

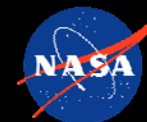


# Machine Learning Applications in Astronomy

Umaa Rebbapragada, Ph.D.  
Machine Learning and Instrument Autonomy Group

Big Data Task Force  
November 1, 2017

Research described in this presentation was carried out at the Jet Propulsion Laboratory under a Research and Technology Development Grant, under contract with the National Aeronautics and Space Administration. Copyright 2017 California Institute of Technology. All Rights Reserved. US Government Support Acknowledged. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement by the United States Government or the Jet Propulsion Laboratory, California Institute of Technology.



**Jet Propulsion Laboratory**  
California Institute of Technology

# Machine Learning (ML) for Astronomy

- Enabling Science
- Transient Science
- Catalog Science
- New Techniques
- Supporting ML in Astronomy

# JPL ML-Astronomy Collaborations

## Current and Past

- Palomar Transient Factory, intermediate Palomar Transient Factory, Zwicky Transient Facility
- The Very Long Baseline Array (VLBA) Fast Radio Transients Experiment (V-FASTR)
- Variables and Slow Transients (VAST) survey, part of Australian Square Kilometre Array Pathfinder (ASKAP)
- RealFast project at the Very Large Array (VLA) radio telescope
- MIT Lincoln Lab collaboration on Space Surveillance Telescope (SST) data
- Dark Energy Survey

# Enabling Science

# Big Telescopes, Big Science, Big Data



Large Synoptic Sky Telescope (LSST)  
15 TB/night



Square Kilometre Array  
160 TB/second



James Webb Space Telescope (JWST)



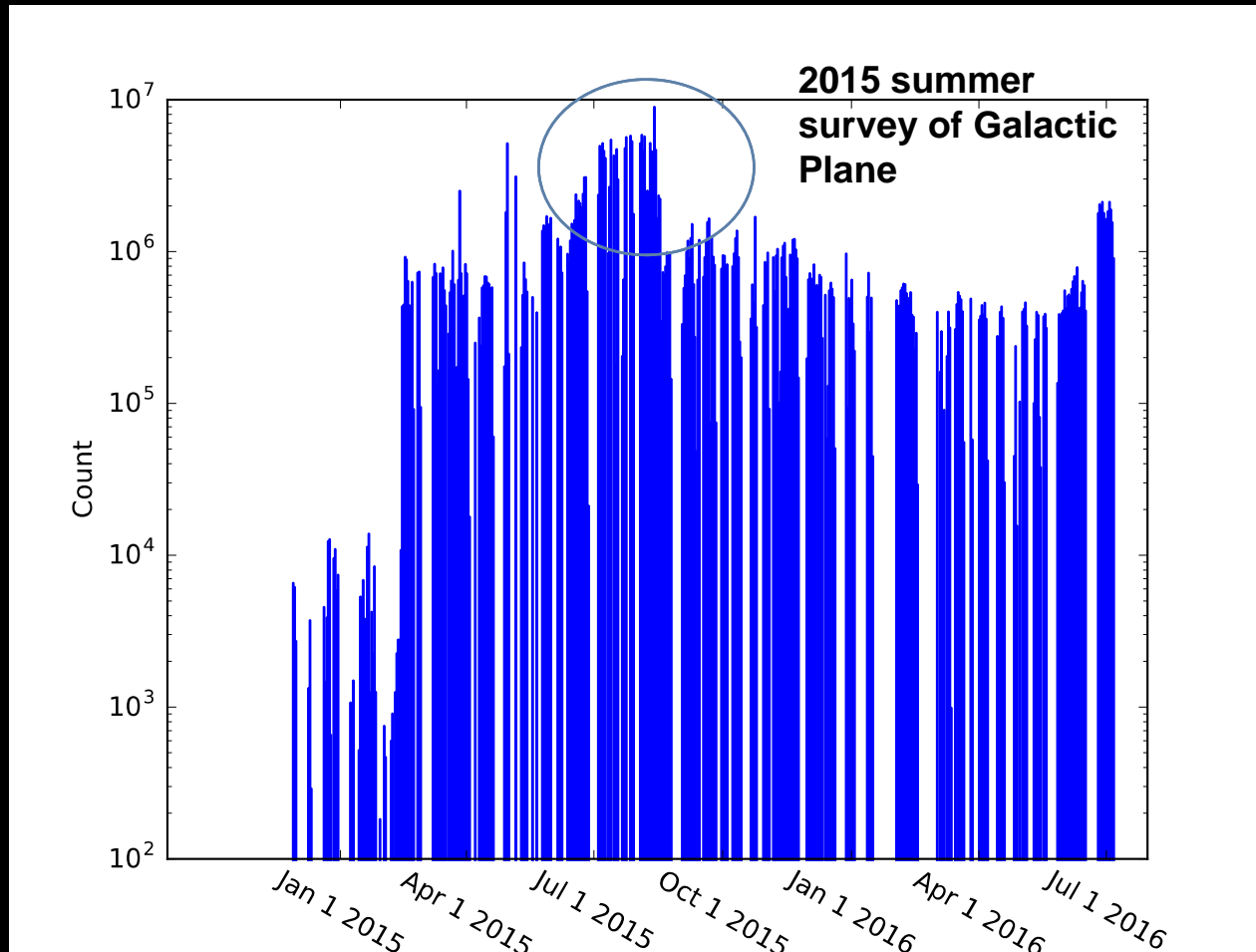
Wide-field Infrared Survey Telescope (WFIRST)



Transiting Exoplanet Survey Satellite (TESS)

# Millions of Detections per Night...in 2015

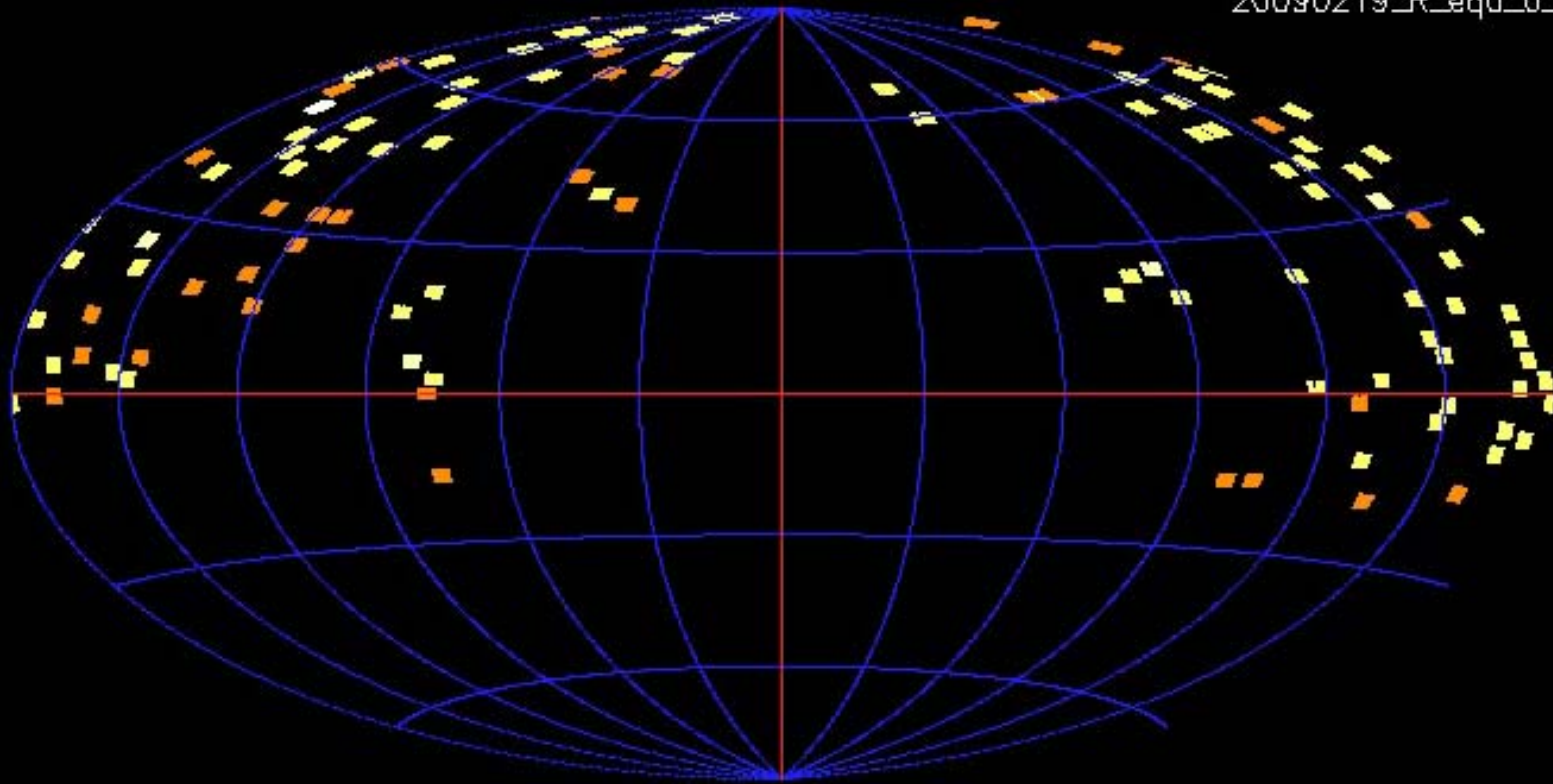
intermediate Palomar Transient Factory 2015-16



# Huge Catalogs from All-Sky Surveys

Observations of millions of objects, including spectra and lightcurves

20090219\_R\_equ\_0\_0.fits





# Science Team Size is Constraining Factor

ptf.nersc.gov/project/deepsky/ptfvet/treasures\_rb245.cg?young=0&tilt=0&coadd=0&new=0&hidrock=0&field=None&chip=None&remove\_smeas=0&verbose=0&date\_lo=20150401&date\_hi=201504016

## iPTF Science Team Could Handle 200 Objects / Night

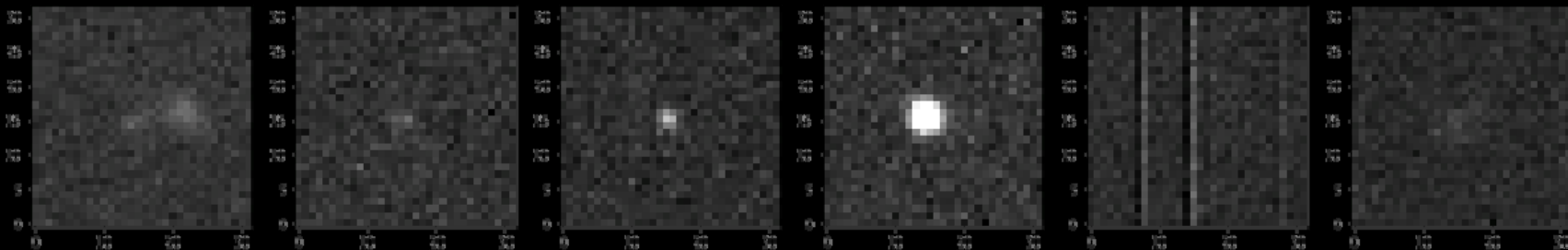
Found 71 candidates with rb2 between 0.30 and 1.00 or rb4 between 0.70 and 1.00 or rb5 between 0.60 and 1.00 Only showing unique candidates

New	Ref	Sub	SDSS	Metadata	Light Curve
				ID: 393281289 <a href="#">Examine</a> , 1171793 <a href="#">Zoom-Sub</a> RB2: 0.80 RB4: 0.72 RB5: 0.96 Mag: 17.91 iPTF <a href="#">15ku</a> Nearby <a href="#">KUG 1232+315</a> Abs Mag: -17.41 26 Matches in iPTF DB before tonight 0 Matches in PTF/best DB	
				ID: 393185333 <a href="#">Examine</a> , 1170882 <a href="#">Zoom-Sub</a> RB2: 0.58 RB4: 0.66 RB5: 0.93 Mag: 17.76 iPTF <a href="#">15wd</a> 0 Matches in iPTF DB before tonight 0 Matches in PTF/best DB	
				ID: 393279360 <a href="#">Examine</a> , 1171777 <a href="#">Zoom-Sub</a> RB2: 0.64 RB4: 0.73 RB5: 0.92 Mag: 17.00 iPTF <a href="#">15xl</a> 0 Matches in iPTF DB before tonight 0 Matches in PTF/best DB	
				ID: 393147669 <a href="#">Examine</a> , 1170645 <a href="#">Zoom-Sub</a> RB2: 0.89 RB4: 0.78 RB5: 0.92 Mag: 16.63 iPTF <a href="#">15afq</a> 0 Matches in iPTF DB before tonight 0 Matches in PTF/best DB	

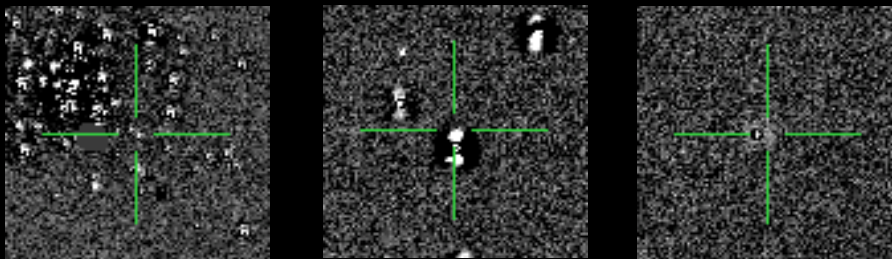
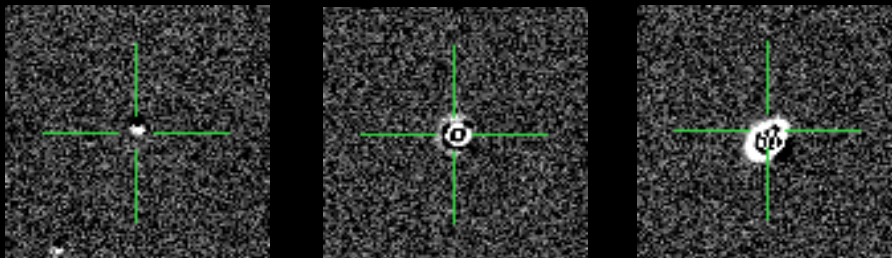


# Machine Learning Classifiers Filter False Detections

Are these candidates optical or pipeline artifacts?



Artifacts of Source Extraction



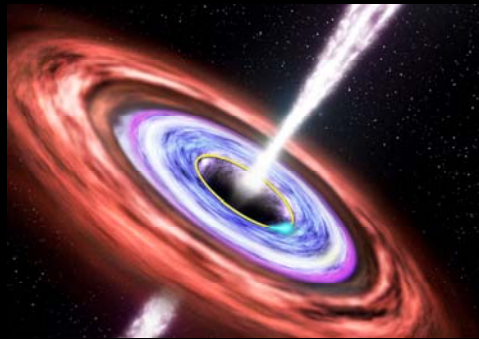
Artifacts of Image Differencing

# Transient Science

# Astronomical Transients

- Explosive events, very short duration

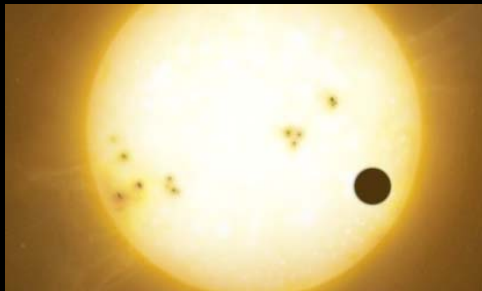
Gamma Ray  
Burst  
(milliseconds to hours)



Supernova  
(weeks to months)



- Transience from our observational perspective



Planetary  
transit

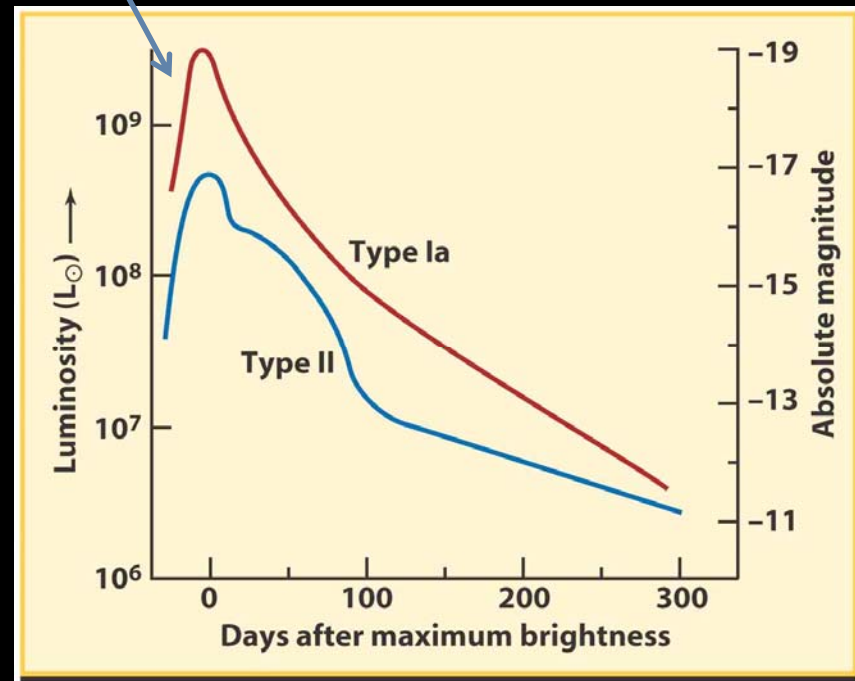


Asteroid

# Transient Science Requires Real-time Algorithms

- Find all scientifically-interesting observations
- Filter all irrelevant observations
- Goal: trigger follow-up of science-rich targets

Ideally, find it here

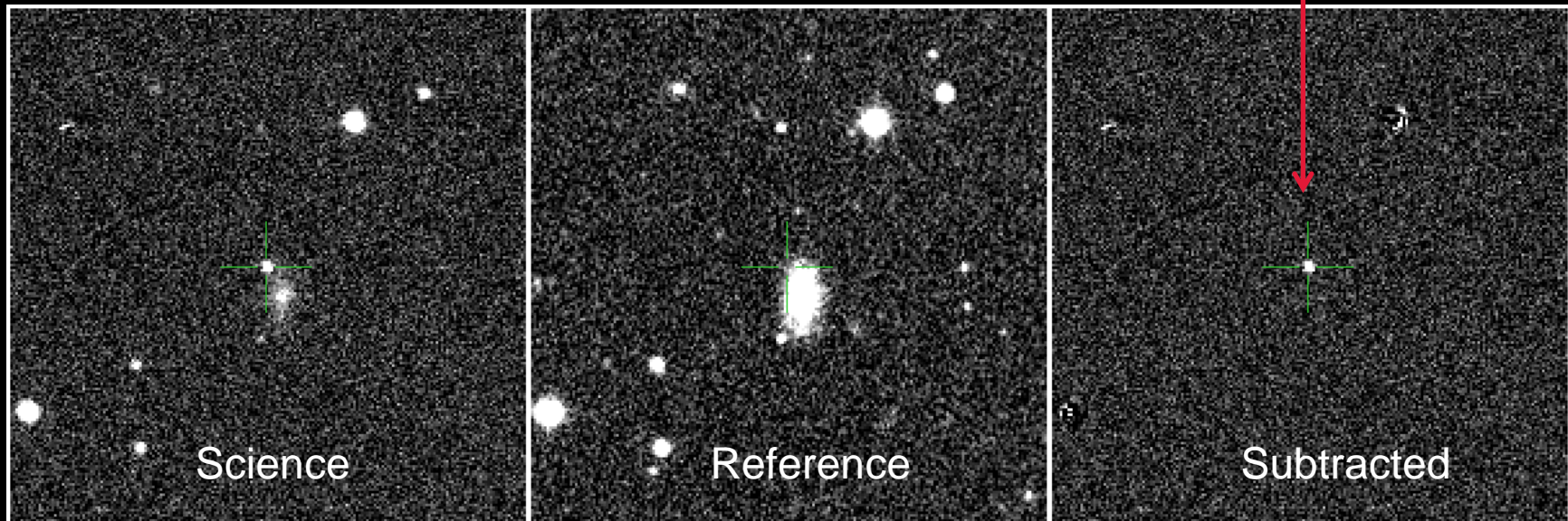


# Real-time Filtering

## Optical Astronomy Example from iPTF

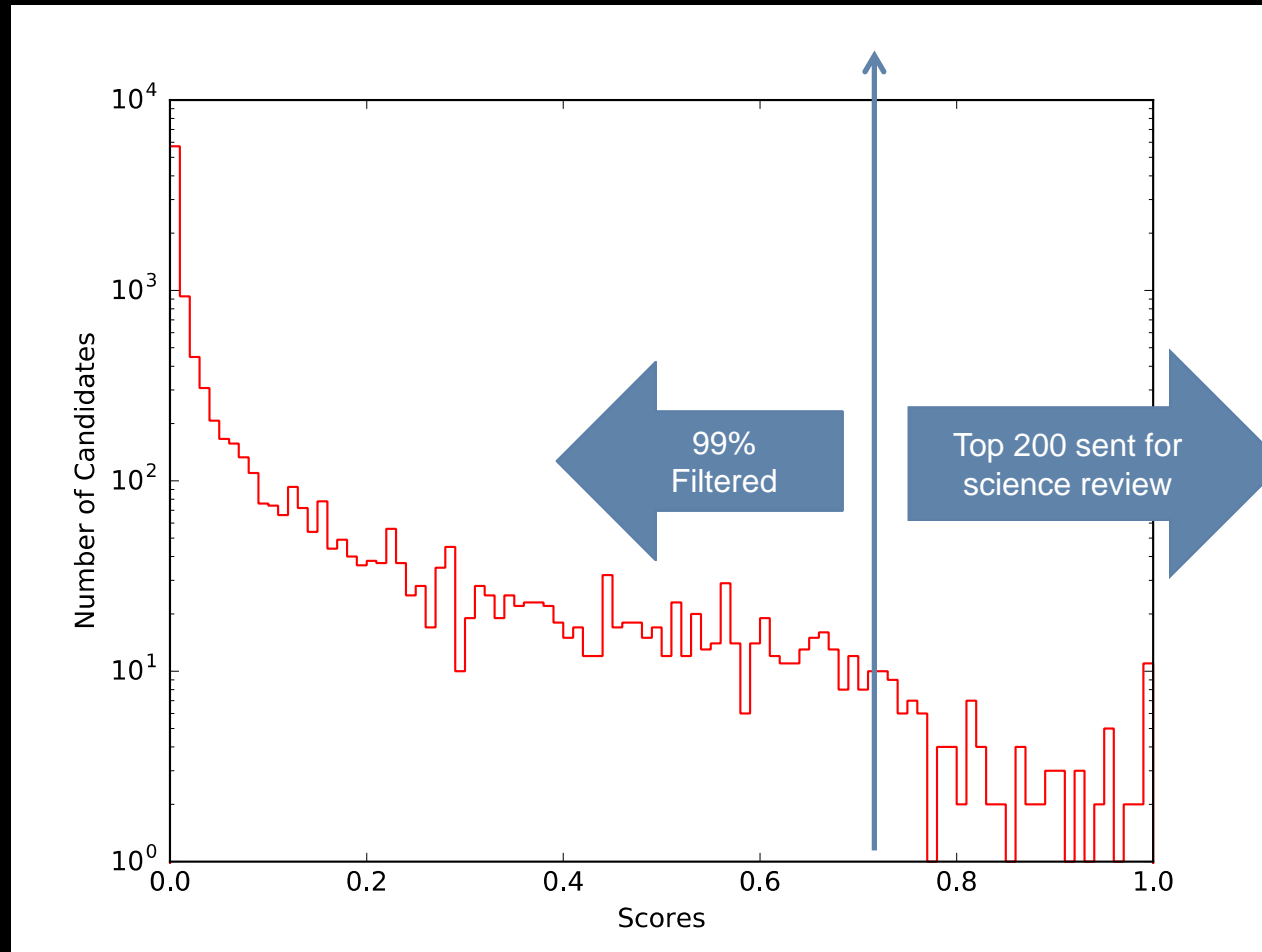
- Automated classifier scores candidates from 0 (bogus) to 1 (real)

Is this candidate real or bogus?



# Real-time Filtering

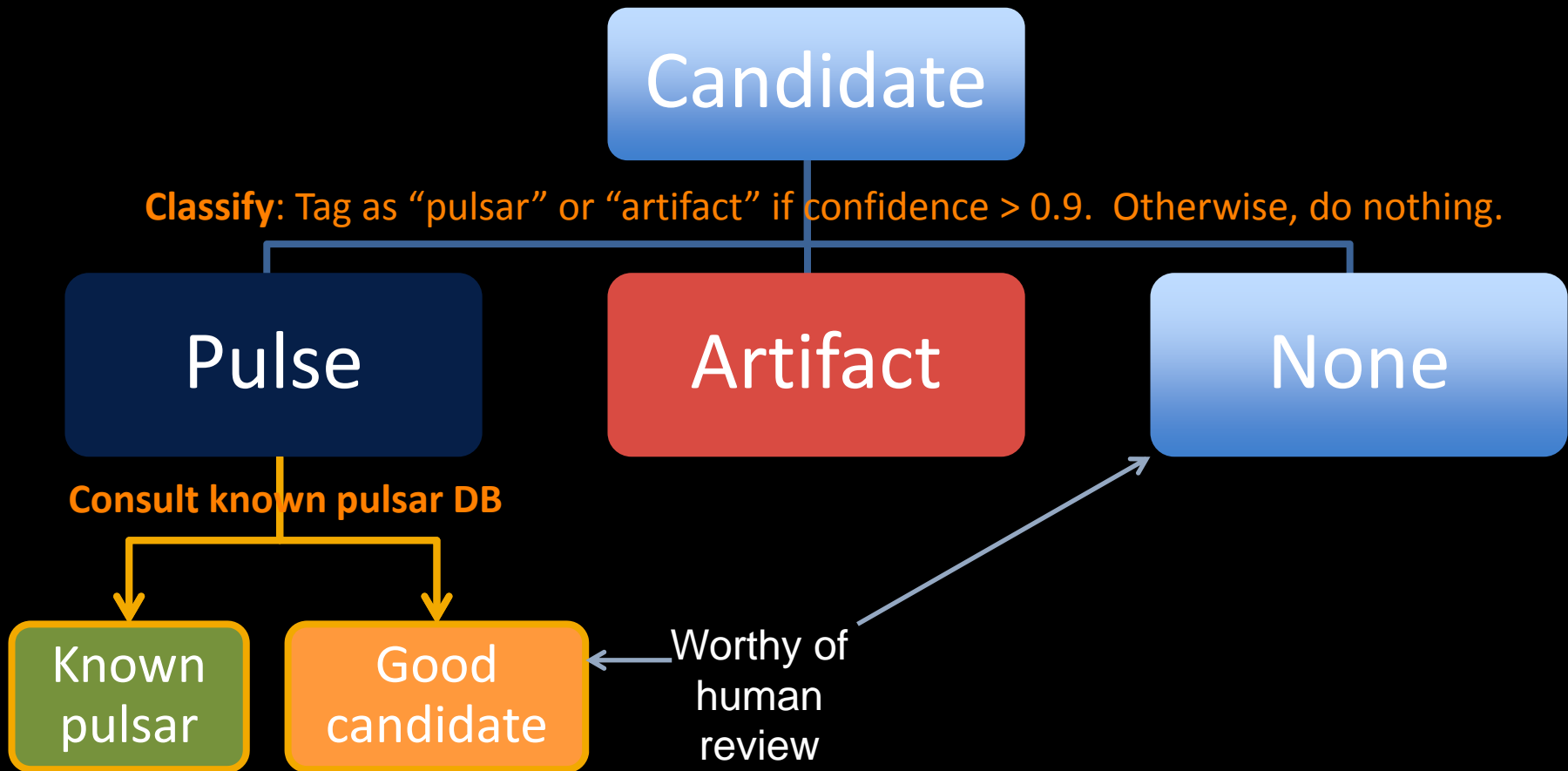
## Optical Astronomy Example from iPTF



Set decision threshold that filters 99%

# Real-time Filtering

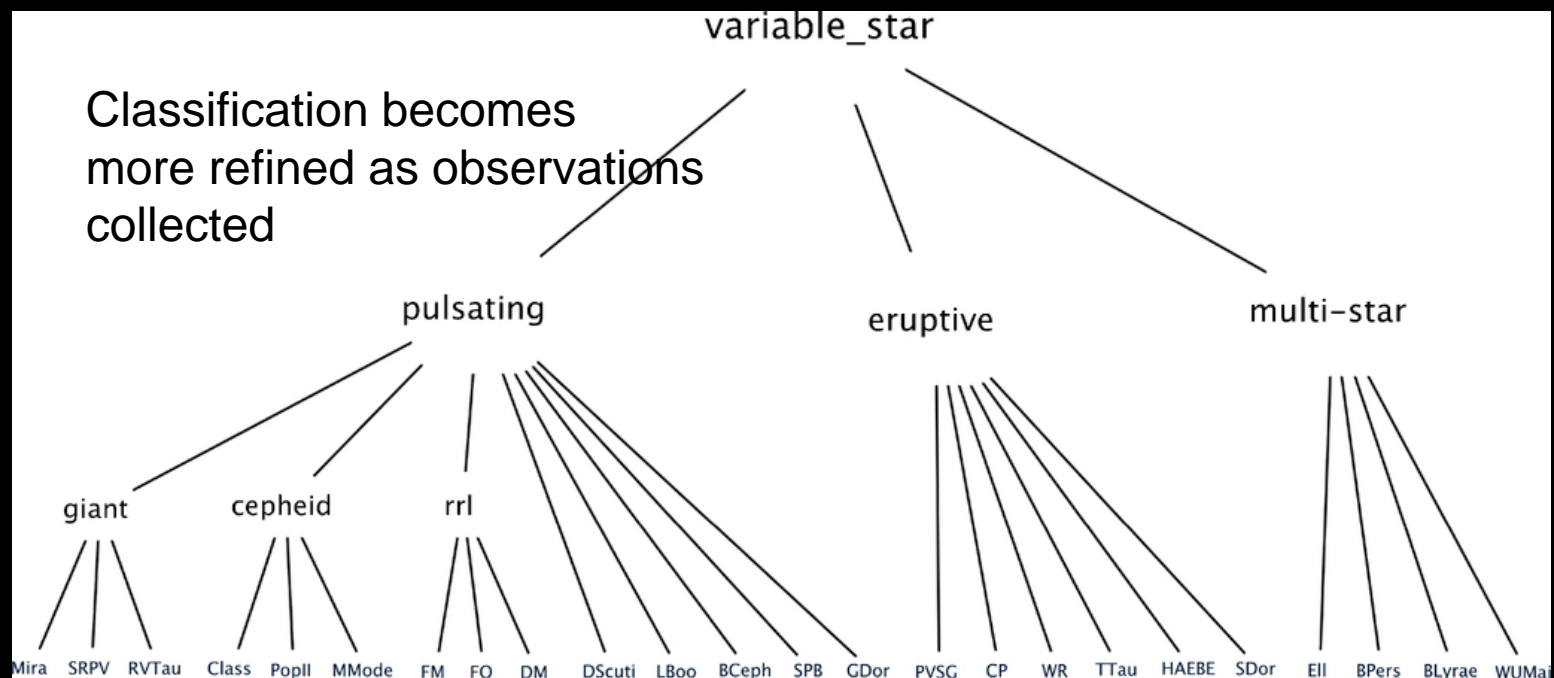
Radio Astronomy Example from V-FASTR / VLBA





# Real-time Classification

- Modern “data brokers” perform real-time classification of objects



Time



Richards, et al. 2011, ApJ

- Science users contribute filters to downselect the information they want

# Catalog Science

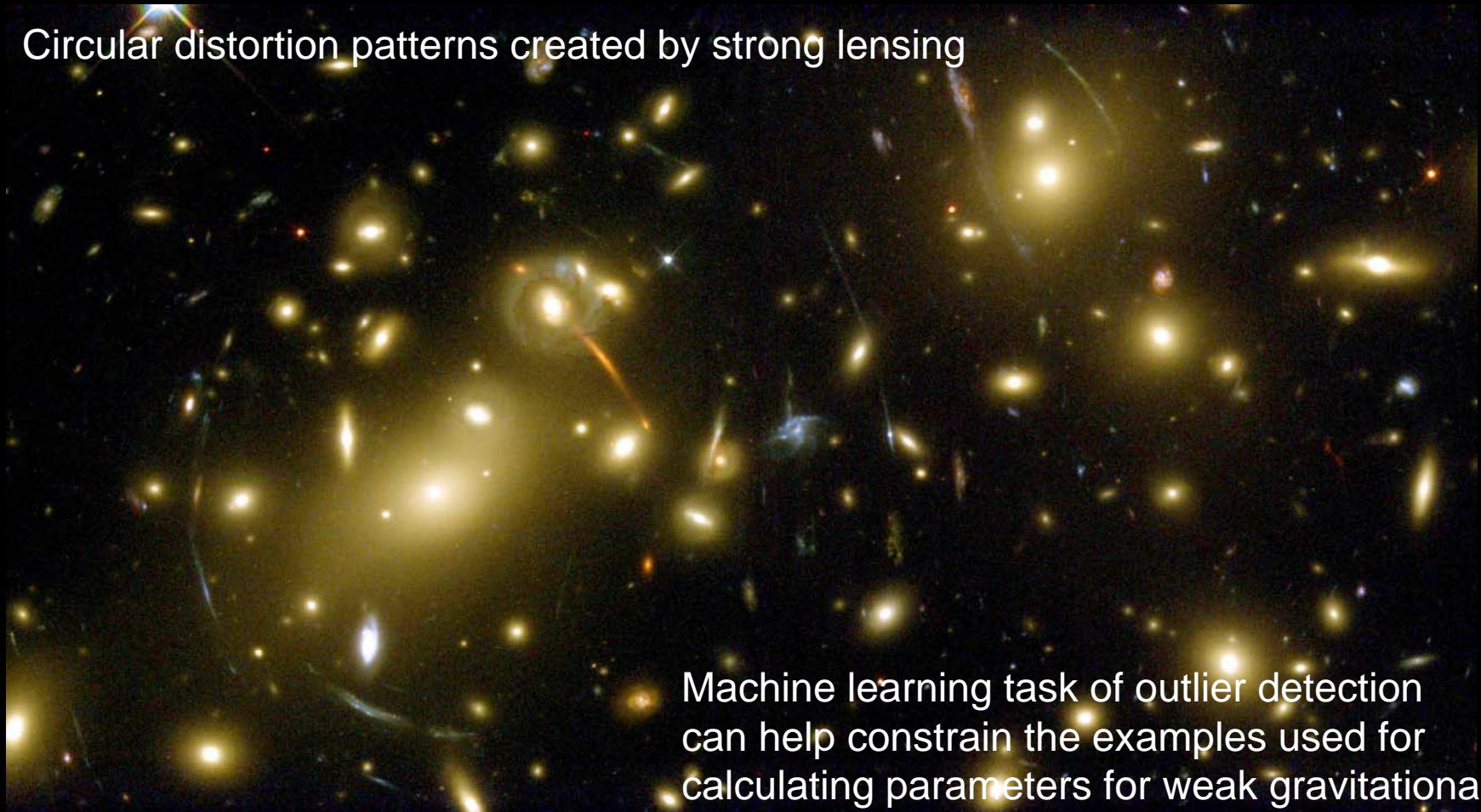
# ML Applications

- Stellar classification
- Star / galaxy separation
- Identification of planetary transits (exoplanets)
- NEO identification / tracking
- Estimating cosmological parameters

# Mining Astronomical Archives

## Weak Lensing Archives

Circular distortion patterns created by strong lensing



Machine learning task of outlier detection can help constrain the examples used for calculating parameters for weak gravitational lensing.

# New Techniques

# Crowdsourcing + ML

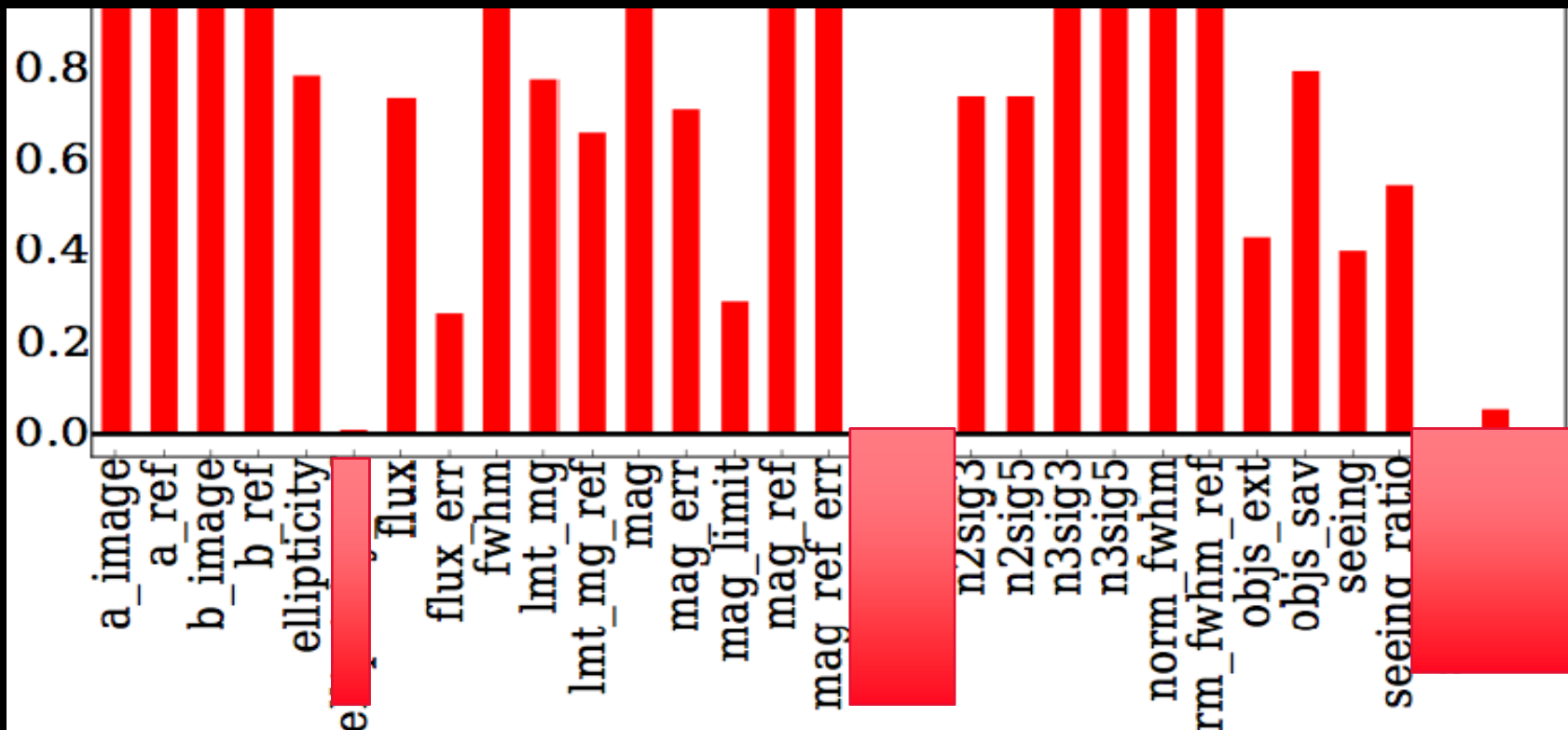
- Crowdsourcing platforms offer interfaces that integrate with ML algorithms and build training data quickly
- Allow interaction with scientists, general public

The screenshot shows the 'Exoplanet Explorers' interface on the Zooniverse website. The browser address bar shows the URL <https://www.zooniverse.org/projects/zenod/exoplanet-explorers/classify>. The page has a navigation bar with 'PROJECTS', 'ABOUT', 'GET INVOLVED', 'TALK', 'BUILD A PROJECT', and 'NEWS'. Below this is a sub-navigation bar with 'EXOPLANET EXPLORERS', 'ABOUT', 'CLASSIFY', 'TALK', and 'COLLECT'. The main content area features three graphs: 'Individual Transits' (a series of light curves), 'Entire Light Curve (Folded)' (a scatter plot of light curves), and 'transit-zoom' (a detailed view of a transit dip). To the right is a 'FIELD GUIDE' with the text: 'Does this look like a transiting planet? Look at the graph on the top-right: is it flat except for a dip at phase 0.0? (like the example image below). Look at the graph on the bottom-right: does the blue line fit the data points well? Check out the field guide to the right for examples of good and bad candidates.' Below the field guide are two radio buttons for 'Yes' and 'No'. At the bottom, there is a 'Need some help with this task?' section with a 'Done' button.

Exoplanet task on Zooniverse

# Responding to Instrument and Survey Changes

- Example of how a pipeline upgrade changed data characteristics at iPTF



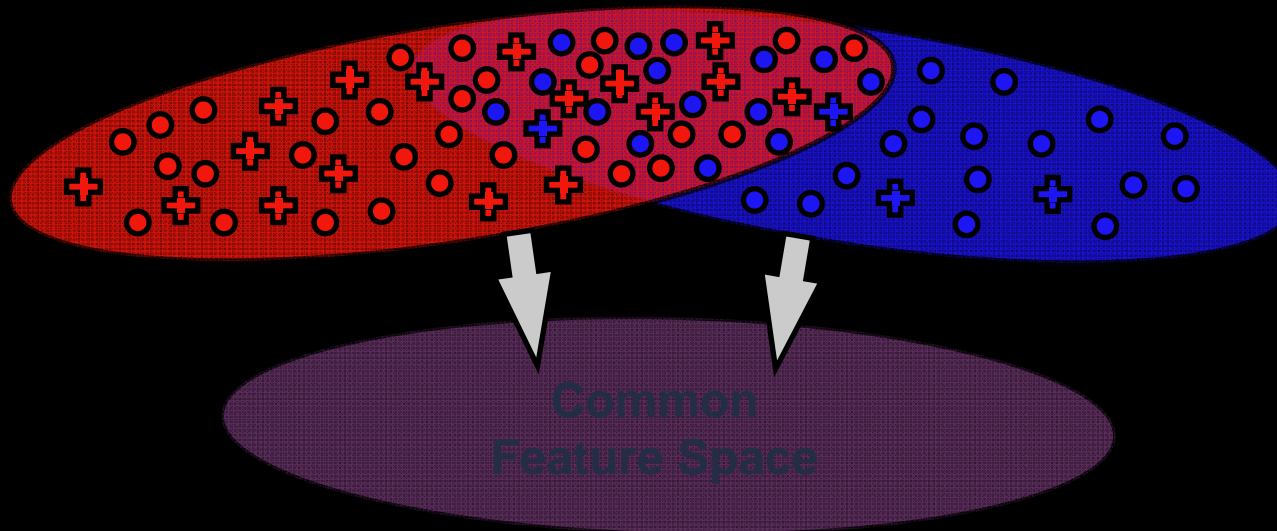


# Domain Adaptation

- Computes mapping between source and target data sources that share common science goals
- Continuous data record between old and new missions

**Old Instrument, pre-upgrade**

**New Instrument, post-upgrade**



# Deep Learning

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed at turpis vitae velit euismod aliquet. Pellentesque et arcu. Nullam venenatis gravida orci. Pellentesque et arcu. Nam pharetra. Vestibulum viverra varius enim.

Nam laoreet dui sed magna. Nunc in turpis ac lacus eleifend sagittis. Pellentesque ac turpis. Aliquam justo lectus, iaculis a, auctor sed, congue in, nisl. Aenean luctus vulputate turpis. Mauris urna sem, suscipit vitae, dignissim id, ultrices sed, nunc.

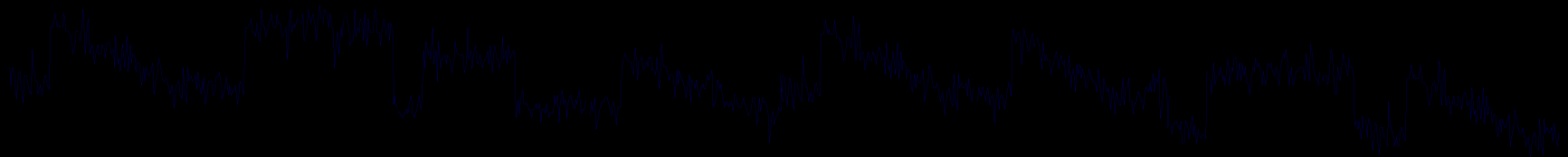
Phasellus nisi metus, tempus sit amet, ultrices ac, porta nec, felis. Quisque malesuada nulla sed pede volutpat pulvinar. Sed non ipsum. Mauris et dolor. Pellentesque suscipit accumsan massa. In consectetur, lorem eu lobortis egestas, velit odio



Pixel values,  
SIFT, HoG,  
histograms of visual words

DFT, wavelets,  
time series statistics

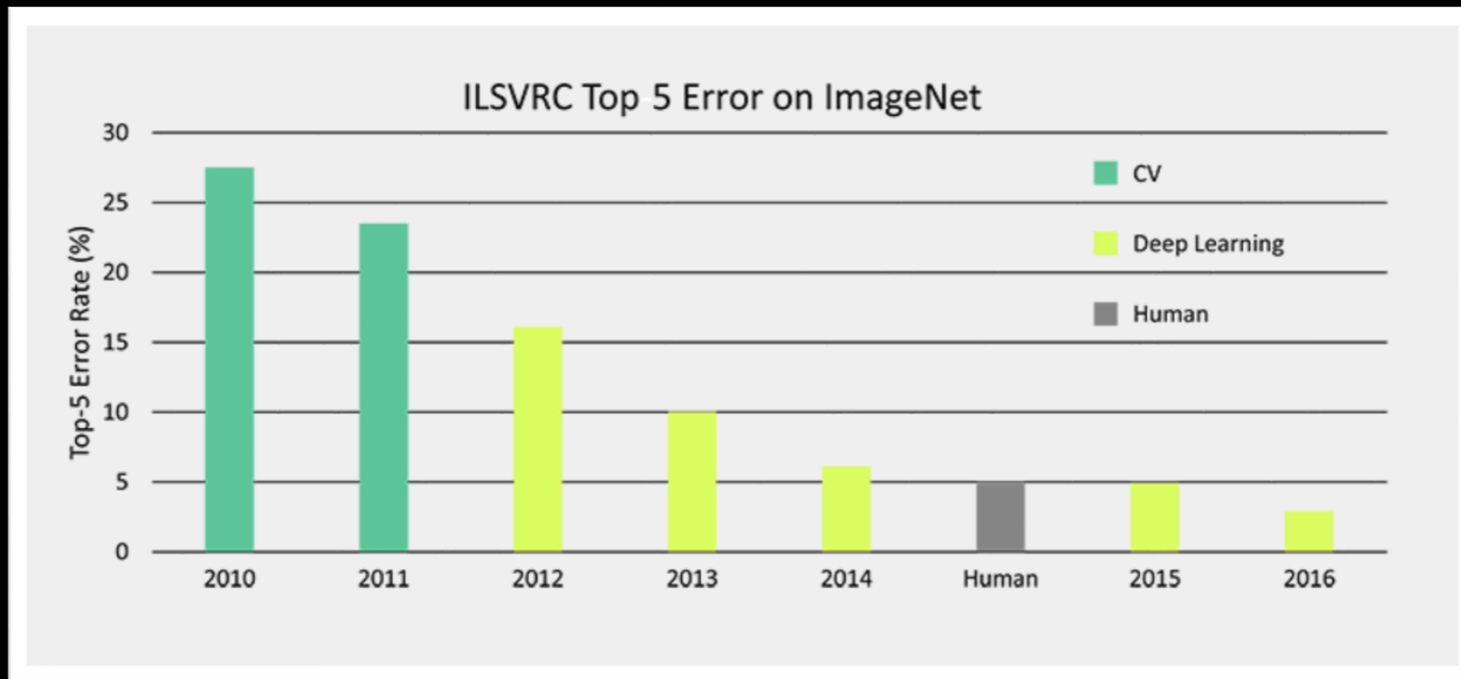
Bag of Words  
TFIDF



# Deep Learning

## ImageNet Large Scale Visual Recognition Challenge (ILSVRC)



<https://www.dsiac.org/resources/journals/dsiac/winter-2017-volume-4-number-1/real-time-situ-intelligent-video-analytics>

# Supporting ML in Astronomy

# Challenges and Opportunities

- Programmatic
  - Space telescope managers don't know we exist
  - ML experts not involved in data pipeline requirements and design
- Financial
  - Few ROSES opportunities that support ML work in astronomy
  - Limited or no budget for high level data products produced by ML
- Cultural
  - "It's just software."
  - "My post-doc can do it."



**Jet Propulsion Laboratory**  
California Institute of Technology

---

[jpl.nasa.gov](http://jpl.nasa.gov)