Draft version November 12, 2015
Preprint typeset using LaTeX style emulateapj v. 01/23/15

PROJECT MILESTONE: SEARCHING FOR ARTIFACTS IN IMAGES FROM THE DARK ENERGY SURVEY

Joseph W. DeRose and Warren R. Morningstar
Kavli Institute for Particle Astrophysics and Cosmology & Physics Department, Stanford University, Stanford, CA 94305, USA
Draft version November 12, 2015

ABSTRACT

In this paper, we present an implementation of a Machine Learning algorithm to be used in identifying and classifying image artifacts in observations taken by the Dark Energy Survey (DES). Specifically, we treat our individual pixels as features, and run a classifier using a Support Vector Machine. We find that our first implementation identifies image artifacts with an accuracy of Some Percent. We discuss future improvements to our modeling, and sources of systematic errors in our analysis, and how these errors may be overcome.

1. INTRODUCTION

In modern cosmology, much attention has been turned toward using large photometric sky surveys to provide strong constraints on structure formation and the composition of our universe. An important component of modern sky surveys is the management of the large volumes of data produced by the survey instruments. Typical data generation of current sky surveys is several GB of data per night, and future sky surveys will produce TB of data in the same time frame. Because of the vast quantities of data produced, the majority of analysis is performed in an automated (or at least semi-automated) manner.

While this provides significant advantages in performance, an obvious disadvantage is that it is prone to overlooking subtle features in the data, which may cause systematic errors in measurements of important quantities. In particular, since the goal of many photometric surveys is to provide precise measurements of the brightness of astronomical objects, any spurious brightness variations within a detector has the potential to be interpreted as a real signal, and thus has the capability of interfering with the accuracy of measurements.

Brightness variations in a detector can be caused in many different ways, and thus have a diverse variety of appearances. A few notable examples that have been observed by ongoing sky surveys include airplanes and satellites passing across the field of view, cosmic rays striking the detector, and improper subtraction of the background noise. Some of these are detected and masked away by the survey data reduction pipelines, but a large number are missed (these are usually referred to as artifacts). Reliable identification of artifacts is an essential component of preparing data for analysis. However, currently much of the identification of artifacts is carried out by members of survey collaborations, a process that uses a significant number of human hours and that could be better spent devoted to more scientifically interesting problems. Therefore an automated means of identifying and classifying image artifacts is desirable.

The Dark Energy Survey (DES) is a sky survey being conducted across many institutions. Survey is being carried out using the 4m Blanco Telescope located at the Cerro Tololo Inter-American Observatory in Chile. The survey saw first light in 2012, and is currently in its 3rd year. In this paper, we implement an algorithm for identification of image artifacts found in observations performed by DES. We describe our data set in Section 2. In Section 3, we describe our choice of features and the implementation of our model. In Section 4, we present the results of preliminary classification runs. In Section 5, we discuss these results, including ways in which our modeling could be improved.

2. DATA

Describe our data set. Include any undersampling of images we used, and how we intend to classify images. Maybe a sample image or two, although I dunno if we're allowed to do that on account of the data is private.

3. METHODS

SVM implementation. Feature selection. Any struggles we had along the way to show the volume of work we've done

4. RESULTS

How good did the SVM do? A few preliminary quantitative results would be useful (e.g. test error for the SVM). Maybe a plot or two showing some key feature that distinguishes artifacts from non-artifacts. Not sure if there will even be such a thing. Maybe cool to include a plot of a population of successfully identified artifacts

5. DISCUSSION

Probably most of the milestone report to go here. Discuss limitations of SVM for the problem we're trying to accomplish. Say what we're gonna do moving forward. If no good results to show, explain what is going wrong currently.