Finding exoplanets with TESS & AI

Megan Ansdell, Yani Ioannou, Hugh Osborn, Michele Sasdelli,

+ Jeff Smith, Jon Jenkins, Doug Caldwell, Adam Lesnikowski, Chedy Raissi, Massimo Mascaro













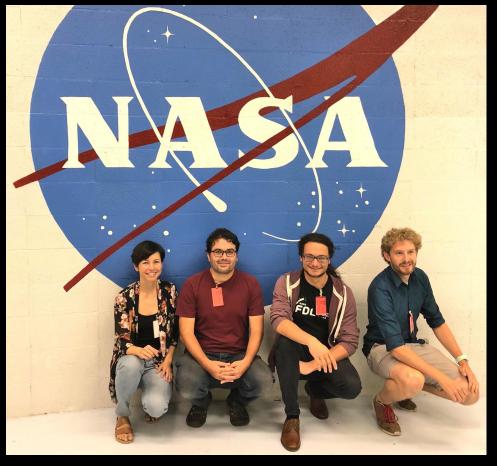


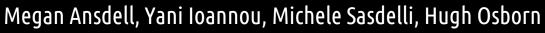






The Team







Frontier Development Lab

Space Science Researchers Machine Learning Researchers Silicon Valley Partners

- 8 week research accelerator hosted at NASA Ames & SETI
- 9 projects across 5 areas proposed by lead mentors

Radical Solutions to problems in Space Science















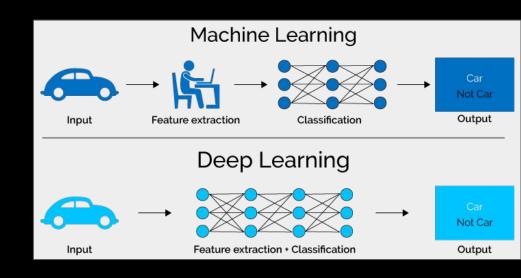






Deep Learning

- Excellent at classification problems when:
 - Lots (N> 10 000s) of labelled data
 - Signal is complex to model (not True for "planet" but true for non-planet!)
- Once trained, a deep learning algorithm is:
 - Far faster than classical at performing classification
- But:
 - Large computing infrastructure to train
 - Must set aside much of the data as a training set





















Deep Convolutional Neural Networks

- Machine Learning (ML):
 models learn features from data
- Deep Learning: layers build increasingly complex features
- Neural Network (NN): model learns weights of nodes
- Convolutional Neural Network (CNN): exploits spatial structure in data
- **Binary Classification**: final layer outputs single number from 0-1









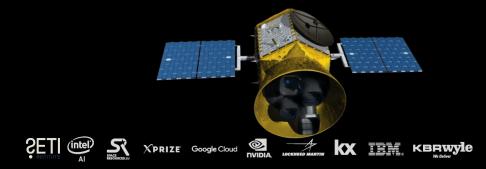






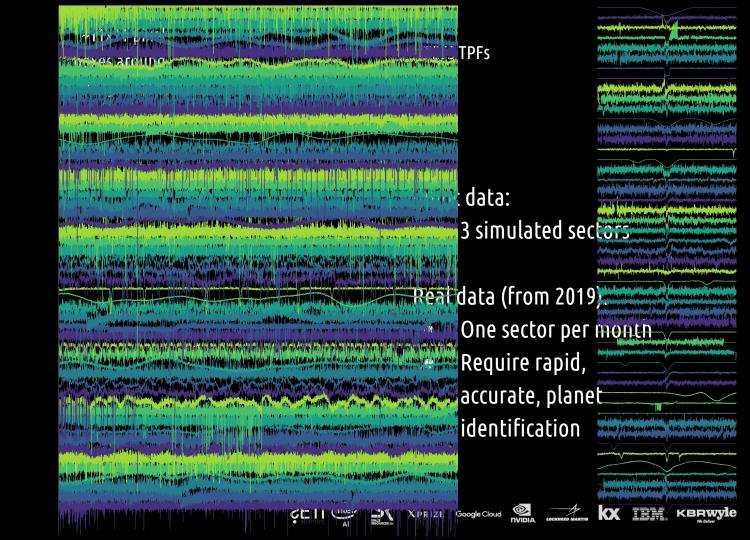


The Data - TESS



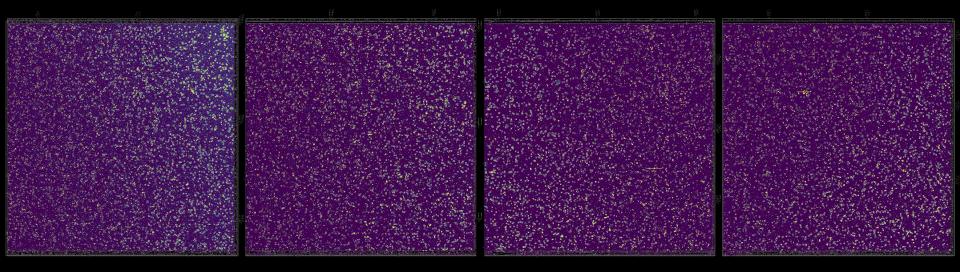
TESS "Data" (TSOP-301 simulations)

Target Pixel Files
(simulated planets)
16k stars per
sector
2min cadence





The Data - TESS



Per month: 4 CCDs ~16 000 target stars

21 000 images

Our dataset: TSOP-301. 4 simulated sectors







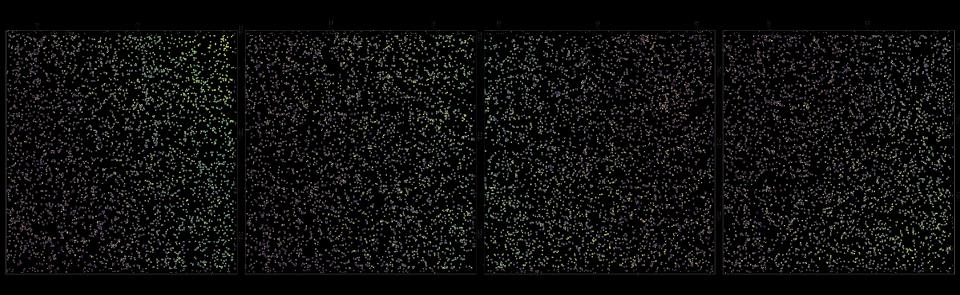








TESS - the data bottleneck



4 simulated sectors ~16 000 stars per month ~4 000 hours of video









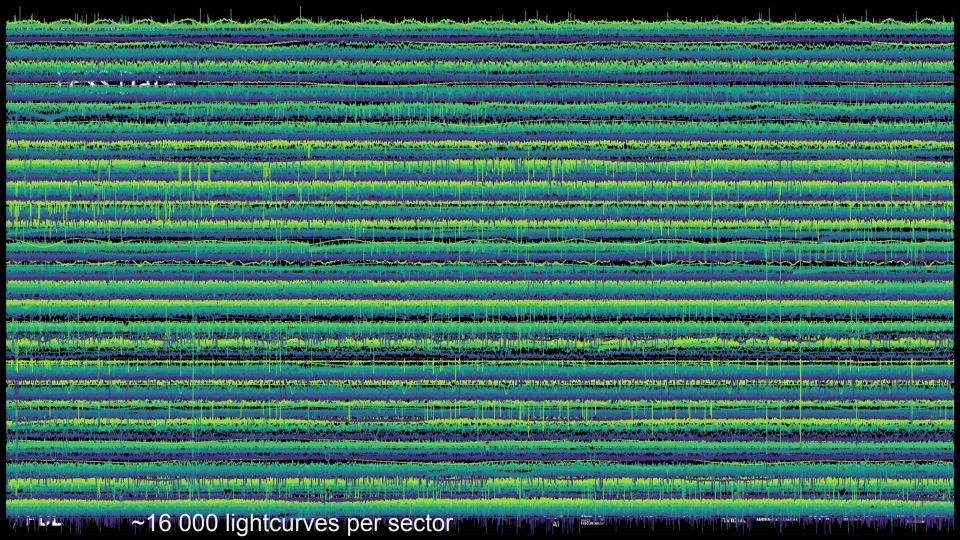




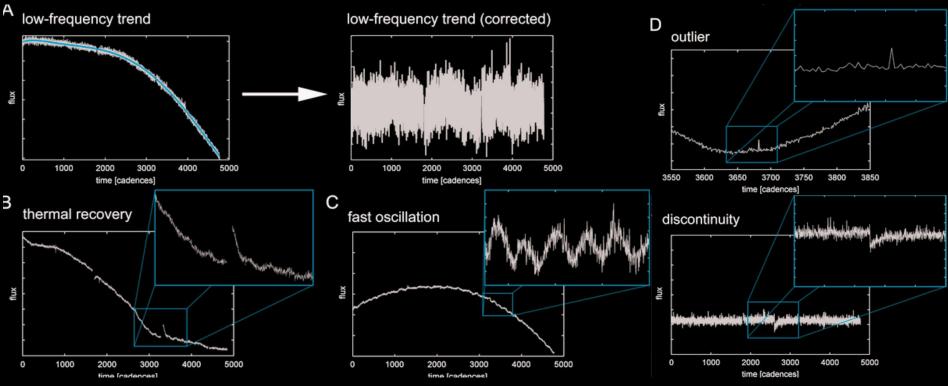








TESS Systematics













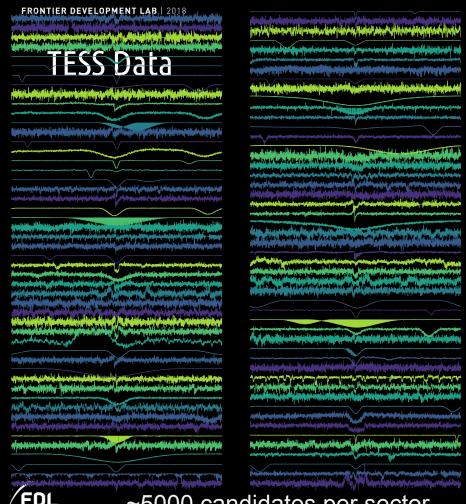












والمتراجع والمراجع والمترجع والمتراجع والمتراجع والتنافي والمتراجل والمتراجع والمتراج والمتراجع والمتراج والمتراجع والمتراج والمتراجع والمتراجع والمتراجع والمتراجع والمتراجع والمتراجع وا alatik pi jilang panalati panalating minang dibanasahan kanasahan alatik panasanca dan menanantah dibanasa Mananantah panalating minangan menangan menangan beradah jura beradah beradah beradah beradah panalati panalat And the state of t A track of the tra ارة بخالية البروجية المناولة وموسطة بعن بالبروجية ومساوية في كان المناولة ال المناولة الروان المراجع المراجع

AND THE PROPERTY OF THE PROPER the february place of the state pull y plate por an de cir la la la mara de competit de cir la place de collection pull de la proper de competit de la competit del la competit de la compet takin taling began beraka paramanan belan beraka paraman belan aparaman beraka belan beraka belan beraka belan Attention of the state of the s وبرايا والمراج والمراجع ويشرف والمرور ويتراوي فالمراج والمراجعة والمراجعة والمراجع والمراجع والمراجع The property of the property o A STATE OF THE PROPERTY OF THE



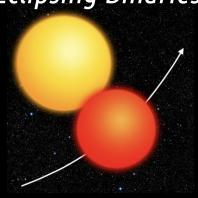




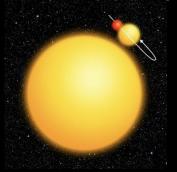


The problem: variety of false positives

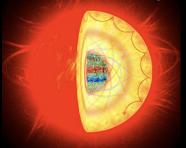
Eclipsing Binaries

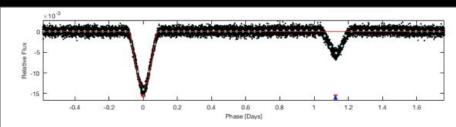


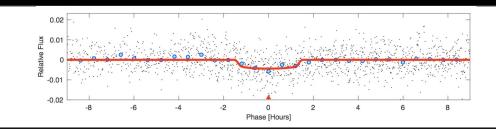
Background Eclipsing Binaries



















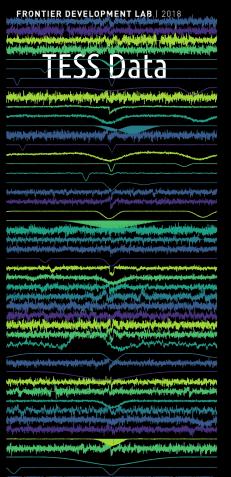












FOL

particle terrentalist manifes description description of the properties and the properties of playing فرون والقالم المراوا والمناز ويستخط والمناط والمناز والمناط والمناط أوا وارتباه فالمراوي والمال والمناطع والمناطع the file of the strong to be destricted from the figure of the strong of أورال أوال ودريان وبالملامات بالماجوني المراجي بالرياد والمراجة المراجة المردية الماجود والمتلاوي والمعاولة Toronte particular descriptions and dispersion descriptions of the properties of the A service the service of the service والمستخدمة والمنافس ومجافدها المساورة والمادية

Proposition destruction proposition and the proposition of the proposi

phylogical distinguish production of the configuration of the physical distribution of the configuration of the co alatik pi jilang panalati panalating minang dibanasahan kanasahan salatik panasanca dan menanantah dibanasa Mananantah panalatin menangan menangan kanasahan dibanasah dibanasah panasah menangan dibanasah menangan beras AND DESCRIPTION OF THE PROPERTY OF THE PROPERT وتناها والمرابط والمراجع والمراجع والمراجع والمراجع والمحافظ والمراجع والمراجع والمراجع والمراجع والمراجع والمراجع والمراجع And the state of t التي بدايل والمراجع والمراجع والمستورة والمستورة والمستورة والمستورة والمستورة والمستورة والمستورة والمستورة و والمراجع والمستورة والمس minimum properties properties of management of the second and the state of t والمنافذة والمنافذة والماسية والمنافذة والمنافذة والمنافذة والمنافذة والمنافذة والمنافذة والمنافذة والمنافذة والمنافذة And the second s A train to the state of a second state of the state of th والمرافية ارة بخالية البروجية المناولة وموسطة بعن بالبروجية ومسهوم في يراح إمادة والمناولة البروجية ومن من المناولة المن المناولة ال المناولة And the first of the second of the first of the first of the first of the second of the first of

ويراج والمناف ويناه والمراج والمراجع ومناهم والمراجع والمناف والمراجع والمراجع والمراجع والمراجع والمراجع

The state of the s Berighed means his played a locate discharge the best live to a second property provides and extended the best and the

إسامه المسابع والمعاولة المرابع والمرابع والمرابع والمرابع والمرابع والمرابع والمرابع والمرابع المرابع

And the state of t things and his and providing residents. It said in a lower of the private standing the first and a second The pilloping on hear later many or print a property of the print of t

of the state of th

organisation large the production of the production of the control of the production between the first

Attention of the best of the best of the best of the second property of the second second second second second

property of the property of th

Hateriale philipping to the property of the property of the property of the party o





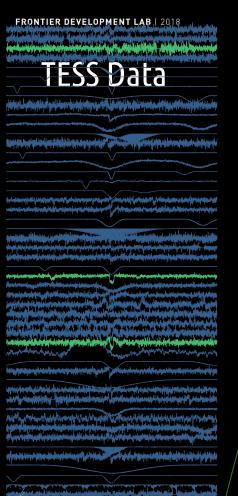












FDL

والمراجع والمناول والمناوران فالمرام والمناوران المناول والمناور والمناور والمناور والمناور والمناور والمناور والمناور	
	AND PERSONAL PROPERTY.
the lift of the state of the st	
Application of the second seco	
and the state of the second se	To And de Control of the Control of
Sanger Control agricultural construction of superior and the best of the superior and the s	
	A second to the second
	المحدود التعالم
	A CONTRACTOR OF THE PERSON OF
	-
the profile the profile and the profile of the prof	A PART AND
the state of the s	mphysial production
	And a state of the
and like the ship to the like a part of the province with the part of the tenter having more at it is a mall o consideration of the like and the same at the same at the same and the same at th	Harry Street of the street of
A starting market and production for the starting dispersion by the first starting	A Participant of the street
he dig the establishment to apply a style state of the establishment that the forest and a story in the state of the state	Parish No.
te distribution distribution and the distribution and the distribution of the second state of the second	
	Appropriate the second
And the state of the best of the state of th	Handalahan
istik gentrus katikationen professionen interesten in der som en statut in der signationen beheins produktione Auf der statut interesten interesten bestieten bestieten in der statut interesten interesten der statut interesten interesten interesten interesten der statut interesten inter	de la linea de la companya de la com
militares proprietares de la comita de la forma de la comita de la comita de la comita de la comita de la comi La comita de la comita del comita de la comita del la comita	shirtelitikanikana
Alfabla .	HANNELPHAN

adiritalista propieta en alla per la propieta de l the control of the state of the second property of the state of the st والمراط المتعلق المدر والتعلق ورواز فاستأفرا مراطان وروعاتك الأطور ويقاف والمراط الماسة a restaurable and beautiful and the second of the second s The same transfer of the same ر به الرائم في الطبيع والمواجع والمنظمة المنظمة المنظمة المنظمة المنظمة المنظمة والمنظمة والمنظمة المنظمة المن والمنظمة والمنظمة المنظمة المنظمة والمنظمة المنظمة المنظمة المنظمة المنظمة المنظمة والمنظمة المنظمة المنظمة ال manyment promote the manyment Printed to the second of the s والمناف والمدور والمراج والمناورة والمناف والمناف والمناف والمناف والمناف والمناف والمناف والمناف والمنافرة والمراجع وال i para mana mana dia mana dia paonina mana dia paonina mana dia mpika mpika mpika mpika mpika mpika mpika mpik Mana mpika mpi Mana mpika mpi transfer for a little plant and the second state of the second state of the second state of the second state of مورة المتناق والمترة والمتراق والمتروق المراواة فيهاده والمائية والمتامن المعارة المعارة والمتراطية أوجه والمام land and a state of the second as the first probability of the second and the first probability of the behavior of the behavio

The sailed and represent the sail the property of the sail of the der the state of t de principal de constitución de constitución de la والمعادية والمحمودة المتعاوية المعاوية والمحمد المالة المحمد المالة المعادية المحمولة والمحمدات المحمدات المحمد in his prikeri ser anga kanta ka dila manaka penggulah kanta a kalanda a kinah dibirti pini balan di Birka kantan panggulan a panggulan manaka manaka panggulan anga panggulan kantan di panggulan panggulan panggu an addition of the second of Attendance because the state of And the life bearing services in the constraint of the services of the service transmit and it little with it ill strawment to pay to the angular properties and an extended by The second secon And the second s Bishgildes frightiggetingster bigger better mit the prist of the biggilder describes



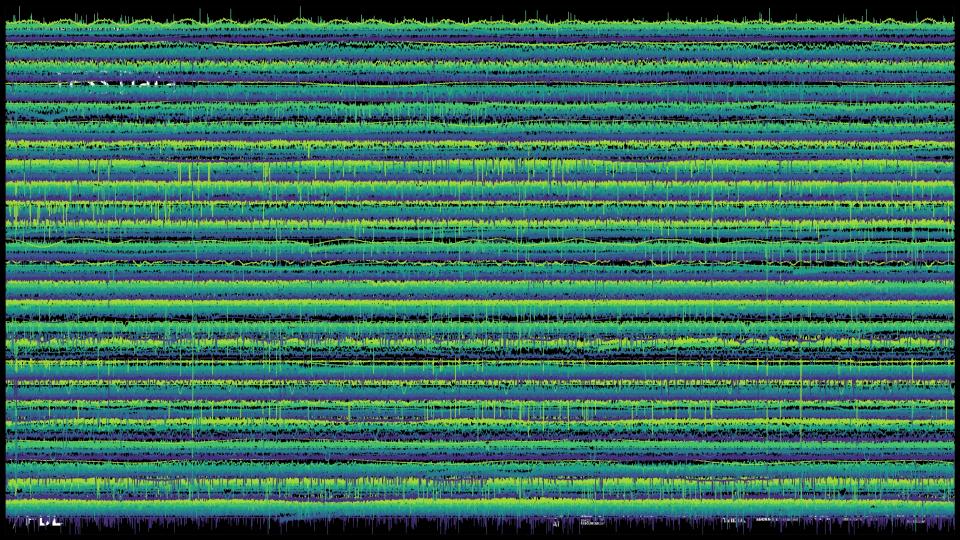
XPRIZE Google Cloud

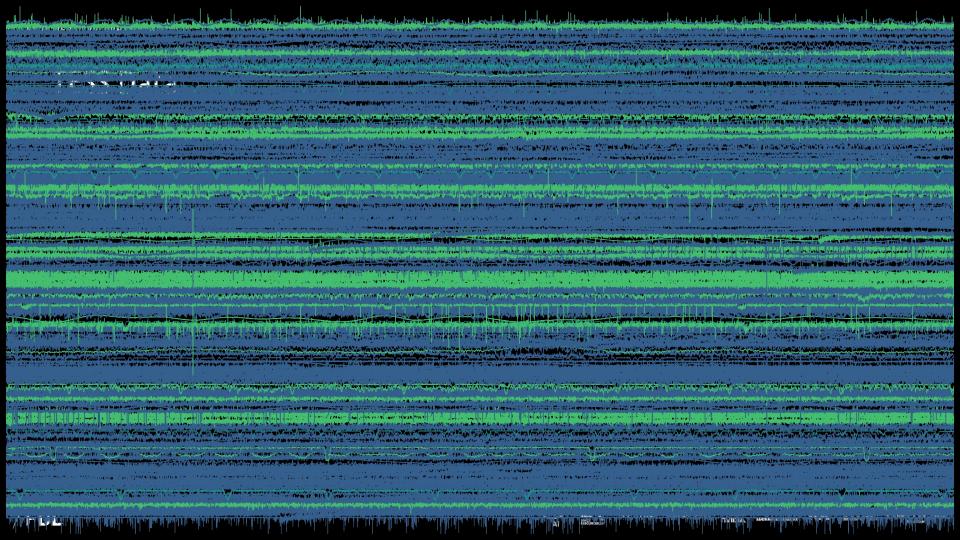




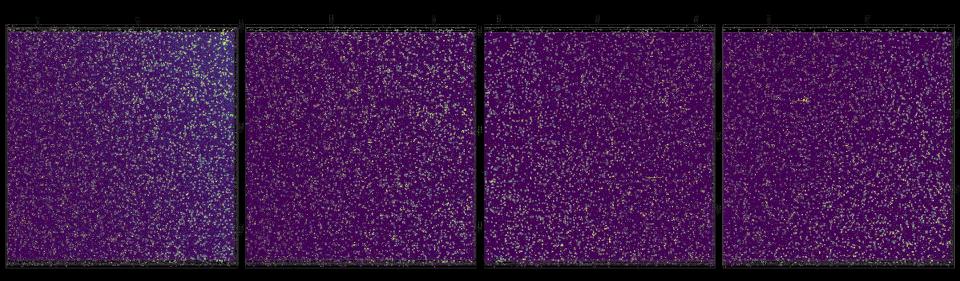








TESS Data











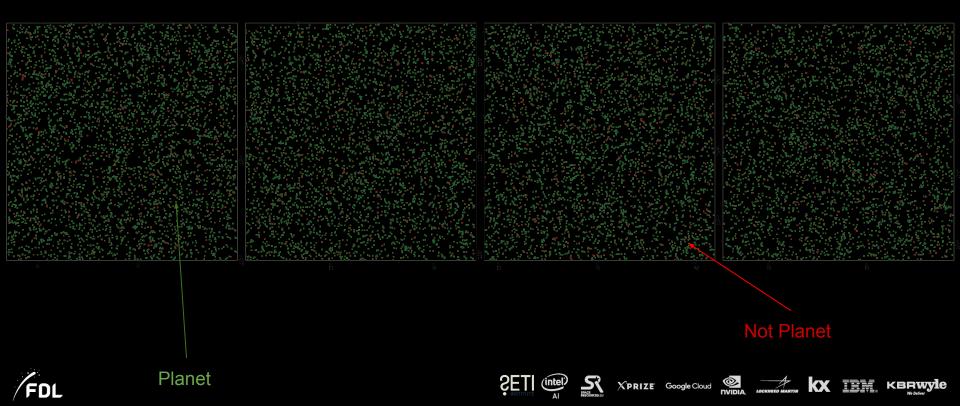








TESS Data



Current Classification Technique

- Statistical/automated methods are used to whittle down candidates
- Manual vetting is still common
- Team of 18 humans: 94% accuracy
- ~300 human hours per sector















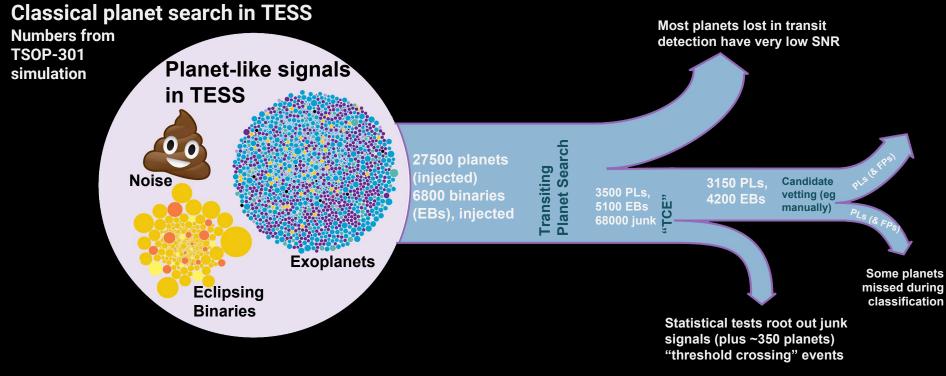
























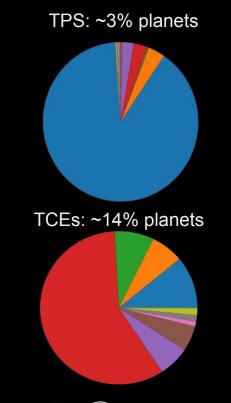






Classifying TESS Candidates

- TESS pipeline produces candidates in two tiers
- 70 000 initial detections from transiting planet search
- 30 000 pass statistical thresholds (TCEs)
- Only ~4000 planets (highly unbalanced)





















the pilling of the property of the party of

والمناول والمناول

これはころかったのかというのはないというというはないかられたいかんというかん

white the second of the second

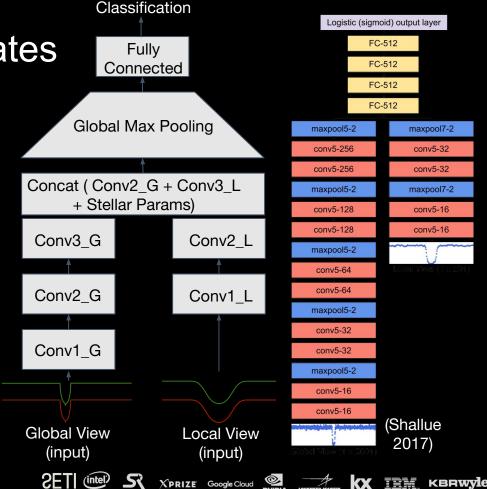
والتفريد فكارس ليبرنا وكالبطوري والطيفا وفيستم فللس فالفور فيفرخ واساس إفيانات



Classifying TESS Candidates

Adapted Shallue et al

- Made model smaller (0.06%)
- Added stellar parameters
- Added motion of star (centroid)
- Mini-batch balancing to account for label imbalance















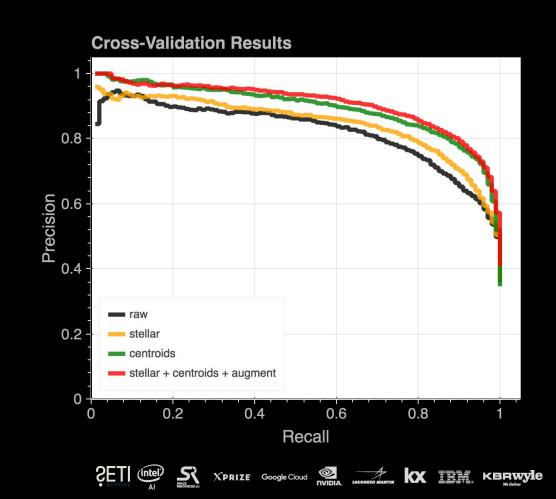






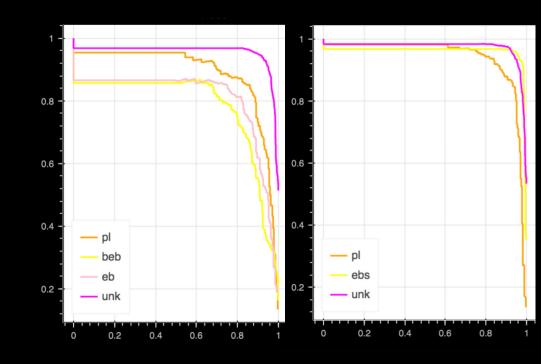


- Average precision on Kepler ~96%
- Recovers ~90 more planets than
 Shallue et al (on single model)
- Model is ~500 times smaller
- Similar precision on TESS data
- Can be run in minutes not hours!





- Also developed multi-class models on TESS data.
- Useful for follow-up!
- 3-class > 4-class
- Slightly lower precision on planets.













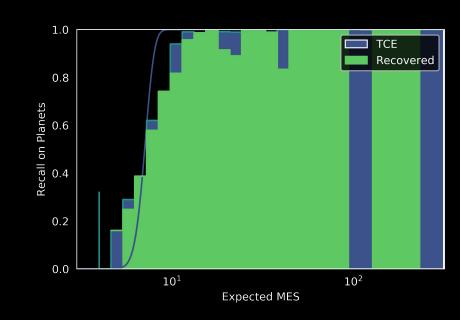






Kepler vs Shallue

Up to 650 more planets recovered (at precision of 0.9 on single model)











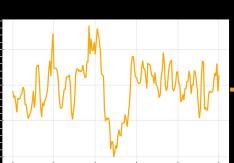


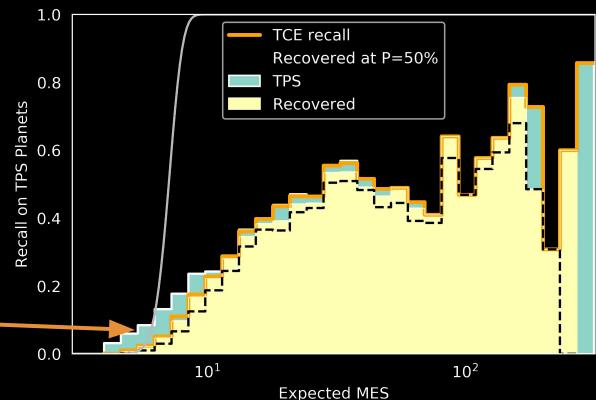






- Can replicate the classical statistical threshold tests
- Recovered planetary signals missed by the classical tests













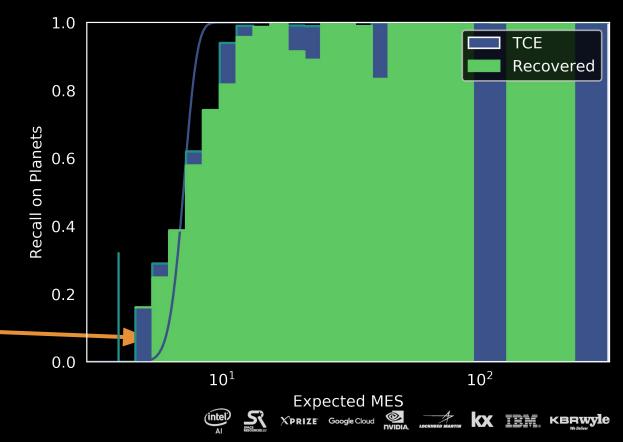






- Can replicate the classical statistical threshold tests
- Recovered planetary signals missed by the classical tests





Future - applying to real TESS data

- Directly transfer network to real flight data...
 - Noisy data
 - o Is the data (& systematics) the same as in training?
- Use human-vetted labels for TESS candidates...
 - Noisy labels
 - Slow
 - Exactly what we're trying to replace!
- Inject known signals into real flight data
 - Can train network rapidly with real data and noise.
 - Simulations and network re-training can be part of pipeline.



















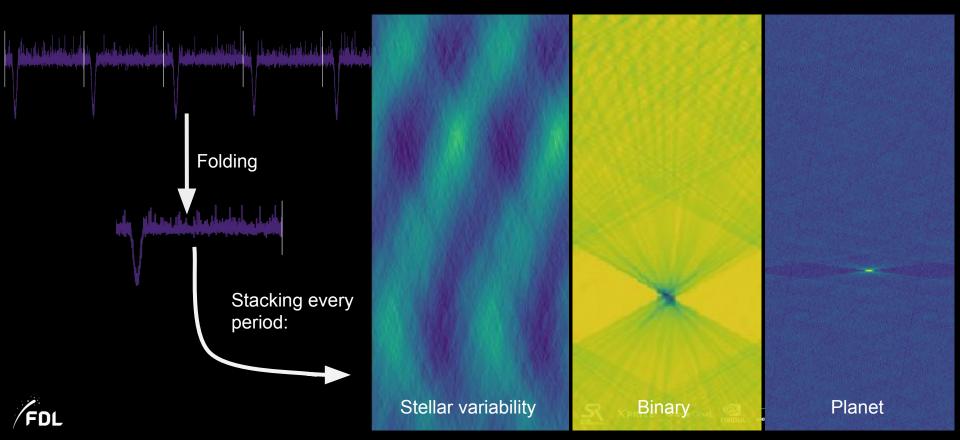






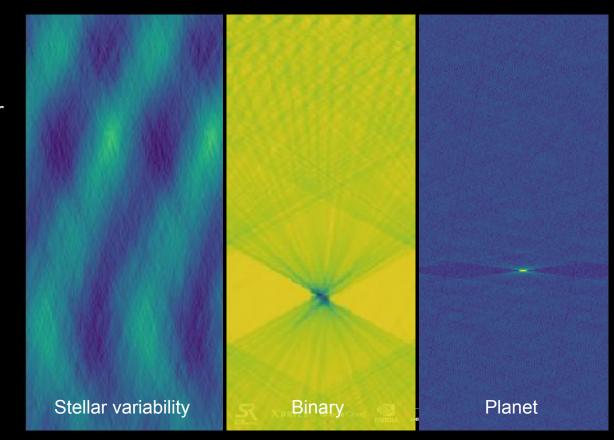


From the lightcurve...



From the lightcurve...

- Applies ResNet50
- 91% accuracy on Kepler candidates
- Promising but need more time.





From pixels...

64 000 month-long videos

Lots of noise!

Promising approach which could by-pass TESS pipeline...

But needs more work.













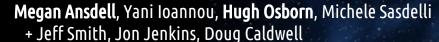




Summary

- Classify TESS planets **faster** & **more precisely** than previous approaches.
- Innovative new avenues for planet hunting direct from light curves & pixels

KX Google Cloud















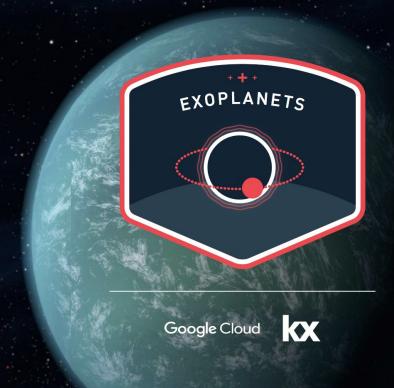












Megan Ansdell, Yani Ioannou, Hugh Osborn, Michele Sasdelli

+ Jeff Smith, Jon Jenkins, Doug Caldwell

















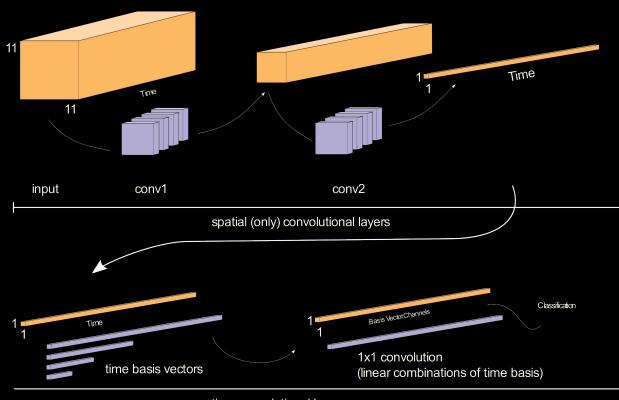


From pixels...

64 000 videos of stars

Promising approach which could by-pass TESS pipeline...

But needs more work

















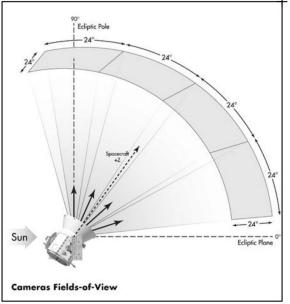


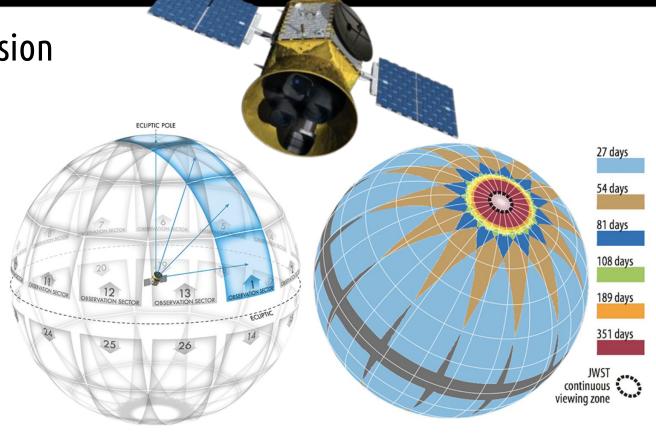






The data: TESS mission



















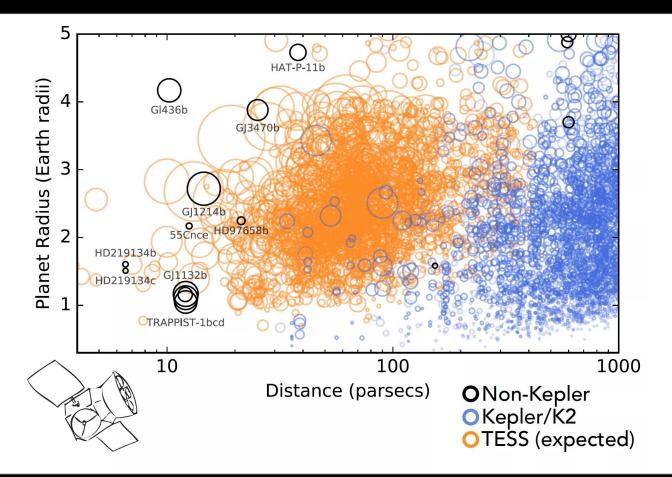






TESS

Exoplanets around nearby stars









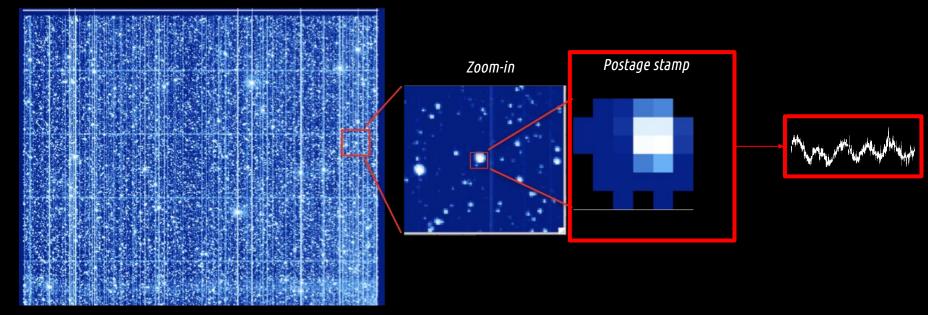








Full-frame images





















Incremental: Classify planet candidates detected by the pipeline

 Train a Neural Network on candidate planets (TCEs) detected by the pipeline Transiting Planet Search

 Use heavily pre-processed domain data generated by the pipeline.

In Shallue et al 2017: Full lightcurve Zoomed-in lightcurve New Input Datasets: Centroids (x & y) Stellar Properties Data augmentation





















Incremental: Classify planet candidates detected by the pipeline

<u>Improvements on Shallue, 2017:</u>

- More input datasets
- Multiple false-positive classes
- Use data balancing
- Augment light curves to increase training examples.
- Lighter NN

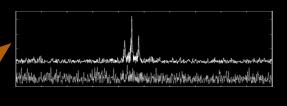
In Shallue et al 2017:
Full lightcurve
Zoomed-in lightcurve

New Input Datasets:

Centroids (x & y)

Stellar Properties

Frequency-space data



















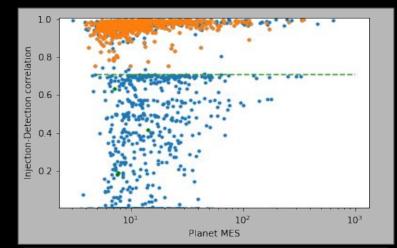


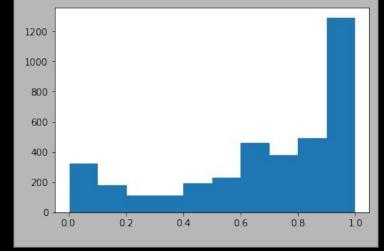


Ongoing work:



- Labels not necessarily intuitive.
- Modified "correlation" metric between injections and detections to include period multiples.
- Still issues with high-SNR planets not being detected



















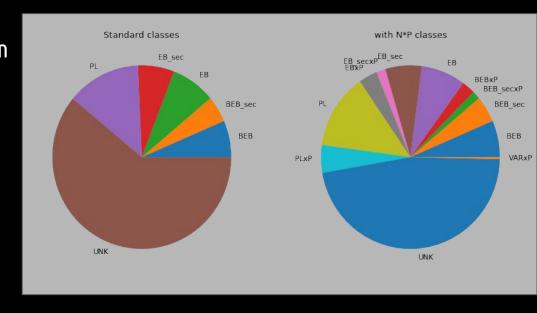




Ongoing work:



- Labels not necessarily intuitive.
- Modified "correlation" metric between injections and detections to include period multiples.
- Still issues with high-SNR planets not being detected













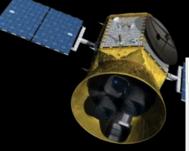




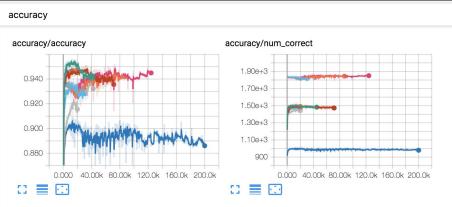




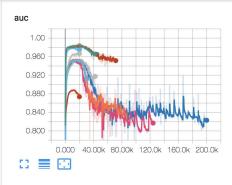
Ongoing work:



- Converted Shallue code from Kepler to TESS
- Improvement with centroids
- Also improved with modified smoothing (spline) techniques.
- Models being trained
- Stellar parameters and frequency space data to be tested and added

























Next steps...

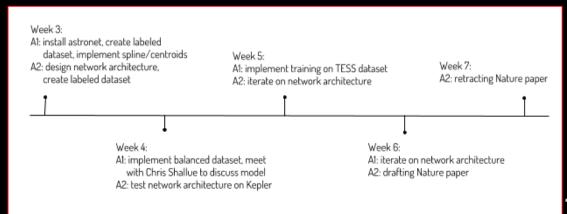
Week 4:

- Avenue 1: Implement balanced labels
- Avenue 2: Test design architecture with Kepler



Week 5:

- Avenue 1: Implement training based on TESS data
- Avenue 2: Iterate on model/architecture





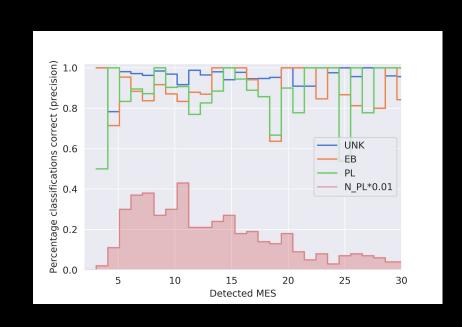








Results on TESS



















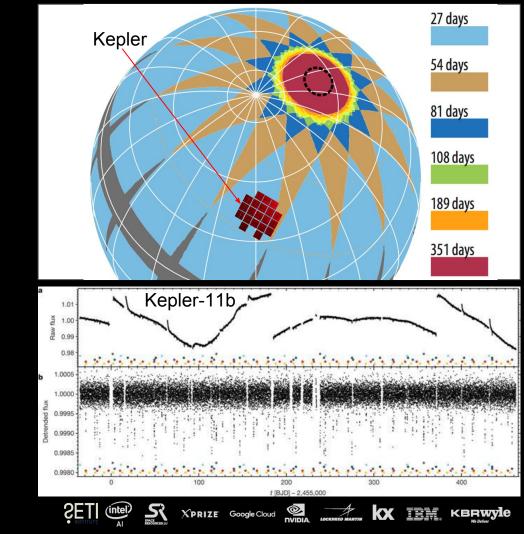




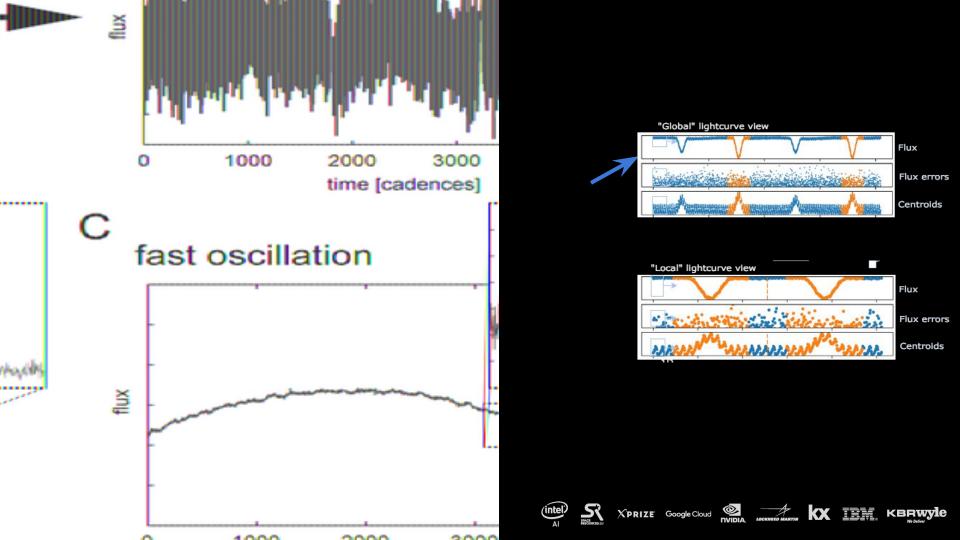
Kepler Data

Real data but noisy labels

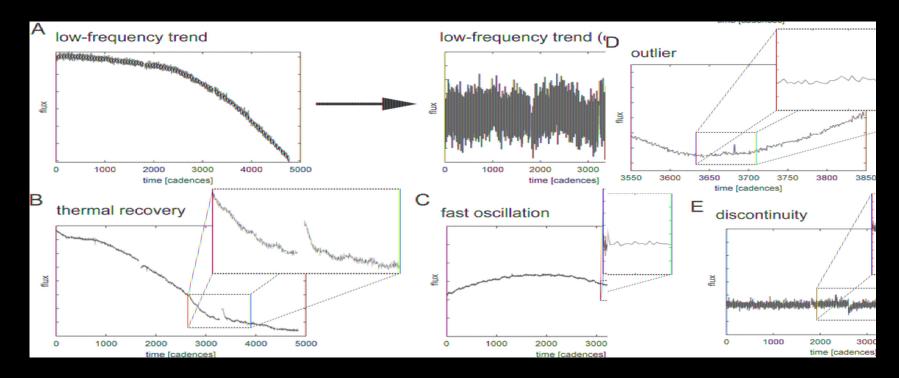
- 150 000 lightcurves
- Lightcurves 4 years in duration
- ~4 000 planets (candidates & confirmed)
- Can augment data to TESS-like 27-day campaigns (up to 7.8 million 'targets')







The data: real light curves















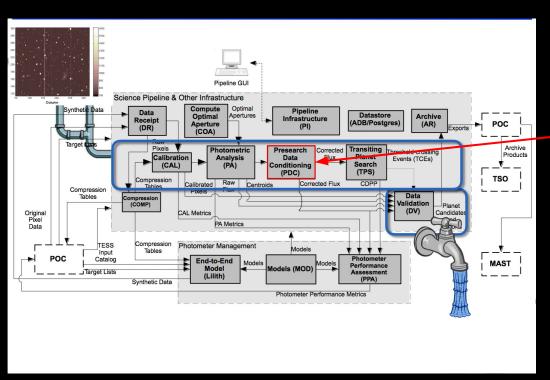


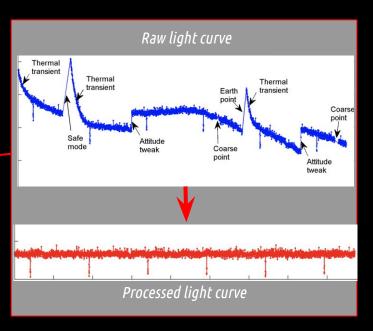






The data: detrended light curves





 Remove trends present on all stars to leave only astrophysical signal













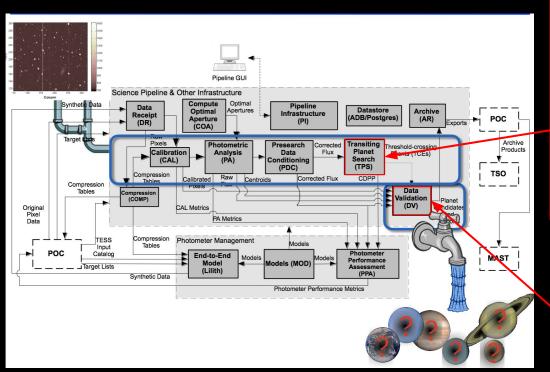


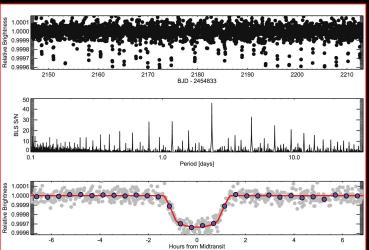






The data: Transiting Planet Search





- Search for transiting planets in frequency domain.
- Strong candidates are analysed using statistical tests (DV)













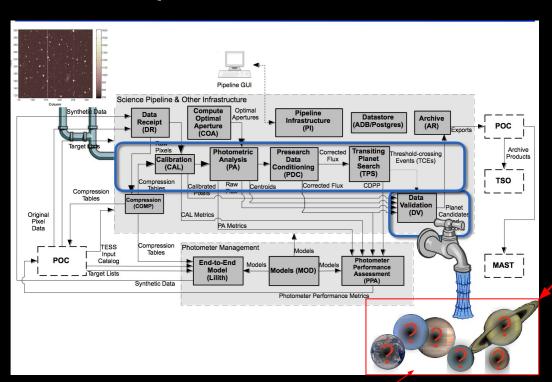








The data: planet candidates



Planet candidates (TCEs) are analysed to determine which are planets.

Often performed:

- By human vetters
- Using classical statistical techniques
- Using Machine Learning













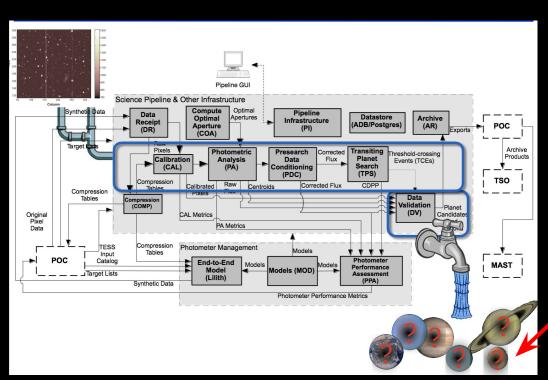








Incremental: Classify planet candidates detected by the pipeline



Classify candidates detected by TESS pipeline

Planet, EB, BEB?

















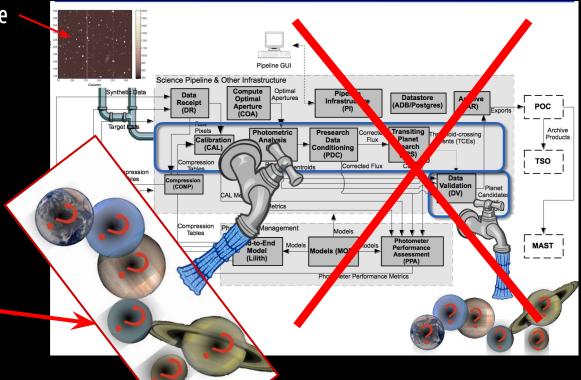


Innovative approach: Classifying directly from pixel data

Target Pixel File

Cut out the pipeline entirely

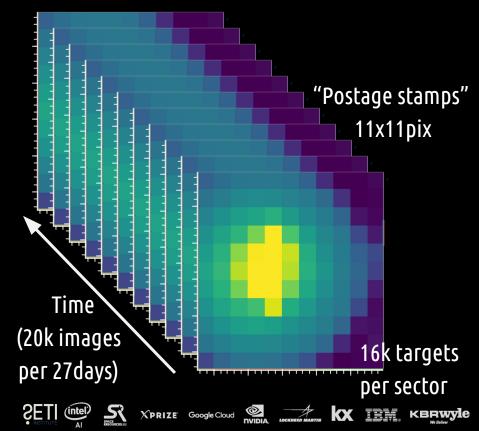
Planet, EB, BEB?



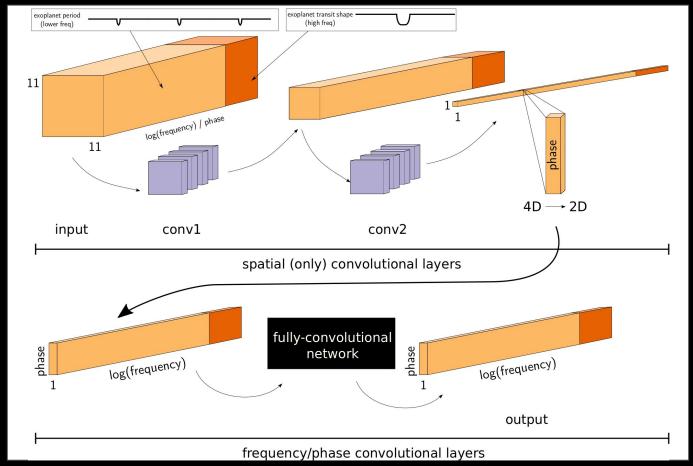


Innovative approach: Classifying directly from pixel data

- Train a Convolutional Neural Network (CNN) directly on pixels
- Transform image to more appropriate representation, i.e. frequency domain for periodic data
- Incorporate our domain knowledge into the network architecture and learning algorithm, rather than the data











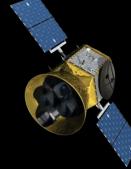












Next steps...

Week 3:

- Avenue 1: Re-produce Shallue & Vanderburg 2017 results
- Avenue 2: Setup data infrastructure/get basic CNN training

Week 4:

- Avenue 1: Incorporate new domain knowledge
- Avenue 2: Iterate on model/implement data balancing methods











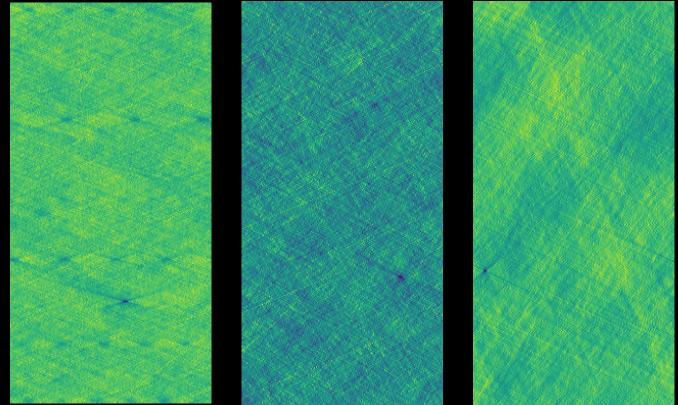








Planets



















No planets

