Intermediate Australian Makeshift Mining Asset Detection

Intermediate Australian Makeshift Mining Asset Detection (I AM MAD)

Idea: Bootstrapping

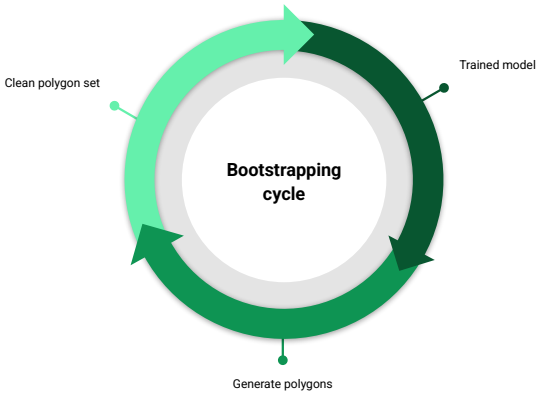
Suspicion: The imprecise labelling is causing issues.

=> Start from a well labelled set of polygons to train a first model

=> Use this model to generate better labels

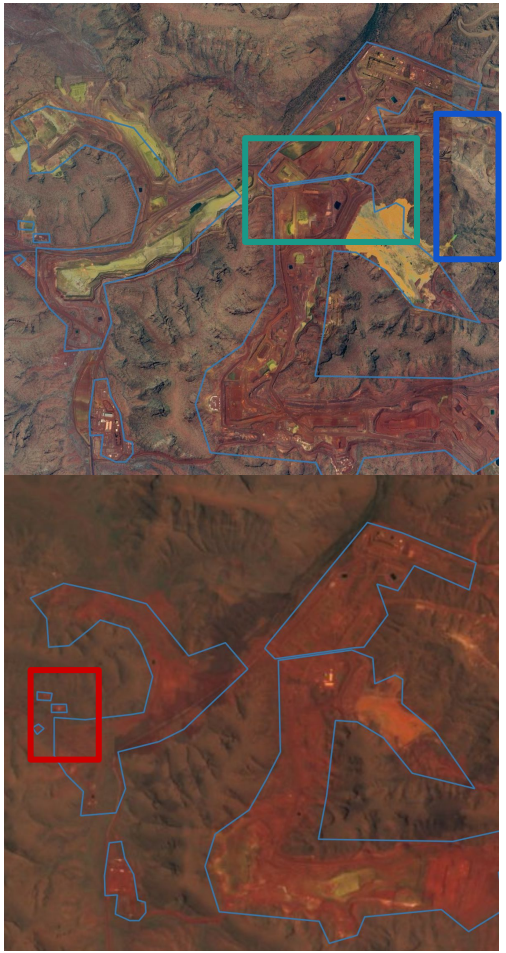
=> Use the better labels to train a better model

....



Cleaning the polygons

- Unification of polygons
  - Ensures a minimal label
  - Lowers labelling error penalty
- Removal of miniscule features
  - What is unobservable can not be learned
- Addition of wrongly unlabelled areas



Cleaning the polygons - Food for thought

What is labelled matters!

- Labelling of small features/additional infrastructure leads to labelling of those features
    - Water collections

Cleaning the polygons - Food for thought

What is labelled matters!

- Labelling of small features/additional infrastructure leads to labelling of those features
  - Water collections
  - Former mining locations

Cleaning the polygons - Food for thought

What is labelled matters!

- Labelling of small features/additional infrastructure leads to labelling of those features
  - Water collections
  - Former mining locations
  - Parking lots
  - Landing strips
- Maybe we should even train on multiple classes
  - What is typical for a mine
  - Combine the multiple labels for a “mining region label”

Training the model

- Due to time pressure only 2100 polygons checked of 3900 in Australia
  - 3 days effort
- Used this halfway cleaned dataset to train a model on Australia (60min)
- Performed inference with this model on:
  - All of Australia (30min)
  - All of Kenya as comparison (30min)
- Performed inference with the previous model on Australia (30min)

Human in the loop reinforcement

With a higher confidence threshold:

1. Generate the labelled set
2. Clean the labels by hand with a human (simple clean no relabelling)
3. Use the improved labelled set to train next model

My remaining time

What I definitely have to do:

1. Document all of the code + cleanup
2. Document general project
3. Ensure all of the code still works

What I could do:

1. Include a supervision pipeline to make the bootstrapping process easier
2. Run one or two loops of bootstrapping
3. Do another experiment?

Some open questions

- With what kind of labelling does the model perform the best?
  - Coarse grained with all additional infrastructure in the label?
  - Fine grained with multiple labels and a combination defining the mine?
- What should be done with unseeable mines?
  - Search for another satellite image solution?
  - Just keep them labelled and hope the model picks up a pattern?
- How can the performance be fairly quantified?
  - Can we even change YOLO and the underlying structure enough?