Astronomy & Astrophysics manuscript no. output
March 19, 2025

Euclid Quick Data Release (Q1)

First visual morphology catalogue

Euclid Collaboration: M. Walmsley[⋆1,2], M. Huertas-Company[3,4,5,6], L. Quilley[7], K. L. Masters[8], S. Kruk[9], K. A. Remmelgas[9], J. J. Popp[10], E. Romelli[11], D. O'Ryan[12], H. J. Dickinson[10], C. J. Lintott[13], S. Serjeant[10], R. J. Smethurst[13], B. Simmons[14], J. Shingirai Makechemu[14], I. L. Garland[15], H. Roberts[16], K. Mantha[16], L. F. Fortson[16], T. Géron[1], W. Keel[17], E. M. Baeten[18], C. Macmillan[13], J. Bovy[1], S. Casas[19], C. De Leo[20], H. Domínguez Sánchez[21], J. Katona[22,23], A. Kovács[23,24], N. Aghanim[25], B. Altieri[9], A. Amara[26], S. Andreon[27], N. Auricchio[28], H. Aussel[29], C. Baccigalupi[30,11,31,32], M. Baldi[33,28,34], A. Balestra[35], S. Bardelli[28], A. Basset[36], P. Battaglia[28], R. Bender[37,38], A. Biviano[11,30], A. Bonchi[39], E. Branchini[40,41,27], M. Brescia[42,43], J. Brinchmann[44,45], S. Camera[46,47,48], G. Cañas-Herrera[49,50,51], V. Capobianco[48], C. Carbone[52], J. Carretero[53,54], F. J. Castander[55,56], M. Castellano[57], G. Castignani[28], S. Cavuoti[43,58], K. C. Chambers[59], A. Cimatti[60], C. Colodro-Conde[3], G. Congedo[61], C. J. Conselice[2], L. Conversi[62,9], Y. Copin[63], F. Courbin[64,65], H. M. Courtois[66], M. Cropper[67], A. Da Silva[68,69], H. Degaudenzi[70], G. De Lucia[11], A. M. Di Giorgio[71], C. Dolding[67], H. Dole[25], F. Dubath[70], C. A. J. Duncan[2], X. Dupac[9], S. Dusini[72], A. Ealet[63], S. Escoffier[73], M. Fabricius[37,38], M. Farina[71], R. Farinelli[28], F. Faustini[39,57], F. Finelli[28,74], P. Fosalba[56,55], S. Fotopoulou[75], M. Frailis[11], E. Franceschi[28], S. Galeotta[11], K. George[38], B. Gillis[61], C. Giocoli[28,34], P. Gómez-Alvarez[76,9], J. Gracia-Carpio[37], B. R. Granett[27], A. Grazian[35], F. Grupp[37,38], S. Gwyn[77], S. V. H. Haugan[78], H. Hoekstra[51], W. Holmes[79], I. M. Hook[14], F. Hormuth[80], A. Hornstrup[81,82], P. Hudelot[83], K. Jahnke[84], M. Jhabvala[85], B. Joachimi[86], E. Keihänen[87], S. Kermiche[73], A. Kiessling[79], R. Kohley[9], B. Kubik[63], K. Kuijken[51], M. Kümmel[38], M. Kunz[88], H. Kurki-Suonio[89,90], O. Lahav[86], Q. Le Boulc'h[91], A. M. C. Le Brun[92], D. Le Mignant[93], P. Liebing[67], S. Ligori[48], P. B. Lilje[78], V. Lindholm[89,90], I. Lloro[94], G. Mainetti[91], D. Maino[95,52,96], E. Maiorano[28], O. Mansutti[11], S. Marcin[97], O. Marggraf[98], M. Martinelli[57,99], N. Martinet[93], F. Marulli[100,28,34], R. Massey[101], S. Maurogordato[102], H. J. McCracken[83], E. Medinaceli[28], S. Mei[103,104], M. Melchior[105], Y. Mellier[106,83], M. Meneghetti[28,34], E. Merlin[57], G. Meylan[107], A. Mora[108], M. Moresco[100,28], L. Moscardini[100,28,34], R. Nakajima[98], C. Neissner[109,54], R. C. Nichol[26], S.-M. Niemi[49], J. W. Nightingale[110], C. Padilla[109], S. Paltani[70], F. Pasian[11], K. Pedersen[111], W. J. Percival[112,113,114], V. Pettorino[49], S. Pires[29], G. Polenta[39], M. Poncet[36], L. A. Popa[115], L. Pozzetti[28], F. Raison[37], R. Rebolo[3,116,117], A. Renzi[118,72], J. Rhodes[79], G. Riccio[43], M. Roncarelli[28], B. Rusholme[119], R. Saglia[38,37], Z. Sakr[120,121,122], A. G. Sánchez[37], D. Sapone[123], B. Sartoris[38,11], J. A. Schewtschenko[61], P. Schneider[98], T. Schrabback[124], M. Scodeggio[52], A. Secroun[73], G. Seidel[84], M. Seiffert[79], S. Serrano[56,125,55], P. Simon[98], C. Sirignano[118,72], G. Sirri[34], L. Stanco[72], J. Steinwagner[37], P. Tallada-Crespí[53,54], D. Tavagnacco[11], A. N. Taylor[61], H. I. Teplitz[126], I. Tereno[68,127], N. Tessore[86], S. Toft[128,129], R. Toledo-Moreo[130], F. Torradeflot[54,53], I. Tutusaus[121], E. A. Valentijn[131], L. Valenziano[28,74], J. Valiviita[89,90], T. Vassallo[38,11], G. Verdoes Kleijn[131], A. Veropalumbo[27,41,40], Y. Wang[126], J. Weller[38,37], A. Zacchei[11,30], G. Zamorani[28], F. M. Zerbi[27], I. A. Zinchenko[38], E. Zucca[28], V. Allevato[43], M. Ballardini[132,133,28], M. Bolzonella[28], E. Bozzo[70], C. Burigana[134,74], R. Cabanac[121], A. Cappi[28,102], D. Di Ferdinando[34], J. A. Escartin Vigo[37], L. Gabarra[13], J. Martín-Fleitas[108], S. Matthew[61], N. Mauri[60,34], R. B. Metcalf[100,28], A. Pezzotta[135,37], M. Pöntinen[89], C. Porciani[98], I. Risso[136], V. Scottez[106,137], M. Sereno[28,34], M. Tenti[34], M. Viel[30,11,32,31,138], M. Wiesmann[78], Y. Akrami[139,140], I. T. Andika[141,142], S. Anselmi[72,118,143], M. Archidiacono[95,96], F. Atrio-Barandela[144], C. Benoist[102], K. Benson[67], D. Bertacca[118,35,72], M. Bethermin[145], L. Bisigello[35], A. Blanchard[121], L. Blot[146,143], H. Böhringer[37,147,148], M. L. Brown[2], S. Bruton[149], F. Buitrago[150,127], A. Calabro[57], B. Camacho Quevedo[56,55], F. Caro[57], C. S. Carvalho[127], T. Castro[11,31,30,138], F. Cogato[100,28], A. R. Cooray[151], O. Cucciati[28], S. Davini[41], F. De Paolis[152,153,154], G. Desprez[131], A. Díaz-Sánchez[155], J. J. Diaz[4,3], S. Di Domizio[40,41], J. M. Diego[21], P.-A. Duc[145], A. Enia[33,28], Y. Fang[38], A. G. Ferrari[34], A. Finoguenov[89], A. Fontana[57], A. Franco[153,152,154], K. Ganga[103], J. García-Bellido[139], T. Gasparetto[11], V. Gautard[156], E. Gaztanaga[55,56,157], F. Giacomini[34], G. Gozaliasl[158,89], M. Guidi[33,28], C. M. Gutierrez[159], A. Hall[61], W. G. Hartley[70], S. Hemmati[119], C. Hernández-Monteagudo[117,3], H. Hildebrandt[160], J. Hjorth[111], J. J. E. Kajava[161,162], Y. Kang[70],

V. Kansal[163,164], D. Karagiannis[132,165], K. Kiiveri[87], C. C. Kirkpatrick[87], J. Le Graet[73], L. Legrand[166,167], M. Lembo[132,133], F. Lepori[168], G. Leroy[169,101], G. F. Lesci[100,28], J. Lesgourgues[19], L. Leuzzi[100,28], T. I. Liaudat[170], A. Loureiro[171,172], J. Macias-Perez[173], G. Maggio[11], M. Magliocchetti[71], F. Mannucci[174], R. Maoli[20,57], C. J. A. P. Martins[175,44], L. Maurin[25], M. Miluzio[9,176], P. Monaco[177,11,31,30], C. Moretti[32,138,11,30,31], G. Morgante[28], C. Murray[103], S. Nadathur[157], K. Naidoo[157], A. Navarro-Alsina[98], S. Nesseris[139], F. Passalacqua[118,72], K. Paterson[84], L. Patrizii[34], A. Pisani[73,178], D. Potter[168], S. Quai[100,28], M. Radovich[35], P.-F. Rocci[25], G. Rodighiero[118,35], S. Sacquegna[152,153,154], M. Sahlén[179], D. B. Sanders[59], E. Sarpa[32,138,31], C. Scarlata[16], J. Schaye[51], A. Schneider[168], M. Schultheis[102], D. Sciotti[57,99], E. Sellentin[180,51], F. Shankar[181], L. C. Smith[182], K. Tanidis[13], G. Testera[41], R. Teyssier[178], S. Tosi[40,136], A. Troja[118,72], M. Tucci[70], C. Valieri[34], A. Venhola[183], D. Vergani[28], G. Verza[184], P. Vielzeuf[73], N. A. Walton[182], E. Soubrie[25], and D. Scott[185]

(Affiliations can be found after the references)

March 19, 2025

ABSTRACT

We present a detailed visual morphology catalogue for *Euclid*'s Quick Release 1 (Q1). Our catalogue includes galaxy features such as bars, spiral arms, and ongoing mergers, for the 378 000 bright ($I_E < 20.5$) or extended (area $\geq$ 700 pixels) galaxies in Q1. The catalogue was created by finetuning the `Zoobot` galaxy foundation models on annotations from an intensive one month campaign by Galaxy Zoo volunteers. Our measurements are fully automated and hence fully scaleable. This catalogue is the first 0.4% of the approximately 100 million galaxies where *Euclid* will ultimately resolve detailed morphology.

Key words. Galaxies: structure – Galaxies: spiral – Catalogs, Galaxies: interactions – Galaxies: elliptical and lenticular – Methods: statistical

1. Introduction

Detailed visual morphology refers to the recognizable features which comprise a galaxy, such as bars, spiral arms, and tidal tails (Hubble 1926; De Vaucouleurs 1959; Toomre & Toomre 1972; Sellwood & Masters 2022). Understanding how galaxies acquire their stellar structure provides key insights into the processes driving mass assembly in the Universe (e.g. Wuyts et al. 2011; Tacchella et al. 2015; Huertas-Company et al. 2016) Visual morphology has historically also described the method of detection; we measure these features visually, by eye. Those eyes may either belong to professional astronomers (Nair & Abraham 2010; Baillard et al. 2011; Buta et al. 2015) or to members of the public taking part in citizen science projects such as Galaxy Zoo (Lintott et al. 2008; Masters 2019) and Galaxy Cruise (Tanaka et al. 2023). Visual morphology complements parametric morphology, such as Sérsic fitting (Sérsic 1963), and non-parametric morphology, such as concentration and asymmetry (Morgan 1958; Conselice et al. 2000; Shimasaku et al. 2001; Abraham et al. 2003), which both use rule-based automated methods to interpret galaxy images. Parametric and non-parametric morphology have historically been together known as 'quantitative' morphology, contrasting with 'qualitative' visual morphology.

The complexity of galaxies is greater than the complexity we are able to express in code. Galaxies have features which are too complex for our rule-based methods, but are real nonetheless (see e.g., Lintott et al. 2009; Rudnick 2021; Bowles et al. 2023; Gordon et al. 2024). Astronomers have therefore faced a trade-off. One can use visual morphology to capture detailed features, or quantitive morphology to make measurements which are scaleable and reproducible (Conselice 2014). There is also a spectrum of work between these two extremes that makes detailed automated measurements under a degree of manual supervision and tuning, e.g., `galfit` (Peng et al. 2002) and Galaxy Zoo Builder (Lingard et al. 2020).

Recent advances in computer vision make it possible, even straightforward, to automate some visual judgements. Seminal work by Dieleman et al. (2015) won the Galaxy Challenge, a Kaggle competition to predict the visual judgements of Galaxy Zoo volunteers, and in doing so introduced deep learning to astronomy. A decade later, deep learning is a ubiquitous tool for measuring visual morphology (e.g., Khan et al. 2019; Abraham et al. 2018; Pearson et al. 2019; Ghosh et al. 2020; Bom et al. 2021; Ćiprijanović et al. 2022 and review by Huertas-Company & Lanusse 2023). Citizen science and deep learning have together underpinned detailed visual morphology catalogues for the *Hubble* Space Telescope (Huertas-Company et al. 2015), the Sloan Digital Sky Survey (Domínguez Sánchez et al. 2018), the Dark Energy Camera Legacy Survey (Walmsley et al. 2022a), and the companion Legacy Surveys (Walmsley et al. 2023b; Ye et al. 2025).

Our core advance here is timing. Morphology catalogues typically follow years after a telescope data release – 3.5 to 5 years for each of Galaxy Zoo's morphology catalogues, for example. Much of this time is needed for volunteers to annotate galaxies and, more recently, to train models.

What if we made morphology measurements at the same time as the survey takes images, just as we already do for other automated measurements? Placing a trained deep learning model within the survey image processing pipeline allows for immediate morphology measurements and immediate use by scientists. Our models can be trained quickly because we use new 'foundation' models (described in Sect. 3.2) that need fewer examples to learn to classify new surveys. In this work, we deliver a detailed visual morphology catalogue for *Euclid* in weeks instead of years.

Euclid will resolve the detailed visual morphology of at least an order of magnitude more galaxies than have ever been measured. The largest current detailed morphology catalogues use

⋆ e-mail: `m.walmsley@utoronto.ca`

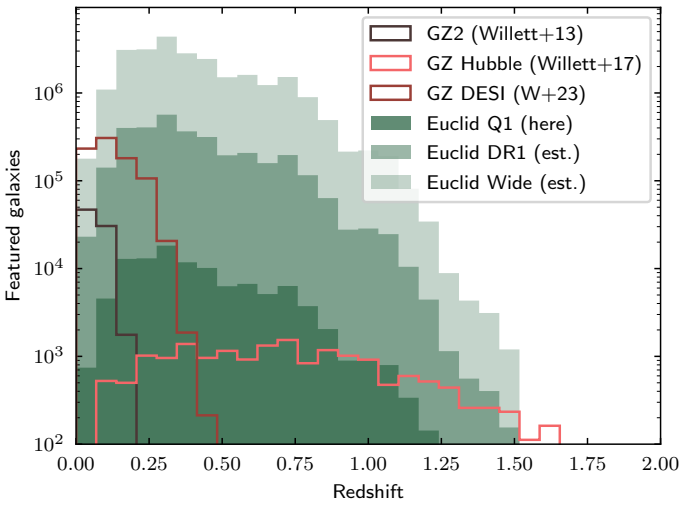

Fig. 1. Galaxies with well-resolved features as a function of redshift, for morphology catalogues from *Euclid* vs. Galaxy Zoo 2 (main sample, Willett et al. 2013), Galaxy Zoo Hubble (Willett et al. 2017), and Galaxy Zoo DESI (Walmsley et al. 2023b). *Euclid* will outscale all current visual morphology catalogues for all redshifts below $z < 1.5$, typically by several orders of magnitude. Q1 morphologies may be especially valuable between $0.3 < z < 0.7$. We define 'featured' as a galaxy with an expected volunteer vote fraction for 'Featured' above 0.5 (see Sect. 5). 'Euclid Q1' is this catalogue; estimates for 'Euclid DR1' and 'Euclid Wide' (EWS) are made by trivially multiplying our catalogue counts by area. *Euclid* redshifts are photometric (Euclid Collaboration: Tucci et al. 2025).

images from the DESI Legacy Surveys (Walmsley et al. 2023b), with 19 000 deg^2 of imaging at 1.″1 seeing; and the Sloan Digital Sky Survey (Willett et al. 2013; Domínguez Sánchez et al. 2018), with 9000 deg^2 of imaging at 1.″3 seeing (DR7, Abazajian et al. 2009). The EWS will cover approximately 14 000 deg^2 with a spatial resolution of 0.″16 (Euclid Collaboration: Cropper et al. 2024) – a comparable area at ten times higher resolution. The final *Euclid* morphology catalogues will include approximately 10^8 galaxies. Here, we measure detailed morphology in the first 0.4% – Euclid Quick Release 1 (Q1, Euclid Quick Release Q1 2025).

Euclid connects low-redshift ground-based morphology measurements with high-redshift space-based measurements, enabling a continuous view of galaxy morphology through time. Figure 1 compares, as a function of redshift, the number of galaxies with visual features in our Q1 catalogue vs. previous catalogues made with the Sloan Digital Sky Survey (Willett et al. 2013), the *Hubble* Space Telescope (HST, Willett et al. 2017), and the Legacy Surveys (Walmsley et al. 2023b). Q1 adds an order-of-magnitude more galaxies between $0.3 < z < 0.7$. Straightforwardly multiplying our results by area, the EWS will ultimately increase the number of galaxies with measured morphology features between $0.3 < z < 0.7$ by around three orders of magnitude.

Our Q1 catalogue is available in two forms. First, our initial trained model is part of the *Euclid* pipeline, and so the morphology measurements from that model are reported as part of the Q1 data release (Euclid Collaboration: Romelli et al. 2025). Those measurements are accessible through the ESA Science Archive Service as with other core measurements such as photometry, redshifts, and so forth. We refer to this as the *pipeline catalogue*. Second, we created a separate catalogue by applying our next generation of models directly to the *Euclid* images, outside of the *Euclid* pipeline. We did this to use the best possible models (which are updated more frequently than is practical within the *Euclid* pipeline) and to create and share our embeddings (vectors which mathematically summarise the visual features of each galaxy). We refer to this as the *dynamic catalogue*.

Our catalogue complements parallel work by Euclid Collaboration: Romelli et al. (2025) and Euclid Collaboration: Quilley et al. (2025) to create a morphology catalogue for Q1 with parametric and non-parametric measurements. We recommend using these traditional measurements for galaxies less extended than around 700 pixels in segmentation area[1], below which *Euclid* cannot reliably resolve detailed features. Euclid Collaboration: Quilley et al. (2025) includes a comparison of disk and bulge measurements using this detailed morphology catalogue and using Sérsic fits and finds consistent results.

Our catalogue was made possible by the efforts of 9976 Galaxy Zoo volunteers who together contributed 2.9M annotations to adapt the `Zoobot` foundation deep learning models for *Euclid* images. These measurements, combined with parallel work using the `Zoobot` models to find strong lenses (Euclid Collaboration: Walmsley et al. 2025; Euclid Collaboration: Rojas et al. 2025; Euclid Collaboration: Lines et al. 2025; Euclid Collaboration: Li et al. 2025; Euclid Collaboration: Holloway et al. 2025), stellar bars (Euclid Collaboration: Huertas-Company et al. 2025), mergers (Euclid Collaboration: La Marca et al. 2025) and AGN (Euclid Collaboration: Margalef-Bentabol et al. 2025) demonstrate the practical value of foundation models in astronomy.

In Sect. 2, we describe our selection function and image processing choices. In Sect. 3.1, we describe how Galaxy Zoo volunteers contributed annotations. In Sect. 3.2, we motivate our use of foundation models and detail the finetuning process. In Sect. 4, we validate the performance of our finetuned models. In Sect. 5, we share our dynamic catalogue, embeddings, and images, and provide practical guidance on how these might be used. They can be downloaded from Zenodo[2] and HuggingFace[3].

2. Data

2.1. Coverage

Euclid will detect approximately 1.5 billion sources (Euclid Collaboration: Bretonnière et al. 2022; Euclid Collaboration: Mellier et al. 2024). The largest sources will be revealed in exquisite detail (Hunt et al. 2024). Most will be barely resolved. In between will be a middle ground of sources which show some suggestion of detailed morphology (the trace of a disc, an arm, a bar, etc.). When choosing which galaxies to measure for detailed morphology, where should we draw the line?

The human annotations guiding the models that, in turn, create our catalogue, come from Galaxy Zoo volunteers – members of the public contributing their time to click through galaxy images (Masters 2019). We need to make the best possible use of Galaxy Zoo volunteers' time, particularly during the one month labelling campaign to produce the pipeline models (Sect. 3.1). We should especially avoid showing a high ratio of featureless

[1] As measured by SourceExtractor++ within the MERge pipeline and reported as 'SEGMENTATION_AREA', (Euclid Collaboration: Romelli et al. 2025; Bertin & Arnouts 1996). This roughly corresponds to 1.″5 in radius. MERge mosaic images have a pixel scale of 0.″1 per pixel.

[2] https://doi.org/10.5281/zenodo.15002907

[3] https://huggingface.co/collections/mwalmsley/euclid-67cf5a80e2a93f09e6e4df2c

galaxies ('blobs') as these are relatively straightforward to classify automatically and may dissuade volunteers. Therefore, we chose the following conservative cut to select galaxies with a moderate chance of showing detailed features.

`segmentation_area` > 1200 pixels
OR
I_E < 20.5 `AND segmentation_area`> 200 pixels

We found segmentation area (the total number of pixels within the segmentation source mask from `SourceExtractor++`, as calculated by Euclid Collaboration: Romelli et al. 2025) to be the critical factor in determining if a galaxy was well-resolved. Segmentation area is a natural proxy for assessing if a galaxy is well-resolved because each morphological feature requires sampling by some number of point spread function full-width-half-maximum (FWHM) to be resolved, and this sampling happens in two dimensions. Results using radii were broadly similar but suffered from orientation effects or asymmetric sources. Our choice of 1200 pixels was a subjective choice with the aim of creating an engaging sample for Galaxy Zoo volunteers (see above), and was ultimately later revised for the dynamic catalogue (below). The magnitude cut follows from the common science requirement for completeness, and is complemented by an alternative (far more generous) segmentation cut to remove galaxies where detailed features are plainly unmeasurable. Overall, this selection cut includes the brightest and most extended 0.8% of galaxies in Q1 (195 716 galaxies). These form the selection shown to volunteers and measured by the pipeline models.

For the dynamic catalogue, we reduce the `segmentation_area` cut from 1200 pixels to 700 pixels (for a total of 380 111 galaxies, 1.5% of Q1). This adds 184 395 galaxies which are fainter and less extended but may still have resolvable features. We do not (currently) show these less extended galaxies to Galaxy Zoo volunteers, and instead rely on our trained models to extrapolate to this regime. The dynamic catalogue includes the column 'in_extrapolated_selection' for users to include or exclude these additional galaxies as desired. Lacking ground truth labels, we cannot make any performance claim for these galaxies, but our expert visual inspection qualitatively suggests the models continue to work similarly well – perhaps because the images are less detailed and therefore present a less challenging computer vision task, because the segmentation area is imprecisely measured, or because the models were pretrained on similar images from other surveys (Walmsley et al. 2024).

While we could make automated measurements of every source in Q1, visually inspecting example images suggested that galaxies with a segmentation area below around 700 pixels are insufficiently resolved to clearly show detailed features, and so we select our lowest area cut as 700 pixels and defer deep learning morphology measurements of smaller galaxies to future work. Figure 2 illustrates our choice of selection cuts.

For both selections, we additionally require `vis_det`= 1 and `spurious_prob`< 0.2, to remove artifacts, and require no *Gaia* cross-match to remove stars.

2.2. Image processing

We create three jpg cutouts from each source. Figure 3 shows examples. The three cutouts are:

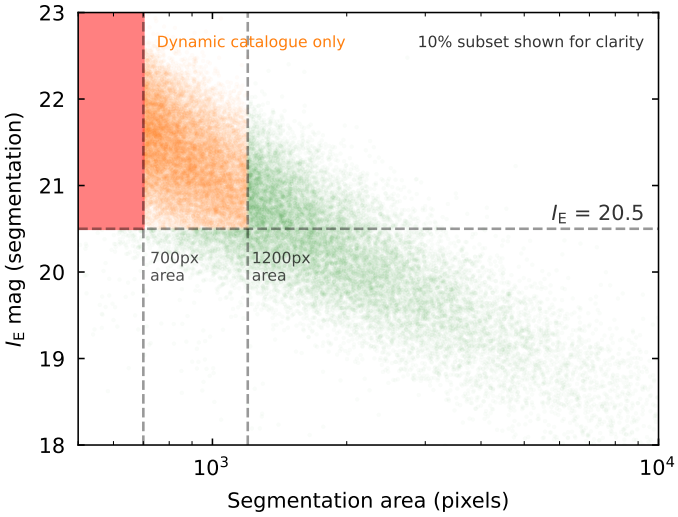

Fig. 2. Cuts applied to select galaxies with resolvable visual morphology. The pipeline catalogue includes all galaxies with area > 1200 pixels or I_E < 20.5 (and area > 200 pixels). The dynamic catalogue reduces the area limit to 700 pixels. The total number of galaxies in the dynamic catalogue only (orange) vs. in both catalogues (green) are roughly equal (195 716 vs. 184 395, respectively)

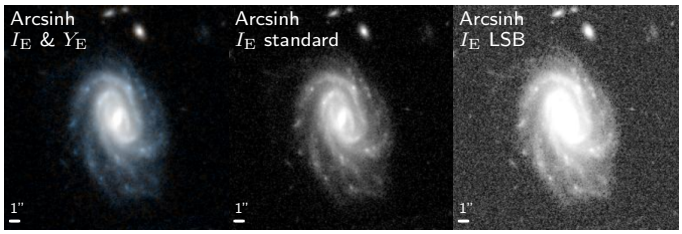

Fig. 3. Example cutouts shown to volunteers. Left to right: I_E/Y_E composite, I_E-only greyscale with standard processing, I_E-only greyscale with enhanced low-surface-brightness processing.

1. A composite RGB image where the R channel is Y_E, the B channel is I_E, and the G channel is the mean of the pixelwise flux in the other two channels, following a 99.85th percentile clip and an arcsinh stretch, i.e., $x' = \mathrm{arcsinh}(Qx)$ with $Q = 100$ where x is the flux in each pixel.
2. A greyscale image where the single channel is identical to the I_E/B channel above, maximising resolution
3. A greyscale image where the single channel is again from I_E, but adjusted to highlight low-surface brightness features. We use the recipe from Gordon et al. (2024) with a stretch of 20 and a power of 0.5, and add a 98th percentile clip.

We designed these processing options to create a complementary set of images for volunteers; a colour image showing the general galaxy features, a maximum-sharpness (but greyscale) image, and an image aimed at highlighting low-surface brightness features which are better revealed when shown on a separate scale to the bright galaxy core. We used Y_E for the colour image as Y_E is the sharpest (lowest PSF FWHM) NISP band. We combined data from different instruments using the aligned and resampled mosaics provided by the MERge pipeline (Euclid Collaboration: Romelli et al. 2025).

Volunteers were shown all three images in a flipbook format with the order above. We then use their responses to adapt our models. The pipeline catalogue is made by a model shown the standard I_E image (only I_E is available). The dynamic catalogue is made by a model shown the composite I_E/Y_E image.

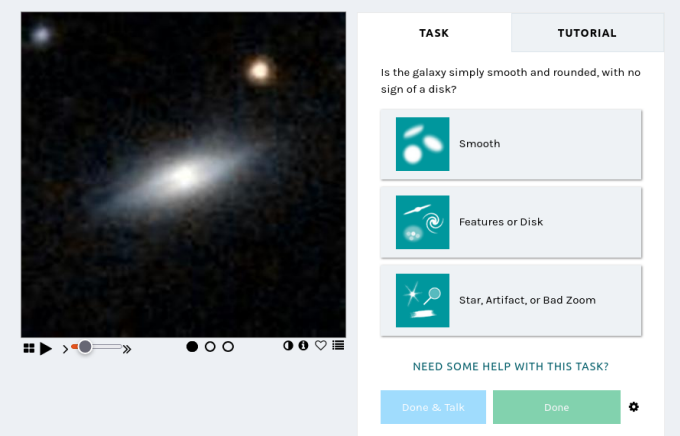

Fig. 4. Galaxy Zoo annotation interface, as shown to volunteers. Volunteers answer a tree of questions, where the next question depends on the previous answer; only the first question is shown here.

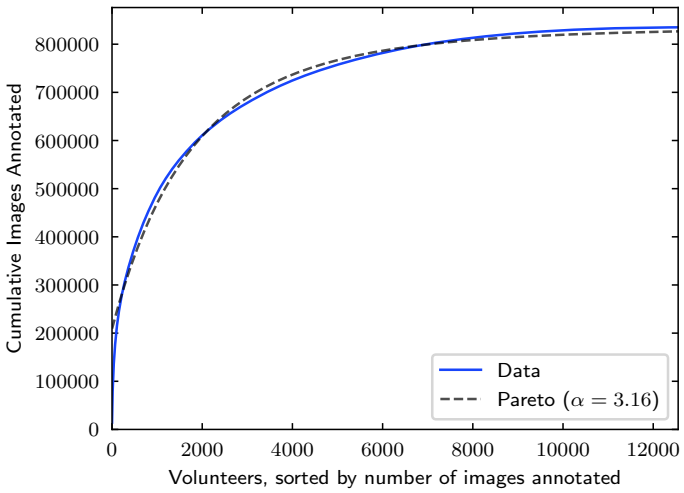

Fig. 5. Cumulative images annotated when selecting the N volunteers who annotated the most images. The bulk of annotations come from highly-engaged volunteers.

3. Methods

3.1. Citizen science

We presented *Euclid* images to Galaxy Zoo volunteers. Volunteers annotated images from the Euclid Wide Survey (EWS), and *not* from Q1. Our pipeline models run within the *Euclid* pipeline that produced the Q1 data release, and so the pipeline models needed to be ready before Q1 was available. We showed these EWS images with permission from ESA and via a Memorandum of Understanding between the Euclid Consortium and the Zooniverse. This MoU created a framework for the Galaxy Zoo team to work with *Euclid* scientists to share a small set of EWS images with the public. These images are ideal for training models that work well on the EWS, and therefore on the vast majority of galaxies *Euclid* will image. We plan on returning to specifically annotate the Euclid Deep fields (including the Q1 area) once full-depth data is available.

The *Euclid* survey images are available as mosaic tiles of $32' \times 32'$ (Euclid Collaboration: Romelli et al. 2025). Galaxies were selected from a set of tiles spread uniformly across the EWS area[4]. All tiles were drawn from the southern half of the EWS (declination < 0) as source catalogue data were not yet available for the northern half.

9976 volunteers contributed 2.9 million annotations of 114 000 galaxies. Of those, 1.56 million annotations were made in the initial one month labelling campaign and used to train the pipeline model. All 2.9 million annotations were used to train the dynamic catalogue model.

Volunteers were presented with a pop-up tutorial, and shown examples in 'help' instructions for each question alongside a site-wide 'field guide'. The annotation interface is shown in Fig. 4. In line with previous Galaxy Zoo projects, a small portion of highly-engaged volunteers contribute the bulk of the annotations. Volunteer contributions are well-modelled by a Pareto distribution (Fig. 5).

We decided to request five volunteer annotations per galaxy for the most galaxies (110 000 of 114 000). Five volunteers is far fewer than typical; Galaxy Zoo has historically collected classifications from 40 volunteers per galaxy (for example Willett et al. 2013). Asking fewer volunteers per galaxy increases the noise in our labels but also increases the diversity of galaxies labelled. We hypothesize that this is a useful trade-off for maximising model performance provided the model loss function can handle uncertain labels (see Sect. 3.2). To accurately measure model performance, we also chose a small random subset (3500) to be annotated by 25 volunteers.

Maximising model performance is our key goal because only a tiny fraction of galaxies imaged by *Euclid* will ever be seen by humans. The complete EWS will include approximately 10^8 galaxies passing our selection cuts above, compared to $\lesssim$ 2M galaxies in all Galaxy Zoo projects over the last 15 years. It is impossible for volunteers to annotate more than a few percent of *Euclid* galaxies. We therefore do not expect scientists to use the volunteer annotations directly, as in all Galaxy Zoo projects prior to Walmsley et al. (2022a), but instead to rely on model predictions. Collecting volunteer annotations to maximise model performance ultimately makes volunteers even more vital because we gain a multiplicative benefit from each annotation; a volunteer annotating one galaxy helps improve a model that annotates all galaxies.

3.2. Foundation models and finetuning

The models used here are the end result of a research project developing adaptable models for galaxy morphology. We briefly summarise the computer science motivation and previous progress below.

Transfer learning is the practice of training on one task to do better at a second task. Image features learned on the first task are hoped to 'transfer' (be relevant) to the second task (Lu et al. 2015). Transfer learning is especially useful where data for the second task is scarce. This was recognised early on as a useful technique in astronomy (Ackermann et al. 2018; Dominguez Sanchez et al. 2019; Tang et al. 2019).

Separately, models trained simultaneously on a diverse set of related tasks often outperform models trained on any single task (Caruana 1997). One explanation is that labels for one task can help models learn general image features relevant to other tasks. Earlier work in this project trained models on multiple morphology tasks in a single survey (Walmsley et al. 2022a) and then expanded to training models on several closely related surveys (Walmsley et al. 2023b).

[4] We selected tiles by picking a random tile, then picking the most distant tile to all previous tiles, repeatedly.

Foundation models (Bommasani et al. 2021; Oquab et al. 2023) combine both transfer learning and multi-task learning. Foundation models involve two model-building phases: 'pretraining' on multiple tasks and then 'downstream finetuning' where the trained model is adapted to a new task. The hope is that the foundation model learns to extract generally useful image features (from multi-task learning) which are then applied to solve the new task (as in transfer learning). Walmsley et al. (2022b) found that the multi-survey pretrained model extracted features that were useful for similarity search (finding similar galaxies to a query galaxy), personalised anomaly recommendation (finding galaxies interesting to a specific user), and new morphology tasks. This motivated the release of `Zoobot` (Walmsley et al. 2023a), the first galaxy foundation models designed to be adapted by other people to new galaxy image tasks. `Zoobot` is part of a recent trend towards foundation models in astronomy (Różański et al. 2023; Leung & Bovy 2023; Koblischke & Bovy 2024; Parker et al. 2024). In related work, Euclid Collaboration: Siudek et al. (2025) experiments with applying the foundation model of Smith et al. (2024b) to Q1.

Model 'scaling laws' (not to be confused with galaxy scaling laws) describe how model performance predictably increases when increasing any of one variable of data, training compute[5], or parameters, provided the other two variables are plentiful. This appears to be true largely independently of model architecture (Kaplan et al. 2020; Hoffmann et al. 2022). Because foundation models are pretrained on diverse tasks with cumulatively plentiful data, they can take advantage of scaling laws by increasing in parameter size and training compute. This underlies the recent success of 'large' language models and recent demand for AI training hardware. Walmsley et al. (2024) investigated model scaling laws for galaxy images (see also Smith et al. 2024b) and released new '`Zoobot 2.0`' models trained on 10^8 volunteer annotations. We use these models here.

The base models used in this work and deployed in the *Euclid* pipeline *were not trained on Euclid data*. They are the `Zoobot` foundation models introduced in Walmsley et al. (2024) and designed to adapt to new tasks and new surveys. They were previously successfully tested on Euclidised HST images morphology in Euclid Collaboration et al. (2024). We use the volunteer annotations to learn a linear mapping, equivalent to logistic regression, projecting the image features extracted by the base model onto *Euclid* morphology measurements. In neural network terminology, we add a new 'head' layer with one unit per morphology answer and freeze the base layers.

4. Results

The most intuitive way to demonstrate the quality of visual morphology measurements is visually.

Figures 6, 7, and 8 demonstrate three challenging visual morphology tasks: identifying strong bars, tidal tails, and galaxies with exactly two spiral arms. Traditional methods for identifying these features are typically only applied to (relatively) small samples of hundreds to thousands of galaxies, e.g., Hoyle et al. (2011); Garcia-Gómez et al. (2017); Consolandi (2016); Lee et al. (2020); Smith et al. (2024a). Figure 9 demonstrates identifying bulgeless edge-on disk galaxies. These are of particular scientific interest as they are likely to be free of recent mergers and hence are useful laboratories for investigating galaxy and supermassive black hole growth (Simmons et al. 2013; Smethurst et al. 2024).

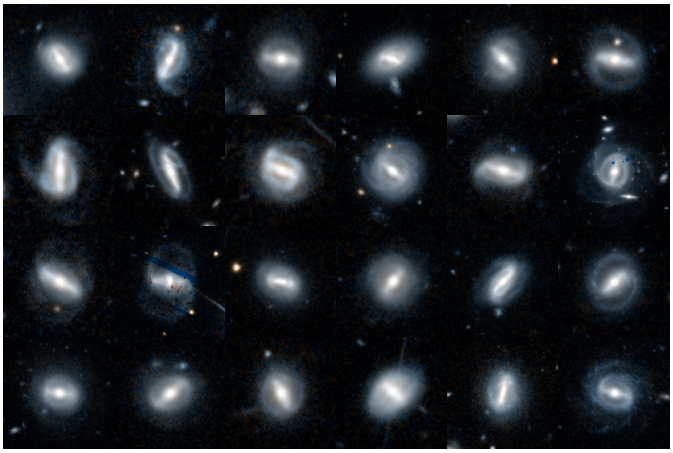

Fig. 6. Q1 galaxies predicted as most likely to host strong bars.

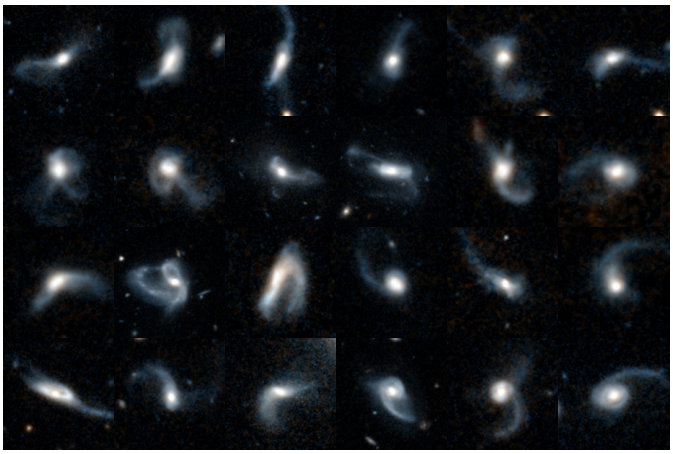

Fig. 7. Q1 galaxies predicted as most likely to have a major disturbance (typically interpreted as tidal tails or similar structures, and distinct from ongoing mergers such as close pairs).

Our catalogue is also useful for measuring less conventional morphology. Figure 10 inverts the previous search for two-armed spirals (Fig. 8) and shows the galaxies which are featured but least likely to be two-armed spirals. This identifies galaxies involved in multiple ongoing mergers. This illustrates how `Zoobot`'s features have generalised beyond the volunteer labels originally used for training; volunteers were not asked to separately identify multiple mergers. Finally, Fig. 11 shows images of dichrotic ghosts, a common artifact (Euclid Collaboration: Jahnke et al. 2024). We identify eight categories of problematic images including stars, saturation features, and bright diffraction spikes.

For a quantitative assessment of the performance of our models, we assess their agreement with volunteers on an intensively-annotated subset of intensively-labelled galaxies created for this purpose (3500 galaxies each with 25 annotations).

We first report classification metrics. These are created by binning the fraction of volunteers giving each answer (for example, if 60% of volunteers answered 'Featured', we bin this label to 'Featured'). We report metrics on, both all galaxies and (following Domínguez Sánchez et al. 2018) on galaxies for whom the class label is confidently known (defined as a volunteer vote

[5] The number of calculations required to train the model, typically measured in floating point operations (FLOPs)

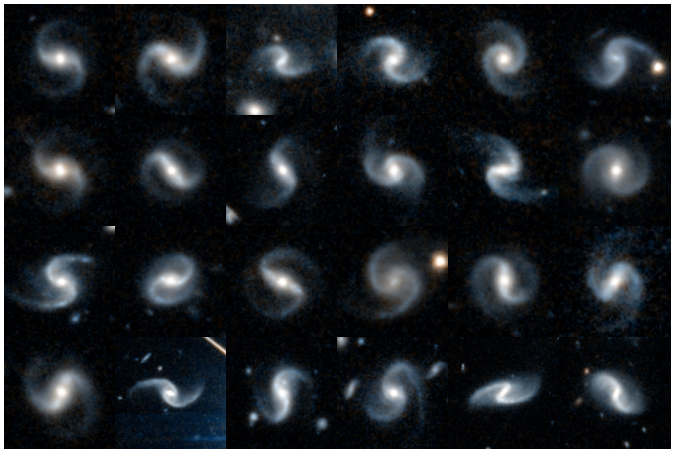

Fig. 8. Q1 galaxies predicted as most likely to have exactly two spiral arms.

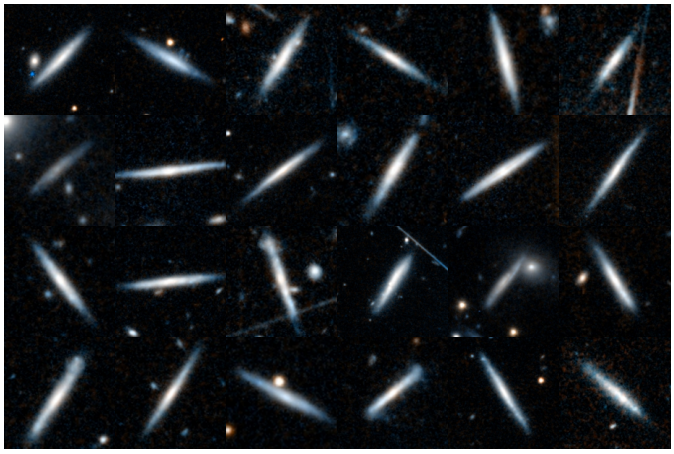

Fig. 9. Q1 galaxies predicted as most likely to be bulgeless edge-on disk galaxies. These are likely to be free of recent mergers and hence are useful laboratories for investigating galaxy and supermassive black hole growth (Simmons et al. 2013; Smethurst et al. 2024).

Fig. 10. Q1 galaxies predicted as featured but *least* likely to have spiral arms. Volunteers were never asked to annotate clusters or multi-mergers, but `Zoobot` has nonetheless learned to extract features that identify such galaxies.

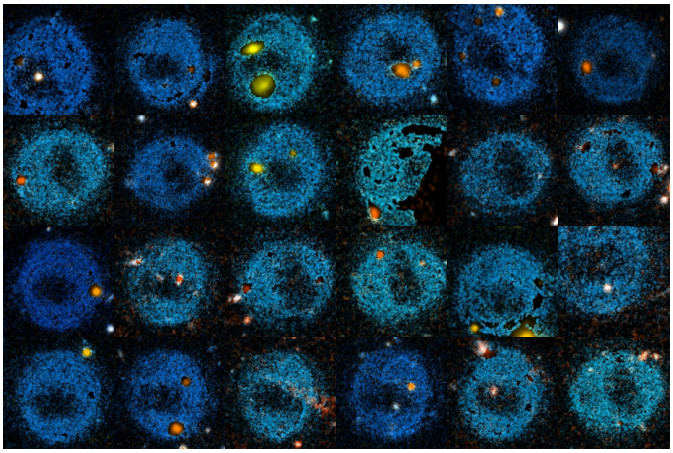

Fig. 11. Q1 sources predicted as most likely to be dichrotic ghosts, a well-known artifact caused by internal reflections in telescopes (Euclid Collaboration: Cropper et al. 2024; Euclid Collaboration: Jahnke et al. 2024).

fraction above 80% for that answer). Figure 12 shows the resulting confusion matrices.

Our model is near-perfect at all questions when evaluated on high-confidence labels, achieving over 99% accuracy on 7 of 13 questions and no lower than 95% accuracy on any question. Performance including lower-confidence labels is more mixed, which likely reflects firstly, more challenging images for both volunteers and models, and secondly, statistical uncertainty in our binned labels. Figure 13 illustrates this for the first morphology question ('smooth or featured?'). When the volunteers give a vote fraction decisively skewed to one answer, our model always predicts that answer. As we move to vote fractions near 0.5, the model begins to make nominally incorrect class predictions – but the binomial uncertainty on the volunteer vote fraction suggests that many binned volunteer labels will fall to one side by chance.

We quantify this by simulating the predictions of a perfect model. We do this by, for each galaxy, drawing a new set of 25 trials from a binomial distribution with p set to the actual volunteer vote fraction. Because the outcome of those new trials includes some uncertainty, the fraction of successful trials is not the same as p. For example, a true vote fraction just below 0.5 (which we should label as 0) will sometimes give a fraction of successful trials above 0.5 and then be given an incorrect label of 1. In our analogy, these correspond to galaxies where our perfect model has made the correct prediction of p, but where the binomial uncertainty in volunteer responses has caused us to record (by misfortune) the wrong label, and so we incorrectly record the model as wrong. When compared to this perfect model, we find that `Zoobot` is only 30% below the best possible classification accuracy (relative error reduction), primarily attributable to a minor bias towards 'Featured'.

We next report regression metrics – the ability of our models to predict the fraction of volunteers giving each answer. For example, if 60% of volunteers answering 'Featured', our model should predict '0.6'. These avoid the noise introduced by either binning uncertain labels or considering only galaxies with confident labels. Figure 14 shows the mean absolute deviation between the predicted and observed volunteer vote fractions (excluding the artifact-related questions, for which we have insufficient examples to calculate reliable metrics). `Zoobot` typically estimates the volunteer vote fraction to within 10%. Consistent with previous work (Walmsley et al. 2023b), increasingly detailed questions are increasingly difficult to precisely predict,

with 'edge-on disk' predicted most accurately (4% error) and 'spiral arm count' least accurately (17% for 2-armed spirals).

5. Data access

All data is available from Zenodo[6]. We also share our data in a machine-learning-friendly format on HuggingFace[7]. This includes the catalogue, the cutout images, the embeddings (vectors summarising the content of each image), and the models. Documentation is provided at those links; below, we provide a summary.

5.1. Catalogues

The dynamic catalogue contains the columns listed below.

- Cross-matching information for the MERge catalogue: object_id ; tile_index.
- id_str. Use to join with embeddings table (below). Formatted like {release_name}_{tile_index}_{object_id}.
- Key MERge catalogue columns for convenience: right_ascension and declination (degrees); kron_radius; mag_segmentation (I_E, from flux_segmentation); segmentation_area.
- Paths to jpg cutouts.
- Detailed morphology measurements formatted like question_answer_fraction (e.g. smooth-or-featured_smooth_fraction), question_answer_dirichlet (e.g. smooth-or-featured_smooth_dirichlet).

We expect that most catalogue users will be primarily interested in the fraction columns. For example, the column 'smooth-or-featured_smooth_fraction' includes the fraction of volunteers predicted by `Zoobot` to give the answer 'Smooth' when asked 'Is this galaxy smooth or featured?'. They can be combined; for example, 'smooth-or-featured_featured-or-disk_fraction > 0.5' and 'disk-edge-on_no > 0.5' would select galaxies which are featured and face-on. Catalogue users might use these columns as selection cuts (to investigate galaxies with specific morphologies) or consider how these measurements correlate with other common measurements like mass, star formation rate, location in the cosmic web, etc. There is no 'best' choice of cuts, because it depends on your aim; increasing the threshold for any cut will make your sample purer but smaller. We suggest starting generously (with thresholds of 0.5 for most questions, or lower for questions with many or rare answers) and raising your thresholds until the sample reaches your desired purity, as judged from the images.

The pipeline catalogue includes only the Dirichlet values and they are named simply as question_answer. For the Dirichlet columns, each value is a parameter for a Dirichlet distribution. The Dirichlet distribution is the multivariate version of the beta distribution[8]. When both beta parameters (that is, both answers to a morphology question) have low values, the beta distribution is flat, and we are uncertain about the galaxy morphology. When one answer is high, and one answer is low, we are confident in that high answer. The '_dirichlet' columns therefore encode both the predicted vote fraction and the uncertainty on that predicted vote fraction. One can calculate the predicted fraction with $\mathrm{E}(X_i) = \alpha_i / \sum \alpha$ and the uncertainty with $\mathrm{Var}(X_i) = \alpha_i(1-\alpha_i)/(1+\sum \alpha)$ where α_i is the Dirichlet concentration of the chosen answer and $\sum \alpha$ is the sum of concentrations for all answers to the chosen question.

The morphology measurements are `Zoobot` predictions, not volunteer answers. `Zoobot` predicts every answer to every question. To avoid providing measurements where a question is not relevant (for example, answering "how many spiral arms? for a smooth galaxy), we set morphology predictions to NaN where the answer is expected to be not relevant. We define this as a leaf probability (that is, the product of all the vote fractions which led to that question) below 0.5.

The detailed morphology measurements in the pipeline catalogue are released as part of the MER (as in, MERged data, see Euclid Collaboration: Romelli et al. 2025) catalogue. The MER catalogue also includes common measurements like photometry, including photometry from other surveys; please refer to the ESA *Euclid* Science Archive website[9] for the latest information.

The detailed morphology measurements in the dynamic catalogue are made outside of the official pipeline. This allows more experimentation and flexibility. For example, the dynamic catalogue can use composite $I_E + Y_E$ images, while the official pipeline uses I_E images only. For another example, we can update the model (such as by introducing new labels) and make new predictions at any time. In general, we expect the dynamic catalogue to include the latest 'bleeding edge' measurements, while the pipeline catalogue will include slowly-changing measurements that match the release cadence of the *Euclid* mission data releases. The difference between the dynamic catalogue and pipeline catalogue should reduce over time as we settle into 'normal operations'.

5.2. Cutouts

We share cutouts of all galaxies in the catalogue, in two formats. For file-based access, we upload our cutouts as part of our Zenodo archive. For machine learning applications, we also share our cutouts on the HuggingFace Hub along with our embeddings (below). Cutouts are saved in native resolution for storage efficiency; you may wish to resize them to a constant size[10].

Images for all of the galaxies in our catalogue are created from MERge mosaics, described in Euclid Collaboration: Romelli et al. (2025) The original Q1 data is available via the ESA *Euclid* Science Archive[11] and described in Euclid Collaboration: Aussel et al. (2025).

We also share reference code for creating our images on GitHub[12]. This code is primarily intended for making *Euclid* cutouts at scale via ESA Datalabs (ESA Datalabs 2021) but also acts as a public record of our exact process.

5.3. Embeddings

We previously (Sect. 3.2) described how foundation models aim to extract image features that summarise the visual content of each image. Each feature is an N-dimensional coordinate vec-

[6] https://doi.org/10.5281/zenodo.15002907

[7] https://huggingface.co/collections/mwalmsley/euclid-67cf5a80e2a93f09e6e4df2c

[8] If helpful, a visualisation tool for the beta distribution is available at https://homepage.divms.uiowa.edu/~mbognar/applets/beta.html

[9] https://eas.esac.esa.int/sas/

[10] For example, with `PIL`, you could use `Image.open(original_loc).resize((300, 300)).save(new_loc)`

[11] https://eas.esac.esa.int/sas/

[12] https://github.com/mwalmsley/bulk-euclid-cutouts

tor (here, $N = 512$) locating (embedding) the image in an N-dimensional space. The linear mapping we describe in Sect. 3.2 aims to learn which volume in this space corresponds to which volunteer answers. But embeddings are also useful for broader tasks such as similarity search, including for galaxy morphology (Stein et al. 2021; Parker et al. 2024; Euclid Collaboration: Siudek et al. 2025).

The `Zoobot` embeddings are presented on Zenodo and the HuggingFace Hub as a table with rows of galaxies and columns like 'feat_pca_n', where feat_pca_n is the N^{th} principle component of our higher-dimensional embedding. We include the first 40 components (preserving 94% variance). We also share the uncompressed higher-dimensional embedding, similarly with columns like 'feat_n'. Bear in mind that many methods may struggle (either performing poorly due to the 'curse of dimensionality' or becoming impractically slow) in high dimensions.

Figure 15 shows two similarity searches made on the `Zoobot` Q1 embeddings. Note that these embeddings were not created using any *Euclid* data (Sect. 3.2).

6. Conclusion and outlook

Our catalogue provides robust visual morphology measurements (e.g., spiral arm counts, bars, mergers) for the bright and extended galaxies in Q1. These measurements complement traditional morphology measurements (Euclid Collaboration: Romelli et al. 2025) and, together, both catalogues provide a comprehensive description of morphology for all Q1 galaxies. Our measurements have already proven useful for addressing a range of science questions; for example, Euclid Collaboration: Huertas-Company et al. (2025) presents new precise measurements of the fraction of barred galaxies out to $z = 1$. We look forward to seeing the work of the wider community.

Deep learning does not automatically resolve the fundamental limitations of visual morphology – it only makes them scale. We use human responses to images as our ground truth throughout and do not attempt to correct for observational biases such as redshift or angular size[13]. The catalogue reports what is visible in each image, as a necessary first step to estimating the intrinsic nature of each galaxy. We anticipate that future progress in large-scale galaxy morphology will come not from more accurate models but from models that take a broader interpretation of what learning means (for example, Cranmer 2023 and Wu 2024) and from the thoughtful combination of models with other essential astronomy tools like simulations and statistics.

Summarising deep learning for survey astronomy, Huertas-Company & Lanusse (2023) write that 'the majority of works are still at the proof-of-concept stage'. In this work, we show that one month of volunteer effort is sufficient to make a science-ready morphology catalogue for 10^8 in *Euclid*, not just as a proof-of-concept but by creating the first rows of that catalogue. The foundation model used was only trained on *Euclid* data for a final linear mapping. We used it to create a galaxy morphology catalogue; we might have chosen another image task. A similar approach has proven successful on strong lenses (Euclid Collaboration: Walmsley et al. 2025), active galactic nuclei (Euclid Collaboration: Margalef-Bentabol et al. 2025), star-forming clumps (Popp et al. 2024), mergers (Margalef-Bentabol et al. 2024), tidal features (Omori et al. 2023; O'Ryan et al. 2023), anomaly searches (Lochner & Rudnick 2024), segmentation Sazonova (2022), etc. We encourage the reader to experiment with using the tools[14] behind our catalogue – the foundation models, and the code to adapt them – to create exactly the catalogue needed for each science case.

[13] In analogy to magnitudes, we report apparent and not absolute morphology.

Acknowledgements. The data in this paper are the result of the efforts of the Galaxy Zoo volunteers, without whom none of this work would be possible. Their efforts are individually acknowledged at http://authors.galaxyzoo.org. The Dunlap Institute is funded through an endowment established by the David Dunlap family and the University of Toronto. MHC acknowledges support from the State Research Agency (AEIMCINN) of the Spanish Ministry of Science and Innovation under the grants "Galaxy Evolution with Artificial Intelligence" with reference PGC2018-100852-A-I00 and "BASALT" with reference PID2021-126838NBI00. This publication uses data generated via the Zooniverse.org platform, development of which is funded by generous support, including a Global Impact Award from Google, and by a grant from the Alfred P. Sloan Foundation. This research was supported by the International Space Science Institute (ISSI) in Bern, through ISSI International Team project #23-584. The Euclid Consortium acknowledges the European Space Agency and a number of agencies and institutes that have supported the development of *Euclid*, in particular the Agenzia Spaziale Italiana, the Austrian Forschungsförderungsgesellschaft funded through BMK, the Belgian Science Policy, the Canadian Euclid Consortium, the Deutsches Zentrum für Luft- und Raumfahrt, the DTU Space and the Niels Bohr Institute in Denmark, the French Centre National d'Etudes Spatiales, the Fundação para a Ciência e a Tecnologia, the Hungarian Academy of Sciences, the Ministerio de Ciencia, Innovación y Universidades, the National Aeronautics and Space Administration, the National Astronomical Observatory of Japan, the Netherlandse Onderzoekschool Voor Astronomie, the Norwegian Space Agency, the Research Council of Finland, the Romanian Space Agency, the State Secretariat for Education, Research, and Innovation (SERI) at the Swiss Space Office (SSO), and the United Kingdom Space Agency. A complete and detailed list is available on the *Euclid* web site (www.euclid-ec.org). This work has made use of the *Euclid* Quick Release Q1 data from the *Euclid* mission of the European Space Agency (ESA), 2025, https://doi.org/10.57780/esa-2853f3b. This research makes use of ESA Datalabs (datalabs.esa.int), an initiative by ESA's Data Science and Archives Division in the Science and Operations Department, Directorate of Science.

References

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543
Abraham, R. G., van den Bergh, S., & Nair, P. 2003, ApJ, 588, 218
Abraham, S., Aniyan, A. K., Kembhavi, A. K., Philip, N. S., & Vaghmare, K. 2018, MNRAS, 477, 894
Ackermann, S., Schawinski, K., Zhang, C., Weigel, A. K., & Turp, M. D. 2018, MNRAS, 479, 415
Baillard, A., Bertin, E., de Lapparent, V., et al. 2011, A&A, 532, A74
Bertin, E. & Arnouts, S. 1996, A&AS, 117, 393
Bom, C. R., Cortesi, A., Lucatelli, G., et al. 2021, MNRAS, 507, 1937
Bommasani, R., Hudson, D. A., Adeli, E., et al. 2021, arXiv e-prints, arXiv:2108.07258
Bowles, M., Tang, H., Vardoulaki, E., et al. 2023, MNRAS, 522, 2584
Buta, R. J., Sheth, K., Athanassoula, E., et al. 2015, ApJS, 217, 32
Caruana, R. 1997, Machine Learning, 28, 41
Conselice, C. J. 2014, ARA&A, 52, 291
Conselice, C. J., Bershady, M. A., & Jangren, A. 2000, ApJ, 529, 886
Consolandi, G. 2016, A&A, 595, A67
Cranmer, M. 2023, arXiv:2305.01582
De Vaucouleurs, G. 1959, in Astrophysik IV: Sternsysteme / Astrophysics IV: Stellar Systems, ed. S. Flügge (Berlin, Heidelberg: Springer), 275–310
Dieleman, S., Willett, K. W., & Dambre, J. 2015, MNRAS, 450, 1441
Dominguez Sanchez, H., Huertas-Company, M., Bernardi, M., et al. 2019, MNRAS, 484, 93
Domínguez Sánchez, H., Huertas-Company, M., Bernardi, M., Tuccillo, D., & Fischer, J. L. 2018, MNRAS, 476, 3661
ESA Datalabs. 2021, http://dx.doi.org/10.13140/RG.2.2.36173.56807
Euclid Collaboration, Aussel, B., Kruk, S., et al. 2024, A&A, 689, A274
Euclid Collaboration: Aussel, H., Tereno, I., Schirmer, M., et al. 2025, A&A, submitted
Euclid Collaboration: Bretonnière, H., Huertas-Company, M., Boucaud, A., et al. 2022, A&A, 657, A90
Euclid Collaboration: Cropper, M., Al Bahlawan, A., Amiaux, J., et al. 2024, A&A, accepted, arXiv:2405.13492

[14] https://github.com/mwalmsley/zoobot

Euclid Collaboration: Holloway, P., Verma, A., Walmsley, M., et al. 2025, A&A, submitted
Euclid Collaboration: Huertas-Company, M., Walmsley, M., Siudek, M., et al. 2025, A&A, submitted
Euclid Collaboration: Jahnke, K., Gillard, W., Schirmer, M., et al. 2024, A&A, accepted, arXiv:2405.13493
Euclid Collaboration: La Marca, A., Wang, L., Margalef-Bentabol, B., et al. 2025, A&A, submitted
Euclid Collaboration: Li, T., Collett, T., Walmsley, M., et al. 2025, A&A, submitted
Euclid Collaboration: Lines, N. E. P., Collett, T. E., Walmsley, M., et al. 2025, A&A, submitted
Euclid Collaboration: Margalef-Bentabol, B., Wang, L., La Marca, A., et al. 2025, A&A, submitted
Euclid Collaboration: Mellier, Y., Abdurro'uf, Acevedo Barroso, J., et al. 2024, A&A, accepted, arXiv:2405.13491
Euclid Collaboration: Quilley, L., Damjanov, I., de Lapparent, V., et al. 2025, A&A, submitted
Euclid Collaboration: Rojas, K., Collett, T., Acevedo Barroso, J., et al. 2025, A&A, submitted
Euclid Collaboration: Romelli, E., Kümmel, M., Dole, H., et al. 2025, A&A, submitted
Euclid Collaboration: Siudek, M., Huertas-Company, M., Smith, M., et al. 2025, A&A, submitted
Euclid Collaboration: Tucci, M., Paltani, S., Hartley, W., et al. 2025, A&A, submitted
Euclid Collaboration: Walmsley, M., Holloway, P., Lines, N., et al. 2025, A&A, submitted
Euclid Quick Release Q1. 2025, https://doi.org/10.57780/esa-2853f3b
Garcia-Gómez, C., Athanassoula, E., Barberà, C., & Bosma, A. 2017, A&A, 601, A132
Ghosh, A., Urry, C. M., Wang, Z., et al. 2020, ApJ, 895, 112
Gordon, A. J., Ferguson, A. M. N., & Mann, R. G. 2024, MNRAS, 534, 1459
Hoffmann, J., Borgeaud, S., Mensch, A., et al. 2022, arXiv:2203.15556
Hoyle, B., Masters, K. L., Nichol, R. C., et al. 2011, MNRAS, 415, 3627
Hubble, E. P. 1926, ApJ, 64, 321
Huertas-Company, M., Bernardi, M., Pérez-González, P. G., et al. 2016, MNRAS, 462, 4495
Huertas-Company, M., Gravet, R., Cabrera-Vives, G., et al. 2015, ApJS, 221, 8
Huertas-Company, M. & Lanusse, F. 2023, PASA, 40
Hunt, L., Annibali, F., Cuillandre, J.-C., et al. 2024, A&A, accepted, arXiv:2405.13499
Kaplan, J., McCandlish, S., Henighan, T., et al. 2020, arXiv: 2001.08361
Khan, A., Huerta, E. A., Wang, S., et al. 2019, Phys. Lett. B, 795, 248
Koblischke, N. & Bovy, J. 2024, arXiv:2411.04750
Lee, Y. H., Park, M.-G., Ann, H. B., Kim, T., & Seo, W.-Y. 2020, ApJ, 899, 84
Leung, H. W. & Bovy, J. 2023, arXiv: 2308.10944
Lingard, T. K., Masters, K. L., Krawczyk, C., et al. 2020, ApJ, 900, 178
Lintott, C., Schawinski, K., Keel, W., et al. 2009, MNRAS, 399, 129
Lintott, C. J., Schawinski, K., Slosar, A., et al. 2008, MNRAS, 389, 1179
Lochner, M. & Rudnick, L. 2024, arXiv:2411.04188
Lu, J., Behbood, V., Hao, P., et al. 2015, Knowledge-Based Systems, 80, 14
Margalef-Bentabol, B., Wang, L., Marca, A. L., et al. 2024, A&A, 687, A24
Masters, K. L. 2019, in Proc. IAU, Vol. 14, 205–212
Morgan, W. W. 1958, PASP, 70, 364
Nair, P. B. & Abraham, R. G. 2010, ApJS, 186, 427
Omori, K. C., Bottrell, C., Walmsley, M., et al. 2023, A&A, 679, A142
Oquab, M., Darcet, T., Moutakanni, T., et al. 2023, arXiv e-prints, arXiv:2304.07193
O'Ryan, D., Merín, B., Simmons, B. D., et al. 2023, ApJ, 948, 40
Parker, L., Lanusse, F., Golkar, S., et al. 2024, MNRAS, 531, 4990
Pearson, W. J., Wang, L., Trayford, J. W., Petrillo, C. E., & van der Tak, F. F. 2019, A&A, 626
Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2002, AJ, 124, 266
Popp, J. J., Dickinson, H., Serjeant, S., et al. 2024, RAS Techniques and Instruments, 3, 174
Rudnick, L. 2021, in Galaxies, Vol. 9 (MDPI), arXiv: 2110.13733
Różański, T., Ting, Y.-S., Jabłońska, M., et al. 2023, arXiv:2306.15703
Sazonova, E. 2022, PhD thesis, Johns Hopkins University, Baltimore
Sellwood, J. A. & Masters, K. L. 2022, ARA&A, 60, 73
Shimasaku, K., Fukugita, M., Doi, M., et al. 2001, AJ, 122, 1238
Simmons, B., Lintott, C., Schawinski, K., et al. 2013, MNRAS, 429, 2199
Smethurst, R. J., Beckmann, R. S., Simmons, B. D., et al. 2024, MNRAS, 527, 10855
Smith, B. J., Watson, M., Giroux, M. L., & Struck, C. 2024a, arXiv:2405.01516
Smith, M. J., Roberts, R. J., Angeloudi, E., & Huertas-Company, M. 2024b, arXiv:2405.14930
Stein, G., Harrington, P., Blaum, J., Medan, T., & Lukic, Z. 2021, arXiv: 2110.13151
Sérsic, J. L. 1963, Boletin de la Asociacion Argentina de Astronomia La Plata Argentina, 6, 41
Tacchella, S., Carollo, C. M., Renzini, A., et al. 2015, Science, 348, 314
Tanaka, M., Koike, M., Naito, S., et al. 2023, PASJ, 75, 986
Tang, H., Scaife, A. M., & Leahy, J. P. 2019, MNRAS, 488, 3358
Toomre, A. & Toomre, J. 1972, ApJ, 178, 623
Walmsley, M., Allen, C., Aussel, B., et al. 2023a, Journal of Open Source Software, 8, 5312
Walmsley, M., Bowles, M., Scaife, A. M. M., et al. 2024, arXiv:2404.02973
Walmsley, M., Géron, T., Kruk, S., et al. 2023b, MNRAS, 526, 4768
Walmsley, M., Lintott, C., Géron, T., et al. 2022a, MNRAS, 509, 3966
Walmsley, M., Scaife, A. M., Lintott, C., et al. 2022b, MNRAS, 513, 1581
Willett, K. W., Galloway, M. A., Bamford, S. P., et al. 2017, MNRAS, 464, 4176
Willett, K. W., Lintott, C. J., Bamford, S. P., et al. 2013, MNRAS, 435, 2835
Wu, J. F. 2024, arXiv:2501.00089
Wuyts, S., Förster Schreiber, N. M., van der Wel, A., et al. 2011, ApJ, 742, 96
Ye, R., Shen, S., de Souza, R. S., et al. 2025, MNRAS, 537, 640
Ćiprijanović, A., Lewis, A., Pedro, K., et al. 2022, arXiv:2211.00677

[1] David A. Dunlap Department of Astronomy & Astrophysics, University of Toronto, 50 St George Street, Toronto, Ontario M5S 3H4, Canada
[2] Jodrell Bank Centre for Astrophysics, Department of Physics and Astronomy, University of Manchester, Oxford Road, Manchester M13 9PL, UK
[3] Instituto de Astrofísica de Canarias, Vía Láctea, 38205 La Laguna, Tenerife, Spain
[4] Instituto de Astrofísica de Canarias (IAC); Departamento de Astrofísica, Universidad de La Laguna (ULL), 38200, La Laguna, Tenerife, Spain
[5] Université PSL, Observatoire de Paris, Sorbonne Université, CNRS, LERMA, 75014, Paris, France
[6] Université Paris-Cité, 5 Rue Thomas Mann, 75013, Paris, France
[7] Centre de Recherche Astrophysique de Lyon, UMR5574, CNRS, Université Claude Bernard Lyon 1, ENS de Lyon, 69230, Saint-Genis-Laval, France
[8] Departments of Physics and Astronomy, Haverford College, 370 Lancaster Avenue, Haverford, PA 19041, USA
[9] ESAC/ESA, Camino Bajo del Castillo, s/n., Urb. Villafranca del Castillo, 28692 Villanueva de la Cañada, Madrid, Spain
[10] School of Physical Sciences, The Open University, Milton Keynes, MK7 6AA, UK
[11] INAF-Osservatorio Astronomico di Trieste, Via G. B. Tiepolo 11, 34143 Trieste, Italy
[12] Centro de Astrobiología (CAB), CSIC-INTA, ESAC Campus, Camino Bajo del Castillo s/n, 28692 Villanueva de la Cañada, Madrid, Spain
[13] Department of Physics, Oxford University, Keble Road, Oxford OX1 3RH, UK
[14] Department of Physics, Lancaster University, Lancaster, LA1 4YB, UK
[15] Masaryk University, Kotlářská 2, Brno, 611 37, Czech Republic
[16] Minnesota Institute for Astrophysics, University of Minnesota, 116 Church St SE, Minneapolis, MN 55455, USA
[17] University of Alabama, Tuscaloosa, AL 35487, USA
[18] Citizen Scientist, Zooniverse c/o University of Oxford, Keble Road, Oxford OX1 3RH, UK
[19] Institute for Theoretical Particle Physics and Cosmology (TTK), RWTH Aachen University, 52056 Aachen, Germany
[20] Dipartimento di Fisica, Sapienza Università di Roma, Piazzale Aldo Moro 2, 00185 Roma, Italy
[21] Instituto de Física de Cantabria, Edificio Juan Jordá, Avenida de los Castros, 39005 Santander, Spain
[22] ELTE Eötvös Loránd University, Institute of Physics and Astronomy, Pázmány P. st. 1/A, H-1171 Budapest, Hungary
[23] MTA-CSFK Lendület Large-Scale Structure Research Group, Konkoly-Thege Miklós út 15-17, H-1121 Budapest, Hungary
[24] Konkoly Observatory, HUN-REN CSFK, MTA Centre of Excellence, Budapest, Konkoly Thege Miklós út 15-17. H-1121, Hungary

[25] Université Paris-Saclay, CNRS, Institut d'astrophysique spatiale, 91405, Orsay, France
[26] School of Mathematics and Physics, University of Surrey, Guildford, Surrey, GU2 7XH, UK
[27] INAF-Osservatorio Astronomico di Brera, Via Brera 28, 20122 Milano, Italy
[28] INAF-Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, Via Piero Gobetti 93/3, 40129 Bologna, Italy
[29] Université Paris-Saclay, Université Paris Cité, CEA, CNRS, AIM, 91191, Gif-sur-Yvette, France
[30] IFPU, Institute for Fundamental Physics of the Universe, via Beirut 2, 34151 Trieste, Italy
[31] INFN, Sezione di Trieste, Via Valerio 2, 34127 Trieste TS, Italy
[32] SISSA, International School for Advanced Studies, Via Bonomea 265, 34136 Trieste TS, Italy
[33] Dipartimento di Fisica e Astronomia, Università di Bologna, Via Gobetti 93/2, 40129 Bologna, Italy
[34] INFN-Sezione di Bologna, Viale Berti Pichat 6/2, 40127 Bologna, Italy
[35] INAF-Osservatorio Astronomico di Padova, Via dell'Osservatorio 5, 35122 Padova, Italy
[36] Centre National d'Etudes Spatiales – Centre spatial de Toulouse, 18 avenue Edouard Belin, 31401 Toulouse Cedex 9, France
[37] Max Planck Institute for Extraterrestrial Physics, Giessenbachstr. 1, 85748 Garching, Germany
[38] Universitäts-Sternwarte München, Fakultät für Physik, Ludwig-Maximilians-Universität München, Scheinerstrasse 1, 81679 München, Germany
[39] Space Science Data Center, Italian Space Agency, via del Politecnico snc, 00133 Roma, Italy
[40] Dipartimento di Fisica, Università di Genova, Via Dodecaneso 33, 16146, Genova, Italy
[41] INFN-Sezione di Genova, Via Dodecaneso 33, 16146, Genova, Italy
[42] Department of Physics "E. Pancini", University Federico II, Via Cinthia 6, 80126, Napoli, Italy
[43] INAF-Osservatorio Astronomico di Capodimonte, Via Moiariello 16, 80131 Napoli, Italy
[44] Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto, CAUP, Rua das Estrelas, PT4150-762 Porto, Portugal
[45] Faculdade de Ciências da Universidade do Porto, Rua do Campo de Alegre, 4150-007 Porto, Portugal
[46] Dipartimento di Fisica, Università degli Studi di Torino, Via P. Giuria 1, 10125 Torino, Italy
[47] INFN-Sezione di Torino, Via P. Giuria 1, 10125 Torino, Italy
[48] INAF-Osservatorio Astrofisico di Torino, Via Osservatorio 20, 10025 Pino Torinese (TO), Italy
[49] European Space Agency/ESTEC, Keplerlaan 1, 2201 AZ Noordwijk, The Netherlands
[50] Institute Lorentz, Leiden University, Niels Bohrweg 2, 2333 CA Leiden, The Netherlands
[51] Leiden Observatory, Leiden University, Einsteinweg 55, 2333 CC Leiden, The Netherlands
[52] INAF-IASF Milano, Via Alfonso Corti 12, 20133 Milano, Italy
[53] Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Avenida Complutense 40, 28040 Madrid, Spain
[54] Port d'Informació Científica, Campus UAB, C. Albareda s/n, 08193 Bellaterra (Barcelona), Spain
[55] Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, 08193 Barcelona, Spain
[56] Institut d'Estudis Espacials de Catalunya (IEEC), Edifici RDIT, Campus UPC, 08860 Castelldefels, Barcelona, Spain
[57] INAF-Osservatorio Astronomico di Roma, Via Frascati 33, 00078 Monteporzio Catone, Italy
[58] INFN section of Naples, Via Cinthia 6, 80126, Napoli, Italy
[59] Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, HI 96822, USA
[60] Dipartimento di Fisica e Astronomia "Augusto Righi" - Alma Mater Studiorum Università di Bologna, Viale Berti Pichat 6/2, 40127 Bologna, Italy
[61] Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK
[62] European Space Agency/ESRIN, Largo Galileo Galilei 1, 00044 Frascati, Roma, Italy
[63] Université Claude Bernard Lyon 1, CNRS/IN2P3, IP2I Lyon, UMR 5822, Villeurbanne, F-69100, France
[64] Institut de Ciències del Cosmos (ICCUB), Universitat de Barcelona (IEEC-UB), Martí i Franquès 1, 08028 Barcelona, Spain
[65] Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig de Lluís Companys 23, 08010 Barcelona, Spain
[66] UCB Lyon 1, CNRS/IN2P3, IUF, IP2I Lyon, 4 rue Enrico Fermi, 69622 Villeurbanne, France
[67] Mullard Space Science Laboratory, University College London, Holmbury St Mary, Dorking, Surrey RH5 6NT, UK
[68] Departamento de Física, Faculdade de Ciências, Universidade de Lisboa, Edifício C8, Campo Grande, PT1749-016 Lisboa, Portugal
[69] Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal
[70] Department of Astronomy, University of Geneva, ch. d'Ecogia 16, 1290 Versoix, Switzerland
[71] INAF-Istituto di Astrofisica e Planetologia Spaziali, via del Fosso del Cavaliere, 100, 00100 Roma, Italy
[72] INFN-Padova, Via Marzolo 8, 35131 Padova, Italy
[73] Aix-Marseille Université, CNRS/IN2P3, CPPM, Marseille, France
[74] INFN-Bologna, Via Irnerio 46, 40126 Bologna, Italy
[75] School of Physics, HH Wills Physics Laboratory, University of Bristol, Tyndall Avenue, Bristol, BS8 1TL, UK
[76] FRACTAL S.L.N.E., calle Tulipán 2, Portal 13 1A, 28231, Las Rozas de Madrid, Spain
[77] NRC Herzberg, 5071 West Saanich Rd, Victoria, BC V9E 2E7, Canada
[78] Institute of Theoretical Astrophysics, University of Oslo, P.O. Box 1029 Blindern, 0315 Oslo, Norway
[79] Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA, 91109, USA
[80] Felix Hormuth Engineering, Goethestr. 17, 69181 Leimen, Germany
[81] Technical University of Denmark, Elektrovej 327, 2800 Kgs. Lyngby, Denmark
[82] Cosmic Dawn Center (DAWN), Denmark
[83] Institut d'Astrophysique de Paris, UMR 7095, CNRS, and Sorbonne Université, 98 bis boulevard Arago, 75014 Paris, France
[84] Max-Planck-Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany
[85] NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA
[86] Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK
[87] Department of Physics and Helsinki Institute of Physics, Gustaf Hällströmin katu 2, 00014 University of Helsinki, Finland
[88] Université de Genève, Département de Physique Théorique and Centre for Astroparticle Physics, 24 quai Ernest-Ansermet, CH-1211 Genève 4, Switzerland
[89] Department of Physics, P.O. Box 64, 00014 University of Helsinki, Finland
[90] Helsinki Institute of Physics, Gustaf Hällströmin katu 2, University of Helsinki, Helsinki, Finland
[91] Centre de Calcul de l'IN2P3/CNRS, 21 avenue Pierre de Coubertin 69627 Villeurbanne Cedex, France
[92] Laboratoire d'etude de l'Univers et des phenomenes eXtremes, Observatoire de Paris, Université PSL, Sorbonne Université, CNRS, 92190 Meudon, France
[93] Aix-Marseille Université, CNRS, CNES, LAM, Marseille, France
[94] SKA Observatory, Jodrell Bank, Lower Withington, Macclesfield, Cheshire SK11 9FT, UK
[95] Dipartimento di Fisica "Aldo Pontremoli", Università degli Studi di Milano, Via Celoria 16, 20133 Milano, Italy

[96] INFN-Sezione di Milano, Via Celoria 16, 20133 Milano, Italy
[97] University of Applied Sciences and Arts of Northwestern Switzerland, School of Computer Science, 5210 Windisch, Switzerland
[98] Universität Bonn, Argelander-Institut für Astronomie, Auf dem Hügel 71, 53121 Bonn, Germany
[99] INFN-Sezione di Roma, Piazzale Aldo Moro, 2 - c/o Dipartimento di Fisica, Edificio G. Marconi, 00185 Roma, Italy
[100] Dipartimento di Fisica e Astronomia "Augusto Righi" - Alma Mater Studiorum Università di Bologna, via Piero Gobetti 93/2, 40129 Bologna, Italy
[101] Department of Physics, Institute for Computational Cosmology, Durham University, South Road, Durham, DH1 3LE, UK
[102] Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Laboratoire Lagrange, Bd de l'Observatoire, CS 34229, 06304 Nice cedex 4, France
[103] Université Paris Cité, CNRS, Astroparticule et Cosmologie, 75013 Paris, France
[104] CNRS-UCB International Research Laboratory, Centre Pierre Binetruy, IRL2007, CPB-IN2P3, Berkeley, USA
[105] University of Applied Sciences and Arts of Northwestern Switzerland, School of Engineering, 5210 Windisch, Switzerland
[106] Institut d'Astrophysique de Paris, 98bis Boulevard Arago, 75014, Paris, France
[107] Institute of Physics, Laboratory of Astrophysics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, 1290 Versoix, Switzerland
[108] Aurora Technology for European Space Agency (ESA), Camino bajo del Castillo, s/n, Urbanizacion Villafranca del Castillo, Villanueva de la Cañada, 28692 Madrid, Spain
[109] Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, 08193 Bellaterra (Barcelona), Spain
[110] School of Mathematics, Statistics and Physics, Newcastle University, Herschel Building, Newcastle-upon-Tyne, NE1 7RU, UK
[111] DARK, Niels Bohr Institute, University of Copenhagen, Jagtvej 155, 2200 Copenhagen, Denmark
[112] Waterloo Centre for Astrophysics, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada
[113] Department of Physics and Astronomy, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada
[114] Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada
[115] Institute of Space Science, Str. Atomistilor, nr. 409 Măgurele, Ilfov, 077125, Romania
[116] Consejo Superior de Investigaciones Cientificas, Calle Serrano 117, 28006 Madrid, Spain
[117] Universidad de La Laguna, Departamento de Astrofísica, 38206 La Laguna, Tenerife, Spain
[118] Dipartimento di Fisica e Astronomia "G. Galilei", Università di Padova, Via Marzolo 8, 35131 Padova, Italy
[119] Caltech/IPAC, 1200 E. California Blvd., Pasadena, CA 91125, USA
[120] Institut für Theoretische Physik, University of Heidelberg, Philosophenweg 16, 69120 Heidelberg, Germany
[121] Institut de Recherche en Astrophysique et Planétologie (IRAP), Université de Toulouse, CNRS, UPS, CNES, 14 Av. Edouard Belin, 31400 Toulouse, France
[122] Université St Joseph; Faculty of Sciences, Beirut, Lebanon
[123] Departamento de Física, FCFM, Universidad de Chile, Blanco Encalada 2008, Santiago, Chile
[124] Universität Innsbruck, Institut für Astro- und Teilchenphysik, Technikerstr. 25/8, 6020 Innsbruck, Austria
[125] Satlantis, University Science Park, Sede Bld 48940, Leioa-Bilbao, Spain
[126] Infrared Processing and Analysis Center, California Institute of Technology, Pasadena, CA 91125, USA
[127] Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências, Universidade de Lisboa, Tapada da Ajuda, 1349-018 Lisboa, Portugal
[128] Cosmic Dawn Center (DAWN)
[129] Niels Bohr Institute, University of Copenhagen, Jagtvej 128, 2200 Copenhagen, Denmark
[130] Universidad Politécnica de Cartagena, Departamento de Electrónica y Tecnología de Computadoras, Plaza del Hospital 1, 30202 Cartagena, Spain
[131] Kapteyn Astronomical Institute, University of Groningen, PO Box 800, 9700 AV Groningen, The Netherlands
[132] Dipartimento di Fisica e Scienze della Terra, Università degli Studi di Ferrara, Via Giuseppe Saragat 1, 44122 Ferrara, Italy
[133] Istituto Nazionale di Fisica Nucleare, Sezione di Ferrara, Via Giuseppe Saragat 1, 44122 Ferrara, Italy
[134] INAF, Istituto di Radioastronomia, Via Piero Gobetti 101, 40129 Bologna, Italy
[135] INAF - Osservatorio Astronomico di Brera, via Emilio Bianchi 46, 23807 Merate, Italy
[136] INAF-Osservatorio Astronomico di Brera, Via Brera 28, 20122 Milano, Italy, and INFN-Sezione di Genova, Via Dodecaneso 33, 16146, Genova, Italy
[137] ICL, Junia, Université Catholique de Lille, LITL, 59000 Lille, France
[138] ICSC - Centro Nazionale di Ricerca in High Performance Computing, Big Data e Quantum Computing, Via Magnanelli 2, Bologna, Italy
[139] Instituto de Física Teórica UAM-CSIC, Campus de Cantoblanco, 28049 Madrid, Spain
[140] CERCA/ISO, Department of Physics, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106, USA
[141] Technical University of Munich, TUM School of Natural Sciences, Physics Department, James-Franck-Str. 1, 85748 Garching, Germany
[142] Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, 85748 Garching, Germany
[143] Laboratoire Univers et Théorie, Observatoire de Paris, Université PSL, Université Paris Cité, CNRS, 92190 Meudon, France
[144] Departamento de Física Fundamental. Universidad de Salamanca. Plaza de la Merced s/n. 37008 Salamanca, Spain
[145] Université de Strasbourg, CNRS, Observatoire astronomique de Strasbourg, UMR 7550, 67000 Strasbourg, France
[146] Center for Data-Driven Discovery, Kavli IPMU (WPI), UTIAS, The University of Tokyo, Kashiwa, Chiba 277-8583, Japan
[147] Ludwig-Maximilians-University, Schellingstrasse 4, 80799 Munich, Germany
[148] Max-Planck-Institut für Physik, Boltzmannstr. 8, 85748 Garching, Germany
[149] California Institute of Technology, 1200 E California Blvd, Pasadena, CA 91125, USA
[150] Departamento de Física Teórica, Atómica y Óptica, Universidad de Valladolid, 47011 Valladolid, Spain
[151] Department of Physics & Astronomy, University of California Irvine, Irvine CA 92697, USA
[152] Department of Mathematics and Physics E. De Giorgi, University of Salento, Via per Arnesano, CP-I93, 73100, Lecce, Italy
[153] INFN, Sezione di Lecce, Via per Arnesano, CP-193, 73100, Lecce, Italy
[154] INAF-Sezione di Lecce, c/o Dipartimento Matematica e Fisica, Via per Arnesano, 73100, Lecce, Italy
[155] Departamento Física Aplicada, Universidad Politécnica de Cartagena, Campus Muralla del Mar, 30202 Cartagena, Murcia, Spain
[156] CEA Saclay, DFR/IRFU, Service d'Astrophysique, Bat. 709, 91191 Gif-sur-Yvette, France
[157] Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth PO1 3FX, UK
[158] Department of Computer Science, Aalto University, PO Box 15400, Espoo, FI-00 076, Finland
[159] Instituto de Astrofísica de Canarias, c/ Via Lactea s/n, La Laguna 38200, Spain. Departamento de Astrofísica de la Universidad de La Laguna, Avda. Francisco Sanchez, La Laguna, 38200, Spain
[160] Ruhr University Bochum, Faculty of Physics and Astronomy, Astronomical Institute (AIRUB), German Centre for Cosmological Lensing (GCCL), 44780 Bochum, Germany

[161] Department of Physics and Astronomy, Vesilinnantie 5, 20014 University of Turku, Finland
[162] Serco for European Space Agency (ESA), Camino bajo del Castillo, s/n, Urbanizacion Villafranca del Castillo, Villanueva de la Cañada, 28692 Madrid, Spain
[163] ARC Centre of Excellence for Dark Matter Particle Physics, Melbourne, Australia
[164] Centre for Astrophysics & Supercomputing, Swinburne University of Technology, Hawthorn, Victoria 3122, Australia
[165] Department of Physics and Astronomy, University of the Western Cape, Bellville, Cape Town, 7535, South Africa
[166] DAMTP, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, UK
[167] Kavli Institute for Cosmology Cambridge, Madingley Road, Cambridge, CB3 0HA, UK
[168] Department of Astrophysics, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland
[169] Department of Physics, Centre for Extragalactic Astronomy, Durham University, South Road, Durham, DH1 3LE, UK
[170] IRFU, CEA, Université Paris-Saclay 91191 Gif-sur-Yvette Cedex, France
[171] Oskar Klein Centre for Cosmoparticle Physics, Department of Physics, Stockholm University, Stockholm, SE-106 91, Sweden
[172] Astrophysics Group, Blackett Laboratory, Imperial College London, London SW7 2AZ, UK
[173] Univ. Grenoble Alpes, CNRS, Grenoble INP, LPSC-IN2P3, 53, Avenue des Martyrs, 38000, Grenoble, France
[174] INAF-Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, 50125, Firenze, Italy
[175] Centro de Astrofísica da Universidade do Porto, Rua das Estrelas, 4150-762 Porto, Portugal
[176] HE Space for European Space Agency (ESA), Camino bajo del Castillo, s/n, Urbanizacion Villafranca del Castillo, Villanueva de la Cañada, 28692 Madrid, Spain
[177] Dipartimento di Fisica - Sezione di Astronomia, Università di Trieste, Via Tiepolo 11, 34131 Trieste, Italy
[178] Department of Astrophysical Sciences, Peyton Hall, Princeton University, Princeton, NJ 08544, USA
[179] Theoretical astrophysics, Department of Physics and Astronomy, Uppsala University, Box 515, 751 20 Uppsala, Sweden
[180] Mathematical Institute, University of Leiden, Einsteinweg 55, 2333 CA Leiden, The Netherlands
[181] School of Physics & Astronomy, University of Southampton, Highfield Campus, Southampton SO17 1BJ, UK
[182] Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK
[183] Space physics and astronomy research unit, University of Oulu, Pentti Kaiteran katu 1, FI-90014 Oulu, Finland
[184] Center for Computational Astrophysics, Flatiron Institute, 162 5th Avenue, 10010, New York, NY, USA
[185] Department of Physics and Astronomy, University of British Columbia, Vancouver, BC V6T 1Z1, Canada

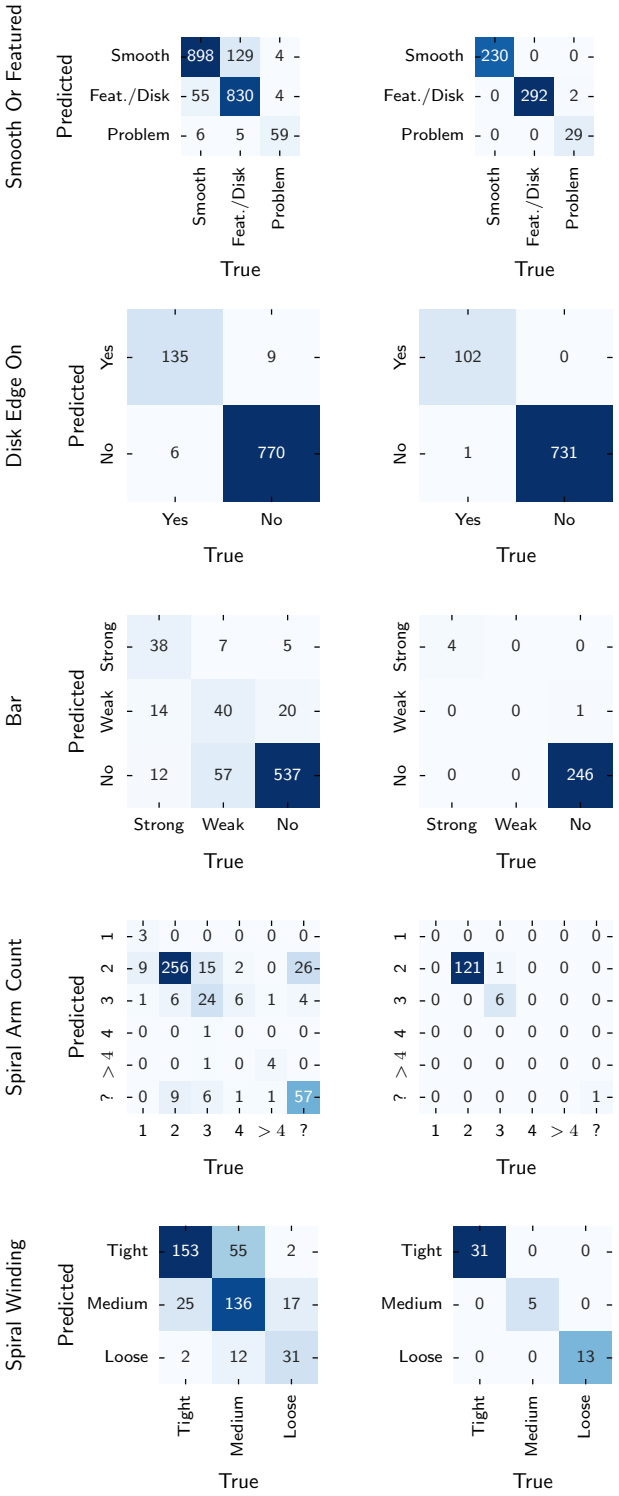

Fig. 12. Confusion matrices for the most commonly-used morphology features. Right column includes only galaxies with high confidence (volunteer vote fraction > 0.8) labels. Our model is near-perfect at all questions when evaluated on high-confidence labels. Performance including lower-confidence labels is more mixed, which likely reflects more challenging images for both volunteers and models.

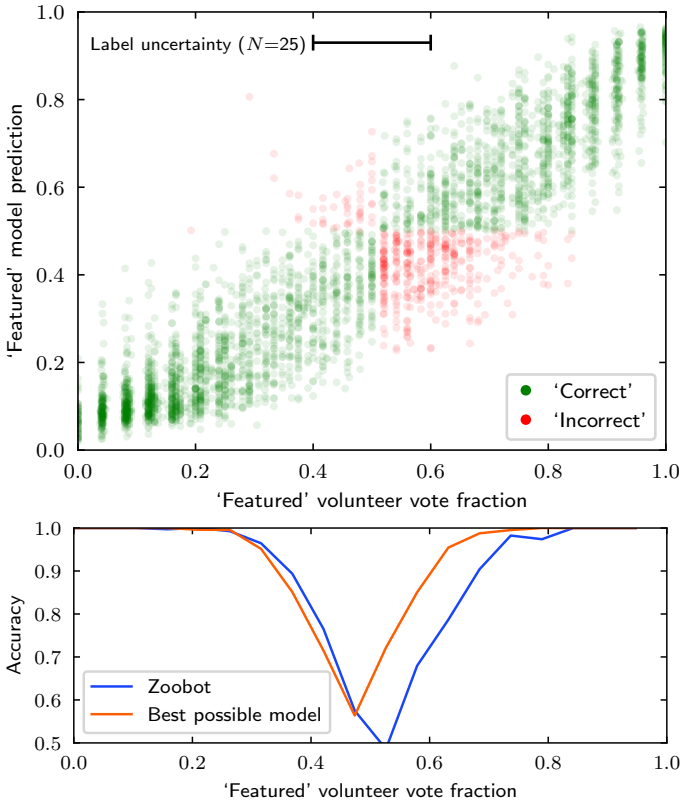

Fig. 13. Detailed model performance for the 'Smooth or Featured?' question, illustrating the limitations of classification metrics. Above, the measured (x-axis) vs. predicted (y-axis) 'Featured' vote fractions, coloured according to whether the binned prediction equals the binned label ('correct') or not ('incorrect'). Classification errors mostly occur for vote fractions where the true label is uncertain (when the label boundary is within our uncertainty on the vote fraction itself). Below, model accuracy as a function of 'featured' vote fraction for either `Zoobot` or a perfect model (see Sect. 4), illustrating that `Zoobot` is near-perfect within the uncertainty on our labels.

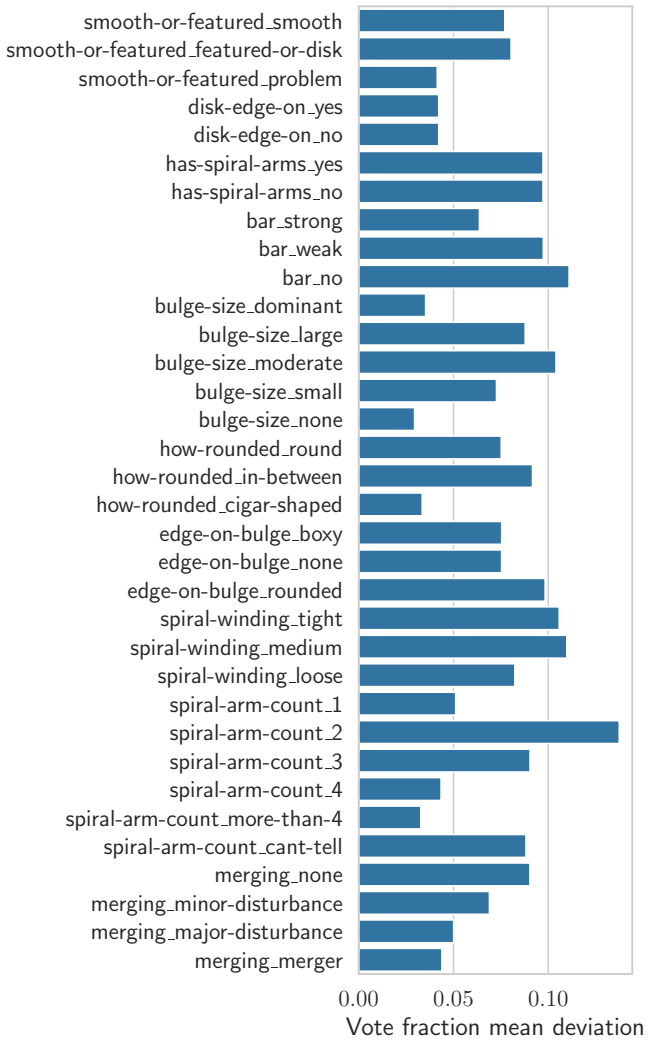

Fig. 14. Mean absolute deviation between the predicted and observed volunteer vote fractions. Increasingly detailed questions are increasingly difficult to precisely predict, with the most challenging task being counting spiral arms (17% error for 2-armed spirals). Errors are typically within 10%. For conciseness, 'clumps' and 'problem' questions are not shown.

Fig. 15. Similarity searches for a pair of stars (above) and strong gravitational lens candidates (below) in Q1.

Table A.1. Classification metrics on all galaxies. See Sect. 4.

Question	Count	Accuracy	Precision	Recall	F1
Smooth Or Featured	3546	0.9041	0.9041	0.9041	0.9041
Disk Edge On	1648	0.9806	0.9806	0.9806	0.9806
Has Spiral Arms	1308	0.9029	0.9029	0.9029	0.9029
Bar	1308	0.8471	0.8471	0.8471	0.8471
Bulge Size	1308	0.8242	0.8242	0.8242	0.8242
How Rounded	1573	0.9091	0.9091	0.9091	0.9091
Edge On Bulge	173	0.8902	0.8902	0.8902	0.8902
Spiral Winding	794	0.7292	0.7292	0.7292	0.7292
Spiral Arm Count	794	0.7746	0.7746	0.7746	0.7746
Merging	3463	0.9090	0.9090	0.9090	0.9090
Clumps	1308	0.7638	0.7638	0.7638	0.7638
Problem	98	0.8673	0.8673	0.8673	0.8673
Artifact	24	0.7083	0.7083	0.7083	0.7083

Table A.2. Classification metrics on high-confidence ($p > 0.8$) galaxies. See Sect. 4.

Question	Count	Accuracy	Precision	Recall	F1
Smooth Or Featured	1136	0.9974	0.9974	0.9974	0.9974
Disk Edge On	1478	0.9993	0.9993	0.9993	0.9993
Has Spiral Arms	787	0.9848	0.9848	0.9848	0.9848
Bar	453	0.9934	0.9934	0.9934	0.9934
Bulge Size	120	1.0000	1.0000	1.0000	1.0000
How Rounded	875	0.9977	0.9977	0.9977	0.9977
Edge On Bulge	69	1.0000	1.0000	1.0000	1.0000
Spiral Winding	85	0.9882	0.9882	0.9882	0.9882
Spiral Arm Count	235	0.9830	0.9830	0.9830	0.9830
Merging	1090	1.0000	1.0000	1.0000	1.0000
Clumps	419	0.9523	0.9523	0.9523	0.9523
Problem	40	0.9750	0.9750	0.9750	0.9750
Artifact	4	1.0000	1.0000	1.0000	1.0000

Appendix A: Classification metric tables

This section records the classification metrics on, split between all galaxies and high-confidence ($p > 0.8$) galaxies.