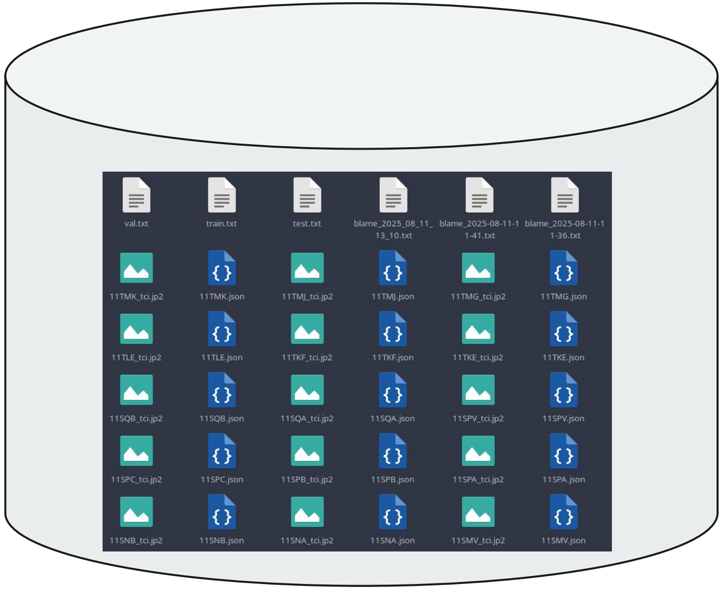
Week 10

Mining Asset Detection (MAD)

Dataset management

A dataset consists of:

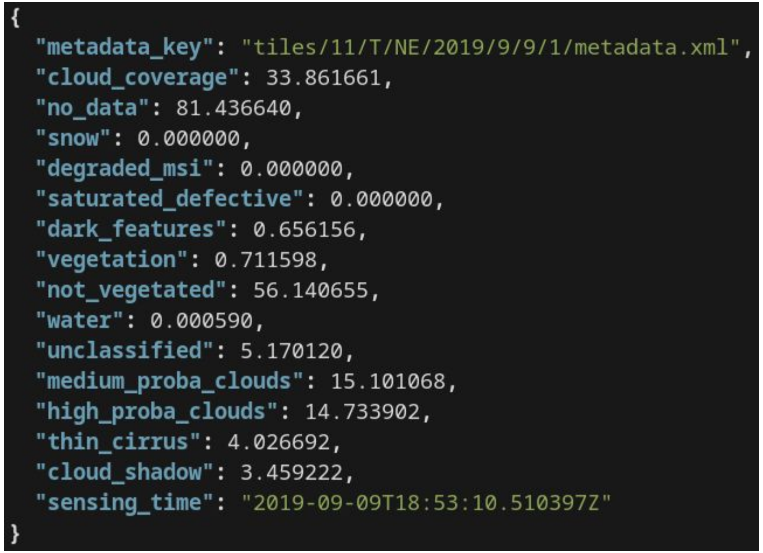
- The respective .jp2 images
- Their corresponding AWS metadata
- Information about the used split
- Blame files



Dataset management - Metadata

Metadata:

- Contains information about all the quality metrics
- Used to retrieve more images/ better quality images for a tile

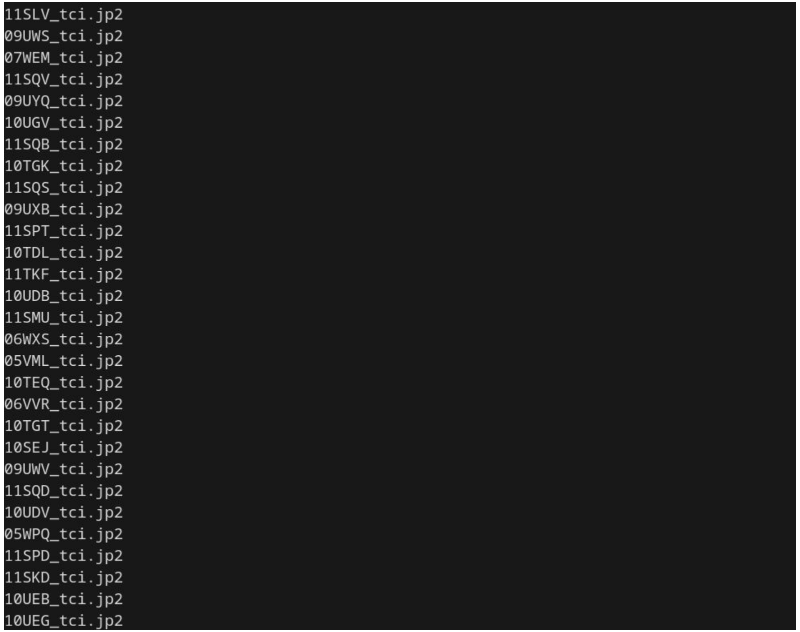


```
{
  "metadata_key": "tiles/11/T/NE/2019/9/9/1/metadata.xml",
  "cloud_coverage": 33.861661,
  "no_data": 81.436640,
  "snow": 0.000000,
  "degraded_msi": 0.000000,
  "saturated_defective": 0.000000,
  "dark_features": 0.656156,
  "vegetation": 0.711598,
  "not_vegetated": 56.140655,
  "water": 0.000590,
  "unclassified": 5.170120,
  "medium_proba_clouds": 15.101068,
  "high_proba_clouds": 14.733902,
  "thin_cirrus": 4.026692,
  "cloud_shadow": 3.459222,
  "sensing_time": "2019-09-09T18:53:10.510397Z"
}
```

Dataset management - Split

Split files:

- Simple index for which files to use for which split

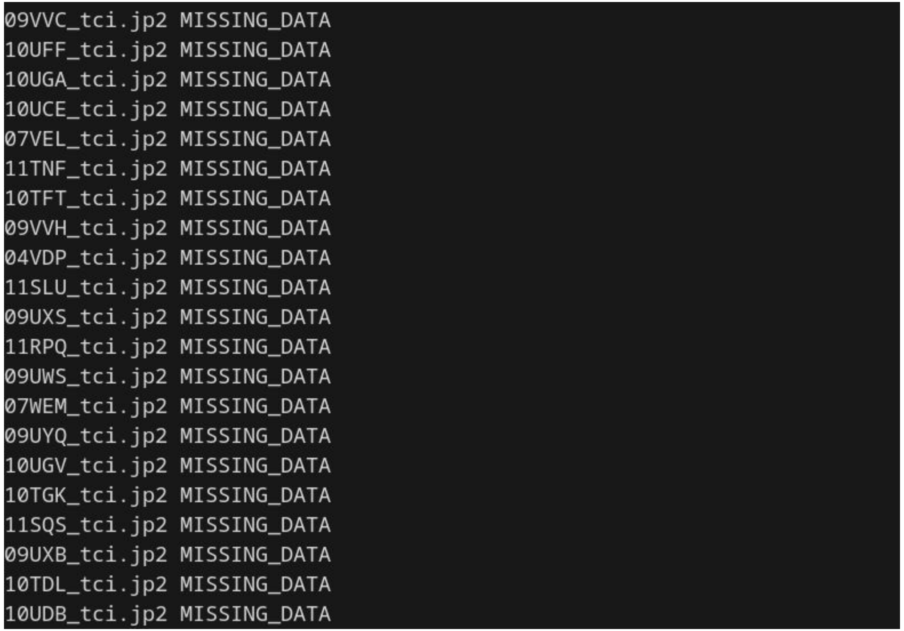


```
11SLV_tci.jp2
09UWS_tci.jp2
07WEM_tci.jp2
11SQV_tci.jp2
09UYQ_tci.jp2
10UGV_tci.jp2
11SQB_tci.jp2
10TGK_tci.jp2
11SQS_tci.jp2
09UXB_tci.jp2
11SPT_tci.jp2
10TDL_tci.jp2
11TKF_tci.jp2
10UDB_tci.jp2
11SMU_tci.jp2
06WXS_tci.jp2
05VML_tci.jp2
10TEQ_tci.jp2
06VVR_tci.jp2
10TGT_tci.jp2
10SEJ_tci.jp2
09UWV_tci.jp2
11SQD_tci.jp2
10UDV_tci.jp2
05WPQ_tci.jp2
11SPD_tci.jp2
11SKD_tci.jp2
10UEB_tci.jp2
10UEG_tci.jp2
```

Dataset management - Blame

Blame files:

- Contain information about what filters failed during yolo set creation
- Using this data we can then lazily resolve the issues in the dataset
- Filters define how to resolve the blame



```
09VVC_tci.jp2 MISSING_DATA
10UFF_tci.jp2 MISSING_DATA
10UGA_tci.jp2 MISSING_DATA
10UCE_tci.jp2 MISSING_DATA
07VEL_tci.jp2 MISSING_DATA
11TNF_tci.jp2 MISSING_DATA
10TFT_tci.jp2 MISSING_DATA
09VVH_tci.jp2 MISSING_DATA
04VDP_tci.jp2 MISSING_DATA
11SLU_tci.jp2 MISSING_DATA
09UXS_tci.jp2 MISSING_DATA
11RPQ_tci.jp2 MISSING_DATA
09UWS_tci.jp2 MISSING_DATA
07WEM_tci.jp2 MISSING_DATA
09UYQ_tci.jp2 MISSING_DATA
10UGV_tci.jp2 MISSING_DATA
10TGK_tci.jp2 MISSING_DATA
11SQS_tci.jp2 MISSING_DATA
09UXB_tci.jp2 MISSING_DATA
10TDL_tci.jp2 MISSING_DATA
10UDB_tci.jp2 MISSING_DATA
```

Dataset management - Resolve example

Missing data:

=> Resolve by finding an image with other no_data field value and mosaic the images

High snow in image:

=> Repoll source image, preferring the one with least snow + least cloud coverage

Over time this should improve the training image quality without making fidelity compromises!

Pipeline in action

Filtered out:

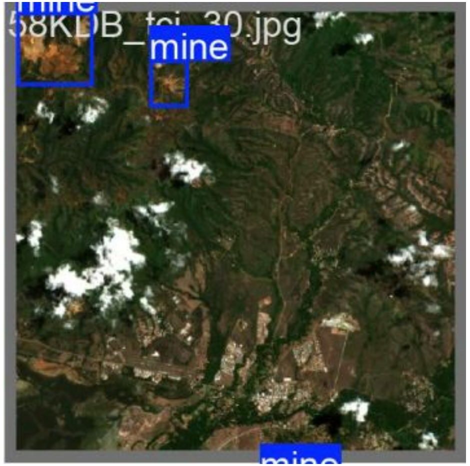
- All windows containing no data (hence the nodata entries on the blame slide)

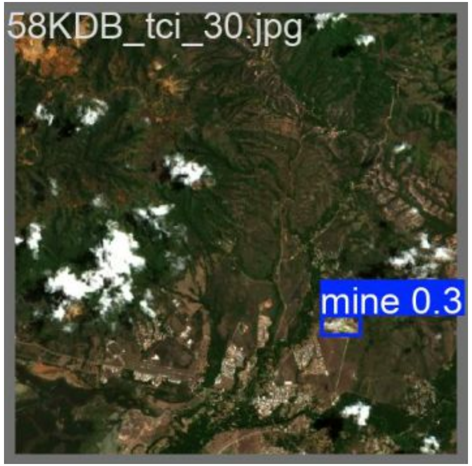
Expansion:

- Simple 10km x 10km grid of the whole images with and without negatives

Duration ~ 2.5 hrs each (40k images)

Pipeline in action - Training run with YOLO



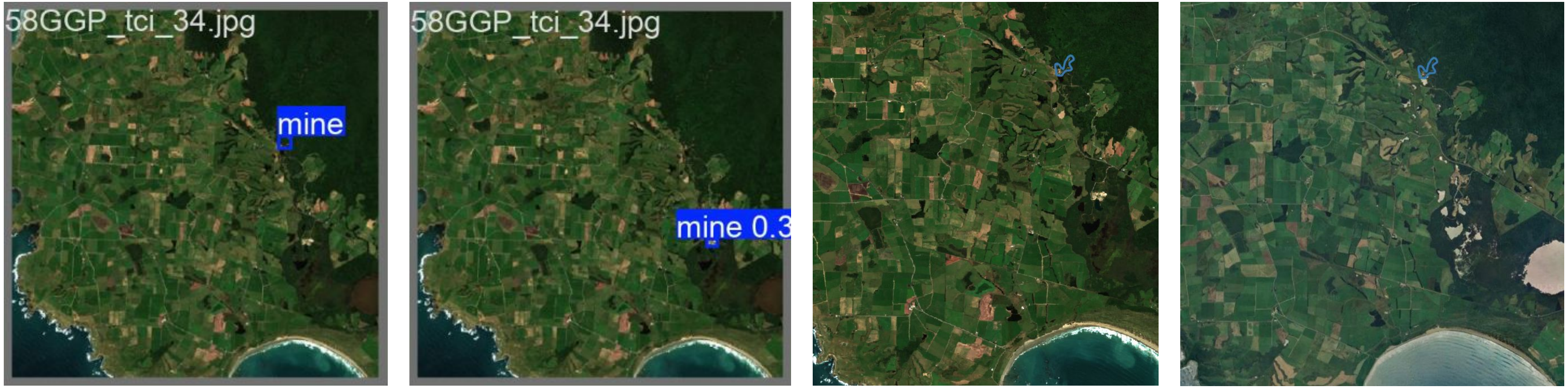


Av. de Koniambo
Rte de la Carrière
Rte de la Carrière

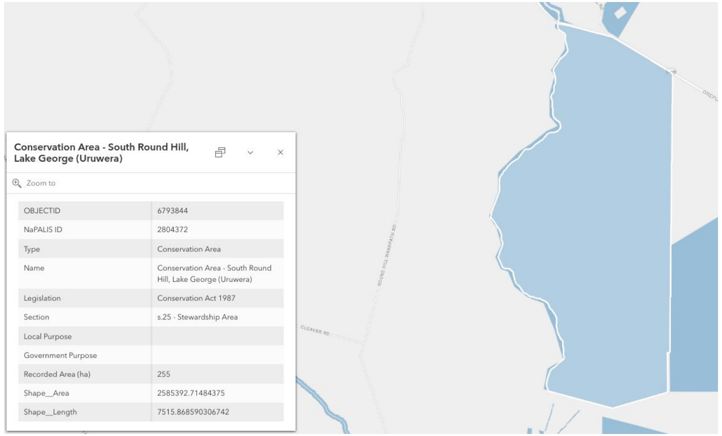
Pipeline in action - Training run with YOLO

Performance results not much different than previous runs…

But this time we can interpret the more thoroughly:

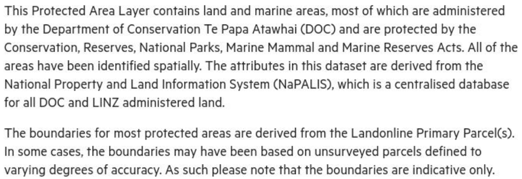


Pipeline in action - Training run with YOLO



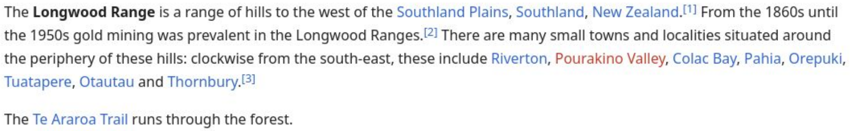
Conservation Area - South Round Hill, Lake George (Uruwera)

OBJECTID	6793844
NaPALIS ID	2804372
Type	Conservation Area
Name	Conservation Area - South Round Hill, Lake George (Uruwera)
Legislation	Conservation Act 1987
Section	s.25 - Stewardship Area
Local Purpose	
Government Purpose	
Recorded Area (ha)	255
Shape__Area	2585392.71484375
Shape__Length	7515.868590306742



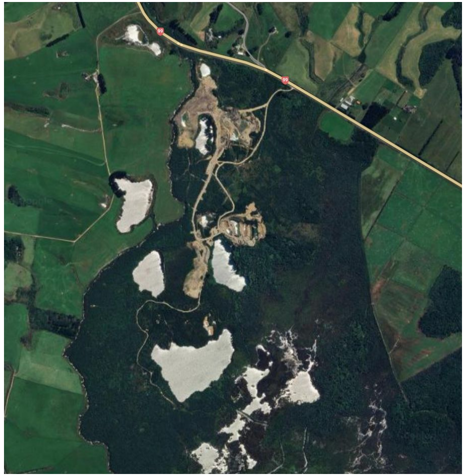
This Protected Area Layer contains land and marine areas, most of which are administered by the Department of Conservation Te Papa Atawhai (DOC) and are protected by the Conservation, Reserves, National Parks, Marine Mammal and Marine Reserves Acts. All of the areas have been identified spatially. The attributes in this dataset are derived from the National Property and Land Information System (NaPALIS), which is a centralised database for all DOC and LINZ administered land.

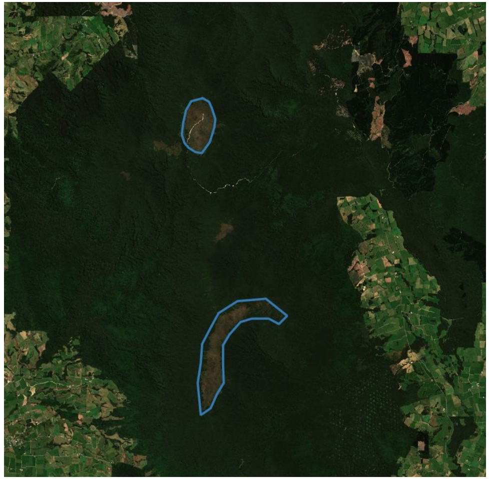
The boundaries for most protected areas are derived from the Landonline Primary Parcel(s). In some cases, the boundaries may have been based on unsurveyed parcels defined to varying degrees of accuracy. As such please note that the boundaries are indicative only.



The **Longwood Range** is a range of hills to the west of the Southland Plains, Southland, New Zealand.[1] From the 1860s until the 1950s gold mining was prevalent in the Longwood Ranges.[2] There are many small towns and localities situated around the periphery of these hills: clockwise from the south-east, these include Riverton, Pourakino Valley, Colac Bay, Pahia, Orepuki, Tuatapere, Otautau and Thornbury.[3]

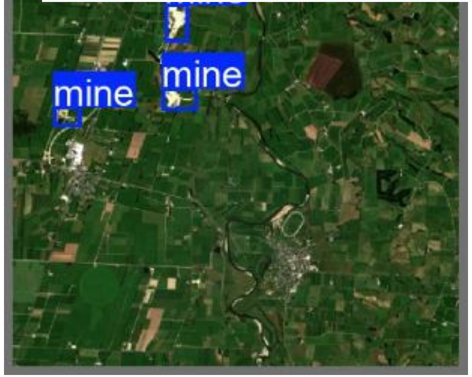
The Te Araroa Trail runs through the forest.

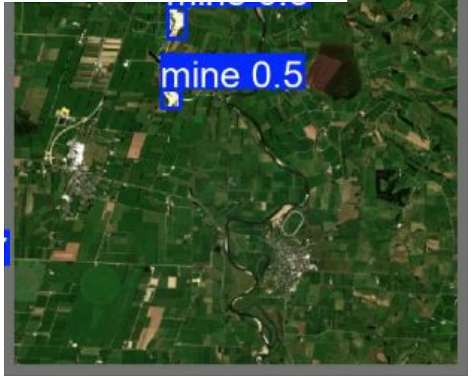


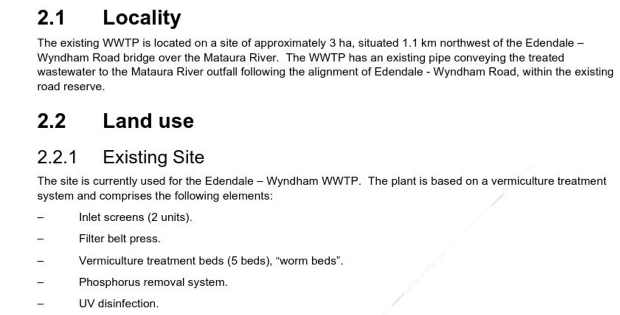


Pipeline in action - Training run with YOLO

Edendale as seen from a distance, the Fonterra dairy factory prominent.







2.1 Locality

The existing WWTP is located on a site of approximately 3 ha, situated 1.1 km northwest of the Edendale – Wyndham Road bridge over the Mataura River. The WWTP has an existing pipe conveying the treated wastewater to the Mataura River outfall following the alignment of Edendale - Wyndham Road, within the existing road reserve.

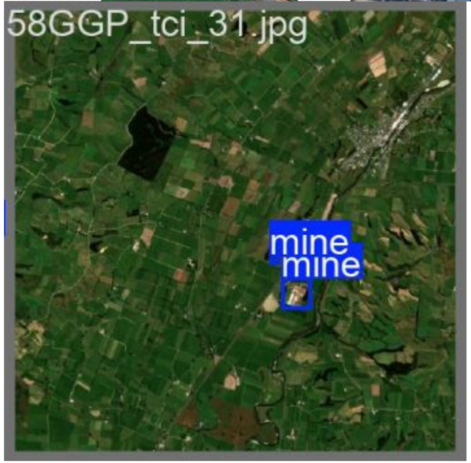
2.2 Land use

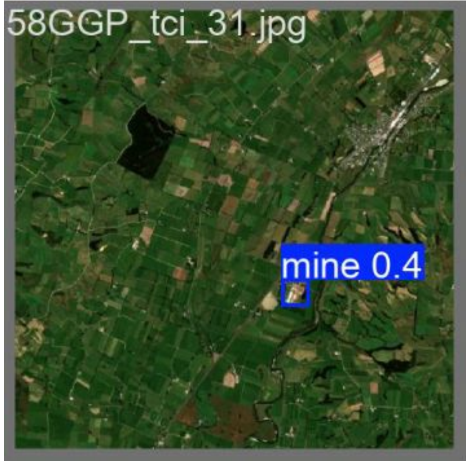
2.2.1 Existing Site

The site is currently used for the Edendale – Wyndham WWTP. The plant is based on a vermiculture treatment system and comprises the following elements:

- Inlet screens (2 units).
- Filter belt press.
- Vermiculture treatment beds (5 beds), "worm beds".
- Phosphorus removal system.
- UV disinfection.

Pipeline in action - Training run with YOLO





Pipeline in action - Training run with YOLO

The model is absolutely not perfect, but…

Maybe the image quality is not really the bottleneck but the image labelling?

- At least hand label the final test set?
- Improve the bounding boxes automatically somehow?
- Hard for the model to learn anything with so many false negatives…

I want to visualize the results of the model

- Output the boxes as .gpkg to overlay in qgis

Pipeline in action - Expansion problems

`gdal / rasterio` really dislike rotating images

- Some ways around it, but maybe not worth the time?

Instead I’d focus on expansion using **translation and different zoom levels**