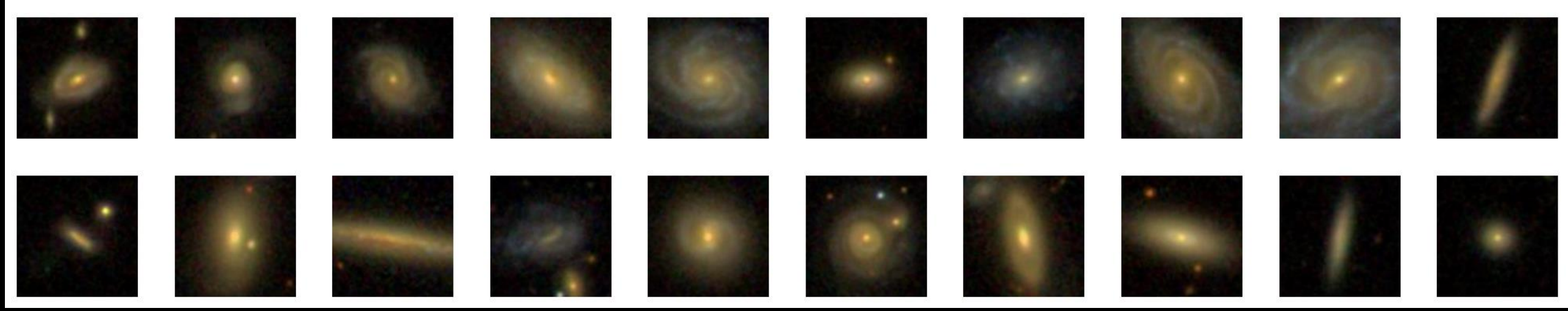


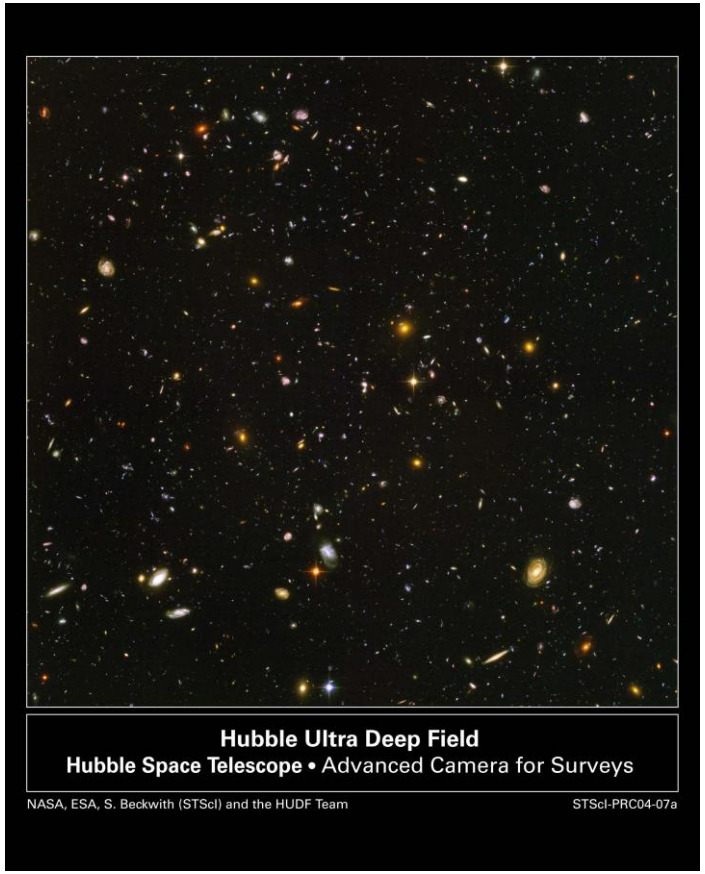
From A to B: AI applications on galaxy images

Shiyin Shen

Shanghai Astronomical Observatory

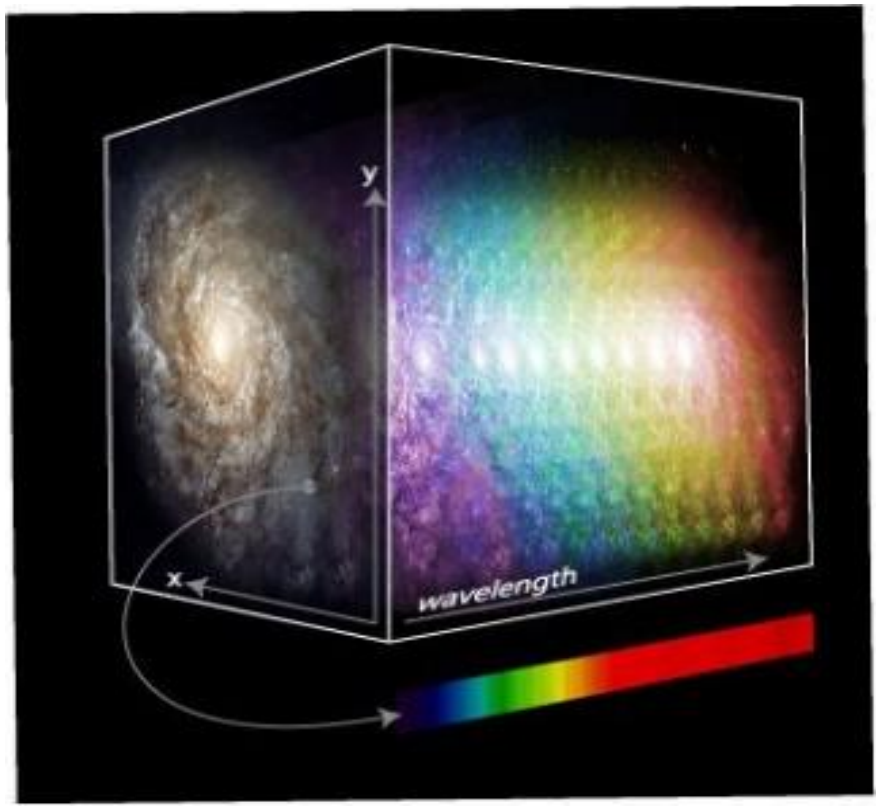
2025/10/21

galaxies：cosmic island



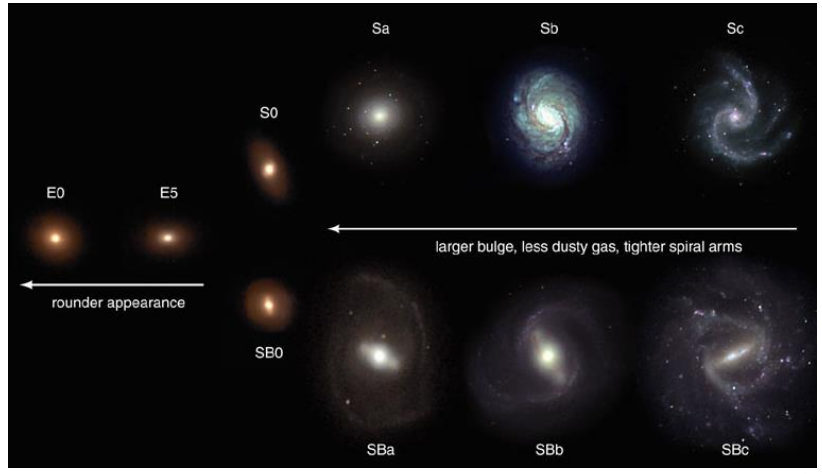
Images of galaxies

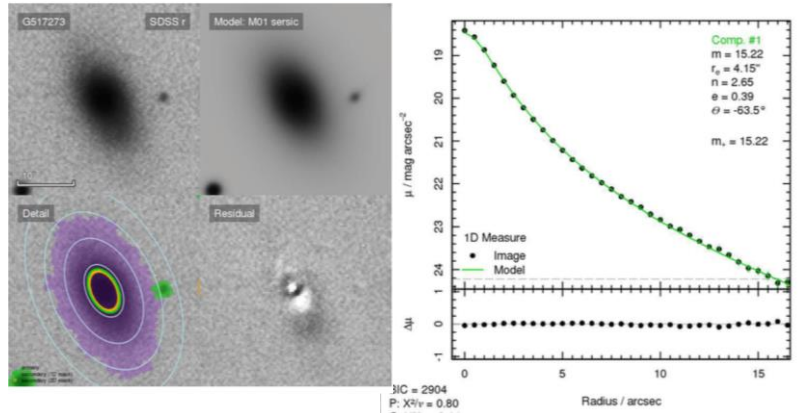
- projection of galaxy struture
  - (3d → 2d)
- energy band dependent
  - radiation transfer process
  - different componens have different emissions
  - Multi-band images/integral-field spectroscopy
- Astronomy: From 2d → 3d



Galaxy morpholog: classical routines

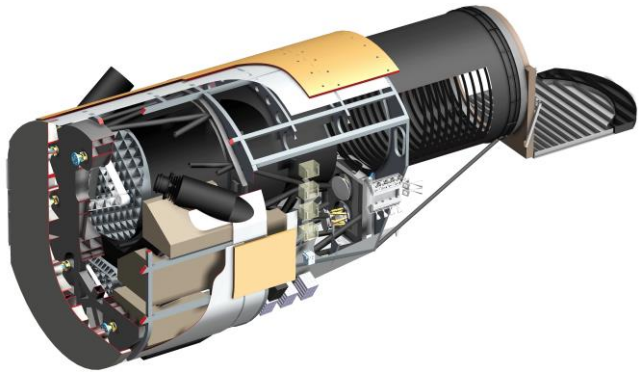
- Morphology classification
  - AI alomost solves everthing~
- Surface brightness profile fitting
  - physical informative (e.g. Sersic profile)
  - Too idealistic, computational-intensive
    - **GALMOSS (our GPU version Galfit)**
- Non-parameteric modelling： morphometry
  - Concentration,Asymmetry,Smoothness,Gini
  - depends on observational effect

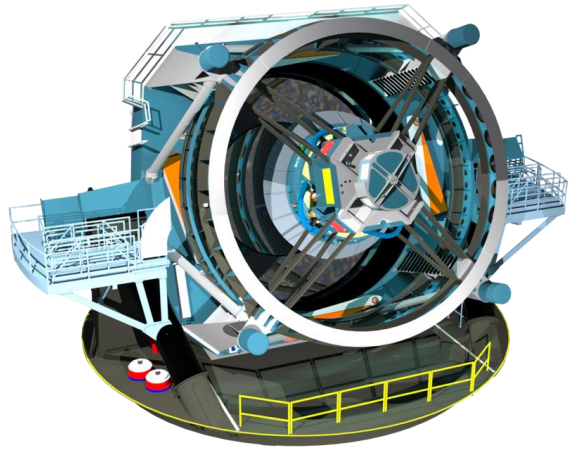




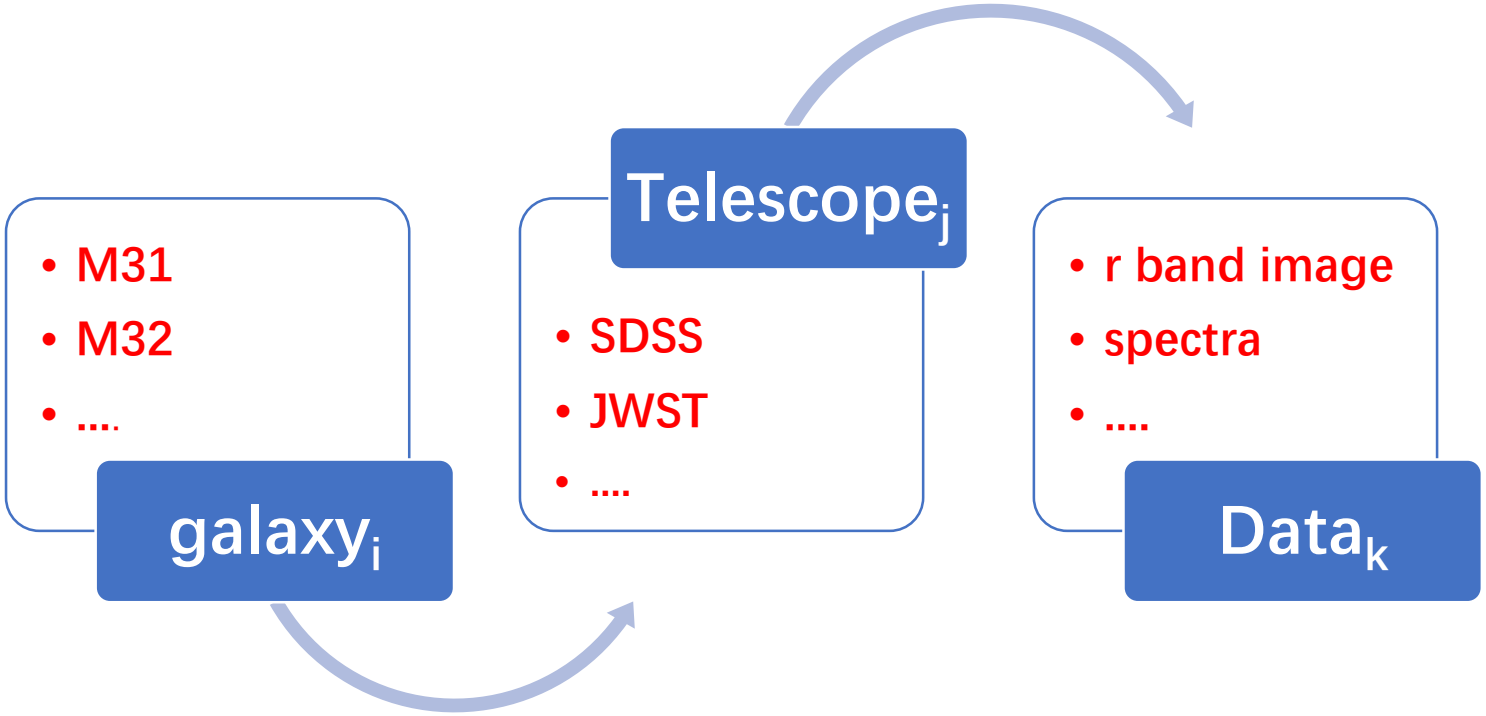
galaxy image：observational effect

- Instrument
  - resolution
  - point spread function
- Background
  - ground based VS space telescope
- Exposure depth
  - shallow VS deep exposure
  - survey VS pointing mode





Galactic astronomy: methodology

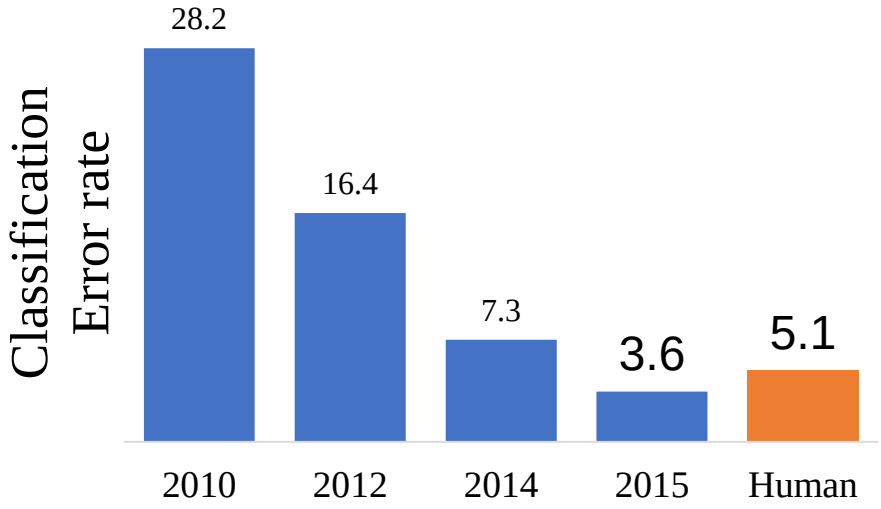


Astronomers always need more observations $D_{i,j,k}$

Universe does not evolve on human time-scale + cosomological principle

redunt information kept in exisiting data: elephant probably already inside the room~

galaxy images meet AI



2015, AI works better than human on classification

2024, Nobel Prizes

2027, CSST

2024, Euclid

2023, ChatGPT

2023, WFST

2022, Galaxy Zoo DECaLS, **260,000 galaxies**

2019, DESI Legacy Survey, **>10 million galaxies with z**

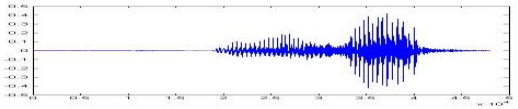
2008, Galaxy Zoo, **800,000 galaxies**

2000, SDSS, **~ 1million galaxies with z**

1926, Hubble Sequence, **400 galaxies**

AI ≈ deep learning ≈ neural network

- **Voice to text**

$f($  $)=$ “How are you”

- **Image classification**

$f($  $)=$ “Cat”

Intelligence/learning process

mapping something unfamiliar to something familiar(understandable)

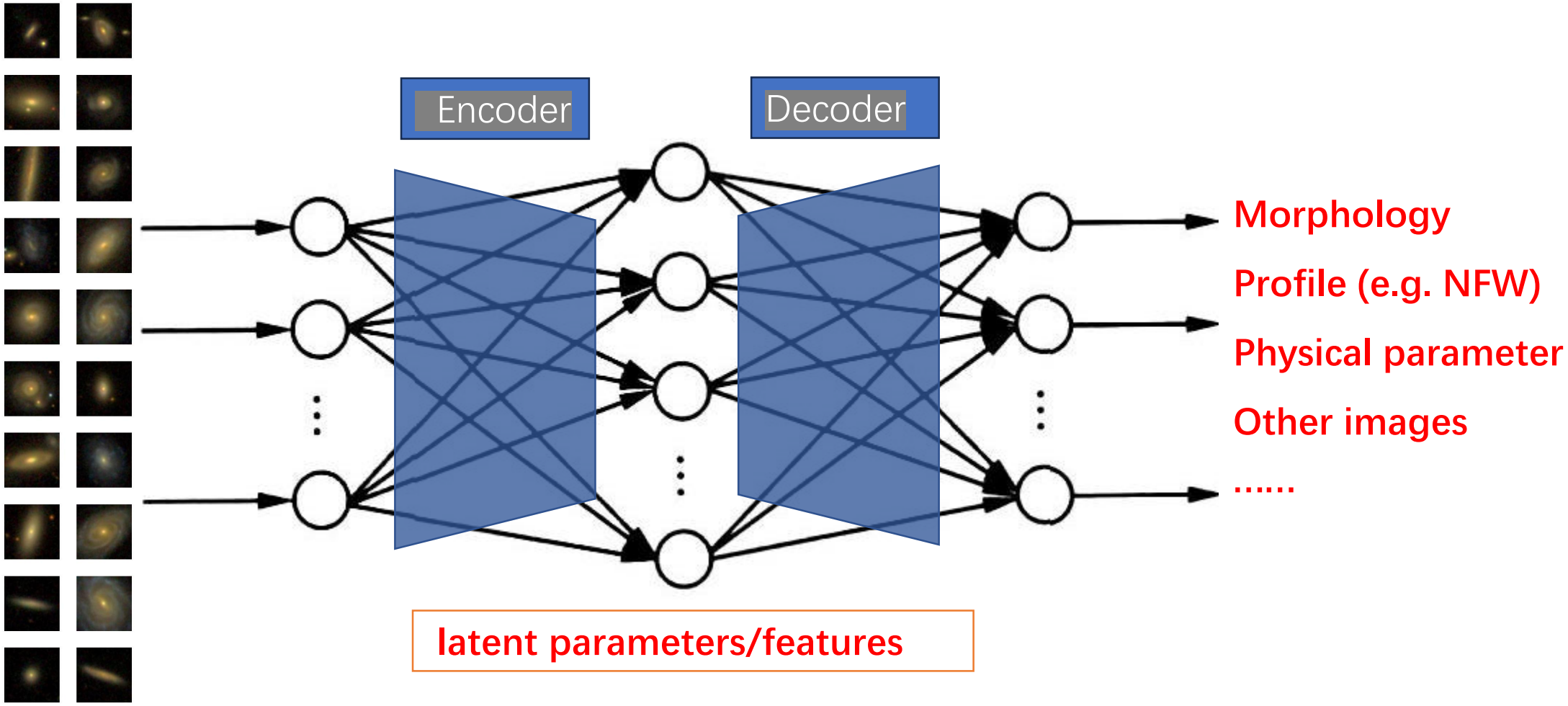
A to B

Can AI understand “galaxy” ?

“Turing Test” in Galactic Astronomy:

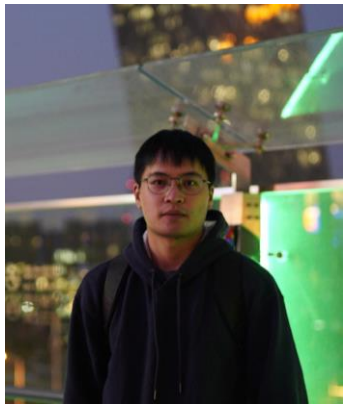
mappings among $Galaxy_i/Telescope_j/Data_k$?

Architecture of a neural network in computer vision



From A to B: our AI applications for galaxy images

- **From images to prameters/features ...** (Xu et al. 2023, Chen et al. 2024)
- **From parameters to images** (Xu et al. 2025, in prep)
- **From Survey A to Survey B** (Ye et al., 2024)
- **From images to images** (a large galaxy image model proposal)

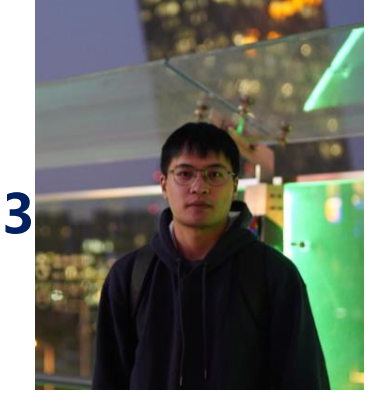

Quangfeng Xu


Renhao Ye


Mi Chen


Shiyuan Chen

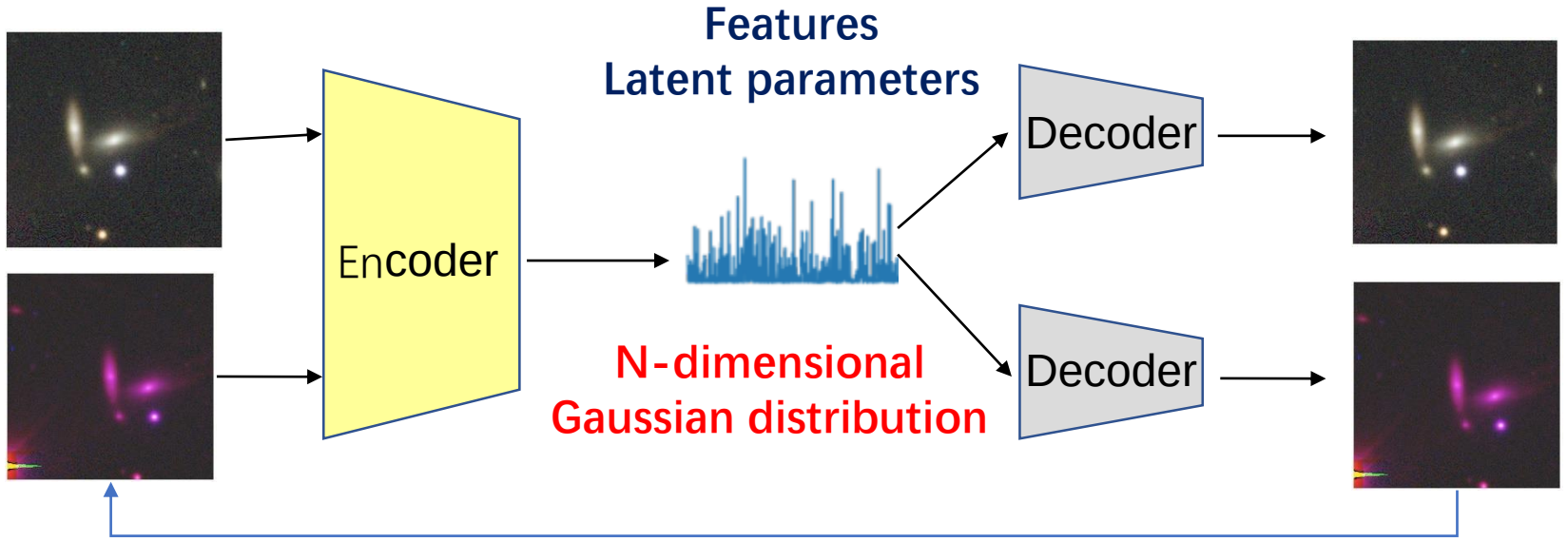
From images to features(Xu,Shen+ et al. 2023



Quangfeng Xu
徐权锋

Compression is Intelligence!

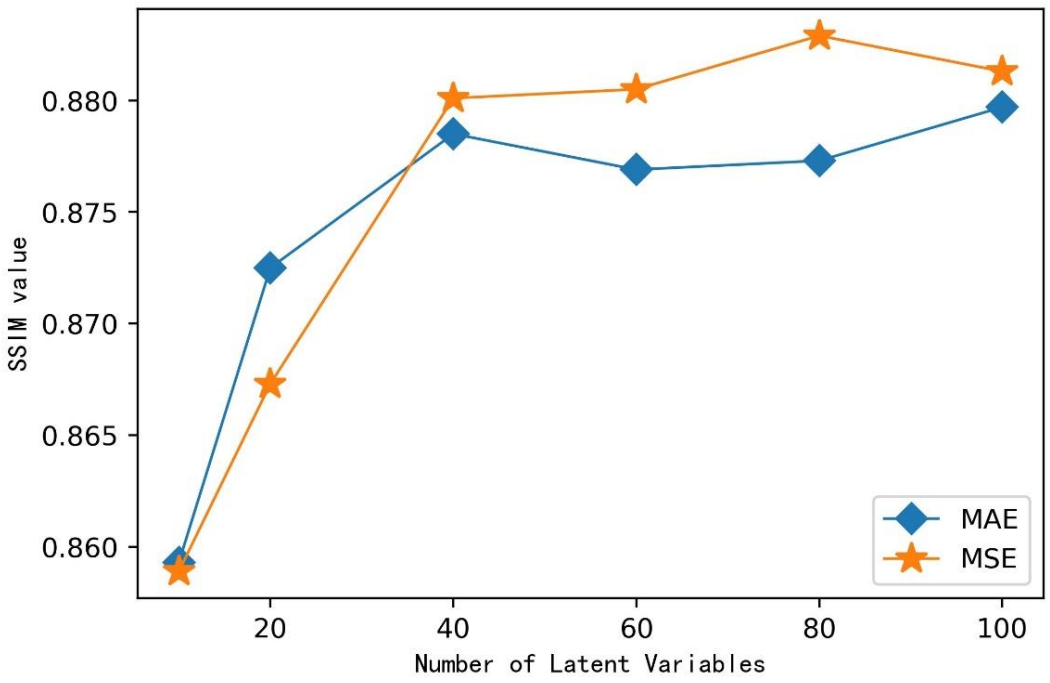
Variational auto-encoder



Q1：How many features (latent parameters) can effectively represent/compress a galaxy image?

Q2：what is the usage of the latent parameters?

Number of features of a DESI galaxy image (128*128)



N ~40

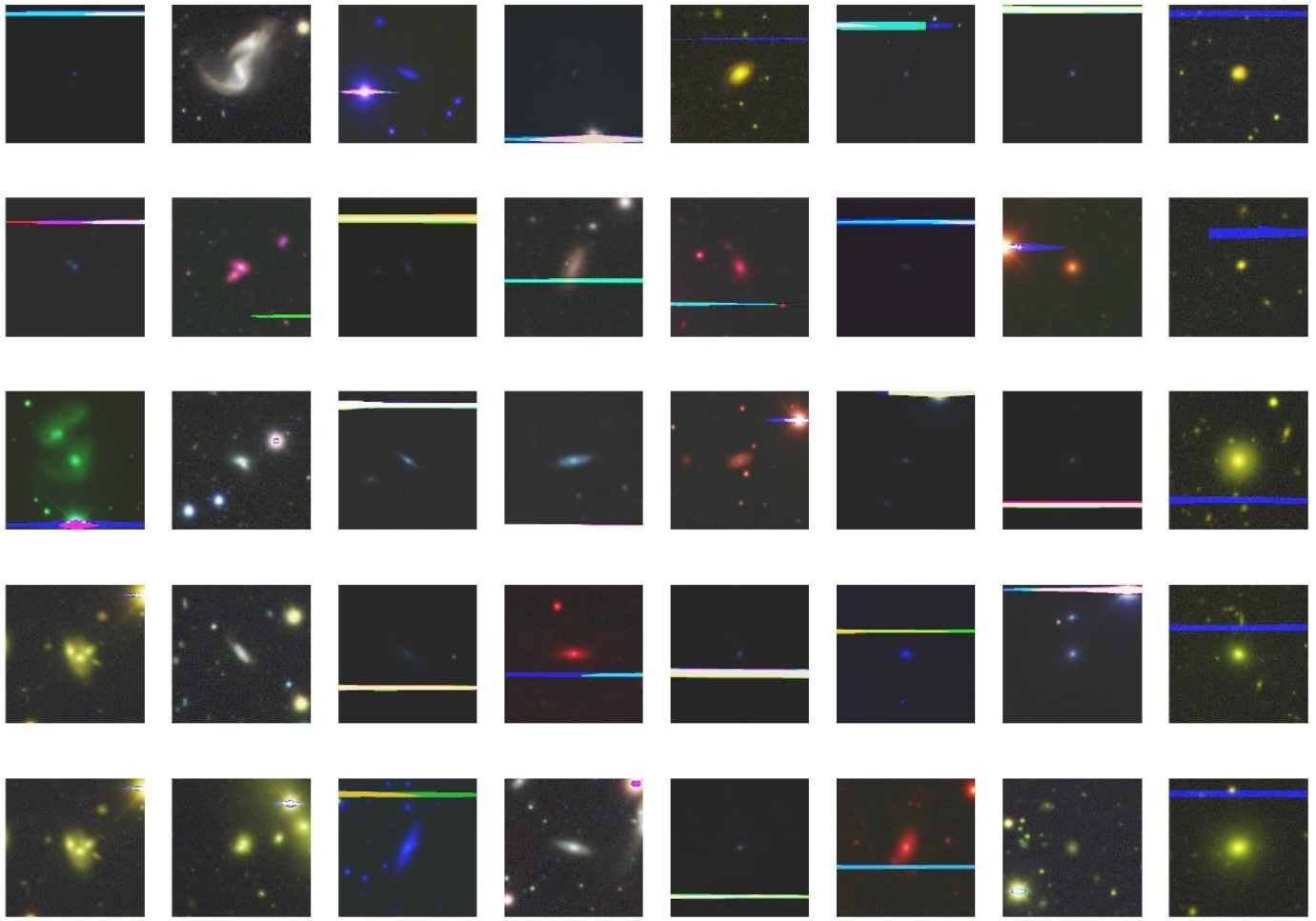
Morphology classification with 40 latent parameters:
Random Forest：Fast and effective

Feature	Classes	Test acc	Precision	Recall	F1
How round	3	90.68 %	90.68 %	90.68 %	90.68 %
Edge-on	2	84.25 %	98.78 %	84.43 %	91.04 %
Bulge Shape	3	84.08 %	84.08 %	84.08 %	84.08 %
Bar	3	76.21 %	76.21 %	76.21 %	76.21 %
Have Arm	2	91.91 %	73.38 %	96.68 %	83.44 %
Arm Tightness	3	75.84 %	75.84 %	75.84 %	75.84 %
Arm Count	4	86.32 %	86.32 %	86.32 %	86.32 %
Bulge Size	4	82.60 %	82.60 %	82.60 %	82.60 %
Merger	2	96.36 %	99.90 %	96.45 %	98.15 %

Comparable to morphology classification on images

Application of latent parameters: outliers

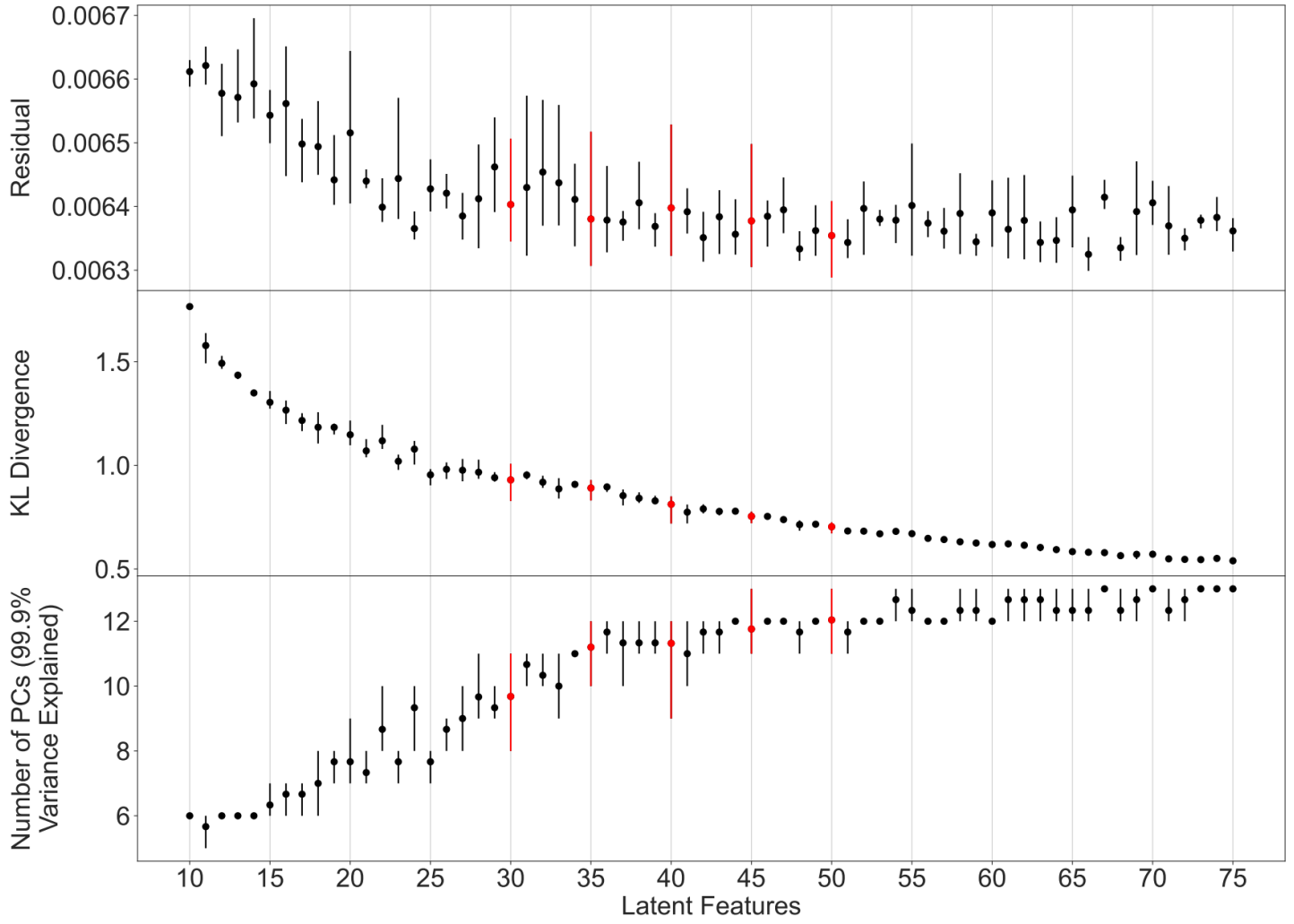
- latent parameters
  - Assumed to follow multi-dimensional Gaussian distribution
- Outlier in latent space
  - >4σ



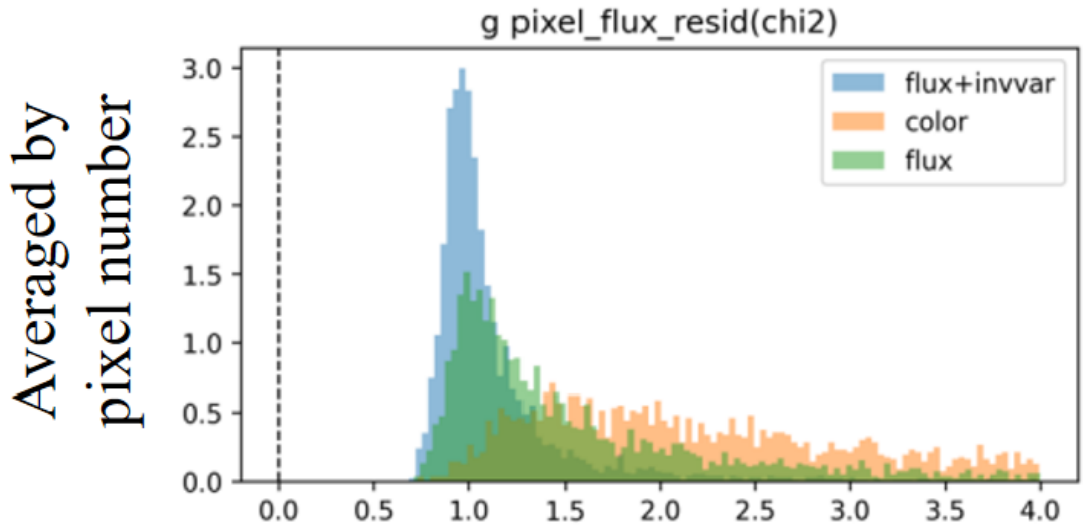
Further dimension reduction?

- arXiv:2509.03640
- PCA analysis on ~40 latent parameter

  n(PC) ~12 （99.9%)

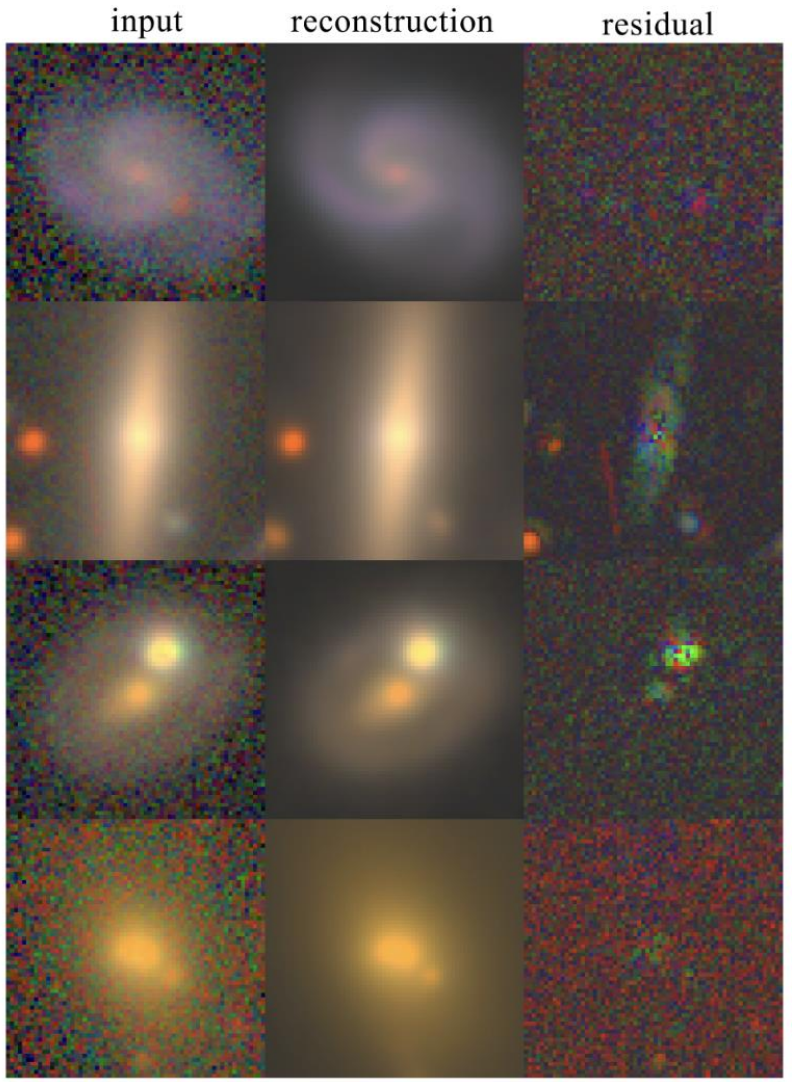


How good the latent parameters represent galaxy images?



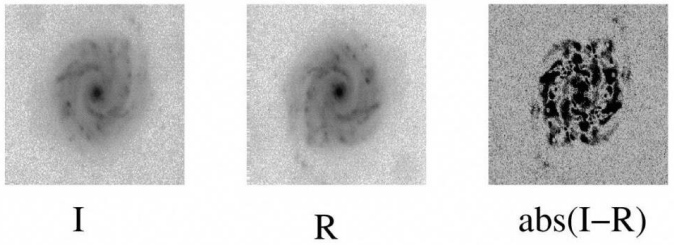
$$\mathcal{L}_{\text{rec}} = \frac{|y - x|}{\sigma^2}$$

observation error

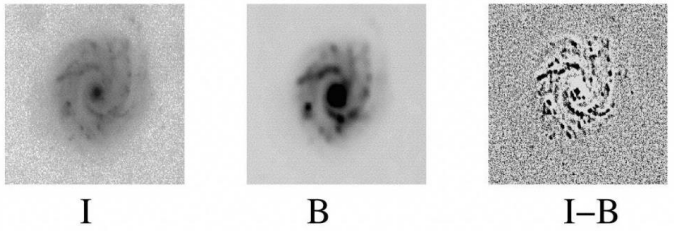


2 Application of latent features: From parameters to images

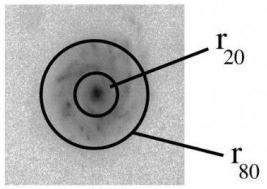
We can not generate galaxy images with a given set of non-parametric metric in a traditional way.



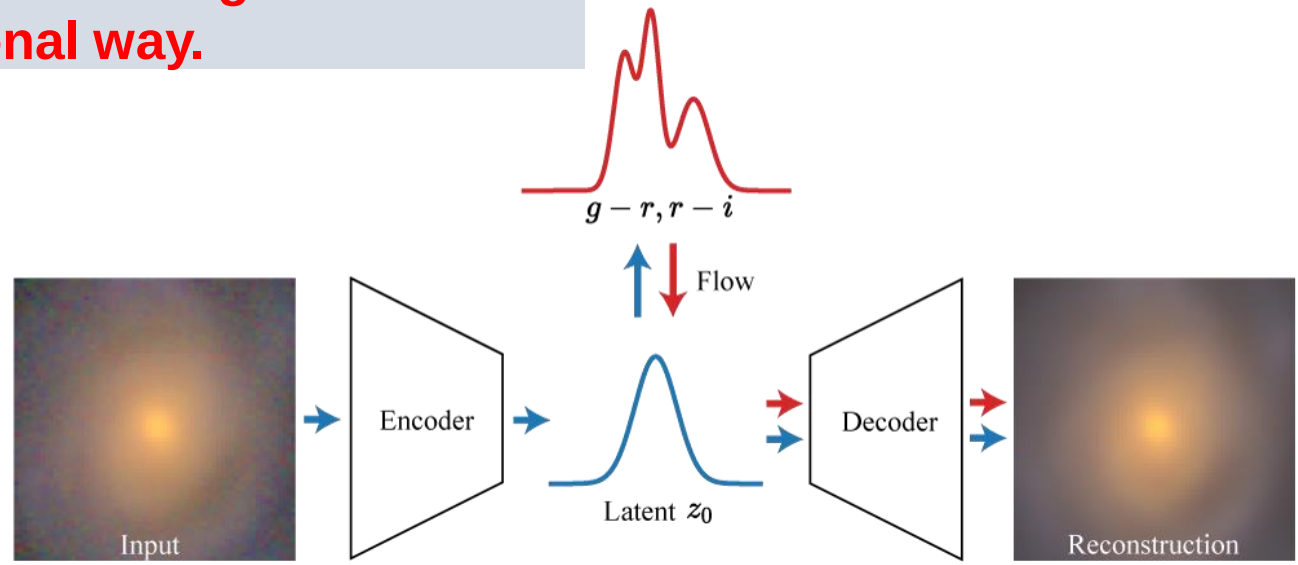
$$A = \frac{\text{abs(I–R)}}{\text{I}}$$

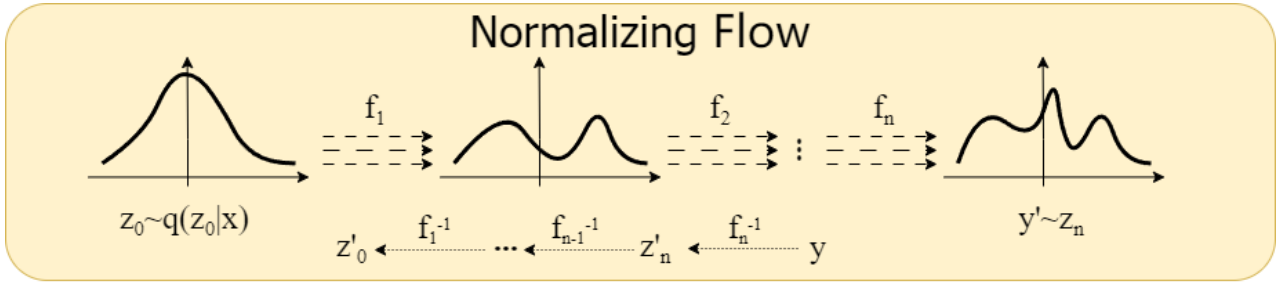


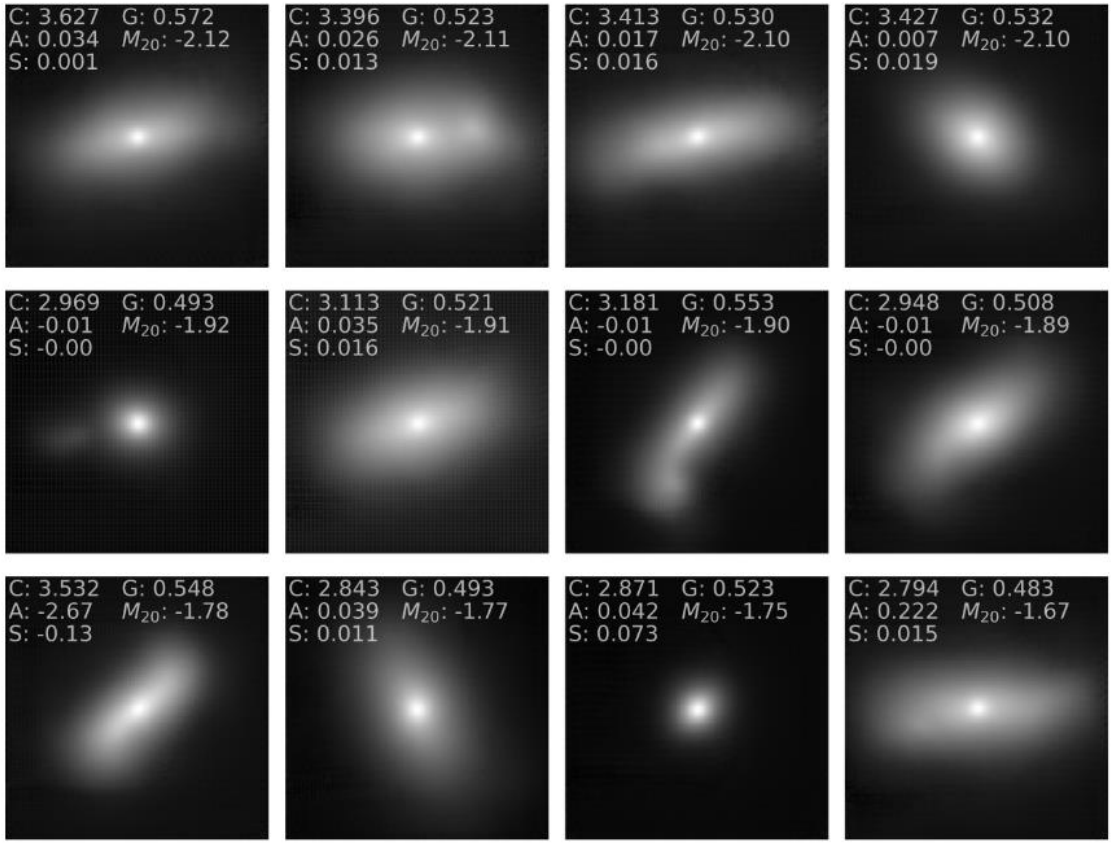
$$S = \frac{\text{I–B}}{\text{I}}$$

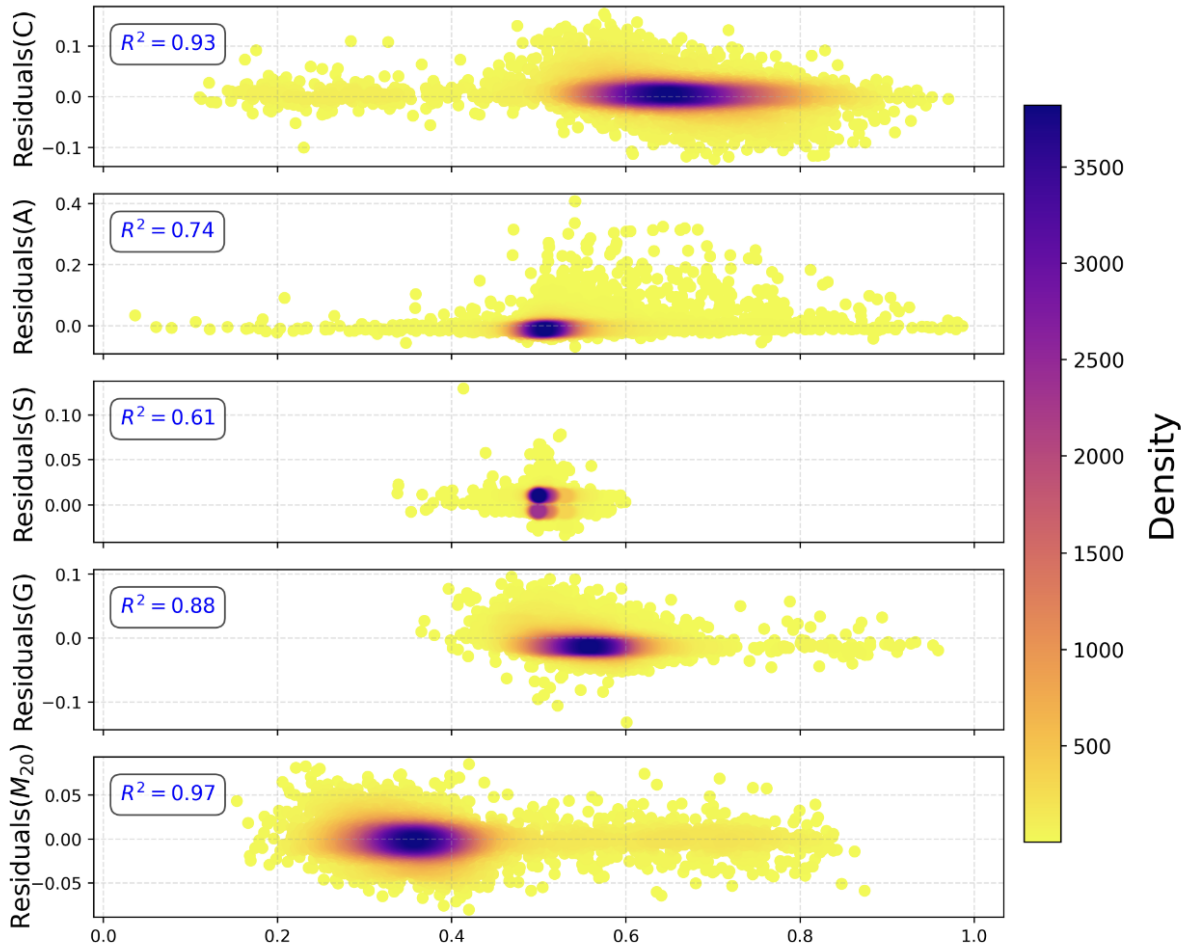


$$C = 5 \log\left(\frac{r_{80}}{r_{20}}\right)$$



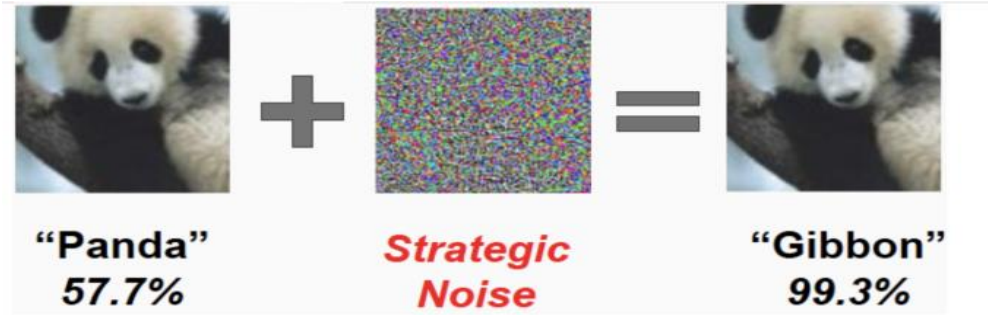






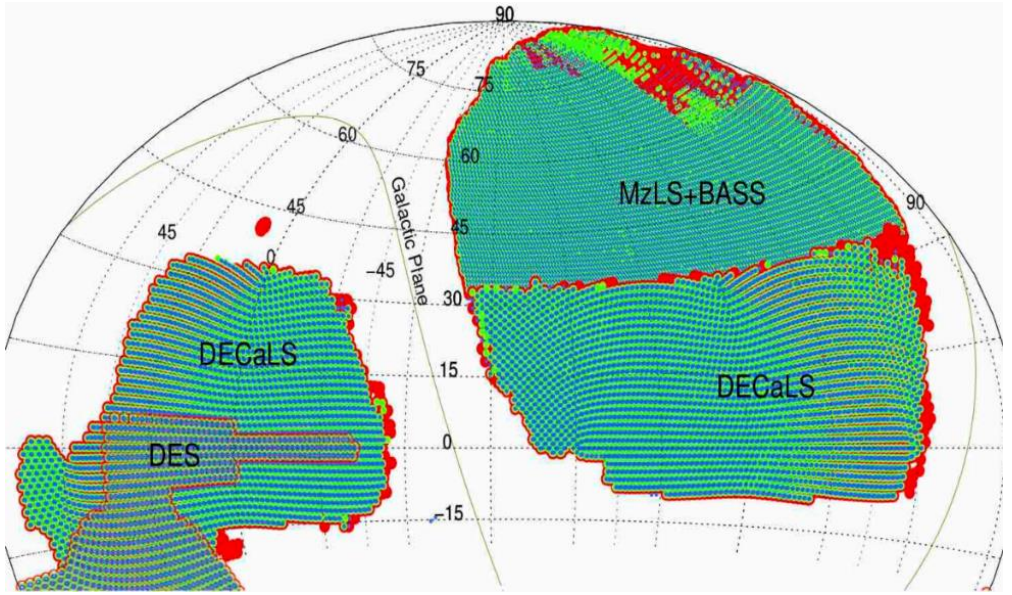
(Xu,Shen et al. 2025, in prep)

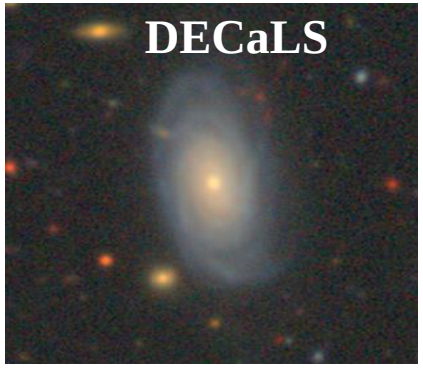
3 From Survey A to Survey B

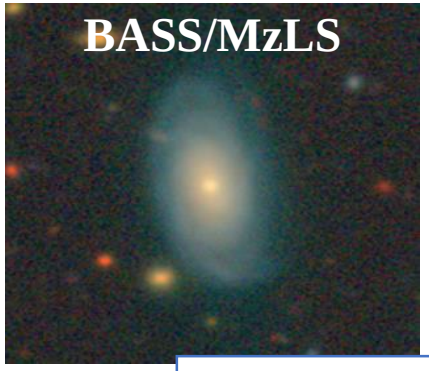


Survey	Instrument	Bands	Area	Pixel Scale
DECaLS	Blanco 4m/DECam	g, r, z	$\sim 9000°$	$0.''262$
BASS	Bok 2.3m/90Prime	g, r	$\sim 5000°$	$0.''454$
MzLS	Mayall 4m/MOSAIC-3	z	$\sim 5000°$	$0.''262$

DECaLS



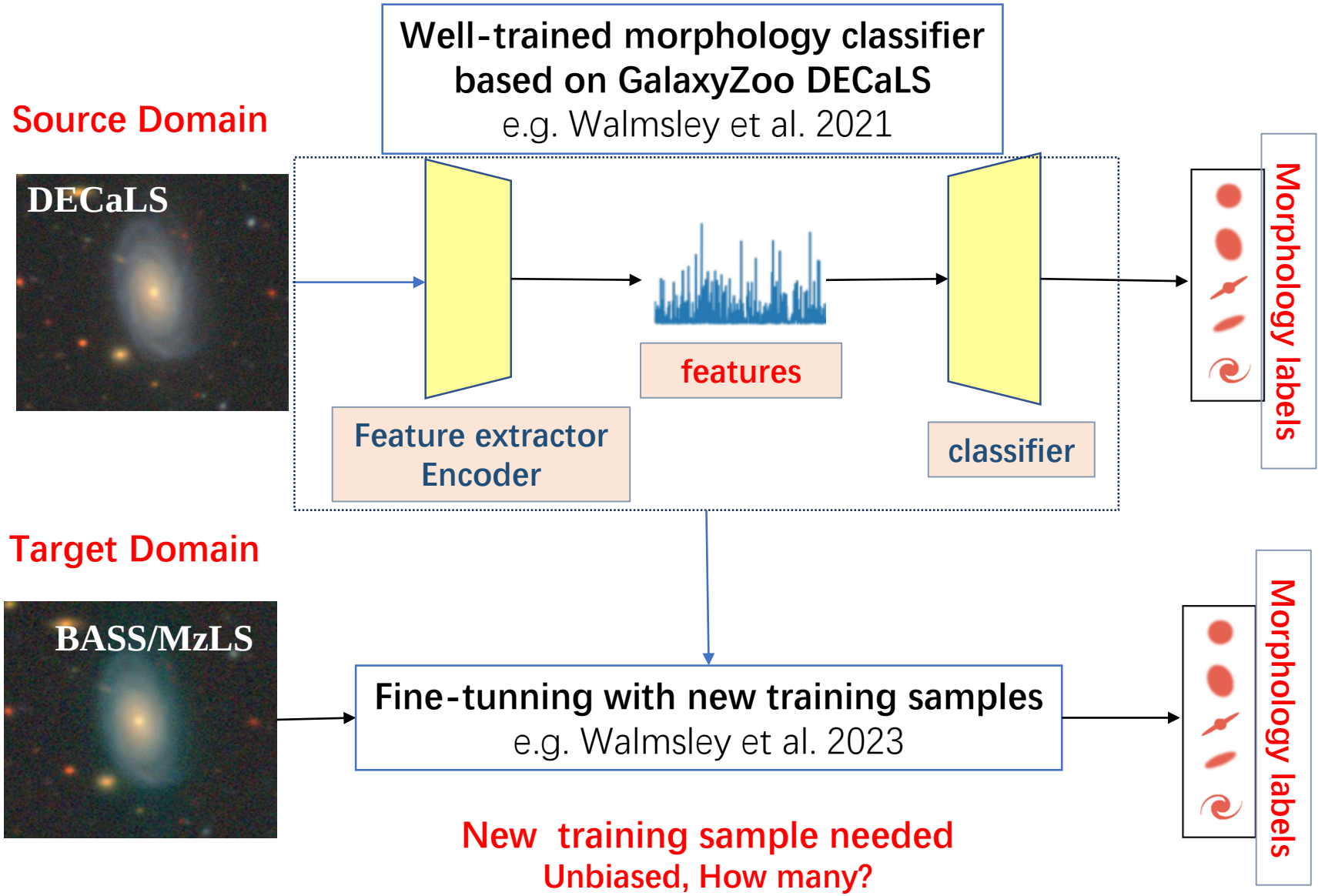




GALAXY ZOO .org

How to transfer AI classification from DECaLS to BASS/MzLS?

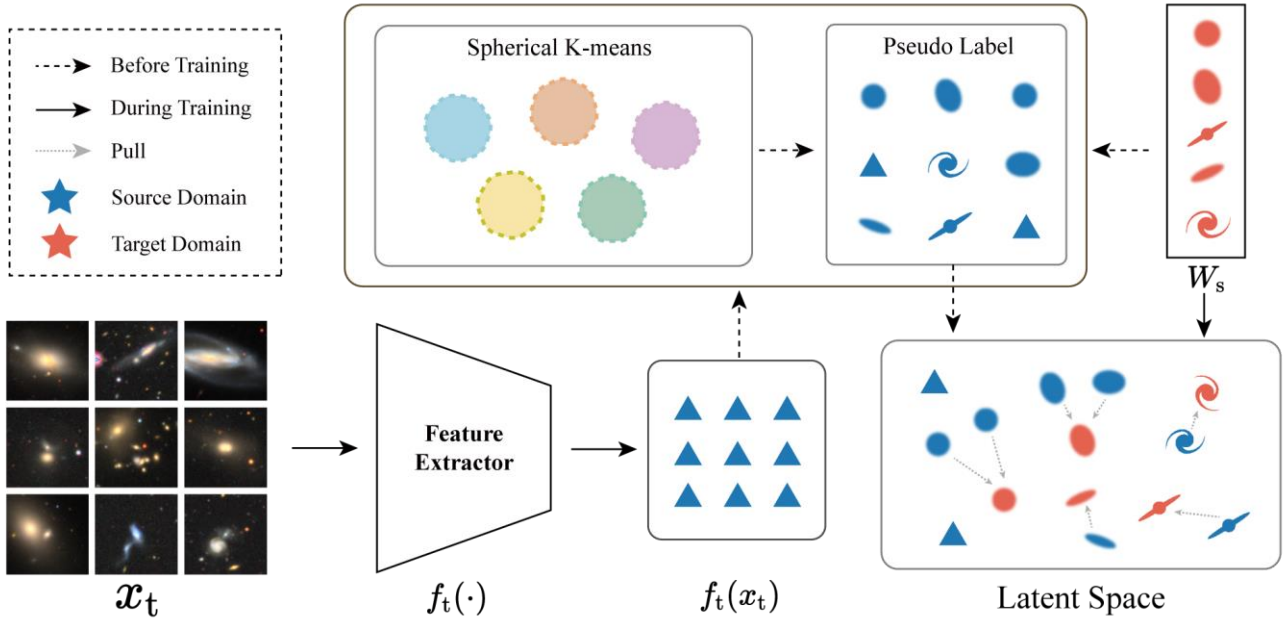
Supervised Domain Adaption



Unsupervised Domain Adaption:
from Galaxy Zoo DECaLS to BASS/MzLS

Renhao Ye
叶人豪

Galaxy images in different surveys should have the same latent parameter distribution, then have the same morphology label distribution.



Ye, Shen+ et al. 2024

Feature	Accuracy
bar	82.1%
DECaLS images	

⬇

Feature	Accuracy
bar	75.9%
BASS/MzLS images	

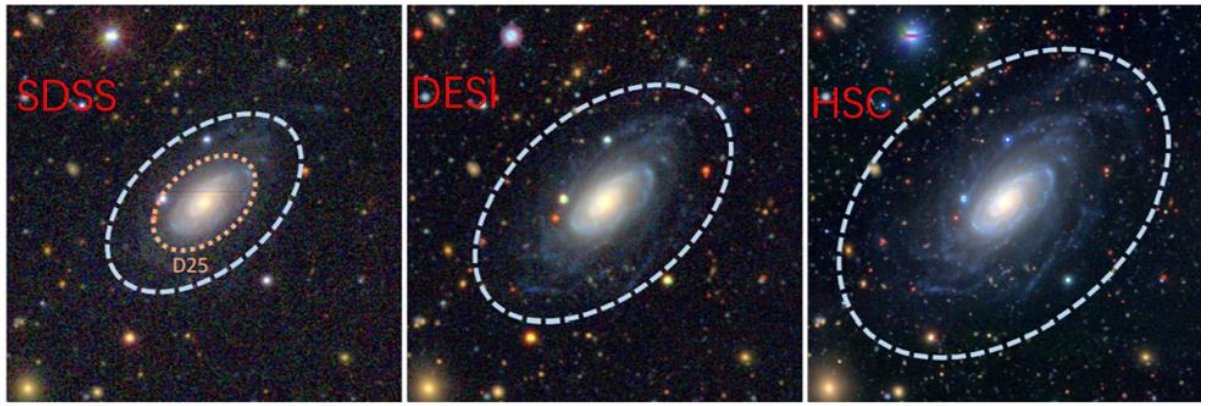
Unsupervised Domain Adaption

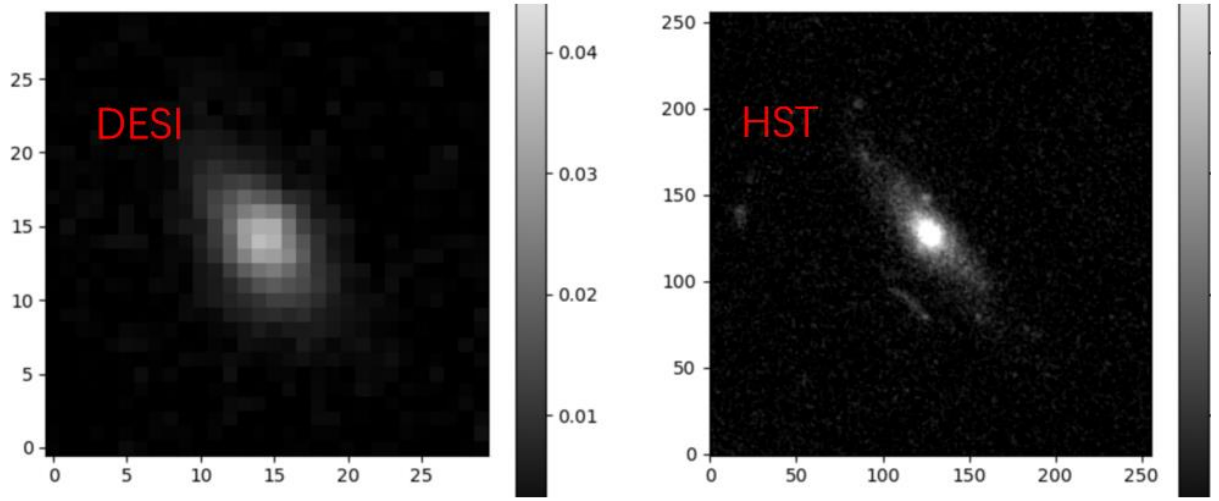
⬇

Feature	Accuracy
bar	**83.4%**
BASS/MzLS images	

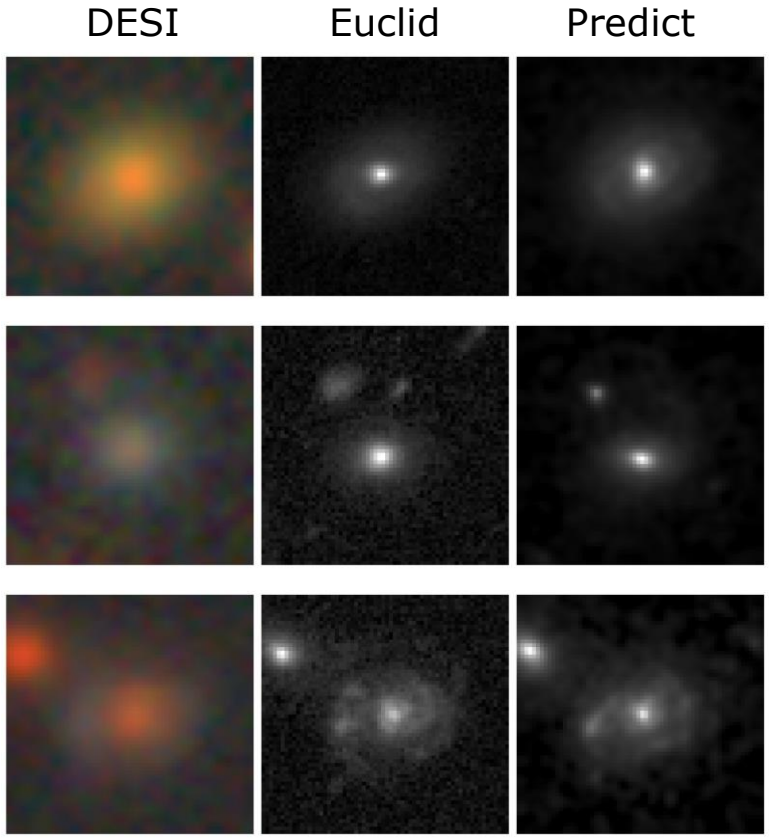
4 From images to images:

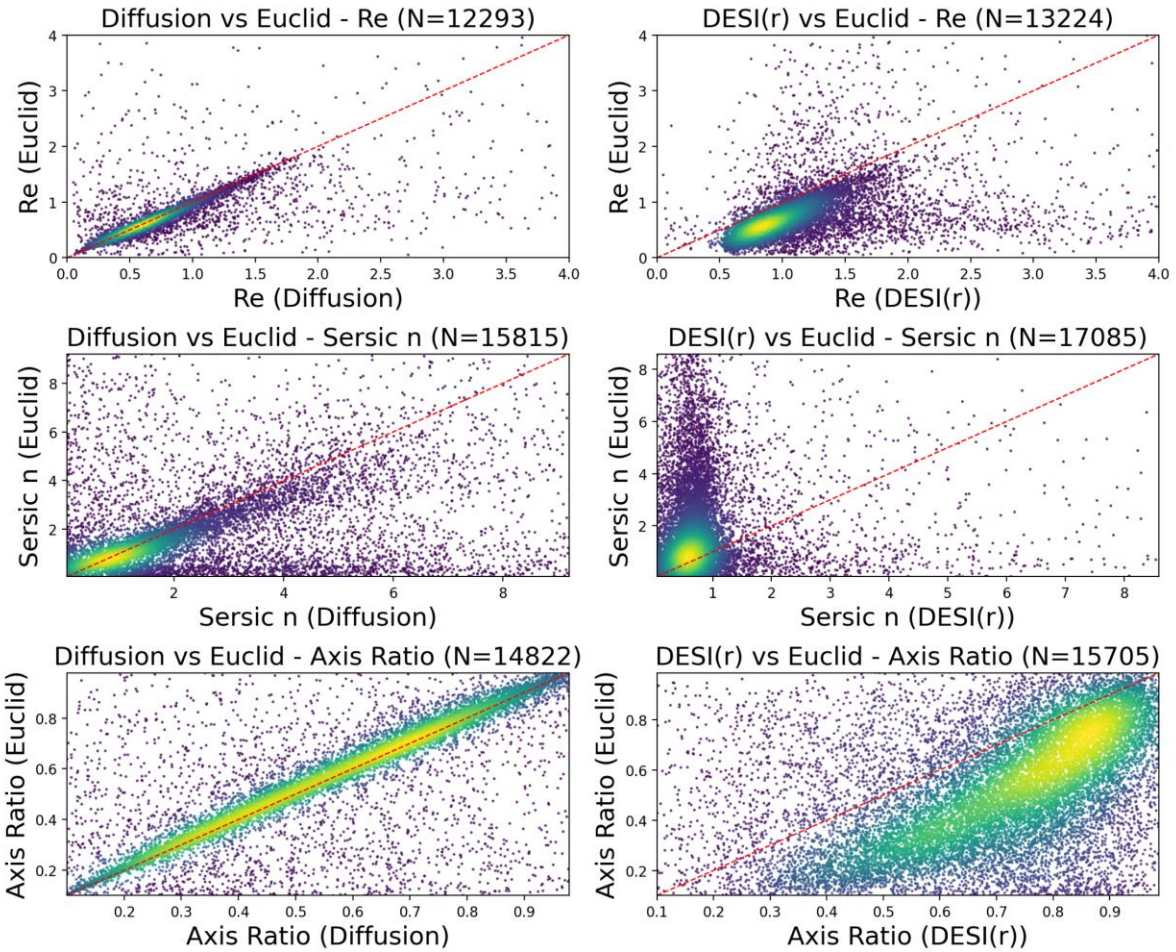
- low resolution →
  - high resolution
- low exposoure →
  - high exposure
- this band →
  - that band





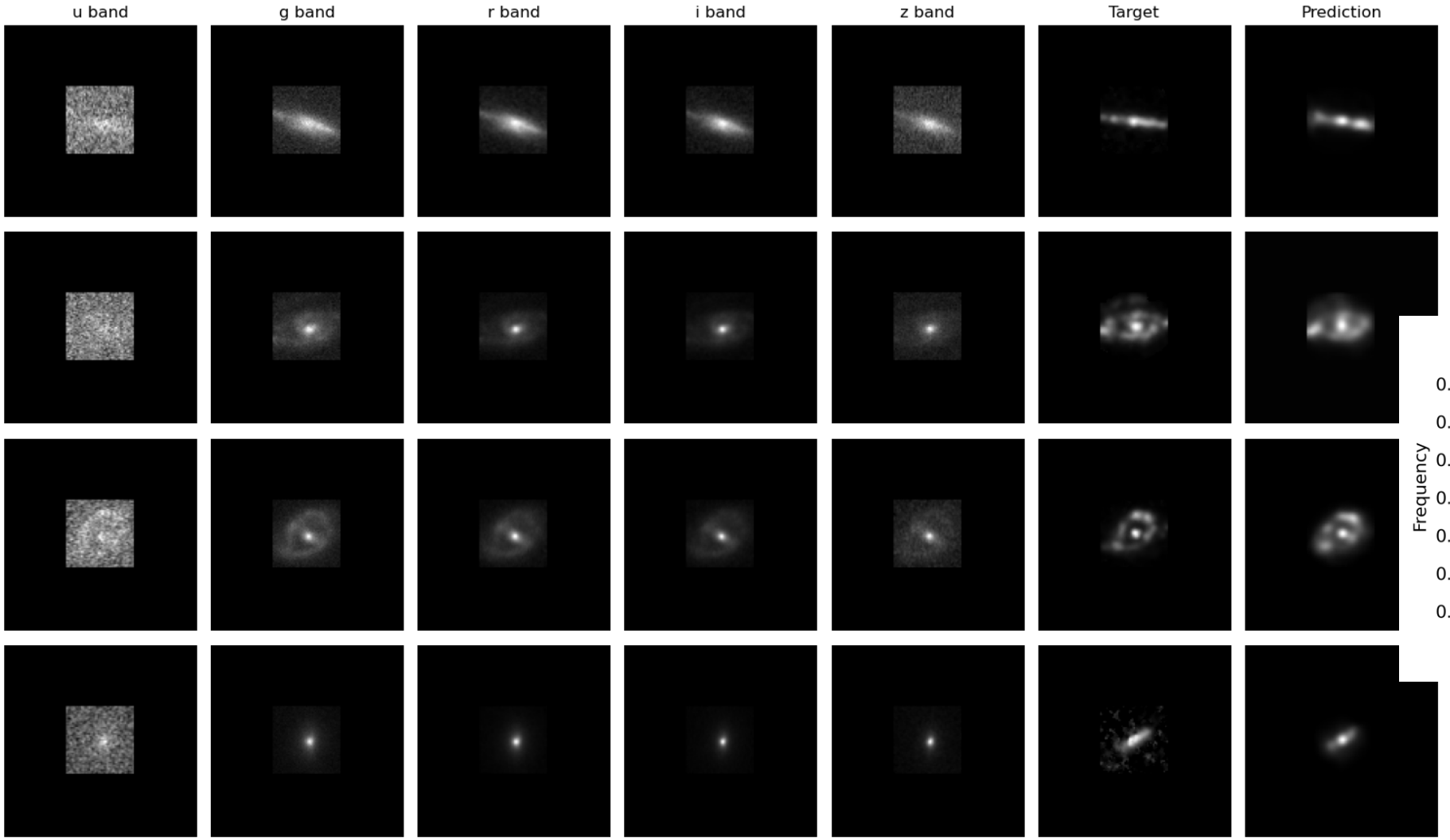
From DESI to Euclid (Ye,Shen+, in prep)





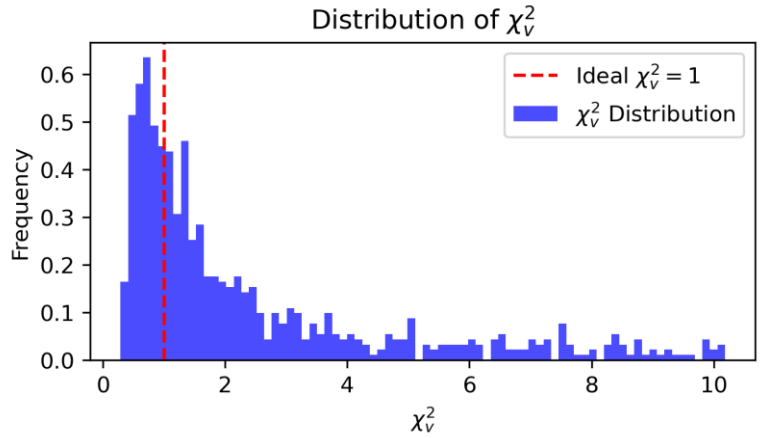
DESI BGS galaxies: from low to high resolution image quaility, revisit galaxy structure study

From u,g,r,iz images to Ha image



Trained on MaNGA data

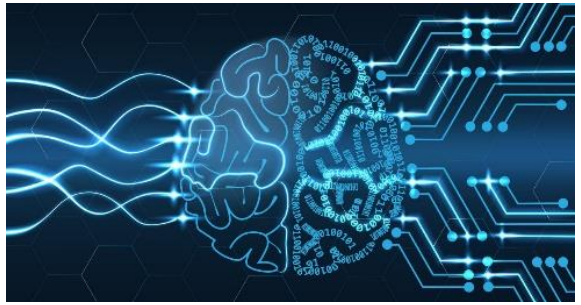
Shiyuan Chen
陈思远



Chen, Shen+ in prep

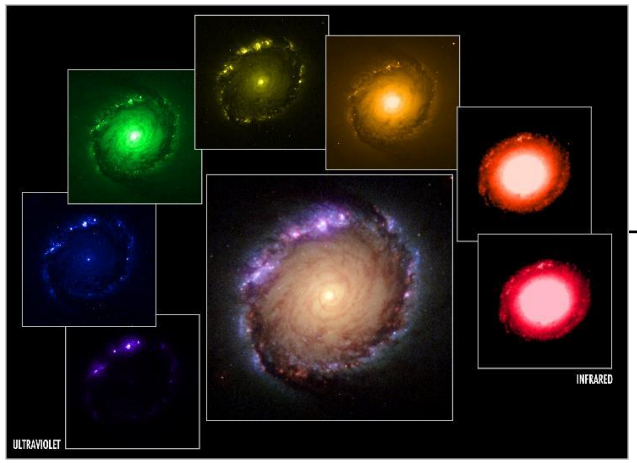
AI VS real observation

- AI: sampling from prior of high resolution images
- Observation: posterior sampling from real images



VS

Outlook: a latent-space baed galaxy image AI model



multi-band images

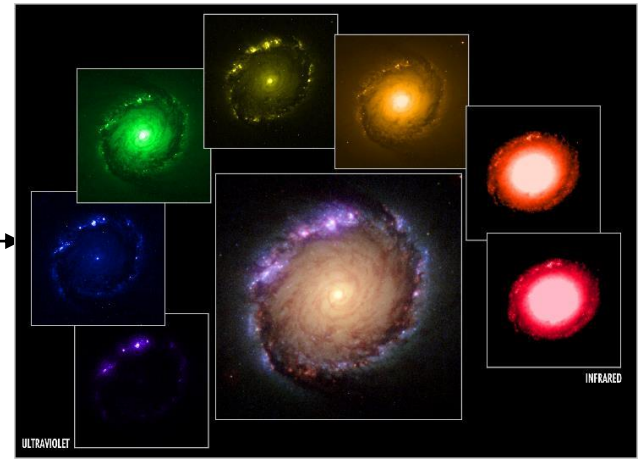
Encoder

instrument features

ideal image features

band features

Decoder

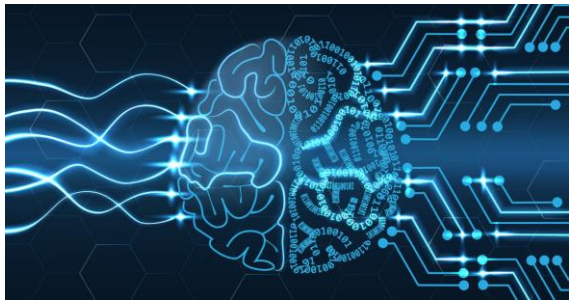


physical features (knowledge)

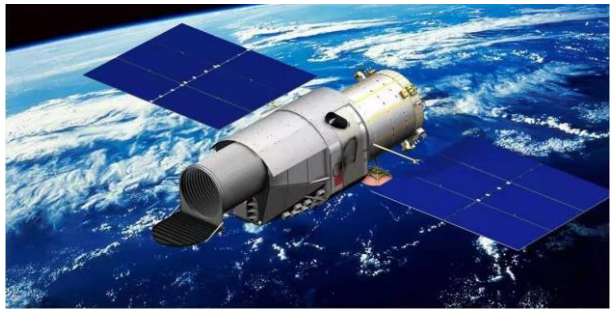
Normalizing Flow

$z_0 \sim q(z_0|x)$ $\xrightarrow{f_1}$ $\xrightarrow{f_2}$ $\cdots$ $\xrightarrow{f_n}$ $y' \sim z_n$

$z'_0 \xleftarrow{f_1^{-1}} \cdots \xleftarrow{f_{n-1}^{-1}} z'_n \xleftarrow{f_n^{-1}} y$



Summary



- Morden AI technic can build excellent mappings among galaxy images/parameters, which helps explore the physical nature of galaxies.

- Porposal: A unified latent-space baed AI galaxy image model

Mappings among all known galaxy image data
(e.g. SDSS ←→DESI ←→HSC ←→ HST ←→ JWST)