Residual Pathway Priors for Soft Equivariance Constraints

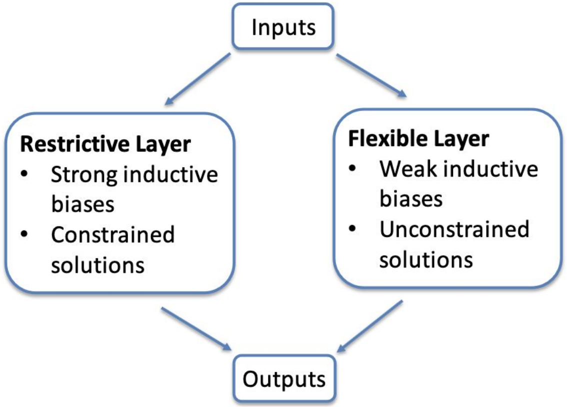
Marc Finzi*, Greg Benton*, Andrew Gordon Wilson

https://proceedings.neurips.cc/paper_files/paper/2021/file/fc394e9935fbd62c8aedc372464e1965-Paper.pdf

Presenter: Niklas Smedemark-Margulies

TL;DR

To avoid issues when symmetry is approximate or mis-specified, create a layer as the sum of two parts: one symmetric part, and one unconstrained part



(b) Structure of RPP Models

Background: Residual Connections

Residual Connections - "ResNet"

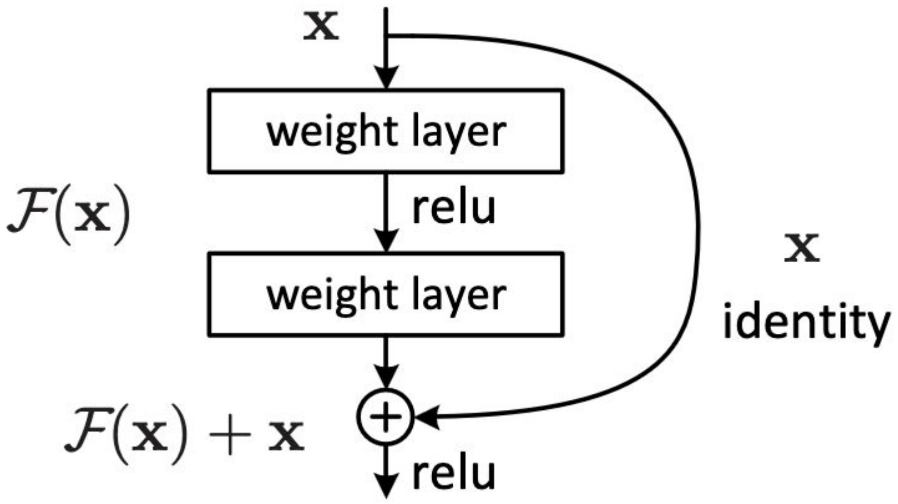


Figure 2. Residual learning: a building block.

Deep Residual Learning for Image Recognition
https://openaccess.thecvf.com/content_cvpr_2016/papers/He_Deep_Residual_Learning_CVPR_2016_paper.pdf

Residual Connections - "Residual RL"

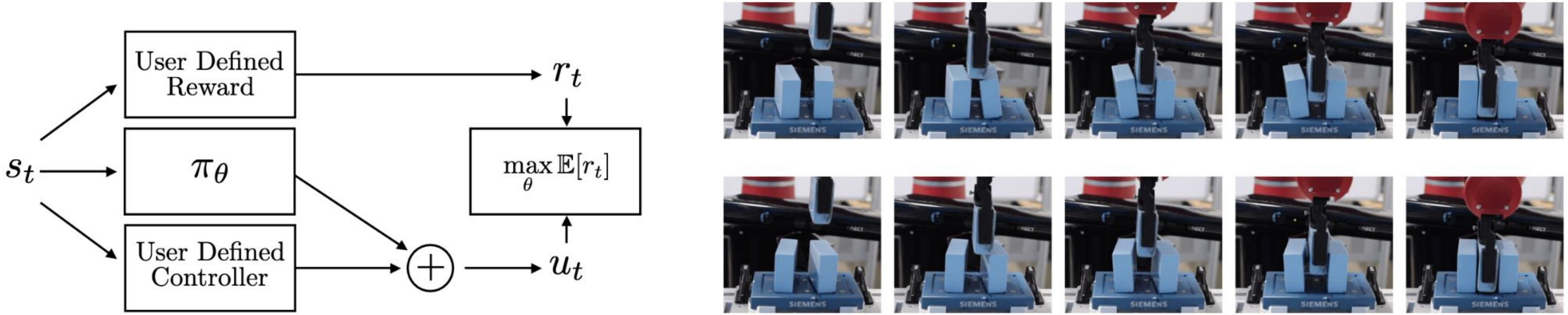


Fig. 1: We train an agent directly in the real world to solve a model assembly task involving contacts and unstable objects. An outline of our method, which consists of combining hand-engineered controllers with a residual RL controller, is shown on the left. Rollouts of residual RL solving the block insertion task are shown on the right. Residual RL is capable of learning a feedback controller that adapts to variations in the orientations of the standing blocks and successfully completes the task of inserting a block between them. Videos are available at `residualrl.github.io`

Residual Reinforcement Learning for Robot Control
https://arxiv.org/pdf/1812.03201.pdf

Residual Connections - "Residual Policy Learning"

learning. Our main idea is to augment arbitrary initial policies by learning *residuals* on top of them. Given an initial policy $\pi : \mathcal{S} \to \mathcal{A}$ with states $s \in \mathcal{S}$ and actions $a \in \mathcal{A} \subseteq \mathbb{R}^d$, we learn a residual function $f_\theta(s) : \mathcal{S} \to \mathcal{A}$ so that we have a *residual policy* $\pi_\theta : \mathcal{S} \to \mathcal{A}$ given by

$$\pi_\theta(s) = \pi(s) + f_\theta(s) \quad .$$

Observe that $\nabla_\theta \pi_\theta(s) = \nabla_\theta f_\theta(s)$, that is, the gradient of the policy does not depend on the initial policy π. We can therefore use policy gradient methods to learn π_θ even if the initial policy π is not differentiable.

Residual Policy Learning
https://arxiv.org/pdf/1812.06298.pdf

Residual Connections - "Physics-Augmented Learning"

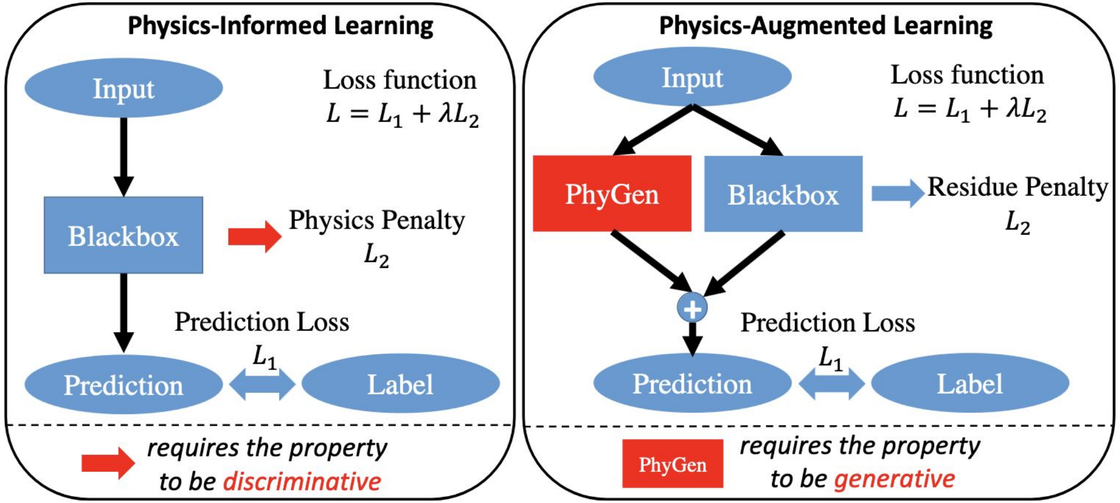
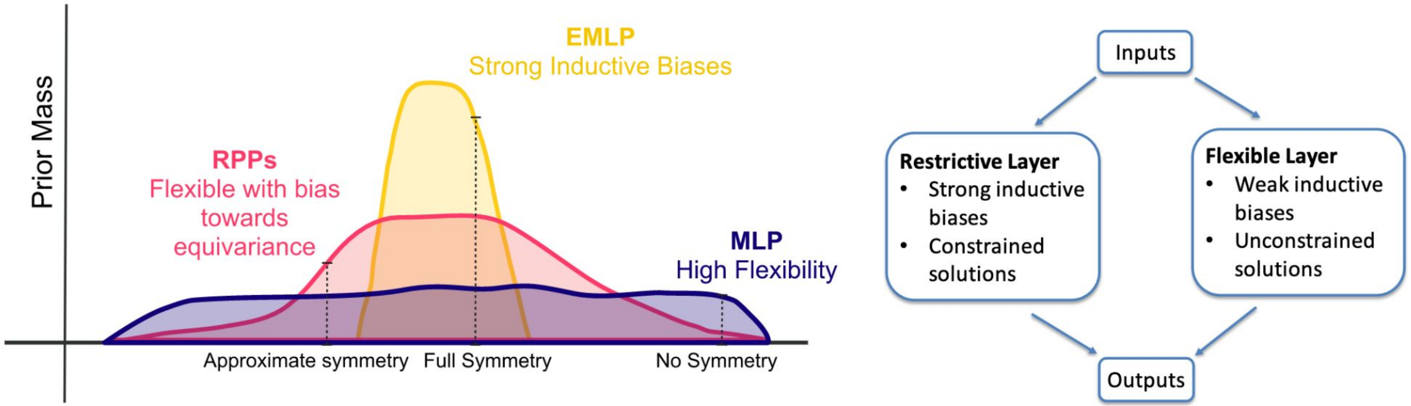


Figure 1: Compare Physics-informed learning (PIL, left) and physics-augmented learning (PAL, right). PIL and PAL apply to discriminative and generative properties respectively.

Physics-Augmented Learning: A New Paradigm Beyond Physics-Informed Learning
https://arxiv.org/pdf/2109.13901.pdf

Method

Basic Method



(a) Priors over Equivariant Solutions

(b) Structure of RPP Models

Figure 1: **Left:** RPPs encode an Occam's razor approach to modeling. Highly flexible models like MLPs lack the inductive biases to assign high evidence to key datasets, while models with strict equivariance constraints are not flexible enough to support problems with only approximate symmetry. **Right:** The structure of RPPs. Expanding the layers into a sum of the constrained and unconstrained solutions while setting the prior to favor the constrained solution, leads to the more flexible layer explaining only the *residual* of what is already explained by the constrained layer.

Recap on EMLP (Finzi et al 2021)

Equivariant MLPs EMLPs provide a method for automatically constructing exactly equivariant layers for any given group and representation by solving a set of constraints. The way in which the vectors are equivariant is given by a formal specification of the types of the input and output through defining their representations. Given some input vector space V_{in} with representation ρ_{in} and some output space V_{out} with representation ρ_{out} the space of all equivariant linear layers mapping $V_{\text{in}} \to V_{\text{out}}$ satisfies

$$\forall g \in G : \quad \rho_{\text{out}}(g)W = W\rho_{\text{in}}(g).$$

These solutions to the constraint form a subspace of matrices $\mathbb{R}^{n_{\text{out}} \times n_{\text{in}}}$ which can be solved for and described by a r dimensional orthonormal basis $Q \in \mathbb{R}^{n_{\text{out}} n_{\text{in}} \times r}$. Linear layers can then be parametrized in this equivariant basis. The elements of W can be parametrized $\text{vec}(W) = Q\beta$ for $\beta \in \mathbb{R}^r$ for the linear layer $v \mapsto Wv$, and symmetric biases can be parametrized similarly.

Defining the Residual Pathway Prior

Define the weight matrix as a sum: $W = A + B$

Consisting of an equivariant part (A) and an unconstrained part (B)

Defining the Residual Pathway Prior

In the case of Equivariant MLP (EMLP) from Finzi et al 2021, this consists of:

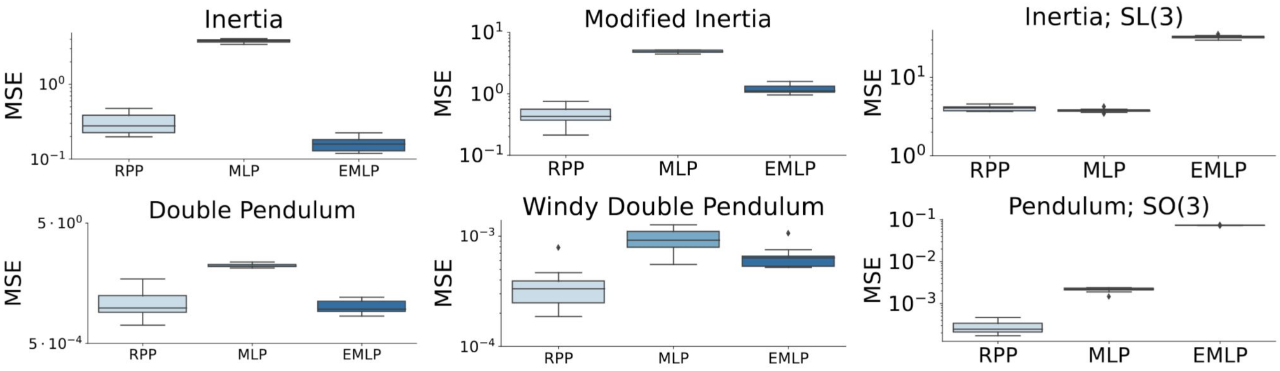
- **equivariant weight matrix** $\mathrm{vec}(A) = Q\beta$ with $\beta \sim \mathcal{N}(0, \sigma_a^2 I)$
  (equivalent to $A \sim \mathcal{N}(0, \sigma_a^2 QQ^\top)$)

- **unconstrained weight matrix** $B \sim \mathcal{N}(0, \sigma_b^2 I)$ with $\sigma_b^2 I = \sigma_b^2 QQ^\top + \sigma_b^2 PP^\top$

In total: $A + B = W \sim \mathcal{N}(0, (\sigma_a^2 + \sigma_b^2)QQ^\top + \sigma_b^2 PP^\top)$

Experiments

Experiment: Inertia and Pendulum Datasets



(a) Exact Symmetries (b) Approximate Symmetries (c) Mis-specified Symmetries

Figure 2: A comparison of test performance over 10 independent trials using RPP-EMLP and equivalent EMLP and MLP models on the inertia (**top**) and double pendulum (**bottom**) datasets in which we have three varying levels of symmetries. The boxes represent the interquartile range, and the whiskers the remainder of the distribution. **Left:** perfect symmetries in which EMLP and the equivariant components of RPP-EMLP exactly capture the symmetries in the data. **Center:** approximate symmetries in which the perfectly symmetric systems have been modified to include some non-equivariant components. **Right:** mis-specified symmetries in which the symmetric components of EMLP and RPP-EMLP do not reflect the symmetries present in the data.

Datasets from: https://proceedings.mlr.press/v139/finzi21a/finzi21a.pdf

Experiments: Sampling from Ensemble, and Varying Prior

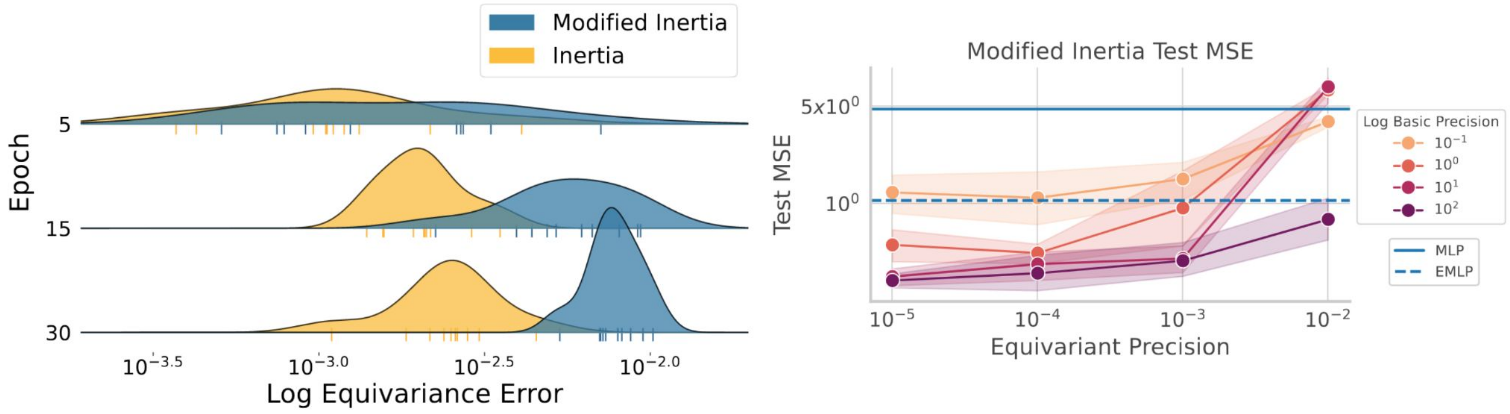


Figure 3: **Left:** Kernel density estimators of log equivariance error across training epochs for 10 independently trained networks. Here the color denotes the dataset these models were trained on. Treating these samples as a proxy for posterior density, we see that on the non-equivariant Modified Inertia dataset, the posterior is shifted upward to match the level of equivariance in the data during training. **Right:** Test MSE as a function of the weight decay parameters on the equivariant and basic weights on the modified inertia dataset. We observe that so long as the prior in the basis of equivariant weights is broad enough, we can achieve low test error with RPPs.

Experiments on Toy Datasets

	CIFAR-10	Energy	Fertility	Pendulum	Wine
MLP	37.61 ± 0.56	0.39 ± 0.48	0.049 ± 0.0044	4.65 ± 0.50	0.66 ± 0.058
RPP	12.62 ± 0.34	0.73 ± 0.44	0.060 ± 0.0097	4.25 ± 0.50	0.69 ± 0.031
Conv	12.03 ± 0.46	1.34 ± 0.38	0.076 ± 0.0157	4.63 ± 0.36	0.79 ± 0.092

Table 1: Mean test classification error on CIFAR-10 and MSE on 4 UCI regression tasks, with one standard deviation errors taken over 10 trials. Similar to Figure 4, we find that whether the constrained convolutional structure is helpful (CIFAR) or not (UCI), RPP-Conv performs similarly to the model with the correct level of complexity.

Experiments on Mujoco Toy Problems

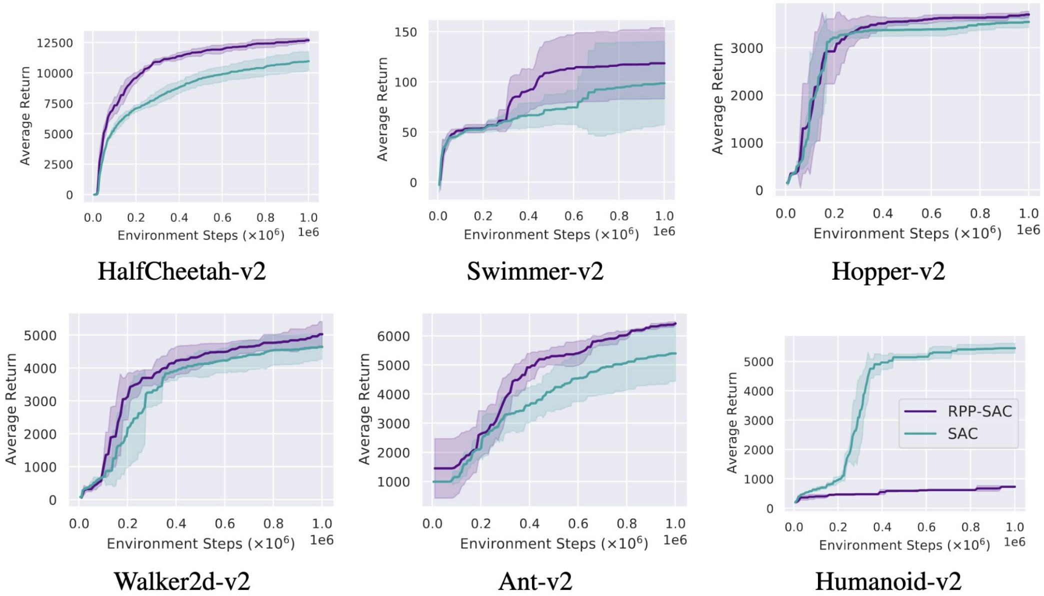


Figure 5: Average reward curve of RPP-SAC and SAC trained on Mujoco locomotion environments (max average reward attained at each step). Mean and one standard deviation taken over 4 trials shown in the shaded region. Incorporating approximate symmetries in the environments improves the efficiency of the model free RL agents.

Discussion and Issues

Main conclusions:

- Overall, a very clever, simple, easily implemented idea!
- The probabilistic interpretation seems possibly weak and maybe obscures the key idea.
  - Was this added after-the-fact to try and make the method seem more fancy?
  - Is there even really a probabilistic method here, or just a weighted sum of two layers, with two different values of weight decay?
- Some experimental details missing from paper (datasets, training scheme, experiments on "posterior equivariance error")