



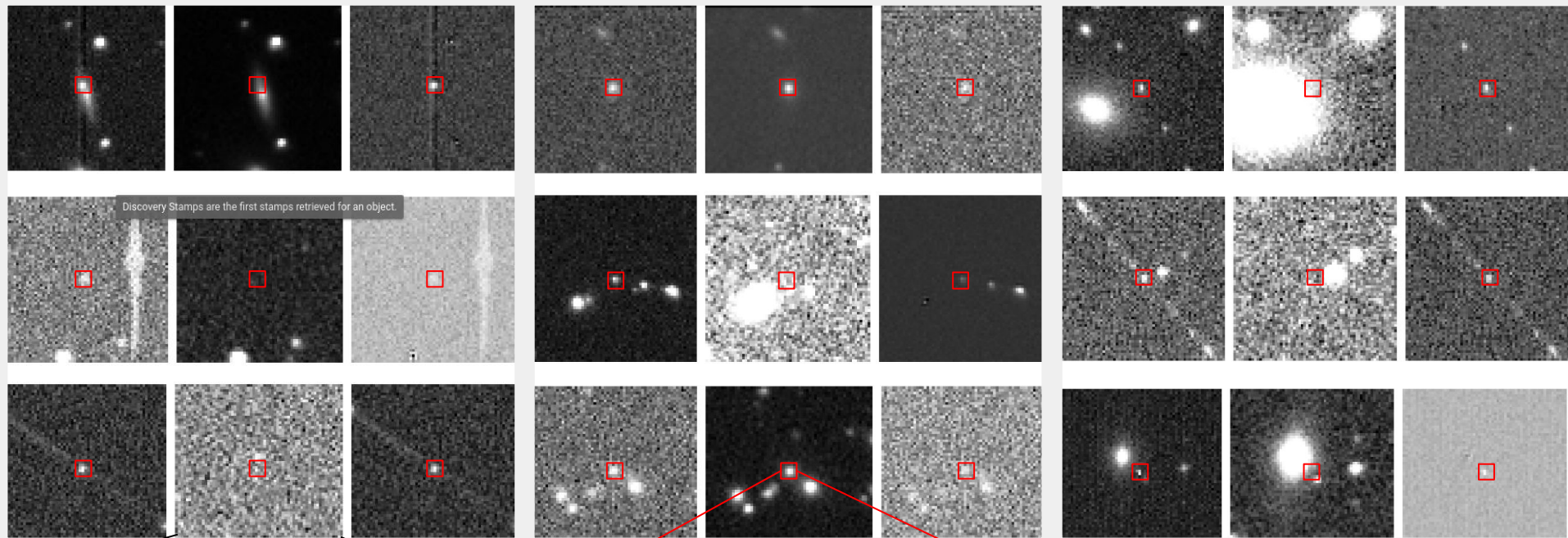
ALeRCE

Automatic Learning for the
Rapid Classification of Events

Real-time automatic host galaxy association or *Fixing the LSST image stamps*

Francisco Förster, Alejandra Muñoz, Guillermo Cabrera-Vives, Amelia Bayo, Paula Sánchez-Sáez & ALeRCE ML team + Ava Polzin, Felipe Figueroa, Dylan Britt, Sara Cuellar, Yara Yousef (LSDSS 2021) + Alex Gagliano

Would you trigger 8 m class follow-up on these ZTF candidates?



ZTF stamp size (61"x61")

LSST stamp size (6"x6")

Automatic host association

Reliable **real-time alert classification** would benefit from **real-time host galaxy identification**.

Two main ways to identify host galaxies:

1. from a **catalogue** of galaxies around the position of an object (e.g. smallest normalized distance)
2. Identifying the host galaxy directly from the **images**, e.g., visual selection or convolutional neural network (CNN)



ALeRCE's sample of host galaxies

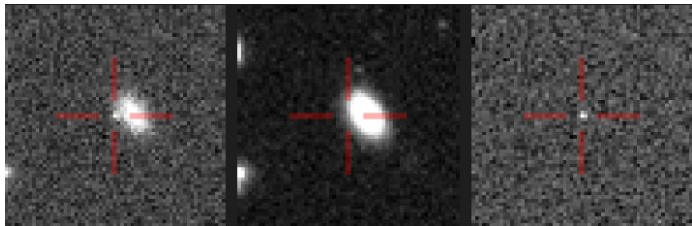
Alert ingestion: ingestion of the ZTF stream.

CNN classification: new objects are classified using their image stamps (SN, AGN, asteroid, variable star, or bogus).

Visual inspection: every new SN candidate is visually inspected and vetoed if necessary.

Galaxy selection: host galaxy of good SN candidates are visually selected.

More than 13 k host galaxies selected so far!



jupyter ALeRCE_ZTF_TNS_send_SNe Last Checkpoint: 38 minutes ago (autosaved)

File Edit View Insert Cell Kernel Navigate Widgets Help Trusted Python 3

Contents

- 1 Notebook to select 1
- 2 Start API
- 3 Load recent candidi
- 4 Select host galaxy
- 4.1 **5. Manual magic**
- 4.2 Manual magic
- 4.3 Save hosts
- 5 Show nearest scate
- 5.1 Prepare the
- 5.2 Sandbox or real
- 5.3 Get API key are
- 5.4 Build the report
- 5.5 Manual magic
- 5.6 Display report
- 5.7 Save report
- 5.8 Send to TNS
- 5.9 Show output
- 5.10 Show feedback
- 5.11 Check the full
- 6 Prepare SkyPortal r
- 6.1 Load coordina
- 6.2 Build the report
- 6.3 Save report
- 6.4 Submit to Skyth
- 6.5 Check SkyPort

Ho. Object Name RA DEC Type Velocity Redshift Redshift Flag Magnitud

21 WISEA 3002852-63-164157-7 7.21925 -16.69942 G 32168 0.1873 17.34

Saving hosts to hosts/20211106a_hosts.csv

```
http://skyserver.sdss.org/dr16/SkyServerWS/SearchTools/Search?cmd=select+top+1+objid,+z,+zerr,+photoerror+class+from+photoz+where+objid=1237673018375077902&format=csv
http://skyserver.sdss.org/dr16/SkyServerWS/SearchTools/Search?cmd=select+top+1+p.objid,+z,+zerr,+s.class,+s.warning+from+photoz+as+p+join+specobj+as+s+on+s.bestobjid=p.objid+where+p.objid=1237673018375077902&format=csv
None
```

host_name host_ra host_dec host_offset host_source host_redshift_spec host_redshift host_redshift_error host_redshift_type

aid

ZTF21acmhp WISEA 3002852-63-164157-7 7.21925 -16.69942 3.875817 NED True 0.1073 NULL

16 candidates remaining

2000 23 45 50.040 +31 14 9.79

100 3.14

/home/forster/anaconda3/lib/python3.8/site-packages/astroquery/simbad/core.py:135: UserWarning: Warning: The script line number 3 raised an error (recorded in the 'errors' attribute of the result table): '23:45:05.0577 +31:14:09.8556': No astronomical object found ; warnings.warn('Warning: The script line number %i raised %s' % (line_number, str(e)), UserWarning)

Saving hosts to hosts/20211106a_hosts.csv

None

In [8]: alerce.candidate_hosts

executed in 13ms, finished 18:38:50 2021-10-24

Out[8]:

	host_name	host_ra	host_dec	host_offset	host_source	host_redshift_spec	host_redshift	host_redshift_error	host_redshift_type
aid									
ZTF21acmhp	SDSS J000246.56-030151.9	8.943993	-3.305628	0.687802	SDSSDR16	False	0.182095	0.074397	photoz
ZTF21acmhp	SDSS J19524426-7323173	7.746119	-50.530916	False	0.038914	0.011058			photoz

https://github.com/alercebroker/TNS_upload

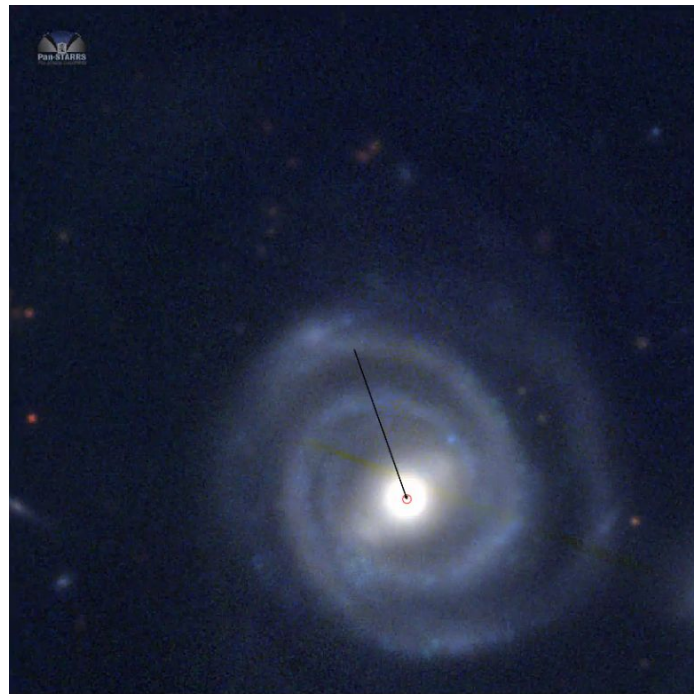
Automatic image detection

Can we **mimic human labelers**?

Training as **regression**: given an image predict a vector with the position of the host.

A **large input field of view** would be needed to account for large nearby galaxies.

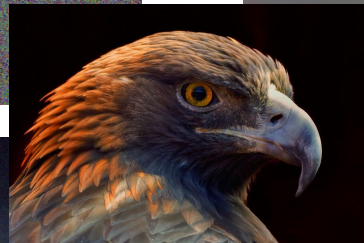
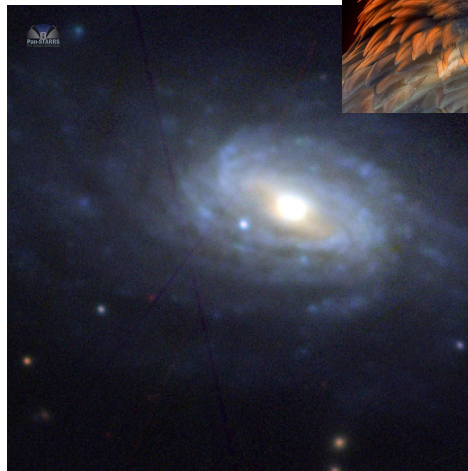
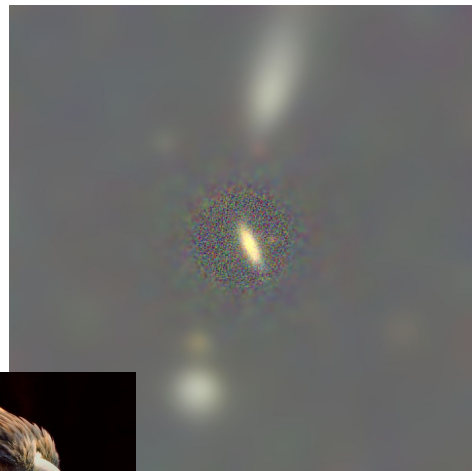
*We can solve the regression problem using **multi resolution images** centered at the alert position.*



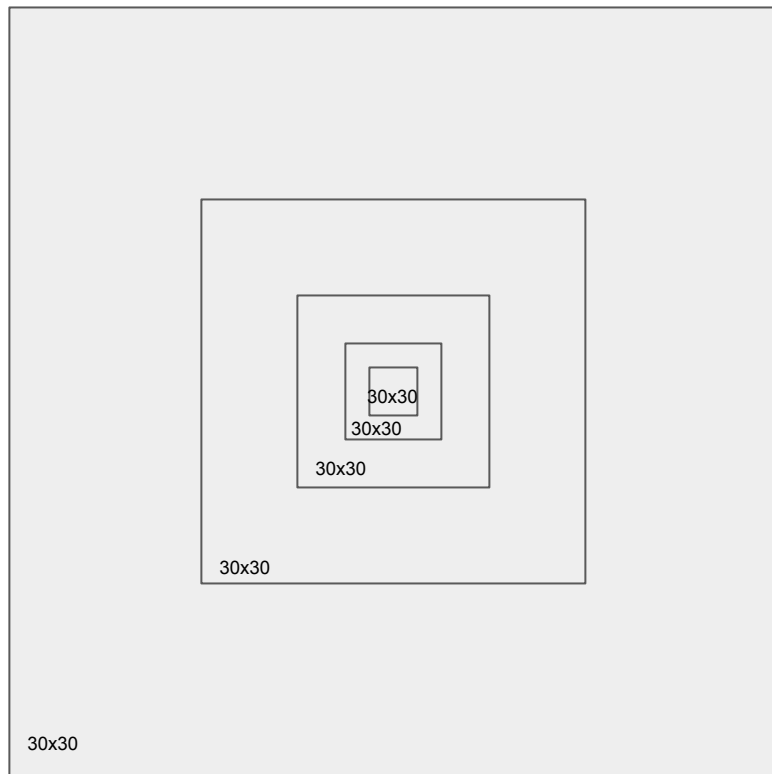
Multi resolution images

We want input images that satisfy the following requirements:

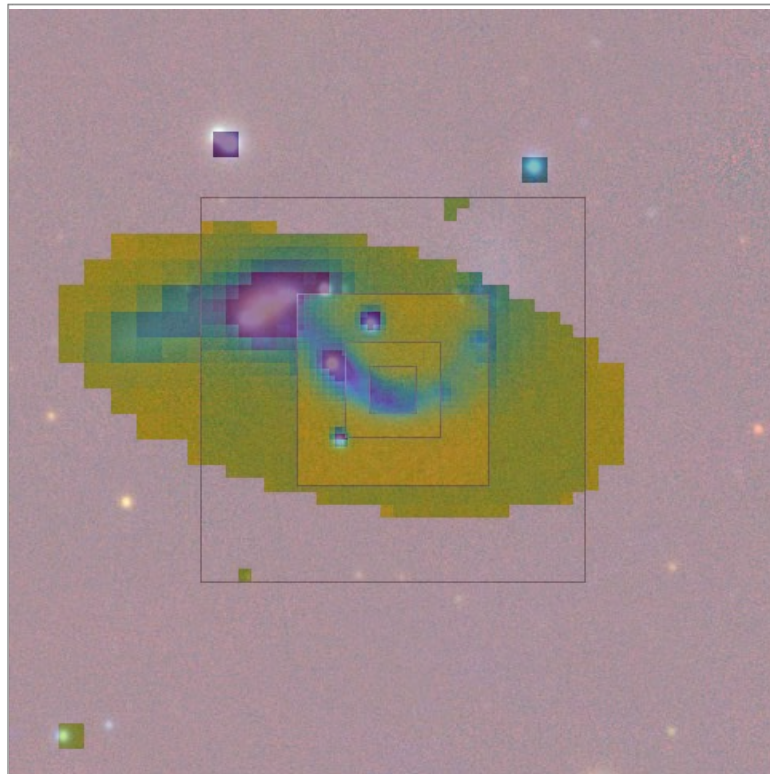
1. Span a **large area** to identify large nearby host galaxies.
2. Have **good central resolution** to identify small host galaxies.
3. Are **lightweight** to be acquired and streamed rapidly.



Multi resolution images (starting from 120"x120"=480x480=2,304,000 pixels)



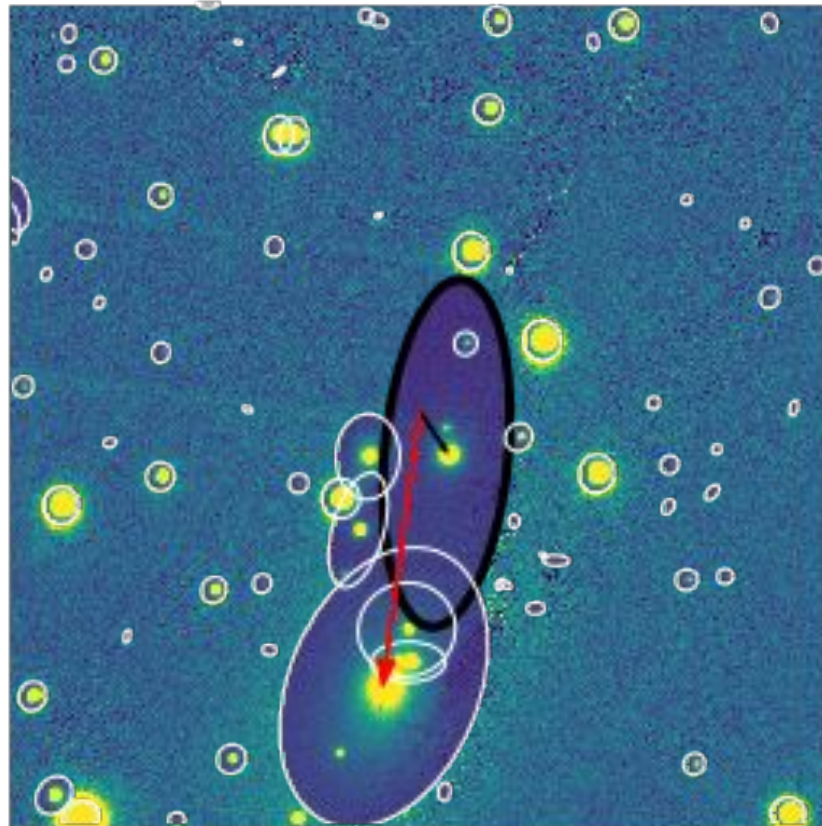
5 levels ($5 \times 30 \times 30 = 4,500$ pixels = 2.0%)



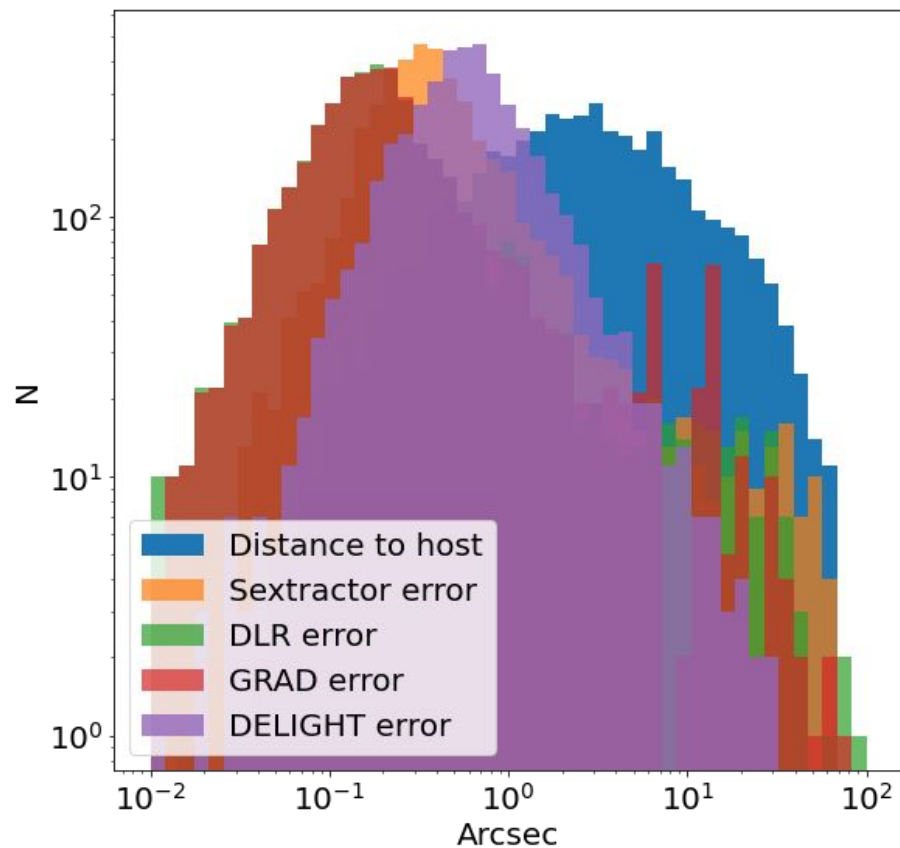
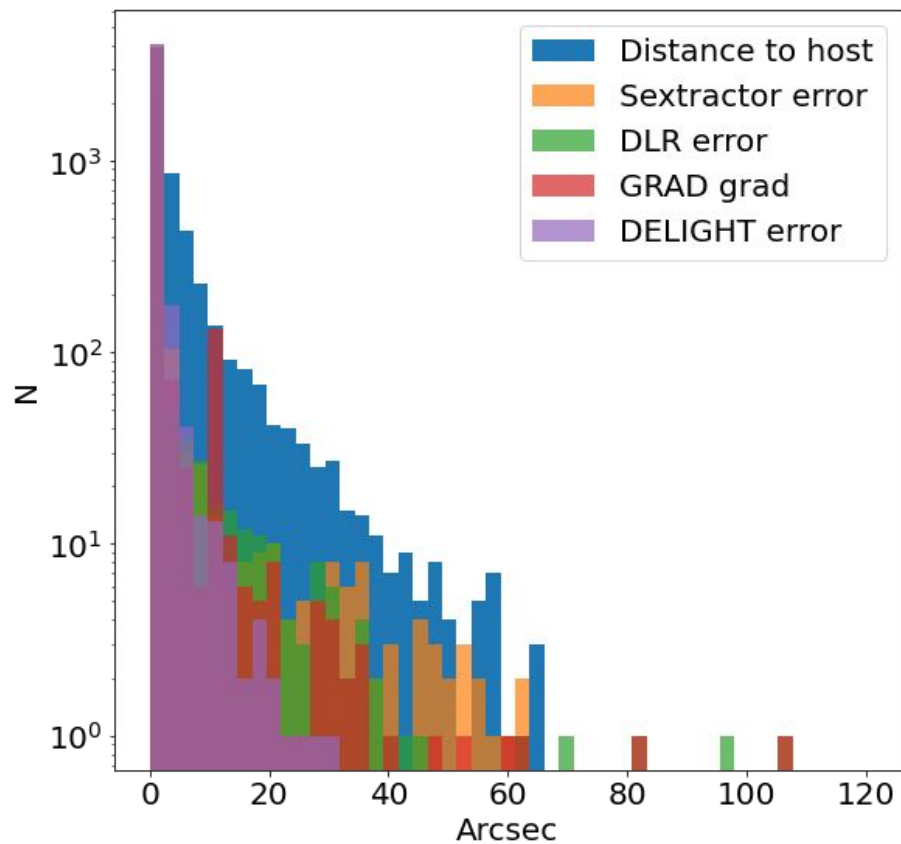
Host association methods

Compare different methods:

1. **True position:** visually selected host (from NED, Simbad or SDSS DR16).
2. **SExtractor predicted:** closest normalized distance using SExtractor ellipses.
3. **Directional light radius (DLR, Gupta+2016):** closest normalized distance using source moments.
4. **Gradient ascent (Grad, Gagliano+21):** start from source and ascend light profile.
5. **DELIGHT (Deep Learning Galaxy Host, this work):** CNN prediction based on multi resolution image.



Distance error



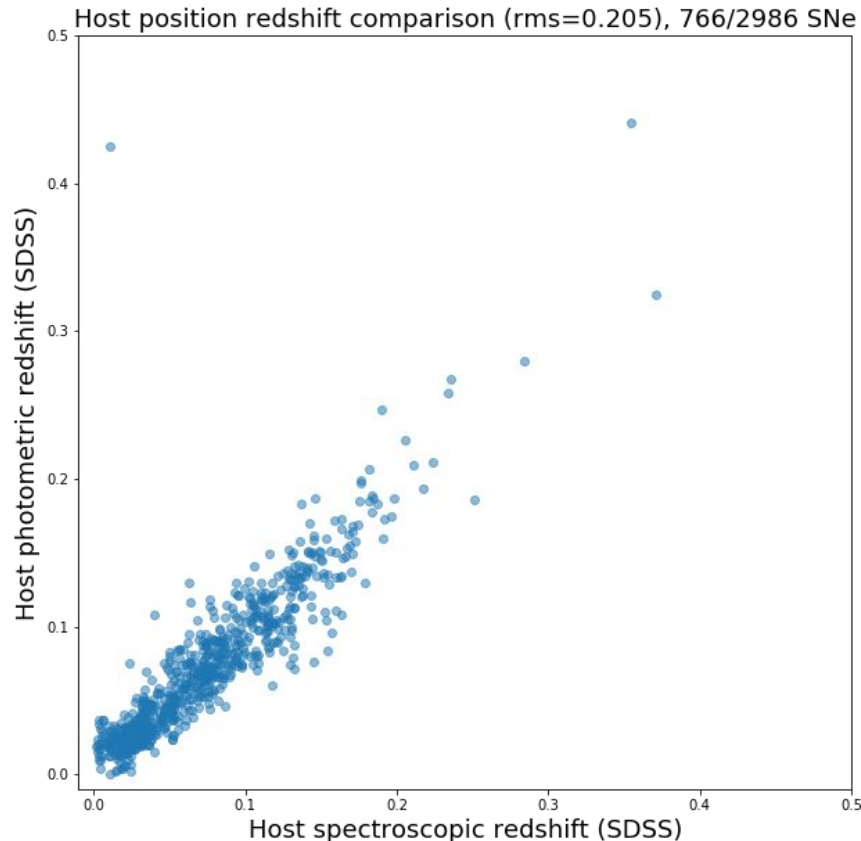
What is the redshift error?

Host galaxy properties can help with **fast classification**, in particular the **redshift**.

What are the sources of error in the redshift?

$$e(z_{\text{tran}}) = e(\text{host_phot}) + e(\text{host_id})$$

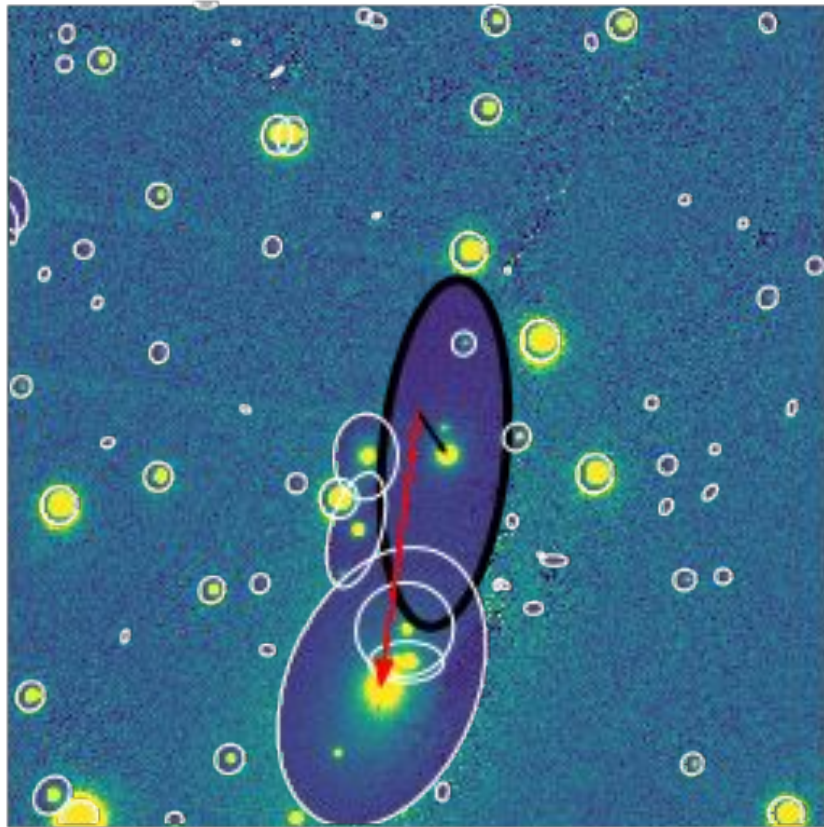
Can we estimate the distribution of the 2nd random variable?



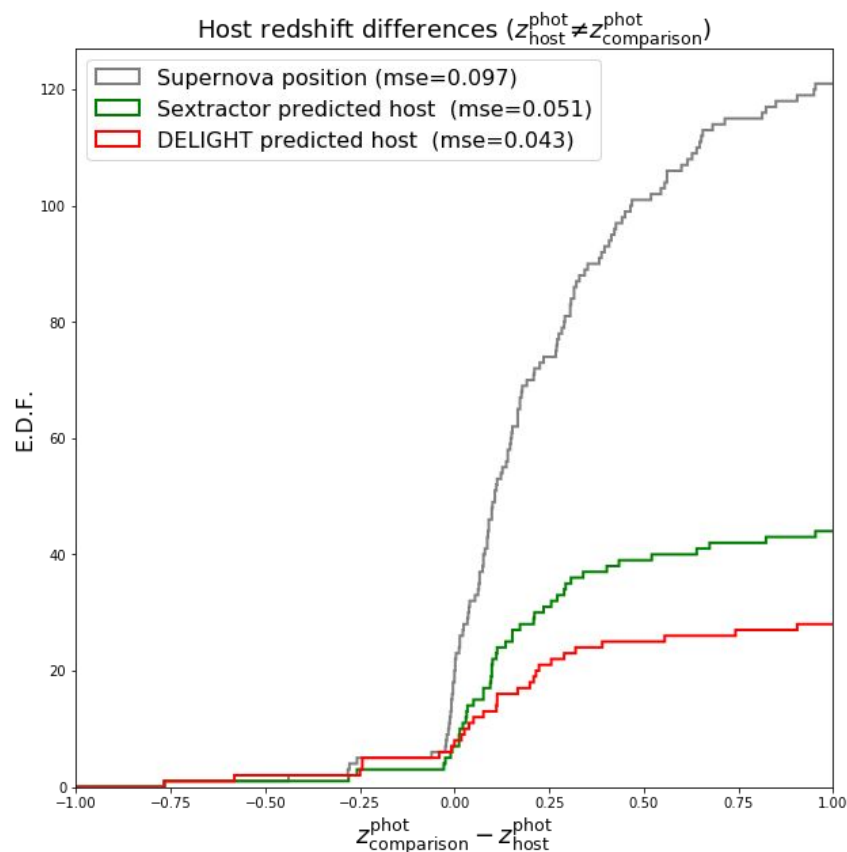
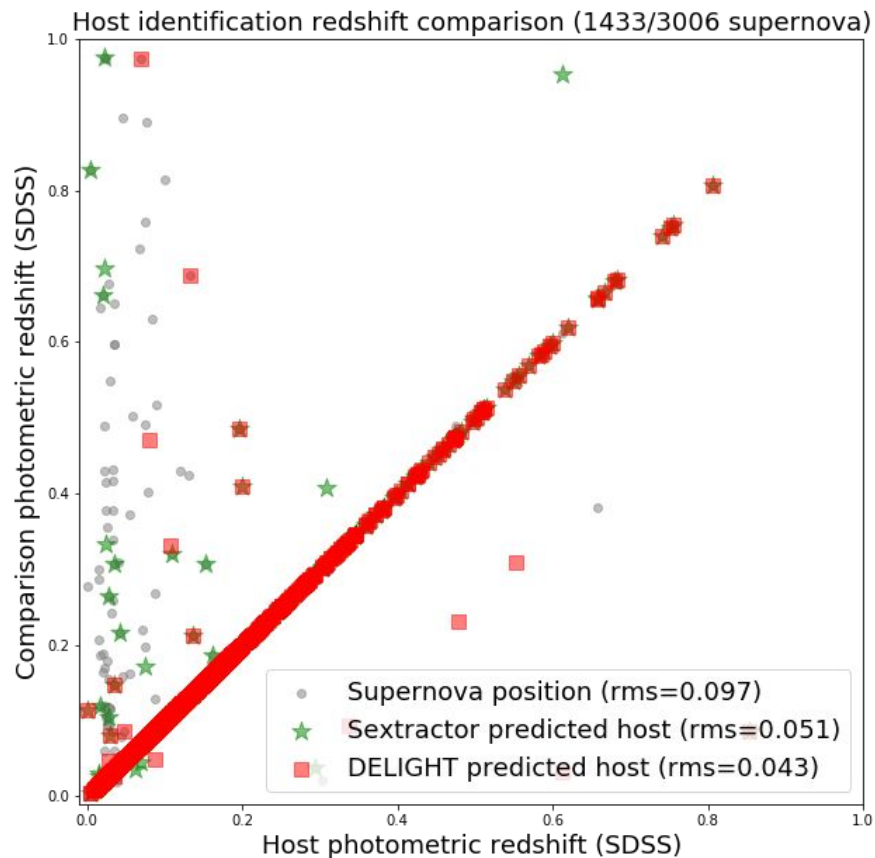
Redshift determination

Compare different methods:

1. **True redshift:** source in SDSS nearest to visually selected host position
2. **Supernova position:** source in SDSS nearest to SN position.
3. **SExtractor predicted:** source in SDSS nearest to closest normalized distance sextractor source.
4. **DELIGHT predicted:** source in SDSS nearest to CNN host prediction.

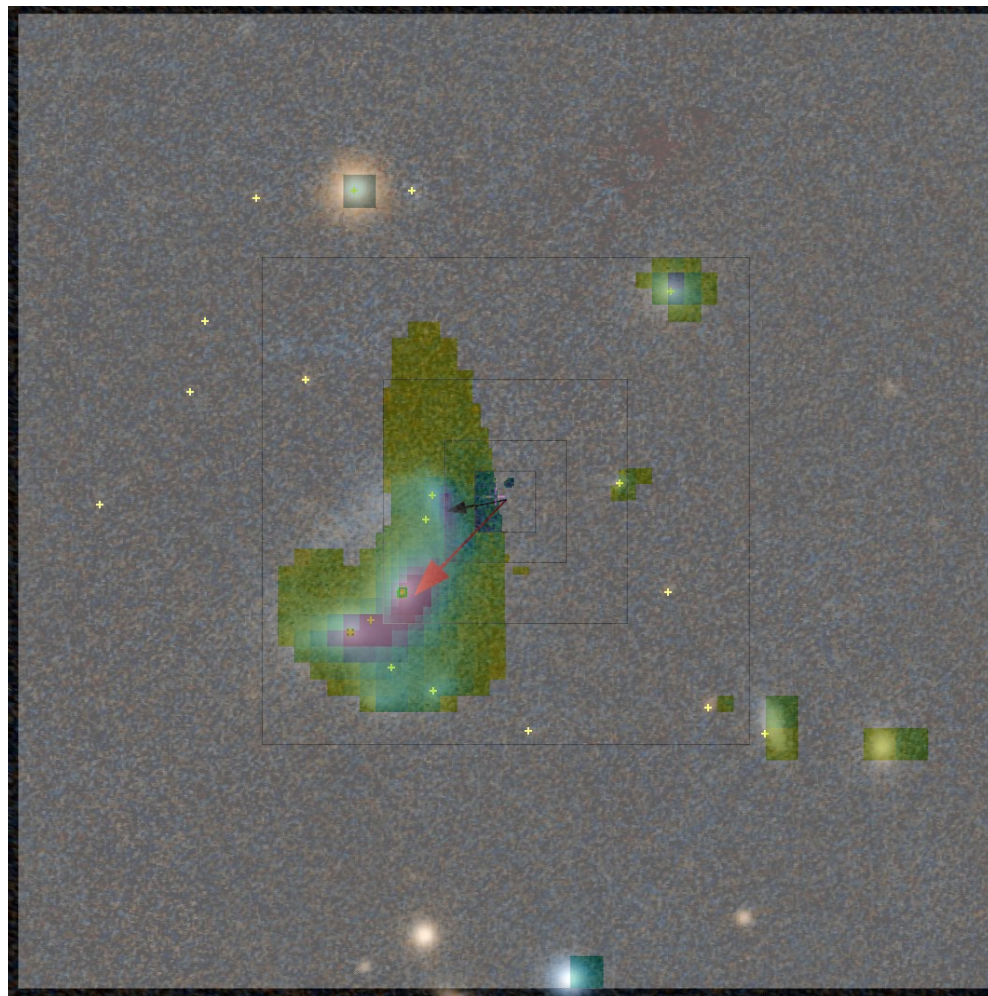


Redshift error

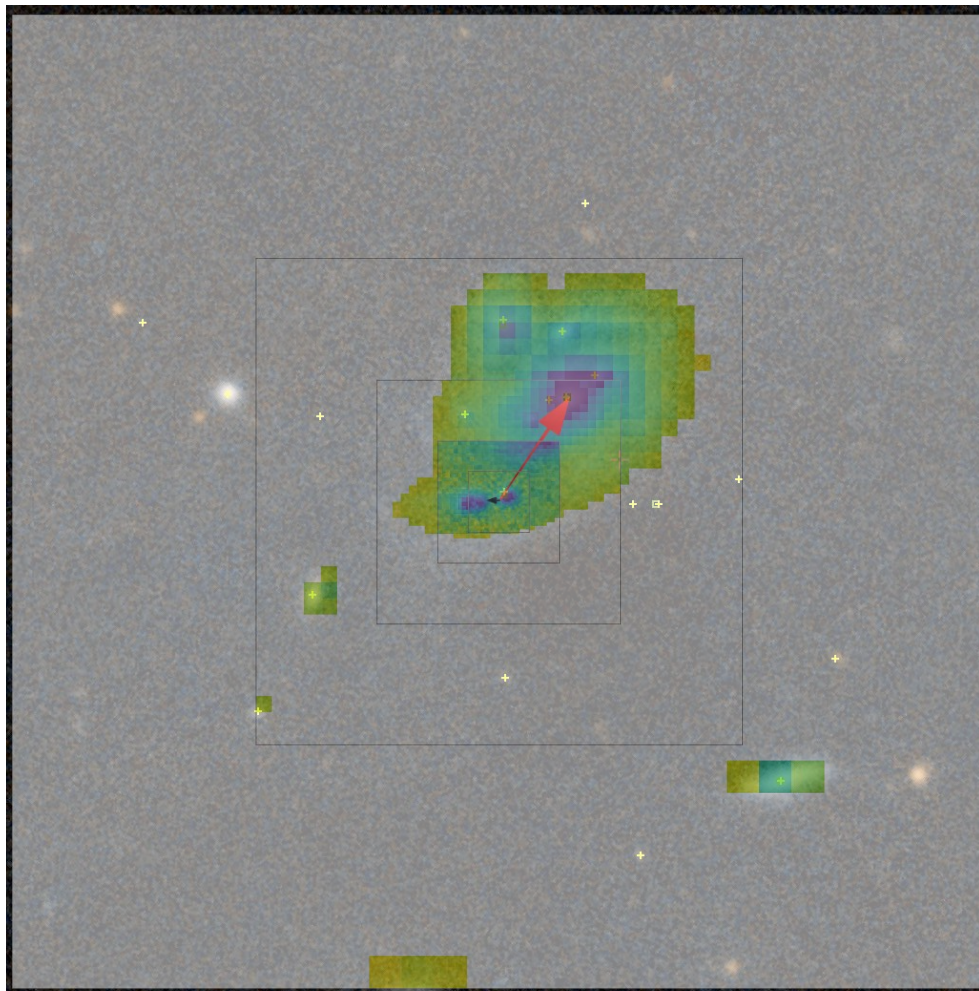


Examples

Sextractor
CNN



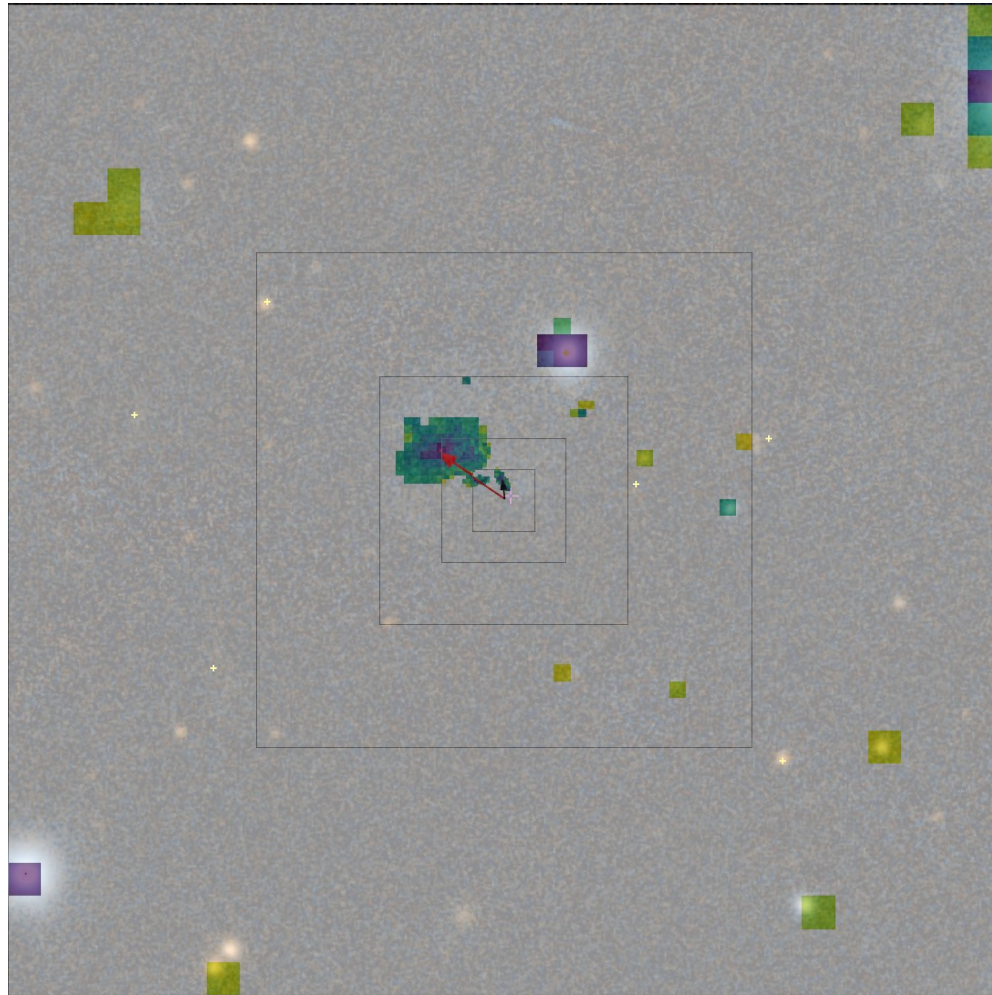
Examples



Sextractor
CNN

Examples

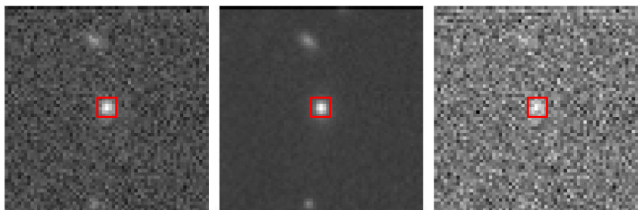
SExtractor
CNN



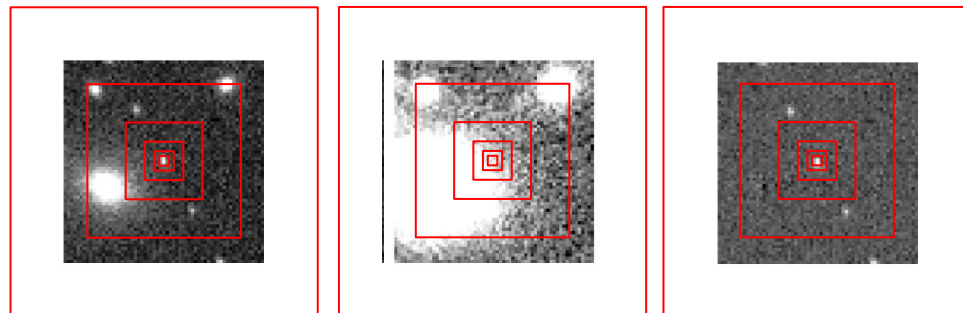
Conclusions & proposal

Multi resolution images are an efficient way of storing information. They allow for **very fast**, lightweight, and reliable **host galaxy association** using CNNs, but also for fast and reliable **alert classification** (c.f. ALeRCE stamp classifier, Carrasco-Davis+21).

LSST alert stamps are too small. We propose to represent LSST alert stamps in a hierarchical way, moving from 6" pixel FoV to $3'' \times 2^{n-1}$ FoV, with n the number of levels.



Current LSST
 $31 \times 31 = \mathbf{961 \text{ pix}}$
FOV=6"



Proposal
 $15 \times 15 * (1 + (n-1) * 3/4) \text{ pix}$
= **900 pix** for $n=5$, FOV = 48"
= **1068 pix** for $n=6$, FOV=96"

"The main challenges ahead of massive time-domain surveys are timely recognition of interesting transients in the torrent of imaging data, and maximizing the utility of the follow-up observations." (Tyson 2006)