

Using Image Processing and Deep Learning to Detect Dwarf Galaxies in NGC 5128

Cameron Leahy
University of Calgary.

cameron.leahy@ucalgary.ca

Abstract: Detecting dwarf galaxies is an important task in the field of observational astronomy. However, since many dwarf galaxies are extremely faint, traditional astronomical object detection software frequently fails to identify them in science imaging. To detect faint dwarf galaxy candidates in our image data, we implement a novel algorithm that uses image processing techniques and a convolutional neural network (CNN). Our imaging comes from the Dark Energy Camera (DECam) and features the region surrounding the giant elliptical galaxy NGC 5128, which is known to host a large population of dwarf galaxies. Using our algorithm, we retrieve 155 promising dwarf galaxy candidates in the imaging; future analyses and follow-up studies will determine whether they represent true dwarf galaxy satellites of the NGC 5128 system.

Introduction

Ever since the discovery of other galaxies beyond our own Milky Way (Hubble, 1925), astronomers have sought to understand the vast population of galaxies in the cosmos. Innovations such as the Schmidt camera enabled the first major wide-field galaxy surveys to be conducted in the 1950s (Minkowski & Abell, 1963), leading to the discovery of numerous small, faint galaxies in the nearby universe. Labeled as ‘dwarf galaxies’ (van den Bergh, 1959), these small galaxies were initially logged but not considered to be especially interesting until observations in the early 1980s, such as of the Draco dwarf (Aaronson, 1983), revealed that their mass budgets were dominated by dark matter. Today, dwarf galaxies are regarded as important probes of the hierarchical structure formation process outlined in our current theory of the evolution of the universe, the Lambda Cold Dark Matter (Λ CDM) paradigm (Bullock & Boylan-Kolchin, 2017; Navarro, 2018). Various discrepancies between dwarf galaxy observations and Λ CDM-based cosmological simulations have arisen in the last three decades, making further dwarf galaxy detection imperative for refining our understanding of structure formation in the universe (Sales et al., 2022).

There is ongoing work to detect faint dwarf galaxies in various regions of the universe (Eigenthaler et al., 2018; Ferrarese et al., 2020). In 2014, the Dark Energy Camera (Flaugher et al., 2015) at the Cerro Tololo Inter-American Observatory in Chile was used to image the giant elliptical galaxy NGC 5128 (also known as Centaurus A), resulting in the Survey of Centaurus A’s Baryonic Structures (SCABS; Taylor et al., 2016, 2017). The survey captures 72 square degrees of sky centered on NGC 5128 in the Sloan Digital Sky Survey (SDSS) optical filters $u'g'r'i'z'$.

In a section of the imaging spanning ~ 3 square degrees, 16 dwarf galaxy candidates were found using a by-eye search method (Taylor et al., 2018) (see [Figure 1](#)). This approach is common in faint dwarf galaxy detection (e.g., Eigenthaler et al.,

2018), since standard astronomical object detection software often misses dwarf galaxy profiles in imaging. However, the process of manually scanning the imaging is slow, often involves multiple researchers, and is not scalable to large datasets such as SCABS. Therefore, we turn to creating our own dwarf galaxy detection algorithm, which uses image processing and machine learning to detect faint dwarf galaxy candidates in the SCABS image data.

Methods

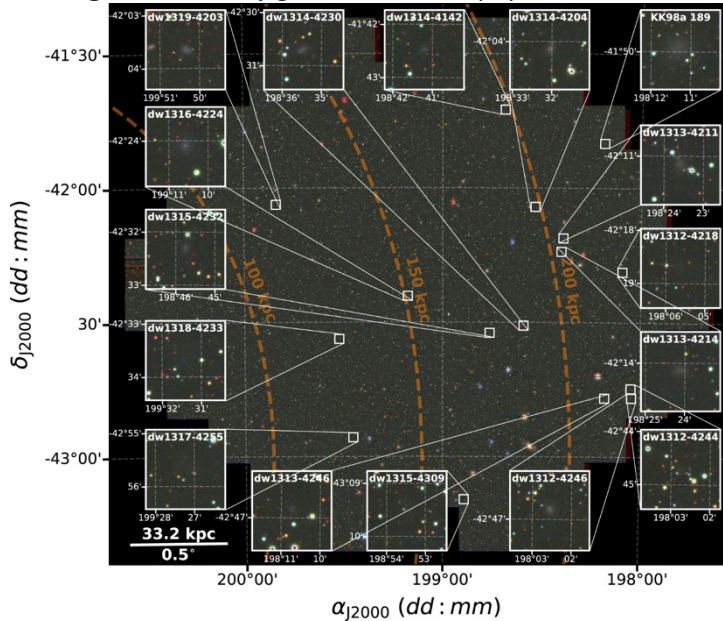
Algorithm design

A flowchart of the dwarf galaxy detection algorithm is shown in [Figure 2](#). The user first inputs a science image to be analyzed. Various image processing operations are performed to expose faint, diffuse structures, including dwarf galaxies. First, operations such as high-pass filtering, Sobel edge detection, and binary dilation are used to create a mask that encompasses stars and other bright objects. The pixel values in the mask are set to the median background value, effectively removing these bright, non-dwarf objects from the image. Then, a median filtering operation is performed, amplifying extended, faint structures in the image while suppressing any remaining point sources. After image processing, the detection software Source Extractor (SE; Bertin, 2006) is applied to the transformed image. Diffuse objects highlighted by the image processing are detected by SE, and their coordinates in the image are output to a catalog file. Together, the image processing routine and SE allow diffuse objects such as dwarf galaxies to be pinpointed in the original input image.

Since many other objects besides dwarf galaxies, such as diffraction spikes, large background galaxies, the faint outer wings of large stars, and image artefacts, are also faint and diffuse, the objects listed in the catalog must be filtered to remove non-dwarf objects. To accomplish this, we use a convolutional neural network (CNN) that processes cutout

images of SE-detected objects and classifies them as dwarfs or non-dwarfs based on their morphology. In the algorithm, small cutout images are created from the original image at the coordinates listed in the catalog and are routed to the CNN for inference (processing). The CNN outputs a much smaller, filtered set of cutouts, which are much more likely to contain dwarf galaxies. These cutouts are manually inspected to remove any remaining non-dwarf objects; the result is a final set of cutouts containing the most promising dwarf galaxy candidates in the imaging.

Figure 1: 16 dwarf galaxies detected by eye in SCABS



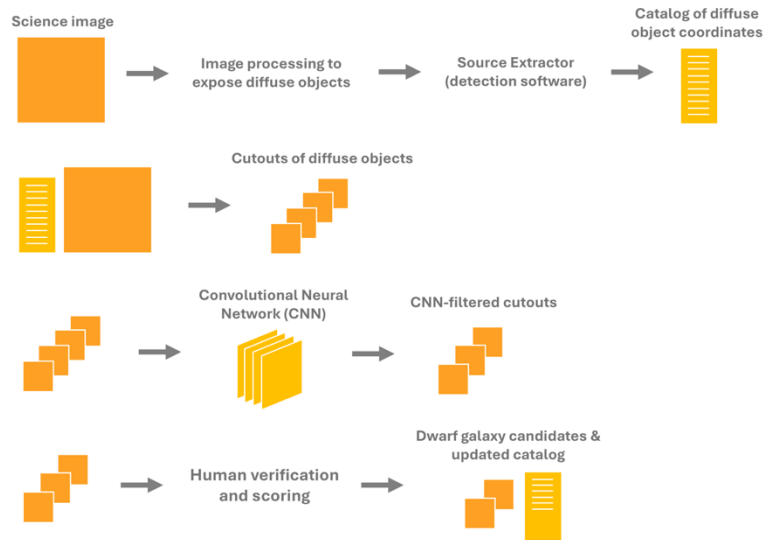
Dwarf galaxy candidates detected by eye in a small section of the SCABS data. Postage stamp images featuring magnified views of the candidates are displayed in the white boxes surrounding the main image. The horizontal and vertical axes give the $J2000$ right ascension (α) and declination (δ) sky coordinates of the image center.

CNN training procedure and architecture

Prior to its use in the algorithm, the CNN is trained on synthetic data to enable it to learn feature representations of dwarf- and non-dwarf objects. Image cutouts of dwarf galaxies are synthesized using a custom utility which can inject dwarf galaxy light profiles into ‘blank’ regions of the SCABS data and make cutouts of the injection locations; non-dwarf cutouts are produced in a similar way, but without artificial dwarf galaxy injection. The light profiles of many dwarf galaxies can be approximated with a Sérsic light distribution (Sérsic, 1963), where $I(r)$ is the light intensity at radius r from the center of the galaxy. The profile is parameterized by the effective intensity I_{eff} , effective radius r_{eff} , and the Sérsic index n , the axis ratio q , and the position angle θ . Equation (1) is used to simulate realistic light profiles of dwarf galaxies and create training data for the CNN.

$$I(r) = I_{\text{eff}} \exp \left\{ -b_n \left[\left(\frac{r(q, \theta)}{r_{\text{eff}}} \right)^{\frac{1}{n}} - 1 \right] \right\}, \quad (1)$$

Figure 2: Schematic of the algorithm



Schematic showing the steps in the dwarf galaxy detection algorithm.

Our CNN is trained on approximately 185,000 cutouts of synthetic dwarf galaxies and non-dwarf objects, split among the training, validation, and test sets. We implement a custom CNN architecture tailored to the extraction of features relevant for dwarf galaxy detection (see Figure 3). Four parallel branches of convolutional blocks, each employing dilated kernels with different sparsities, promote the extraction of extended spatial information at different scales. Pooling layers are omitted to avoid excessive early downsampling, enabling the retention of diffuse information. Downsampling is accomplished later in the network with a cropping layer, which also serves as an attention mechanism, emphasizing information in the center of cutouts, where dwarf galaxies, if present, are most likely to reside, due to how the cutouts are created.

We train our CNN for 14 epochs on our training set, achieving a binary accuracy of 95.86% and a recall of 99.50% (see Table 1). A decision threshold $\tau = 0.1$ is used, resulting in a higher recall (absence of false negatives, or missed dwarf galaxy detections) and a lower precision (the absence of false positives, or non-dwarfs that get labelled as dwarfs). This tradeoff is favourable, as missed detections are more deleterious than false positives, which can be eliminated in the manual verification step that follows CNN inference.

Results and conclusions

Running the algorithm on the entire SCABS survey, we detect 157 promising objects. Cross-referencing these objects with the NASA/IPAC Extragalactic Database (Mazzarella & NED Team, 2007) and other dwarf galaxy catalogs of NGC 5128 (Karachentsev et al., 2013; Crnojević et al., 2016; and Müller et al., 2017), we find that two of the objects are background galaxies, while eight represent previously detected dwarf satellites of NGC 5128.

