A new method for computational cosmological data analysis

Pale Blue Dot Team:
Alyssa Funk
Farnaz Golnaraghi
Thomas Thayer
Jose Zamora

Advisory Board: Dr. David Brown, Dr. Peter Nugent
Faculty Mentor: Dr. Juan Meza

April 28, 2017
Background

- **Automated**: minimal human interference
- **Transient identification**: find short-lived cosmological phenomena (like supernovae)
DES-SN transient detection pipeline

Generate single-epoch transient search images → Mask cosmic rays + bad pixels, detrend → Perform astrometry for each CCD (scamp)

Subtract templates from new images (hotpants) → Transform templates to align with new images (swarp) → Inject fake SNe Ia to monitor efficiency

Extract objects from difference images (SExtractor) → Make selection cuts on extracted objects (see Table 1) → Identify artifacts with machine learning algorithm (autoScan)

Select targets for spectroscopy → Visually scan candidates with at least two nights of non-artifact detections → Spatially associate objects into transient candidates
DES-SN transient detection pipeline

1. Generate single-epoch transient search images
2. Mask cosmic rays + bad pixels, detrend
3. Perform astrometry for each CCD
   (scamp)
4. Subtract templates from new images
   (hotpants)
5. Transform templates to align with new images
   (swarp)
6. Inject fake SNe Ia to monitor efficiency
7. Extract objects from difference images
   (SExtractor)
8. Make selection cuts on extracted objects
   (see Table 1)
9. Identify artifacts with machine learning algorithm
   (autoScan)
10. Select targets for spectroscopy
11. Visually scan candidates with at least two nights of non-artifact detections
12. Spatially associate objects into transient candidates
DES-SN transient detection pipeline

1. Extract objects from difference images (SExtractor)
2. Make selection cuts on extracted objects (see Table 1)
3. Identify artifacts with machine learning algorithm (autoScan)
4. Select targets for spectroscopy
5. Visually scan candidates with at least two nights of non-artifact detections
6. Spatially associate objects into transient candidates
Autoscan Disadvantages

- Dependencies are proprietary
- Designed for specific dataset
Pale Blue Dot workflow

1. Autoscan
   - CSV file
   - Bash script
   - unorganized.jpegs
   - Bash script
   - organized.jpegs
   - Caffe
   - Object or Artifact prediction

2. Autoscan
   - CSV file
   - Python
   - Simplified features file
   - SciKit
   - Object or Artifact prediction
Machine Learning

● “[Machine Learning is the] field of study that gives computers the ability to learn without being explicitly programmed” - Arthur Samuel, 1959

● Combination of multiple fields

● Basic problems include clustering, regression, and classification

● Multiple potential algorithms:
  ○ Deep Neural Networks, Random Forests
Deep Neural Networks

- Ensemble of many layers of neurons with a specified structure
- Benefits - Potentially very accurate, great for image data
- Caffe
  - Allows quick redefinition of neural net geometry
  - Fast computation using nvidia GPUs
  - Open source
Random Forests

- Ensemble of many random decision trees
- Benefits - less likely to overfit, works best with feature-based data
- Used by Autoscan to classify artifacts
- Scikit
  - Different ensemble methods
  - Open source
Results

<table>
<thead>
<tr>
<th>Caffe</th>
<th>SciKit</th>
</tr>
</thead>
<tbody>
<tr>
<td>67,000 training images</td>
<td>10,000 sample size</td>
</tr>
<tr>
<td>92% accuracy</td>
<td>95.6% accuracy</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>True Positive</th>
<th>False Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caffe</td>
<td>99.7%</td>
<td>0.3%</td>
</tr>
<tr>
<td>SciKit</td>
<td>82.4%</td>
<td>17.6%</td>
</tr>
</tbody>
</table>
Advantages for our methods

- Achieving > 90% accuracy
- No need for proprietary dependencies, e.g. Oracle
- Flexibility with data types
Future direction

- Find essential features for accurate prediction (SciKit)
- Test different machine learning algorithms on SciKit
- Ensemble of Caffe and SciKit
- Test our dual method on new data set
- Classification of objects
Thank you from Earth

Special thanks to:
David Brown, Ph.D
Peter Nugent, Ph.D
Juan Meza, Ph.D