MINERAI DE FER QUÉBEC
QUEBEC IRON ORE

UBC Master of Data Science

QIO Geo-Detection Project

Segmenting rock types using Machine Learning

Jingyun Chen, Socorro Dominguez, Milos Milic, Aditya Sharma

I.R.O.N.N.

Table of Contents

1. Business Problem
2. Data Science Objectives
3. Data Product/Scientific techniques
4. Challenges
5. Future Steps

Business Problem

Currently, a trained geologist is needed to identify different rock types.

QIO needs a prototype that will help:

- Segmenting rock types from an active mining face image

Data Science Objectives

Labeling the data:

- LabelMe

Understanding the data:

- Colour and camera feature analysis

Building the machine learning model:

- Image segmentation and classification

Improving the model:

- Design tools to improve data quality

Labeling the Data: Origin State

MINERAI DE FER QUÉBEC
QUEBEC IRON ORE

UBC
Master of Data Science

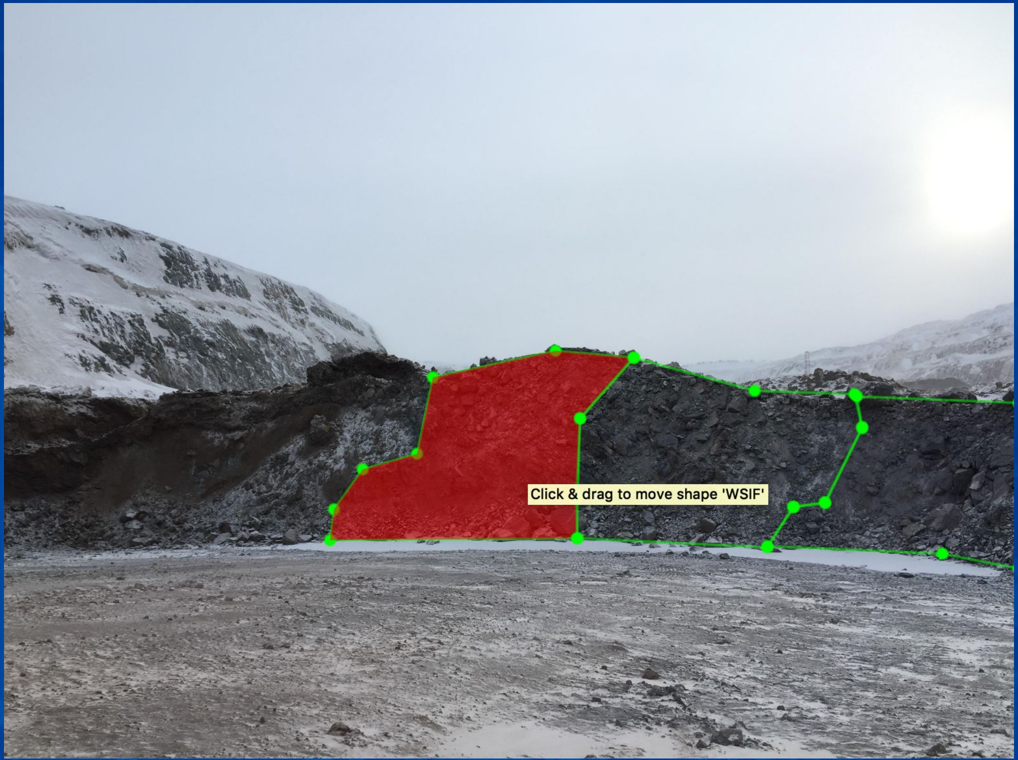
Hand annotations

- cannot be used to train a machine learning model

Labeling the Data: Solution

[LabelMe]

- Obtain coordinates that are useful for EDA and model training



Understanding the Data: Image 101

MINERAI DE FER QUÉBEC
QUEBEC IRON ORE

UBC
Master of Data Science

Pixel:
- a small block on an image

Pixel value (0~255):

0: black
255: white

RGB Image:
- each image consists of 3 layers/channels (Red, Green, Blue)

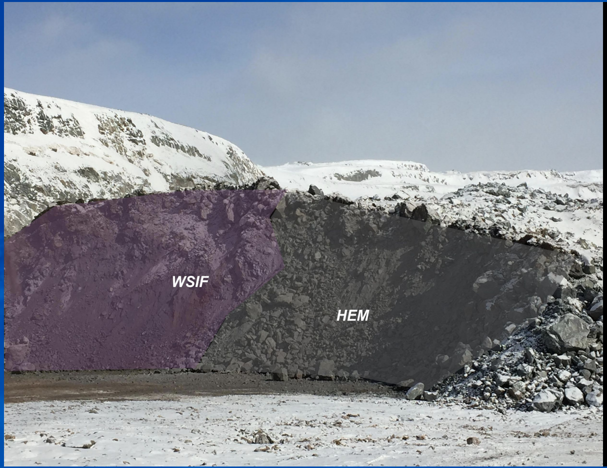
MINERAI DE FER QUÉBEC
QUEBEC IRON ORE

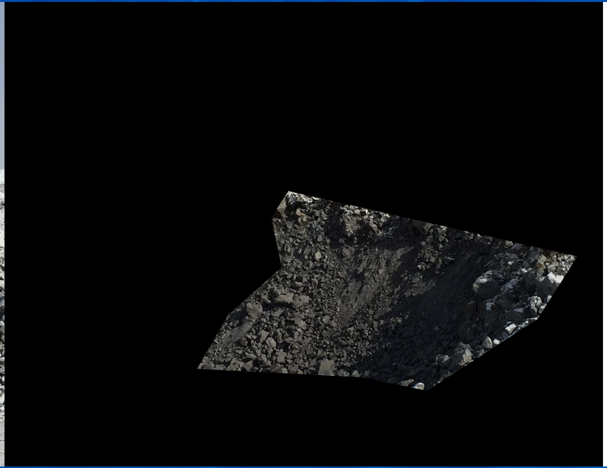
UBC
Master of Data Science

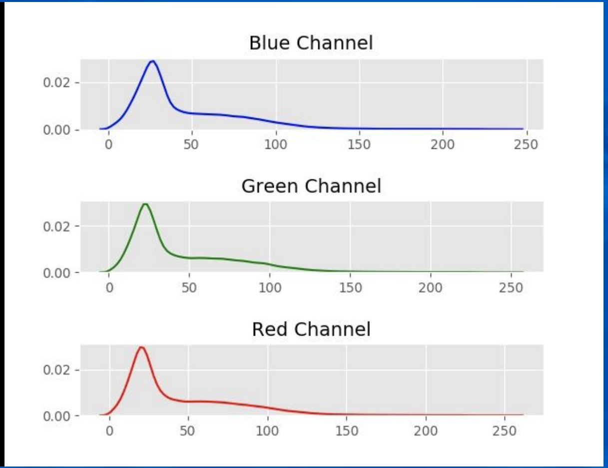
Understanding the Data: Colour Features

Purpose: intuitive way to understand image data

Method: extracted 9 features including mean pixel on each of the 3 channels for each rock type in every image



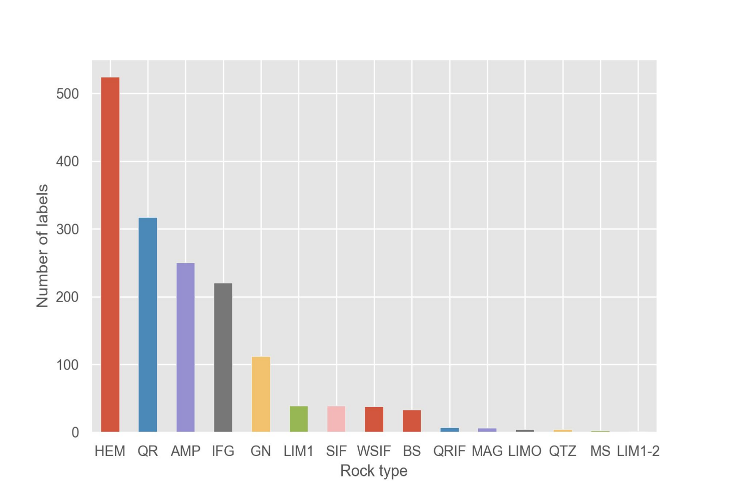


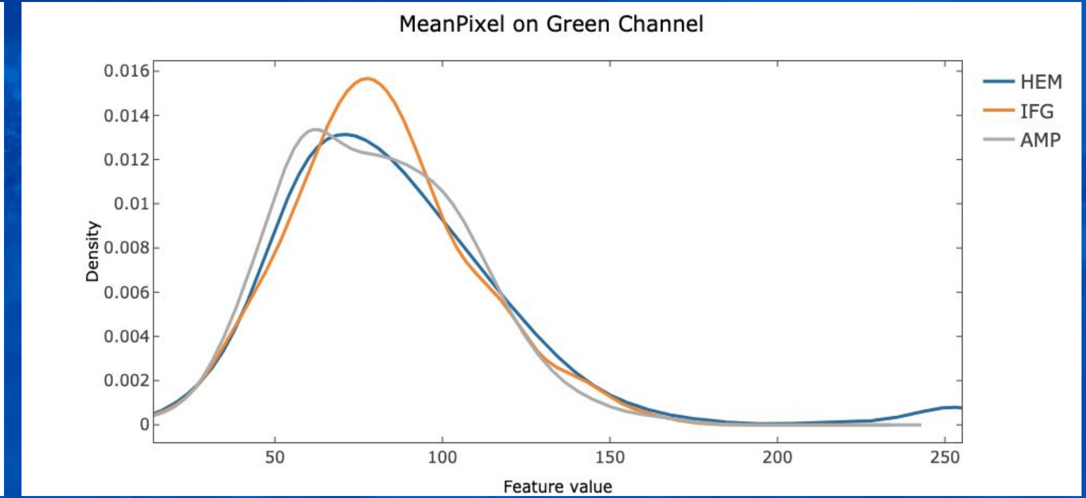


Understanding the Data: Colour Features

Statistical test: analyze if the rocks are “easily differentiable” from each other

Problem: hard to distinguish due to imbalanced data





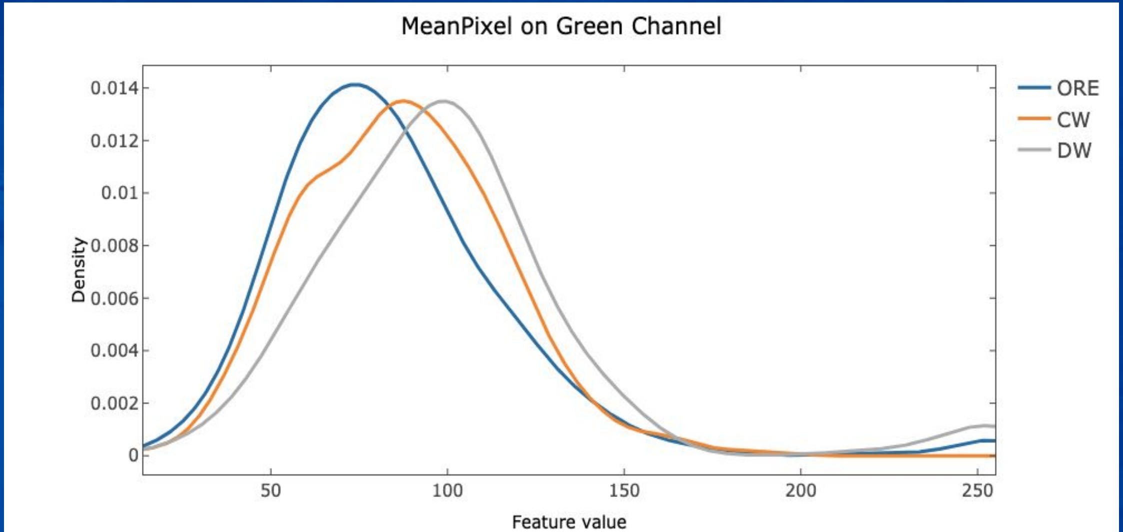
Understanding the Data: Colour Features

MINERAI DE FER QUÉBEC
QUEBEC IRON ORE

UBC Master of Data Science

Combine rocks into 3 categories:

- Ore
- Dilution waste
- Contamination waste



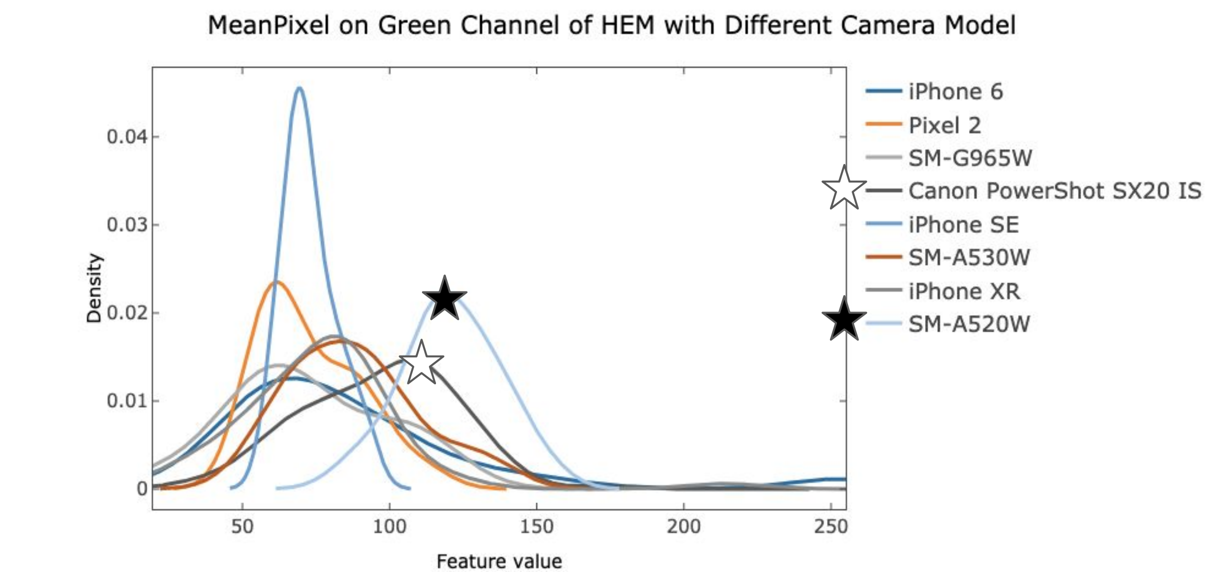
Understanding the Data: Camera Features

Purpose:

- Check camera effect

Method:

- Collected data such as phone type, exposure time, focal length, etc.



Camera Features: Impact

Images on the right:

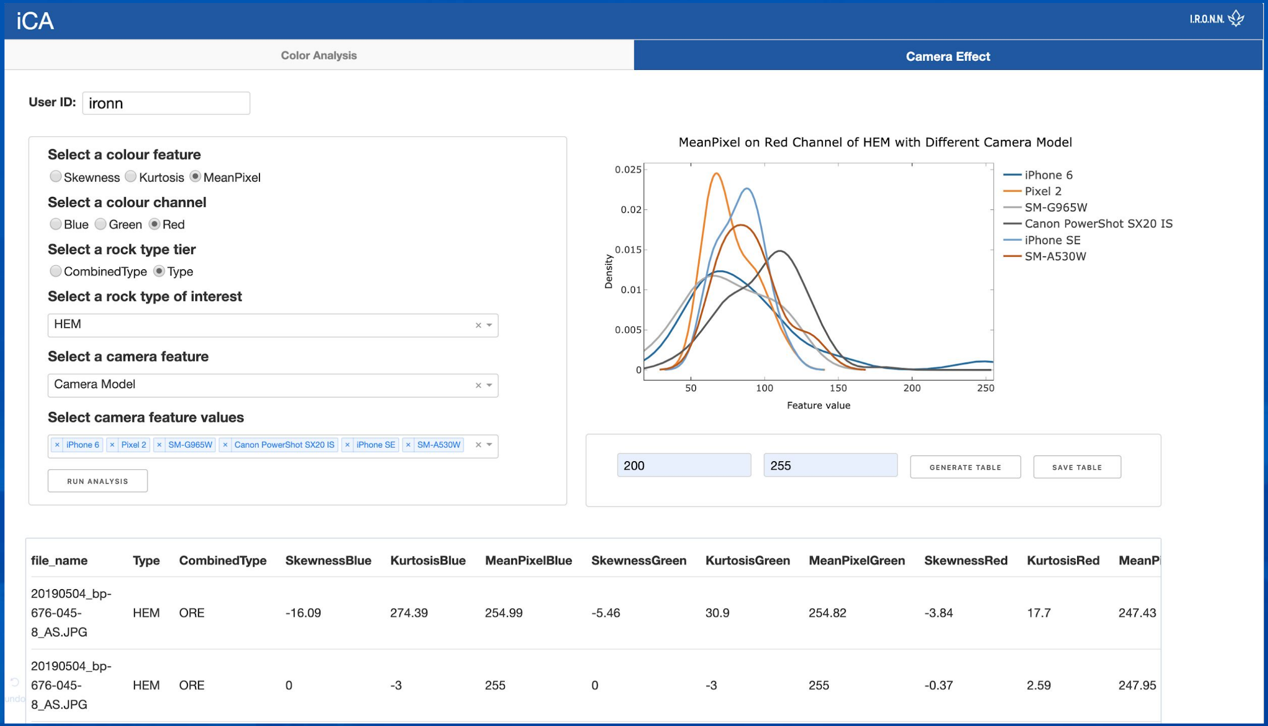
- Similarity: contain *hematite*
- Difference: red colour profile

Image	Camera	MeanPixelRed
	SM-A520W	148.13
	iPhone 6	70.53

Data Product: iCA

Plotly Dashboard

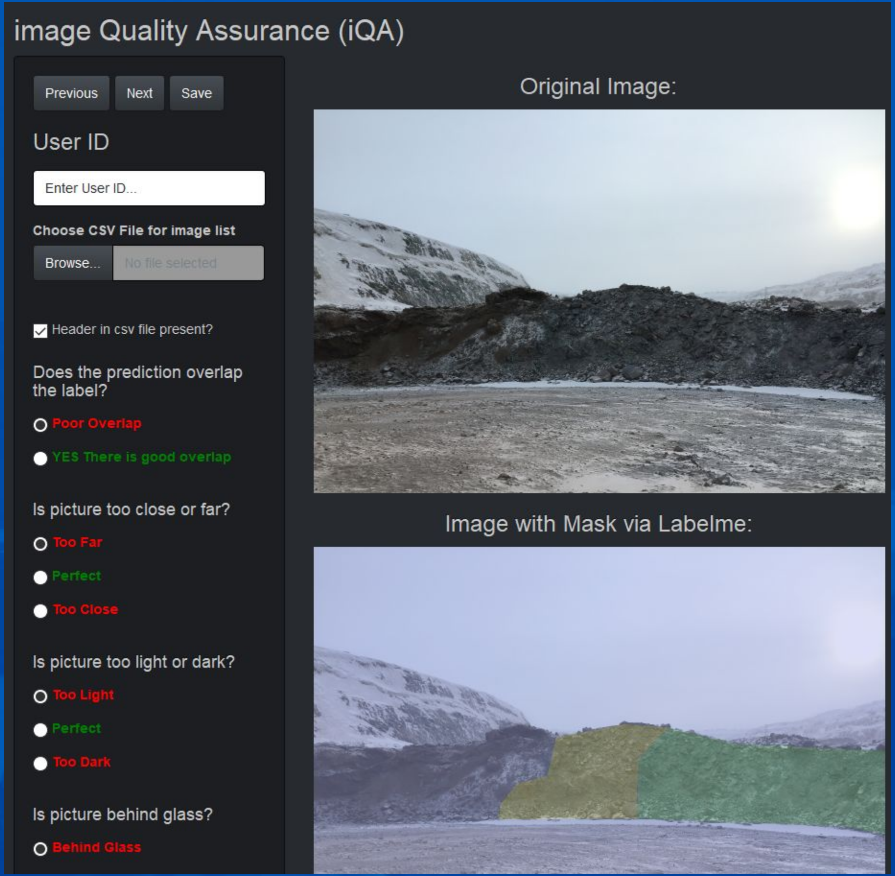
- Interactively identify images that have outstanding features (outliers)
- Can be used for image quality check with iQA



Data Product: iQA

Shiny Dashboard

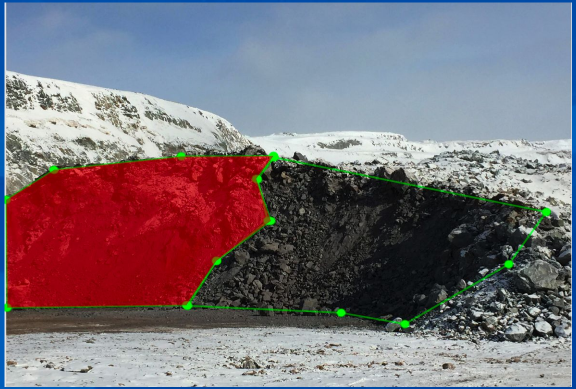
- Check if there are any qualitative issues with original images, labels, predictions (blurriness, machinery present, etc)

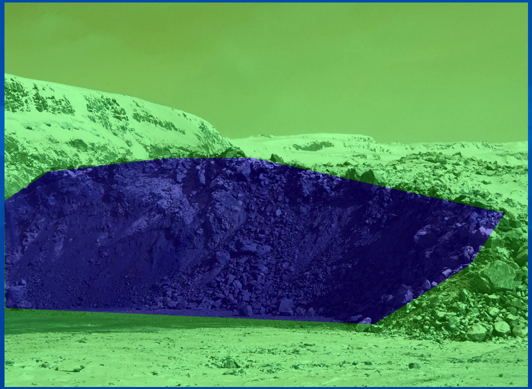


Main Data Product: I.R.O.N.N.

First model: I.R.O.N.N. V0.1

- See if we can distinguish whole rock face from background (sky, ground, etc.)
- Created convex hulls for polygons/blasted face
- Trained on ~430 images



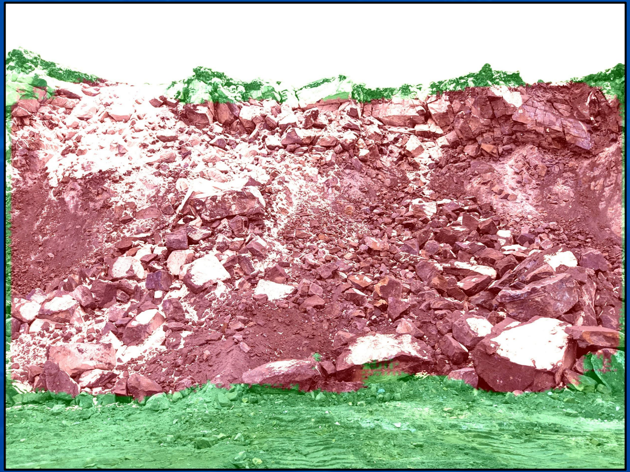


I.R.O.N.N. V0.1: Result

Original image and the model output for an example image where:

- **Red**: blasted face
- **Green**: rest of the image/background

Testing the model:



MINERAI DE FER QUÉBEC
QUEBEC IRON ORE

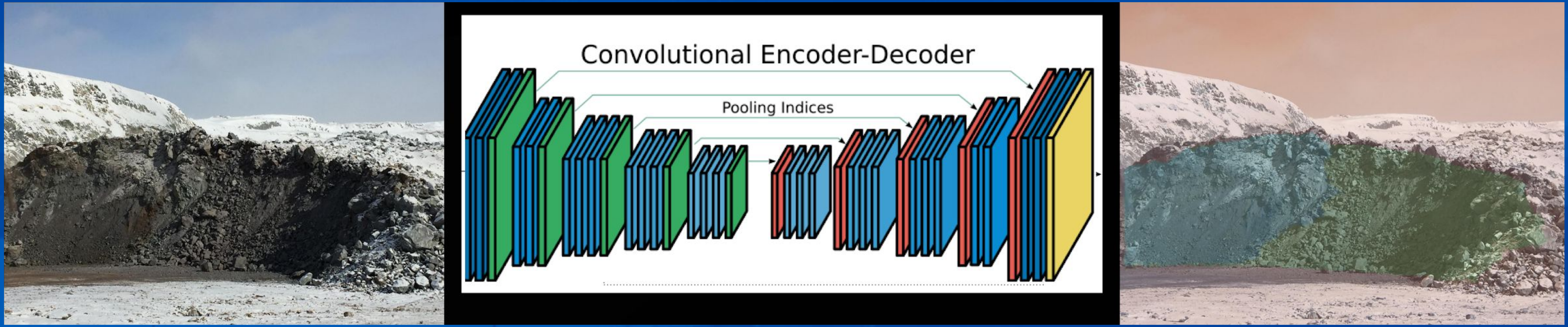
UBC
Master of Data Science

I.R.O.N.N. V1.0: Architecture

Currently using Fully Convolutional Networks for Image Segmentation

Takes in images and spits out images

Framework: PyTorch



I.R.O.N.N. V1.0

MINERAI DE FER QUÉBEC
QUEBEC IRON ORE

UBC Master of Data Science

1. Original unlabeled image

I.R.O.N.N. V1.0

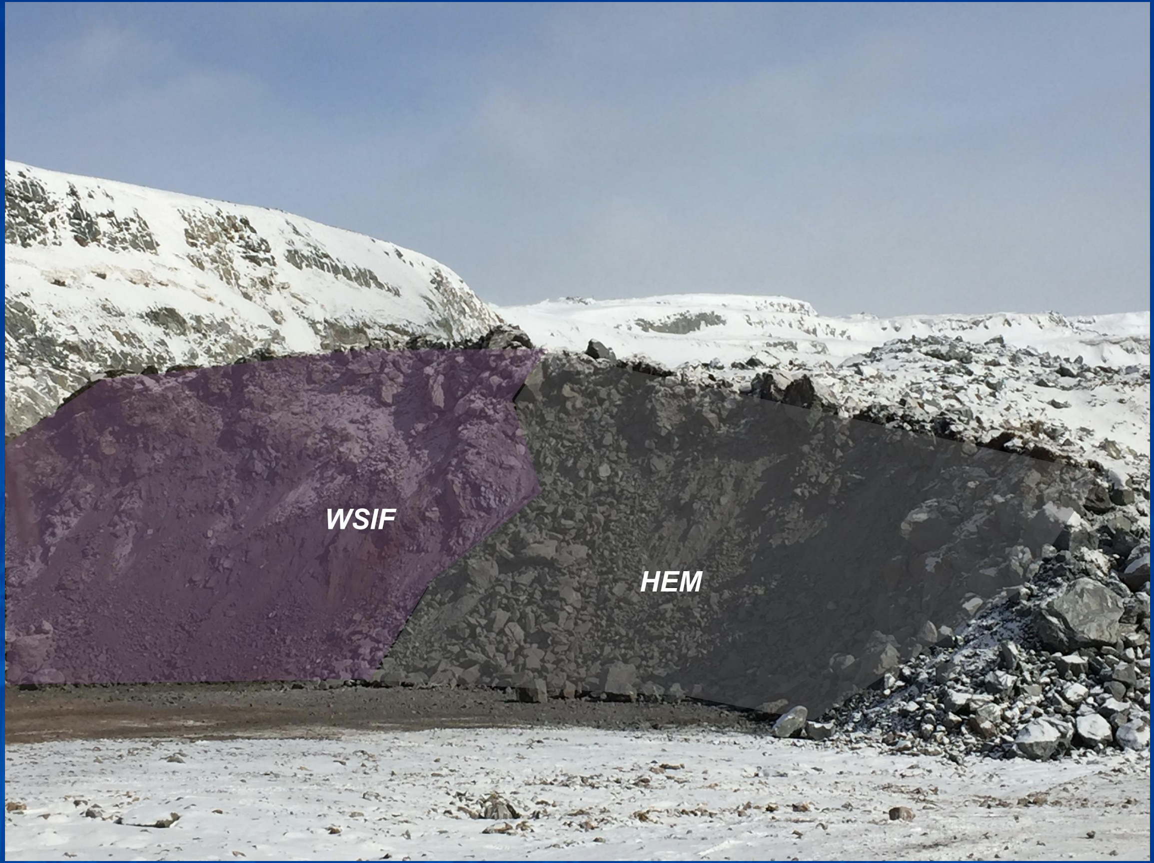
MINERAI DE FER QUÉBEC
QUEBEC IRON ORE

UBC
Master of Data Science

2. Labeled image

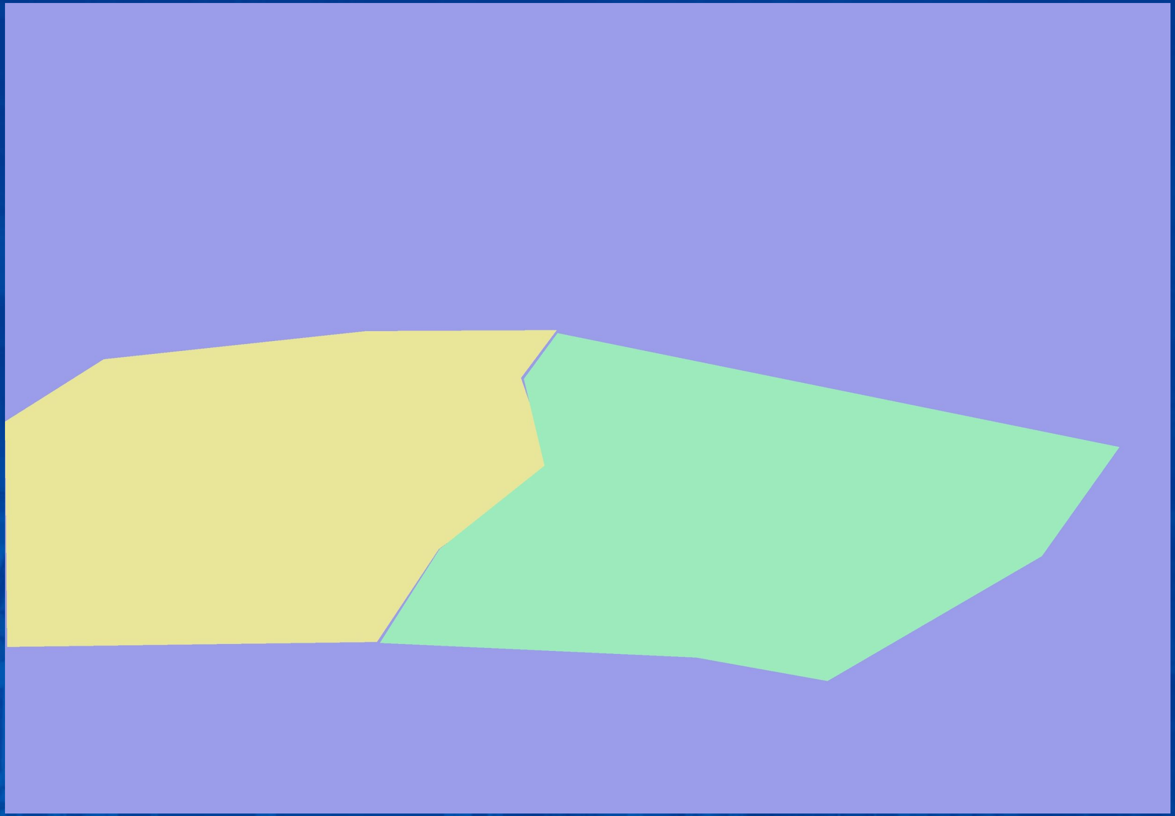
HEM: Ore

WSIF: Contamination waste



I.R.O.N.N. V1.0

3. Created mask from the label

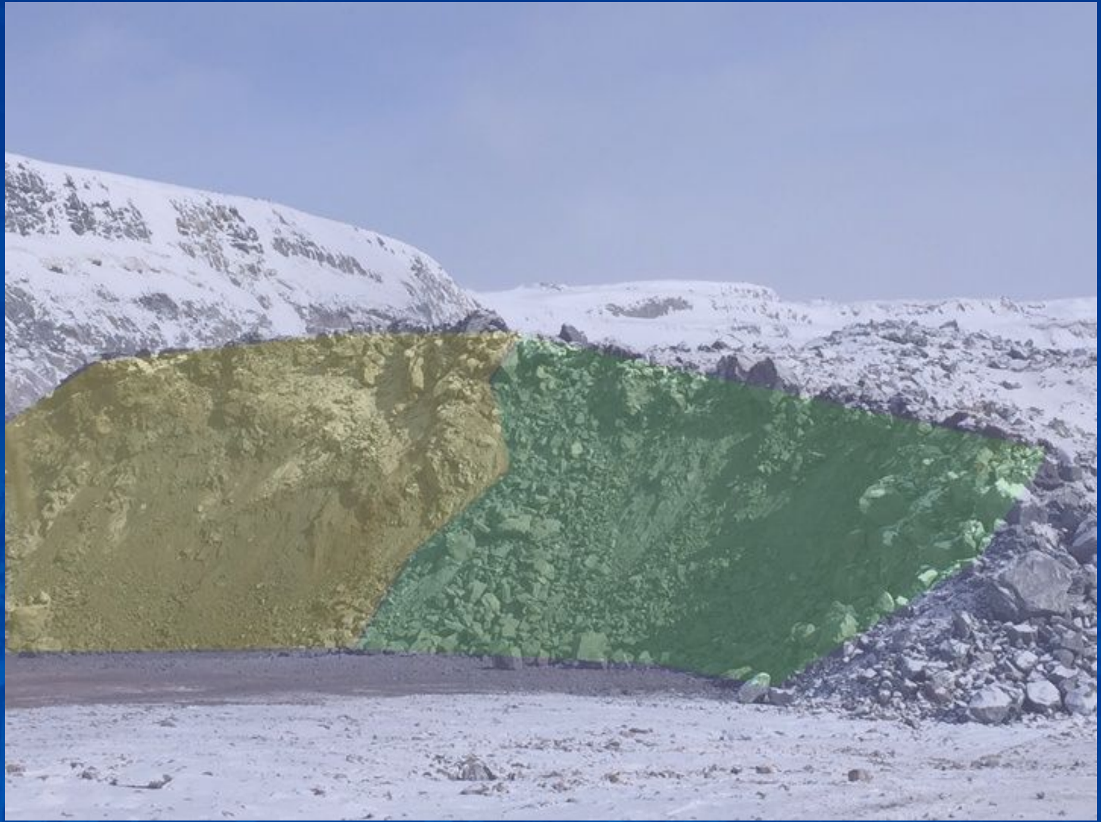


I.R.O.N.N. V1.0

MINERAI DE FER QUÉBEC
QUEBEC IRON ORE

UBC
Master of Data Science

4. Mask is superimposed over the original image

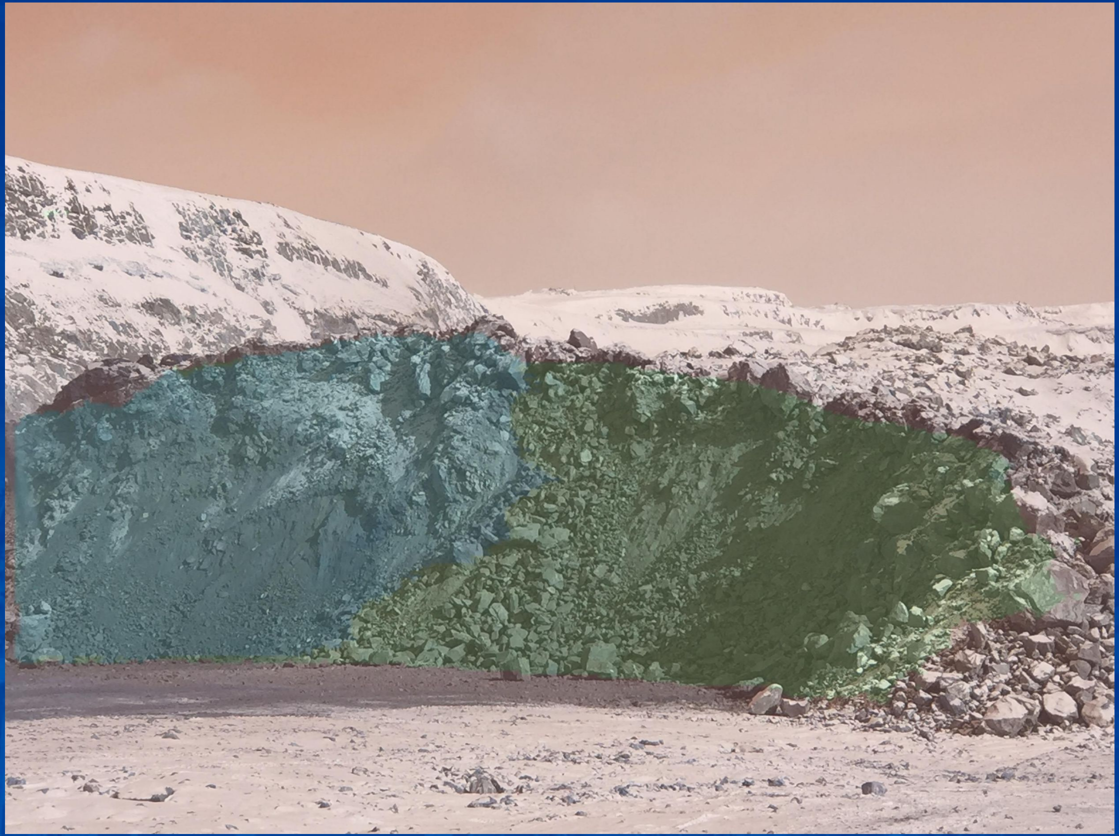


I.R.O.N.N. V1.0

MINERAI DE FER QUÉBEC
QUEBEC IRON ORE

UBC
Master of Data Science

5. Prediction after training



I.R.O.N.N. V1.0

MINERAI DE FER QUÉBEC
QUEBEC IRON ORE

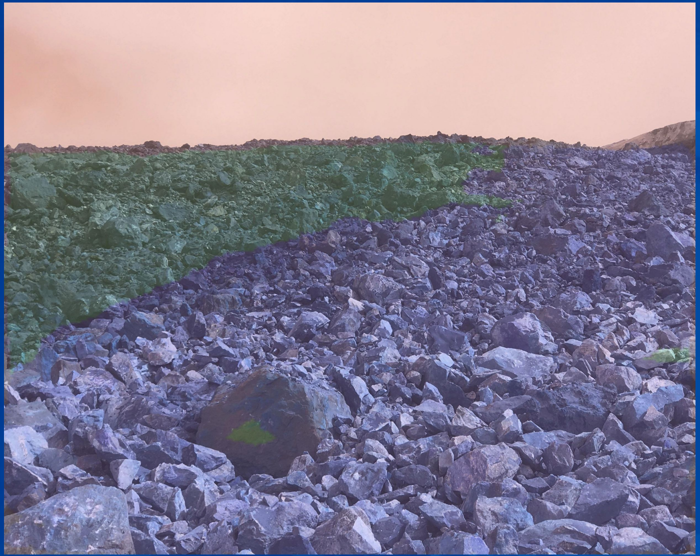
UBC Master of Data Science

An example on an unseen image

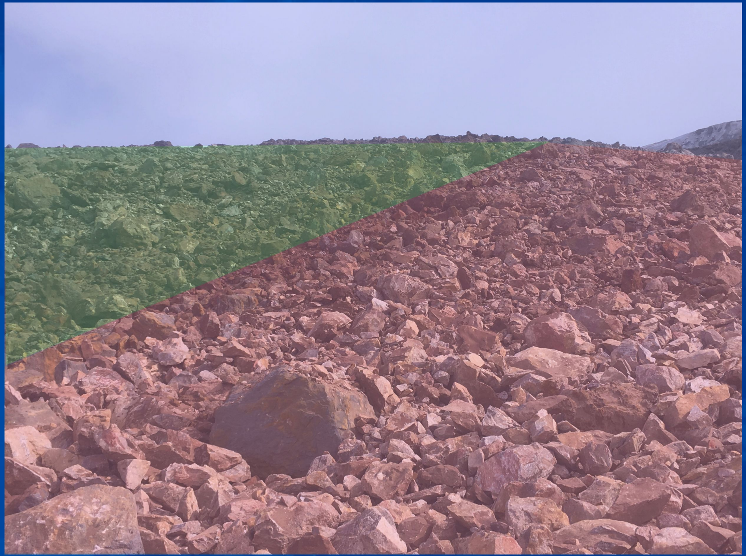
Model Results

I.R.O.N.N. V1.0: pixel-wise accuracy of **77%**

I.R.O.N.N. V2.0 (Outliers removed): pixel-wise accuracy of **85%**



Prediction



Actual superimposed label

Cost of Training on AWS:

- Cost of using the GPU ~ $1 per hour of training (active use time)
- Time to train (using 600 images): 2.5 - 3 hours
- Cost of GPU + Cost of storage ~ $4 + $1 = $5
- Expected training frequency = 2 times a month
- Total a month: Approx $15 (with 650-700 images)

Difficulties & Challenges

MINERAI DE FER QUÉBEC
QUEBEC IRON ORE

UBC
Master of Data Science

- Some images are unsuitable for training (bad lighting, angles, and background noise)
- May not have enough training data due to a large number of rock types
- Model's current performance might not be the same when we introduce more classes
- Model might not be able to handle fine-grained differences from textures between rock types

Next Steps

Pipeline has been created

- Train a model with all the different rock types
- Improve model’s performance

Train mine operators to capture images in a standardized format

Develop a user friendly interface that can be used by the operators and predict in real-time

Summary of Pipeline

MINERAI DE FER QUÉBEC
QUEBEC IRON ORE

UBC
Master
of Data
Science

MINERAI DE FER QUÉBEC
QUEBEC IRON ORE

UBC
Master of Data Science

Merci beaucoup!