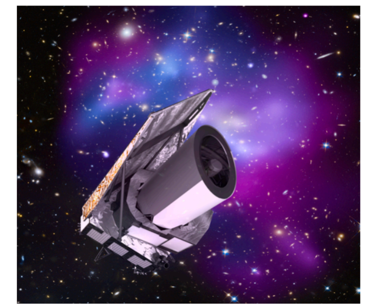
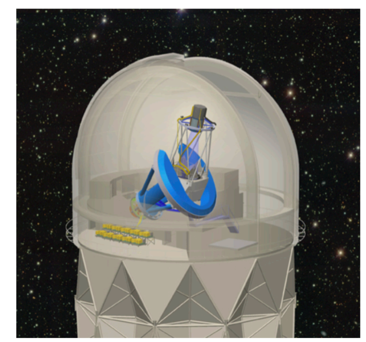
Machine Learning the Universe with Large Sky Surveys



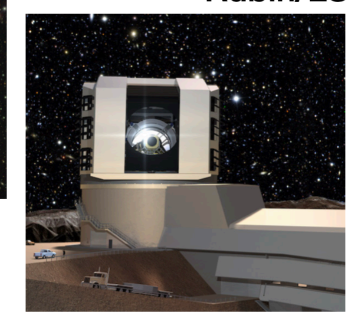
Euclid

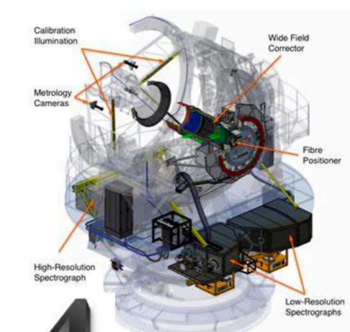
CSST



DESI

Rubin/LSST





Nicola R. Napolitano

Department of Physics “E. Pancini”
University of Naples Federico II

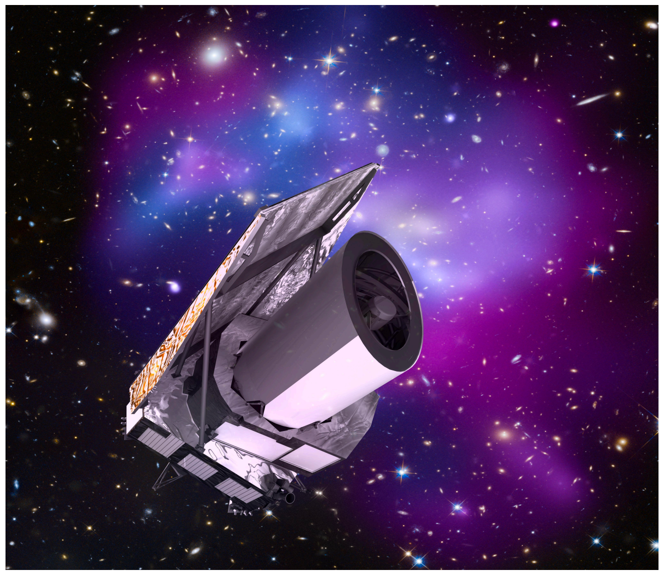
Formerly: Sun Yat-sen University, Zhuhai Campus (China)

My Team: Rui Li (SYSU->NAOC->Zhengzhou) Fucheng Zhong (SYSU/ UniNA), Sirui Wu (SYSU), Lanlan Qiu (SYSU), Ruibiao Luo (SYSU), Cheng Qiu, Xincheng Zhu, Linghua Xie, Leyao Wei

+ Kilo Degree Survey Consortium + Department of Physics University of Naples

Next Generation Surveys will collect billions of galaxies

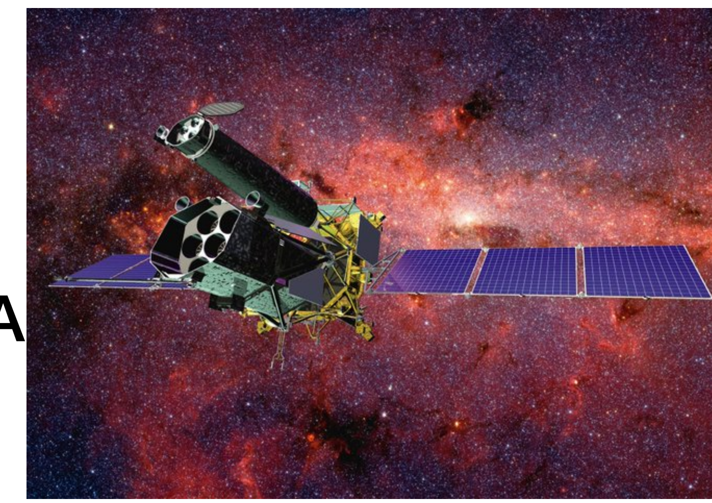
IV Stage surveys are observing the full sky in all wavelengths

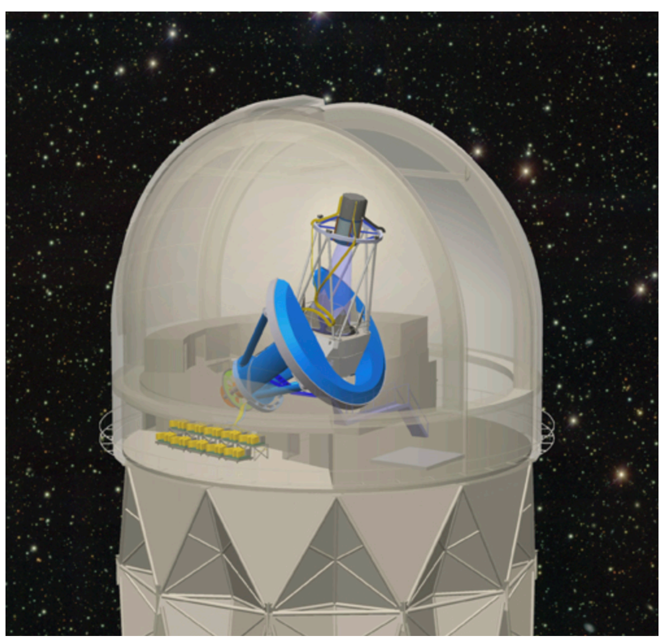


Euclid

CSST

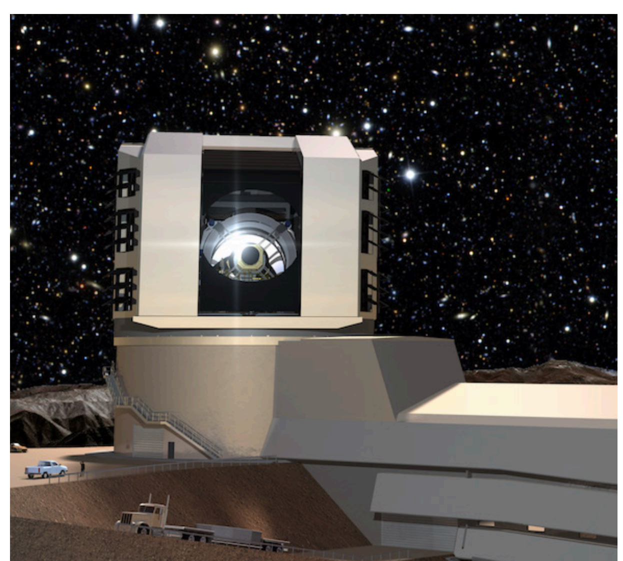
eROSITA

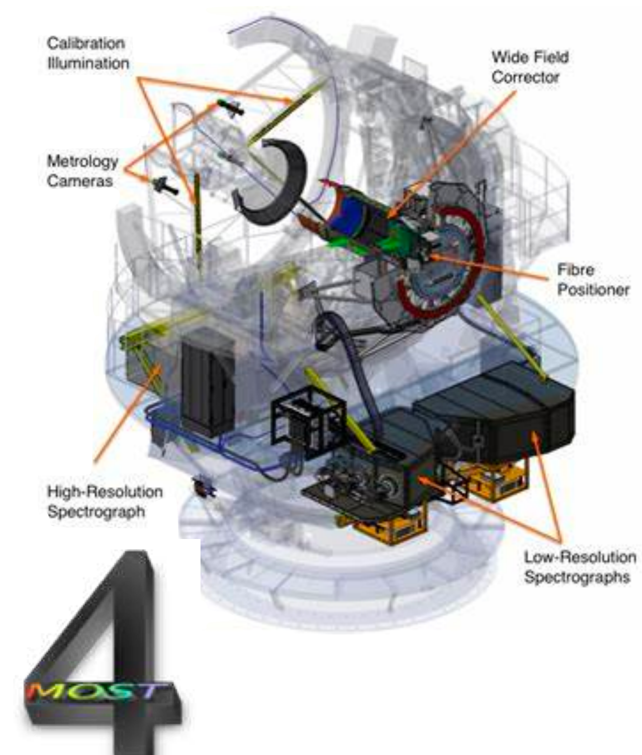




DESI

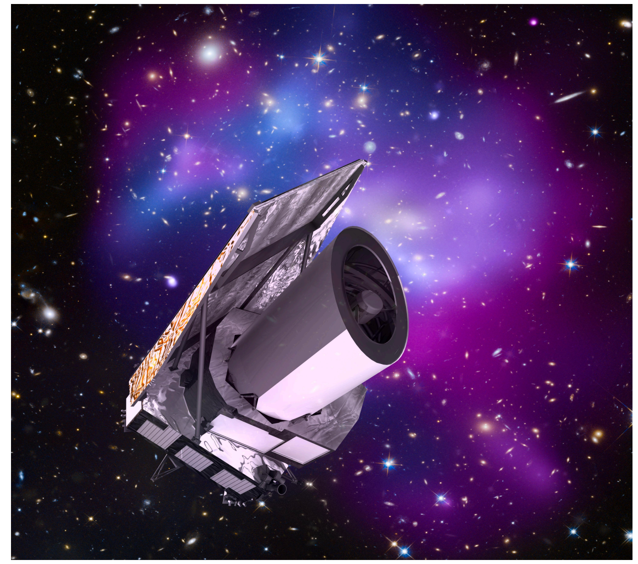
Rubin/LSST



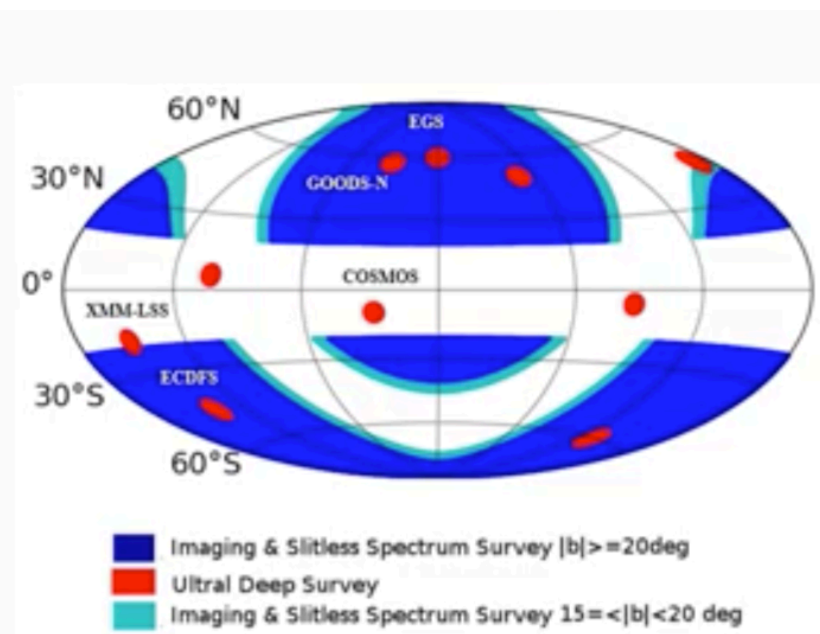


Next Generation Surveys will collect billions of galaxies

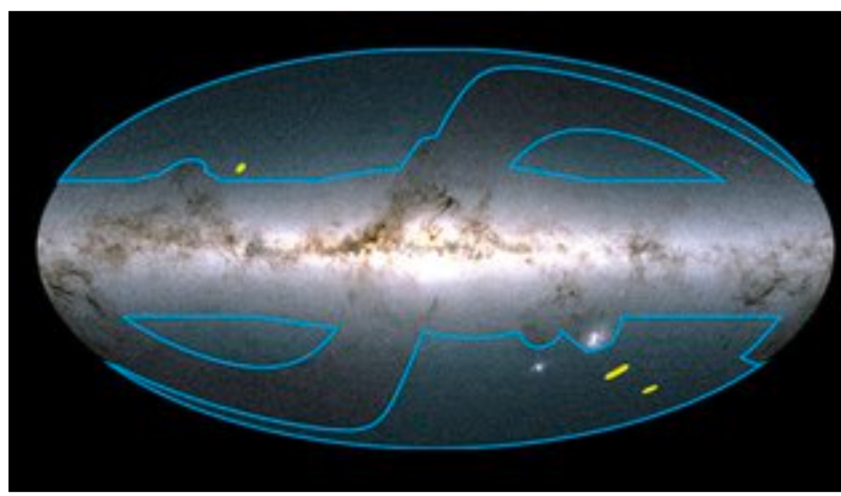
NUV u g r i z y → ***VIS YJH***



Galaxies at 1<z<3 with good mass estimates and morph.	~2x10^8
Massive galaxies (1<z<3) with spectra	~few x 10^3
Hα emitters/metal abundance at z~1-2	~4x10^7/10^4
Galaxies in massive clusters at z>1	~(2-4)x10^4
Type 2 AGN (0.7<z<2)	~10^4
Galaxy mergers	~10^5-few x 10^6
Strong galaxy-scale lenses	~300,000
z > 8 QSOs	~30
Dwarf LSB	~75,000
Ultra Cool Dwarf stars	~few 10^3
Solar System Objects	~1.5 x 10^5

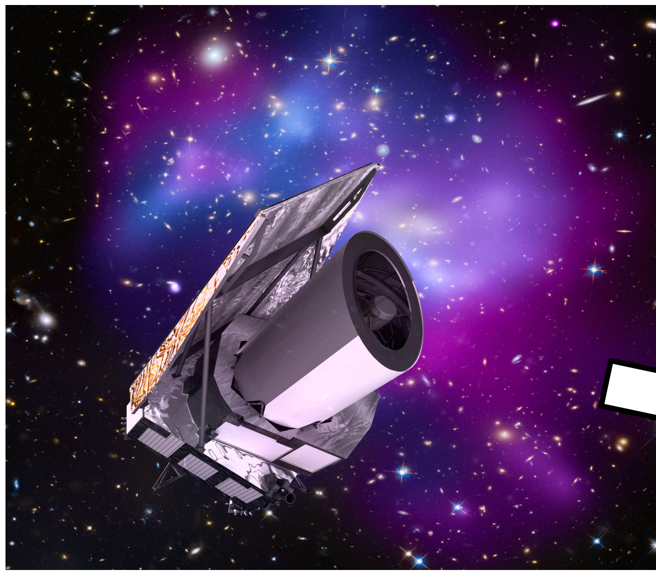


Ecliptic Coord.
Deep fields will be finalized later; sim results for demo only.



Next Generation Surveys will collect billions of galaxies

IV Stage surveys are observing the full sky in all wavelengths

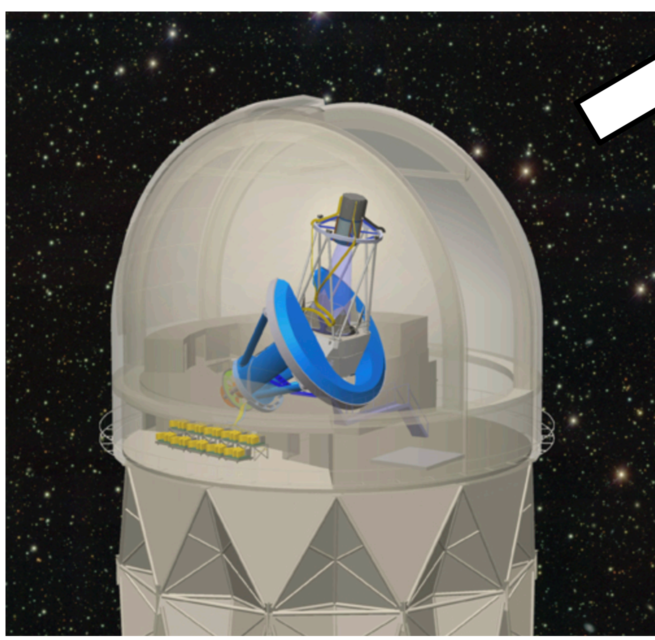


Euclid

CSST

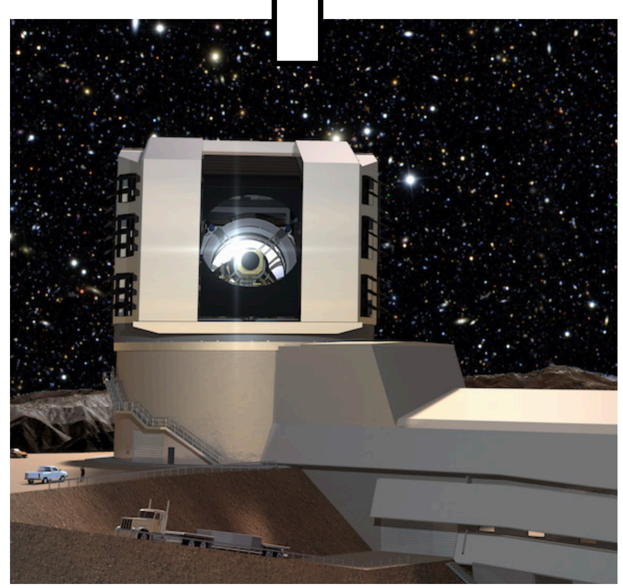
Billions of galaxies

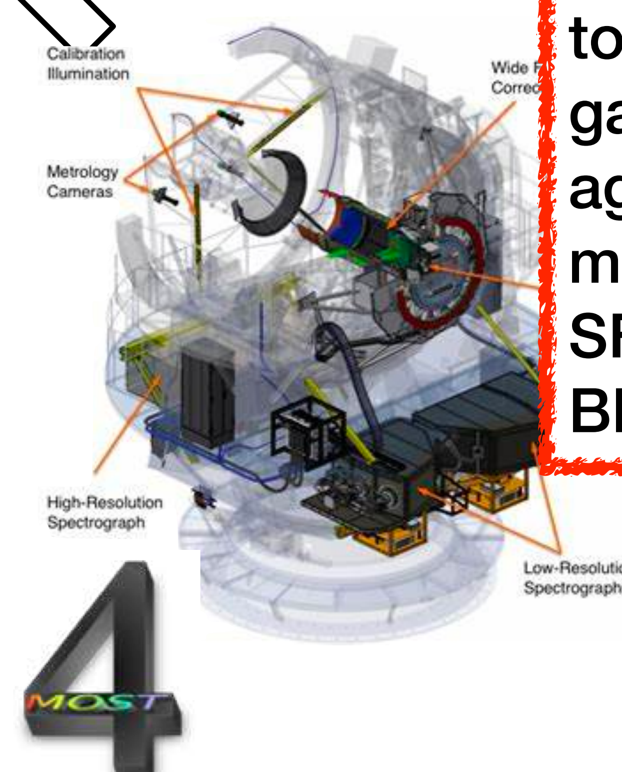
tens of params:
multi-mag
multi-col (opt+NIR)
sizes
morphology
redshifts
velocities
kinematics
stellar masses
total masses
gas mass
ages
metallicity
SFR
BH masses



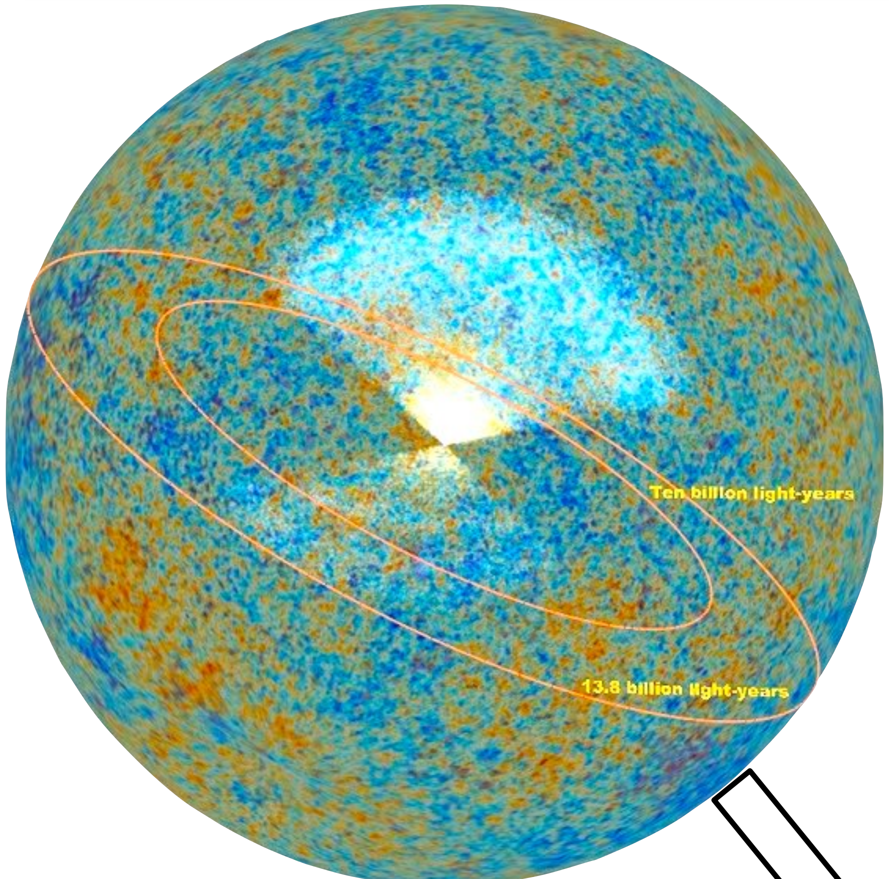
DESI

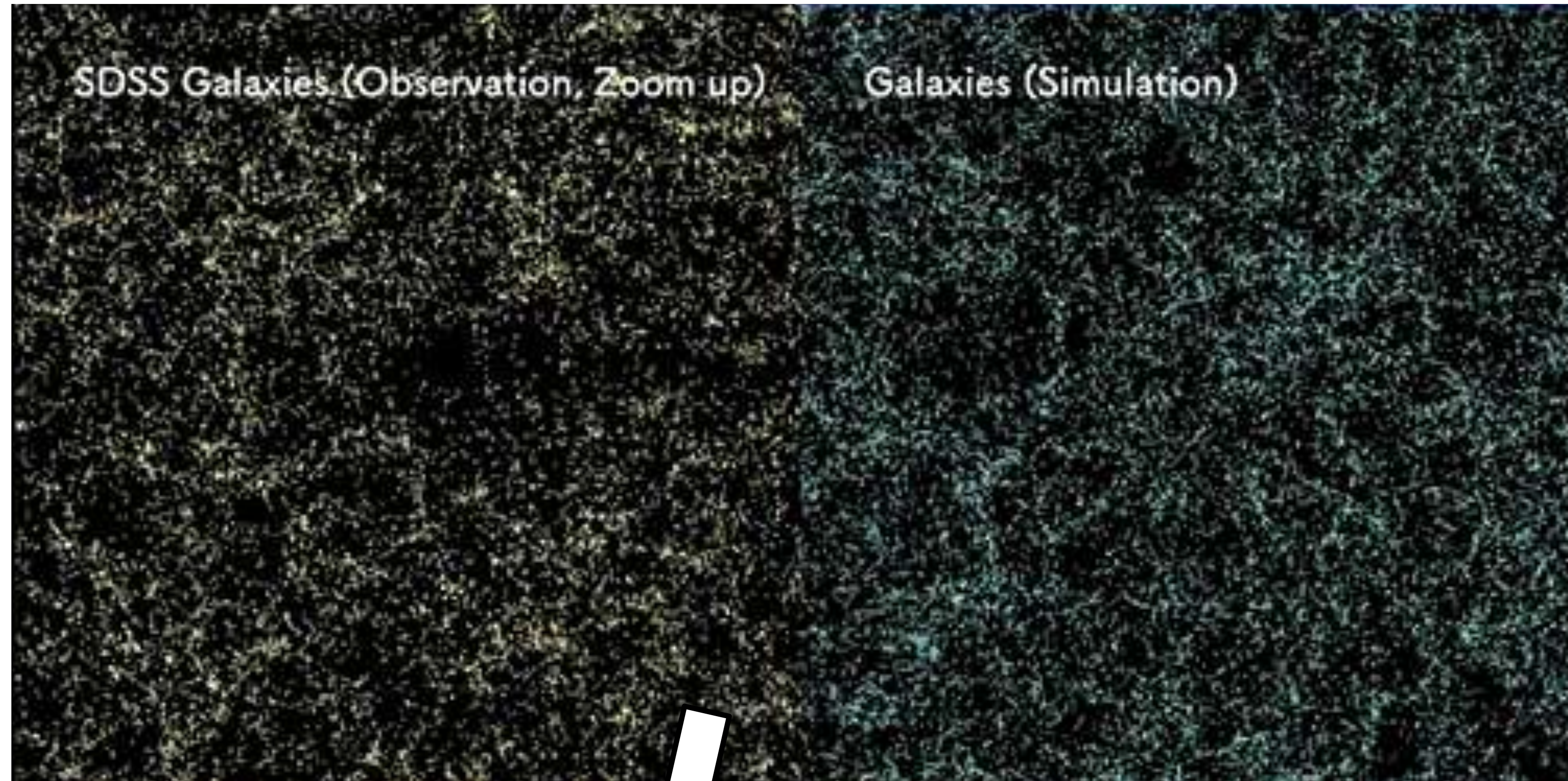
Rubin/LSST



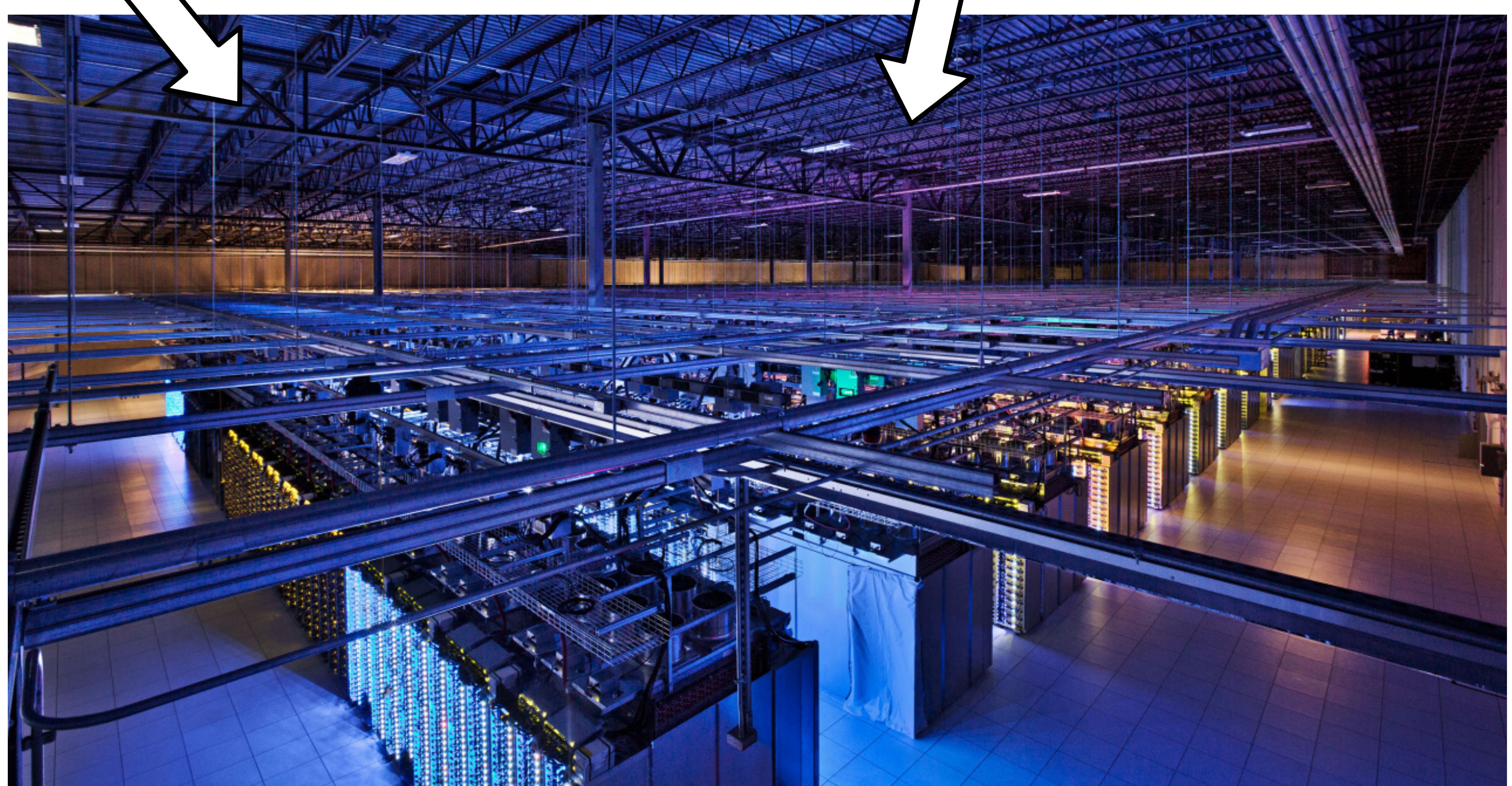


Physical (multi-wav observations) vs. artificial (simulations) universe





Planck+eBOSS



Google

Two main problems

1) How can we efficiently measure billions of galaxy parameters?

   Do we really need all of them (feature importance)?

2) How can we optimise the Science outcome from this tsunami of data?

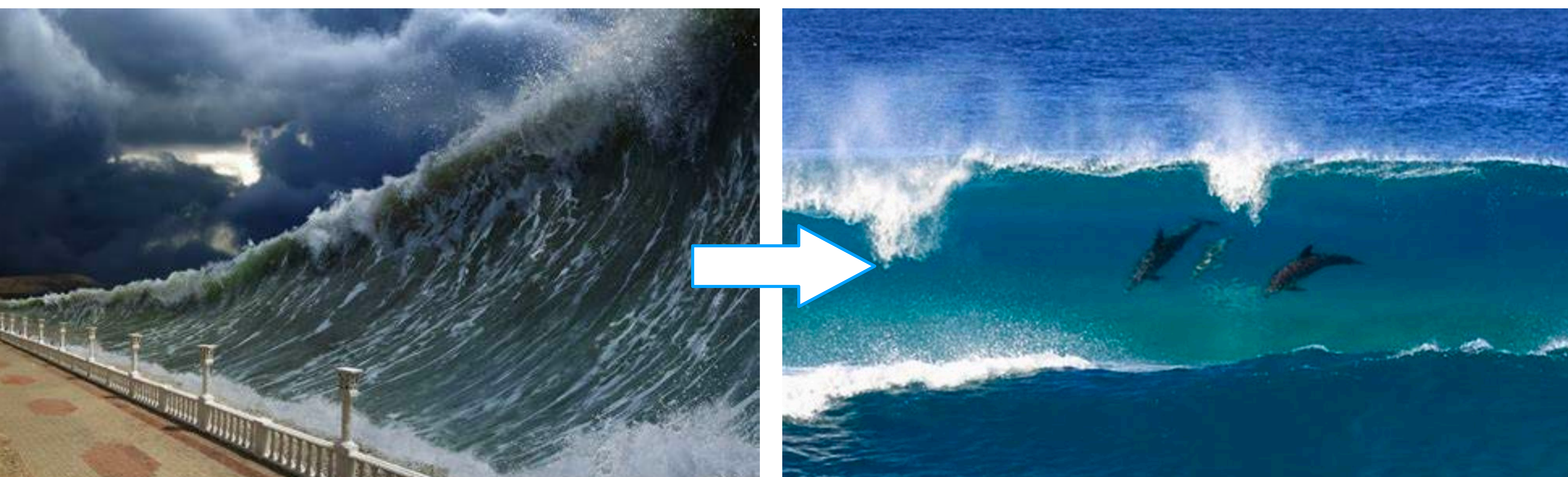
Two main problems

1) How can we efficiently measure billions of galaxy parameters?

Do we really need all of them (feature importance)?

2) How can we optimise the Science outcome from this tsunami of data?

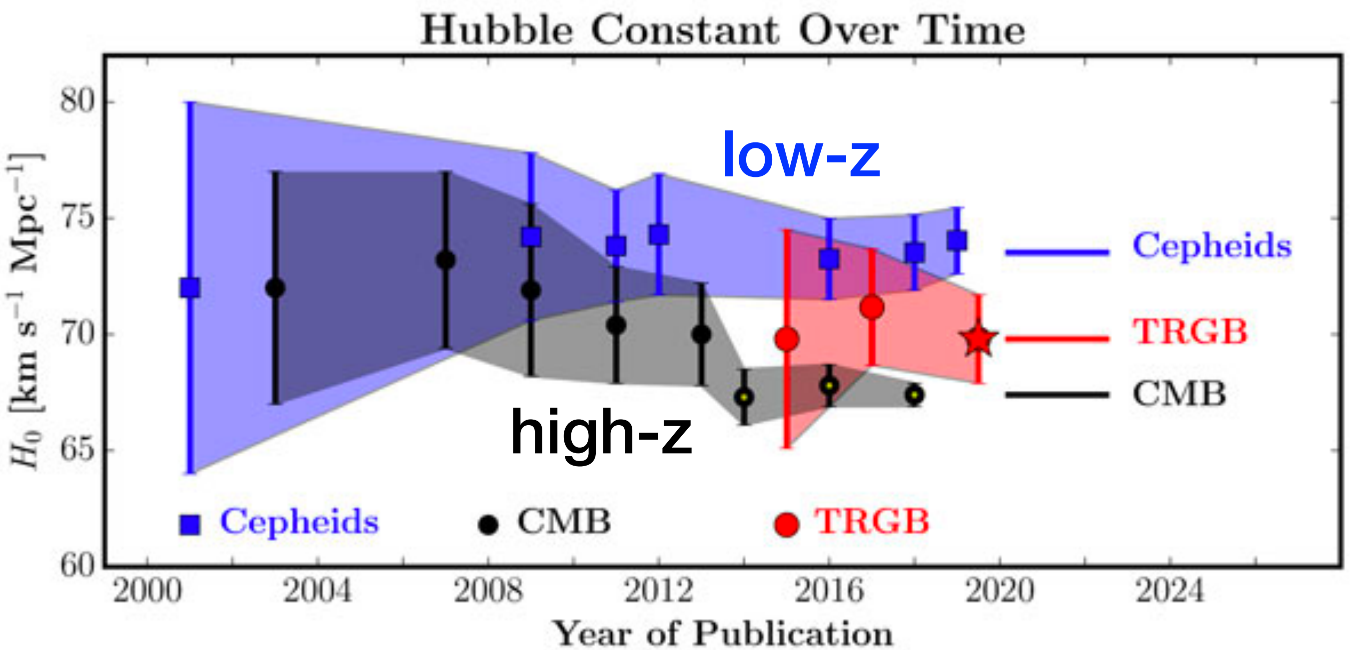
How to go from here **To here**



Can we use galaxies to solve the cosmo parameter tensions?

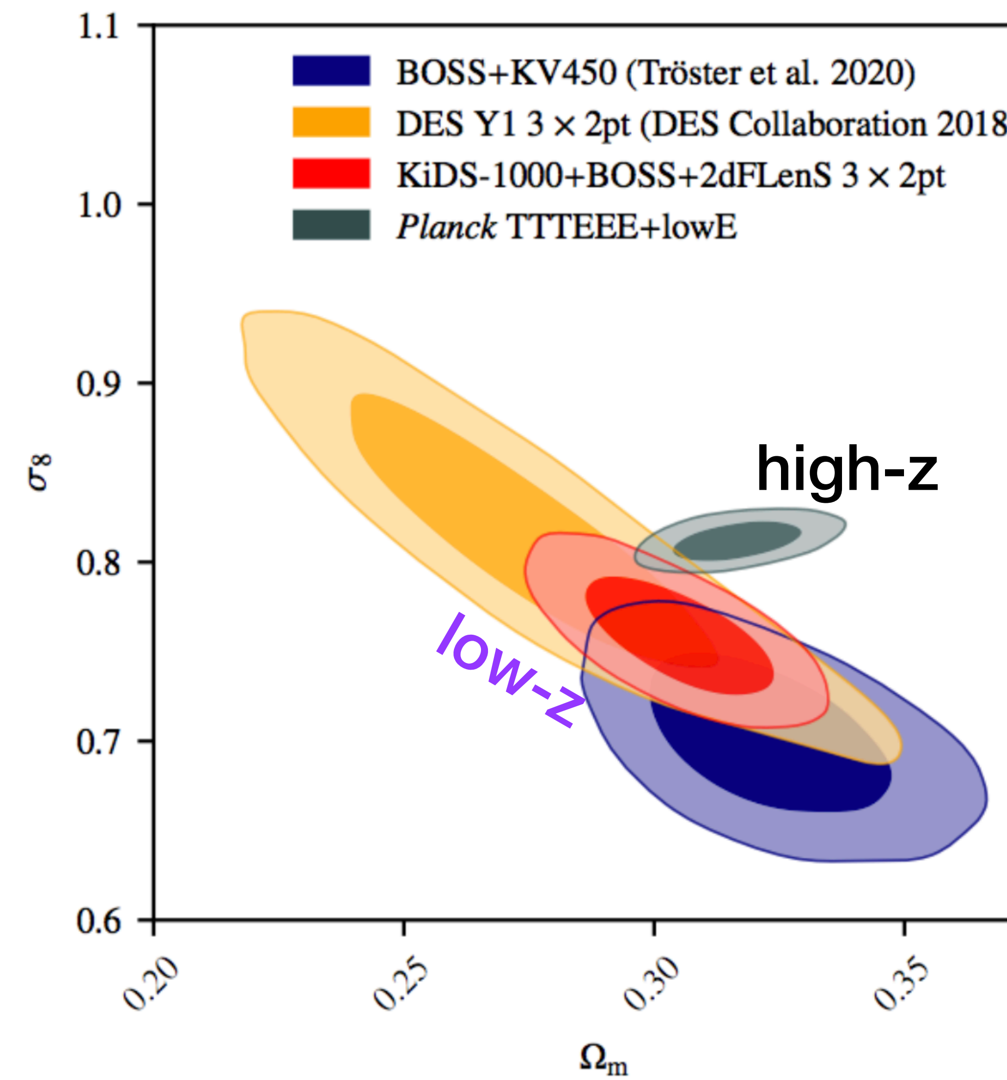
Lately some tensions have emerged in the parameter estimations from high and low-redshift universe

H_0



Freedman et al. 2019

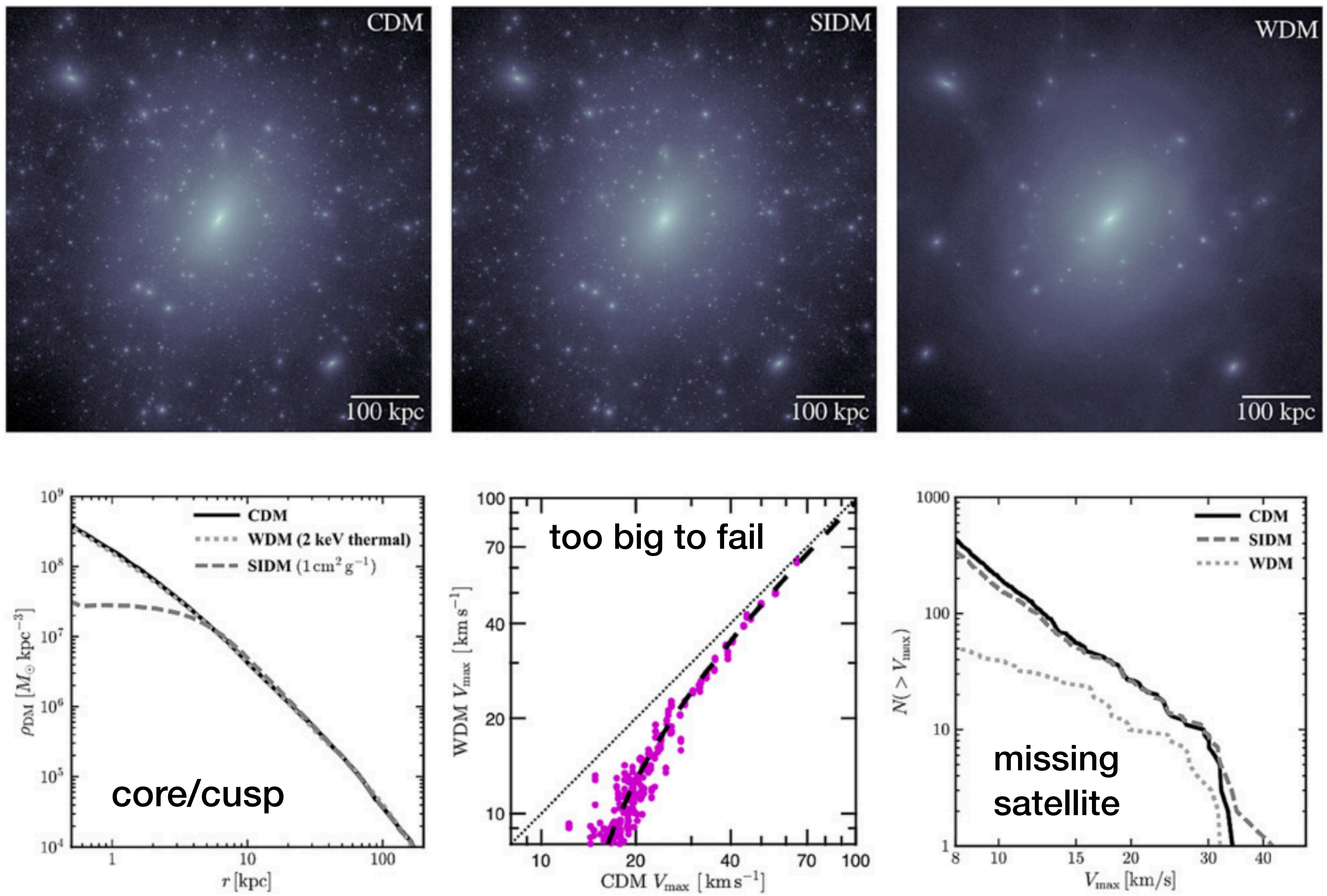
$\Omega_m - \sigma_8$ (S_8)



KiDS collaboration

Can we use galaxies to test the Nature of the Dark Matter?

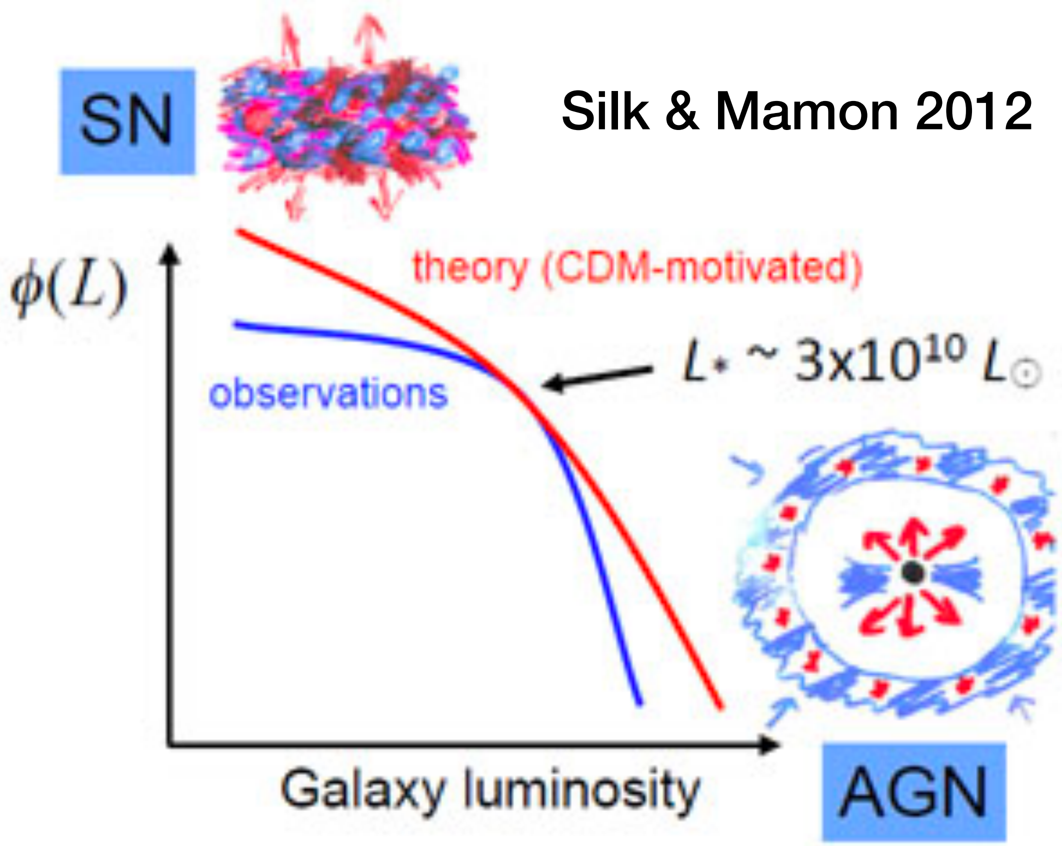
Different DM “flavors” would produce different predictions and alleviate other CDM headaches like the core/cusp, the “too big to fail” and the satellite crisis that we find at galaxy scales.



Mayer 2022

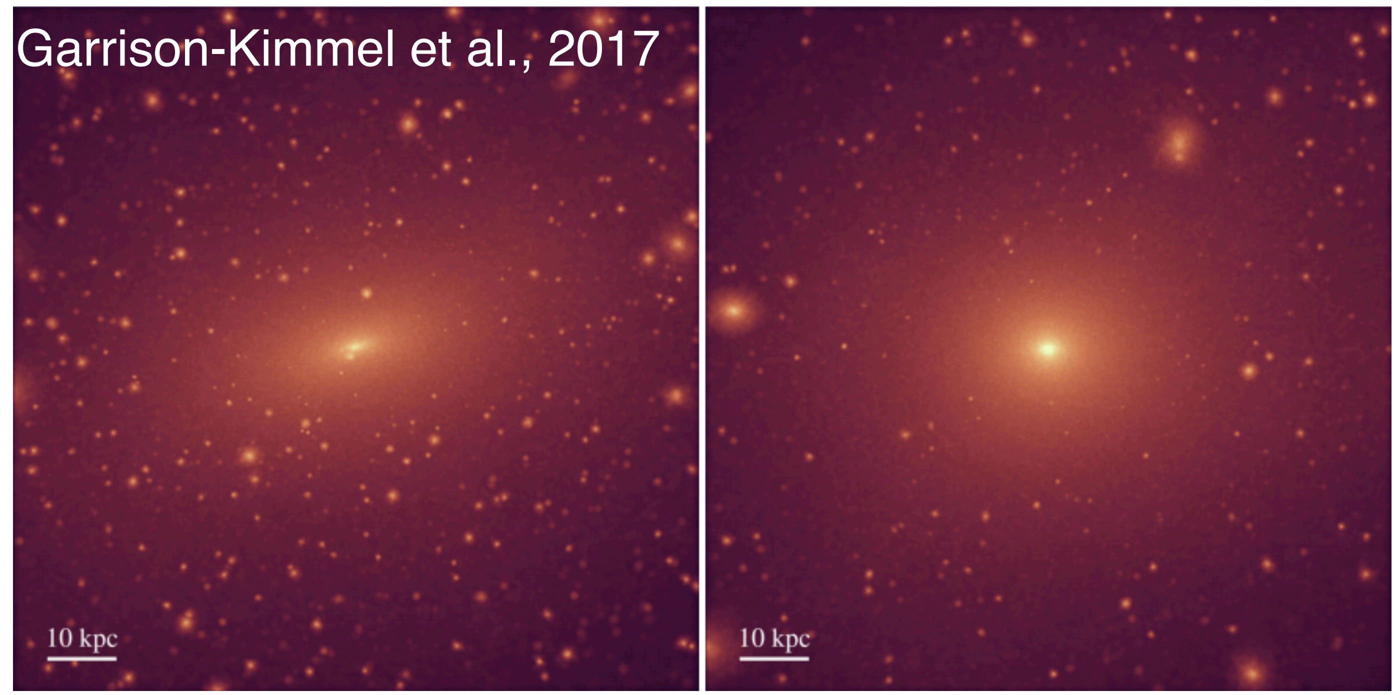
Can we use galaxies to test the Baryon Physics? (Feedback)

And yet we need to fully understand the impact of the baryon physics, as feedbacks from e.g. AGN or supernovae, strongly affect the galaxy formation and determine the number and shape of the galaxies actually formed in the dark haloes



Silk & Mamon 2012

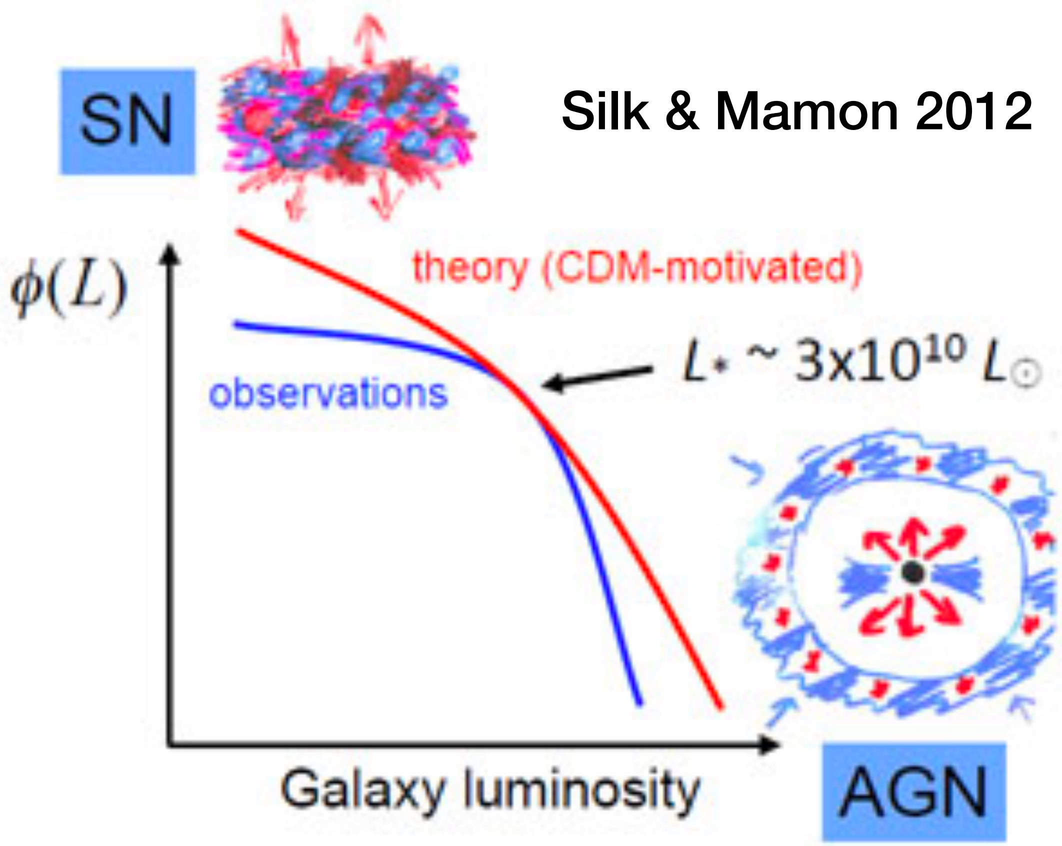
Garrison-Kimmel et al., 2017

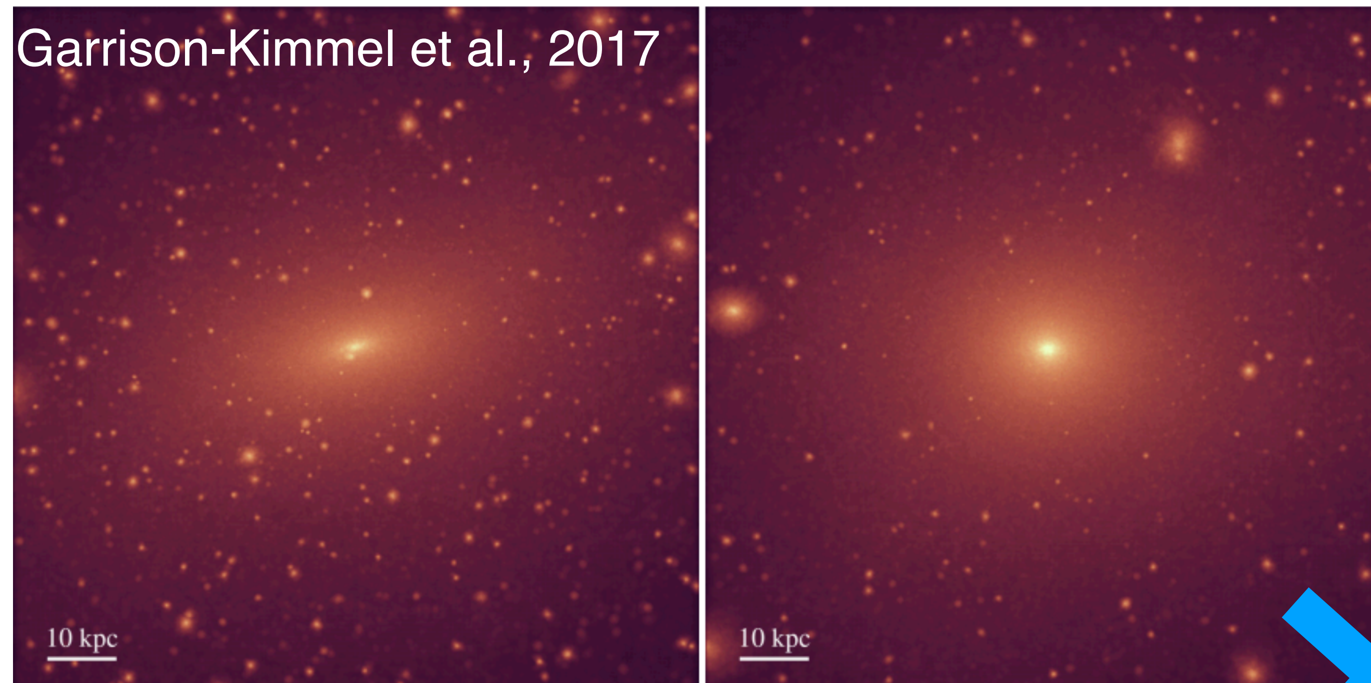


Milky Way-mass host halo (left) and in a hydrodynamic simulation of the same system from the FIRE project

Can we use galaxies to test the Baryon Physics? (Feedback)

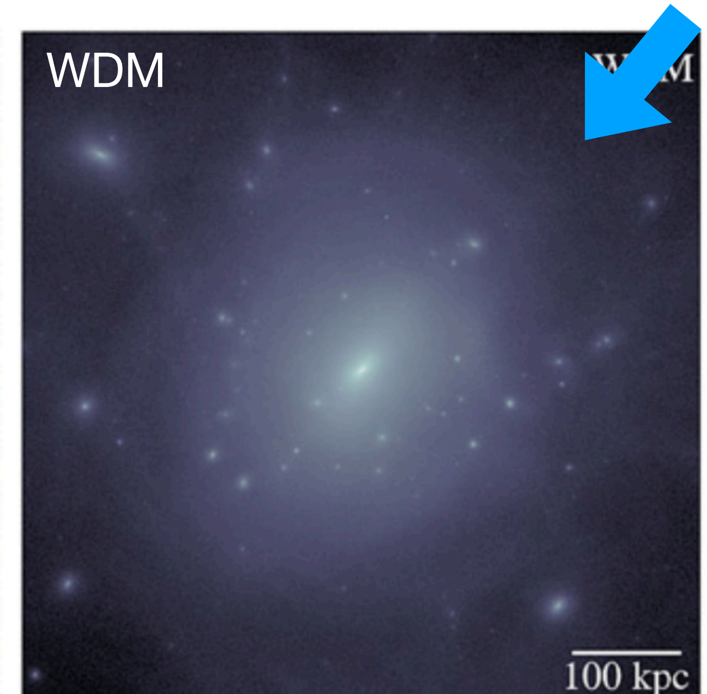
And yet we need to fully understand the impact of the baryon physics, as feedbacks from e.g. AGN or supernovae, strongly affect the galaxy formation and determine the number and shape of the galaxies actually formed in the dark haloes





Milky Way-mass host halo (left) and in a hydrodynamic simulation of the same system from the FIRE project

vs.

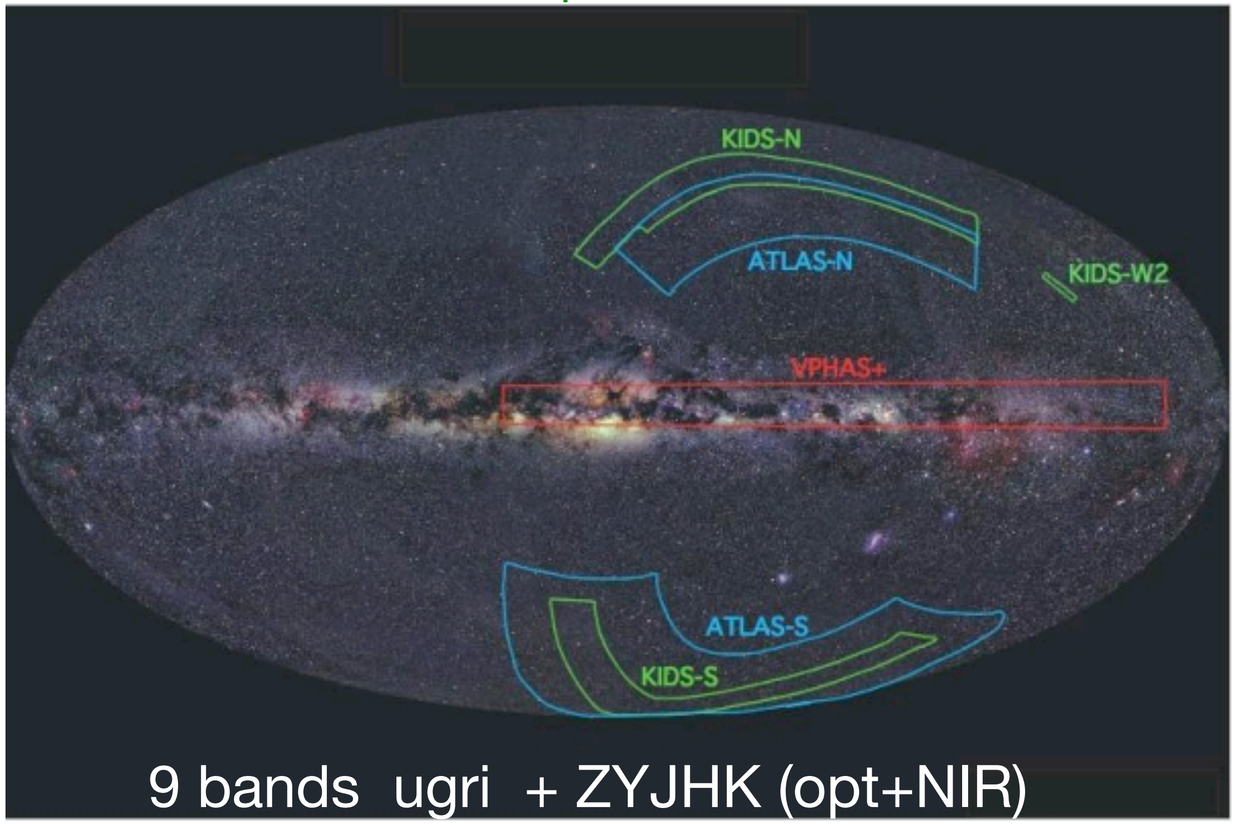


Our reference dataset: the Kilo-Degree Survey

To get prepared for this challenge we have used current high-quality ground-based surveys with image quality comparable to LSST

VLT Survey Telescope
(Made in Naples)

KiDS + VIKING footprint



Kilo Degree Survey (KiDS) @ VST

PI: Kuijken (Leiden) + ~140 members

1350 deg^2 sky

9 bands ugri (optical) + ZYJHK (near infrared) from VIKING @ VISTA (PI: A. Edge)

high-quality imaging (FWHM~0.7” in r-band)

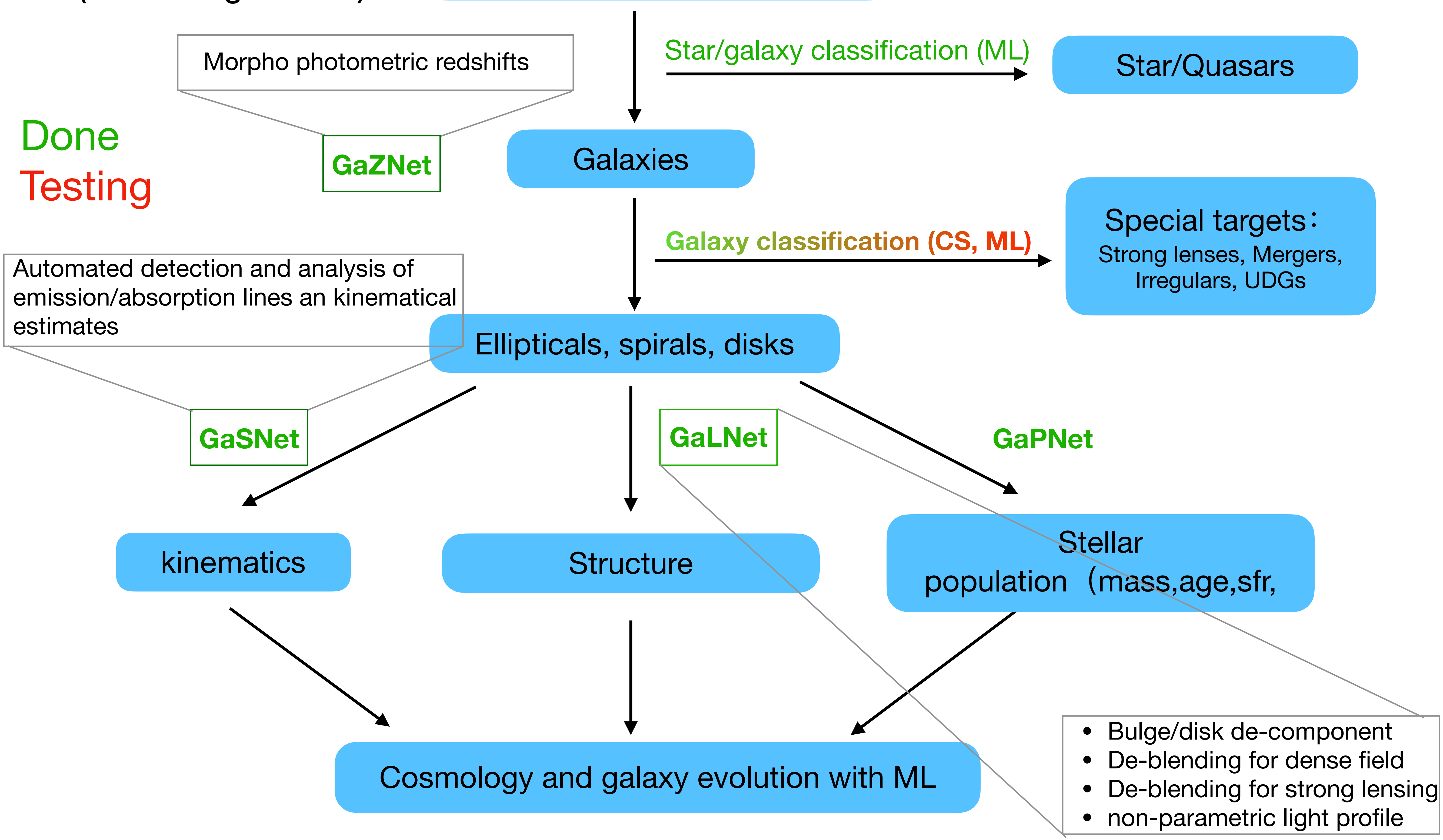
~200 million sources

~60 million galaxies with redshifts

The GaXNet Environment

Rui Li & Nicola R. Napolitano

Rui Li (now Zhengzhou Un)



GaLNets: Convolutional Neural Networks for galaxy structure

Li, NRN, et al. 2021, ApJ

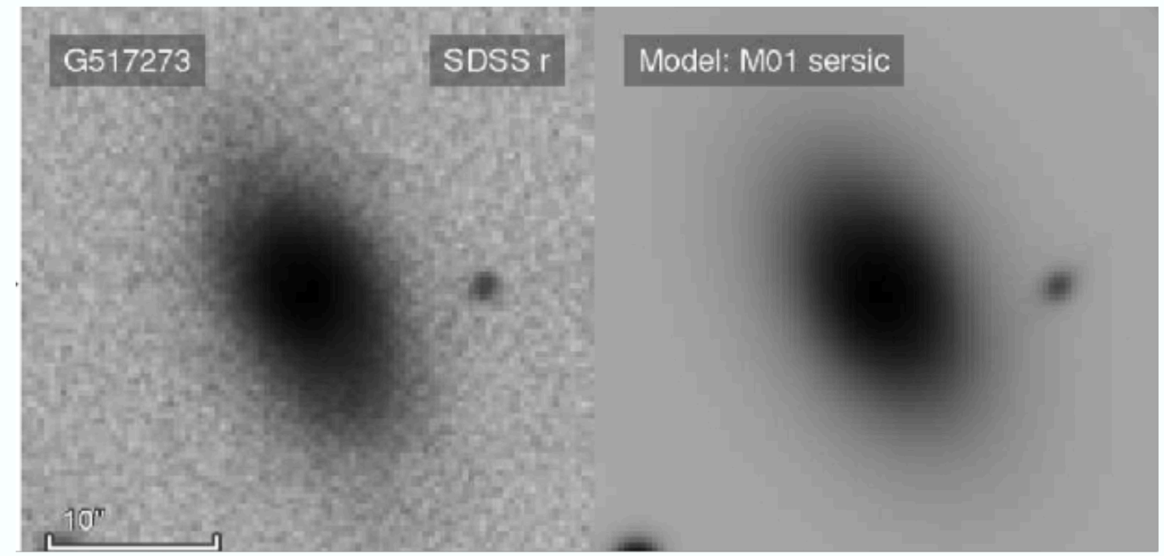
Sersic profile

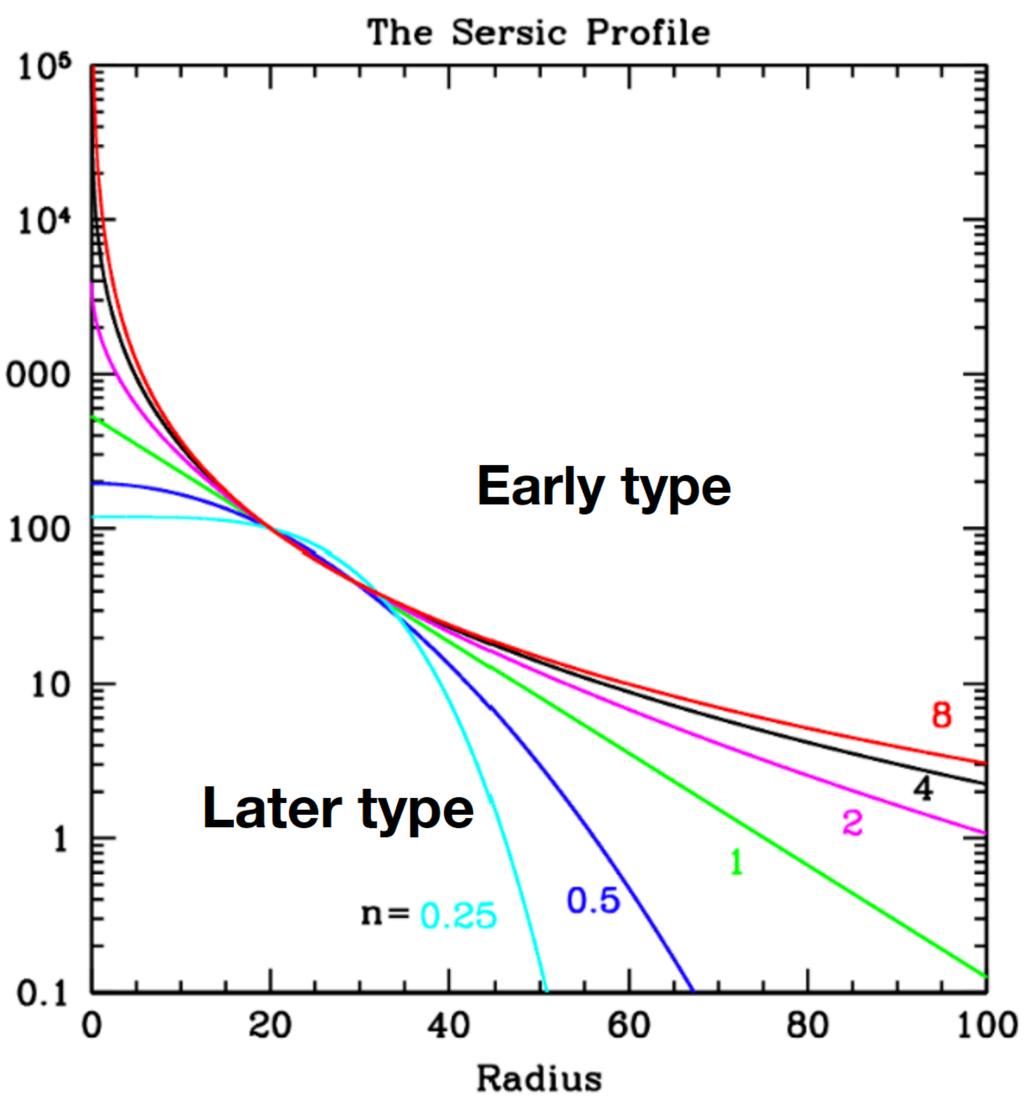
$$I(R) = I_e \exp\{-b_n[(\frac{\sqrt{qx^2 + y^2/q}}{R_{eff}})^{\frac{1}{n}} - 1]\}$$

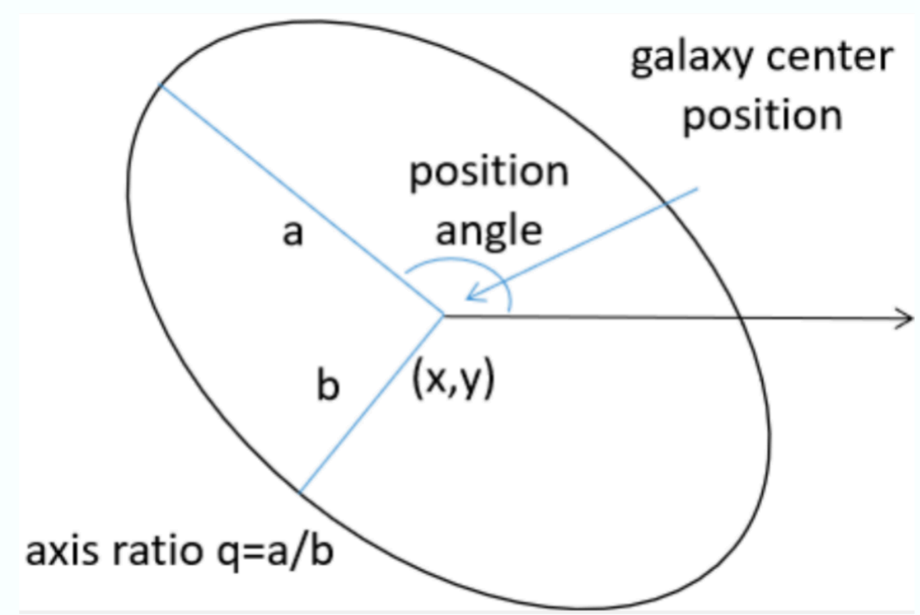
I_e: intensity

R_{eff}: effective radius

n: Sersic index







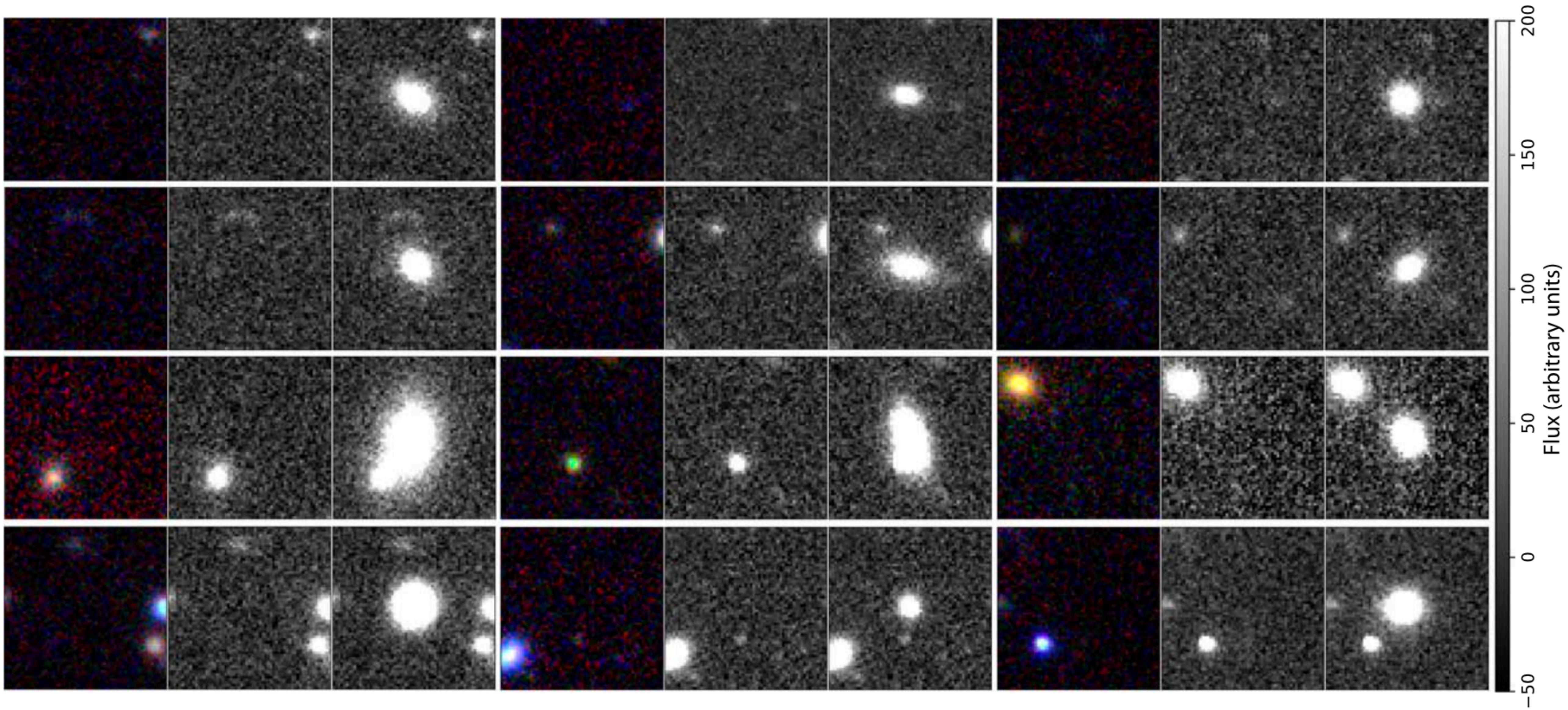
In large surveys: chi2 minimization (Galfit, 2DPHOT) or MCMC (Stalder et al. 2017, using GALPHAT).

GaLNets: Convolutional Neural Networks for galaxy structure

Simulation of training/testing galaxies:

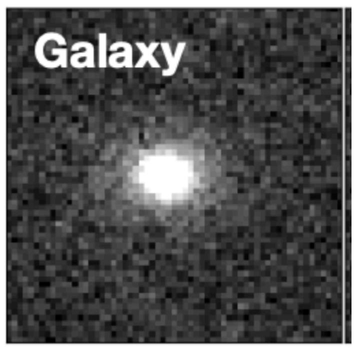
- **Noise**: randomly cutouts from KiDS DR4
- **PSFs**: fitted from the stars using 2DPHOT.
- **Real-like galaxies**: Sersic profile Convolved by PSF, then add noise.

	GaLNet-1	GaLNet-2
Training data	200k	200k
Testing data	20k	20k
Testing data	25k	25k



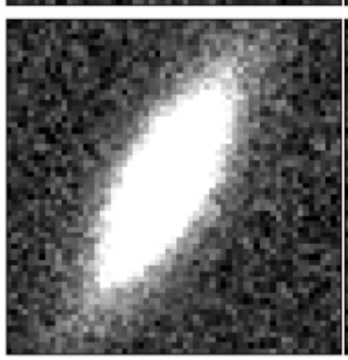
GaLNets: Convolutional Neural Networks for galaxy structure

Regressor CNN

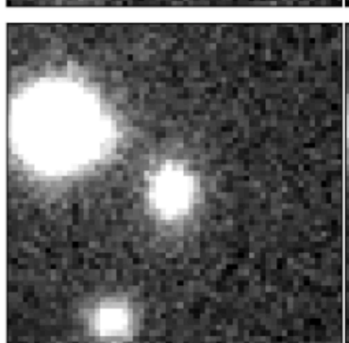


label 1 (p_{11}, p_{12}, p_{13}, p_{14}…)

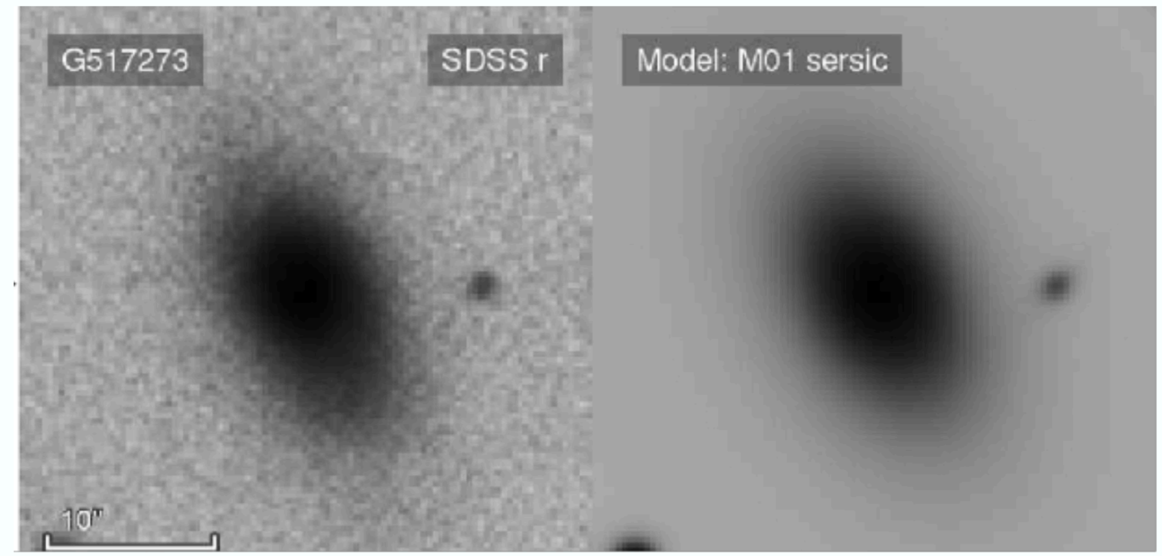
label 2 (p_{21}, p_{22}, p_{23}, p_{24}…)

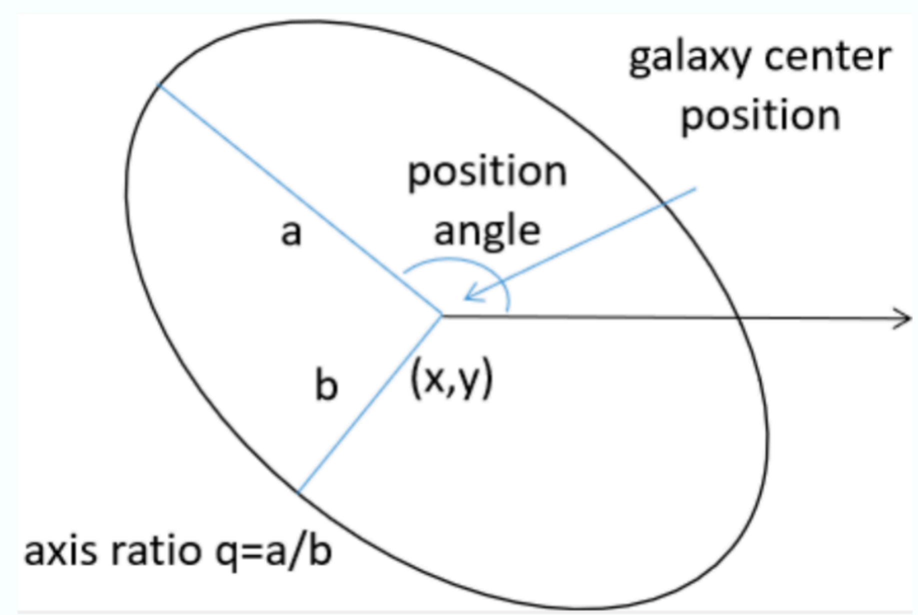


label 3 (p_{31}, p_{32}, p_{33}, p_{34}…)



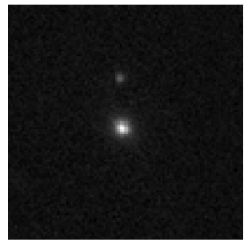
label 4 (p_{41}, p_{42}, p_{43}, p_{44}…)

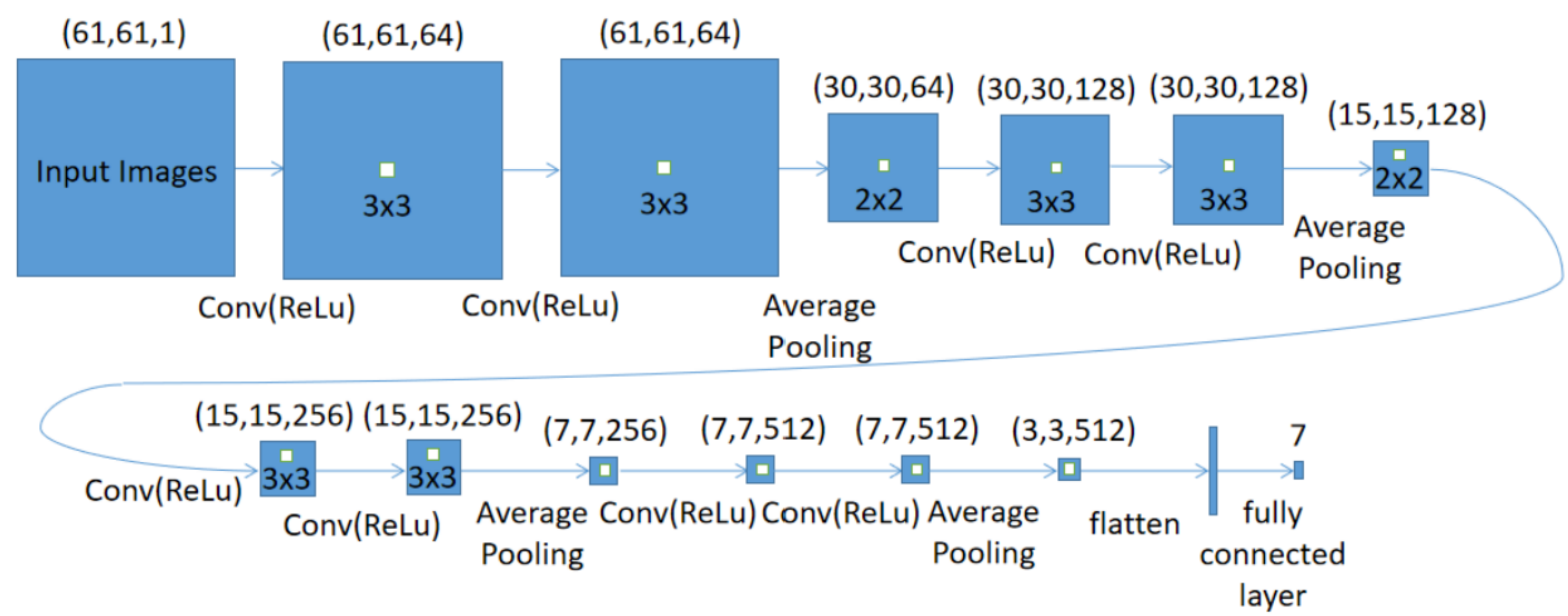




GaLNets: Convolutional Neural Networks for galaxy structure

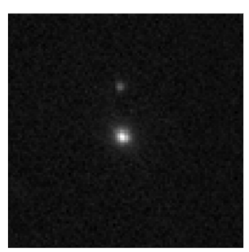
GaLNet-1

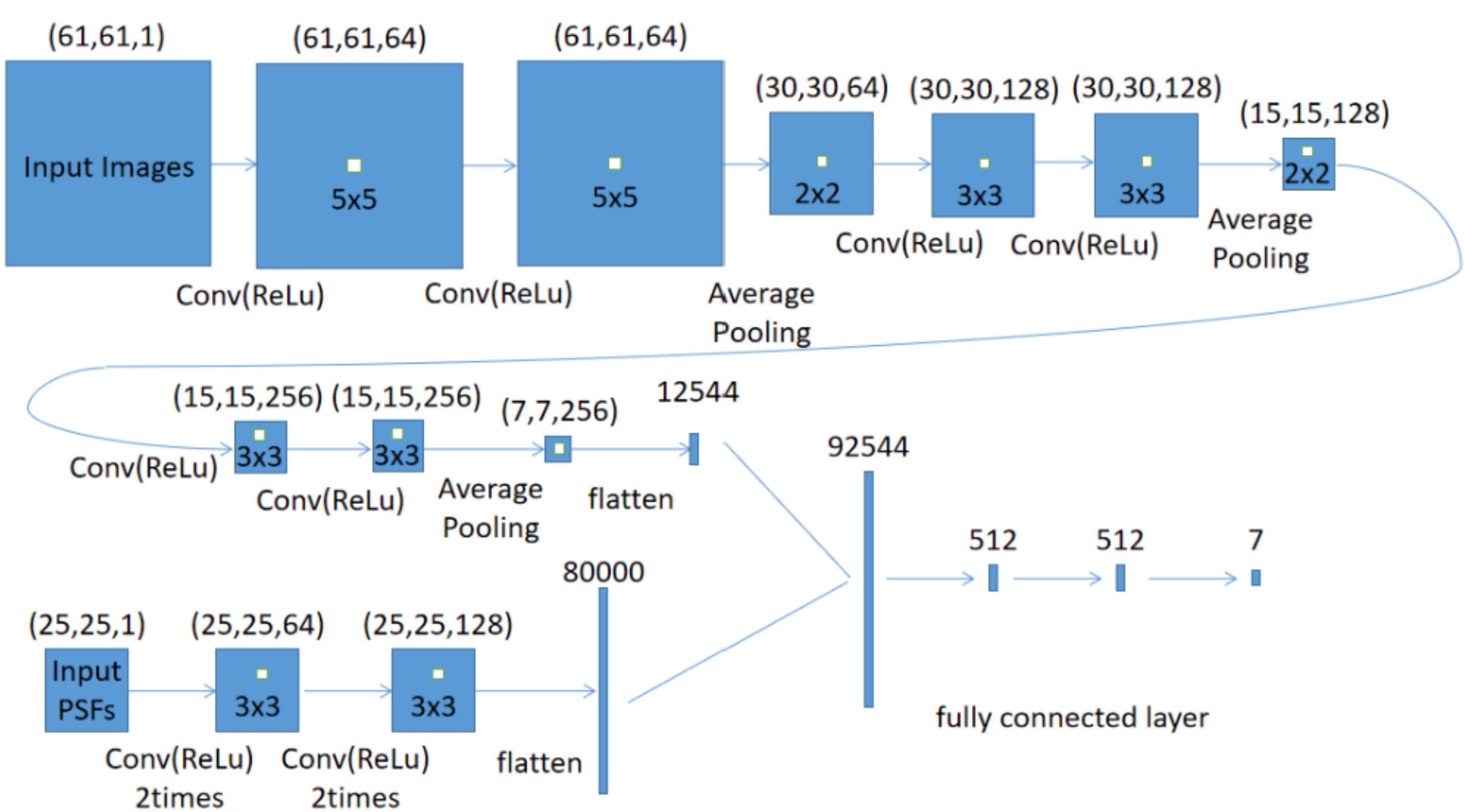




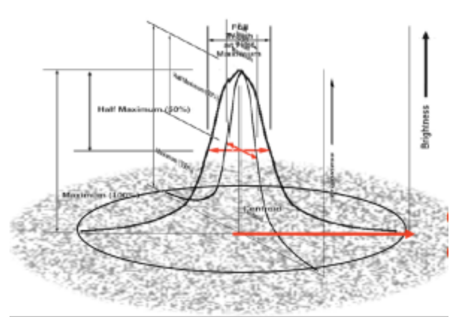
GaLNet-2

galaxy



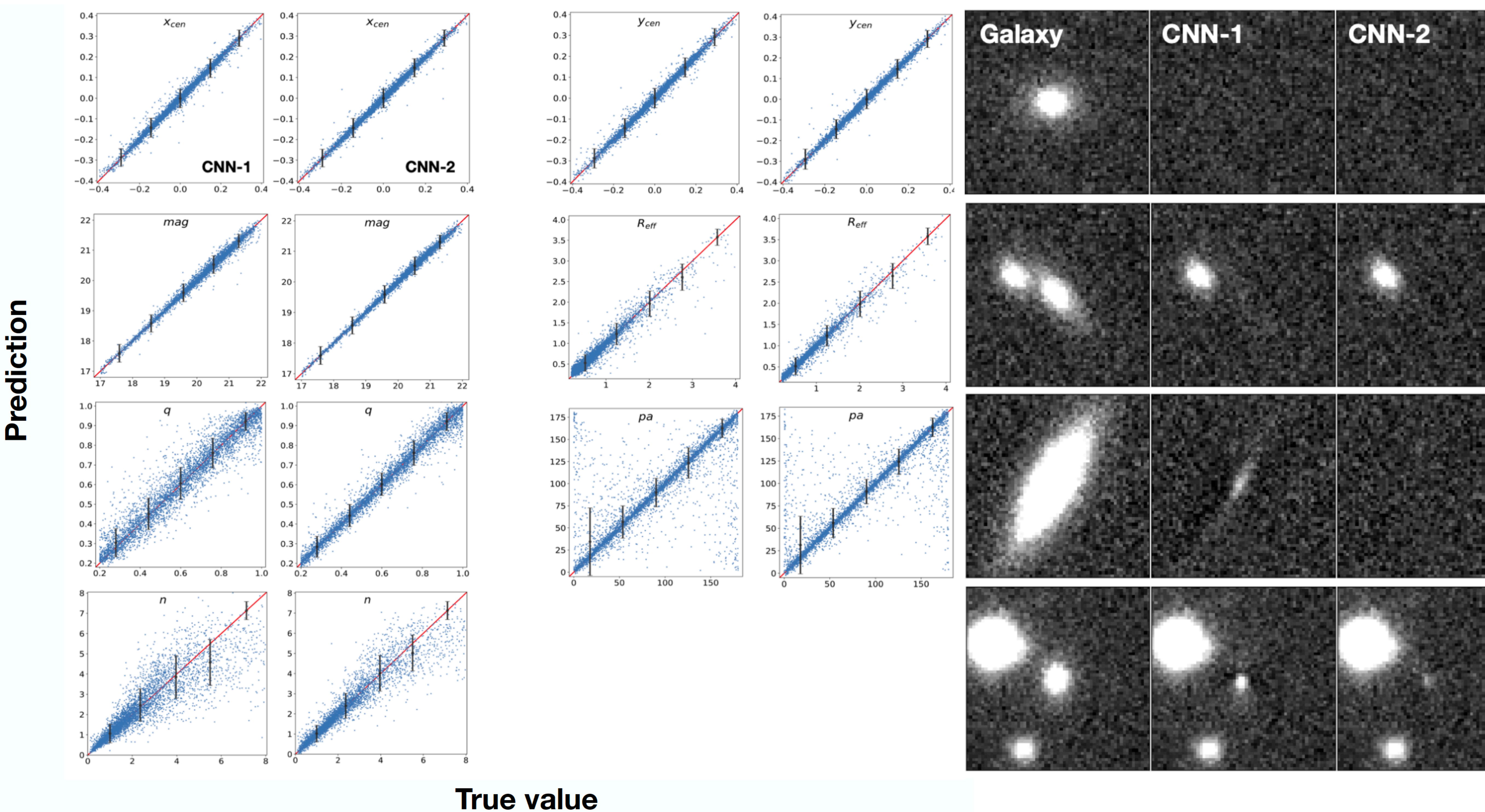


psf

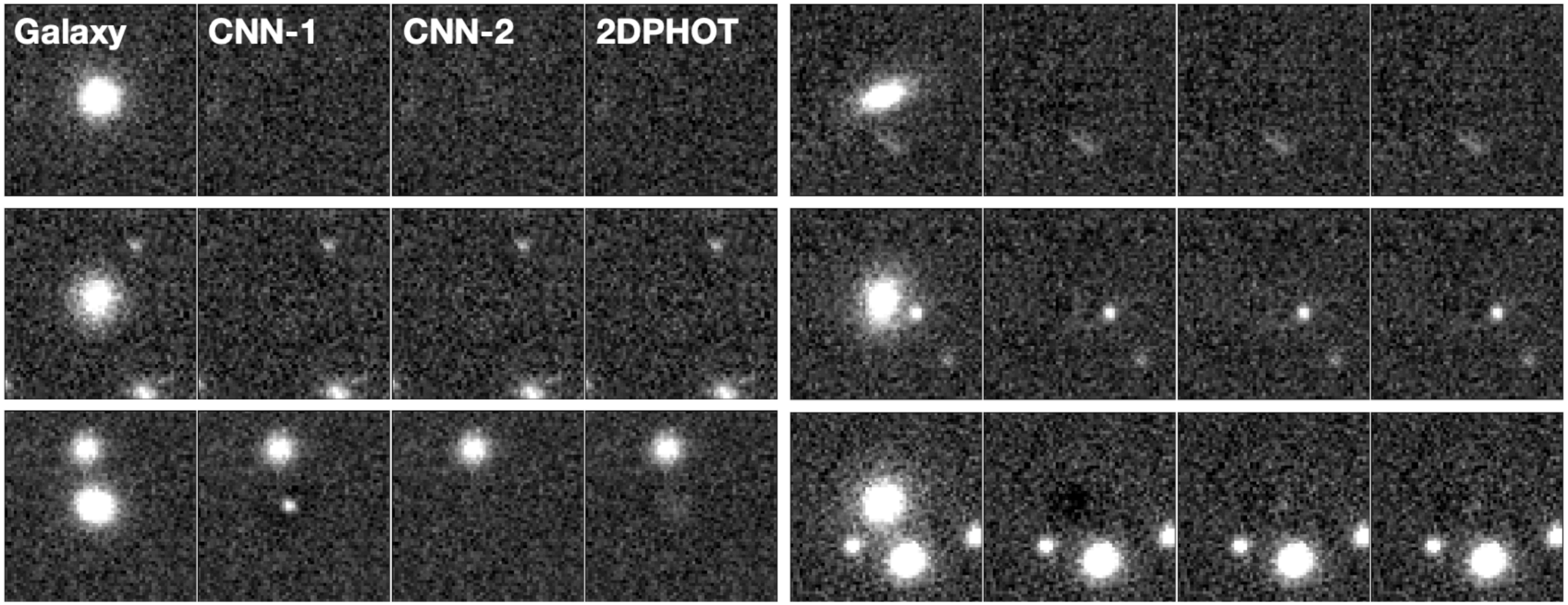


Takes into account the local PSF

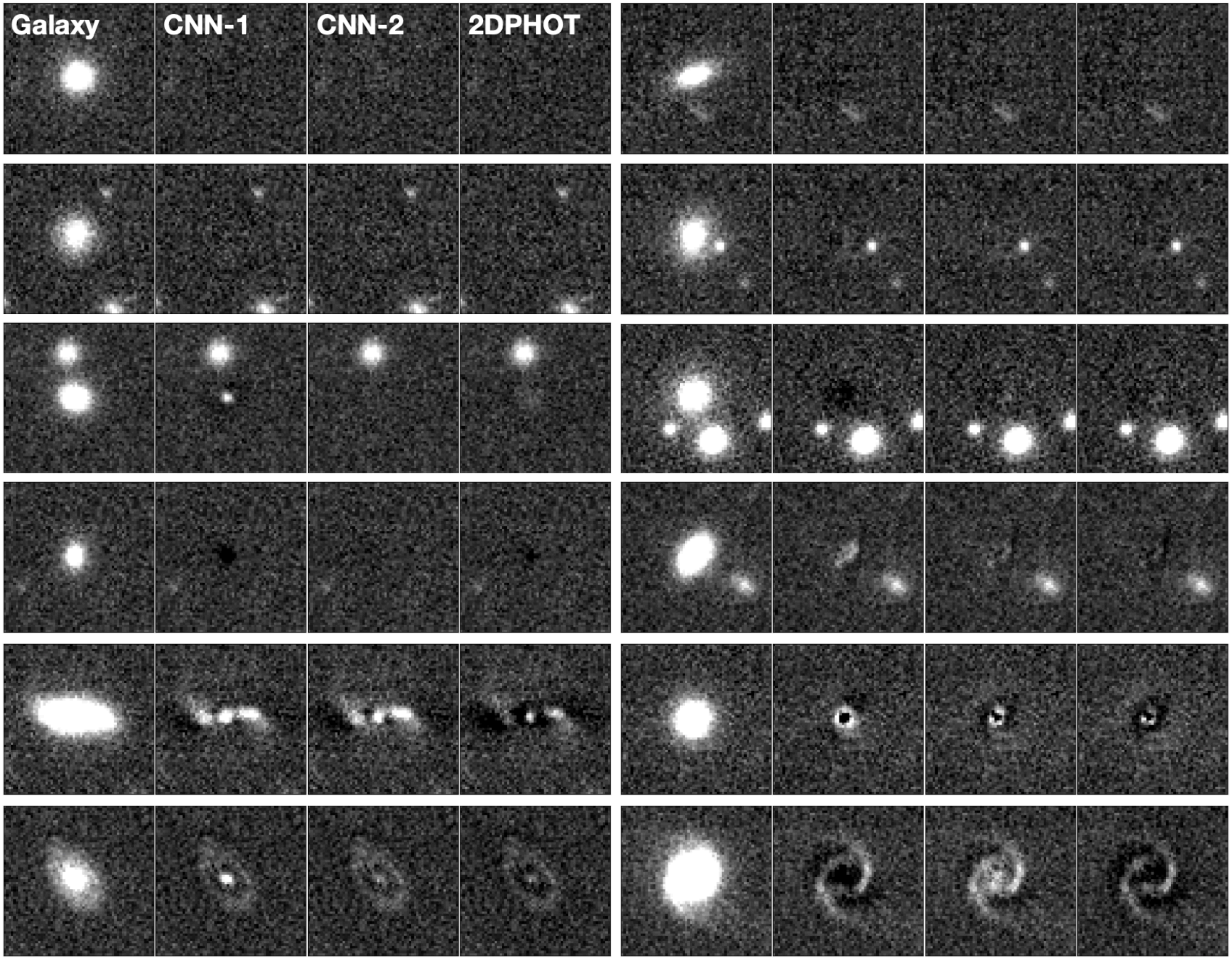
GaLNets: Convolutional Neural Networks for galaxy structure



Comparison of GaLNets vs Standard on KiDS “real galaxies”



Comparison of GaLNets vs Standard on KiDS “real galaxies”



Comparison of GaLNet-1 and GaLNet-2 vs Standard

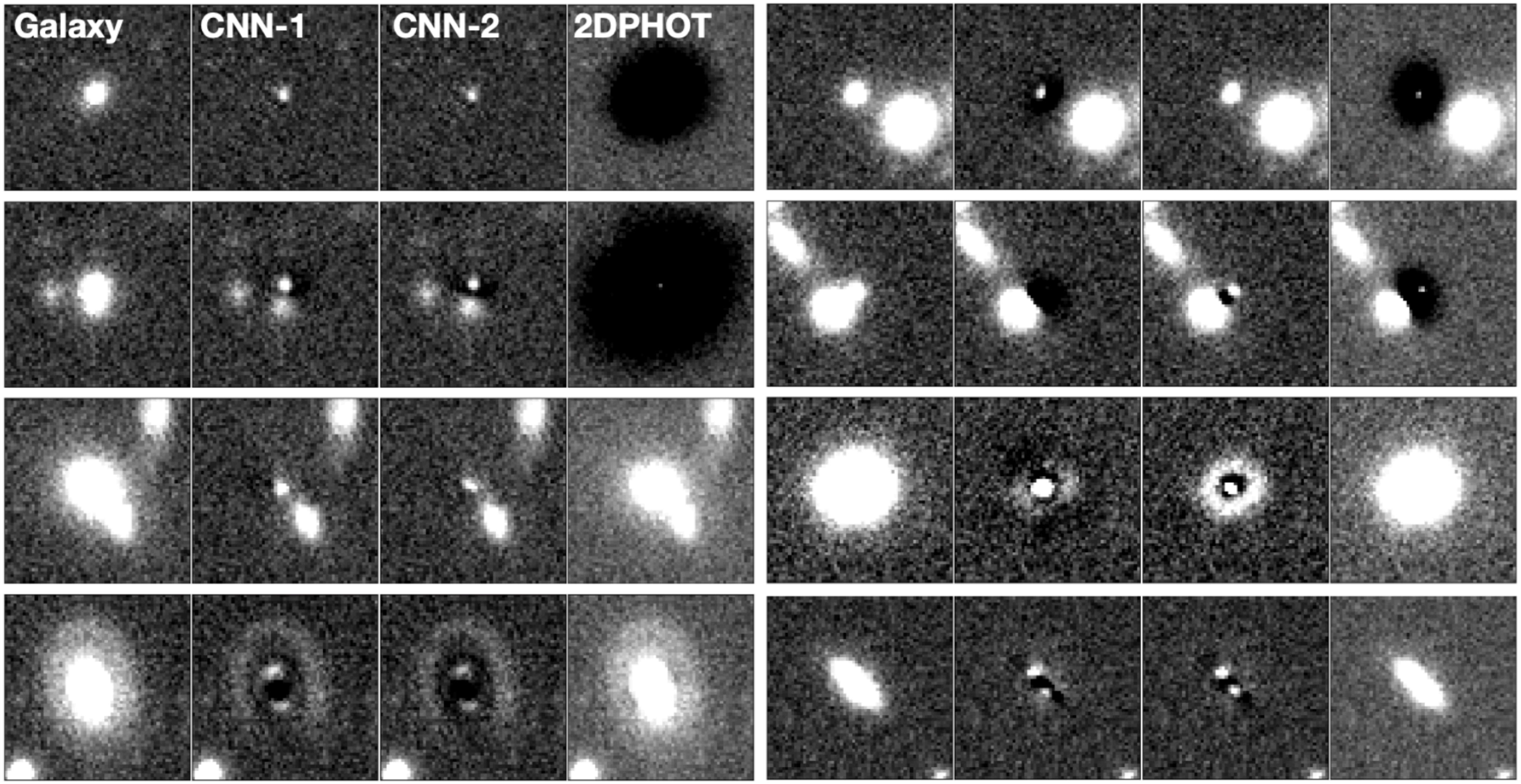
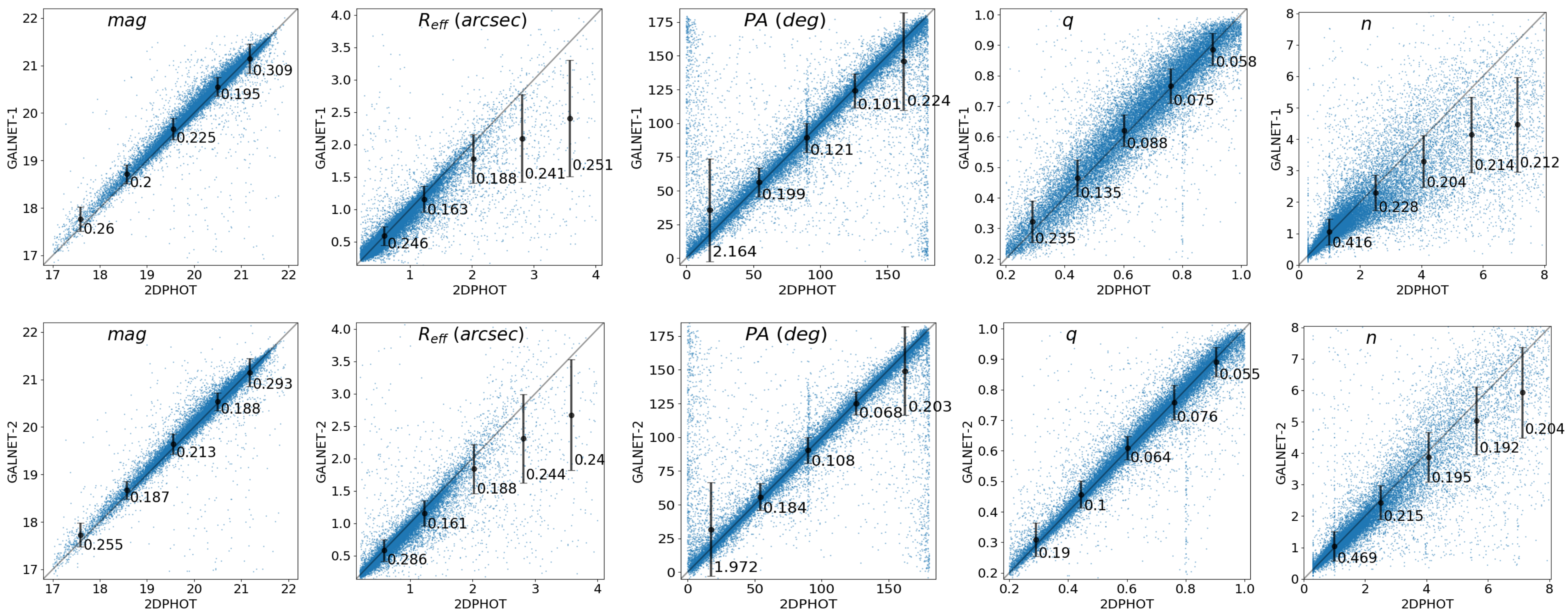
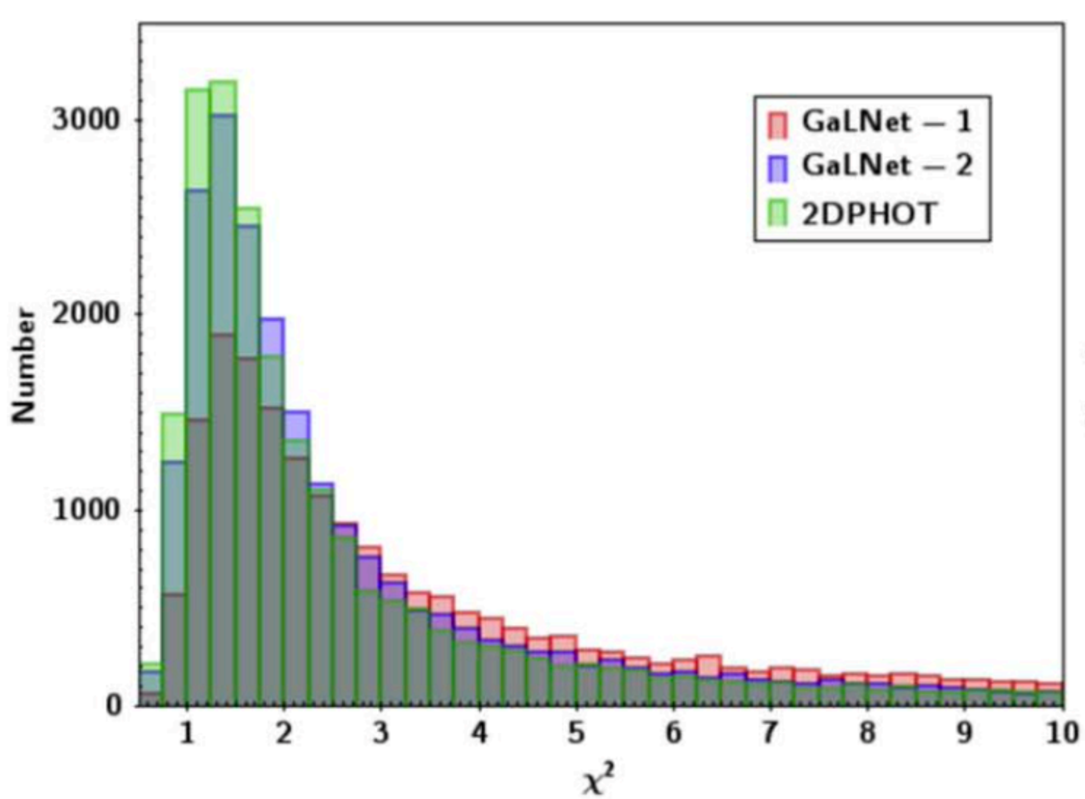


FIG. 6.— Real galaxies and their residuals Residuals. In each panel, A is the r−band image of real galaxy, B is the residual obtained from the CNN prediction, C is the residual obtained from 2DPHOT fitting.

Why CNN?
Performs as well as standard tool but it is much faster

1 million galaxies: from 1 month to 1 hour!
2DPHOT/Galfit: ~ 6s/galaxy
GaLNets with CPU: ~0.04s/galaxy
GaLNets with GPU (RTX2070): ~0.004s/galaxy



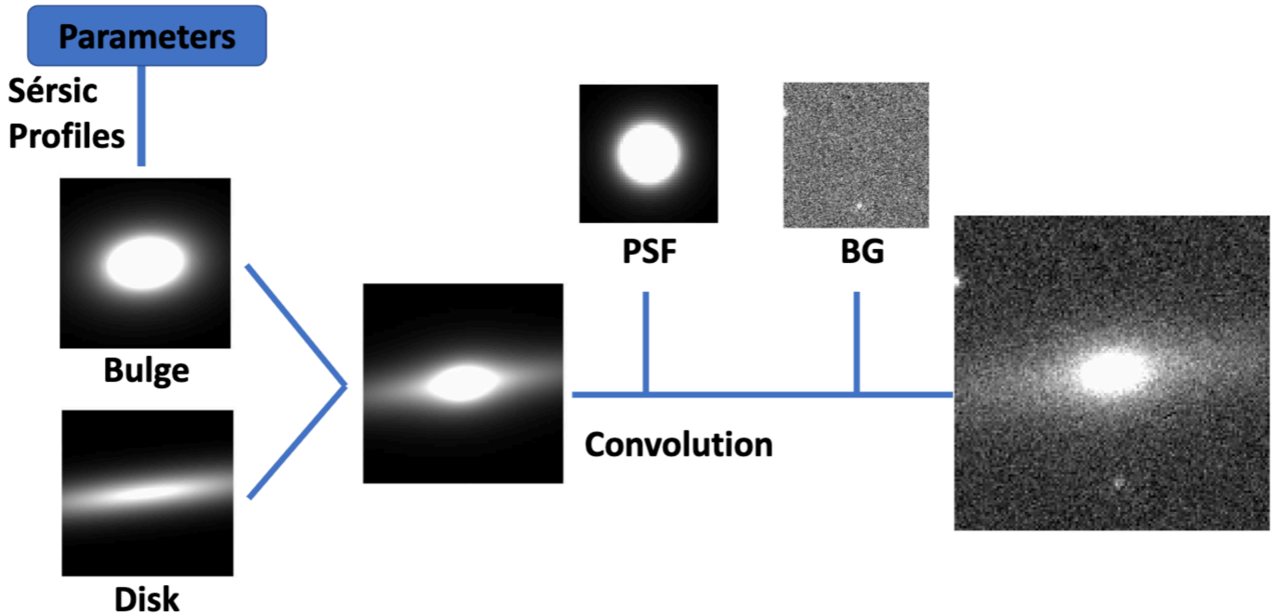


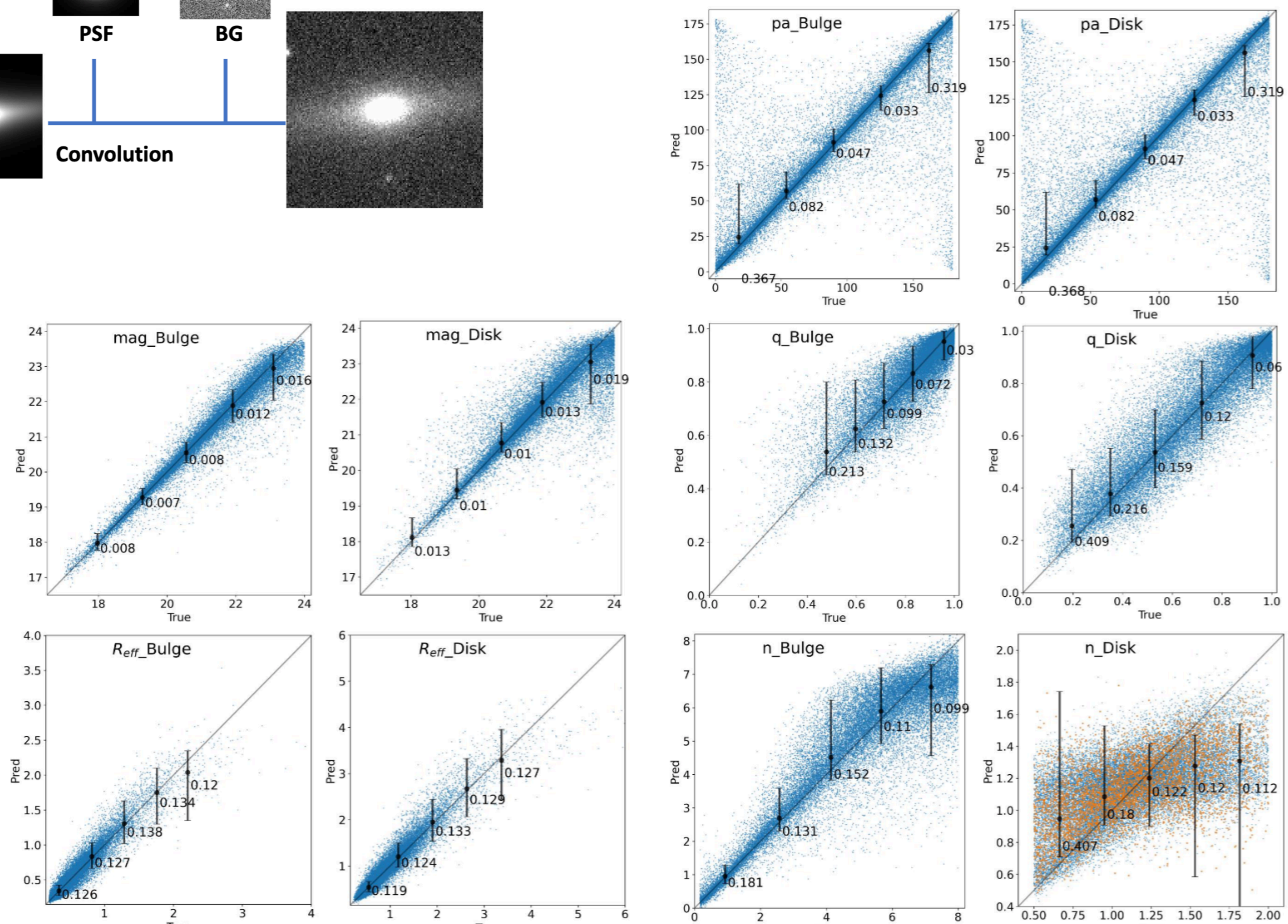
Advatages:

1. Unbiased parameters can be got by both GaLNet-1 and GaLNet-2.
2. GaLNet-2 performs general better than GaLNet-1.
3. PSFs affect more on q and n.
4. 2DPHOT performs the best, however, GaLNet-1 have a close performance.
5. GaLNets are 1000 times faster than traditional codes (e.g. 2DPHOT).

GaLNet for Bulge/Disk Decomposition

Qiu, Li, NRN, et al. submitted

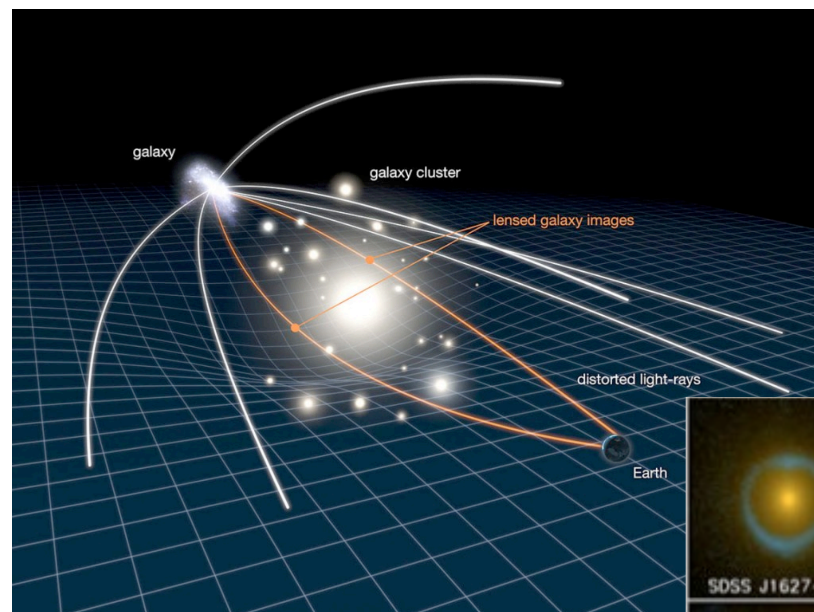




Strong Lensing with Machine learning

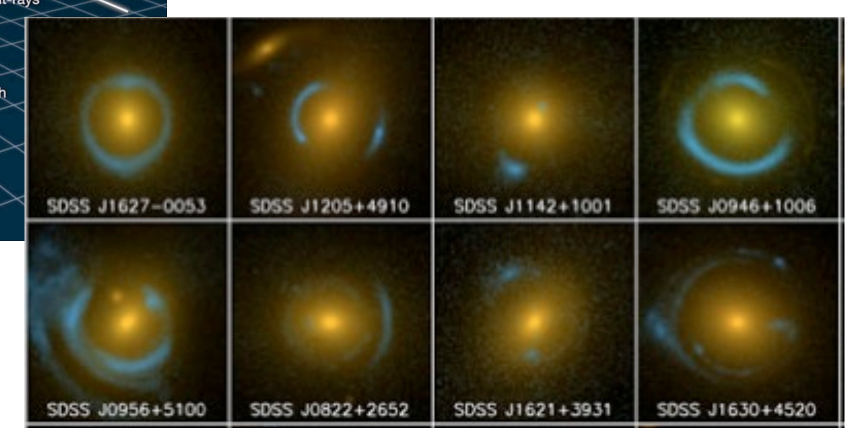
We have seen that EUCLID and CSST will provide up to 10^5 SGLs over ~1Billion of observed galaxies

To get prepared to this challenge we have used current high-quality ground-based surveys



KiDS@VST

- ~1arc/deg^2 or
- 0.1 lensed quasar/deg^2
- typically 120k source/deg2
- 40k galaxies/deg^2
- ~10% being ETGs with Mass>$10^{10.5}$M$_{sun}$
- ~5M candidates in 1350deg^2 (and ~1000 real lenses)



Strong Lensing with Machine learning

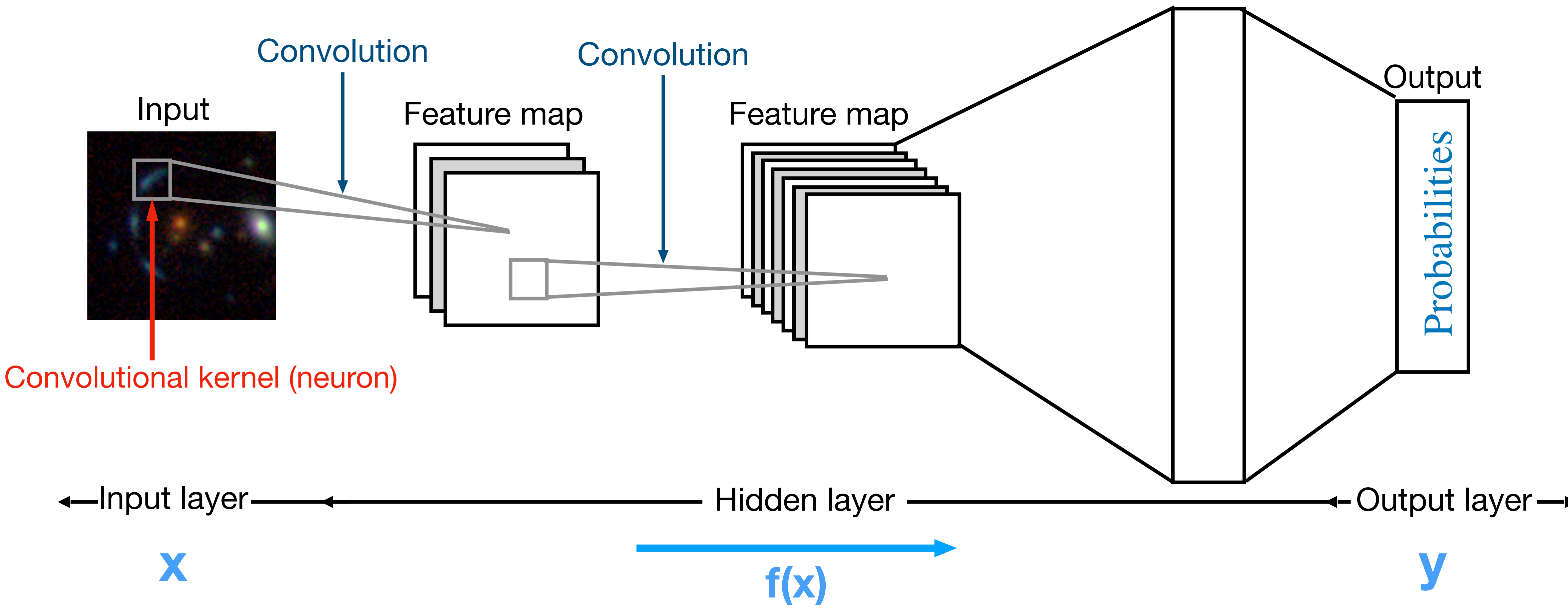
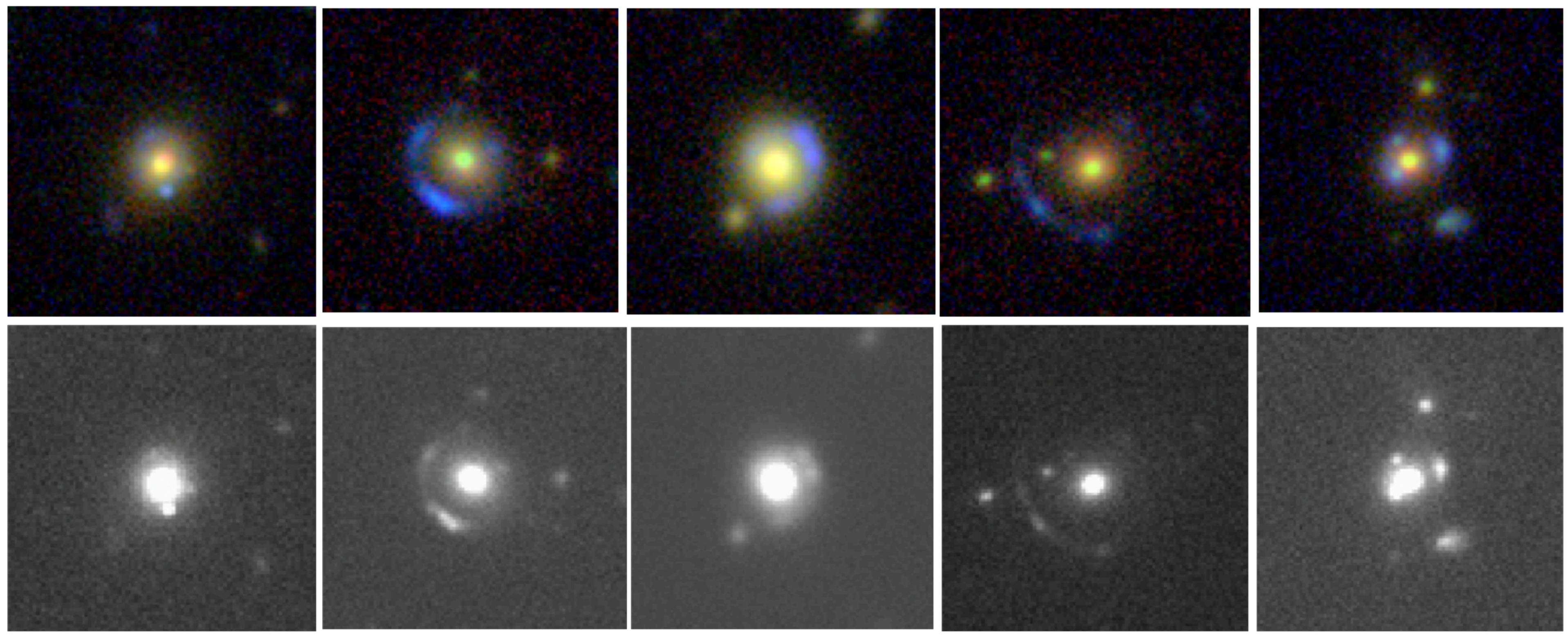


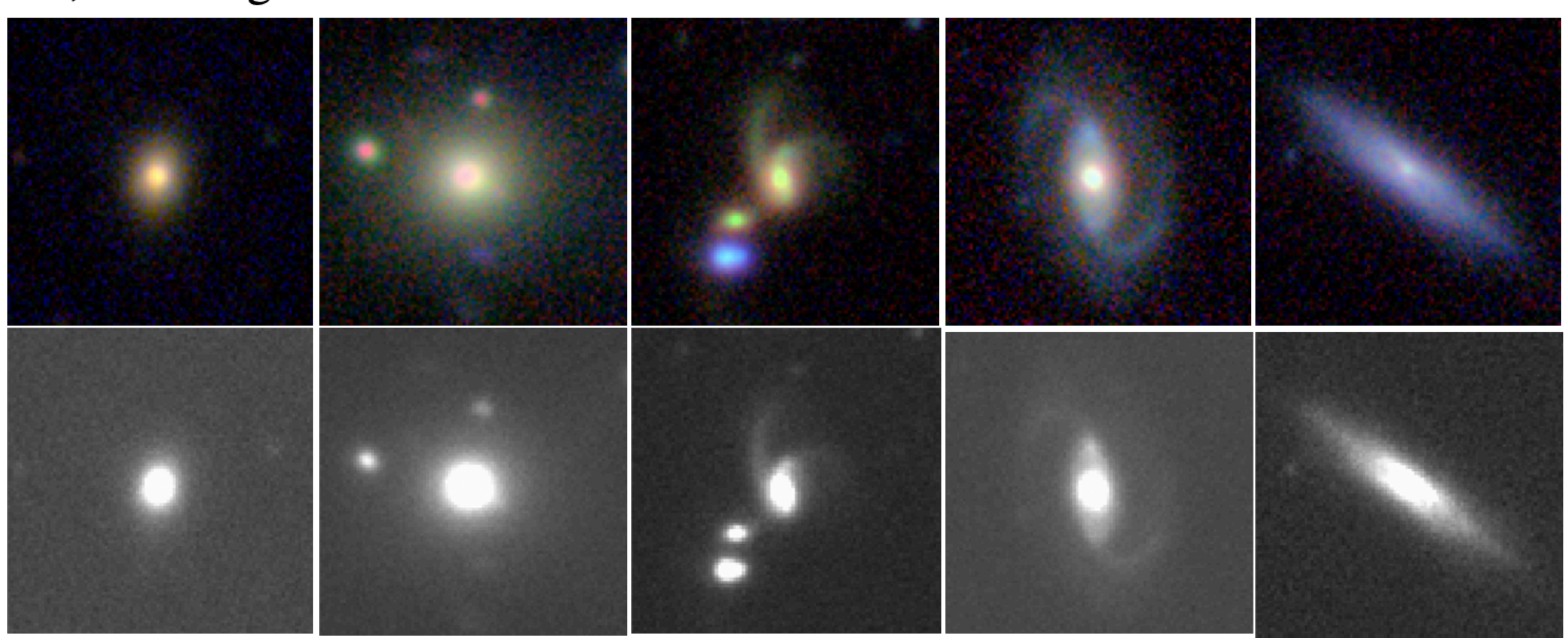
Fig.—A simple CNN architecture with input layer, hidden layer and outout layer. In the hidden layer, the CNN extract feature maps with convolution kernels.

Training data

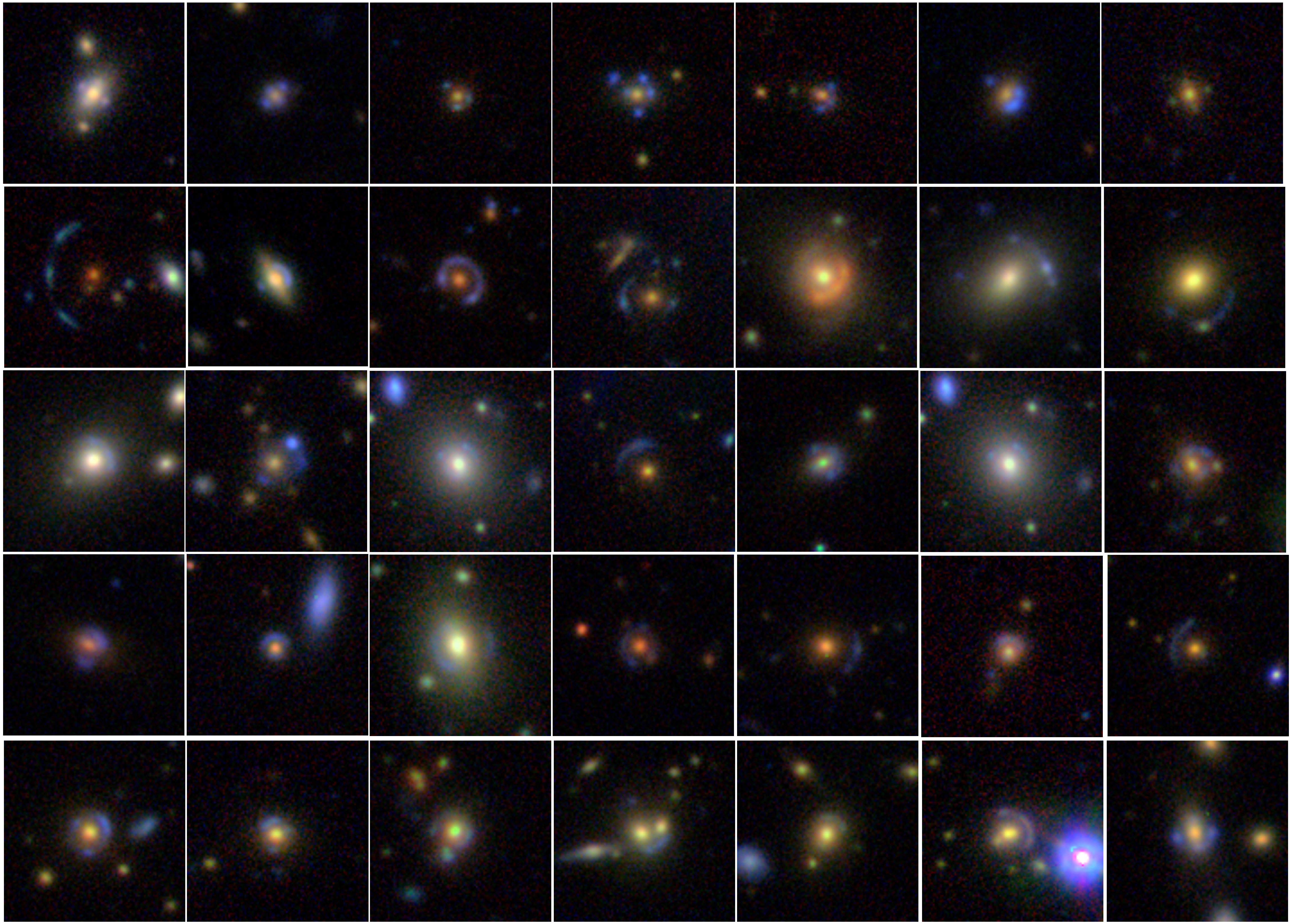
Positives: 45,000 simulated lenses (adding 45,000 simulated lenses).



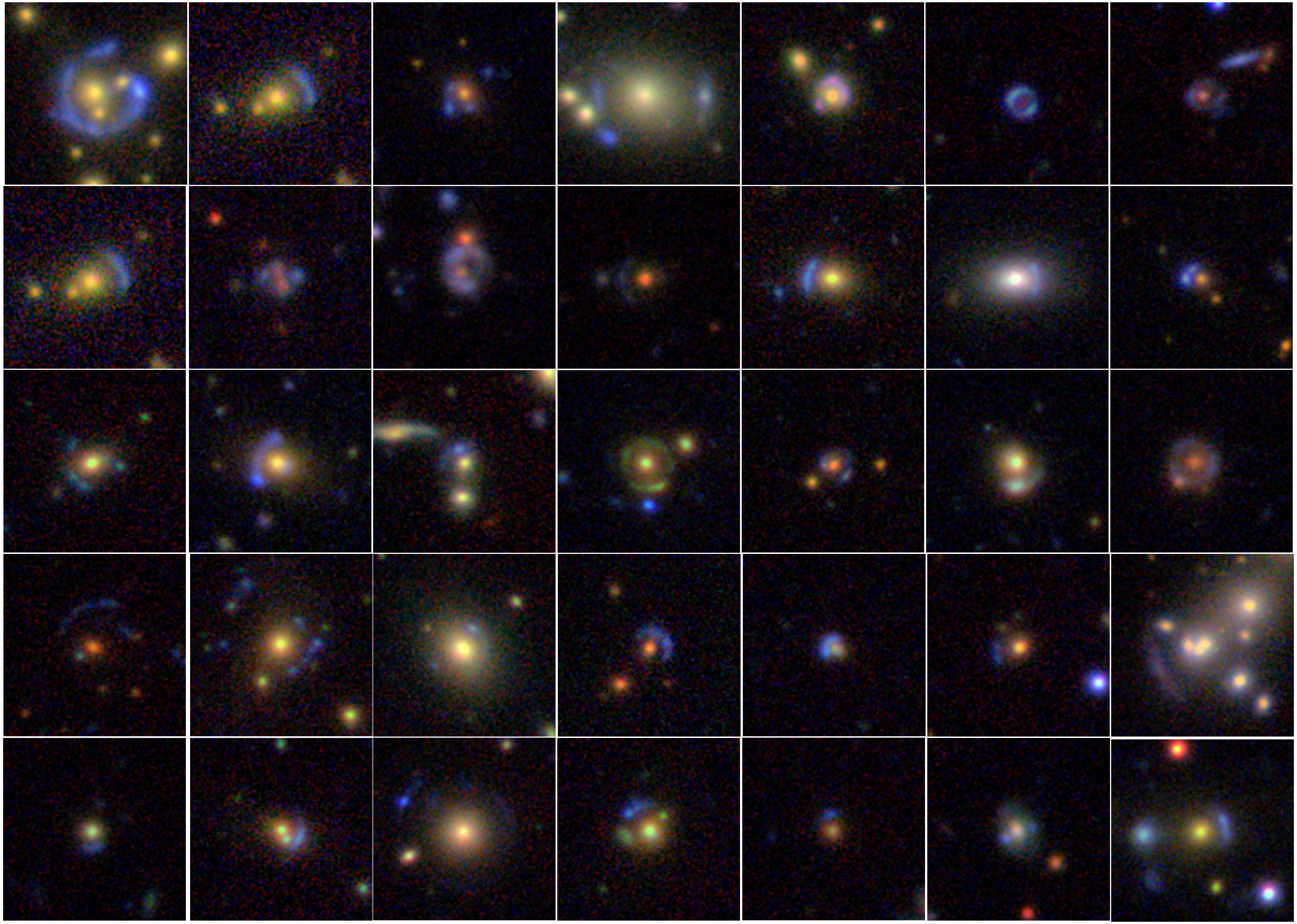
Negatives: 45,000 real galaxies.



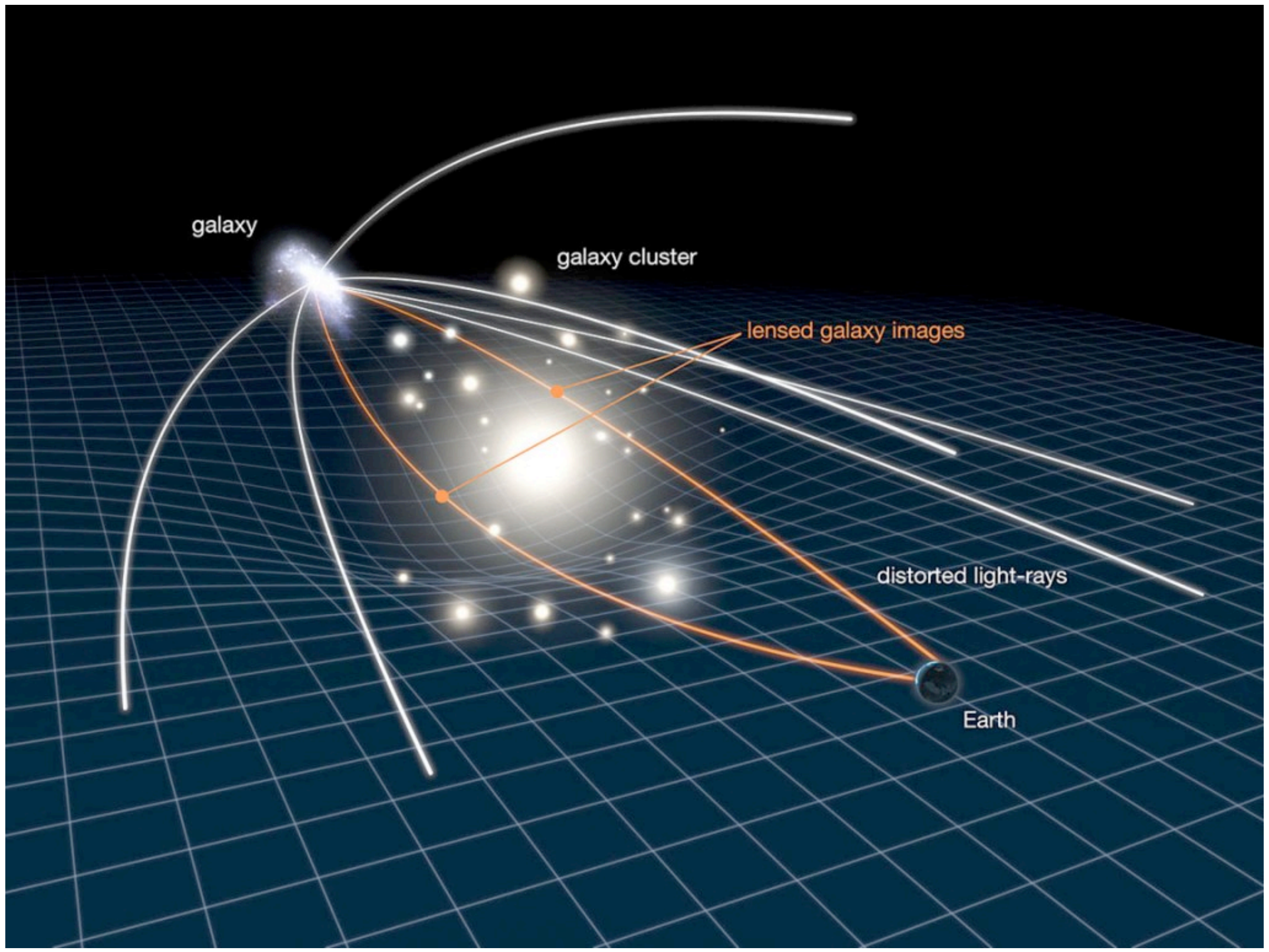
Samples in KiDS DR4 (From Petrillo 2019 and Li 2020)

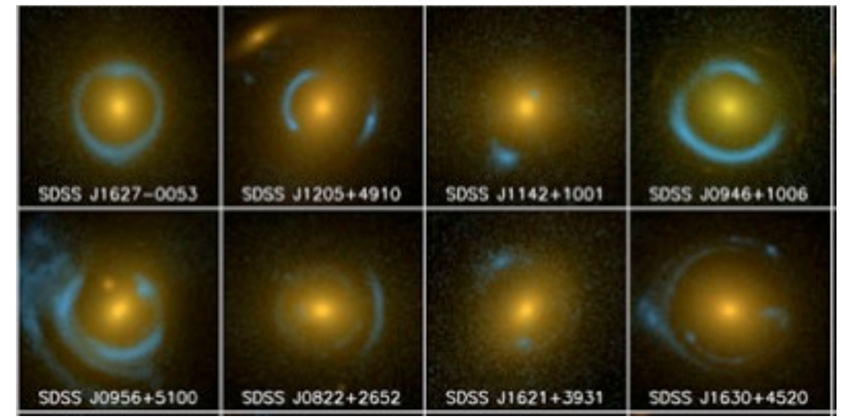


New HQ strong lenses in KiDS DR5 (Li, NRN et al. 2021, ApJ)



Modeling Gravitational lenses

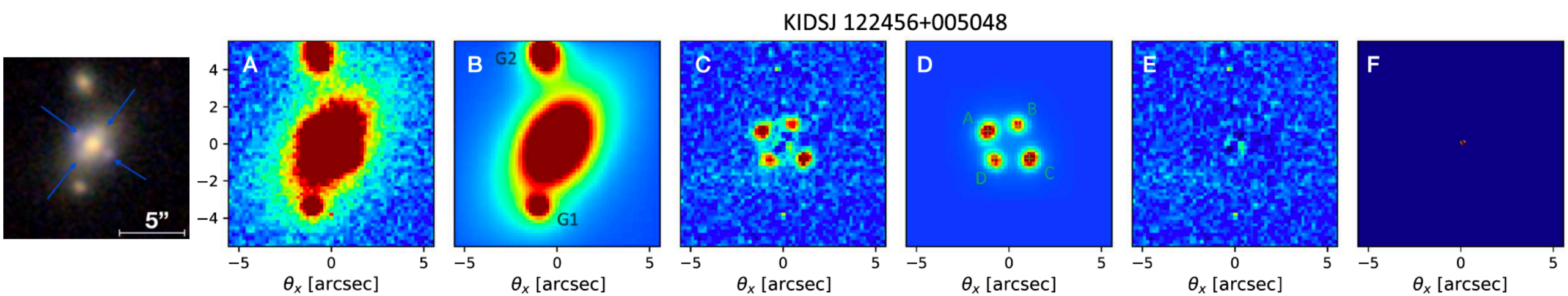




To model a lens one need to determine :

1) **The properties of the mass acting as lens**
2) The light of the lens
3) The position and light distribution of the source

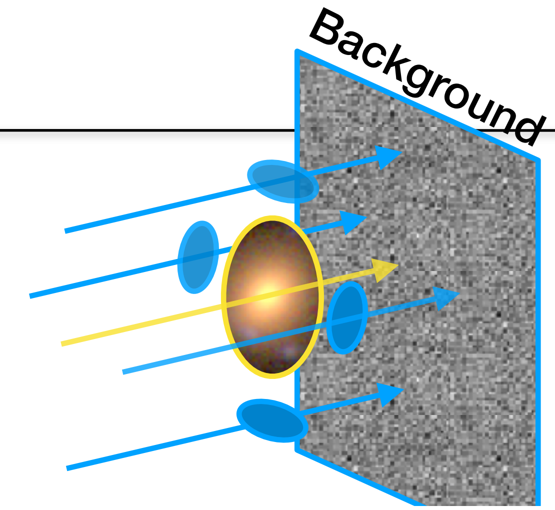
Typically 20 parameters



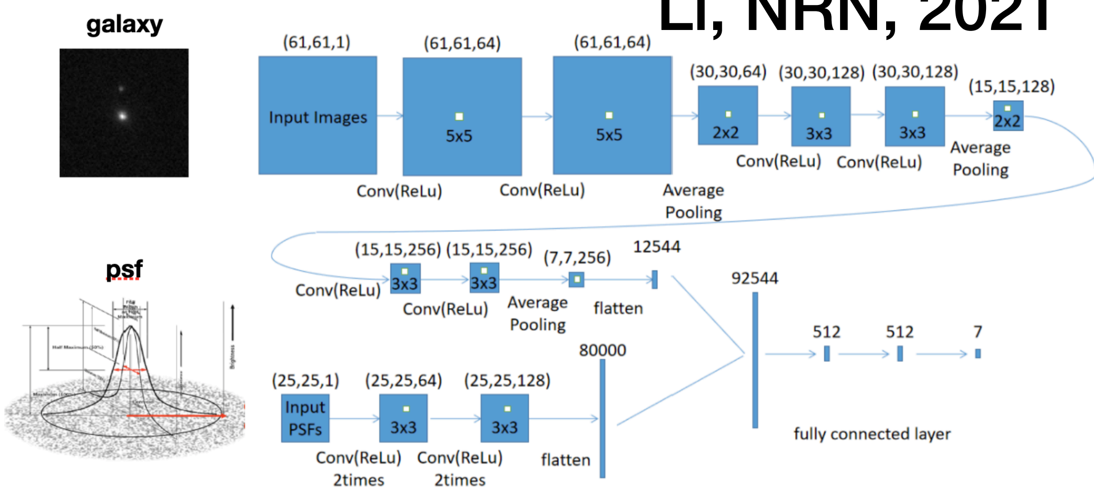
NRN et al. 2020, ApJL

Modeling Quads with Machine Learning

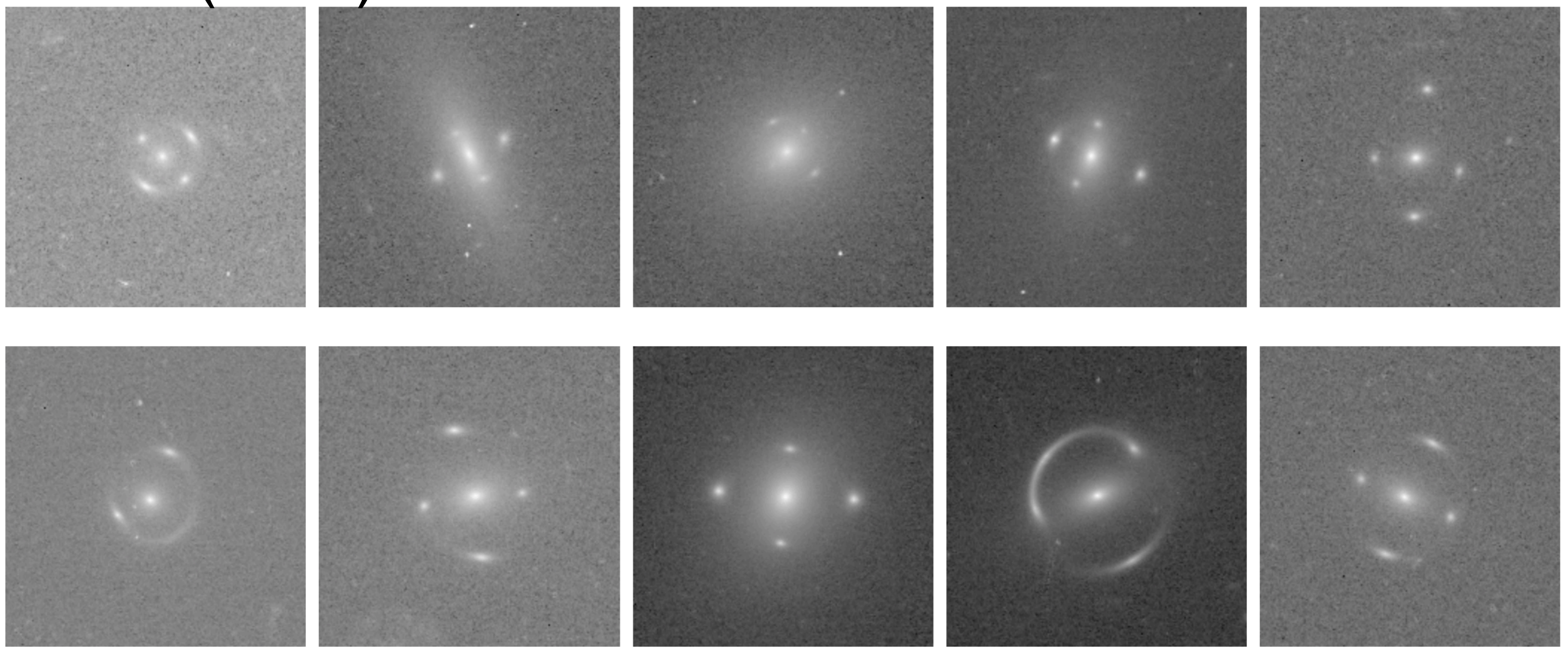
Zhu, Li, NRN et al. to be submitted



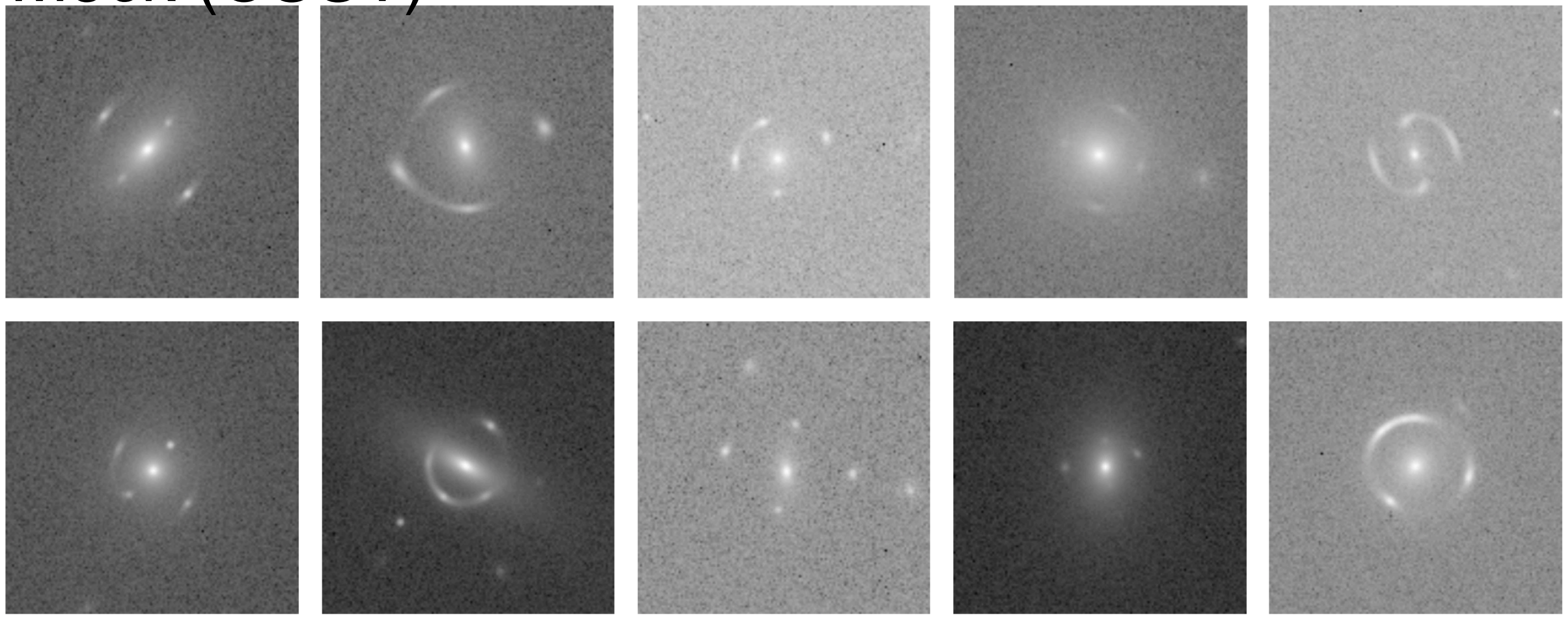
Li, NRN, 2021



Mock (HST)



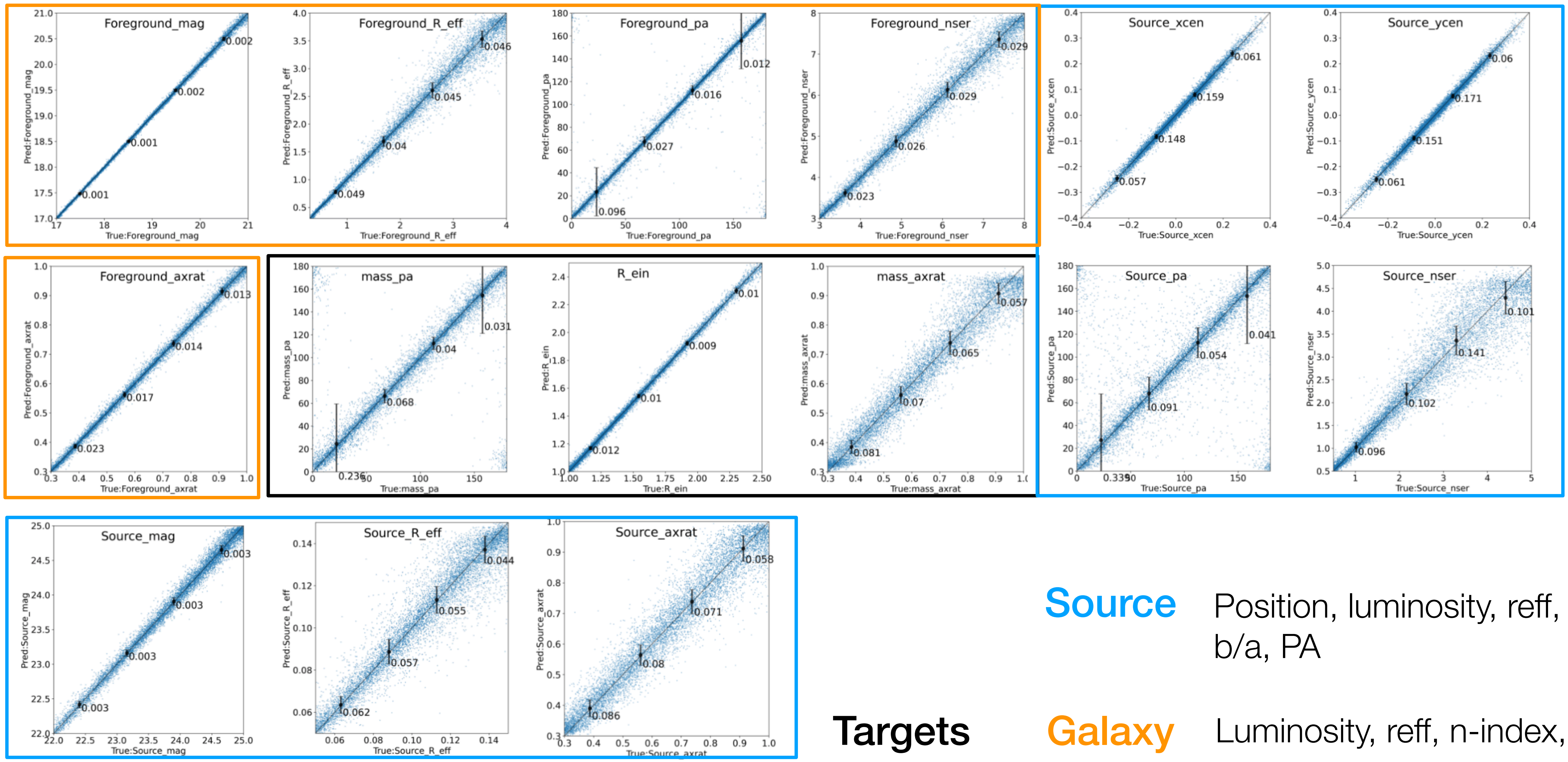
Mock (CSST)



Parameter	Range	Units	Distribution
		Foreground Galaxy (FG)	
x_{center}	-0.3~0.3	arcseconds	normal (μ = 0.0, σ = 0.15)
y_{center}	-0.3~0.3	arcseconds	normal (μ = 0.0, σ = 0.15)
mag	17~21		uniform
R_{eff}	0.3~4	arcseconds	uniform
pa	0.0~180.0	degrees	uniform
n	3~8		uniform
q	0.3~1		uniform
		Lens mass	
x_{center}			= x_{center} of FG
y_{center}			= y_{center} of FG
pa			(pa of FG)±20 (uniformly)
R_{ein}	1~2.5		exponential (s=1)
q	0.3-1		uniform
		Source Galaxy	
x_{center}			= x_{center} of FG±0.2 (uniformly)
y_{center}			= y_{center} of FG±0.2 (uniformly)
mag	22~25		uniform
R_{eff}	0.05~0.15	arcseconds	uniform
pa	0.0~180.0	degrees	uniform
n	0.5~5		uniform
q	0.3~1		uniform

Modeling Quads with Machine Learning

Results (HST) 200k training 10k test

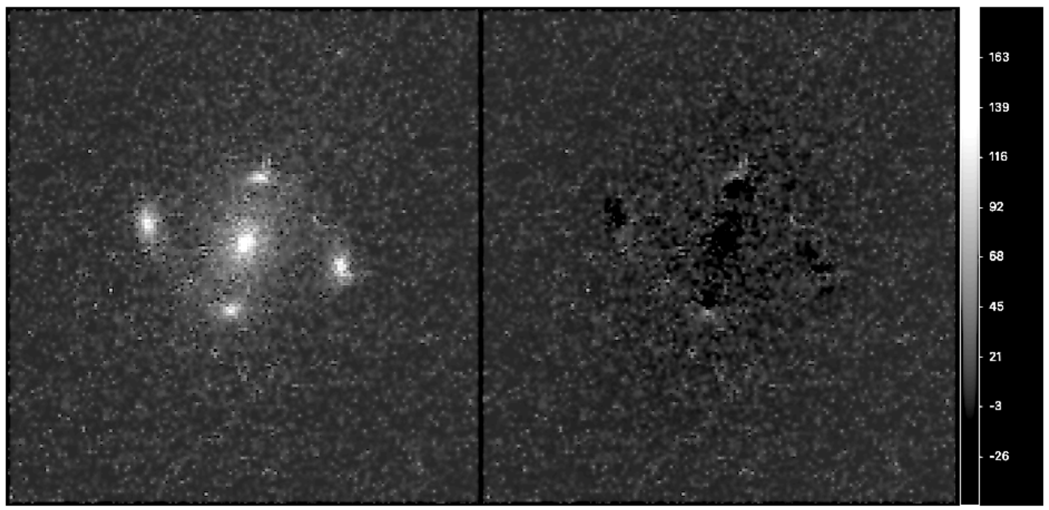


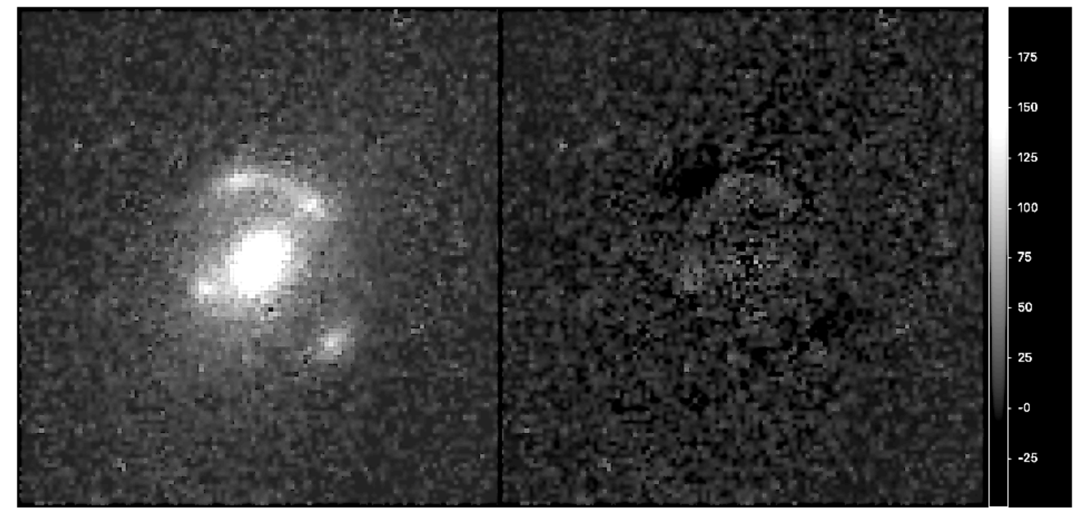
Targets

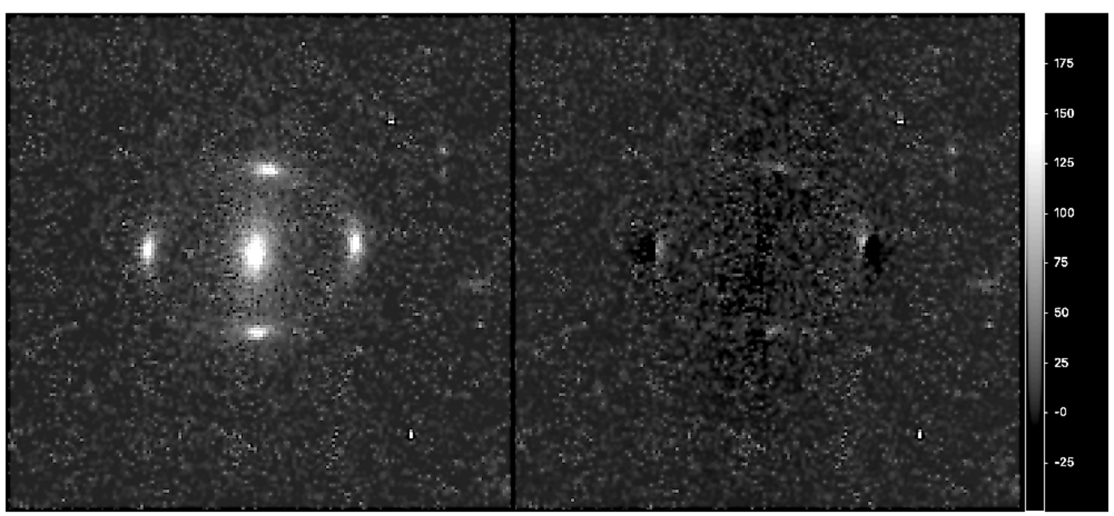
Source	Position, luminosity, reff, n-index, b/a, PA
Galaxy	Luminosity, reff, n-index, b/a, PA
Mass	Einst. rad., b/a, PA

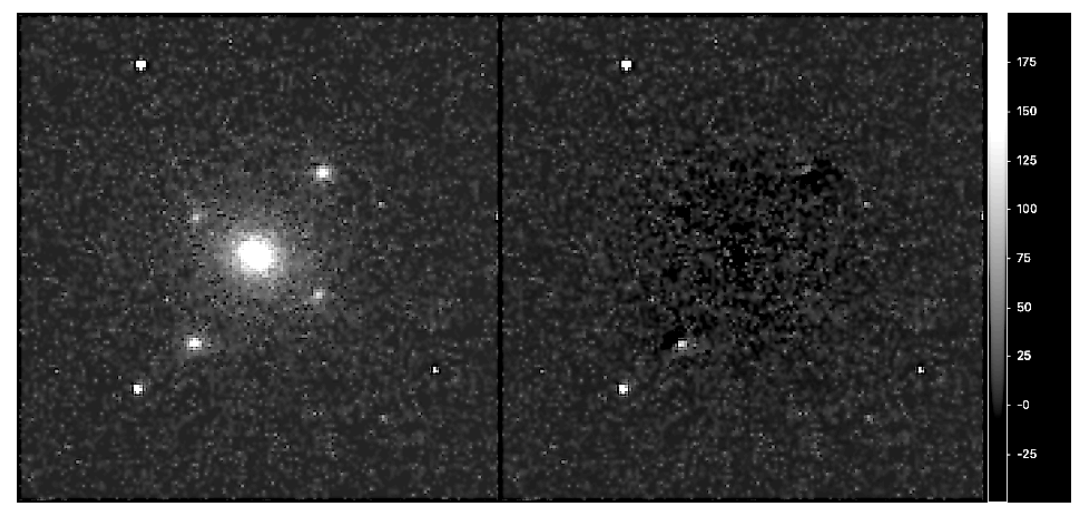
Modeling Quads with Machine Learning

Residual Maps









What About Science?

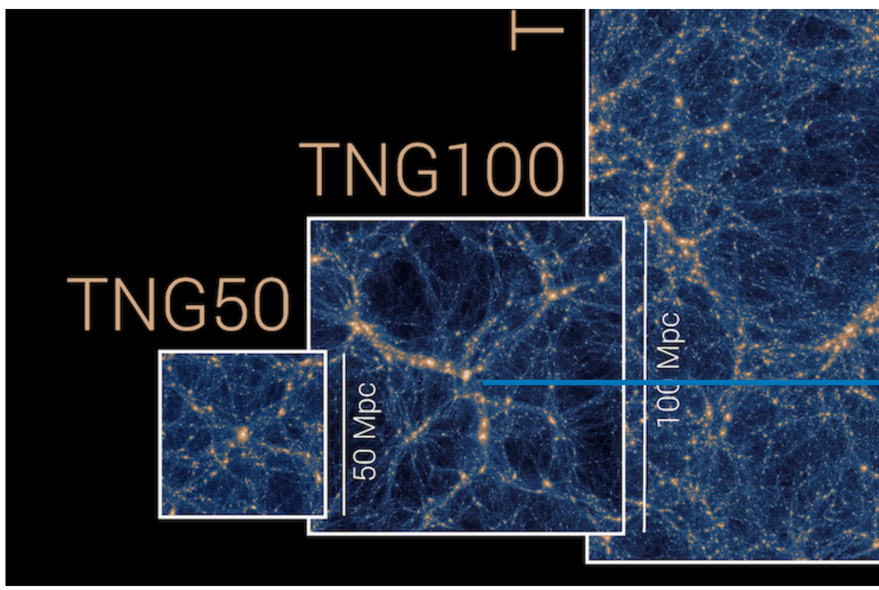
Galaxy Mass Estimate machine Learning Algorithm (MELA)

A&A, 686, A80 (2024)
https://doi.org/10.1051/0004-6361/202348152

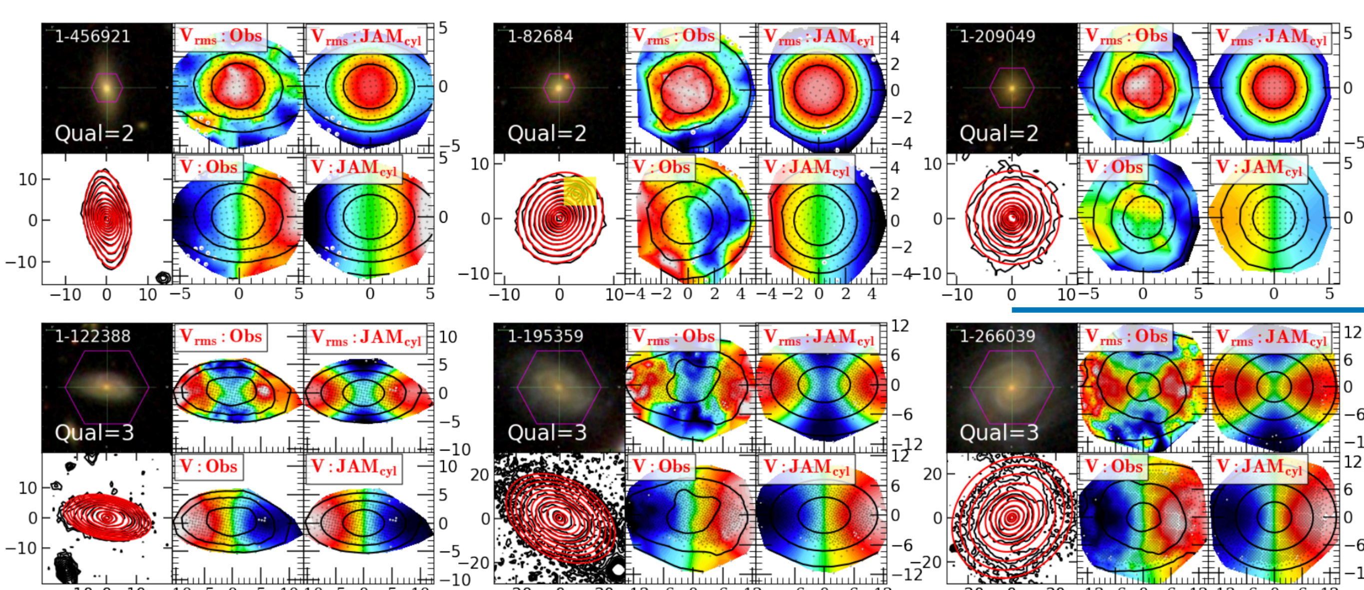
Astronomy & Astrophysics

Total and dark mass from observations of galaxy centers with machine learning

Sirui Wu[1,2], Nicola R. Napolitano[1,2,3], Crescenzo Tortora[4], Rodrigo von Marttens[5,6], Luciano Casarini[7], Rui Li[8,9], and Weipeng Lin[1,2]



Train/Validation/Test



Predictive sample

MaNGA-Dynpop: Zhu et al. 2023,

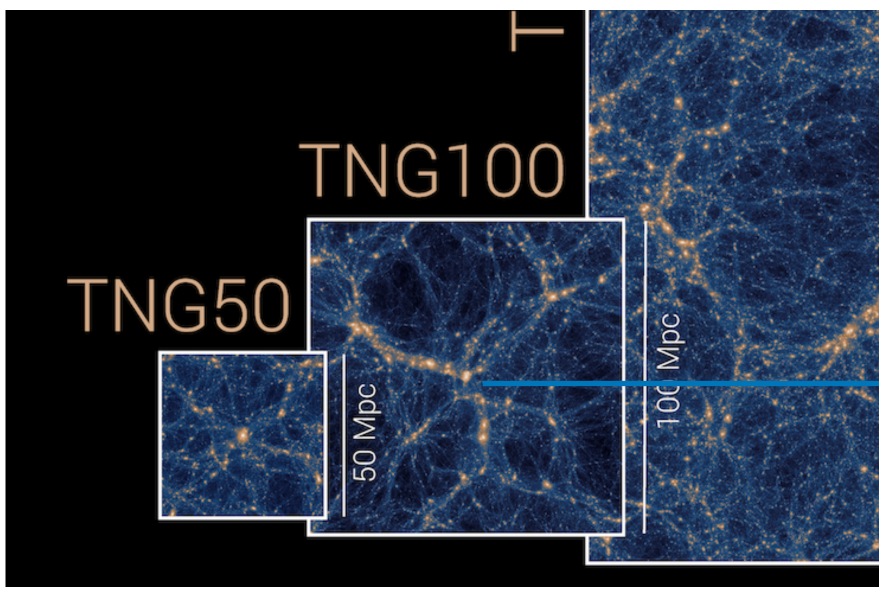
Galaxy Mass Estimate machine Learning Algorithm (MELA)

A&A, 686, A80 (2024)
https://doi.org/10.1051/0004-6361/202348152

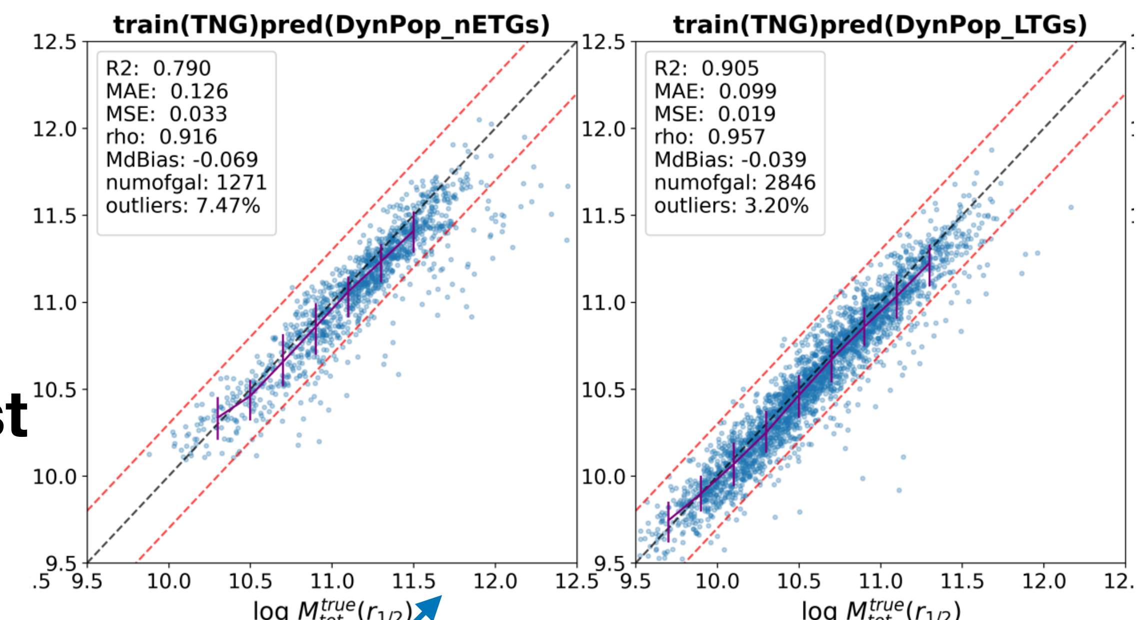
Astronomy & Astrophysics

Total and dark mass from observations of galaxy centers with machine learning

Sirui Wu[1,2], Nicola R. Napolitano[1,2,3], Crescenzo Tortora[4], Rodrigo von Marttens[5,6], Luciano Casarini[7], Rui Li[8,9], and Weipeng Lin[1,2]



Train/Validation/Test

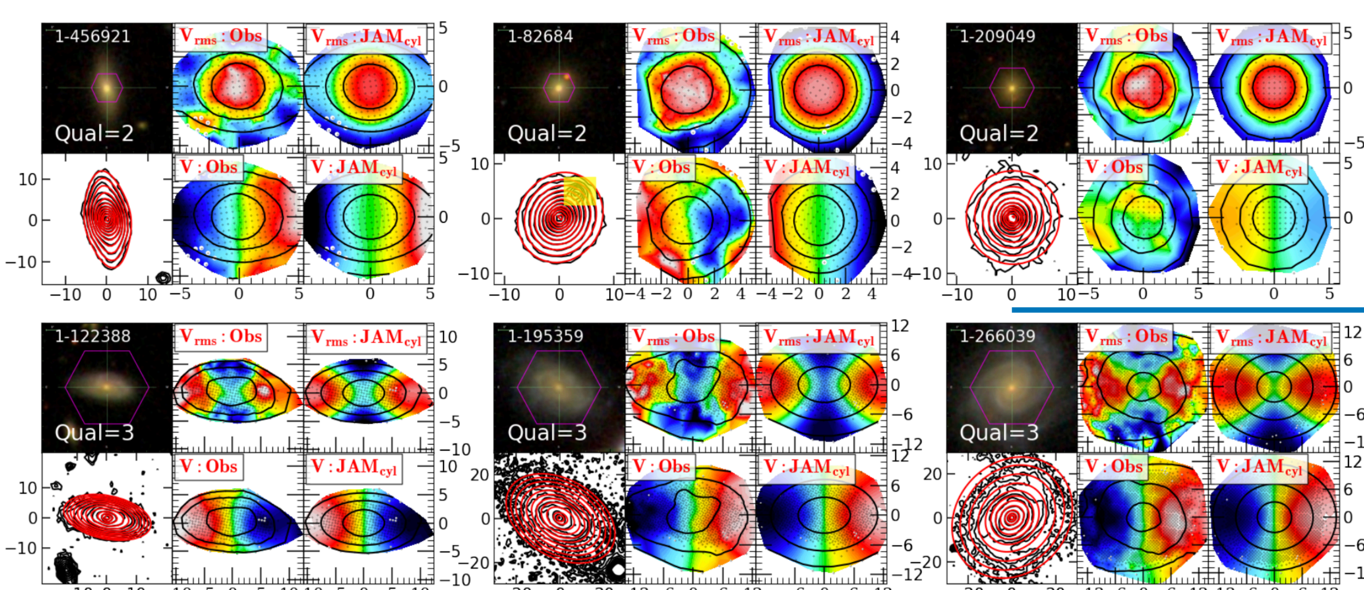


Important Features

effective redii

stellar masses

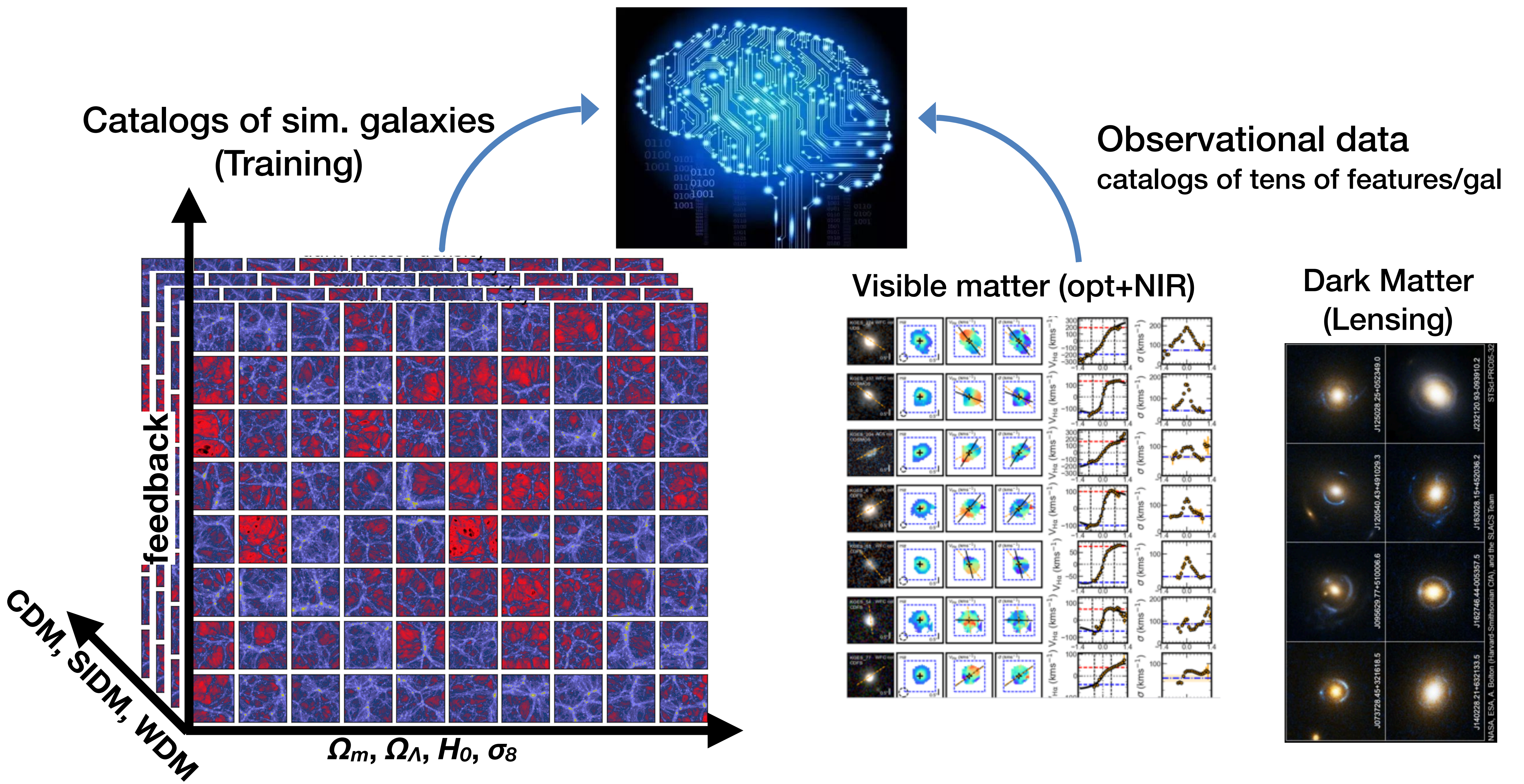
velocity dispersion

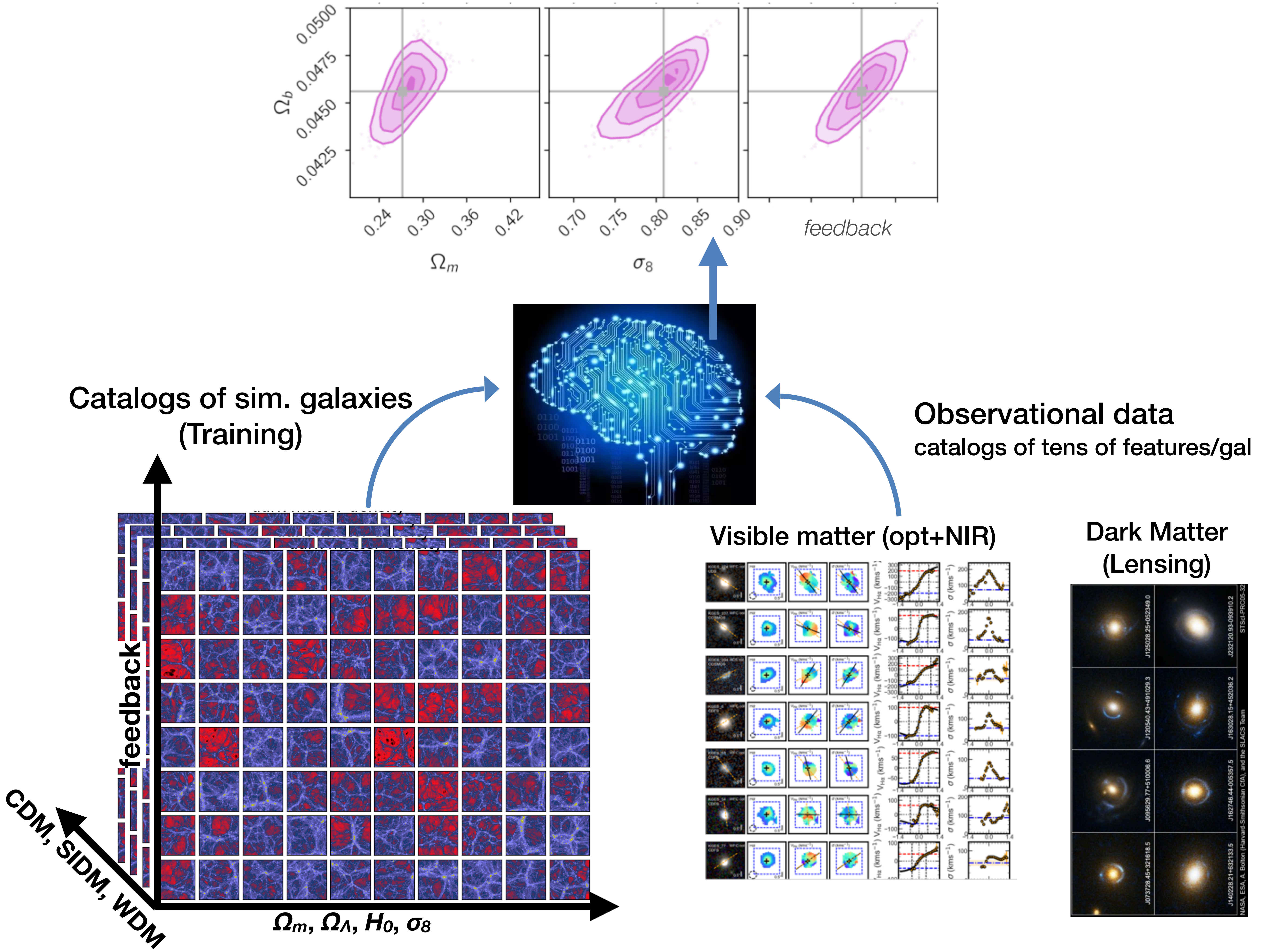


Predictive sample

MaNGA-Dynpop: Zhu et al. 2023,

Using Machine Learning to match simulations and data and predict cosmology (see CAMELS and DREAMS — Villaescusa-Navarro talk)

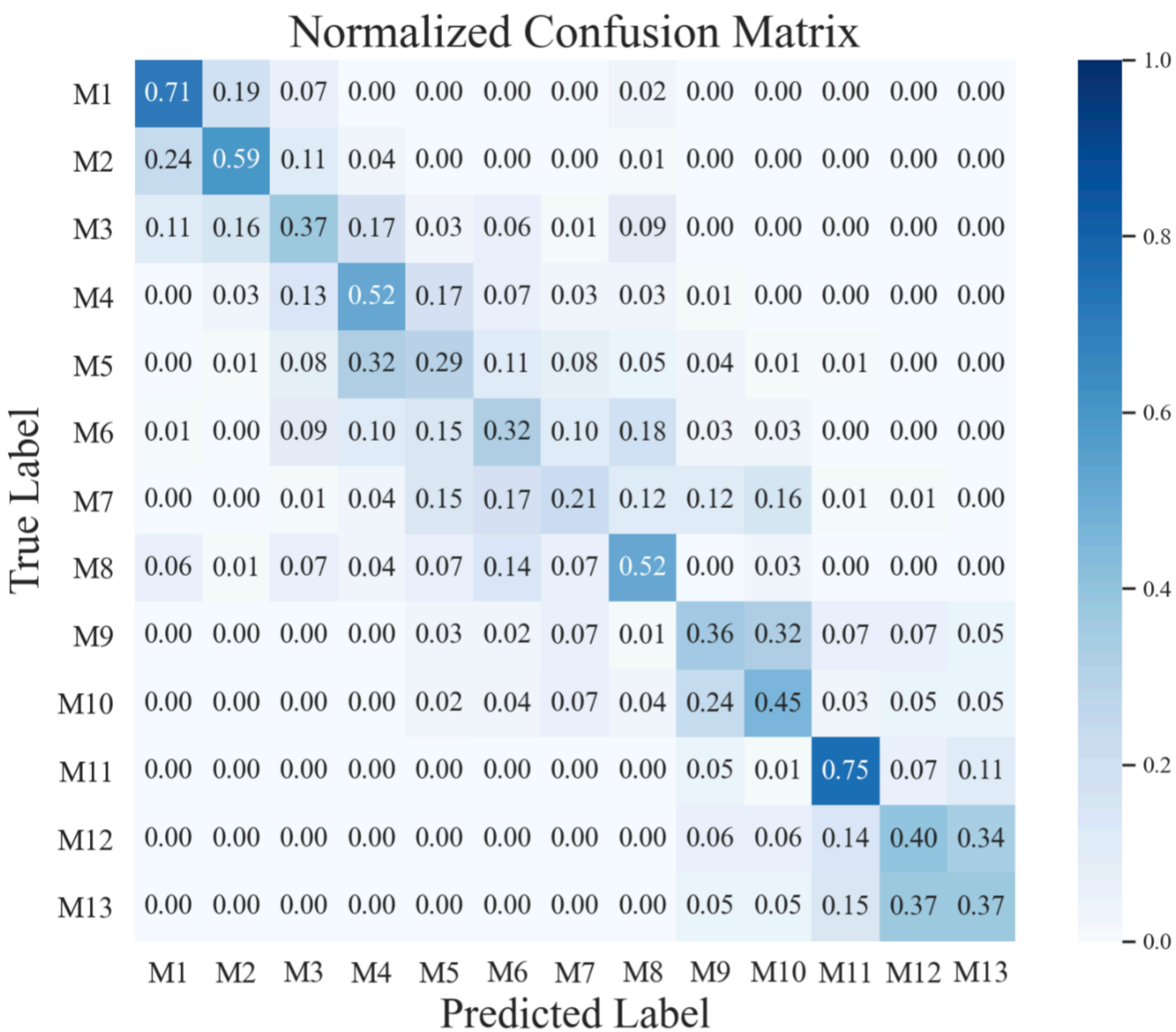




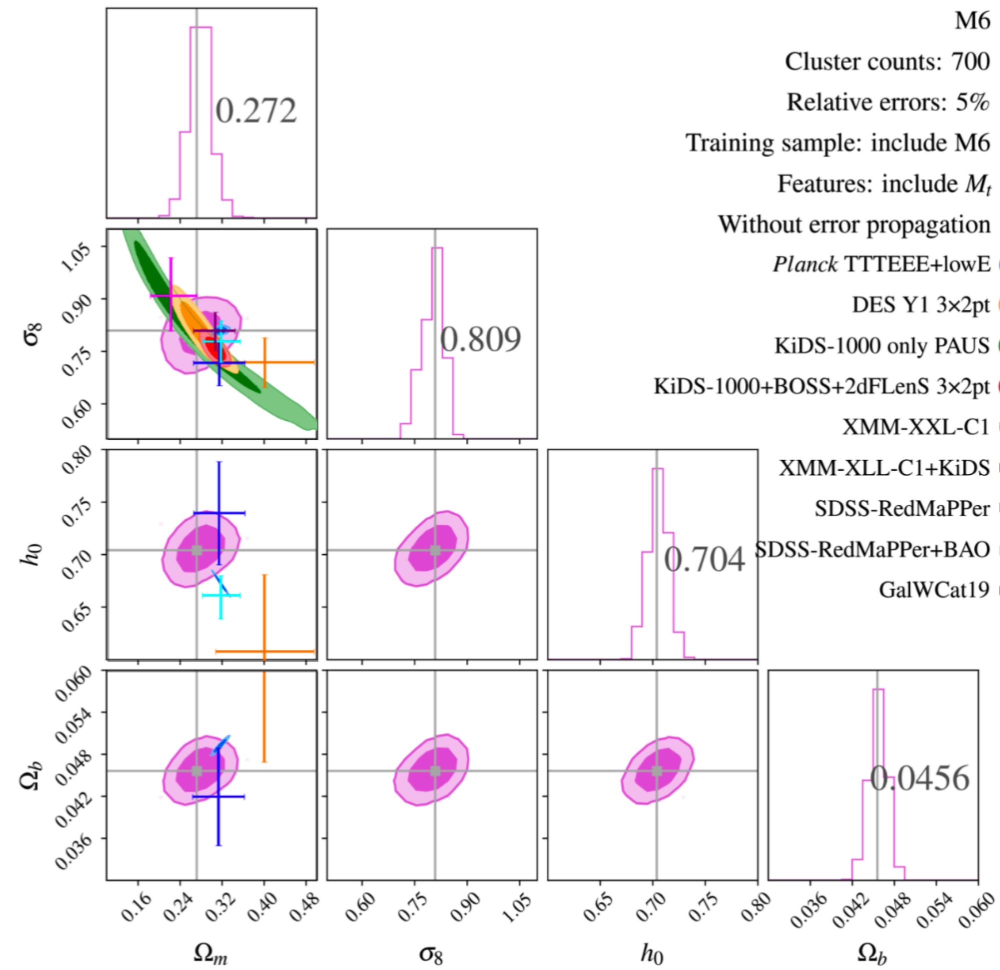
Ω_b
0.0500
0.0475
0.0450
0.0425
0.24
0.30
0.36
0.42
Ω_m
0.70
0.75
0.80
0.85
0.90
σ_8
feedback
Catalogs of sim. galaxies
(Training)
Observational data
catalogs of tens of features/gal
Visible matter (opt+NIR)
Dark Matter
(Lensing)
feedback
CDM, SIDM, WDM
Ω_m, Ω_Λ, H_0, σ_8

Galaxy Clusters for cosmology using ML

How well the classifier is able to recognise a given mock catalog to come from the right cosmology



Cosmological parameter inferences



Important Features
stellar mass
gas mass
total mass

Qiu, L., NRN et al. 2024, A&A, arXiv:2304.09142

Take away messages

Machine/Deep Learning are groundbreaking techniques for large surveys

Data are coming and it is time to move from proof-of-concept to real data applications

Publications

1. Qiu C., **NRN**, Li R., et al., **Galaxy Light profile neural Networks (GaLNets). II. Bulge-Disc decomposition in optical space-based observations**, 2023, ApJ subm, arXiv:2306.05909
2. Qiu L., **NRN**, Borgani S., et al., **Cosmology with Galaxy Cluster Properties using Machine Learning**, 2024, A&A, arXiv e-prints, arXiv:2304.09142
3. Wu, S., **NRN**, Tortora, C., et al., **Total and dark mass from observations of galaxy centers with machine learning,** 2024, A&A, 686, 80
4. Zhong F., **NRN**, et al. **Galaxy Spectra Neural Networks (GaSNets). II,** 2024, MNRAS, 532, 643
5. Li R., **NRN**, Feng H., et al., **Galaxy morphoto-Z with neural Networks (GaZNets). I. Optimized accuracy and outlier fraction from imaging and photometry**, 2022, A&A, 666, 85
6. Zhong F., Li R., **NRN**, **Galaxy Spectra Neural Networks (GaSNets). I. Searching for Strong Lens Candidates in eBOSS Spectra Using Deep Learning**, 2022, RAA, 22, 065014
7. Li R., **NRN**, Roy N., et al., **Galaxy Light Profile Convolutional Neural Networks (GaLNets). I. Fast and Accurate Structural Parameters for Billion-galaxy Samples**, 2022, ApJ, 929, 152
8. Li R., **NRN**, Spiniello C., et al., **High-quality Strong Lens Candidates in the Final Kilo-Degree Survey Footprint**, 2021, ApJ, 923, 16

…others in preparation/ArXiv