DEEP LEARNING AT SCALE FOR THE CONSTRUCTION OF GALAXY CATALOGS WITH THE DARK ENERGY SURVEY

Allocation: Illinois/600 Knh
PI: Eliu Huerta[1]
Co-PIs: Gabrielle Allen[1], Roland Haas[1], Ed Seidel[1], Zhizhen Zhao[1]
Collaborators: Asad Khan[1], Sibo Wang[1], Robert Gruendl[1], Elise Jennings[2], Huihuo Zheng[2]

[1]University of Illinois at Urbana–Champaign
[2]Argonne Leadership Computing Facility

EXECUTIVE SUMMARY

The scale of ongoing and future electromagnetic surveys poses formidable challenges to classifying astronomical objects. Pioneering efforts on this front include citizen science campaigns adopted by the Sloan Digital Sky Survey (SDSS). SDSS data sets have recently been used to train neural network models to classify galaxies in the Dark Energy Survey (DES) that overlap the footprint of both surveys. The research team has demonstrated that knowledge from deep learning algorithms, pretrained with real-object images, can be transferred to classify galaxies that overlap both SDSS and DES surveys, achieving a state-of-the-art accuracy of 99.6%. In addition, the team has demonstrated that this process can be completed within just eight minutes using distributed training. The researchers also used their neural network classifier to label 10,000 DES galaxies that do not overlap previous surveys. Further, the team has shown that these new data sets can be combined with recursive training to create DES galaxy catalogs in preparation for the Large Synoptic Survey Telescope era.

RESEARCH CHALLENGE

The classification of astrophysical objects has been pursued in the past using a diverse set of tools. For instance, galaxies have been classified using their photometric properties, achieving classification accuracies of around 85% [1]. Other methods of classifying galaxies according to their morphology have taken into account their physical properties across multiple wavelengths. For instance, the method introduced in [2] considered a sample of galaxies from the Sloan Digital Sky Survey (SDSS) [3] using the five SDSS filters (u, g, r, i, z) and then used a combination of shapelet decomposition and principal components analysis (PCA). Other methods for galaxy classification include Concentration–Asymmetry–Smoothness [4] and machine learning, including artificial neural networks and PCAs [5].

In recent years, citizen science campaigns have played a key role in classifying thousands of celestial objects in astronomical surveys. SDSS is an archetypical example of a successful approach to classifying hundreds of thousands of galaxies. As electromagnetic surveys continue to increase their depth and coverage, campaigns of this nature may lack scalability. For example, within six years of operation, the Dark Energy Survey (DES) [6] observed over three hundred million galaxies, a number that will be surpassed by the observing capabilities of the Large Synoptic Survey Telescope (LSST) [7]. In brief, there is a pressing need to explore new approaches to maximize the science throughput of next-generation electromagnetic surveys. A promising paradigm is the convergence of deep learning and large-scale computing to address the imminent increase in data volume, complexity, and latency of observations of LSST-type surveys.

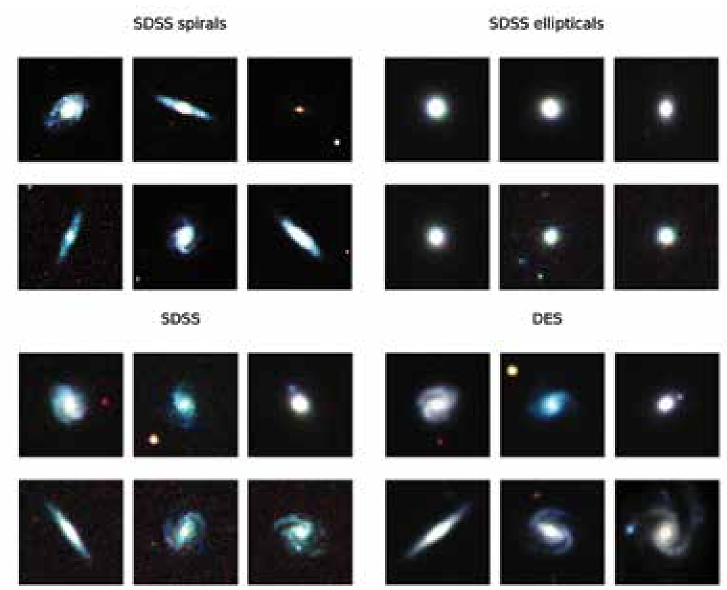

Figure 1: Top panels: Labeled images of the SDSS training data set. Bottom panels: Sample of galaxies from SDSS DR7 and the corresponding cross-matched galaxies from DES DR1.

Figure 2: t-SNE visualization of the clustering of HP SDSS and DES test sets, and unlabeled DES test set.

METHODS & CODES

The research team used a subset of SDSS Data Release (DR) 7 images for which it had high confidence classifications through the Galaxy Zoo project; *i.e.*, the team only chose galaxies with a debiased probability greater than 0.985 for combined spirals and 0.926 for ellipticals, respectively. Samples of these images are shown in Fig. 1. The team chose these cutoff thresholds to ensure that: (1) the galaxies used for training the neural network had robust and accurate classifications; and (2) the representation of both classes in the training and test data sets were balanced. The team then divided these images into three separate data sets for training, validation, and testing. The validation set was used to monitor the accuracy and loss when training and fine-tuning the deep neural network, and hence served to optimize hyperparameters such as learning rate and number of epochs for training.

Two test sets were carefully constructed so that the images in each set lay in both the SDSS and DES footprints. The first test set consisted of images with a Galaxy Zoo classification confidence similar to that of the training set, *i.e.*, a high probability cut-off was introduced. This test set was hence labeled High Probability (HP) Test Set, and there were two versions, one for each survey: HP SDSS and HP DES. Just as in the training set, the images for SDSS were obtained from SDSS DR7 and the corresponding images for DES were obtained from the DES DR1 data release. Furthermore, a second test set was created without introducing any probability thresholds on the Galaxy Zoo classification confidence. This set consisted of almost all galaxies lying in both the SDSS and DES footprints, and hence was labeled Full Overlap (FO) Test Set. Again, there were two versions: FO SDSS and FO DES. The motivation behind creating this second test set was that the galaxy profiles in the unlabeled DES data set would more closely match those in the FO test sets. Hence, the FO test set served as a good evaluation metric of the performance of the neural net on the ultimate task of classifying all unlabeled galaxies in the DES catalogue.

The team used open-source software stacks for its studies. The deep learning APIs used were Keras [8] and TensorFlow [9]. For the classification problem, the team did transfer learning starting with the Xception model [10], which had been pretrained with the ImageNet data set [11]. The researchers chose this neural network model because it outperforms many other state-of-the-art neural network models, including Inception-v3, ResNet-152, and VGG16, on the ImageNet validation data set, and it has been suggested that better ImageNet architectures are capable of learning better transferable representations [12]. More importantly, the research team carried out several experiments and found that Xception performed as well as or nominally better on the validation and testing galaxy data sets compared to many other state-of-the-art architectures.

RESULTS & IMPACT

This is the first application of deep transfer learning combined with distributed training for the classification of DES galaxies that overlap the footprint of the SDSS survey, achieving state-of-the-art accuracies of 99.6%. The research team has also used its neural network classifier to label over 10,000 DES galaxies that had not been observed in previous surveys. Using t-SNE visualizations (see [13] and Fig. 2), the research group found that deep transfer learning was effective to abstract morphological information from the galaxy images to clearly identify two distinct classes of galaxies in the unlabeled DES data set.

WHY BLUE WATERS

Blue Waters was essential to extract and curate the data sets used to train, validate, and test the team's neural network models at scale. Staff provided support to deploy the software stacks, to use them at scale, and to carry out all the analyses reported in this study.

PUBLICATIONS & DATA SETS

A. Khan, E. A. Huerta, S. Wang, R. Gruendl, E. Jennings, and H. Zheng, "Deep learning at scale for the construction of galaxy catalogs with the Dark Energy Survey," *Phys. Lett. B*, vol. 795, pp. 248–258, 2019, doi: 10.1016/j.physletb.2019.06.009.