

Spin parity of spiral galaxies II: a catalogue of 80 k spiral galaxies using big data from the Subaru Hyper Suprime-Cam survey and deep learning

Ken-ichi Tadaki¹,¹★ Masanori Iye,¹ Hideya Fukumoto,² Masao Hayashi¹,¹ Cristian E. Rusu¹,¹ Rhythm Shimakawa¹ and Tomoka Tosaki³

¹National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan

²The Open University of Japan, 2-11 Wakaba, Mihama-ku, Chiba 261-8586, Japan

³Joetsu University of Education, Yamayashiki-machi, Joetsu, Niigata 943-8512, Japan

Accepted 2020 June 23. Received 2020 May 12; in original form 2020 March 24

ABSTRACT

We report an automated morphological classification of galaxies into S-wise spirals, Z-wise spirals, and non-spirals using big image data taken from Subaru/Hyper Suprime-Cam (HSC) Survey and a convolutional neural network (CNN)-based deep learning technique. The HSC *i*-band images are about 36 times deeper than those from the Sloan Digital Sky Survey (SDSS) and have a two times higher spatial resolution, allowing us to identify substructures such as spiral arms and bars in galaxies at $z > 0.1$. We train CNN classifiers by using HSC images of 1447 S-spirals, 1382 Z-spirals, and 51 650 non-spirals. As the number of images in each class is unbalanced, we augment the data of spiral galaxies by horizontal flipping, rotation, and rescaling of images to make the numbers of three classes similar. The trained CNN models correctly classify 97.5 per cent of the validation data, which is not used for training. We apply the CNNs to HSC images of a half million galaxies with an *i*-band magnitude of $i < 20$ over an area of 320 deg². 37 917 S-spirals and 38 718 Z-spirals are identified, indicating no significant difference between the numbers of two classes. Among a total of 76 635 spiral galaxies, 48 576 are located at $z > 0.2$, where we are hardly able to identify spiral arms in the SDSS images. Our attempt demonstrates that a combination of the HSC big data and CNNs has a large potential to classify various types of morphology such as bars, mergers, and strongly lensed objects.

Key words: techniques: image processing – catalogues – galaxies: spiral.

1 INTRODUCTION

Spiral arms are one of the most beautiful structures in galaxies and attract the interest of many people. Density wave theory is a traditionally accepted concept to explain spiral patterns in galaxies (Lin & Shu 1964) whereas recent numerical simulations support that spiral arms are formed due to a swing amplification associated with galactic shear motion (e.g. Baba, Saitoh & Wada 2013; Dobbs & Baba 2014), which was originally suggested by Goldreich & Lynden-Bell (1965). The winding direction of spiral arms with respect to the galaxy rotation direction has also been a classic subject of controversial debates in observational and theoretical studies of spiral galaxies until around the decade of the 1960s. Iye, Tadaki & Fukumoto (2019) demonstrate a corroborative evidence that all galaxies are trailing spirals provided that the dark lane dominant side is the side of the disc near to us. Once the winding direction of spiral arms is identified, we can infer the spin vector of galaxies, which is one of important physical properties in the process of galaxy formation. In the framework of the tidal torque theory (e.g. Peebles 1969; Doroshkevich 1970; White 1984), galaxies acquire angular momentum by the tidal fields of their neighbours in the linear stage of structure formation. Thus, the winding direction, which is the

spin vectors, of galaxies would have been randomly located, leading to the isotropic distribution that the number of S-wise spirals (S-spirals) is identical to that of Z-wise spirals (Z-spirals). If there is a global anisotropy in the spatial distribution of the winding direction, a large-scale vorticity such as galaxy-cluster tidal interaction would affect the spin of galaxies (Sugai & Iye 1995). However, a statistical analysis of the winding direction was not well developed in the past two decades, except for a few studies (Hayes, Davis & Silva 2017; Shamir 2017).

We can immediately identify spiral arms in a galaxy and judge whether it is a S-spiral or a Z-spiral by visual inspection. However, it is not easy to repeat this procedure 10 000 times or more. Another problem is that visual classification depends on the expertise and experience of people who look at images. In the Galaxy Zoo project (Lintott et al. 2011; Willett et al. 2013), about 100 000 volunteers classified the morphology of $\sim 900\,000$ galaxies at $0.001 < z < 0.25$, drawn from the Sloan Digital Sky Survey (SDSS; York et al. 2000). Masters et al. (2010) studied 5433 face-on spiral galaxies at $0.03 < z < 0.085$ from the Galaxy Zoo data base. To make a similarly large sample of more distant spiral galaxies at $z > 0.1$, we require higher sensitivity and higher resolution imaging data set over a wide area.

We are conducting a multiband imaging survey by using Hyper Suprime-Cam (HSC) in Subaru Strategic Program (HSC-SSP; Aihara et al. 2018). The HSC has the largest field of view of 1.5-deg diameter on 8-m class telescopes. The Wide layer of the survey

* E-mail: tadaki.ken@nao.ac.jp

covers 1400 deg^2 in five broad-bands (*grizy*) with a $5\text{-}\sigma$ point-source depth of $i \sim 26.2$, which is about 3.9 magnitudes deeper than SDSS (York et al. 2000). The increased sensitivity allows us to characterize spiral arms in distant galaxies at $z > 0.1$ while at the same time the wide survey produces images of more than one million galaxies. We therefore need to develop an automated method for morphological classification in the big data era. Commonly used parametric methods such as Sérsic model fitting (e.g. Peng et al. 2010) and non-parametric ones such as the concentration (C), asymmetry (A), clumpiness (S) method, and the Gini/ M_{20} parameters (e.g. Conselice 2014) are not suitable for identifying substructures in galaxies, such as spiral arms, bars, and tidal streams. Currently, only several studies succeed in automatically extracting spiral structures (Davis & Hayes 2014; Kuminski & Shamir 2016; Hart et al. 2017).

In 2012, deep learning has been dramatically developed enough to correctly recognize the picture of a cat as a cat with high accuracy of ~ 84 per cent (Krizhevsky, Sutskever & Hinton 2012). The accuracy of image classification has exceeded human accuracy in 2015 (~ 95 per cent; He et al. 2016). A convolutional neural network (CNN) is now a commonly used technique for classifying images into multiple categories (e.g. Fukushima 1980; LeCun et al. 1998; Russakovsky et al. 2015). CNNs convolve images with multiple kernels (filters) to reduce the amount of information and efficiently extract local features in images. Dieleman, Willett & Dambre (2015) have applied a CNN technique to astronomical images for galaxy morphology classification and successfully reproduced the results from the Galaxy Zoo with the accuracy of 99 per cent (see also e.g. Huertas-Company et al. 2015; Abraham et al. 2018; Domínguez Sánchez et al. 2018). These approaches are mostly supervised learning, which requires a training data set with pre-labelled images. On the other hand, there are some studies that adopt an unsupervised learning approach for automated morphological classification (e.g. Hocking et al. 2018; Martin et al. 2019). Furthermore, Schawinski et al. (2017) generate super resolution images from artificially degraded low-resolution images using a generative adversarial network (Goodfellow et al. 2014) although they caution about application to unknown galaxy population, which is not included in a training data set. CNNs and other deep learning techniques are becoming increasingly common in astronomy.

In this paper, we present CNN models to identify spiral arms in galaxies by using the HSC imaging data. In Section 2, we describe the imaging data taken from the HSC-SSP survey and build a training data set. We show the architecture of CNNs and estimate the accuracy of morphological classification by using validation data sets in Section 3. We apply the trained CNNs to an unlabelled data set of a half million galaxy images and present a catalogue of 80k spiral galaxies in Section 4.

2 DATA

2.1 Subaru Hyper Suprime-Cam data sample

This work is based on data from the second public data release (PDR-2) of the HSC-SSP for the Wide layer (Aihara et al. 2019). For morphological classification, we use *i*-band images, which have reached an exposure time of about 20 min. Fig. 1 shows the images from the HSC-SSP data and SDSS for two spiral galaxies, demonstrating the superb image quality with which we can identify the spiral winding sense even in distant galaxies at $z > 0.1$.

We use galaxies with 5.7 arcsec aperture magnitudes $i < 20$ so that we can visually classify their morphology. It is still possible

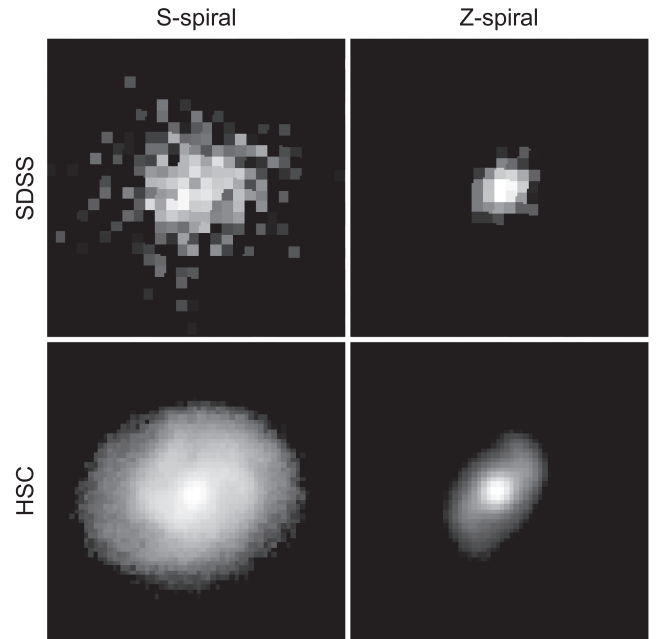


Figure 1. Comparisons between SDSS and HSC *i*-band images for an S-wise spiral galaxy (left) with $i = 18.7$ at $z = 0.16$ and a Z-wise spiral galaxy (right) with $i = 18.8$ at $z = 0.19$. The image sizes are all $10.8 \text{ arcsec} \times 10.8 \text{ arcsec}$.

to identify spiral arms in HSC images even for galaxies with $i \sim 21$, allowing for morphological classification of galaxies at higher redshift. However, as such cases are rare, the inclusion of fainter objects makes it difficult to identify a larger number of spiral galaxies for training CNNs. We therefore choose a magnitude cut of $i < 20$ in this work.

Stars are removed in advance by the flag of `i_extendedness_value` in the HSC-SSP data base. As most galaxies are observed in other four broad-band filters (*grzy*), their photometric redshift is available, provided by the Direct Empirical Photometric code (DEMP: Hsieh & Yee 2014). For galaxies with $i < 20$, the photometric redshift error of $\Delta z = (z_{\text{phot}} - z_{\text{spec}})/(1 + z_{\text{spec}})$ and the outlier fraction of $|\Delta z| > 0.15$ is $\sigma_{\Delta z} = 0.02$ and 5–10 per cent, respectively (Tanaka et al. 2018; Nishizawa et al. 2020). We remove nearby galaxies with spectroscopic redshift of $z_{\text{spec}} < 0.05$, provided by VIPERS (Garilli et al. 2014), SDSS (Alam et al. 2015), Wiggle-Z (Drinkwater et al. 2010), GAMA (Liske et al. 2015), and PRIMUS (Cool et al. 2013), as the physical scale resolution becomes significantly different from that at $z > 0.1$. Edge-on-like objects with a major-to-minor axial ratio of less than 0.1 are also removed in advance because it is hard to distinguish between spirals and non-spirals. For a subsample of 56 787 galaxies with $i < 20$ in the Wide XMM Large Scale Structure survey (XMM-LSS) field, the redshift and stellar mass distributions are shown in Fig. 2. The stellar mass is estimated from multiwavelength photometry, empirically given by the DEM code (Hsieh & Yee 2014). We also derive the completeness as a function of redshift and stellar mass by calculating the ratio of the number of galaxies with $i < 20$ to the number of fainter galaxies with $i < 22$. The stellar mass 90 per cent completeness limits are $\log(M_*/M_\odot) \sim 10$ at $z = 0.2$ and $\log(M_*/M_\odot) \sim 11$ at $z = 0.4$.

We convert FITS images of galaxies to Joint Photographic Experts Group (JPEG) format by using STIFF software (Bertin 2012). We adopt GAMMA = 2.2 to automatically adjust the contrast and

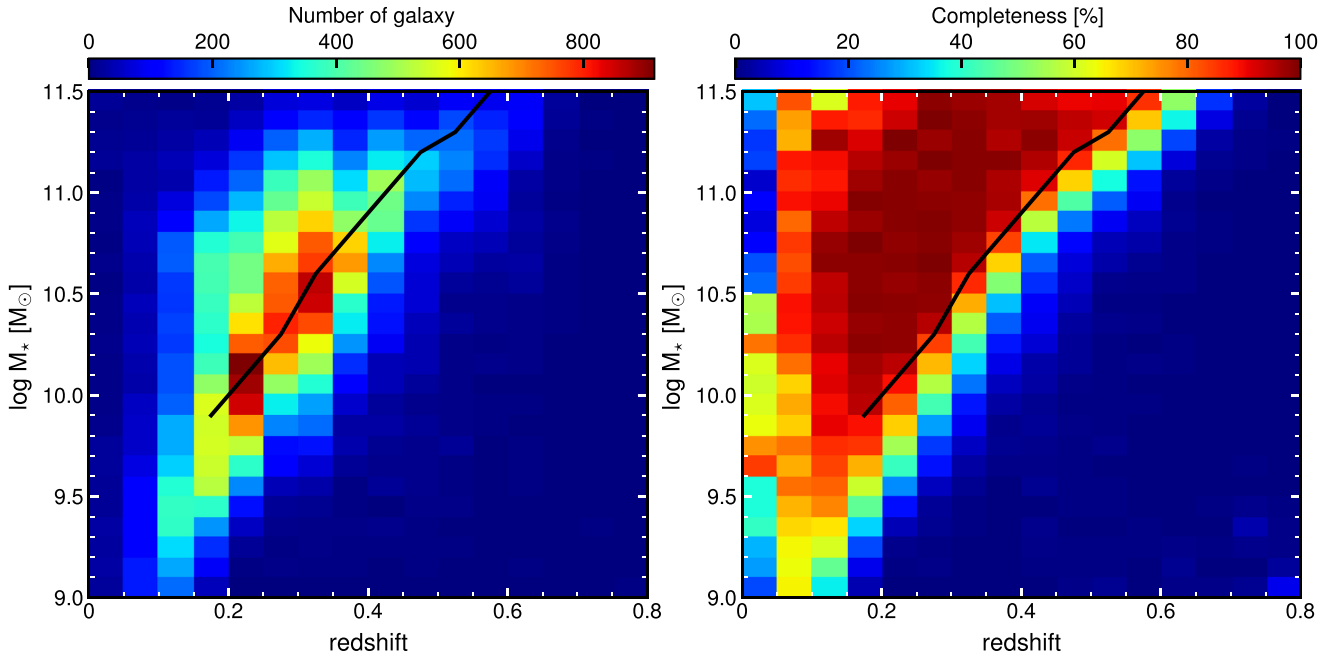


Figure 2. Left: stellar mass versus redshift for galaxies in a subsample of 56 787 galaxies with $i < 20$ in the Wide XMM-LSS field. Right: completeness as a function of stellar mass and redshift. Black lines denote the completeness of 90 per cent.

brightness of JPEG images for classification and slightly change this parameter for data augmentation (Section 2.2). Converting images to JPEG images could potentially lose information respect to the original FITS images. Optimization of the grey-scale images is one of the key challenges for improving classification with deep learning but is beyond the scope of our work. In this paper, we simply use JPEG images by following the previous works (e.g. Dieleman et al. 2015; Huertas-Company et al. 2015; Abraham et al. 2018; Domínguez Sánchez et al. 2018). The size of post-stamp images is 64 pixel \times 64 pixel, covering 10.8 arcsec \times 10.8 arcsec, where main spiral features are covered for most galaxies.

Although colour composite images are often used for classifying galaxy morphology, we use monochromatic images in the i -band for two reasons. First, i -band observations are executed in the best observing conditions for cosmic shear measurements (Mandelbaum et al. 2018). The median seeing is 0.6 arcsec in the i -band, corresponding to 1.1 kpc at $z = 0.1$ and 2.1 kpc at $z = 0.2$, while it is 0.7 ± 0.8 arcsec in other bands. Secondly, composite images have the information of galaxy colours as well as galaxy morphology. There is a strong correlation between colour and morphology: blue galaxies tend to have a disc with spiral arms while red galaxies are ellipticals (e.g. Strateva et al. 2001). Red spiral galaxies are likely to be an important population for understanding transitions from blue to red galaxies (Masters et al. 2010) but the number density is smaller compared to blue spiral galaxies. If colour information is taken into account, the trained models would tend to classify red galaxies into non-spirals rather than spirals. We therefore use i -band images to avoid the colour bias and keep morphology information independent from colours.

2.2 Training data set

We chose the XMM-LSS field (Pierre et al. 2004) over an area of ~ 28 deg² to make a training data set for CNNs. Among 56 787

objects, we confirm 1447 spiral galaxies with clear S-wise spiral structure and 1382 with clear Z-wise spiral structure by visual inspection. They are visually checked by all of the authors. Additional 1177 and 1131 galaxies are identified to have S-wise and Z-wise spiral structure with somewhat reduced confidence level. To define as clearly as possible spirals, we classify these galaxies into a category of unclear/dubious, which is not used for training CNNs. The remaining 51 650 galaxies are non-spiral galaxies. In this work, we do not distinguish between mergers and non-mergers. Even if spiral galaxies are clearly affected by tidal interactions with their companions, they are categorized as S-spirals or Z-spirals. We show example images of randomly selected 100 galaxies in each class from the training data set in Appendix A.

We adopt a K-fold cross-validation technique to evaluate the performance of CNN models. We randomly divide the original sample in each class into five subsets, which consist of 289 S-spirals, 276 Z-spirals, and 10 330 non-spirals. Four of them (1156 S-spirals, 1104 Z-spirals, and 41 320 non-spirals) are used for the training and the remaining one is used for validation. As the numbers of S-spirals and Z-spirals are much smaller than that of non-spirals, we augment the data of spiral galaxies. We add horizontally flipped images of Z-spirals and S-spirals to the S-spiral and Z-spiral classes, resulting in the same number in S-spirals and Z-spirals. Flipping spiral galaxies is also important for making an unbiased training data set. When CNNs are trained by a sample biased to S-spirals, the trained model would naturally give more S-spirals than Z-spirals. We furthermore rotate the images by 90 deg, 180 deg, and 270 deg, and rescale the brightness of the images with GAMMA = 2.0, 2.1, 2.3, and 2.4 parameters (Fig. 3). The data augmentation increases the number of spiral galaxies by 40 times and makes the numbers of three classes similar. The training data set therefore contains images of 45 200 S-spirals, 45 200 Z-spirals, and 41 320 non-spirals. We eventually make five different training data sets by selecting a different subset for validation.

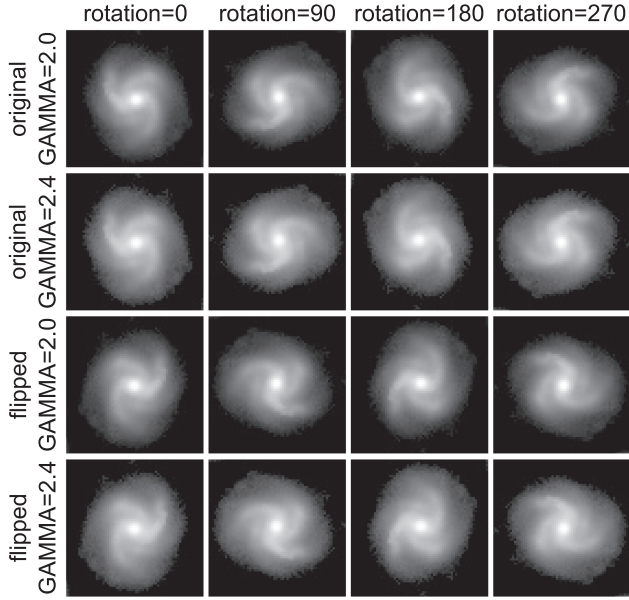


Figure 3. Examples of data augmentation. A FITS image of one galaxy is converted into multiple JPEG images with a different grey-scale value, which is adjusted by the GAMMA parameter. The original images of S-spirals are horizontally flipped and are treated as Z-spiral images. Furthermore, these images are rotated by 90, 180, and 270 deg.

Table 1. Structure of CNNs used in this paper.

	Layer	Output shape
1	Input	(64 pix, 64 pix, 1 map)
2	Convolution (5 pix × 5 pix)	(60 pix, 60 pix, 32 maps)
3	Convolution (5 pix × 5 pix)	(56 pix, 56 pix, 64 maps)
4	MaxPooling (4 pix × 4 pix)	(14 pix, 14 pix, 64 maps)
5	Convolution (3 pix × 3 pix)	(12 pix, 12 pix, 128 maps)
6	Convolution (3 pix × 3 pix)	(10 pix, 10 pix, 128 maps)
7	MaxPooling (2 pix × 2 pix)	(5 pix, 5 pix, 128 maps)
8	Dropout	(5 pix, 5 pix, 128 maps)
9	Flatten	(3200 features)
10	Dense	(3200 features)
11	Dropout	(3200 features)
12	Dense	(3200 features)
13	Dropout	(3200 features)
14	Dense	(3 classes)

3 CONVOLUTIONAL NEURAL NETWORKS

We make CNN models to classify galaxy morphology into non-spirals, S-spirals, and Z-spirals in a similar way to previous works (e.g. Dieleman et al. 2015). Table 1 summarizes the configuration of the CNN used in this paper. The size of input images is 64 pixel × 64 pixel. There are four convolutional layers with kernel sizes of 5 pixel × 5 pixel, 5 × 5, 3 × 3, and 3 × 3, respectively. The number of convolutional filters is 32, 64, 128, and 128, respectively. Each filter generates a feature map. We add two pooling layers, which take the maximum value in 4 pixel × 4 pixel and 2 × 2. The maximum pooling efficiently extracts important features like edges as well as reduces the amount of information by resampling. After convolutional layers, 128 feature maps with 5 pixel × 5 pixel are flattened and fed into a fully connected layer with 3200 features. These features are combined in dense layers. We also include three dropout layers to avoid overfitting of the CNNs (Srivastava et al.

2014). In these layers, 20 per cent, 50 per cent, and 50 per cent of input units are randomly set to zero. The final layer uses the Softmax function, which is computed as

$$p_{i,c} = e^{s_{i,c}} / \sum_{c=1}^3 e^{s_{i,c}}, \quad (1)$$

where $s_{i,c}$ is the output score for the c -th category for morphology classification (non-spiral, S-spiral, and Z-spiral) of the i -th image and $p_{i,c}$ corresponds to the predicted probabilities of each class. We eventually adopt the class with the highest probability to determine the morphology.

We train the CNNs using the Keras library (Chollet et al. 2015) with a single GPU, NVIDIA GeForce GTX 1080 Ti. A total of 20 769 539 trainable parameters of the model are determined by minimizing a loss function, which expresses inconsistency between actual classes and predicted probabilities. We adopt a cross-entropy (CE) loss function defined as,

$$\text{CE loss} = - \sum_{i=1}^{256} \sum_{c=1}^3 t_{i,c} \log p_{i,c}, \quad (2)$$

where $t_{i,c}$ is the ground truth label (1 if true and 0 if false). We use Adam algorithm (Kingma & Ba 2014), which optimizes the parameters based on the gradient descent of the loss function with a subsample of 256 images randomly selected from the training data set. This method is called the mini-batch stochastic gradient descent. One epoch ends with 33 s when the entire training data set has been used once for the calculation of the loss function. We repeat this 60 times and derive the accuracy, which is simply a ratio of correctly predicted images to all the images, and the CE loss in each epoch. The results of one CNN model are shown in Fig. 4. The training accuracy continues to increase to almost 100 per cent while the validation accuracy saturates at about 20 epochs. The loss function takes the minimum at 32 epochs and turns to increase at the later epochs. As this is clearly overfitting the training data set, we adopt the CNN models where the validation loss is minimized. Other four CNN models reach the minimum at 29, 32, 41, and 28 epochs. The validation data set is not directly used for the training of CNNs but indirectly affects the choice of the best model. We therefore compute the average accuracy of five CNN models from the cross-validation (see Section 2.2).

The average accuracy and standard deviation is 97.48 ± 0.14 per cent. For galaxy images with the predicted probability of >0.95 , the accuracy is increased to 99.37 ± 0.10 per cent. Note that the accuracy is different among the morphology classes. We show the confusion matrix, which is the fraction of correct or incorrect predictions in each predicted class, in Table 2. Most of the failures are for the case that non-spirals are misclassified as either S-spirals or Z-spirals and vice versa. The fraction that S-spirals (Z-spirals) are misclassified as Z-spirals (S-spirals) is only 0.2 per cent.

For about 2.5 per cent of the validation data set, the predicted class is different from the labelled one. Fig. 5 shows examples of misclassification. Some objects have the second highest probability of 0.1–0.5 in the labelled class while others are misclassified with the high probability of >0.95 . Non-spirals with different predictions seem to have some substructures, suggesting that they potentially have spiral arms with low contrast. Edge-on galaxies seem to be often misclassified, compared to face-on ones. The accuracy in the validation data set depends on the major-to-minor axis, which can be interpreted as an inclination angle (Fig. 6). For galaxies with an axial ratio of <0.2 , the accuracy decreases to 96 per cent. This may be due

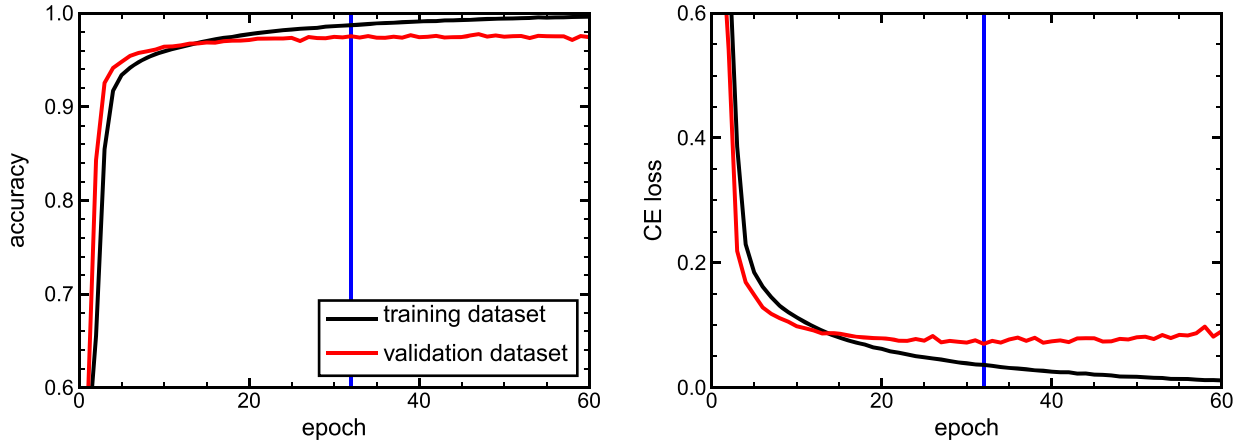


Figure 4. The accuracy and the CE loss function for the training (black lines) and validation data set (red lines) for one of the six cross-validation tests. The blue horizontal lines indicate the epoch when the validation loss is minimized.

Table 2. The fractions of images with the predicted class in each labelled class from the cross-validation.

Predicted class	Labelled class		
	Non-spiral	S-spiral	Z-spiral
Non-spiral	$96.31 \pm 0.47\%$	$1.86 \pm 0.20\%$	$1.83 \pm 0.34\%$
S-spiral	$1.71 \pm 0.17\%$	$98.12 \pm 0.24\%$	$0.16 \pm 0.10\%$
Z-spiral	$1.90 \pm 0.22\%$	$0.19 \pm 0.13\%$	$97.91 \pm 0.34\%$

to a lack of edge-on objects in the training data set as it becomes more difficult to identify spiral arms by visual inspection.

We evaluate the accuracy as a function of i -band magnitude, photometric redshift, and FWHM size (Fig. 6). There is also a weak trend that brighter objects are more correctly classified. On the other hand, the accuracy tends to be constant across a redshift range to $z \sim 0.5$. The FWHM size is estimated from Gaussian-weighted second-order moment in the i -band images (Bernstein & Jarvis 2002). The moment is stored as `i_sdssshape_shape11, 22, 12` in the PDR2 data base. We compute the determinant radius as $r_{\text{det}} = (\text{shape11} \times \text{shape22} - \text{shape12}^2)^{0.25}$. Under the assumption

of Gaussian, we convert the radius to FWHM by applying $2\sqrt{2 \ln 2}$. The high accuracy in compact galaxies with $\text{FWHM} \sim 1$ arcsec is due to the fact that most galaxies are classified as non-spirals in the HSC images.

We also look at how the sample size affects the performance of CNNs. We train CNNs by using images of randomly selected 100, 200, 400, and 800 galaxies in each class and measure the accuracy of validation data. When only original images are used for training, the accuracy gradually increases from 52 per cent at 100 to 90 per cent at 800. The data augmentation including horizontal flipping, rotation, and rescaling significantly improves the accuracy from 52 per cent to 89 per cent with the same training data set of 100 images. It requires at least 100 images to reach an accuracy of more than 90 per cent with the data augmentation.

4 A CATALOGUE OF SPIRAL GALAXIES

Now, we apply the trained CNN models to a large data set in other HSC-SSP fields, where 561 251 galaxy images are available over an area of $\sim 320 \text{ deg}^2$. We use five CNN models made from the

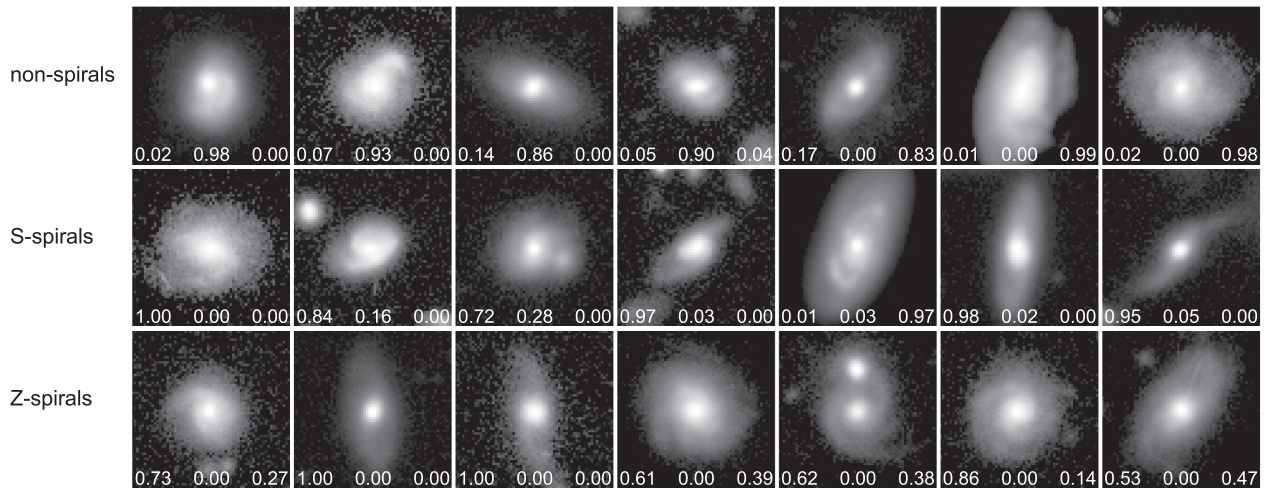


Figure 5. Examples of HSC images of misclassification in each class. From left to right in the bottom of each images, we show the predicted probabilities of non-spiral, S-spiral, and Z-spiral.

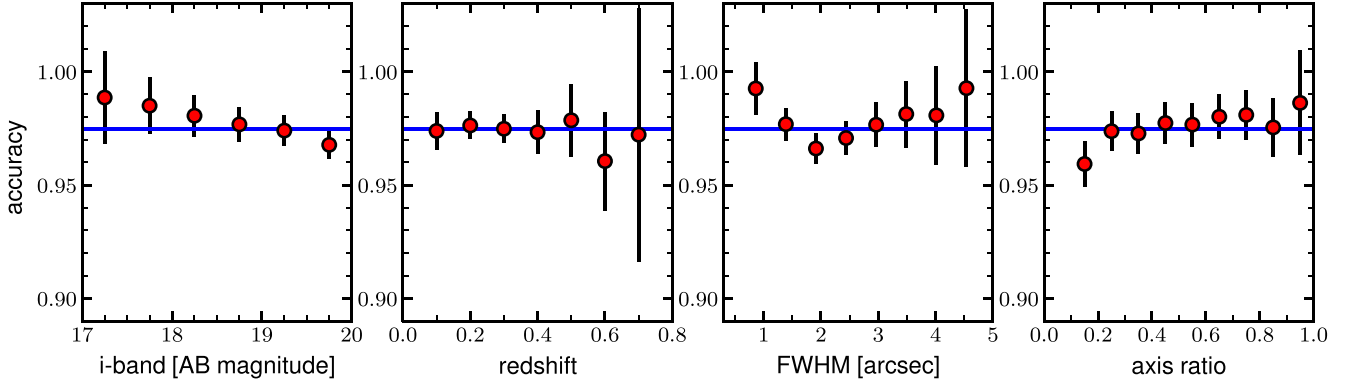


Figure 6. The accuracy of the CNN models for the validation data set as a function of *i*-band magnitude, photometric redshift, FWHM size, and major-to-minor axial ratio. A blue line shows the overall accuracy, 0.9748.

Table 3. A spin parity catalogue of spiral galaxies. Object ID is the same in the PDR-2 of the HSC–SSP for the wide layer (Aihara et al. 2019). p_0 , p_1 , and p_2 indicate the predicted probabilities of non-spirals, S-spirals, and Z-spirals, respectively. The full table is available online.

Object ID	Class flag ^a	p_0	p_1	p_2
40959011452899104	2	0.091	0.000	0.909
40959011452899880	2	0.093	0.000	0.907
40959011452901552	1	0.272	0.728	0.000
40959015747870352	2	0.009	0.000	0.991
40959015747871064	1	0.011	0.989	0.000
40959020042835800	2	0.006	0.000	0.994
40959020042837000	1	0.000	1.000	0.000

^aFlag: 1 = S-spiral; 2 = Z-spiral.

cross-validation to derive average predicted probabilities of each class. We identify 37 917 S-spirals and 38 718 Z-spirals and provide the catalogue including the predicted probabilities in Table 3. The remaining 484 616 galaxies are non-spirals. The difference between the numbers of S-spirals and Z-spirals is $N_S - N_Z = -801$. The significance level of is 2.9σ when only Poisson statistics is applied to estimate the uncertainties in the number of galaxy images (Gehrels 1986). However, the uncertainties on these numbers are likely to be dominated by misclassification of CNN-based classification, rather than Poisson errors. S-spirals and Z-spirals are in principle affected to the same degree by the contamination. In the validation data set, the fraction that S-spirals are misclassified as non-spirals is similar to the fraction that Z-spirals are misclassified as non-spirals while there are some variations (0.20 per cent, 0.34 per cent) between the CNN models (Table 2). This is non-negligible because the vast majority of a half million galaxies is non-spirals. Considering the uncertainties in the misclassification, the error of the difference would be $\Delta(N_S - N_Z) = 1932$, which is larger than the actual measurement. We also use five individual trained CNNs for classification of 561 251 images to calculate the average and the standard deviation of the numbers of spirals, $N_S = 38625 \pm 1138$ and $N_Z = 39537 \pm 1479$, corresponding to the error of $\Delta(N_S - N_Z) = 1866$. A stable performance of 0.04 per cent in misclassification is required so that the uncertainty is dominated by Poisson errors in the HSC imaging data. We would need to train the model with more validation data.

We do not find a significant difference of the numbers between S-spirals and Z-spirals. On the other hand, there is a significant excess

of S-spirals over Z-spirals in our training data set and in the Galaxy Zoo Catalogue (Land et al. 2008; Lintott et al. 2008). This is likely to be caused by a human selection bias (Hayes et al. 2017). Visual inspection by human eyes may unconsciously select more S-spirals. We also calculate the significance level of the number difference between S-spirals and Z-spirals, $(N_S - N_Z)/\Delta(N_S - N_Z)$, in bins of *i*-band magnitude, photometric redshift, FWHM size, and major-to-minor axial ratio (Fig. 7). We take into account the incompleteness of CNN-based classification (Table 2) as well as the Poisson errors. We do not find a significant excess of S-spirals or Z-spirals with $|(N_S - N_Z)| > 3\Delta(N_S - N_Z)$.

The fraction of both spirals including S-spirals and Z-spirals to all the galaxies is 13.7 per cent though it depends on galaxy properties. Note that we can identify only galaxies with visible spiral arms, depending on the sensitivity and spatial resolution of images used in the classification. In the Galaxy Zoo project, the fraction of galaxies with features such as spiral arms is 10 per cent at $z = 0.1$ and ~ 0 per cent at $z = 0.2$ (Willett et al. 2013). In the deeper HSC images, spiral arms are visible in ~ 20 per cent of galaxies at $z = 0.1$ – 0.2 (Fig. 7). The measured fraction of spiral galaxies should be still a lower limit because more galaxies with fainter or lower contrast spiral arms can be identified in even deeper and higher resolution images. We actually find that ~ 50 per cent of extended galaxies with $\text{FWHM} \sim 3$ arcsec are classified as spirals since the HSC resolution is high enough to identify their spiral arms.

Fig. 8 presents examples of S-spirals and Z-spirals with a spectroscopic redshift of $z_{\text{spec}} = 0.2$ – 0.3 . 48 576 of 76 635 spiral galaxies are located at $z_{\text{phot}} > 0.2$, where we are not able to identify spiral arms with the SDSS images. The fraction of spiral galaxies decreases from 20 per cent at $z_{\text{phot}} = 0.2$ to 10 per cent at $z_{\text{phot}} = 0.6$. The redshift dependence is strongly affected by the cosmological dimming of the surface brightness, which decreases as $(1+z)^{-4}$. It becomes difficult to detect an extended substructure such as spiral arms in high-redshift galaxies. The magnitude dependence of the spiral fraction is coupled with the redshift dependence since faint sources tend to be at higher redshift. The decrease of spirals with a small axial ratio of < 0.2 is likely to be caused by human bias since it becomes hard to visually identify spiral arms in edge-on galaxies.

2524 galaxies are identified to have spiral arms at $z_{\text{phot}} = 0.5$ – 0.7 in spite of the strong effect of the cosmological dimming. 1455 of them have a stellar mass of $\log(M_*/M_\odot) > 10.8$, which is similar to that of Andromeda (M31: Gehean et al. 2006). The existence of spiral arms indicates that the galaxies are still forming stars. The majority of M31-mass galaxies have an early-type morphology without spiral

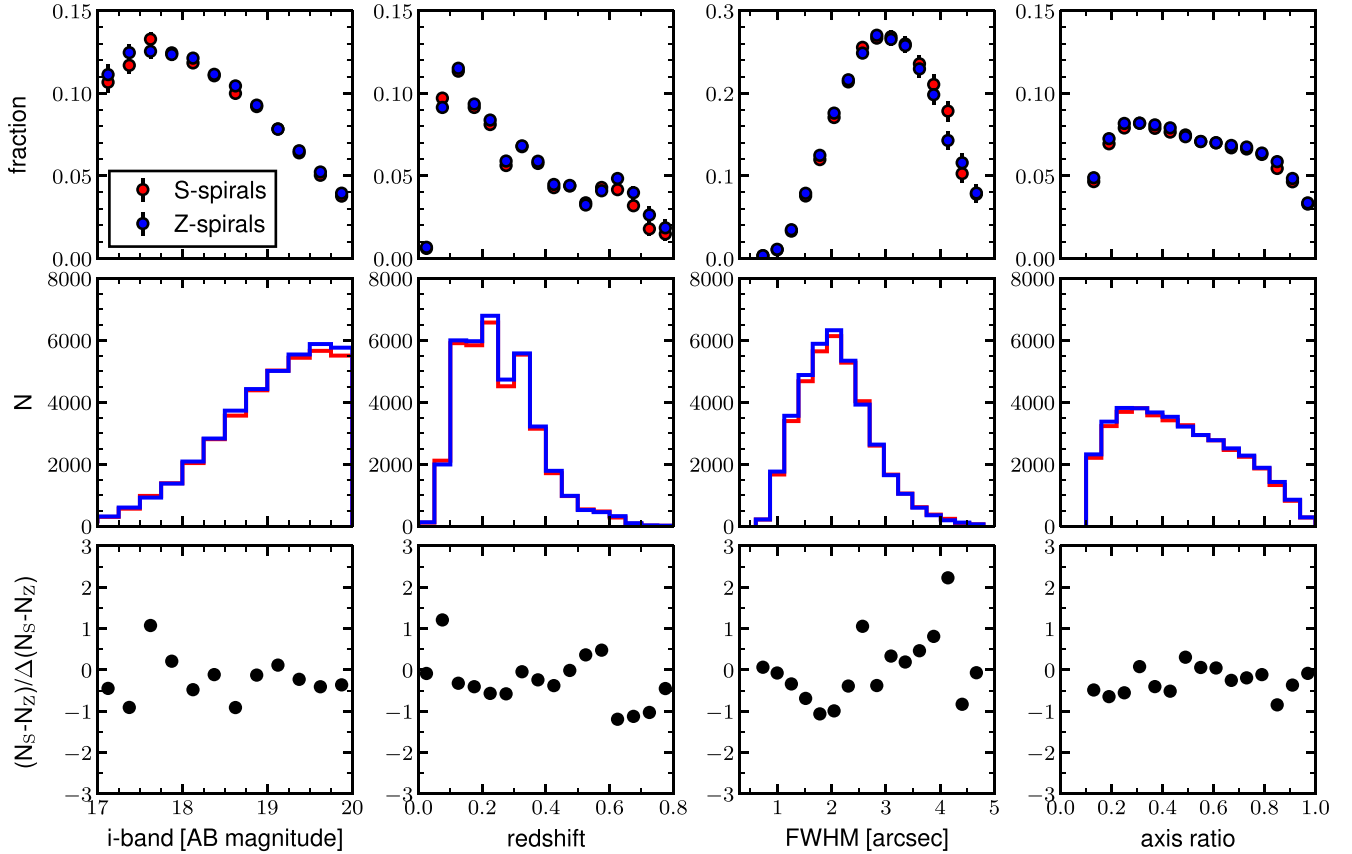


Figure 7. The fraction of S-spirals (red circles) and Z-spirals (blue circles) to all the galaxies as a function of *i*-band magnitude, photometric redshift, FWHM size, and major-to-minor axial ratio (top four panels). The middle four panels and the bottom four panels show the number of spiral galaxies and the significance level of the number difference between S-spirals and Z-spirals, respectively. The error, $\Delta(N_S - N_Z)$, takes into account both the Poisson error and the incompleteness of CNN-based classification.

arms at $0 < z < 0.7$ and quench the star formation (Papovich et al. 2015). The identified massive spiral galaxies are likely to be the progenitors of M31.

5 SUMMARY

We have developed a CNN model to classify galaxy morphology into three categories (S-spiral, Z-spiral, and non-spiral) by using images taken by the Subaru HSC survey. The superb image quality allows us to identify spiral arms in faint galaxies with $i \sim 20$ by visual inspection. We have used a total of 0.2 million images after data augmentation such as flipping, rescaling, and rotation for training the model. The trained model successfully classifies the test data set, which is not used for training and validation, and results in an accuracy of 97.5 ± 0.1 per cent. The accuracy decreases to ~ 90 per cent in the case that the training data set consists of less than 100 images in each class. This would become more of a problem when one finds rare objects.

We have applied the trained CNN model to 561 251 galaxy images over an area of $\sim 320 \text{ deg}^2$. Our automated classification efficiently picks up spiral galaxies and determines their winding direction of spiral arms, providing 37 917 S-spirals and 38 718 Z-spirals. We do not find a significant excess of S-spirals over Z-spirals, which is seen in the training data set and the Galaxy Zoo Catalogue. We have also identified 1455 massive spiral galaxies with $\log(M_*/M_\odot) > 10.8$ at $z_{\text{phot}} = 0.5\text{--}0.7$, which are likely the progenitors of M31.

There are some limitations to our CNN-based classification. The criterion of spiral arms is defined by the training data set, which is selected by our visual inspection. Although we have used the sample of galaxies whose spiral arms are clearly seen for the training to minimize the contamination of non-spirals, the criterion of clear spirals is still somewhat ambiguous. The ambiguous definition is probably one of the reasons that 2.5 per cent of the validation data set is misclassified. It would be important to make a clean training sample. Creating mock images from numerical simulations is one of the several efficient methods to prepare a large data set for training models (e.g. Huertas-Company et al. 2018; Metcalf et al. 2019). Another direction to define a repeatable class of morphology is an unsupervised learning approach, which does not require visually classified training data sets (e.g. Hocking et al. 2018; Martin et al. 2019). Nevertheless, our attempt already demonstrates that CNN is powerful for making a large sample of galaxies with particular substructures such as spiral arms from a large data set and efficiently picking up rare objects such as massive spiral galaxies.

ACKNOWLEDGEMENTS

We are very grateful to the referee for constructive suggestions to improve the paper. The Hyper Suprime-Cam (HSC) collaboration includes the astronomical communities of Japan and Taiwan, and Princeton University. The HSC instrumentation and software were developed by the National Astronomical Observatory of Japan

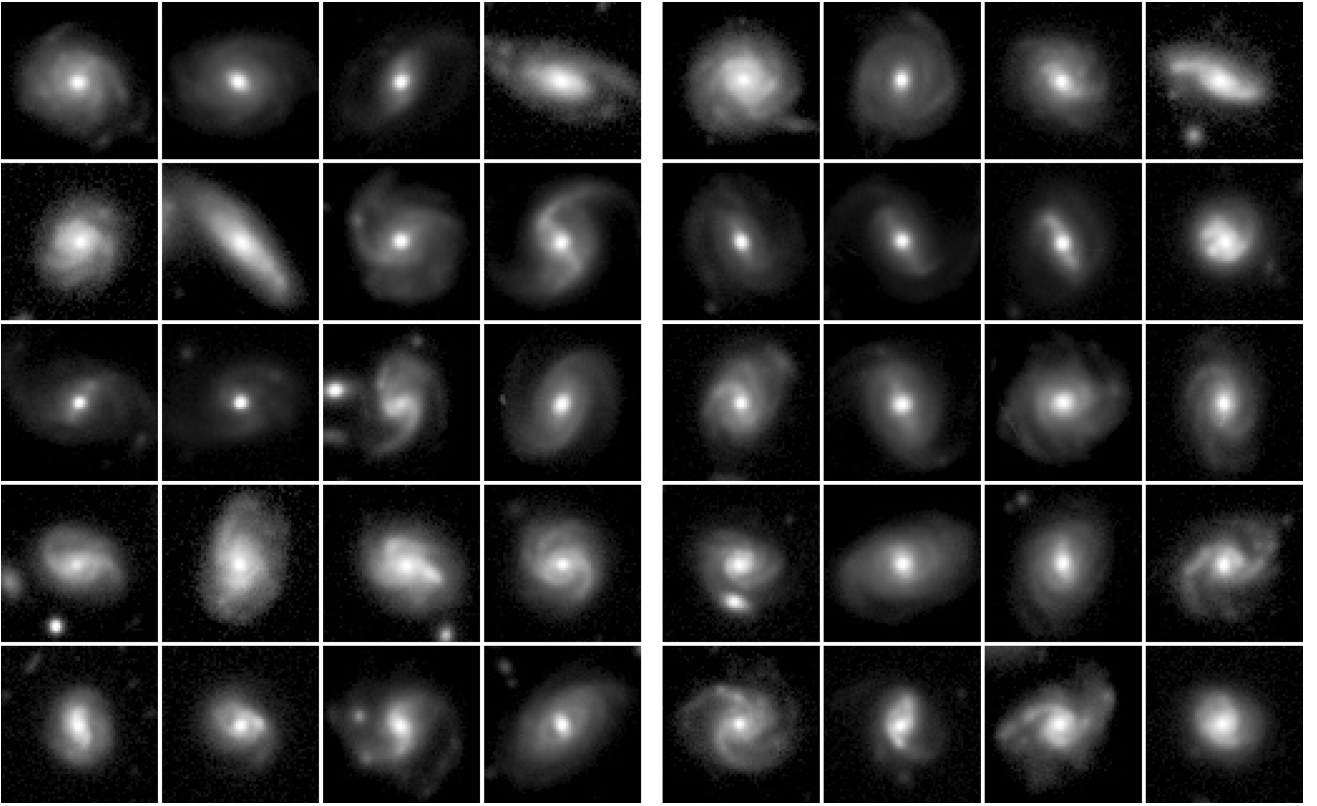


Figure 8. Examples of HSC images of S-spirals (left) and Z-spirals (right) with the predicted probability of >0.95 . They are randomly selected from spectroscopically confirmed galaxies at $z_{\text{spec}} = 0.2\text{--}0.3$.

(NAOJ), the Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU), the University of Tokyo, the High Energy Accelerator Research Organization (KEK), the Academia Sinica Institute for Astronomy and Astrophysics in Taiwan (ASIAA), and Princeton University. Funding was contributed by the FIRST program from the Japanese Cabinet Office, the Ministry of Education, Culture, Sports, Science and Technology (MEXT), the Japan Society for the Promotion of Science (JSPS), Japan Science and Technology Agency (JST), the Toray Science Foundation, NAOJ, Kavli IPMU, KEK, ASIAA, and Princeton University.

This paper makes use of software developed for the Large Synoptic Survey Telescope. We thank the LSST Project for making their code available as free software at <http://dm.lsst.org>

This paper is based on data collected at the Subaru Telescope and retrieved from the HSC data archive system, which is operated by Subaru Telescope and Astronomy Data Center (ADC) at NAOJ. Data analysis was in part carried out with the cooperation of Center for Computational Astrophysics (CfCA), NAOJ.

The Pan-STARRS1 Surveys (PS1) and the PS1 public science archive have been made possible through contributions by the Institute for Astronomy, the University of Hawaii, the Pan-STARRS Project Office, the Max Planck Society and its participating institutes, the Max Planck Institute for Astronomy, Heidelberg, the Max Planck Institute for Extraterrestrial Physics, Garching, The Johns Hopkins University, Durham University, the University of Edinburgh, the Queen's University Belfast, the Harvard-Smithsonian Center for Astrophysics, the Las Cumbres Observatory Global Telescope Network Incorporated, the National Central University of Taiwan, the Space Telescope Science Institute, the National Aeronautics and Space Administration under grant no. NNX08AR22G issued through the

Planetary Science Division of the NASA Science Mission Directorate, the National Science Foundation grant no. AST-1238877, the University of Maryland, Eotvos Lorand University (ELTE), the Los Alamos National Laboratory, and the Gordon and Betty Moore Foundation.

CER acknowledges Anupreeta More for providing a tool for visual inspection of images.

Funding for the Sloan Digital Sky Survey (SDSS) has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Aeronautics and Space Administration, the National Science Foundation, the U.S. Department of Energy, the Japanese Monbukagakusho, and the Max Planck Society. The SDSS Web site is <http://www.sdss.org/>. The SDSS is managed by the Astrophysical Research Consortium (ARC) for the Participating Institutions. The Participating Institutions are The University of Chicago, Fermilab, the Institute for Advanced Study, the Japan Participation Group, The Johns Hopkins University, Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, University of Pittsburgh, Princeton University, the United States Naval Observatory, and the University of Washington.

DATA AVAILABILITY

The data underlying this article are available in the Hyper Suprime-Cam Subaru Strategic Program Data Release Site at <https://hsc-release.mtk.nao.ac.jp/>.

REFERENCES

- Abraham S., Aniyan A. K., Kembhavi A. K., Philip N. S., Vaghmare K., 2018, *MNRAS*, 477, 894
- Aihara H. et al., 2018, *PASJ*, 70, S4
- Aihara H. et al., 2019, *PASJ*, 71, 114
- Alam S. et al., 2015, *ApJS*, 219, 12
- Baba J., Saitoh T. R., Wada K., 2013, *ApJ*, 763, 46
- Bernstein G. M., Jarvis M., 2002, *AJ*, 123, 583
- Bertin E., 2012, in Ballester P., Egret D., Lorente N. P. F., eds, ASP Conf. Ser. Vol. 461, Astronomical Data Analysis Software and Systems XXI. Astron. Soc. Pac., San Francisco, p. 263
- Chollet F. et al., 2015, Keras, Available at: <https://keras.io>
- Conselice C. J., 2014, *ARA&A*, 52, 291
- Cool R. J. et al., 2013, *ApJ*, 767, 118
- Davis D. R., Hayes W. B., 2014, *ApJ*, 790, 87
- Dieleman S., Willett K. W., Dambre J., 2015, *MNRAS*, 450, 1441
- Dobbs C., Baba J., 2014, *PASA*, 31, e035
- Domínguez Sánchez H., Huertas-Company M., Bernardi M., Tuccillo D., Fischer J. L., 2018, *MNRAS*, 476, 3661
- Doroshkevich A. G., 1970, *Astrophysics*, 6, 320
- Drinkwater M. J. et al., 2010, *MNRAS*, 401, 1429
- Fukushima K., 1980, *Biol. Cybern.*, 36, 193
- Garilli B. et al., 2014, *A&A*, 562, A23
- Geehan J. J., Fardal M. A., Babul A., Guhathakurta P., 2006, *MNRAS*, 366, 996
- Gehrels N., 1986, *ApJ*, 303, 336
- Goldreich P., Lynden-Bell D., 1965, *MNRAS*, 130, 125
- Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y., 2014, in Ghahramani Z., Welling M., Cortes C., Lawrence N. D., Weinberger K. Q., eds, Advances in Neural Information Processing Systems 27, p. 2672
- Hart R. E. et al., 2017, *MNRAS*, 472, 2263
- Hayes W. B., Davis D., Silva P., 2017, *MNRAS*, 466, 3928
- He K., Zhang X., Ren S., Sun J., 2016, in 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016. Las Vegas, NV, USA, p. 770
- Hocking A., Geach J. E., Sun Y., Davey N., 2018, *MNRAS*, 473, 1108
- Hsieh B. C., Yee H. K. C., 2014, *ApJ*, 792, 102
- Huertas-Company M. et al., 2015, *ApJS*, 221, 8
- Huertas-Company M. et al., 2018, *ApJ*, 858, 114
- Iye M., Tadaki K.-i., Fukumoto H., 2019, *ApJ*, 886, 133
- Kingma D. P., Ba J., 2015, the 3rd International Conference for Learning Representations, preprint ([arXiv:1412.6980](https://arxiv.org/abs/1412.6980))
- Krizhevsky A., Sutskever I., Hinton G. E., 2012, in Pereira F., Burges C. J. C., Bottou L., Weinberger K. Q., eds, Advances in Neural Information Processing Systems 25., p. 1097
- Kuminski E., Shamir L., 2016, *ApJS*, 223, 20
- Land K. et al., 2008, *MNRAS*, 388, 1686
- LeCun Y., Bottou L., Bengio Y., Haffner P., 1998, *Proc. IEEE*, 86, 2278
- Lin C. C., Shu F. H., 1964, *ApJ*, 140, 646
- Lintott C. J. et al., 2008, *MNRAS*, 389, 1179
- Lintott C. et al., 2011, *MNRAS*, 410, 166
- Liske J. et al., 2015, *MNRAS*, 452, 2087
- Mandelbaum R. et al., 2018, *PASJ*, 70, S25
- Martin G., Kaviraj S., Hocking A., Read S. C., Geach J. E., 2019, *MNRAS*, 491, 1408
- Masters K. L. et al., 2010, *MNRAS*, 405, 783
- Metcalf R. B. et al., 2019, *A&A*, 625, A119
- Nishizawa A. J., Hsieh B.-C., Tanaka M., Takata T., 2020, preprint ([arXiv:2003.01511](https://arxiv.org/abs/2003.01511))
- Papovich C. et al., 2015, *ApJ*, 803, 26
- Peebles P. J. E., 1969, *ApJ*, 155, 393
- Peng C. Y., Ho L. C., Impey C. D., Rix H.-W., 2010, *AJ*, 139, 2097
- Pierre M. et al., 2004, *J. Cosmology Astropart. Phys.*, 2004, 011
- Russakovsky O. et al., 2015, *Int. J. Comput. Vis. (IJCV)*, 115, 211
- Schawinski K., Zhang C., Zhang H., Fowler L., Santhanam G. K., 2017, *MNRAS*, 467, L110
- Shamir L., 2017, *PASA*, 34, e044
- Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdinov R., 2014, *J. Mach. Learn. Res.*, 15, 1929
- Strateva I. et al., 2001, *AJ*, 122, 1861
- Sugai H., Iye M., 1995, *MNRAS*, 276, 327
- Tanaka M. et al., 2018, *PASJ*, 70, S9
- White S. D. M., 1984, *ApJ*, 286, 38
- Willett K. W. et al., 2013, *MNRAS*, 435, 2835
- York D. G. et al., 2000, *AJ*, 120, 1579

SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

Tadaki_2020_table3.cat

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

APPENDIX A: SOME EXTRA MATERIAL

In Figs A1 and A2, we show example images of S-spirals and Z-spirals, which are randomly selected from the training data set. We use them for training a CNN model after data augmentation such as flipping, rescaling, and rotation of images.

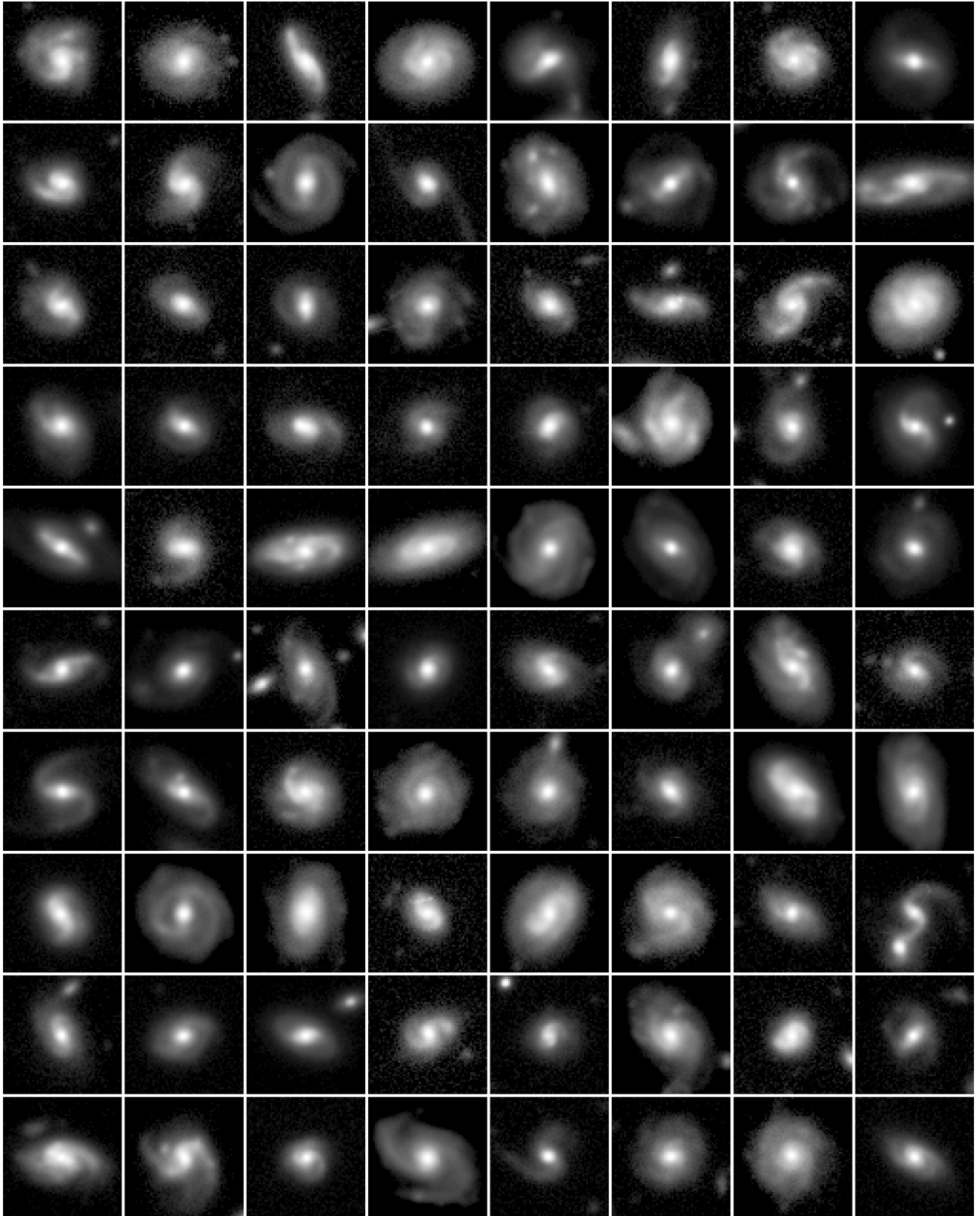


Figure A1. Example images of S-spirals.

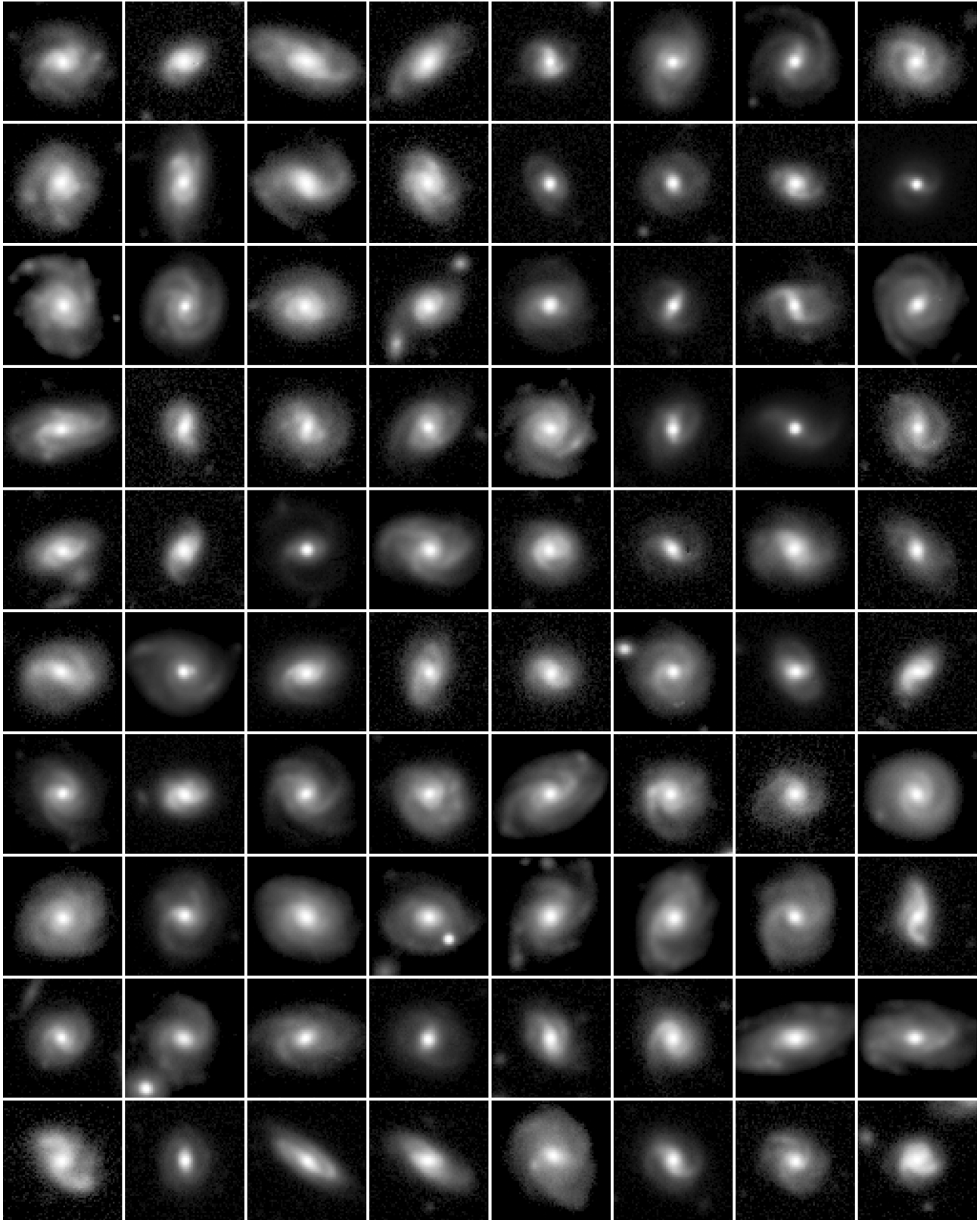


Figure A2. Example images of Z-spirals.

This paper has been typeset from a \LaTeX file prepared by the author.