What Comparing HST and JWST Images Reveals About Stellar Clusters Without Labels

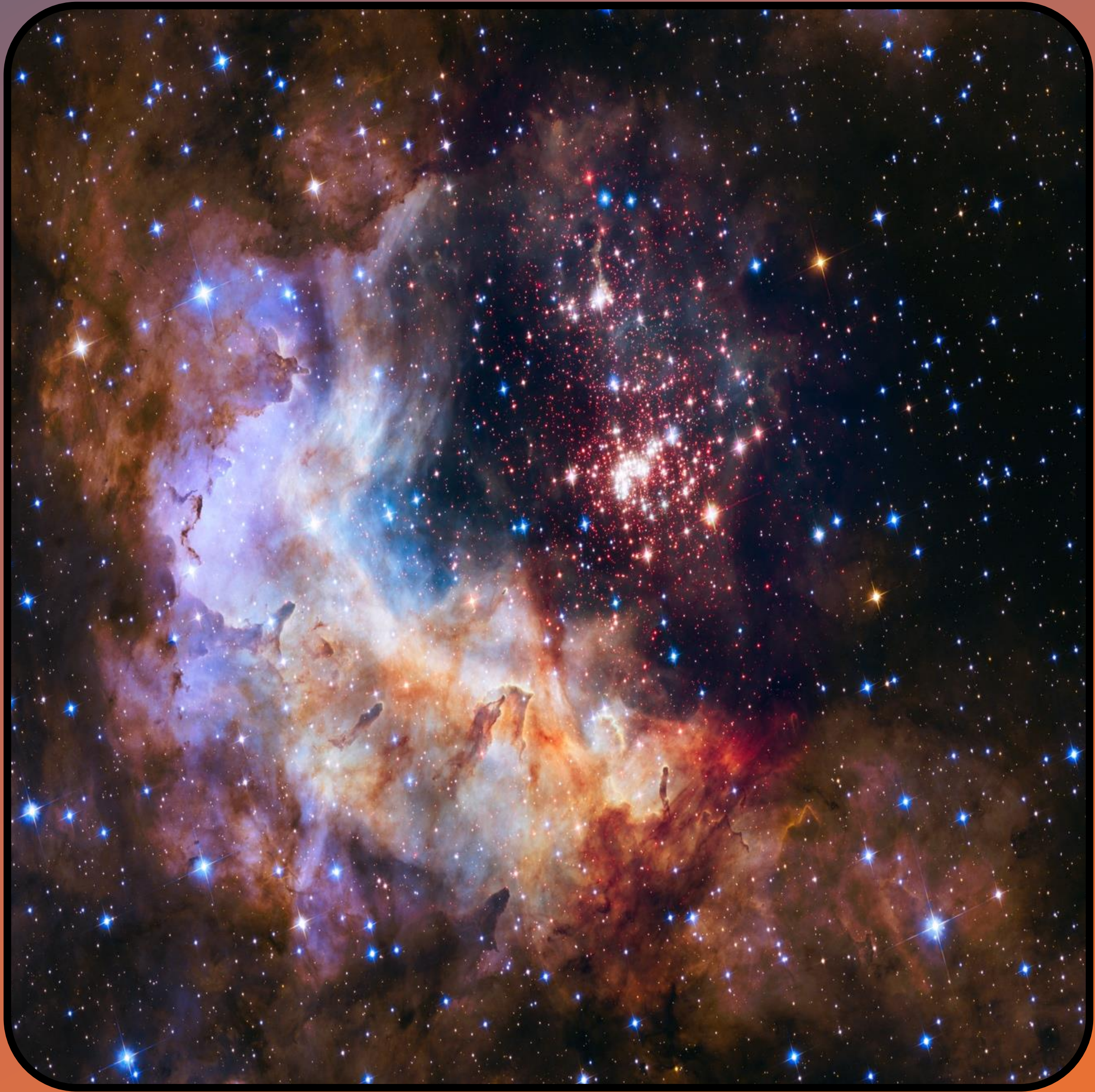
Conference ML4Astro 3

Nils Candebat

Institut für Theoretische Astrophysik
Zentrum für Astronomie der Universität Heidelberg (ZAH-ITA)

Star Formation group [Machine Learning] :
Ralf Klessen, Simon Glover, Phillipp Girichidis, Victor Ksoll, Daniel Walter, Jorge Bastidas, Shuyu Tan

nils.candebat@uni-heidelberg.de

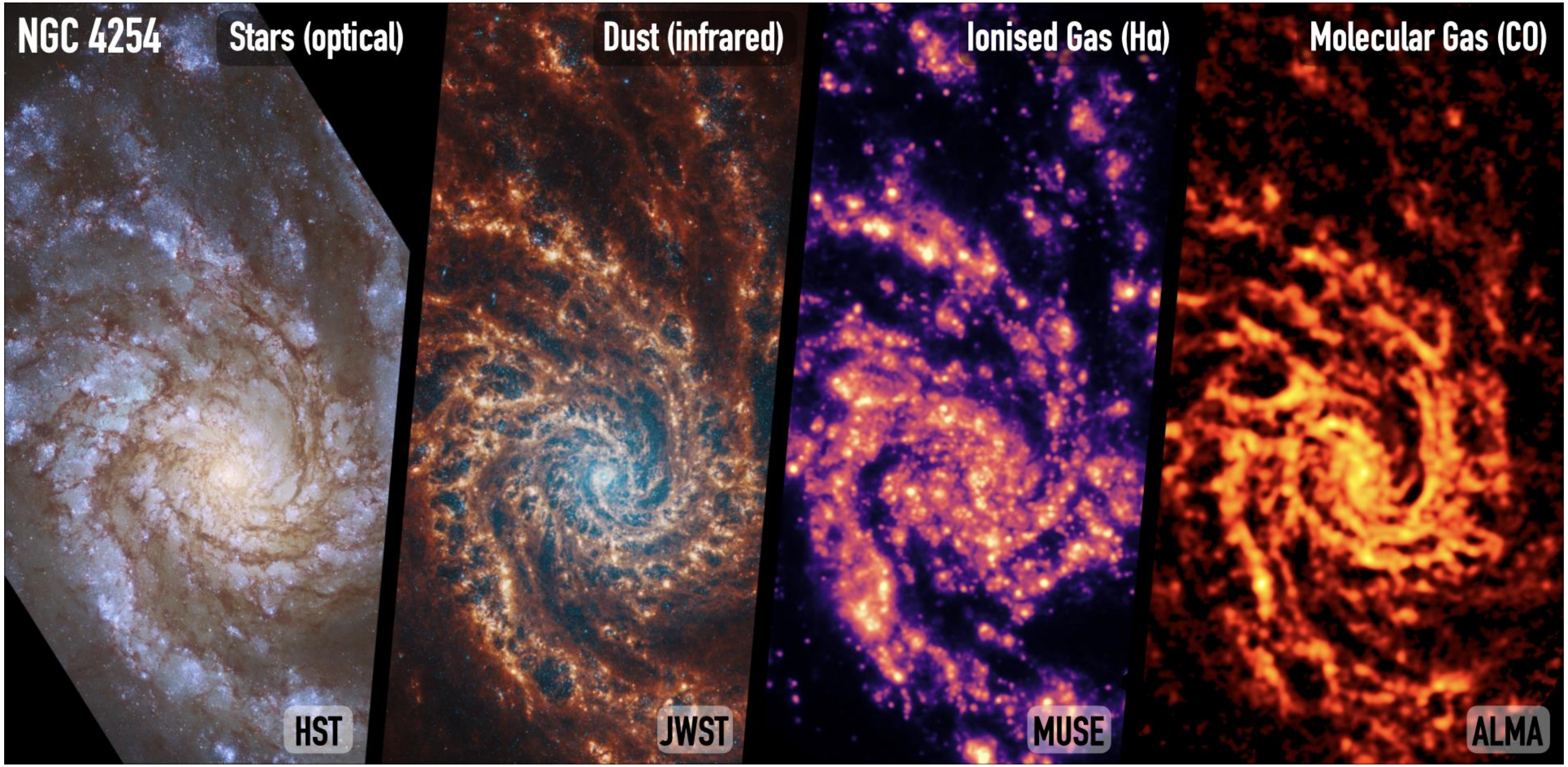


Westerlund 2 (Hubble), NASA

31 August 2026

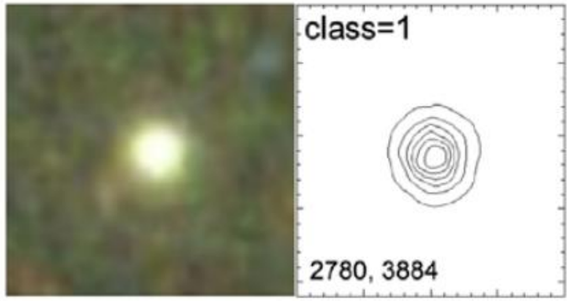
PHANGS

- 38 galaxies observed around a wide range of instruments and wavelength.
- 95, 992 stellar clusters and compact associations (classes 1, 2, 3)
- 35,071 stellar clusters (classes 1 and 2)
- Two cycles of observations (cycle 2107 – cycle 3707)

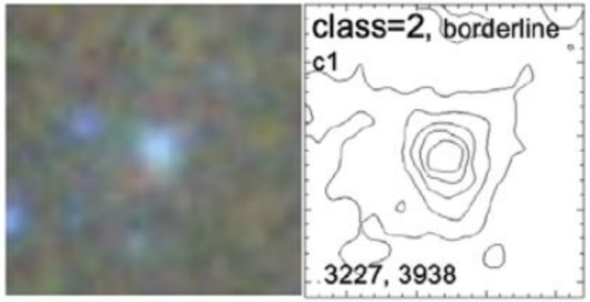


Cluster in nearby galaxies:

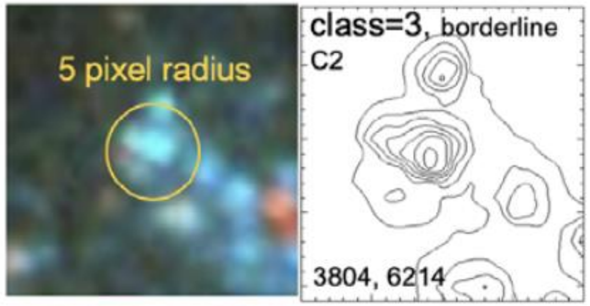
Symmetric 1



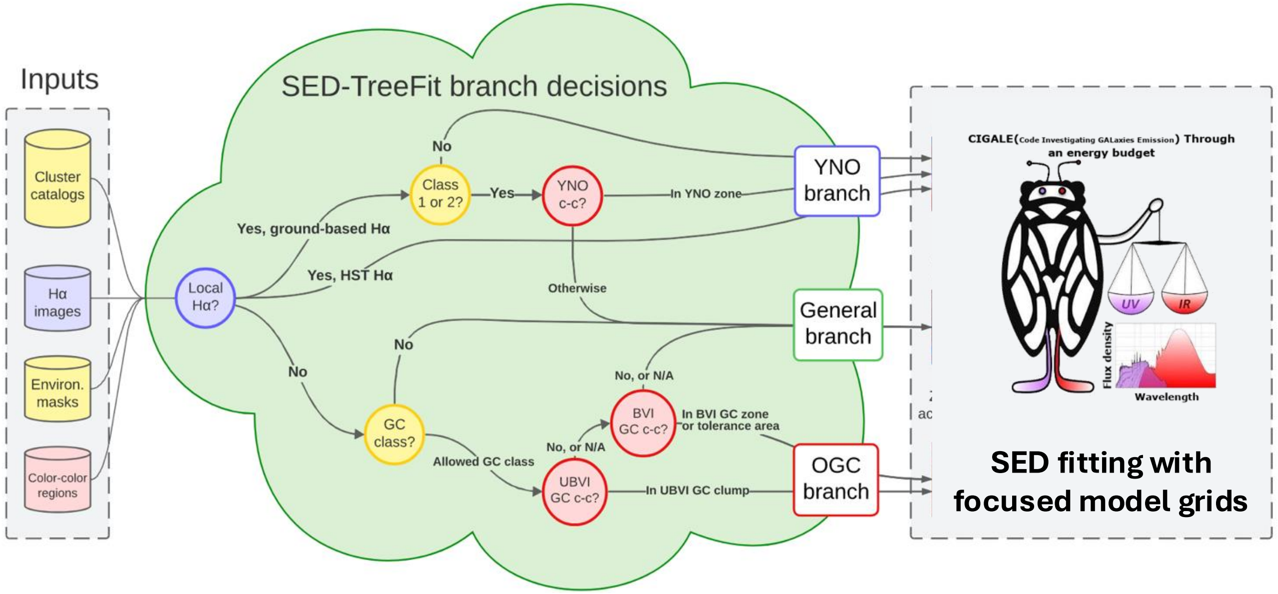
Asymmetric 2



Compact association 3

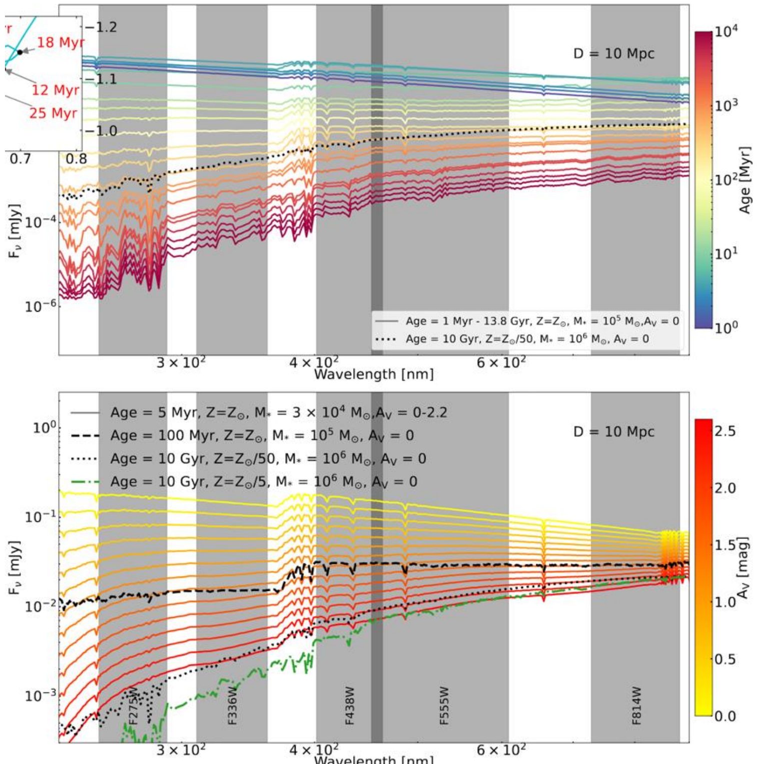


Determination of age



Thilker et al. 2025

SoTA age determination algorithm

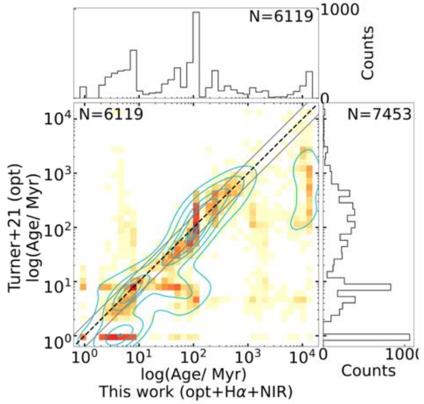


Turner et al. 2021

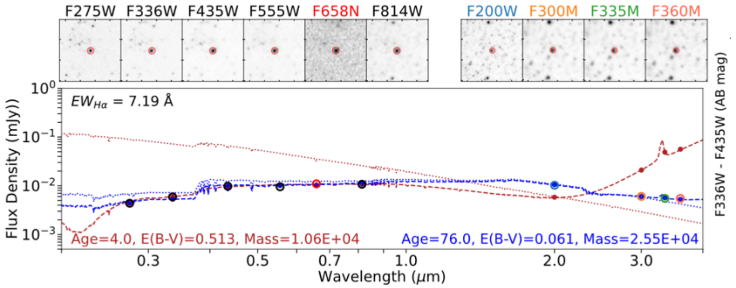
Degeneracies

Problems:

- Current age determination is messy with arbitrary choices
- Age problem is multiple degenerate
- No good agreements between methods



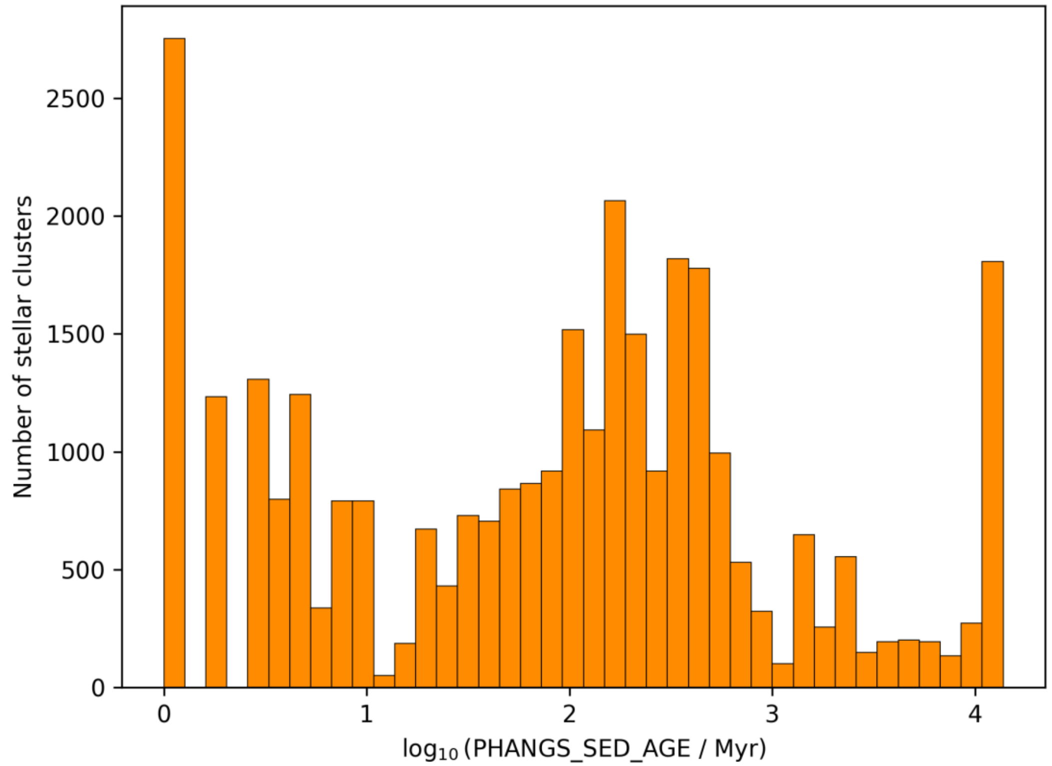
Henny et al. 2025



Henny et al. 2025

Filter dependence

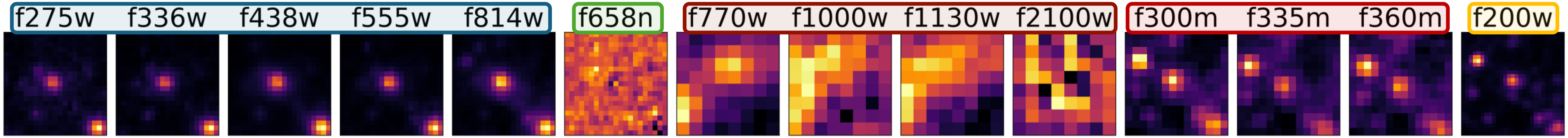
Dataset



Instruments	pixel size (arcsec)	Cycle 2107	Cycle 3707
HST	0.04	**f275w, f336w, f438w, f555w, f814w**	
HST Hα	0.04	(F658n), (f657n)	f657w
JWST NIRCam SW	0.031	**f200w**	**(f200w)**
JWST NIRCam LW	0.063	**f300m, f335m,** f360m	**f300m, f335m**
JWST MIRI	0.11	**f770w**, f1000w, f1130w, **f2100w**	**f770w, f2100w**

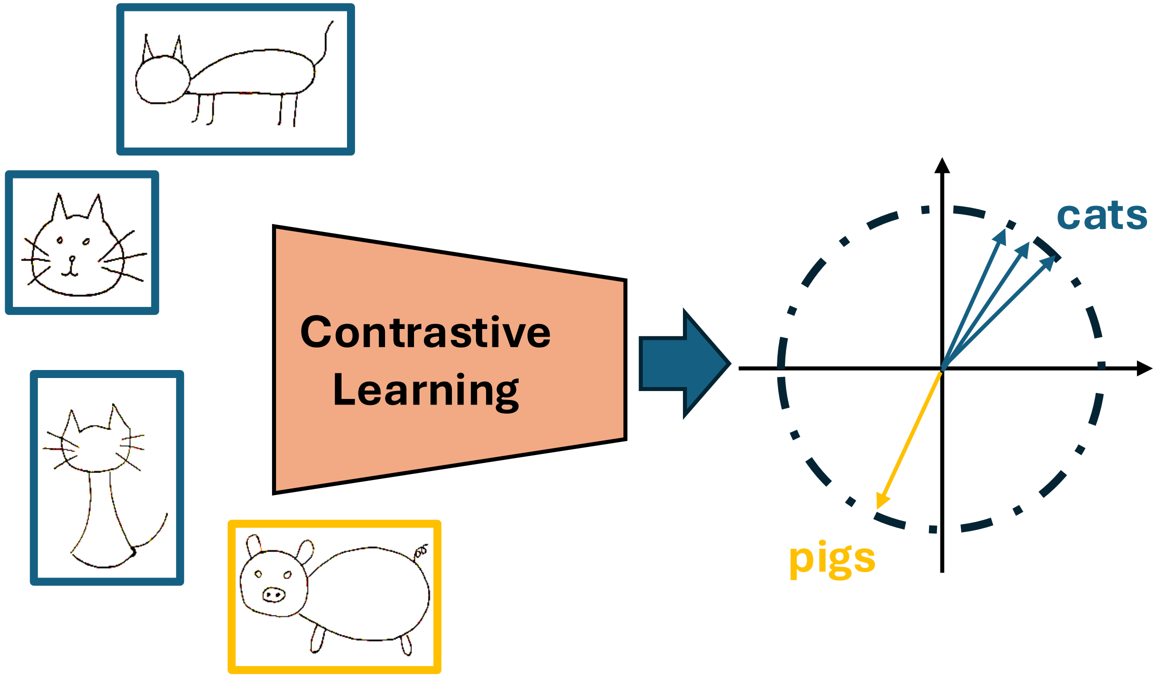
Age distribution
Poor label quality

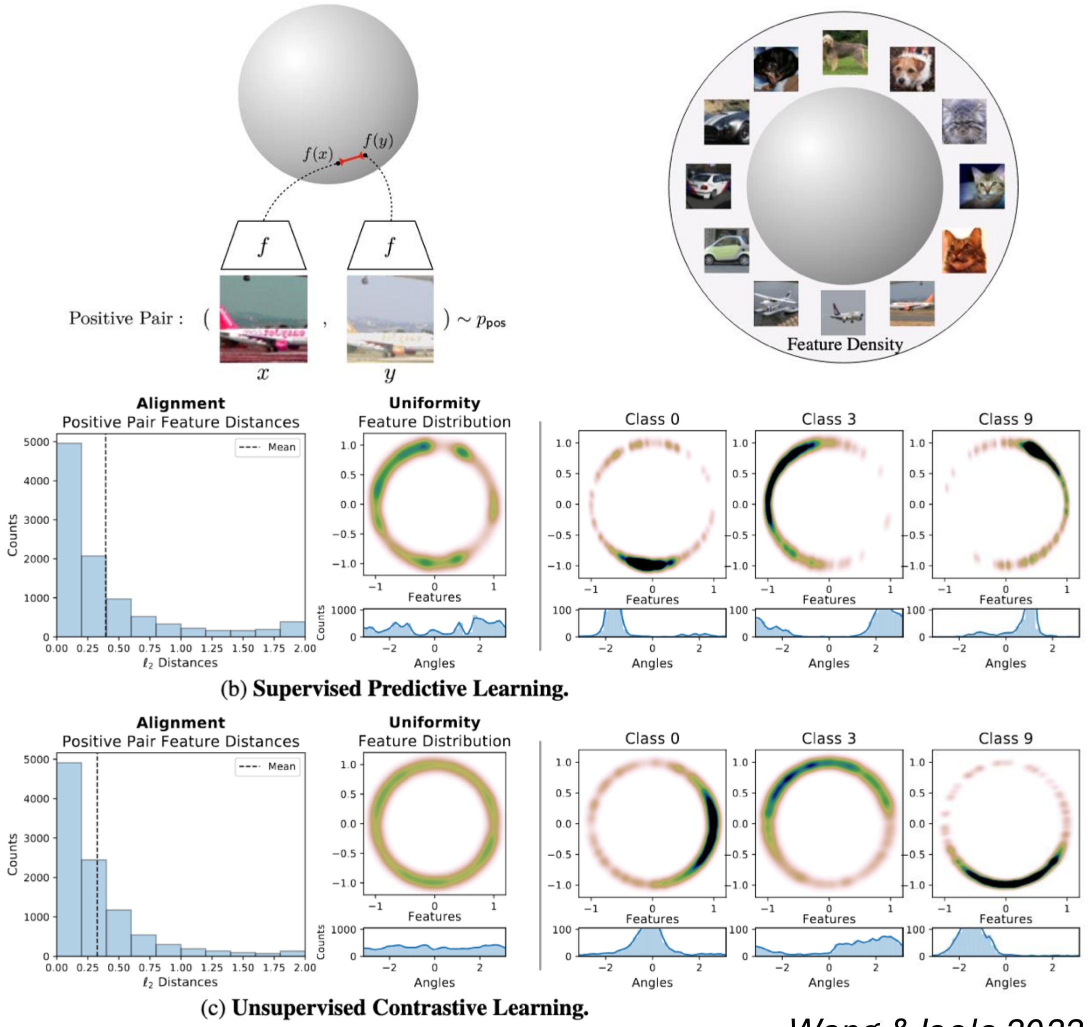
Age distribution
Missing Modality problem



Unsupervised Contrastive Learning (1/2)

In contrastive methods applied to joint embedding architectures, the output embeddings for a sample and its distorted version are brought close to each other, while other samples and their distortions are pushed away.





Alignment: two samples forming a positive pair should be mapped to nearby features and thus be (mostly) invariant to unneeded noise factors.
Uniformity: feature vectors should be roughly uniformly distributed on the unit hypersphere, preserving as much information of the data as possible.

Wang & Isola 2022

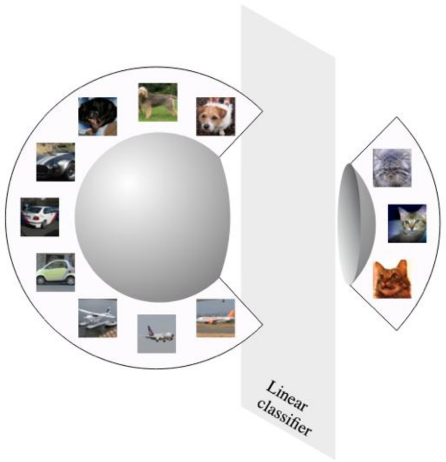
Unsupervised Contrastive Learning (2/2)

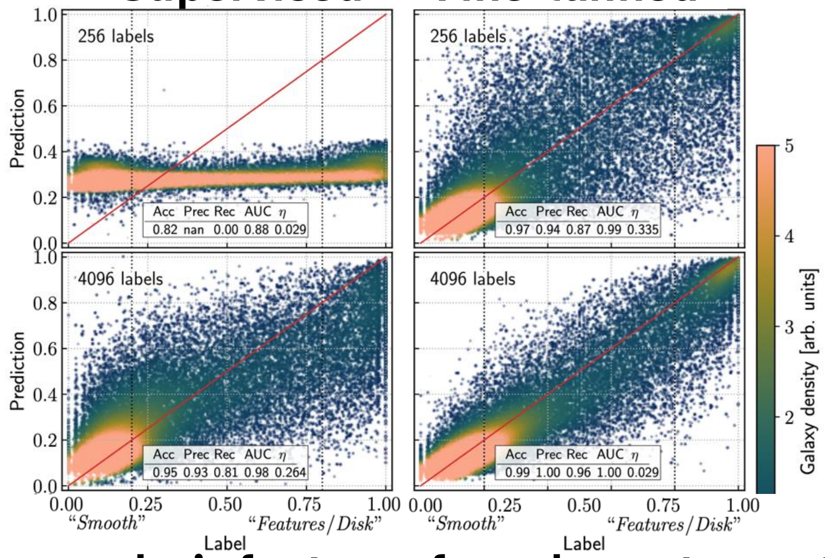
Wang & Isola 2022
Chen et al. 2020
Hayat et al. 2021

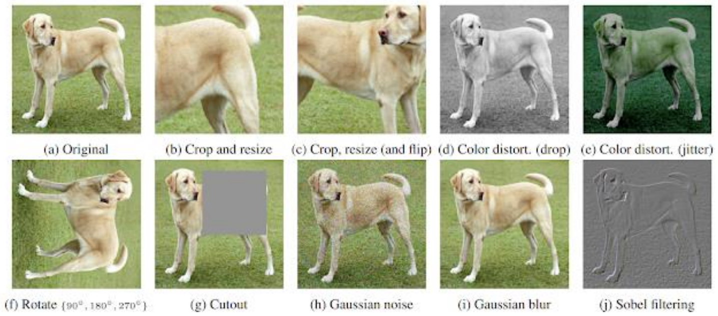
Multiple algorithm: MoCO, SimCLR, CMC, CLIP…

One loss: $$\mathcal{L}_{\text{InfoNCE}} = -\mathbb{E}_{x,x^+,x_i^-}\left[\log \frac{\exp\left(f(x)^T f(x^+)/\tau\right)}{\exp\left(f(x_i)^T f(x_i^+)/\tau\right)+\sum_i \exp\left(f(x)^T f(x_i^-)/\tau\right)}\right]$$

Pros	Cons
• Simple • Clean theoretical grounding • No need for label • allow to marginalize over known nuisance effects.	• Need negative pairs • Need large Batch (<4096) • False negative Pairs • Sensitive to hyperparameters



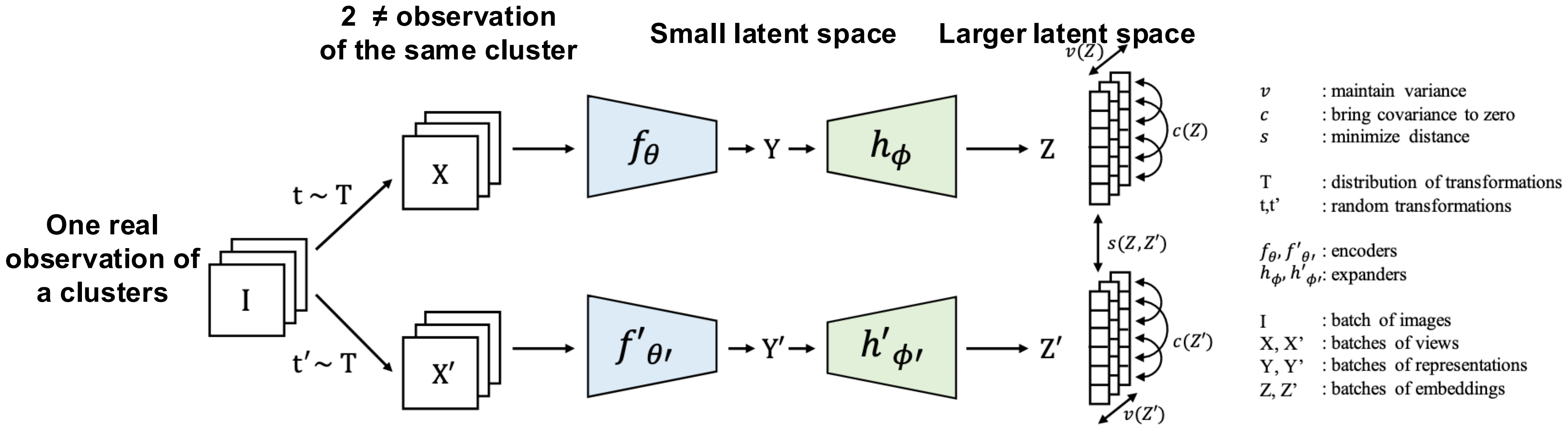




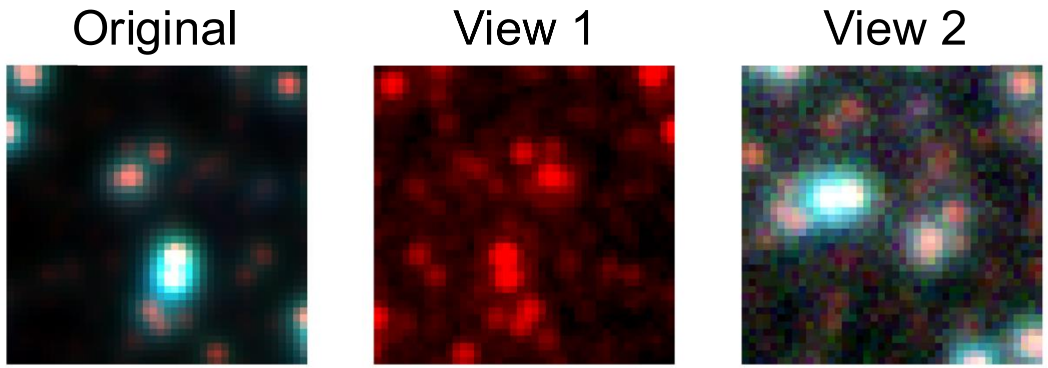
Fine-tuning reuses the pre-trained encoder's features for a downstream task

VICReg

VICReg: Variance-Invariance-Covariance Regularization for Self-Supervised Learning, *Bardes et al. 2022*



$$\boxed{\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{inv}} + \mathcal{L}_{\text{var}} + \mathcal{L}_{\text{cov}}}$$



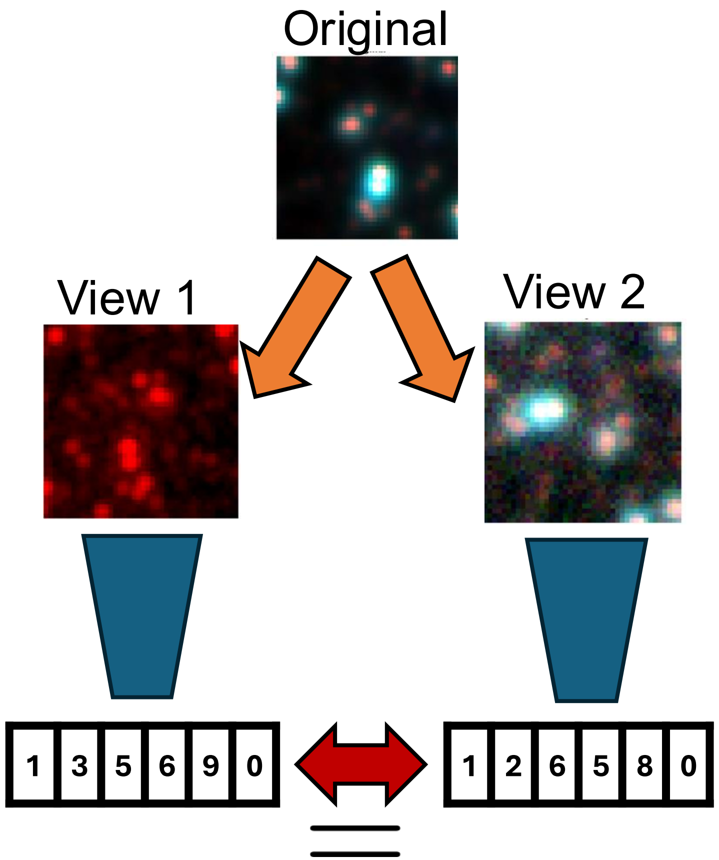
Advantages:

- **No negative Pair**
- Possibility to have different branches and input data
- No memory bank, No large batch needed
- Very stable
- No batch wise or feature normalization

VICReg loss

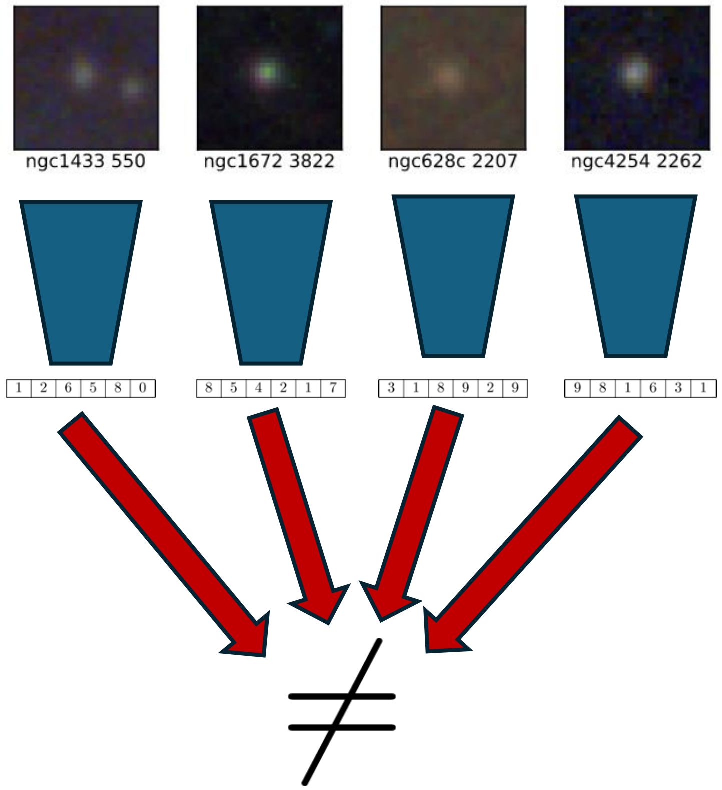
Loss to minimize: $\boxed{\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{inv}} + \mathcal{L}_{\text{var}} + \mathcal{L}_{\text{cov}}}$

Invariance (-)



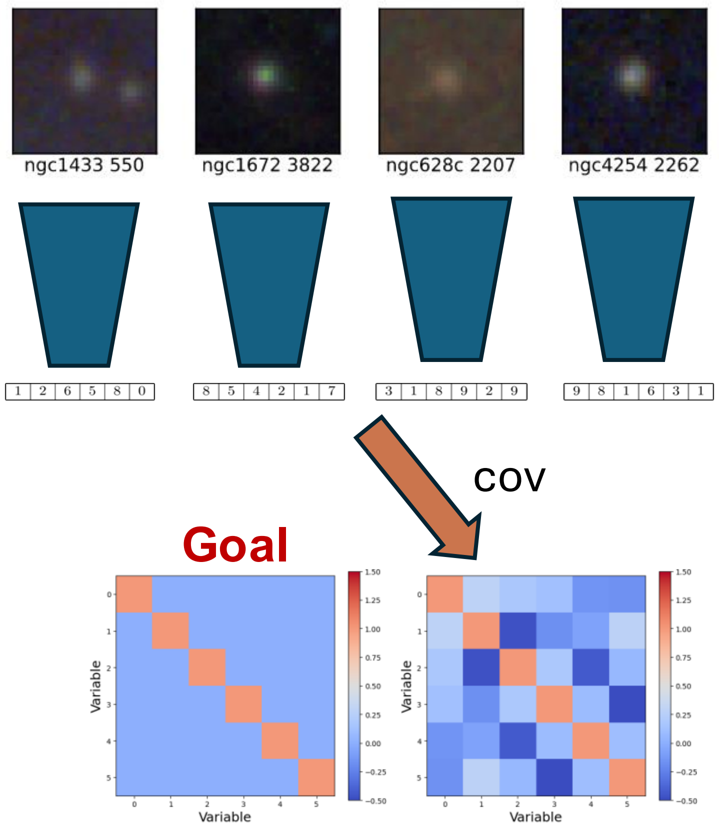
Same cluster => Same encoding

Variance (+)



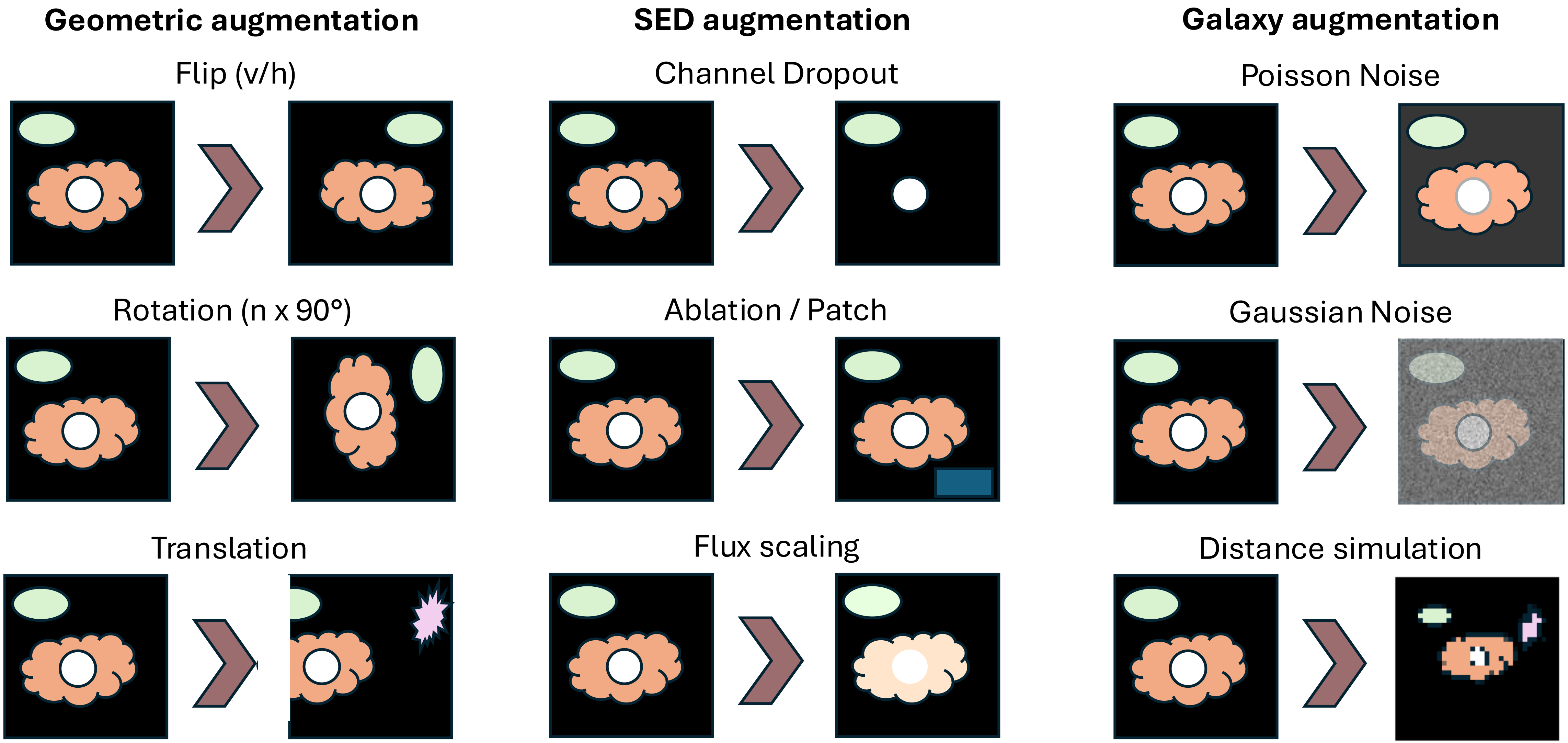
Different clusters => Different encodings

Covariance (-)



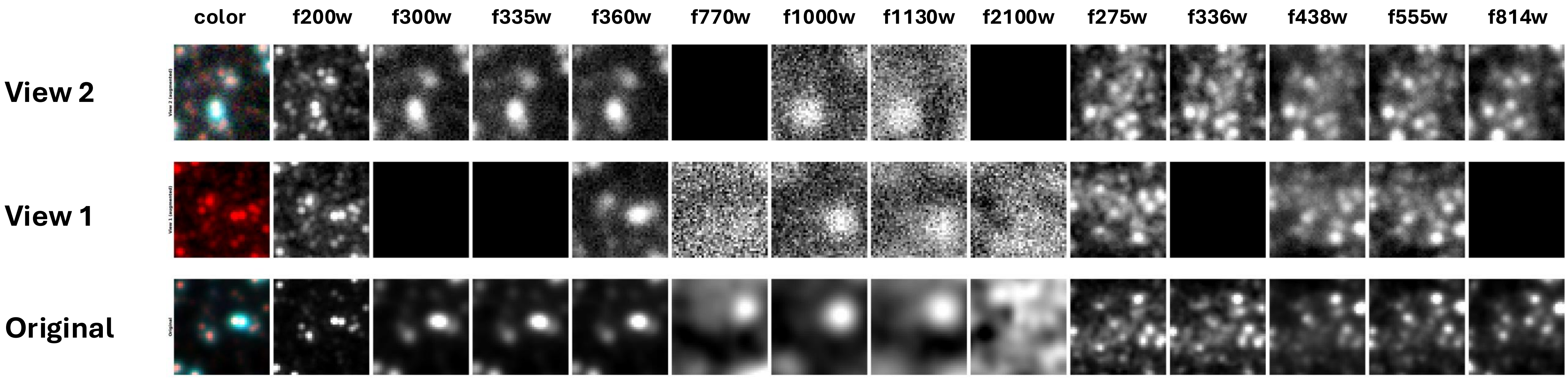
Each variable is a different parameter

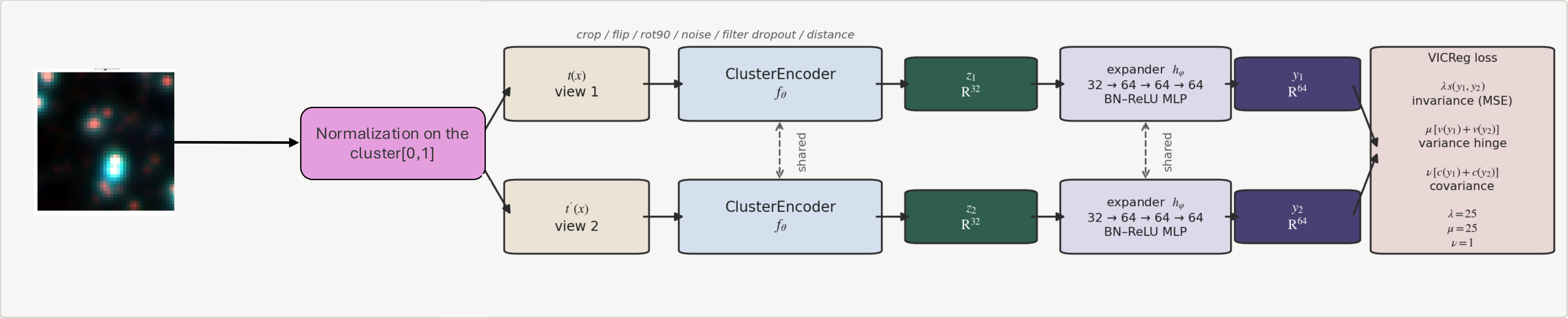
Augmentation



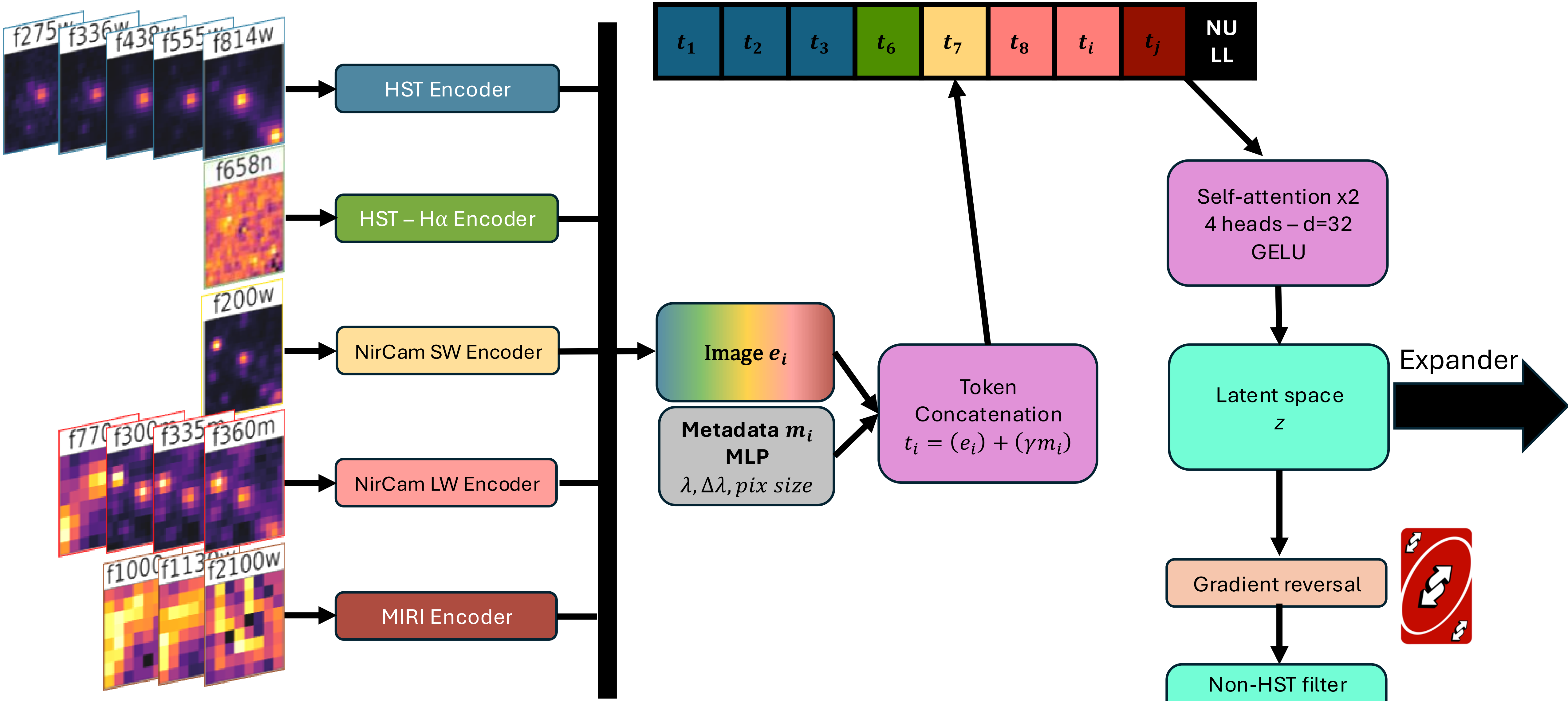
Augmentations create two views of the same object, so the model learns features invariant to nuisance transformations

Exemple of augmentation

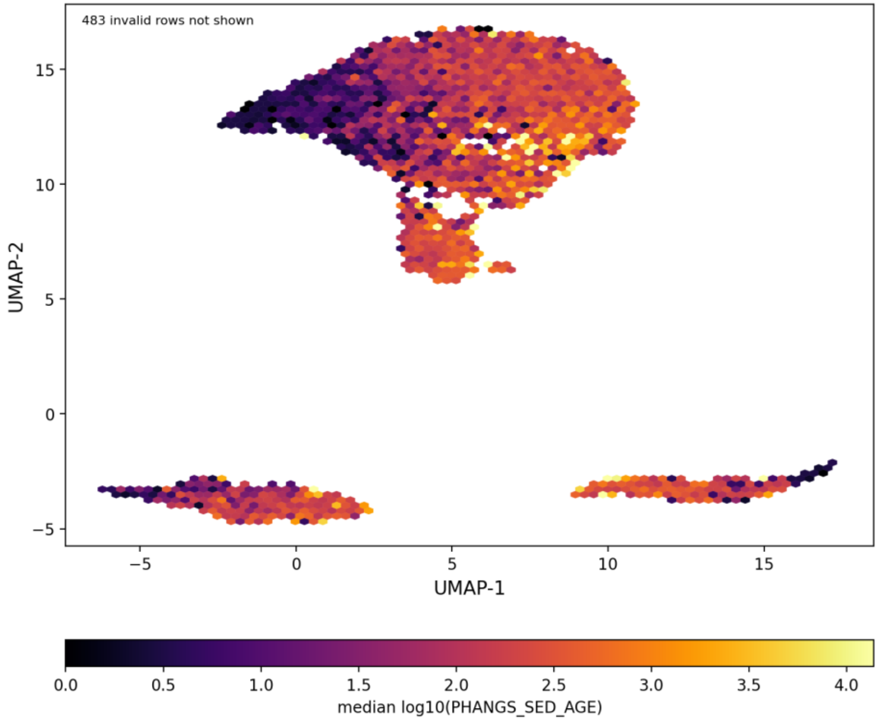




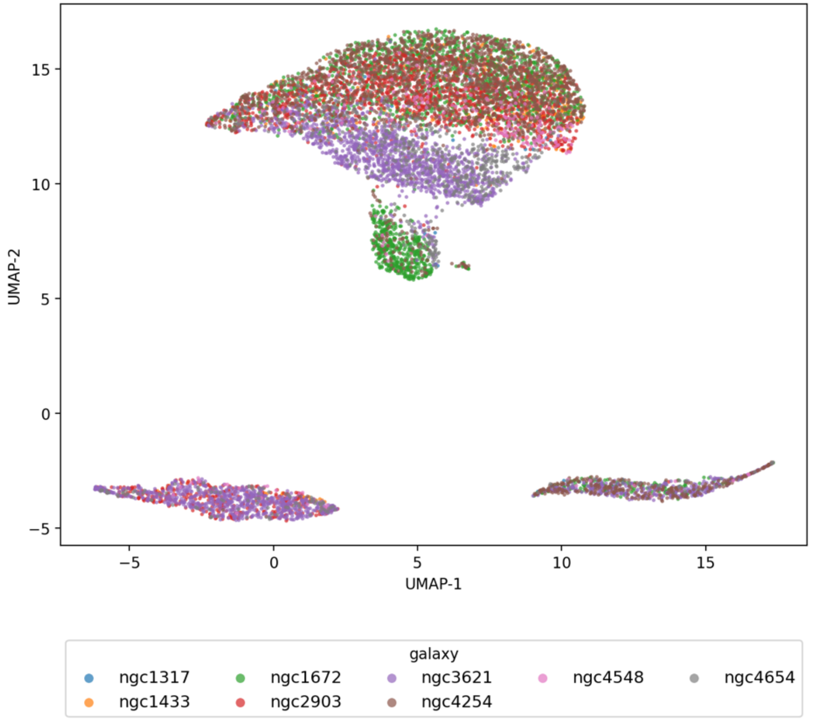
Encoder architecture



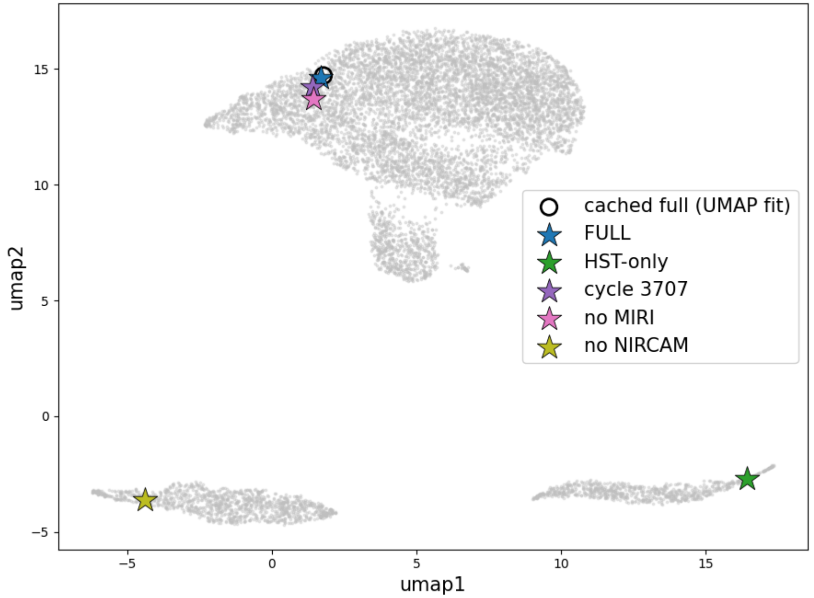
Results – UMAP



Age Structuration



Galaxy Structuration

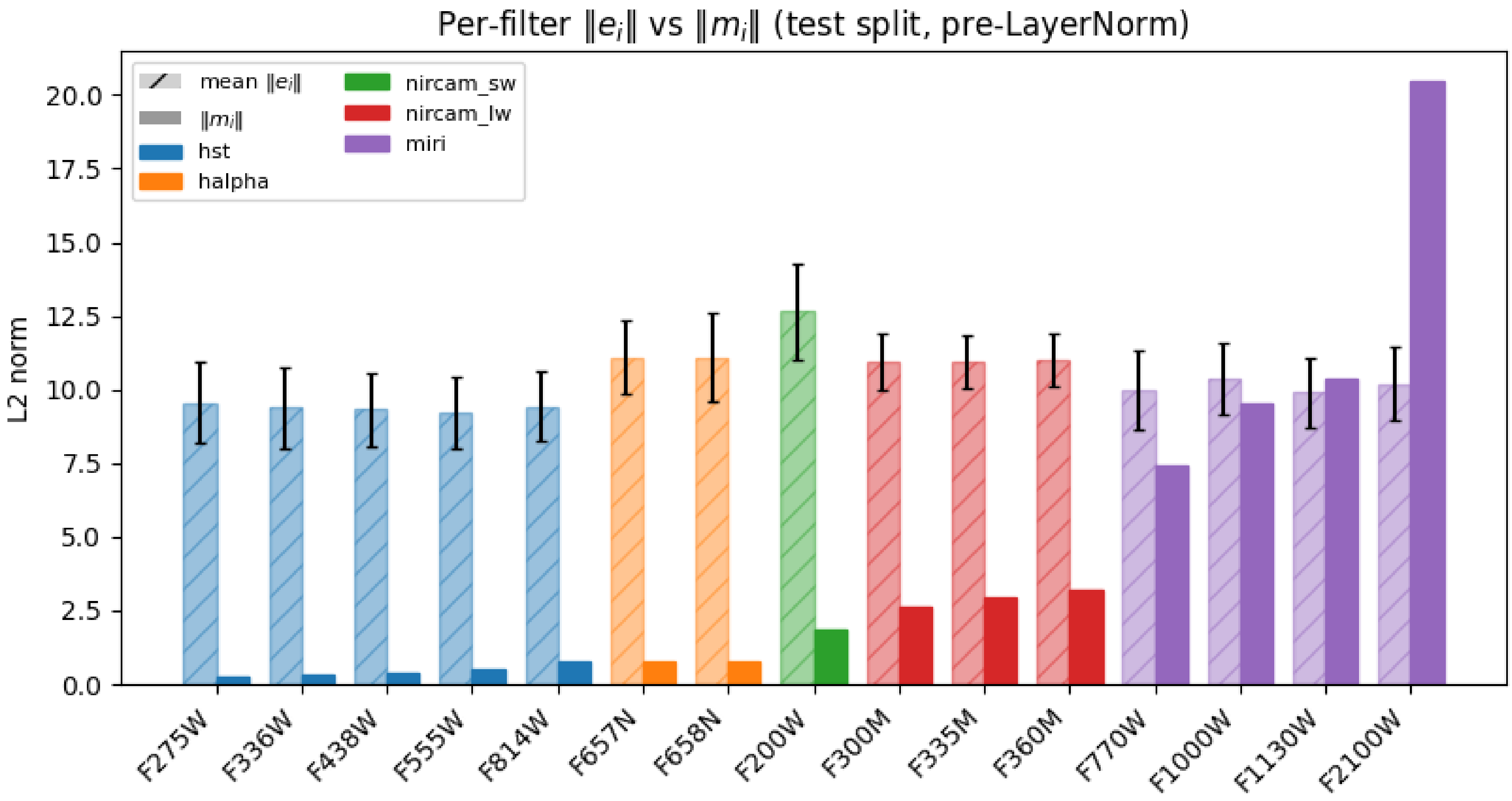


Ablation filters study

Candebat et al in prep

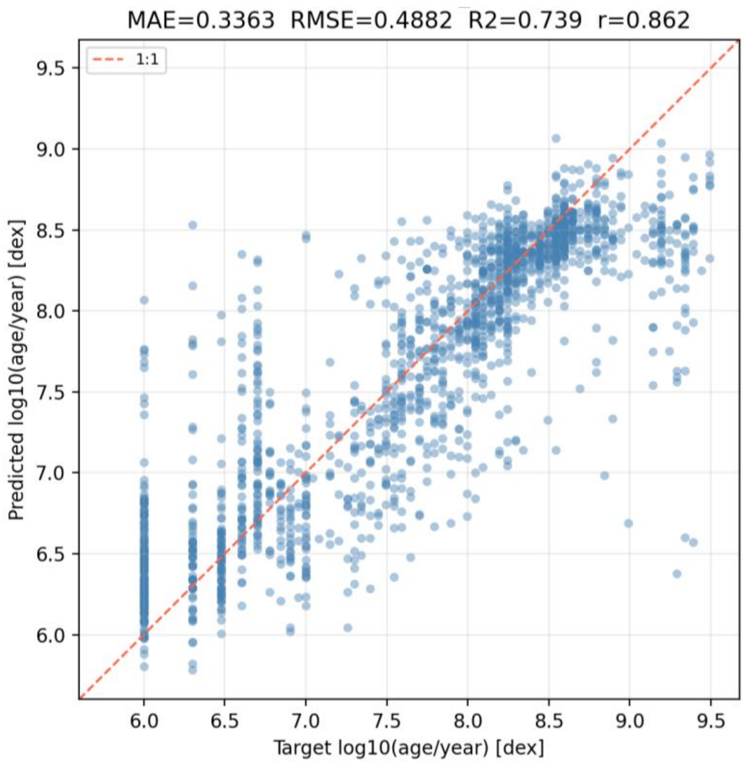
Results – Features importance

Ratio between the norm of the vector that come from the image $\| e_i \|$and the metadata $\| \gamma m_i \|$

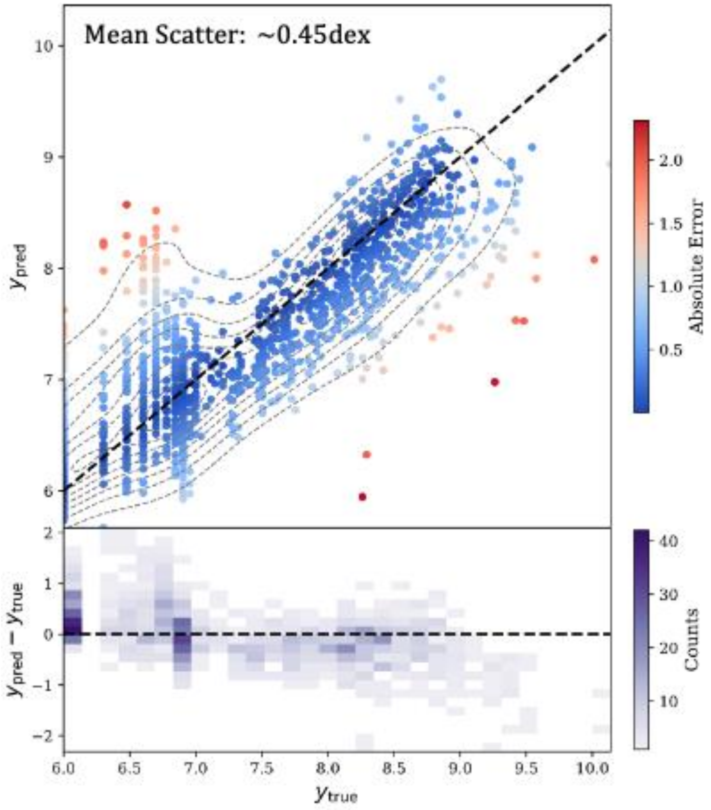


Candebat et al in prep

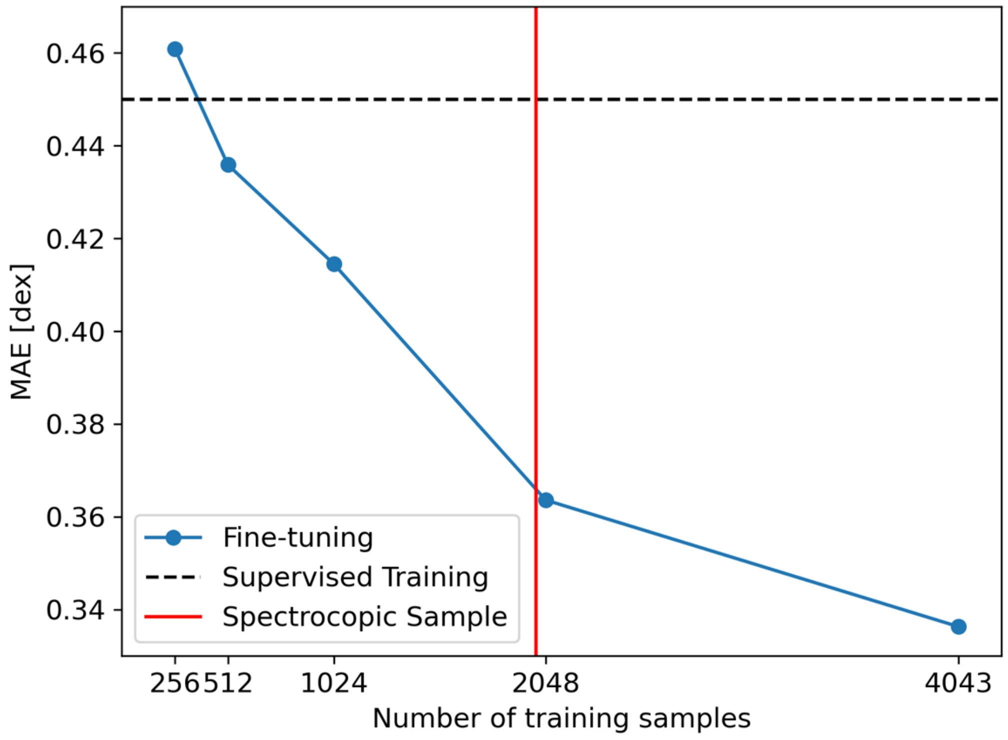
Results – fine-tunning



Fine-tuned Training
4,043 sources



Supervised Training
9,000 of the brightest sources
Viaña et al. 2026

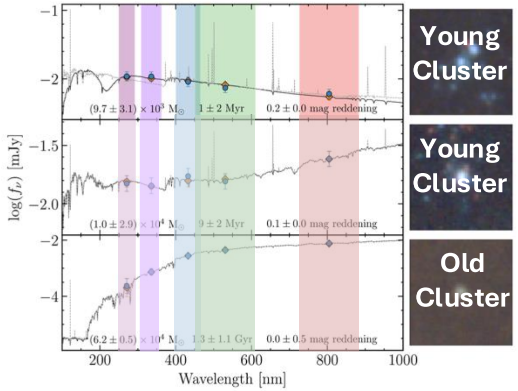


Fine-tuned Evolution

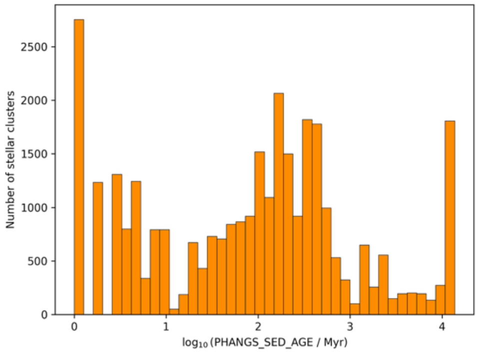
Candebat et al in prep

Conclusion

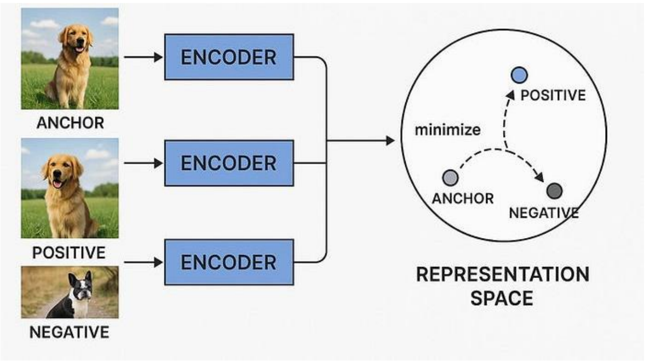
Stellar cluster age is important to understand ISM and galaxy evolution in nearby galaxies.



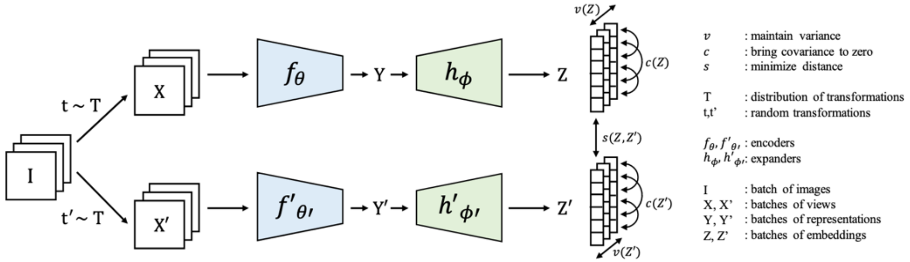
Stellar cluster PHANGS dataset is becoming bigger, but labels are currently unreliable



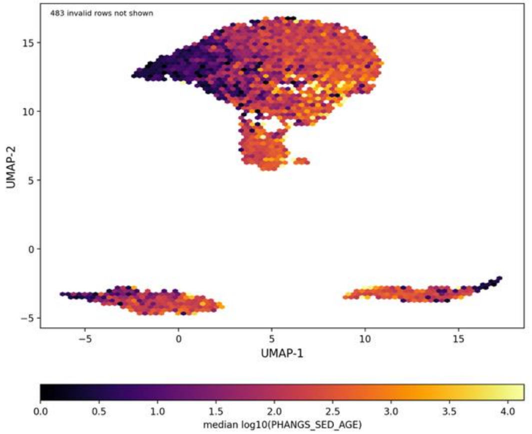
Contrastive Learning overcome the need of labels the dataset



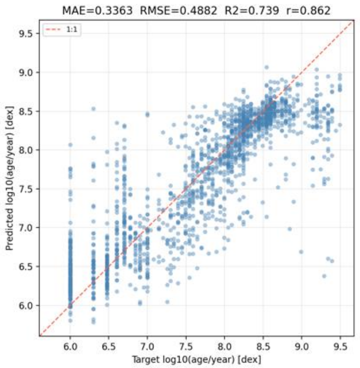
VICReg offers a compelling and simple solution by defining which transformations the network must be robust to.



The network learned what age is

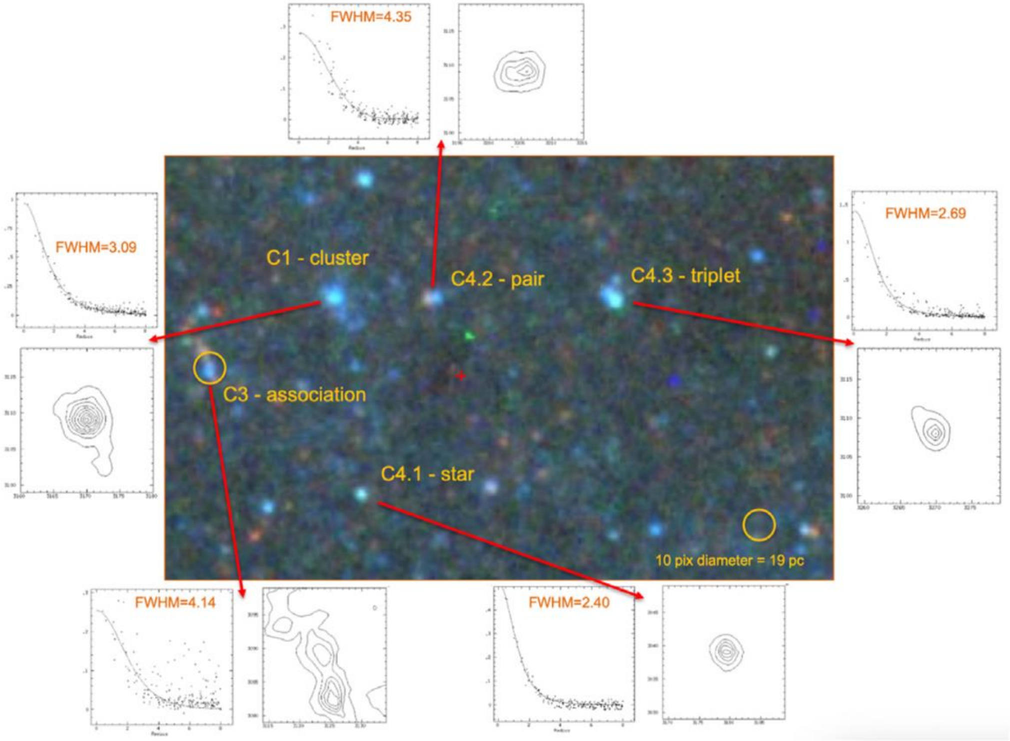


Fine-tuning gives better results on the benchmark, but this must be confirmed



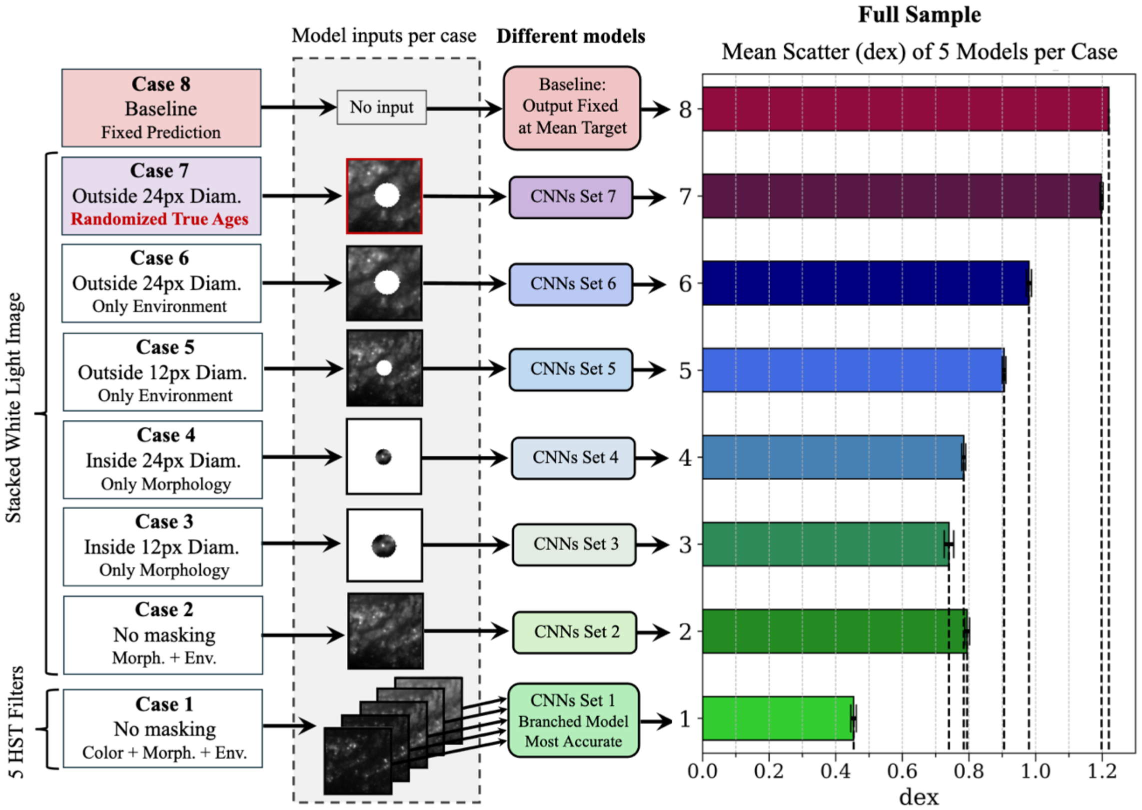
Thanks !!!!

Machine Learning for Stellar Cluster



Whitmore et al. 2021

Detection of Stellar Clusters

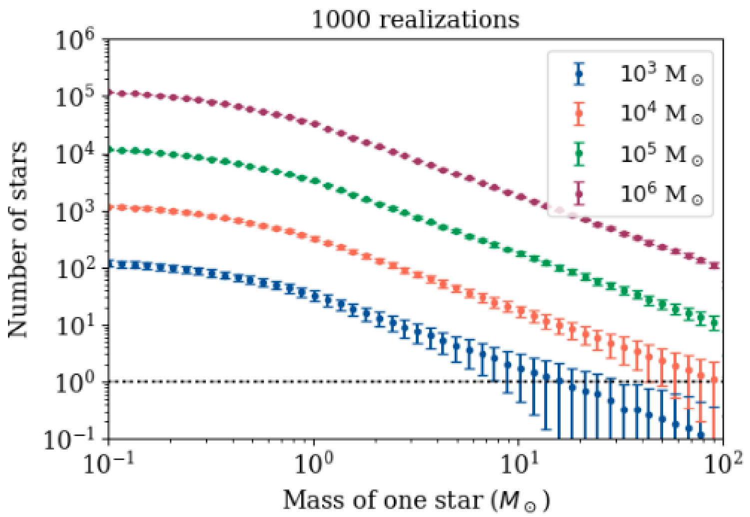


Viaña et al. 2026

Age estimation (supervised learning)

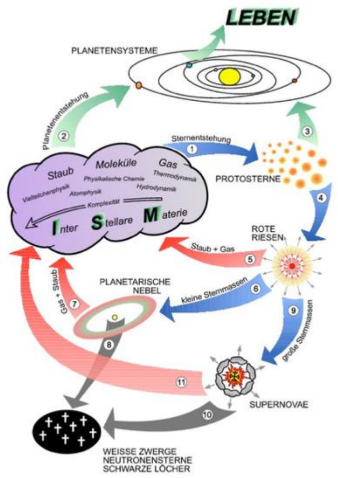
Why study stellar cluster ?

SFR: Star Formation Rate
GMC : Giant Molecular Cloud
IMF: Initial Mass Function
ISM: Interstellar Medium



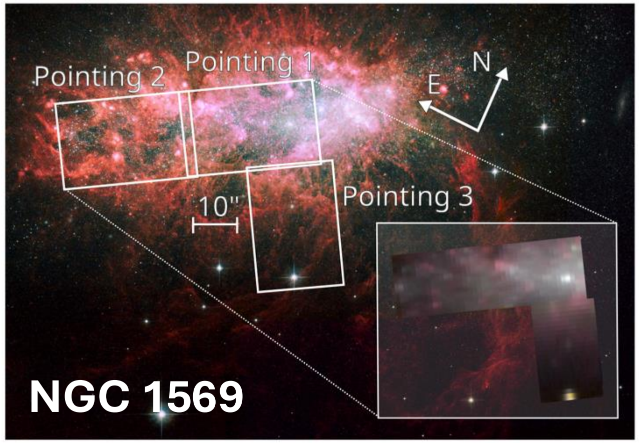
Orozco-Duarte et al .2021

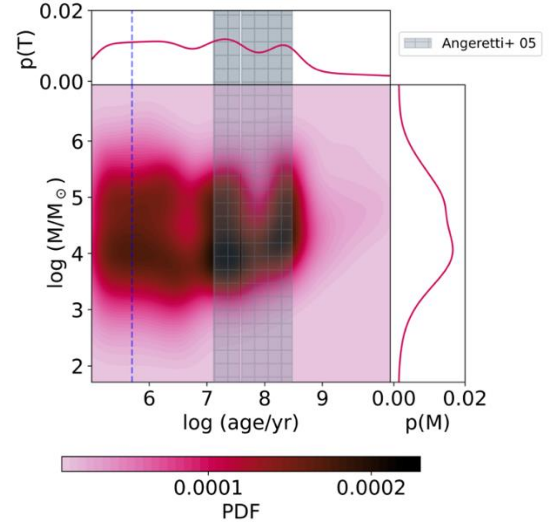
Star (IMF)



Klessen

ISM



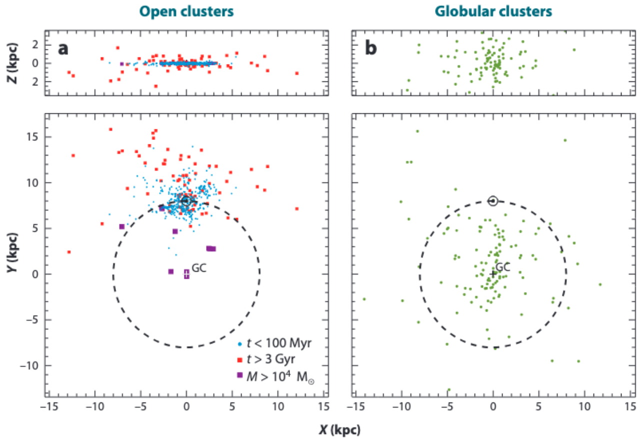


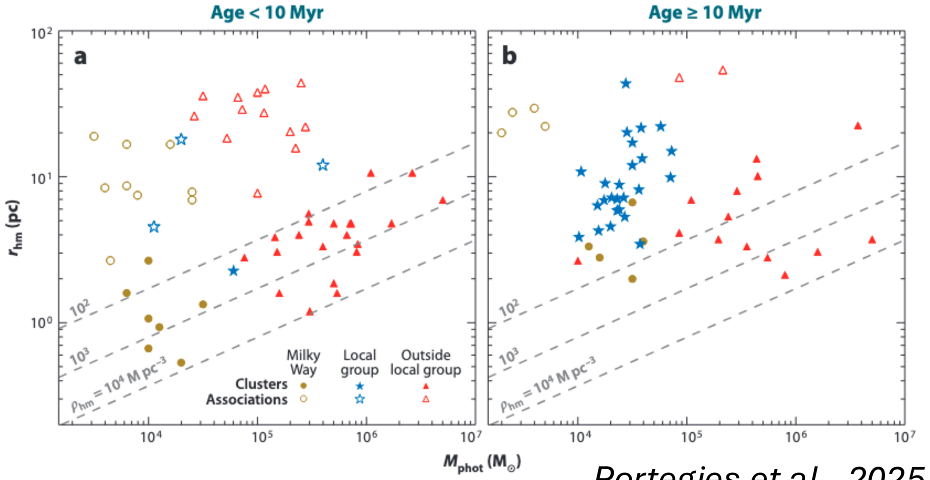
Björgvinsson et al. 2026

Galaxies

Problem in Milky Way:

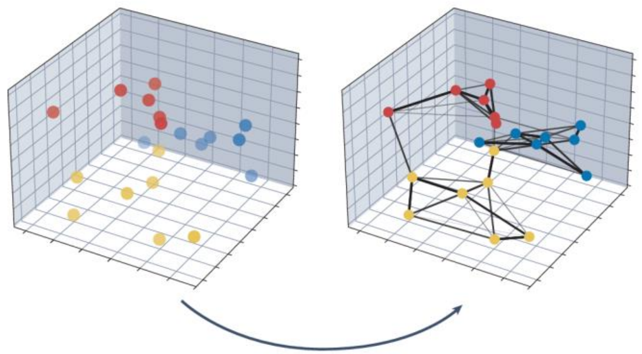
- Low number, completeness
- Properties are not independent of galaxies



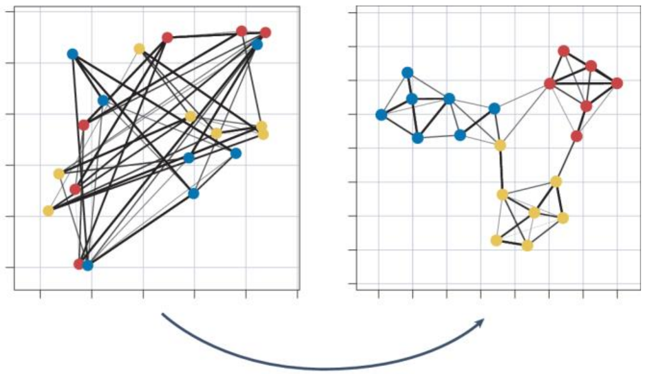


Portegies et al., 2025

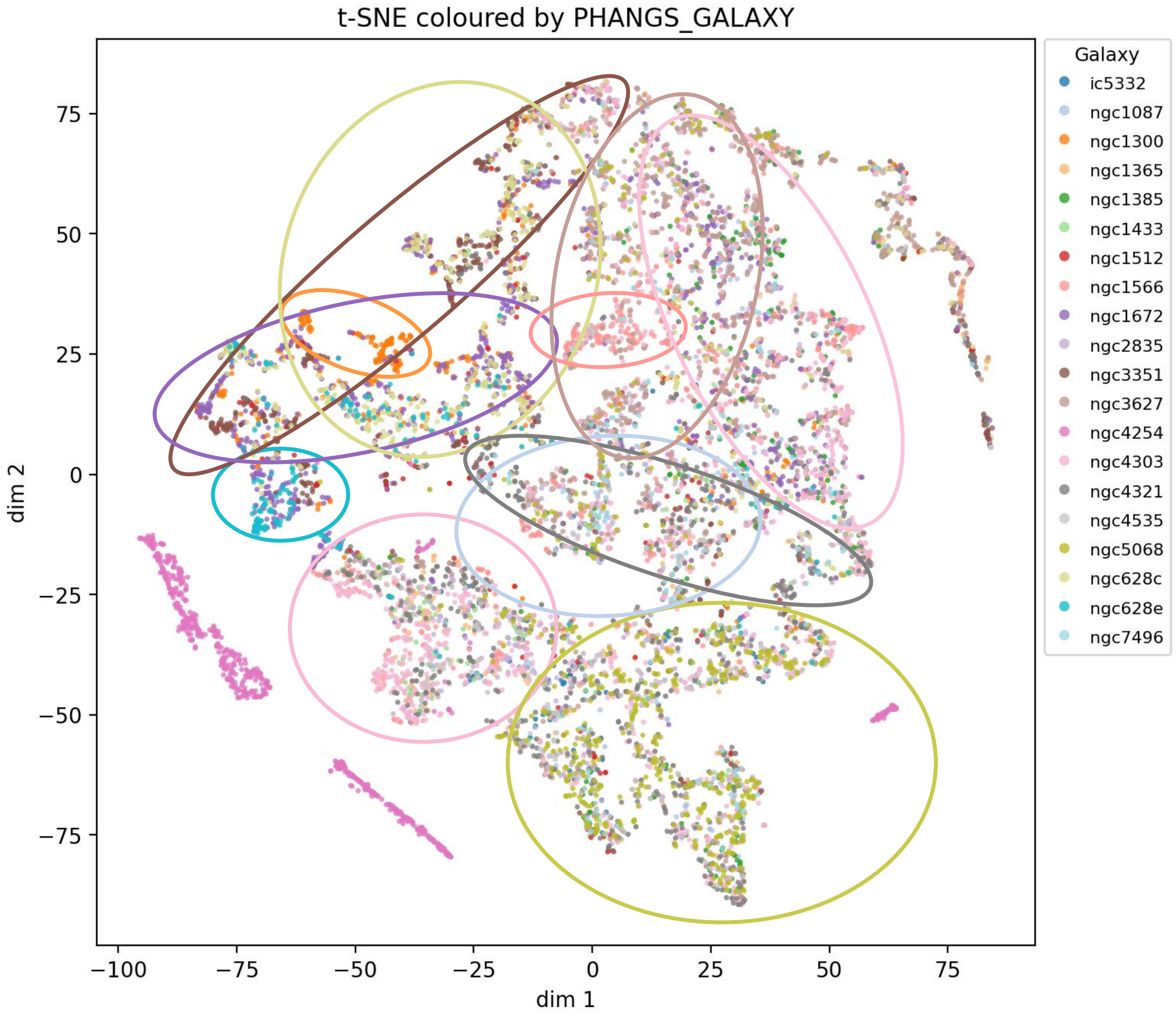
Umaps

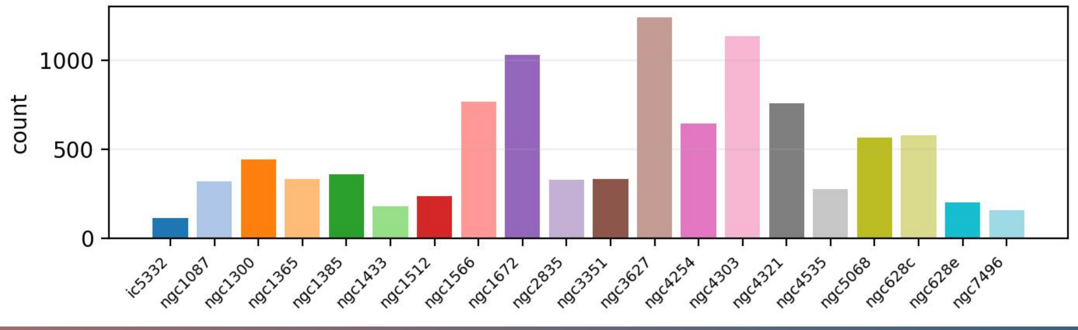


Build a topological representation of
the high-dimensional data

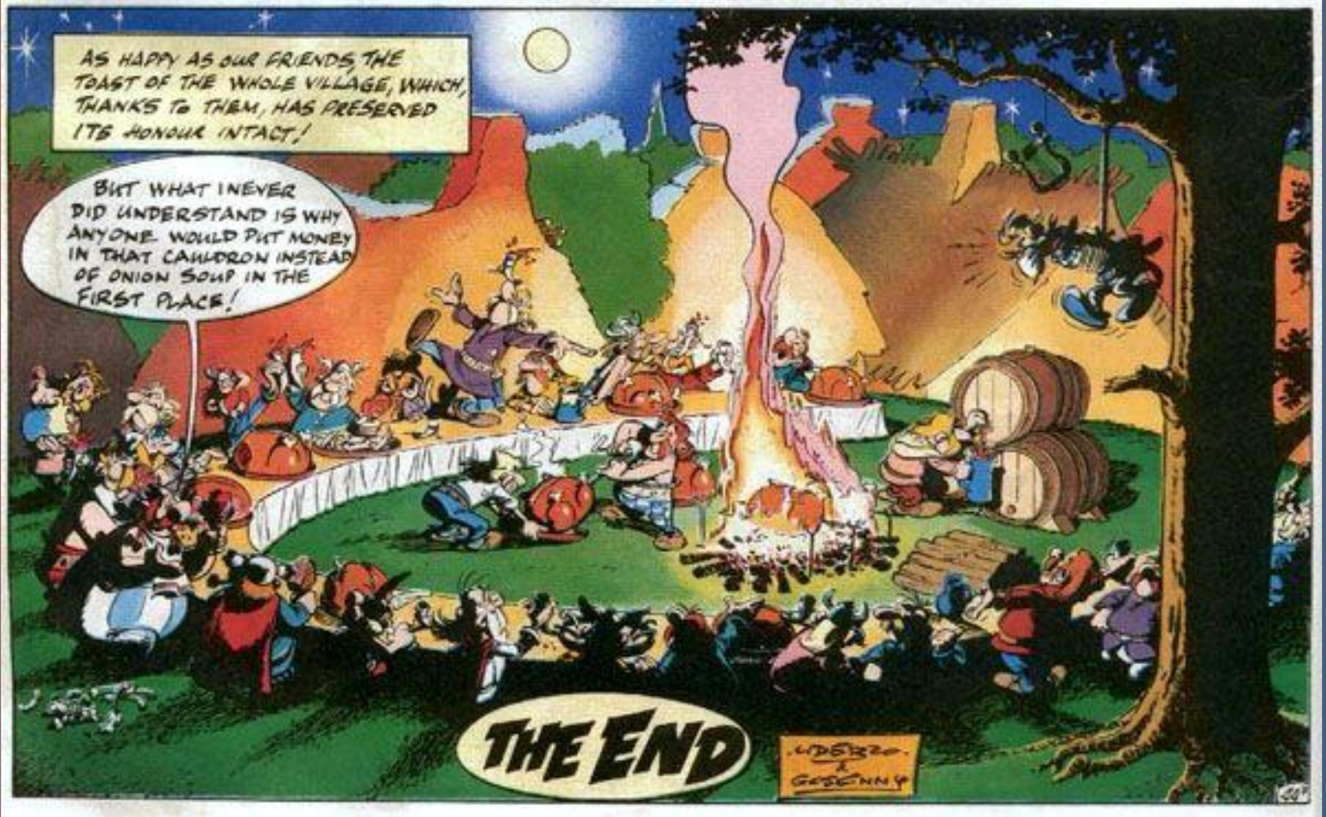


Optimize a low-dimensional embedding
to have similar topological representation



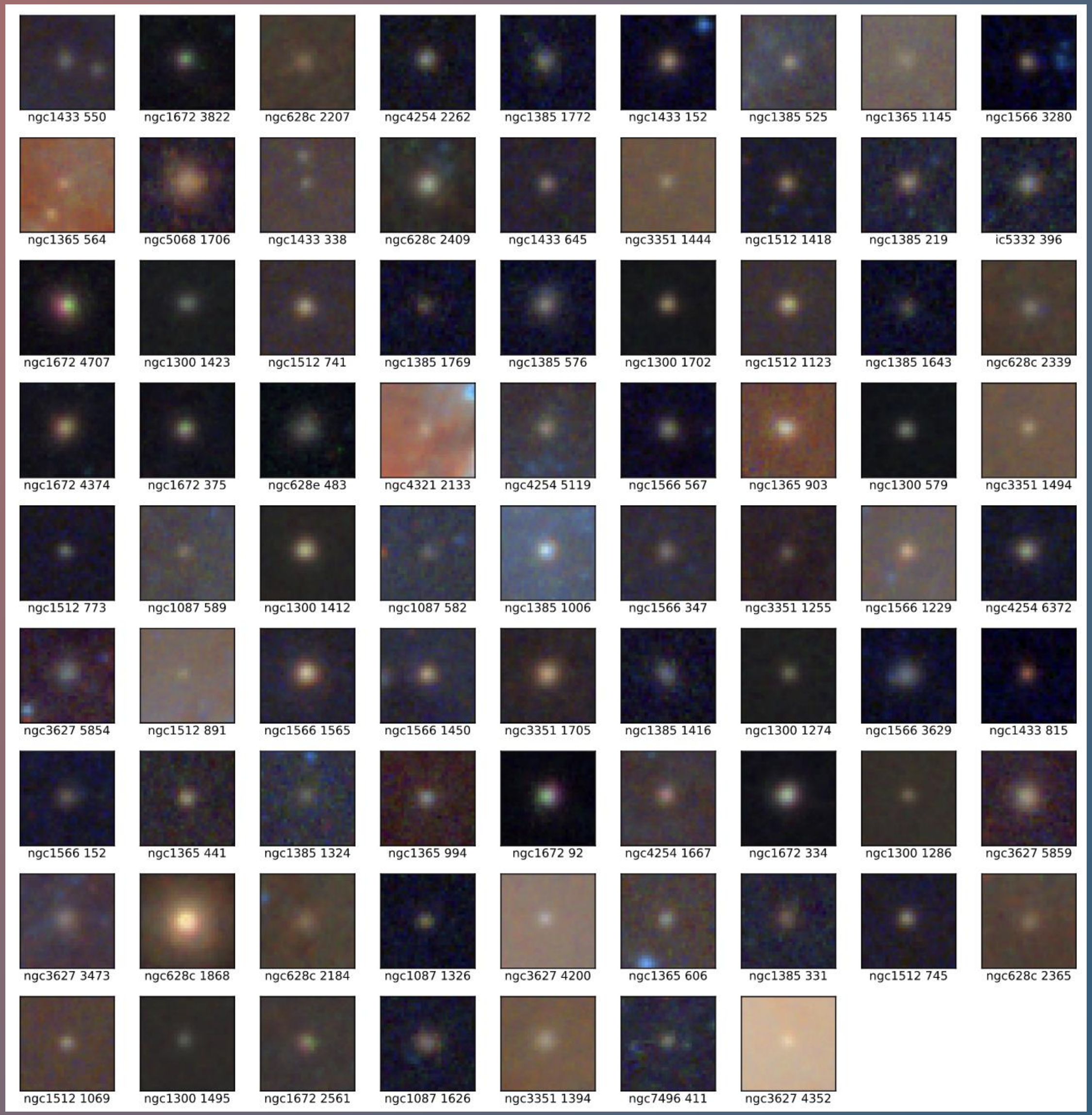


Conclusion



Asterix and Obelix and the cauldron

Project


ngc1433 550
ngc1672 3822
ngc628c 2207
ngc4254 2262
ngc1385 1772
ngc1433 152
ngc1385 525
ngc1365 1145
ngc1566 3280
ngc1365 564
ngc5068 1706
ngc1433 338
ngc628c 2409
ngc1433 645
ngc3351 1444
ngc1512 1418
ngc1385 219
ic5332 396
ngc1672 4707
ngc1300 1423
ngc1512 741
ngc1385 1769
ngc1385 576
ngc1300 1702
ngc1512 1123
ngc1385 1643
ngc628c 2339
ngc1672 4374
ngc1672 375
ngc628e 483
ngc4321 2133
ngc4254 5119
ngc1566 567
ngc1365 903
ngc1300 579
ngc3351 1494
ngc1512 773
ngc1087 589
ngc1300 1412
ngc1087 582
ngc1385 1006
ngc1566 347
ngc3351 1255
ngc1566 1229
ngc4254 6372
ngc3627 5854
ngc1512 891
ngc1566 1565
ngc1566 1450
ngc3351 1705
ngc1385 1416
ngc1300 1274
ngc1566 3629
ngc1433 815
ngc1566 152
ngc1365 441
ngc1385 1324
ngc1365 994
ngc1672 92
ngc4254 1667
ngc1672 334
ngc1300 1286
ngc3627 5859
ngc3627 3473
ngc628c 1868
ngc628c 2184
ngc1087 1326
ngc3627 4200
ngc1365 606
ngc1385 331
ngc1512 745
ngc628c 2365
ngc1512 1069
ngc1300 1495
ngc1672 2561
ngc1087 1626
ngc3351 1394
ngc7496 411
ngc3627 4352

Long term goals

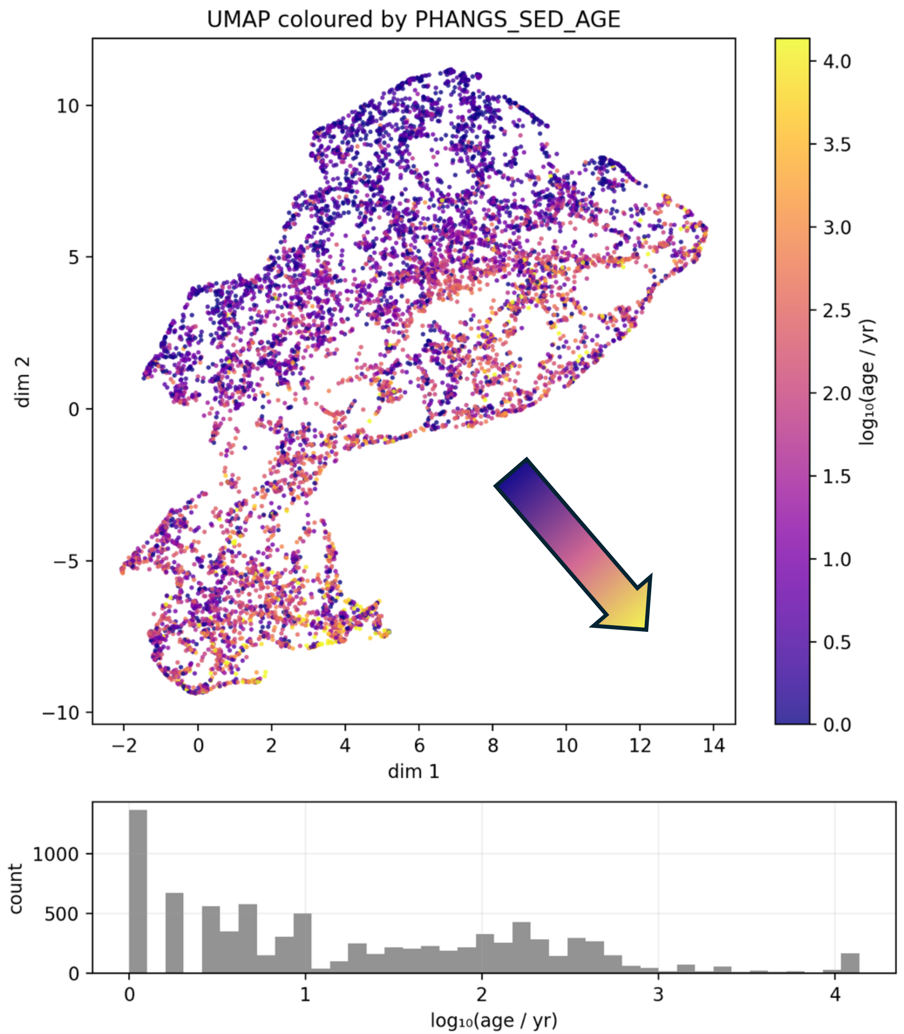
1. Study the embedding space

2. Fine-tunes with high quality data

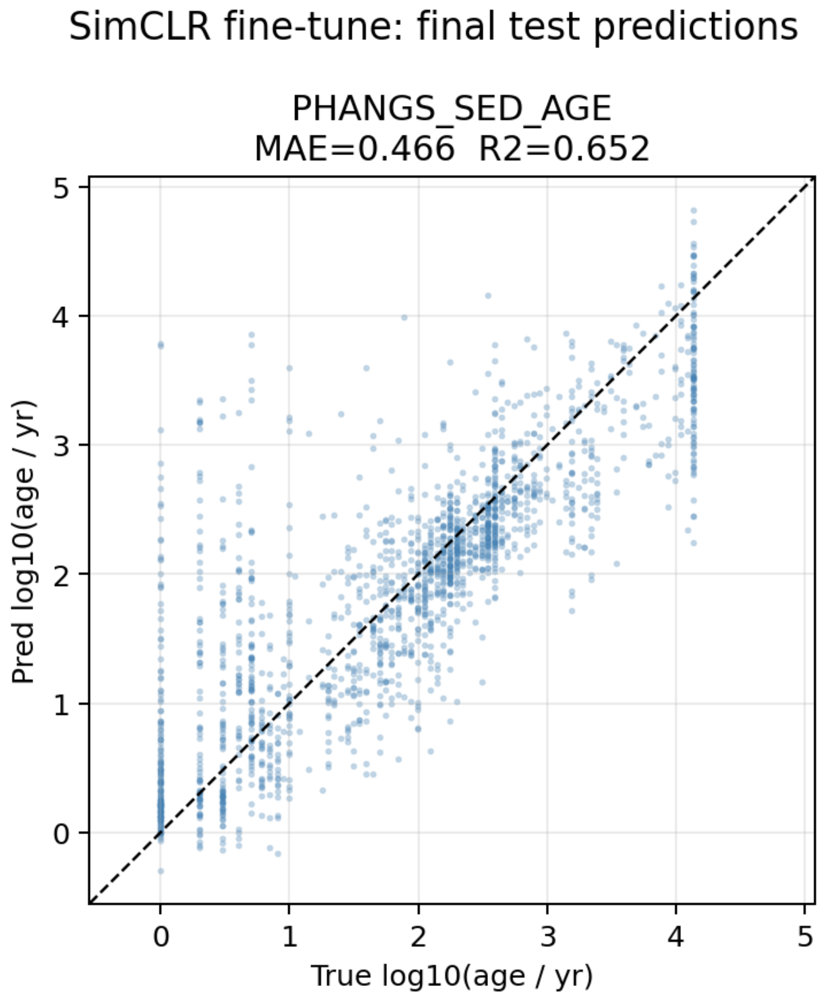
3. Go to more complex parameters / quantities with nearby galaxies (M33 / LMC) as a fine-tuning dataset

4. Make the Neural Network reliable with different combinations of filters (different cycle)

Results – fine-tunning

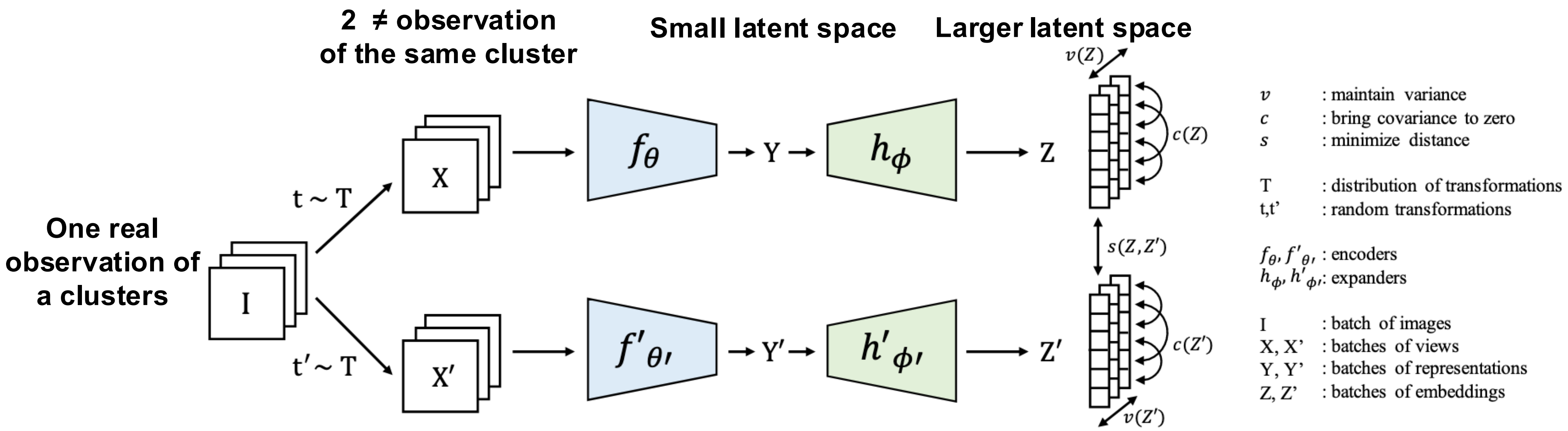


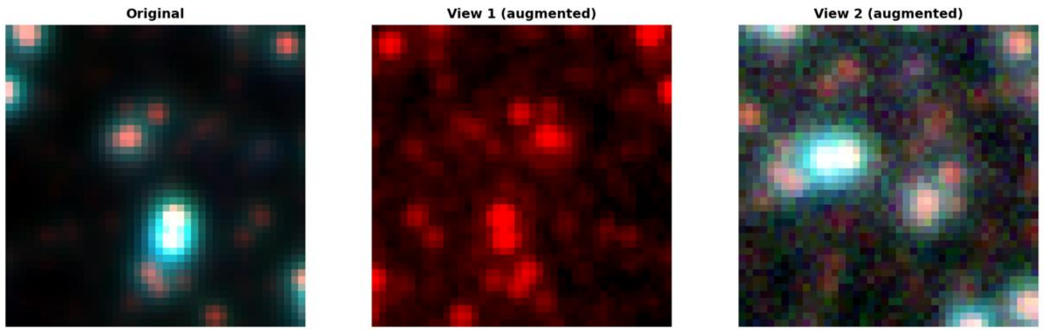
UMAP (Uniform Manifold Approximation and Projection) is a dimensionality reduction technique that preserves both local and global data structure



Fine tunning

Neural Network





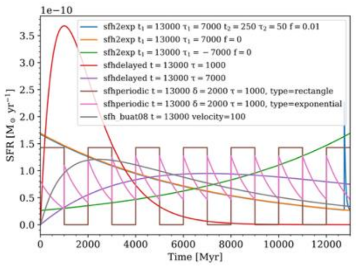
- **Invariance**: the mean square distance between the embedding vectors.
- **Variance**: a hinge loss to maintain the standard deviation (over a batch) of each variable of the embedding above a given threshold. This term forces the embedding vectors of samples within a batch to be different.
- **Covariance**: a term that attracts the covariances (over a batch) between every pair of (centered) embedding variables towards zero. This term decorrelates the variables of each embedding and prevents an *informational collapse* in which the variables would vary together or be highly correlated.

SED fitting

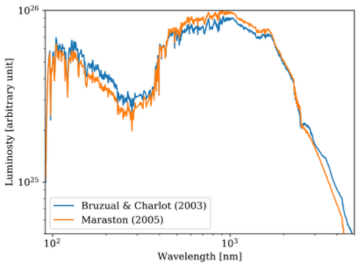
SFH: Star Formation History
SSP: Single Stellar Population
Z = Metallicity
IMF: Initial Mass Function
E(B−V) : reddening

1 - Create a model:

1. SFH

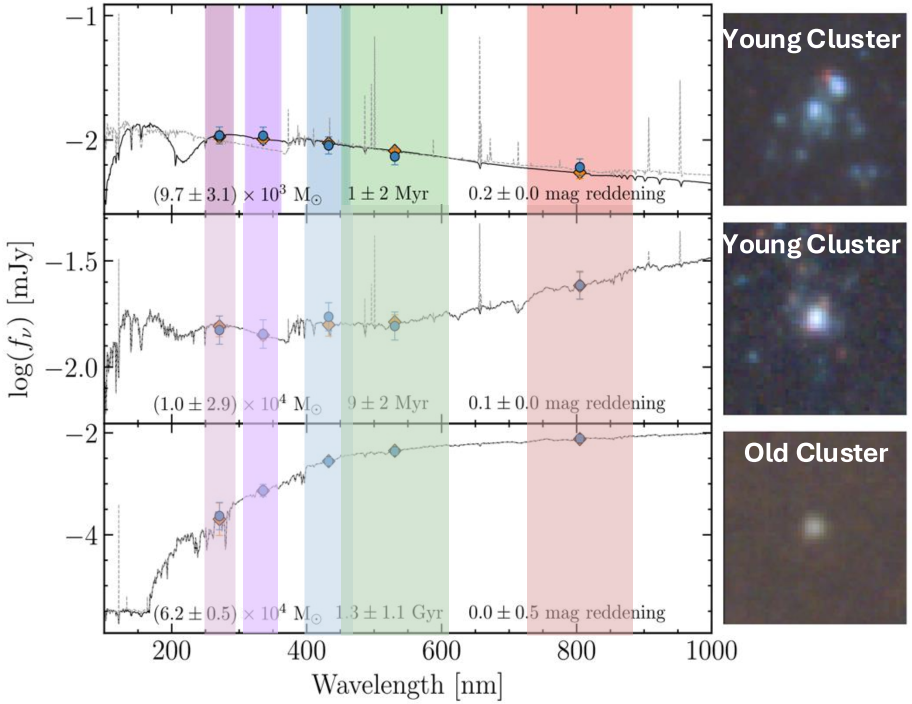


2. SSP (Z + IMF)



3. Interactions:

- Dust contribution
- Nebular emission
- Attenuation



Turner et al. 2021

2 – Fit the model to the data:

These models can be used to evaluate physical properties based on the likelihood.

CIGALE (Code Investigating the GALaxies Emission) THROUGH AN ENERGY BUDGET

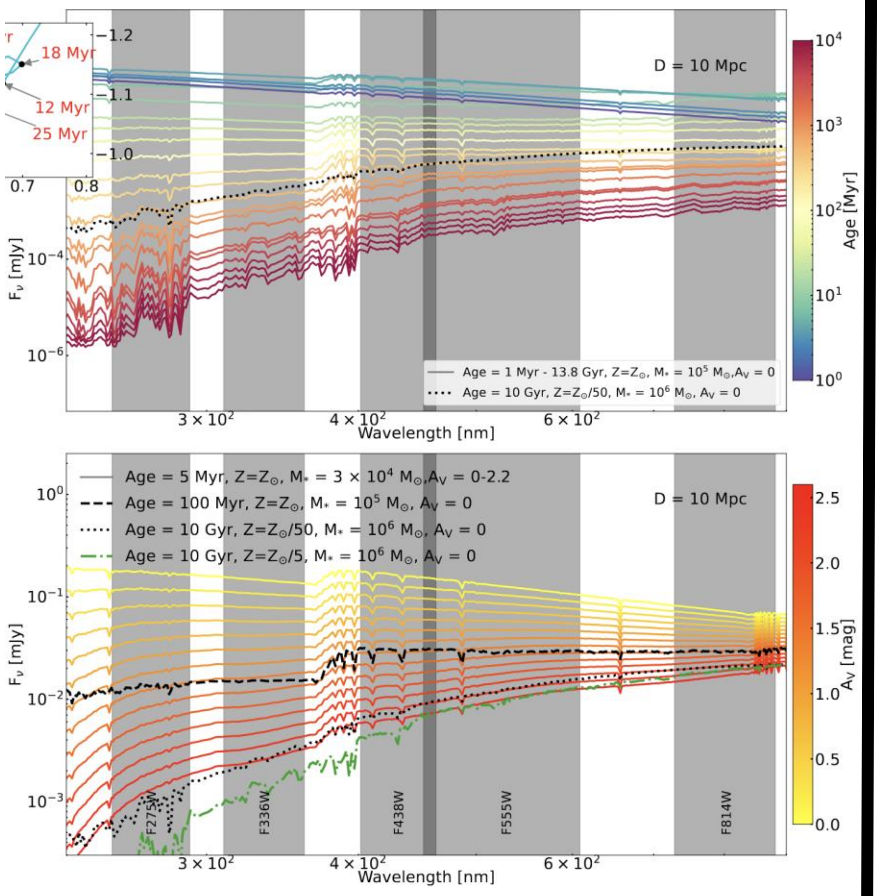
Boquien et al. 2019

Parameters:

- Age
- Mass
- E(B-V)

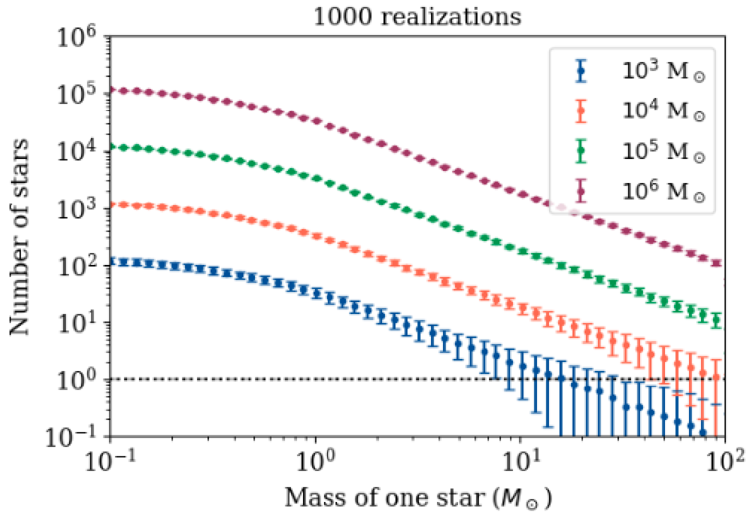
Problems

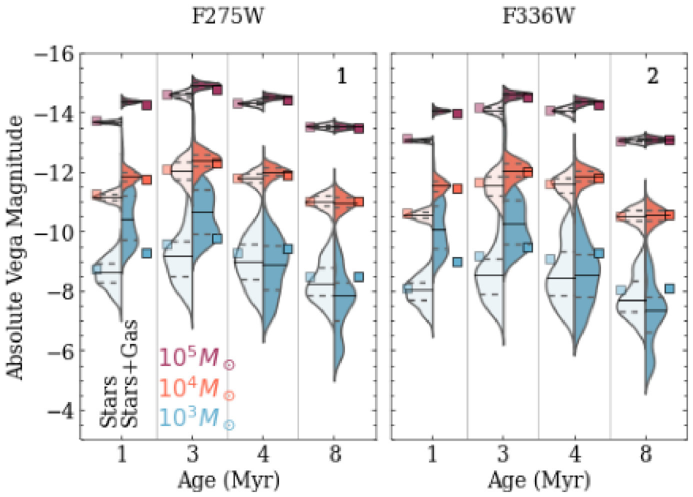
Degeneracies



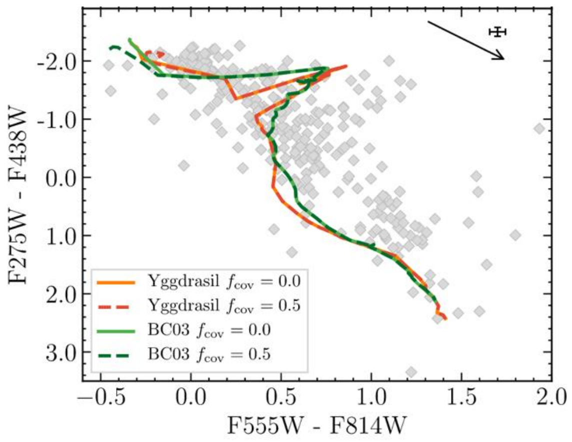
Thilker et al. 2025

Model assumptions (IMF, component contribution, model, …)





Orozco-Duarte et al .2021



Turner et al .2021