High performance image analysis workflow designs for automated mapping of ice-wedge polygons from high-resolution satellite imagery

Mahendra Rajitha Udawalpola Department Natural Resources and the Environment, University of Connecticut

Amit Hasan Department Natural Resources and the Environment, University of Connecticut

Chandi Witharana Department Natural Resources and the Environment, University of Connecticut

Anna Liljedahl Woodwell Climate

Upsurge of peta-byte scale sub-meter resolution satellite imagery archives has been radically redefining the azimuth of Arctic permafrost research. Data processing challenges combined with clamant quest for pan-Arctic scale permafrost modelling efforts spontaneously creating the vacuum for pursuit of artificial intelligence (AI) algorithms, such as deep learning (DL) convolution neural nets (CNNs). Despite the remarkable performances of DLCNNs in everyday image understanding, bottlenecks still exist when translated to geo-object detection from remote sensing imagery. Among others, image dimensions, multiple spectral channels, spatial reference, seasonality, and most importantly semantic complexity aggregated into multiple spatial scales pose greater friction on the performance of DLCNN model predictions. Scalability of automated analysis over millions of square kilometers comprising heterogeneous landscapes reverberates the need for efficient image-to-assessment workflows that center on high performance computing resources (HPC). We have developed a novel high performance image analysis framework - Mapping application for Arctic Permafrost Land Environment (MAPLE) – that enables the integration of operational-scale GeoAI capabilities into Arctic science applications. The MAPLE workflow is three-fold: image preprocessing, DLCNN prediction, and post-processing. While the first and last segments involve CPU implementations, the prediction operates on GPUs. In this study, we have investigated highly-parallelized 4 different workflow designs to implement MAPLE on heterogeneous HPC systems. It is equally important to understand how different workflow designs interact with underlying service unit (SU) accounting models of HPC systems. We systematically analyzed the execution time of image-pre processing, DLCNN model execution, geospatial data generation of each design when predicting ice-wedge polygons and water bodies from satellite imagery. Currently, the MAPLE combines DLCNN algorithms with HPC resources to automatically identify ice-wedge polygons from thousands of commercial satellite imagery at an unprecedented spatial scale and also to detect small scale water bodies in the Arctic permafrost tundra.

Interactive, geospatial visualization of high-resolution, pan-Arctic permafrost features in the Permafrost Discovery Gateway

Matthew Jones National Center for Ecological Analysis and Synthesis, University of California

Robyn Thiessen-Bock National Center for Ecological Analysis and Synthesis, University of California

Christopher Jones National Center for Ecological Analysis and Synthesis, University of California

Lauren Walker National Center for Ecological Analysis and Synthesis, University of California

Anna Liljedahl Woodwell Climate

Chandi Witharana Department Natural Resources and the Environment, University of Connecticut

Mahendra Rajitha Udawalpola Department Natural Resources and the Environment, University of Connecticut

Jason Cervenec Byrd Polar and Climate Research Center, The Ohio State University

Michael Gravina Byrd Polar and Climate Research Center, The Ohio State University

Luigi Marini National Center for Supercomputing Applications, University of Illinois

Todd Nicholson National Center for Supercomputing Applications, University of Illinois

The Permafrost Discovery Gateway (PDG) is a new visualization portal to help researchers understand the extent, spatio-temporal variations, and impact of pan-Arctic permafrost thaw and other phenomena. Geospatial visualization is key to understanding permafrost features, but is challenging at pan-Arctic scales for high-resolution permafrost features of interest. We present the PDG portal, an online application that provides interactive visualization, exploration, and access to geospatial permafrost data products derived from high-resolution satellite imagery (Big Imagery). We incorporate the Fluid Earth Viewer to enable global and regional visualization of Arctic data products over time. We also incorporate Cesium as a tile-based Imagery Viewer that allows exploration of pan-Arctic, sub-meter map products over time and can be exported as publication-quality map images. Workflows will enable a Plot Viewer for interactive data exploration in which researchers can produce their own publication-ready 2D and 4D graphs from aggregated statistics from the geospatial data products. We discuss the initial data layer example of pan-Arctic ice-wedge polygons derived using machine learning from sub-meter imagery and processed into both raster and vector tiles. Future extensions of the portal platform will enable researchers to create publication quality maps and publication quality plots, as well as the ability to download specific subsets of high-resolution data for further analysis locally. Finally, the data processing pipeline has been built so that it can be extended to create visualization layers for new permafrost features of interest to the research community.

Permafrost Discovery Gateway: A project overview

Anna Liljedahl	Woodwell Climate
Matthew Jones	National Center for Ecological Analysis and Synthesis, University of California
Kenton McHenry	National Center for Supercomputing Applications, University of Illinois
Chandi Witharana	Department Natural Resources and the Environment, University of Connecticut
Jason Cervenec	Byrd Polar and Climate Research Center, The Ohio State University
Ingmar Nitze	Alfred Wegener Institute – Helmholtz Centre for Polar and Marine Research
Galina Wind	Science Systems and Applications Inc.
Michael Brubaker	Alaska Pacific University
Benjamin Jones	University of Alaska Fairbanks
Mahendra Rajitha Udawalpola	Department Natural Resources & the Environment, University of Connecticut
Luigi Marini	National Center for Supercomputing Applications, University of Illinois
Todd Nicholson	National Center for Supercomputing Applications, University of Illinois
Amit Hasan	Department Natural Resources and the Environment, University of Connecticut
Amber Budden	National Center for Ecological Analysis and Synthesis, University of California
Robyn Thiessen-Bock	National Center for Ecological Analysis and Synthesis, University of California
Christopher Jones	National Center for Ecological Analysis and Synthesis, University of California
Lauren Walker	National Center for Ecological Analysis and Synthesis, University of California
Michael Gravina	Byrd Polar and Climate Research Center, The Ohio State University
Jennifer Moss	University of Alaska Fairbanks
Guido Grosse	Alfred Wegener Institute – Helmholtz Centre for Polar and Marine Research

The remote and expansive Arctic region is experiencing rapid changes to its permafrost landscapes in response to a warming climate. Simultaneously, more data than ever is created of this region in terms of satellite imagery. The Permafrost Discovery Gateway (PDG) aims to enable the creation of pan-Arctic geospatial data products from sub-meter and monthly scales, while also making the data accessible and discoverable to allow knowledge-generation by scientists and the public. About half-way into our National Science Foundation Navigating the new Arctic funded project, we now have developed operational hybrid machine learning pipelines using Cloud and High Performance Computing Resources for initial use cases based on Maxxar and Landsat pan-Arctic imagery. Newly developed imagery products include first versions of near pan-Arctic sub-meter mapping of ice-rich permafrost as well as a selection of complementing NASA atmospheric data. In the visualization portion of the PDG, we have integrated examples of coarse permafrost geospatial products into the Fluid Earth Viewer (FEV) global visualization tool and incorporated Cesium as a tile-based Imagery Viewer to allow exploration of pan-Arctic, sub-meter map products over time. In addition to refining FEV and the Imagery Viewer, developments will also include workflows that will enable a Plot Viewer for interactive data exploration in which researchers can produce their own publication-ready 2D and 4D graphs from aggregated statistics from the geospatial data products. This presentation aims to give an overview of the multi-institutional project, while additional presentations will give more in-depth insights of the different components under development within the PDG. We greatly welcome the opportunity to interact with the PDG user community through these presentations.

Developing Hybrid Machine Learning Pipelines Using Cloud and HPC Resources for the Permafrost Discovery Gateway

Todd Nicholson National Center for Supercomputing Applications, University of Illinois
Luigi Marini National Center for Supercomputing Applications, University of Illinois
Anna Liljedahl Woodwell Climate
Kenton McHenry National Center for Supercomputing Applications, University of Illinois
Chandi Witharana Department Natural Resources and the Environment, University of Connecticut
Mahendra Rajitha Udawalpola Department Natural Resources and the Environment, University of Connecticut
Ingmar Nitze Alfred Wegener Institute – Helmholtz Centre for Polar and Marine Research

The aim of this work is to build interactive scientific gateways to manage extensible pipelines for large scale execution of ML models on satellite imagery using hybrid Cloud and HPC resources. Initial use cases include ice wedge polygon identification from ArcticDEM data (MAPLE) and thaw slump, lake and fire identification from Landsat data via Google Earth Engine (LandSatRes). We leverage the Clowder data management framework to expose the existing models as interactive web interfaces and to visualize intermediate geospatial datasets. Our current implementation leverages the NCSA Radiant OpenStack Cluster and Kubernetes to manage data locally and provide a container based execution environment for the models. The gateway leverages the XSEDE Bridges2 resource for large scale execution of MAPLE using GPUs (not available on Radiant). Individual files uploaded to the gateway can be submitted for execution directly from the gateway web interface, or a path to existing data available on Bridges2 and moved there off band by the Polar Geospatial Center (in case where the data cannot be widely shared). Outputs are then uploaded back to the gateway using the Clowder web service APIs. Geospatial visualizations are available in the PDG using Clowder geospatial extractors and previewers, and integrate with other PDG visualization tools currently under development, including Arctic wide visualizations of the results. Our hybrid approach reinforces the idea that a single solution does not often fit all use cases and integration of different solutions is often required.

Understanding the effect of image augmentation on deep learning convolutional neural net algorithms

Amit Hasan Department Natural Resources and the Environment, University of Connecticut

Mahendra Rajitha Udawalpola Department Natural Resources and the Environment, University of Connecticut

Chandi Witharana Department Natural Resources and the Environment, University of Connecticut

Anna Liljedahl Woodwell Climate

Easy access to sub-meter resolution imagery repositories positions the permafrost science at the precipice of revolution. We now have the imagery-enabled power to map, monitor, and document even individual permafrost features and their change over time. Artificial intelligence (AI) methods such as deep learning (DL) convolutional neural nets (CNNs), demonstrate a remarkable success in automated analysis of semantically complex imagery from multiple domains. By design, inferential strengths of CNN models are largely fuelled by the quality and volume of hand-labelled training data. Production of hand-annotated samples is a daunting task. This is particularly true for regional-scale mapping applications, such as permafrost feature detection across the Arctic, where landscape complexity would spontaneously inflate the semantic complexity of sub-meter resolution imagery. Additionally, image dimensions, multispectral channels, imaging conditions, seasonality, coupled with multiscale organization of geo-objects pose extra challenges on the generalizability of DLCNN models. Image augmentation is a strategic 'data-space' solution to synthetically inflate the size and quality of training samples without additional investments on hand-annotations. A plethora of augmentation methods have been proposed under the auspices of two general categories: data warping and oversampling. The performances of image augmentations methods largely depend on the image recognition problem in hand and characteristics of underlying data. In this study, we systematically investigate the effectiveness of a spectrum of augmentation methods falling into different cohorts, such as color space, geometrical transformations. feature space, and kernel filter when applied to CNN algorithms to recognize ice-wedge polygons from commercial satellite imagery.

Automated recognition of ice-wedge polygon troughs and human-built infrastructure in the Arctic permafrost landscapes using commercial satellite imagery

Elias Manos Department of Geography, University of Connecticut, Storrs, CT, USA.

Amit Hasan Department Natural Resources and the Environment, University of Connecticut

Mahendra Rajitha Udawalpola Department Natural Resources and the Environment, University of Connecticut

Chandi Witharana Department Natural Resources and the Environment, University of Connecticut

Anna Liljedahl Woodwell Climate

Very high spatial resolution (VHSR) satellite imagery affords permafrost scientists the ability to map and monitor both the natural and human environments at finer scales. Geo-AI mapping applications based on the deep learning (DL) convolutional neural networks (CNNs) have been successful in translating big imagery resources into Arctic science-ready products. However, frontline DLCNN models are computationally intensive, which in turn hampers the scalability of imagery-enabled mapping efforts. Automated permafrost feature recognition is further challenged by the scarcity of large amounts of hand-annotated training samples to cope with inherent landscape complexities in regional-scale model applications. In this exploratory study, we investigated the ability of a lightweight U-Net DLCNN to efficiently perform semantic segmentation of VHSR commercial satellite imagery with a limited amount of training data in automated mapping of: 1) ice-wedge polygon troughs and 2) human-built infrastructure in the permafrost-affected regions of the Arctic. We conducted a multidimensional experiment to understand how data augmentation, image pre-processing, and imagery-derived indices improve the performances of DL-based semantic segmentation of VHSR imagery. In both tasks, we employed data augmentation and tested the impacts of different types on trough and infrastructure recognition to determine the optimal set of augmentations. In the case of trough recognition, we additionally tasked mathematical morphological operations to exploit the geometric structure of the image. We utilized imagery-derived spectral indices and tested the impacts of normalized difference vegetation index (NDVI) and normalized difference water index (NDWI) on trough recognition. With a relatively low number of model parameters, limited labelled training data, short training time, and high segmentation accuracy, our findings suggest that overall, the U-Net DLCNN, coupled with data augmentation and mathematical morphology, could serve as an accurate and efficient method for mapping both natural and human-built features in the Arctic permafrost environment without compromising spatial details and geographical extent.