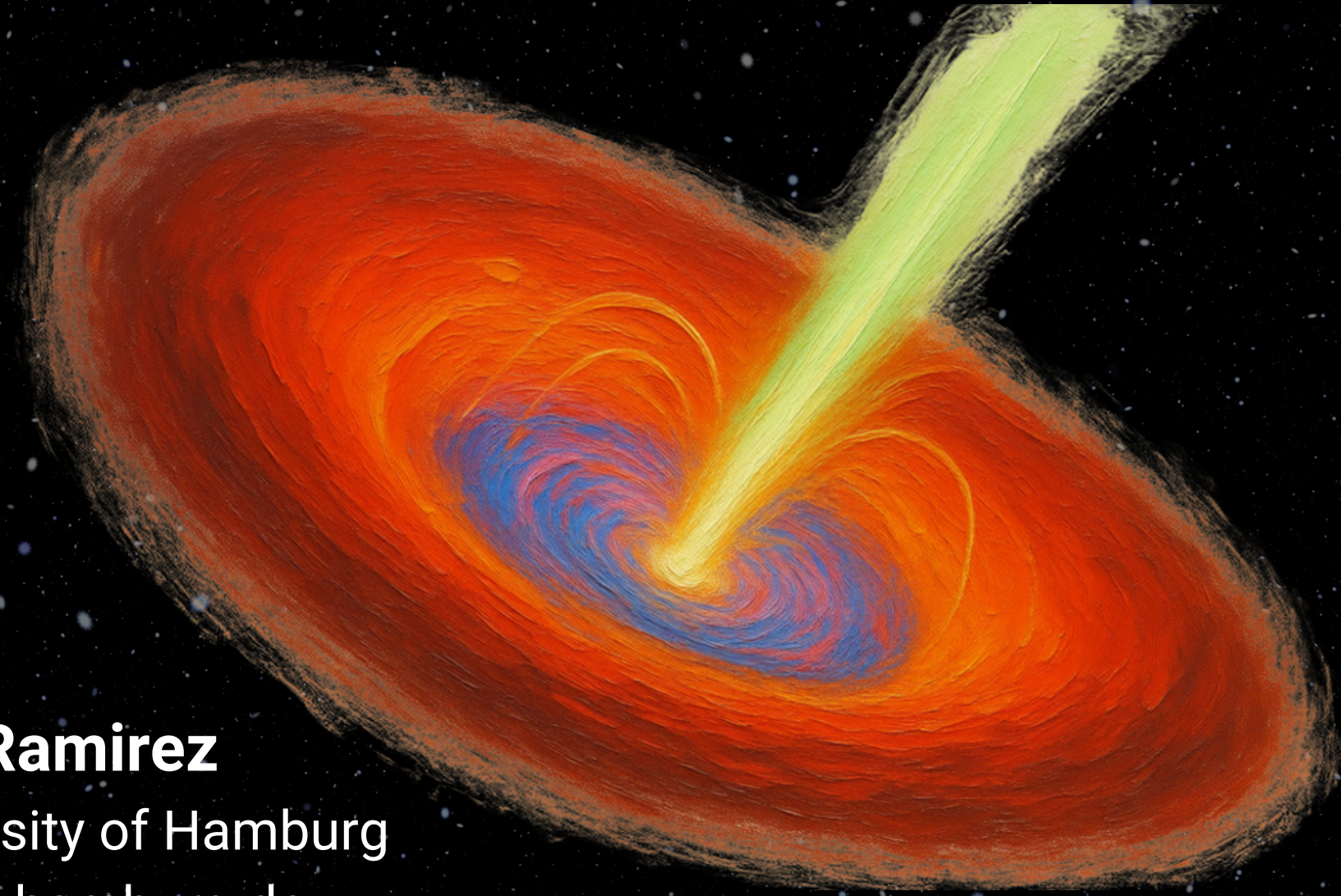


NEEDLES IN THE COSMIC HAYSTACK:

Innovating the search for Reionization-Era quasars with self-supervised ML



Laura Natalia Martinez Ramirez

Hamburg Observatory, University of Hamburg

*Laura.martinez.ramirez@uni-hamburg.de

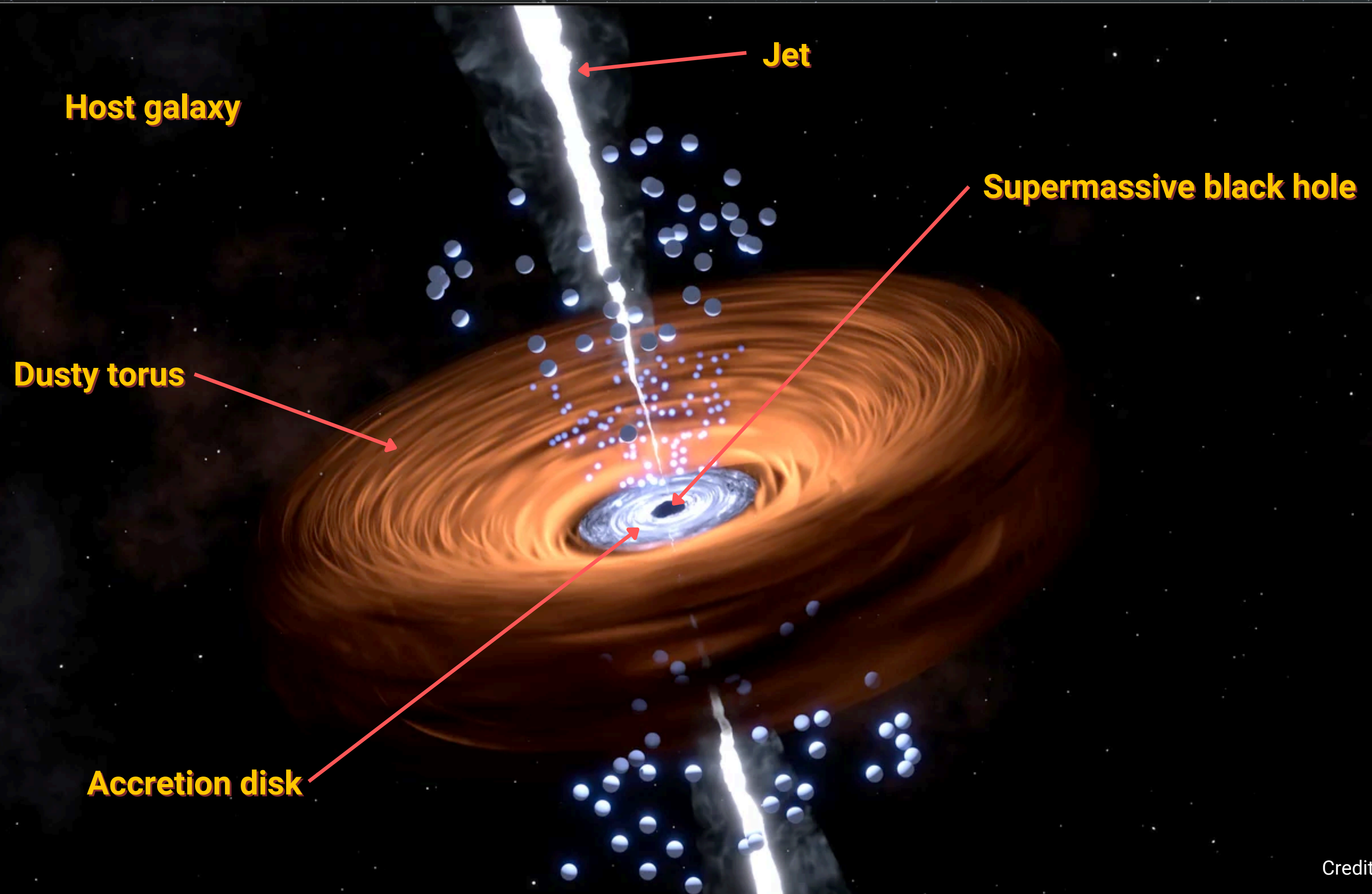
Team: Jan-Torge Schindler, Francesco Guarneri, Tatsuyuki Sekine



**MACHINE LEARNING
FOR ASTROPHYSICS**
3RD EDITION VALLETTA, 31 AUG - 4 SEP, 2026

The most luminous sources of the Universe

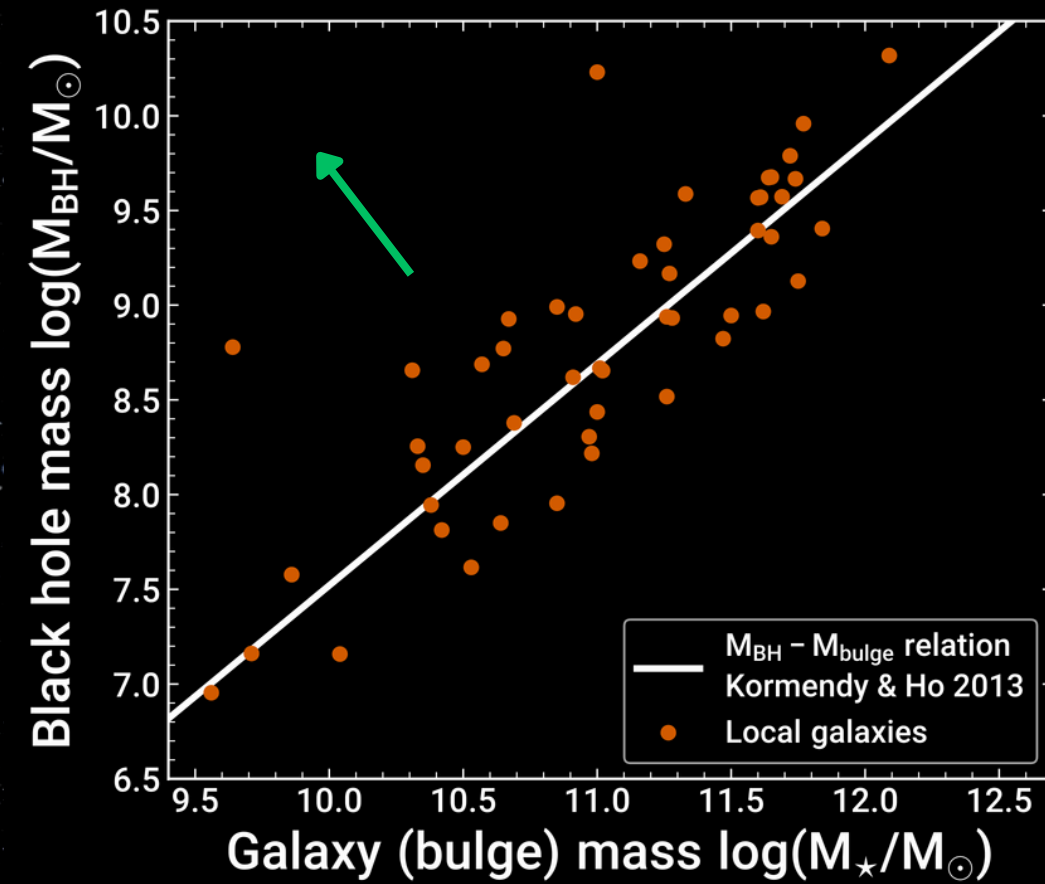
2



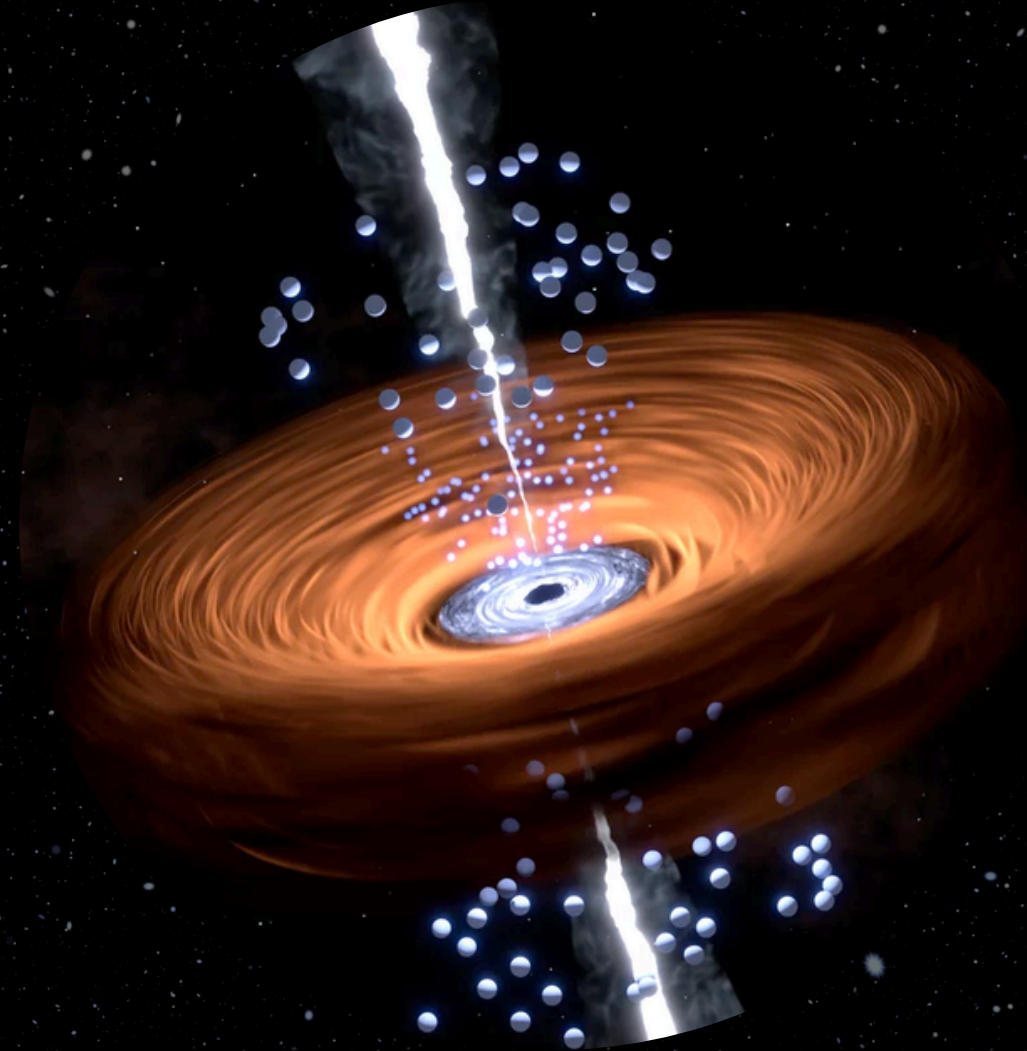
Why study high-z quasars?

3

- SMBHs and host galaxies coevolution

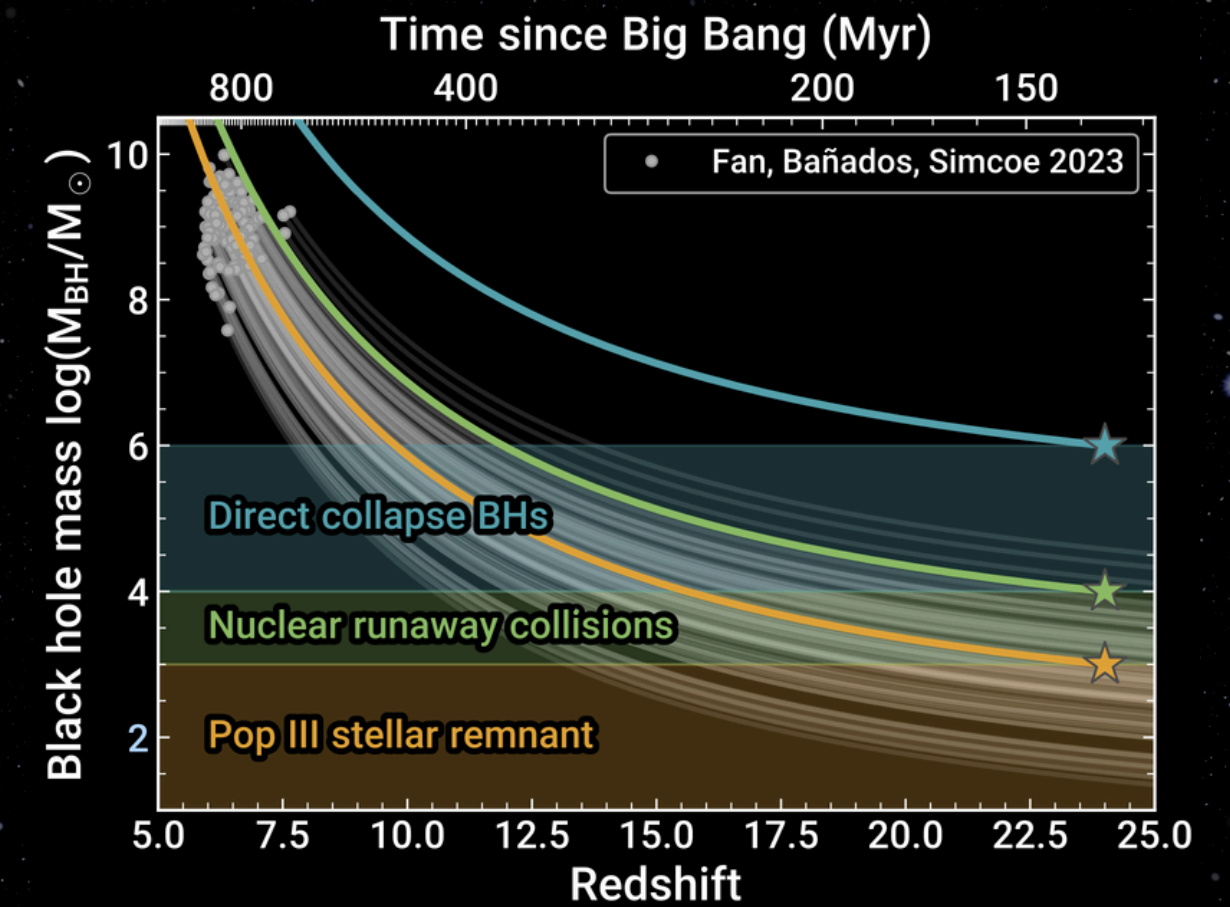


Credit: Schindler



Credit: T. Müller / MPIA

- Early black hole seeds and growth

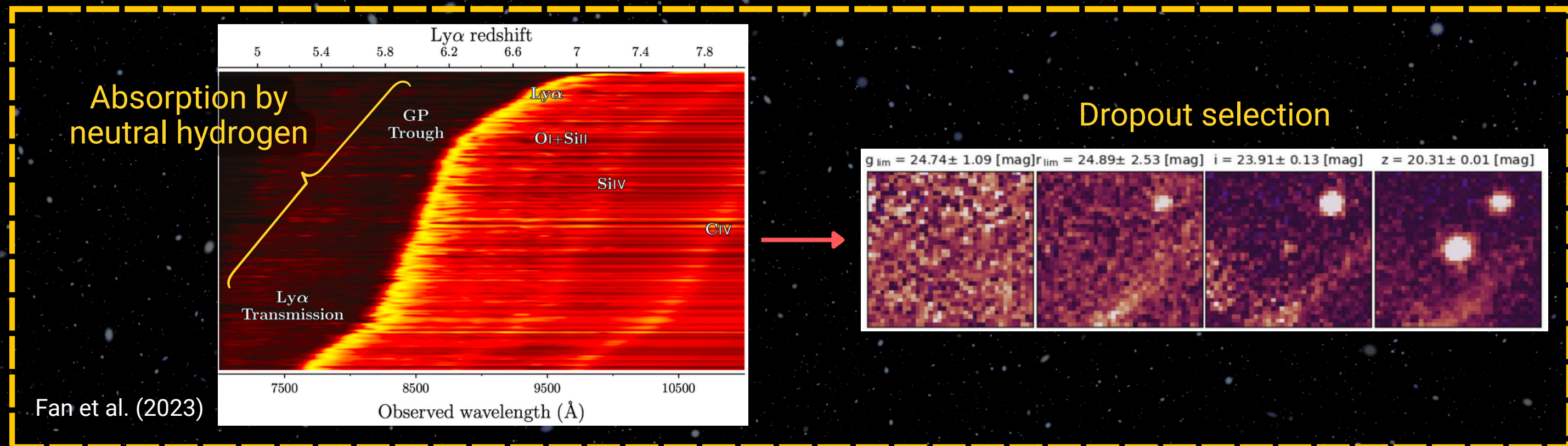
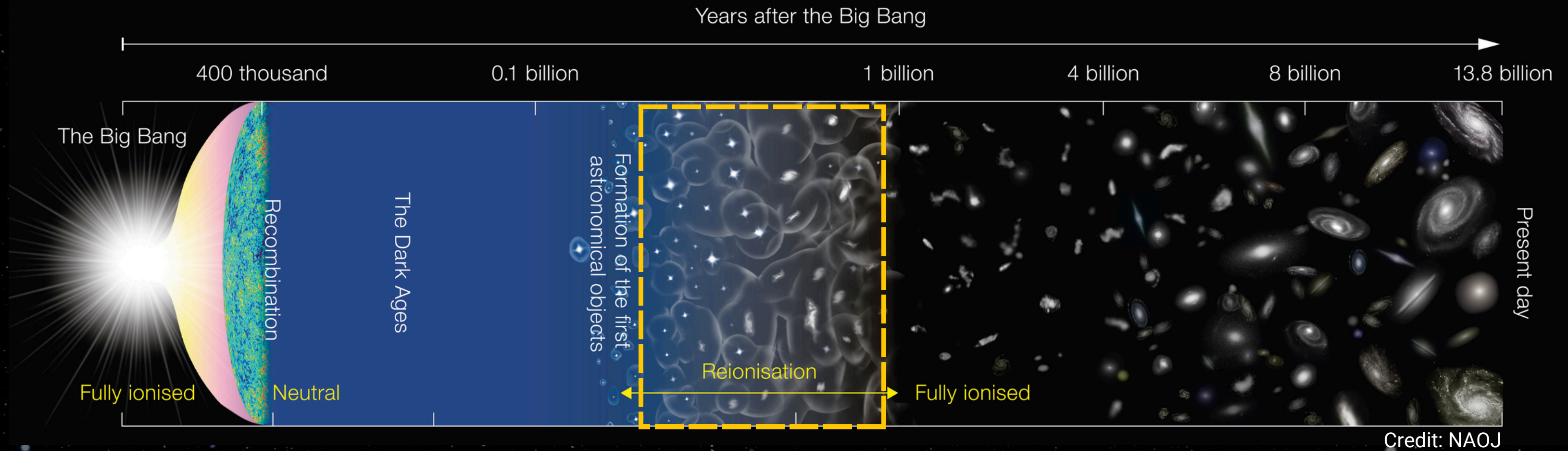


Credit: Schindler

- Ionization and chemical enrichment of IGM
- Probe of cosmological models

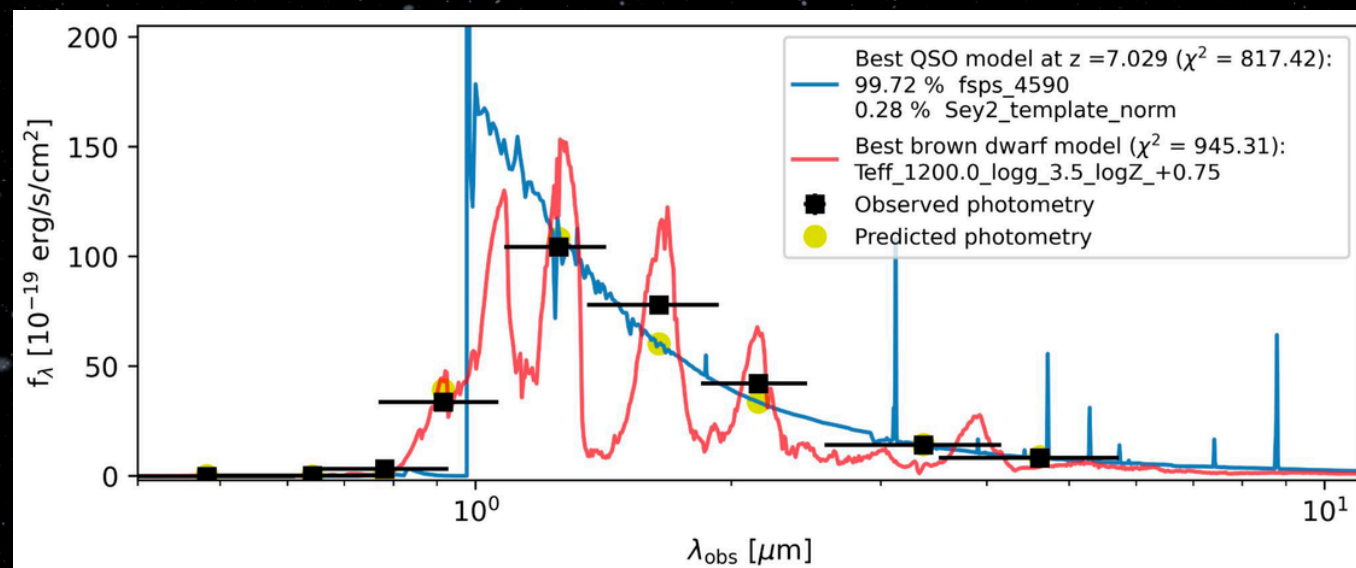
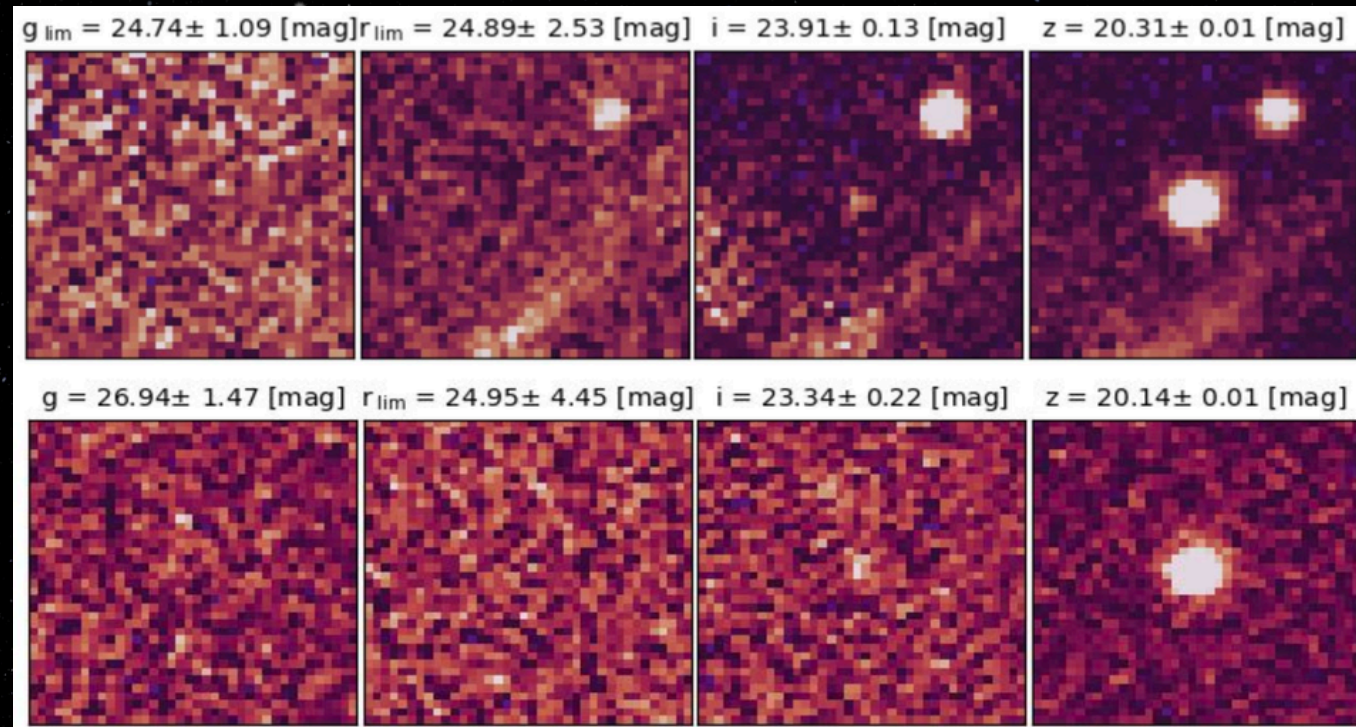
The reionization era

4



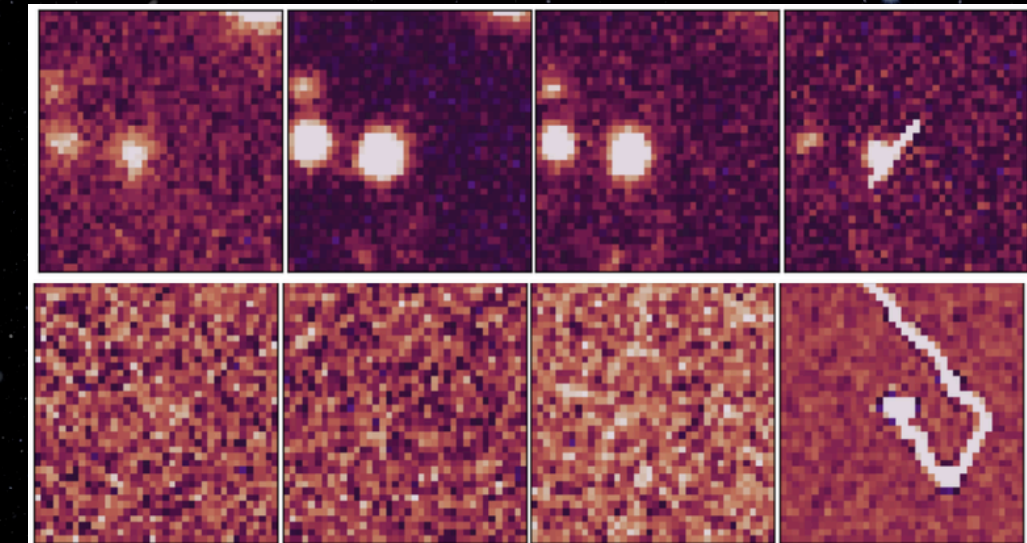


- < 1 QSOs per Gpc^3 at $z > 6$
- Overwhelming contamination by brown dwarfs



Martinez-Ramirez et al. 2026

and even not real sources...

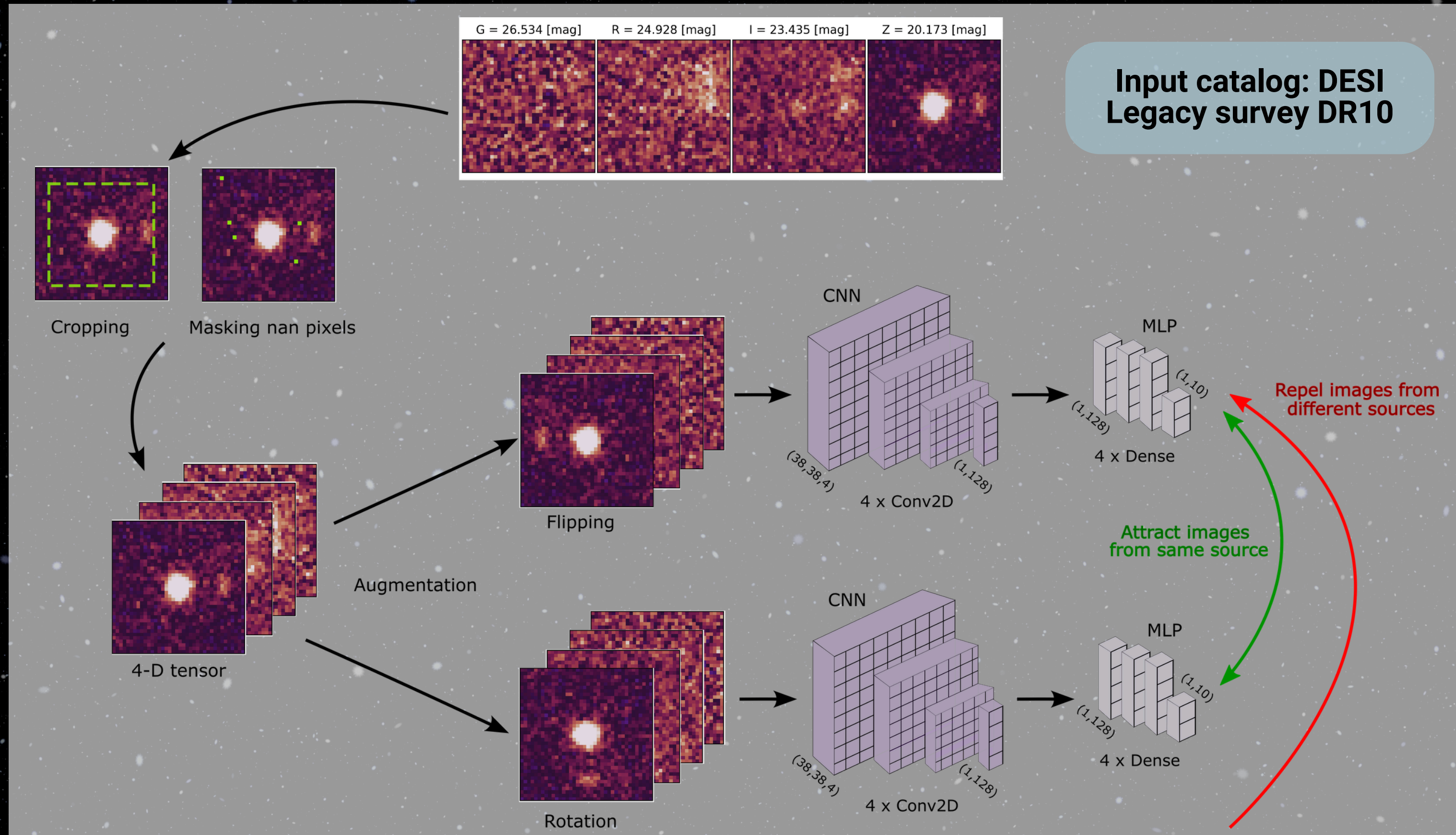


Cosmic rays and artifacts are not fully flagged on catalogs

Traditional selections are biased towards bright, blue QSOs.. and still contaminated by artifacts.

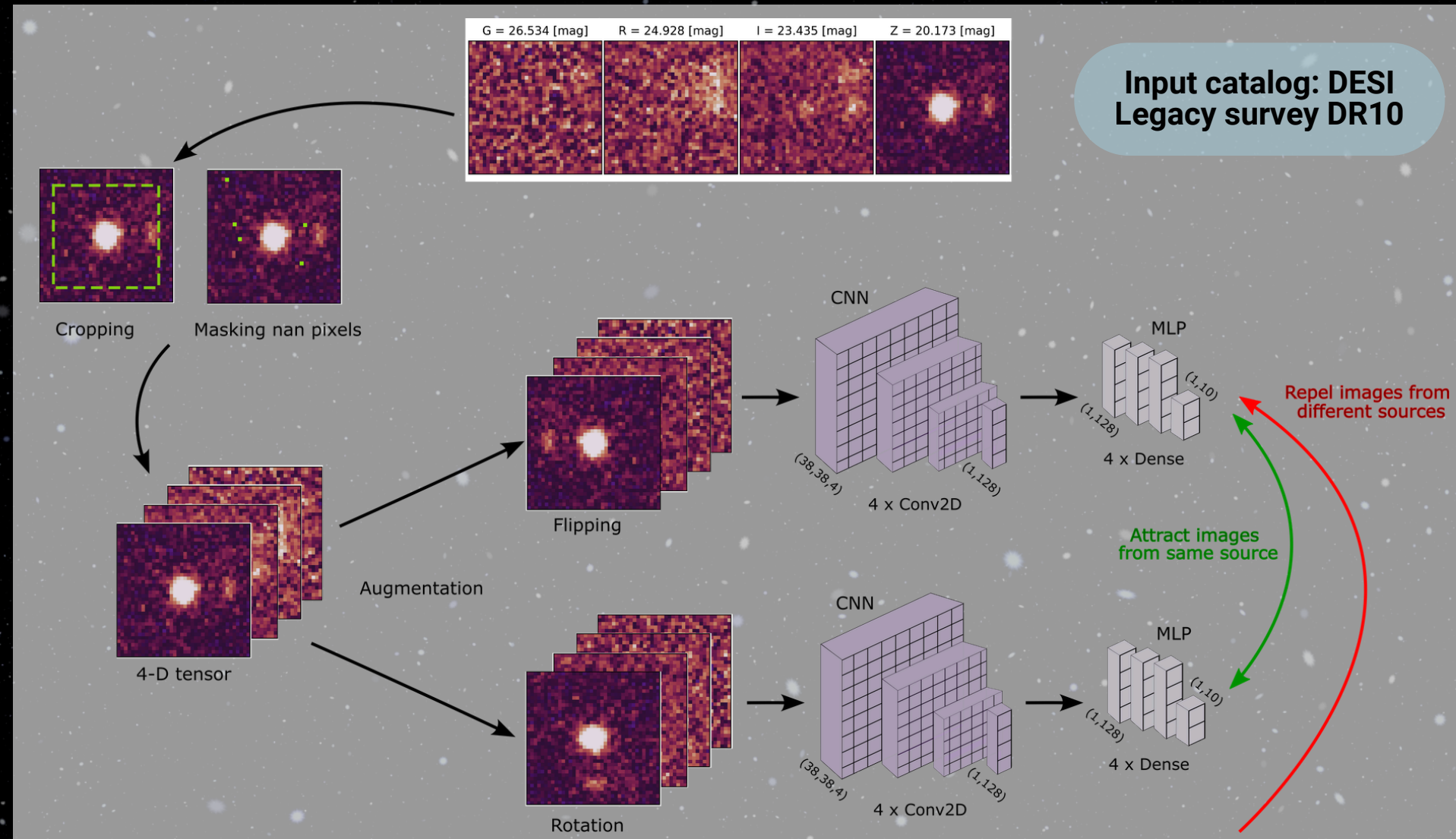
Contrastive learning for high-z quasar selection

6



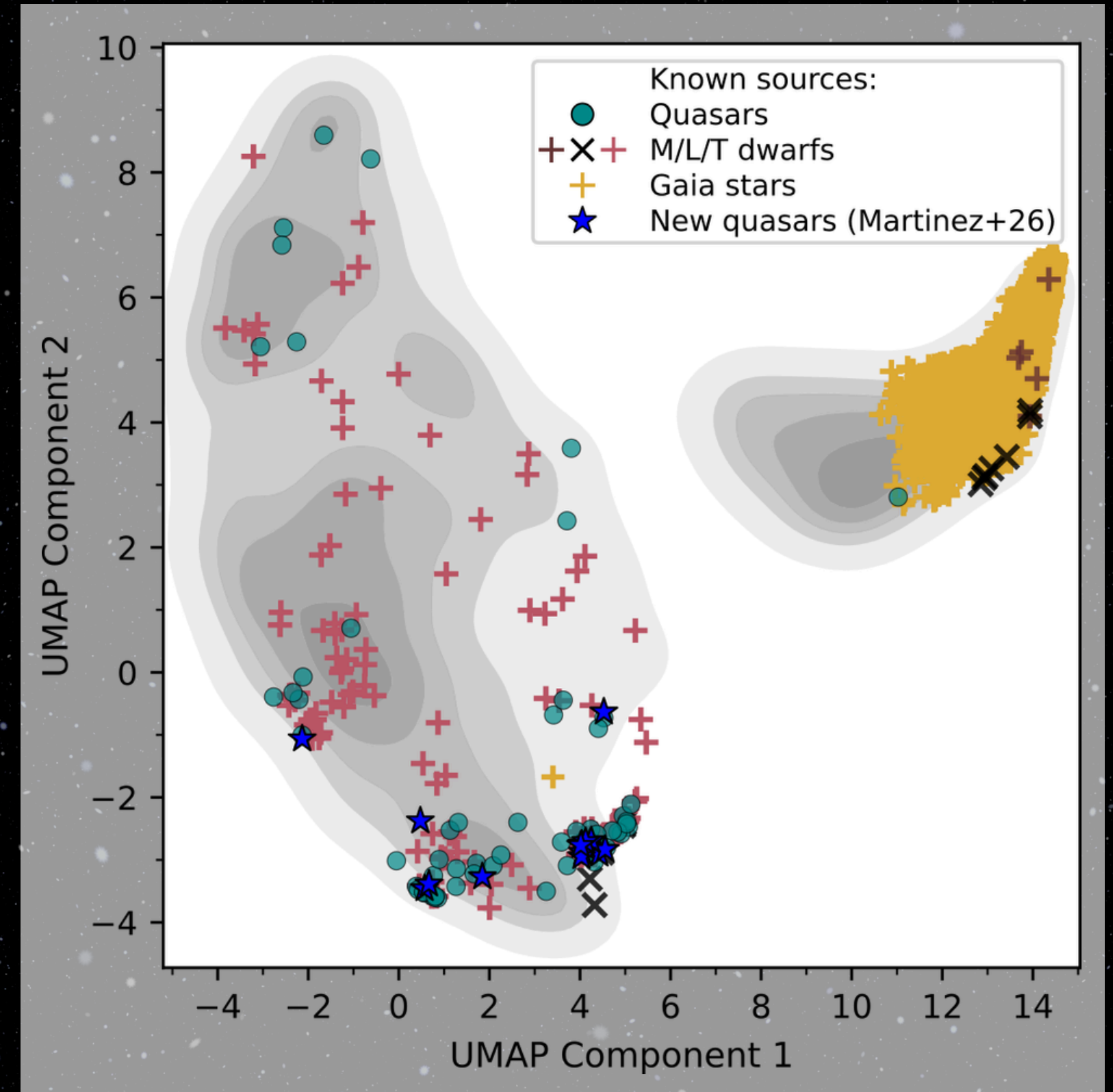
Contrastive learning for high-z quasar selection

7



Martinez-Ramirez et al. 2026

**Latent space
representation!**

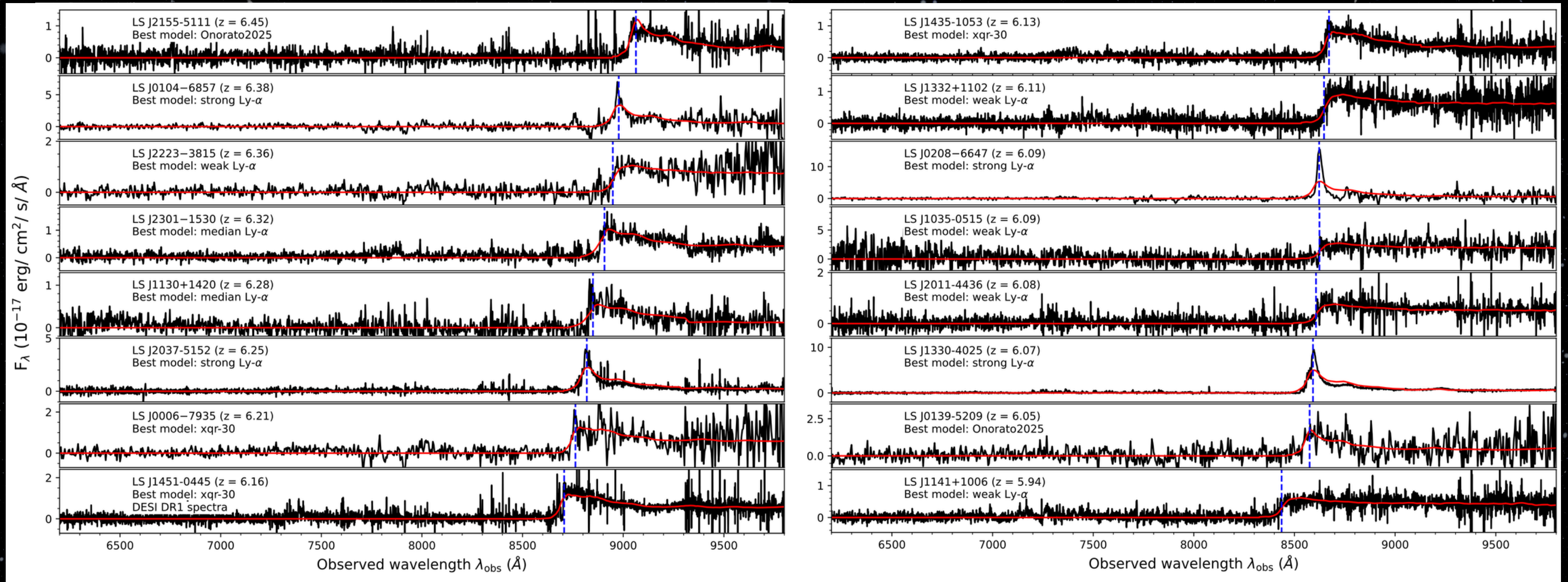


Contrastive learning for high-z quasar selection

8

By selecting candidates from a region in the latent space already populated by known quasars, CL leads to the

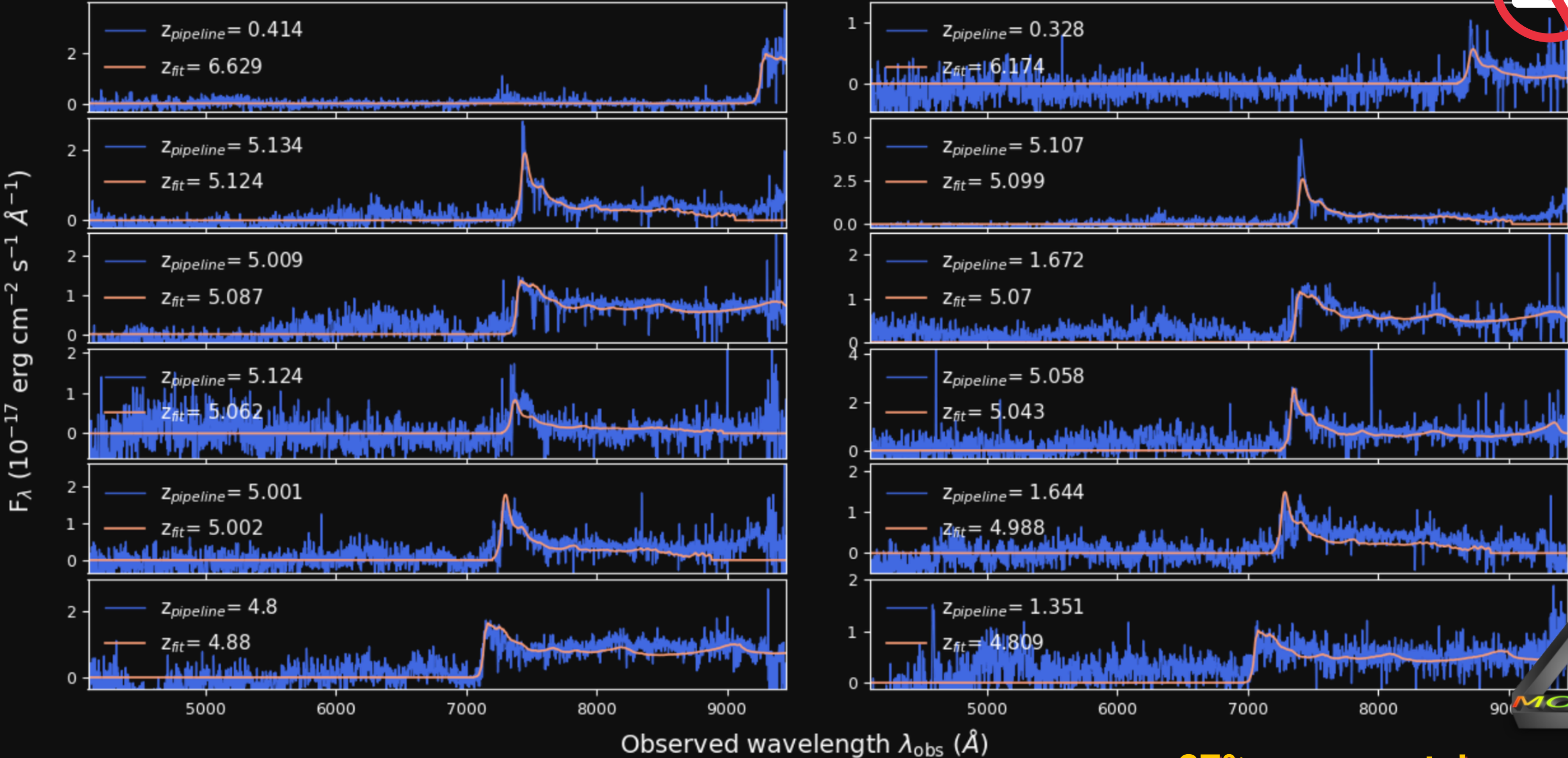
Discovery of 16 new QSOs with CL!



Martinez-Ramirez et al. 2026

45% success rate on spectroscopic follow-up!

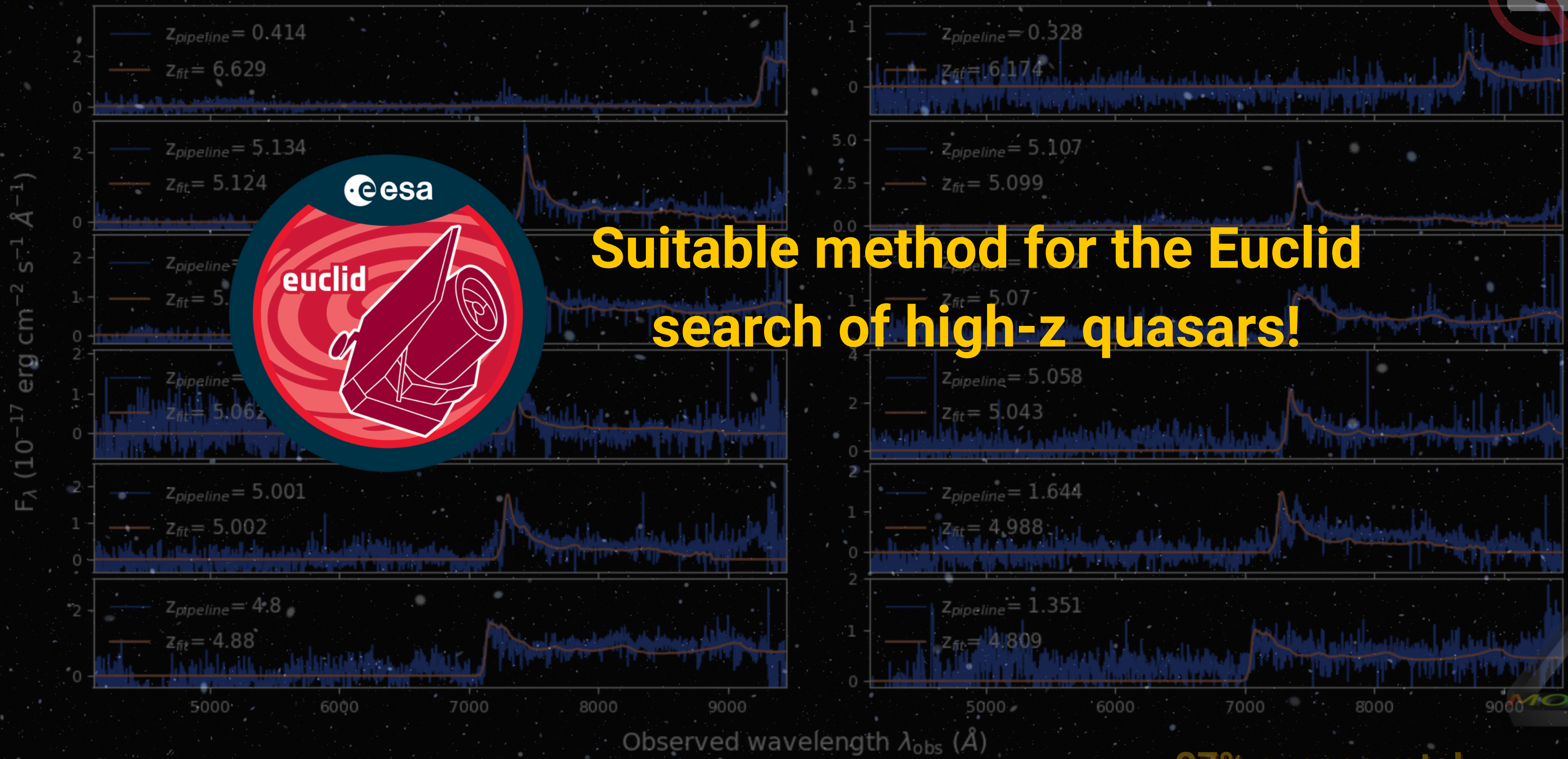
4MOST Preliminary!



27% sucess rate!



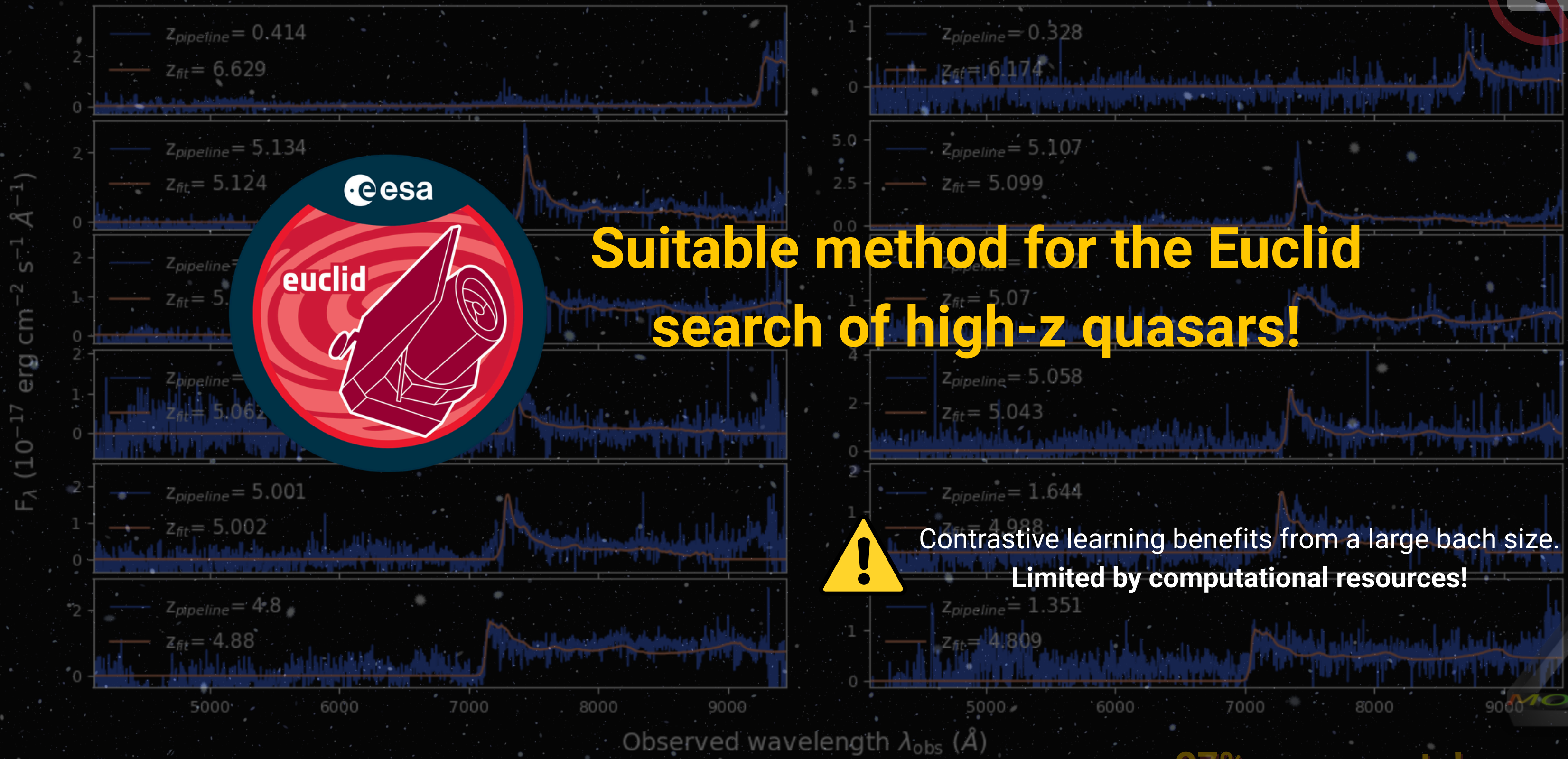
4MOST Preliminary!



Suitable method for the Euclid search of high-z quasars!

27% sucess rate!

4MOST Preliminary!



Suitable method for the Euclid search of high-z quasars!



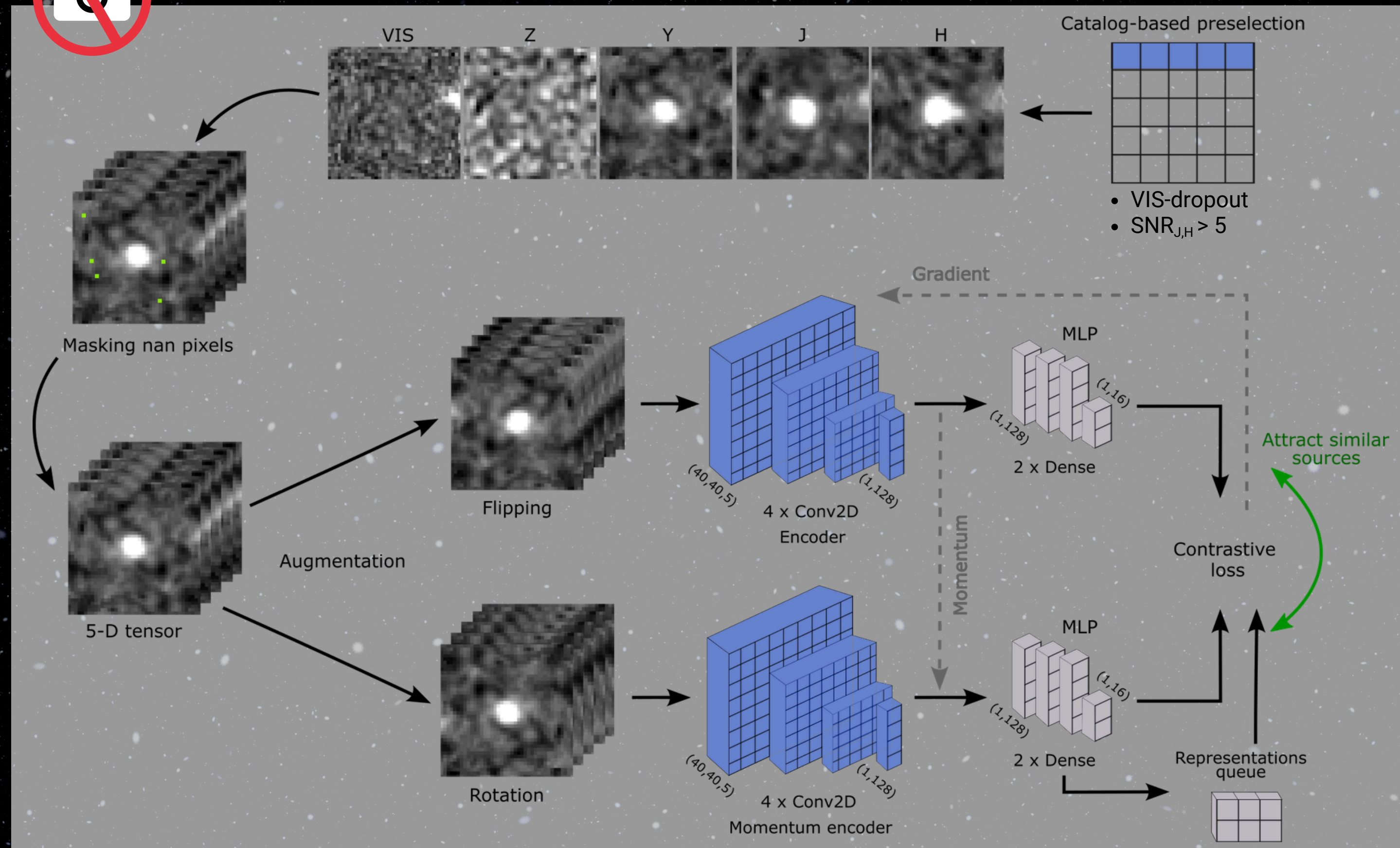
Contrastive learning benefits from a large batch size.
Limited by computational resources!

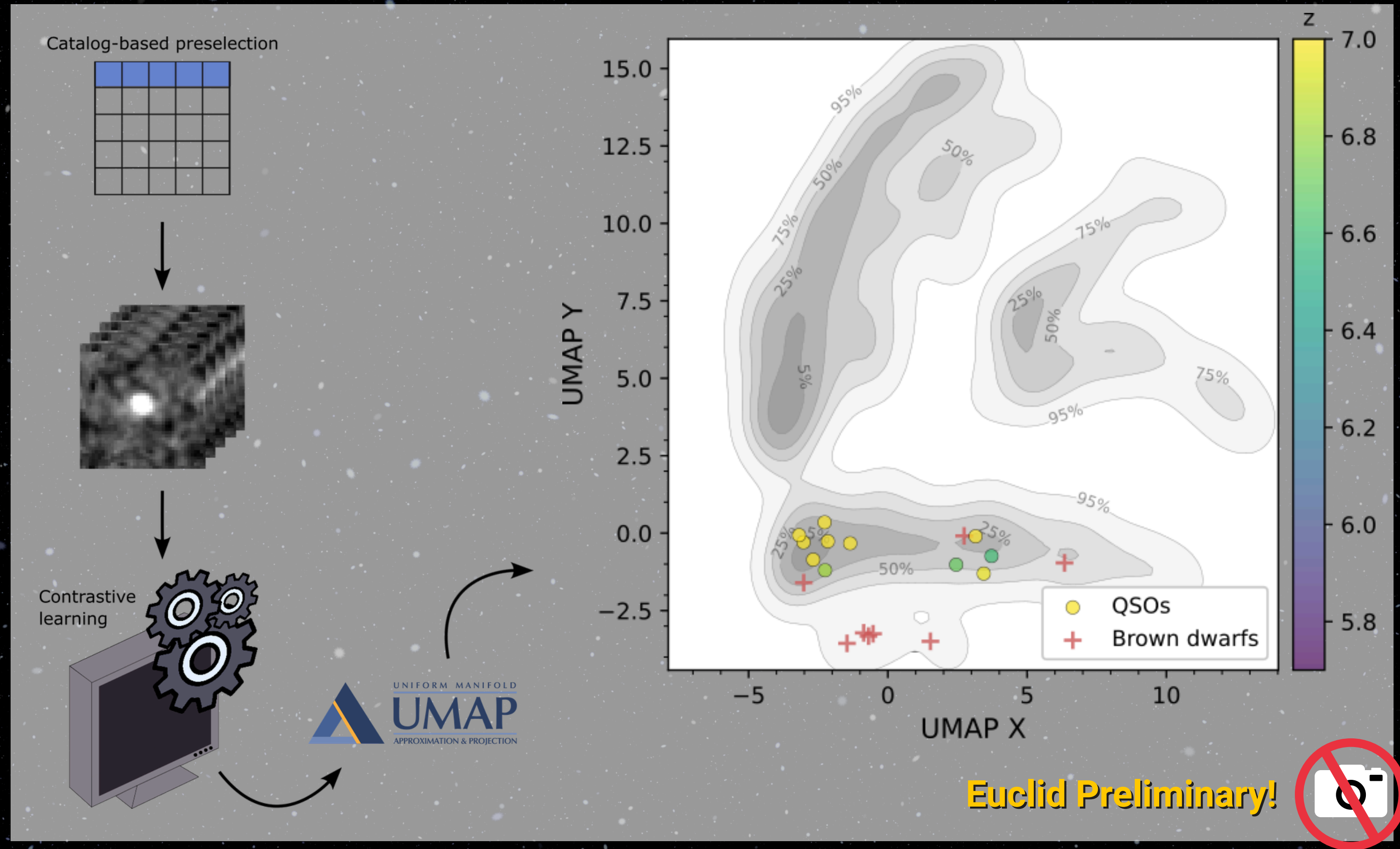
27% sucess rate!

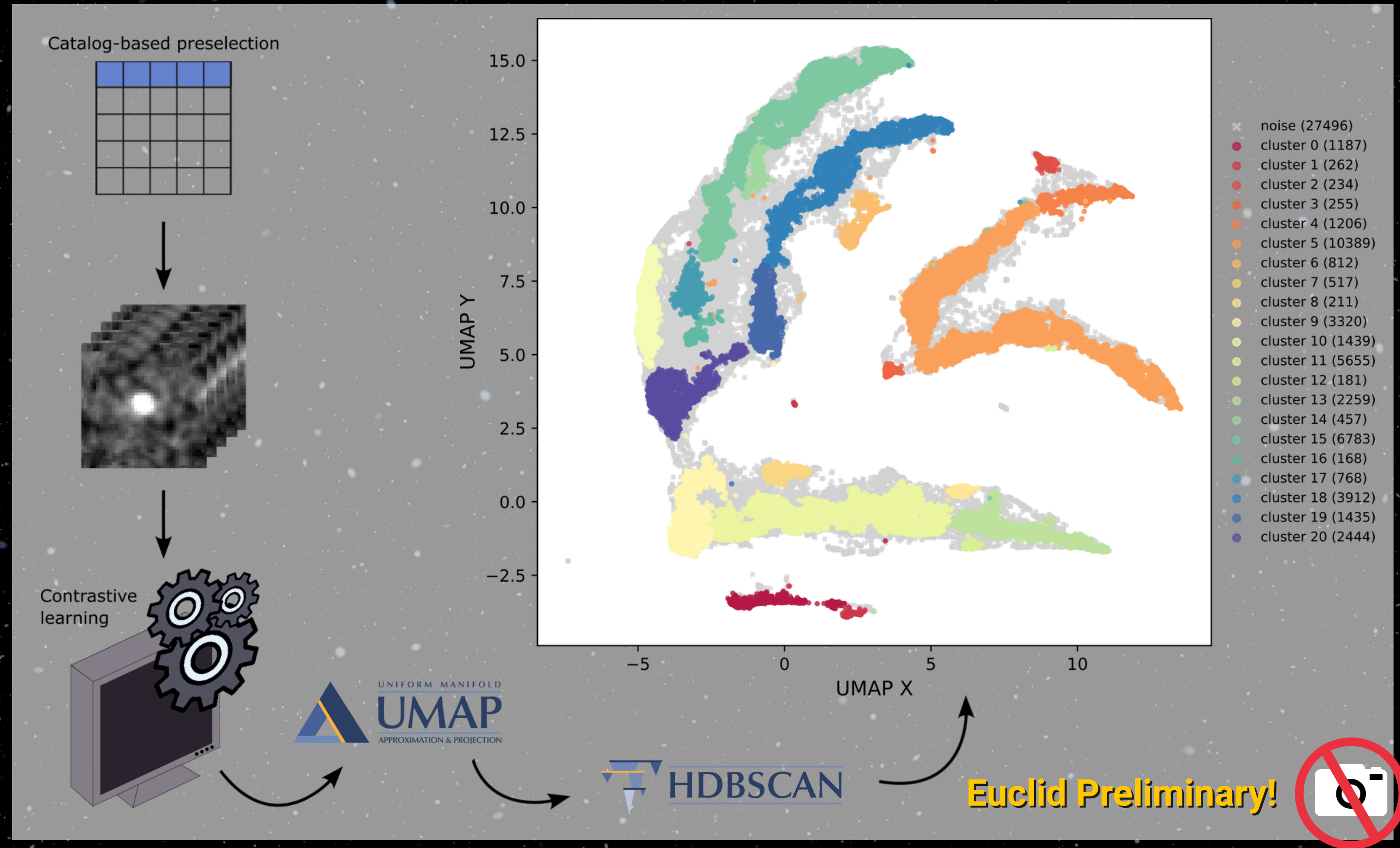


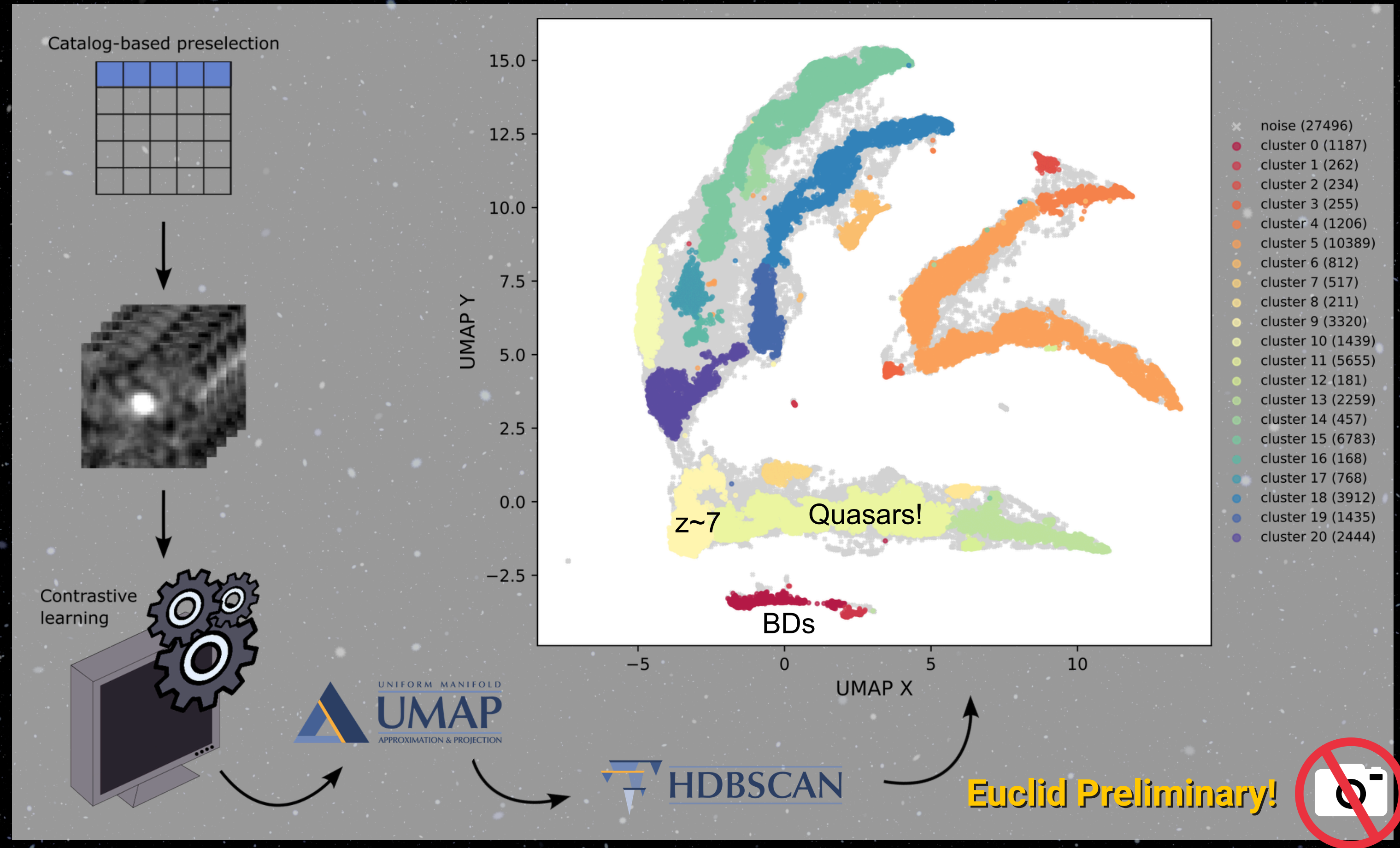


Euclid Preliminary!

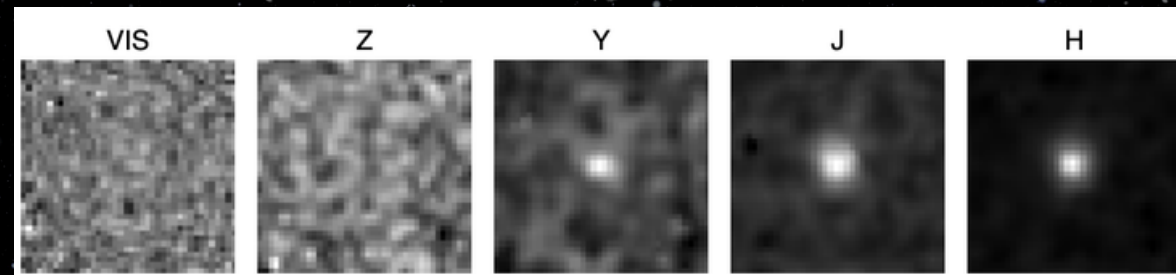
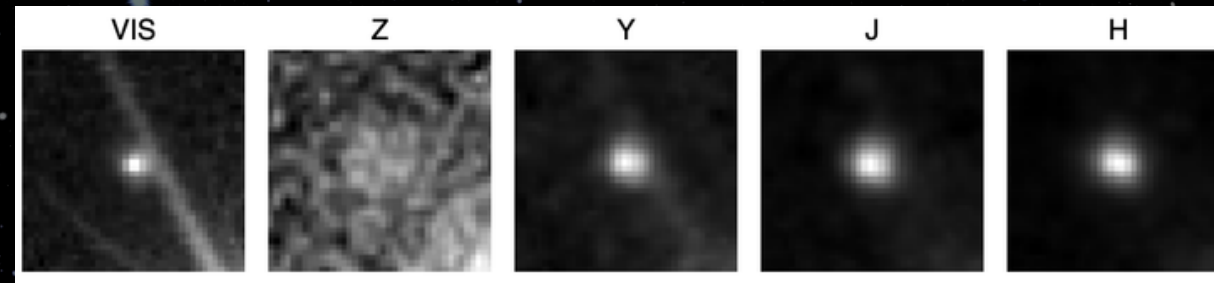




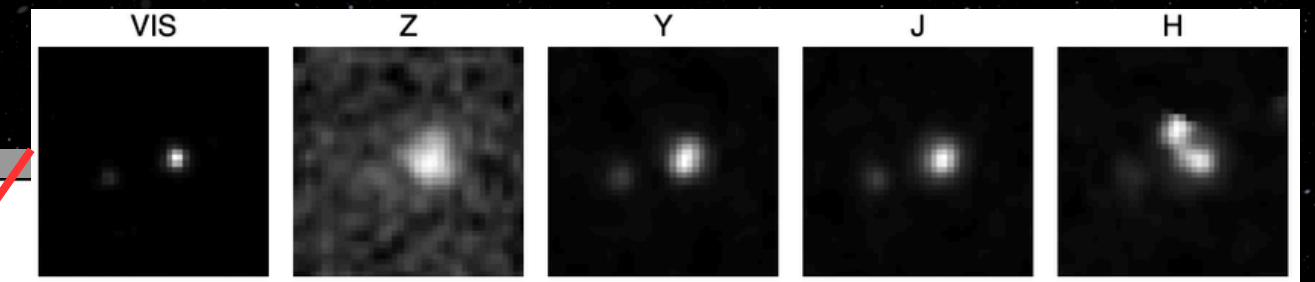




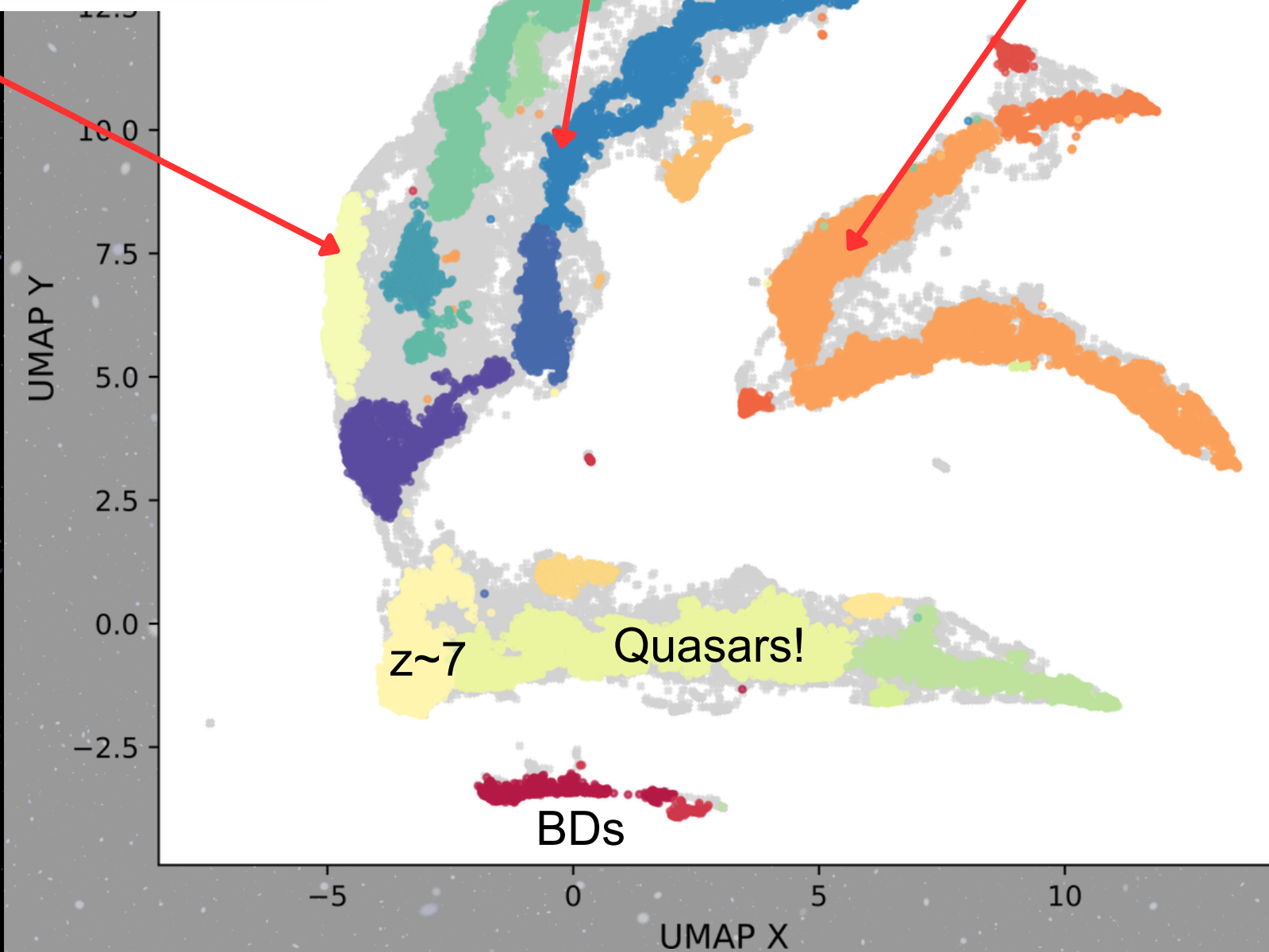
Diffraction spikes



More dropouts!



Sources with companions

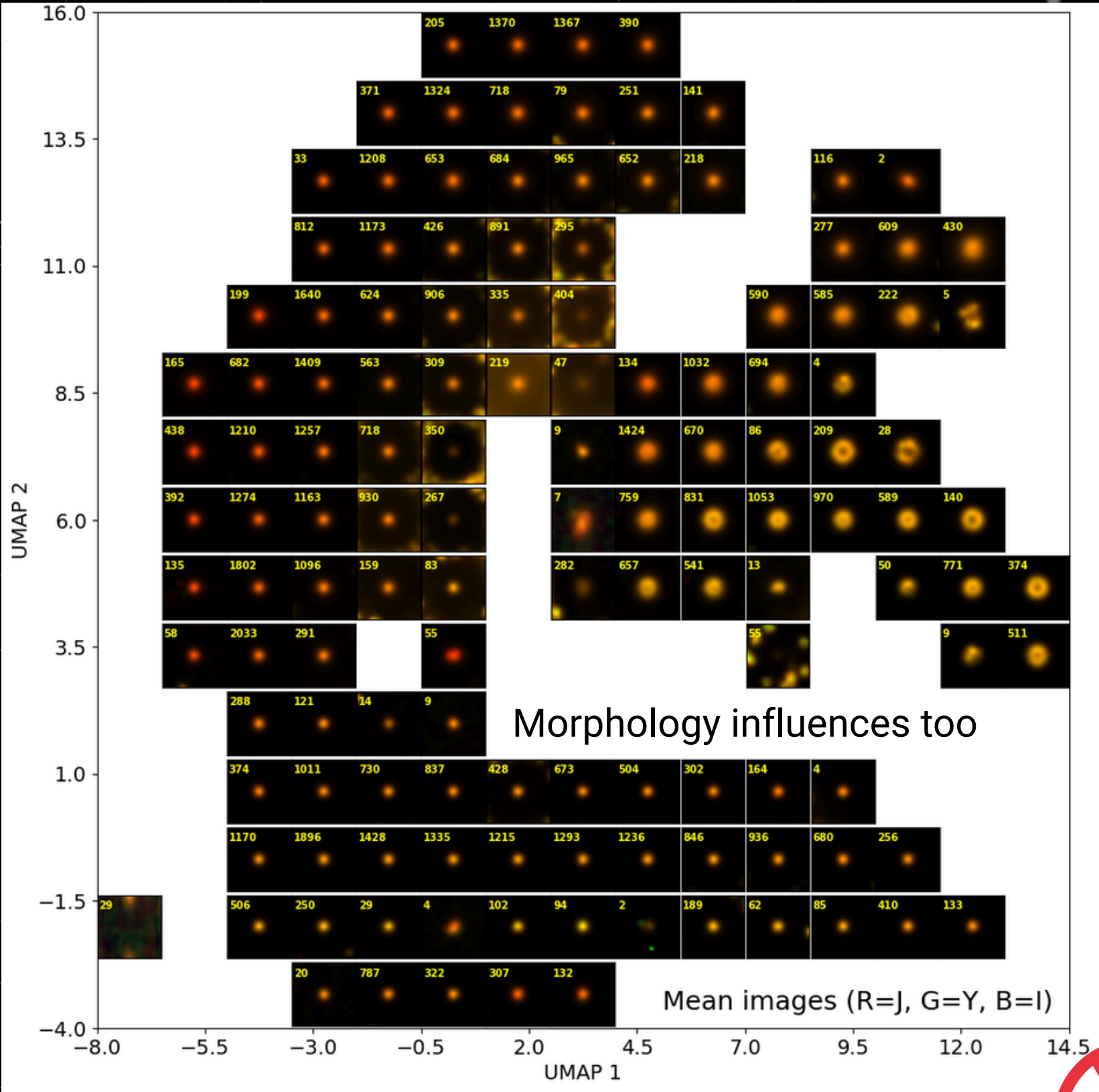
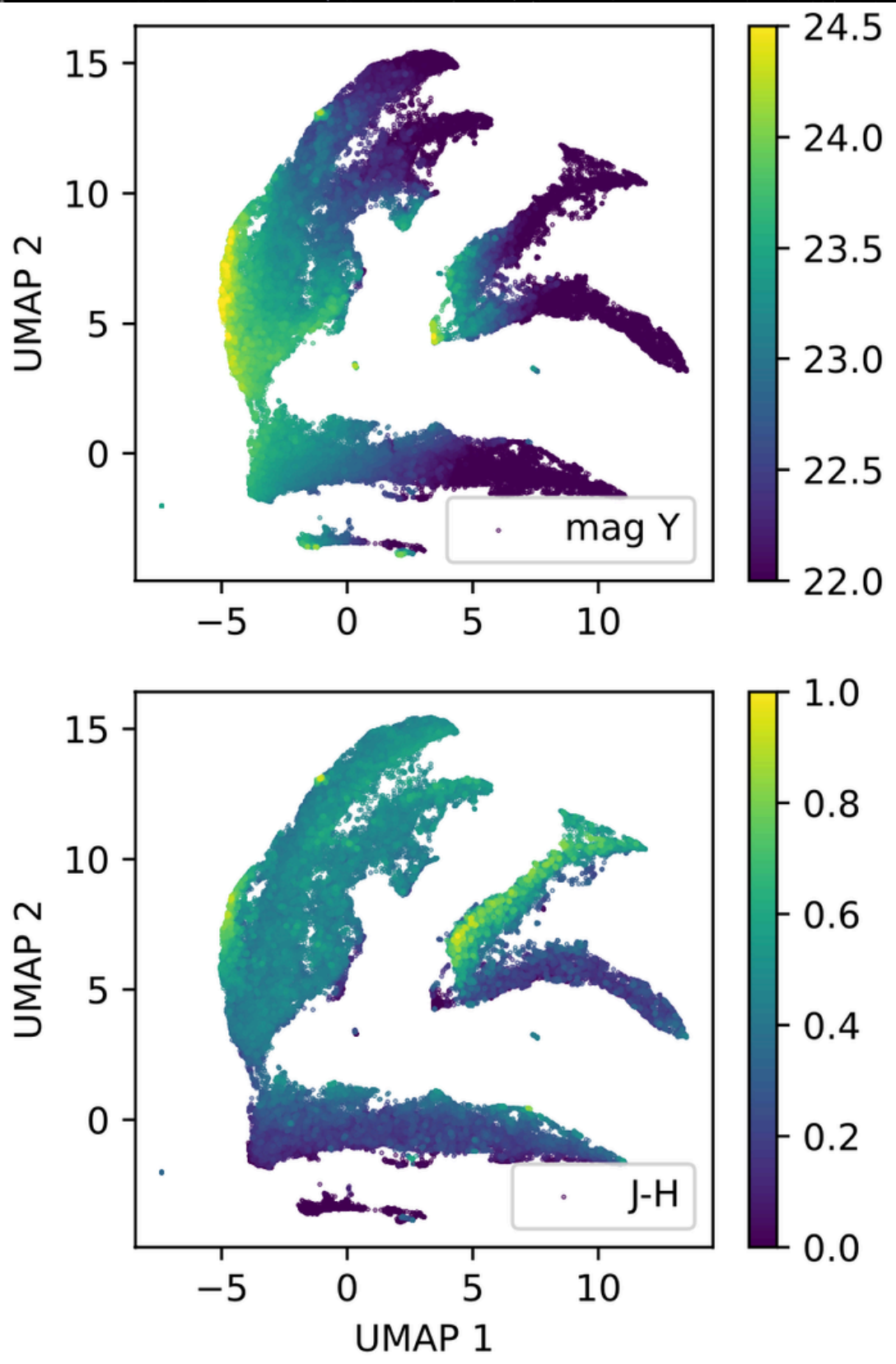


- × noise (27496)
- cluster 0 (1187)
- cluster 1 (262)
- cluster 2 (234)
- cluster 3 (255)
- cluster 4 (1206)
- cluster 5 (10389)
- cluster 6 (812)
- cluster 7 (517)
- cluster 8 (211)
- cluster 9 (3320)
- cluster 10 (1439)
- cluster 11 (5655)
- cluster 12 (181)
- cluster 13 (2259)
- cluster 14 (457)
- cluster 15 (6783)
- cluster 16 (168)
- cluster 17 (768)
- cluster 18 (3912)
- cluster 19 (1435)
- cluster 20 (2444)



Euclid Preliminary!

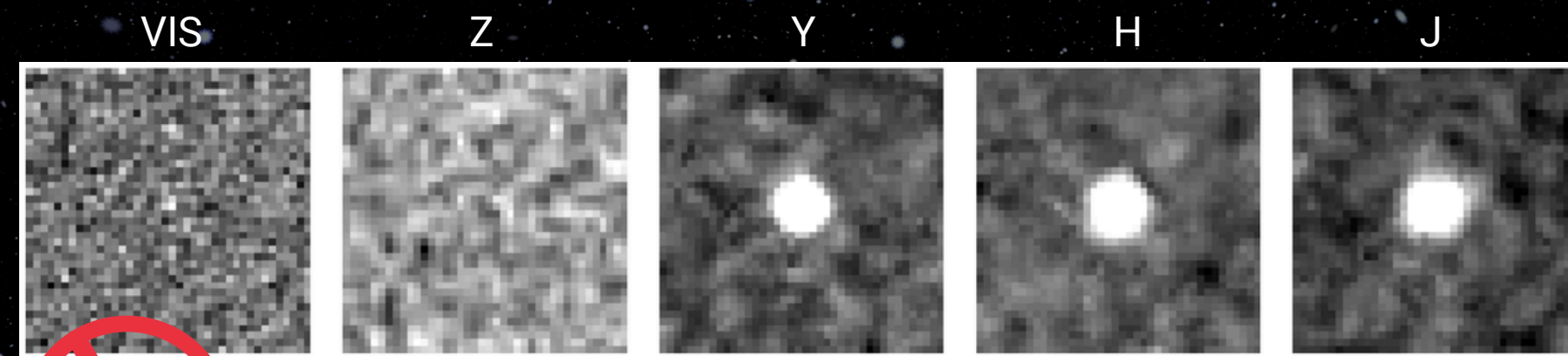
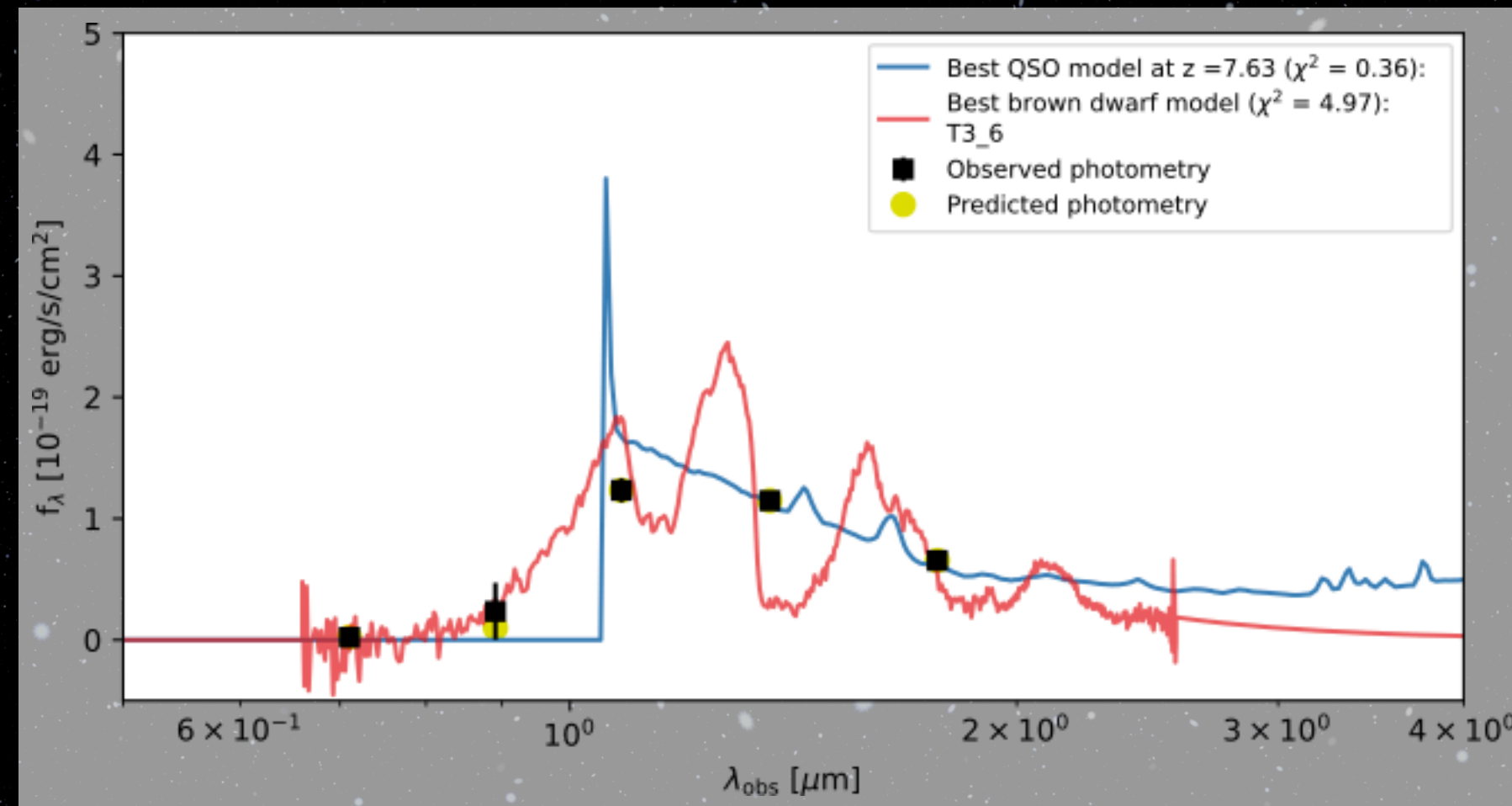
Gradient in brightness and colors



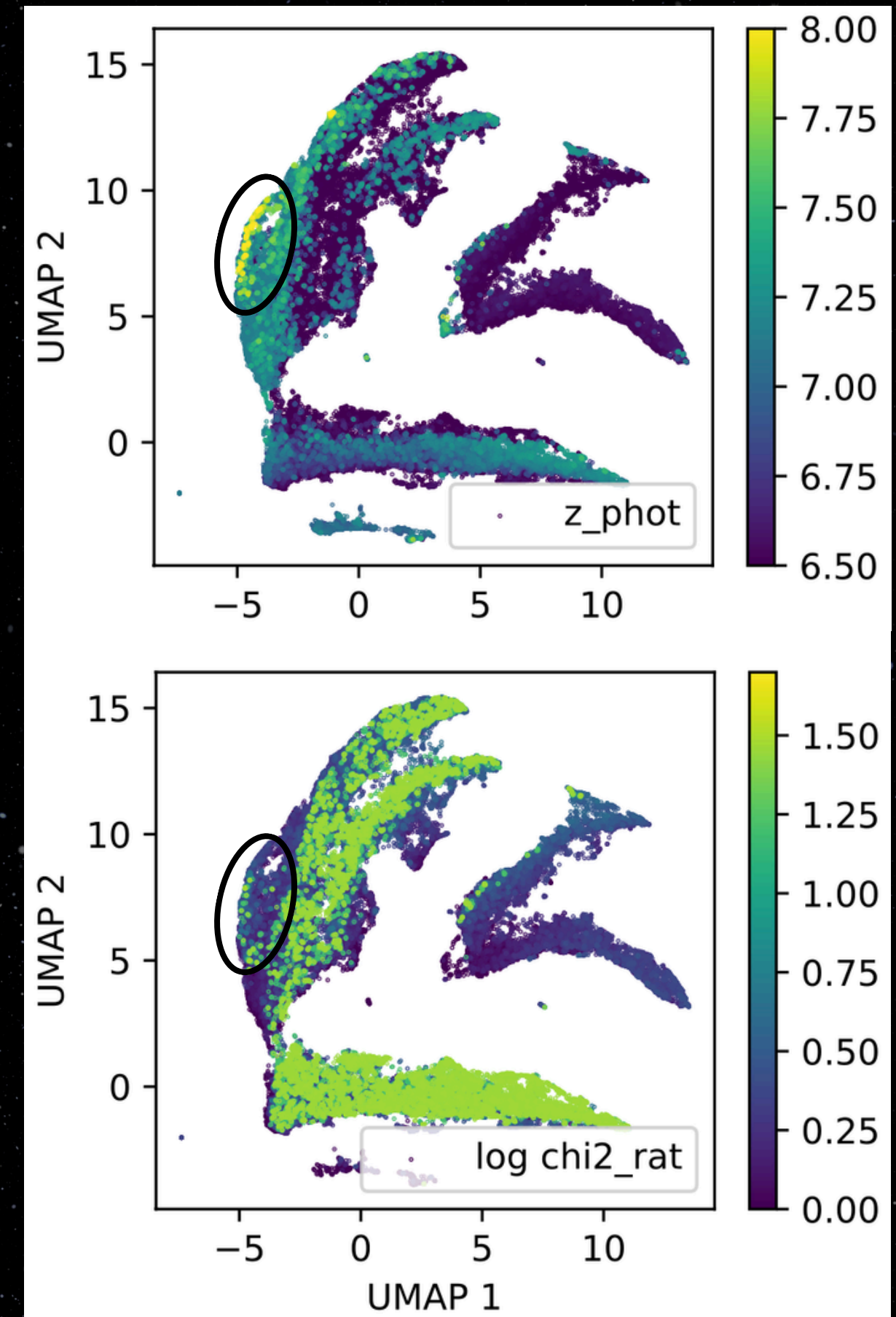
Euclid Preliminary!



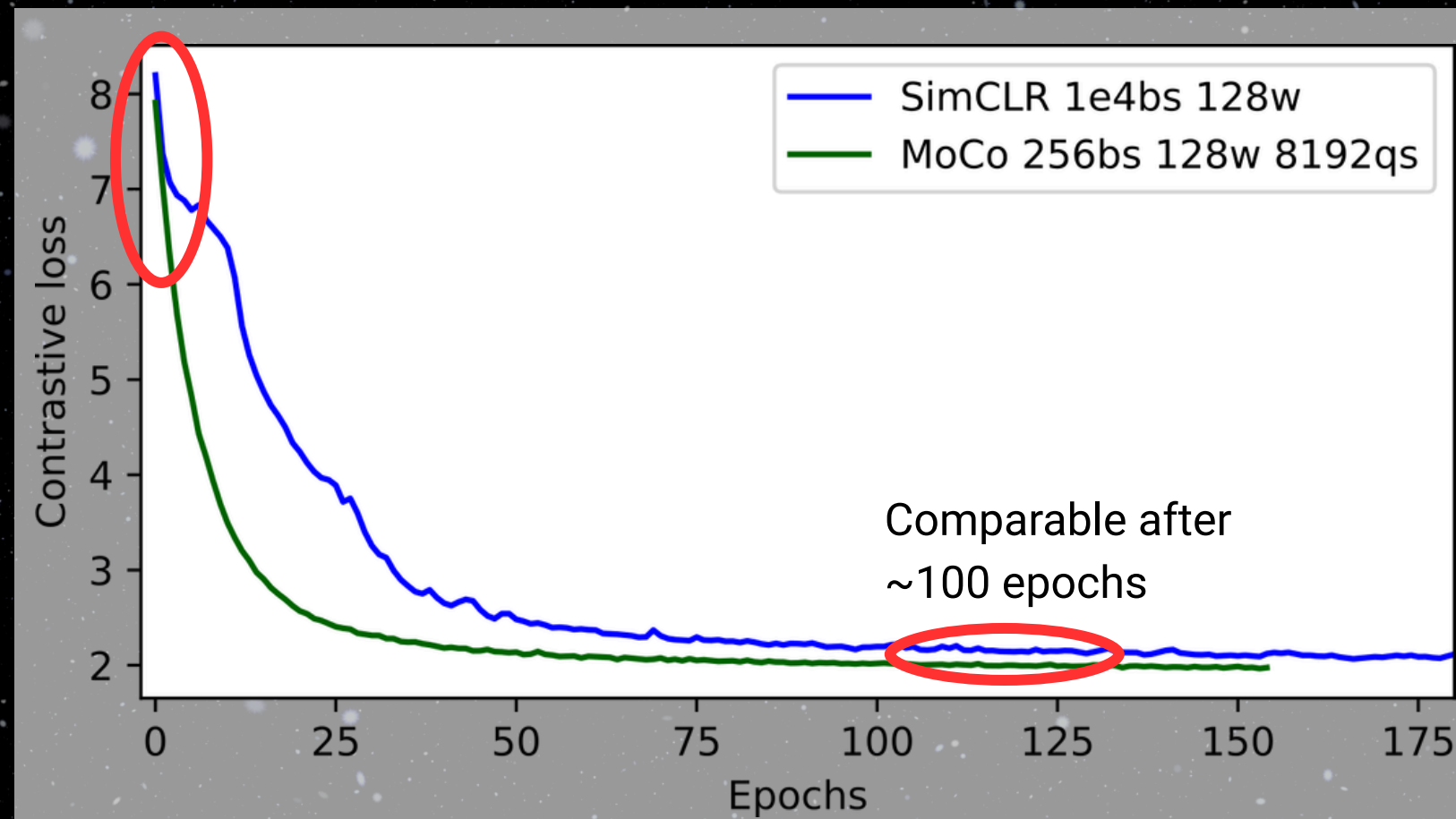
SED fitting for prioritization



Euclid Preliminary!



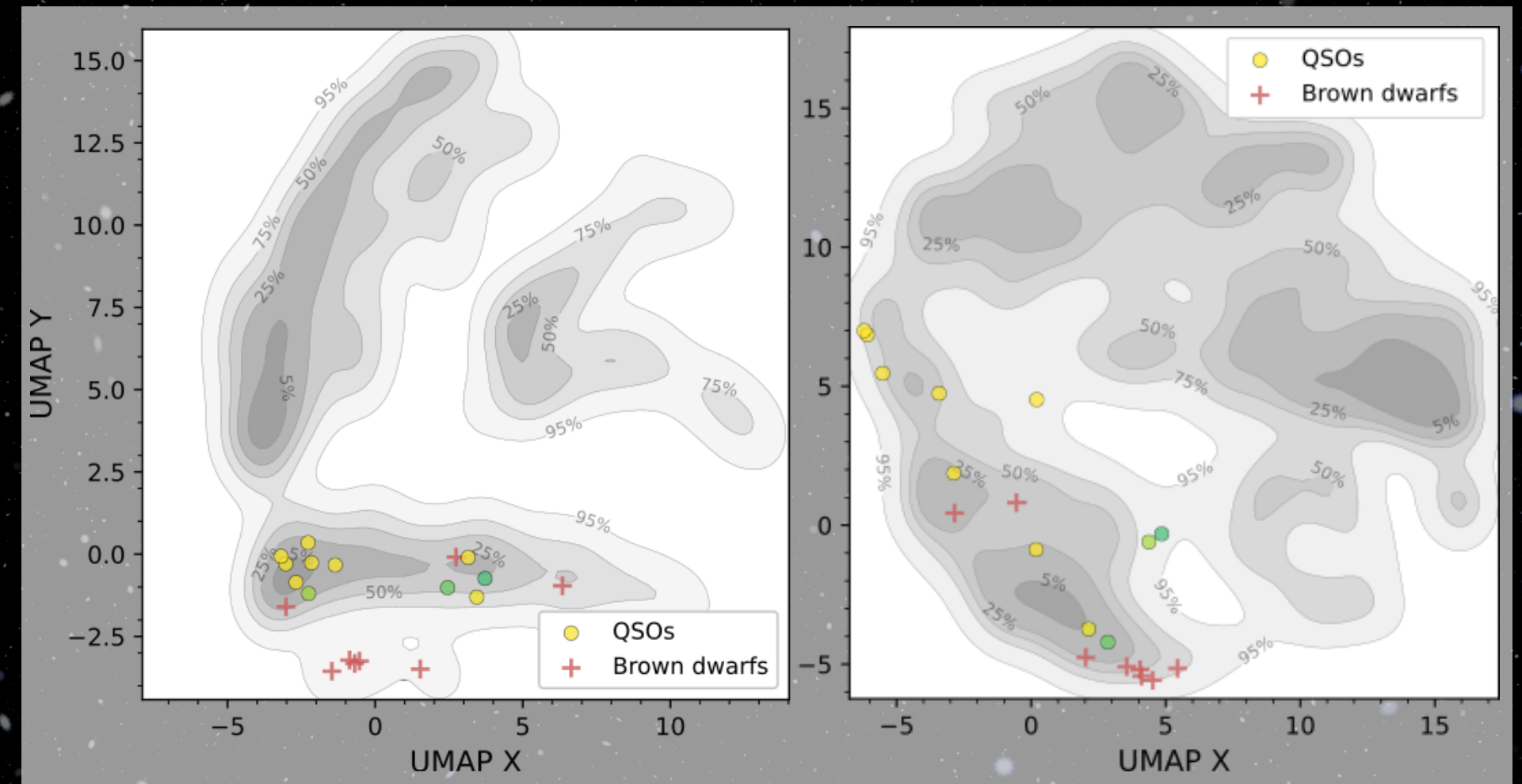
Effects of larger negative pairs!



MoCo converges faster.



Euclid Preliminary!



But MoCo representation:

- Has more clustered quasar population
- is highly informative with clear effects of morphology and environment

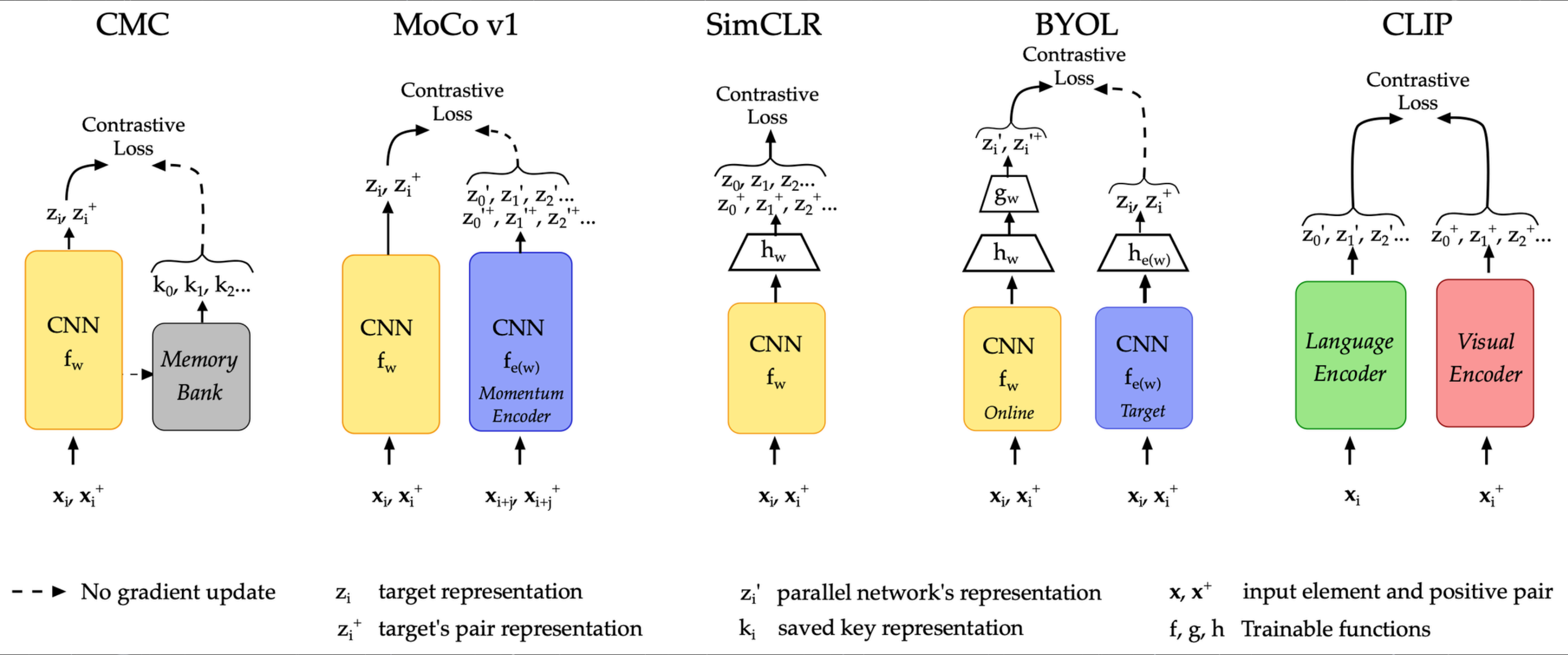
TAKE-HOME MESSAGES

- The era of large photometric and spectroscopic surveys provides unprecedented opportunities for **astrophysical data mining**.
- **Self supervised machine learning** provides a powerful, label-efficient framework for high- z quasar searches beyond the limits of traditional selections. CL is a very promising avenue!
- The first 4MOST spectra and individual campaigns already validate the high- z selection strategies, with 28 quasars confirmed at $4.8 < z < 6.6$.
- While labels guide interpretation, the latent space can reveal both known populations and rare astrophysical objects.

Stay tuned for future quasar discoveries!

BACKUP

Contrastive learning flavors



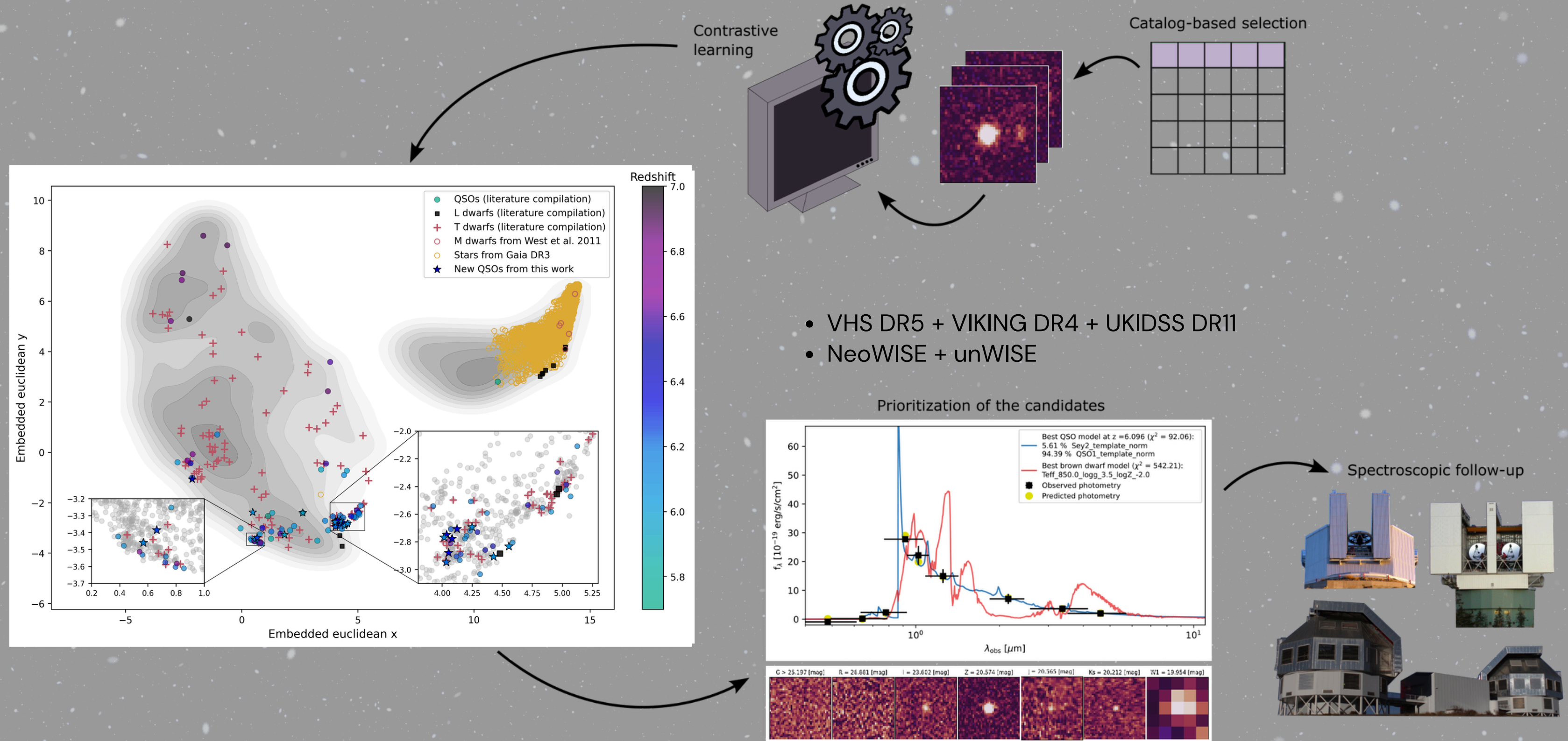
Old embeddings $\rightarrow$
noise in loss function

More negatives,
stable training

Limited by
batch size

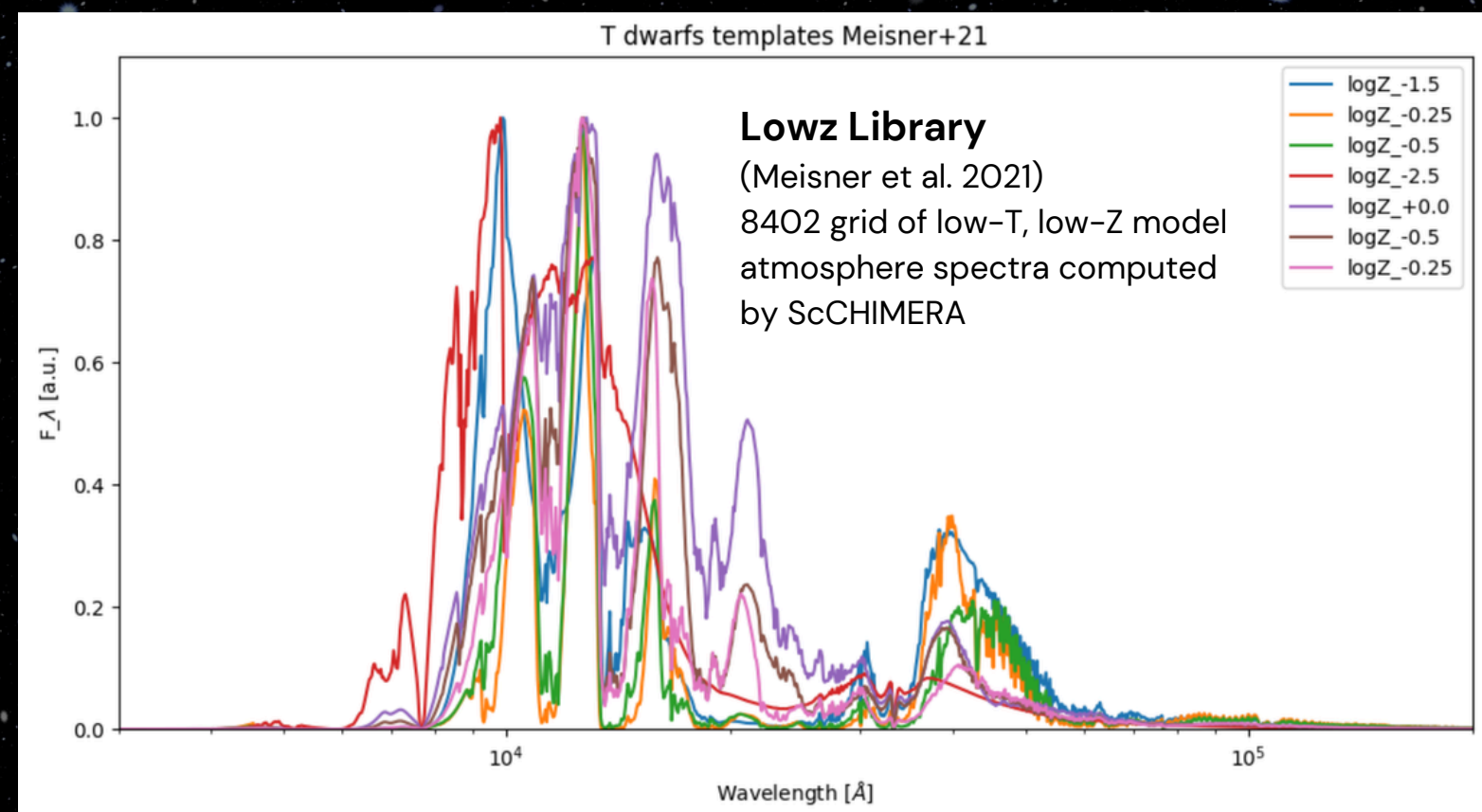
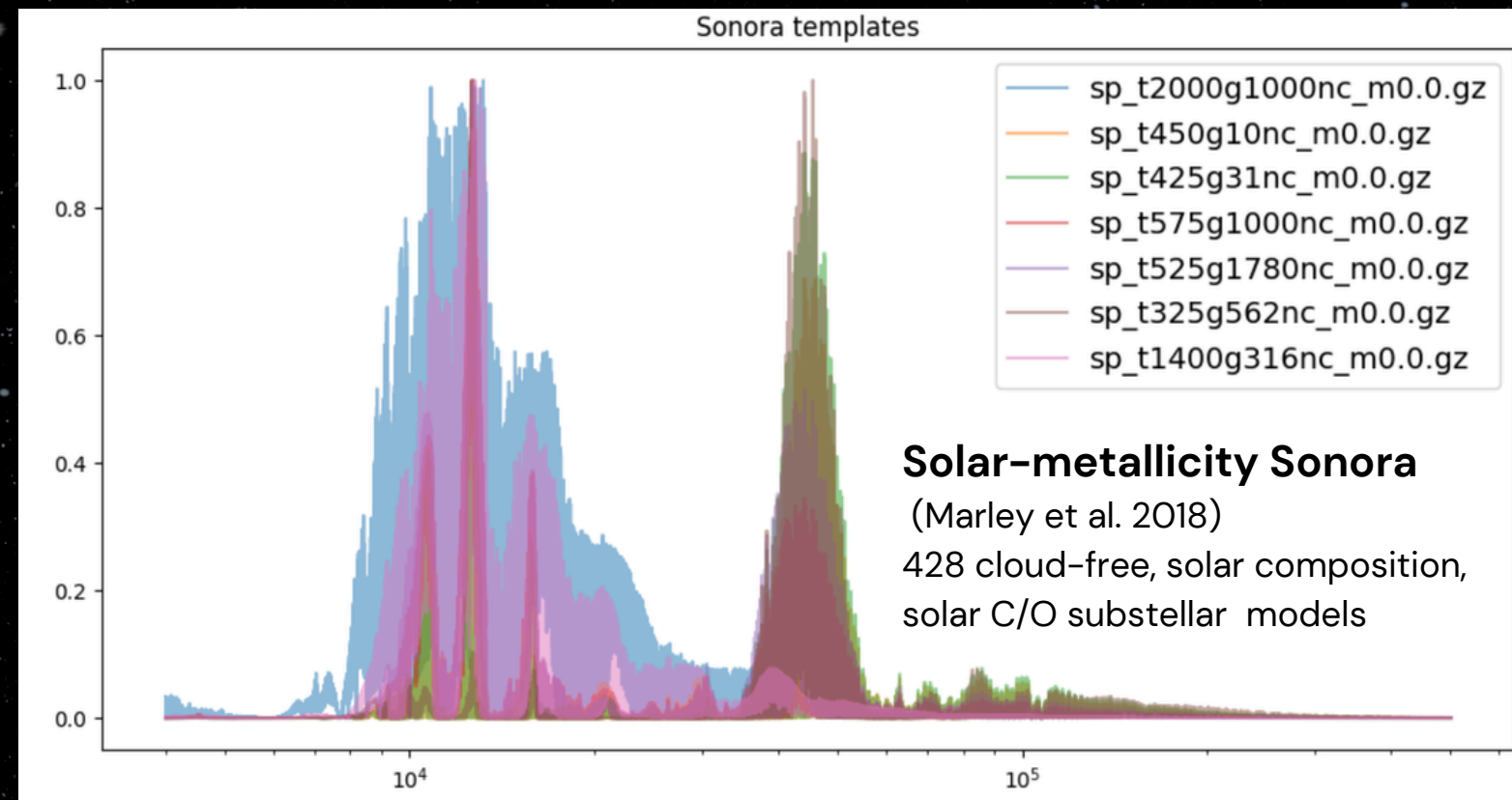
No negative pairs $\rightarrow$
Critical for quasar vs
dwarfs separation

Selection pipeline with simCLR

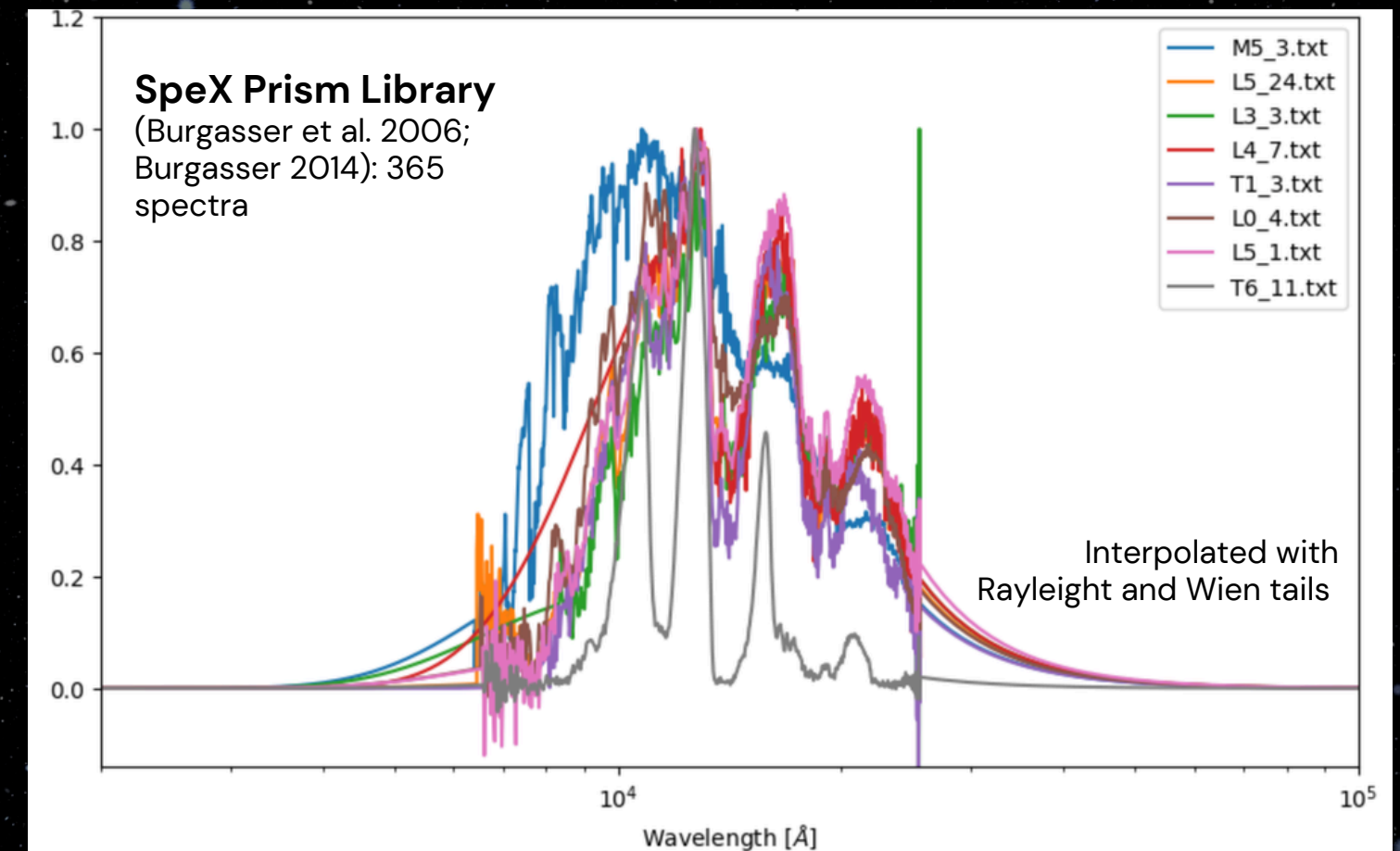


*This strategy was also applied for the selection of $z > 4$ quasar candidates for 4MOST-ChaNGES survey.

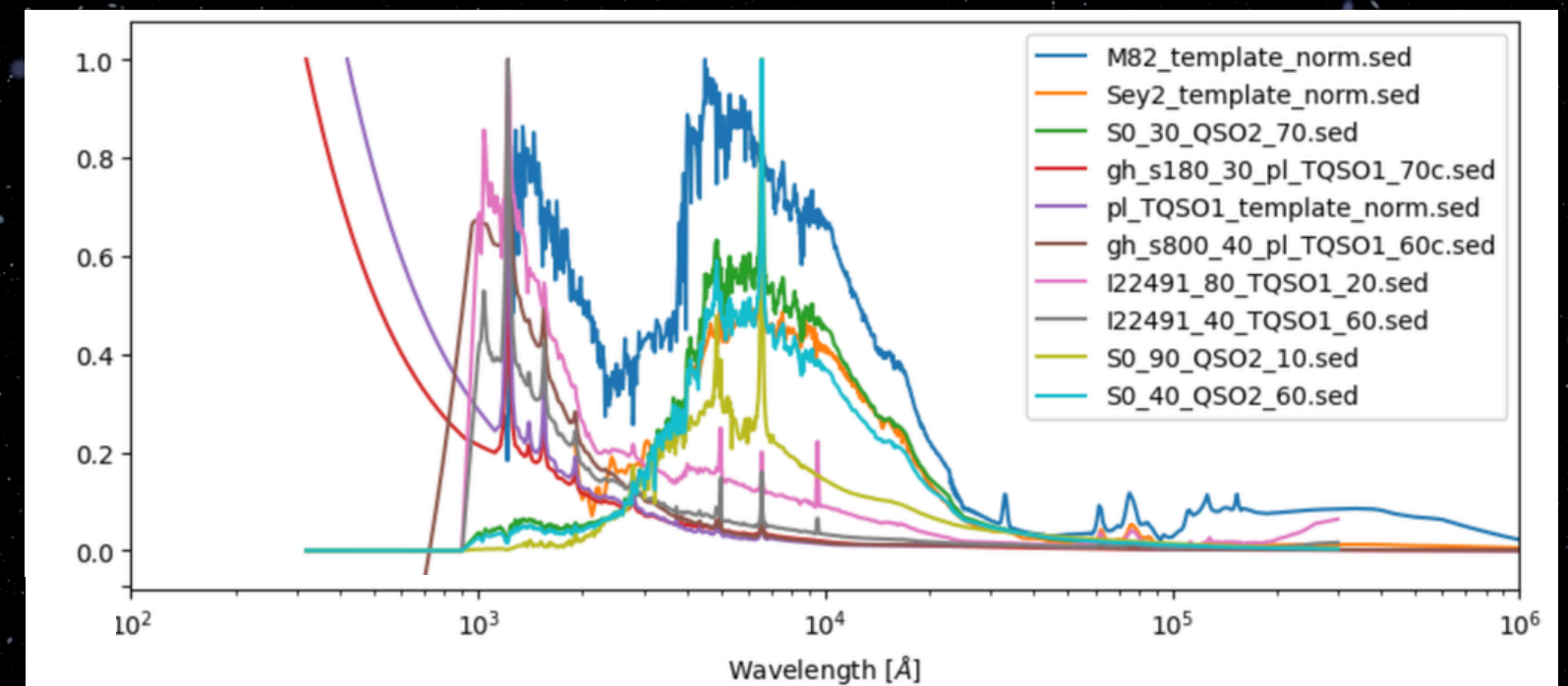
Templates in Eazy-py



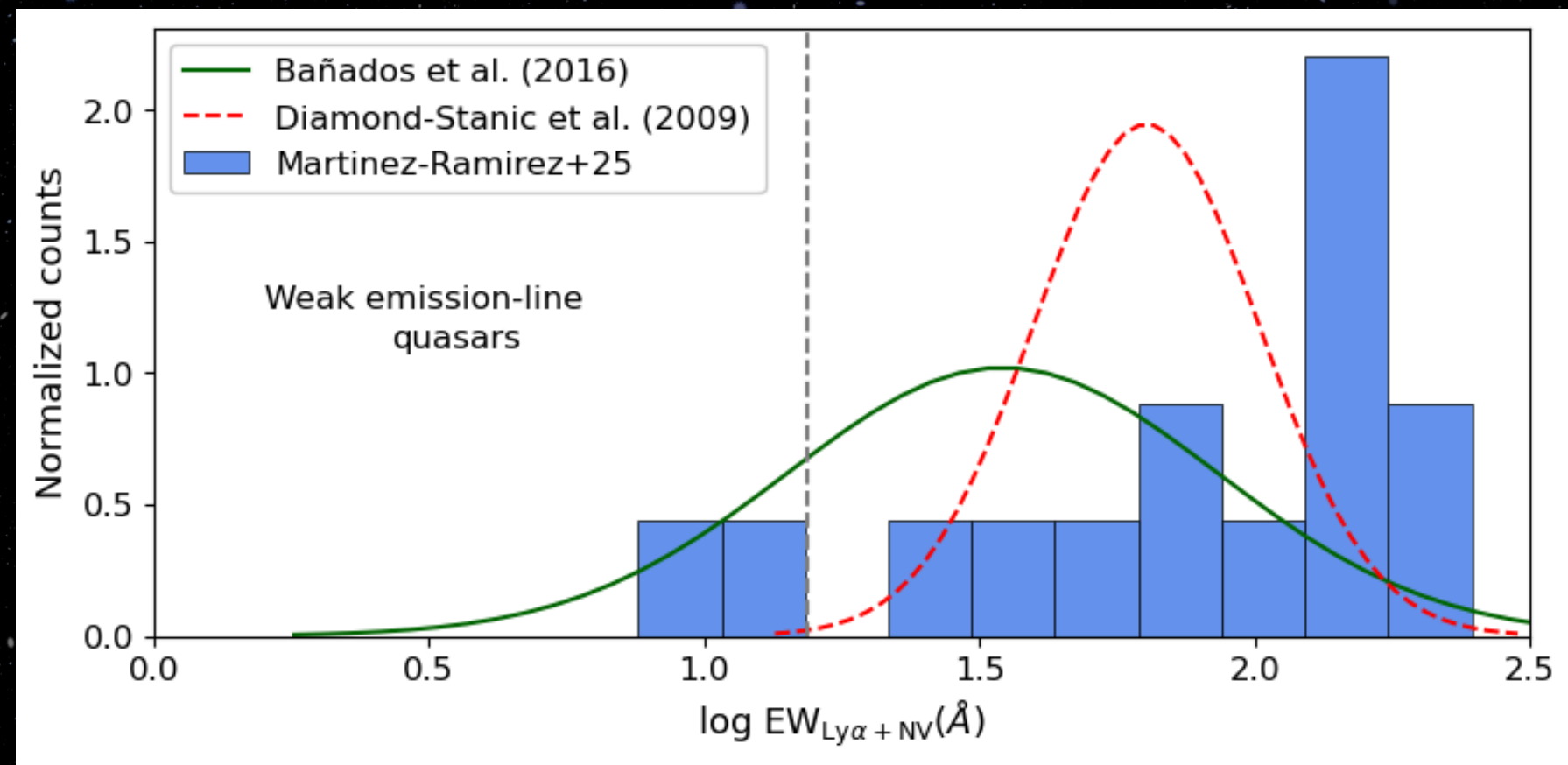
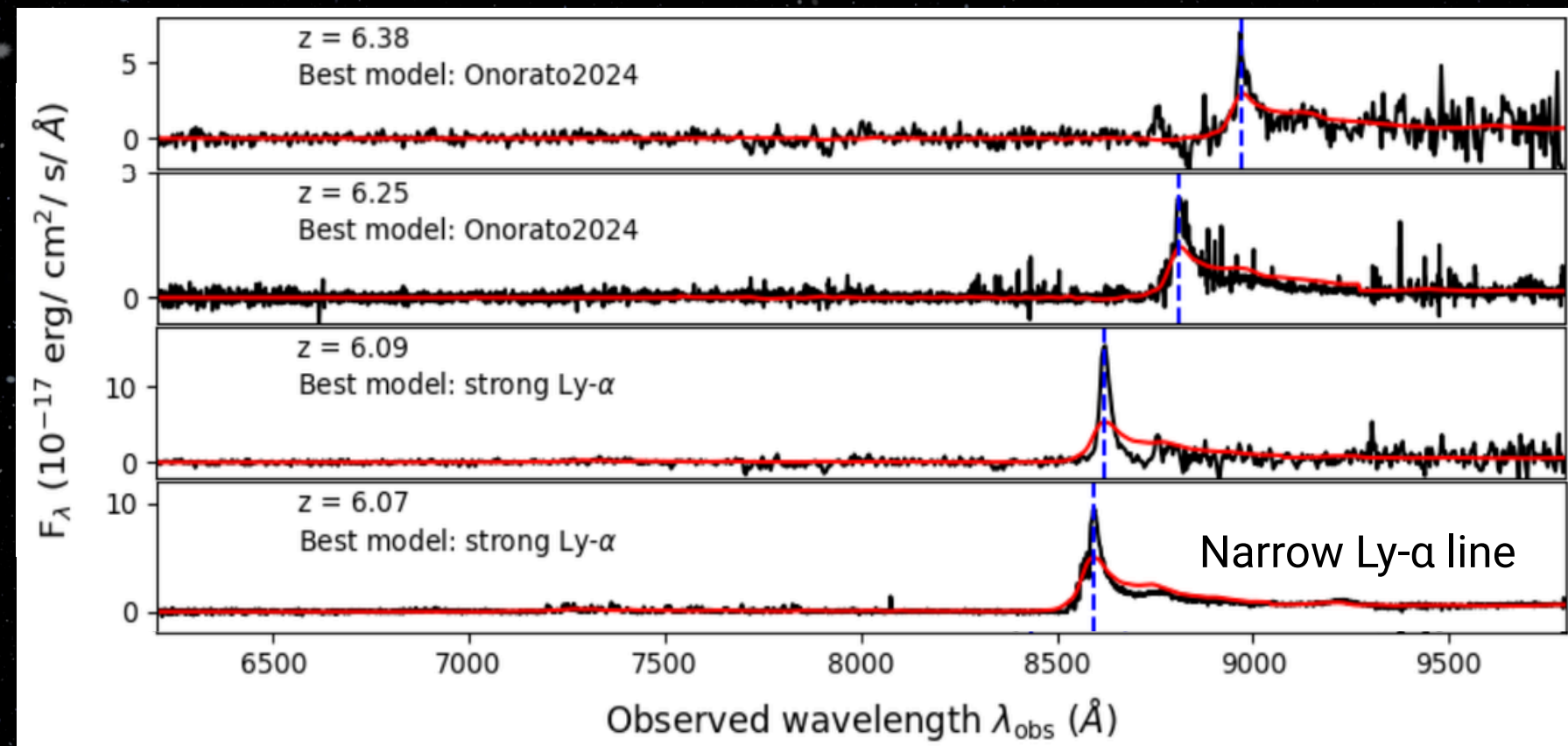
TRDS Pickles Atlas, AMES-Cond, AMES-Dusty, NextGen



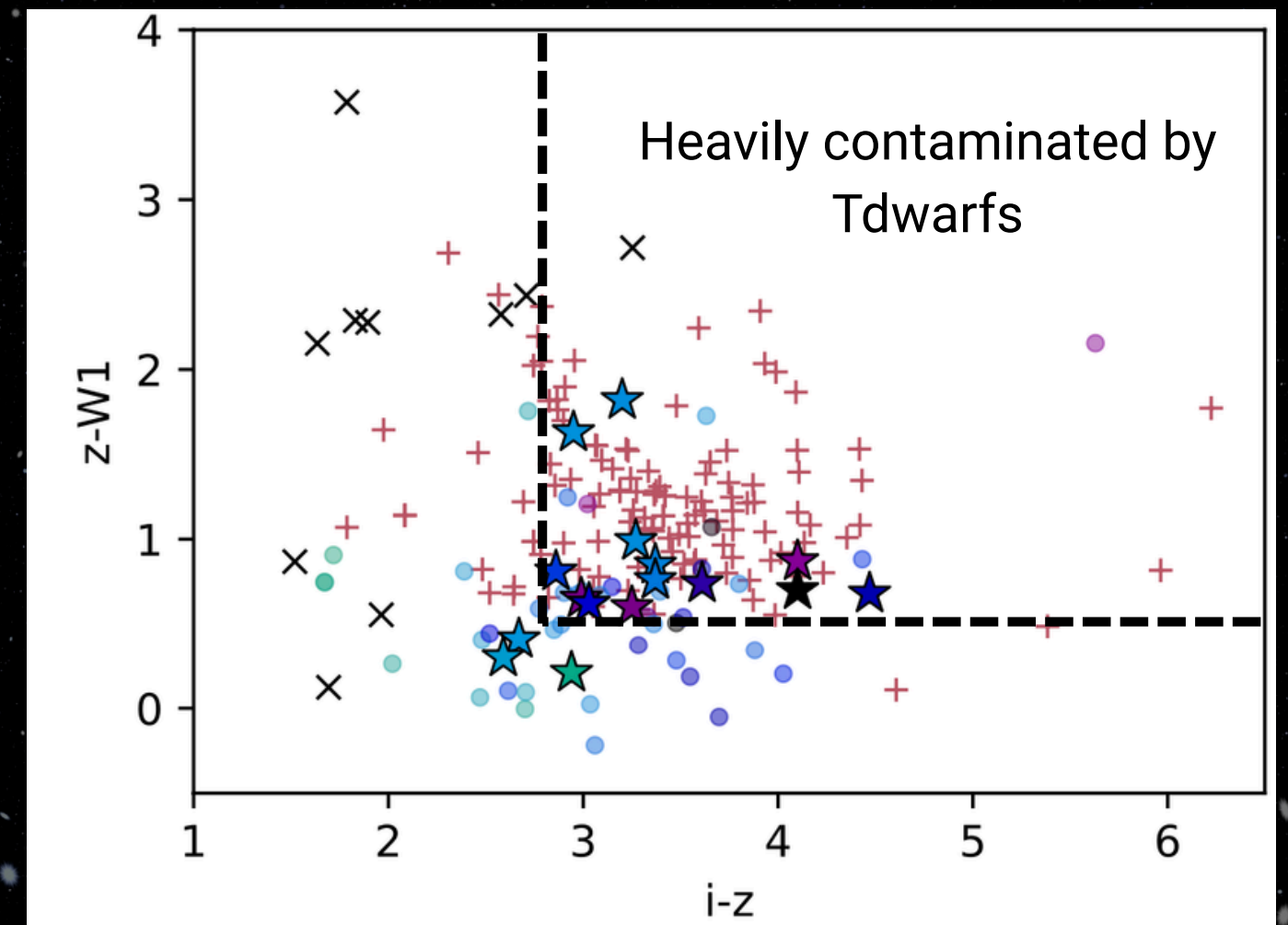
Individual and hybrid AGNs (Salvato et al. 2022)



Selection pipeline with simCLR



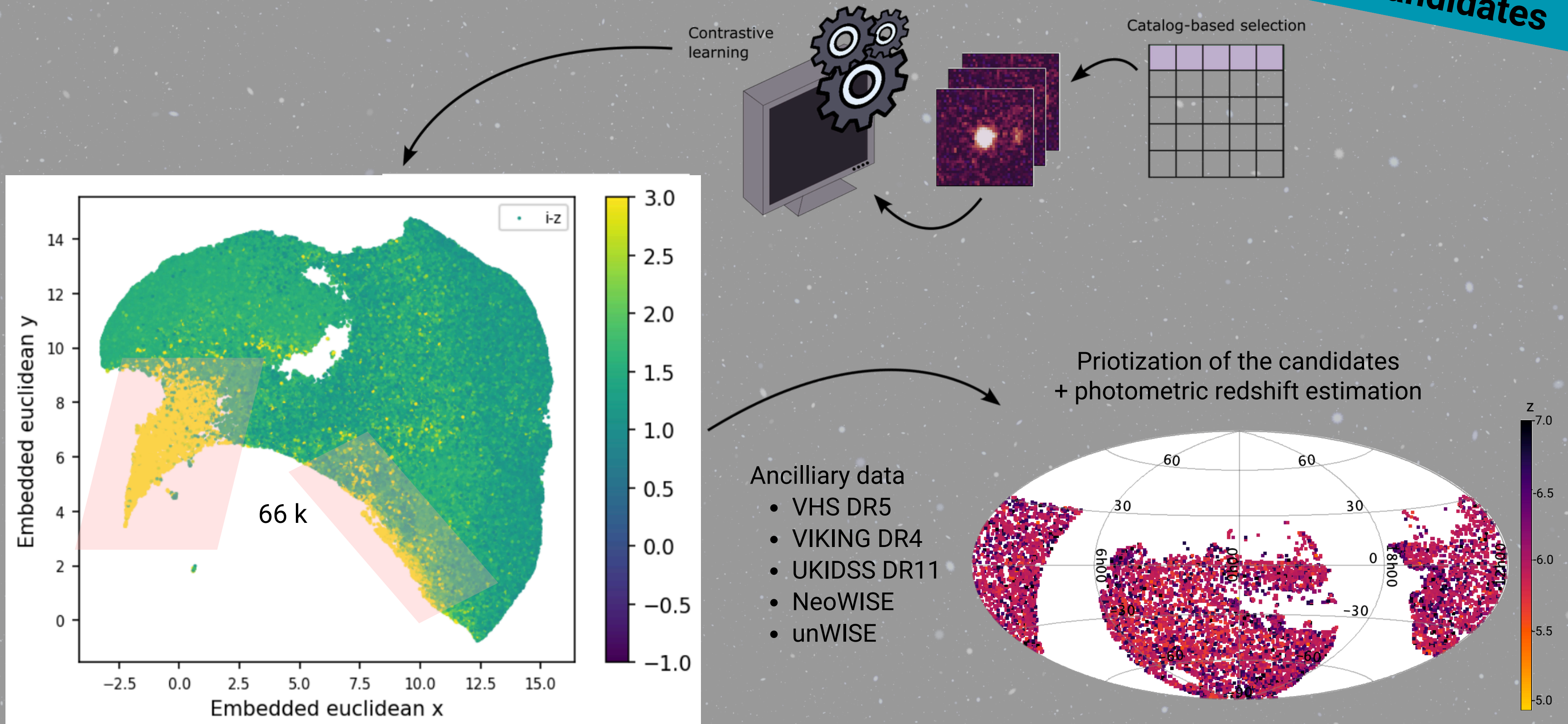
Interesting features of the new quasars



Martinez-Ramirez et al.. (2026)

The high-z subsurvey in ChANGES: CL + SED fitting prioritization

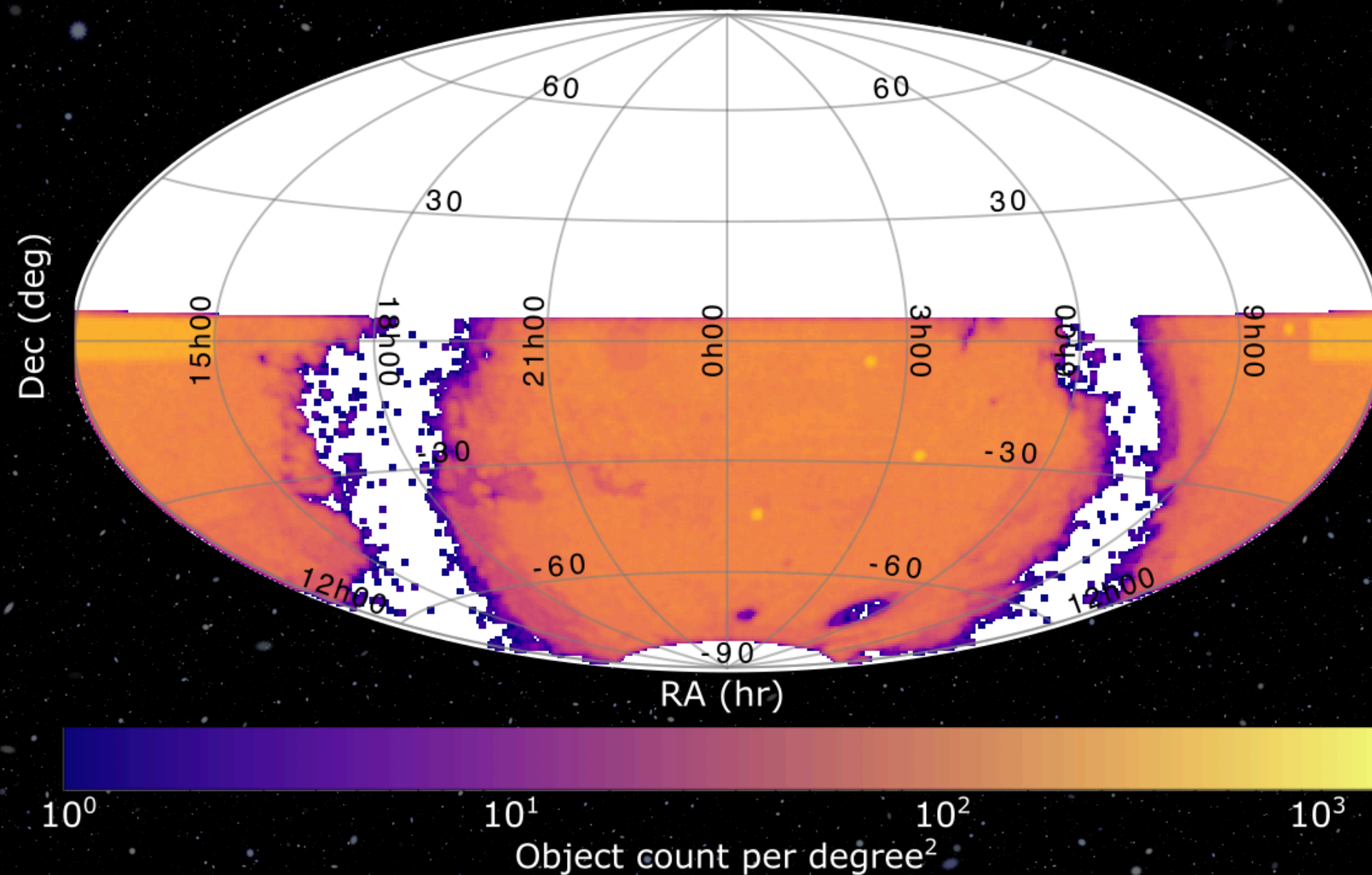
8k candidates



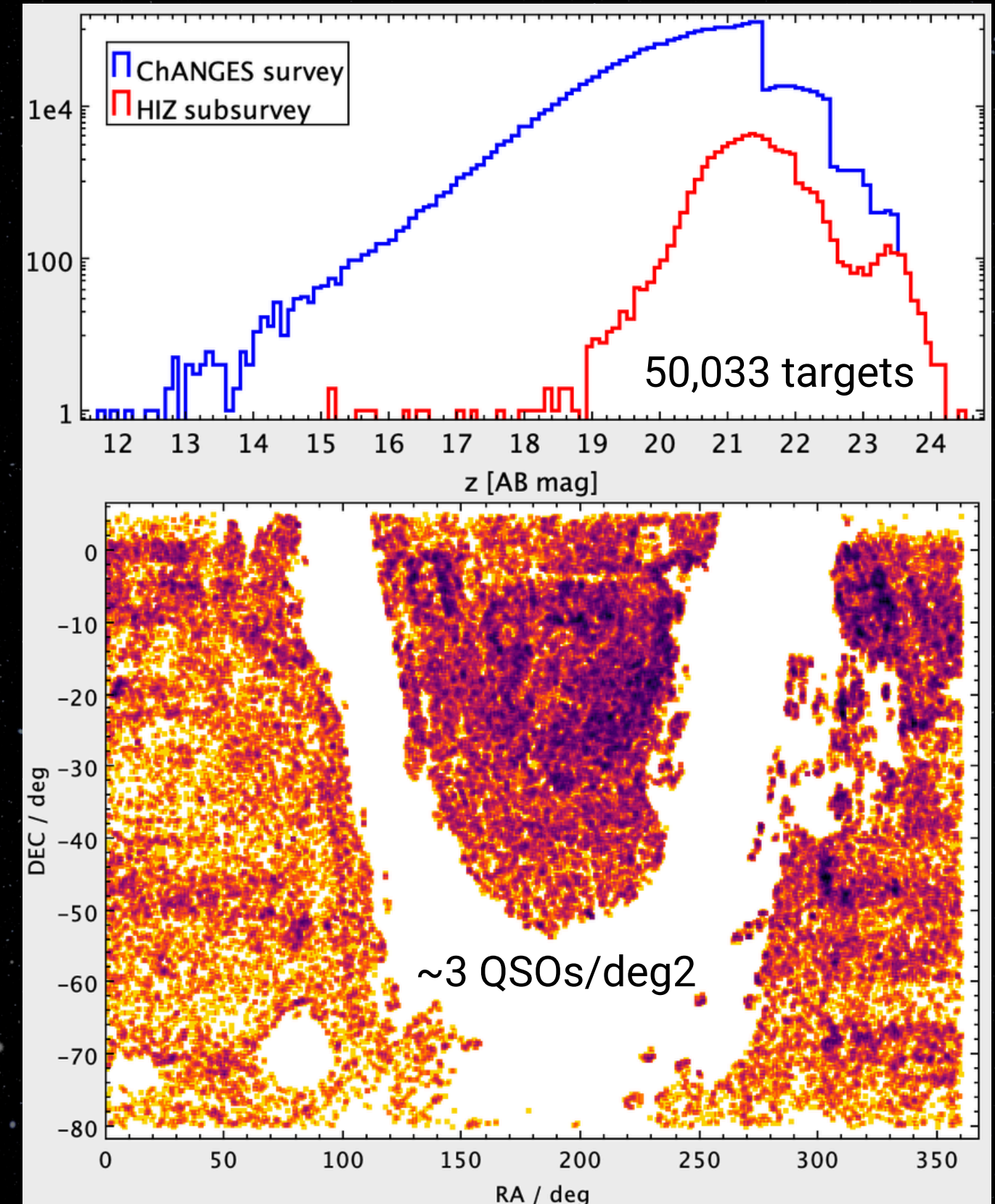
The high-z subsurvey in ChANGES

ChANGES (2268206 targets)

- Exclude low-latitude sky ($AV < 0.5$)
- ~ 200 AGN/deg² all-sky
- ~ 500 AGN/deg² WAVES

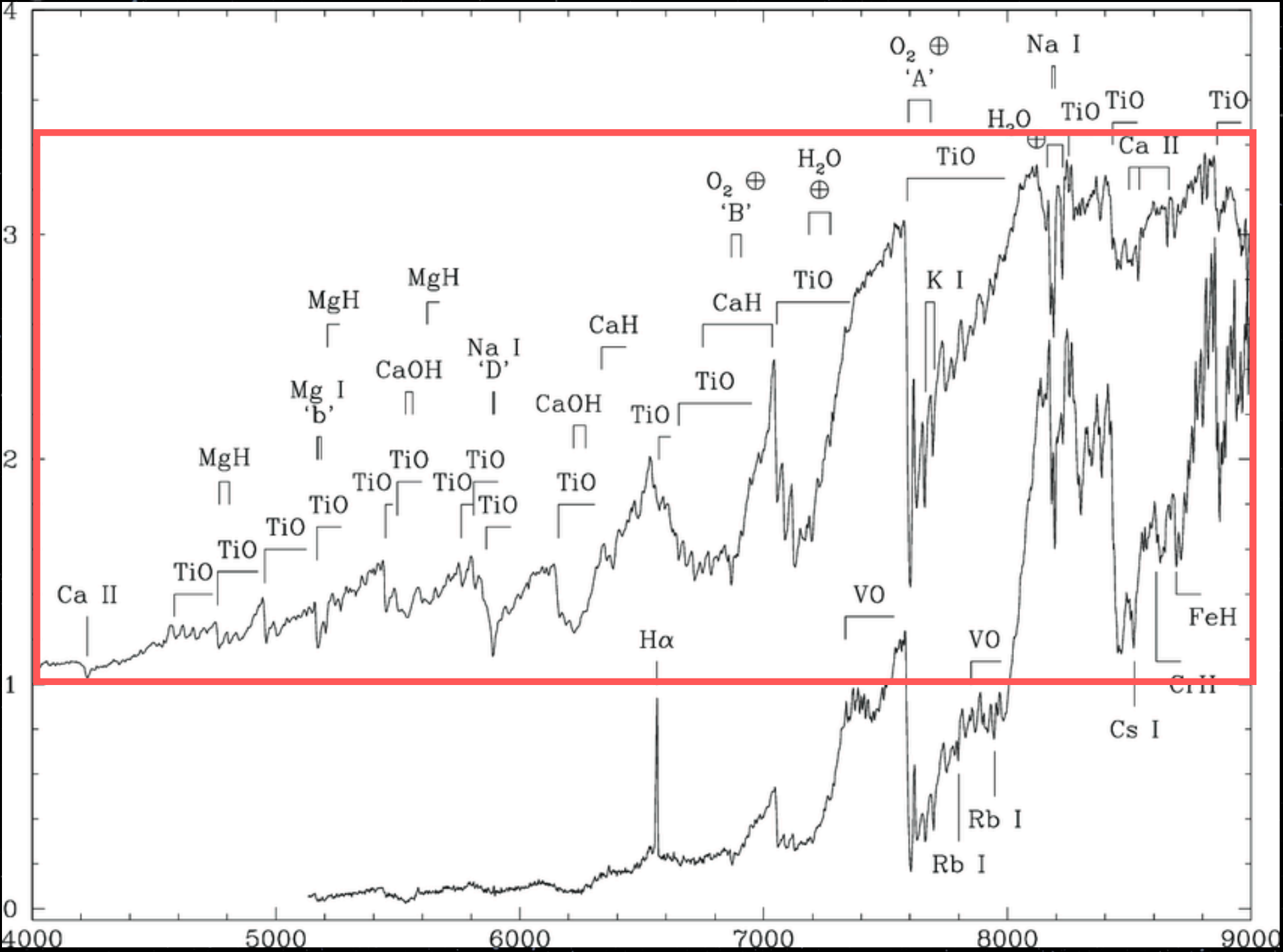


HIZ sub survey



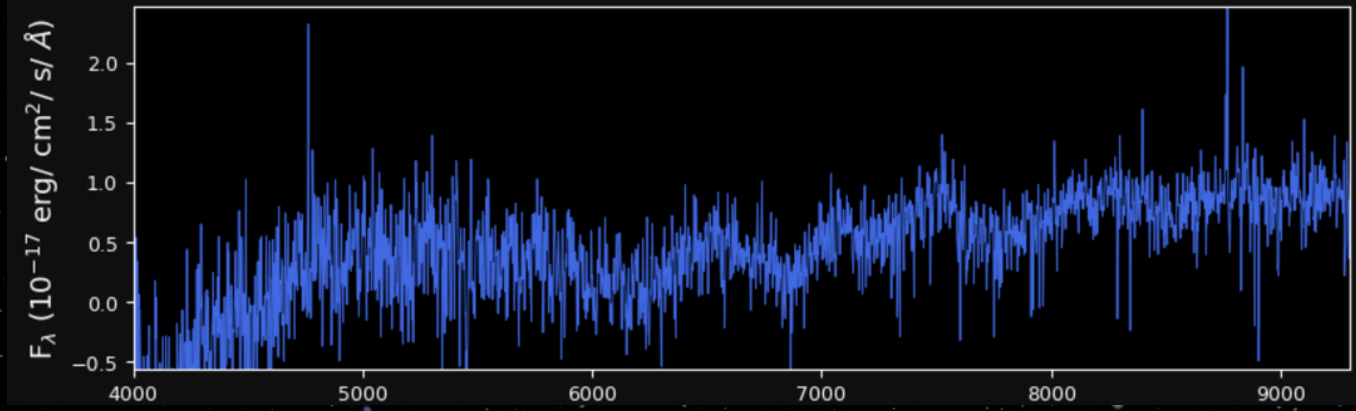
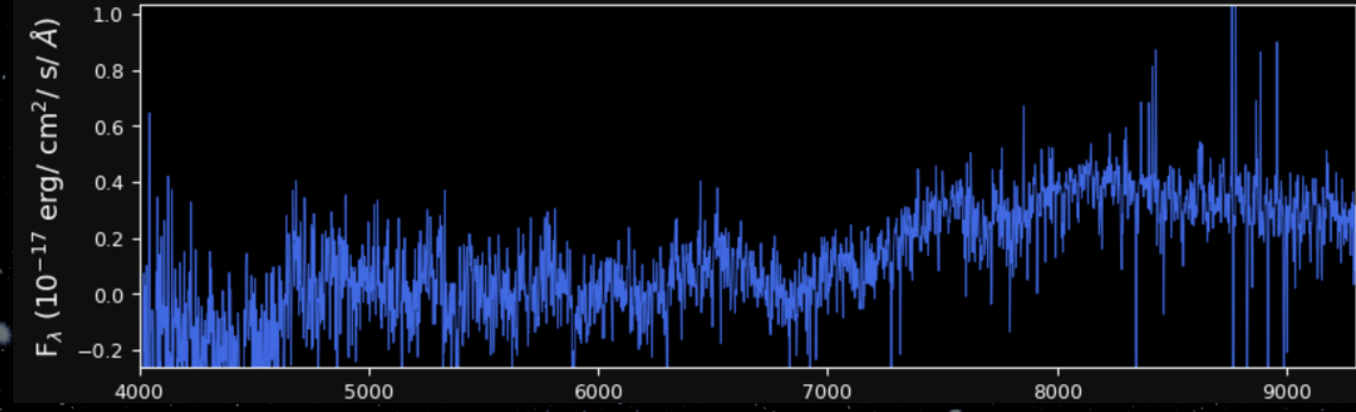
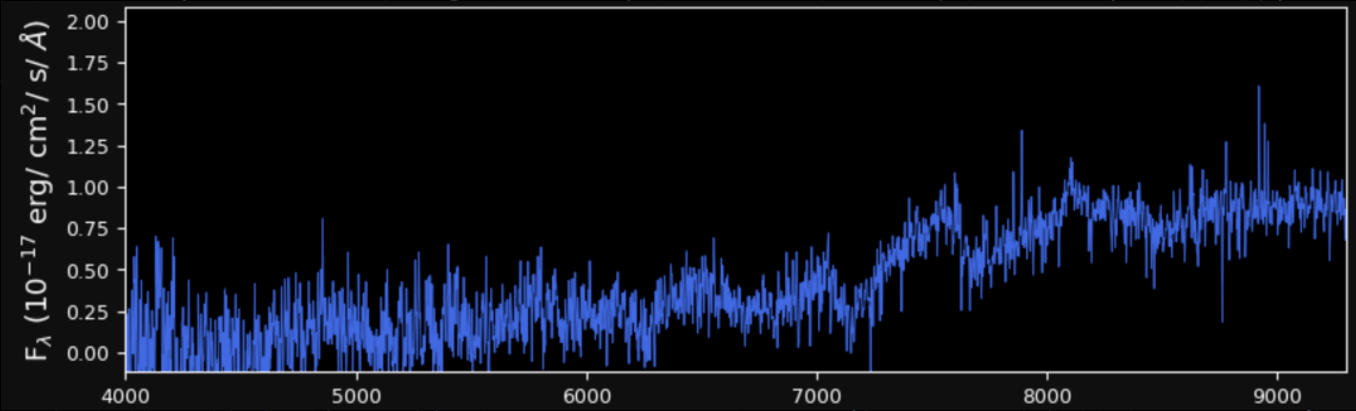
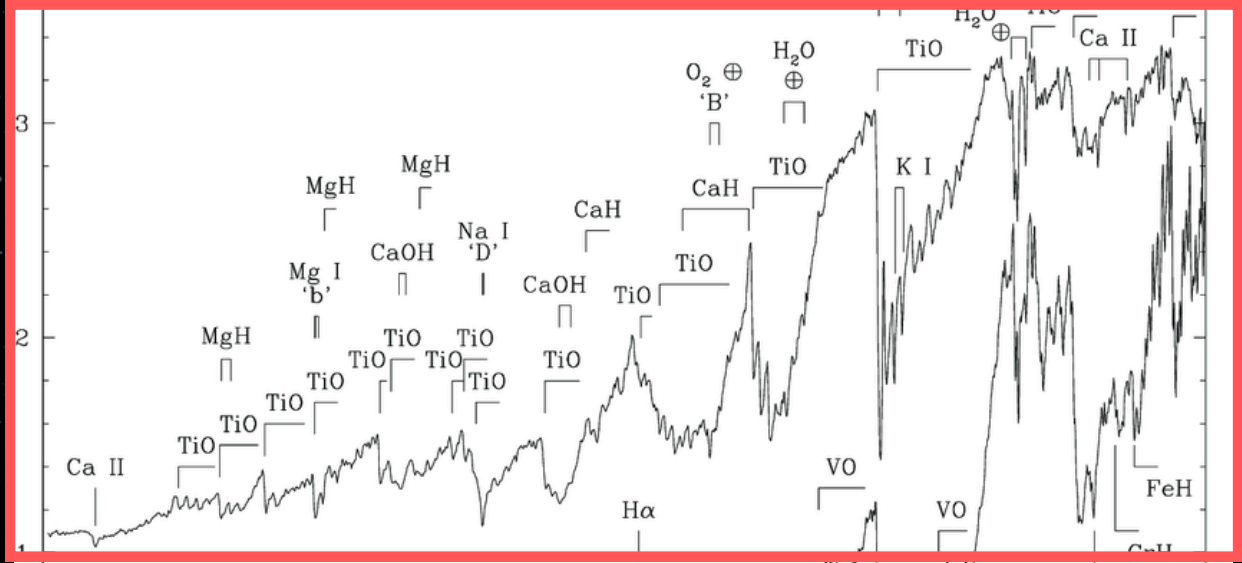
$\sim 75\%$ completeness goal

Ultracool dwarfs discoveries



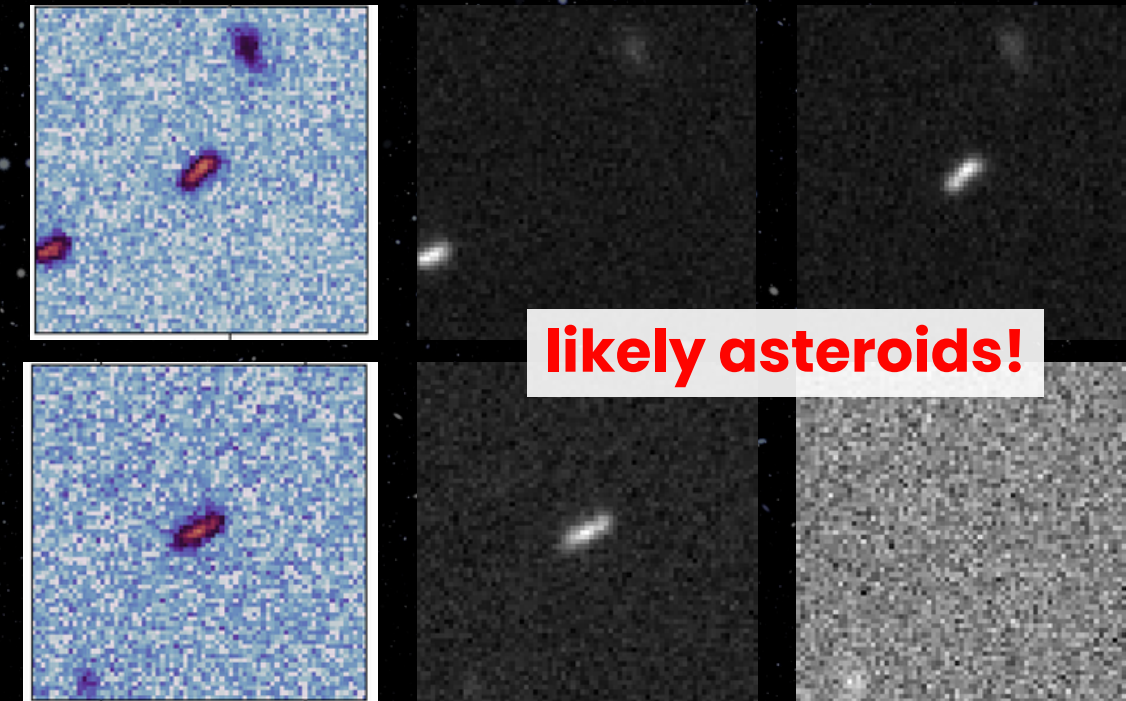
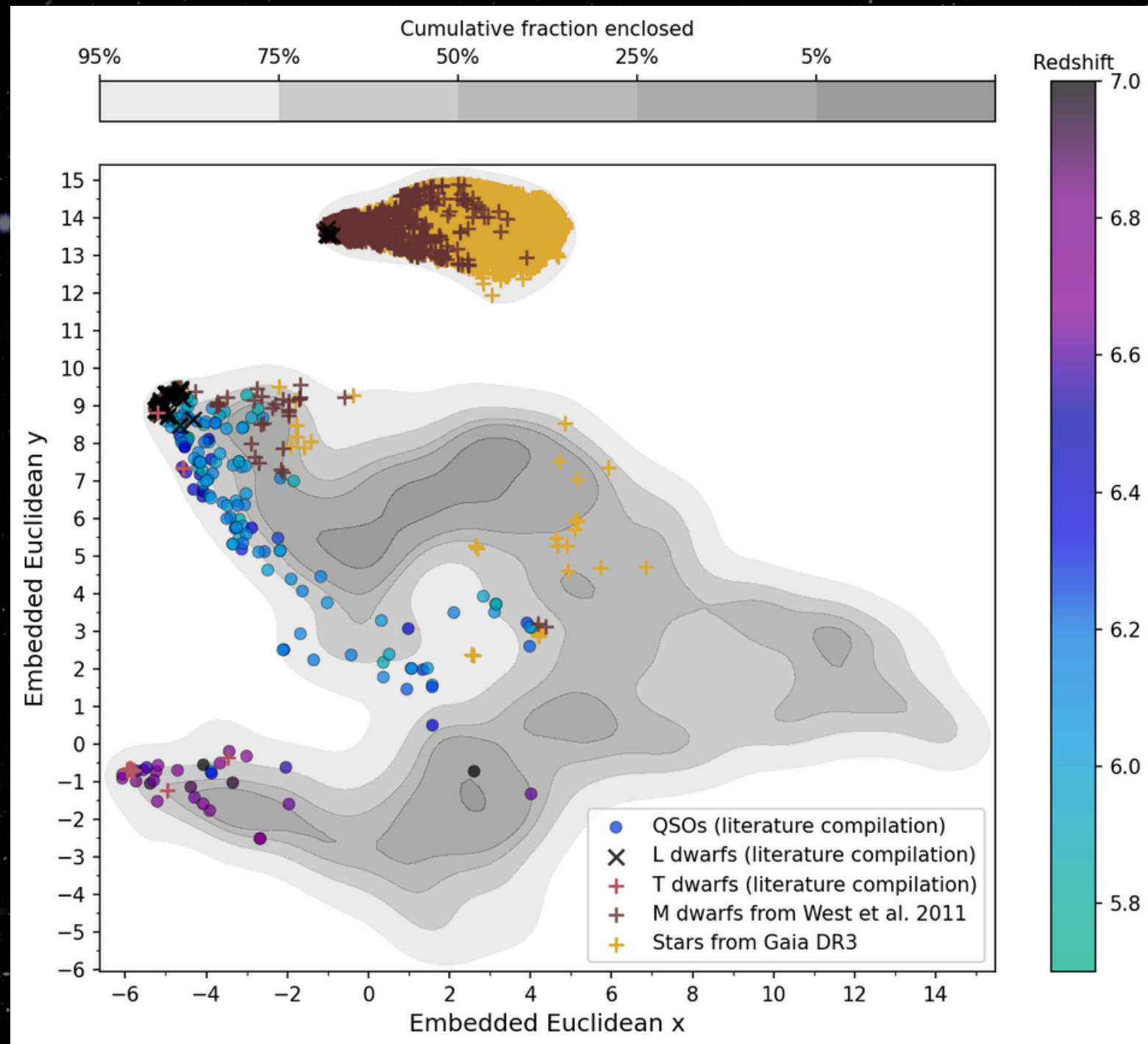
Gray (2009)

M dwarfs

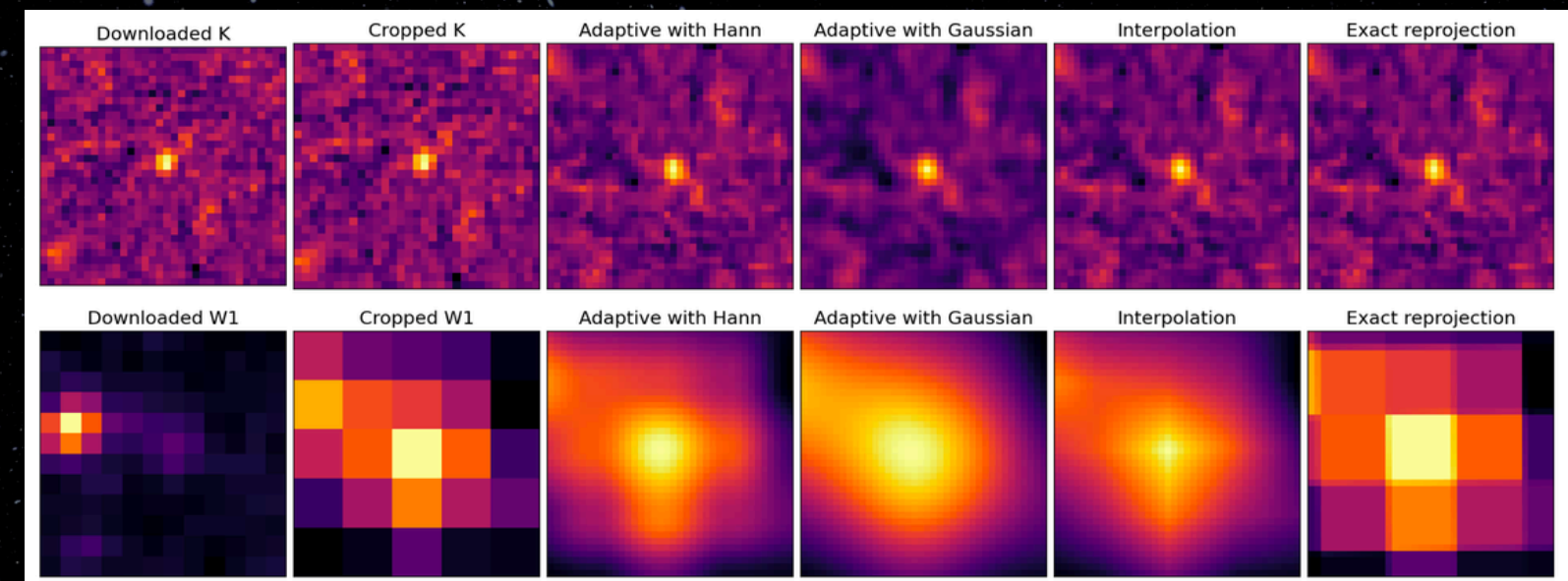


SimCLR applied to other wide-field surveys

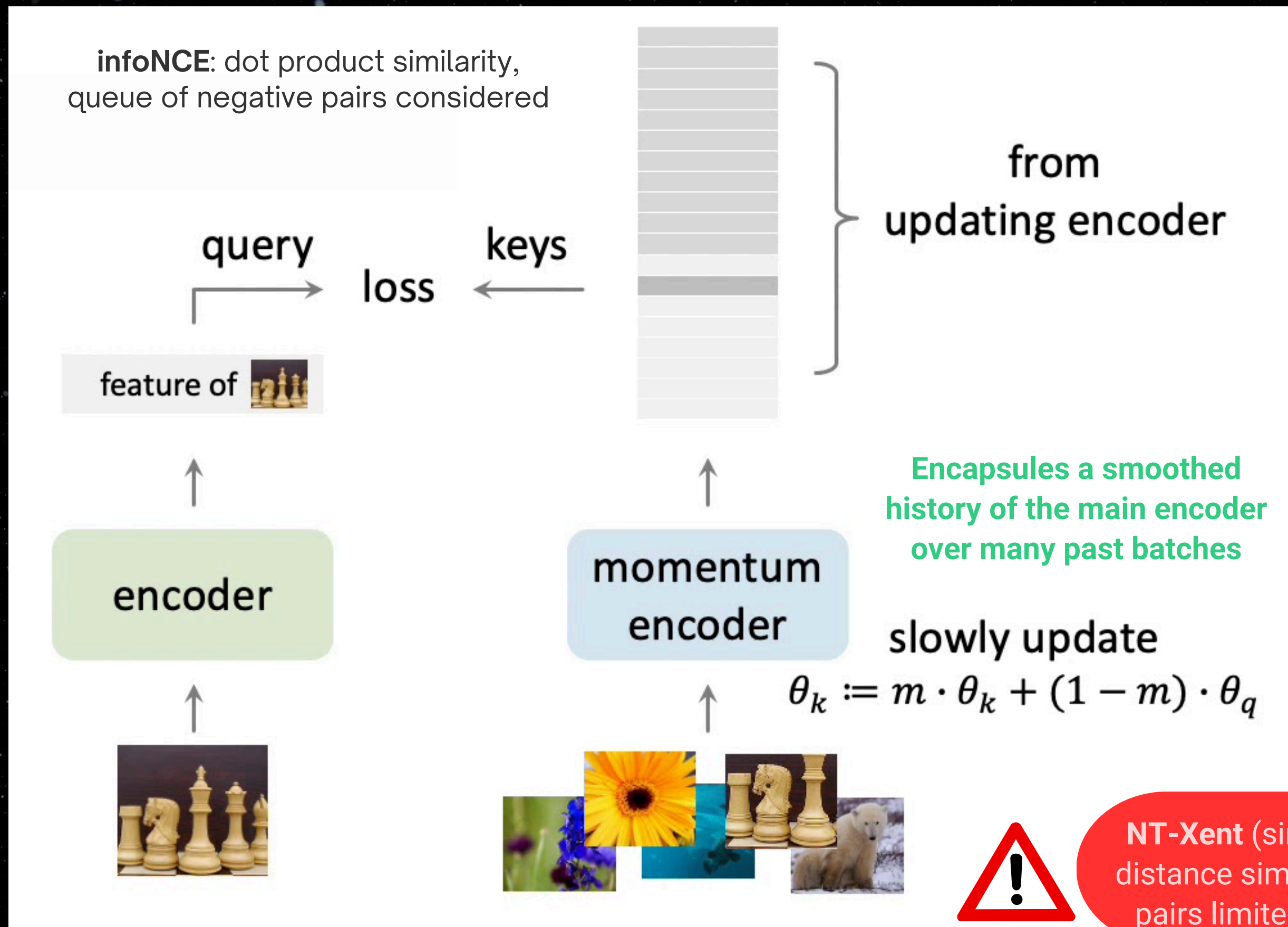
- Preliminary test with wide HSC-SSP DR3



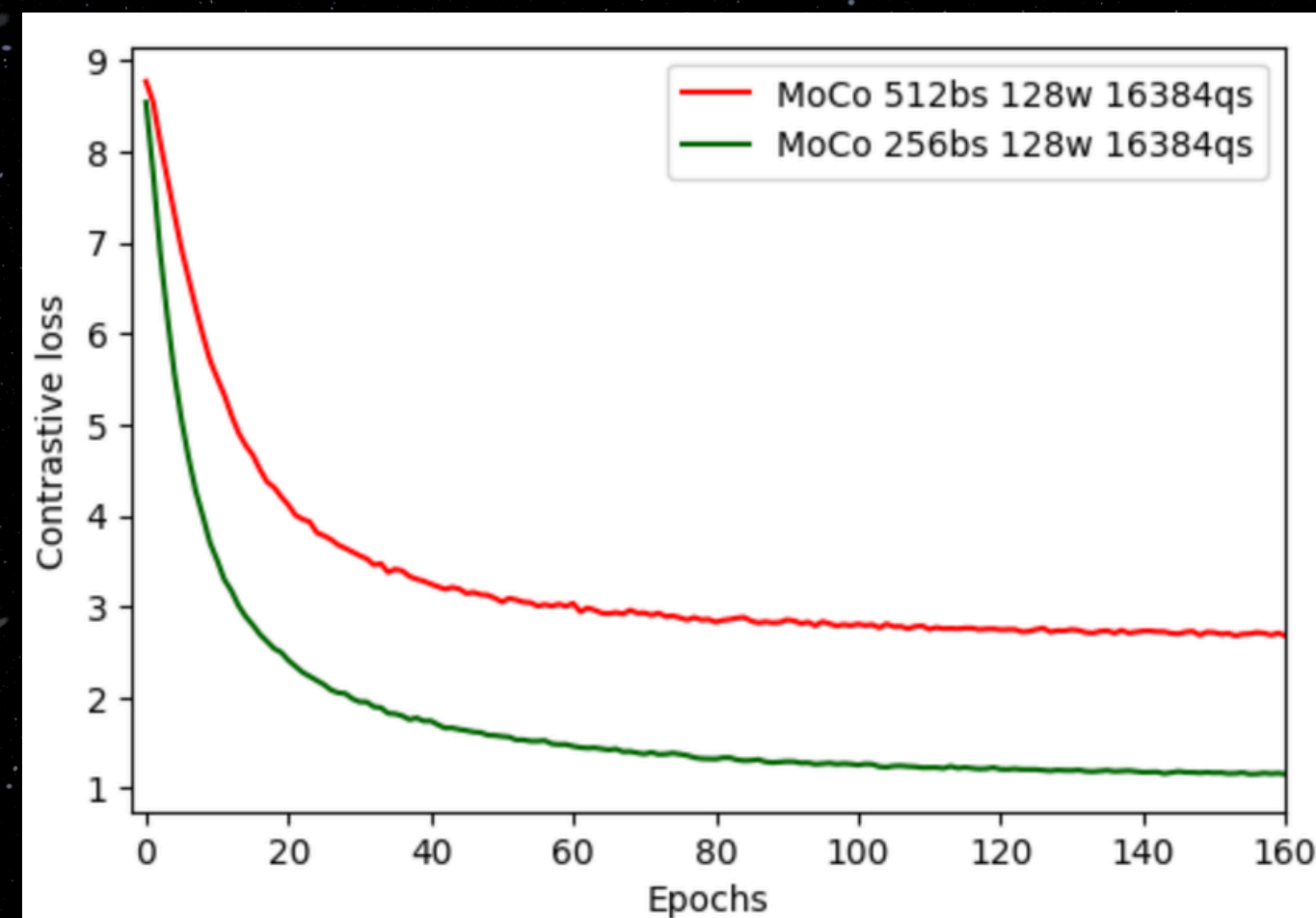
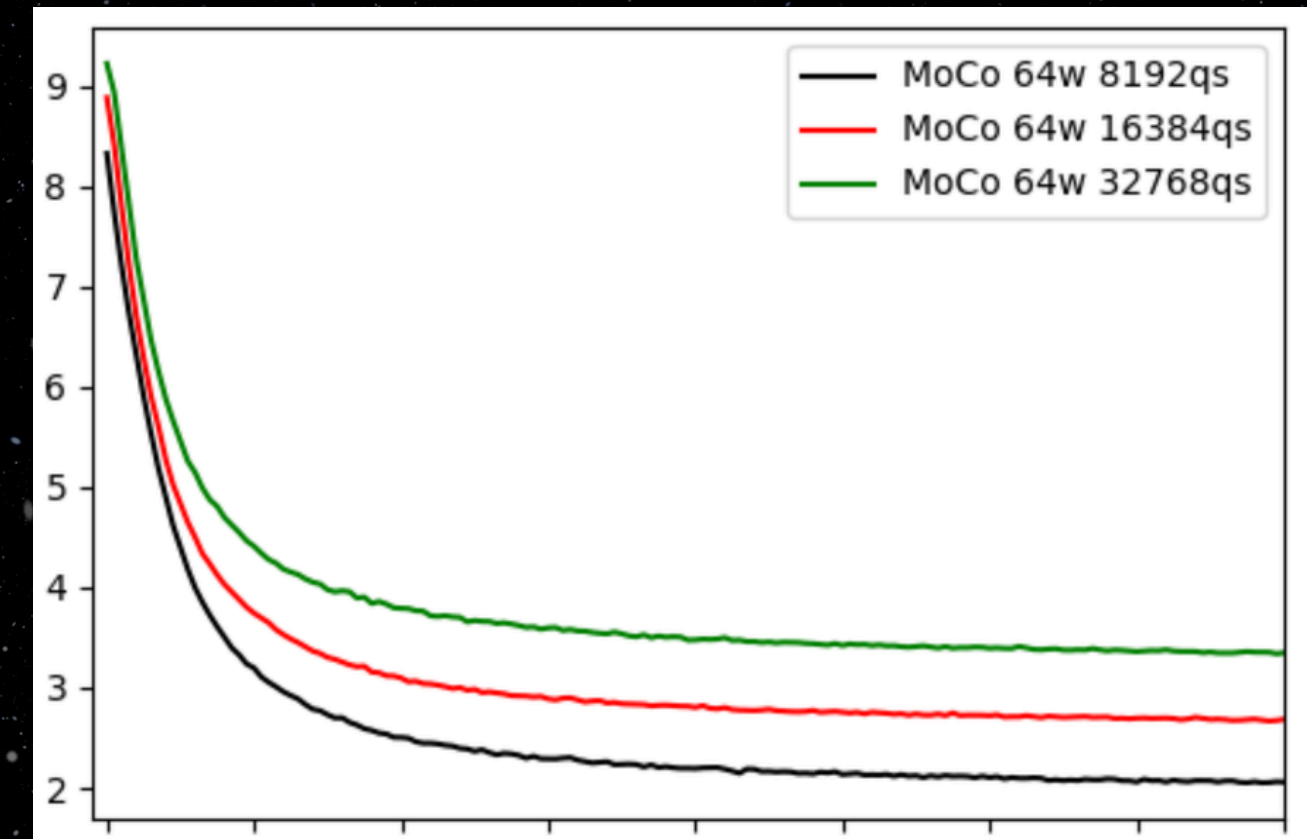
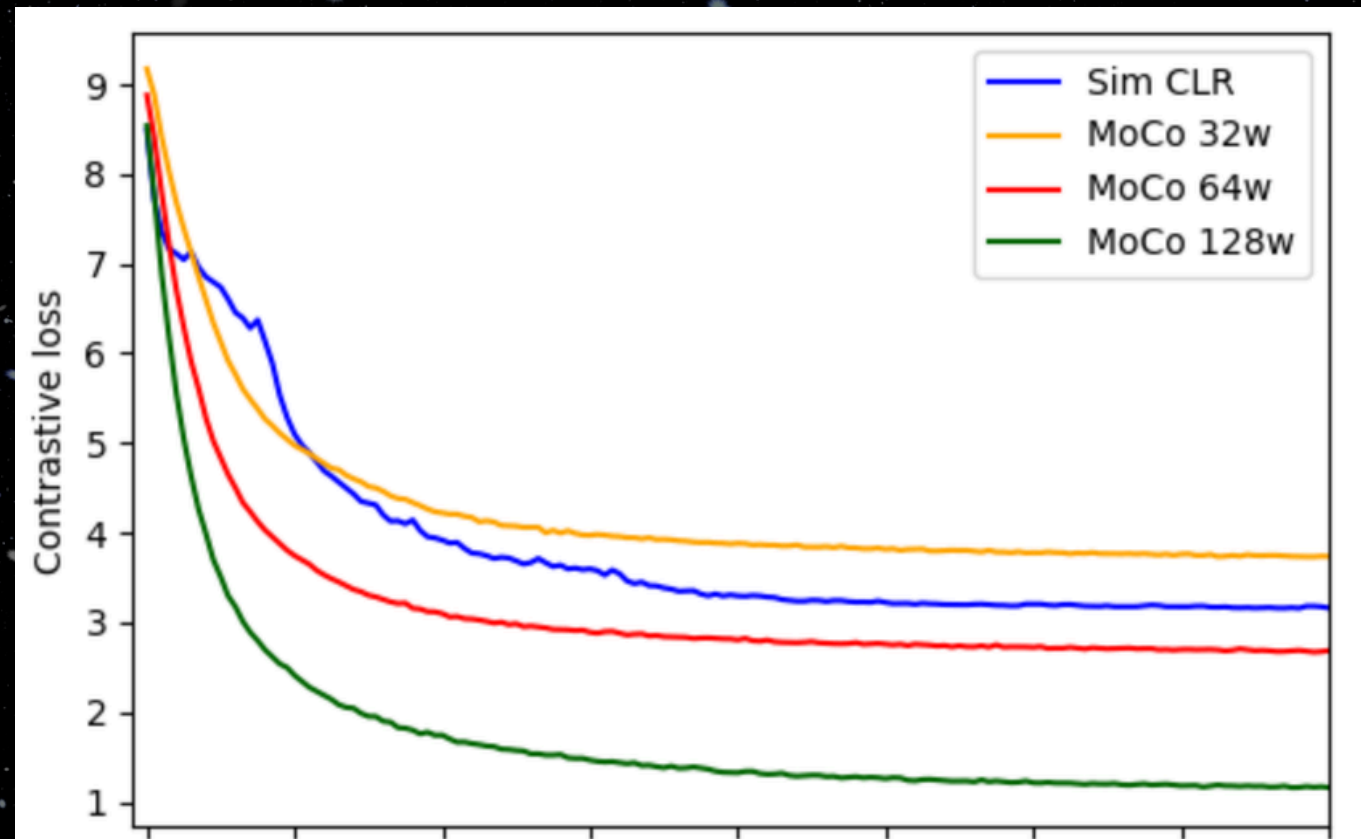
- Discoveries with 4MOST based on LSDR10
- First attempt to include VHS and WISE imaging



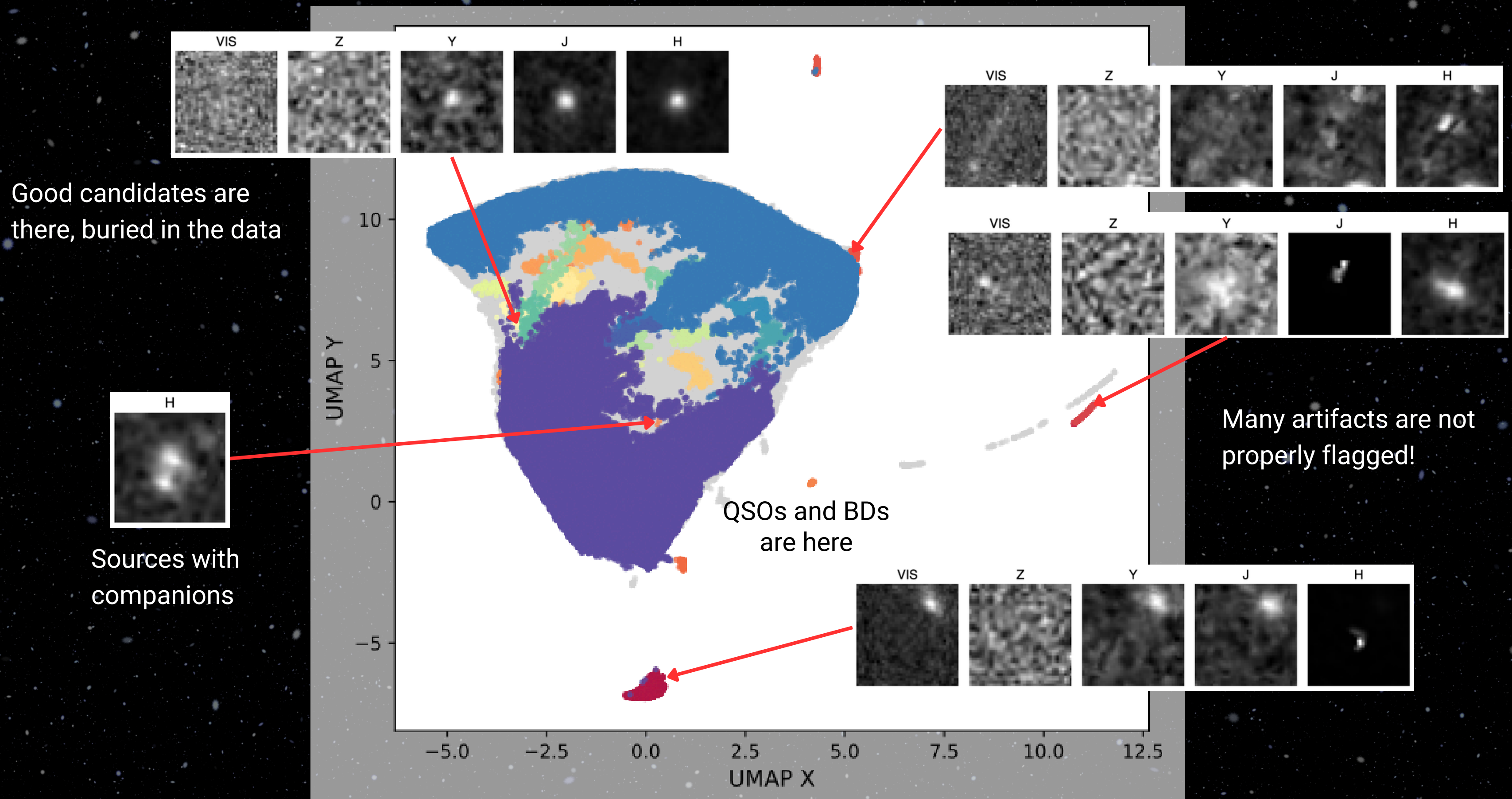
Momentum Contrastive Learning



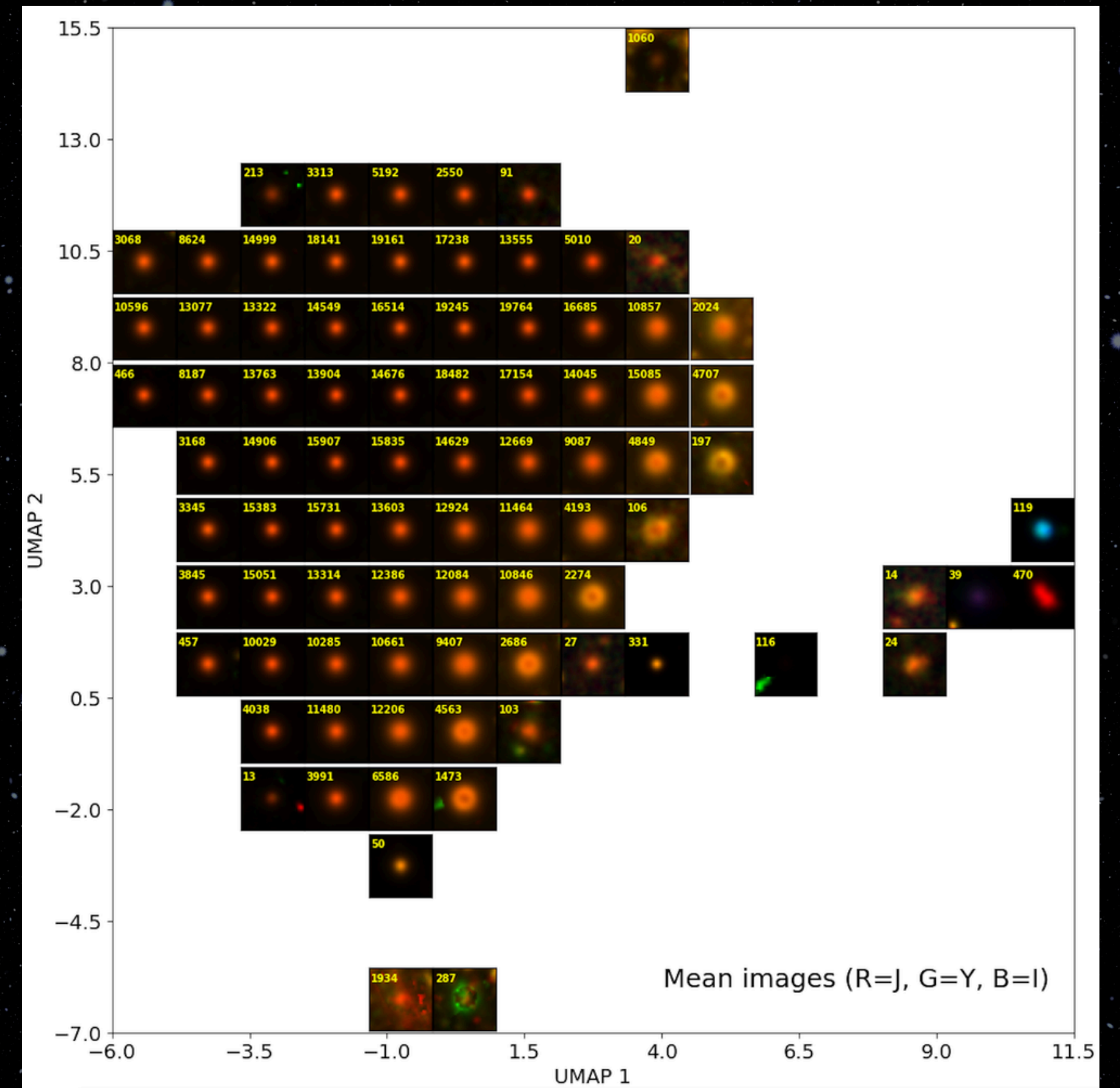
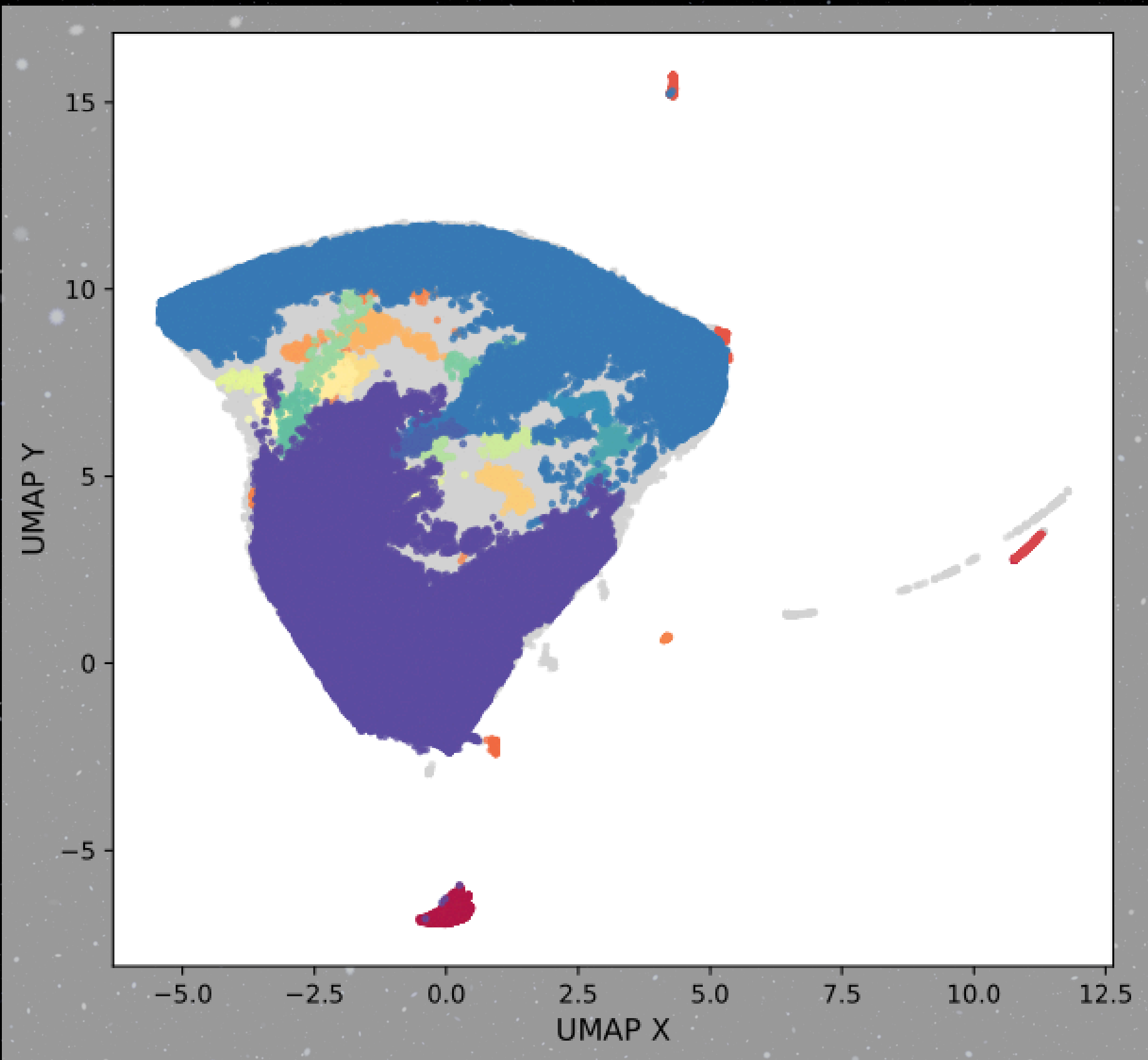
Hyperparameters for MoCo



Preliminary results of MoCo applied to DR1 data

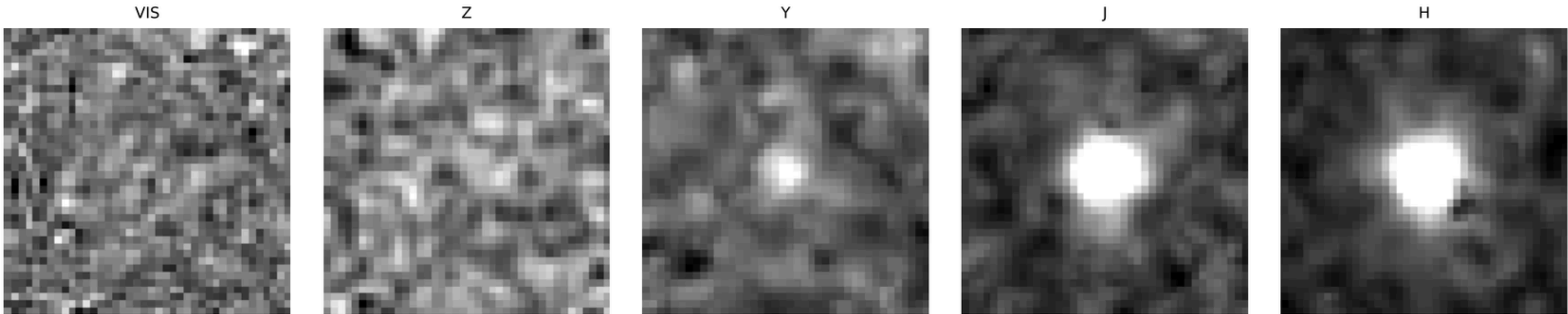


Morphologies in DR1 latent space

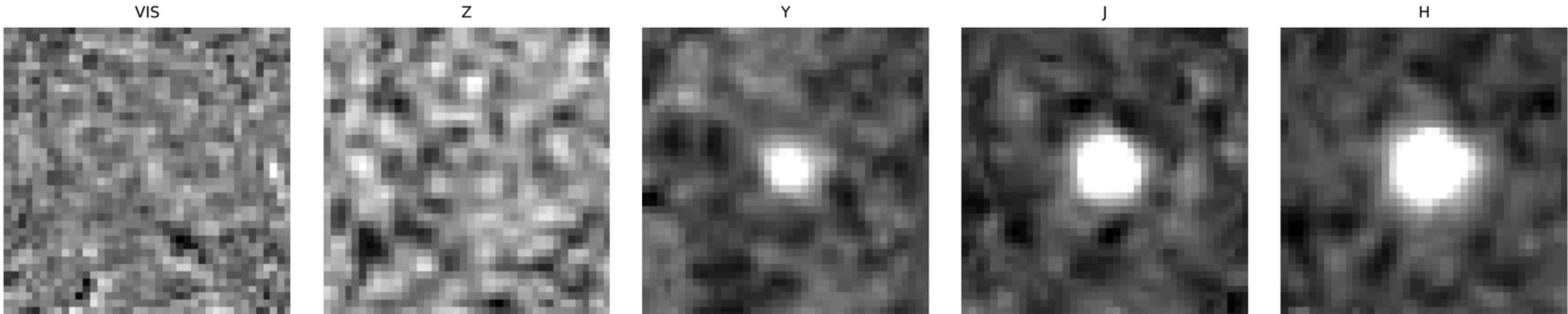


Some $z > 8$ quasar candidates from DPS

ID: 4339_71093 | RA:
151.188632 | DEC:
68.740001
proba: 0.9276 | z_{ml} :
8.6168 | mag_J :
22.8783



ID: 1330_26734 | RA:
200.660017 | DEC:
69.077079
proba: 0.9229 | z_{ml} :
8.3509 | mag_J :
23.0914



ID: 4322_18126 | RA:
161.395789 | DEC:
73.206787
proba: 0.8339 | z_{ml} :
8.6509 | mag_J :
22.9990

