

Foundation Models Hackathon

Using AstroClip, astroPT, AION with Euclid Q1 images and DESI spectra.

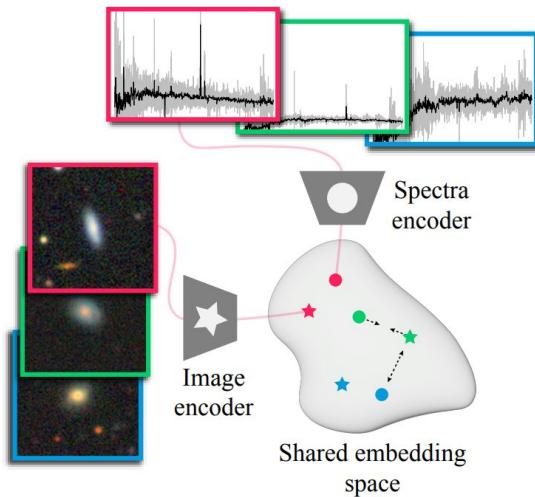
AstroCLIP : Simone, Anna, Clara

astroPT : Miguel, Julien, Gosia, Marc

AION : Wassim, Maxime, Sybille

3 different FM models

AstroClip



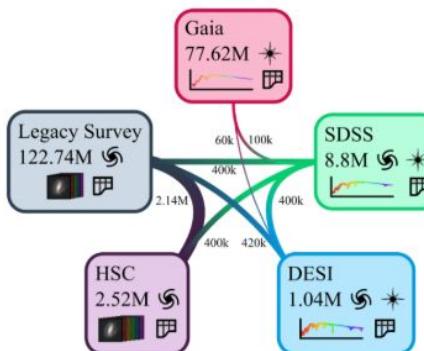
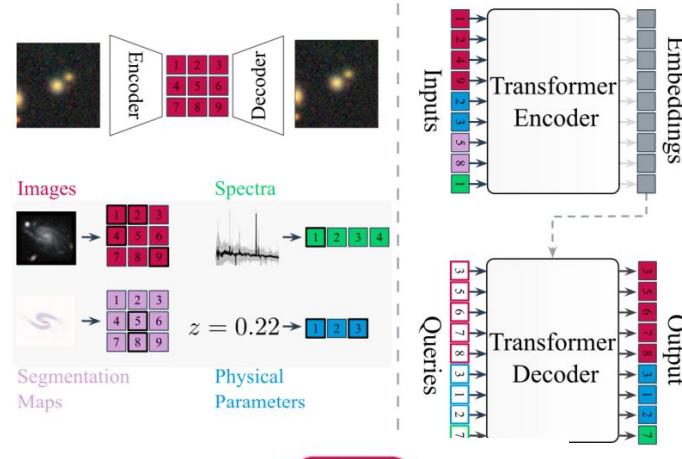
Legacy Survey
122.74M



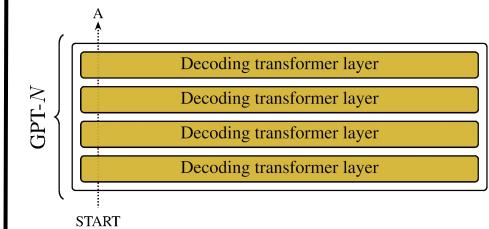
DESI
1.04M



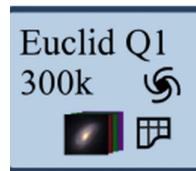
AION



AstroPT

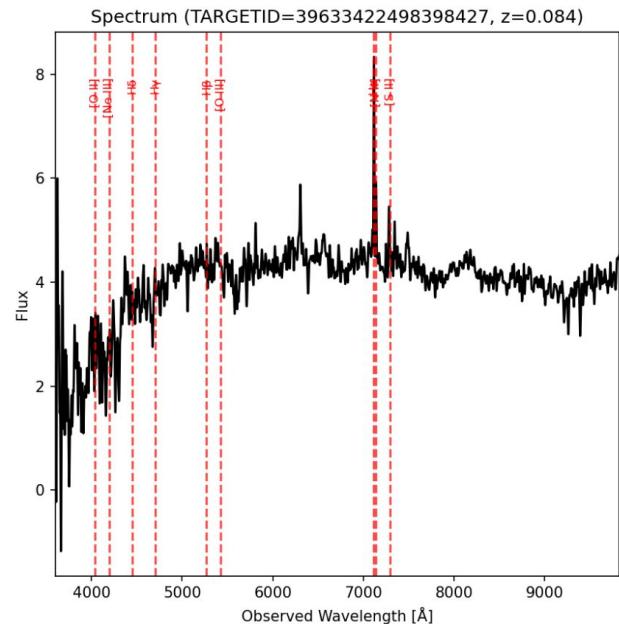


16	15	14	13	12
17	04	03	02	11
18	05	00	01	10
19	06	07	08	09
20	21	22	23	24

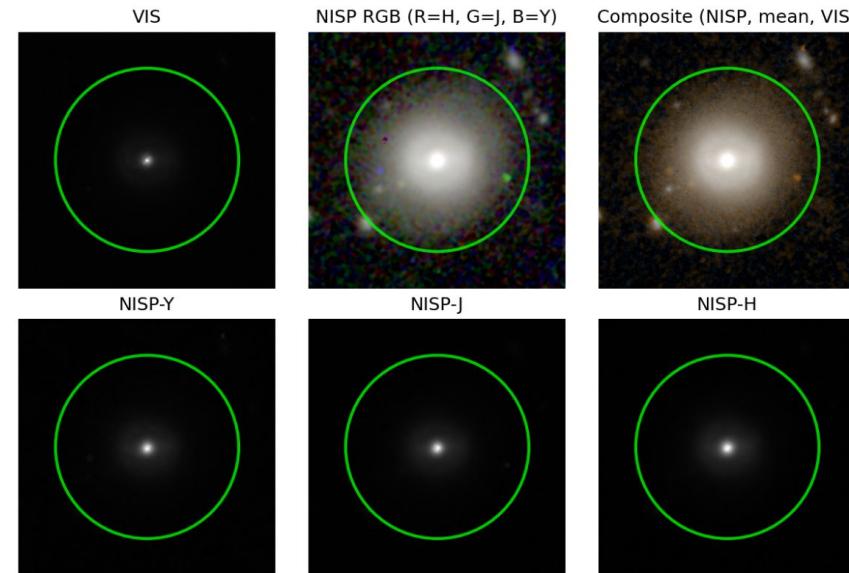


1 same dataset

Euclid (HR images)



DESI (HR spectra)



Hugging Face

Size of downloaded dataset files:
16.4 GB

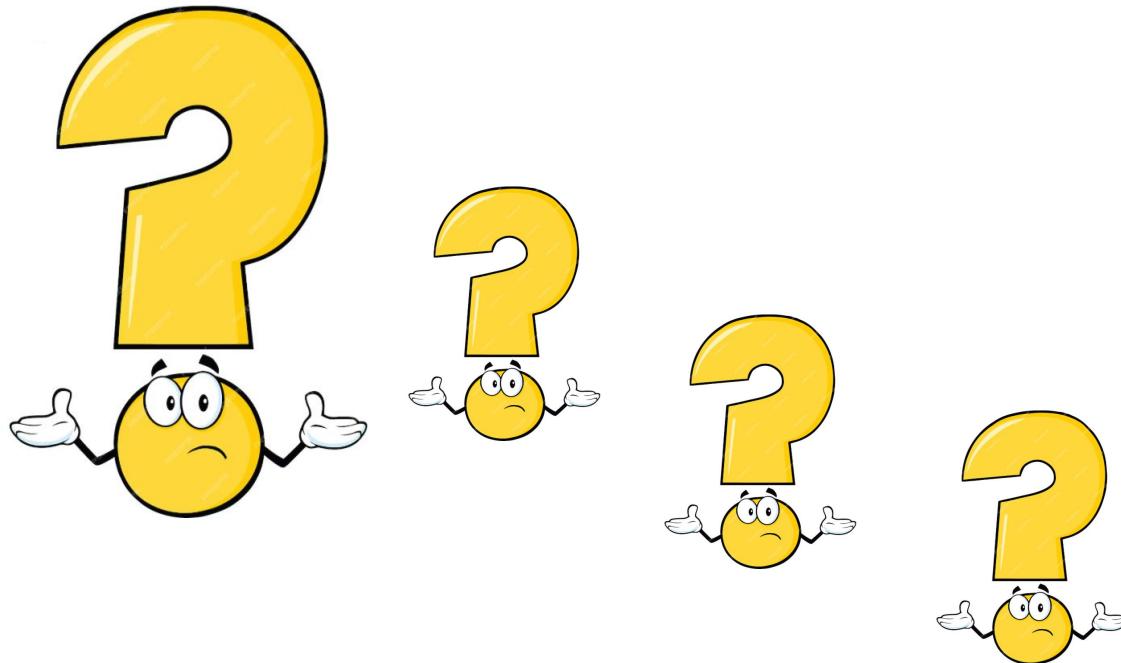
Size of the auto-converted Parquet files:
16.4 GB

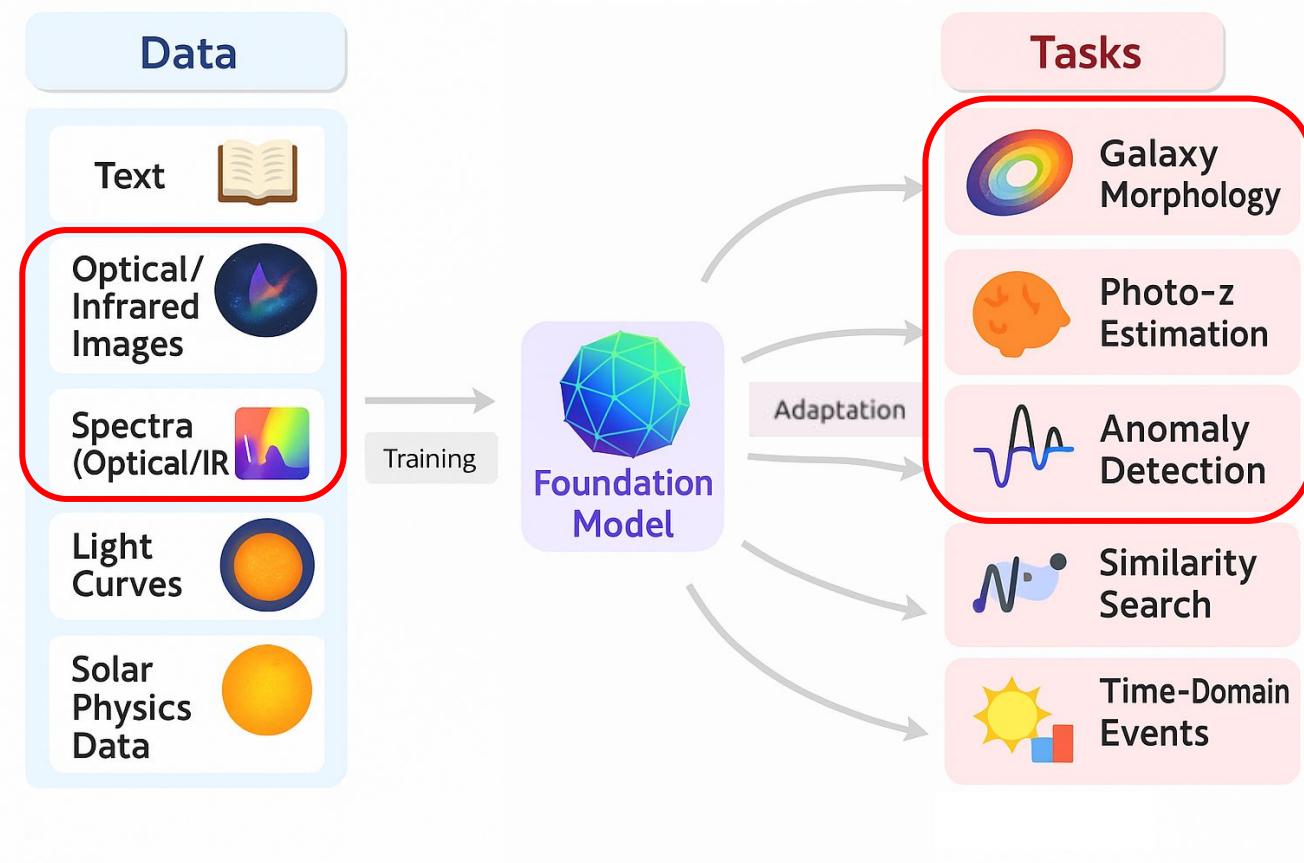
Number of rows:
15,263

[dataset online](#)

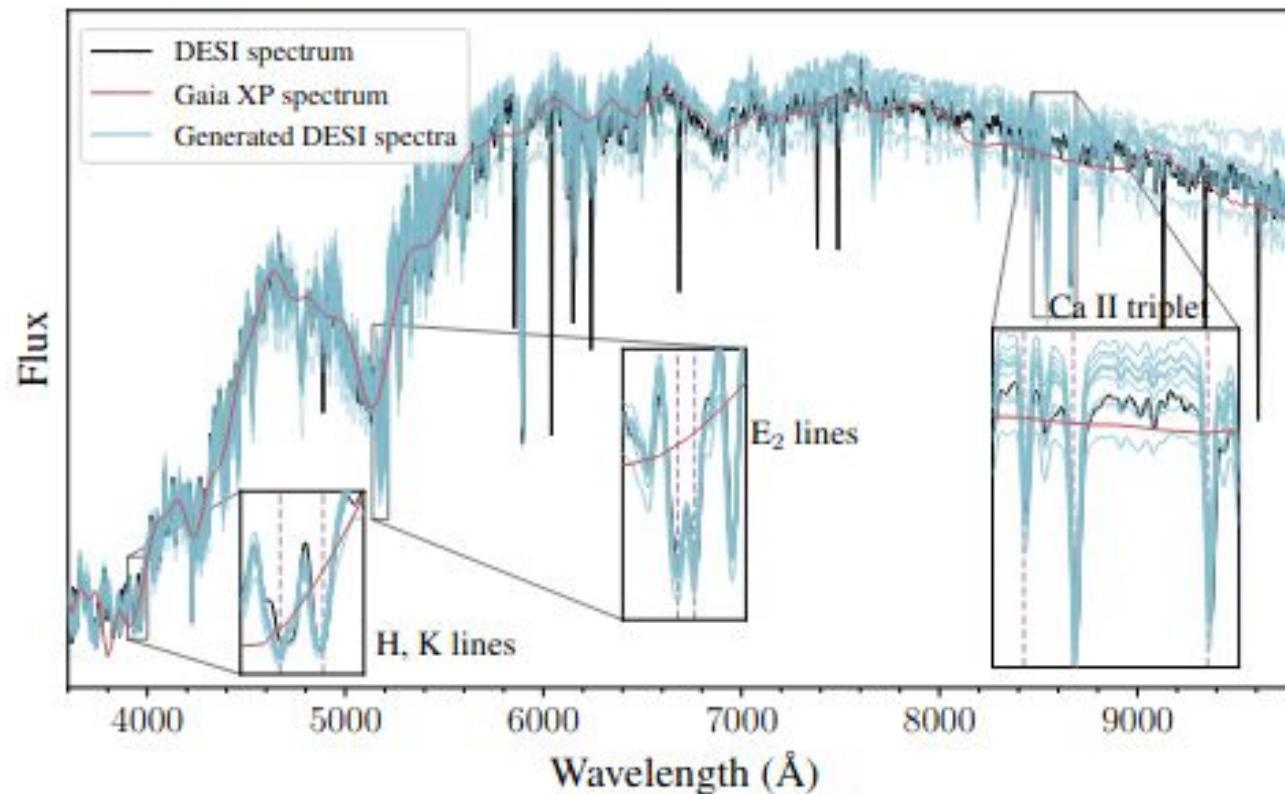
Machine learning

Astro - applications





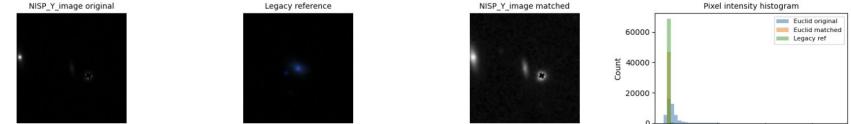
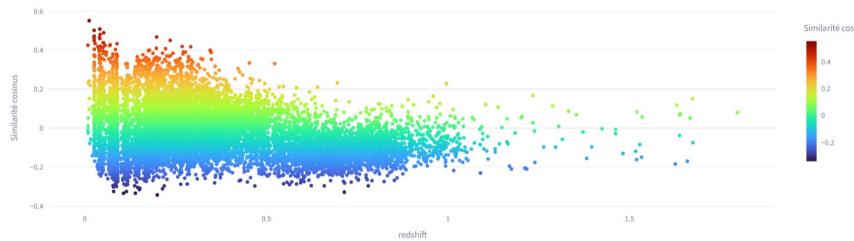
Cross modality : Spectral Super-Resolution



Training: AstroCLIP

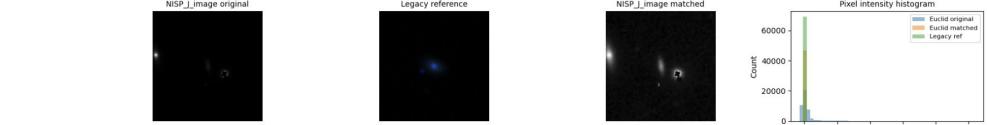
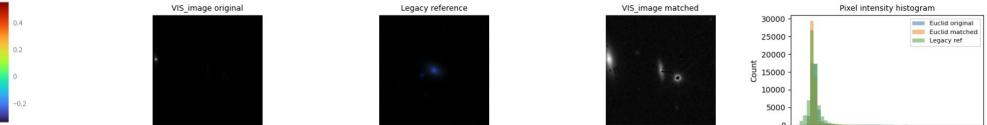
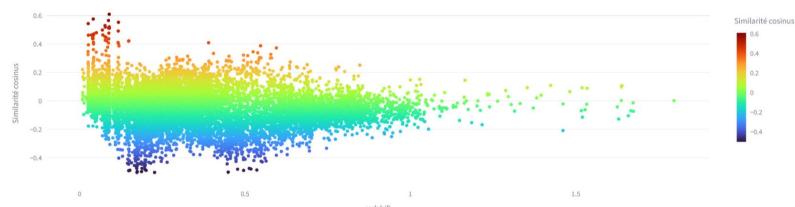
With fine-tuning

Similarité vs Redshift

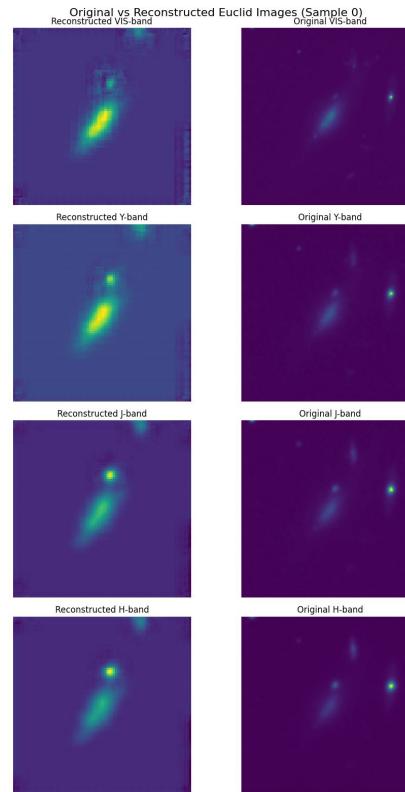
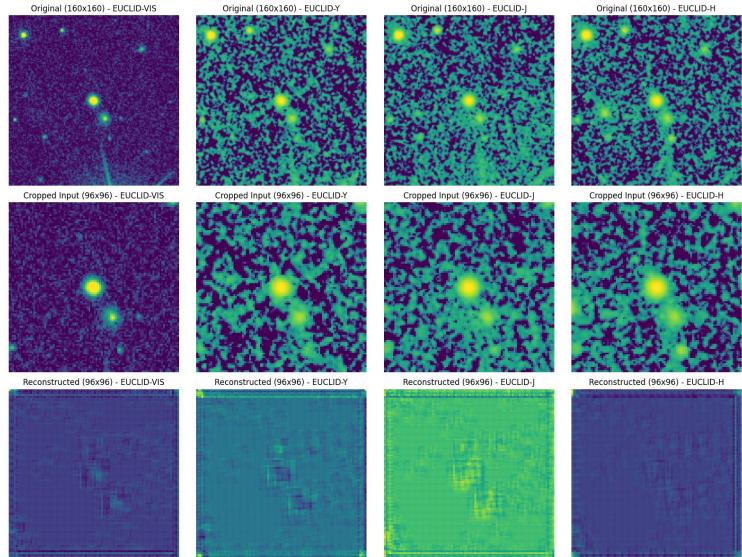


Without fine-tuning

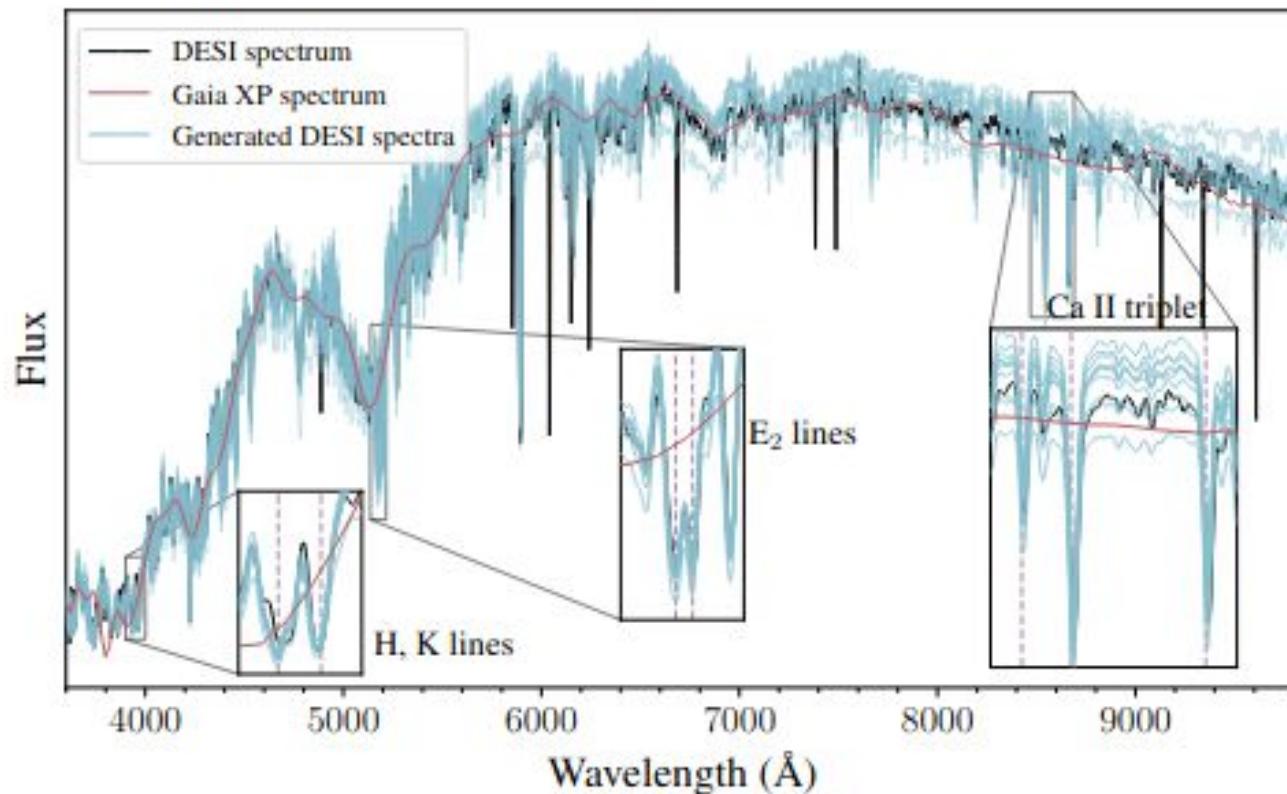
Similarité vs Redshift



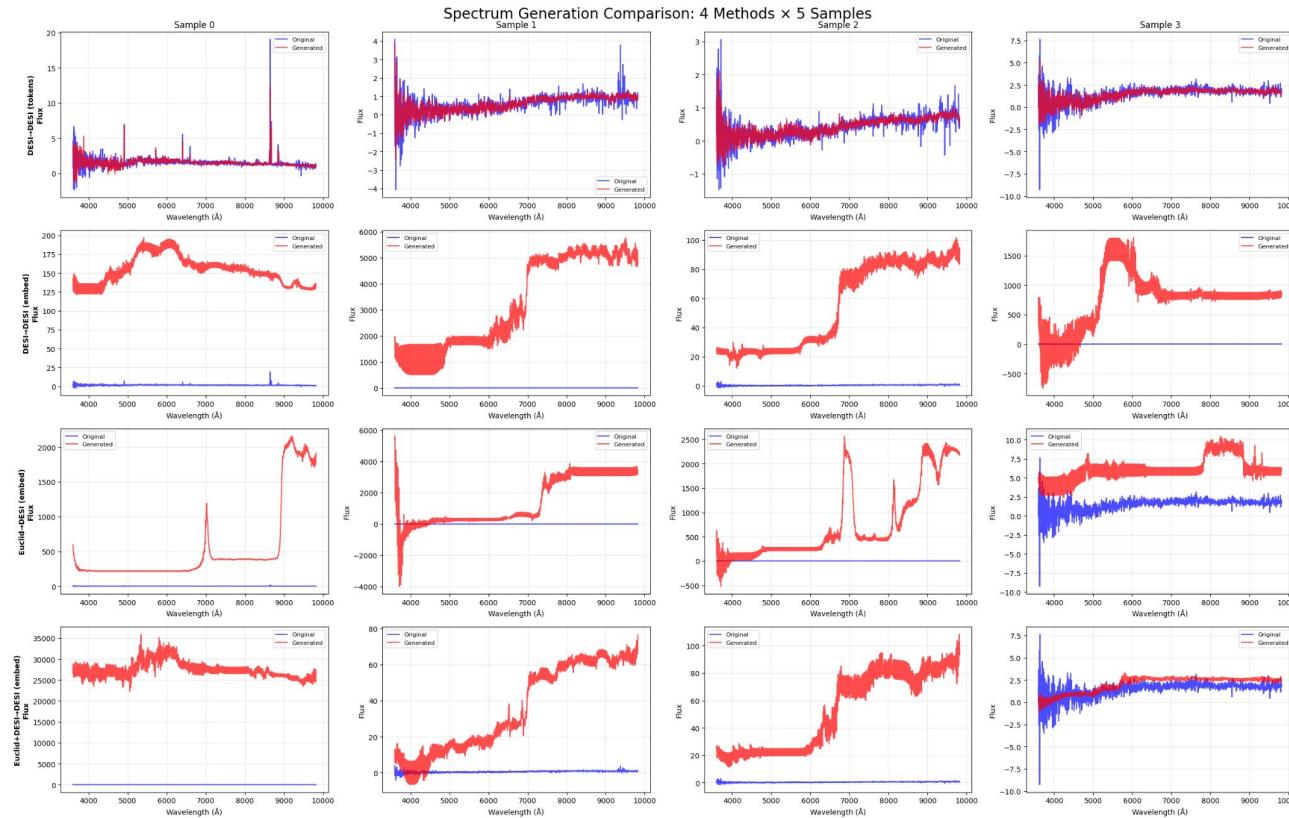
Retraining Encoder



Cross modality : the dream

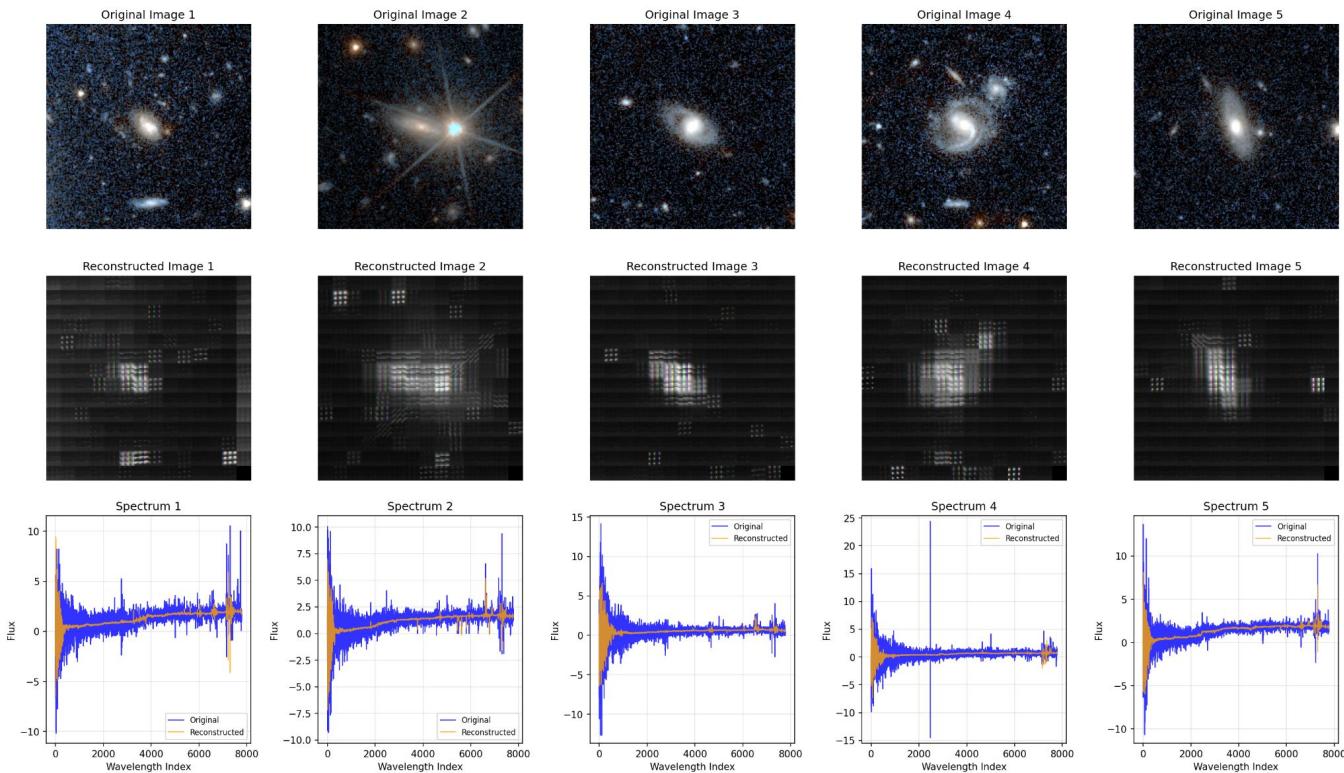


Cross modality : the nightmare

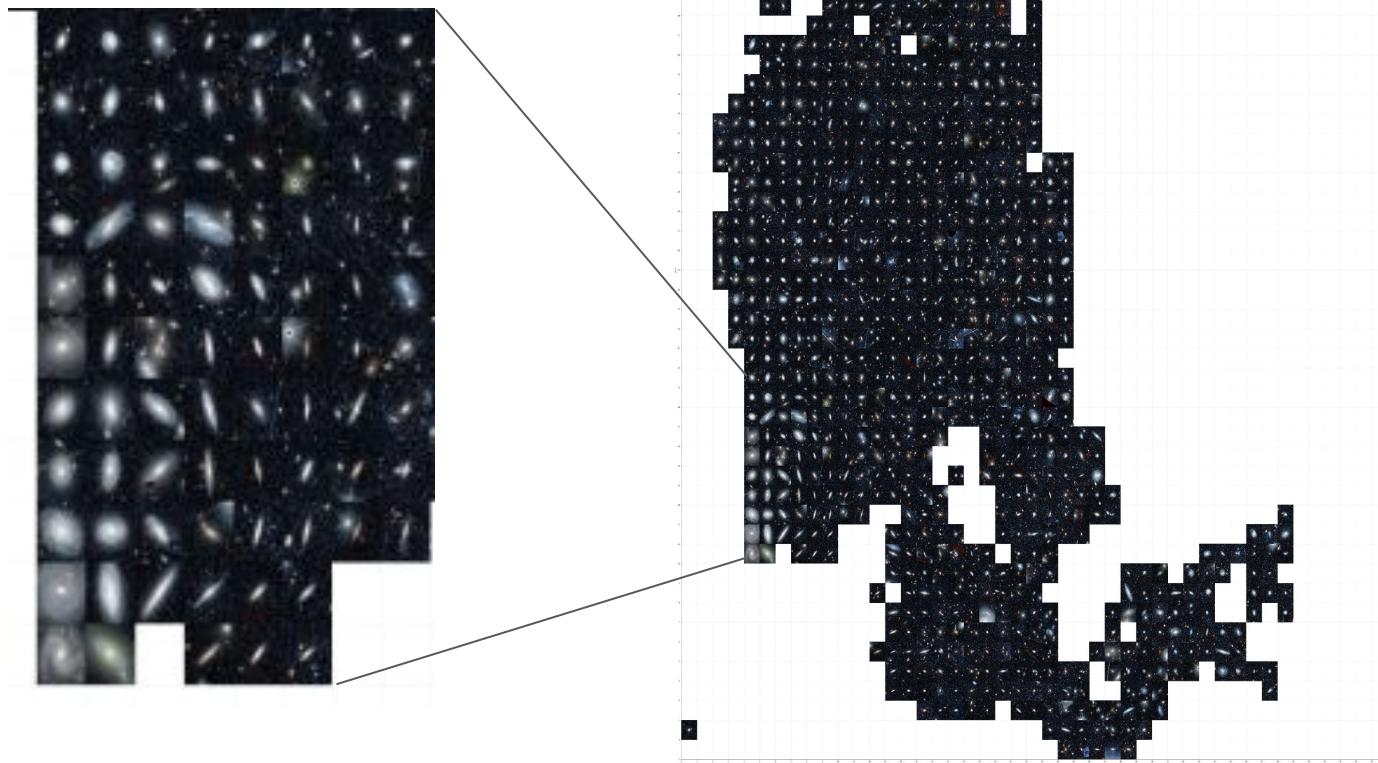


Training: AstroPT

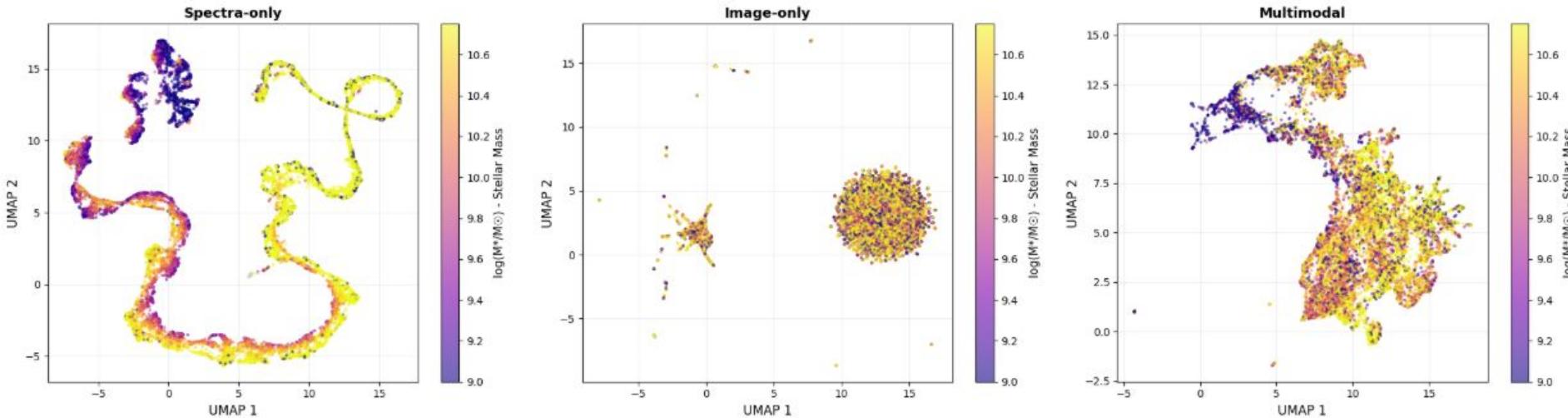
Reconstructions at Iteration 20000



UMAPs

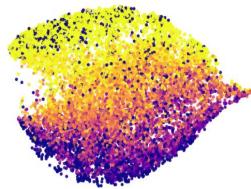


UMAP: correlation across modalities

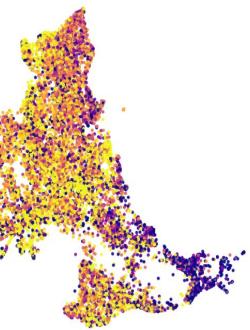


astroPT

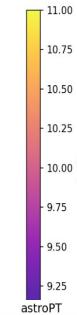
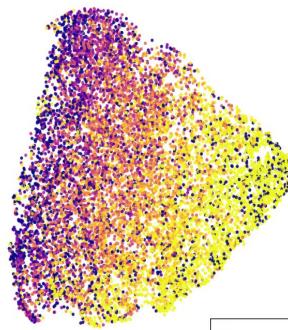
logM



AION



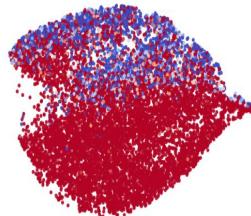
AstroCLIP



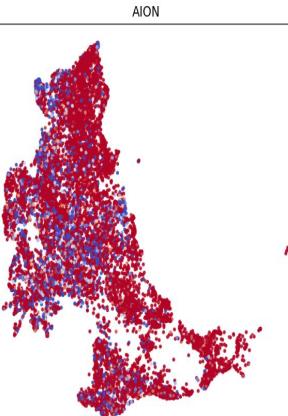
UMAP: correlations with galaxy properties (spectra driven)

astroPT

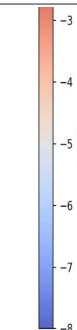
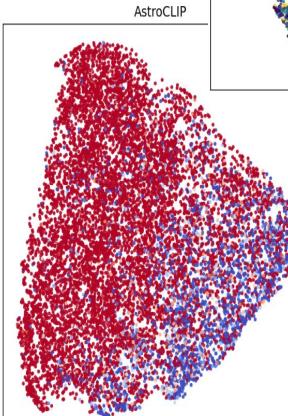
logSFR



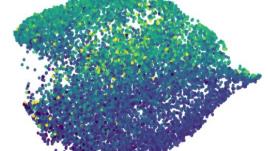
AION



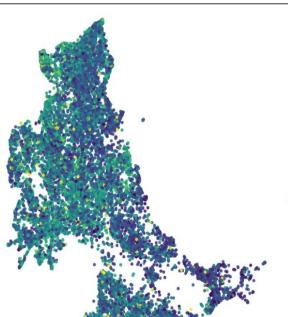
AstroCLIP



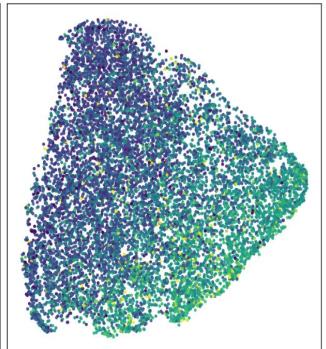
AGE



AION

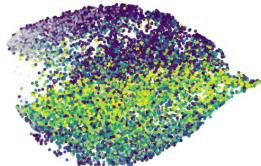


AstroCLIP

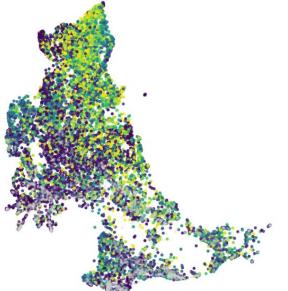


astroPT

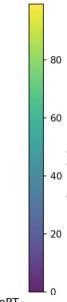
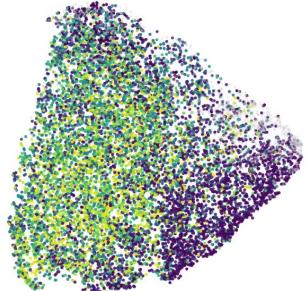
Has spiral
arms



AION



AstroCLIP

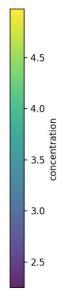
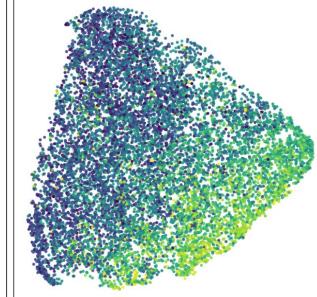


UMAP: correlations with galaxy properties (image driven)

AION



AstroCLIP

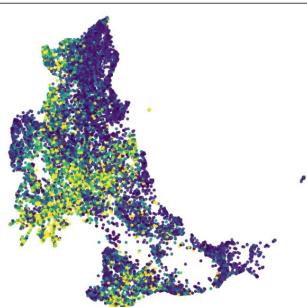


Sersic index

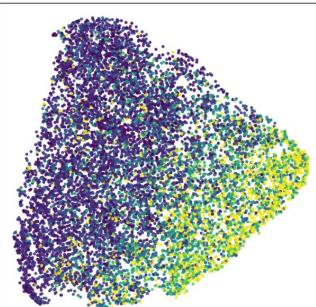
astroPT



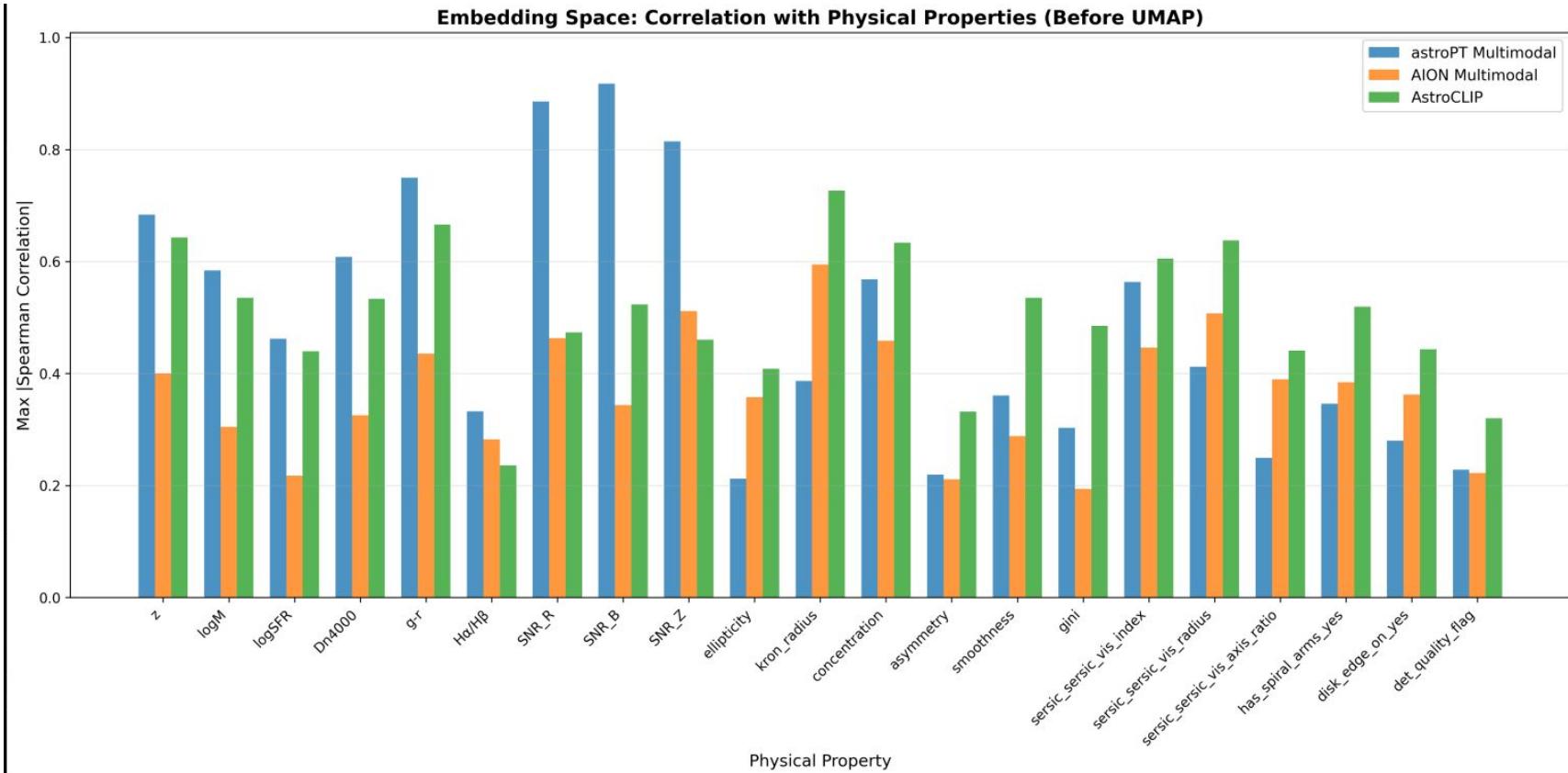
AION



AstroCLIP



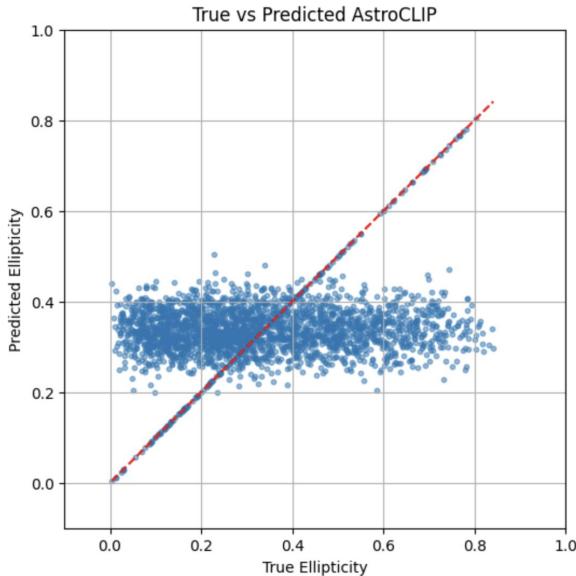
Embeddings: correlations with galaxy properties (score)



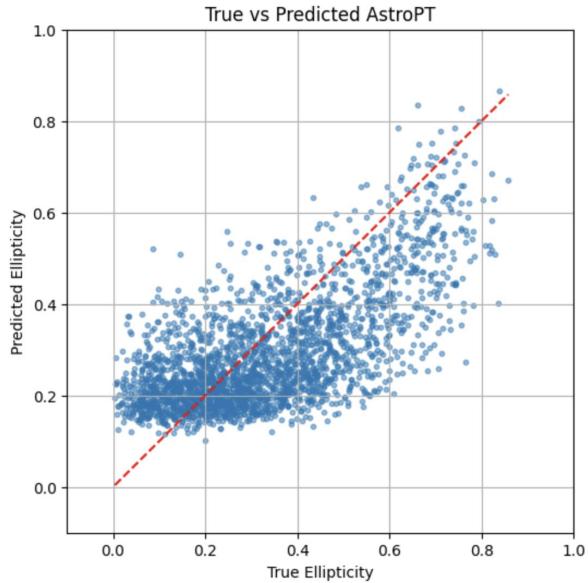
Forecasting features

Regression to predict the **Ellipticity** for each model

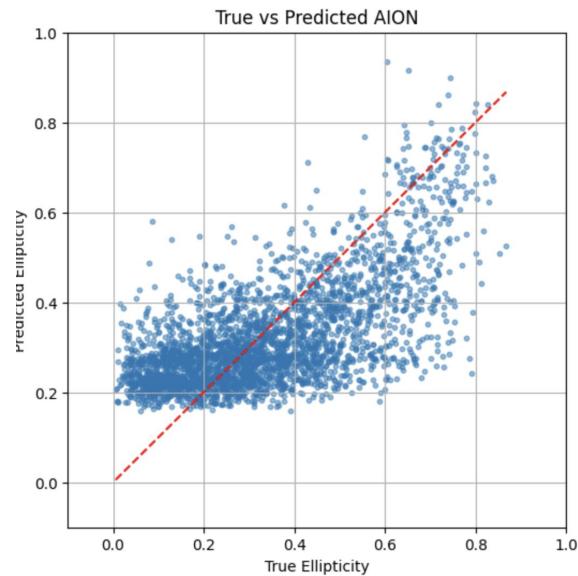
AstroCLIP



AstroPT



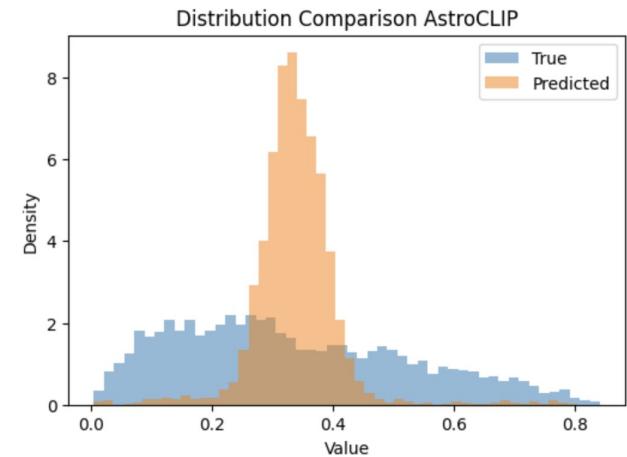
AION



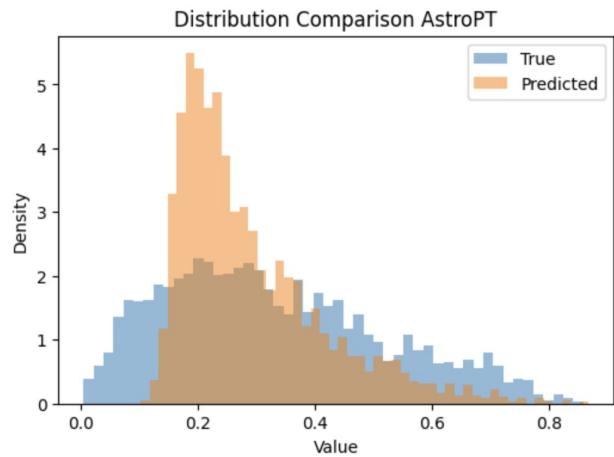
Forecasting features

Regression to predict the **Ellipticity** for each model

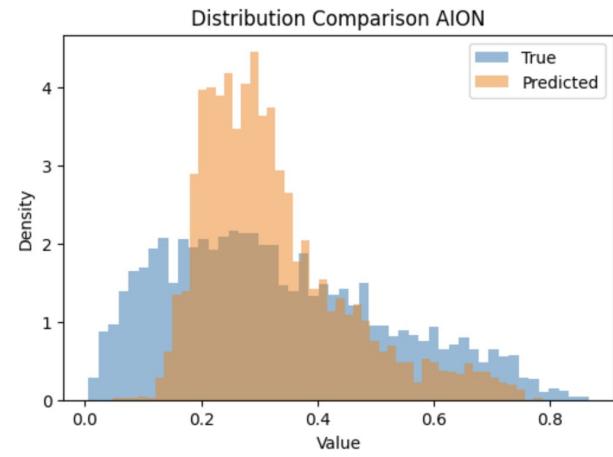
AstroCLIP



AstroPT

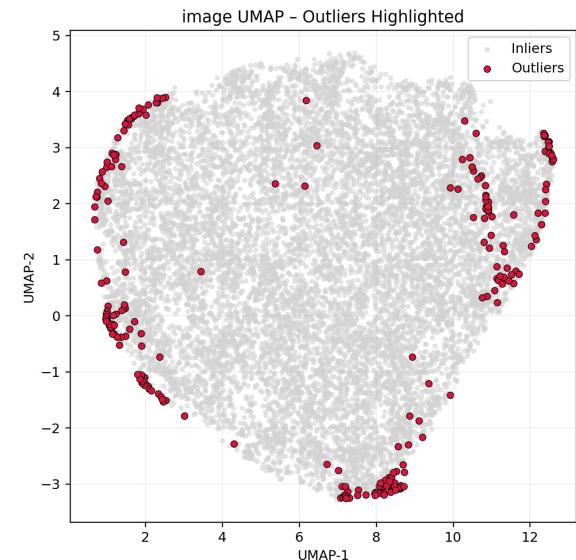


AION

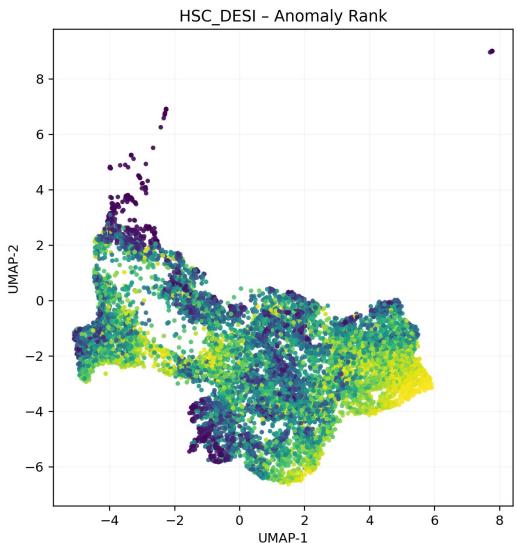


Anomaly detection

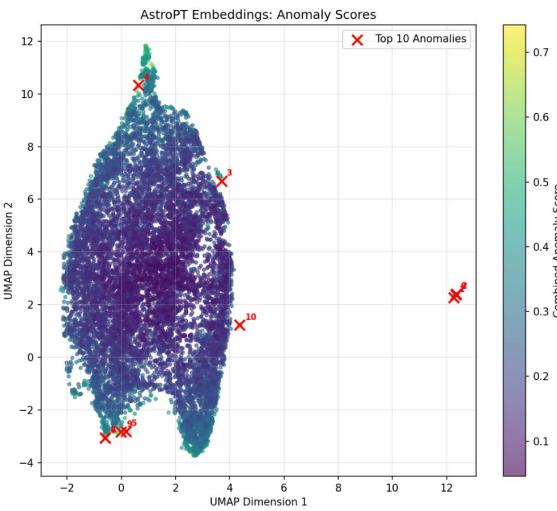
AstroCLIP



AION



AstroPT

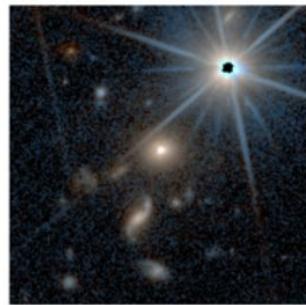


Some anomalies (AION)

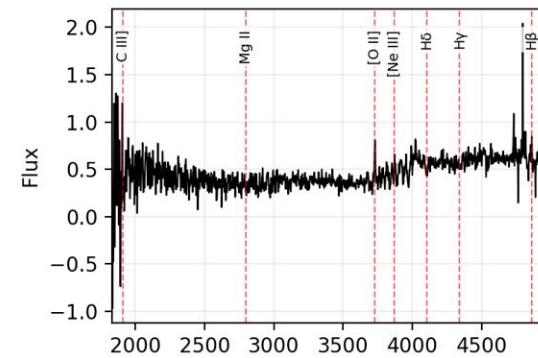
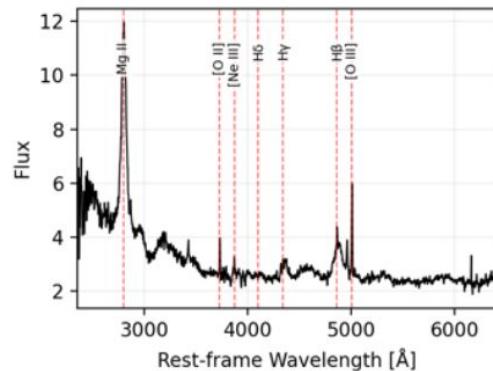
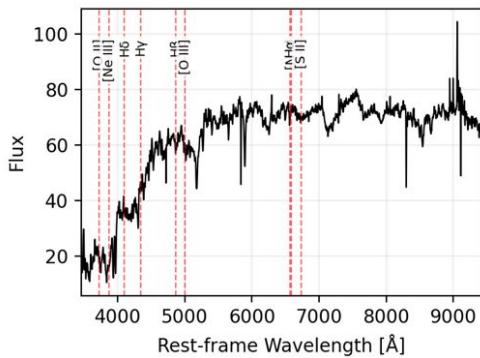
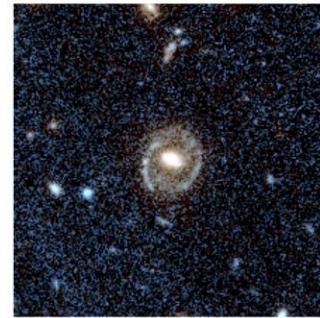
Rest-frame Wavelength [Å]
2677813028655307424 ($z=0.042$)



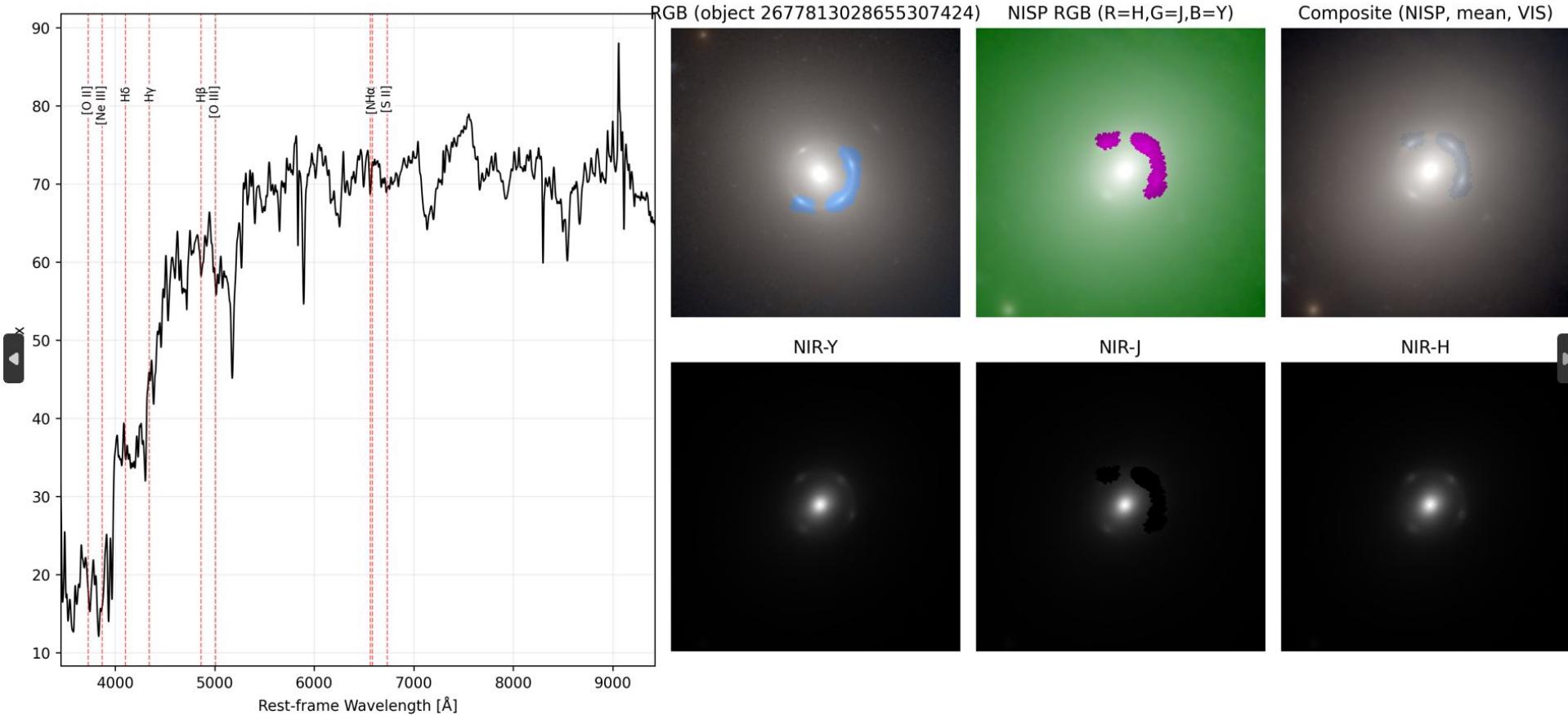
2688459062653158253 ($z=0.532$)



2682983371638126599 ($z=0.969$)



Some anomalies (AION)



Some anomalies (AstroPT)

2717133723668032087 ($z=0.026$)



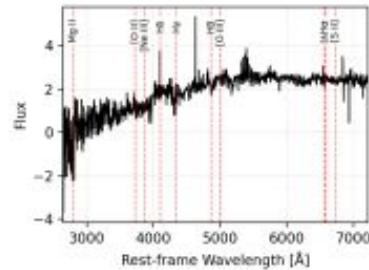
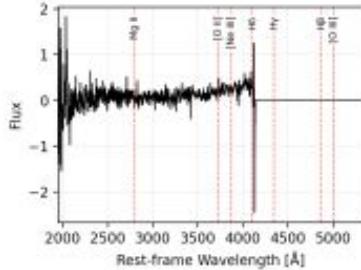
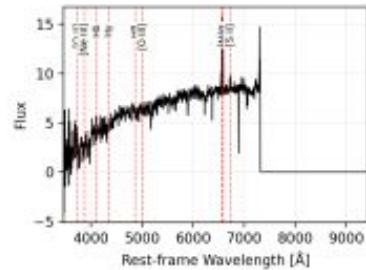
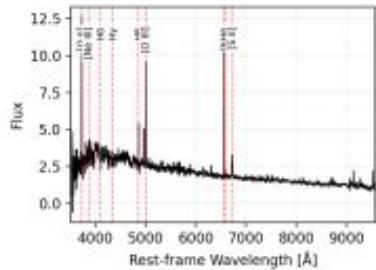
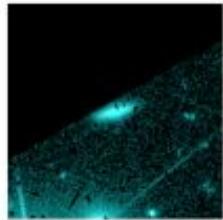
2662565963665897225 ($z=0.041$)



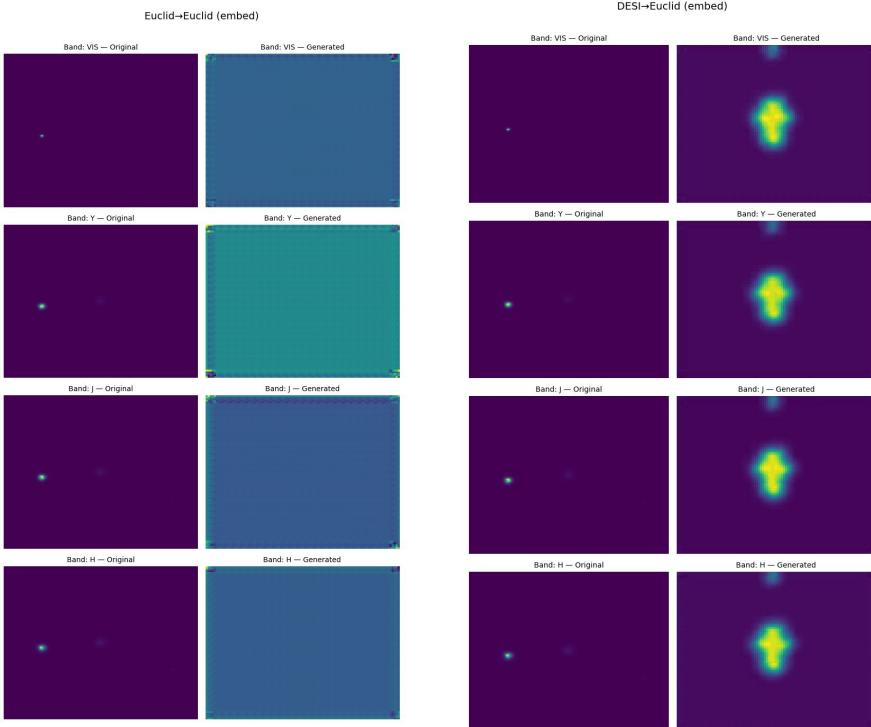
2667323413664460141 ($z=0.843$)



2707818497633976424 ($z=0.361$)



Cross modality : Euclid to Euclid via embeddings



- Using spectra to Images gives likely but incorrect results
- Using Images (that the model has not been trained on) does not work

Take away :

- simply predicting another modality from embeddings does not work
- Foundation model are not generative models

Cross modality : what we should have done

<https://arxiv.org/pdf/2202.04200>

