SCALABLE HYBRID LEARNING FOR SCIENTIFIC DATA COMPRESSION

Author's: P. Suvathi[1], S. Rajeshkumar[2], K. Sujith[3], N. Vanjulavalli[4]

Abstract

The exponential growth of scientific data generated from simulations, remote sensing platforms, sensor networks, and high-resolution numerical models has created significant challenges in storage, transmission, and real-time analytics. Efficient compression techniques are essential to reduce data volume while preserving critical structural and statistical properties required for downstream scientific analysis. Traditional linear compression methods such as Principal Component Analysis (PCA) and low-rank approximations offer scalability and interpretability but often fail to capture complex nonlinear patterns inherent in multi-dimensional scientific datasets. Conversely, deep learning–based autoencoders provide powerful nonlinear representation learning capabilities but may demand higher computational resources and reduced interpretability.This project proposes a Scalable Hybrid Learning Framework that integrates classical dimensionality reduction techniques with deep neural network–based residual modeling for efficient scientific data compression. The framework first applies a low-rank linear projection (PCA) to capture dominant global structures in multi-channel spatiotemporal data. Subsequently, a lightweight convolutional autoencoder is trained to model and compress the residual components, enabling accurate reconstruction of fine-grained nonlinear details. By combining linear global compression with nonlinear residual refinement, the hybrid approach achieves superior reconstruction fidelity at competitive compression ratios.