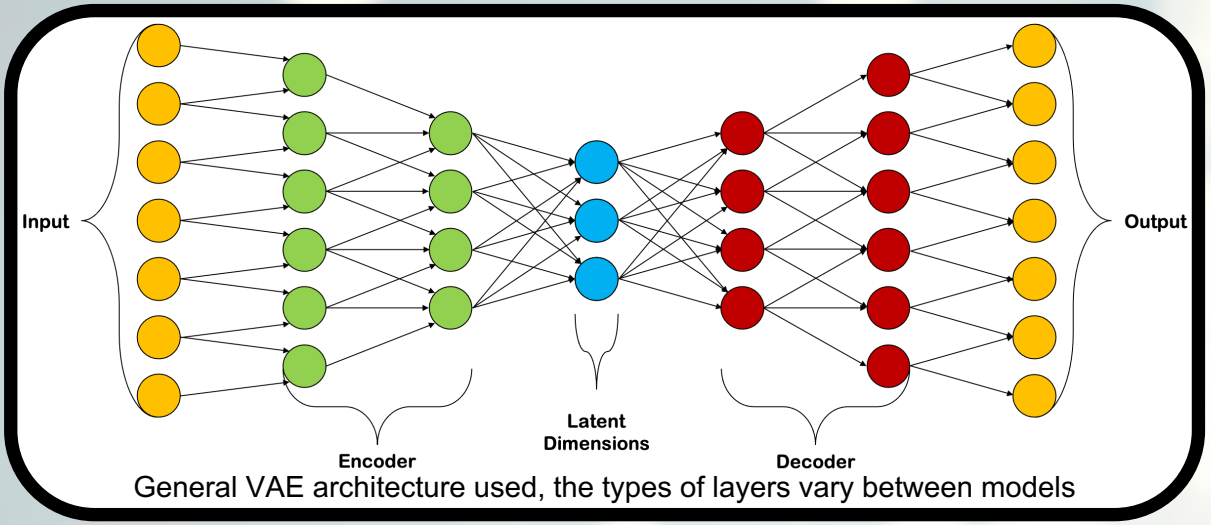
Maximum Information Retrieval From Optical Spectra

Matthew Scourfield

UCL

- This project investigates improving the SNR of galaxy spectra collected in the optical surveys, such as DESI, using **variational autoencoders** (VAEs), a type of neural network.



General VAE architecture used, the types of layers vary between models

- Autoencoders work by using an encoder to reduce a multi-dimensional input down to some smaller number of dimensions.
- A decoder then takes these latent dimensions and uses them to reconstruct the input, or in this case a de-noised version of it.

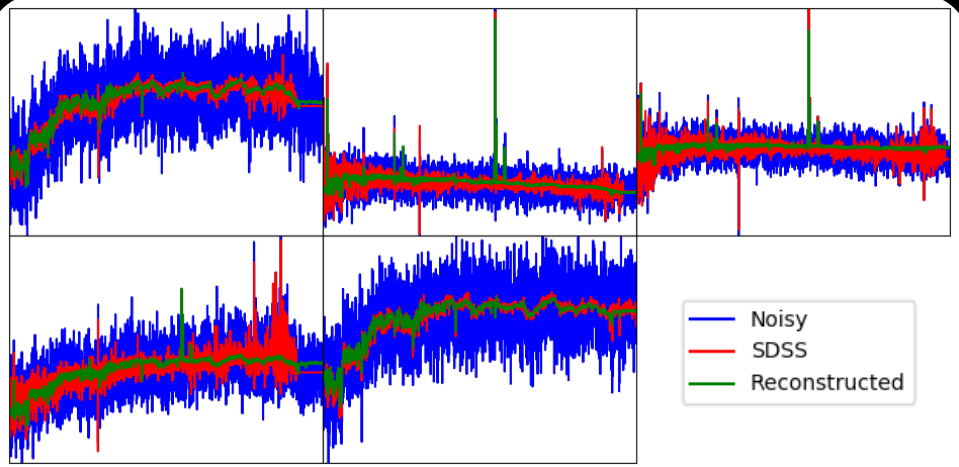
- We use 8,000 **SDSS spectra** to train our model and a further 2,000 for testing. These are processed by adding artificial noise as well as de-redshifting and normalising them.
- Two models are produced, a **convolutional** and a **dense model**, each of which uses two layers in the encoder and decoder. The output layer has a sigmoid activation function, all other layers use ReLU functions.

matt.scourfield.18@ucl.ac.uk

Maximum Information Retrieval From Optical Spectra

Matthew Scourfield

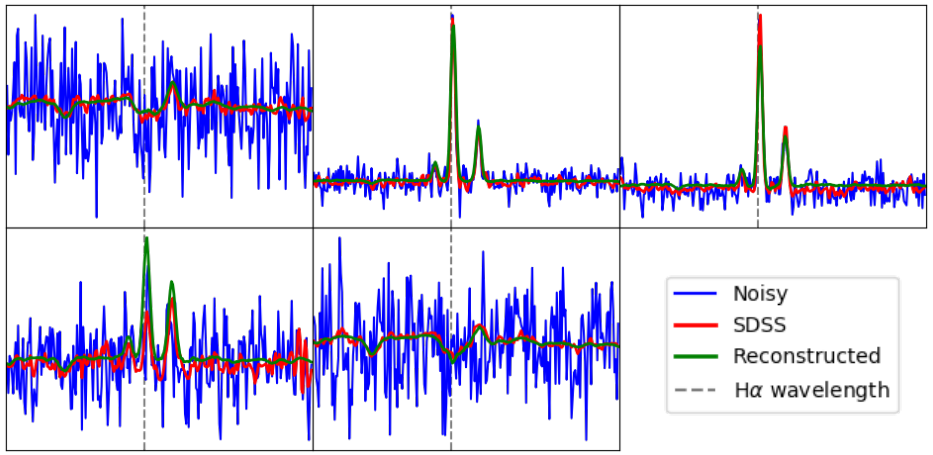
UCL



Examples of noisy, SDSS and VAE reconstructed spectra

- The convolutional model uses two **convolutional layers** in the encoder, and two **transposed convolutional layers** in the decoder.
- This model performs better at **de-noising**, producing spectra with continuum close to the general shapes of the SDSS spectra, though with less noise overlayed.

- The model is also capable of reliably identifying the position of several spectral lines and in the case of those such as Hα even distinguish between cases of absorption and emission; however, the exact amplitude of the reproduced lines is not always accurate.

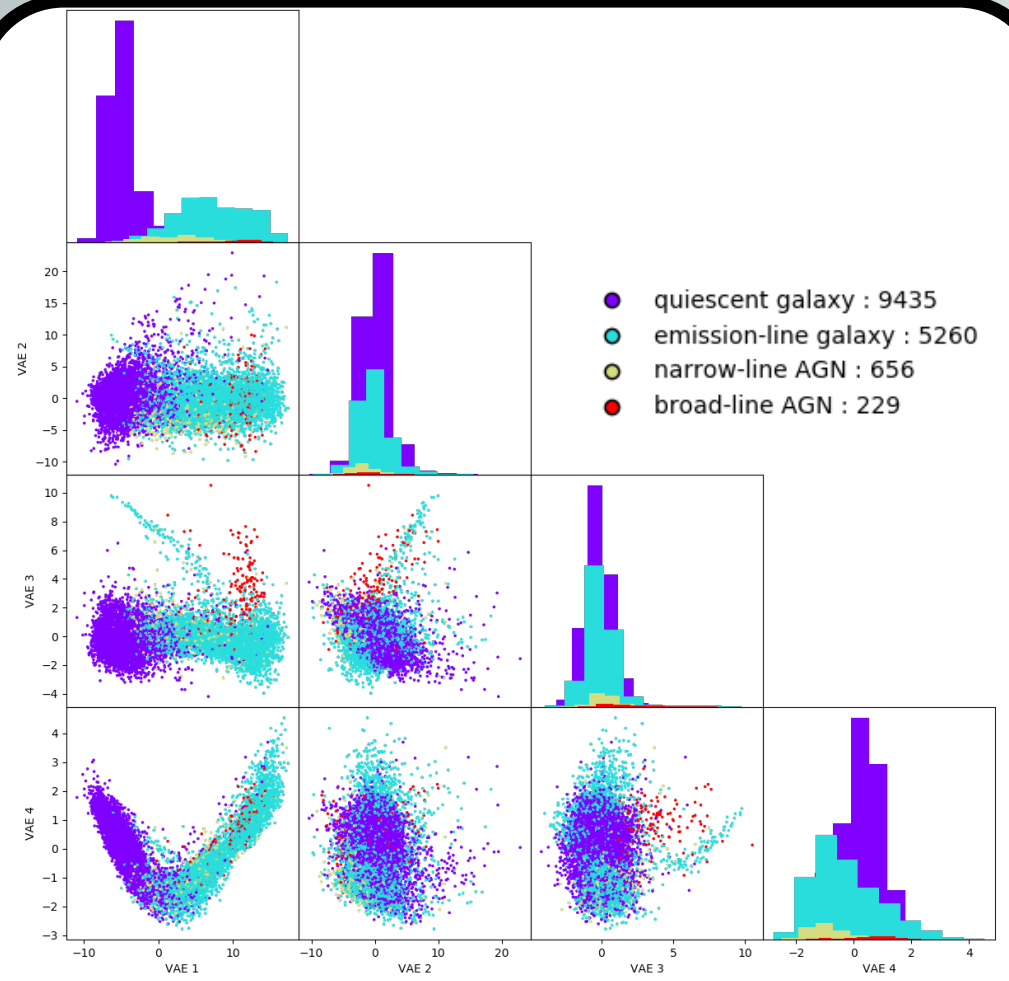


Examples of noisy, SDSS and VAE reconstructed spectra around the Hα line

matt.scourfield.18@ucl.ac.uk

Maximum Information Retrieval From Optical Spectra

Matthew Scourfield



VAE model latent space corner plot, colours correspond to spectral subclasses as in the SDSS DR14 SpecObj catalogue

- The dense model uses a pair of **dense layers** in both the encoder and the decoder. A **dropout** rate of 20% is also used with these layers to prevent overfitting.
- The **latent space** produced by the dense model more effectively separates out the different classifications of SDSS spectra in the data set, despite these not being used to train the network.
- Narrow-line AGN are not separated out, due to their low numbers and similar spectra to emission-line galaxies.

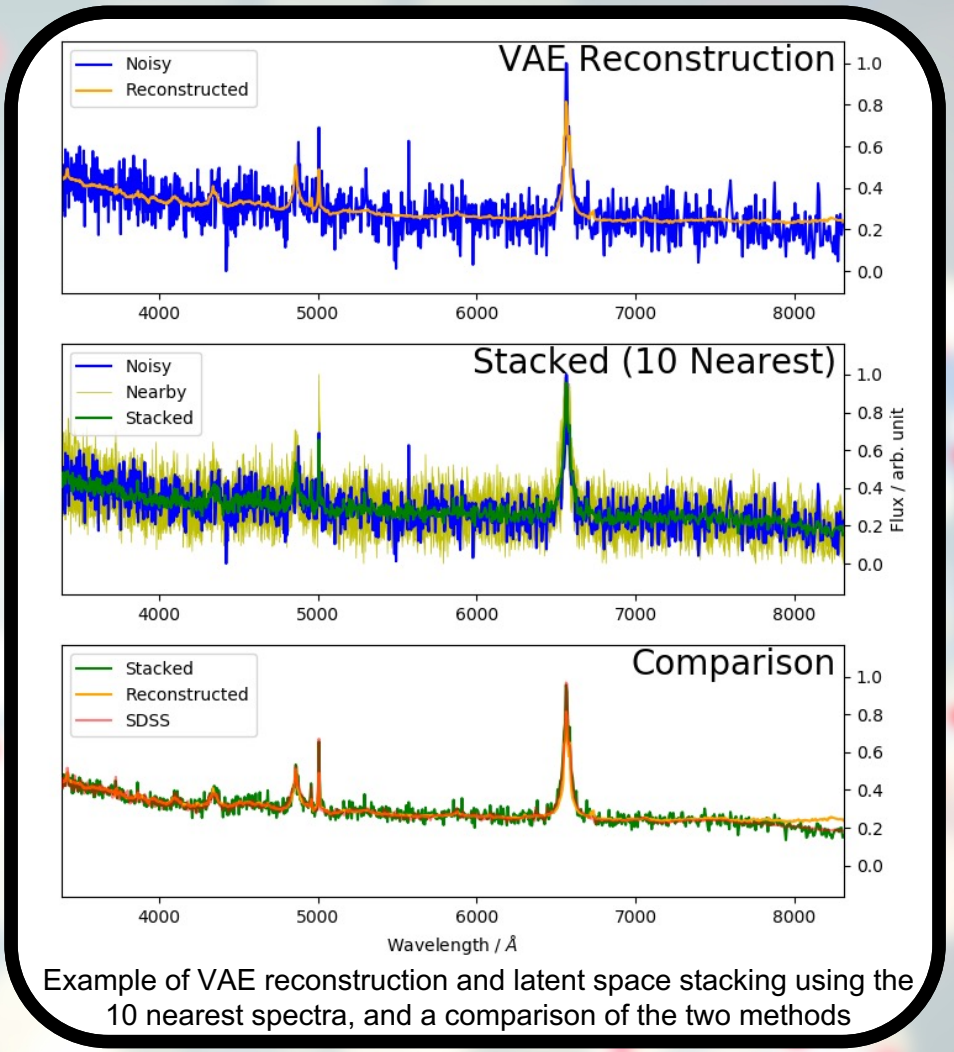
matt.scourfield.18@ucl.ac.uk

Maximum Information Retrieval From Optical Spectra

Matthew Scourfield

UCL

- Using the latent space, we can identify similar spectra by looking at their proximity within the space.
- This can be used to create **automated stacking** methods, without the need to manually select spectra.
- VAE reconstructed spectra tend to have less continuum noise, while stacked spectra more accurately reproduce line amplitudes.



Example of VAE reconstruction and latent space stacking using the 10 nearest spectra, and a comparison of the two methods

matt.scourfield.18@ucl.ac.uk