Machine Learning and Galaxies

John F Wu STScI · JHU

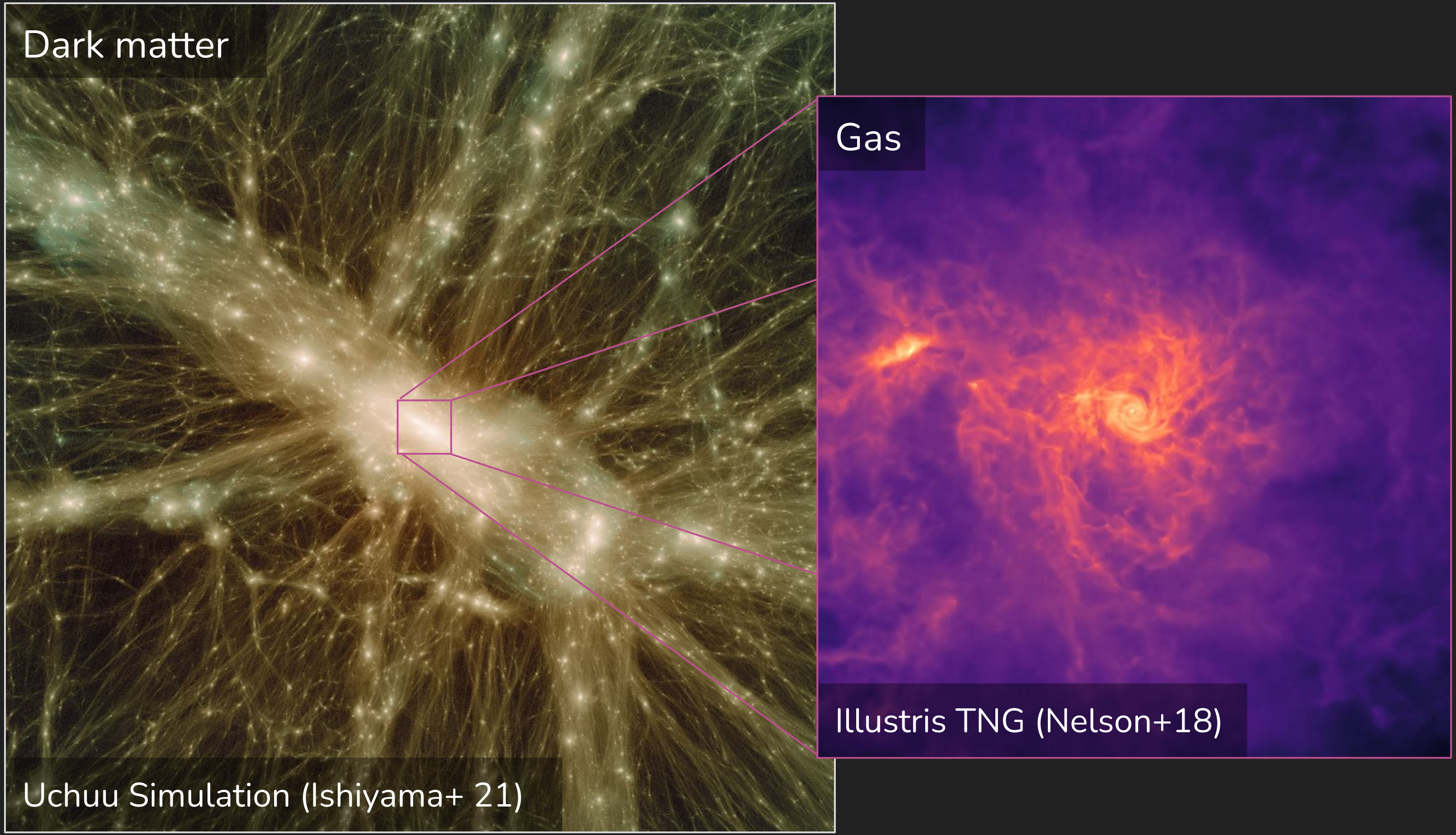
Roadmap

I. The growth and evolution of galaxies
II. Studying satellite galaxies with CNNs
III. Learning relationships with graphs

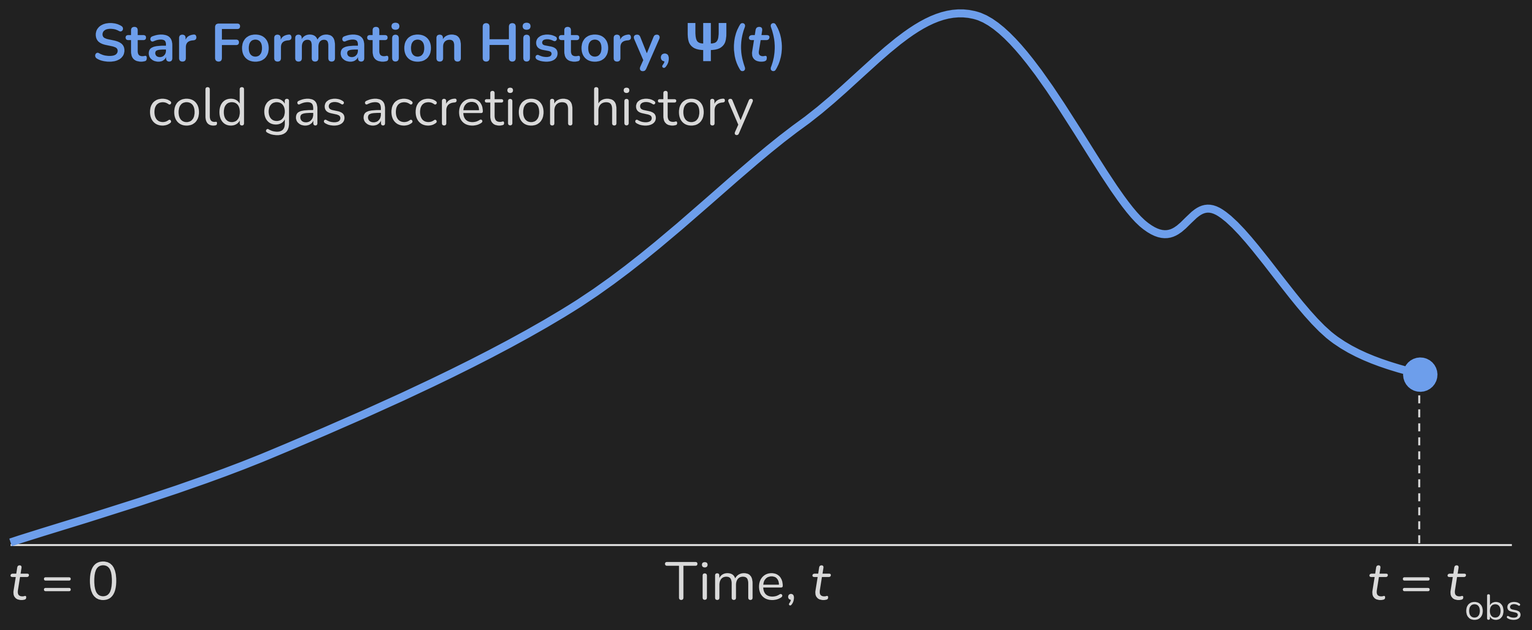
Roadmap

I. **The growth and evolution of galaxies**
II. Studying satellite galaxies with CNNs
III. Learning relationships with graphs

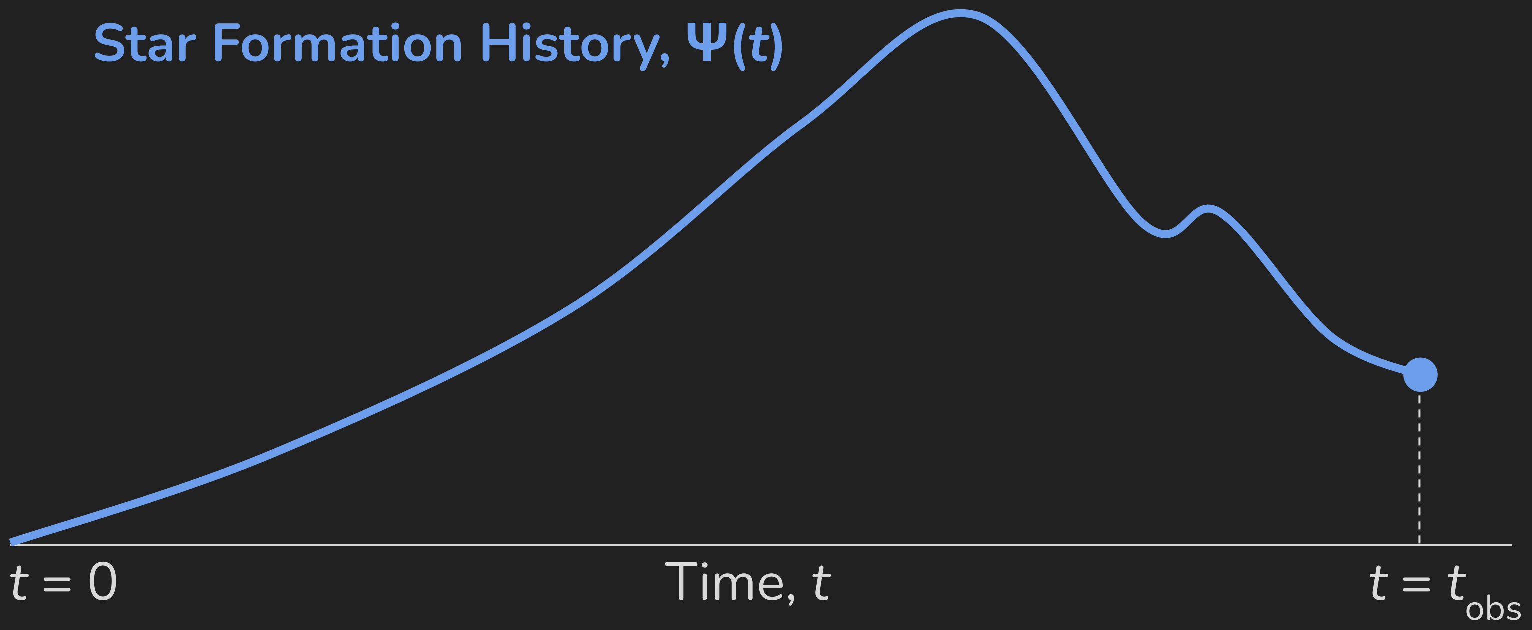
Galaxies grow via gas accretion, star formation, and merging



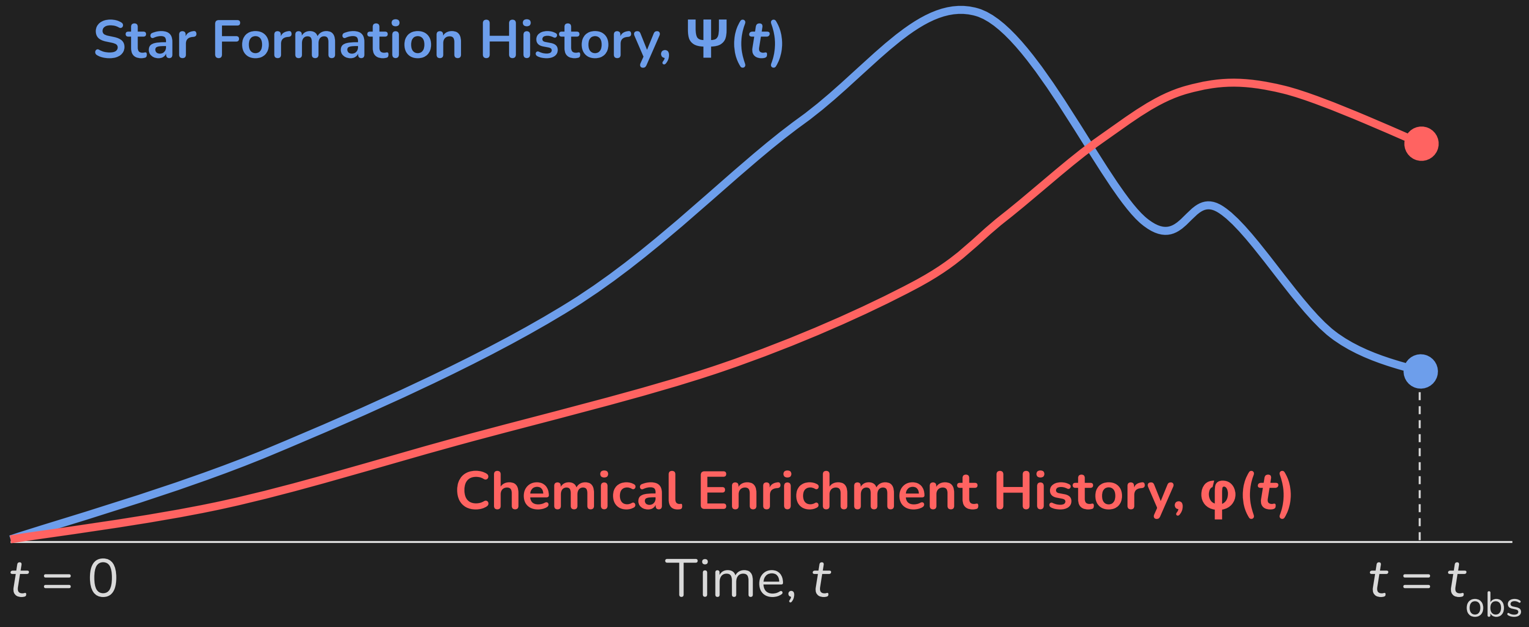
Galaxies grow via gas accretion, star formation, and merging



Galaxies grow via gas accretion, star formation, and merging



Heavy element production follows star formation

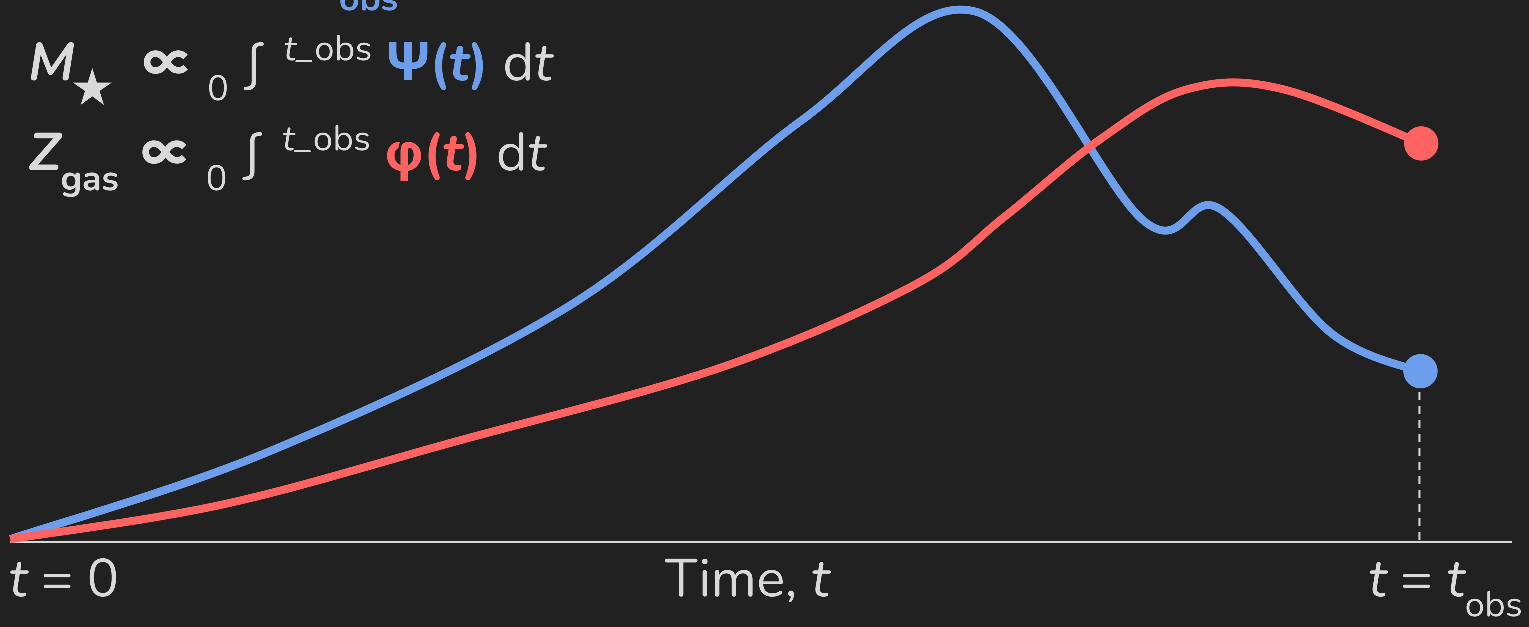


Models map physical processes to observables

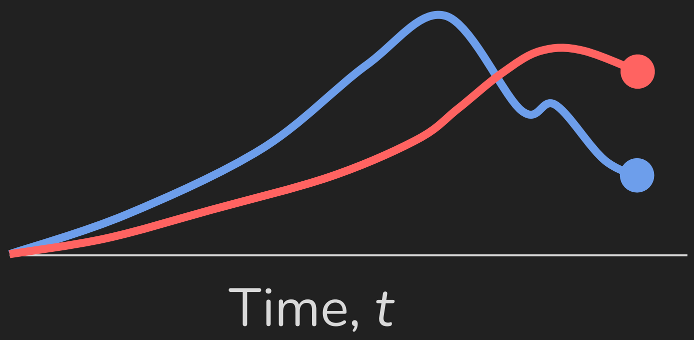
$$\mathrm{SFR} = \Psi(t=t_{obs})$$

$$M_\star \propto \int_0^{t\_obs} \Psi(t)\, \mathrm{d}t$$

$$Z_{gas} \propto \int_0^{t\_obs} \varphi(t)\, \mathrm{d}t$$



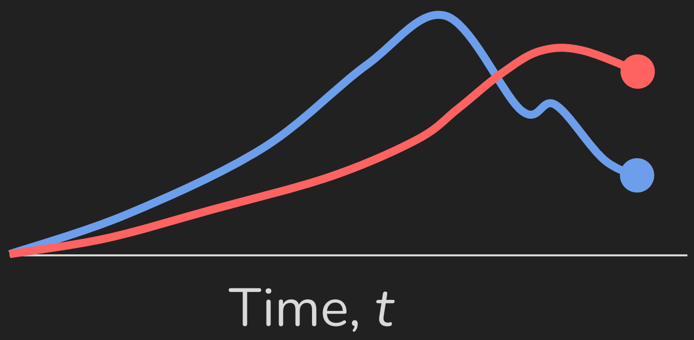
Models map physical processes to observables



$$M_\star \propto {}_0\int^{t\_obs} \Psi(t)\, dt$$

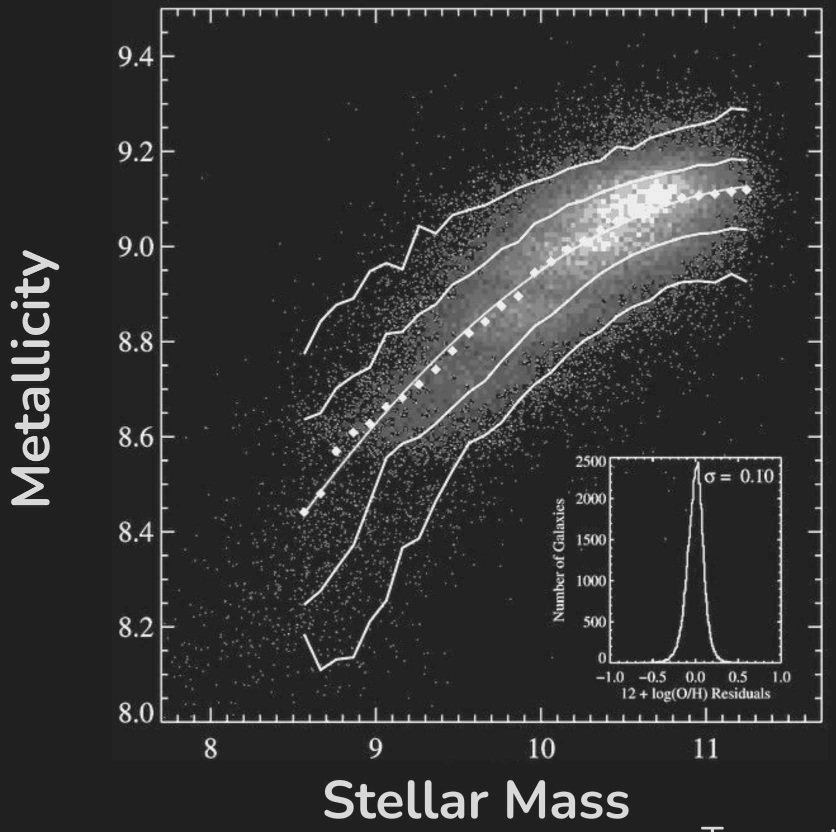
$$Z_{gas} \propto {}_0\int^{t\_obs} \varphi(t)\, dt$$

Models map physical processes to observables



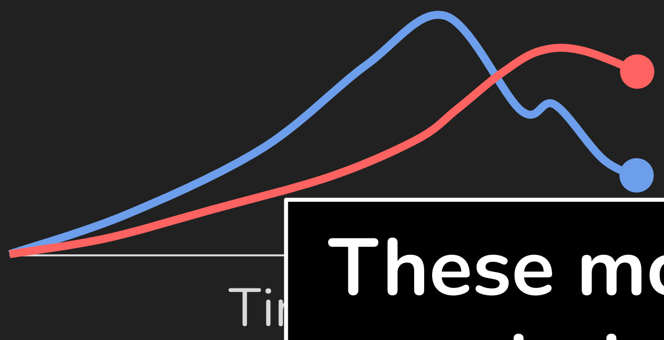
$$M_\star \propto \int_0^{t\_obs} \Psi(t)\, dt$$

$$Z_{gas} \propto \int_0^{t\_obs} \varphi(t)\, dt$$



Tremonti+ 04

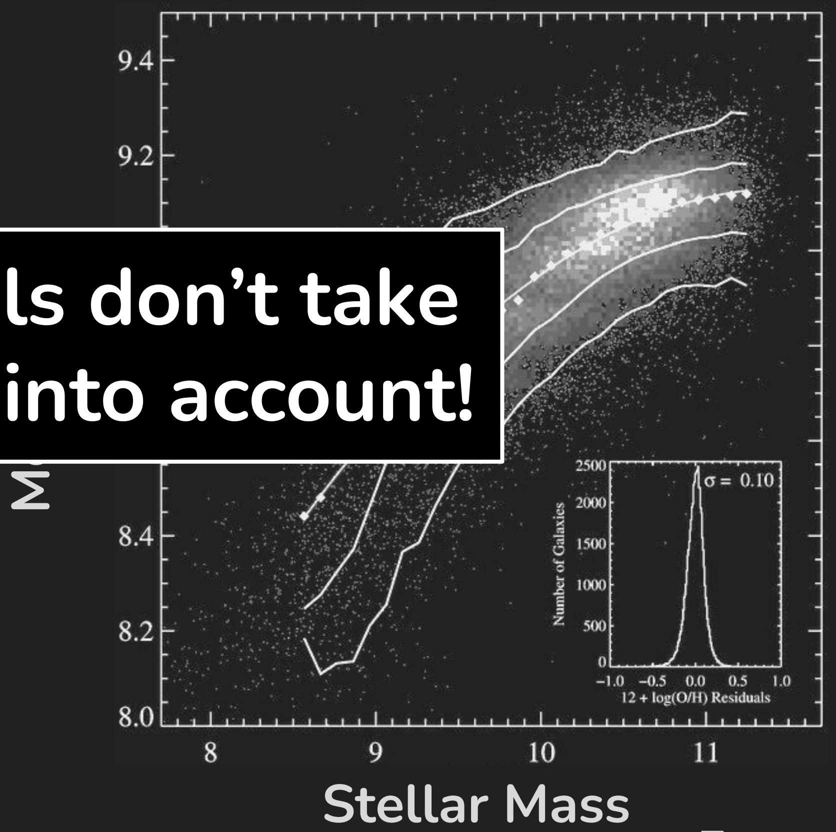
Models map physical processes to observables



These models don't take morphology into account!

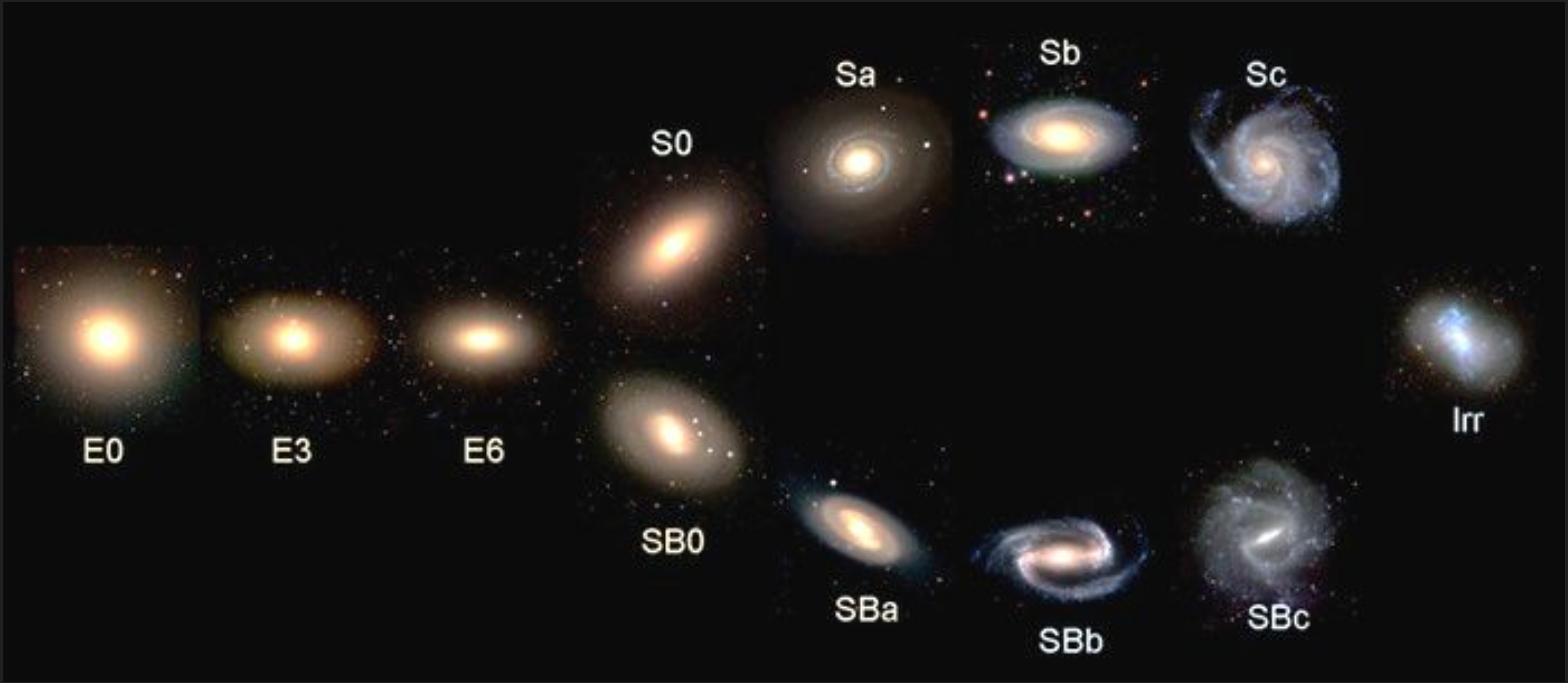
$$M_\star \propto \int_0^{t\_obs} \Psi(t)\, dt$$

$$Z_{gas} \propto \int_0^{t\_obs} \varphi(t)\, dt$$



Tremonti+ 04

Physical processes are imprinted on galaxies' morphologies



U. Iowa

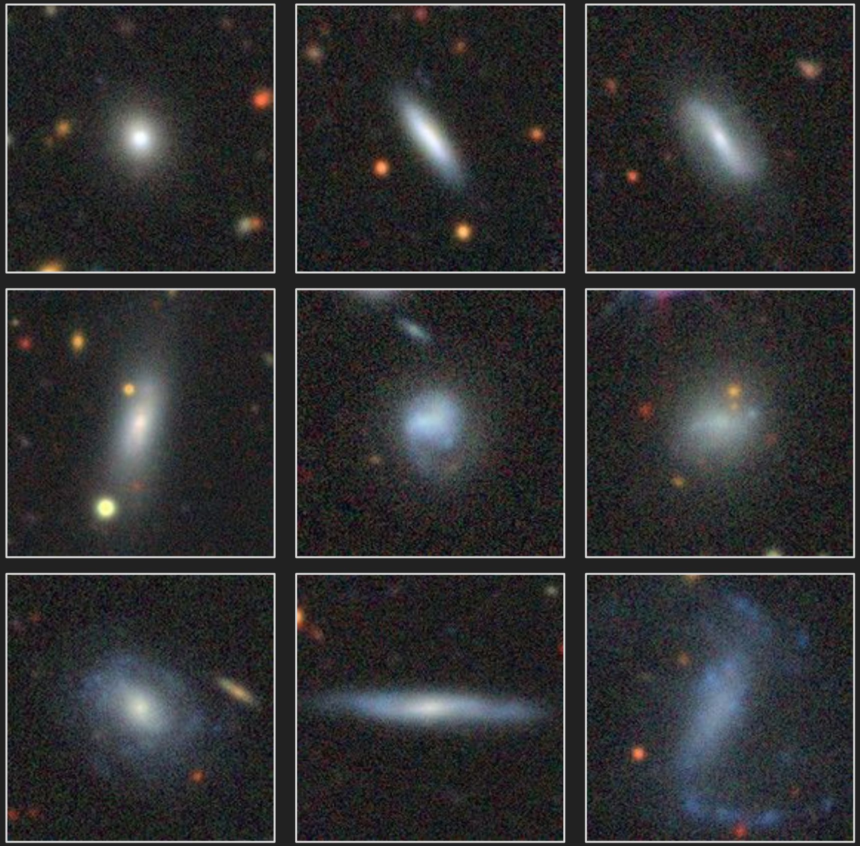
An image is more informative than a row in a photometric catalog

g mag	*r* mag
17.50	16.99
17.47	16.97
17.50	17.00
17.46	16.95
17.43	16.93
17.48	16.97
17.42	16.92
17.46	16.95
17.47	16.97

Legacy Survey DR9 (Dey+ 19)

An image is more informative than a row in a photometric catalog

g mag	*r* mag
17.50	16.99
17.47	16.97
17.50	17.00
17.46	16.95
17.43	16.93
17.48	16.97
17.42	16.92
17.46	16.95
17.47	16.97



Legacy Survey DR9 (Dey+ 19)

Roadmap

I. **The growth and evolution of galaxies**
II. Studying satellite galaxies with CNNs
III. Learning relationships with graphs

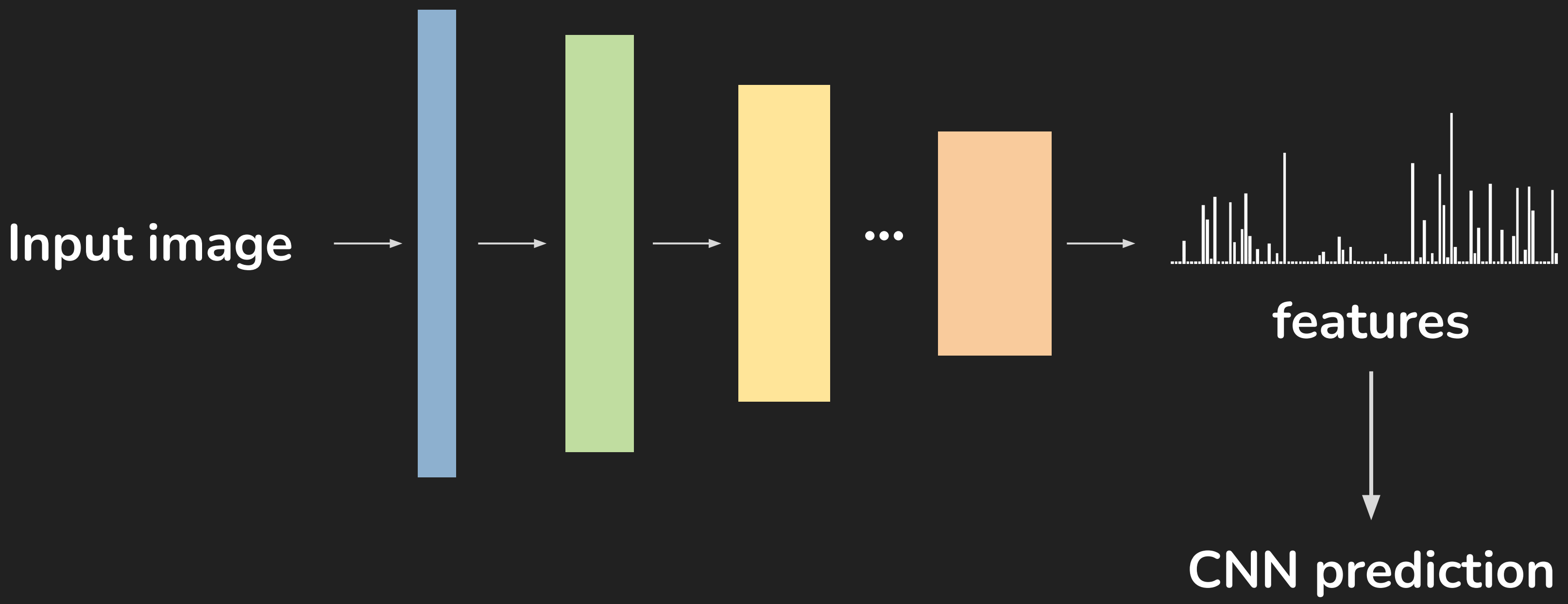
Roadmap

I. The growth and evolution of galaxies

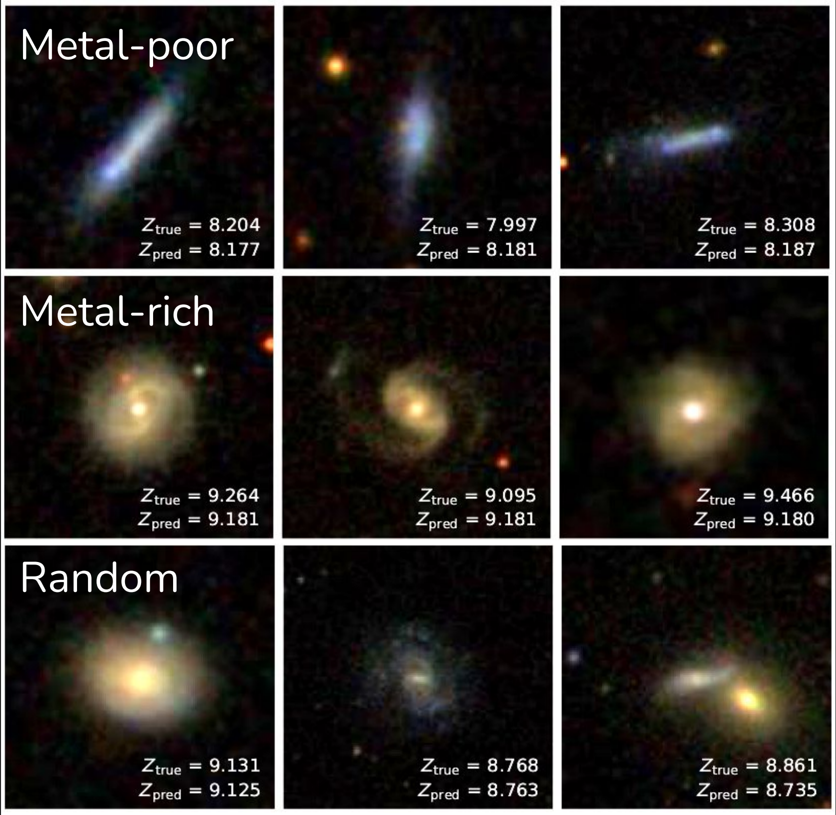
II. Studying satellite galaxies with CNNs

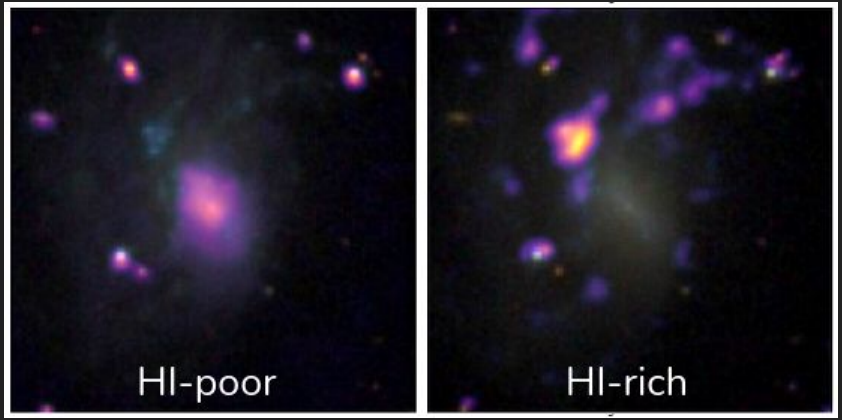
III. Learning relationships with graphs

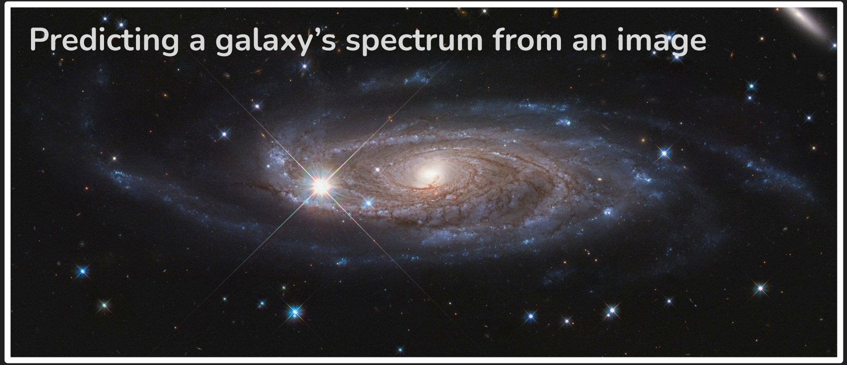
CNNs are just sequential morphological feature finders



Just a sample of what can be done with CNNs...



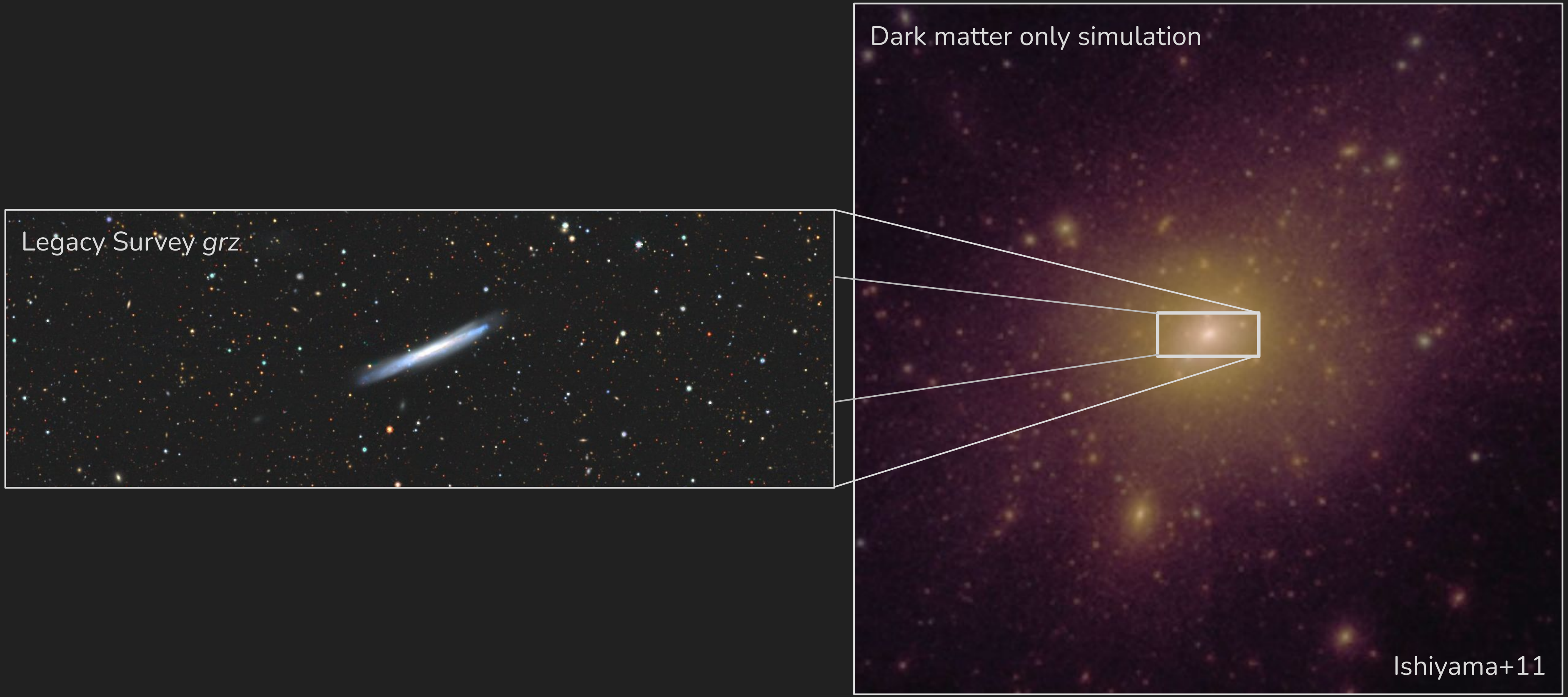




Wu & Boada 19; **Wu** 2020; **Wu** & Peek 2020; Holwerda+ 21; Guo+22

Identifying dwarf (satellite) galaxies is hard…

… but important for galaxy formation theory.



SAGA is the premier spectroscopic survey of low-z satellites

378 satellites around 101 host galaxies using >75,000 spectra

Mao+24, Geha+24, Wang+24

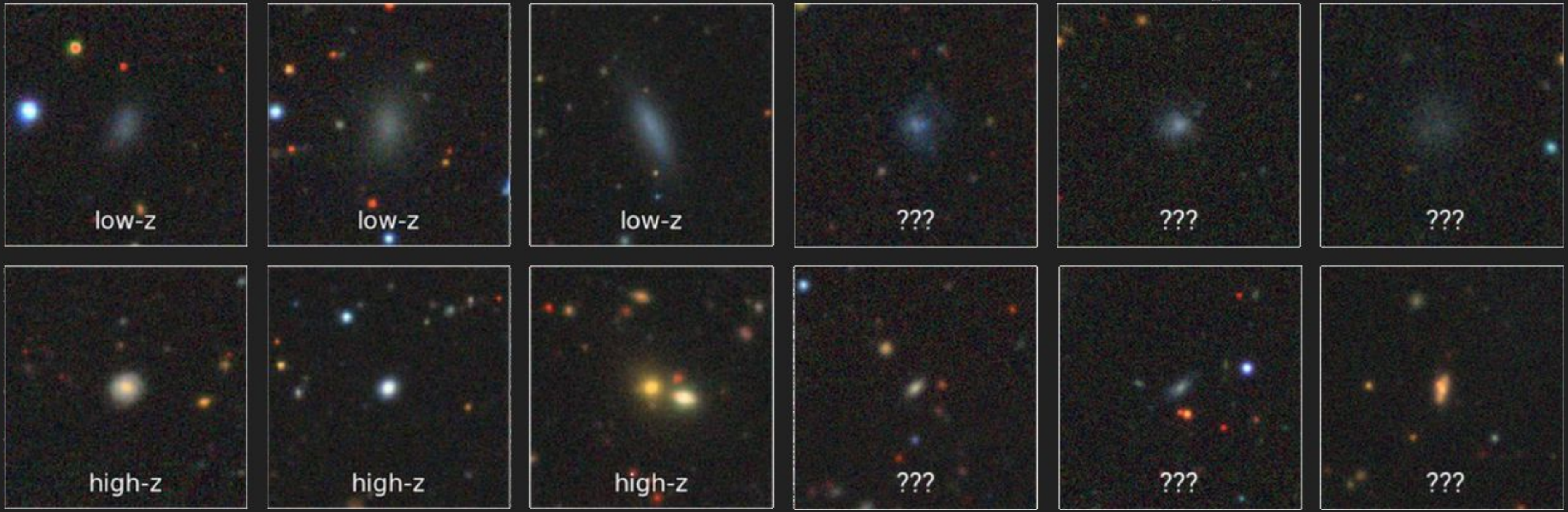
A CNN robustly selects low-z (z < 0.03) galaxies

SAGA training sample

A CNN robustly selects low-z (z < 0.03) galaxies

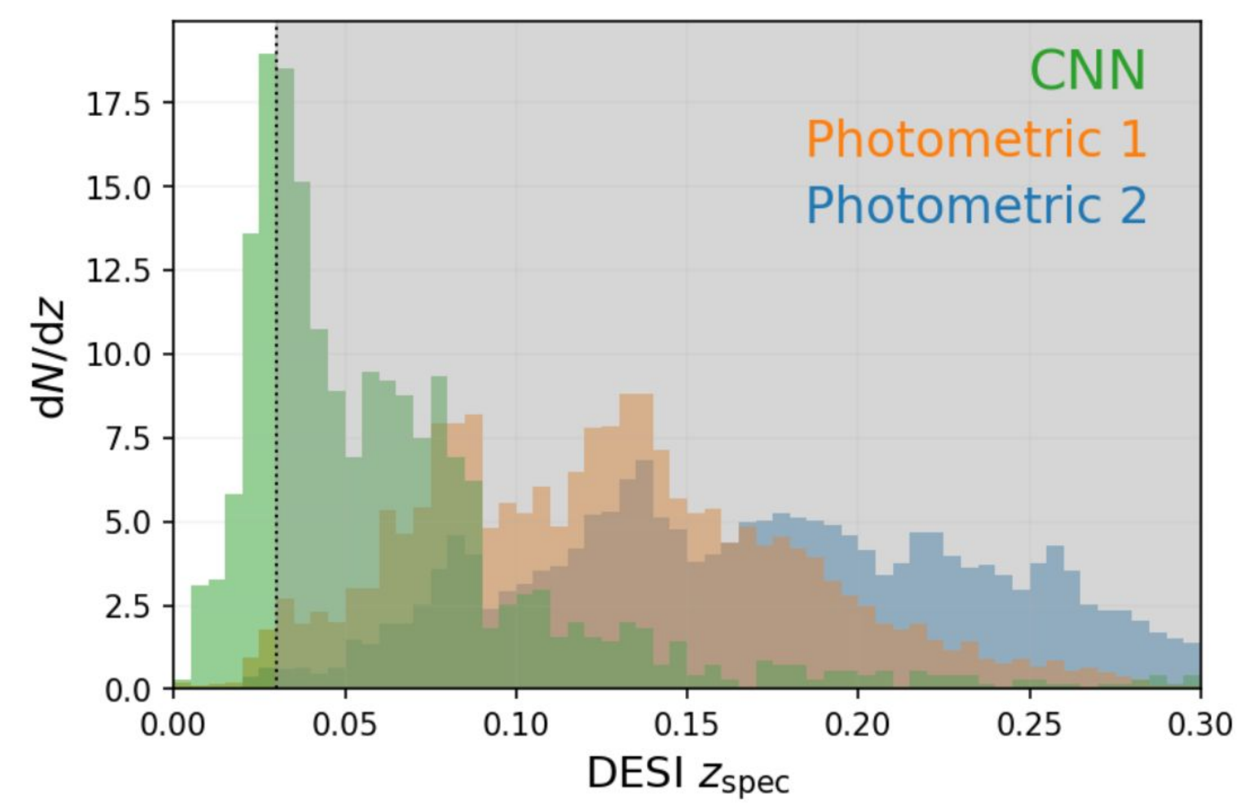
SAGA training sample

xSAGA test sample



Wu+ 22

We can *validate* CNN performance with observations!

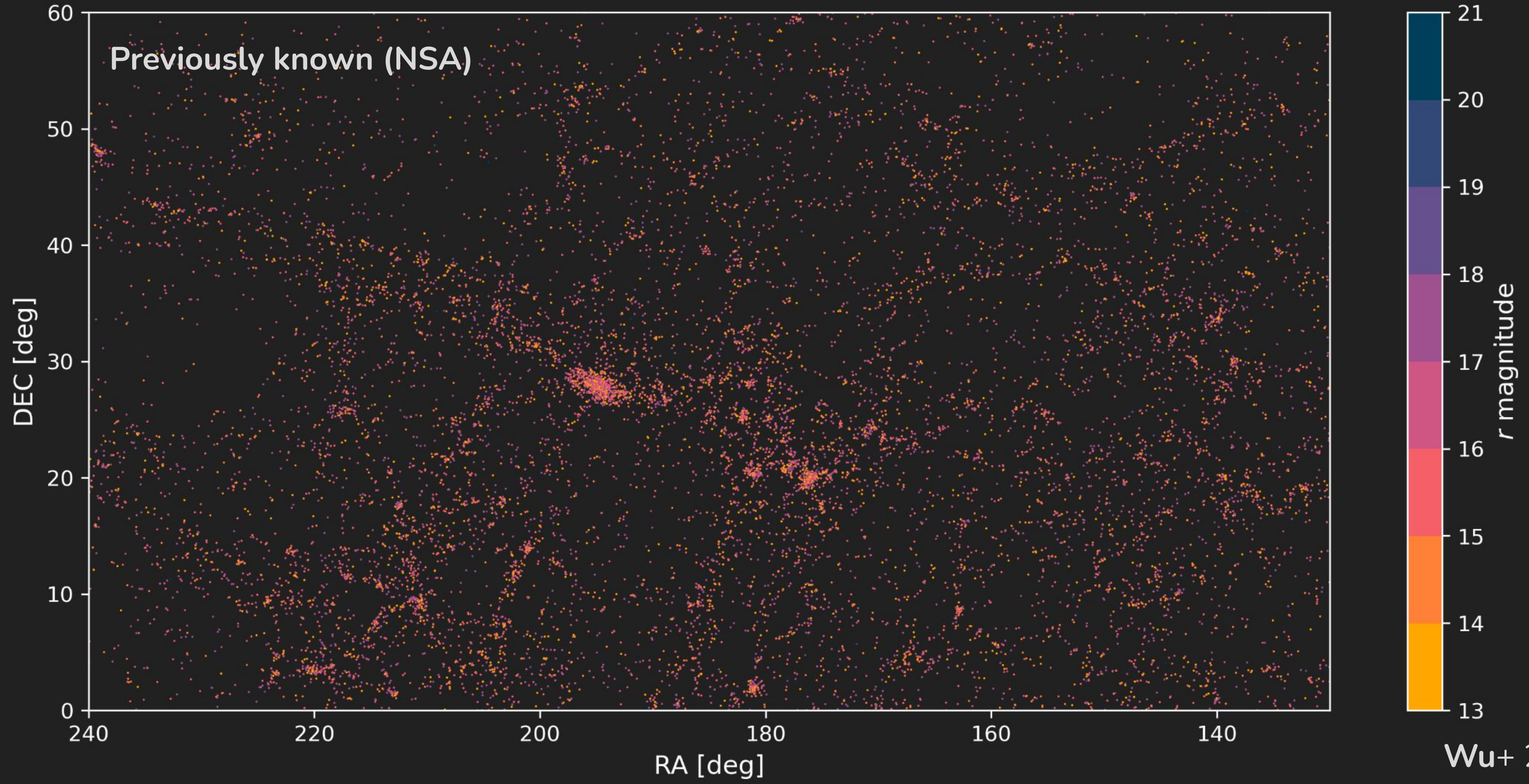


~ 19% (Tier 1)
~ 1.3% (Tier 2)
~ 0.5% (Tier 3)

from DESI LOW-Z spectroscopy!

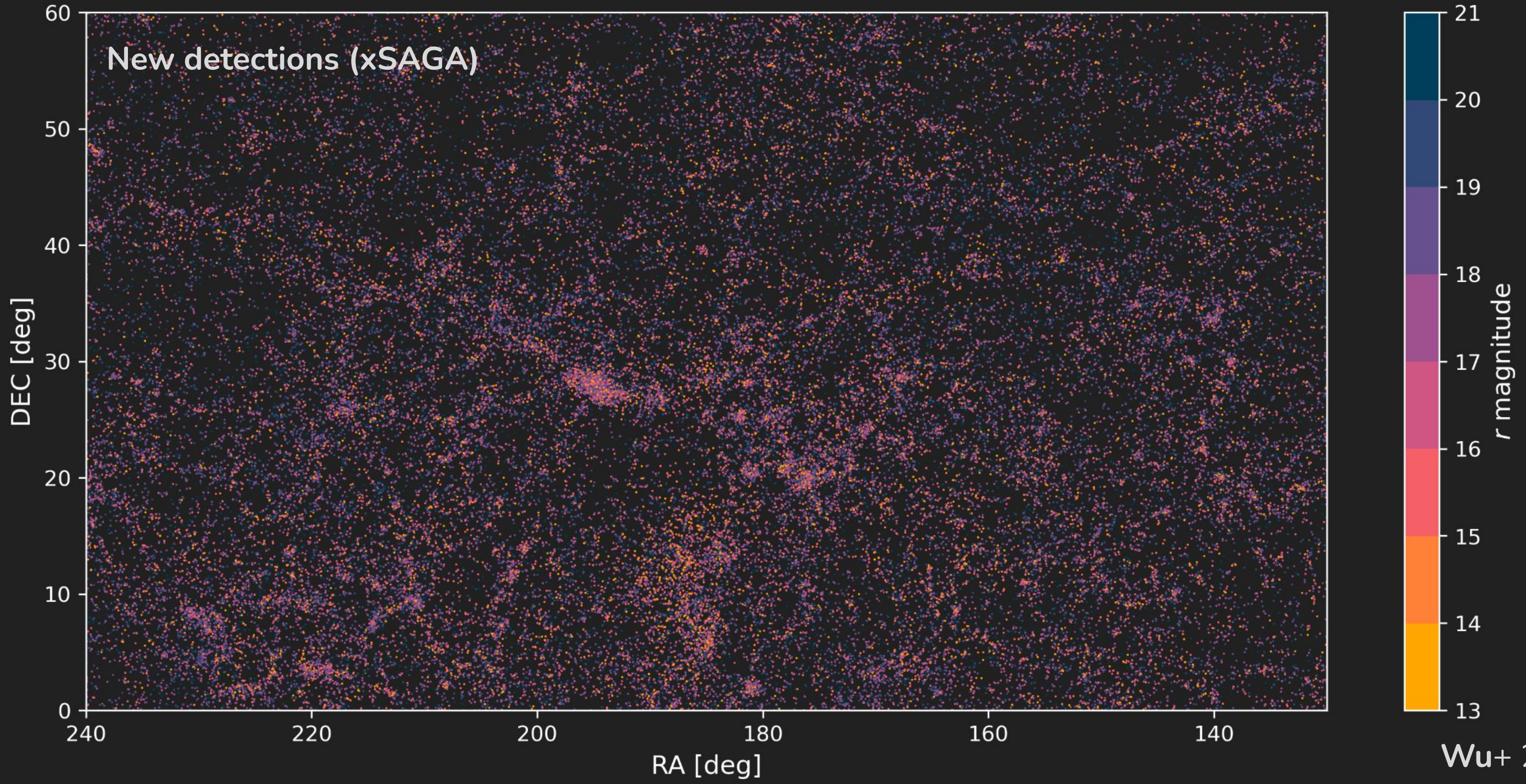
Darragh-Ford+ 23

SDSS found bright $z < 0.03$ galaxies



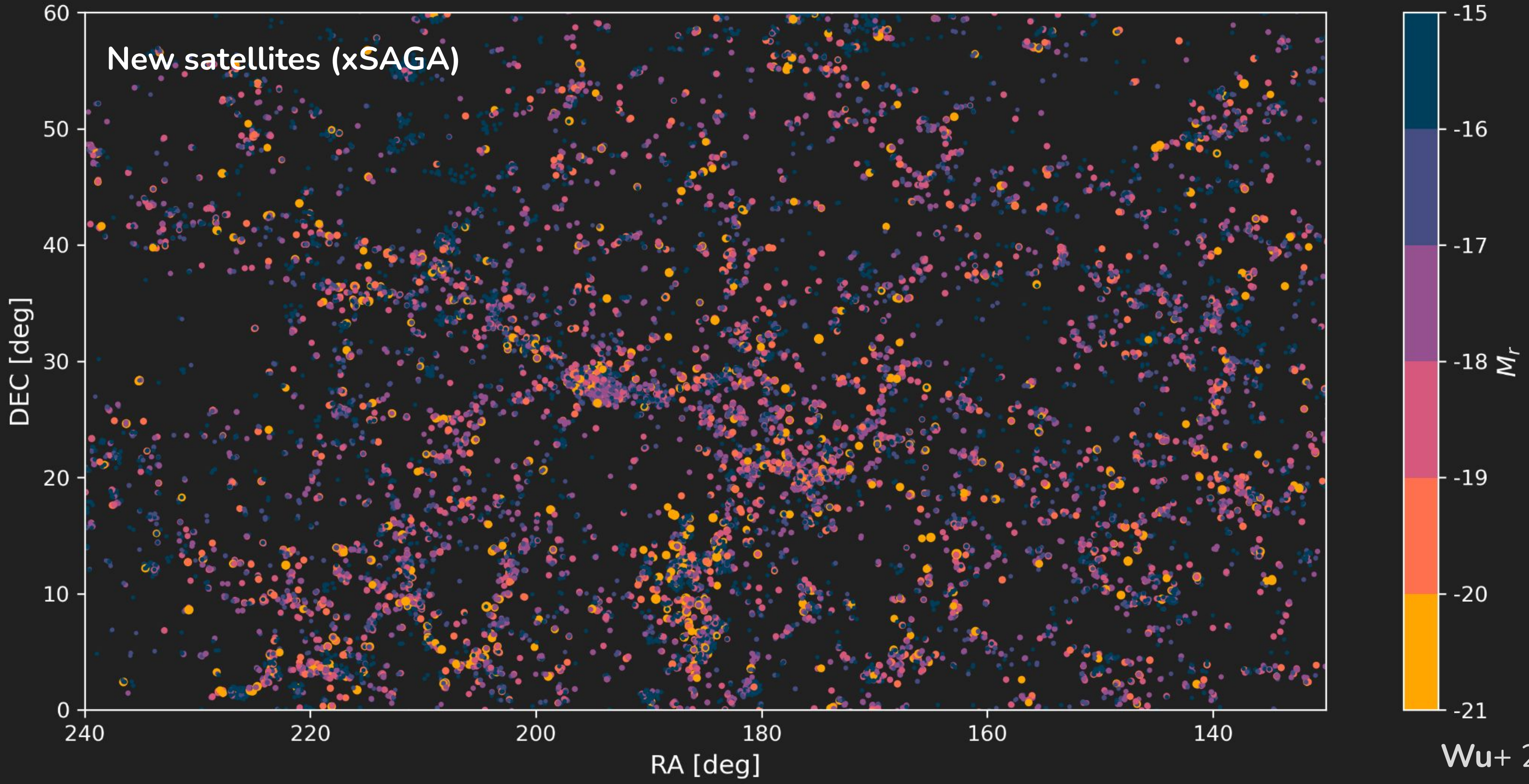
Wu+ 22

Here are another >100k low-z candidates found via CNN



Wu+ 22

xSAGA: >100x as many satellite systems as before



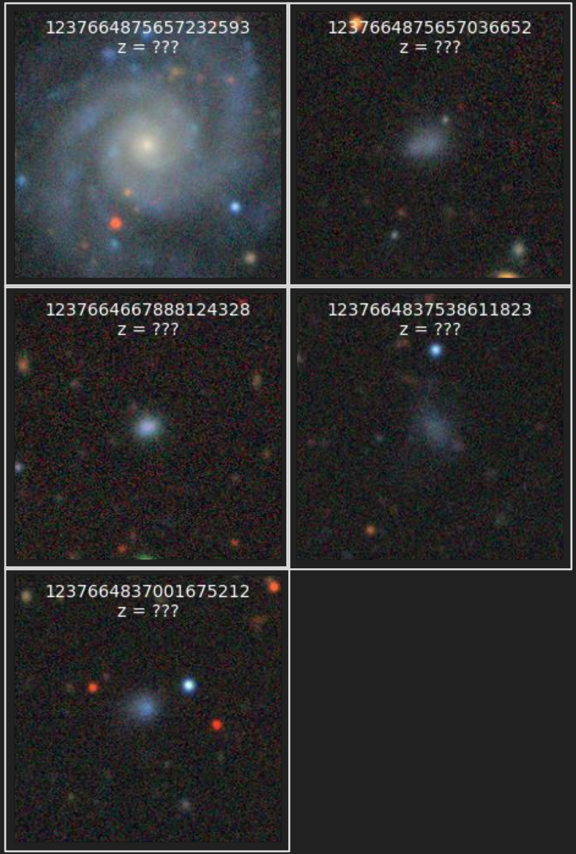
Wu+ 22

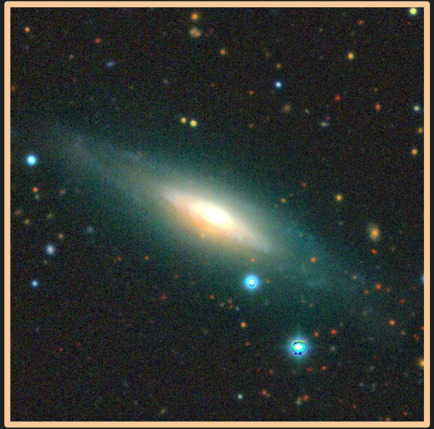
Studying satellites around z~0.03 Milky Way analogs

spectroscopically confirmed

no redshift confirmed

NSAID 407998
z = 0.029

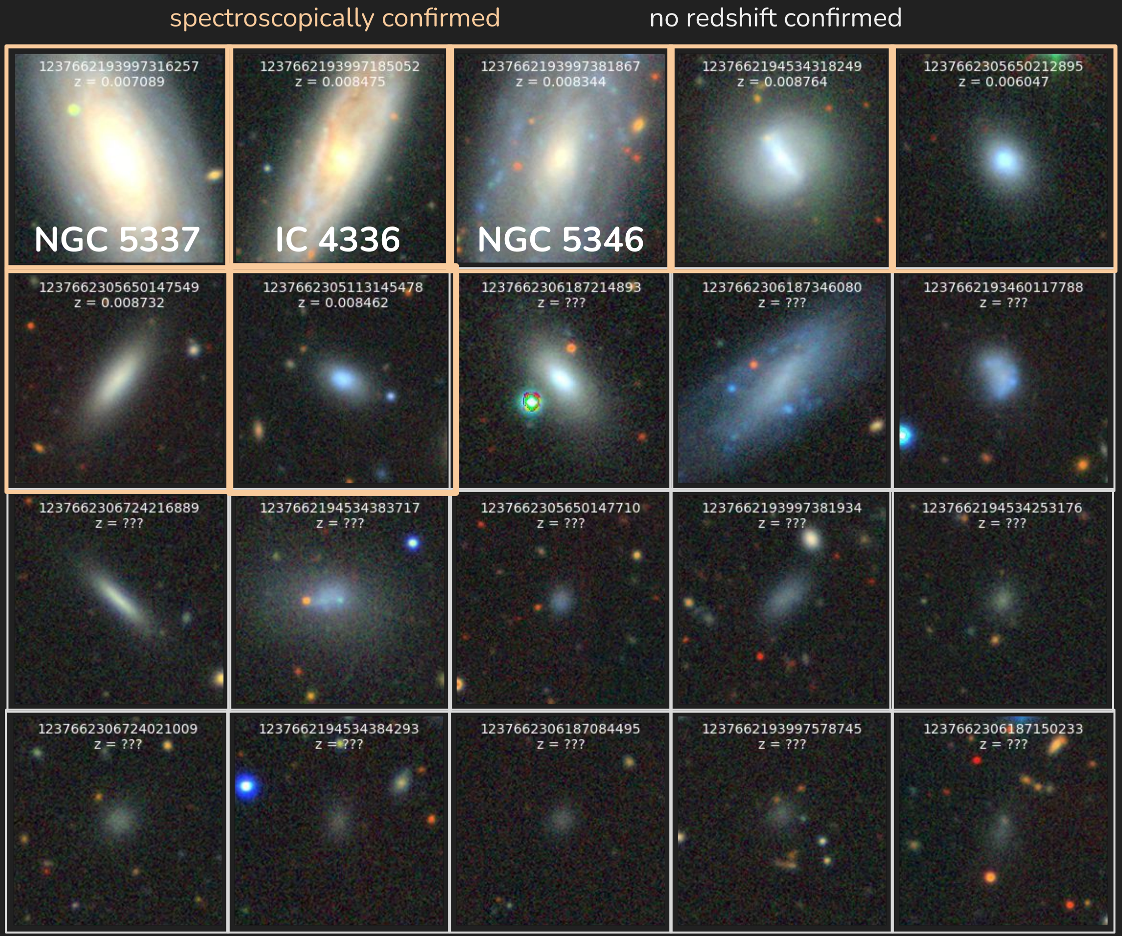




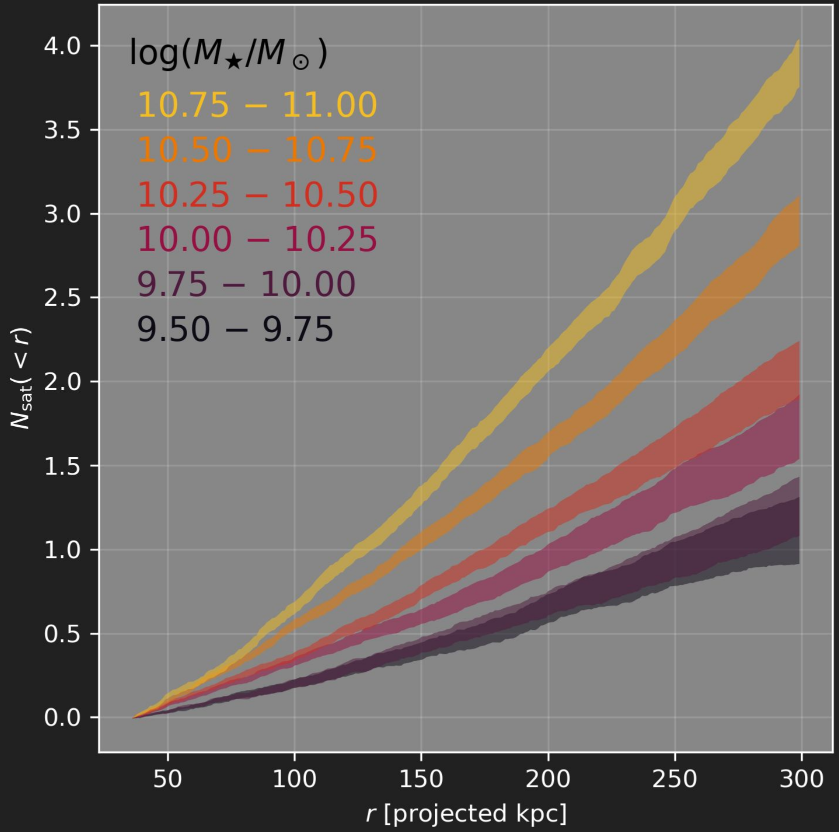
NGC 1234
z = 0.020

Studying z ~ 0.008 satellite groups and their dwarfs

NGC 5326 $z = 0.008$

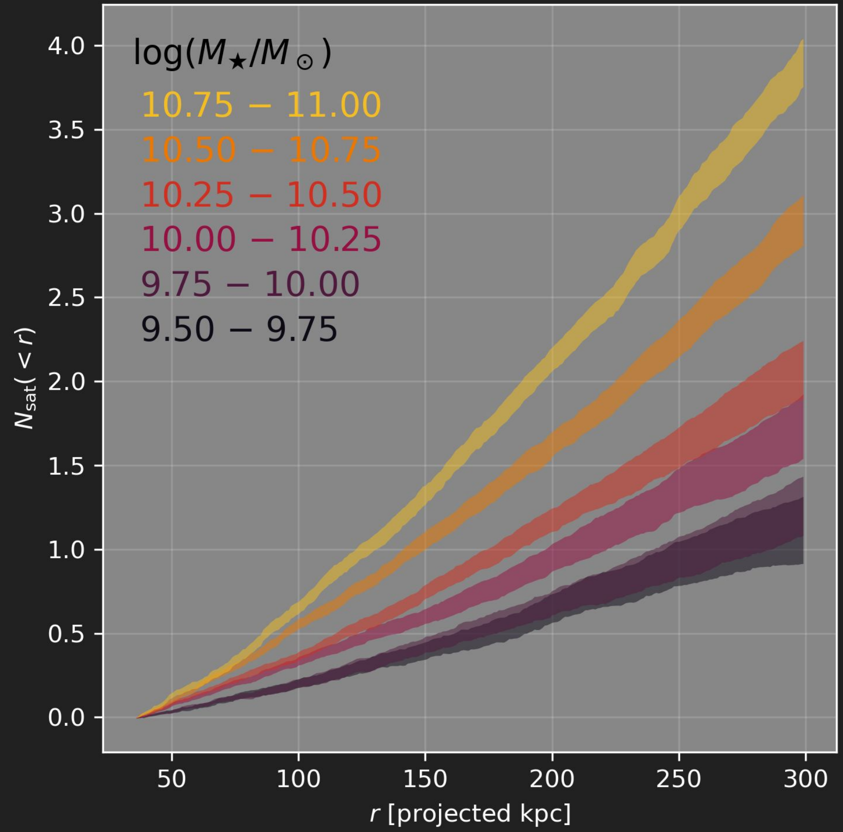


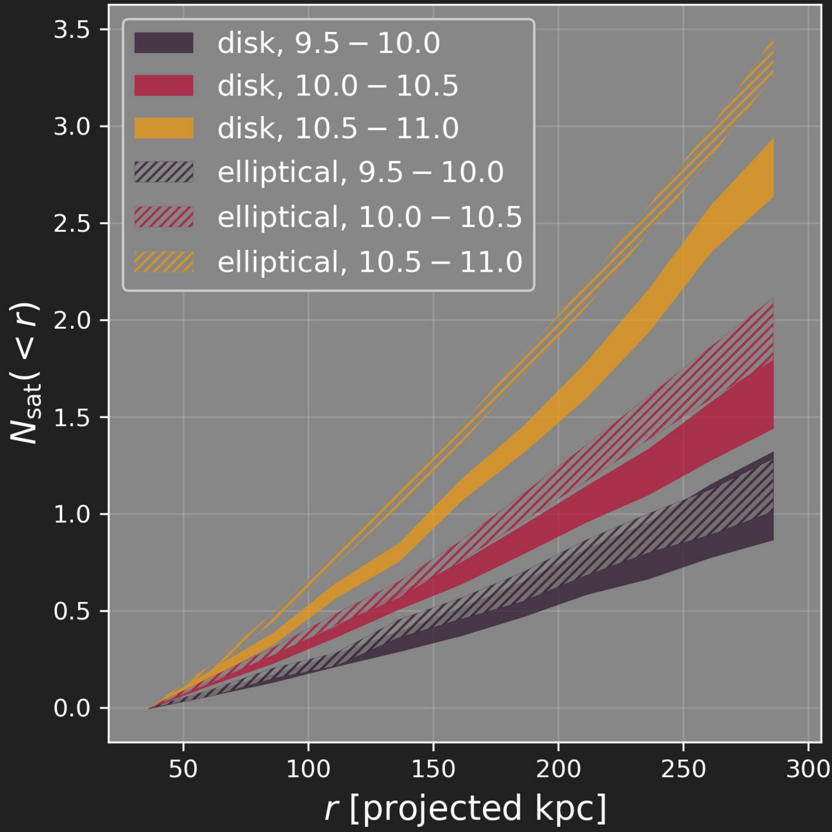
Exquisite statistics on satellite radial profiles with host mass!



Wu+ 22

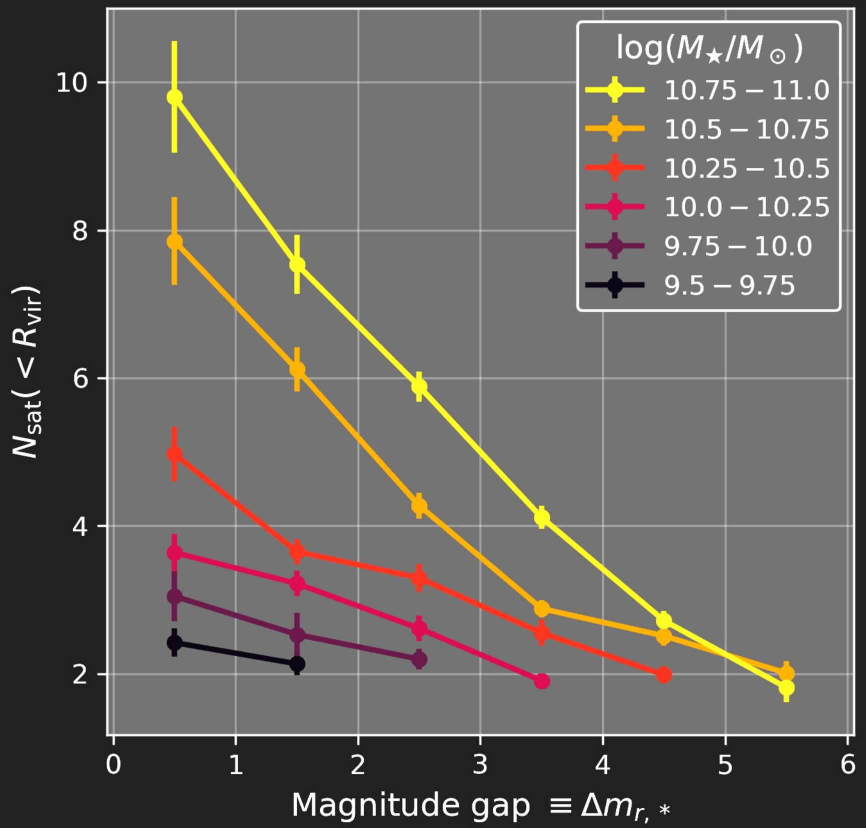
Exquisite statistics on satellite radial profiles with host mass!





Wu+ 22

Satellites probe the halo accretion history



Time since last halo accretion event →

Wu+ 22

Roadmap

I. The growth and evolution of galaxies

II. Studying satellite galaxies with CNNs

III. Learning relationships with graphs

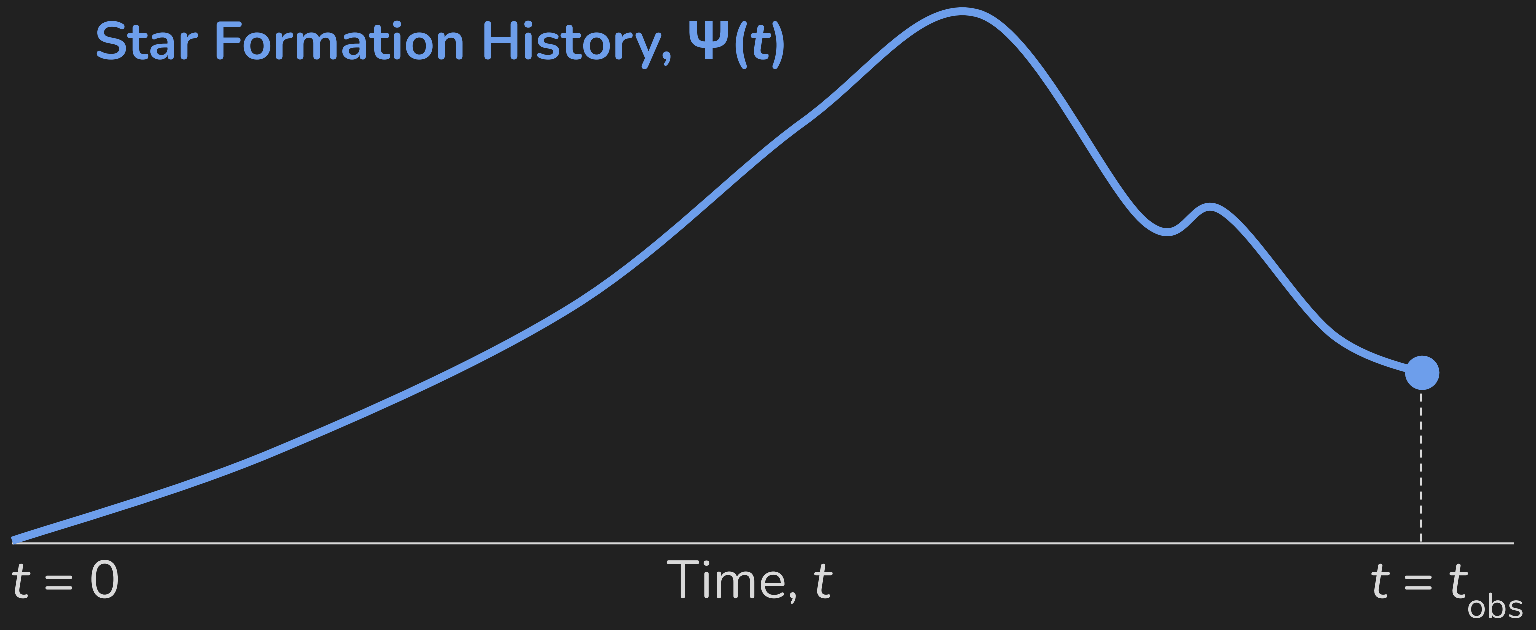
Roadmap

I. The growth and evolution of galaxies

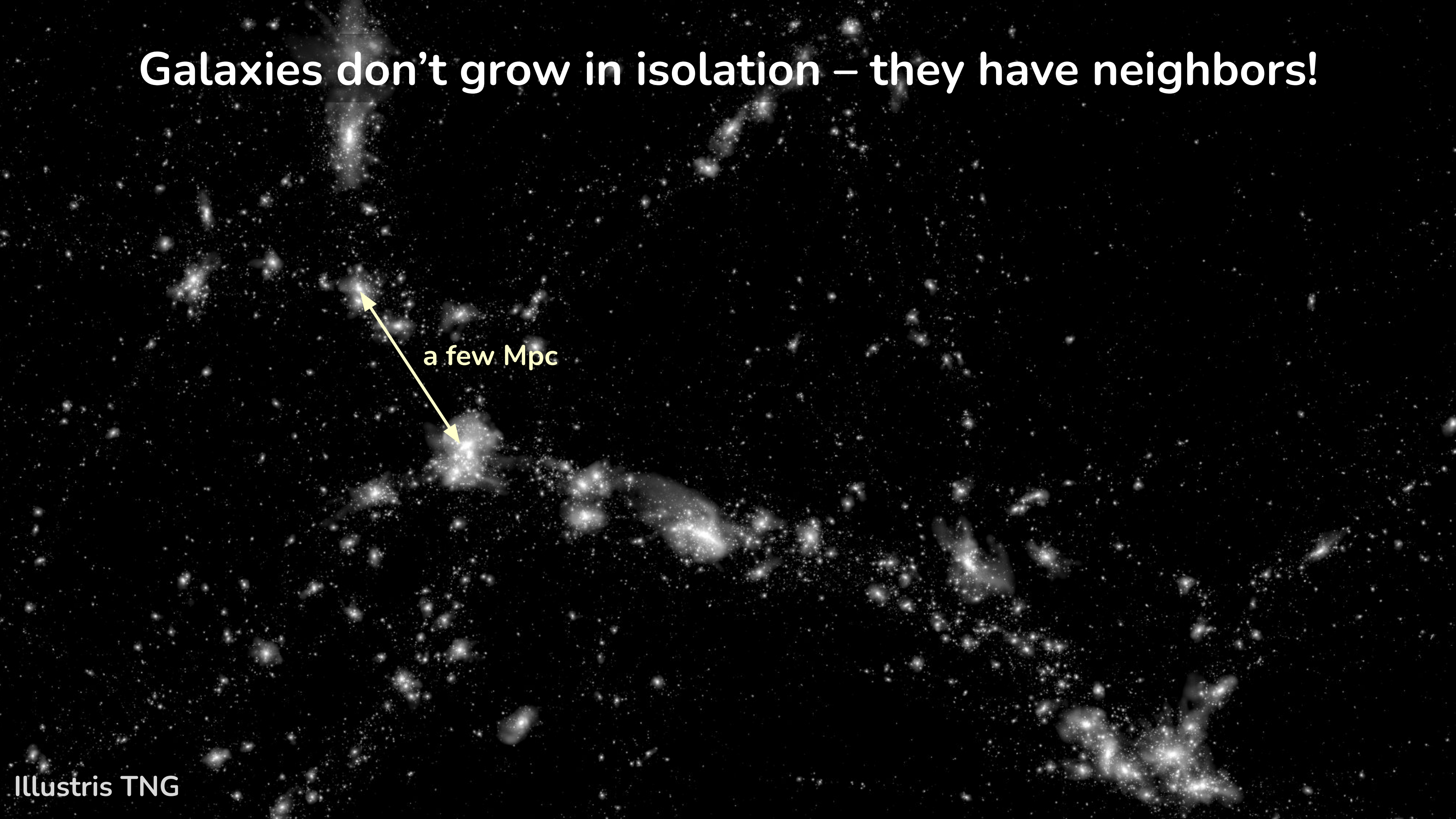
II. Studying satellite galaxies with CNNs

III. Learning relationships with graphs

Galaxies don't grow in isolation



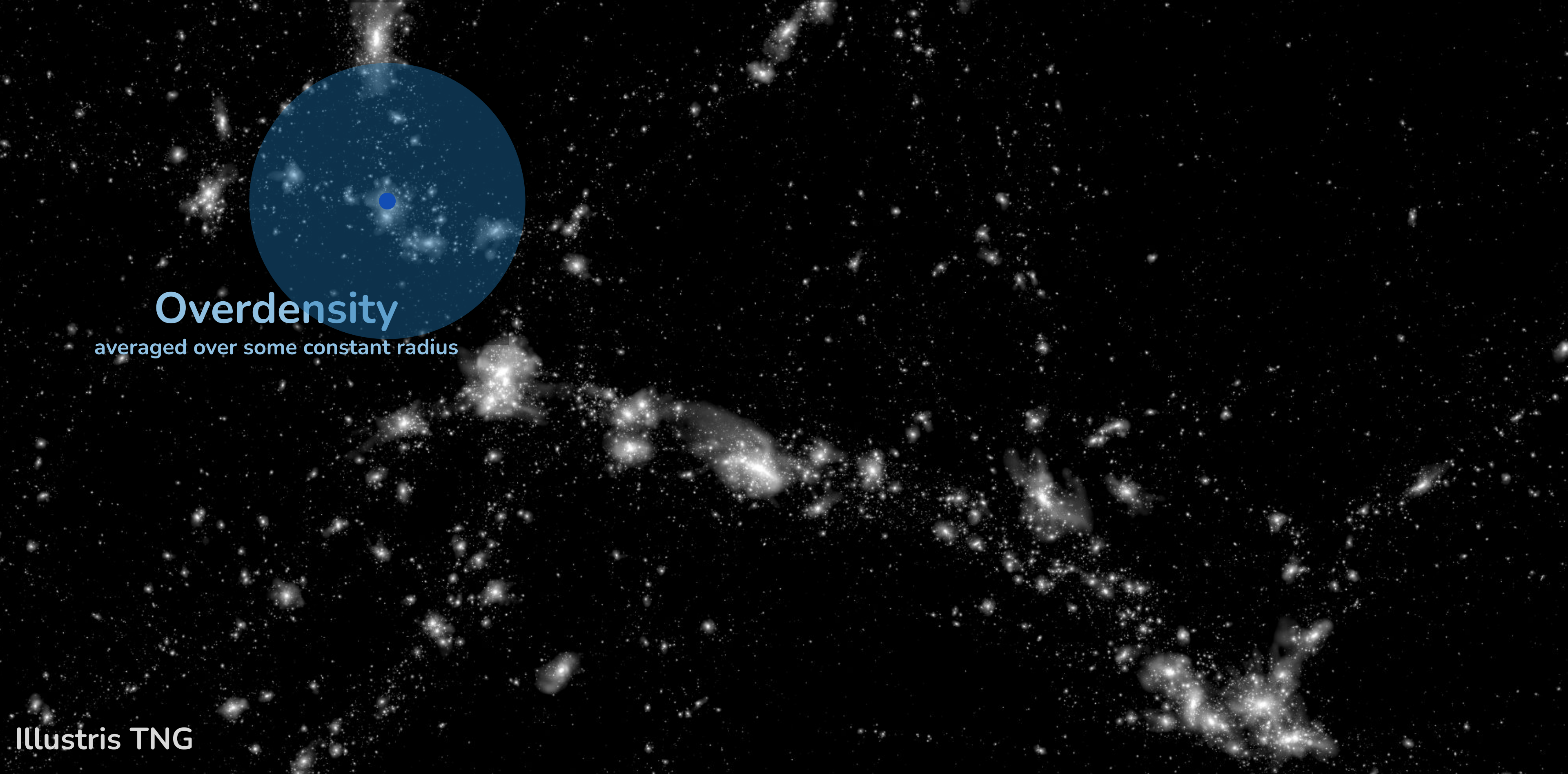
Galaxies don’t grow in isolation – they have neighbors!

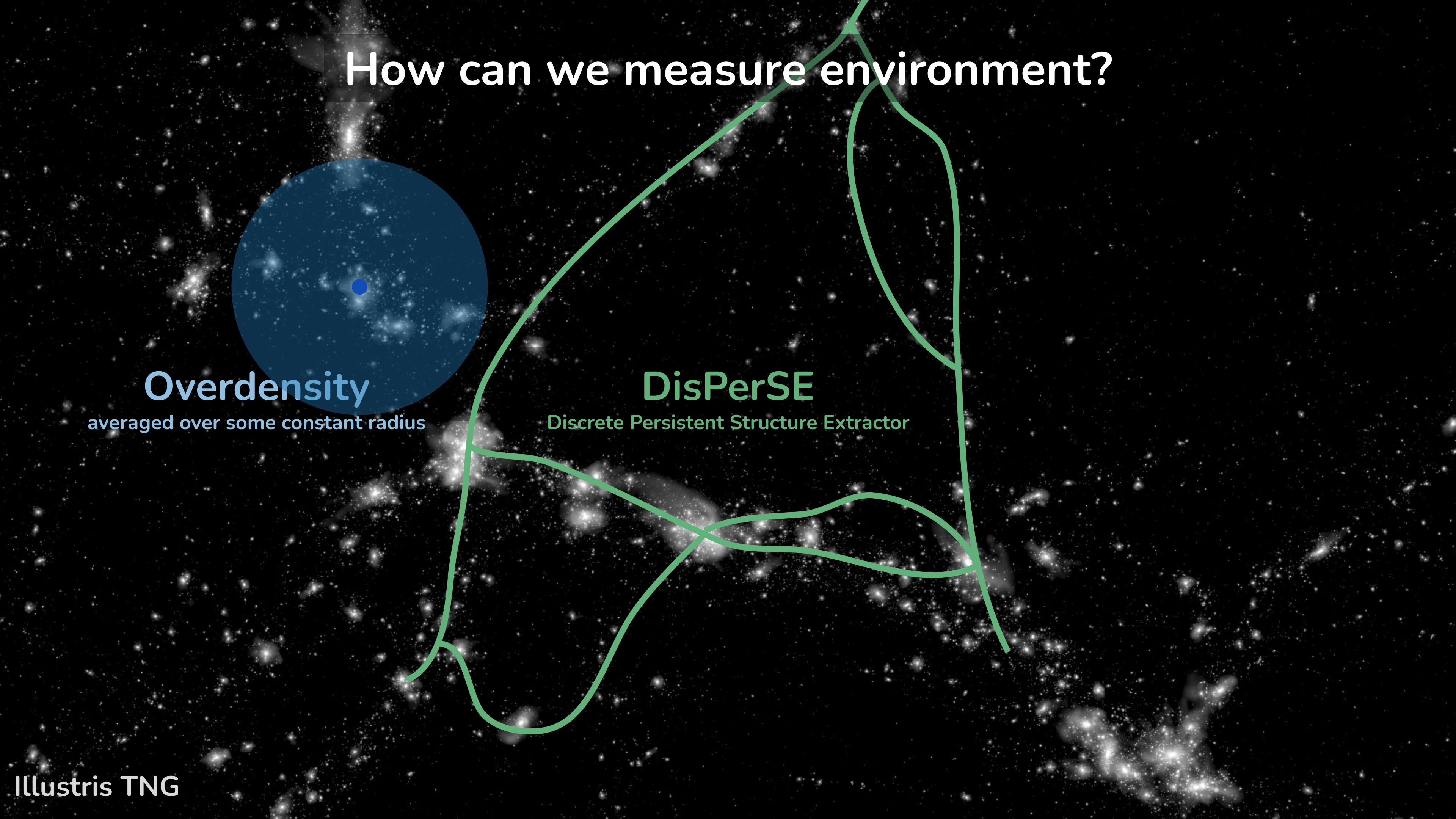


How can we measure environment?

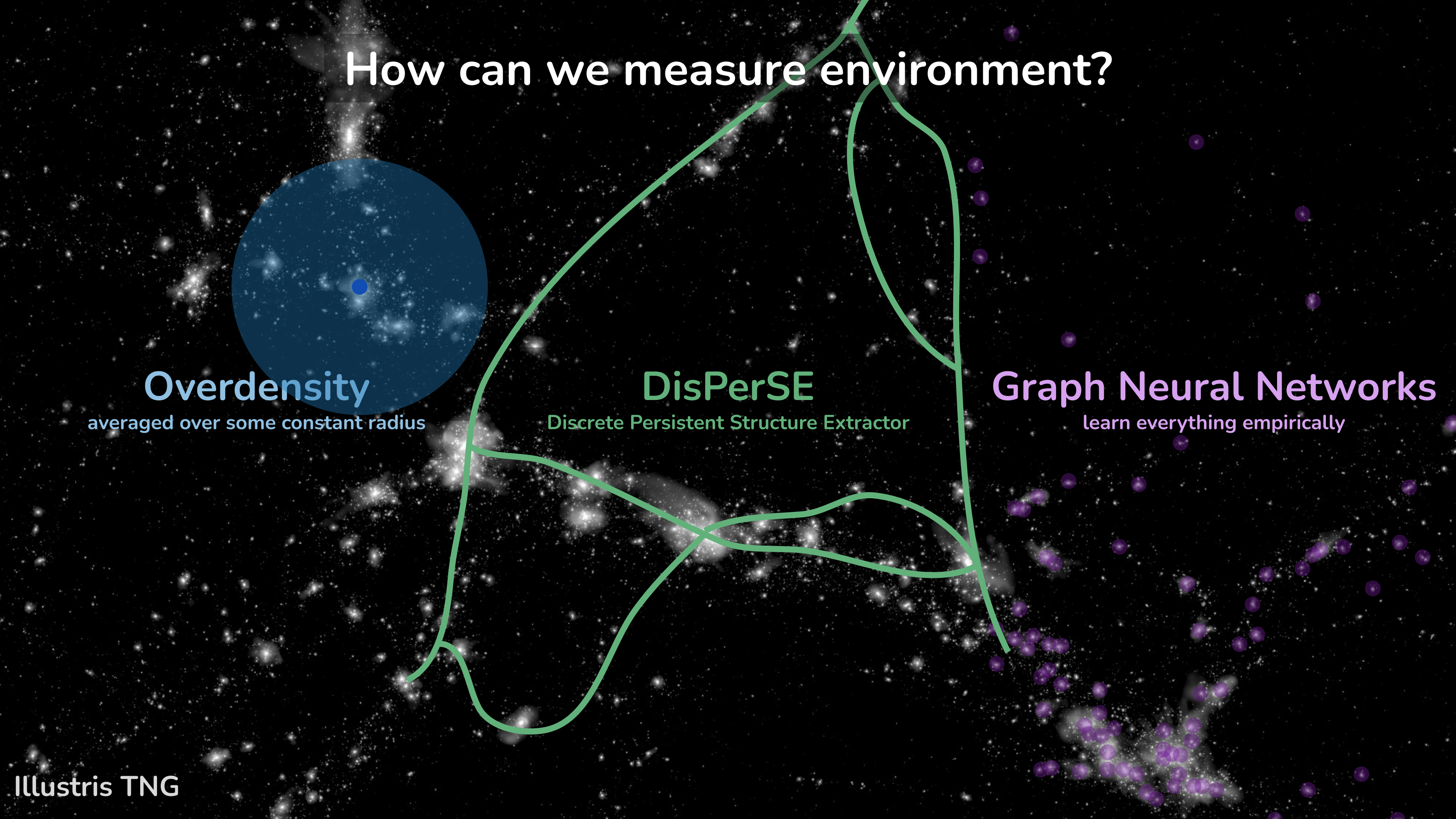
Illustris TNG

How can we measure environment?

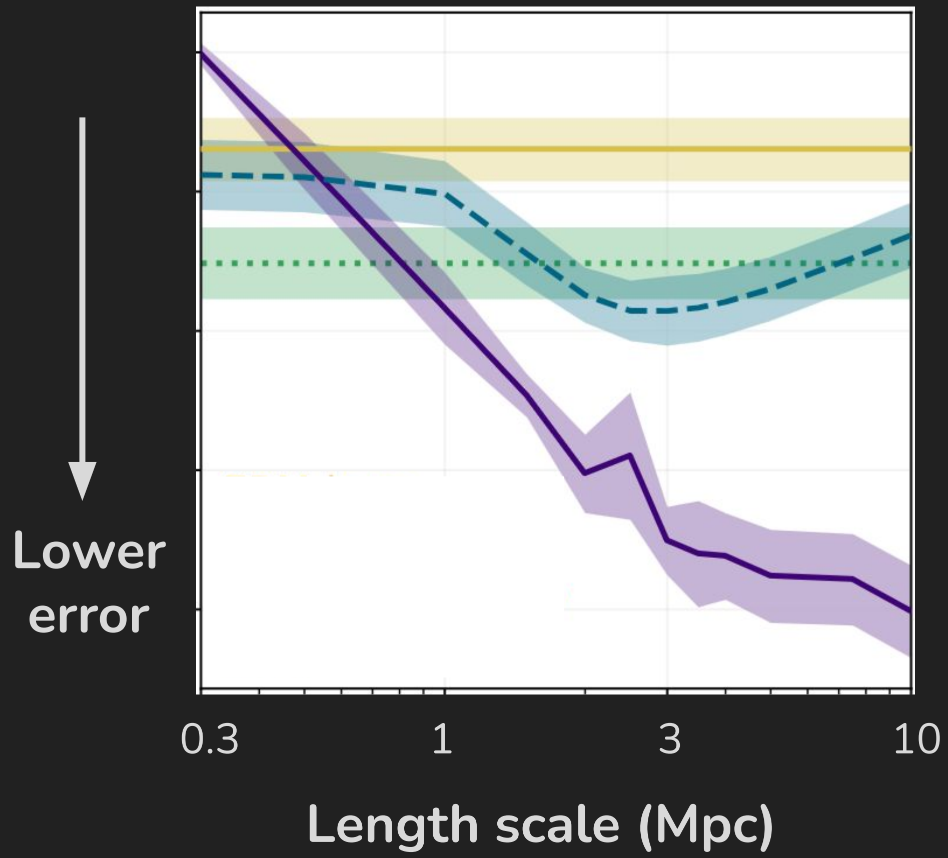



How can we measure environment?
Overdensity
averaged over some constant radius
DisPerSE
Discrete Persistent Structure Extractor
Illustris TNG

How can we measure environment?



Environment tells you which galaxy lives in which halo



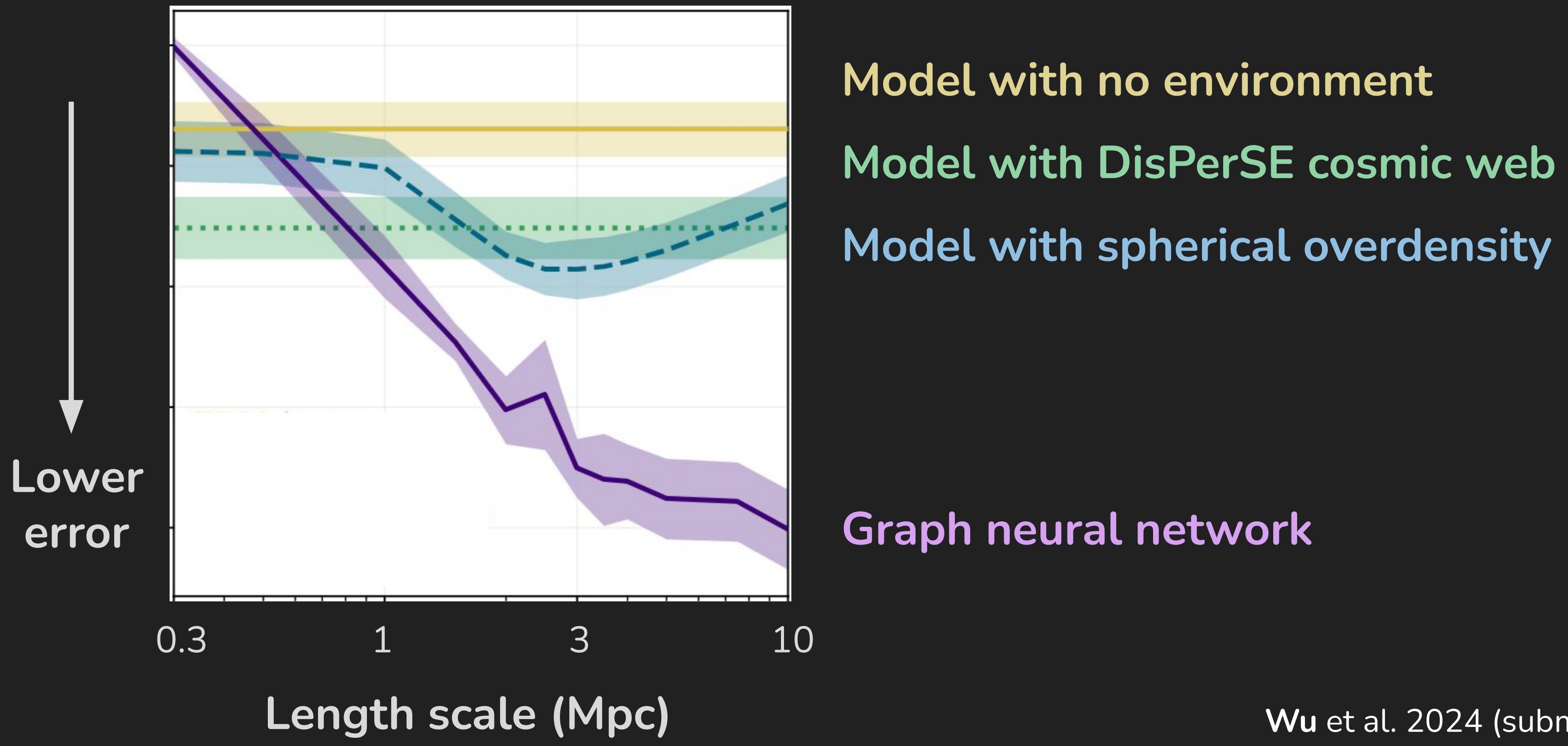
Model with no environment

Model with DisPerSE cosmic web

Model with spherical overdensity

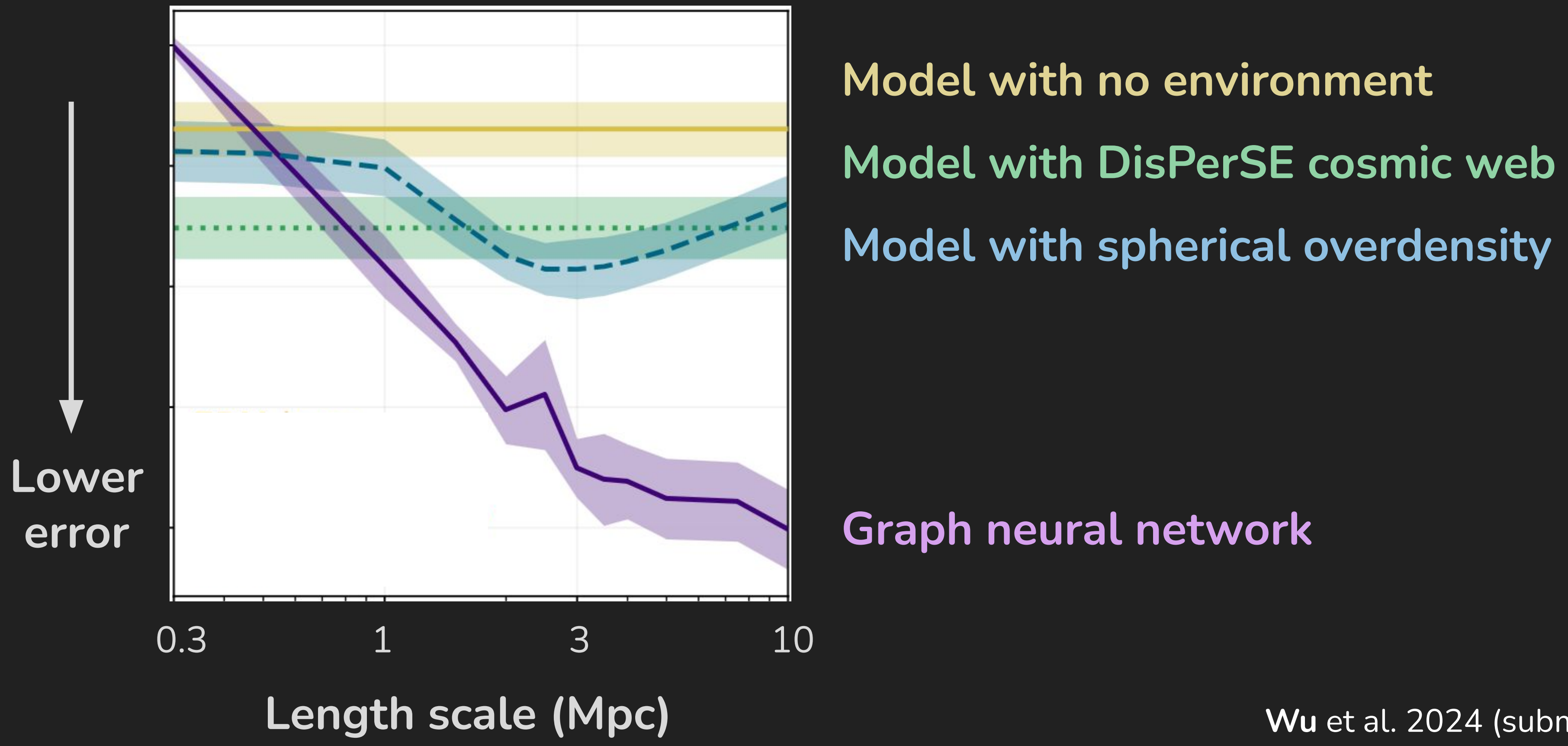
Wu et al. 2024 (subm.)
arXiv:2402.07995

GNNs are most informative at capturing environment



Wu et al. 2024 (subm.)
arXiv:2402.07995

GNNs are most informative at capturing environment

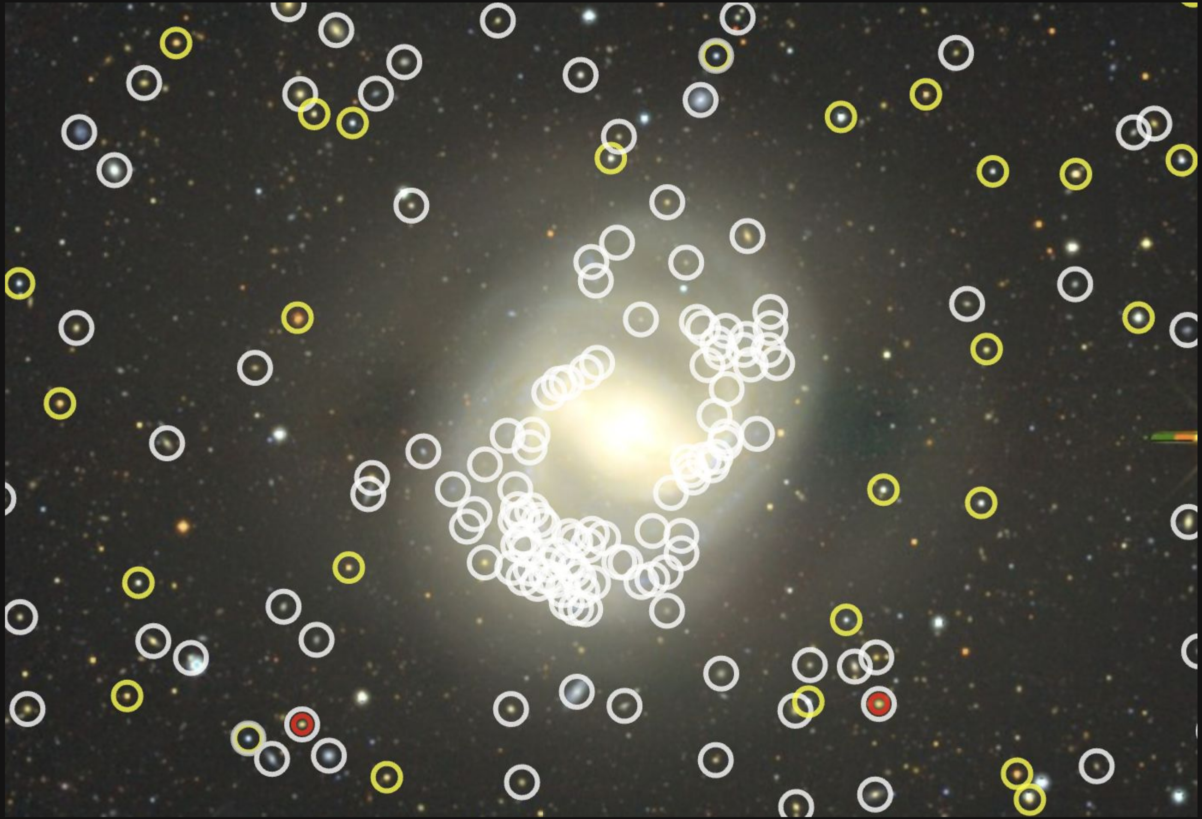


Wu et al. 2024 (subm.)
arXiv:2402.07995

Anomaly detection at scale with graph learning

M91 – Legacy Survey DR10 *grz*

Anomaly detection at scale with graph learning



M91 – Legacy Survey DR10 *grz* + **DESI bright time fiber targets**

Summary

I. We can learn a lot about galaxies using advanced ML methods and astronomical survey data.

Summary

I. We can learn a lot about galaxies using advanced ML methods and astronomical survey data.

II. The morphologies of galaxies tells us about their physical evolution – and helps to identify rare gems!

Summary

I. We can learn a lot about galaxies using advanced ML methods and astronomical survey data.

II. The morphologies of galaxies tells us about their physical evolution – and helps to identify rare gems!

III. xSAGA/DESI LOWZ gives us an entirely new way to study substructure of the low-redshift cosmos.

Summary

I. We can learn a lot about galaxies using advanced ML methods and astronomical survey data.

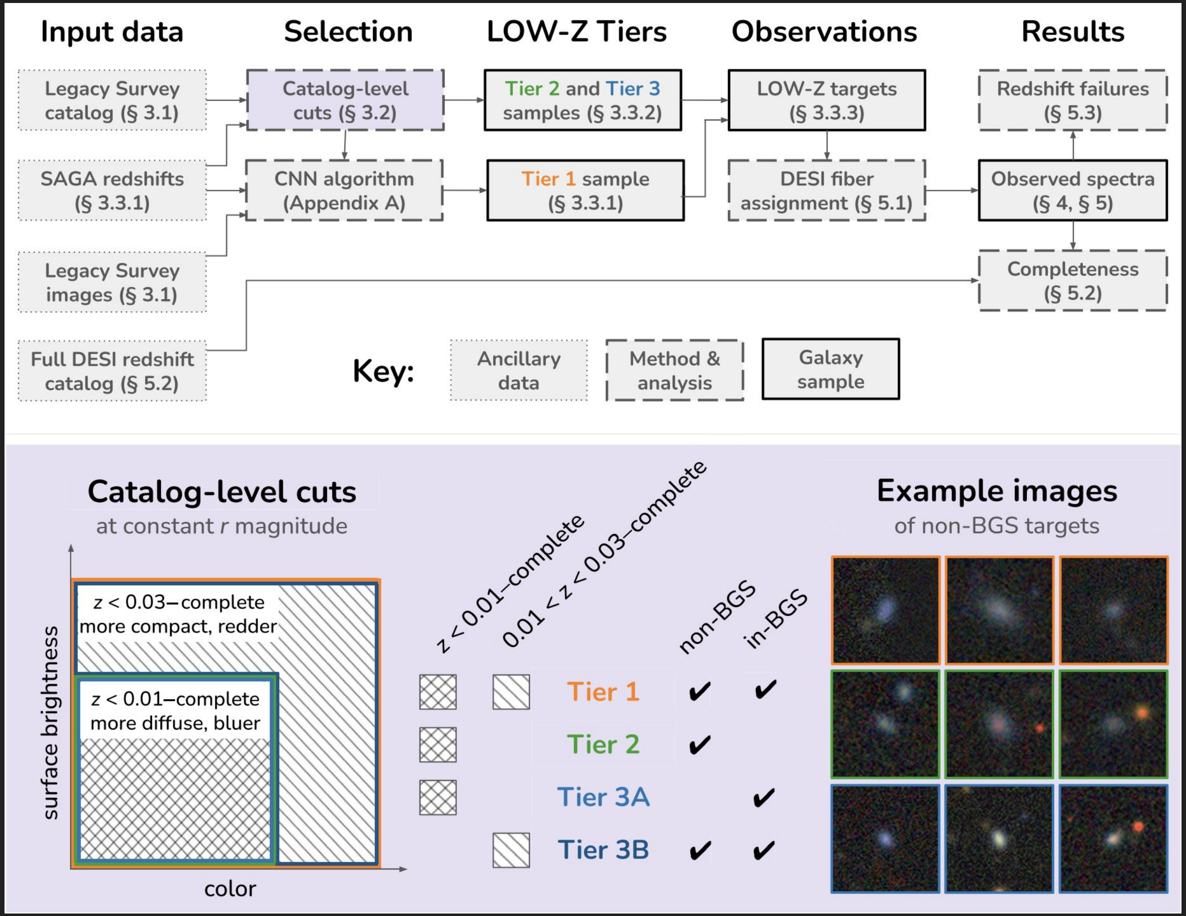
II. The morphologies of galaxies tells us about their physical evolution – and helps to identify rare gems!

III. xSAGA/DESI LOWZ gives us an entirely new way to study substructure of the low-redshift cosmos.

IV. GNNs powerful for learning relationships, like the galaxy-halo-environment connection.

John F Wu · jfwu@stsci.edu · jwuphysics.github.io

DESI LOW-Z: survey design



Darragh-Ford+ 23

DESI LOW-Z: target densities

Table 1
Target Density for the Three Tiers in the LOW-*Z* Survey

Tier	Y1 Targets All	Y1 Targets BGS Overlap	Y1 Observed All	Y1 Observed BGS Overlap	Y2 Targets All	Y2 Targets BGS Overlap
Tier 1	22 deg^{-2}	6 deg^{-2}	11 deg^{-2}	6 deg^{-2}	97 deg^{-2}	1.7 deg^{-2}
Tier 2	80 deg^{-2}	…	41 deg^{-2}	…	325 deg^{-2}	1.3 deg^{-2}
Tier 3A	120 deg^{-2}	120 deg^{-2}	120 deg^{-2}	120 deg^{-2}	…	…
Tier 3B	80 deg^{-2}	…	30 deg^{-2}	…	…	…
$z < 0.03$	…	…	3.7 deg^{-2}	1.6 deg^{-2}	…	…

Notes. Columns 1 and 2: submitted target densities for the Y1 survey. Columns 3 and 4: observed target densities for Y1 survey. All objects in Columns 1–4 are between the Y1 LOW-*Z* magnitude cuts of $19 < r < 21$. Columns 5 and 6: submitted target densities for the Y2 survey. All objects are between the Y2 LOW-*Z* magnitude cuts of $19 < r < 21.15$. The full BGS target density is 1400 targets per square degree (864 deg^{-2} in the Bright sample and 533 deg^{-2} in the Faint sample).

Table 2
Color Cuts for the BGS Bright and BGS Faint Samples (Hahn et al. 2022)

BGS Sample	r	r_{fib}	Color	Density
BGS Bright	$r < 19.5$	$r_{fib} < 22.9$	…	864 deg^{-2}
BGS Faint	$19.5 < r < 20.175$	$r_{fib} < 21.5$ if color $\geqslant 0$ or $r_{fib} < 20.75$	$(z - W1) - 1.2(g - r) + 1.2$	533 deg^{-2}

Darragh-Ford+ 23

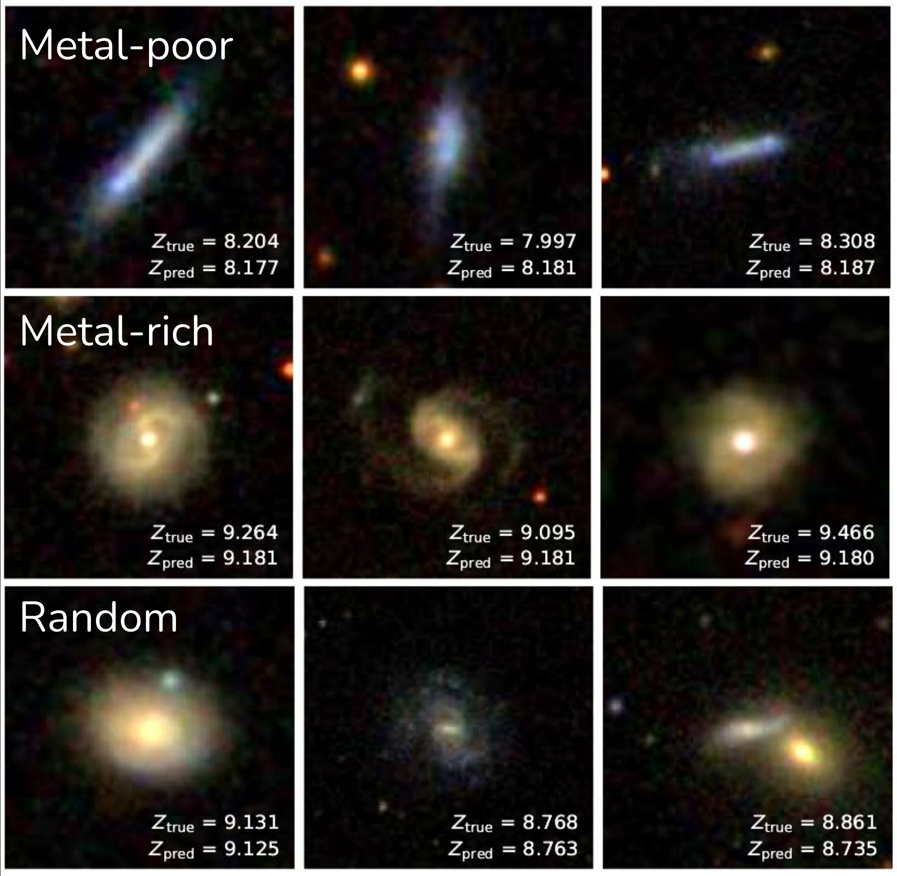
DESI LOW-Z: target densities

Table 3
Observed Number of Targets in LOW-Z Survey Split by Tier and Redshift

	$z < 0.01$	$z < 0.03$	All Redshifts
One-Percent Survey			
Tier 1	26	382	2015
Tier 1 (excl. BGS)	12	167	992
Tier 2	5	100	7445
Tier 3	3	179	27,021
Tier 3 (excl. BGS)	2	37	5906
Main Survey			
Tier 1	53	875	4618
Tier 1 (excl. BGS)	4	34	163
Tier 2	1	22	2034
Tier 3	3	461	100,353
Tier 3 (excl. BGS)	0	4	1413

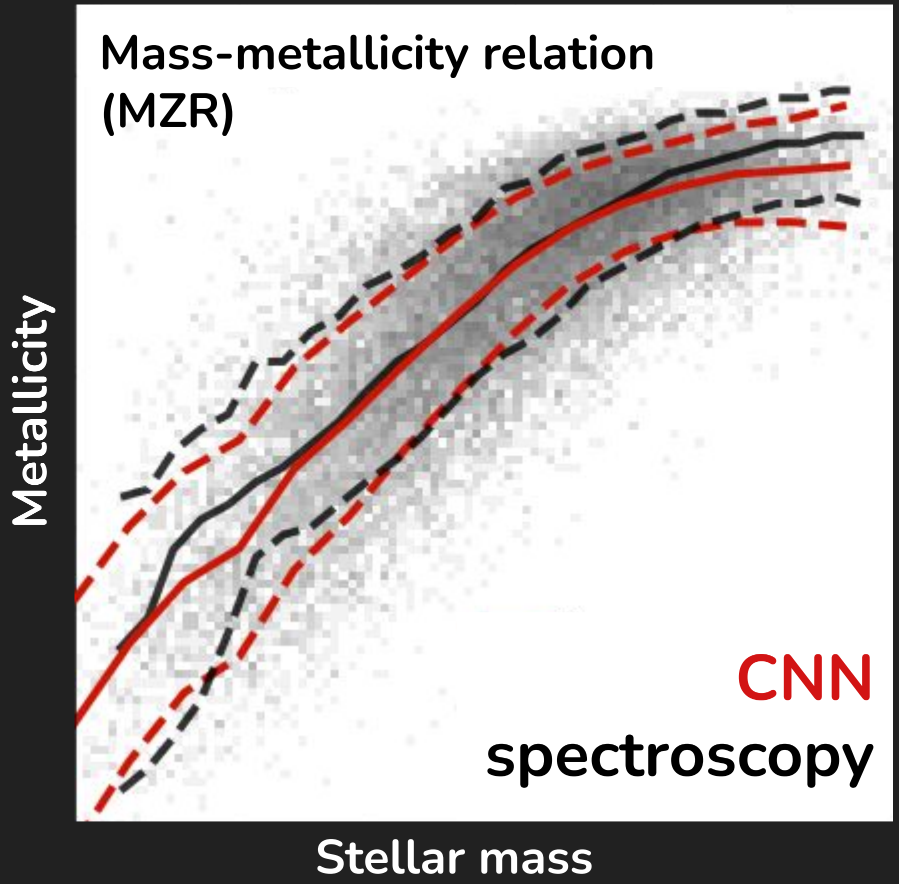
Darragh-Ford+ 23

CNNs can estimate spectroscopic properties like metallicity!



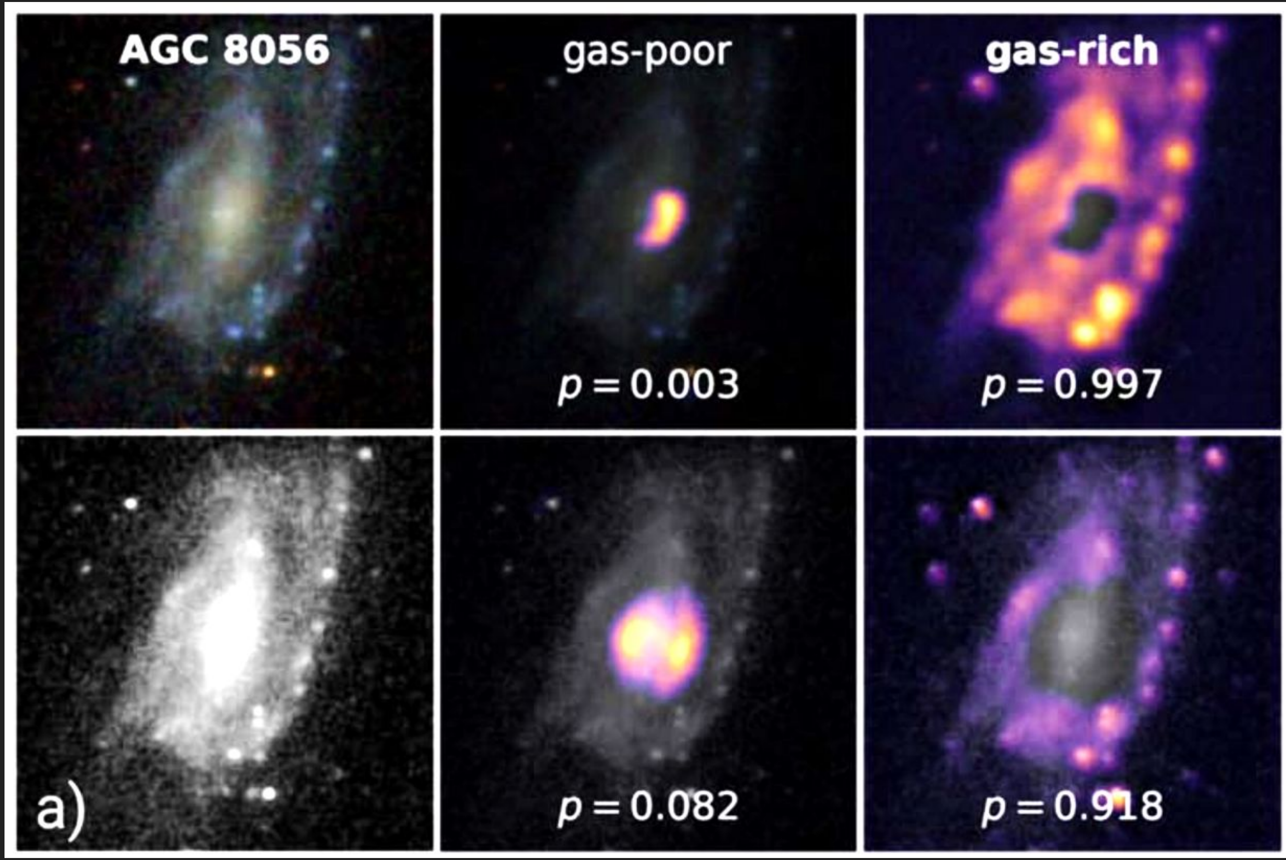
Wu & Boada 19

Re-constructing the MZR *without any spectroscopy*



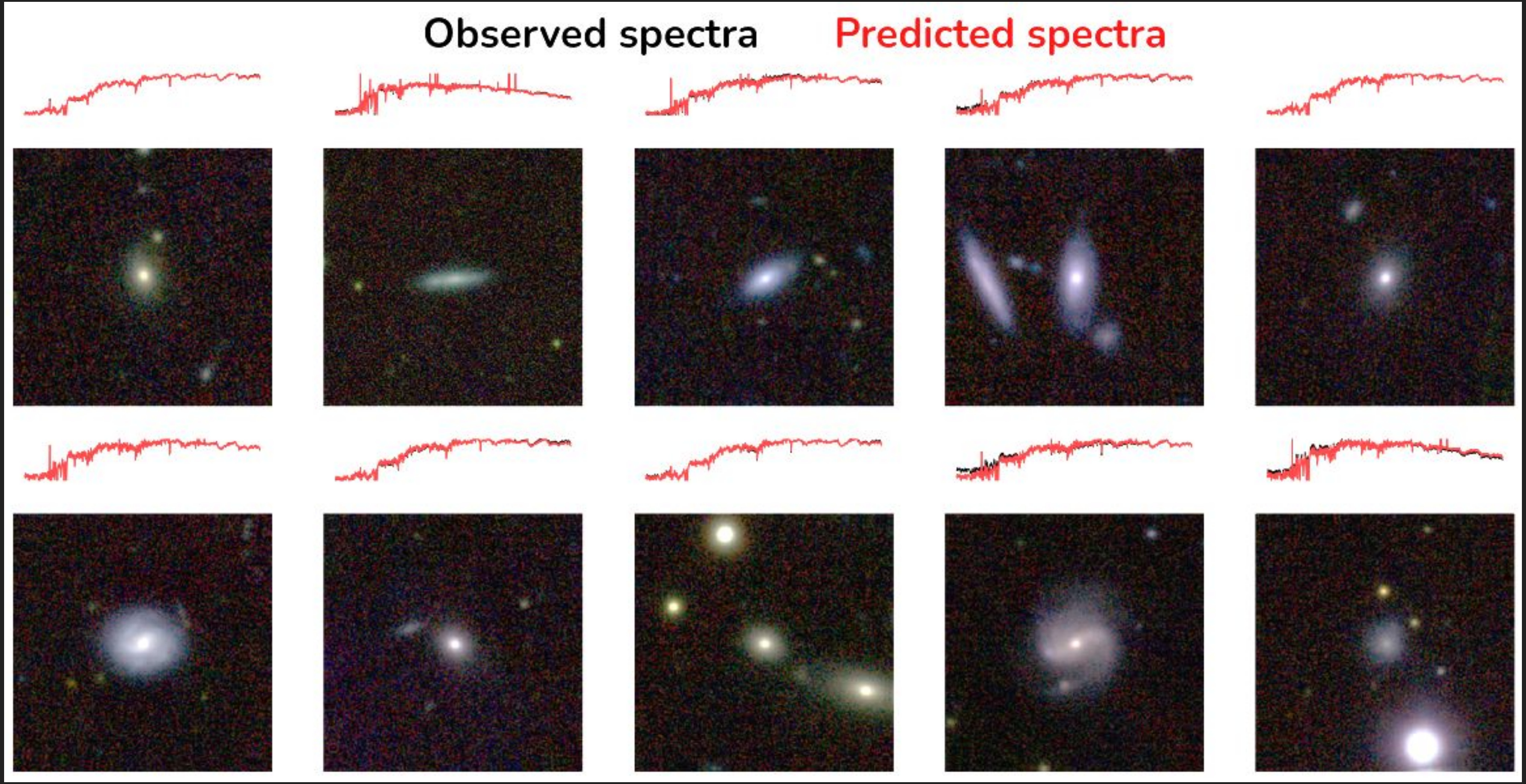
Wu & Boada 19

We know what CNNs are "looking" at!



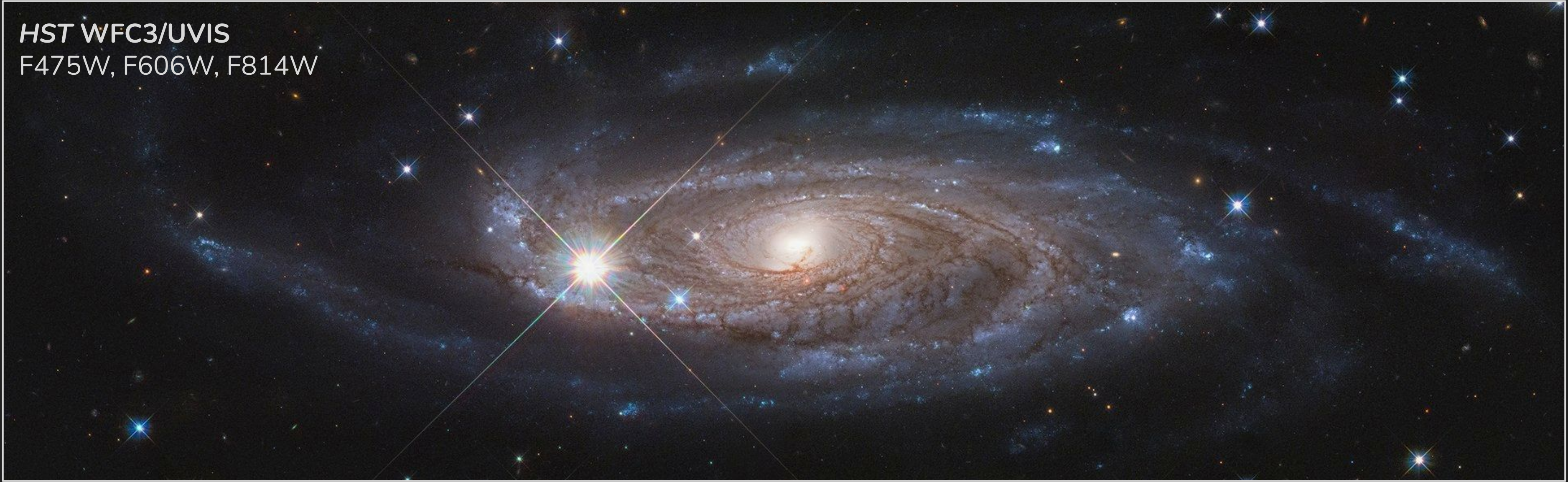
Wu 20

Predict the entire optical spectrum from Pan-STARRS imaging



Wu & Peek 20

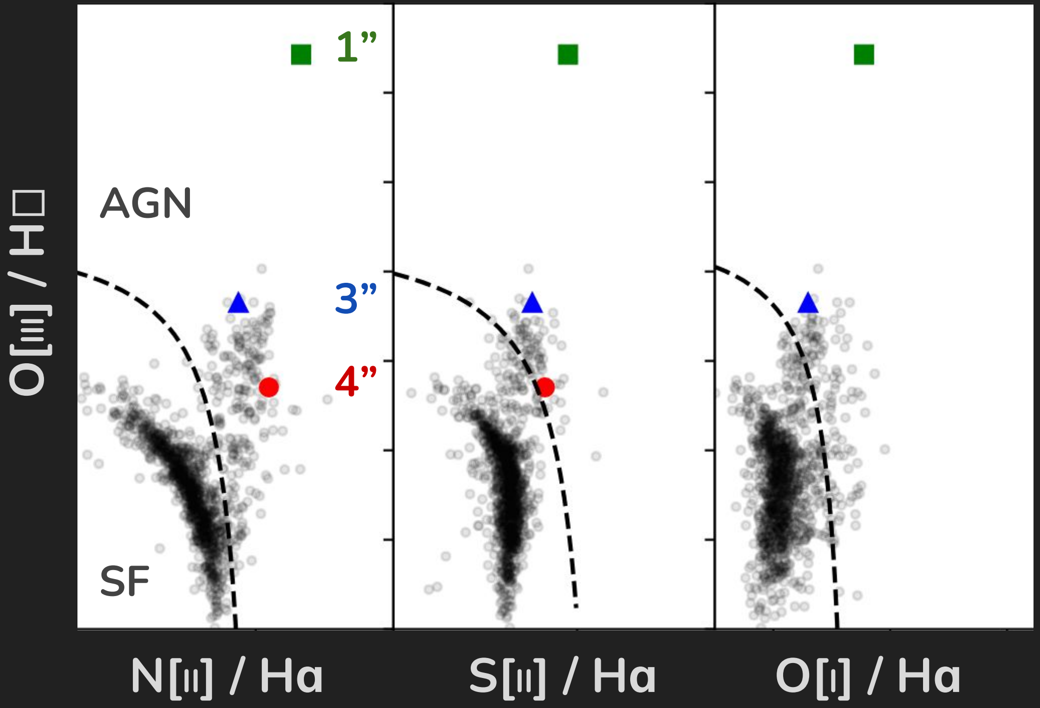
Passing the test: a weak AGN detected in an outlier galaxy



70 kpc

Holwerda+ 21

Passing the test: a weak AGN detected in an outlier galaxy



VIRUS-P + KPNO 2.1m + Mount Lemmon 60in

CNN prediction

MMT Binospec

Holwerda+ 21