RUTGERS
THE STATE UNIVERSITY OF NEW JERSEY

I.C.E.B.E.R.G.

Imagery Cyberinfrastructure and Extensible Building Blocks to Enhance Research in the Geosciences

Jake Lewandowski, Alex Dewey, Ash Kondapalli and Raj Patel

Shantenu Jha and Matteo Turilli

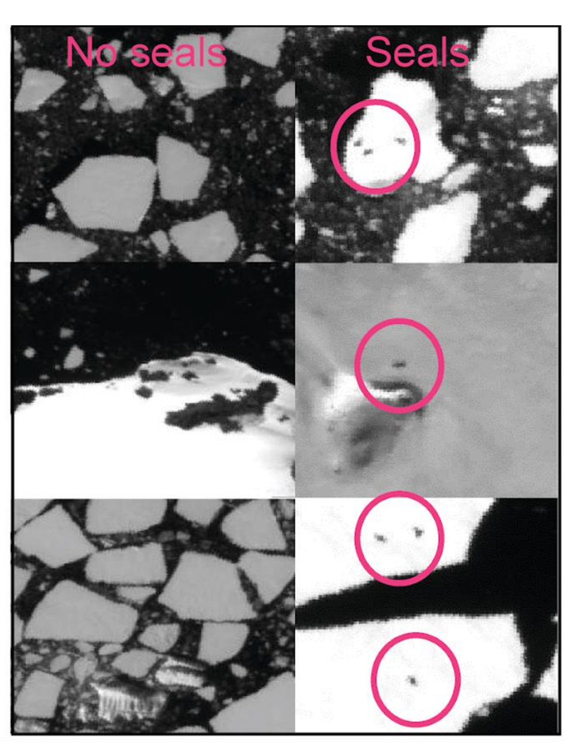
Goal

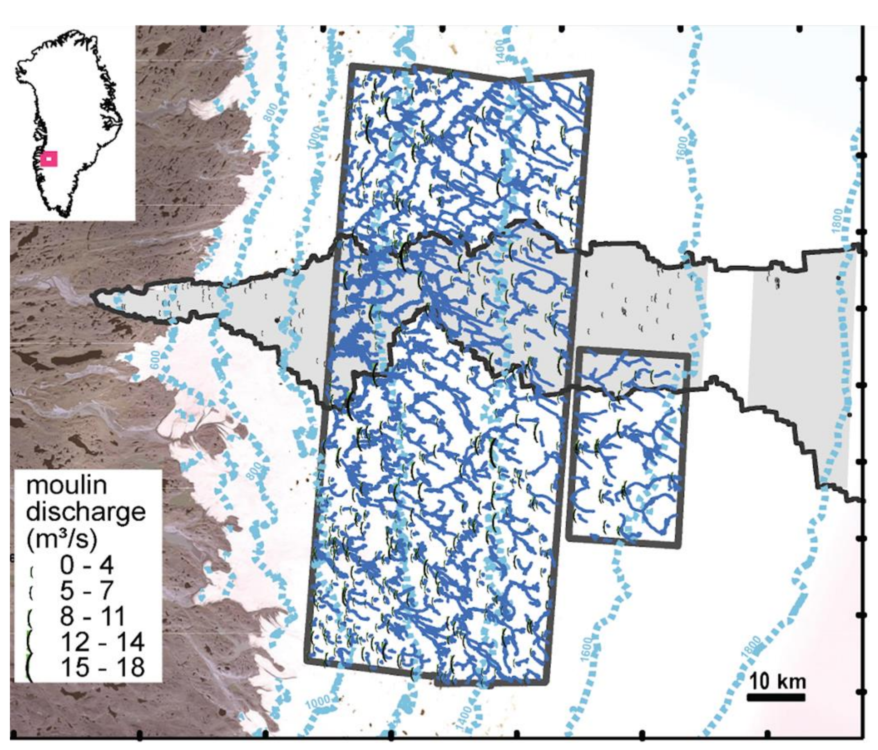
- The goal of ICEBERG is actual quite simple. Our focus is to help uncover the Arctic for scientific research. The implementation is the challenging part. With access to satellite images (WorldView-3), from the National Science Foundation, and allocation on the Bridges supercomputer, hosted at the Pittsburgh Supercomputing Center, we are capable of processing these images. Using parallelization techniques, processing a large set of images, efficiently, to identify biological life and geology in the Arctic can be realized.

File Transfer → Training → Validation → Classification → File Output

Motivations and Objectives

- Motivations
  - With global warming becoming an increasing concern in the world, helping scientists study the Arctic as much as possible is prevalent in understanding what the future holds.
  - As there are thousands and thousands of Arctic images that must be analyzed, providing an efficient, scalable method to process these images is essential
- Objectives
  - Identify seals from satellite images of the Arctic to study: migration patterns, population dynamics and abundance of life.
  - Delineate glacial streams from lakes, slush, and other water features in images of the Greenland Ice Sheet
  - Understand how to run jobs on a supercomputer and utilize it's computational efficiency to process a pipeline of code

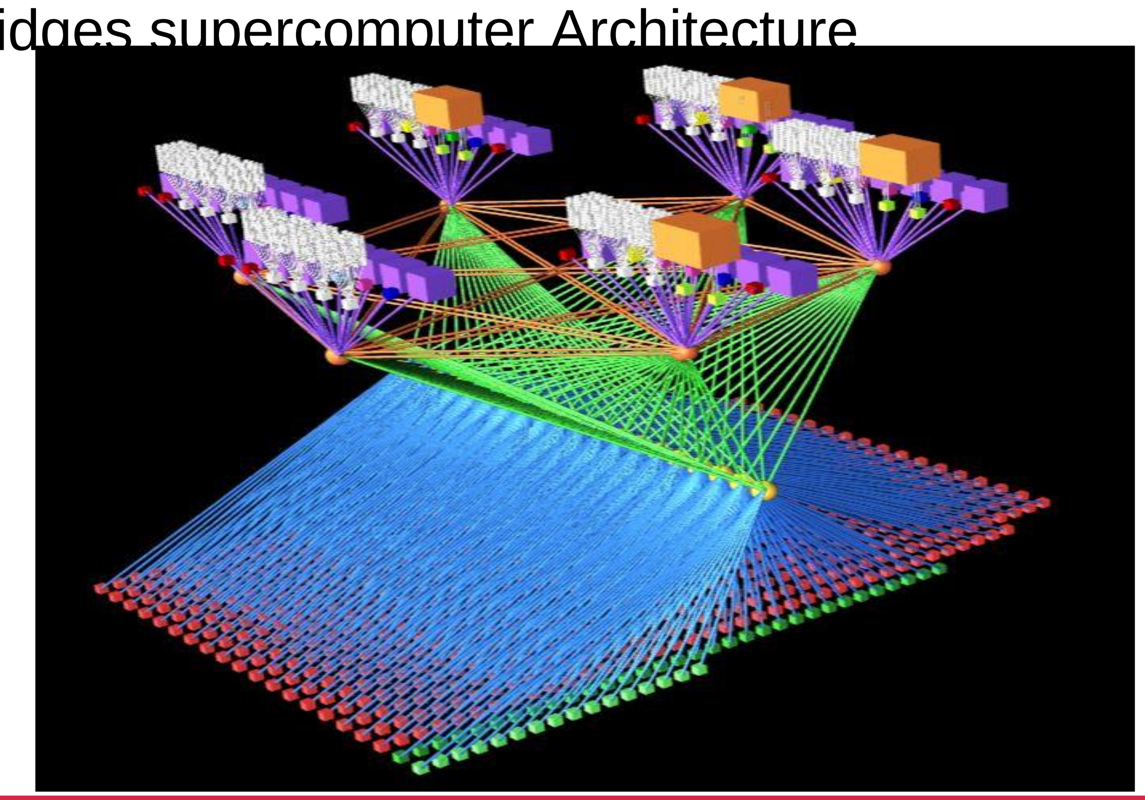




Supercomputing

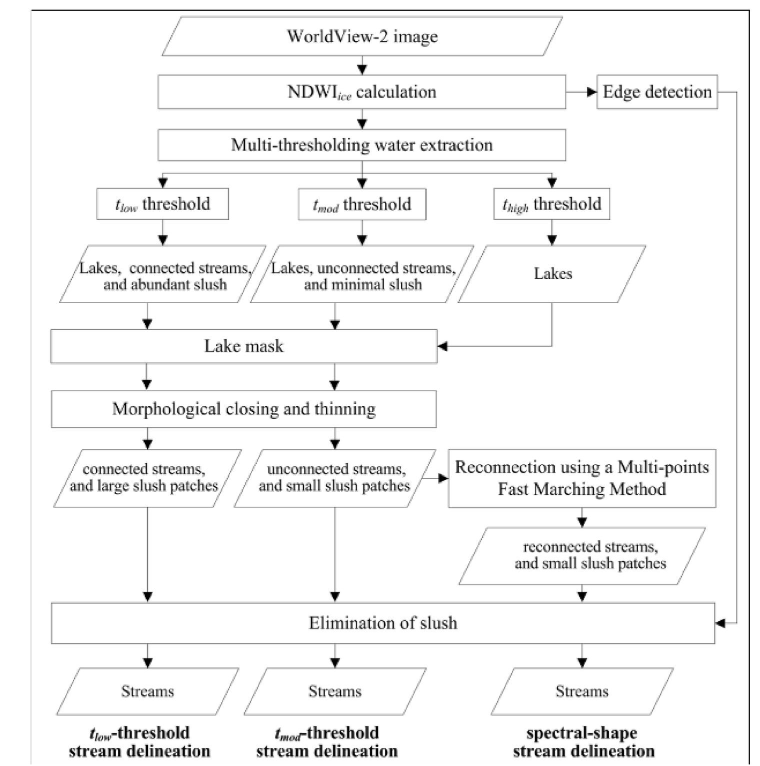
Explanation

- A supercomputer is a computer with a high level of performance compared to a general-purpose computer.
- The following diagram provides a model of the Bridges supercomputer Architecture



Methodology

- Workflow representing the different stages of the stream delineation



- Diagram explaining the PST (Pipeline, Stage, Task) Model. Tasks are contained within stages, which are contained within a pipeline.

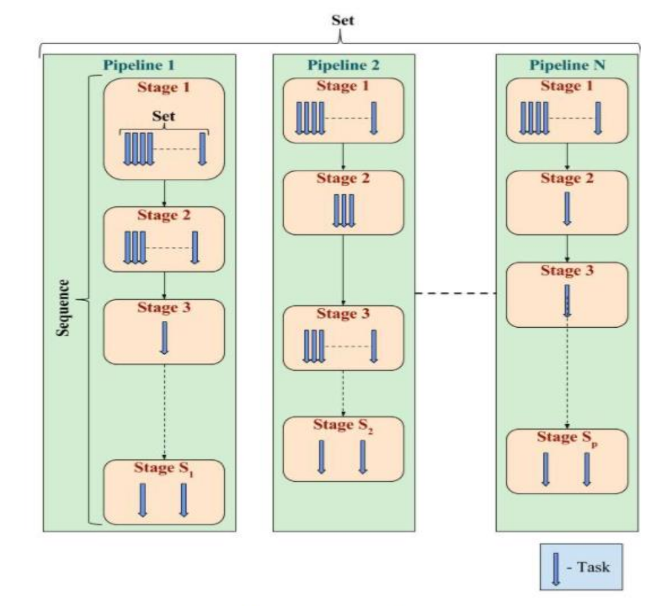
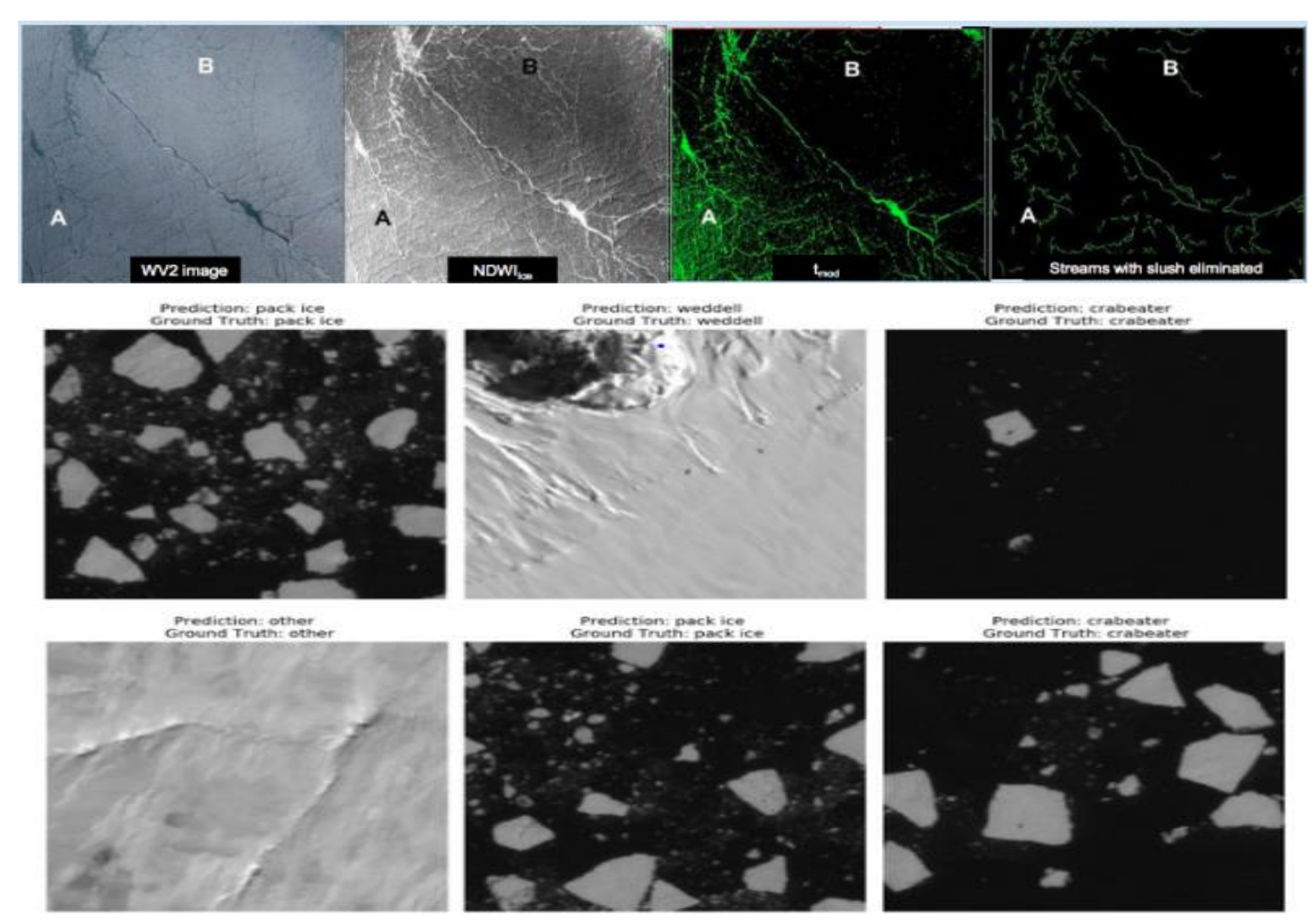


Figure 1: PST Model

Results

- Currently able to predict at 95.83 % validation accuracy with across 5 classes.
- Test different ratios to optimize amount of images processed in a single run and time frame.



References

[1]https://www.psc.edu/resources/computing/bridges
[2] https://portal.xsede.org/lsu-supermic
[3]https://radical-cybertools.github.io/

Research Challenges

- Learning how to access and run jobs on a supercomputer
- Integrating RADICAL Cybertools (Pilot, Ensemble Toolkit) to create a pipeline to process images
- Optimizing parallelization techniques to process images in the most efficient manner
- Training a neural network to identify images

Acknowledgement

We would like to thank the Pittsburgh Supercomputing Center, National Science Foundation, Heather Lynch and Bento from Stony Brook, and, of course, Professors Jha and Turilli.