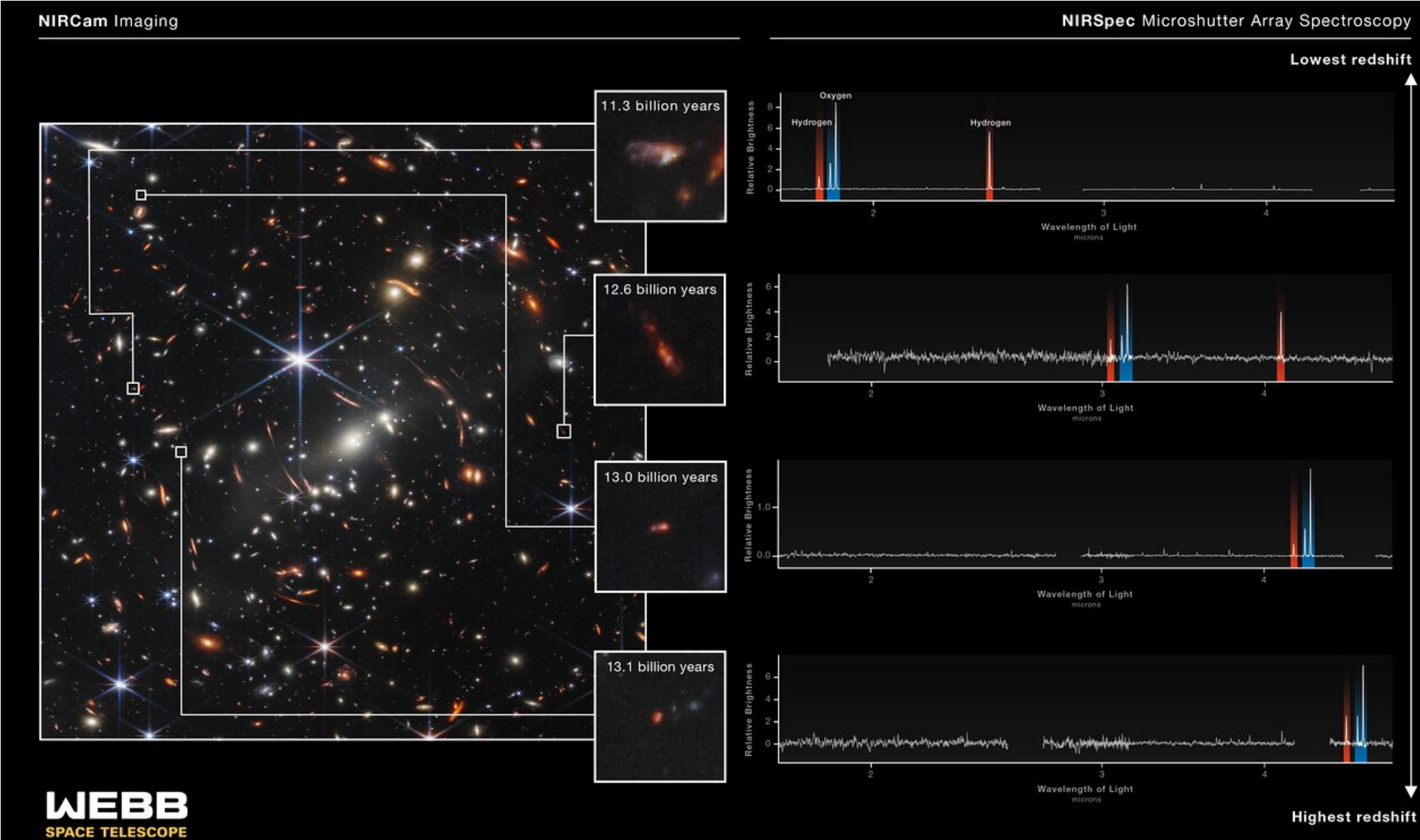


Ivelina Momcheva with Yang Cheng, Morgan Fouesneau, Tobias Buck

More Than the Sum of Its Parts

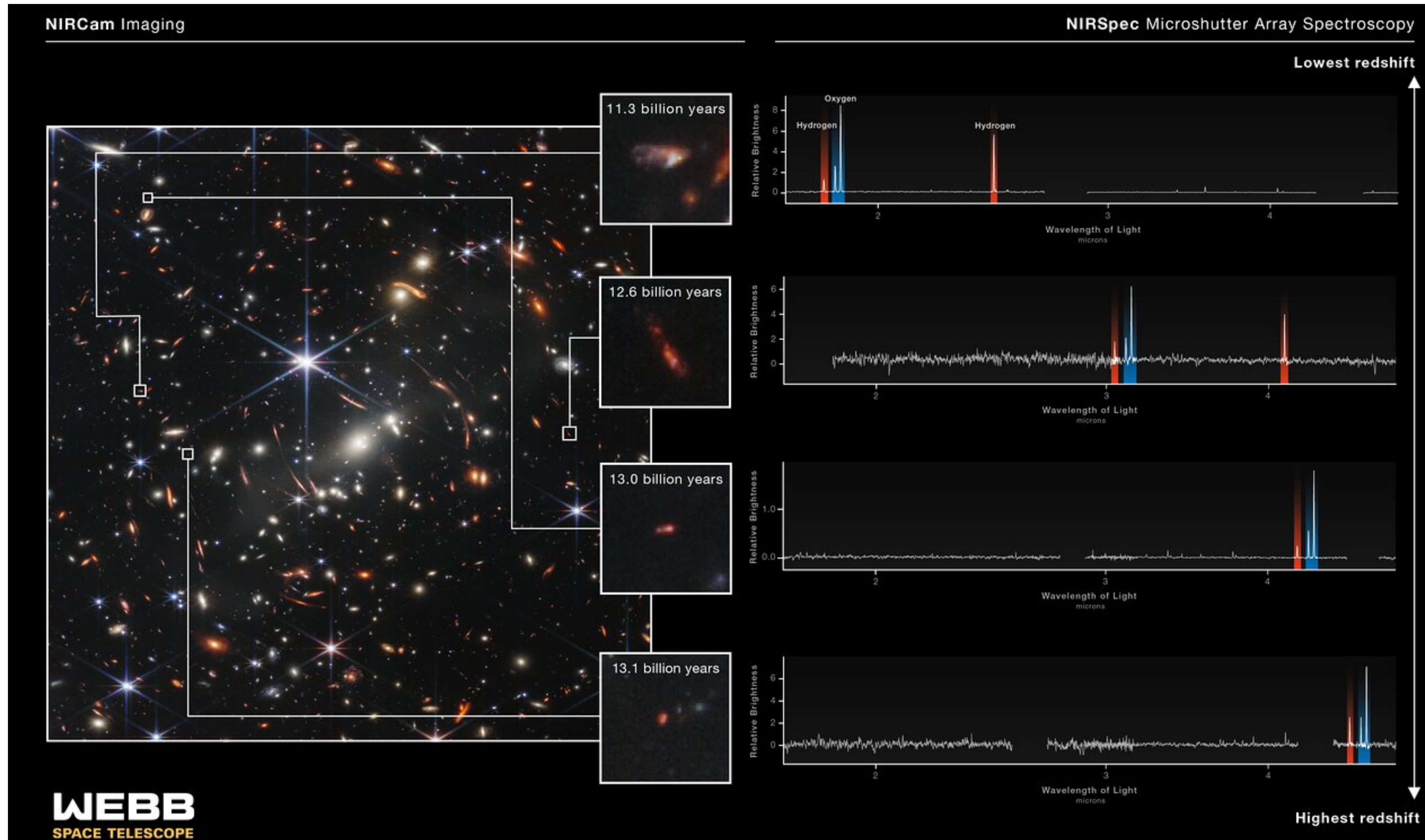
Multimodal Learning for Galaxy Evolution in JWST Pure Parallel Observations

JWST Data is Exquisite



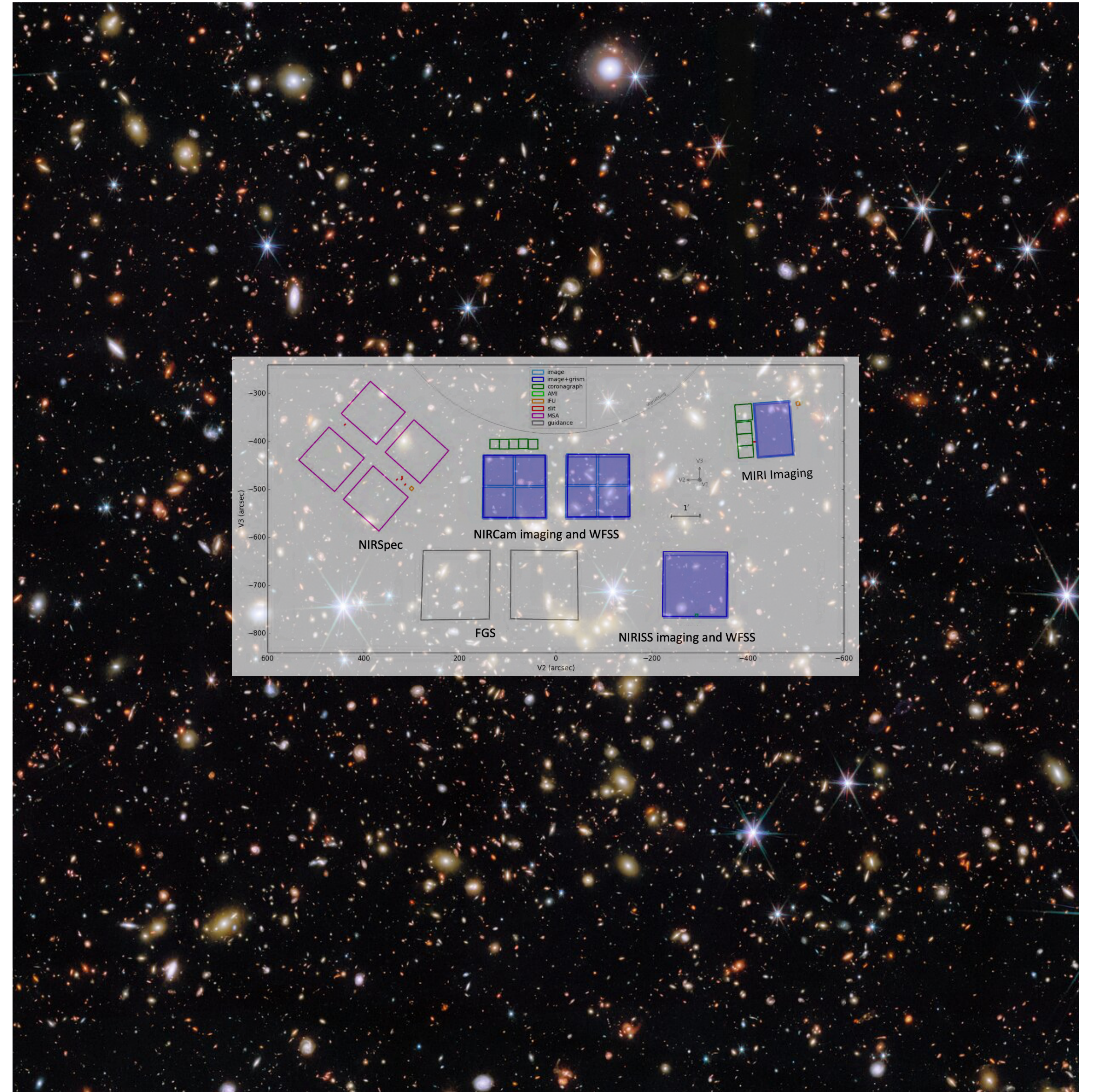
JWST Data is Exquisite, but it is also...

- Narrow FOV
- Limited to well-known areas
- Deep



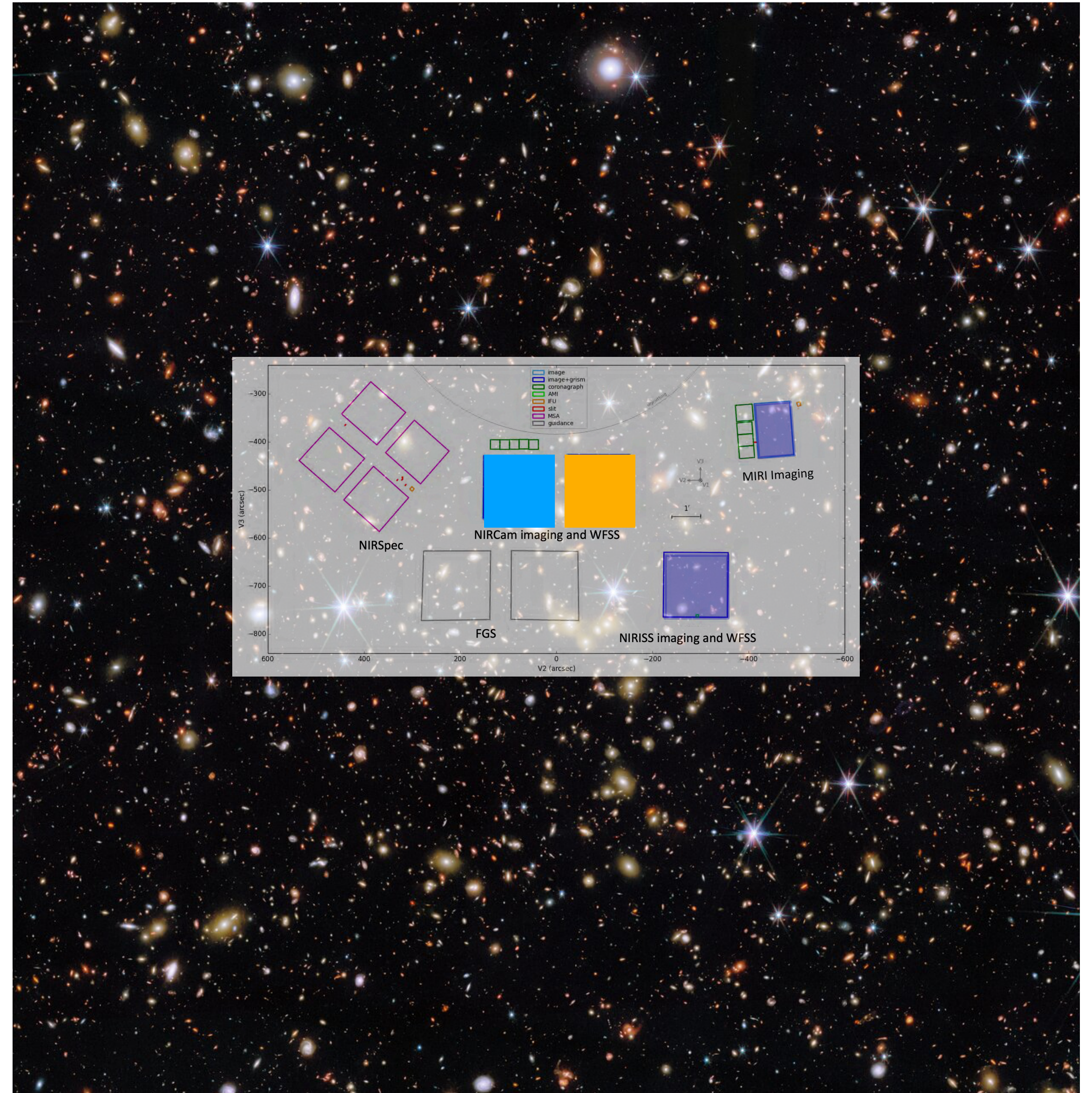
JWST Pure Parallel

- Allow for parallel observations with a second instrument
- Optimize time use on the telescope
- Cannot select position/exposure but can control optical elements
- ... but many/most programs use wide field slitless modes
- Have accumulated >1000 sq. arcmin and > 1000 hours of data!!!



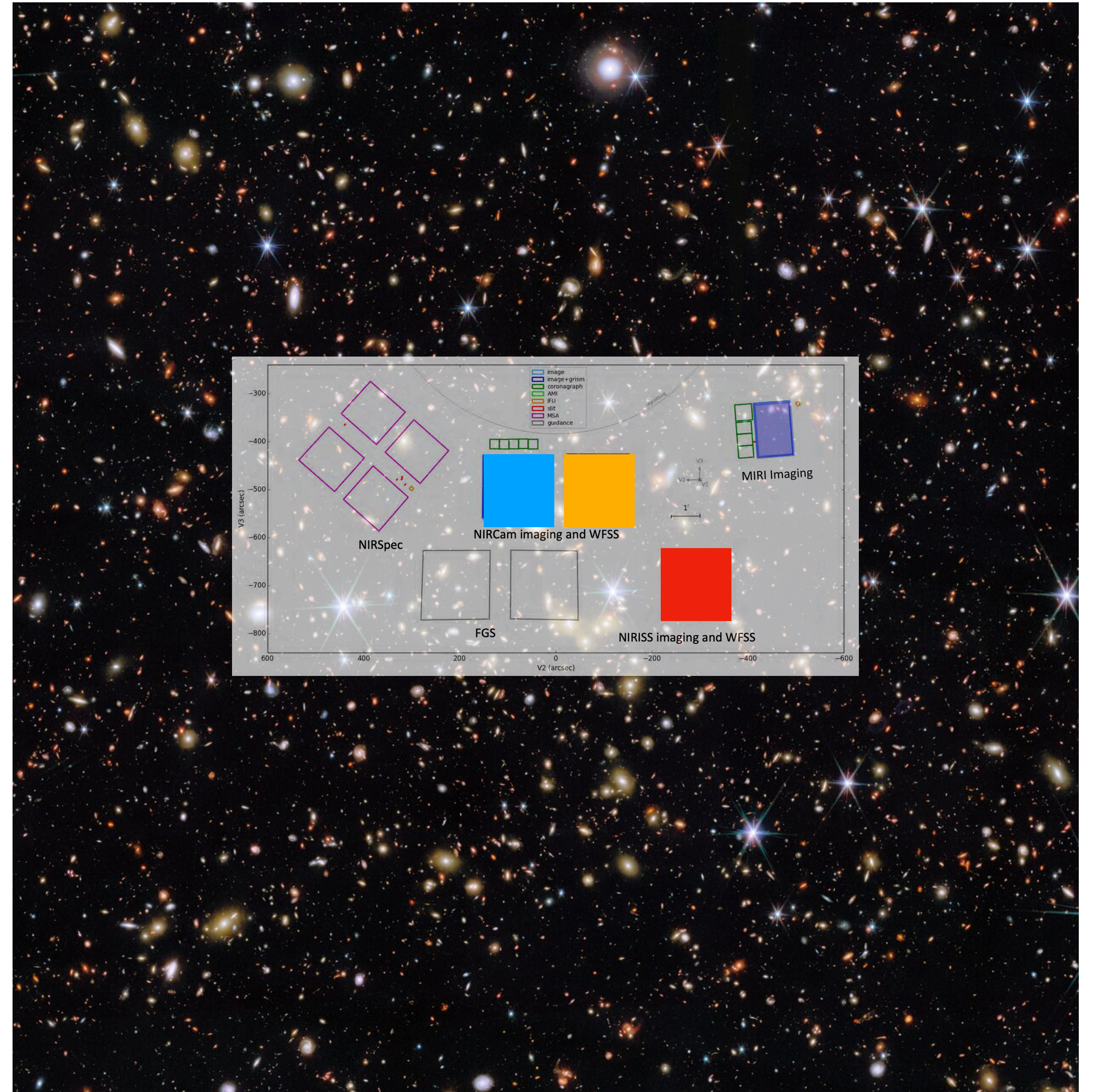
JWST Pure Parallel

- Allow for parallel observations with a second instrument
- Optimize time use on the telescope
- Cannot select position/exposure but can control optical elements
- ... but many/most programs use wide field slitless modes
- Have accumulated >1000 sq. arcmin and > 1000 hours of data!!!



JWST Pure Parallel

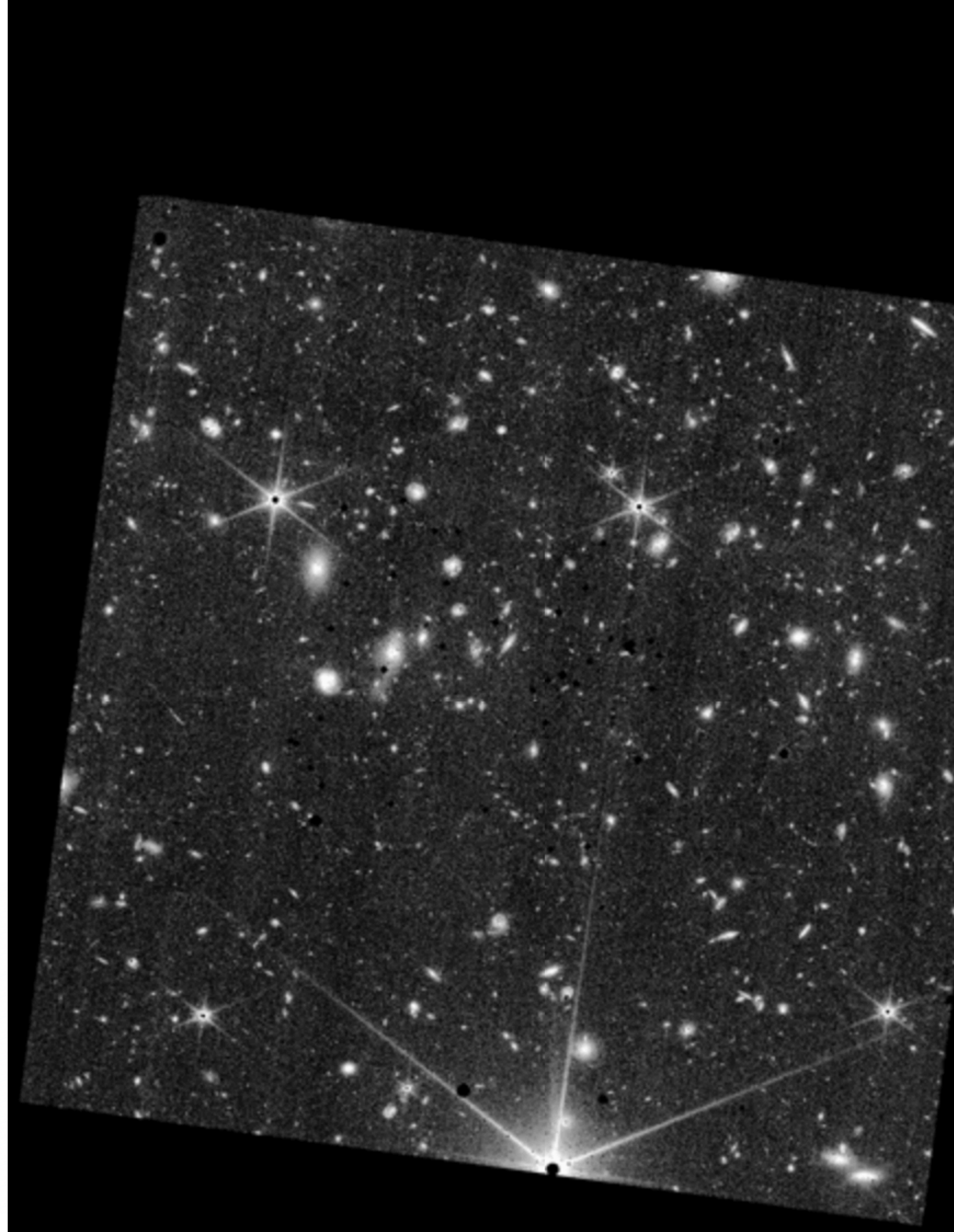
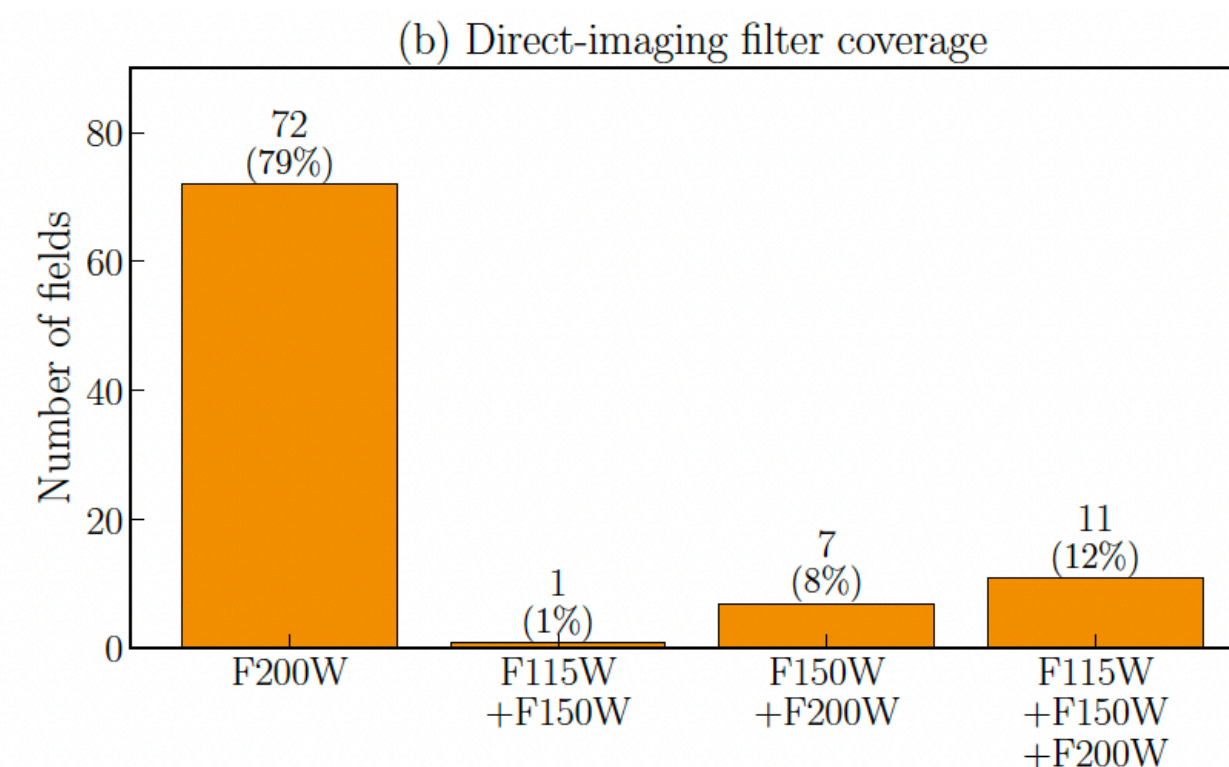
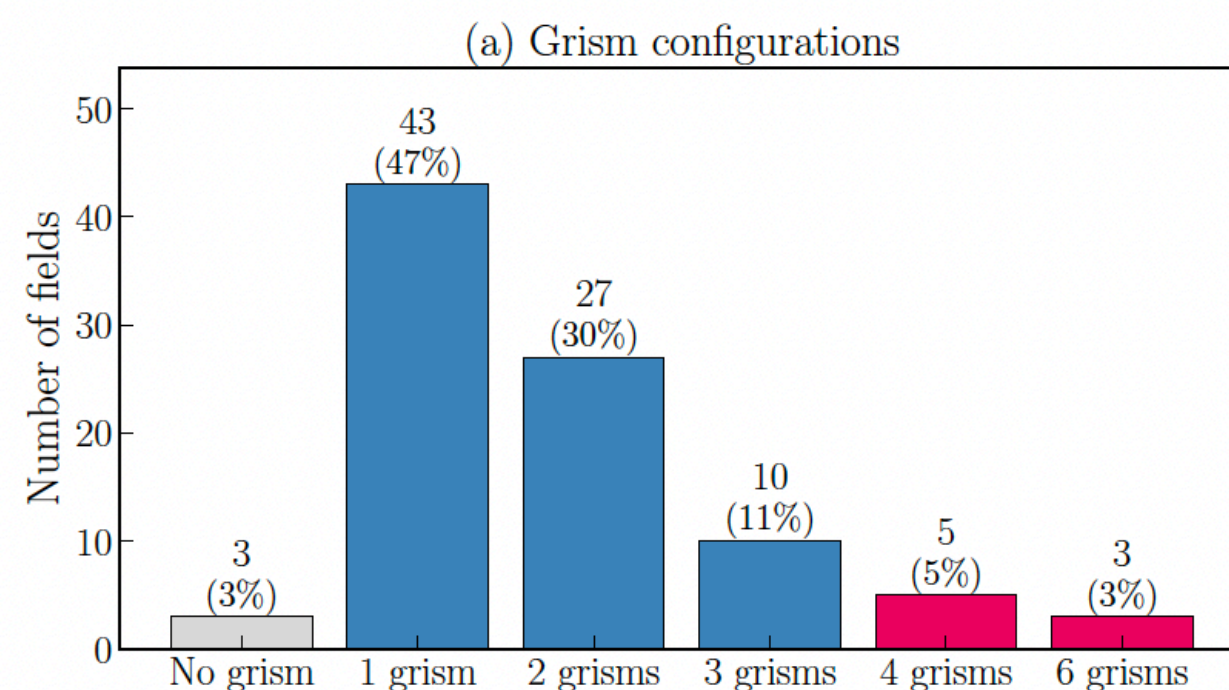
- Allow for parallel observations with a second instrument
- Optimize time use on the telescope
- Cannot select position/exposure but can control optical elements
- ... but many/most programs use wide field slitless modes
- Have accumulated >1000 sq. arcmin and > 1000 hours of data!!!



OutThere + Passage

Pure Parallel Surveys w/ NIRISS

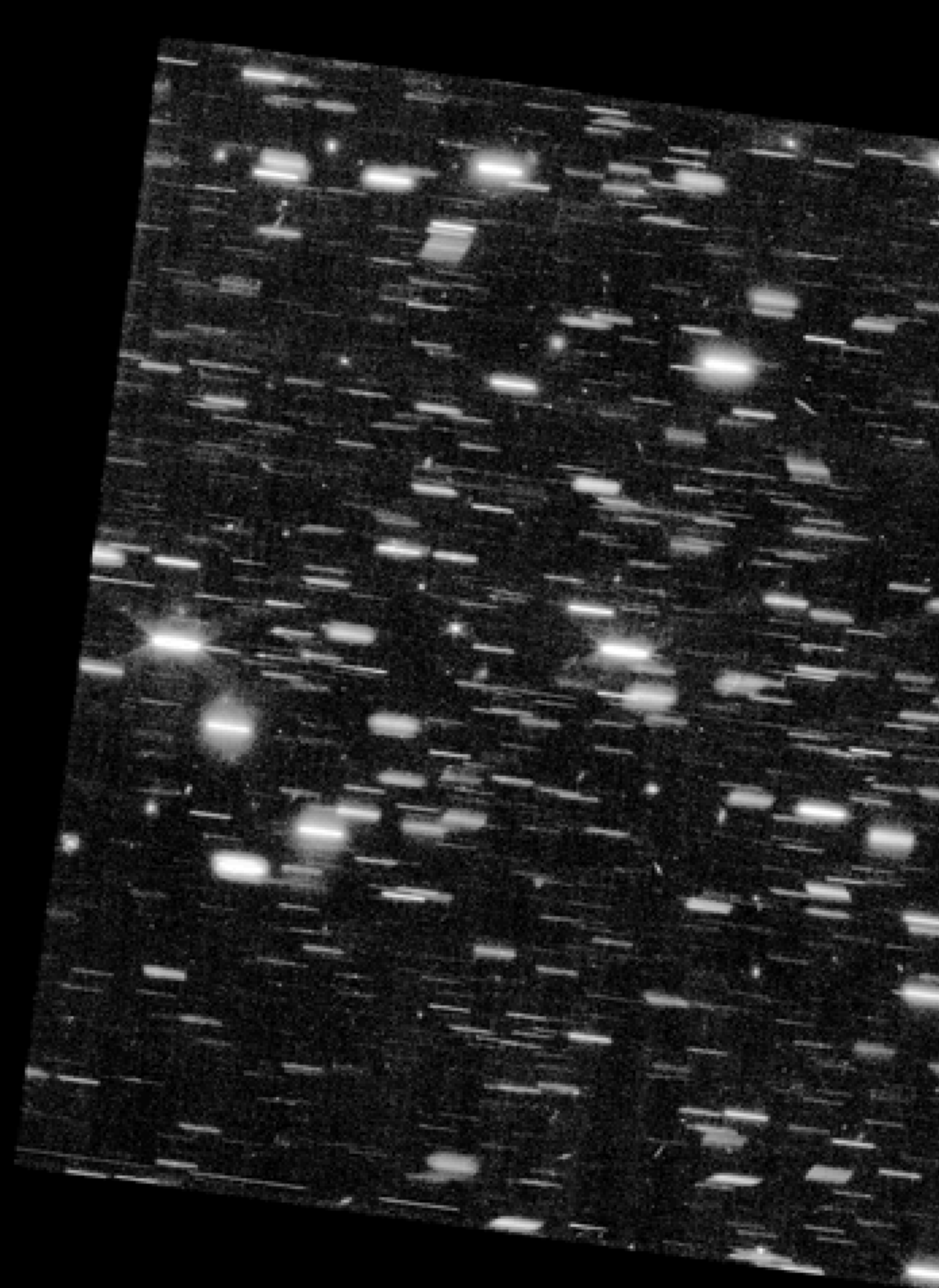
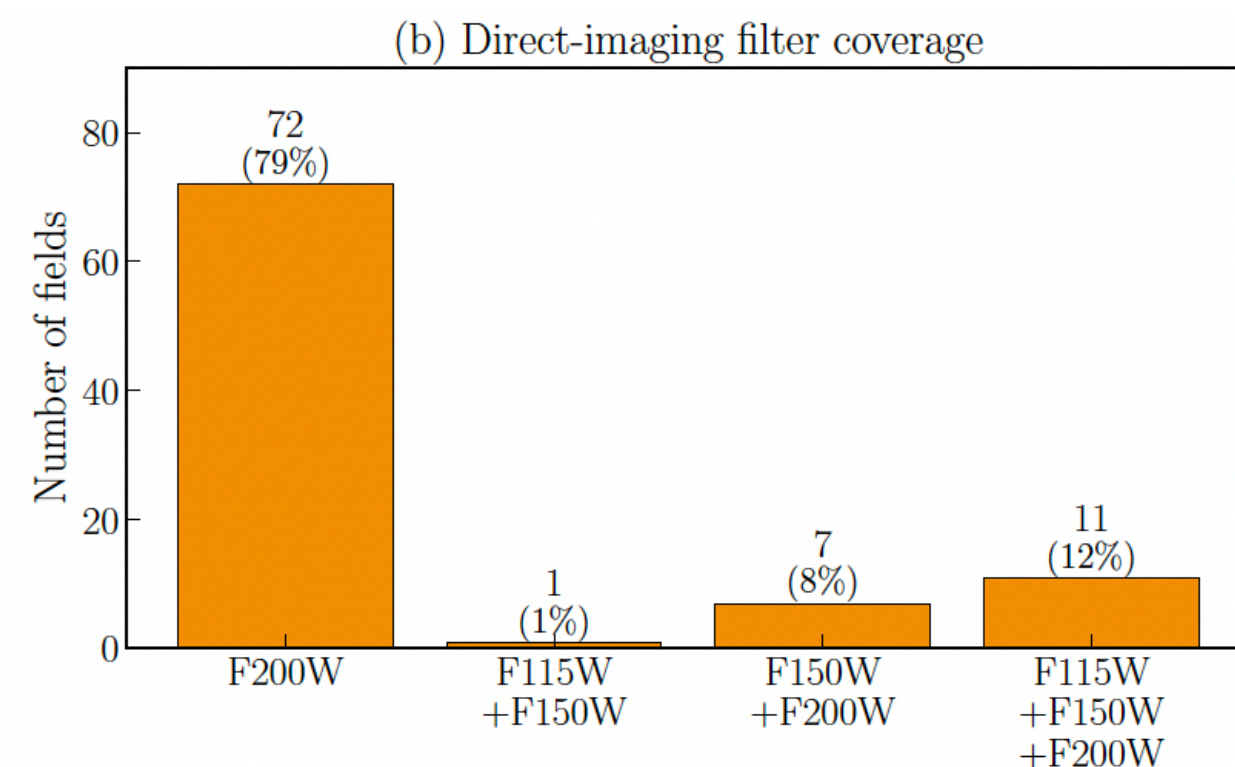
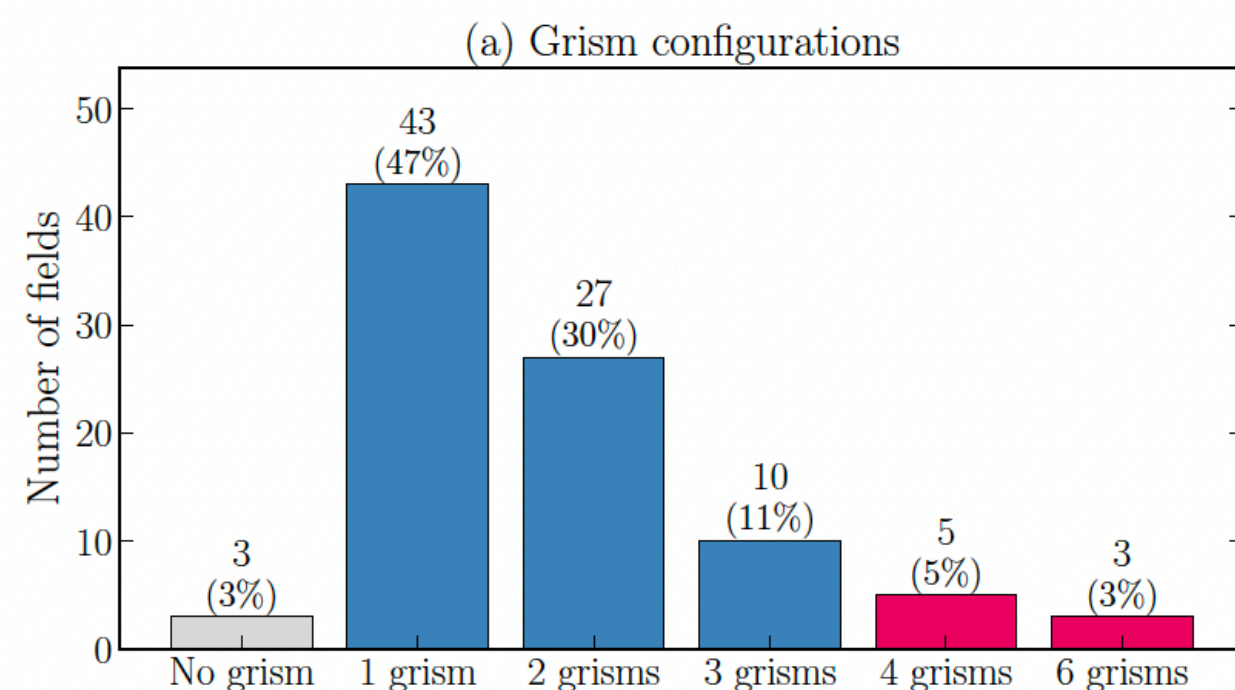
- >100 independent fields
- Total area: > 1000 arcmin²
- Median field area: 4.8 arcmin²
- Includes GOODS-N, COSMOS, UDS
- Science drivers: clustering, star formation, quenching, environments, galaxy structure, rare populations



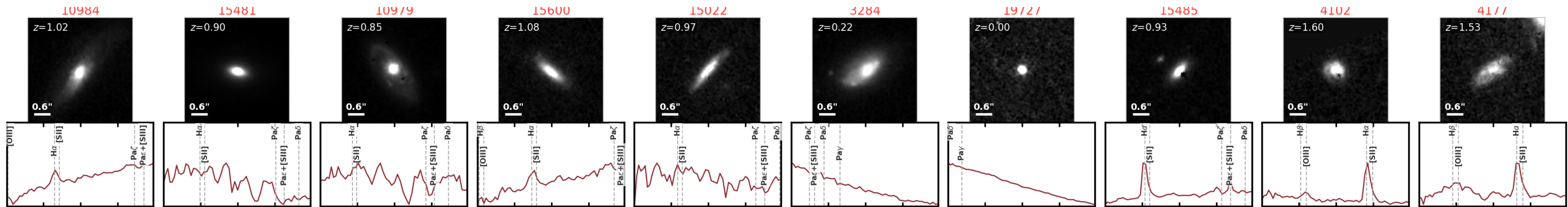
OutThere + Passage

Pure Parallel Surveys w/ NIRISS

- >100 independent fields
- Total area: > 1000 arcmin²
- Median field area: 4.8 arcmin²
- Includes GOODS-N, COSMOS, UDS
- Science drivers: clustering, star formation, quenching, environments, galaxy structure, rare populations



The Problem: (Almost) No Ancillary Data



The Solution: Multimodal Learning



Yang Cheng
MSc Student



Morgan Fouesneau
Data Science MPIA



Tobias Buck
ML/CS

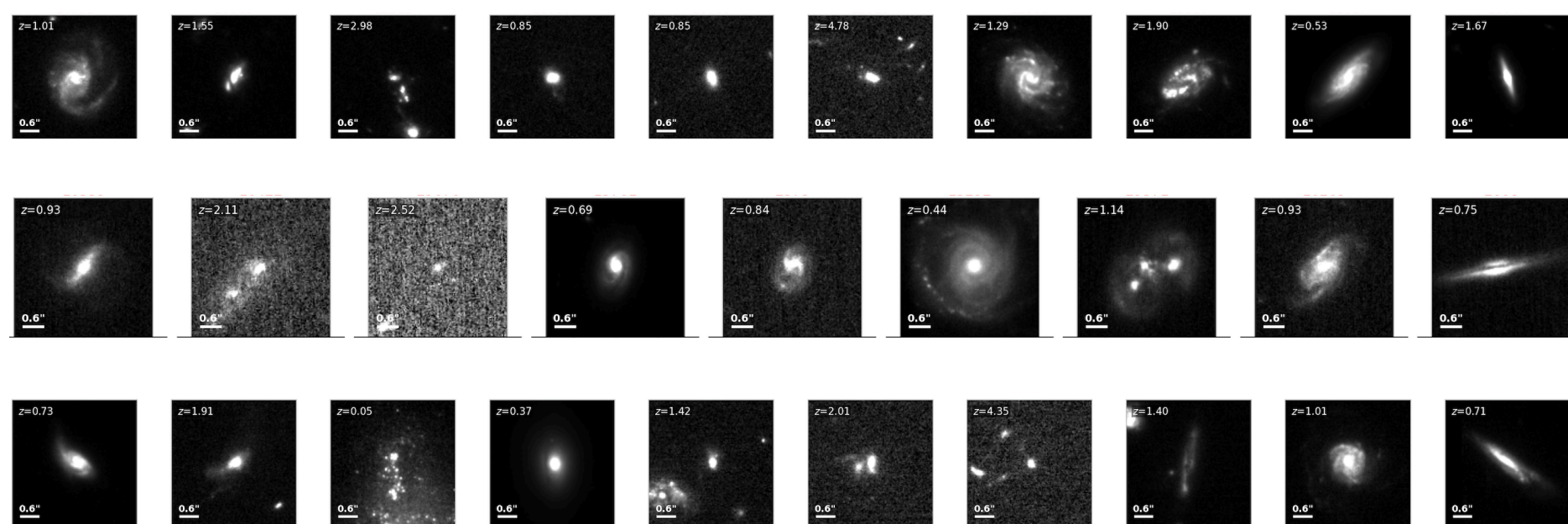
The Solution: Multimodal Learning

Data

Explore the Archive

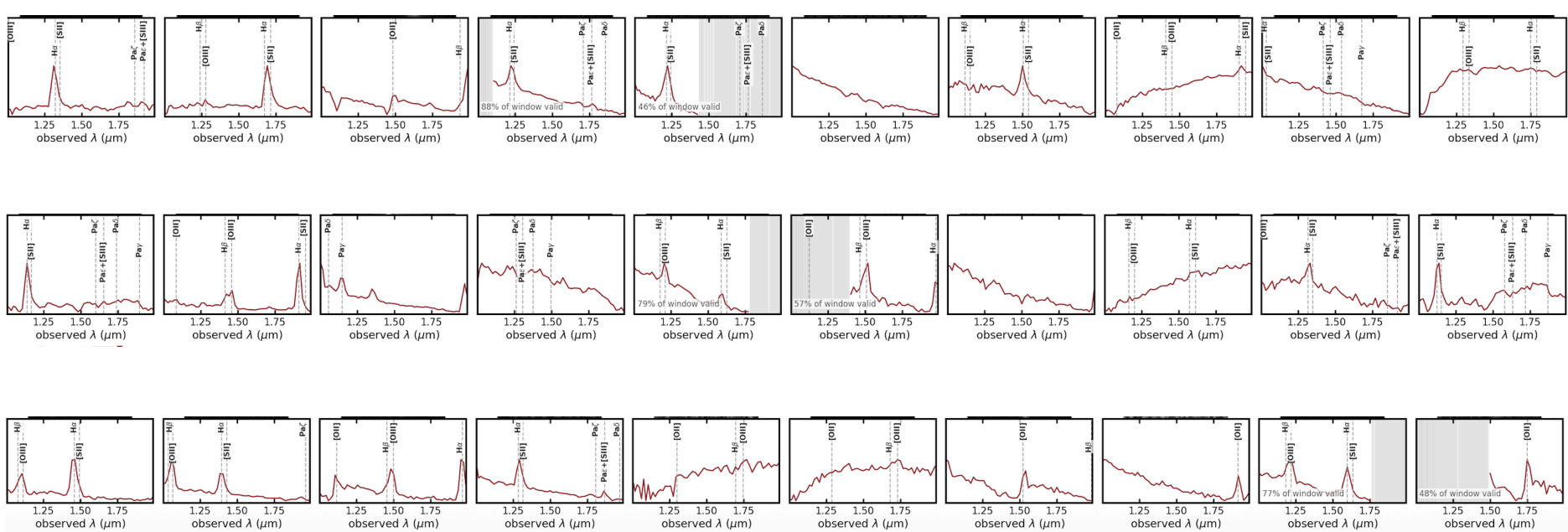
Imaging:

1,400,000 NIRCam + NIRISS F150W cutouts from CEERS, COSMOS-Web, OutThere, 128x128 px @ 30 mas/pix



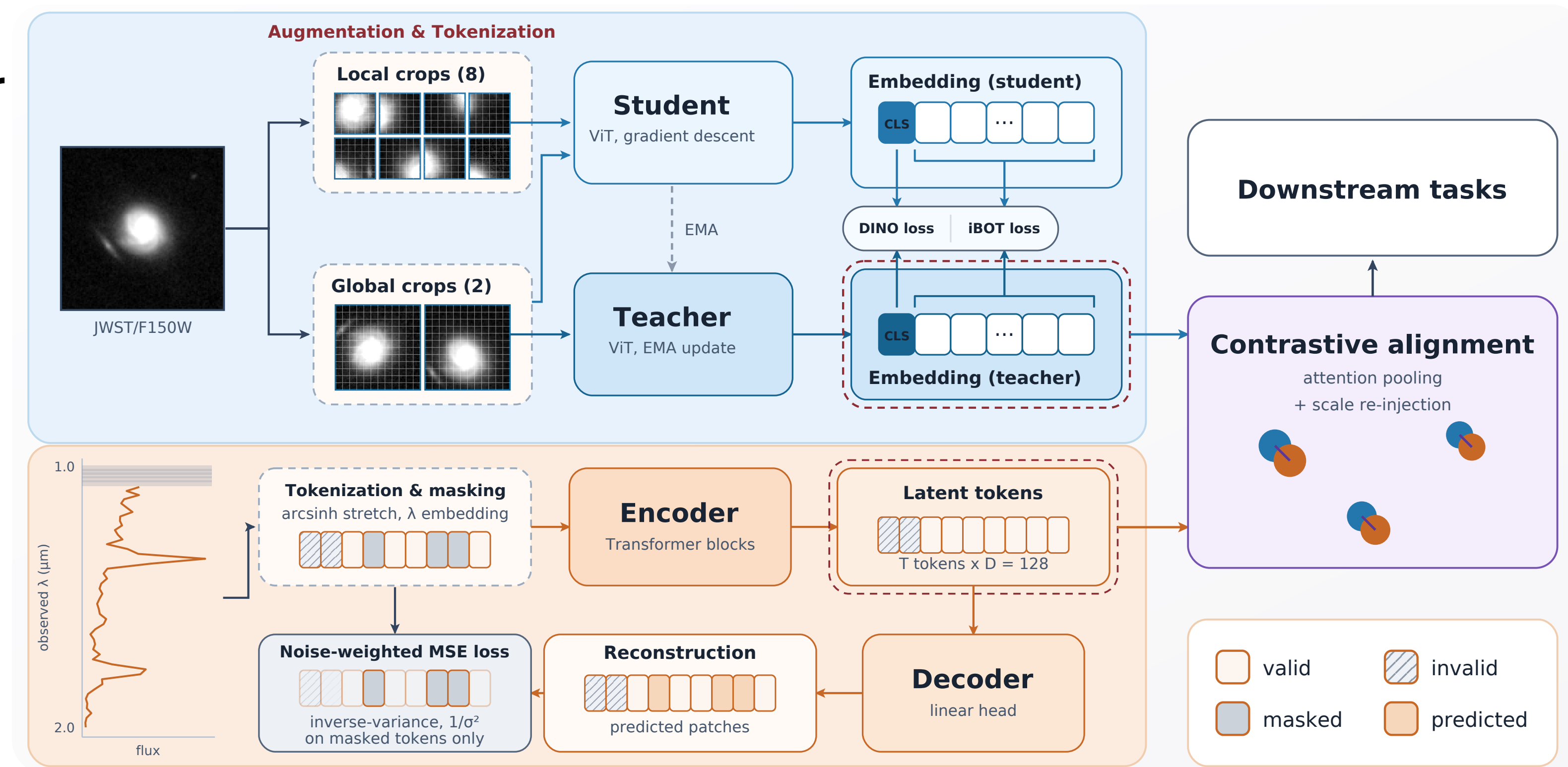
Spectra:

25,000 spectra, R~100 PRISM NIRSpec cropped to 1 - 2 μ m, 56 pixels

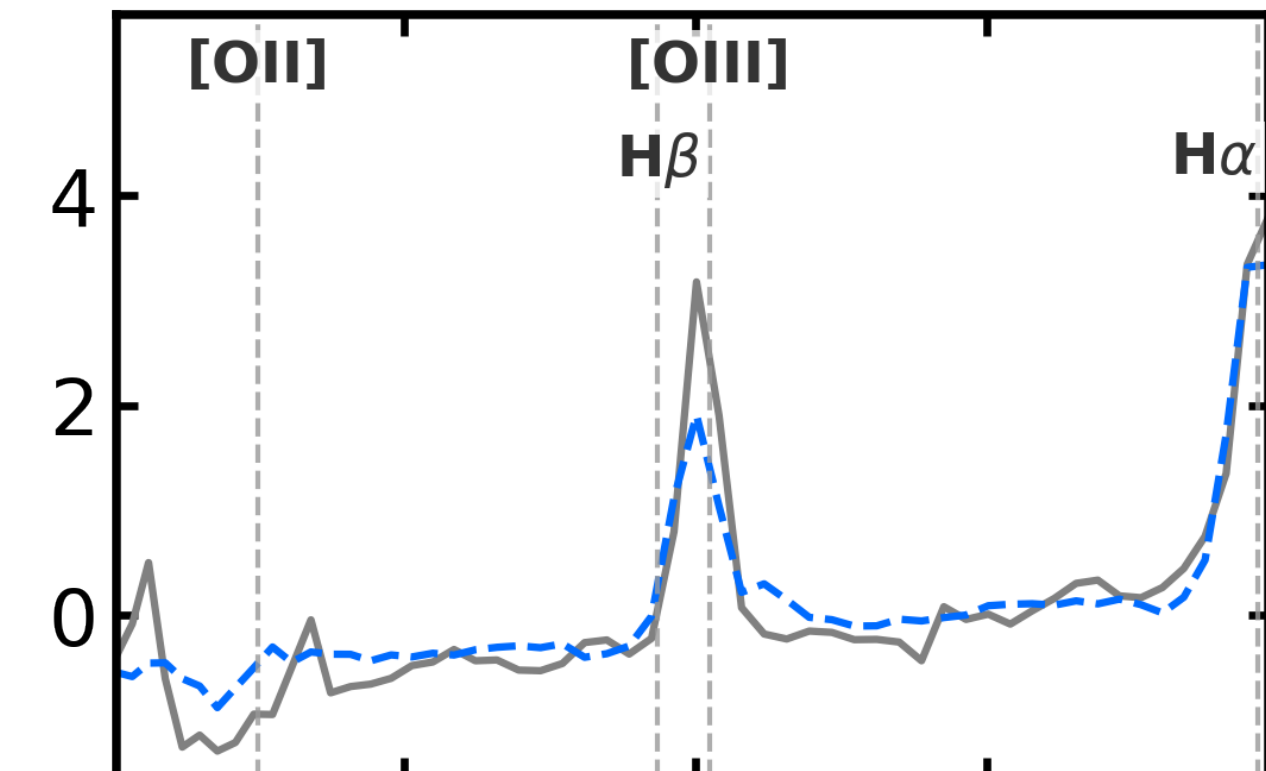
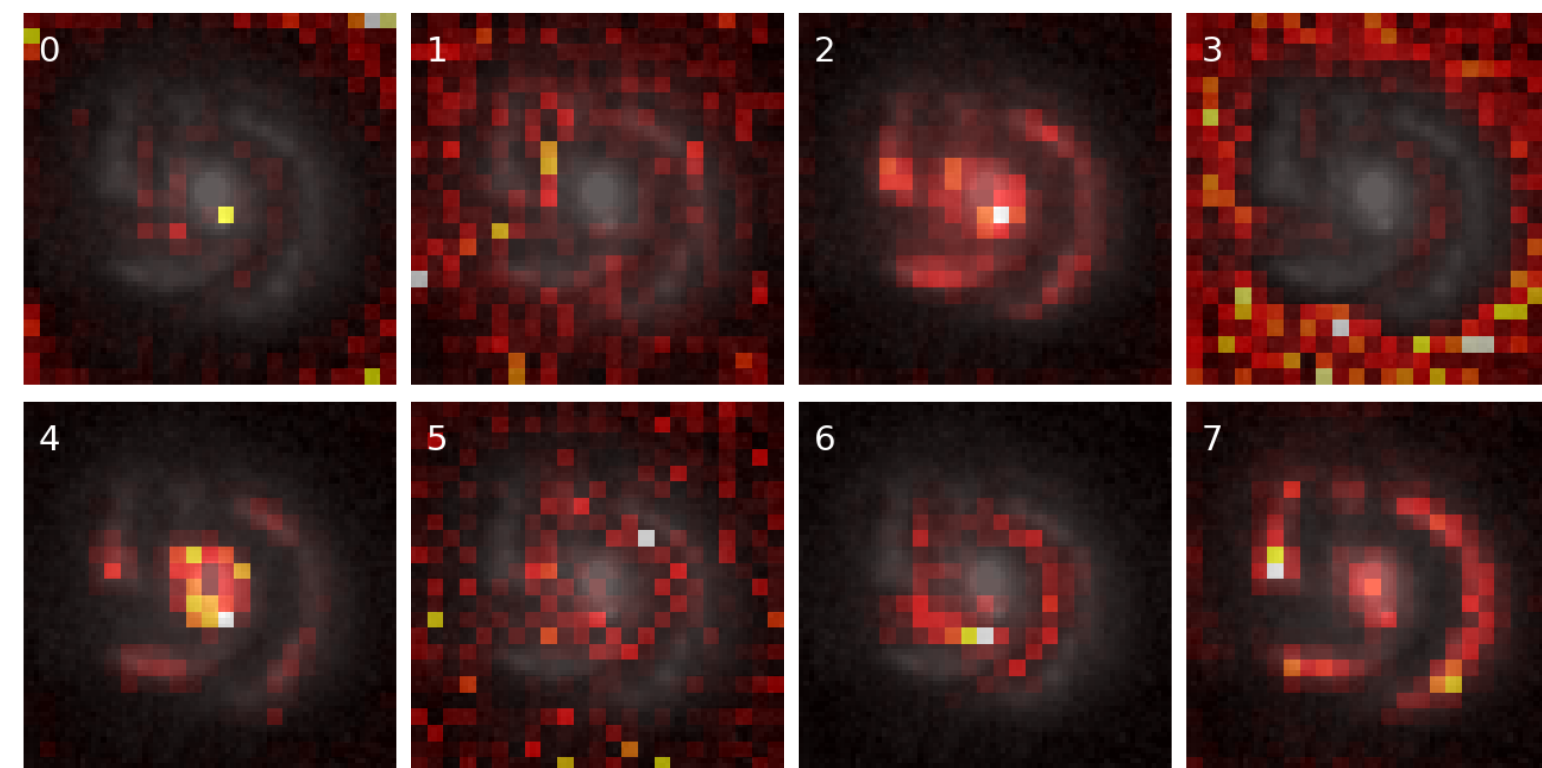
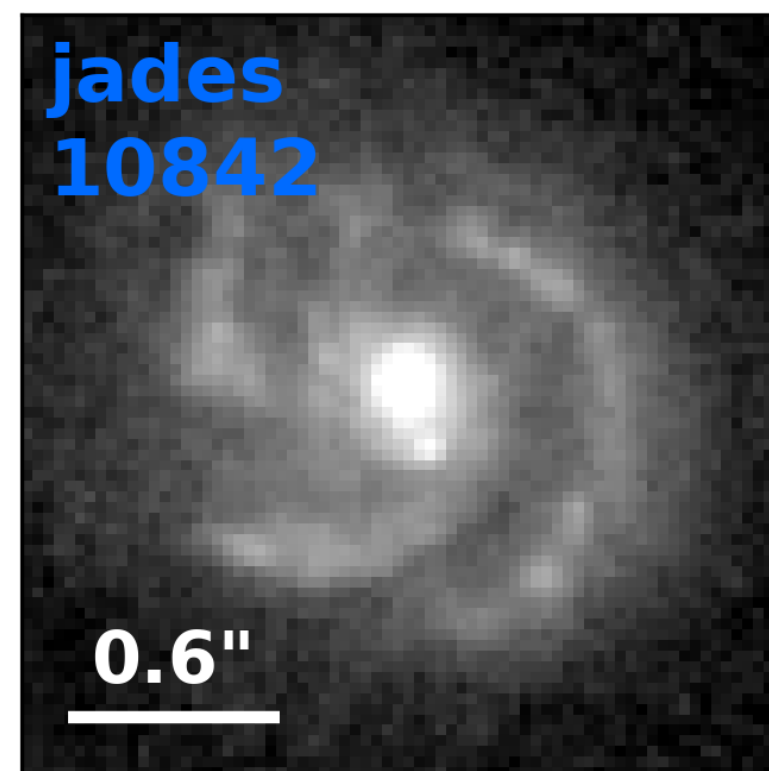
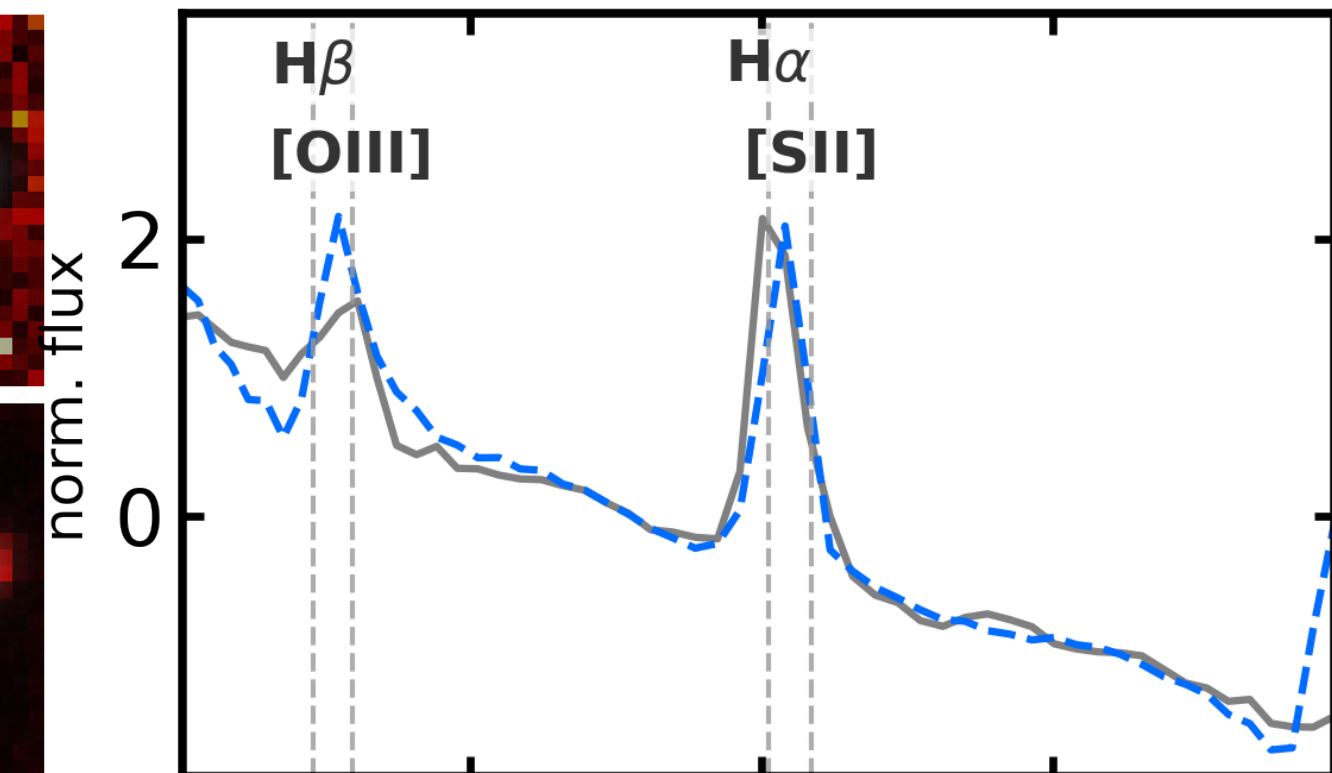
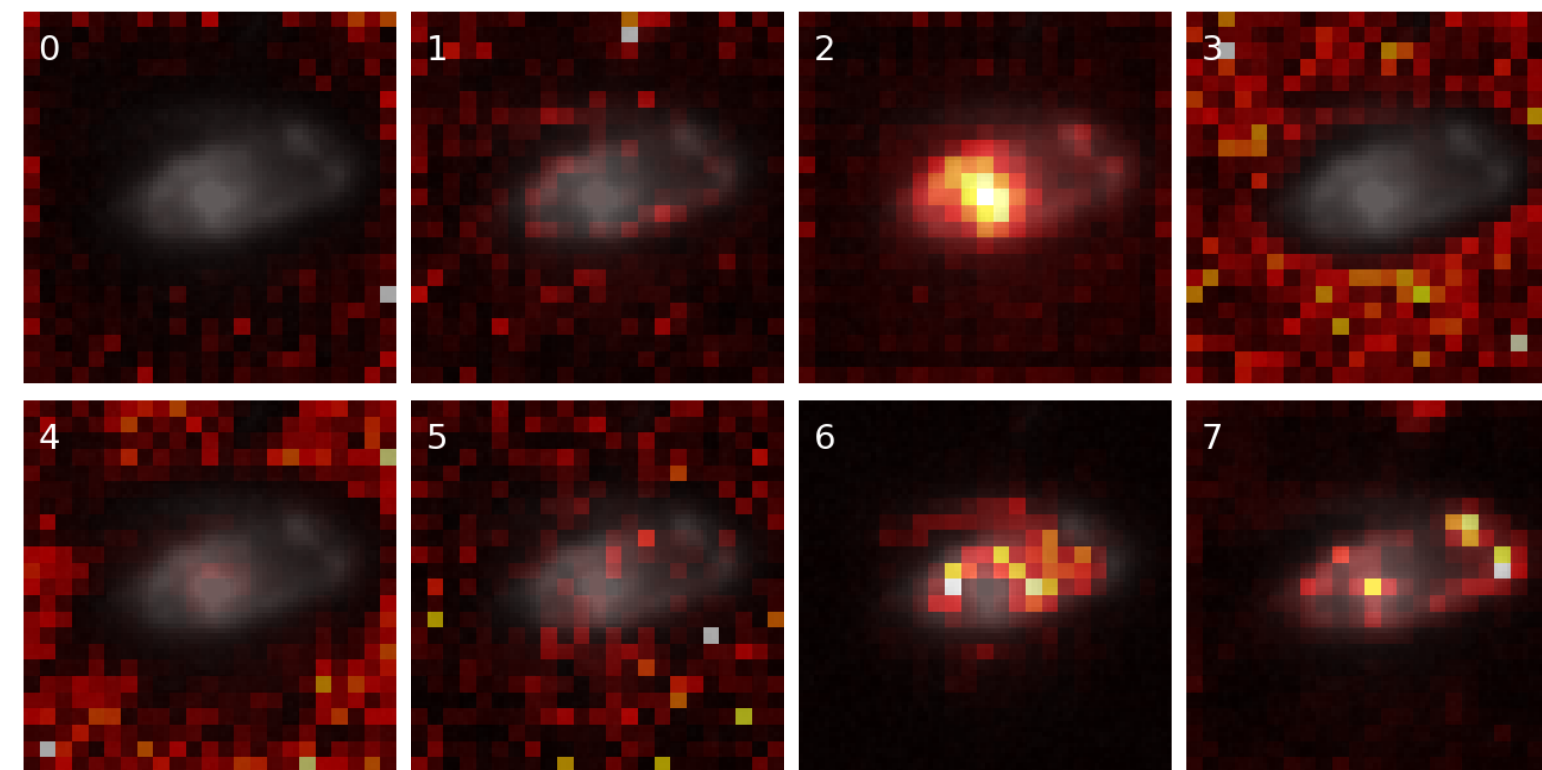
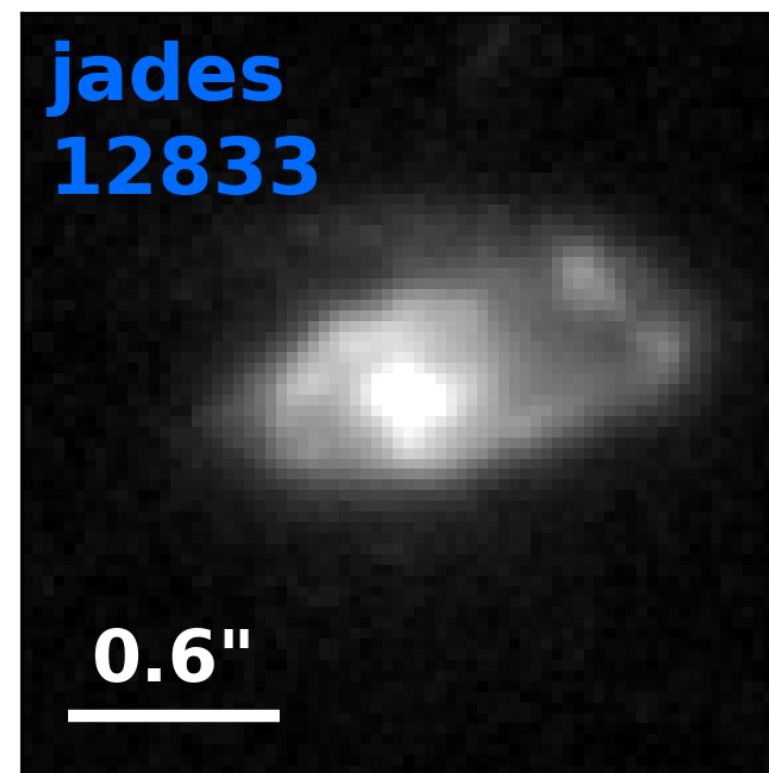


The Framework

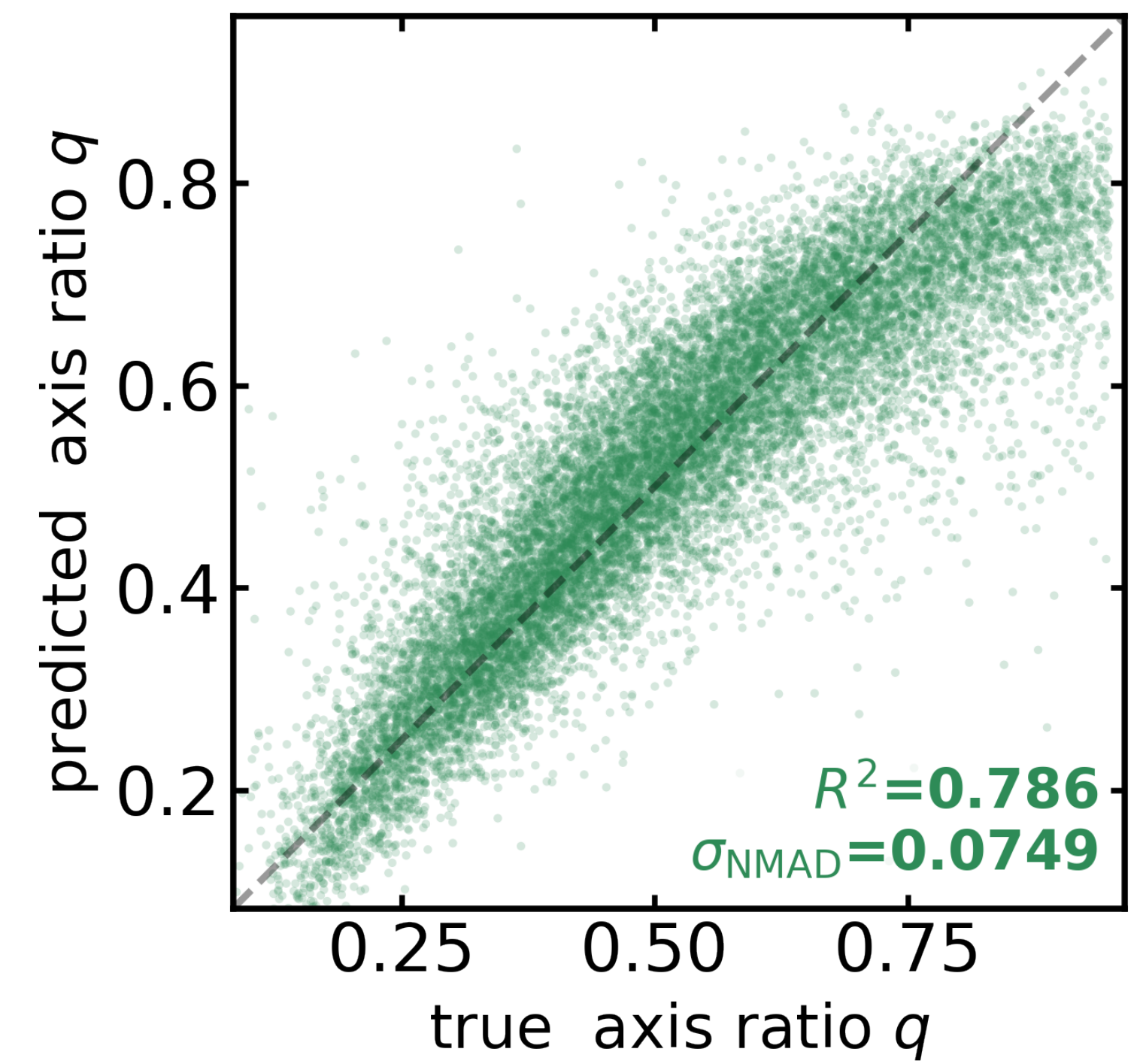
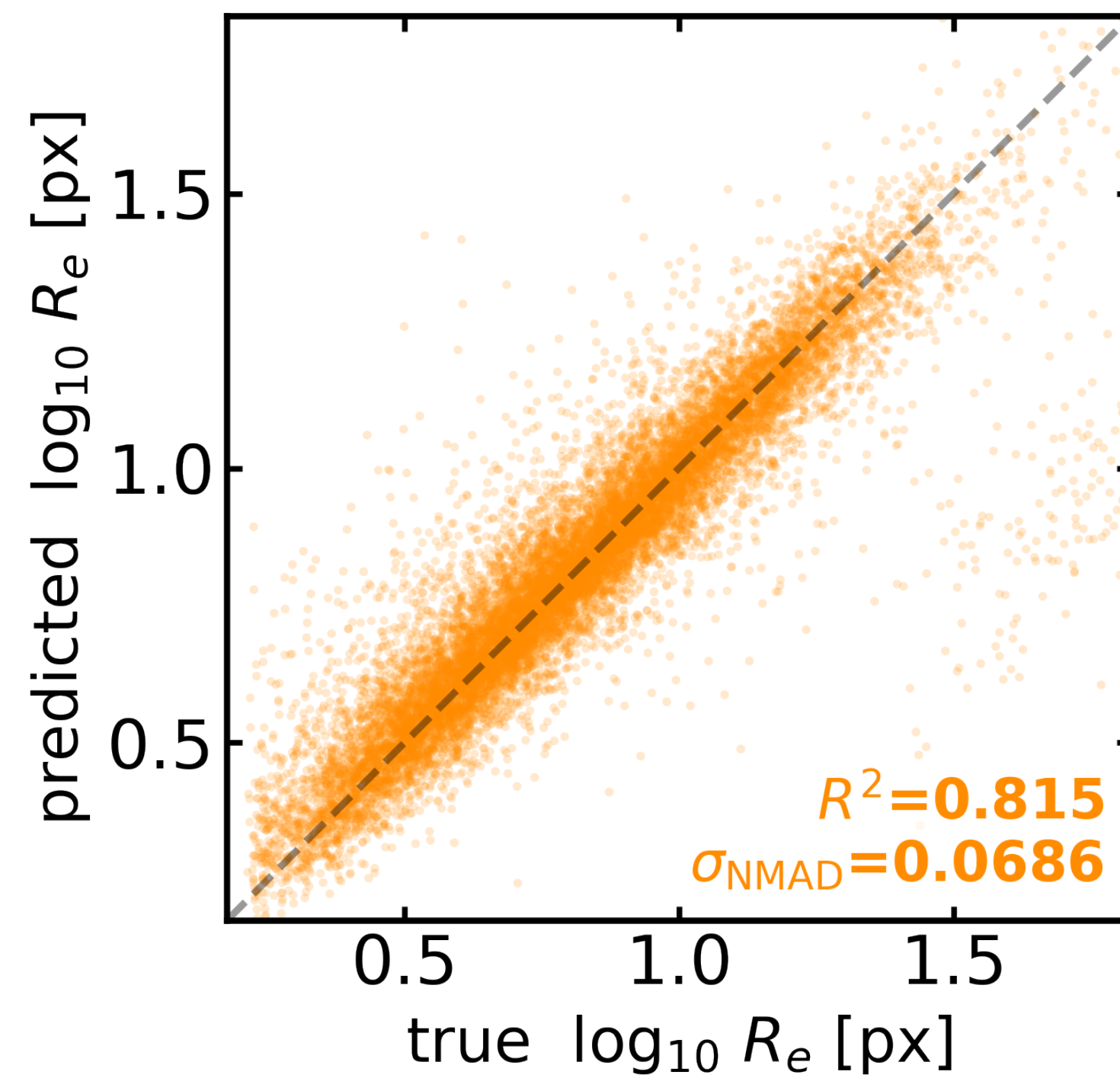
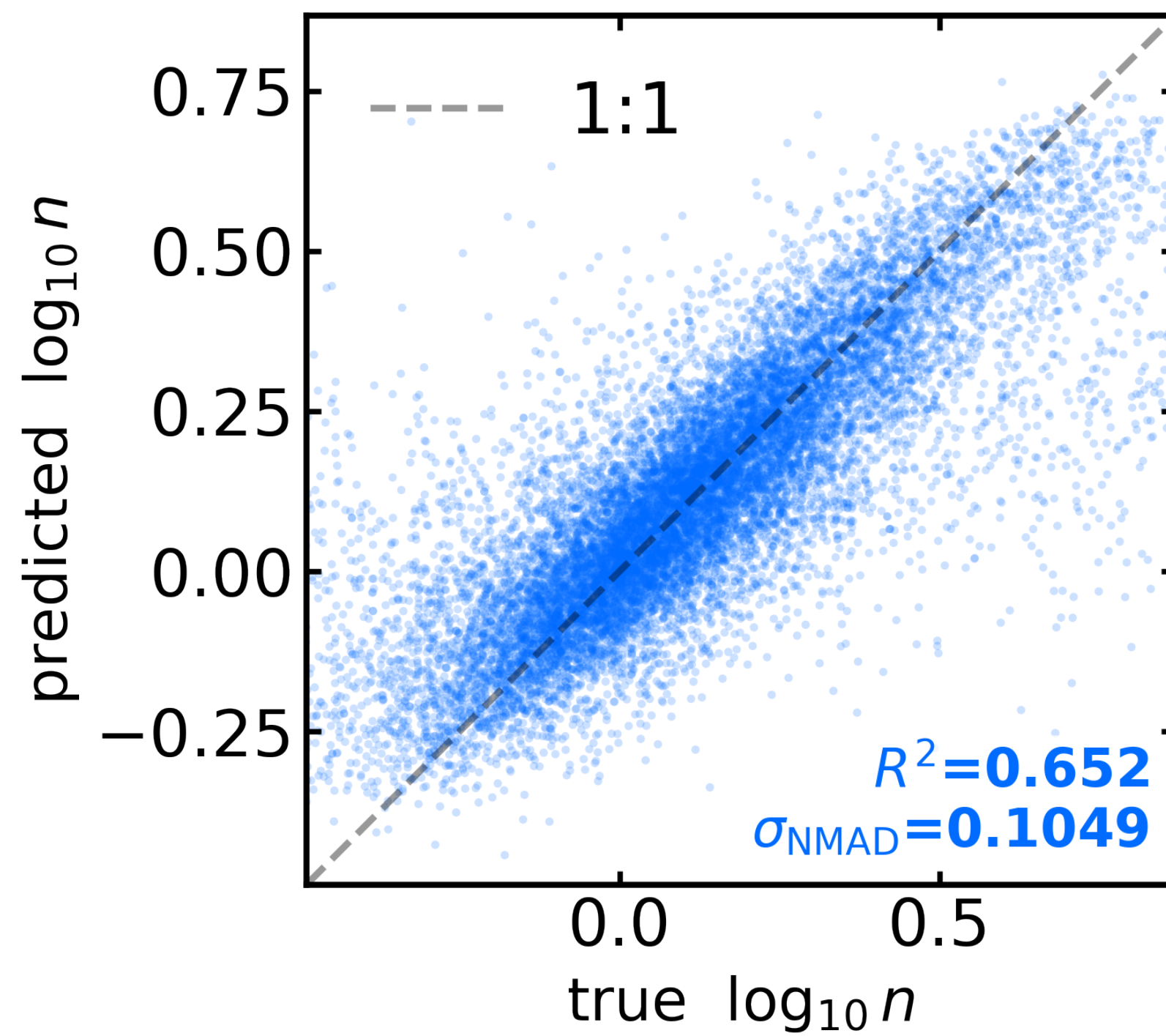
- Imaging: single-channel DINOv2 vision transformer
- Spectra: transformer auto encoder with masked-modeling objective
- CLIP alignment: frozen encoders + attention pooling head



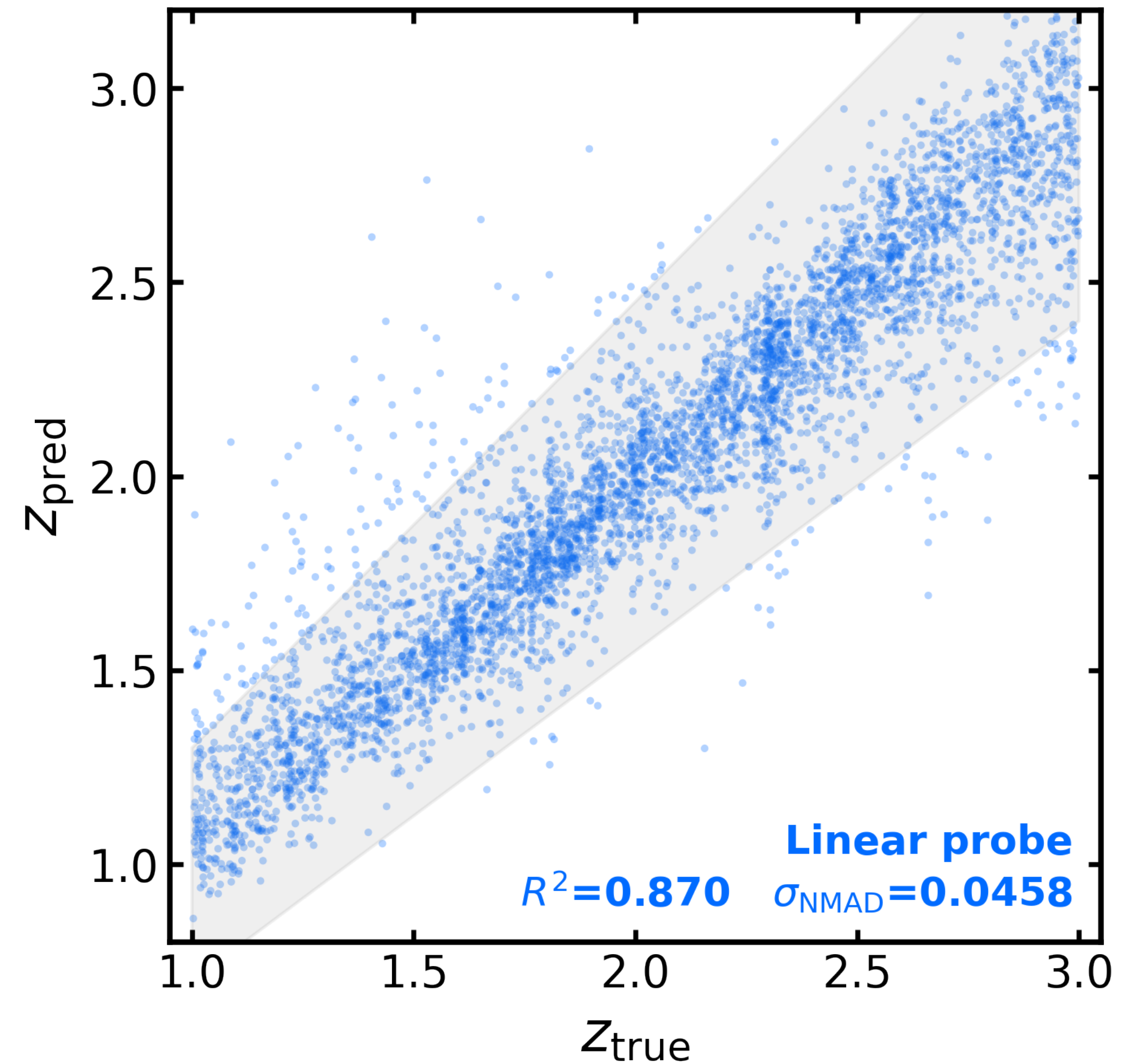
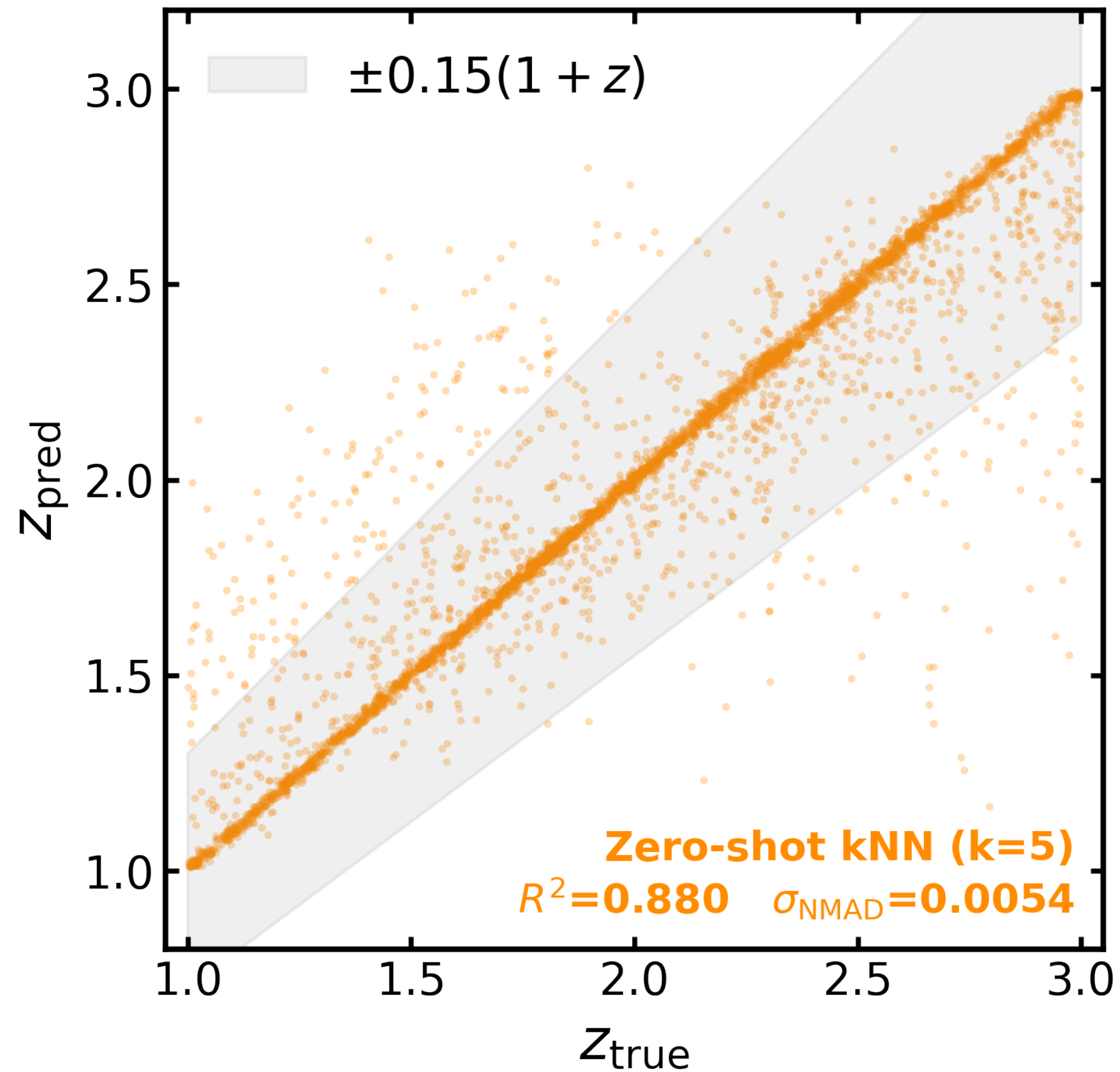
What do the models “see”?



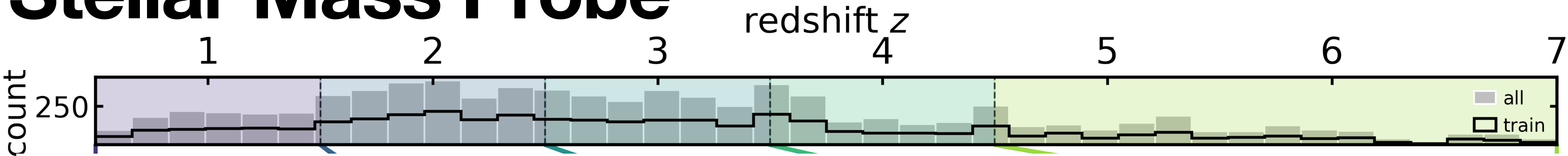
Single Encoder Linear Probe: Morphology



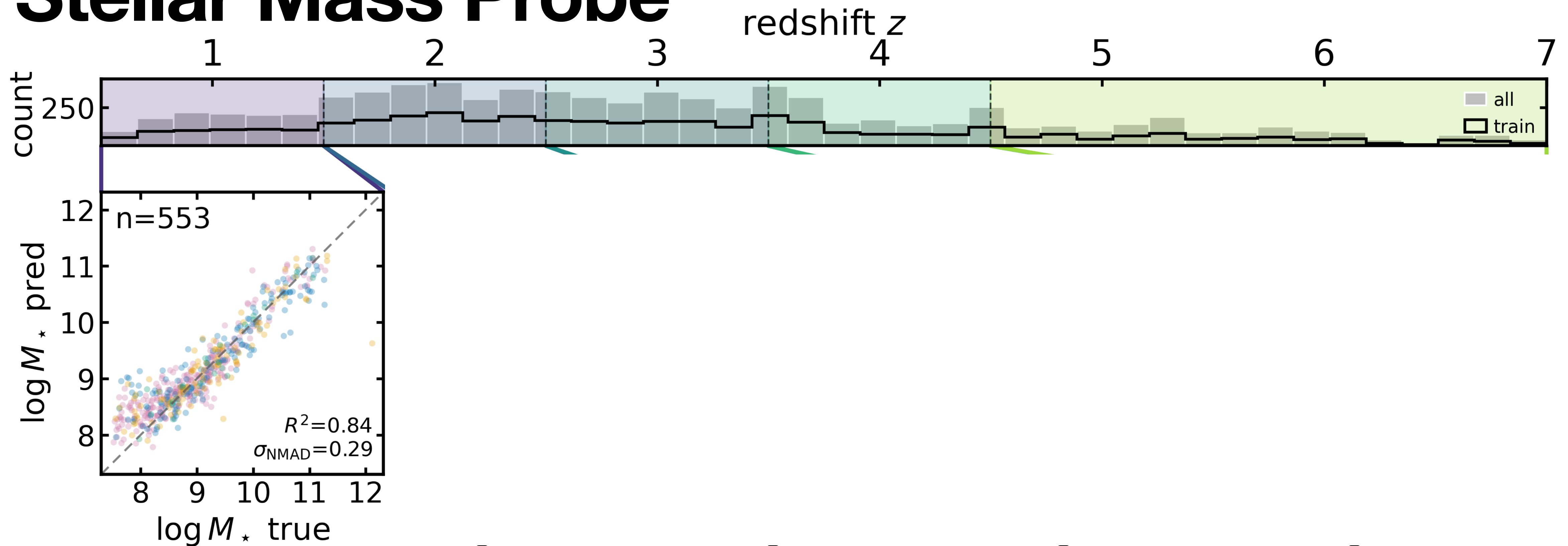
Single Encoder Linear Probe: Redshift



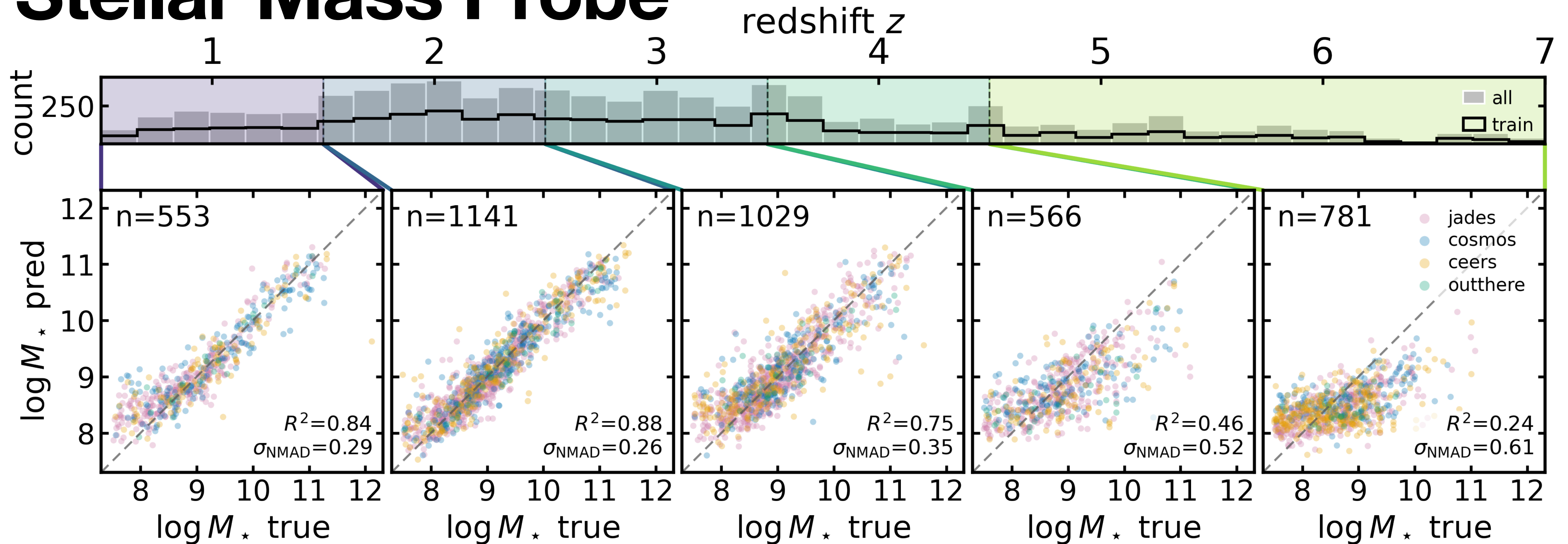
Stellar Mass Probe



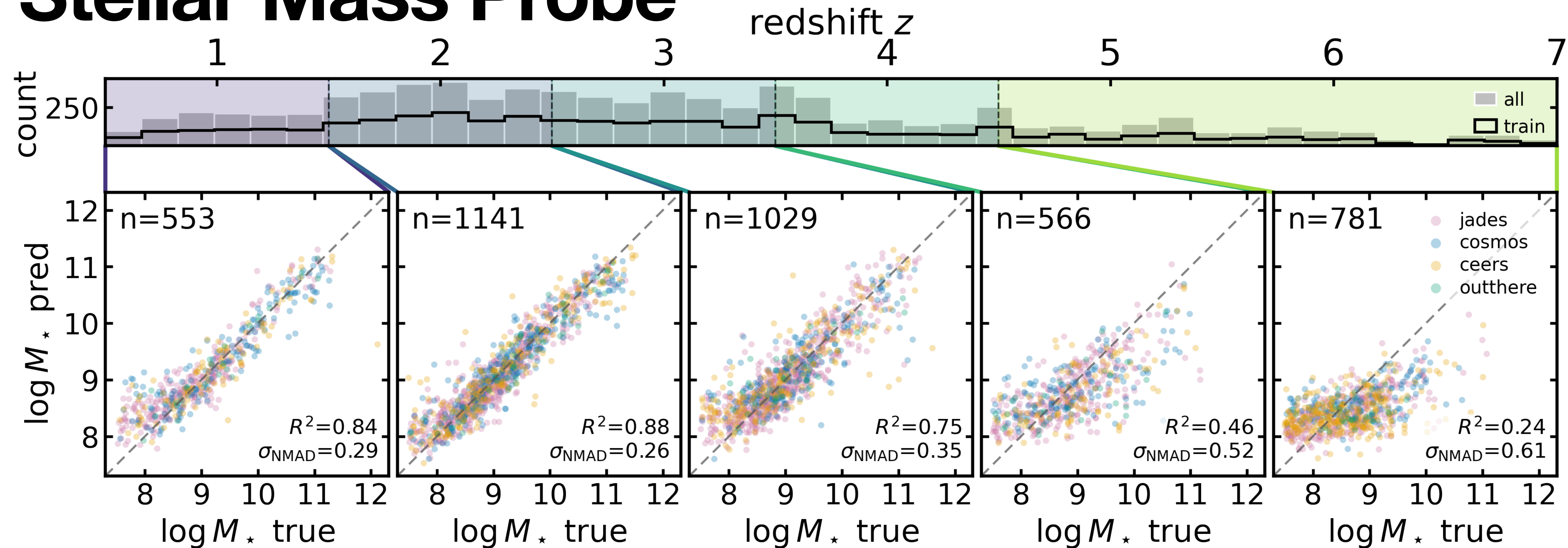
Stellar Mass Probe



Stellar Mass Probe



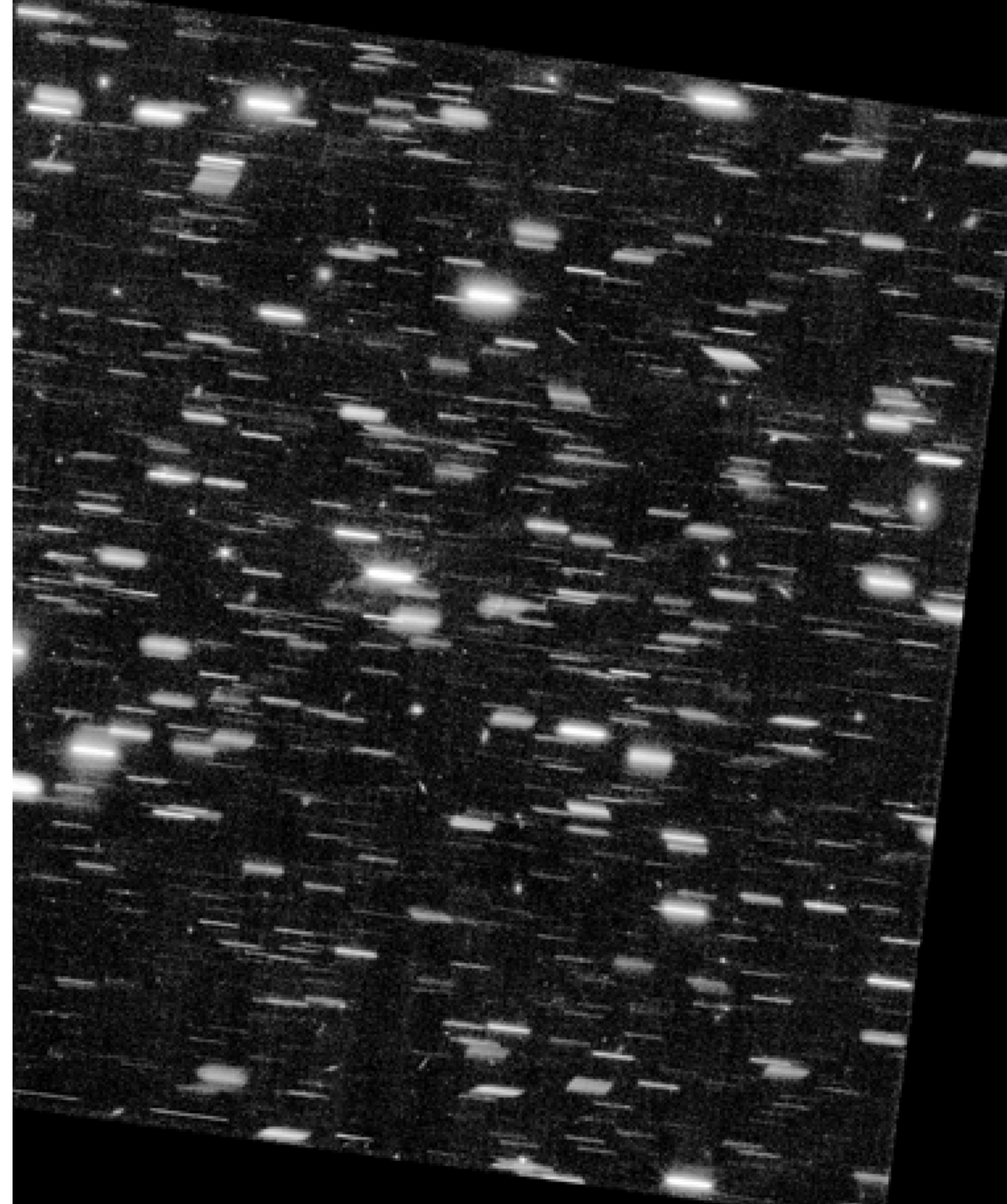
Stellar Mass Probe



Representation	R^2		σ_{NMAD} [dex]		$f_{>0.5}$ [%]	
	un.	al.	un.	al.	un.	al.
F150W photometry	0.750		0.358		19.5	
Image	0.780	0.818 $\uparrow$	0.376	0.342 $\downarrow$	20.1	17.1 $\downarrow$
Spectrum	0.764	0.856 $\uparrow$	0.391	0.303 $\downarrow$	22.4	13.4 $\downarrow$
Image + spectrum	0.896	0.902	0.242	0.244	8.1	8.7

Future Work

- Application to other physical properties
- ... and to actual OutThere survey
- Euclid
- Roman
- How would the addition of other modalities improve the model?
More images? Catalogs?



Summary

- Self-supervised encoders trained separately on sparse JWST data recover mass with $R^2 \approx 0.9$, $\sigma_{\text{NMAD}} = 0.24$ dex at $1 < z \leq 3$
- CLIP alignment narrows the gap between single-modality and joint performance
- Makes physical property estimation feasible for a growing population of sources @ scale without needing (many) labels

