Exact and approximate symmetries in machine learning models

Soledad Villar

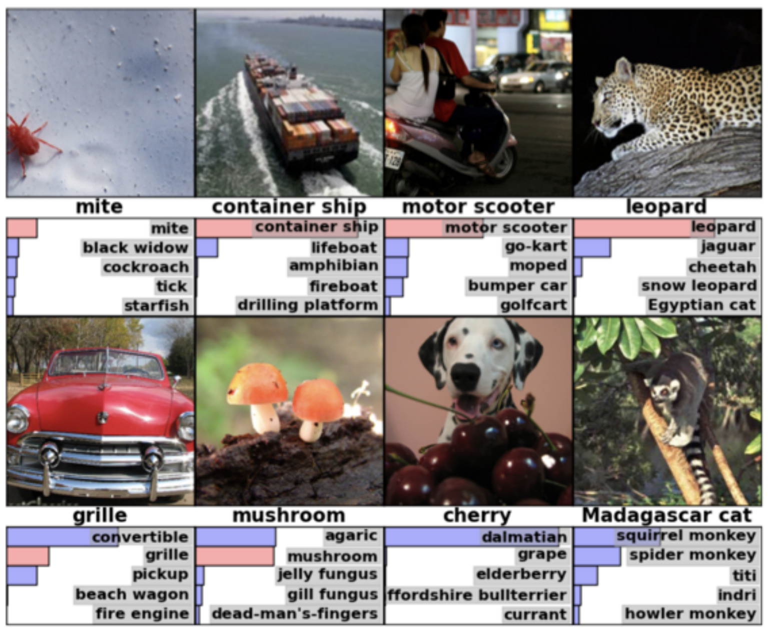
Johns Hopkins University

IAIFI, MIT, March 22, 2024

Deep learning

State of the art performance in several domains

ChatGPT '23

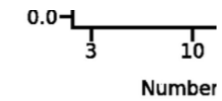


Krizhevsky et al. '12

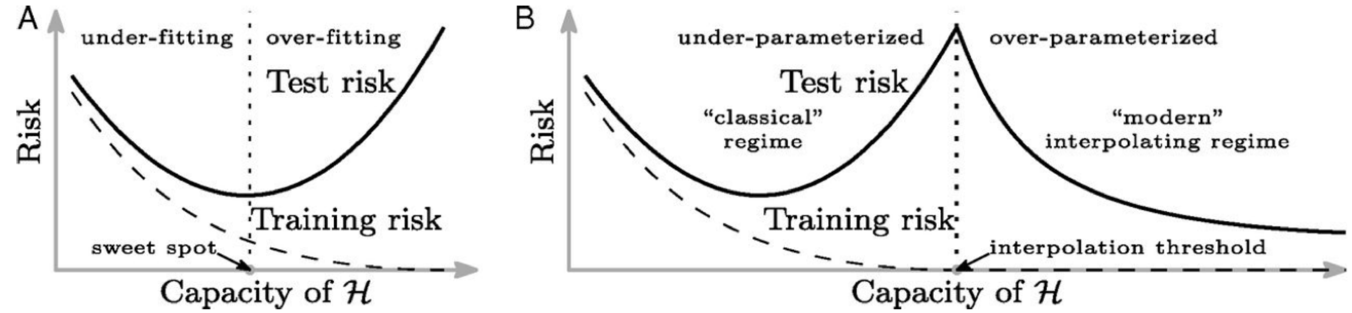
Silver et al. '16

Jumper et al. '21

Motivation : Deep learning inductive bias



Belkin et al '19
"double descent" →



- Very overparameterized models

  # parameters >> # data points

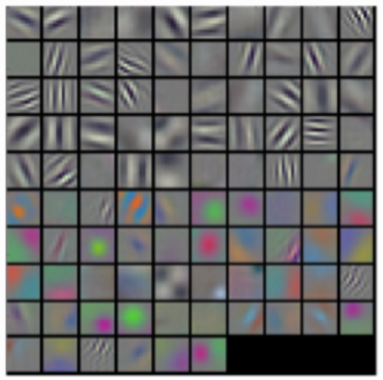
- Many functions in the hypothesis class fit the data

  How to choose the right one?

- Define the hypothesis class with the correct inductive bias

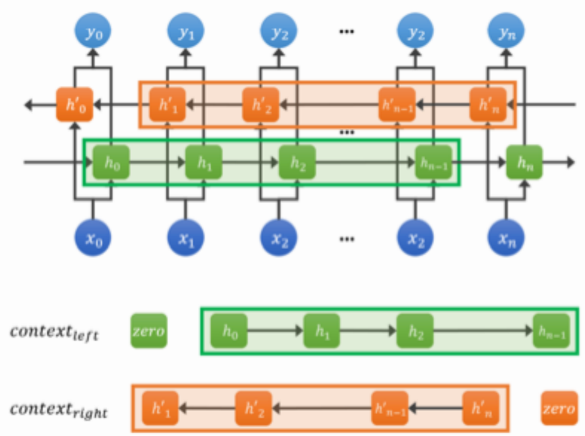
  Meaning that local optimization converges to a good solution (ie with good generalization)

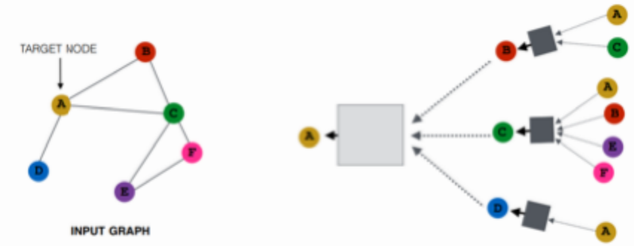
Deep learning archictectures with good inductive bias



CNN
spatial translation symmetry

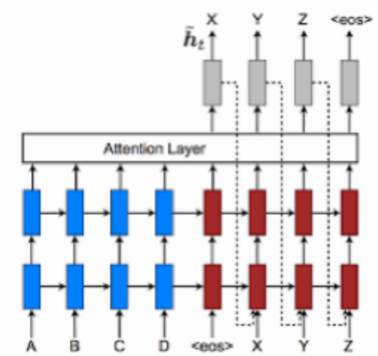
RNN
time translation symmetry





GNN
graph permutation symmetry

Transformers
flexible symmetries



(permutations and more!)

They exploit symmetries or approximate symmetries
[or more generally the physical structure of problems]

Related with geometric deep learning (Bronstein et al '22)

Motivation

Symmetries in the physical sciences

2 types of symmetries:

ACTIVE

- Symmetries that come from observed regularities of physics

Emmy Noether

conservation of energy — time translation symmetry

conservation of momentum — translation symmetry

conservation of angular momentum — rotation symmetry

PASSIVE

- Symmetries that come from choice of mathematical representation of physical objects

coordinate freedom

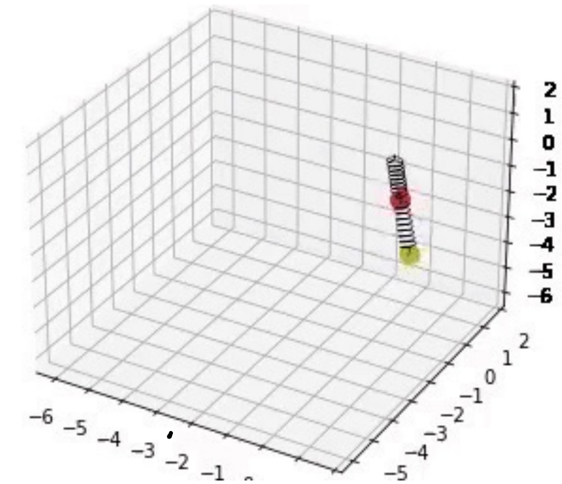
units equivariances

gauge invariances / equivariances

CLAIM: ML / data science methods should be consistent with these

Definition

Invariance / Equivariance



Exact symmetries

G a group acting on dataset X

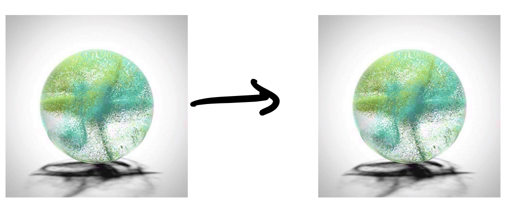
$F: X \to Y$ invariant if $F(g \cdot x) = F(x)$ $\forall g \in G \quad x \in X$

 $\to$ "Rose"

If G also acts in Y

$H: X \to Y$ equivariant $H(g \cdot x) = g \cdot H(x)$ $\forall g \in G, \; x \in X$

$$\begin{tikzcd} X \arrow[r, "H"] \arrow[d, "\varphi(g)"'] & Y \arrow[d, "\psi(g)"] \\ X \arrow[r, "H"] & Y \end{tikzcd}$$

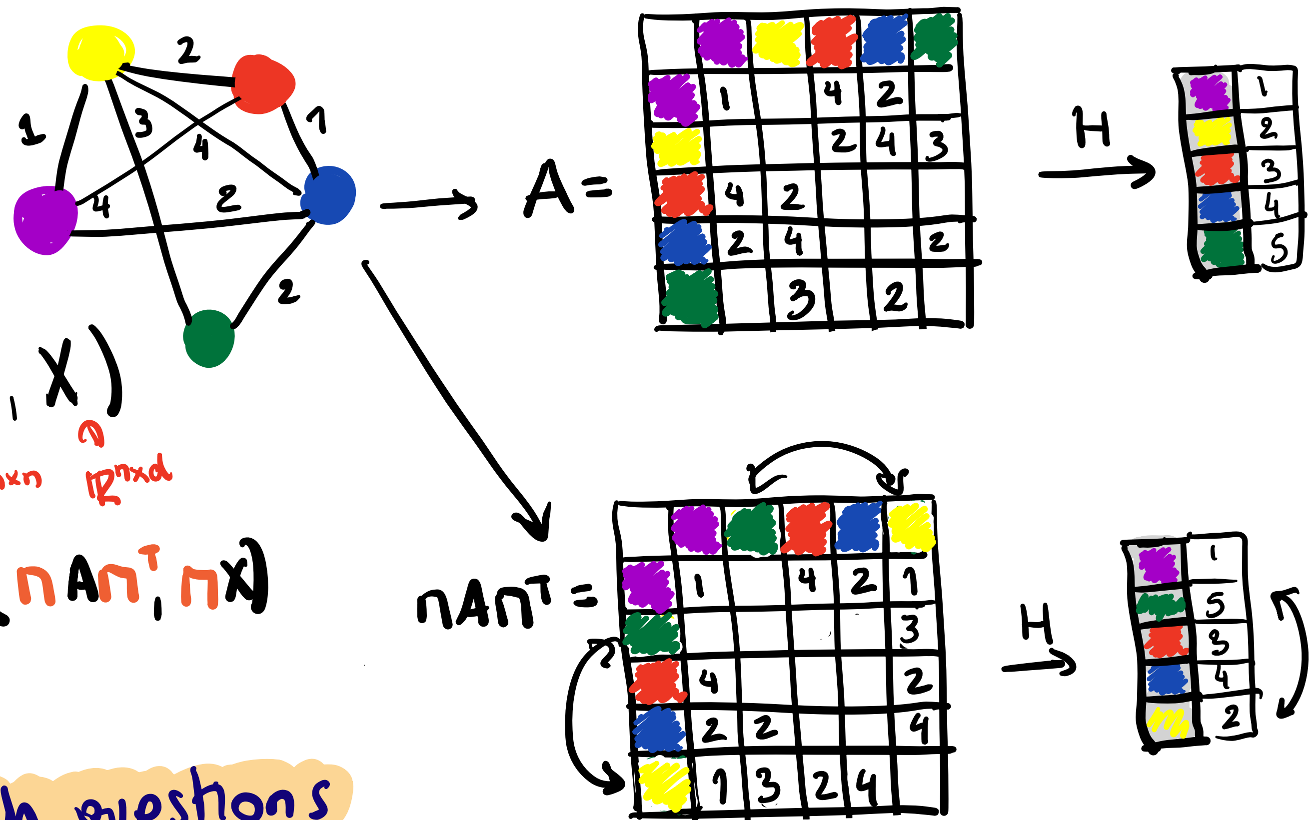


All functions in this space are (approx) invariant or equivariant

EQUIVARIANT ML : $\mathcal{H} = \{ f_\theta : X \to Y \}$

Example passive symmetries in graph learning

G graph



$G = (A, X)$, $A \in \mathbb{R}^{n \times n}$, $X \in \mathbb{R}^{n \times d}$

$\Pi G = (\Pi A \Pi^T, \Pi X)$

Example:

shortest path starting at ●

- length (invariant)
- path (equivariant)

Node embedding: $G \mapsto \mathbb{R}^{n \times d}$

$H(\Pi A \Pi^T, \Pi X) = \Pi \cdot H(A) = \Pi \cdot P$ equivariant

Graph classification / regression: $G \to \mathbb{R}$

$F(\Pi A \Pi^T, \Pi X) = F(A)$ invariant

$\forall A \in \mathbb{R}^{n \times n}$

$\Pi \in$ permutation

Research questions

- How to parameterize equivariant functions efficiently for ML
- Optimization, generalization
- Expressive power
  - Connections to graph isomorphism
- Approximate symmetries and bias-variance trade-offs

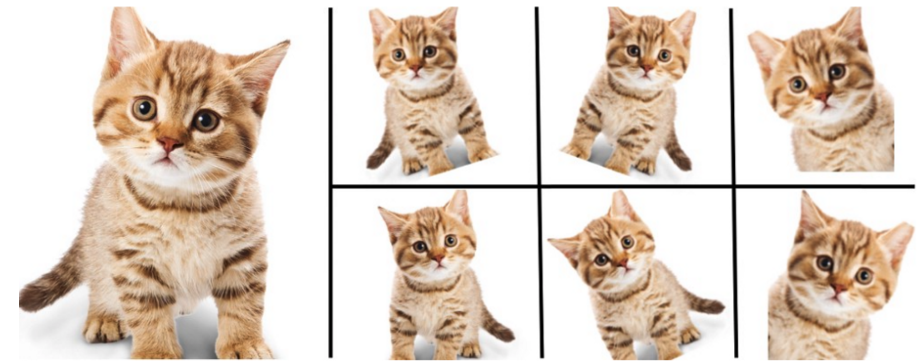
This talk

- Equivariant functions as gradients of invariant functions
  Blum-Smith, V.
  - Examples
    - O(d), SO(d), O(1,d) equivariance
      V., Hogg, Storey-Fisher, Yao, Blum-Smith
    - Units equivariance
      V., Yao, Hogg, Blum-Smith, Dumitrescu
- Applications to approximately equivariant contrastive learning
  Gupta, Robinson, Lim, V., Jegelka
- Invariant functions on point clouds with applications to cosmology
  Blum-Smith, Huang, Cuturi, V.
  Storey-Fisher, Hogg, Genel, V
- Approximate equivariance in graph neural networks
  Huang, Levie, V.

How are symmetries implemented?

- Data augmentation
- Loss function penalties

Approximate



Enlarge your Dataset

Credit: Bharath Raj

- Representation theory (Kondor, Cohen, Welling, Smidt...)
- Group convolutions / weight sharing (Yu, Trivedi...)
- Invariant theory (Villar, Blum-Smith)

Exact

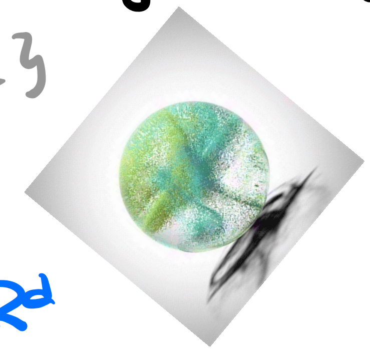
Restrict the class of functions to functions that satisfy the symmetries

Bonus if universal

Invariant theory approach

Example $O(d)$ Orthogonal group

$\{R \in \mathbb{R}^{d\times d} : RR^T = R^TR = I\}$



$\xrightarrow{f} \mathbb{R}$

$O(d)$- invariant functions $f : (\mathbb{R}^d)^n \to \mathbb{R}$

$$f(Rv_1, \dots Rv_n) = f(v_1, \dots v_n) \quad \forall\ R \in O(d) \quad v_1 \dots v_n \in \mathbb{R}^d$$

(Weyl 1946) first fundamental theorem of invariant theory

$f : (\mathbb{R}^d)^n \to \mathbb{R}$ is $O(d)$-invariant if and only if

$$f(v_1, \dots v_n) = \tilde{f}\Big(\underbrace{(v_i^T v_j)_{i,j=1}^n}_{\text{inner products}} \Big)$$

$$\begin{tikzcd} (\mathbb{R}^d)^n \arrow[r, "f"] \arrow[d, "\Phi"] & \mathbb{R} \\ \mathbb{R}^{n\times n} \arrow[ur, dashed, "\tilde{f}"'] & \end{tikzcd}$$

$(v_1, \dots v_n)$

$M \quad M_{ij} = v_i^T v_j$

- Similar characterizations for
  - Lorentz
  - Rotations
  - Symplectic group
  - Unitary group

(Villar et al '21)
Neurips '21

Characterization of SO(d)-invariant functions

$f:(\mathbb{R}^d)^n \to \mathbb{R}$ is SO(d)-invariant if and only if

$$f(v_1,\dots,v_n) = \tilde{f}\left((v_i^T v_j)_{i,j=1}^n ,\ \det(v_{i_1} \dots v_{i_d})_{i_1 \dots i_d \in \binom{[n]}{d}} \right)$$

$$V = \begin{bmatrix} | & & | \\ v_1 & \dots & v_n \\ | & & | \end{bmatrix}$$

Characterization of Lorentz-invariant functions

$f:(\mathbb{R}^{d+1})^n \to \mathbb{R}$ is Lorentz-invariant if and only if

$$f(v_1,\dots,v_n) = \tilde{f}\left(\langle v_i, v_j \rangle_M {}_{i,j=1}^n \right)$$

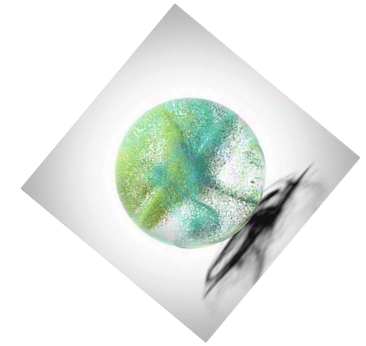
where $\langle (t,x),(t',x') \rangle_M = tt' - x^T x'$

Minkowski "inner product"

$$O(1,d) = \{ R : R \Lambda R^T = \Lambda \ : \ \Lambda = \begin{pmatrix} -1 & 0 \\ 0 & I \end{pmatrix} \}$$

Example

$O(d)$ Orthogonal group



$\xrightarrow{h} \mathbb{R}^d$

equivariant functions $h : (\mathbb{R}^d)^n \to \mathbb{R}^d$

$$h(Rv_1, \dots Rv_n) = R \cdot h(v_1, \dots v_n) \quad \forall \; R \in O(d) \qquad v_1 \dots v_n \in \mathbb{R}^d$$

Proposition

$h : (\mathbb{R}^d)^n \to \mathbb{R}^d$ is $O(d)$-equivariant if and only if

$$h(v_1, \dots v_n) = \sum_{s=1}^{n} \underbrace{f_s\Big(\underbrace{(v_i^T v_j)_{i,j=1}^{n}}_{\text{inner products}} \Big)}_{\text{invariant scalar functions}} v_s$$

Proof (by Schwarz, Malgrange)

$f : (\mathbb{R}^d)^{\times n} \times \mathbb{R}^d \to \mathbb{R}$ inv linear in last coordinate

$\dfrac{\partial f}{\partial v_{n+1}} : (\mathbb{R}^d)^n \to \mathbb{R}$ equivariant

$f : (\mathbb{R}^d)^n \to \mathbb{R}^d$ equiv $(O(d))$

$h : (\mathbb{R}^d)^n \times \mathbb{R}^d \to \mathbb{R}$ invariant

$h(v_1 \dots v_n, v^*) = \frac{\partial}{\partial v^*} \underline{\langle f(v_1 \dots v_n), \cancel{v^*} \rangle}$ inv

$= \frac{\partial}{\partial v^*} g\left(\underline{\langle v_i, v_j \rangle_{i,j=1}^n , \langle v_s, \cancel{v^*} \rangle} \right)$

$= \frac{\partial}{\partial v^*} \sum_{s=1}^n g(\langle v_i, v_j \rangle) \cdot \langle v_s, \cancel{v^*} \rangle$

General theory INVARIANCE → EQUIVARIANCE

(Malgrange) Schwarz ('75, '80)

$V \to W$

$V \times W^* \to \mathbb{R}$

Knowledge of invariant maps $V \times W^* \to \mathbb{R}$

$\Rightarrow$ knowledge of equivariant maps $V \to W$

Ben Blum-Smith

If $h: V \to W$ equivariant $\Rightarrow$ $f: V \times W^* \to \mathbb{R}$ invariant

$$(v, \ell) \longmapsto \ell(f(v))$$

$$\mathrm{Hom}(V, W) \cong \mathrm{Hom}(V \otimes W^*, \mathbb{R}) \cong \mathrm{Hom}(V \times W^*, \mathbb{R})$$

(linear maps) (bilinear maps)

Formally

$\left.\begin{array}{l} G \text{ acts on } V \text{ by } \phi \\ G \text{ acts on } W \text{ by } \psi \end{array}\right\} \Rightarrow G$ acts on maps(V, W)

$$gf := \psi(g) \circ f \circ \phi^{-1}(g)$$

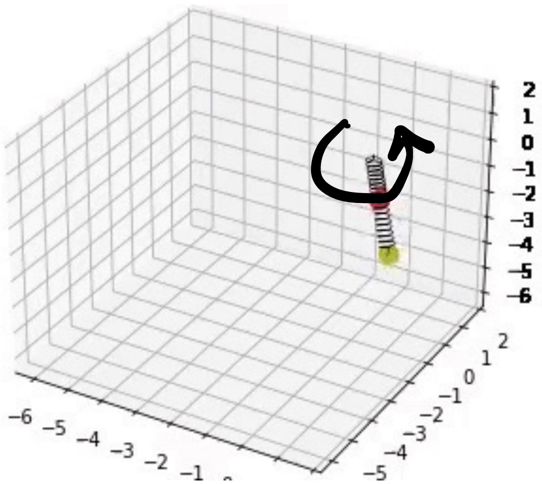
$$\begin{tikzcd} V \arrow[r, "f"] & W \arrow[d, "\psi(g)"] \\ V \arrow[u, "\phi(g)"] \arrow[r, "gf"] & W \end{tikzcd}$$

f is a fixed point of the action on maps(V, W) $\Longleftrightarrow$ f is equivariant

$f = gf \Leftrightarrow f$ equivariant

$$\begin{tikzcd} V \arrow[r, "f"] \arrow[d, "\phi(g)"'] & W \arrow[d, "\psi(g)"] \\ V \arrow[r, "f"] & W \end{tikzcd}$$

Toy example : double pendulum with springs



Credit : EMLP (Finzi et al '21)

Weichi Yao

data: $(q_1(t), p_1(t))$, m_1, L_1, K_1

$(q_2(t), p_2(t))$ m_2, L_2, K_2

problem: predict the dynamics

$$KE = \frac{|p_1|^2}{2m_1} + \frac{|p_2|^2}{2m_2} \quad PE = \frac{1}{2}K_1(|q_1| - L_1)^2 - m_1 p_1 \cdot g + \frac{1}{2}K_2(|q_1 - q_2| - L_2)^2 - m_2 p_2 \cdot g$$

$H = KE + PE$ conserved quantity $\longleftrightarrow$ time translation symmetry

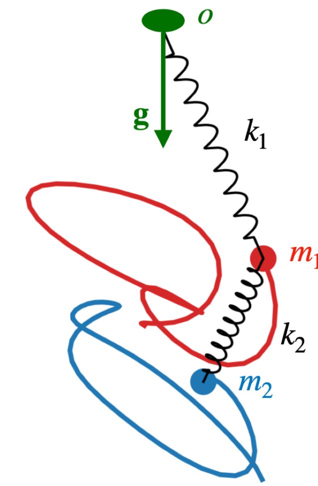
(Hamiltonian)

O(3)-equivariant (passive symmetry)

$$F: (\mathbb{R}^3)^5 \times \mathbb{R} \longrightarrow (\mathbb{R}^3)^4$$

$$(q_1(0), p_1(0), q_2(0), p_2(0), g, \Delta t) \longmapsto (q_1(\Delta t), p_1(\Delta t), q_2(\Delta t), p_2(\Delta t))$$

Computational approaches:



Goal: Predict $z(t) = (q_1(t), q_2(t), p_1(t), p_2(t))$

- Neural ODEs

$$(z(0), g) \longrightarrow \begin{matrix} O(3)\text{-equivariant} \\ \text{function } F_\theta \end{matrix} \longrightarrow \frac{dz(t)}{dt} = F_\theta(z, g) \xrightarrow[\text{ODE solver}]{} \hat{z}(t)$$

$$\text{input} \longrightarrow \text{scalars} \xrightarrow{f_1} \vdots \text{(MLP)} \rightarrow \oplus \rightarrow \text{output}$$

$$q_1, q_2, p_1, p_2, g \qquad \begin{matrix} q_1^T q_1 \\ q_1^T q_2 \\ q_1^T p_1 \end{matrix}$$

$$F_\theta(q_1, q_2, p_1, \vdots, p_2, g) = \sum_{v} f_v\left((v_i^T v_j)_{v_i, v_j \in V}\right) \cdot v$$

$$v_1 \in V = \{q_1, q_2, p_1, p_2, g\}$$

- Hamiltonian neural network (HNNs)

$$(z(0), g) \rightarrow \text{O(3) - invariant function } H_\theta \rightarrow \begin{aligned} \frac{dp_1}{dt} &= -\frac{dH_\theta}{dq_1} \\ \frac{dq_1}{dt} &= \frac{dH_\theta}{dp_1} \end{aligned} \xrightarrow[\text{ODE solver}]{} \hat{z}(t)$$

(symplectic integrator)

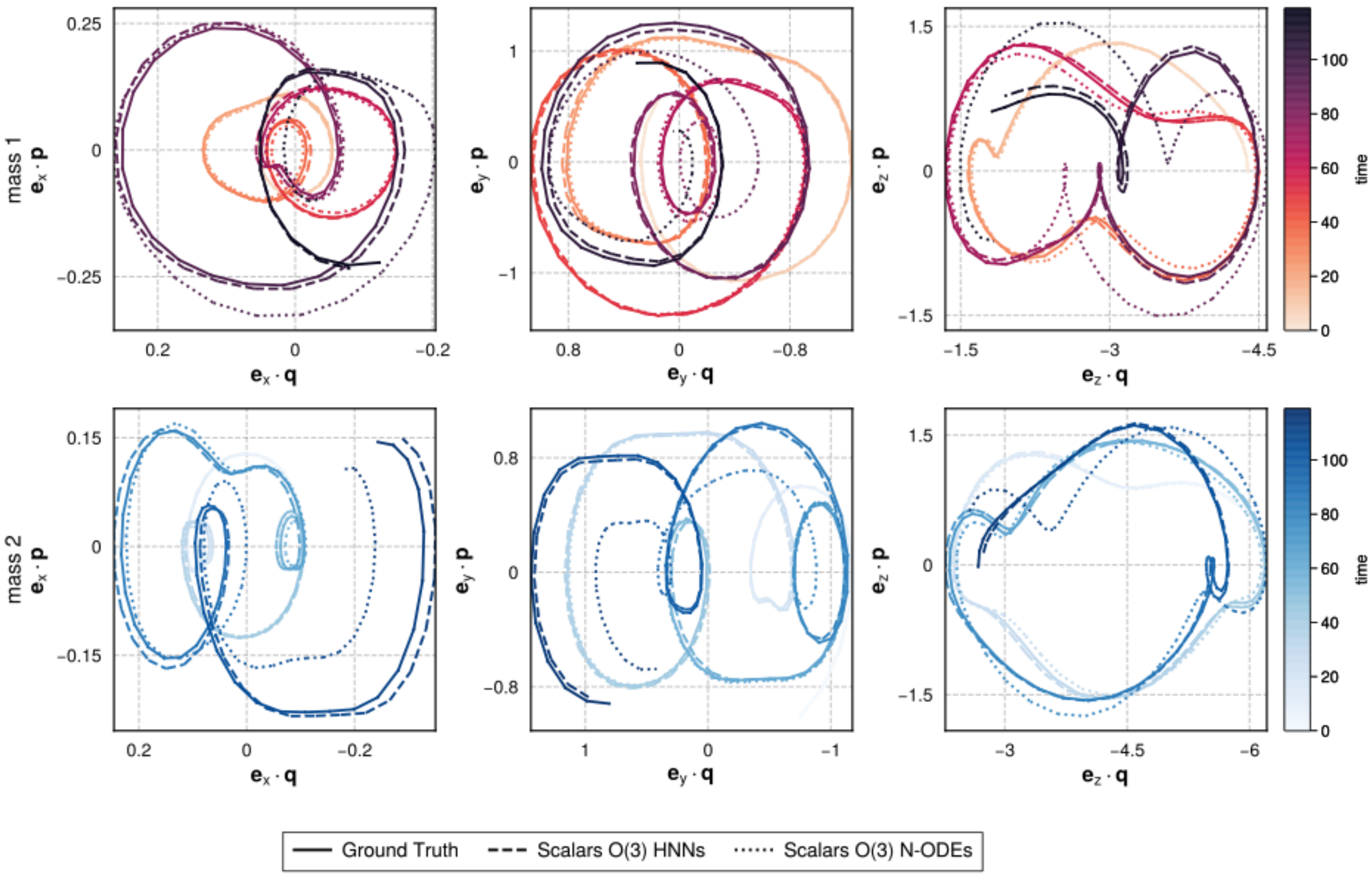
$$H_\theta(q_1, q_2, p_1, p_2, q_0, g) = h\,(\text{inner products})$$

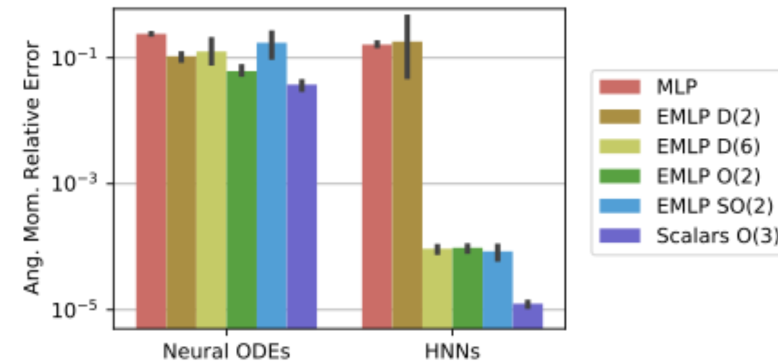
Learned scalar invariant function

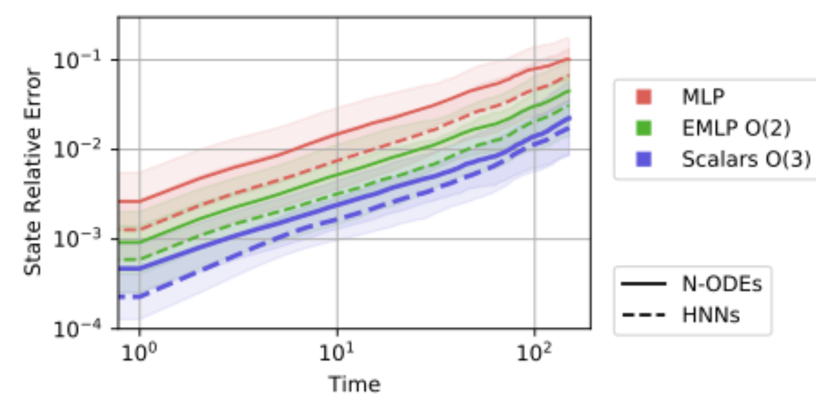
Results

	Scalars O(3)	EMLP				MLP
		O(2)	SO(2)	D_2	D_6	
N-ODEs:	**.009 ± .001**	.020 ± .002	.051 ± .036	.023 ± .002	.036 ± .025	.048 ± .000
HNNs:	**.005 ± .002**	.012 ± .002	.016 ± .003	.111 ± .167	.013 ± .002	.028 ± .001

Finzi ICML '21







Units-equivariance

Non-compact groups
(Villar et al '22)

Double pendulum:

$$PE = \frac{1}{2}k_1(|q_1| - L_1)^2 - m_1 p_1 \cdot g + \frac{1}{2}k_2(|q_1 - q_2| - L_2)^2 - m_2 p_2 \cdot g$$

$$KE = \frac{|p_1|^2}{2m_1} + \frac{|p_2|^2}{2m_2}$$

Energy has units: $kg\ m^2\ s^{-2}$

Predictions should be equivariant with respect to rescalings
↳ Passive symmetry from dimensional analysis

Approach:

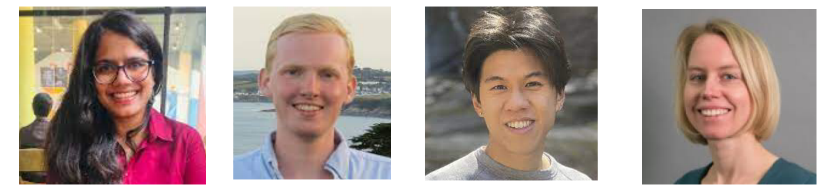
inputs: $\mathbb{R}^d$ $\xrightarrow{\Phi}$ dimensionless features $\mathbb{R}^s$ $\xrightarrow{\text{NNET}}$ dimensionless labels $\xrightarrow{\text{Fixed scaling}}$ dimensional outputs

each feature has units

Based on classical dimensional analysis (Buckingham PI theorem)

Application

Approximate symmetries in contrastive learning



S. Gupta J. Robinson D. Lim S. Jegelka

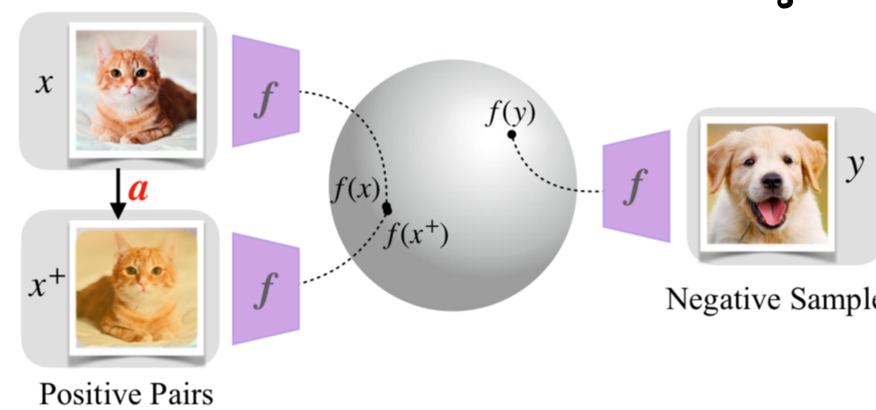
Contrastive learning (self-supervised)

$$f : X \to S^{d-1}$$

data embedding

$$\mathcal{L}_{\text{infoNCE}} = \mathbb{E}_{x, x^+, \{x_i^-\}_{i=1}^{N}} - \log \frac{e^{\frac{f(x)^T f(x^+)}{\tau}}}{e^{\frac{f(x)^T f(x^+)}{\tau}} + \sum_{i=1}^{N} e^{\frac{f(x)^T f(x_i^-)}{\tau}}}$$

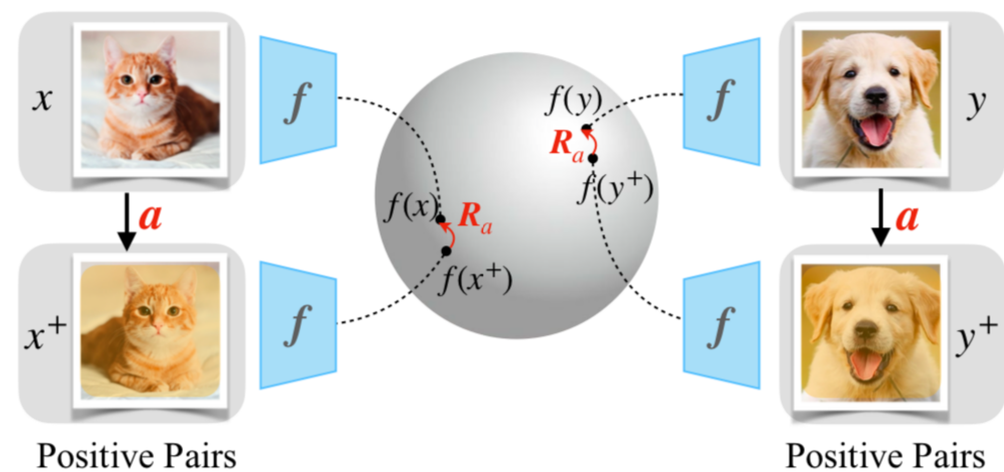
$a(x)$ augmented versions of points



Invariant Contrastive Learning (SimCLR)

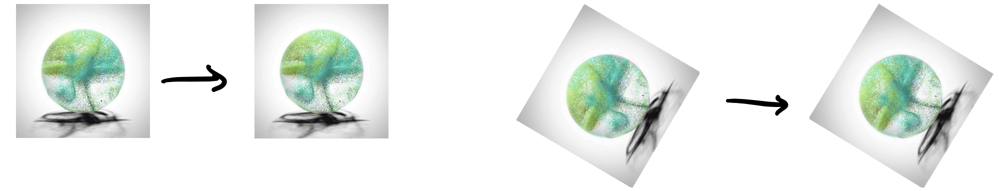
Equivariance to augmentation

- $f(a(x)) = T_a f(x)$
- Equivariance should be expressed in terms of pairs of data points



Orthogonally Equivariant Contrastive Learning (CARE)

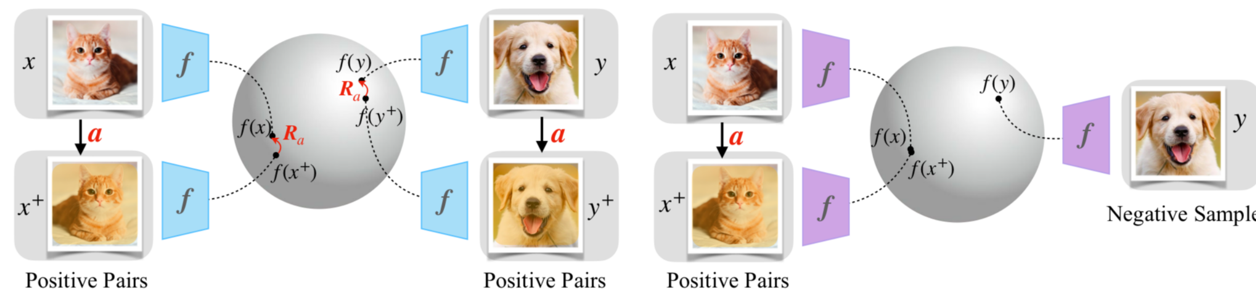
O(d)-equivariant contrastive learning



Recall $y_1, y_2, z_1, z_2 \in \mathbb{S}^{d-1}$ satisfy $\underbrace{y_1^T y_2 = z_1^T z_2}_{\text{same inner product}}$

iff there exists $R \in O(d)$ such that $Ry_1 = z_1$, $Ry_2 = z_2$

Equivariant contrastive learning:



Orthogonally Equivariant Contrastive Learning (CARE) **Invariant Contrastive Learning (SimCLR)**

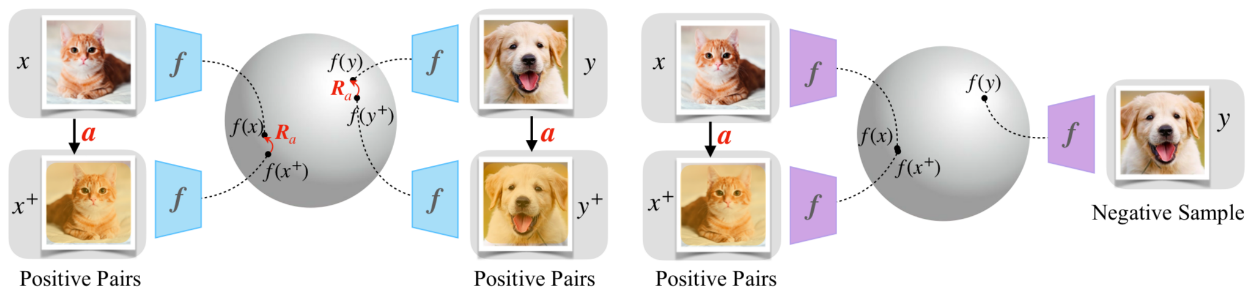
$f: X \to \mathbb{R}^d$ $\quad f(a(x'))^T f(a(x)) = f(x')^T f(x) \quad \forall x, x' \in X$

$\iff$ there exists $R_a \in O(d)$ st $f(a(x)) = R_a f(x) \quad \forall x \in X$

Augmentations in $X \Rightarrow$ Orthogonal transformations in embedding space

Contrastive Augmentation-induced Rotation Equivariance (CARE)

Augmentations on inputs $\longleftrightarrow\ \approx O(d)$ transformations in embeddings



Orthogonally Equivariant Contrastive Learning (CARE) **Invariant Contrastive Learning (SimCLR)**

$$\mathcal{L}_{CARE}(f) = \mathcal{L}_{inv}(f) + \mathcal{L}_{unif}(f) + \lambda\, \mathcal{L}_{equiv}(f)$$

$$\mathcal{L}_{inv}(f) = \mathbb{E}_{a,a'\sim A} \|f(a(x)) - f(a'(x))\|, \quad \mathcal{L}_{unif}(f) = \log \mathbb{E}_{x,x'\in X} \exp\left(f(x)^T f(x')\right)$$

$$\mathcal{L}_{equi}(f) = \mathbb{E}_{a\sim A}\, \mathbb{E}_{x,x'\sim X} \left[f(a(x'))^T f(a(x)) - f(x')^T f(x) \right]^2$$

PROP $\mathcal{L}_{equi}(f) = 0 \iff$ for almost all $a \in A$ $\exists\, R_a \in O(d)$ st $f(a(x)) = R_a f(x)$

$\Rightarrow \rho : A \to O(d)$, $a \mapsto R_a$ is a group homorphism $\Rightarrow A \leq O(d)$

Results :

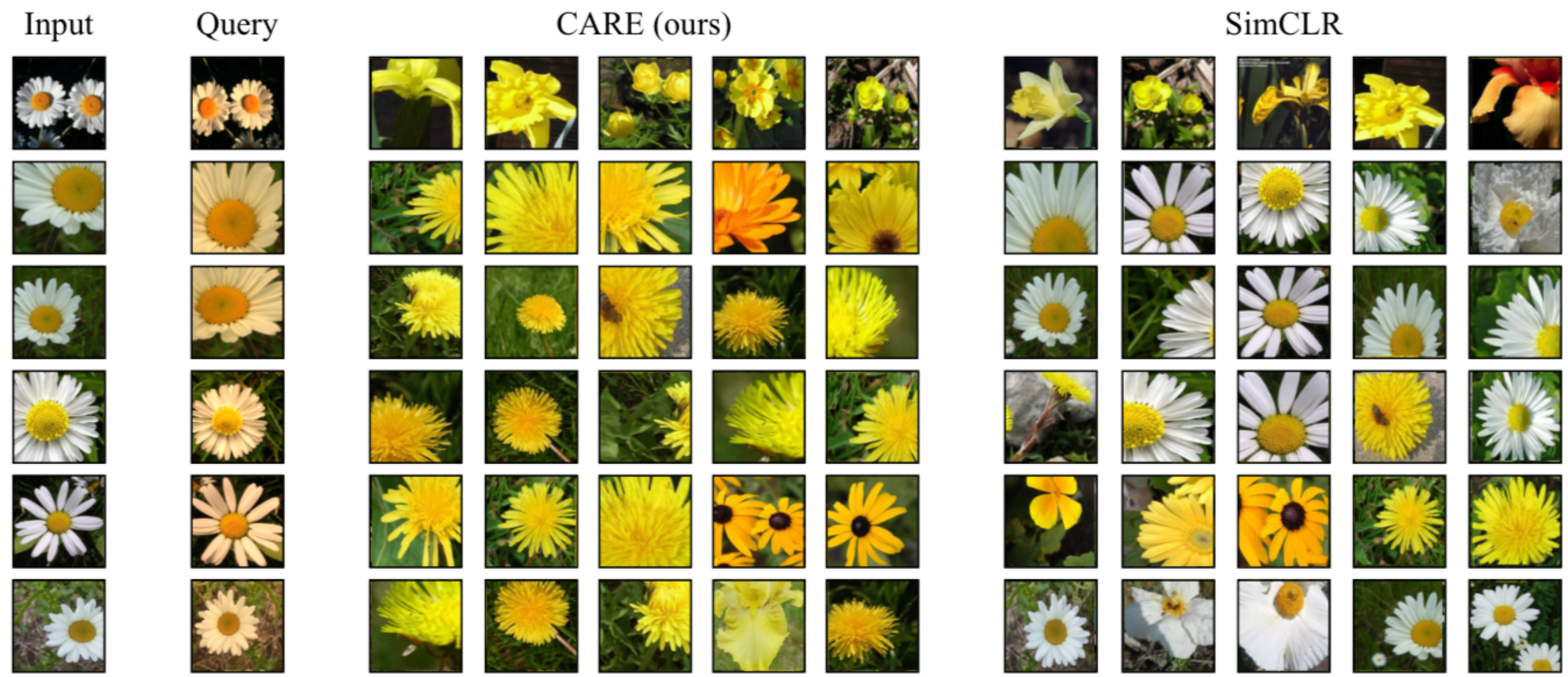


Figure 5: CARE exhibits sensitivity to features that invariance-based contrastive methods (e.g., SimCLR) do not. For each input we apply color jitter to produce the query image. We then retrieve the 5 nearest neighbors in the embedding space of CARE and SimCLR.

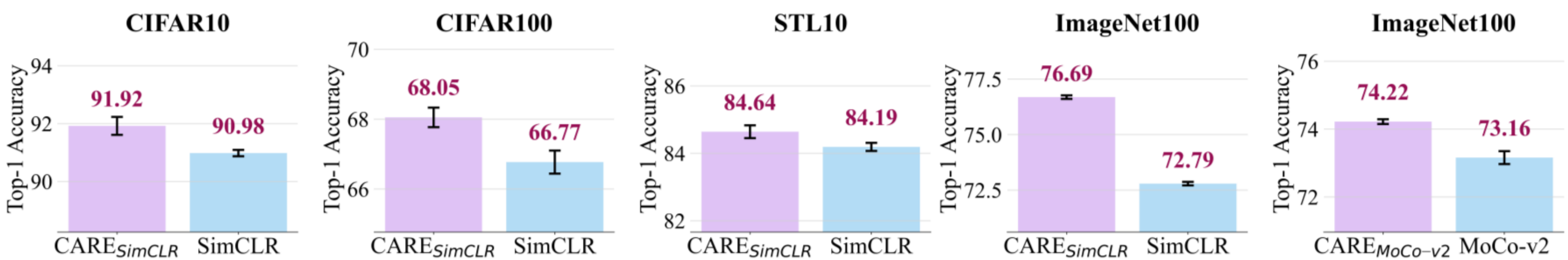
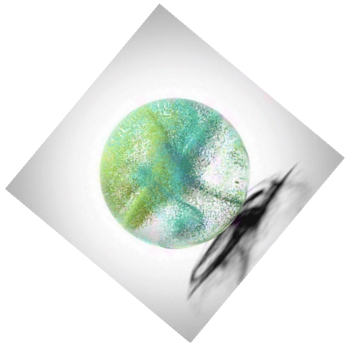


Figure 8: Top-1 linear readout accuracy (%) on CIFAR10, CIFAR100, STL10 and ImageNet100. All results are from 5 independent seed runs for the linear probe.

Extension

Invariant functions on point clouds

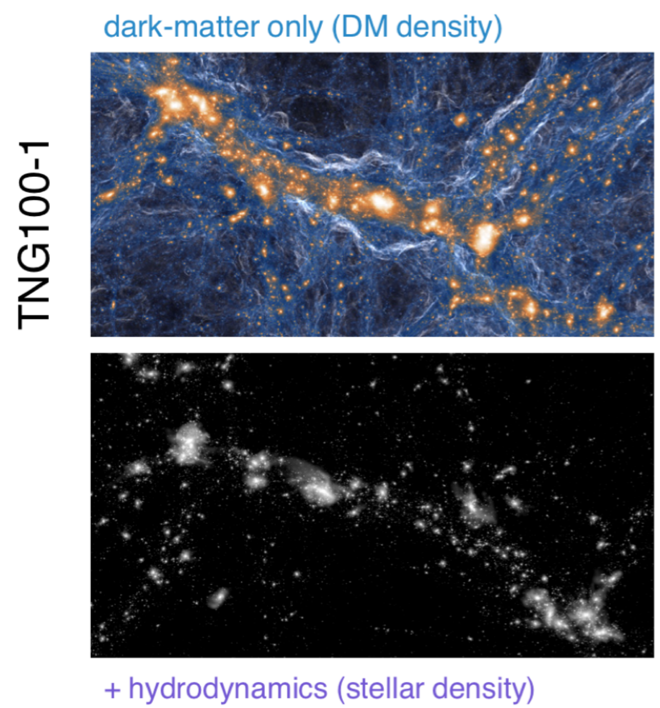
O(d) + permutation symmetry

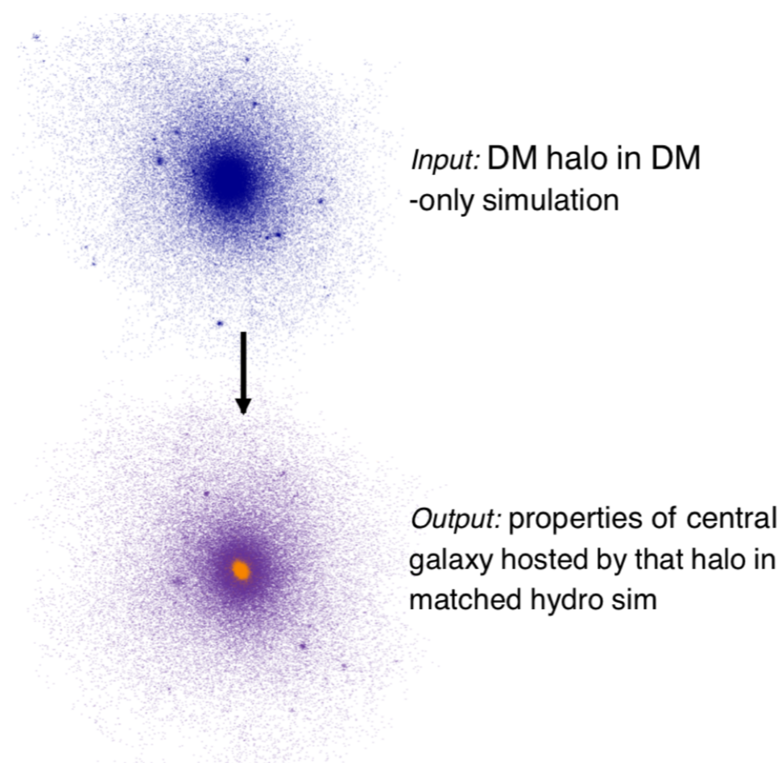


Kate Storey-Fisher

Motivation:

Emulation of cosmological simulations





Galaxy properties predictions from dark-matter only sims:

Storey-Fisher, Hogg, Genel, Hattori, V.

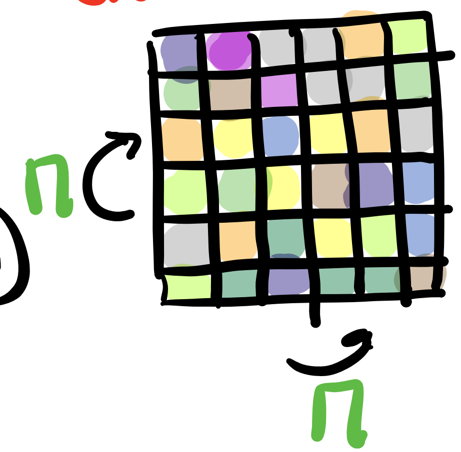
Invariant functions on point clouds

$$X = \begin{pmatrix} | & & | \\ p_1 & \cdots & p_n \\ | & & | \end{pmatrix} \in \mathbb{R}^{d \times n} \qquad O(d) \backslash X / S_n$$

$f : X \to \mathbb{R}$ $O(d)$-invariant $\iff f(X) = \tilde{f}(\underbrace{X^T X}_{\text{Gram matrix}})$

$$f(X\Pi) = f(X) \Rightarrow \tilde{f}(\Pi^T X^T X \Pi) = \tilde{f}(X^T X)$$

($\tilde{f}$ invariant by permutations acting by conjugation)



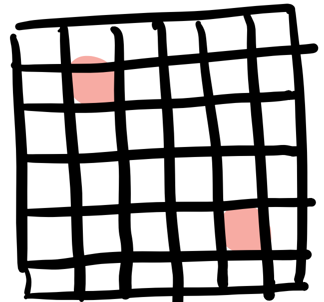
Rank d symmetric

Symmetry of GNNs !

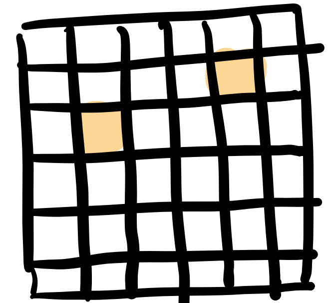
→ Current work: $n(d+1)$ generators for the field of invariant functions

Main idea:

2 orbits



$x_{ii} \to x_{ij}$



$x_{ij} \to x_{kl}$ $\quad i \neq j, \ k \neq l$

S_n field invariant functions by conjugation $\quad f(\mathrm{set}(x_{ii}), \mathrm{set}(x_{ij}), f^*)$

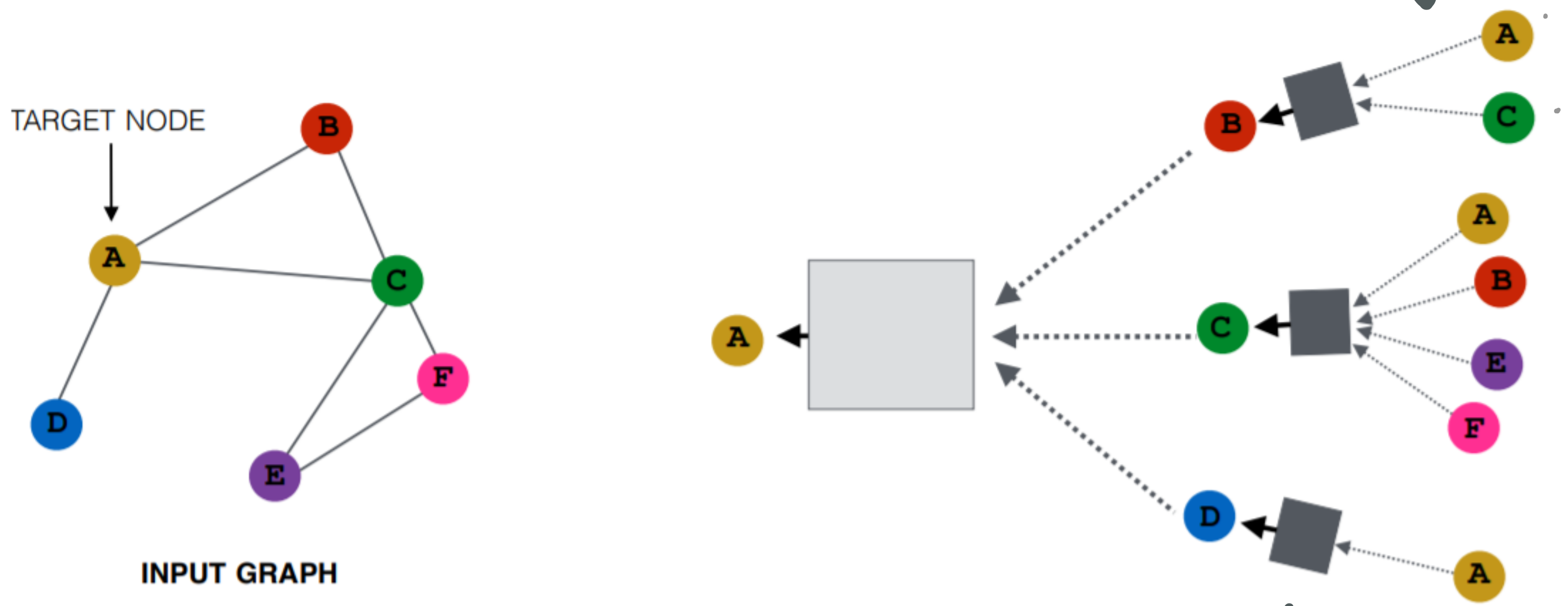
$\cap$ $\quad$ ← GALOIS → $\quad \cup$ $\quad$ ↑ Field extension $K[f^* = \sum_{ij} x_{ii} x_{ij}]$

$S_n \times S_{\binom{n}{2}}$ field invariant functions permuting diag → diag, off-diag → off-diag $\quad \leftrightarrow \quad f(\mathrm{set}(x_{ii}), \mathrm{set}(x_{ij})_{i \neq j})$

[Reduction to $n(d+1)$ generators is obtained using that $X^T X$ is low rank]

Message passing graph neural networks

message function

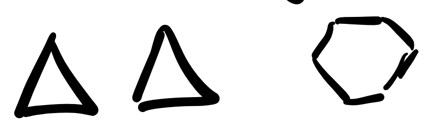


aggregation function

Equivariance is obtained by using the same message and aggregation function everywhere in the graph

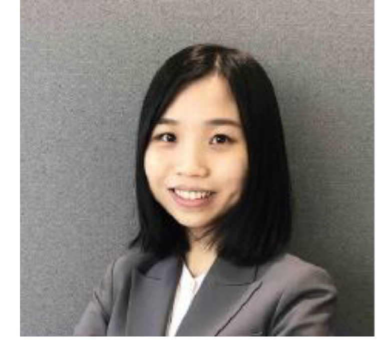
Expressive power

- Not every invariant / equivariant function can be expressed this way example :



- Connections to graph isomorphism : (Chen, Villar, Chen, Bruna, NeurIPS '19)
- They cannot count substructures (Chen, Chen, Villar, Bruna, NeurIPS '20)

Our approach : Break the symmetries



Teresa Huang

Approximate equivariant graph networks

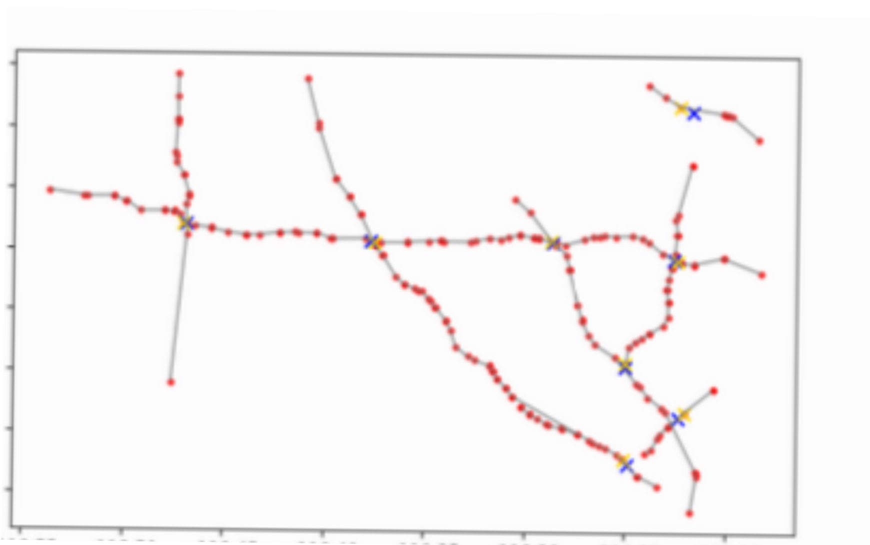
- When the graph is fixed decouple the action from the domain and the signal

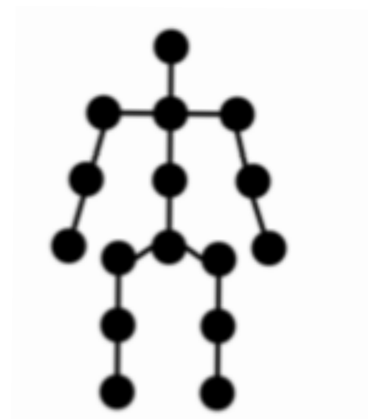
$$\pi G = (\pi A \pi^T, \pi X) \to \pi G = (A, \pi X)$$

$$\pi \in S_n \qquad \pi \in \mathcal{G} < S_n$$

$$H(\pi G) = \pi H(G) \qquad H(\pi G) \approx \pi H(G)$$

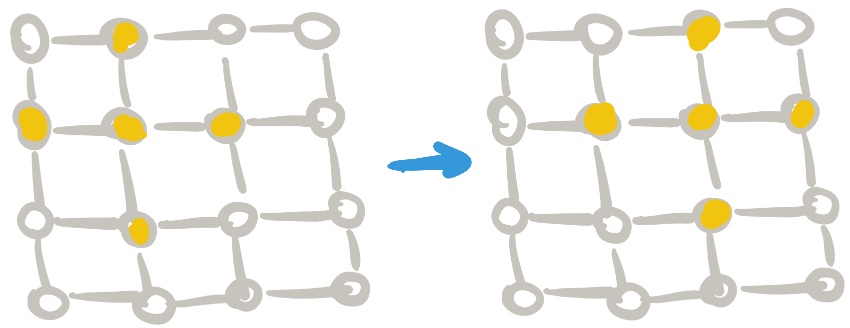
- Choose different subgroups $\mathcal{G}$





Intuition

CNN symmetry

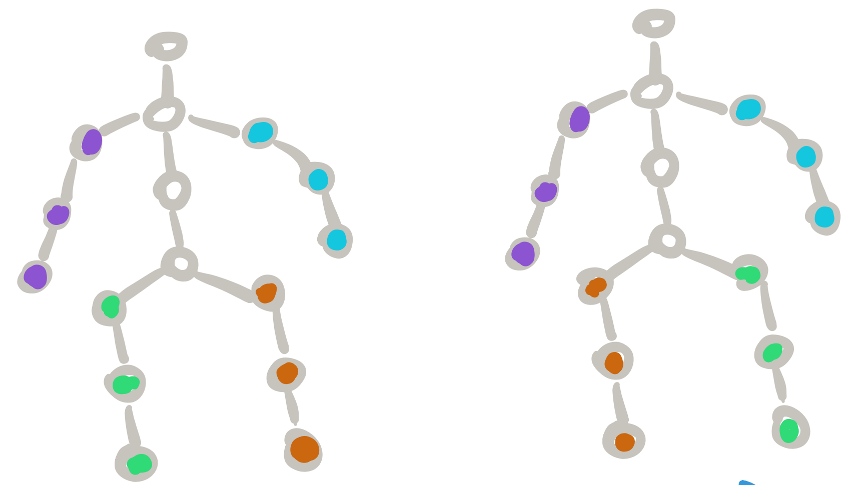


domain stays the same

signal (image) shifts

$$\underbrace{X}_{\text{grid signal (image)}} : \mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z} \to \mathbb{R}$$

GNN symmetry



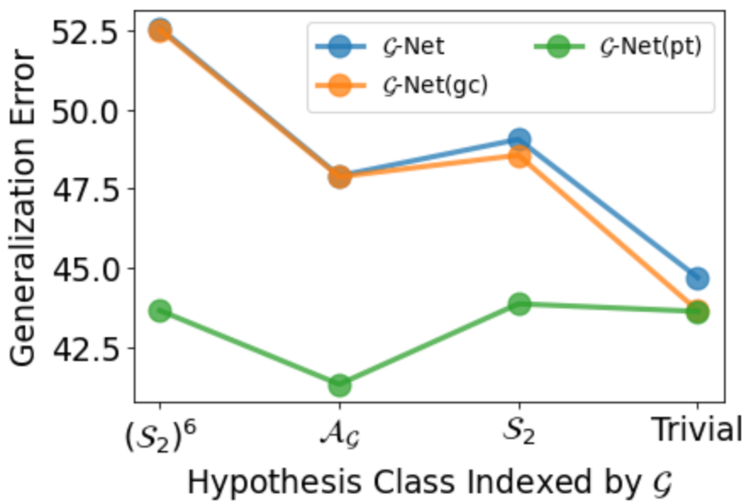
both domain (graph) and signal (color) move together (the object doesn't change)

$$G = (\underbrace{[n]}_{\text{nodes}}, \underbrace{A}_{\text{adj matrix}}) \qquad X : \underbrace{[n]}_{\text{graph signal}} \to \mathbb{R}^d$$

What we do: Fix the graph and implement different symmetries on the signal

- How to implement the approximate symmetries?
  - → Cluster graph nodes and share functions on clusters
    Group = $(S_{c_1} \times \dots S_{ck}) \rtimes A_G$
  - → Representation theory

- Bias - variance tradeoff



Lemma $G \leq S_n$ $\quad X \sim \mu$ $\quad S_n$ - invariant

$$y = f^*(x) + \xi \qquad f = \bar{f}_G + f_G^\perp$$

$$\underbrace{\Delta(f, \bar{f}_G)}_{\text{risk gap}} = \mathbb{E}_{x \sim \mu} \| y - f(x) \|^2 - \mathbb{E}_{x \sim \mu} \| y - \bar{f}_G(x) \|^2$$

$$= \underbrace{-2 \langle f^*, f_G^\perp \rangle_\mu}_{\text{mismatch}} + \underbrace{\| f_G^\perp \|_\mu^2}_{\text{constraint}}$$

Applications

- Explicit bias // variance tradeoff for linear regression with approx symmetry
- Approx guarantees for graph coarsening
- Examples showing imposing more symmetry may reduce the risk (bias ↑ variance ↓)

cf Eleredy & Zaidi ICML '21

Bias-variance for linear regression

$$
\begin{tikzcd}
\mathcal{X} \cong \mathbb{R}^N \arrow[r] \arrow[d, "\phi"'] & \mathcal{Y} \cong \mathbb{R}^N \arrow[d, "\psi"] \\
\mathcal{X} \arrow[r] & \mathcal{Y}
\end{tikzcd}
$$

$$Y = X\theta + \eta$$

$$\mathbb{E}(\eta\eta^T) = \sigma^2$$

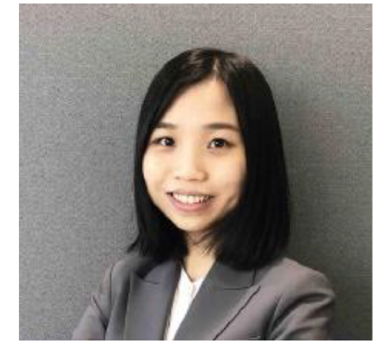
$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \ \|X\theta - Y\|^2$$

$\Psi_G(\hat{\theta})$ = projection of $\hat{\theta}$ onto equivariant maps

$$\Psi_G(\hat{\theta}) = \int_G \phi(g)\,\theta\,\psi(g)^{-1}\,dg$$

$$\mathbb{E}\left(\Delta(\hat{\theta}, \Psi_G(\hat{\theta}))\right) = \underbrace{-\sigma^2 \|\Psi_G^{\perp}(\theta)\|_F^2}_{\text{bias}^2} + \underbrace{\frac{\sigma^2 N^2 - (\chi_\psi | \chi_\phi)}{n - Nd - 1}}_{\text{variance}}$$

$$(\chi_\psi, \chi_\phi) = \int_G \chi_\psi(g)\,\chi_\phi(g)\,dg \leftarrow$$ dimension of space of linear equivariant functions



Teresa Huang

Numerical experiments

human pose estimation

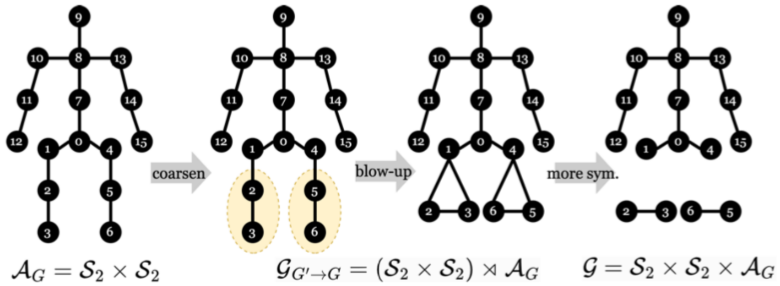
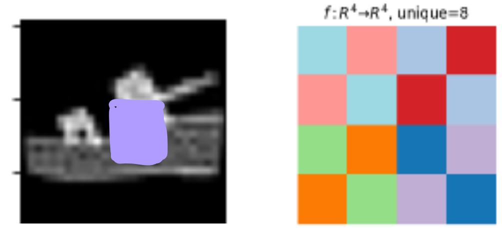


Figure 3: Human skeleton graph G, its coarsened graph G' (clustering leg joints), and blow-up of G'

$\mathcal{G}$-Net(gc+ew)	$\mathcal{S}_{16}$	$(\mathcal{S}_6)^2$	Relax-$\mathcal{S}_{16}$	$\mathcal{A}_G = (\mathcal{S}_2)^2$	Trivial
MPJPE ↓	42.55 ± 0.88	43.33 ± 0.99	$\mathbf{39.87 \pm 0.46}$	42.18 ± 0.49	41.60 ± 0.3
P-MPJPE ↓	34.48 ± 0.44	34.87 ± 0.48	$\mathbf{31.38 \pm 0.14}$	32.08 ± 0.20	31.69 ± 0.1

Image inpainting



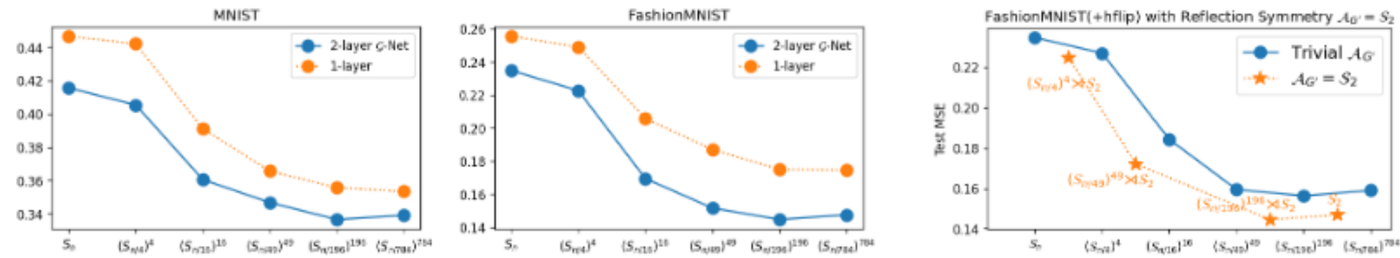
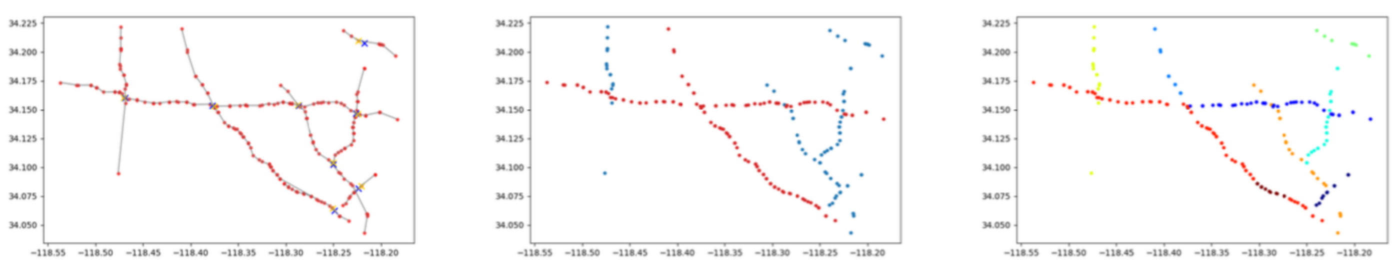


Figure 6: Bias-variance tradeoff via graph coarsening. Left:2-layer $\mathcal{G}$-Net(blue) and 1-layer linear $\mathcal{G}$-equivariant functions (orange), assuming the coarsened graph is asymmetric; Right: 2-layer $\mathcal{G}$-Net with both trivial and non-trivial coarsened graph symmetry.

MSE $(\times 1e^{-2})\downarrow$	$\mathcal{S}_{28^2} = \mathcal{S}_n$	$(\mathcal{S}_{14^2})^4 = (\mathcal{S}_{n/4})^4$	$(\mathcal{S}_{7^2})^{16} = (\mathcal{S}_{n/16})^{16}$	$(\mathcal{S}_{4^2})^{49} = (\mathcal{S}_{n/49})^{49}$	$(\mathcal{S}_{2^2})^{196} = (\mathcal{S}_{n/196})^{196}$	Trivial $= (\mathcal{S}_{n/784})^{784}$
MNIST	41.56 ± 0.16	40.53 ± 0.26	36.06 ± 0.24	34.68 ± 0.5	$\mathbf{33.67 \pm 0.07}$	33.92 ± 0.04
Fashion	23.48 ± 0.14	22.26 ± 0.02	16.94 ± 0.08	15.16 ± 0.1	$\mathbf{14.47 \pm 0.11}$	14.75 ± 0.11

Traffic flow prediction



(a) Our faithful traffic graph (b) Graph clustering (2 clusters) (c) Graph clustering (9 clusters)

$\mathcal{G}$-Net(gc)	$\mathcal{S}_N$	$\mathcal{S}_{c_1} \times \mathcal{S}_{c_2}$	$\mathcal{S}_{c_1} \times \ldots \times \mathcal{S}_{c_9}$
Graph G_s	3.173 ± 0.013	$\mathbf{3.150 \pm 0.008}$	3.204 ± 0.006
Graph G	3.106 ± 0.013	$\mathbf{3.092 \pm 0.008}$	3.174 ± 0.013

Conclusions

- (Approximate) symmetries give a good inductive bias for ML
  → Physical sciences (cosmology)
  → Engineering (self-supervised learning)

- Tools: invariant theory, representation theory

- Rethinking the roles of symmetries as "model selection"
  - bias-variance tradeoff on graph learning by relaxing symmetries

- Future work
  - Coordinate free models on vector fields
    → Ocean dynamics
  - Interactions with differential geometry

Thank you!

- Villar, Hogg, Storey-Fisher, Yao, Blum-Smith
  NeurIPS 2021
- Yao, Hogg, Storey-Fisher, Villar
  NeurIPS workshop ML 4 Physics 2021
- Villar, Yao, Hogg, Blum-Smith, Dumitrascu
  JMLR, 2023
- Blum-Smith, Villar
  Notices of the AMS 2023
- Villar, Hogg, Yao, Kevrekidis, Schölkopf
  TMLR 2024
- Gupta, Robinson, Lim, Villar, Jegelka
  ICLR 2024

Funding

NSF

SIMONS FOUNDATION

Office of Naval Research
ONR
Science & Technology

amazon

Questions?