Galaxy Zoo: Hubble - crowdsourced morphological classifications for 169,944 galaxies at $0 < z < 2.5$

Kyle Willett, Melanie Galloway, Chris Lintott, Brooke Simmons, Steven Bamford, Karen Masters, Tom Melvin, Edmond Cheung, Kevin Schawinski, Claudia Scarlata, Lucy Fortson, Melanie Beck, & the Galaxy Zoo Science Team

Minnesota (KW,MG,CS,LF,MB); Oxford (CL,BS); Nottingham (SB); Portsmouth (KM, TM); Kavli IPMU (EC); ETH Zürich (KS)

Abstract

The morphology of a galaxy is a powerful tracer of its dynamical state, and provides evidence of the evolutionary physical processes that have shaped it. Measuring galaxy morphologies for large datasets that span a range of redshifts probes galaxy evolution in a statistical framework. Multiple surveys (including AEGIS, COSMOS, GEMS, and GOODS) based on observed-frame optical imaging from the *Hubble Space Telescope* have been completed and released, and provide a powerful community resource for studying galactic physics.

Automated classification of galaxy morphology, however, is still limited in accuracy and detail, while the total number of images makes individual annotations highly impractical. The **Galaxy Zoo: Hubble (GZH)** project provides visual morphologies of galaxies from crowdsourced classifications generated by volunteers through a web-based interface focusing on Hubble images from the ACS camera. We present an early version of the aggregated and calibrated dataset that measures the detailed morphology — including disk/elliptical/clumpy separations, plus smaller-scale structures such as bars, spiral arms, and mergers — for more than 100,000 galaxies at a median redshift of z=0.7.

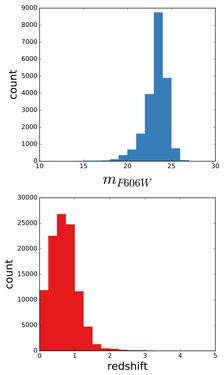


Figure 1. Distributions of physical parameters for the ~10^5 galaxies in GZH. Top: apparent magnitudes as measured in the *Hubble* F606W (V-band) filter. Bottom: redshifts (spectroscopic + photometric) for the same data.

ACS imaging

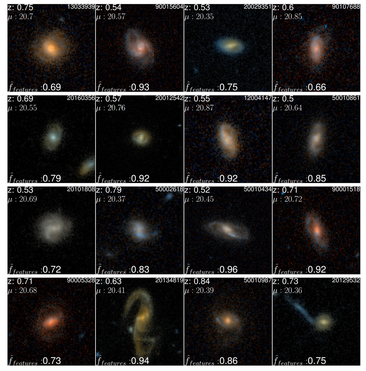
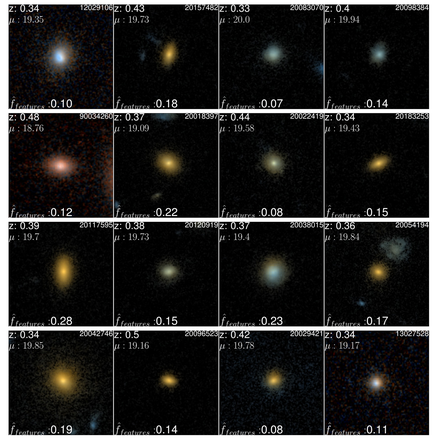
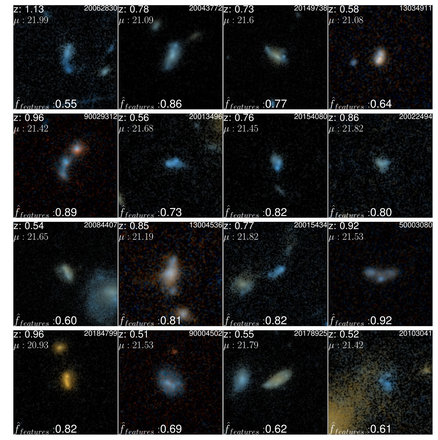


Figure 3a) Two-armed spiral disks



b) Elliptical galaxies



c) Three clumps in a chain-like arrangement

Classification scheme and interface

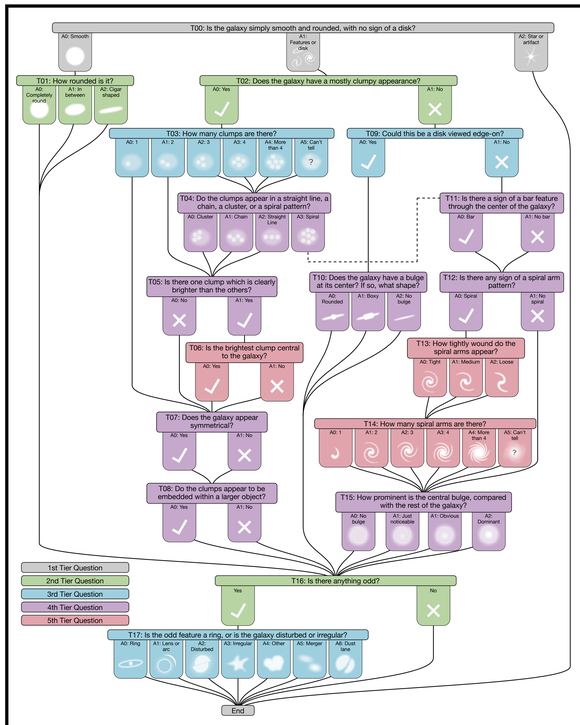


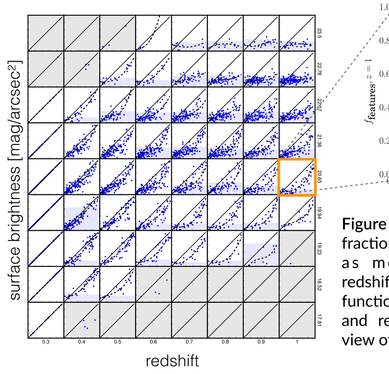
Figure 2. Classification taxonomy for Galaxy Zoo: Hubble. Questions are asked in a hierarchical fashion, beginning with the distinction between smooth galaxies, featured/disk galaxies, or stars/artifacts. The schema includes the main features of the Hubble sequence, along with distinctions between clumpy morphologies and rarer irregular classes.

Morphological classifications were collected using a web-based interface (see zoo3.galaxyzoo.org for an archived version). Volunteers were shown randomly-selected, color-composite images centered on a galaxy selected from the *Hubble* ACS general catalog (Griffith+12). Users responded to a series of increasingly-detailed questions on the galaxy's morphology; the number of questions asked depended on the specific answers given at various branching nodes (Figure 2).

Data collection for Galaxy Zoo: Hubble ran from Apr 2010 — Sep 2012. Each image of a galaxy was classified multiple times by independent users; the median for the sample is 48 classifications. GZH includes a total of more than 70 million classifications.

The data are aggregated and initially weighted based on the level of consensus with other classifiers. Weighted vote fractions for each question form the basis of the morphology categories in the GZH catalog.

Calibration and debiasing



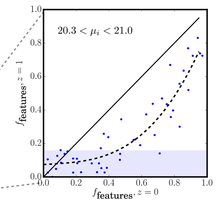


Figure 4. Left: change in vote fraction for feature/disk galaxies as measured in artificially redshifted calibration images, as a function of surface brightness (μ_i) and redshift (z). Top: zoomed-in view of one μ_i-z bin.

Raw morphological classifications are subject to known biases as a function of size, luminosity, and distance — smaller and fainter galaxies are more difficult to classify visually and thus have proportionally lower vote fractions. The GZH data is calibrated using classifications from a set of SDSS test galaxies at z < 0.1. Those images are processed using the FERENGI code (Barden+08) to appear as if they lay at a range of simulated redshifts out to z=1. The change in morphology as a function of surface brightness and redshift is applied to the HST galaxy data as a correction to their top-level morphologies. This allows direct comparisons between morphologies over the entire sample.

$$f_{\text{corr}} = 1 - (1 - f_{\text{raw}})e^{-(z - z_0)/\zeta}$$

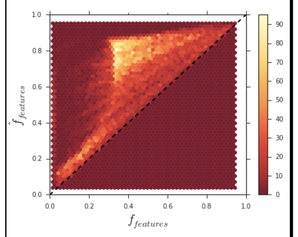


Figure 5. Effect of the debiased vote fractions on the GZH top-level data. The overall effect boosts the features/disk probabilities above the one-to-one (*dashed*) line.

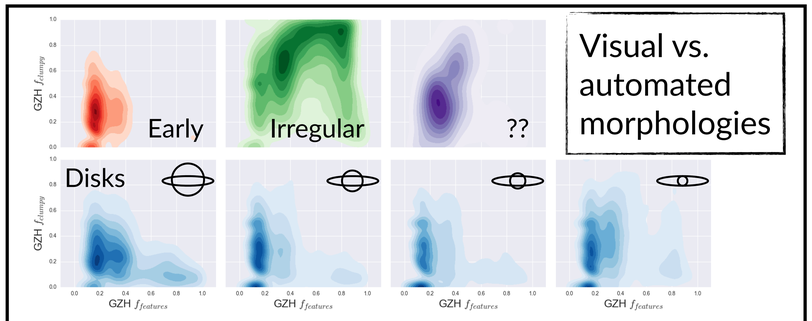


Figure 6. Comparison of Galaxy Zoo: Hubble crowdsourced morphologies against non-parametric classifications (**ZEST**; Scarlata+07) for ~75,000 COSMOS galaxies. GZH broadly agrees with the automated classes of elliptical and irregular galaxies, but predicts smooth morphologies for many galaxies where **ZEST** identifies a disk.

Data release

Galaxy Zoo: Hubble is made possible by the efforts of more than 87,000 volunteers. Their contributions are individually acknowledged at authors.galaxyzoo.org.

	hubble_id	t01_a01_count	t01_a01_weight	t01_a01_frac	t01_a01_wfrac	t01_a01_debiased	t01_a01_flag
1	10000034	17	16.01	0.548	0.609	0.608	0
2	10000162	6	6.0	0.200	0.214	0.214	0
3	10000169	3	3.0	0.091	0.111	0.111	0
4	10000189	84	84.0	0.712	0.853	0.853	1

Example of the GZH catalog structure

The full catalog will be submitted shortly as both a paper and online at data.galaxyzoo.org.

This side: Melanie Galloway's poster on **red disk galaxies** in GZ: Hubble

www.galaxyzoo.org

This side: Brooke Simmons' poster on the **Galaxy Zoo: CANDELS** project