RAPID

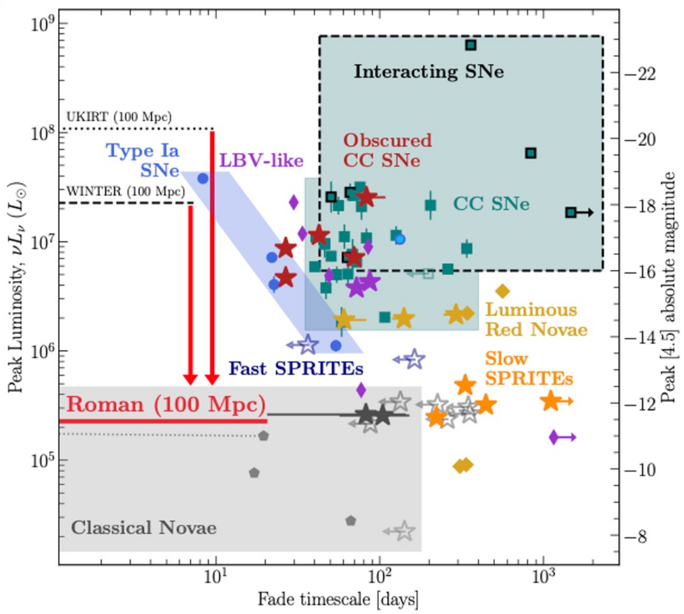
(Roman Alerts Promptly from Image Differencing)

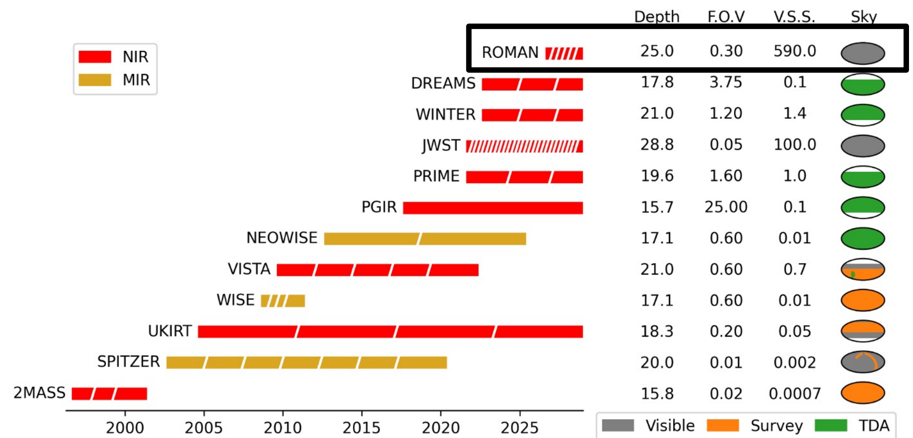
A Roman Space Telescope Project Infrastructure Team (PIT)

Mansi Kasliwal (PI; Caltech)
Ben Rusholme (Pipeline Development Lead; IPAC)
Schuyler Van Dyk (Science Definition Lead; IPAC)
Roberta Paladini (Galactic transients, Inclusion Lead; IPAC)
Lin Yan (Testing and Validation Lead; Caltech)
Ashish Mahabal (Machine Learning Lead; Caltech)
Joe Masiero (Solar System Objects Lead; IPAC)
Russ Laher (Senior Applications Developer; IPAC)
Jacob Jencson (Testing and Validation Scientist; IPAC)

Science motivation for RAPID: Enabling Time-Domain

Transient discovery in Roman HLTDS, HLWAS, GBTDS, and GAS





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

CCS: HLTDS=High-Latitude Time Domain Survey + HLWAS=High-Latitude Wide-Area Survey + GBTDS=Galactic Bulge Time Domain Survey (~75%)
GAS=General Astrophysics Surveys (~25%)

Our goal is to provide four services to the Roman community:

- Rapid image-differencing of every new Roman image from a reference image
- Prompt public alert stream of all Roman transient and variable candidates
- Source-matched light curve files for every identified Roman candidate
- Forced-photometry service for photometric history at any observed location

There was no existing plan by the Roman project to deliver a rapid time-domain discovery alert system

The Importance of Starting Now

- Develop a prompt Roman time-domain alert system years before launch
  - ➢ ***We need to be ready by Day One***
- Pipelines, algorithms, schema need to be defined and tested well in advance
- Product specifications and example alert packets need to be communicated to downstream consumers (astronomers, event brokers) to allow enough time to prepare for our products
- The prompt alert system will need to be validated and tuned in operations …
- … but, the system needs to be in place before Roman data are acquired

The Technical Plan

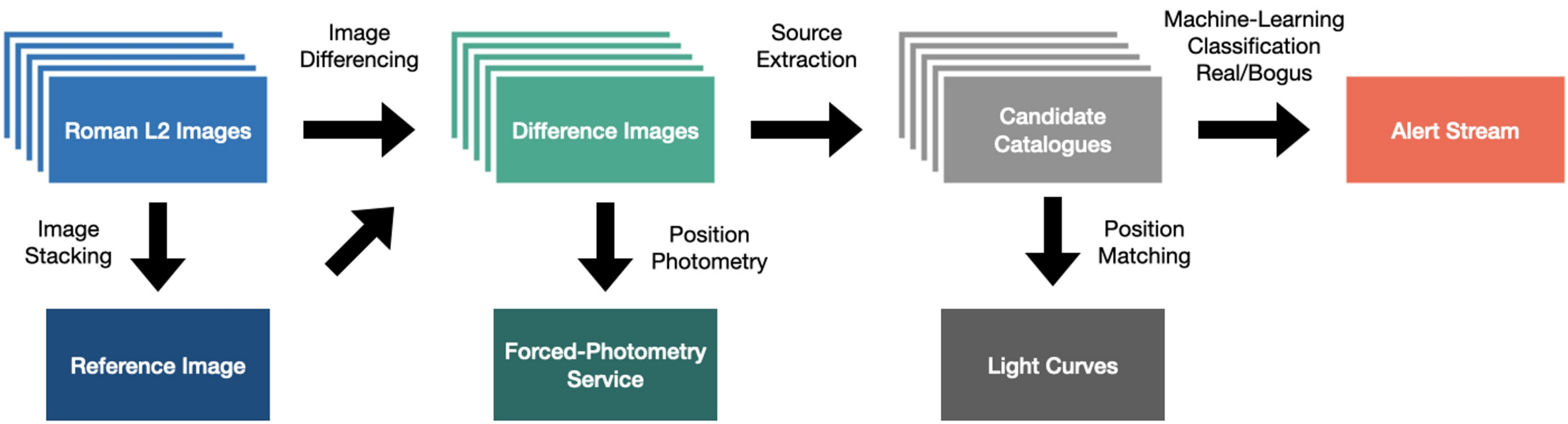
- Pull calibrated Level-2 WFI data from the SOC staging location

  (< 48 hour turnaround)
- Execute image differencing and prompt public alert broadcasting (< 1 hour)
  - Includes initial source classification via machine learning
- Offer forced photometry on difference images
- Append and release light curve history or sources

  (< 24 hour)
- Archive public alerts via STScI MAST

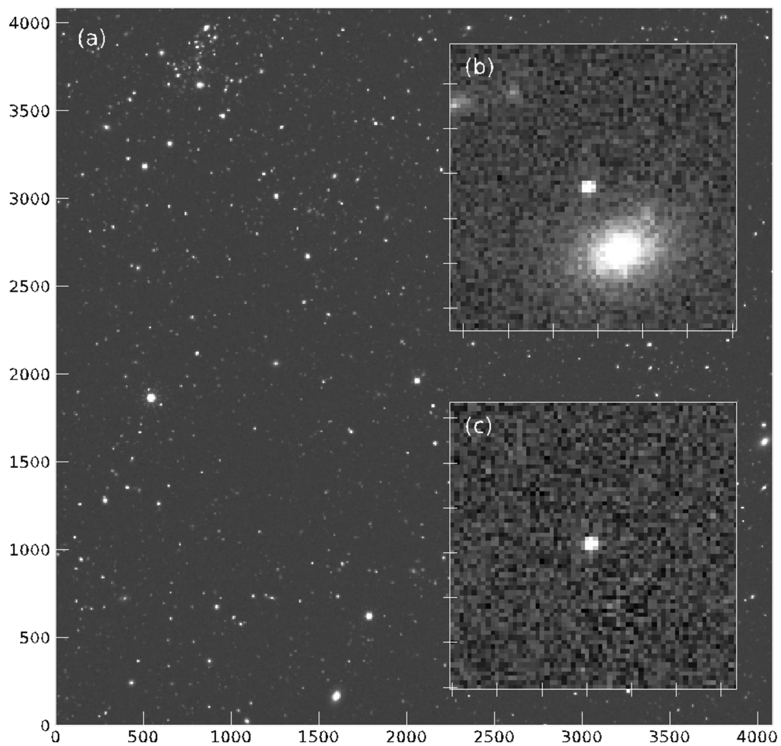
RAPID Pipeline Flow

Leveraging our previous experience with the Zwicky Transient Facility (ZTF)



<u>End-to-end testing and validation</u>

Building Reference Images

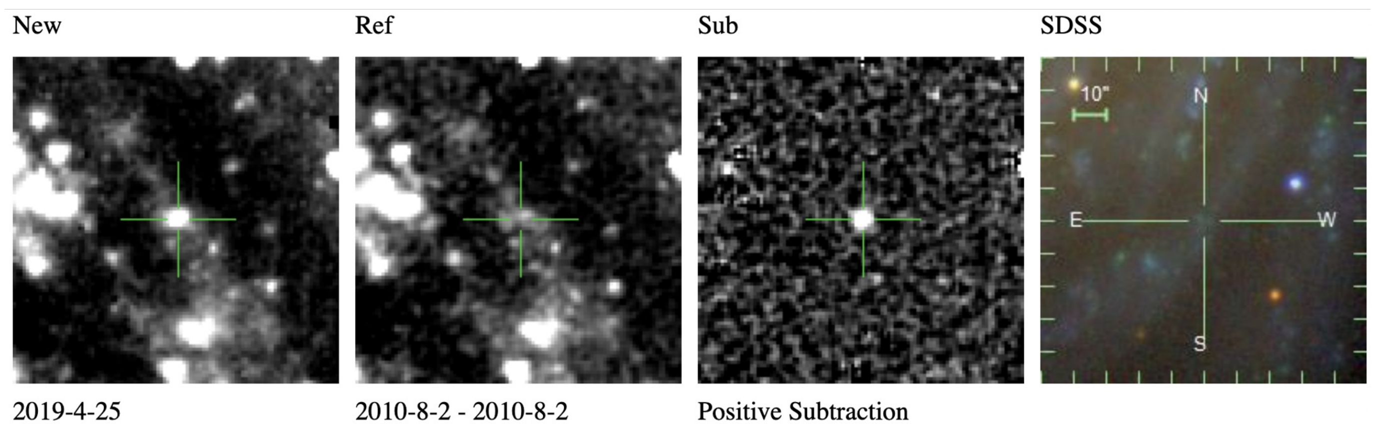


Proof-of-concept image differencing of a single simulated Y-band Roman SCA. From Wang et al. (2022).

- Challenges:
  - ➢ No fixed on-sky field grid
  - ➢ No set orientation
  - ➢ No constant exposure time
  - ➢ Repeat visits of the same sky will only partially overlap, with different depths
- Define fixed sky grid with cells 1/9 of Roman FoV *(per available band)*
- For each new image, identify which cells completely or partially overlap fixed grid
- If overlap >50%, add to counter of cells and increment the exptime at sky position
- **Nominally, first epoch of a given cell becomes the reference image**
- That reference image superseded by next deeper exposure
- Next shallower exposure triggers image differencing
- *May build reference mosaics*

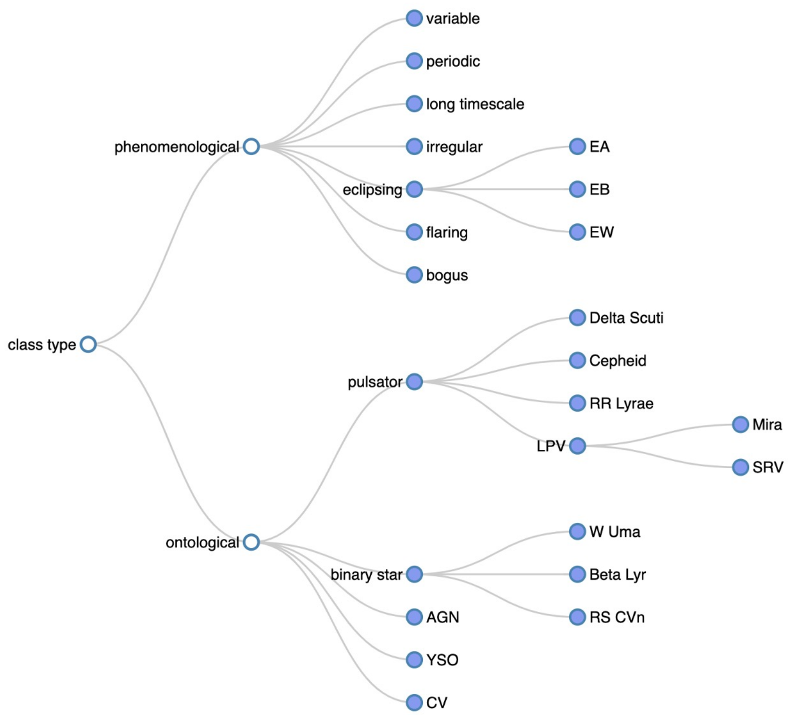
Optimizing Image Differencing

- *Challenges*: mismatch in SCA, mismatch in roll angle, undersampled PSF
- Currently testing various difference imaging algorithms to optimize
- Will experiment whether drizzling, blotting are necessary
- *Metric*: maximal recovery of injected simulated light curves and minimal false positives



Proof-of-concept image differencing of mismatched, partially covered, and sparse PSF sampled data. From Kasliwal et al. (2017).

Source Classification via Machine Learning



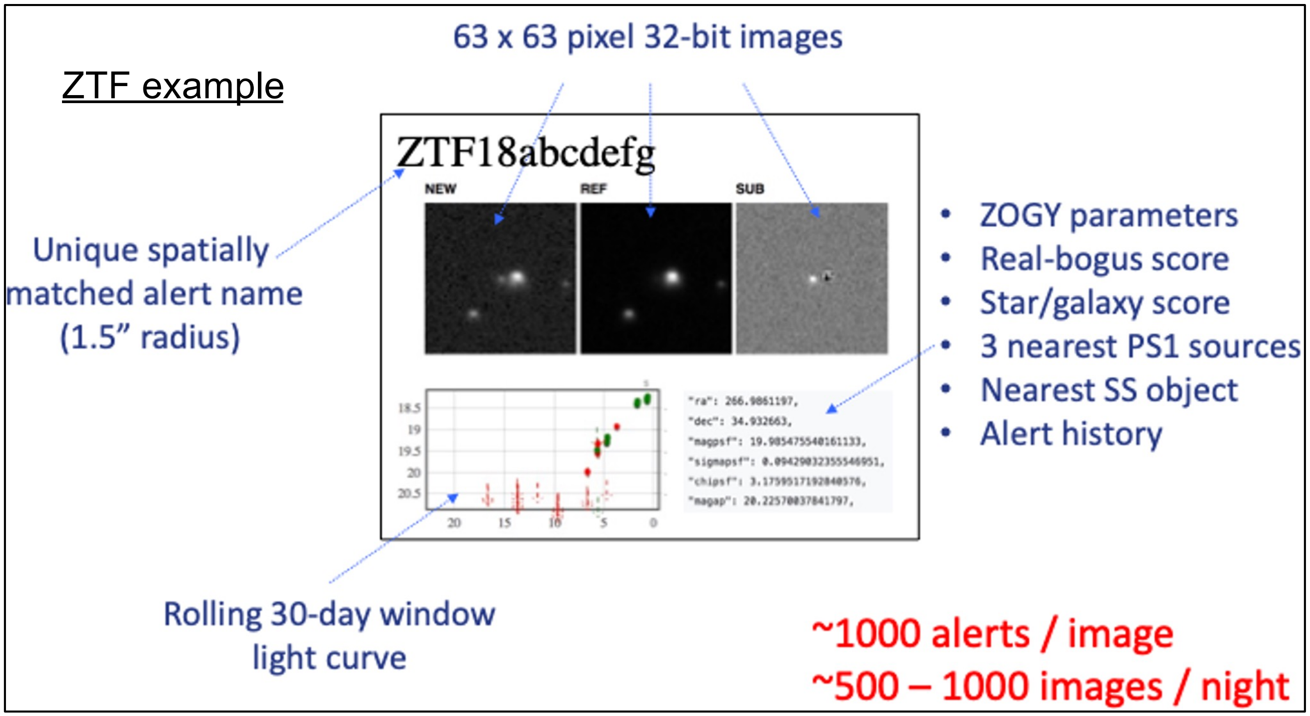
EA/EB/EW - classes of binaries; LPV - Long Period Variable; SRV - Semi-Regular Variable; AGN - Active Galactic Nucleus; YSO - Young Stellar Object; CV - Cataclysmic Variable; W Uma - W Ursa Majoris; RS CVn - RS Canum Venaticorum

- Hierarchical binary source deep-learning classifiers robustly and successfully implemented in ZTF
- Allows users to choose thresholds
- Real/bogus score to flag artifacts
- Classification included in alert packet
- *Work to be done before launch to optimize classification for Roman*

Pre-launch development and testing underway

- Existing HST WFC3/IR and JWST public data
  - ➢ RAISIN, SIRAH, other programs
- Existing simulated Roman WFI data
  - ➢ Cross-team Rubin/Roman NASA Open Universe sims (2024)
  - ➢ SSC MSOS/RGES Galactic bulge sims
- Need to inject additional transients, variables, Solar System objects

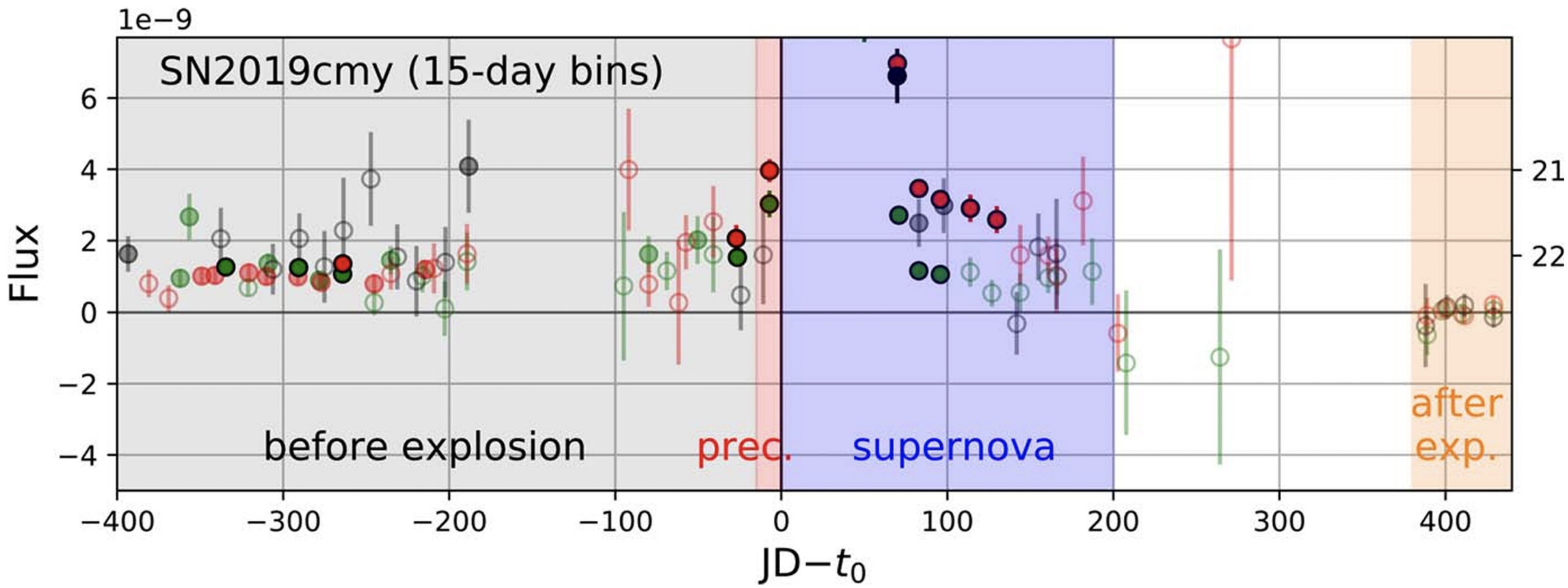
Alert Stream: prototyping underway



Raw stream to be picked up by brokers (ALeRCE, Fink, ANTARES, Lasair, etc.)

Stream could easily be extended for use by HLTDS and GBTDS

Forced Photometry and Light Curves



Complete ZTF light curve of SN 2019cmy – upper limits (open circles) and detections (solid circles) via forced photometry (from Strotjohann et al. 2021).

Data Products/Deliverables

Product	**Format**	**Distribution**	**Access**	**Size**
Simulated Images	ASDF	S3/MAST	Public	$\mathcal{O}$(5 TB)
Reference Images	ASDF	S3/MAST	Public	$\mathcal{O}$(10 TB)
Difference Images	ASDF	S3/MAST	Public	$\mathcal{O}$(100 TB)
Alert Stream	Avro	Kafka	Public	$\mathcal{O}$(1 GB/day)
Alert Archive	Avro/tarball	MAST	Public	$\mathcal{O}$(20 TB)
Light Curves	Parquet	S3/MAST	Public	$\mathcal{O}$(10 TB)

Optimizing Image Differencing

Considering **ZOGY** (Zackay, Ofek & Gal-Yam 2016) and **SFFT** (Hu+ 2022)

ZOGY: Cross-convolution based

✓ Computationally less expensive

✓ Numerically stable and symmetric

✓ “Optimal” for transient detection

✓ Minimal tuning, significant IPAC experience from ZTF

O Requires PSFs

✗ Spatial variation requires subdivisions

SFFT: Convolution-kernel based (à la Alard & Lupton)

✗ Computationally expensive and memory intensive

✓ Fast on GPUs

✗ Numerically unstable and asymmetric

O Many parameters, requires tuning and more preprocessing

✓ No PSFs required*

✓ Handles spatial variation automatically*

Optimizing Image Differencing

Considering **ZOGY** (Zackay, Ofek & Gal-Yam 2016) and **SFFT** (Hu+ 2022)

ZOGY: Cross-convolution based

✔ Computationally less expensive

✔ Numerically stable and symmetric

✔ "Optimal" for transient detection

✔ Minimal tuning, significant IPAC experience from ZTF

O Requires PSFs

✗ Spatial variation requires subdivisions

Hybrid: Cross-convolution + SFFT (see Hu & Wang 2024 on JWST/NIRCAM)

✗ Computationally expensive and memory intensive

O Fast on GPUs

✗ Numerically unstable and asymmetric

O Many parameters, requires tuning and more preprocessing

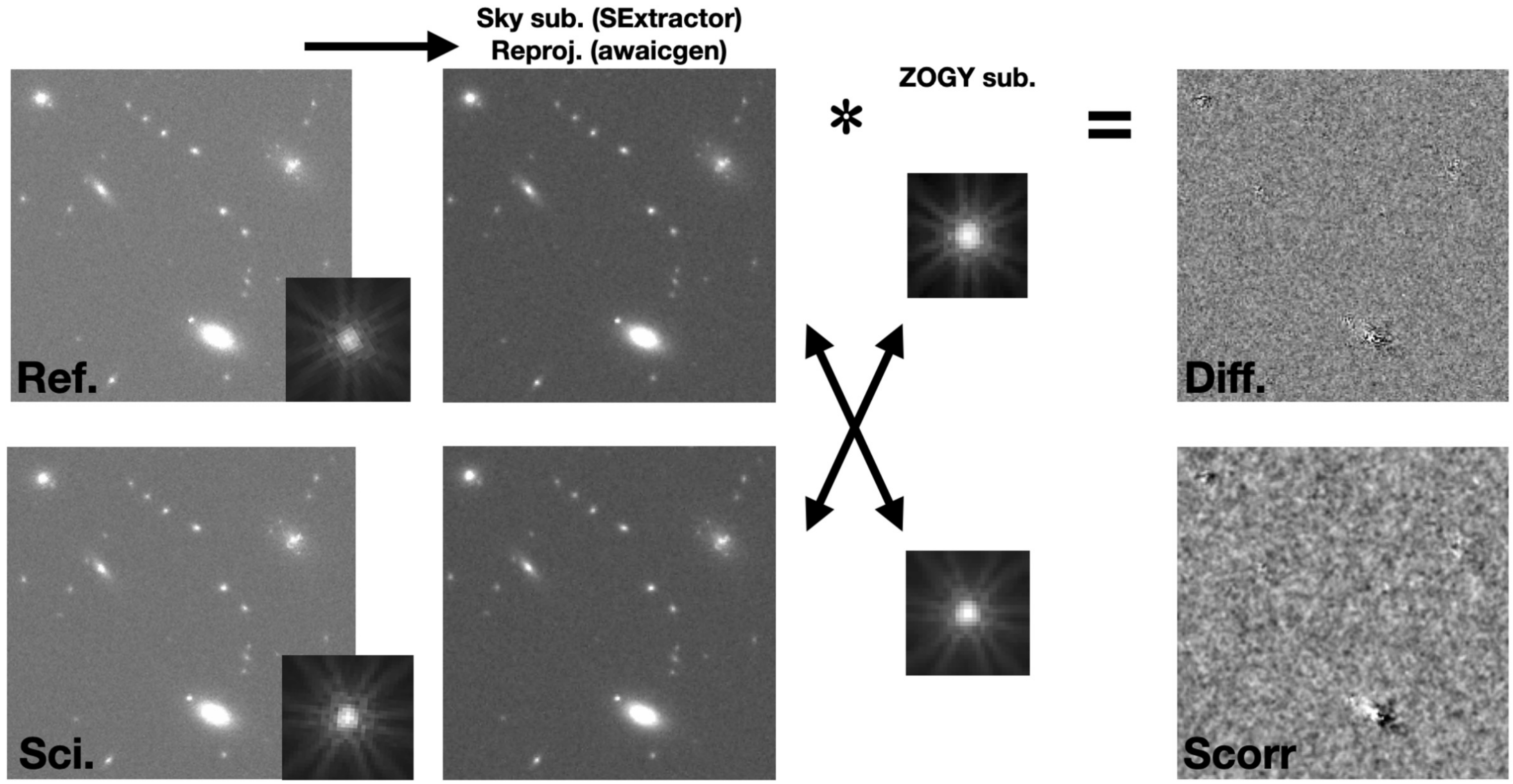
O Uses PSFs

✗ Deconvolution requires subdivisions

Optimizing Image Differencing

Testing underway with NASA OpenUniverse HLTDS Simulations

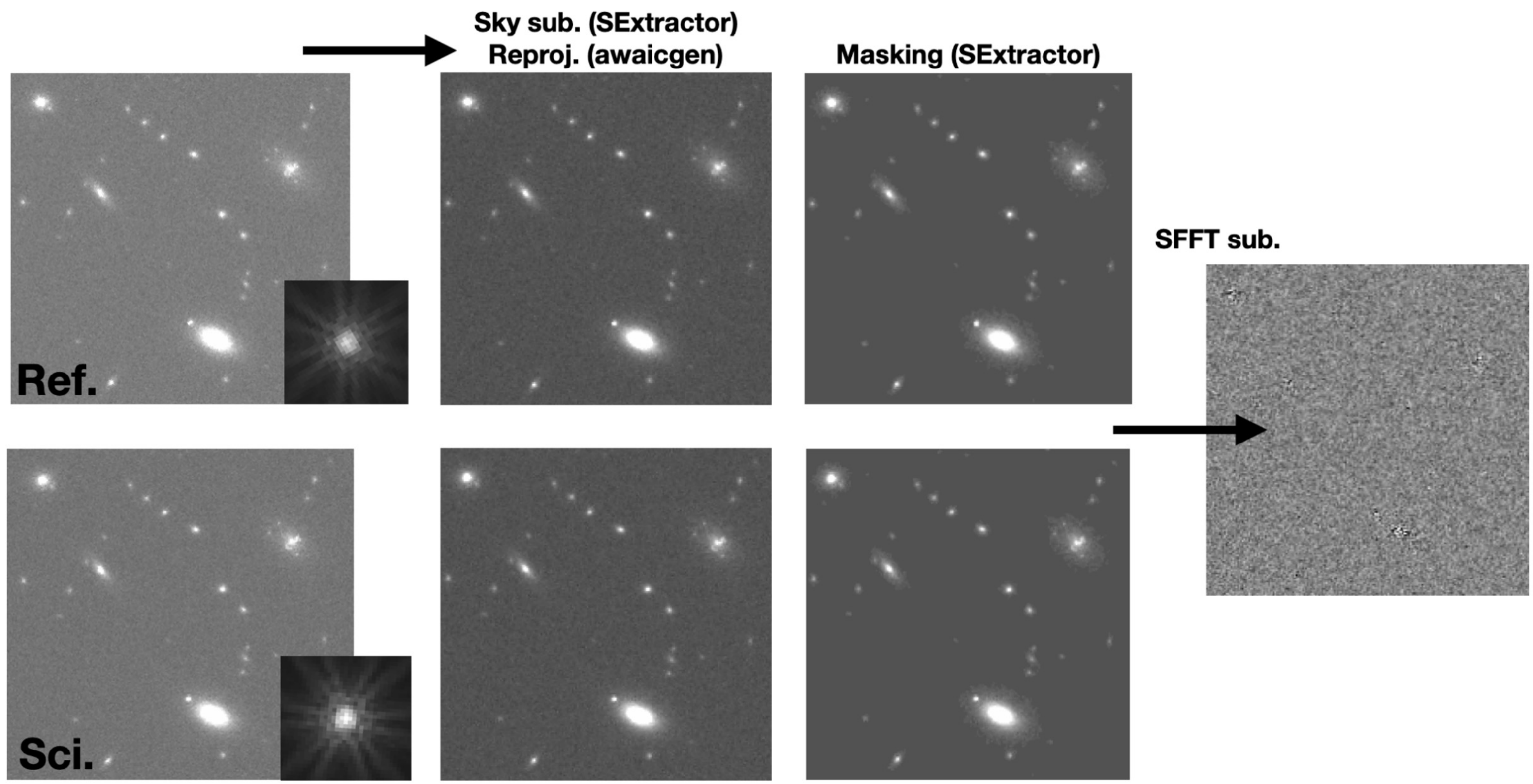
(Thank you Alina Kiessling, Michael Troxel, Dan Scolnic, Rebekah Hounsell, Rick Kessler)



Optimizing Image Differencing

Testing underway with NASA OpenUniverse HLTDS Simulations

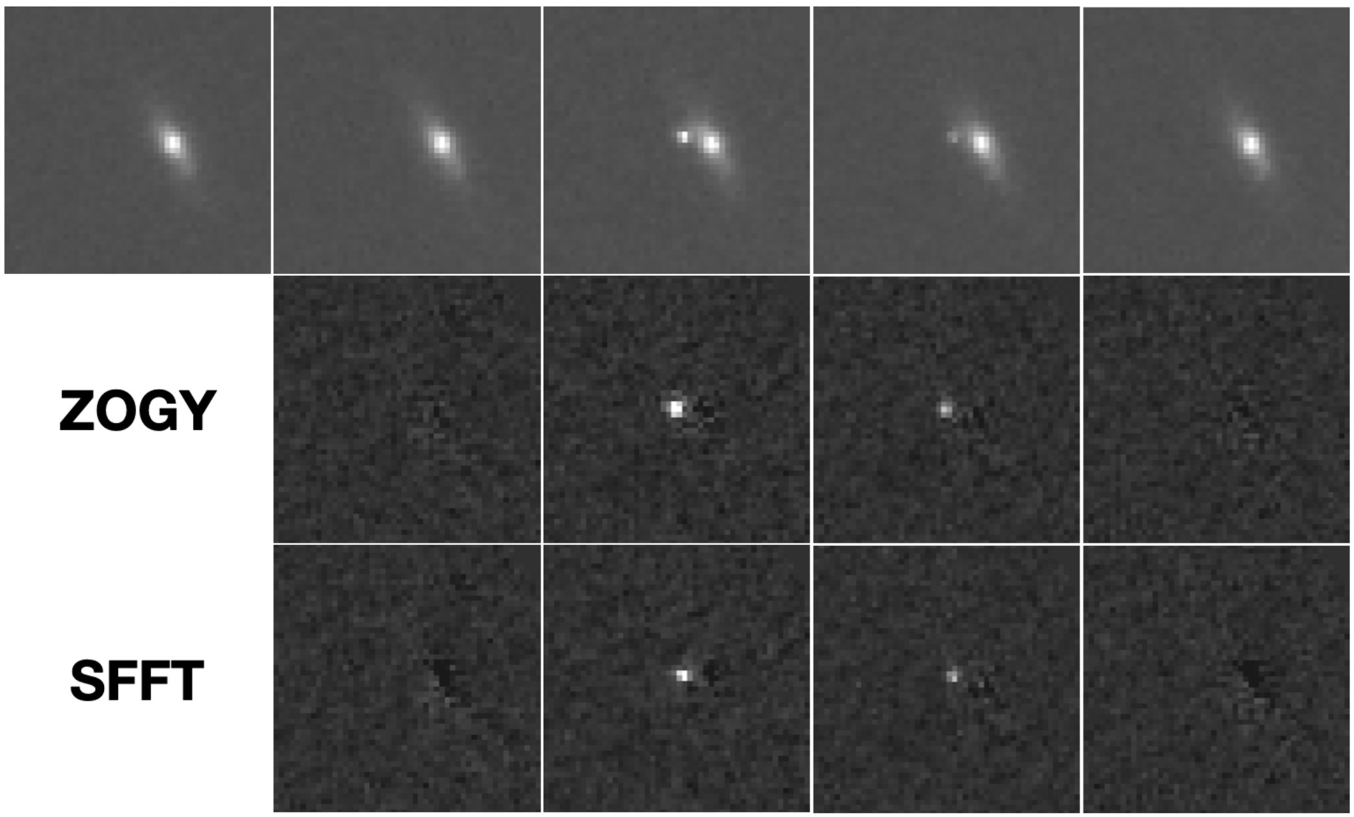
(Thank you Alina Kiessling, Michael Troxel, Dan Scolnic, Rebekah Hounsell, Rick Kessler)



Optimizing Image Differencing

Testing underway with NASA OpenUniverse HLTDS Simulations

(Thank you Alina Kiessling, Michael Troxel, Dan Scolnic, Rebekah Hounsell, Rick Kessler)



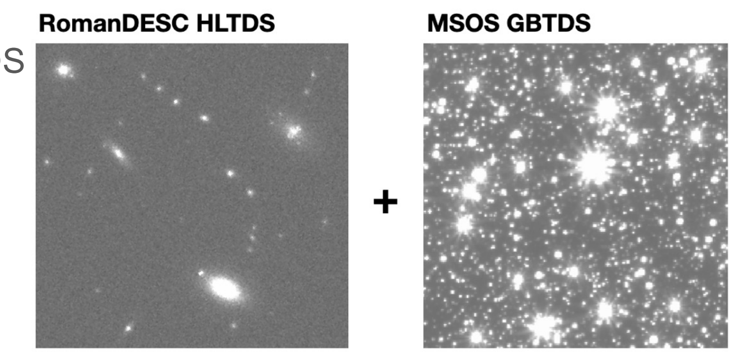
Optimizing Image Differencing

Evaluation Metrics

- Basic difference image metrics, e.g., χ2 (noise properties, artifacts)
- Set concrete and achievable requirements for:
    - Transient injection recovery statistics: completeness & purity
    - Recovered photometric accuracy
- Cost, speed, and resources

Other considerations and planned work

- Performance across variety of environments/situations
    - Expand to Microlensing Science Operations System (MSOS simulations of GBTDS *(Thank you MSOS and RGES PIT)*
    - Faintest transients, nuclear events, variables
    - Compare with SN PIT work on SFFT for extragalactic SNe
- Current simulations are “perfect”
    - Explore impact of imperfect astrometry / registration errors
    - Explore PSF model inaccuracies
    - Cosmic rays and detector effects
    - Tests on real data: HST and JWST
- Algorithmic improvements, e.g. Hybrid + B-spline SFFT (Hu & Wang 2024), TRANSLIENT (Springer+ 2024)



Machine Learning Progress

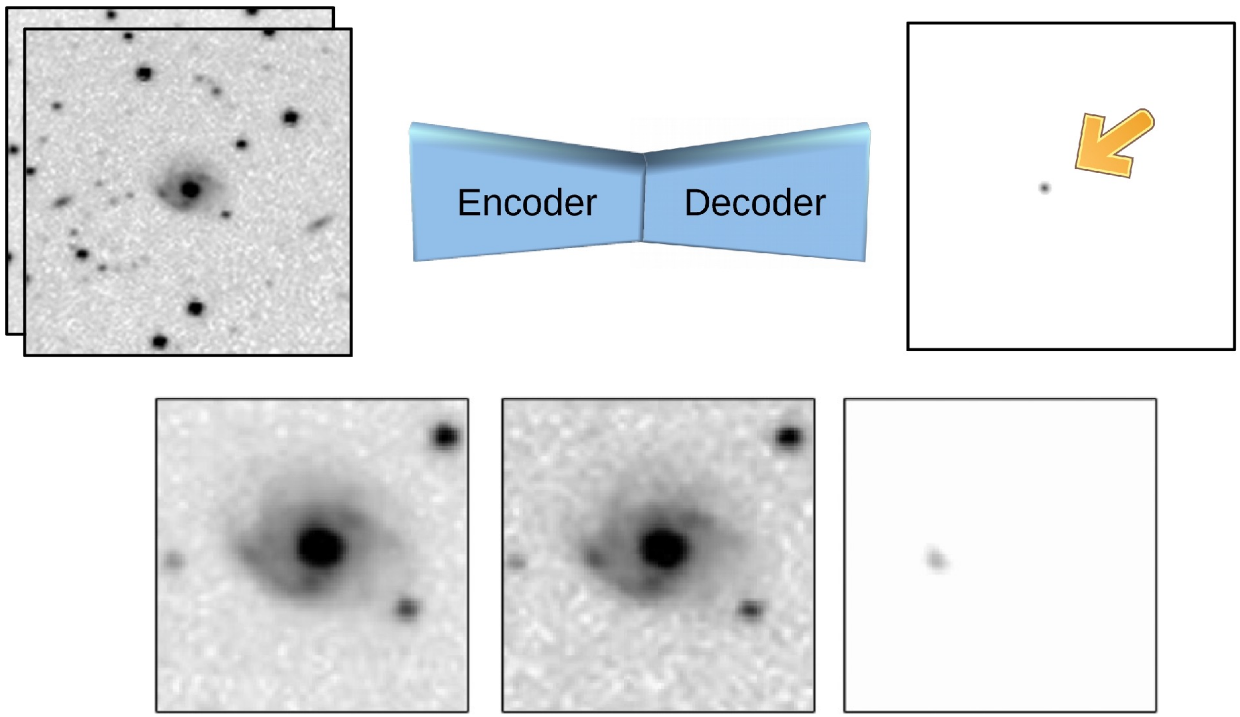
We have started exploring current available simulations
The plan is to:

- start detecting injected transients
- Start classifying them if there is a large enough set
- Understand issues near edges, etc.
- While we expect very few instrumental artifacts, we will monitor for artifacts due to saturated transients, etc.
- May need to perform most training at the catalog, rather than the image, level

We will use convolutional neural networks (CNNs) and may also explore other ML techniques/workflows

If enough data are available we may attempt transfer learning of TraniNet, a transient detector without image subtraction

Image subtraction for hunting transients without subtraction

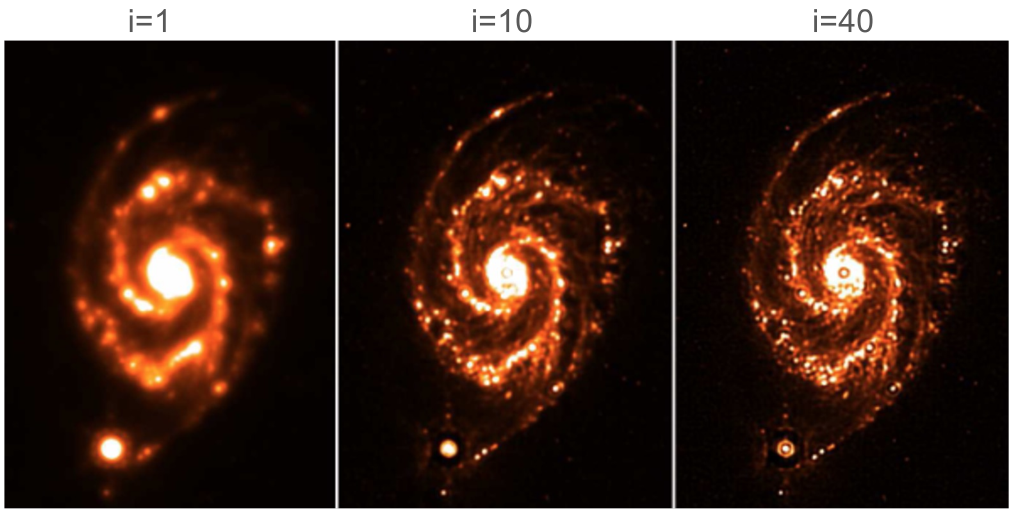


TransiNet, based on convolutional neural networks (CNNs)

Sedaghat & Mahabal (2017)

AWAICGEN Software

Developers (IPAC): F. Masci and J. Fowler



- Image reprojection to common grid
  - Use WCS with distortion (SIP keywords)
  - Common zero point & background matching
  - Optional bad-pixel mask & sigma image
  - Optional area-weighted interpolation
  - Upsampling option
- Capable of image co-addition
  - PRF-weighted averaging or optional inverse-variance weighting
  - Optional Hi-Res method (maximum SNR correlation)
    - Allows for non-isoplanatic PRFs
    - Prior noise-variance weighting
    - Uncertainty estimation
    - Ringing-suppression
    - Iterative to improve resolution (see upper right)
- Products are coadd, coverage, and sigma images
- https://web.ipac.caltech.edu/staff/fmasci/home/icore.pdf
- May use SWarp (Bertin)

Plan for Solar System Objects

- Leveraging IPAC-based connections to NEO Surveyor, RAPID will make use of new software tools for predicting Solar system object positions and brightnesses
- For simulations, software tools will use the MPC orbit catalog to predict asteroid positions based on image headers, and assumed physical properties for fluxes
  - Sources will be injected and Romanized like other transient sources
- For operations, spacecraft position and pointing will be taken from image headers to generate lists of objects near the field of view
  - Association with known objects will be done after image differencing and source extraction prior to issuing an alert
  - MPC catalog will be updated regularly to ensure fresh source lists in the era of rapid growth in known catalog due to Rubin and NEO Surveyor
- Tasks to be addressed:
  - Setup asteroid position prediction software
  - Perform test queries for a Roman-like field of view and sensitivity to determine scope
  - Begin setting up interfaces with Roman simulation data

Additional Simulation Sets Required

Open Universe sims already incredibly useful (thanks!)

However, … further simulations ***vital*** beyond scientific validation:

- Develop/debug pipelines
- Profile realistic performance
- Determine technical budget/tradeoffs (CPU, RAM, network, AWS cost)

Deliverables?

- Schedule
- Contents
- Format
- Volume

Prefer regular iteration over polish!

Community Engagement: External Advisory Board

Members:

- Eric Bellm, UW (chair)
- Kishalay De, MIT
- Ori Fox, STScI
- Anais Moller, Swinburne U
- Armin Rest, STScI
- David Trilling, NAU

Meeting quarterly (started 2024-05-29)

Community Engagement: Time Domain Working Group

Ashish Mahabal from RAPID chairing

Initial charter created

Strategic **T**ime-domain **R**esearch and **I**nfrastructure **D**evelopment for Roman **E**xploration (STRIDE)

For open communications across PITs, SOC, SSC, and other teams

Monthly meetings

First meeting being convened

Objectives:

Facilitate Communication, Technology Development, Streamline Development and Optimize Processes

Focus areas:

Feasibility checks, Algorithm development, Prioritization feedback

Community Engagement: WE NEED YOU!

- Adding transients to existing Roman simulated images
  - Low-lum SNe II-P
  - Low-lum SNe Ib/IIb
  - ILRTs
  - LRNe
  - FBOTs
  - AGN flares and other nuclear transients
  - LBVs/impostors/pre-SN eruptions
  - Novae
  - SLSN-II
  - "normal" IIn
  - Ia-CSM
  - Double-peaked SNe
  - Galactic variables and transients

Community Engagement: Interact with RAPID!

Sign up for announcements
from RAPID

Visit our website for
more info and updates

Send us email:
rapid@ipac.caltech.edu

Join RAPID Slack

We want your input and feedback!

RAPID is hiring!

Currently an open position:

- Staff Scientist

Please forward to any potential candidates!
Further opportunities in FY25