Transfer learning for galaxy morphology from one survey to another

H. Domínguez Sánchez[1]⋆, M. Huertas-Company[1,2,3,4,5], M. Bernardi[1], S. Kaviraj[6], J.L. Fischer[1],
T. M. C. Abbott[7], F. B. Abdalla[8,9], J. Annis[10], S. Avila[11], D. Brooks[8], E. Buckley-Geer[10],
A. Carnero Rosell[12,13], M. Carrasco Kind[14,15], J. Carretero[16], C. E. Cunha[17], C. B. D'Andrea[1],
L. N. da Costa[12,13], C. Davis[17], J. De Vicente[18], P. Doel[8], A. E. Evrard[19,20], P. Fosalba[21,22],
J. Frieman[10,23], J. García-Bellido[24], E. Gaztanaga[21,22], D. W. Gerdes[21,22], D. Gruen[17,25],
R. A. Gruendl[14,15], J. Gschwend[12,13], G. Gutierrez[10], W. G. Hartley[8,26], D. L. Hollowood[27],
K. Honscheid[28,29], B. Hoyle[30,31], D. J. James[32], K. Kuehn[33], N. Kuropatkin[10], O. Lahav[8],
M. A. G. Maia[12,13], M. March[1], P. Melchior[34], F. Menanteau[14,15], R. Miquel[16,35], B. Nord[10],
A. A. Plazas[36], E. Sanchez[18], V. Scarpine[10], R. Schindler[25], M. Schubnell[20], M. Smith[37],
R. C. Smith[7], M. Soares-Santos[38], F. Sobreira[39,12], E. Suchyta[40], M. E. C. Swanson[15], G. Tarle[20],
D. Thomas[11], A. R. Walker[7], and J. Zuntz[41]

18 January 2019   *Author affiliations are listed at the end of this paper*

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 labeled 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 dataset, 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 5000 galaxies with a similar redshift distribution to SDSS. Applying the models directly to DES data provides a reasonable global accuracy ($\sim 90\%$), but small completeness and purity values. A fast domain adaptation step, consisting in a further training with a small DES sample of galaxies ($\sim$500-300), is enough for obtaining an accuracy $> 95\%$ and a significant improvement in the completeness and purity values. This demonstrates that, once trained with a particular dataset, 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: galaxies: structure – methods: observational – surveys

1 INTRODUCTION

Astronomy is entering the Big Data era. We are experiencing a revolution in terms of available data thanks to surveys such as COSMOS (Scoville et al. 2007), SDSS (Eisenstein et al. 2011), DEEP2 (Newman et al. 2013), DES (DES Collaboration et al. 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 closely 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; Kaviraj 2014; Bournaud et al. 2014; Belfiore et al. 2015; Dubois et al. 2016). It is therefore crucial to have accurate morphological classifications for large samples of galaxies.

⋆ Corresponding author: helenado@sas.upenn.edu

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[1], a Kaggle competition for reproducing the Galaxy Zoo 2, where the top three algorithms used CNNs (e.g. Dieleman et al. 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 5000-10000 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 dataset, i.e., can the features learned by an unsupervised process on a

given dataset be *transferred* to a new dataset 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 dataset)?

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., Shermin et al. 2018; Guo et al. 2018; Kornblith 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 the ability of DL models to *transfer knowledge* for different datasets. 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 labeled images needed for training.

The paper is organised 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 summarise 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 et al. (2015, 2016)
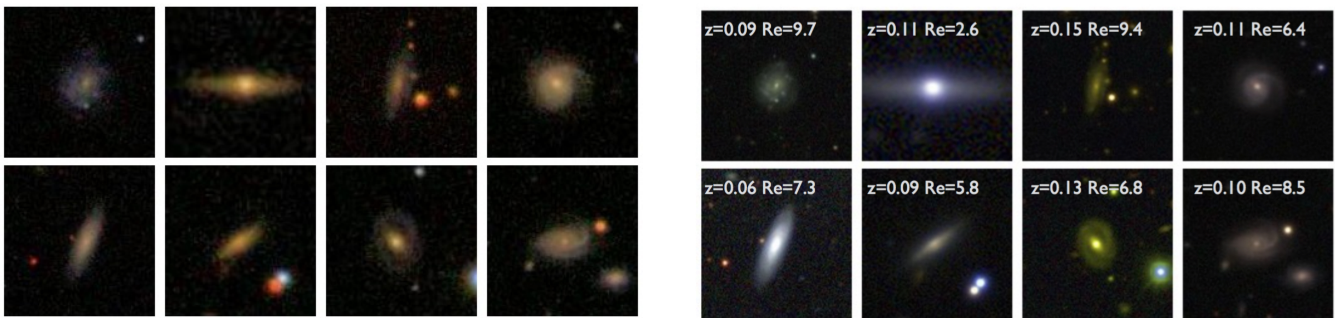
---

[1] https://www.kaggle.com/c/galaxy-zoo-the-galaxy-challenge

Figure 1. Examples of 6 galaxies observed by SDSS-DR7 (left panels) and DES survey (right panels). The cutouts are zoomed in to 1/2 of the size of the images used for training the models. They have a variable angular size of approximately $5 \times R_{90}$, where R_{90} is the Petrosian radius of each galaxy (shown in each cutout - in arcsec -, as well as their redshift). The galaxies are randomly selected from the common sample of the two surveys, with the only requirement of having high probability of being disk, 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.

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 color SDSS-DR7 images. We obtain T-Types and a series of GZ2 type questions (disk/features, edge-on galaxies, bar signature, bulge prominence, roundness and mergers) for a sample of galaxies with r-band Petrosian magnitude limits $14 \leq m_r \leq 17.77$ mag and z < 0.25. The SDSS images are the standard cutouts downloaded from the SDSS DR7 server[2], with a resolution of 0.396 ″/pixel.

2.2 Image data: Dark Energy Survey

The images used to test how DL models can adapt to new surveys characteristics come from the Dark Energy Survey (DES; DES Collaboration et al. 2016). DES is an international, collaborative effort designed to probe the origin of the accelerating universe and the nature of dark energy by measuring almost the 14-billion-year history of cosmic expansion with high precision. The survey will map ∼ 300 million galaxies. This huge number demands to find automated methods for morphological classification of galaxies.

DES 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 deg^2 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 ″/pixel. The magnitude limits and median PSF 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 cutouts from the internal Y1A1 data release.

2.3 Morphological catalogue: Dark Energy Camera Legacy Survey

Unfortunately, there is no morphological classification available for DES galaxies to date. Instead, we take advantage of the Galaxy Zoo Dark Energy Camera Legacy Survey (DE-CaLS) morphological catalogue to assign a classification for DES galaxies. This is necessary for quantifying the performance of the DL models, as well as for labeling the training sample in the fine-tunning or domain adaptation step (see section 3). The DECaLS survey (Dey et al. 2018) is observed with the same camera as the DES survey and with a similar depth (g=24.0, r=23.4, z=22.5 mag at 5σ level), and so (average) observing conditions are very similar to the DES ones. The DECaLS Galaxy Zoo catalogue (private communication) contains morphological classifications for ∼ 32,000 objects up to z ∼ 0.15. The redshift range and most of the classification tasks are the same as for the GZ2 catalogue, which was used for training the DL models from DS18. Therefore, it is the perfect catalogue to test the performance of the SDSS-based DL models on DES images. The main difference of DES/DECaLS with respect to SDSS images is the use of a larger telescope and better seeing conditions, which allow to get deeper images (∼ 1.5 mag) with significantly better data quality than SDSS. This effect can be seen in Figure 1, where we show 6 examples of galaxies as observed by SDSS and DES.

The DES sample used in this work are the 4,938 galaxies with a DECaLS - Galaxy Zoo classification (obtained with a match of 1 arcsec separation). Note that, since our final aim will be to provide a morphological catalogue for DES, we use the DECaLS classification catalogue as the ground truth to test (and train) our models on DES images. Given the similarities between DES and DECaLS surveys, the Galaxy Zoo classifications will be identical or very similar, which allows us to perform this exercise.

3 METHODOLOGY

The objective of this paper is to assess if knowledge acquired by a DL algorithm from an existing survey can be exported to a new dataset with different characteristics in terms of

---

[2] http://casjobs.sdss.org/ImgCutoutDR7

depth, PSF and 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 disks/features versus smooth galaxies. We consider as positive examples galaxies with disk 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 summarise the DL model architecture, which consists on 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 2,602,849 (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 $10 \times R_{90}$, where R_{90} 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-tunning* 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 the fully connected layer.
d) Training the models from scratch using a DES training sample with the same size as in (b) and (c).

Layer type	Output shape	Num. parameters
Conv2D (6x6)	(32, 69, 69)	3488
Dropout (0.5)	(32, 69, 69)	0
Conv2D (5x5)	(64, 69, 69)	51264
MaxPooling	(64, 34, 34)	0
Dropout (0.25)	(64, 34, 34)	0
Conv2D (2x2)	(128, 34, 34)	32896
MaxPooling	(128, 17, 17)	0
Dropout (0.25)	(128, 17, 17)	0
Conv2D (3x3)	(128, 17, 17)	147584
Dropout (0.25)	(128, 17, 17)	0
Flatten	(36992)	0
Dense	(64)	2367552
Dropout (0.5)	(64)	0
Dense	(1)	65
Total num. parameters		2602849

Table 1. DL model architecture. It consists of 4 convolution layers with different filter sizes (6, 5, 2 and 3, as shown in brackets) and one fully connected layer, also referred to as the dense layer. Dropout is performed after each convolutional layer (the reduction factor is shown in brakets) and MaxPooling is used after the second and third layers. The output shape and the number of free parameters in each layer are also shown.

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 datasets, not 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 uneven proportion of positive and negative examples for this two classes. We only use in the training DES galaxies with a robust classification, i.e. galaxies with a large agreement - *a(p)* - between Galaxy Zoo classifiers (roughly corresponding to P > 0.7 in one of the two answers) and with at least 5 votes. [Reader can refer to DS18 for a description of the agreement parameter, *a(p)*.] This methodology has demonstrated to be a more efficient way to train the models, but it strongly limits the statistics of our train and test samples. For example, only 624 out of 4938 galaxies (∼ 13%) have $P_{edge-on} > 0.7$ and at least 5 votes. This number is even smaller (103, ∼ 2%) for the barred galaxies.

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 9 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:

$$\text{TPR} = \frac{\text{TP}}{(\text{TP} + \text{FN})}; \qquad \text{FPR} = \frac{\text{FP}}{(\text{FP} + \text{TN})}$$

$$\text{P} = \frac{\text{TP}}{(\text{TP} + \text{FP})}; \qquad \text{Acc} = \frac{\text{TP} + \text{TN}}{\text{Total}} \tag{1}$$

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 form 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 Figure 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 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 $P_{th}= 0.5$ (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%), reaching 93% and 95% 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. This quantities are strongly dependent on the classification task. For example, for Q1 the precision value is very high (92%), but the completeness is less than 50%. On the other hand, the SDSS model recovers 91% of the DES edge-on galaxies, but the precision value for this task is 76%. 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 Figure 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% and the purity values are 91% and 86%, 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 8 out of 9 barred galaxies (TP) and there are only 3 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 3 FP cases have $P_{bar} > 0.6$ according to model (b), and that galaxy shows a bright central feature which could be a distorted bar or a dust lane (see Figure 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 4 convolutional layers), most of the trainable weights actually come from the fully connected layer (235232 vs 2367617 for the convolutional layer and the fully connected layer, respectively). Despite of this, the performance of the models after *fine-tunning* 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 applicable to many datasets 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 initialization. As expected, the performance of the models trained from scratch is worst than the performance of the models after *fine-tunning*. This demonstrates that using a 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% 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 as small sample is unable to learn and separate the features related to the presence of a bar,
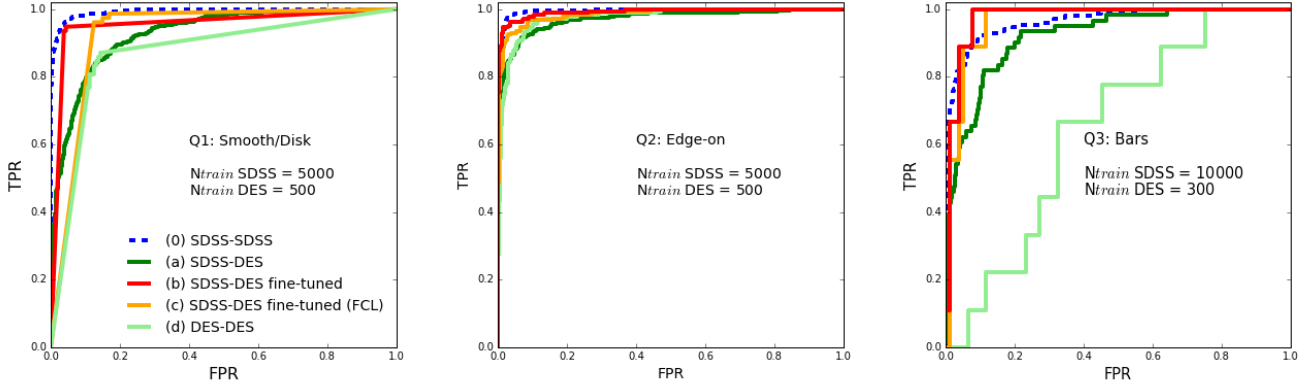
Figure 2. True positive rate (TPR, i.e., fraction of well classified positive cases) vs. false positive rate (FPR, i.e., fraction of wrongly classified positive cases) for different P_{th} values for the three classification task studied in this work, as stated in the legend. We show the performance of the DL models for the four experiments explained in section 3.2 (labeled (a) to (d), color-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-tunning* 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).

Question		Experiment	N$train$	N$test$	Npos	TPR	Prec.	Acc.
	(0)	SDSS-SDSS	5000	3370	674	0.93	0.91	0.97
Q1	(a)	SDSS-DES	0	2409	797	0.48	0.92	0.81
Smooth/Disk	(b)	SDSS-DES fine-tuned	500	238	78	0.95	0.91	0.95
	(c)	SDSS-DES fine-tuned (FCL)	500	238	78	0.96	0.78	0.90
	(d)	DES-DES	500	238	78	0.81	0.77	0.85
	(0)	SDSS-SDSS	5000	2687	396	0.98	0.80	0.96
Q2	(a)	SDSS-DES	0	2851	536	0.91	0.76	0.93
Edge-on	(b)	SDSS-DES fine-tuned	500	738	187	0.96	0.86	0.95
	(c)	SDSS-DES fine-tuned (FCL)	500	738	187	0.97	0.77	0.92
	(d)	DES-DES	500	238	78	0.97	0.70	0.89
	(0)	SDSS	10000	1806	169	0.76	0.79	0.96
Q3	(a)	DES	0	1768	61	0.57	0.35	0.95
Bar sign	(b)	SDSS-DES fine-tuned	300	86	9	0.89	0.73	0.95
	(c)	SDSS-DES fine-tuned (FCL)	300	86	9	1.0	0.5	0.89
	(d)	DES -DES	300	86	9	1.0	0.1	0.1

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$train$ is the number of galaxies used for training. When N$train$=0, it means the SDSS model is directly applied to DES data. N$test$ 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 Npos are the positive cases (e.g., galaxies showing disk/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.

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-tunning* (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 and TRP values are smaller. This is because the values in Table 2 are given for P_{th}=0.5, while the optimal performance for model (a) is obtained by setting P_{th}=0.1. This means that the knowledge transferred for different datasets needs to be *recalibrated*.
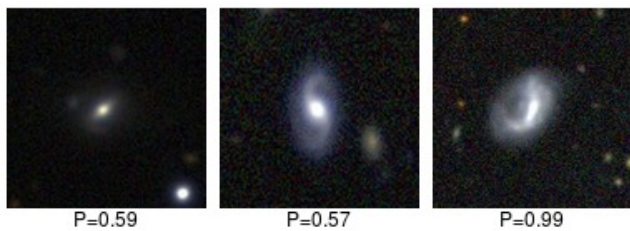
Figure 3. The 3 FP cases of barred galaxies, according to the DES-fine-tuned model (b). The numbers shown in the cutouts are the predicted P_{bar} given by our model. The cutouts are zoomed in to $\sim 1/2$ of the input to the models (approximately $5 \times R_{90}$). There is a clear bright central structure in all of them, which may be difficult to distinguish from a true bar, even for non-expert human classifiers.

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-tunning* 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 labeled galaxies) can be reduced by an order of magnitude, once the models are trained with a different dataset, 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 the 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 to this respect using cosmological hydro-dynamical simulations such as Horizon-AGN (Kaviraj et al. 2017) in a future work.

REFERENCES

Abazajian K. N., et al., 2009, ApJS, 182, 543
Ackermann S., Schawinski K., Zhang C., Weigel A. K., Turp M. D., 2018, MNRAS, 479, 415
Banerji M., et al., 2010, MNRAS, 406, 342
Barchi P. H., da Costa F. G., Sautter R., Moura T. C., Stalder D. H., Rosa R. R., de Carvalho R. R., 2017, preprint, (arXiv:1705.06818)
Belfiore F., et al., 2015, MNRAS, 449, 867
Bengio Y., 2012, in Unsupervised and Transfer Learning - Workshop held at ICML 2011, Bellevue, Washington, USA, July 2, 2011. pp 17–36, http://jmlr.csail.mit.edu/proceedings/papers/v27/bengio12a.html
Bournaud F., et al., 2014, ApJ, 780, 57
Conselice C. J., 2003, ApJS, 147, 1
DES Collaboration et al., 2016, MNRAS, 460, 1270
Deng J., Dong W., Socher R., Li L.-J., Li K., Fei-Fei L., 2009, in CVPR09.
Dey A., et al., 2018, preprint, (arXiv:1804.08657)
Dieleman S., Willett K. W., Dambre J., 2015, MNRAS, 450, 1441

Domínguez Sánchez H., Huertas-Company M., Bernardi M., Tuccillo D., Fischer J. L., 2018, MNRAS, 476, 3661
Drlica-Wagner A., et al., 2017, preprint, (arXiv:1708.01531)
Dubois Y., Peirani S., Pichon C., Devriendt J., Gavazzi R., Welker C., Volonteri M., 2016, MNRAS, 463, 3948
Eisenstein D. J., et al., 2011, AJ, 142, 72
Flaugher B., et al., 2015, AJ, 150, 150
Grogin N. A., et al., 2011, ApJS, 197, 35
Guo Y., Shi H., Kumar A., Grauman K., Rosing T., Feris R., 2018, preprint, (arXiv:1811.08737)
Huertas-Company M., et al., 2015, ApJS, 221, 8
Kaviraj S., 2014, MNRAS, 440, 2944
Kaviraj S., et al., 2017, MNRAS, 467, 4739
Kornblith S., Shlens J., Le Q. V., 2018, CoRR, abs/1805.08974
LSST Science Collaboration et al., 2017, preprint, (arXiv:1708.04058)
Lahav O., et al., 1995, Science, 267, 859
Lahav O., Naim A., Sodré Jr. L., Storrie-Lombardi M. C., 1996, MNRAS, 283, 207
Lintott C., et al., 2011, MNRAS, 410, 166
Meert A., Vikram V., Bernardi M., 2015, MNRAS, 446, 3943
Meert A., Vikram V., Bernardi M., 2016, MNRAS, 455, 2440
Nair P. B., Abraham R. G., 2010, ApJS, 186, 427
Newman J. A., et al., 2013, ApJS, 208, 5
Peng C. Y., Ho L. C., Impey C. D., Rix H.-W., 2002, AJ, 124, 266
Pérez-Carrasco M., Cabrera-Vives G., Martinez-Marín M., Cerulo P., Demarco R., Protopapas P., Godoy J., Huertas-Company M., 2018, preprint, (arXiv:1810.07857)
Postman M., et al., 2012, ApJS, 199, 25
Powers D. M. W., Ailab 2011.
Racca G. D., et al., 2016, in Space Telescopes and Instrumentation 2016: Optical, Infrared, and Millimeter Wave. p. 99040O (arXiv:1610.05508), doi:10.1117/12.2230762
Scoville N., et al., 2007, ApJS, 172, 1
Shermin T., Murshed M., Lu G., Teng S. W., 2018, preprint, (arXiv:1811.07459)
Simmons B., Lintott C., Masters K., Willett K., Kartaltepe J. S., Closson Ferguson H., Faber S. M., Galaxy Zoo Team C. T., 2016, in American Astronomical Society Meeting Abstracts. p. 342.42
Tajbakhsh N., Shin J. Y., Gurudu S. R., Hurst R. T., Kendall C. B., Gotway M. B., Liang J., 2016, IEEE Transactions on Medical Imaging, 35, 1299
Tuccillo D., Huertas-Company M., Decencière E., Velasco-Forero S., Domínguez Sánchez H., Dimauro P., 2017, preprint, (arXiv:1711.03108)
Willett K. W., et al., 2013, MNRAS, 435, 2835
Yosinski J., Clune J., Bengio Y., Lipson H., 2014, preprint, (arXiv:1411.1792)

AFFILIATIONS

[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), 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 Hertford-shire, 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, P. O. Box 500, Batavia, IL 60510, USA
[11] Institute of Cosmology & Gravitation, University of Portsmouth, Portsmouth, PO1 3FX, UK
[12] Laboratório Interinstitucional de e-Astronomia - 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
[16] Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, 08193 Bellaterra (Barcelona) Spain
[17] Kavli Institute for Particle Astrophysics & Cosmology, P. O. Box 2450, Stanford University, Stanford, CA 94305, USA
[18] Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain
[19] Department of Astronomy, University of Michigan, Ann Arbor, MI 48109, USA
[20] Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA
[21] Institut dÉstudis Espacials de Catalunya (IEEC), 08193 Barcelona, Spain
[22] Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, 08193 Barcelona, Spain
[23] Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA
[24] Instituto de Fisica Teorica UAM/CSIC, Universidad Autonoma de Madrid, 28049 Madrid, Spain
[25] SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA
[26] Department of Physics, ETH Zurich, Wolfgang-Pauli-Strasse 16, CH-8093 Zurich, Switzerland
[27] Santa Cruz Institute for Particle Physics, Santa Cruz, CA 95064, USA
[28] Center for Cosmology and Astro-Particle Physics, The Ohio State University, Columbus, OH 43210, USA
[29] Department of Physics, The Ohio State University, Columbus, OH 43210, USA
[30] Max Planck Institute for Extraterrestrial Physics, Giessenbachstrasse, 85748 Garching, Germany
[31] Universitäts-Sternwarte, Fakultät für Physik, Ludwig-Maximilians Universität München, Scheinerstr. 1, 81679 München, Germany
[32] Harvard-Smithsonian Center for Astrophysics, Cambridge, MA 02138, USA
[33] Australian Astronomical Observatory, North Ryde, NSW 2113, Australia
[34] Department of Astrophysical Sciences, Princeton Univer-

sity, Peyton Hall, Princeton, NJ 08544, USA

[35] Institució Catalana de Recerca i Estudis Avanccats, E-08010 Barcelona, Spain

[36] Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr., Pasadena, CA 91109, USA

[37] School of Physics and Astronomy, University of Southampton, Southampton, SO17 1BJ, UK

[38] Brandeis University, Physics Department, 415 South Street, Waltham MA 02453

[39] Instituto de Física Gleb Wataghin, Universidade Estadual de Campinas, 13083-859, Campinas, SP, Brazil

[40] Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831

[41] Institute for Astronomy, University of Edinburgh, Edinburgh EH9 3HJ, UK

This paper has been typeset from a TEX/LATEX file prepared by the author.