Latent space representations of galaxy images

and their relationship with galactic properties.

Nagoya University B4

Dillon Loh

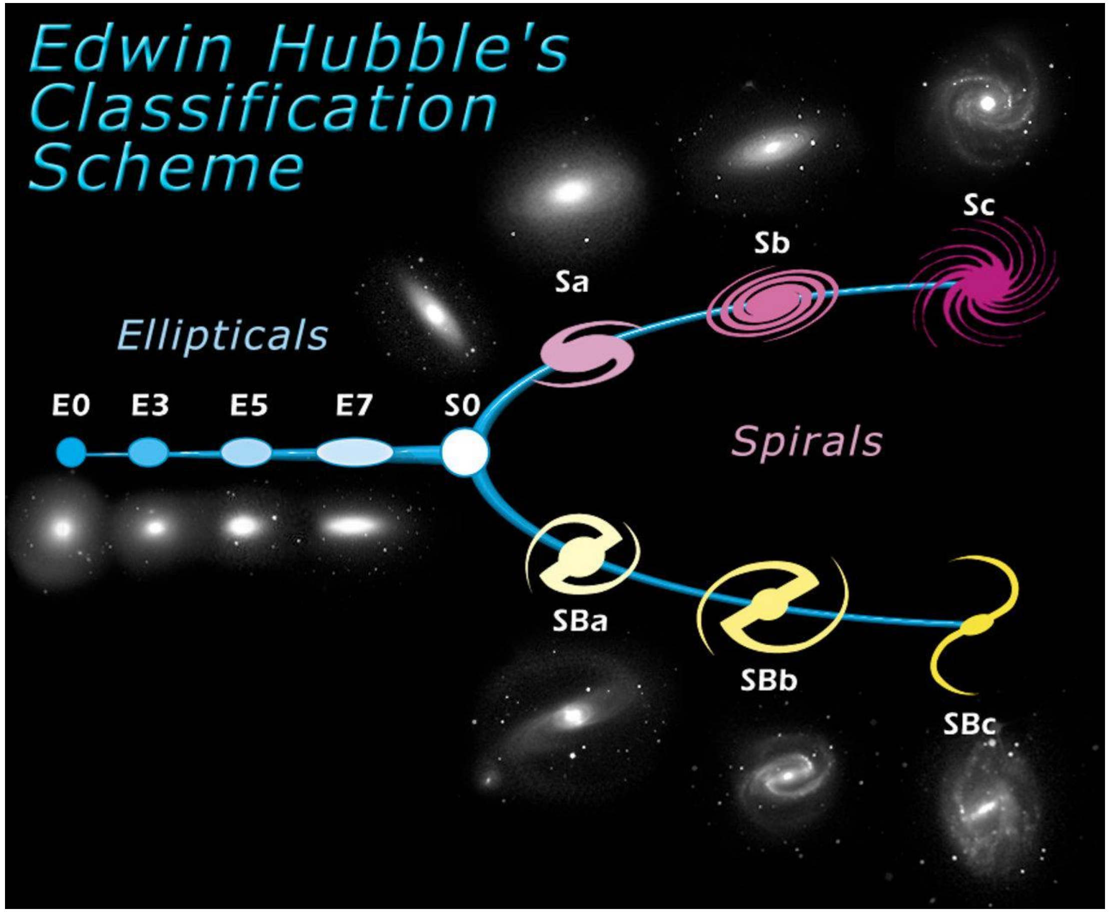
Takeuchi Tsutomu, Suchetha Cooray, Iwasaki Daiki

Flow of today's presentation

1. **Introduction**
   - Galactic Morphologies
   - Morphological vs Physical Properties
2. **Data**
   - Data Source
   - Preprocessing
3. **Method**
   - What are Neural Networks?
   - What are Variational Autoencoders?
4. **Application to Astrophysics**
5. **Conclusion/Future Work**

Galaxy Morphologies

- Galaxies are systems of stars and interstellar matter.
  - Come in all sorts of shapes and have defining features
- Morphology of a galaxy refers to its shape/features (i.e. stellar mass distribution)
- Galaxies that do not fall under Ellipticals or Spirals are called ‘Irregulars’



NASA

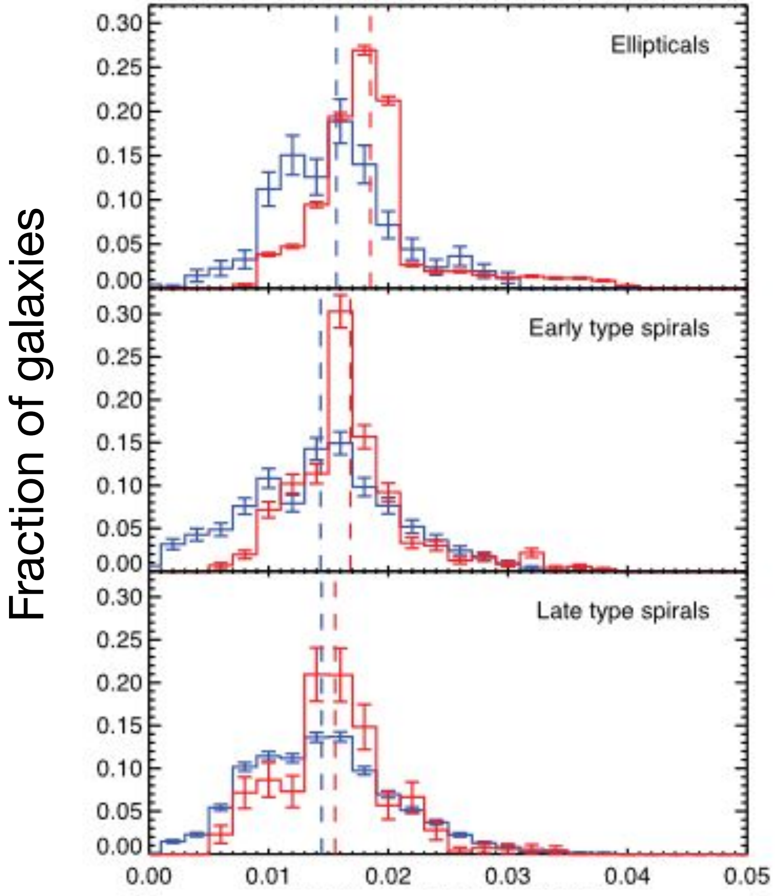
Morphology vs Physical Properties

In general, there are well-defined connections between the morphological properties of a galaxy and its physical property. *

- Ellipticals tends to be older.*
- Ellipticals tend to have higher metallicity
- Spirals tend to have higher SFR than Ellipticals**
- Spirals’ bulges tend to be older than their disc components

Can we predict a galaxy’s physical properties just from the features in its image?

Fraction of galaxies with different metallicities for Ellipticals and Blue/Red Spirals



Mass-weighted Stellar Metallicity

Tojeiro et al., 2013

Why do we want to predict a galaxy's physical properties just from the features in its image?

Direct calculation of physical properties is:

1. **Calculation/time intensive.**
2. **Requires us to collect large amounts of other data.**

Method

What needs to be done?

We need a way to extract features from the images autonomously.

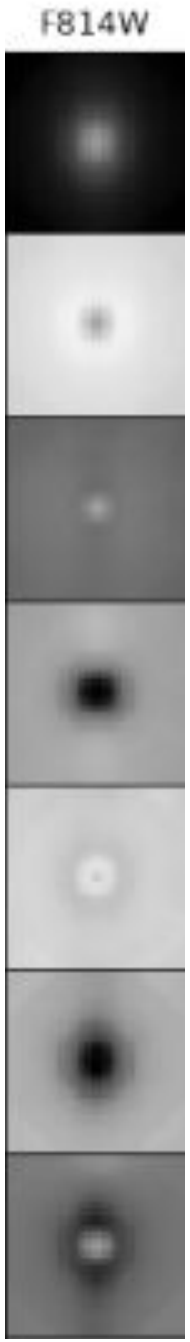
Past works

"Eigengalaxies: describing galaxy morphology using principal components in image space" – 2020

- Demonstrated that galactic images could be decomposed into a linear sum of 'eigen-galaxies' via PCA.

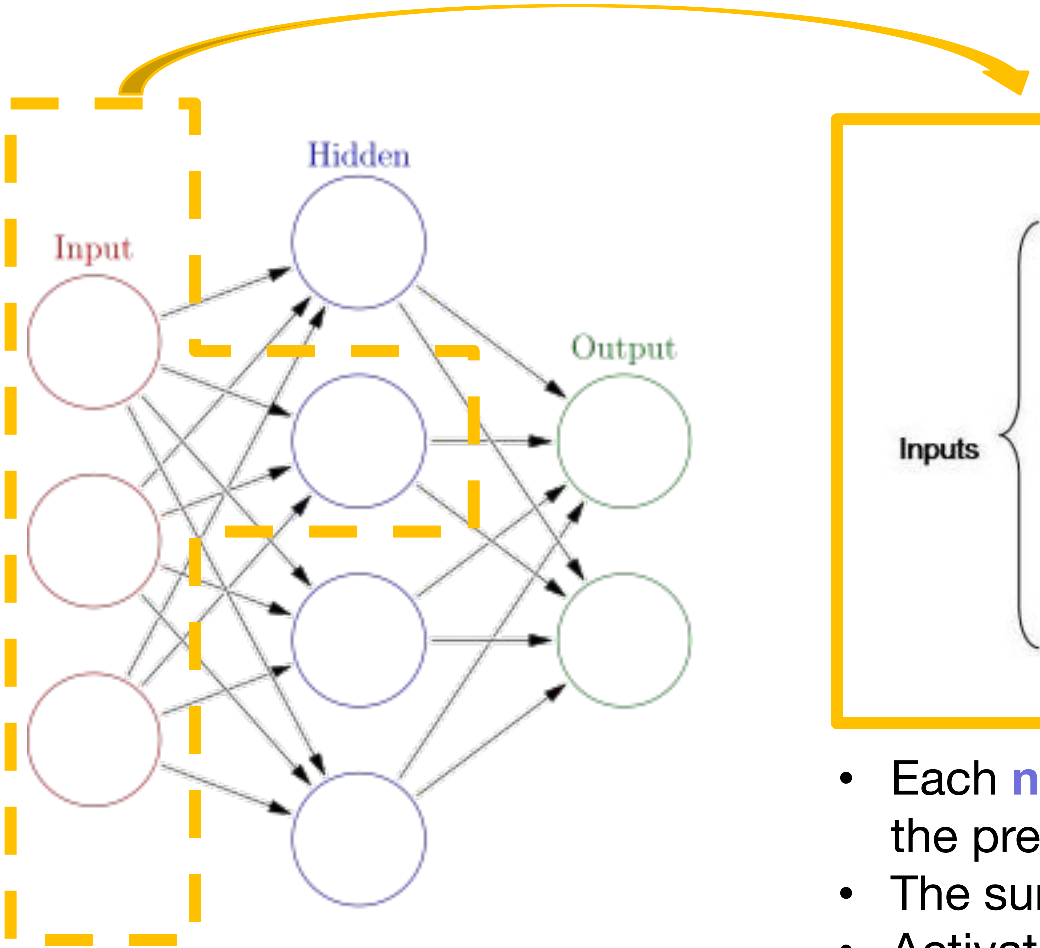
This work

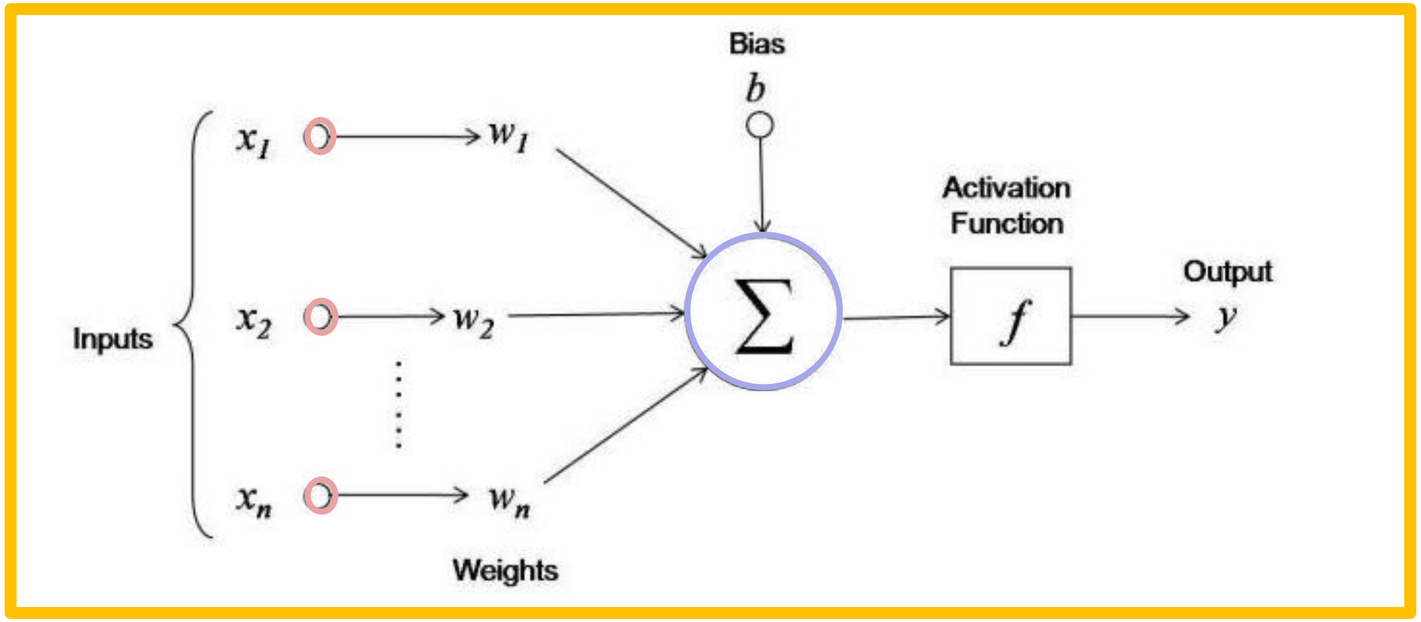
We use unsupervised deep learning (autoencoders).



(Uzeirbegovic et al.,

Neural Networks

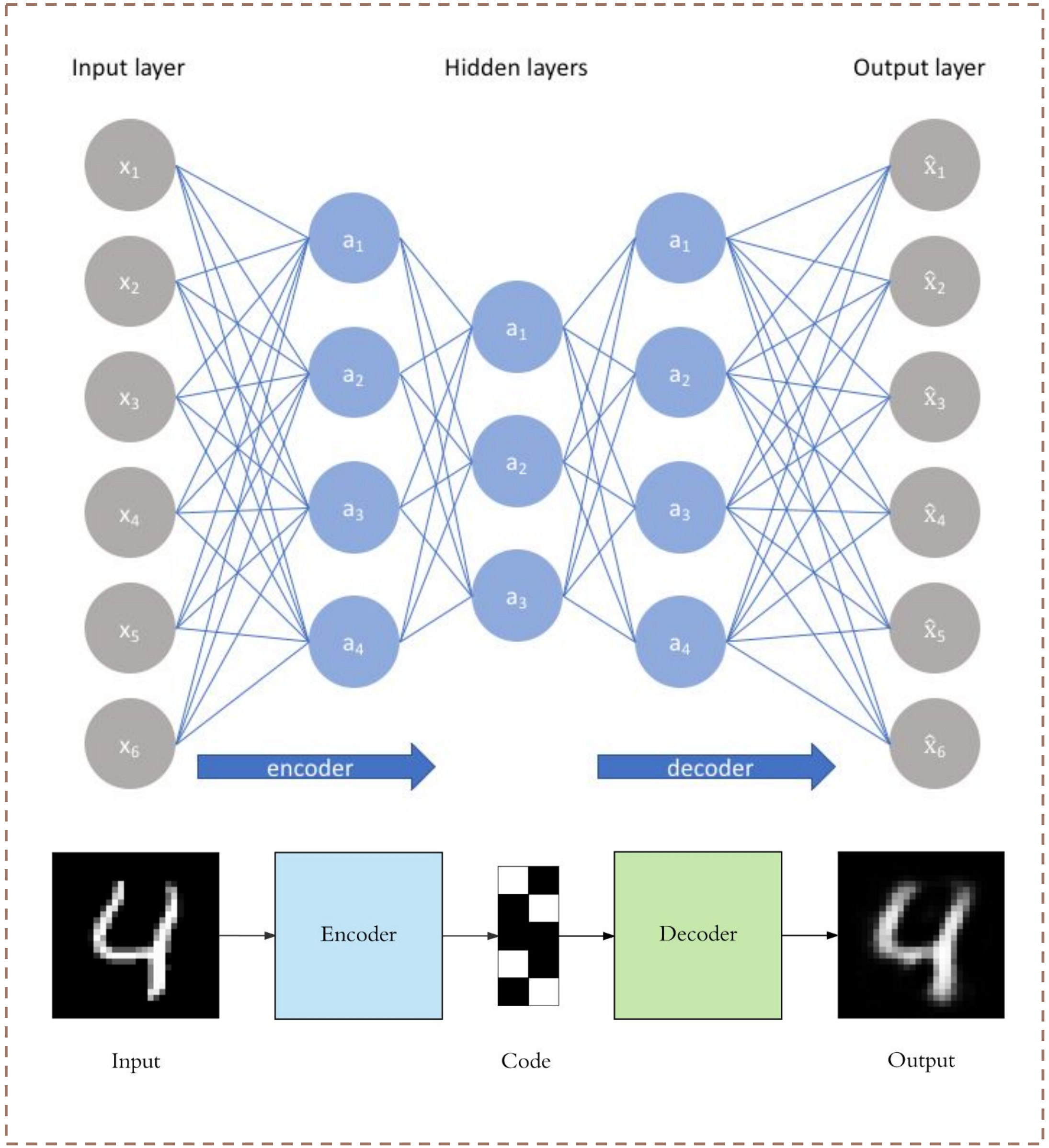




- Each **node** sums the (**inputs*** **weight**) from ALL the nodes of the previous layer.
- The sum is passed through an activation function.
- Activation function output is passed as inputs to next layer's nodes.
- "Training" the NN involves adjusting the **weight** until it gives "good" outputs from the **final layer**.

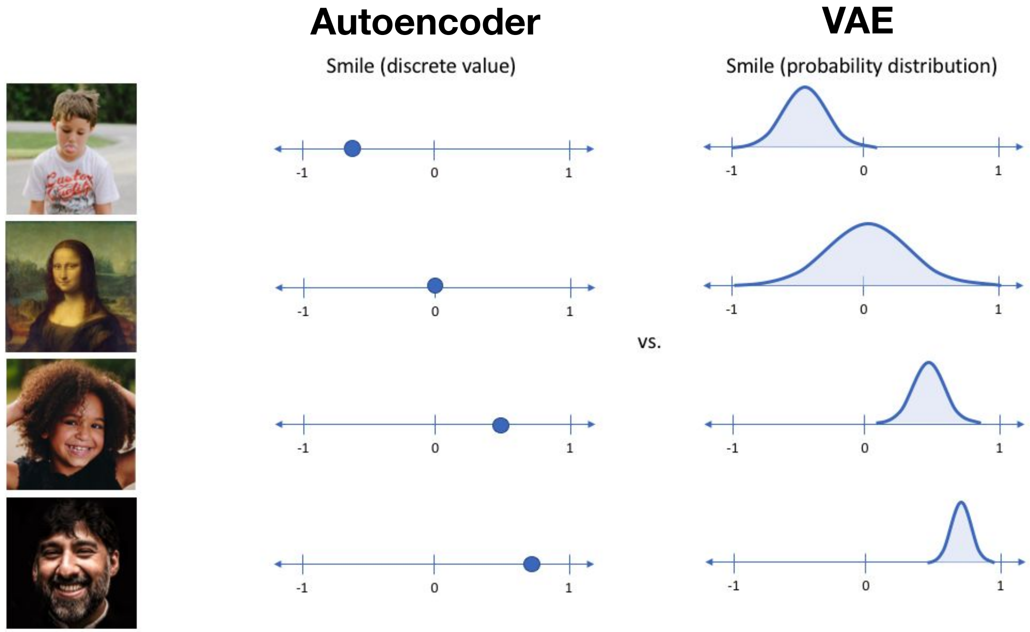
Autoencoders

- Autoencoders are a type of feed forward neural network.
- Consists of a ‘encoder’ and ‘decoder’.
- Reconstruction-loss function helps check and guide training progress
- The encoder is trained to extract features from the image.
  - Each node in the bottleneck layer is a ‘feature’.
  - When an image is passed into the network, it outputs one fixed value from each node.
  - These values are the ‘amounts’ of each feature in the image.

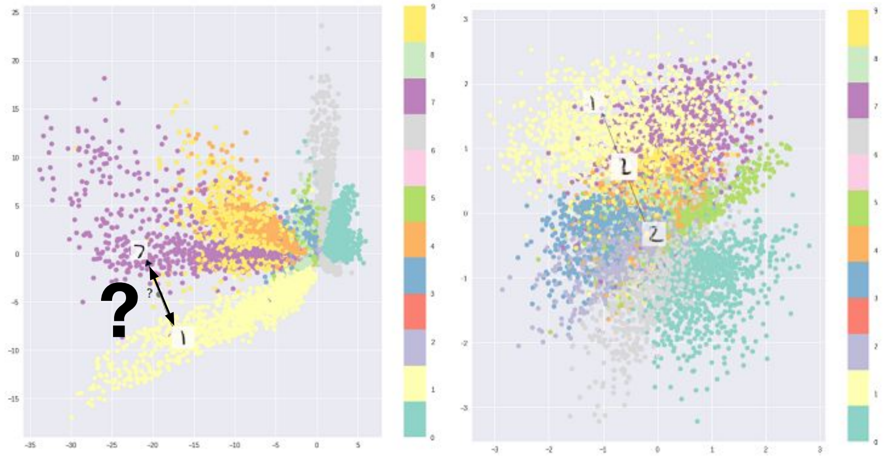


Variational Autoencoders

- Unlike a basic autoencoder, a variational autoencoder (VAE) represents each latent variable as a **probability distribution**.
- We can ‘shape’ our distribution to make it smooth.
- Typically, we use the very simple Gaussian distribution.
- To encourage our latent variables to resemble a Gaussian, we add a loss term that measures the difference between the latent and Gaussian distributions
  - Kullback-Leibler Divergence Term



https://www.jeremyjordan.me/variational-autoencoders
/



Data

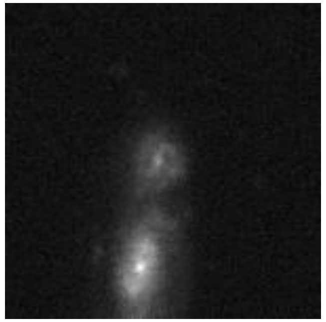
Data used

- Sloan Digital Sky Survey (SDSS) is a multi-spectral imaging and redshift survey.
  - Data Release 17 optical images are downloaded via SkyView*.
- SDSS *i* Band Images Used
  - Near-infrared wavelengths sensitive to long-living stars' emission
  - Representative of stellar mass distribution
- 150x150 pixel images

Example Images

Example Image

Relatively small object



Center object disturbed by extra objects

Presence of extra objects that have negligible disturbance

https://skyview.gsfc.nasa.gov/current/cgi/titlepage.pl

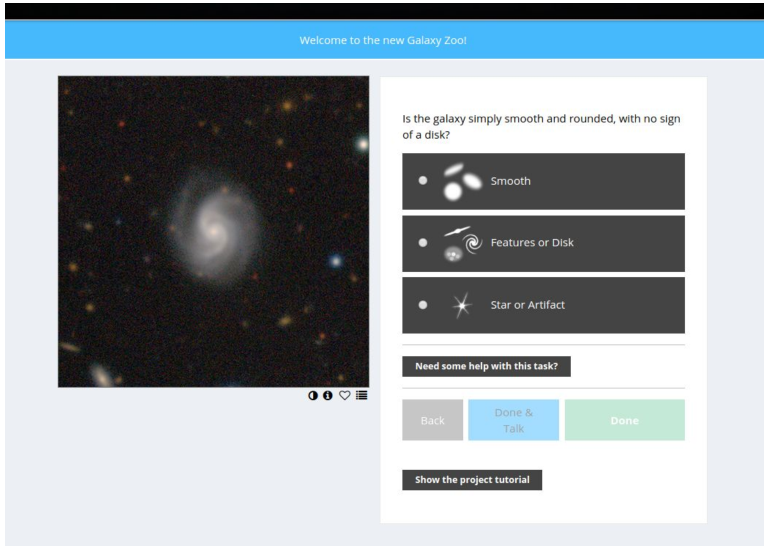
Filtering by Galaxy Morphology with Galaxy Zoo 2

Galaxy Zoo is a project **where humans assist in the manual identification of morphological features.**

Using this data, we filtered out the following:

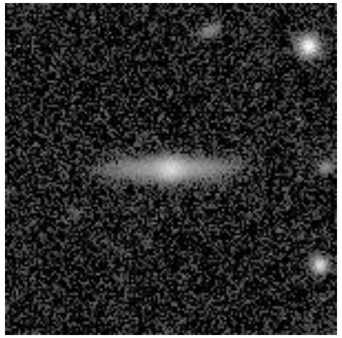
- Edge-on galaxies
  - Difficult to observe morphological features.
- Galaxies with ‘oddities’
  - Galactic merging, overlap, lensing, dust lane, etc.

Final Dataset: <u>12539 galaxies</u>

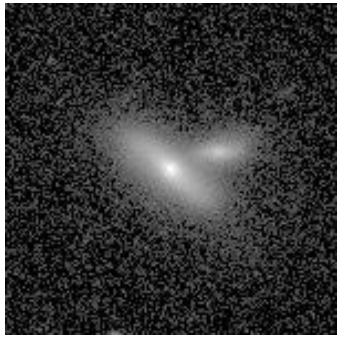


https://www.zooniverse.org/projects/zookeeper/galaxy-zoo/

Edge-on

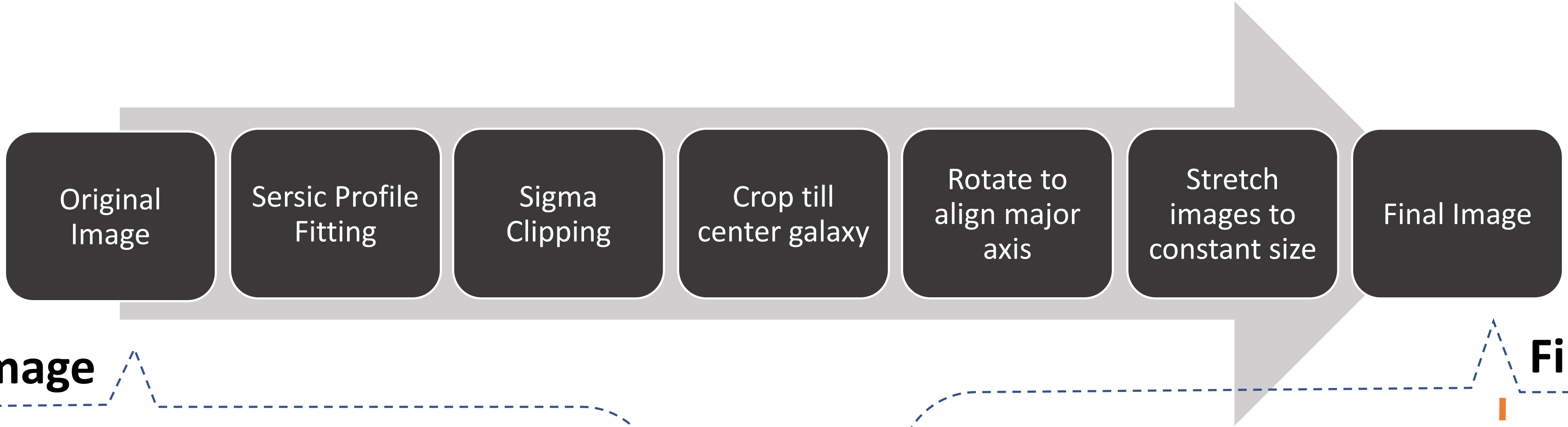


Oddities

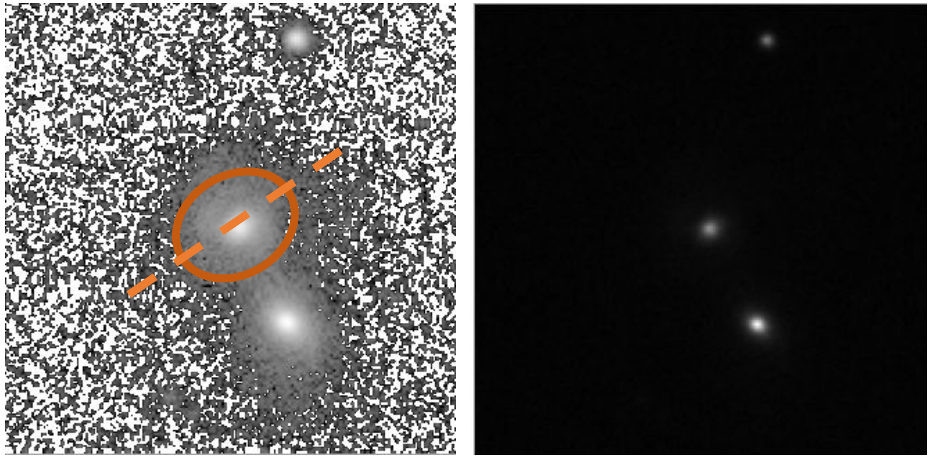


https://skyview.gsfc.nasa.gov/current/cgi/titlepage.pl

Data Preprocessing

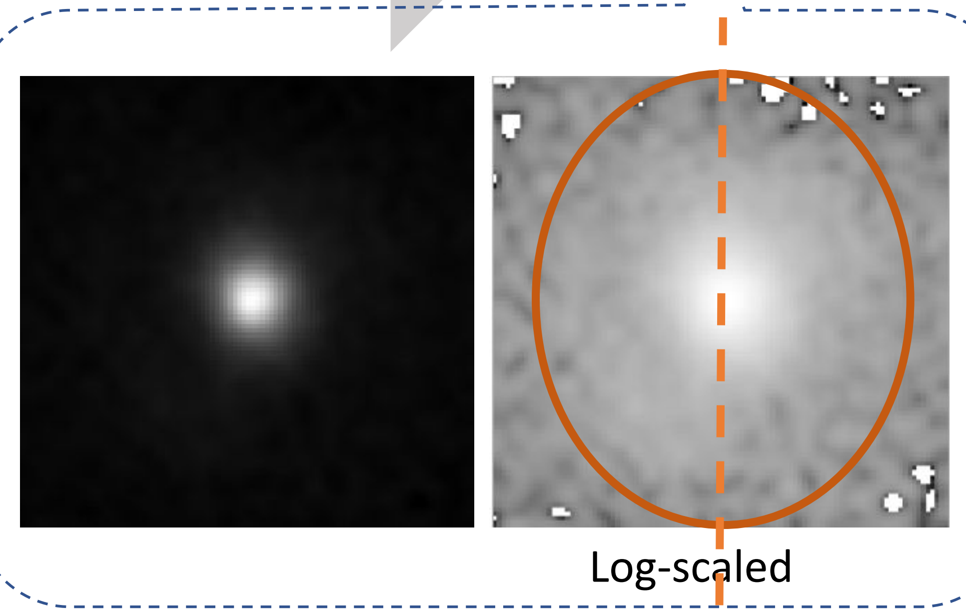


Original Image



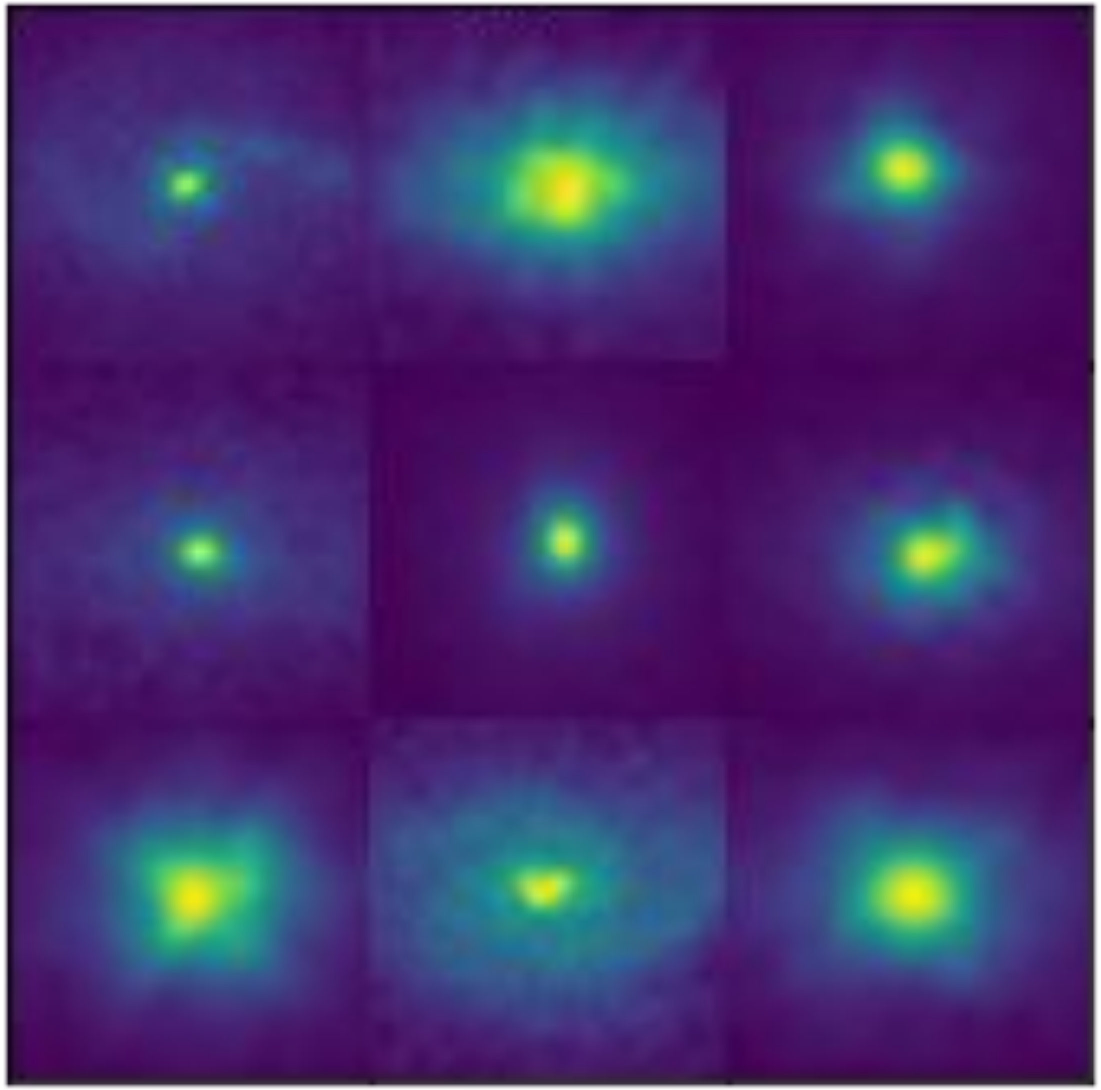
Log-scaled

Final Image



Log-scaled

Data Examples

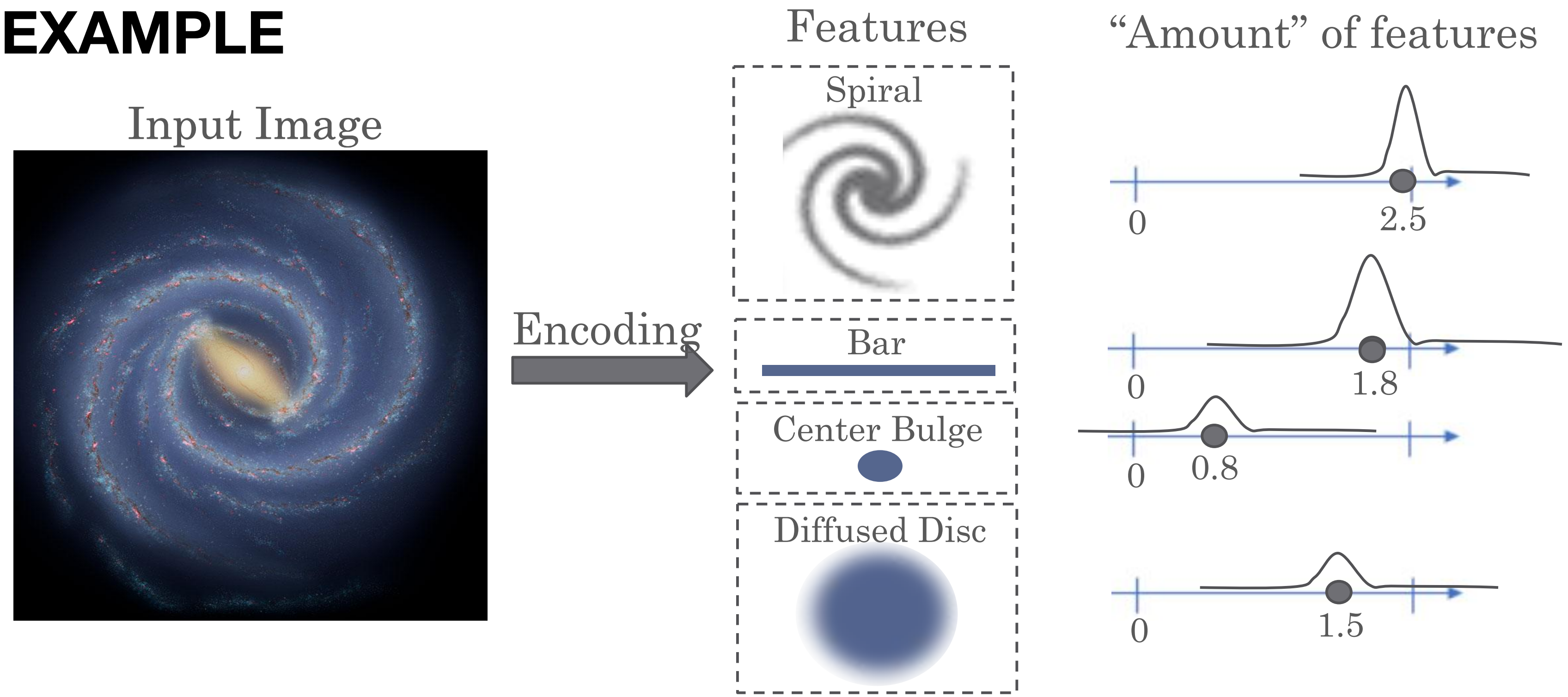


Astrophysics Applications

How did we use our autoencoders?

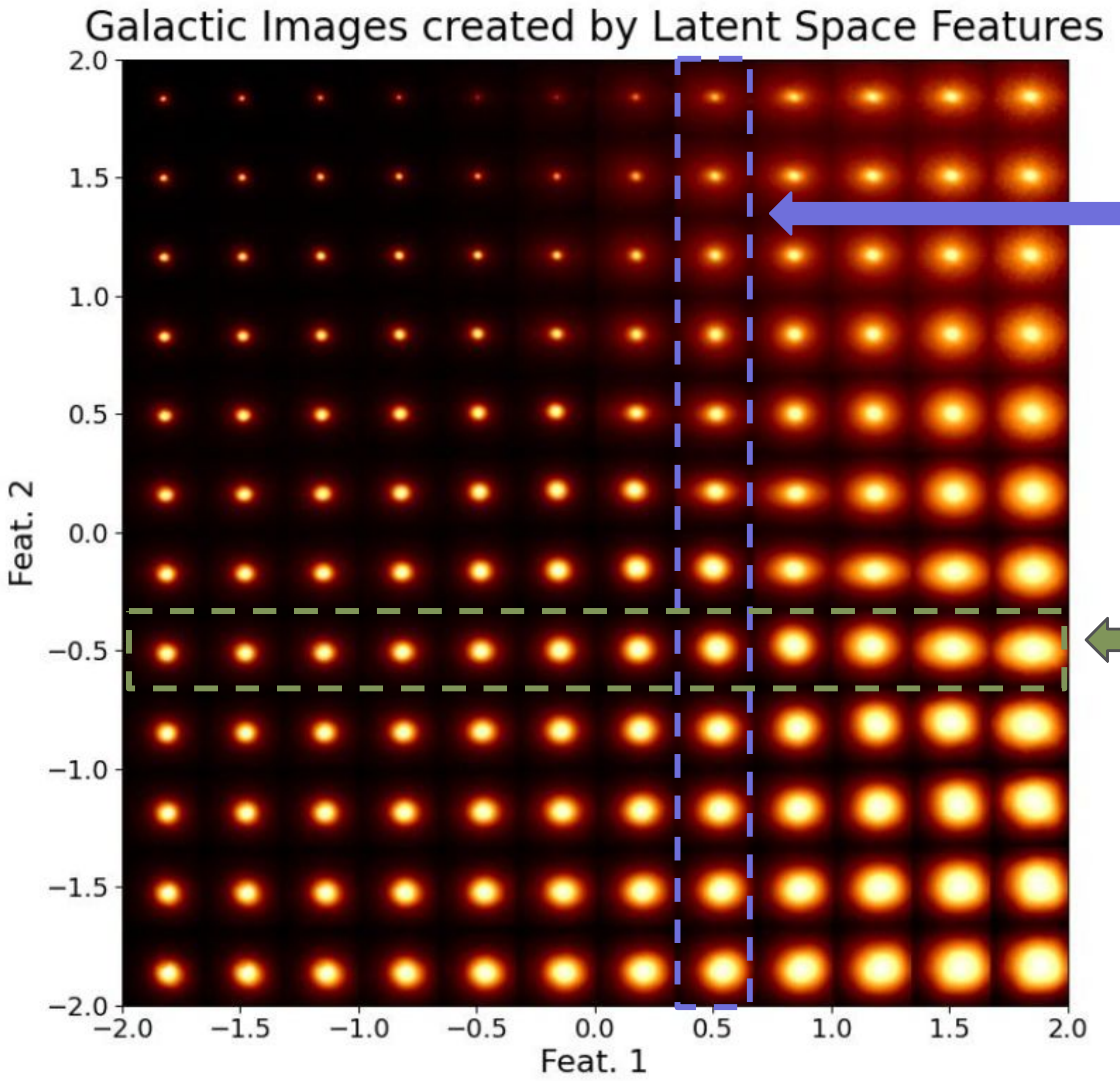
- We trained our autoencoder to find a set of the most common features in all our galactic images
- If we pass a galactic image into this autoencoder, it will give us a probability distribution of values for each feature.
  - If we passed it into a normal AE, it would just give a single value for each feature.
- The values represent 'how much' of that feature is present in the input galactic image.

EXAMPLE



Extraction of Morphological Features

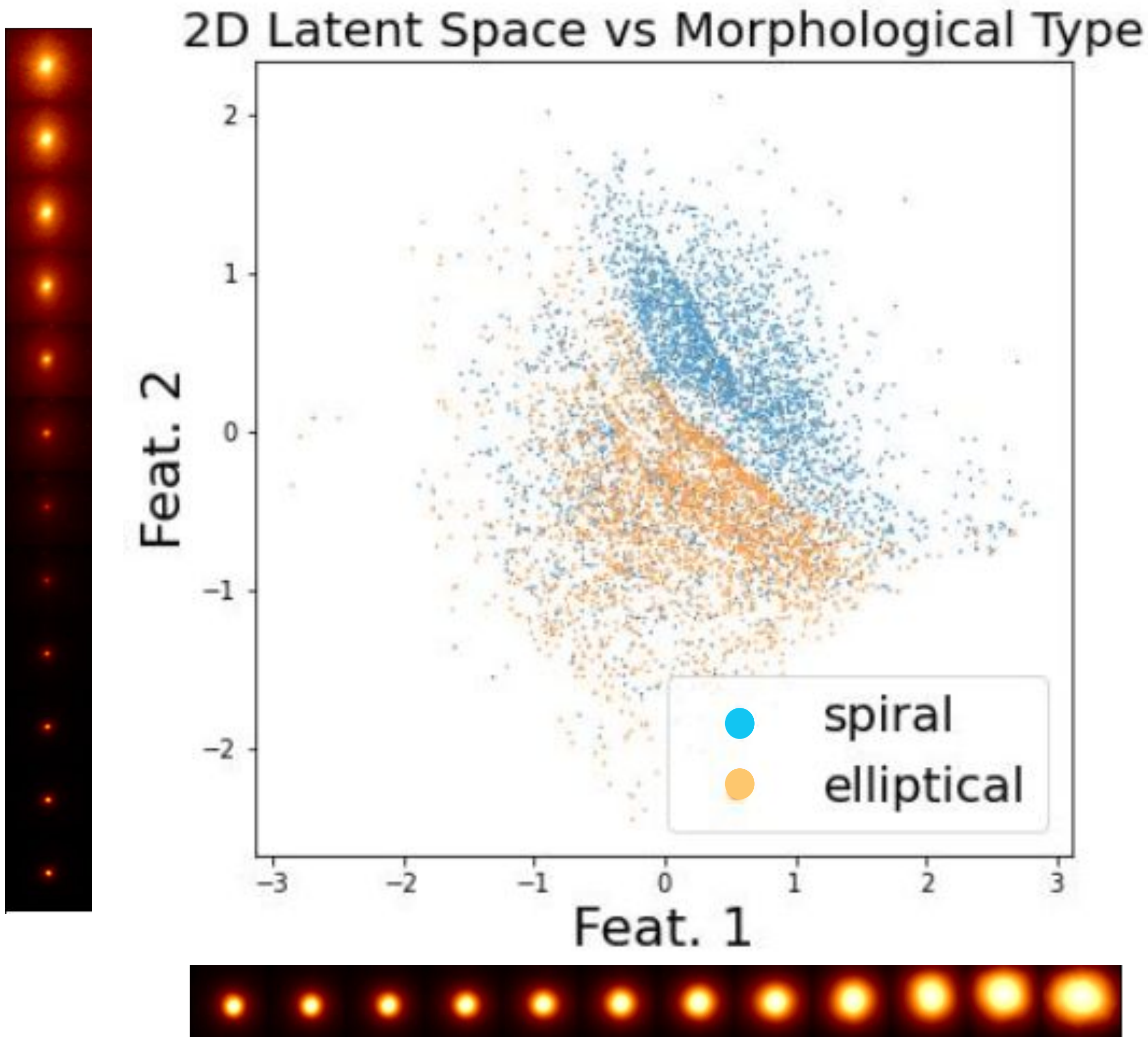
We set our VAE to find only 2 latent features that best describes the images.



Higher Feat. 2 corresponds to bigger diffused disc-like structures.

Higher Feat. 1 corresponds to bigger center bulges

Morphological Class Distribution in Latent Space



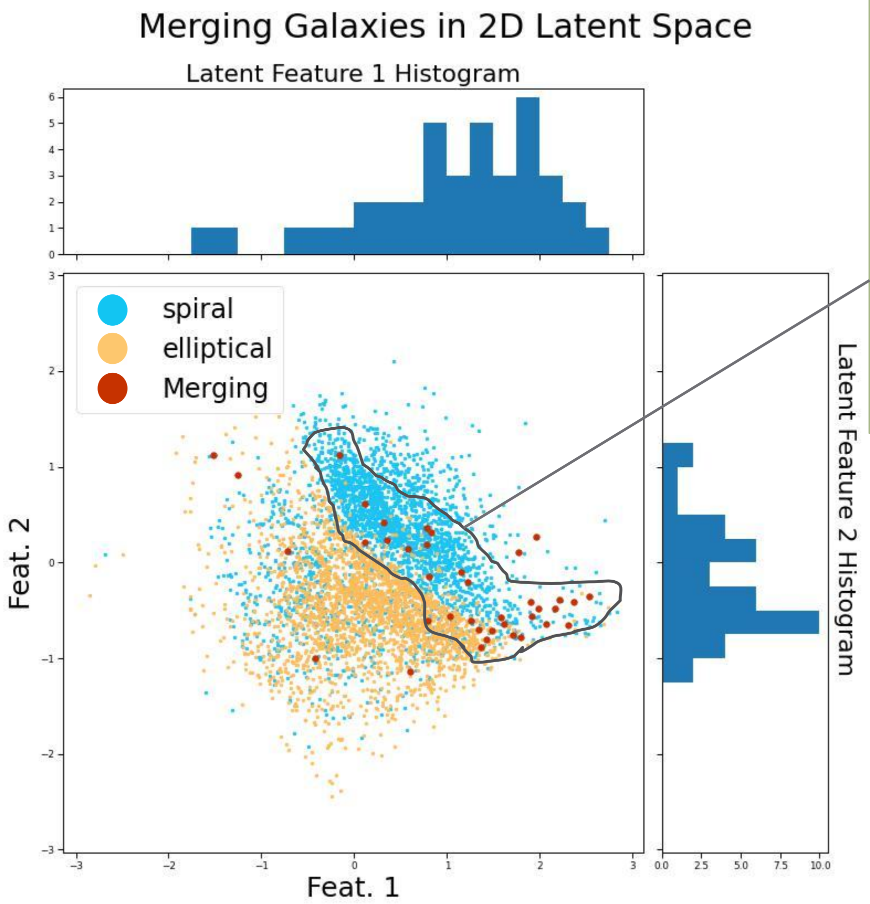
	Feat. 1	Feat. 2
spiral	0.322450	0.198545
elliptical	0.137961	-0.344143

- Spirals tend to have higher Feat. 2 while Ellipticals have lower.
- Ellipticals tend to have higher amounts of Feat. 1
- Surprisingly, distribution of spirals is skewed towards greater Feat. 1

What about irregular galaxies?

What if we now encoded images of irregular/merging galaxies?

- Recall that we trained only with ‘regular’ galaxies (Spirals/Ellipticals)

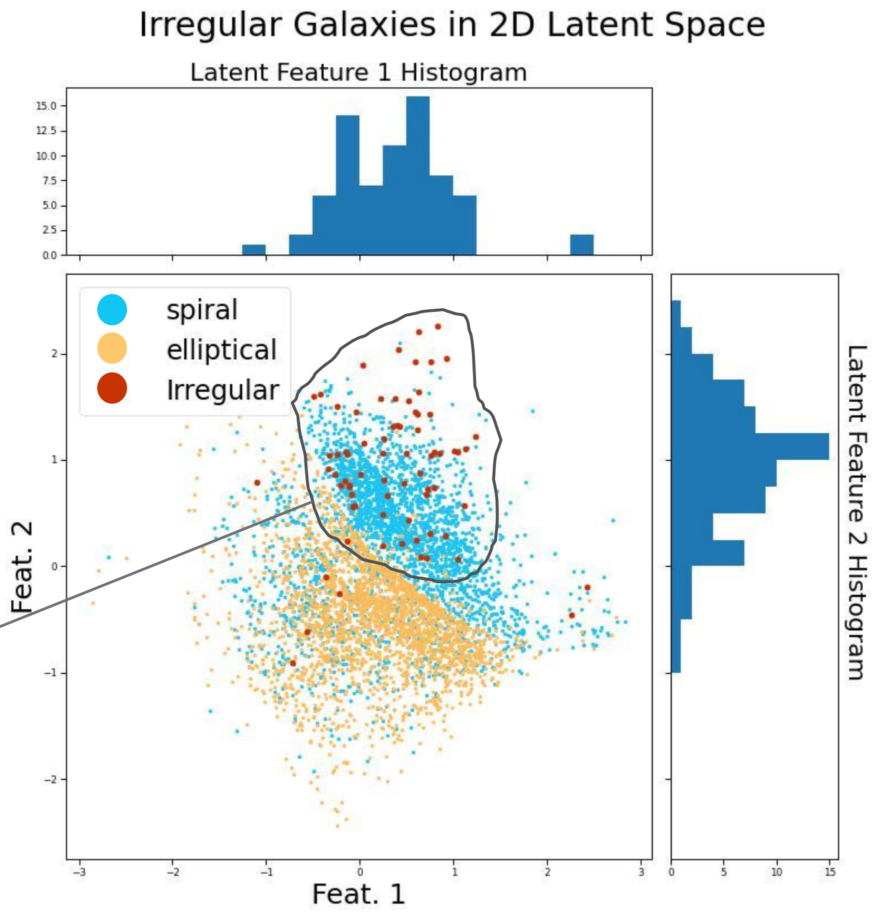


Merging galaxies found along spiral-elliptical border.

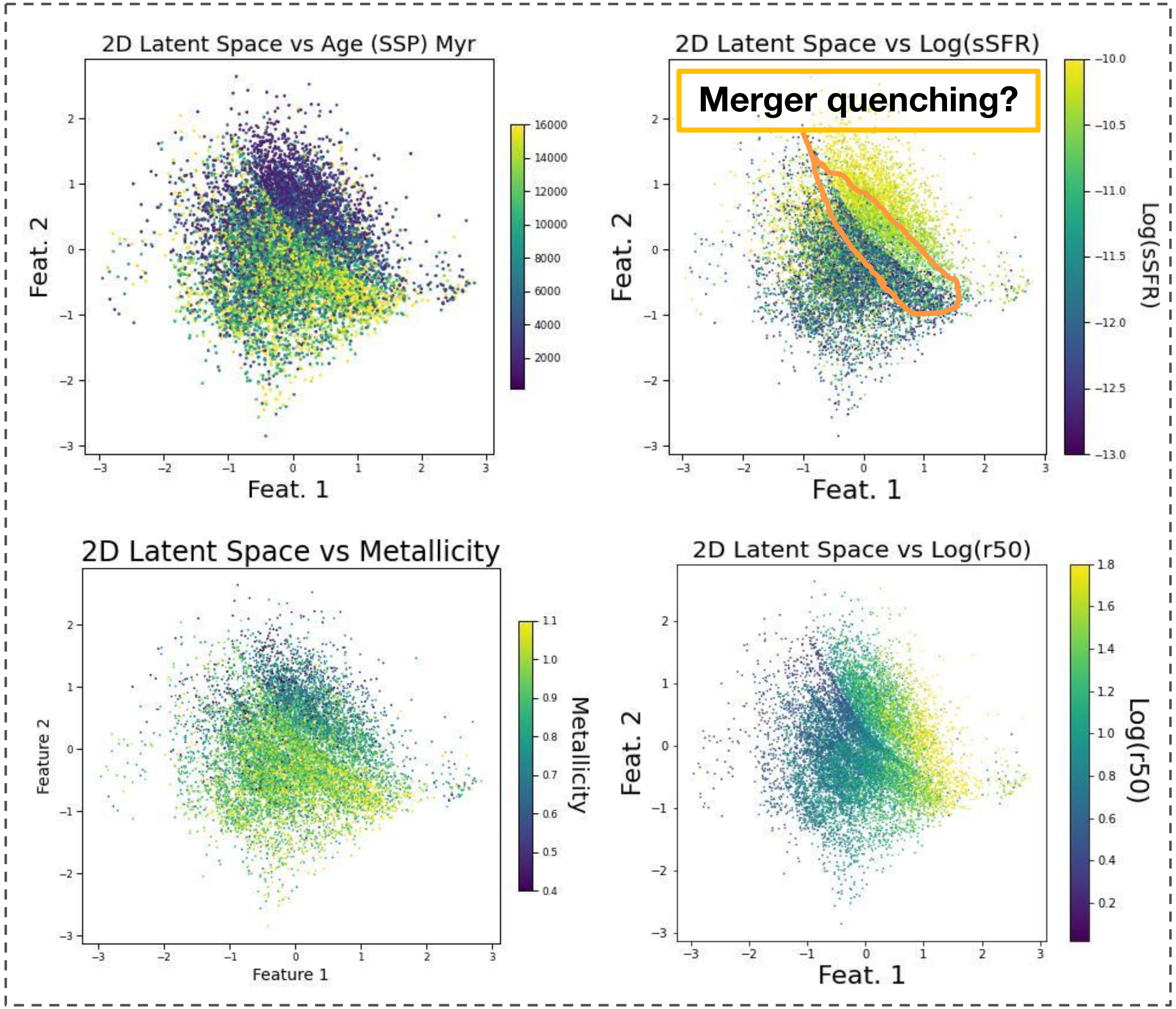
- Spirals merge to form ellipticals.
- Suggests merger quenching? (Refer to next slide)

Irregular Galaxies in top right (higher Feat.1, Feat.2)

- Higher sSFR region
- ??



Latent Space vs Physical Properties

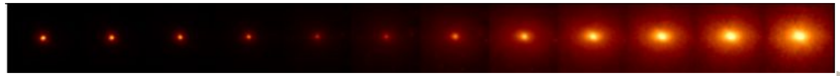


- We see that there are gradients in the physical properties across Latent Space.
- Some of these make intuitive sense based on our knowledge of galactic structures.

For example

Feat. 1 ↑

- Since stars in center bulges tend to be older, a larger bulge implies an overall older galaxy (assuming equally sized galaxies)



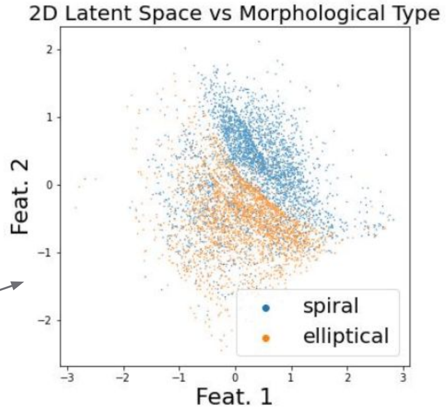
Feat. 2 ↑

- Since SFR in spiral arms/discs tend to be higher, larger proportion of them in galaxy => higher sSFR.

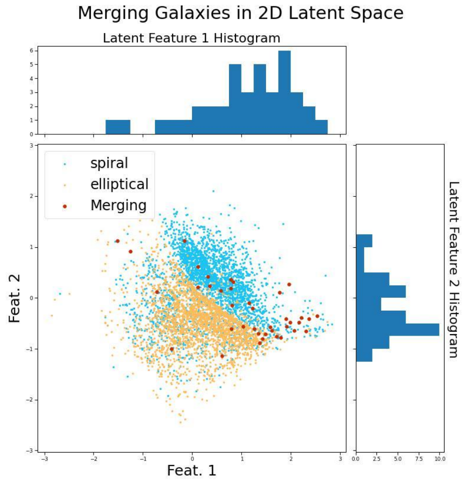
Conclusion/Discussion

Conclusion

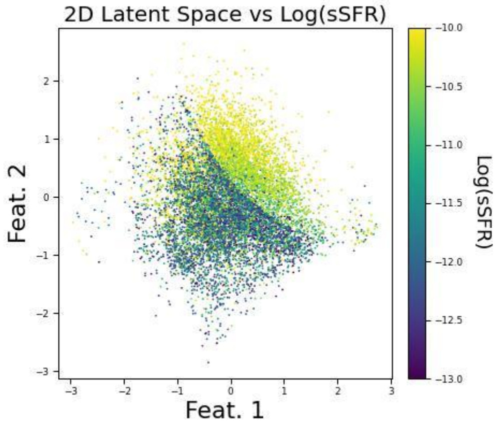
- Our VAE was able to find a set of just <u>2</u> features that distinctly defines the morphology of spiral and elliptical galaxies.



- Even though we only trained with ‘regular’ galaxies, our VAE could still be applied to ‘irregular’ galaxies with physically interpretable results.



- The image latent features our VAE found also had monotonic relationships with actual physical properties.



Future Work

- Use multiple band images to better capture other sources of radiation/physical properties.
- Use maps of other physical properties such as emission line maps or stellar velocity dispersion maps.
- Build parametric models based on these latent feature/physical property relationships.

I would greatly appreciate further suggestions to improve on this work!

Thank you

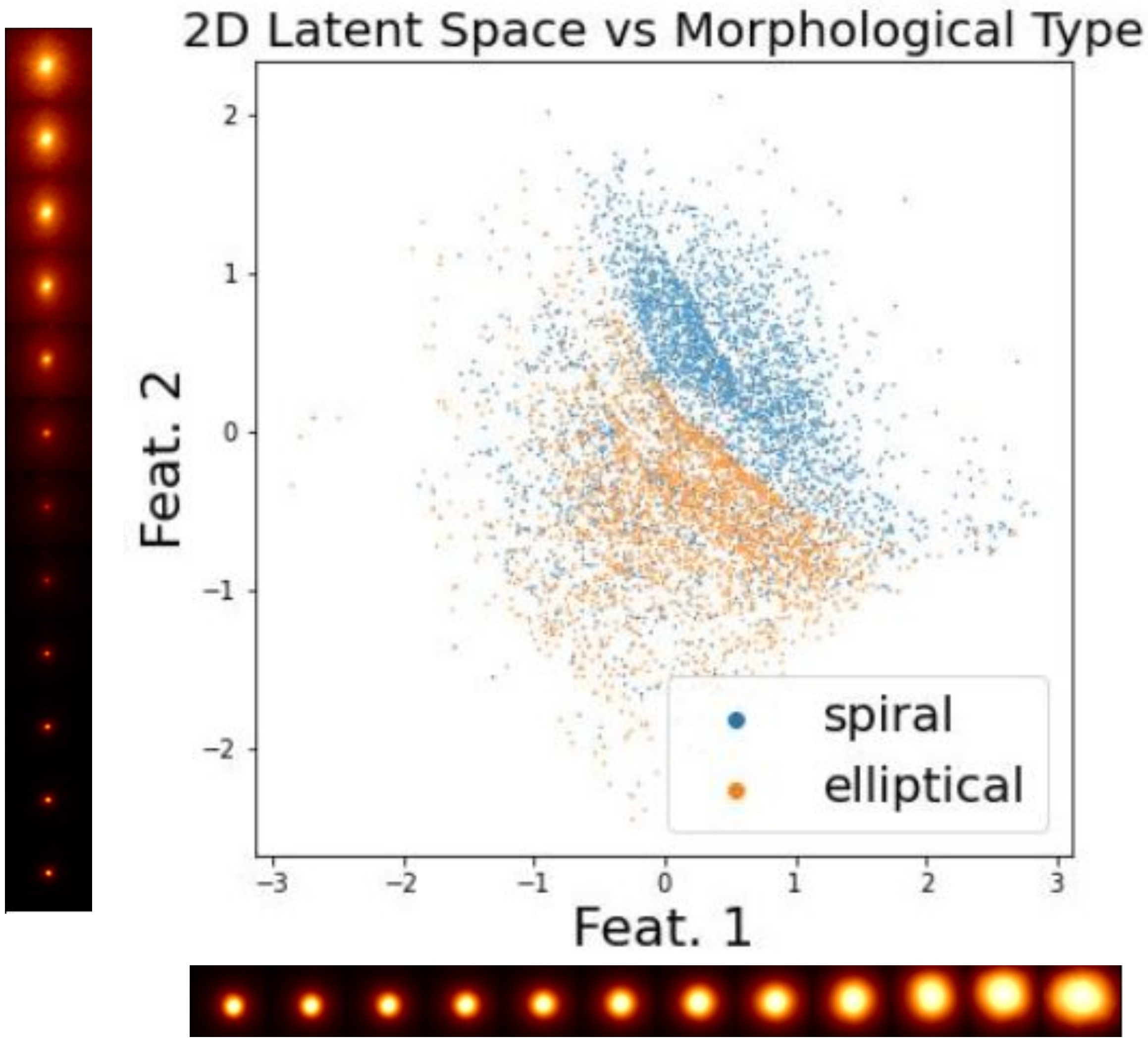
Correlations Tables

	Feat. 1	Feat. 2
Vel. Dispersion (SSP) km/s	-0.051947	-0.465508
Age (SSP) Myr	0.034974	-0.417873
Metallicity (SSP)	0.019536	-0.455701
Vel. Dispersion (exp SFH) km/s	-0.058639	-0.472199
Age (exp SFH) Myr	-0.049727	0.466863
Metallicity (exp SFH)	0.057441	-0.445452
Petrosian 50 Radius arcsec	0.717488	-0.033900
LOGSFRSED	-0.058900	0.045080
LOGMSTAR	-0.069458	-0.008703
LOGSSFR	0.110948	0.334421

	Feat. 1	Feat. 2
spiral	0.018326	-0.390460
elliptical	0.282897	-0.045579
uncertain	-0.249939	0.376698

	Feat. 1	Feat. 2
Has Smooth Profile	-0.123796	0.563262
Has Features or Disk	0.126552	-0.561337
Has Spiral Arms	-0.022620	-0.425911
Has Obvious Bulge	0.292862	0.088417
Is Completely Round	0.172631	0.014948

Morphological Class Distribution in Latent Space

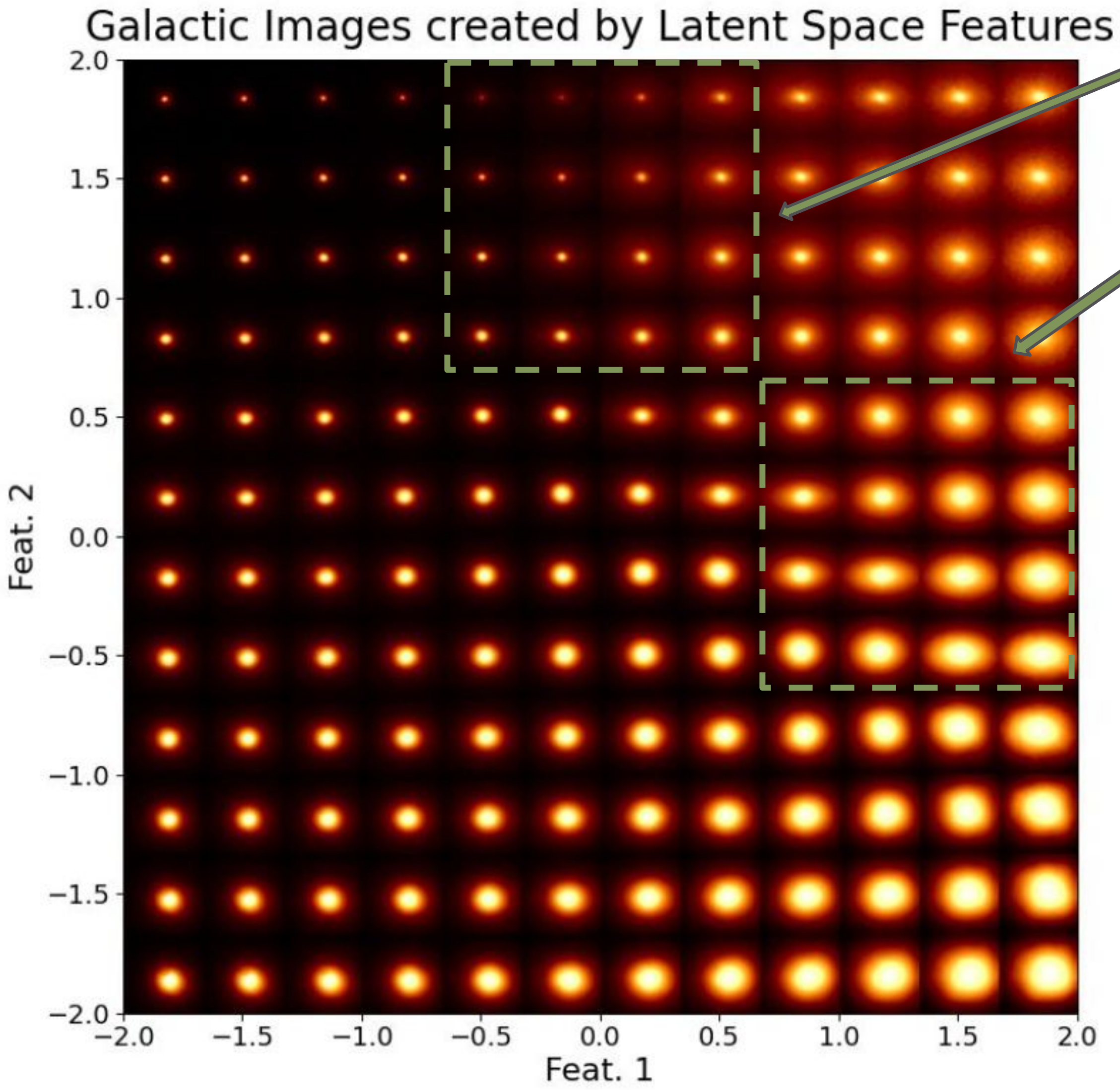


	Feat. 1	Feat. 2
spiral	0.322450	0.198545
elliptical	0.137961	-0.344143

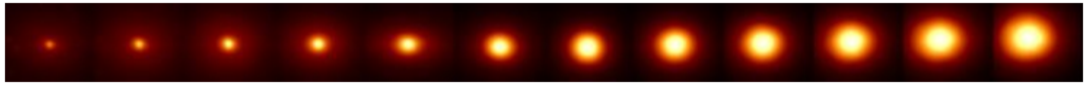
- Spirals tend to have higher Feat. 2 while Ellipticals have lower.
- Ellipticals tend to have higher amounts of Feat. 1
- Surprisingly, distribution of spirals is skewed towards greater Feat. 1

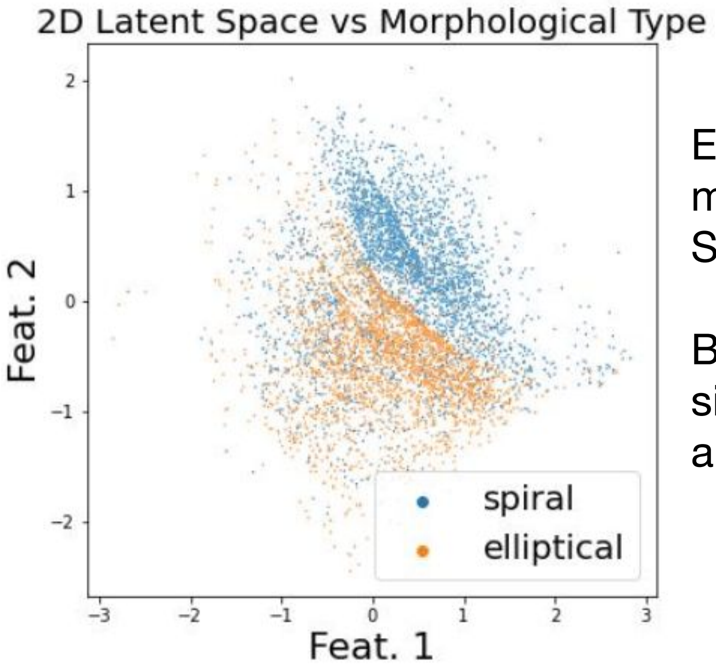
Morphological Class Distribution in Latent Space

Higher Feat. 2 corresponds to more diffused disc-like structures.



Higher Feat. 1 corresponds to bigger center bulges

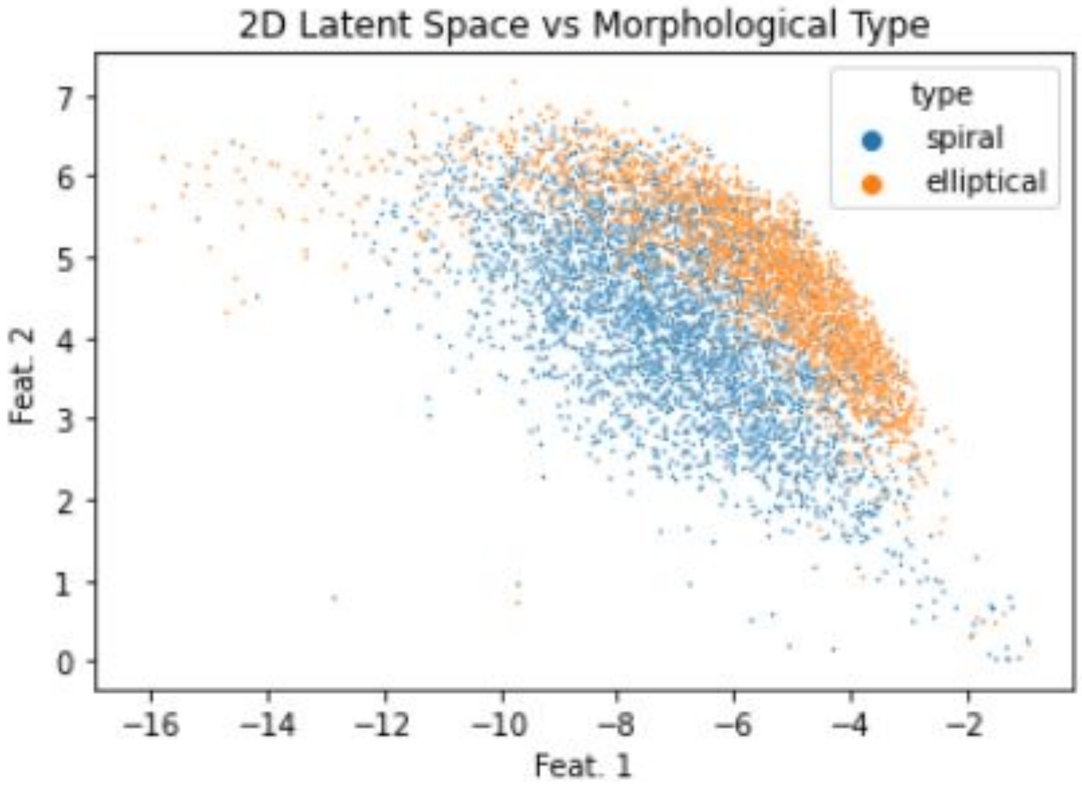




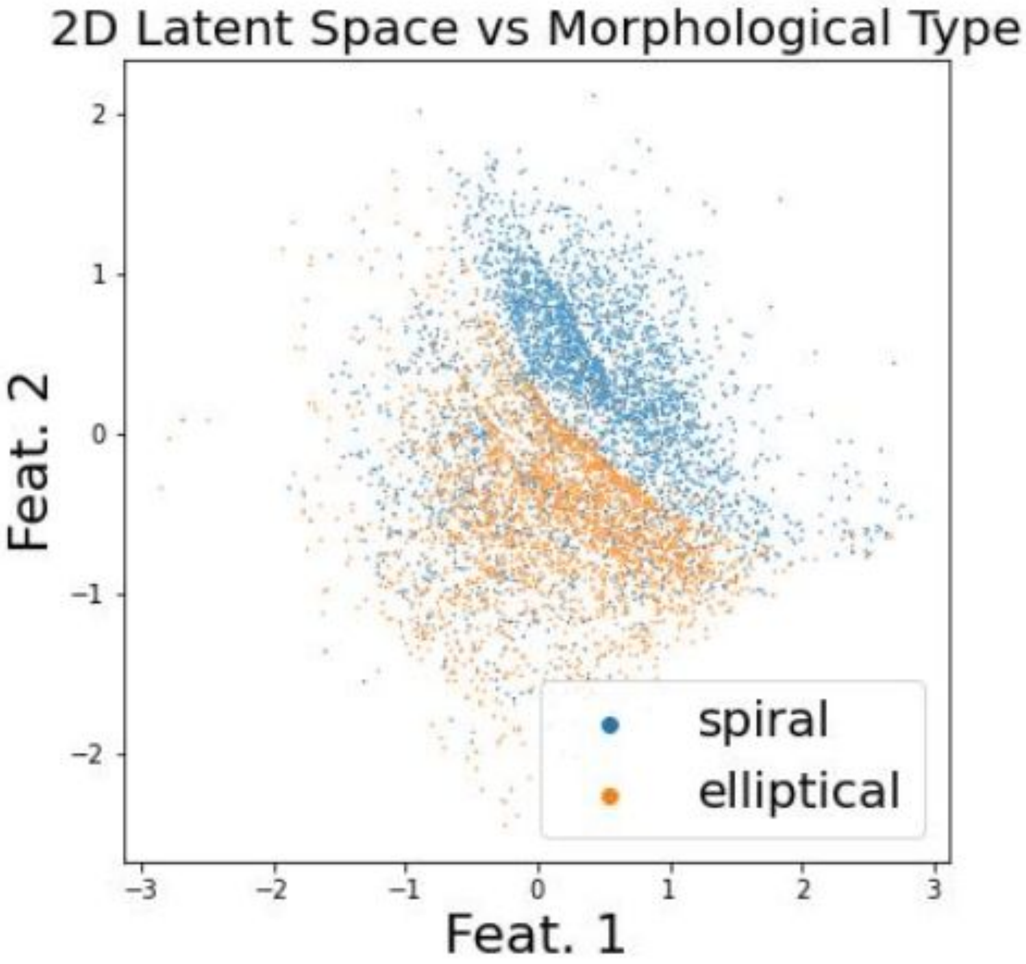
Elliptical galaxies have more less Feat. 2 while Spirals have more.

Both classes are similarly distributed along Feat. 1

Morphological Class Distribution in Latent Space



Autoencoder



Variational Autoencoder