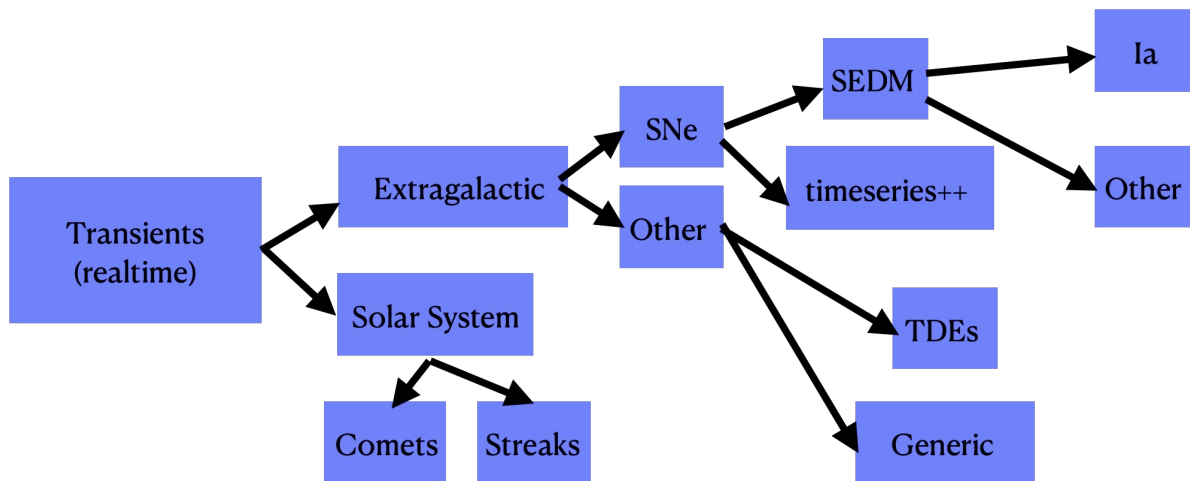


RAPID Insights: Real-Time Transient Classification for Roman



Caltech



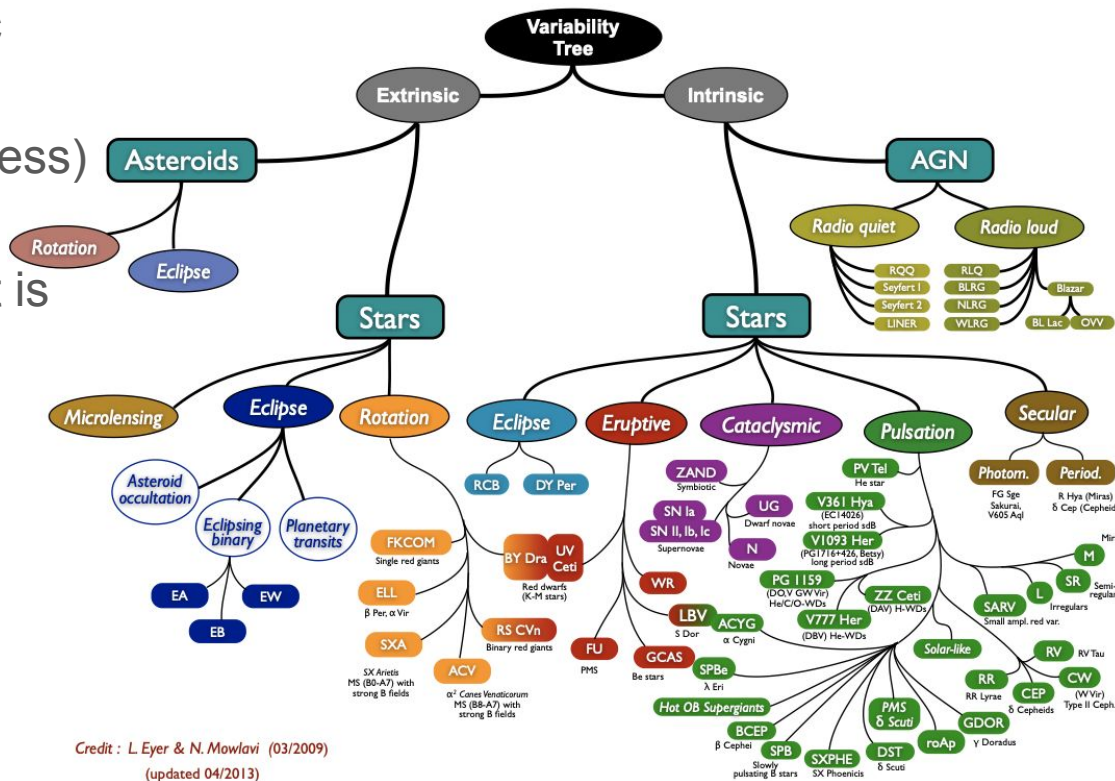
Ashish Mahabal on behalf of RAPID
Deputy Director, Center for Data Driven Discovery, Caltech
Transients from Space, STScI, Baltimore, 13 Mar 2025

Outline

- Brief intro to Roman/RAPID
- Classification needs (of the community)
- Availability of data/methodology/resources
- Ongoing efforts/status

Brief intro to Roman

- Surveys: High and low Galactic latitudes; Time Domain
- Depth: Exquisite (though area less)
- Types of objects: Everything
- False positives: Everything that is not in training samples



Credit : L. Eyer & N. Mowlavi (03/2009)
(updated 04/2013)

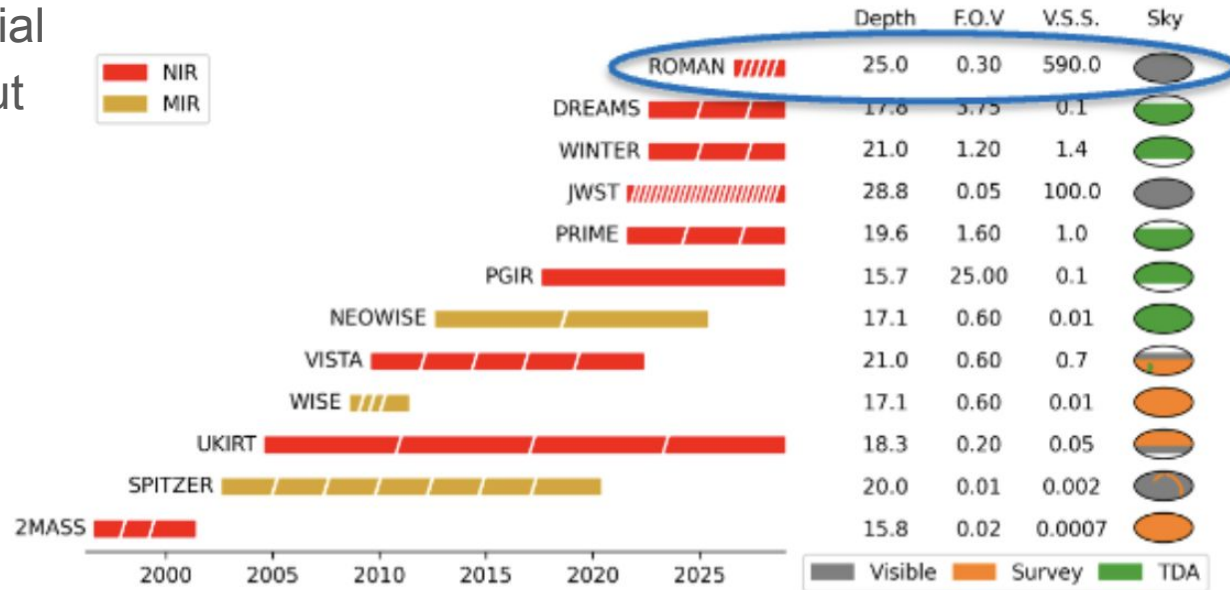
Resolution: 0".11 arcsec
Depth: 26th mag

RAPID

- Roman Alerts Promptly from Image Differencing - Project Infrastructure Team
 - PI: Kasliwal. Also see poster by Schuyler Van Dyk
- Tasks (from L2 data):
 - Image tiling,
 - image differencing,
 - rapid detection and classification,
 - and low latency event alerts

Data ... Methodology ... Resources

- Detailed classification and subclassification non-trivial
- Not just single images but also coadds
- We will use existing data and modern machine learning methods



V.S.S. = volumetric survey speed (normalized to PRIME*)
(*PRIME = PRime-focus IR Microlensing Experiment)

Availability of methodology

- CNNs and vision transformers for image-based data
- XGBoost and transformers for time series data (and derived features)
- Foundation Models
- Detection
- Classification
- Clustering
- Transfer learning

Extensive training sets needed

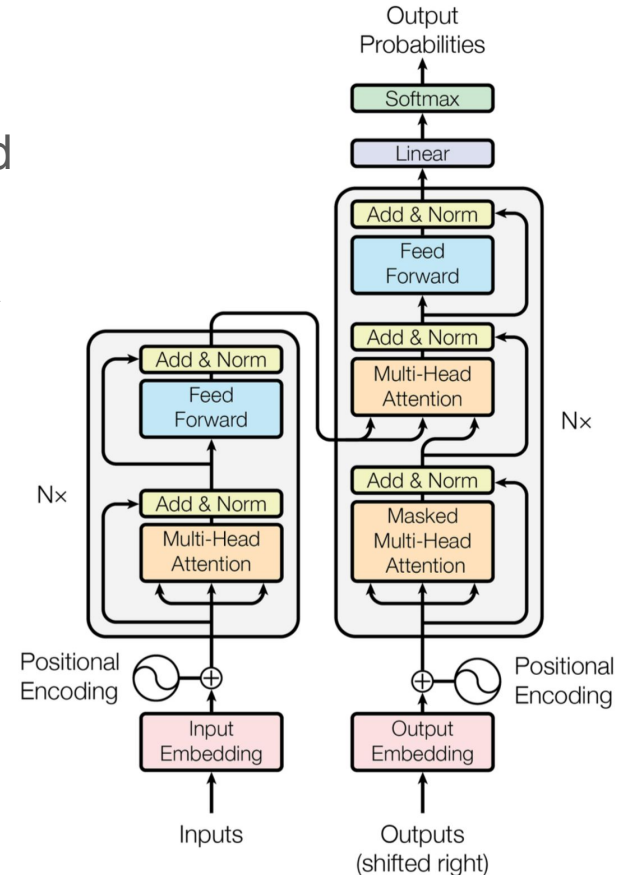


Figure 1: The Transformer - model architecture.

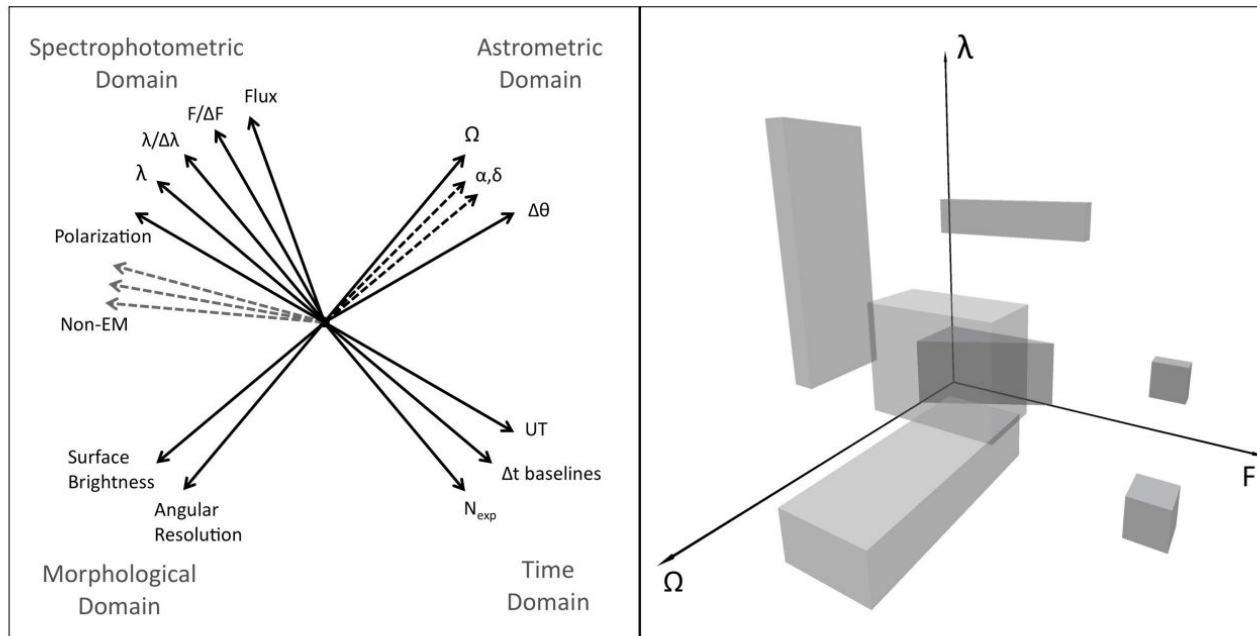


Figure 1. A schematic illustration of the Observable Parameter Space (OPS). All axes of the OPS corresponding to independent measurements are mutually orthogonal. Every astronomical observation, surveys included, carves out a finite hypervolume in this parameter space. *Left:* Principal axes of the OPS, grouped into four domains; representing such high-dimensionality parameter spaces on a 2D paper is difficult. *Right:* A schematic representation of a particular 3D representation of the OPS. Each survey covers some solid angle (Ω), over some wavelength range (λ), and with some dynamical range of fluxes (F). Note that these regions need not have orthogonal, or even planar boundaries.

[arxiv:1203.5111](https://arxiv.org/abs/1203.5111)

Djorgovski, Mahabal et al.

IR Surveys included:

2MASS
UKIRT
VISTA
WISE
SWIRE
GLIMPSE

One big difference today is machine learning

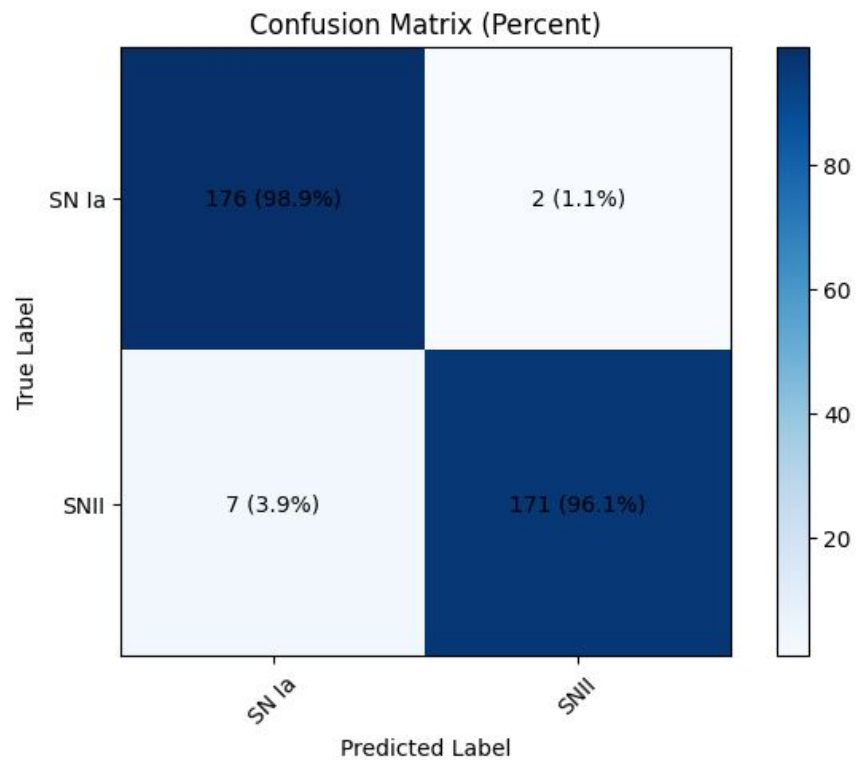
- It can be used in many ways
- Transfer learning from brighter samples will do only so much
- Simulating all types for the depth of Roman will be needed
- Right now limited simulations
- Explosive events mostly

<https://arxiv.org/abs/2501.05632>

Table 1. List of transient models and summary information for each model, including number of events, rest-frame phase T_{rest} (days) with $T_{\text{rest}} = 0$ at peak brightness, and artificial rate multiplier.

model	Number of generated: events SEDs		T_{rest} range (days)	integer type	rate multi- plier
SN Ia	224,575	67,070,894	−20, +300	10	1
SNIax	165,703	31,978,600	−50, +300	12	1
SNII	694,064	144,330,844	−50, +200	32	1
SNIb	148,722	33,604,658	−50, +200	21	1
SNIc	148,724	33,753,628	−50, +200	26	1
SLSN-I	1,128	398,547	−50, +300	40	1.5
TDE	3,784	583,063	−100, +250	42	10
PISN-H	113	33,694	−50, +200	57	4
PISN-He	112	35,318	−50, +200	58	4
KN	53	12,785	−5, +30	50	100
RanMag	27,884	138,855	−100, +300	99	—
Total	1,386,978	311,802,031	—	—	—

Easy classification



With Alex Kumar, Caltech

Easy classification?

But what if this classifier sees something else?

What if the proverbial blindfolded men feel this animal's trunk or ears?

They will think it is an elephant!



Colossal Biosciences would like to bring back extinct animals.
They recently took the woolly genes of a woolly mammoth



And ...

And injected them into a mouse to get a wooly mouse!



Don't believe what
Saturday Night Live said
3 days back

That the next step is to
mate this mouse with an
elephant to get wooly
mammoth

Because ...

That would be a bit like transfer learning without being adaptive

Colossal Biosciences plan to instead introduce the same genes into elephants

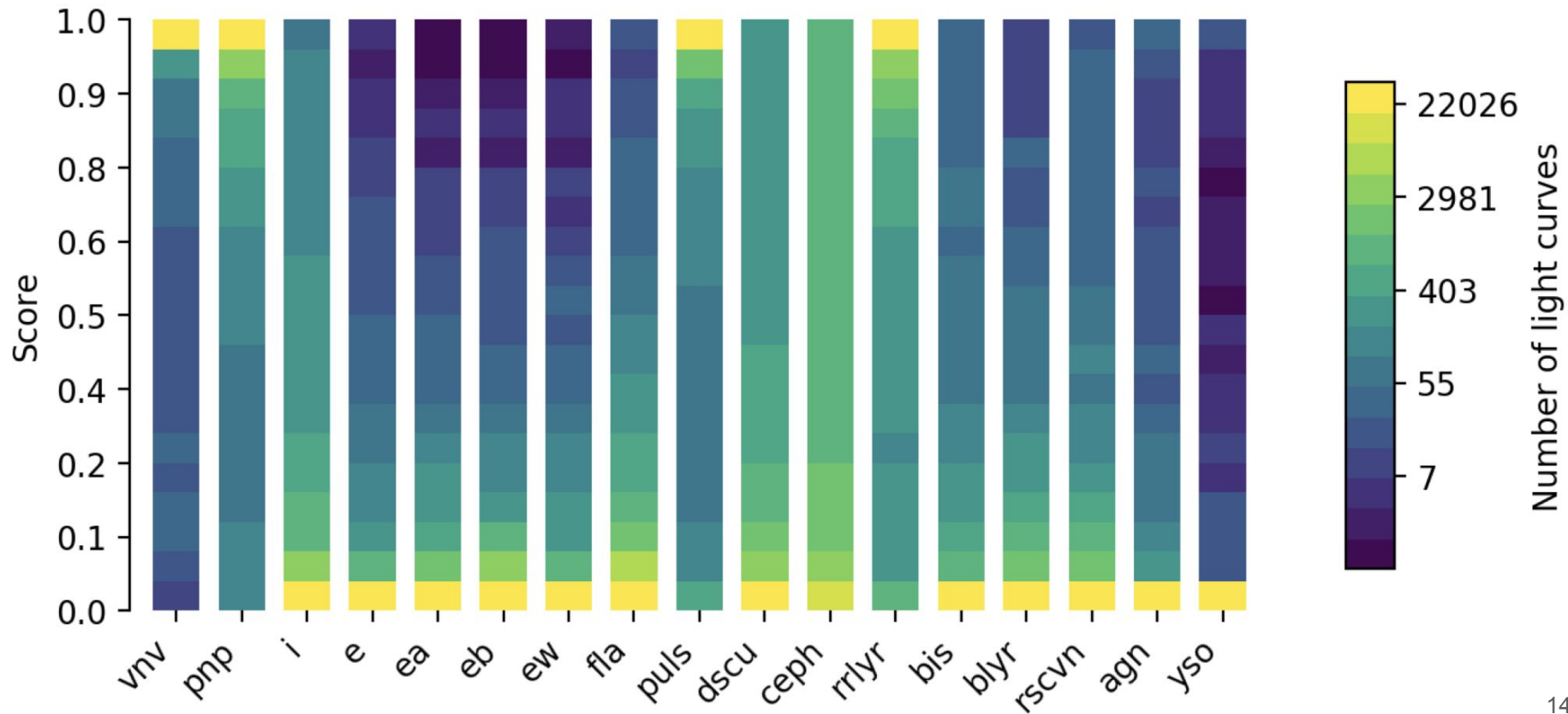
We want exactly that!

Besides the explosive events we want every other kind of known object in our simulated data.

Then we can build multiple one-versus-a classifiers like we have done for ZTF



Using multiple classifiers together



STRIDE - Roman's Time Domain Working Group

Monthly meetings - 10 AM PT, first Fridays

Co-chairs: Ashish Mahabal, Tyler Pritchard

Encouraging collaboration across different stakeholders

Solar System, stellar sources, extragalactic events ...



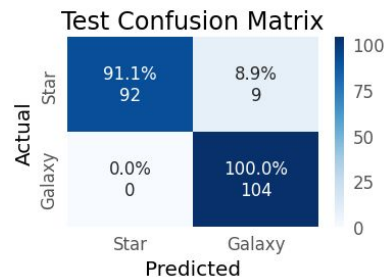
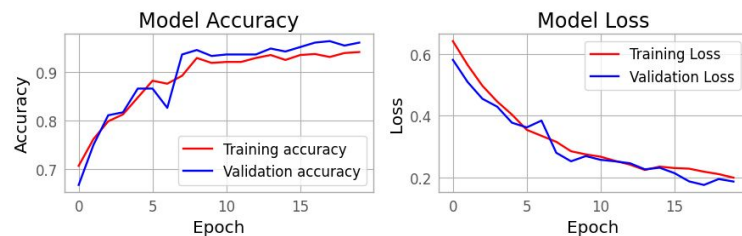
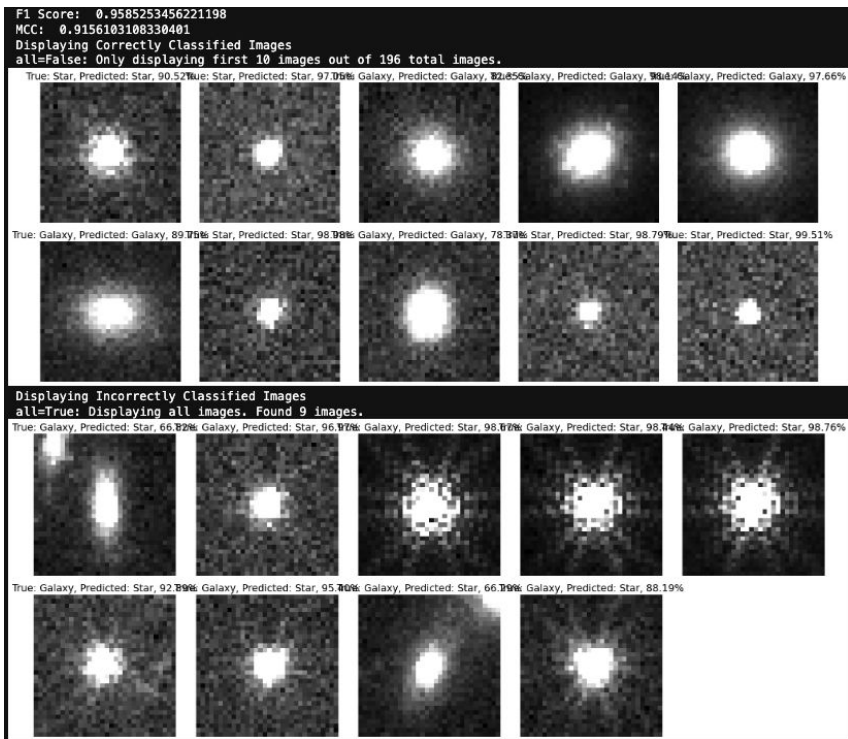
<https://outerspace.stsci.edu/display/RSWGS> to join the Roman Science Working Groups

ROMANTDAWG@MAILLIST.STSCI.EDU to join the mailing list

Roman Space Telescope Slack: # wg-tda-stride

https://join.slack.com/t/romanspacetelescope/shared_invite/zt-2mcscgq0x-24pp8xJr5gqiq9yBvvutJA

Simple star/Galaxy Classifier

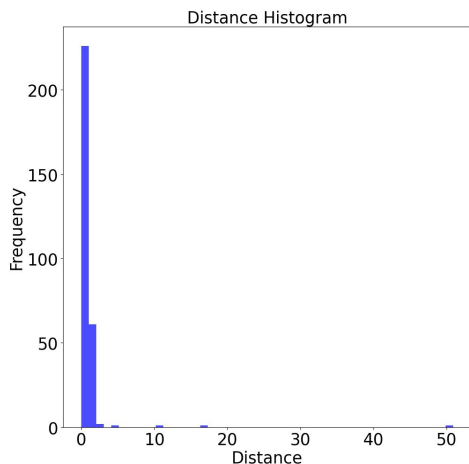


Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 32, 32, 25)	250
max_pooling2d (MaxPooling2D)	(None, 16, 16, 25)	0
dropout (Dropout)	(None, 16, 16, 25)	0
flatten (Flatten)	(None, 6400)	0
dense (Dense)	(None, 40)	256,040
dense_1 (Dense)	(None, 2)	82

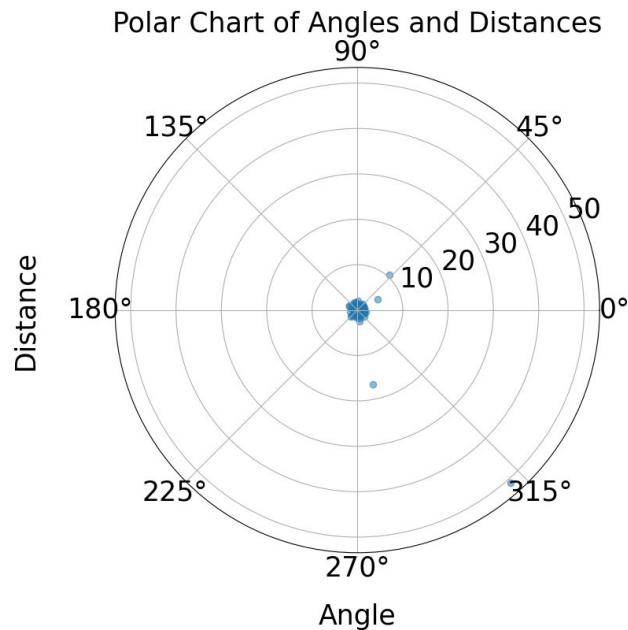
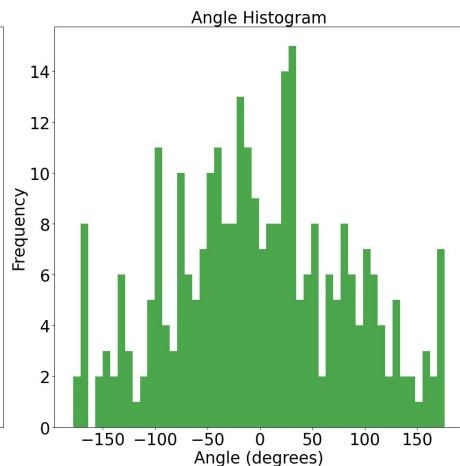
Image subtraction without subtraction

- 98% within 2 pixels
- Angle Histogram has normal distribution, no angular bias
- Polar chart confirms lack of bias

Distance



Angle

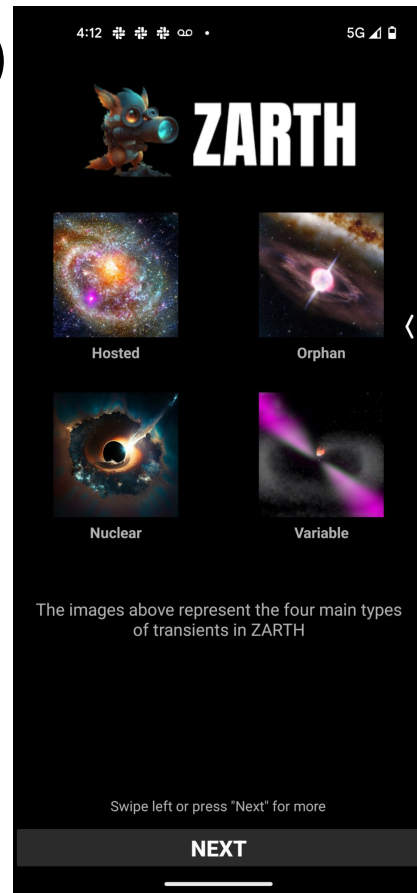


With Heramb Patil, Caltech

Data challenges, Training sets, (citizen science?)

Could duplicate ZARTH for Roman

ZARTH has ambiguous sources that
players could help create training
sets with



Summary

RAPID is moving along with existing sets

Extensive training samples required to avoid false positives

Many modern ML algorithms will provide great results after that