Harnessing high performance computing for large-scale imagery-enabled polar research

HEATHER J. LYNCH[1], SHANTENU JHA[2], BENTO GONCALVES[1], BRADLEY SPITZBART[1], MICHAEL WILLIS[3], MICHAEL MACFERRIN[3], VENA CHU[4], MARK SALVATORE[5], MATTEO TURILLI[2]

[1]Stony Brook University, [2]Rutgers University, [3]University of Colorado, Boulder, [4]University of California, Santa Barbara, [5]Northern Arizona University

The polar regions are critical to our understanding of climate and biogeochemical cycling, but traditional constraints imposed by their remoteness have made it difficult to even map some of these areas, no less to understand the mechanisms that link the region's geology, hydrology, and biology. Over the last decade, however, there has been an extraordinary increase in the capture and use of high-resolution satellite imagery in polar areas. As the community moves from smaller-scale projects to demonstrate feasibility, to regular (or even real-time) pan-Arctic and pan-Antarctic surveys, it has become clear that further progress in imagery-enabled science requires the development of cyberinfrastructure to unite high-performance and distributed computing resources with polar imagery and the tools required for their study (software and code for analysis). Here we demonstrate the use of our new developing cyberinfrastructure (ICEBERG - Imagery Cyberinfrastructure and Extensible Building-Blocks to Enhance Research in the Geosciences) for a pan-Antarctic pack-ice seal survey. To accomplish this survey, we are using convolutional neural networks for imagery annotation, an approach of broad utility for a range of biological and geological applications involving imagery interpretation and one that requires the careful and efficient coordination of imagery and high performance and distributed computing. We will also introduce several of the other use cases being used to develop ICEBERG's functionality, which we expect will include much of the functionality required by the larger EarthCube community.

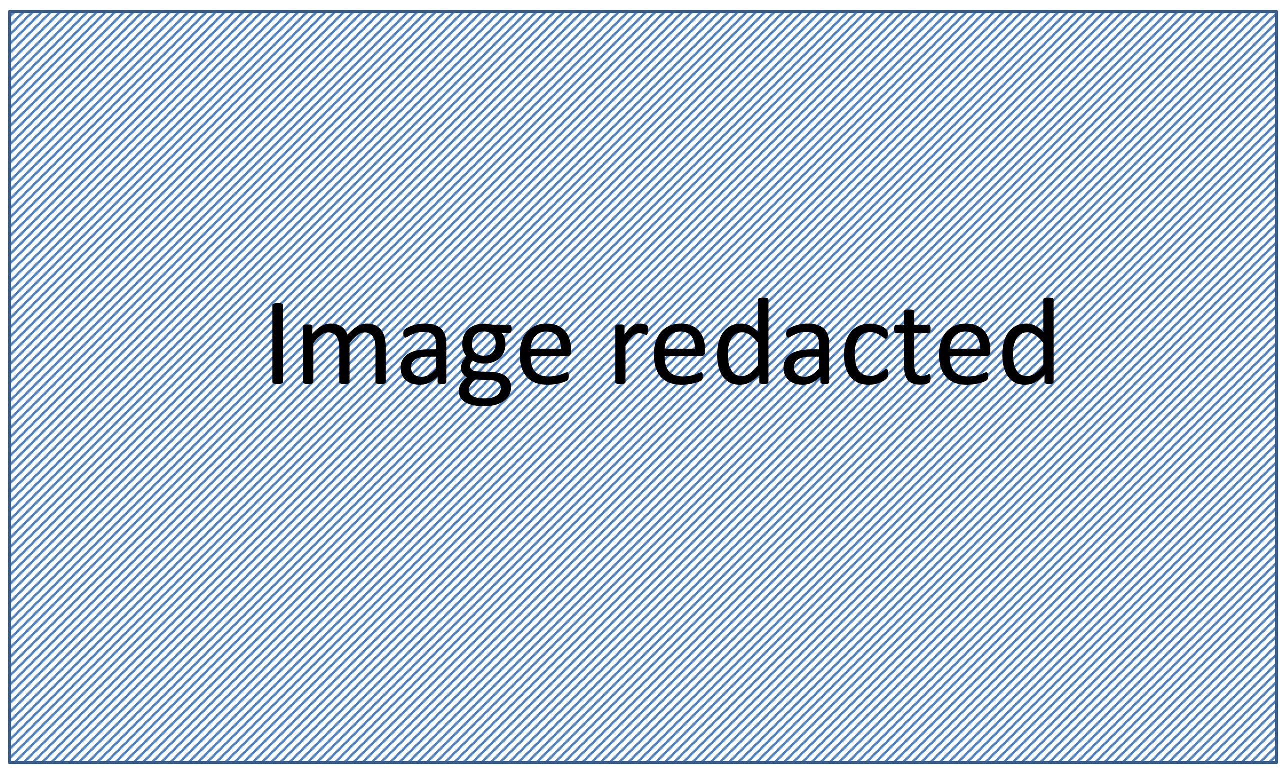
PROJECT AIMS

Aim 1 To develop open source image classification tools tailored to high-resolution satellite imagery of the Arctic and Antarctic to be used on HPDC resources

Aim 2 To create easy-to-use interfaces to facilitate the development and testing of algorithms for application specific geoscience requirements

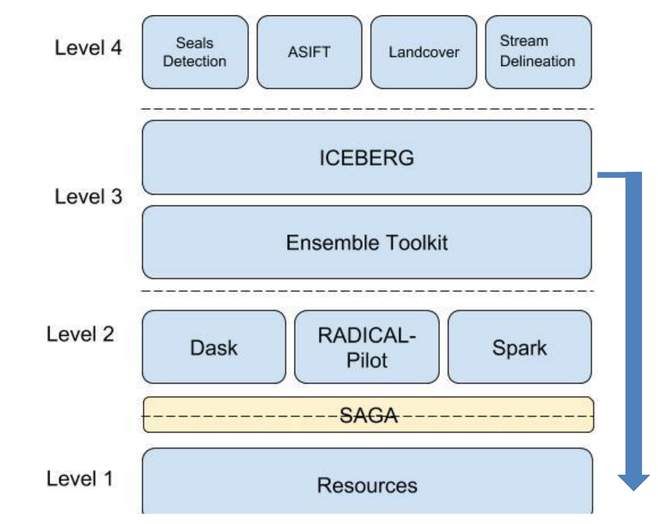
Aim 3 To apply these tools through four use cases that span the biological, hydrological, and geoscience needs of the polar community

Aim 4 To transfer these tools to the larger (non-polar) EarthCube community for continued community driven development

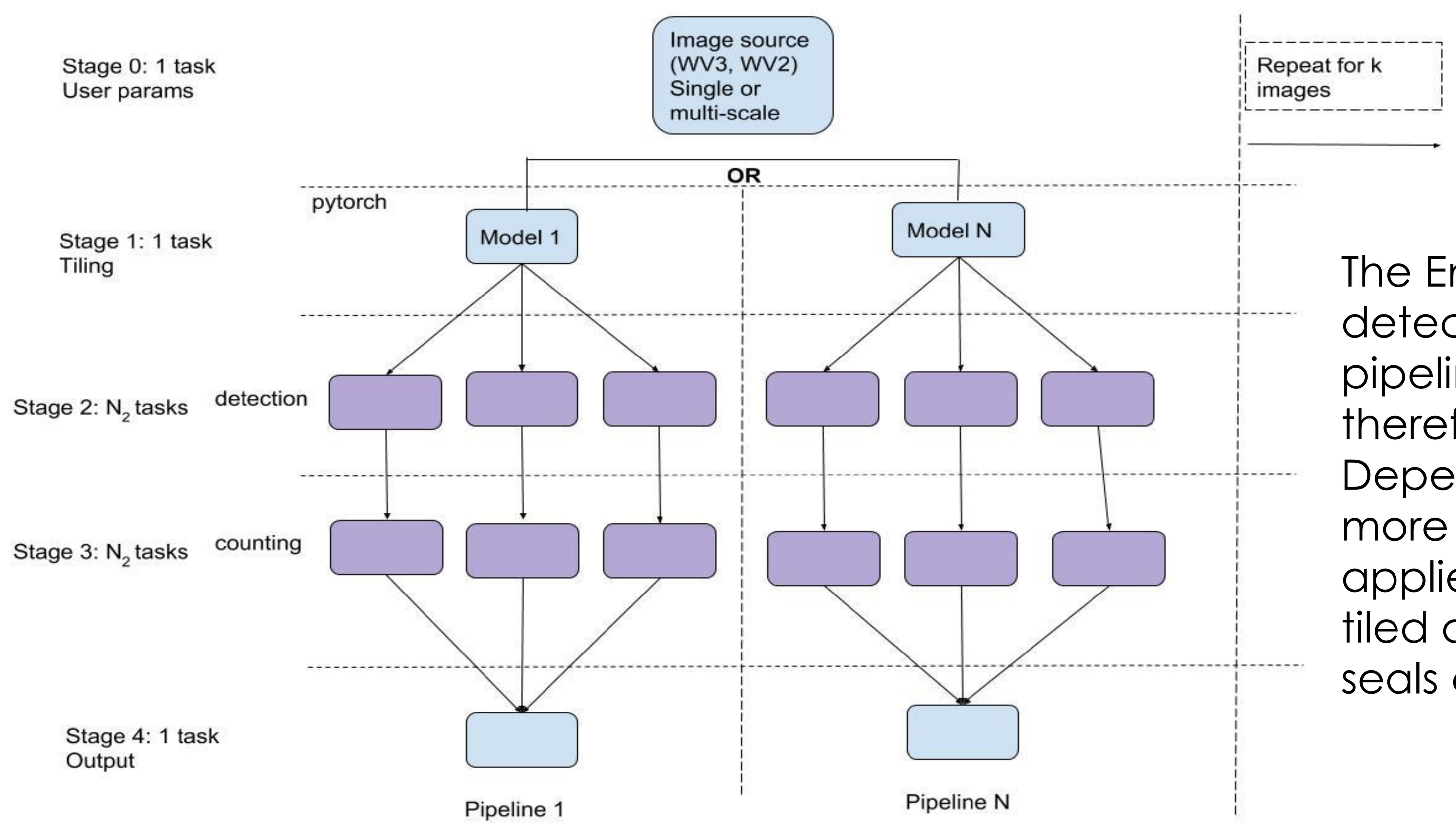


Examples of imagery-enabled geosciences motivating ICEBERG's development. (A) Identifying photosynthetic microbial colonies in the McMurdo Dry Valleys using Worldview-based Normalized Difference Vegetation Index (NDVI). (B) Moulin discharge in Greenland classified with high-resolution satellite imagery. (C) Test images demonstrating how scenes with and without seals can be differentiated. (D) Volume changes from the Vavilov Ice Cap, Russia calculated by co-registration and differencing of DEMs.

DESIGN & ARCHITECTURE



ICEBERG is the first middleware step in processing the pipeline. For each use case in Level 4, ICEBERG determines the tasks and resources needed to pass on to the Ensemble Toolkit (EnTK) (Level 3). EnTK assembles the pipelines to pass to a pilot (Level 2) which then interfaces with XSEDE compute resources (Level 1) through SAGA (A Simple API for Grid Applications).



The EnTK architecture for detecting seals consists of kN pipelines, each independent and therefore can run in parallel. Depending on user input, one or more training model can be applied. The full scale image is tiled and each tile is classified and seals counted in parallel.

EARLY RESULTS

Model training takes 2 – 24 hours, depending on the CNN, with PyTorch and one GPU (NVIDIA GeForce 1080).

At right is the confusion matrix for a CNN achieving 90-95% accuracy across classes.

	crabeater	weddell	pack-ice	other	emperor	open-water	ice-sheet	marching-emperor	crack	glacier	rock
rock	1	1	9	9	1	17	28	1	17	2	382
glacier	0	0	1	0	0	0	88	0	13	466	18
crack	0	0	0	0	0	0	10	1	109	10	1
marching-emperor	4	11	0	0	0	0	0	106	2	1	0
ice-sheet	0	1	0	2	0	0	2042	2	42	48	12
open-water	0	0	10	2	0	987	173	1	2	1	2
emperor	7	3	1	0	678	0	3	9	0	1	2
other	0	1	1	156	0	4	1	0	0	0	0
pack-ice	8	2	1613	22	0	34	0	0	12	12	12
weddell	0	99	1	0	0	0	0	1	0	1	0
crabeater	459	1	4	0	0	0	0	0	0	0	0

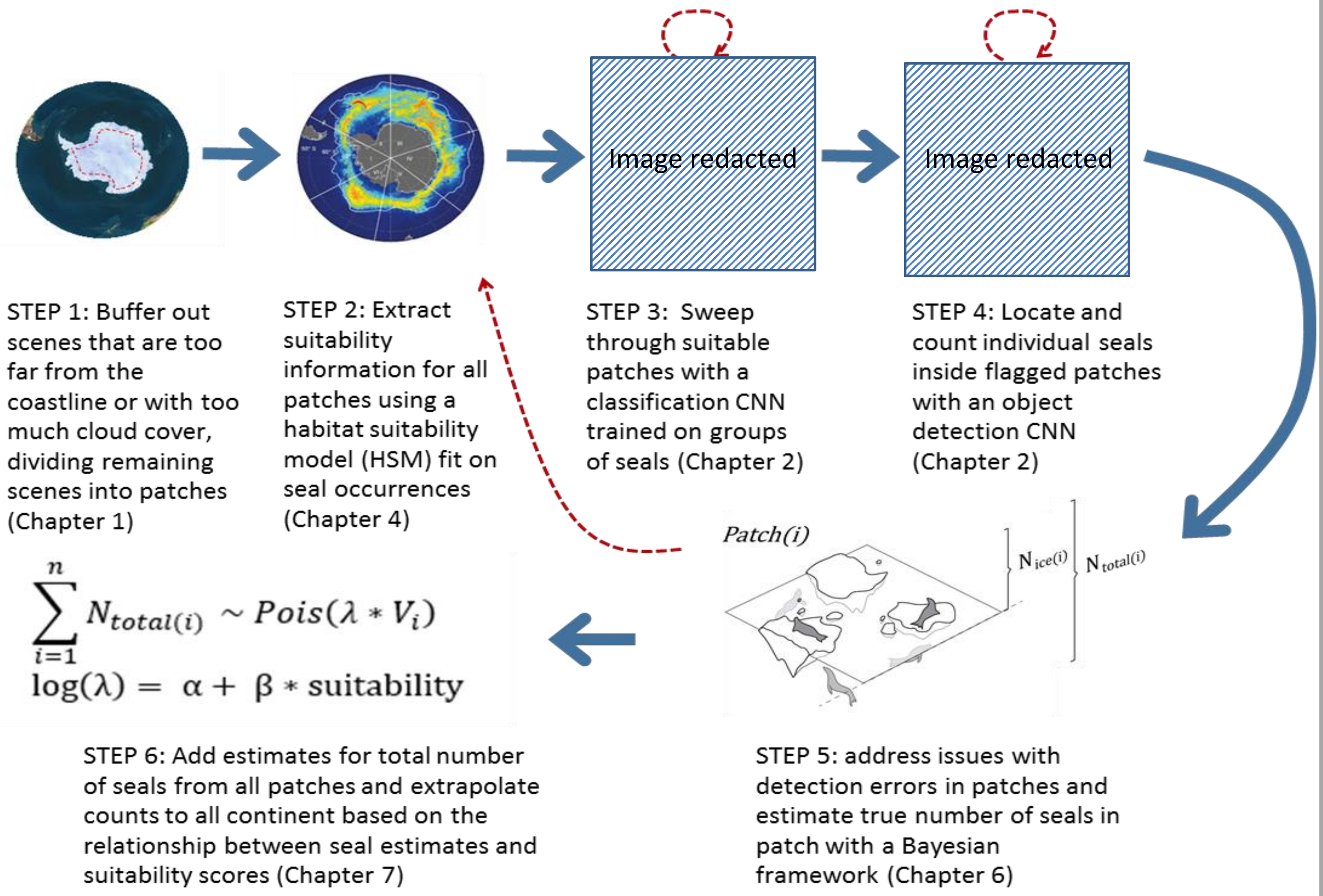
EDUCATION AND OUTREACH

Thanks to Rutgers Electrical and Computer Engineering Capstone 2018 students: Jake Lewandowski, Aashrit Kondapalli, Alexander Dewey, and Raj Patel for laying the groundwork in setting up pipelines, transferring files, writing batch scripts, and documentation.

USE CASES IN DEVELOPMENT

Pack-ice Seal Detection (Bento Goncalves, SBU)

- Produce a functional, extensible, open-source pipeline for seal detection using machine learning and high-performance computers that will be immediately useful to ongoing international efforts to monitor pack-ice seals, and which can be easily adapted for other large-bodied species visible from high-resolution satellite imagery.

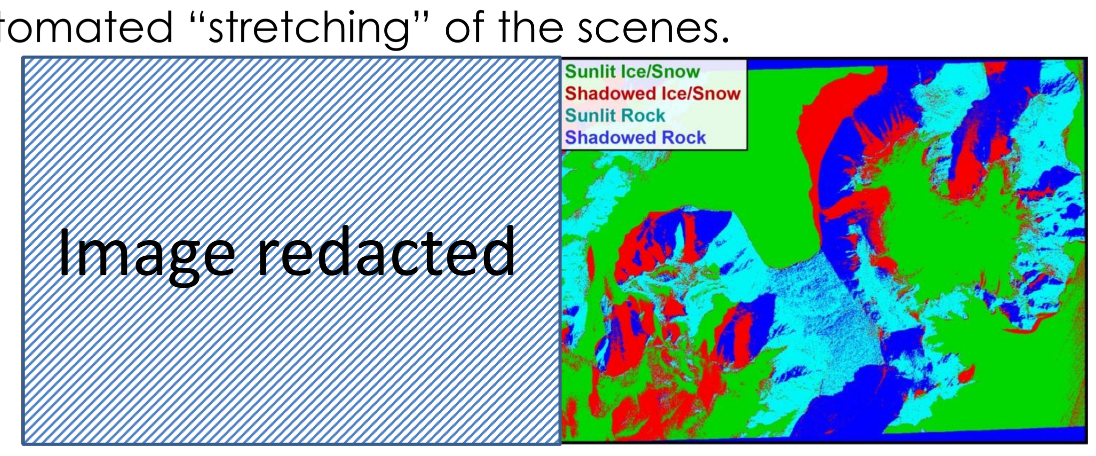


Penguin Colony Detection (Hieu Le, SBU)

- Automated recognition of penguin colonies using WorldView imagery and convolutional neural networks

Large-Scale Spectral Mapping (Mark Salvatore, NAU)

- Develop a pipeline for automated processing of satellite imagery, automated detection and removal of snow, ice, water, and shadows from the scene, automated atmospheric characterization and removal, and automated "stretching" of the scenes.



(Left) Visible image of Ong Valley, Miller Range, Central Transantarctic Mountains. (Right) An automatically generated surface unit map produced using multispectral data and parameters.

Stream Delineation (Vena Chu, UCSB)

- Develop a pipeline to recognize glacial streams (as distinct from other glacial features in Worldview-2 imagery) over large spatial scales using EnTK's PST Model

ASIFT Automated ground-control and ortho-rectification of non-/poorly-geolocated aerial photos (Mike Willis, Mike MacFerrin, John Ohman (UC Boulder)

- Match non-geolocated (historical) airborne imagery to ortho-rectified WorldView satellite imagery and apply ortho-rectification

For more information:

https://iceberg-project.github.io/
http://radical-cybertools.github.io/