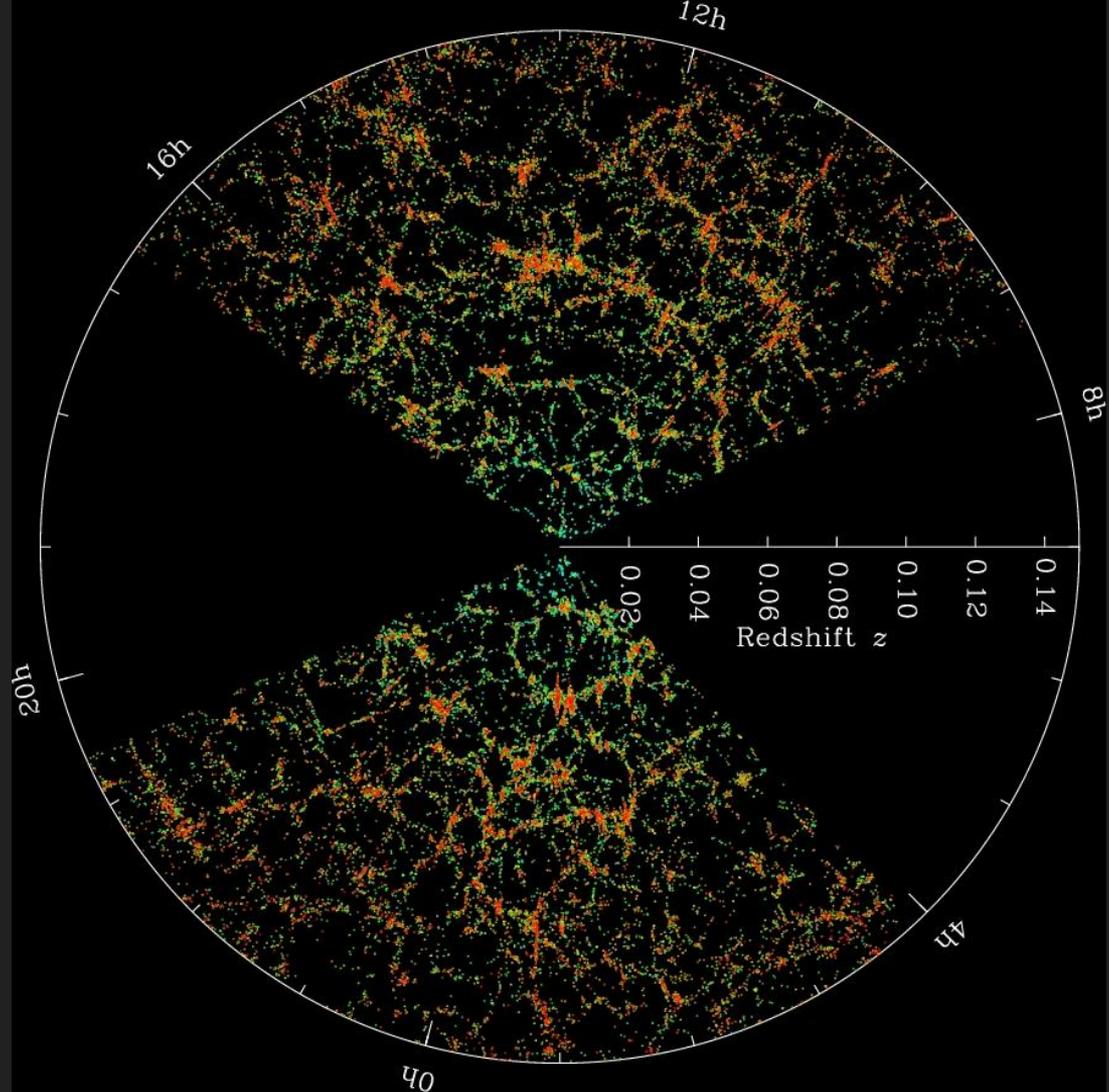
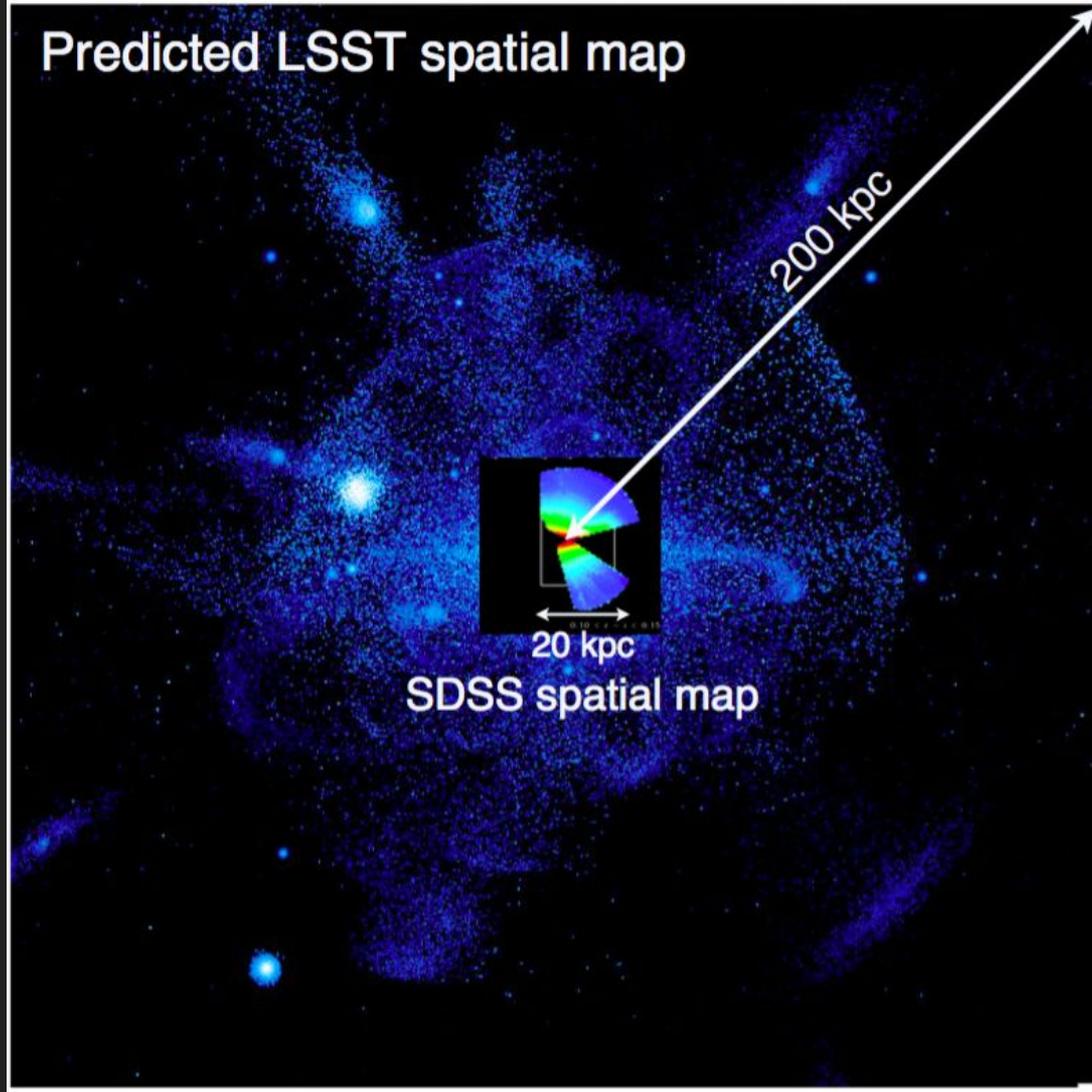


# Computer Vision Experiments in Crowdsourced Astronomy

Rasmi Elasmr  
re2300@columbia.edu



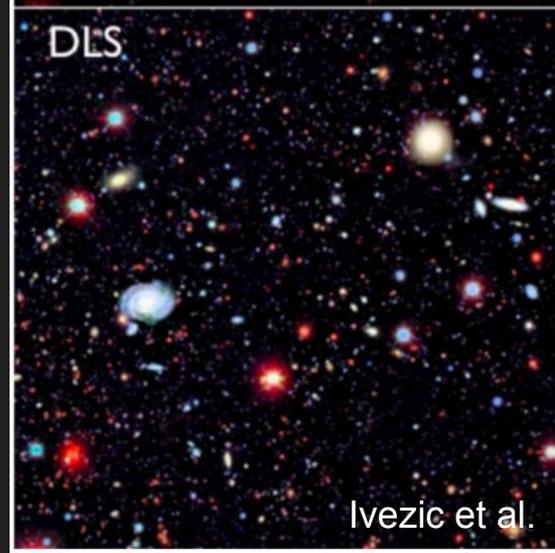
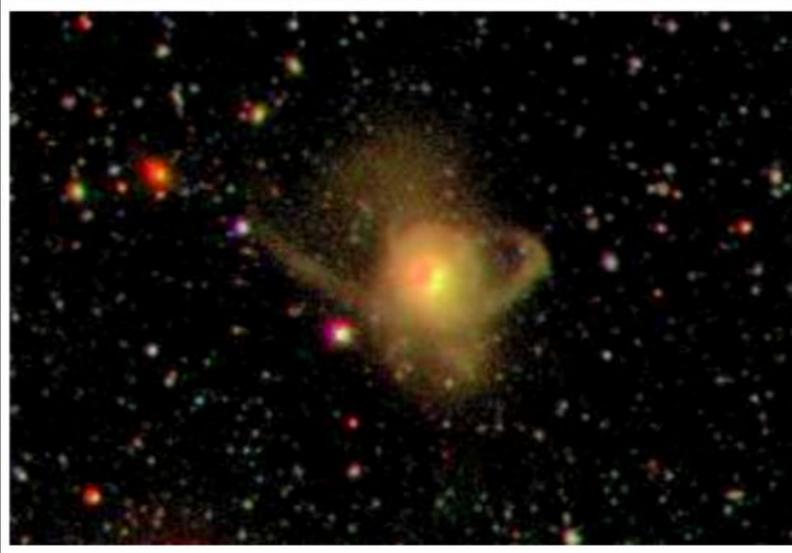
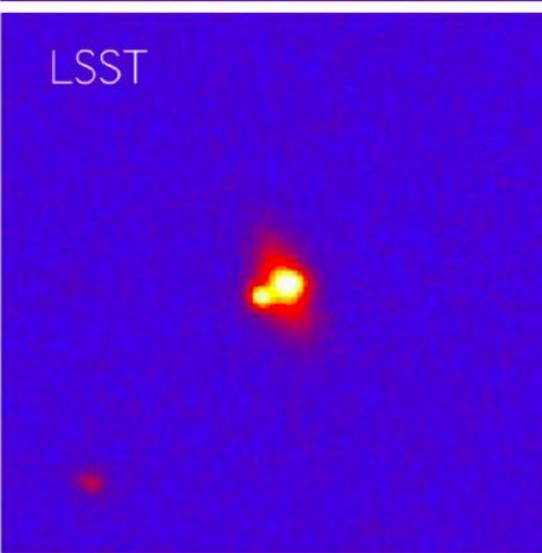
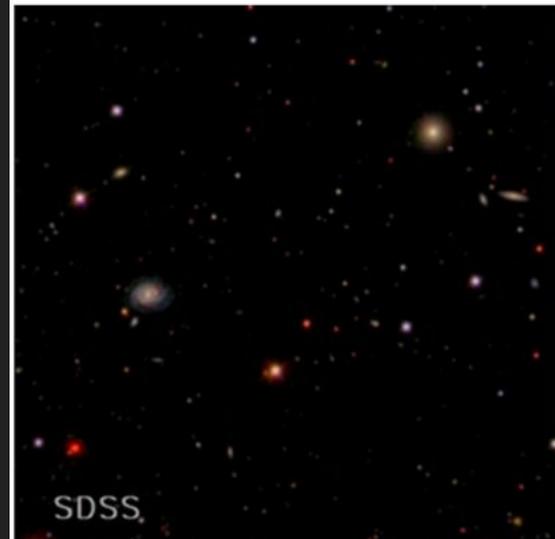
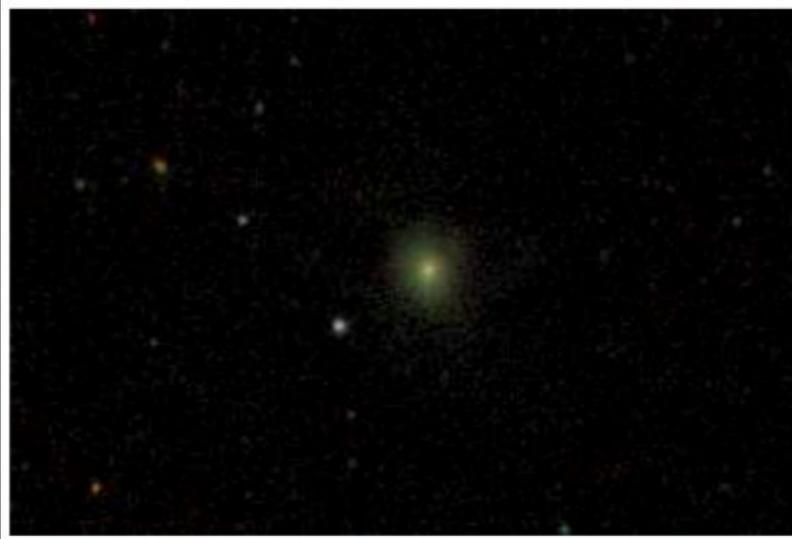
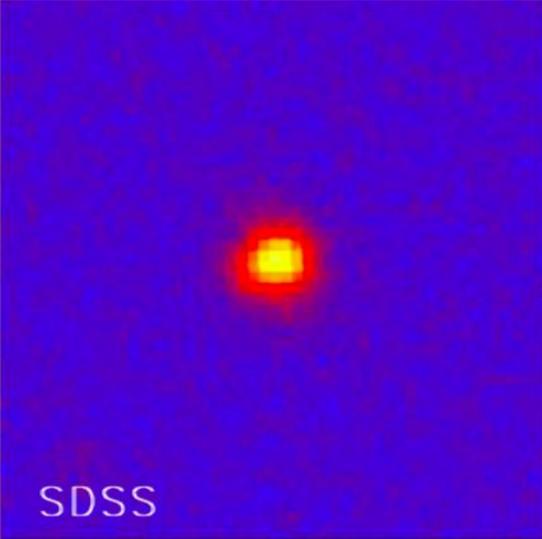
Predicted LSST spatial map



200 kpc

20 kpc

SDSS spatial map



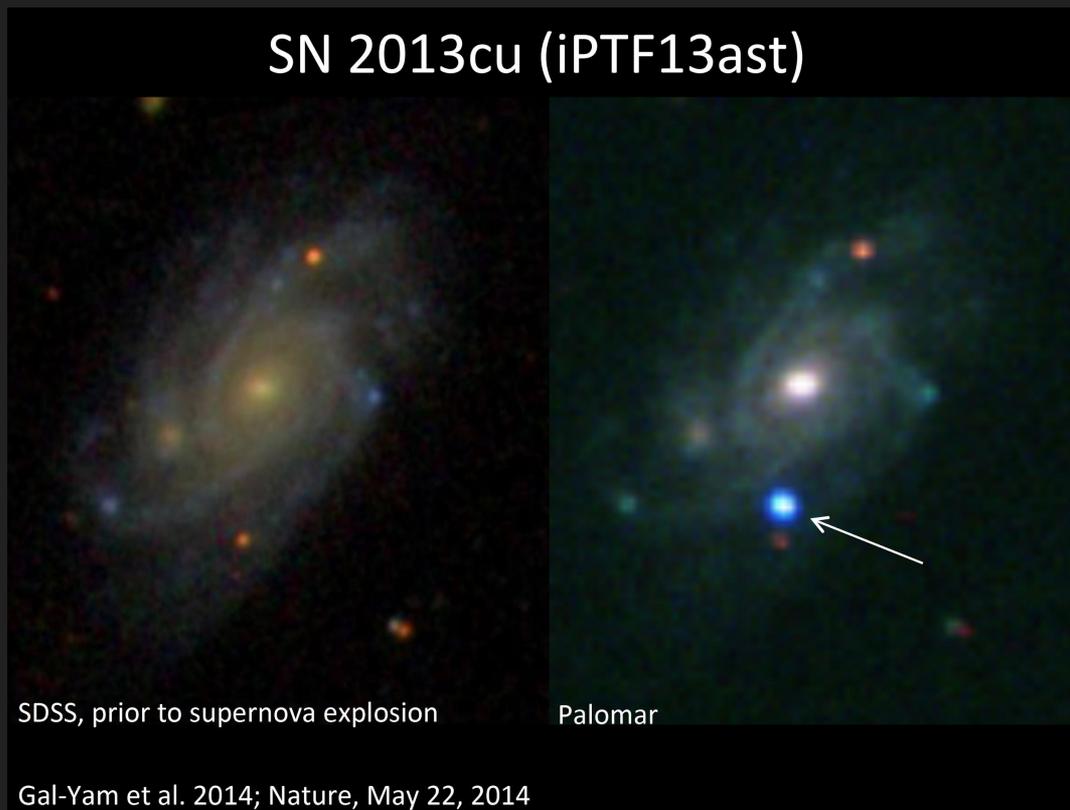
30 terabytes  
(~1 SDSS)  
*every night*

~500 petabytes total

How do you deal with  
the universe's dataset?

# Requirements

- Processing
- Transient object detection for quick (<60 seconds) follow-up
- Classification/cataloging
- Archiving



“Machine learning is really good at partially solving just about any problem.”

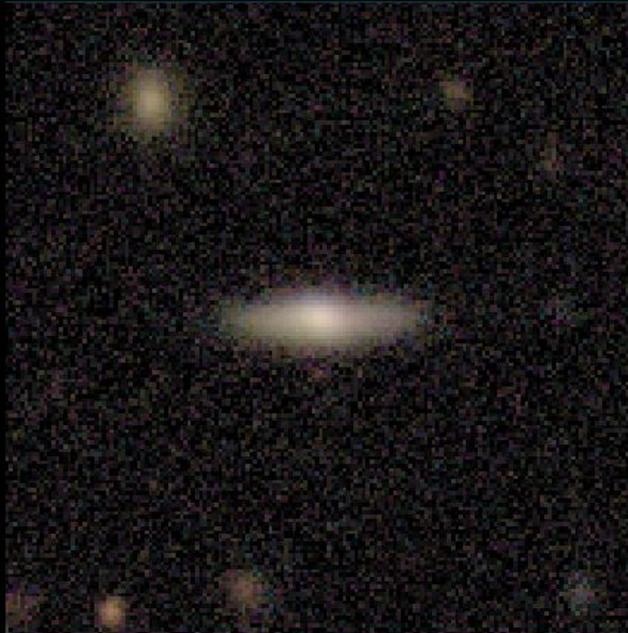
- Chris Dixon

Crowdsourcing analysis?

# Galaxy Zoo

The first time a human will  
see these galaxies.

Easy-to-answer questions.



# Classify



Examples Restart

Note: Please always classify the galaxy in the centre of the image.

## SHAPE

Is the galaxy simply smooth and rounded, with no sign of a disk?

		
Smooth	Features or disk	Star or artifact

## ODD

Do you see any of these odd features in the image?

			
None	Ring	Lens or arc	Irregular
			
Other	Dust lane	Overlapping	Done

## ROUND

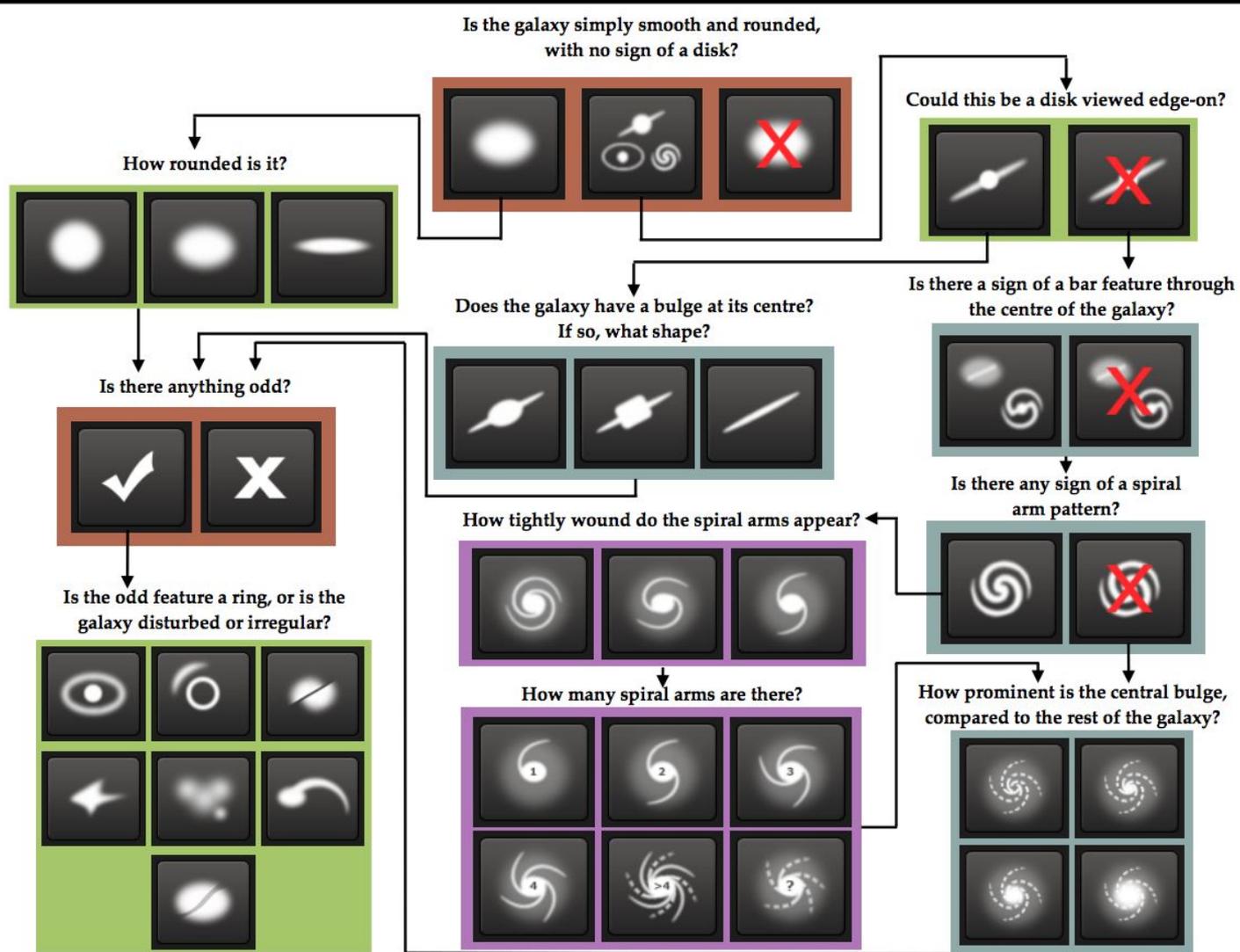
How rounded is it?

		
Completely round	In between	Cigar shaped

## MERGERS

Is the galaxy currently merging or is there any sign of tidal debris?

			
Merging	Tidal debris	Both	Neither



# Benefits

Deal with data that doesn't typically work well with ML.

Double-check and assess ML predictions.

Use classifications to train better models.

Highlight potentially interesting objects and events for expert follow-up.

Run large-scale experiments.

More detailed analysis than simple classification.

# Accurate!

Amateur consensus agrees with experts 97% of the time.

(Experts agree with each other 98% of the time.)

# Adaptable

All Projects



Medicine



Nature



Physics



Social Science



Space



# Make your own!

A screenshot of the Zooniverse Project Builder interface. The background is a dark, textured image of a building facade. The text 'ZOONIVERSE' is centered at the top in a white, sans-serif font. Below it, the text 'Zooniverse Project Builder' is centered in a smaller, white, sans-serif font.

ZOONIVERSE

Zooniverse Project Builder

[zooniverse.org/lab](https://zooniverse.org/lab)

More advanced analysis?

# Galaxy Zoo Challenge

kaggle

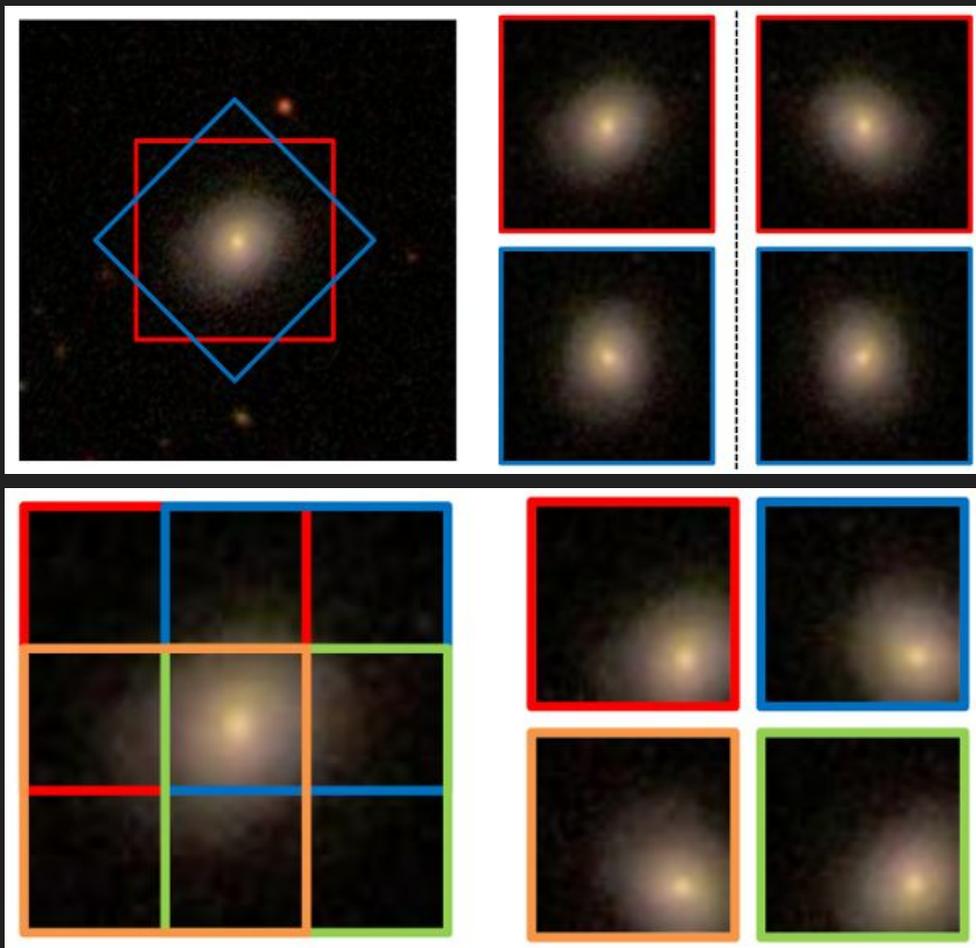
“This competition asks you to analyze the JPG images of galaxies to find automated metrics that reproduce the probability distributions derived from human classifications. For each galaxy, determine the probability that it belongs in a particular class. Can you write an algorithm that behaves as well as the crowd does?”

# Tricks

Winning solution used a 42-million layer convolutional neural network to achieve ~75% accuracy.

Most performance gains were from reducing overfitting (dropouts, data modification).

Space has no preferred direction — rotating, translating, zooming, mirroring images adds more valid training data.

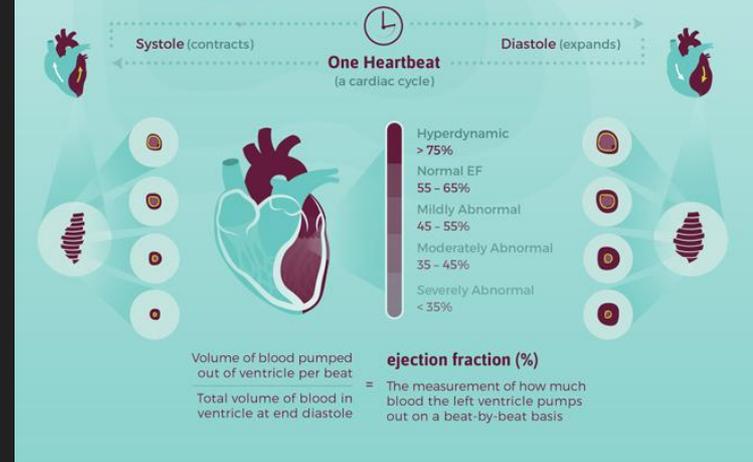
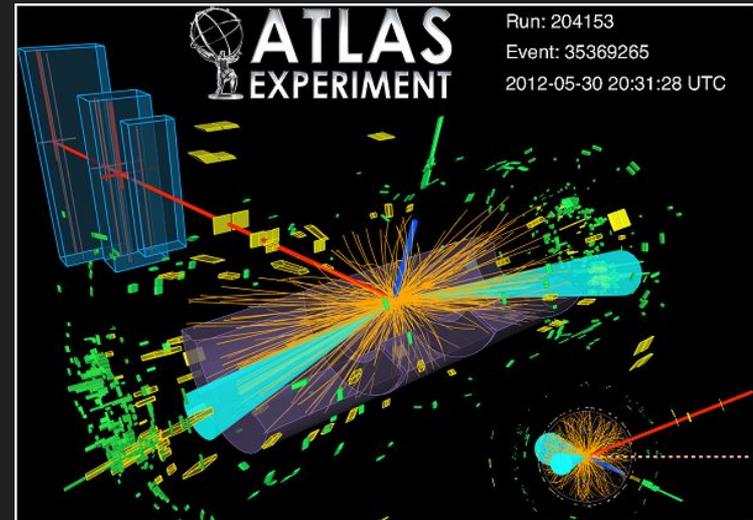


# Proven in other fields

“Use the ATLAS experiment to identify the Higgs boson”

“Transforming How We Diagnose Heart Disease”

“Predict ocean health, one plankton at a time”



Crowdsourcing data?

# Why?

Free, useful data — people are always watching.

Solves problems of alignment, combination, normalization...

Pushes current limits of machine learning with messy data.

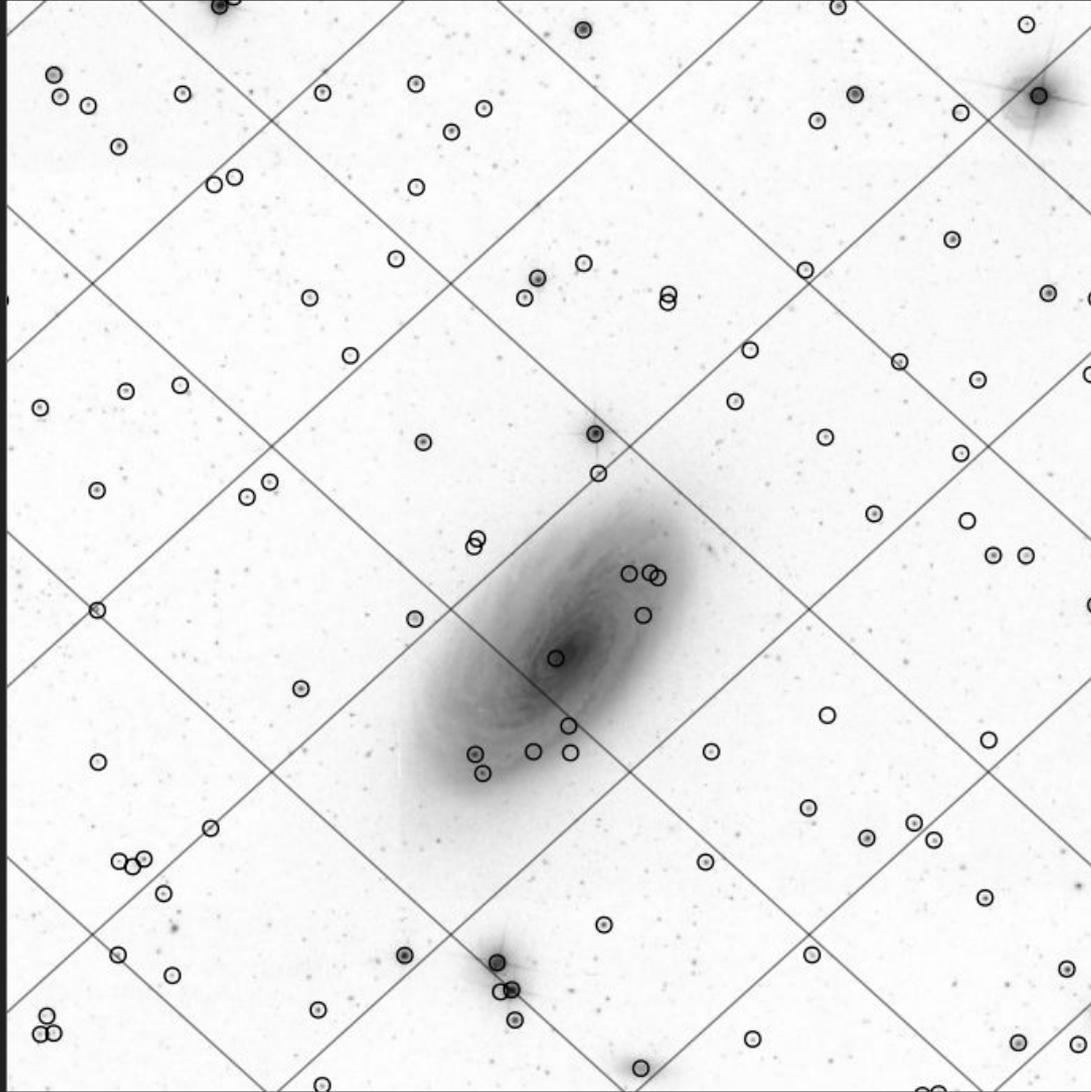
How do we know where an  
image is in the sky?

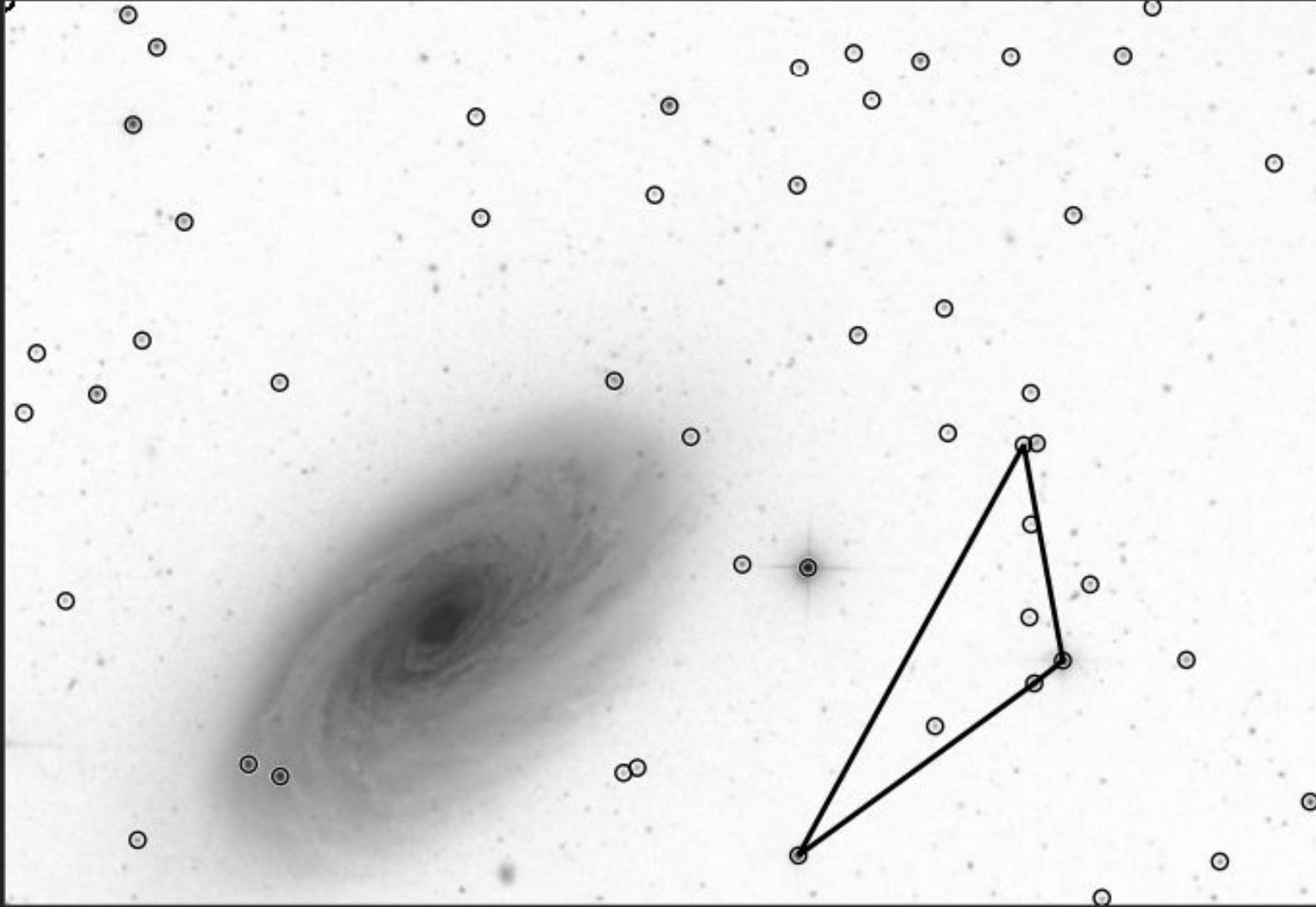
A search problem.

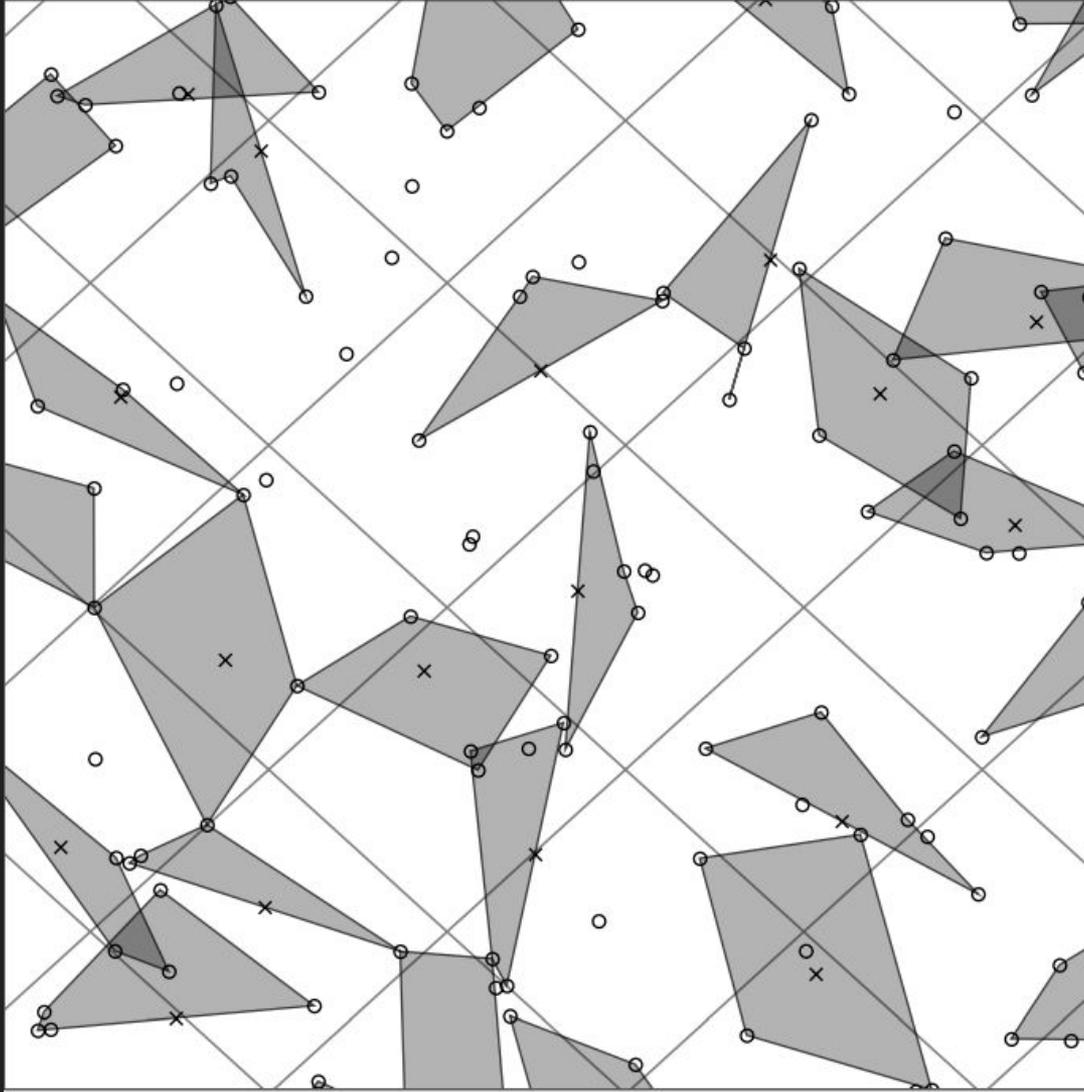
Astrometry.net

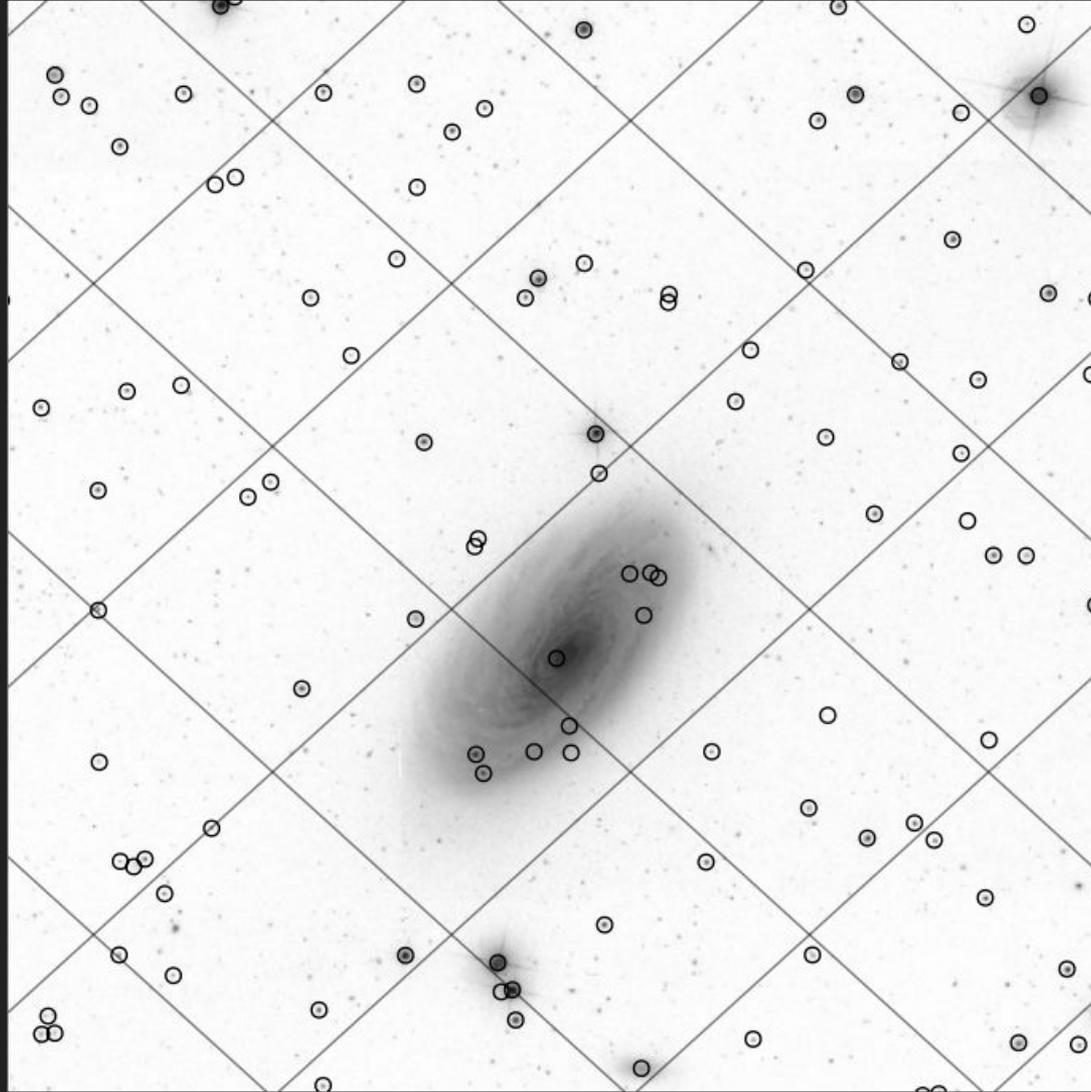
# Reverse-lookup

Indexed hash, just like a Google search.



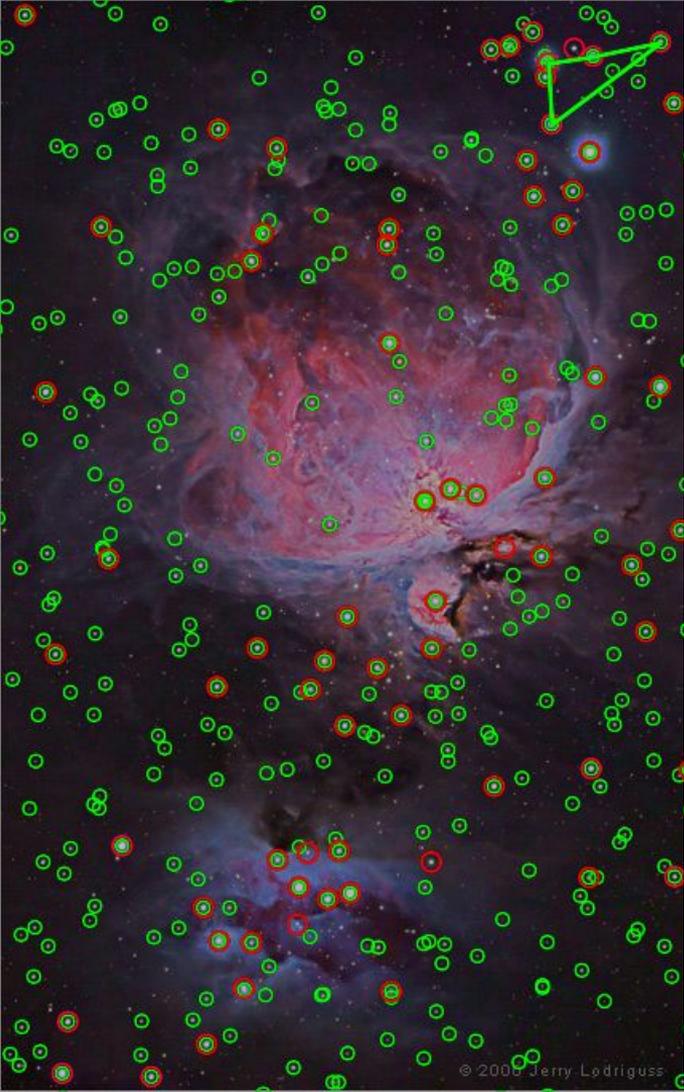




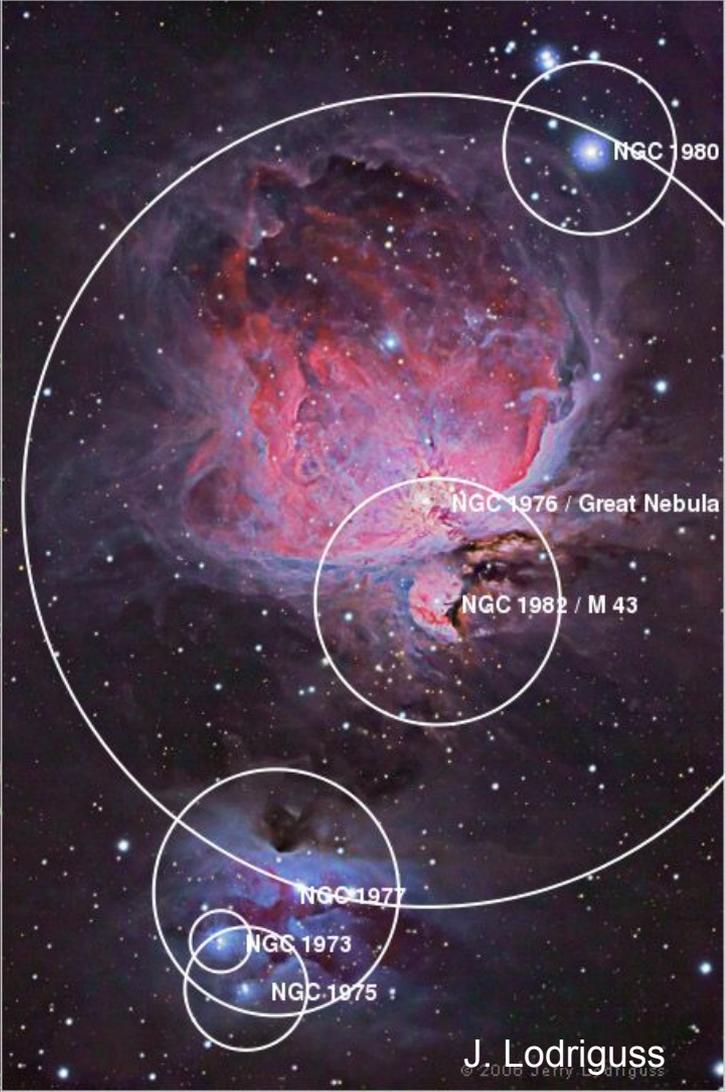




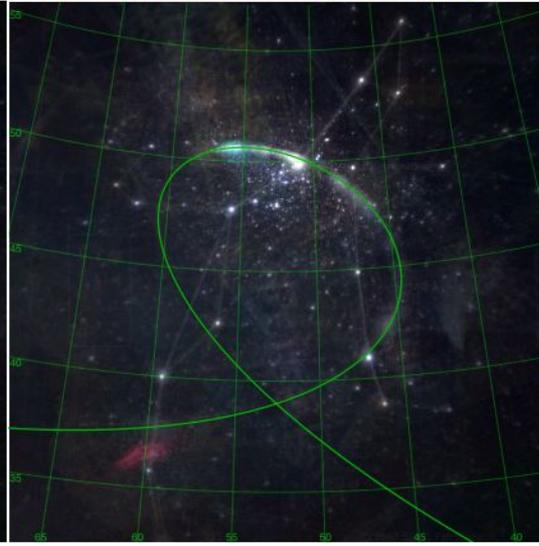
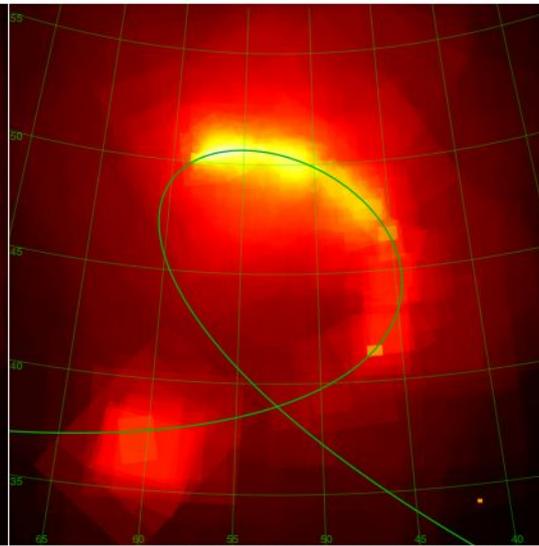
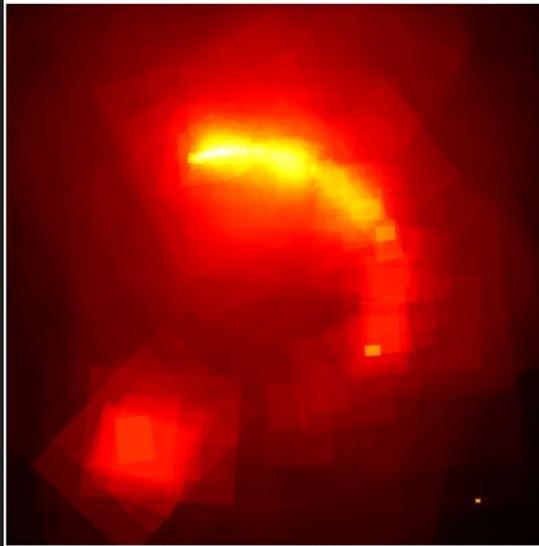
© 2006

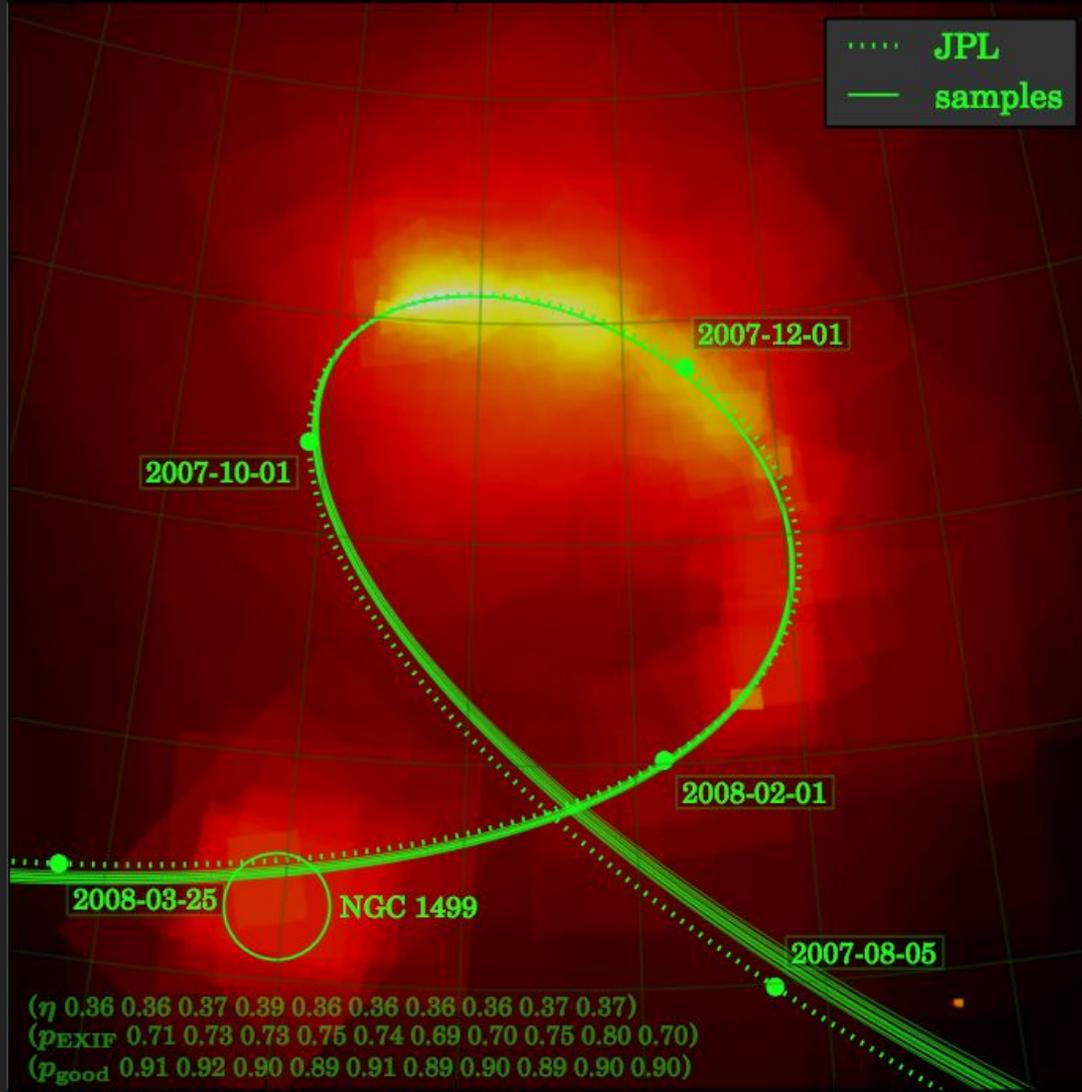


© 2006 Jerry Lodriguss

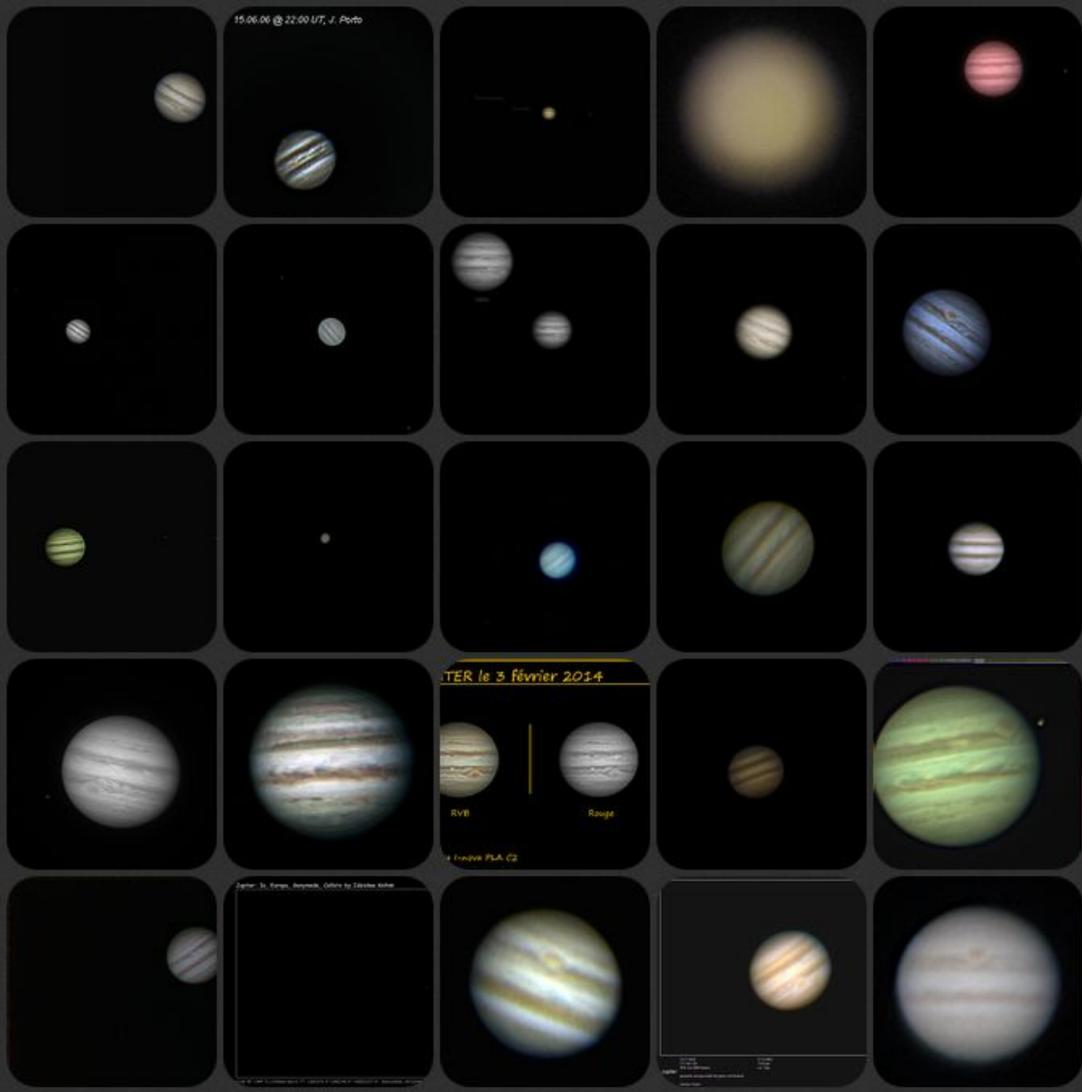


J. Lodriguss  
© 2006 Jerry Lodriguss





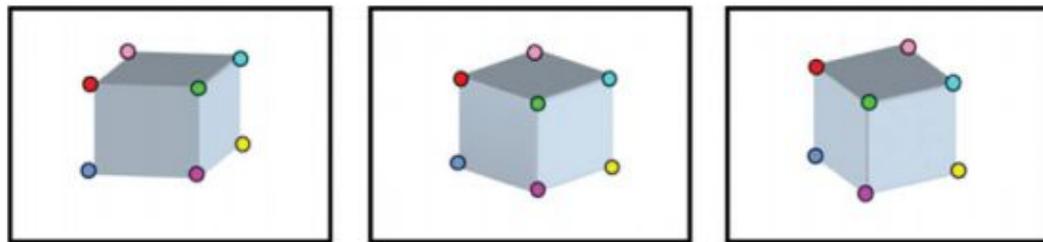
AstroBin



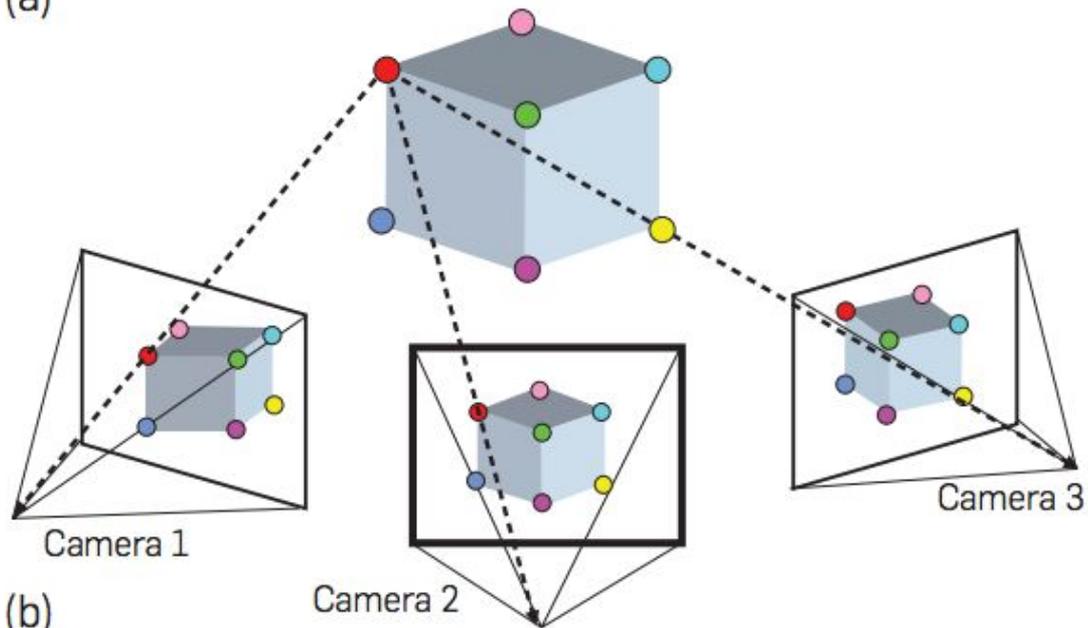
# How do we place Jupiter in the sky?

How can we combine images from different sources?

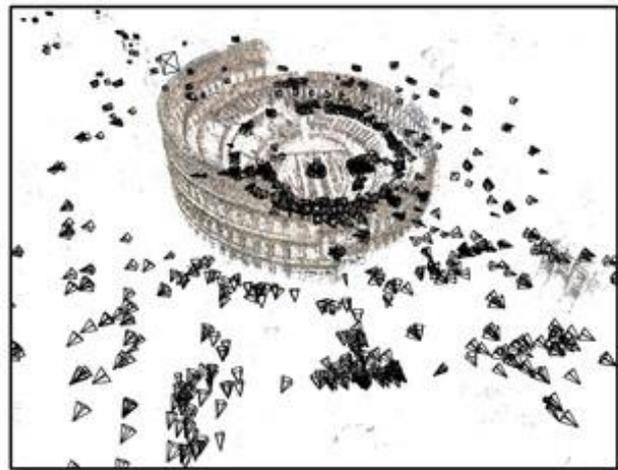
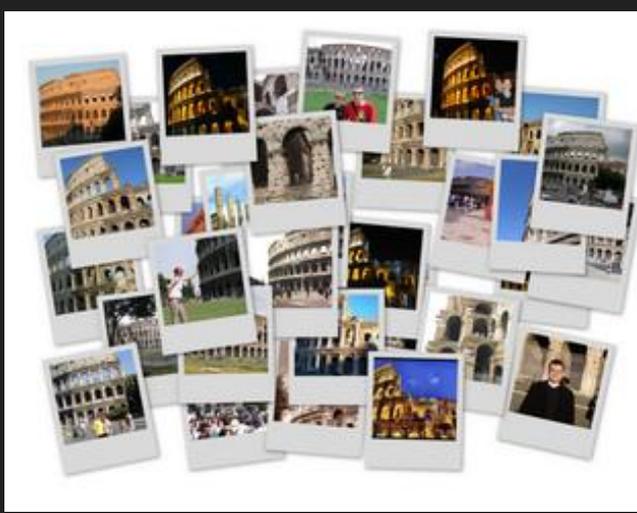
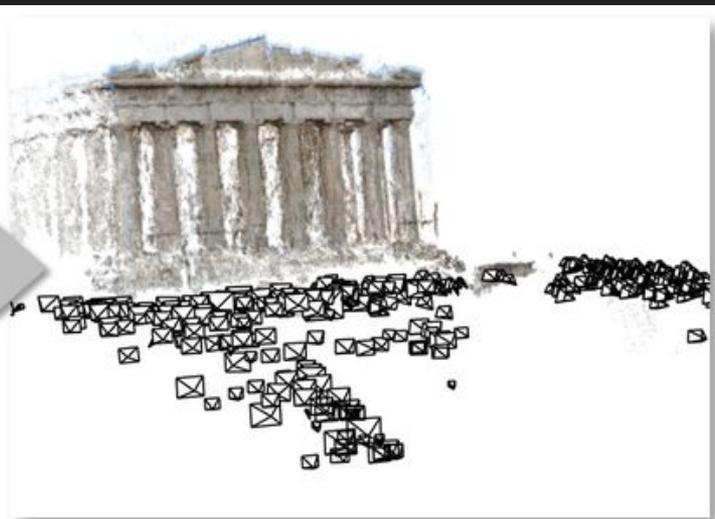
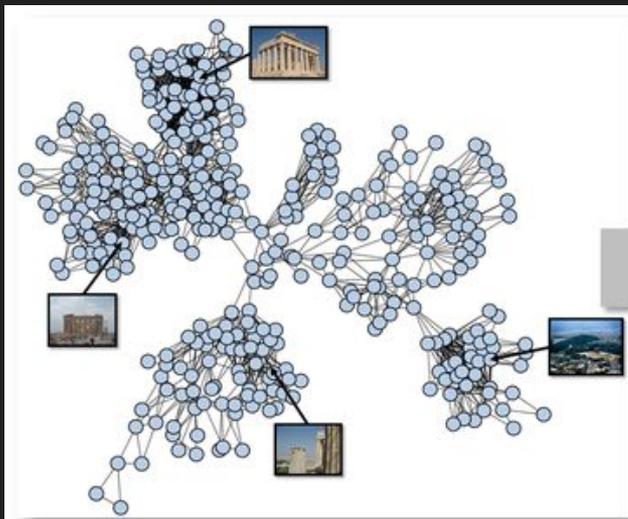
# Structure-from-Motion



(a)



(b)



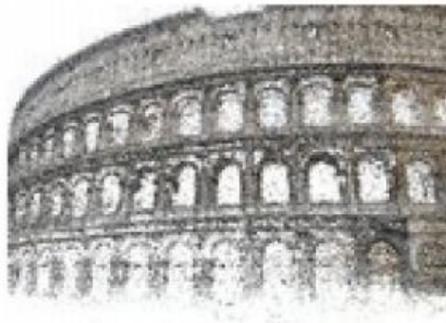
D. Crandall  
N. Snavely

Input images

SfM points

MVS points

Colosseum



St. Peter's



Timelapses!





2007

2008

2009

2010

Goldman Sachs Tower, New York City

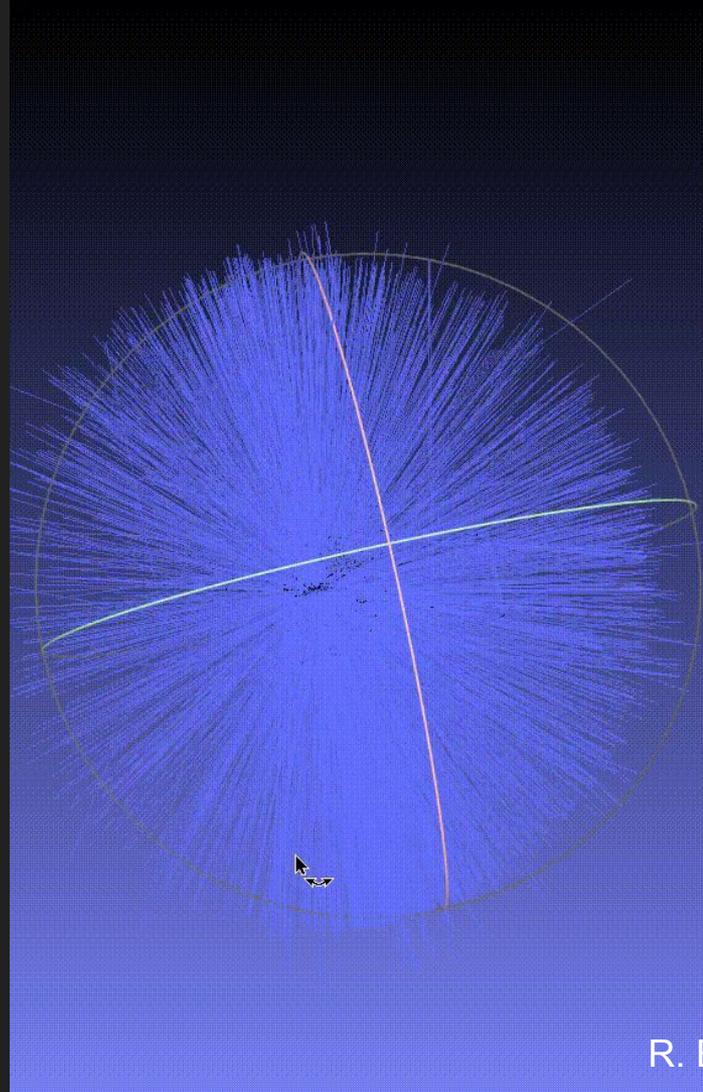
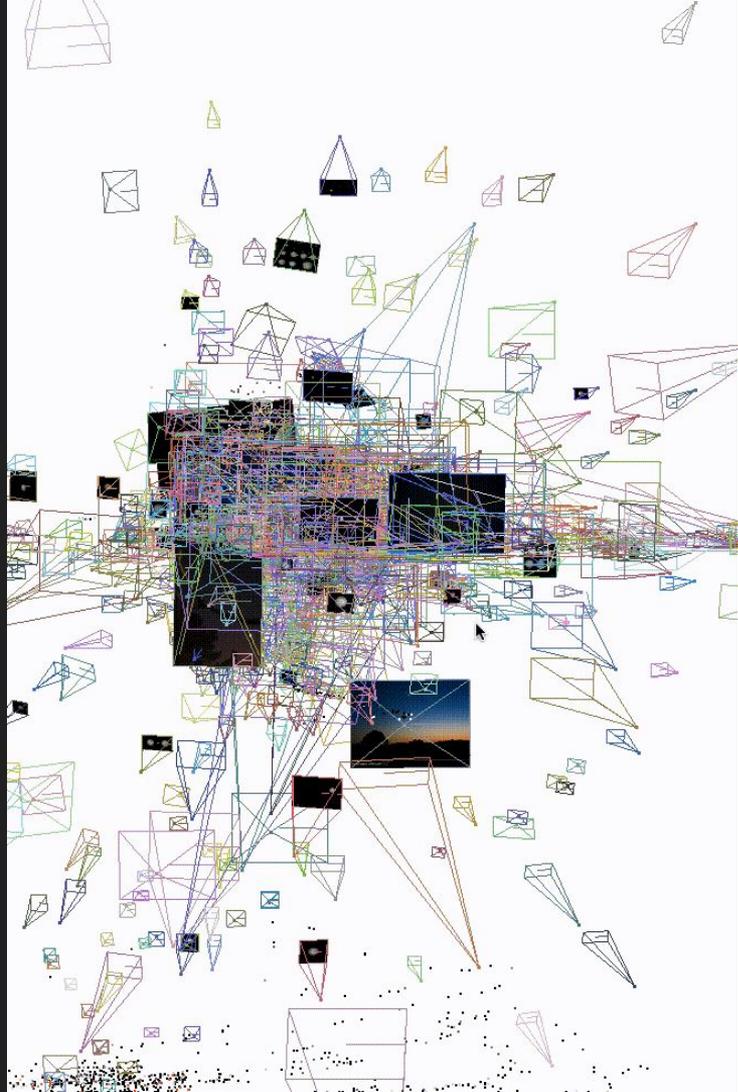
Martin-Brualla et al.



Martin-Brualla et al.



Koh Nang Yuan, Thailand





# Questions?

re2300@columbia.edu