CrossMark

The Evryscope Fast Transient Engine: Real-time Detection for Rapidly Evolving Transients

Hank Corbett[1] , Jonathan Carney[1] , Ramses Gonzalez[1] , Octavi Fors[2] , Nathan Galliher[1] , Amy Glazier[1] ,
Ward S. Howard[3] , Nicholas M. Law[1] , Robert Quimby[4,5] , Jeffrey K. Ratzloff[1] , and Alan Vasquez Soto[1]

[1] University of North Carolina at Chapel Hill, 120 E. Cameron Avenue, Chapel Hill, NC 27514, USA; htc@unc.edu
[2] Dept. de Física Quàntica i Astrofísica, Institut de Ciències del Cosmos (ICCUB), Universitat de Barcelona, IEEC-UB, Martí i Franquès 1, E-08028 Barcelona, Spain
[3] Department of Astrophysical and Planetary Sciences, University of Colorado, 2000 Colorado Avenue, Boulder, CO 80309, USA
[4] San Diego State University, 5500 Campanile Drive, San Diego, CA 92182, USA
[5] Kavli Institute for the Physics and Mathematics of the Universe (WPI), The University of Tokyo Institutes for Advanced Study, The University of Tokyo, Kashiwa, Chiba 277-8583, Japan

Abstract

Astrophysical transients with rapid developments on subhour timescales are intrinsically rare. Due to their short durations, events like stellar superflares, optical flashes from gamma-ray bursts, and shock breakouts from young supernovae are difficult to identify on timescales that enable spectroscopic follow-up. This paper presents the Evryscope Fast Transient Engine (EFTE), a new data reduction pipeline that is designed to provide low-latency transient alerts from the Evryscopes—a north–south pair of ultra-wide-field telescopes with an instantaneous footprint covering 38% of the entire sky—and tools for building long-term light curves from Evryscope data. EFTE leverages the optical stability of the Evryscopes by using a simple direct image subtraction routine that is suited to continuously monitoring the transient sky at a cadence of a minute. Candidates are produced within the base Evryscope 2 minute cadence for 98.5% of images, and internally filtered using VETNET, a convolutional neural network real–bogus classifier. EFTE provides an extensible and robust architecture for transient surveys probing similar timescales, and serves as the software test bed for the real-time analysis pipelines and public data distribution systems for the Argus Array, a next-generation all-sky observatory with a data rate 62 times higher than that of Evryscope.

1. Introduction

Optical transients evolving on short, subhour timescales are difficult to study using the multiwavelength, multi-facility approaches that are typically used for longer-lived transients. For the fastest events, including prompt optical flashes from long gamma-ray bursts (GRBs; Fox et al. 2003; Cucchiara et al. 2011; Martin-Carrillo et al. 2014; Vestrand et al. 2014; Troja et al. 2017), shock breakout in young supernovae (Garnavich et al. 2016; Bersten et al. 2018), and stellar flares (Aizawa et al. 2022; Howard & MacGregor 2022; Pietras et al. 2022), the duration of the event can be $\leqslant 1$ hr, shorter than the base observing cadence of conventional tiling surveys, such as the Zwicky Transient Facility (ZTF; Bellm et al. 2019), PanSTARRS (Kaiser et al. 2010), the Catalina Sky Survey (Larson et al. 2003) and Catalina Real-Time Transient Survey (Drake et al. 2009), SkyMapper (Keller et al. 2007), the Asteroid Terrestrial-impact Last Alert System (ATLAS; Tonry et al. 2018), the All-Sky Automated Survey for Supernovae (ASAS-SN; Shappee et al. 2014), the Dark Energy Survey (DES; Dark Energy Survey Collaboration et al. 2016), the Gravitational Wave Optical Transient Observatory (GOTO; Dyer et al. 2018), and the Mobile Astronomical System of

Telescope-Robots (Lipunov et al. 2004). These surveys tile the sky on timescales of days to maximize their likelihood of detecting supernova-like transients, which evolve over the course of days and months.

Faster events, occurring on minute to hour timescales, are detected in conventional tiling surveys, but with frequently undersampled light curves. Tiling surveys are also not typically optimized for minute-scale latency between detection and reporting, precluding spectroscopic follow-up on timescales comparable to the lifetime of the transient. As a result, searches for short-lived events typically require simultaneous coordinated observations of small sky regions, as in the Deeper-Wider-Faster program (DWF; Andreoni et al. 2020). However, previous searches for fast transients in this regime, by the DWF team (Andreoni et al. 2020), as well as from PanSTARRS (Berger et al. 2013), the Intermediate Palomar Transform Factory (iPTF; Ho et al. 2018), the Palomar Transient Factory (PTF; van Roestel et al. 2019), Tomo-e-Gozen (Richmond et al. 2020), and the Organized Autotelescopes for Serendipitous Event Survey (Arimatsu et al. 2021), have only produced upper limits on the extragalactic event rate of fast transients, suggesting that increased areal survey rates are necessary to observe any new populations of high-speed transients.

An alternative approach to probing the dynamic sky at short timescales is to survey an extreme field of view, typically sacrificing some depth and resolution relative to conventional tiling surveys, in exchange for rapid-cadence monitoring. This

approach enables time-resolved detections of fast optical transients, and poses unique challenges and opportunities for real-time data reduction. Rapid localizations of transients across wide fields of view have recently been used to make spectroscopic observations of flares using the Ground-based Wide-Angle Camera system, with latencies as low as 20 minutes (Wang et al. 2021; Xin et al. 2021).

Galactic transients at minute to hour timescales are plentiful, with stellar flares from M-dwarf stars making up the majority of these events (Kulkarni & Rau 2006). Flares are caused by reconnections in the stellar magnetic field, producing radiation across the electromagnetic spectrum on timescales ranging from seconds to hours. The radiation from the largest events, the so-called "superflares," can reach energies $\geqslant 10^{33}$ erg— orders of magnitude greater than that of the largest solar flares (Schaefer et al. 2000). Flares are responsible for much of the UV environment of rocky planets orbiting cool stars (Walkowicz et al. 2008), potentially providing the bioactive UV flux that is necessary for prebiotic chemistry (Ranjan et al. 2017), or even eroding Earth-like atmospheres (Segura et al. 2010; Howard et al. 2018). Spectroscopic observations taken during the initial stages of a flare can reveal temperature and emission line evolutions during their most impulsive phases, which are valuable for constraining fundamental flare physics, as well as the potential impacts of flare activity on exoplanet atmospheres. Such observations are crucial for assessing the habitability of Earth's closest neighbors: the nearest exoplanet to Earth, Proxima Centauri b, is subject to frequent high-energy flare activity (Howard et al. 2018).

In addition to time resolution constraints on survey design, searches for subhour transients like flares require software data pipelines that are capable of rapidly identifying candidates for spectroscopic follow-up and classification. Examples of bespoke pipelines optimized for minimal latency are presented in Perrett et al. (2010), Kumar et al. (2015), Cao et al. (2016), Förster et al. (2016), and Andreoni et al. (2017). Such pipelines are often built around difference image analysis, a method for isolating sources with variable flux by subtracting an earlier reference image of the field, which is complicated by the need to match the seeing-limited point-spread functions (PSFs) of images from multiple epochs. Methods for subtracting images in the presence of variable PSFs include deconvolution with a matching kernel (Alard & Lupton 1998; Bramich 2008; Becker 2015), which can be computationally expensive and numerically unstable, and, more recently, the statistically optimal Zackay, Ofek, and Gal-Yam (ZOGY) method (Zackay et al. 2016), which requires a robust and static model of the image PSF, and the Saccadic Fast Fourier Transform method (Hu et al. 2022).

In this paper, we present the Evryscope Fast Transient Engine (EFTE), a real-time discovery pipeline for the Evryscopes. The Evryscopes are a pair of gigapixel-scale survey instruments that continuously image 38% of the celestial sphere at a 2 minute cadence. EFTE is optimized for sensitivity to short-duration transients, including stellar flares and flash-like optical counterparts to multimessenger or multiwavelength events. Using EFTE, we are able to produce transient candidates within the survey's cadence, with actionable alerts indexed into our database before the next image in the sequence. EFTE also provides a processing workflow for the batch processing of Evryscope image data, and forms the basis of the Evryscope precision photometry pipeline, providing years-long light curves for every star brighter than $g \sim 15$.

One goal for the EFTE pipeline is to minimize the computational resources that are necessary for data analysis; a single colocated compute node supports each Evryscope site. Low resource requirements are particularly necessary when looking toward next-generation sky surveys, such as the Argus Optical Array (Law et al. 2021, 2022). The upcoming Argus Array Pathfinder instrument, consisting of 38 telescopes with 20 cm apertures, will produce up to 180 TiB of data per night at 1 s cadence and 6 TiB of data per night at the base 30 s cadence; maximizing the science returns from data-intensive systems like Argus will require time- and cost-efficient algorithms and pipelines. For Argus, all images must be reduced within the observing cadence, to provide sufficiently low latency for follow-up and to avoid a backlog of data, which can require runaway compute resources for "catch-up." Incoming Argus images are resampled to a predefined HEALPix (Górski & Hivon 2011) grid using a custom GPU-based code. By parallelizing direct subtraction based on the EFTE algorithm, the Argus pipelines are able to reduce each image into transient candidates and compressed images in an average of of 925 ms. Corbett et al. (2022) present a full description of the Argus Array pipelines and data reduction strategy.

The paper is organized as follows. In Section 2, we give an overview of the Evryscope instruments and survey strategy. In Section 3, we describe the EFTE pipeline and present algorithms for data analysis and transient discovery in ultra-wide-field systems, including a simple image subtraction method that is suitable for time-sensitive searches (see Figure 1). In Section 4, we describe the selection metrics and machine-learning (ML) approaches that are used to select candidates from the event stream. In Section 5, we characterize the photometric, astrometric, and latency performance of the pipeline, including the expected survey completeness and characterization of the convolutional neural network (CNN) that is used for vetting candidates. In Section 6, we summarize the early science returns from the EFTE pipeline, including a characterization of the impact of satellite glints on rapid response transient surveys and rapid response observations of stellar flares using the Goodman High Throughput Spectrograph on the Southern Astrophysical Research (SOAR) 4.1 m telescope (Clemens et al. 2004). We summarize, consider the extensibility of the EFTE pipeline to data from other surveys, and describe the next steps toward producing a public event stream in Section 7.

2. Evryscope Survey Overview

2.1. Instrument Description

The Evryscopes are a pair of multiplexed wide-field survey telescopes that are located at Cerro Tololo Inter-American Observatory (CTIO) in Chile and Mount Laguna Observatory (MLO) in California. Each site consists of up to 27 camera units with 6.1 cm apertures, arranged to simultaneously observe the majority of the sky above an airmass of ~ 2. Collectively, the Evryscopes have an instantaneous field of view of 16,512 square degrees (15,929 square degrees, accounting for overlaps between adjacent cameras), with a resolution of $13\rlap{.}''2$ per pixel across a 1.24 gigapixel combined image plane. The telescopes observe at a constant 2 minute cadence, with a 97% duty cycle, and collect an average of 600 GiB of data per night. While the primary Evryscope survey has been conducted in the Sloan g
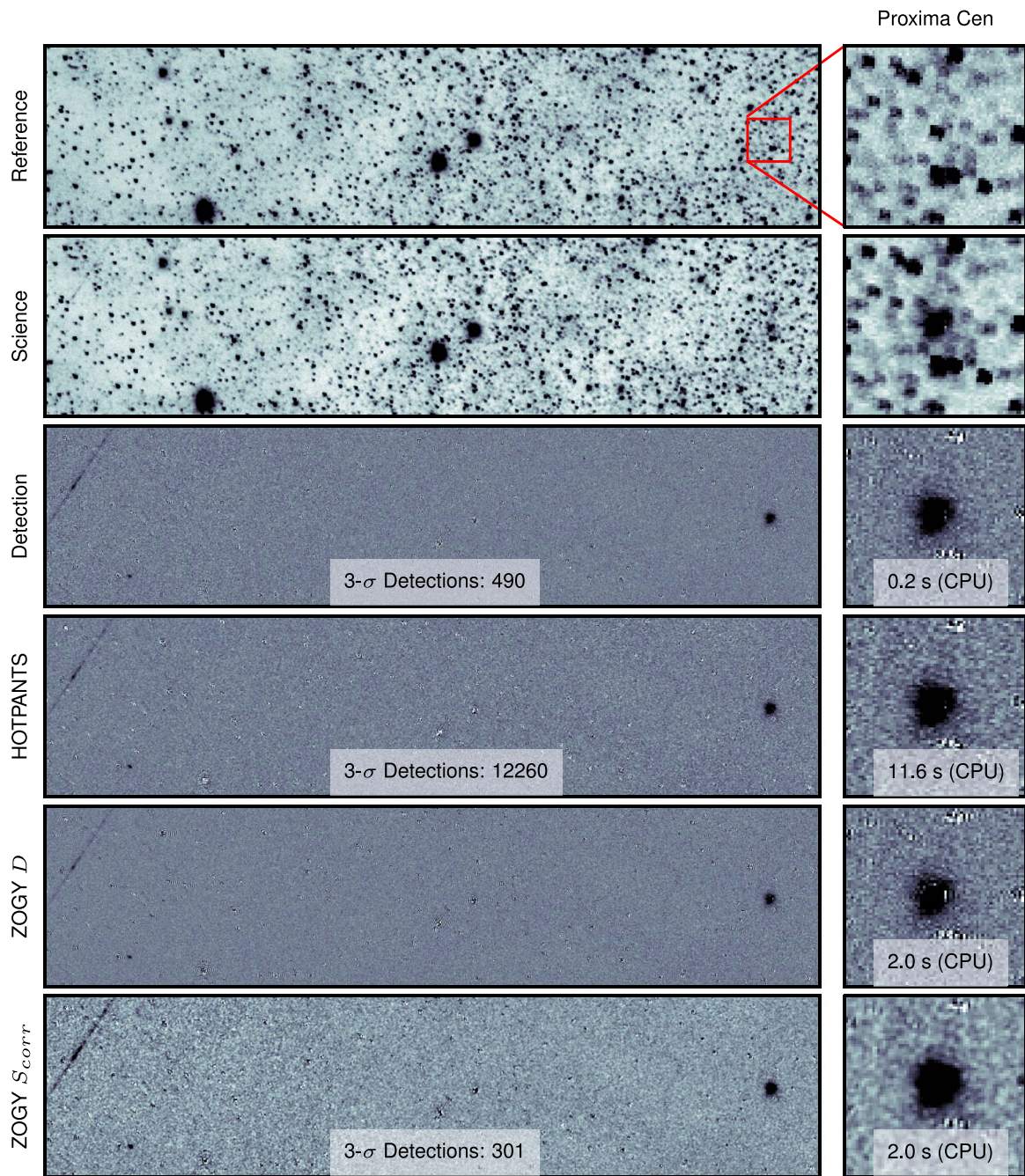
Figure 1. Left: a $4° \times 1°$ region near Proxima Centauri from reference, science, and subtraction images from EFTE, the HOTPANTS algorithm (Becker 2015), and the ZOGY algorithm (Zackay et al. 2016). For ZOGY, we include both the scaled S_{corr} image used for point-source detection and the proper difference image D. In the subtraction images, a faint satellite streak (left), a variable star (bottom left), and an M-dwarf superflare (far right; Howard et al. 2018) are successfully recovered. The EFTE direct subtraction produces 47% more 3σ detections than ZOGY for this field, and it can be computed 10 times faster for this image size. Right: cutouts from the images on the left showing the $1' \times 1'$ region centered on Proxima Cen.

band, the Northern Hemisphere system is also equipped with a Sloan r filter, for use in future surveys. All the telescopes at each site are attached on a single mount, tracking the sky in 2 hr increments.

The instruments are fully robotic, operating autonomously, based on a local weather station. Evryscope-South has been in operation since 2015 May, while Evryscope-North began science operations in 2019 January. For full descriptions of the instrument and the Evryscope science programs, see Ratzloff et al. (2019) and Law et al. (2015). The instrument parameters are summarized in Table 1.

2.2. Evryscope Observation Strategy

The Evryscopes utilize two distinct strategies for determining the 2 hr observing fields that are observed over the course of the night:

1. Semirandom, with the easternmost edge of the field placed 30° above the horizon to the east at the start of each 2 hr observation.
2. Fixed pointings, chosen from 48 overlapping regions, separated by 7°.5 in R.A.

In both scenarios, the duration of a single pointing is limited by the time that it takes for the westward edge of the field to pass

Table 1
System Properties of the Evryscopes

Property	Evryscope-South	Evryscope-North
Field of View (Deduplicated)	8520 sq. deg	7409 sq. deg
Field of View (Total)	8832 sq. deg	7680 sq. deg
Detector Size	662.4 MPix	576 MPix
Cadence	2 minutes	
Aperture	6.1 cm	
Pixel Scale	$13\rlap{.}''2$ pixel^{-1}	
Data Rate	165 Mbps (1.2 GiB minute^{-1})	

Note. For further information, see Ratzloff et al. (2019).

beneath an airmass of ~ 2. This timescale (a "ratchet") is typically on the order of 2 hr. Each Evryscope tracks continuously at the sidereal rate. Minimal drift (of a few arcminutes) due to polar alignment is present over the course of a ratchet, but the visible field between consecutive 2 minute exposures is consistent, point-like, and unstreaked.

Semirandom pointings (currently used for Evryscope-South) are preferred for long-term photometric performance, as diverse field positions allow sensor plane effects to average out over the duration of the survey. Because of the commercial off-the-shelf optics used, individual cameras exhibit up to 50% vignetting at the edge of the sensor's field of view. Randomized pointings also minimize the effects of camera-to-camera periodic noise, provide some resilience against CCD sensor defects, and limit the prevalence of pathological coordinates that are always located at the edge of a sensor and are thus unduly affected by optical vignetting. The trade-off is that individual fields repeat only on timescales of months (cross camera) or years (single camera), and only to a precision of a few degrees.

Fixed fields, by contrast, are used for Evryscope-North, resulting in fields that repeat with arcminute precision on 1–3 day timescales. This repeatability is convenient for transient searches, as it allows us to build up an archive of reference frames to use for image subtraction. For fields above an airmass of 2, 76% are observed within 2 days of the previous visit, while 97% are observed within a week. Because adjacent fields overlap by $\sim 95\%$, a given sky region will appear in many different pointings, meaning that the field recurrence time is independent of the observing cadence.

3. EFTE

EFTE is a pipeline for searching Evryscope images for bright and rapidly changing sources in real time, identifying and recording candidates across the full Evryscope field of view within the 2 minute observing cadence. The primary goal of EFTE is to provide a reliable event stream with sufficiently minimal latency to enable multiwavelength follow-ups of events with subhour durations. EFTE also provides useful general purpose utilities for interacting with and analyzing Evryscope data, including a quick-look photometry pipeline that is independent of the general purpose precision photometry pipeline, a custom astrometric solver, and CCD calibration functions. EFTE is mostly written in Python, with some compute-intensive routines (stamp extraction and photometry) being implemented in C and wrapped as Python extensions using Cython (Behnel et al. 2011).

EFTE is a hierarchical + distributed system, with two analysis servers on site at MLO and CTIO, which stream reduced data products back to a central PostgreSQL[6] relational database on campus at the University of North Carolina at Chapel Hill (UNC-CH). The analysis servers each have dual AMD EPYC processors (36 CPU cores for Evryscope-North, 48 for Evryscope-South), and 512 GiB of RAM at Evryscope-South and 384 GiB at Evryscope-North. The asymmetry between the two sites is due to the additional 4 yr of archival data from Evryscope-South. The central database for reduced data products is hosted on a 36 core server, with 24 TiB of flash storage, which is located on campus at UNC-CH. This server also hosts a backend application for pipeline monitoring, associating EFTE transients with external alerts, and end-user reporting via a Slack-based[7] web interface.

In real-time operations, EFTE instances on each analysis server communicate with the Evryscope data acquisition system via a transmission control protocol (TCP) socket connection, receiving notifications for each incoming image once it has been written to a shared network file system. EFTE maintains an in-memory, per-ratchet database of recent images to be matched for image subtraction, spawning subprocesses for all analysis tasks. Figure 2 shows the primary components of the EFTE pipeline, from the moment that an image is written to disk to the reporting of candidates.

3.1. Image Quality Monitor and CCD Calibration

Once an exposure has been completed, the Evryscope observation daemon sends a TCP packet to an EFTE instance running on the analysis server. Upon receipt of this notification, EFTE will asynchronously record basic metadata—including the camera, timing, origin, and instrument configuration—to the central database that is located at UNC-CH. Before further reduction, the image goes through a series of general quality assurance steps, including:

1. verification of the file against the checksum recorded by the acquisition system;
2. instrument configuration checks for camera cooling, dome status, and exposure type;
3. autocorrelation-based checks for tracking errors and alignment failures; and
4. sky background measurement for saturation and linearity checks.

If these conditions are satisfied (as they are for 98% of images), the image is converted from ADU to electron units, and matched to dark and flat fields for CCD calibration.

Dark frames are regularly regenerated using frames taken at the beginning and ending of each night. Cameras are cooled to a constant -20°C during observing, but a drift in the bias level, of a few percent, is observed as a function of the camera's external temperature. We believe that this is caused by temperature gradients across the readout electronics, so make a quadratic correction to the bias level as a function of the camera electronics temperature, as measured by the onboard sensors in each camera. Additionally, a small ($<1\%$) linearity correction is applied per pixel, based on a cubic fit to pixel values versus the exposure time in lab testing. The linearity

---

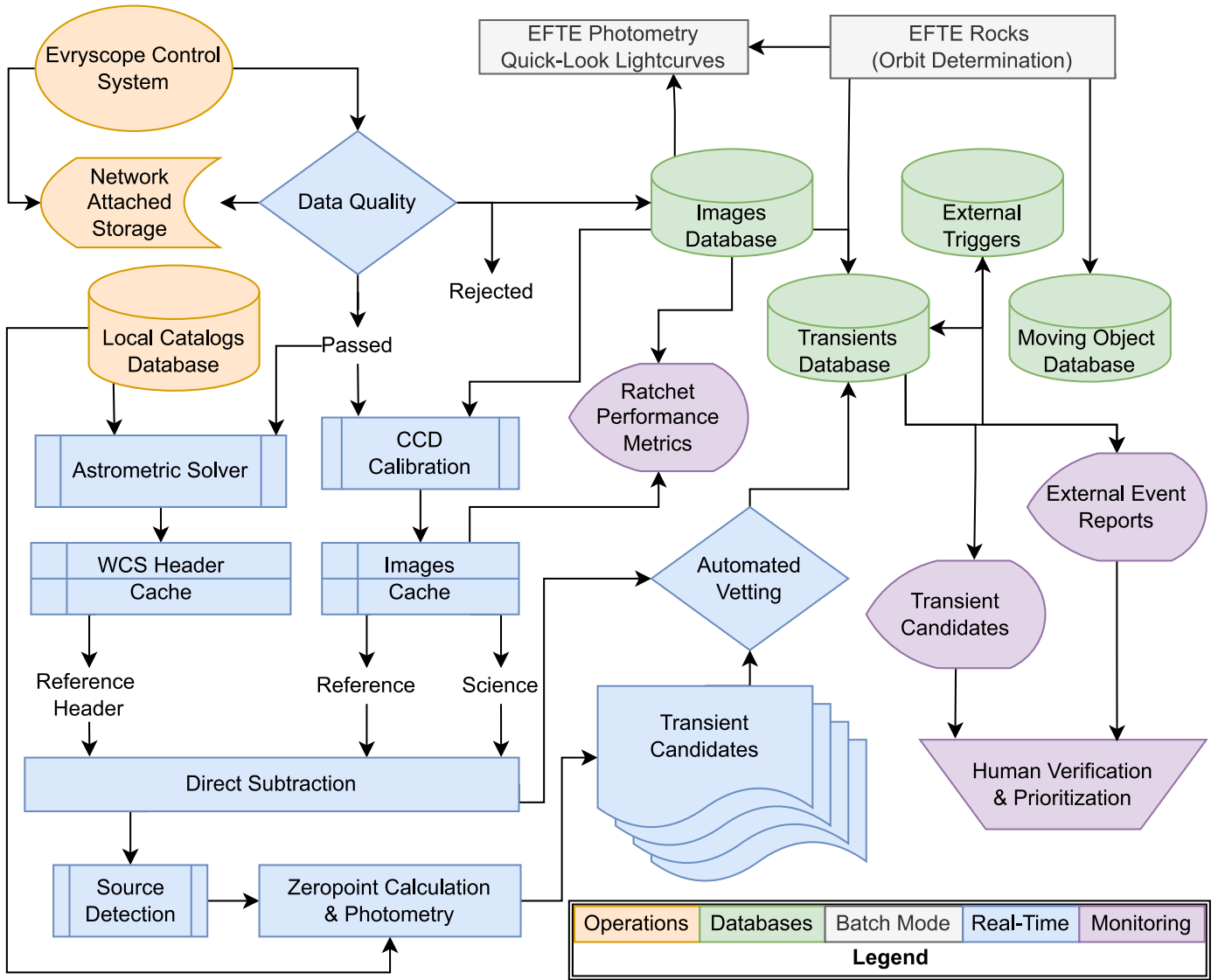6 http://www.postgresql.org/
7 https://slack.com/

Figure 2. Data flow and layout of the EFTE pipelines. The operations (orange) and real-time reduction (blue) components are independent for each observatory, while the pipeline monitoring (purple) and shared databases (green) collate data from both Evryscope-North and Evryscope-South.

corrections were determined to be near-identical for all our sensors.

Because of the extreme single-camera field of view, twilight flats contain significant sky gradients that the Evryscope is unable to compensate for through diverse pointings, because of its fixed camera positions. Instead, we use photometric flats, which are calculated based on a seventh-order polynomial fit to the normalized flux offsets of reference stars relative to the g-band catalog photometry from the ATLAS All-Sky Stellar Reference Catalog (ATLAS-REFCAT2; Tonry et al. 2018). These frames capture the average vignetting patterns of the individual cameras, which can change sharply at the edge of the field. The photometric flat fields are stable at the 1% level on months-long timescales, due to the focus stability of the Evryscope Robotilter alignment system (Ratzloff et al. 2020), and are regenerated only when the instruments are cleaned, which typically requires the replacing or removing of the outer optical windows on each camera.

Bad pixels are replaced with the median of the surrounding 3×3 pixel block, then assigned an arbitrarily high uncertainty in the resulting noise image that is used for photometry and source detection. Parts of each camera's field of view, particularly those near the center of the frame, will have undersampled PSFs. Simple bad pixel masking (i.e., assigning such pixels a `NaN` or 0 value) will produce sharp artifacts in the subsequent analysis, requiring pixel resampling, like the image subtraction described in Section 3.3.

3.2. Astrometric Solutions

In parallel with the science frame calibration steps, the EFTE pipeline produces an astrometric solution for the image, using a custom solver developed for the highly distorted Evryscope focal plane. The Evryscope astrometric solutions begin with an initial solve, based on the center 512×512 pixel region, using a local install of `astrometry.net` (Lang et al. 2010). This solution is only used to locate the center of the image. The sources in the image are then crossmatched against the Tycho-2 catalog (Høg et al. 2000), and the offsets are used to optimize a polynomial distortion solution to fifth order in each of the x, y, and radial positions on the sensor, plus cross terms. The solution is then verified against a subset of bright stars from

Gaia Data Release 2 (DR2; Gaia Collaboration et al. 2018), based on the crossmatch performance against detections in an even grid of 15 different sensor regions, using the following requirements:

1. >80% recovery in at least seven regions;
2. < 50% recovery in no regions; and
3. Uncertain recovery (due to source confusion or non-detections) in no more than two regions.

We selected Gaia DR2 for solution verification due to its reference epoch (J2015.5) coinciding with the beginning of the Evryscope observations. The typical rms offsets from the Gaia DR2 positions are $\sim 4''$, or 0.3 pixels.

The complete solution is written into a world coordinate system (WCS) header using the TPV convention for distortion polynomials (Calabretta et al. 2004). The TPV representation is an extension of the standard tangential projection, including additional terms for a general polynomial correction.[8] Due to atmospheric refraction and tracking errors, the solution must be recalculated for each image, but the solver is able to start with a precomputed baseline distortion solution, averaged over dozens of fields for each camera. We found that starting with an averaged solution decreased the time required for the final optimization by a factor of several, on average.

This header is archived to network storage, and serialized and passed back into the in-memory EFTE matching database, where it is associated with the camera and active field. Finally, a footprint of the image, discretized as a GeoJSON (Butler et al. 2016) polygon, is stored in the central database, where it is indexed using the PostGIS[9] extension of PostgreSQL, which provides a variety of spatial object types. These footprints support a variety of use cases, and allow users to easily query for images containing a given target, or to search for images intersecting with arbitrary sky regions that can be represented as polygons, such as probability skymaps for gravitational-wave and GRB triggers.

3.3. Direct Image Subtraction

Like most optical transient surveys, EFTE isolates objects with changing flux by subtracting each science image from an earlier reference frame of the same field from each image. We optimize our subtraction algorithm for speed rather than statistical optimality, electing per-pixel operations requiring no additional intermediate data products beyond those produced in the initial photometric pipeline. Due to the short (subhour) timescales of interest, and the dominance of instrumental optical effects on the system PSF, the Evryscope images do not require PSF-matching techniques, which are addressed by standard difference image analysis routines, like the High Order Transform of PSF and Template Subtraction Algorithm (HOTPANTS; Becker 2015) and ZOGY (Zackay et al. 2016). However, EFTE was built for extensibility, and implementations of both HOTPANTS and ZOGY are included in EFTE.

The resolution of the Evryscope images is limited by optical distortions from the camera lenses and pixel scale, rather than by atmospheric effects, under most practical observing conditions. The PSFs vary greatly across the image plane of each camera, as illustrated in Figure 3; however, the Evryscope image quality metrics have been measured as being stable at the few percent

level on timescales of many months (Ratzloff et al. 2020), creating highly repeatable PSFs for each individual camera.

As a result, we adopt a straightforward algorithm for image subtraction, in which the reference and science images are aligned, matched in flux, and subtracted directly. The difference between the two images is then weighted by a propagated uncertainty image to identify significant changes in flux. This approach is valid only if the following conditions are satisfied for the reference and science image couplet:

1. the observed PSFs are dominated by telescope optics and pixel scale, and do not vary significantly as a function of observing conditions on the timescale of the lag between the images;
2. all sources have near-identical pixel coordinates in both images, offset by no more than the PSF coherence scale, which we define as the pixel distance over which the spatial PSF variation is less than the combined thermal and atmospheric effects over a baseline of a few minutes, or a 1% maximum change in the normalized PSF—this scale is typically $\sim 10'$, or ~ 50 pixels at the Evryscope pixel scale; and
3. the global flux scaling between the two images is smooth.

In the following subsections, we describe the process by which we match image couplets for subtraction, as well as the custom method that we use to subtract the images, which is optimized for the unique resolution and time domain covered by EFTE.

3.3.1. Reference and Science Image Selection

The primary science targets for the EFTE survey are stellar flares, which have characteristic optical rise times of minutes. As a result, there is minimal benefit to producing reference frames that are widely spaced in time from our science images, to maximize sensitivity to slowly varying objects. Instead, the image-matching daemon uses a sliding reference frame, taken from the same pointing as the science image. The Evryscopes maintain a consistent pointing over the course of a ratchet, with drift of only a few arcminutes even at the equator, meaning that the PSF for a given star is essentially constant during each 2 hr tracking period, up to resampling effects being caused by its subpixel position in the $13''2$ pixels. Additionally, using a reference image from the same pointing means that the science and reference frames are taken under near-identical sky conditions, minimizing the amount of flux scaling necessary.

In the most aggressive case, we could simply subtract consecutive images to achieve near-ideal consistency between the new and reference frames. However, the immediate reuse of science images as reference images limits the survey's sensitivity only to transients with detectable changes over the 2 minute image interval, making confirmation images of highly impulsive events unlikely. In practice, we enforce a short lag, Δt_D, between the reference and science images. Δt_D is typically chosen to be 10 minutes. Over 10 minutes, the field drift due to polar alignment error is consistently less than $10'$ (50 pixels). The PSF variability over 50 pixels typically produces subpercent subtraction artifacts, and image registration between the two images can be done with simple transformations, with a minimal loss of astrometric precision (see Section 5.2).

Image reuse effects are also evident at $\Delta t_D = 10$ minutes, but only after the potential fourth confirmation image. Figure 4 shows this image reuse effect on a real flare that is seen on-sky,
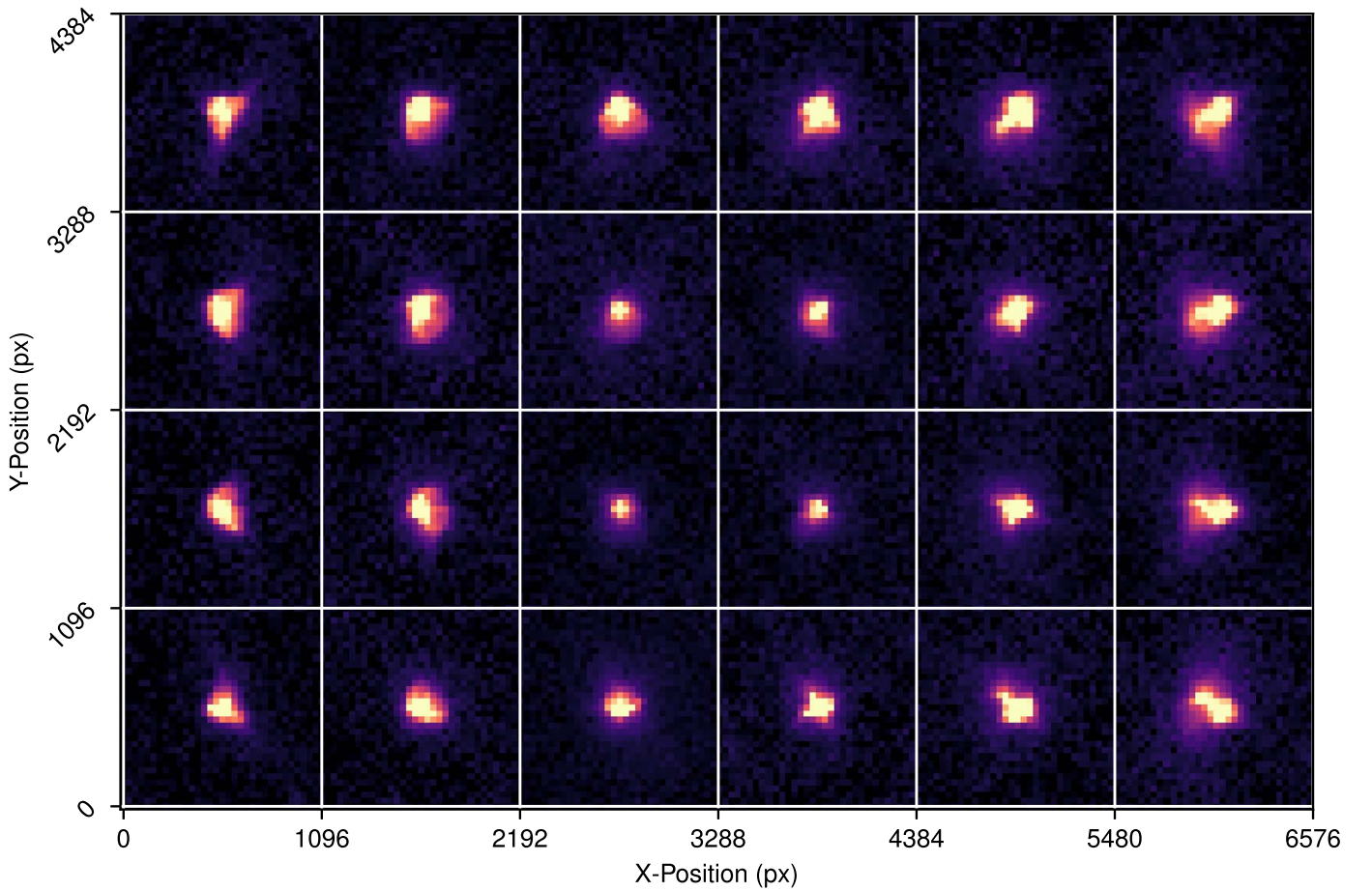

Figure 3. Median PSFs across a 6 × 4 grid of sensor regions. PSF variability as a function of chip position is evident; however, long-term measurements of the Evryscope optical stability (Ratzloff et al. 2020) indicate that the PSFs are repeatable over time, despite aberrations.

in which the same image is both the first science image and the last reference image for the transient. While the amplitude and rise time of this event enabled multiple detections up to 8 minutes after the initial detection, events with shorter rise times or lower amplitudes may only be detected in a single epoch.

Additionally, the sliding reference frame causes the photometry in the unscaled difference image (i.e., the numerator of Equation (2)) to be relative in time. The light curves of EFTE candidates are computed using forced aperture photometry in the science images, as described in Section 3.4.

Over the course of a ratchet, beginning with images A, B, C, D, E, and F, the pipeline will perform the following subtractions: B–A, C–A, D–A, E–A, and F–B. The rise time sensitivity of the pipeline increases as a function of the time delay Δt_D between the science image and the previous image from the same pointing chosen as a reference image. Δt_D is in general a tunable parameter of the pipeline, which could be increased to trade the viability of the assumptions enumerated above (thus resulting in higher false-positive/false-negative alert rates) for increased sensitivity to slower rise times. Given sufficient computing, multiple instances of EFTE can run in parallel, enabling sensitivity to different science targets.

3.3.2. Image Registration

Because of the sliding reference frame selection, the drift between the science and reference images amounts to a maximum of a few pixels during real-time operations, which must be corrected. Additionally, small offsets can have significant effects on the sampled PSF. As such, the images must be carefully aligned and resampled, so that they match in both position and PSF.

For the first image in a ratchet, EFTE must wait for an astrometric solution. However, the astrometric solutions for the subsequent images from each camera can be inferred by alignment to the first image. The effects of image registration on astrometry performance are addressed in Section 5.2. Bootstrapping the astrometric solution in this way reduces the delays in the real-time subtraction process due to the astrometric solver to once per ratchet, on the first image. For image alignment and resampling, we use WCS-independent asterism matching, utilizing the Python AstroAlign[10] package (Beroiz et al. 2020) to calculate a rigid transformation between the two images and perform quadratic resampling.

Alternatively, in cases where a full WCS solution is available for both images (e.g., in batch reductions that are not conducted in real time), the reference and science images can be aligned by resampling the images to a common grid using their WCS solutions, by utilizing the Astropy-affiliated package Reproject. This has the advantage of allowing for nonrigid transformations and accounting for the effects of varying per-pixel sky areas across the sensor plane. While astrometric warping due to atmospheric diffraction is negligible for the typical Δt_D values that are used for real-time reduction,

---

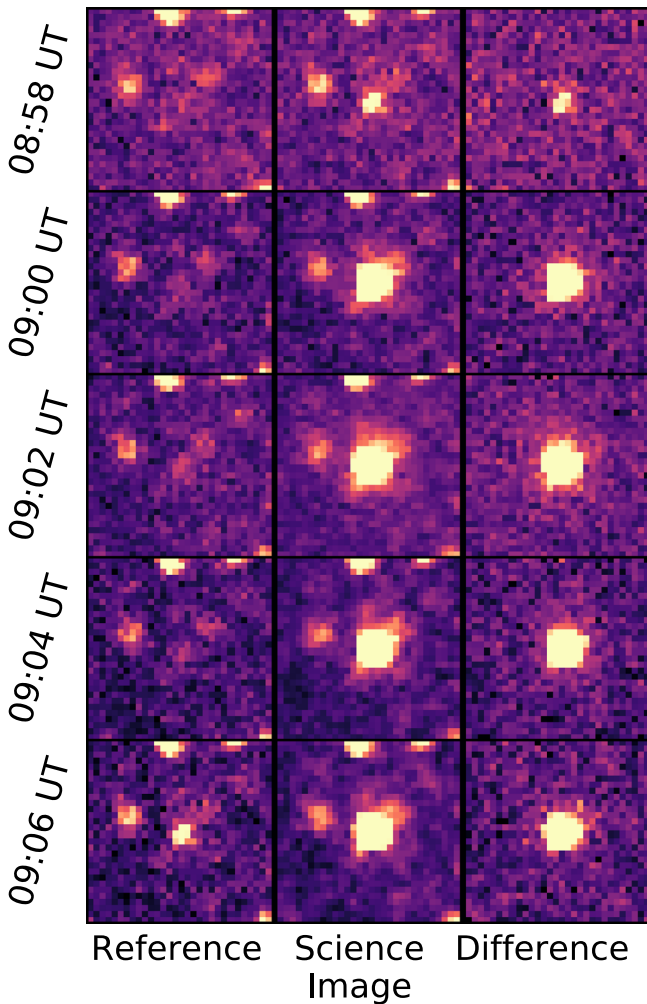[10] https://github.com/toros-astro/astroalign

Figure 4. EVRT-192099, a 5.5 mag flare from a star associated with 1RXS J174441.6-531551 in the ROSAT All-Sky Survey Bright Source Catalogue (Voges et al. 1999; Fresneau & Osborn 2009). The reference frame from 2019 October 4 at 9:06 UT is the science image from 8:58 UT, showing the sliding reference frame used by our direct subtraction algorithm.

WCS-based resampling is necessary for longer baselines and internight comparisons for fixed fields.

3.3.3. Flux Scaling

Despite the minimal baseline between the reference and science images, we fit a multiplicative flux scaling factor to the reference image, to remove any discrepancies with respect to the science image, due to variations in transparency and sky brightness, which is particularly important for observations during twilight conditions. Because each individual Evryscope camera covers a large sky area, we allow the flux scaling factor to vary across the image, based on the results of the forced aperture photometry.

First, we divide each image into 24×16, square regions of that are 274 pixels across, and select several thousand bright stars from ATLAS-REFCAT2. We then calculate the sigma-clipped mean flux ratio between the science and reference images for stars in each of the 384 image sectors, and interpolate this back to full resolution using cubic splines. Finally, the flux-matched reference image $R_m(x, y)$ is calculated

from the original calibrated reference image $R(x, y)$ and the spatially varying flux ratio $F(x, y)$, as

$$R_m(x, y) = R(x, y) * F(x, y). \tag{1}$$

To calculate the uncertainty in the flux ratio, we calculate the standard deviations of the flux ratios for the reference stars in each region, then interpolate across the full field of the image using cubic splines. The flux ratio uncertainty is propagated forward into the noise characterization for the reference image. Figure 5 shows a typical low-resolution flux ratio map for a pair of images with $\Delta t_D = 10$ minutes, showing an average 1.7% change in transparency between the images, with some internal structure. The magnitude of the scaling is consistent with transparency changes due to airmass for a camera placed at the edge of the array. Small-scale structure, when present, tends to move smoothly between images, and is likely caused by high clouds.

Depending on the science program, the flux scaling can be skipped during reduction, to minimize latency. For consecutive image subtraction ($\Delta t_D = 2$ minutes), we neglect the flux scaling effects, as the uncertainties in the flux scaling dominate the final noise budget for the image, and the flux scaling is typically subpercent under normal observing conditions. The primary driver of these uncertainties is likely the subpixel response function, which is highly local on the Evryscope image sensors, causing the effect to not average out beyond the 1%–3% level when interpolating across the image plane. Instead, multiple slightly offset measurements of the sample star must be modeled simultaneously, as they are for the precision photometry pipeline and in the coaddition of multiple images of the same field. However, we include the flux ratio for longer-baseline subtractions, where background variations can dominate over systematics.

3.3.4. Error Analysis and the Detection Image

To identify significant changes in the difference image, we need robust accounting of the noise sources in each image. For each science and reference image, we model a spatially varying background, based on a sigma-clipped and interpolated mesh, using sep, a Python implementation of the core routines from SExtractor (Bertin & Arnouts 1996; Barbary 2016). The standard deviation of the background s_B is also measured in this step, based on the sigma-clipped standard deviation. s_B is treated as an empirical measure of the Gaussian noise contributions to each image, including the readout noise and dark current uncertainty. We note that this approach can overestimate the noise, due to Poisson contributions from unresolved background starlight, and, as a result, the detection significance in the direct subtraction image will tend to be an underestimate, particularly in crowded fields (e.g., near the galactic plane).

For a given combination of a science image $S(x, y)$ and a flux-matched reference image $R_m(x, y)$, both in electron units, the detection image $D(x, y)$ is defined as

$$D(x, y) = \frac{S(x, y) - R_m(x, y)}{\sqrt{s_S^2(x, y) + s_R^2(x, y)}}, \tag{2}$$
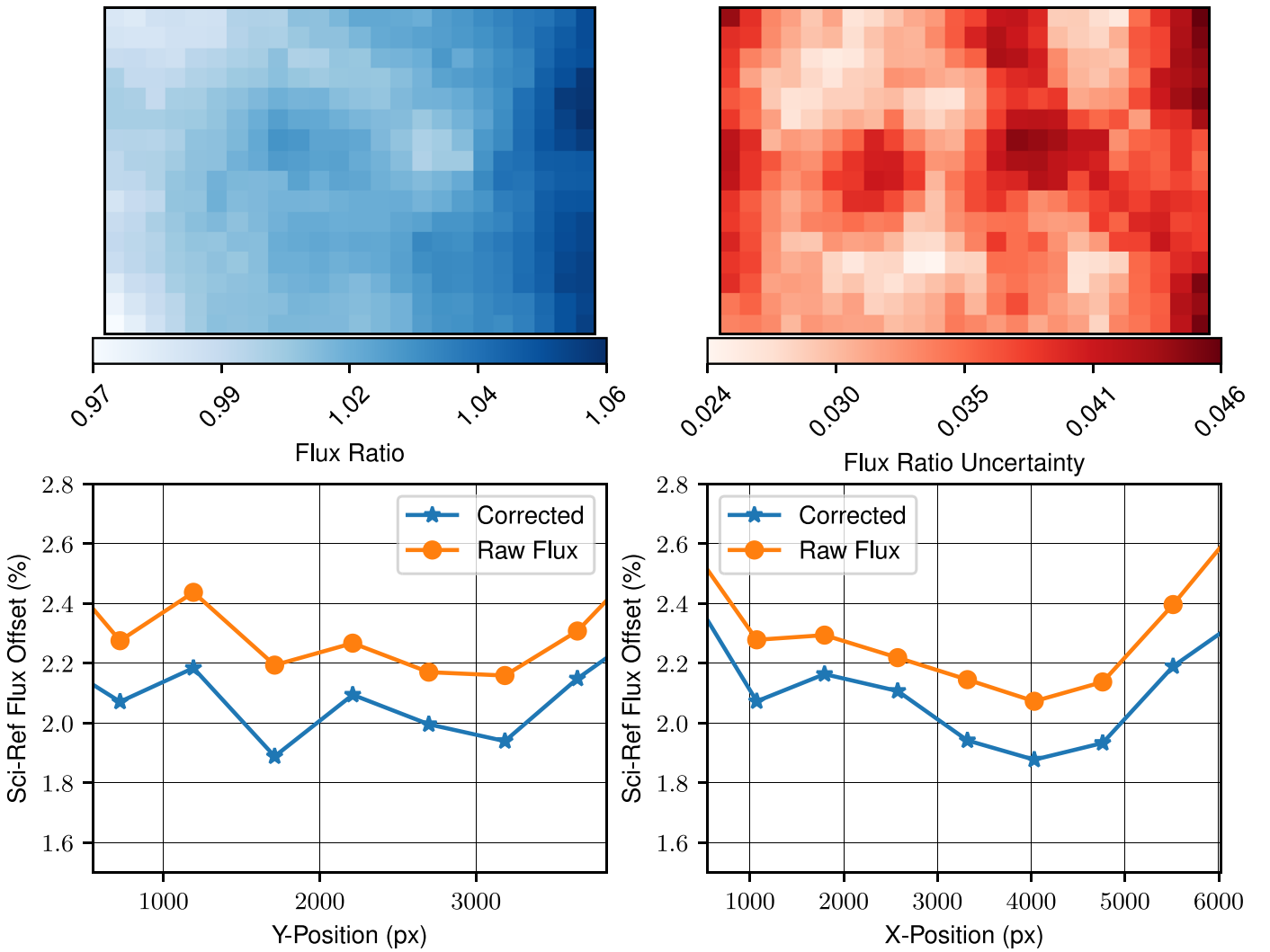
Figure 5. Top left: map of the flux ratio between a science and reference image couplet. Top right: the corresponding flux ratio error map. The flux ratio is measured based on forced aperture photometry of several thousand reference stars, and it is interpolated across each single-camera full-resolution image based on the observed flux ratio, in an equally spaced grid of 384 image sectors. Bottom: relative aperture flux residuals before and after correcting with the interpolated flux ratio, as a function of the x- and y-positions on the sensor. The correction improves the flux match between the images by <1%, which is not significant for short ($\Delta t_D = 2$ minute) subtraction baselines. The increases in the flux offset at the edges of the image are caused by aperture losses due to the variable Evryscope PSF.

where s_S^2 and s_R^2 are the total noise images for S and R_m, given by

$$s_S(x, y) = \sqrt{S(x, y) + s_{\mathrm{BS}}^2(x, y)} \qquad (3)$$

and

$$s_R(x, y) = R(x, y)\sqrt{\frac{R(x, y) + s_{\mathrm{BR}}^2(x, y)}{R^2(x, y)} + \frac{s_F^2(x, y)}{F^2(x, y)}}, \quad (4)$$

where $s_{\mathrm{BS}}(x, y)$ and $s_{\mathrm{BR}}(x, y)$ are the measured background standard deviation maps of the science and reference images, respectively, $F(x, y)$ is the flux ratio between the two images, and $s_F(x, y)$ is the spatially varying uncertainty in the flux ratio. The $D(x, y)$ image is the simple difference between the two images, scaled by the combined per-pixel uncertainty. The detection image has the unit of standard deviation. We again use sep to mask the detection image at the desired threshold and identify sources.

Figure 1 shows an example of an image couplet in a crowded field, along with the resulting direct subtraction image. We also include subtraction images produced using the HOTPANTS and ZOGY algorithms (Becker 2015; Zackay et al. 2016). Direct subtraction produces more false positives (488) than ZOGY (299 in S_{corr}) at the 3σ threshold, but two orders of magnitude fewer than HOTPANTS (12,258). While ZOGY is a computationally efficient approach, direct subtraction is faster by a factor of 10, largely due to the requirement for a PSF model to be calculated for ZOGY. Using the direct detection image, we successfully identify three astronomical transients with the EFTE pipeline, following the vetting procedures described in Section 4.

3.4. Photometric Zeropoints and Forced Photometry

To calibrate the magnitudes of the EFTE transient candidates to a standard photometric system, we build a spatially varying photometric zeropoint, based on a subset of ATLAS-REF-CAT2, a composite catalog consisting of *griz* data from the AAVSO Photometric All-Sky Survey (Henden et al. 2016),

PanSTARRS Data Release 1 (Flewelling et al. 2020), Sky-mapper Data Release 1.1 (Wolf et al. 2018), Tycho-2 (Høg et al. 2000; Pickles & Depagne 2010), the Yale Bright Star Catalog (Hoffleit & Jaschek 1995), and GAIA DR2 (Gaia Collaboration et al. 2018), plus original data from the ATLAS Pathfinder survey. To ensure that the stars used for determining the zeropoint of the image are well exposed, but not saturated, we select a subset of the catalog between $10 < g' < 12$. We also exclude stars with colors redder than $g - r = 1.5$, which might bias the photometry due to unconstrained chromatic aberrations affecting the PSF. We calculate an instrumental magnitude for each of the reference stars using a forced aperture at the catalog position in the background-subtracted and calibrated science image, typically with an aperture radius of 3 pixels ($40''$).

To model the variation in the photometric zeropoint across the field of view, the science image is divided into an 8×12 grid of square subframes, each of which subtends 4 square degrees and contains O(100) reference stars. Within each subframe, we calculate the sigma-clipped median offset between the instrumental magnitudes calculated via aperture photometry and the catalog values. The offsets in each region are then smoothly interpolated over the rectangular mesh of the full-sized grid, using quintic splines to produce a spatially varying zeropoint $z(x, y)$. The resulting image has the unit of magnitude and it has as its values the photometric zeropoint at each pixel, defined such that

$$m_g = -2.5 \log_{10} F_{\mathrm{aper}} - z(x, y), \tag{5}$$

where F_{aper} is the measured flux from aperture photometry.

Finally, we calculate the magnitudes for each candidate detected, as described in Section 3.3.4, based on their centroid positions in the detection image. The centroids are calculated for each candidate by computing their value-weighted average position ("center of mass"). This process uses a custom aperture photometry routine, implemented in Cython for the Evryscope precision photometry pipeline (Ratzloff et al. 2019), on the science image.

4. Automated Vetting

Despite the optical consistency of the Evryscope images chosen for subtraction, the direct subtraction process produces thousands of false positives per image. The observed sources of false positives are plentiful, from inside the CCD sensors out to Earth's orbit, including:

1. cosmic-ray muon tracks;
2. Compton recoil electrons from radionuclides in materials at the observatory;
3. optical ghosts;
4. registration and astrometric errors;
5. persistent residual charge from bright stars remaining after cycling the detectors;
6. flat-fielding errors;
7. aircraft strobes;
8. tumbling satellites and debris (Corbett et al. 2020); and
9. noise artifacts from both photon and astrometric noise.

In total, the event rates from these sources can outnumber the real on-sky rates of astrophysical transients by orders of magnitude. Human candidate inspection remains standard, but it is not scalable to surveys producing hundreds of thousands of candidates per night. As a result, a reliable, efficient, and automated vetting system for candidates is a core component of any transient survey producing an actionable event stream that can be delegated to follow-up resources.

Some false-positive sources in the list above can be identified with simple filters: bright streaks from satellites involve thousands of pixels, while residual charges from bright stars can be flagged based on previous astrometric solutions. Other classes of observed signals can be difficult to identify from simple metrics, in all scenarios. To account for this, we use a combination of data cuts based on explicit filters and ML methods.

4.1. Initial Candidate Filters

While ML techniques can be comprehensive, simple filters grounded in domain knowledge can be both more efficient as well as more easily interpreted. Starting from an initial deep source extraction (signal-to-noise ratio or S/N > 3 in a minimum of 1 pixel) of the detection image $D(x, y)$, we implement three first-order quality cuts, removing candidates that meet any of the following conditions:

1. centroid within 15 pixels of the edge of the CCD;
2. a ratio of negative to positive pixels within a 6 pixel circular aperture >0.4; or
3. more than 750 pixels above the detection threshold.

Detections near the edge of a CCD are typically caused by small amounts of mount drift between the science and reference images. Large ratios of negative to positive pixels typically indicate a photon or astrometric noise artifact. Extended events >750 pixels are commonly bright streaks, predominantly caused by aircraft and low Earth orbit (LEO) satellites.

We apply an additional filter after the ML vetting described below; we reject any candidates coming from a subtraction with more than 500 high-confidence candidates. These failed subtractions rarely occur, and are caused by a doubled or streaked image, due to wind shake at the instrument, or by a breakdown of the assumption of a slow and smoothly varying sky background, required for direct image subtraction, as described in Section 3.3.

With no additional vetting, these simple filters reduce the per-image candidate count to O(10^2), using the baseline values stated above; however, these numbers are readily tunable to the science case and corresponding false-positive tolerance, either by modifying a configuration file for the EFTE pipeline instance running at each observatory or by filtering the database queries that are used to regularly report candidates to end users. Candidates that pass the thresholds for these filters at the pipeline instance level are inserted into the central database, including small 30×30 pixel "postage stamp" cutouts around their detection positions.

4.2. VETNET: Real–Bogus Classification with CNNs

For the additional reduction of the EFTE false-positive rate, we use an ML model based on 2D convolutional layers (LeCun et al. 1989), with weights conditioned directly on the image data. This model is a binary ("real–bogus," or RB) classifier, which assigns each candidate a score between 0 and 1, where a score of 1 indicates that the candidate is likely real. RB classifiers have seen long-standing use in transient surveys, starting with the model that was built by Bailey et al. (2008) for
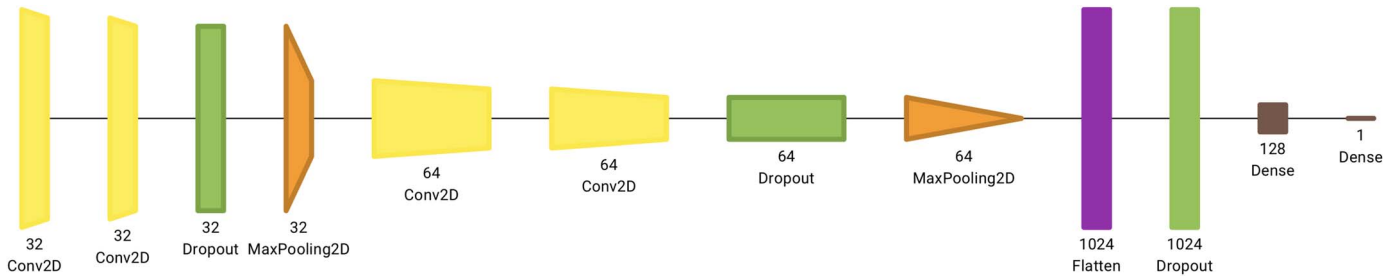
Figure 6. Architecture of VETNET, a convolutional RB classifier used by EFTE. The inputs to the network are a triplet of 30×30 pixel cutouts around the center of each candidate, taken from the reference, science, and direct difference images. All trainable layers, except for the final dense unit, use ReLU activation. The pairs of convolution layers are all followed by max-pooling layers, with 20% dropout for regularization. The network visualization was generated with Net2Vis (Bäuerle et al. 2021).

the Nearby Supernova Factory (Aldering et al. 2002). Similar approaches have been used for PTF; Bloom et al. 2008; Law et al. 2009), iPTF (Brink et al. 2013), DES (Goldstein et al. 2015), and, most recently, ZTF (Duev et al. 2019; Mahabal et al. 2019) and GOTO (Killestein et al. 2021).

Deep learning is a type of ML in which "deep" stacks of artificial neural network layers (McCulloch & Pitts 1943) are used to transform input data into latent space encodings that can be mapped to the desired output quantities. CNNs (LeCun & Bengio 1995) comprise a subclass of artificial neural networks that build up a latent space representation of pixel data using convolutions, which identify increasingly compressed features of the input as the depth of the network increases, as opposed to requiring the preselection of computed—and potentially suboptimal—features to represent the data. CNNs have found widespread use in astronomy for tasks including source detection and deblending (Burke et al. 2019; Stoppa et al. 2022), in addition to transient RB vetting (Förster et al. 2016; Duev et al. 2019; Killestein et al. 2021; Makhlouf et al. 2022).

In this section, we describe VETNET, a CNN-based vetting algorithm that has been trained to assign RB probabilities to EFTE candidates directly from 30×30 pixel cutouts from the reference, science, and direct subtraction difference images.

4.2.1. Training Set and Data Labeling

Supervised ML classifiers require large data sets of labeled examples to identify complex latent associations during training. In general, there are two options for producing these data sets: simulation or human classification. Exclusively training on simulated data is risky, because the efficacy of the final model is dependent on how representative the simulations are of real data. However, human classification is labor-intensive and prohibitive at the level of producing thousands or even millions of labeled examples across a representative sample of the survey.

As a result, we have adopted a compound approach, using both hand-labeled on-sky data and simulated events, which are produced via spatially varying PSF injection. The simulated data set was used to train intermediate models to prescreen events for human labeling, including the prototype CNN used in Corbett et al. (2020). We manually classified the moderate-purity on-sky sample generated by the intermediate model to produce a smaller, but minimally contaminated and representative, data sample for training the production models.

4.2.2. Network Architecture

VETNET uses a sequential VGGNet-like (Simonyan & Zisserman 2014) model with six trainable layers—four convolutional layers and two fully connected output layers. Each set of convolution layers is subject to 20% dropout, to prevent overfitting, encouraging the model to build a diverse set of representations of the data distribution. The dropout fraction at each layer is determined using the HyperBand band algorithm (Li et al. 2018), with a binary cross-entropy loss function. Further regularization is provided by a pooling layer, which reduces the dimensionality of each convolution block output by a factor of 4. The outputs of each pooled layer are normalized and recentered on zero using batch normalization (Ioffe & Szegedy 2015), to improve the training performance and model stability. All convolution layers use 3 pixel square kernels and ReLu activation (Agarap 2018), except for the final fully connected node, which has a sigmoid activation function that produces an output value normalized between 0 and 1. This output, the VETNET RB score, can broadly be interpreted as the probability that a given candidate is real. Figure 6 depicts the architecture of the model, including the filter depths and resulting dimensionality.

VETNET is implemented in Tensorflow (Abadi et al. 2015), using the high-level Keras API (Chollet et al. 2015).

4.2.3. Dropout and Model Uncertainty

CNNs can have an arbitrarily large number of free parameters, and are accordingly able to overfit the training data. As a result, methods of regularizing the training process and the weights that are assigned to the convolutional filters are necessary to maximize the performance on the actual data at inference time. Dropout (Srivastava et al. 2014) is one common technique, in which a tunable fraction of the outputs from a layer are chosen at random and set to zero, preventing them from contributing to the final network outputs.

In addition to slowing overfitting, dropout also can be interpreted as an approximation of Bayesian inference (Gal & Ghahramani 2015a). In this framework, each random sampling of layer outputs can also be considered as a sample from the distribution, representing the network weights in a fully Bayesian network. Evaluating a given sample through these different dropout-induced realizations of the network enables us to similarly approximate the posterior distribution of the network output. The advantage of this approach, called Monte Carlo (MC) Dropout, is that the output distribution includes the systematic uncertainty in the network output due to model selection, which is distinct from the random uncertainty
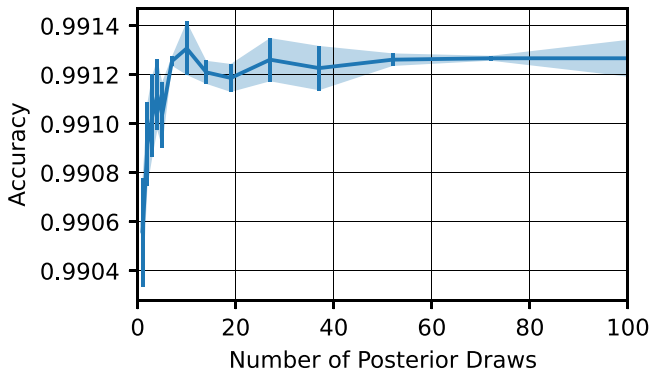
Figure 7. Network accuracy vs. number of forward passes through the network for the validation set. The performances converge at 10 samples. The error bars represent the standard deviations of the results across 20 iterations.

produced by the variance of the training set (Gal & Ghahramani 2015b). To produce an output from the network, each candidate is processed through multiple dropout-induced realizations of the network, producing a distribution of resulting RB scores. We use the median of this distribution as the RB probability for each source.

Interpretations of MC Dropout are unsettled in the literature (namely, whether it represents a genuinely Bayesian approximation; Le Folgoc et al. 2021). However, it can be used to produce a number that scales with the degree of consensus within the network and the amount of support for a sample within the training set, which can be interpreted as a confidence metric. This is similar to the interpretation of the sigmoid activation of the network as a whole as an RB probability, despite not representing a normalized probability density function. We adopt the entropy-based metric from Killestein et al. (2021) to quantify the network confidence:

$$\mathbb{C} = 1\frac{1}{N}\sum_{i=1}^{N} - p_i \log_2 p_i - (1 - p_i)\log_2(1 - p_i), \qquad (6)$$

where N is the number of samples from the posterior distribution and p_i is the network output for the ith sample. The metric $\mathbb{C}$ is the binary entropy of the Bernoulli process representing the RB classification, averaged across posterior draws, and is bounded on the interval [0, 1]. In Section 5.3, we demonstrate that $\mathbb{C}$ also matches the subjective confidence of human vetting.

The number of forward passes used to approximate the network output posterior distribution is determined empirically from the validation set. Figure 7 shows the accuracy of the classifier as a function of the number of forward passes through the network. The accuracy of the median RB score converges after 10 inferences, which is consistent with the findings in Killestein et al. (2021), despite the dropout rate here being two orders of magnitude higher.

4.2.4. Training Set and Data Augmentation

Two data sets were used to train the VETNET classifier; the simulated data set of 435,452 candidates described in Section 4.2.5 and a human-annotated sample of on-sky detections, containing 31,092 candidates flagged as probably real by an earlier iteration of VETNET itself (Corbett et al. 2020). Unlike the simulated data set, the on-sky data set is

heavily class-imbalanced, with only 9.6% of the examples (2,976) being human-labeled as real. To account for this class imbalance, we randomly exclude 25,140 of the bogus samples from the on-sky data set, noting that the simulated data set only contains simulated examples of the real class. The bogus examples in the simulated data set are drawn from the same population as the bogus examples in the on-sky data set. Our approach to maximizing the return from this relatively small sample of on-sky data is described in Section 4.2.6.

We divided the simulated and annotated on-sky data sets into training, validation, and testing subsets, using an 80:10:10 ratio. We used the validation set to tune the MC Dropout fractions and the number of posterior draws, as well as for monitoring the training process.

To extend the effective size of the data set, random flips and rotations are applied to each batch of training samples. As noted by Killestein et al. (2021) and Dieleman et al. (2015), rotations (other than in 90° increments) require interpolation, and thus distort the data from the pixel grid; however, in our use case, the data have previously been resampled with interpolation by the image alignment process (see Section 13). No data augmentation is applied during validation or model evaluation, nor for the training during the fine-tuning with human-annotated data.

4.2.5. Simulated Data Generation

We generated a base training set by injecting simulated transients into 300 randomly selected images across the first 2 yr of full Evryscope science operations at each site. The images were selected uniformly in time, meaning that the moon phase, sky conditions, focus changes due to temperature variations, and dust accumulation on the instrument (leading to measurable changes in the background level and limiting the magnitude on timescales of a few months) are uniformly represented. Each image was calibrated as in the pipeline (see Section 3.3).

For each image, we generated a uniform sample of 5000 positions within the image, then deduplicated the sample, so that no position was within 50 pixels of any other position, to avoid overlapping transients, resulting in an average of 1200 injections per image. While this does bias the initial training set against contemporaneous spatially coincident events, these are sufficiently rare for us to neglect this scenario. Each injection was assigned a random magnitude, drawn from a uniform distribution bounded between the typical saturation limit of $g \sim 7$ and the 1.5σ detection limit at the injection position (determined by the photometric zeropoint interpolation procedure described in Section 3.4).

A second round of injections was done to simulate transients with known visible progenitors. From the catalog stars within each image, 500 stars that were minimally separated by 50 pixels were selected as additional positions for injection. The variability amplitude was uniformly sampled between magnitudes of 0.25 and 8. The upper limit was set by the maximum contrast visible for a predetected star in an Evryscope image, i.e., a star at the dim limit of the survey that reaches the single-exposure saturation limit.

Evryscope PSFs are heavily impacted by optical aberrations, and they exhibit a wide variety of morphologies, both between the cameras and across the fields of view of individual cameras (Ratzloff et al. 2020), making common analytic profiles (such as Moffat, Gaussian, and Lorentzian) untenable. Further, the

coarse pixel scale results in more complex linear models—such as the ePSF (Anderson & King 2000) and those used by PSFEX (Bertin 2011) and `PSFMachine` (Hedges et al. 2021)—being prone to poor fits, due to aliasing and source confusion. We found that the most robust method for simulating transients with morphologically plausible point-like profiles was to build a model PSF based on nearby isolated stars. For each injection position, up to 100 nearby stars with distances less than 137 pixels and significances of 10σ above the local noise were extracted with 30×30 "postage stamp" windows. Each stamp was then multiplied by a smoothly varying (Hanning) window and normalized. The final PSF, to be flux scaled and added into the image, was then the median of the nearby stamp templates, weighted by the relative normalized distance from the injection position and the relative flux uncertainty of the template star. Figure 8 shows examples of simulated PSFs using this technique across a typical Evryscope focal plane for a range of magnitudes, alongside the resulting signal in a direct subtraction image with a consecutive epoch.

At the end of the transient injection process, the image is "decalibrated," by adding back in the expected dark current, bias, and background levels, and the image is converted back into ADU units, with pixel values beyond the range of an unsigned 16 bit integer being truncated, matching the histogram of the simulated images to the distribution that is expected for science images.

To produce a simulation-augmented data set, the transient-injected images are reduced using the EFTE pipeline, and any candidates that are identified within 2 pixels of an injection location are labeled as real, while all others are labeled as bogus. Despite the large number of injected sources, this process produces an unbalanced data set, with artifact detections outnumbering injections at rates up to 1000 to 1. To balance the data set, we randomly select a number of bogus candidates equal to the number of recovered injections for inclusion in the final data set. This results in a data set with noisy labels, due to both background transients (likely dominated by short-duration reflections from Earth satellites; Corbett et al. 2020; Nir et al. 2021) and candidates that were injected below the difference image threshold and recovered coincidentally. From a visual inspection of 10,000 injection candidates, uniformly sampled from both known injections and predicted artifacts, we estimate label contamination to affect $\leqslant 2.7\%$ of the candidates in the data set of 435,452 candidates. Deep convolutional models have been observed in prior works to be robust to many times this level of label noise (Ghosh et al. 2016; Rolnick et al. 2017).

4.2.6. Staged Training Methodology

A common approach to building specialized models with limited training data is to utilize transfer learning, by leveraging the pretrained feature representations of existing models that have been built with massive related data sets. Rather than training an entire model from scratch, which requires a large annotated data set over the domain of interest, a pretrained model can be selectively "fine-tuned" over a representative data set in a new domain. We adopt a similar approach, to make use of the considerable diversity of the observing conditions and sky regions that are represented in the simulated data set described in Section 4.2.4, while minimizing the risk of optimizing for the properties of the transient injection process

(see Section 4.2.5), rather than the properties that are transferable to on-sky data.

Our training curriculum for VETNET was as follows:

1. train the full model—including all convolutional and fully connected layers—on the simulated data set, until convergence, to create the synthetic base model;
2. freeze the weights on all convolutional layers from the synthetic base, and retrain the fully connected layers from scratch using on-sky data; and
3. unfreeze the convolutional layers, then train at a minimal learning rate using on-sky data, until convergence, to produce the final on-sky model.

We used the Adam optimizer (Kingma & Ba 2014) with a binary cross-entropy loss function for all three stages. For the first two training stages, we started with a maximum learning rate of 0.0003, slowed by a factor of 2, whenever the loss on the validation set plateaued for 10 epochs. Scheduling the learning rate in this way helped the network to converge to a local minimum when near a global minimum, and decreased the oscillation around the minimum of the loss function. For the final fine-tuning of the entire model, we reduced the initial learning rate to 0.00001, while maintaining the same scheduled rate decay.

The synthetic base model converged after $\sim$100 epochs, taking about 6.5 hr when training on an Intel Xeon E5-2695v4 CPU. The fine-tuning with on-sky data converged after $\sim$150 epochs, but took less than 10 minutes, due to the smaller quantity of data.

In the final stage, we use a single iteration of the semisupervised relabeling routine suggested by Killestein et al. (2021); the samples in the training set that are differently classified by the network than by human labelers are flipped to the model classification, in cases where the model confidence $\mathbb{C}$ is greater than the median. On review, these samples are generally difficult to classify by hand, with low-significance peaks relative to the surrounding noise, areas of the sensor plane with pathological PSFs, or likely errors in the original labeling process. Figure 9 shows three representative examples that were relabeled by VETNET during this process. In total, fewer than 4.7% of the samples were changed in the training set after the labels were updated. We note that this is comparable to the label noise in the synthetic data set (2.7%).

4.3. Candidate Crossmatching and Source Association

EFTE candidates from both sites and their corresponding metadata are stored in a relational database. On insert, candidates are associated with previous candidates at the same position, which collectively form an "event," via an insert trigger within the database. If a candidate has no antecedent, a new event is created, and an additional trigger crossmatches the new event's position with a variety of externally produced reference catalogs. At the time of writing, these reference catalogs include the International Variable Star Index (Watson et al. 2015), the Galaxy List for the Advanced Detector Era (GLADE; Dálya et al. 2018), ATLAS-REFCAT2, and the ASAS-SN Catalog of Variable Stars (Jayasinghe et al. 2018). Stellar sources are crossmatched with a radius of $26''$ (corresponding to 2 Evryscope pixels and the worst-case astrometric performance for EFTE detections—see Figure 13), and galactic sources from GLADE are crossmatched with a $1'$ radius.
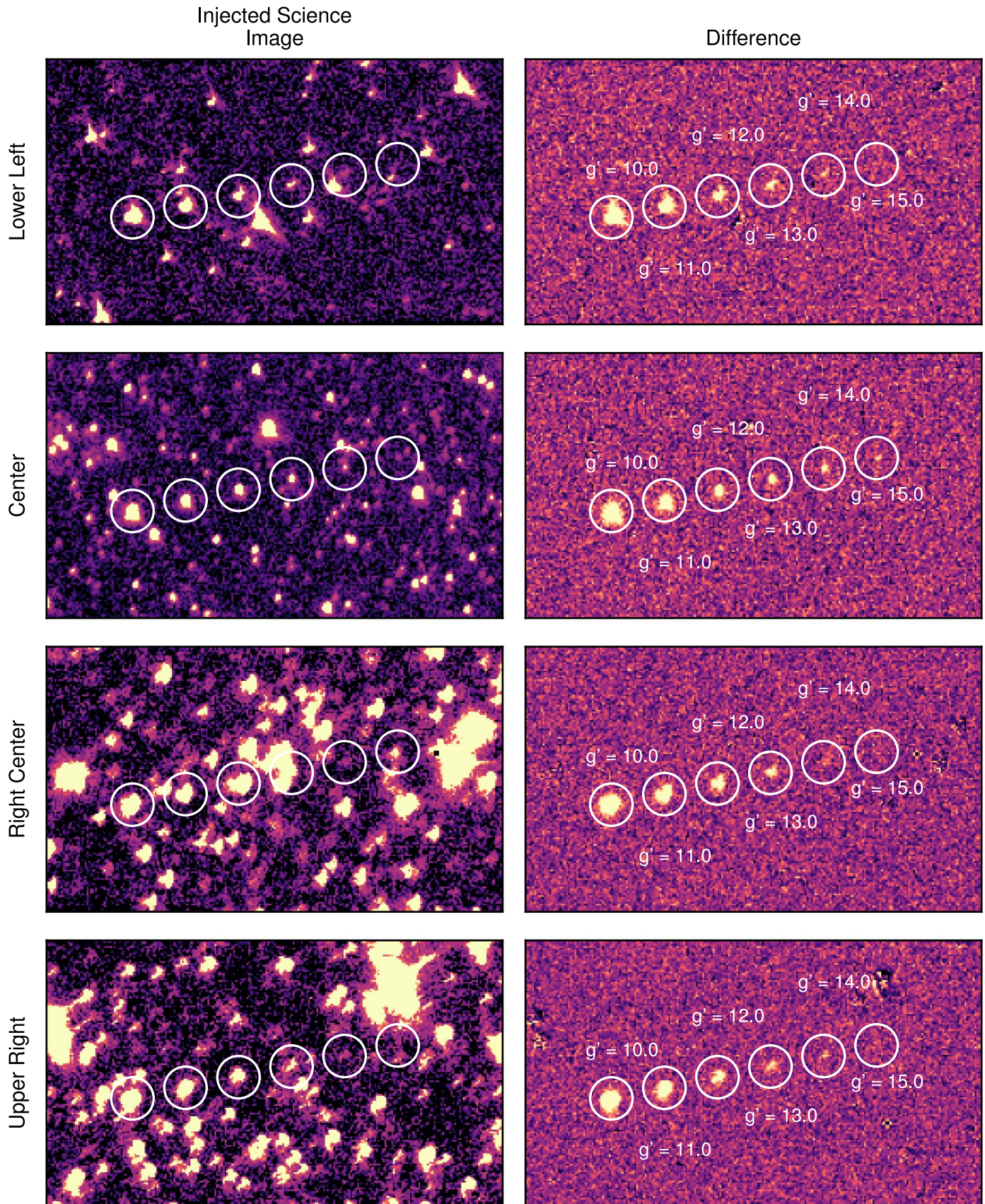
Figure 8. Left: examples of simulated transients with magnitudes $g' = 10, 11, 12, 13, 14, 15$ at the center and edges of a typical mid-galactic-latitude Evryscope image. The PSFs at each position are modeled as normalized, aligned, and sigma-clipped combinations of nearby isolated stars, producing morphologically plausible star-like injections. These simulations are used for the initial conditioning of our ML vetting system. Right: difference images for each subimage, using a reference frame taken 2 minutes before the sunjected science image.
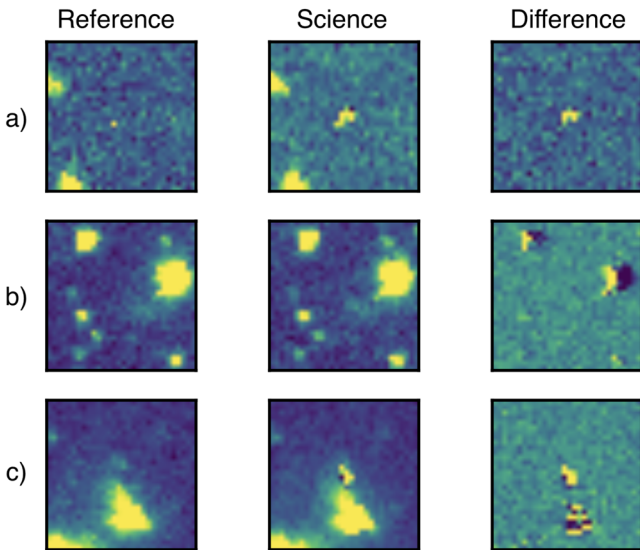
Figure 9. Misclassified training set samples relabeled based on the entropy-based confidence of the VETNET predictions. Network reclassifications typically affect samples that are difficult to classify manually. Samples (a) and (c) were initially classified as real by human vetters, but were relabeled as bogus by the algorithm. Both have pathological PSFs, likely caused by interpolation artifacts from resampling near a cosmic ray, a particle strike, or an unmasked bad pixel. Sample (b) is of low significance and off-center, but was manually classified as bogus, before being confidently relabeled as real by VETNET.
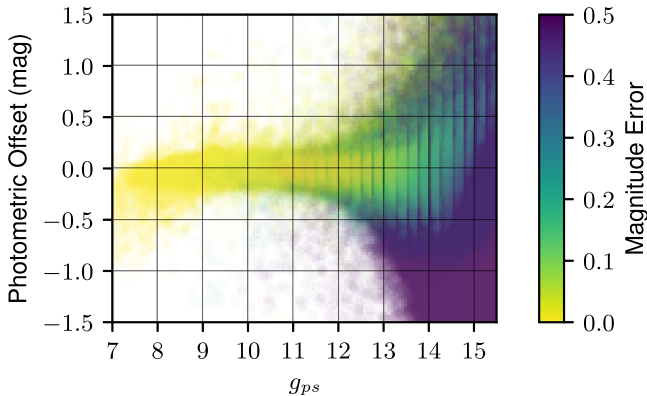
Figure 10. Photometric calibration offsets between ATLAS-REFCAT2 and EFTE. The median rms offset in the region $8 < m_g < 14.5$ is 0.06 mag. The anomalously bright and high-precision measurements (upper right) are due to source confusion and blending. The underreporting of magnitudes due to saturation is evident for stars brighter than $g' = 8$.

To accelerate the in-database crossmatching and candidate queries, all candidates, events, and reference catalogs are indexed using the Quad Tree Cube (Q3C) pixelization scheme,[11] a PostgreSQL extension for efficient spherical crossmatching and radial queries (Koposov & Bartunov 2006). Sky areas, such as the on-sky footprints of images or probability contour regions for multimessenger transient events, are indexed using PostGIS, with a custom nongeodetic projection. This projection does not include the WGS-84 (Kumar 1988) reference ellipsoid, and represents the R.A. and

decl. in the standard barycentric celestial reference system in all EFTE application code.

Candidates can also be crossmatched against external triggers received by EFTE, via automated circulars from the NASA Gamma-ray Coordinates Network/Transient Astronomy Network. Alerts are inserted in the central database by an automated ingest microservice, where they are indexed by position, by using Q3C for tightly localized triggers, as PostGIS polygons for events that are distributed as polygon skymaps, like LIGO/Virgo skymaps (Abbott et al. 2009), or as GRB alerts from the Fermi Gamma Burst Monitor (Bhat et al. 2009).

5. Pipeline Performance Evaluation

5.1. Photometric Solutions

To evaluate the performance of the photometric calibration using a smoothly varying zeropoint, as described in Section 3.4, we compare single-epoch forced photometry from 3,217,215 catalog stars from ATLAS-REFCAT2 across 500 randomly selected images from the 2018 observation year. The images were required to pass the quality assurance metrics described in Section 3.1, but were not otherwise filtered for sky or instrumental conditions. Figure 10 gives the distribution of the photometric offsets and the offset rms as a function of magnitude in the individual images. The resulting photometry is calibrated to the reference catalog, with an RMS offset of 0.05 mag between $8 < m_g < 14.0$, measured using five iterations of a 5σ clip, to remove outliers due to single-epoch failures.

These numbers likely represent an upper limit on the photometric rms for isolated and dim events, as the distribution is dominated by source confusion beyond $g' = 14$ (e.g., dim catalog stars with a brighter star near or within the 6 pixel aperture), causing anomalously bright and high-precision measurements of dim catalog sources. Sources brighter than $g' = 9$ are occasionally saturated when they appear near the center of the image, though sources as bright as $g' = 8$ are typically well calibrated and linear. There is a noise floor around ~5% for single-epoch detections from Evryscope, due to the variation in the subpixel response across the image plane. These effects are modeled in data products from the Evryscope precision photometry pipeline (Ratzloff et al. 2019), but they are prominent in the raw single-epoch bright star photometry from EFTE.

Additional color and airmass terms can be applied to the light curves of EFTE photometry as needed, using the equation

$$g_{\mathrm{EVR}} = g_{\mathrm{PS}} + A + B(g_{\mathrm{PS}} - r_{\mathrm{PS}}) + k_1 X + k_2 X(g_{\mathrm{PS}} - r_{\mathrm{PS}}),$$

(7)

where g_{EVR} is the magnitude in the Evryscope g band, g_{PS} and r_{PS} are the PanSTARRS magnitudes from the ATLAS reference catalog, X is the airmass of the star, and A, B, k_1, and k_2 are fitted photometric conversion factors between the Evryscope and PanSTARRS bandpasses. Based on fits to forced aperture light curves using a robust estimator (Fischler & Bolles 1981), the photometric conversion terms are $A = 0.037 \pm 0.002$, $B = -0.051 \pm 0.004$, $k_1 = 0.021 \pm 0.002$, and $k_2 = -0.051 \pm 0.003$. The light curves used to fit these parameters were chosen from a random sample of 25,000

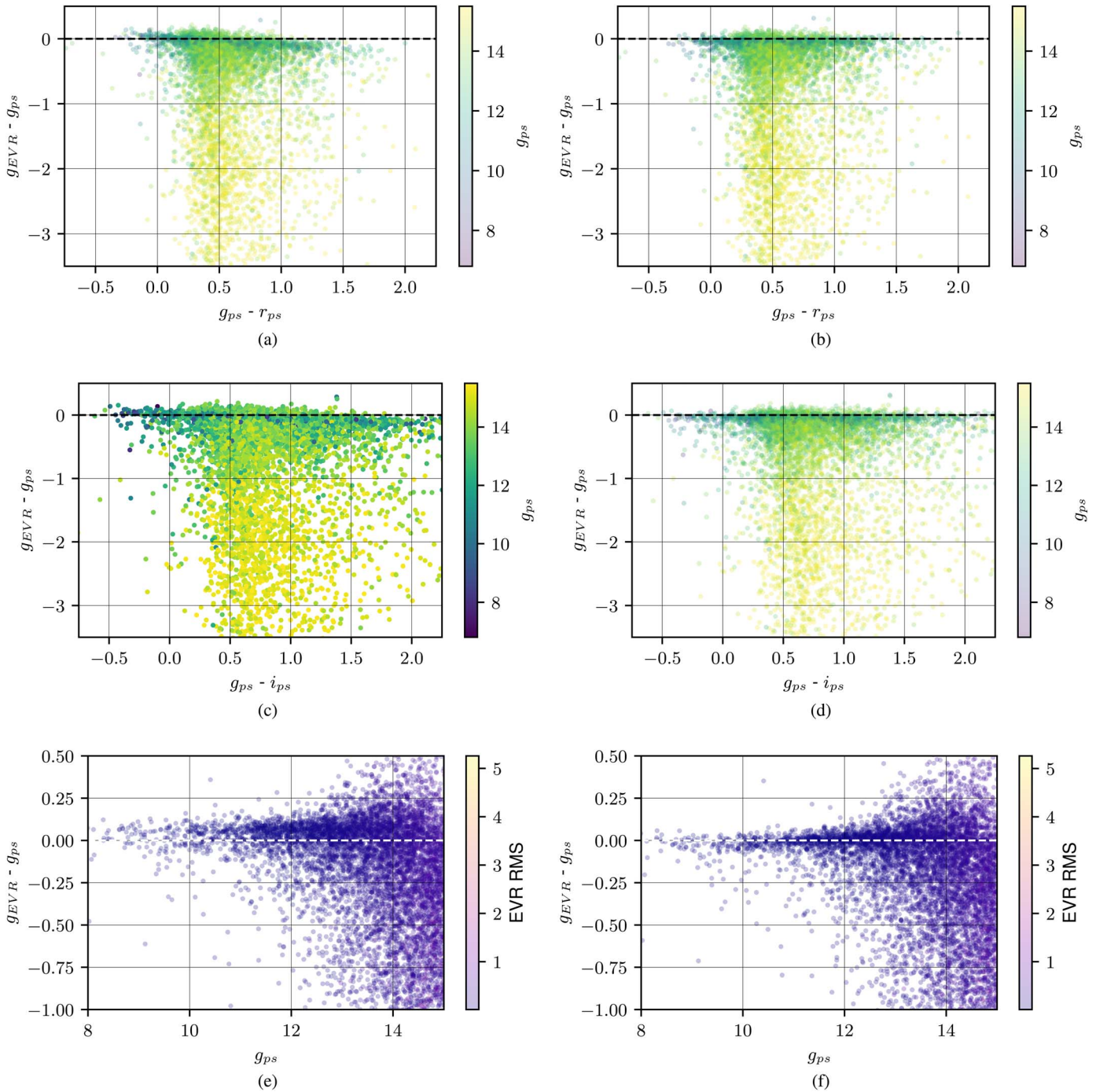[11] See https://ascl.net/1905.008 (Koposov & Bartunov 2019).

Figure 11. (a) and (b) Photometric offsets between the Evryscope and PanSTARRS g bands as a function of $g - r$ before and after calibration with the color and airmass terms above, respectively. The large offsets below the linear trend are caused by blended sources and low-S/N detections that were not filtered based on the light-curve quality metric described in Section 5.1. (c) and (d) The same as above, but with $g - i$ colors in place of $g - r$, to demonstrate the performance over a wider variety of colors. No calibration fits are made as a function of the i-band colors. (e) and (f) The photometric calibration performance for many-epoch light curves as a function of magnitude. The sigma-clipped rms photometric offset decreases from 0.16 to 0.06 mag for sources between $8 < m_g < 14.5$ with the application of color terms.

northern hemisphere stars, which we then filtered based on a quality metric, including the source variability relative to nearby stars, the saturation, and the shape of the aperture flux growth curve, leaving a cleaner sample of 10,671 light curves with an average of 17,140 epochs,

Figure 11 shows the photometric offset as a function of $g - r$ and $g - i$ colors, before and after applying the calibration offset, as well as the resulting impact on the long-term

photometric accuracy of the light curves. The application of the color and airmass correction brings the rms calibration accuracy of the long-term light curves from 0.16 to 0.06 mag, in line with the single-epoch measurements above.

Public EFTE data products, including both transient alerts and long-term photometric light curves, do not include color and airmass terms. For the light curves, calibration for photometric precision, rather than accuracy, is prioritized.
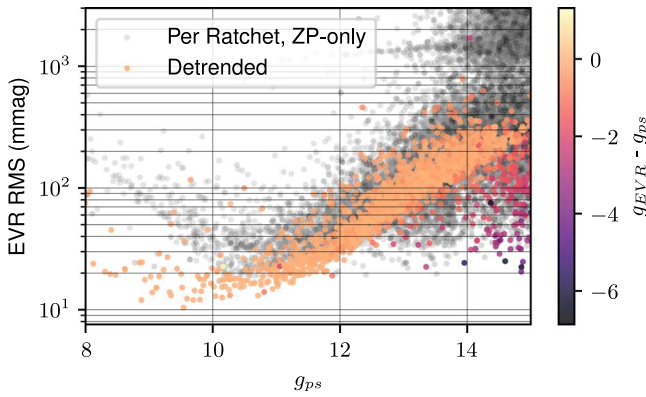
Figure 12. The measured rms values of 10,671 randomly selected long-term Evryscope light curves. The performance in the detrended light curves ranges from 20 mmag at the bright end to 20% for dim sources. The rms values for the raw light curves are averages of the measured rms in each pointing, neglecting zeropoint offsets between ratchets.
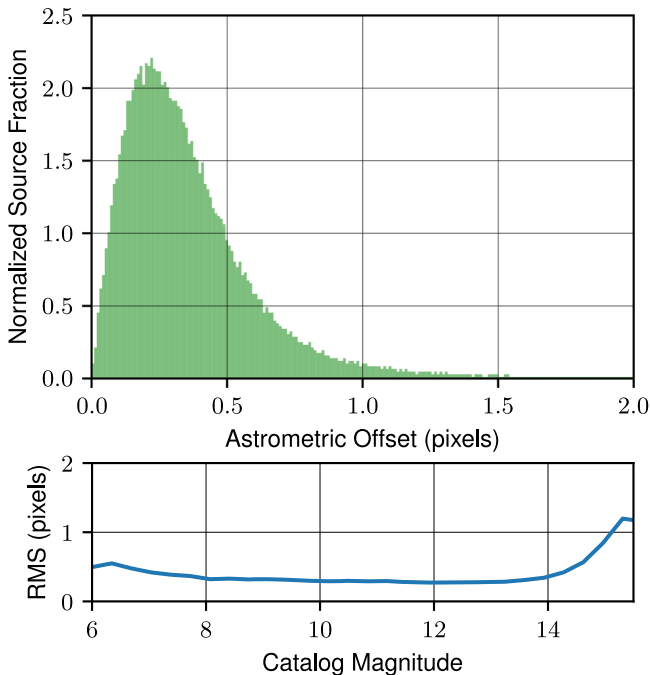
Figure 13. Top: astrometric localization performance for the EFTE pipeline. Bottom: rms localization scatter as a function of magnitude. The performance is subpixel at the 99th percentile, with a typical rms scatter of $7''$, excepting stars brighter than 7 mag, which are typically saturated, and those dimmer than 14 mag, where source confusion dominates in the source extraction.

The Evryscope light curves are first decorrelated from subpixel PSF variations, then detrended using a customized version of the SysREM algorithm (Tamuz et al. 2005), to correct for systematics, ultimately producing light curves that are self-consistent at the ≤20 mmag level at the bright end of Evryscope's operating range, as shown in Figure 12.

5.2. Astrometric Localization

The custom astrometry routines developed for the Evryscopes are capable of providing $1''$–$2''$ (0.08–0.15 pixel) rms astrometry over the field of view of the Evryscopes in single images; however, astrometric localizations from EFTE also depend on the quality of the alignment procedure that is used

for real-time reduction, and must therefore be characterized separately. In addition to the subpixel scatter induced by photon noise, EFTE localizations depend on the consistency of the pointing between consecutive images and the accuracy of the rigid transformations that are calculated to align each image with the previous image for which a WCS solution is available.

We evaluated the quality of this alignment routine by performing source detection in individual science images, aligned to previous target images in the same pointing. We chose target images with a typical Δt_D of 10 minutes, with samples of $\Delta t_D < 10$ minutes being representative of what would occur in the first few images of a ratchet.

The detected sources were then crossmatched with the sources in ATLAS-REFCAT2 (Tonry et al. 2018). As in Section 5.1, 500 science images for testing were randomly selected from the 2018 observing data set, across all weather and moon conditions. Figure 13 shows a histogram of the offsets between the catalog positions and the recovered positions in the aligned science images with reused WCS solutions. The astrometric performance was subpixel for 99% of the detected sources, with an rms scatter of less than $4''$ between 8 and 14 mag. As for the photometry, the precision was limited by saturation effects at the bright end and by source confusion for sources dimmer than 14.5. In all cases, the localization was accurate to within 2 pixels.

5.3. VETNET Model Evaluation

We evaluated the VETNET RB model using both the held back test set, described in Section 4.2.4, and an injection recovery program over a sample of randomly selected images.

5.3.1. On-sky Test Set

Figure 14 shows postage stamp cutouts and classification histograms from the held back test set of on-sky transients, divided evenly between cases where the VETNET classifications and the human-assigned labels agreed and cases where they disagreed. In both categories, the entropy-based confidence score scales with the subjective appraisal of the candidates; candidates (c), (h), (k), and (i) are faint borderline detections, which are accordingly assigned low confidence scores. Candidate (g) is a linear particle collision. Notably, candidates (f) and (l) have anomalously sharp PSFs, which were both counted as real by human labellers, but were assigned bogus scores by the network, suggesting that they are morphologically more similar to cosmics. Even in cases where the labels are consistent, the confidence drops in areas with pathological PSFs, as in example (b) and especially in example (a).

Figure 15 shows the performance of the model on the on-sky test set. The magnitude-integrated precision and recall at the VETNET score threshold of 0.5 are 95.4% and 94.4%, respectively, with a false-positive rate of 5.1%. Depending on the science case and false alarm tolerance of follow-up resources, these numbers can be tuned using a combination of the VETNET RB score and the $\mathbb{C}$ rating; for instance, a subpercent false-positive rate is measured above an RB threshold of 0.7.

5.4. Candidate Production Latency

To enable rapid follow-up, EFTE must produce candidates on timescales that are comparable to the earliest and most impulsive phases of the astrophysical events of interest, ideally
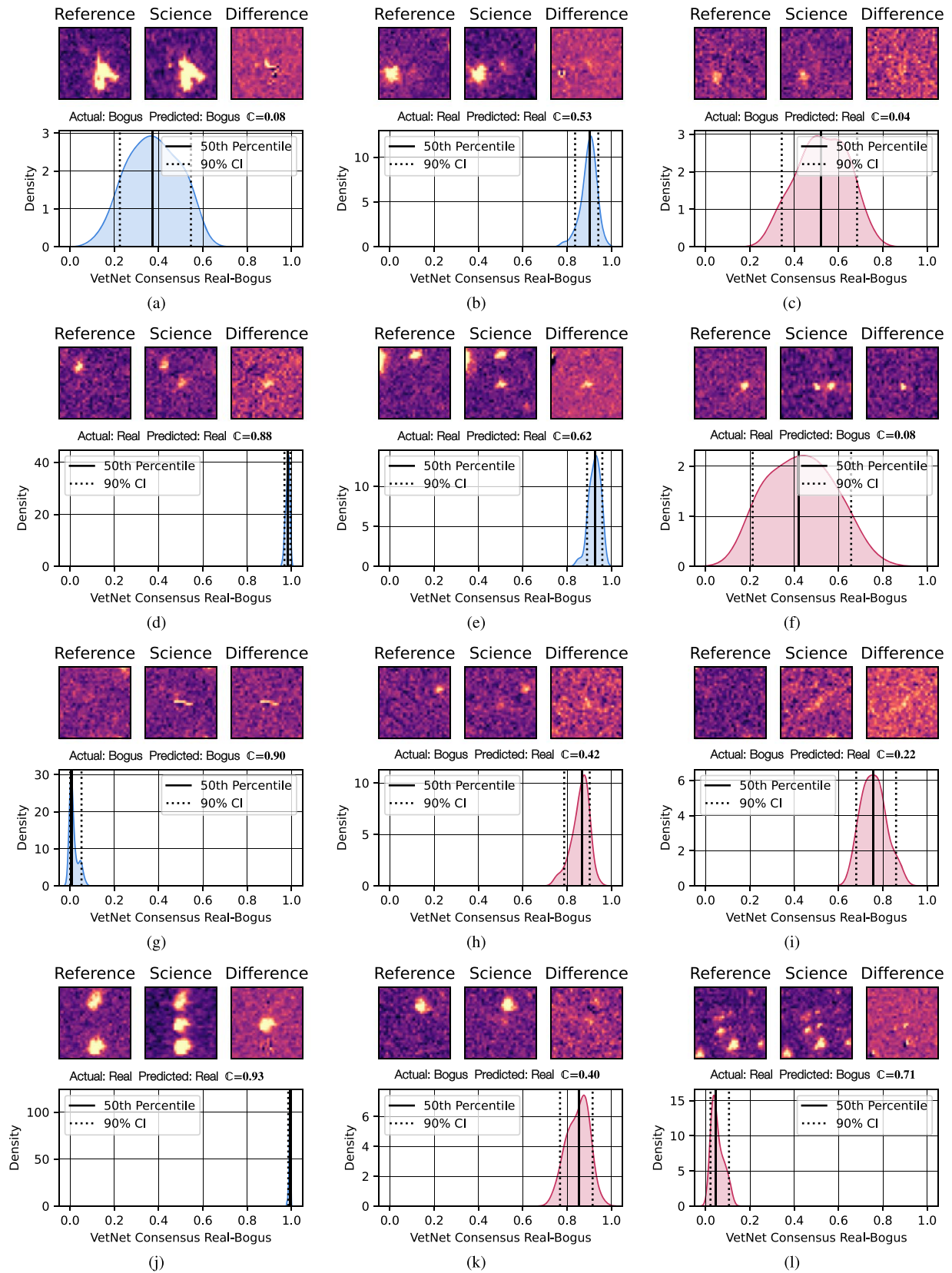
Figure 14. Sample on-sky candidates taken from the VETNET test set. The "Actual" and "Pred" values represent the sky truth class as determined by a human inspector and the predicted class as determined by the network, respectively. Each histogram is an approximation of the normalized Bayesian posterior distribution for the probability of the candidate representing a real astrophysical event, quantified using the entropy-based $\mathbb{C}$ metric from Killestein et al. (2021). In cases where the sky truth and the network prediction disagree, $\mathbb{C}$ is typically <0.5, or extenuating circumstances exist, such as the anomalous PSFs in panels (f) and (l), or the potential human misclassifications in (h), (i), and (k).
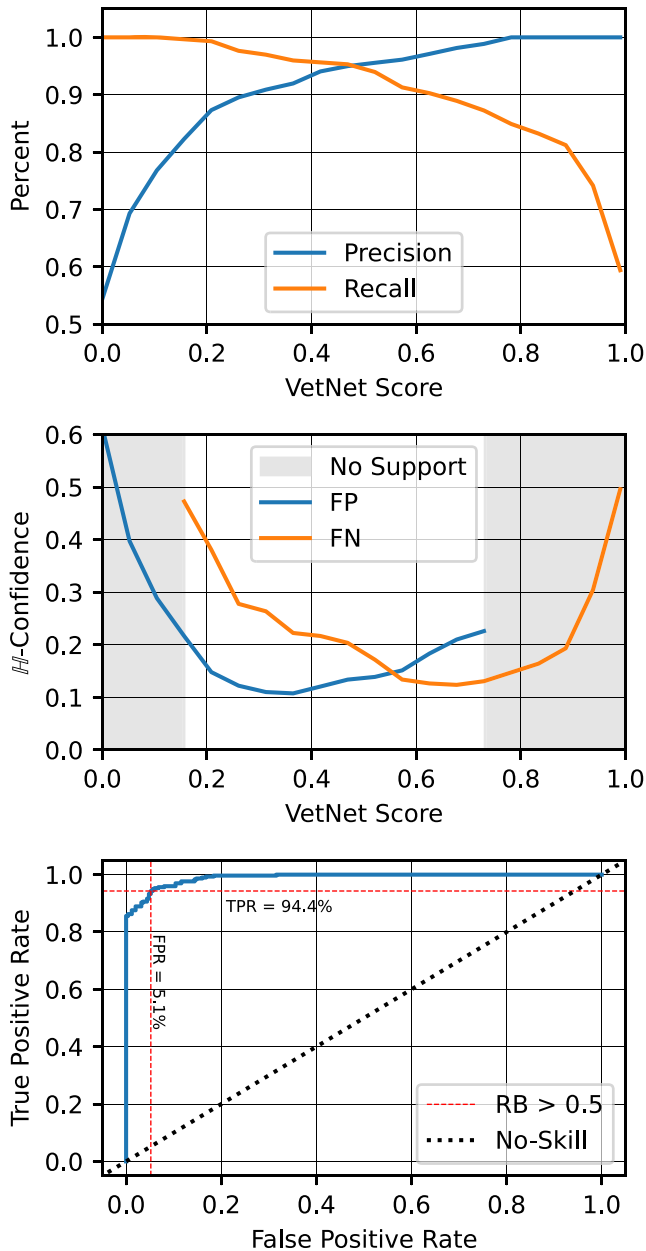
Figure 15. The performance of VETNET on the on-sky test set. Top: precision and recall as a function of the VETNET RB score. At the RB = 0.5 threshold, the observed precision and recall are 95.4% and 94.4%, respectively. Middle: entropy-based confidence scores for false positives and false negatives as a function of the VETNET RB score threshold. The shaded regions indicate the score regions where no false positives or false negatives occur within the on-sky test set. Bottom: receiver operating characteristic (ROC) curve for VETNET. The area under the ROC curve is 0.99, representing the probability that a random real candidate will receive a higher RB score than a random bogus candidate. The "no-skill" line indicates the expected performance curve for a random classifier.

within the base cadence of the survey. For Evryscope, this means adding candidates to an actionable event stream within 2 minutes of the end of each exposure. We consider the candidate production latency to be our figure of merit for speed, defined here as the time delay between the shutter close time for the image and the candidates being fully inserted into the central EFTE database at UNC-CH, with all automated vetting and in-database source association and deduplication actions being complete.

Figure 16 presents histograms of the candidate production latency for both Evryscope-North and Evryscope-South during early on-sky testing of EFTE, between 2019 November 25 and 2020 January 1. Some variation is seen between Evryscope-North and Evryscope-South, which we attribute to a combination of the differences between the on-site compute hardware specifications, the camera counts, and the varying network connectivity to each observatory. Cumulatively, between both sites, EFTE is able to meet the subcadence latency requirement for 98.5% of images.

5.5. Injection Recovery Testing for Completeness

To estimate the expected completeness of the survey, we selected 800 images from the 2021 Evryscope-North data set at random, injected simulated sources using the routine described in Section 4.2.5, and evaluated the recovery probability as a function of magnitude by using the routine described in Corbett et al. (2020). The ratio of variables (injected with a minimum contrast of 0.25 mag) to transients without a counterpart in the reference image was 1:6. In total, 960,000 transients were added to the images.

Figure 17 shows the fraction of simulated transients recovered from the test set, and the corresponding recovery fraction from Corbett et al. (2020), which used an earlier version of the VETNET model. We note that dropping the VETNET RB score threshold to 0.0 has a marginal effect on the dim end of the recovery curve, indicating that the decreased depth (50% at $m_g = 14$ instead of 50% at $m_g = 14.2$) is a property of the slightly different image sample, rather than one of the updated VETNET model. Sources brighter than $m_g = 13.2$ are successfully recovered in all images.

6. Science Results from EFTE

6.1. Rapid Follow-up of Stellar Flares with SOAR

EFTE's latency is fast enough for flare candidates to be observed by other telescopes in the minutes immediately following the flare's detection. The SOAR telescope is a 4.1 m telescope located at Cerro Pachon in Chile, which hosts the Goodman High Throughput Spectrograph (Clemens et al. 2004). The entirety of Evryscope-South's field of view is within SOAR's observable area, and a band of Evryscope-North's field of view below declinations of 10° is accessible to SOAR. Ongoing studies are pairing SOAR and Goodman with the EFTE alert stream to acquire spectra of stellar flares within minutes of the detection of those flares, allowing the spectral evolutions of flares to be characterized during the most impulsive phases of the flare, with the time resolution being limited by the exposure time that is necessary to obtain a spectrum with an adequate S/N (typically less than a minute for EFTE-detected stellar flares).

6.1.1. EVRT-3586872: A $\Delta m = 4.2$ Flare from a mid-M Dwarf

In an exposure beginning at 5:39:56 UTC on 2020 February 15, EFTE detected a new source from Evryscope-South, which was then confirmed in the two consecutive images, at magnitudes of 12.7 and 12.8, respectively, indicating that this source was both astrophysical in nature and potentially detected near its peak. The source, which was assigned the identifier EVRT-3586872, crossmatched to a star in the ATLAS reference catalog with a red color ($g - r = 1.186$), suggesting
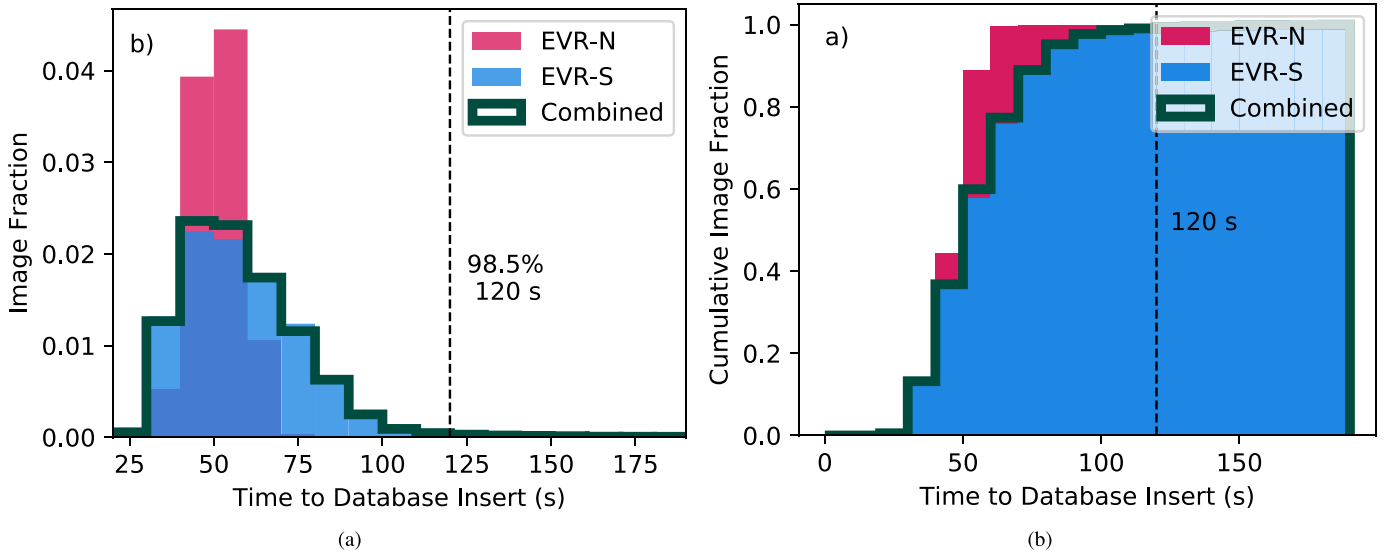
Figure 16. Cumulative histogram (a) and distribution (b) of the time delay between the exposure and the insertion of vetted candidates into the remote database, between 2019 November 25 and 2020 January 1. 98.5% of the images were fully reduced into lists of transient candidates within 120 s, before the next image was complete.
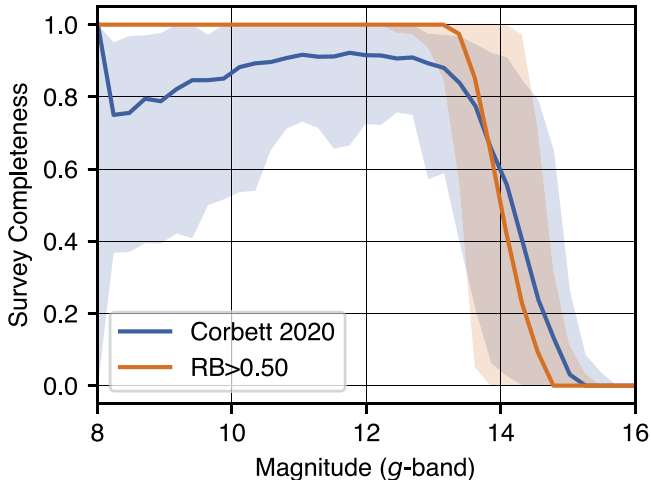

Figure 17. Survey completeness as a function of magnitude for both the current EFTE, with a VETNET RB threshold score of 0.5, and the previous version described in Corbett et al. (2020). The completeness is measured on a synthetic sample of injected transient and variable sources. The recovery probability for sources brighter than $m_g = 13.2$ is 99.9%, which rapidly falls to 50% at $m_g = 14$. The shaded regions represent the 90% confidence intervals of each curve, based on the percentiles of the per-image recovery functions.

a possible M-dwarf origin. The offset between the catalog star, 2MASS08593584-2340201, and the EFTE detection is $4.''0$, well within the expected astrometric error. The star is also cataloged in Heinze et al. (2018) as an irregular sinusoidal variable, consistent with an M-dwarf rotational signature.

Upon receiving the notification of consecutive detections of a cataloged red source via our web interface, we worked with SOAR staff to slew to its location, and began observing the target 14.9 minutes after the end of the first Evryscope detection image. We used the 400 line grating in the M1 configuration, approximately covering the wavelength range from 300 to 705 nm. Figure 18 shows a spectrum of the flare extracted from a 120 s exposure +15 minutes after the initial trigger, with the flux calibrated to the spectrum of LTT2445 (Hamuy et al. 1992, 1994).

We fit a two-component scaled blackbody model to the flare spectrum. This includes both a fixed contribution from the quiescent star, based on the Bayesian estimate of $T_{\rm eff}$ from `StarHorse2` (Anders et al. 2022), and thermal flare emission, with a best-fit temperature of 22,852 K. At almost 23,000 K, this temperature is larger than those typically assumed in flare models (Osten & Wolk 2015), but consistent with temperatures that have recently been inferred from broadband light curves from the NASA Transiting Exoplanet Survey Satellite (TESS) and Evryscope (Howard et al. 2020). We also note that this measurement is subject to known systematics, including Balmer continuum emission features (Kowalski et al. 2013) and increasing uncertainty as the true temperature increases, due to the optical bandpass primarily sampling the Rayleigh–Jeans tail of the spectrum at temperatures beyond 10^4 K (Arcavi 2022). Analysis of the results from the EFTE–SOAR follow-up program is ongoing.

6.2. EFTE Light Curves

In addition to transient alerts, EFTE enables users to produce long-term light curves for targets that are not included in the input catalog that is used for the Evryscope high-precision forced photometry pipeline (Ratzloff et al. 2019). EFTE light curves have been included in multiple publications, both by the Evryscope team and external collaborators. Publications using EFTE light curves include analyses of the galactic novae V1674 Her (Quimby et al. 2021) and V906 Car (Wee et al. 2020), measurements of the rotation periods of TESS exoplanet hosts, including one example with a 2.1 mmag amplitude that was measured from EFTE photometry (Howard et al. 2021), and the long-term monitoring of a mysterious dust-emitting object that is orbiting the star TIC 400799224 (Powell et al. 2021).

6.3. Satellite Glint Foregrounds for Fast Transient Surveys

Image contamination by Earth artificial satellites takes two forms: streaks, with uniform illumination over extended trajectories, and glints, which appear as short-duration flashes. These two morphologies are frequently degenerate, and depend
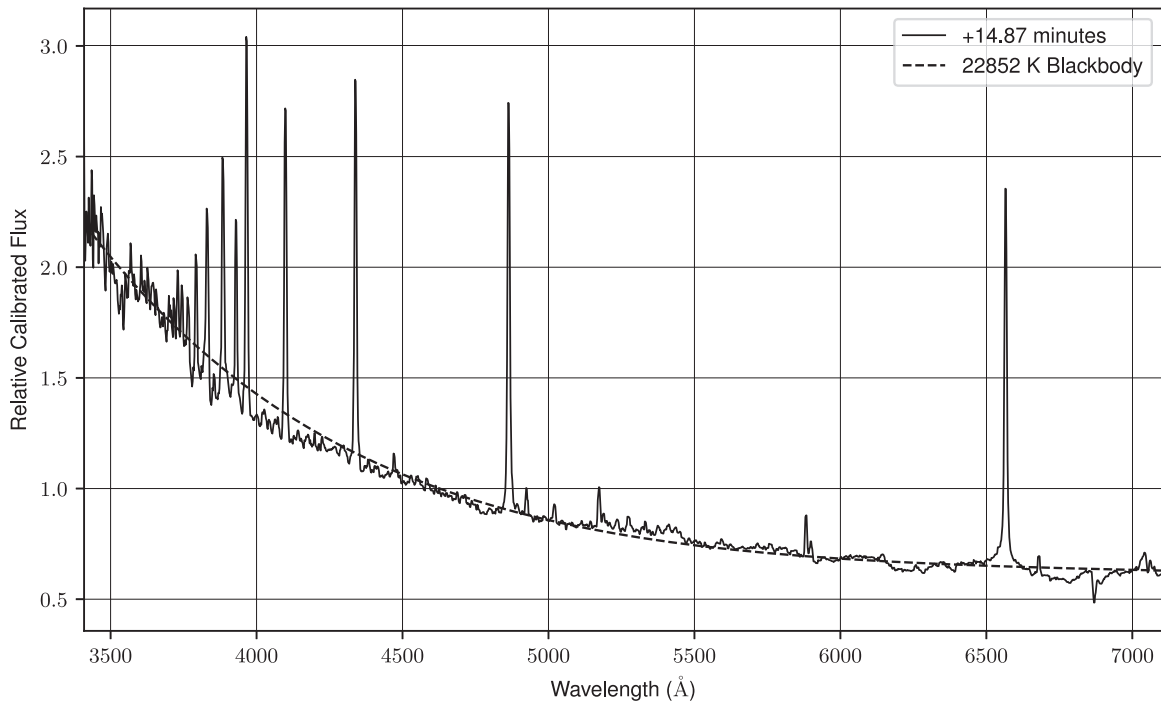
Figure 18. Spectrum of EVRT-3586872 at +14.87 minutes from the end of the first Evryscope detection of a flare candidate. The spectrum has been flux-calibrated and normalized. The overlaid curve is a two-component blackbody spectrum, consisting of a fitted 22,852 K flare continuum and a 3256 K quiescent thermal spectrum.

on the structure and orbit of the reflector. Glints have been mistaken for astrophysical events, due to their short durations, relative to their motions on the sky, and their sharp contrasts with their associated streaks (Maley 1987; Schaefer et al. 1987; Maley 1991; Rast 1991; Shamir & Nemiroff 2006). During the first six months of EFTE operations, we identified 1,415,722 likely satellite glints and modeled an all-sky event rate of 1800^{+600}_{-280} sky^{-1} hr^{-1}, peaking at $m_g = 13.0$ (Corbett et al. 2020). This rate is orders of magnitude higher than the combined rate of the public alerts from all the active all-sky fast-timescale transient searches, including neutrino, gravitational-wave, gamma-ray, and radio observatories. A subsequent study, using the Weizmann Fast Astronomical Survey Telescope (Nir et al. 2021), revealed that this event rate increases sharply with depth, reporting an event rate of 9100^{+3000}_{-2000} sky^{-1} hour^{-1} for $9 < M_{BP} < 11$ around the equator, 2.3 times the value that we measured for $m_g < 9$. As the majority of the events observed by Evryscope appear to be generated at LEO, we expect the event rate for satellite glints to correlate with the rapidly growing number of LEO satellites.[12]

7. Summary and Conclusions

In this paper, we have presented EFTE, the real-time transient discovery pipeline for the Evryscopes. The pipeline is a fully custom data analysis tool, which is suited to the unique parameter space inhabited by the Evryscopes and capable of identifying transient candidates in real time, with alerts available for each image within the 2 minute cadence of the Evryscopes for 98.5% of images. To accomplish this, we have reduced the complex image subtraction process that has been adopted by seeing-limited surveys to a simple direct subtraction of near-consecutive images. The astrometric performance for transient alerts is subpixel at the 99th

percentile, and the photometric performance is within 0.06 mag rms of the ATLAS-REFCAT2 catalog for reference stars within the $8 < m_g < 14.5$ sensitivity range of the survey. Using a convolutional RB classifier, we are able to recover 99.9% of sources brighter than $m_g = 13.2$, with a false-positive rate of 5.1%.

While EFTE is specifically adapted to Evryscope data, the infrastructure, algorithms, and ML models were developed to enable portability to instruments with similar survey strategies, such as TESS (Ricker et al. 2014), or those with stringent data throughput and latency reduction requirements. The core algorithms from EFTE have been adapted for usage in the pipeline of the Argus Optical Array, a 5 m class multiplexed 55 GPix synoptic survey instrument that is currently in development (Law et al. 2021, 2022). Argus will observe a field of view equivalent to that of Evryscope in alternating 1 and 30 s cadences, which will produce up to 4.3 PiB and 145 TiB of raw data per night, respectively. To support this data rate, the Argus data will be reduced in real time, producing low-latency transient alerts, photometry, and calibrated image data for distribution and storage. In Corbett et al. (2022a), we describe the Argus pipeline and data products, and demonstrate the direct subtraction algorithm— which is described in Section 3.3—on the Argus Array Technology Demonstrator (Corbett et al. 2022b). The VETNET model described in Section 4 is optimized for direct image subtraction and Evryscope data; however, the framework for performing this optimization and for staged model training using both on-sky and simulated data is similarly portable to Argus.

A public alert stream from EFTE is in development, based on the evolving community standard, adopted by the Zwicky Alert Distribution System (Patterson et al. 2019) and planned for the Rubin Observatory's Legacy Survey of Space and Time,[13] with serialized alert packets being distributed via Apache Kafka


(Kreps et al. 2011). Alerts will be available via the Arizona-NOIRLab Temporal Analysis and Response to Events System (Matheson et al. 2021). Details of the alert distribution system and the alert schema contents will be addressed in future work.

ORCID iDs

Hank Corbett ⓘ https://orcid.org/0000-0002-6339-6706
Jonathan Carney ⓘ https://orcid.org/0000-0001-8544-584X
Ramses Gonzalez ⓘ https://orcid.org/0000-0001-5083-8272
Octavi Fors ⓘ https://orcid.org/0000-0002-4227-9308
Nathan Galliher ⓘ https://orcid.org/0000-0001-8105-1042
Amy Glazier ⓘ https://orcid.org/0000-0001-9981-4909
Ward S. Howard ⓘ https://orcid.org/0000-0002-0583-0949
Nicholas M. Law ⓘ https://orcid.org/0000-0001-9380-6457
Robert Quimby ⓘ https://orcid.org/0000-0001-9171-5236
Jeffrey K. Ratzloff ⓘ https://orcid.org/0000-0001-8791-7388

References

Abbott, B. P., Abbott, R., Adhikari, R., et al. 2009, RPPh, 72, 076901
Abadi, M., Agarwal, A., Barham, P., et al. 2015, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, https://www.tensorflow.org/
Agarap, A. F. 2018, arXiv:1803.08375
Aizawa, M., Kawana, K., Kashiyama, K., et al. 2022, PASJ, 74, 1069
Alard, C., & Lupton, R. H. 1998, ApJ, 503, 325
Aldering, G., Adam, G., Antilogus, P., et al. 2002, Proc. SPIE, 4836, 61
Anders, F., Khalatyan, A., Queiroz, A. B. A., et al. 2022, A&A, 658, A91
Anderson, J., & King, I. R. 2000, PASP, 112, 1360
Andreoni, I., Cooke, J., Webb, S., et al. 2020a, MNRAS, 491, 5852
Andreoni, I., Jacobs, C., Hegarty, S., et al. 2017, PASA, 34, e037
Arcavi, I. 2022, ApJ, 937, 75
Arimatsu, K., Tsumura, K., Usui, F., Ootsubo, T., & Watanabe, J.-i. 2021, AJ, 161, 135
Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
Bailey, S., Aragon, C., Romano, R., et al. 2008, AN, 329, 292
Barbary, K. 2016, JOSS, 1, 58
Bäuerle, A., van Onzenoodt, C., & Ropinski, T. 2021, in IEEE Transactions on Visualization and Computer Graphics, Vol. 27 (Piscataway, NJ: IEEE), 2980
Becker, A. 2015, HOTPANTS: High Order Transform of PSF ANd Template Subtraction, Astrophysics Source Code Library, ascl:1504.004
Behnel, S., Bradshaw, R., Citro, C., et al. 2011, CSE, 13, 31
Bellm, E. C., Kulkarni, S. R., Graham, M. J., et al. 2019, PASP, 131, 018002
Berger, E., Leibler, C. N., Chornock, R., et al. 2013, ApJ, 779, 18
Beroiz, M., Cabral, J., & Sanchez, B. 2020, A&C, 32, 100384
Bersten, M. C., Folatelli, G., García, F., et al. 2018, Natur, 554, 497
Bertin, E. 2011, in ASP Conf. Ser. 442, Astronomical Data Analysis Software and Systems XX, ed. I. N. Evans et al. (San Francisco, CA: ASP), 435
Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393
Bhat, P. N., Meegan, C. A., Lichti, G. G., et al. 2009, in AIP Conf. Ser. 1133, Gamma-Ray Burst: Sixth Huntsville Symp., ed. C. Meegan, C. Kouveliotou, & N. Gehrels (Melville, NY: AIP), 34
Bloom, J. S., Starr, D. L., Butler, N. R., et al. 2008, AN, 329, 284
Bramich, D. 2008, MNRAS, 386, L77
Brink, H., Richards, J. W., Poznanski, D., et al. 2013, MNRAS, 435, 1047
Burke, C. J., Aleo, P. D., Chen, Y.-C., et al. 2019, MNRAS, 490, 3952
Butler, H., Daly, M., Doyle, A., et al. 2016, The GeoJSON Format, 2070-1721, Internet Engineering Task Force
Calabretta, M. R., Valdes, F., Greisen, E. W., & Allen, S. L. 2004, in ASP Conf. Ser. 314, Astronomical Data Analysis Software and Systems (ADASS) XIII, ed. F. Ochsenbein, M. G. Allen, & D. Egret (San Francisco, CA: ASP), 551
Cao, Y., Nugent, P. E., & Kasliwal, M. M. 2016, PASP, 128, 114502
Chollet, F., et al. 2015, Keras, GitHub, https://github.com/fchollet/keras
Clemens, J. C., Crain, J. A., & Anderson, R. 2004, Proc. SPIE, 5492, 331
Corbett, H., Law, N. M., Soto, A. V., et al. 2020, ApJL, 903, L27
Corbett, H., Soto, A. V., Machia, L., et al. 2022a, Proc. SPIE, 12189, 1218910
Corbett, H., Soto, A. V., Machia, L., et al. 2022b, Proc. SPIE, 12182, 121824D
Corbett, H., Vasquez Soto, A., Machia, L., et al. 2022, Proc. SPIE, 12189, 1218910
Cucchiara, A., Cenko, S. B., Bloom, J. S., et al. 2011, ApJ, 743, 154
Dálya, G., Galgóczi, G., Dobos, L., et al. 2018, MNRAS, 479, 2374
Dark Energy Survey Collaboration, Abbott, T., Abdalla, F. B., et al. 2016, MNRAS, 460, 1270
Dieleman, S., Willett, K. W., & Dambre, J. 2015, MNRAS, 450, 1441
Drake, A. J., Djorgovski, S. G., Mahabal, A., et al. 2009, ApJ, 696, 870
Duev, D. A., Mahabal, A., Masci, F. J., et al. 2019, MNRAS, 489, 3582
Dyer, M. J., Dhillon, V. S., Littlefair, S., et al. 2018, Proc. SPIE, 10704, 107040C
Fischler, M., & Bolles, R. 1981, Communications of the ACM, 24, 381
Flewelling, H. A., Magnier, E. A., Chambers, K. C., et al. 2020, ApJS, 251, 7
Förster, F., Maureira, J. C., Martín, J. S., et al. 2016, ApJ, 832, 155
Fox, D. W., Price, P. A., Soderberg, A. M., et al. 2003, ApJL, 586, L5
Fresneau, A., & Osborn, W. H. 2009, A&A, 503, 1023
Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, A&A, 616, A1
Gal, Y., & Ghahramani, Z. 2015a, arXiv:1506.02142
Gal, Y., & Ghahramani, Z. 2015b, arXiv:1506.02158
Garnavich, P. M., Tucker, B. E., Rest, A., et al. 2016, ApJ, 820, 23
Ghosh, A., Manwani, N., & Sastry, P. S. 2016, arXiv:1605.06296
Goldstein, D. A., D'Andrea, C. B., Fischer, J. A., et al. 2015, AJ, 150, 82
Górski, K. M., & Hivon, E. 2011, HEALPix: Hierarchical Equal Area isoLatitude Pixelization of a sphere, Astrophysics Source Code Library, record ascl:1107.018
Hamuy, M., Suntzeff, N. B., Heathcote, S. R., et al. 1994, PASP, 106, 566
Hamuy, M., Walker, A. R., Suntzeff, N. B., et al. 1992, PASP, 104, 533
Hedges, C., Luger, R., Martinez-Palomera, J., Dotson, J., & Barentsen, G. 2021, AJ, 162, 107
Heinze, A. N., Tonry, J. L., Denneau, L., et al. 2018, AJ, 156, 241
Henden, A. A., Templeton, M., Terrell, D., et al. 2016, yCat, 2336, 0
Ho, A. Y. Q., Kulkarni, S. R., Nugent, P. E., et al. 2018, ApJL, 854, L13
Hoffleit, D., & Jaschek, C. 1995, yCat, 5050, 0
Høg, E., Fabricius, C., Makarov, V. V., et al. 2000, A&A, 355, L27
Howard, W. S., Corbett, H., Law, N. M., et al. 2020, ApJ, 902, 115
Howard, W. S., & MacGregor, M. A. 2022, ApJ, 926, 204
Howard, W. S., Teske, J., Corbett, H., et al. 2021, AJ, 162, 147
Howard, W. S., Tilley, M. A., Corbett, H., et al. 2018, ApJL, 860, L30
Hu, L., Wang, L., Chen, X., & Yang, J. 2022, ApJ, 936, 157
Ioffe, S., & Szegedy, C. 2015, arXiv:1502.03167
Jayasinghe, T., Kochanek, C. S., Stanek, K. Z., et al. 2018, MNRAS, 477, 3145
Kaiser, N., Burgett, W., Chambers, K., et al. 2010, Proc. SPIE, 7733, 77330E
Keller, S. C., Schmidt, B. P., Bessell, M. S., et al. 2007, PASA, 24, 1

Killestein, T. L., Lyman, J., Steeghs, D., et al. 2021, MNRAS, 503, 4838
Kingma, D. P., & Ba, J. 2014, arXiv:1412.6980
Koposov, S., & Bartunov, O. 2006, in ASP Conf. Ser. 351, Astronomical Data Analysis Software and Systems XV, ed. C. Gabriel et al. (San Francisco, CA: ASP), 735
Koposov, S., & Bartunov, O. 2019, Q3C: A PostgreSQL Package for Spatial Queries and Cross-matches of Large Astronomical Catalogs, Astrophysics Source Code Library, ascl:1905.008
Kowalski, A. F., Hawley, S. L., Wisniewski, J. P., et al. 2013, ApJS, 207, 15
Kreps, J., Narkhede, N., & Rao, J. 2011, Kafka: a Distributed Messaging System for Log Processing, in NetDB Workshop '11, 1
Kulkarni, S. R., & Rau, A. 2006, ApJL, 644, L63
Kumar, M. 1988, MarGe, 12, 117
Kumar, S., Gezari, S., Heinis, S., et al. 2015, ApJ, 802, 27
Lang, D., Hogg, D. W., Mierle, K., Blanton, M., & Roweis, S. 2010, AJ, 139, 1782
Larson, S., Beshore, E., Hill, R., et al. 2003, AAS/DPS Meeting, 35, 36.04
Law, N. M., Corbett, H., Galliher, N. W., et al. 2021, PASP, 134, 035003
Law, N. M., Fors, O., Ratzloff, J., et al. 2015, PASP, 127, 234
Law, N. M., Kulkarni, S. R., Dekany, R. G., et al. 2009, PASP, 121, 1395
Law, N., Vasquez Soto, A., Corbett, H., et al. 2022, Proc. SPIE, 12182, 121824H
Le Folgoc, L., Baltatzis, V., Desai, S., et al. 2021, arXiv:2110.04286
LeCun, Y., & Bengio, Y. 1995, in The Handbook of Brain Theory and Neural Networks, ed. M. A. Arbib (Cambridge, MA: MIT Press)
LeCun, Y., Boser, B., Denker, J. S., et al. 1989, Neural Computation, 1, 541
Li, L., Jamieson, K., DeSalvo, G., Rostamizadeh, A., & Talwalkar, A. 2018, Journal of Machine Learning Research, 18, 1
Lipunov, V. M., Krylov, A. V., Kornilov, V. G., et al. 2004, AN, 325, 580
Mahabal, A., Rebbapragada, U., Walters, R., et al. 2019, PASP, 131, 038002
Makhlouf, K., Turpin, D., Corre, D., et al. 2022, A&A, 664, A81
Maley, P. 1991, AdSpR, 11, 33
Maley, P. D. 1987, ApJ, 317, L39
Martin-Carrillo, A., Hanlon, L., Topinka, M., et al. 2014, A&A, 567, A84
Matheson, T., Stubens, C., Wolf, N., et al. 2021, AJ, 161, 107
McCulloch, W., & Pitts, W. 1943, Bulletin of Mathematical Biophysics, 5, 127
Nir, G., Ofek, E. O., Ben-Ami, S., et al. 2021, MNRAS, 505, 2477
Osten, R. A., & Wolk, S. J. 2015, ApJ, 809, 79
Patterson, M. T., Bellm, E. C., Rusholme, B., et al. 2019, PASP, 131, 018001

Perrett, K., Balam, D., Sullivan, M., et al. 2010, AJ, 140, 518
Pickles, A., & Depagne, É. 2010, PASP, 122, 1437
Pietras, M., Falewicz, R., Siarkowski, M., Bicz, K., & Preś, P. 2022, ApJ, 935, 143
Powell, B. P., Kostov, V. B., Rappaport, S. A., et al. 2021, AJ, 162, 299
Quimby, R. M., Shafter, A. W., & Corbett, H. 2021, RNAAS, 5, 160
Ranjan, S., Wordsworth, R., & Sasselov, D. D. 2017, ApJ, 843, 110
Rast, R. H. 1991, Icar, 90, 328
Ratzloff, J. K., Law, N. M., Corbett, H. T., Fors, O., & del Ser, D. 2020, JATIS, 6, 018002
Ratzloff, J. K., Law, N. M., Fors, O., et al. 2019, PASP, 131, 075001
Richmond, M. W., Tanaka, M., Morokuma, T., et al. 2020, PASJ, 72, 3
Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2014, Proc. SPIE, 9143, 914320
Rolnick, D., Veit, A., Belongie, S., & Shavit, N. 2017, arXiv:1705.10694
Schaefer, B. E., Barber, M., Brooks, J. J., et al. 1987, ApJ, 320, 398
Schaefer, B. E., King, J. R., & Deliyannis, C. P. 2000, ApJ, 529, 1026
Segura, A., Walkowicz, L. M., Meadows, V., Kasting, J., & Hawley, S. 2010, AsBio, 10, 751
Shamir, L., & Nemiroff, R. J. 2006, PASP, 118, 1180
Shappee, B. J., Prieto, J. L., Grupe, D., et al. 2014, ApJ, 788, 48
Simonyan, K., & Zisserman, A. 2014, arXiv:1409.1556
Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. 2014, Journal of Machine Learning Research, 15, 1929
Stoppa, F., Vreeswijk, P., Bloemen, S., et al. 2022, A&A, 662, A109
Tamuz, O., Mazeh, T., & Zucker, S. 2005, MNRAS, 356, 1466
Tonry, J. L., Denneau, L., Flewelling, H., et al. 2018, ApJ, 867, 105
Tonry, J. L., Denneau, L., Heinze, A. N., et al. 2018, PASP, 130, 064505
Troja, E., Lipunov, V. M., Mundell, C. G., et al. 2017, Natur, 547, 425
van Roestel, J., Groot, P. J., Kupfer, T., et al. 2019, MNRAS, 484, 4507
Vestrand, W. T., Wren, J. A., Panaitescu, A., et al. 2014, Sci, 343, 38
Voges, W., Aschenbach, B., Boller, T., et al. 1999, A&A, 349, 389
Walkowicz, L. M., Johns-Krull, C. M., & Hawley, S. L. 2008, ApJ, 677, 593
Wang, J., Xin, L. P., Li, H. L., et al. 2021, ApJ, 916, 92
Watson, C., Henden, A. A., & Price, A. 2015, yCat, 02027, 0
Wee, J., Blagorodnova, N., Penprase, B. E., et al. 2020, ApJ, 899, 162
Wolf, C., Onken, C. A., Luvaul, L. C., et al. 2018, PASA, 35, e010
Xin, L. P., Li, H. L., Wang, J., et al. 2021, ApJ, 909, 106
Zackay, B., Ofek, E. O., & Gal-Yam, A. 2016, ApJ, 830, 27