

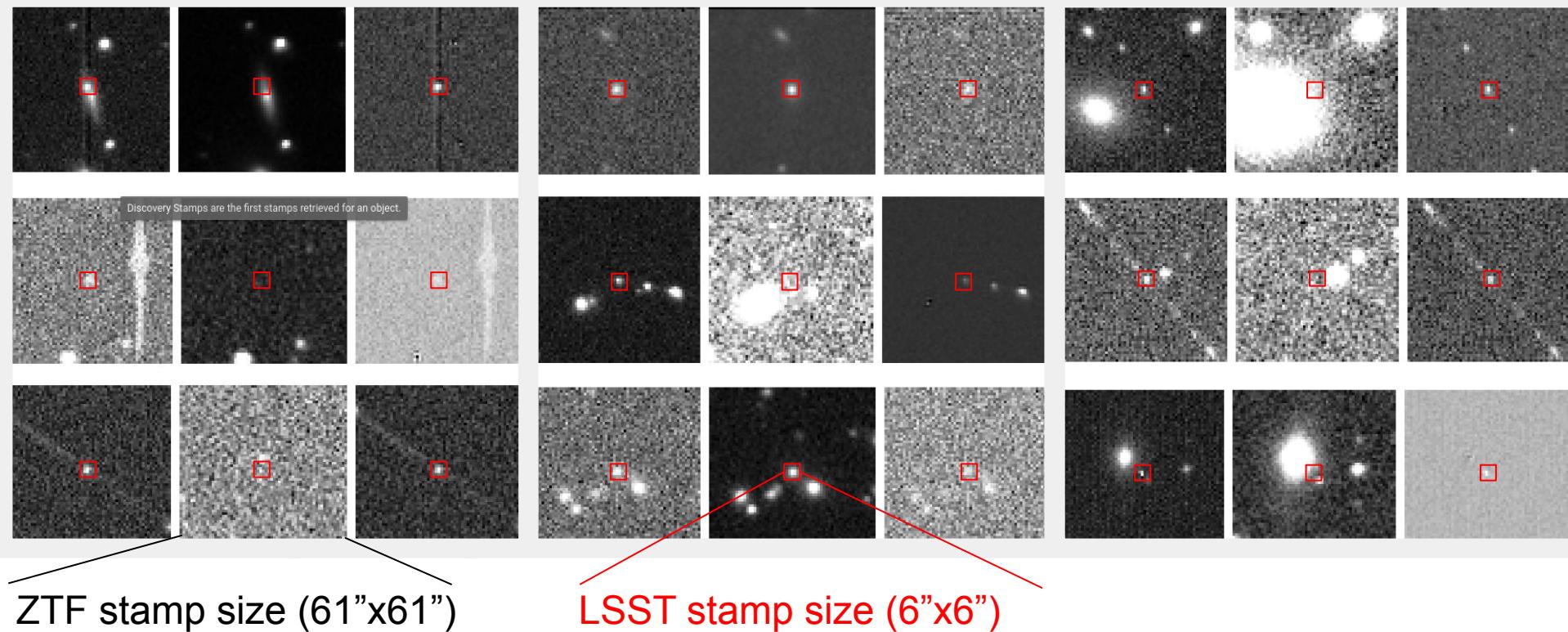


ALeRCE
Automatic Learning for the
Rapid Classification of Events

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?



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

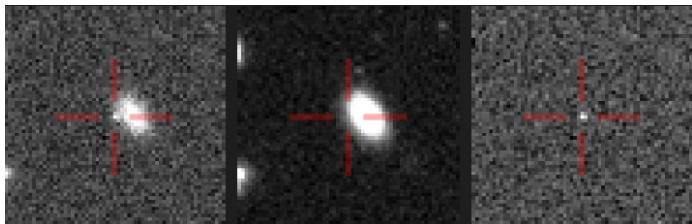
Alert ingestion: ingestion of the ZTF stream.

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_SNs Last Checkpoint: 38 minutes ago (autosaved)

File Edit View Insert Cell Kernel Navigate Widgets Help

Contents 2 o

- 1 Notebooks to select 1
- 2 Stats API
- 3 Load recent candidate
- 4 Select point galaxies
- 4.1 Do manual veto**
- 4.2 Manual moditic
- 4.3 Manual veto
- 5 Show nearest source
- 5.0.1 Prepare TN
- 5.1 Sandbox or real
- 5.2 Load the file type
- 5.3 Confirm report
- 5.4 Build the report
- 5.5 Add to moditic
- 5.6 Display report
- 5.7 Save report
- 5.8 Send to TNS
- 5.9 Send to moditic
- 5.10 Show feedback!
- 5.11 Check the folio
- 6 Preprocess**
- 6.1 Load credential
- 6.2 Build the report
- 6.3 Save report
- 6.4 Submit to SkyServer
- 6.5 Check SkyServer

Pos: 7

No.	Object Name	RA	DEC	Type	Velocity	Redshift	Redshift Flag	Magnitude	Associations: 0
1	WISEA J002052.63-164157.7	7.2105	-16.6994	G	32168	0.1073	True	17.34	

Saving hosts to hosts_20211106a.hosts.csv

<http://skyserver.cssss.org/dr18/SkyServerSQL/SearchTools/SqlSearch?cmd=select+top+1+objid,+z,+zerr,+photoerrorclass+from+photoz+where+objid=1237673018375077982&format=csv>

None

oid	host_name	host_ra	host_dec	host_offset	host_source	host_redshift_spec	host_redshift	host_redshift_error	host_redshift_type
ZTF21acmhjh	WISEA J002052.63	7.21929	-16.69942	3.875817	NED	True	0.1073	NULL	

16 candidates remaining

Photo: 3.14

/home/forster/Downloads/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): '23:45:05.8577 > 31:14:09.8556': No astronomical object found : warnings.warn("Warning: The script line number %i raised an error (%s)" % (self.line_number, self.errors[-1]))

Saving hosts to hosts_20211106a.hosts.csv

None

In [8]: aлерce.candidate_hosts

executed in 19ms, finished 16:38:55 2021-10-24

Out[8]:

oid	host_name	host_ra	host_dec	host_offset	host_source	host_redshift_spec	host_redshift	host_redshift_error	host_redshift_type
ZTF21acmhjw	SDSS J002052.63-001810.9	8.943993	-0.305529	0.687802	SDSSDR16	False	0.182095	0.074357	photz
ZTF21acmhjp	SDSS J09524406-7323173	7.749119	-73.23173	0.038914	SDSSDR16	False	0.038914	0.011058	photz

https://github.com/alercebroker/TNS_upload

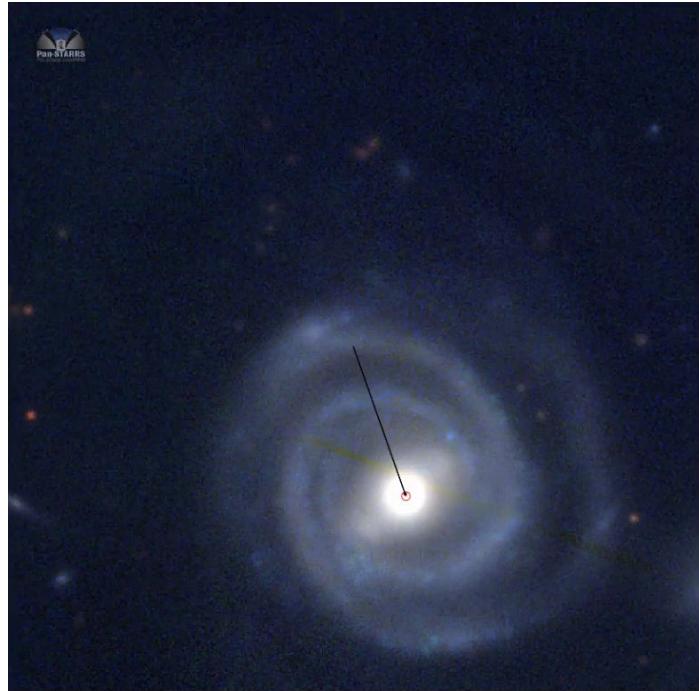
Automatic image detection

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.

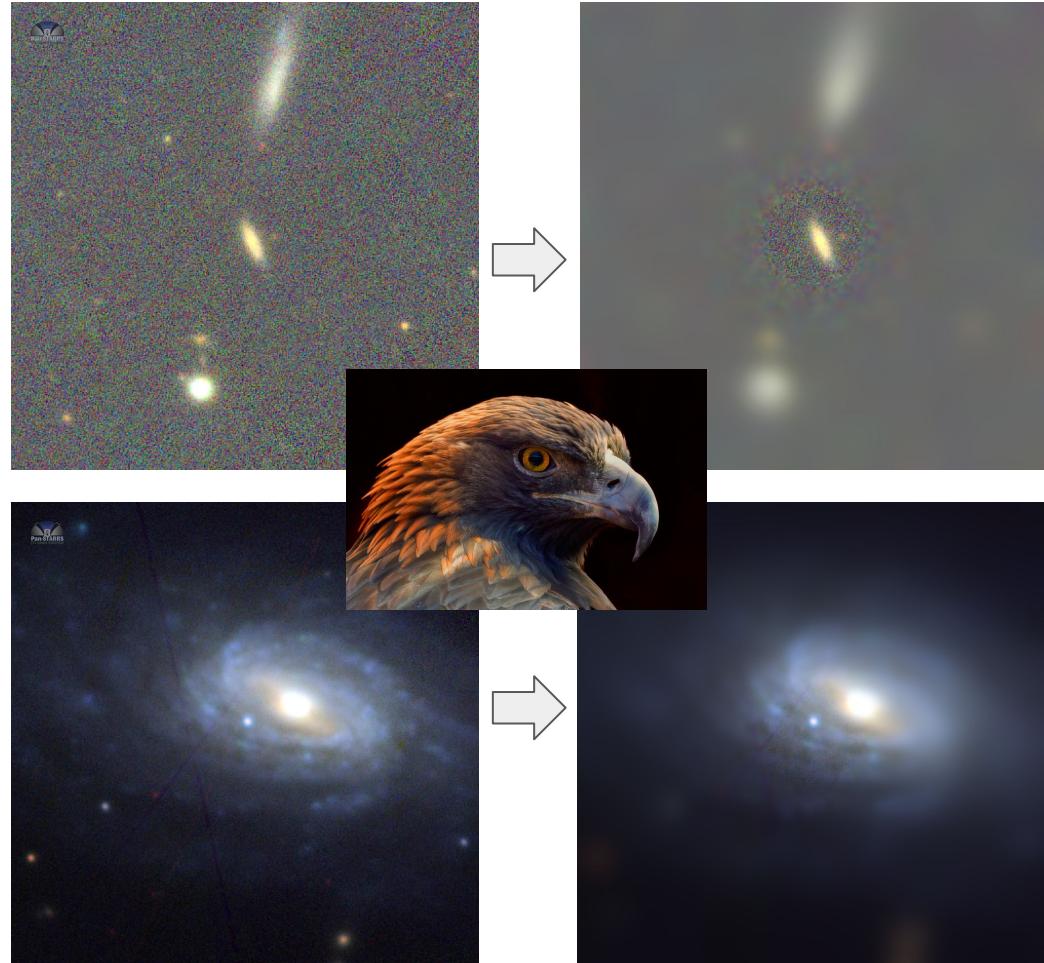
*We can solve the regression problem using **multi resolution images** centered at the alert position.*



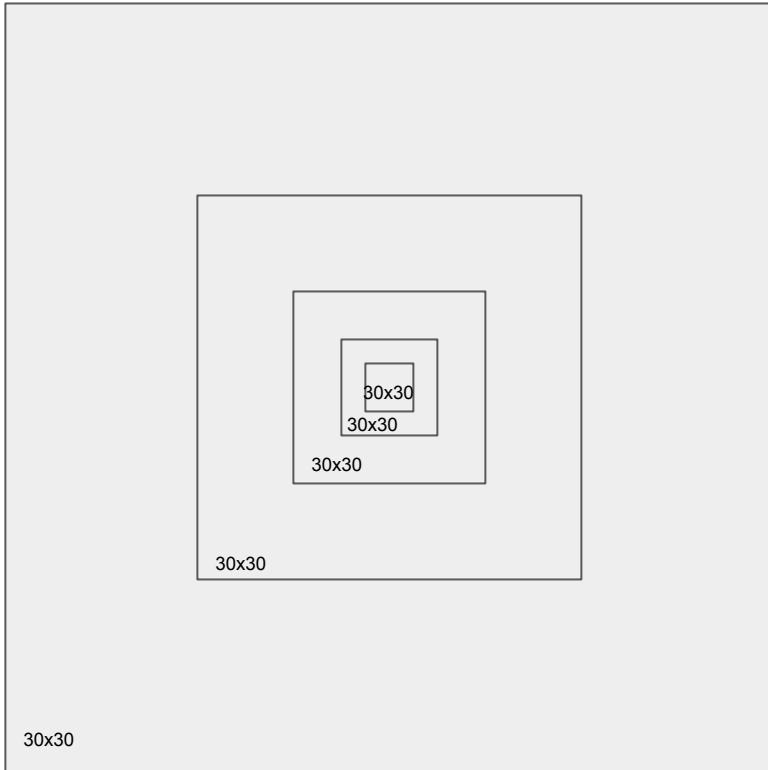
Multi resolution images

We want input images that satisfy the following requirements:

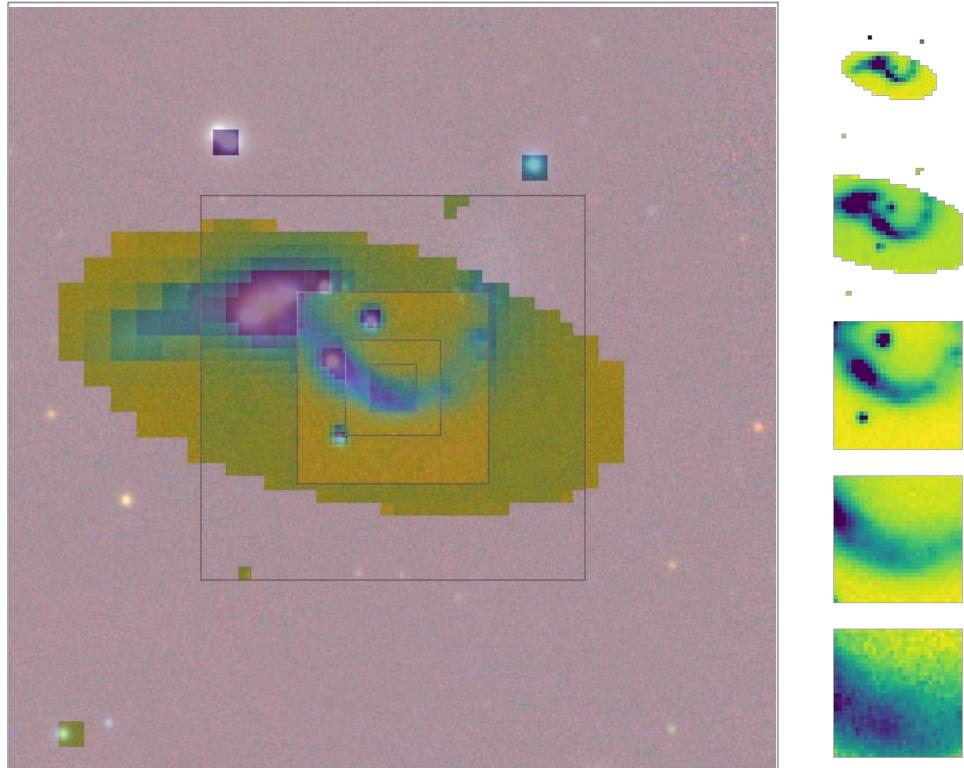
1. Span a **large area** to identify large nearby host galaxies.
2. Have **good central resolution** to identify small host galaxies.
3. Are **lightweight** to be acquired and streamed rapidly.



Multi resolution images (starting from $120'' \times 120'' = 480 \times 480 = 2,304,000$ pixels)



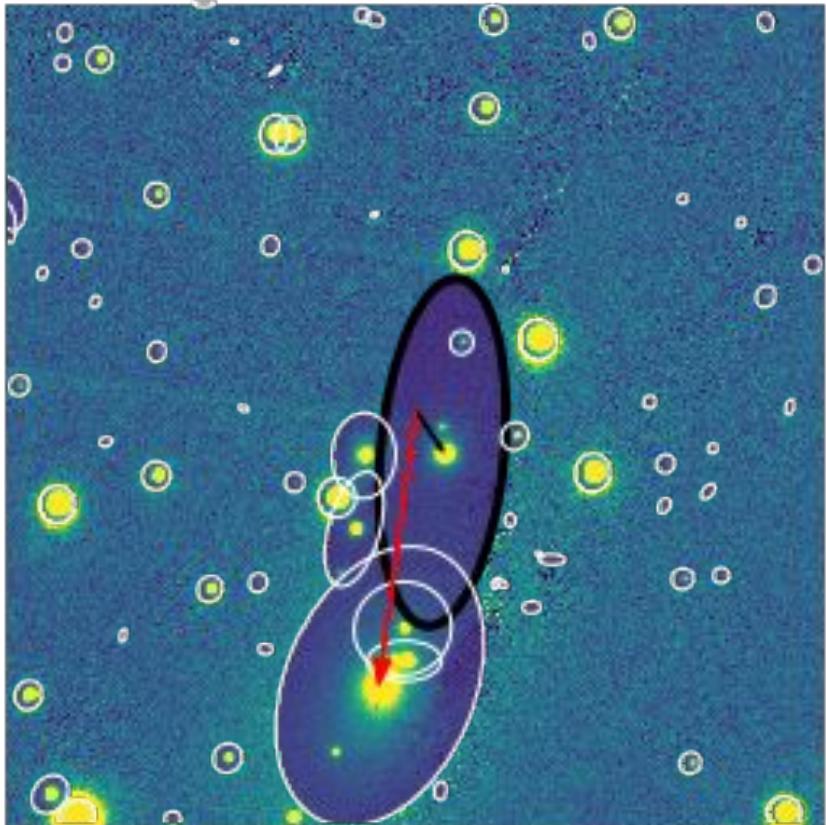
5 levels ($5 \times 30 \times 30 = 4,500$ pixels=2.0%)



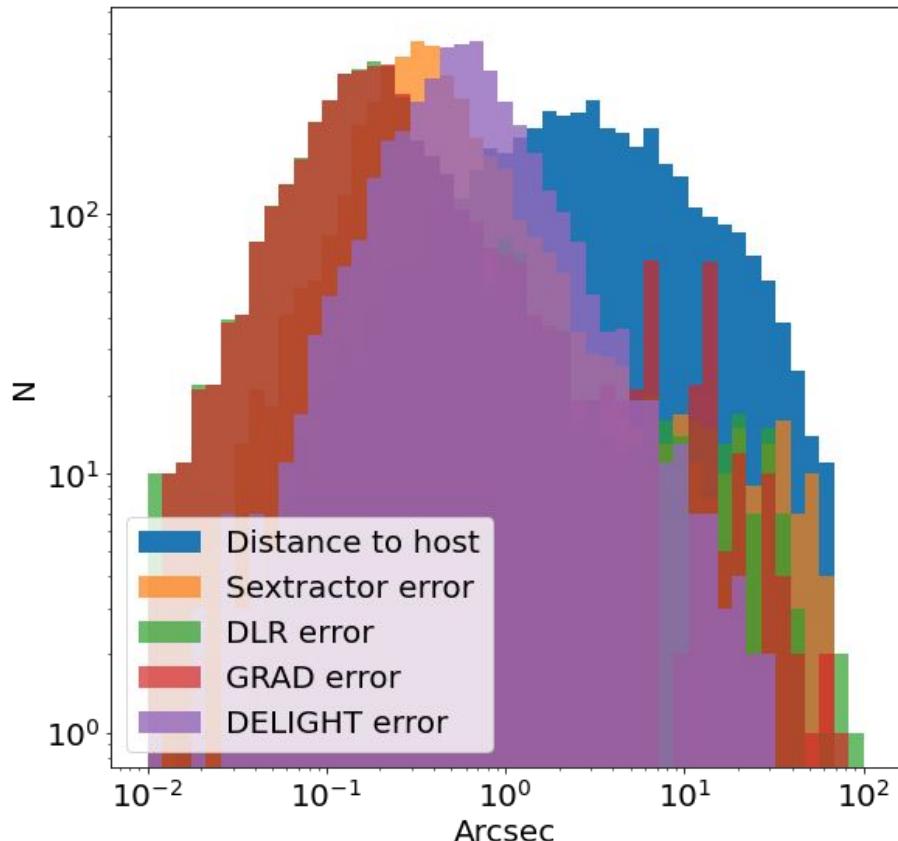
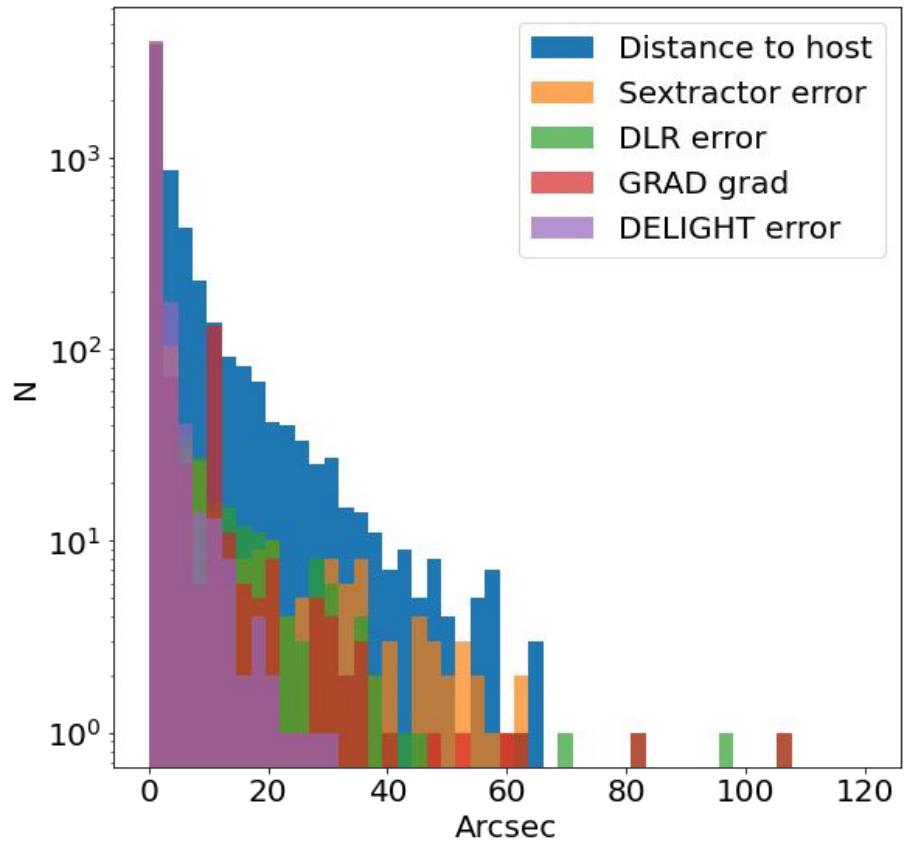
Host association methods

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



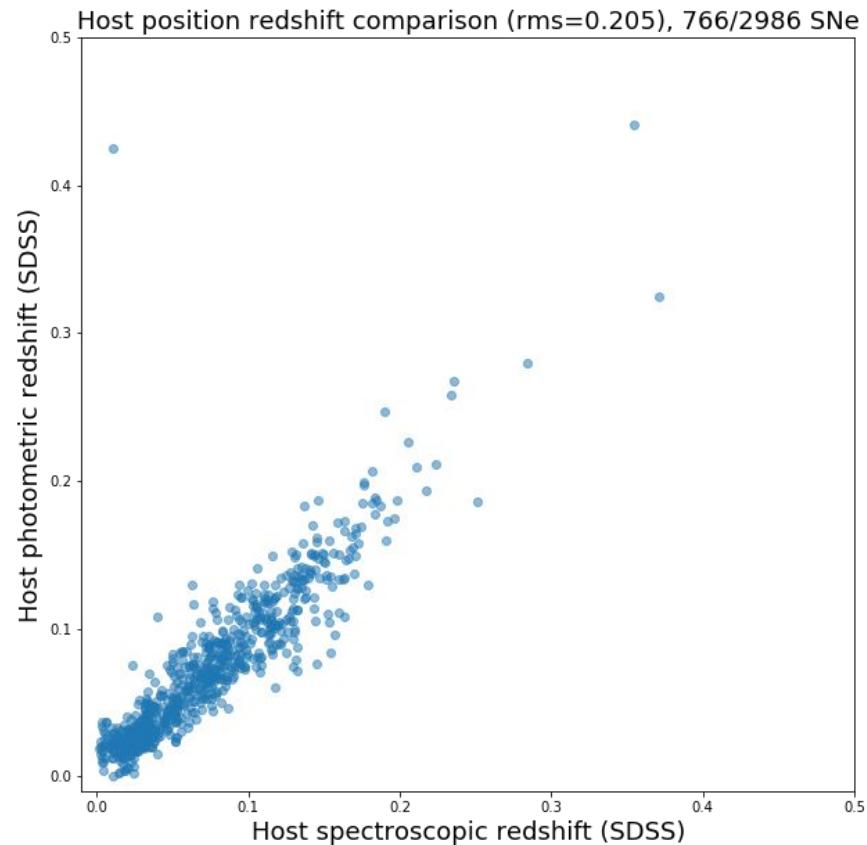
What is the redshift error?

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

What are the sources of error in the redshift?

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

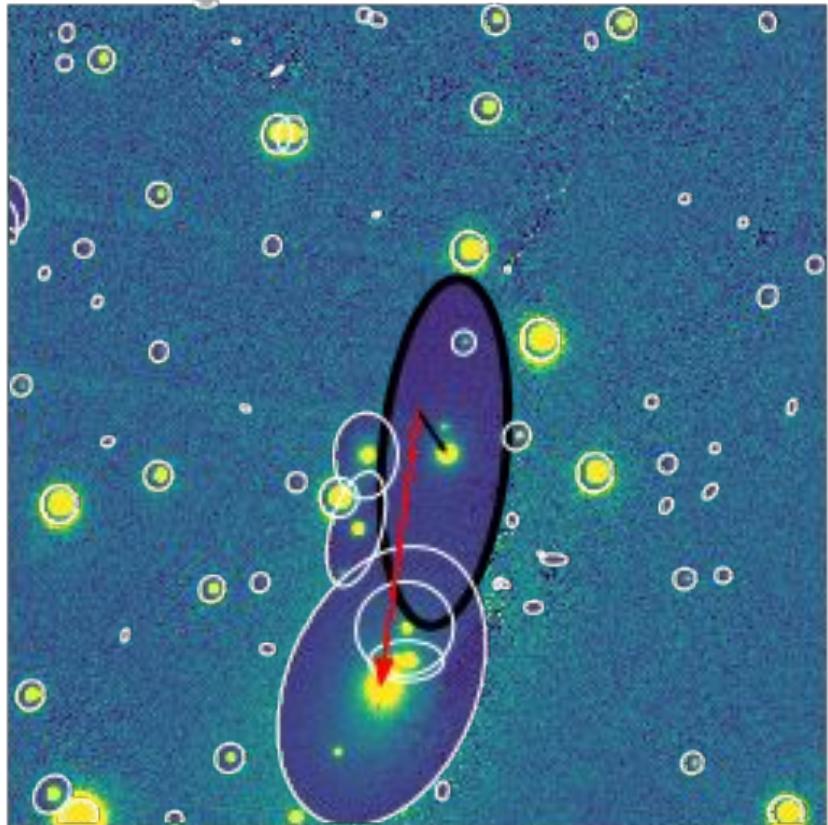
Can we estimate the distribution of the 2nd random variable?



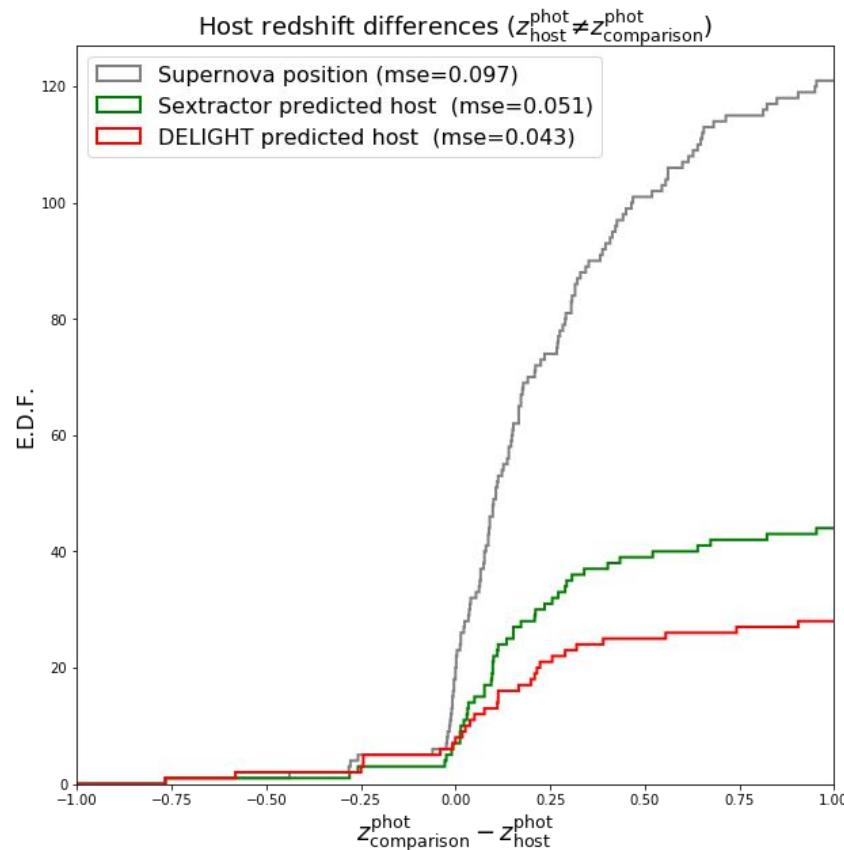
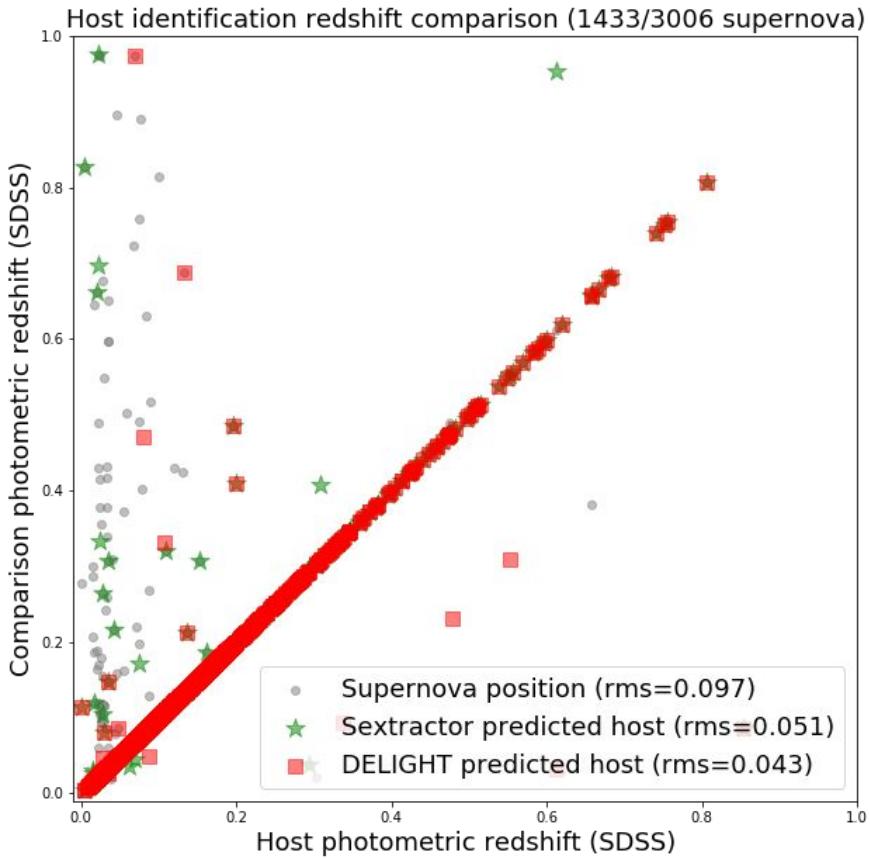
Redshift determination

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.

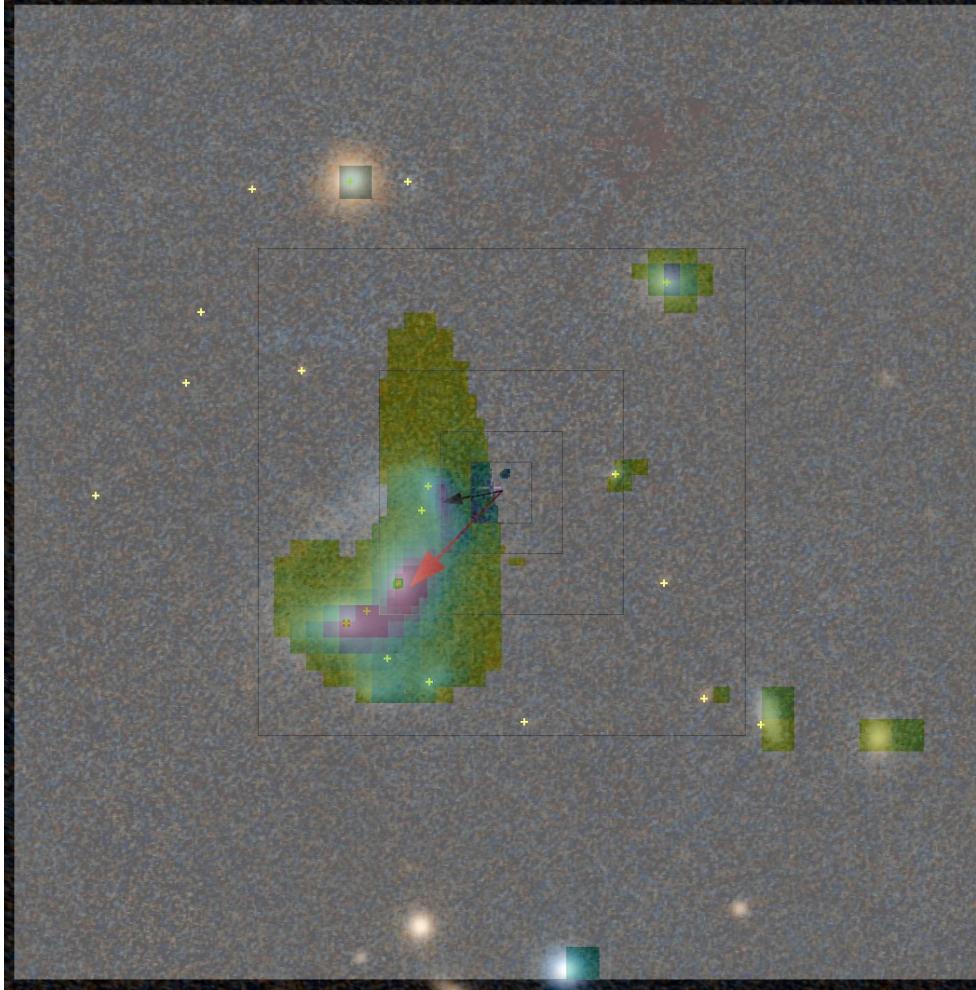


Redshift error



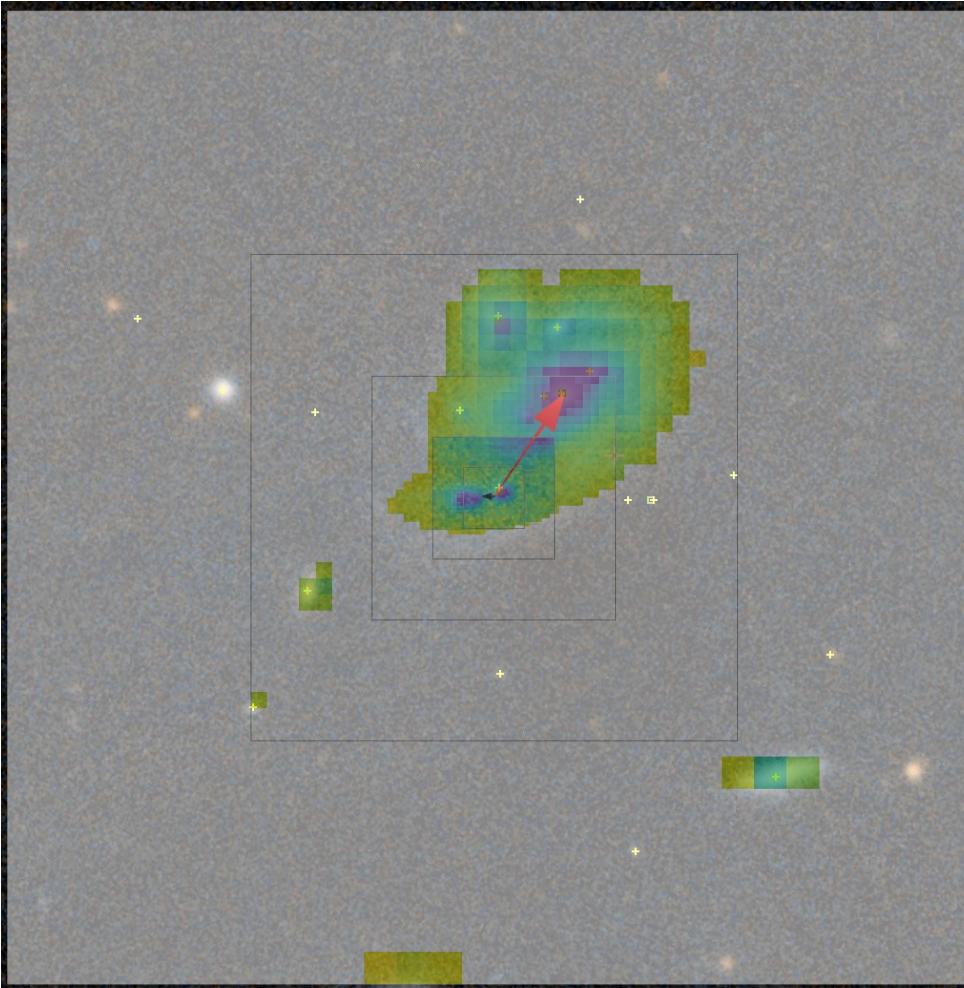
Examples

SExtractor
CNN



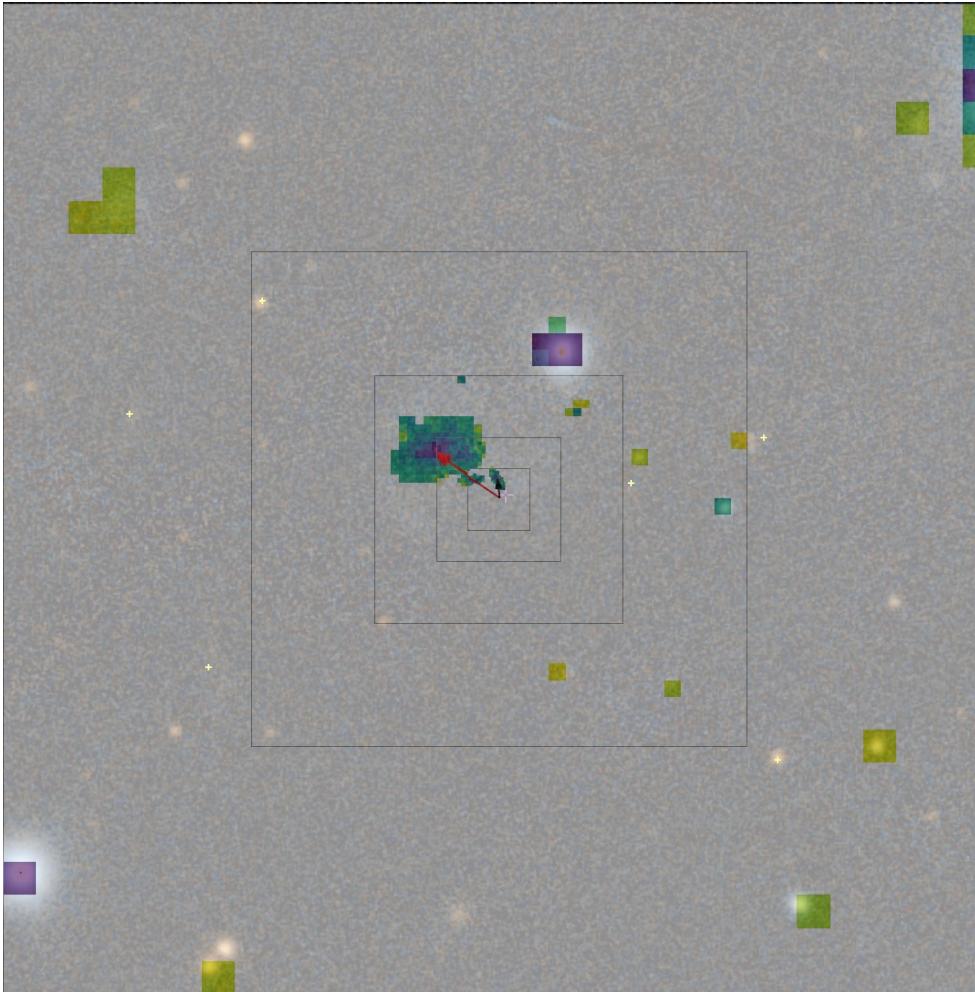
Examples

SExtractor
CNN



Examples

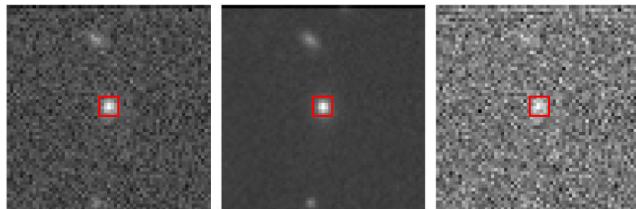
SExtractor
CNN



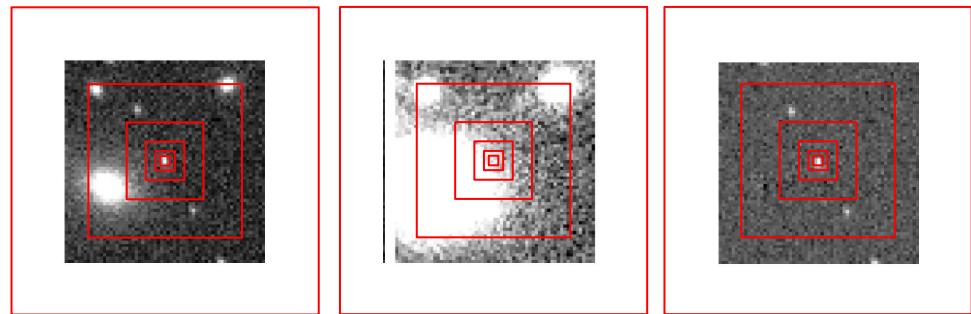
Conclusions & proposal

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
 $31 \times 31 = 961$ pix
FOV=6"



Proposal
 $15 \times 15 * (1 + (n-1) * 3/4)$ 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)