Approximately Equivariant Networks for Imperfectly Symmetric Dynamics

Rui Wang [*1] Robin Walters [*2] Rose Yu [1]

Abstract

Incorporating symmetry as an inductive bias into neural network architecture has led to improvements in generalization, data efficiency, and physical consistency in dynamics modeling. Methods such as CNNs or equivariant neural networks use weight tying to enforce symmetries such as shift invariance or rotational equivariance. However, despite the fact that physical laws obey many symmetries, real-world dynamical data rarely conforms to strict mathematical symmetry either due to noisy or incomplete data or to symmetry breaking features in the underlying dynamical system. We explore approximately equivariant networks which are biased towards preserving symmetry but are not strictly constrained to do so. By relaxing equivariance constraints, we find that our models can outperform both baselines with no symmetry bias and baselines with overly strict symmetry in both simulated turbulence domains and real-world multi-stream jet flow.

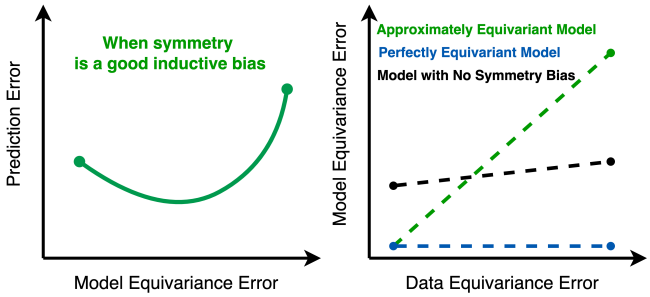

Figure 1: Left: When symmetry is a good inductive bias, prediction performance increases as equivariance or invariance is imposed in the model. But real-world data is very rarely perfectly symmetric, and so relaxing the strict constraint in equivariant networks to balance inductive bias and expressivity can further improve predictive performance. Right: Highly flexible models have trouble achieving zero equivariance error without the guide of appropriate symmetry biases when the data is symmetric. Perfectly equivariant models maintain zero equivariance error, which is overly restricted when data is not perfectly symmetric. An ideal model for real world dynamics should be approximately equivariant and automatically learn the correct amount of symmetry in the data.

1. Introduction

Symmetry and equivariance are fundamental to the success of deep learning (Bronstein et al., 2021). The canonical examples are translation invariance in convolutional layers (Fukushima & Miyake, 1982; LeCun et al., 1989; Krizhevsky et al., 2012), and permutation invariance in graph neural networks (Bruna et al., 2013; Battaglia et al., 2018; Maron et al., 2018). Recently, equivariant networks, which encode symmetry information in network architectures, have gained significant attention for modeling structured and complex data (Ravanbakhsh et al., 2017; Zaheer et al., 2017; Kondor & Trivedi, 2018; Cohen & Welling, 2016a; Worrall et al., 2017; Thomas et al., 2018; Cohen et al., 2018; Maron et al., 2020; Walters et al., 2021).

However, existing equivariant networks assume perfect symmetry in the data. The network is approximating a function that is strictly invariant or equivariant under a given group action. However, real-world data are very rarely perfectly symmetric. For example, in turbulence modeling, even though the governing equations of turbulence satisfy many different symmetries such as scale invariance (Holmes et al., 2012), effects such as varying external forces, certain boundary conditions, or the presence of missing values would break these symmetries to varying degrees. This significantly hinders the potential applications of equivariant networks. Approximately equivariant networks could outperform both strictly equivariant networks and highly flexible models in learning many dynamics in the real world, as shown in Figure 1.

Relaxing the rigid assumption in equivariant networks to balance inductive bias and expressivity in deep learning has been the pursuit of a few recent works. For example, Elsayed et al. (2020) showed that spatial invariance can be

[*]Equal contribution [1]University of California San Diego [2]Northeastern University. Correspondence to: Rui Wang <ruw020@ucsd.edu>.

Proceedings of the 39 [th] International Conference on Machine Learning, Baltimore, Maryland, USA, PMLR 162, 2022.

overly restrictive, and relaxing spatial weight sharing in standard convolution can improve image classification accuracy. d'Ascoli et al. (2021) enforce a convolutional inductive bias in self-attention layers at initialization to improve vision Transformers. Residual Pathway Priors (Finzi et al., 2021b) convert hard architectural constraints into soft priors by placing a higher likelihood on the "residual". The residual explains the difference between the structure in the data and the inductive bias encoded by an equivariant model. Wang et al. (2021) proposed Lift Expansion which factorizes the data into equivariant and nonequivariant components and models them jointly. Despite progress, a formal definition of approximate symmetry does not exist. While existing research focuses on translation symmetry, the rich groups of symmetry in high-dimensional dynamics learning problems are unexplored.

In this paper, we first define approximate symmetry. It gives rise to a new class of approximately equivariant networks that avoid stringent symmetry constraints while maintaining favorable inductive biases for learning. Specifically, we generalize the weight relaxation scheme originally proposed by (Elsayed et al., 2020). We study three symmetries that are common in dynamics: rotation $SO(2)$, scaling $R_{>0}$, and Euclidean $E(2)$. For group convolution, we relax the weight-sharing scheme by expressing the kernel as a weighted combination of multiple filter banks. For steerable CNNs, we introduce dependencies on the input to the kernel basis. We apply our approximate symmetry networks to the challenging problem of forecasting fluid flow and observe significant improvements for both synthetic and real-world datasets [1]. Our contributions include:

- We formally characterize the notion of approximate equivariance, which interpolates between no inductive bias and a strong inductive bias from equivariance.
- We introduce a new class of approximately equivariant networks for modeling imperfectly symmetric dynamics by relaxing equivariance constraints.
- We demonstrate that our approximately equivariant models can outperform baselines with no symmetry bias, baselines with overly strict symmetry, and SoTA approximately equivariant models in both simulated smoke simulations and real experimental jet flow data.

2. Mathematical Preliminaries

2.1. Equivariant Functions and Neural Networks

Equivariant neural networks incorporate an explicit symmetry constraint. They are typically employed when a priori knowledge, such as first principles from physics, imply the ground truth function also respects a symmetry.

Equivariance and Invariance. Formally, a function $f : X \to Y$ may be described as respecting the symmetry coming from a group G using the notion of equivariance. Assume that an input group representation $_{in}$ of G acts on X and an output representation $_{out}$ acts on Y. We say a function f is G-equivariant if

$$f(_{in}(g)(x)) = _{out}(g)f(x) \tag{1}$$

for all $x \in X$ and $g \in G$. The function f is G-invariant if $f(_{in}(g)(x)) = f(x)$ for all $x \in X$ and $g \in G$. This is a special case of equivariance for the case $_{out}(g) = 1$.

Strictly Equivariant Neural Networks. Given an equivariant $f : X \to Y$, learning can be accelerated by optimizing within a model class of functions ffg which are restricted to be equivariant. Since the composition of equivariant functions is again equivariant, in general a neural network will be strictly equivariant if all of its layers, linear, nonlinear, pooling, aggregation, and normalization, are equivariant. Most of the variation and challenge in this area is in designing trainable equivariant linear layers. Two strategies, involving weight sharing and weight tying, are G-convolution and G-steerable CNN. See Bronstein et al. (2021) for more details.

G-Equivariant Group Convolution. A G-equivariant group convolution (Cohen & Welling, 2016a) takes as input a c_{in}-dimensional feature map $f : G \to R^{c_{in}}$ and convolves it with a kernel $: G \to R^{c_{out}c_{in}}$ over a group G,

$$[f ?_G](g) = \sum_{h \in G} f(h) (g^{-1}h): \tag{2}$$

Here, we assume G finite, however, G may also be taken to be compact if the sum is replaced with an integral. Group convolution achieves equivariance by weight sharing since the kernel weight at $(g; h)$ depends only on $g^{-1}h$ and thus pairs with equal $g^{-1}h$ share weights.

G-Steerable Convolution. (Cohen & Welling, 2017) Let f be the input feature map $f : R^2 \to R^{c_{in}}$. Fix a subgroup $H \subset O(2)$, which acts on R^2 by matrix multiplication and on the input and output channel spaces $R^{c_{in}}$ and $R^{c_{out}}$ by representations $_{in} : G \to R^{c_{in}c_{in}}$ and $_{out} : G \to R^{c_{out}c_{out}}$ respectively.

We may convolve f with a matrix-valued kernel $: R^2 \to R^{c_{out}c_{in}}$. In practice, we discretize the input as $f : Z^2 \to R^{c_{in}}$ and kernel $: Z^2 \to R^{c_{out}c_{in}}$ and compute the H-action by interpolation after rotation. By (Weiler et al., 2018a), the standard 2D convolution $f ?_{Z^2}$ is H-equivariant and Z^2-translation equivariant when

$$(hx) = _{out}(h)(x)_{in}(h^{-1}); \ 8h \in H: \tag{3}$$

[1] We open-source our code https://github.com/Rose-STL-Lab/Approximately-Equivariant-Nets

Again, (hx) is computed based on the group and the choice of input and output representations. This linear constraint induces dependence in the weights, which is called weight tying. Solving for a basis of solutions to (3) gives an equivariant kernel basis $f_l g_{l=1}^{L}$ which can be combined using trainable coefficients $= \sum_{l=1}^{L} w_l l$ to learn any element of the solution space formed by (3).

2.2. Approximate Equivariance

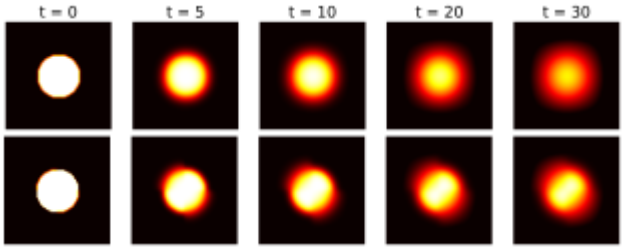

Figure 2: Simulated diffusion of heat in a metal plate with (top) uniform diffusion coefficient resulting in perfect symmetry and (bottom) slightly varying diffusion coefficient resulting in approximate symmetry.

Real-world dynamics data may not satisfy the strict equivariance as in (1). However, since many of the governing equations contain symmetry, the resulting system may still be approximately equivariant, as defined below. For example, while the heat equation itself is fully rotationally symmetric, in practice, imperfections in the thickness of the metal or the composition of the metal can lead to imperfect symmetry, as shown in Figure 2. Below, we give the formal definition of approximate symmetry:

Definition 2.1 (Approximate Equivariance). Let $f : X \to Y$ be a function and G be a group. Assume that G acts on X and Y via representations $_X$ and $_Y$. We say f is $-$approximately G-equivariant if for any $g \in G$,

$$\| f(_X(g)(x)) \quad _Y(g) f(x) \| \quad :$$

Note that strictly equivariant functions are $= 0$ approximately equivariant.

3. Approximately Equivariant Networks

Symmetry in equivariant networks is enforced by strict constraints on the weights. Here we propose relaxing weight-sharing and weight-tying to model approximate symmetries. Elsayed et al. (2020) showed that relaxing spatial weight sharing in standard convolution neural nets can improve image classification accuracy. Whereas 2D convolutions are shift equivariant, this relaxed 2D convolution is only approximately equivariant. Our method generalizes this approach to other symmetry groups, including rotation $SO(2)$, scaling $\mathbb{R}_{>0}$, and Euclidean $E(2)$. Specifically, we relax the strict weight-sharing and weight-tying constraints in both group convolution and steerable CNNs.

3.1. Relaxed Group Convolution

The G-equivariance of group convolution results from the shared kernel $(g^{-1}h)$ in (2). To relax this and consequently relax the G-equivariance, we replace the single kernel with a set of kernels $f \; _l g_{l=1}^{L}$. We define the new kernel as a linear combination of $_l$ with coefficients that vary with h. Thus, we introduce symmetry-breaking dependence on the specific pair $(g; h)$,

$$(g; h) = \sum_{l=1}^{L} w_l(h) \quad _l(g^{-1}h): \tag{4}$$

We define the relaxed group convolution by multiplication with as such

$$\begin{aligned} [f ?_G \;](g) &= \sum_{h \in G} f(h) \quad (g; h) \\ &= \sum_{h \in G} \sum_{l=1}^{L} f(h) w_l(h) \quad _l(g^{-1}h): \end{aligned} \tag{5}$$

By varying the number of kernels L, we can control the degree of equivariance. Small L imposes stronger symmetry, while large L gives more flexibility. In our experiments, we found that $L = 3$ gave the best prediction performance in most cases. The weights $w_i(h) \in \mathbb{R}$ and the kernels $_l(g^{-1}h) \in \mathbb{R}^{c_{out} c_{in}}$ can be learnt from data. Relaxed group convolution reduces to group convolution and is fully equivariant if and only if $g_1^{-1}h_1 = g_2^{-1}h_2$ implies $(g_1; h_1) = (g_2; h_2)$. In particular, this occurs if $w_l(h_1) = w_l(h_2)$ for all $h_1; h_2 \in G$ and for all $l = 1; \; ; L$.

In dynamics learning, we consider velocity vectors as inputs. To apply group convolution over the discrete rotation group C_n, we first lift these velocity vectors to feature maps $f : C_n \to \mathbb{R}$ as described in Walters et al. (2021) Table 1. Given $v = (a; b) \in \mathbb{R}^2$, for $i \in C_n$ we define $f(i) = ca \cos(2i=n) + cb \sin(2i=n)$ where $c \in \mathbb{R}$ is a trainable weight. This amounts to mapping the irreducible $_1$ representation of C_n to the regular representation. This process can also be extended to lifting velocity fields to features $f : C_4 n(\mathbb{Z}^2; +) \to \mathbb{R}$ over the group of discrete rotations and translations. As an additional advantage, the feature maps f are compatible with element-wise non-linearity.

3.2. Relaxed Steerable Convolution

Although relaxed group convolution does not require precomputing an equivariant kernel basis, it is limited to discrete (or compact) groups and is inefficient when the group order is large. Thus, we also propose relaxed steerable convolutions.

G-Steerable 2D Convolution: First, we explicitly write out the formula for G-steerable 2D convolution described by (3). Let $\{\kappa_l\}_{l=1}^{L}$ be an equivariant kernel basis of L non-trainable kernels that satisfy (3) for given input and output representations ρ_{in} and ρ_{out}. Denote $K = \{k_1, \dots, k_n\}$. Denote the input feature as $f_{in} : \mathbb{Z}^2 \to \mathbb{R}^{c_{in}}$, predetermined equivariant kernels $\kappa_l : K^2 \to \mathbb{R}^{c_{out} c_{in}}$, and a trainable weight tensor $w \in \mathbb{R}^{c_{out} c_{in} L}$. Then a G-steerable convolution produces an output $f_{out} = f_{in} \star_{\mathbb{Z}^2} : \mathbb{Z}^2 \to \mathbb{R}^{c_{out}}$ defined as

$$f_{out}(x) = \sum_{y \in \mathbb{Z}^2} \sum_{l=1}^{L} (w_l \kappa_l(y)) f_{in}(x + y) \quad \text{for a} \tag{6}$$

position $x \in \mathbb{Z}^2$ in the input. Here $\odot$ denotes element-wise product in $\mathbb{R}^{c_{out} c_{in}}$, and $y \in \mathbb{Z}^2$ is a spatial location in the kernel.

Relaxed G-Steerable 2D Convolution: We relax (6) and break symmetry by introducing a weight w that depends on y. As w is freely trainable, this breaks the strict positional dependence of κ_l imposed by (3). Formally, letting $w : K^2 \to \mathbb{R}^{c_{out} c_{in} L}$ be the weight, we define the relaxed steerable convolution $f_{out} = f_{in} \tilde{\star}_{\mathbb{Z}^2}$ by

$$f_{out}(x) = \sum_{y \in \mathbb{Z}^2} \sum_{l=1}^{L} (w_l(y) \kappa_l(y)) f_{in}(x + y). \tag{7}$$

When G is a rotation group and $k > 0$, we can define $w_l(y) = w_l(\theta)$, where $\theta = \arctan2(y)$. Since the weight depends only on the angle of the vector y, we use fewer parameters. To prevent the model from becoming overly relaxed, we initialize $w_l(y)$ equally for every y and penalize the value differences in $w_l(y)$ during training, which we describe in Section 3.3.

For the translation group, we can relax the steerable convolution by further allowing $w : \mathbb{Z}^2 \times K^2 \to \mathbb{R}^{c_{out} c_{in} L}$ to vary with the input position x as well:

$$f_{out}(x) = \sum_{y \in \mathbb{Z}^2} \sum_{l=1}^{L} (w_l(x; y) \kappa_l(y)) f_{in}(x + y). \tag{8}$$

However, the above equation is impractical as the space of the trainable weight is too large. We propose using a low-rank factorization of w to reduce dimensionality,

$$w_l(x; y) = \sum_{r=1}^{R} a_r(x) b_{r;l}(y)$$

where $a_r : \mathbb{Z}^2 \to \mathbb{R}$ and $b_{r;l} : K^2 \to \mathbb{R}^{c_{out} c_{in}}$. Then (8) becomes a combination of relaxed translation group convolution and relaxed steerable convolution.

3.3. Soft Equivariance Regularization

To encourage equivariance and prevent the model from becoming over-relaxed, we add regularization terms to the loss function on the symmetry-independent weights w during training. For relaxed group convolution, we add the following regularizer to constrain w in (4),

$$L_{gconv} = \sum_{i=1}^{L} \sum_{g;h \in G} k w_i(h) \quad w_i(g) k.$$

For relaxed steerable convolution, we impose the following loss term to prevent the $w : K^2 \to \mathbb{R}^{c_{out} c_{in} L}$ in (7) from varying too much across the kernel spatial domains,

$$L_{sconv} = \frac{\partial w(m; n)}{\partial m} + \frac{\partial w(m; n)}{\partial n} :$$

The hyperparameter ϵ does not directly control how equivariant the model is, it only a places a equivariance prior on the model to be as equivariant as possible given the data.

3.4. Other Alternatives for Approximate Symmetry

We also explored three alternative ways of building approximately equivariant models.

Lift expansion. Wang et al. (2021) proposed *Lift Expansion* for modeling partial symmetry, in which the input space can be factorized into an equivariant subspace and a non-equivariant subspace. The model uses a non-equivariant encoder that is tiled across the equivariant dimensions of the feature map as additional channels in equivariant neural nets. This method can also model approximate symmetry when both the non-equivariant encoder and the main equivariant backbone are fed with the same input. The encoder can extract non-equivariant features that are then treated as having a trivial representation type and included in the main equivariant model to break perfect equivariance. Note that while treating the output data as having a trivial representation type enforces invariance, treating the input data this way imposes no constraints.

Constrained locally connected neural nets (CLCNN). Another way of building approximately equivariant models is using a very flexible model while imposing soft equivariance constraints on the kernels. We use a locally connected neural network that has the same locality property as convolution but the weights are not shared across the spatial domain (Wadekar, 2019). Thus, it does not have translation equivariance and employs many more parameters than convolution. Suppose $\kappa : \mathbb{Z}^2 \times K^2 \to \mathbb{R}^{c_{out} c_{in}}$ is the filter bank. In addition to the prediction loss, we use the equivari-ant kernel constraint (3) in the objective with a hinge loss instead of solving the constraints explicitly before training:

$$L_{hinge} = \sum_{h \in G} k \rho_{out}(h) \kappa(x) \rho_{in}(h^{-1}) \quad (hx) k:$$

Combination of non-equivariant and equivariant layers. We also build models that begin with non-equivariant layers

followed by equivariant layers. The early layers of the model map observations with approximate symmetries to a space with an explicit symmetry actions.

3.5. Equivariant Error Analysis

Our hypothesis is that if the ground truth function f is approximately equivariant, then a model class with a similar degree of approximate equivariance would better approximate f than a strictly equivariant class or a class without bias towards symmetry.

We define equivariance error, which quantifies how much a function f is approximately equivariant.

Definition 3.1 (Equivariance Error). Let $f : X \to Y$ be a function and G be a group. Assume that G acts on X and Y via representation $_X$ and $_Y$. Then the equivariance error of f is

$$kf k_{EE} = \sup_{x;g} kf(_X(g)(x)) \quad _Y(g)f(x)k:$$

That is, f is -approximately equivariant if and only if $kf k_{EE} < :$

We note that a strictly equivariant model cannot perfectly learn an approximately equivariant function. As stated by the following proposition, such a model would make errors at least proportional to the equivariance error. This motivates our choice to use the model class of approximately equivariant networks.

Proposition 3.2. Let $f : X \to Y$ where G acts on X and Y by $_X$ and $_Y$ which are norm-preserving. Assume f is approximately equivariant with $kf k_{EE} \quad 0$. Assume f is a G-equivariant approximator for f. Then there exists $x_0 \ 2 \ X$ such that

$$kf(x_0) \quad f(x_0)k \quad kf k_{EE}=2:$$

For simplicity, we assume representations which are norm preserving, as with rotations, reflections, and permutations, although this assumption can be removed by inserting a factor to account for the operator norm kgk.

By similar logic, we can also show that given a model class which contains -approximately equivariant functions for varying , the equivariance error of the approximator will converge to the equivariance error of the ground truth function as they converge in model error. Although the supremum norm is used for equivariance and model error, the result holds for other norms as well.

Proposition 3.3. Let ffg be an approximately equivariant model class with varying $kfk_{EE} \ 2 \ R_0$. Assume a G-invariant norm. Let $f : X \to Y$ be a function with $kf k_{EE} =$. Assume $kf \quad fk_1 \quad c$. Then $jkfk_{EE} \quad kfk_{EE}j \quad 2c + :$

The proofs can be found in Appendix A.3.

4. Related Work

4.1. Equivariance and Invariance

Symmetry has long been implicitly used in DL to design networks with known invariances and equivariances. Convolutional neural networks enabled breakthroughs in computer vision by leveraging translational equivariance (Zhang, 1988; LeCun et al., 1989; Zhang et al., 1990). Similarly, re-current neural networks (Rumelhart et al., 1986; Hochreiter & Schmidhuber, 1997), graph neural networks (Maron et al., 2019; Satorras et al., 2021), and capsule networks (Sabour et al., 2017; Hinton et al., 2011) all impose symmetries. Equivariant DL models have achieved remarkable success in learning image data (Cohen et al., 2019; Weiler & Cesa, 2019b; Cohen & Welling, 2016a; Chidester et al., 2018; Lenc & Vedaldi, 2015; Kondor & Trivedi, 2018; Bao & Song, 2019; Worrall et al., 2017; Cohen & Welling, 2016b; Finzi et al., 2020; Weiler et al., 2018b; Dieleman et al., 2016; Ghosh & Gupta, 2019; Sosnovik et al., 2020b).

There is also a deep connection between symmetries and physics. Noether's law gives a correspondence between conserved quantities and groups of symmetries. Thus, the study of equivariant nets in learning dynamical systems has gained popularity. Walters et al. (2021) proposed a rotationally-equivariant continuous convolution model for improved pedestrian and vehicle trajectory predictions. Holderrieth et al. (2021) introduced Steerable Conditional Neural Processes for learning stochastic processes in physics that have invariances and equivariances. Wang et al. (2020b) designed fully equivariant models with respect to symmetries of scaling, rotation, and uniform motion in physical dynamics.

But most dynamics in real world do not have perfect symmetry and thus the proposed models might be overly-constrained. Recently, some work explored the idea of building approximately equivariant networks (van der Ouderaa et al., 2022; Romero & Lohit, 2021). Elsayed et al. (2020) showed that spatial invariance may be overly restrictive and relaxing the spatial weight sharing could outperform both convolution and local connectivity. Finzi et al. (2021b) proposed a mechanism that sums equivariant and non-equivariant MLP layers for modeling soft equivariances, but it cannot handle large data like images or high-dimensional physical dynamics due to the number of weights in the fully connected layers. Our method in contrast has more efficient convolutional layers and uses relaxed constraints to achieve approximate equivariance.

4.2. Learning Dynamical Systems

There is an increasing number of works in modeling dynamical systems with deep learning (Shi et al., 2017; Chen

et al., 2018; Kolter & Manek, 2019; Azencot et al., 2020; Xie et al., 2018; Tompson et al., 2017; Pfaff et al., 2021). An essential topic is physics-guided deep learning (Raissi et al., 2017; Lutter et al., 2018; de Bezenac et al., 2018; Li et al., 2021; Wang et al., 2020a) which integrates inductive biases from physical systems to improve learning. For example, Wang et al. (2020a) proposed a hybrid model marrying the RANS-LES coupling method and the custom-designed U-net. Greydanus et al. (2019) and Cranmer et al. (2020) build models on Hamiltonian and Lagrangian mechanics that respect conservation laws. Guen et al. (2021) proposed a framework that augments physics-based models with deep data-driven models for forecasting dynamical systems. In this work, we encode approximate symmetries as inductive biases into DL models to improve dynamics prediction without over-constraining the representation power.

5. Experiments

Baselines We compare with several state-of-the-art (SoTA) methods from those without symmetry bias to perfect symmetry and SoTA approximately symmetric models.

- MLP: multi-layer perceptrons, an non-equivariant baseline with a weaker inductive bias than convolution neural nets.
- ConvNet: standard convolutional neural nets that have full translation symmetry.
- Equiv: fully equivariant convolutional models. It is same as ConvNet for translation symmetry. We use E2CNN (Weiler & Cesa, 2019a) for rotation and SESN (Sosnovik et al., 2020a) for scaling symmetry.
- Rpp (Finzi et al., 2021b): Residual Pathway Priors, a SOTA approximate equivariance model that sums up the outputs from equivariant and non-equvariant layers while posing constraints on the non-equivariant layer in the loss function. We use the combination of MLP and ConvNet for translation, ConvNet and E2CNN for rotation, and ConvNet and SESN for scaling.
- Combo: models that start with non-equivariant layers followed by equivariant layers, discussed in Section 3.4.
- CLCNN: locally connected neural networks with equivariance constraints imposed in the loss function.
- Lift (Wang et al., 2021): Lift expansion for modeling partial symmetry. Both the encoder and the main equivariant backbone are fed with the same input.

EMLP (Finzi et al., 2021a) is also a SoTA equivariant model, but it cannot handle large data like images as stated in the paper, so we do not include it as a baseline.

Experiments Setup All models are trained to perform forward prediction of raw velocity fields given historical data. For all datasets, we use a sliding window approach to generate sequence samples. We perform a grid hyperparameter search as shown in Table 3, including learning rate, batch size, hidden dimension, number of layers, number of prediction errors steps for training. We also tune the number of filter banks for group convolution-based models and the coefficient of weight constraints for relaxed weight-sharing models. The input length is fixed as 10. Meanwhile, we make sure that the total number of trainable parameters for every model is less than 10^7 for a fair comparison.

We test all models under two scenarios. For test-future, we train and test on the same tasks but in different time steps. For test-domain, we train and test on different simulations/regions with an 80%-20% split. All models are trained to make the prediction of the next step given the previous steps as input. The first scenario evaluates how well the models can extrapolate into the future for the same task. The second scenario estimates the capability of the models to generalize across different simulations/regions. We forecast in an autoregressive manner to generate multi-step predictions during inference and evaluate them based on 20-step prediction RMSEs. All results are averaged over 3 runs with random initialization.

5.1. Experiments on Synthetic Smoke Plumes

Data Description: The synthetic 6464 2-D smoke datasets are generated by PhiFlow (Holl et al., 2020) and contain smoke simulations with different initial conditions and external forces. We explore three symmetry groups: 1) Translation: 35 smoke simulations with different inflow positions. We also horizontally split the entire domain into two separate sub-domains that have different buoyant forces. Although the inflow positions are translation equivariant, the closed boundary and the two different buoyant forces would break the equivariance. 2) Rotation: 40 simulations with different inflow positions and buoyant forces. Both the inflow location and the direction of the buoyant forces have a perfect rotation symmetry with respect to C_4 group, but the buoyancy factor varies with the inflow positions to break the rotation symmetry. 3) Scaling: It contains 40 simulations generated with different spatial steps x and temporal steps t. And the buoyant force varies across the simulations to break the scaling symmetry.

Prediction Performance: Table 1 shows the prediction RMSEs in three synthetic smoke plume datasets with different approximate symmetries by our proposed models and baselines. CNNs are translation-equivariant because CNNs are inherently group convolution, where the group is the translation group, so we do not have a relaxed steerable model for translation. We can see, on the approximate translation dataset, our relaxed group convolution (RGroup) significantly outperforms baselines on both test sets. And for rotation and scaling, the proposed relaxed steerable convolution (RSteer) always achieves the lowest RMSE and

Table 1: Prediction RMSE on three synthetic smoke plume datasets with approximate symmetries. Our proposed RGroup and RSteer methods demonstrate competitive performance. Future means testing data lies in the future time of the training data. Domain means training and test data are from different spatial domain.

Model		MLP	Conv	Equiv	Rpp	Combo	CLCNN	Lift	RGroup	RSteer
Translation	Future	$1.56_{0.08}$	—–	$0.94_{0.02}$	$0.92_{0.01}$	$1.02_{0.02}$	$0.92_{0.01}$	$0.87_{0.03}$	$0.71_{0.01}$	—– $0.68_{0.05}$
	Domain	$1.79_{0.13}$	—–			$0.93_{0.01}$	$0.98_{0.01}$	$0.89_{0.01}$	$0.70_{0.00}$ $0.62_{0.02}$	—–
Rotation	Future	$1.38_{0.06}$	$1.21_{0.01}$	$1.05_{0.06}$	$0.96_{0.10}$	$1.07_{0.00}$	$0.96_{0.05}$	$0.82_{0.08}$	$0.82_{0.01}$ Domain	$0.80_{0.00}$
			$1.34_{0.03}$	$1.10_{0.05}$	$0.76_{0.02}$	$0.83_{0.01}$	$0.82_{0.02}$	$0.84_{0.10}$	$0.68_{0.09}$ $0.73_{0.02}$	$0.67_{0.01}$
Scaling	Future	$2.40_{0.02}$	$0.83_{0.01}$	$0.75_{0.03}$	$0.81_{0.09}$	$0.78_{0.04}$	$1.03_{0.01}$	$0.85_{0.01}$	$0.76_{0.04}$ Domain	$0.70_{0.01}$
			$1.81_{0.18}$	$0.95_{0.02}$	$0.87_{0.02}$	$0.86_{0.05}$	$0.85_{0.01}$	$0.83_{0.05}$	$0.77_{0.02}$ $0.86_{0.12}$	$0.73_{0.01}$

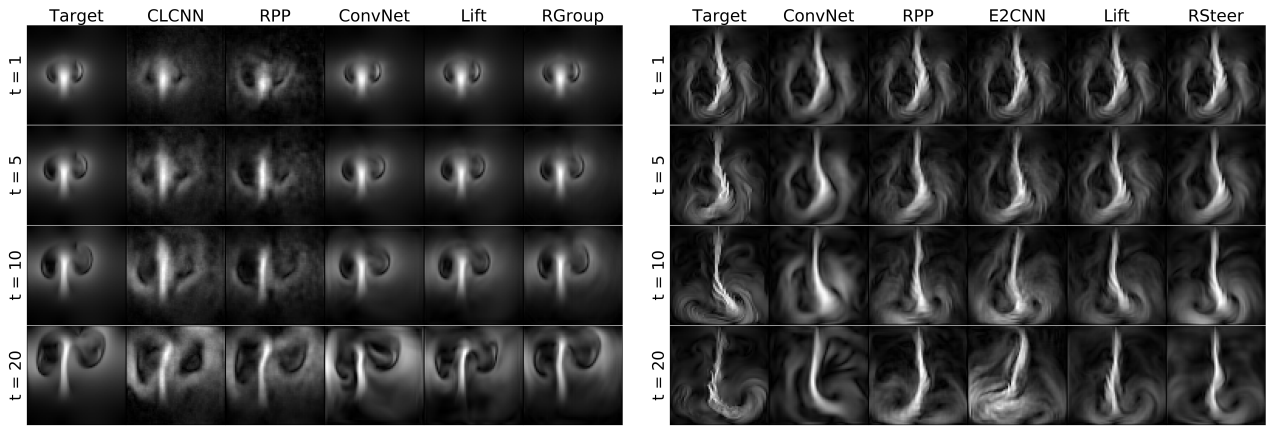

Figure 3: Target (ground truth) and model predictions comparison at time step 1, 5, 10, 20 for smoke simulation with approximate translation (left) and rotation (right) symmetries.

RGroup can outperform most baselines.

Figure 3 shows the target and predictions of our proposed models and the best baselines at time step 1, 5, 10, 20 for smoke simulation with approximate translation (left) and rotation (right) symmetries. From the shape and frequency of the flows, predictions from our approximately equivariant models are much closer to the target than the baselines. Moreover, we can see that E2CNN predicts the smoke flowing to the wrong direction at time step 20, which could be a consequence of over-constraining from equivariance.

Figure 7 in Appendix A.4 shows the prediction performance of a scaling RSteer model trained with different regularization parameter discussed in the section 3.3. We see that the soft equivariance regularization can further improve its prediction performance on both test sets but large may also hinder its learning.

5.2. Learning Different Levels of Equivariance

We use PhiFlow (Holl et al., 2020) to create 10 small smoke plume datasets with different levels of rotational equivariance. In each data set, both the inflow location and the direction of the buoyant forces have a perfect rotation symmetry with respect to the C_4 group. By varying the amount of difference in buoyant force between simulations with different inflow positions, we can control the amount of equivariance error in the data. The data equivariance error of each dataset is the mean absolute error between the simulations after they are all rotated back to the same inflow position.

We trained two-layer ConvNet, E2CNN and our relaxed rotation equivariant steerable convolution RSteerR on these 10 datasets. We calculate the equivariance error of each well-trained model based on Definition 2.1, where G = C_4 and the norm is L1 norm. From Figure 4, we see that E2CNN always has zero equivariance error due to the overly restrictive symmetry constraint even if the data does not have perfect symmetry. And our RSteerR can learn dif-ferent levels of equivariance in the data more accurately than other baselines. Since the prediction errors are not zeros, the equivariance errors in the model and data are not the same. This experiment demonstrates that our proposed methods based on relaxed weight sharing can learn the cor-rect amount of inductive biases from data while avoiding the stringent symmetry constraints.

Table 2: Prediction RMSEs on experimental jet flow data for different models. The proposed RSteer and RSteer are designed for the corresponding assumed symmetry group. RSteerTR and RSteerTS combines relaxed translation group convolution with relaxed rotation and scale steerable convolution.

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

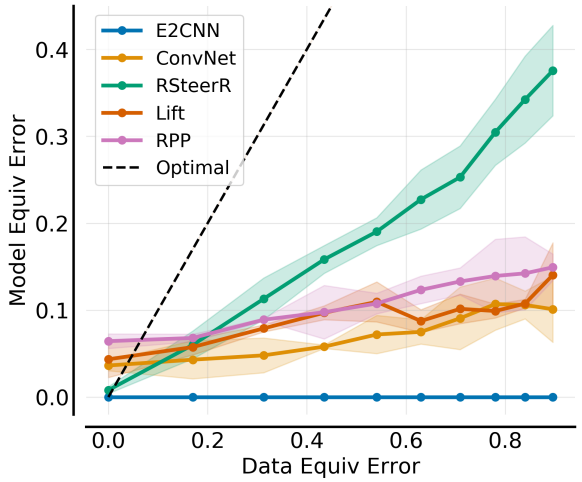

Figure 4: Model equivariance errors vs. data equivariance errors on synthetic smoke plume with different levels of rotational equivariance. We see that our RSteer can learn different levels of equivariance in the data much more accurately than other baselines.

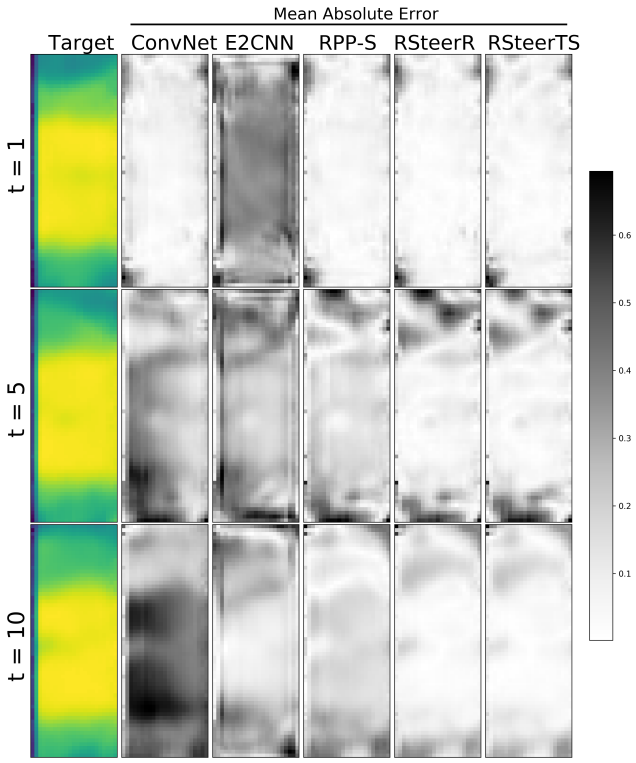

Figure 5: Target jet flow velocity norm fields and the prediction errors (MAE) of different models over 10 time steps.

5.3. Experiments on Experimental Jet Flow Data

Data Description. We use real experimental data on 2D turbulent velocity in NASA multi-stream jets that are measured using time-resolved particle image velocimetry (Bridges & Wernet, 2017). Figure 6 in the Appendix A.4 visualizes the measurement system of the jet flow. The white boxes show fields of view acquired on the streamwise plane at the jet centerline for multi-stream flows. There are three vertical stations at each axial location/white box, as illustrated by the pink lines. In other words, the dataset was acquired by 24 different stations at different locations. Since the data collected at the different locations are not acquired concur-rently, we do not have the complete velocity fields of entire jet flows at each time step. Thus, we trained and test models on 24 6223 sub-regions of jet flows.

Prediction Performance. We compare three equivariant models, three best-performing approximately equivariant baselines in the previous experiment as well as our proposed relaxed steerable convolution and relaxed group convolution. Table 2 shows the prediction RMSEs on the jet flow dataset, and we group the results by each symmetry in the table. For each symmetry, our models based on relaxed weight sharing achieve lower errors than not only the fully equivariant model but also approximately equivariant baselines. We also experimented with combining relaxed translation group convolution with relaxed rotation and scale steerable convolution, which correspond to RSteerTR and RSteerTS respectively in the table. We observe that RSteerTR outperform both RGroup with relaxed group convolution and RSteer with relaxed steerable convolution. This implies relaxing more than one equivariance constraint can potentially lead to even better performance. Figure 5 visualizes the target and mean absolute errors between model jet flow predictions and the ground truth (target), and we can see that our relaxed steerable CNNs achieve the lowest errors.

We also performed experiments on real-world ocean dynamics. We observe that all models have very close prediction performance after fine-tuning. Unlike the smoke plume or jet flow experiments in which our model's approximate equivariance bias better matched the ground truth than either the strictly equivariant model or the non-equivariant model, in this case all three levels of equivariance bias perform similarly. We hypothesize that, while strict rotational symmetry is a feature of ocean currents, imposing it as a strict inductive bias does not provide a significant advantage over the baseline CNN. Therefore, imposing a soft approximate equivariance bias also does not provide an advantage. For additional results, see Appendix A.2.

6. Discussion

We propose a new class of approximately equivariant networks that avoid stringent symmetry constraints to better fit real-world scenarios. Our methods strike a good balance between inductive biases and model flexibility by relaxing the weight-sharing and weight-tying schemes in group convolution and steerable convolution. Based on the experiments on smoke plume simulations and real-world jet flow data, we observe that our proposed approximate equivariant networks can outperform many state of the art baselines with no symmetry bias or with overly strict symmetry constraints. Future work includes applying our relaxed weight sharing design to graph neural networks and theoretical analysis for approximately equivariant networks, including universal approximation and generalization.

Acknowledgements

This work was supported in part by U.S. Department Of Energy, Office of Science, U. S. Army Research Office under Grant W911NF-20-1-0334, Google Faculty Award, and NSF Grant #2134274. R. Walters is supported by a Postdoctoral Fellowship from the Roux Institute and NSF grants #2107256 and #2134178. We also thank Nicholas J. Georgiadis and Mark P. Wernet and the NASA Glenn Research Center for providing the jet flow data and the instructions. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

References

Azencot, O., Erichson, N., Lin, V., and Mahoney, M. W. Forecasting sequential data using consistent koopman autoencoders. In International Conference on Machine Learning, 2020.

Bao, E. and Song, L. Equivariant neural networks and equivarification. arXiv preprint arXiv:1906.07172, 2019.

Battaglia, P. W., Hamrick, J. B., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V., Malinowski, M., Tacchetti, A., Raposo, D., Santoro, A., Faulkner, R., et al. Relational inductive biases, deep learning, and graph networks. arXiv preprint arXiv:1806.01261, 2018.

Bridges, J. and Wernet, M. P. Measurements of turbulent convection speeds in multistream jets using time-resolved piv. In 23rd AIAA/CEAS aeroacoustics conference, 2017.

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

Bruna, J., Zaremba, W., Szlam, A., and LeCun, Y. Spectral networks and locally connected networks on graphs. arXiv preprint arXiv:1312.6203, 2013.

Chen, R. T., Rubanova, Y., Bettencourt, J., and Duvenaud, D. Neural ordinary differential equations. In Proceedings of the 32nd International Conference on Neural Information Processing Systems, pp. 6572–6583, 2018.

Chidester, B., Do, M. N., and Ma, J. Rotation equivariance and invariance in convolutional neural networks. arXiv preprint arXiv:1805.12301, 2018.

Cohen, T. and Welling, M. Group equivariant convolutional networks. In Proceedings of the International Conference on Machine Learning (ICML), 2016a.

Cohen, T., Geiger, M., and Weiler, M. A general theory of equivariant cnns on homogeneous spaces. arXiv preprint arXiv:1811.02017, 2018.

Cohen, T. S. and Welling, M. Steerable CNNs. arXiv preprint arXiv:1612.08498, 2016b.

Cohen, T. S. and Welling, M. Steerable cnns. In International Conference on Learning Representations, 2017.

Cohen, T. S., Weiler, M., Kicanaoglu, B., and Welling, M. Gauge equivariant convolutional networks and the icosahedral CNN. In Proceedings of the 36th International Conference on Machine Learning (ICML), volume 97, pp. 1321–1330, 2019.

Cranmer, M., Greydanus, S., Hoyer, S., Battaglia, P., Spergel, D., and Ho, S. Lagrangian neural networks. ArXiv, abs/2003.04630, 2020.

d'Ascoli, S., Touvron, H., Leavitt, M., Morcos, A., Biroli, G., and Sagun, L. Convit: Improving vision transformers with soft convolutional inductive biases. ICML, 2021.

de Bezenac, E., Pajot, A., and Gallinari, P. Deep learning for physical processes: Incorporating prior scientific knowledge. In International Conference on Learning Representations, 2018. URL https://openreview.net/forum?id=By4HsfWAZ.

Dieleman, S., Fauw, J. D., and Kavukcuoglu, K. Exploiting cyclic symmetry in convolutional neural networks. In International Conference on Machine Learning (ICML), 2016.

Elsayed, G. F., Ramachandran, P., Shlens, J., and Kornblith, S. Revisiting spatial invariance with low-rank local connectivity. In Proceedings of the 37th International Conference of Machine Learning (ICML), 2020.

Finzi, M., Stanton, S., Izmailov, P., and Wilson, A. G. Generalizing convolutional neural networks for equivariance to lie groups on arbitrary continuous data. arXiv preprint arXiv:2002.12880, 2020.

Finzi, M., Welling, M., and Wilson, A. G. A practical method for constructing equivariant multilayer perceptrons for arbitrary matrix groups. International Conference on Machine Learning, 2021a.

Finzi, M. A., Benton, G., and Wilson, A. G. Residual pathway priors for soft equivariance constraints. In Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W. (eds.), Advances in Neural Information Processing Systems, 2021b. URL https://openreview.net/forum?id=k505ekjMzww.

Fukushima, K. and Miyake, S. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In Competition and cooperation in neural nets, pp. 267–285. Springer, 1982.

Ghosh, R. and Gupta, A. K. Scale steerable filters for locally scale-invariant convolutional neural networks. arXiv preprint arXiv:1906.03861, 2019.

Greydanus, S., Dzamba, M., and Yosinski, J. Hamiltonian neural networks. ArXiv, abs/1906.01563, 2019.

Guen, V. L., Yin, Y., Dona, J., Ayed, I., de B'ezenac, E., Thome, N., and Gallinari, P. Augmenting physical models with deep networks for complex dynamics forecasting. Journal of Statistical Mechanics: Theory and Experiment, 2021, 2021.

Hinton, G. E., Krizhevsky, A., and Wang, S. D. Transforming auto-encoders. In International conference on artificial neural networks, pp. 44–51. Springer, 2011.

Hochreiter, S. and Schmidhuber, J. Long short-term memory. Neural computation, 9(8):1735–1780, 1997.

Holderrieth, P., Hutchinson, M. J., and Teh, Y. W. Equivariant learning of stochastic fields: Gaussian processes and steerable conditional neural processes. In Meila, M. and Zhang, T. (eds.), Proceedings of the 38th International Conference on Machine Learning, volume 139 of Proceedings of Machine Learning Research, pp. 4297–4307. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/holderrieth21a.html.

Holl, P. M., Um, K., and Thuerey, N. phiflow: A differentiable pde solving framework for deep learning via physical simulations. In Workshop on Differentiable Vision, Graphics, and Physics in Machine Learning at NeurIPS, 2020.

Holmes, P., Lumley, J. L., Berkooz, G., and Rowley, C. W. Turbulence, coherent structures, dynamical systems and symmetry. Cambridge university press, 2012.

Kolter, J. Z. and Manek, G. Learning stable deep dynamics models. In Advances in Neural Information Processing Systems (NeurIPS), volume 32, pp. 11128–11136, 2019.

Kondor, R. and Trivedi, S. On the generalization of equivariance and convolution in neural networks to the action of compact groups. In Proceedings of the 35th International Conference on Machine Learning (ICML), volume 80, pp. 2747–2755, 2018.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25: 1097–1105, 2012.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D. Backpropagation applied to handwritten zip code recognition. Neural computation, 1(4):541–551, 1989.

Lenc, K. and Vedaldi, A. Understanding image representations by measuring their equivariance and equivalence. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 991–999, 2015.

Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., and Anandkumar, A. Fourier neural operator for parametric partial differential equations. International Conference on Learning Representations, 2021.

Lutter, M., Ritter, C., and Peters, J. Deep lagrangian networks: Using physics as model prior for deep learning. In International Conference on Learning Representations, 2018.

Madec, G. Nemo ocean engine. 2008.

Maron, H., Ben-Hamu, H., Shamir, N., and Lipman, Y. Invariant and equivariant graph networks. arXiv preprint arXiv:1812.09902, 2018.

Maron, H., Ben-Hamu, H., Shamir, N., and Lipman, Y. Invariant and equivariant graph networks. arXiv preprint arXiv:1812.09902, 2019.

Maron, H., Litany, O., Chechik, G., and Fetaya, E. On learning sets of symmetric elements. In International Conference on Machine Learning, pp. 6734–6744. PMLR, 2020.

Pfaff, T., Fortunato, M., Sanchez-Gonzalez, A., and Battaglia, P. Learning mesh-based simulation with graph networks. In International Conference on Learning Representations, 2021. URL https://openreview.net/forum?id=roNqYL0_XP.

Raissi, M., Perdikaris, P., and Karniadakis, G. E. Physics informed deep learning (part i): Data-driven solutions of nonlinear partial differential equations. arXiv preprint arXiv:1711.10561, 2017.

Ravanbakhsh, S., Schneider, J., and Poczos, B. Equivariance through parameter-sharing. In International Conference on Machine Learning, pp. 2892–2901. PMLR, 2017.

Romero, D. W. and Lohit, S. Learning equivariances and partial equivariances from data. arXiv preprint arXiv:2110.10211, 2021.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. Learning representations by back-propagating errors. nature, 323(6088):533–536, 1986.

Sabour, S., Frosst, N., and Hinton, G. E. Dynamic routing between capsules. arXiv preprint arXiv:1710.09829, 2017.

Satorras, V. G., Hoogeboom, E., and Welling, M. E(n) equivariant graph neural networks. arXiv preprint arXiv:2102.09844, 2021.

Shi, X., Gao, Z., Lausen, L., Wang, H., Yeung, D., Wong, W., and chun Woo, W. Deep learning for precipitation nowcasting: A benchmark and a new model. In Advances in neural information processing systems, 2017.

Sosnovik, I., Szmaja, M., and Smeulders, A. Scale-equivariant steerable networks. In International Conference on Learning Representations, 2020a. URL https://openreview.net/forum?id=HJgpugrKPS.

Sosnovik, I., Szmaja, M., and Smeulders, A. Scale-equivariant steerable networks. In International Conference on Learning Representations, 2020b. URL https://openreview.net/forum?id=HJgpugrKPS.

Thomas, N., Smidt, T., Kearnes, S., Yang, L., Li, L., Kohlhoff, K., and Riley, P. Tensor field networks: Rotation-and translation-equivariant neural networks for 3d point clouds. arXiv preprint arXiv:1802.08219, 2018.

Tompson, J., Schlachter, K., Sprechmann, P., and Perlin, K. Accelerating Eulerian fluid simulation with convolutional networks. In ICML'17 Proceedings of the 34th International Conference on Machine Learning, volume 70, pp. 3424–3433, 2017.

van der Ouderaa, T. F., Romero, D. W., and van der Wilk, M. Relaxing equivariance constraints with non-stationary continuous filters. arXiv preprint arXiv:2204.07178, 2022.

Wadekar, S. N. Locally connected neural networks for image recognition. Purdue University Graduate School, Dec 2019. doi: 10.25394/PGS.11328404.v1. URL https://hammer.purdue.edu/articles/thesis/LOCALLY_CONNECTED_NEURAL_NETWORKS_FOR_IMAGE_RECOGNITION/11328404/1.

Walters, R., Li, J., and Yu, R. Trajectory prediction using equivariant continuous convolution. International Conference on Learning Representations, 2021.

Wang, D., Walters, R., Zhu, X., and Platt, R. Equivariant q learning in spatial action spaces. In 5th Annual Conference on Robot Learning, 2021. URL https://openreview.net/forum?id=IScz42A3iCI.

Wang, R., Kashinath, K., Mustafa, M., Albert, A., and Yu, R. Towards physics-informed deep learning for turbulent flow prediction. Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2020a.

Wang, R., Walters, R., and Yu, R. Incorporating symmetry into deep dynamics models for improved generalization. In International Conference on Learning Representations (ICLR), 2020b.

Weiler, M. and Cesa, G. General E(2)-Equivariant Steerable CNNs. In Conference on Neural Information Processing Systems (NeurIPS), 2019a.

Weiler, M. and Cesa, G. General E(2)-equivariant steerable CNNs. In Advances in Neural Information Processing Systems (NeurIPS), pp. 14334–14345, 2019b.

Weiler, M., Geiger, M., Welling, M., Boomsma, W., and Cohen, T. 3d steerable cnns: Learning rotationally equivariant features in volumetric data. In NeurIPS, 2018a.

Weiler, M., Hamprecht, F. A., and Storath, M. Learning steerable filters for rotation equivariant CNNs. Computer Vision and Pattern Recognition (CVPR), 2018b.

Worrall, D. E., Garbin, S. J., Turmukhambetov, D., and Brostow, G. J. Harmonic networks: Deep translation and rotation equivariance. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5028–5037, 2017.

Xie, Y., Franz, E., Chu, M., and Thuerey, N. tempoGAN: A temporally coherent, volumetric GAN for super-resolution fluid flow. ACM Transactions on Graphics (TOG), 37(4):95, 2018.

Zaheer, M., Kottur, S., Ravanbakhsh, S., Poczos, B., Salakhutdinov, R., and Smola, A. Deep sets. arXiv preprint arXiv:1703.06114, 2017.

Zhang, W. Shift-invariant pattern recognition neural network and its optical architecture. In Proceedings of annual conference of the Japan Society of Applied Physics, 1988.

Zhang, W., Itoh, K., Tanida, J., and Ichioka, Y. Parallel distributed processing model with local space-invariant interconnections and its optical architecture. Applied optics, 29(32):4790–4797, 1990.

A. Experiments Details

A.1. Hyperparameter Tuning

We perform grid hyperparameters search as shown in Table 3, including learning rate, batch size, hidden dimension, number of layers, number of steps of prediction errors for training. We also tune the number of filter banks for group convolution models and the coefficient of weight constraints for relaxed weight sharing models. The input length is fixed as 10. In the meanwhile, we make sure the total number of trainable parameters for every model is fewer than 10^7 in order to make fair comparison.

Table 3: Hyperparameter Tuninig Range.

LR	Batch size	Hid-dim	Num-layers	Num-banks	#Steps for Backprop	
10 ^{2} 10 ^{5}	8 64	64 512	3 6	1 4	3 6	0; 10 ^{2}; 10 ^{4}; 10 ^{6}

A.2. Additional Experiments on Real-world Ocean Dynamics

Table 4: Prediction RMSE on ocean currents data.

Model	Conv	LiftT	RGroupT	E2CNN	LiftR	RSteerR	SESN	RppS	RSteerS	RSteerTR	RSteerTS
Future	0.52$_{0.02}$	0.52$_{0.01}$	0.51$_{0.00}$	0.51$_{0.03}$	0.56$_{0.06}$	0.51$_{0.01}$	0.51$_{0.01}$	0.50$_{0.01}$	0.50$_{0.01}$	0.50$_{0.01}$	0.50$_{0.02}$
Domain	0.46$_{0.02}$	0.46$_{0.01}$	0.45$_{0.01}$	0.45$_{0.01}$	0.53$_{0.04}$	0.45$_{0.02}$	0.45$_{0.02}$	0.42$_{0.03}$	0.42$_{0.01}$	0.42$_{0.01}$	0.42$_{0.01}$

Data Description: We use the reanalysis ocean current velocity data generated by the NEMO ocean engine (Madec, 2008). We selected an area (-180 - 150, -30 0)from the Pacific Ocean from 01/01/2021 to 12/31/2021 and extracted 36 64×64 sub-regions for our experiments. We not only test all models on the test sets with different time range and spatial domain from the training set.

Prediction Performance: We compare three equivariant models, three best approximately equivariant baselines as well as our proposed relaxed steerable CNNs and relaxed group convolutions.

A.3. Equivariance Error Analysis

Proposition A.1. Let $f : X \to Y$ where G acts on X and Y by $_X$ and $_Y$ which are norm-preserving. Assume that f is approximately equivariant with $\|f\|_{\varepsilon} \quad 0$. Assume f is a G-equivariant approximator for f. Then there exists $x_0 \ 2 \ X$ such that

$$\|f(x_0) \quad f(x_0)\| \quad \|f\|_{\varepsilon}=2:$$

Proof. We leave implicit the action maps $_X$ and $_Y$. By definition there exists $x \ 2 \ X$ and $g \ 2 \ G$ such that $\|f(gx) \quad gf(x)\| = \|f\|$, whereas $f(gx) \quad gf(x) = 0$: Thus by triangle inequality

$$\begin{aligned}\|f\|_{\varepsilon} &= \|f(gx) \quad gf(x)\| \\ &= \|f(gx) \quad gf(x) \quad f(gx) + gf(x)\| \\ &\|f(gx) \quad f(gx)\| + \|gf(x) \quad gf(x)\|\end{aligned}$$

As the G-action is norm-preserving,

$$\|f\| \ \|f(gx) \quad f(gx)\| + \|f(x) \quad f(x)\|:$$

Thus either $\|f(gx) \quad f(gx)\|$ or $\|f(x) \quad f(x)\|$ is greater than $\|f\|=2$ in which case set x_0 to be gx or x respectively. □

Proposition A.2. Let $f f g$ be an approximately equivariant model class with varying $\|f\|_{\varepsilon} \ 2 \ R_0$. Assume a G-invariant norm. Let $f : X \to Y$ be a function with $\|f\|_{\varepsilon} = \ .$ Assume $\|f \quad f\|_1 \quad c$. Then $\|\|f\|_{\varepsilon} \quad \|f\|_{\varepsilon}\| \ 2c + \ :$

Proof. By triangle inequality and invariance of the norm,

$$\begin{aligned} &\|gf(x) - f(gx)\| \le \|gf(x) - gf(x)\| + \\ &\quad \|gf(x) - f(gx)\| + \|f(gx) - f(gx)\| \\ &\quad 2c + \varepsilon : \end{aligned}$$

□

A.4. Additional Figures

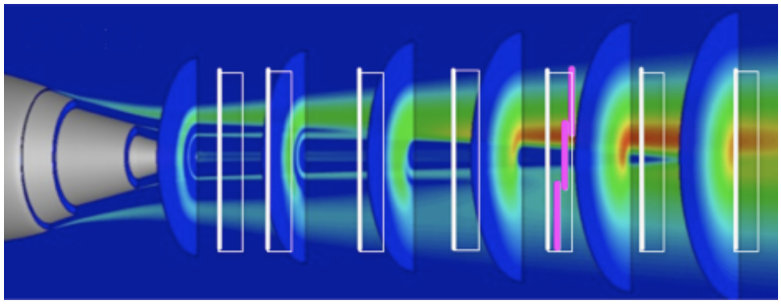

Figure 6: Visualization of axial measurement locations. White boxes show fields of view acquired on streamwise plane at jet centerline for multistream flows. There are three vertical stations at each axial locations/white box, as illustrated by the pink lines. Figure taken from (Bridges & Wernet, 2017).

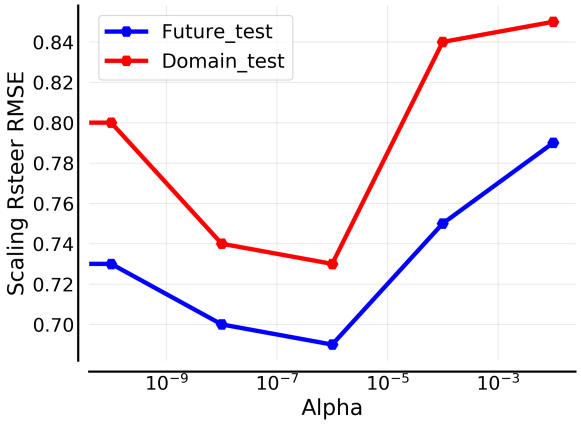

Figure 7: The prediction RMSEs on test sets of a scaling RSteer model trained with different regularization parameter