Deep learning tools for detecting and segmenting galactic structures and contaminants in low surface brightness images

Adeline Paiement, Pierre-Alain Duc, Felix Richards, Elisabeth Sola, Renaud Vancoellié, Worachit Ketrungsri

Observatoire astronomique de Strasbourg

Swansea University Prifysgol Abertawe

Observatoire astronomique de Strasbourg

ioa

LIS Laboratoire d'Informatique & Systèmes

Université de Toulon

LIS Laboratoire d'Informatique & Systèmes

Université de Toulon

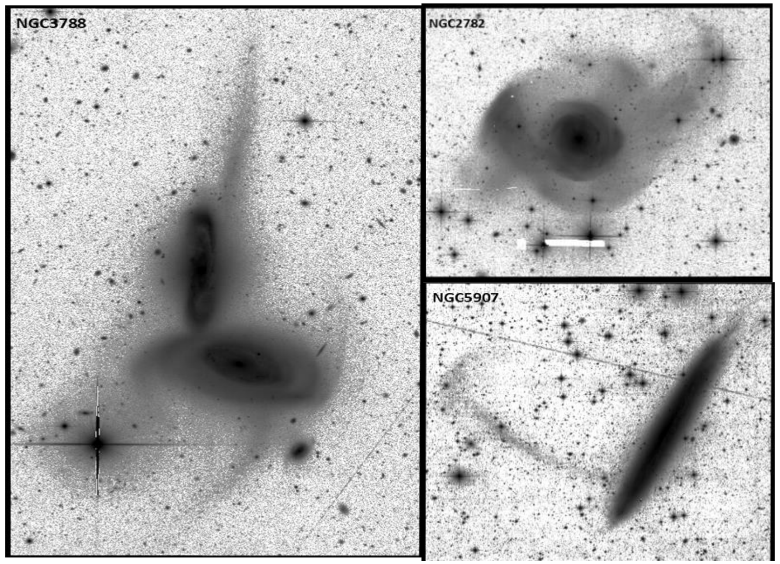
Galaxy morphology: different levels of analysis

A. Classification of **morphology types**

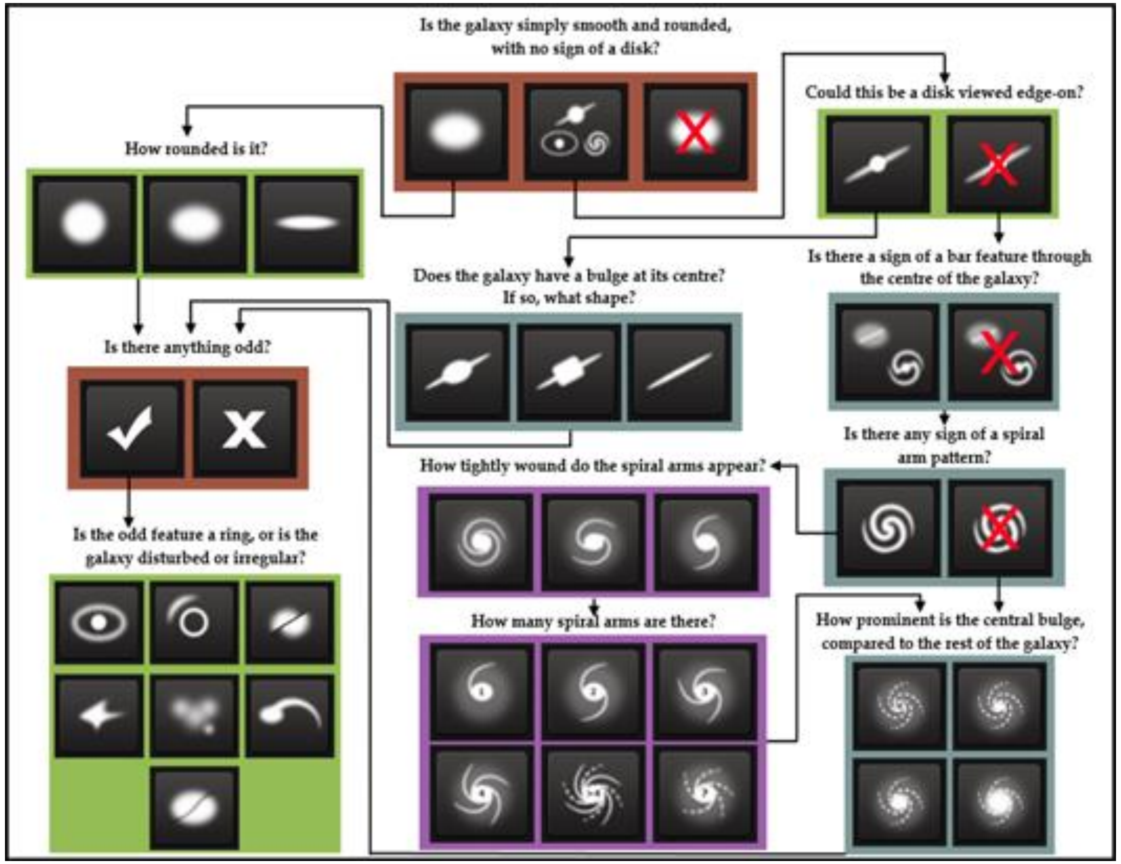
B. Classification/regression of key **morphology parameters**

C. Identification of **low brightness tidal features**

- → insights into the galaxy's evolution history



Examples of tidal features in CFIS images.



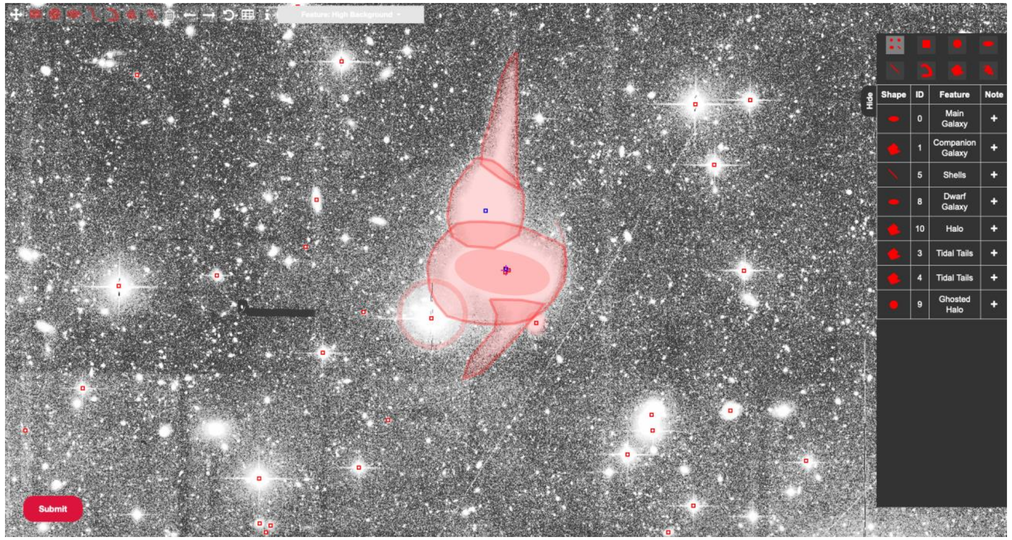
Galaxy Zoo analysis process

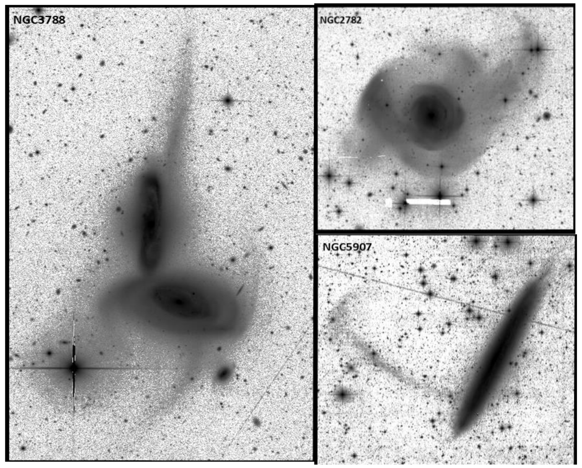
Tidal features: different levels of analysis

A. Classification of **presence vs. absence**
   - ➢ Weak constraint on the tool's focus

B. **Fine localisation**
   - ➢ Strong focus on the tidal structures

1. Bounding box detection
2. Segmentation: pixel-wise localisation





Examples of tidal features in CFIS images.

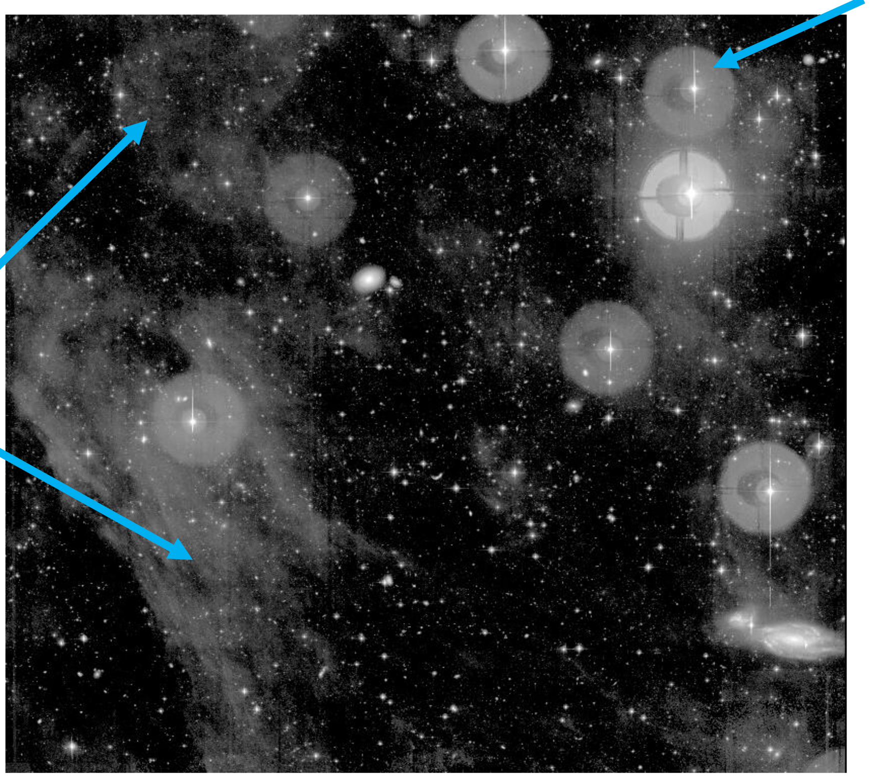
Challenges of noisy & crowded images

Imaging reveals low surface brightness tidal structures…

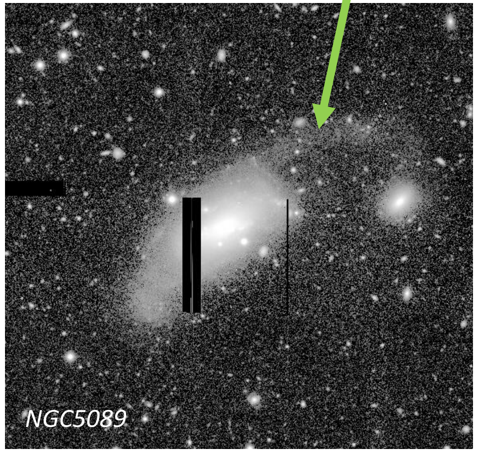
… but also dust clouds (cirrus) and imaging artefacts

Ghost halos
(diffraction figures around stars)

Cirrus



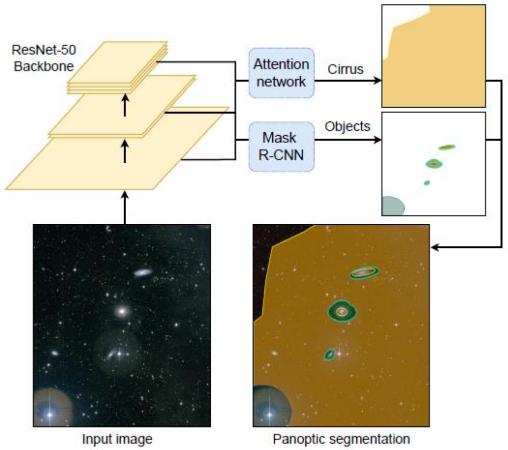
Elongated tidal feature



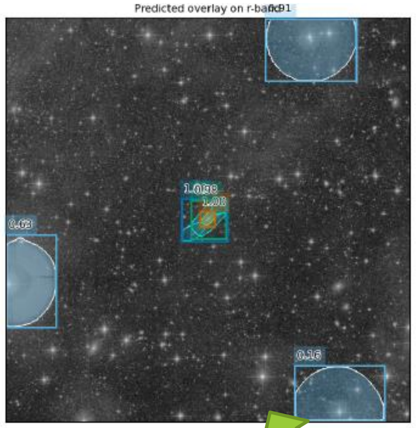
Images from the MATLAS survey (Mass Assembly of early-Type GaLAxies with their fine Structures), CFHT MegaCam instrument

Method overview

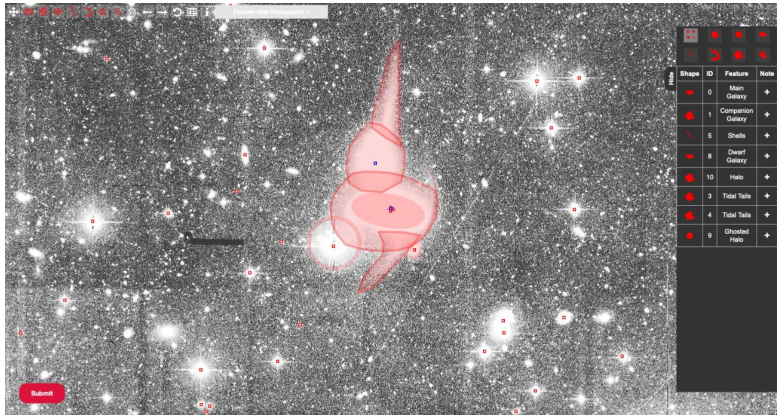
1. Purpose-designed neural network



3. **Combined** detection and segmentation of galactic substructures and image contaminants



2. Dataset creation



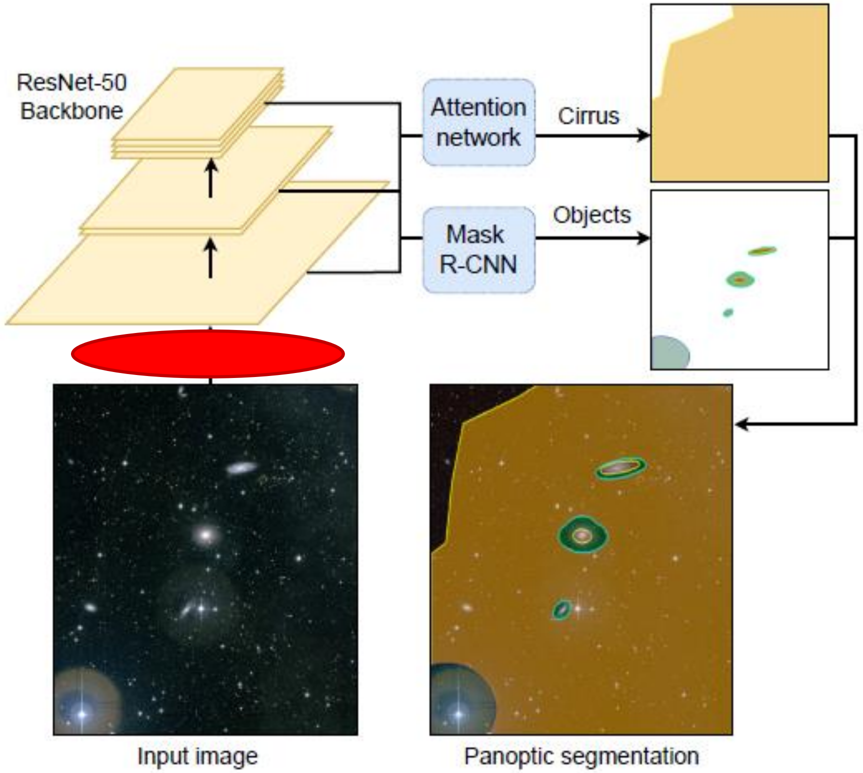
Purpose-designed neural network

Combined detection and segmentation of:
- galactic features
- image contaminants

Helps distinguish between:
- tidal features and cirrus
- galaxy halos and ghost halos

New architecture:
- sensitivity to low brightness structures [1]
- sensitivity to the oriented textures of cirrus [2,3]

New training strategy:
- Consensus between annotators
- Human-in-the-loop training



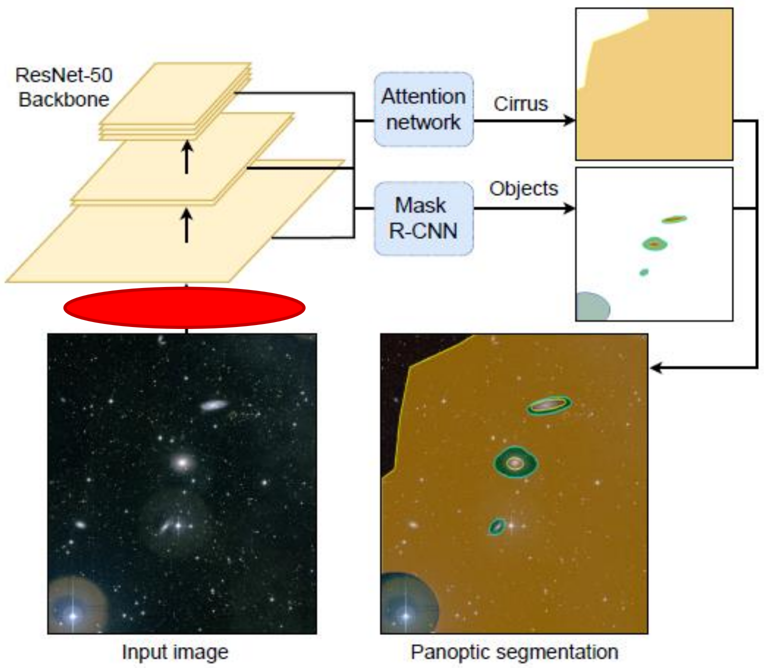
[1] F. Richards, A. Paiement, X. Xie, E. Sola, P.-A. Duc: Panoptic Segmentation of Galactic Structures in LSB Images. *International Conference on Machine Vision Applications (MVA)*, 2023
[2] F. Richards, E. Sola, A. Paiement, X. Xie, P.-A. Duc: Multi-scale gridded Gabor attention for cirrus segmentation. *IEEE International Conference on Image Processing (ICIP)*, 2022
[3] F. Richards, A. Paiement, X. Xie, P.-A. Duc: Learnable Gabor modulated complex-valued networks for orientation robustness. Under review with *Image and Vision Computing*, 2023

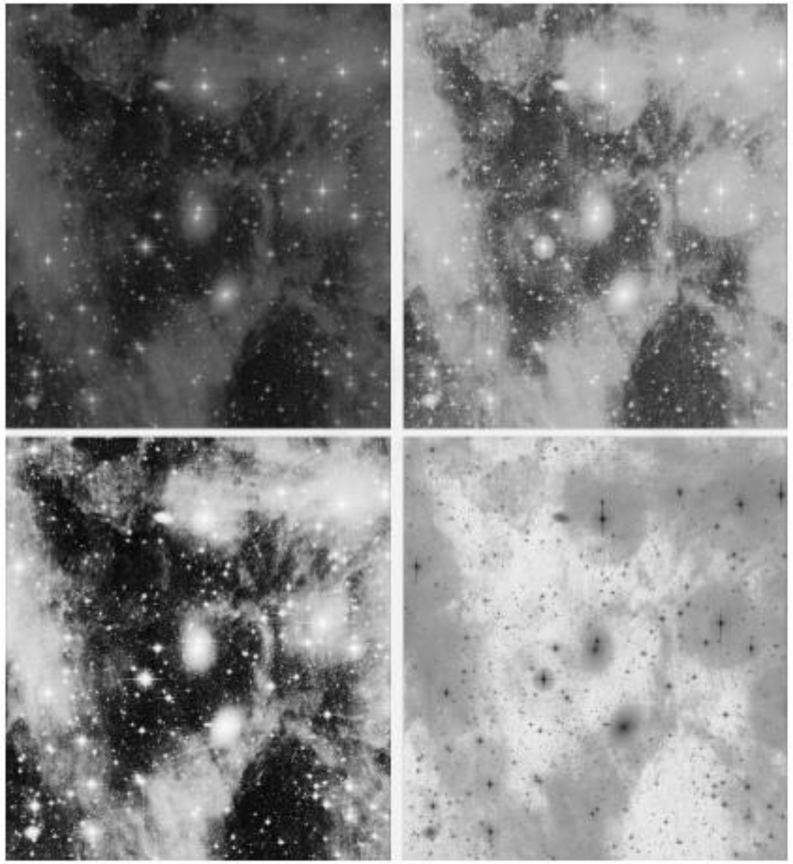
Adapting to low surface brightness images

A new pre-processing layer that adaptively scales image intensity: $X_s = \text{arcsinh}\,(aX + b)$, where $a, b \in \mathbb{R}$ are learned

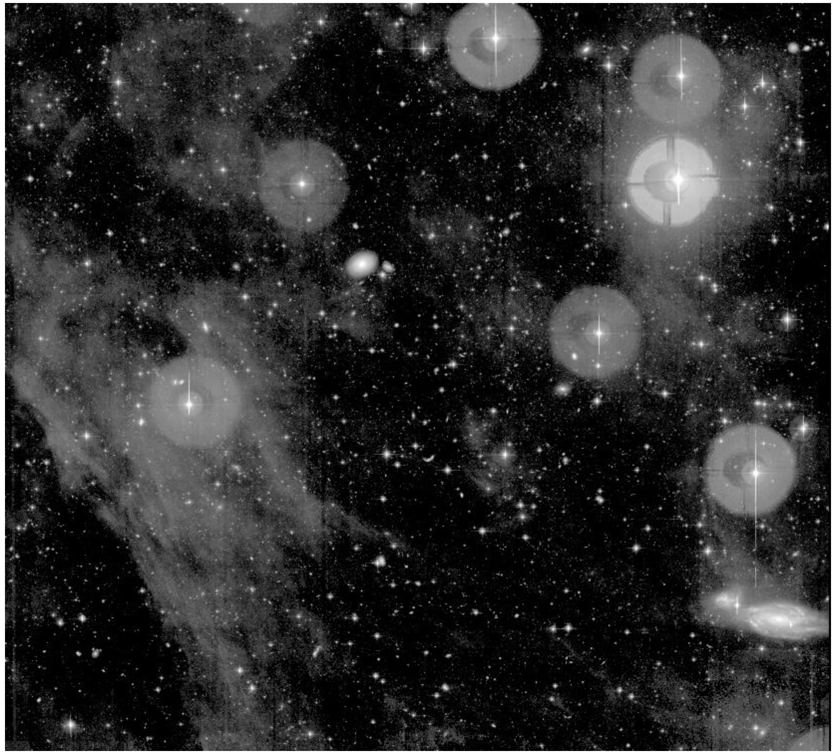
a, b are optimised for the entire dataset

➢ Discovers the portions of the image's dynamic range that are relevant to identify and distinguish cirrus and tidal features





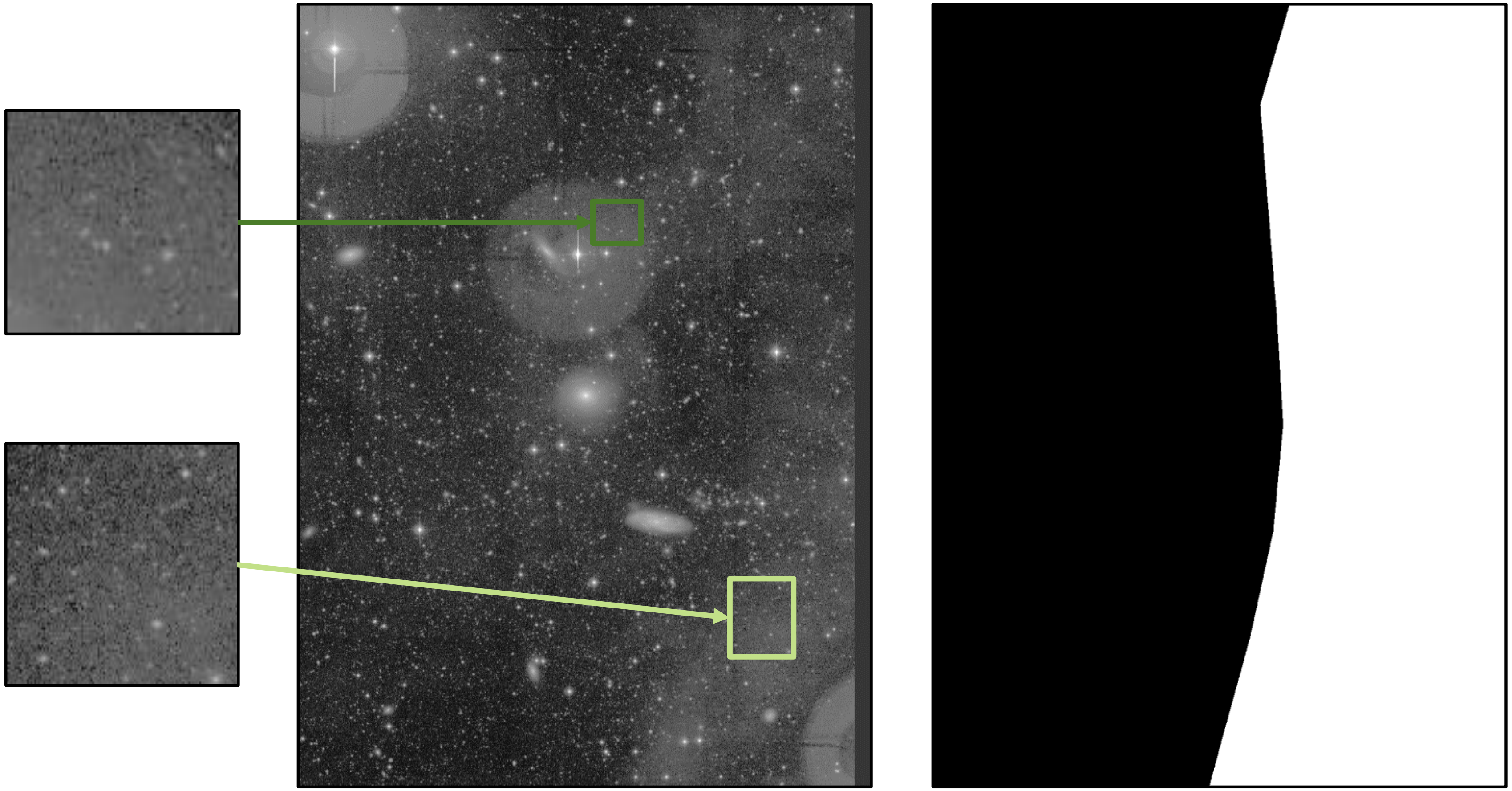
Detecting cirrus contaminants



Example images contaminated with cirrus (MATLAS).

Detecting cirrus contaminants

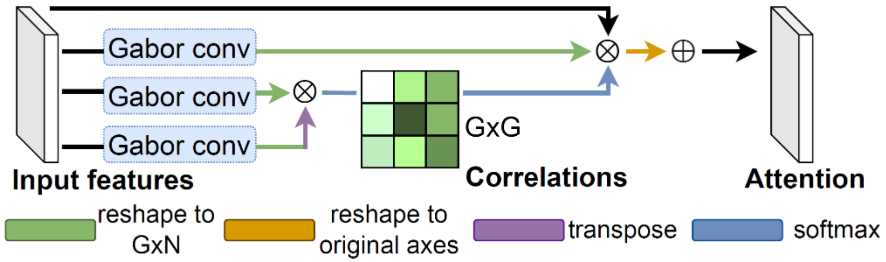
Precise segmentation of cirrus requires ample **global context** alongside understanding of **textural patterns**



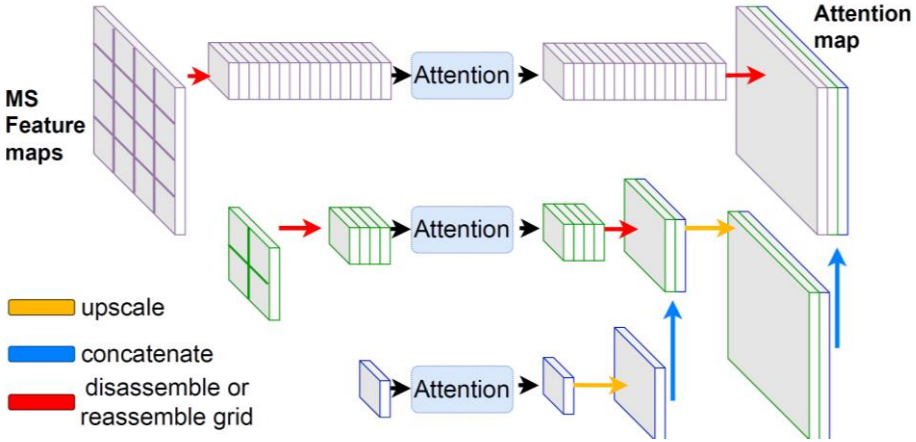
Comparison of localised regions (left), cirrus segmentation label (right)

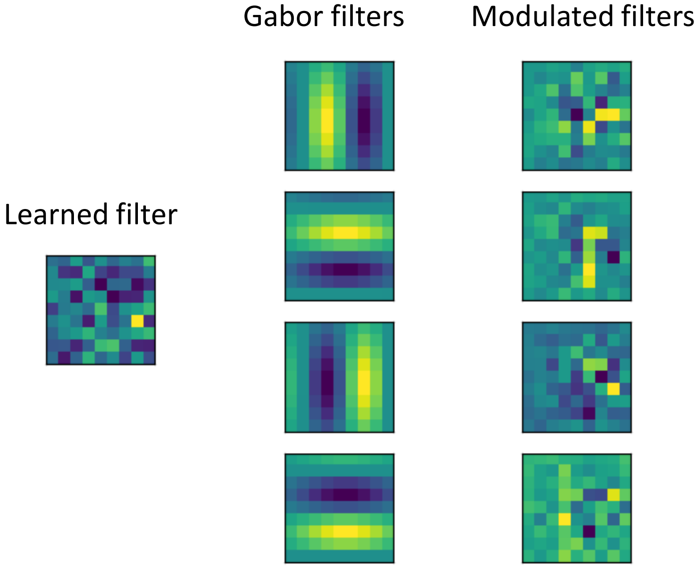
Detecting cirrus contaminants: Sensitivity to large-scale oriented textures

1. Orientation-sensitive attention module



2. Multi-scale analysis

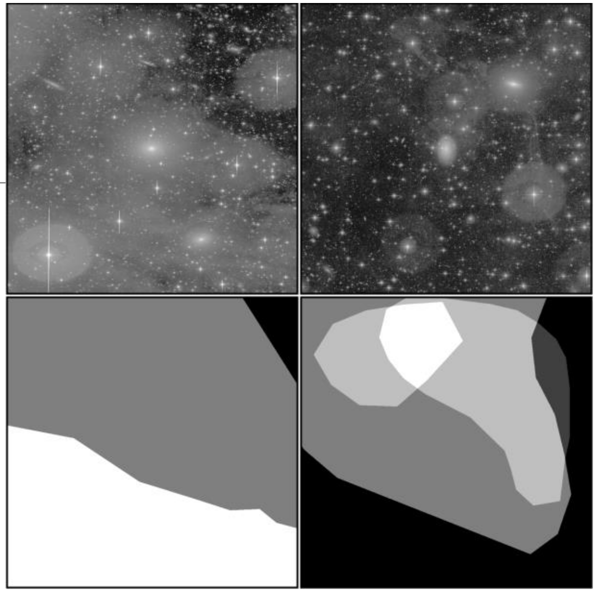




Example of learned convolutional filter modulated with four Gabor filters of varying orientation.

- Each branch handles a spatial scale
  - ➢ Local and global context

Training with uncertain labels



- Consensus between annotators is often not perfect
- Labels may be considered as probabilistic
- Example for cirrus contaminants:

Focal loss L_f puts focus on difficult examples

$$L_{\mathsf{SML}} = \begin{cases} \beta \cdot L_f(x, y) & \text{if } y \geq 0.75. \\ L_f(x, y) & \text{if } 0.5 \leq y < 0.75. \\ 0 & \text{if } 0.25 < y < 0.5. \\ L_f(x, y) & \text{otherwise.} \end{cases}$$

Definitely cirrus: super majority consensuses are prioritised ($\beta = 1.25$)

Probably cirrus

Uncertain targets are ignored

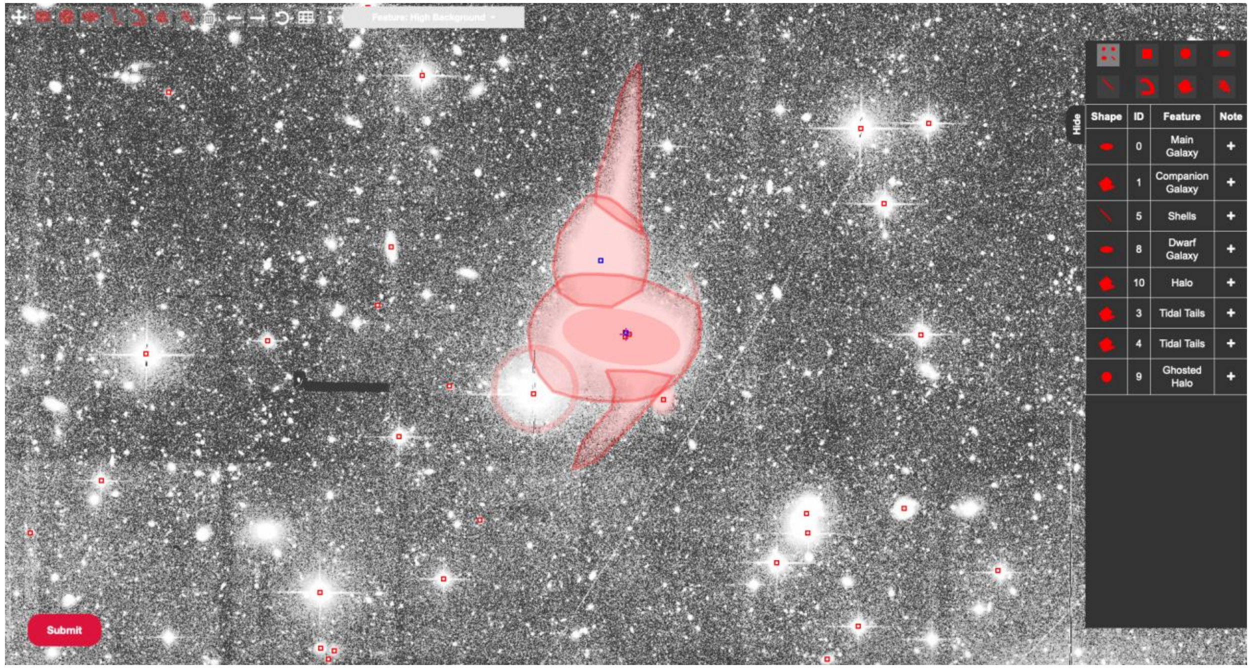
Probably not cirrus

- Tidal features: see Renaud Vancoellie's flash talk on Wednesday (11:45)

Dataset creation

Home-made annotation tool for LSB structures

- **Online** tool
- Based on Aladin Lite
- Goal: **Draw** with precision the **shapes** of LSB structures

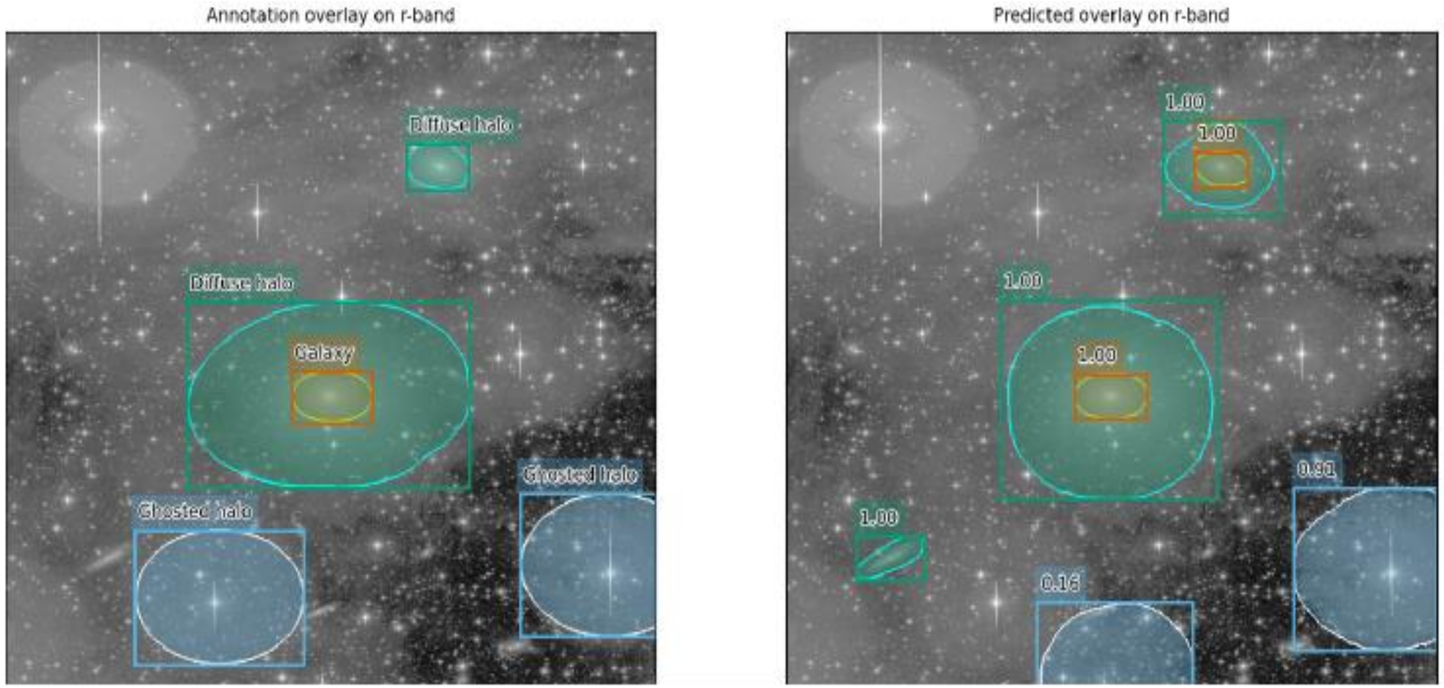


➢ **Dataset** of manually labelled LSB structures and image contaminants
  - 186 MATLAS LSB images (6000px², two spectral channels)
  - On average 1.7 (std 0.9) galaxies annotated per image

[4] E. Sola, P.-A. Duc, F. Richards, A. Paiement, M. Urbano, J. Klehammer, M. Bílek, J.-C. Cuillandre, S. Gwyn, A. McConnachie: Characterization of Low Surface Brightness structures in annotated deep images, Astronomy & Astrophysics, 662, A124, 2022

Semi-automatic augmentation of the dataset

Human in the loop training

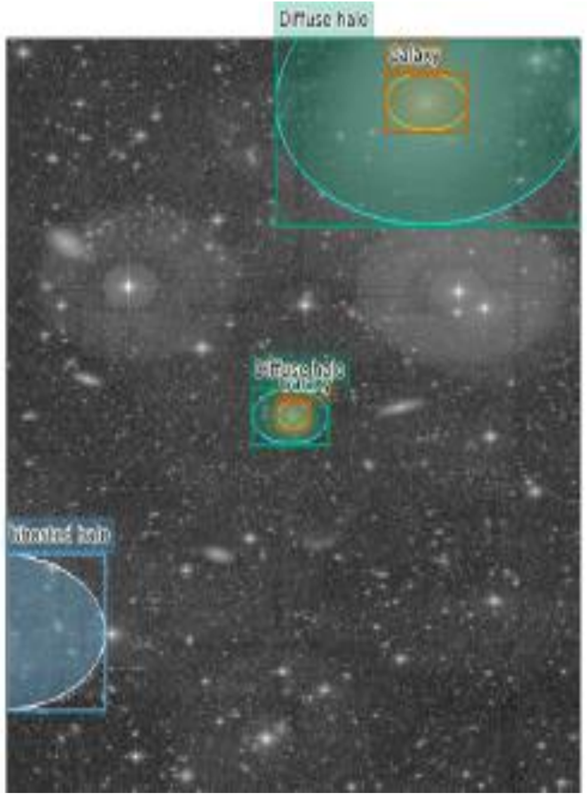


1. Train our deep neural network on the manual annotations
2. Manually select true detections that were not manually labelled
3. Re-train with more complete dataset
4. Repeat if needed, until convergence

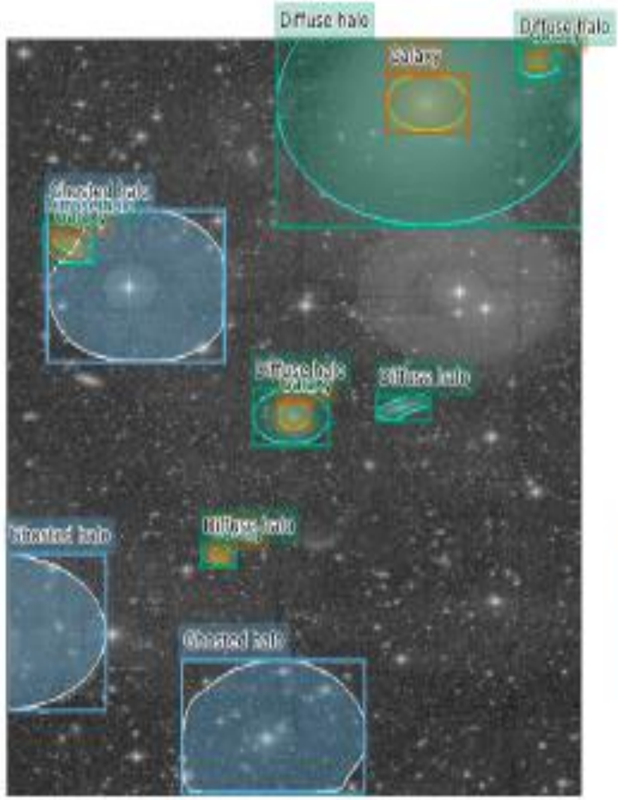
Semi-automatic augmentation of the dataset

Human in the loop training

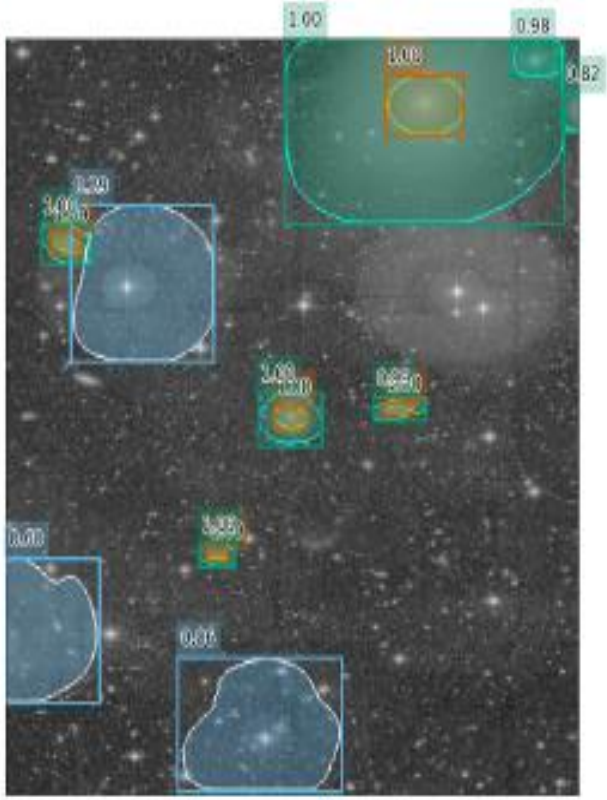
Example of results



Initial manual annotations



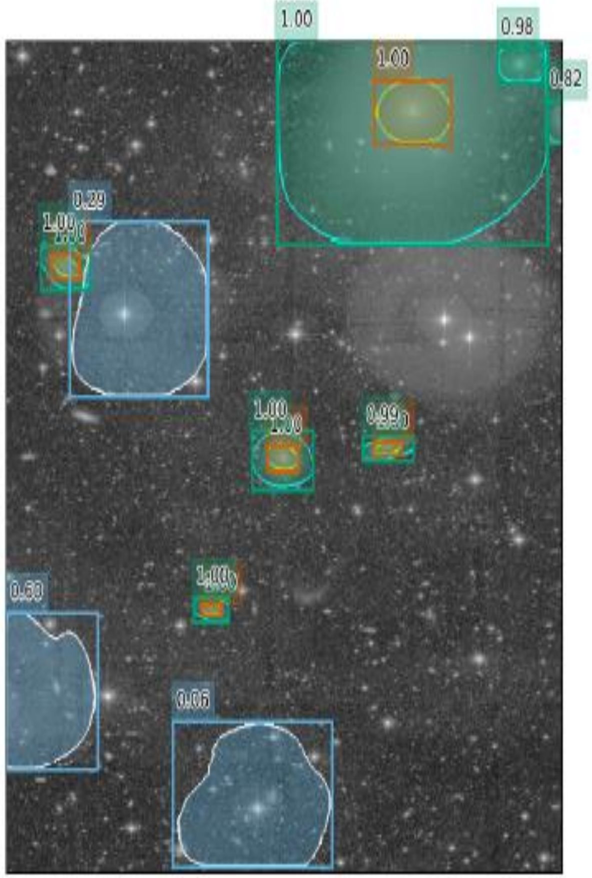
HITL augmented annotations

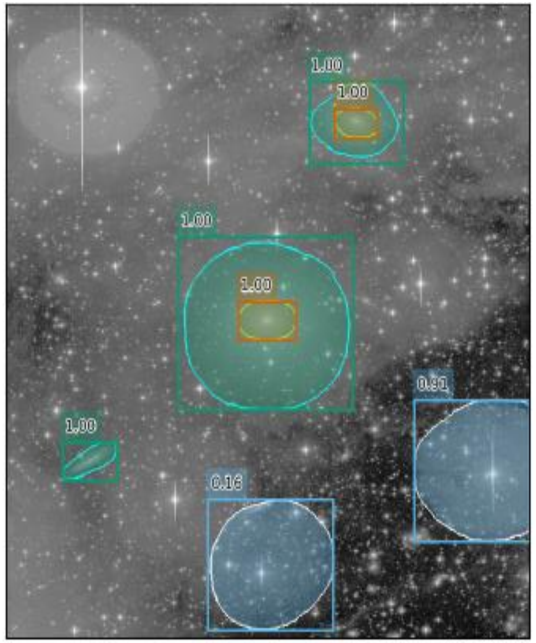


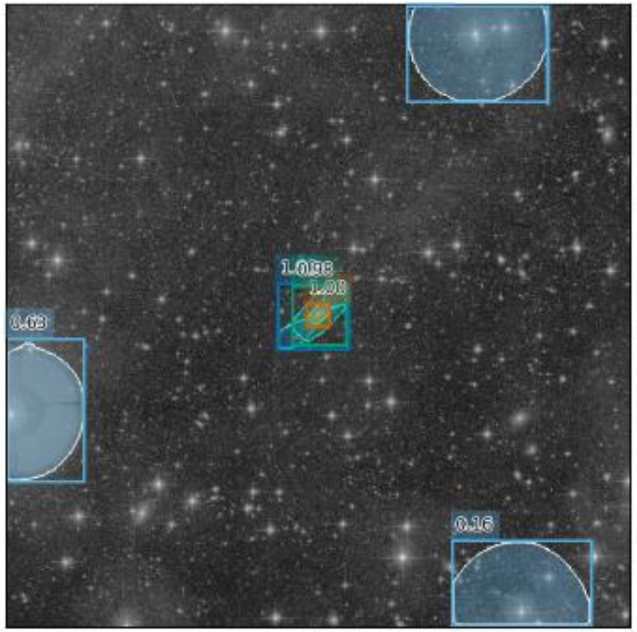
Final DNN detections

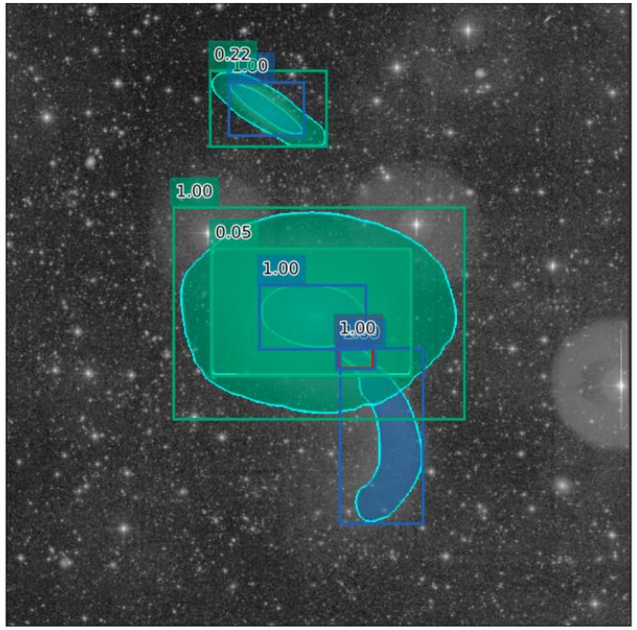
Red: inner galaxy
Green: extended halo
Dark blue: elongated tidal features (streams, tidal tails, plumes)
Light blue: ghosted halo (contaminant)

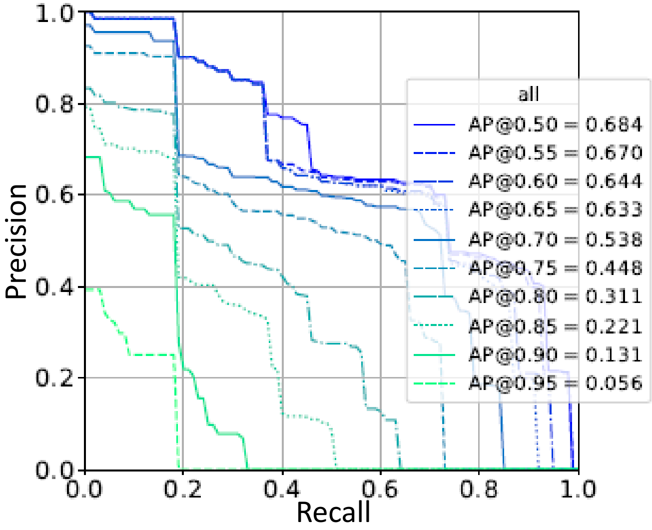
Example results



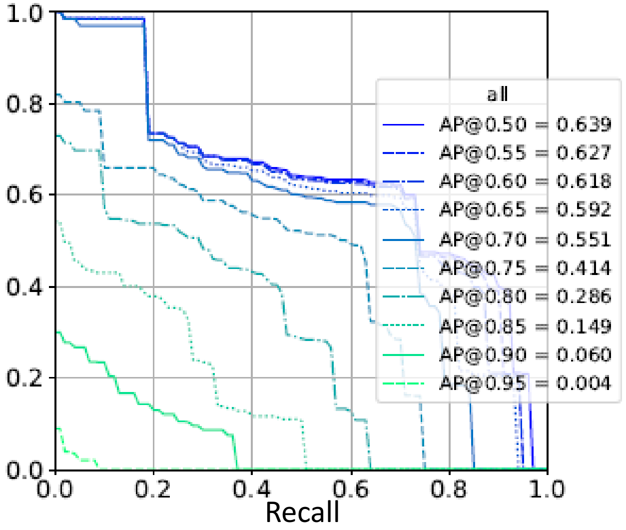








(a) Bounding box score



(b) Segmentation score

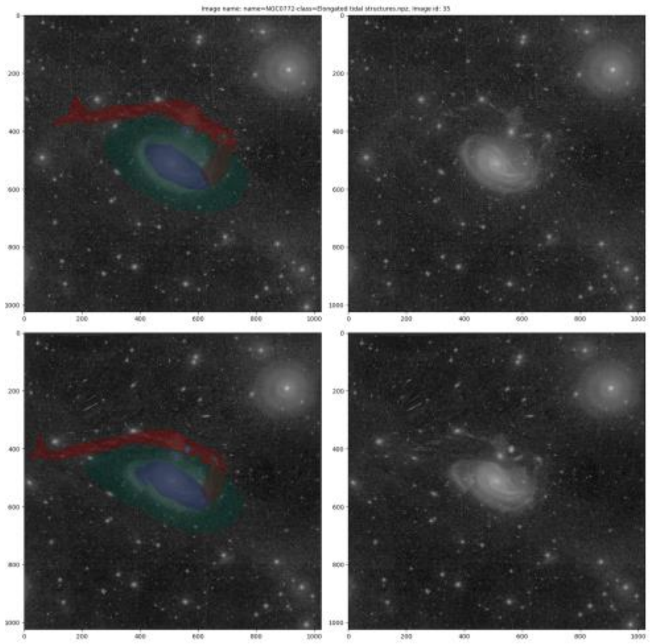
Red: inner galaxy
Green: extended halo
Dark blue: elongated tidal features
Light blue: ghosted halo (contaminant)
Cirrus not shown for readability

Next steps, work in progress

- Increase and balance the size of the dataset
  - Annotation of more images
  - Pre-training by self-supervision
  - Data augmentation

Original

Deformed



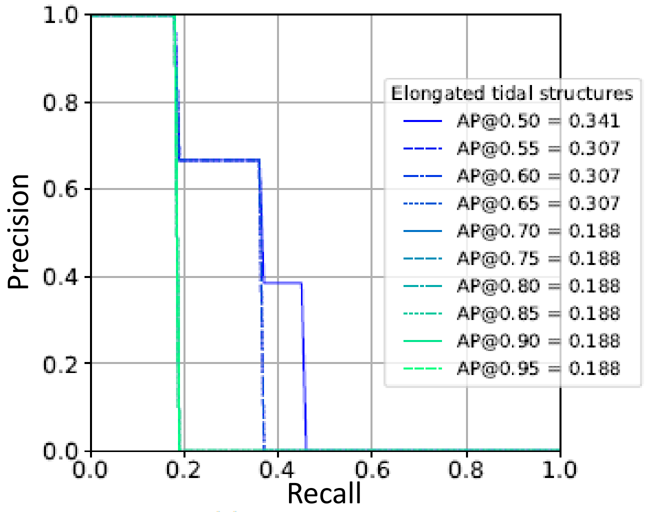
- Test on different surveys and image bands

Training set

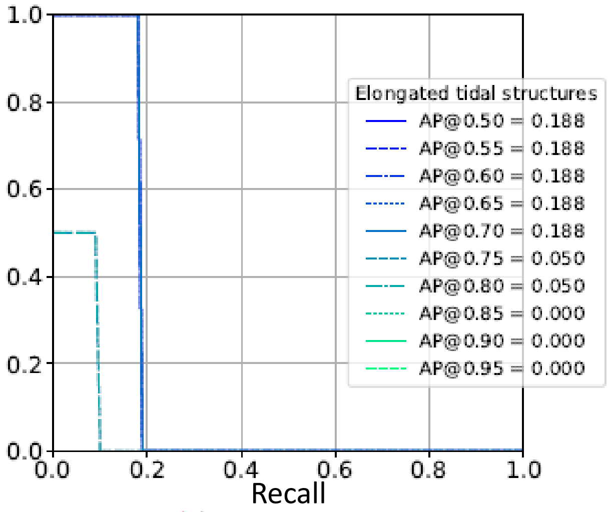
- 'Diffuse halo': 144 images (80%)
- 'Galaxy': 147 images (79.89%)
- 'Elongated tidal structures': 45 images (81.81%)
- 'Ghosted halo': 117 images (80.68

Testing set

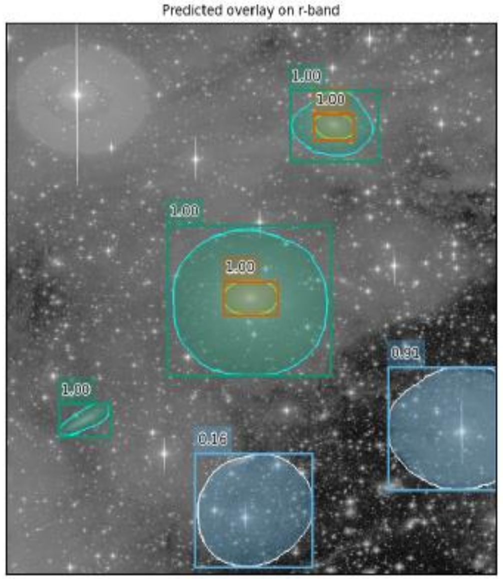
- 'Diffuse halo': 36 images (20%)
- 'Galaxy': 37 images (20.1%)
- 'Elongated tidal structures': 10 images (18.18%)
- 'Ghosted halo': 28 images (19.31%)

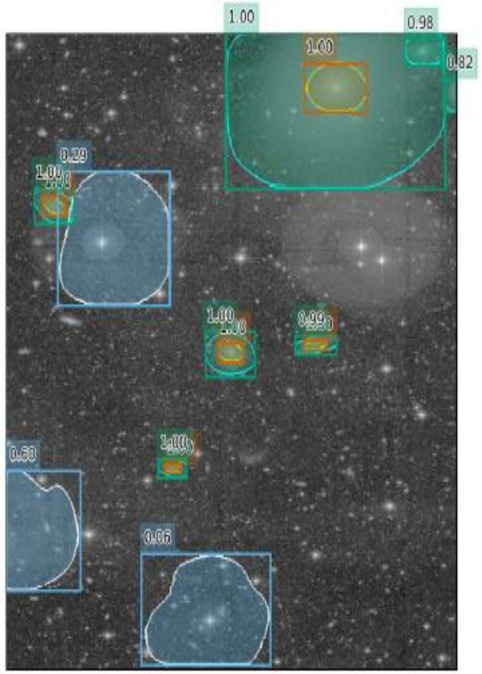


(a) Bounding box score



(b) Segmentation score





Thank you for your attention