Galaxy Spin Classification. I. Z-wise versus S-wise Spirals with the Chirality Equivariant Residual Network

He Jia (朱家) 1, Hong-Ming Zhu (朱弘明) 2, and Ue-Li Pen (潘威礼) 2,3,4,5,6

1 Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544, USA; hejia@princeton.edu
2 Canadian Institute for Theoretical Astrophysics, 60 St. George Street, Toronto, Ontario M5S 3H8, Canada
3 Institute of Astronomy and Astrophysics, Academia Sinica, Astronomy-Mathematics Building, No. 1, Section 4, Roosevelt Road, Taipei 10617, Taiwan
4 Canadian Institute for Advanced Research, 661 University Avenue, Toronto, Ontario M5G 1M1, Canada
5 Dunlap Institute for Astronomy and Astrophysics, University of Toronto, 50 St. George Street, Toronto, Ontario M5S 3H4, Canada
6 Perimeter Institute for Theoretical Physics, 31 Caroline Street North, Waterloo, Ontario N2L 2Y5, Canada

Received 2022 October 9; revised 2022 November 21; accepted 2022 December 3; published 2023 January 23

Abstract

The angular momentum of galaxies (galaxy spin) contains rich information about the initial condition of the universe, yet it is challenging to efficiently measure the spin direction for the tremendous amount of galaxies that are being mapped by ongoing and forthcoming cosmological surveys. We present a machine-learning-based classifier for the Z-wise versus S-wise spirals, which can help to break the degeneracy in the galaxy spin direction measurement. The proposed chirality equivariant residual network (CE-ResNet) is manifestly equivariant under a reflection of the input image, which guarantees that there is no inherent asymmetry between the Z-wise and S-wise probability estimators. We train the model with Sloan Digital Sky Survey images, with the training labels given by the Galaxy Zoo 1 project. A combination of data augmentation techniques is used during the training, making the model more robust to be applied to other surveys. We find an ∼30% increase in both types of spirals when Dark Energy Spectroscopic Instrument (DESI) images are used for classification, due to the better imaging quality of DESI. We verify that the ∼7σ difference between the numbers of Z-wise and S-wise spirals is due to human bias, since the discrepancy drops to <1.8σ with our CE-ResNet classification results. We discuss the potential systematics relevant to future cosmological applications.

Unified Astronomy Thesaurus concepts: Large-scale structure of the universe (902); Convolutional neural networks (1938); Galaxy classification systems (582); Astronomy image processing (2306)

1. Introduction

In tidal-torque theory, the angular momentum of galaxies (galaxy spin) is generated by the tidal torque due to the misalignment between the protohalo inertia tensor and the local gravitational tidal shear (Peebles 1969; Doroshkevich 1970; White 1984). Cosmological simulations have confirmed that the direction of dark matter halo spin is well described by the tidal-torque theory (Porciani et al. 2002), and that disk galaxies generally follow dark matter and gain similar spin directions as their host halos (Teklu et al. 2015; Jiang et al. 2019). This makes galaxy spin a promising cosmological probe of various parameters, including the initial condition of the universe, primordial chirality, and the neutrino mass (see e.g., Lee & Pen 2000, 2001; Yu et al. 2019, 2020; Motloch et al. 2021, 2022a, 2022b, for recent investigations).

Recently, Motloch et al. (2021) found a correlation between the observed galaxy spins and the initial density field of the universe. The same galaxy catalog has also been used to search for primordial chirality violations (Motloch et al. 2022a). However, the signal-to-noise ratio is limited by the number of galaxies (∼15,000) with their spin directions observed. In order to optimally extract the information from galaxy spin data to constrain the evolution of our universe, it is necessary to develop new methods to measure the direction of galaxy spin accurately and efficiently.

Assuming that spiral galaxies are well approximated by circular disks, their three-dimensional spin directions can be determined through the position angles and axis ratios, which are readily available from photometric observations, up to a fourfold degeneracy (see Figure 2 of Motloch et al. 2021, for a visual illustration). Iye et al. (2019) visually inspect 842 spiral galaxies and confirm that (1) all the spirals are trailing, i.e., Z-wise spirals rotate clockwise, and (2) the dark, dust-lane-dominant side of the minor axis is closer to us. Therefore, one can break the fourfold degeneracy if one can determine (1) whether the galaxy is a Z-wise or S-wise spiral, and (2) which side of the minor axis is darker and redder.7 In this paper, we focus on the classification of Z-wise versus S-wise spirals, and leave the dark side versus bright side classification for future research.

Galaxy Zoo 1 (GZ1) is a citizen science project that classifies about 9 × 10⁷ galaxies by members of the public (Lintott et al. 2008, 2011).8 It provides information on clockwise or counterclockwise (Z-wise or S-wise spiral pattern) for galaxies from Sloan Digital Sky Survey (SDSS; Abazajian et al. 2009; Almada et al. 2020) data, which leads to a ∼3σ detection of the correlation between the galaxy spin field and cosmological initial conditions (Motloch et al. 2021) and preliminary results for primordial chirality violations (Motloch et al. 2022a).

7 Alternatively, one could use the information regarding which side of the major axis is approaching us (Han et al. 1995; Pen et al. 2000; Motloch et al. 2021). This may be less ambiguous than visually determining the dark side of the minor axis, but requires spectroscopic data and thus cannot be directly obtained from photometric surveys.

8 http://zoo1.galaxyzoo.org/
However, ongoing and forthcoming cosmological surveys, such as the Dark Energy Spectroscopic Instrument (DESI; DESI Collaboration et al. 2016a, 2016b), will map tens of more galaxies than those classified in GZ1, which are prohibitive to be again classified by humans. Machine-learning (ML) based classification methods are required to efficiently identify the morphological properties of galaxies.

Deep convolutional neural networks (CNN) have led to a series of breakthroughs in computer vision during the past ten years (LeCun et al. 1989; Krizhevsky et al. 2012; He et al. 2016; Tan & Le 2019). They are now regarded as the state-of-the-art image classification method and are widely used in general astrophysics applications (Banerji et al. 2010; Huertas-Company et al. 2011; Dieleman et al. 2015; He et al. 2019; Abul Hayat et al. 2020; Yao-Yu Lin et al. 2020). In this paper, we present chirality equivariant residual network (CE-ResNet), an ML-based classifier for the Z-wise versus S-wise spirals. The remainder of this paper is organized as follows: we introduce the data sets in Section 2 and the architecture of our model in Section 3. We present the training and classification results of our model in Section 4 with SDSS images, and in Section 5 with DESI images. The known asymmetry between Z-spirals and S-spirals in GZ1 is discussed in Section 6. We conclude this paper in Section 7. The source code of our CE-ResNet model and the classification catalogs are publicly available.

2. Galaxy Spin Data Sets

In the GZ1 project, public volunteers are asked to classify the SDSS galaxy images into six categories: ellipticals, Z-wise spirals, S-wise spirals, edge-on spirals, star/’do not know, and mergers. The catalog includes the vote counts of the six morphological types for 667,944 galaxies with SDSS spectra data available, and for 225,268 galaxies with no spectra data available. The empirical probability that a galaxy belongs to each category can be estimated by the fraction of votes, while only the probabilities of Z-wise and S-wise spirals, $p_z$ and $p_s$, are relevant for this paper. For simplicity, we effectively treat all the mergers as non-spirals, based on their vote fractions, despite that some mergers may indeed contain Z(S)-spirals. We refer to the empirical probabilities from the GZ1 catalog as the true probability, in contrast to the predicted probability given by the classifiers. To keep the number of different kinds of galaxies roughly balanced, we downsample the galaxies with $p_m \equiv \max(p_z, p_s) \in [0, 0.1]$ by a factor of 20 (i.e., only keep one of 20 such galaxies), the galaxies with $p_m \in (0.1, 0.2]$ by a factor of 5, and the galaxies with $p_m \in (0.2, 0.3]$ by a factor of 2. We then query the SDSS DR16 SQL database with the R.A. and decl. of each galaxy in the GZ1 catalog, and apply the following cuts to remove the galaxies that are unlikely to be clear enough to identify their morphology:

1. There should be exactly one PhotoObj within 1″ of the location in the GZ1 catalog.
2. The error of r-band magnitude should be in (0, 1).
3. The r-band half-light radius $r_h$ should be larger than 1″.
4. The r-band relative error of radius $\sigma_{r_h}/r_h$ should be in (0°, 0.25°).

We find 173,097 galaxies that meet all the criteria (dubbed the “reduced GZ1” catalog henceforth), and use 70% of them for training, 15% for validation, and 15% for testing. We obtain the jpeg images of these galaxies from both SDSS and DESI surveys using the Legacy Survey Sky Viewer tool (Dey et al. 2019). The SDSS DR16 images are generated from the gri bands, while the DESI DR9 images are generated from the grz bands. The numbers of $p_z$ and $p_s$ galaxies with different choices of cutoff values $p_{cut}$ for the vote fraction are listed in Table 1. Sample images for galaxies with different GZ1 morphology classification probabilities can be found in Figure 9.

To assess whether our model can be robustly applied to surveys other than SDSS (with which the model is trained), we also collect all the galaxies in the DESI Legacy Survey Sweep Catalogs that are larger than 1″ in half-light radius and have spectroscopic redshifts available, leading to the “preliminary DESI” catalog of 1,953,246 galaxies. For our model, the image field of view (FOV) is chosen as a multiple of the galaxy size, which however may deviate significantly between different survey measurements (see Figure 8). This requires that our classification model should be insensitive to the image FOV, which will be justified later in this paper (see Figure 4 and the discussion therein).

3. Network Architecture

Empirically, deeper neural networks are more expressive than their shallower analogs, at the cost of being more difficult to train (Mhaskar et al. 2017; Mehta et al. 2019): as the depth of the network increases, its accuracy may get saturated and then eventually decreases rapidly, which is known as the degradation problem (He & Sun 2015; Srivastava et al. 2015). This issue can be mitigated by deep residual networks (ResNet; He et al. 2016), which allow the construction of extremely deep convolution networks and is still considered the state-of-the-art method for computer vision tasks. The key insight of ResNet is that, in principle, deeper networks should perform at least as well as their shallower analogs, since if the subsequent layers are all identity, a deep network becomes equivalent to a shallow network. The degradation problem then implies that it is not easy for neural networks to approximate the identity function with nonlinear layers. Therefore, instead of fitting the target function $f(x)$ directly, it is advantageous to have each nonlinear layer fit $f(x) := f(x) - x$, which can be implemented by a simple shortcut connection (see, e.g., Figure 2 of He et al. 2016). Then, if the linear weights and biases are all zero, we have $f(x) \equiv 0$ and $f(x)$ is identity. This leads to an explicit solution for deeper networks such that they should perform no worse than shallower networks.

We implement our CE-ResNet in pytorch (Paszke et al. 2019), based on the ResNet-50 model in He et al. (2016). The network structure is shown in Table 2. We use the ReLU activation function for the convolution layers and the tanh activation function for the fully connected layers. Our model is

9 https://github.com/k3jia/galaxy_spin_classifier/
10 DOI:10.5281/zenodo.7170929.
11 Note that $p_{z,s}$ throughout this paper is the probability that a random volunteer in the GZ1 project will decide that the galaxy is a Z(S)-spiral, which is not exactly the same as the probability that the galaxy is actually a Z(S)-spiral. However, these two types of probabilities should be positively correlated, and the galaxies with $p_{z,s} \rightarrow 1$ are indeed very likely to be real Z(S)-spirals (see the sample images in Appendix A). Therefore, the empirical probability estimated by the vote fractions is still a useful quantity for the classification of galaxy morphology.
12 If one object has a large $p_{z,s}$, its $p_{z,s}$ should be small, since the vote fractions of the six categories add up to 1.
13 https://www.legacysurvey.org/
A galaxy is classified as a Z(S)-wise spiral if its $p_z(p_s)$ is larger than the cutoff value $p_{cut}$. See the corresponding sections for the details of each combination. We also list the significance values for chirality violation based on the number of Z-wise and S-wise spirals; see Section 6 for more details.

### Table 1

Number of Z-wise and S-wise Spirals in the Different Classification Catalogs

| Catalog       | Image Source | Classifier   | $p_{cut}$ | Z-wise | S-wise | $\sigma$-value | See       |
|---------------|--------------|--------------|-----------|--------|--------|----------------|-----------|
| Reduced GZ1   | SDSS         | GZ1 human    | 0.5       | 35068  | 37022  | 7.278          | Section 2 |
|               |              |              | 0.6       | 27753  | 29535  | 7.445          |           |
|               |              |              | 0.7       | 22064  | 23488  | 6.672          |           |
|               |              |              | 0.8       | 16674  | 17921  | 6.704          |           |
|               |              |              | 0.9       | 10269  | 11283  | 6.907          |           |
| Reduced GZ1   | SDSS         | CE-ResNet    | 0.5       | 35327  | 35667  | 1.276          | Section 4 |
|               |              |              | 0.6       | 27437  | 27737  | 1.277          |           |
|               |              |              | 0.7       | 21036  | 21283  | 1.201          |           |
|               |              |              | 0.8       | 15433  | 15542  | 0.619          |           |
|               |              |              | 0.9       | 9218   | 9442   | 1.640          |           |
| Reduced GZ1   | DESI         | CE-ResNet    | 0.5       | 44362  | 45432  | 3.571          | Section 5 |
|               |              |              | 0.6       | 36892  | 37619  | 2.663          |           |
|               |              |              | 0.7       | 29826  | 30217  | 1.596          |           |
|               |              |              | 0.8       | 22525  | 22804  | 1.310          |           |
|               |              |              | 0.9       | 13583  | 13775  | 1.161          |           |
| Preliminary DESI | DESI     | CE-ResNet    | 0.5       | 55243  | 55526  | 0.850          | Section 5 |
|               |              |              | 0.6       | 41977  | 42269  | 1.006          |           |
|               |              |              | 0.7       | 31508  | 31693  | 0.736          |           |
|               |              |              | 0.8       | 22012  | 22143  | 0.623          |           |
|               |              |              | 0.9       | 11649  | 11919  | 1.759          |           |

**Note.** A galaxy is classified as a Z(S)-wise spiral if its $p_z(p_s)$ is larger than the cutoff value $p_{cut}$. See the corresponding sections for the details of each combination. We also list the significance values for chirality violation based on the number of Z-wise and S-wise spirals; see Section 6 for more details.

The input image size is now $3 \times 160 \times 160$, and the output size for each layer is changed accordingly. (2) We add four additional fully connected layers to improve its expressivity. (3) Each galaxy image is fed into the neural network twice: the same network predicts the scores for Z-spirals and non-spirals given the original image, and the scores for S-spirals and non-spirals given the flipped image. We then average the two estimates of the non-spiral score, and apply a softmax function to get the probabilities for the three categories (see Figure 1). This guarantees that the network is purely parity even, which is crucial for cosmological galaxy spin analysis. In practice, we flip the images about the vertical axis (i.e., left to right), as illustrated in Figure 1, while we note that the direction of the flip should not affect the performance of our model, which is insensitive to the orientation of the input image (see Figure 4).

Note that we directly use the network to predict the scores and probabilities; we do not select a cutoff value and divide the galaxies into discrete categories of Z-spirals, S-spirals, and non-spirals before training. Unlike standard ML data sets, such as MNIST (Lecun et al. 1998), where it is obvious whether one image is a handwritten number 8 or not, many galaxy images are not clear enough so that one can assert they belong to a certain category of morphology. Although galaxies with larger $p_{z/s}$ are more likely to be real Z/S-spirals, there is no simple cutoff value such that all the galaxies with $p_{z/s}$ above it are Z/S-spirals while all the galaxies below it are not: as shown in Figure 9, the images just change continuously with respect to $p_{z/s}$. The value of $p_{z/s}$ in the GZ1 catalog is indeed a stochastic variable: assuming that the vote on one single galaxy from each person follows some i.i.d. distribution based on its morphology and image quality, the fraction of total votes for each category does asymptotically converge with infinite vote size, but will always have finite noise when the vote size is limited. If one pre-divides the galaxies into discrete categories, there are always mislabeled galaxies around the cutoff probability, which will be confusing to the classifier. Therefore, we stick...
with estimating the scores directly, and leave the interpretation of the output probabilities to the user.

4. Training and Results

The inputs of our model are three-channel RGB images of shape $3 \times 160 \times 160$, with the target galaxy centered and the image FOV equal to 5 times the galaxy half-light diameter. Since the chirality of a spiral galaxy should have no dependence on its location, orientation, size, and color, we apply the following data augmentation during training, which are illustrated in Figure 2:

1. The relative location of the galaxy is moved by up to 25% of its half-light radius along both directions, with the exact translation randomly sampled from the uniform distribution $U(-0.25 \, r_{50}, 0.25 \, r_{50})$.
2. The galaxy is rotated by a random angle, sampled from the uniform distribution $U(0^\circ, 360^\circ)$.
3. The exact FOV of training images is sampled from a uniform distribution between 4 and 6 times the galaxy half-light diameter.
4. We apply a random permutation to the three color channels, to mitigate the potential overfitting to the correlation between galaxy color and morphology (e.g., Bamford et al. 2009). This will also improve the model’s generalizability to images from other surveys, with possibly different filters, instrumental responses, and color scales.

Here we stick with three-channel images and do not average the different bands or use only the band with the best signal-to-noise ratio: although the color of the galaxy should contain no chirality information, it is possible that, for example, the spiral arms where more new stars are forming look bluer than the region between the arms, so that the spiral structure is less clear in single-channel images compared with the full three-channel images.

We train our CE-ResNet model for 120 epochs using the cross-entropy loss and the AdamW optimizer (Loshchilov & Hutter 2019), which takes about 1 day on one NVIDIA V100 GPU. We set the weight decay coefficient to 1 and use an initial learning rate of 0.0001, and after every five epochs, we reduce the learning rate by a factor of 15%. The training and validation losses are shown in Figure 3. Sample images of different $P_s/s, p, n$ are shown in Figure 10, which are similar to those from the original GZ1 classification (Figure 9). The total number of spiral galaxies is also close to the original human classification (Table 1).

We test the robustness of our model under the data augmentation transforms shown in Figure 4, with only one type of transform activated each time. Note that in the first panel, we apply a stochastic transform similar to the one used during training, with the maximum translation indicated by the horizontal axis. For the other three panels, the same deterministic transform is applied to all the images in each case. The loss function is almost constant within the extent of training data augmentation indicated by the gray bands, meaning that our model is indeed insensitive to these transforms. Beyond the gray bands, the loss function does increase when a large translation or a small FOV is used, probably because part of the galaxies are cut out from the images. This implies that when the model is applied to another catalog without reliable galaxy radius measurements, one should consider setting the image FOV slightly larger rather than smaller, to make sure that the whole galaxy is contained in the image.

In Figure 5, we show two-dimensional histograms of the true versus predicted GZ1 vote probabilities for the Z-wise and S-wise spirals, which is roughly diagonal with a small dispersion of $\leq 0.1$. Is this a good result? Since our training data is from the volunteer vote fractions, the dispersion cannot be smaller than the Poisson noise in the data. Assuming that the votes follow an i.i.d. binomial distribution, the empirical vote fraction will have a standard error of $\Delta p = \sqrt{p(1 - p)/n}$, which equals 0.073 if $p = 0.8$ and $n = 30$. Therefore, the performance of our model is already close to the best allowed by the training data. We also note an excess of galaxies in the $0 < P_s/s, true < 0.1, 0.1 < P_s/s, pred < 0.3$ bins, due to the relatively large number of galaxies with $0 < P_s/s, true < 0.1$. However, such small $P_s/s$ galaxies are rarely relevant to practical applications, since their spin directions are mostly undetermined.

We check whether the classification accuracy systematically depends on certain galaxy parameters in Figure 6. All galaxies in the test data set are binned according to their redshift $z$, half-light radius $r_{50}$, $r$-band magnitude $m_r$, color $m_g - m_r$, aspect ratio $b/a$, orientation $\phi$, as well as the true $P_s/s$. We plot the error of the $P_m$ prediction and the number of galaxies for each bin. Generally, the error is the smallest for...
We also apply our model to the preliminary DESI catalog, which indeed has a large overlap with the GZ1 catalog, as SDSS contributed most of the current galaxy spectra data before DESI spectra become available. We find that 150,283 of the 173,097 galaxies in the reduced GZ1 catalog can be matched to one galaxy within 1″ in the preliminary DESI catalog. However, these galaxies have slightly different locations and radius measurements in DESI Legacy Surveys, which are required to determine the image cut for our model input. We thus use the preliminary DESI catalog to validate the accuracy of our model with DESI photometric measurements, since the future DESI spectra catalog will have the galaxy location and radius measured with the same pipeline.

We compare the predicted $p_m$ for the galaxies matched between the reduced GZ1 and preliminary DESI catalogs. In principle, $p_{m,pred}$ should be close between these two cases, since we are classifying the same galaxies using the same imaging survey but just with different image cuts. As shown in Table 1, the total number of Z-spirals and S-spirals is similar to that in the reduced GZ1 catalog with DESI images, with slightly fewer galaxies if one chooses $p_{cut} = 0.9$ but more galaxies if one chooses $p_{cut} \in [0.5, 0.7]$. We find that $\Delta p_{m,pred}$ has a strong correlation with the difference in the measured galaxy radius: most of the galaxies have their radii measured larger by about 30% in the DESI catalog, which however only leads to a small impact on the predicted $p_m (<0.05)$ since our model is insensitive to a reasonable number of changes in the image FOV (Figure 4). On the other hand, some large, dim galaxies have $>100\%$ difference in radius measurements, which leads to a larger dispersion in $\Delta p_{m,pred}$. This should only have limited effects on cosmological applications of our model, however, as the number of such galaxies is relatively small according to Figure 6. Although in practice, one may further improve the performance by applying an empirical correction for the galaxy radii to compensate for the difference between surveys, based on, for example, the general trend of $r_{DESI}/r_{SDSS}$ in Figure 8, which we leave for future research.

6. Chirality Violation Due to Human Bias

There is a well-known bias toward S-wise spirals in the GZ1 classification catalog, which has been intensively investigated in a study of bias by Galaxy Zoo (Lintott et al. 2008, 2011). The volunteers were presented with flipped galaxy images but still found an excess of S-wise spirals. It is shown that such bias is due to the human selection effect, while the chirality of spiral galaxies is not violated once the human bias is appropriately accounted for (Land et al. 2008; Hayes et al. 2017).

To validate that our CE-ResNet model is parity even, we check the symmetry between the numbers of Z-spirals and S-spirals when the desired galaxy counts in Table 1 are matched rather than when $p_{cut} = 0.7$ for the original human classification catalog, which has been intensively investigated in a study of bias by Galaxy Zoo (Lintott et al. 2008, 2011). The volunteers were presented with flipped galaxy images but still found an excess of S-wise spirals. It is shown that such bias is due to the human selection effect, while the chirality of spiral galaxies is not violated once the human bias is appropriately accounted for (Land et al. 2008; Hayes et al. 2017).
Spirals using the classification catalogs in Table 1. We use the following statistics to determine the significance of chirality violation,

\[ T = \frac{n_z - n_s}{\sqrt{n_z + n_s}}, \]

which should asymptotically follow the standard Gaussian distribution under the null hypothesis of no chirality violation. See Appendix B for the derivation.

Note that our reduced GZ1 catalog contains all the \( p_m > 0.3 \) galaxies in the full GZ1 catalog, while the \( p_m \leq 0.3 \) galaxies are downsampled. Similar to Hayes et al. (2017), we find a \( \sim 7\sigma \) asymmetry in the GZ1 human classification catalog, which disappears if the same SDSS images are classified by the parity-even CE-ResNet, implying that the asymmetry is due to a slight underestimate of \( p_z \) relative to \( p_s \) by GZ1 humans. However, when DESI instead of SDSS images are used for the reduced GZ1 galaxies, there are again slightly more S-wise than Z-wise galaxies with \( p_{\text{cut}} = 0.5 \). This is likely because some of the \( p_z \sim 0.5 \) (with DESI images) galaxies are cut out from the reduced GZ1 catalog since they may have \( p_z < 0.3 \) by GZ1 humans. When applied to the preliminary DESI catalog, our CE-ResNet finds an equal number (\( \sim 1.8\sigma \) asymmetry) of Z-spirals and S-spirals for all different choices of \( p_{\text{cut}} \), confirming that no real chirality violation between these two types of spirals exists in nature.

7. Discussions

In this paper, we present CE-ResNet, an ML-based model for the classification of Z-wise versus S-wise spirals. Trained with GZ1 data, our model gives similar predictions on the chirality of galaxies as the volunteers in the GZ1 project, but can be efficiently applied to the millions or even billions of galaxies that will be mapped in the near feature. Our model is manifestly parity even, since basically the same estimator is used to predict the probabilities of Z-spirals and S-spirals, using the technique in which one gets an S-spiral if one flips the image of a Z-spiral. Heuristically, humans have different selection functions for Z- and S-spirals, while our model learns the average of these two selection functions from humans and uses it to classify both types of spirals, which explains how we can train an unbiased classifier from a biased data set. We validate our model in Section 4, and verify that our model can be directly applied to DESI images even though it is trained with SDSS images in Section 5. We confirm that no real asymmetry between the numbers of Z-wise and S-wise spirals exists in Section 6.

We note a few related works in the literature. Hayes et al. (2017) develop an unbiased selector to demonstrate that the excess of S-spirals in the GZ1 catalog is due to the human selection effect. However, their unbiased machine spirality selector finds significantly fewer spirals than GZ1 humans, such that it cannot maximally extract the information in the survey data for cosmological analysis. Recently, Tadaki et al. (2020) studied the classification of spiral galaxies with CNN.
However, their data set only includes galaxies that are unambiguously Z-spirals, S-spirals, and non-spirals, whereas real-world survey catalogs also include galaxies with unclear morphological type due to the limitation of image quality. Their model predictions on these unclear galaxies can be undefined as the network has never seen such galaxies during training, making it risky to be directly applied to full survey catalogs. Also, the classifier in Tadaki et al. (2020) is not manifestly parity even, so one should be cautious about the possible inherent asymmetry between Z-spirals and S-spirals in their model.

The network used in this paper is equivariant under a reflection of the input image, which eliminates the potential bias caused by the difference between the Z- and S-type estimators. Additionally, we augment the data set by a random translation, rotation, scaling, and permutation of color channels, to mitigate potential overfitting onto the irrelevant position, orientation, size, and color information. According to Figure 4, our network is indeed stable under such transforms. We note that in principle, one can implement a more advanced equivariant network that manifestly accounts for all the relevant symmetries (e.g., Sosnovik et al. 2019; Weiler & Cesa 2019; Zhang 2019; Cesa et al. 2022). Also, domain adaptation techniques may help improve the performance when the model needs to be applied to data from different surveys (Ben-David et al. 2010; Alexander et al. 2021). We leave these directions for future research, since our current architecture already works well in the various benchmarks demonstrated in this paper.

We thank Jiaxuan Li, Mingwei Ma, and Zhehao Xu for helpful discussions. Ue-Li Pen receives support from Ontario Research Fund-Research Excellence Program (ORF-RE), Natural Sciences and Engineering Research Council of Canada (NSERC) [funding reference number RGPIN-2019-067, CRD 523638-18, 555585-20], Canadian Institute for Advanced Research (CIFAR), Canadian Foundation for Innovation (CFI), the National Science Foundation of China (grant No. 11929301), Thoth Technology Inc, Alexander von Humboldt Foundation, and the National Science and Technology Council.
Computations were performed on the Mist supercomputer at the SciNet HPC Consortium and the SOSCIP Consortium’s GPU computing platform. SciNet is funded by the Canada Foundation for Innovation; the Government of Ontario; Ontario Research Fund—Research Excellence; and the University of Toronto (Loken et al. 2010). SOSCIP is funded by the Federal Economic Development Agency of Southern Ontario, the Province of Ontario, IBM Canada Ltd., Ontario Centers of Excellence, Mitacs, and 15 Ontario academic member institutions.

The Legacy Survey consists of three individual and complementary projects: the Dark Energy Camera Legacy Survey (DECaLS; Proposal ID #2014B-0404; PIs: David Schlegel and Arjun Dey), the Beijing-Arizona Sky Survey (BASS; NOAO Prop. ID #2015A-0801; PIs: Zhou Xu and Xiaohui Fan), and the Mayall z-band Legacy Survey (MzLS; Prop. ID #2016A-0453; PI: Arjun Dey). DECaLS, BASS, and MzLS together include data obtained, respectively, at the Blanco telescope, Cerro Tololo Inter-American Observatory, NSF’s NOIRLab, the Bok telescope, Steward Observatory, University of Arizona, and the Mayall telescope, Kitt Peak National Observatory, NOIRLab. Pipeline processing and analyses of the data were supported by NOIRLab and the Lawrence Berkeley National Laboratory (LBNL). The Legacy Survey project is honored to have been permitted to conduct astronomical research on Iolkam Du’ag (Kitt Peak), a mountain with particular significance to the Tohono O’odham Nation.

NOIRLab is operated by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation. LBNL is managed by the Regents of the University of California under contract to the U.S. Department of Energy.

This project used data obtained with the Dark Energy Camera (DECam), which was constructed by the DES Collaboration. Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundacao Carlos Chagas Filho de Amparo, Financiadora de Estudos e Projetos, Fundacao Carlos Chagas Filho de Amparo e Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministerio da Ciência, Tecnologia e Inovação, the Deutsche Forschungsgemeinschaft and the Collaborating Institutions in the Dark Energy Survey. The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenössische Technische Hochschule (ETH) Zurich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciencies de l’Espai (IEEC/CSIC), the Institut de Física d’Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig Maximilians Universität München and the associated Excellence Cluster Universe, the University of Michigan, NSF’s NOIRLab, the University of Nottingham, the Ohio State University, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford...
University, the University of Sussex, and Texas A&M University.

BASS is a key project of the Telescope Access Program (TAP), which has been funded by the National Astronomical Observatories of China, the Chinese Academy of Sciences (the Strategic Priority Research Program “The Emergence of Cosmological Structures” grant No. XDB09000000), and the Special Fund for Astronomy from the Ministry of Finance. The BASS is also supported by the External Cooperation Program of the Chinese Academy of Sciences (grant No. 114A11KYSB20160057), and the Chinese National Natural Science Foundation (grant Nos. 12120101003 and 11433005).

The Legacy Survey team makes use of data products from the Near-Earth Object Wide-field Infrared Survey Explorer (NEOWISE), which is a project of the Jet Propulsion Laboratory/California Institute of Technology. NEOWISE is funded by the National Aeronautics and Space Administration.

The Legacy Survey imaging of the DESI footprint is supported by the Director, Office of Science, Office of High Energy Physics of the U.S. Department of Energy under contract No. DE-AC02-05CH1123, by the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility under the same contract; and by the U.S. National Science Foundation, Division of Astronomical Sciences under contract No. AST-0950945 to NOAO.

Appendix A
Sample Galaxy Images

We show sample galaxy images for the various classification catalogs in Figures 9–11. They all come from the GZ1 catalog, but are classified by GZ1 humans with SDSS images, by CE-ResNet with SDSS images, and by CE-ResNet with DESI images, respectively. In Figures 9 and 10, we randomly choose 10 galaxies in each $p_{z/s}$ bin of width 0.1. Figure 11 shows the same galaxies as Figure 10; it is clear that the better images of DESI led to the increase in the $p_{z/s}$ prediction.
Figure 9. Sample SDSS images for the reduced GZ1 catalog, with classification probabilities given by GZ1 humans.
Figure 10. Sample SDSS images for the reduced GZ1 catalog, with classification probabilities predicted by CE-ResNet.
Figure 11. The same galaxies as in Figure 10 but classified with DESI images. The better imaging quality enables more confident morphology classification for many galaxies.
Appendix B
Derivation of Equation (1)

Suppose that we have observed $n$ galaxies in total, of which $n_z$ are Z-wise spirals and $n_s$ are S-wise spirals. Let the operator $1_z$ be equal to 1 if a galaxy is Z-wise and 0 otherwise, and similarly for the operator $1_s$. Under the null hypothesis of no chirality violation, we have $\exp(1_z - 1_s) = 0$, and

$$\text{var}(1_z - 1_s) = \exp \left[ \left(1_z - 1_s \right)^2 \right] - \left[ \exp \left(1_z - 1_s \right) \right]^2 = \frac{n_z + n_s}{n} \rightarrow \infty,$$

(B1)

where we have used $\exp(1_z) = 0$ as one galaxy cannot be both Z-wise and S-wise. According to the central limit theorem, in the limit of large $n$,

$$\sqrt{n} \left(1_z - 1_s\right) \sim N \left(0, \left(\frac{n_z + n_s}{n}\right)^2\right),$$

(B2)

from which it is straightforward to show that the $T$ statistics in Equation (1) follow the standard Gaussian distribution.

ORCID iDs

He Jia (贾赫) © https://orcid.org/0000-0002-9958-7758

Hong-Ming Zhu (朱弘明) © https://orcid.org/0000-0002-8202-8642

Ue-Li Pen (彭威禮) © https://orcid.org/0000-0003-2155-9578

References

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543

Abul Hayat, M., Stein, G., Harrington, P., Lukić, Z., & Mustafa, M. 2020, ApJL, 911, 33

Ahumada, R., Prieto, C. A., Almeida, A., et al. 2020, ApJS, 249, 3

Alexander, S., Gleyzer, S., Reddy, P., Tidball, M., & Toomey, M. W. 2021, arXiv:2112.12121

Bamford, S. P., Nichol, R. C., Baldry, I. K., et al. 2009, MNRAS, 393, 1524

Banerji, M., Lahav, O., Lintott, C. J., et al. 2010, MNRAS, 406, 542

Ben-David, S., Blitzer, J., Crammer, K., et al. 2010, Machine Learning, 79, 151

Cesa, G., Lang, L., & Weiler, M. 2022, in 10th Int. Conf. on Learning Representations, ICLR https://openreview.net/forum?id=BkgDRCyQ7Y

Mehta, P., Bukov, M., Wang, C.-H., et al. 2019, PhR, 810, 1

Mhaskar, H., Liao, Q., & Poggio, T. 2017, in Proc. AAAI Conf. on Artificial Intelligence, 31

Motloch, P., Pen, U.-L., & Yu, H.-R. 2022a, PhRvD, 105, 083512

Motloch, P., Pen, U.-L., & Yu, H.-R. 2022b, PhRvD, 105, 083504

Motloch, P., Yu, H.-R., Pen, U.-L., & Xie, Y. 2021, NatAs, 5, 283

Paszke, A., Gross, S., Massa, F., et al. 2019, in Advances in Neural Information Processing Systems, 32, ed. H. Wallach et al. (Curran Associates Inc.), 8024, https://papers.nips.cc/paper/2019/hash/bdbca288e7922fde9d7f0212773740-Abstract.html

Peelers, P. J. E. 1969, ApJ, 155, 393

Pen, U.-L., Lee, J., & Seljak, U. 2000, ApJL, 543, L107

Porciani, C., Dekel, A., & Hoffman, Y. 2002, MNRAS, 332, 325

Sosnovik, I., Sznaja, M., & Smeudders, A. 2019, in 8th Int. Conf. on Learning Representations https://openreview.net/forum?id=H1gppugKPS

Srivastava, R. K., Greff, K., & Schmidhuber, J. 2015, in Advances in Neural Information Processing Systems, 28, https://proceedings.neurips.cc/paper/2015/hash/215a71a2769b656c3e32e7299f1c5ed-Abstract.html

Tadaki, K.-i., Iye, M., Fukushima, H., et al. 2020, MNRAS, 496, 4276

Tan, M., & Le, Q. 2019, in 36th Int. Conf. on Machine Learning Research, 97, ed. K. Chaudhuri & R. Salakhutdinov (PMLR), 6105, https://proceedings.mlr.press/v97/tan19a.html

Tsku, A. F., Remus, R.-S., Dolag, K., et al. 2015, ApJ, 812, 29

Weiler, M. & Cesa, G. 2019, in Advances in Neural Information Processing Systems, 32, https://papers.nips.cc/paper/2019/hash/45df6637b718df024a237069f41b0d4-Abstract.html

White, S. D. M. 1984, ApJ, 286, 38

Yao-Yu Lin, J., Wong, G. N., Prather, B. S., & Gammie, C. F. 2020, arXiv:2007.00794

Yu, H.-R., Motloch, P., Pen, U.-L., et al. 2020, PhRvL, 124, 101302

Yu, H.-R., Pen, U.-L., & Wang, X. 2019, PhRvD, 99, 123532

Zhang, R. 2019, in Proc. 36th Int. Conf. on Machine Learning Research, 97, (PMLR), 7324, https://proceedings.mlr.press/v97/zhang19a.html