Deep learning for the Zwicky Transient Facility (ZTF): real/bogus classification and identification of fast-moving objects

Dmitry A. Duev

Research Scientist | Astro Dept | Caltech duev@caltech.edu

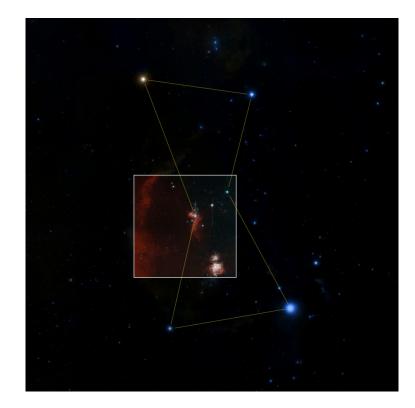
On behalf of the Caltech/IPAC ZTF team

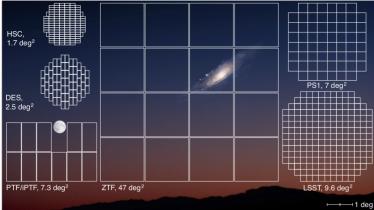


ZTF at a glance

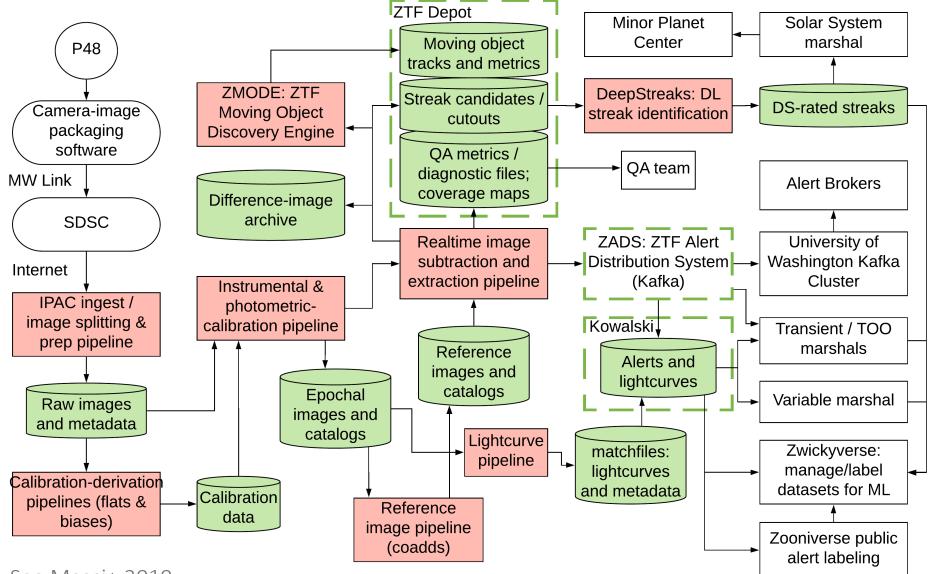
- Telescope: Samuel Oschin 48-inch Schmidt, Palomar Observatory
- Active detector area: ~ 47 deg²
- Areal survey rate: 4300 deg² / hour
- Nominal survey duration: 3 years
- Filters: g / r / i
- Nominal exposure time: 30 sec
- Single exposure depth (5σ): 20.8 / 20.6 / 19.9
- Image quality (FWHM): 2.1" / 2.0" / 2.1"
- CCDs: 16 x 6k x 6k, 1.0" / pix
- Survey entire Northern visible sky to $\delta \simeq -30^{\circ}$
- A fast, wide-area time-domain survey:
 - fast, young, and rare flux transients
 - counterparts to gravitational wave sources
 - low-z Type Ia SNe for cosmology
 - variable stars & eclipsing binaries
 - Solar System objects

https://ztf.caltech.edu





ZTF data/processing flow



See Masci+ 2019

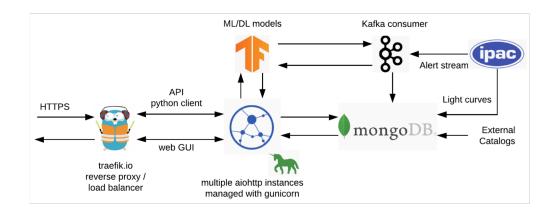
Data volumes and source stats

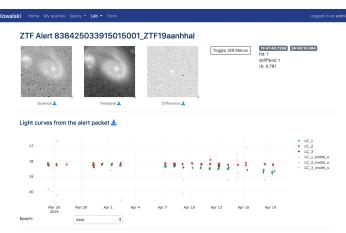
Single night	8h40m	Nominal survey	3 x 26
# on-sky exposures	~700	Volume of data products	~3.2
Raw image data	~1 TB	Volume of ref images	~60 -
Real-time data products	~4 TB	# CCD quad ref images	~2.8 x
# unvetted 5σ alerts	~10 ⁵ - 10 ⁶	Volume of matchfiles	~50 1
# ML-vetted alerts	~10 ³ - 10 ⁵	# matchfiles	~2.8 x
# unvetted streaks	~10 ⁴ - 10 ⁶	<pre># single-epoch PSF-fit source measurements</pre>	~800
# ML-vetted streaks	~10 ² - 10 ³	# single-epoch aperture source measurements	~230

DL -> genuine astrophysical events/objects

Kowalski: store+access ZTF alerts/LC

- "Swiss knife" system for time domain astronomy (with ZTF)
 - ..
 - API, python client
 - 115M+ ZTF alerts, 2.5B+ ZTF light curves
- ZTF Alert Lab
 - Filter and display/access alerts
 - Packet and compound LC, corrected for reference flux if necessary





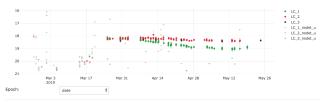
LC3: (telescope: 'PO-12m,' instrument: 'ZFF, 'filer', ' comment' 'no corrections applied, using raw magst, sigmapst, and diffmaglim') LC3: (telescope: 'PO-12m,' instrument': ZFF, 'filer', ' comment' to corrections applied, using raw magst, sigmapst, and diffmaglim') LC3: (telescope: 'PO-12m,' instrument': ZFF, 'filer', 3, comment': 'no corrections applied, using raw magst, sigmapst, and diffmaglim')

Note Using ?match_radius_arcsec=1.5&star_galaxy_threshold=0.4

Alert packet contents 去



Composite light curves for ZTF19aanhhal 去



LG:1: (felescope': PO:1.2m, 'instrument': ZTF, 'iller': 1, 'comment': 'no corrections applied, using raw magpst, sigmapst, and diffmaglim') LG:2: (felescope': PO:1.2m, 'instrument': ZTF, 'iller': 2, 'comment': 'no corrections applied, using raw magpst, sigmapst, and diffmaglim') LG:3: (felescope': PO:1.2m, 'instrument': ZTF, 'iller': 3, 'comment': 'no corrections applied, using raw magpst, sigmapst, and diffmaglim')

Note Using ?match_radius_arcsec=1.5&star_galaxy_threshold=0.4

© ZTF 2018 🗘

https://github.com/dmitryduev/kowalski

Duev+ in prep.

Zwickyverse: manage/label datasets for ML

- Efficiently label large amounts of image data
- Collaborative
- API, python client to get data in/out

r ickyverse Home Projects Docs	
+ New project	
DeepStreaks: rb (bogus vs real) edit remove	
id: 5b96af9c0354c9000b0aea36 Separating streaks from bogus detections in ZTF data Project metadata 🛓	
Classes	
Users	
Datasets	
DeepStreaks: sl (long vs short) edit remove id: 5b99b2c6aec3c500103a14de Once separated from bogus detections, streaks are classified as long/short Project metadata	
Classes	

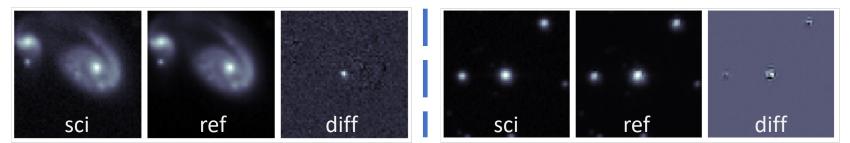
Zwickyverse Home Projects Docs	Logged in as admin	Zwickyverse Home Projects Docs Logged in as admin *
Dataset: i_band.20190412.batch1 600: rb<0.4 + 300: rb>0.4 Image: The transmission of transmission of the transmission of the transmission of transmissi		DeepStreaks: rb (bogus vs real) edit remove id: 5b96af9c0354c9000b0aea36 Separating streaks from bogus detections in ZTF data
Object	bogus r	Project metadata 🛓
sci ref zogy fit residual		Classes Users Datasets Long streaks classify inspect edit remove id: 5b96b0670354c9000b0eea37 Long streaks from QZ
Stowing 1 to 100 of 300 rows 100 ~ rows per page	(123)	id: 5b98ec/05ec848000c70a870 Mostly bogus cutouts. Imported September 10, 2018

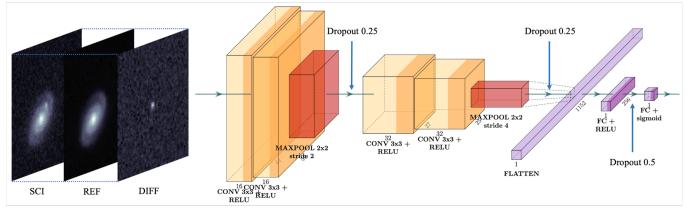
https://github.com/dmitryduev/zwickyverse

Duev+ in prep.

braai: real/bogus classifier for ZTF

- braai: bogus/real adversarial artificial intelligence
- CNN-based architecture ("VGG6"), TensorFlow/Keras
- Input: stacked triplets [science, reference, ZOGY]
- 30K+ training examples; Zwickyverse for labeling; trained on GPU; active learning
- Data sources: Kowalski; ZTF transient and variable marshals, Zooniverse

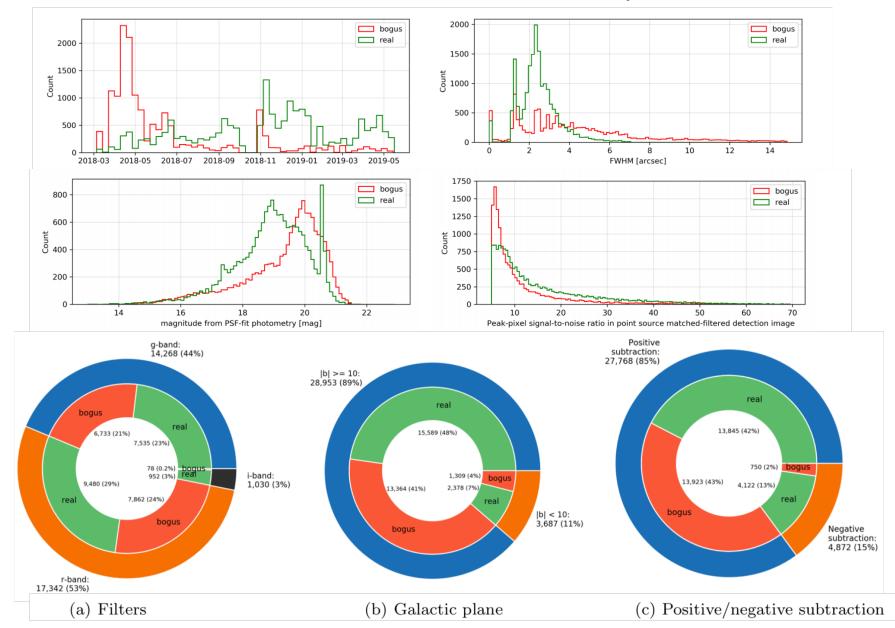




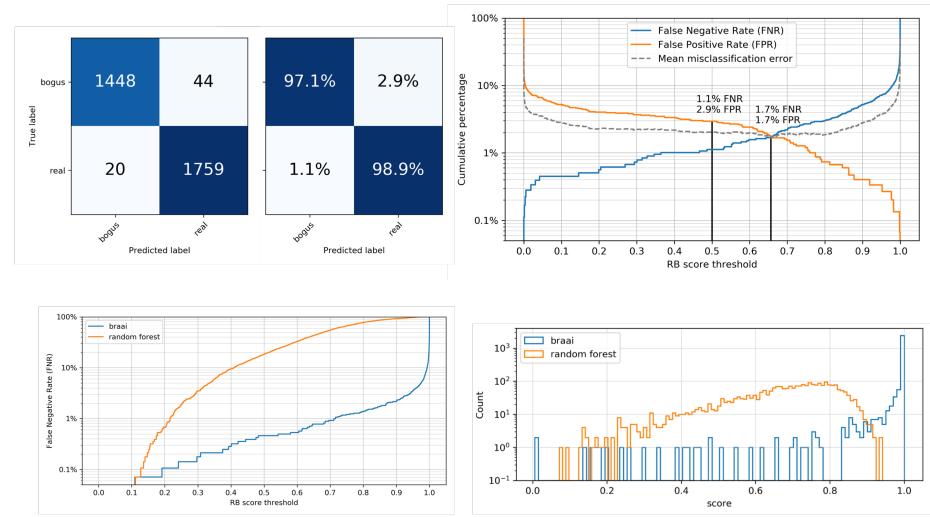
https://github.com/dmitryduev/braai

Duev+ in prep.

braai: data diversity

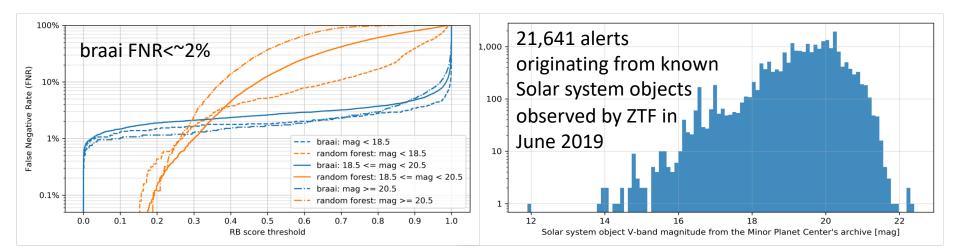


braai: performance



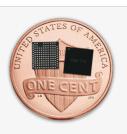
2,820 alerts from a set of 140 SN detected in 2019. 0.5% braai FNR vs 20% RF RB FNR 2,633 ZTF alerts from the night of May 14, 2019 from 921 (vetted) real objects on GROWTH marshal. 0.7% braai FNR vs 11% RF RB FNR

braai: performance



Google's Coral Edge TPUs:

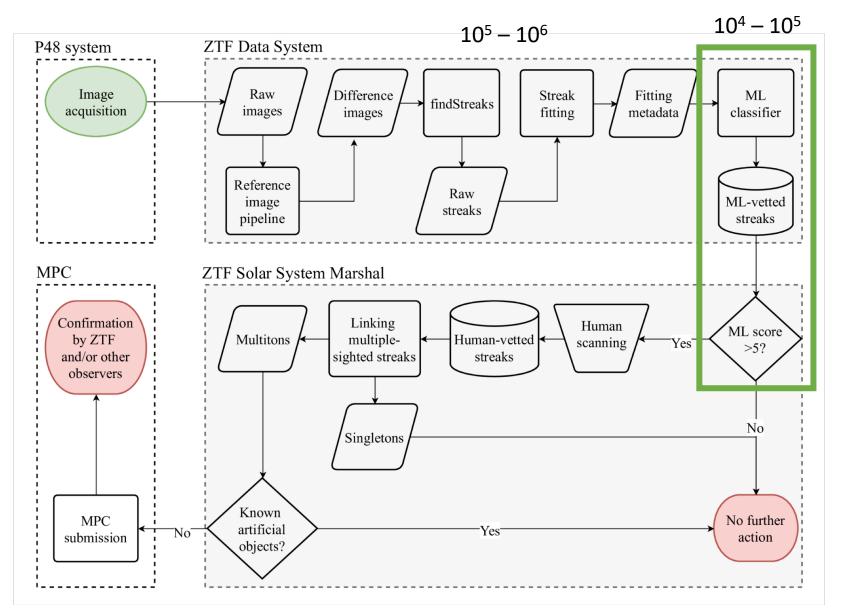
- Compiled (uint8) model -> same performance
- 3-5 minutes (including I/O) to process 200k alerts, ~50x beefy multi-core desktop
- ~\$100-150







ZTF Solar System framework

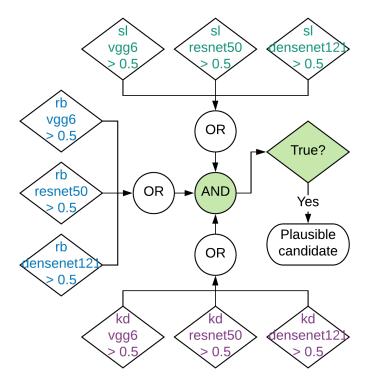


DeepStreaks: architecture

Convolutional-neural-network, deep-learning system designed to efficiently identify streaking FMOs/NEOs

"rb": bogus or real streak?
 Identify all streak-like objects,
 including actual streaks from
 FMOs, long streaks from
 satellites, and cosmic rays

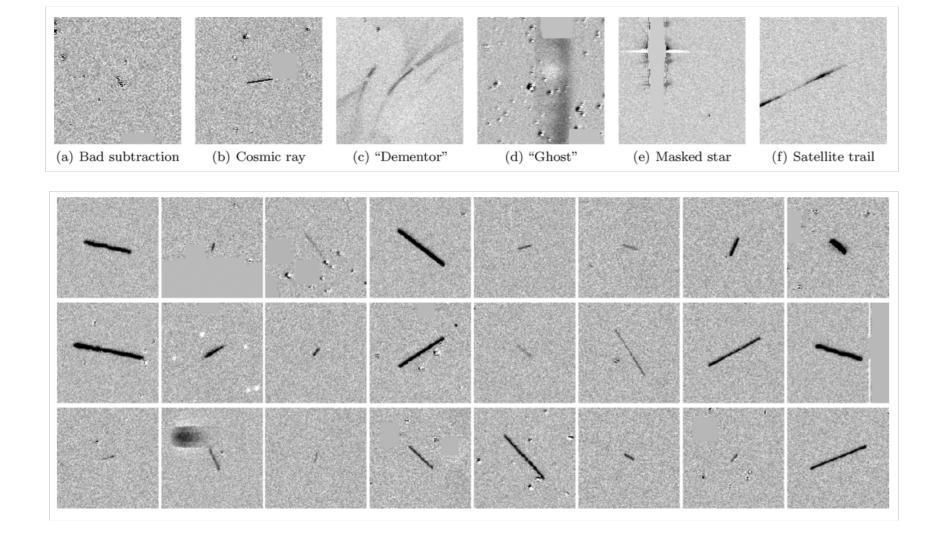
- "sl": long or short streak?
- "kd": ditch or keep? Is this a real streak, or a cosmic ray/some other artifacts?



 Three different CNN architectures within each family: VGG6, ResNet50, and DenseNet121

DeepStreaks: dataset

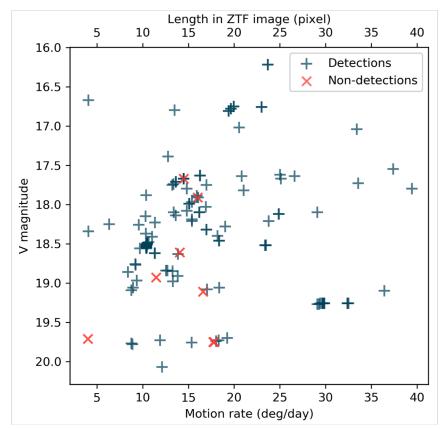
30k training examples; used Zwickyverse for labeling; trained on GPU



DeepStreaks: performance/results

- 96-98% true positive rate, depending on the night
 - Quantified by performance on test data sets and using known NEOs observed by ZTF
- Below 1% false positive rate, 50x-100x improvement over original RF classifier
 - Sanity check: 0% FPR on 8,000 random ImageNet images
- Near-real-time operations; below 10 min per day spent by human scanners vs ~hours with original RF classifier
- 60+ confirmed new NEAs
- Another 50+ "lost" due to insufficient follow-up

Duev+ 2019 (MNRAS, arXiv: 1904.05920)



Completeness identifications using known NEOs observed by ZTF in October 2018 – January 2019. Out of 210 streaks from real NEOs detected by the ZTF Streak pipeline at IPAC, 202 (96%) are correctly classified.