Predicting Galaxy Parameters with Machine Learning

Clár-Bríd Tohill*, Steven Bamford, Chris Conselice, Leonardo Ferreira
clar-brid.tohill@nottingham.ac.uk

School of Physics and Astronomy, University of Nottingham, University Park, Nottingham, NG7 2RD, UK

University of Nottingham
UK | CHINA | MALAYSIA

Background

Distant galaxies are difficult to classify based on their morphology due to the lack of features caused by cosmological dimming. One method chosen to classify these galaxies is hence based on their stellar light distributions. We investigate two such parameters, the concentration and asymmetry of a galaxy. Future 'Big Data' surveys will image billions of galaxies making it computationally infeasible to calculate these values using current methods. One possible solution to this is machine learning. Machine learning networks can process much larger datasets at a much quicker rate than conventional methods. In this work we have trained a network to predict both the concentration and asymmetry values of a galaxy. Our trained network can reproduce these parameters within a reasonable scatter and is ~10,000 times faster than previous algorithms.

Data

We use ~100,000 galaxy images from the CANDELS fields imaged by the Hubble Space telescope within our work. We have measured the concentration and asymmetry values for all. We then use these images to train our network.

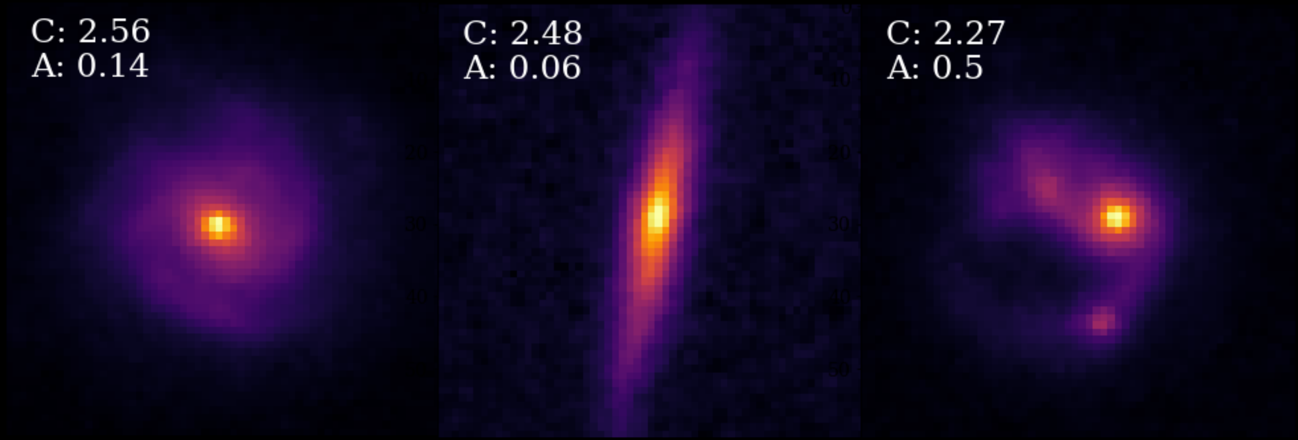


Figure: Examples of galaxy images from our sample. The Concentration and asymmetry value are indicated in each image

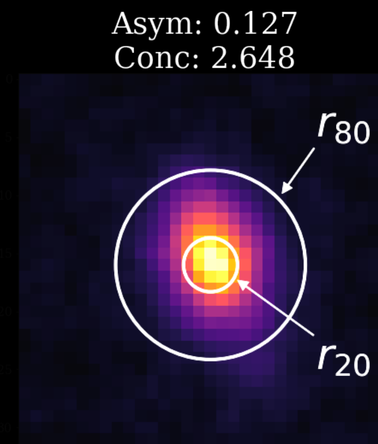
Concentration and Asymmetry

University of Nottingham
UK | CHINA | MALAYSIA

The Concentration (C) parameter is simply a measure of how concentrated the light from the central region of a galaxy is compared to the outer parts. This can tell us if a galaxy has a disk structure, is more elliptical, or is an irregular type galaxy.
It is calculated using the ratio of two radii that contain a predefined amount of light from the galaxy.
For our measurements we use the ratio of the radii containing 80% and 20% of the total light.

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

$$A = \left(\frac{|I - R_{180}|}{I}\right)$$



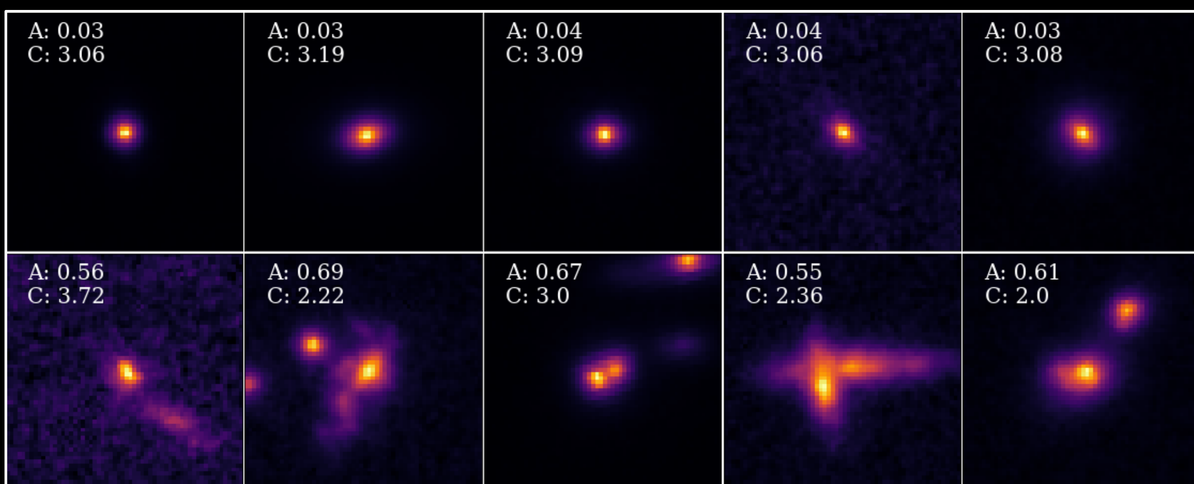


Figure: Top row: Images of the galaxies with high concentration values. These galaxies appear to be compact, spheroidal and have no close neighbours. Bottom row: Images of galaxies with low concentrations, and high asymmetry values. As it can be seen from these images these galaxies appear to be undergoing mergers. The concentration (C) and asymmetry (A) are indicated above each galaxy stamp.

The Asymmetry (A) value is a measure of how asymmetric the galaxy light is, meaning galaxies that are undergoing a merging event with another galaxy will have high A values most of the time. The A value of a large sample of galaxies can then be used to estimate the merger fraction which is important in galaxy evolution models.
It is calculated by rotating the image 180° and subtracting it from the pre-rotated image. The residuals are then summed and divided by the original galaxy flux.

Machine Learning

University of Nottingham
UK | CHINA | MALAYSIA

The idea behind machine learning (ML) and artificial intelligence is the ability to teach a machine to 'think' like the human brain does. This is achieved by using a large amount of data to train the machine. This allows the machine to learn and improve without being explicitly programmed.

Within our work we utilise a type of deep learning network known as a Convolutional Neural Network (CNNs). These are networks that are best suited for image classification problems. CNNs are made up of convolutional layers which extract and learn features from images by applying multiple filter matrices to the image (see figure). These filters can extract features like shape, size, edges etc.
Machine learning has been used in many areas of astronomy ranging from galaxy classification to merger detection.

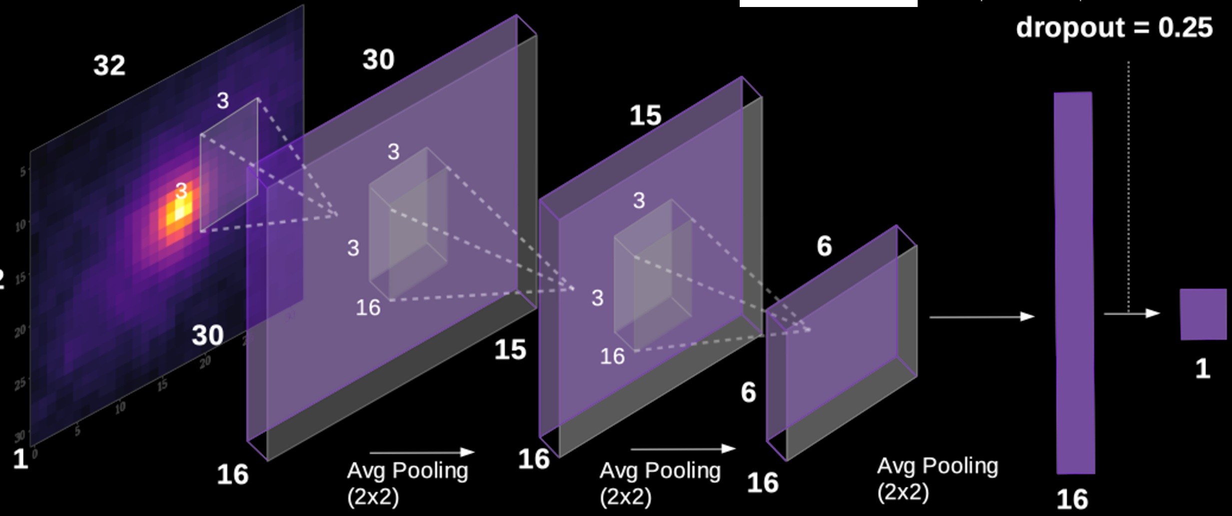


Figure: Architecture of a convolutional neural network that takes a galaxy image as the input and outputs a single value corresponding to that galaxy. This is either the concentration or asymmetry value for our network.

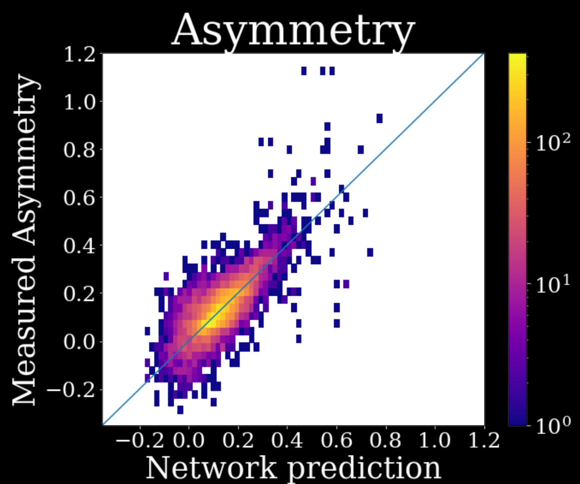
ML has several advantages over previous computational methods including:

- Much more efficient – our network is ~10,000 times faster than the original algorithm.
- Can improve over time with more data
- Less susceptible to human bias (still have this bias with labelled data sometimes)
- Requires fewer pre-processing steps as feature extraction and classification are preformed within the network.

Results

University of Nottingham
UK | CHINA | MALAYSIA

Our trained network reproduces both the C and A values within a reasonable scatter. The average difference between the network predicted value and the measured value is less than the average error on both measurements.



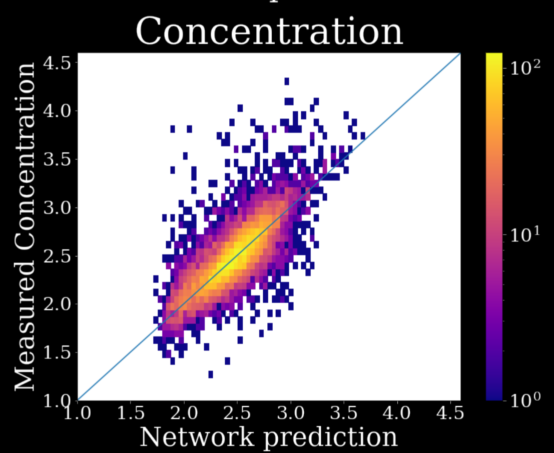


Figure: Results from the trained network showing the network's prediction of the C and A values of our galaxy sample.

When investigating those galaxies with large differences between the measured value and the networks prediction it was found they had very low signal to noise (SNR) values. This is an issue when dealing with very distant galaxies as they have much lower SNRs than local galaxies.

We tested the impact of noise on both the network and the original algorithm separately by simulating a sample of galaxies at different SNRs (see example bellow) and remeasuring their C and A values. We found that while the C measures are similar for both, the asymmetry network is more stable than the original algorithm in the low SNR regime.

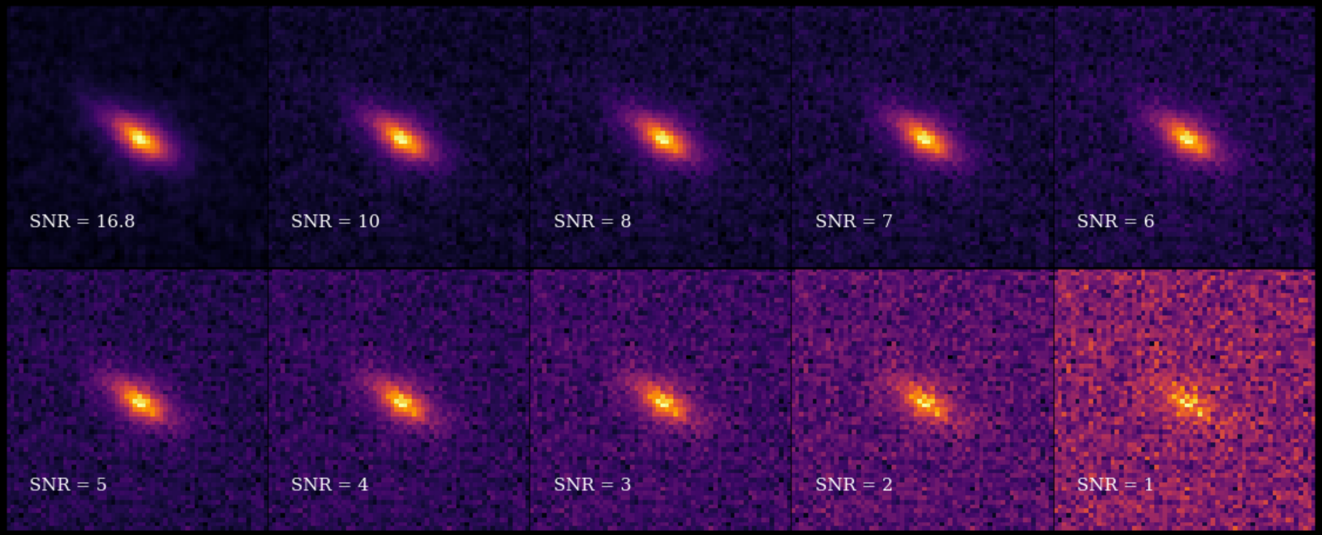


Figure: Example of a galaxy simulated at varying signal to noise levels. The SNR is indicated in each image.