Symmetry-Preserving Variational Autoencoder and Latent Space Rectified Flow Diffusion for Accelerated Materials Discovery With Geometric Graph Neural Networks and Integrated Property Network Multi-Layer Perceptron

Rezaei, Ryan (School: Canyon Crest Academy)

Advancements in materials science have driven technological progress, yet discovering new materials remains challenging due to resource-intensive laboratory experiments, costly characterization techniques, and time-consuming synthesis processes that limit the systematic exploration of vast chemical spaces. Generative machine learning models offer a promising alternative by autonomously navigating chemical spaces to identify novel, stable materials. In this work, I present a two-stage framework that combines a Variational Autoencoder (VAE) with a latent diffusion model to generate new materials. The VAE, utilizing adapted geometric deep learning architectures, encodes critical geometric and chemical properties—including lattice parameters, atomic species, and coordinates—into a compact latent space that respects crystallographic symmetries and material invariances such as permutation, translation, rotation, and periodicity. I implement a Fourier time-conditioned latent diffusion process enhanced with rectified flow sampling, based on the denoising formulation in Stable Diffusion 3, with a time-adapted modified GemNet architecture for noise removal. Leveraging a curated dataset of nearly 1.5 million materials from the Materials Project and Alexandria databases, my model better captures complex crystallographic properties. By enabling property-targeted generation through integrated gradient-based Multi-layer Perceptrons trained on this new dataset, the generated crystals can be guided toward fine-tuned properties, demonstrating capabilities extensible to sustainable energy, electronics, and environmental technologies, with potential of these models accelerating material discovery by orders of magnitude as this field expands.

Awards Won: