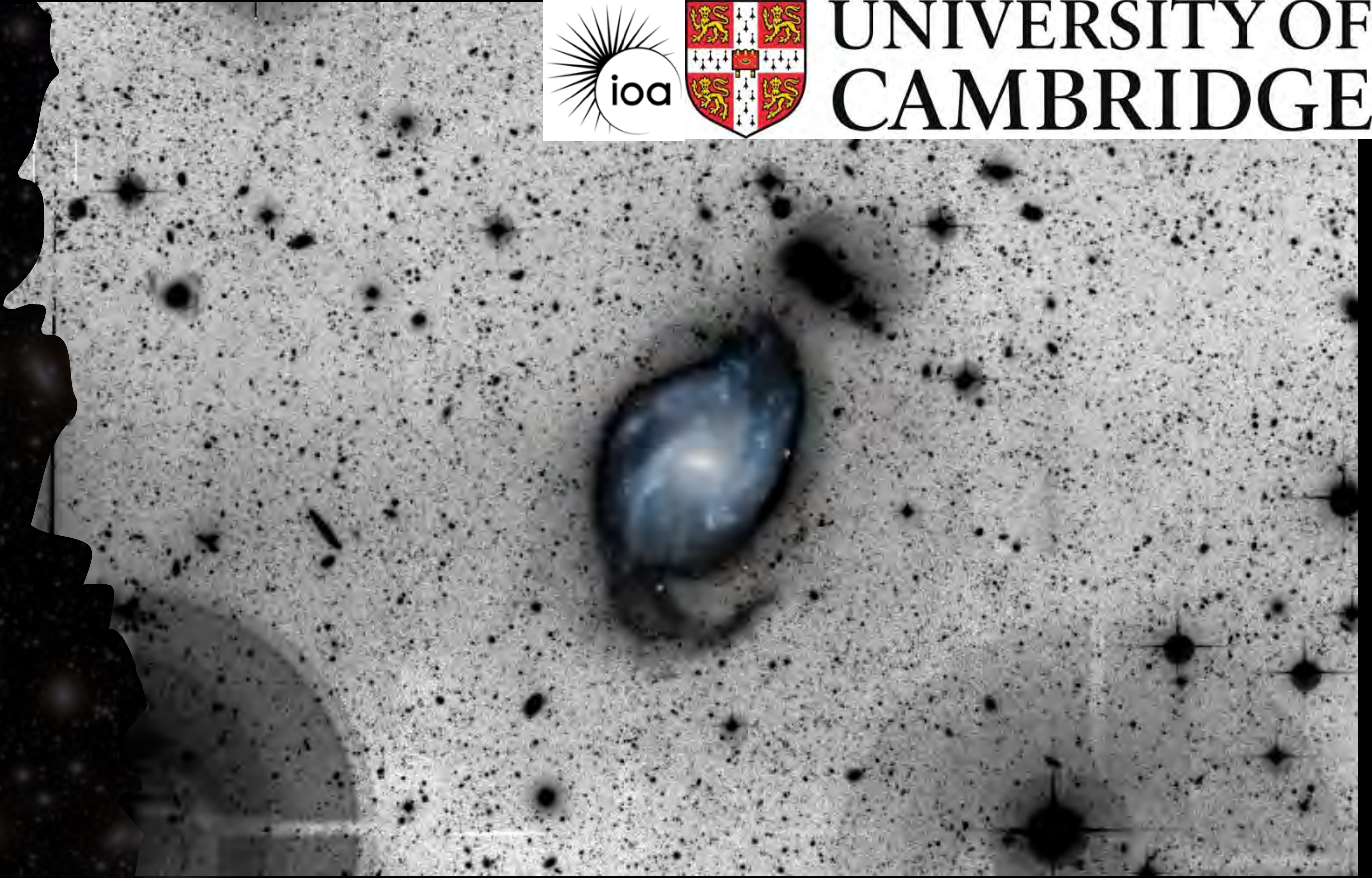
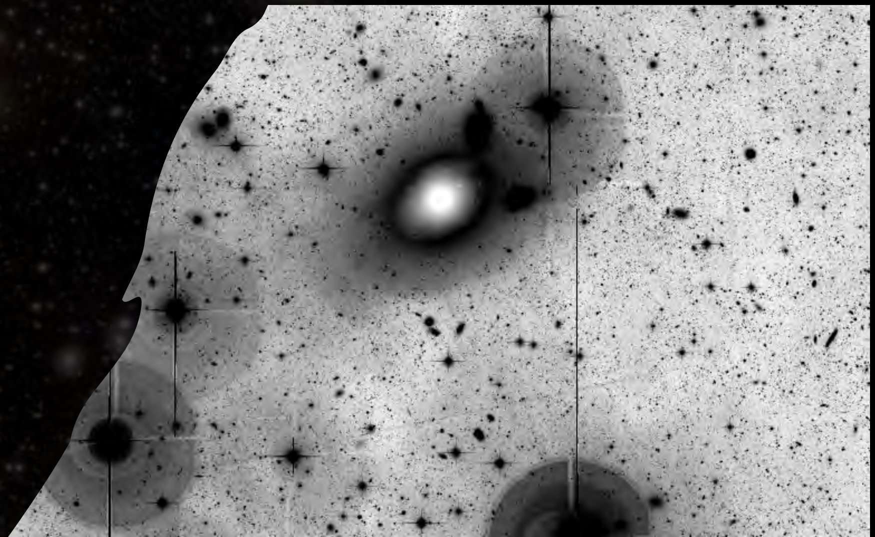
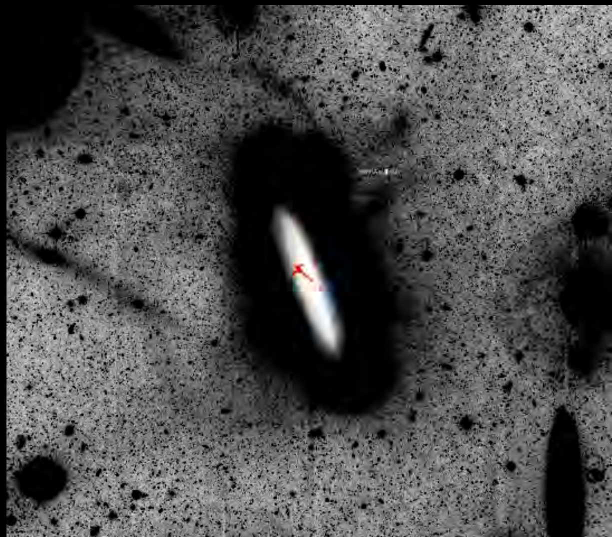
DIS MPhil : Lecture 13

LSB features and galactic evolution







Dr Elisabeth Sola
es2074@cam.ac.uk, IoA, Hoyle 34

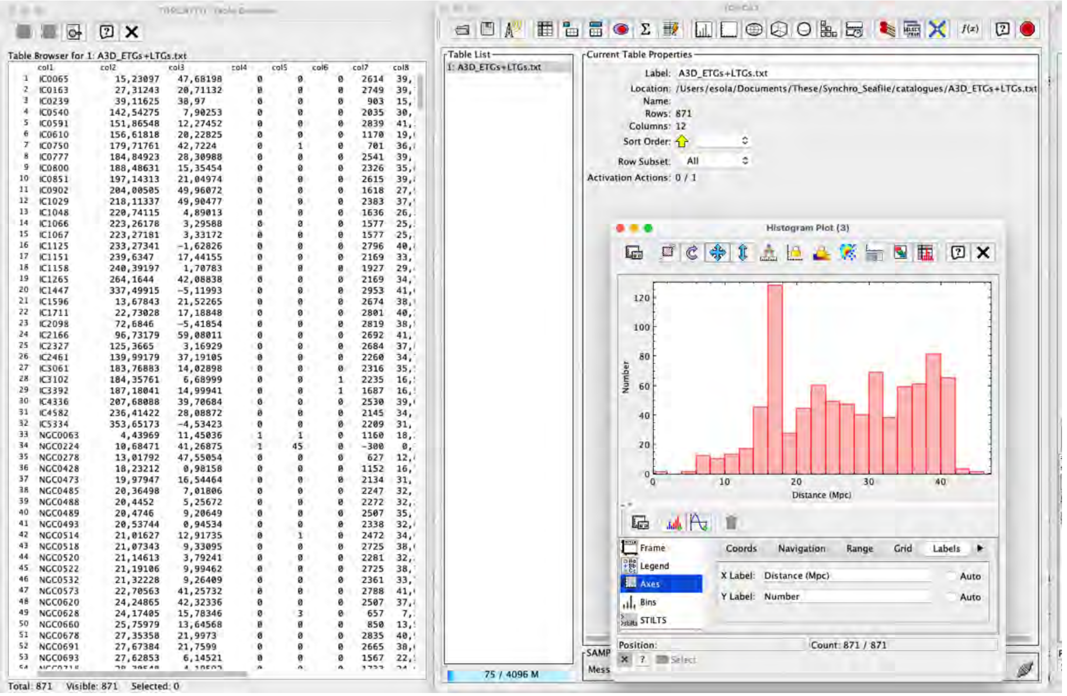
Outline of this block

Block C: Low Surface Brightness Universe & image analysis

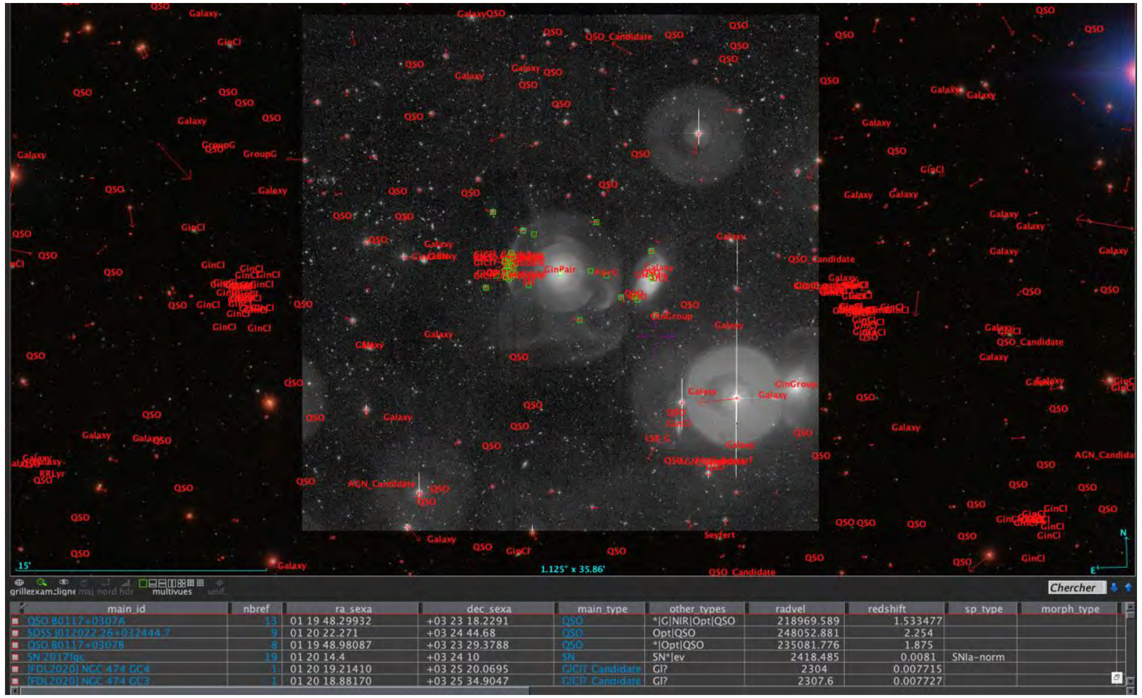
1. **28/02/2025: Galaxies in the Local Universe:** Introduction to models of galaxy formation and evolution. Galaxy mergers and tidal features. Tidal features in the Local Group and beyond.
2. **07/03/2025: Exploring the LSB Universe**: Deep astronomical imaging and data processing techniques. The LSB Universe. Hunting for tidal features in the LSB realm.
3. **07/03/2025: LSB features and galactic evolution**: LSB features as probes of galactic evolution. Ongoing and future surveys. Machine learning for LSB features.
4. **11/03/2025: Hands-on session**: explore astronomical images and catalogues with Aladin, DS9, Topcat, Python

Recap of Lecture 12

- Astronomical images: data + metadata (header)
- Many tools to manipulate FITS files (e.g., DS9) and catalogues (e.g., TOPCAT)
- Many databases of catalogues and images are publicly available online
- Cross-matching catalogues must be done carefully



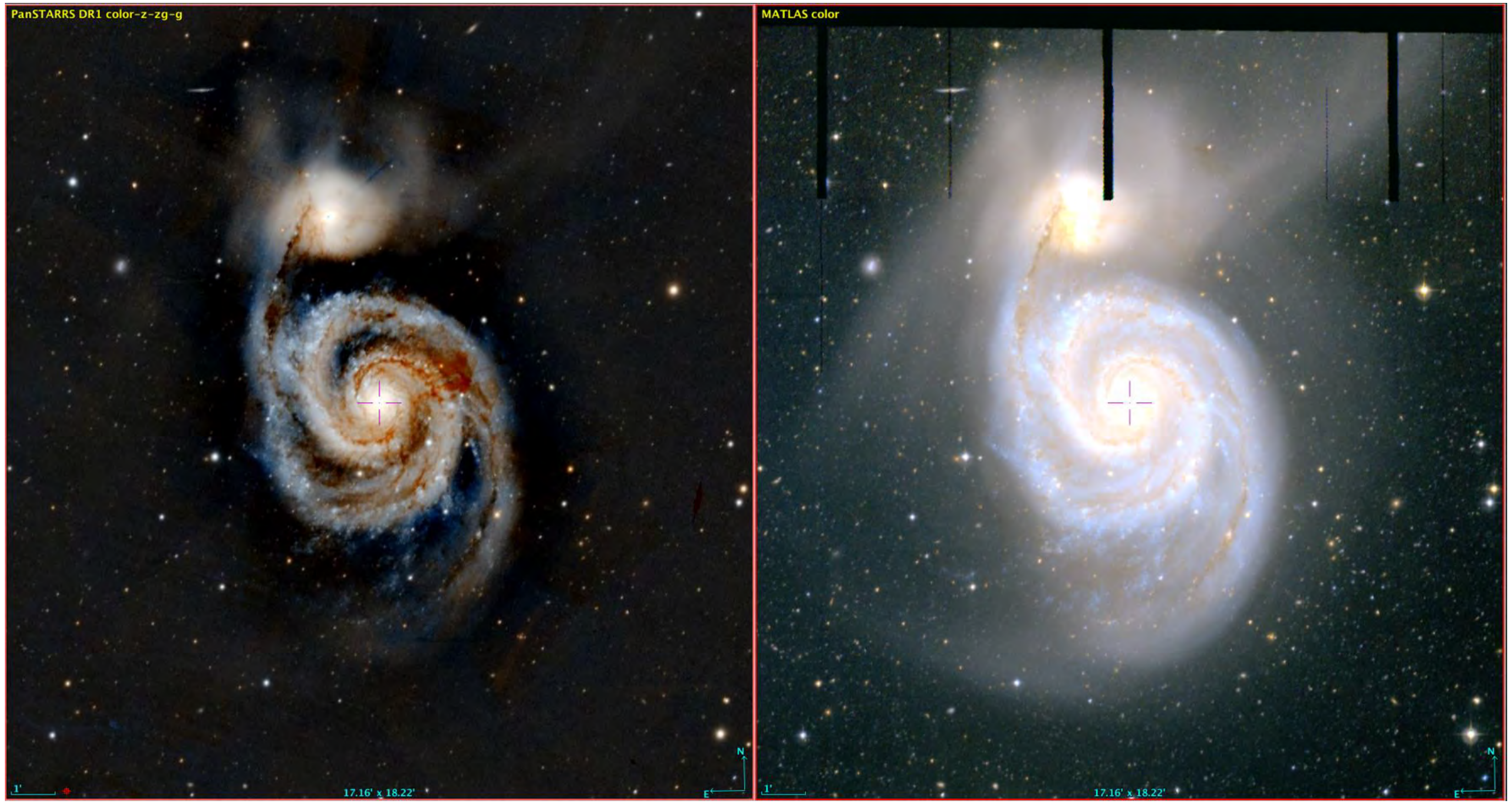
Topcat



Aladin

Recap of Lecture 12

- Careful reduction is needed to preserve the LSB signal

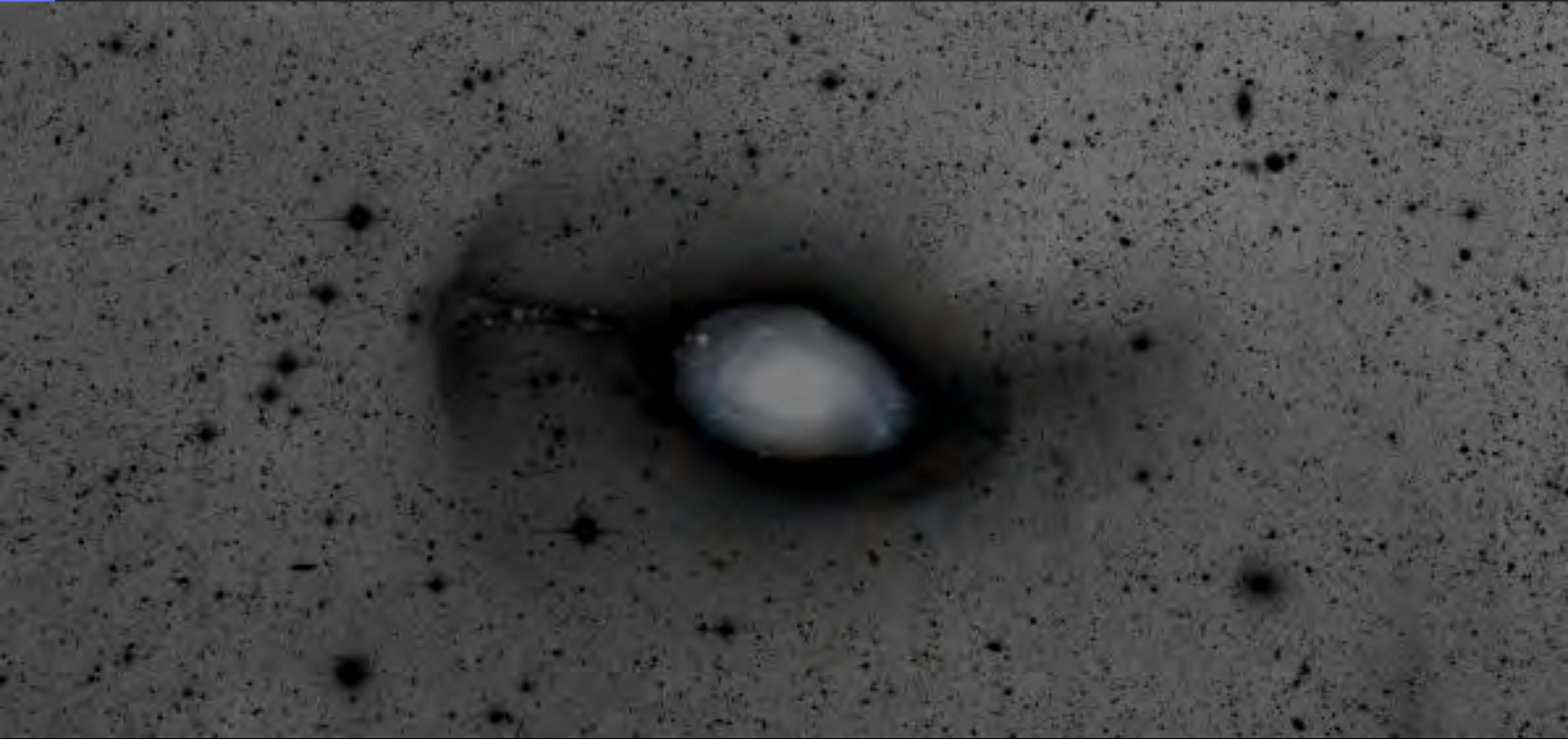


- Identification of tidal features is mostly performed visually… but this will no longer be the case for large surveys to come !

Outline Lecture 13

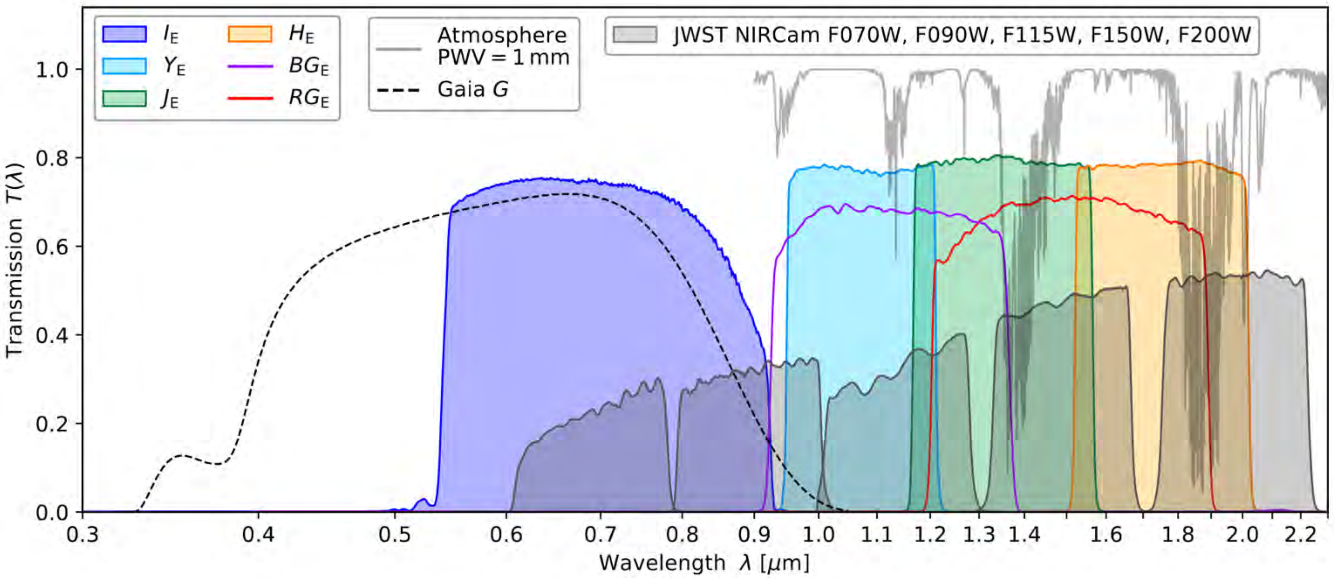
- Euclid and future surveys
- Machine learning and the faint Universe
- Tidal feature and galactic evolution

Euclid and future large surveys

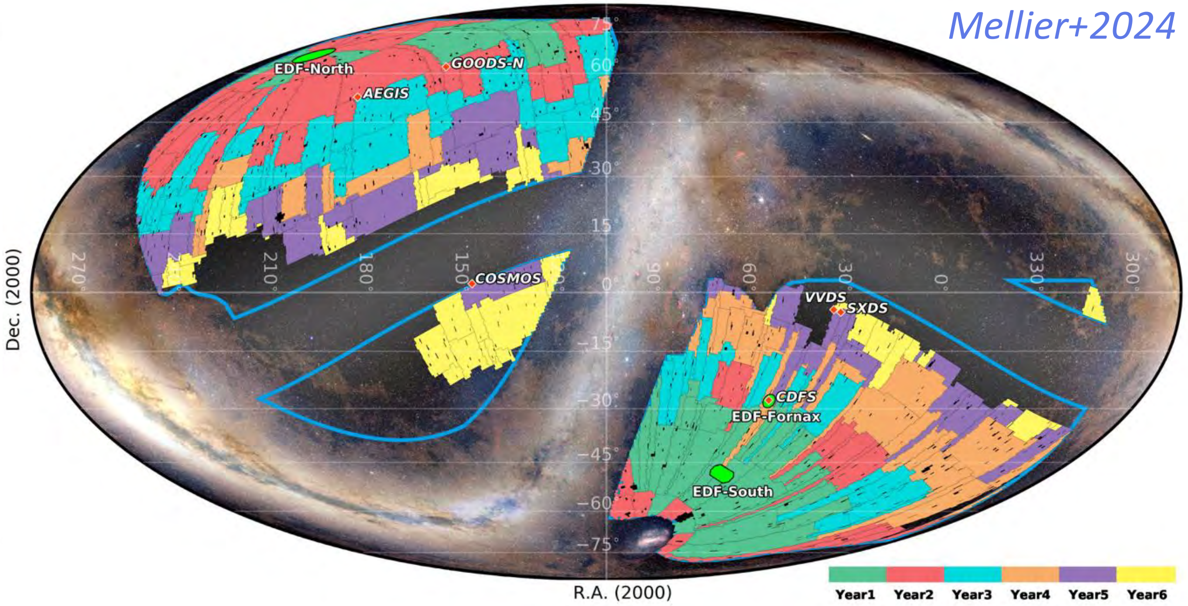


Euclid

- Euclid: unprecedented combination of:
- Large spatial **coverage** (15,000 deg^2) : many different environments probed
- Spatial **resolution** (0.1'' and 0.3'') **+** wide **FoV** (0.57 deg^2) + great PSF stability: possible to have distance estimation (GCs and SBF)
- **Depth** (29.5 mag/arcsec^2, *Borlaff+2022*)
- **Colour** : VIS + NIR (Y, J, H): constraining stellar populations



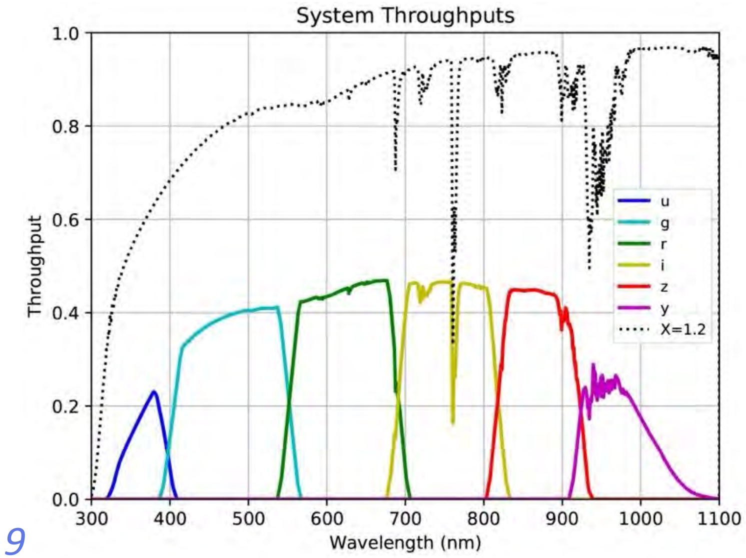
Mellier+2024



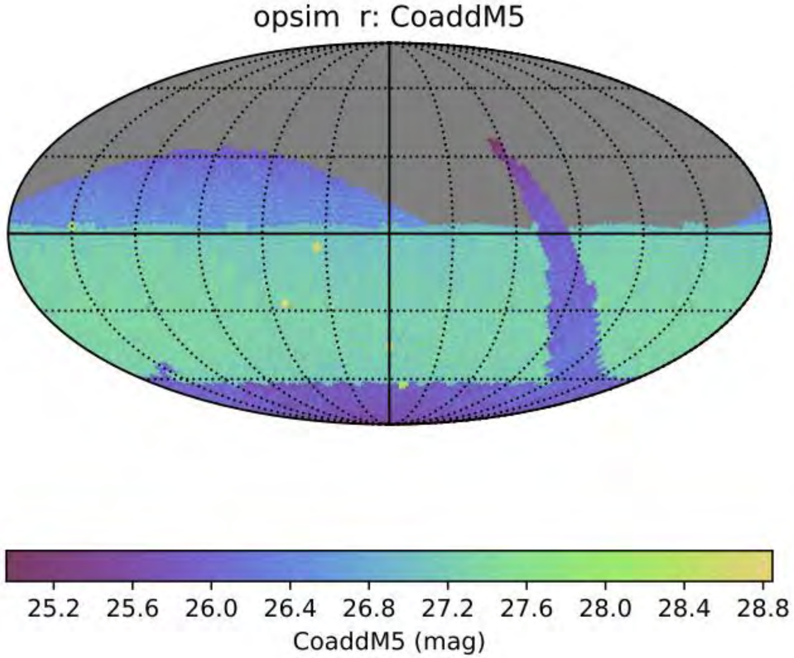
Rubin

- Rubin will cover **18,000 deg^2** of the Southern hemisphere with very deep imaging (30.5 mag/arcsec^2 after 10 years)
- Spatial resolution: 0.2’’/pixel
- **Multi-bands**: u,g,r,i,z,y
- Complementary to Euclid

© Rubin

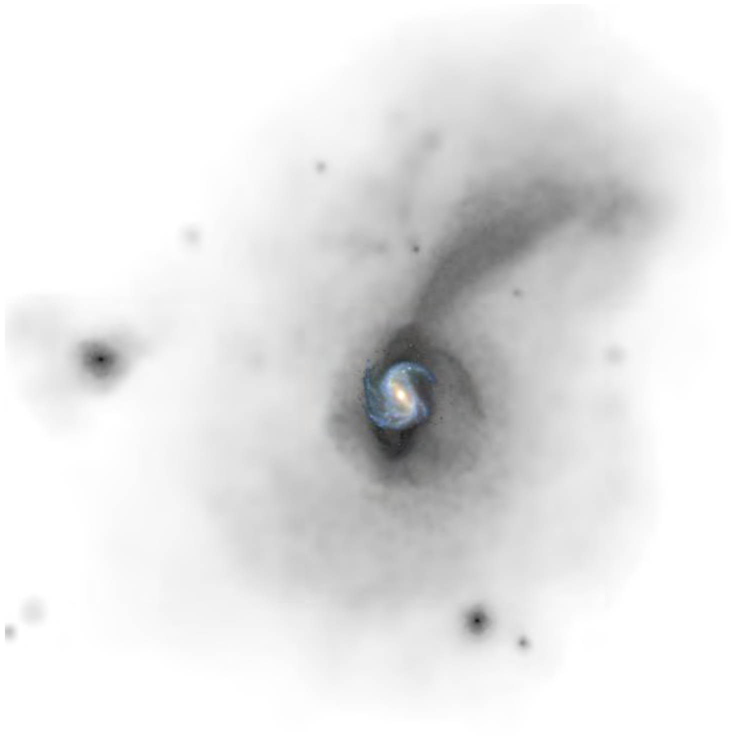


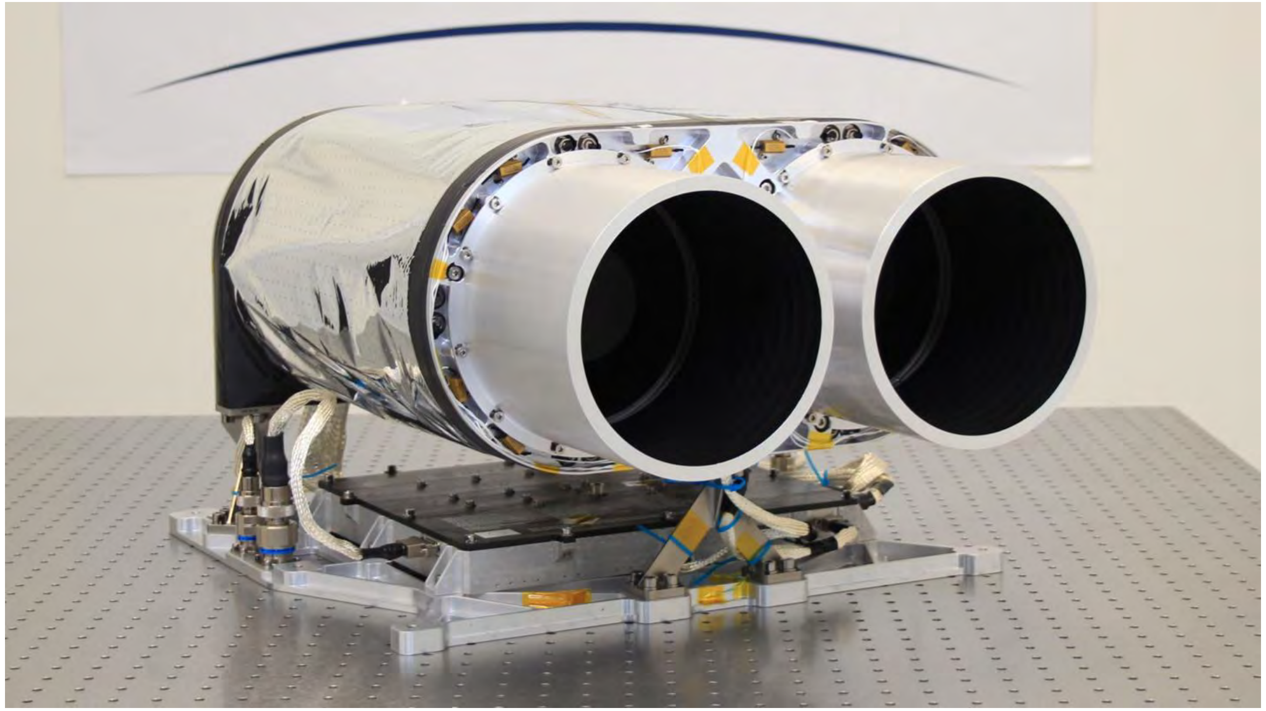
Ivezić+2019



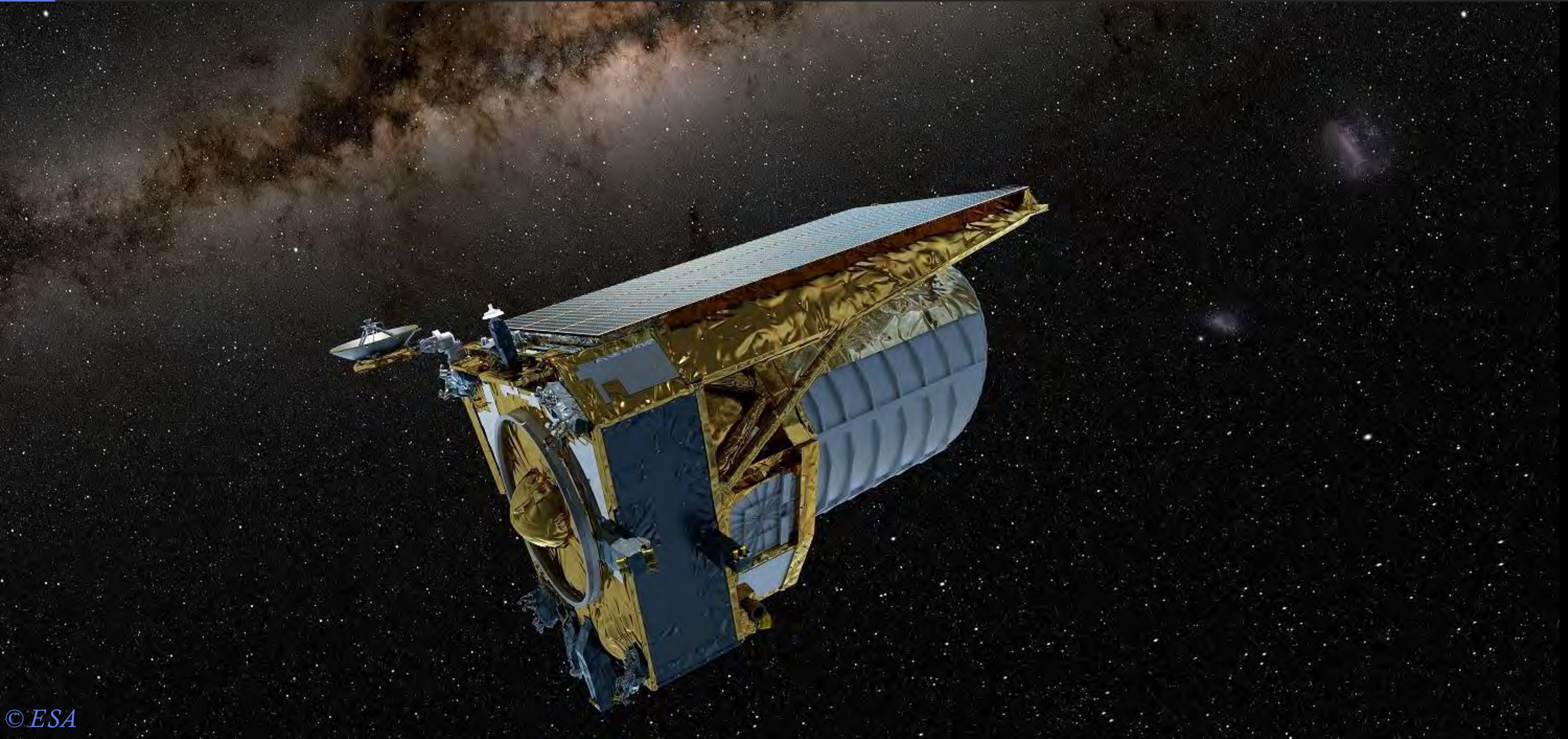
ARRAKIHS

- **ARRAKIHS** : ESA's next F-class ('Fast') mission ~2030s, probing the dark matter in the Universe by studying faint, **LSB haloes and tidal features** of Milky Way-like galaxies.
- Multi-band, very deep (below 30 mag/arcsec^2), targeting nearby spirals





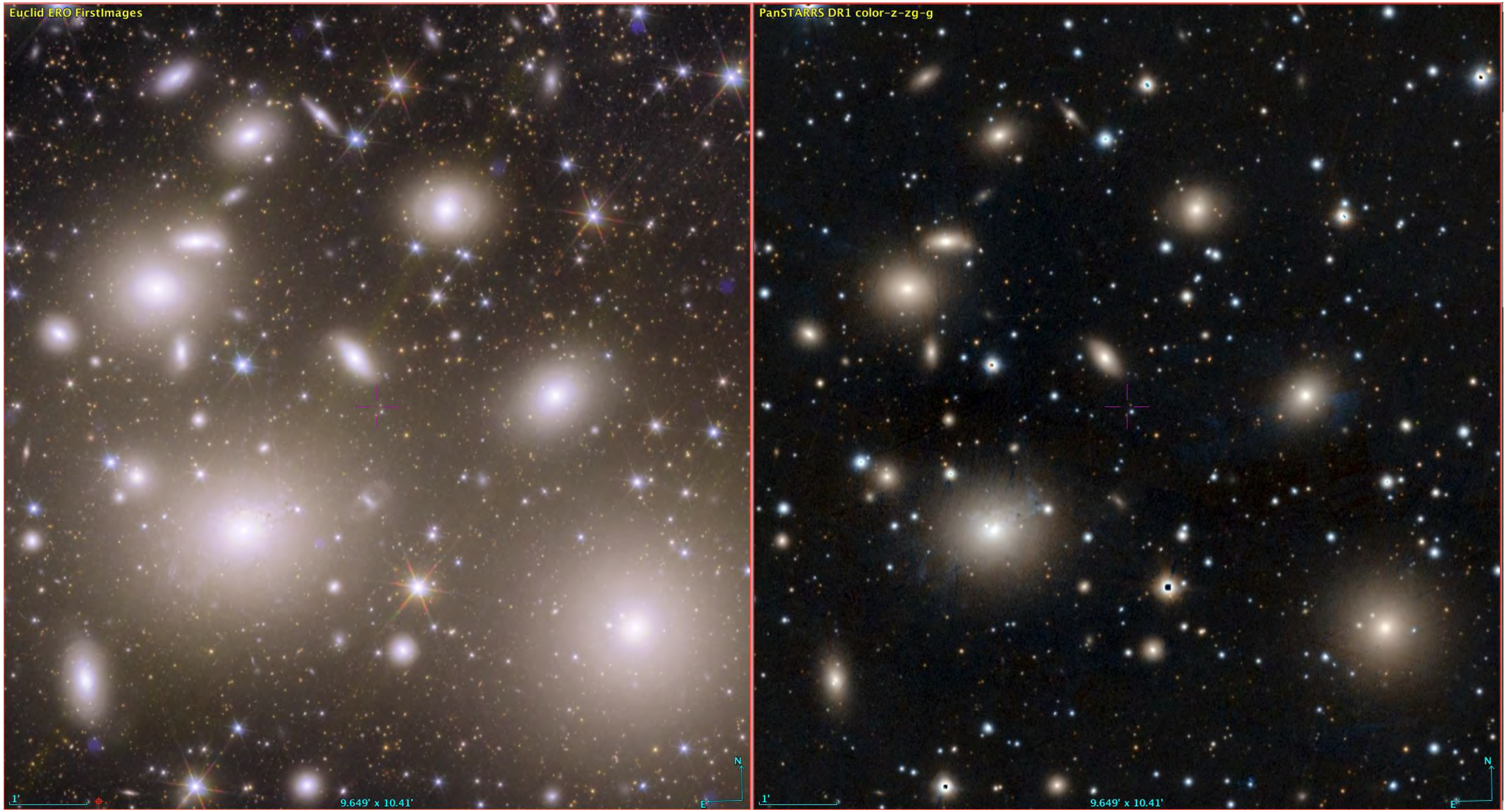
Euclid's view of the LSB Universe



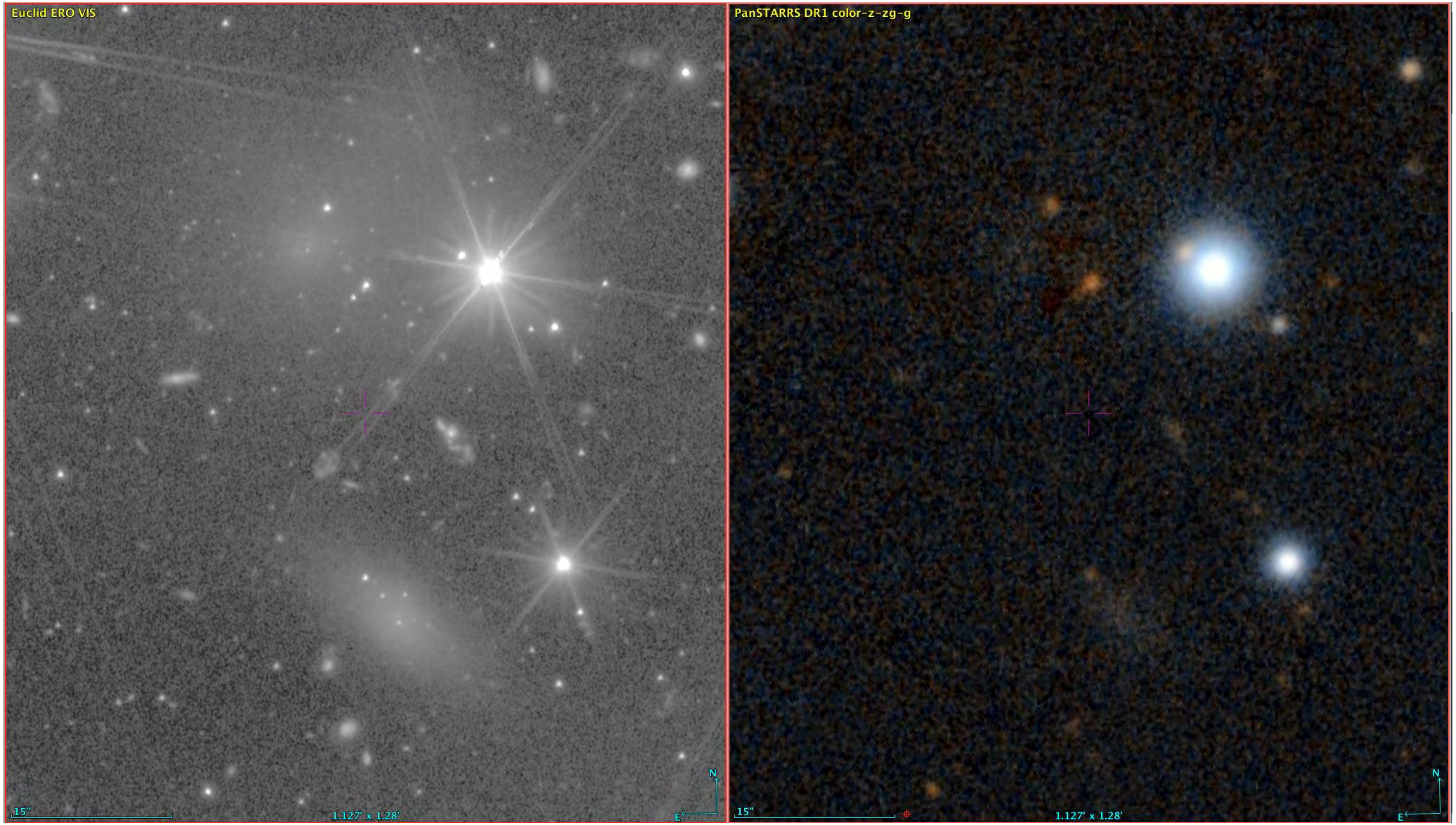
Euclid Early Release Observations (ERO)

- ERO: first images showcasing the depth and diversity of science Euclid will provide
- 17 targets, including star-forming regions, globular clusters, nearby galaxies and galaxy clusters

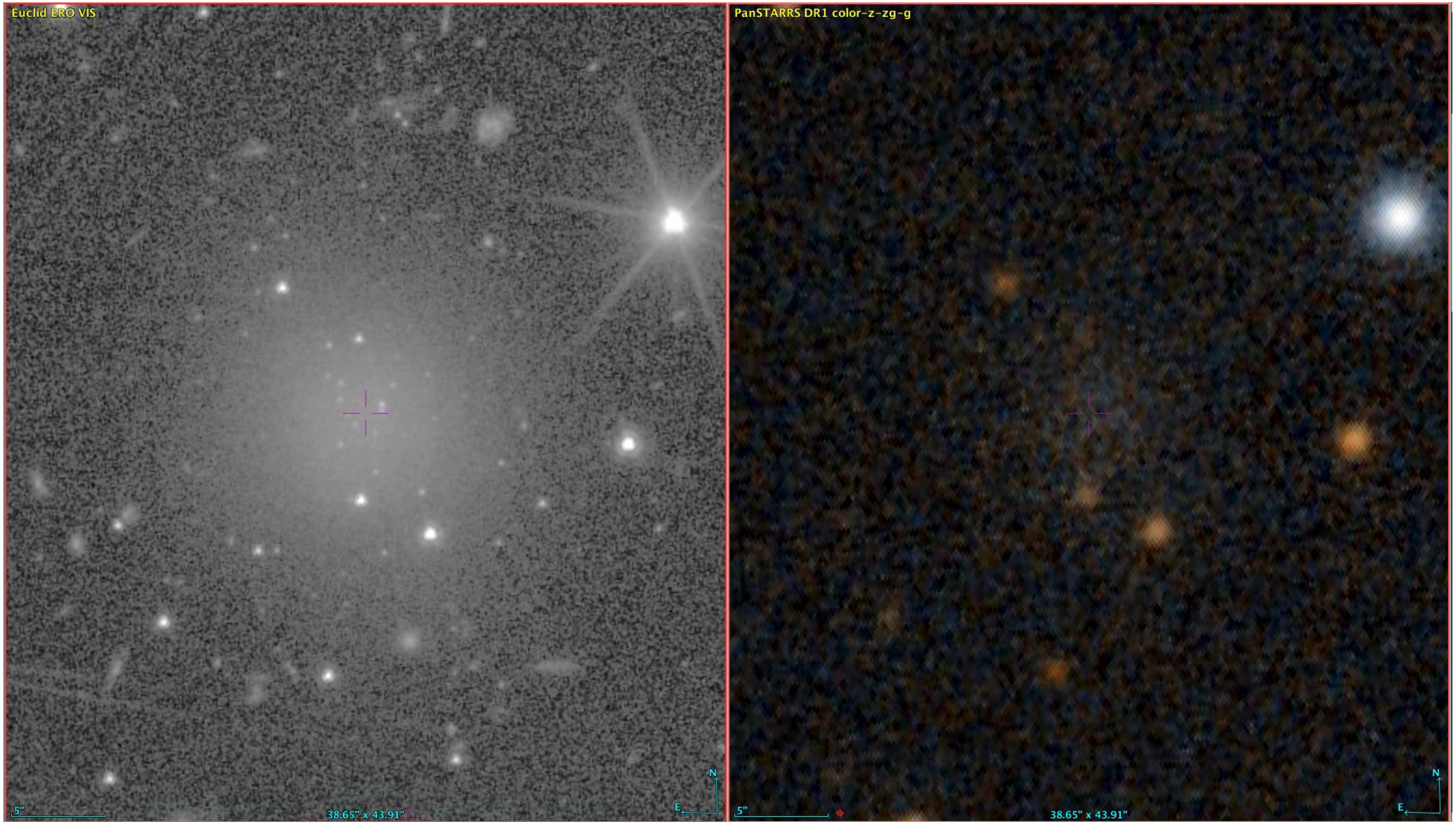
Euclid ERO


Euclid ERO FirstImages
1'
9.649' x 10.41'
N
E
PanSTARRS DR1 color-z-zg-g
1'
9.649' x 10.41'
N
E

Euclid ERO


Euclid ERO VIS
15"
1.127' x 1.28'
N
E
PanSTARRS DR1 color-z-zg-g
15"
1.127' x 1.28'
N
E

Euclid ERO


Euclid ERO VIS
5"
38.65" x 43.91"
N
E
PanSTARRS DR1 color-z-zg-g
5"
38.65" x 43.91"
N
E

Euclid ERO


Euclid ERO VIS
15"
1.127' x 1.28'
N
E


PanSTARRS DR1 color-z-zg-g
15"
1.127' x 1.28'
N
E

Euclid ERO Perseus

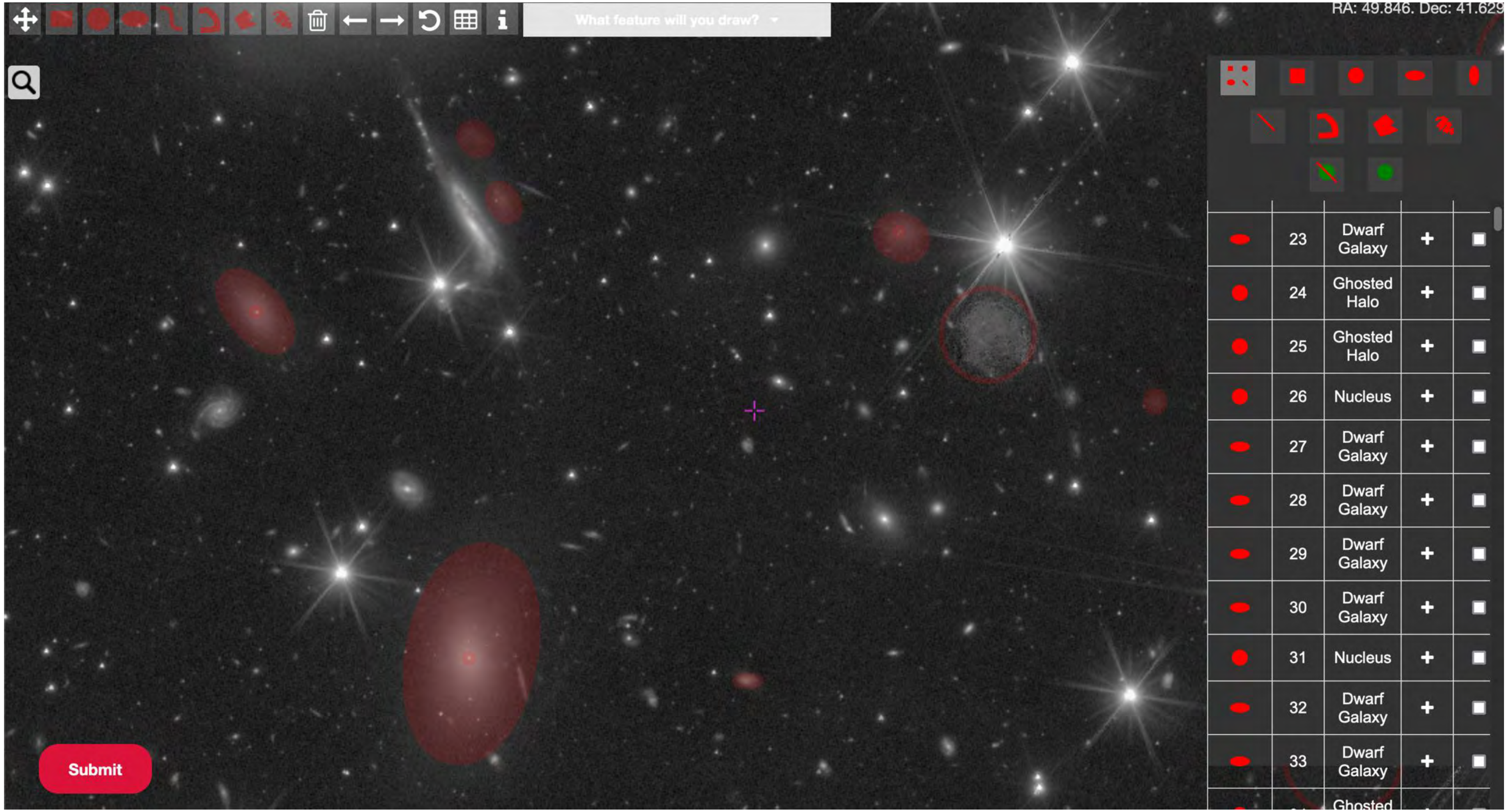
- Perseus galaxy cluster:
- 72 Mpc, dynamically young, signs of ongoing interactions
- Dominated by ETGs, lack of LTGs + large population of dwarf galaxies

- ERO images:
- 0.7 deg^2 located within the inner $0.25r_{200}$
- I_E SB limit of 30.1 $mag/arcsec^2$ and point-source depth of 28 mag.

What can Euclid tell us about dwarf galaxies in this well-studied cluster ?
▶ *Marleau+2024*

Dwarf galaxy catalogue

- Visual detection of the field by seven contributors using the *Jafar* annotation tool (Sola+2022)



- Indicate the presence of nuclei and ghost haloes

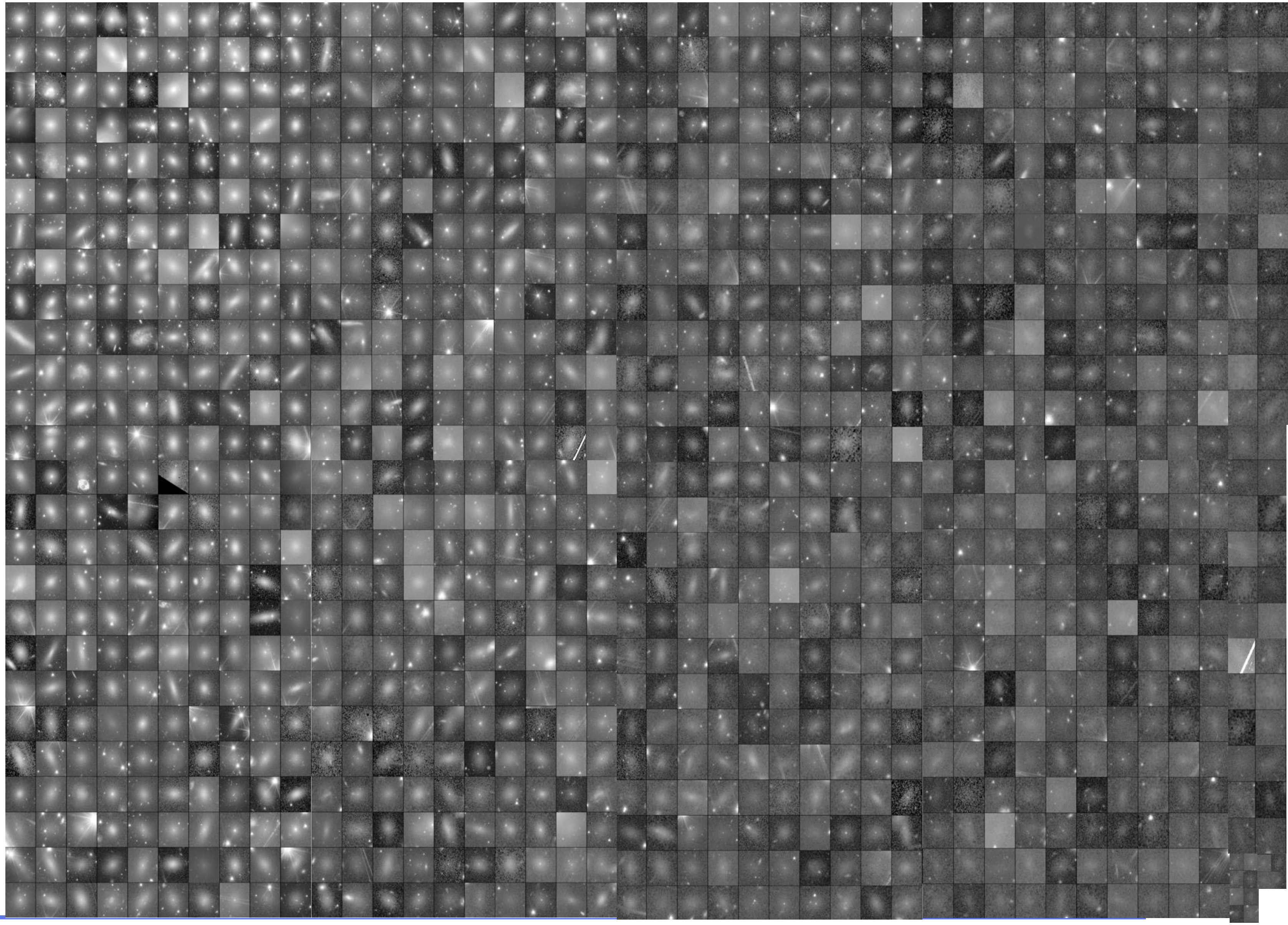
- Lot of time spent to be very selective, **purity** favoured over completeness

Marleau+2024

Dwarf catalogue

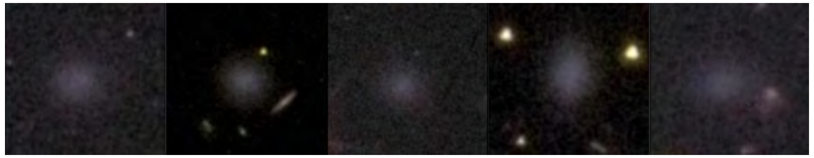
- 7 individual catalogues cross-matched
- Keep objects identified by at least 2 people
- Re-check

► Final list of **1100 dwarf** galaxies candidates, including 638 newly discovered !



Morphologies

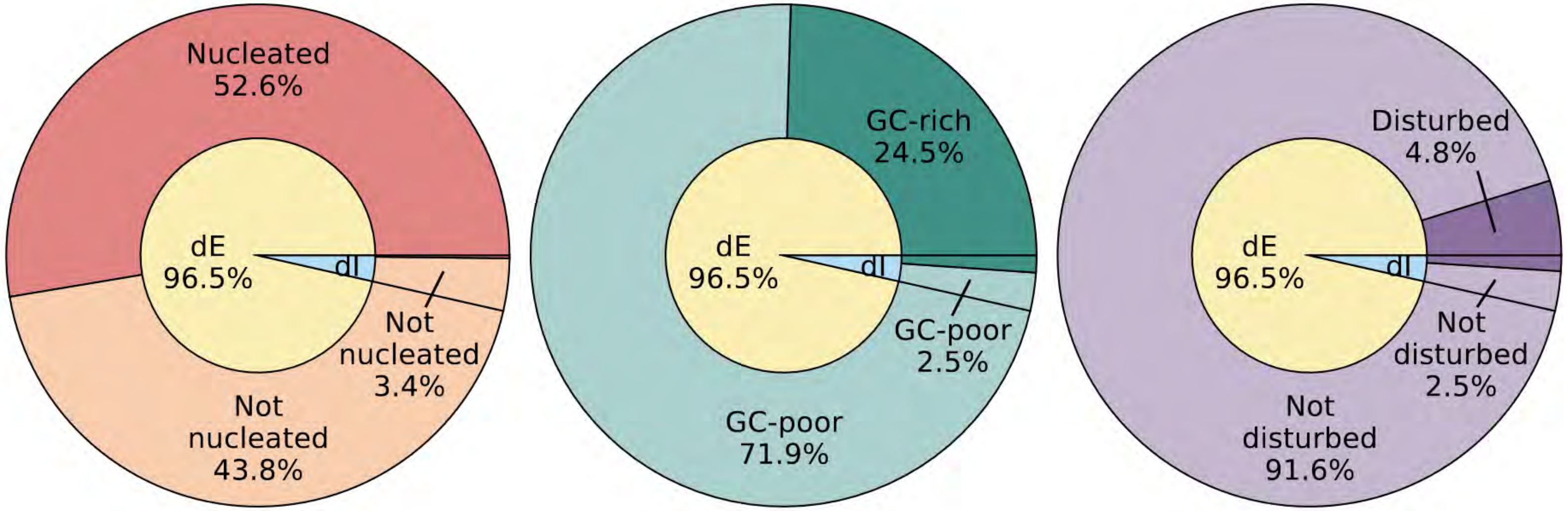
dE



96% dE

53% nucleated

dIrr



Nucleated

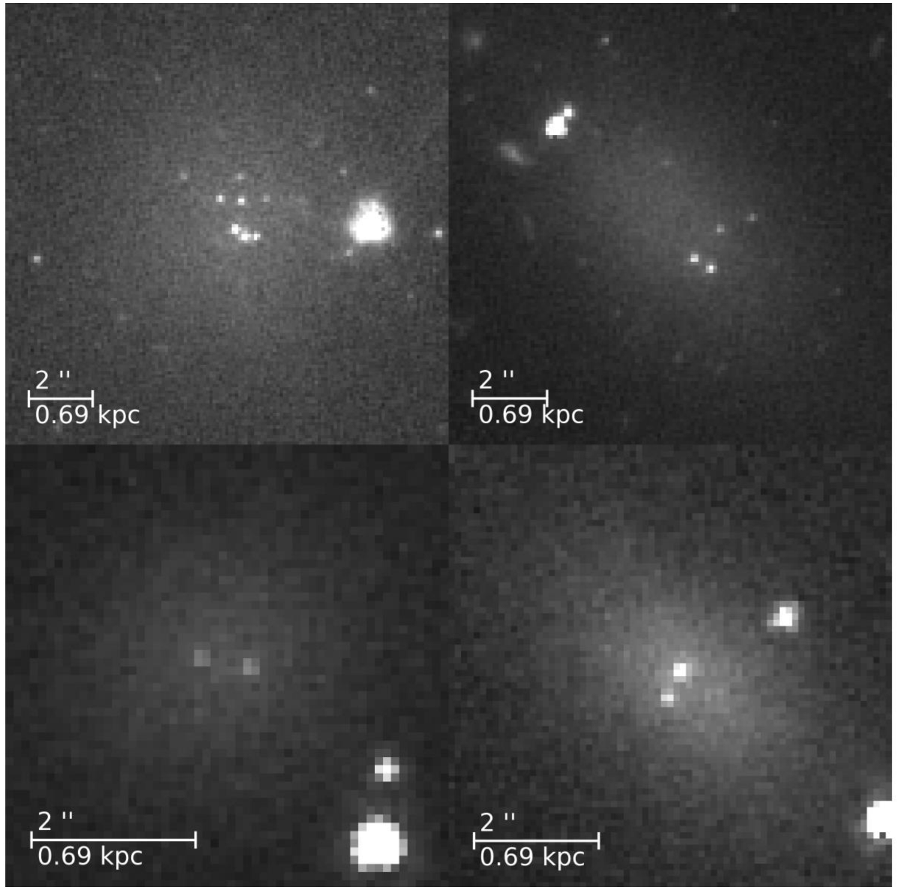
26% GC-rich

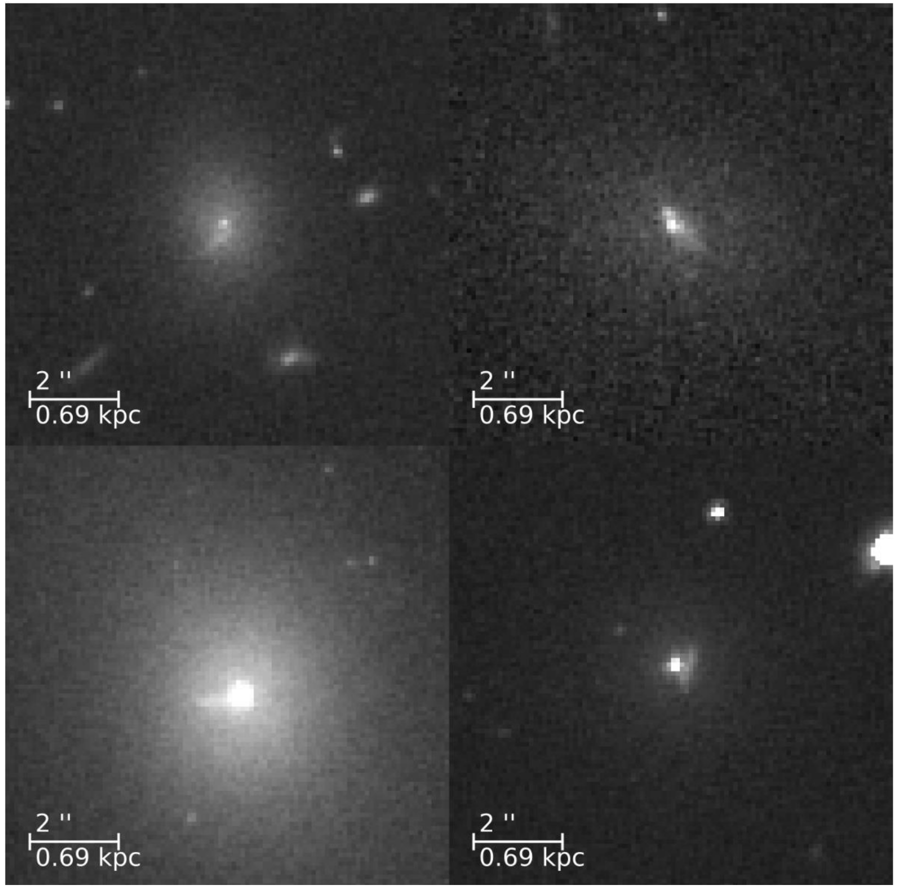
6% disturbed morphology

GC-rich

Nucleated dwarfs

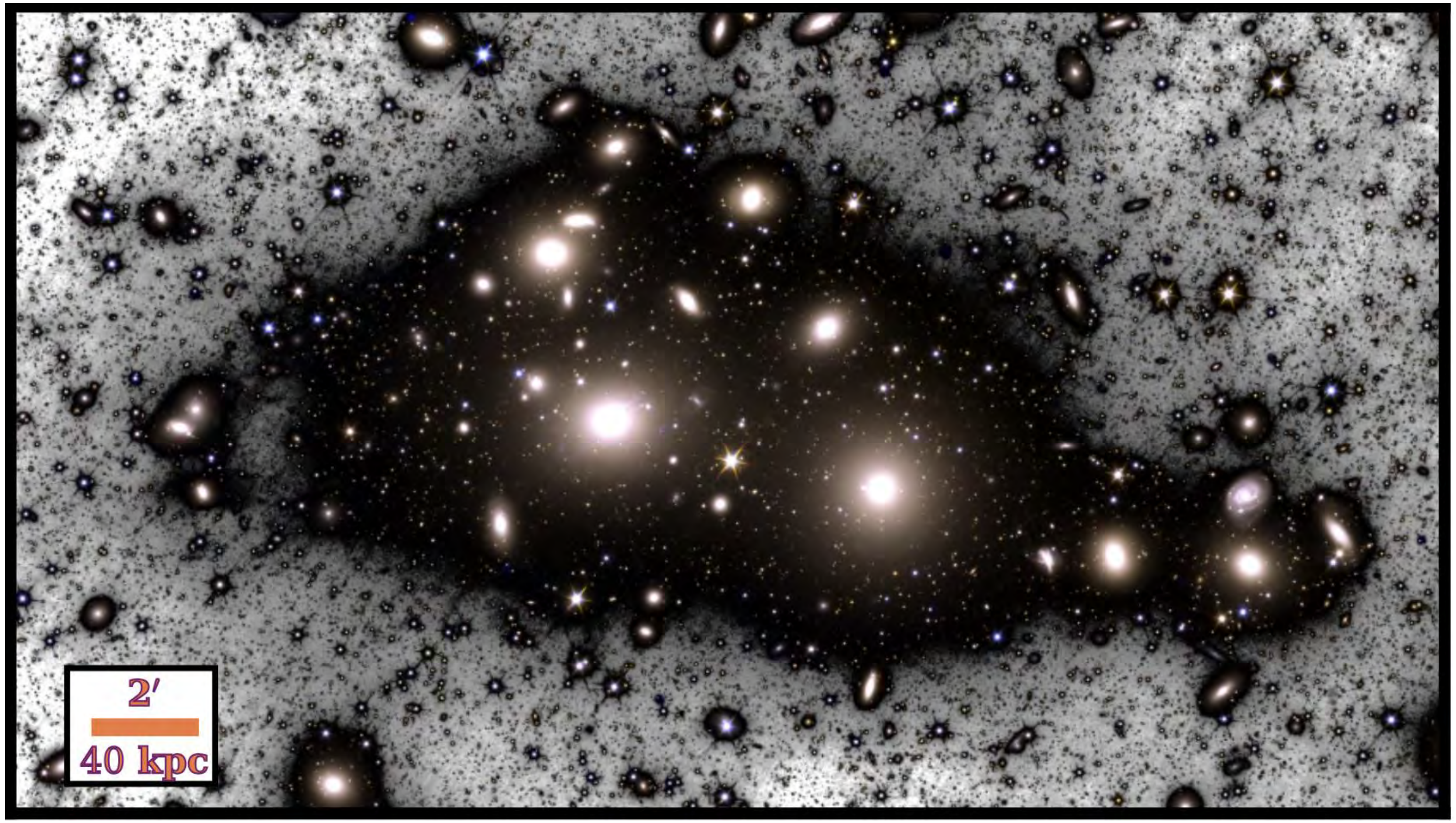
- Some dwarfs with multiple nuclei: dwarf-dwarf merger ? Massive GC infall ? Nucleus with stellar disc ? Contamination ?

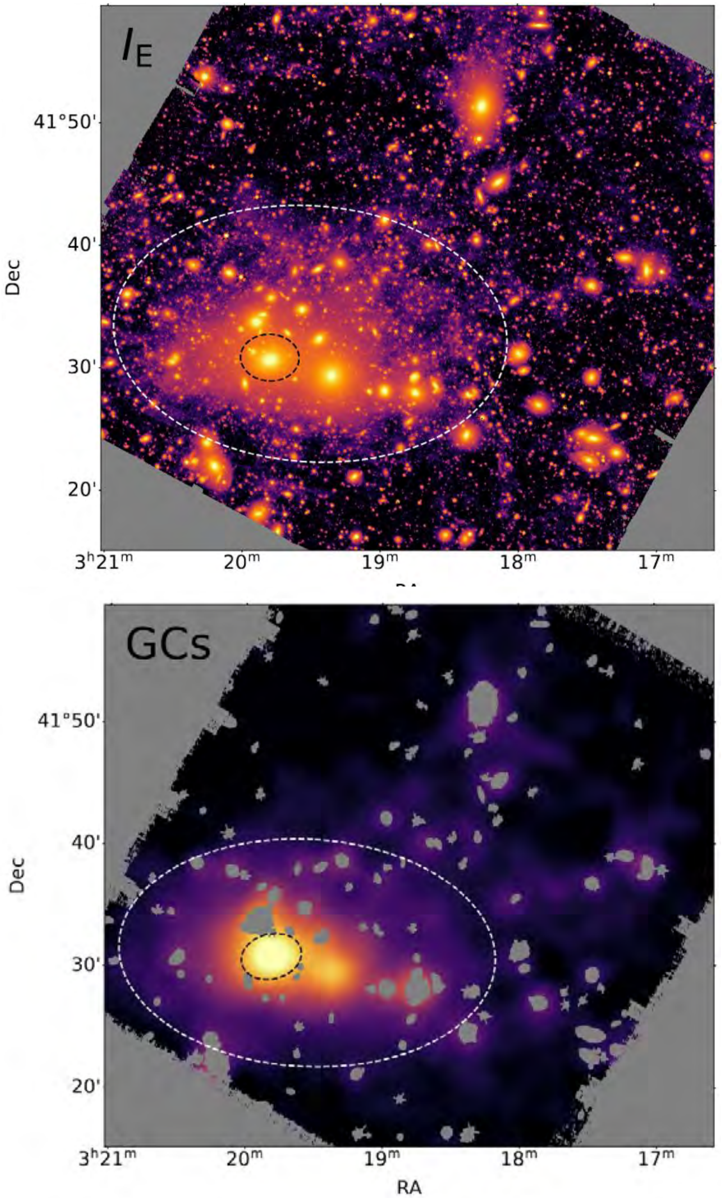




ERO Perseus ICL

- Intracluster light in Perseus cluster requires very careful processing and background subtraction
- Study the globular cluster and light distribution

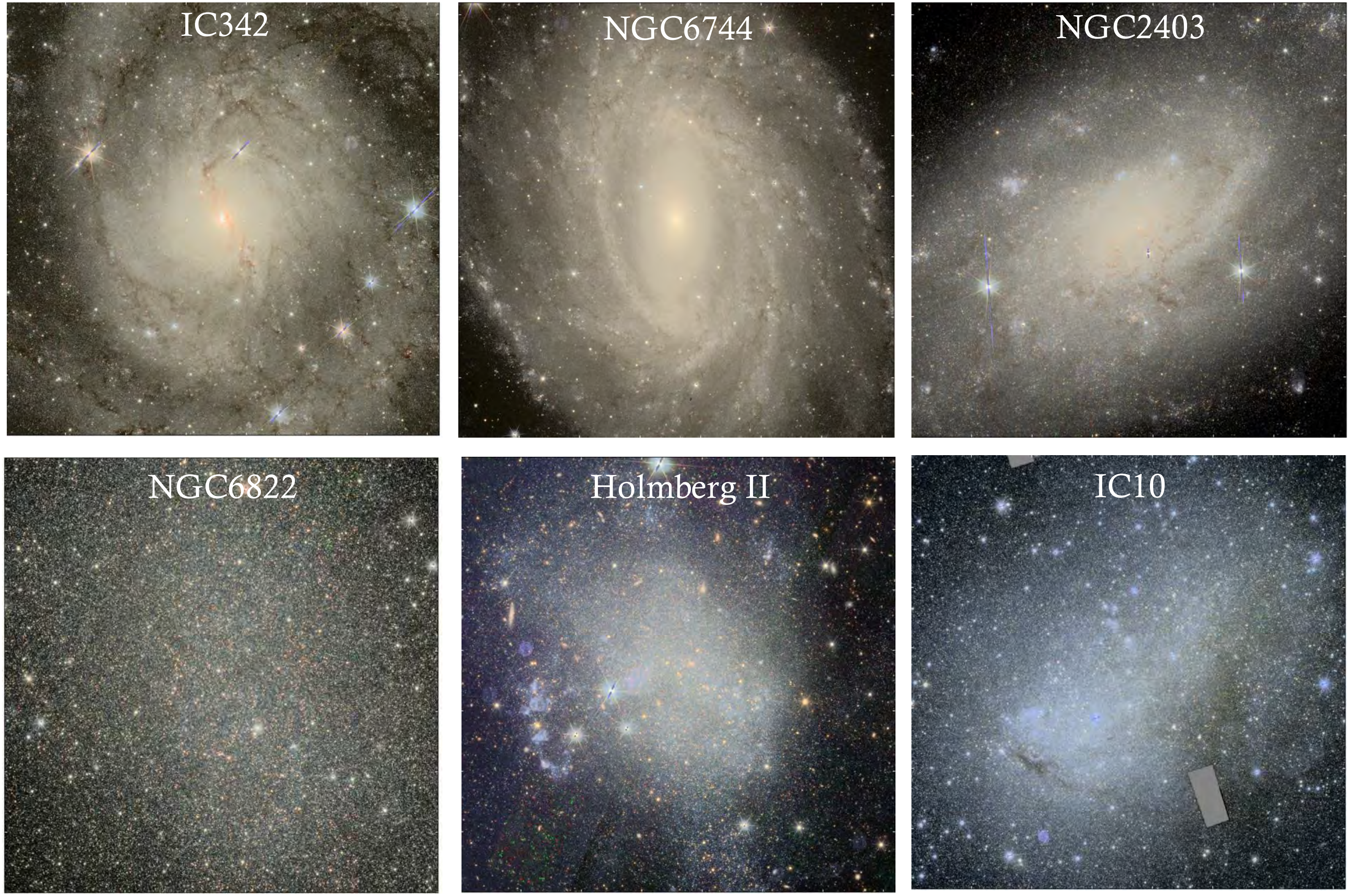




EC+ Hunt+2024

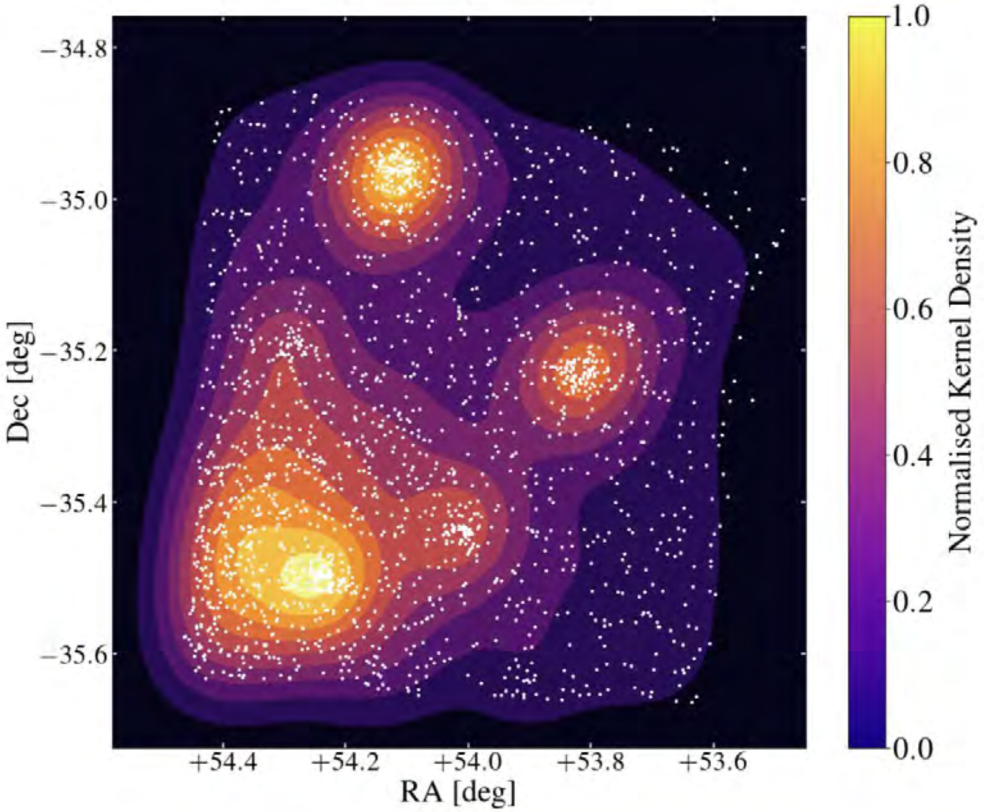
ERO nearby galaxy showcase

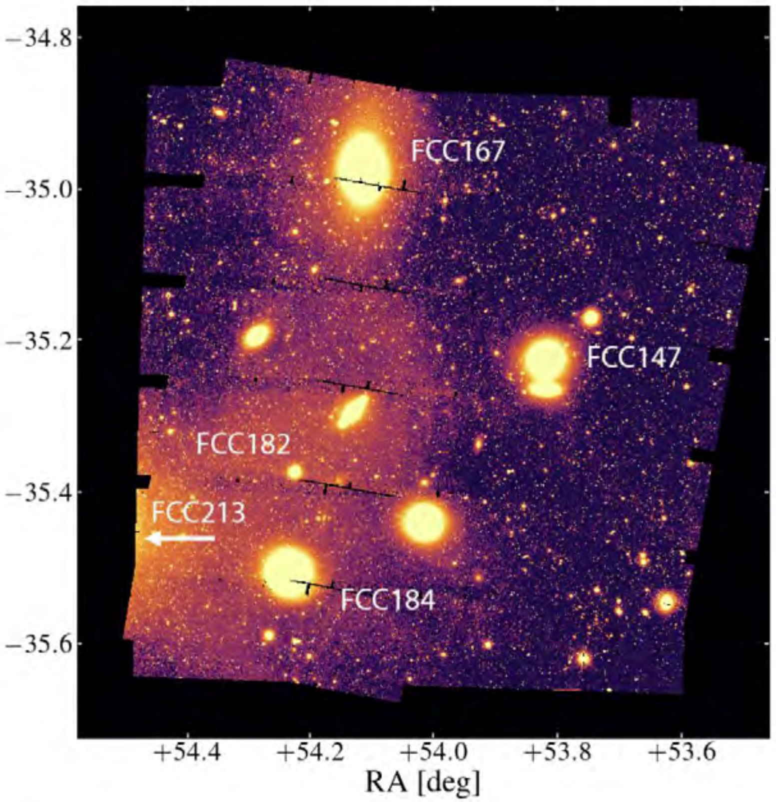
- 3 dwarf galaxies and 3 spirals
- D ~ 0.5-8.8 Mpc
- Resolved stellar photometry
- Study dwarf satellites systems and globular clusters



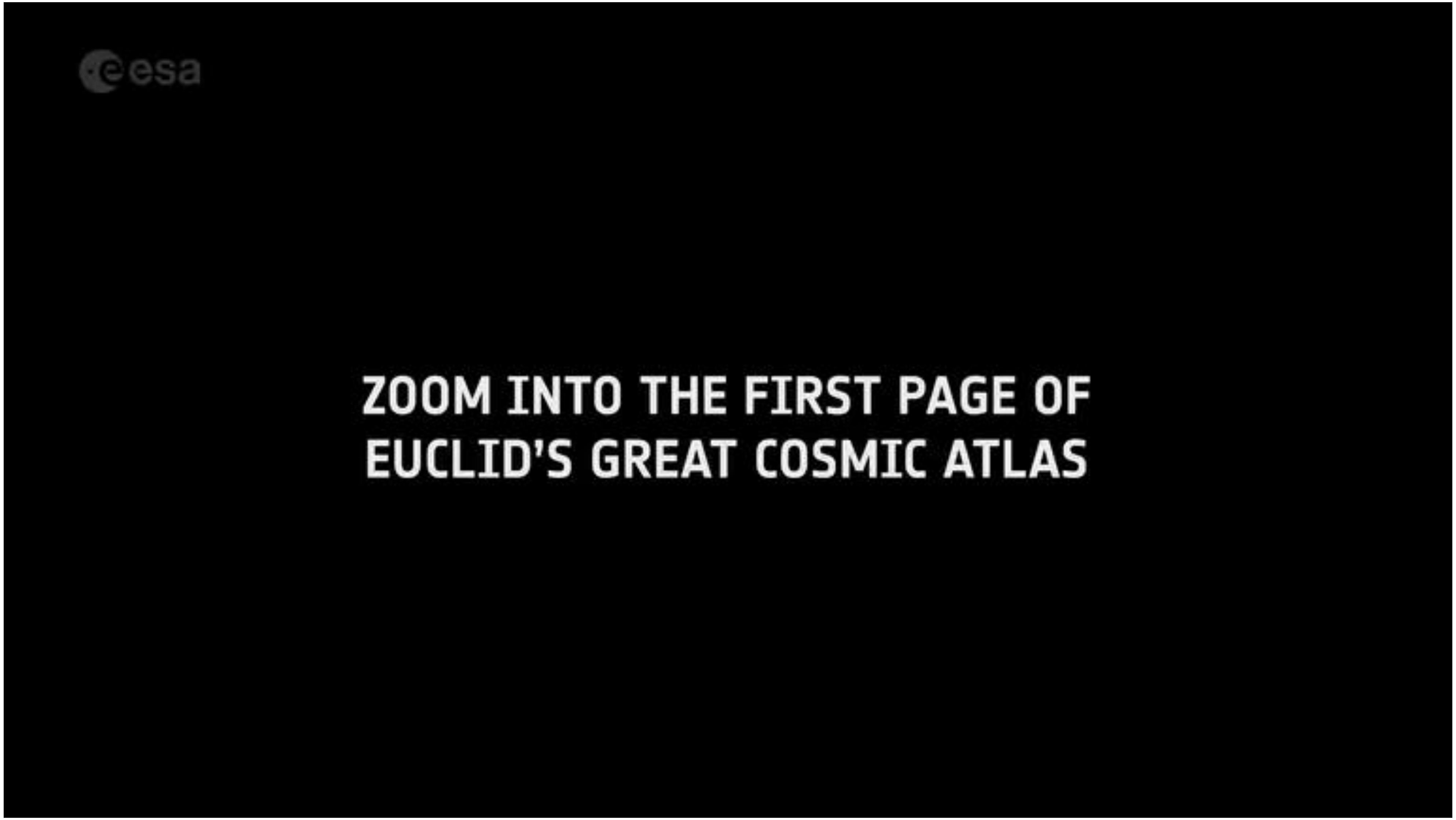
Globular clusters in the Fornax cluster

- Fornax galaxy cluster (20 Mpc)
- Goal: assess the detectability of globular clusters at 20 Mpc (need simulated data)

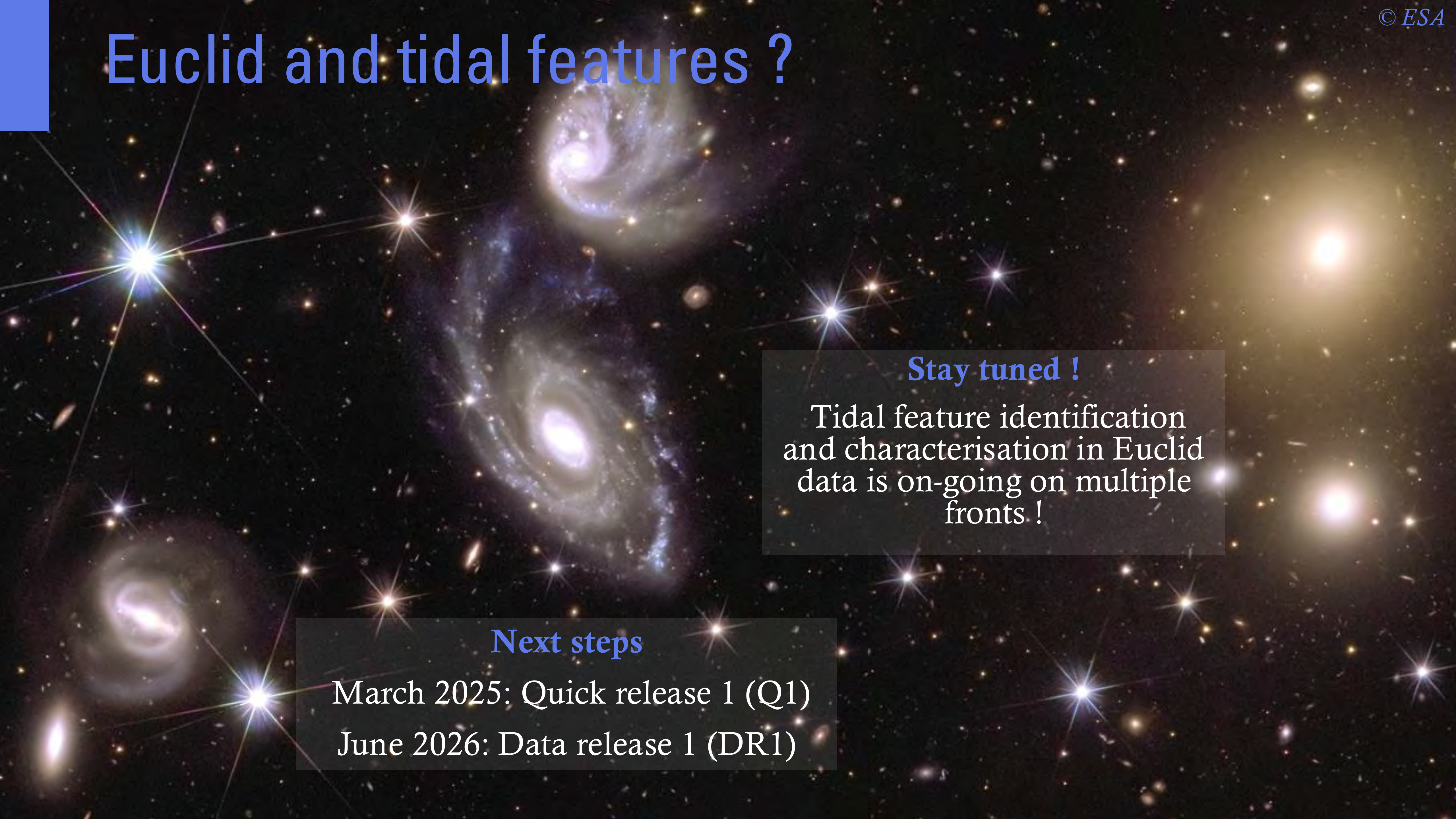




Euclid and tidal features



Euclid and tidal features ?



Stay tuned !

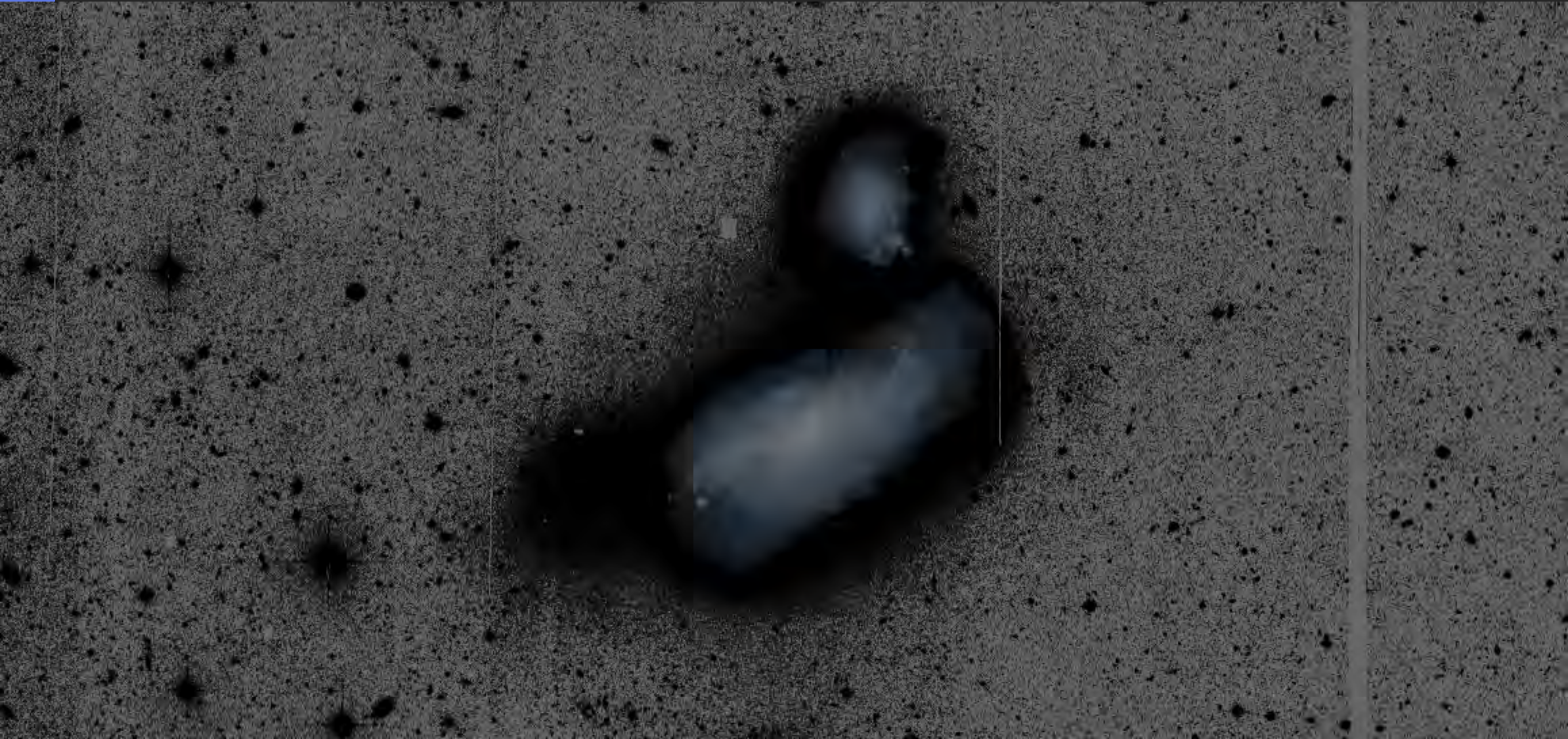
Tidal feature identification and characterisation in Euclid data is on-going on multiple fronts !

Next steps

March 2025: Quick release 1 (Q1)

June 2026: Data release 1 (DR1)

Finding tidal features



Deep images and tidal features

- Once the deep images are available, how to characterise tidal features ?

- Several approaches: **visual** (largely used), automated methods, machine learning

<table>
<tr><td>

- Large sample of galaxies
- Many contributors
- Decision trees

</td><td>Census of simple morphological features</td><td>

- Limited sample of galaxies
- One expert contributors

</td><td>Detailed analyses</td></tr>
</table>

- **Our approach**: obtain detailed analyses of LSB features on hundreds of galaxies with few contributors

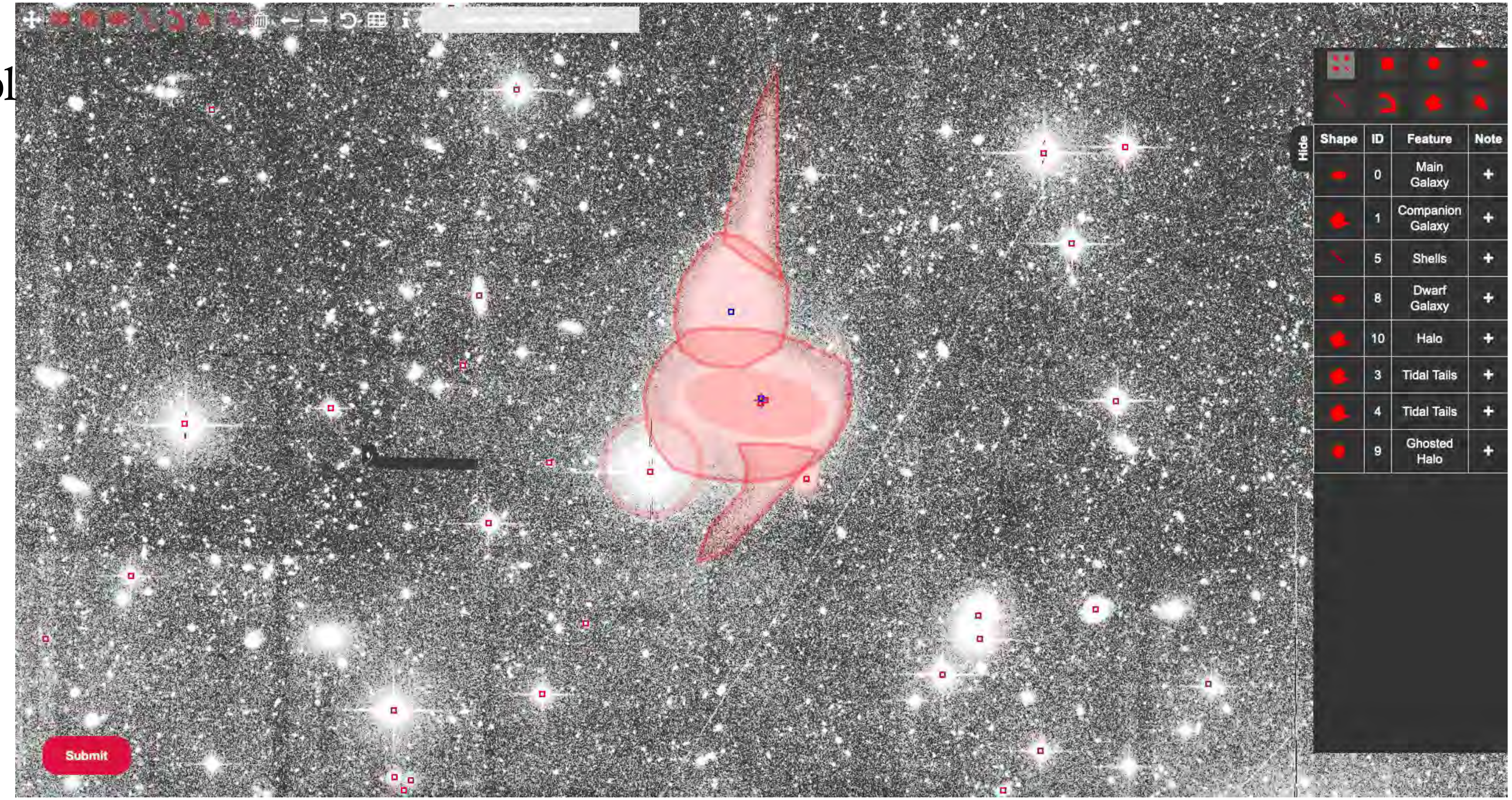
Jafar: an annotation tool for LSB structures

- Development of Jafar *(Sola+2022,2025)* an **online** tool *(with F. Richards*) to easily annotate and classify LSB structures in deep images

- Based on CDS Aladin Lite

Goals:

- **Draw** with precision the **shapes** of LSB structures
- A few expert contributors
- Obtain **quantitative** results about LSB structures



Interface of the annotation tool

Jafar: an annotation tool for LSB structures

Home Surveys Tutorial test

My annotations-observations

My annotations-simulations TNG50

Results and analysis

Logout

Welcome on the deep-imaging annotation server

Low Surface Brightness structures

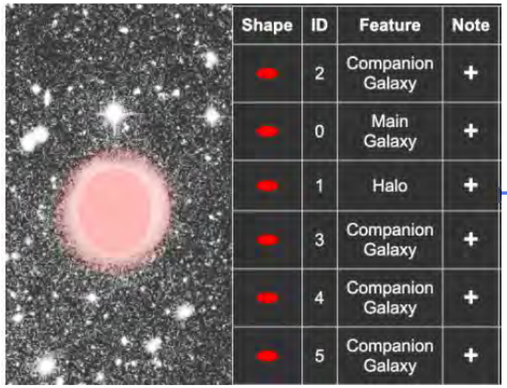
The study of the **morphology of galaxies** is essential to constrain models of galaxy evolution. Indeed, in the standard model, galaxies are formed through successive mergers with galaxies and other phenomena such as dark matter accretion or continuous process of cold gas. Vestiges of these past interactions between galaxies remain today, and they are very important as their shape and number informs us about the past merging history of a galaxy, and may change its apparent morphology.

However, these structures, these collisional debris, are very faint so their detection is complicated. They are called **Low Surface Brightness (LSB) structures**, as their flux per unit area is low. That is the reason why they have not been studied much outside the Local Group, until the use of powerful enough telescopes made their studies possible.

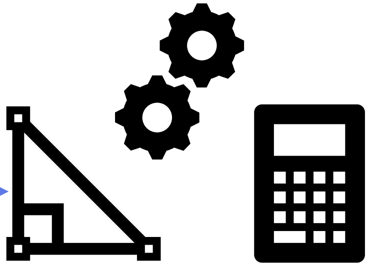
In particular, the **Canada-France-Hawaii Telescope** (CFHT) provided deep enough images of some galaxies that were studied with the MATLAS survey: many LSB structures around elliptical galaxies were discovered and classified. More recently, CFIS (Canada-France Imaging Suvey) is a large CFHT program aiming at mapping a large part of the Northern hemisphere (5,000 square degrees) with deep images, where MATLAS only

Analyses

- Need to take into the annotations of **several users**
- **Quantitative** measurements from the annotation database + **automated aperture photometry**



Contours + labels of the annotated shapes



Area
Length
Width
Surface brightness

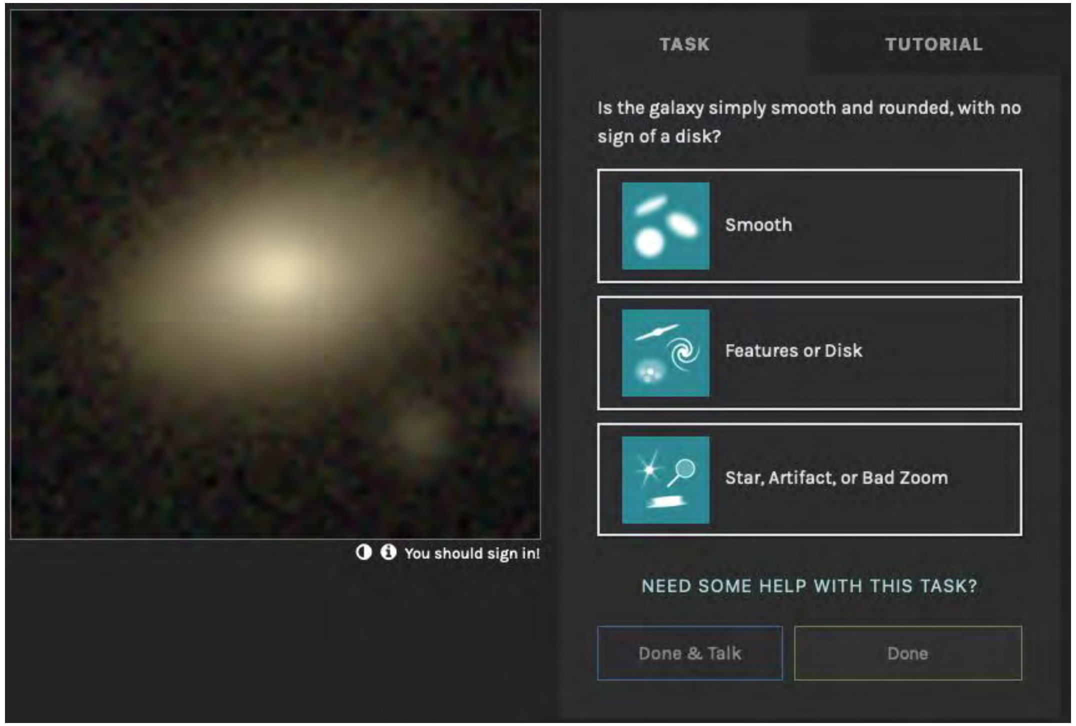
Masks of the annotations

Aperture photometry

Photometry
Colours
Mass

What about citizen science ?

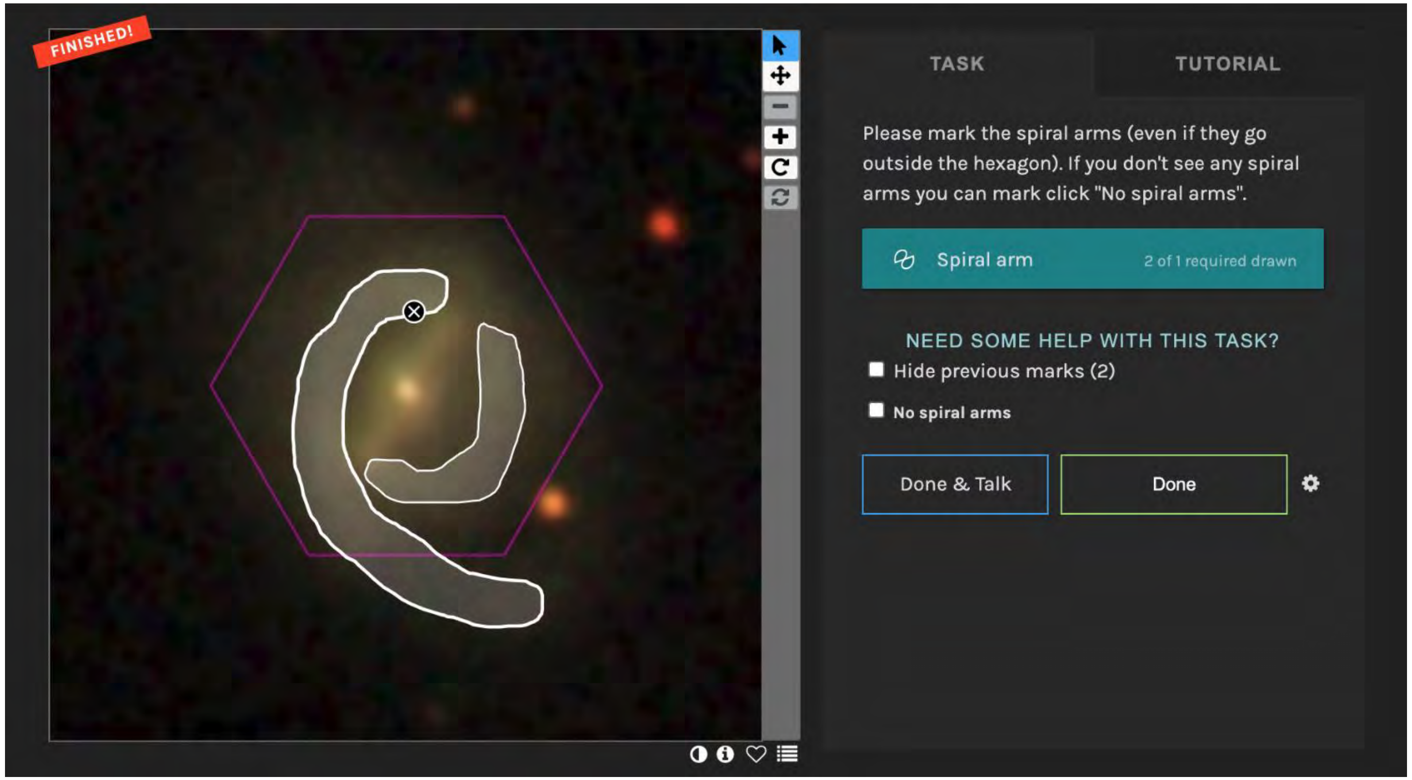
- Why not use the help of the general public for this task ? **Citizen scientists**
- A crowd of citizen scientists can classify hundreds of thousands of galaxies for simple tasks.
- Galaxy zoo is a great example: +60 million classifications about galaxy morphology

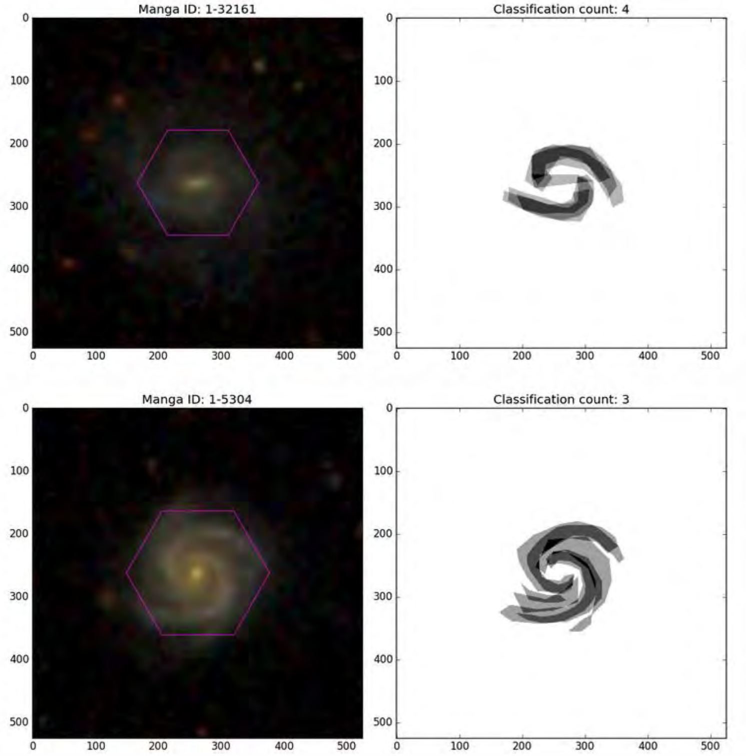


Example of the Galaxy Zoo interface

What about citizen science ?

- Galaxy Zoo 3D: recently added the possibility to draw polygons: used to count the number of spiral arms or the shape of the bar

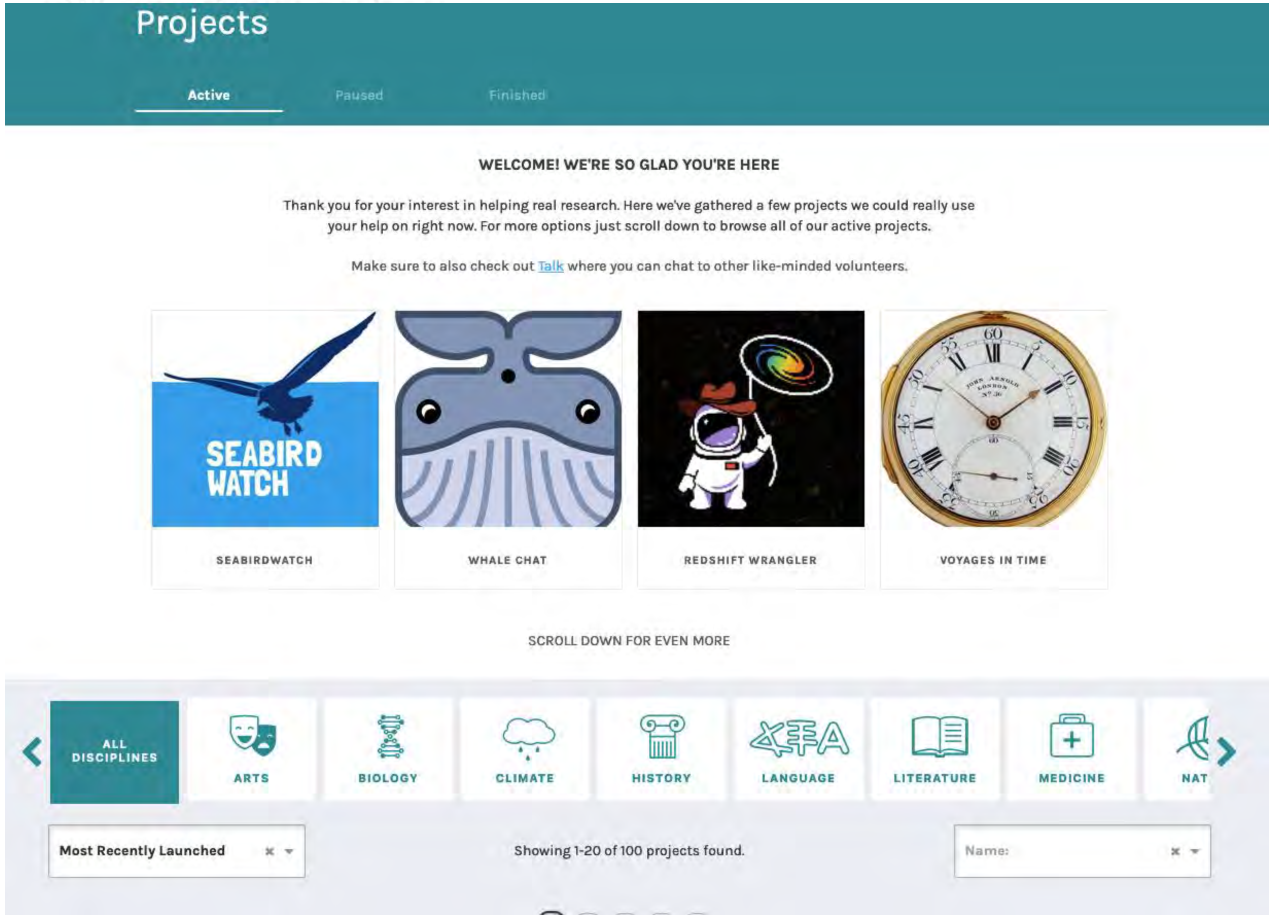




Masters+2021

Citizen science projects

- Zooniverse: extension to many different fields of science, largest and most popular platform but many others exist (e.g., NASA..)



Zooniverse

NASA

Citizen science projects

- Tidal feature identification: Galaxy Cruise project to classify the shapes of interacting galaxies

Fig. 1. GALAXY CRUISE's classification screen. Participants classify the target at the center, and then 'sail' to the next target. Passport stamps, souvenirs and other features are available to motivate the participants.

public. We thus choose to adopt the simplified classification scheme. Some interacting galaxies show significantly distorted shapes and their original morphology can be difficult to infer. In such a case, the participant can choose 'not sure'. When a target galaxy is too close to a nearby bright star and a fair classification

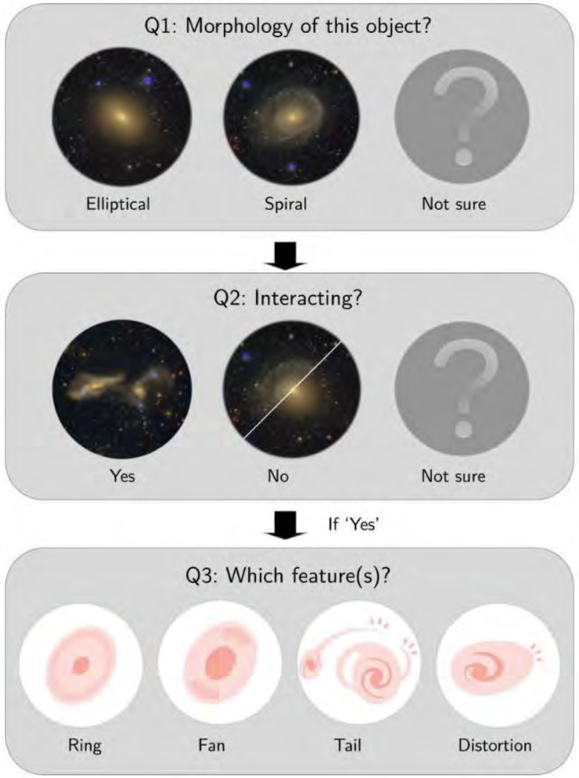


Fig. 2. GALAXY CRUISE's classification scheme. The first question is whether a galaxy is spiral or elliptical. The second question asks if the galaxy is interacting or not. If yes, the participant can choose observed feature(s).

Tanaka+2023

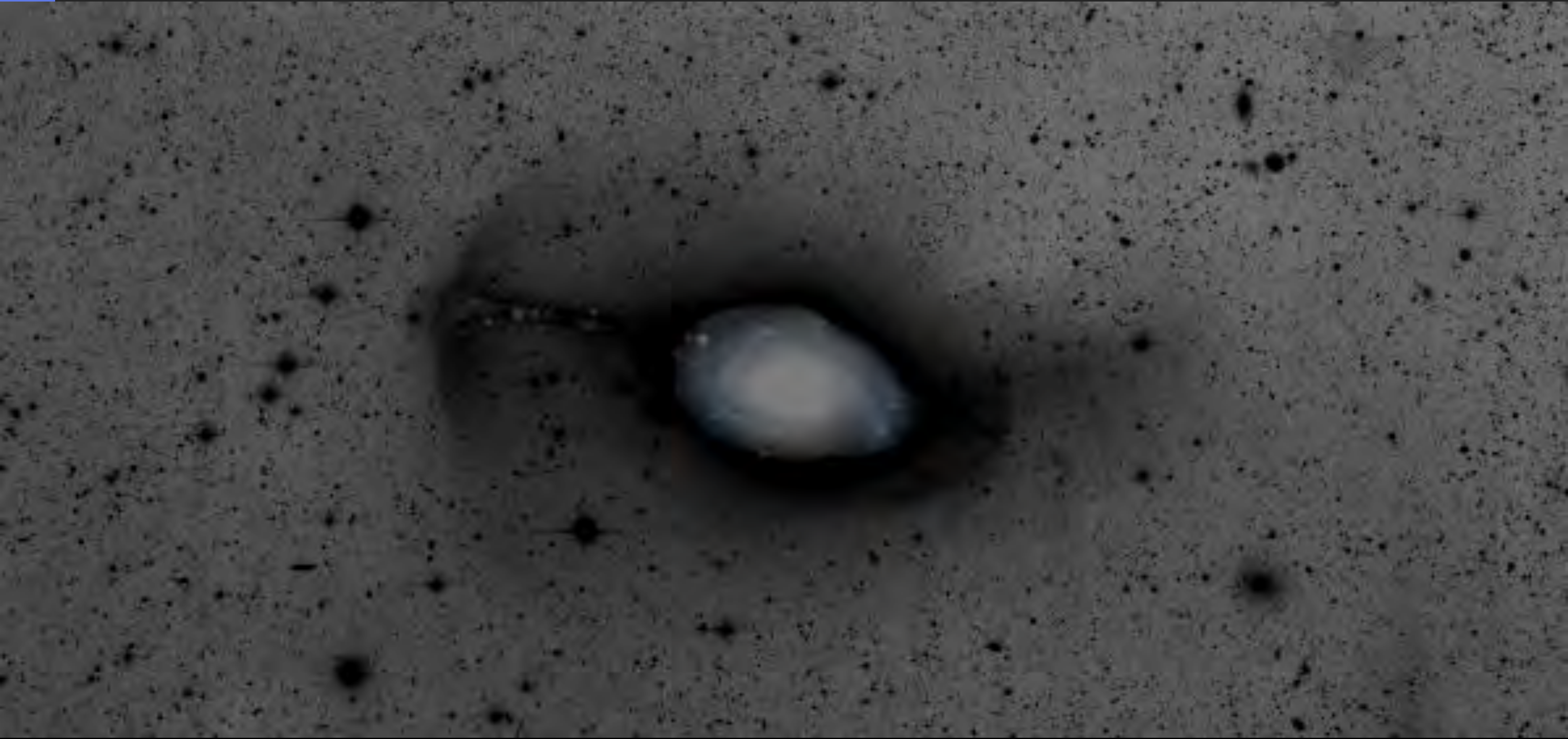
Then, why not use citizen scientists ?

- Highly specialized task: requires previous training and knowledge
- Subtle disentanglement between tidal tails and streams
- **Segmentation** and not just identification !
- Very time consuming

► Limited team of a few expert contributors

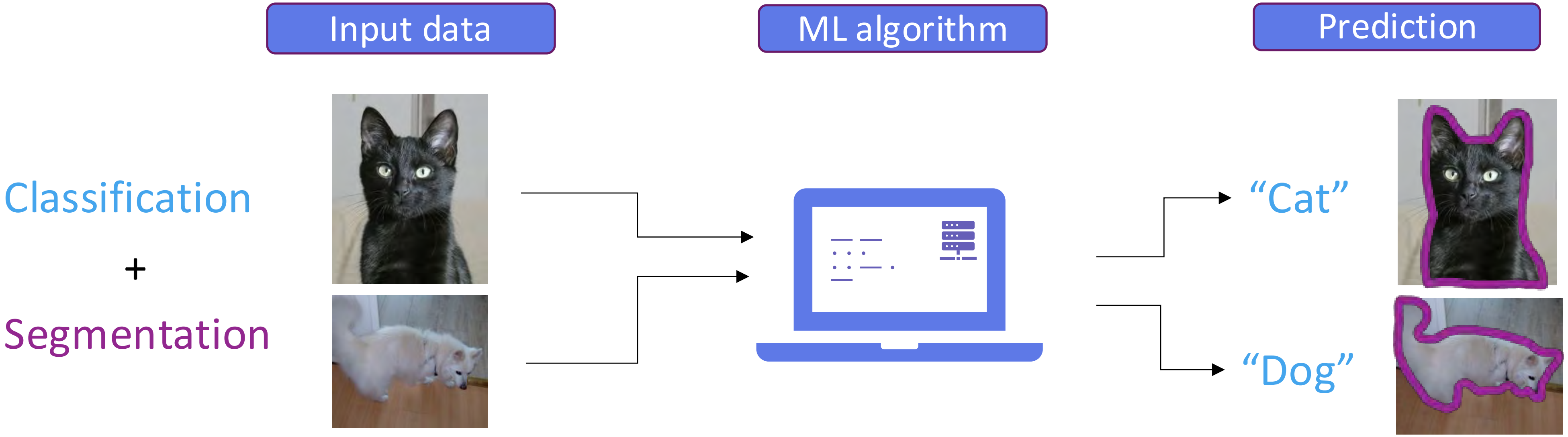
- But our manual approach is not possible for future surveys to come where thousands of galaxies will be studied.
- Need automated methods and **machine learning** approaches !

Machine learning and the LSB Universe



Machine learning

- Future surveys such as **Euclid** or **Rubin**/LSST will cover thousands of square degrees on the sky with exquisite depth
- Will push further the limit of the LSB Universe, but impossible to manually annotate
  - ►Need automated methods: **Machine Learning (ML)** to automatically find patterns in data



- Need **datasets of annotated structures** to train the algorithms (for supervised learning)

Supervised learning

- Several methods widely used in astronomy, in particular CNN (convolutional neural networks) to classify the morphology of galaxies *(e.g., Dieleman+2015, Huertas-Company+2015, Vega-Ferrero+2021, Walmsley+2022…).* **Supervised** learning using labelled data from e.g., Galaxy Zoo !
- Unsupervised techniques also used *(e.g., Martin+2020, Cheng+2021, Spindler+2021)*
- Networks focusing on the detection of tidal features *(e.g., Walmsley+2019, Storey-Fisher+2020, Martin+2022, Dominguez-Sanchez+2023, Gordon+24)*

Tidal feature identification with CNNs 3863

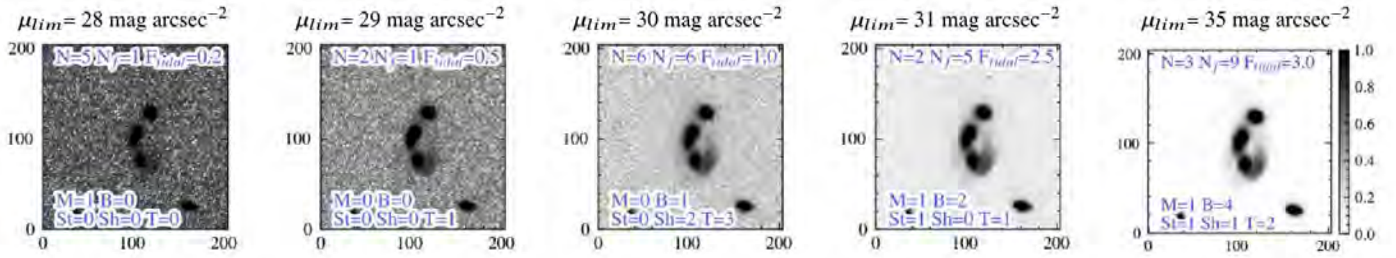
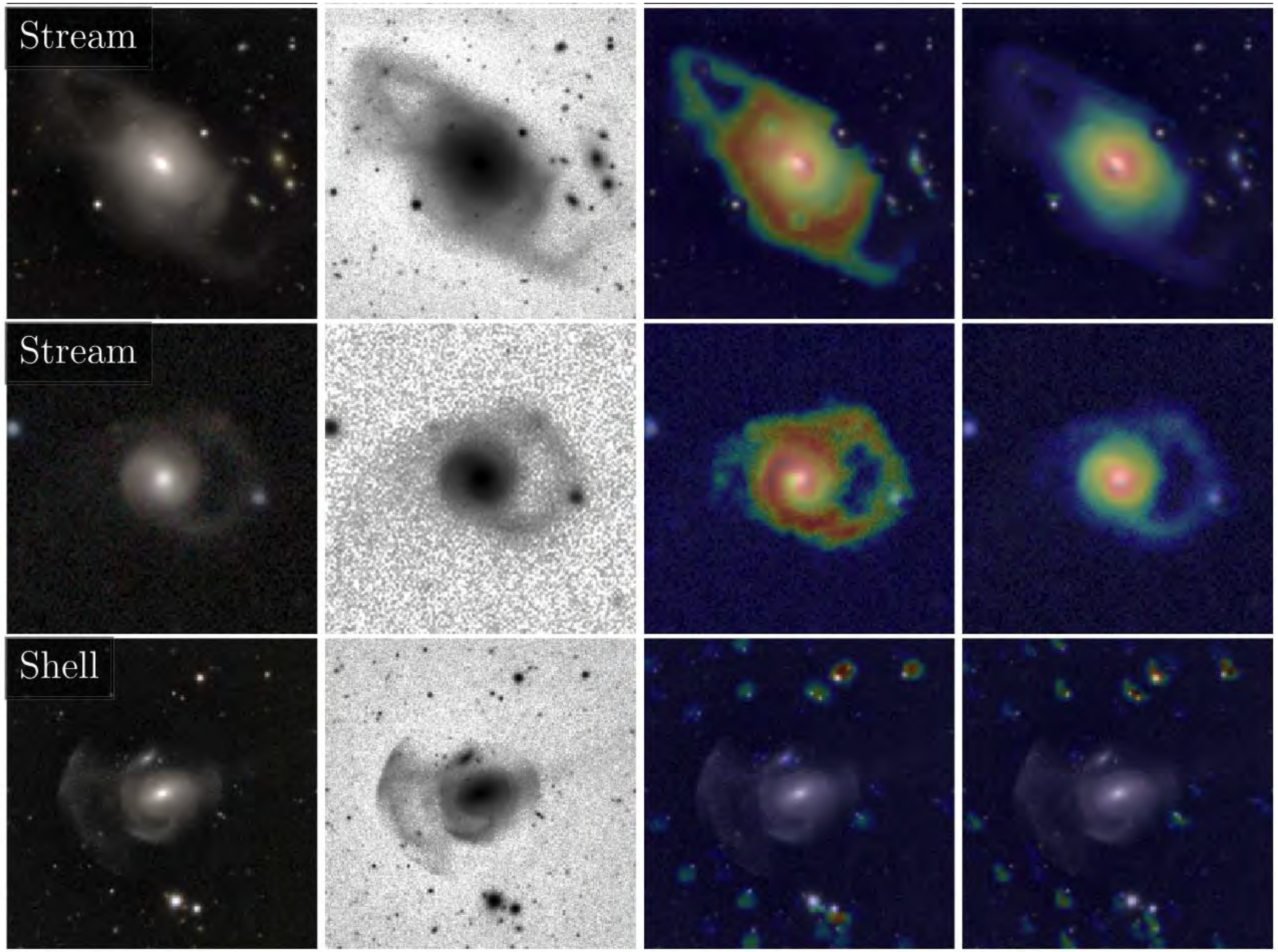


Figure 1. Example of the classification performed in M22. Images of the same galaxy observed at $z = 0.05$ with different $\mu_{\rm lim}$ (28, 29, 30, 31, 35 mag arcsec^{-2}, form left to right) were classified by a varying number of astronomers (N) into different categories. The number of observed features of each class is reported in the cut-outs (St = streams, Sh = shells, T = tails, M = mergers, B = bridges) along with the total number of features ($N_{\rm f}$) and $F_{\rm Tidal} = N_{\rm f}/N$. In this work, we consider as positive examples to those images with $F_{\rm Tidal} > 1$, negative those with $F_{\rm Tidal} = 0$, and uncertain otherwise. Following this criteria, the images with $\mu_{\rm lim} = 28$ and 29 have an uncertain classification, while those with $\mu_{\rm lim} > 29$ are classified as showing tidal features. The cut-outs are normalized in the (0, 1) range using *arcsinh stretch*, as described in Section 3.1.

Dominguez-Sanchez+2023

Supervised learning

- Supervised CNNs applied to tidal features (visually classified before) e.g., Gordon+24



Gordon+2024

Zoobot

- Zoobot *(Walmsley+24)* : CNN trained on 840k images and more than 100 million classifications from Galaxy Zoo volunteers
- Available on Github
- Possible to **retrain** Zoobot to adapt it to your own task

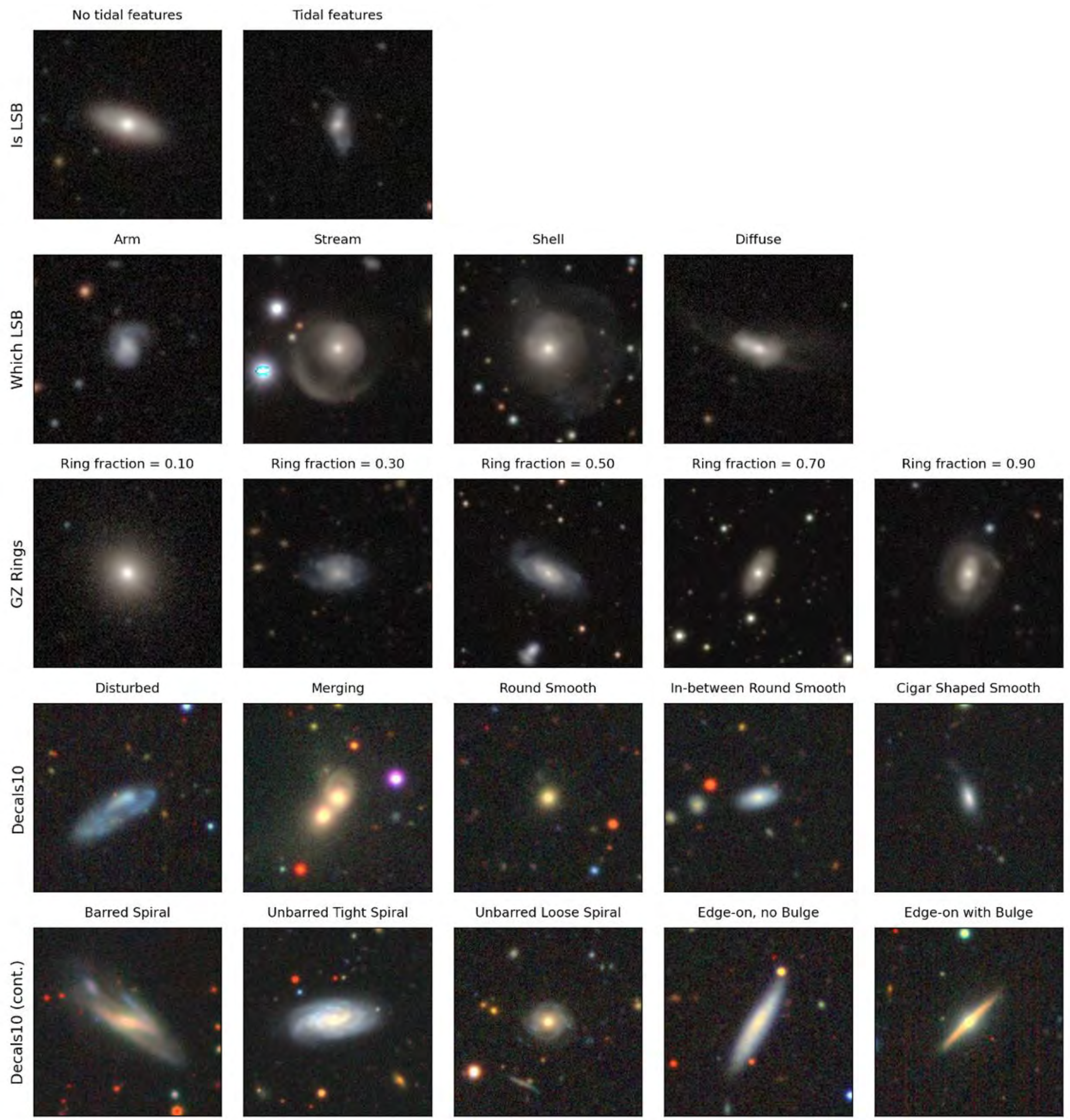
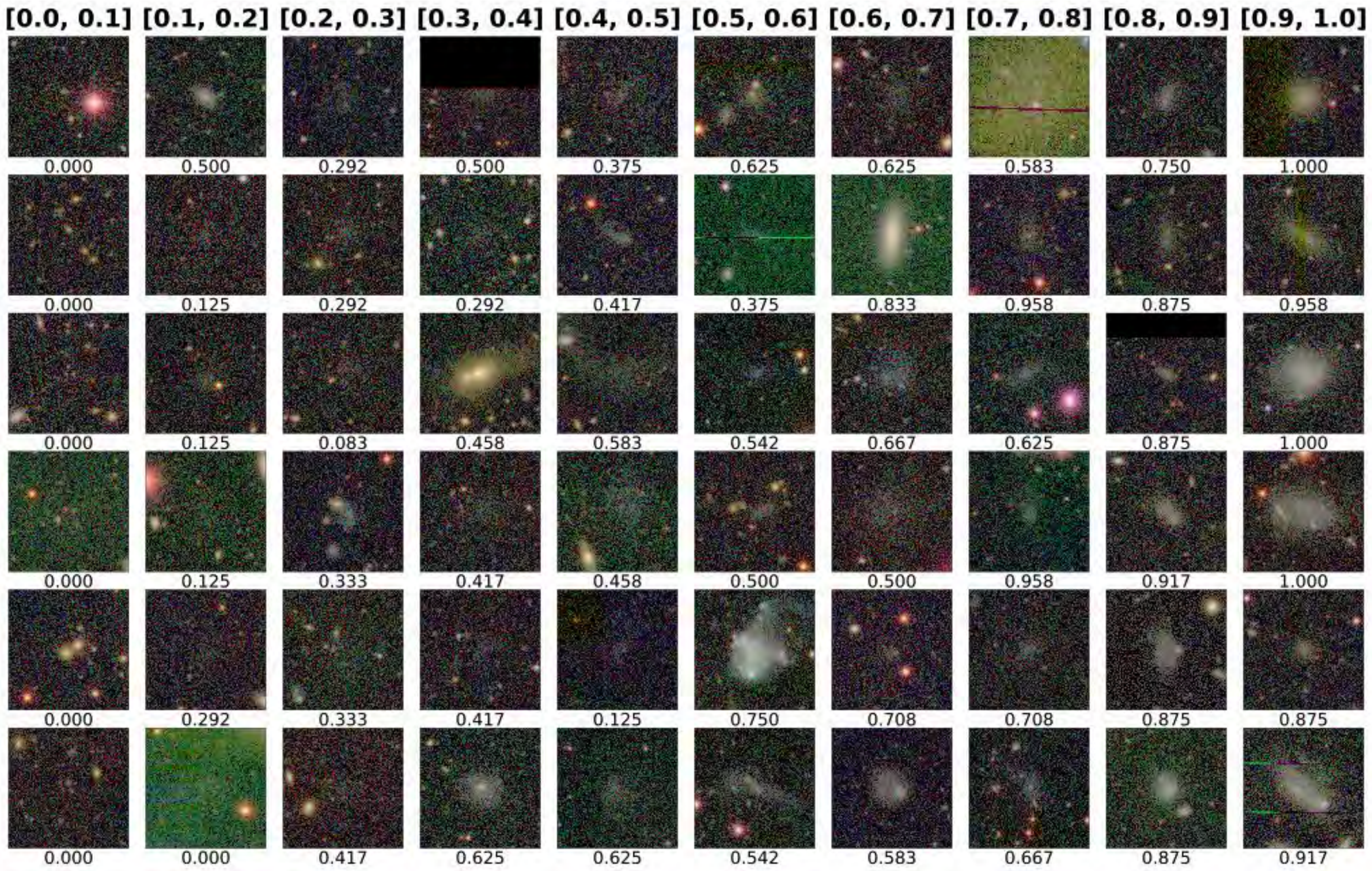


Figure 12: Random examples of labelled galaxies in each downstream dataset. Each dataset is shown by row (two rows for Galaxy10 DECaLS' 10 classes). The downstream label is shown above each galaxy.

Walmsley+24

Example CNN and dwarf galaxies

- CNN used to find dwarfs: e.g., Zoobot retrained to find dwarfs *(Hesteers+, in prep)*



Hesteers+, in prep

Self-supervised learning

- Another example: self-supervised learning: no labelled data for the training of the encoder but paired with a linear classifier trained on a small labelled dataset *(e.g., He+2019, Chen+2020, Walmsley+2022, Desmons+2023)*

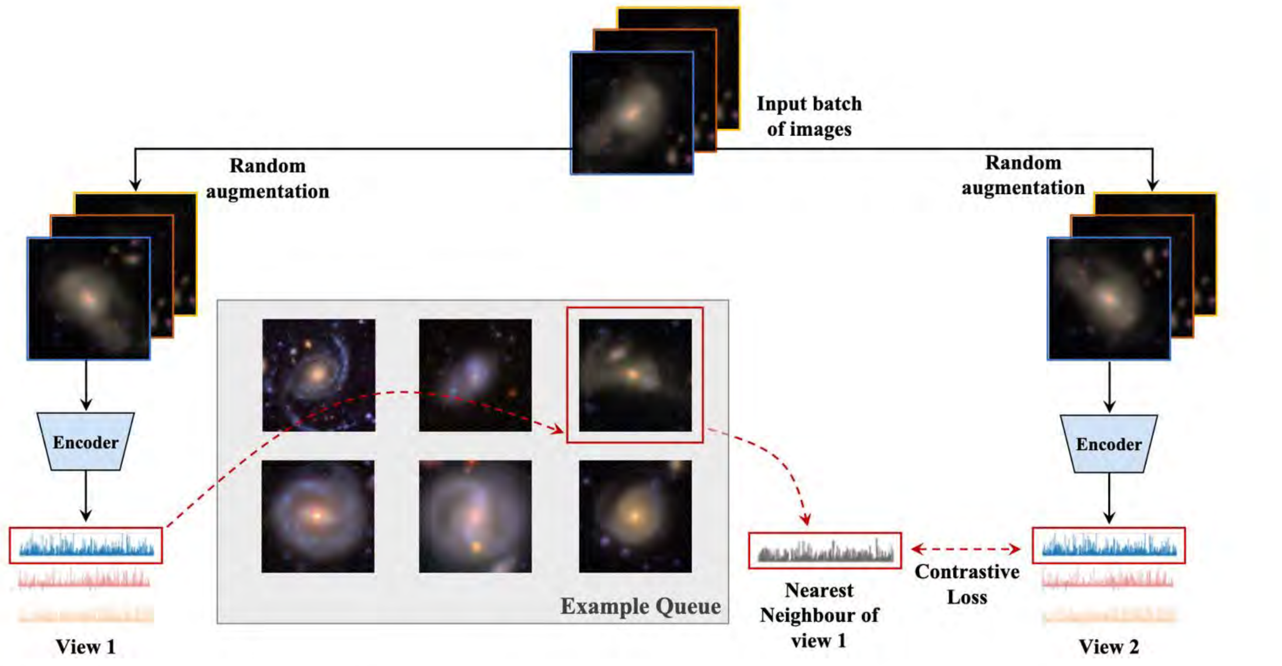
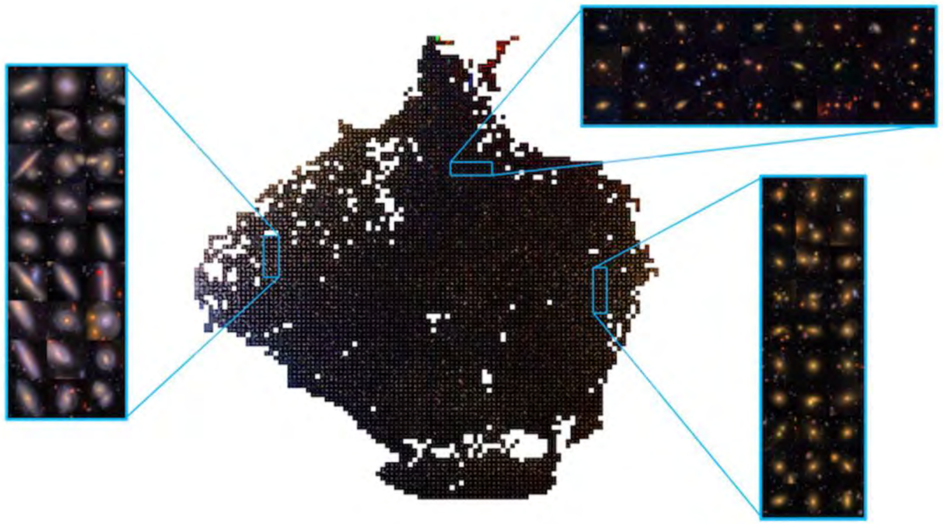


Figure 2. Illustration of the self-supervised model architecture. The model takes in a batch of images as input and creates two different views of the batch using random augmentations. Each view is encoded and the nearest neighbour of view 1 is located from the queue of example images. The loss between view 2 and the nearest neighbour of view 1 is calculated using a contrastive loss function which is minimised for similar pairs of images and maximised otherwise.

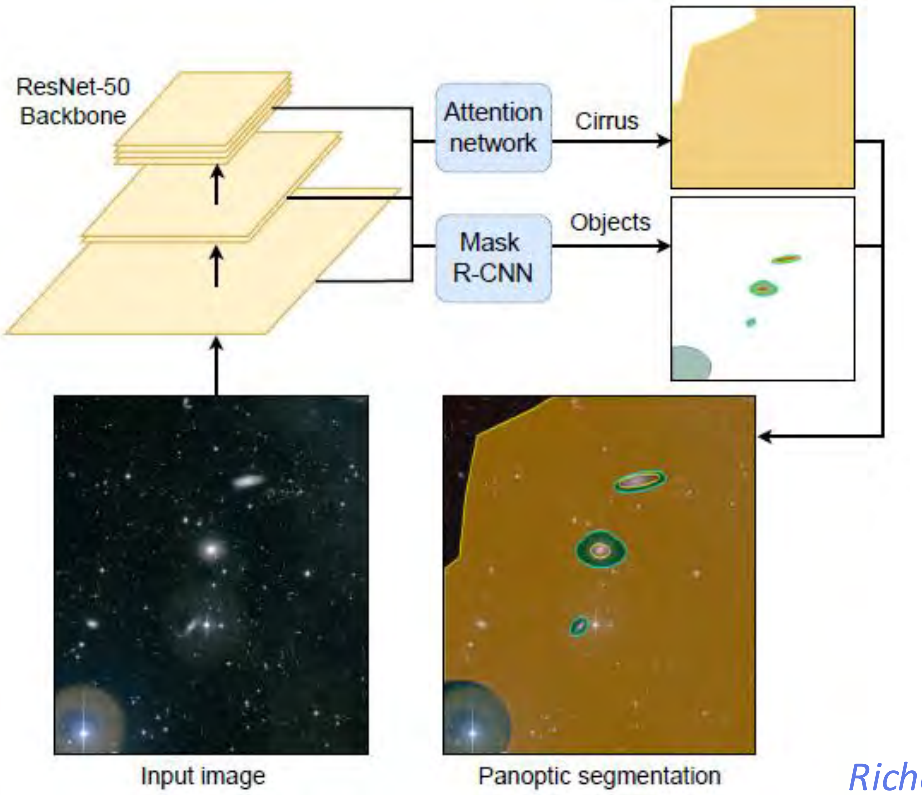


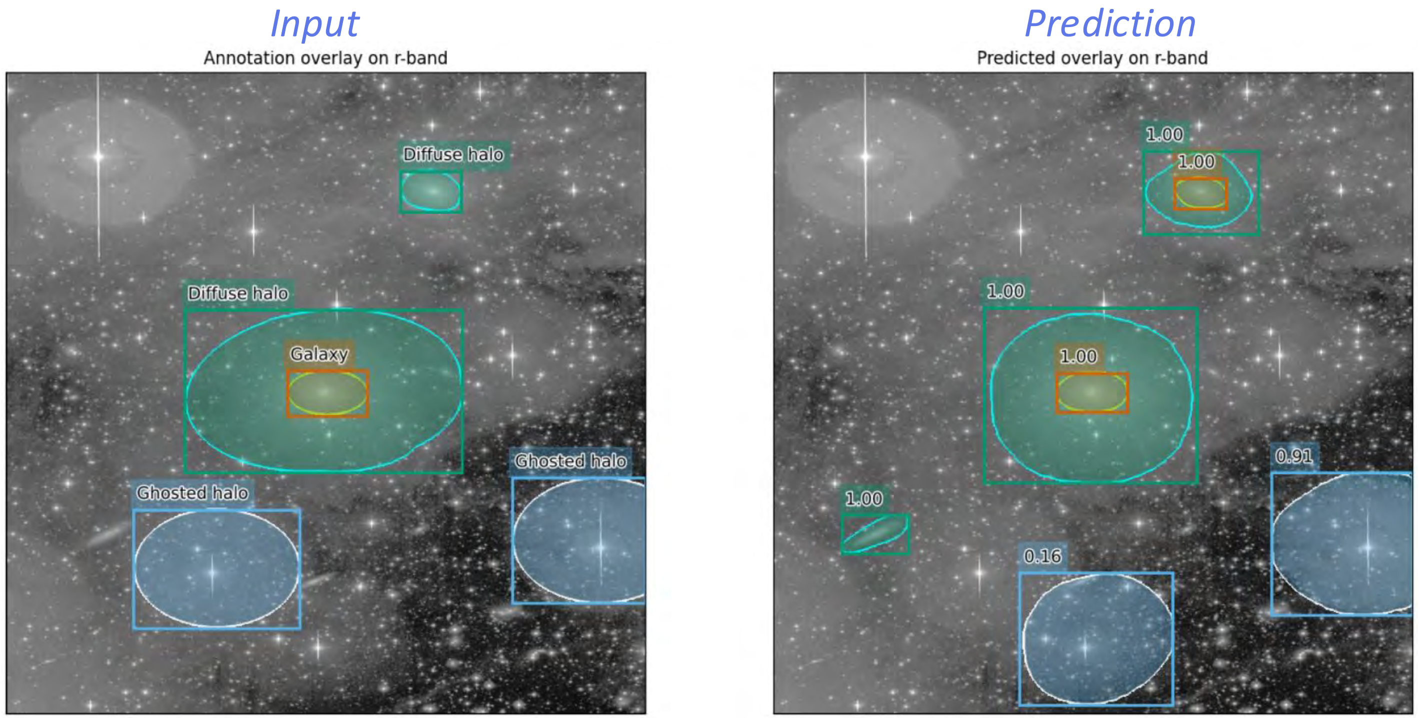
2D UMAP projection of the self-supervised representations

Desmons+2023

Automated feature segmentation

- Development of a Deep Learning algorithm trained on our **annotation database** to automatically detect and **segment** tidal features in deep images:
    - *Richards+2022, 2023*: network **adapted** to our segmentation problem, with **small dataset**, captures features orientation (e.g., cirrus)





Richards+2022, 2023

Automated feature segmentation

- Several approaches currently studied to automatically find and segment faint features in deep images (tidal features, dwarf galaxies…) ►big challenge !

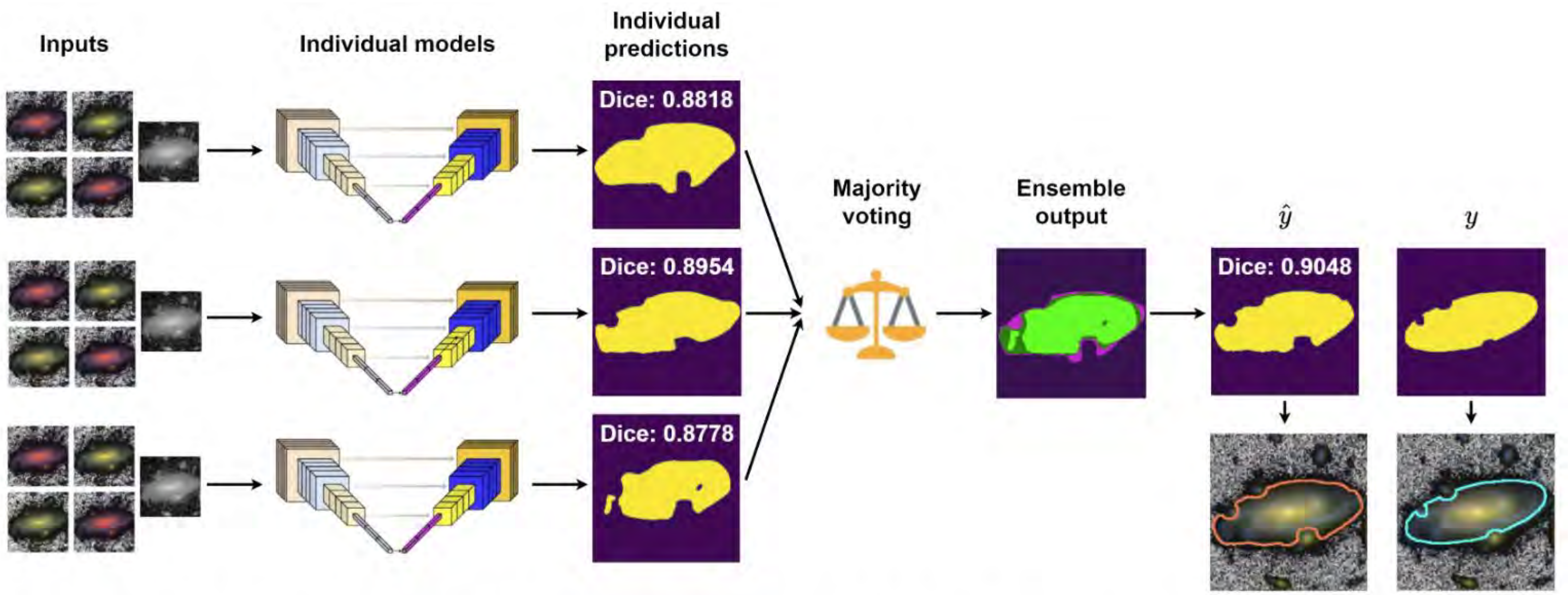


Fig. 9: Designed ensemble architecture. Predictions from the individual models are combined using a majority vote criterion to provide a final output.

Fernandez+2023

Foundation models

- Several approaches currently studied to automatically find and segment faint features in deep images (tidal features, dwarf galaxies…) ►big challenge !
- **Foundation models** (= general-purpose AI) investigated as one way to solve this problem, e.g., Segment Anything Model (SAM)

►

https://segment-anything.com/demo

Foundation models

Xu & Zhu+2024

- SAM applicable to many different types of images

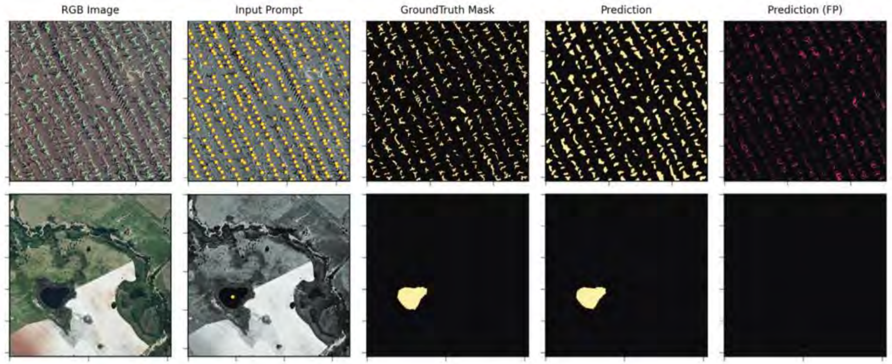


Figure 32: From Osco et al. (2023), depictions of images segmented by SAM utilizing point prompts. The initial column displays the RGB

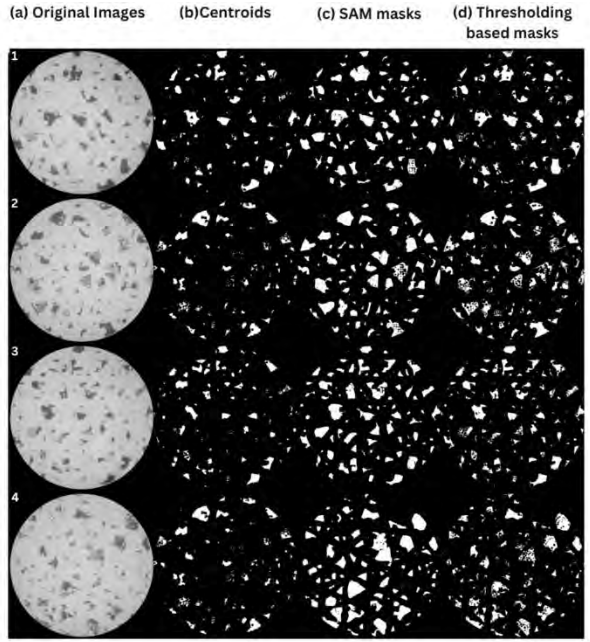
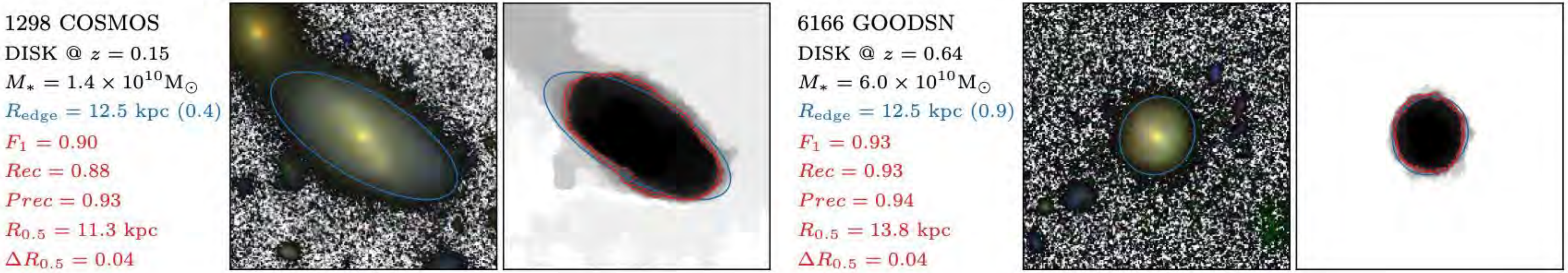
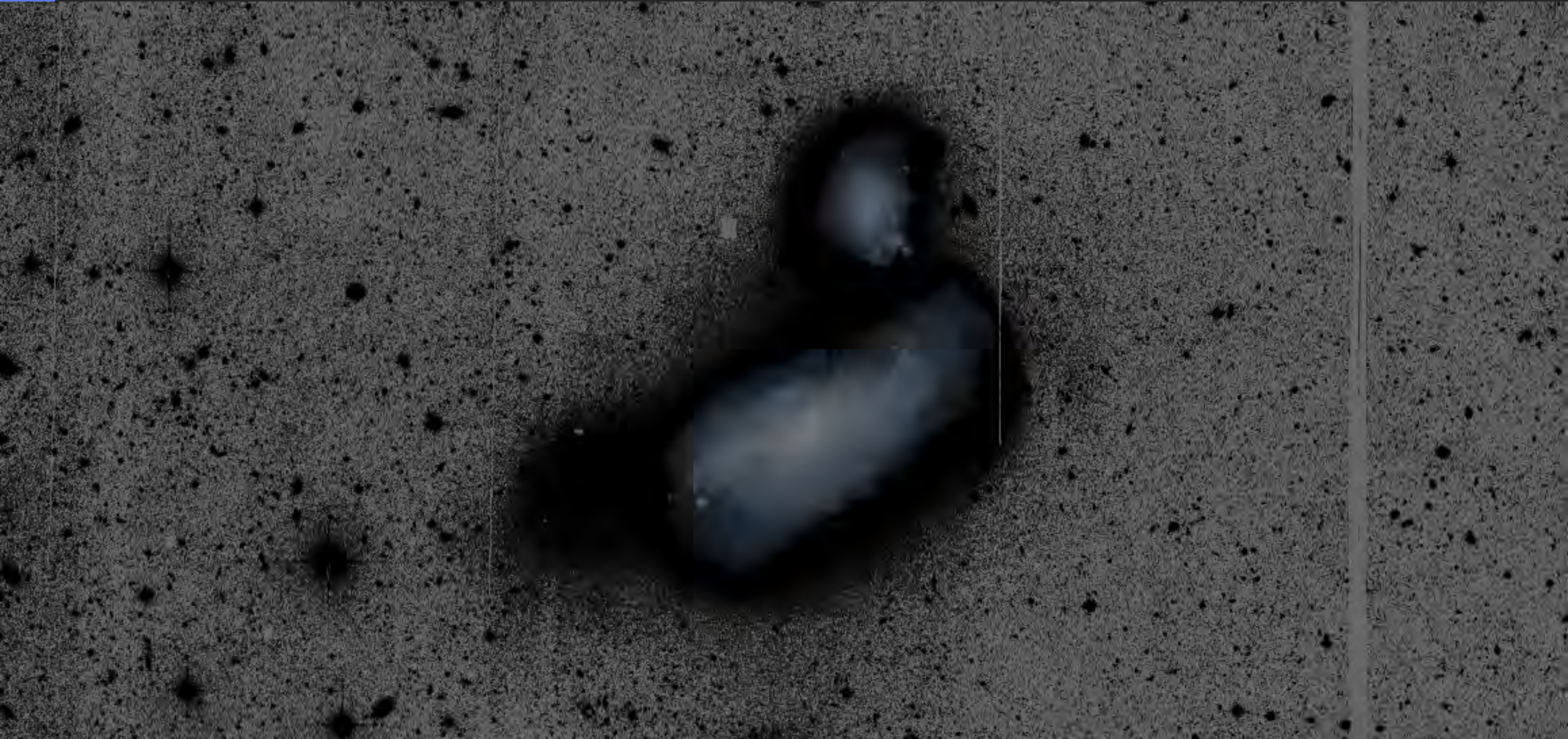


Figure 31: From Zarin Era et al. (2023), a demonstration of SAM's application in segmenting defects on a metal surface.

- SAM to identify galaxy truncations *(e.g. Jesus Vega-Ferrero+24)*



Tidal features and galactic growth



Tidal features and galactic growth

- NB: the following slides are biased towards some of my results, biased towards diffuse light (unresolved), and they are not exhaustive
- Non-exhaustive list of tidal feature properties that can be studied:
  - Census
  - Geometry
  - Photometry
  - Colour…

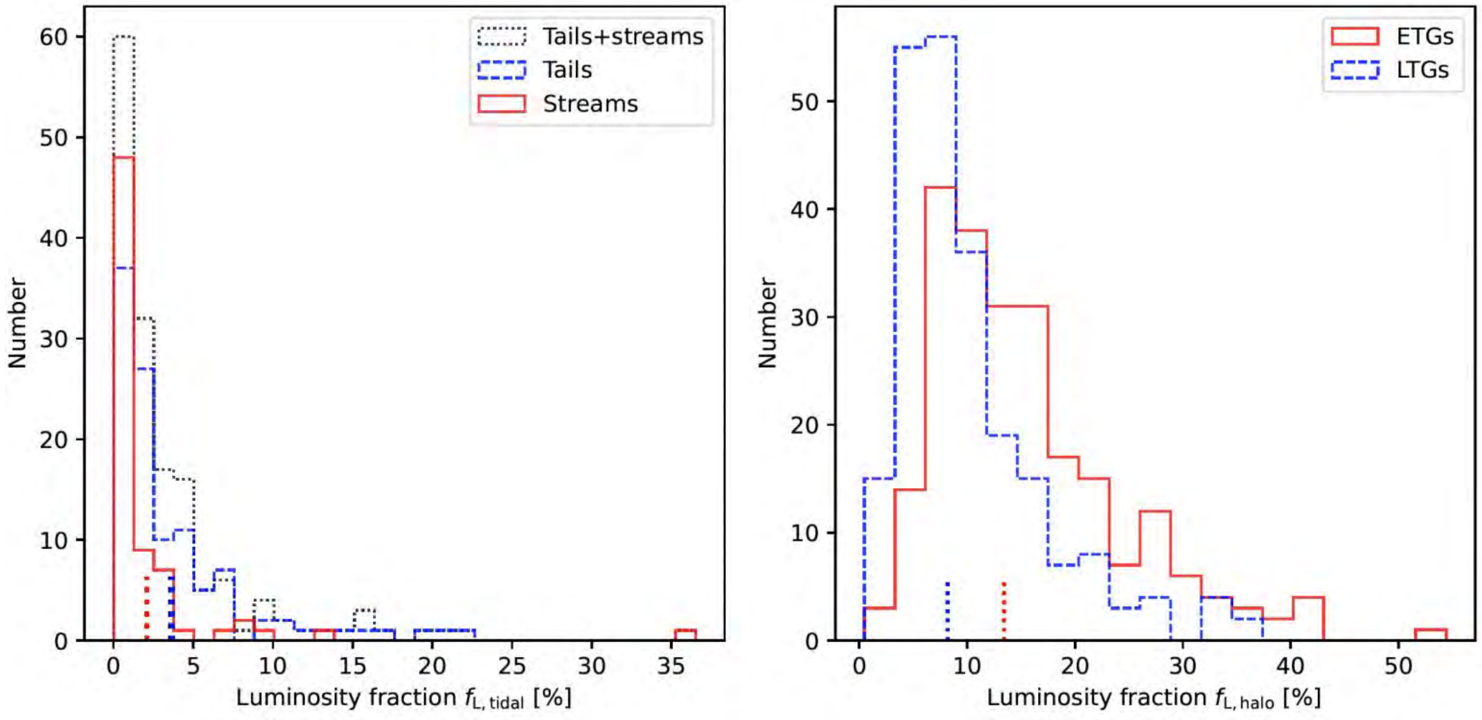
VS:

- Mass
- Environment
- Morphological type
- Kinematics
- …

Global statistical results

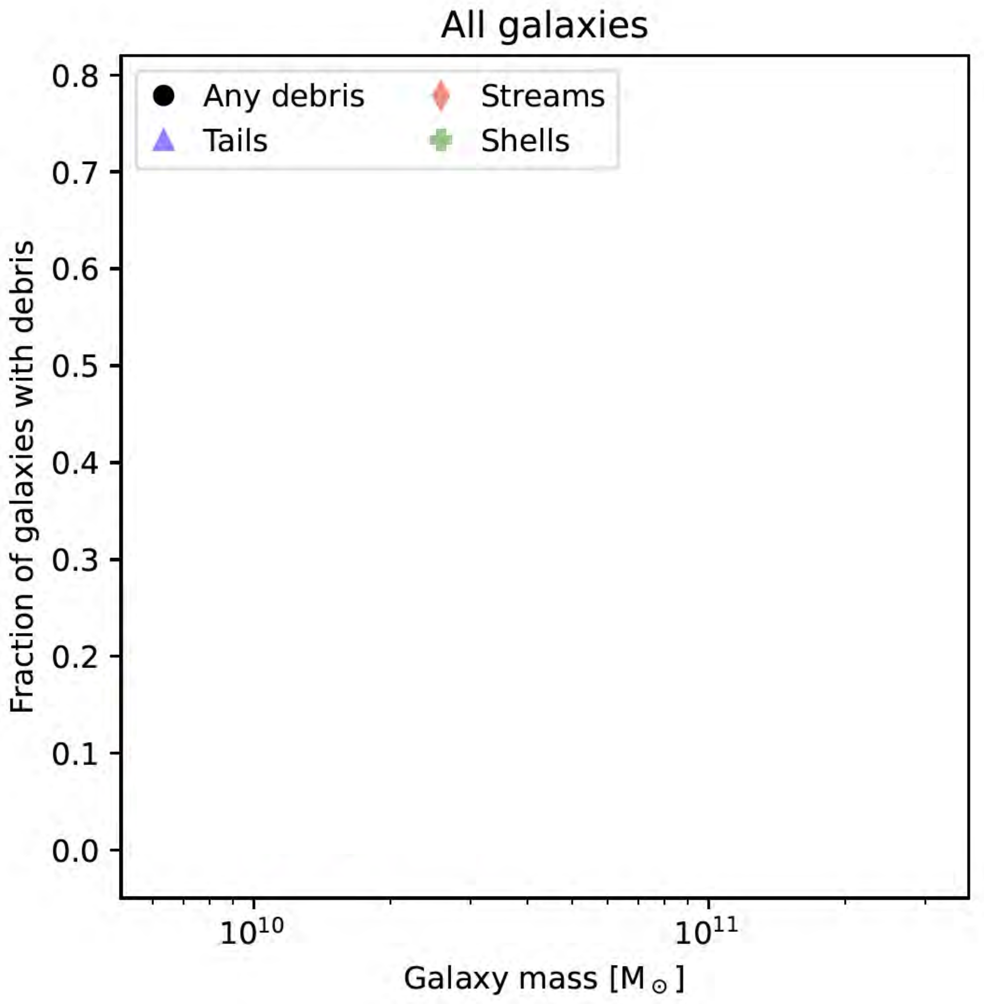
- *Sola+2022,2025*: study tidal features of 475 nearby (D<45Mpc), massive galaxies
- Database of +10,000 annotated features, including **199 tails, 100 streams and 455 haloes**
- **36%** of galaxies host tidal features: 23% host tails, 15% streams and 12% shells

- Tails+streams account for **2-4% of the total galaxy luminosity**, while haloes account for **10%**



Trends as a function of galaxy mass

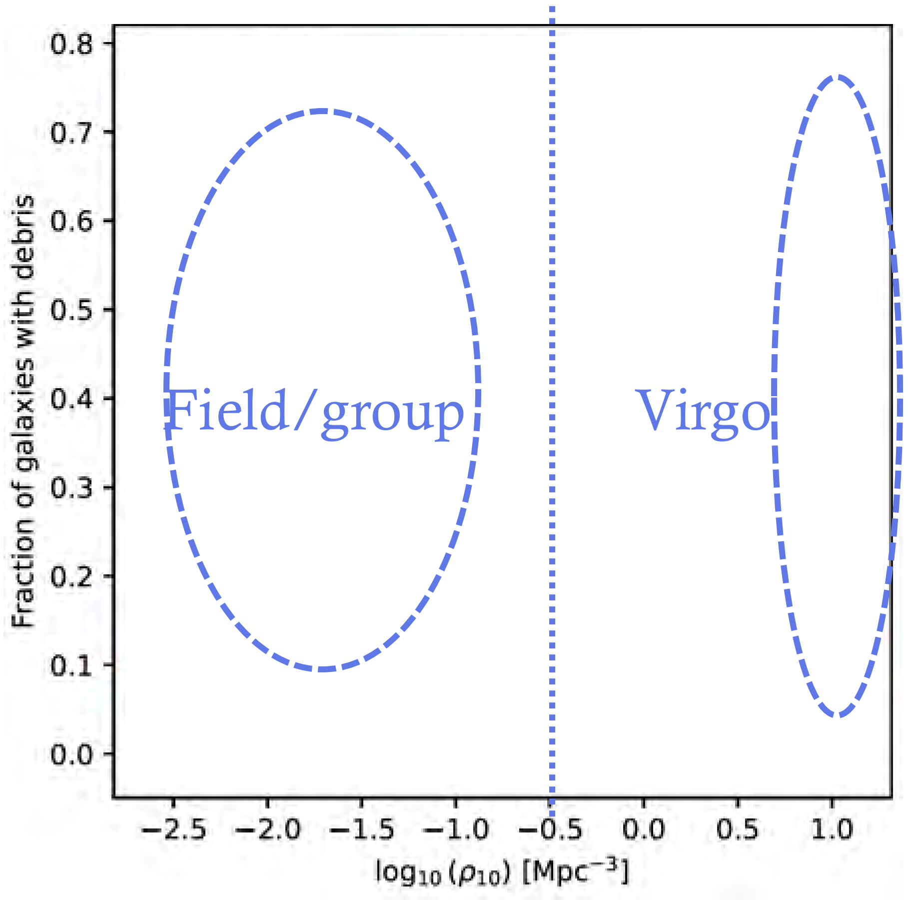
- The fraction of galaxies hosting tidal features more than **doubles** from 25% to 60% for the **highest-mass** galaxies.
- **Increase** visible for all **feature types**
- Similar trends for ETGs and LTGs separately



(Sola+ under review)

Trends as a function of the environment

- Large-scale environment: Virgo cluster vs field/group



Overall similar fractions of galaxies with debris in Virgo cluster and in field/group

But two peaks: on-going interacting galaxies ?

►Separate between isolated and in-pair galaxies

No peak, no trend with ρ_{10}

2 peaks: on-going interactions

(Sola+ under review)

LSB features vs galaxy mass and environment

Mass

- **More massive** galaxies have **more tidal features**
- **More massive ETGs** have **more luminous haloes** (not LTGs)
- Presence of a mass threshold ($\sim$4-7$\times 10^{10}$ $M_{\odot}$) above which the increase is steeper

► Support the **hierarchical** paradigm: more massive galaxies have undergone more merger events and their assembly is driven by mergers (rather than accretion)

Environment

- Large-scale environment: does not impact much tidal features
- **Small-scale environment** (presence of companion): mergers between interacting galaxies trigger tidal features
- **LTGs** have more extended haloes in the Virgo **cluster** (no evolution for ETGs): sign of **ram-pressure stripping**

► Galaxy **mass** seems to be the dominant factor that affects tidal features and haloes

Comparison to literature

- **Comparison** between works **far from being straightforward**: not the same image depth, definition of what a tidal feature is, not the same methods, not the same metrics *(e.g., Atkinson+2013, Hood+2018)* …
- Can lead to **large discrepancies**: e.g., the fraction of galaxies with debris can range from a few percents up to 70% *(e.g., Malin & Carter 1983, Bridge+2010, Miskolzi+2011, Atkinson+2013, Bílek+2020, Yoon+2023…)*
- But still the **increase** of the fraction of galaxies with debris and their luminosities **with stellar mass** has also been reported in observations and simulations *(e.g., Atkinson et al. 2013; Duc et al. 2015; Bílek et al. 2020; Yoon & Lim, 2020; Huang & Fan 2022; Vázquez-Mata et al. 2022, Rodriguez-Gomez et al. 2015; Martin et al. 2022)*
- Previous slides: examples of how tidal features can be used to probe galactic assembly. But what else can we learn ?

Tidal features and dark matter haloes

- Streams: useful probes of the properties of the **Dark Matter** (DM) haloes surrounding galaxies. With 6D information ► strong constraints can be put on DM shape and mass *(e.g., Springel and White 1999, Dubinski+1999, Ibata+2001, Helmi+2004, Johnston+2005, Koposov+2010, Bovy+2016, Koposov+2023….)*

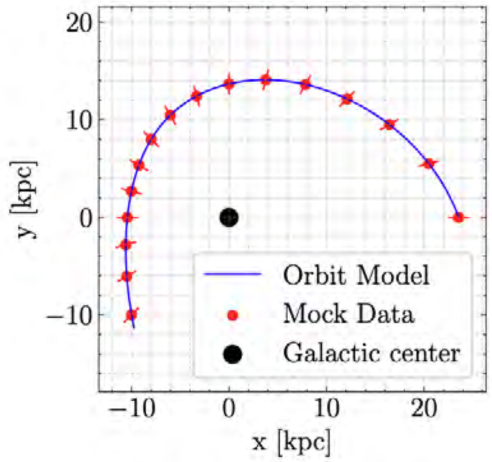
- What about streams with only **2D information** (spatial position + photometry) ?
  - ➔ *Nibauer+ 2023* : light constraint on DM shape only from 2D position
  - ➔ What about a population of streams ?

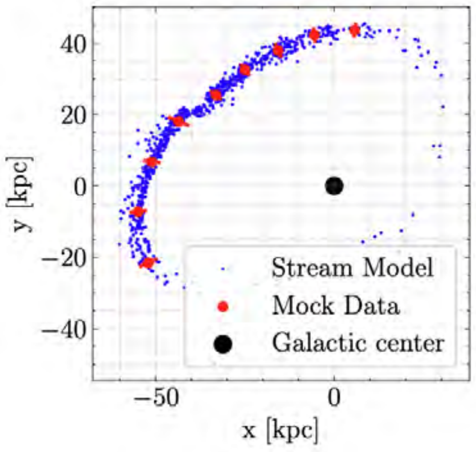
- *Chemaly+, in prep*: constrain the flattening of DM haloes from 2D information of streams (position + photometry), at a population level, using hierarchical Bayesian inference

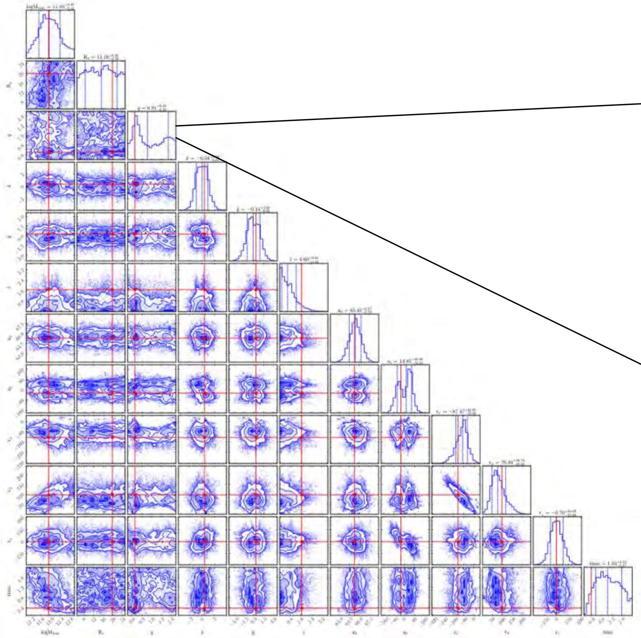
Streams and DM flattening

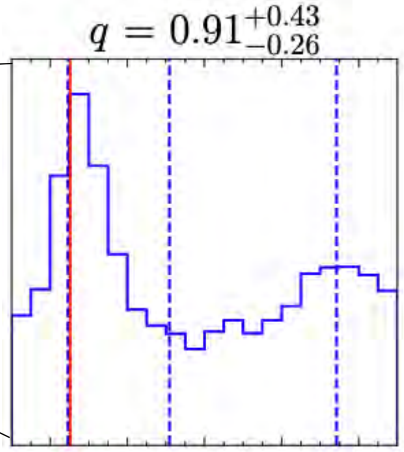
Mock data → Hierarchical Bayesian Inference ← Model

Either orbits or streams









Posterior on the flattening for only one orbit

► Next: apply to real data on a large sample of 'best'-looking streams (long, curved, faint) but no consensus in the literature, no large catalogue of these streams

► *Sola+, in prep*: catalogue of tidal features and modelable streams in DESI Legacy Imaging surveys

Summary

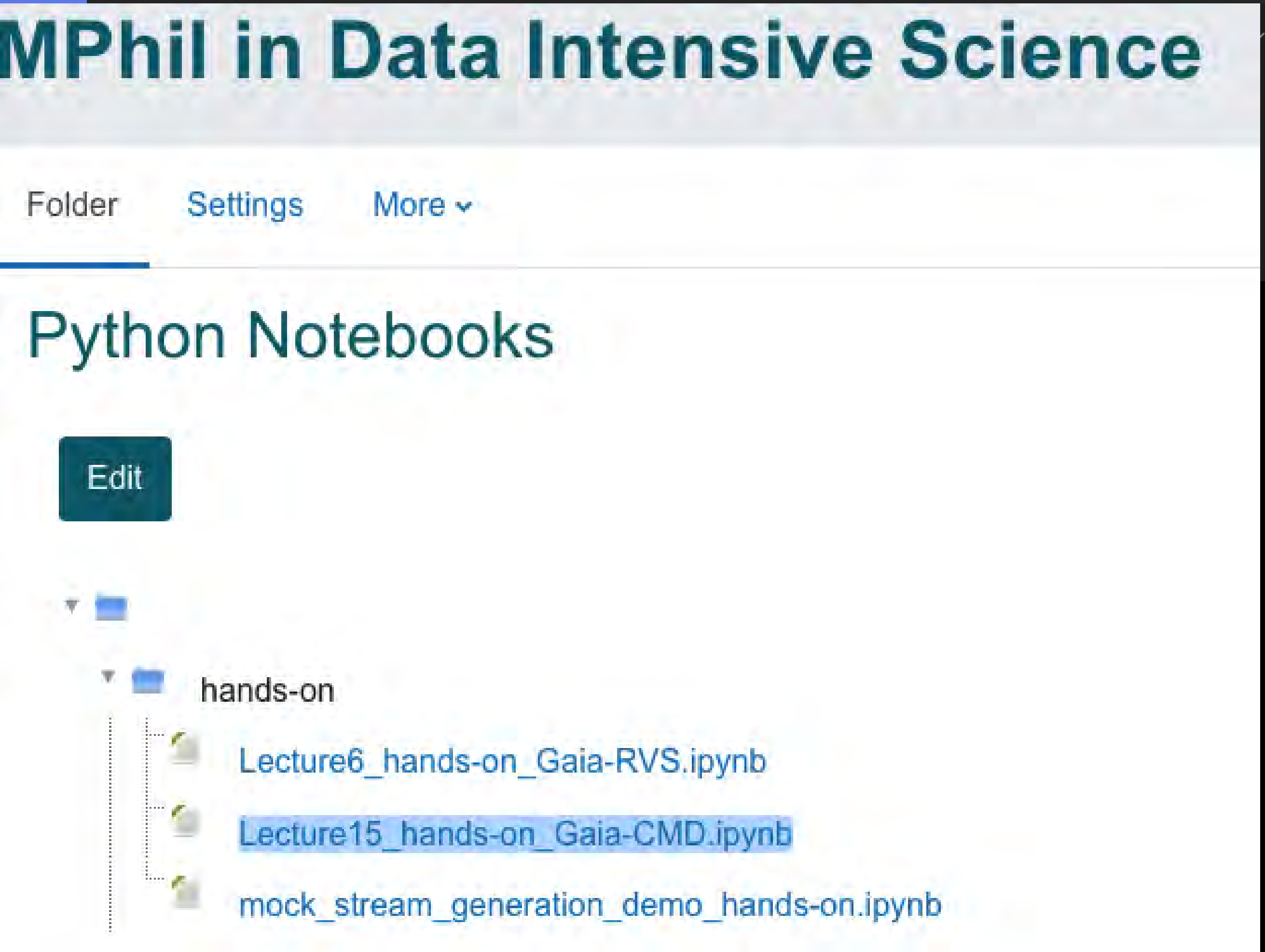
- Future/ongoing large sky surveys, depth multi-band, great resolution ▶push further the limit of the LSB Universe
- **Euclid**: already proved its wonderful capabilities with the Early Release Observations, soon the Quick data release (Q1) will be public. So many papers about LSB science !
- Tidal features in deep images found mostly by visual inspection…
- … but impossible to do in ongoing/future large surveys: need **machine learning**
- CNN-based networks largely used but other methods, including using foundation models
- Tidal features fraction, geometry and luminosity give clues about the **late mass assembly** of galaxies and can be used to constrain **dark matter haloes**’ shapes

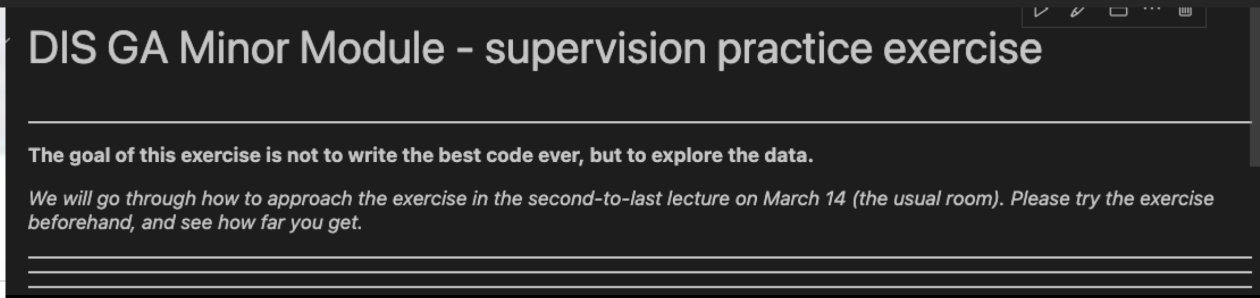
Be ready for the last lecture !

WikiCommons

- Install SAOImage **DS9**, **TOPCAT** and **Aladin** desktop
  (at least look up at Aladin Lite online)
  https://www.star.bris.ac.uk/~mbt/topcat/
  https://aladin.cds.unistra.fr/java/nph-aladin.pl?frame=downloading
  https://sites.google.com/cfa.harvard.edu/saoimageds9
- Look up at **NED** and **SIMBAD**
- Have **Python** ready with astropy and astroquery packages installed (+ other standard packages) + **Jupyter** notebook
- Bring your computer !

Supervision session on 14/03





- Try to do as much as you can for this supervision practice exercice
- Lecture15_hands-on_Gaia-CMD.ipynb