Distributed Cross-Channel Hierarchical Aggregation for Foundation Models

Background/Objective

Scientific foundation models using Vision Transformers (ViTs) face severe compute and memory challenges when handling hyperspectral images with hundreds of channels. Existing distributed methods do not scale along the channel dimension, limiting model size. This work introduces Distributed Cross-Channel Hierarchical Aggregation (D-CHAG) to overcome that limitation, enabling scalable multi-channel learning for scientific imaging.

Approach

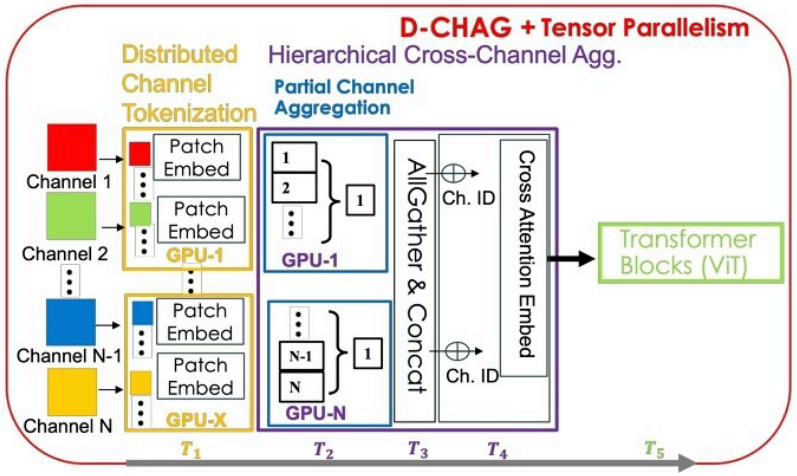
- The D-CHAG method distributes both tokenization and channel aggregation across GPUs with hierarchical cross-attention using data from the Advanced Plant Phenotyping Laboratory (APPL).
- This approach reduces quadratic complexity and memory costs by combining partial-channel aggregation (local) with global aggregation (shared across ranks).
- It is compatible with any ViT or distributed parallelism training strategy (TP, SP, FSDP, DP) and is implemented on the Frontier supercomputer using multichannel imaging datasets.

Results

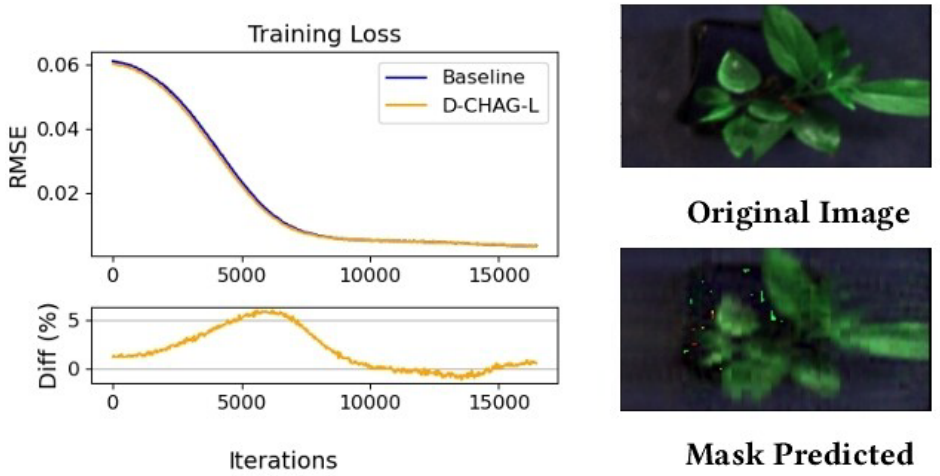
- Up to 75% lower memory use and >2× throughput compared to tensor parallelism alone.
- Efficient scaling to 1,024 AMD GPUs.
- Enables training of 26B-parameter models on datasets with 500+ channels per image.
- Less than 1% loss in accuracy on plant hyperspectral imaging and weather forecasting tasks.

Significance/Impacts

This model represents the first framework to explicitly tackle channel-dimension scaling in scientific foundation models. It can make large, multimodal ViTs feasible for Earth system and biological imaging. Its general, exascale-ready design allows for efficient training of data-rich scientific AI models that can be applied to hyperspectral plant images from APPL, Unmanned Aerial Vehicles (UAVs), and satellites.



D-CHAG method applied in the base architecture



Training loss compared to baseline (left) and results of image prediction from masked autoencoder (right)

Tsaris, A. et al., *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*. ACM. (2025). doi: 10.1145/3712285.3759870