SEAL – A Symmetry Encouraging Loss for High Energy Physics

Pradyun Hebbar
Physics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
Department of Particle Physics, University of Geneva, Geneva 1205, Switzerland
Pradyun.Hebbar@unige.ch

Thandikire Madula
University College London, Gower Street, London, WC1E 6BT, UK
thandi.madula.17@ucl.ac.uk

Vinicius Mikuni
Nagoya University, Kobayashi-Maskawa Institute, Aichi 464-8602, Japan
vmikuni@hepl.phys.nagoya-u.ac.jp

Benjamin Nachman
Department of Particle Physics and Astrophysics, Stanford University, Stanford, CA 94305, USA
Fundamental Physics Directorate, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA
nachman@stanford.edu

Nadav Outmezguine
Leinweber Institute for Theoretical Physics, University of California, Berkeley, CA 94720, USA
Physics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
nadav.out@gmail.com

Inbar Savoray
Leinweber Institute for Theoretical Physics, University of California, Berkeley, CA 94720, USA
Physics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
inbar.savoray@berkeley.edu

Abstract

Physical symmetries provide a powerful inductive bias for machine learning models in science, improving robustness, data efficiency, and interpretability. However, building models that explicitly enforce symmetries requires specialized architectures, and real-world experiments often break these symmetries through finite detector granularity and energy thresholds. We introduce SEAL (Symmetry Encouraging Loss), a family of soft-constraint loss terms that encourage equivariance, requiring no architectural modifications. We present two complementary variants: GSEAL, which penalizes output differences under random group transformations of the input, and δSEAL, which penalizes the model's gradients along directions corresponding to infinitesimal symmetry transformations. Using top quark jet tagging and Lorentz equivariance as a case study, we observe that the addition of the soft constraints leads to more robust performance while requiring minimal computational costs.

39th Conference on Physics and AI at Stanford University (PAI 2026).

1 Introduction

Machine learning is now central to data analysis in high energy physics (HEP), yet particle physics data poses unique challenges: events are naturally variable-length point clouds of four-momenta transforming under the Poincaré group. A well-studied strategy is to design *equivariant* architectures that respect these symmetries by construction [Gong et al., 2022, Qiu et al., 2023, Bogatskiy et al., 2024, Batatia et al., 2023, Spinner et al., 2024]. While effective, such models face limitations: constrained layer operations increase complexity and training cost, scaling behavior may lag behind general-purpose architectures, and strict equivariance assumes the symmetry is exactly realized in the data. In practice, detector thresholds, beam geometry, and reconstruction algorithms introduce symmetry-breaking effects that are difficult to model within a rigid equivariant framework [Spinner et al., 2024, Bogatskiy et al., 2024]. Alternative approaches combine equivariant and non-equivariant components [Murnane et al., 2023, Nabat et al., 2025], but still require dedicated architectural design.

We propose **SEAL** (Symmetry Encouraging Loss)[1], a penalty term added to any standard loss function that encourages symmetry without modifying the model architecture. SEAL is agnostic to the network design, incurs no additional cost at inference time, and gracefully accommodates approximate symmetries. We focus on Lorentz invariance, which is highly relevant to HEP and rarely encountered in other domains, but the formulation is general and applicable to any Lie group.

We introduce two variants. **GSEAL** (Group SEAL) penalizes discrepancies between the model's output on original and randomly transformed inputs. δ**SEAL** (infinitesimal SEAL) penalizes the model's gradients along directions corresponding to symmetry transformations of the inputs. Related ideas have been explored for E_3 invariance [Elhag et al., 2024] and physics-informed networks [Akhound-Sadegh et al., 2023]; our work develops the framework for the Lorentz group and demonstrates it on a realistic collider physics benchmark.

2 Method

Setup. A function $f : X \to Y$ is equivariant with respect to a group G if $f(g \odot x) = (g \odot f)(x)$ for all $x \in X$ and $g \in G$, where $g\odot$ denotes the group action. Invariance is the special case where $g \odot f = f$. Given input-target pairs (x, y) and a neural network f_θ, we augment the standard training loss $\mathcal{L}$ with a symmetry penalty Γ such that the training objective becomes:

$$\min_\theta \; \mathbb{E}\Big[\mathcal{L}(f_\theta(x),\, y) + \lambda\, \Gamma(f_\theta,\, x)\Big], \tag{1}$$

where λ controls the relative strength of the symmetry constraint. The general form of Γ integrates the squared equivariance violation over all group elements and inputs; in practice, we approximate this via two complementary strategies.

GSEAL. For a mini-batch of N inputs, we sample a random Lorentz boost g_i for each input x_i and define:

$$\Gamma_G = \frac{1}{N} \sum_{i=1}^{N} \left| f(g_i \odot x_i) - f(x_i) \right|^2 . \tag{2}$$

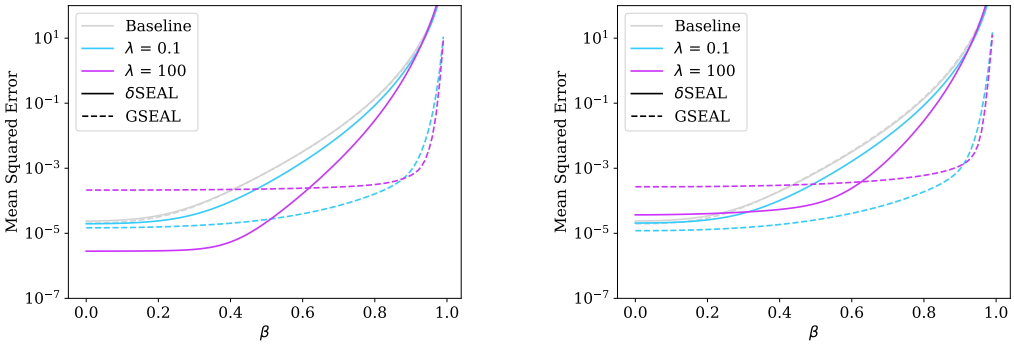
Each boost is parameterized by a velocity vector $\vec{\beta}$ with direction sampled uniformly on the unit sphere and magnitude $\|\vec{\beta}\|^2$ sampled uniformly in $[0, \beta^2_{\max}]$, with $\beta^2_{\max} = 0.95$ in our experiments. A different transformation is drawn per data point per epoch.

δ**SEAL.** For continuous groups, infinitesimal transformations are generated by the Lie algebra generators L^a. The variation of the function under infinitesimal symmetry transformations is given by $\nabla f \cdot \delta x$, with $\delta x = Lx$ the variation of the input. Therefore, for an invariant function we define

$$\Gamma_\delta = \frac{1}{Nn} \sum_{i=1}^{N} \sum_{a=1}^{n} \left| \nabla f \cdot L^a_x \cdot x_i \right|^2 , \tag{3}$$

where n=6 is the number of Lorentz generators (three rotations, three boosts) and L^a_x are the generators in the representation matching that of the input, which in this work will be the four-vector

[1] Code available at https://github.com/inbarsavoray/SEAL.git.



(a) Exact symmetry

(b) Approximate symmetry

Figure 1: MSE as a function of the Lorentz boost β applied to test data in the toy regression task. Solid lines: δSEAL; dashed lines: GSEAL. Different colors correspond to different values of the constraint strength λ.

representation. This penalty requires computing gradients of f with respect to its inputs, which is efficient via automatic differentiation.

Both penalties add cost only during training (roughly $2\times$ per training step), while inference is unchanged. SEAL requires only knowledge of the symmetry group and how inputs transform – no architectural changes are needed. While SEAL shares similarities with data augmentation [Quiroga et al., 2019, Gerken et al., 2022], augmentation transforms the inputs and assigns them the same label, whereas SEAL directly penalizes the model through a loss term derived from the symmetry.

3 Experiments

3.1 Toy regression

We first validate SEAL on a controlled regression task. We generate 10^5 random four-vectors $\{p_i\}$ with components uniform in $[-0.5, 0.5]$ and regress a second-degree polynomial in Lorentz-invariant products. We study two scenarios: (i) exact Lorentz invariance, where the target depends only on $m_i^2 = p_i^\alpha \eta_{\alpha\beta} p_i^\beta$; and (ii) approximate symmetry, where a small spurion $s = (0, 0, 0, 10^{-3})$ breaks invariance as $f(p_i) = f\left(m_i^2, p_i^\alpha \eta_{\alpha\beta} s^\beta\right)$. The model is a 3-layer multilayer perceptron (width 300, GELU activations [Hendrycks and Gimpel, 2016]) trained with mean squared error (MSE) task-loss.

Figure 1 shows the MSE on test data boosted in the z-direction by various β values. Without SEAL, performance degrades sharply with boost magnitude. Both GSEAL and δSEAL dramatically lower and flatten the error curve, with GSEAL providing superior robustness at large boosts and δSEAL excelling at moderate boosts. Even with approximate symmetry (Figure 1b), modest values of λ improve accuracy and extrapolation.

3.2 Top quark jet tagging

We apply SEAL to a realistic collider physics benchmark: classifying jets initiated by top quark decays versus light QCD background in the ATLAS Top Tagging dataset [ATLAS Collaboration, 2022], using 16M training jets and 4M validation jets. Lorentz invariance is a natural requirement for the classifier output, as the jet's label should not depend on the spatial orientation or the boost of the jet. However, the symmetry is violated in practice by the colliding beams, detector effects, imperfect reconstruction and clustering schemes.

Our classifier is a transformer encoder [Vaswani et al., 2023] (3 layers, dimension 256, 4 attention heads, $\sim$1.3M parameters) taking seven kinematic features per jet constituent as input. We compare: (i) Baseline (cross-entropy only), (ii) Baseline + GSEAL (λ=1.0), (iii) Baseline + δSEAL (λ=0.01), and (iv) PELICAN [Bogatskiy et al., 2024], an architecture with hard Lorentz invariance built in.

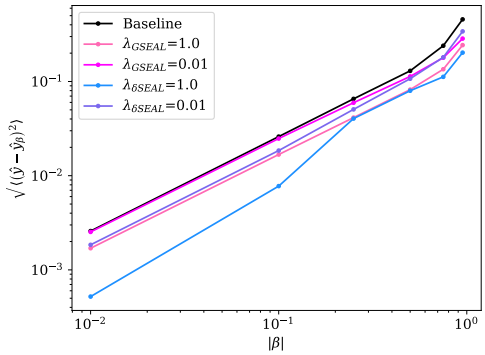


Figure 2: Prediction change under Lorentz boosts of test jets. SEAL-trained models are substantially more invariant than the baseline across the full range of boost parameters.

Table 1: Classification performance and computation time per batch on the ATLAS Top Tagging dataset. Uncertainties are standard deviations over 5 training runs. $1/\epsilon_b^{0.3}$ is the inverse background acceptance at 30% signal efficiency.

Model	Bal. Accuracy	AUC	$1/\epsilon_b^{0.3}$	Train Step [s]	Evaluation Step [s]
Baseline	0.891 ± 0.001	0.959 ± 0.001	652 ± 37	0.03	0.005
+ GSEAL	0.891 ± 0.001	0.959 ± 0.001	638 ± 37	0.06	0.005
+ δSEAL	0.890 ± 0.002	0.959 ± 0.001	620 ± 48	0.06	0.005
PELICAN	0.890	0.959	630	0.4	0.2

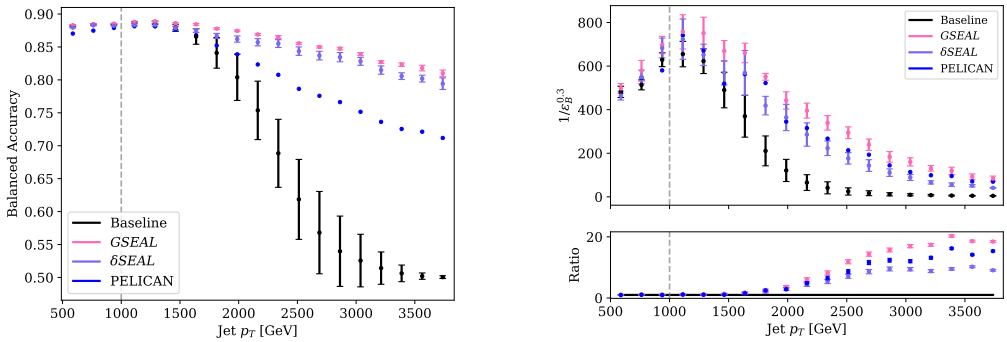
Invariance. Figure 2 quantifies the change in model predictions under Lorentz boosts of the test data. Both SEAL variants significantly reduce prediction variance under boosts compared to the unconstrained baseline. For smaller λ, GSEAL is more effective at large boosts, while δSEAL is more effective at small boosts. While at $\lambda = 1.0$ δSEAL becomes more invariant than GSEAL even at large boosts, this may come at a greater cost to the model's accuracy.

Classification performance. Table 1 shows that all models achieve comparable accuracy (balanced accuracy $\approx$0.891) and AUC ($\approx$0.959) on the standard test set, implying SEAL does not sacrifice classification power. Training with SEAL roughly doubles the per-step training time (0.06 s vs. 0.03 s per batch of 256 jets on an A100 GPU), while inference time is unchanged (0.005 s). In contrast, PELICAN requires $\sim$13$\times$ longer training and $\sim$40$\times$ longer inference per batch.

Extrapolation. To test generalization, we train all models only on jets with transverse momentum $p_T \leq 1$ TeV and evaluate on the full p_T spectrum (Figure 3). The baseline accuracy collapses above $\sim$1.5 TeV, while both SEAL variants maintain robust performance with significantly reduced variance across training seeds. GSEAL achieves the best extrapolation, with background rejection 10-20$\times$ higher than the baseline in the extrapolation region. This demonstrates that soft symmetry constraints act as an effective regularizer for out-of-distribution generalization in kinematic variables.

4 Conclusion

We introduced SEAL, a general framework for encouraging symmetry equivariance in machine learning models through soft loss penalties, requiring no architectural modifications. Applied to Lorentz invariance in top quark jet tagging with the realistic ATLAS dataset, SEAL matches the classification performance of both unconstrained and hard-equivariant baselines while substantially improving invariance under Lorentz boosts and generalization to unseen kinematic regions. The approach is architecture-agnostic, adds no inference cost, and gracefully handles the approximate symmetries common in experimental physics.



(a) Balanced accuracy

(b) Background rejection ($1/\epsilon_b^{0.3}$)

Figure 3: Extrapolation from $p_T \leq 1$ TeV training data to higher p_T. The vertical dashed line marks the training cut. SEAL-trained models ($\lambda = 1.0$ for both variants) maintain performance far beyond the training region, while the baseline degrades rapidly.

Several directions remain for future work. The hyperparameters λ and $\beta_{\max}$ could be made learnable or scheduled dynamically during training. SEAL can be extended to equivariant (rather than invariant) outputs, enabling applications to tasks predicting four-momenta or other Lorentz-covariant quantities. Since SEAL does not use truth labels, it may also be useful beyond supervised learning. Finally, SEAL can be readily combined with foundation models during fine-tuning [Mikuni and Nachman, 2025a,b], adding symmetry awareness to large pre-trained architectures without retraining from scratch.

References

Shiqi Gong, Qi Meng, Jue Zhang, Huilin Qu, Congqiao Li, Sitian Qian, Weitao Du, Zhi-Ming Ma, and Tie-Yan Liu. An efficient lorentz equivariant graph neural network for jet tagging. *Journal of High Energy Physics*, 2022(7), July 2022. ISSN 1029-8479. doi: 10.1007/jhep07(2022)030. URL `http://dx.doi.org/10.1007/JHEP07(2022)030`.

Shikai Qiu, Shuo Han, Xiangyang Ju, Benjamin Nachman, and Haichen Wang. Holistic approach to predicting top quark kinematic properties with the covariant particle transformer. *Phys. Rev. D*, 107(11):114029, 2023. doi: 10.1103/PhysRevD.107.114029.

Alexander Bogatskiy, Timothy Hoffman, David W. Miller, Jan T. Offermann, and Xiaoyang Liu. Explainable equivariant neural networks for particle physics: PELICAN. *JHEP*, 03:113, 2024. doi: 10.1007/JHEP03(2024)113.

Ilyes Batatia, Mario Geiger, Jose Munoz, Tess Smidt, Lior Silberman, and Christoph Ortner. A general framework for equivariant neural networks on reductive lie groups. *Advances in Neural Information Processing Systems*, 36:55260–55284, 2023.

Jonas Spinner, Victor Bresó, Pim de Haan, Tilman Plehn, Jesse Thaler, and Johann Brehmer. Lorentz-Equivariant Geometric Algebra Transformers for High-Energy Physics. 10 2024.

Daniel Murnane, Savannah Thais, and Jason Wong. Semi-Equivariant GNN Architectures for Jet Tagging. *J. Phys. Conf. Ser.*, 2438(1):012121, 2023. doi: 10.1088/1742-6596/2438/1/012121.

Seth Nabat, Aishik Ghosh, Edmund Witkowski, Gregor Kasieczka, and Daniel Whiteson. Learning broken symmetries with approximate invariance. *Phys. Rev. D*, 111(7):072002, 2025. doi: 10.1103/PhysRevD.111.072002.

Ahmed A. Elhag, T. Konstantin Rusch, Francesco Di Giovanni, and Michael Bronstein. Relaxed equivariance via multitask learning, 2024. URL `https://arxiv.org/abs/2410.17878`.

Tara Akhound-Sadegh, Laurence Perreault-Levasseur, Johannes Brandstetter, Max Welling, and Siamak Ravanbakhsh. Lie point symmetry and physics-informed networks. *Advances in Neural Information Processing Systems*, 36:42468–42481, 2023.

Facundo Quiroga, Franco Ronchetti, Laura Lanzarini, and Aurelio F. Bariviera. *Revisiting Data Augmentation for Rotational Invariance in Convolutional Neural Networks*, page 127–141. Springer International Publishing, March 2019. ISBN 9783030154134. doi: 10.1007/978-3-030-15413-4_10. URL http://dx.doi.org/10.1007/978-3-030-15413-4_10.

Jan E. Gerken, Oscar Carlsson, Hampus Linander, Fredrik Ohlsson, Christoffer Petersson, and Daniel Persson. Equivariance versus augmentation for spherical images, 2022. URL https://arxiv.org/abs/2202.03990.

Dan Hendrycks and Kevin Gimpel. Bridging nonlinearities and stochastic regularizers with gaussian error linear units. *CoRR*, abs/1606.08415, 2016. URL http://arxiv.org/abs/1606.08415.

ATLAS Collaboration. Constituent-Based Top-Quark Tagging with the ATLAS Detector. Technical report, CERN, Geneva, 2022. URL https://cds.cern.ch/record/2825328.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2023. URL https://arxiv.org/abs/1706.03762.

Vinicius Mikuni and Benjamin Nachman. Solving key challenges in collider physics with foundation models. *Phys. Rev. D*, 111(5):L051504, 2025a. doi: 10.1103/PhysRevD.111.L051504.

Vinicius Mikuni and Benjamin Nachman. Method to simultaneously facilitate all jet physics tasks. *Phys. Rev. D*, 111(5):054015, 2025b. doi: 10.1103/PhysRevD.111.054015.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction claim that SEAL improves invariance and extrapolation without sacrificing classification performance; these claims are supported by the experimental results in Section 3.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The conclusion discusses limitations including the need for hyperparameter tuning (λ, $\beta_{\max}$) and the fact that experiments are currently limited to one symmetry group (Lorentz) and one benchmark task.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [NA]

   Justification: The paper does not contain formal theorems or proofs. The method is presented as a practical framework with empirical validation.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: Section 3 provides model architectures, hyperparameters, dataset details, and training procedures. Code is available via an anonymized repository.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: Code is publicly available at an anonymized URL. The ATLAS Top Tagging dataset is publicly available.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: Section 3 specifies architectures, data splits (16M train, 4M validation), key hyperparameters (λ, $\beta_{\max}$), and activation functions.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: Table 1 and Figures 3–4 report standard deviations over 5 independent training runs with different random seeds.

8. **Experiments compute resources**

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: Per-batch training and inference times on A100 GPUs are reported in Section 3.2.

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics?

   Answer: [Yes]

   Justification: This work involves fundamental physics research with no direct ethical concerns.

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [NA]

    Justification: This is fundamental research on symmetry-aware machine learning for particle physics, with no foreseeable negative societal impacts.

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse?

    Answer: [NA]

    Justification: The models and data are specific to particle physics and pose no risk for misuse.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets used in the paper properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: The ATLAS Top Tagging dataset and PELICAN model are properly cited.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: The SEAL code repository includes documentation and instructions for reproducing experiments.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots?

    Answer: [NA]

    Justification: No human subjects or crowdsourcing were involved.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether IRB approvals were obtained?

    Answer: [NA]

    Justification: No human subjects research was conducted.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research?

Answer: [NA]

Justification: LLMs were not used as a component of the core methodology.