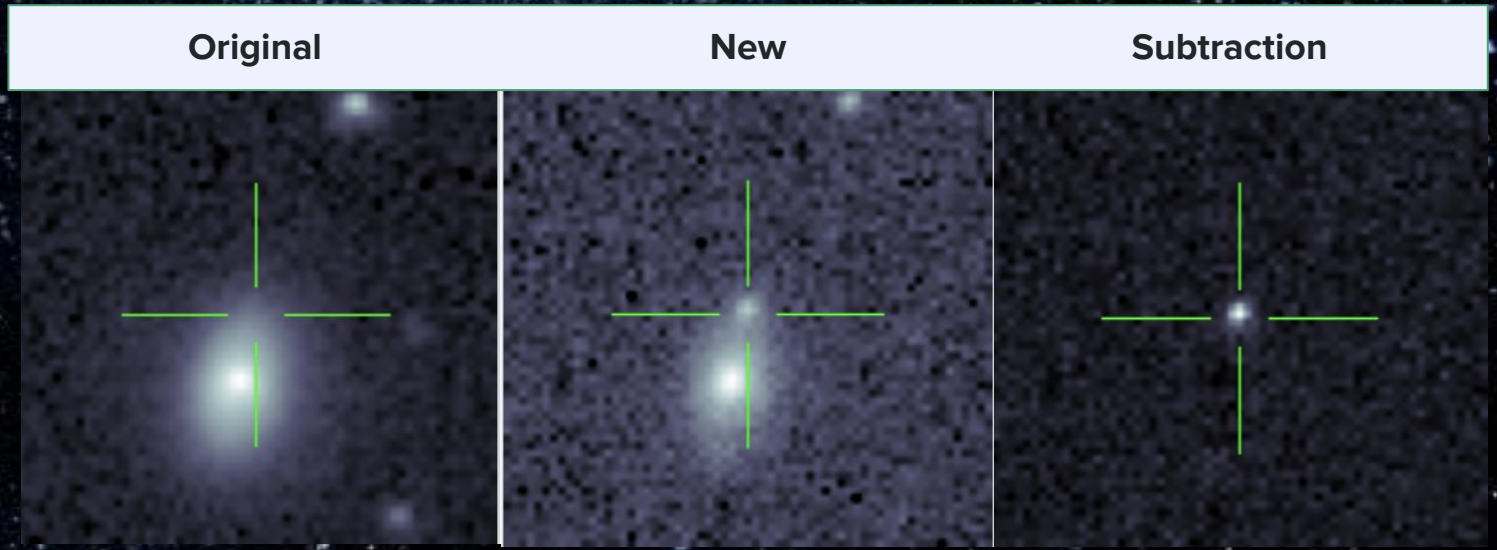
Kasliwal Lab: Creating a Dataset of “Reals” for the WINTER Real-bogus Classifier

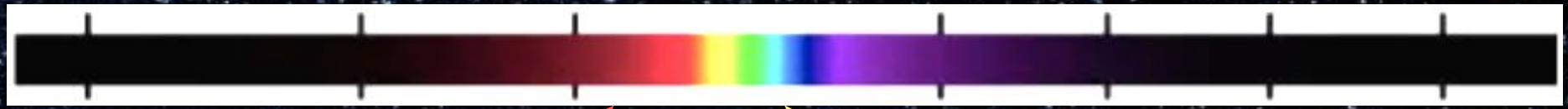
Student: Mallika Sheshadri
Mentor: Dr. Tomás Ahumada
Professor Mansi Kasliwal

What is time-domain astronomy?

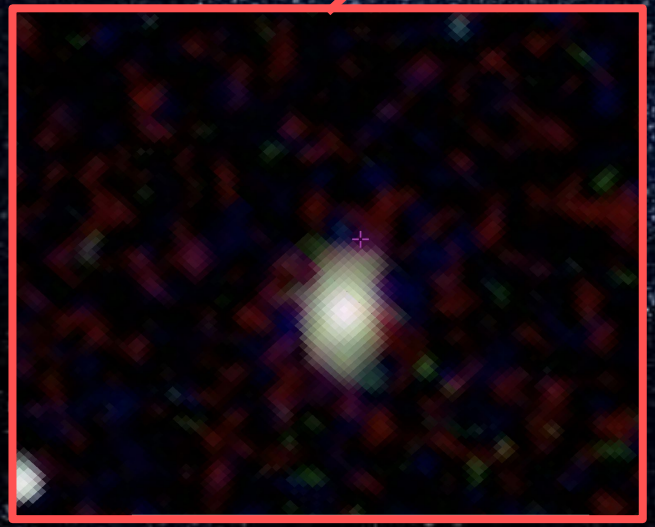
- Time-domain astronomy is the study how celestial objects change over time.
  - Objects that change are called transients.



Electromagnetic Spectrum



Infrared



Optical

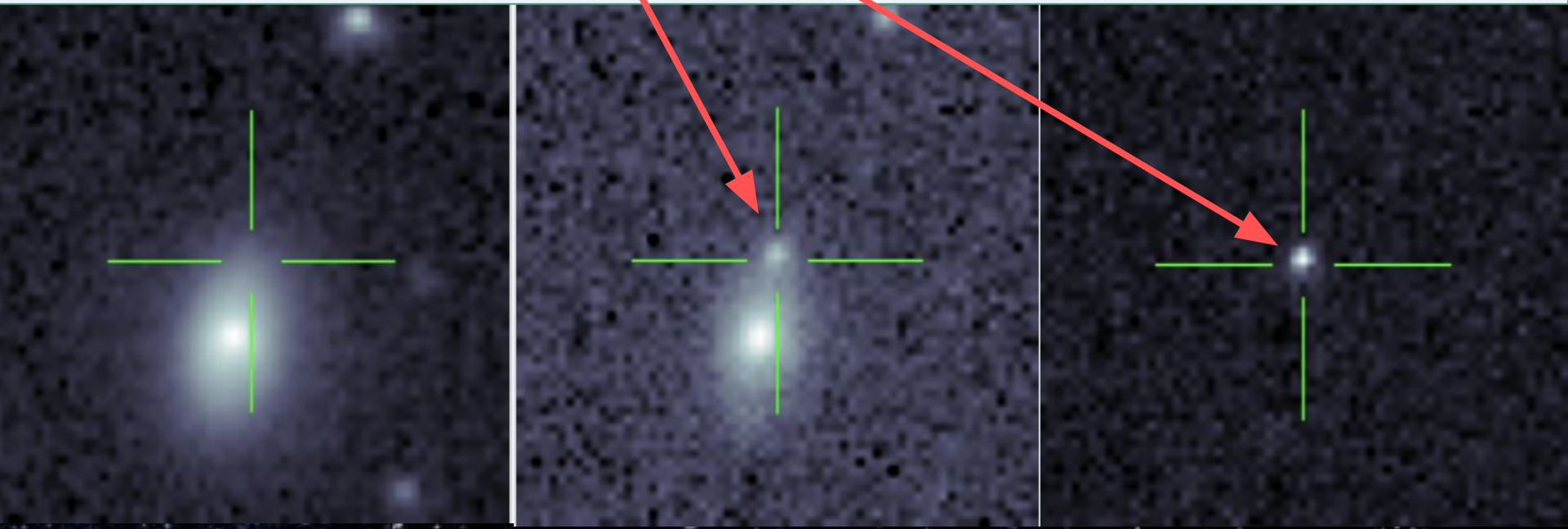
What is WINTER?

The Wide-field Infrared Transient Explorer, or WINTER, is a time-domain survey instrument on a 1 meter telescope at the Palomar Observatory in California.

Implemented in mid-2021, it has been designed to observe astronomical transients in the infrared spectrum.

What could this be?

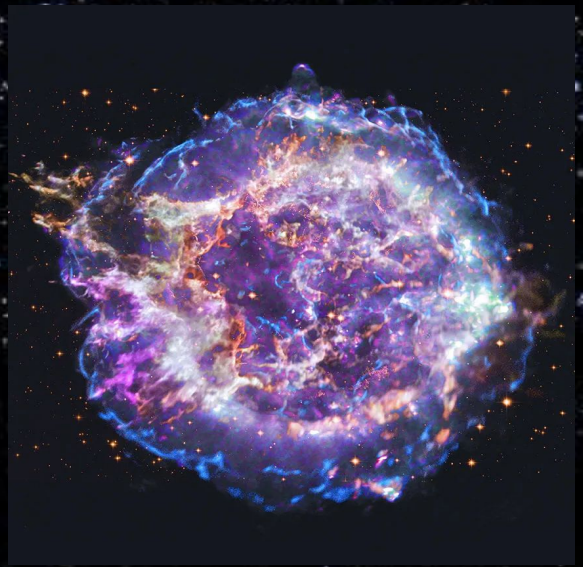
Original	New	Subtract the difference



Some things it could be:

Active Galactic Nuclei (AGNs)

Variable Stars



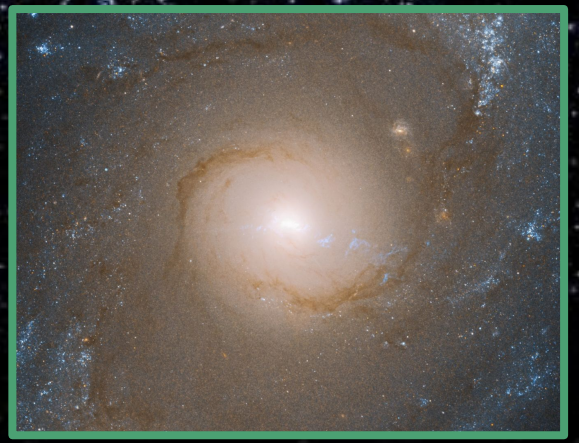
Supernovae

Active Galactic Nuclei (AGN)

AGNs are extremely luminous regions found at the center of some galaxies. They are powered by black holes at their centers that pull in and heat up gas and dust. This makes them shine for long periods of time.

They can be classified into several types:

- Seyfert Galaxies, which are extremely bright centers due to active supermassive black holes.
- Radio Galaxies, emitting strong radio waves from powerful jets of particles.
- Quasars, galaxies fueled by a enormous giant black holes producing high-energy jets in various directions.
- Blazars are a subclass of quasars. Their jets always point toward Earth.



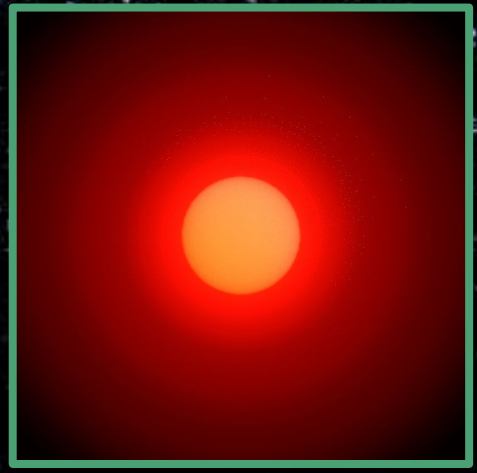
Variable Stars

Variable stars are stars whose brightness change over time.

There are two types of variable stars:

- Extrinsic variables: Brightness changes due to outside factors such as eclipsing binaries (top), where one star blocks light emitting from the other.

- Intrinsic variables: Brightness changes due to characteristics of the stars themselves, such as pulsating variables (bottom).



Supernovae

Supernovae are the explosions of supermassive stars.

(Often times becoming black holes).

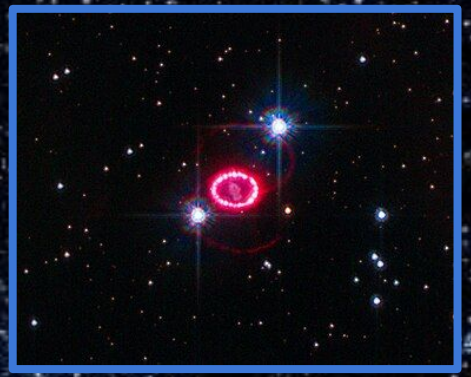
Two types of Supernovae:

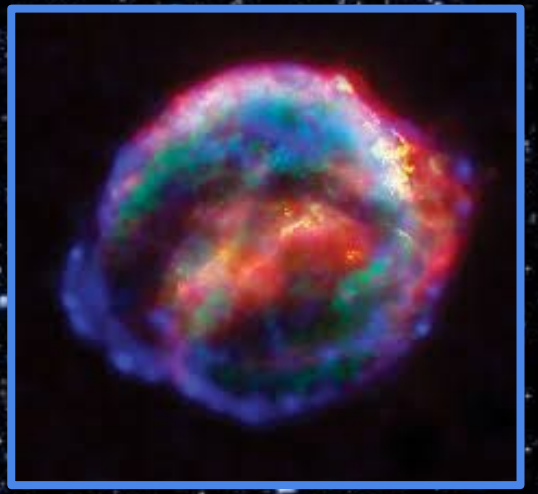
Type I Supernovae:

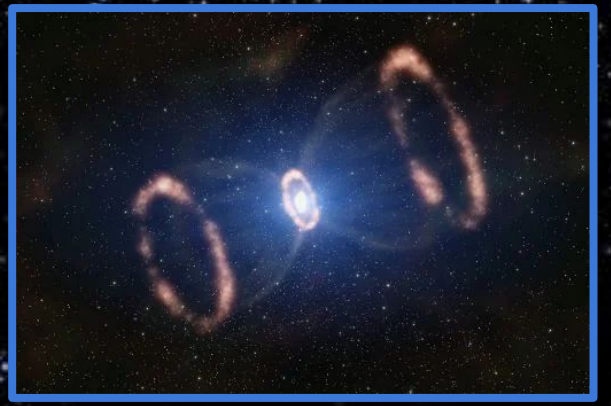
- **Cause:** A white dwarf star accumulates too much matter with other stars in the system, causing an explosion

Type II Supernovae:

- **Cause:** A big star runs out of fuel and collapses, which can create neutron stars or black holes.





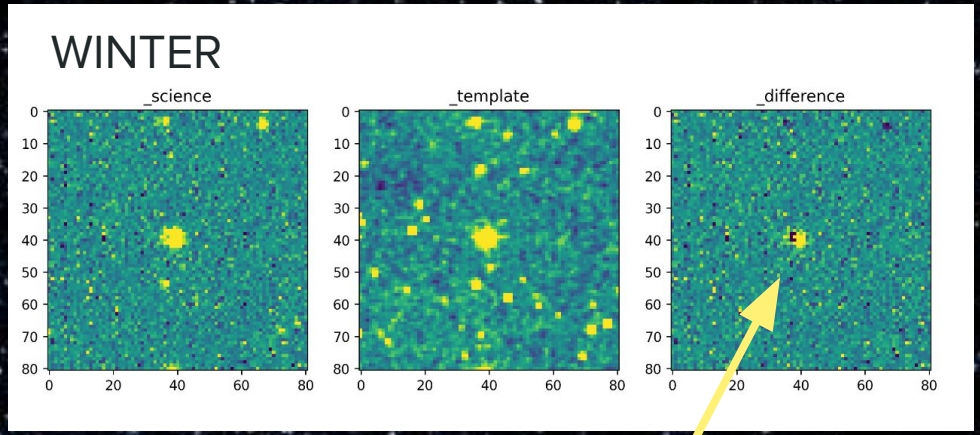


The Problem

Subtraction can help us find transients and classify them as different objects (supernovae, variable stars, AGNs, etc)

Image subtraction is complicated and sometimes generates false transients/sources.

With so many sources, it is difficult to sit down and sift through all of them to see which are real or not.



The Solution

A machine learning model was built to classify these sources as real or bogus (fake).

The real-bogus classification model can be trained with sample data sets of “reals” and “boguses” to be able to determine what each source actually is.

Prior to this summer, the model’s dataset had a lot of bogus samples but not very many real samples. This imbalance prevents it from being accurate.

For the last six weeks, we have been going through all of the WINTER data to make a dataset of “real” sources to train the real-bogus model.

The Roadmap to Get to a Better ML Real-Bogus Classifier

Acquire WINTER data + image subtractions + detect sources

Downloading catalogs and crossmatching

Train a machine learning algorithm to be able to detect reals on it's own

Read the source files to get coordinates

Analyze crossmatches to determine what's real and what's bogus

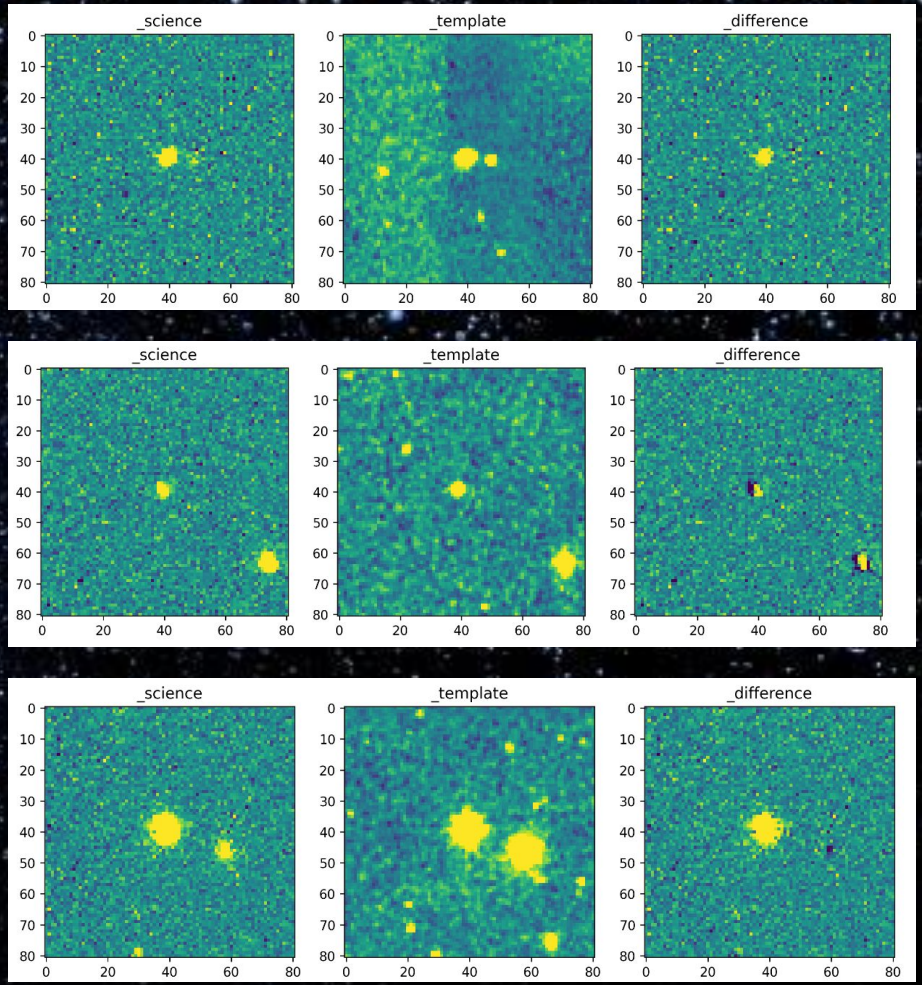
We are here

Getting the Data

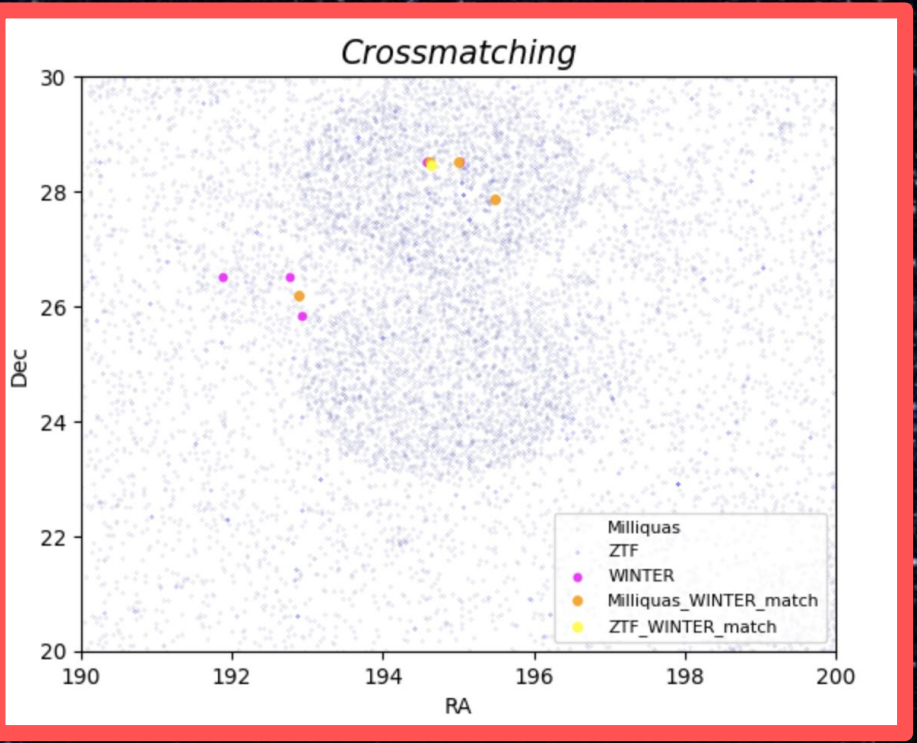
The first thing we had to do was to detect sources with WINTER and get the coordinates and subtractions for each source.

This data was already collected and available to us, since WINTER had been looking for infrared sources for a year.

These are what cutouts look like →



Plotting and Crossmatching



Then, we had to get the source files and read the coordinates.

Right Ascension (RA) and Declination (Dec) are a coordinate system to position objects in the sky

Crossmatching: Identifying objects that have the same coordinates but are in different catalogs.

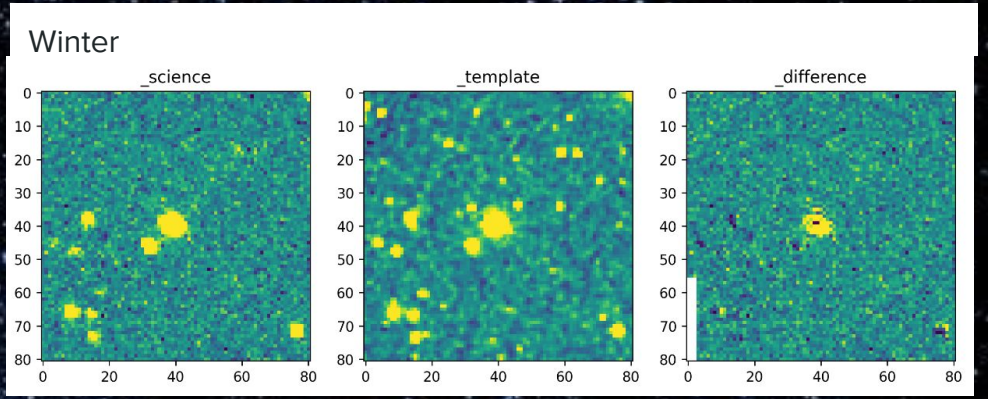
We used *astropy* and *matplotlib* to work with astronomical data and plot the data.

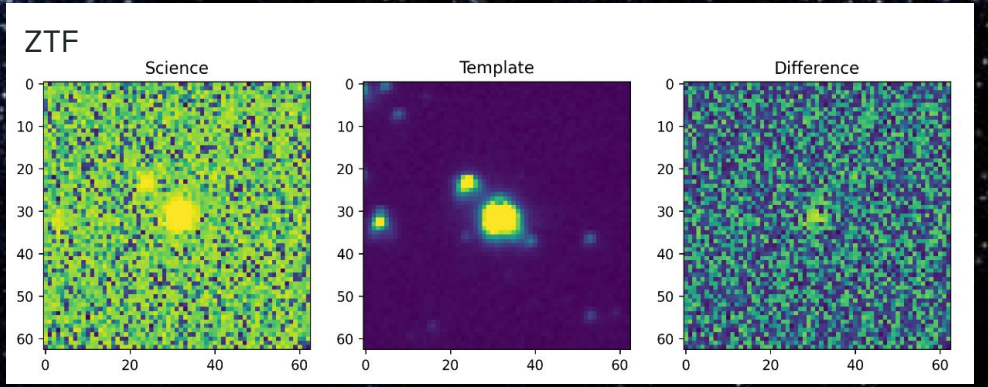
Finding Reals

To figure out what was real and what wasn't in the WINTER data, we crossmatched the WINTER coordinates with sources from the Zwicky Transient Facility (ZTF).

ZTF is an optical survey that images the northern sky every two nights, similarly to WINTER.

If a source in WINTER shared coordinates with a source in ZTF, there is a high chance they are both detecting the same source, so it's probably real.





Analysis of the Night of June 15th, 2024

32,000 WINTER sources → 400 crossmatches to ZTF

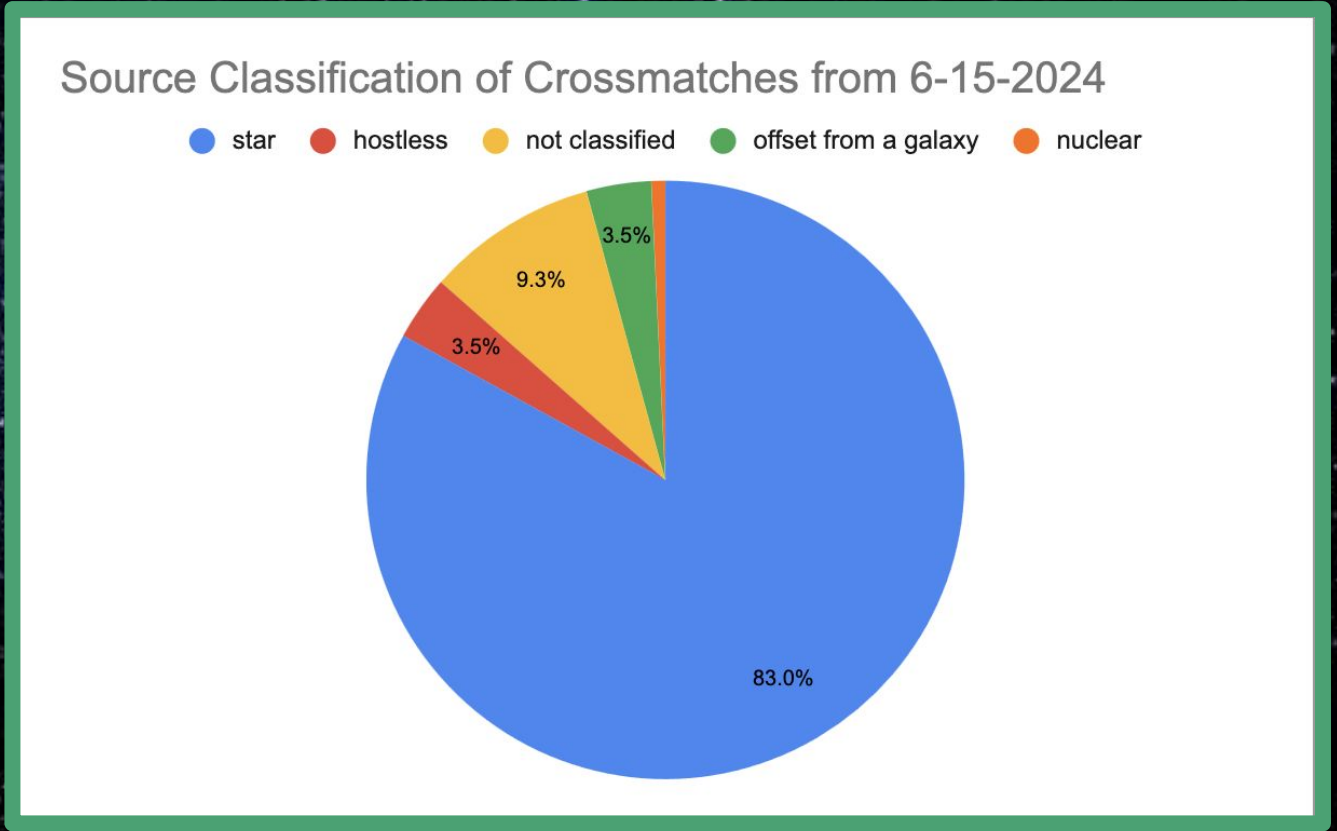
	Unnamed: 0	WINTER_name	ra	dec	file	packet_idx	ztf_xmatch
0	18	WNTR24arjeq	192.646207	26.710739	winter_test/avro/WINTERcamera_20240614-034205-...	18	ZTF18aavrpts
1	61	WNTR24eyraw	192.358804	26.630885	winter_test/avro/WINTERcamera_20240614-034205-...	61	ZTF20aauzzfu
2	274	WNTR24ezaoi	192.501092	26.380356	winter_test/avro/WINTERcamera_20240614-034413-...	52	ZTF18aasvurt
3	276	WNTR24abkwn	192.568611	26.337719	winter_test/avro/WINTERcamera_20240614-034413-...	54	ZTF18aahvnoe
4	281	WNTR24afdyk	192.600873	26.238923	winter_test/avro/WINTERcamera_20240614-034413-...	59	ZTF18aahvnnl
...	...	...	...	...	...	...	...
395	31325	WNTR24csloa	334.195971	36.666697	winter_test/avro/WINTERcamera_20240614-114759-...	761	ZTF18abzovzh
396	31389	WNTR24cnnwq	334.311559	36.622389	winter_test/avro/WINTERcamera_20240614-114759-...	825	ZTF18abctzyk
397	31438	WNTR24ccylx	334.009178	36.586844	winter_test/avro/WINTERcamera_20240614-114759-...	874	ZTF24aawbqkp
398	31461	WNTR24eyqtx	334.674775	36.580617	winter_test/avro/WINTERcamera_20240614-114759-...	897	ZTF22abhzjly
399	32128	WNTR24eylwh	333.973376	36.437859	winter_test/avro/WINTERcamera_20240614-114759-...	525	ZTF19aboblml

400 rows × 7 columns

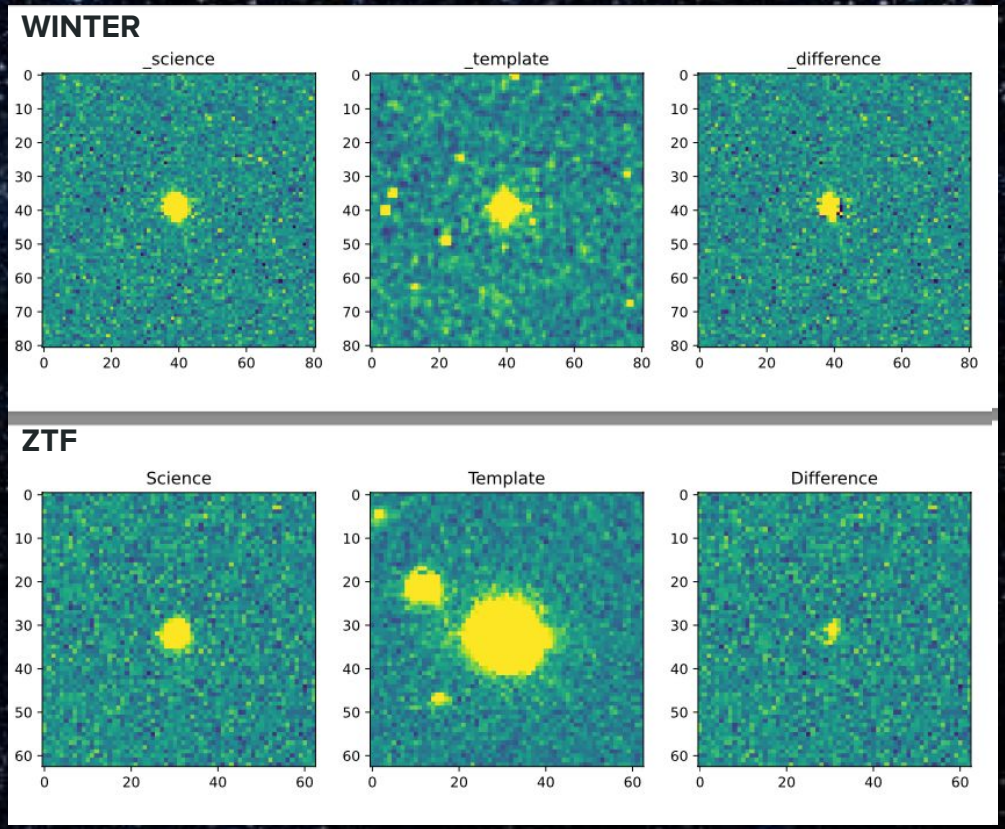
Analysis of the Night of June 15th, 2024

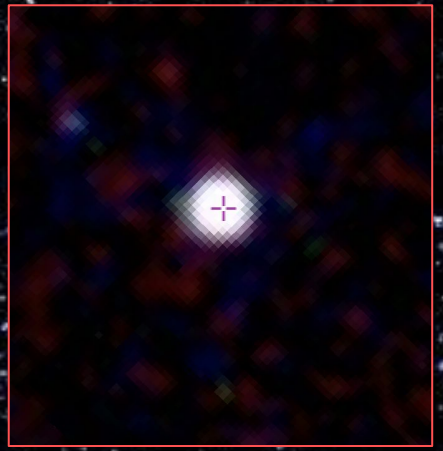
- 332 are stars
- 3 are nuclear
- 14 are offset from a galaxy
- 14 are hostless

37 were not classified.

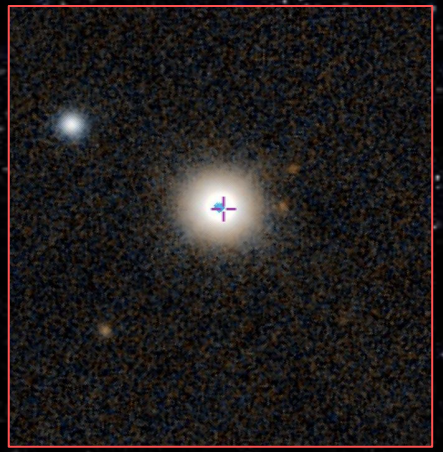


Subtractions and Cutouts of a Star



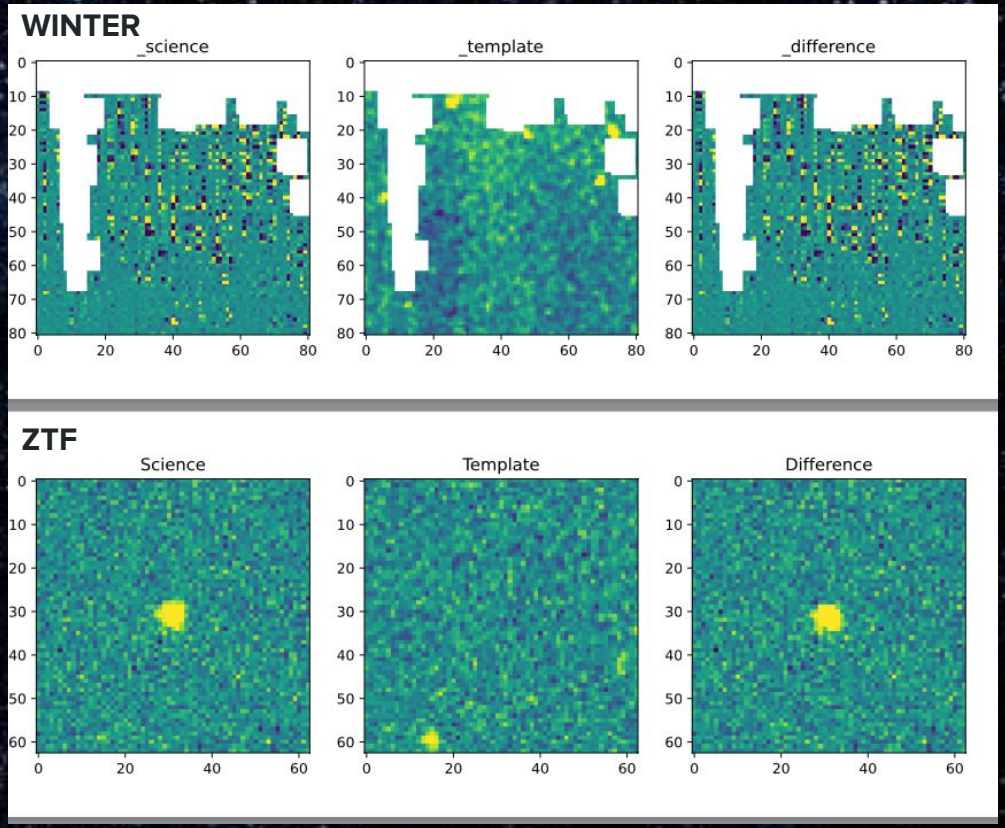


2MASS (infrared)



PanSTARRS (optical)

Bad Cutouts and Subtractions



The Roadmap to Get to a Better ML Real-Bogus Classifier

Acquire WINTER data + image subtractions + detect sources

Downloading catalogs and crossmatching

Train a machine learning algorithm to be able to detect reals on it's own

Read the source files to get coordinates

Analyze crossmatches to determine what's real and what's bogus

We are here

Next Steps and Future Work

The next step is to run our code on the entire WINTER database instead of just a few nights, to give us as much data as possible.

Then we can use our dataset of reals and boguses to train the model to continue to do this work on its own.