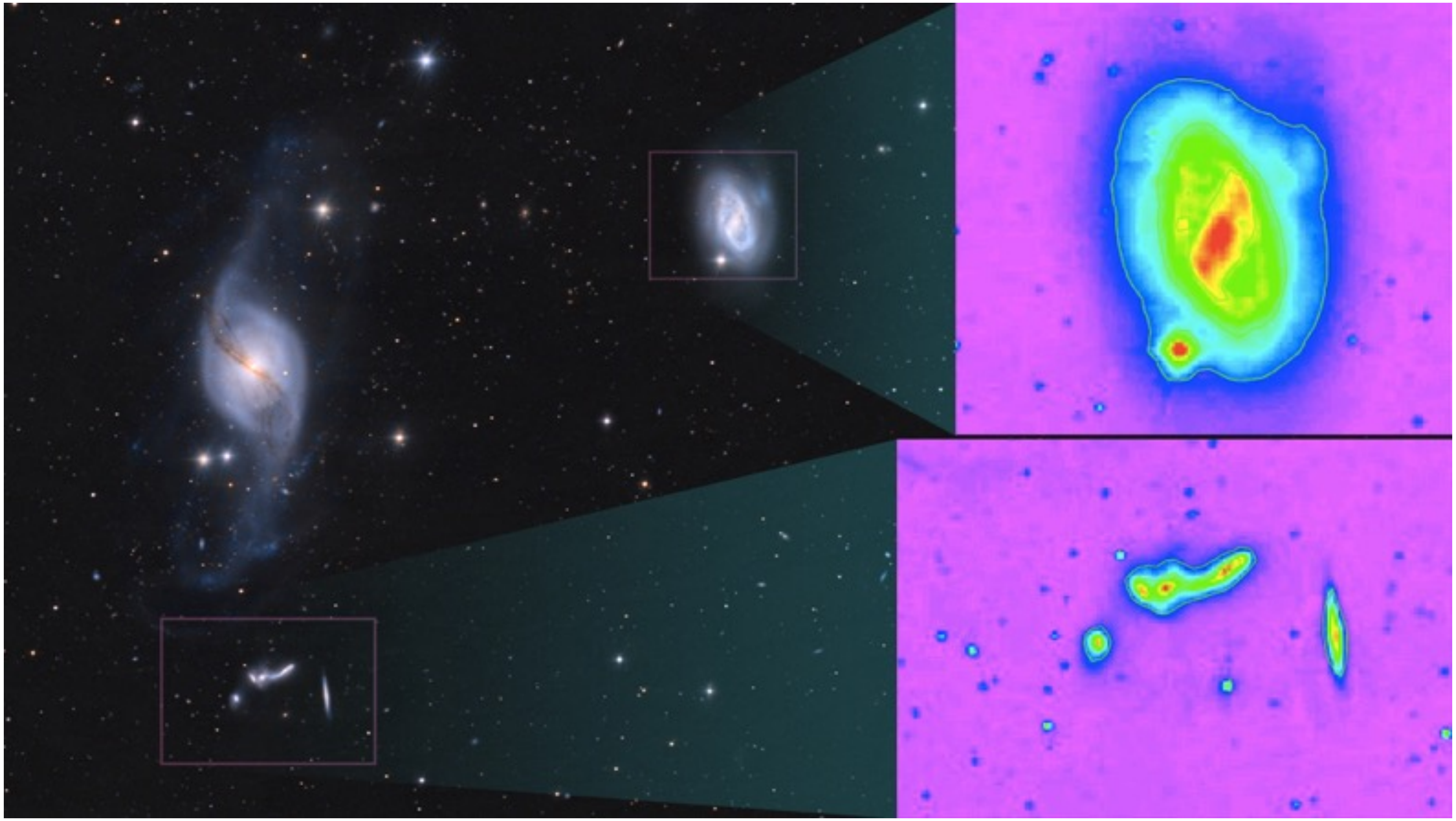
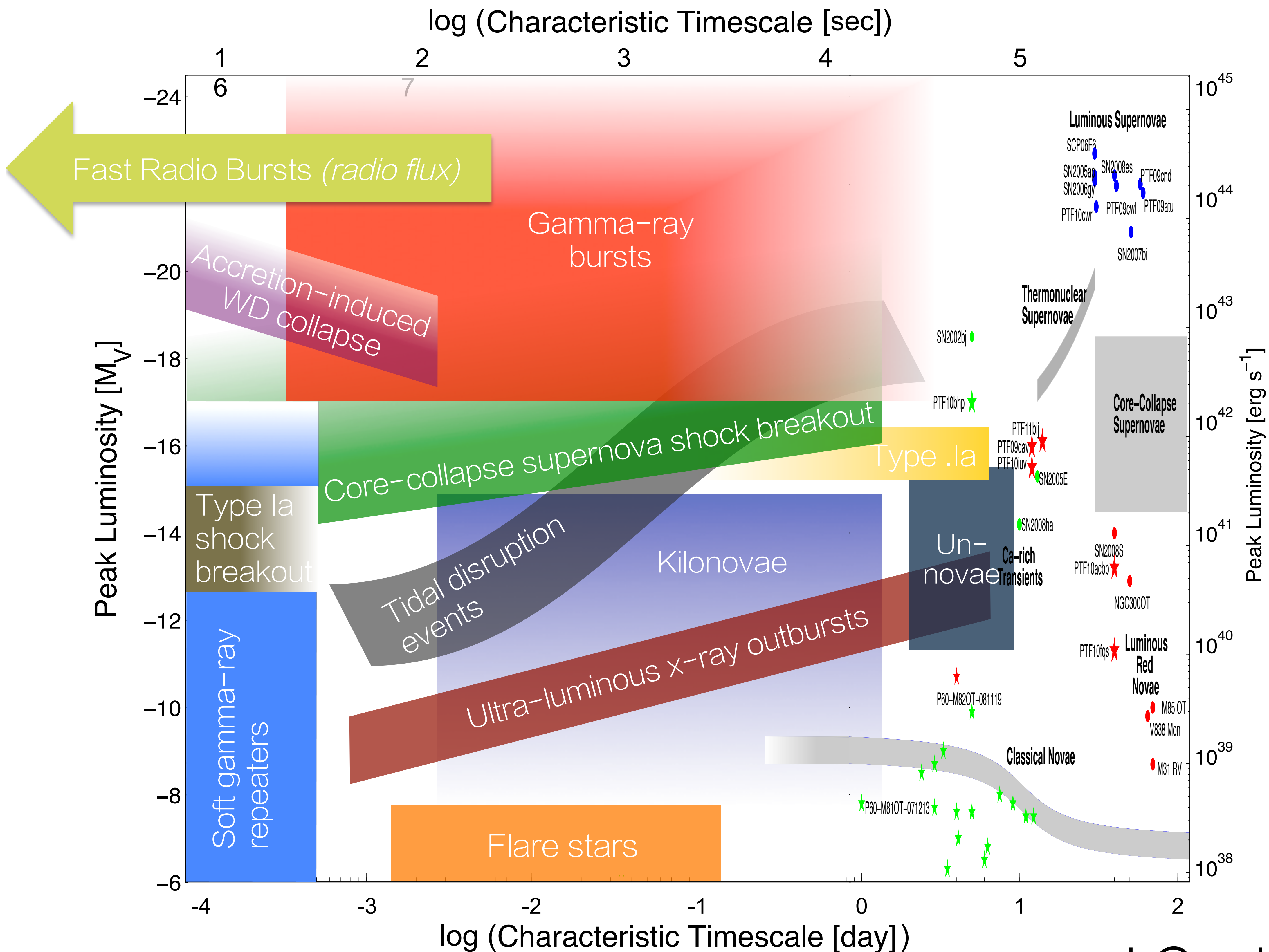
Astronomical Applications of Machine Learning and Neural Networks



Ashish Mahabal <aam@astro.caltech.edu>
Center for Data Driven Discovery, Caltech
LISA, KISS, Caltech, 20180117

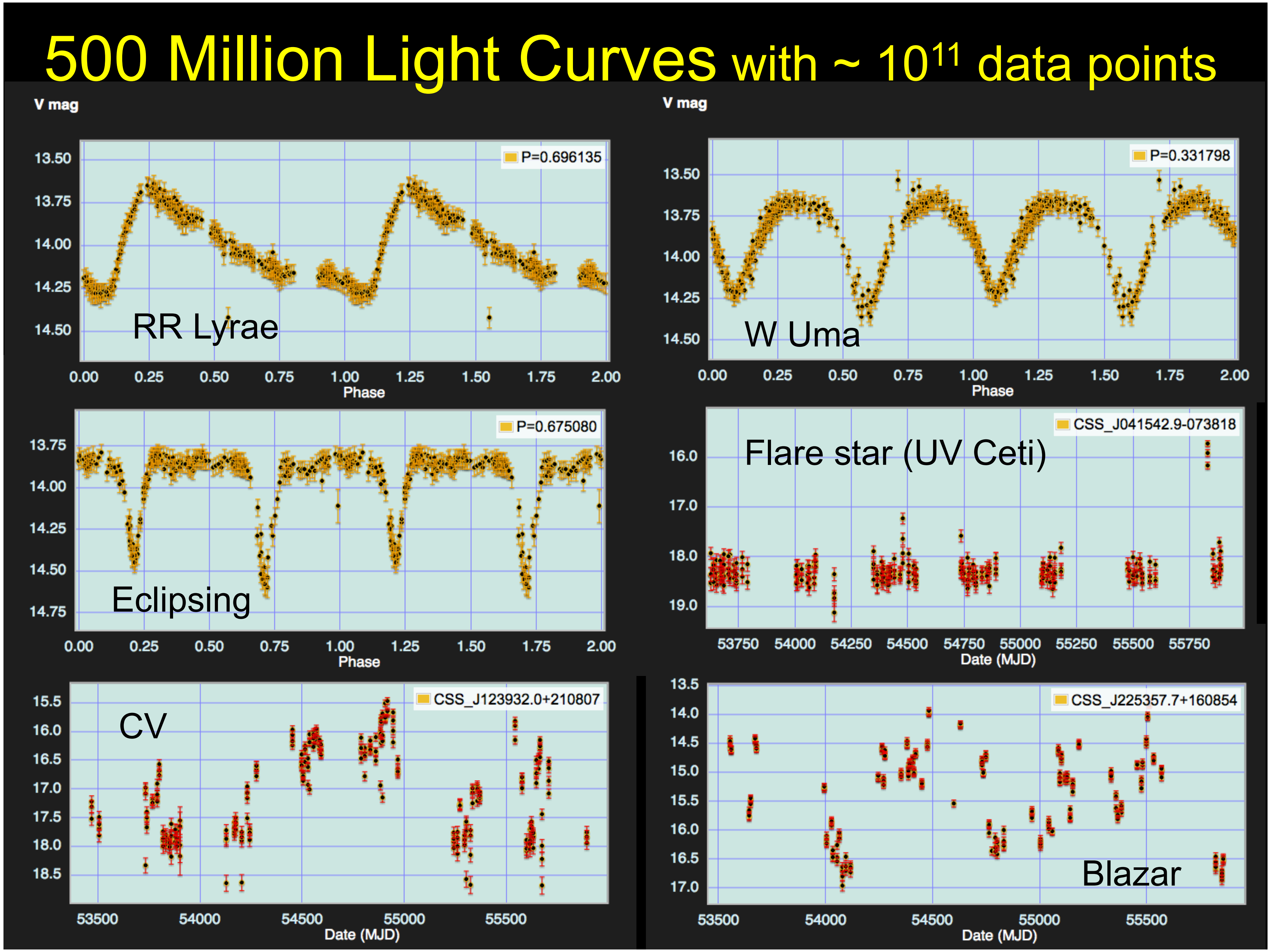


log (Characteristic Timescale [sec])
1
2
3
4
5
6
7
Fast Radio Bursts (radio flux)
Gamma-ray bursts
Accretion-induced WD collapse
Core-collapse supernova shock breakout
Type .Ia
Type Ia shock breakout
Tidal disruption events
Kilonovae
Un-novae
Ultra-luminous x-ray outbursts
Soft gamma-ray repeaters
Flare stars
Luminous Supernovae
Thermonuclear Supernovae
Core-Collapse Supernovae
Ca-rich Transients
Luminous Red Novae
Classical Novae
Peak Luminosity [M_V]
Peak Luminosity [erg s^{-1}]
log (Characteristic Timescale [day])

J Cooke

Overwhelming (amounts of) data

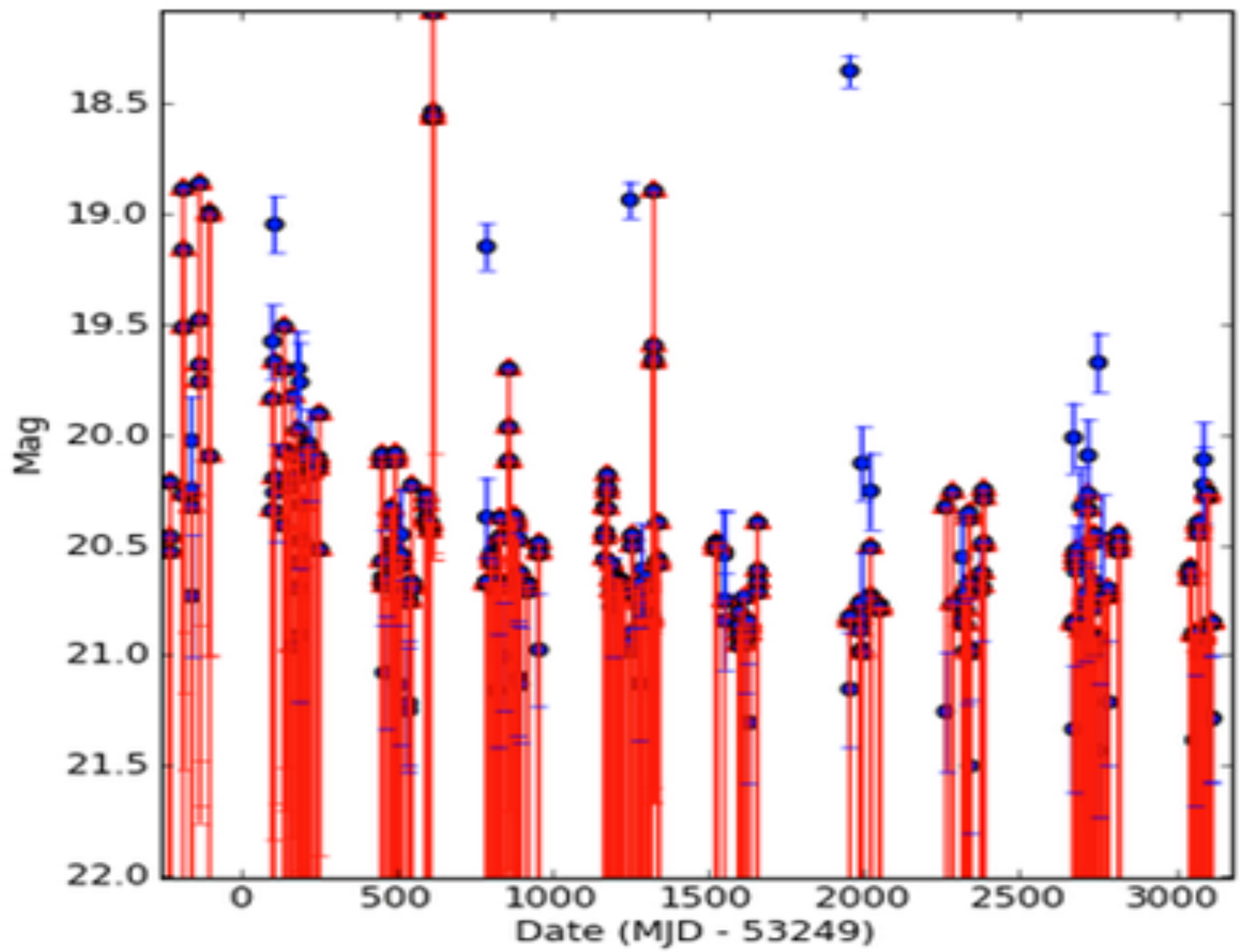
CRTS



ZTF, LSST, SKA

Properties of light-curves

- Gappy
- Irregular
- Heteroskedastic



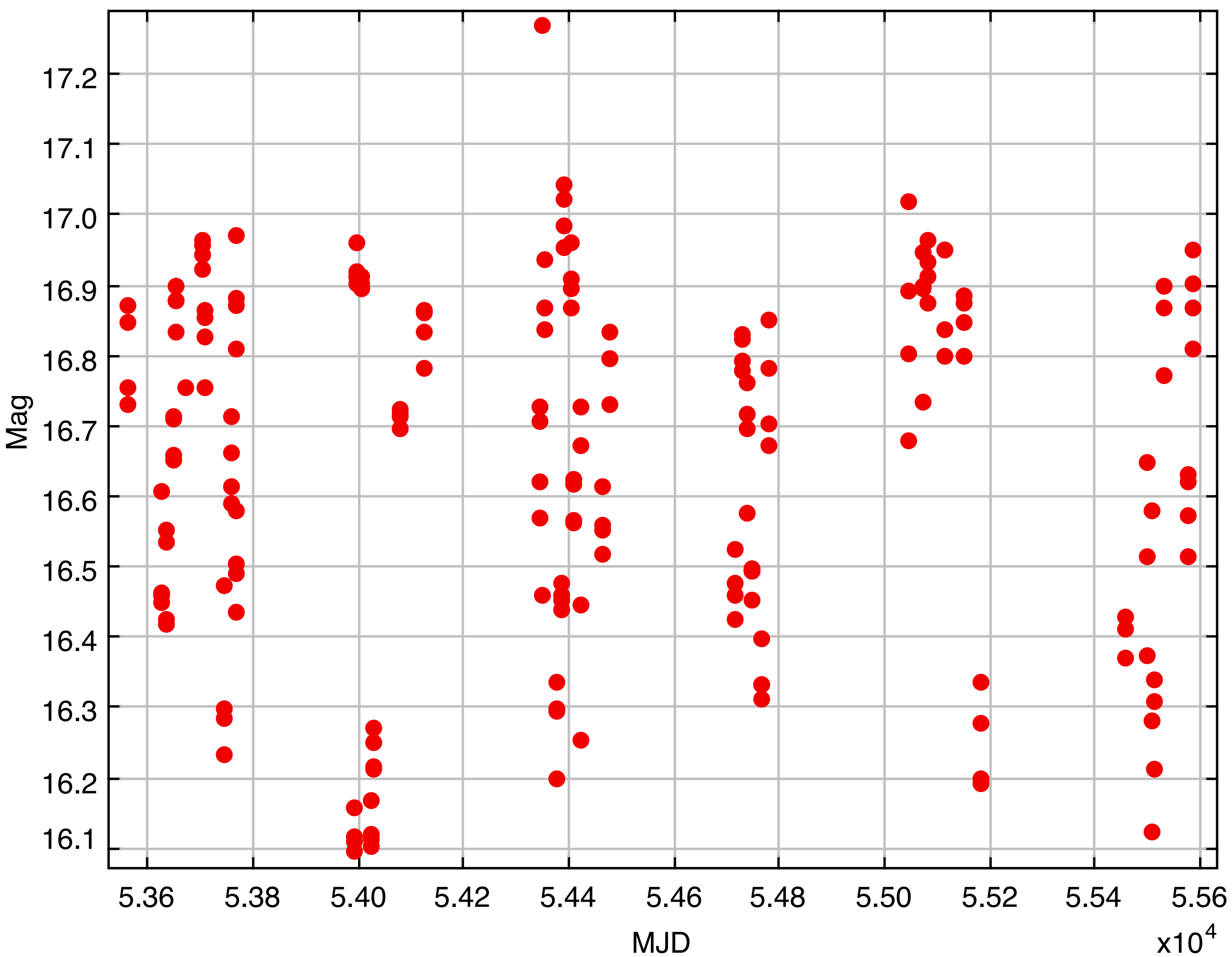
Reasons:

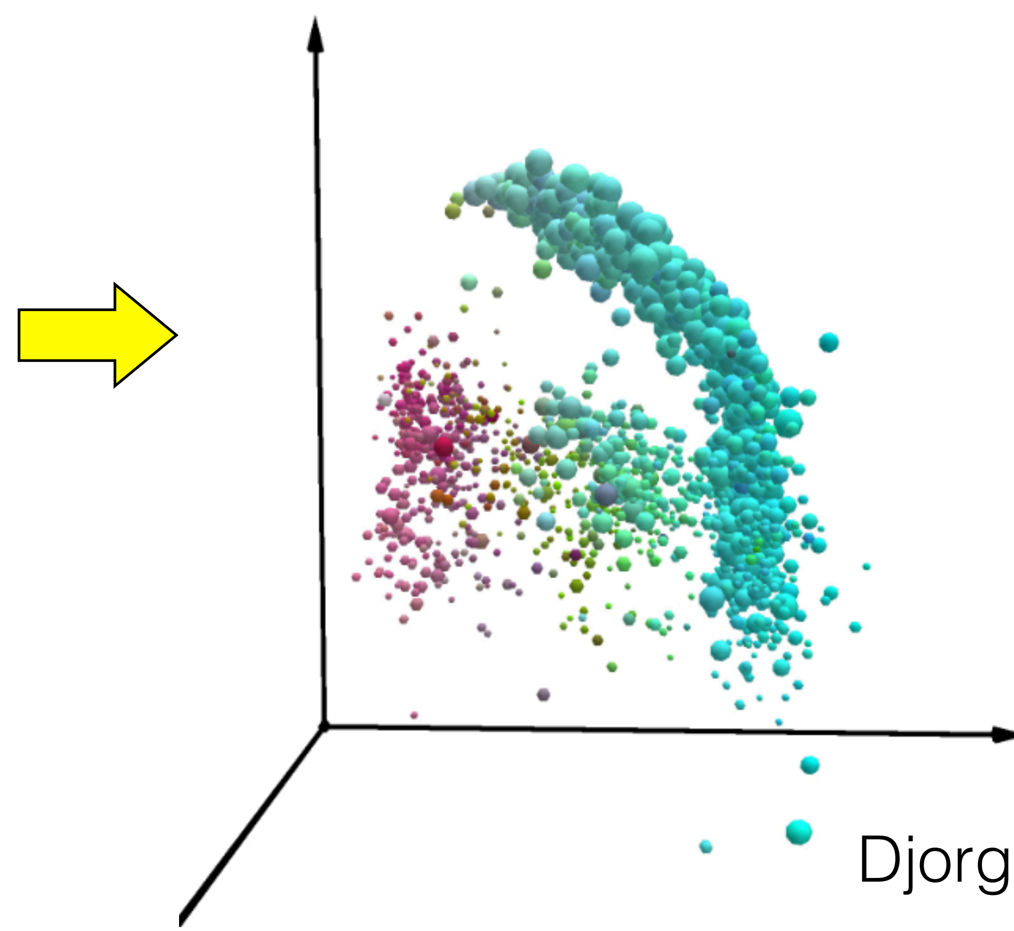
- **expense, rotation/revolution of Earth, moon**
- **science objectives, weather, moon**
- **weather, moon, airmass**

errors ignored by many methods

Statistical features

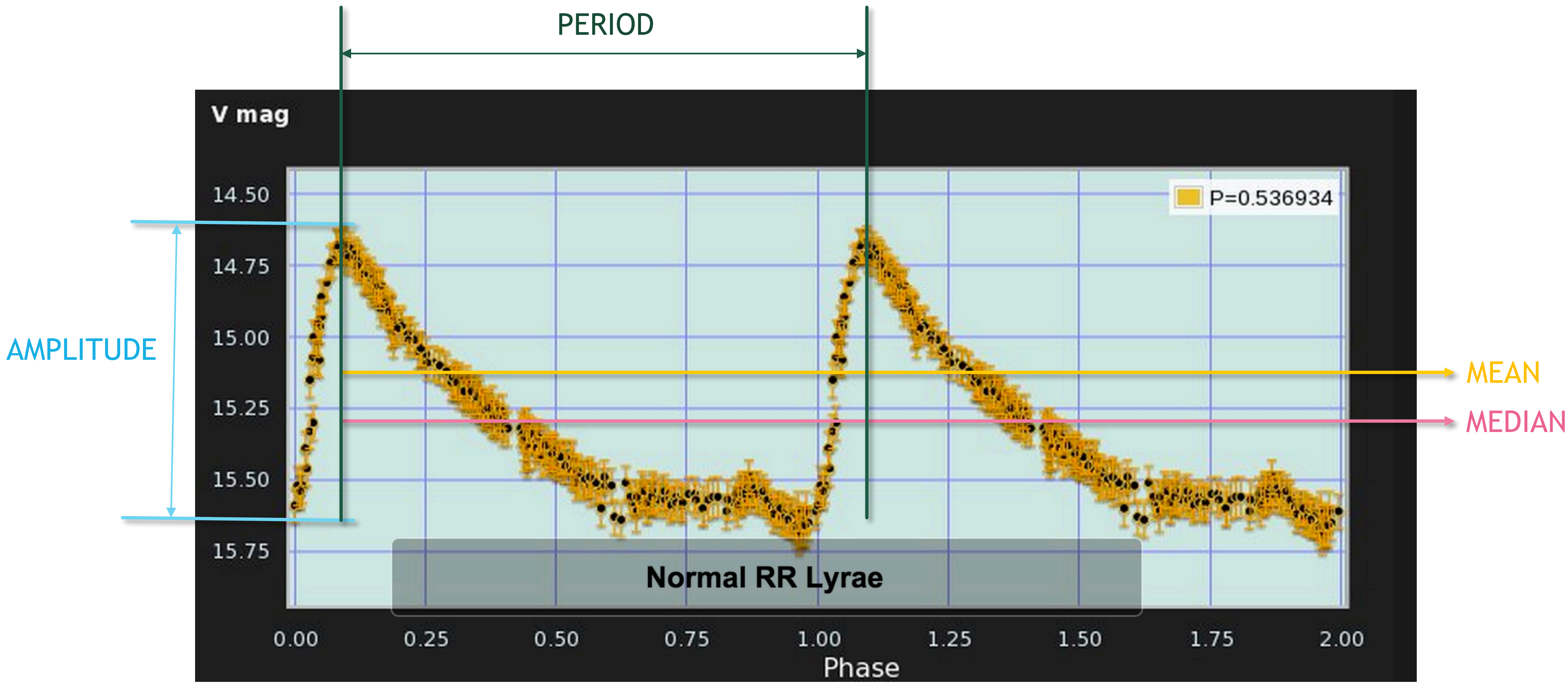
Compute features (statistical measures) for each light curve:
amplitudes, moments, periodicity, etc.
Converts heterogeneous light curves into homogeneous
feature vectors in the parameter space
Apply a variety of automated classification methods





Djorgovski

Light-curve features



Statistical features

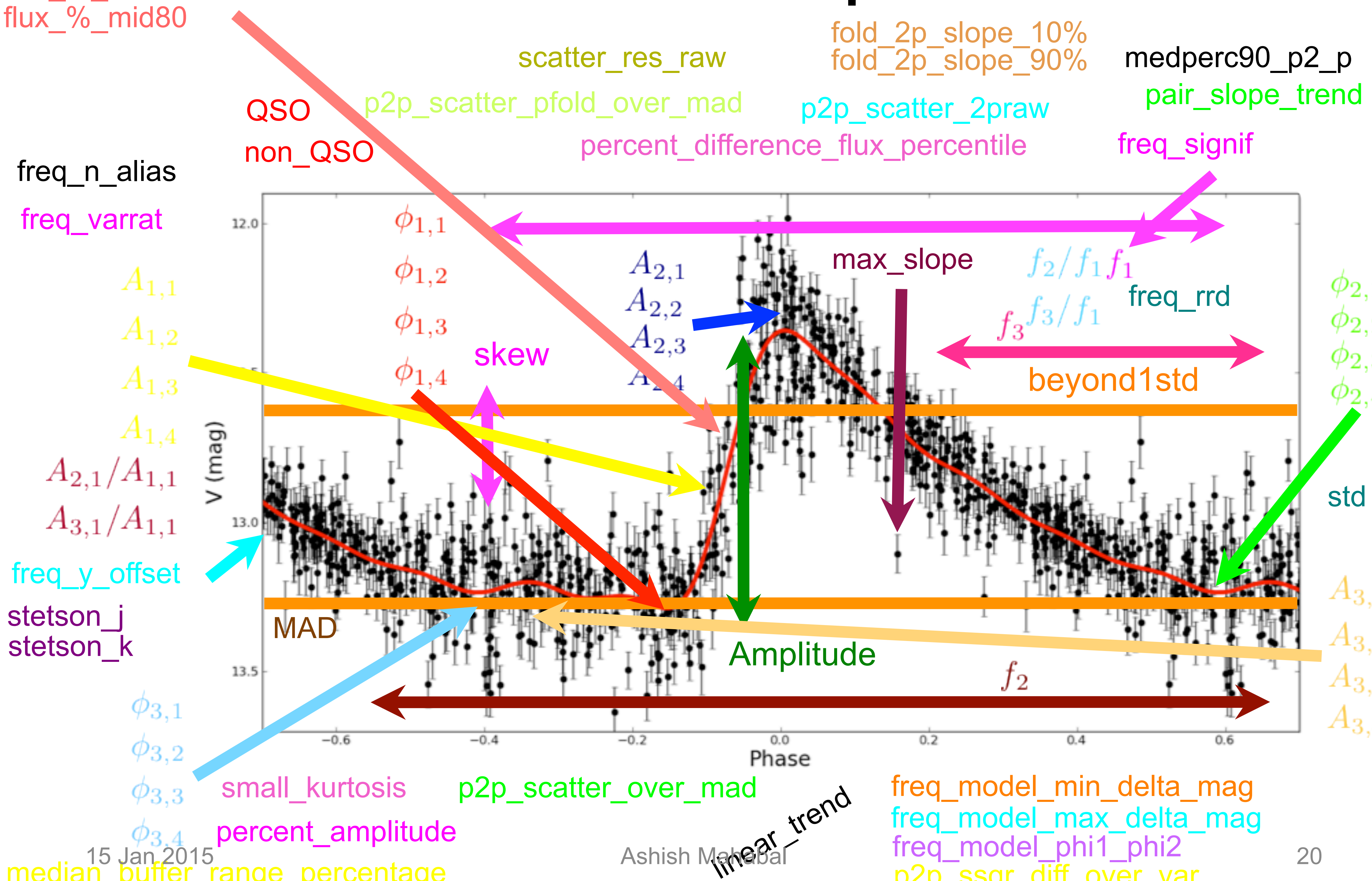
TABLE VI

RANDOM FOREST FEATURES. THE FIRST THREE ARE NOT USED IN RF WHEREAS THE REMAINING 18 ARE FAIRLY GENERIC FEATURES OFTEN USED IN CLASSIFICATION E.G. [1], [2], [3]. FORMULAE FOR THE FEATURES ARE FROM $http://nirgun.caltech.edu:8000/scripts/description.html\#method\_summary$

Feature	Formula
meanmag	$< mag >$
minmag	mag_{min}
maxmag	mag_{max}
amplitude	$0.5 * (mag_{max} - mag_{min})$
beyond1std	$p(\|(mag- < mag >)\| > \sigma)$
flux percentile ratio mid20	$(flux_{60} - flux_{40})/(flux_{95} - flux_5)$
flux percentile ratio mid35	$(flux_{67.5} - flux_{32.5})/(flux_{95} - flux_5)$
flux percentile ratio mid50	$(flux_{75} - flux_{25})/(flux_{95} - flux_5)$
flux percentile ratio mid65	$(flux_{82.5} - flux_{17.5})/(flux_{95} - flux_5)$
flux percentile ratio mid80	$(flux_{90} - flux_{10})/(flux_{95} - flux_5)$
linear trend	b where mag = a * t + b
max slope	$max(\|(mag_{i+1} - mag_i)/(t_{i+1} - t_i)\|)$
median absolute deviation	$med(flux - flux_{med})$
median buffer range percentage	$p(\|flux - flux_{med}\| < 0.1 * flux_{med})$
pair slope trend	$p(flux_{i+1} - flux_i > 0; i = n - 30, n)$
percent difference flux percentile	$(flux_{95} - flux_5)/flux_{med}$
skew	μ_3/σ^3
small kurtosis	μ_4/σ^4
std	σ
stetson j	var_j (mag)
stetson k	var_k (mag)

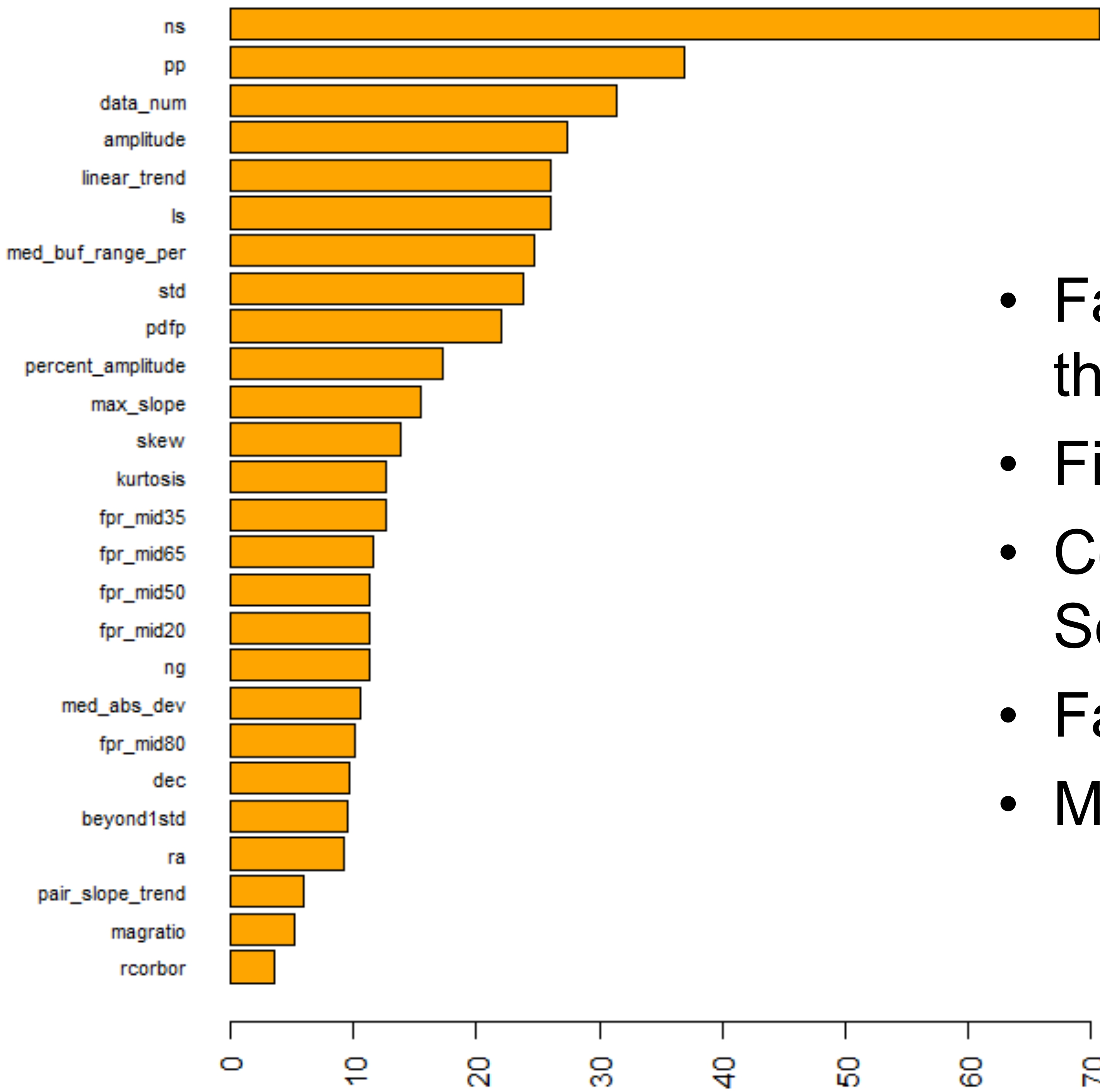
Many features
- not all are independent

Adam Miller



15 Jan 2015 Ashish Mahabal 20

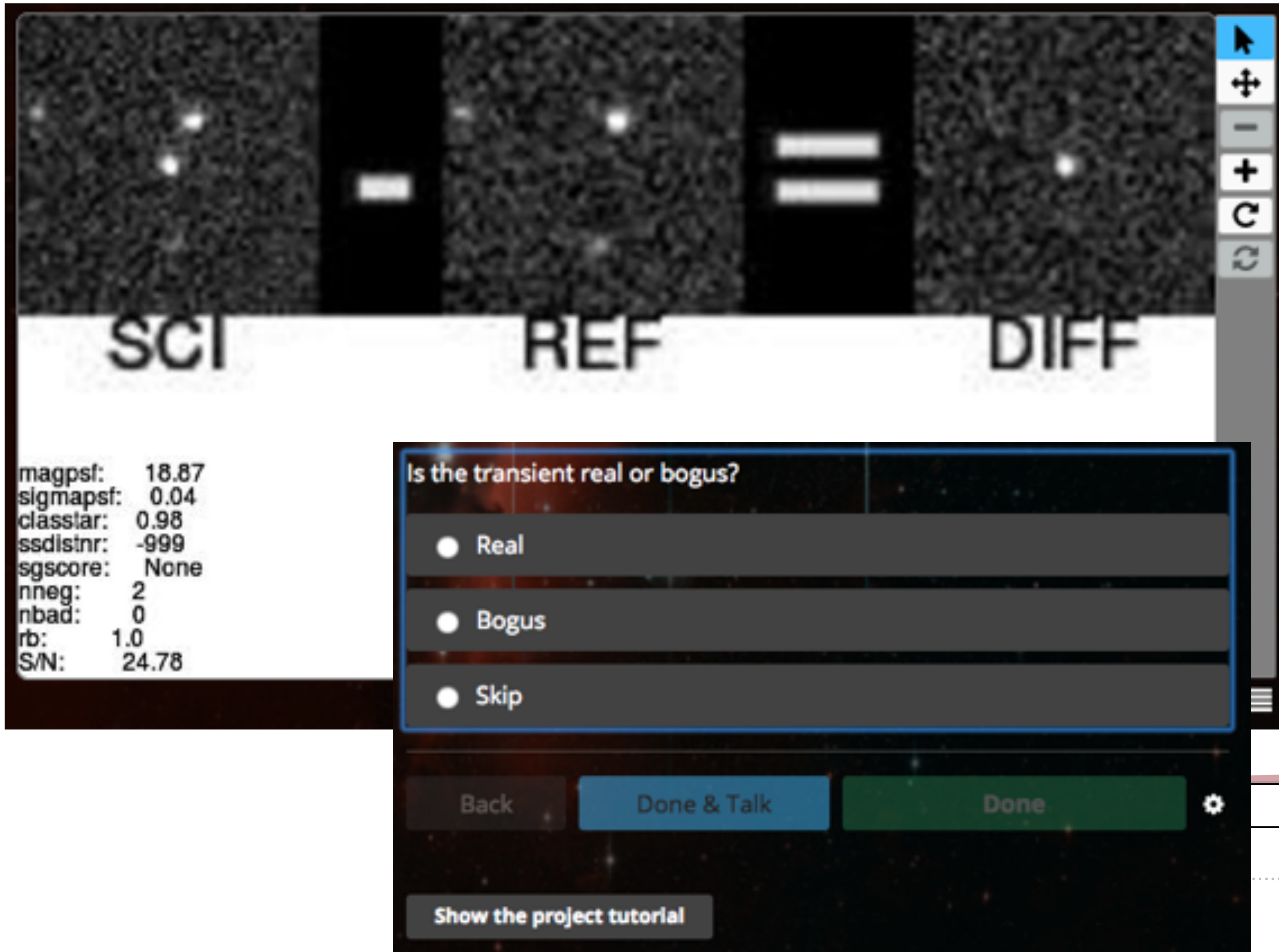
Feature selection strategy



- Fast Relief Algorithm (wt and threshold)
- Fisher Discriminant Ratio
- Correlation based Feature Selection
- Fast Correlation Based Filter
- Multi Class Feature Selection

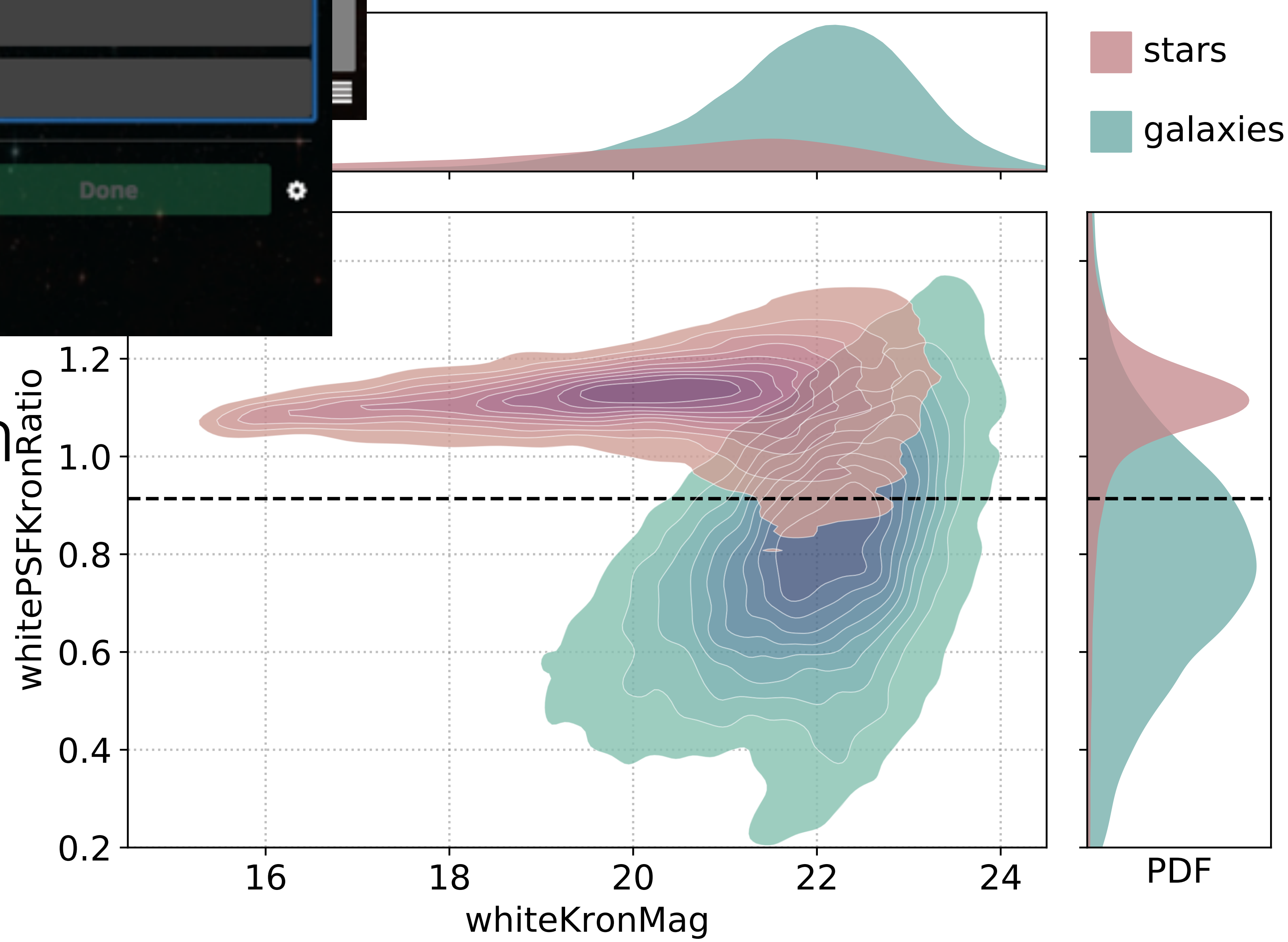
Donalek, .., Mahabal, … arxiv:1310.1976

Machine Learning for ZTF



Adam Miller

Star-galaxy separation

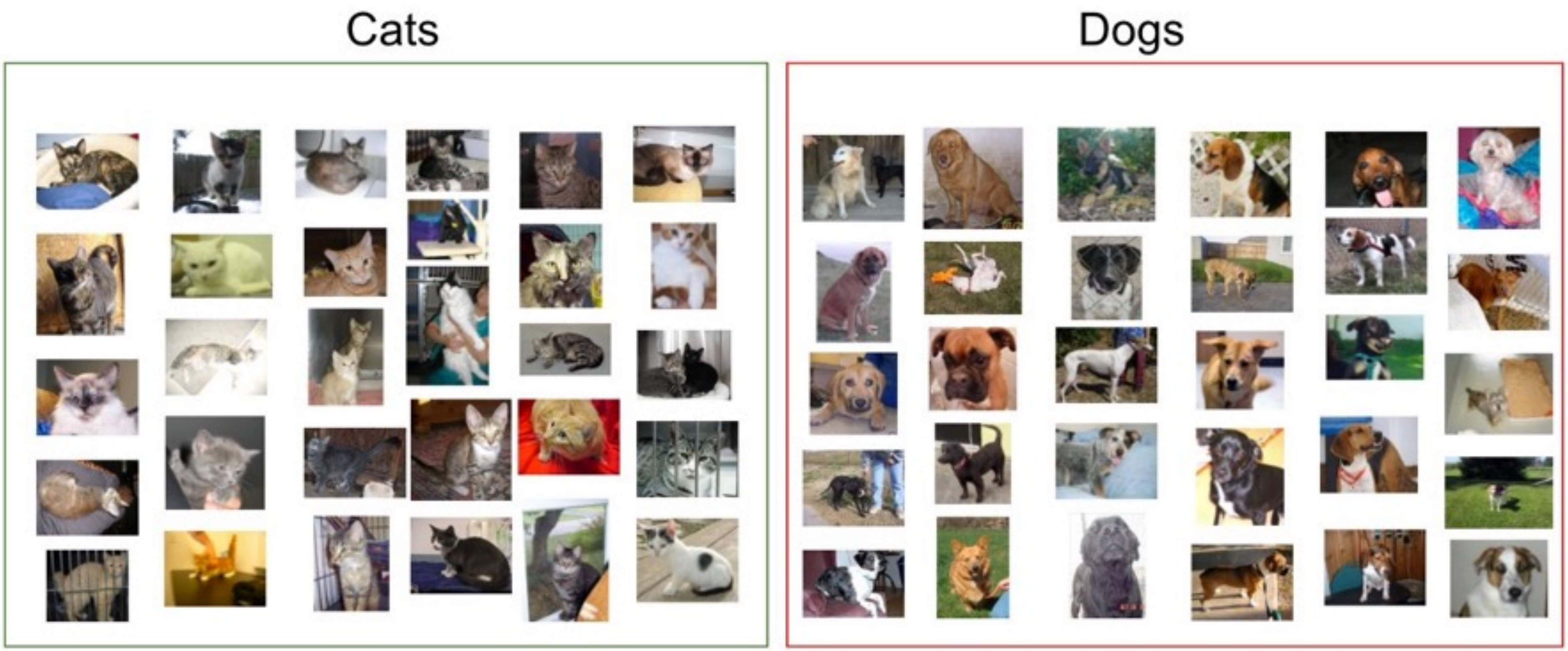


Real-Bogus classification

Citizen Science (zooniverse)
+ random forests

Umaa Rebbapragada

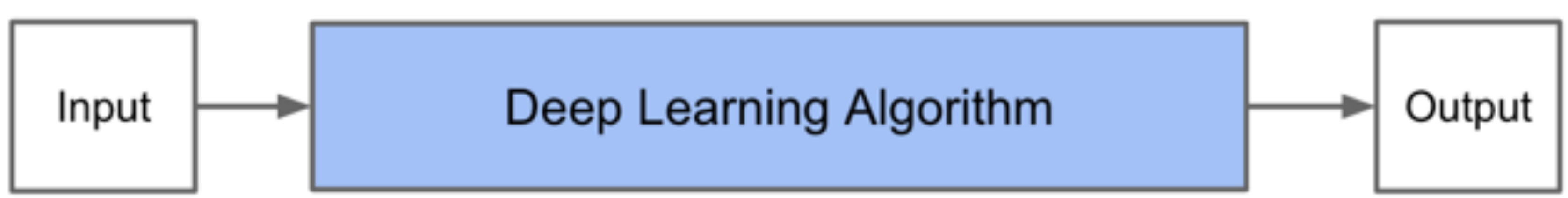
ZTF



Sample of cats & dogs images from Kaggle Dataset

Traditional Machine Learning Flow

Promise:
Works better

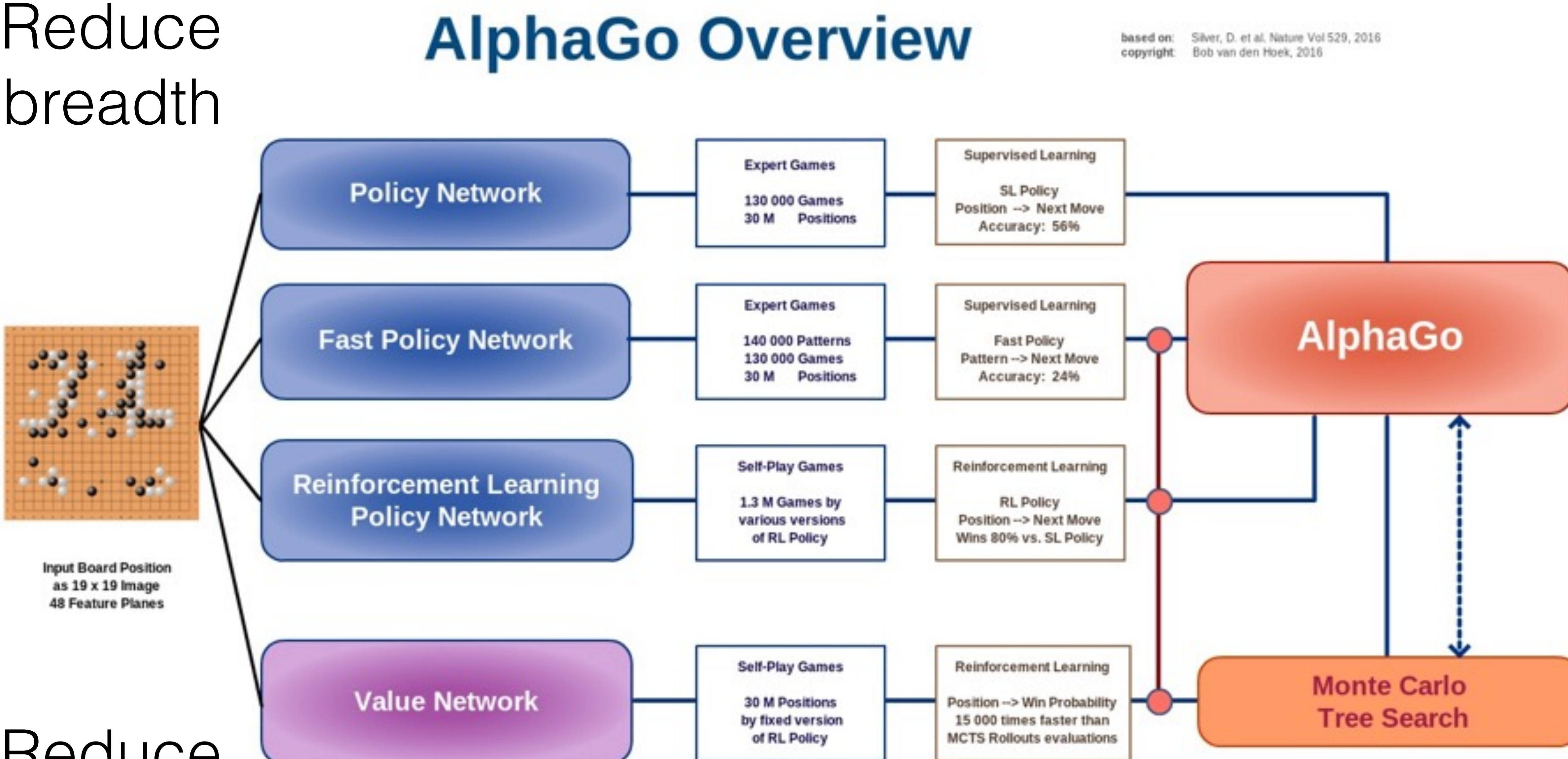


Deep Learning Flow

Pitfall:
Blacker box

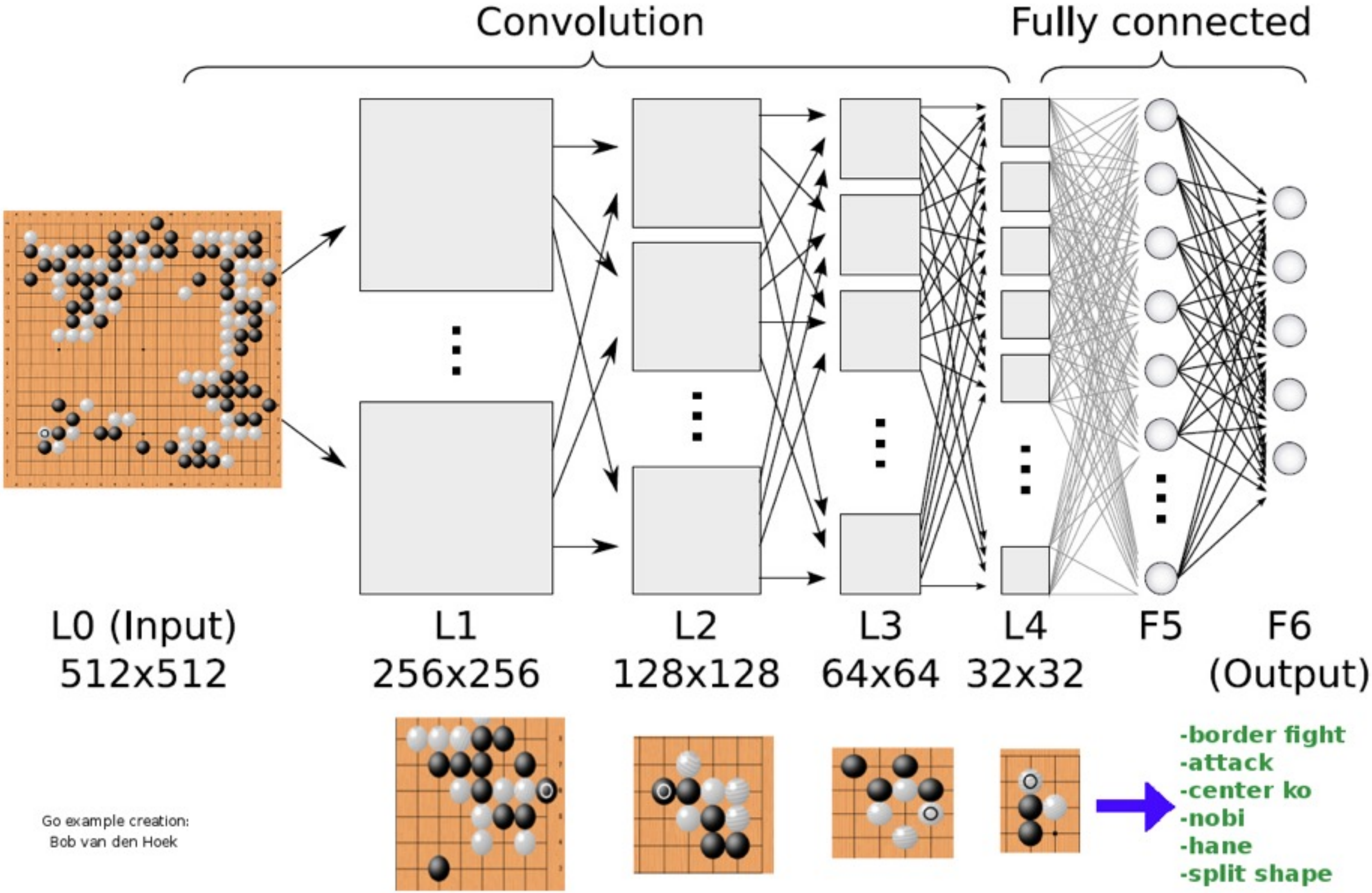
Adil Moujahid

Reduce breadth



Reduce depth

http://deeplearningskysthelimit.blogspot.com/2016/04/part-2-alphago-under-magnifying-glass.html

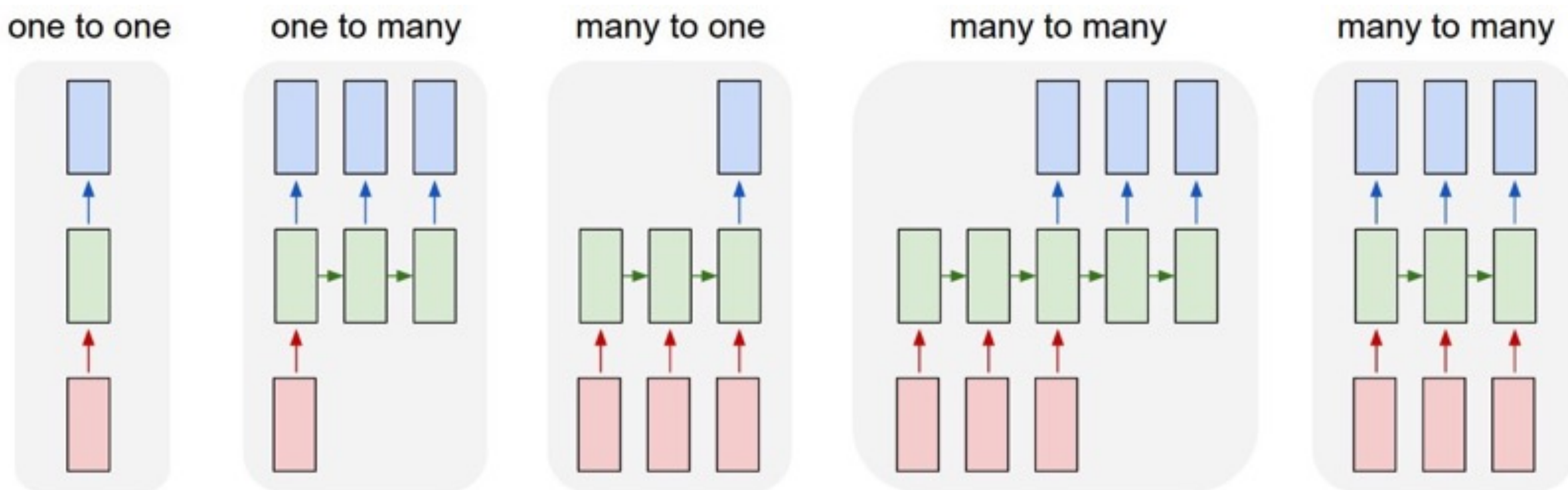


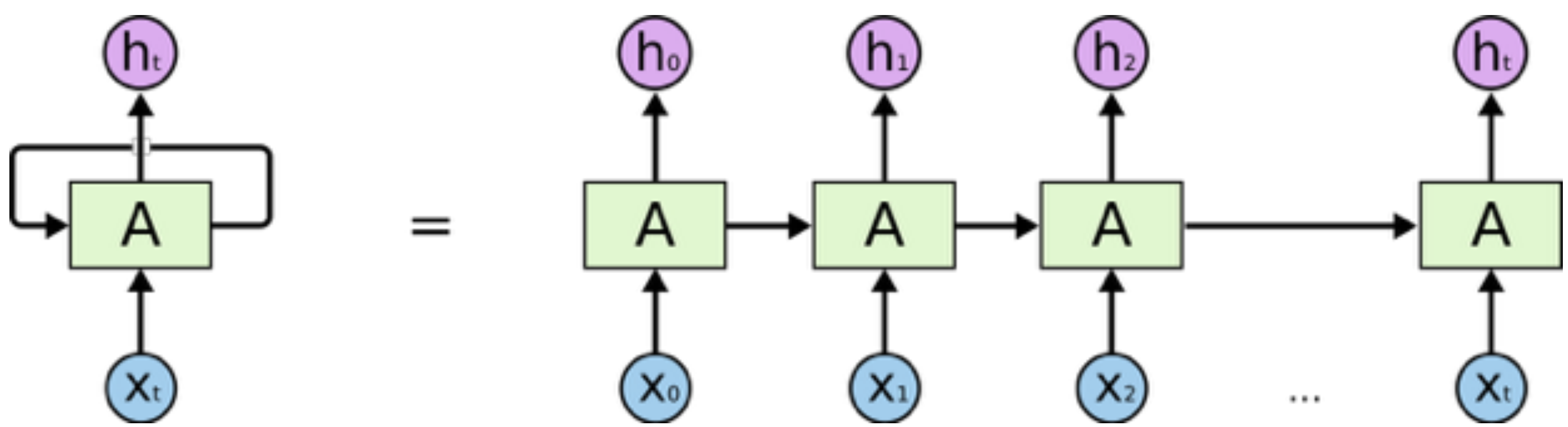
http://gobase.org/online/intergo/?query=%22hane%20nobi%22

http://deeplearningskysthelimit.blogspot.com/2016/04/part-2-alphago-under-magnifying-glass.html

non-image deep networks

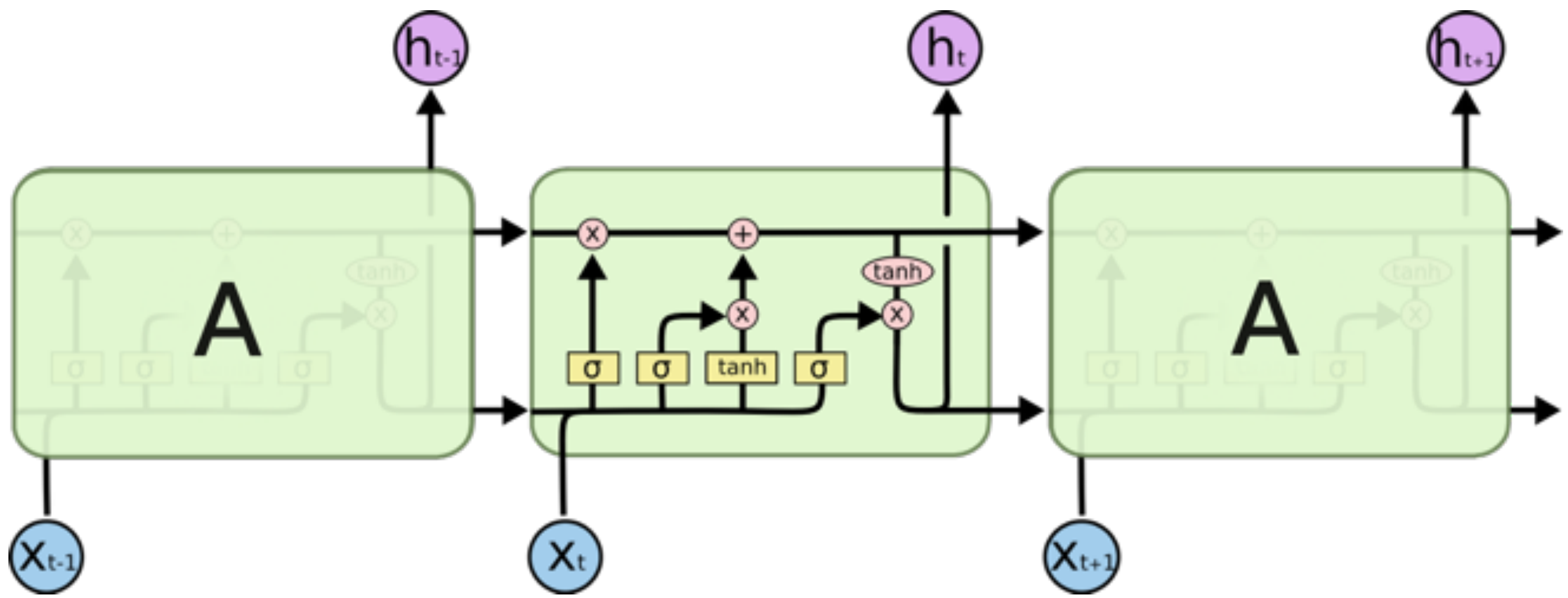
Recurrent Neural Networks





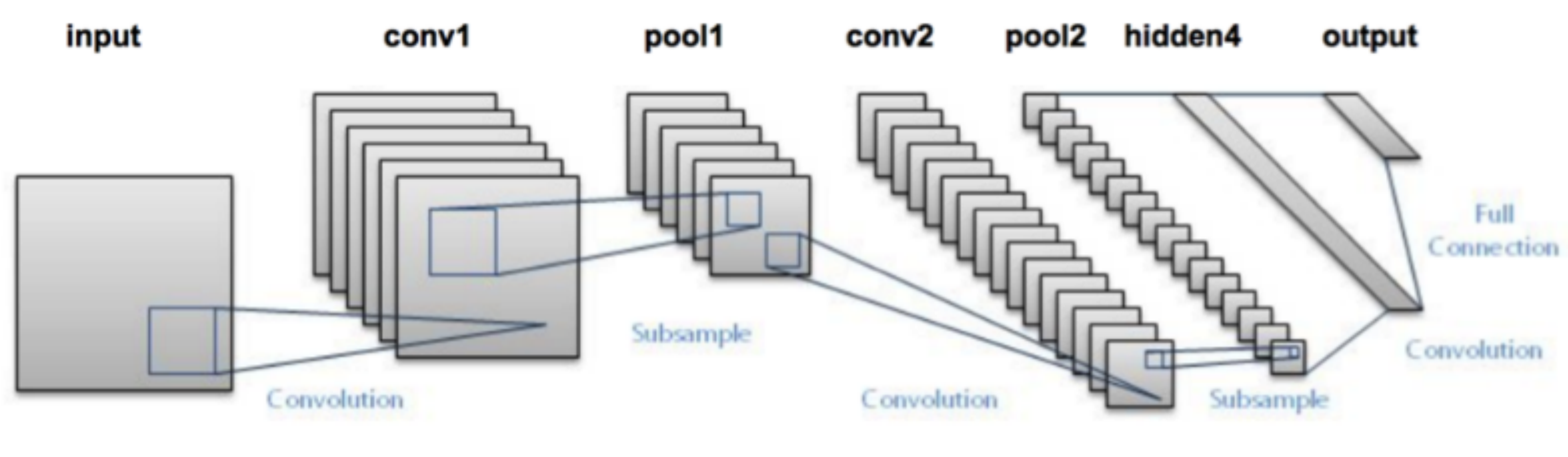
http://colah.github.io/posts/2015-08-Understanding-LSTMs/

Long Short Term Memory (LSTM)



Can be used for light-curves and other time-series

Convolutional network (single slide) primer



analyticsvidhya.com

INPUT IMAGE

18	54	51	239	244	188
55	121	75	78	95	88
35	24	204	113	109	221
3	154	104	235	25	130
15	253	225	159	78	233
68	85	180	214	245	0

WEIGHT

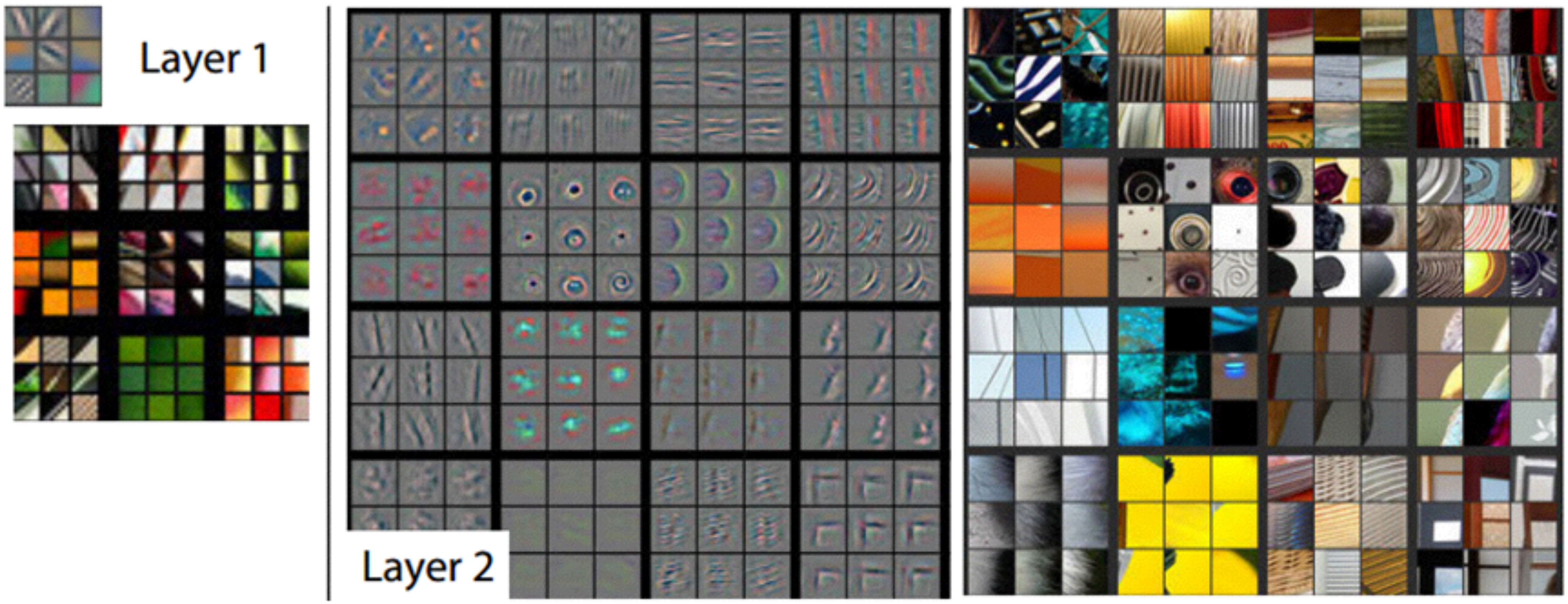
1	0	1
0	1	0
1	0	1

429

429	505	686	856
261	792	412	640
633	653	851	751
608	913	713	657

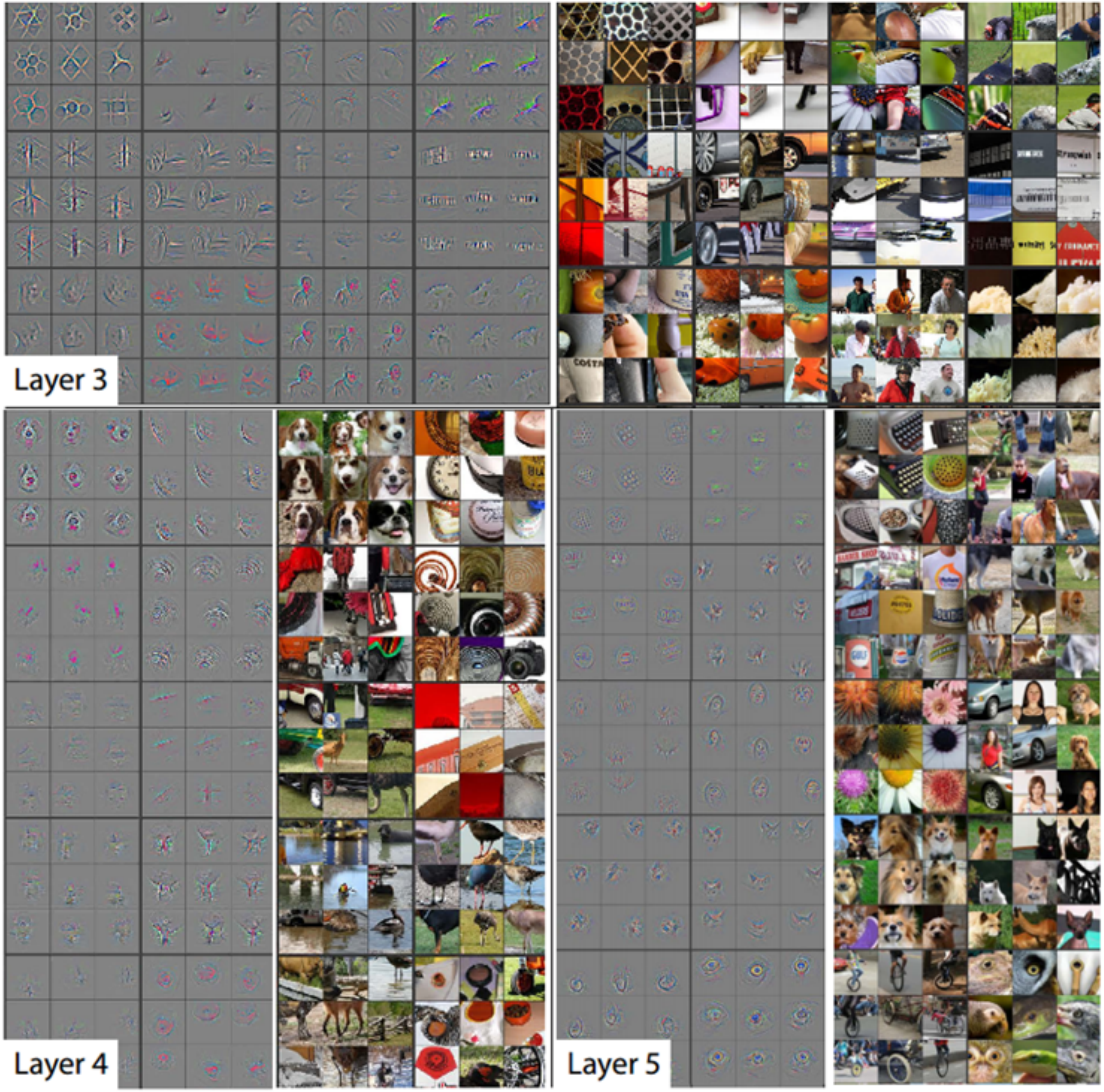
792	856
913	851

deconvnets

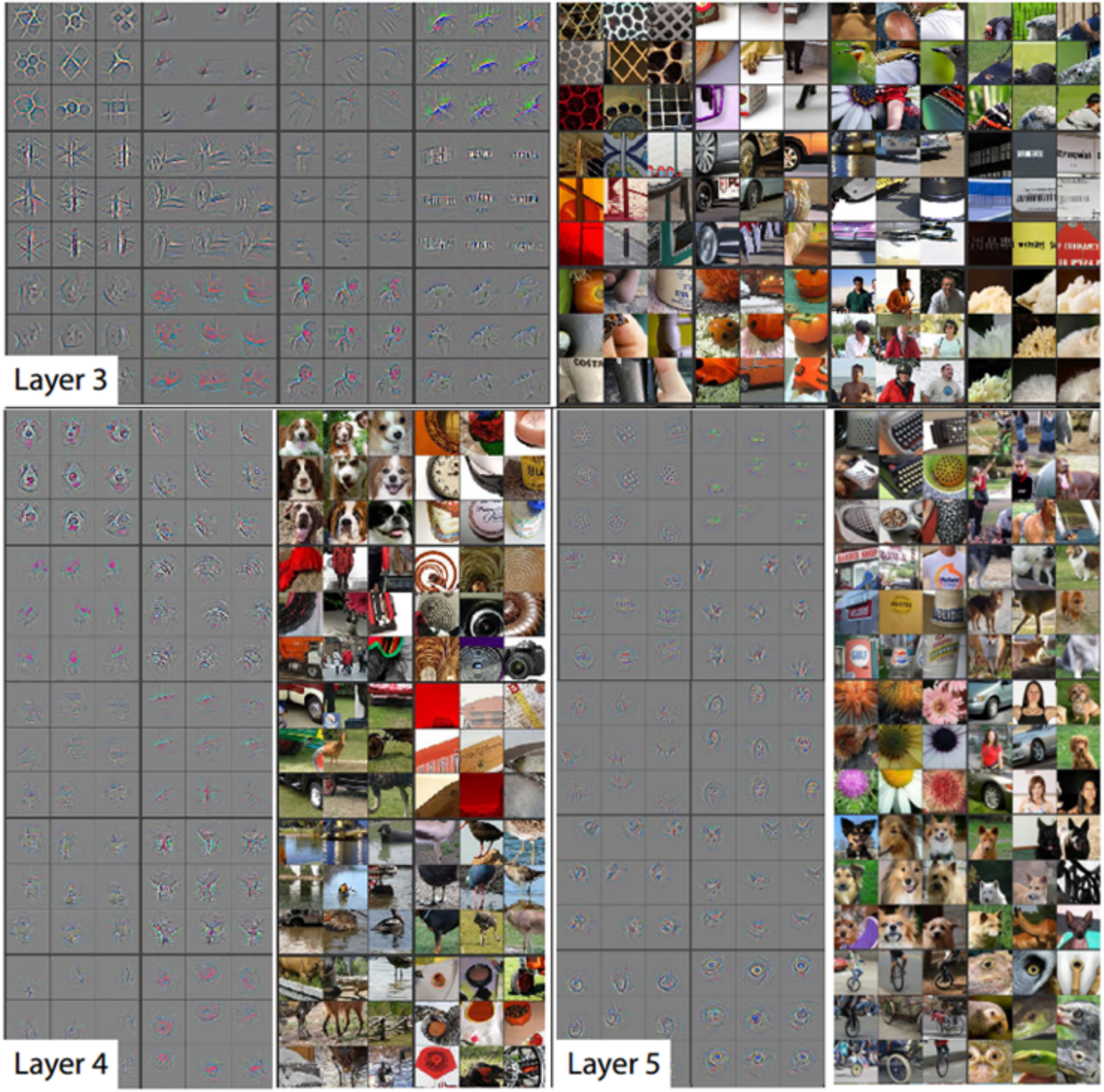


Visualizations of Layer 1 and 2. Each layer illustrates 2 pictures, one which shows the filters themselves and one that shows what part of the image are most strongly activated by the given filter. For example, in the space labled Layer 2, we have representations of the 16 different filters (on the left)

https://arxiv.org/pdf/1311.2901v3.pdf



 Visualizations of Layers 3, 4, and 5 https://arxiv.org/pdf/1311.2901v3.pdf

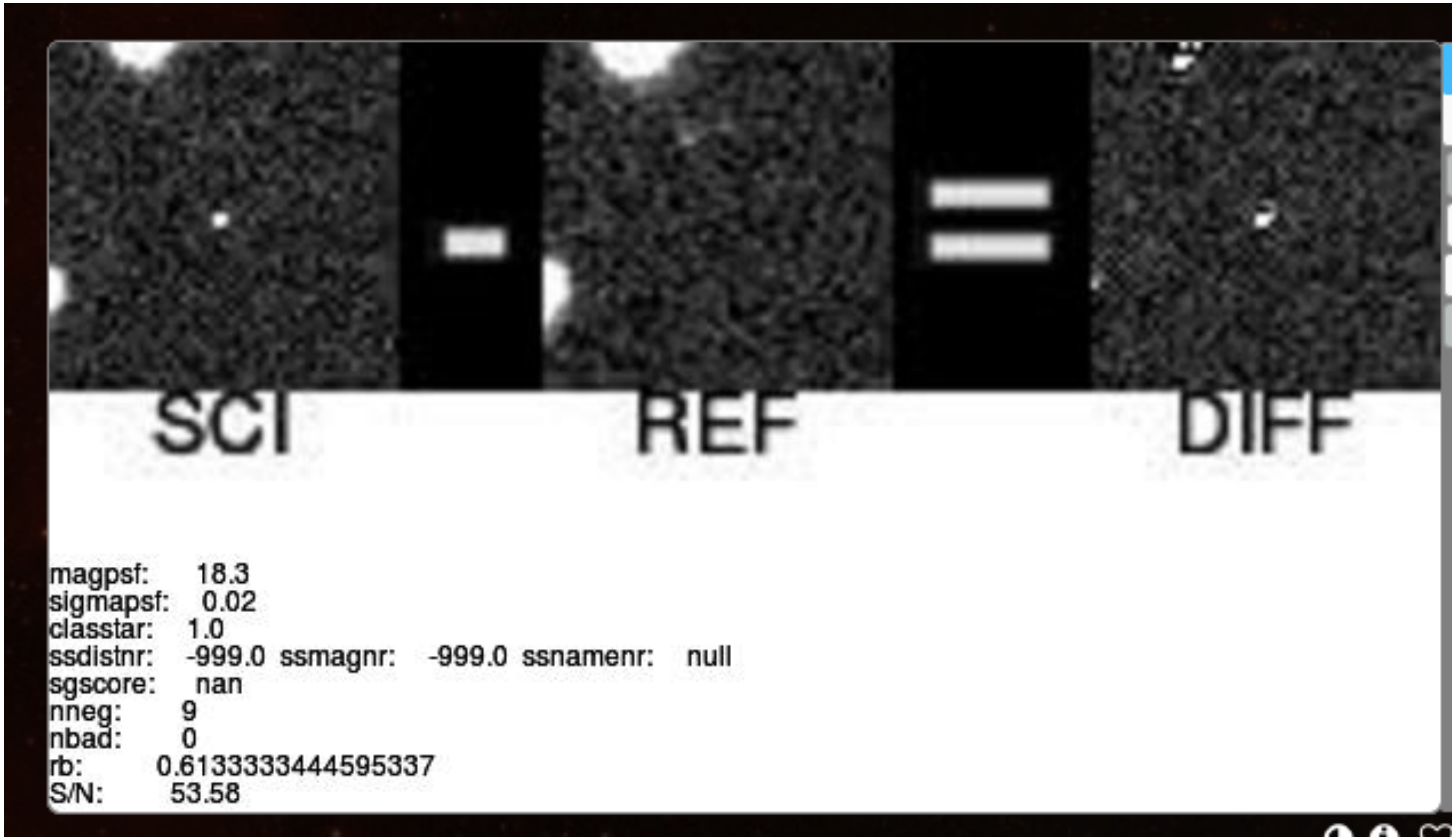


Promise: New Features

Visualizations of Layers 3, 4, and 5

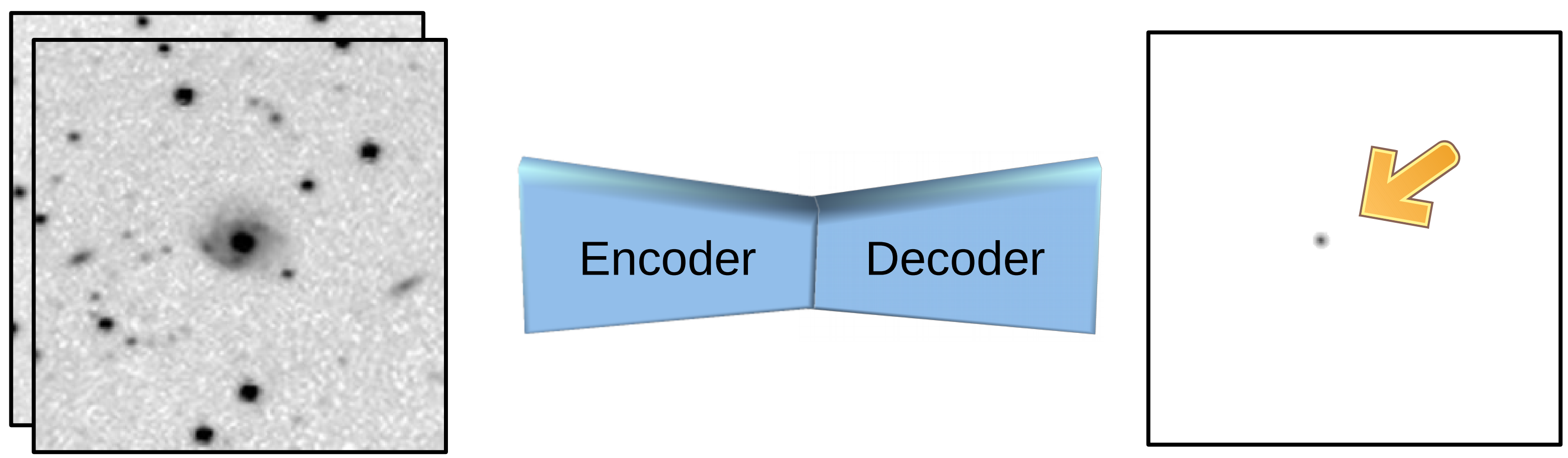
https://arxiv.org/pdf/1311.2901v3.pdf

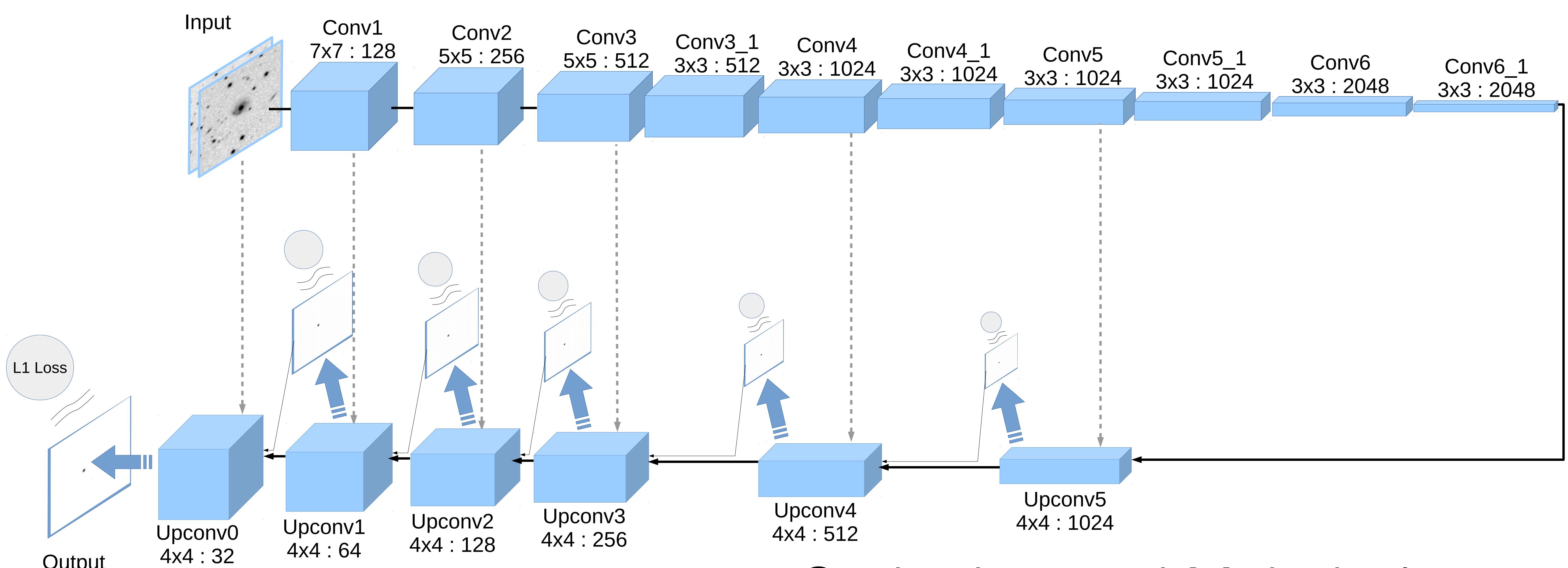
Transient hunting through image differencing



reorient, match background, match PSF, eliminate artifacts

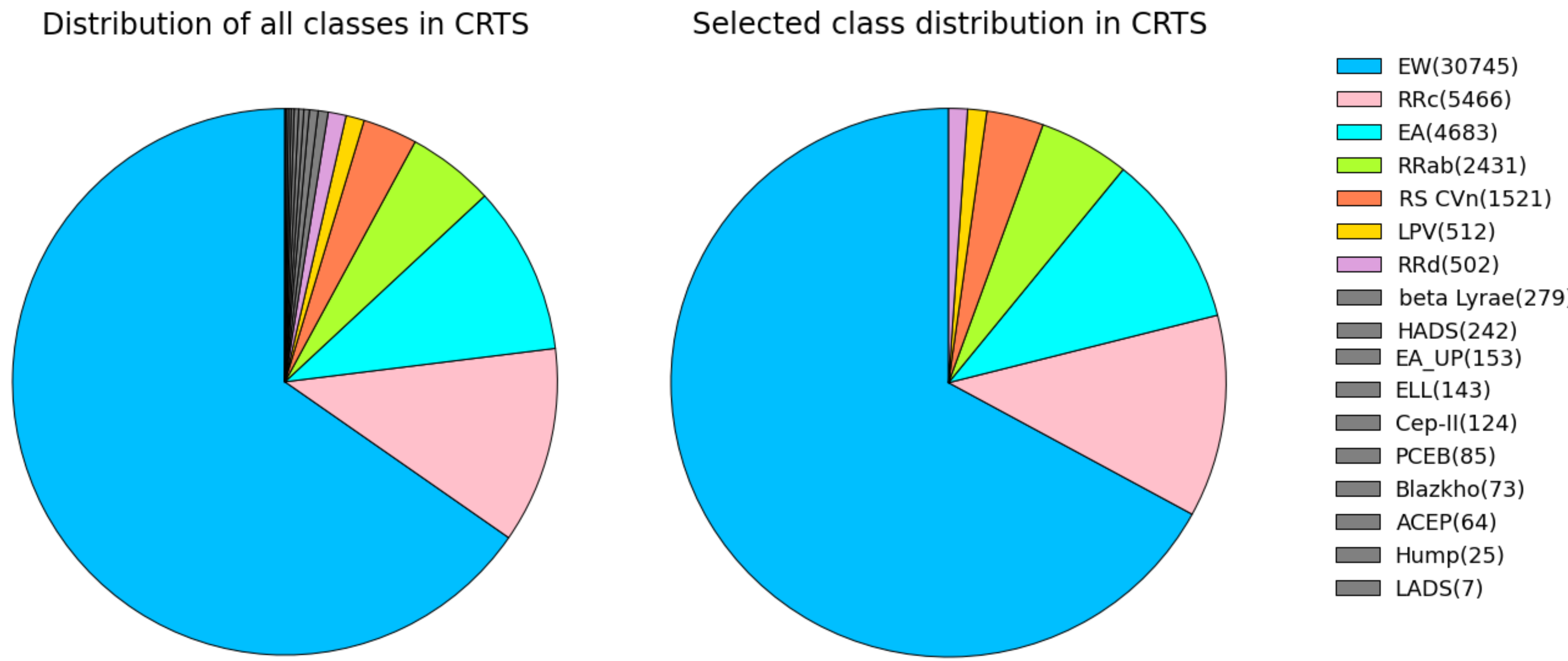
Image subtraction for hunting transients without subtraction





Sedaghat and Mahabal, 2017

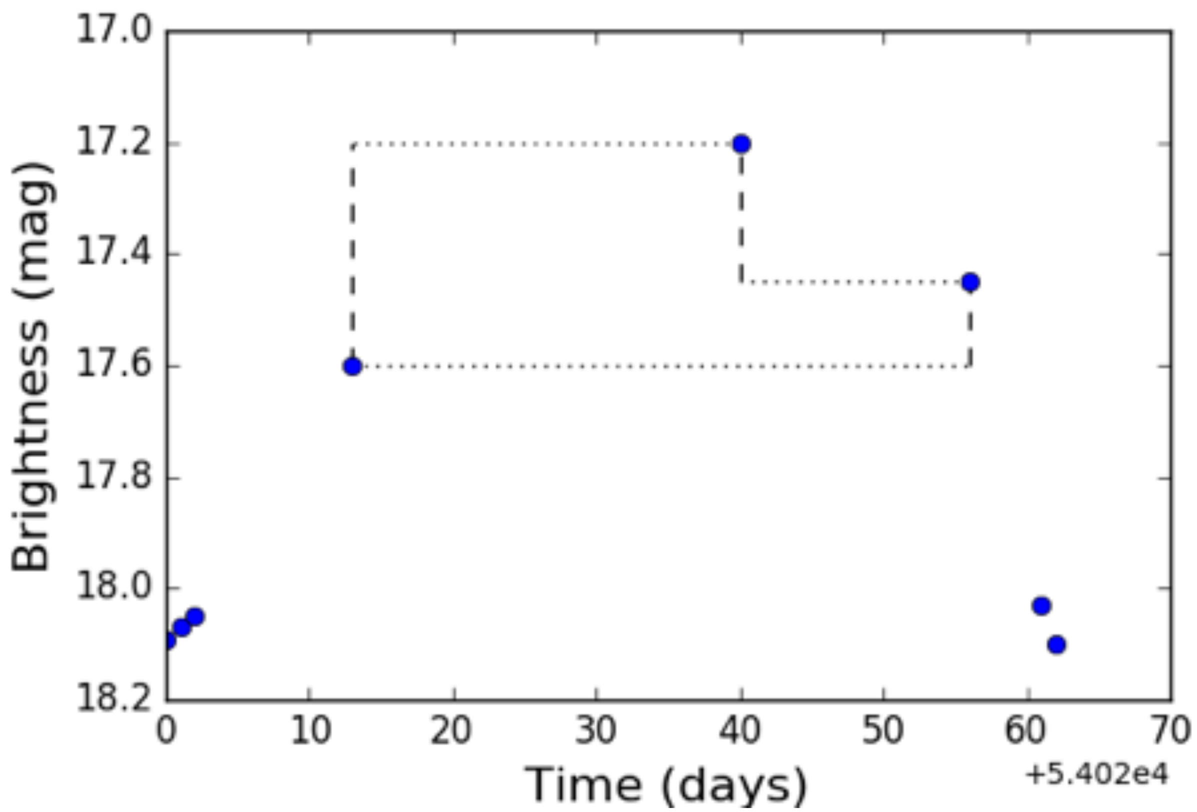
50K Periodic Variables from CRTS



Drake et al. 2014

(dmdt) Image representation

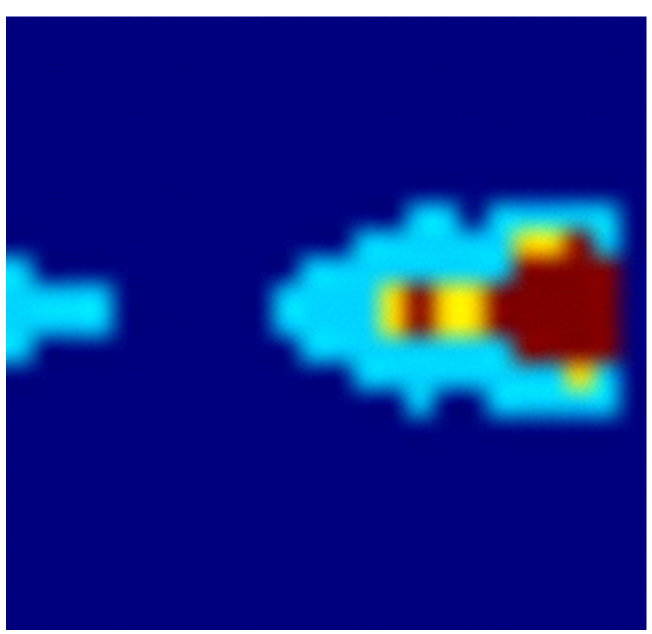
Mahabal et al., 2017

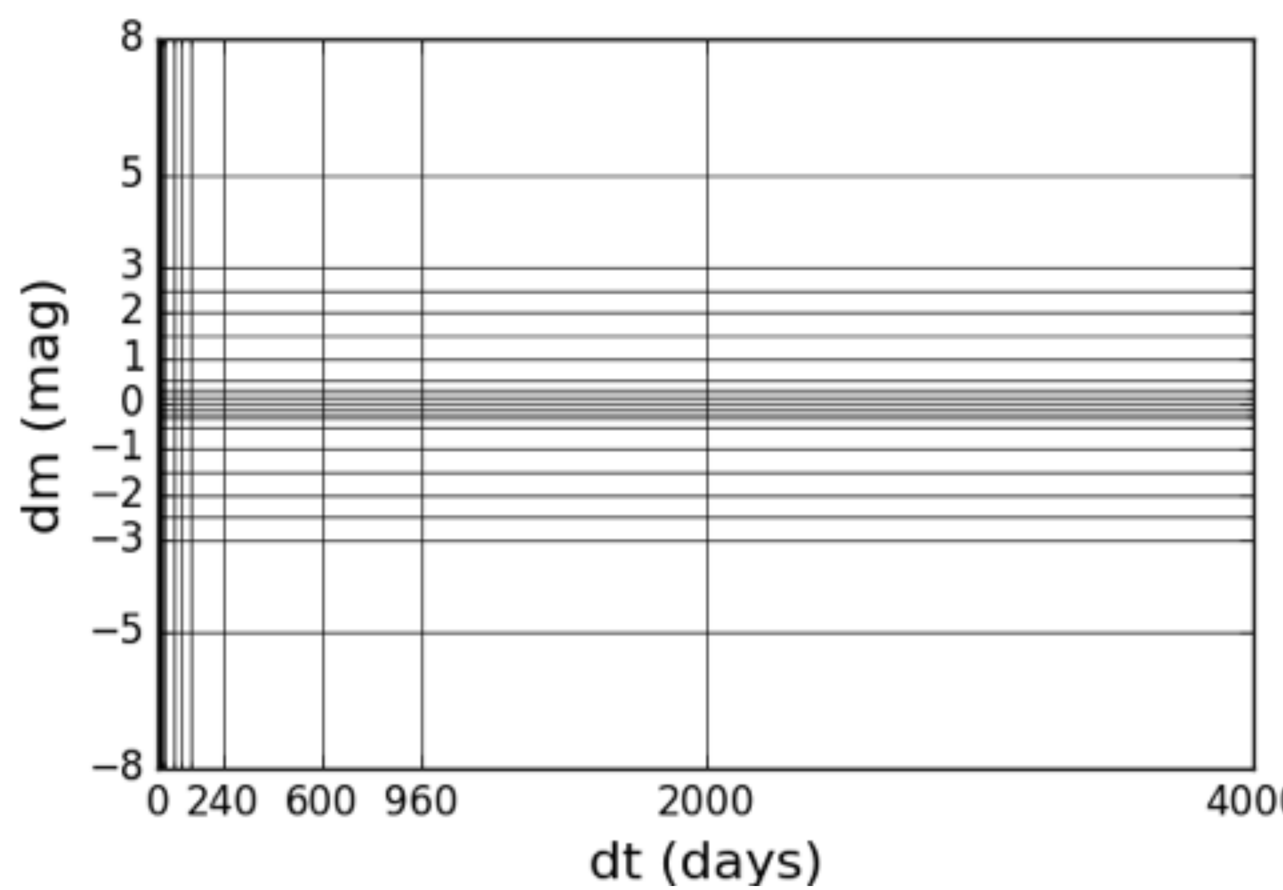


light curve with n points

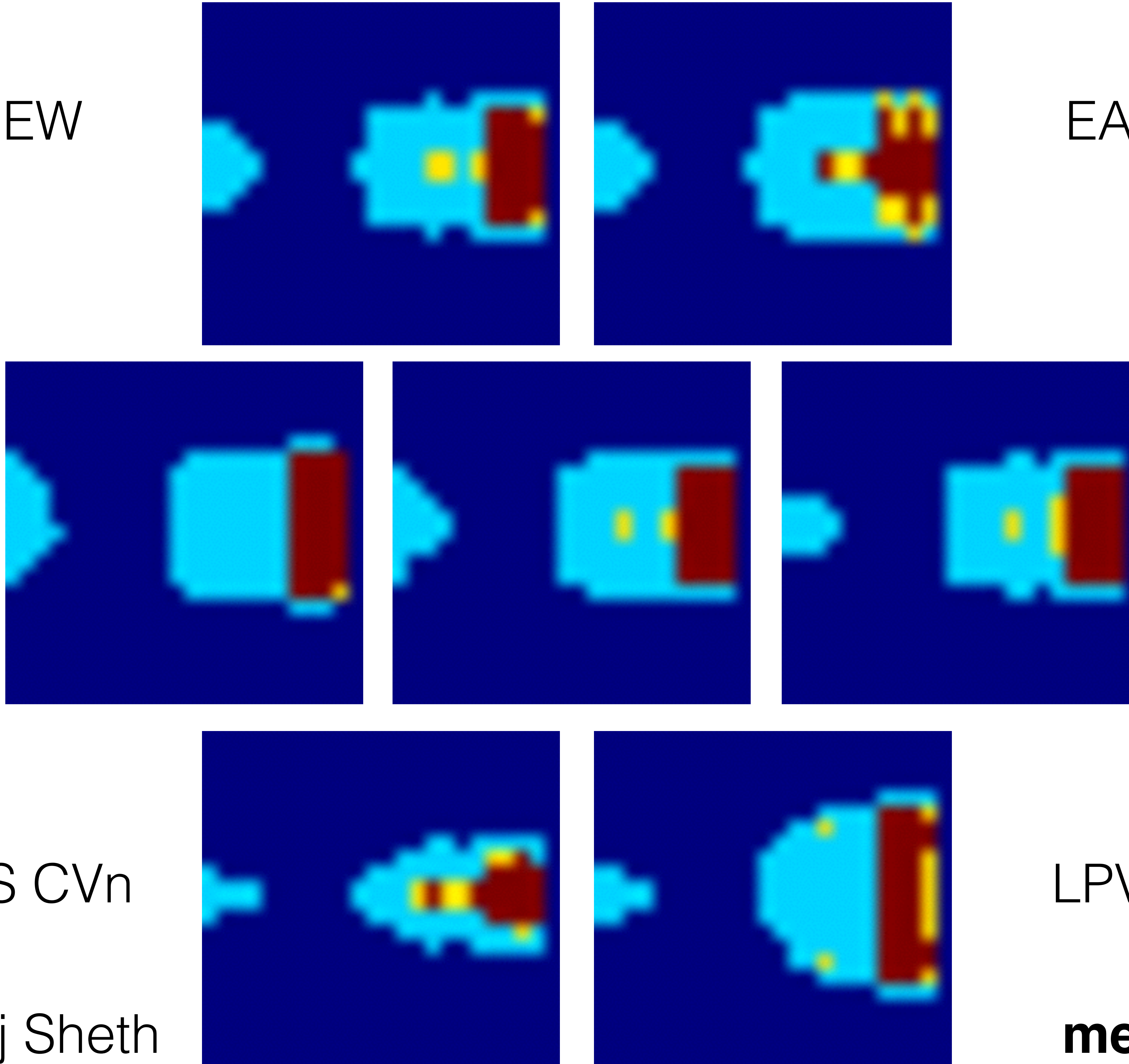
23 x 24 output grid

n * (n-1)/2 points



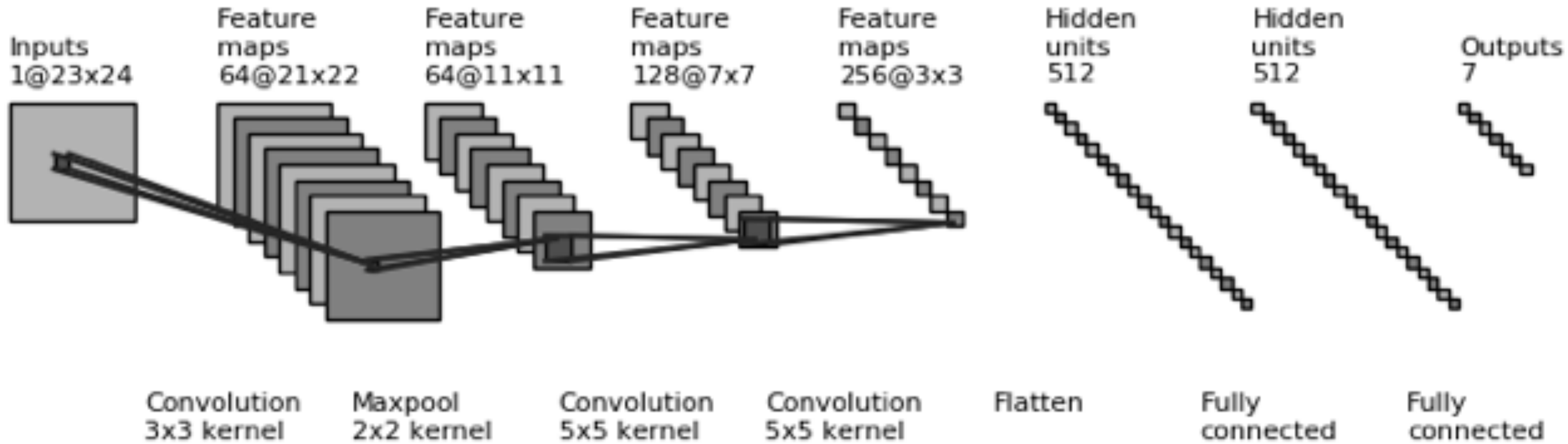


Area equalized pixels



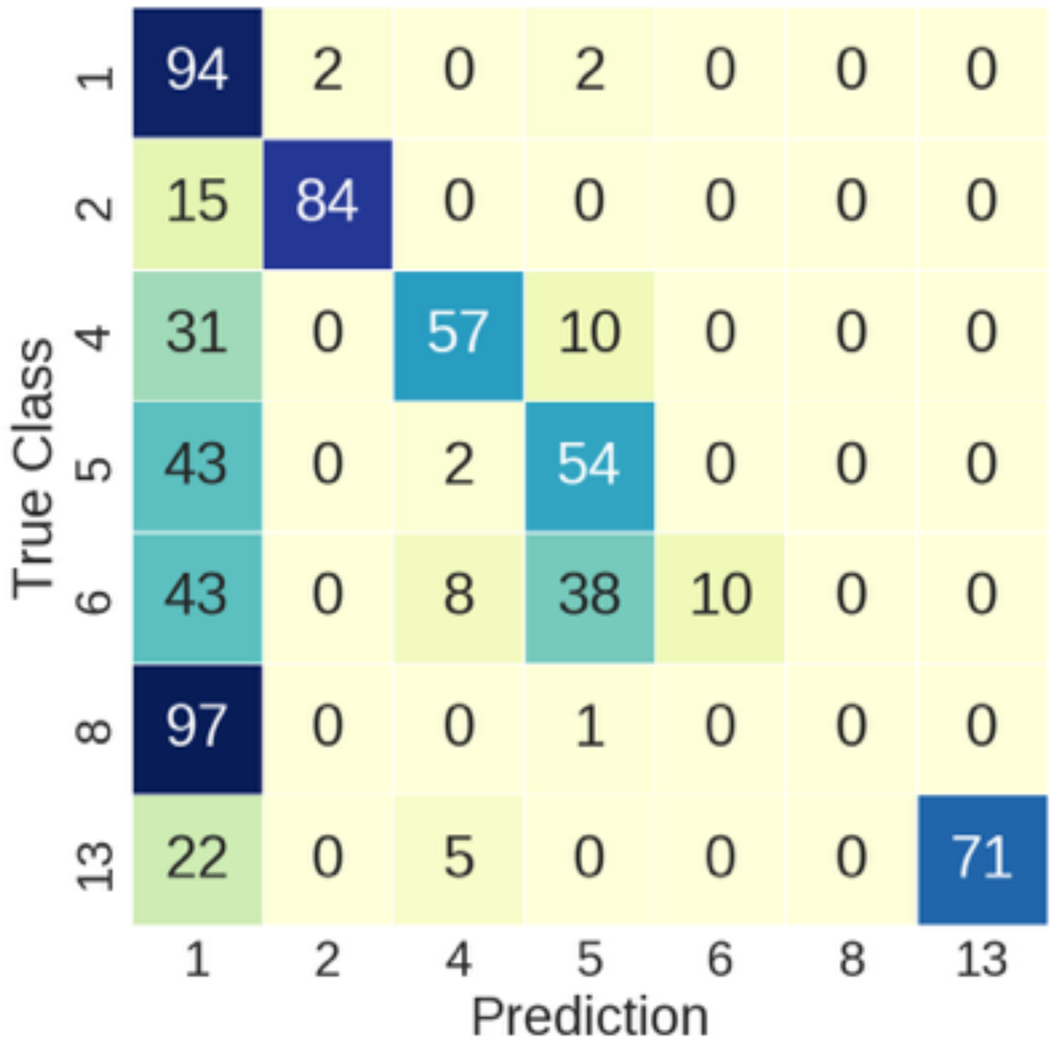
Kshiteej Sheth

Network architecture



In this case shallow works well too

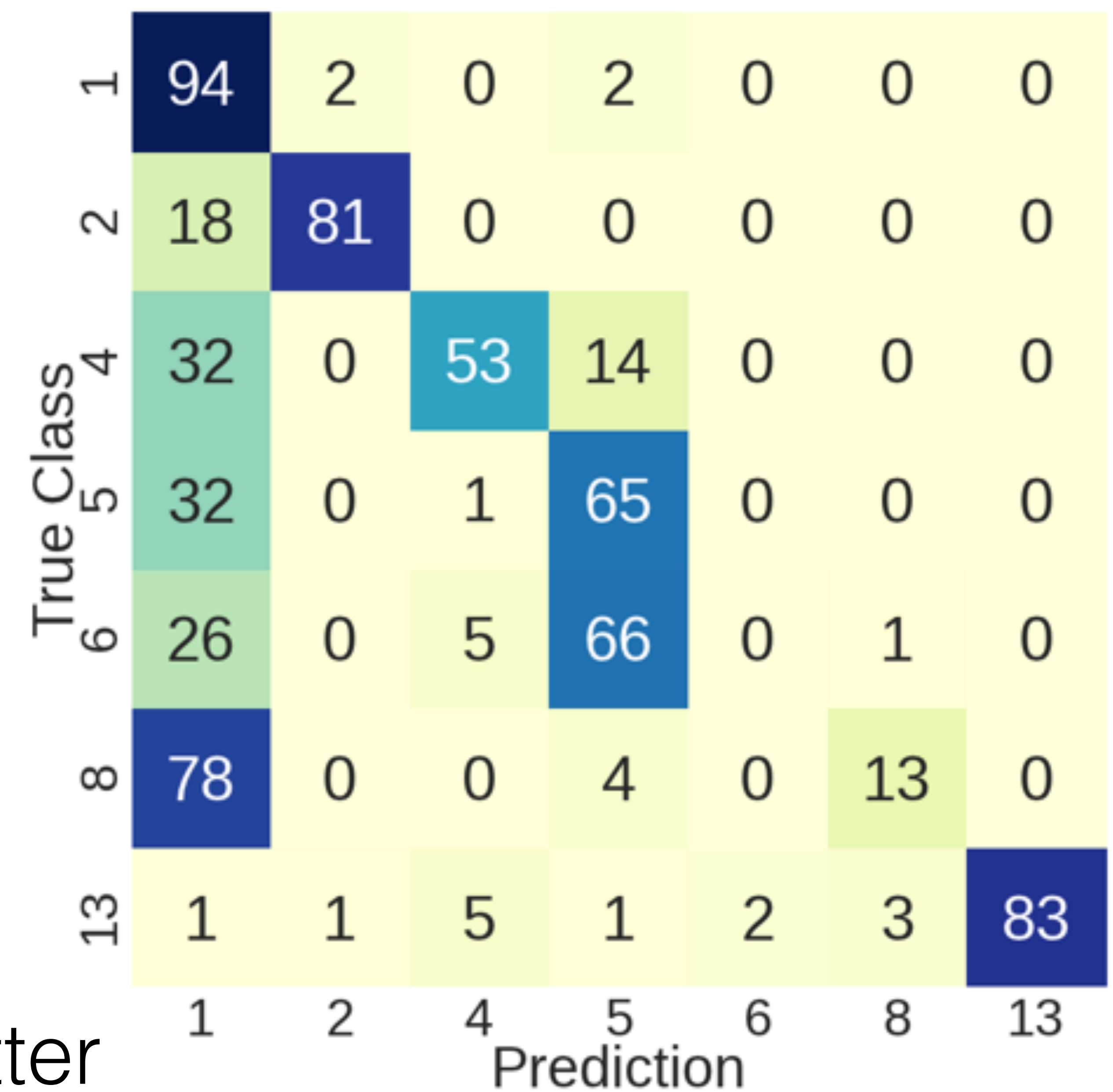
```
layers = [
    InputLayer,
    Conv2DLayer(32, size:3x3, rectify),
    DropoutLayer(0.1),
    DenseLayer(128),
    DropoutLayer(0.25),
    DenseLayer(128),
    DenseLayer(all, softmax),
```



Random Forest using standard features

no features
no dimensionality reduction
comparable results

Convolutional Network



1	EW/EB
2	EA
4	RRab
5	RRc
6	RRd
8	RS CVn
13	LPV

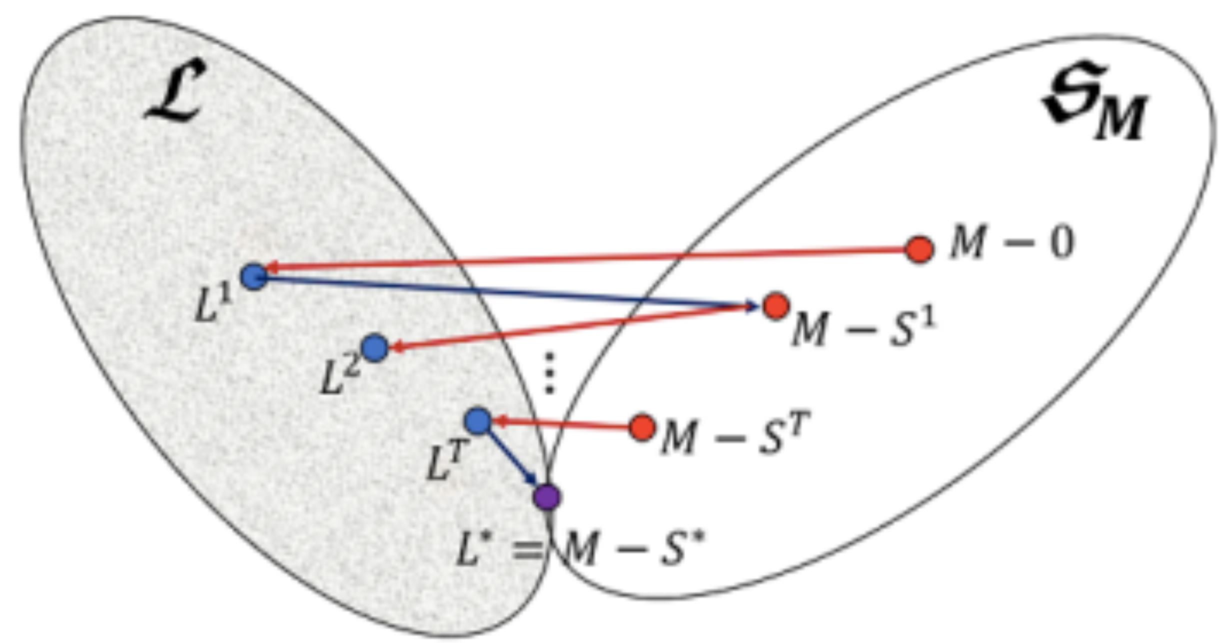
Binary probabilities are better

$dmdt$–image $= b + ci + s$

- background (survey, cadence)
- class background
- individual object (specific)

$$\underset{L,S}{\mathbf{Min}} \|M - L - S\|_2$$

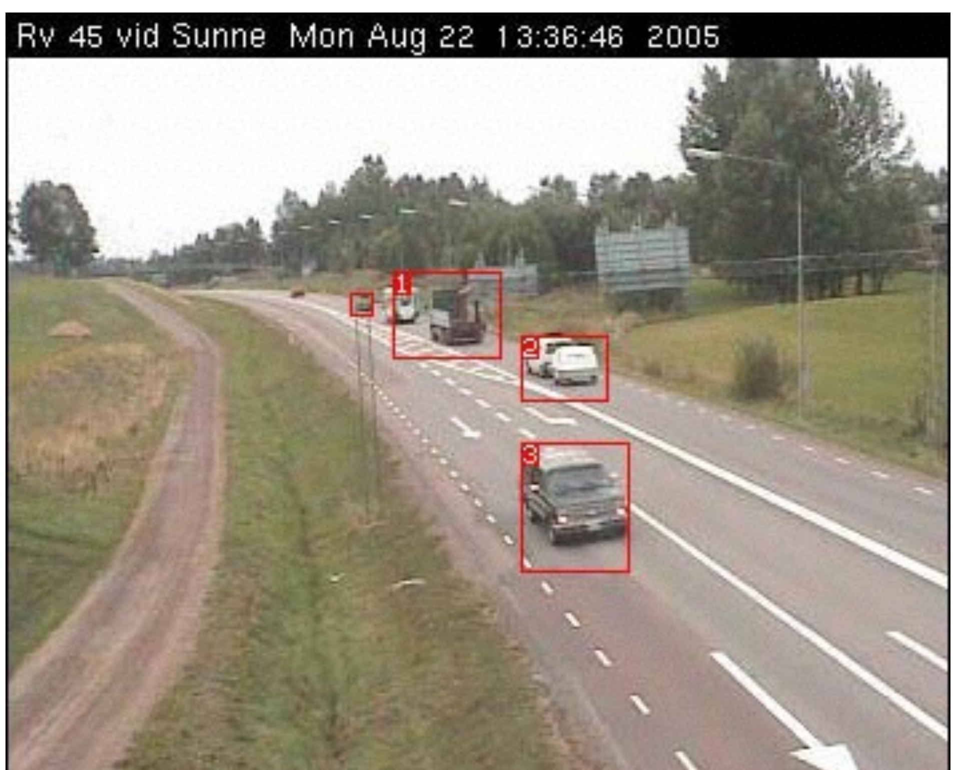
1. L lies in the set of low-rank matrices,
2. S lies in the set of sparse matrices.



non-convex robust PCA
Netrapalli et al., 2014

Video Surveillance Anology

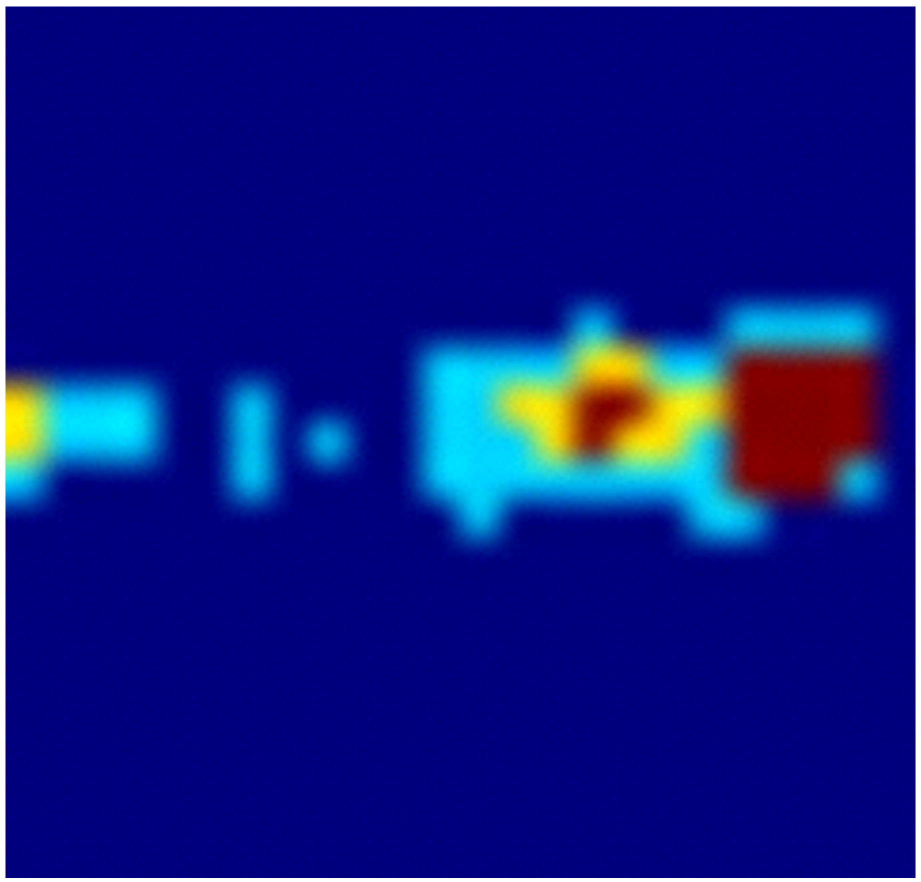
non-convex robust PCA
Netrapalli et al., 2014

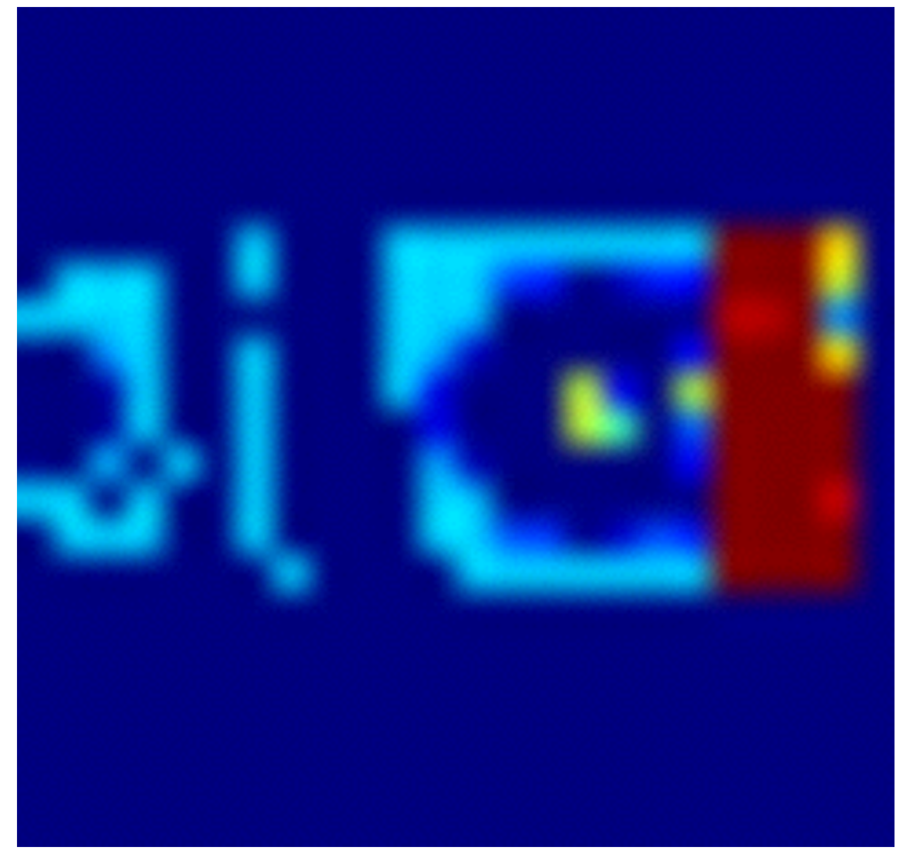


Andrew Kirillov

Each class is like a different road
Each individual object has/is perturbations over it

EW/EB separation?

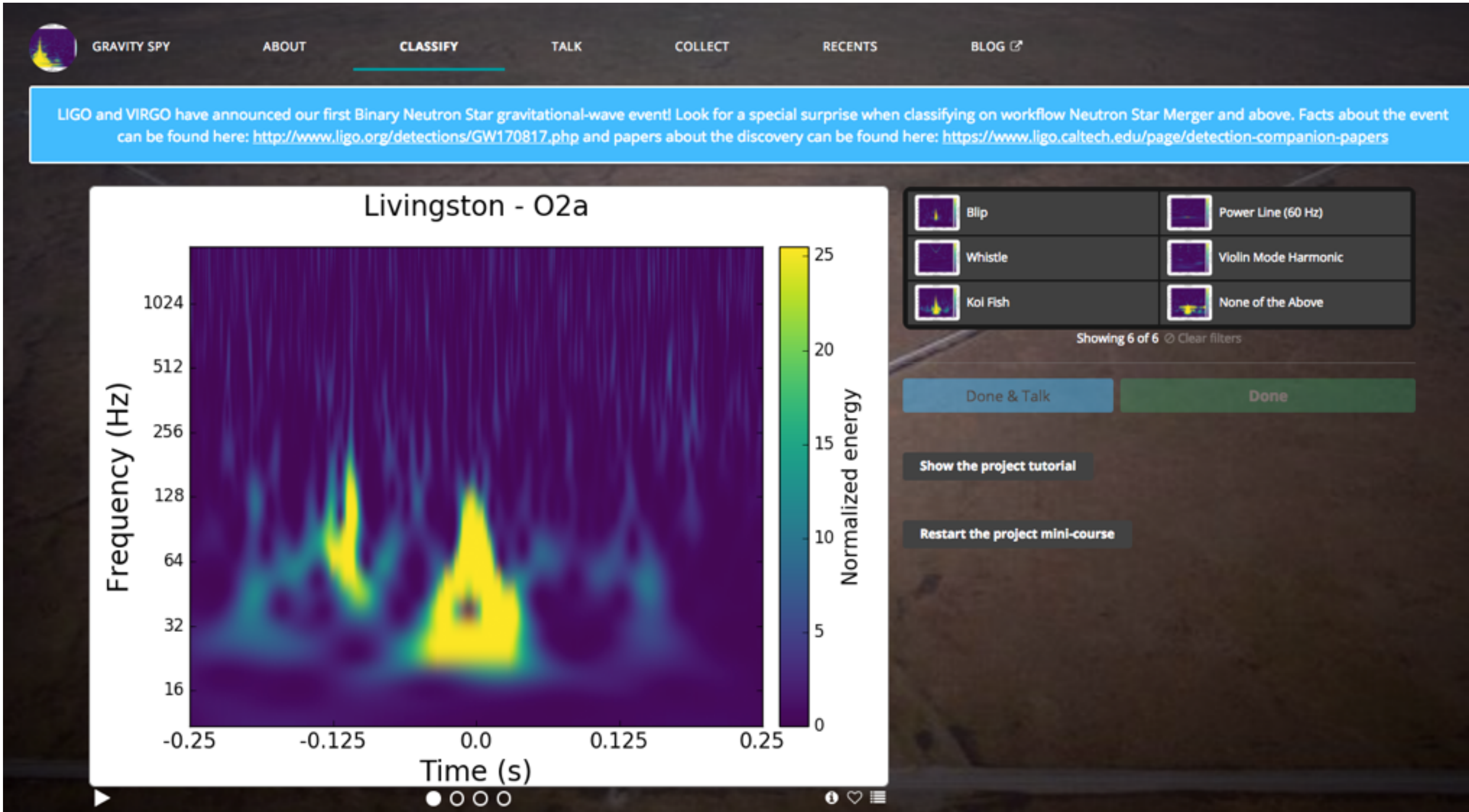




Two separate backgrounds emerged for class 1

Gravity Spy

Zevin et al., 2017



GW deep filtering

diff. defn of SNR?

Daniel George, E. A. Huerta
1711.03121, 1701.00008

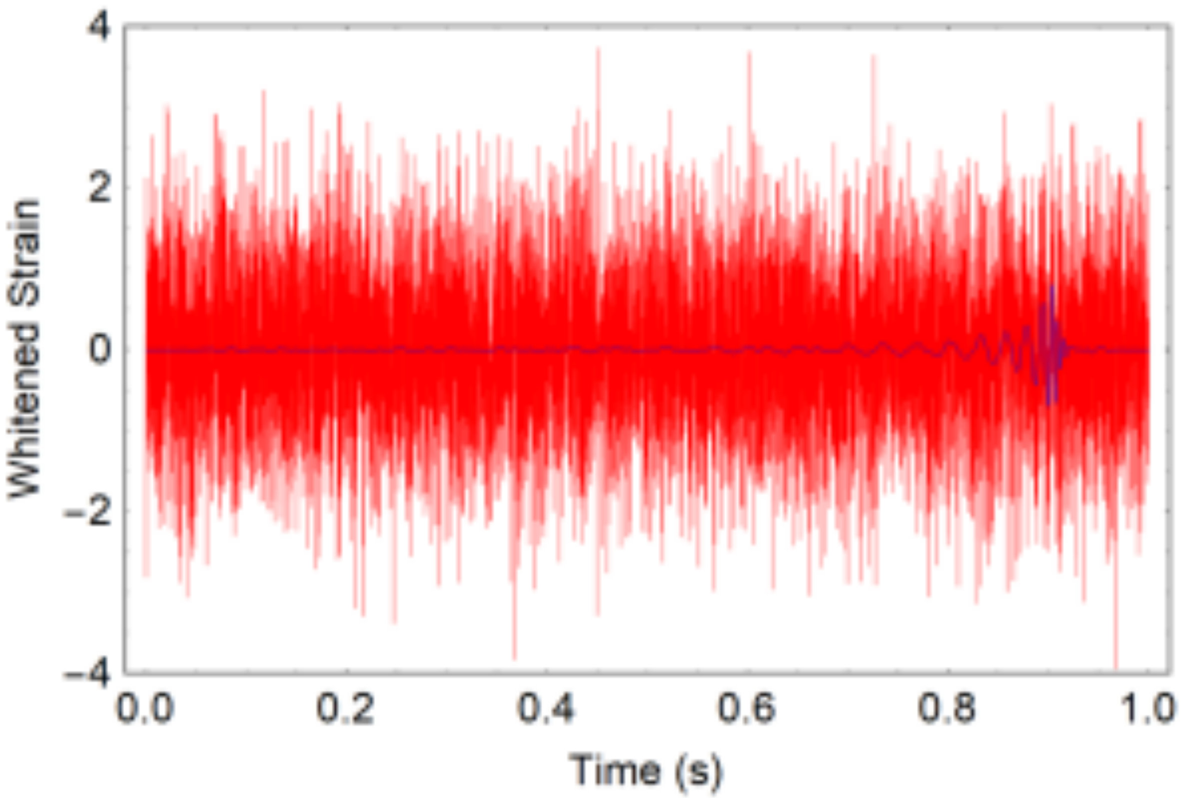


FIG. 1. **Sample signal injected into real LIGO noise.** The red time-series is an example of the input to our `Deep Filtering` algorithm. It contains a hidden BBH GW signal (blue) from our test set which was superimposed in real LIGO noise from the test set and whitened. For this injection, the optimal matched-filter SNR = 7.5 (peak power of this signal is 0.65 times the power of background noise). The component masses of the merging BHs are $57M_{\odot}$ and $33M_{\odot}$. The presence of this signal was detected directly from the (red) time-series input with over 99% sensitivity and the source's parameters were estimated with a mean relative error less than 10%.

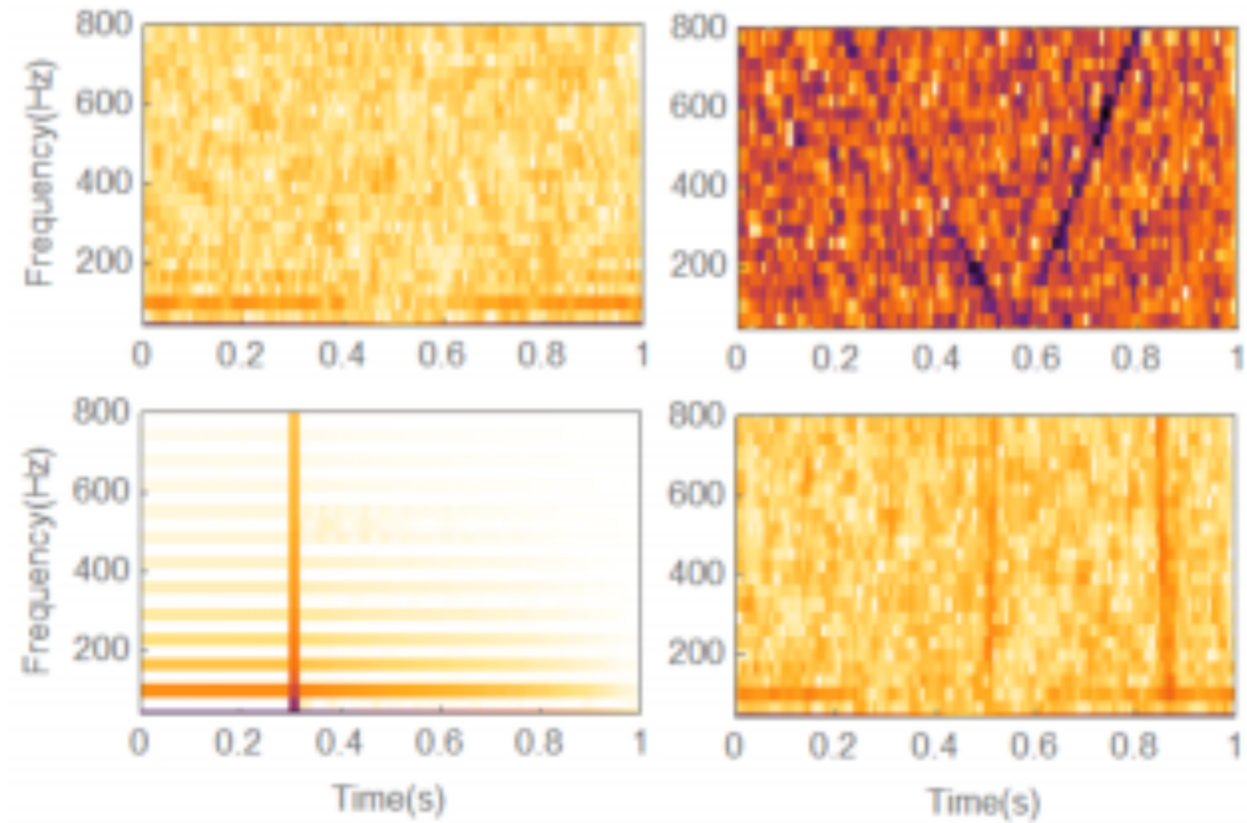
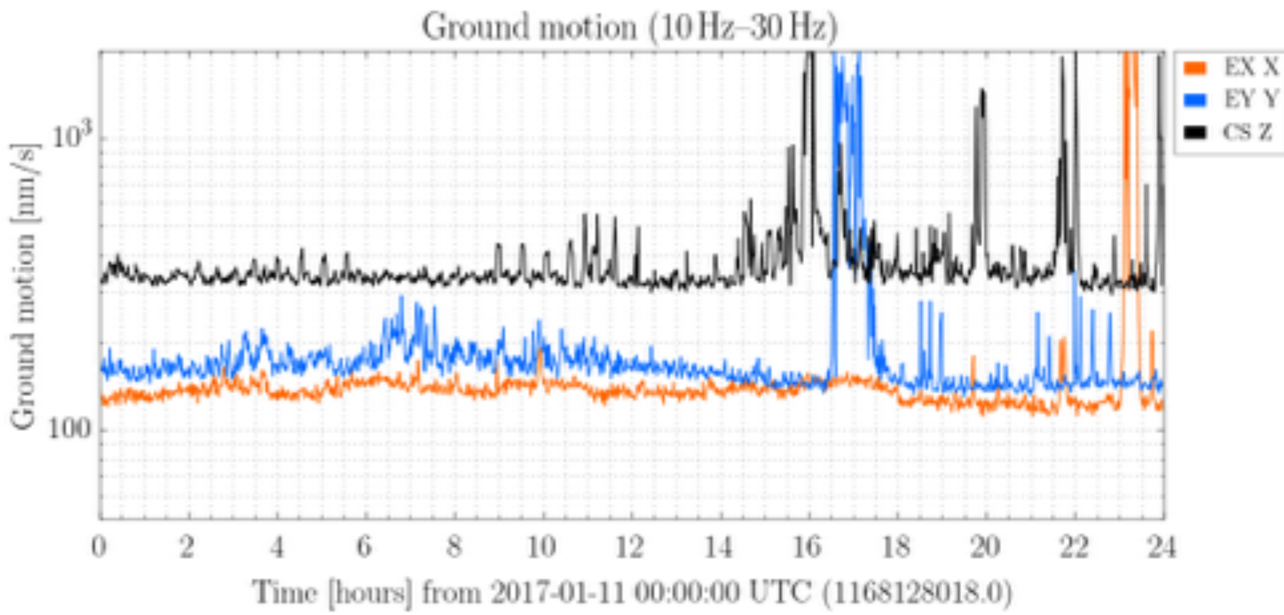
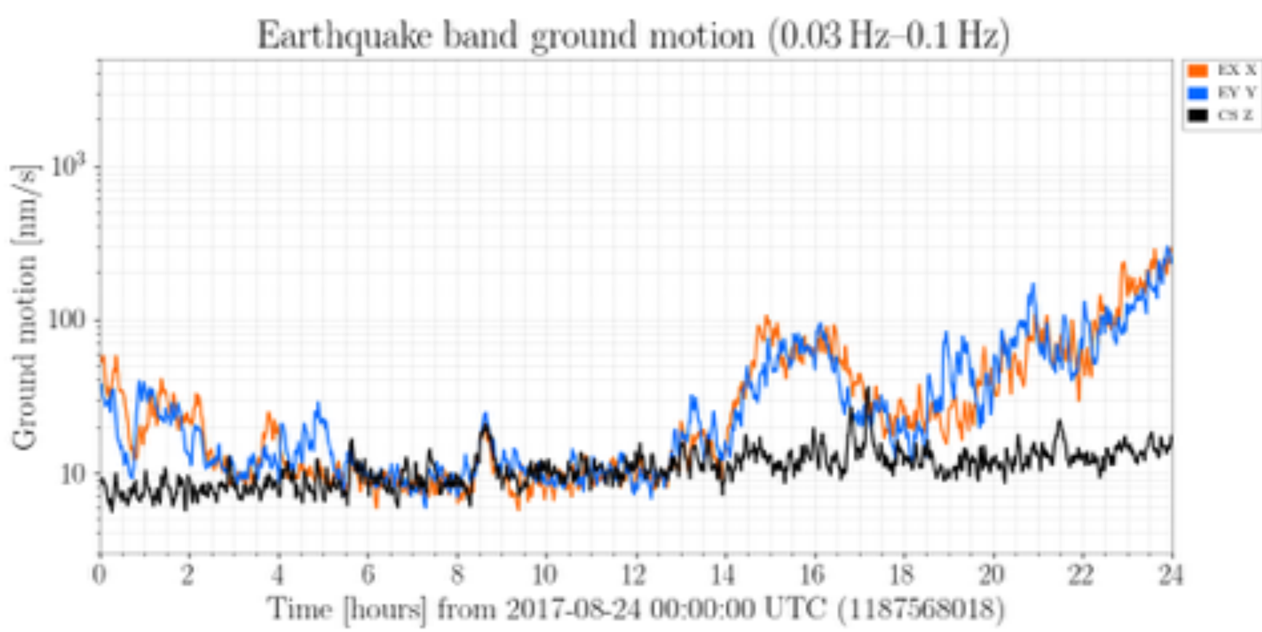


FIG. 2. **Spectrograms of real LIGO noise test samples.** We used signals injected into real data from the LIGO detectors in this article, ensuring that the training and testing sets did not contain noise from the same events. These are some random examples of real glitches that were present in our test set of LIGO noise. The `Deep Filtering` method takes the 1D strain directly as input and is able to correctly classify glitches as noise and detect true GW signals as well as simulated GW signals injected into these highly non-stationary non-Gaussian data streams, with similar sensitivity compared to matched-filtering.

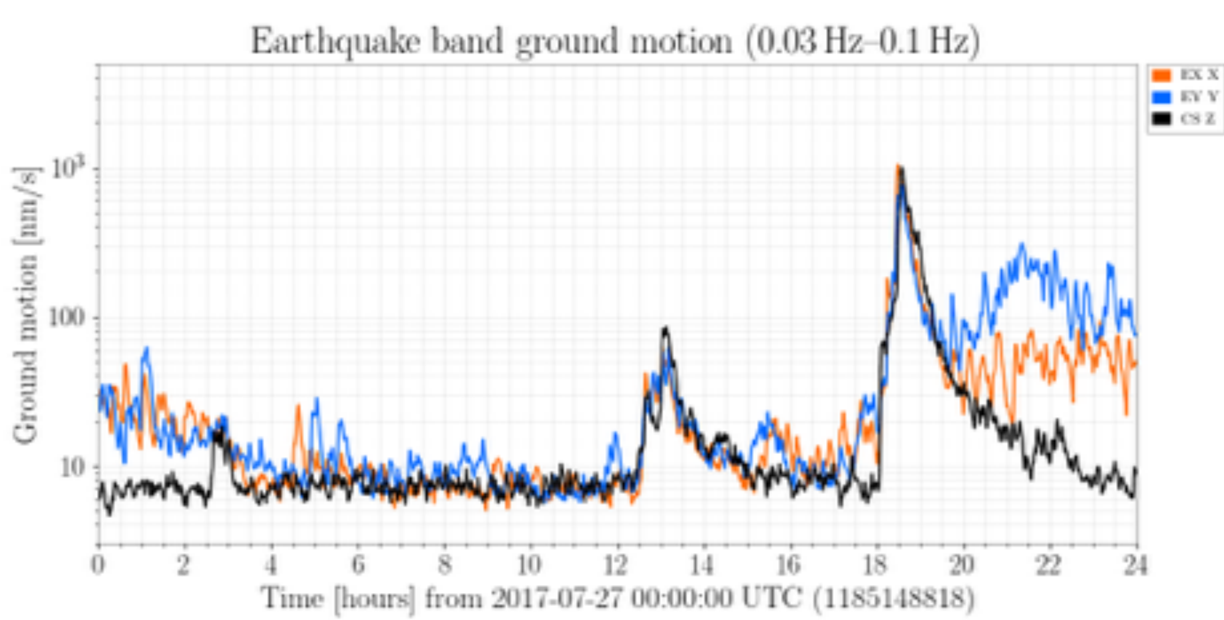


LIGO auxiliary channels

Snow plows
Earthquakes
High Winds



different frequencies/binning
different time signatures



Characterize and identify
in streaming data

With Jess McIver

Other astro applications

- Supernova classification, Charnock and Moss, 1606.07442
- Supernova real-bogus, Cabrera-Vives et al., 1701.00458
- Star-galaxy separation, Kim and brunner, 1608.04369
- Radio galaxies, Aniyan and Thorat, 1705.03413
- Galaxy bars, Abraham et al., 1711.04573

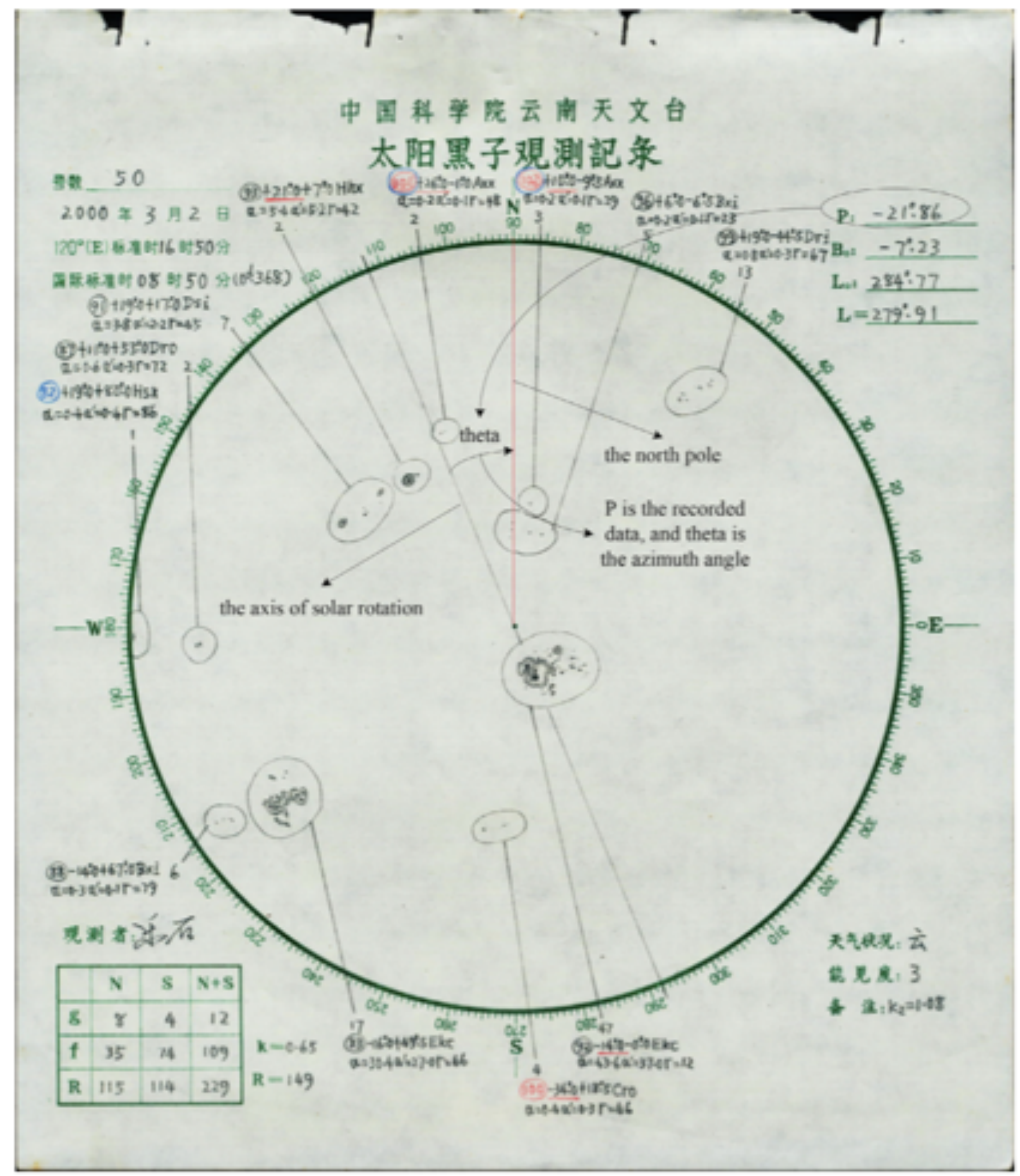


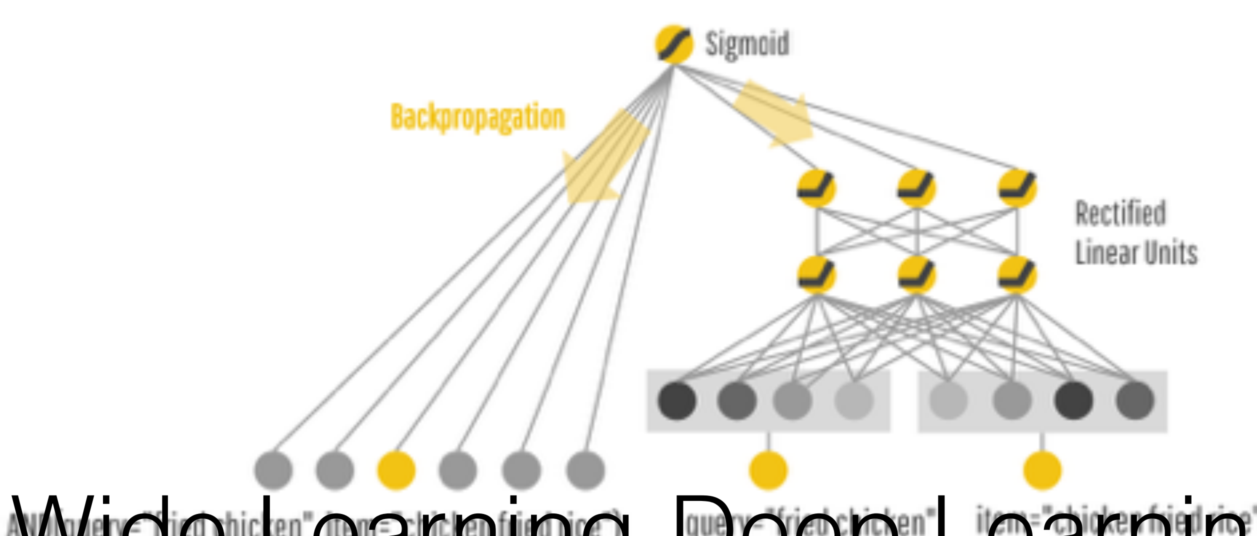
Fig. 1. One example of sunspot drawings preserved by Yunnan Observatory.

Summary

- Direct light curve classification using DL
- Adapting to other surveys
- Applicability to radio, x-ray etc.
- Applicability to transients (with sparse lc)

Extension to other forms e.g. spectra possible

Plans to apply to gravitational wave data



Wide Learning Deep Learning

Deep learning is here to stay!

aam@astro.caltech.edu