Meeting the challenge of blending in LSST DESC

Cécile Roucelle & Bastien Arcelin

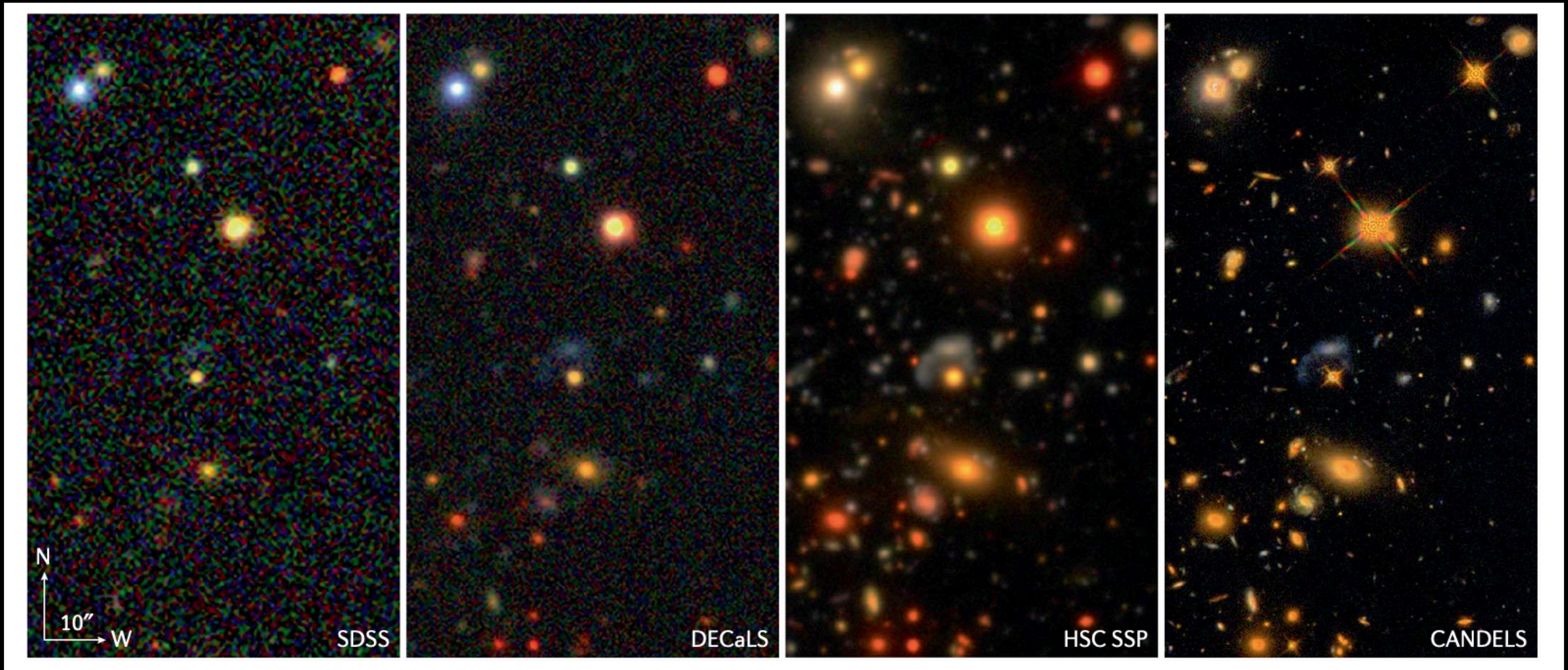
Blending ?

- Overlapping light sources on the line of sight (stars and galaxies)
- Nothing new but…

pixel scale 0.396 arcsec

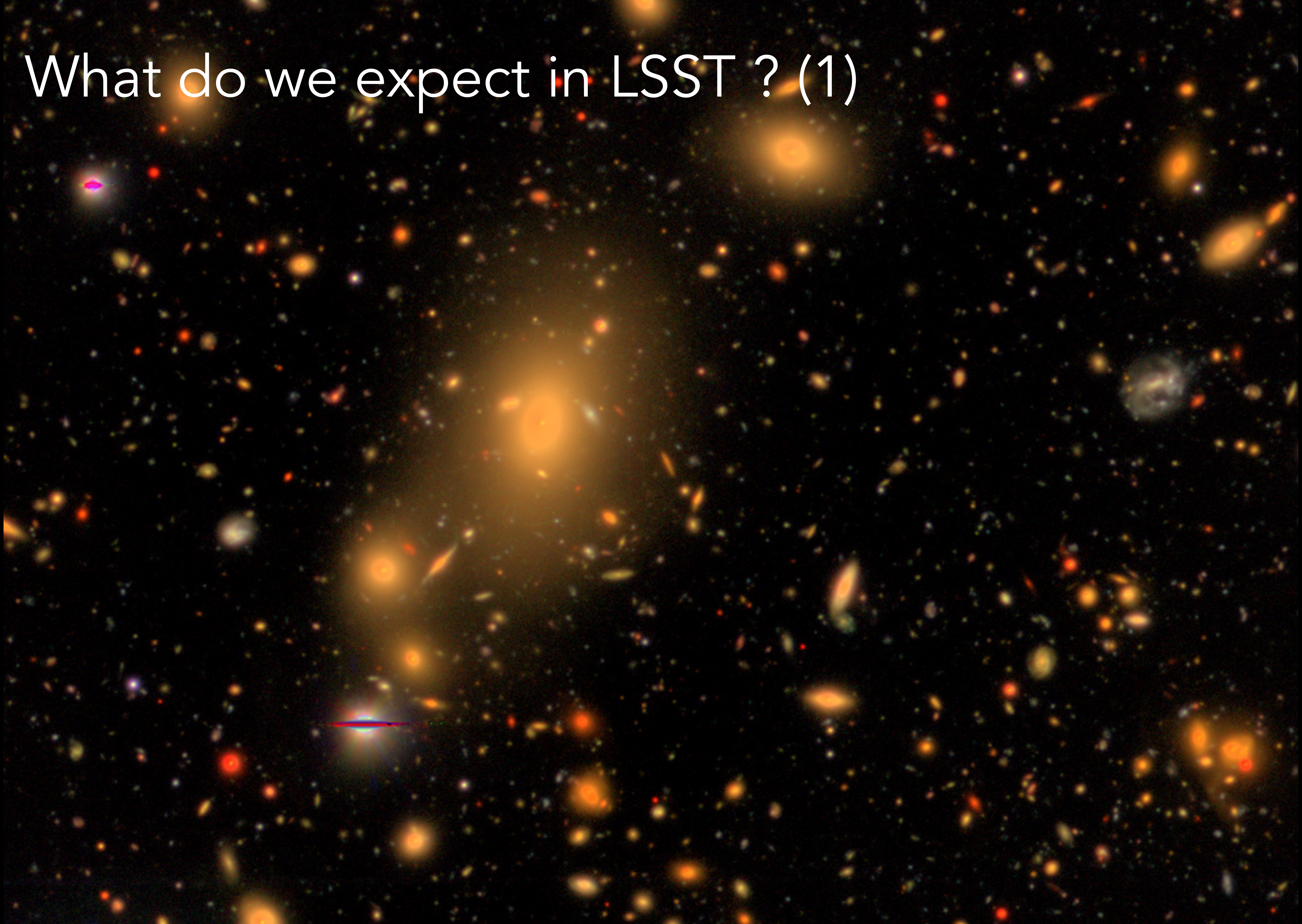
pixel scale 0.168 arcsec

pixel scale 0.06 arcsec



Credit R. Joseph et al *Nat Rev Phys* 3, 712–718 (2021)

What do we expect in LSST ? (1)

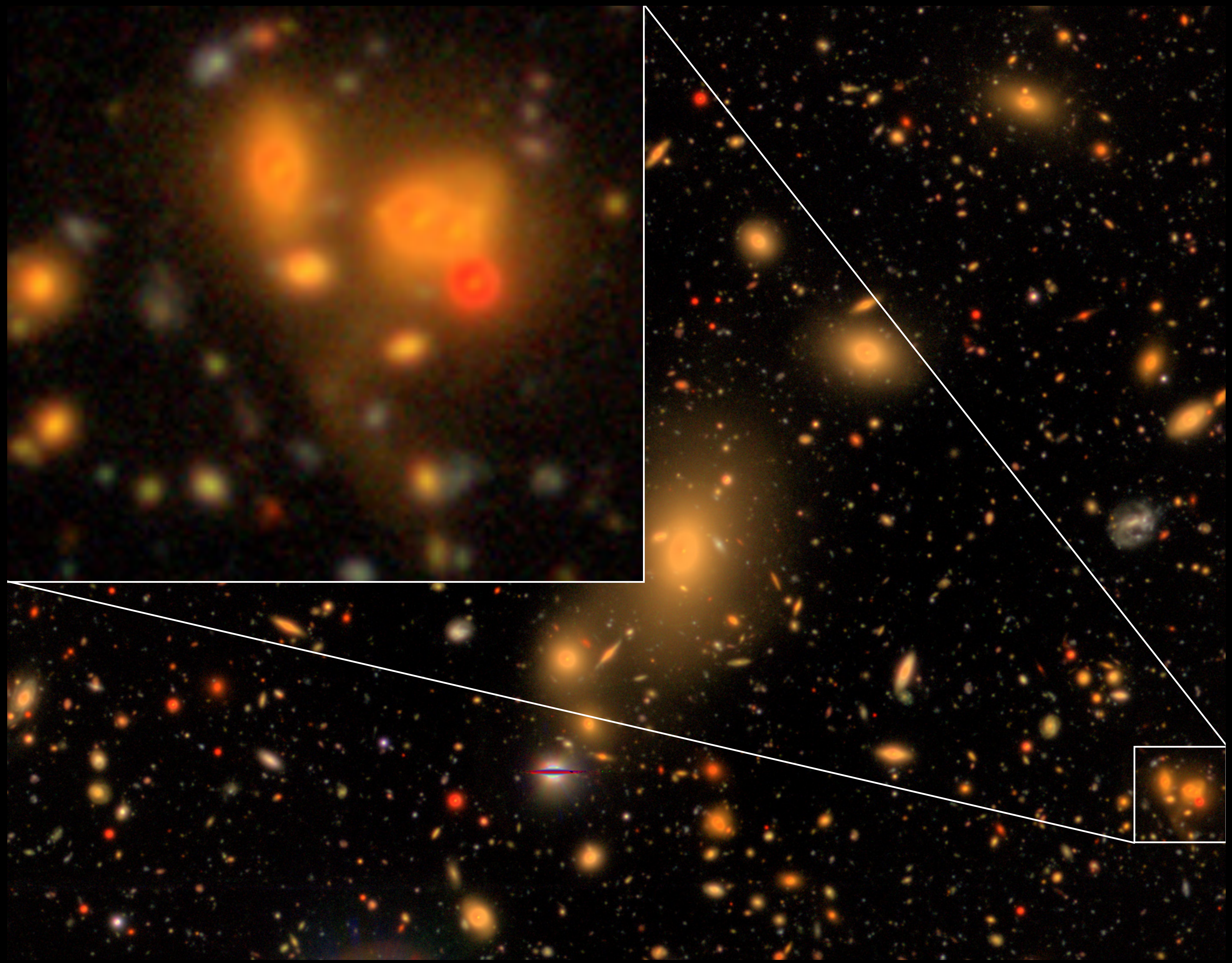


Credit : HSC

What do we expect in LSST ? (2)

Blending figures :

- HSC: 58% of the detected objects are identified as blended (Bosch+2017) - worsens in deep & ultra deep layers
- LSST : at least **62%** (Sanchez+2021, Dawson+2016)

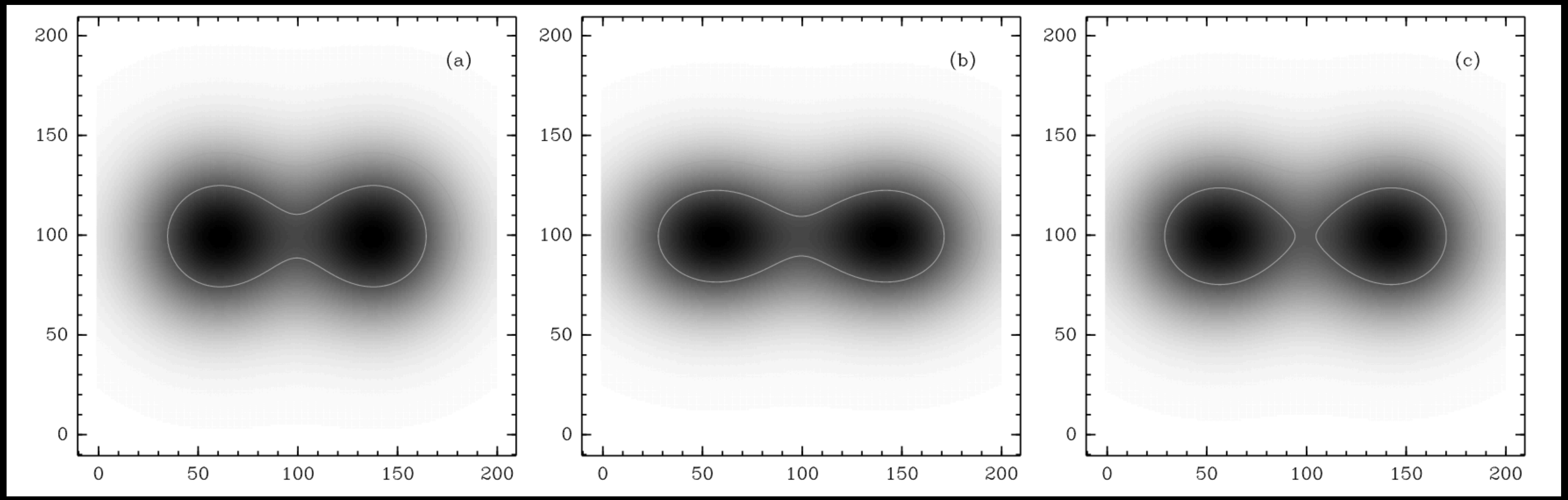


Credit: HSC

With a portion of unrecognized blends... (15%)

Why is this such a concern ?

- It affects **all** aspects of the analysis (detection, selection effect => n_{eff}, measured object properties : shapes & photometry, so photo-z…) in ways that are **coupled**.



In Sheldon+2020

➡We do have a problem as we know only how to deal with isolated objects

➡ Impact on shear analysis, individual measurements in crowded areas…

- + Affects the matching between space based and ground based photometric surveys (LSST)

➡ The blending effect has to be addressed and/or carefully evaluated ; the strategy depends on your analysis

What do we call deblending ?

- Solving the inverse problem of blending. Covers different efforts :
  - Extracting sources to get individual properties with (more or less) usual pipelines

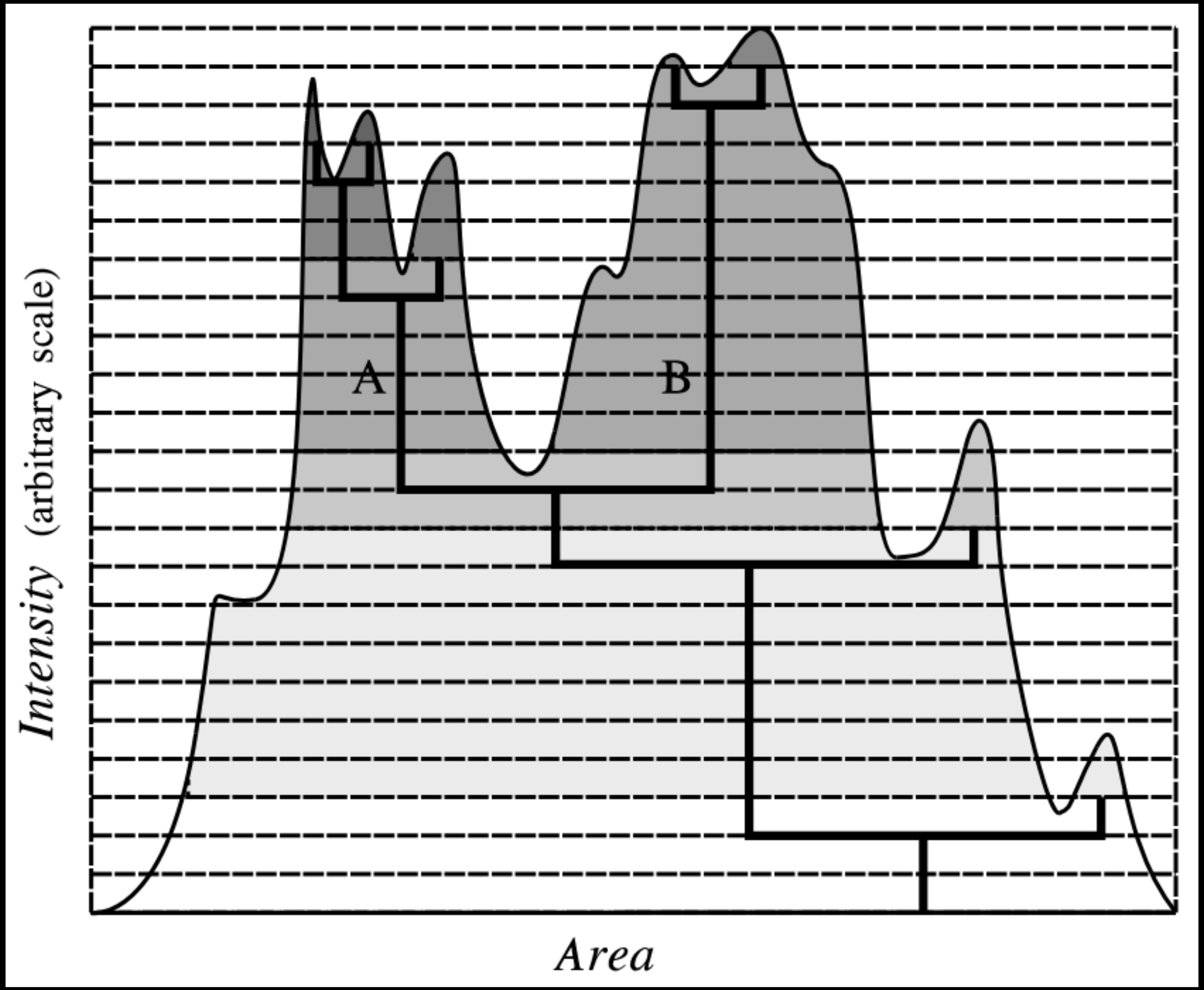
    ➡ Never perfect but the better it is; the smaller the corrections for calibration

  - Dealing with blended objects to extract relevant properties without separation
  - Calibration of the effect on an analysis (or part of the analysis) + characterization of the blending impact

SExtractor

- Detection of a blend by connecting zones at intermediate thresholds (multi-thresholding) on single band images
- Attributes pixels to one object or the other
- Suitable set of thresholds to work out the majority of cases ?…

➡ Important fraction of unrecognized blends

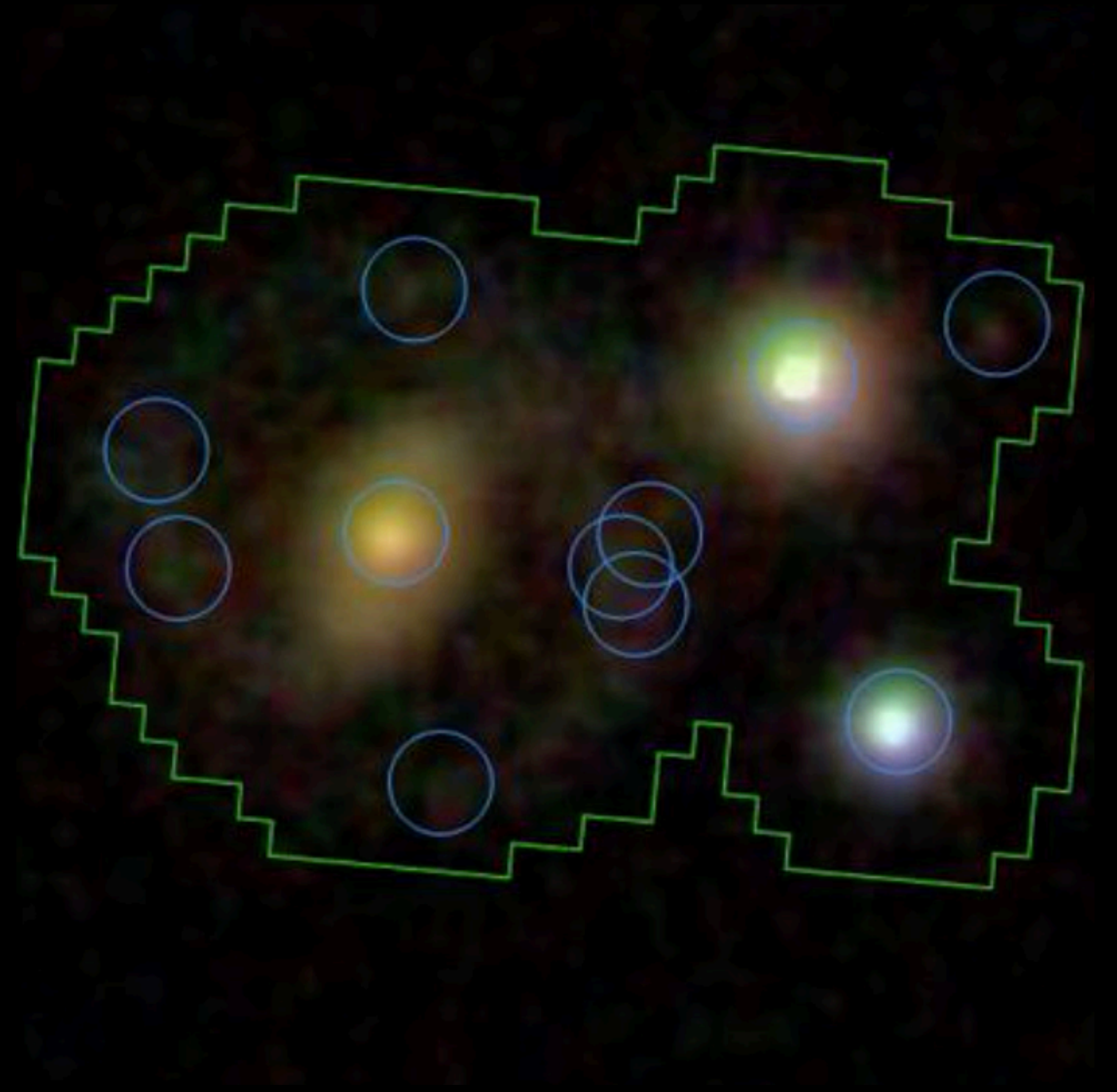


E.Bertin et al, 1996

HSC-SDSS Photo deblender

https://www.astro.princeton.edu/~rhl/photo-lite.pdf

- Designed for multi-band images (initially SDSS)
- Used in HSC (Bosch+ 2018)
- Identification of peaks and symmetrization to the lowest value



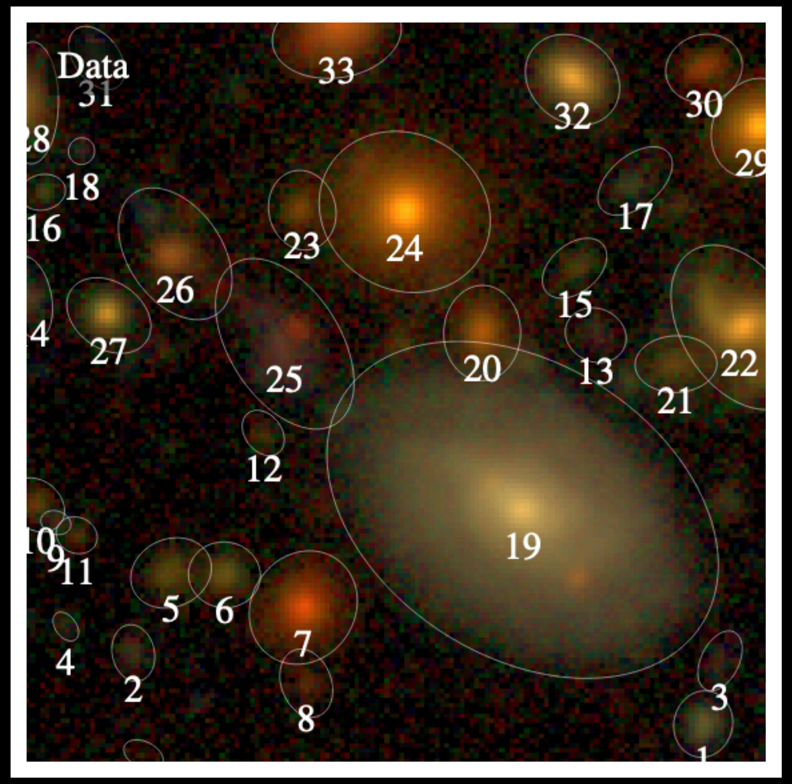
- Supposes non negative, optically thin objects, symmetrical
- Symmetry assumption tend to have it struggle with symmetric blends + it fails on irregular morphology.

Baseline algorithm for LSST : SCARLET

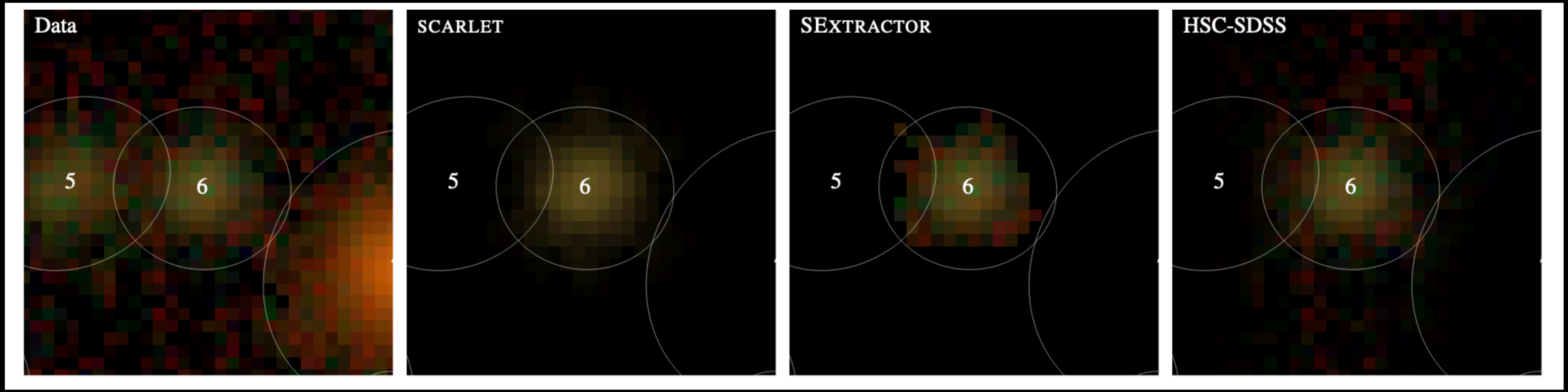
Melchior+ 2018
https://arxiv.org/pdf/1802.10157.pdf

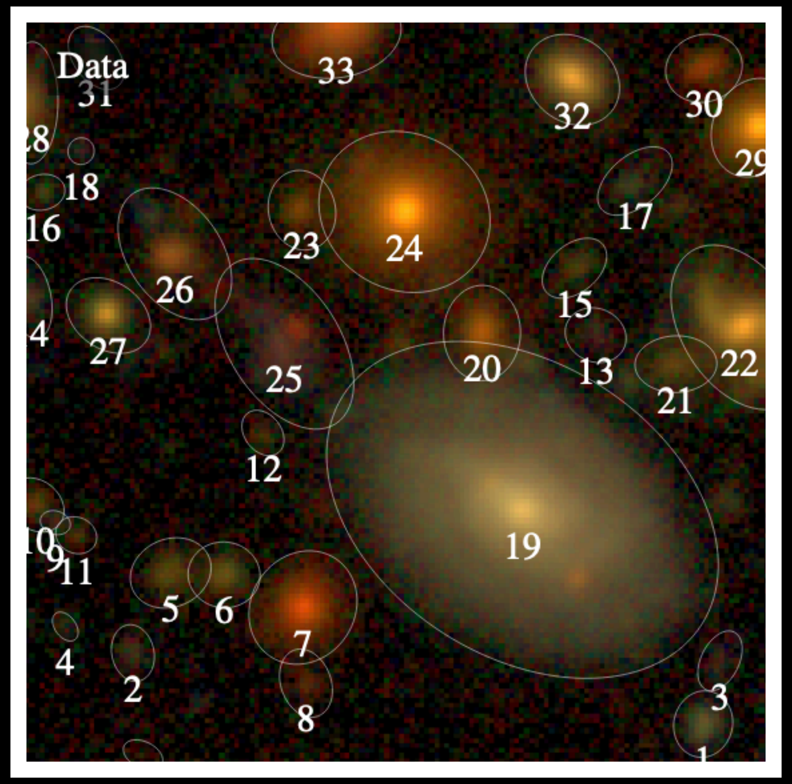
- Deblending as a mixture model
- Fit of a multi-band model $M = \sum_{k=1}^{K} A_k^T \times S_k = AS$
- Imposes a central symmetry of the sources, non-negativity and declining monotonicity from the peak pix.
- Other priors can be included using for exemple deep generative models (Lanusse+ 2019)
- Requires the position of the sources (detection step)

Performances

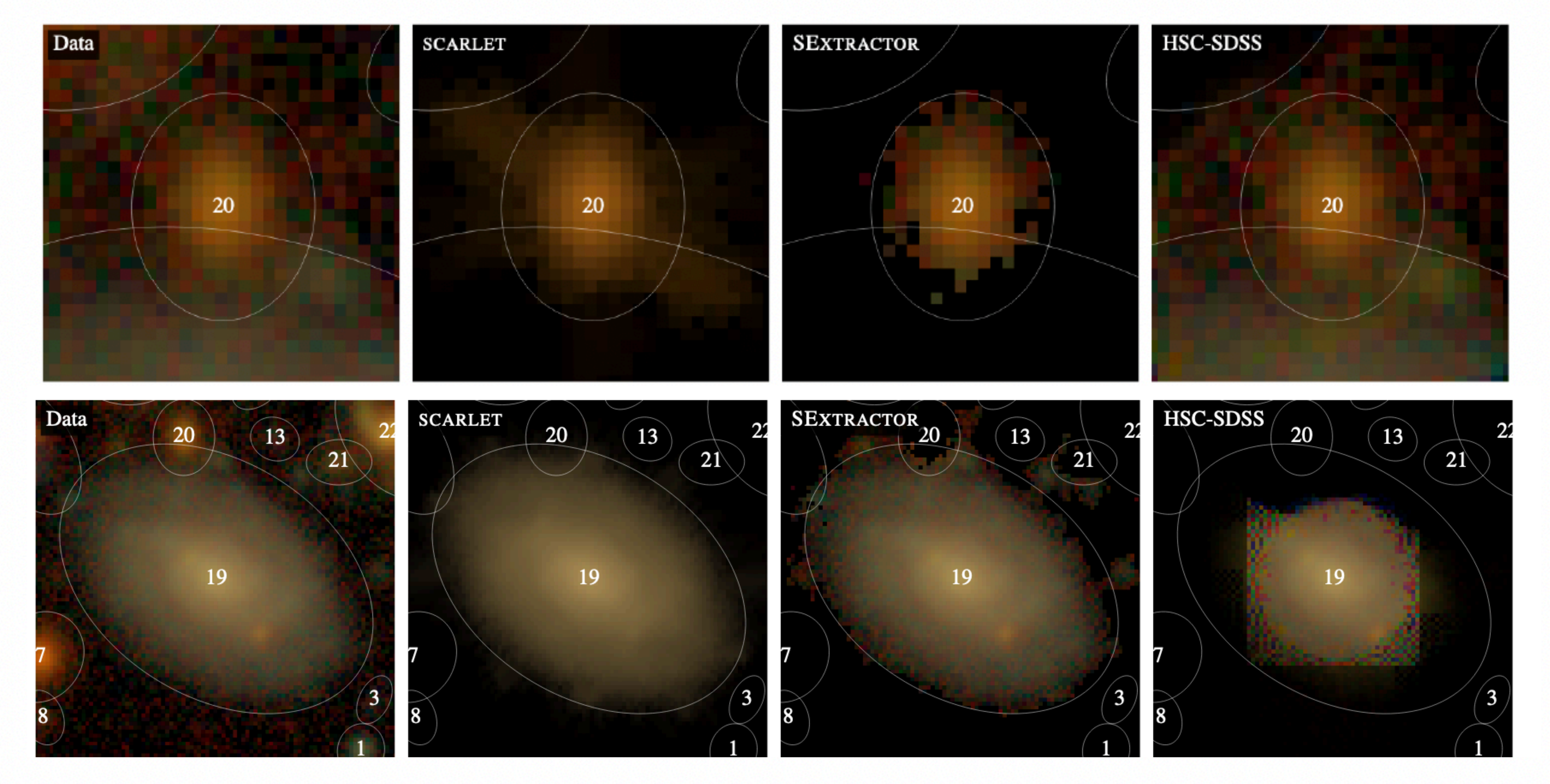


HSC UltraDeep COSMOS with an arcsinh stretch
25x25 arcsec
(Melchior+2018)





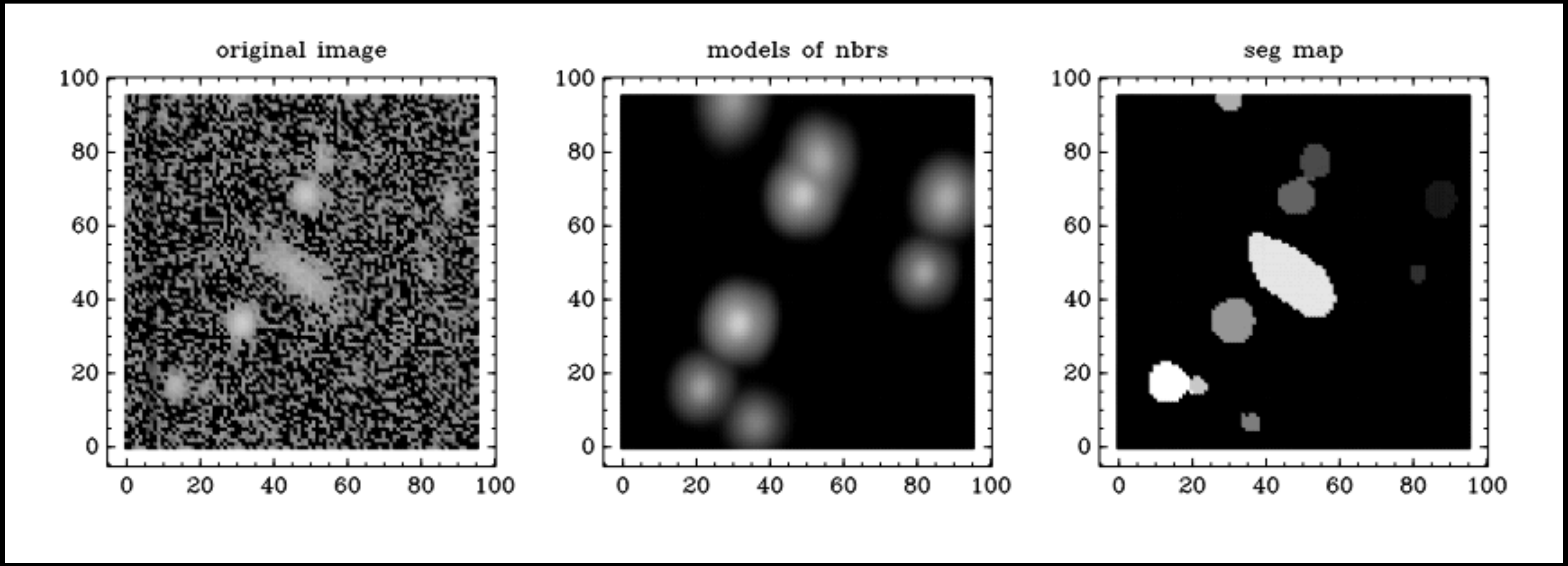
Performances



Melchior+ 2018, https://arxiv.org/pdf/1802.10157.pdf

DES algorithm

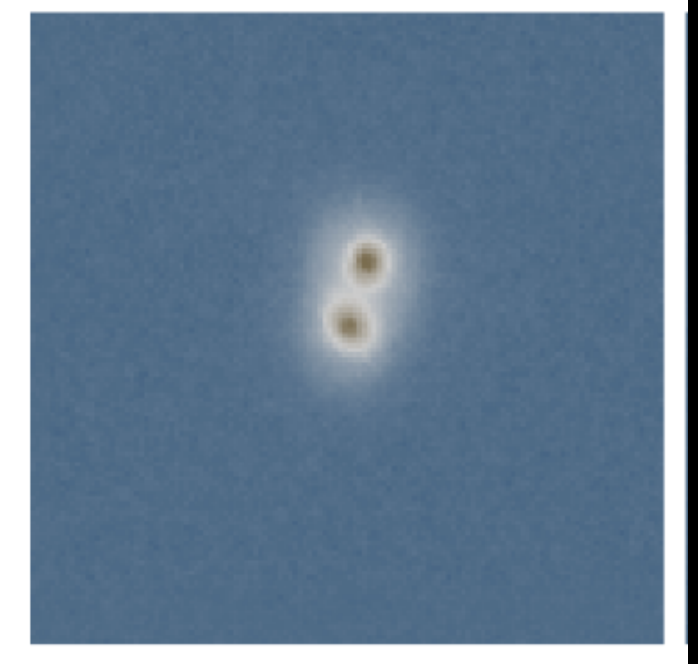
- Masking neighboring pixels (Jarvis+2016) Multi object fitting (MOF) (Drlica-Wagner+2018)
- **Forward modeling approach** fitting bulge + disk models to the blended sources
- Limited by the simplicity of the model

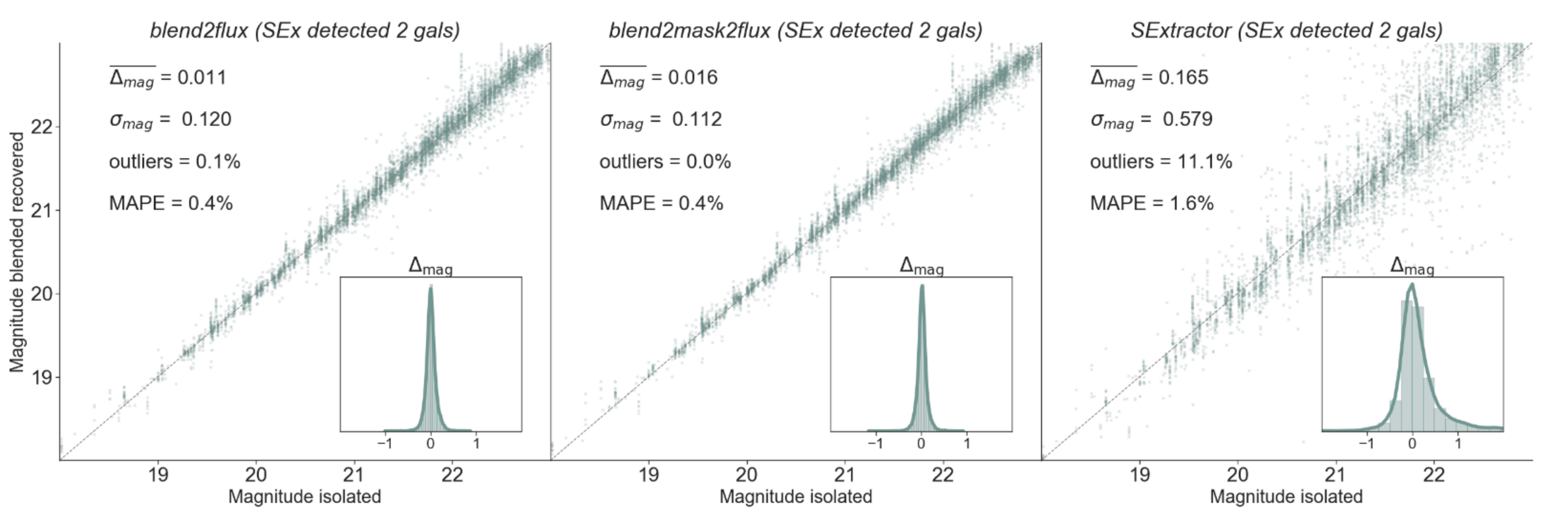


Using deep generative models

- Modeling of images is more refined (Régier+2015, Lanusse+2020, Bretonnière+2021)

- Use of non parametric methods for modeling > possible generalization





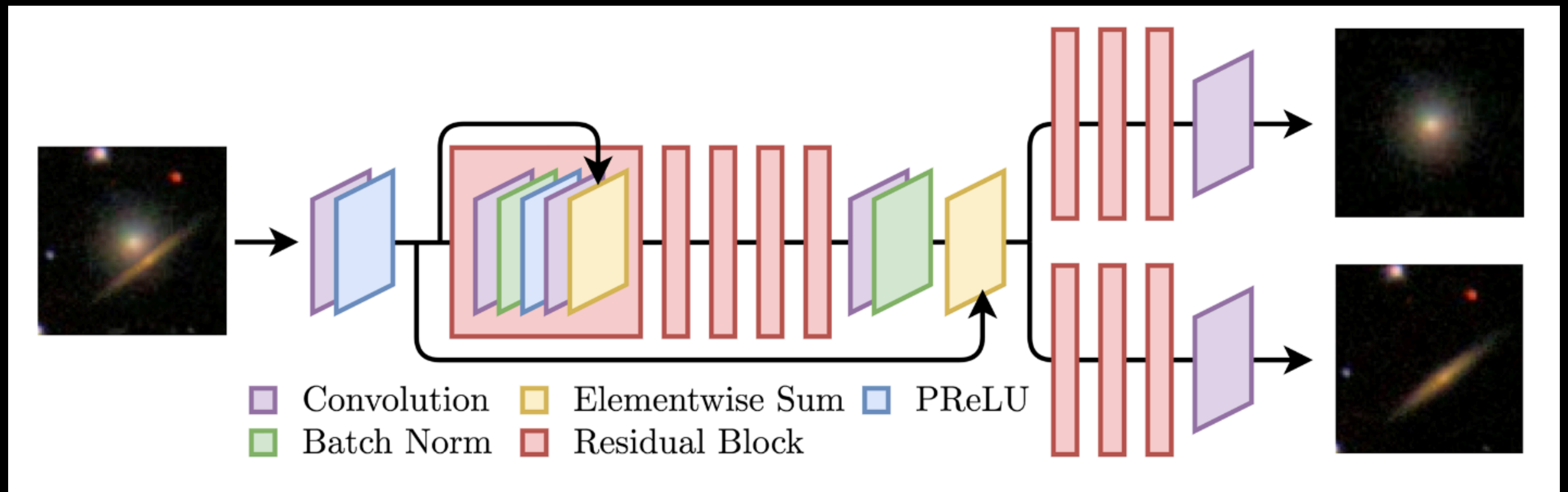
Boucaud+2019

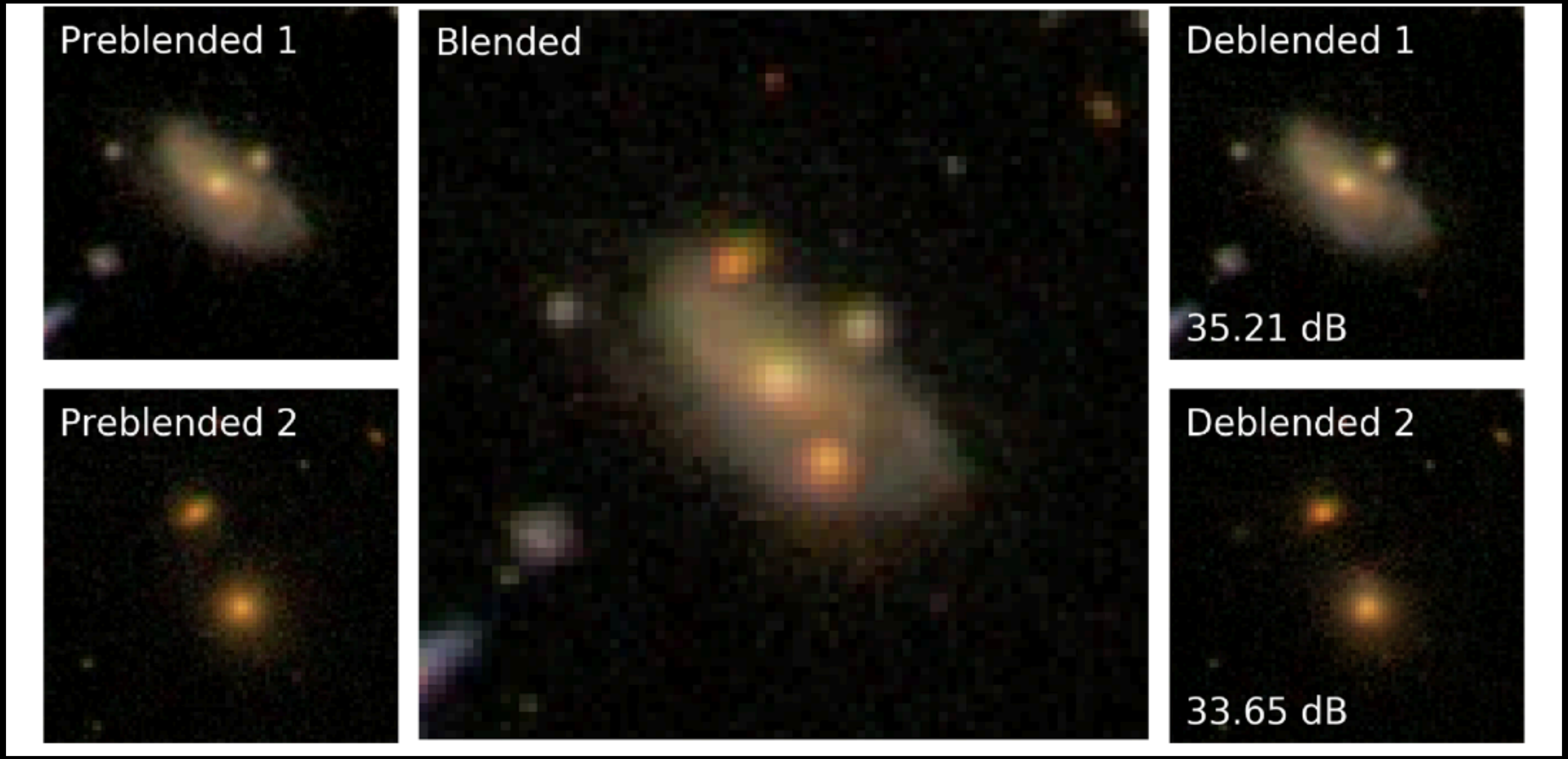
- Naturally allows multi-band use and allows to get to multi instrument/resolution analysis (Joseph+2021)

➡ Possible joint analysis at the pixel level (on overlapping regions)

Source extractions with generative nets

- GANs (Reiman+2019)

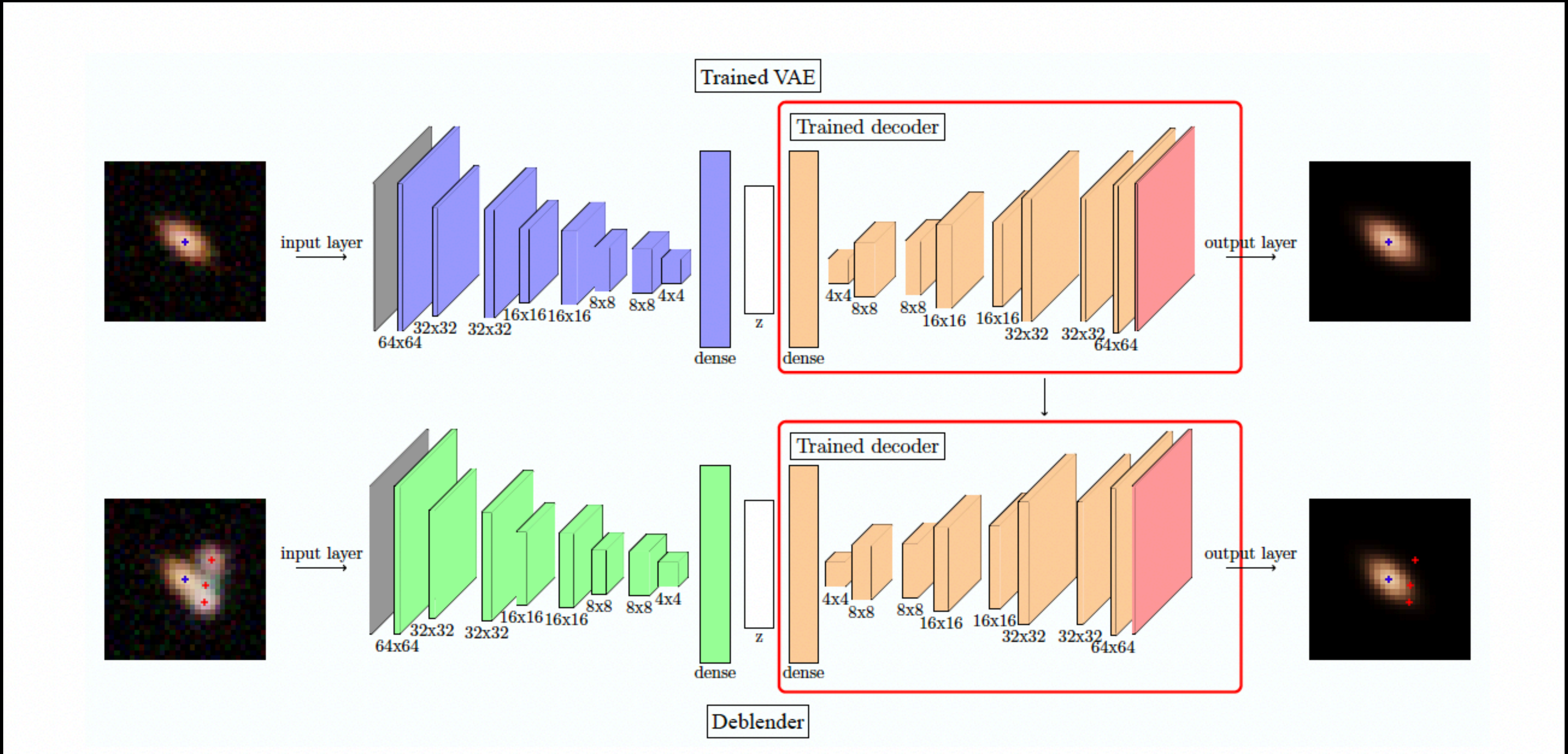




- VAE (Kingma+2013)

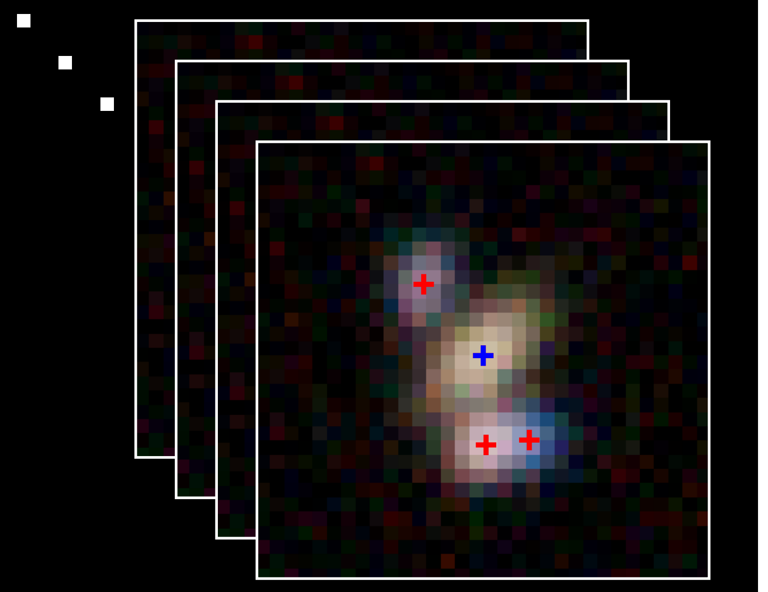
DebVAder

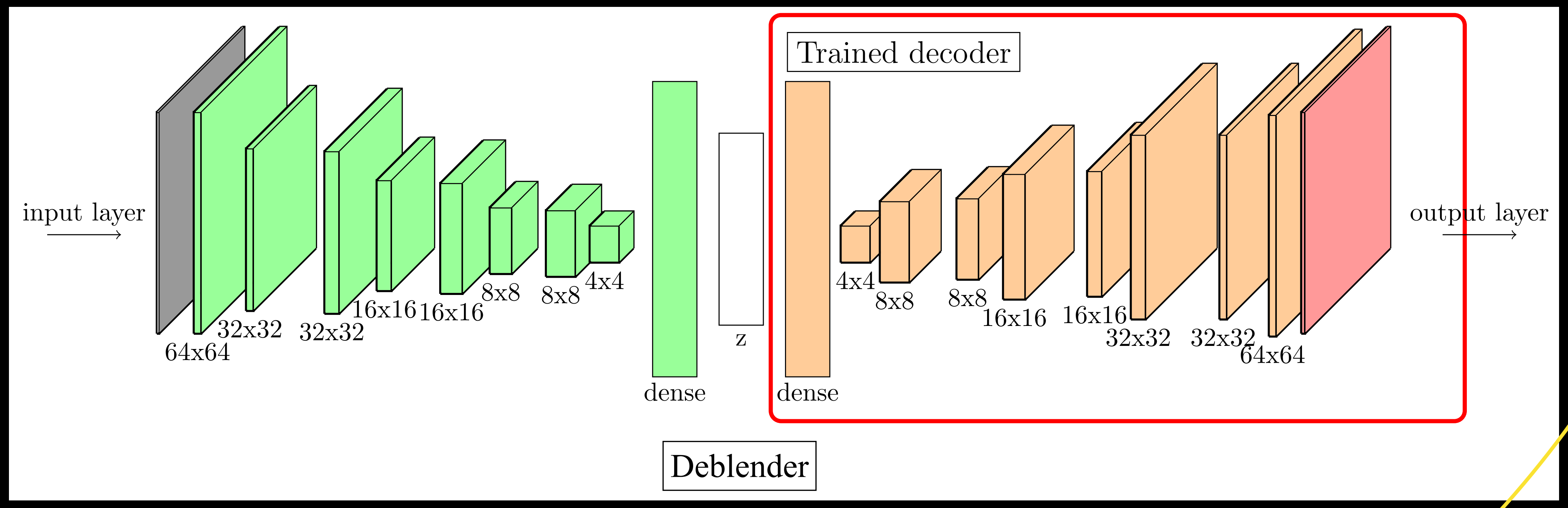
B. Arcelin PhD thesis



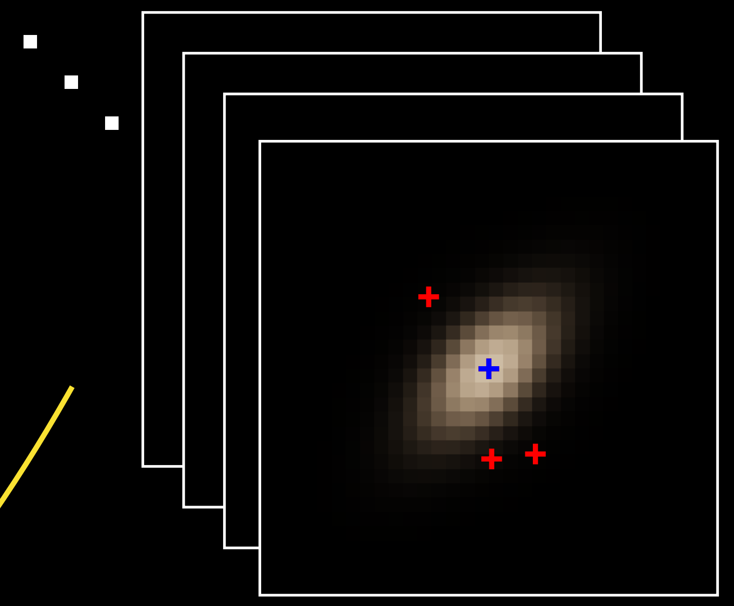
Variational autoencoders for deblending

Input





Output



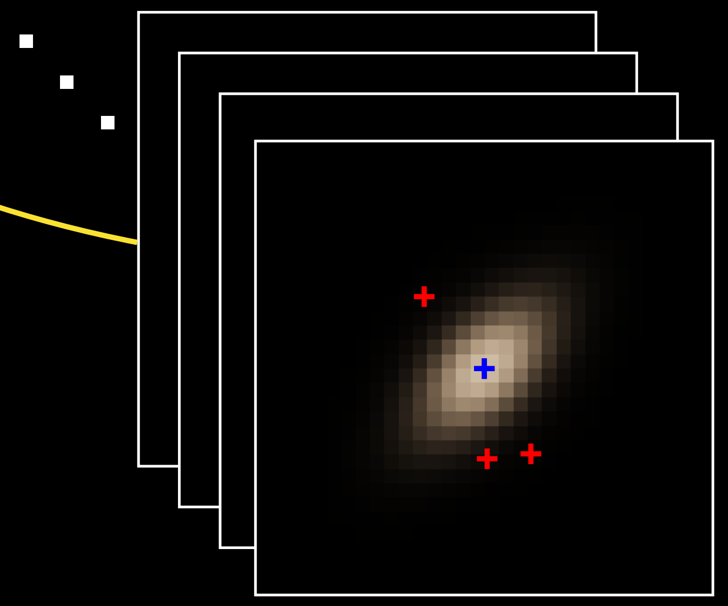
KSB (Kaiser+1995)

$$(e_{1,meas}, e_{2,meas})$$

$$\Delta |e| = \sqrt{e_{1,true}^2 + e_{2,true}^2} - \sqrt{e_{1,meas}^2 + e_{2,meas}^2}$$

$$(e_{1,true}, e_{2,true})$$

Target image



KSB (Kaiser+1995)

<u>Blend rate (Melchior+2018)</u>

$$B_{tot,i} = 1 - \frac{\sum_p s_{ip} s_{ip}}{\sum_p (s_{ip} \sum_j s_{jp})}$$

for galaxy i and neighbors j, p the pixels.

Variational autoencoders for deblending

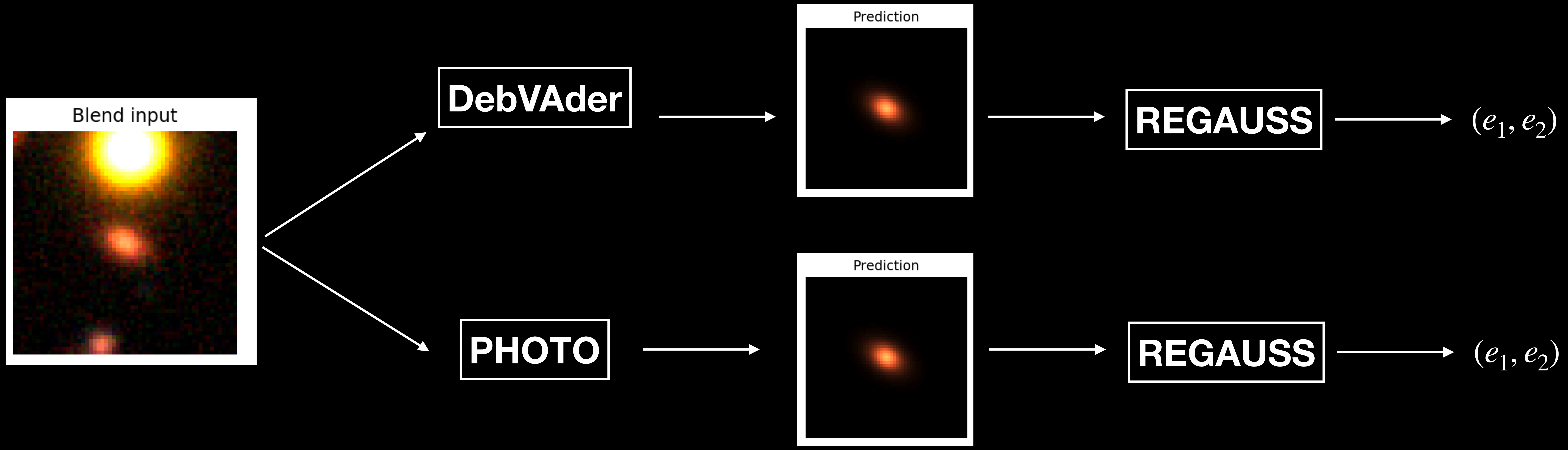
VAEs applied to overlapped galaxies

B. Arcelin

- Comparison with the PHOTO deblender implemented within the LSST DESC measurement pipeline:

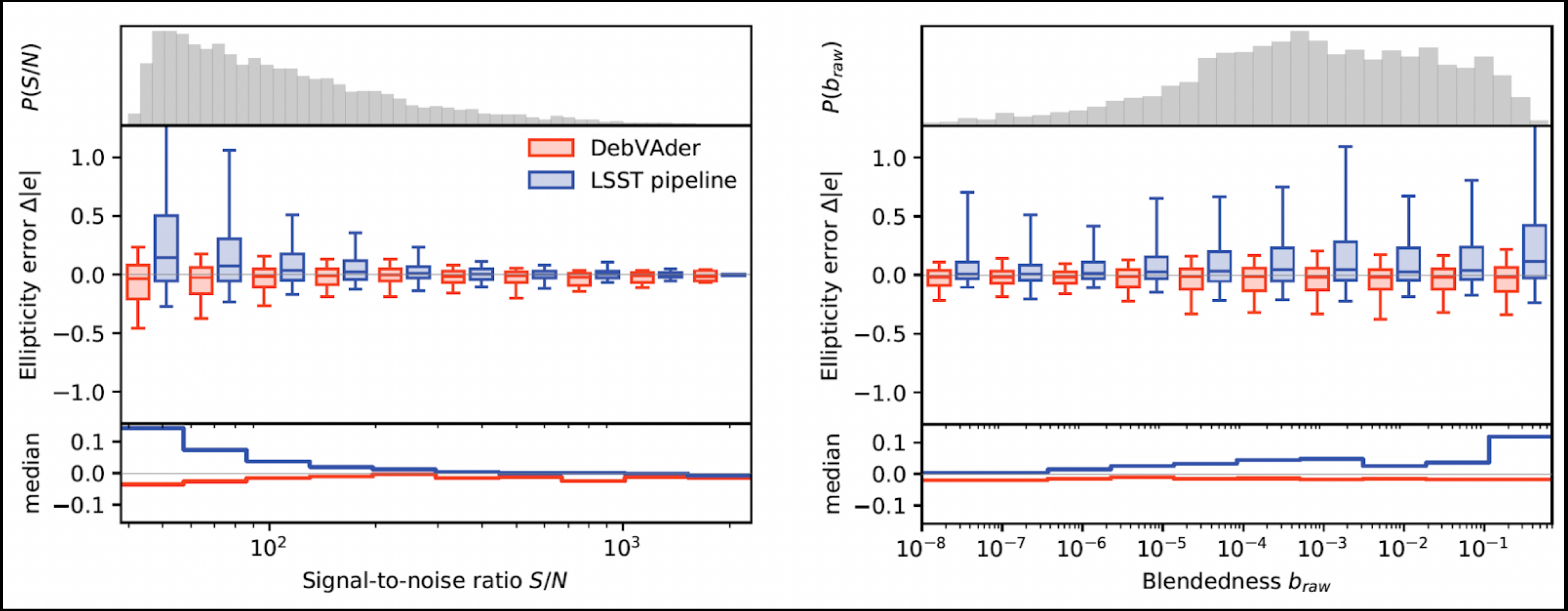
  1. Deblending with DebVAder or PHOTO

  2. Measuring the shapes with the REGAUSS implementation in the GalSim package



DebVAder vs Photo on DC2

B. Arcelin



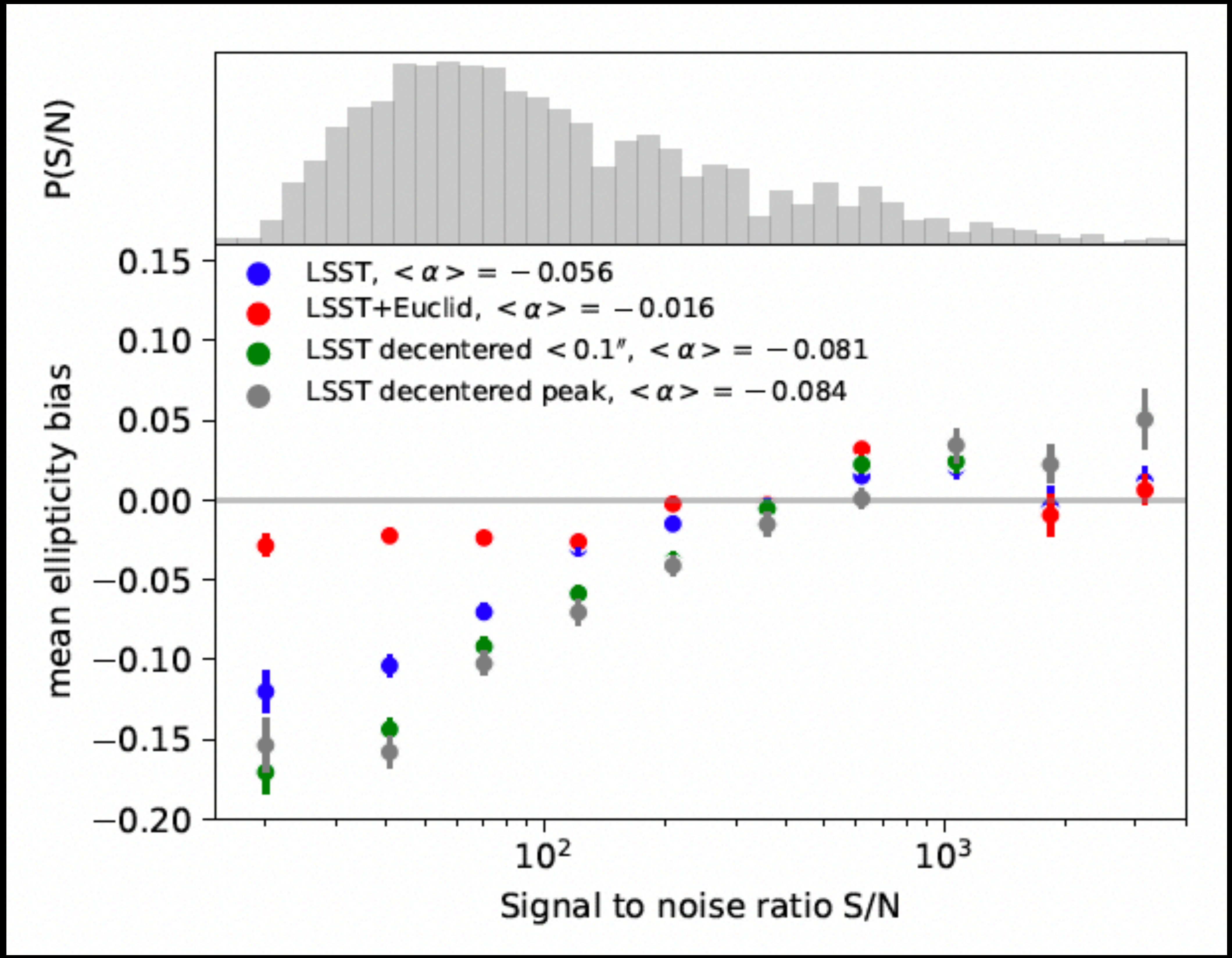
MAGNITUDE CUT : 24.5

DebVAder decreases median ellipticity errors by 70 and 120% over the test dataset on DC2 simulations.

DebVAder : multi-instrument approach

- Adding space-based survey (Euclid) to LSST
- Reduction of the ellipticity bias on Galsim simulations

Arcelin+2021



Ellipticity bias:

$$e_i^{out} \approx (1+\alpha_i)e_i^{int} + (1+m_i)\gamma_i + c_i$$

for galaxy i.

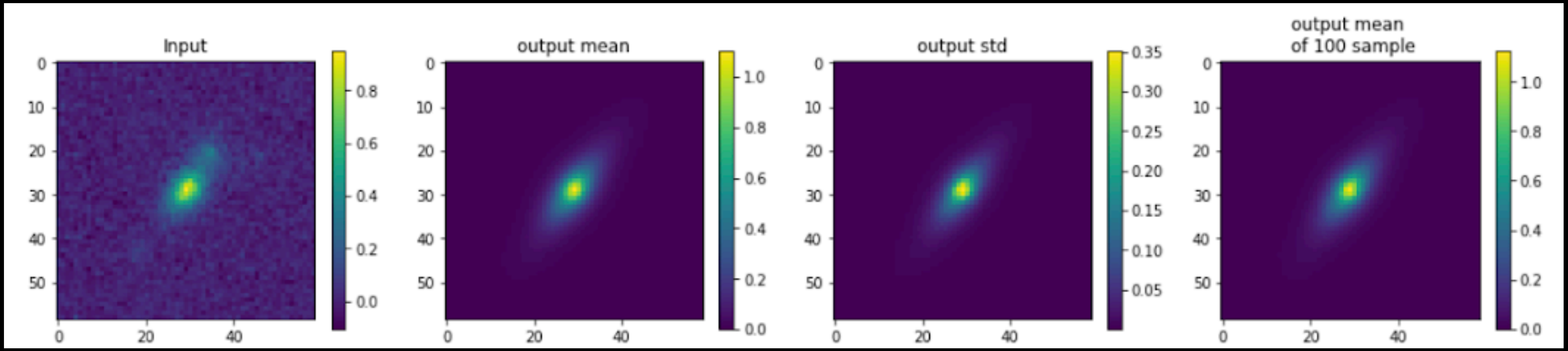
Going for multiple sources

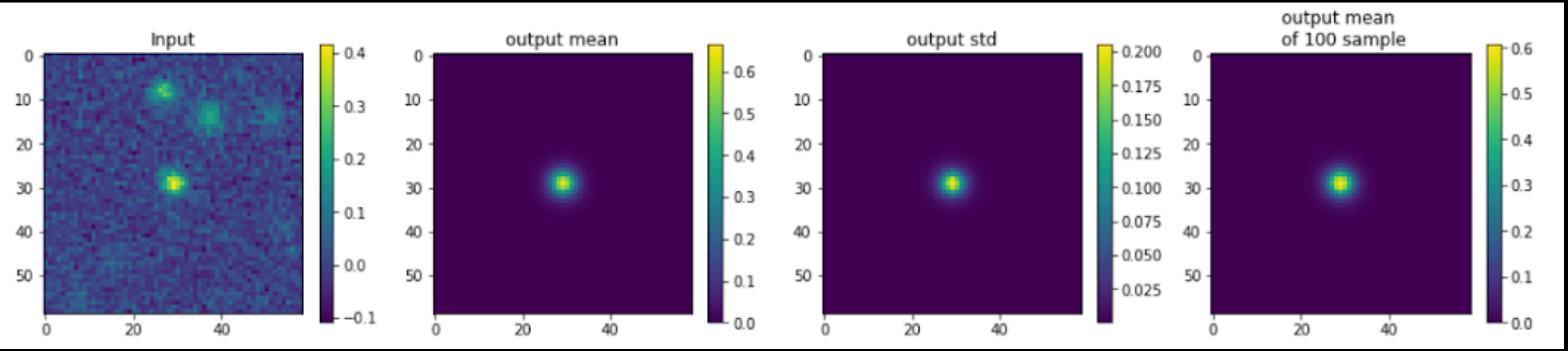
- Goal : get the whole image treated without prior assumptions on the location of the sources.
  - Solve for multiple sources at once
  - Iterative procedure based on DebVAder : get to work with a field of galaxies

  - H.Wang+2021 RDN + classifier

DebVAder cont'd

- Probabilistic outputs taking uncertainty into account





- DebVAder pip package under release (6/12 - Blending working group presentation)

Blending characterization

- The characterization of performances and the comparison of algorithms is a central point to establish strengths in analyses :
  - The Blending Toolkit (BTK) provides with an environment for simulation and benchmarking (I. Mendoza, A. Boucaud, T. Sainrat) - the first release is about to be made - https://lsstdesc.org/BlendingToolKit/index.html
  - Quickly generate datasets of blended objects for testing different detection, deblending and measurement algorithms, as well as training samples for machine learning algorithms.

What are we left with ?

- Other data driven approach : METACALIBRATION (even METADETECTION) for shear analyses

- Principle : infer shear response matrix of objects by applying artificial known shear

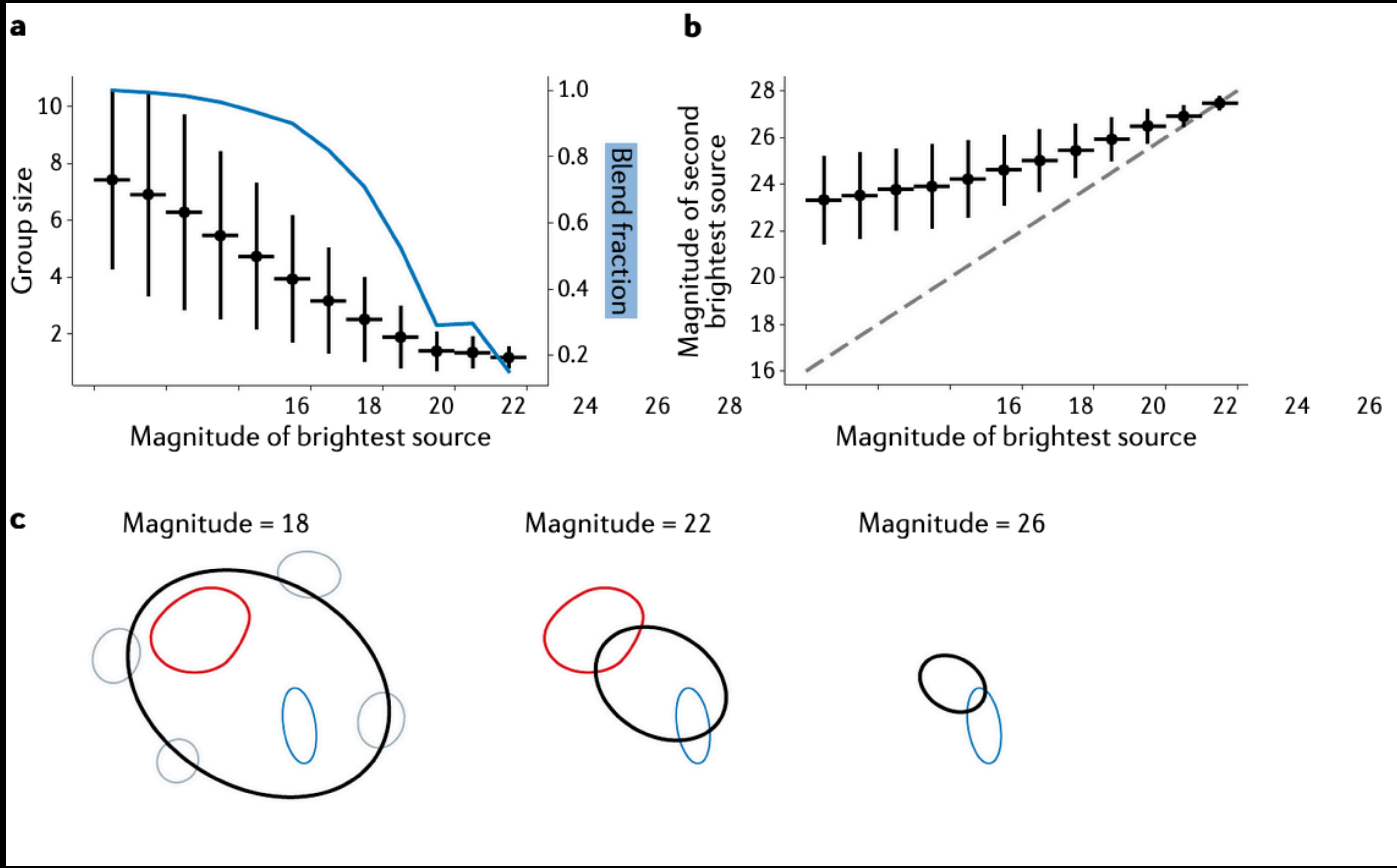
  - Corrects for shear dependant detection biases - unaffected by blends https://arxiv.org/pdf/1911.02505.pdf

- Remaining problem : the blended galaxies have different z => they are lensed differently and have different photo-z…relying on photometry : define n_γ(z) accounting for both effects but very heavy with realistic simulations

Conclusion

- Deep surveys like LSST have to deal with blending
  - Extract shapes and fluxes
  - Mitigate any remaining effect per calibration (METACALIBRATION et METADETECTION for shear estimation)
  - Variety of methods with strong development of DL based methods : non parametric and more realistic solutions
- Further thoughts :
  - Probabilistic cataloging (Feder+2020) and faster StarNet (Liu+2021)
    - Accounting and characterization of uncertainties
  - Pixel based ground + space joint analyses (Joseph+2021)

En +



Owing to their apparent size on the sky, the brightest galaxies (*i*-band magnitude ~16) are effectively guaranteed to be blended (blue line in figure part **a**), typically forming groups of 8 ± 3 sources (black data points in figure part **a**), with the second brightest source in such a group usually being 8 magnitudes (that is, a factor of 1,500) fainter (figure part **b**). The measurement of the brightest sources is, thus, mostly unaffected by blending, but blending almost certainly hinders the detection and measurement of fainter sources in their vicinity.

Galaxies with intermediate brightness still have blending rates of more than 50% and group sizes of 2–4 (figure part **a**), but if they are blended, the secondary source is only 2–4 magnitudes (factors 6–40) fainter (figure part **b**). This is the regime in which blending is still very common and can be very noticeable in measurements.

At the faintest end, sources appear very small and blending becomes rare (figure part **a**), but if it happens, it involves another source of almost equal brightness (figure part **b**). This is the regime in which measurements, and even detection probabilities, are most strongly affected by blending.

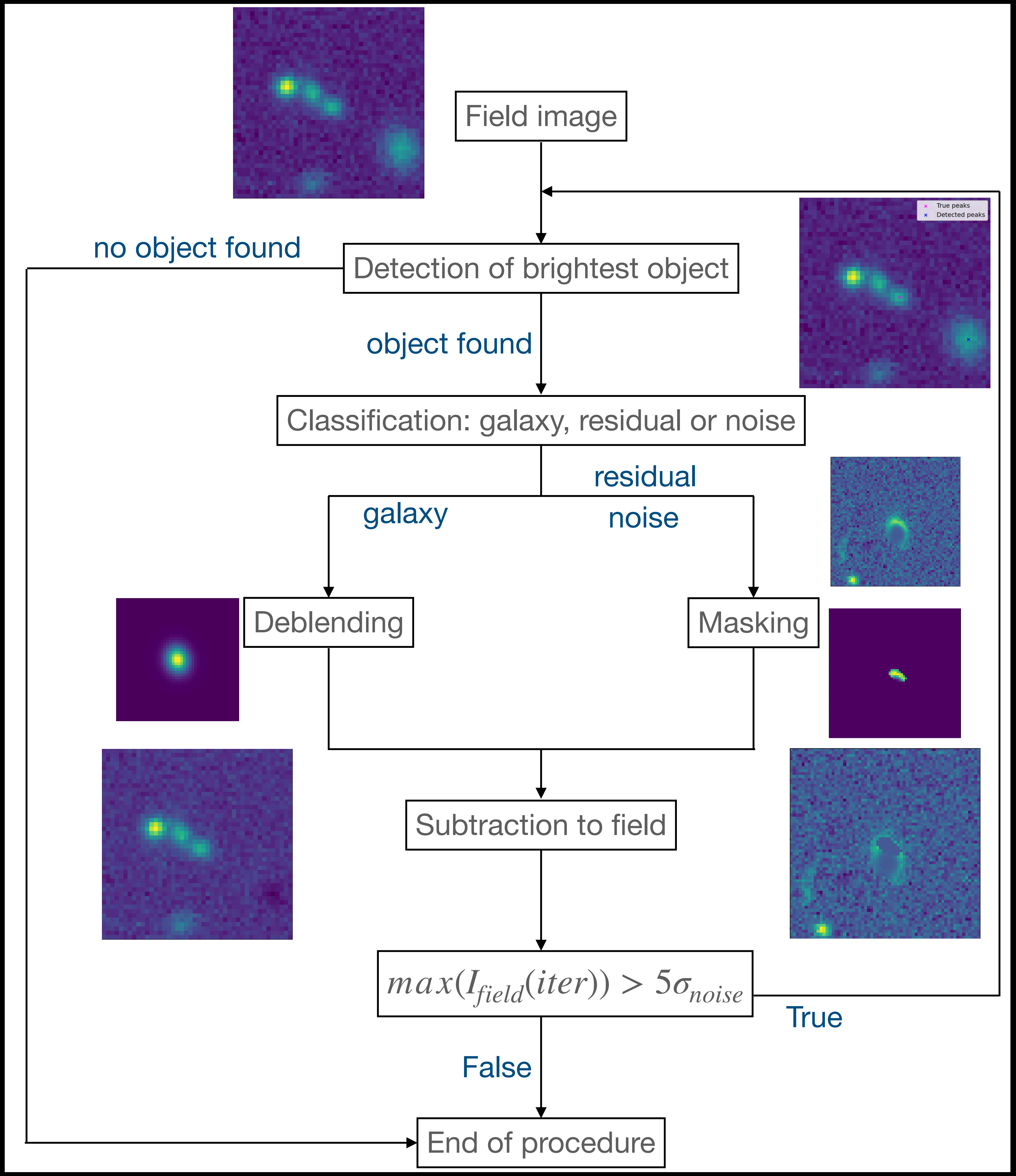
A sketch to summarize these findings from blending statistics is shown in figure part **c**. Because brighter galaxies are intrinsically larger and their outskirts remain visible further away from their centres, they occupy more area on the sky (not shown to scale). With the same secondary source population in all three cases, blended groups get smaller with decreasing brightness of the primary source, but the secondaries become similarly bright.

In figure parts **a** and **b**, vertical error bars denote symmetrical 68% confidence intervals and horizontal error bars indicate the bin width of 1 magnitude. The dashed line in part **b** indicates equality between brightest and second brightest sources as a guide to the eye.

Variational autoencoders for deblending

VAEs applied to overlapped galaxies

Work with Thomas Sainrat (intern from July to December 2020)



Combining space and ground data

To combine images, no pretreatment: directly simulated images