Noether Networks: Meta-Learning Useful Conserved Quantities

Ferran Alet*[1], Dylan Doblar*[1], Allan Zhou[2], Josh Tenenbaum[1], Kenji Kawaguchi[3], Chelsea Finn[2]; [1]MIT, [2]Stanford, [3]NUS; *Equal contribution

Noether Networks meta-learn inductive biases in the form of useful conserved quantities that improve predictions when optimized inside the prediction function.

Motivation and background

Successful **inductive biases** often exploit **symmetries**, which are often **unknown** and **hard to discover**, or **difficult to encode**.

We learn **conserved quantities**, inspired by **Noether's Theorem**:

> *For every **continuous symmetry** property of a dynamical system, there is a corresponding **conservation law**.*

Tailoring framework [1]: impose inductive biases at **prediction time**

- **Optimize** unsupervised *tailoring loss* **inside** prediction function
- **Fine-tune** the model to the **particular** query
- **Drawback**: tailoring losses must be **hand-coded**

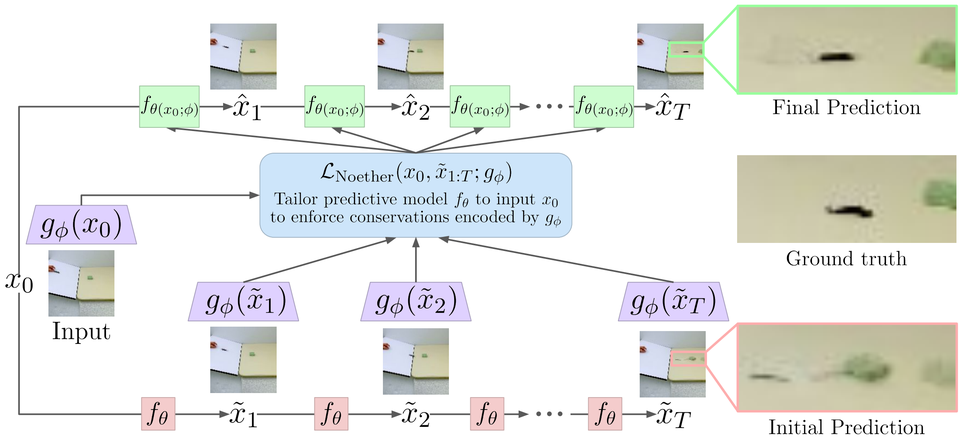
Prediction and training with neural Noether loss

$$\mathcal{L}_{\text{Noether}}(x_0, \tilde{x}_{1:T}; g_\phi) = \sum_{t=1}^{T} \left| g_\phi(x_0) - g_\phi(\tilde{x}_t) \right|^2$$

procedure PREDICTSEQUENCE($x_0; \theta, \phi$)
  $\tilde{x}_0, \hat{x}_0 \leftarrow x_0, x_0$
  $\tilde{x}_t \leftarrow f_\theta(\tilde{x}_{t-1})\ \forall t \in \{1, \dots, T\}$
  $\theta(x_0; \phi) \leftarrow \theta - \lambda_{\text{in}} \nabla_\theta \mathcal{L}_{\text{Noether}}(x_0, \tilde{x}_{1:T}; g_\phi)$
  $\hat{x}_t \leftarrow f_{\theta(x_0;\phi)}(\hat{x}_{t-1})\ \forall t \in \{1, \dots, T\}$
  return $\hat{x}_{1:T}$

procedure TRAIN
  $\phi \leftarrow$ randomly initialized embedding weights
  $\theta \leftarrow$ randomly initialized predictive model weights
  while not done **do**
    Sample batch $x^{(0)}_{0:T}, \dots, x^{(N)}_{0:T} \sim \mathcal{D}_{\text{train}}$
    for $0 \le n \le N$ **do**
      $\hat{x}^{(n)}_{1:T} \leftarrow$ PREDICTSEQUENCE($x^{(n)}_0; \theta, \phi$)
    $\phi \leftarrow \phi - \lambda_{\text{emb}} \nabla_\phi \sum_{n=0}^{N} \mathcal{L}_{\text{task}}(\hat{x}^{(n)}_{1:T}, x^{(n)}_{1:T})$
    $\theta \leftarrow \theta - \lambda_{\text{out}} \nabla_\theta \sum_{n=0}^{N} \mathcal{L}_{\text{task}}(\hat{x}^{(n)}_{1:T}, x^{(n)}_{1:T})$
  return ϕ, θ

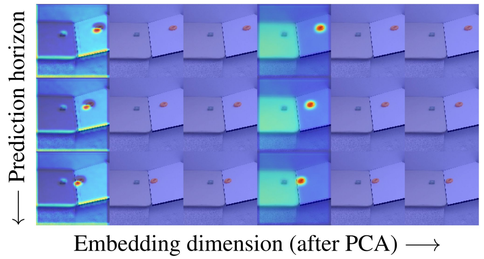
Optimizing Noether conservation loss at prediction time

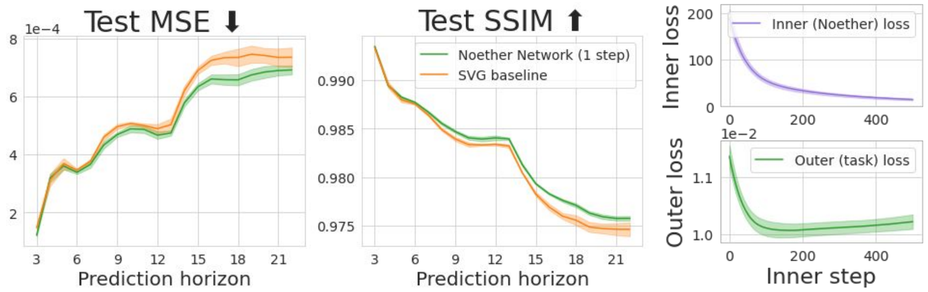


Imposing meta-learned inductive biases improves video prediction quality.

Learning from raw pixels in real-world video prediction

- Physics 101 [5]: colliding objects
- Grad-CAM heatmaps: embeddings attend to relevant pixels (right)
- Improves performance, esp. with long horizon (below, left and center)
- Generalizes beyond a single test-time inner step (below, right)





Theoretical advantages: enforcing conservations

Theorem 1. *Let $\rho \in \mathbb{N}^+$. Then, for any $\delta > 0$, with probability at least $1 - \delta$ over an iid draw of n examples $((x_i, y_i))_{i=1}^n$, the following holds for all $\theta \in \Theta$:*

$$G(\theta) \le C \sqrt{\frac{\xi \ln(\max(\sqrt{\xi}, 1)) + \xi \ln(2R(\zeta^{1-1/\rho})\sqrt{n}) + \ln(1/\delta)}{2n}} + \mathbb{1}\{\xi \ge 1\} \sqrt{\frac{\zeta^{2/\rho}}{n}}. \quad (1)$$

where $\xi = m$ if $g_\phi(f_\theta(x)) = g_\phi(x)$ for any $x \in \mathbb{R}^d$ and $\theta \in \Theta$, and $\xi = d$ otherwise.

Each conservation law decreases ξ by 1, lowering the train-test gap.

Noether Nets recover known conservation laws

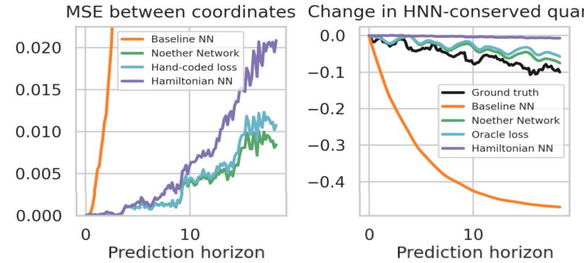
Noether loss is parameterized as symbolic formula in a simple DSL.

Method	Description	RMSE
Vanilla MLP	N/A	0.0563
Noether Nets	$p^2 - 2.99\cos(q)$	0.0423
True $\mathcal{H}$ [oracle]	$p^2 - 3.00\cos(q)$	0.0422

Ideal pendulum

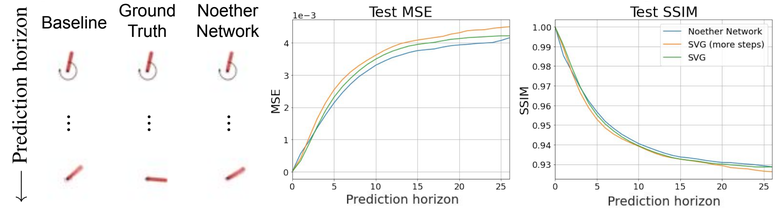
Method	Description	RMSE
Vanilla MLP	N/A	.0174
Noether Nets	$q^2 + 1.002\,p^2$	.0165
True $\mathcal{H}$ [oracle]	$q^2 + 1.000\,p^2$	.0166

Ideal spring



Real (dissipative) pendulum system

Improving predictions with controlled dynamics



Related work

[1] Tailoring: encoding inductive biases by optimizing losses optimized at prediction time; Alet et al.
[2] Meta-learning symmetries by reparameterization; Zhou et al.
[3] Hamiltonian neural networks; Greydanus et al.
[4] Stochastic video generation with a learned prior; Denton and Fergus. (Used as base video prediction model)
[5] Physics 101: Learning physical object properties from unlabeled videos; Wu et al.
[6] AI Poincaré: Machine learning conservation laws from trajectories; Liu and Tegmark.