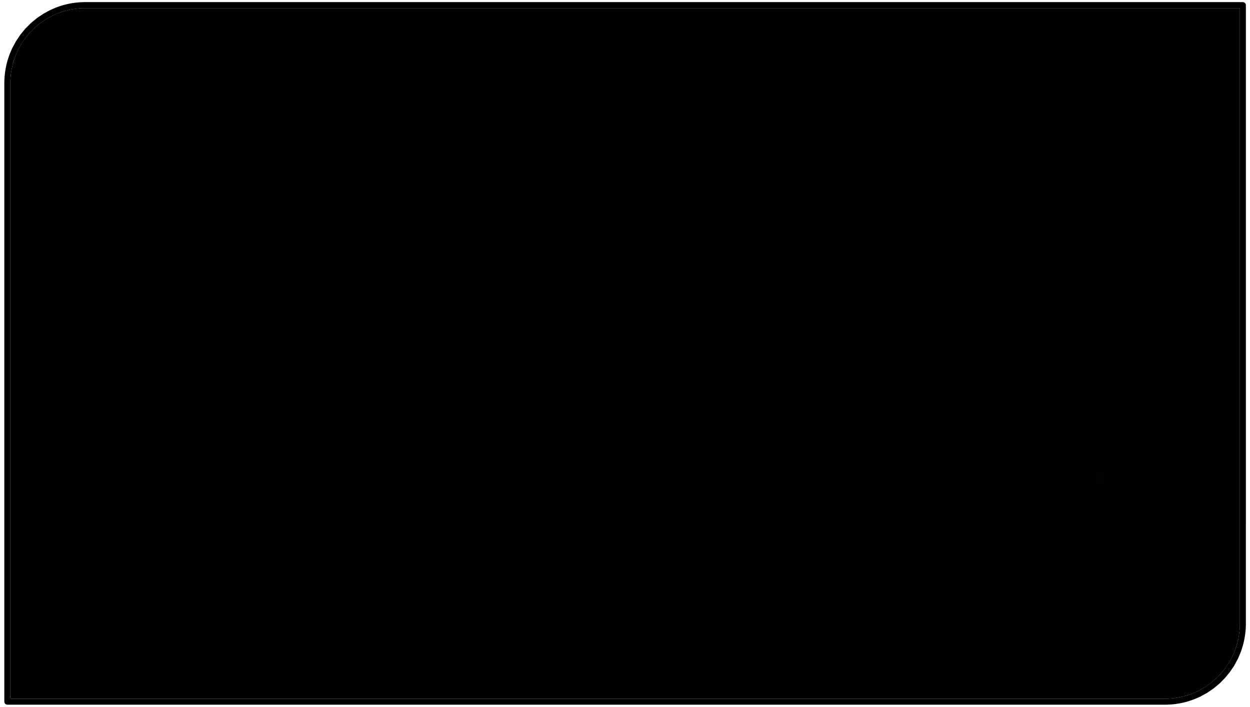
How can machine learning contributing to mining Kepler Data?

Megan Ansdell
CIPS Postdoctoral Fellow, UC Berkeley
233rd AAS Meeting, Seattle

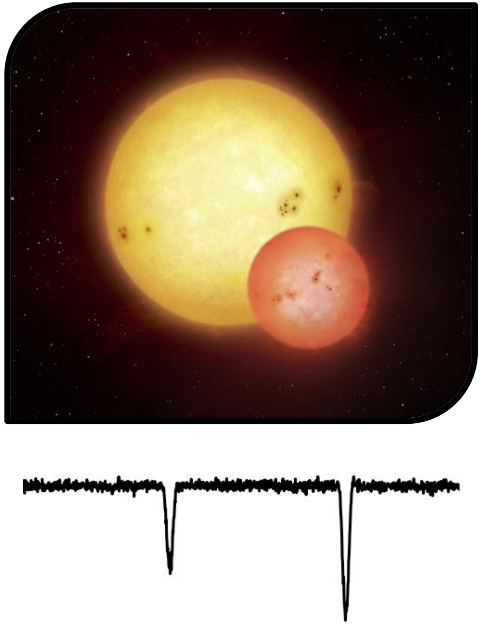
Exoplanet Transit Signals



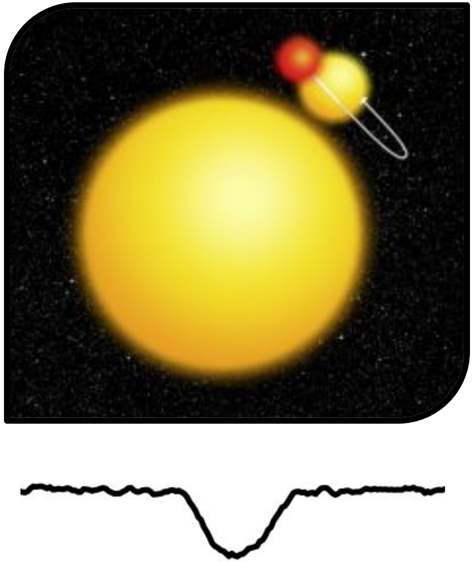
Astronet / C. Shallue

- Orbiting exoplanet transits in front of host star
- Distinct box-shaped transits
- Very shallow 0.01%–1.0% drops in stellar flux

False-positive Transit Signals



Eclipsing Binaries (EBs)

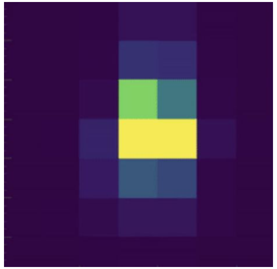


Background Eclipsing Binaries (BEBs)

Stellar Variability / Instrumental Noise

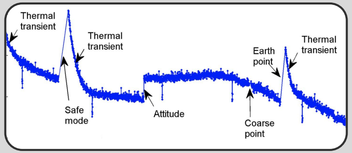
The Kepler Pipeline

Target Pixel File (**TPF**)



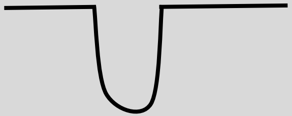
Smith+2012, Stumpe+2012

Aperture Photometry & Systematics Correction



Jenkins+2010, Seader+2013

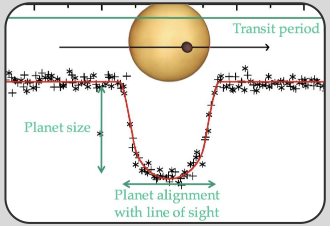
Transiting Planet Search



Threshold Crossing Event (**TCE**)

Wu+2010

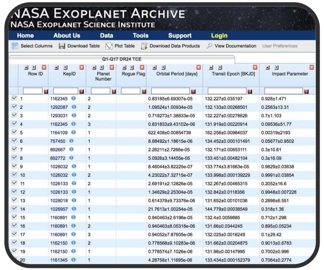
Data Validation



Batalha+2013, Burke+2014, Rowe+2015, Mullally+2015

Kepler TCE Review Team [human vetting]

Exoplanet Catalogues



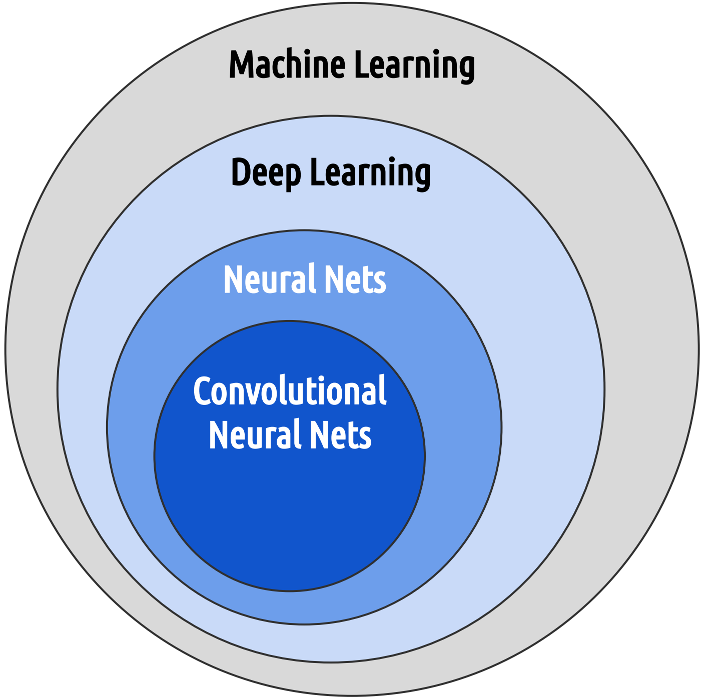
NASA Exoplanet Archive
NASA Exoplanet Science Institute
Home About Us Data Tools Support Login
Select Columns Download Table Plot Table Download Data Products View Documentation User Preferences

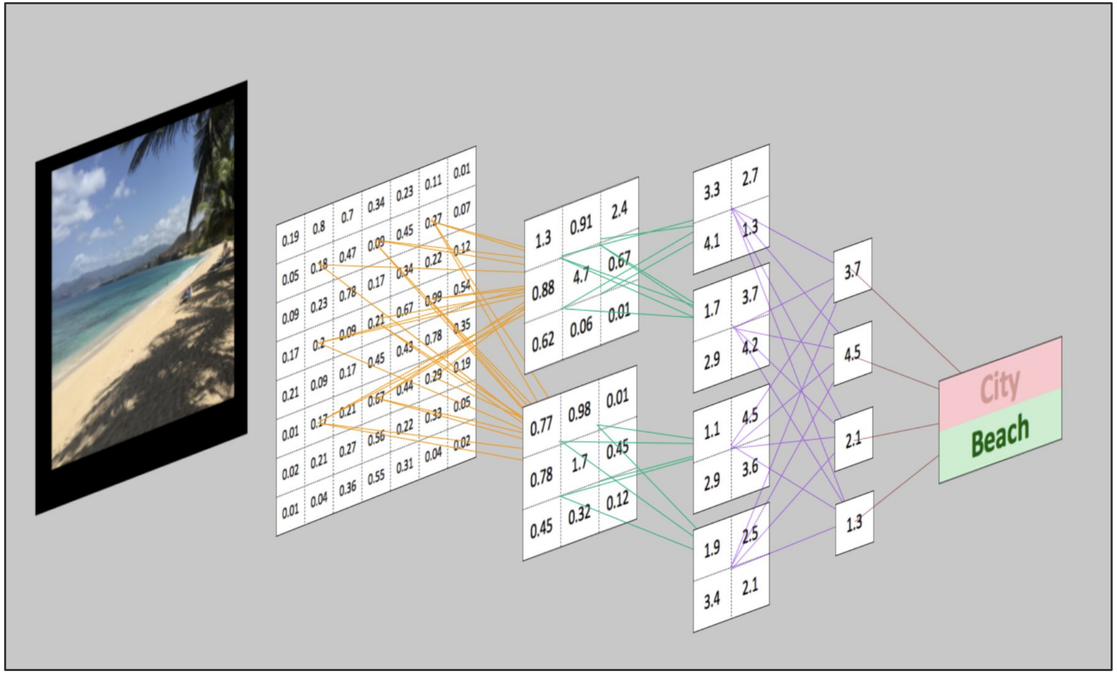
Q1-Q17 DR24 TCE

Row ID	KepID	Planet Number	Rogue Flag	Orbital Period [days]	Transit Epoch [BKJD]	Impact Parameter
1	1162345	2		0.83185±6.69307e-05	132.227±0.035197	0.926±1.471
2	1292087	2		1.09524±1.00934e-05	132.133±0.00268501	0.2563±13.31
3	1293031	2		0.719273±1.38833e-05	132.227±0.00278626	0.7±1.103
4	1162345	3		0.831833±9.43102e-06	131.919±0.00220914	0.09536±51.77
5	1164109	1		622.408±0.00854739	162.256±0.00984037	0.00319±2193
6	757450	1		8.88492±1.18615e-06	134.452±0.000101491	0.05677±0.9502
7	892667	1		2.26211±2.7266e-05	132.171±0.00653111	0.3±10.61
8	892772	1		5.0926±3.14455e-05	133.451±0.00482104	0.3±16.09
9	1026032	1		8.46044±3.82225e-07	133.774±3.81663e-05	0.9829±0.03638
10	1026032	2		4.23022±7.32715e-07	133.998±0.000139229	0.9991±0.03854
11	1026133	2		2.69191±2.12826e-05	132.267±0.00465315	0.2052±16.6
12	1026133	1		1.34629±2.25304e-05	132.842±0.0118356	0.9948±0.007228
13	1028018	1		0.614378±9.73376e-06	131.652±0.00101036	0.2898±6.551
14	1026957	1		21.7613±1.00254e-05	144.779±0.00038549	0.318±1.36
15	1160891	1		0.940463±2.6198e-05	132.4±0.0059885	0.712±1.298
16	1160891	2		0.940463±8.05318e-06	131.66±0.0344245	0.895±0.05234
17	1160891	3		0.94052±7.87605e-06	132.025±0.0016249	0.1±29.42
18	1162150	2		0.778568±9.10283e-06	131.662±0.00204875	0.9013±0.6783
19	1162150	1		0.778574±7.1026e-06	131.98±0.00147995	0.7002±0.996
20	1161345	1		4.28758±1.11695e-06	133.434±0.000152379	0.7064±0.2774

Where machine learning can (and is!) helping

Classifying Transit Signals with Deep Learning





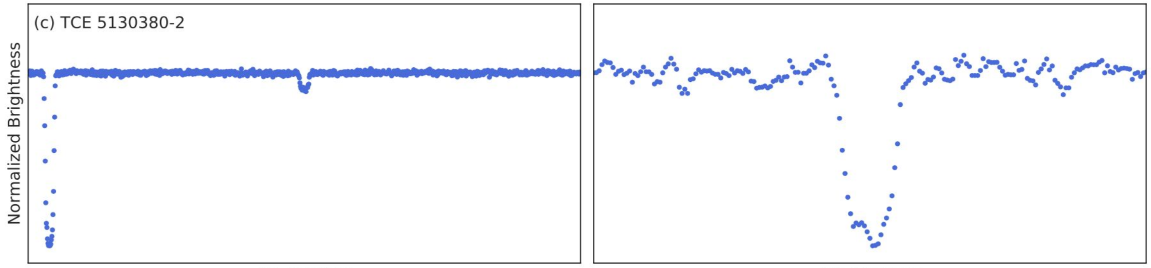
Classifying Transit Signals with Deep Learning

- **Quick** ➔ trained models take seconds to apply to new data
- **Systematic** ➔ important for calculating exoplanet occurence rates
- **Upgradable** ➔ re-doing analysis with upgrades is easy/quick
- **Quantifiable** ➔ can assign probabilities/uncertainties to planet candidates

Astronet

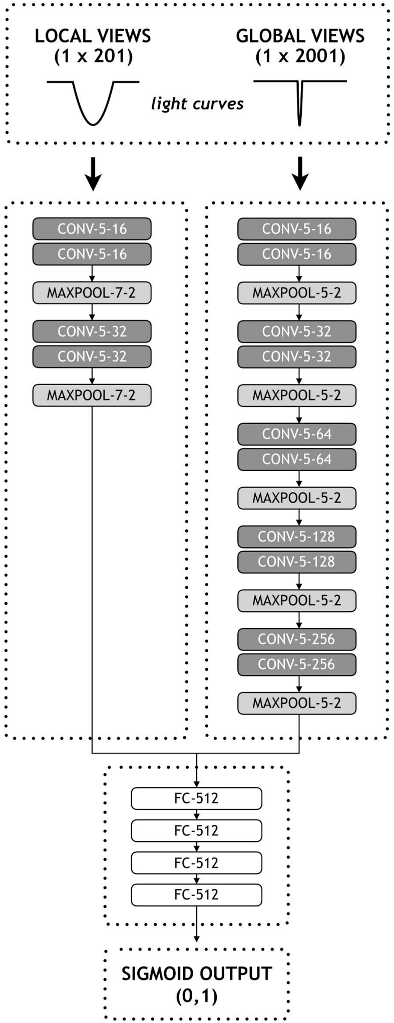
Shallue & Vanderburg 2018 ➔ [https://github.com/google-research/exoplanet-ml]

- Deep convolutional neural network written in TensorFlow
- Inputs are “local” and “global” views of each phase-folded TCE
- Two disjoint 1D convolutional columns + 4 fully connected layers
- Output is binary classifier in the range [0,1]



Global View

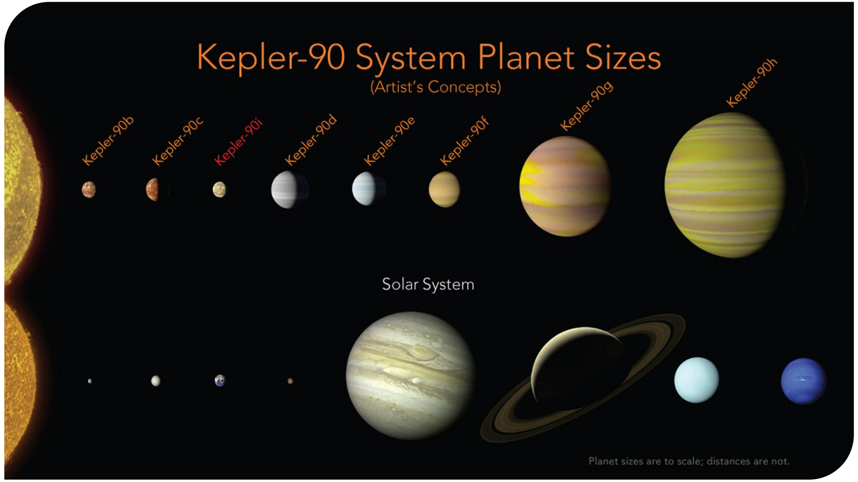
Local View



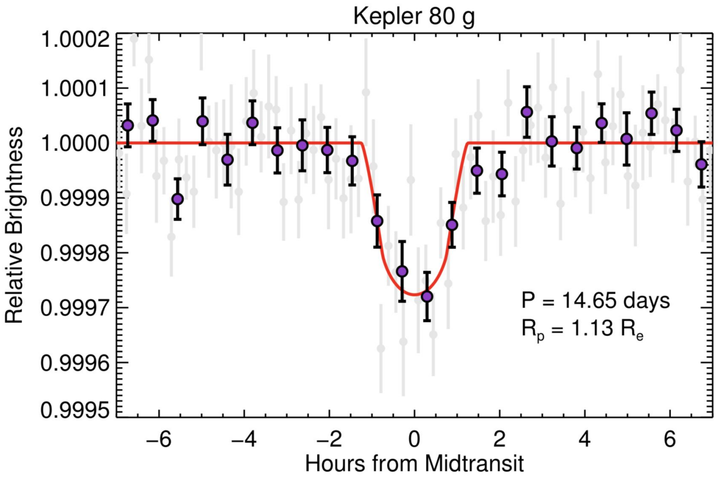
Astronet

Shallue & Vanderburg 2018 ➔ [https://github.com/google-research/exoplanet-ml]

- Applied Astronet to subset (670) of known multi-planet systems ➔ know *a priori* that system is edge-on
- Used lower SNR cutoff of 5.0 ➔ many more spurious TCEs, but ML vetting is quick/automated
- Discovered 2 new Earth-sized planets in high-multiplicity systems ➔ Kepler 80g & Kepler 90i



JPL/NASA

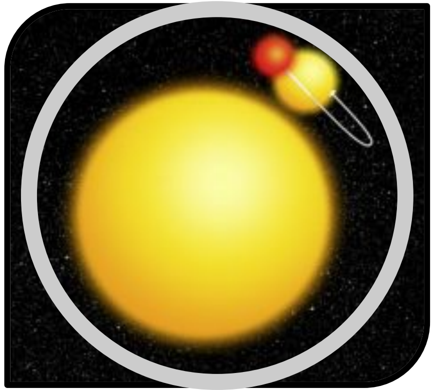


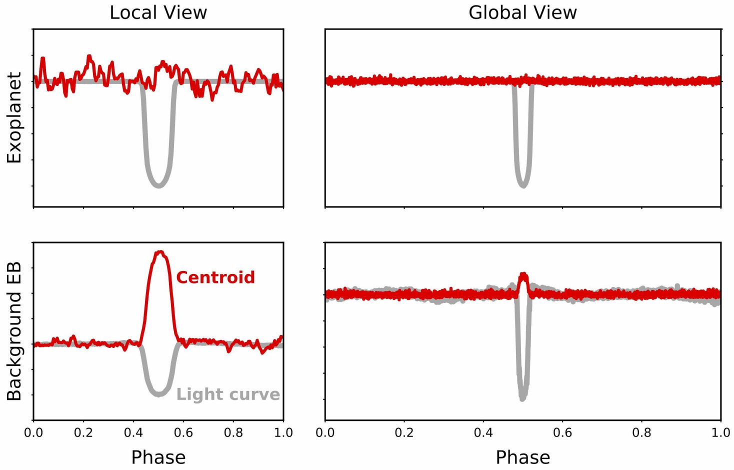
Exonet [Astronet + Domain Knowledge]

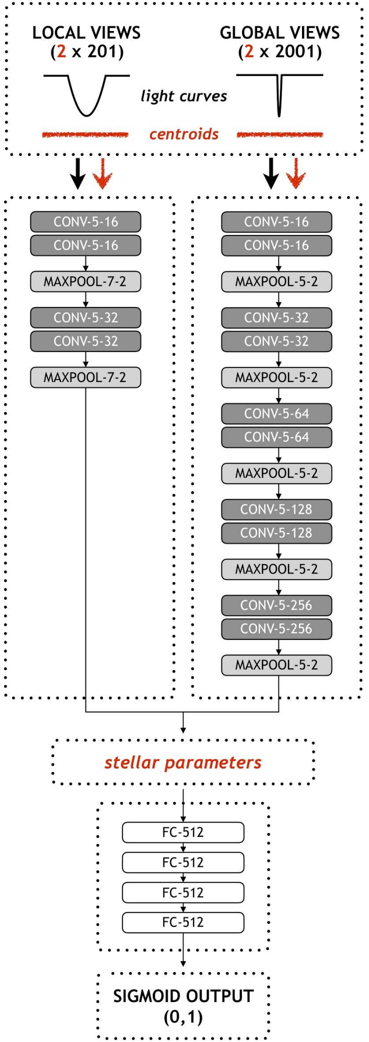
Ansdell, Ioannou, Osborn, Sasdelli, et al. 2018 ➔ [https://gitlab.com/frontierdevelopmentlab/exoplanets]

Centroid Time-series

- Pixel position of center of light in TPF as function of time
- Important for identifying EBs and BEBs





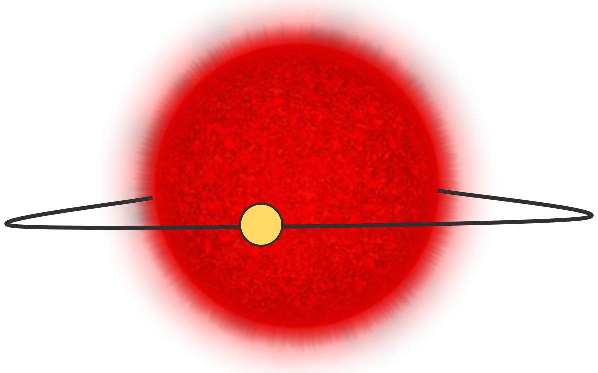


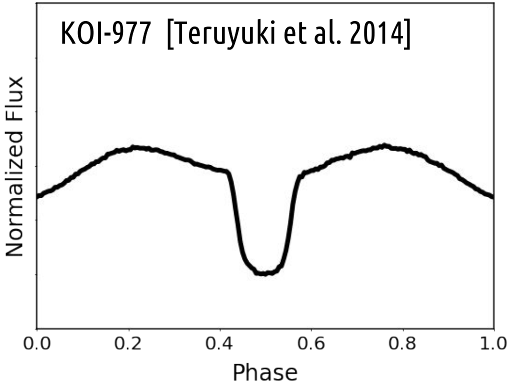
Exonet [Astronet + Domain Knowledge]

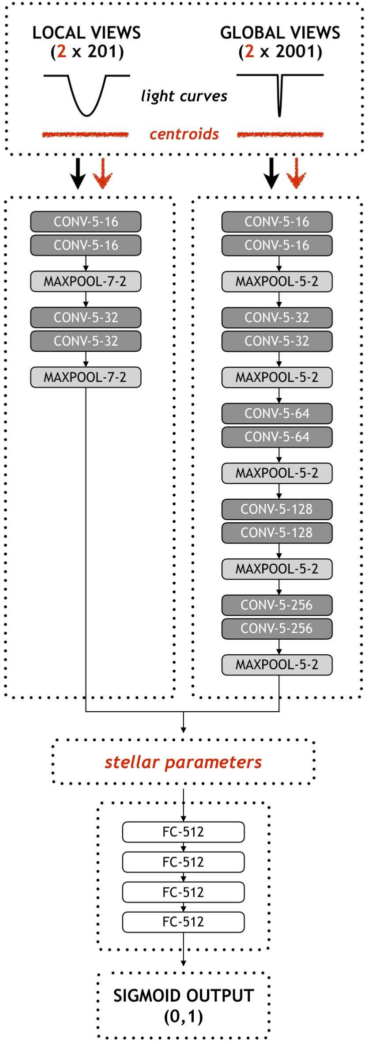
Ansdell, Ioannou, Osborn, Sasdelli, et al. 2018 ➔ [https://gitlab.com/frontierdevelopmentlab/exoplanets]

Stellar Properties

- From KOI catalog: mass, radius, density, surface gravity, metallicity
- Important for identifying, e.g., giant star eclipsing binaries







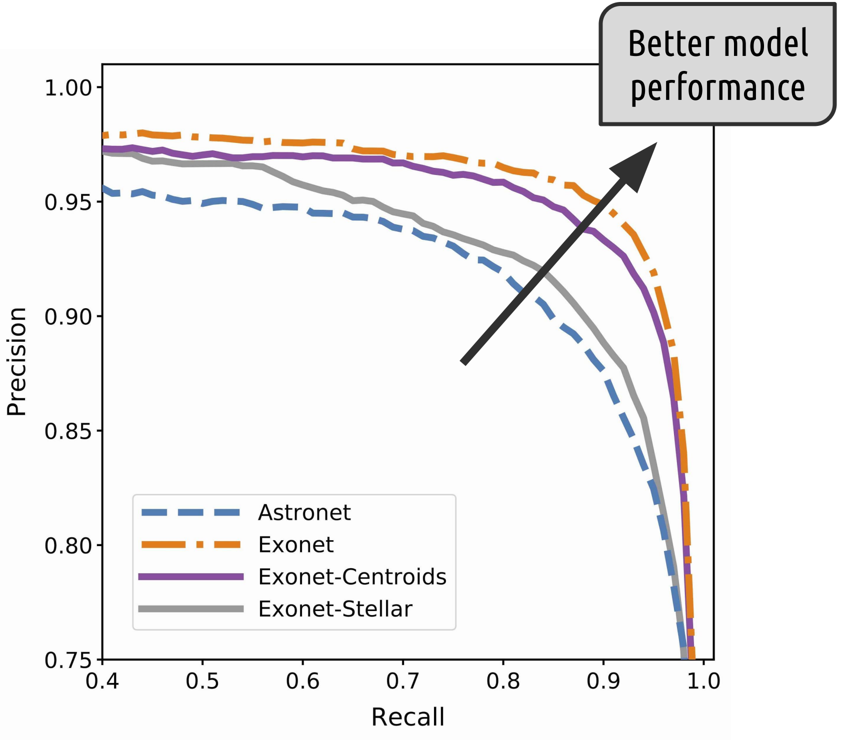
Exonet [Astronet + Domain Knowledge]

Ansdell, Ioannou, Osborn, Sasdelli, et al. 2018 → [https://gitlab.com/frontierdevelopmentlab/exoplanets]

Improved Overall Performance

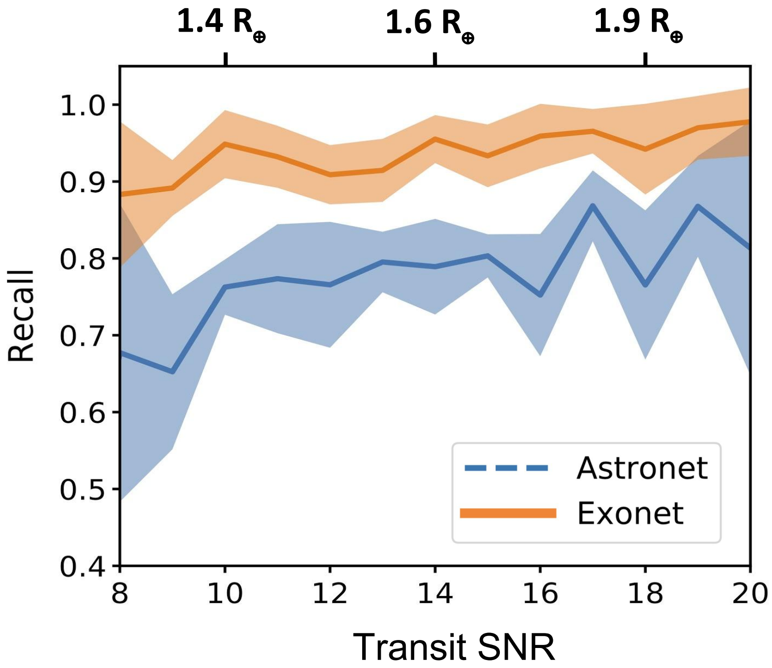
	Accuracy	Avg. Precision
Astronet	95.8%	95.5%
Exonet	**97.5%**	**98.0%**

- Accuracy = % of correct classifications
- Precision = % of classified planets that are true planets
- Recall = % of planets recovered by model



Earth-sized Planets still lurking in Kepler data?

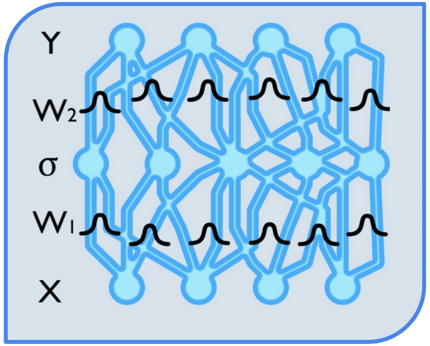
- **Exonet** promises to find new Earth-sized exoplanets still hiding in Kepler data
  - ➔ 20% higher recall for small exoplanets
  - ➔ Apply to SNR~5 TCEs



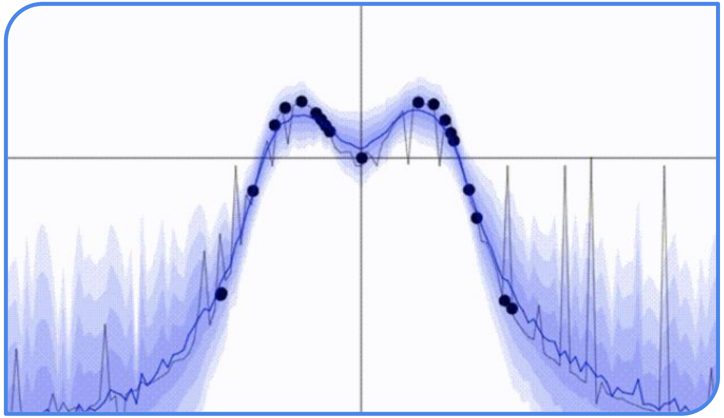
Incorporating Bayesian Deep Learning

- **Bayesian Deep Learning** leverages dropout to efficiently produce uncertainties on probabilities
  - ➔ Incorporate into exoplanet occurence rates
  - ➔ Useful for prioritizing follow-up observations

Bayesian Deep Learning ≈



≈ **Gaussian Processes**

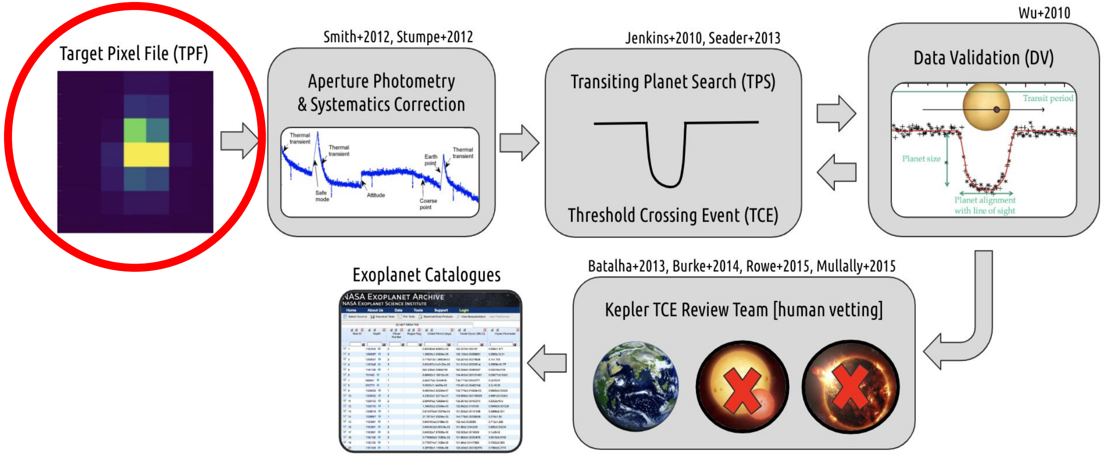


http://www.cs.ox.ac.uk/people/yarin.gal/website/blog.html

Detection of Transits Directly in Target Pixel Files

- **TPFs** (image time series) are minimally processed; find exoplanets traditionally missed?

  ➔ Can deep learning find more informative representations beyond highly processed light curves?
  ➔ Computationally expensive; unclear if effective on low-SNR TCEs that require phase-folding




Questions?