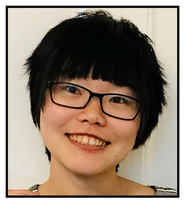


Beyond the Hubble Sequence — Exploring galaxy morphology with unsupervised machine learning

Ting-Yun Cheng (Durham), Marc Huertas-Company[1,2], Christopher J. Conselice[3], Alfonso Aragón-Salamanca[4], Brant E. Robertson[5], Nesar Ramachandra[6]

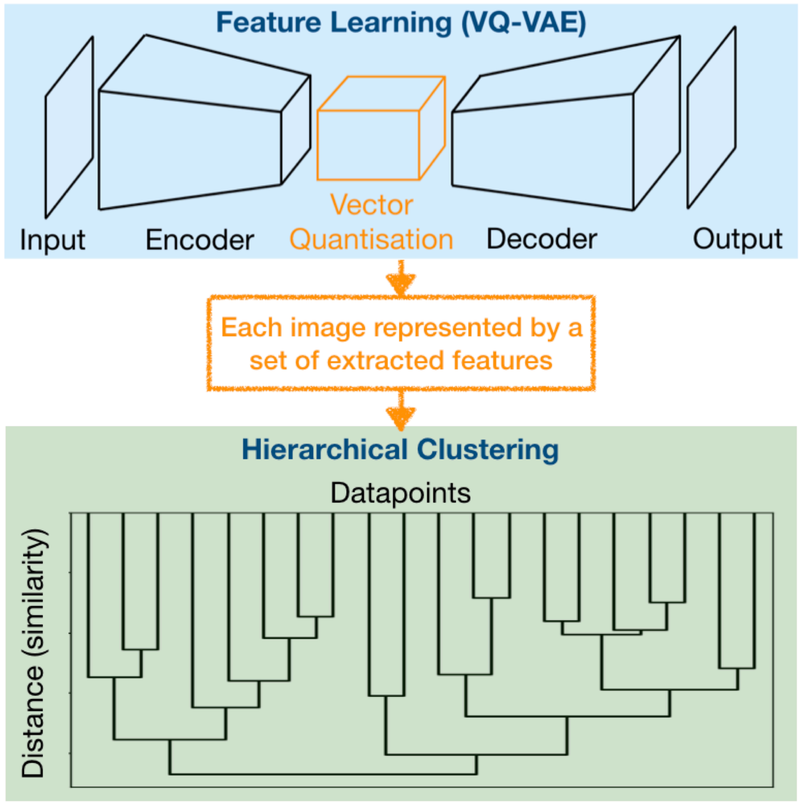
Abstract: Galaxy morphology is a fundamental property which is strongly connected with the assembly history of a galaxy. For the past century, a visual morphological classification system — *Hubble sequence* — is dominant in galaxy studies. Since visual classification can be intrinsically biased due to the subjective judgement of human classifiers, we propose the use of ***unsupervised machine learning (ML)*** to construct an unbiased machine classification scheme for future galaxy studies. In this starting work, we explore galaxy morphology from SDSS imaging data with an unsupervised technique composed of a feature extractor using vector-quantisation variational autoencoder (VQ-VAE) and hierarchical clustering. Three novel strategies are proposed in this work and results in 27 machine-defined classes which are physically distinctive from each other in stellar mass, absolute magnitude, physical size, and colour.

Paper (MNRAS)

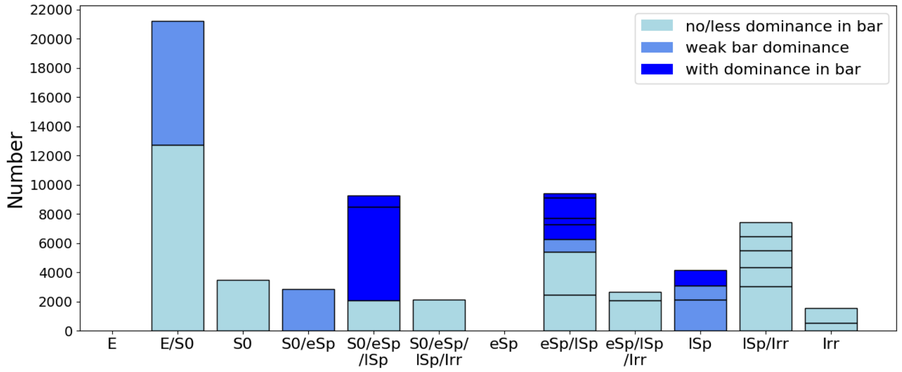
Galaxy dataset — We use *r*-band imaging data from Sloan Digital Sky Survey (SDSS) DR7 with a redshift cut z < 0.2. T-Type values from Dominguez Sanchez et al. 2018 are used to define each Hubble types (E, S0, eSp, lSp, Irr, and barred galaxies) for examination.

Methodology — The workflow (shown as below) includes a feature learning phase using VQ-VAE and a clustering step using hierarchical clustering (with hamming distance). Three following novel strategies are proposed in this work:

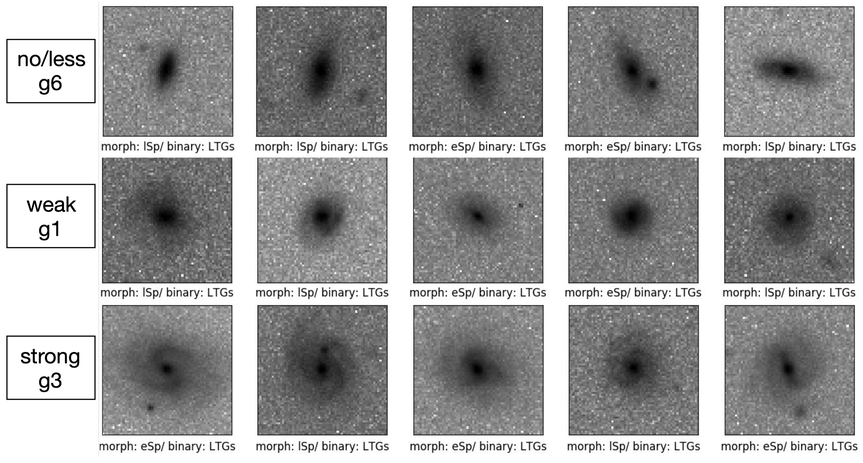
- Considering a preliminary clustering result simultaneously when learning features from images;
- Different distance thresholds are applied to different branches to allow adequate number of clusters to be generated;
- The variation caused by galaxy rotation is used to decide the final number of clusters.

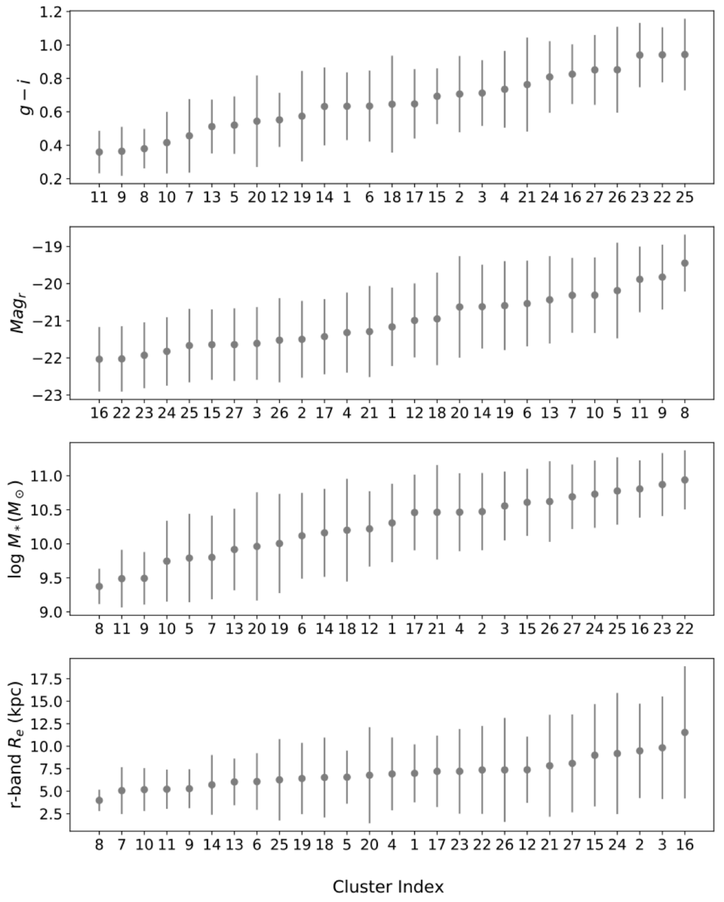


Results — Our unsupervised approach results in ***27 machine-classified classes*** without human error and bias. These classes are separated based on galaxy structure and shape. When associating the machine classes with the visual features reflected in visual morphologies, we conclude that galaxies with similar structures can be visually classified as multiple different morphology types. For example, as shown below, the structure of galaxies in two machine classes can be visually classified as either E or S0 (two leftmost bars). In other words, this indicates that the decision boundary of Hubble sequence is not precisely defined.



Results — In addition to the association with typical Hubble types (E, S0, eSp, lSp, Irr), galaxies with ***different bar strengths*** are distinguished by our unsupervised machine and categorise them into an individual class. Examples of machine classes containing galaxies with different bar strengths are shown below.





Results — Each machine class shows ***distinctive physical properties***, i.e. colour, r-band absolute magnitude, stellar mass, and physical size of galaxies. A smooth change in physical properties between different machine classes is presented.

Summary & Future Plan — Hubble sequence, as a visual classification system, has been widely used in galaxy studies. We reveal the ambiguity in its decision boundary using unsupervised ML. With the coming era of high precision astronomical imaging facilities (e.g. JWST), a revolution in ***how to classify galaxies*** is essential for future galaxy studies. A main future plan is to build a machine classification scheme that can be used unbiasedly by different individuals and different surveys using unsupervised ML.

1. Instituto de Astrofisica de Canarias (IAC), Spain. 2. LERMA, Observatoire de Paris, CNRS, PSL, Université Paris Diderot, France.
3. Jodrell Bank Centre for Astrophysics, University of Manchester, UK. 4. School of Physics and Astronomy, University of Nottingham, UK.
5. Department of Astronomy and Astrophysics, UCSC, USA. 6. High Energy Physics Division, Argonne National Laboratory, USA.