Transfer learning for galaxy morphology from one survey to another


Affiliations are listed at the end of the paper

Accepted 2018 December 18. Received 2018 November 27; in original form 2018 April 30

ABSTRACT

Deep learning (DL) algorithms for morphological classification of galaxies have proven very successful, mimicking (or even improving) visual classifications. However, these algorithms rely on large training samples of labelled galaxies (typically thousands of them). A key question for using DL classifications in future Big Data surveys is how much of the knowledge acquired from an existing survey can be exported to a new data set, i.e. if the features learned by the machines are meaningful for different data. We test the performance of DL models, trained with Sloan Digital Sky Survey (SDSS) data, on Dark Energy Survey (DES) using images for a sample of \( \sim 5000 \) galaxies with a similar redshift distribution to SDSS. Applying the models directly to DES data provides a reasonable global accuracy (\( \sim 90 \) per cent), but small completeness and purity values. A fast domain adaptation step, consisting of a further training with a small DES sample of galaxies (\( \sim 500–300 \)), is enough for obtaining an accuracy >95 per cent and a significant improvement in the completeness and purity values. This demonstrates that, once trained with a particular data set, machines can quickly adapt to new instrument characteristics (e.g. PSF, seeing, depth), reducing by almost one order of magnitude the necessary training sample for morphological classification. Redshift evolution effects or significant depth differences are not taken into account in this study.

Key words: methods: observational – methods: photometric – surveys – galaxies: structure.

1 INTRODUCTION

Astronomy is entering the Big Data era. We are experiencing a revolution in terms of available data due to surveys such as COSMOS (Scoville et al. 2007), SDSS (Eisenstein et al. 2011), DEEP2 (Newman et al. 2013), DES (DES Collaboration 2016), etc.

The close future is even brighter with missions like EUCLID (Racca et al. 2016) or LSST (LSST Science Collaboration et al. 2017), offering photometric, quasi-spectroscopic data of millions/billions of galaxies.

One key measurement severely affected by this Big Data transition is galaxy morphology estimated from images. Galaxies exhibit a great variety of shapes and their morphology is intimately related to their stellar content. In addition, the light profiles provide information about their mass assembly, interactions, accretion,
quenching processes or feedback (e.g. Conselice 2003; Bournaud et al. 2014; Kaviraj 2014; Belfiore et al. 2015; Dubois et al. 2016). It is therefore crucial to have accurate morphological classifications for large samples of galaxies.

Galaxy morphological catalogues have been usually based on visual classifications. Unfortunately, visual classification is an incredible time-consuming task. The size of present and future Big Data surveys, containing millions of galaxies, make this approach a near impossible task. One beautiful solution to this problem was the Galaxy Zoo project (Lintott et al. 2011), which involved more than 100k volunteer citizens to morphologically classify the full SDSS sample and has now been extended to other higher redshifts and surveys (e.g. CANDELS survey, Simmons et al. 2016; DECaLS survey). However, with the next generation of surveys, we are reaching the limit of applicability of these approaches. It is estimated that about a hundred years would be needed to classify all data from the EUCLID mission with a Galaxy Zoo-like approach, unless the number of people involved is significantly increased. A question naturally arises: can human classifiers be replaced by algorithms?

Automated classifications using a set of parameters that correlate with morphologies, e.g. CAS-methods (Concentration-Asymmetry-Smoothness, Conselice 2003) or Principal Component Analysis (Lahav et al. 1995, 1996; Banerji et al. 2010, and references therein) have been attempted. However, the parameter extraction also requires large amounts of time. DL algorithms where, in contrast to classic machine learning algorithms, no image pre-processing is needed, have come to the rescue for image analysis of large data surveys. The use of convolutional neural networks (CNNs) to learn and extract the most meaningful features at pixel level have been shown to produce excellent results for pattern recognition in complex problems and are widely used by many technology giants such as Google. CNNs have demonstrated their success for morphological classification of galaxies in The Galaxy Challenge, a Kaggle competition for reproducing the Galaxy Zoo 2, where the top three algorithms used CNNs (e.g. Dieleman, Willett & Dambre 2015). At higher redshifts, Huertas-Company et al. (2015) also showed that CNNs represent a major improvement with respect to CAS-based methods.

In a companion paper, Domínguez Sánchez et al. (2018, DS18 hereafter), we combine the best existing visual classification catalogues with DL algorithms to provide the largest (670 000 galaxies from DR7-SDSS survey) and most accurate morphological catalogue to date. The catalogue includes two flavours: T-type, related to the Hubble sequence, and Galaxy Zoo 2 classification scheme. One of the main improvements with respect to previous works (Dieleman et al. 2015) is that only galaxies with robust classifications (large agreement between Galaxy Zoo classifiers) are used for training each task. This helps the models to detect the relevant features for each question and a smaller training sample is required for the models to converge.

In spite of this improvement on the training approach, these algorithms still rely on large training sets (around 5000–10 000 galaxies, depending on the classification task). A key question, in view of using DL-based algorithms to assess the morphologies of galaxies in future Big Data surveys, is therefore how much of the knowledge acquired from an existing survey can be exported to a new data set, i.e. can the features learned by an supervised process on a given data set be transferred to a new data set with different properties? And – if not – what is the cost of updating those features (in terms of new objects to be classified from the new data set)?

This process, usually referred to as transfer learning or fine-tuning in the literature, is becoming popular for general image recognition (e.g. Bengio 2012; Yosinski et al. 2014; Tajbakhsh et al. 2016) and several recent works explore the optimal strategy to transfer knowledge (e.g. Guo et al. 2018; Kornbluth, Shlens & Le 2018; Shermin et al. 2018 and references therein). However, transfer learning using astronomical data has not been yet fully explored.

Some preliminary tests have been performed by our team to assess the performance of DL algorithms, trained with simulated data, on real data. In a recent paper (Tuccillo et al. 2017) we show that a DL machine trained on one-component Sérsic galaxy simulations (with real HST/CANDELS F160W PSF and noise) can accurately recover parametric measurements of real HST galaxies with at least the same quality as GALFIT (Peng et al. 2002), but several orders of magnitude faster. It shows indications that DL is able to transition from simplistic simulations to real data without seriously impacting the results.

In a recent paper, Ackermann et al. (2018) investigate transfer learning for galaxy merger detection by retraining CNNs first trained on pictures of everyday objects (i.e. ImageNet data set, Deng et al. 2009). In this work, we study transfer learning for morphological classification of galaxies between different astronomical surveys. To that end, we take advantage of the DL models trained with SDSS data to test their performance when applied to DES survey, with and without training on DES images. This is, to the best of our knowledge, the first work addressing the ability of DL models to transfer knowledge for different data sets. In a recent work, Pérez-Carrasco et al. (2018) provide a morphological catalogue of CLASH (Postman et al. 2012) galaxies by fine-tuning a CNN pre-trained on CANDELS survey (Grogin et al. 2011). They confirm the result presented in this paper: that transfer learning reduces the number of labelled images needed for training.

The paper is organized as follows: in Section 2, we describe the SDSS-based DL models, DES images, and morphological catalogue used in this work; in Section 3, we explain our methodology; in Section 4, we discuss the results and in Section 5 we summarize our conclusions.

2 DATA

In this paper, we test the performance of DL models, trained with SDSS-DR7 data (Abazajian et al. 2009), on DES images. The morphological classification of DES galaxies comes from the DECaLS - Galaxy Zoo catalogue. In this section we describe the SDSS DL models, DES images, and the morphological catalogue used throughout the paper.

2.1 Deep learning models trained with SDSS-DR7 data

In DS18 we morphologically classify ~670 000 SDSS-DR7 galaxies with automated DL algorithms. The galaxies correspond to the sample for which Meert, Vikram & Bernardi (2015, 2016) provide accurate photometric reductions. Reader can refer to DS18 for a detailed explanation on the data and methodology but, in short, we use two visual classification catalogues, Galaxy Zoo 2 (GZ2 hereafter, Willett et al. 2013) and Nair & Abraham (2010), for training CNNs with colour SDSS-DR7 images. We obtain T-types and a series of GZ2-type questions (disc/features, edge-on galaxies, bar signature, bulge prominence, roundness, and mergers) for a sample of galaxies with $r$-band Petrosian magnitude limits $14 \leq$
or very similar, which allows us to perform this exercise. DECaLS surveys, the Galaxy Zoo classifications will be identical models on DES images. Given the similarities between DES and classification catalogue as the ground truth to test (and train) our of 1 arcsec separation). Note that, since our final aim will be to adapt automated methods for morphological classification of galaxies.

DE is a photometric survey utilizing the Dark Energy Camera (DECam; Flaugher et al. 2015) on the Blanco-4m telescope at Cerro Tololo Inter-American Observatory (CTIO) in Chile to observe ∼5000 deg² of the southern sky in five broad-band filters, g, r, i, z, and Y (∼400 nm to ∼1060 nm) with a resolution of 0.263 arcsec pixel⁻¹. The magnitude limits and median PSF full width at half-maximum (FWHM) for the first year data release (Y1A1 GOLD) are 23.4, 23.2, 22.5, 21.8, 20.1 mag and 1.25, 1.07, 0.97, 0.89, 1.07 arcsec, respectively (from g to Y, see Drlica-Wagner et al. 2017 for a detailed description of the survey). In this work, we use standard DES cut-outs from the internal Y1A1 data release.

The main difference of DES/DECaLS with respect to SDSS images is the use of a larger telescope and better seeing conditions, which so (average) observing conditions are very similar to the DES instrumental effects. This work aims to be a first proof of concept and not a full morphological classification catalogue. The redshift distribution of the DES galaxies used in this work is very similar to the SDSS (see 2.3), so no evolution effects are included: we are only changing the instrument and survey depth and spatial resolution. We leave for a forthcoming work a thoughtful study on the brightness and redshift effect on the models performance.

We focus our analysis on the binary questions from the GZ2 scheme, since they are the easiest to evaluate. We note that there is one model per question. The three classification tasks that we evaluate are:

- Q1: Galaxies with disc/features versus smooth galaxies. We consider as positive examples galaxies with disc or features (Y = 1 in our input label matrix).
- Q2: Edge-on galaxies versus face-on galaxies. Edge-on galaxies are considered positive cases.
- Q3: Galaxies with bar signature versus galaxies with no bar presence. Barred galaxies are positive cases.

3.1 Deep learning architecture

The methodology used in this paper (in terms of training sample selection, model input, and DL model architecture) is exactly the same as in DS18, where the reader can find a detailed explanation about the procedure. In this study, we do not aim to maximize absolute model performance, but rather to study knowledge transfer on a well-known architecture. To facilitate the reader, in Table 1 we summarize the DL model architecture, which consists of four convolutional layers (with ReLU activation, Max Pooling, and dropout) and one fully connected layer (also referred to as the dense layer). The total number of free parameters is 2602849 (see also figure 1 in DS18).

To keep the methodology as similar as possible to DS18, the input for the models are the same as in DS18, i.e. 424 × 424 pixel size images (from DES in this case), which are down-sampled into (69, 69, 3) RGB matrices, with each number representing the flux per pixel at each filter (g, r, i). The flux values are normalized to the maximum value in each filter for each galaxy. The angular size of the images is variable, approximately 10 × R_{P}, where R_{P} is the Petrosian radius of each galaxy (from SDSS).

3.2 Training and transfer learning

In order to assess how much knowledge from one survey can be exported to another, we carry out four experiments:

(a) Apply the models trained on SDSS data directly to DES images, without any further training or fine-tuning on DES data.

(b) Load the weights trained on SDSS data and fine-tune them by training the models with a small DES sample (300–500 galaxies). The training is performed for all the layers in the DL model.

(c) Same as (b) but freezing all the layers (i.e. fixing the weights learned by SDSS) except for the fully connected layer.

(d) Training the models from scratch using a DES training sample with the same size as in (b) and (c).

We compare these experiments with the results presented in DS18 for models trained and tested on SDSS data. Note that in this work we focus on knowledge transfer between different data sets, not

\[^2\text{http://casjobs.sdss.org/ImgCutoutDR7}\]
between different tasks. This means that, for experiments (a) to (c), we use the SDSS models trained for each particular task.

For test (a), the algorithm applies the weights learned by the SDSS models and returns a probability value for each task. For tests (b) to (d) the training procedure is identical to the one used in DS18. We train the models in binary mode. Data augmentation (as explained in DS18) is applied to the DES images to help avoiding overfitting. Balanced weights are used for Q2 and Q3 due to the requirement of having high probability of being disc, edge-on, or barred galaxies. The better quality of DES images reveals with higher detail some galaxy features, such as bulge component or spiral arms.

We only use in the training DES galaxies with a robust uneven proportion of positive and negative examples for this two overfitting. Balanced weights are used for Q2 and Q3 due to the explained in DS18) is applied to the DES images to help avoiding overfitting. Balanced weights are used for Q2 and Q3 due to the requirement of having high probability of being disc, edge-on, or barred galaxies. The better quality of DES images reveals with higher detail some galaxy features, such as bulge component or spiral arms.

We only use in the training DES galaxies with a robust uneven proportion of positive and negative examples for this two overfitting. Balanced weights are used for Q2 and Q3 due to the requirement of having high probability of being disc, edge-on, or barred galaxies. The better quality of DES images reveals with higher detail some galaxy features, such as bulge component or spiral arms.

We test the fine-tuned models on a sample of DES galaxies not used for training. Although this limits the statistics, specially in the case of Q3 (bar signature), it is important to properly evaluate the models. Since we need at least 300 galaxies for training Q3 (and the training sample should include a reasonable number of positive cases), we only have nine barred galaxies left for testing our models (see Table 2). The code used in this work is publicly available at https://github.com/HelenaDominguez/DeepLearning.

4 RESULTS AND DISCUSSION

We use a standard method for testing the performance of our models: receiver operating characteristic (ROC) curve, true positive rate (TPR, also known as recall), precision (P), and accuracy values (e.g. Powers & Ailab 2011; Dieleman et al. 2015; Barchi et al. 2017). For binary classifications, where only two input values are possible (positive or negative cases), the true positives (TP) are the correctly classified positive examples. One can define, in an analogous way, true negatives, false positives, and false negatives (TN, FP, FN, respectively). The true positive rate (TPR), false positive rate (FPR), precision (P), and accuracy (Acc) are expressed as

$$TPR = \frac{TP}{TP + FN}; \quad FPR = \frac{FP}{FP + TN}; \quad P = \frac{TP}{TP + FP}; \quad Acc = \frac{TP + TN}{Total}$$

TPR is a completeness proxy (how many of the true examples are recovered), precision is a contamination indicator (what fraction of the output positive cases are really positive), and accuracy is the fraction of correctly classified objects among the test sample. Since the output of the model is a probability (ranging from 0 to 1), a probability threshold ($P_{th}$) value must be chosen to separate positive and negative cases. The ROC curve represents the TPR and FPR values for different $P_{th}$. A perfect classifier would yield a point in upper left corner or coordinate (0,1) of the ROC space, (i.e. no false negatives and no false positives), while a random classifier would give a point along a diagonal line.

In Fig. 2, we show the ROC curve for the three classification tasks studied in this work for the SDSS model applied to SDSS data (0), for the SDSS model applied to DES data without any training on DES (a), for the model fine-tuned on a small DES sample with...
Table 2. Performance of the models according to the TPR, precision, and accuracy values for the three classification tasks studied in this work. The experiment column specifies the approach used, as explained in Section 3.2. $N_{\text{train}}$ is the number of galaxies used for training. When $N_{\text{train}} = 0$, it means the SDSS model is directly applied to DES data. $N_{\text{test}}$ are the number of galaxies used for testing the models (they fulfill the requirement of having a robust morphological classification, as the training sample), of which $N_{\text{pos}}$ are the positive cases (e.g. galaxies showing disc/features for Q1). Galaxies used for training are not included in the testing sample. This explains the scarcity of barred galaxies used for testing the models with DES training.

<table>
<thead>
<tr>
<th>Question</th>
<th>Experiment</th>
<th>$N_{\text{train}}$</th>
<th>$N_{\text{test}}$</th>
<th>$N_{\text{pos}}$</th>
<th>TPR</th>
<th>Prec.</th>
<th>Acc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>(a) SDSS–DES</td>
<td>0</td>
<td>2409</td>
<td>797</td>
<td>0.48</td>
<td>0.92</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td>(b) SDSS–DES fine-tuned</td>
<td>500</td>
<td>238</td>
<td>78</td>
<td>0.95</td>
<td>0.91</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>(c) SDSS–DES fine-tuned (FCL)</td>
<td>500</td>
<td>238</td>
<td>78</td>
<td>0.96</td>
<td>0.78</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>(d) DES–DES</td>
<td>500</td>
<td>238</td>
<td>78</td>
<td>0.81</td>
<td>0.77</td>
<td>0.85</td>
</tr>
<tr>
<td>Q2</td>
<td>(a) SDSS–DES</td>
<td>0</td>
<td>2851</td>
<td>536</td>
<td>0.91</td>
<td>0.76</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>(b) SDSS–DES fine-tuned</td>
<td>500</td>
<td>738</td>
<td>187</td>
<td>0.96</td>
<td>0.86</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>(c) SDSS–DES fine-tuned (FCL)</td>
<td>500</td>
<td>738</td>
<td>187</td>
<td>0.97</td>
<td>0.77</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>(d) DES–DES</td>
<td>500</td>
<td>238</td>
<td>78</td>
<td>0.97</td>
<td>0.70</td>
<td>0.89</td>
</tr>
<tr>
<td>Q3</td>
<td>(a) DES</td>
<td>0</td>
<td>1768</td>
<td>61</td>
<td>0.57</td>
<td>0.35</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>(b) SDSS–DES fine-tuned</td>
<td>300</td>
<td>86</td>
<td>9</td>
<td>0.89</td>
<td>0.73</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>(c) SDSS–DES fine-tuned (FCL)</td>
<td>300</td>
<td>86</td>
<td>9</td>
<td>1.0</td>
<td>0.5</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>(d) DES–DES</td>
<td>300</td>
<td>86</td>
<td>9</td>
<td>1.0</td>
<td>0.1</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Figure 2. True positive rate (TPR, i.e. fraction of well-classified positive cases) versus false positive rate (FPR, i.e. fraction of wrongly classified positive cases) for different $P_{\text{th}}$ values for the three classification tasks studied in this work, as stated in the legend. We show the performance of the DL models for the four experiments explained in Section 3.2 [labelled (a) to (d), colour-coded as shown in the legend], as well as the results of the models trained with SDSS galaxies and applied to SDSS images (blue dashed line). The number of galaxies used in the training for each question for the SDSS and DES samples are shown in the legend. The knowledge transfer from SDSS plus fine-tuning provides results comparable to the SDSS–SDSS models, but the training sample size can be reduced at least one order of magnitude. The ‘apparent’ better performance of the fine-tuned DES models with respect to the SDSS–SDSS one for Q3 is caused by the small size of the barred test sample (see Table 2).

transfer knowledge from the SDSS model [allowing all layers to be trained (b) or freezing all layers but the fully connected layer (c)], and for the model trained with a random initialization on a small DES sample (d).

In Table 2, we show the TPR, precision, and accuracy values for the same experiments. For simplicity, we only list the values obtained for $P_{\text{th}} = 0.5$ (the standard value for separating positive and negative cases). Both the train and test DES samples are required to have a robust classification in the morphological catalogue (see Section 3). The number of galaxies used for training and testing (and the positive cases), are also given in Table 2.

Our first main result is that, when applying the SDSS models directly to DES images, with no training at all on DES data, the accuracy values obtained are reasonable (>80 per cent), reaching 93 per cent and 95 per cent for Q2 and Q3. However, the accuracy can be misleading when few positive cases are included in the test sample and it is important to consider completeness and purity of the classification. These quantities are strongly dependent on the classification task. For example, for Q1 the precision value is very high (92 per cent), but the completeness is less than 50 per cent. On the other hand, the SDSS model recovers 91 per cent of the DES edge-on galaxies, but the precision value for this task is 76 per cent. For Q3, both the completeness and purity values obtained with the SDSS model are small (0.57 and 0.35, respectively). This indicates that bar identification is a very sensitive task to resolution and depth, while, on the other hand, inclination is less dependent on the survey characteristics.
The second main result is that, after a fast domain adaptation step (i.e. training the models with a small sample – less than 500 – of highly reliable DES galaxies), the models are able to adapt to the new data characteristics and quickly converge, providing results comparable to the ones obtained for the SDSS models applied to SDSS data (see Table 2 and Fig. 2). We tested the performance of the models with DES training samples of different sizes and we found that the presented here are an optimal trade-off between models’ results and training sample size. The accuracy values are ≥ 0.95 for all the classification tasks. For both Q1 and Q2 the completeness reaches at least 95 per cent and the purity values are 91 per cent and 86 per cent, respectively. The TPR and precision values for Q3 are smaller (0.89 and 0.73, respectively), but are severely affected by the test sample statistics. In fact, the model recovers eight out of nine barred galaxies (TP) and there are only three FP cases. After visual inspection, we found that the FN case is not a real barred galaxy but a bulge-dominated galaxy. On the other hand, only one of the three FP cases have $P_{\text{bar}} > 0.6$ according to model (b), and that galaxy shows a bright central feature which could be a distorted bar or a dust lane (see Fig. 3).

Regarding the comparison between experiments (b) and (c), the results are slightly better for all tasks when training both the convolutional filters and the fully connected layer, rather than training the fully connected layer alone. Given the ‘simplicity’ of the CNN used (only four convolutional layers), most of the trainable weights actually come from the fully connected layer (235232 versus 2367617 for the convolutional layer and the fully connected layer, respectively). Despite this, the performance of the models after fine-tuning all the layers is improved. It has been suggested in the literature (e.g. Yosinski et al. 2014) that the first-layer features of deep neural networks are general, in the sense that they can be applied to many data sets and tasks. The results from this work indicate that the features learned by the convolutional layers are in fact important to improve the classification. Note that Yosinski et al. (2014) work is based on different classification tasks using the same input images, while in this work we want to transfer knowledge between different surveys. Our results suggest that the differences arising from different data sets (i.e. the survey image characteristics) have an effect on the features learned by the CNN, and not only by the dense layer.

To better understand the impact of transfer learning from the SDSS models, we train the models with the same DES training sample as in the previous exercises, but now with a random weight initialization. As expected, the performance of the models trained from scratch is worst than the performance of the models after fine-tuning. This demonstrates that using an SDSS initialization leads to a better local minimum during training. However, the results are strongly dependent of the task being trained. For example, the accuracy for Q2 is 89 per cent and the ROC curve is comparable to (even above) the one obtained when applying the SDSS models to the DES data without training (a). On the other hand, a model trained with such a small sample is unable to learn and separate the features related to the presence of a bar, as can be seen from the ROC curve shape and Table 2. This reveals that CNNs efficiency is related to the difficulty level of the classification task being trained (identifying edge-on galaxies is a much easier exercise than detecting bars).

Another interesting point is the fact that the models trained with a small DES sample provide similar results to the SDSS models applied to DES data without fine-tuning (except for Q3, as previously discussed). It suggests that transfer learning is equivalent to a small training step. Note that, for Q1, the area below the ROC curve of model (a) – dark green – is larger (i.e. better performance) than the one for model (d) – light green –, although the accuracy TRP values are smaller. This is because the values in Table 2 are given for $P_{\text{th}} = 0.5$, while the optimal performance for model (a) is obtained by setting $P_{\text{th}} = 0.1$. This means that the knowledge transferred for different data sets needs to be recalibrated.

5 CONCLUSIONS

In this paper, we demonstrate that deep-nets can transfer knowledge from one survey to another and quickly adapt to new domains and data characteristics such as depth, PSF, and instrumental effects. The combination of transfer learning and fine-tuning boosts the models performance and allows for a significant reduction of the training sample size.

The fact that the training sample (and therefore the a priori labelled galaxies) can be reduced by an order of magnitude, once the models are trained with a different data set, is a major discovery in order to apply DL models to future surveys, such as EUCLID or LSST. It means that we will be able to recycle models from previous surveys (within the same redshift distribution), preventing from the huge effort of visually classifying a large sample of galaxies from that particular survey.

It is beyond the scope of this paper to test the effect of the models on more complicated aspects of galaxy surveys, such as redshift evolution. We leave for a forthcoming work this mandatory step to release a reliable morphological catalogue, which will certainly be an add-on value to DES. Also, a major advance of extremely deep future surveys will be the detection of features which are invisible in surveys such as SDSS or DES (e.g. tidal features and debris). Machines trained on shallower data are unlikely to produce robust results on very deep images. We plan to carry out a thorough study in this respect using cosmological hydrodynamical simulations such as Horizon-AGN (Kaviraj et al. 2017) in a future work.

ACKNOWLEDGEMENTS

The authors would like to thank the anonymous referee for the useful comments which helped to improve the content of the paper. This work was supported in part by NSF AC1816330. Funding for the DES Projects has been provided by the DOE and NSF(USA), MEC/MICINN/MINECO(Spain), STFC(UK), HEFCE(UK), NCSA(UIUC), KICP(U. Chicago), CCAPP(Ohio State), MIFPA(Texas A&M), CNPQ, FAPERJ, FINEP (Brazil), DFG(Germany), and the Collaborating Institutions in the Dark Energy Survey. The Collaborating Institutions are Argonne Lab, UC Santa Cruz, University of Cambridge, CIEMAT-Madrid, University...
Transferring learning for galaxy morphology

REFERENCES


1 Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA

2 LERMA, Observatoire de Paris, PSL Research University, CNRS, Sorbonne Universités, UPMC Univ. Paris 06, F-75014 Paris, France

3 University of Paris Denis Diderot, University of Paris Sorbonne Cité (PSC), F-75205 Paris Cedex 13, France

4 Instituto de Astrofísica de Canarias, E-38200 La Laguna, Tenerife, Spain

5 Departamento de Astrofísica, Universidad de La Laguna, E-38206 La Laguna, Tenerife, Spain

6 Centre for Astrophysics Research, University of Hertfordshire, College Lane, Hatfield, Herts AL10 9AB, UK

7 Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory, Casilla 603, La Serena, Chile

8 Department of Physics & Astronomy, University College London, Gower Street, London WC1E 6BT, UK

9 Department of Physics and Electronics, Rhodes University, PO Box 94, Grahamstown 6140, South Africa

10 Fermi National Accelerator Laboratory, PO Box 500, Batavia, IL 60510, USA

11 Institute of Cosmology & Gravitation, University of Portsmouth, Portsmouth PO1 3FX, UK

12 Laboratorio Interinstitucional de e-Astronomía - LIneA, Rua Gal. José Cristino 77, Rio de Janeiro, RJ 20921-400, Brazil

13 Observatório Nacional, Rua Gal. José Cristino 77, Rio de Janeiro, RJ 20921-400, Brazil

14 Department of Astronomy, University of Illinois at Urbana-Champaign, 1002 W. Green Street, Urbana, IL 61801, USA

15 National Center for Supercomputing Applications, 1205 West Clark St., Urbana, IL 61801, USA
This paper has been typeset from a \TeX/\LaTeX file prepared by the author.