal{Y}}(g)\mathbf{y} = f(\rho_{\mathcal{X}}(g)\mathbf{x})$. We may also directly write group elements instead of their representations for simplification: $g\mathbf{y} = f(g\mathbf{x})$.

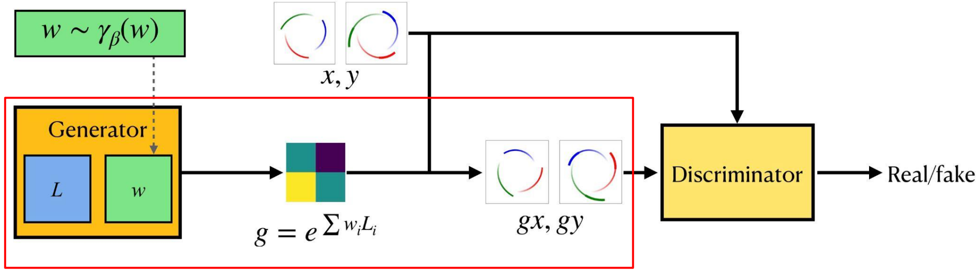
Generative Model for Symmetry Discovery

- LieGAN framework



Structure of the proposed LieGAN model. 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, where the generator could learn rotation symmetry.

Model Design



- GAN generator

$$g \sim \mu_\beta(g)$$
$$\Phi(x, y) = (\rho_{\mathcal{X}}(g)x, \rho_{\mathcal{Y}}(g)y)$$

- Loss function

$$\min_{\Phi} \max_{D} L(\Phi, D) = \mathbb{E}_{x,y\sim p_d, g\sim\mu_\beta}[\log D(x,y) + \log(1 - D(\Phi(x,y)))]$$
$$= \mathbb{E}_{x,y\sim p_d}[\log D(x, y)] + \mathbb{E}_{x,y\sim p_g}[\log(1 - D(x, y))]$$

Proof of Correctness

Theorem 1. *The generator can achieve zero JS divergence by learning a maximal subgroup* $G^* \subset \mathrm{GL}(n)$ *with respect to which* $y = f(x)$ *is equivariant if* $p_d(x)$ *is distributed proportionally to the volume of inverse group element transformation along each orbit of* G^**-action on* $\mathcal{X}$*, that is,* $p_d(gx_0) \propto |\rho_{\mathcal{X}}(g^{-1})||\rho_{\mathcal{Y}}(g^{-1})|$.

Theorem 2. *Under assumptions 1,2 and 3, the GAN loss function under the ideal discriminator* $L(\Phi, D^*)$ *is lower with a generator that learns a subspace of the true Lie algebra* $\mathfrak{g}^*$ *than a generator with an orthogonal Lie algebra to* $\mathfrak{g}^*$*. That is, if* $\mathfrak{g}_1 \cap \mathfrak{g}^* \neq \{\mathbf{0}\}$, $\mathfrak{g}_2 \cap \mathfrak{g}^* = \{\mathbf{0}\}$, *then* $L(\mathfrak{g}_1, D^*) < L(\mathfrak{g}_2, D^*) = 0$.

Model Design

- How to parameterize a distribution over a matrix group?
- Consider the following density function:

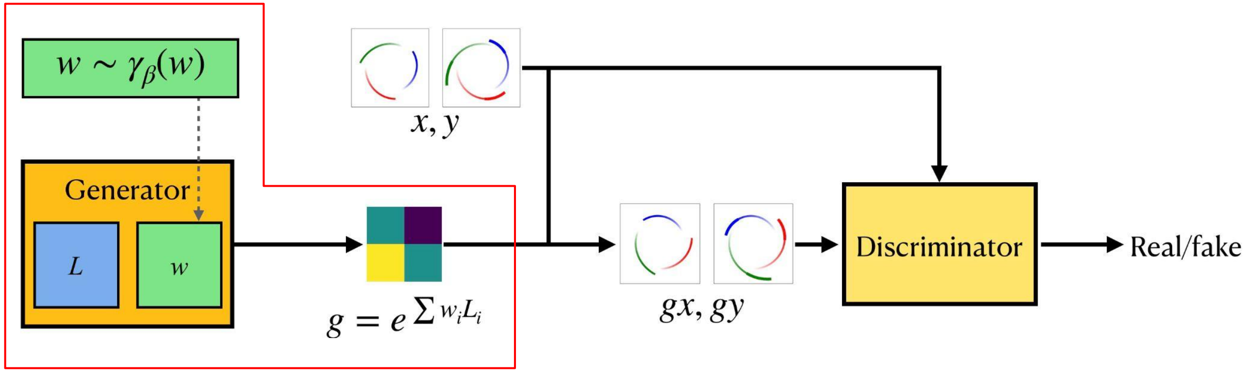
$$p\left(\begin{bmatrix}2 & 0\\ 0 & 1\end{bmatrix}\right) = 1,\ p\left(\begin{bmatrix}4 & 0\\ 0 & 1\end{bmatrix}\right) = 0$$

Arbitrary distribution over general linear group may not respect the group axioms!

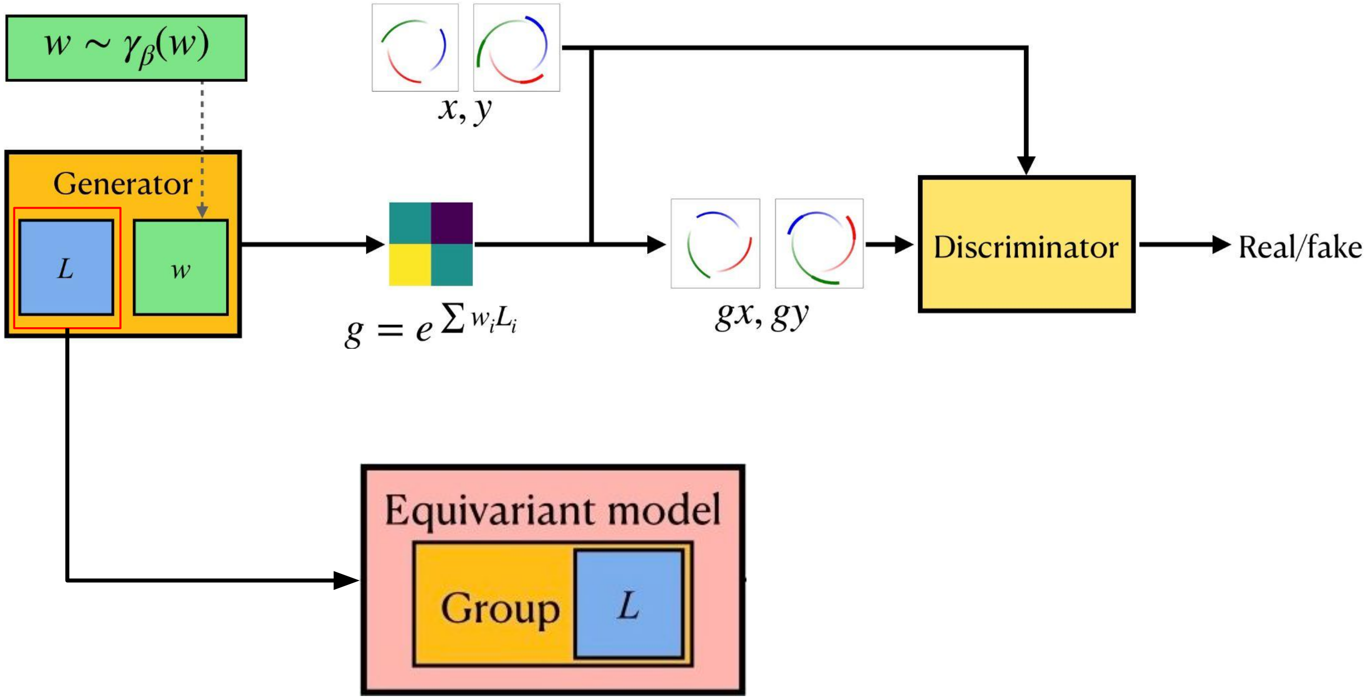
Symmetry Discovery

- Parameterizing the distribution over continuous Lie group

$$g \sim \mu_\beta(g) \longrightarrow w \sim \gamma_\beta(w), \quad g = \exp\left[\sum_i w_i L_i\right]$$



Symmetry Discovery for Prediction



Example: Equivariant GNN

- E-GNN enforces E(n) symmetry by invariant features [Satorras et al, 2022]

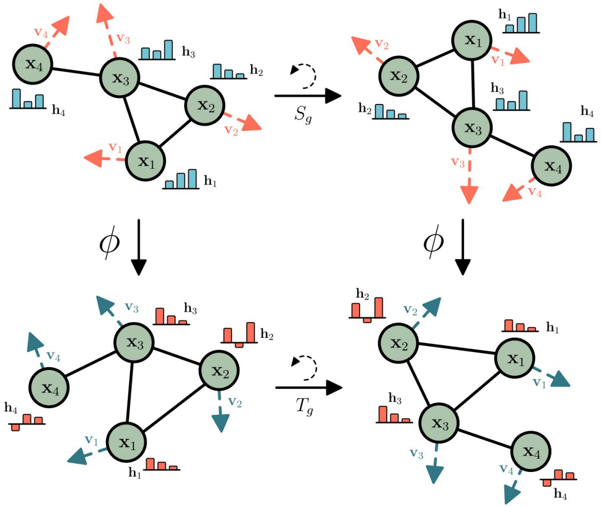


Figure 1. Example of rotation equivariance on a graph with a graph neural network ϕ

EGNN
$\mathbf{m}_{ij} = \phi_e(\mathbf{h}_i^l, \mathbf{h}_j^l, \Vert\mathbf{r}_{ij}^l\Vert^2, a_{ij})$ $\hat{\mathbf{m}}_{ij} = \mathbf{r}_{ij}^l \phi_x(\mathbf{m}_{ij})$
$\mathbf{m}_i = \sum_{j \neq i} \mathbf{m}_{ij}$ $\hat{\mathbf{m}}_i = C \sum_{j \neq i} \hat{\mathbf{m}}_{ij}$
$\mathbf{h}_i^{l+1} = \phi_h\left(\mathbf{h}_i^l, \mathbf{m}_i\right)$ $\mathbf{x}_i^{l+1} = \mathbf{x}_i^l + \hat{\mathbf{m}}_i$
E(n)-Equivariant

Example: Equivariant GNN

- Replacing E(n) with discovered symmetry from LieGAN
- Use invariant feature

$$m_{ij} = \phi_e(h_i, h_j, \|x_i - x_j\|_J^2, \langle x_i, x_j \rangle_J)$$

$$\text{where } \|u\|_J = \sqrt{u^T J u}, \quad \langle u, v \rangle_J = u^T J v$$

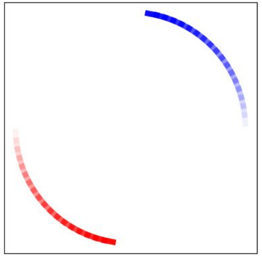
- Compute general group invariant metric tensor

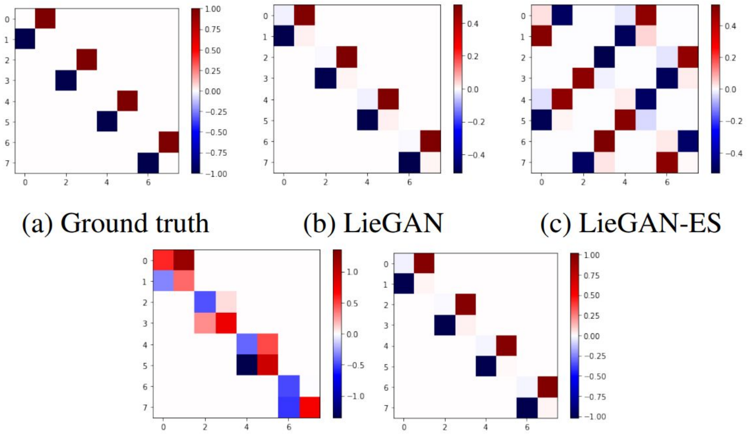
Proposition 1. *Given a Lie algebra basis $\{L_i \in \mathbb{R}^{k\times k}\}_{i=1}^c$, $\eta(u, v) = u^T J v$ $(u, v \in \mathbb{R}^k, J \in \mathbb{R}^{k\times k})$ is invariant to infinitesimal transformations in the Lie group G generated by $\{L_i\}_{i=1}^c$ if and only if $L_i^T J + J L_i = 0$ for $i = 1, 2, ..., c$.*

$$\arg\min_J \sum_{i=1}^{c} \|L_i^T J + J L_i\|^2 - a \cdot \|J\|^2$$

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.

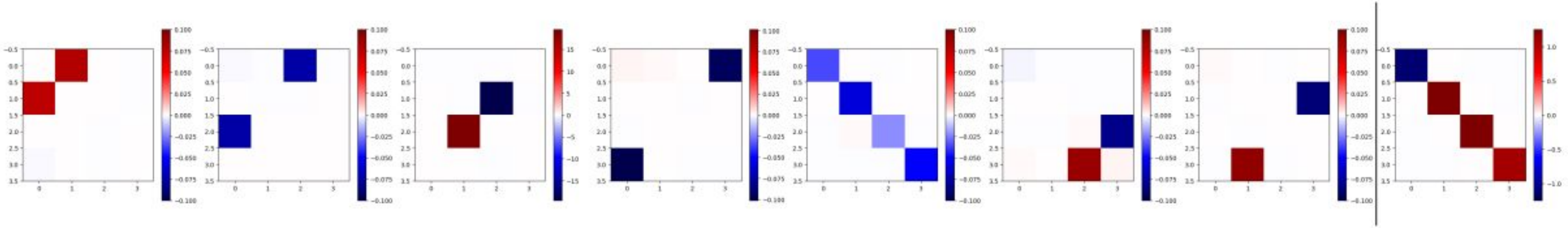
Predicting 2-Body Trajectory

- 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))



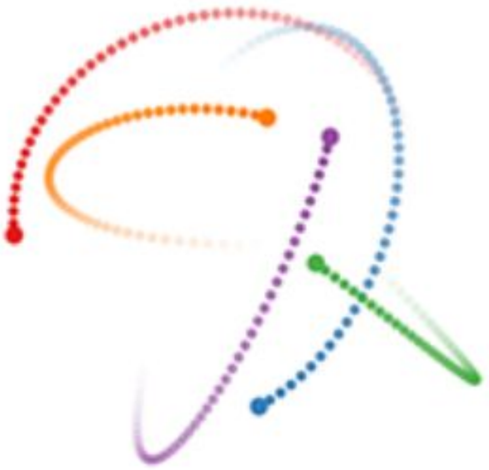
- Left: LieGAN discovers an approximate restricted Lorentz group symmetry
- Right: Computed invariant metric of the discovered symmetry

Top Quark Tagging

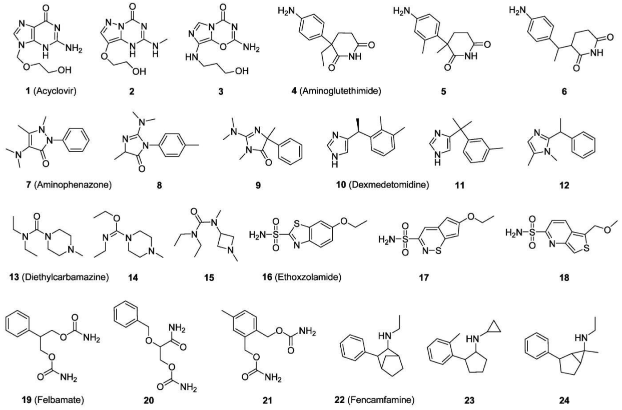
- Test accuracy and AUROC for top tagging
- LieGNN reaches the performance with LorentzNet which explicitly encodes Lorentz symmetry

Model	Accuracy	AUROC
LorentzNet	**0.940**	**0.9857**
LieGNN	0.938	0.9848
LorentzNet (w/o)	0.934	0.9832
EGNN	0.922	0.9760

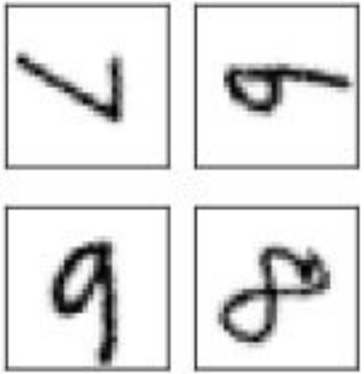
Applications



Dynamics



Molecules



Vision

Conclusion

- Discover general linear symmetries from data with LieGAN
- Interpretable Lie algebra basis as discovery result
- Larger search space than previous works
- Pipeline for utilizing discovered symmetry to downstream prediction tasks
- Scientific discovery with machine learning

Thank you!

Contact: Jianke Yang <jiy065@ucsd.edu>