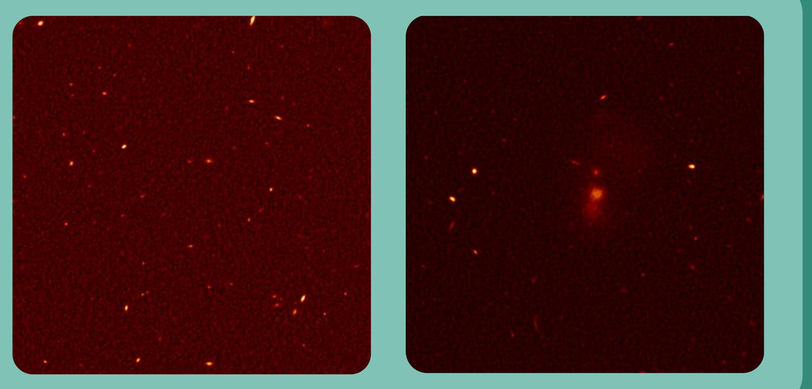
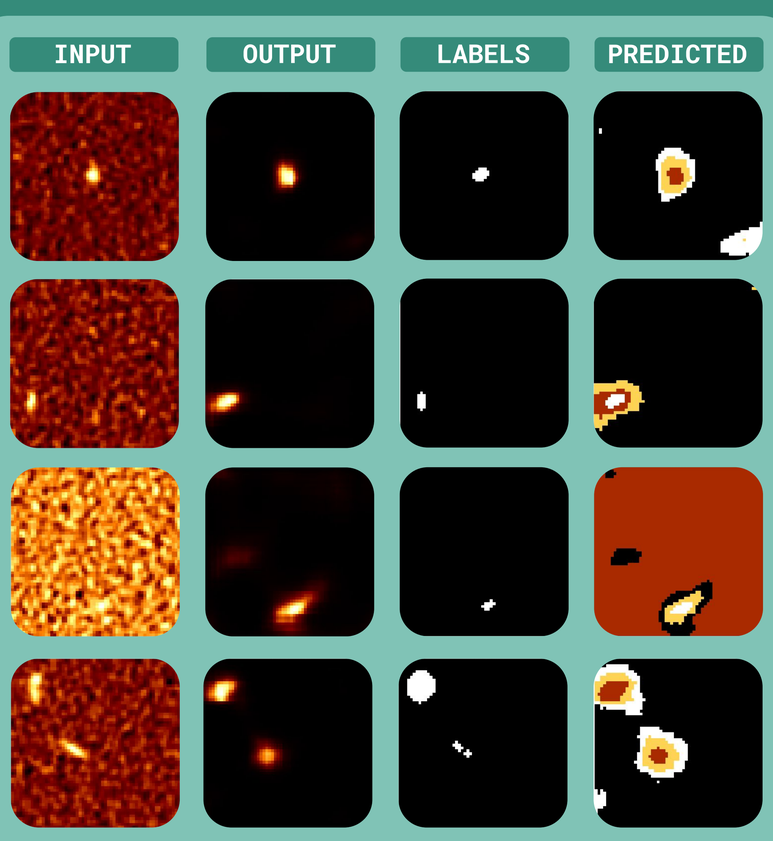
AUTOENCODERS FOR THE RADIO SKY

HATTIE STEWART, JASON YEUNG,
MARK BIRKINSHAW

RADIO ASTRONOMY'S BIG DATA PROBLEM

We are entering a new era of astronomy. With more and more large scale observatories in the pipeline, we are anticipating the influx of huge volumes of data. We hope that with Machine Learning (ML) we may be able to tackle this big data problem. The Square Kilometre Array (SKA) is a new radio interferometer currently under construction in South Africa and Western Australia. In the lead up to its launch, the SKAO are releasing model datasets for analysis by the community. The images below show tiny snapshots of the radio continuum data from SDC1, and demonstrate the variety of sources within it. Our task is to detect and classify the full source population within these datasets in an optimised and automated way.



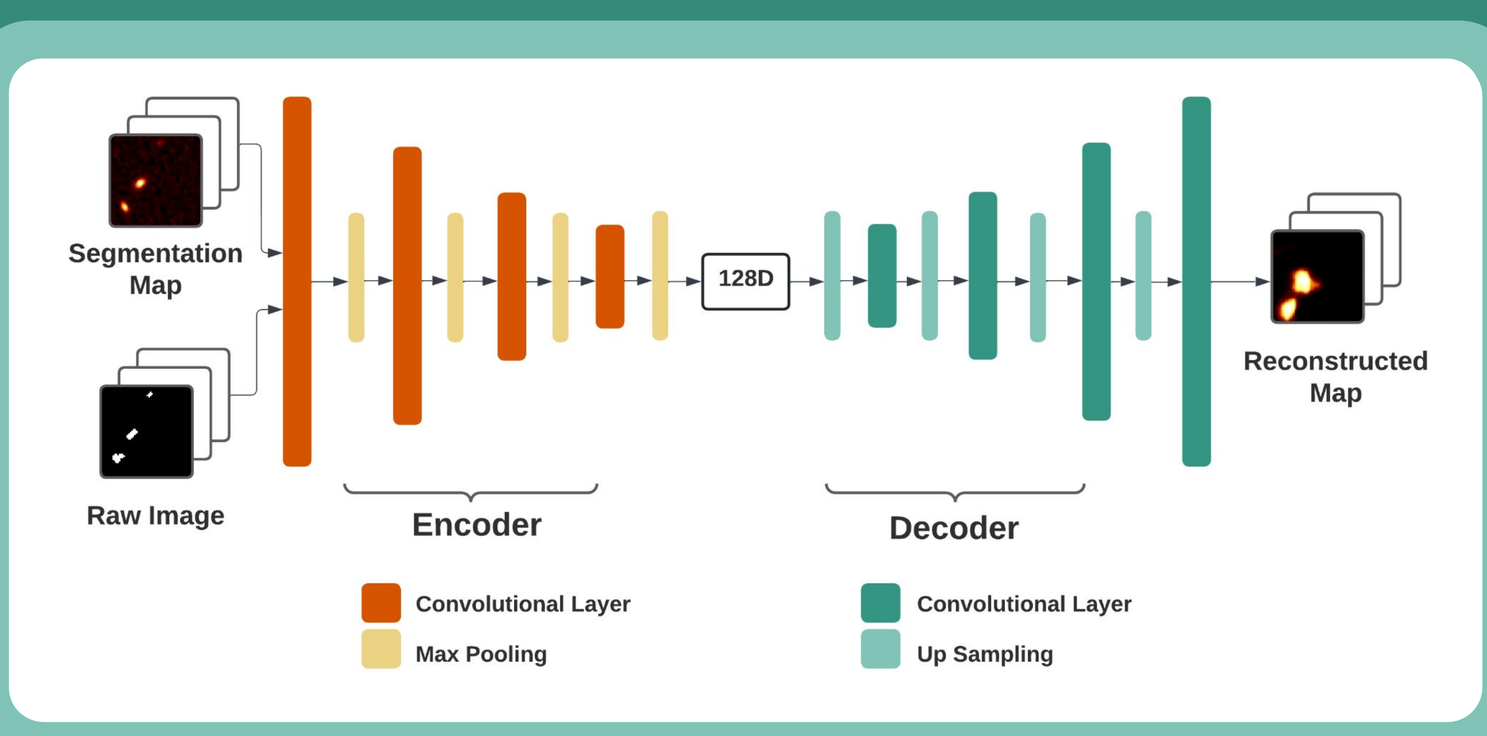


DETECTING SOURCES

We propose a Convolutional Autoencoder (CAE), presented in the figure below, as a method of image segmentation to detect the sources in the radio continuum. This network takes input data and trains against its corresponding binary segmentation map, attempting to reconstruct the map and thus learn the segmentation. New data can then be parsed to the network to recover the location of the sources in an image. The grid to the left shows the network input and output for some test cutouts. *Input* is the raw data, *output* is the reconstruction and *labels* are binary segmentation maps for each patch. *Predicted* shows the labels for the reconstruction generated by a Gaussian Mixture Model.

WHERE TO GO FROM HERE

The network has proven accurate at detecting the location of sources with a high surface brightness; detecting ~65% of the source population within the test set. We must attempt to recover the faint sources that have a low signal/noise; this could be achieved using an iterative training process. We can see that the network has struggled to capture the shape of the sources accurately. We need to recover the source parameters: axial ratio, angular size and position angle. To do this, we will train a U-Net on bright, well defined sources. We will generate new cutouts from the detected locations and predict the segmentation hyperparameters with the trained U-Net. With this combined ML approach we hope to recover the full source catalogue of the SDC1 data set.



SKAO (2022). *SKA Science Data Challenge #1.*

Dalca, A. v., Guttag, J., & Sabuncu, M. R. (2018). *Anatomical Priors in Convolutional Networks for Unsupervised Biomedical Segmentation.*

Yu, E. M., Iglesias, J. E., Dalca, A. v., & Sabuncu, M. R. (2020). *An Auto-Encoder Strategy for Adaptive Image Segmentation*.

Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-Net: Convolutional Networks for Biomedical Image Segmentation*

hattie.stewart@bristol.ac.uk

University of BRISTOL