

# Identifying Astrophysical Anomalies in the *Hubble* Legacy Archive With *AnomalyMatch*

---

David O’Ryan

Pablo Gómez, Laslo Ruhberg, Pedro Mas Buitrago, Maria Teresa Nardone, Sandor Kruk

SCI-SAS / SCI-EF

10/06/2025

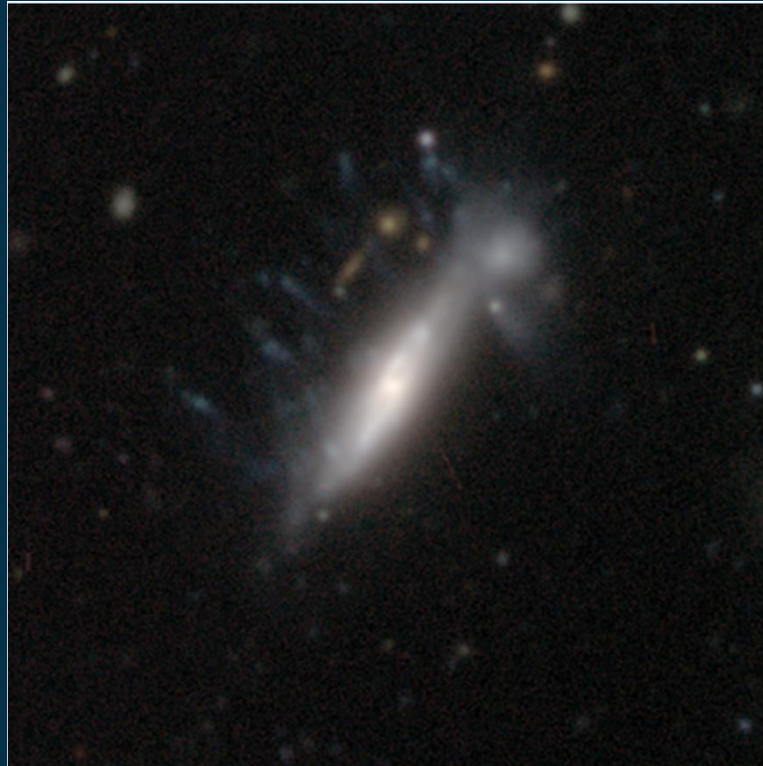
JADES Mini-Symposium

# Astrophysical Anomalies

- Astrophysical sources with morphologies that are rare in large morphology catalogues.



ESA/Hubble: *Antennae Galaxy Merger*



Zooniverse: *Jellyfish Galaxy*



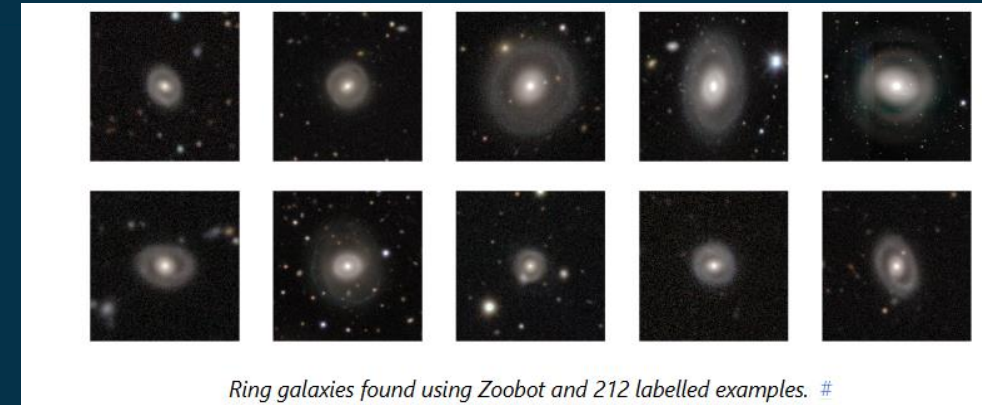
ESA/Hubble: *Gravitational Lens*

Excellent probes of galaxy evolution and cosmology.

# Current Search Efforts

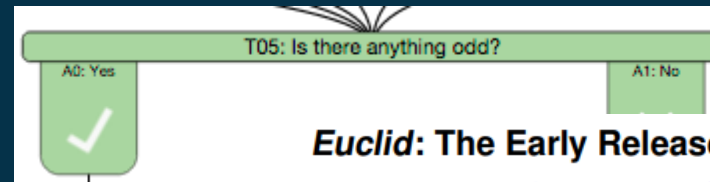
## 1. Foundation models.

- Algorithms such as *Zoobot*.



Ring galaxies found using Zoobot and 212 labelled examples. #

## 2. Citizen Scientists / Expert Labelling.

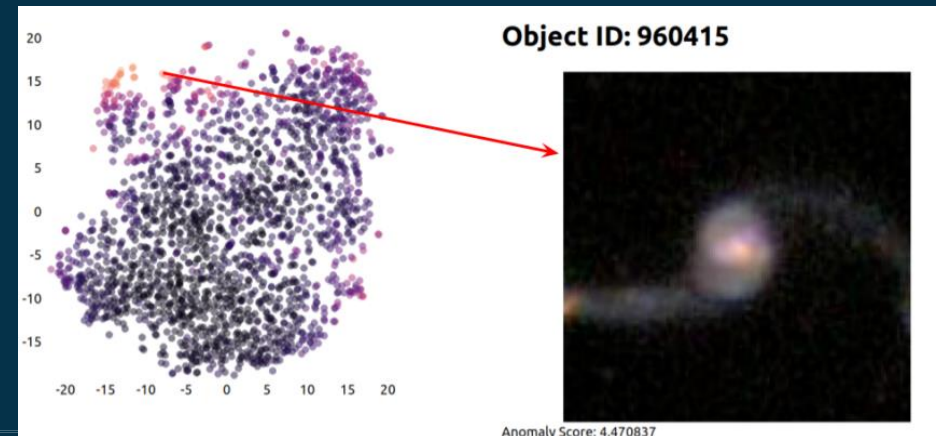


### Euclid: The Early Release Observations Lens Search Experiment★

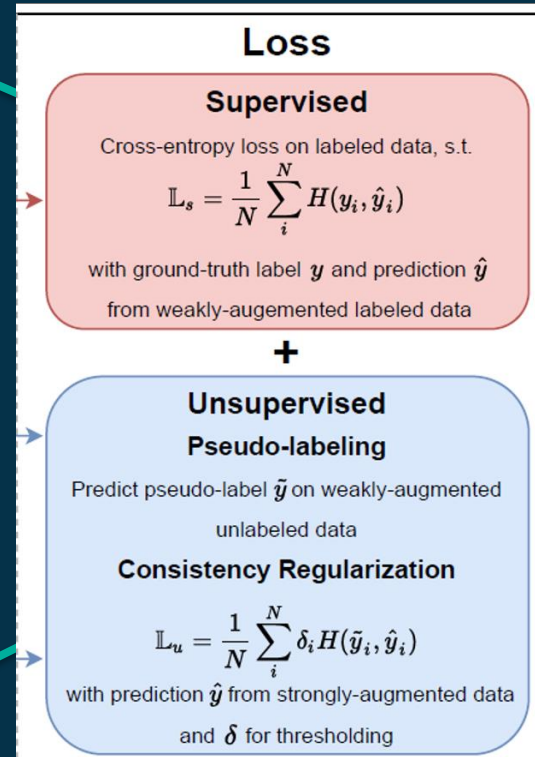
J. A. Acevedo Barroso<sup>\*\*1</sup>, C. M. O’Riordan<sup>2</sup>, B. Clément<sup>1,3</sup>, C. Tortora<sup>4</sup>, T. E. Collett<sup>5</sup>, F. Courbin<sup>1,6,7</sup>, R. Gavazzi<sup>8,9</sup>, P. B. Metcalfe<sup>10,11</sup>, V. Rucillo<sup>4,12,13</sup>, I. T. Andika<sup>14,2</sup>, R. Cabanac<sup>15</sup>, H. M. Courtois<sup>16</sup>, J. Crook-Mansour<sup>17</sup>

## 3. Novel software methods.

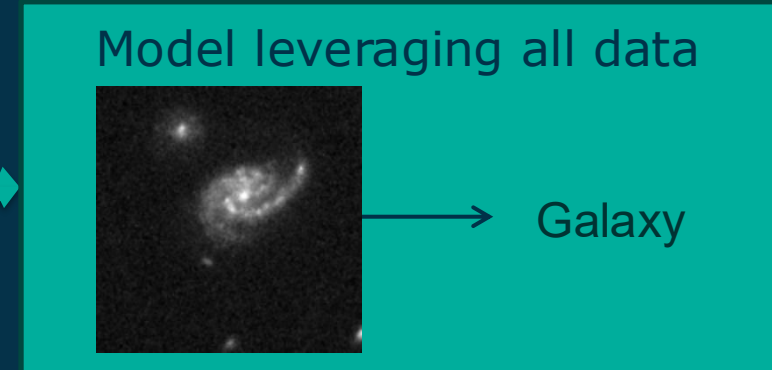
- Astronomy – exploring feature spaces.



# AnomalyMatch: Semi-Supervised Learning



+  
AL



Result: need a tiny (< 10) labelled anomaly set to find more.

# An Intuitive UI

Save Model | Load Model | Save Labels | Load Top Files

Most Anomalous | Most Nominal | Closest to Mean | Closest to Median

Filename: 4001012821032.jpeg | Score: 1.0000 | Index: 0 / 9197 | Label: None

Top 10 Mispredicted Test Images | Acc: 90.9% | AUROC: 1.000 | AUPRC: 1.000

4000956847935.jpeg | Pred: 0.26 | Label: 1.00  
6000412898716.jpeg | Pred: 0.63 | Label: 1.00  
4236963598.jpeg | Pred: 0.04 | Label: 0.00  
4000737238068.jpeg | Pred: 0.00 | Label: 0.00  
4000862191589.jpeg | Pred: 1.00 | Label: 1.00  
4001136568575.jpeg | Pred: 0.00 | Label: 0.00  
6000405941848.jpeg | Pred: 0.00 | Label: 0.00  
4001035733503.jpeg | Pred: 0.00 | Label: 0.00  
4001220646553.jpeg | Pred: 0.00 | Label: 0.00  
4001031017348.jpeg | Pred: 0.00 | Label: 0.00

ROC Curve | Precision-Recall Curve

True Positive Rate | False Positive Rate | ROC (AUC=1.000)

Precision | Recall | PRC (AUC=1.000)

Invert Image | Restore | Apply Unsharp Mask | Remember

Brightness: 1.00 | Contrast: 1.00 | RGB:  R  G  B

Previous | Anomalous | Nominal | Next

Train Iterations: 500 | Batch Size: 10000

Reset Model | Next Batch | Train | Evaluate Search Dir

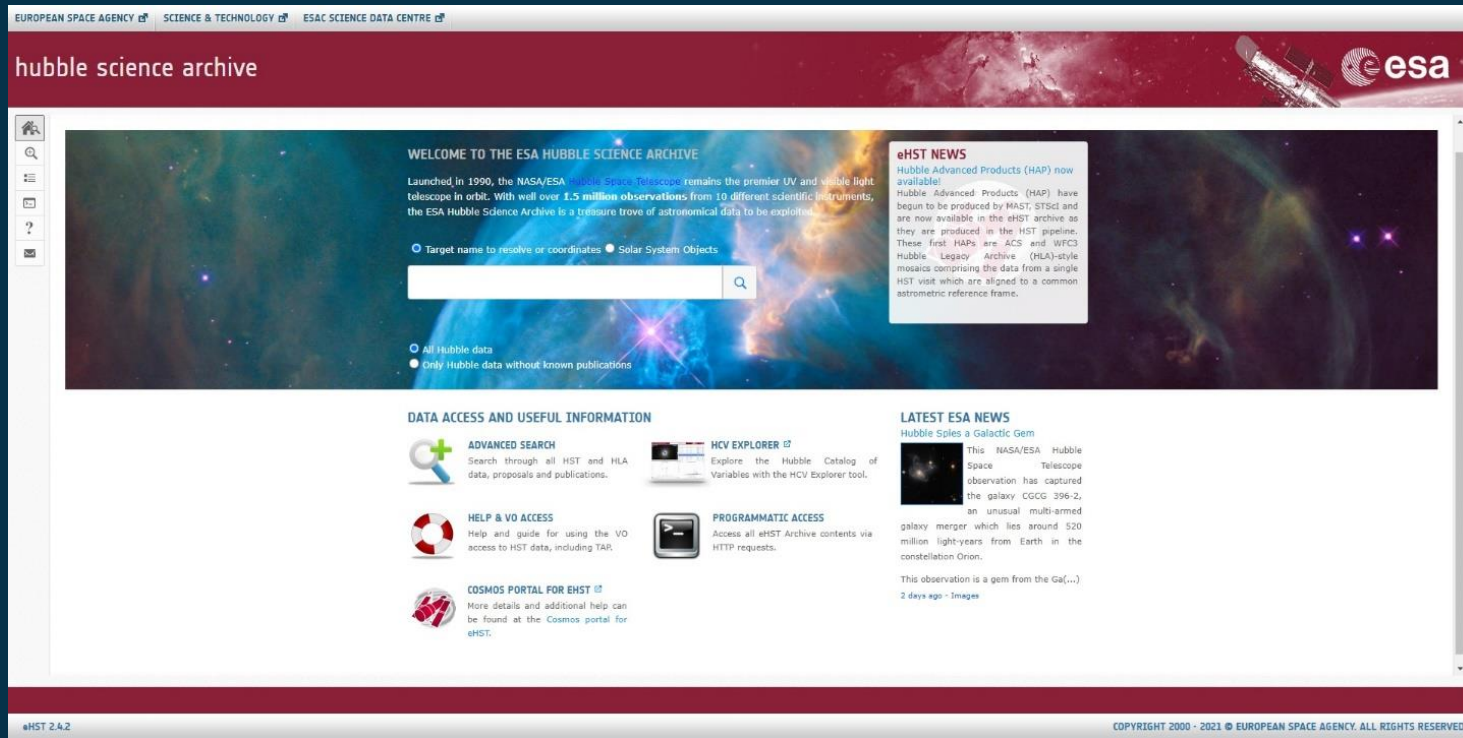
Version: 1.0.0

Memory: 10.12 GB | GPU: 0.05 GB [12:50:43]

Training Complete. Eval Acc: 90.91%

- Fully available as its own python package + example Jupyter Notebooks.

# Applying to the HLA

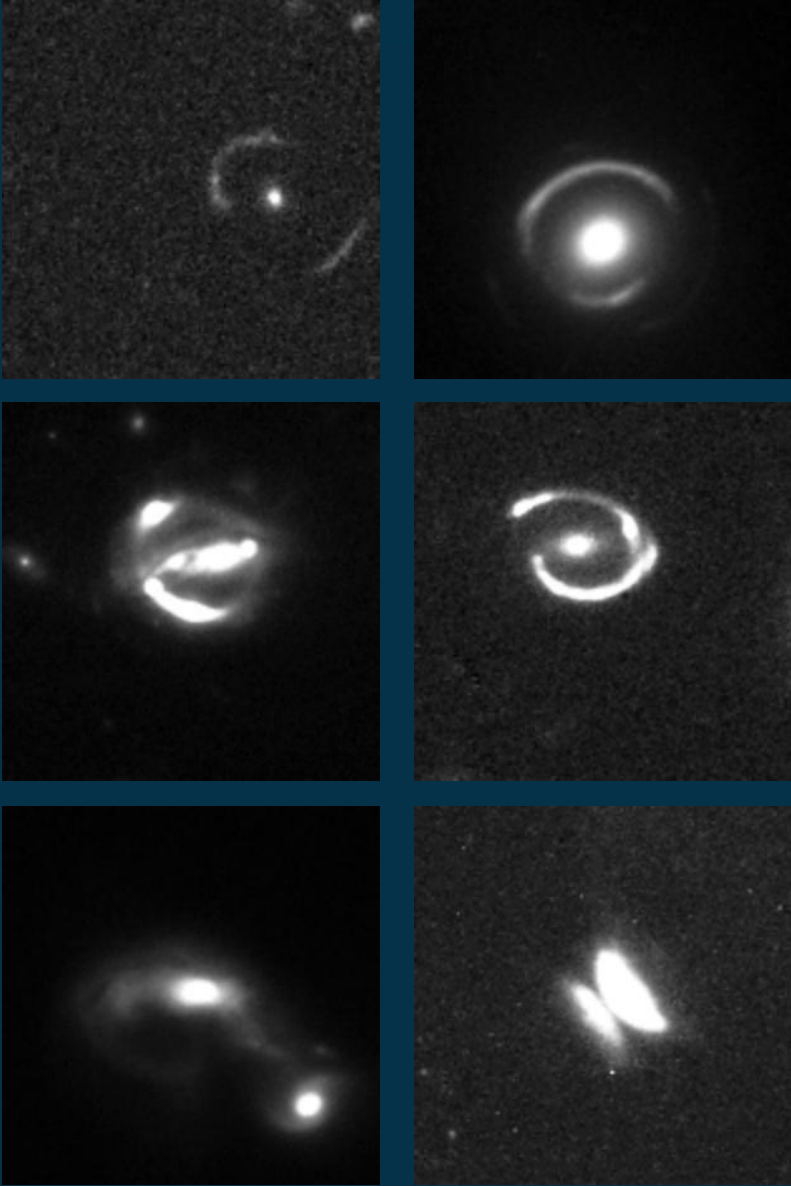


**ESA DATALABS**

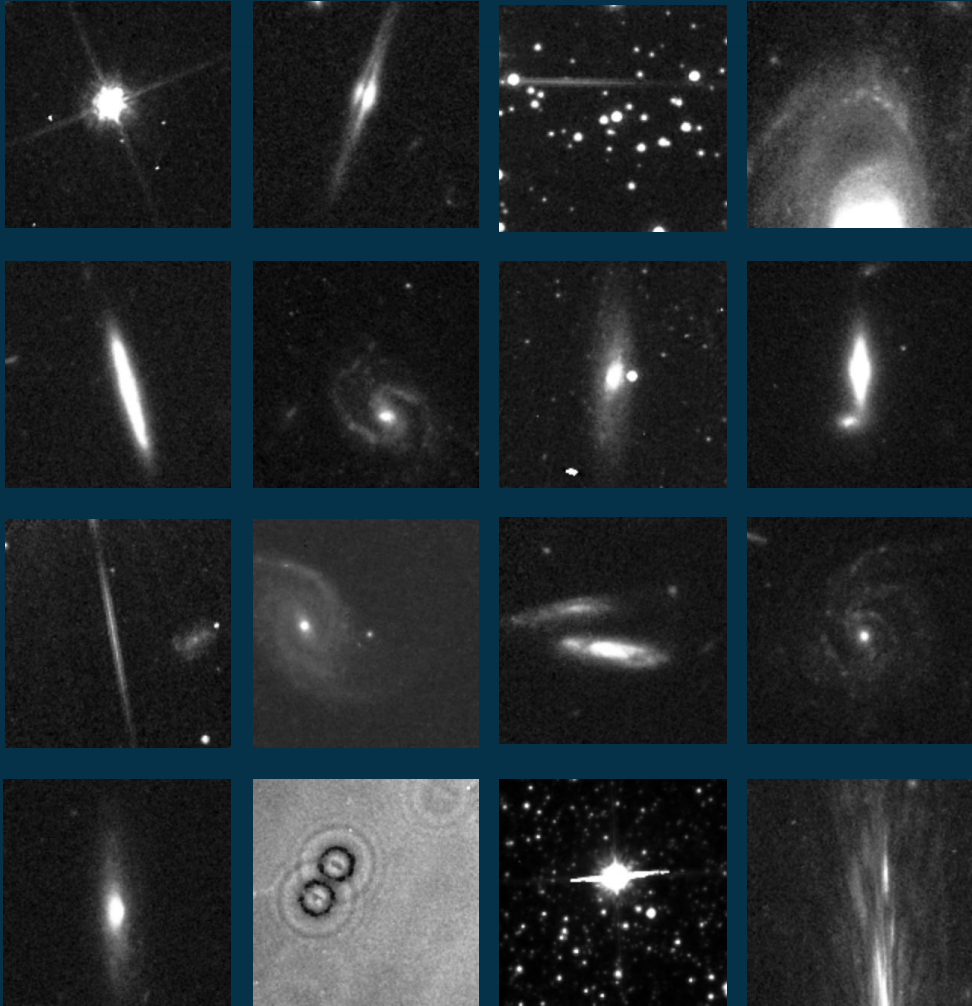
- Contains all observations from *HST*.
  - All filters, from raw data to science mosaics, all instruments.
- Select *F814W* filters, only final product data.

~99.6 million sources.

# A Small Training Set

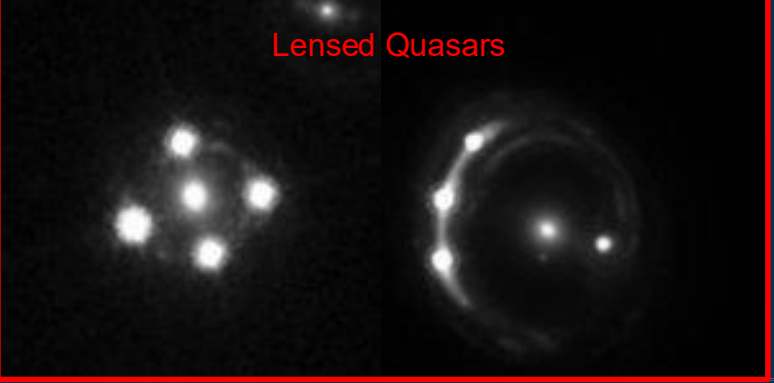
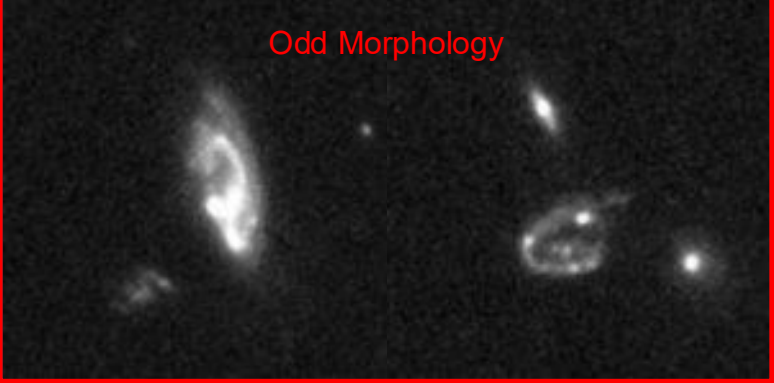
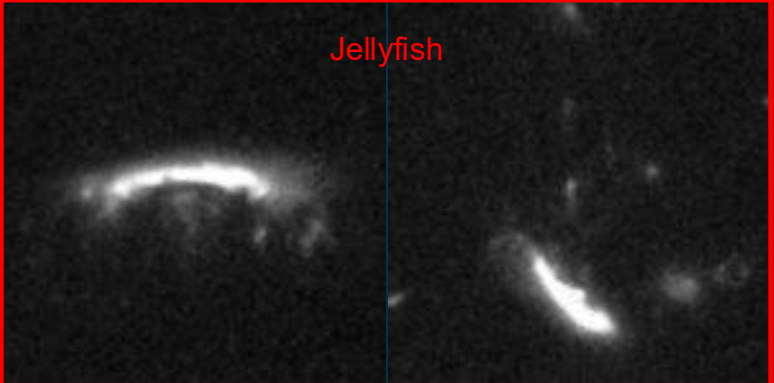
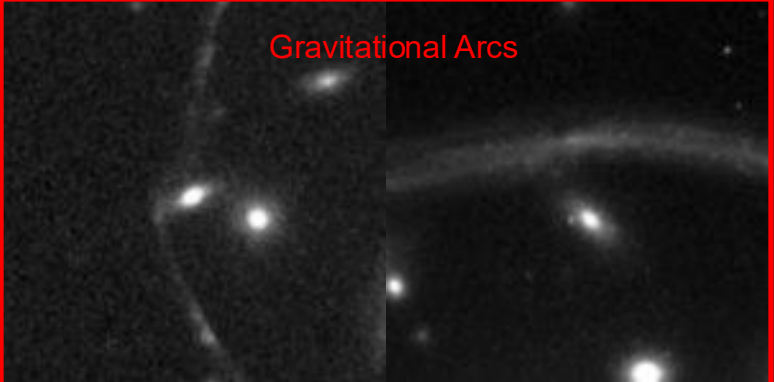
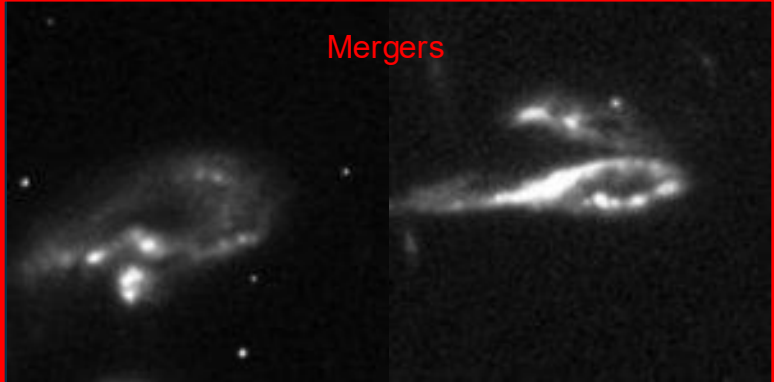


Anomalous Images ~160



Nominal Images ~1000

# Trawling *Hubble* Legacy Archive



# Trawling the *HLA*: Table

**Table 1.** Breakdown of anomalies found in the development of the *AnomalyMatch* algorithm.

Classification	N Found	N Referenced
Merging Galaxy	629	212
Gravitational Lenses	140	54
Odd Galaxy	164	60
Nominal Galaxy	163	60
Unknown Morphology	43	0
High Redshift Galaxy	28	7
Jellyfish Galaxy	35	17
Overlapping Galaxy	39	26
Gravitational Arc	39	19
Clumpy Galaxy	11	5
Galaxy Hosting a Jet	13	5
Galaxy Hosting AGN	8	4
Ring Galaxy	12	5
Lensed Quasar	5	5
Collisional Ring Galaxy	2	0
Edge-On Protoplanetary Disk	2	2
Galaxy Hosting Supernova	2	2
Submillimetre Galaxy	1	1

Full paper on ArXiv:  
O’Ryan & Gómez 2025 (2505.03508)

*AnomalyMatch* description paper:  
Gómez et al. 2025 (2505.03509)

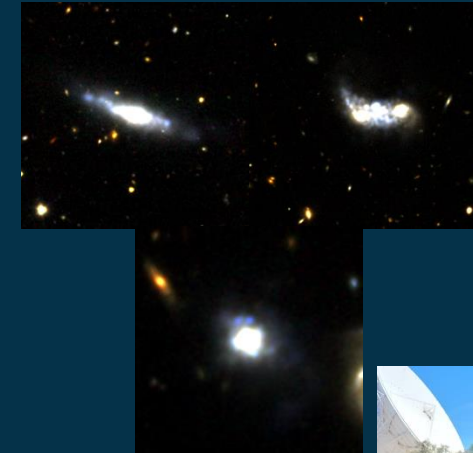
# Applications to Euclid Q1

- Searching for jellyfish galaxies.
- Measure jellyfish properties vs cluster & field galaxies.
- Searching for merging galaxies.
- Aim: measure merger rates & benchmark *AnomalyMatch*.



Euclid Q1 Data

IllustrisTNG Mock Data



- Both projects showing excellent progress with objects of interest being discovered.





# Conclusions & Future Work

- Current work on anomaly detection is plagued by many issues.
  - Limited examples for training.
  - Subjectivity.
  - Previous knowledge of morphology required.
- *AnomalyMatch* addresses this with semi-supervised methods to leverage both labelled and unlabelled data.
  - ‘Plug & Play’ model – fully integrated into ESA Datalabs.
  - Updates also coming.

- Initial tests on *HLA* show potential.

- Next step: *Euclid* and *JWST*.

Thank you for listening!

Contact:

David O’Ryan: [David.Oryan@esa.int](mailto:David.Oryan@esa.int)

Pablo Gómez: [Pablo.Gomez@esa.int](mailto:Pablo.Gomez@esa.int)

Sandor Kruk: [Sandor.Kruk@esa.int](mailto:Sandor.Kruk@esa.int)

Jan Reerink: [Jan.Reerink@esa.int](mailto:Jan.Reerink@esa.int)





# Benchmarking

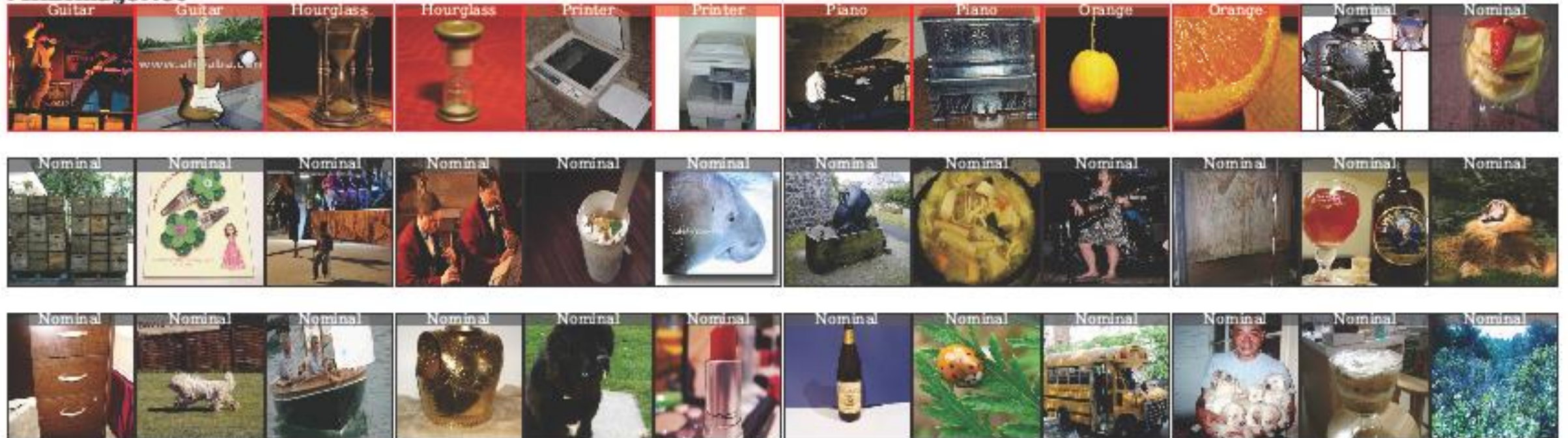
- Use MiniImageNet and GalaxyMNIST.

## GalaxyMNIST

Anomaly Classes in Red

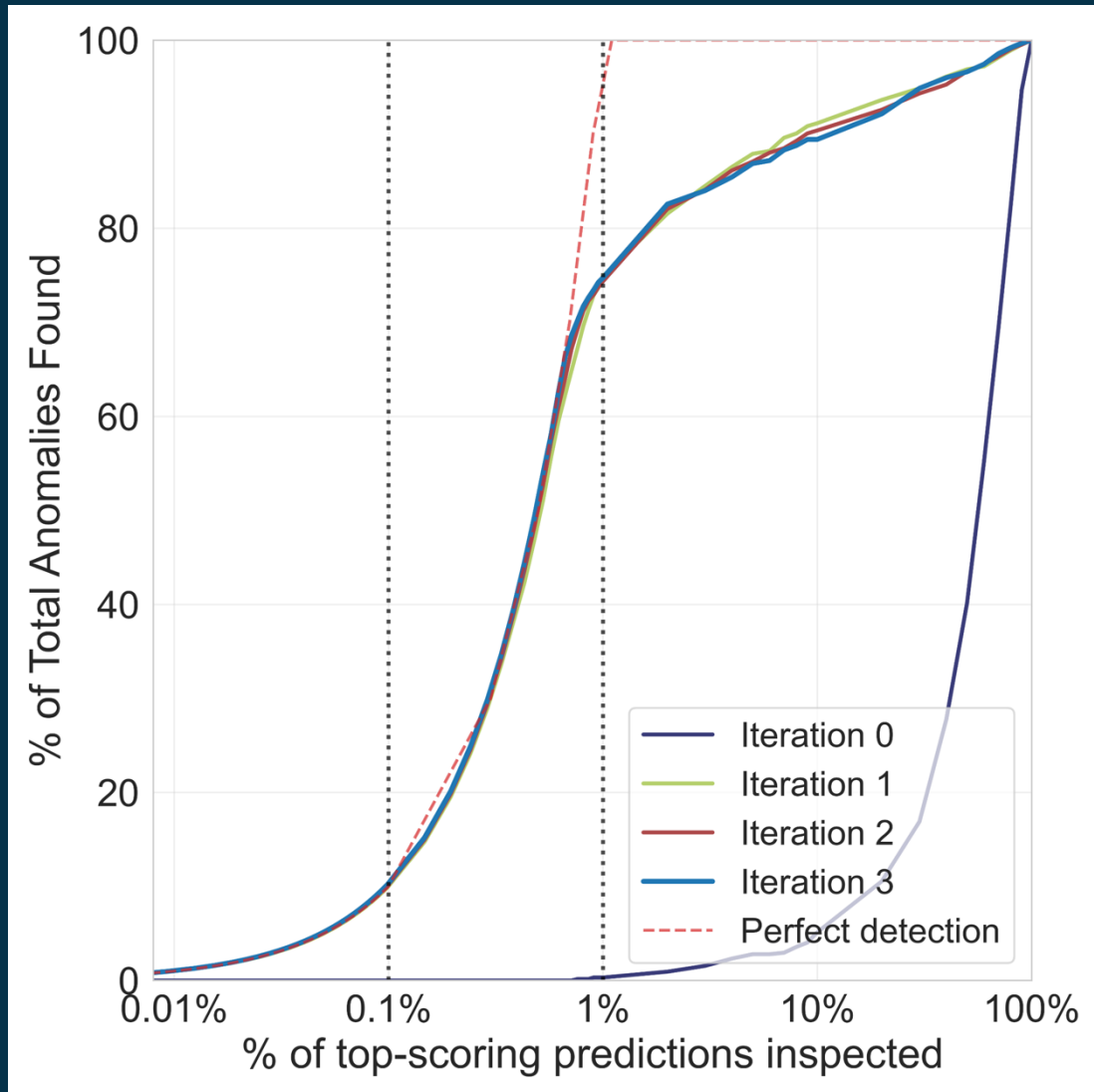


## MiniImageNet

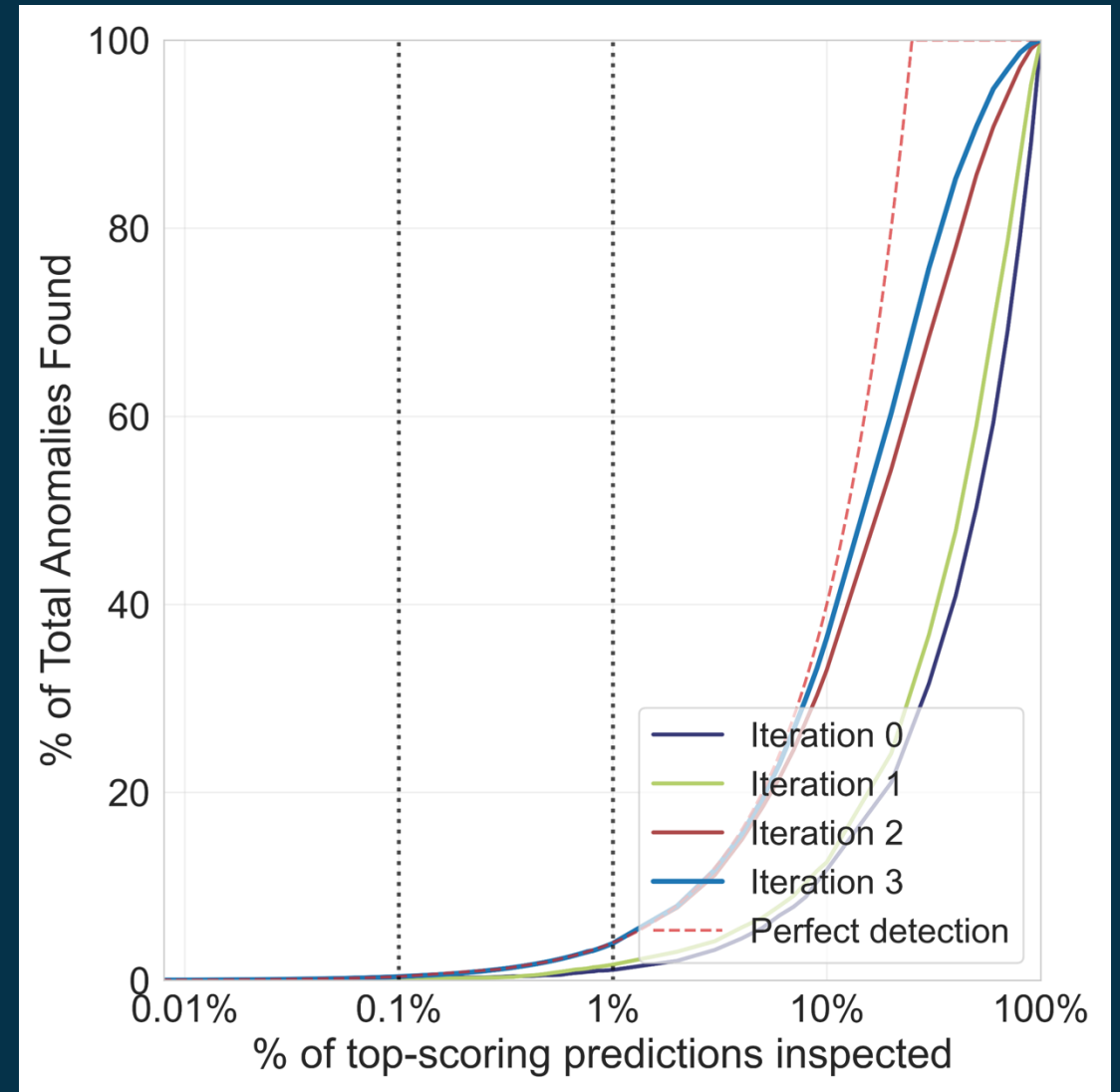


# Benchmarking

## Mini-ImageNet



## GalaxyMNIST



MiniImageNet detection performance: hourglass.

GalaxyMNIST detection performance: unbarred spiral.

