A new method for computational cosmological data analysis

Pale Blue Dot Team:

Alyssa Funk

Farnaz Golnaraghi

Thomas Thayer

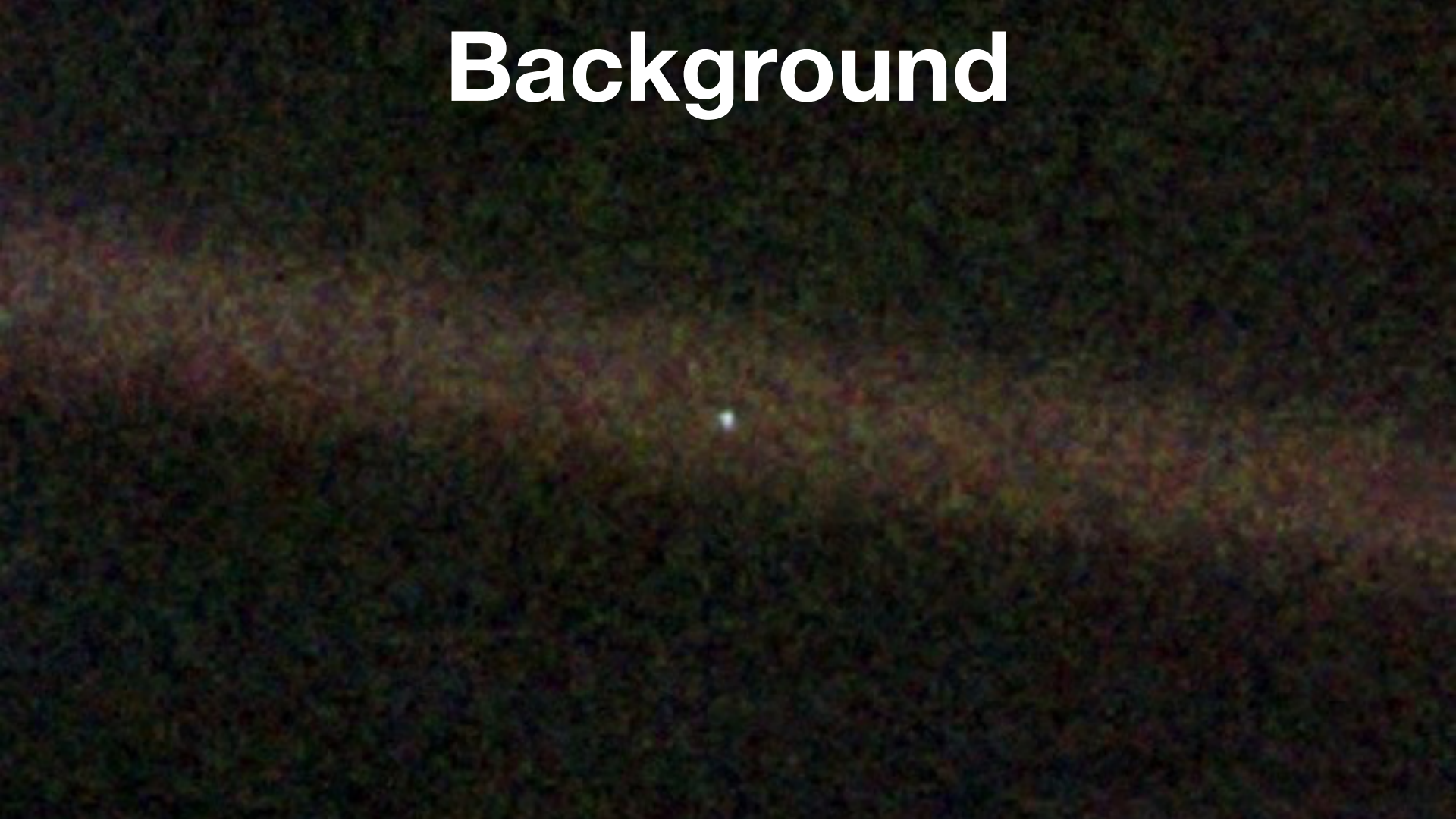
Jose Zamora

Advisory Board: Dr. David Brown, Dr. Peter Nugent

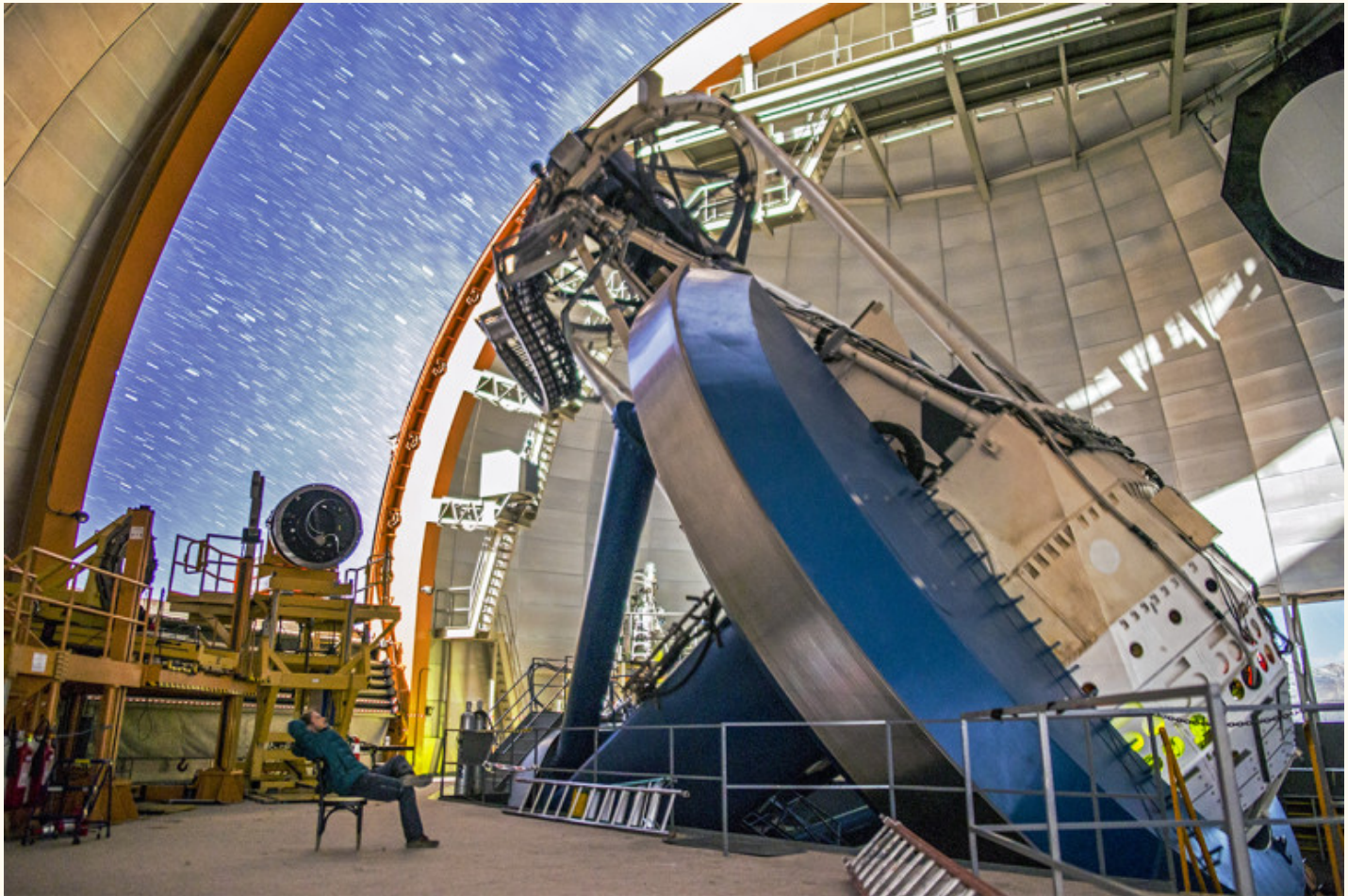
Faculty Mentor: Dr. Juan Meza

April 28, 2017

Background



Dark Energy Survey Telescope



Background

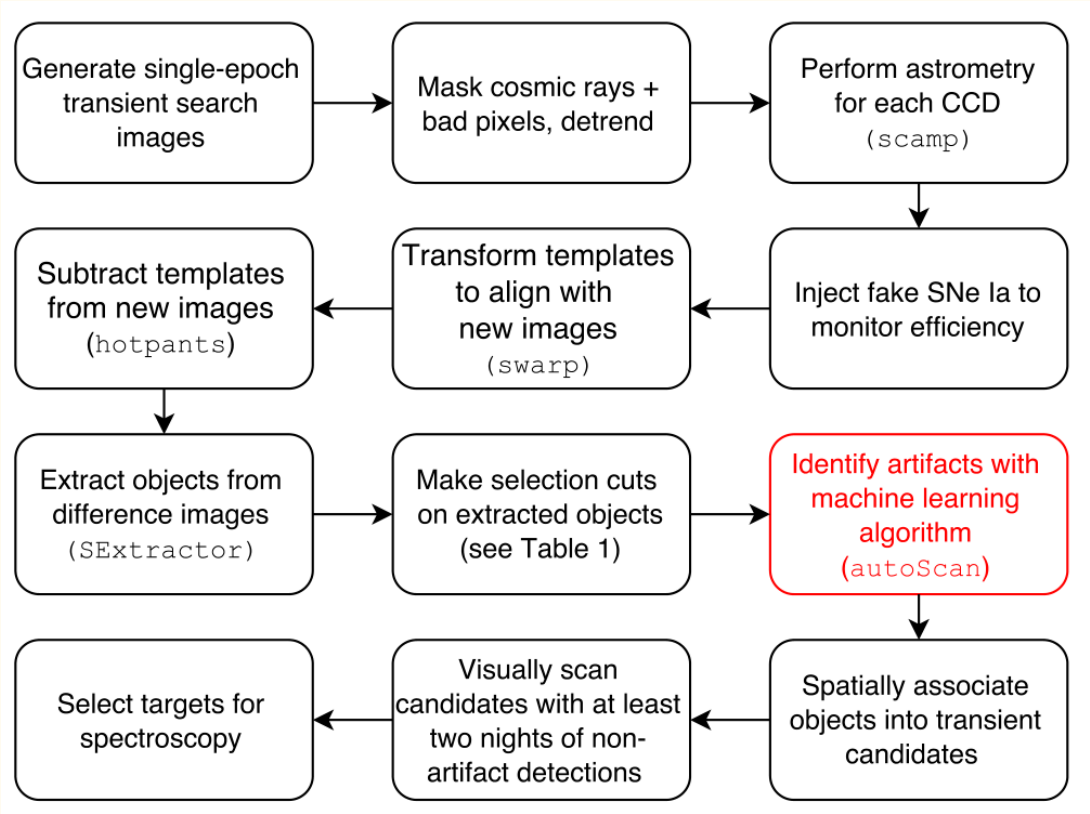
THE ASTRONOMICAL JOURNAL, 150:82 (15pp), 2015 September doi:10.1088/0004-6256/150/3/82

AUTOMATED TRANSIENT IDENTIFICATION IN THE DARK ENERGY SURVEY

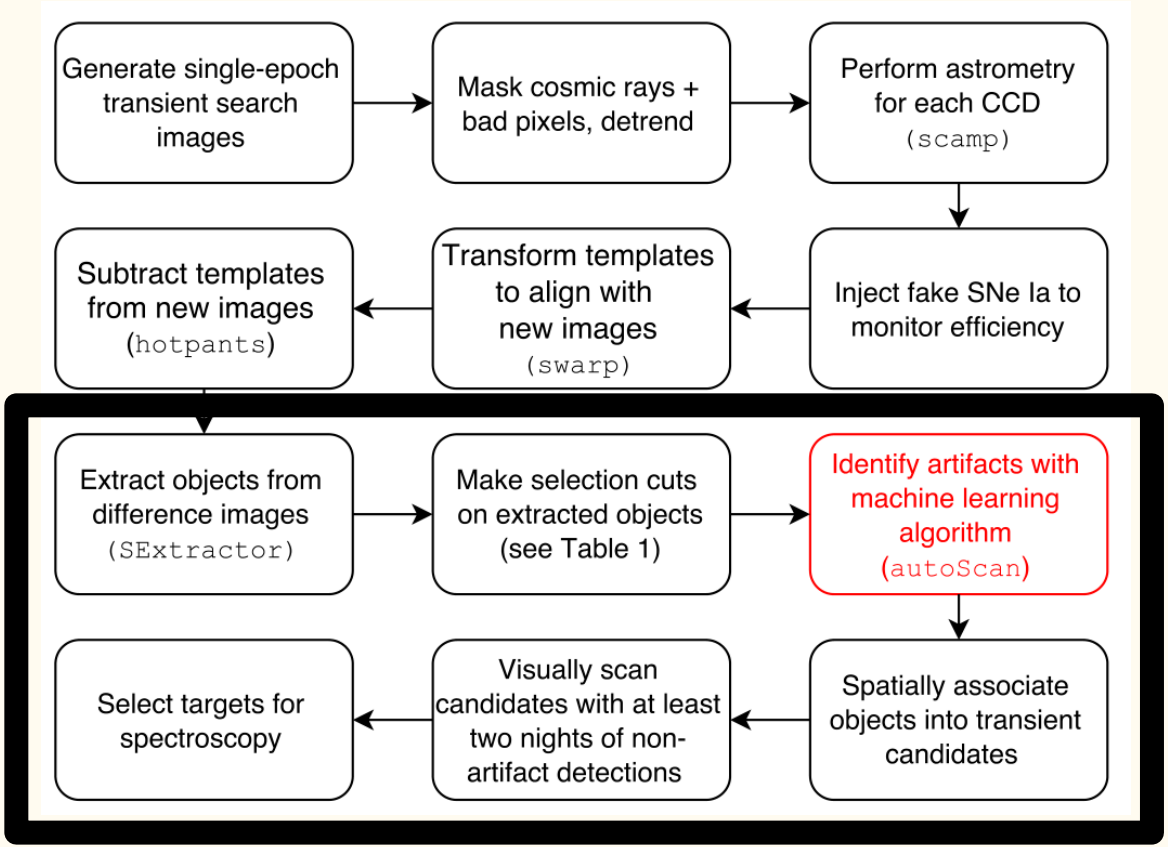
D. A. GOLDSTEIN[1,2], C. B. D'ANDREA[3], J. A. FISCHER[4], R. J. FOLEY[5,6], R. R. GUPTA[7], R. KESSLER[8,9], A. G. KIM[2], R. C. NICHOL[3],

- Automated: minimal human interference
- Transient identification: find short-lived cosmological phenomena (like supernovae)

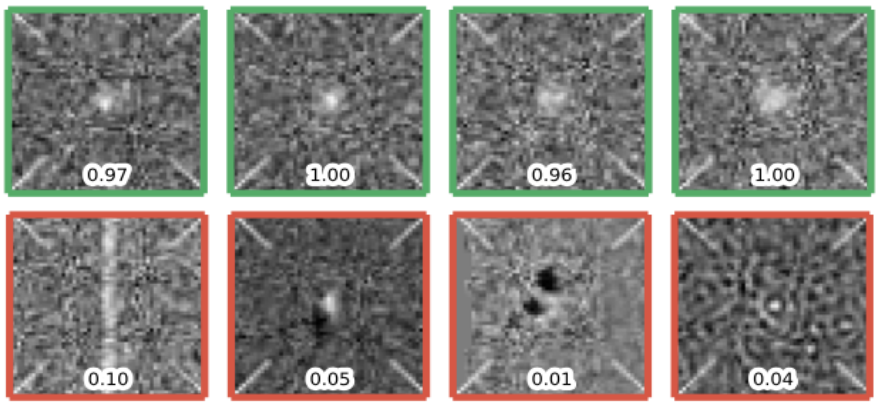
DES-SN transient detection pipeline

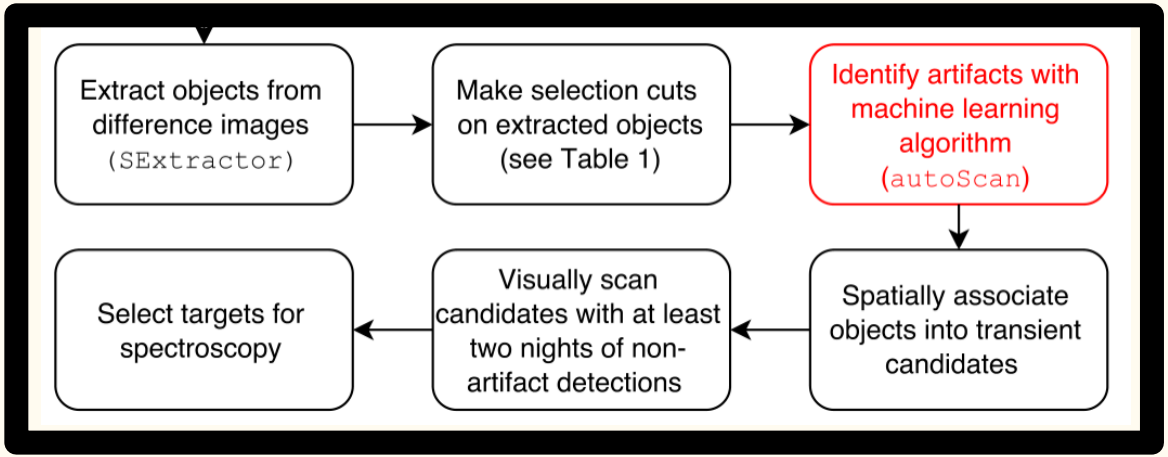


DES-SN transient detection pipeline



DES-SN transient detection pipeline

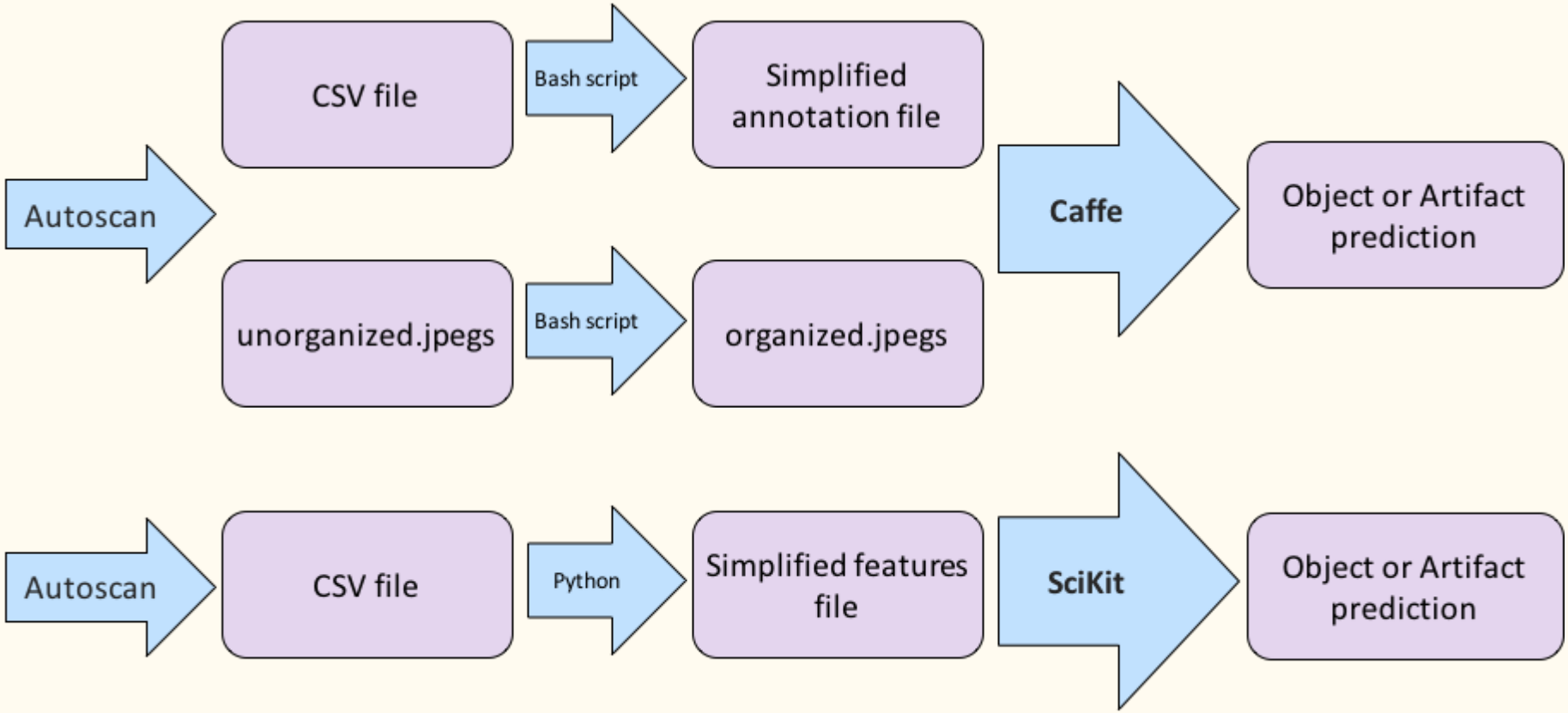




Autoscan Disadvantages

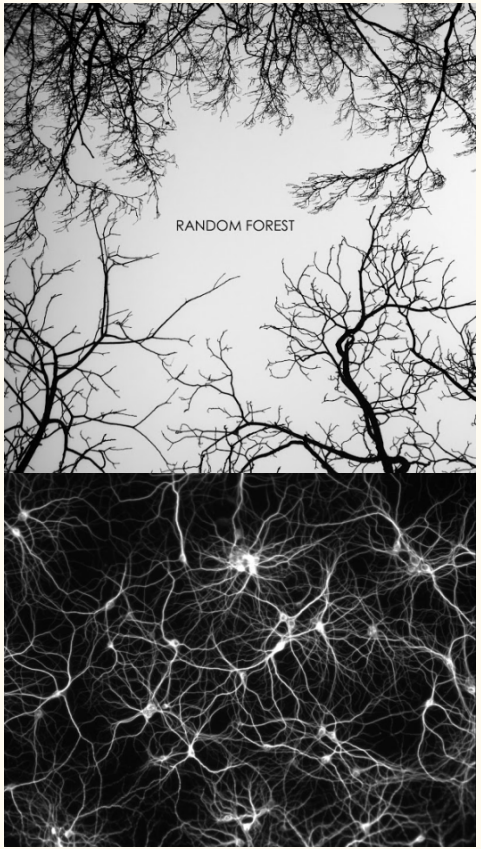
- Dependencies are proprietary
- Designed for specific dataset

Pale Blue Dot workflow



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

machine learning in Python

Results

	Caffe	SciKit
	67,000 training images	10,000 sample size
	92% accuracy	95.6% accuracy
True Positive	99.7%	82.4%
False Positive	0.3%	17.6%

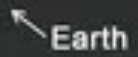
Advantages for our methods

- 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

UCMERCED
UNIVERSITY OF CALIFORNIA

GRADUATE
DIVISION
ICGE Program

BERKELEY LAB