MINERAI DE FER QUÉBEC
QUEBEC IRON ORE

UBC Master of Data Science

QIO Geo-Detection Project

Implementing a CNN to segment rock types

QIO Partner: Veronique Mulholland
MDS Mentor: Simon Goring
Contributors: Jingyun Chen, Socorro Dominguez, Milos Milic, Aditya Sharma

I.R.O.N.N.

Agenda

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

Introduction

Mining company based in Québec that specializes in iron ore extraction and processing

Around 500 specialists in various fields of iron ore extraction

Collect 100GB of production data per day

Want to employ machine learning to:

- increase efficiency (6M to 7M tonnes of iron)
- leverage stored data to reduce operational costs

Business Problem

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

QIO needs a prototype (proof of concept) that will help:

- Segmenting rock types from an active mining face image

The Data

MINERAI DE FER QUÉBEC
QUEBEC IRON ORE

UBC
Master of Data Science

Data Science Objectives/Tools

MINERAI DE FER QUÉBEC
QUEBEC IRON ORE

UBC Master of Data Science

Labeling the data:

- LabelMe

Understanding the data:

- Colour and camera feature analysis

Building the machine learning model:

- Image segmentation and classification

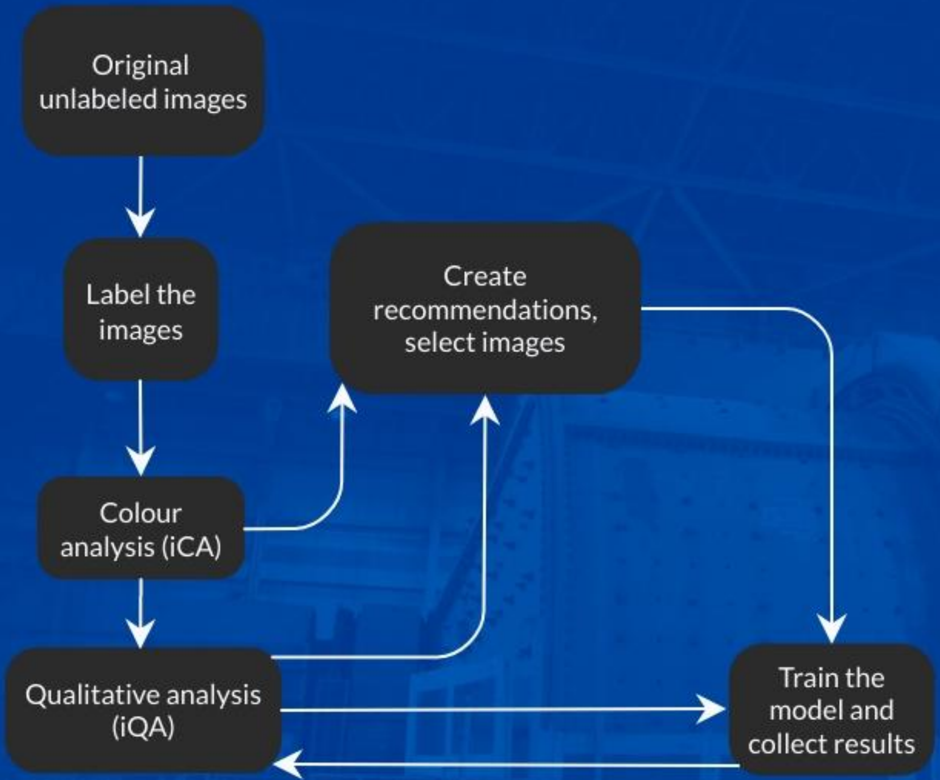
Improving the model:

- Design tools to improve data quality

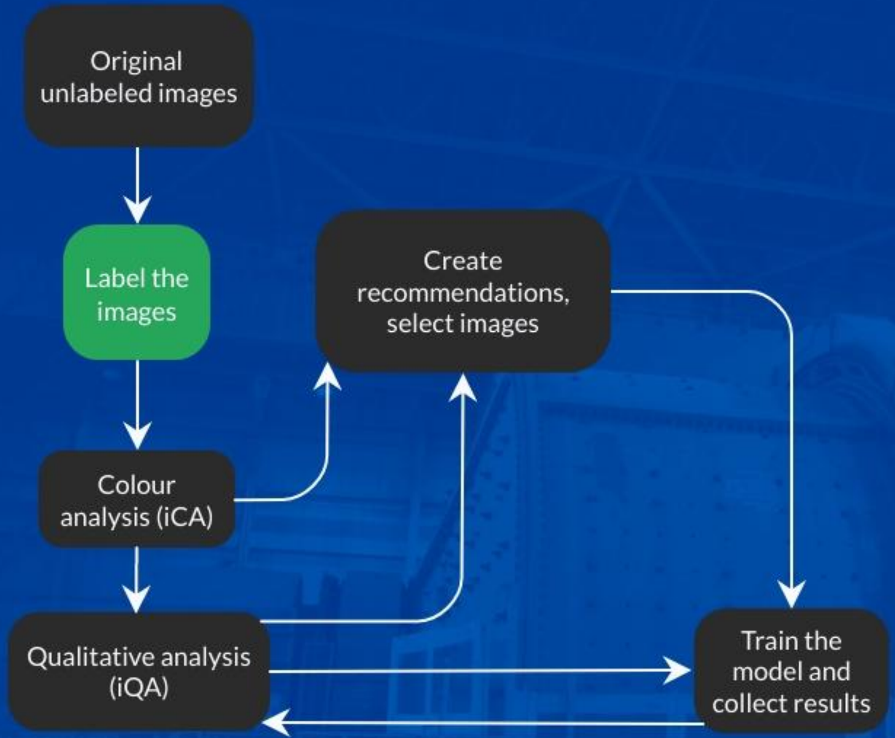
Data Pipeline

MINERAI DE FER QUÉBEC
QUEBEC IRON ORE

UBC Master of Data Science



Data Pipeline



Labeling the Data: Origin State

Hand annotations

- Cannot be used to train a machine learning model

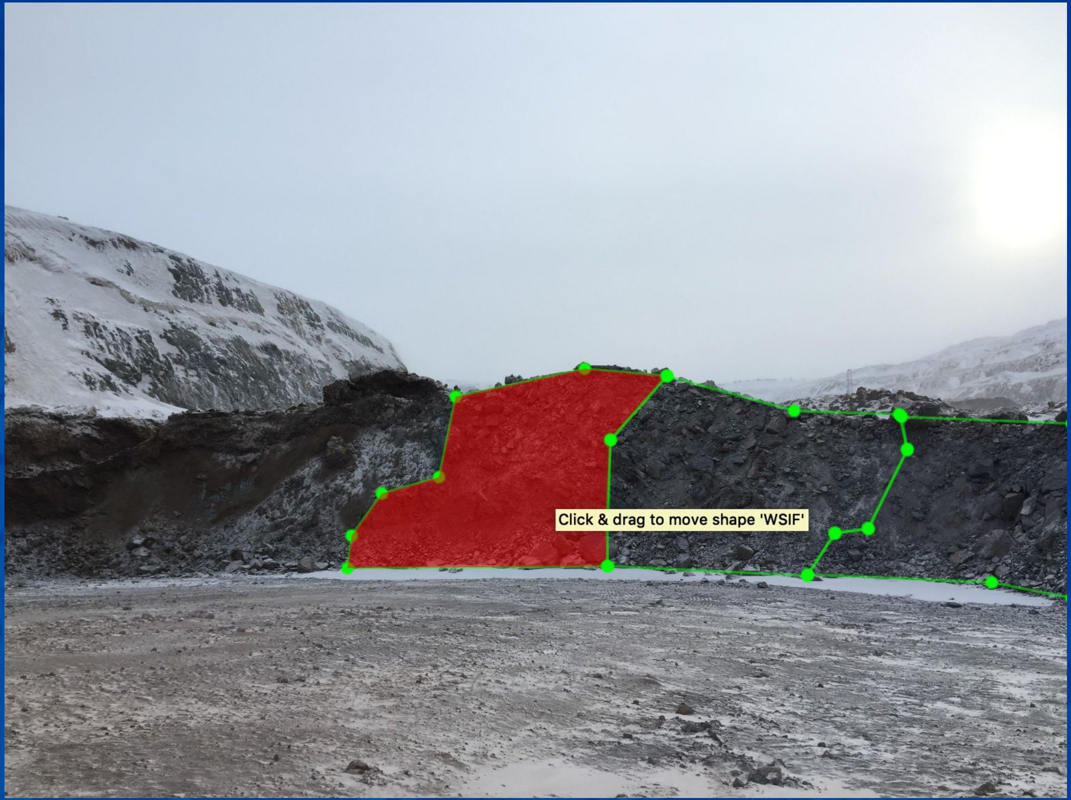
Labeling the Data: Solution

MINERAI DE FER QUÉBEC
QUEBEC IRON ORE

UBC
Master
of Data
Science

LabelMe

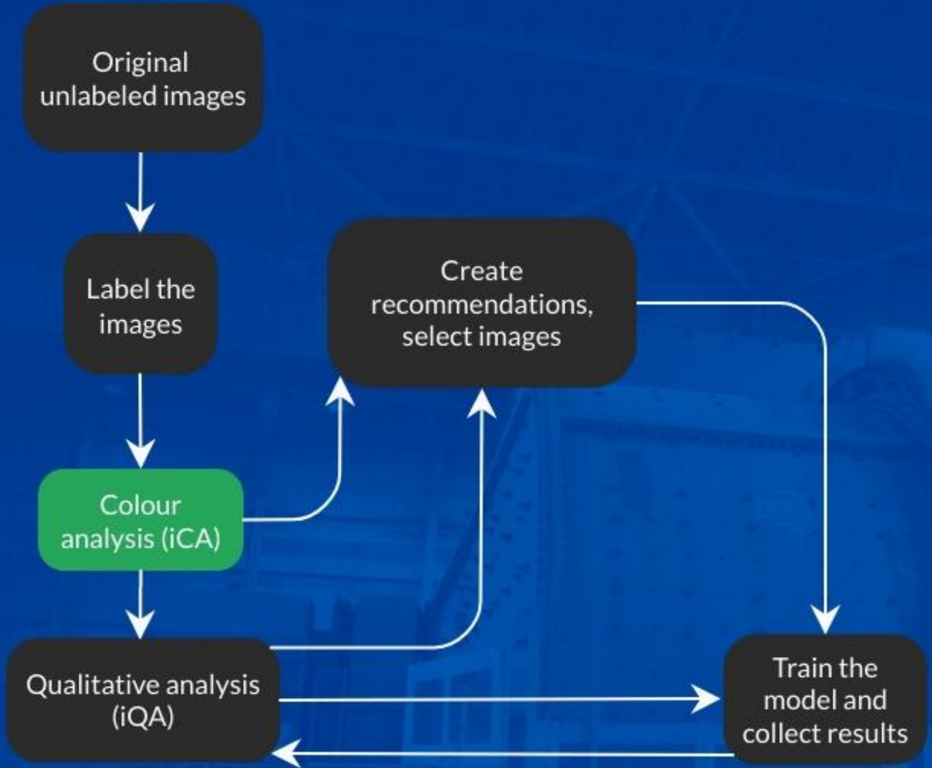
- Obtain coordinates that are useful for EDA and model training



Data Pipeline

MINERAI DE FER QUÉBEC
QUEBEC IRON ORE

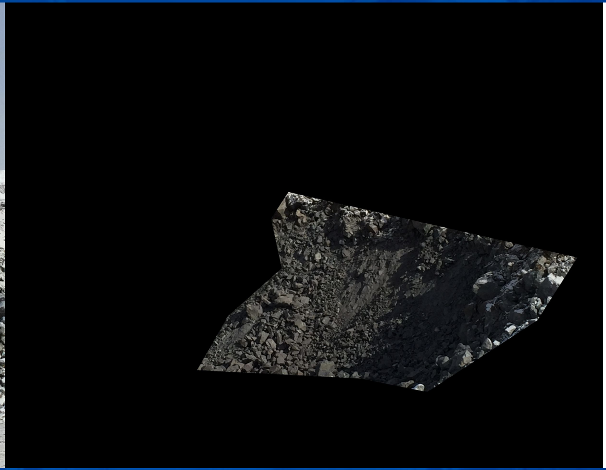
UBC
Master of Data Science

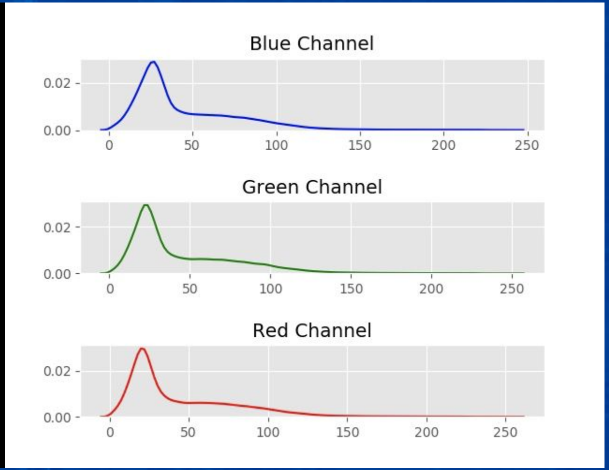


Understanding the Data: Colour Features

Purpose: intuitive way to understand image data

Method: extracted 9 features (skewness, kurtosis, mean pixel value on 3 channels)

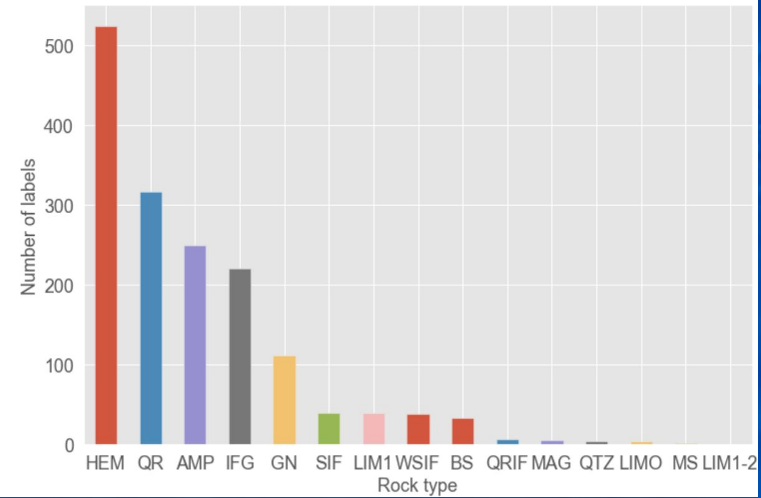


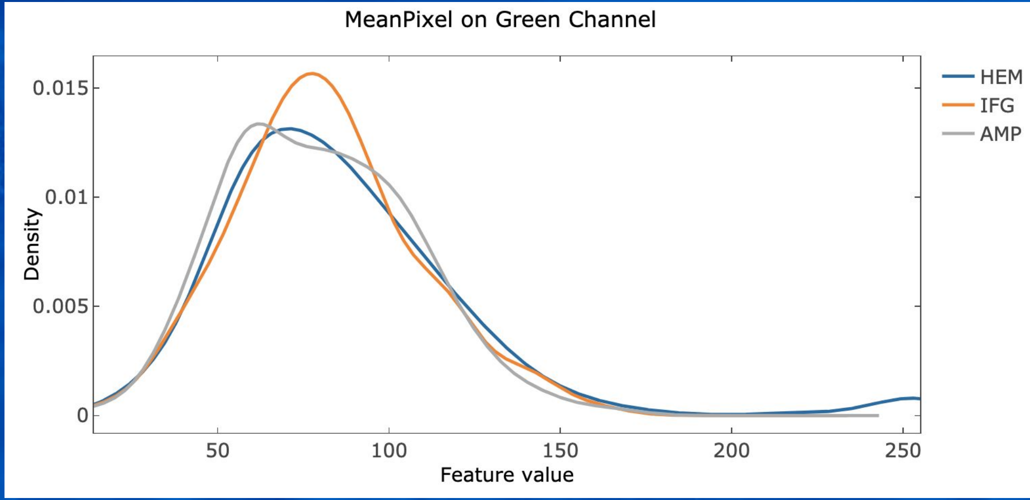


Understanding the Data: Colour Features

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

Problem: hard to distinguish due to imbalanced data

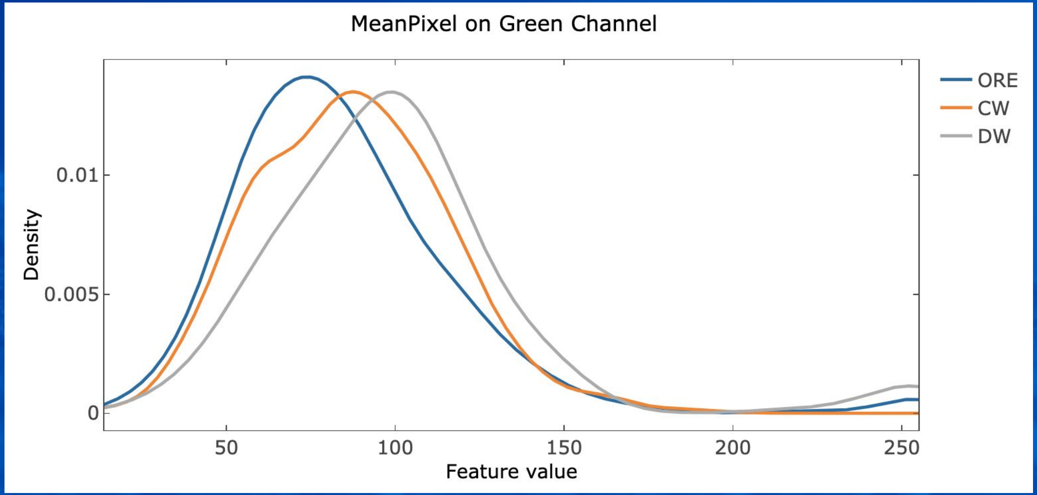




Understanding the Data: Colour Features

Combine rocks into 3 categories:

- Ore
- Dilution waste
- Contamination waste



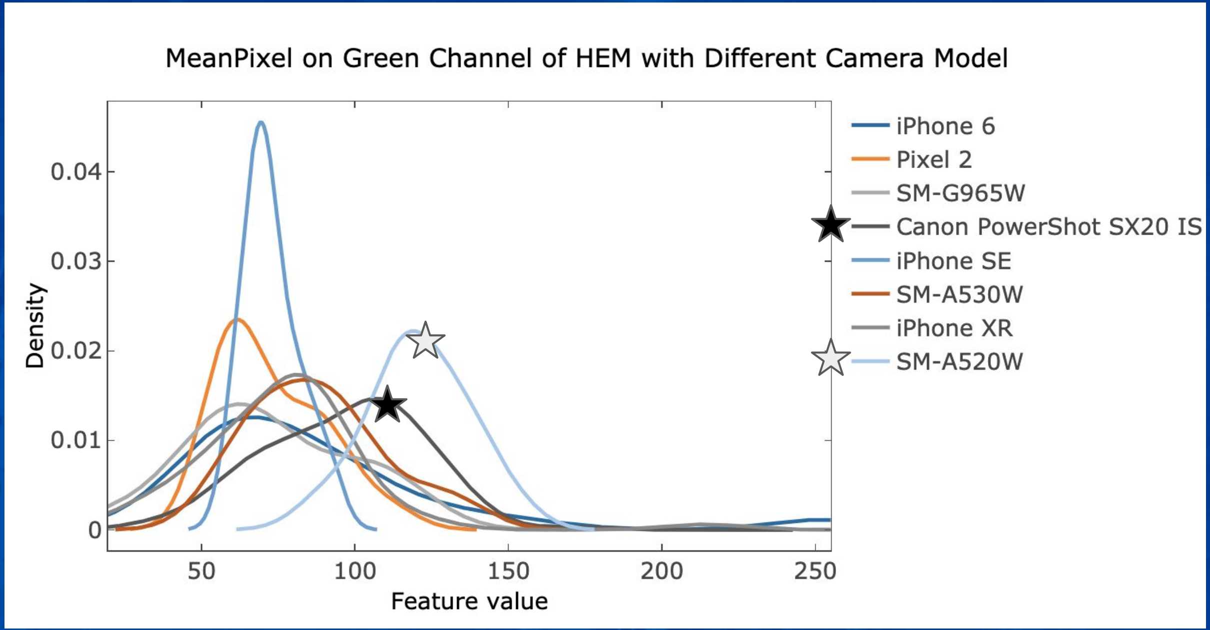
Understanding the Data: Camera Features

Purpose:

- Check camera effect

Method:

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



Camera Features: Impact

MINERAI DE FER QUÉBEC
QUEBEC IRON ORE

UBC
Master of Data Science

Images on the right:

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

Image	Camera	MeanPixelRed
	SM-A520W	148.13
	iPhone 6	70.53

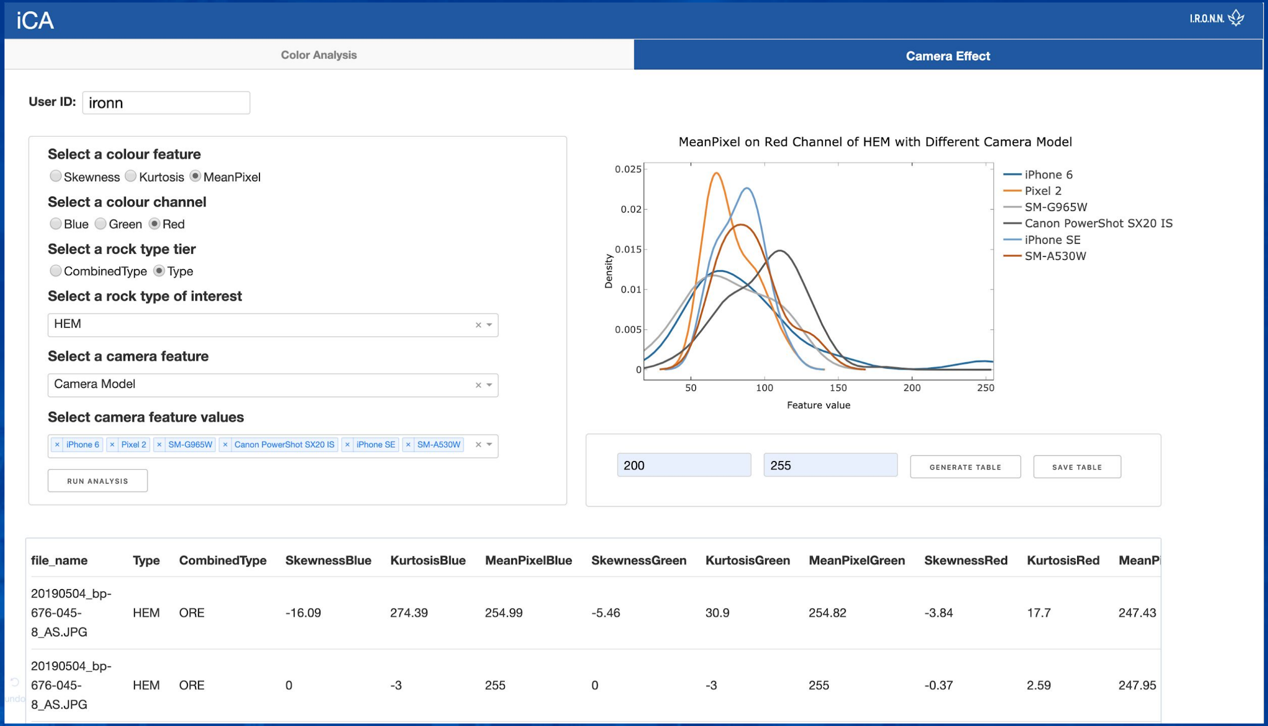
Data Product: iCA

MINERAI DE FER QUÉBEC
QUEBEC IRON ORE

UBC Master of Data Science

Plotly Dashboard

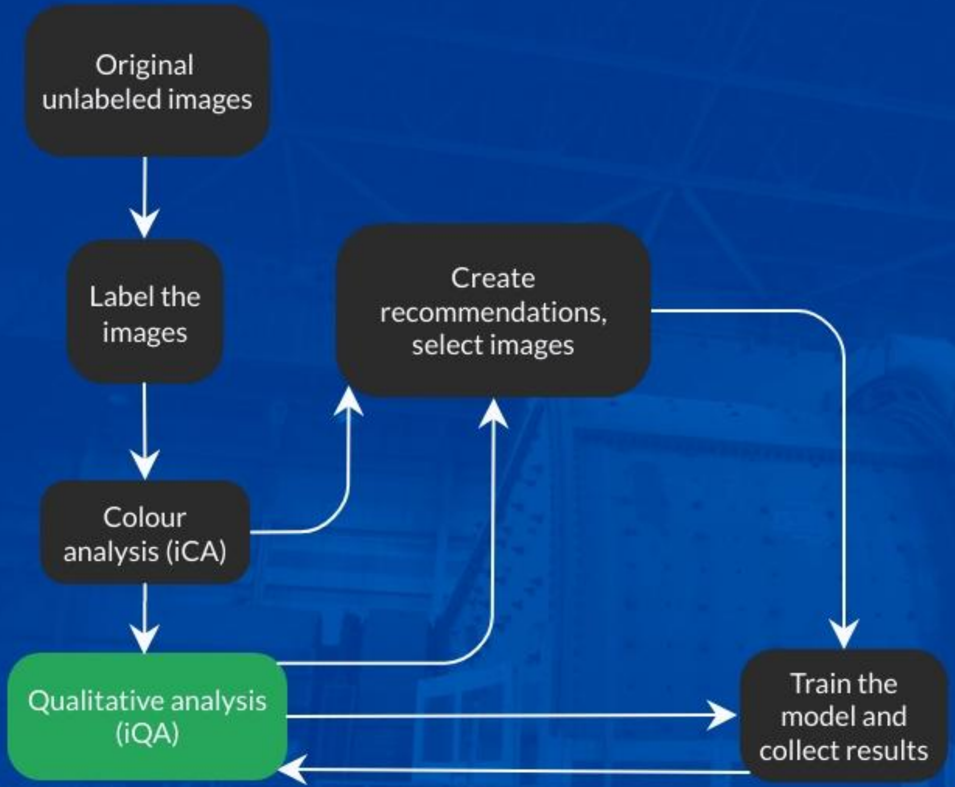
- Interactively identify images that have outstanding features (outliers)
- Can be used for image quality check with iQA



Data Pipeline

MINERAI DE FER QUÉBEC
QUEBEC IRON ORE

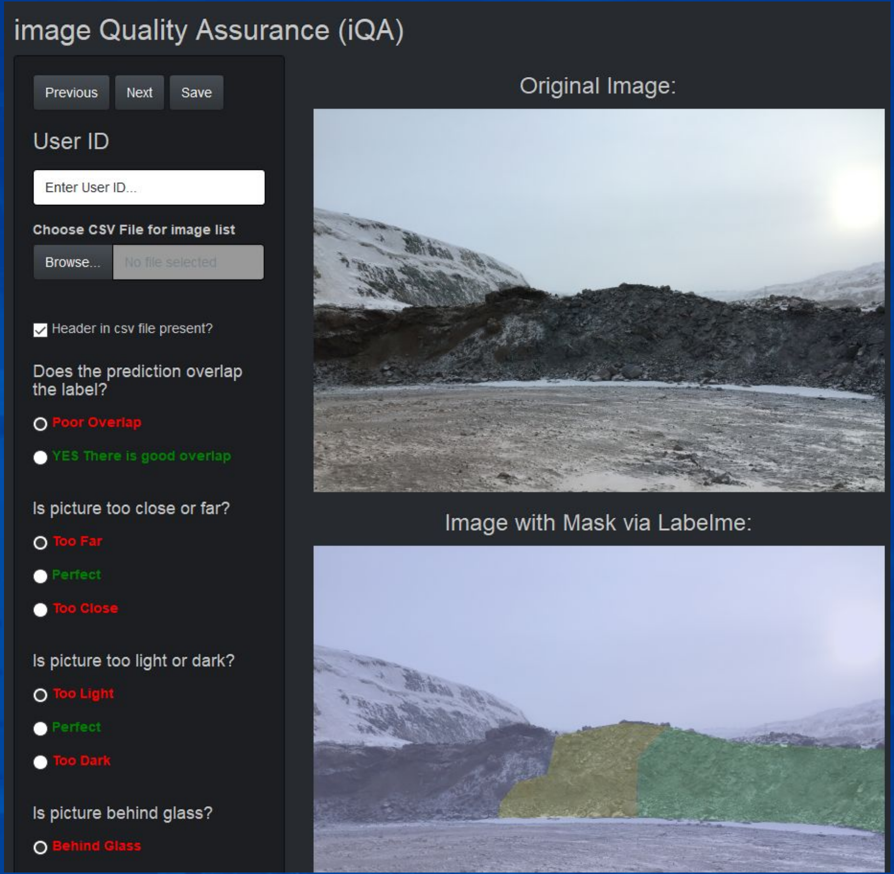
UBC Master of Data Science



Data Product: iQA

Shiny Dashboard

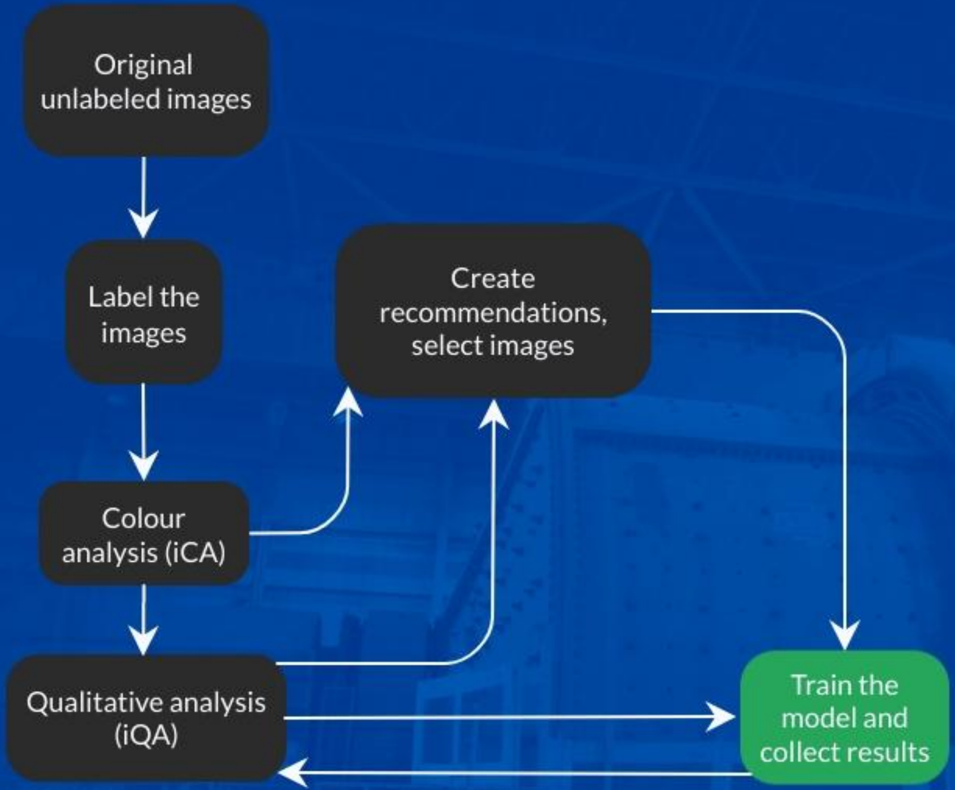
- Check if there are any qualitative issues with original images, labels, predictions (blurriness, machinery present, etc)



Data Pipeline

MINERAI DE FER QUÉBEC
QUEBEC IRON ORE

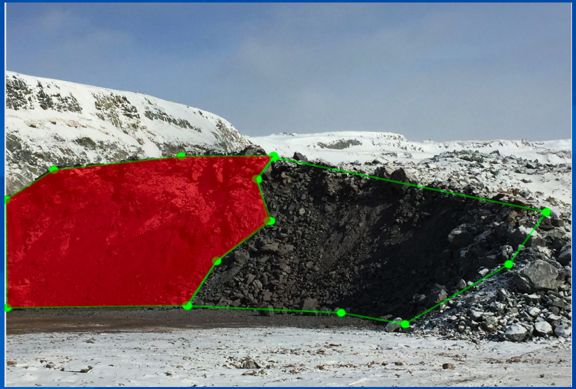
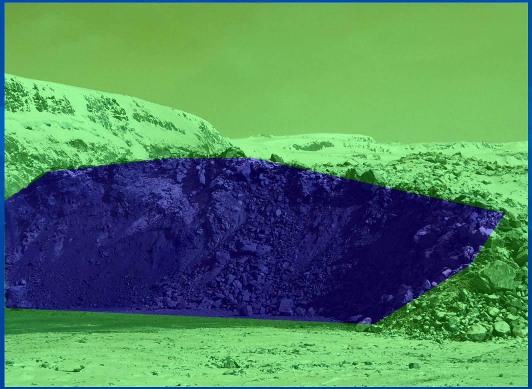
UBC
Master of Data Science



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 from polygons as blasted faces
- Trained on ~430 images

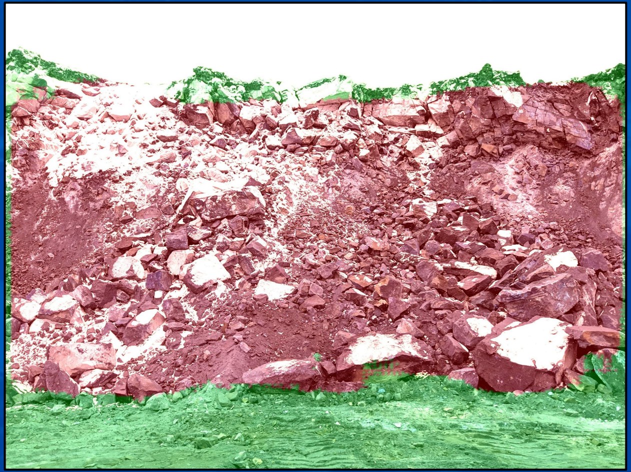




I.R.O.N.N. v0.1: Result

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

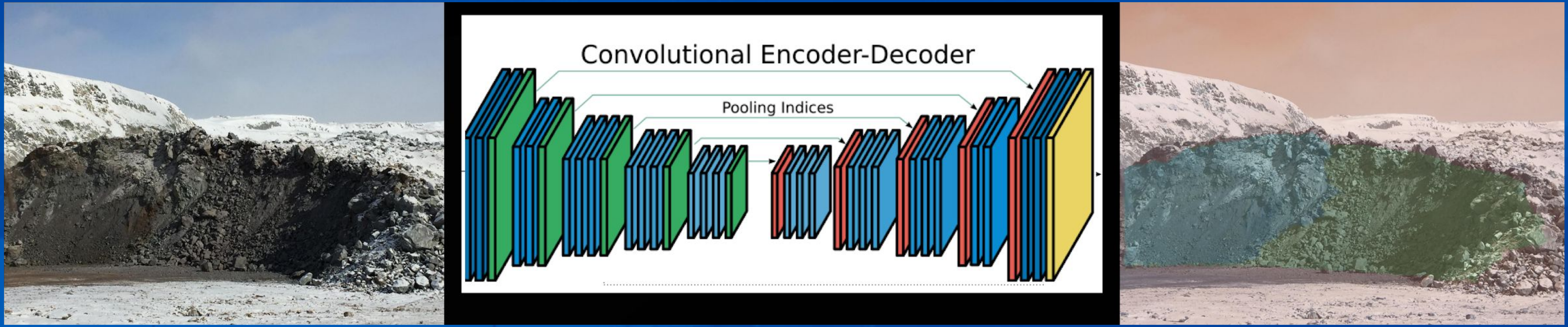
UBC
Master of Data Science

I.R.O.N.N. Architecture

Currently using Fully Convolutional Networks for Image Segmentation

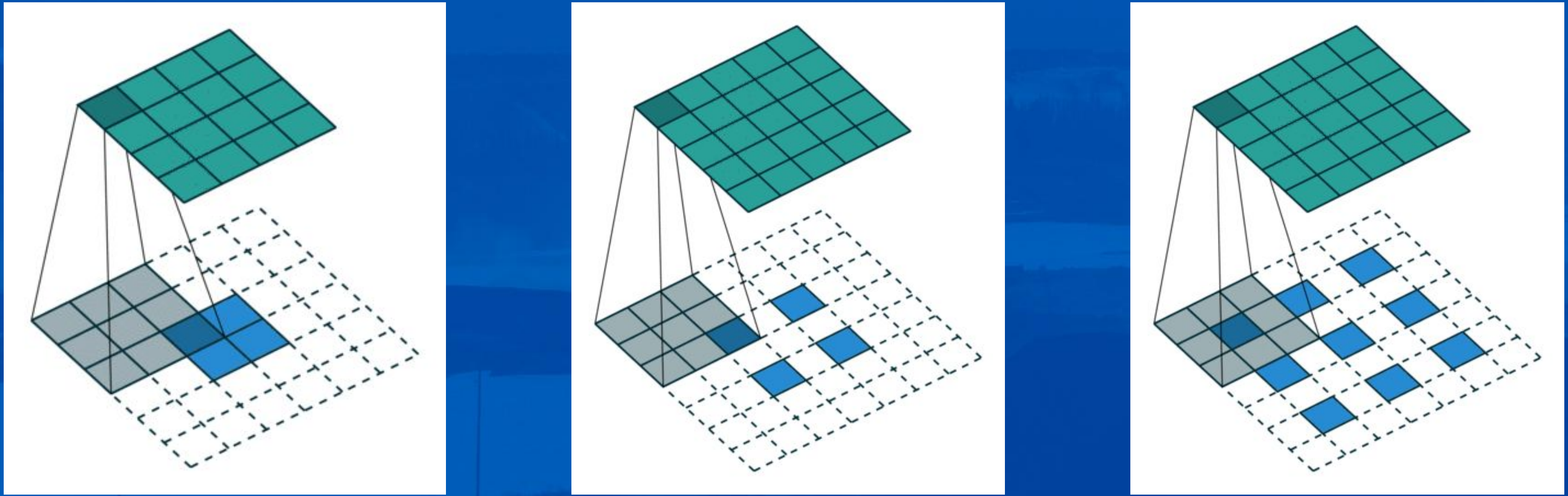
Takes in images and returns images as predictions

Framework: PyTorch



How does Upsampling work?

- Used when we want to generate images from neural networks
- Creates high resolution images from low resolution images
- Blue: Input, Cyan: Output



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

MINERAI DE FER QUÉBEC
QUEBEC IRON ORE

UBC
Master of Data Science

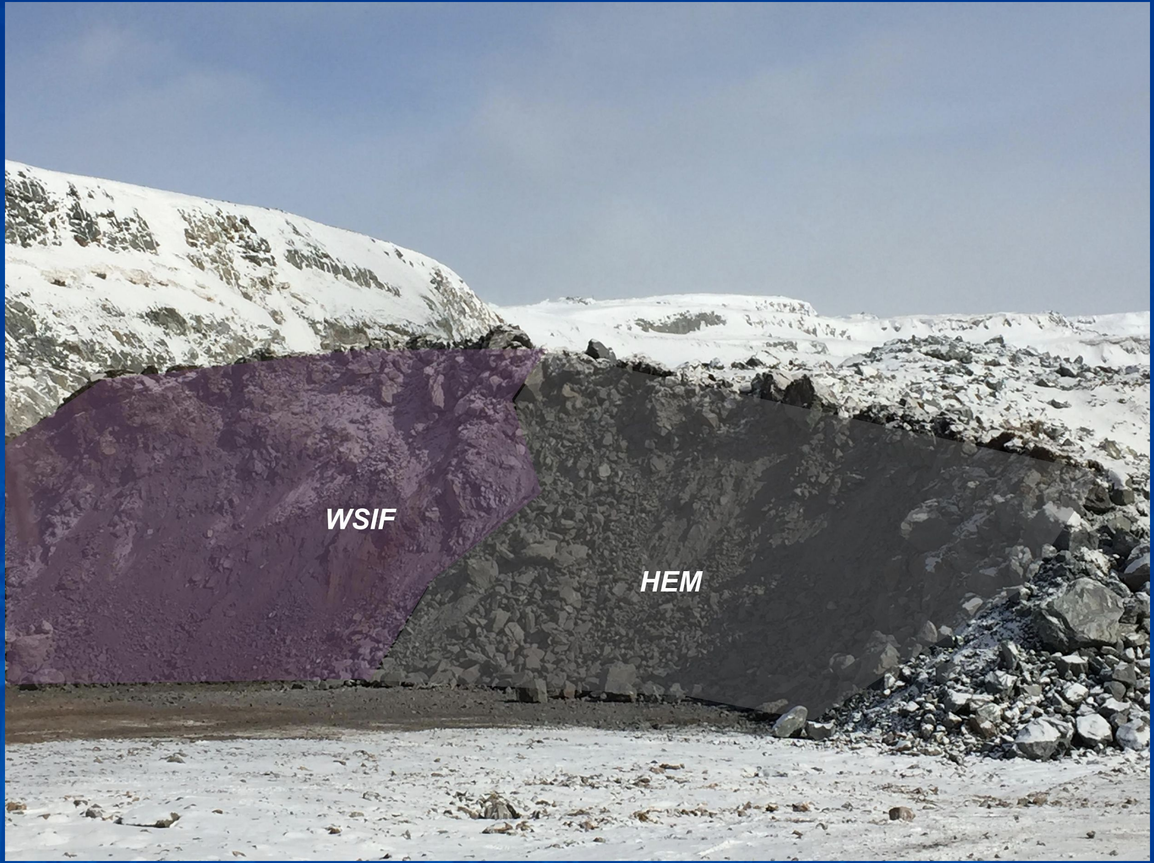
1. Original unlabeled image

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

2. Labeled image

HEM: Ore

WSIF: Contamination waste

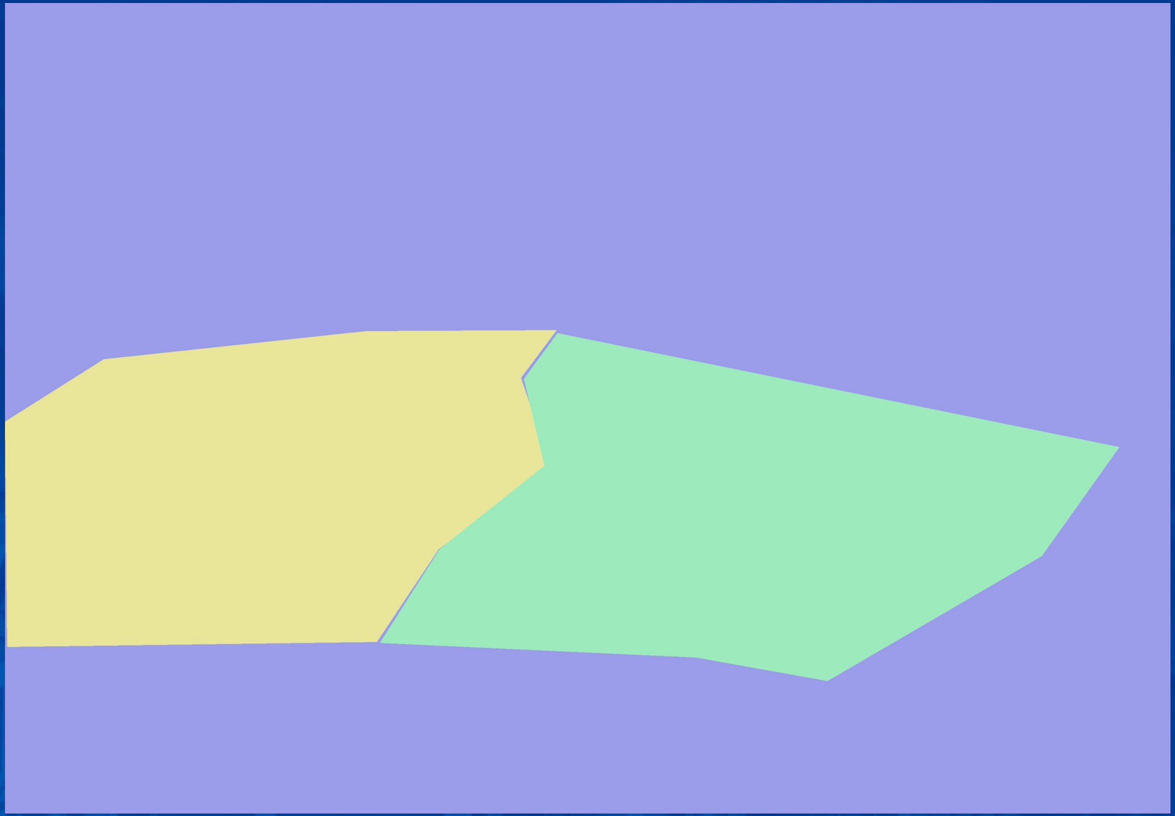


I.R.O.N.N. V1.0

MINERAI DE FER QUÉBEC
QUEBEC IRON ORE

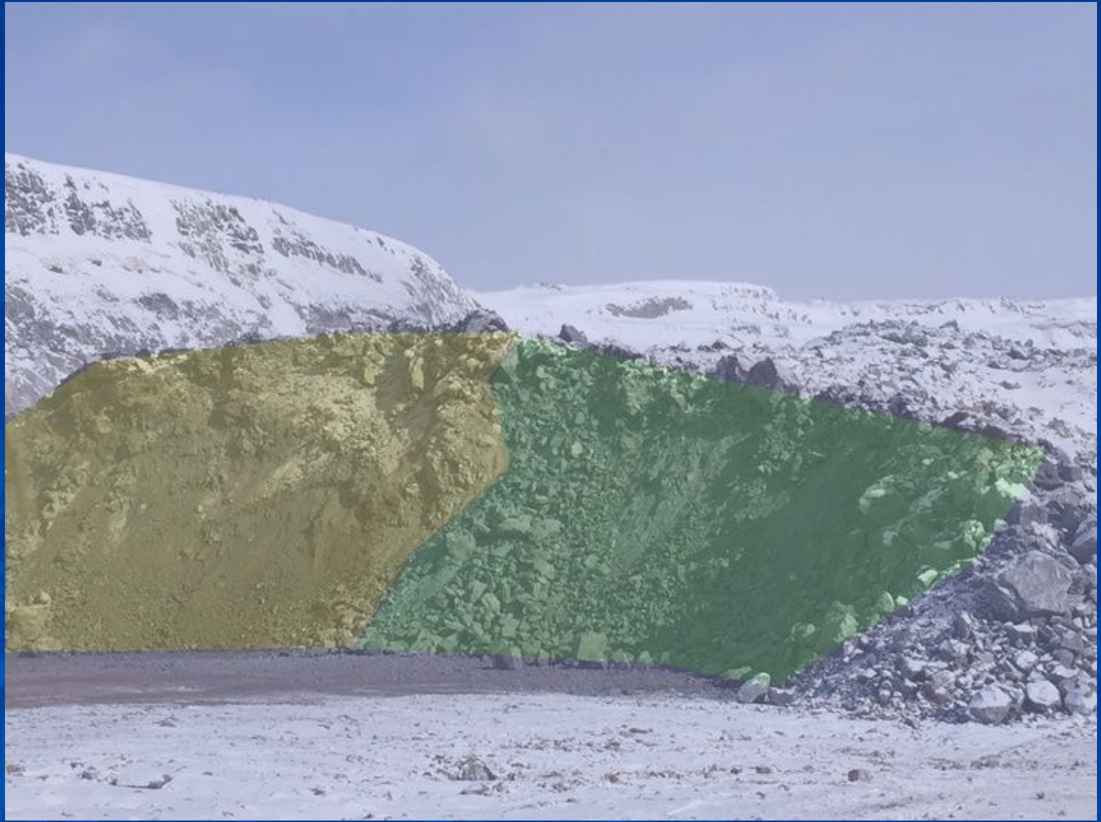
UBC
Master of Data Science

3. Created mask from the label



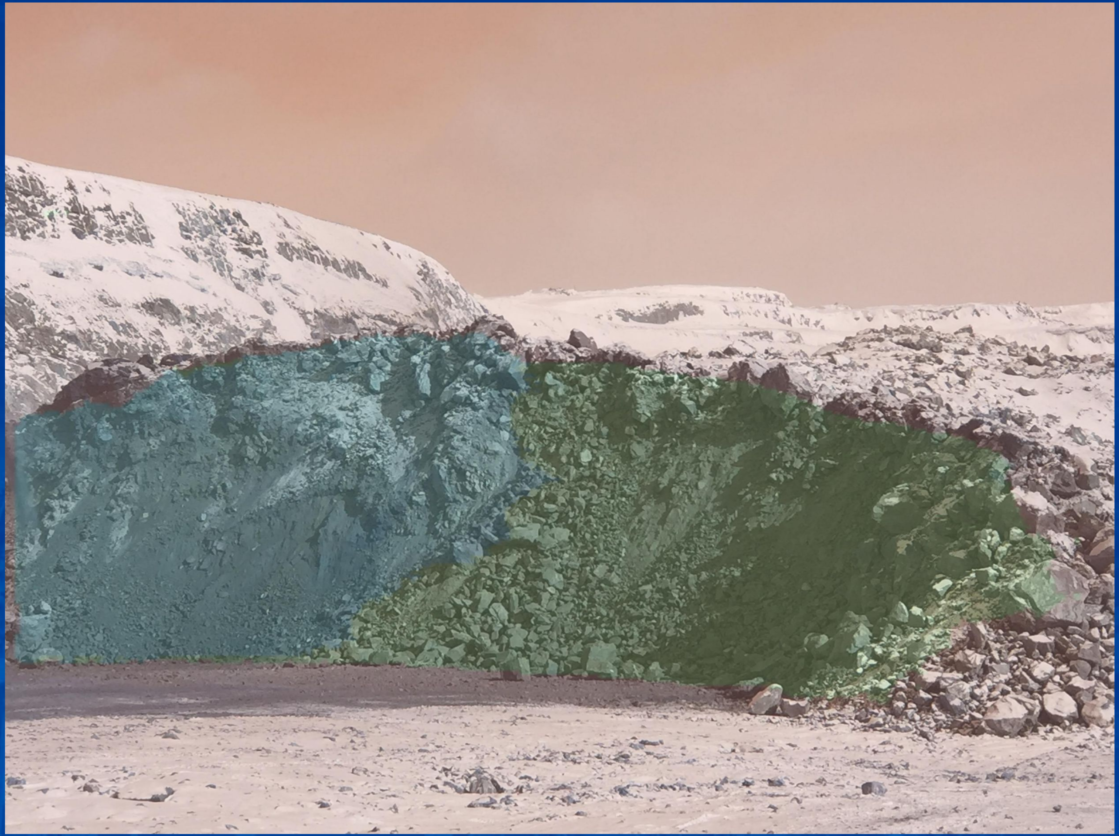
I.R.O.N.N. V1.0

4. Superimposed mask over the original image for sanity check



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

5. Prediction after training



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

MINERAI DE FER QUÉBEC
QUEBEC IRON ORE

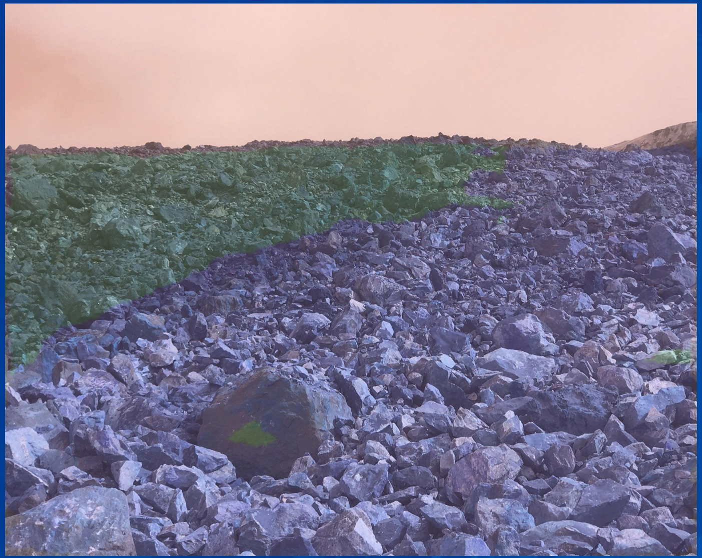
UBC
Master of Data Science

An example from validation set

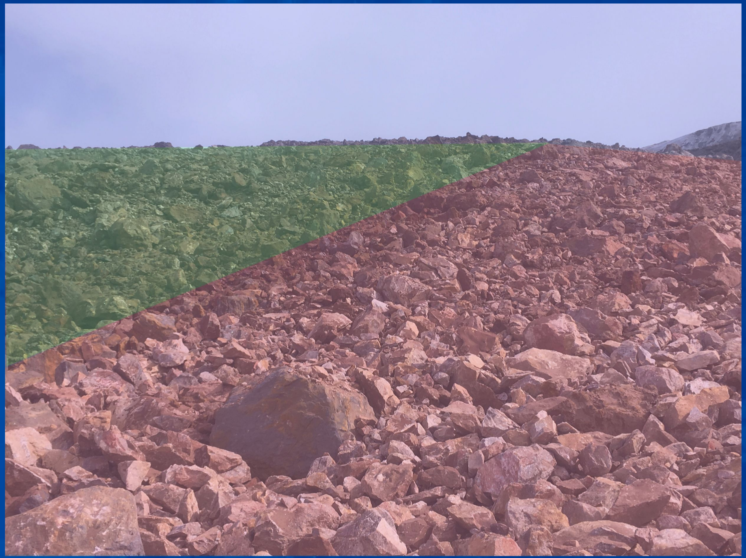
Model Results

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%**

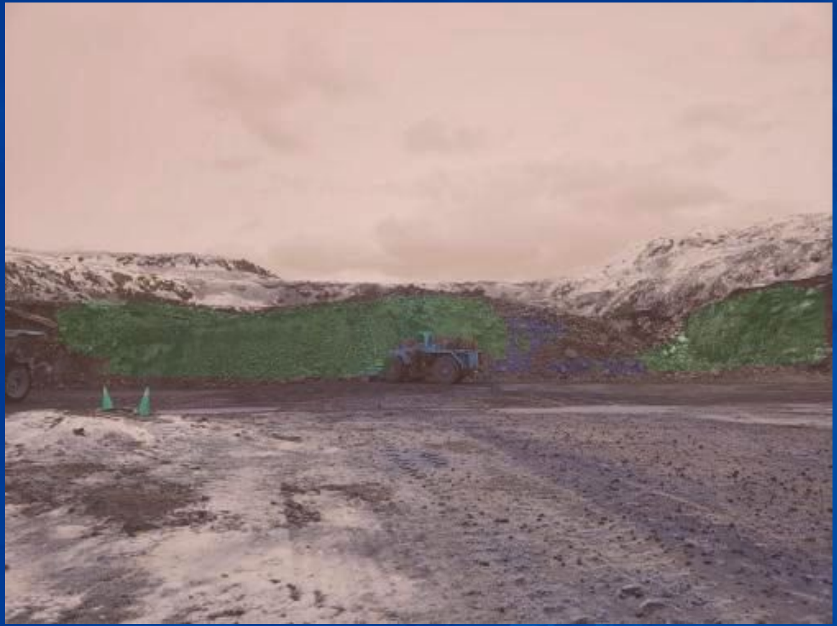


Prediction



Actual superimposed label

A bad example



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

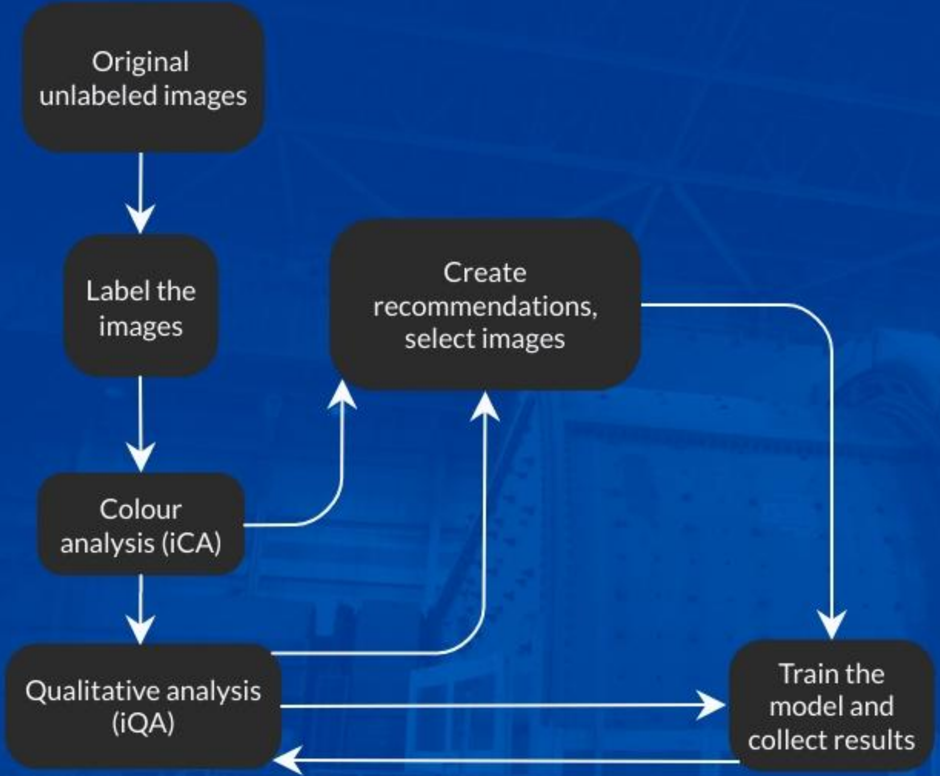
- Some images are unsuitable for training (bad lighting, angles, background noise and low resolution)
- Insufficient and imbalanced training data for a large number of rock types
- Missing EXIF data
- Limited online resources for plotly dash

Future Work

MINERAI DE FER QUÉBEC
QUEBEC IRON ORE
UBC Master of Data Science

- Collect more data
- Train a model with all the different rock types
- Assess and improve model performance
  - Data Augmentation
  - New and complex model architectures
- Capture images in a standardized format
- Develop an app for getting predictions in real-time

Summary of Pipeline



Any Questions?