

Thesis Topic: Distributed training of deep learning models for Copernicus data

Background

Deep learning, characterized by a collection of computational neural network models that are composed of multiple processing layers capable of learning distributed representations of data with multiple levels of abstraction, has revolutionized and advanced the state-of-the-art for many research domain problems in the computer vision and geospatial research community. Its techniques seek to automatically discover the representations of large amounts of raw data fed into a machine for performing actions such as prediction and classification tasks. However, the computational complexity of training the multiple processing layers of the end-to-end deep learning models and finding the correct combination of weights from layer to layer and the parameters that change the input data becomes substantially compute-intensive.

This thesis work is part of the H2020 project "[ExtremeEarth: From Copernicus Big Data to Extreme Earth Analytics](http://earthanalytics.eu/)". To read more about the ExtremeEarth project, check out this link: <http://earthanalytics.eu/>.

Problem Statement

The training of sophisticated, large-scale machine learning and deep learning models is very compute-intensive and a critical challenge in the machine learning community. This has fundamentally led to the emergence and increasing demand for powerful compute-resources. Hence, designing an approach for parallelization and distributed training of deep learning models across multiple compute devices connected by a network is important to speed up the task, improve the performances, and make the system scalable and fault-tolerant.

Tasks and Expected Results

- Looking into existing state-of-the-art deep learning frameworks for distributed training.
- Describing the problem statement in detail followed by different parallelization and distributed training strategies for deep learning models.
- Analyzing which among the existing state-of-the-art distributed deep learning frameworks are most suitable for distributed learning in an earth observation settings.
- Designing and implementing a scalable distributed deep learning training model for Copernicus data.

Requirements

This thesis work will require a deeper knowledge of Python programming, machine learning, deep learning models, TensorFlow or PyTorch libraries, and a basic understanding of statistics.

Contact details

If you are interested in working on this suggested thesis topic, feel free to send an email to the following contact details.

Full name	E-mail address	Phone Number	Office Number
Vladimir Vlassov (PhD, Professor, KTH)	vladv@kth.se	+46 736441465	2490
Desta Haileselassie Hagos (PhD, Postdoc, KTH)	destah@kth.se	+46 762555021	4412