November 27th - December 1st 2023
IAP, Paris / Flatiron institute, New York

Data-driven galaxy morphologies at $z > 3$ with contrastive learning and cosmological simulations

Jesús **Vega-Ferrero** and Marc **Huertas-Company**

https://arxiv.org/abs/2302.07277v2
(Accepted in ApJ)

Collaborators

L. Costantin, P.G. Pérez-González, R. Sarmiento, A. Pillepich

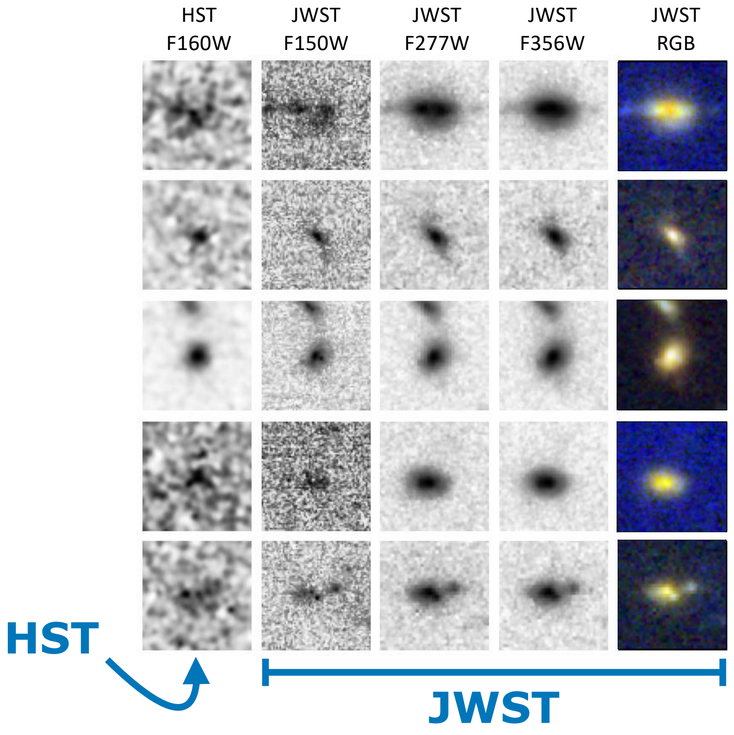
+ JWST/CEERS collaboration + TNG collaboration + external collaborators

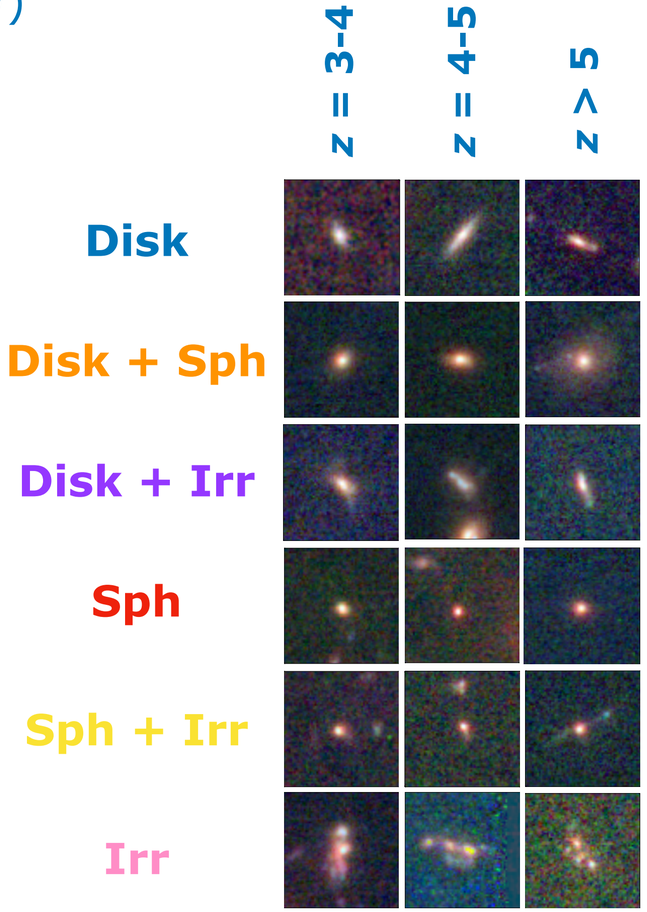
Universidad de Valladolid

Data-driven galaxy morphologies at $z > 3$

JWST: a galaxy morphology machine at high-z

- **High spatial resolution**: 0.03''/px (SW) and 0.06''/px (LW)
- ~0.5 kpc resolution at $z \sim 3$
- **Optical rest-frame at $z > 3$**





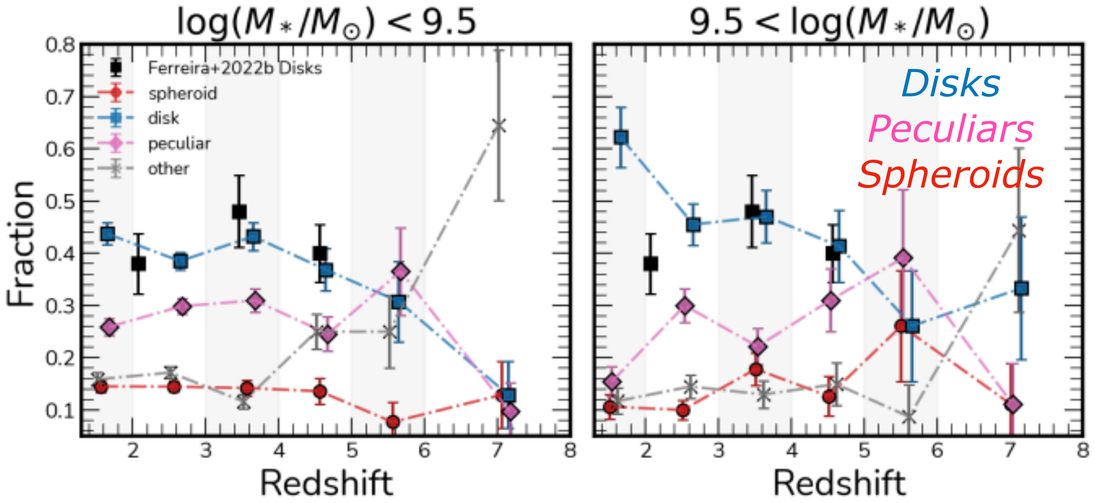
Credits: *Kartaltepe et al. 2023*

JWST: a galaxy morphology machine at high-z

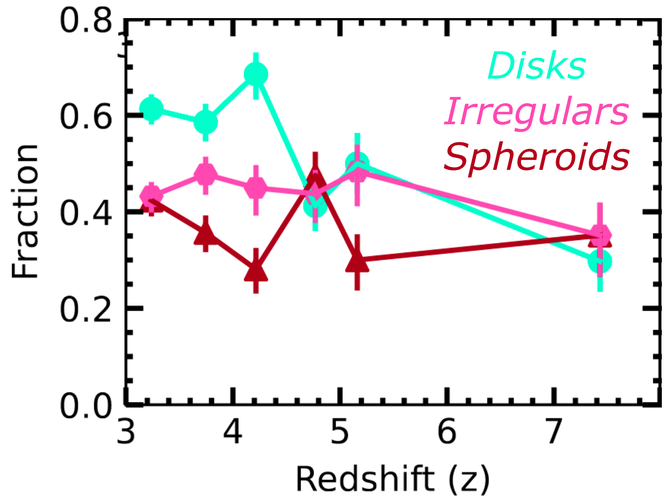
- First JWST observations: **disk galaxies in place very early on**

  (*Kartaltepe et al. 2023, Ferreira et al. 2023, Robertson et al. 2023, Huertas-Company et al. 2023b*)

Visual classifications



Ferreira et al. 2023



Kartaltepe et al. 2023

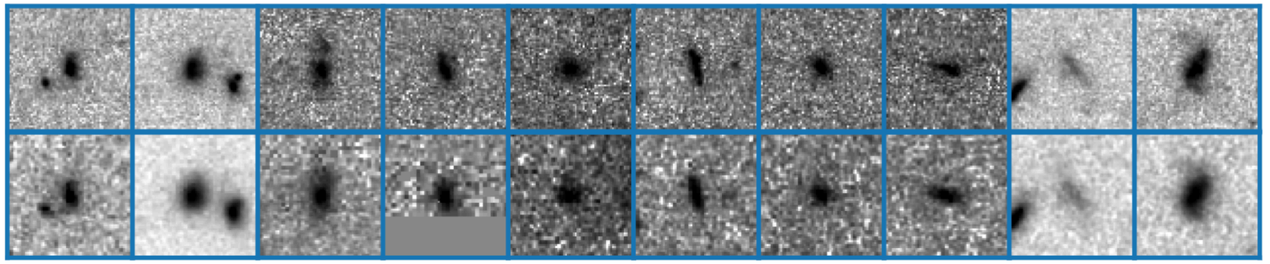
Morphological classification of high-z JWST galaxies

- **Visual classifications are less reliable at high-z**

  PSF, cosmological dimming, noise, contaminants,...

- **'No secure' labelling** of high-z galaxies, even at depths and resolution of JWST

Disks at $z > 3$ (*Kartaltepe et al. 2023*)



Can we represent galaxy images in a space that is robust to these perturbations?

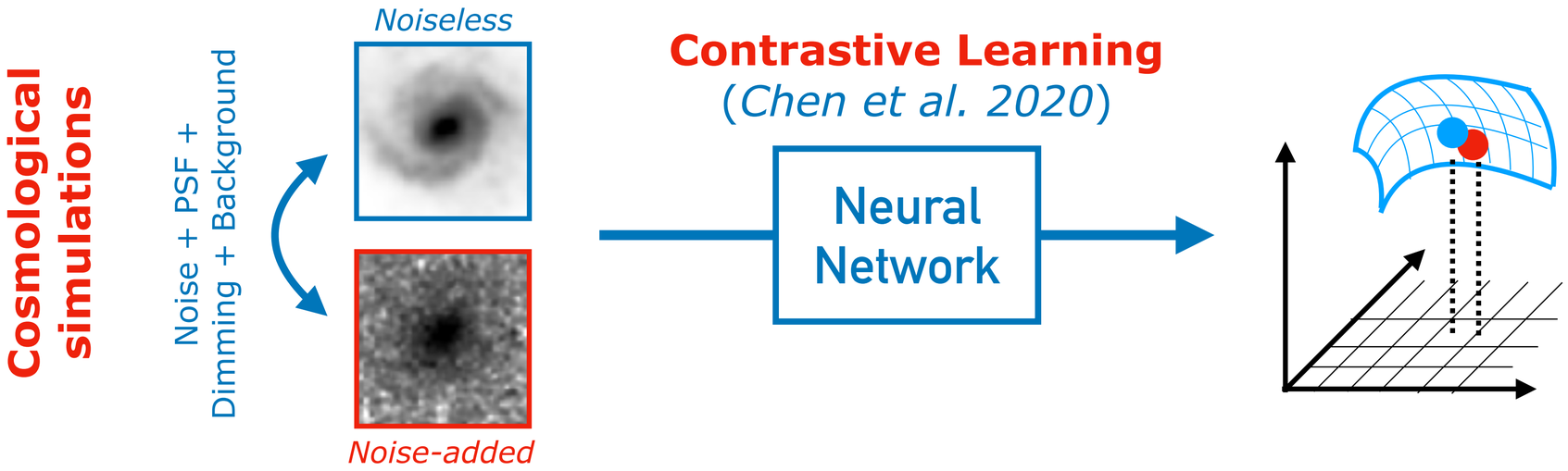
Can we say more about the nature of these objects?

Morphological classification of high-z JWST galaxies

- **Visual classifications are less reliable at high-z**

  PSF, cosmological dimming, noise, contaminants,...

- **'No secure' labelling** of high-z galaxies, even at depths and resolution of JWST



Can we represent galaxy images in a space that is robust to these perturbations?

Can we say more about the nature of these objects?

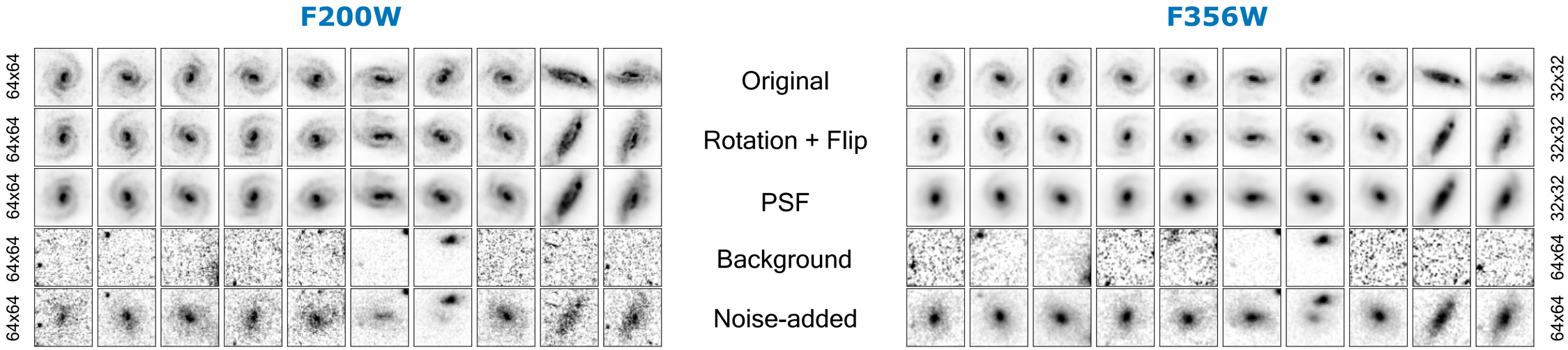
Data-driven galaxy morphologies at $z > 3$

Mock JWST images of TNG50 galaxies (*Costantin et al. 2022*)

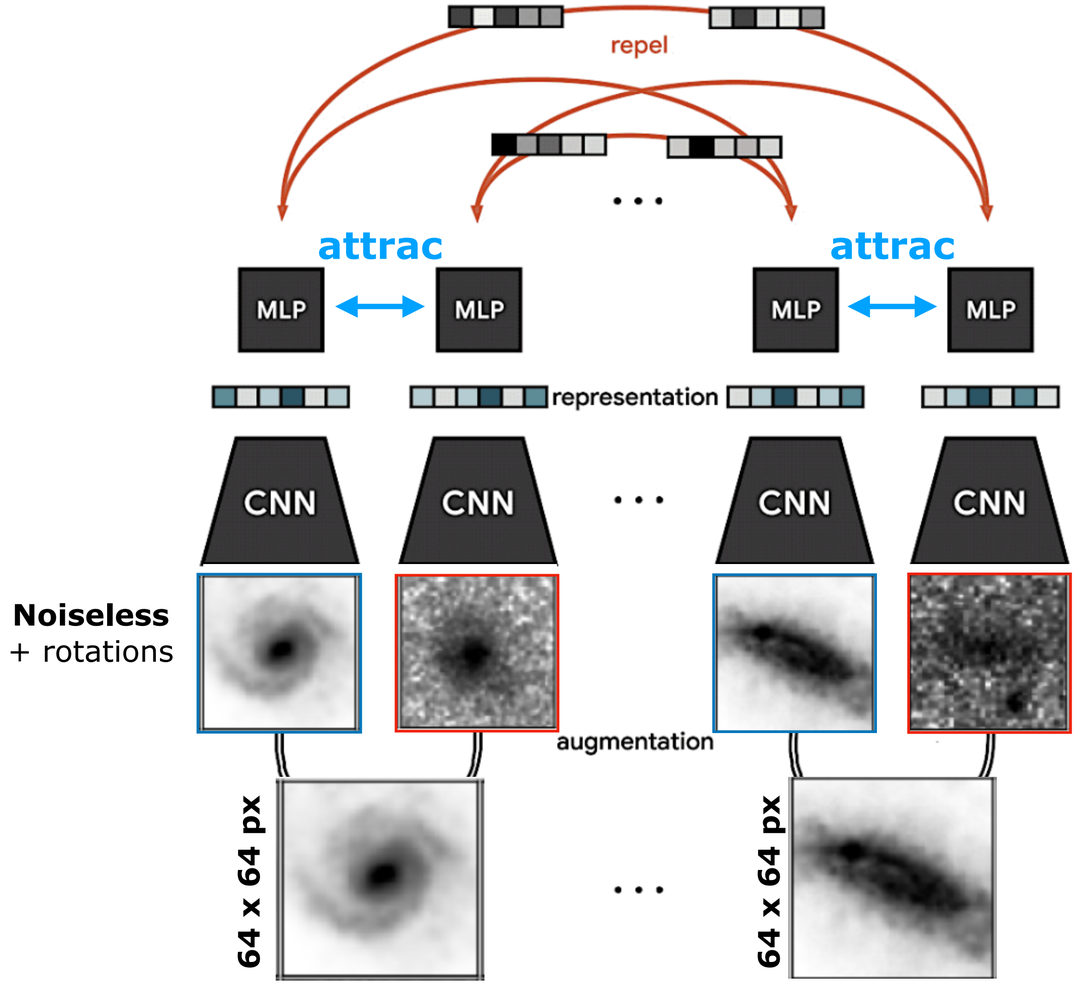
https://www.tng-project.org/data/docs/specifications/#sec5v

- 1,238 TNG50 galaxies with $\mathbf{M_\star > 10^9\ M_\odot}$
- Redshift range: $\mathbf{3 < z < 6}$
- 20 l.o.s. projections per galaxy

- **24,760 galaxy images**
- **F200W** (SW) and **F356W** (LW) filters
- **Image augmentations** (rotations, source noise, etc.)



Contrastive learning for morphological classification



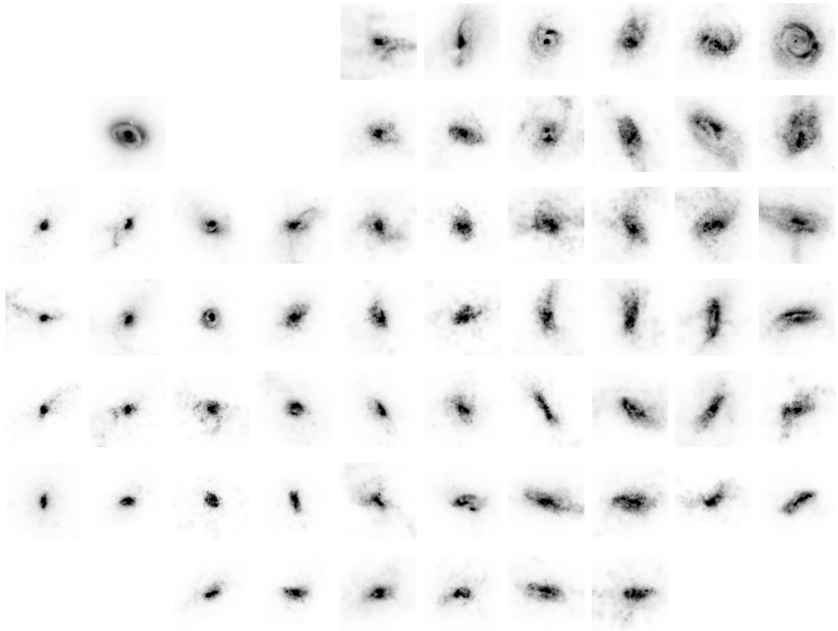
Vector encoding galaxy images (2048 features)
(invariant to perturbations)

Noise + PSF + Dimming
+ rotations + background

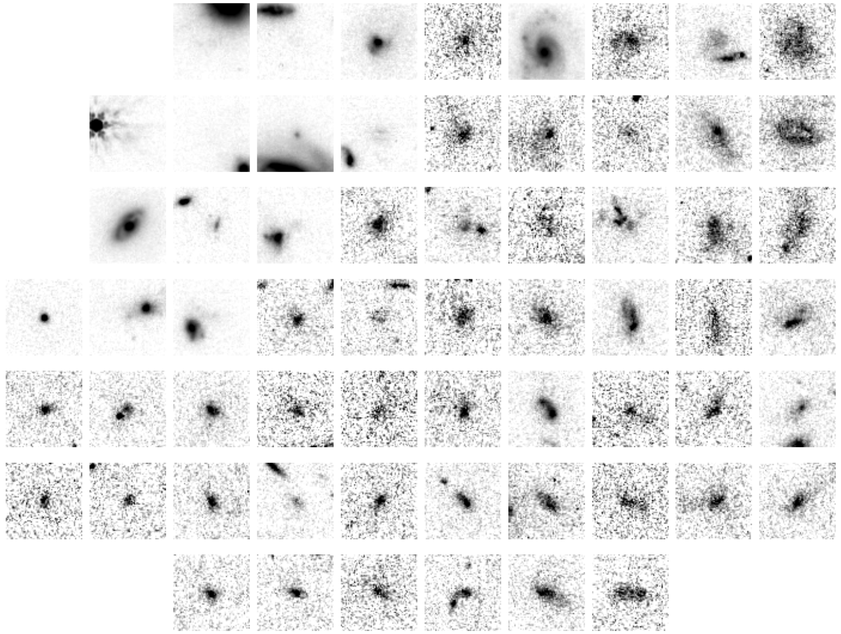
Contrastive learning for morphological classification

UMAP 2D projection

Noiseless (TNG50)
F200W



Noise-added (TNG50)
F200W

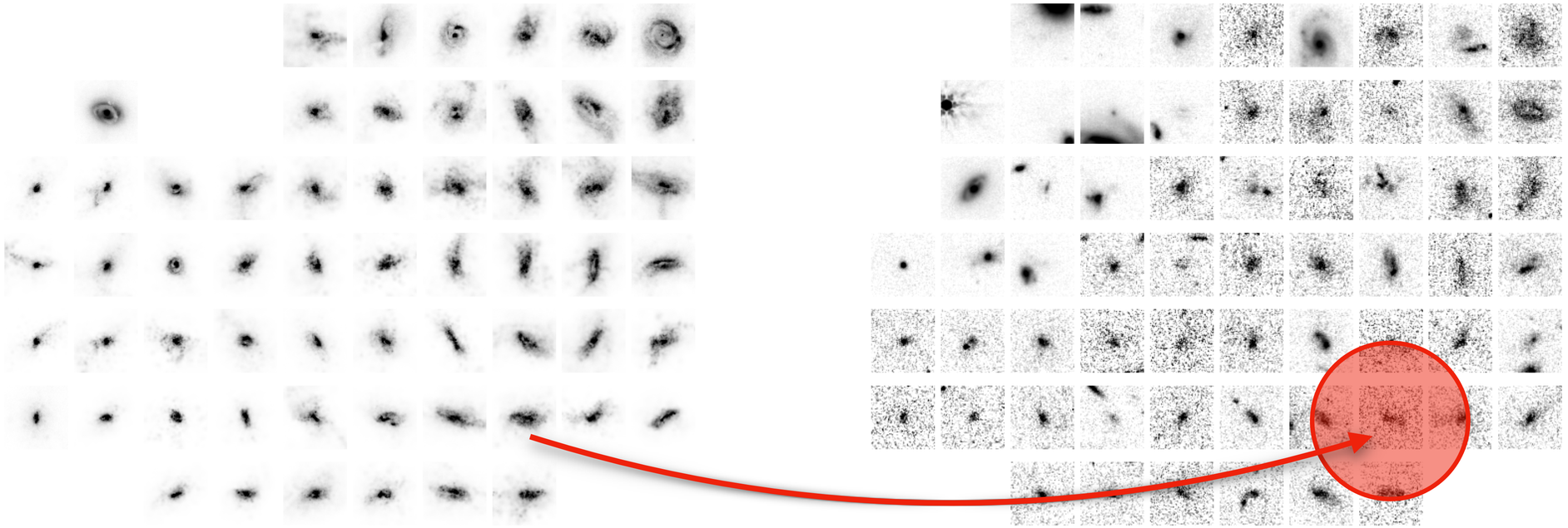


Contrastive learning for morphological classification

A morphological representation robust to noise

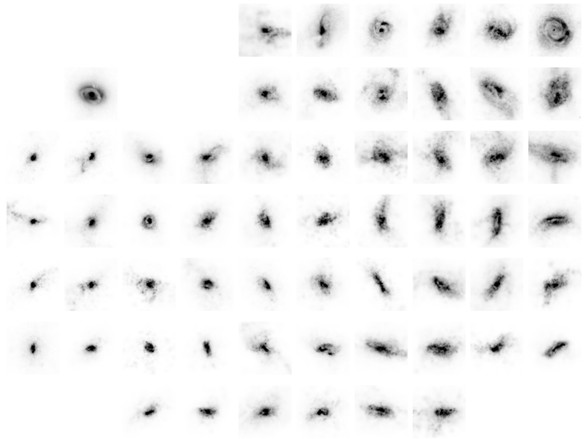
Noiseless (TNG50)
F200W

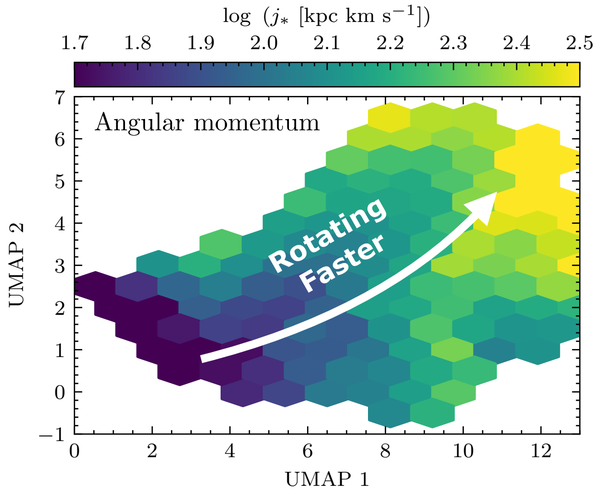
Noise-added (TNG50)
F200W

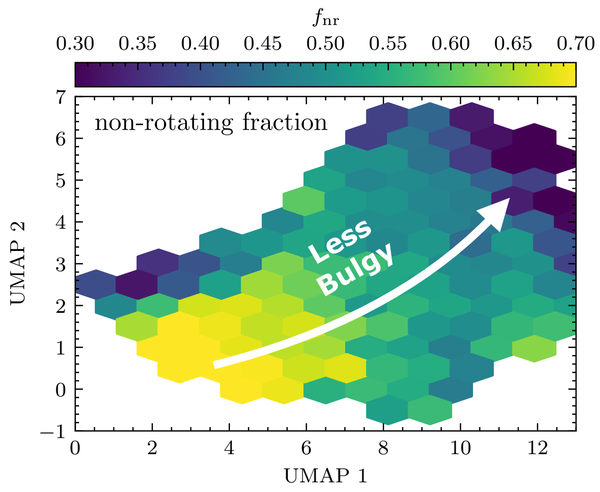


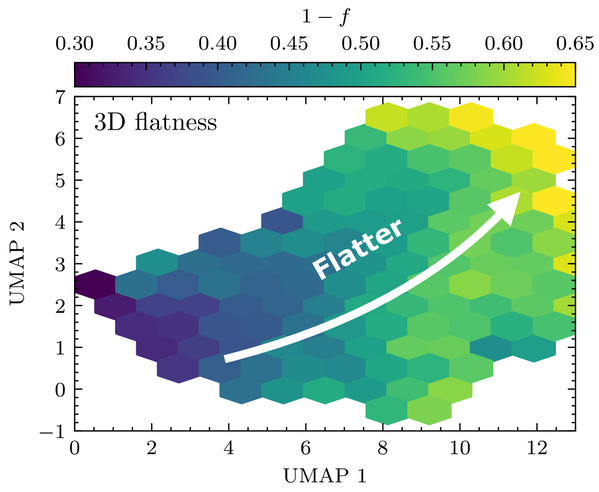
Contrastive learning for morphological classification

Correlation with physical properties



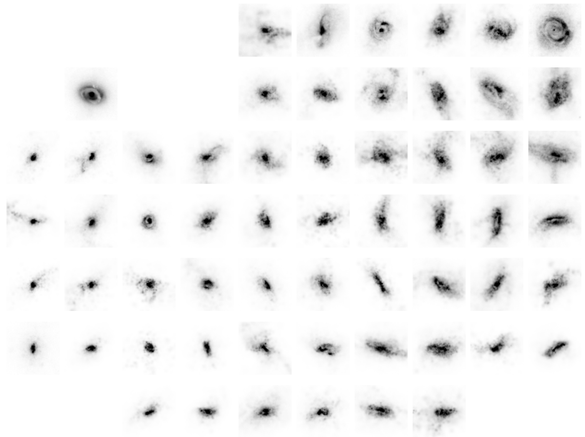


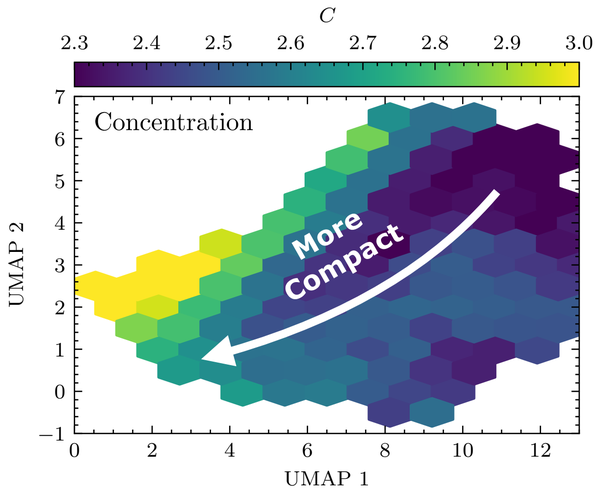


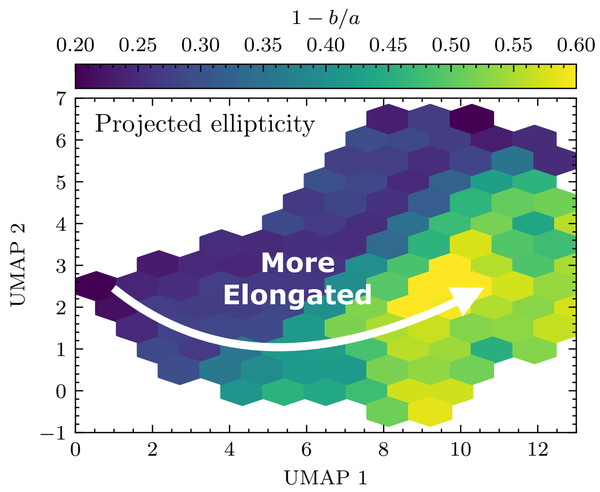


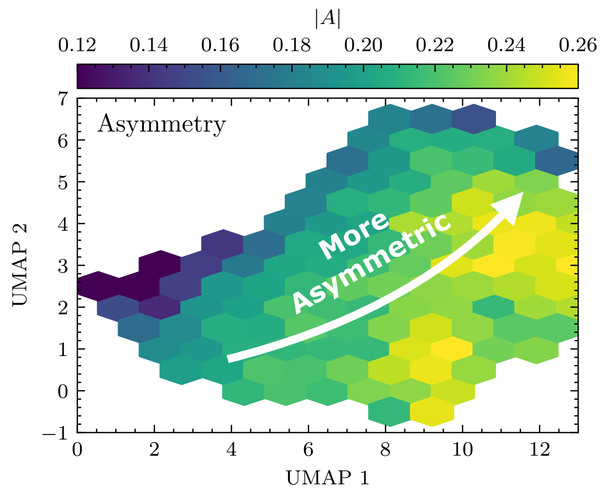
Contrastive learning for morphological classification

Correlation with standard morphological parameters



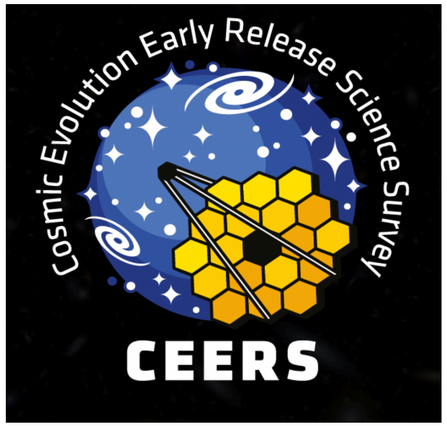


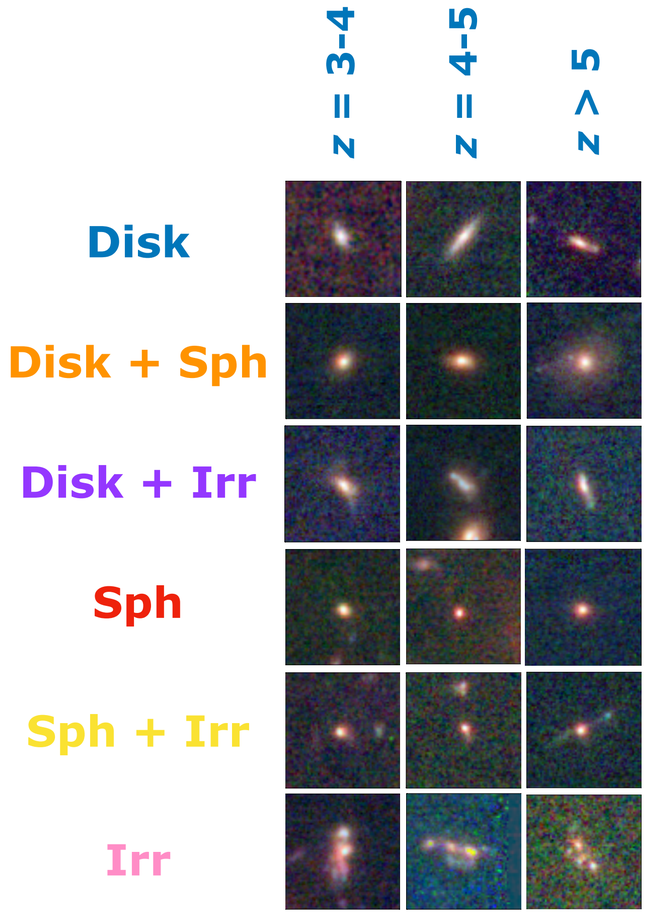




Contrastive learning representations of observed JWST galaxies

- **CEERS** (*Finkelstein et al. 17,22,23; Bagley et al. 22*)
- NIRCam imaging of 100 arcmin^2 with **depth 28.8-29.7 mag**
- Mass-complete **($M_\star > 10^9\ M_\odot$)**
- Redshift-selected sample **($3 < z < 6$)**

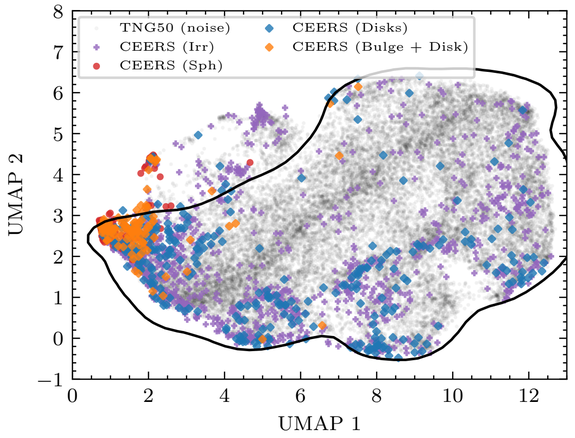


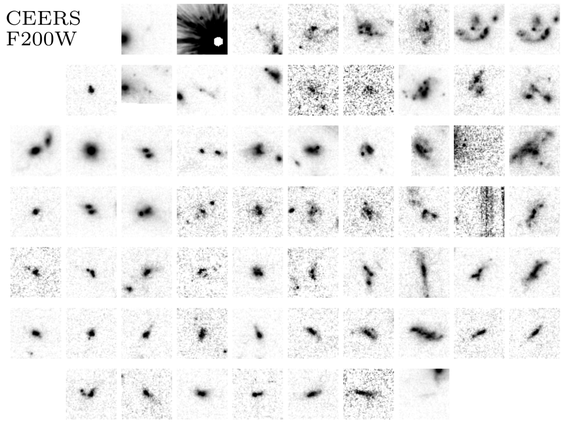


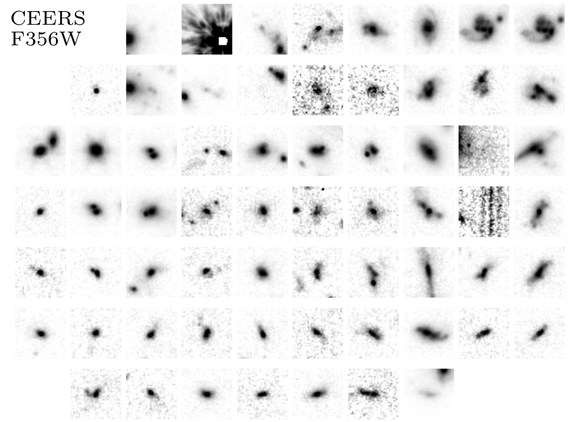
Credits: *Kartaltepe et al. 2023*

Contrastive learning representations of observed CEERS galaxies

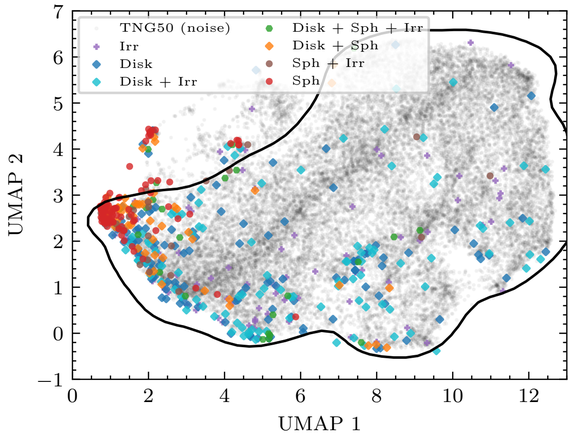
MHC+2023b
CNN-based
Domain adaptation
from CANDELS labels

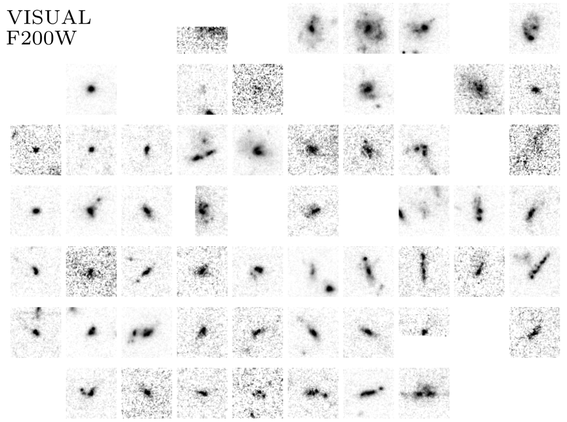


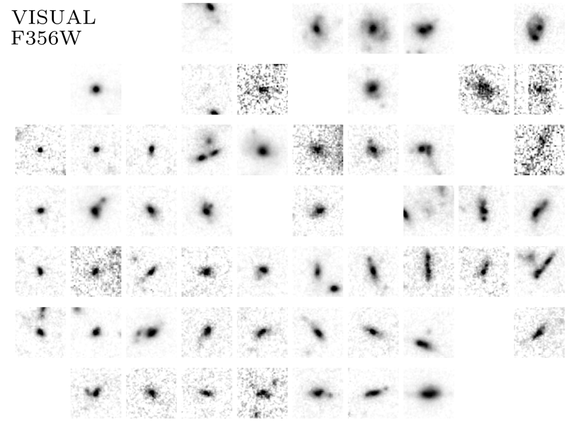




Kartaltepe+2023
Visual classifications





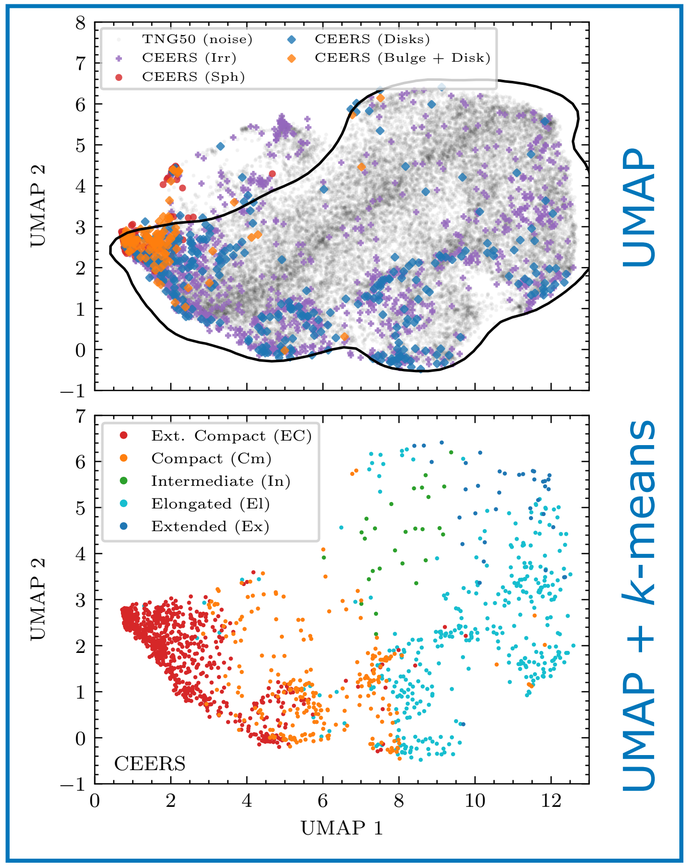


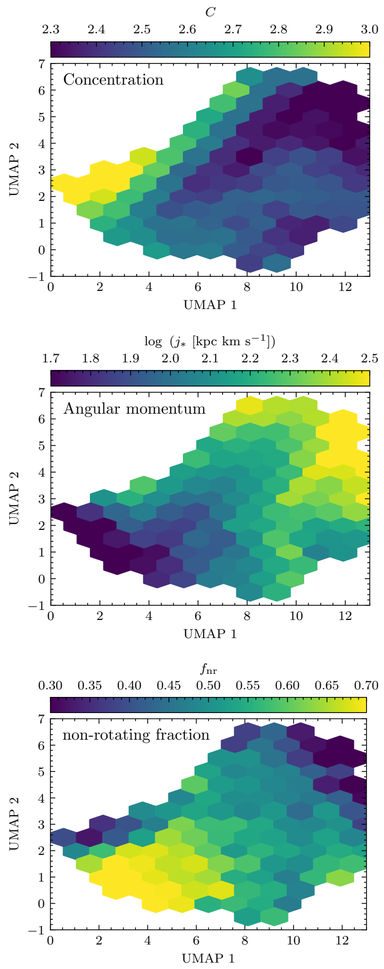
Not a good overlap between simulations and observations

Data-driven galaxy morphologies at $z > 3$

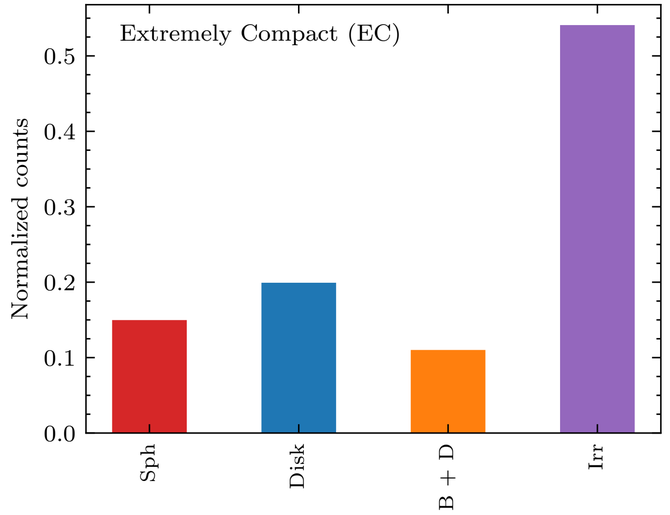
Can we say more about the nature of the observed disk galaxies?

CEERS
(*MHC+2023b*)





Extremely Compact (EC)

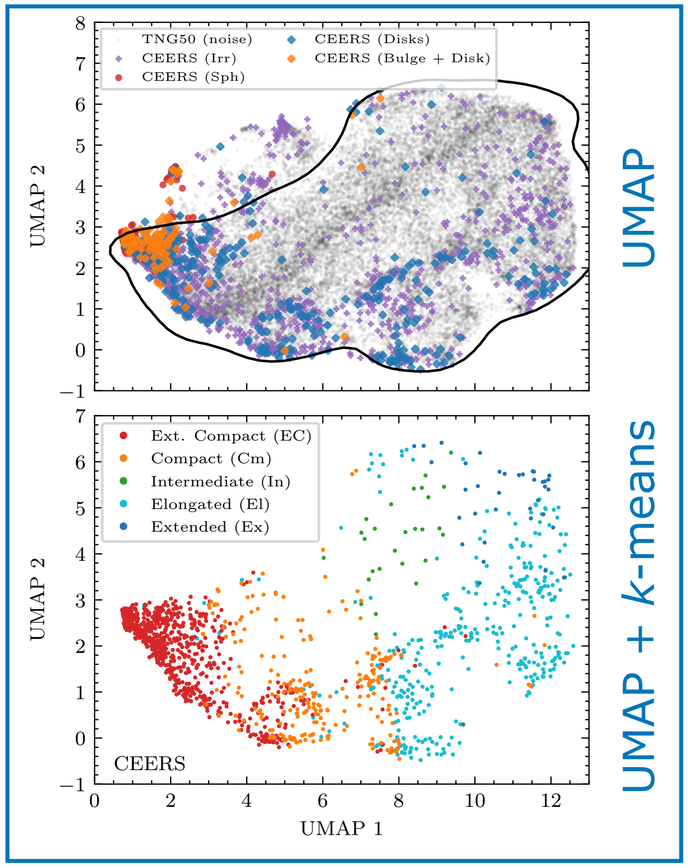


~50% of Irr and Disk are EC, slow rotators, bulge dominated

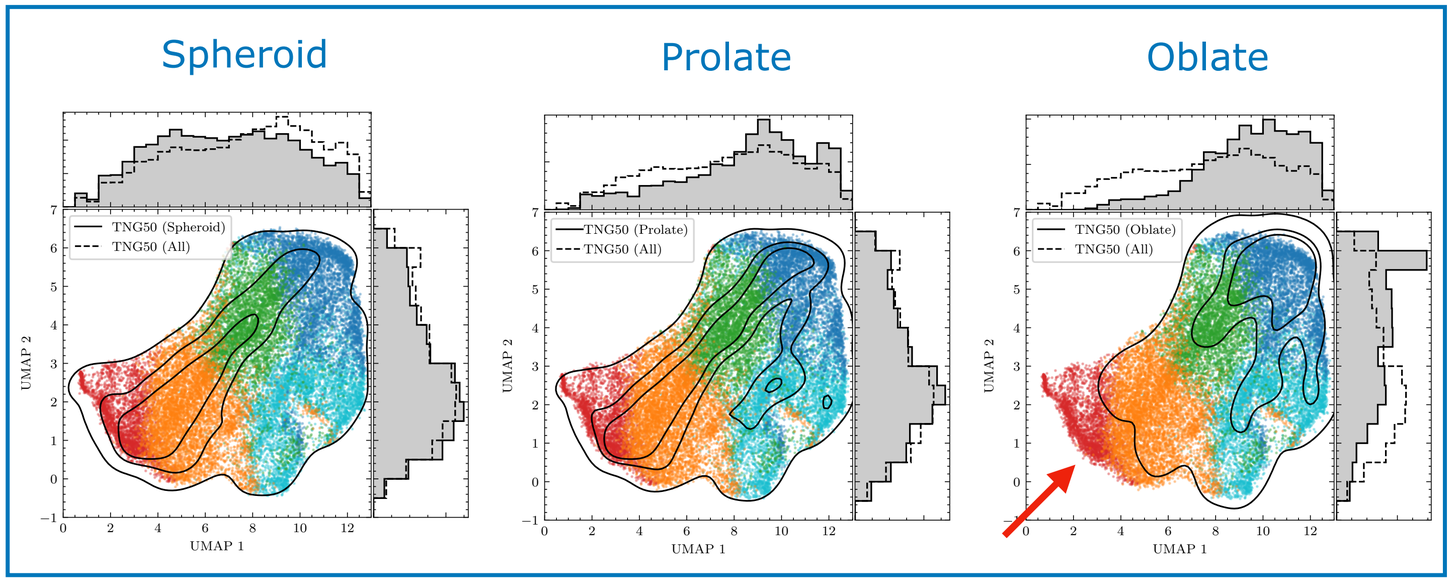
Data-driven galaxy morphologies at $z > 3$

Can we say more about the nature of the observed disk galaxies?

CEERS
(*MHC+2023b*)



TNG50



~50% of Disks not likely to be pure Disks, consistent with prolate stellar structure

Pandya et al. 2023
50%-80% prolate dwarfs at z = 3-8

Conclussions

- **Morphologies of JWST galaxies at $z > 3$ with contrastive learning** (data-driven)
- Method **calibrated on mock JWST galaxy images of TNG50 galaxies**
- Representations robust to noise, color, orientation, S/N
- Representations correlate with physical, photometric and structural properties
- Morphological distributions of **CEERS and TNG50 galaxies are different** (observed galaxies are more compact and elongated than simulated ones)
- **CNN-based and visually classified Disks** similar to **TNG50 compact galaxies** with **low angular momentum** and **non-oblate structure**
- **Disk fractions at z > 3 uncertain and possibly overestimated**
- **Deeper imaging** and/or **spectroscopic follow-ups** (also model calibrations with **other simulations**) needed to determine the true nature of these galaxies

https://arxiv.org/abs/2302.07277v2