# Application of unsupervised neural networks for predicting galactic properties via image features

Dillon Loh,*,[1] Takeuchi Tsutomu,[1,2] Suchetha Cooray,[1] Iwasaki Daiki[1]

[1] Division of Particle and Astrophysical Science, Nagoya University
[2] The Research Center for Statistical Machine Learning, The Institute of Statistical Mathematics
*Email: loh.dillon@gmail.com

## ABSTRACT

We want to determine the various physical properties of a galaxy in order to understand its evolution. However, typical methods of direct calculation using photometric/spectral data have two drawbacks:

1. Computationally expensive
2. Requires large amounts of collected data.

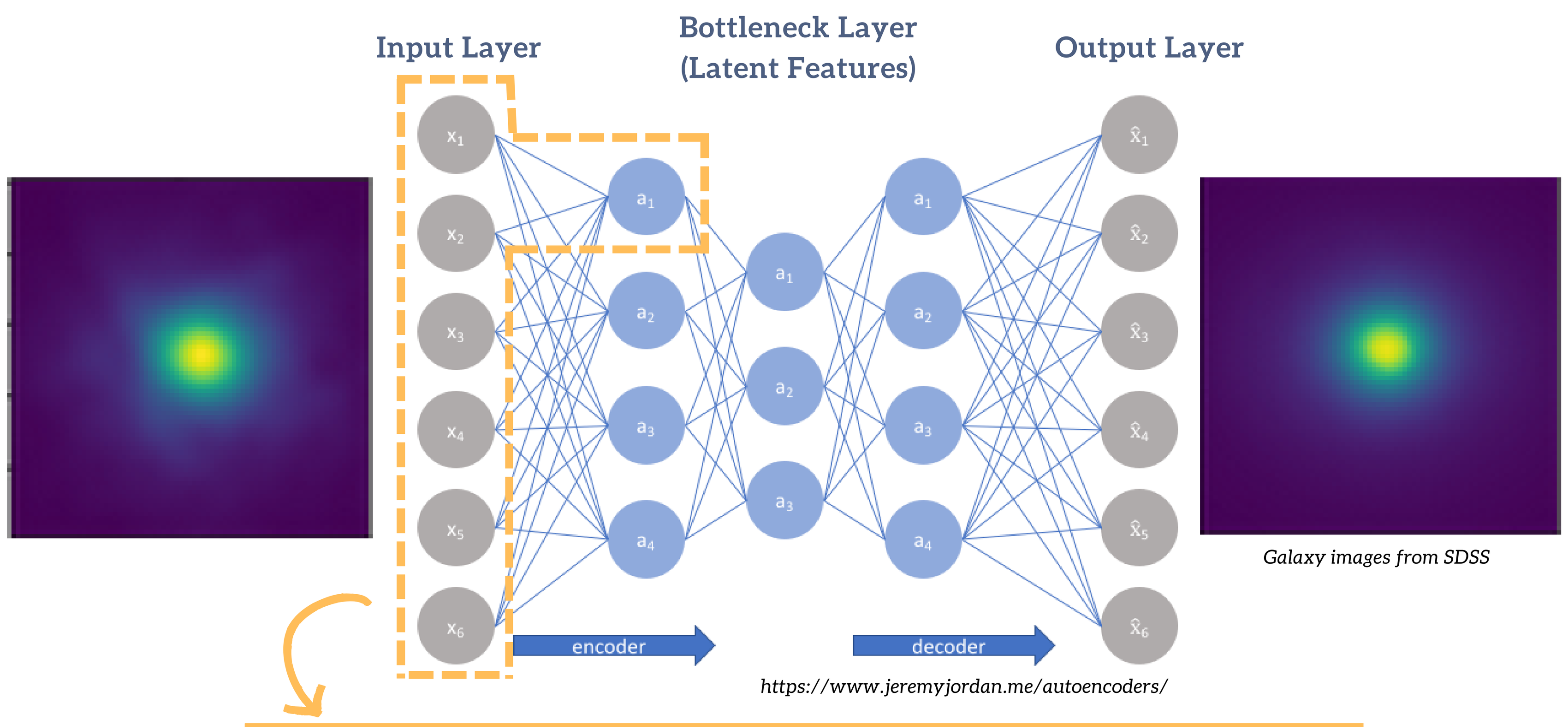
Since there are well-defined relationships between the morphology of a galaxy and its physical properties, **we propose a method of estimating a galaxy's properties by examining the presence of morphological features in their images instead**. We can then make estimations on a galaxy's properties from just its appearance, even in the absence of other data.

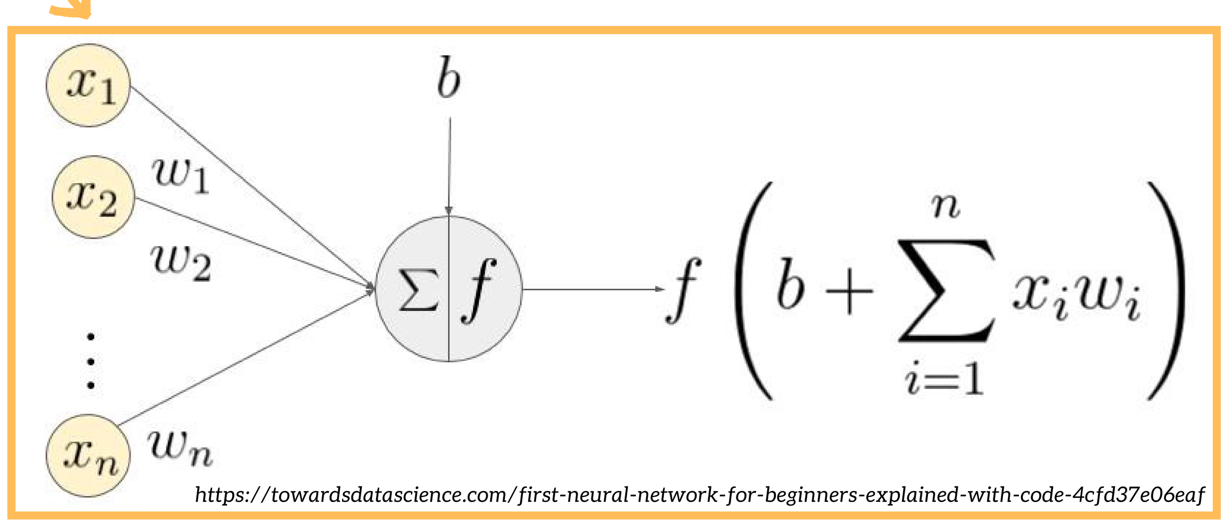
Previous works have used Principal Component Analysis (PCA) to extract features from images (Uzeirbegovic et al., 2020). **We expand on this by training a non-linear unsupervised neural network to find such latent features instead.**

We then discuss in our work the relationships between the latent features we find and a galaxy's various properties, along with its use in predictive modelling.

## METHOD

### AUTOENCODER




Autoencoders are a type of neural network that:

- **Learns to find a compressed low-dimensional representation (vector) of our input image + how to reconstruct it from that representation.**
- Trains by comparing the input and output images and adjusts the weights ($w_1$, $w_2$ ...) between each layer's nodes to minimise the difference/reconstruction loss.

### VARIATIONAL AUTOENCODER (VAE)

Unlike in basic autoencoders:

- VAEs find a probability distribution for each latent feature instead of a fixed value.
- This leads to smoother distributions of galaxies along each latent feature, which allows for better physical interpretation.
- We impose a new loss term (Kullback-Leibler Divergence) in our loss function to encourage our latent features to be normally distributed.

## RESULTS

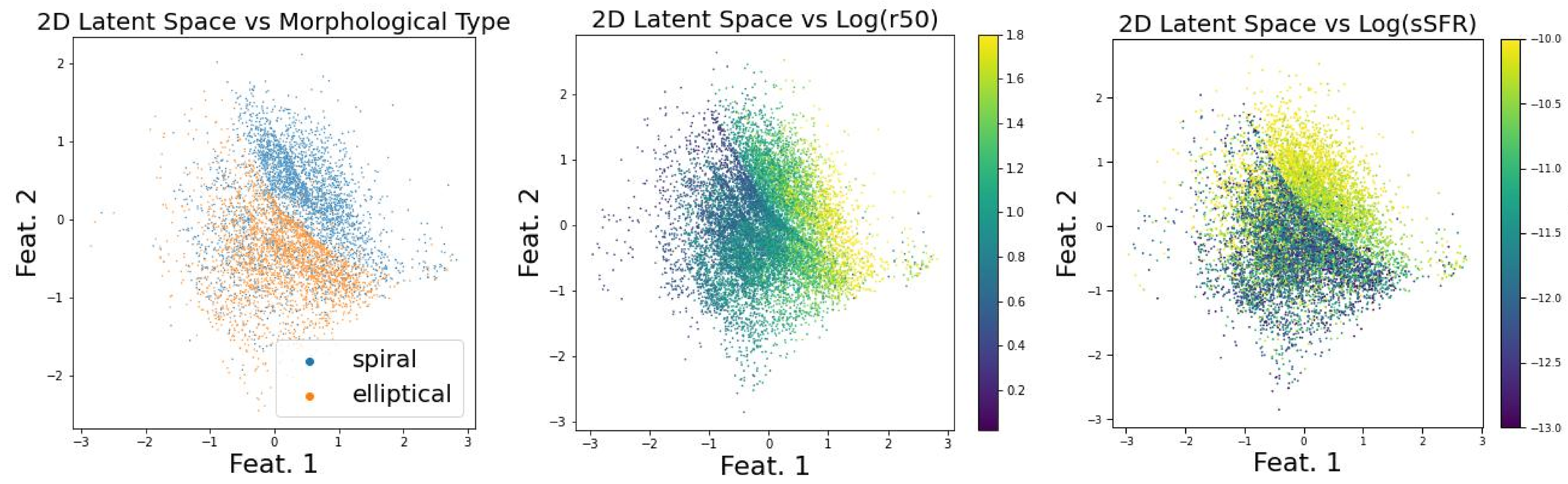



*Fig.1: Galaxy morphological classes (spirals/ellipticals) are distinctly clustered.*
*Fig. 2, 3: Physical properties like effective half light radius (r50) and specific stellar formation rate (sSFR) gradate smoothly across the latent features.*

We trained our autoencoder with images of non-irregular galaxies, and found a latent space where the way galaxy distribution show relations with different morphological and physical properties:

- Clear difference in where different morphological classes are found in the latent space. (Ellipticals are at lower Feat. 1/Feat.2 )
- **The position of the galaxy in the latent space (i.e. how much of each feature it has) is related to its physical properties.**

We see that these results make intuitive sense, as **greater proportion of Feat.1 and lower proportion of Feat. 2 corresponds to more significant diffused disc-like structures (Fig. 4)** where we would theoretically expect to be characteristic of spirals and have higher star formation rates.

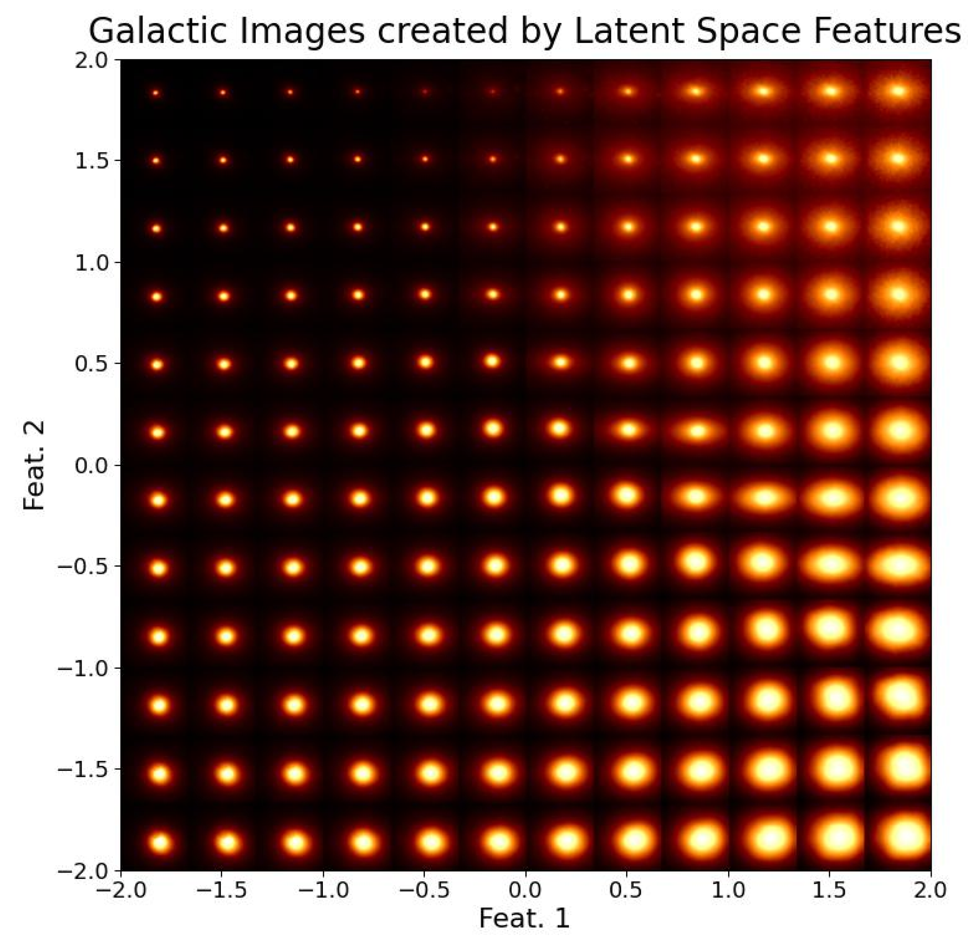



*Fig. 4: Galactic images generated from linear sums of the 2 latent features.*

It is important to note that in our training process, these properties were not known to the neural network. **Rather, it independently learnt to find features that correlated with these properties.**

In fact, even though we did not train with images of merging galaxies, we note that **merging galaxies tend to be distributed along the border between the spiral and elliptical clusters** if we encode with the same VAE.

This makes physical sense, as spiral galaxies merge to form elliptical galaxies naturally.

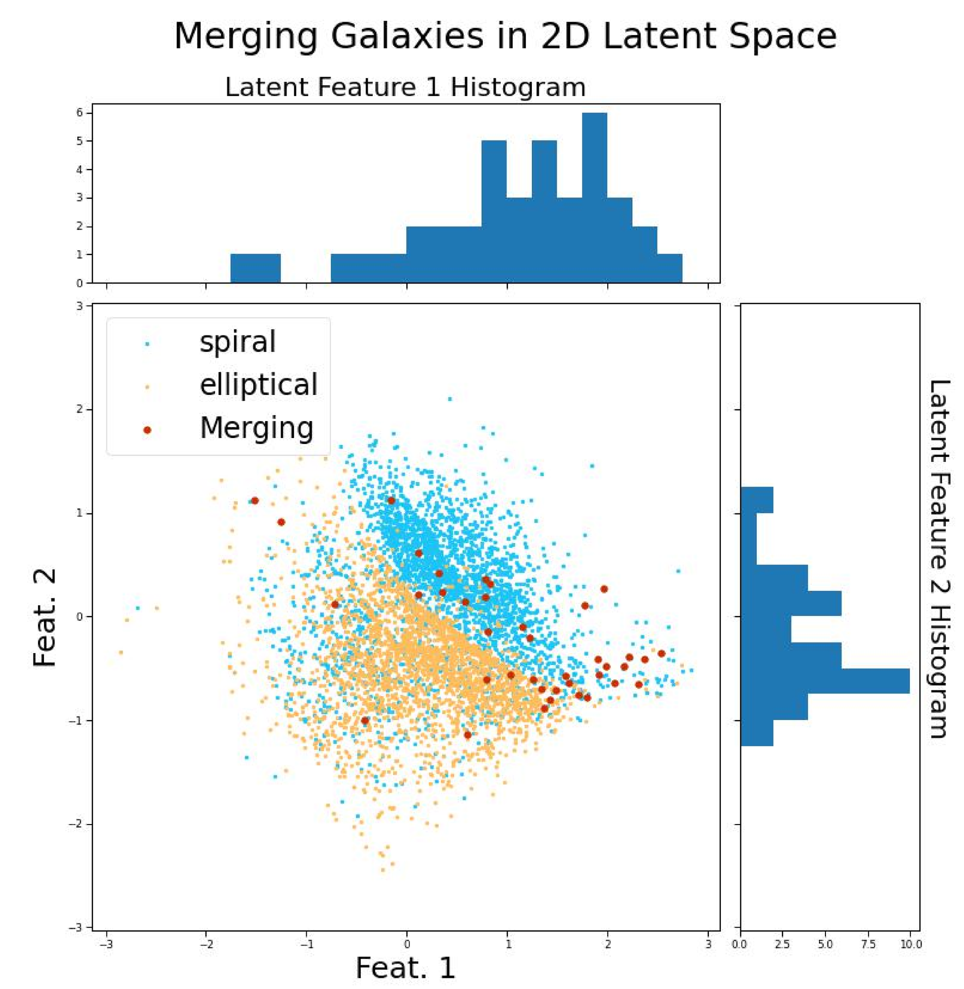



*Fig. 5: Distribution of merging galaxies in the same latent space*

## FUTURE WORK

In the future, we propose two possible developments:

1. Surface fitting on the (Latent Features) + (Physical Property) space, which will allow us to mathematically model their relationship.
2. Apply PCA on the latent features, in order to impose linear independence constraints on them.