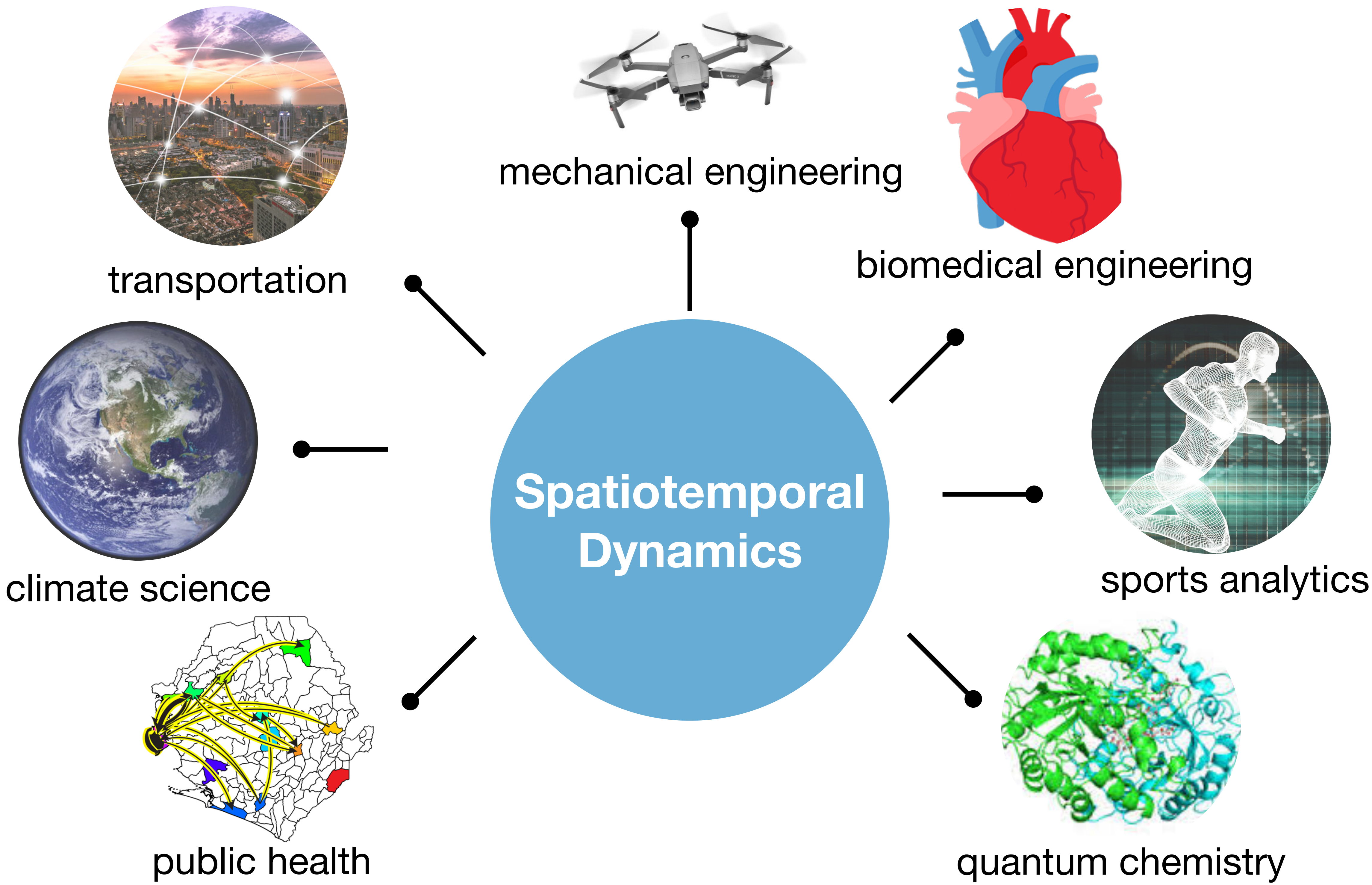
UC San Diego

Incorporating Symmetry for Learning Spatiotemporal Dynamics

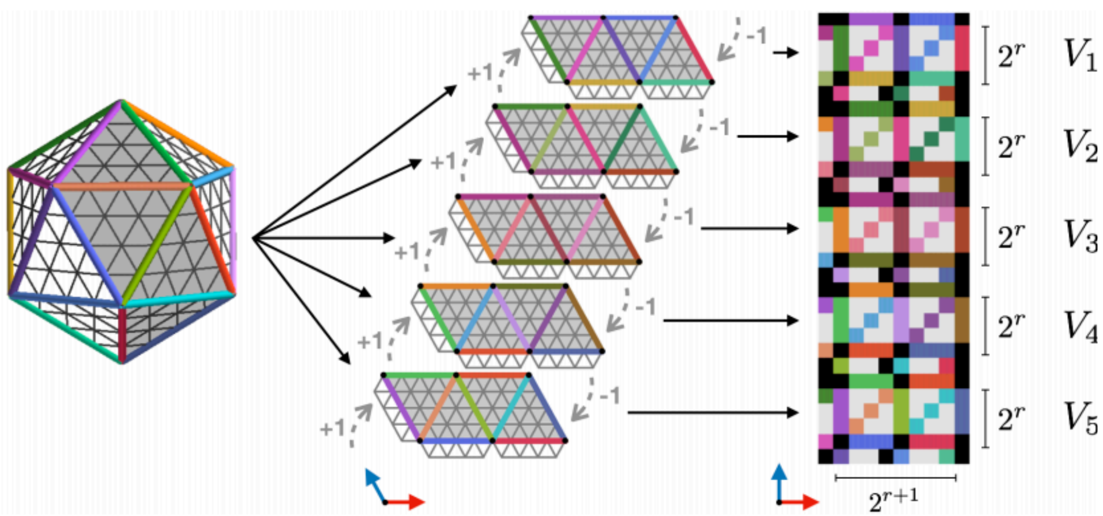
Rose Yu

Assistant Professor
University of California, San Diego

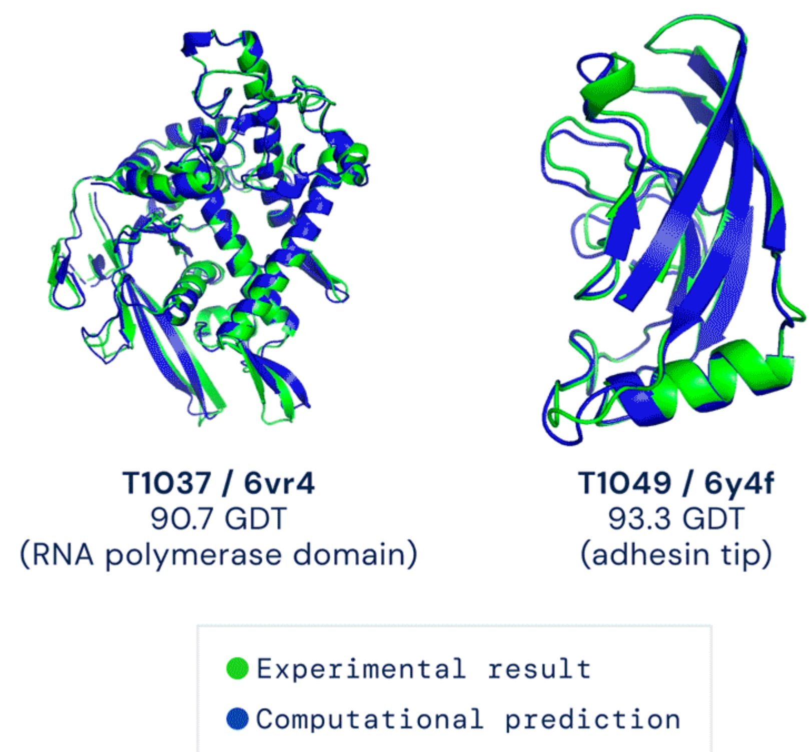
Learning Spatiotemporal Dynamics



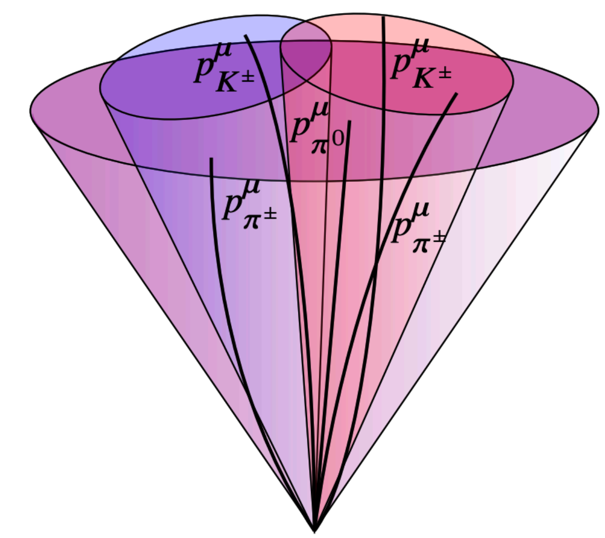
Success of Symmetry



Cohen et al 2019



Jumper et al 2021

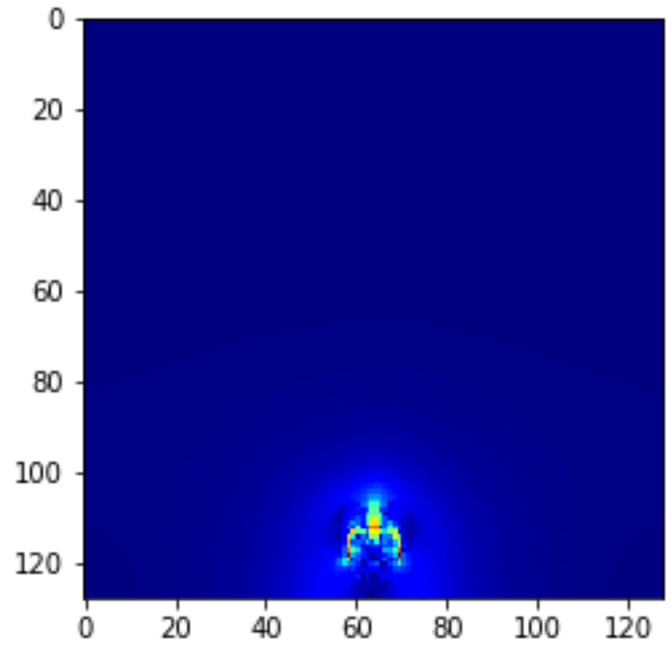


Bogatskiy et al 2020

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

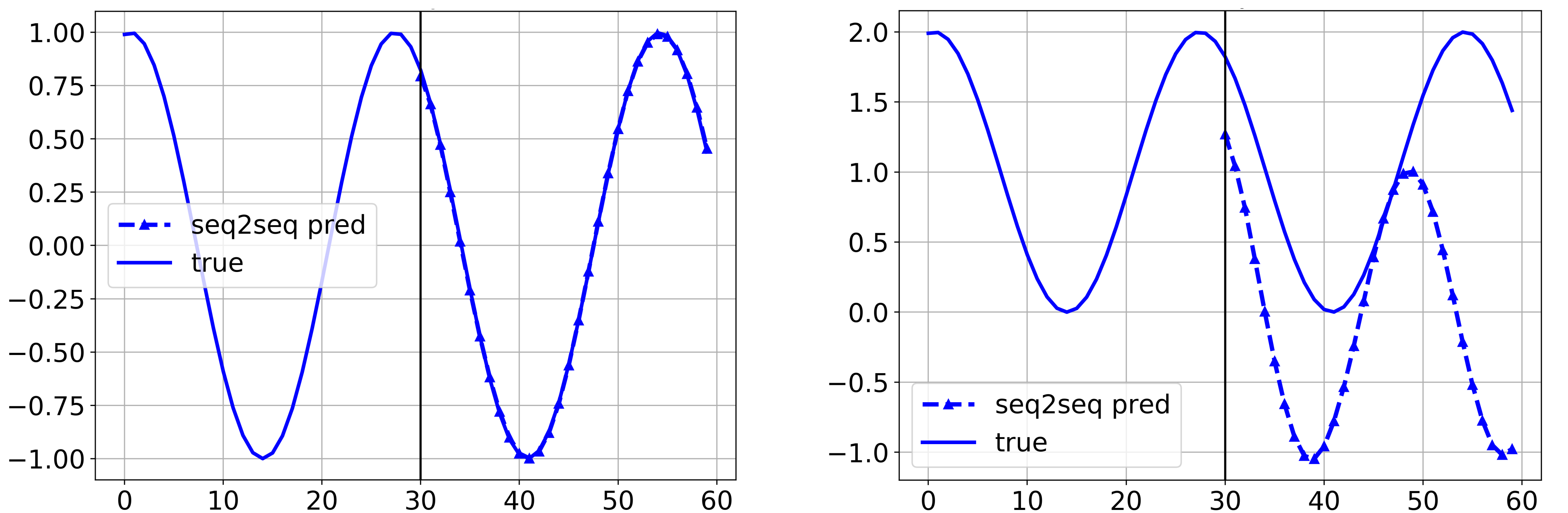


Rui Wang

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



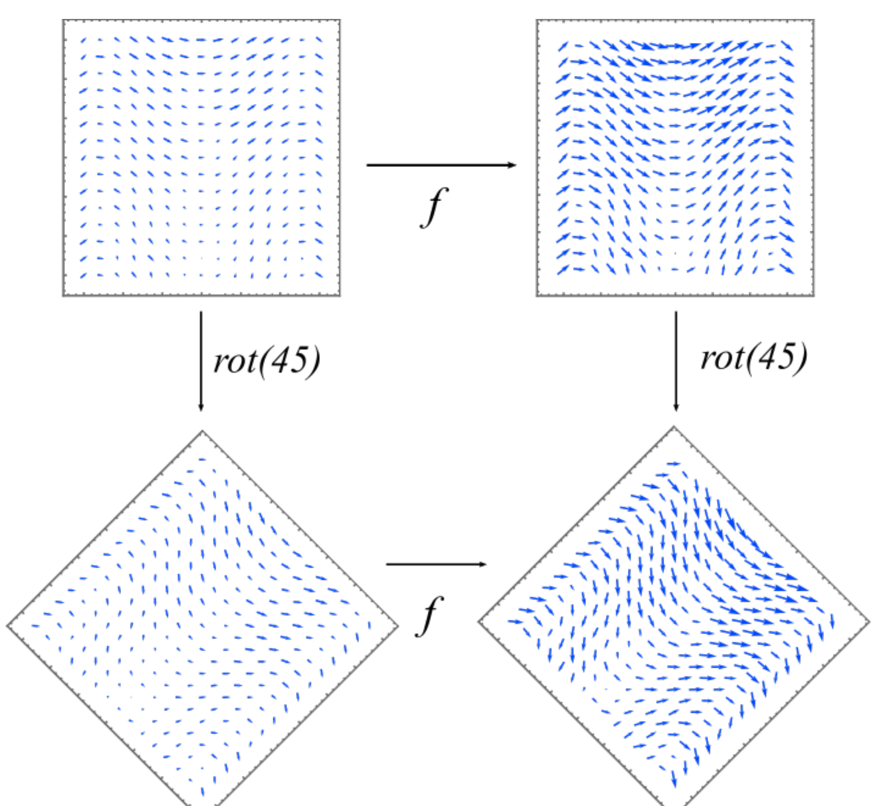
$\theta^{\textbf{tr}} \in [a, b]$ $\theta^{\textbf{te}} \in [a, b]$ $\theta^{\textbf{tr}} \in [a, b]$ $\theta^{\textbf{te}} \in [b, c]$

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

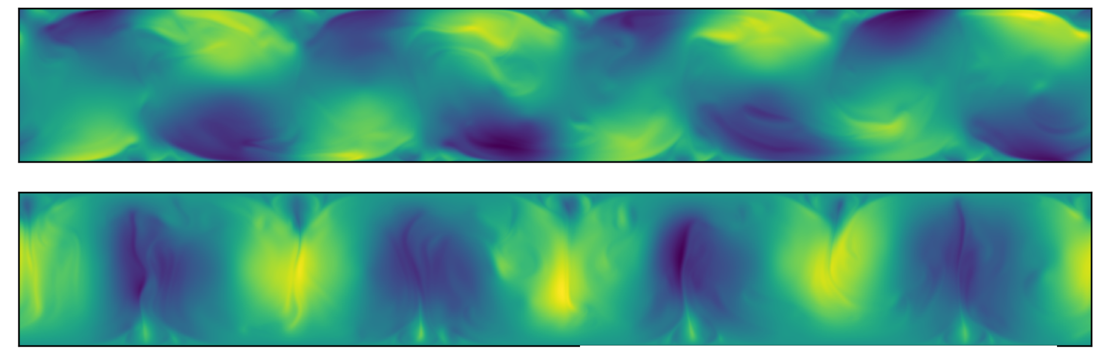
Symmetry in Dynamical Systems

- 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**

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

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

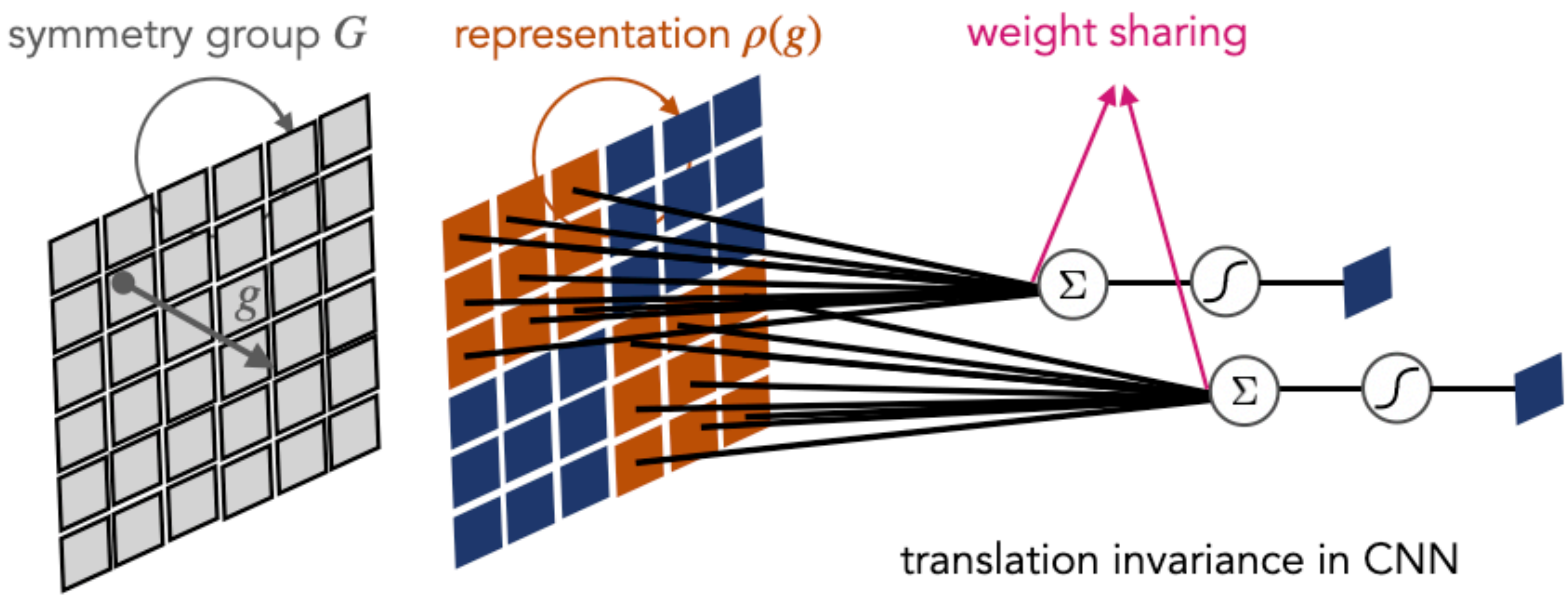
$$\frac{\partial T}{\partial t} + (\mathbf{w} \cdot \nabla)T = \kappa \nabla^2 T$$



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	$\lambda \boldsymbol{w}(\lambda \boldsymbol{x}, \lambda^2 t)$

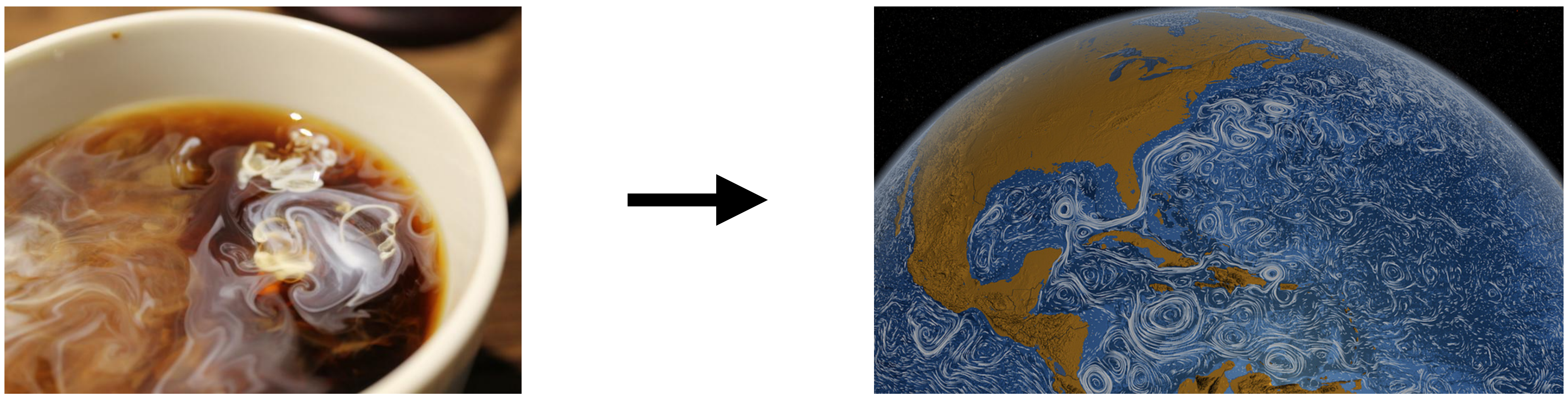
Scaling Law

Weight Symmetry



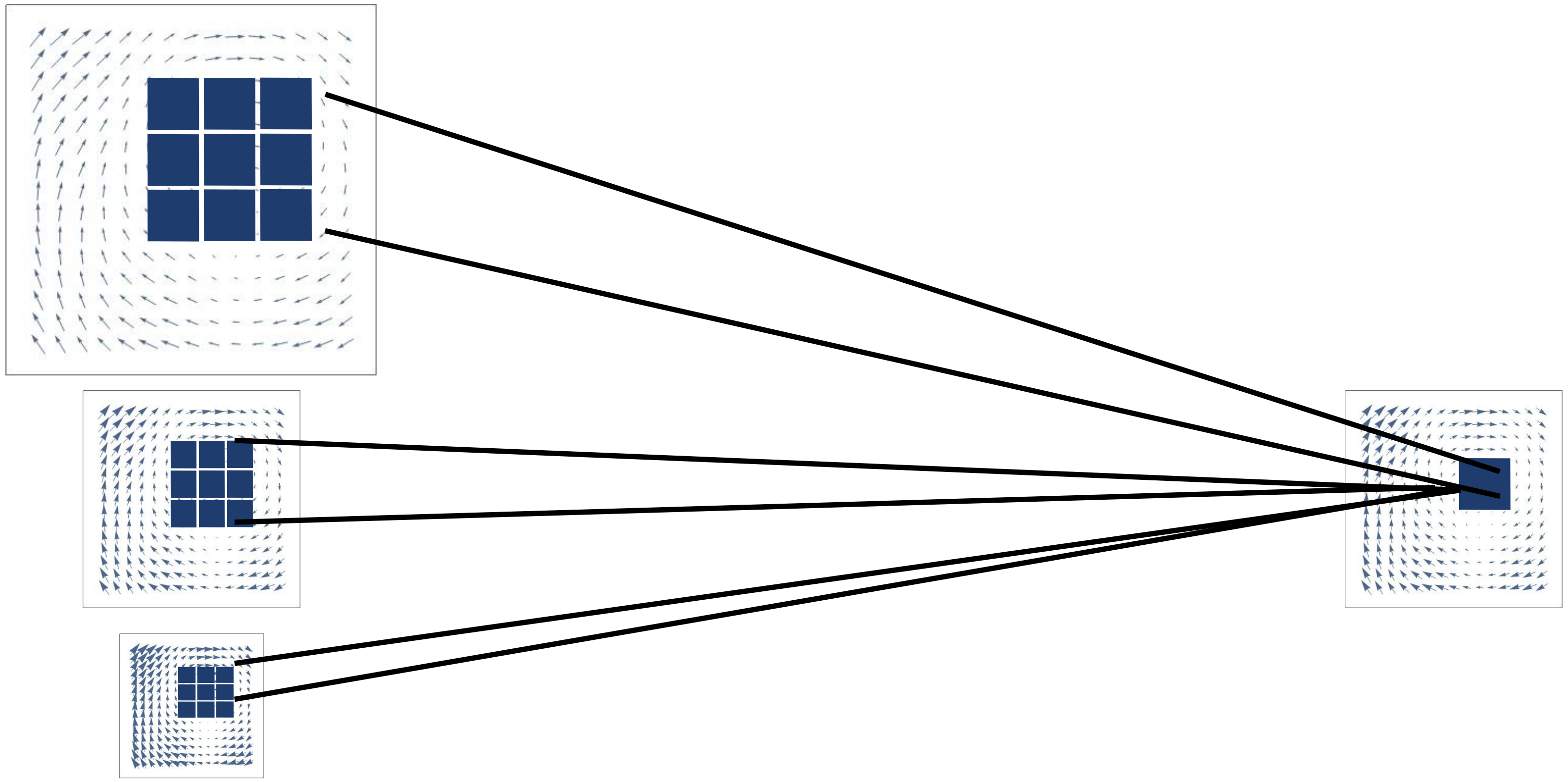
Theorem (Weiler & Cesa 2019): a convolutional layer is G-equivariant if and only if the kernel satisfies $K(gv) = \rho_{out}^{-1}(g)K(v)\rho_{in}(g)$ for all $g \in G$, with action maps ρ_{in} and ρ_{out} .

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

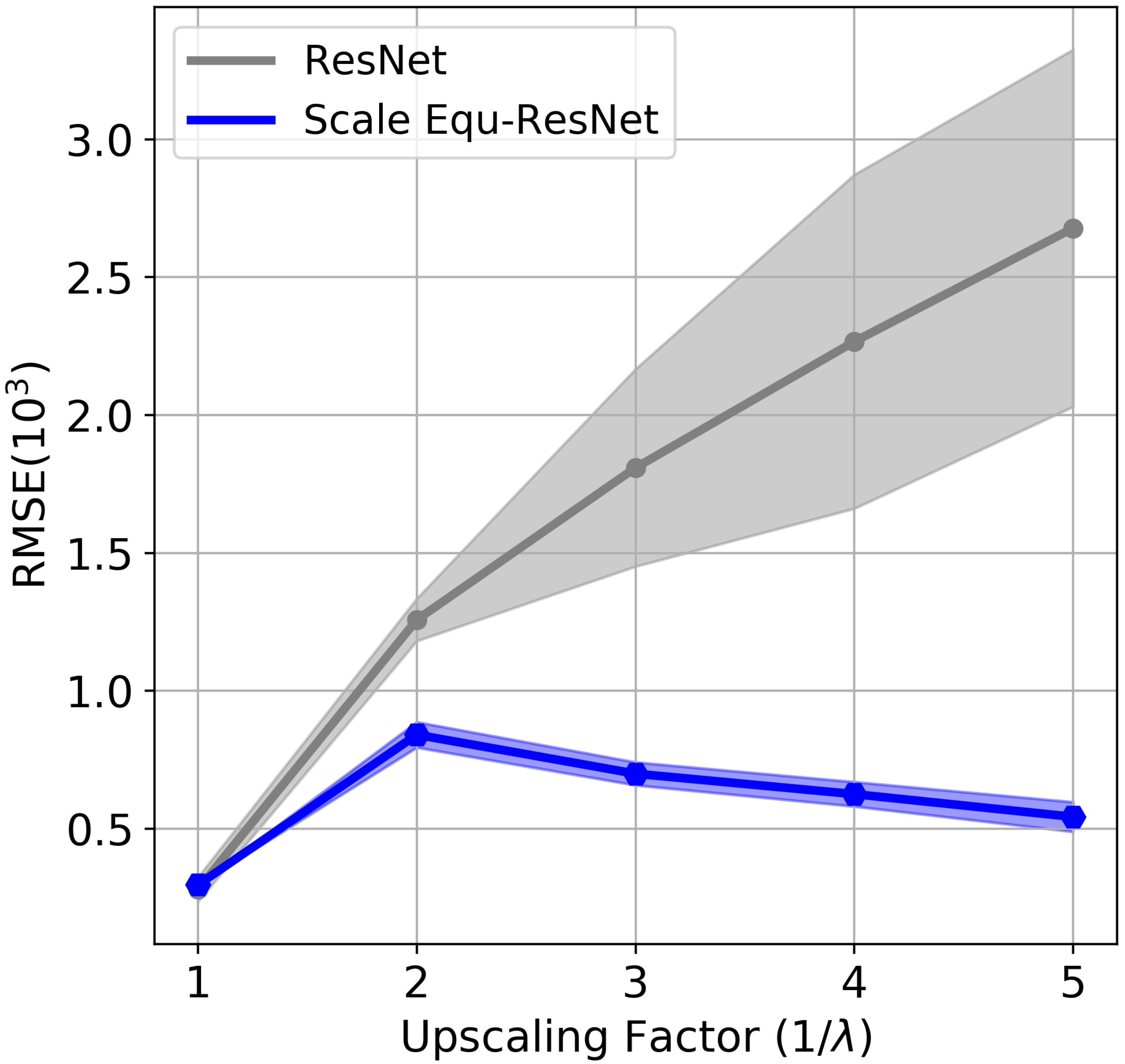


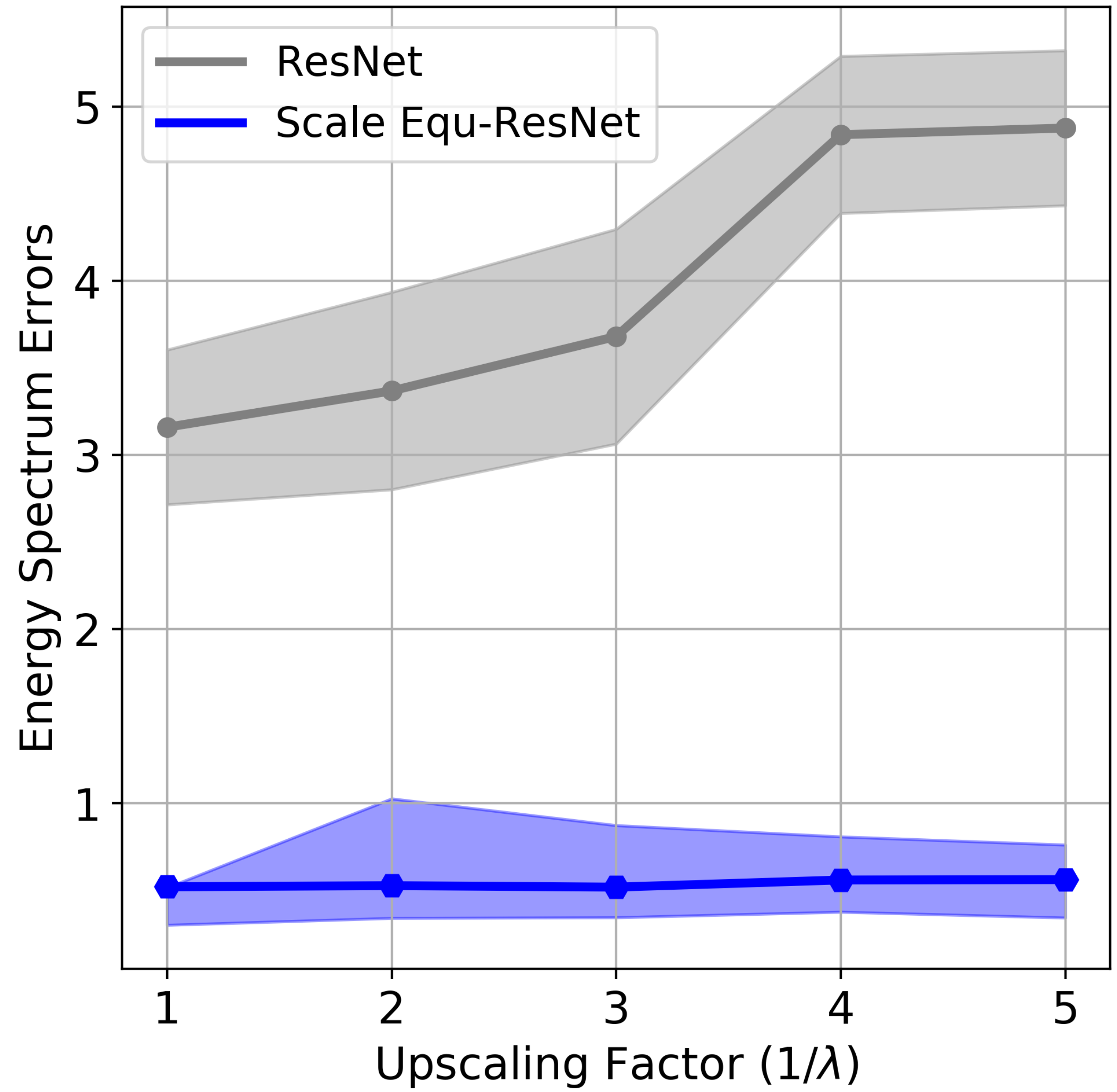
- *Scale* equivariant

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

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

Ocean Currents Forecast





Physically Consistent Predictions!

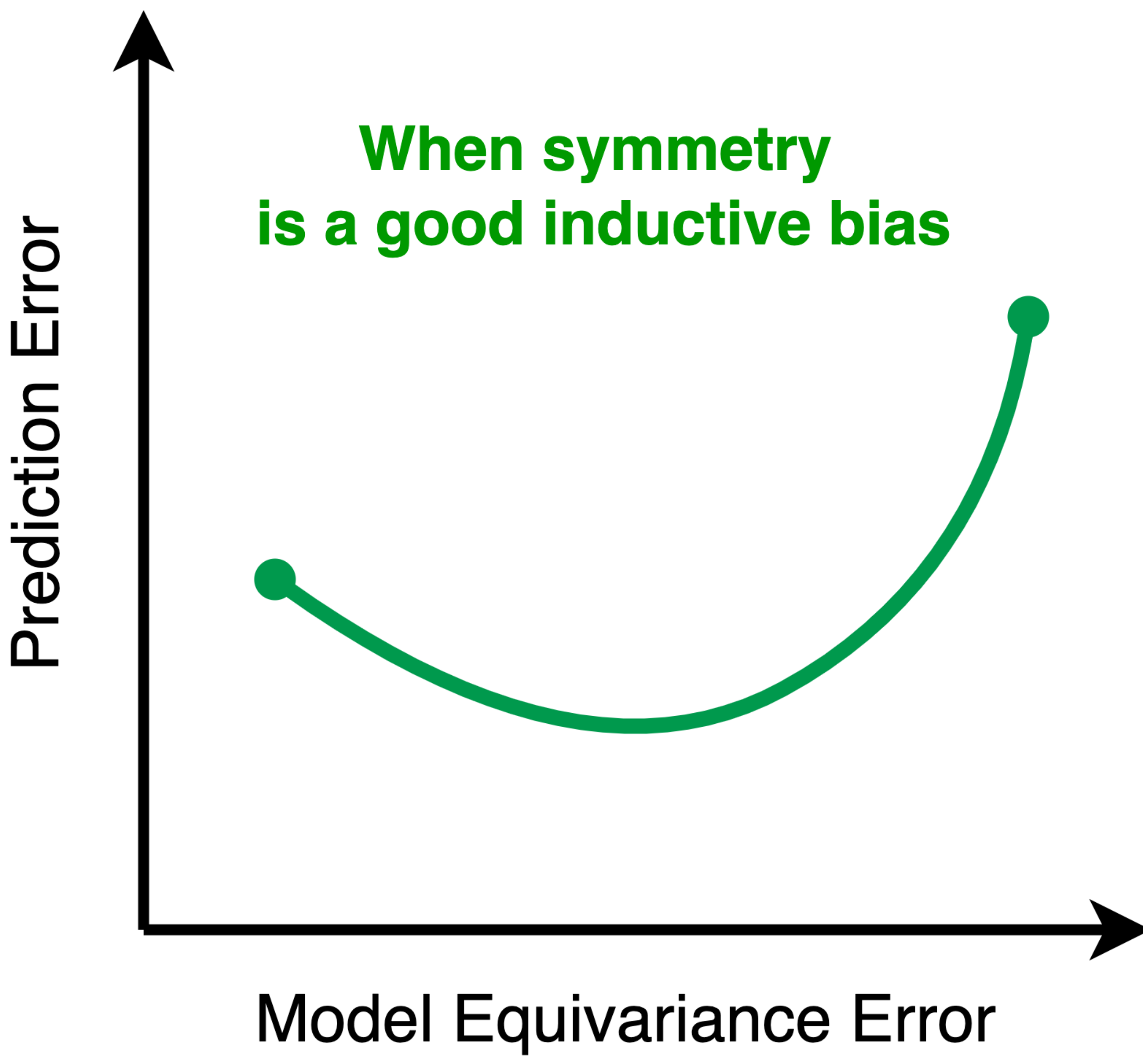
Approximately Equivariant Networks

Rui Wang

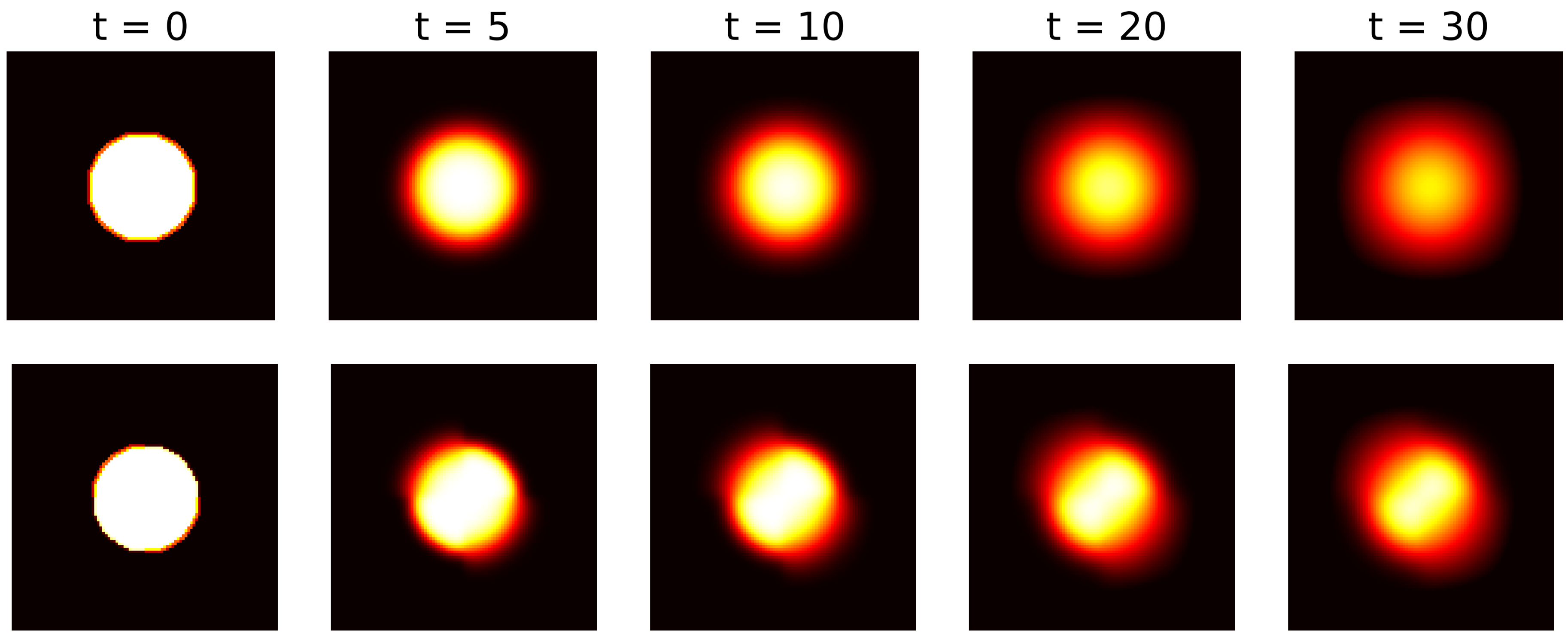
Robin Walters

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

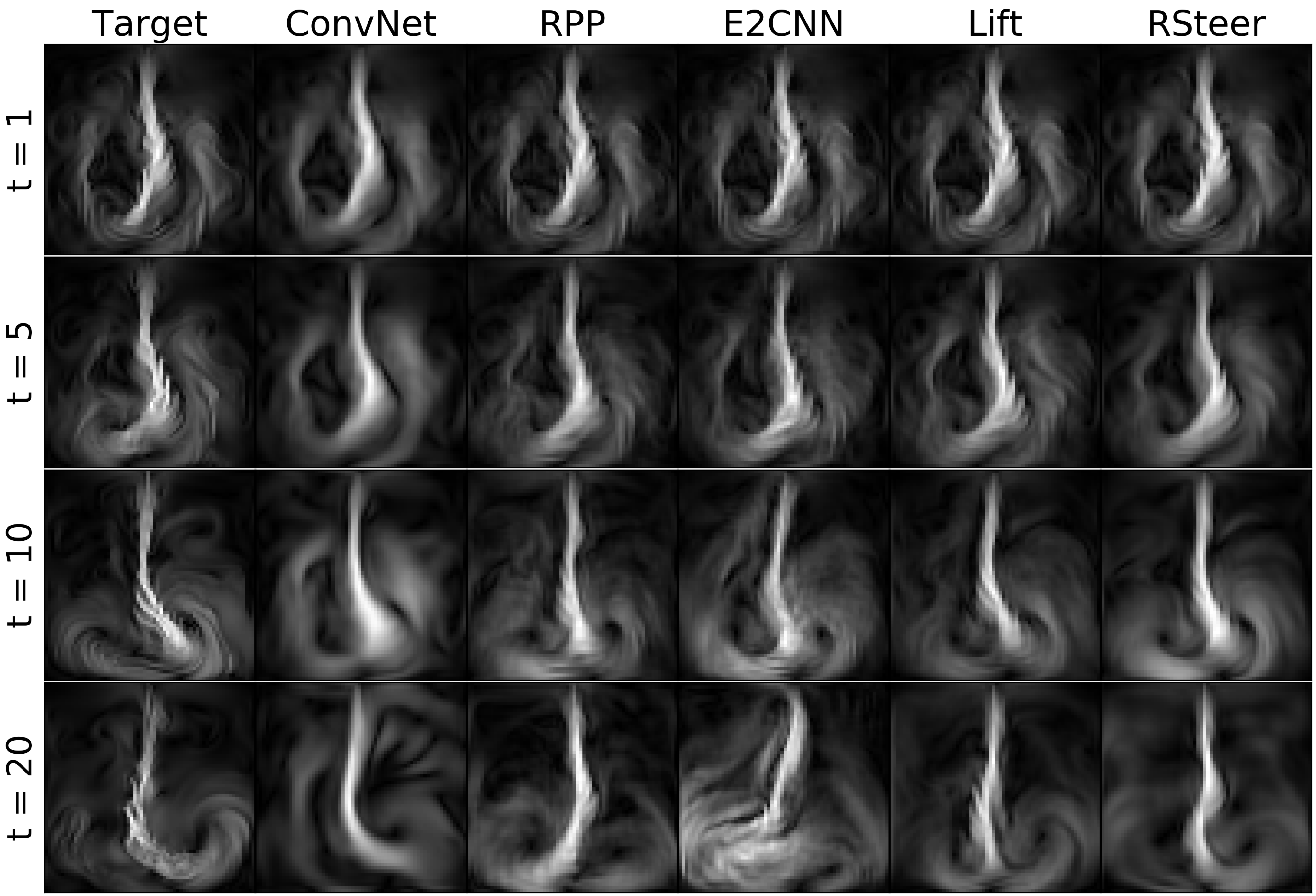
- Relaxed G-conv (**RGroup**):

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

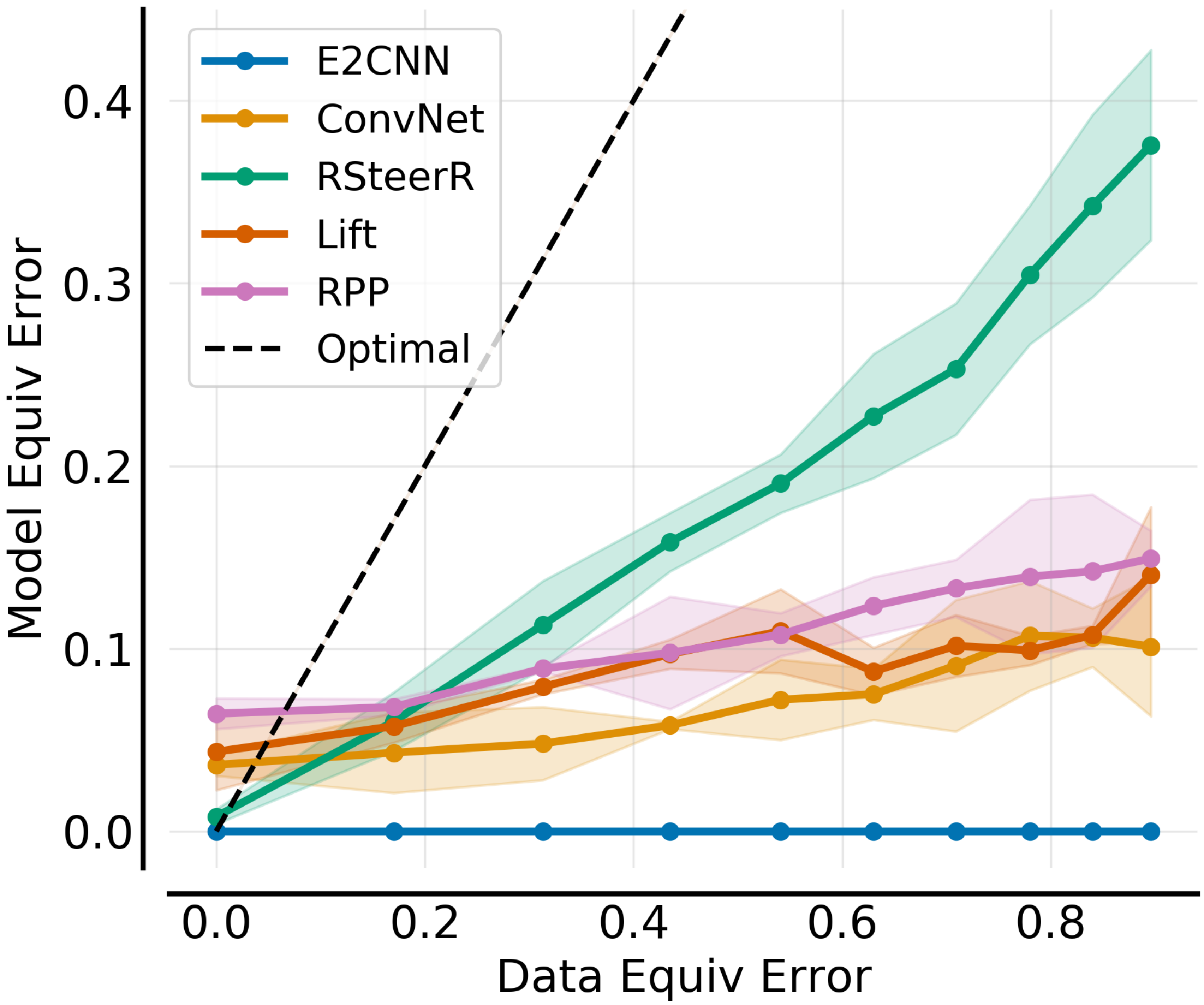
- Relaxed Steerable (**RSteer**):

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

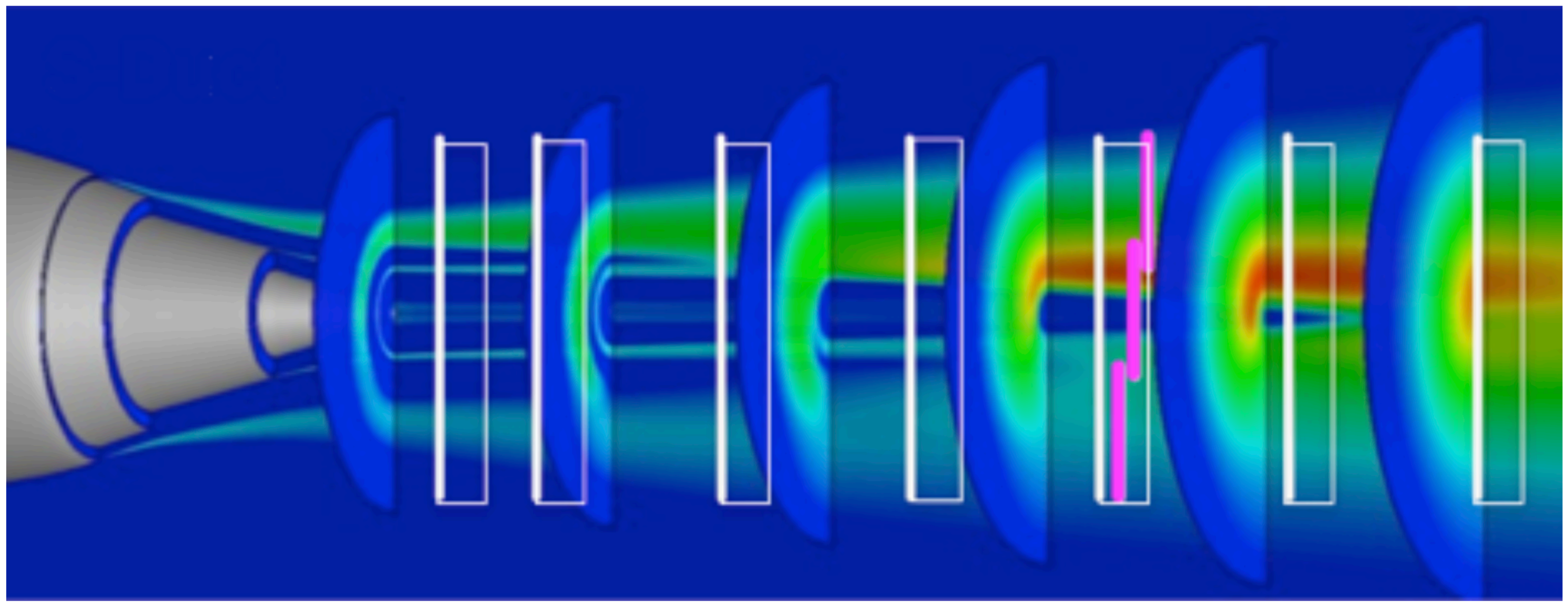
Smoke Plume



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

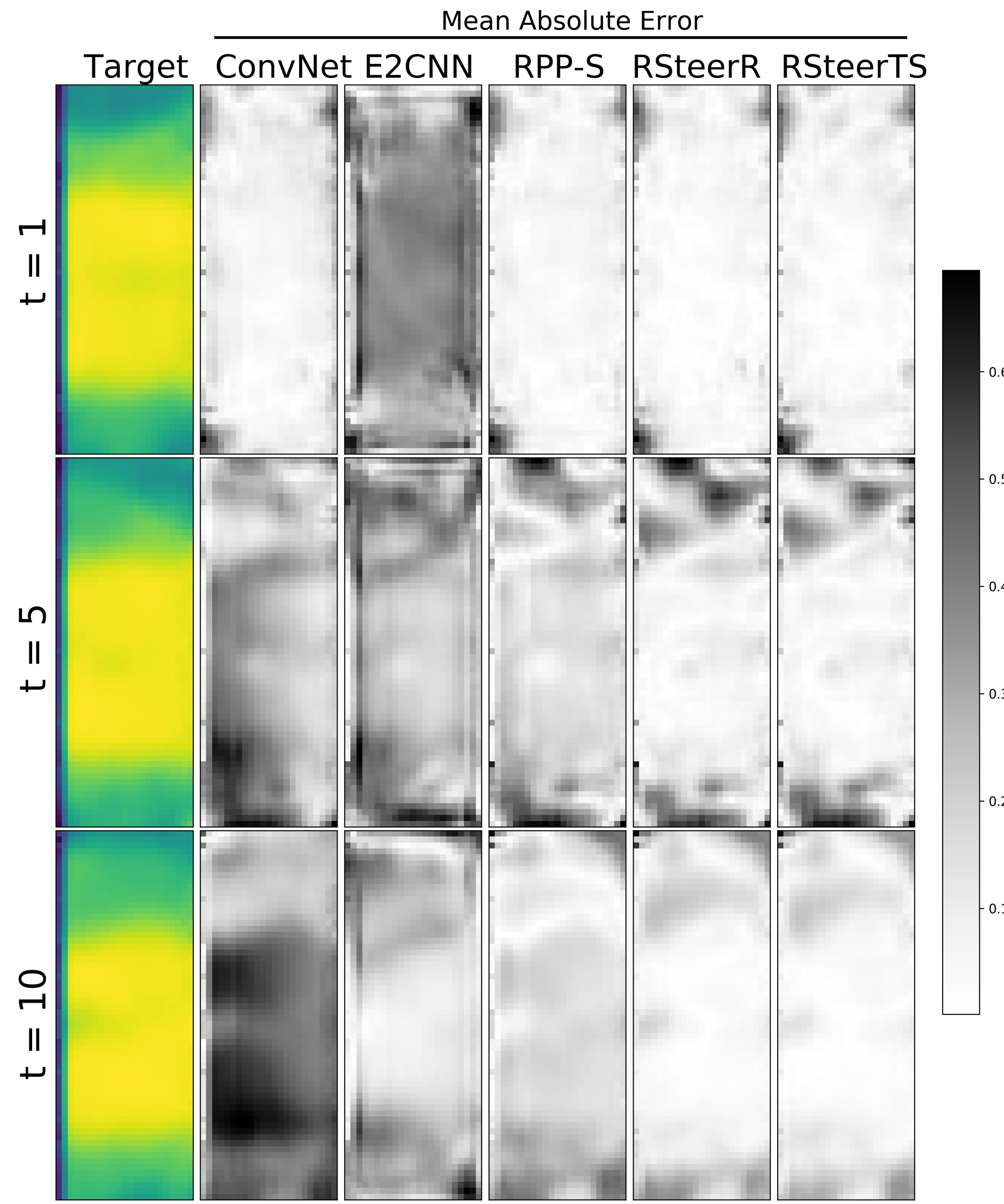
Model	Conv	Lift	RGroup
	Translation		
Future	0.22±0.06	0.17±0.02	0.15±0.00
Domain	0.23±0.06	0.18±0.02	0.16±0.01

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

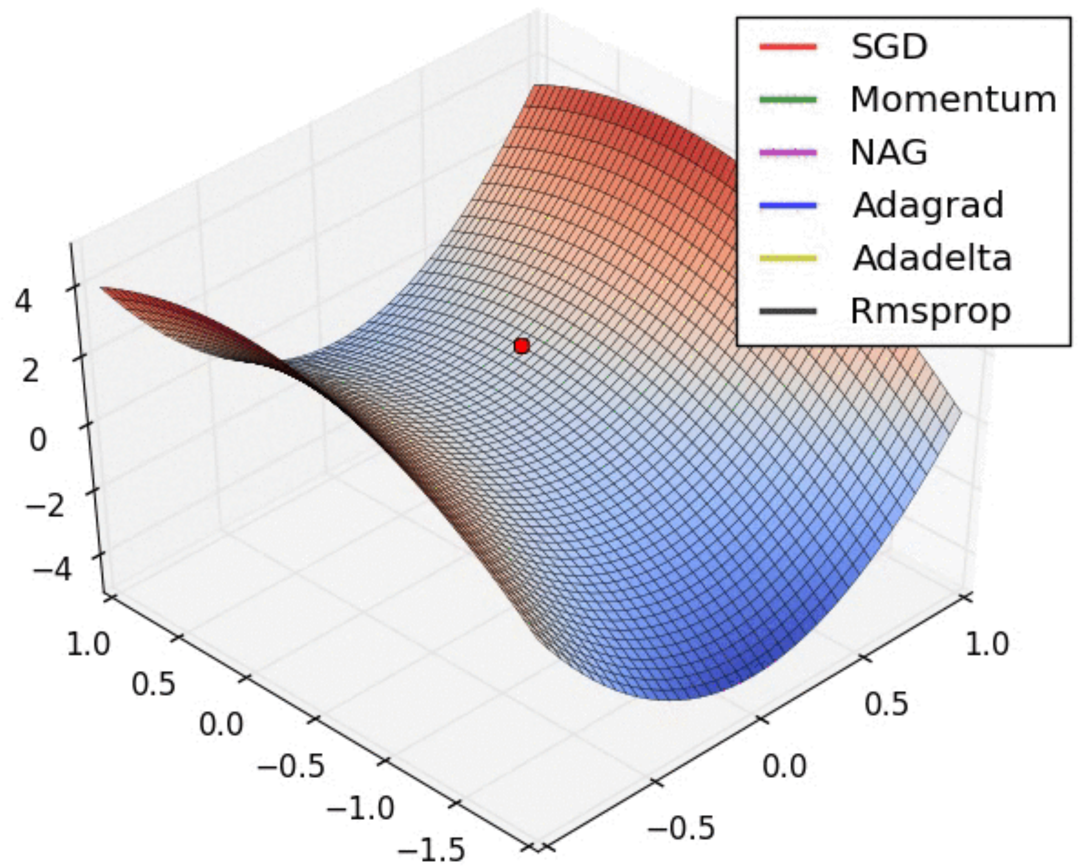
20% better

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



Symmetry Teleportation for Accelerated Optimization

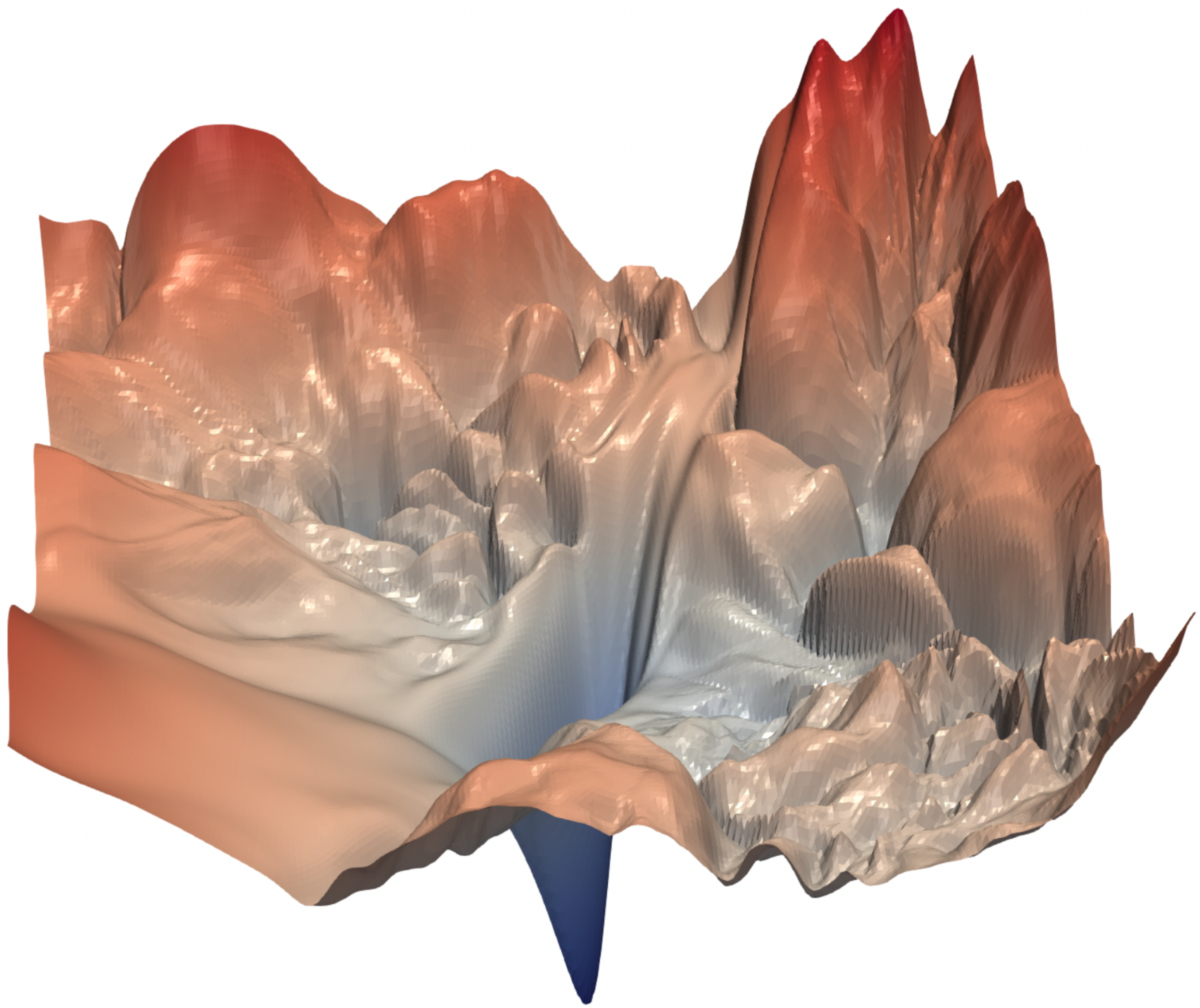


Bo Zhao

Nima Dehmamy

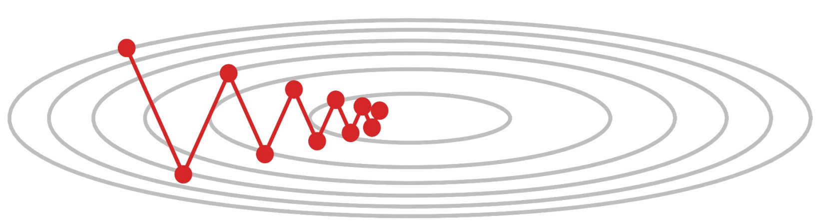
Symmetry Teleportation for Accelerated Optimization
Bo Zhao, Nima Dehmamy, Robin Walters, and <u>Rose Yu</u>.
Advances in Neural Information Processing Systems (NeurIPS), 2022.

Optimization in DL is Hard!

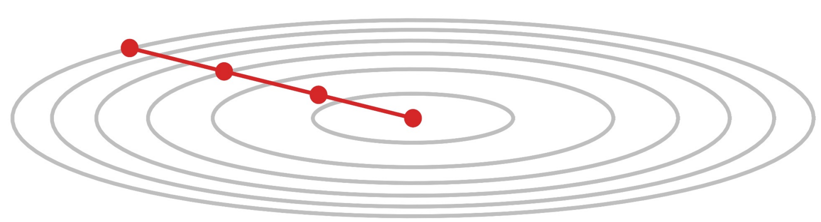


Li, Hao, et al. "**Visualizing the loss landscape of neural nets**."
Advances in neural information processing systems 31 (2018).

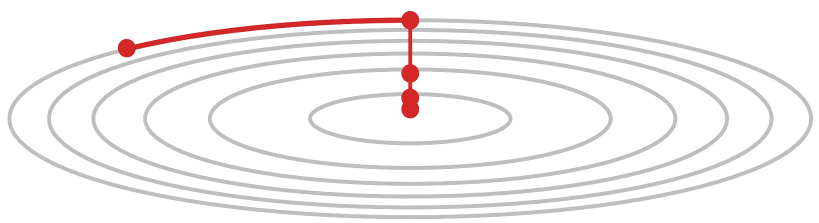
Symmetry Can Help



Gradient Descent



Second-order



GD + Teleportation

Symmetry Teleportation

Algorithm 1: Symmetry Teleportation

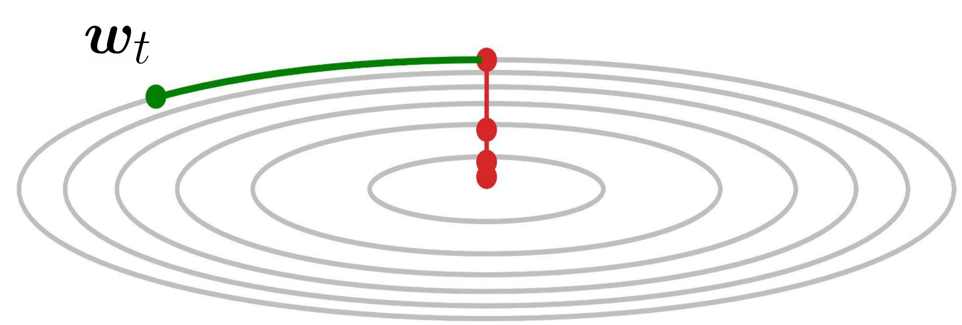
Input: Loss function $\mathcal{L}(\boldsymbol{w})$, learning rate η, number of epochs t_{max}, initialized parameters $\boldsymbol{w}_0$, symmetry group G, teleportation schedule K.

Output: $\boldsymbol{w}_{t_{max}}$.

for $t \leftarrow 0$ **to** $t_{max} - 1$ **do**
  if $t \in K$ **then**
    $g \leftarrow \text{argmax}_{g \in G} \|(\nabla \mathcal{L})|_{g \cdot \boldsymbol{w}_t}\|^2$
    $\boldsymbol{w}_t \leftarrow g \cdot \boldsymbol{w}_t$
  end if
  $\boldsymbol{w}_{t+1} \leftarrow \boldsymbol{w}_t - \eta(\nabla \mathcal{L})|_{\boldsymbol{w}_t}$
end for
return $\boldsymbol{w}_{t_{max}}$

teleportation

gradient descent

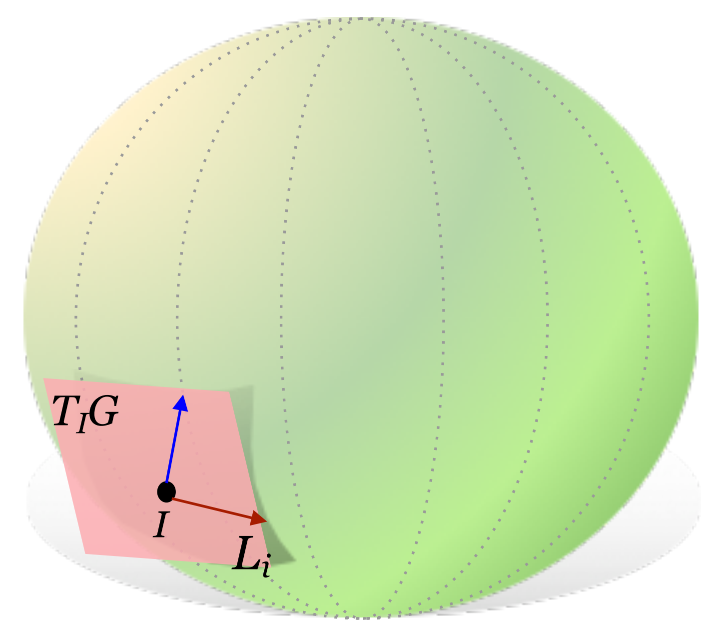


Finding teleportation destination

$$g \leftarrow \mathrm{argmax}_{g \in G} \|(\nabla \mathcal{L})|_{g \cdot \boldsymbol{w}_t}\|^2$$

Parametrize and gradient ascent on *g.*

- Example 1: SO(2), gradient ascent on rotation angle θ
- Example 2: $GL_d(\mathbb{R})$, g ≈ $I + \varepsilon T$ where $\varepsilon << 1$ and $T \in \mathbb{R}^{dxd}$, gradient ascent on T

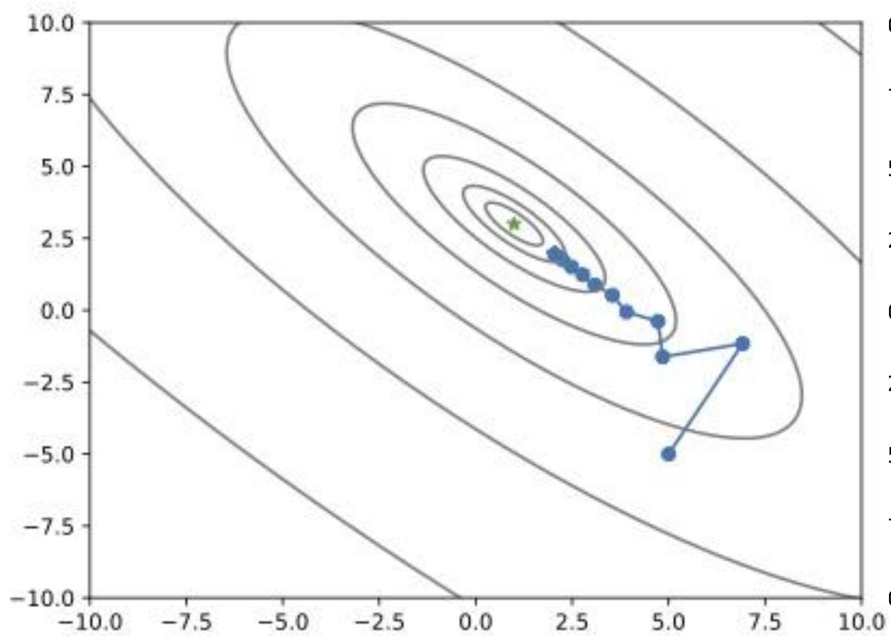


Test function: Booth

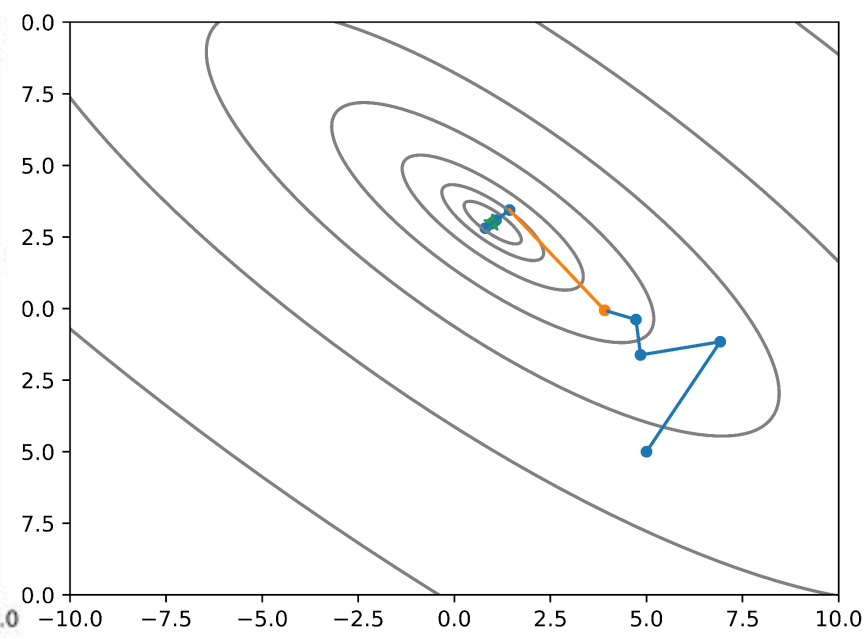
$$\mathcal{L}_b(x_1, x_2) = (x_1 + 2x_2 - 7)^2 + (2x_1 + x_2 - 5)^2$$

Symmetry group: SO(2)

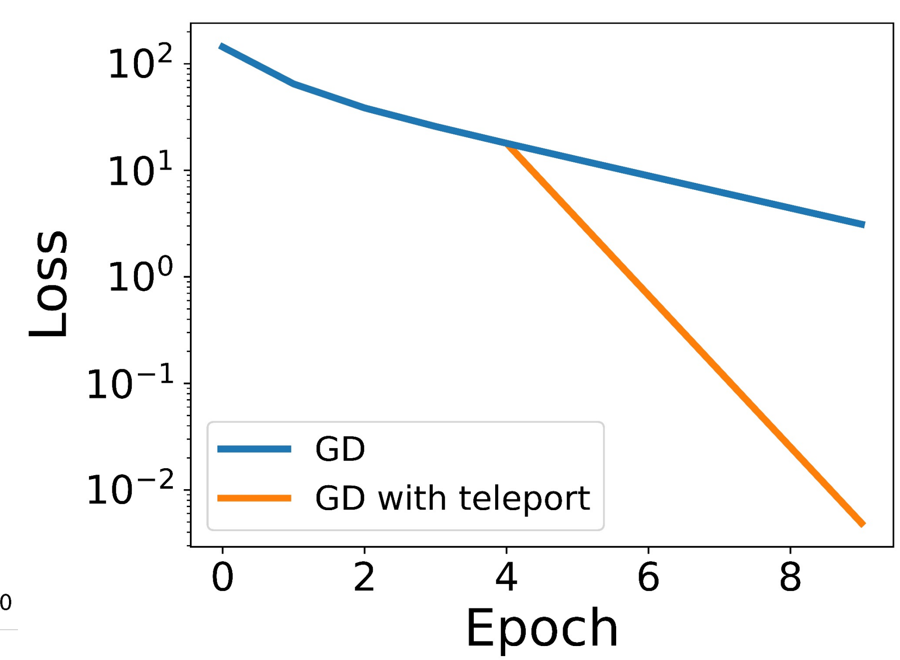
GD



GD + teleportation



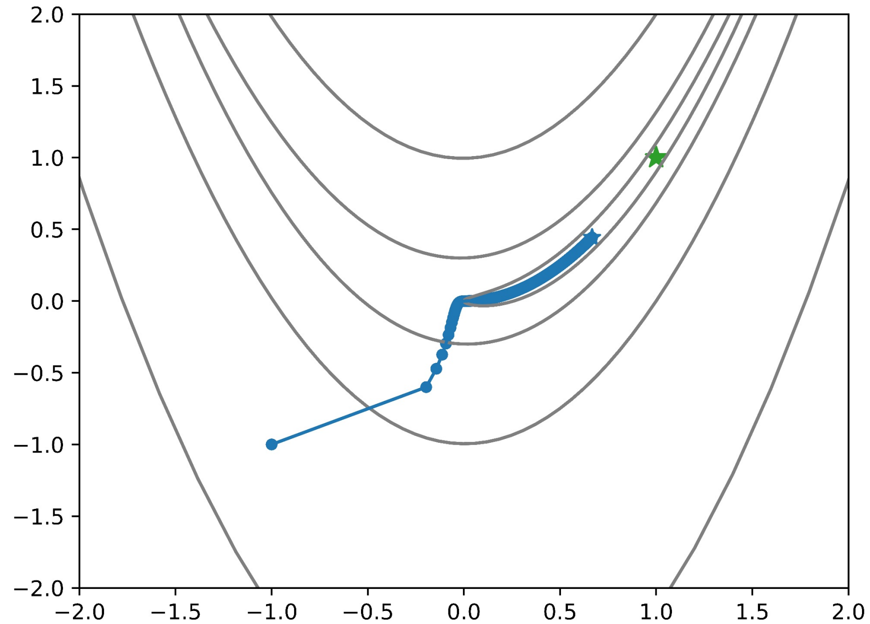
Training curves



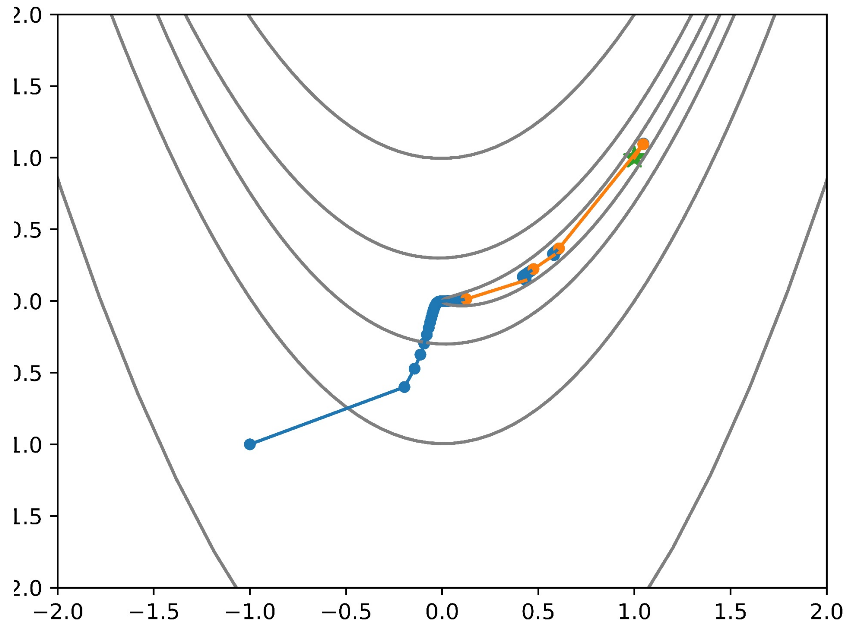
Test function: Rosenbrock

$$\mathcal{L}_r(x_1, x_2) = 100(x_1^2 - x_2)^2 + (x_1 - 1)^2$$

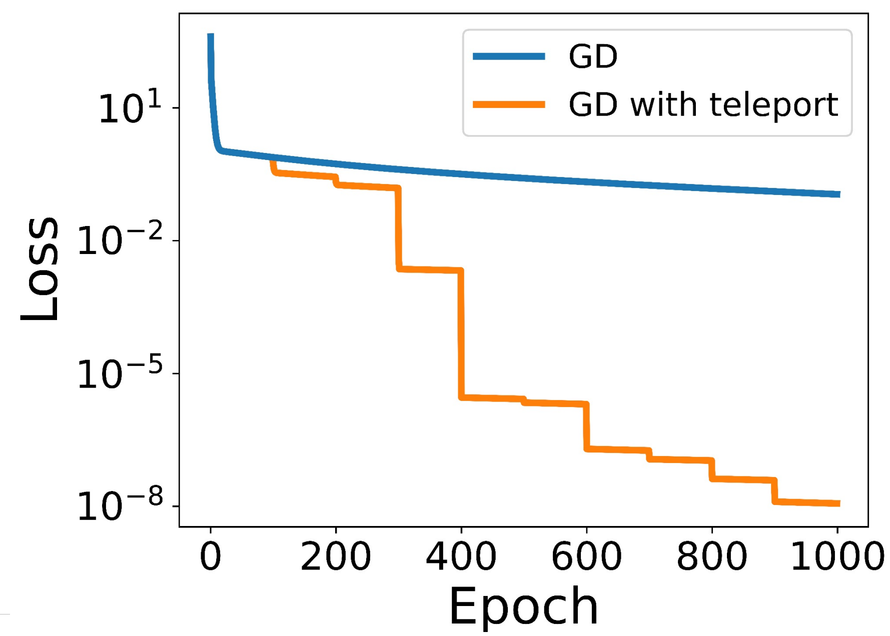
GD



GD with teleportation

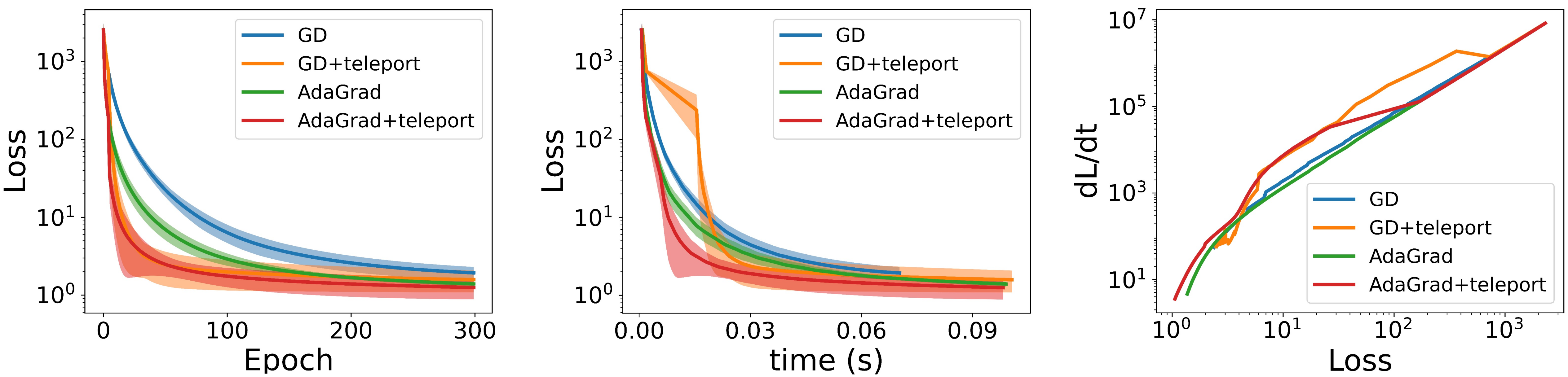


Training curves



Multi-layer neural network regression

$$\mathcal{L}(W_1, W_2, W_3) = \|Y - W_3\sigma(W_2\sigma(W_1 X))\|_2$$



Conclusion

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

On Molecules…

SCAN ME

Acknowledgment

Open Source Code and Data: roseyu.com

@yuqirose