MACHINE LEARNING FOR ASTROPHYSICS
2ND EDITION
CATANIA, 8-12 JULY, 2024

ML Methods For Space Debris Detection in Bistatic Radar Data

Miguel Andrea Zammit

In collaboration with

Dr Andrea DeMarco, Dr Denis Cutajar & Prof. Alessio Magro

L-Università ta' Malta
Institute of Space Sciences & Astronomy

Xjenza
The Malta Council for Science & Technology

Outline

1. Scientific Motivation
2. BIRALES: Current Setup & MSDS Detector
3. Dataset: Curation & Preparation
4. Model Selection
5. Preliminary Results
6. Next Steps & Future Work

Scientific Motivation

Scientific Motivation

Tracking Resident Space Objects (RSOs)

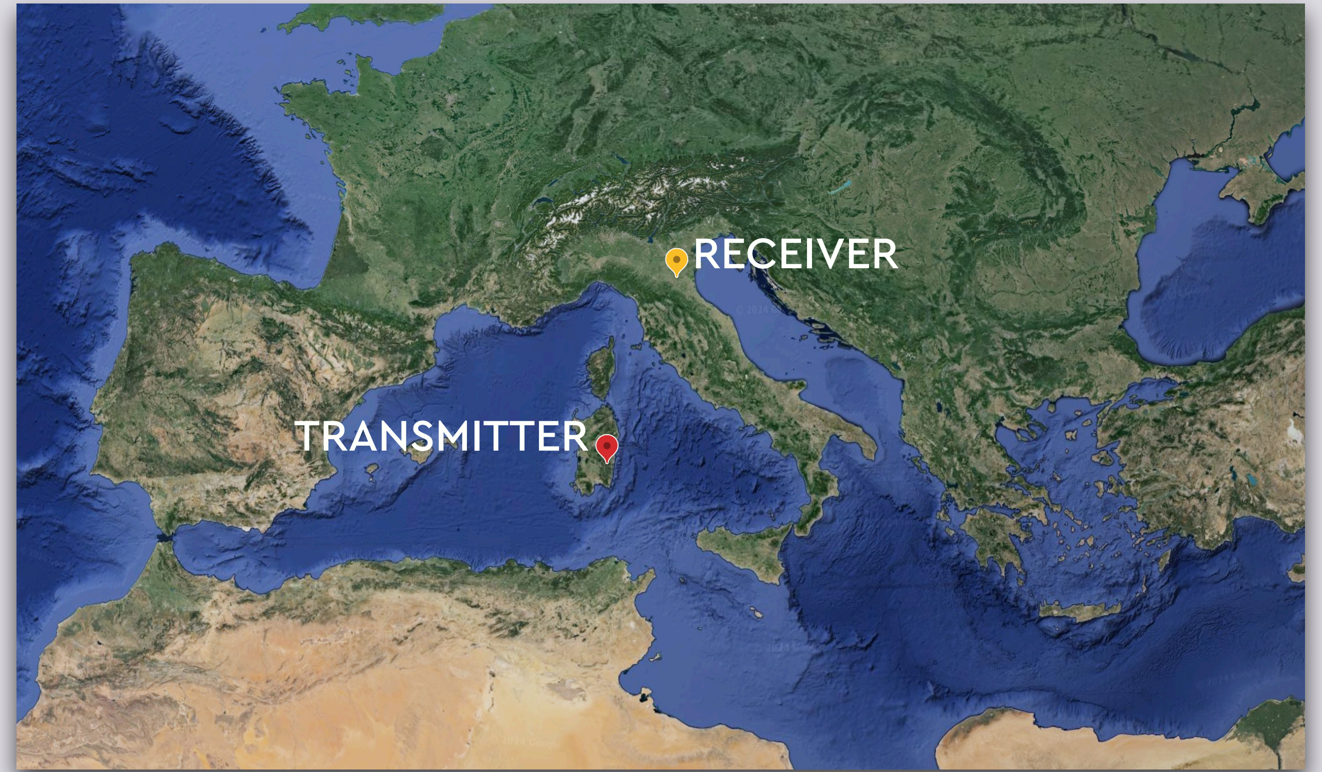
- Since the launch of the first satellite in 1957, the near-earth population has been increasingly steadily.
- As the satellite population grew, so did the population of orbital debris.
- M**onitoring RSOs** has become ever more important in order to reduce the risk of a collision.
- Measurements provide us with a deeper understanding of the current environment, including **growth trends and accumulation regions.**
- Ground-based radars play a critical role in space debris monitoring and space situational awareness (SSA) programs.

BIRALES

Current Setup & MSDS Detector

BIstatic Radar for LEo Survey

Current Setup



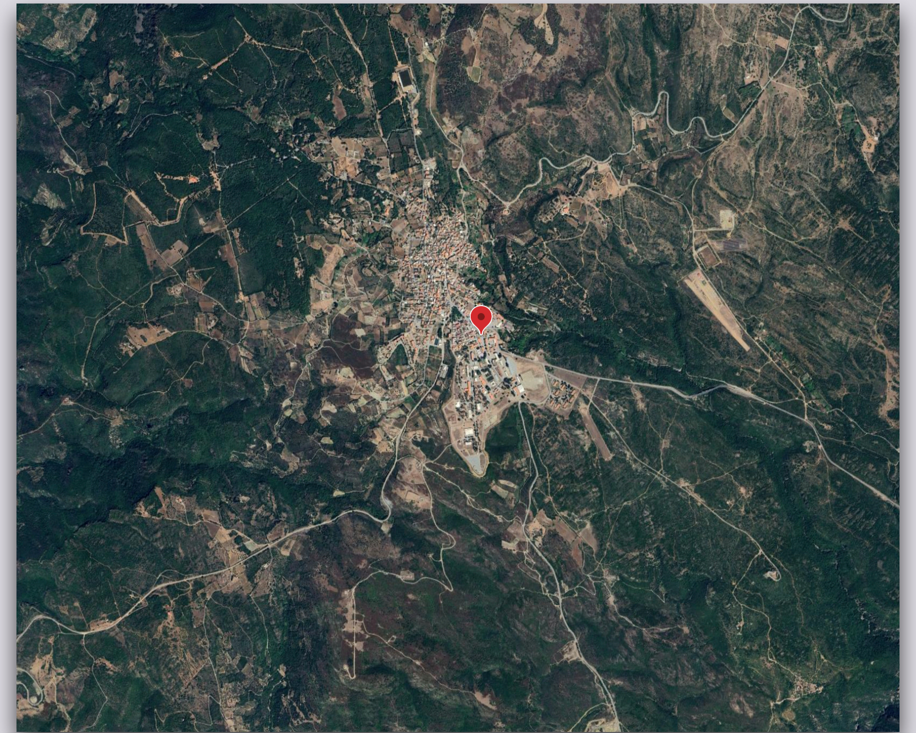
BIstatic Radar for LEo Survey

Current Setup

- 2 Modes: unmodulated Continuous Wave mode & a Compressed Chirp mode
- Sensitivity: small objects with a size of 10 cm at 1,000 km
- Transmitter
    - Diameter 7 m
    - Maximum Power 10 kW
    - Bandwidth 410–415 MHz
    - Beam Size 6 degrees

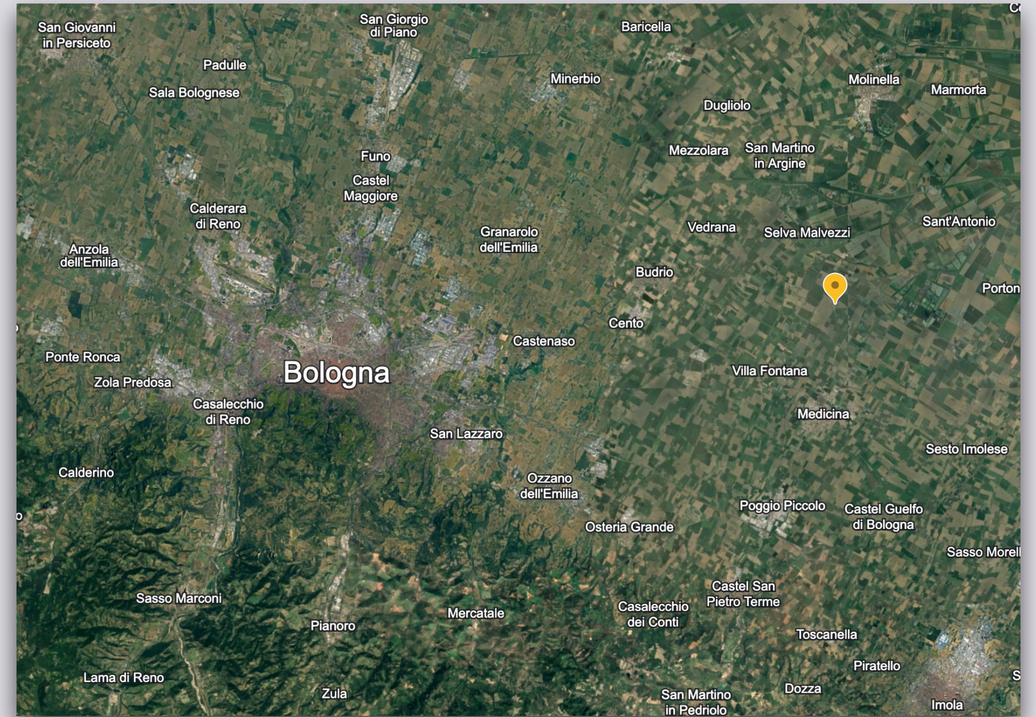
BIstatic Radar for LEo Survey

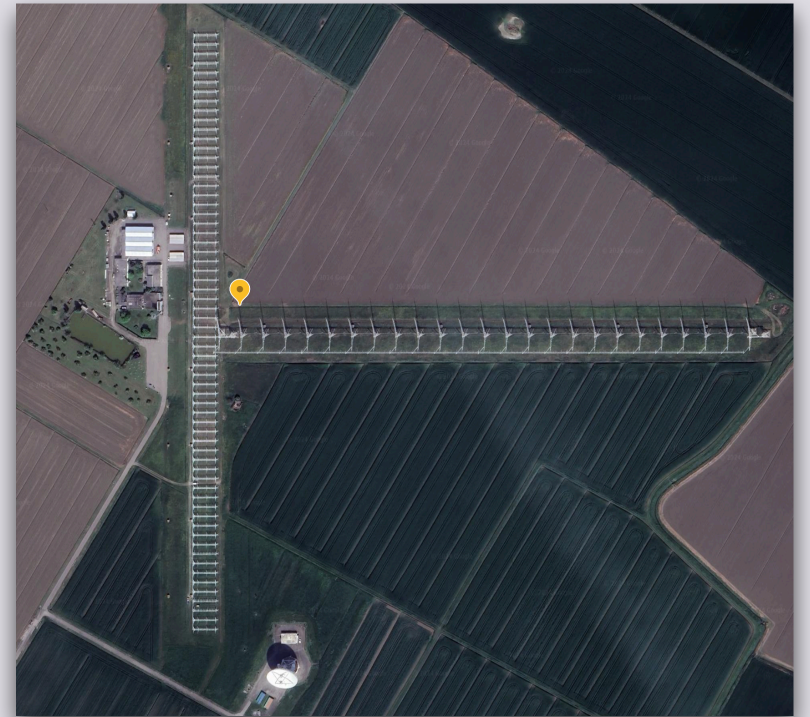
Transmitter



BIstatic Radar for LEo Survey

Receiver: The Northern Cross





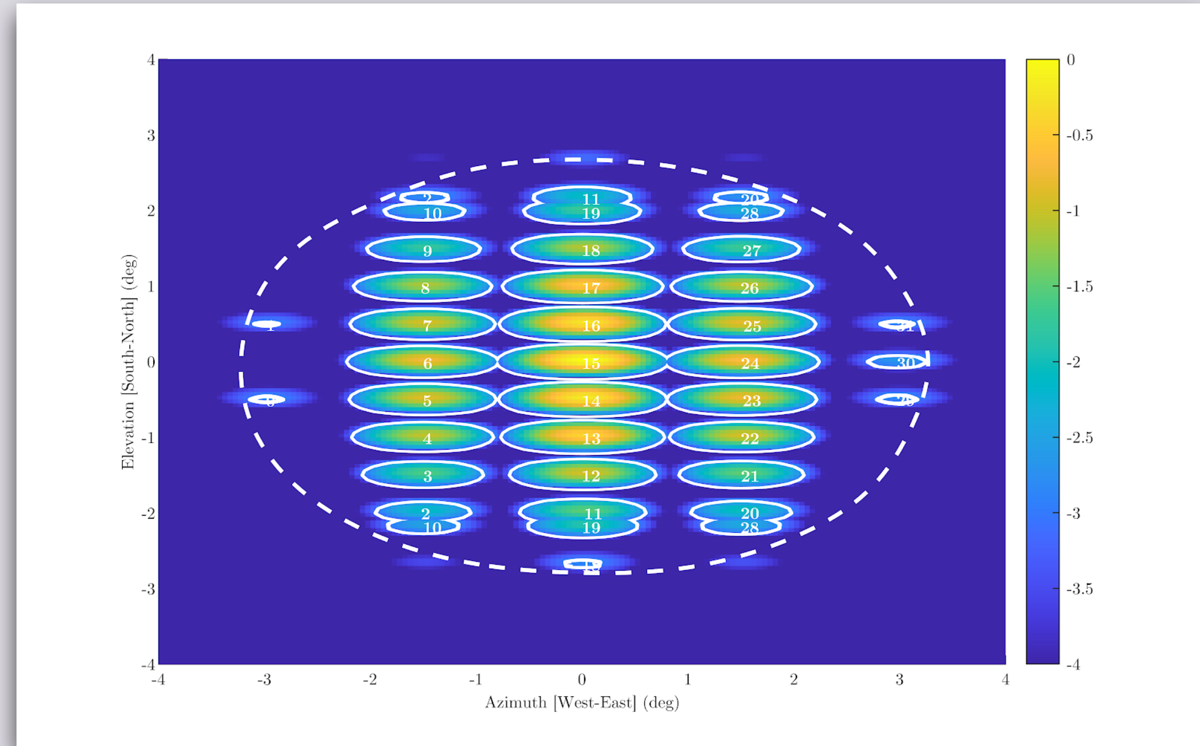
BIstatic Radar for LEo Survey

The Northern Cross

- T-shape array operating at the 408 MHz
- Steerable in declination only.
- The E-W arm is a 564 by 29.4m cylindrical parabolic reflector having a total collecting area of 1,600 m^2.
- The N-S arm is composed of 64 parallel parabolic-shaped cylindrical antennas (22.6 by 7.5 m) spaced 10 m apart having a total geometrical area of 10,800 m^2.
- As part of the Square Kilometre Array Design Studies, 32 receivers in 8 cylinders were upgraded to use new analog fibre-optic and coaxial digital links

BIstatic Radar for LEo Survey

Multi-Pixel Beam Mapping



BIstatic Radar for LEo Survey

Detector Pipeline

1.	2.	3.	4.
Image Segmentation	**Track Initiation**	**Track Association**	**Track Termination**
RFI and background noise filtering	Track detection algorithm	Compare tracks together in order to find similar streaks	Terminate validated tracks or cancel false positives
Triangle or Sigma Clip	DBScan or MSDS	Cosine similarity or Coefficient of Variation	

BIstatic Radar for LEo Survey

Detector Pipeline

1.	2.	3.	4.
Image Segmentation	**Track Initiation**	**Track Association**	**Track Termination**
RFI and background noise filtering	Track detection algorithm	Compare tracks together in order to find similar streaks	Terminate validated tracks or cancel false positives
Triangle or Sigma Clip	DBScan or MSDS	Cosine similarity or Coefficient of Variation	

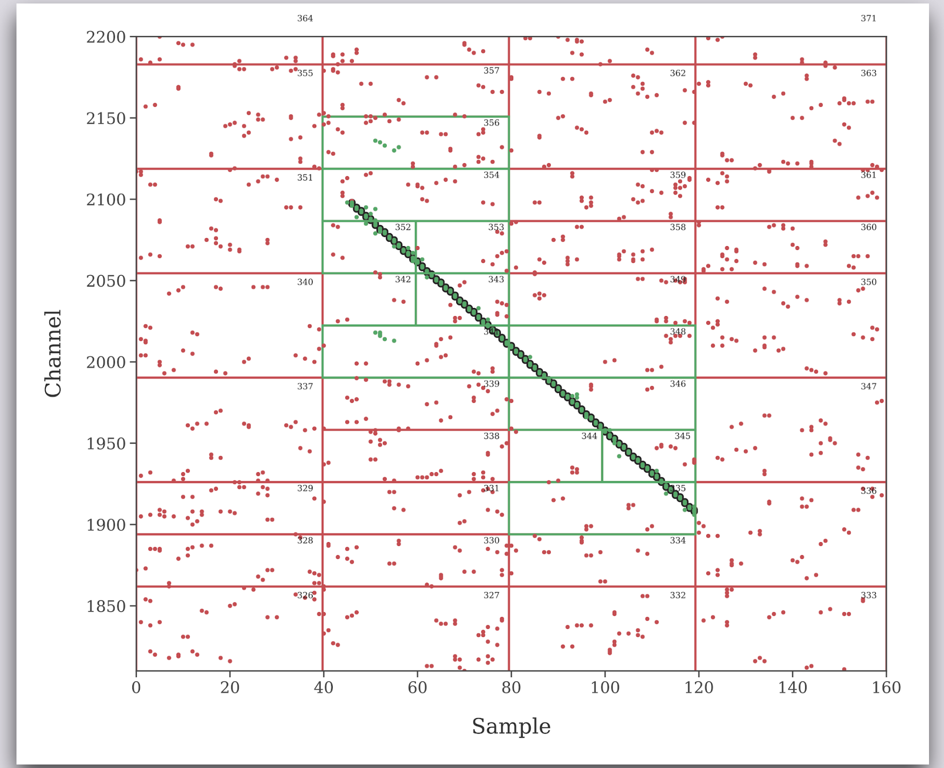
BIstatic Radar for LEo Survey

MSDS Detector

- The current pipeline uses a clustering approach
- Sub-divides the de-noised domain in rectangular boxes using kd binary tree.
    - Each rectangle is further split along an axis (vertical or horizontal) recursively until the number of points within the rectangle reaches a predefined leaf size.
    - Each leaf is checked for a line streak using an Agglomerative hierarchical clustering algorithm.
    - The similarity between clusters is determined, penalising two points whose gradient is not within the expected range.
    - Cluster's shape is checked for linearity using the inertia ratio. In this formulation, an inertia ratio of 1 indicates a circle while a perfect line has an value of 0.
    - The next stage is to correctly merge the individual clusters detected across the leaves into a single larger track.

BIstatic Radar for LEo Survey

MSDS Detector



Cutajar et al (2022)

BIstatic Radar for LEo Survey

Planned Expansion

- An upgrade is scheduled for the BIRALES receiver in Medicina
  - Slated for 2025

- Number of antennas will increase drastically (32 antennas → 684 antennas)
- Sensitivity will increase ~30 times
- Beam will be larger

- Expected Data Rate will mean MSDS will not be able to keep up in real-time.

Dataset

Curation & Preparation

Building a Dataset for the Model

Data Collection & Curation

- Data acquired from the BIRALES archive.
- ~100 Observations, (~7% Tracking Observations, ~93% Survey Observations) from April to November 2023
- Every observation was labelled using the offline detection pipeline already installed in Medicina.
- The MSDS algorithm was used to label the data by providing the channel/sample values of all detected streaks for each beam. It serves as the current benchmark with which the new algorithms will be compared against.

Building a Dataset for the Model

Data Preparation

- From the collated detections for every observation, the data was prepared as follows:
    - As a general format, all beams were tiled as 1024×1024 images, normalised as grayscale images.
        - Will need further tiling prior to being fed to specific models to avoid resizing.
    - Data Augmentation to further increase the number of annotated images was done by taking randomly shifted tiles across the face of the beam.
    - If any annotations are present within a tile, values are converted
        - time sample and receiver frequency → pixel coordinates
    - COCO format was used for all image, annotations, and category mapping

Building a Dataset for the Model

Data Preparation

- To avoid introducing systematics and ensure that the training, validation and test sets are balanced, stratified shuffle splitting was done based on:
  - Beam
  - ratio of single/multiple annotations
  - Maximum Raw Power prior to normalisation, stratified by interquartile range bins.
  - Maximum Width of annotations in image, stratified by interquartile range bins.

Building a Dataset for the Model

Data Preparation

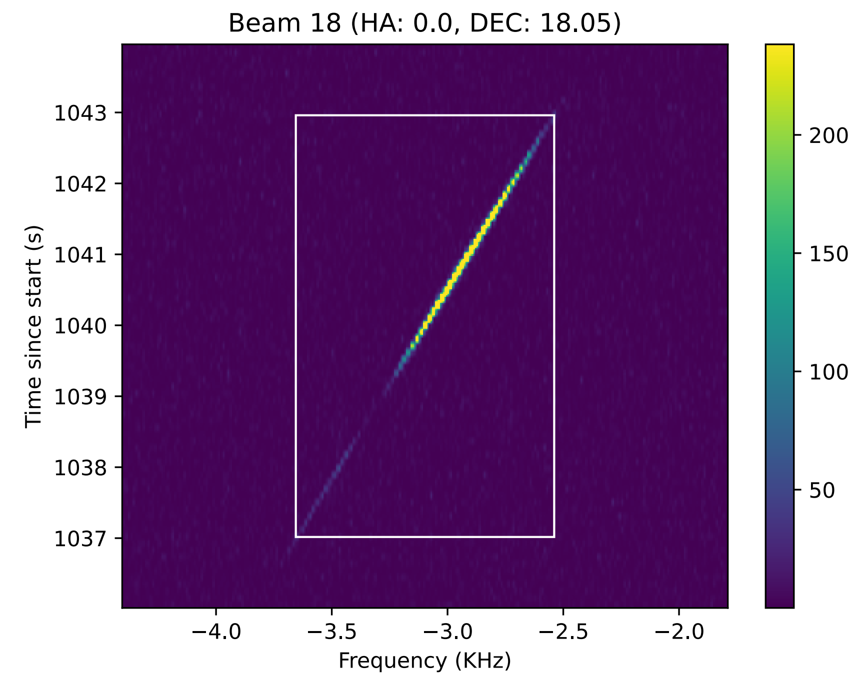
- Full dataset size with annotations: 50,749 samples

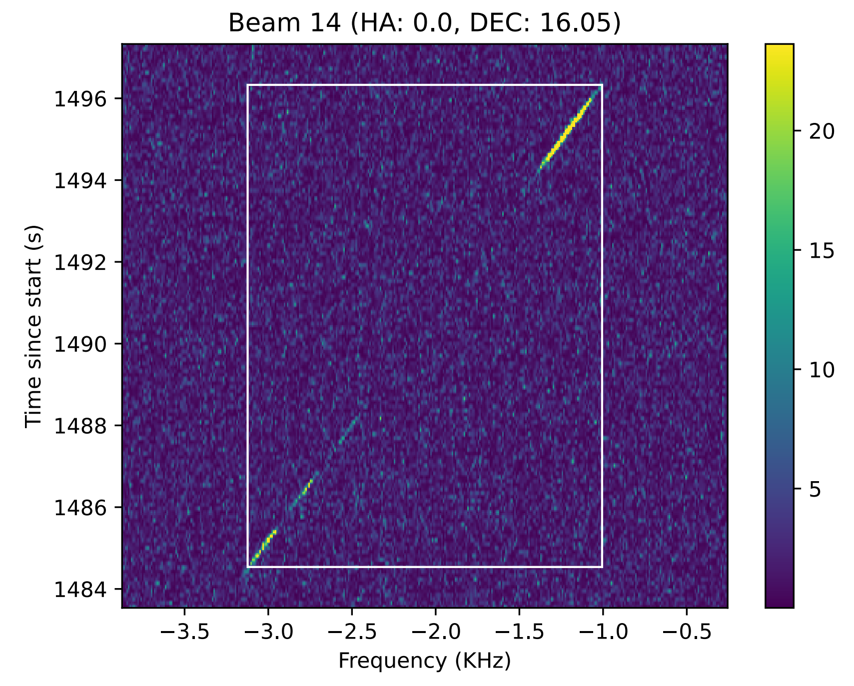
- Final Dataset sizes:
  - Training Set: 30,447 samples
  - Validation Set: 10,152 samples
  - Test Set: 10,150 samples

Building a Dataset for the Model

Things to Keep in Mind About the Data

- The visibility, duration and nature of the debris echoes will vary substantially.

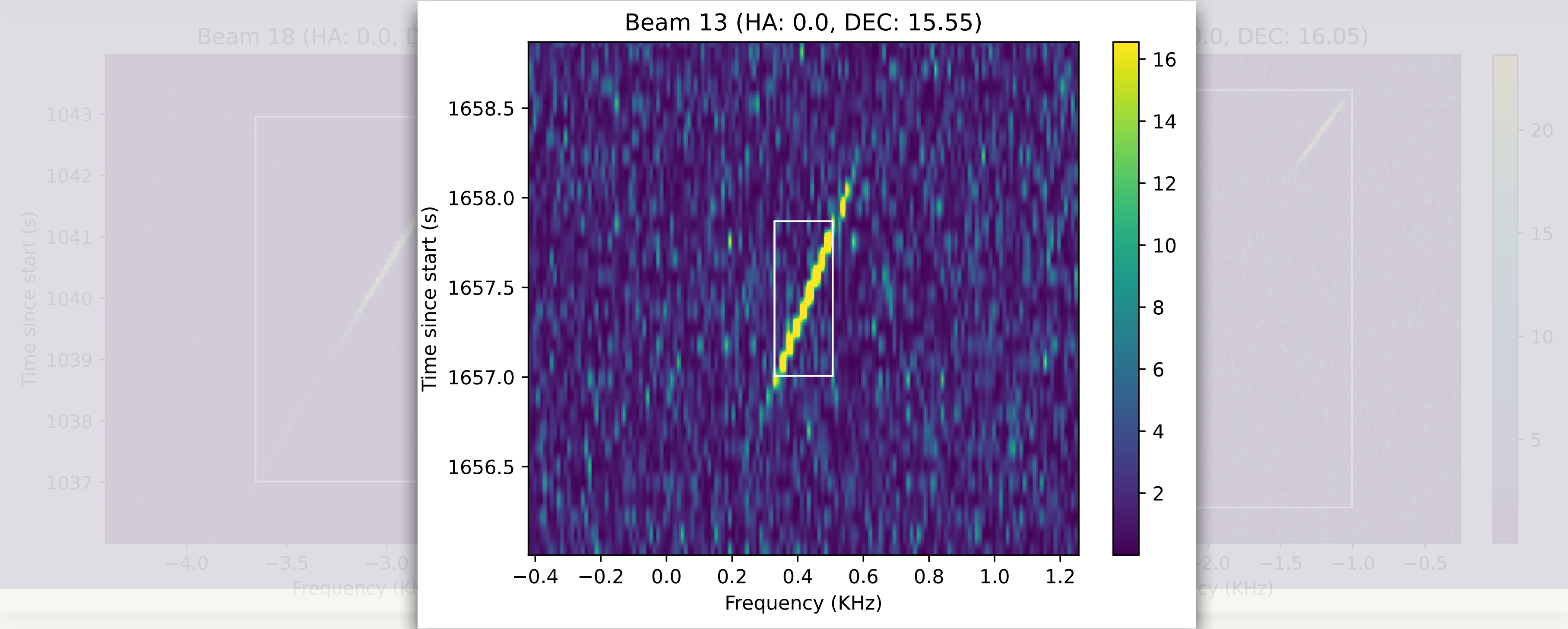




Building a Dataset for the Model

Things to Keep in Mind About the Data

- The visibility, duration and nature of the debris echoes will vary substantially.



Building a Dataset for the Model

Things to Keep in Mind About the Data

- The visibility, duration and nature of the debris echoes will vary substantially.
- A large amount are a pixel wide, some only having a width and height of ~20 pixels.
    - Image Rescaling prior to being fed into any model needs to be avoided.
- Any model used will need significant adjustment for this use case. So a large dataset and substantial training time/resources will be required.
- Since the data is labelled by the MSDS detector, it will be limited to a degree by the detector's recall and precision.
    - Important that any final model is significantly regularised.

Model Selection

Choosing the Right ML Model

Object Detection

- Care needs to be taken about selecting the right model for the job.
- Important that training and running inference from the model used is not computationally expensive and number of trainable parameters is manageable with our resources.
- Transfer learning from a pre-existing model that is versatile and reliable for different use cases would be ideal.

You Only Look At One Sequence (YOLOS)

YOLO Vision Transformer (ViT)

- YOLOS leverages the ViT architecture, enabling it to effectively capture spatial relationships in images.
  - Particularly beneficial for detecting irregularly shaped and positioned objects
- Originally trained using DETR loss on ImageNet-1k (200 epochs) and COCO2017 (150 epochs).
- The base-size model achieves an AP of 42 on COCO2017 validation.
- We select the small variant (*hustvl/yolos-small*) for a good balance between performance and manageability.
  - Number of trainable parameters: 30.7M

You Only Look At One Sequence (YOLOS)

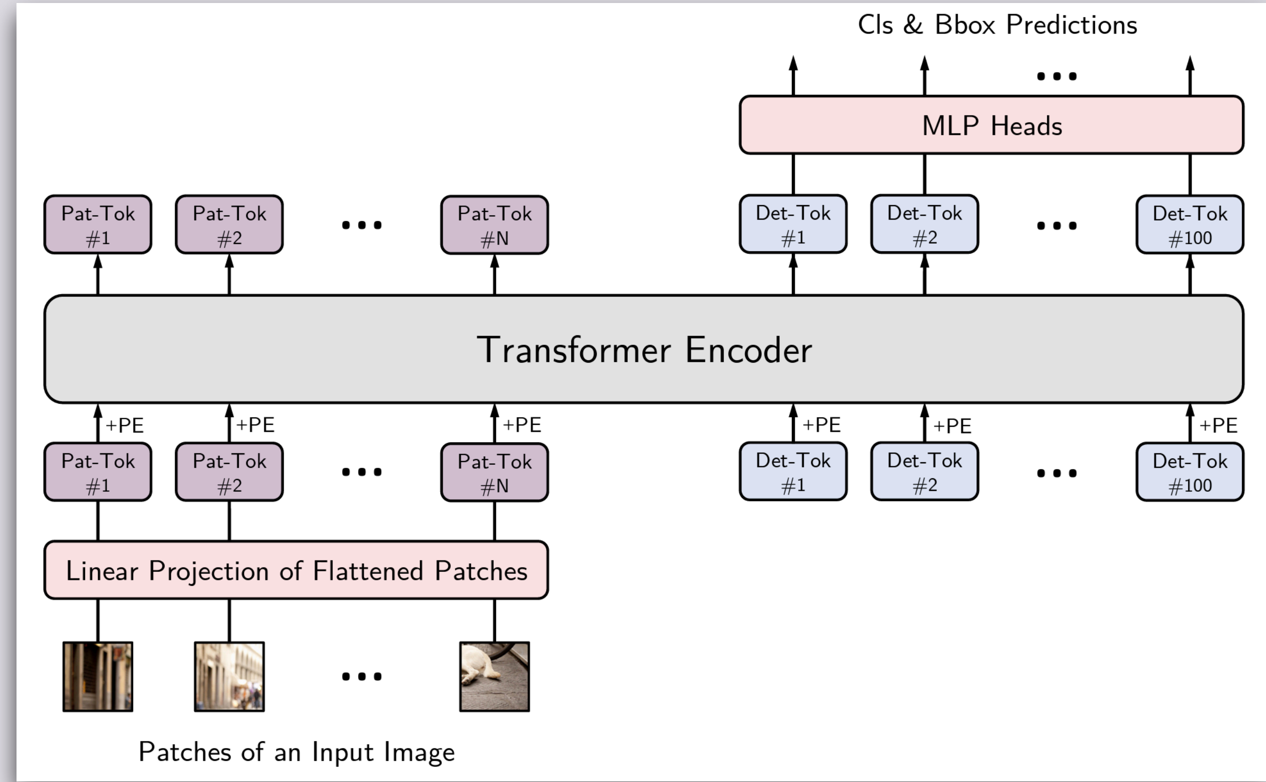
YOLO Vision Transformer (ViT)

- Trained using a bipartite matching loss
  - Hungarian matching algorithm achieves a 1–1 mapping from the 100 object queries given by the model to a padded list of annotations in the image.
  - To optimise model parameters, a combined loss of the cross-entropy (label classifier) and the L1 & Generalised IoU loss (bbox predictor) is used.

$$\text{Loss} = \lambda_{class}\text{ClfLoss} + \lambda_{box}L_1\text{Loss} + \lambda_{gIoU}\text{gIoULoss}$$

You Only Look At One Sequence (YOLOS)

YOLO Vision Transformer (ViT)



Fang et al (2021)

You Only Look At One Sequence (YOLOS)

YOLO Vision Transformer (ViT)

- Some further preprocessing on our data is required to ensure it is suitable as an input for YOLOS.
    - Images are 'converted' from grayscale to RGB shape.
    - To ensure z-score normalisation is consistent across the dataset, a sample of 2000 images are taken to calculate the average image mean and std.
    - Images are further tiled to match the 512×864 shape input. This ensures that all resizing is avoided.
    - Only tiles with annotations are kept.

Preliminary Results

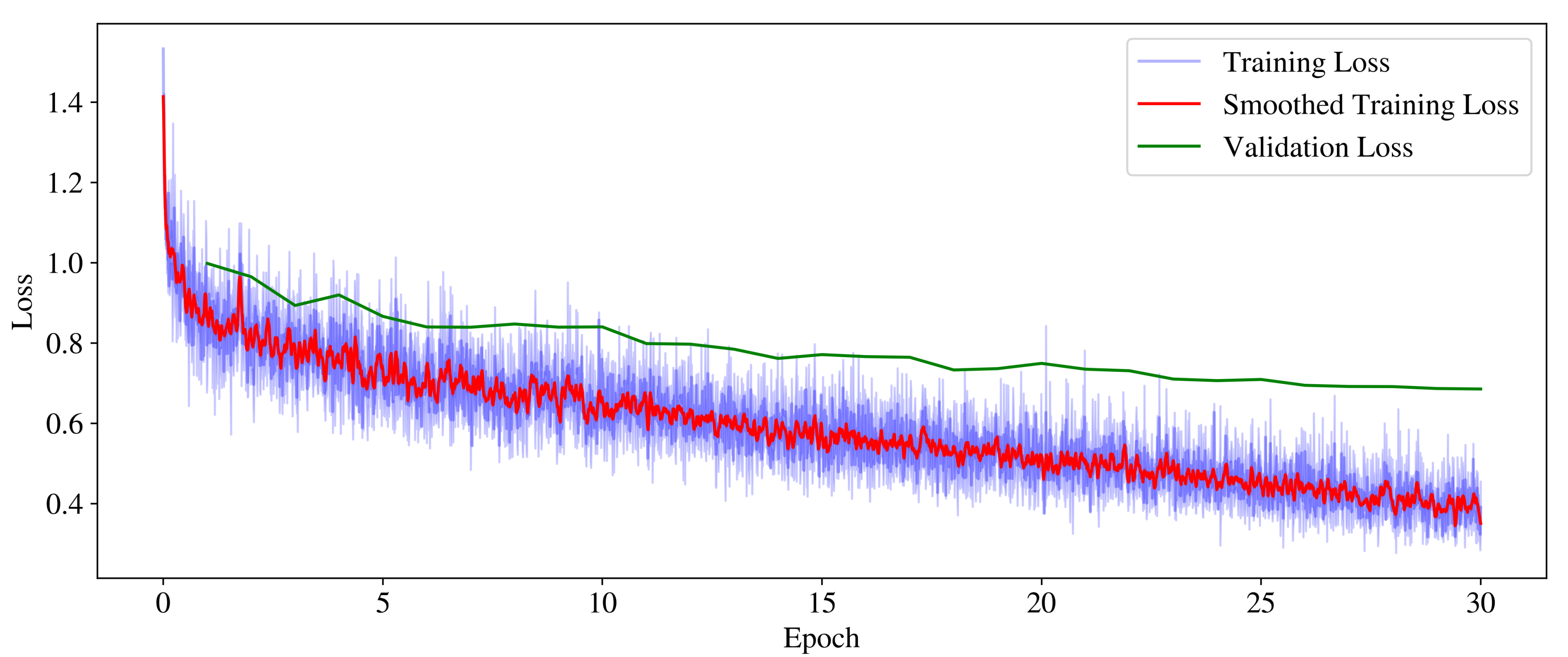
(Very) Preliminary Results

Training Curves

- We are still in the stage of trying to minimise the training loss
- The results presented here have been trained on 20,643 tiles (10k images) and validated on 15,682 tiles (10k images).
- Training was done on a single Tesla P100-SXM2–16GB GPU
    - Batch size = 8
    - Linear scheduling with a starting learning rate of the order of 10^{-5}.
    - ~34 hours for 30 epochs

(Very) Preliminary Results

Training Curves



(Very) Preliminary Results

Training Curves

- Further Hyperparameter tuning is needed to first continue to minimise the training loss as much as possible
    - Both in terms of the model configuration and optimisation during training
- The loss has not plateaued just yet, so with more epochs and more data, further minimisation is likely.
- Regularising the model to further minimise overfitting will be the next important step.

(Very) Preliminary Results

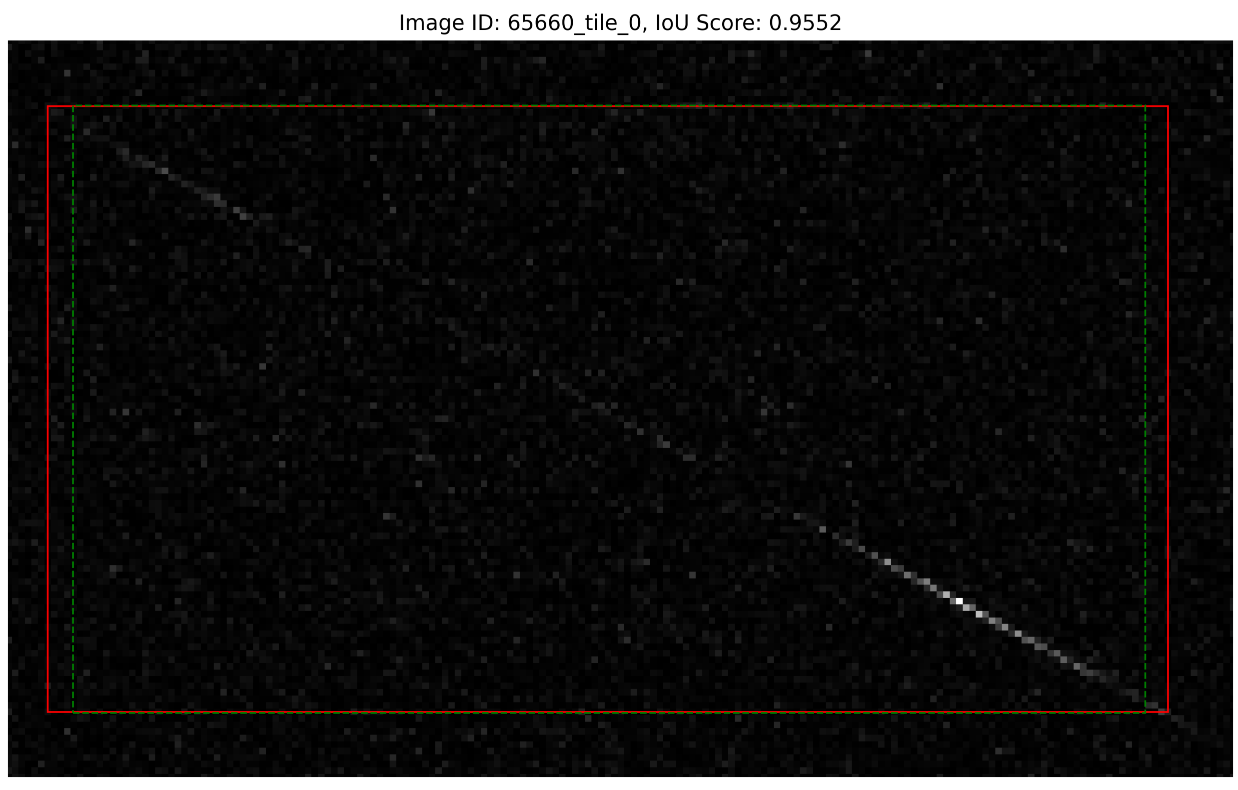
Promising Inference

- After 30 epochs, the number of FPs, even on the training set, is still very high, highlighting the need to minimise the loss further.
- However, there are consistently a number of inferences with promising Intersection over Union (IoU) scores that show that the model is starting to learn.
- Tends to perform best with longer, wider echoes
    - Expected since the pre-trained model was trained on datasets in which the targets are either on the foreground or take up a significant portion of the image.

(Very) Preliminary Results

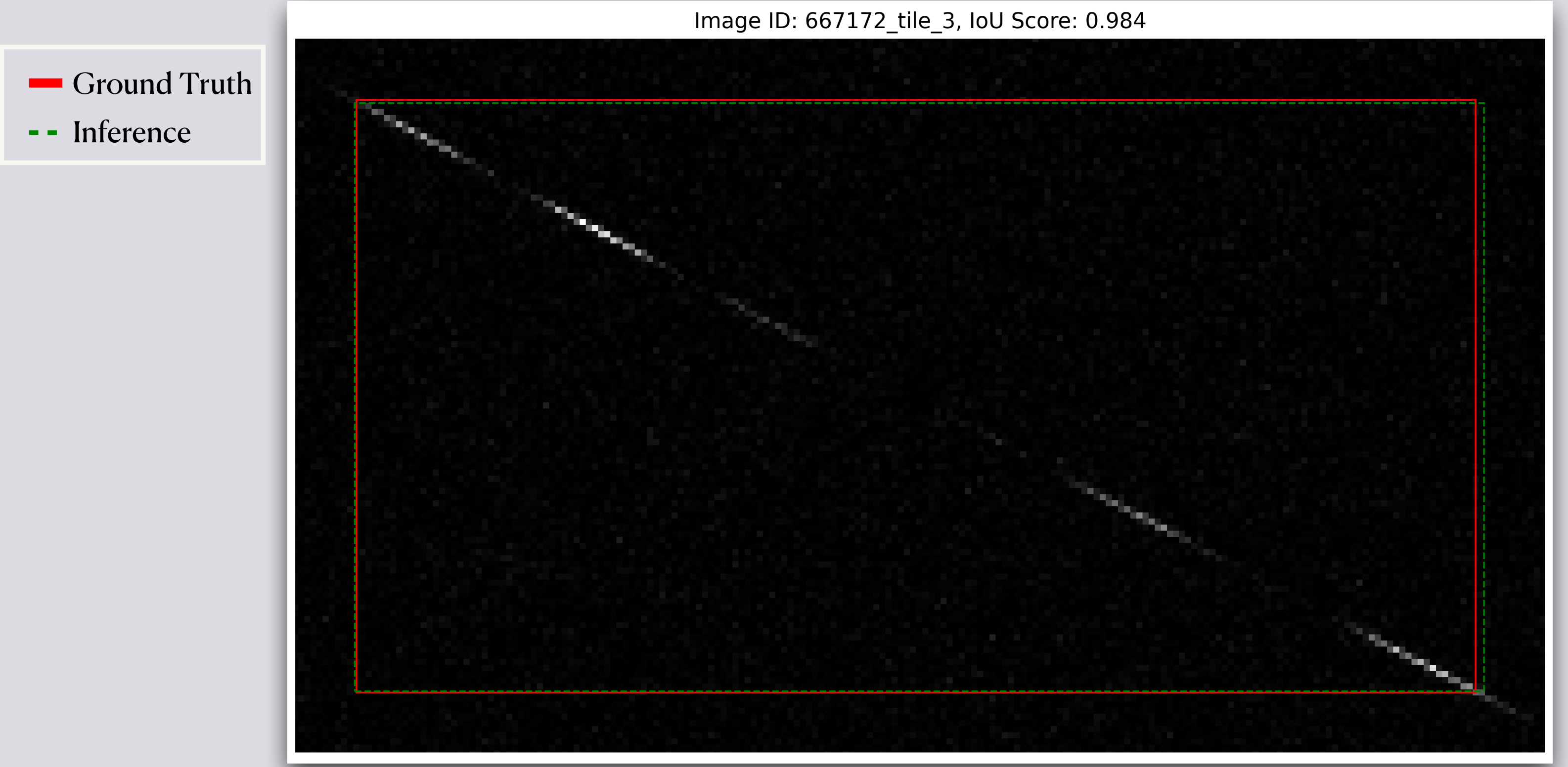
Promising Inference

Ground Truth
Inference



(Very) Preliminary Results

Promising Inference

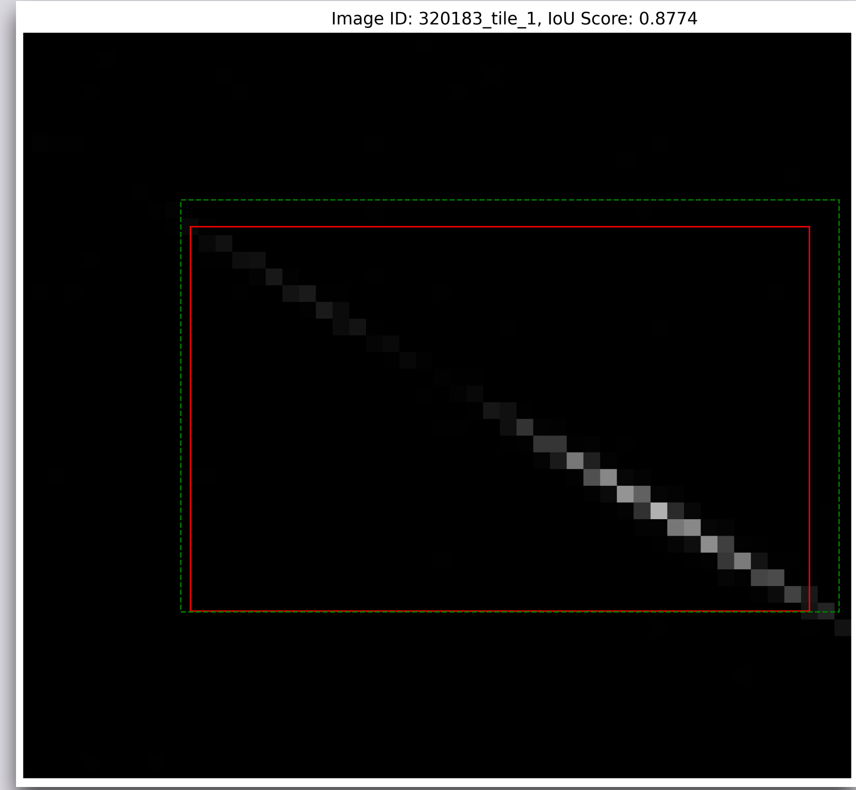


(Very) Preliminary Results

Promising Inference

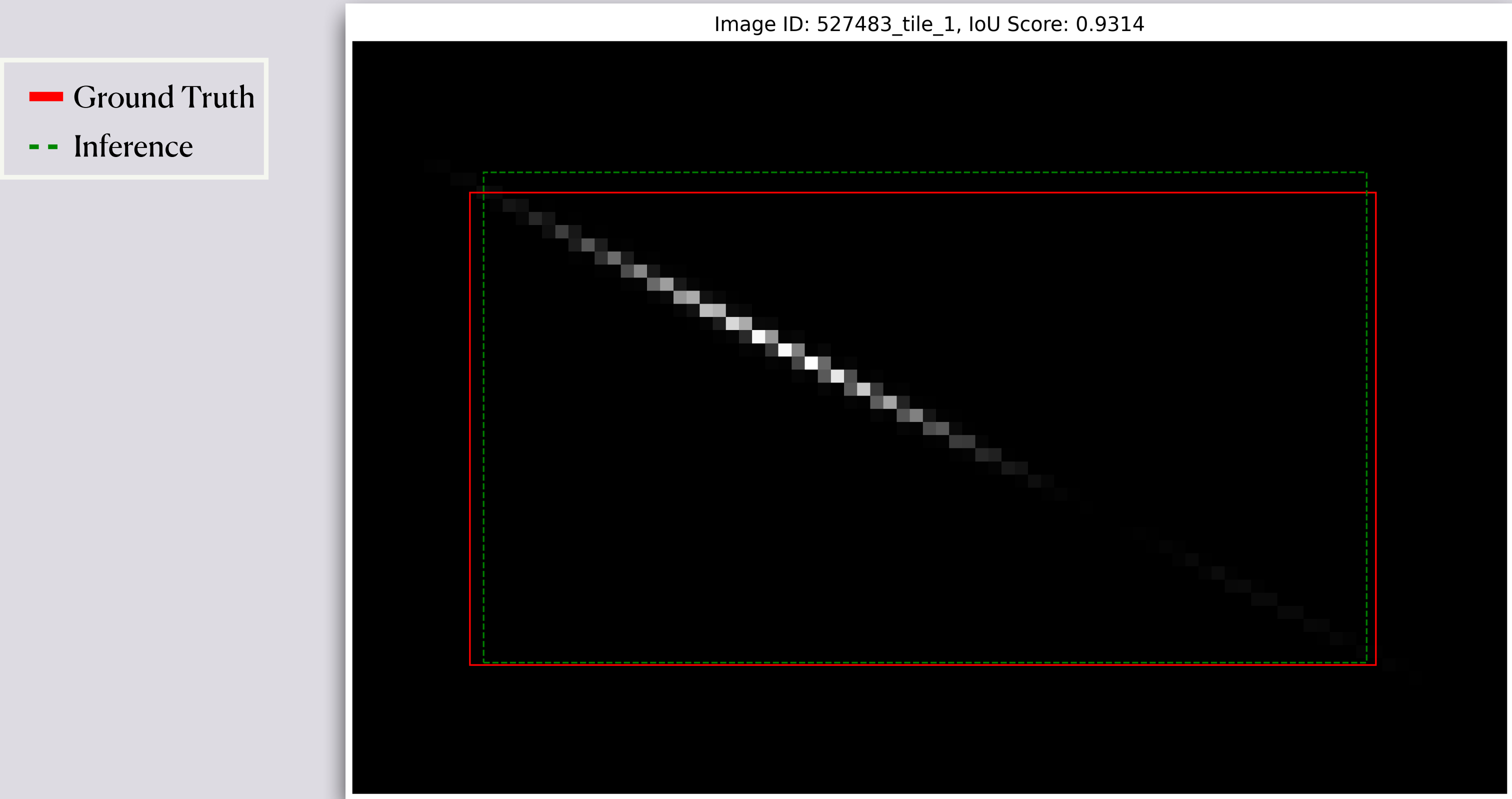
Ground Truth

Inference



(Very) Preliminary Results

Promising Inference



Next Steps & Future Work

Next Steps & Future Work

Short-Term

- Hyperparameter Tuning, both in terms of model and training configuration
    - Get the Training Loss down
    - Regularise the model for stable generalisation
- Explore model performance across observation SNR, duration & beams.
- Test a number of other YOLO (CNN-based & transformer-based) pre-trained models, as well as other popular model architectures.

Next Steps & Future Work

Longer-Term

- Develop a simulator to mitigate the reliance on observational data.
    - A mix of observational and synthesised instances will be ideal.
- Explore whether models trained entirely on our dataset leads to stronger & more stable performances.
- Investigate whether segmentation models perform better or worse than object detection models.
- Integrate the most successful models into the BIRALES detection pipeline and ensure real-time performance is maintained.

Thank You For YourAttention!

Feel free to ask any questions