Galaxy detection and identification using deep learning and data augmentation

González, R. E., Munoz, R. P., & Hernández, C. A. (2018). Galaxy detection and identification using deep learning and data augmentation. *Astronomy and computing*, *25*, 103-109. <10.1016/j.ascom.2018.09.004> Accessed 10 Nov 2020.

Abstract

We present a method for automatic detection and classification of galaxies which includes a novel data-augmentation procedure to make trained models more robust against the data taken from different instruments and contrast-stretching functions. This method is shown as part of AstroCV, a growing open source computer vision repository for processing and analyzing big astronomical datasets, including high performance Python and C++ algorithms used in the areas of image processing and computer vision.

The underlying models were trained using convolutional neural networks and deep learning techniques, which provide better results than methods based on manual feature engineering and SVMs in most of the cases where training datasets are large. The detection and classification methods were trained end-to-end using public datasets such as the Sloan Digital Sky Survey (SDSS), the Galaxy Zoo, and private datasets such as the Next Generation Virgo (NGVS) and Fornax (NGFS) surveys.

Training results are strongly bound to the conversion method from raw FITS data for each band into a 3-channel color image. Therefore, we propose data augmentation for the training using 5 conversion methods. This greatly improves the overall galaxy detection and classification for images produced from different instruments, bands and data reduction procedures.

The detection and classification methods were trained using the deep learning framework DARKNET and the real-time object detection system YOLO. These methods are implemented in C language and CUDA platform, and makes intensive use of graphical processing units (GPU). Using a single high-end Nvidia GPU card, it can process a SDSS image in 50 ms and a DECam image in less than 3 s.

We provide the open source code, documentation, pre-trained networks, python tutorials, and how to train your own datasets, which can be found in the AstroCV repository. https://github.com/astroCV/astroCV.

Keywords

Galaxies, General, Techniques, Image processing, Computing methodologies, Machine learning