Pre-training Surrogate Models for Hybrid AI Climate Simulations

Abhnil Prasad, David Fuchs, Steven Sherwood, Kirill Trapeznikov, UNSW

Machine Learning (ML) approaches have become popular in producing surrogates replacing physical parameterizations in Global Climate Models (GCMs). While current implementations emulate existing parameterizations, it would clearly be of interest to improve on their ML-based approaches. However, obtaining high-quality training data is expensive. We first test the similarity and generalization of the physics packages from a simpler model (Community Atmosphere Model or CAM4) vs a more sophisticated model (superparameterized CAM or SPCAM) by training two Deep Neural Networks (NNs), one on data from each model, and evaluating each on both datasets. Each physics package is predicted well by its own emulator, and less accurately by the emulator of the other GCM. SPCAM was harder to emulate than CAM4, but the SPCAM emulator generalizes better, probably because SPCAM visits more states. The emulators reproduce different precipitation near the mid-latitude storm tracks, and very different heating and moistening tendencies in the lower troposphere between the two physics models, indicating emulation errors were much smaller than the uncertainty in physics they were emulating. We introduce a pre-training and fine-tuning strategy where a smaller high-fidelity data sample is used to retrain a NN pre-trained on low-fidelity data from an existing model. Fine-tuning the CAM4 emulator with SPCAM data required 10-100 times less SPCAM data to achieve the same skill in emulating SPCAM than training only on SPCAM, showing that this strategy can efficiently utilise scarce high-fidelity data. Finally we discuss results of testing these inline using the recently developed TorchClim plugin, which allows the ML surrogates to be run inside CAM6.