Predicting bulge to total luminosity ratio of galaxies using deep learning

Harsh Grover\textsuperscript{1}\#, Omkar Bait\textsuperscript{2}, Yogesh Wadadekar\textsuperscript{2} and Preetish K. Mishra\textsuperscript{2,3}

\textsuperscript{1}Birla Institute of Technology and Science, Pilani, Rajasthan 333031, India
\textsuperscript{2}National Centre for Radio Astrophysics, TIFR, Post Bag 3, Ganeshkhind, Pune 411007, India
\textsuperscript{3}Inter-University Centre for Astronomy and Astrophysics, Post Bag 4, Ganeshkhind, Pune 411007, India

Accepted XXX. Received YYY; in original form ZZZ.

ABSTRACT

We present a deep learning model to predict the $r$-band bulge-to-total light ratio ($B/T$) of nearby galaxies using their multi-band JPEG images alone. Our Convolutional Neural Network (CNN) based regression model is trained on a large sample of galaxies with reliable decomposition into the bulge and disk components. The existing approaches to estimate the $B/T$ use galaxy light-profile modelling to find the best fit. This method is computationally expensive, prohibitively so for large samples of galaxies, and requires a significant amount of human intervention. Machine learning models have the potential to overcome these shortcomings. In our CNN model, for a test set of 20000 galaxies, 85.7 per cent of the predicted $B/T$ values have absolute error (AE) less than 0.1. We see further improvement to 87.5 per cent if, while testing, we only consider brighter galaxies (with $r$-band apparent magnitude < 17) with no bright neighbours. Our model estimates $B/T$ for the 20000 test galaxies in less than a minute. This is a significant improvement in inference time from the conventional fitting pipelines, which manage around 2-3 estimates per minute. Thus, the proposed machine learning approach could potentially save a tremendous amount of time, effort and computational resources while predicting $B/T$ reliably, particularly in the era of next-generation sky surveys such as the Legacy Survey of Space and Time (LSST) and the Euclid sky survey which will produce extremely large samples of galaxies.

Key words: galaxies: structure – methods: data analysis – techniques: image processing

1 INTRODUCTION

Artificial Intelligence (AI) is the domain of computer science that deals with computers performing tasks which usually need human-level intelligence. Machine Learning (ML) is the branch of AI in which programs learn from experience to progressively get better at a task (Mitchell 1997). A ML model develops its own implicit set of rules for a certain task using a learning algorithm with experience. When learning from labelled datasets, a branch of ML called supervised learning is used. The aim of supervised learning is to learn a latent function that maps the inputs to the ground-truth labels in a given dataset. Depending on the distribution of ground-truth labels, supervised learning is broadly divided into classification and regression problems. If labels belong to a discrete space (or are categorical) the problem is defined as a classification problem. Whereas, if they belong to a continuous space, the problem is defined as a regression problem.

Over the years, many studies have utilised the competence of ML for classification, parameter estimation and outlier detection problems in data-driven astronomy (see Baron 2019, for a detailed overview). One of the popular supervised learning algorithm is Support Vector Machines (SVM) which has been applied successfully to extract useful information from the data coming from large sky surveys. Particularly in the field of extragalactic astronomy, ML and SVM based techniques have been used for star-galaxy classification (Philip et al. 2002), photometric redshifts (Wadadekar 2005), galaxy morphology classifications (Huertas-Company, M. et al. 2011) etc. In recent years, a new ML technique – Deep Learning – has emerged as the best performing supervised learning technique across many applications in different disciplines (LeCun et al. 2015).

Deep learning (DL) is the paradigm of machine learning which uses multi-layer neural networks. Neural networks are ML models inspired from the network of brain cells or neurons. The differentiating factor between DL and conventional ML is in the process of feature selection. In conventional ML, performance strongly depends on the features used. More often than not, new features are created for a task to better capture correlations in the data. This is called feature engineering. ML models perform poorly if not given good features to learn from. Deep learning eliminates this dependence by defining its own features to learn from, which are relevant to the
task at hand. This makes deep learning a versatile, high performing solution for a variety of supervised ML tasks. The Artificial Neural Networks (ANNs) have found application in not only the traditional tasks in extragalactic astronomy such as photometric redshift estimation (e.g., Firth et al. 2003; Collister & Lahav 2004; Tagliaferri et al. 2003; Vanzella et al. 2004; Sadler et al. 2016; Bilić et al. 2018; Pasquet et al. 2019), but also in more specialised problems such as predicting infrared luminosity of galaxies (Ellison et al. 2016), estimation of star formation properties (Surana et al. 2020) and ranking the quehuman parameters of galaxies (Teimoorinia et al. 2016). In particular, Convolutional Neural Networks (CNN) are now extremely popular in studies using imaging data on galaxies. The CNNs are deep learning models designed to extract features from images. They have provided state-of-the-art performance in majority of computer vision tasks in recent times (Krizhevsky et al. 2012; He et al. 2017). CNNs have been utilised for galaxy morphological classification e.g., classifying the optical morphologies broadly into spheroidal, disk and irregular types (Huertas-Company et al. 2015) and even for classification of various radio galaxy morphologies (Wu et al. 2019). Recently, Ribeil et al. (2019) have shown that CNNs can be used to predict galaxy shapes, needed for weak lensing studies, from a wide field but shallow sky survey images using the ‘ground truth’ images from a deeper but narrower field survey. CNNs have also found their application in the automated detection of features in sky-survey images such as galactic bars (Abraham et al. 2018) and strong gravitational lenses (Li et al. 2020; Canameras et al. 2020). Further, ML and CNNs in particular have also been used to detect outliers in large area Sloan Digital Sky Survey (SDSS) data (e.g., Baron & Poznanski 2017; Sharma et al. 2019). In this study, we want to determine the bulge to total luminosity ratio \( B/T \) of a galaxy using its optical multi-band images as input. Due to the use of a dataset with galaxy images labelled by \( B/T \in [0, 1] \) in a continuous space \( \mathbb{R} \), we have performed CNN based regression in this work.

The bulge to total light ratio \( B/T \) of a galaxy is the fraction of total flux coming from the bulge component of the galaxy. Bulges are the central component of disc galaxies which appear as central bright cores in galaxy images or as excess-light over the disc light in the inner region of galaxy light profiles. The \( B/T \) ratio is an important indicator of galaxy structure and is well correlated with several other physical quantities of interest in studies of galaxy evolution, such as galaxy morphology (Graham & Worley 2008), kinematics (Cappellari et al. 2013), stellar mass and star formation rate (Bluck et al. 2014). The \( B/T \) is also directly related to the bulge luminosity which in turn correlates with the mass of the central supermassive black hole of the galaxy (Kormendy & Richstone 1995; Marconi & Hunt 2003; Kormendy & Ho 2013). Galaxies having different \( B/T \) ratios are thought to have undergone evolution along different evolutionary paths. A high value of \( B/T \) of a galaxy generally indicates its evolutionary history dominated by galaxy mergers (Hopkins et al. 2010). On the other majority of low \( B/T \) systems are often pseudobulge hosting galaxies (Fisher & Drory 2008; Gadotti 2009) which are thought to undergo slow evolution through internal, secular evolution processes (Kormendy & Kennicutt 2004). It must be noted that unlike visual morphology which is a qualitative measure, \( B/T \) is a quantitative measure of the galaxy morphology. It is the only reliable way to separate ellipticals and disc galaxies when morphological features such as spiral arms cannot be resolved by the telescope (due to resolution or sensitivity limitations, particularly at high redshifts). The quantitative nature of the \( B/T \) ratio also allows direct comparison of galaxies with their counterparts in cosmological simulations. One can study these simulated galaxies to understand the origin of various properties of real galaxies. All these factors make the \( B/T \) an important parameter to measure in studies of galaxy formation and evolution.

The \( B/T \) ratio of galaxies has been traditionally measured by modelling their surface brightness profile. Such modelling allows one to find the light-profiles of individual photometric components of galaxies and then measure their properties such as luminosity, shape, size etc. The earliest works aimed at photometric decomposition of galaxies involved fitting for the azimuthally averaged 1-D surface brightness profile (Kent 1985). However, this method leads to systematic errors as creation of 1-D light profile by either averaging over the galaxy image or by taking a single cut along the major axis do not properly take into account non-axisymmetric features like a bar or isophotal twist (for more details see Gao & Ho (2017)). Due to these shortcomings, techniques to fit the whole 2D image of the galaxy, pixel-by-pixel, were developed (Wadadekar et al. 1999). There are now a number of implementations of 2D bulge disk decomposition available – GIM2D: Simard (2010), GALFIT: Peng et al. (2002a), IMFIT: Erwin (2015), PooFr: Robotham et al. (2016), BUDDA: de Souza et al. (2004), GASPP2D: Méndez-Abreu et al. (2008) – which fit the 2D images of galaxies. They usually employ different algorithms and methods to find the best fit model for a galaxy image. In the last two decades, some pipeline codes have been developed (PYMORPH: Vikram et al. (2010), GALAPAGOS: Barden et al. (2012)) to carry out bulge disc decomposition for large galaxy samples, in a mostly automated fashion. This 2D bulge-disc decomposition approach, while being powerful, is time consuming. The employed fitting algorithm and code carries out a search in parameter space defined by various model parameters and then finds the best fit. Most of the time one also needs to carry out additional steps (masking and providing PSF images) before the actual fitting of galaxies take place. Thus estimating \( B/T \) for a single galaxy involves a significant amount of time and computational cost. More importantly, the fitting procedure scales only linearly with the size of the galaxy sample. This limitation will be a major problem in the era of next generation sky surveys (e.g., LSST, Euclid sky surveys) which will observe several billion galaxies.

For an accurate and fast estimation of \( B/T \) for large galaxy samples, one can adopt a Machine Learning approach. One can develop a ML algorithm to learn the latent mapping from the galaxy images to the \( B/T \) ratio which will reduce the inference time by orders of magnitude. The only time consuming part in this approach is training the model, which is not an issue because it has to be done only once. Such a ML based approach has been found to be useful in estimating galaxy structural parameters in the past. For example, Tuccillo et al. (2017) have used CNN to fit the surface brightness profile of galaxies. This study demonstrates the computational efficiency and speed of deep learning over conventional fitting methods. However, their analysis is limited to single component (Sérsic) fit. In this study, we propose a CNN based regression model to estimate \( r \)-band \( B/T \) ratio of galaxies using their \( g,r,i \)-band images from the SDSS. We have used images of galaxies coming from the SDSS legacy survey DR 15 (Aguado et al. 2019) and their \( B/T \) values as listed in Meert et al. (2015) which is a large bulge-disc decomposition catalogue of SDSS galaxies.

The organisation of this paper is as follows. In Section 2 we describe the data used in our study. In Section 3, we describe the deep learning model and the training methodology used for the analysis. In Section 4, we discuss the results of our study with various metrics and error analysis. We end this paper by providing a summary and conclusion in Section 5.

MNRAS 000, 1–11 (2019)
2 DATA AND SAMPLE SELECTION

In order to train and test our CNN based model capable of predicting $B/T$ ratio of galaxies, we needed a statistically significant number of galaxies with available $B/T$ measurements. To select our sample for this purpose, we started with the data from Meert et al. (2015) (hereafter Meert catalogue) which is a catalogue of 2D photometric decompositions of about ~670,000 spectroscopically selected galaxies coming from Sloan Digital Sky Survey (SDSS) Data Release 7 (Abazajian et al. 2009). This is a flux limited sample of low-redshift galaxies lying within the redshift range $0.05 < z < 0.3$ and having $r$-band Petrosian magnitude between $14.0 < r < 17.77$. The fits in the Meert catalogue are performed on $r$-band SDSS images using the GALFIT code with the PYMORPH analysis pipeline (Vikram et al. 2010). Each galaxy in this catalogue is fitted by four different models of galaxy light profiles, two of which are single component de Vaucouleurs and Sérsic profiles and the other two are two-component de Vaucouleurs+Exponential and Sérsic+Exponential profiles. The galaxies having bad fits, or having erroneous magnitude and size measurements are indicated by setting bit 14 and 20 in the 'finalflag' column in the Meert catalogue. The galaxies having good fits in this catalogue are divided into three categories: those having good bulge fits only, the ones with good bulge+disc fits and those having good disc fits only. Since the aim of this work is to predict the $B/T$ ratio of galaxies, we chose to work with galaxies having good two component bulge+disc fits only. These galaxies have bits 10 AND 12 set in the 'finalflag' column of Meert catalogue. Additionally, we also removed galaxies having bulge Sérsic index $\geq 8$ as recommended by Meert et al. (2015) in order to exclude galaxies with unreliable measurements of subcomponent properties.

Using the above mentioned constraints, we obtained a sample of 103,757 galaxies having measurements of $r$-band $B/T$ ratio from the Meert catalogue. We treat these $r$-band $B/T$ ratio measurements as ground truth labels for our analysis. As mentioned previously, we aim to come up with a CNN based regression model to predict the $B/T$ of galaxies via their $g,r,i$-band images. In order to complete the training and testing requirement of our model, we downloaded a gri colour composite image in JPEG format using the SDSS finding chart tool 1 for every galaxy in our sample. The size of each image is $128 \times 128$ pixels. Scale of the image was kept at the natural scale of SDSS, which is $0.396$ arcsec pixel$^{-1}$. Each image is therefore approximately, $50.7 \times 50.7$ arcsec$^2$ in angular extent. All the subsequent analysis presented in this paper was done using these composite gri images of 103,757 galaxies labelled by their respective $r$-band $B/T$ values.

We tried both SDSS DR15 and DR7 images and found that our CNN performs better with DR15 images. Hence, all the analysis in this paper uses DR15 images. This may be due to the fact that the JPEG images in DR15 are made using the Lupton colour scheme which is a more correct representation of galaxy colour (Lupton et al. 2004).

We also experimented with rescaling images to $128 \times 128$ pixels in proportion to the Petrosian radius of each galaxy. However, our CNN performs better when natural scale is used for all images. We therefore, have used the natural scale images in our analysis.

We show some of the characteristics of our sample galaxies in Figure 1. The top-left panel of this figure contains the distribution of the $B/T$ ratio of our sample galaxies, which is seen to be fairly broad and peaking around $B/T = 0.6$. Apparent magnitudes of galaxies

Figure 1. Distribution of $B/T$, apparent magnitude and Petrosian radius for our galaxy sample as measured in the SDSS $r$ band. We selected galaxies based on good two component fits as described in Section 2.1. Our sample of ~103,000 galaxies is about 15 per cent of the Meert catalogue.

1 https://skyserver.sdss.org/dr15/en/tools/chart/chartinfo.aspx
in our sample range from 13.13 at the bright end to 19.16 at the faint end, however, most of them lie in the range of 16 to 18 as seen in the top-right panel of Figure 1. The bottom-left panel of Figure 1 depicts the distribution of the Petrosian radius of galaxies in our sample. The full range of Petrosian radius goes from 1.61 arcsec to 68.31 arcsec but majority of them are smaller than 8 arcsec. Lastly, it should be noted that our requirement of a good two component fit excludes the extreme cases, pure ellipticals ($B/T = 1$) and bulgeless disc galaxies ($B/T = 0$) from our sample. Therefore, any result presented in this work should be taken to be valid for regular disc galaxies only.

3 CONVOLUTIONAL NEURAL NETWORKS

CNNs have now established themselves as the go to learning algorithm in computer vision, with state of the art performance in almost every task in this field (Krizhevsky et al. 2012; He et al. 2017). Since input to our model will be images, our task comes within the field of computer vision.

A CNN is a neural network with convolution layers. A convolution layer performs 2D convolution operations on the input image. The filters for these convolution are the learning parameter for the convolution layer. A typical CNN architecture has first few layers as convolution layers, followed by few fully connected layers (Lecun et al. 1998). As convolution operations work well for extracting image features, the part of the network containing convolution layers is also called Feature Extractor.

The difference in using a feature extractor of a CNN and extracting features using regular image processing is that the filters in a convolution layer are learned by the network, based on the task. The first few layers extract low-level features, like edges and color shifts. As we go deeper in the network, high-level abstract features are extracted using the features from previous layers. These high-level task-specific feature extraction is what makes CNNs quite effective for computer vision tasks.

The learning is generally done using backpropagation and gradient descent algorithms. The fully connected layers, which follows the feature extractor, flatten the different features learned into a single vector. The design of the final fully connected layer and the loss function used for training depends on the functionality required and varies from task to task. This last layer can be a classifier or can perform regression on the extracted features.

Since we are learning from galaxy images to predict $B/T \in (0, 1)$, we have used a CNN model for image regression. KERAS (Chollet et al. 2015) – an open source neural network library written in PYTHON – is used for building and training the CNN model.

3.1 Model Architecture

A number of different CNN architectures exist based on the number of convolution layers, kinds of convolution performed by them and their placement in the network. We tested the architectures with good performance on ImageNet ILSVRC challenge (Russakovsky et al. 2015), i.e. InceptionV3 (Szegedy et al. 2015), Resnet50 (He et al. 2015), Xception (Chollet 2016) and Inception-ResnetV2 (Szegedy et al. 2016). After extensive experimentation we found that the Xception architecture provides the best performance for our dataset. All the models were trained from scratch with random initialisation. Inception-ResNetV2 performed slightly worse than Xception but with more than double the parameters. Therefore, Xception architecture was chosen due to its high performance as well as parameter efficiency.

The Xception network is a stack of depth-wise separable convolution layers with residual connections (He et al. 2015), as shown in Figure 2. Depth-wise separable convolution separates the spatial and cross-channel correlations completely, by breaking normal convolution operation into two sub-operations, point-wise and depth-wise convolution. Using two operations, instead of one normal convolution also results in fewer parameters and faster computation. Reduced number of parameters, in turn, reduces variance (overfitting), which is crucial for models with millions of parameters. Global Average Pooling (GAP) as described in Lin et al. (2013) is used after the Xception backbone to further reduce parameter count and variance without much loss of relevant information. The output of this layer is directly connected to a single output neuron with sigmoid activation. Sigmoid activation is used for the output layer because all possible values of $B/T \in [0, 1]$.

3.2 Training Methodology

3.2.1 Data Pre-processing

We processed the images before passing them as input to the model. This can be a crucial step to improve performance with the same model and dataset. We normalise the image pixel values for faster optimization during training. Data augmentation is also performed while training to regularise the model. A train-val-test split of 60-20-20 is used to quantify the training results and measure training performance after each epoch.

Data augmentation is a process to create more data samples for training from the existing samples. We do this by exploiting the invariances in our dataset. Galaxy images have rotational invariance, i.e. rotating an image by an arbitrary angle would not affect the B/T. Similarly, they are invariant to reflection, i.e. flipping the image horizontally or vertically would result in same B/T. We use these invariances to create multiple images from a single image. While training, every input image is rotated by a randomly selected angle between 0 and 360 degrees. They are also flipped horizontally and vertically with probability of 0.5 for each flip. We also normalise every pixel value to be in range [0, 1], by dividing each value by 255, the maximal value of an 8 bit pixel. No other processing was done to the images. It is to be noted that in our dataset all the target galaxies are at the centre of the image. We used the ImageDataGenerator class in KERAS to implement the augmentation and normalisation operations.

We use 20000 images each for test and validation sets. These images are chosen randomly from the 103757 galaxies selected as per section 2. The remaining 63757 samples constitute the training set. None of the images in the test or validation set are used for training. After each epoch, we validate the model on the validation set to measure its performance on unseen data. This also helps us to determine bias and variance of the model. This is necessary as we do not want the model to suffer from high bias (under-fitting) or high variance (over-fitting). We selected the model with best performance on the validation set, after the training. We then tested the selected model on the test set to gauge its performance in real world scenarios.

3.2.2 Transfer Learning

Pre-trained networks are known to converge much faster than a randomly initialised network with potential gain in performance.
similar fine-tuning and pre-training datasets. In our case, because the pre-training dataset (imagenet) and fine-tuning dataset (galaxies) are not similar, we do not observe better performance by using pre-training. Hence, we have trained the model from scratch using random initialisation as our main model. However, one can still benefit from better weight initialisation from pre-training, which leads to faster convergence when training on our dataset. This in turn leads to much less use of computational resources to achieve comparable results. This trade-off is discussed further in Section 4.3.1.

3.2.3 Early Stopping

Early stopping is used to cease training after considerable over-fitting and to restore weights from the model with the best performance on the validation set. The number of epochs the system waits till there is no further improvement while training is called patience.

3.2.4 Tuning Hyperparameters

(i) **Loss Function**: Loss function is a hyperparameter which determines how we want to train a model (LeCun 1988). It is a function of predicted values $\hat{y}$ and ground-truth $y$, and measures the difference between them. Mean Absolute Error (MAE) is a commonly used loss function for regression tasks. It was found to perform better than Mean Squared Error and Huber Loss, which are two other widely used loss functions used for regression. Therefore, we have used the MAE loss function.

(ii) **Optimizer**: The aim of an optimizer is to find the global minimum for loss function in the parameter space. Adaptive Moment Estimation or Adam (Kingma & Ba 2014) is currently the most widely used variant of Gradient Descent (GD) optimizer. It uses exponentially weighted average of gradients, or momentum, and root mean square propagation (RMSProp) with GD for adaptive estimation of lower-order moments. We have used Adam for our analysis. Learning rate of 5e-4 is used, which is determined empirically from the learning curves.

(iii) **Mini-batch size**: Using the entire training set for one update of the optimizer can be impractical due to limited computational resources. It is more practical and faster to use only a subset of training examples for GD update. This is called Mini-batch Gradient Descent and the subsets of training set are called mini-batches. We determine the mini-batch size as the largest number of samples which can be processed without exhausting our GPU RAM. We have used 128 images as the mini-batch size due to limited GPU RAM.

(iv) **Early stopping patience**: As described in Section 3.2.3, patience determines number of epochs the system waits after there is no improvement while training. It is determined empirically from the learning curves, based on the learning rate. We never observed the loss decrease after over-fitting for more than 15 epochs, without using early stopping. Therefore, we used a patience of 20 epochs.

4 RESULTS AND ANALYSIS

4.1 Performance Metric

The metric used to determine the performance of the model is based on absolute error in prediction of the B/T ratio. Performance is measured as the fraction of test samples with $AE$ less than a threshold absolute error ($AE$). We have defined accuracy as performance for $AE$ cut-off = 0.1. This choice of the $AE$ cutoff is somewhat arbitrary; with threshold $AE = 0.1$, for a galaxy with true $B/T = 0.4$. 

---

Figure 2. Xception architecture (Chollet 2016) used in our work. SConv represents separable convolution. Number of filters are specified with every Conv and SConv operations. Every Conv and SConv operation uses filter size of 3x3 unless explicitly specified. Every Conv 1x1 and first Conv 32 use a stride of 2x2, rest all Conv and SConv use stride of 1x1.
we label the prediction a success if the predicted value $(B/T)_P$ is $0.3 < (B/T)_P < 0.5$

4.2 Learning Curve

The curves in Figure 3 show training and validation loss for each training epoch with learning rate = 5e-4. The first curve is for Xception model trained from scratch and the second curve is for pre-trained Xception model. We have trained using Google Colaboratory\textsuperscript{2} environment with NVIDIA Tesla T4 GPU (12.7 GB GPU RAM) and 12 GB CPU RAM. Each epoch takes about seven minutes to complete with the specified hardware. Early stopping is used with patience of 20 epochs. We observed the lowest validation loss value of 0.054 at epoch 89 for the Xception model trained from scratch with random initialisation. After that the validation loss stopped decreasing. The learning stopped at epoch 109 with the use of early stopping, as the model had started over-fitting. The weights at epoch 89 are hereby used for analysis. The complete training took close to 13 hours. The pre-trained model is observed to converge much quicker with the same learning rate. The pre-training was done on ImageNet ILSVRC challenge dataset (Russakovsky et al. 2015). This subset of ImageNet dataset contains over a million labelled images from 1000 classes. We then retrain the whole model with our training data. We observed the lowest validation loss value of 0.0539 at epoch 39 for the pre-trained Xception model.

4.3 Bulge-to-Total Luminosity Ratio

4.3.1 Performance

With the Xception model trained from scratch with random initialisation, we record an accuracy of 85.7 per cent and MAE of 0.0543 on the test set. (see fraction (y axis) at absolute error (AE) = 0.1 in Figure 4). This model is subsequently referred as the main model. The accuracy with InceptionV3, Resnet50 and Inception-ResnetV2 are 84.7, 84.9 and 85 per cent respectively. These models are also trained from scratch with random initialisation.

Using a pre-trained Xception model, we record an accuracy of 85.2 per cent and MAE of 0.055 on the test set. The pre-trained model reaches loss minimum in just 39 epochs as opposed to 89 epochs for the main model, as shown in Figure 3. There is hence a trade-off between the performance increase and the additional computational resources required to achieve the higher performance. The test set is kept the same for both pre-trained accuracy measurement and main model accuracy measurement.

The main model takes 44.1 seconds to predict $B/T$ for the 20000 test images on an NVIDIA Tesla P100 GPU with 12.7 GB GPU RAM. On a more modest Intel Xeon CPU at 2.3GHz with 12 GB RAM, the main model takes 17 minutes to predict $B/T$ ratio for the 20000 test images.

It is important to test the sensitivity of our predictions to the fact that the galaxy of interest is always located at the centre of the image. Shifting the target galaxies during testing, by as few as 10 pixels in either X or Y direction reduces the accuracy to ~55 per cent and shifting them by 20 pixels reduces it further to ~30 per cent. Note that with our acceptance criterion of $AE < 0.1$ and with $0 < B/T < 1$, about 20 per cent accuracy would be achieved by random chance, even without any learning. Thus the strong dependence of accuracy in $B/T$ ratio prediction on the distance of the target galaxy from the centre, shows that our model is not shift-invariant and has learned the position of the target galaxies.

Some randomly sampled galaxies with their real and predicted $B/T$ values are shown in Figure 5 and Table 1. It is clear that the $B/T$ ratio is estimated well over a wide range of galaxy size, brightness and morphology.

4.3.2 Error Analysis

The predicted versus true $B/T$ values, as well as the mean and standard deviation of predicted samples are shown in Figure 6. We observe that the model consistently underestimates the $B/T$ ratio for samples with true value higher than 0.65.

Error with respect to true $B/T$ can be seen in Figure 7. It is observed that the number of outliers increases as we move to the edges of the $B/T$ distribution. This is in line with the fact that fewer training samples were present at the edges of the $B/T$ distribution (see Figure 1). It is possible that based on our training set there are certain regions in the galaxy size-apparent magnitude relation which are not well trained and hence have higher errors. In Figure 7 we plot the distribution of MAE with galaxy size and apparent magnitude. Notice that the regions of higher error typically lie at the edges which have fewer training samples (see Figure 1). Along with

\textsuperscript{2} https://colab.research.google.com/
Figure 4. Performance as a function of absolute error cut-off. For any point \((x,y)\) on this graph, \(x\) is the absolute error and \(y\) represent fraction of test samples with absolute error less than \(x\). On the entire test set of 20000 images, we achieved accuracy of 85.7 per cent with the main model. Accuracy is defined as the fraction of test samples with absolute error less than 0.1.

| Image | True \(B/T\) | Pred \(B/T\) | AE | AE |
|-------|--------------|--------------|----|----|
| I     | 0.79         | 0.73         | 0.06|    |
| II    | 0.31         | 0.38         | 0.07|    |
| III   | 0.65         | 0.65         | 0.00|    |
| IV    | 0.32         | 0.33         | 0.01|    |
| V     | 0.55         | 0.64         | 0.09|    |
| VI    | 0.23         | 0.19         | 0.04|    |
| VII   | 0.58         | 0.60         | 0.02|    |
| VIII  | 0.25         | 0.36         | 0.11|    |
| IX    | 0.63         | 0.55         | 0.08|    |
| X     | 0.28         | 0.30         | 0.02|    |
| XI    | 0.57         | 0.58         | 0.01|    |
| XII   | 0.67         | 0.67         | 0.00|    |
| XIII  | 0.27         | 0.29         | 0.02|    |
| XIV   | 0.6          | 0.73         | 0.13|    |
| XV    | 0.56         | 0.54         | 0.04|    |

Table 1. Details of predictions for samples in Figure 5 with the main model. Predicted \(B/T\) and AE are rounded off to 2 decimal places.

these are a few "islands" which are outliers in the size-apparent magnitude relation and thus have quite high errors. The higher signal to noise ratio in the images of brighter galaxies is likely to make the prediction of the \(B/T\) ratio with CNN easier. Ignoring galaxies fainter than \(r\)-band mag of 17 in the test set, we get a sample of 7557 bright galaxies. Accuracy on this set is 87 per cent with MAE=0.0518 using the main model.

4.3.3 Effect of Bright Neighbours

With the high sensitivity of modern digital sky surveys, it is quite common to have unrelated stars and galaxies in an image cutout of the galaxy of interest. Occasionally a bright object may lie close to the galaxy of interest. It is important to analyses the effect of bright neighbouring objects on our predictions. To test this, we pick images which have their brightest pixel outside the central region. We define this central region as 15 pixel distance from the centre, or about 6 arcsec. Specifically, if the brightest pixel in an image is more than 6 arcsec away from the centre, we select that image for the bright neighbour set. The 6 arcsec cutoff is based on the median Petrosian radius of our dataset, which is \(\sim 4.8\) arcsec. The impact of the presence of a bright object on our \(B/T\) predictions is shown in Table 2. We observe that the performance is reduced by around 3.5 per cent for galaxies with bright neighbouring objects. This also means that our results can be improved further to 86.33 per cent accuracy, if we only consider galaxies with no bright neighbours. Further, ignoring galaxies fainter than \(r\)-band mag of 17 in the no bright neighbour set, we get a sample of 6460 bright galaxies with no bright neighbours. Accuracy on this set is 87.5 per cent with MAE=0.051. We note that the impact on the accuracy of prediction by considering only bright galaxies and/or those without bright neighbours is fairly small, but measurable.

4.3.4 Catastrophic Failures

We treat galaxies with absolute error in \(B/T\) estimation greater than 0.2 as failures. From the 20000 test images, there are a total of 239 failures with the main model. This shows that only \(\sim 1.2\) per cent of galaxies suffer such failure in our \(B/T\) estimation.

We treat galaxies with absolute error in \(B/T\) estimation greater than 0.3 as catastrophic failures. There are 10 such galaxies using the main model, as shown in Figure 8 and Table 3. This shows that 0.05 per cent of galaxies suffer catastrophic failure in our \(B/T\) estimation. We observed no obvious patterns in the properties of the galaxy images which were in the failure or catastrophic failure categories. Many images that were visually similar to the ones present in these failure sets had accurate \(B/T\) estimates. Only 3 of these failures had bright neighbours.

5 SUMMARY AND CONCLUSIONS

In this work, we have attempted to apply deep learning techniques to a tough regression problem in galaxy morphology. Traditionally, galaxy morphological classification, when done by human experts, is via careful examination of broad-band galaxy images. Inspite of great experience and expertise of the human classifiers, a standard deviation of about 0.5 in the Hubble T-type – where the same galaxy is classified multiple times by the same expert – is about the best performance that can be achieved (Nair & Abraham 2010). The alternative to the human expert based approach, is the quantitative bulge-disk decomposition which assumes different analytic profiles for the galaxy components. This approach has the benefit of being
quantitative and reproducible, although the time taken for classification scales linearly with the number of galaxies. For human classifiers to guess the $B/T$ luminosity ratio of a galaxy, by visual examination of broad-band images, is considered sufficiently error prone, that it is never attempted in scientific analysis. In this work, we have demonstrated that a task which is considered too difficult for humans, is being performed with reasonably high accuracy by a carefully chosen deep learning network.

We predict the $r$-band $B/T$ ratio of galaxies using end-to-end machine learning, with 85.7 per cent predictions having $AE < 0.1$ with no manual intervention. We used the Xception CNN architecture (Chollet 2016) with GAP (Lin et al. 2013) as the machine learning model. We used composite colour images constructed from individual images in the $(g, r, i)$ bands from SDSS DR15 for training and testing the model. This study only deals with $B/T$ ratio estimates for two-component galaxies as defined in the Meert Catalogue (Meert et al. 2015). The main model takes less than a minute to predict the $B/T$ ratio for 20000 test images. A similar sized sample
would take ~100 hours with conventional bulge-disk decomposition pipelines like GALFIT (Peng et al. 2002b) using wrappers like Py-MORPH (Vikram et al. 2010), without even including time for data pre-processing. The decomposition codes also require some amount of human intervention. It all adds up to a significant overhead over our method. This overhead would become enormous while predicting properties of billions of galaxies with the next generation of galaxy surveys during the LSST and Euclid eras. Thus our proposed machine learning approach has the potential to save tremendous amount of time, computational and financial resources, along with reduced human effort in predicting an important parameter that characterises galaxies.

ACKNOWLEDGEMENTS

We thank the anonymous referee whose insightful comments helped improve both the content and presentation of this paper. HG thanks NCRA-TIFR for hosting him while the bulk of this work was completed. We would like to thank the Google Colaboratory for providing hardware and software resources for this study. **MATPLOTLIB** (Hunter 2007) was used for producing the plots in this paper.

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 Insti-
Table 3. Catastrophic failures with the main model. Refer to Figure 8 for the image labels.

| Image | True B/T | Pred B/T | AE |
|-------|----------|----------|----|
| I     | 0.81     | 0.49     | 0.32|
| II    | 0.72     | 0.40     | 0.32|
| III   | 0.67     | 0.30     | 0.37|
| IV    | 0.74     | 0.22     | 0.52|
| V     | 0.74     | 0.40     | 0.34|
| VI    | 0.51     | 0.16     | 0.35|
| VII   | 0.76     | 0.45     | 0.31|
| VIII  | 0.64     | 0.34     | 0.30|
| IX    | 0.26     | 0.64     | 0.38|
| X     | 0.73     | 0.32     | 0.41|

REFERENCES
Abazajian K. N., et al., 2009, ApJS, 182, 543
Abraham S., Aniyam A. K., Kembhavi A. K., Philip N. S., Vaghmare K., 2018, MNRAS, 477, 894
Aguado D. S., et al., 2017, ApJS, 249, 23
Barden M., Hauder B., Peng C. Y., McIntosh D. H., Guo Y., 2012, MNRAS, 422, 449
Baron D., 2019, arXiv e-prints, p. arXiv:1904.07248
Baron D., Poznanski D., 2017, MNRAS, 465, 4530
Bilicki M., et al., 2018, A&A, 616, A69
Bluck A. F. L., Mendel J. T., Ellison S. L., Moreno J., Simard L., Patton D. R., Starkenburg E., 2014, MNRAS, 441, 599
Canameras R., et al., 2020, arXiv e-prints, p. arXiv:2004.13048
Cappellari M., et al., 2013, MNRAS, 432, 1862
Chollet F., 2016, Xception: Deep Learning with Depthwise Separable Convolutions (arXiv:1610.02357)
Chollet F., et al., 2015, Keras, https://github.com/fchollet/keras
Collister A. A., Lahav O., 2004, PASP, 116, 345
Ellison S. L., Teimoorinia H., Rosario D. J., Mendel J. T., 2016, MNRAS, 455, 370
Erwin P., 2015, ApJ, 799, 226
Firth A. E., Lahav O., Somerville R. S., 2003, MNRAS, 339, 1195
Fisher D. B., Drory N., 2008, AJ, 136, 773
Gadotti D. A., 2009, MNRAS, 393, 1531
Gao H., Ho L. C., 2017, ApJ, 845, 114
Graham A. W., Worley C. C., 2008, MNRAS, 388, 1708
He K., Zhang X., Ren S., Sun J., 2015, Deep Residual Learning for Image Recognition (arXiv:1512.03385)
He K., Gkioxari G., Dollár P., Girshick R., 2017, Mask R-CNN (arXiv:1703.06870)
Hopkins P. F., et al., 2010, ApJ, 715, 202
Huertas-Company, M., Aguerri, J. A. L. Bernardi, M. Mei, S. Sánchez Almeida, J. 2011, A&A, 525, A157
Huertas-Company, M., et al., 2015, ApJS, 221, 8
Hunter J. D., 2007, Computing in Science & Engineering, 9, 90
Kent S. M., 1985, ApJS, 59, 115
Kingma D. P., Ba J., 2014, Adam: A Method for Stochastic Optimization (arXiv:1412.6980)
Kormendy J., Ho L. C., 2013, ARA&A, 51, 511
Kormendy J., Kennicutt Robert C. J., 2004, ARA&A, 42, 603
Kormendy J., Richstone D., 1995, ARA&A, 33, 581
Krizhevsky A., Sutskever I., Hinton G. E., 2012, in Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1, NIPS’12, Curran Associates Inc., USA, pp 1097–1105, http://dl.acm.org/citation.cfm?id=2999134.2999257
LeCun Y., 1988.
LeCun Y., Bengio Y., Hinton G., 2015, Nature, 521, 436
Lecun Y., Bottou L., Bengio Y., Haffner P., 1998, Proceedings of the IEEE, 86, 2278
Li R., et al., 2020, arXiv e-prints, p. arXiv:2004.02715
Lin M., Chen Q., Yan S., 2013, Network In Network (arXiv:1312.4400)
Lupton R., Blanton M. R., Fekete G., Hogg D. W., O’Mullane W., Szalay A., Wherry N., 2004, PASP, 116, 133
Marconi A., Hunt L. K., 2003, ApJ, 589, L21
Meert A., Vikram V., Bernardi M., 2015, MNRAS, 446, 3943
Méndez-Abreu J., Aguerri J. A. L., Corsini E. M., Simonneau E., 2008, A&A, 478, 353
Mitchell T. M., 1997, Machine Learning. McGraw-Hill, New York
Nair P. B., Abraham R. G., 2010, ApJS, 186, 427
Pasquet J., Bertin E., Treyer M., Arnouts S., Fouchez D., 2019, A&A, 621, A26
Peng C. Y., Ho L. C., Impy C. D., Rix H.-W., 2002a, AJ, 124, 266
Peng C. Y., Ho L. C., Impy C. D., Rix H.-W., 2002b, AJ, 124, 266
Philip N. S., Wadadekar Y., Kembhavi A., Joseph K. B., 2002, A&A, 385, 1119
Ribli D., Dobos L., Csabai I., 2019, MNRAS, 489, 4847
Robotham A. S. G., Taranu D. S., Tobar R., Moffett A., Driver S. P., 2016, MNRAS, 466, 1513–1541
Russakovskiy O., et al., 2015, International Journal of Computer Vision (IJCV), 115, 211
Sadegh I., Abdalla F. B., Lahav O., 2016, PASP, 128, 104502
Sharma K., Kembhavi A., Sivarani T., Abraham S., 2019, Bulletin de la Societe Royale des Sciences de Liege, 88, 174
Simard L., 2010, GIM2D: Galaxy Image 2D (ascl:1004.001)
Suruna S., Wadadekar Y., Bait O., Bhosale H., 2020, MNRAS, 493, 4808
Szegedy C., Vanhoucke V., Ioffe S., Shlens J., Warden J., 2015, Rethinking the Inception Architecture for Computer Vision (arXiv:1512.00567)
Tagliaferri R., Jaffe D., Andreon S., Capozziello S., Donalek C., Giordano G., 2003, Neural Networks for Photometric Redshifts Evaluation. pp 226–234, doi:10.1007/978-3-540-45216-4_26
Teimoorinia H., Bluck A. F. L., Ellison S. L., 2016, MNRAS, 457, 2086
Taccino D., Huertas-Company M., Decencière E., Velasco-Forero S., Domínguez Sánchez H., Dimauro P., 2017, MNRAS, 475, 894–909
Vanzella E., et al., 2004, A&A, 423, 761
Vazza F., et al., 2014, PASP, 126, 936
Viana C. R., et al., 2012, ApJS, 201, 6
Vikram V., Bernardi M., 2015, MNRAS, 455, 370
Wadadekar Y., Bluck A. F. L., Kembhavi A., Vijayagovindan G. V., 2010, MNRAS, 409, 1379–1392
Wadadekar Y., 2005, PASP, 117, 79–85
Wadadekar Y., Robbason B., Kembhavi A., 1999, AJ, 117, 1219
Wu C., et al., 2019, MNRAS, 482, 1211
de Souza R. E., Gadotti D. A., dos Anjos S., 2004, ApJS, 153, 411–427

This paper has been typeset from a \TeX/\LaTeX{} file prepared by the author.