TriForces: Augmenting Atomistic GNNs for Transferable Representations

Joseph Musielewicz[1], Ali Ramlaoui[1,2], Alexandre Duval[1], Hannah Bull[1], Victor Schmidt[1], Hugues Talbot[2], Fragkiskos Malliaros[2]
[1]Entalpic, Paris, France
[2]CentraleSupélec, Université Paris-Saclay, France

joseph.musielewicz@entalpic.ai

Abstract

Geometric Graph Neural Networks (GNNs) have emerged as the dominant paradigm for Machine Learning Interatomic Potentials (MLIPs), achieving high fidelity on large-scale Density Functional Theory (DFT) datasets. However, standard architectures often entangle composition and geometry into a single latent representation optimized solely for energy and force prediction. Consequently, these models suffer from brittle transfer performance when adapted to new chemical domains or properties, often requiring extensive fine-tuning that leads to catastrophic forgetting [1]. Furthermore, the resulting embeddings are frequently ill-suited for chemical exploratory analysis, such as similarity retrieval.

To address these limitations, we present TriForces, a model-agnostic framework that augments existing atomistic GNNs (such as MACE [5], Orb [6], or eSEN [7]) to enforce disentangled and transferable representations. As illustrated in Figure 1, TriForces structurally separates the node-level representation into three distinct streams:

- A *Composition Stream*, utilizing a Transformer with count-weighted attention to encode chemical stoichiometry independent of coordinates.
- A *Structure Stream*, employing rotation-invariant descriptors and message passing to capture geometric motifs independent of element identity.
- An *Interaction Stream*, which is the base geometric GNN capturing the coupling between chemistry and geometry.

To effectively train this decomposed architecture without requiring DFT labels, we introduce a multi-objective Self-Supervised Learning (SSL) strategy, detailed in Figure 2. We combine reconstruction-based objectives—specifically position denoising to stabilize geometry and masked atom prediction to learn compositional patterns—with a non-reconstruction objective, LeJEPA [2], that aligns representations across augmented views.

We demonstrate that TriForces significantly outperforms single-stream baselines on transfer learning benchmarks. On the Open Materials 2024 (OMat24) benchmark [3], TriForces improves data efficiency dramatically, reducing energy Mean Absolute Error (MAE) by 57% in the limited-data regime (20K samples) compared to standard fine-tuning. We further show state-of-the-art results on MatBench [4] and QM9 tasks. Beyond predictive accuracy, the disentangled embedding space enables interpretable, query-specific retrieval—allowing users to search for materials based solely on structural similarity or compositional similarity, a capability absent in standard MLIPs.

References

[1] Niblett, S. P., et al. "Transferability of data sets between machine-learned interatomic potential algorithms." Journal of Chemical Theory and Computation, 2025.
[2] Balestriero, R. and LeCun, Y. "LeJEPA: Provable and scalable self-supervised learning without the heuristics." arXiv preprint arXiv:2511.08544, 2025.
[3] Barroso-Luque, L., et al. "Open materials 2024 (OMat24) inorganic materials dataset and models." arXiv preprint arXiv:2410.12771, 2024.
[4] Dunn, A., et al. "Benchmarking materials property prediction methods: the matbench test set." Npj Computational Materials, 6(1):138, 2020.
[5] Batatia, I., et al. "MACE: Higher order equivariant message passing neural networks for fast and accurate force fields." NeurIPS, 2022.
[6] Rhodes, B., Vandenhaute, S., Šimkus, V., Gin, J., Godwin, J., Duignan, T., and Neumann, M. "Orb-v3: atomistic simulation at scale." arXiv preprint arXiv:2504.06231, 2025.
[7] Fu, X., Wood, B. M., Barroso-Luque, L., Levine, D. S., Gao, M., Dzamba, M., and Zitnick, C. L. "Learning Smooth and Expressive Interatomic Potentials for Physical Property Prediction." In Pro-ceedings of the 42nd International Conference on Machine Learning (ICML), Proceedings of Machine Learning Research, Vol. 267, pp. 17875–17893, PMLR, 2025.

Figures

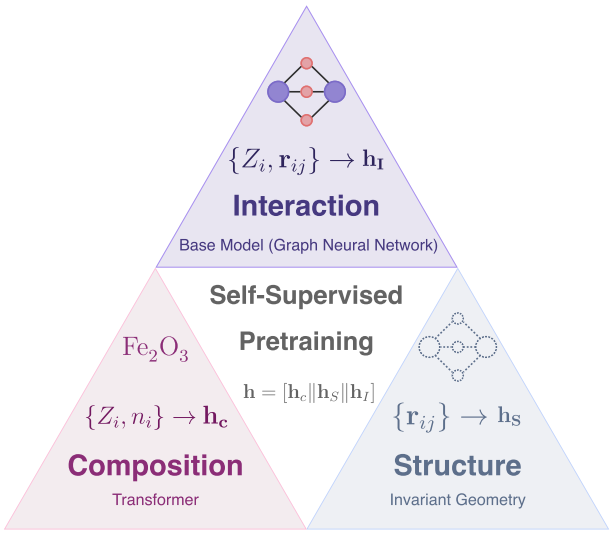

Figure 1. The TriForces framework separating Composition, Structure, and Interaction streams.

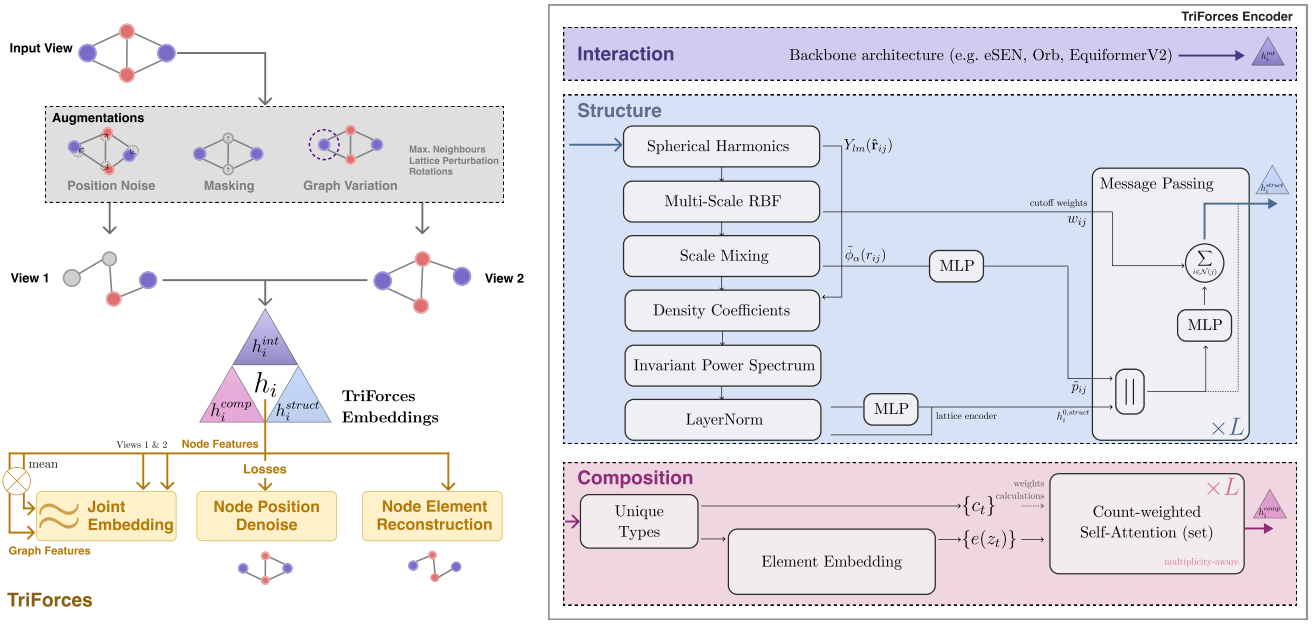

Figure 2. The self-supervised pretraining pipeline.