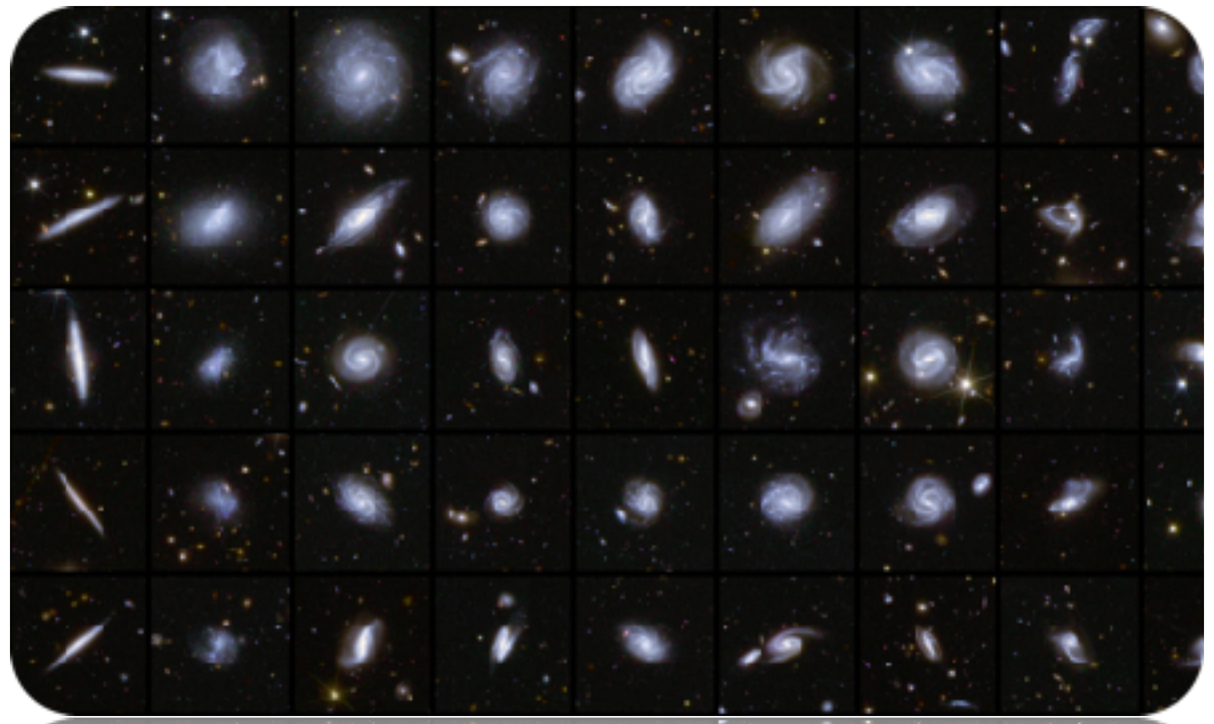
Using AI to Probe the Invisible Universe

Machine learning methods for weak gravitational lensing

Samuel Farrens

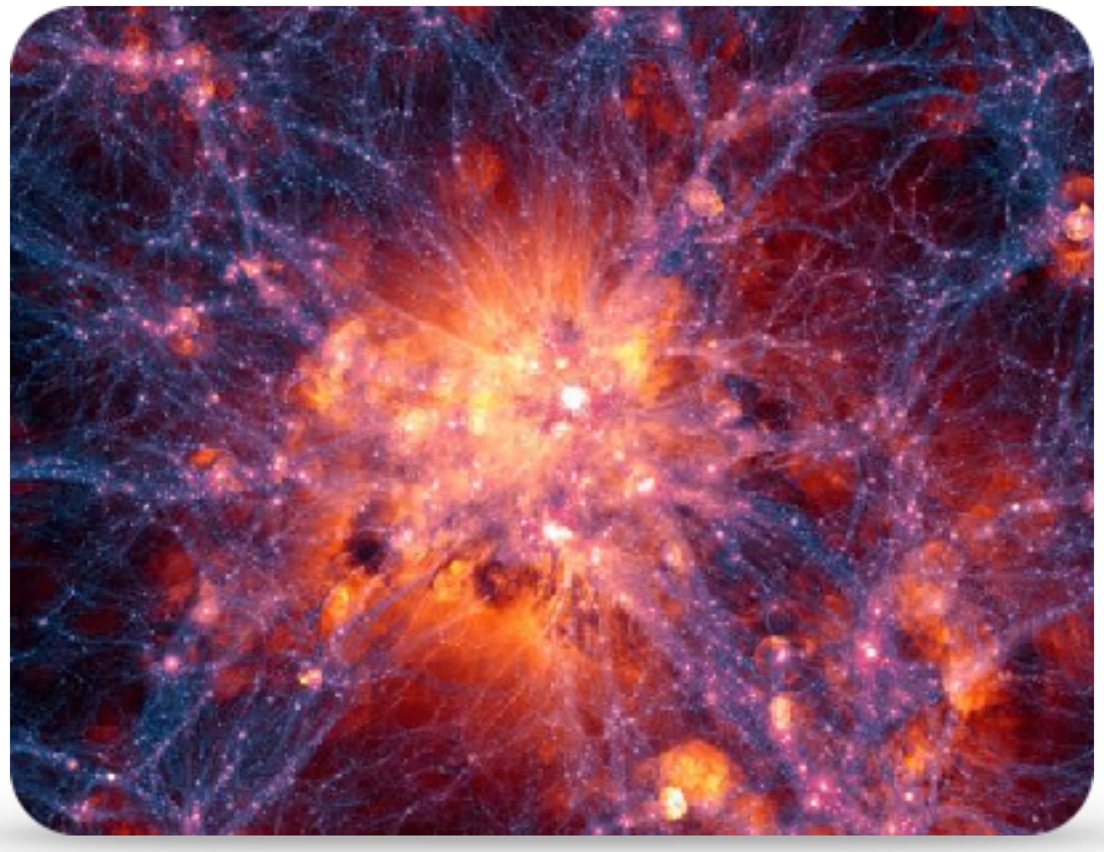


About Me

Sam Farrens

- Staff researcher at CEA
- Key research topics
- Project involvement
  - Euclid Consortium
  - ARGOS

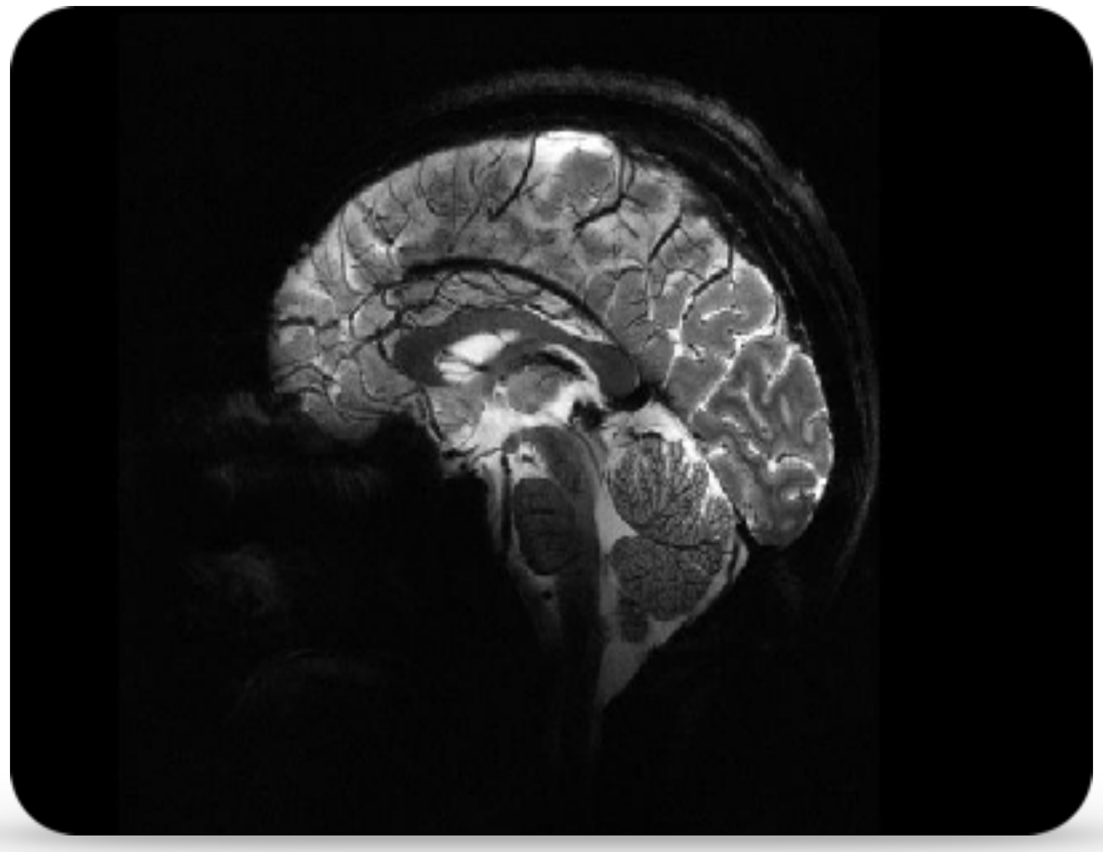
sfarrens.github.io/



Cosmology

Artificial Intelligence

Scientific software



Signal processing

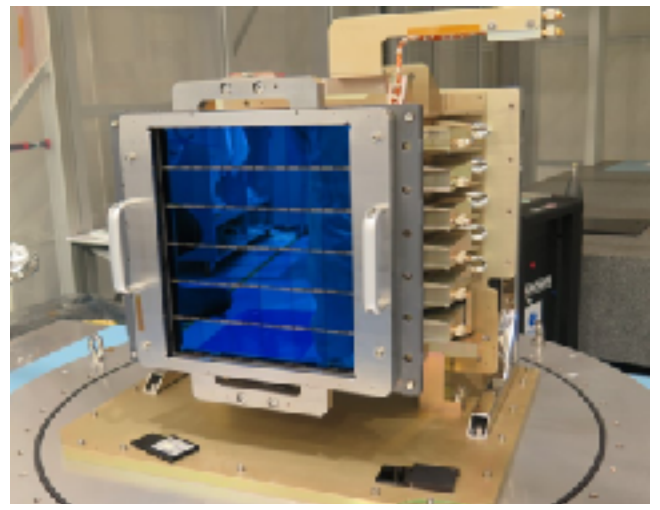
Where I Work

www.cea.fr & irfu.cea.fr

- **CEA:** French Alternative Energies and Atomic Energy Commission
- **IRFU:** Institute of research into the fundamental laws of the Universe

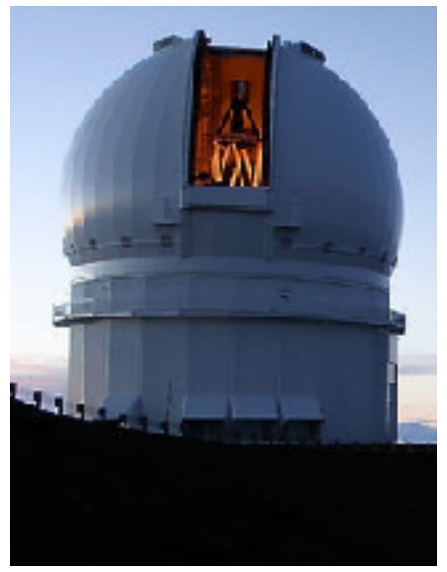
irfu.cea.fr/dap/

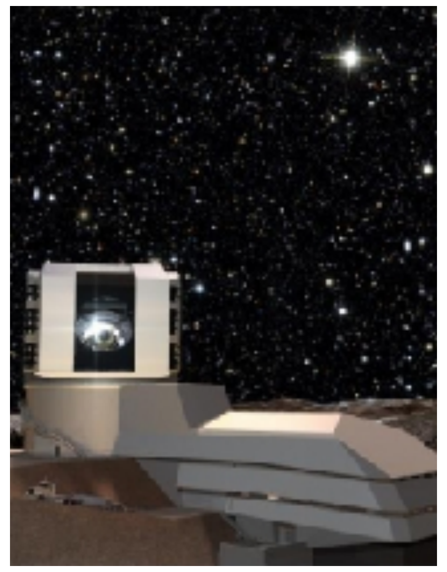
- **AIM:** Astrophysics, Instrumentation and Modelling
  - Cosmology
  - Galaxy Formation
  - Star Formation
  - Exo-planets
  - High-energy events
  - Plans modelling
  - Instrumentation (e.g. JWST, Euclid)



www.cosmostat.org

- **CosmoStat:** Cosmology and Statistics
  - Weak gravitational lensing
  - Machine learning
  - Signal/image processing
  - Inference
  - Euclid, Rubin LSST, SKAO, UNIONS





Outline

1. Cosmological motivation

- What are the **problems** we need to solve?
- How does **data** play a role in this?
- Why do we need **AI** and how can we trust the results?

2. Weak gravitational lensing

- ‘Observing’ **invisible matter**
- Technical **challenges**

3. AI for cosmological data analysis

- Estimating **stellar SEDs** from photometric data
- Identify of **blended galaxies** in survey data
- Going beyond

4. Future perspectives

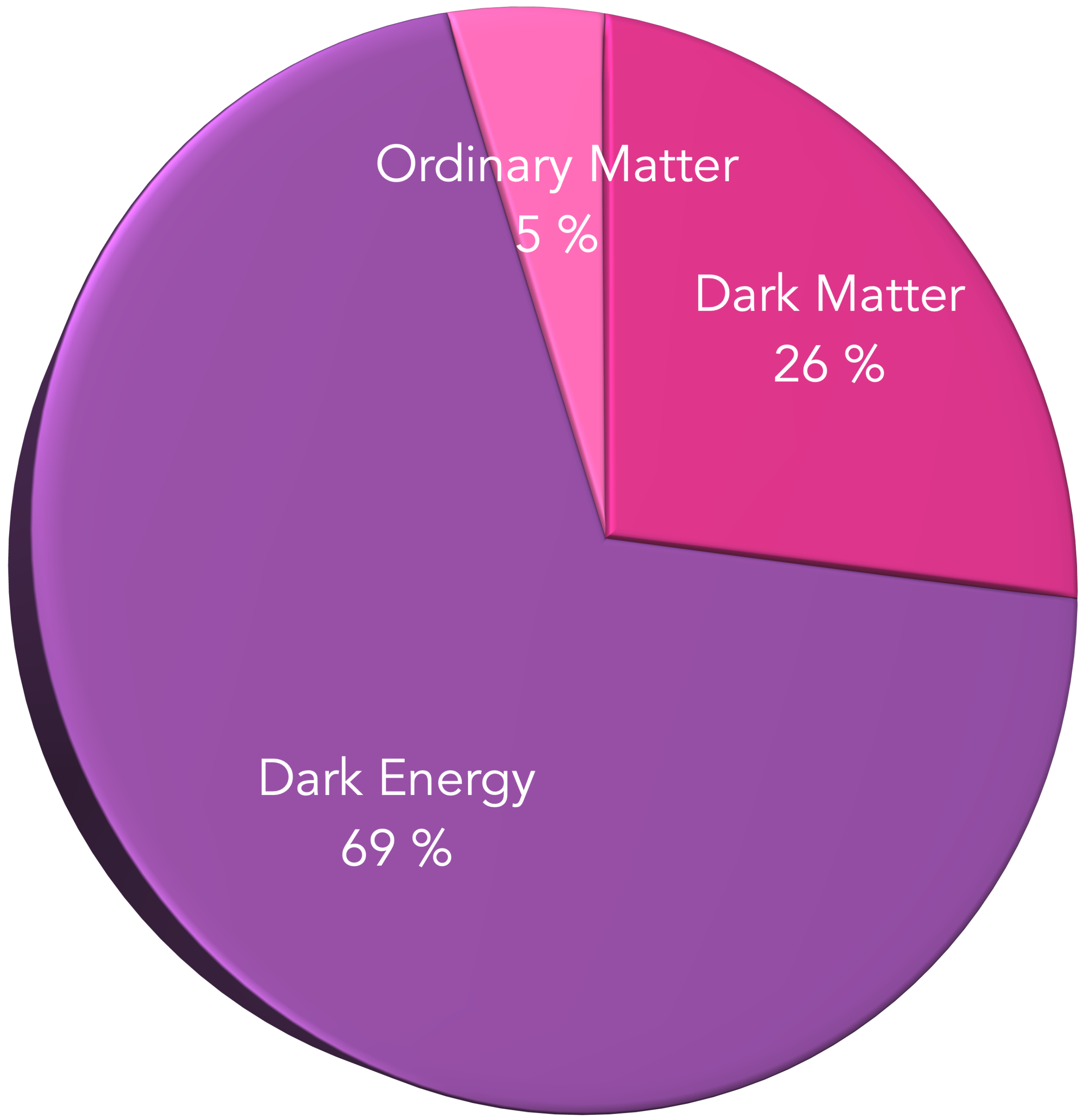
- Where can AI take us?
- Accessible **tools**
- Conclusions

Outline

1. Cosmological motivation

- What are the **problems** we need to solve?
- How does **data** play a role in this?
- Why do we need **AI** and how can we trust the results?

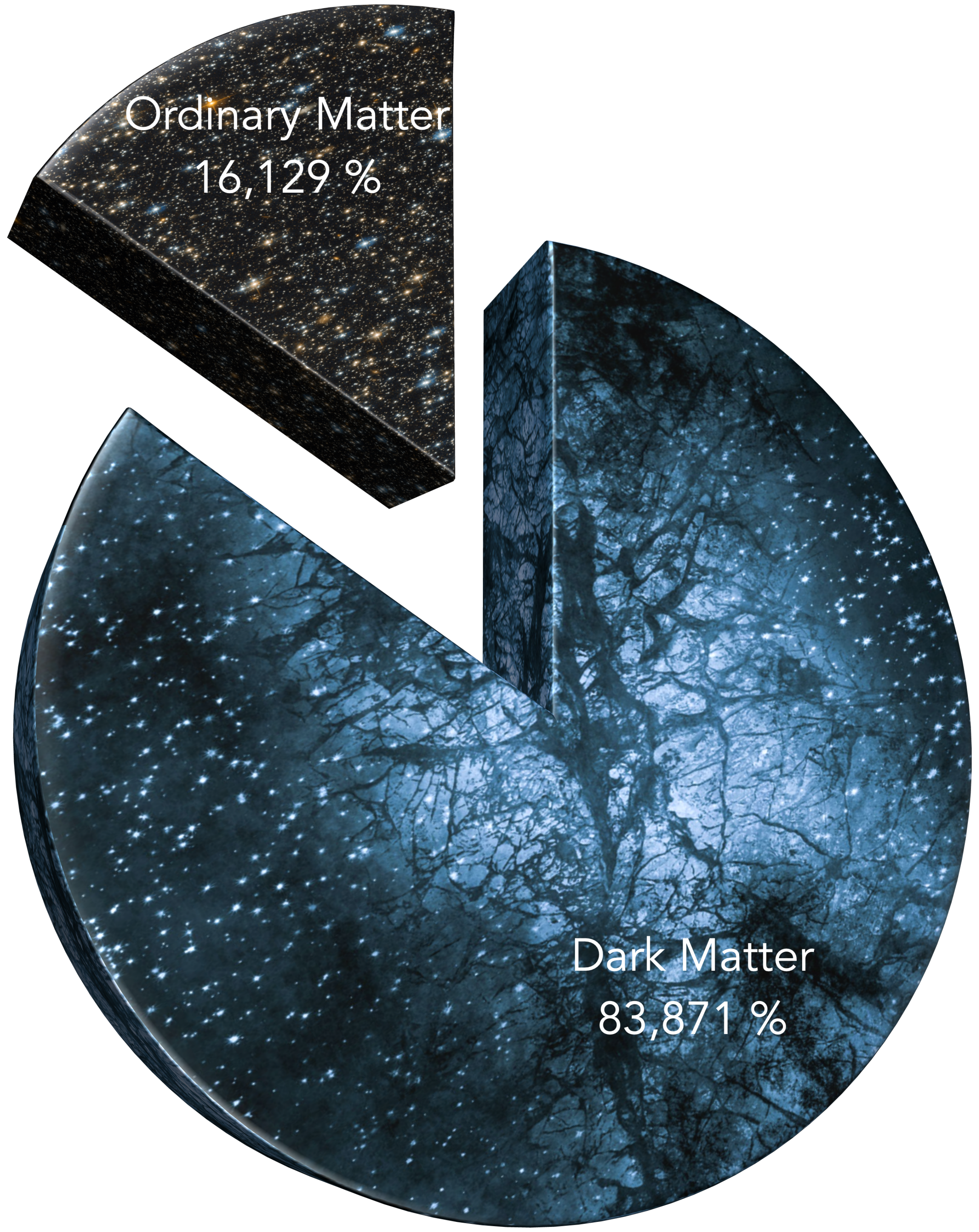
Challenges in Cosmology



Energy Density

Challenges in Cosmology

- Baryonic effects?
- Nature of dark matter?
- Nature of dark energy?
- Is ΛCDM complete?

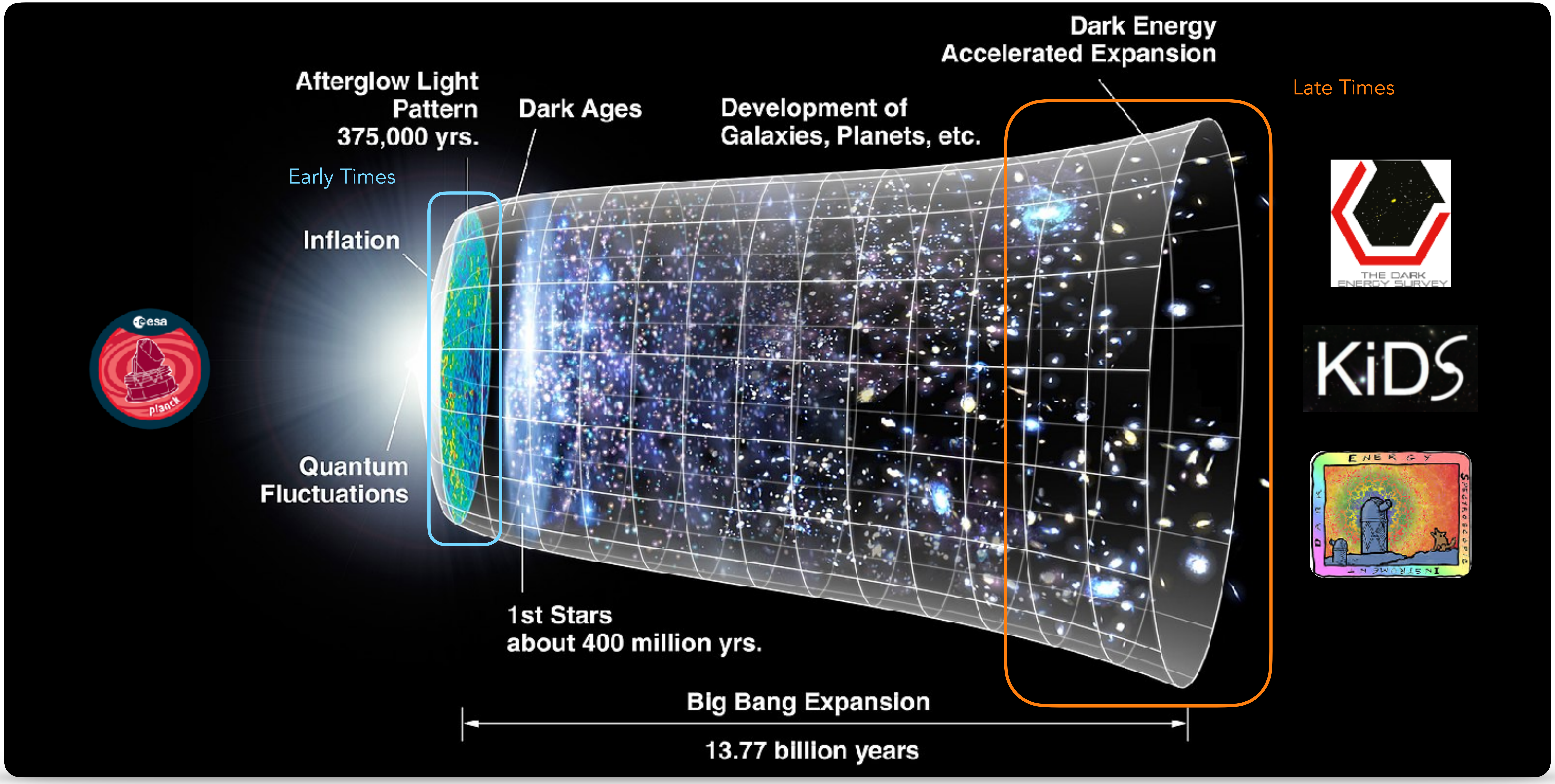


Matter Density

- Is General Relativity complete?
- New physics to be found?
- Do we really understand systematics?
- Neutrino masses?

Challenges in Cosmology

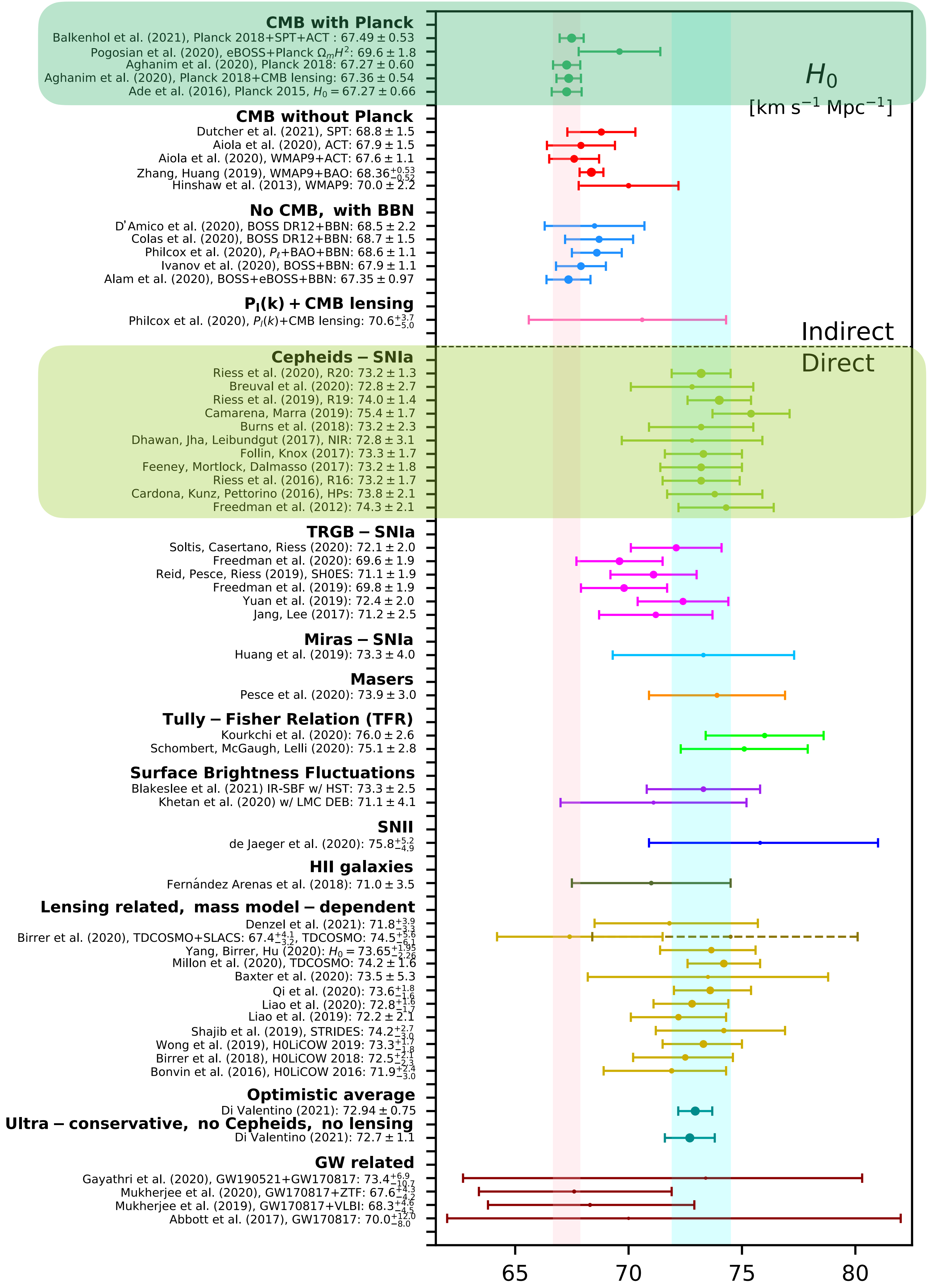
Tension between early and late-times



Credit: NASA/WMAP

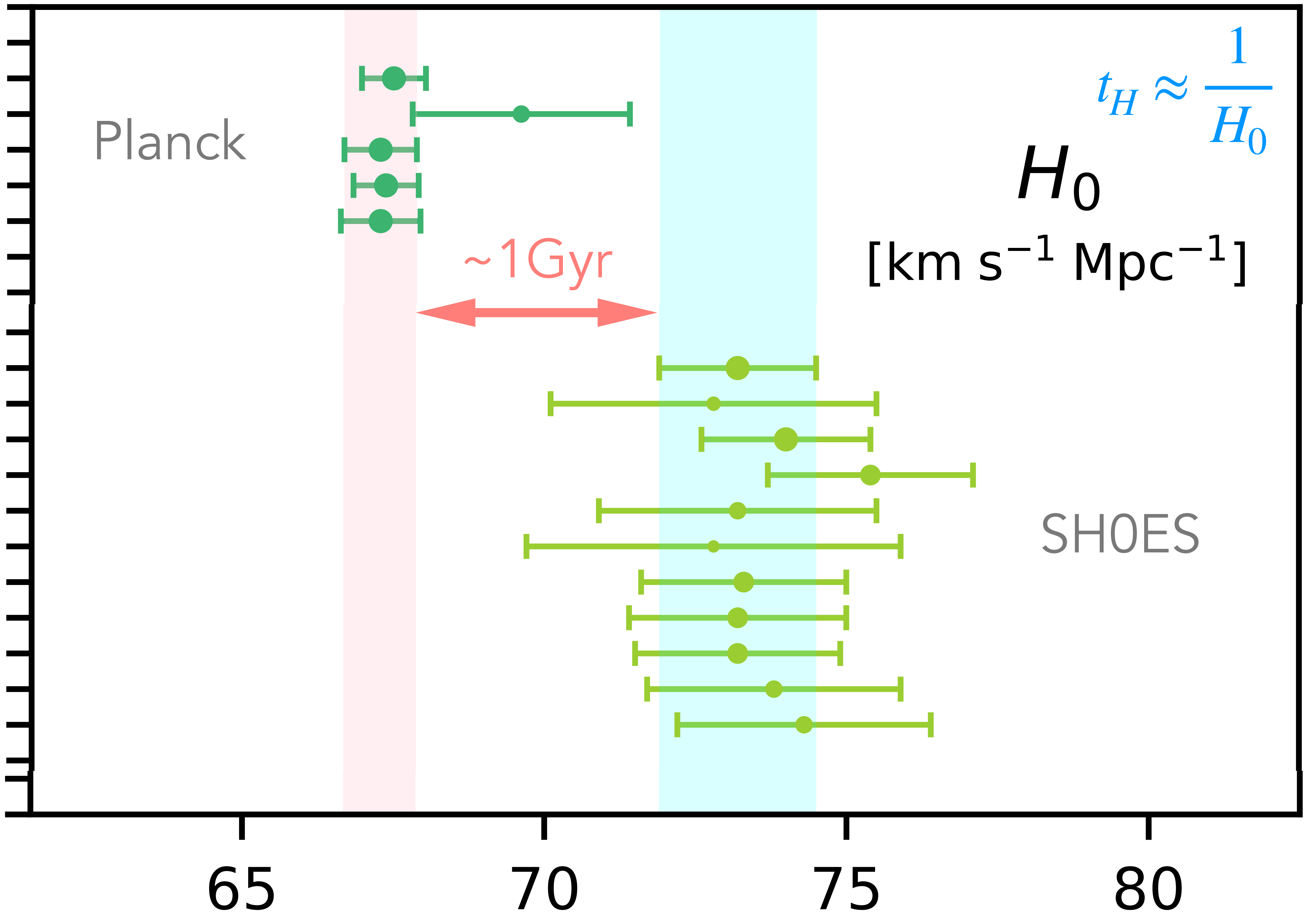
Challenges in Cosmology

Hubble Tension



Di Valentino et al. (2021)

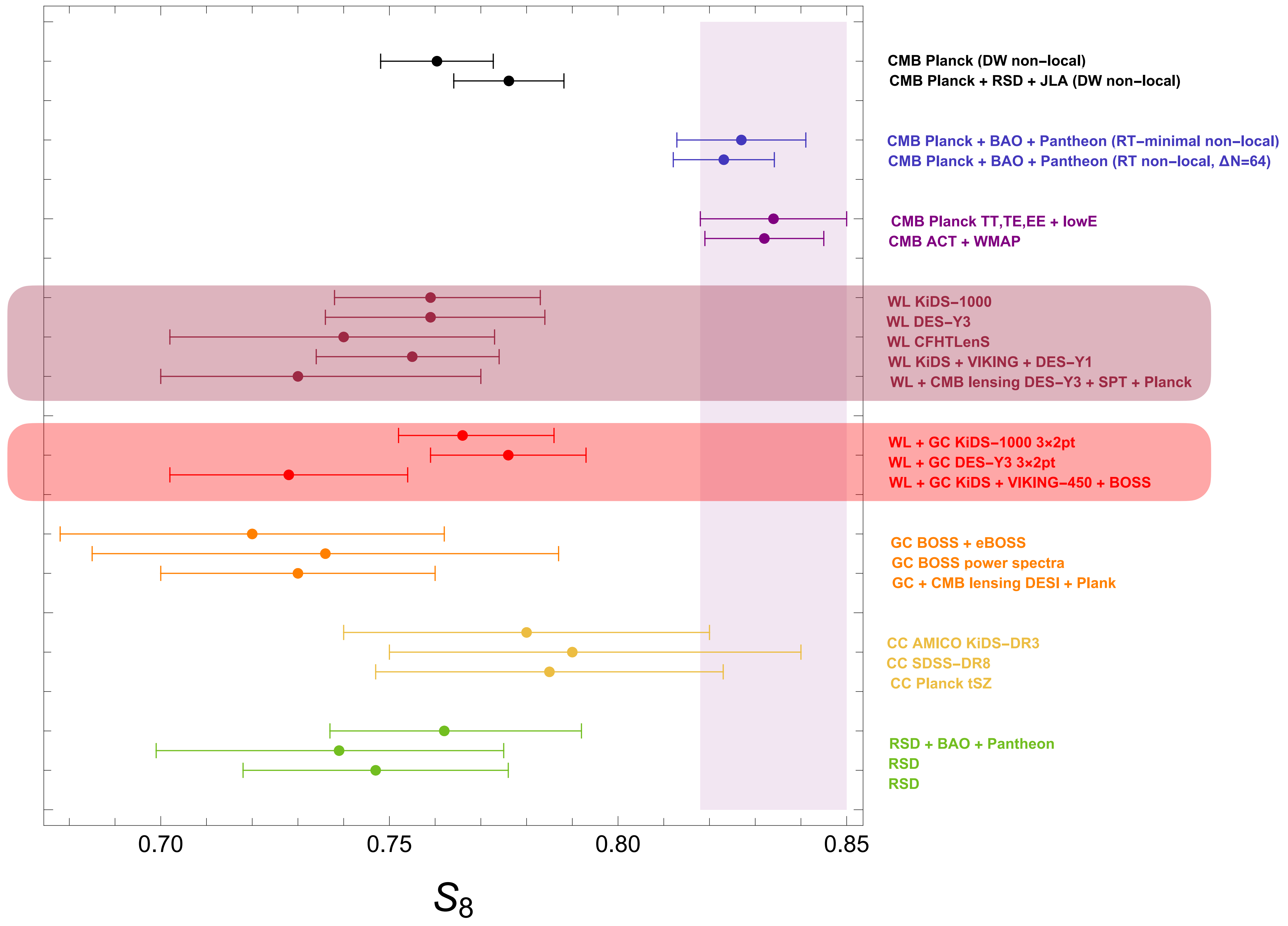
Age of the Universe



Challenges in Cosmology

Mass and Structure of the Universe

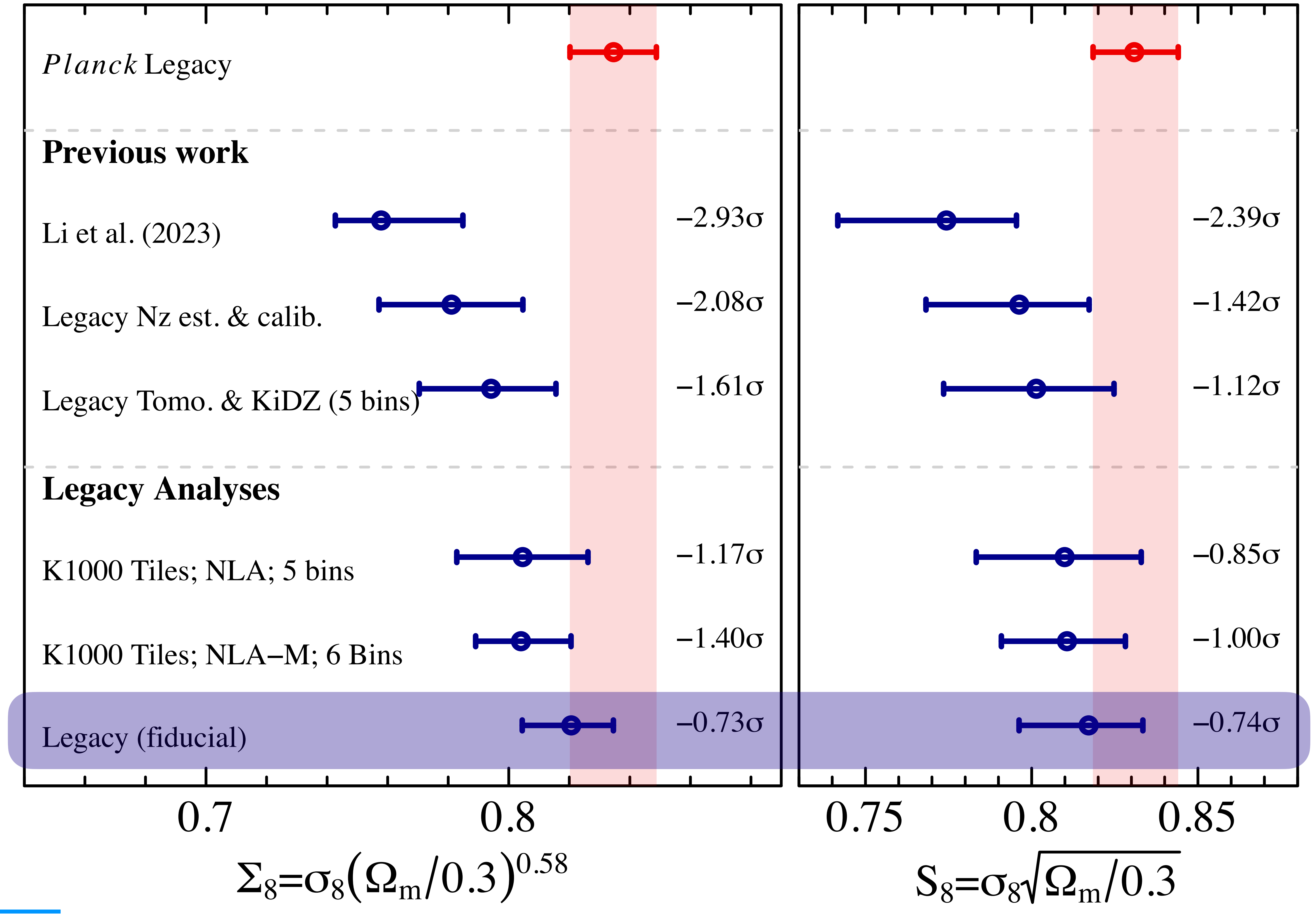
S_8 Tension



Bouché et al. (2022)

$$S_8 = \sigma_8\sqrt{\Omega_m/0.3}$$

KiDS Legacy



Wright et al. (2022)

Challenges in Cosmology

Missions

~2-5 PB

~1 PB

~10-20 PB

~30-60 PB

~20-40 PB

~600 PB/year

~10-15 PB

~1-2 PB

Outline

1. Cosmological motivation

- What are the **problems** we need to solve?
- How does **data** play a role in this?
- Why do we need **AI** and how can we trust the results?

2. Weak gravitational lensing

- 'Observing' **invisible matter**
- Technical **challenges**

3. AI for cosmological data analysis

- Estimating **stellar SEDs** from photometric data
- Identify of **blended galaxies** in survey data
- Going beyond

4. Future perspectives

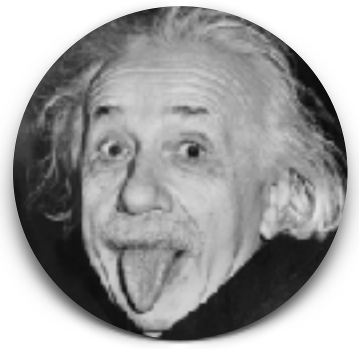
- Where can AI take us?
- Accessible **tools**
- Conclusions

Weak Lensing

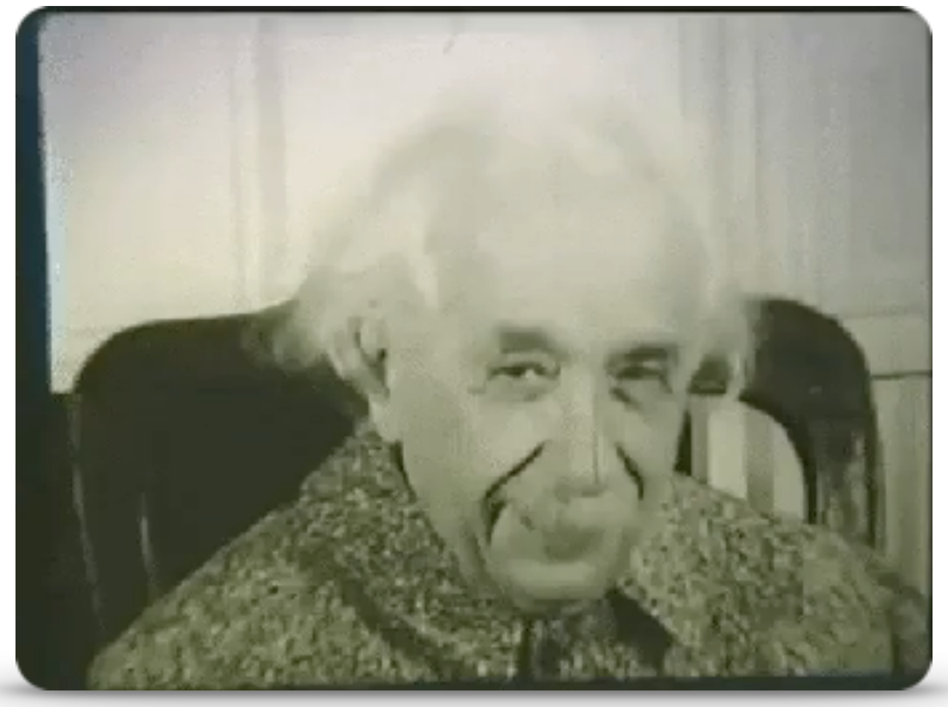
2. Weak gravitational lensing

- ‘Observing’ **invisible matter**
- Technical **challenges**

Weak Lensing



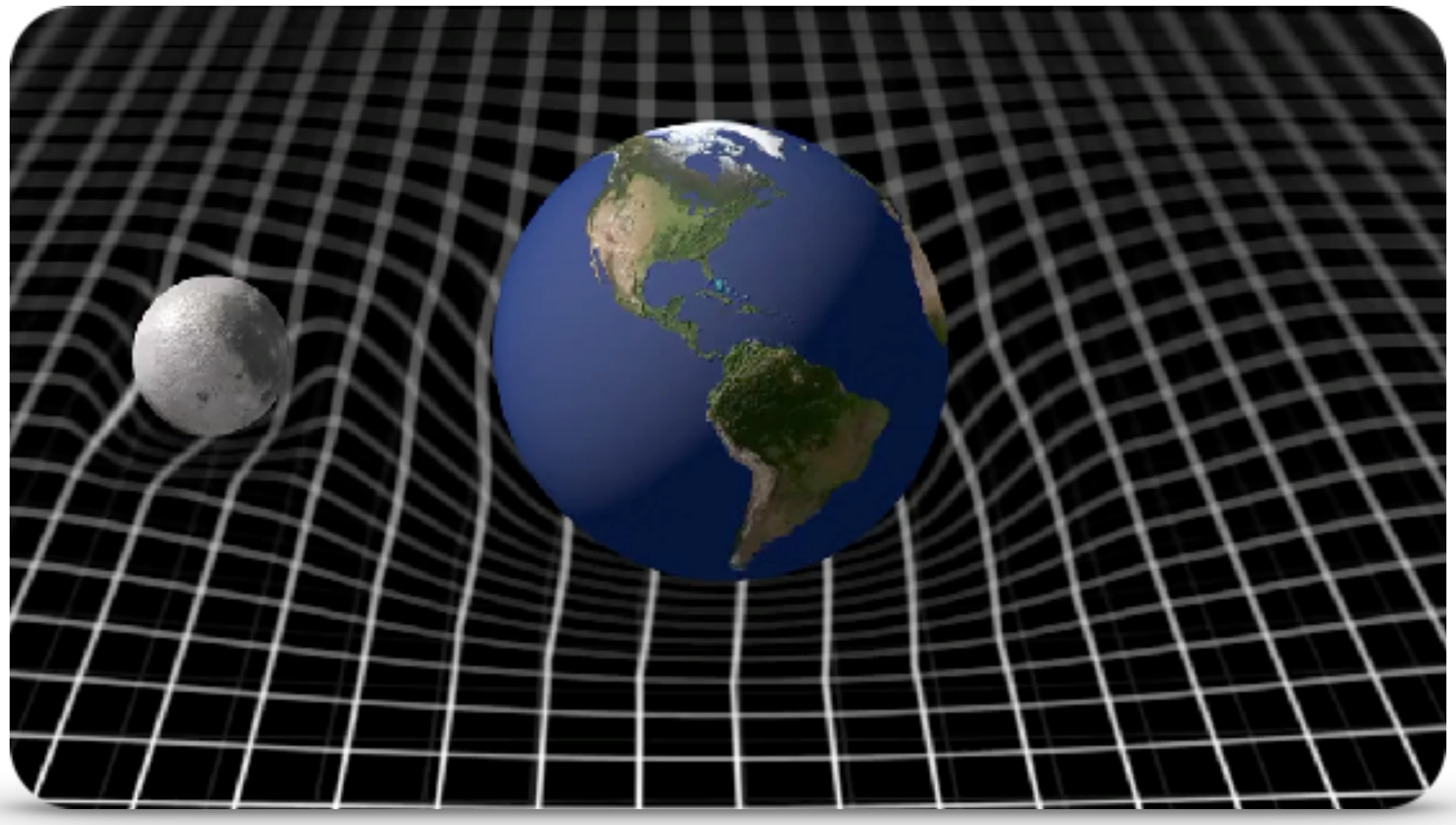
Albert Einstein

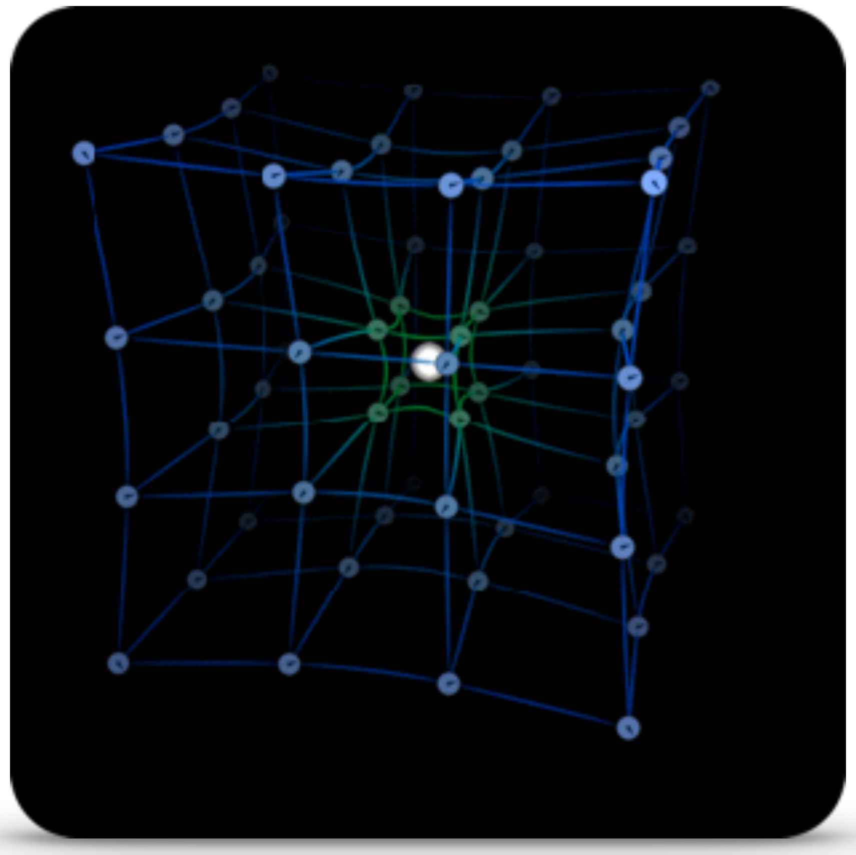


General Relativity

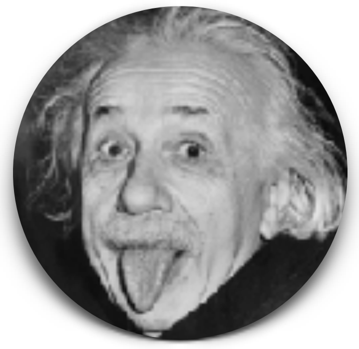
$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu}$$

'Spacetime tells matter how to move;
matter tells spacetime how to curve.'
- John Wheeler



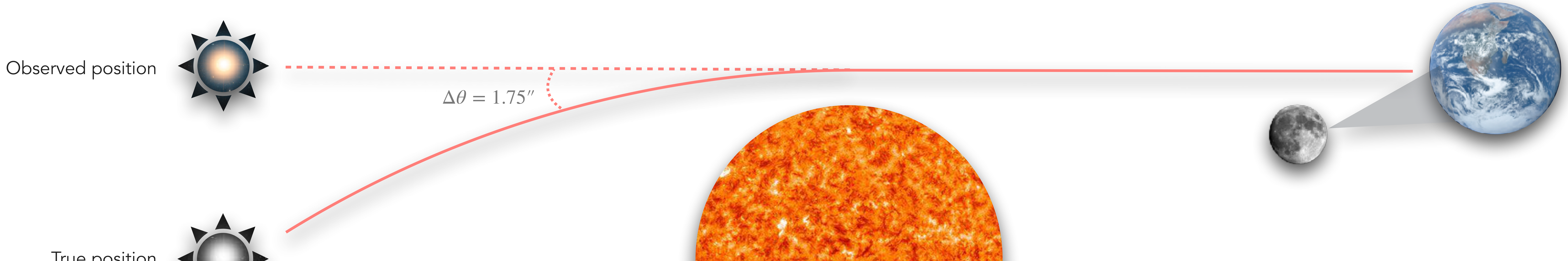


Weak Lensing



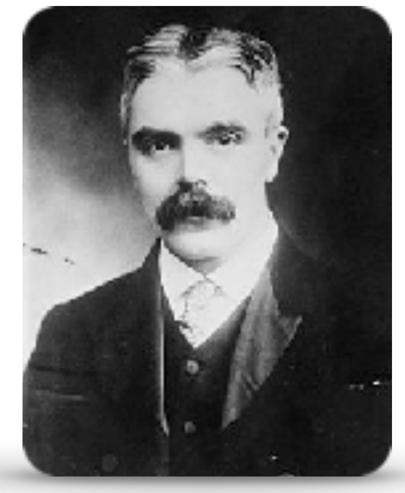
Albert Einstein

General Relativity



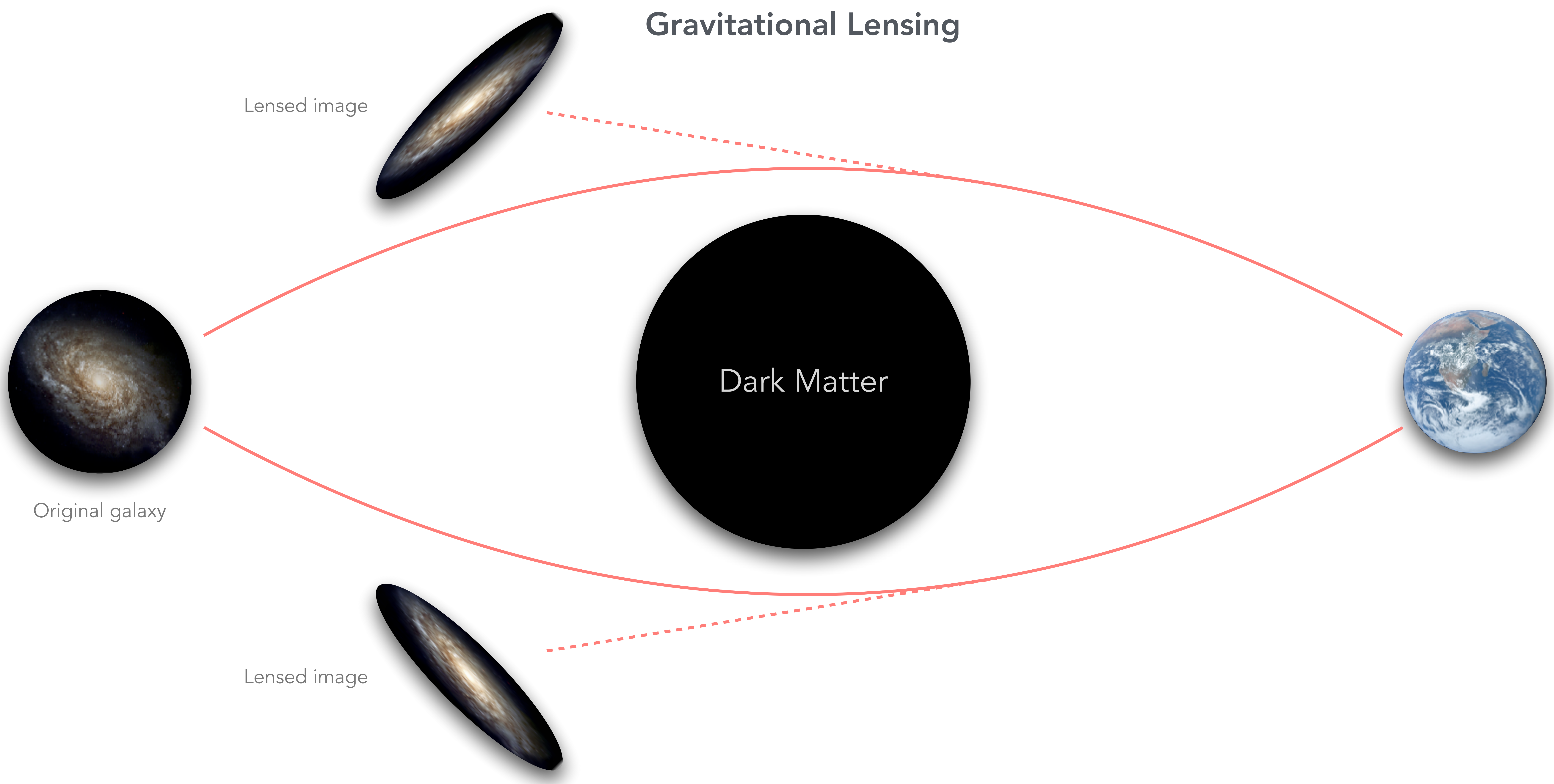
Eddington experiment 1919

Eddington



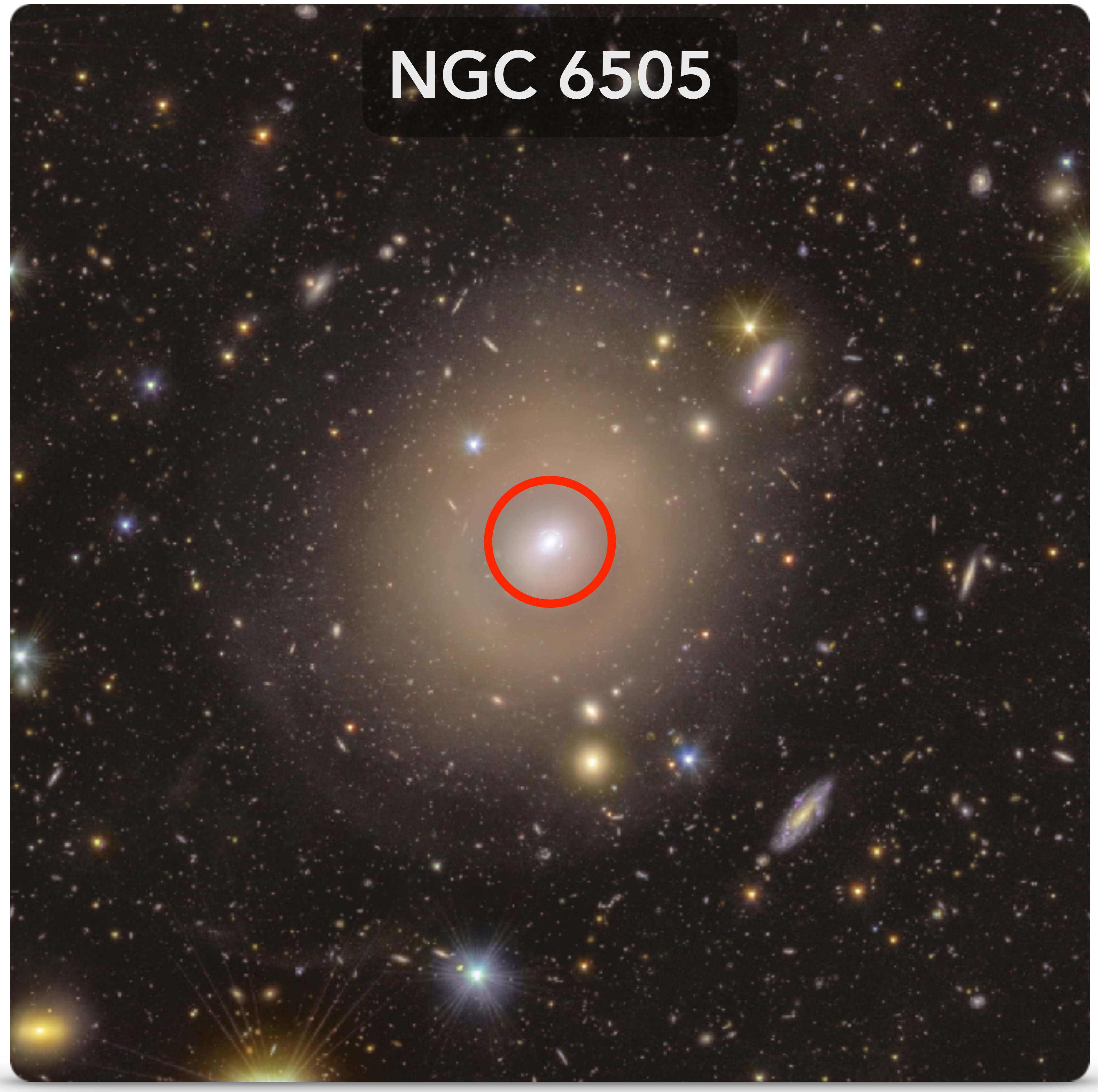
Dyson

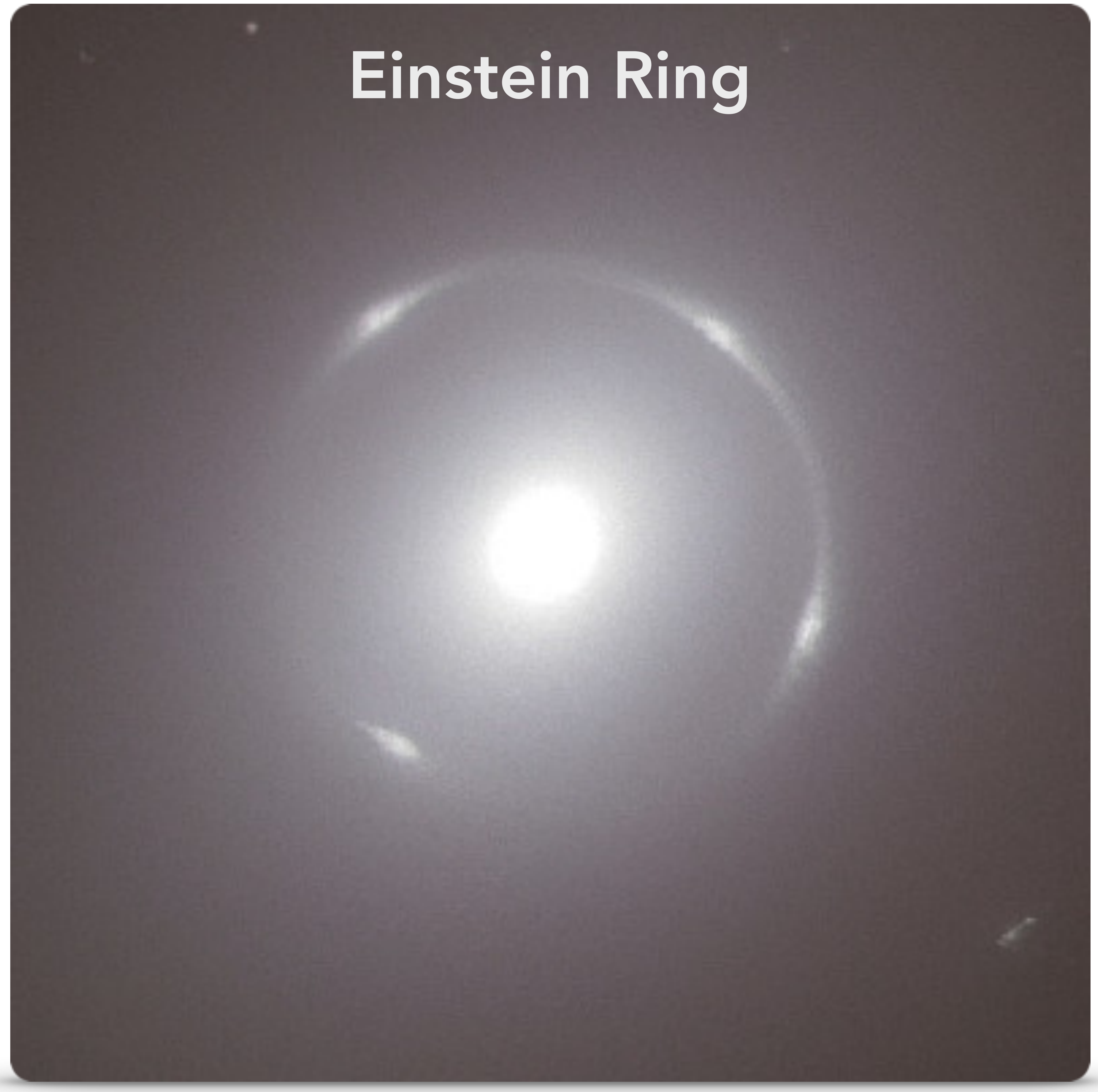
Weak Lensing



Weak Lensing

Strong Gravitational Lensing

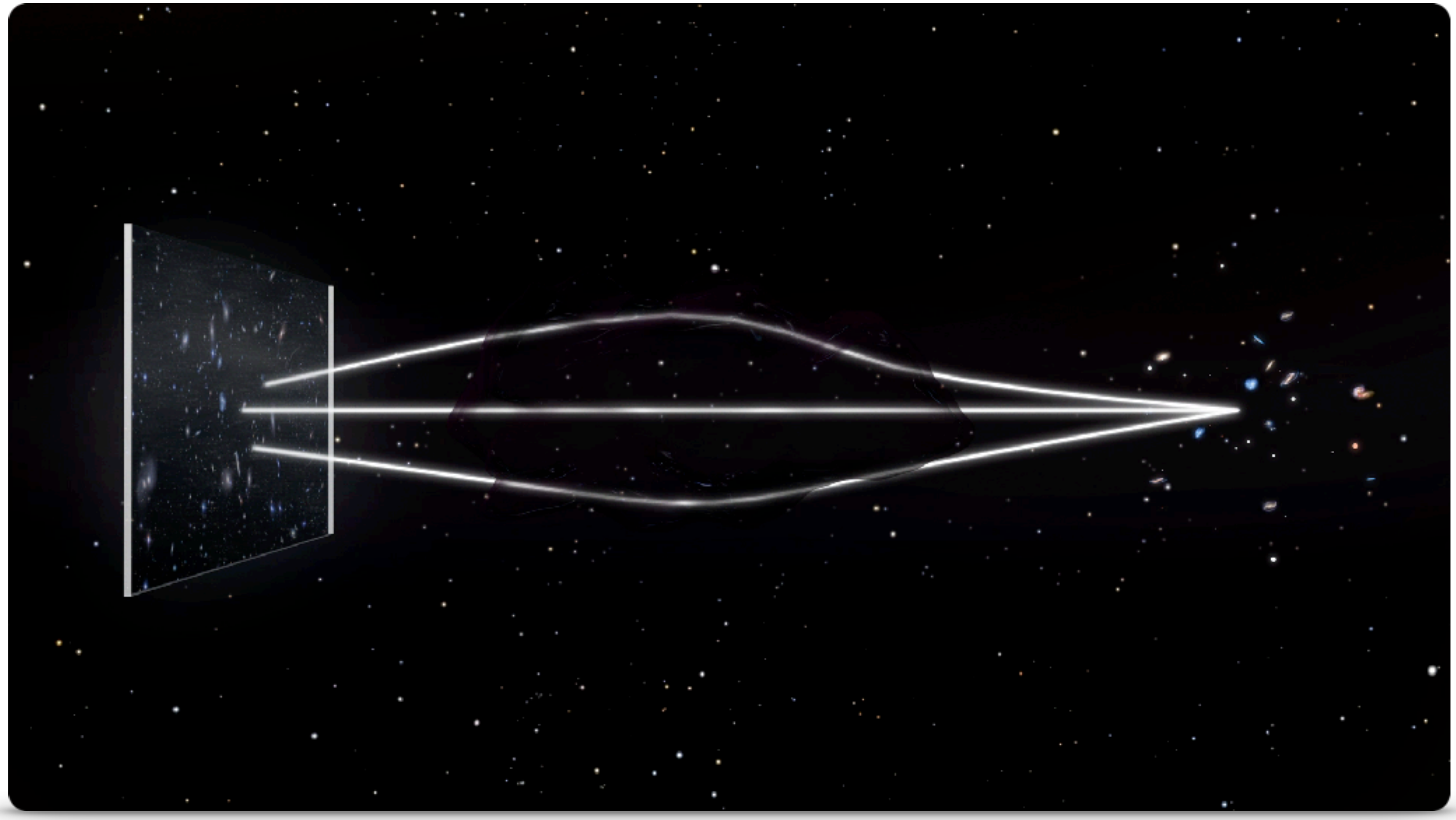




Credit: ESA/Euclid/Euclid Consortium/NASA, image processing by J.-C. Cuillandre, T. Li

Weak Lensing

Weak Gravitational Lensing



Credit: NASA's Goddard Space Flight Center Conceptual Image Lab

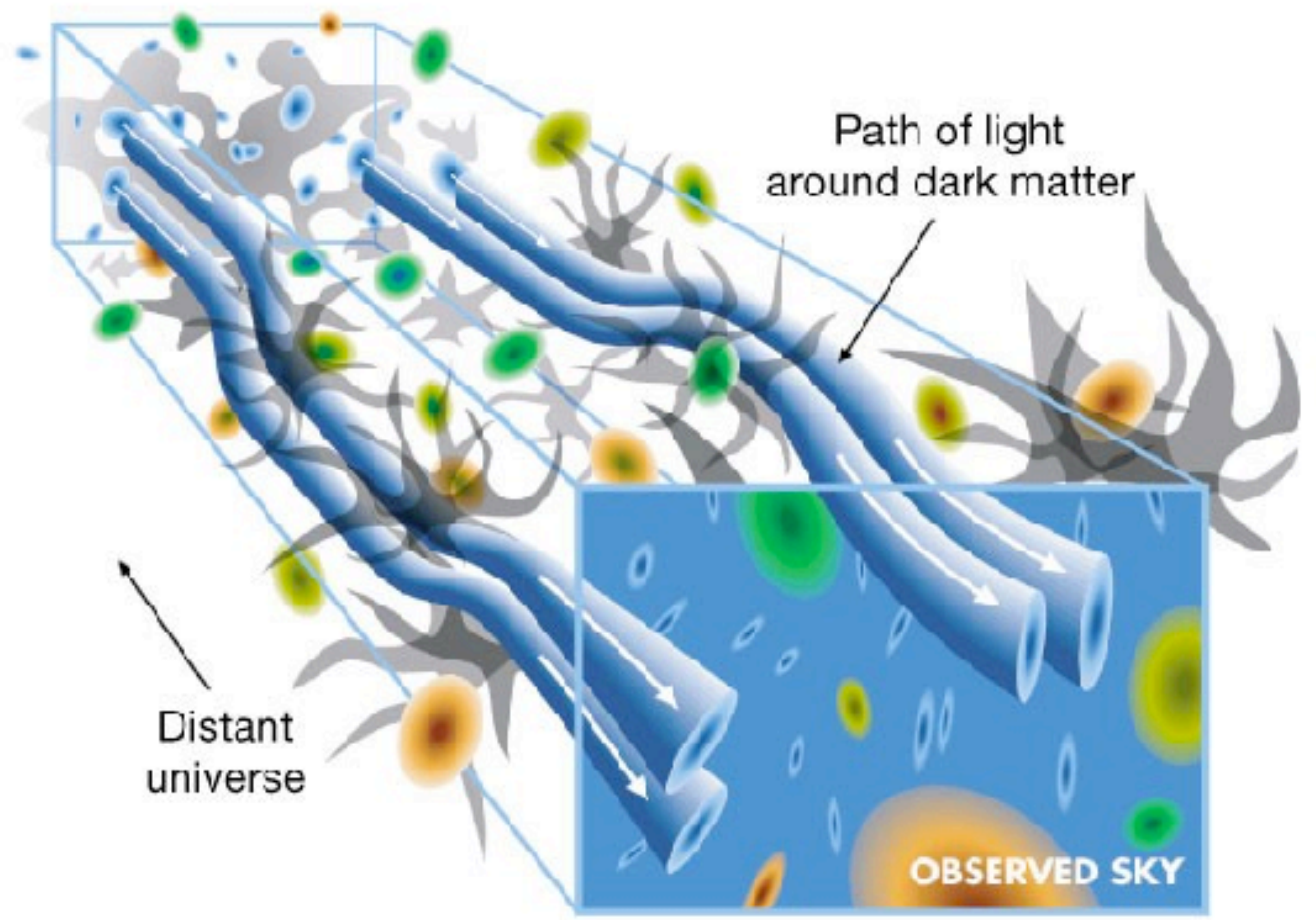
Credit: ESA/Hubble (M. Kornmesser & L. L. Christensen)

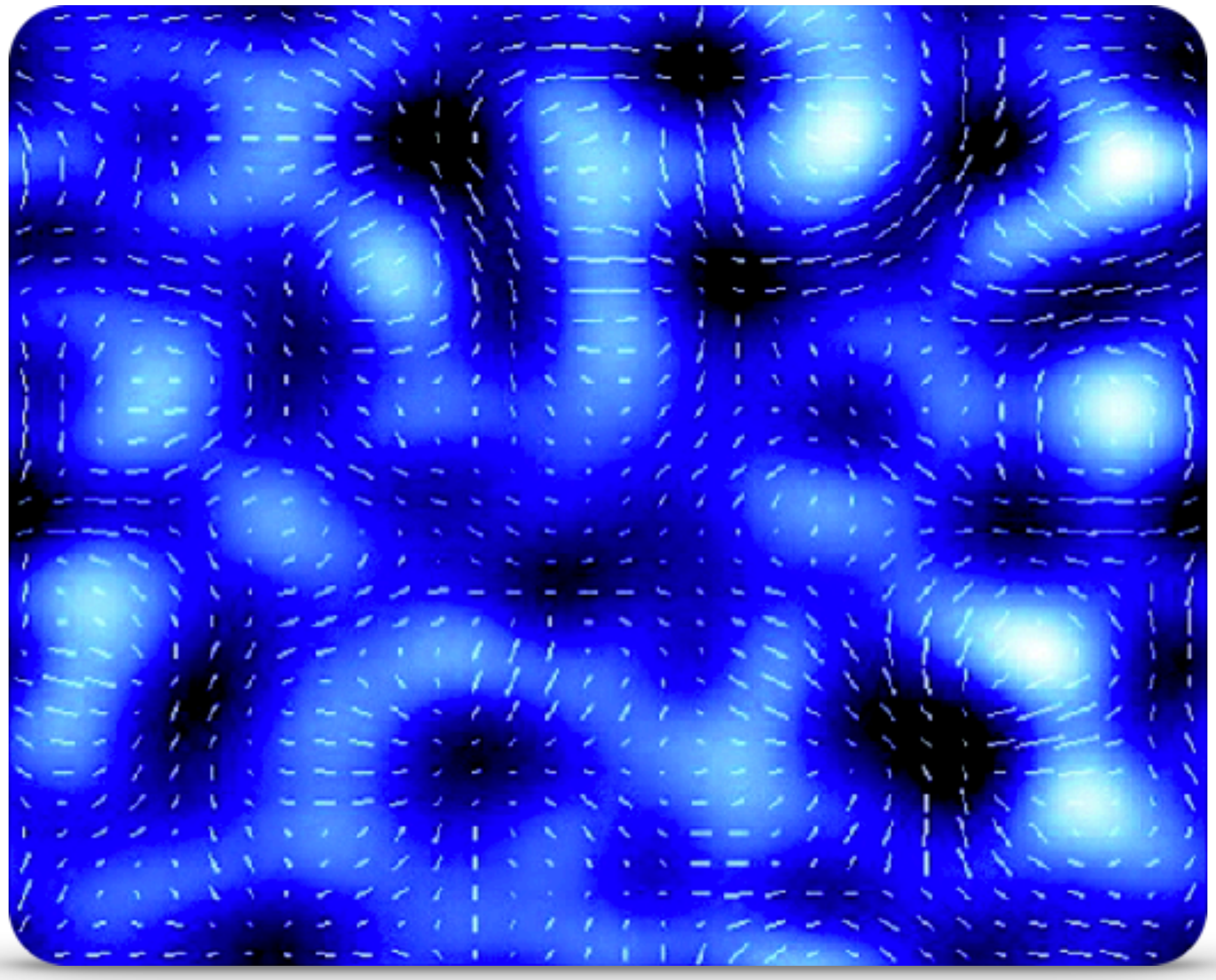
Weak Lensing

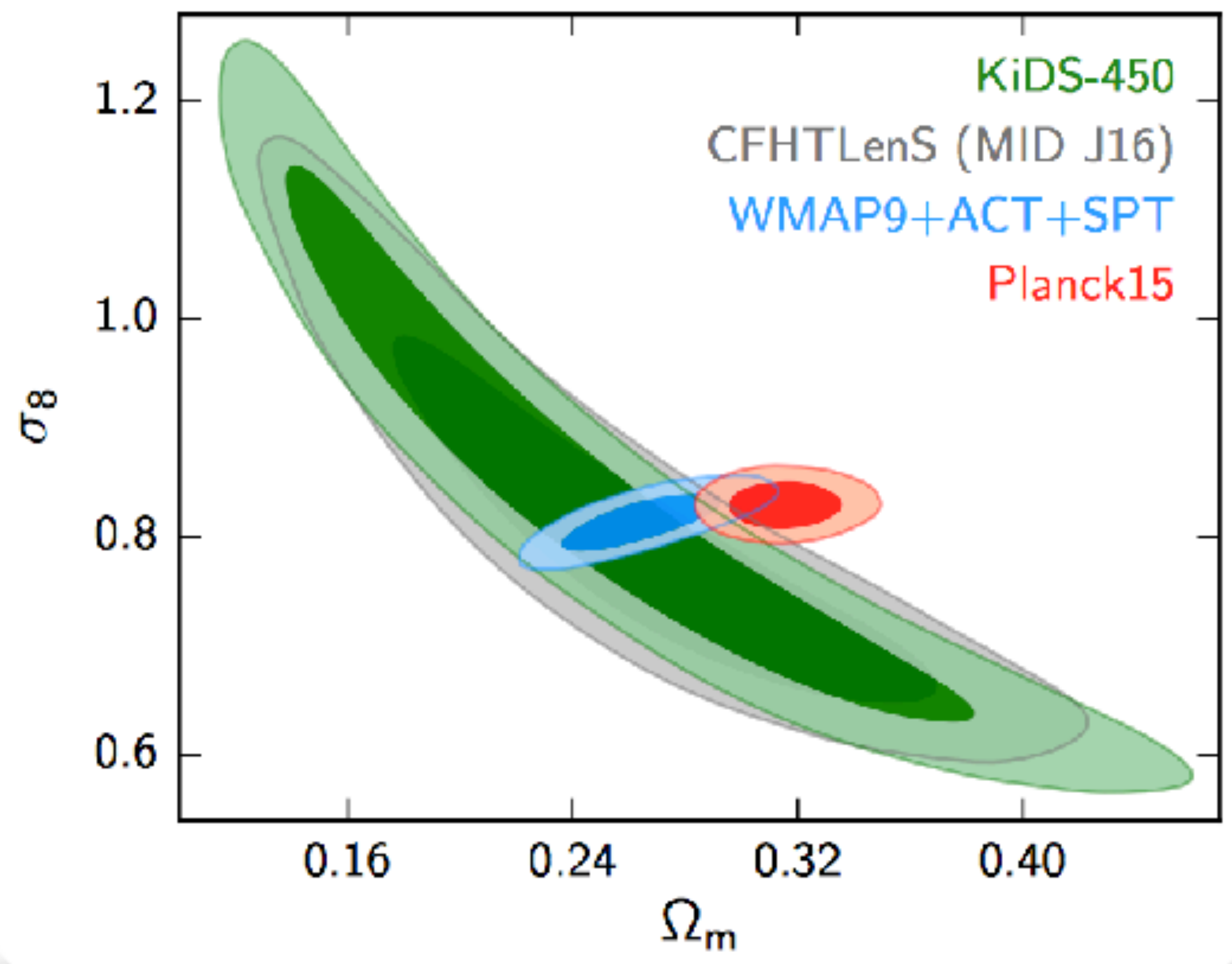
Weak Gravitational Lensing

Weak Lensing

Weak Gravitational Lensing







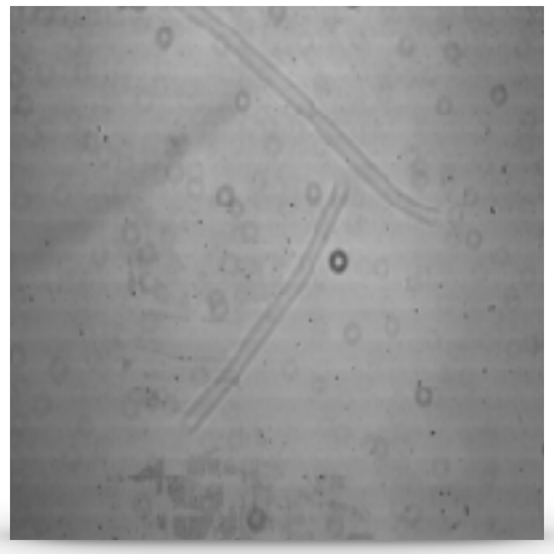
- Consider all galaxies and all the matter the light encounters along the line of sight

- Measure the shapes of galaxies to infer the amount of (mostly dark) matter needed to induce the 'squishing' we observe
- There is a ~1% change to the shape on average

- Use the statistics of millions of galaxies to put constraints on cosmological parameters and hence determine the amount of dark matter in the Universe

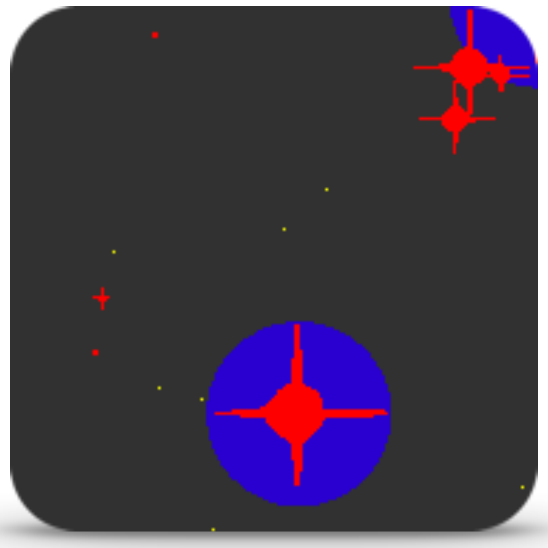
Weak Lensing

Raw images



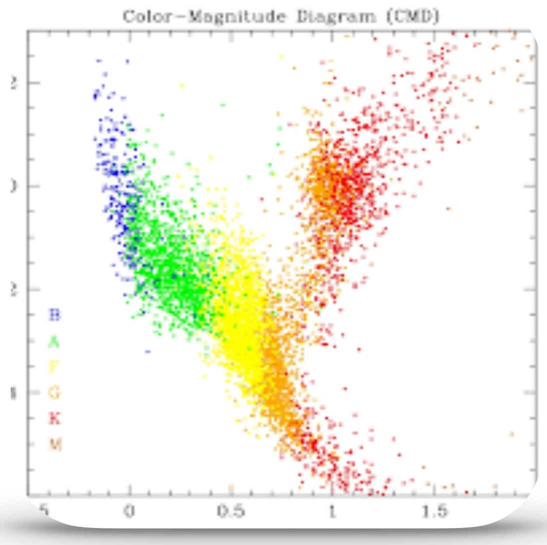
Reduced images

Masking



Source detection

Photometry



Source selection

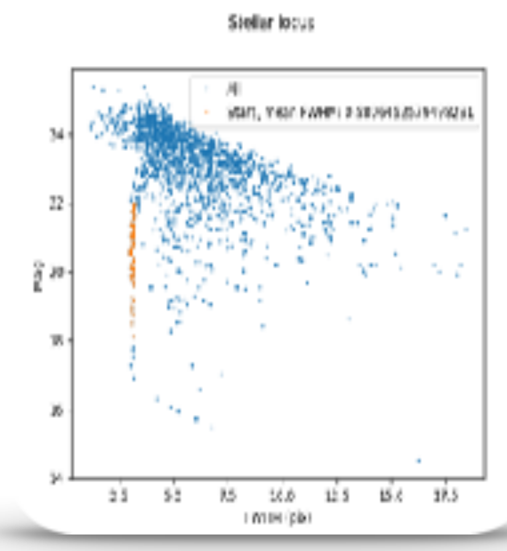
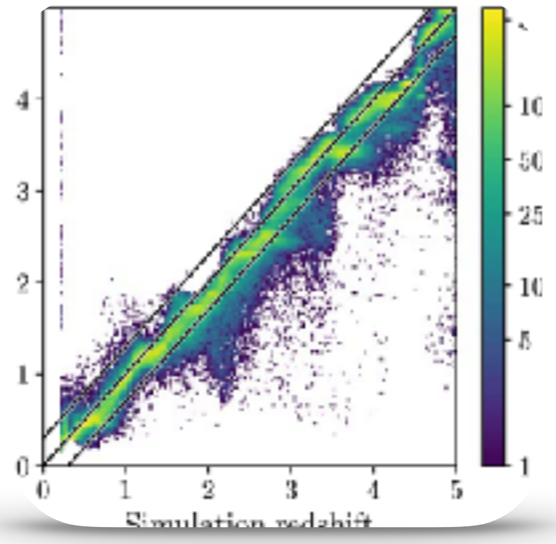


Photo-z



Just an illustration. Many steps not listed and order of steps more nuanced in reality.

Lots of Software!

We need to be able to trust and understand all of these steps

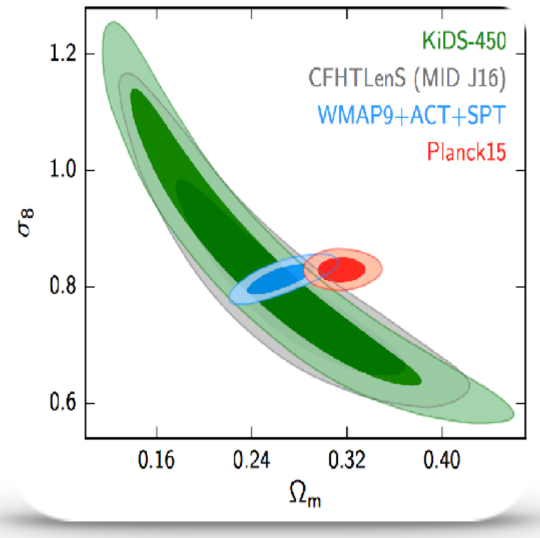
PSF estimation

PSF validation

Shape measurement

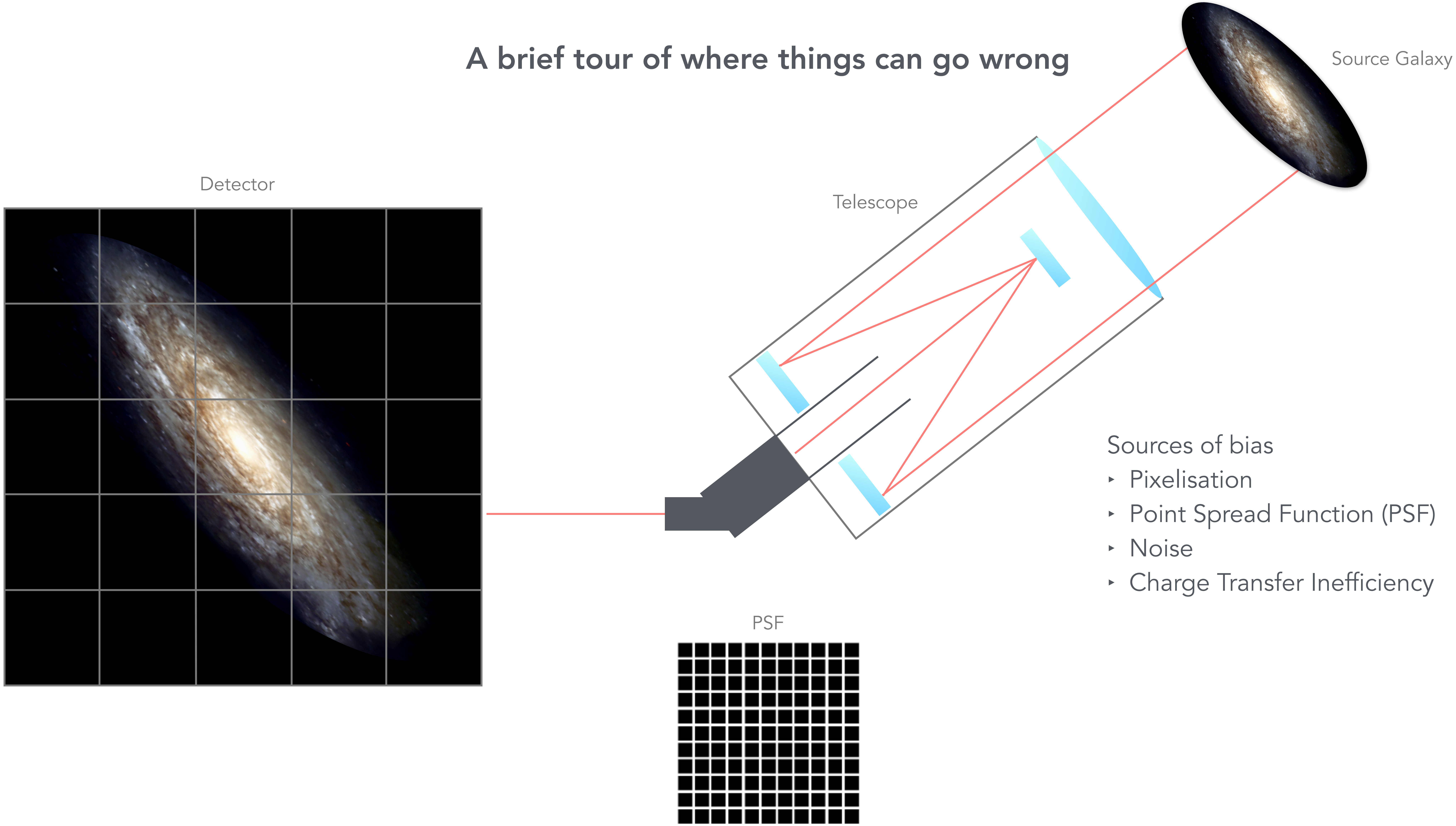
Shape calibration

Cosmology



Weak Lensing

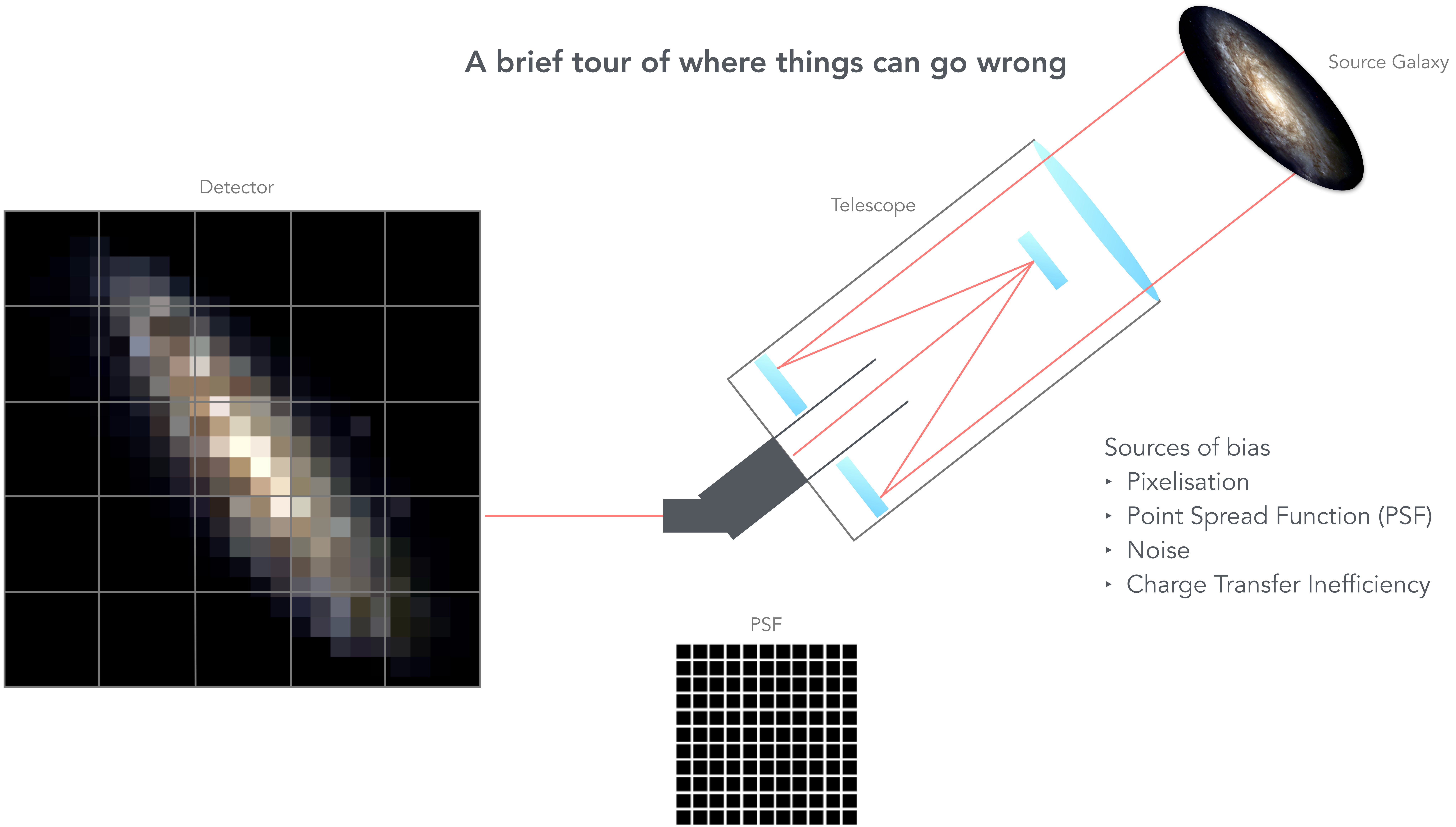
A brief tour of where things can go wrong



Sources of bias
- Pixelisation
- Point Spread Function (PSF)
- Noise
- Charge Transfer Inefficiency

Weak Lensing

A brief tour of where things can go wrong

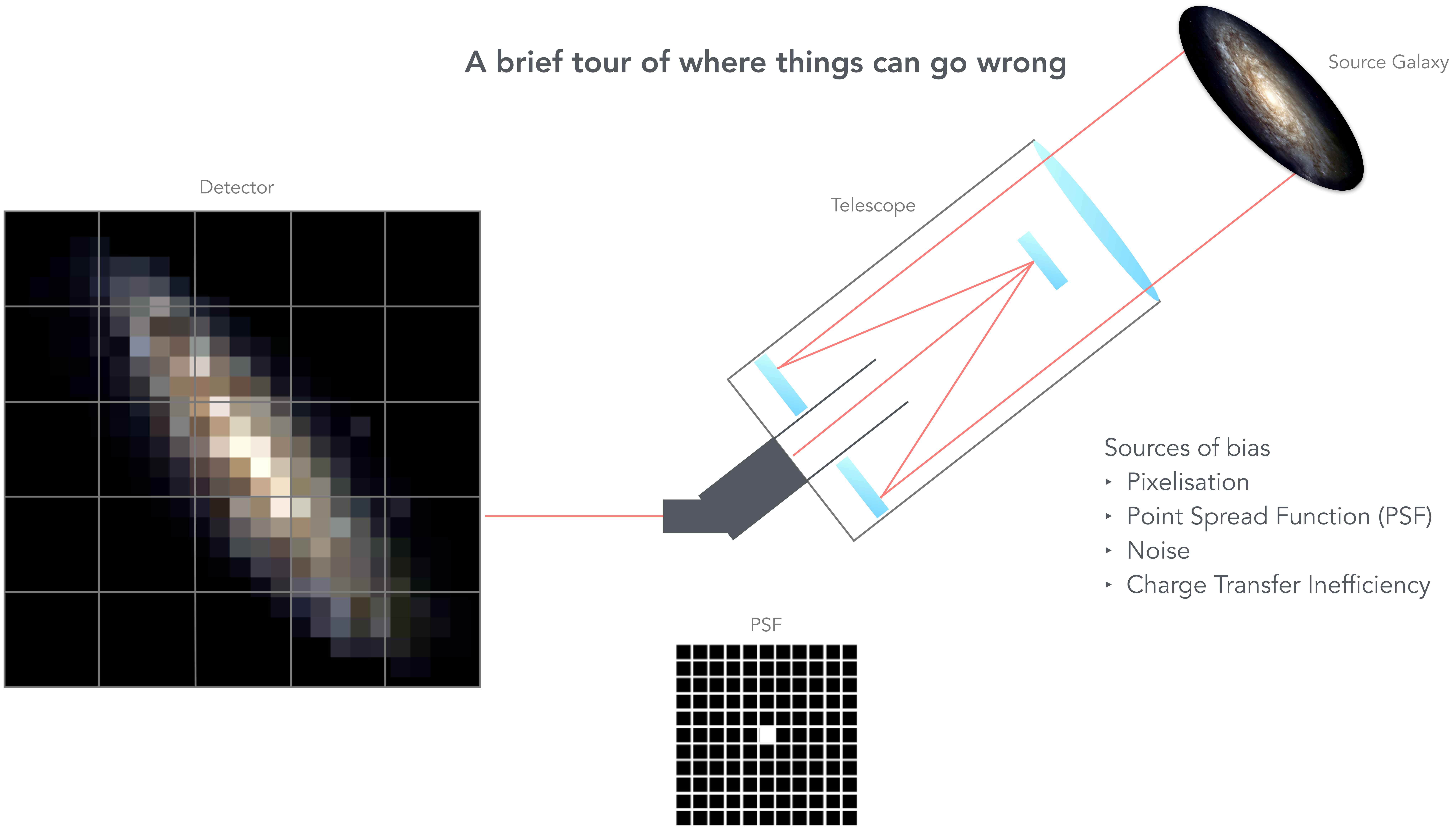


Sources of bias

- Pixelisation
- Point Spread Function (PSF)
- Noise
- Charge Transfer Inefficiency

Weak Lensing

A brief tour of where things can go wrong

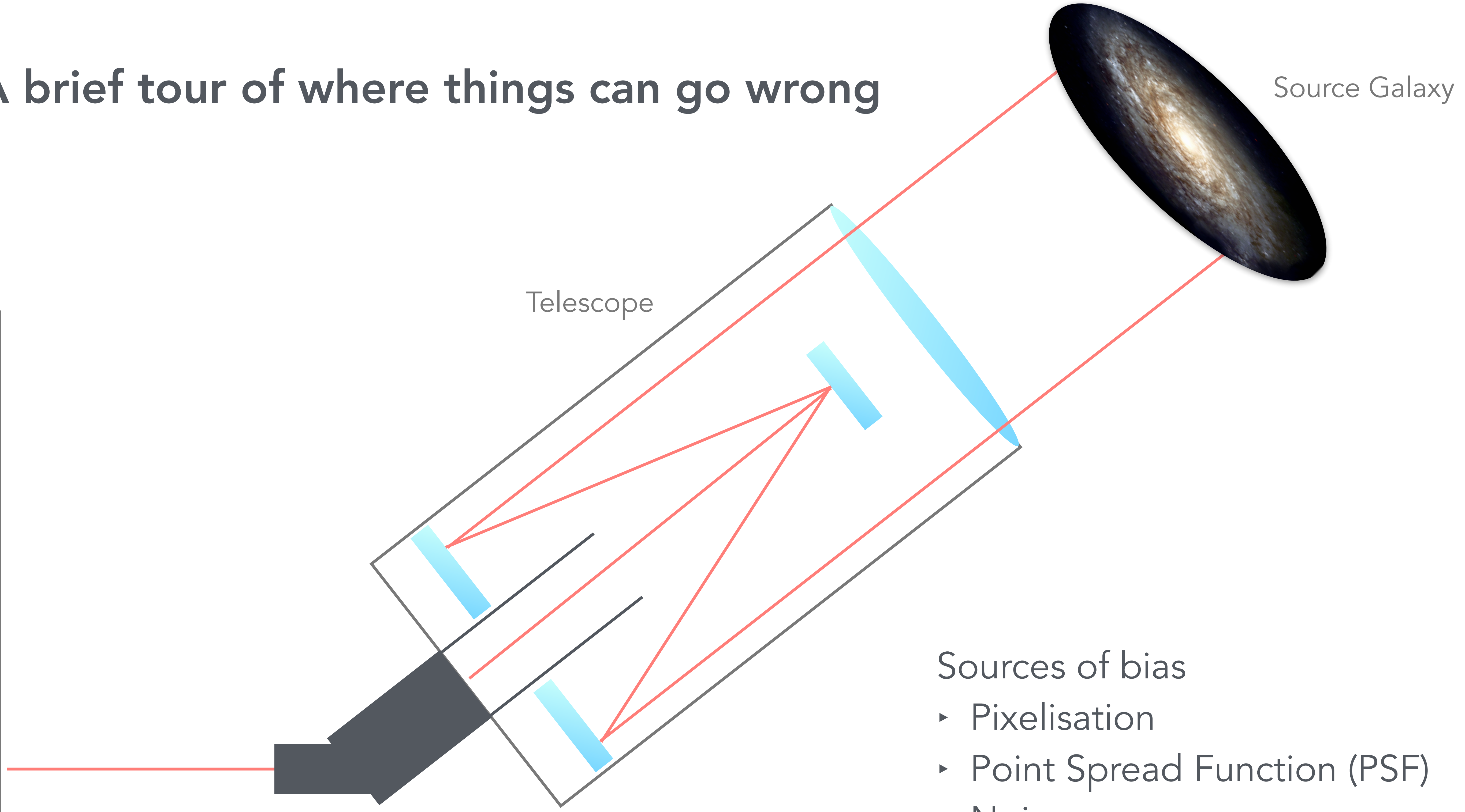


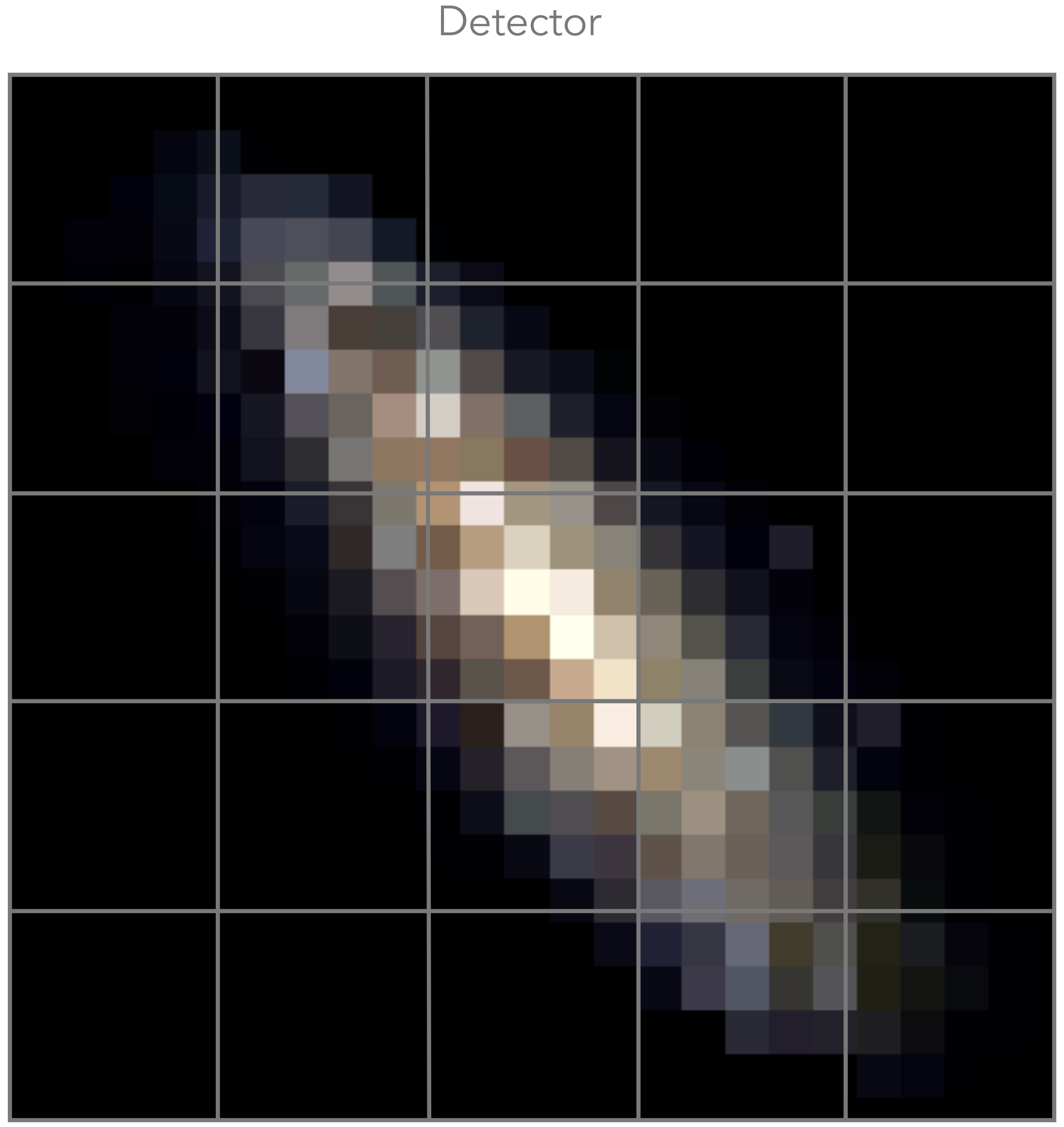
Sources of bias

- Pixelisation
- Point Spread Function (PSF)
- Noise
- Charge Transfer Inefficiency

Weak Lensing

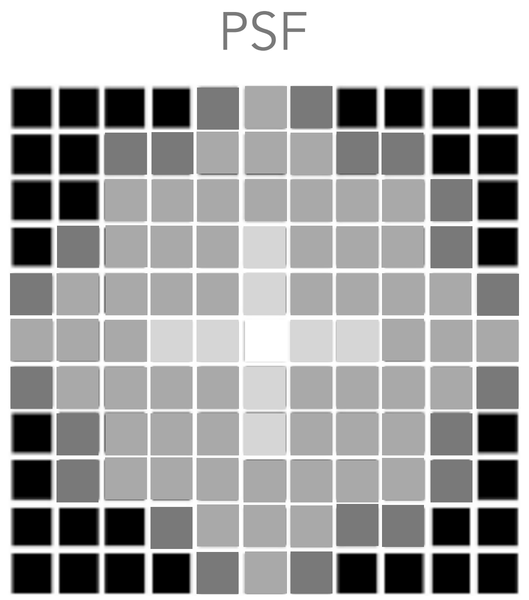
A brief tour of where things can go wrong





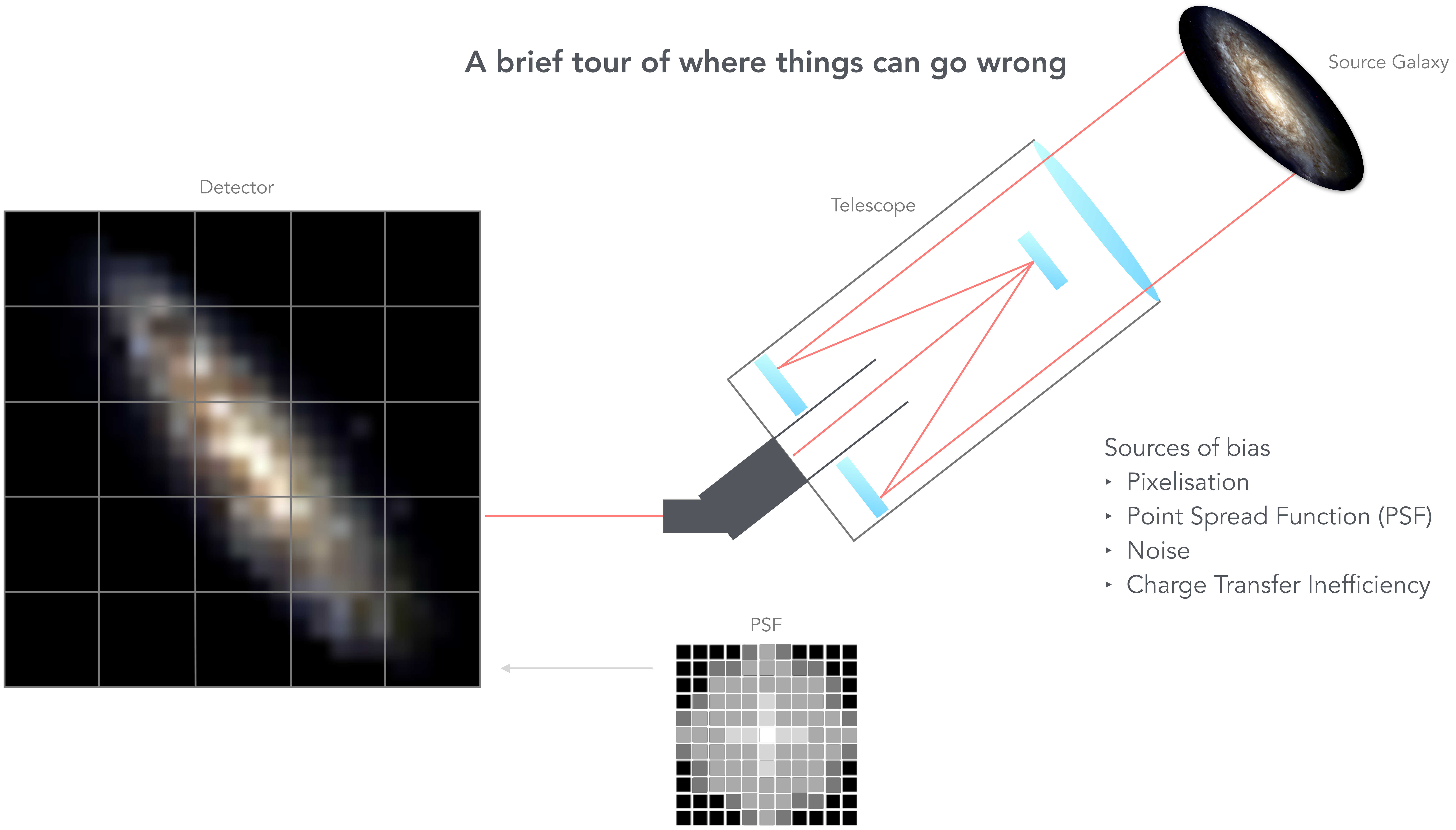
Sources of bias

- Pixelisation
- Point Spread Function (PSF)
- Noise
- Charge Transfer Inefficiency



Weak Lensing

A brief tour of where things can go wrong

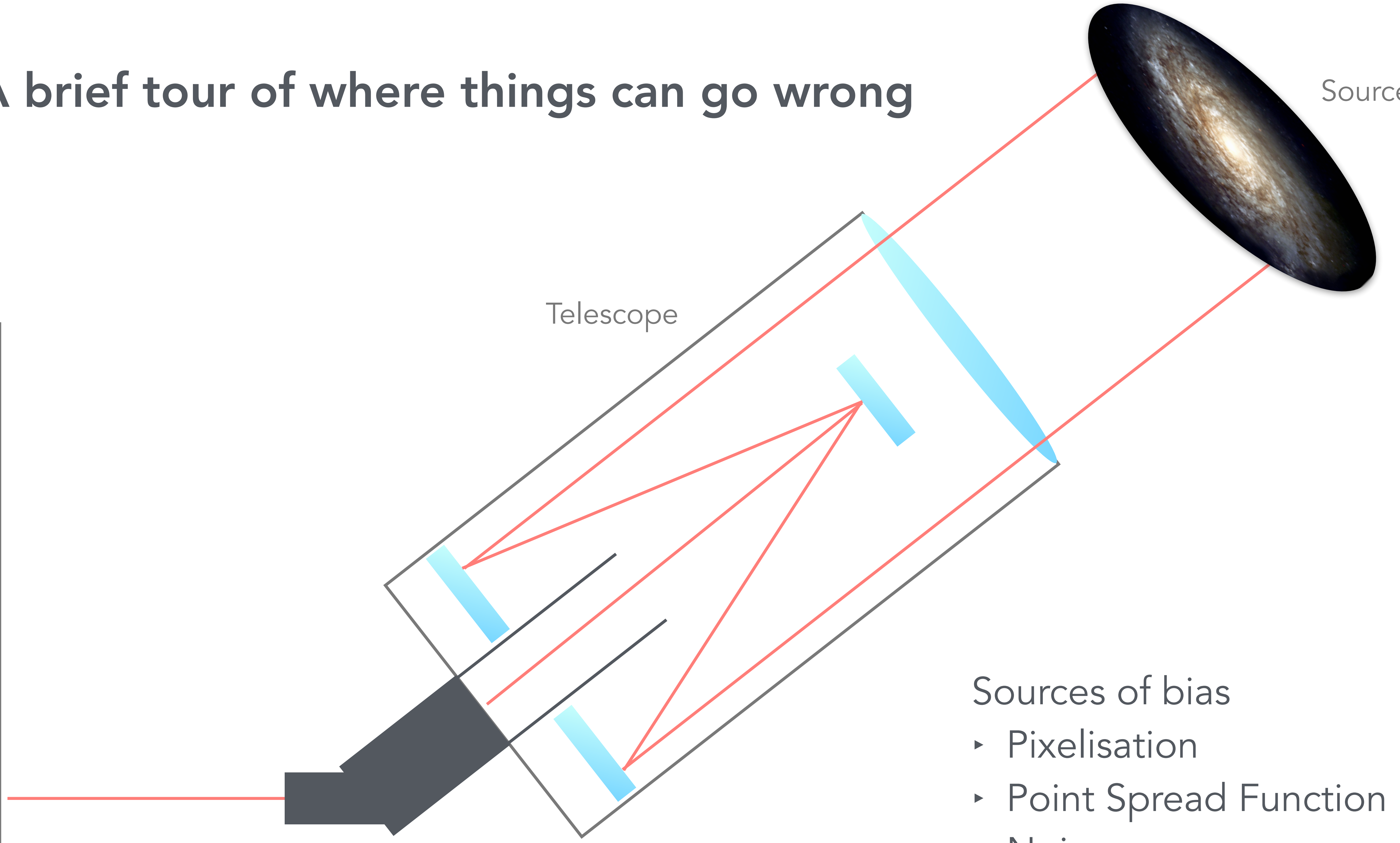


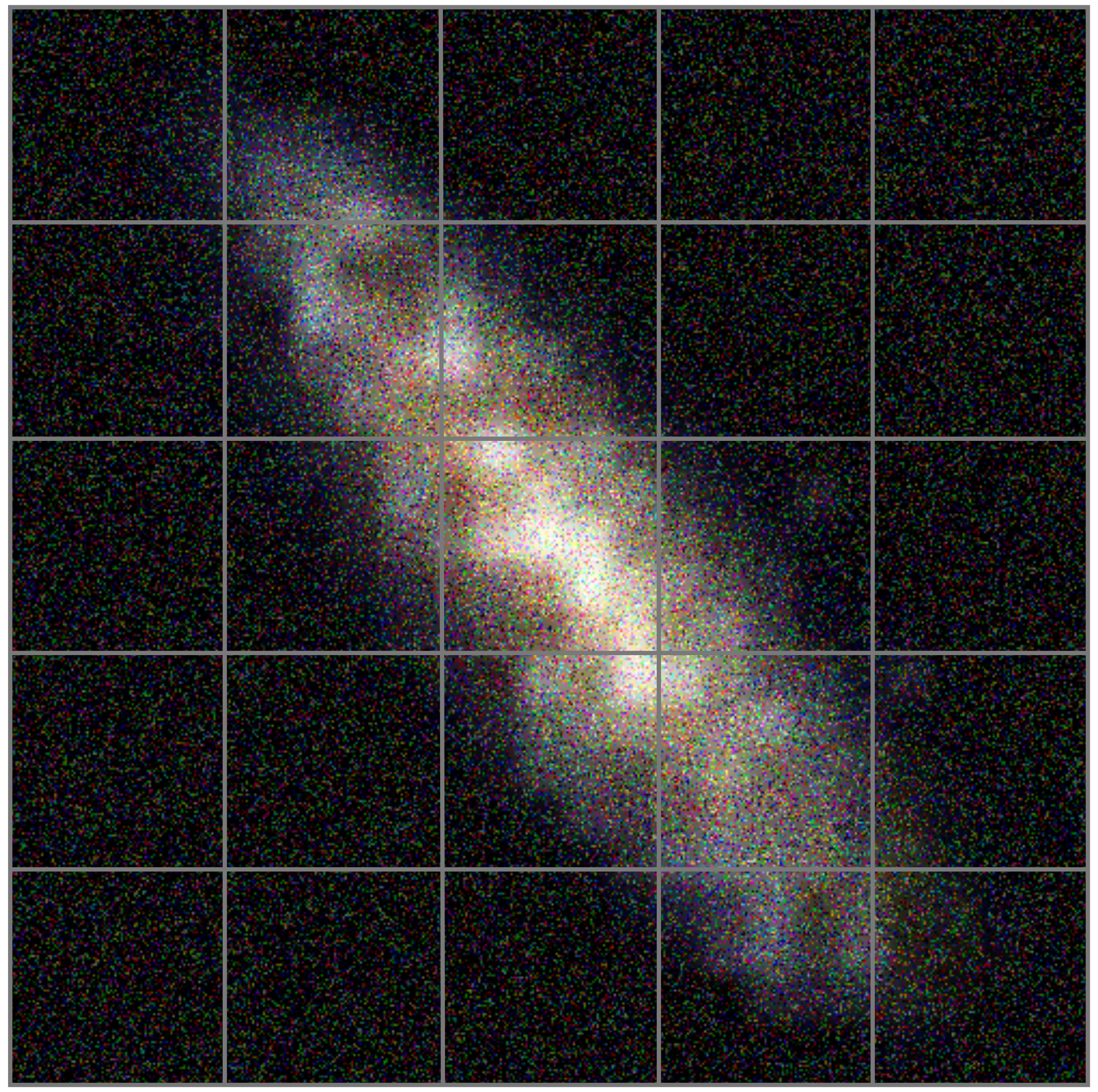
Sources of bias

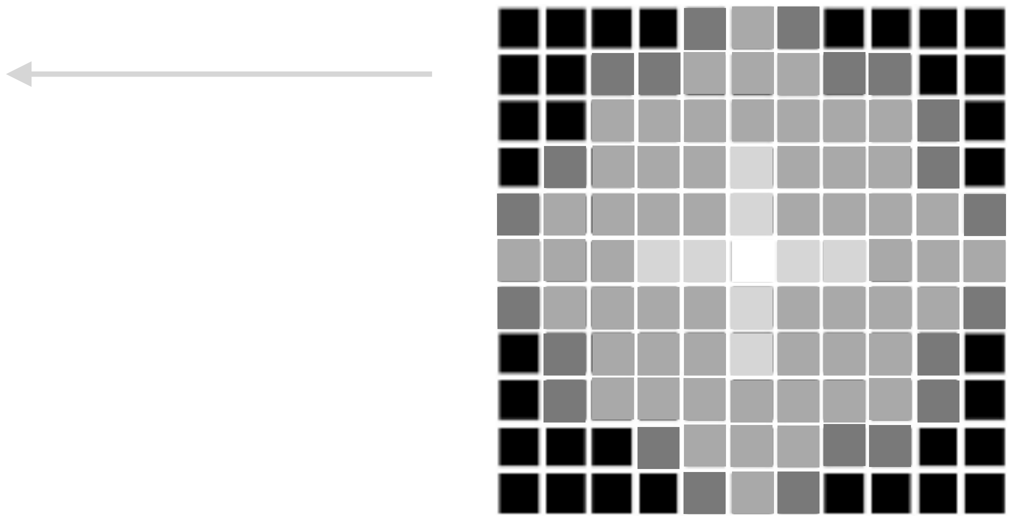
- Pixelisation
- Point Spread Function (PSF)
- Noise
- Charge Transfer Inefficiency

Weak Lensing

A brief tour of where things can go wrong





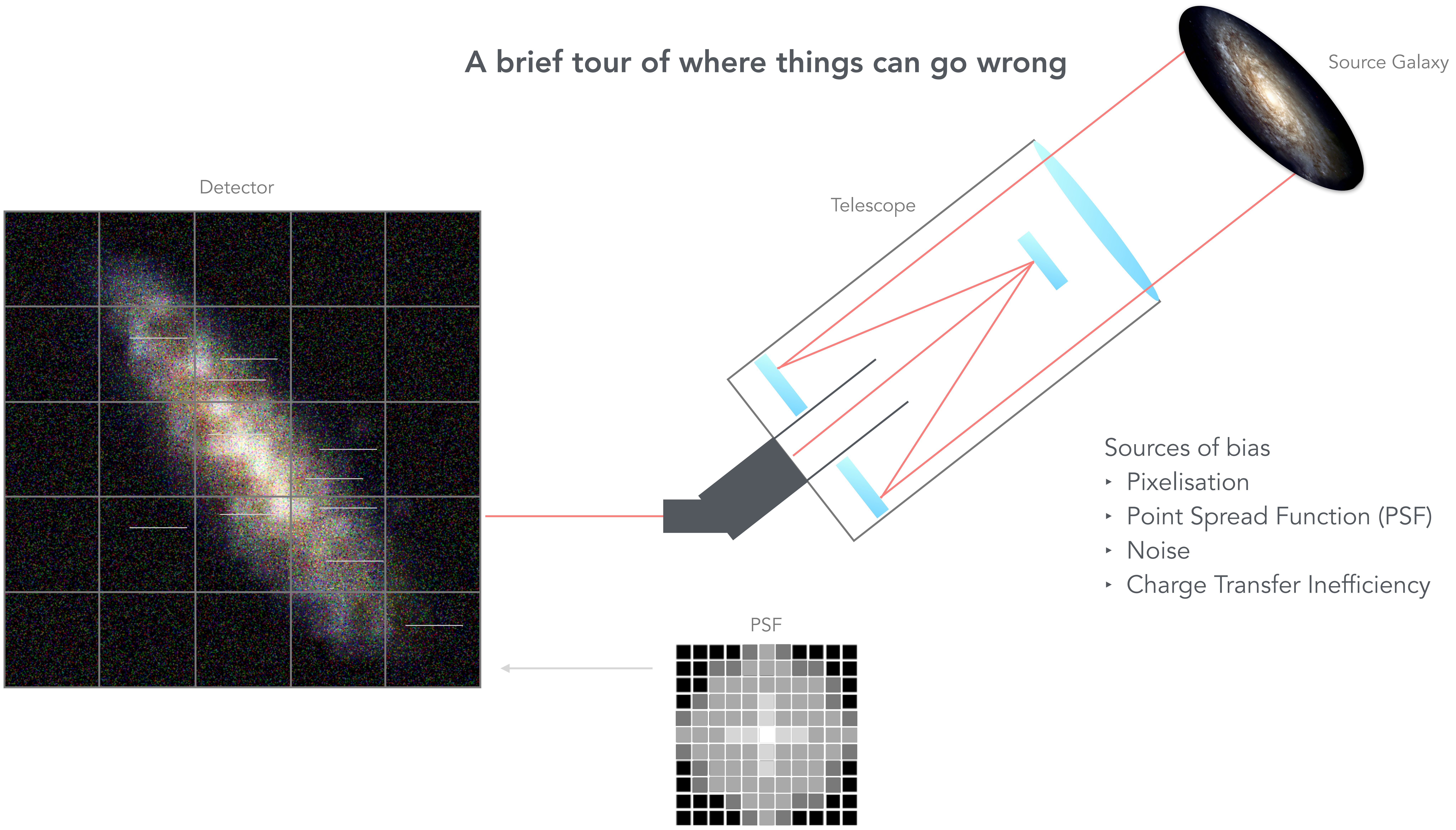


Sources of bias

- Pixelisation
- Point Spread Function (PSF)
- Noise
- Charge Transfer Inefficiency

Weak Lensing

A brief tour of where things can go wrong



Sources of bias
- Pixelisation
- Point Spread Function (PSF)
- Noise
- Charge Transfer Inefficiency

Weak Lensing

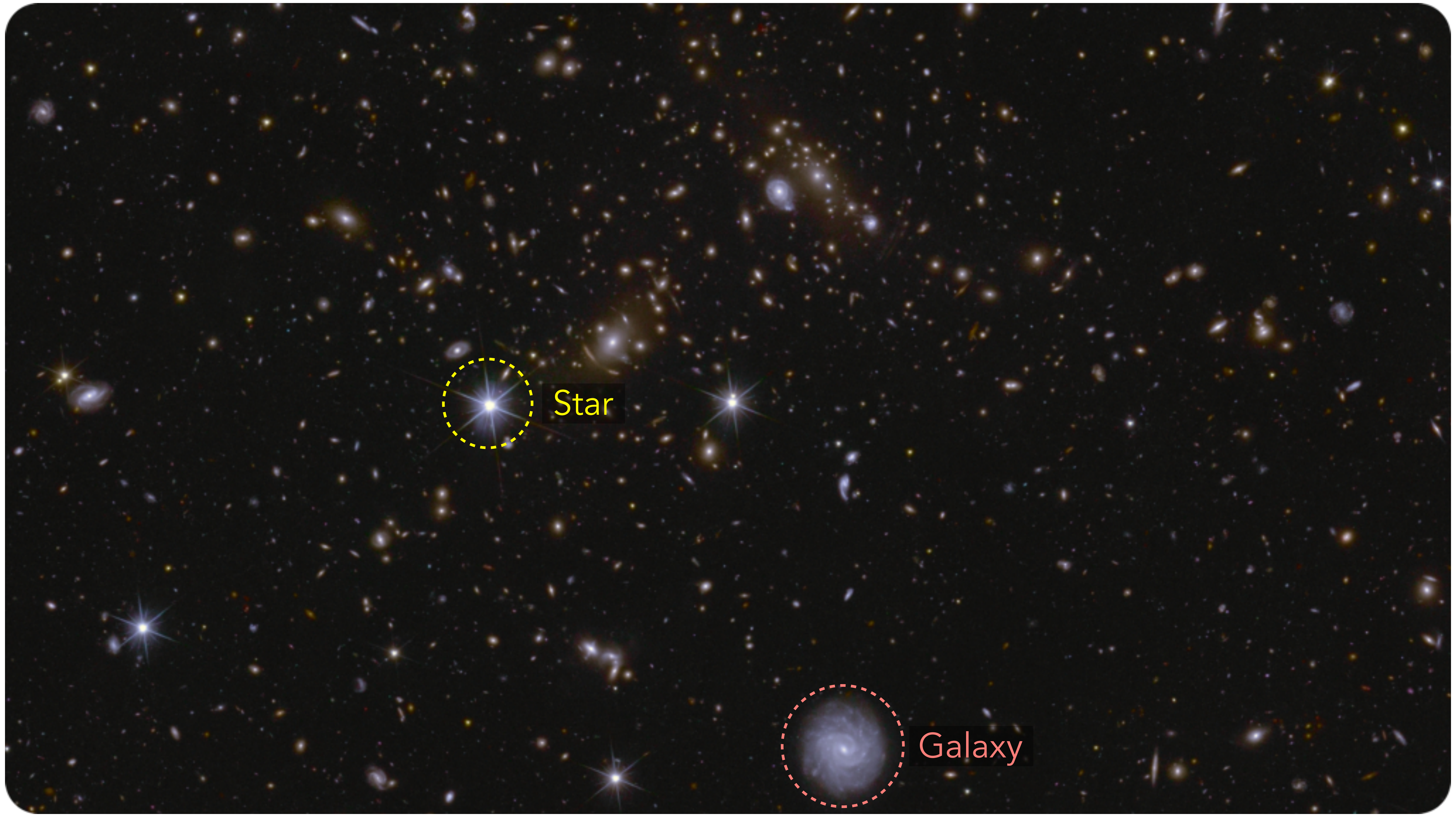
A brief tour of where things can go wrong

Sources of bias
- Detection
- Blending
- Masking

Credit: ESA/Euclid/Euclid Consortium/NASA, image processing by J.-C. Cuillandre, E. Bertin, G. Anselmi

Weak Lensing

A brief tour of where things can go wrong

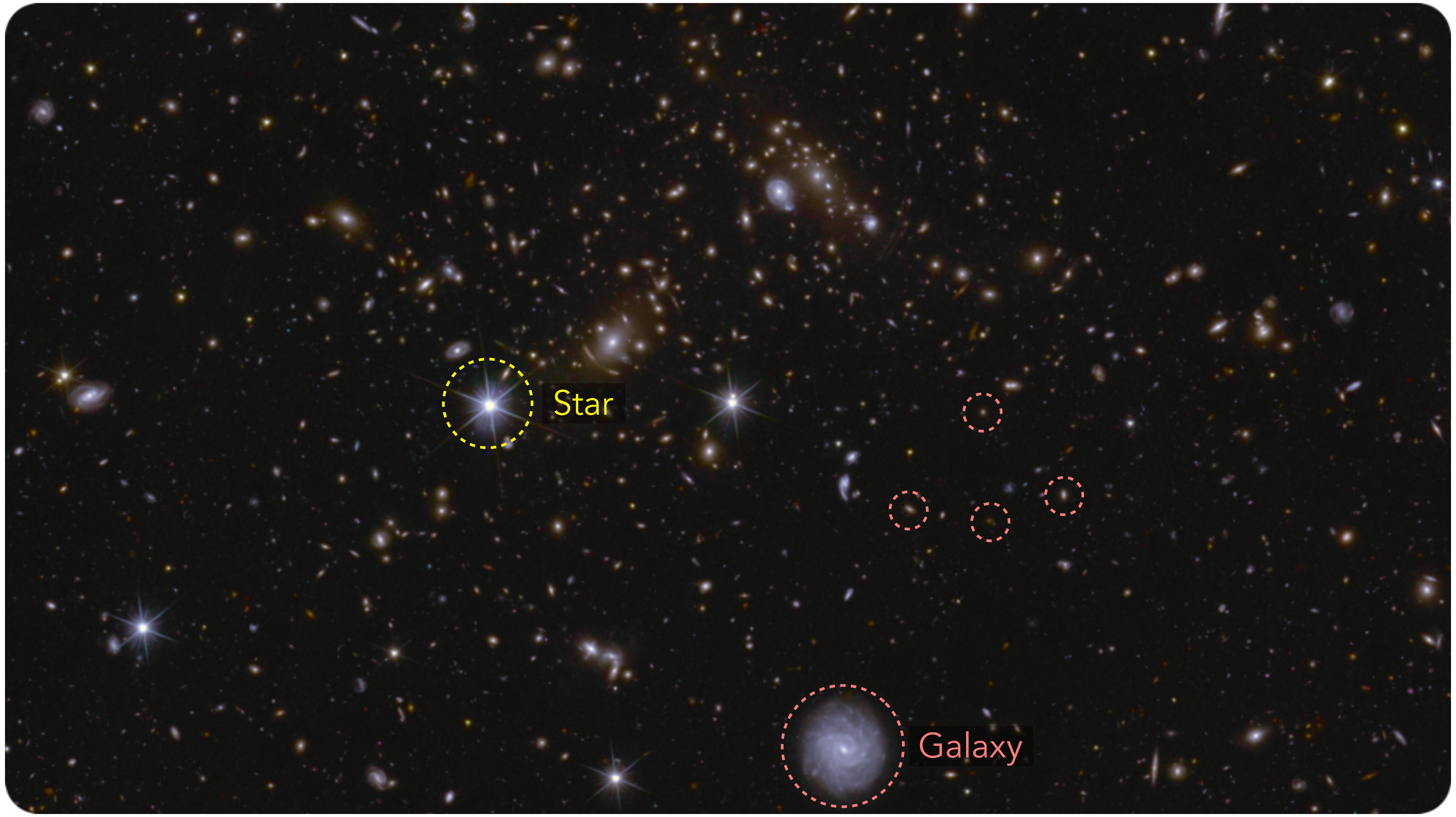


Credit: ESA/Euclid/Euclid Consortium/NASA, image processing by J.-C. Cuillandre, E. Bertin, G. Anselmi

Sources of bias
- Detection
- Blending
- Masking

Weak Lensing

A brief tour of where things can go wrong

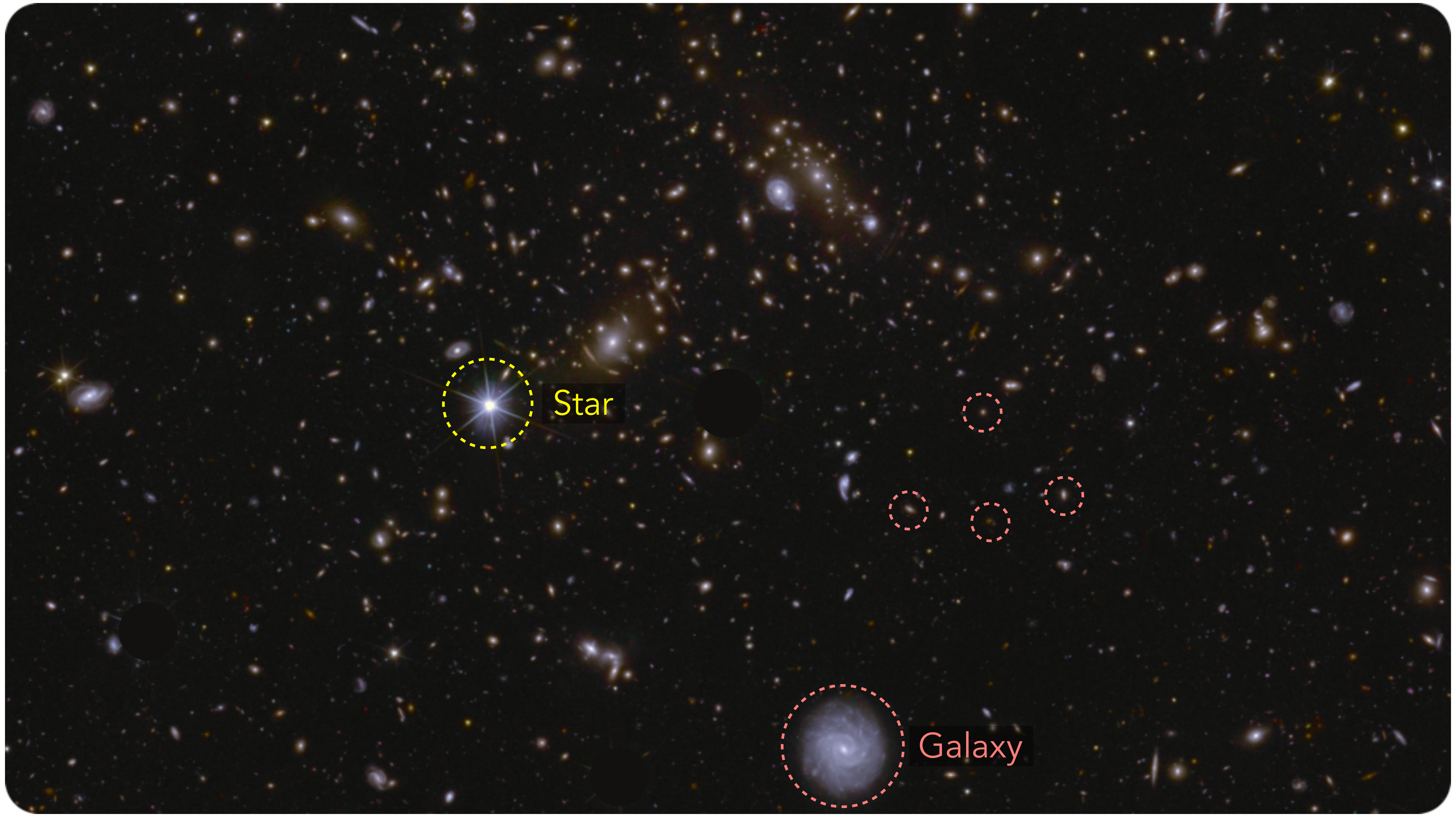


Sources of bias
- Detection
- Blending
- Masking

Credit: ESA/Euclid/Euclid Consortium/NASA, image processing by J.-C. Cuillandre, E. Bertin, G. Anselmi

Weak Lensing

A brief tour of where things can go wrong



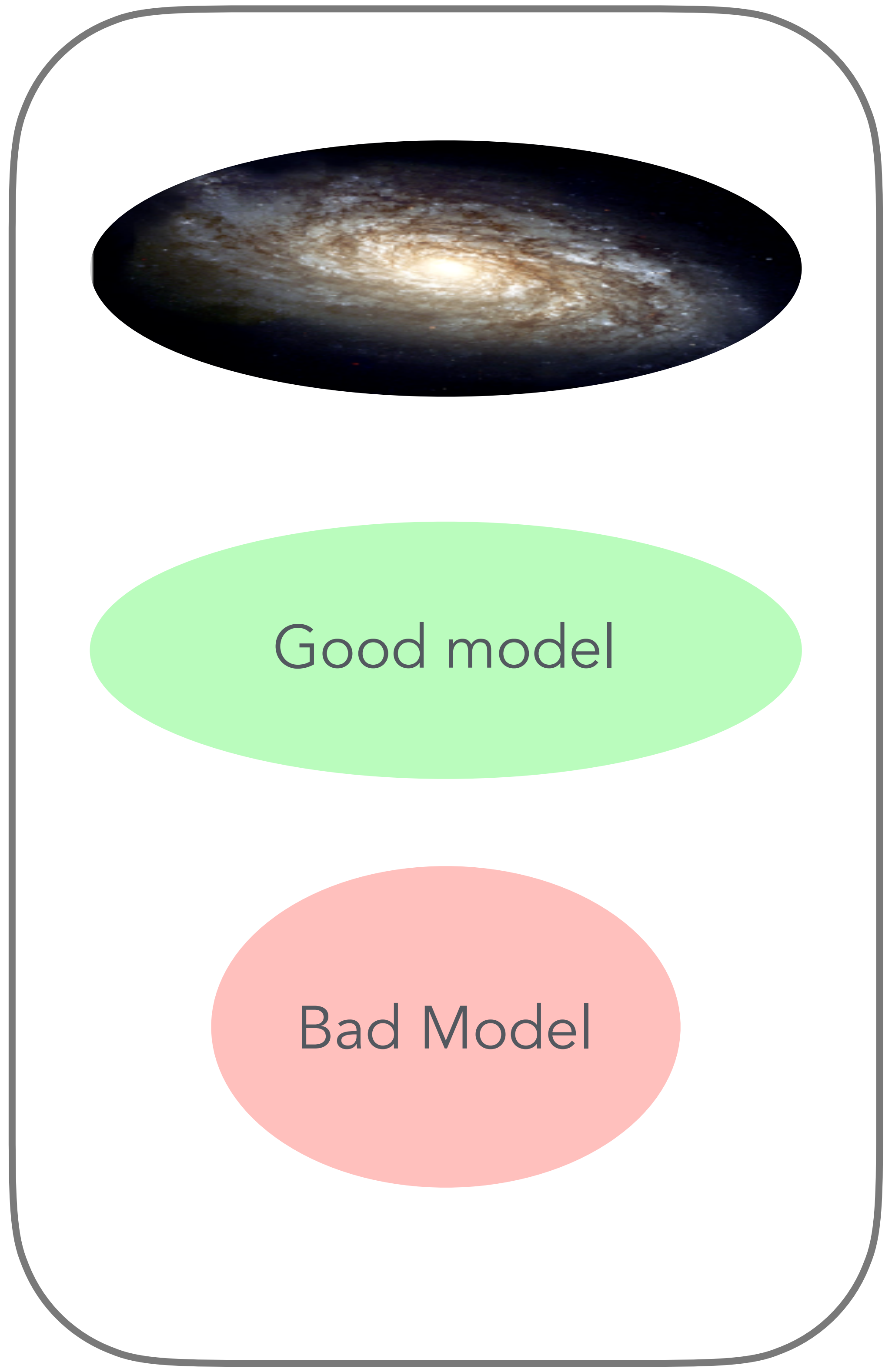
Credit: ESA/Euclid/Euclid Consortium/NASA, image processing by J.-C. Cuillandre, E. Bertin, G. Anselmi

Sources of bias
- Detection
- Blending
- Masking

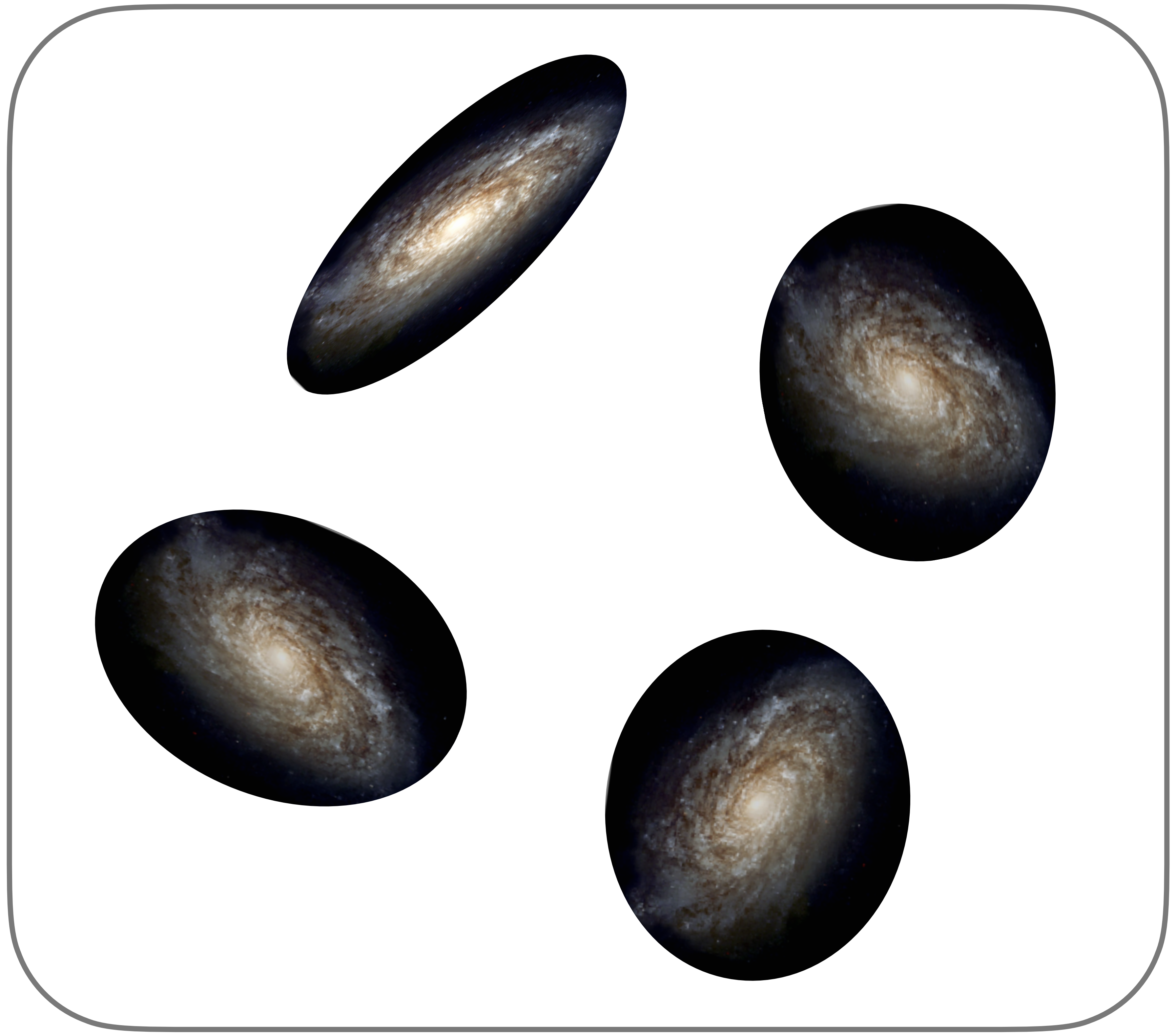
Weak Lensing

A brief tour of where things can go wrong

Model Fitting



Source Selection



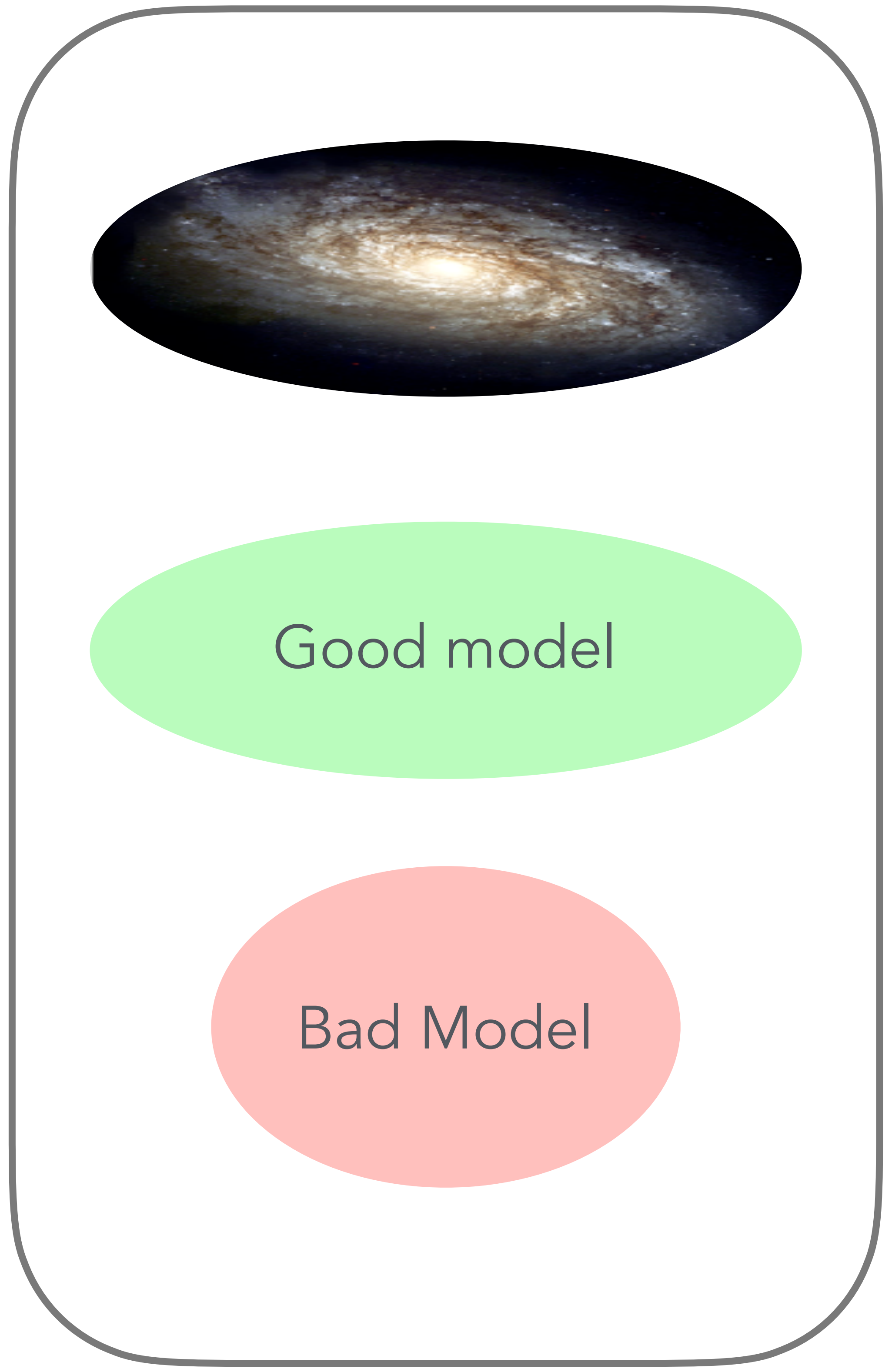
Sources of bias

- Model
- Selection

Weak Lensing

A brief tour of where things can go wrong

Model Fitting



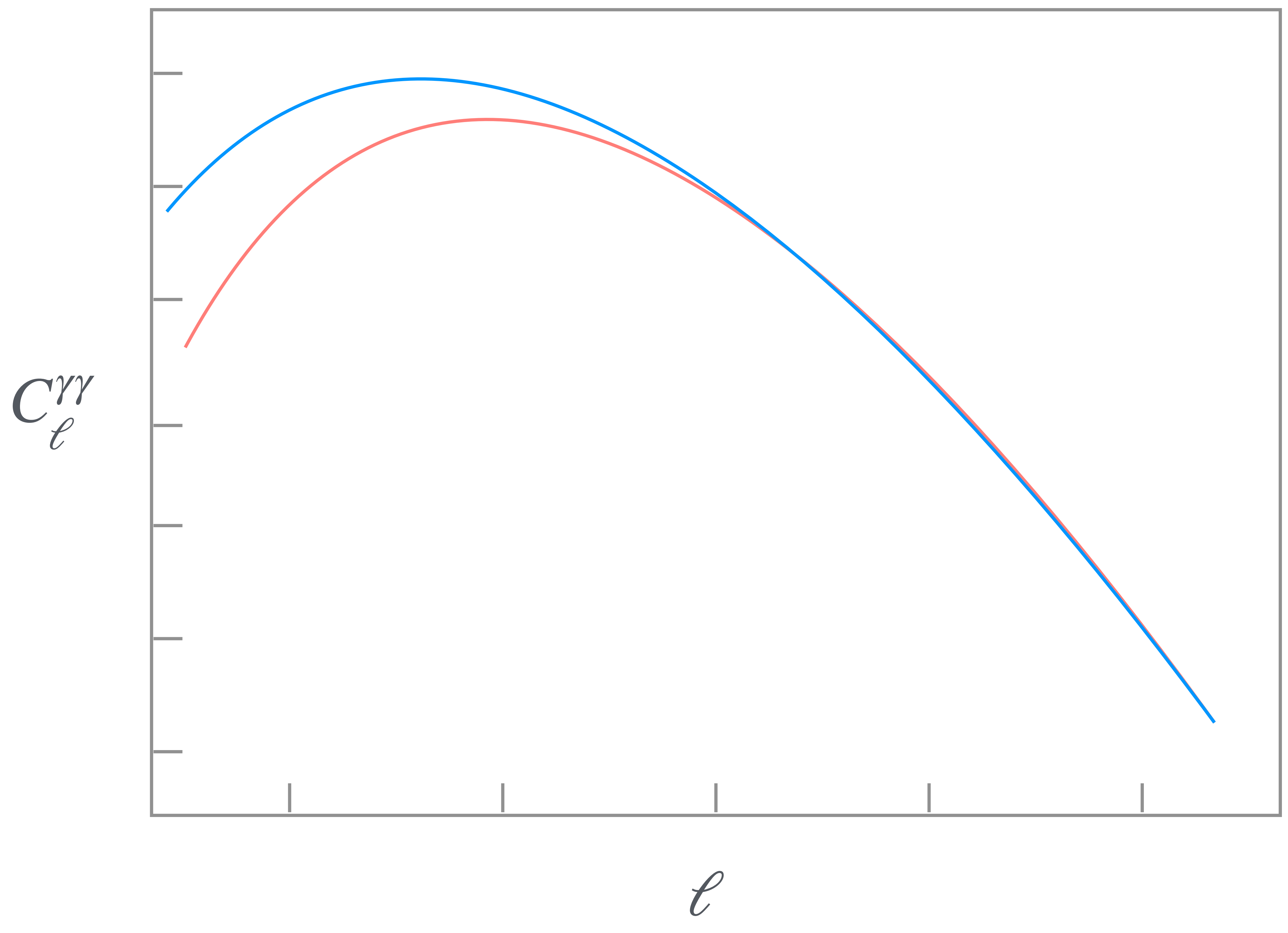
Source Selection

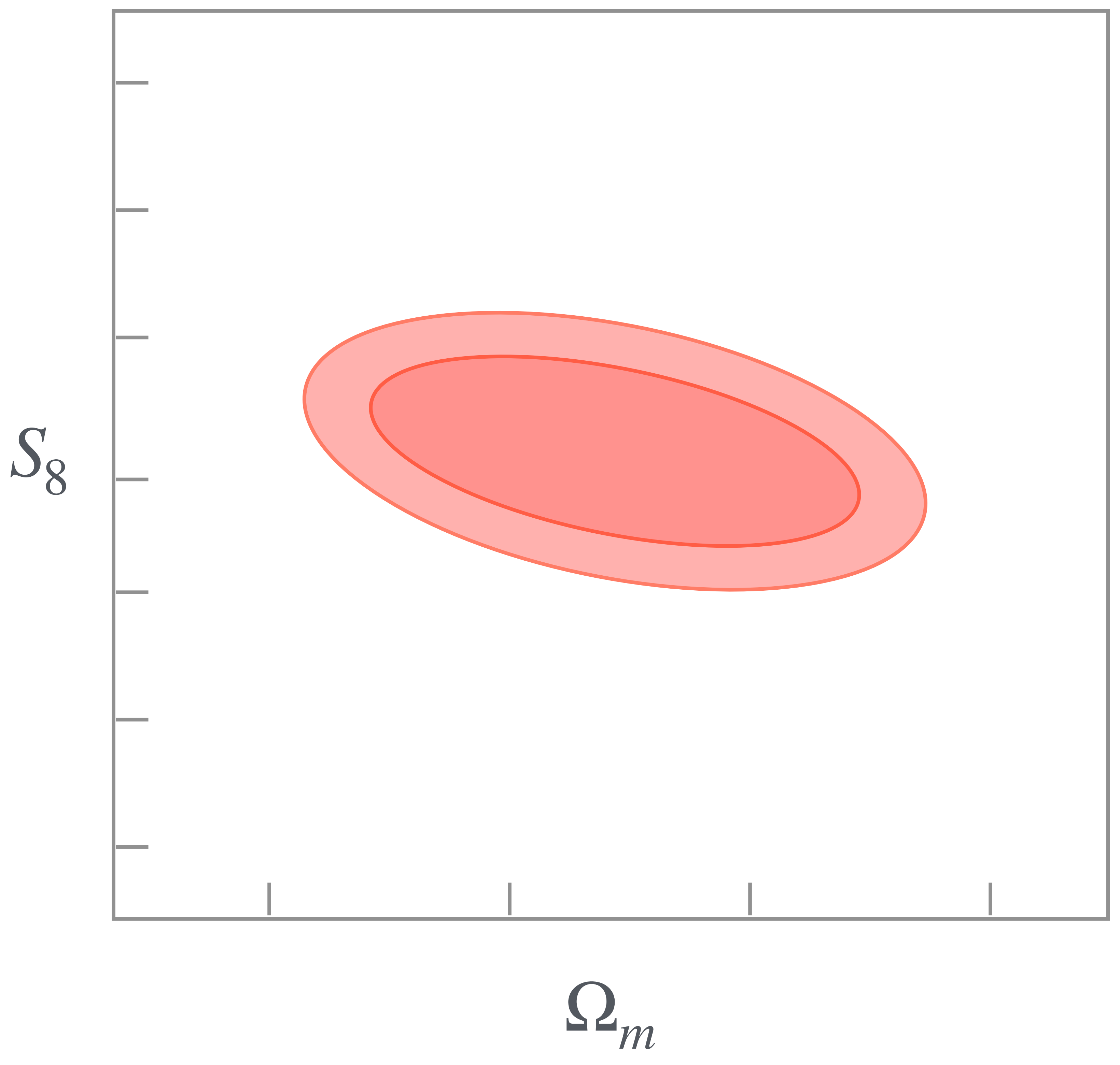
Sources of bias
- Model
- Selection

Weak Lensing

A brief tour of where things can go wrong

Summary Statistics





Sources of bias

- Theoretical modelling
- Parameter inference

Outline

1. Cosmological motivation

- What are the **problems** we need to solve?
- How does **data** play a role in this?
- Why do we need **AI** and how can we trust the results?

2. Weak gravitational lensing

- 'Observing' **invisible matter**
- Technical **challenges**

3. AI for cosmological data analysis

- Estimating **stellar SEDs** from photometric data
- Identify of **blended galaxies** in survey data
- Going beyond

4. Future perspectives

- Where can AI take us?
- Accessible **tools**
- Conclusions

Outline

3. AI for cosmological data analysis

- Estimating **stellar SEDs** from photometric data
- Identify of **blended galaxies** in survey data
- Going beyond

SED Classification

E. Centofanti

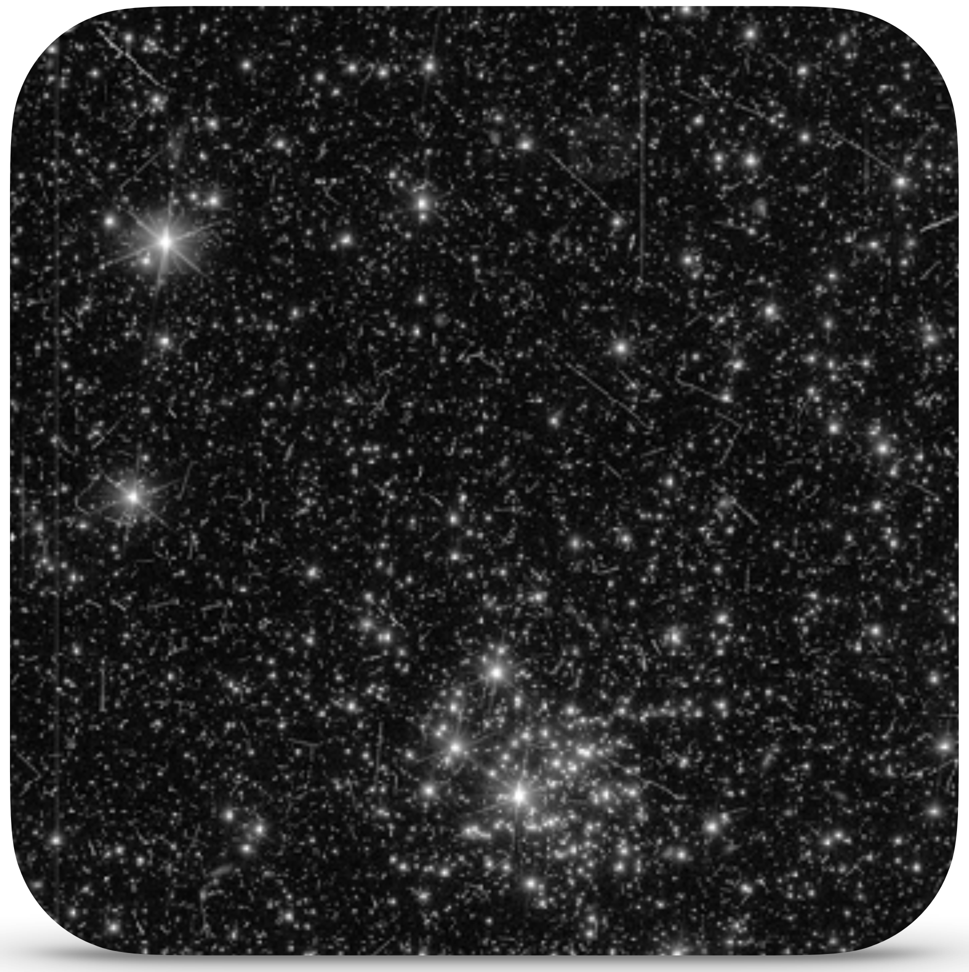
A. Szapiro

J. Pollack

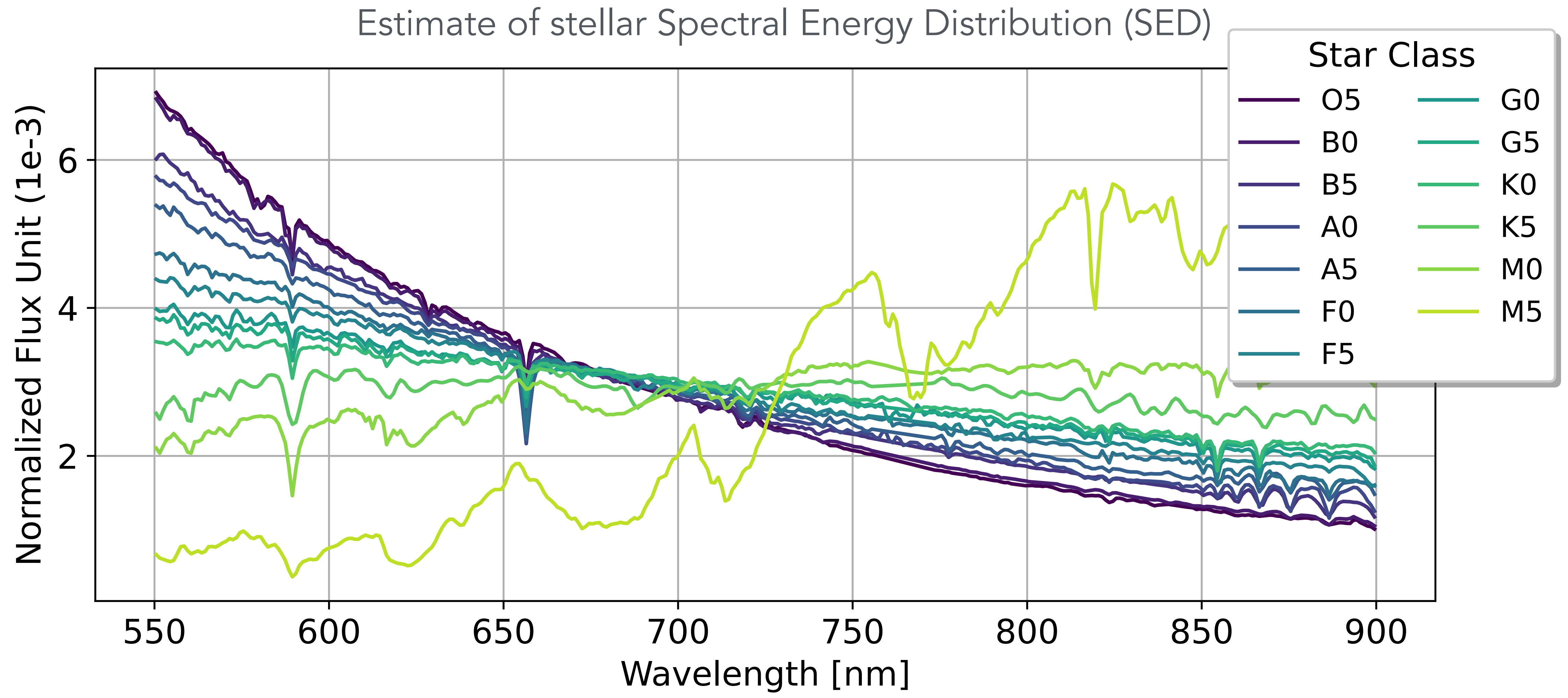
arXiv

Breaking the degeneracy in stellar spectral classification from single wide-band images

- **Ezequiel Centofanti**, **Samuel Farrens**, Jean-Luc Starck, Tobias Liaudat, **Alex Szapiro**, **Jennifer Pollack**
- Published in A&A in 2025 [arxiv.org/abs/2501.16151]



Euclid early commissioning VIS instrument image © ESA/Euclid/Euclid Consortium/NASA



Flux-normalised spectral templates for 13 stellar classes (Pickles 1998)

SED Classification

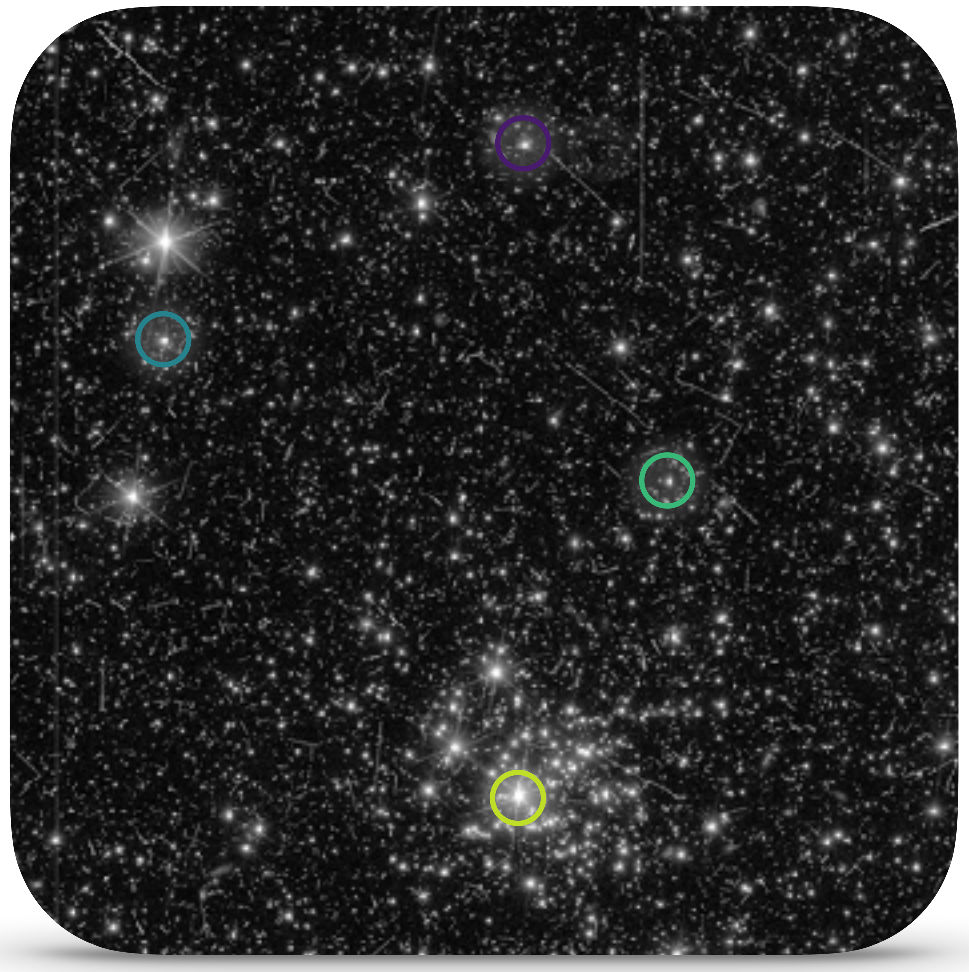
E. Centofanti

A. Szapiro

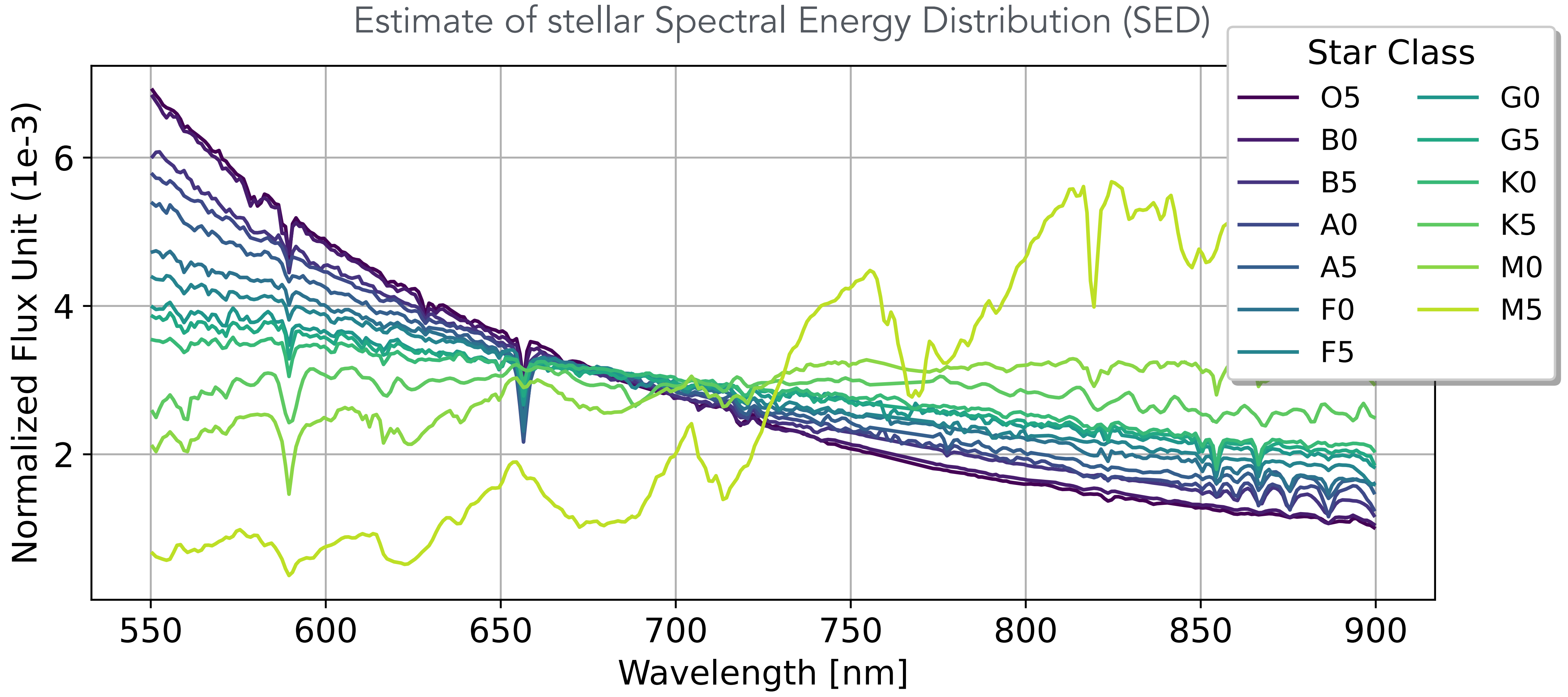
J. Pollack

Breaking the degeneracy in stellar spectral classification from single wide-band images

- **Ezequiel Centofanti**, **Samuel Farrens**, Jean-Luc Starck, Tobias Liaudat, **Alex Szapiro**, **Jennifer Pollack**
- Published in A&A in 2025 [arxiv.org/abs/2501.16151]



Euclid early commissioning VIS instrument image © ESA/Euclid/Euclid Consortium/NASA



Flux-normalised spectral templates for 13 stellar classes (Pickles 1998)

SED Classification

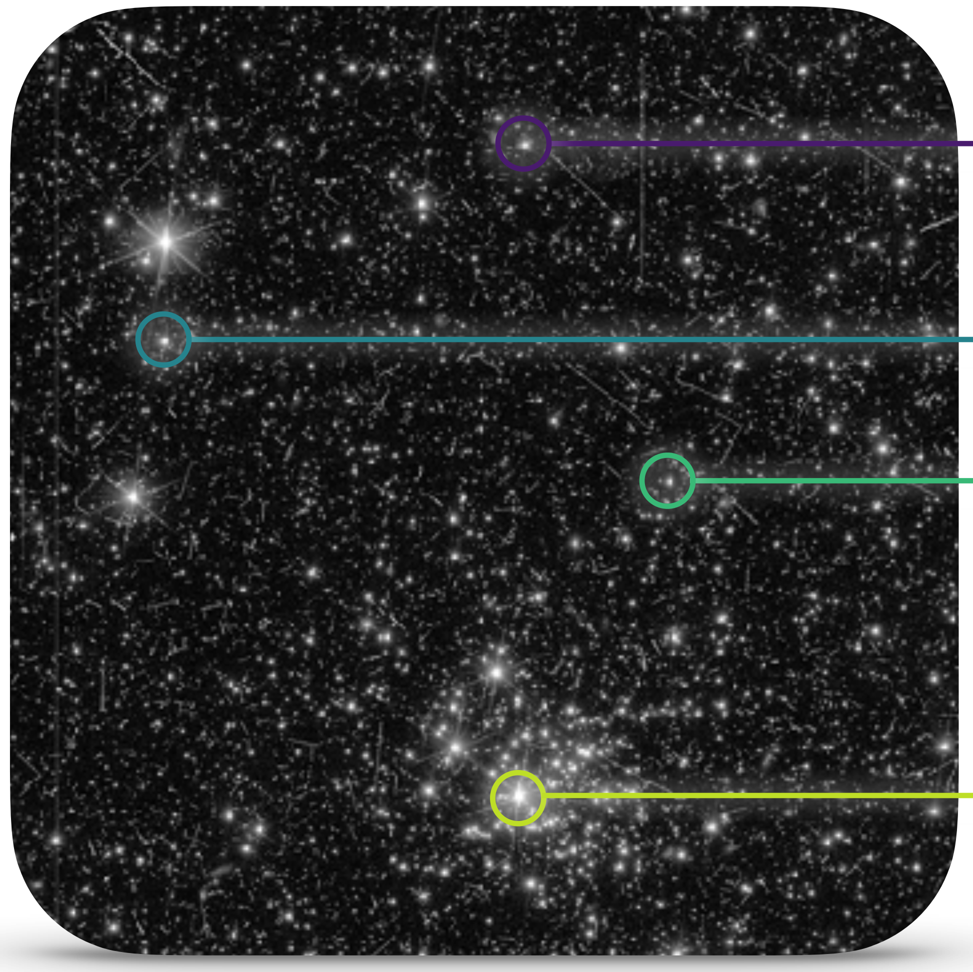
E. Centofanti

A. Szapiro

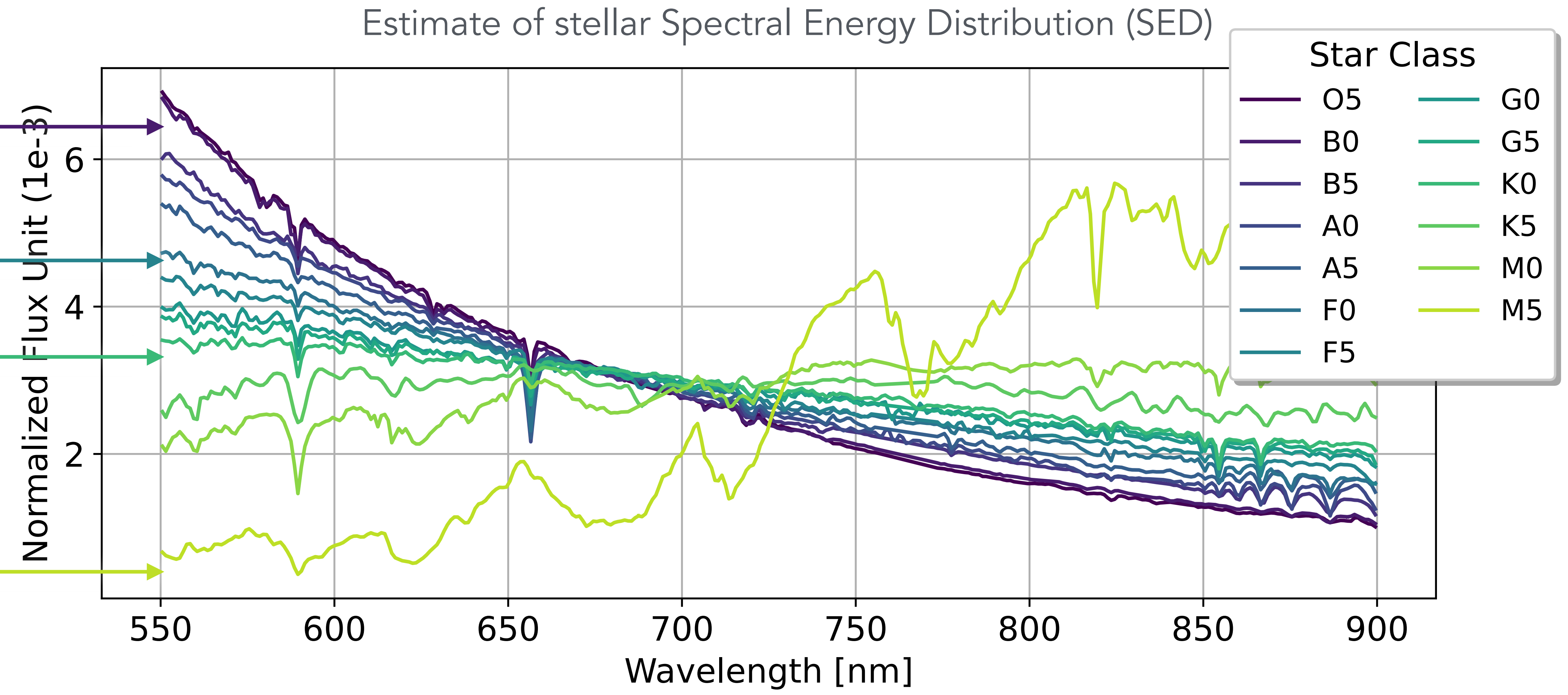
J. Pollack

Breaking the degeneracy in stellar spectral classification from single wide-band images

- Ezequiel Centofanti, **Samuel Farrens**, Jean-Luc Starck, Tobias Liaudat, **Alex Szapiro**, **Jennifer Pollack**
- Published in A&A in 2025 [arxiv.org/abs/2501.16151]



Euclid early commissioning VIS instrument image © ESA/Euclid/Euclid Consortium/NASA



Flux-normalised spectral templates for 13 stellar classes (Pickles 1998)

SED Classification

T. Liaudat

J. Pollack

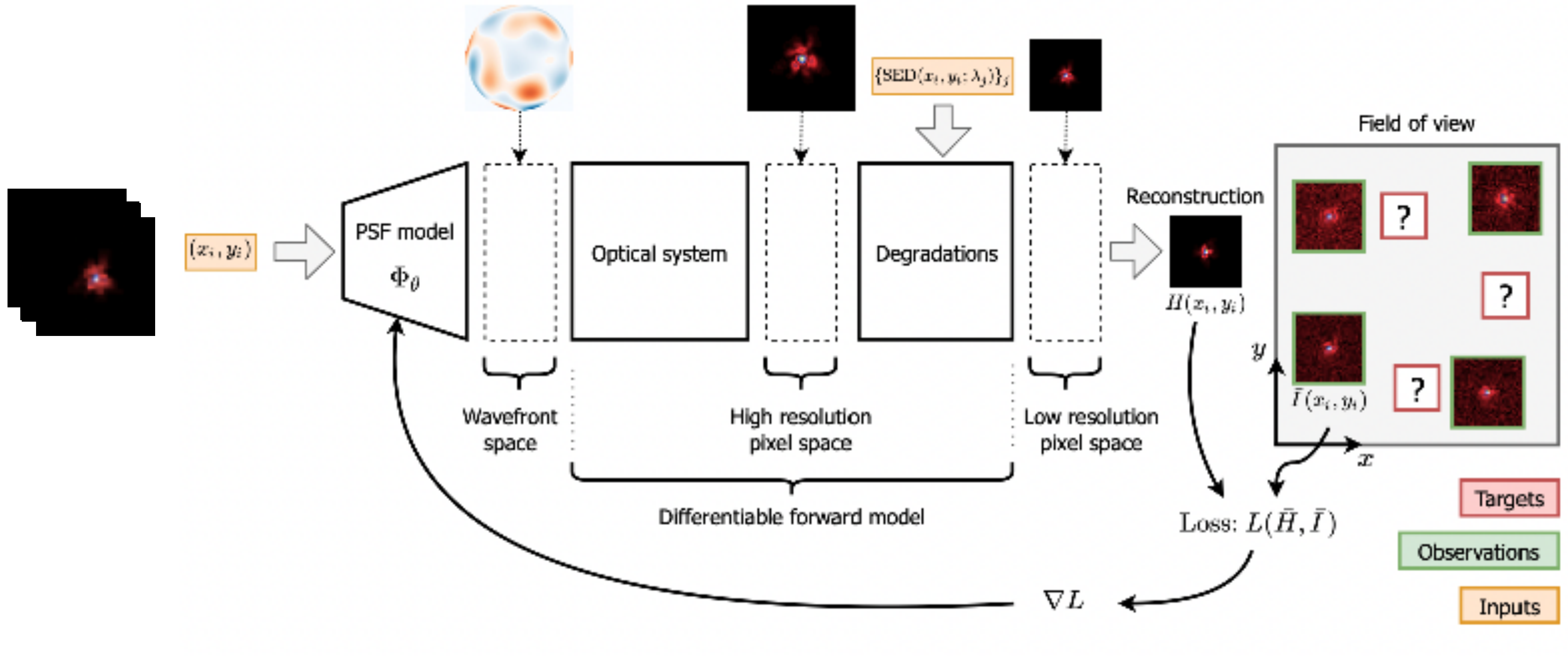
N. Moukaddem

E. Centofanti

Data-driven PSF modelling

- Accurate modelling of the **point spread function** (PSF) is critically important for galaxy shape measurement and hence weak-lensing analyses
- **Parametric modelling** of (in particular) space-based telescopes (like *Euclid*) is **very challenging**
- The data-driven wavefront-based **WaveDiff** software (**Liaudat** et al. 2023) **requires stellar SEDs** to model the PSF
- We **won't have spectra** for all the stars in the field
- Can we do better if we can estimate **SEDs from the star images**?

WaveDiff

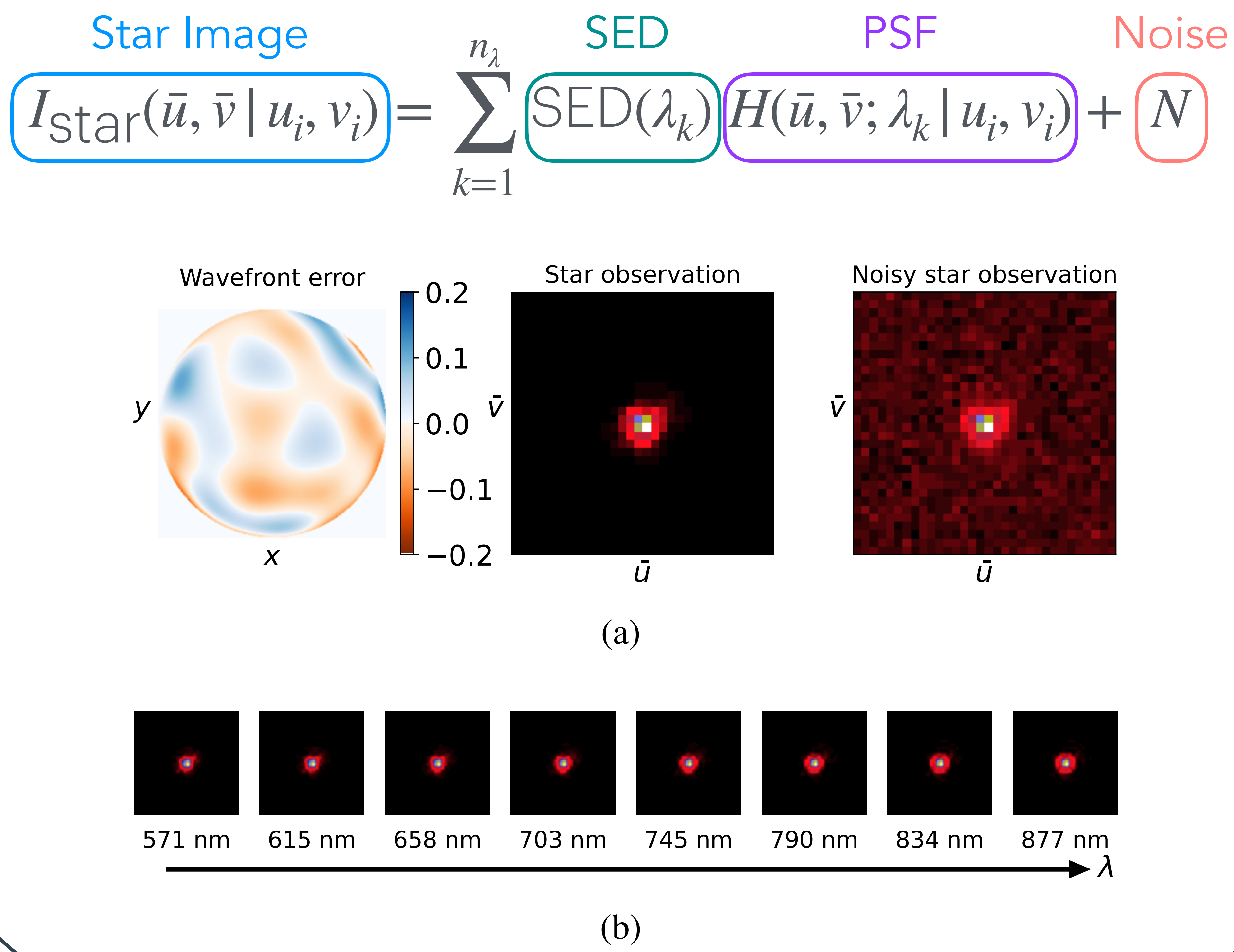


Liaudat et al. (2023) [arxiv.org/abs/2203.04908]

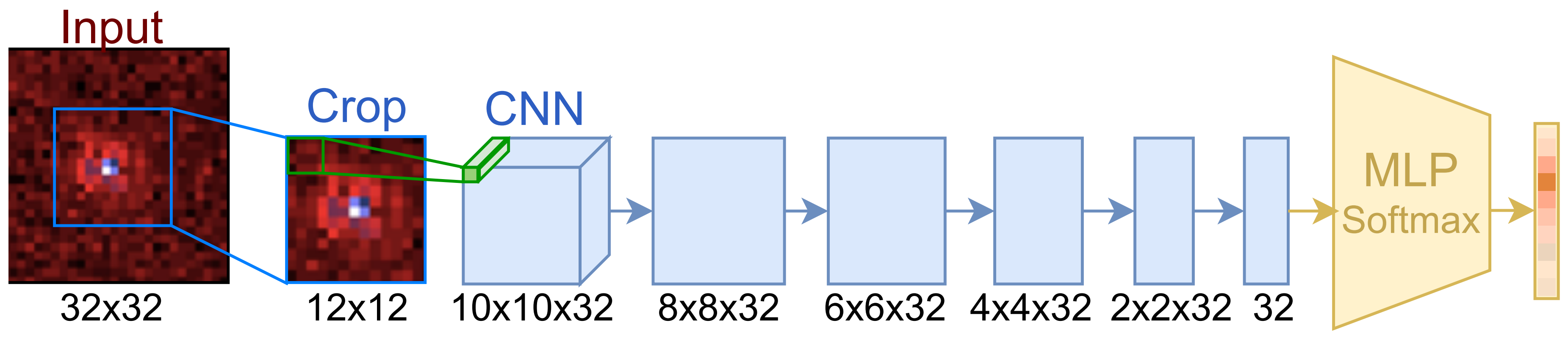
github.com/CosmoStat/wf-psf

SED Classification

WaveDiff Star Image Simulator



Convolutional Neural Network (CNN)



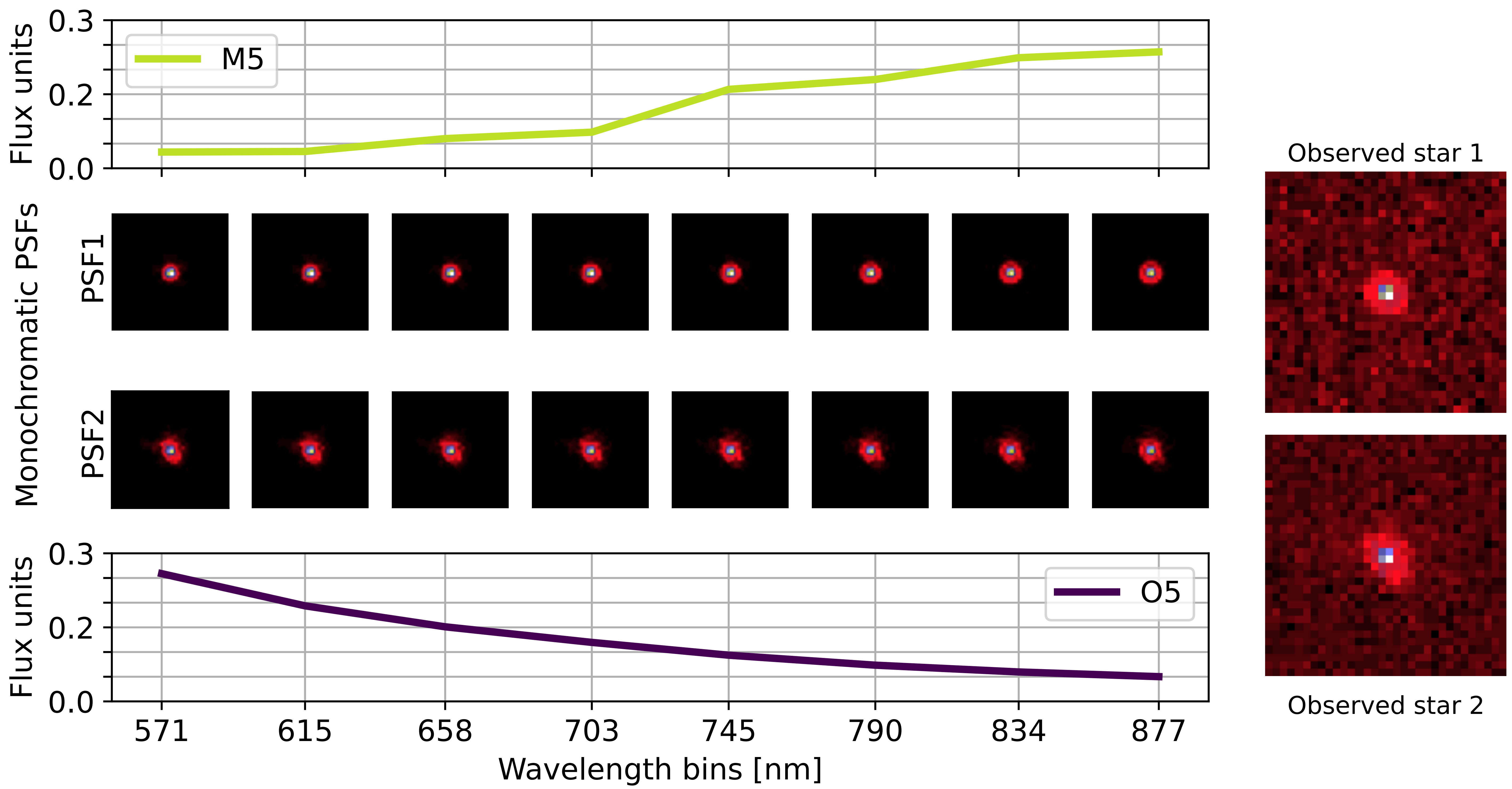
- Train with simulated star images with known SEDs
- Compare performance with PCA + CNN results presented in Kuntzer et al. (2016)
- Results were very consistent, but both approaches run into an **inherent limitation**

SED Classification

E. Centofanti (PhD)

Degeneracy between the PSF and the stellar SED

- It is difficult to disentangle the PSF size from the contribution of the SED in the observer stars
- The CNN (or any other) will struggle to learn appropriate labels for the images



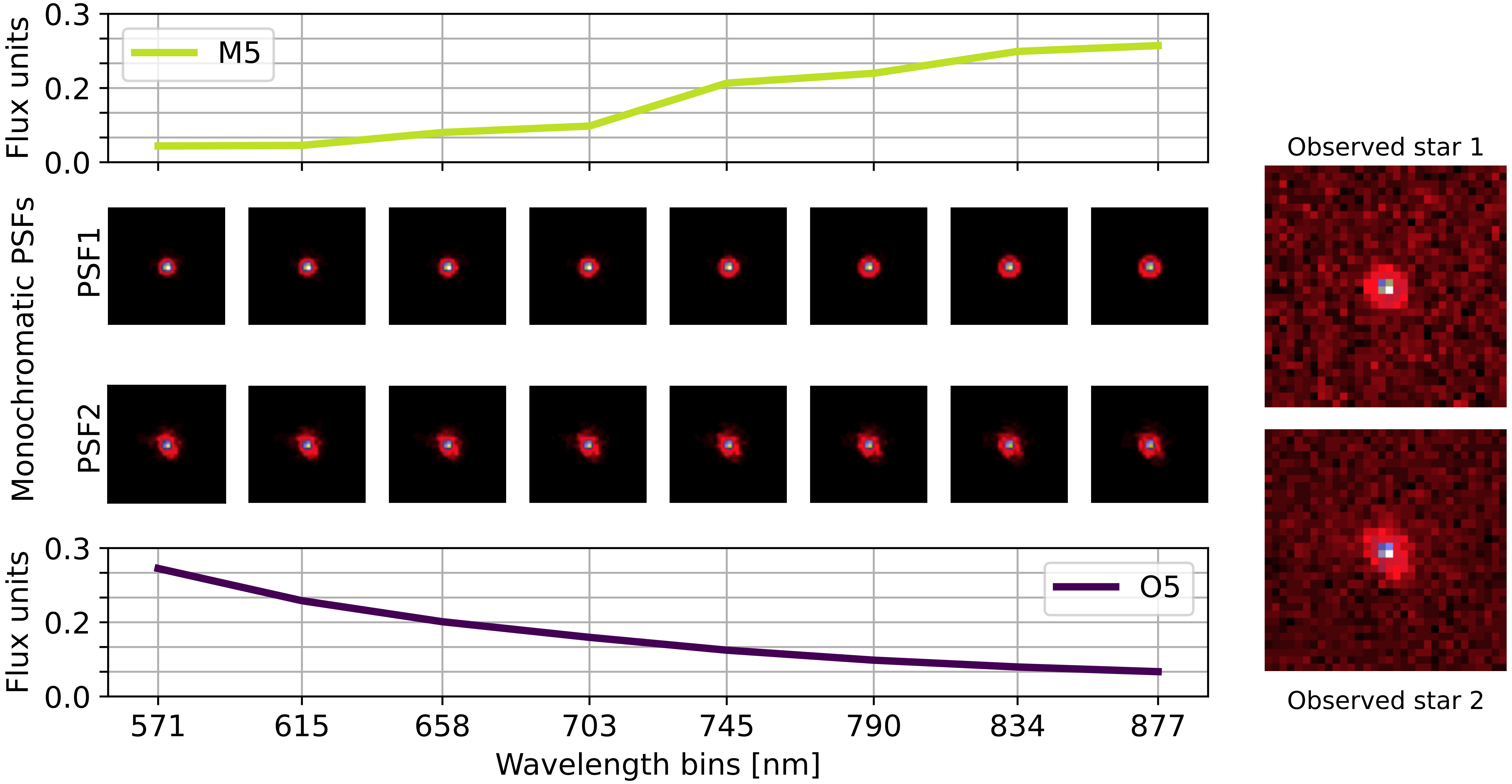
Monochromatic PSFs for two positions in the FOV and 8-bins spectral energy distribution of two stars, a *M5* star (top) and an *O5* star (bottom)

SED Classification

E. Centofanti (PhD)

Degeneracy between the PSF and the stellar SED

- It is difficult to disentangle the PSF size from the contribution of the SED in the observer stars
- The CNN (or any other) will struggle to learn appropriate labels for the images



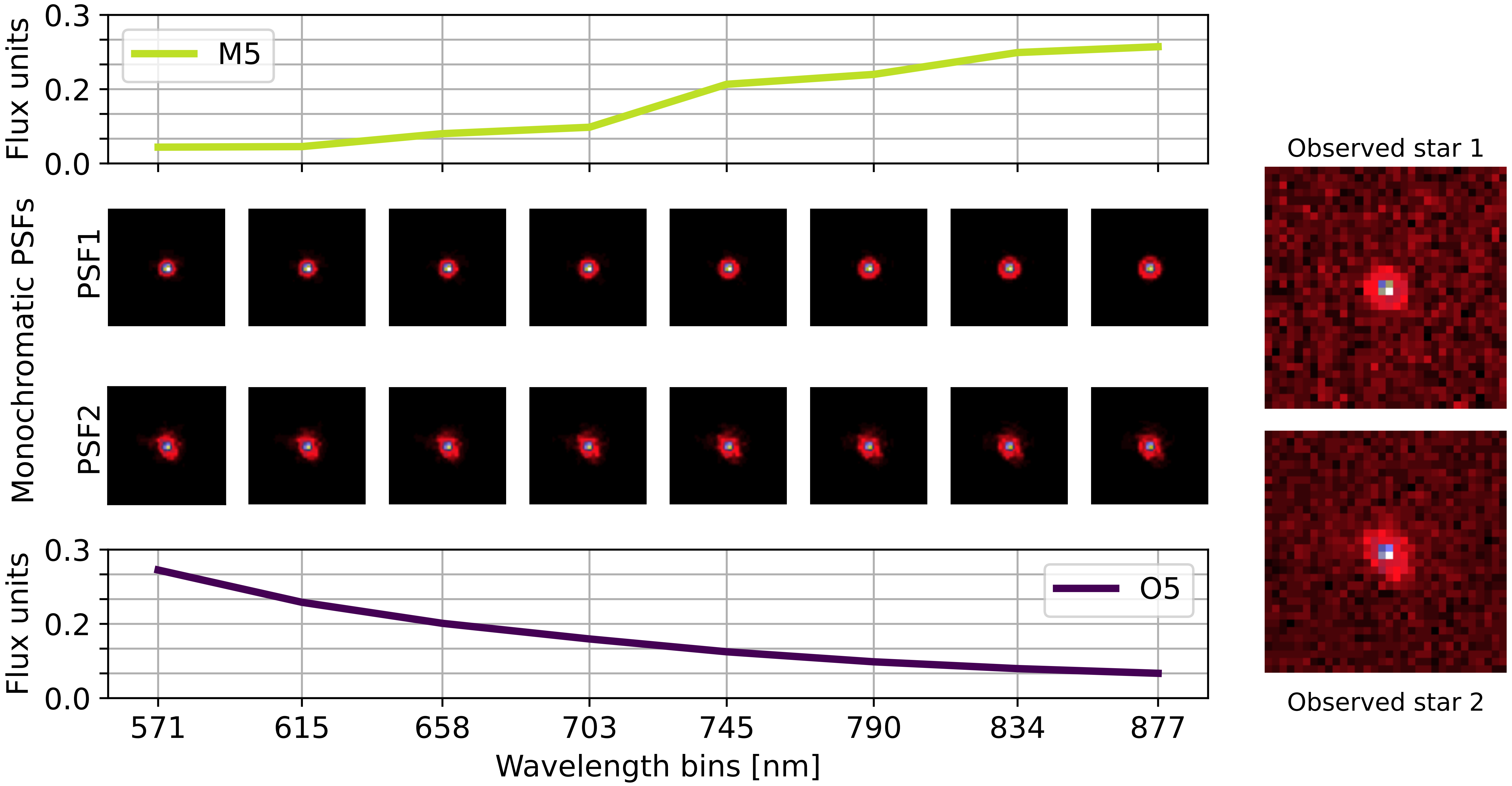
Monochromatic PSFs for two positions in the FOV and 8-bins spectral energy distribution of two stars, a *M5* star (top) and an *O5* star (bottom)

SED Classification

E. Centofanti (PhD)

Degeneracy between the PSF and the stellar SED

- It is difficult to disentangle the PSF size from the contribution of the SED in the observer stars
- The CNN (or any other) will struggle to learn appropriate labels for the images



Monochromatic PSFs for two positions in the FOV and 8-bins spectral energy distribution of two stars, a *M5* star (top) and an *O5* star (bottom)

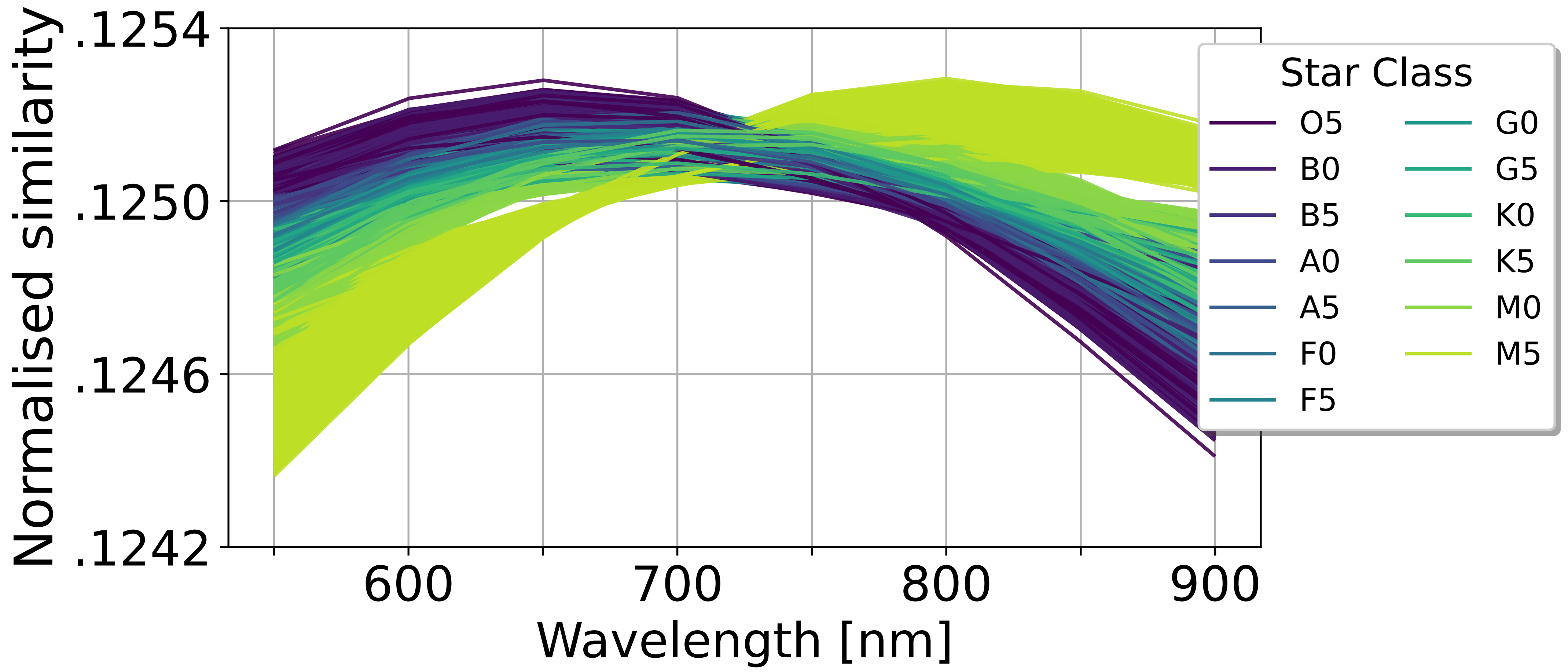
SED Classification

E. Centofanti (PhD)

Breaking the degeneracy

- Use the stars with measured spectra (and hence SEDs) to obtain an **approximate PSF model**
- Use this approximate PSF model to measure **similarity features**
- Train a support vector machine (**SVM**) **classifier** on the similarity features

$$\underbrace{\mathrm{SF}\left\langle I_{\mathrm{star}}(\bar{u},\bar{v}\,|\,u_i,v_i)\tilde{H}(\bar{u},\bar{v};\lambda_k\,|\,u_i,v_i)\right\rangle(\lambda_k\,|\,u_i,v_i)}_{\text{Similarity Features}} = \frac{1-\|\overbrace{I_{\mathrm{star}}(\bar{u},\bar{v}\,|\,u_i,v_i)}^{\text{Star Image}}-\overbrace{\tilde{H}(\bar{u},\bar{v};\lambda_k\,|\,u_i,v_i)}^{\text{Approx PSF}}\|^2_{\bar{u}\bar{v}}}{n_\lambda-\sum_{j=1}^{n_\lambda}\|I_{\mathrm{star}}(\bar{u},\bar{v}\,|\,u_i,v_i)-\tilde{H}(\bar{u},\bar{v};\lambda_j\,|\,u_i,v_i)\|^2_{\bar{u}\bar{v}}}$$

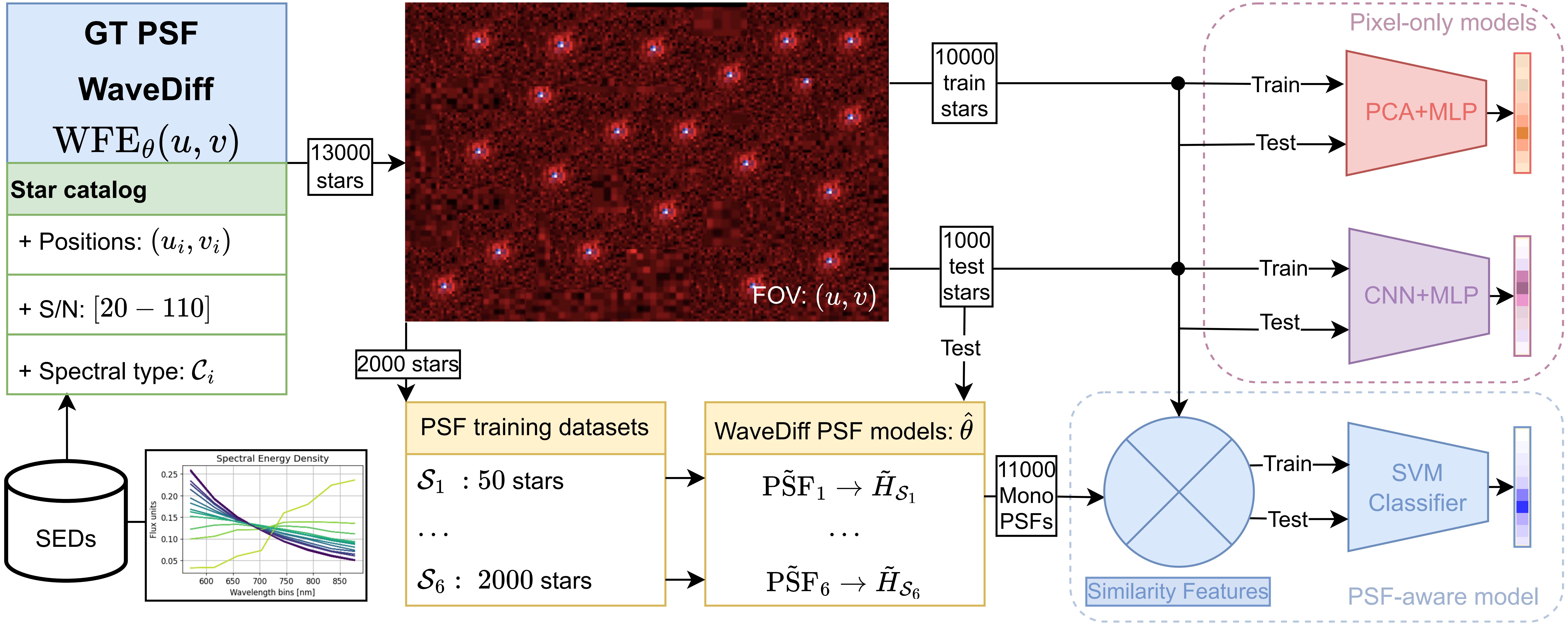


Normalised similarity features as a function of monochromatic PSF wavelength

SED Classification

E. Centofanti (PhD)

Experimental Setup

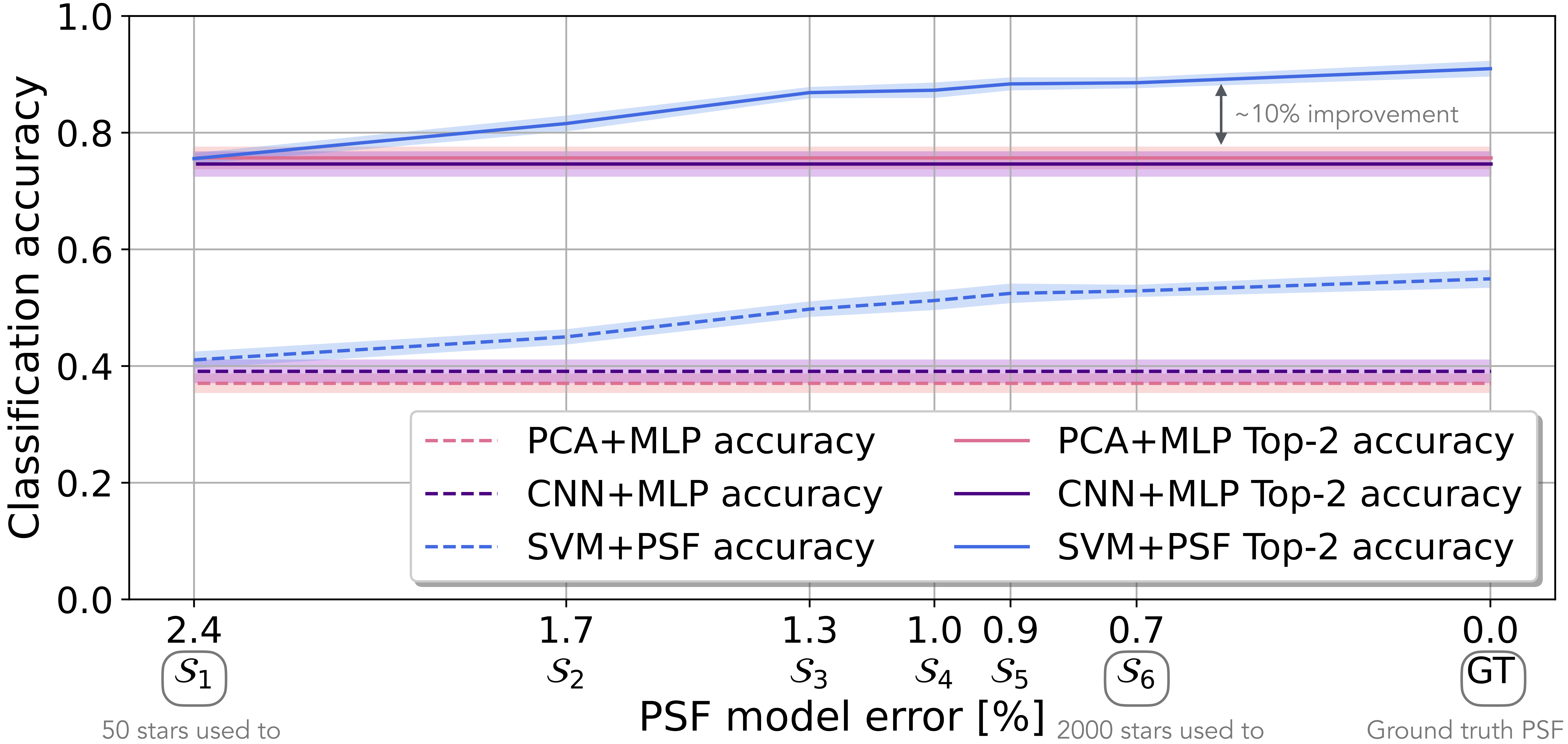


- GT = Ground Truth (i.e. the PSF model used to simulate the star images)
- $\mathcal{S}_n$ are nested subsamples of the total 2000 stars available for training the approximate PSF model

SED Classification

E. Centofanti (PhD)

Results



Estimated SED is within half a stellar class from the simulation

Estimated SED exactly matches the simulation

50 stars used to train the approx PSF model

2000 stars used to train the approx PSF model

Ground truth PSF model used to obtain the similarity features

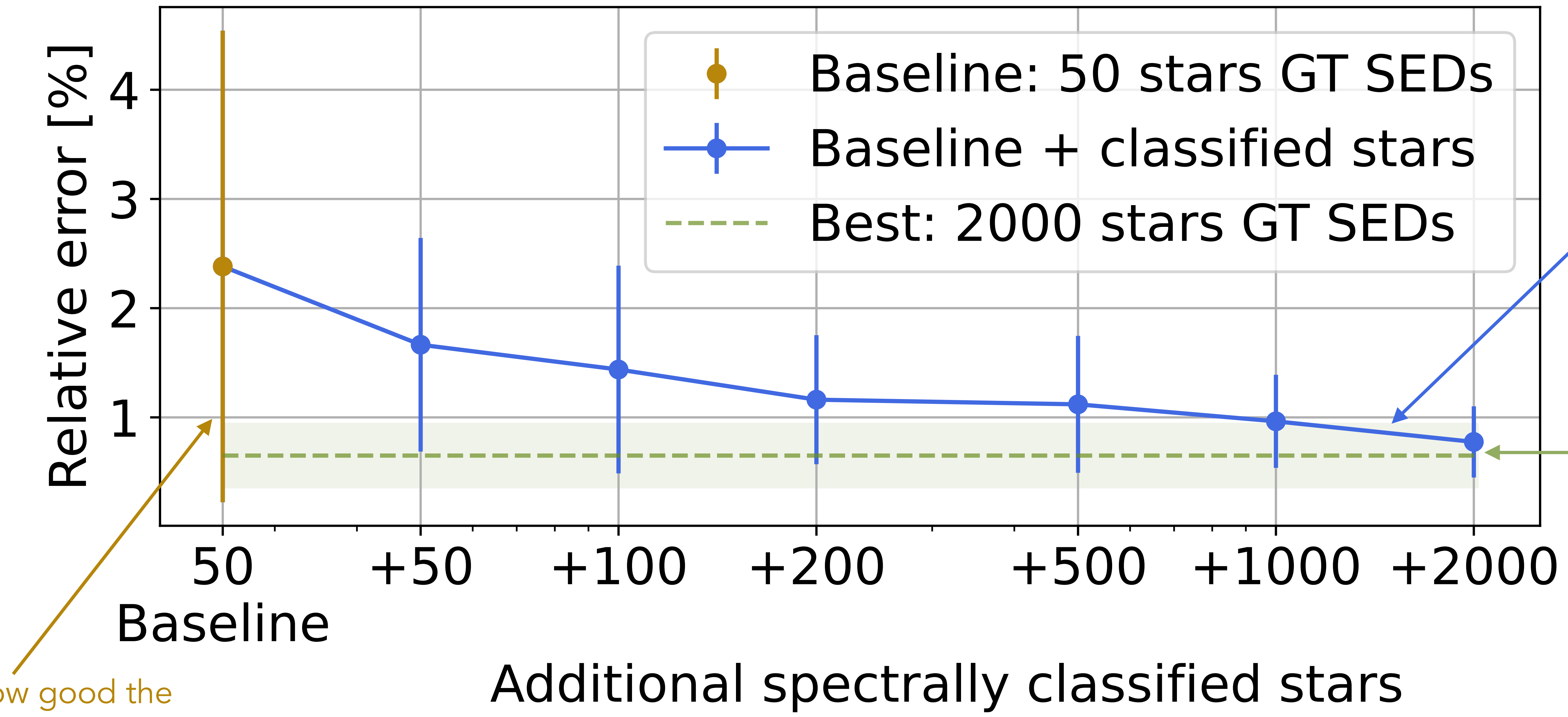
- The PSF-aware approach consistently outperforms pixel-only classification methods

SED Classification

E. Centofanti (PhD)

Results

Quality of the WaveDiff PSF model with respect to the ground truth



This is how much the WaveDiff PSF improves if we estimate additional SEDs from the images

This is how good the WaveDiff PSF would be if we had 2000 measured SEDs

This is how good the WaveDiff PSF would be if we had 50 measured SEDs

- Initial results indicate that this PSF-aware approach to SED classification could be used to improve data-driven PSF models
- Future work → see if we can make this work with real survey data!

Blend Identification


A. Lacan


A. Zamorano Vitorelli

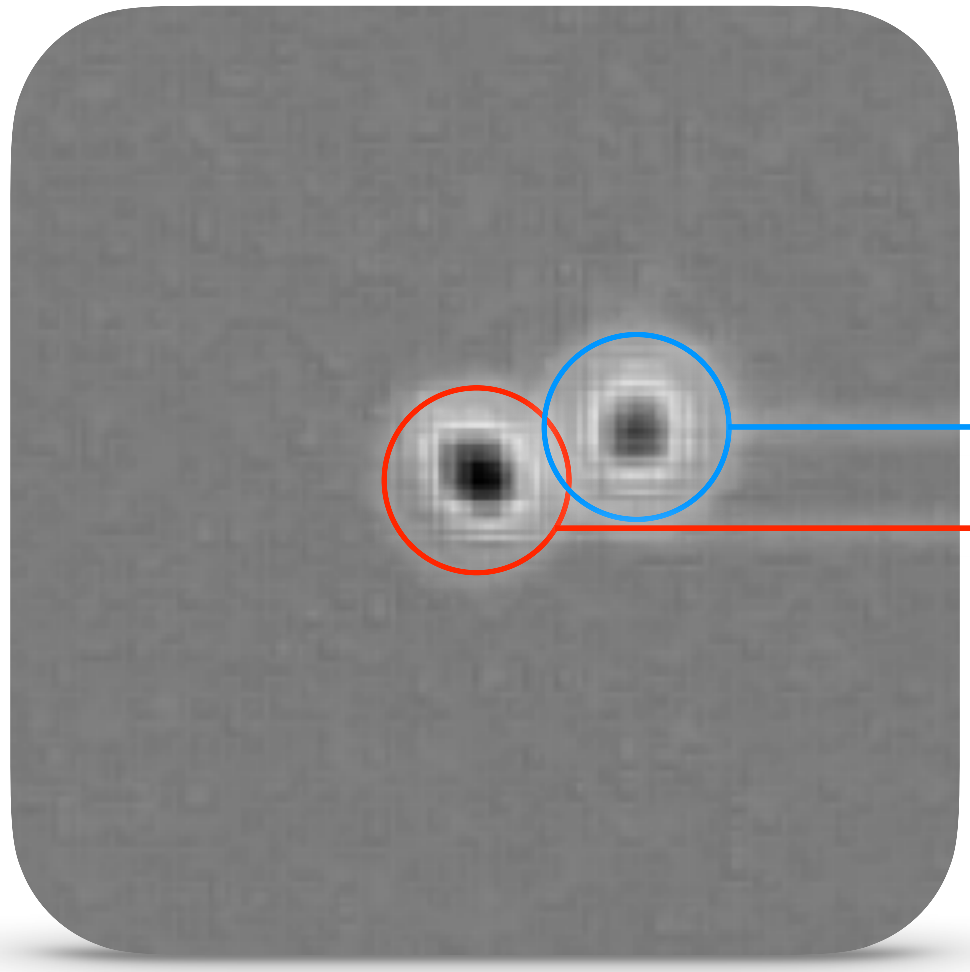

A. Bruckert

arXiv

Deep transfer learning for blended source identification in galaxy survey data

- **Samuel Farrens**, **Alice Lacan**, Axel Guinot, **Andre Zamorano Vitorelli**
- Published in A&A in 2022 [arxiv.org/abs/2110.08180]

github.com/CosmoStat/BlendHunter



Source 1
Source 2

Blend Identification

UNIONS

CFHT

Pan-STARRS

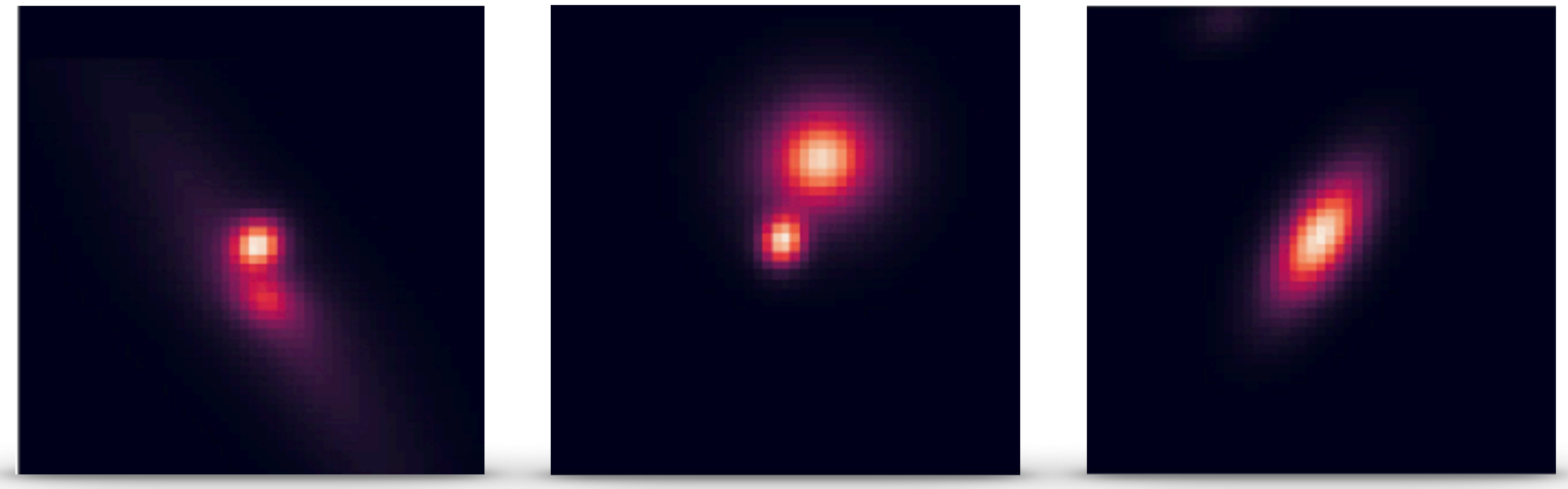
Subaru

AstroDeep

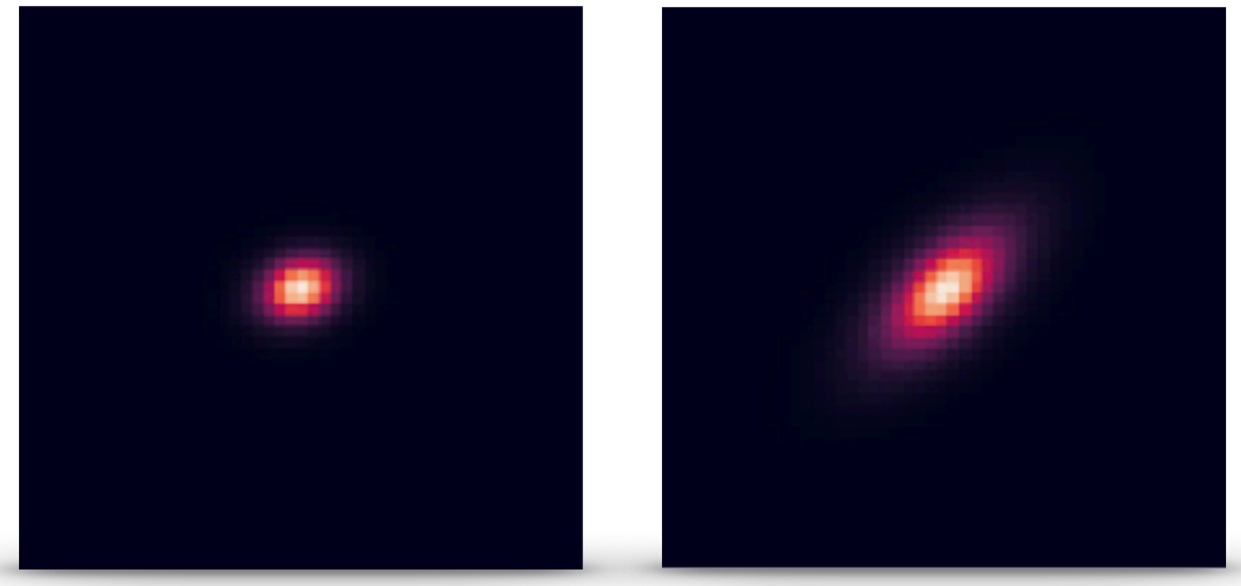
Blend Identification

The Challenge

Simulated blended galaxies



Simulated isolated galaxies

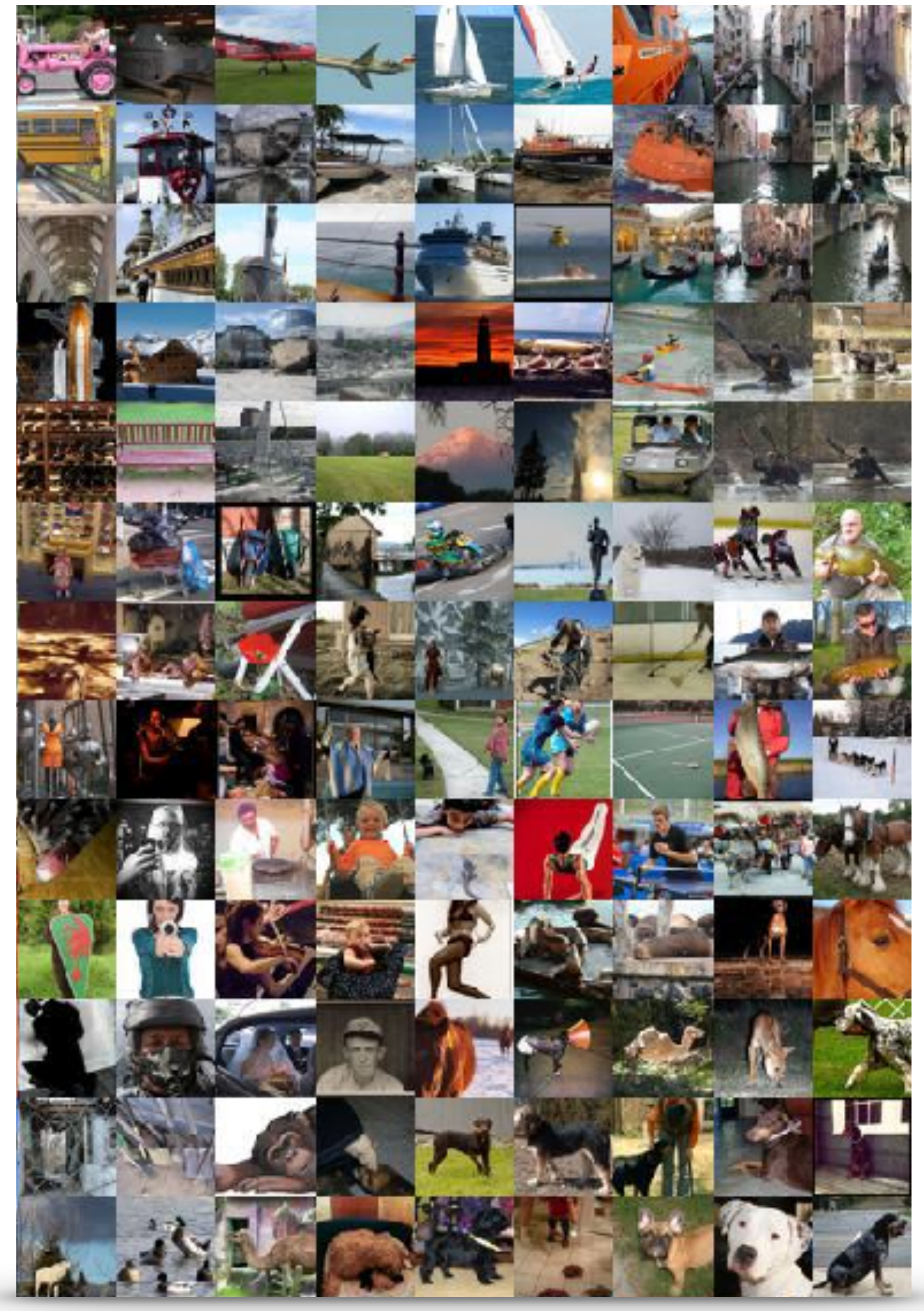


Simulating blended sources

- Assume a UNIONS-like survey where galaxies are selected from r-band images
- Need to have **complete control** of which sources are blended or not → we simulate galaxy postage stamps (51×51 pixels) with and without blends
- Need to define **what constitutes a blend** → we assumed any secondary object in the postage stamp
- Need a **benchmark for comparison** → we used Source Extractor (Bertin & Arnouts 1996)
- How do you **prevent overfitting** to the simulations?

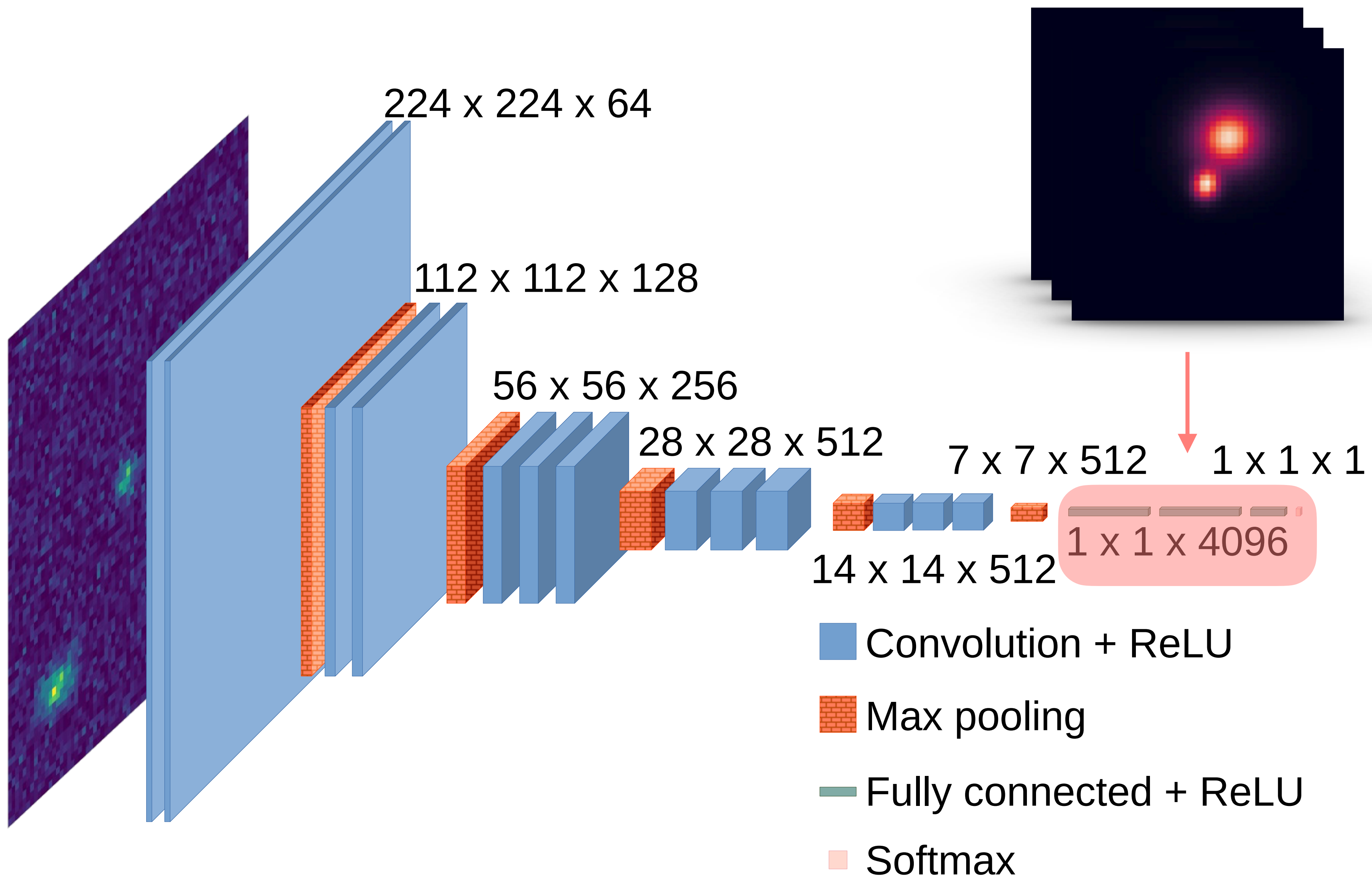
Blend Identification

IMAGENET



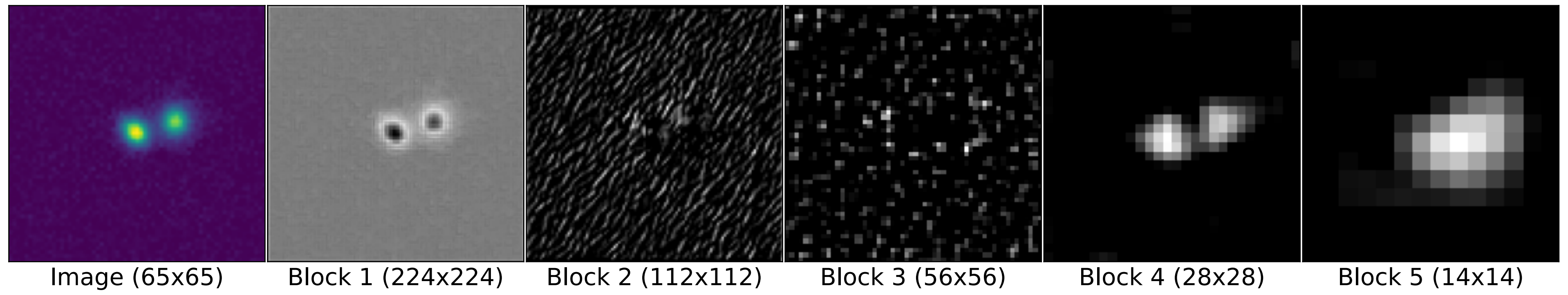
Convolutional layers pre-trained with natural images

Deep Transfer Learning



Fully connected layers trained with simulated galaxy images

Example features extracted by the VGG-16 network

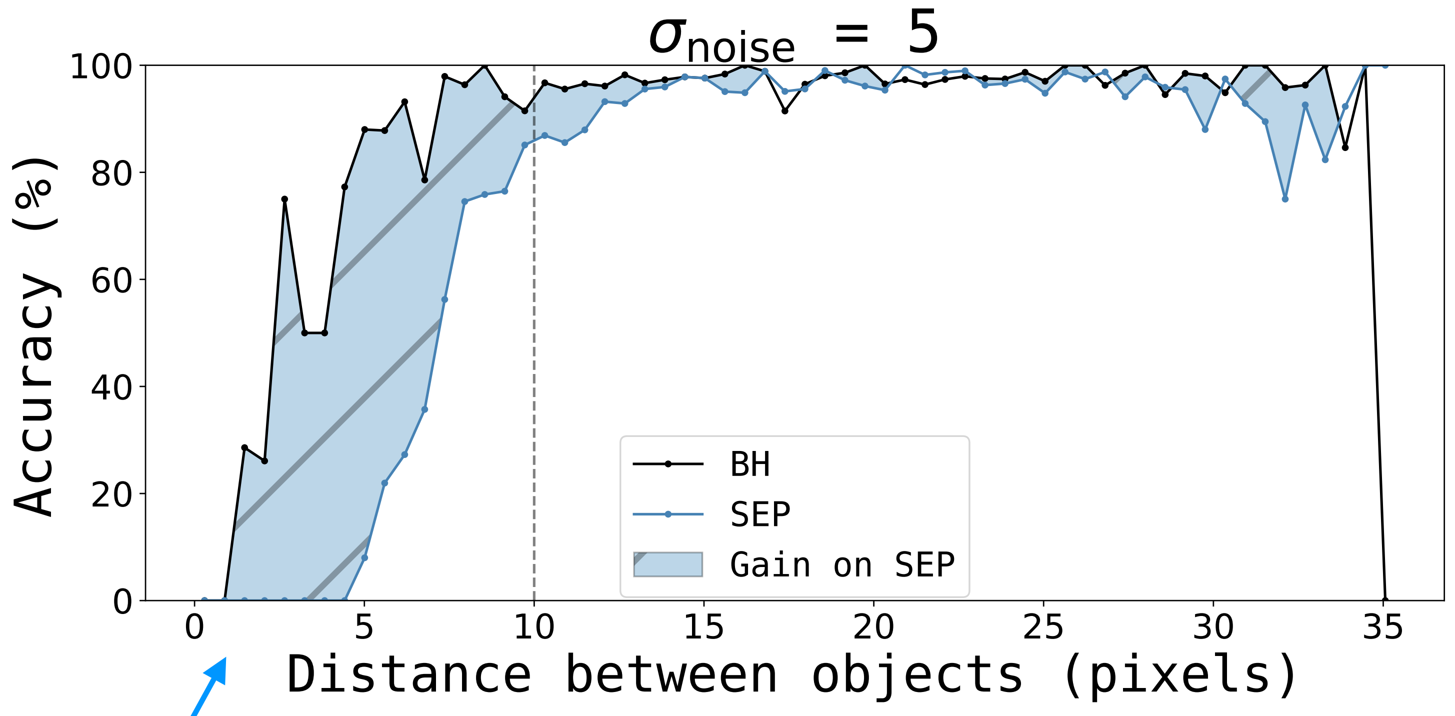


- Aim to avoid over-fitting by limiting which parts of the network can change
- The use of natural images avoids fitting to simulation-specific parameters

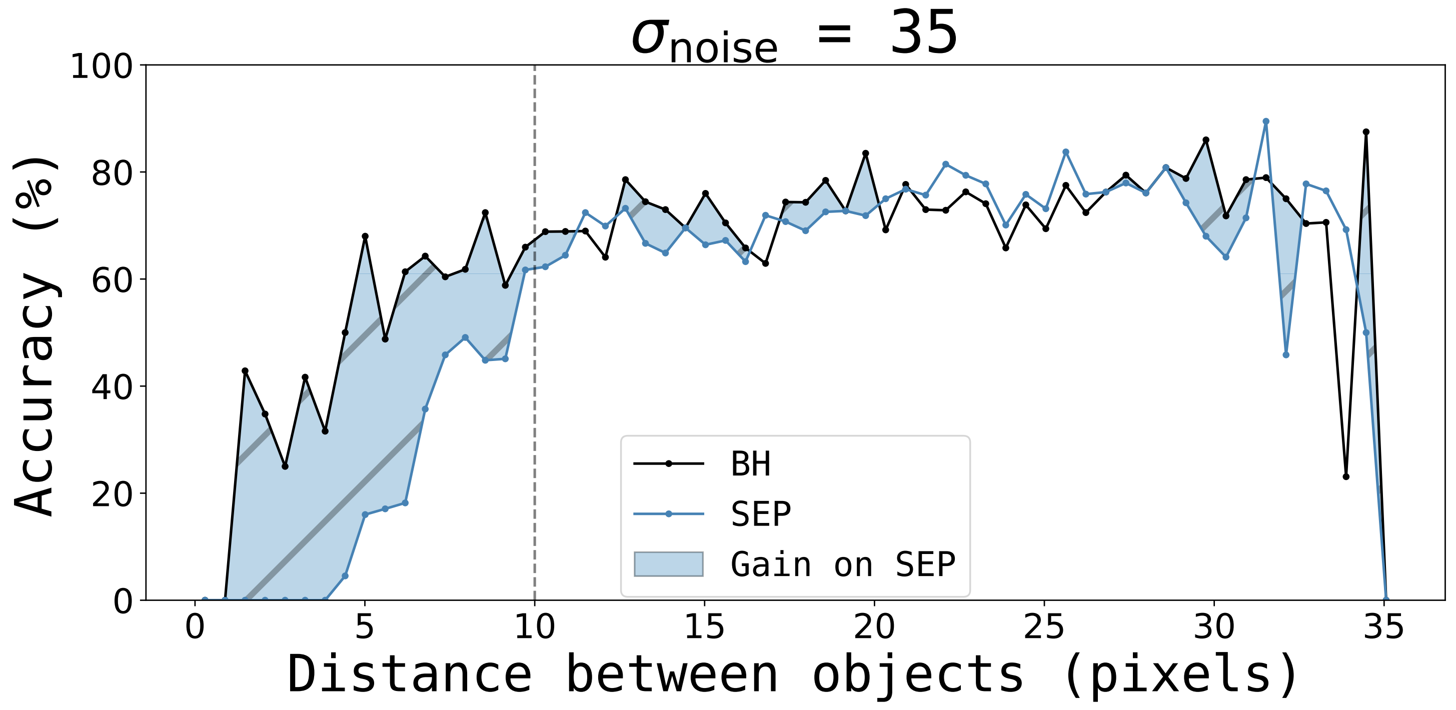
Blend Identification

Results

Small amount of noise added to simulated images



Larger amount of noise added to simulated images

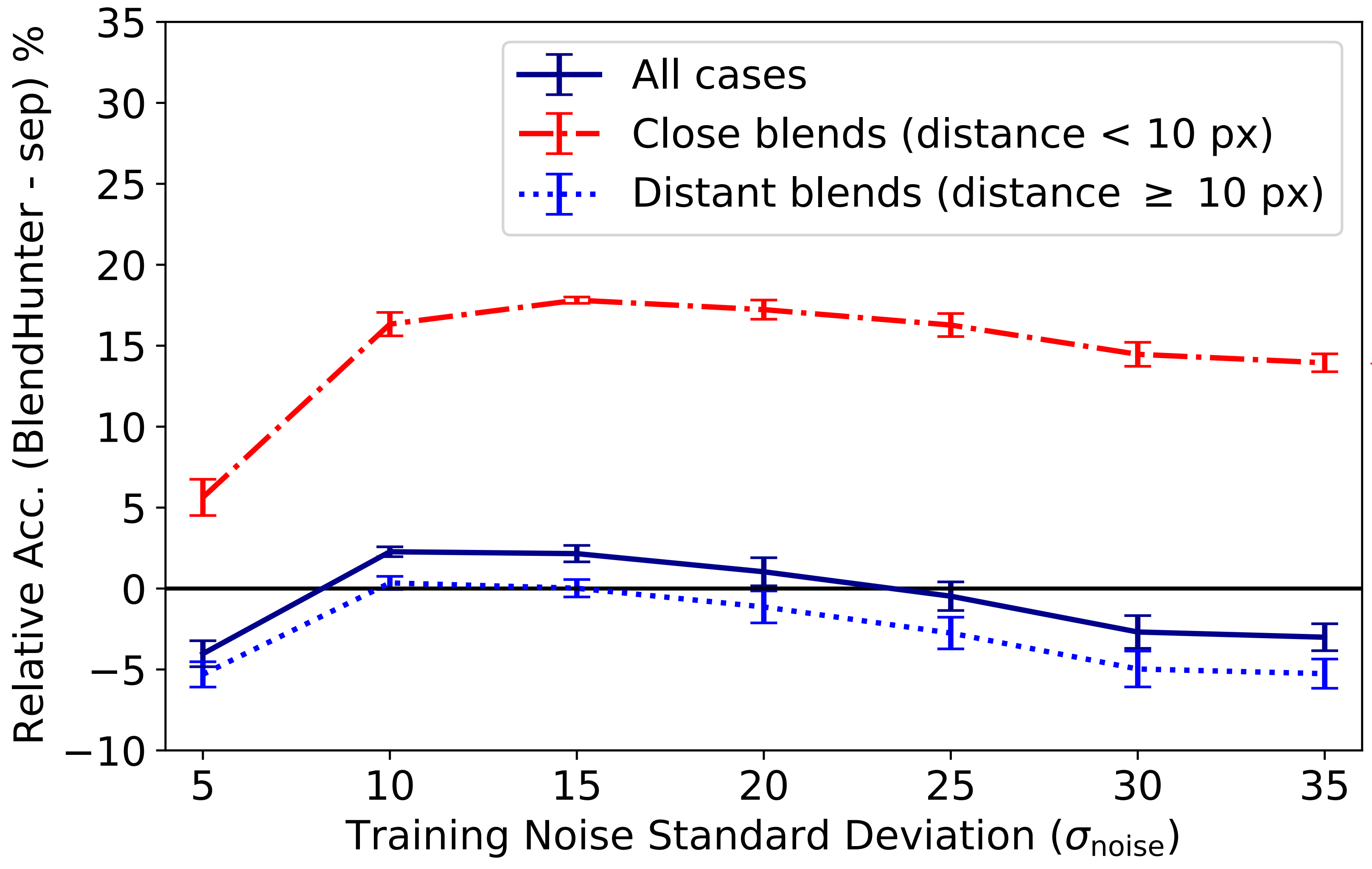


Results from simple parametric models

Transfer weights

Results from UNIONS-CFIS-like galaxy images

Gain in accuracy with respect to Source Extractor



The deep transfer learning approach outperforms Source Extractor for close blends by ~15%

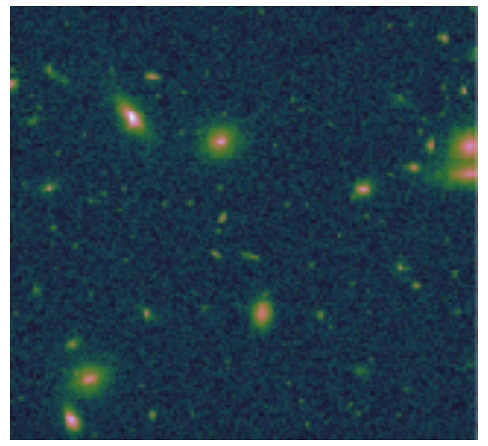
Forward Modelling


E. Centofanti

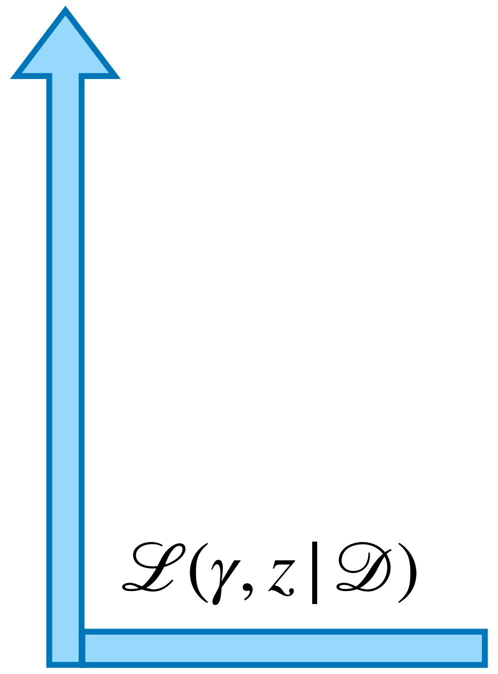

A. Ayçoberry

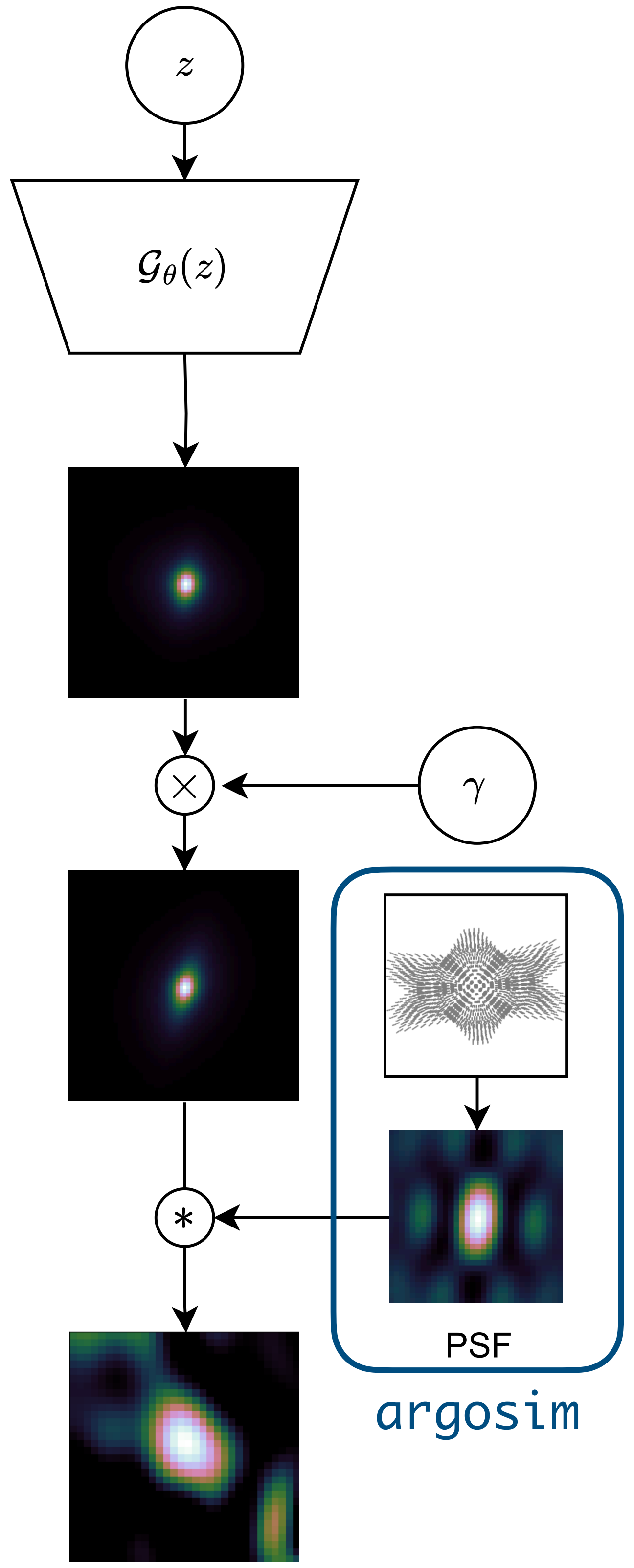

B. Remy


F. Lanusse

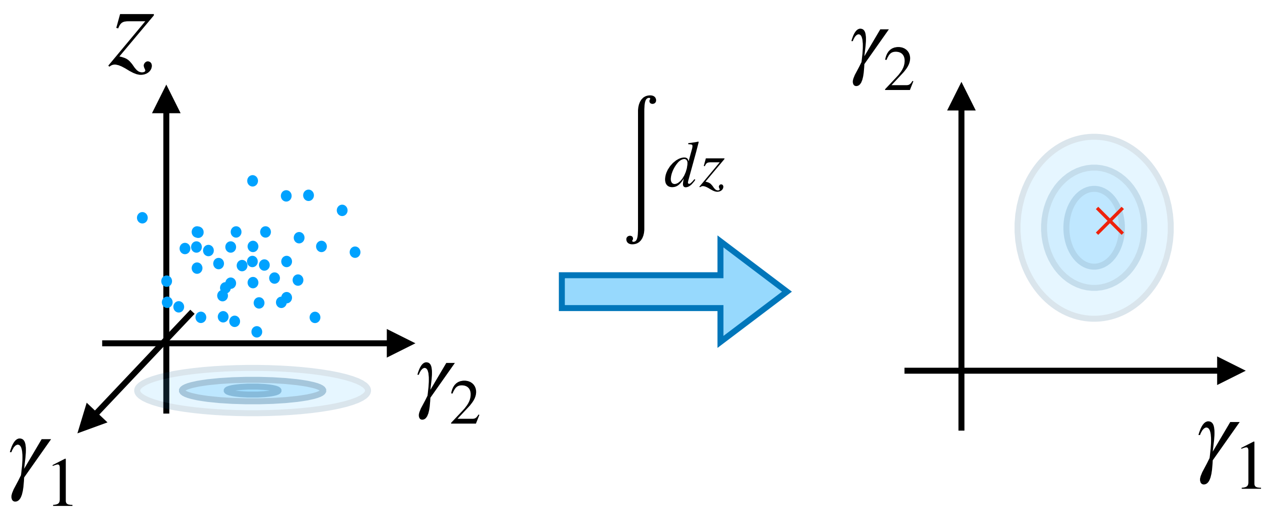


Observed tile
- Shared shear
- Dataset: $\mathcal{D}$





- Sample $(\boldsymbol{z}, \boldsymbol{\gamma})$
- Generate galaxy $\mathcal{G}(\boldsymbol{z})$
- Apply shear and PSF
- Evaluate likelihood
- Accept / reject $\boldsymbol{z}, \boldsymbol{\gamma}$



Bypass shapes?

Do we even need to measure the shapes of galaxies if we can directly model the shear field?

Outline

1. Cosmological motivation

- What are the **problems** we need to solve?
- How does **data** play a role in this?
- Why do we need **AI** and how can we trust the results?

2. Weak gravitational lensing

- 'Observing' **invisible matter**
- Technical **challenges**

3. AI for cosmological data analysis

- Estimating **stellar SEDs** from photometric data
- Identify of **blended galaxies** in survey data
- Going beyond

4. Future perspectives

- Where can AI take us?
- Accessible **tools**
- Conclusions

Outline

4. Future perspectives

- Where can AI take us?
- Accessible **tools**
- Conclusions

AI Prospects for WL

AI Pros

- Many of the technical challenges presented are **well suited for machine learning solutions** as many of them correspond to complex classification, regression, segmentation, etc. problems.
- The volume of data we will be dealing with in upcoming surveys demands **efficient ways** to extract the **maximum amount of information**.
- Novel new methods are coming out daily for **optimally compressing data** and **inferring cosmological parameters** without needing an analytical likelihood (e.g. SBI, field-level, etc.)

AI Cons

- Lots of ***proof-of-concept* work** being done and more full **practical applications are needed**. Do these tools work with real survey data?
- Need **end-to-end pipeline** implementations incorporating AI tools. Replace individual components of traditional pipelines or skip multiple steps?
- New developments in **uncertainty quantification** need to be incorporated into the tools we actually use. We need to trust the results to answer the big questions.
- Need to work on providing user-friendly **community standard AI tools** for WL analyses. If you build it they will come!

Conclusions

- Despite our ignorance as to the composition of Dark Matter, we can **map out the distribution** of this invisible component of the Universe **using weak gravitational lensing**.
- Despite the simplicity of the premise, there are **many places where things can go wrong in a WL analysis** introducing biases that limit our ability to answer the pressing questions in cosmology.
- **AI tools** offer efficient ways to **solve many challenges we face** in the analysis. Example include:
  - Data-driven modelling of the PSF of the instrument
  - Identification of blended sources
  - Directly inferring the shear field
  - And more
- Plenty of **work is still needed to demonstrate that these tools can be trusted** when we are working towards percent-level measurements of cosmological parameters.
- With surveys like Euclid preparing their first public data releases (~Oct 2026), there is no better time to get ready!