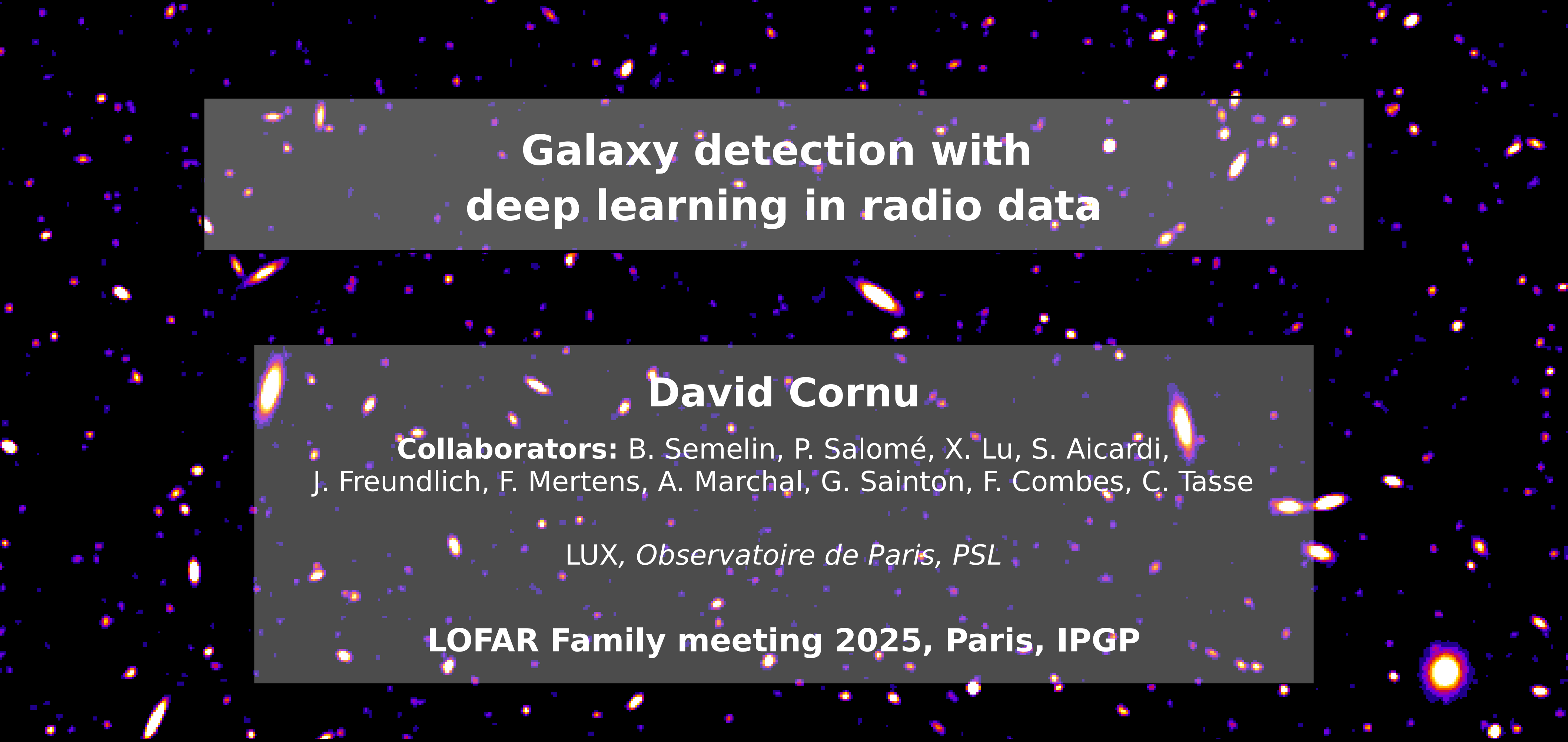


Galaxy detection with deep learning in radio data

David Cornu

Collaborators: B. Semelin, P. Salomé, X. Lu, S. Aicardi, J. Freundlich, F. Mertens, A. Marchal, G. Sainton, F. Combes, C. Tasse

LUX, *Observatoire de Paris, PSL*

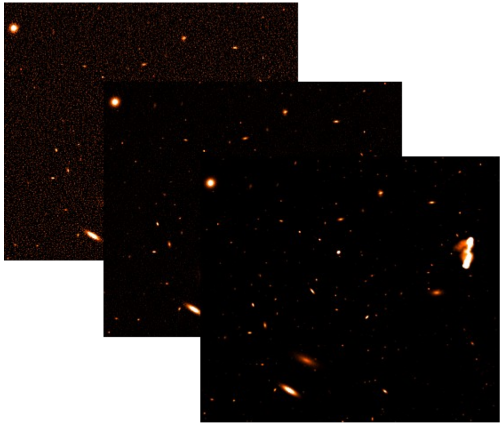
LOFAR Family meeting 2025, Paris, IPGP

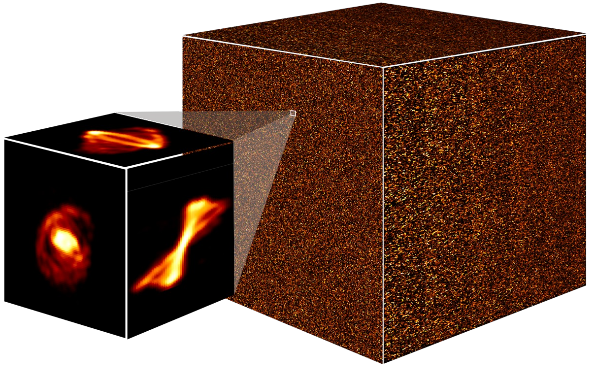
LUX Observatoire de Paris | PSL MINERVA

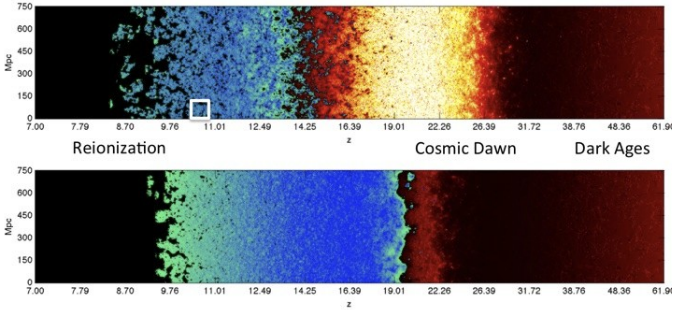
SKAO Science Data Challenges (SDCs)

Simulated datasets that should resemble typical SKA data products

Source detection and characterization







Florent Mertens' talk on Monday

SDC1: Continuum 2D images
3 integration times x 3 bands
Each image = 4 GB
From Dec 2018 to April 2019

SDC2: Hyperspectral cube of HI emission
Full cube = 1 TB
From Feb 2021 to July 2021

SDC3: 21 cm emission
Visibility and Image
Full size ∼ 7 TB
EoR Focused, 2023-2025

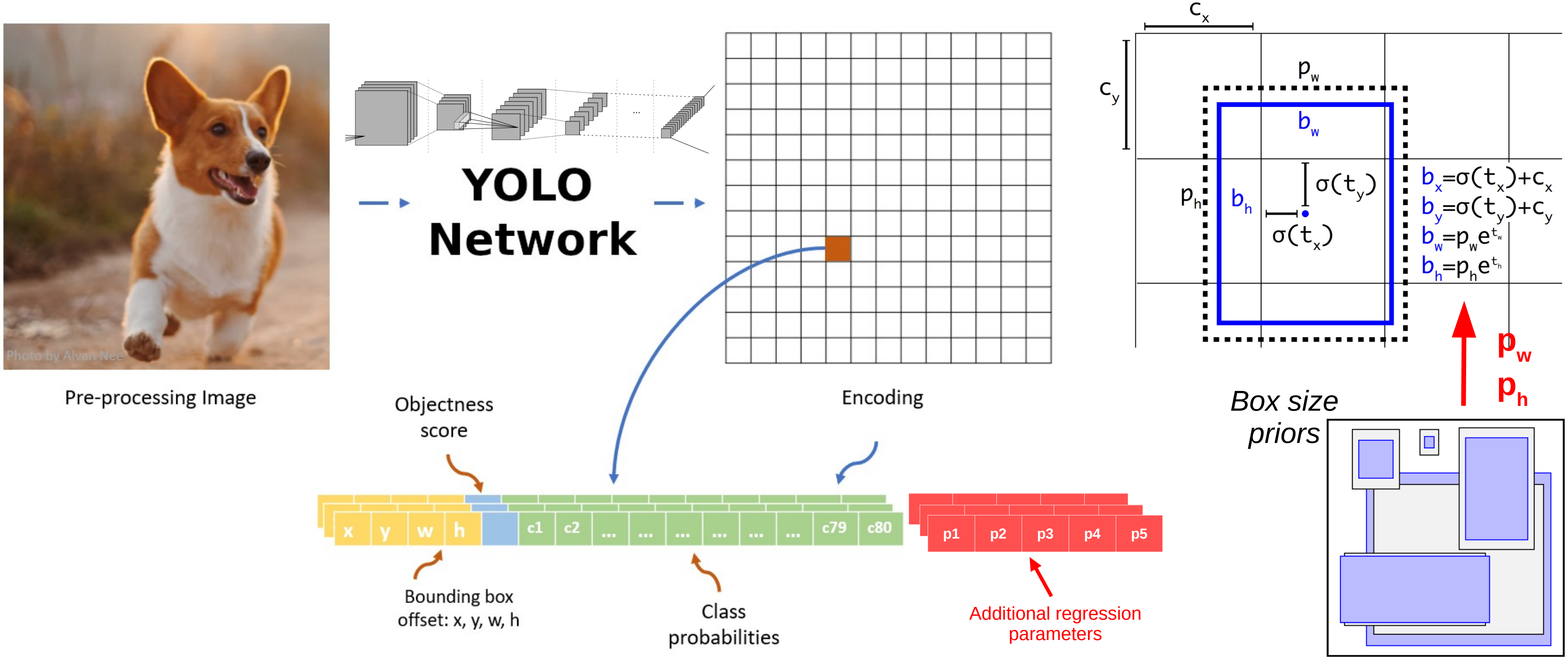
Our objective → develop a deep learning approach to tackle both the SDC1 and SDC2

You Only Look Once (YOLO)

Regression-base deep learning object detector

Originally introduced in Redmon et al. 2015 (V1), 2016 (V2), 2018 (V3)

**Images from blog post and Redmon's papers*



Supervised method → **learns from a list of bounding box examples**

Application to SKAO SDC1

SKA SDC1 summary paper, Bonaldi et al. 2021

Data:

Simulated continuum image:

- 5.5 square degree area (pixel size 0.6’’)
- 560 MHz, 1000h integration time
- **4GB image (32,768 pixel square)**

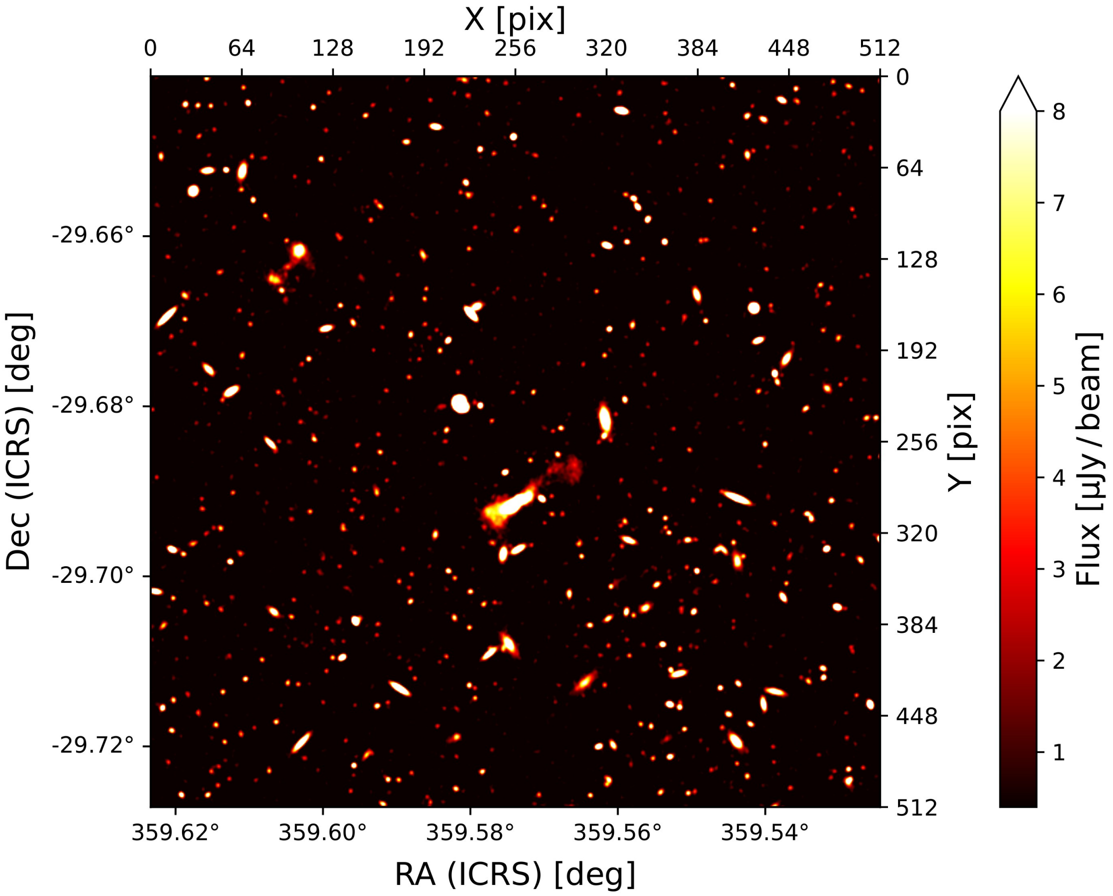
The challenge:

1. Find the sources (RA, Dec)
2. Characterize each source:
   → (Flux, Bmaj, Bmin, PA, ...)

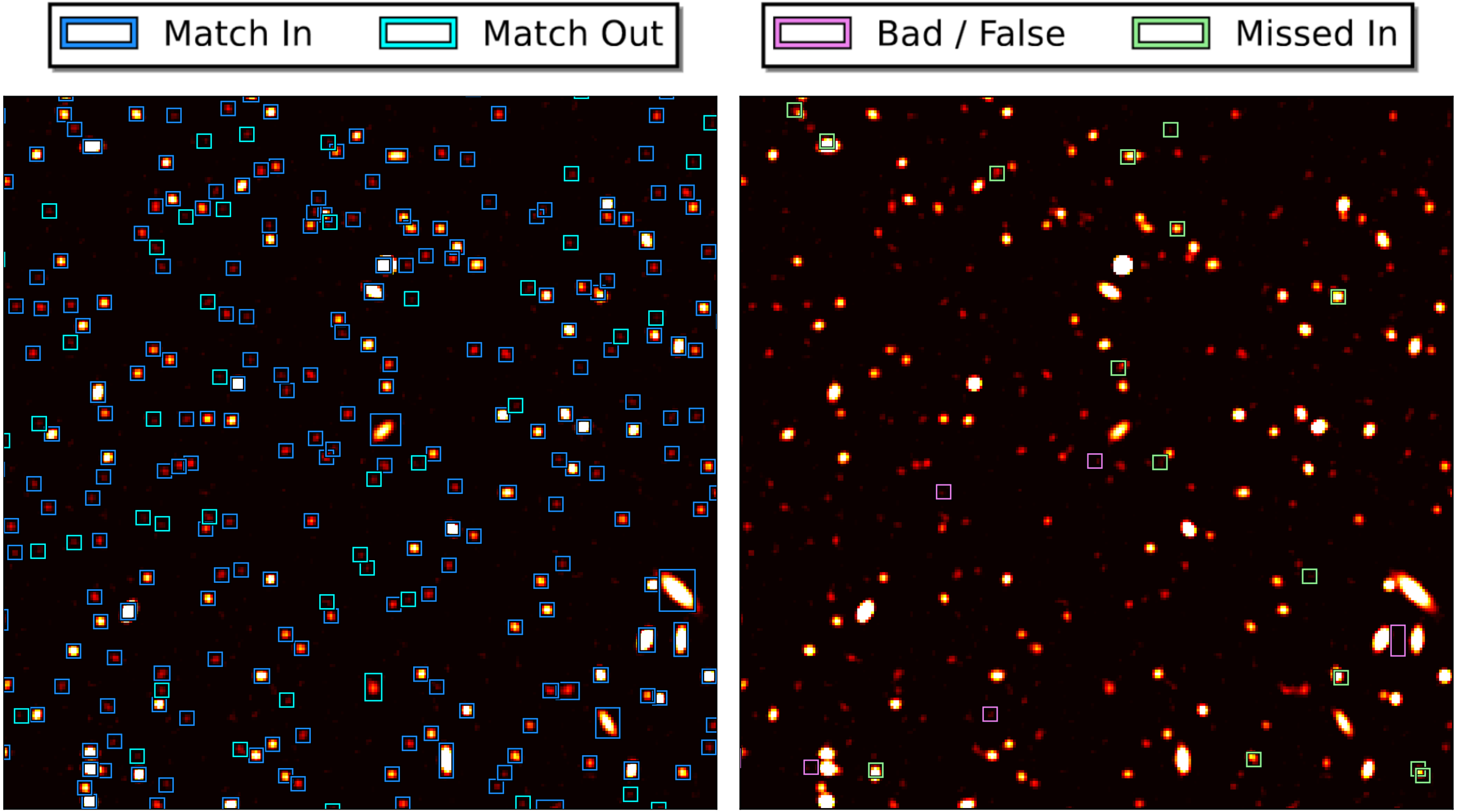
Training labels provided for a subpart of the image (5% of the surface, ~34 000 sources).

SKA SDC1 took place early 2020. Challenge data are publicly available on the dedicated web-page.

Example 512² sub-field

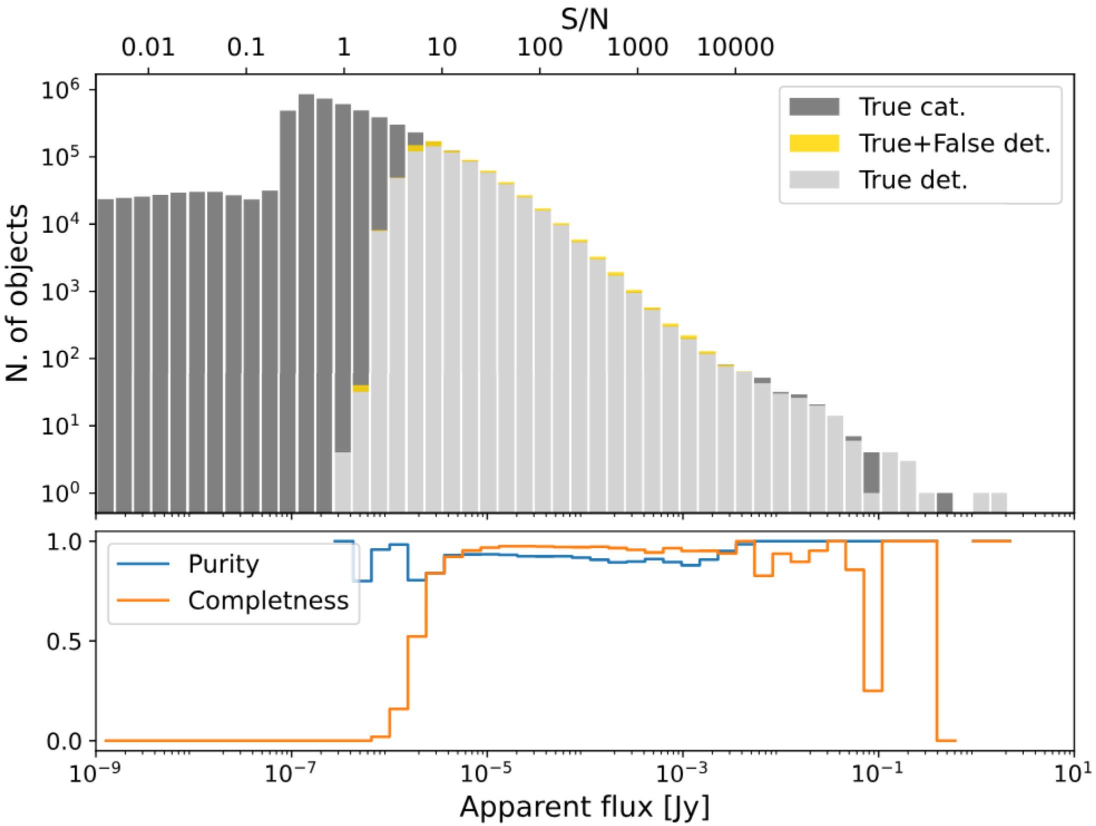


Detection example fields



Global result

MINERVA team paper SDC1, YOLO-CIANNA → Cornu et al. 2024, A&A 690 A211



Comparison to other teams

- Challenge score **2.4 times higher** than the original **SDC1 winning team.**
  - Detect 60% more sources
  - Best characterization accuracy
- Challenge score **1.6 times higher** than the other post-challenge score published.

Prediction time for the full image ~8 sec

Using a single RTX 6000 ada GPU

SCIENCE DATA CHALLENGE 2

SKA SDC2 summary paper, Hartley et al. 2023

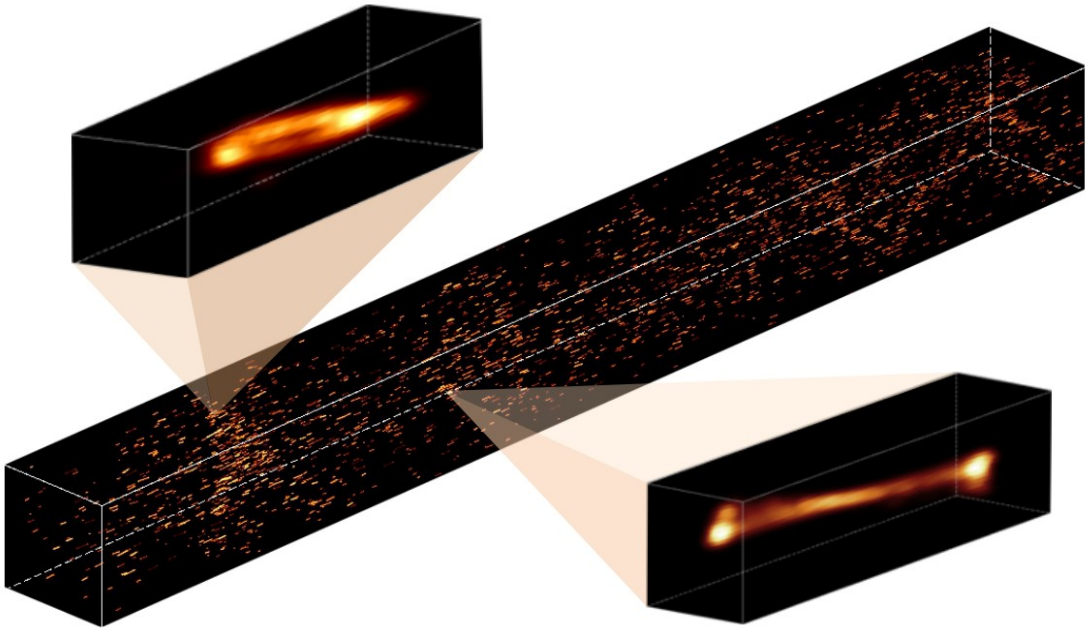
Data: a 3D cube of simulated HI emission

- 20 square degree area
- 950 to 1150 MHz frequency (30KHz res; z = 0.235–0.495)
- **2000h integration time**
- **Near 1 TB cube (5851 x 5851 x 6668)**

The challenge:

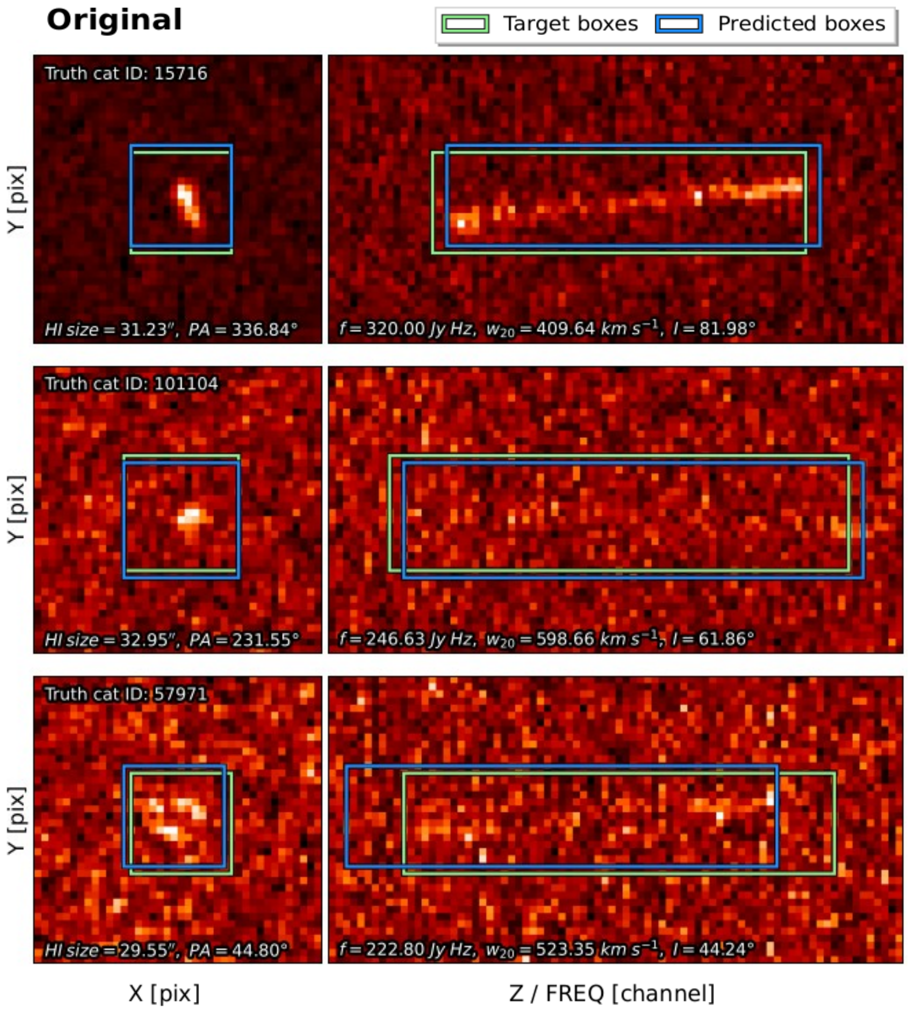
1. Find the sources (RA, Dec, Freq)
2. Characterize each source:
   → Flux, HI size, line width, PA, Inclination

Training labels available for a secondary 40GB cube (1 sq deg, ~1600 sources)



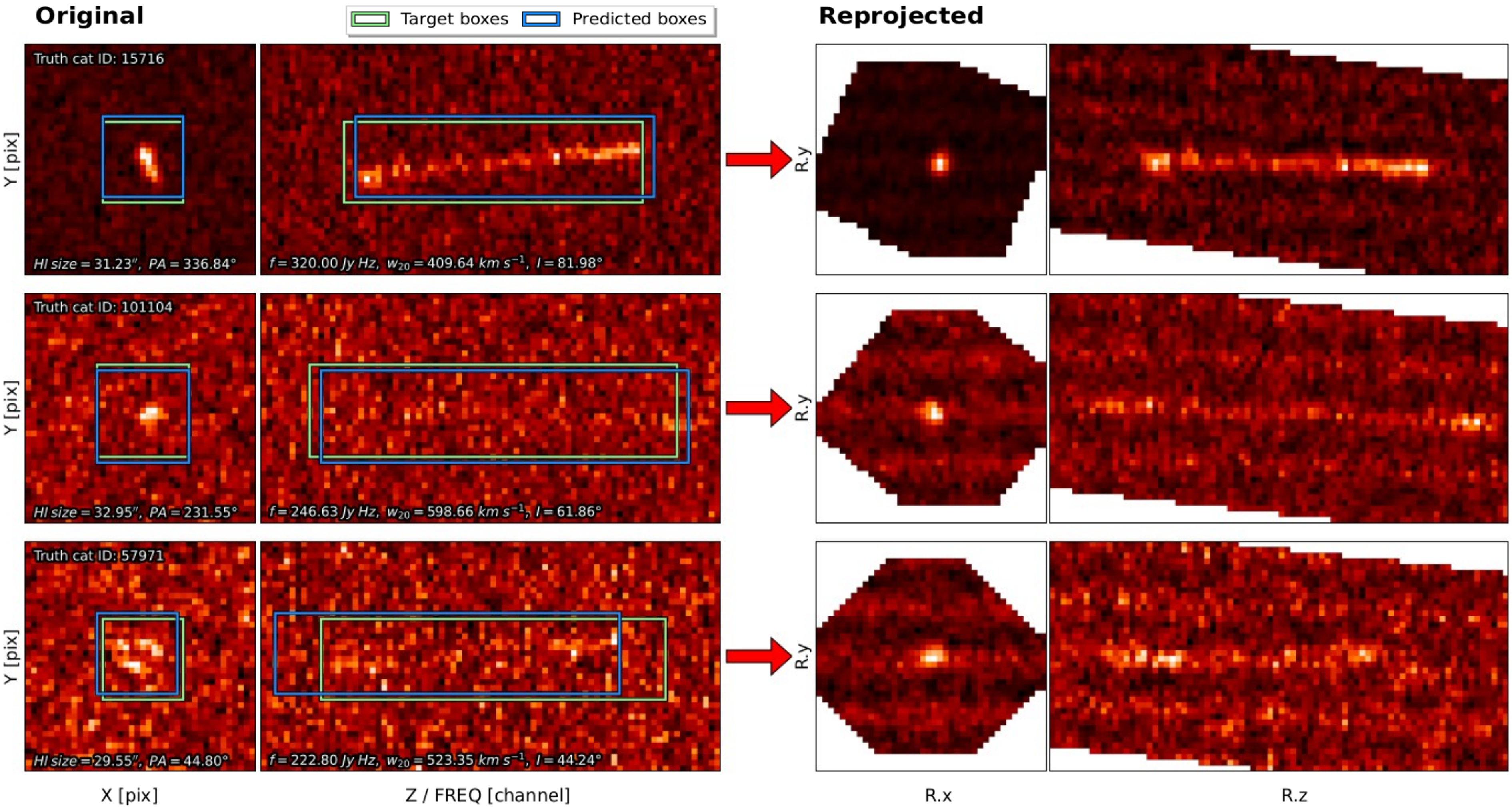
SKA SDC2 took place in 2021. Challenge data are publicly available on the dedicated web-page.

Detection examples



Images are based on 40x40x120 cutouts centered on a source. Signal is averaged over the source dimension in the projected axis.

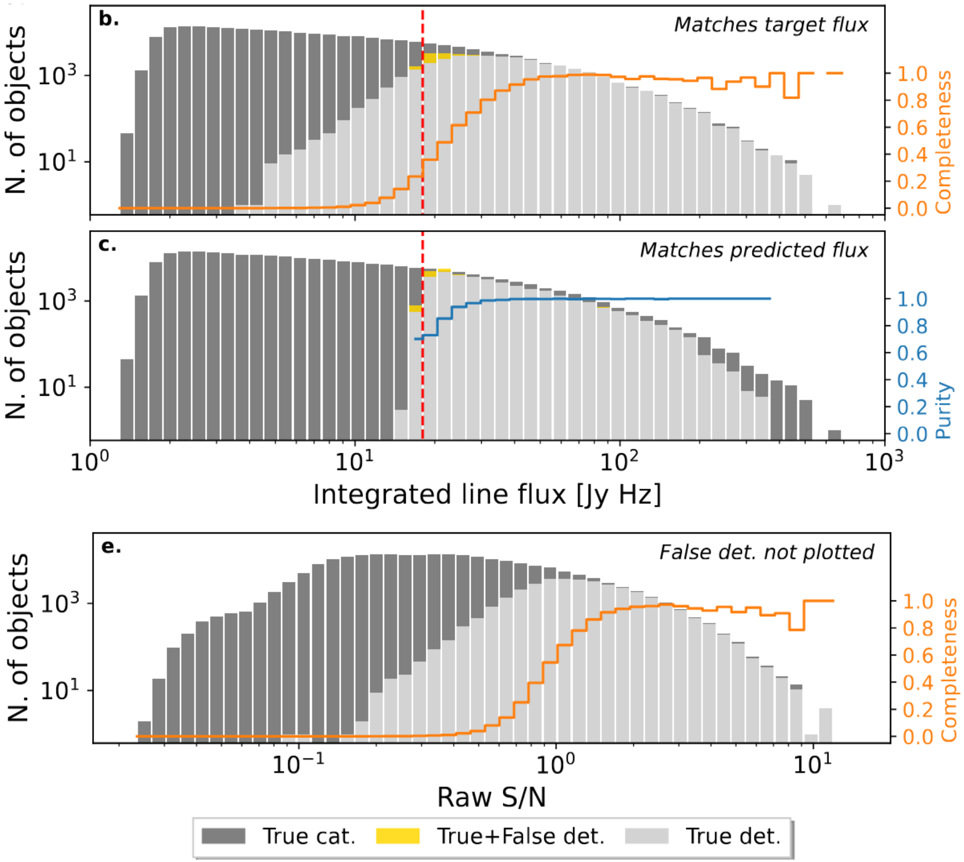
Detection examples



Images are based on 40x40x120 cutouts centered on a source. Signal is averaged over the source dimension in the projected axis.

Detection examples

MINERVA team paper SDC2, YOLO-CIANNA-3D → Cornu et al. 2025, submitted (arXiv:2509.12082)



Comparison to other teams

- **Won the original SDC2**
- The updated version of the method improves our challenge score by 10%
- Highest characterization score

Prediction time for the ~1TB cube ~30min (dominated by data loading)

Using a single RTX 6000 Ada GPU

Generalizing to SKA precursors

LOFAR

ASKAP

MeerKAT

On going application of our method to the LoTSS and RACS surveys

PhD thesis starting next month
Student => **Adam Zarka**

Preliminary work on generalizing to the WALABY and LADUMA surveys

PhD thesis started last spring
Student => **Adrien Anthore**

Main difficulty → building robust training sample for each survey

Collaboration propositions from survey experts are welcome !