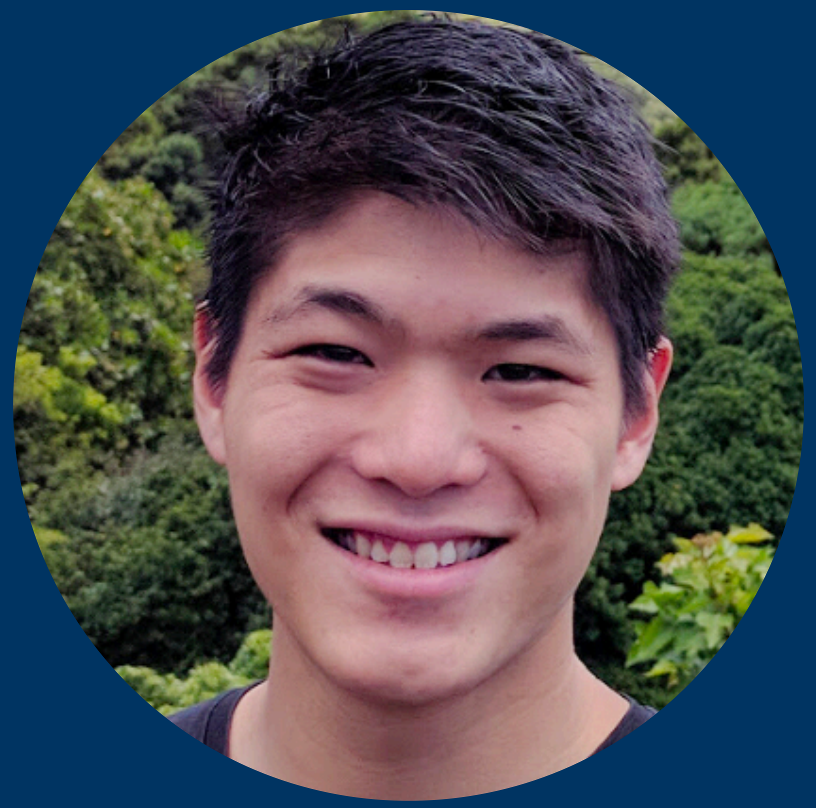
An Introduction to Machine Learning & Astronomy

Josh Peek
Head of Data Science
Space Telescope Science Institute

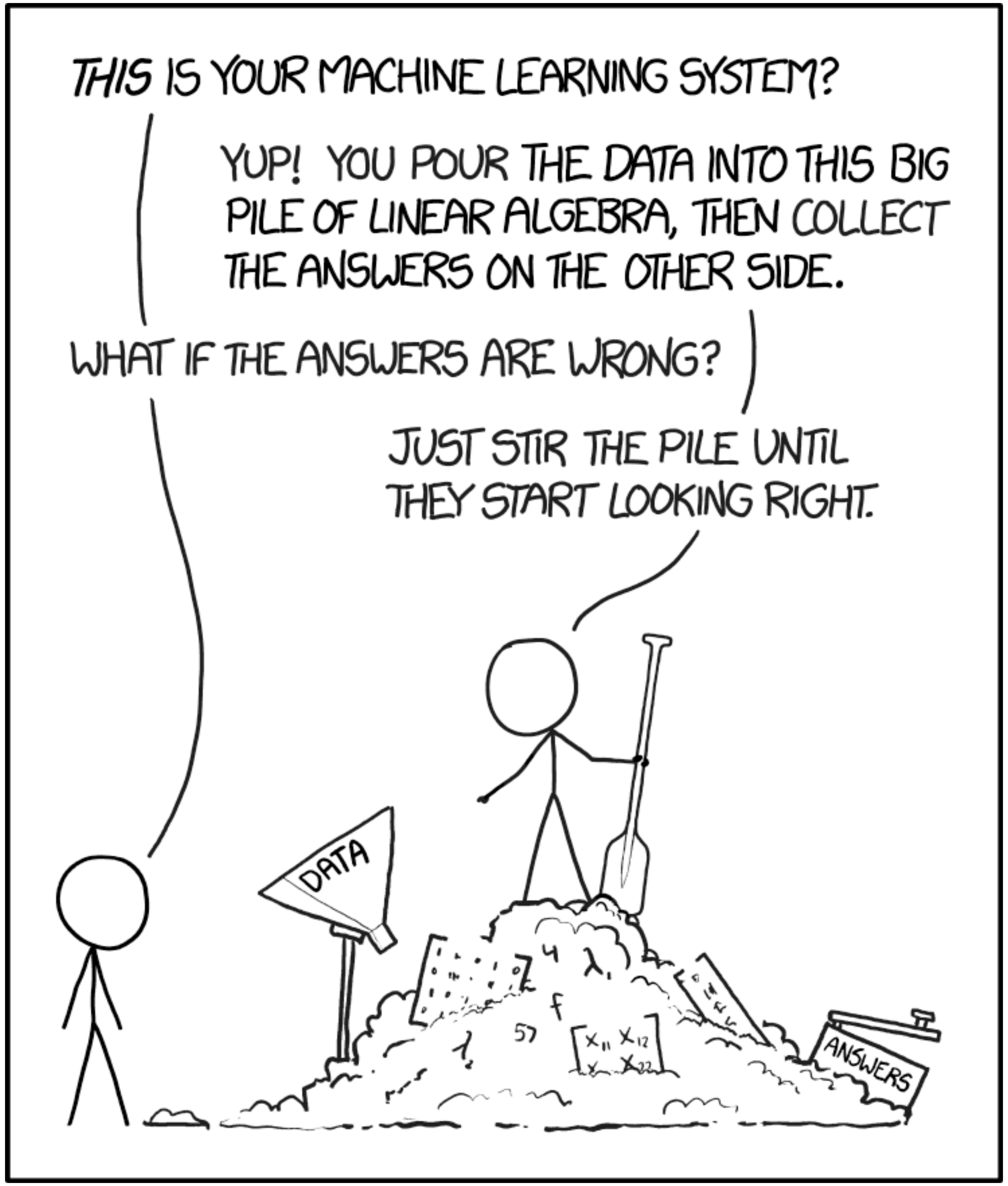


John Wu

Rick White

2024-05-20 Rare Gems 💎

What is Machine Learning, Anyway?



Randall Munroe, xkcd #1838

What is Machine Learning, Anyway?

"I saw the best minds of my generation solve deep, longstanding problems in AI in order to serve better ads"

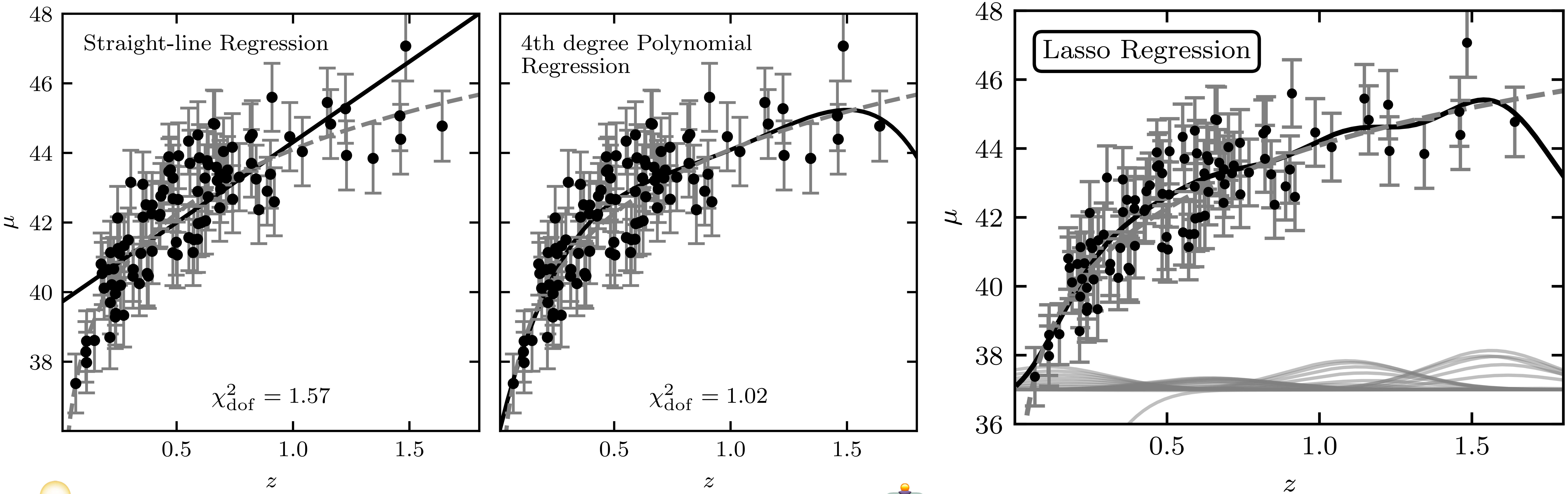
~Jeff Hammerbacher, Facebook

What is Machine Learning *in the astronomy context*?

Machine Learning is* statistics, but where *every* parameter is a *nuisance* parameter

**(in the astronomy context)*

Sometimes you want 💡, sometimes you want 🤖



💡 (z, few parameters I care about) = μ

🤖(z, many garbage parameters) = μ

Ivezic, Connolly, ***VanderPlas*** *& Gray 2020, astroml*

What is ML in Astronomy?

Supervised vs. Unsupervised

Supervised Learning in Astronomy

Supervised Learning: *Galaxy Images*

Unsupervised Learning in Astronomy

Unsupervised Learning: *Search By Image*

What is ML in Astronomy?

Supervised vs. Unsupervised

Supervised Learning in Astronomy

Supervised Learning: *Galaxy Images*

Unsupervised Learning in Astronomy

Unsupervised Learning: *Search By Image*

Machine Learning has two main branches

Supervised Learning:

Regression (= fitting)

Classification

Unsupervised Learning:

Dimensionality reduction

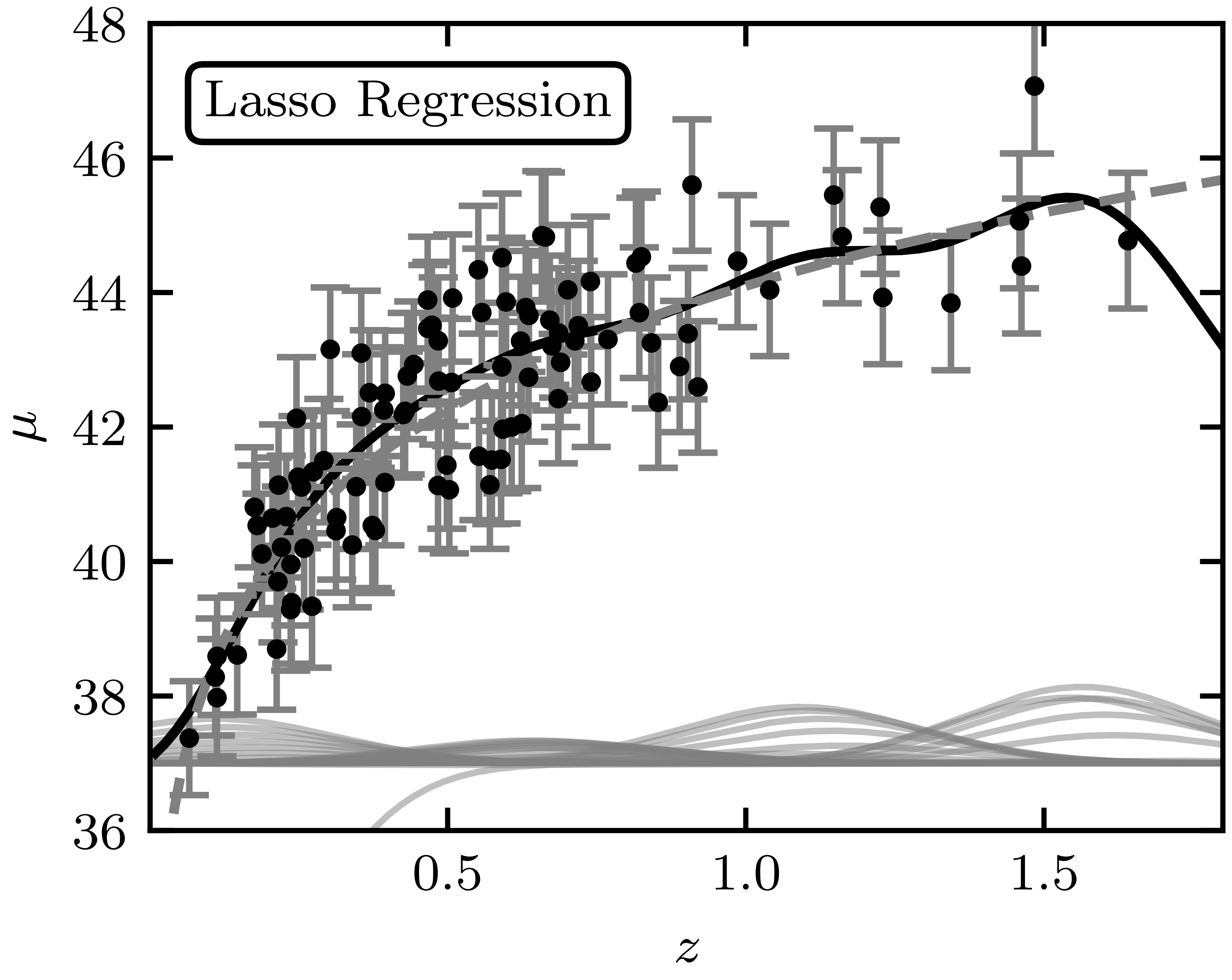
Clustering

Outlier Detection

Machine Learning has two main branches

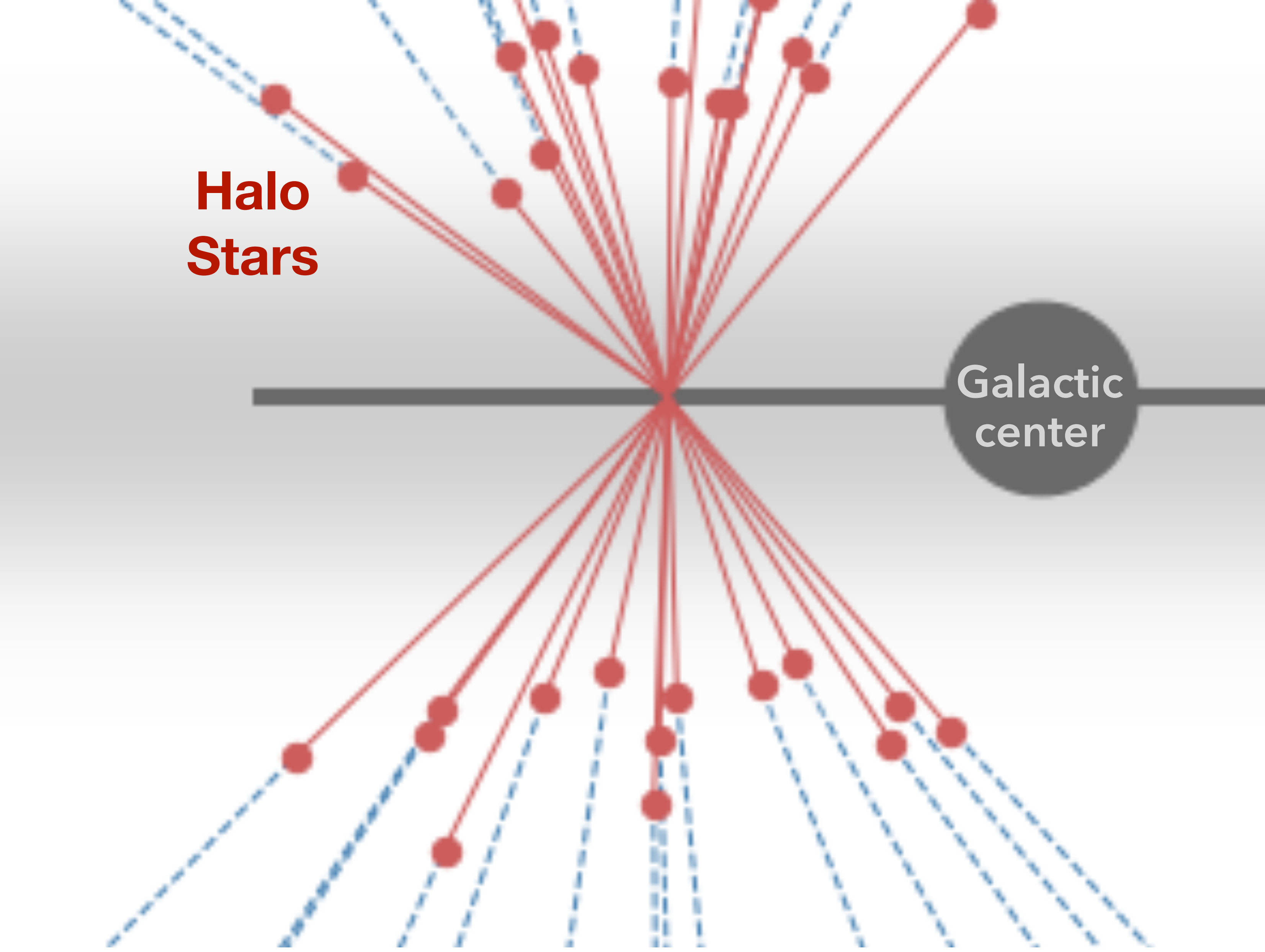
Supervised Learning:
- Regression (= fitting)
- Classification

Unsupervised Learning:
- Dimensionality reduction
- Clustering
- Outlier Detection



*Ivezic, Connolly, **VanderPlas** & Gray 2020, astroml*

Machine Learning has two main branches



Supervised Learning:

Regression (= fitting)

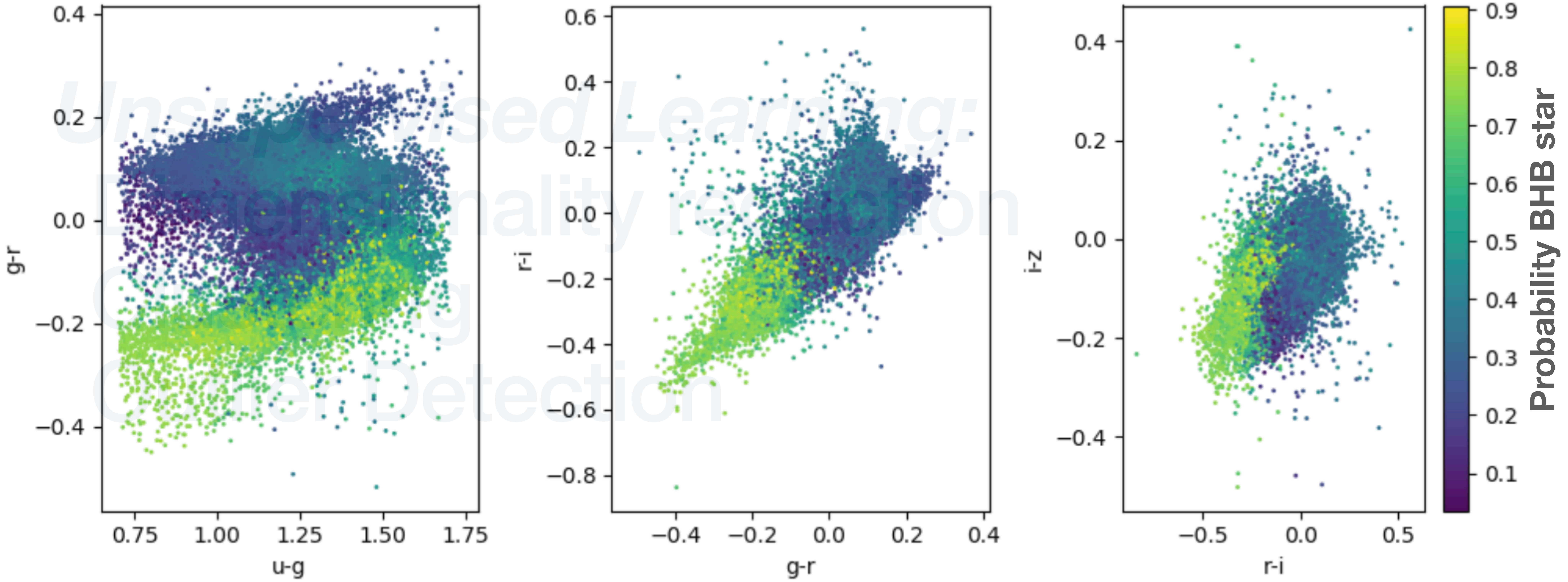
Classification

Unsupervised Learning:

Dimensionality reduction

Clustering

Outlier Detection



JEGP, HST 15656

Machine Learning has two main branches

Supervised Learning:

Regression (= fitting)

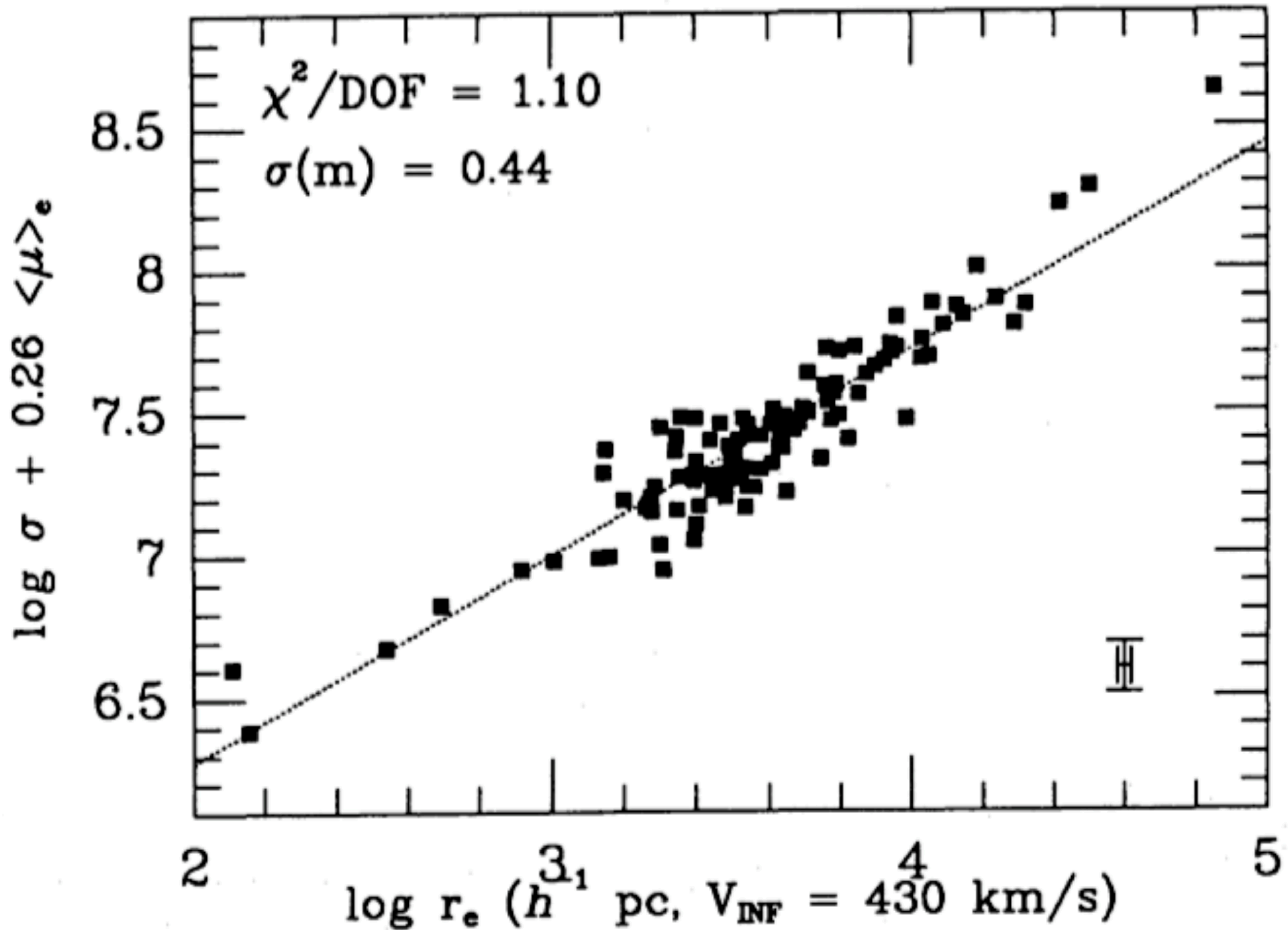
Classification

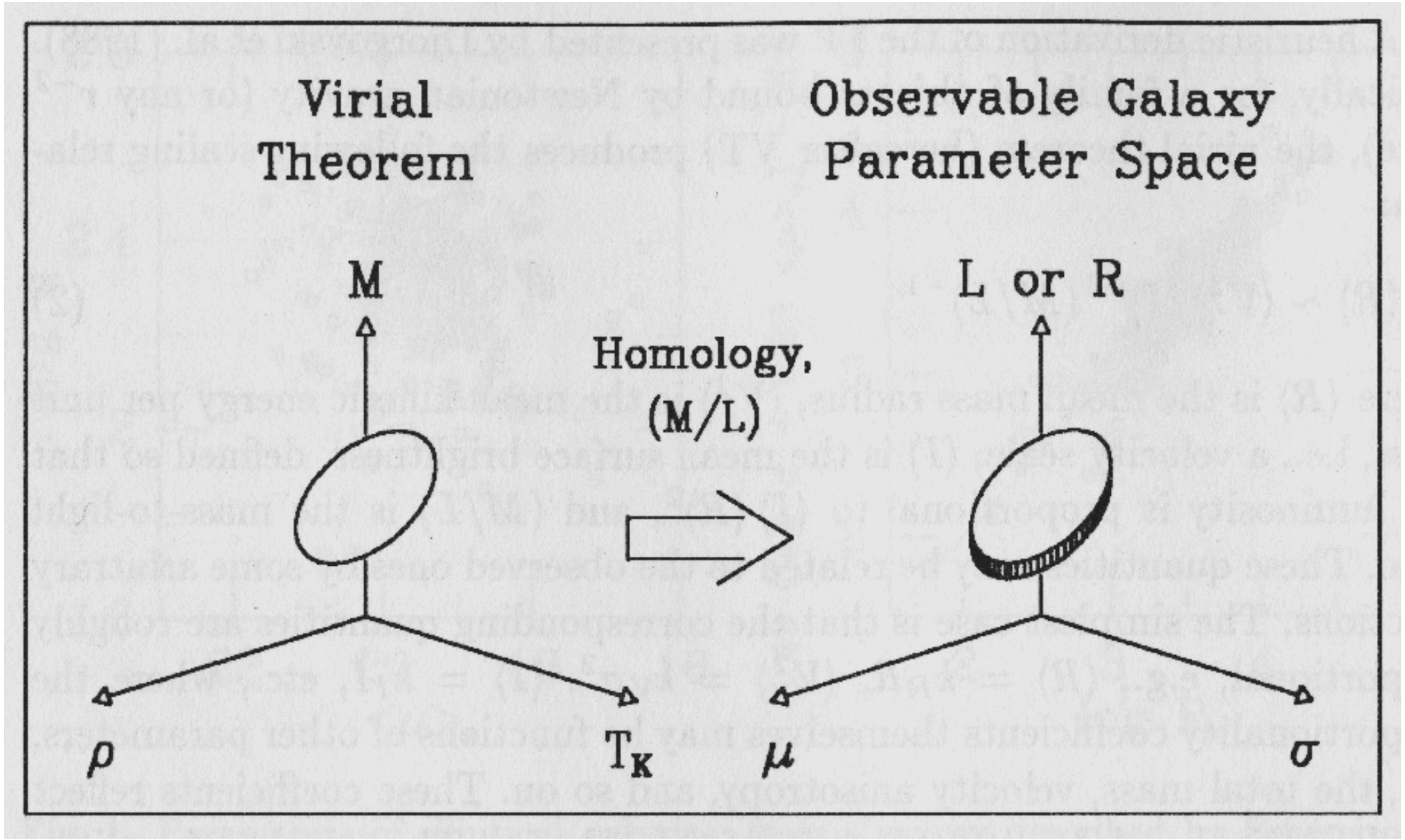
Unsupervised Learning:

Dimensionality reduction

Clustering

Outlier Detection





Djorgovski & Davis 1987

Machine Learning has two main branches

Supervised Learning:

Regression (= fitting)

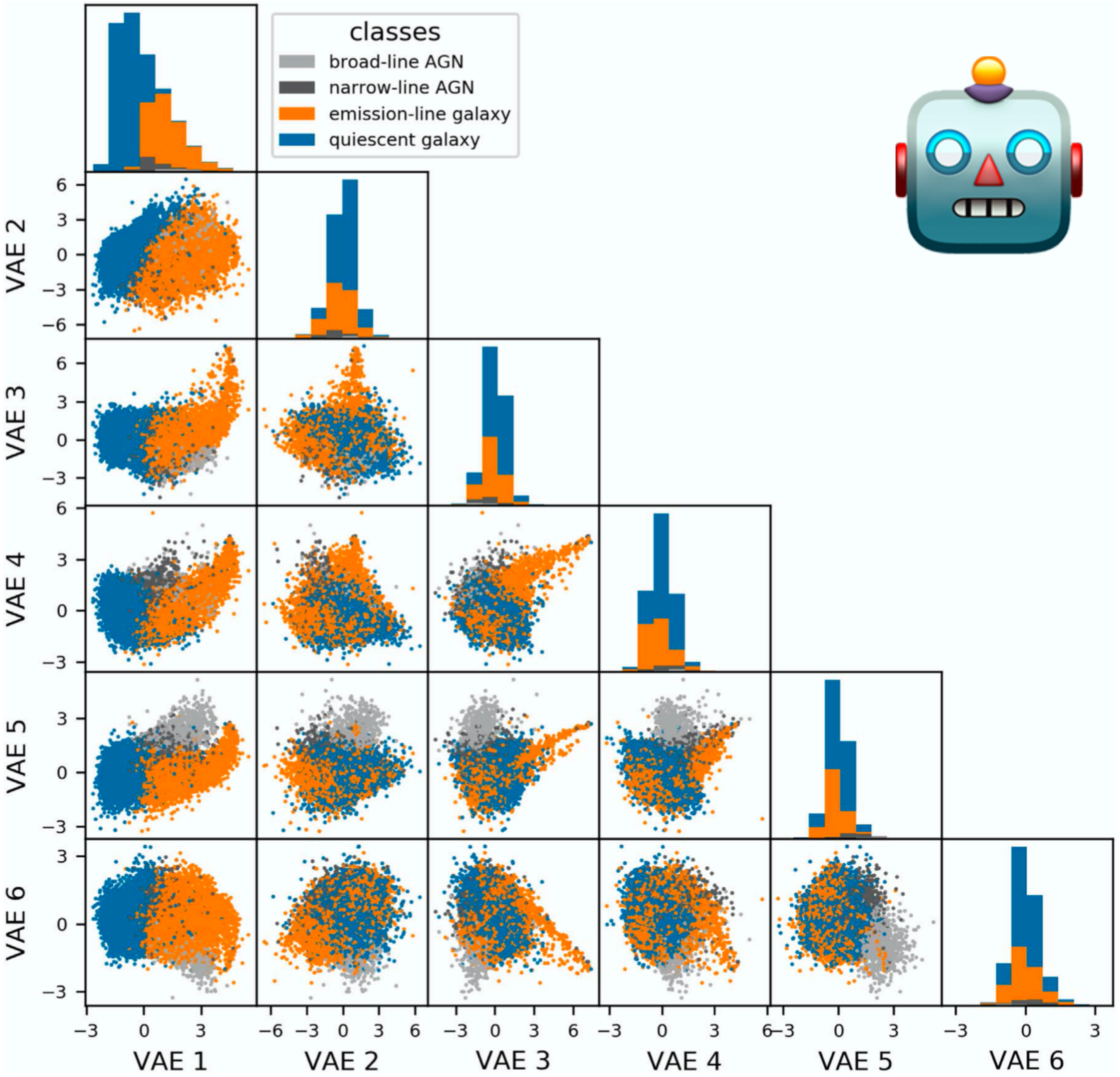
Classification

Unsupervised Learning:

Dimensionality reduction

Clustering

Outlier Detection

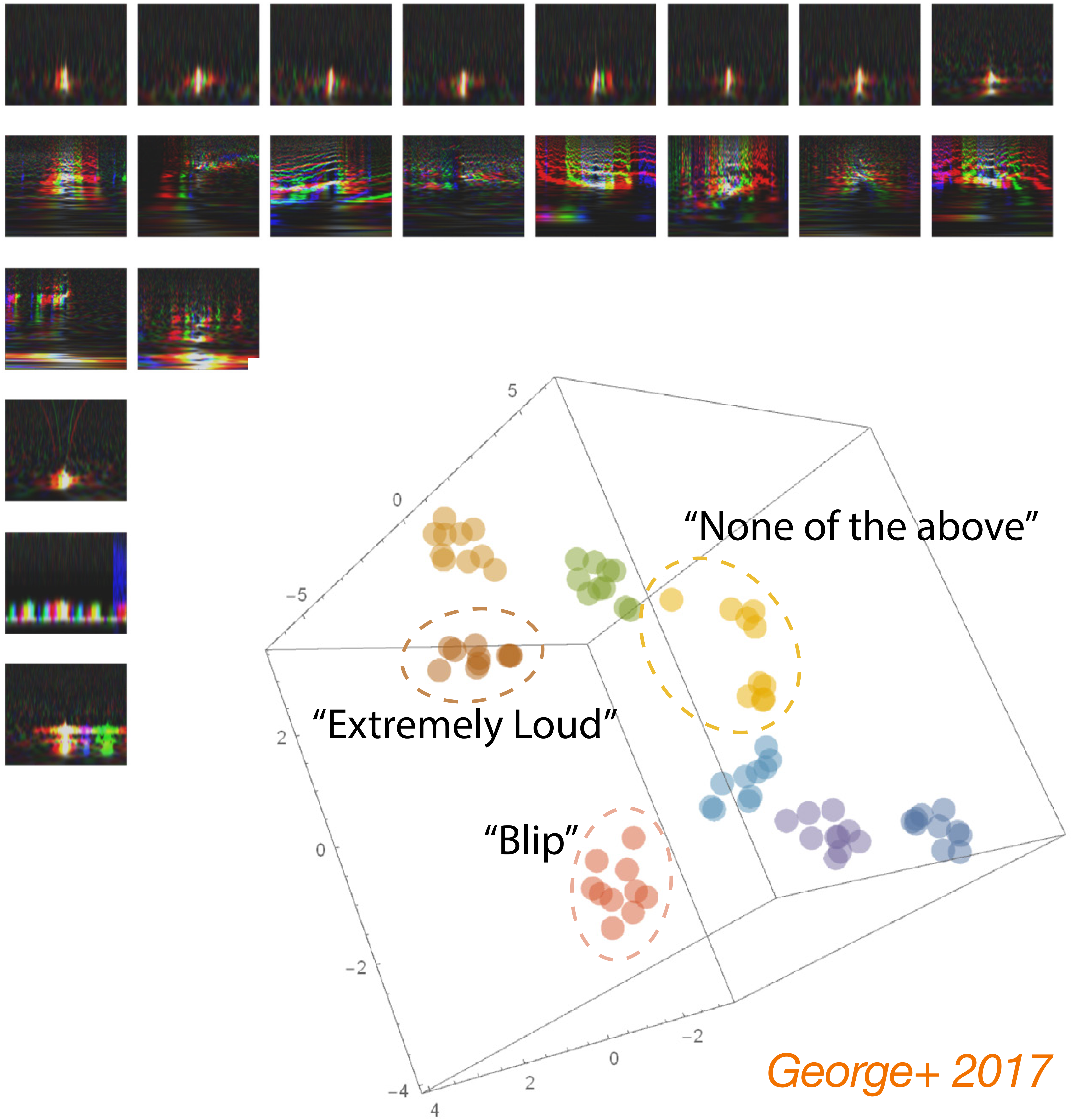


Portillo+ 2020

Machine Learning has two main branches

Supervised Learning:
- Regression (= fitting)
- Classification

Unsupervised Learning:
- Dimensionality reduction
- Clustering
- Outlier Detection



George+ 2017

A Brief Comment on Modality in Astronomy:

*The cosmos is **weakly** modal compared with human experience*

Machine Learning has two main branches

Supervised Learning:

Regression (= fitting)

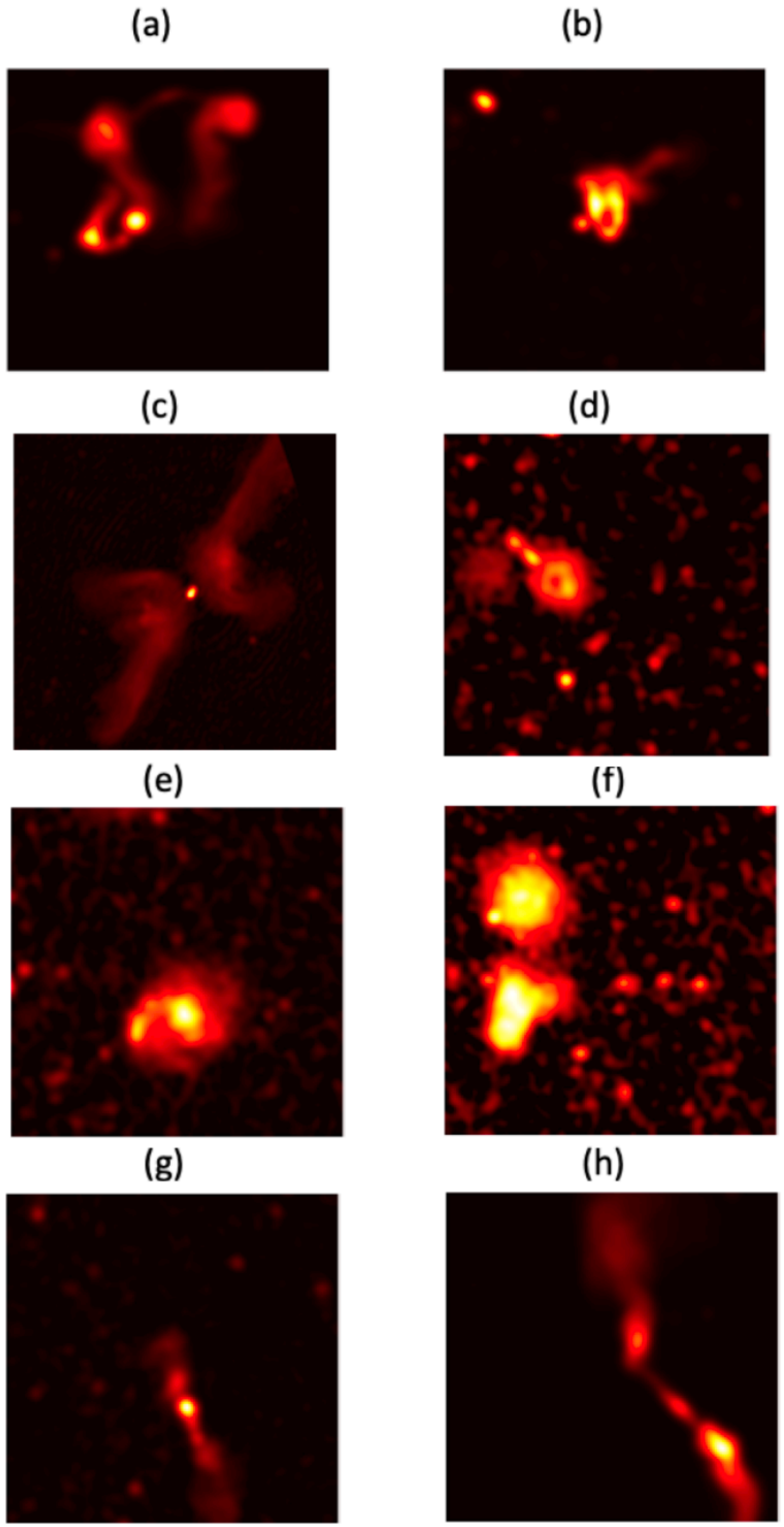
Classification

Unsupervised Learning:

Dimensionality reduction

Clustering

Outlier Detection



Segal+2023

Machine Learning ~~has two main branches~~

has been eaten by self-supervised learning!

Supervised Learning:

I propose an investigation into the role of dark matter mini-halos in the formation and evolution of globular clusters (GCs) in the Milky Way. This study would combine the high-precision astrometric data from Gaia EDR3 and DR3 (Gaia Collaboration et al. 2016, 2020) with spectroscopic data from large ground-based surveys like APOGEE, GALAH, SDSS SEGUE, and LAMOST to characterize the dynamics of stars in the peripheral regions of GCs. The aim would be to determine whether these GCs are embedded in dark matter mini-halos, which could provide critical insights into their origins (Peebles 1984; Peñarrubia et al. 2017). The proposed research would build upon the probabilistic approach developed by Kuzma et al. (2021) for studying the peripheral regions of GCs, which utilizes a mixture model in spatial and proper motion space to model cluster, extra-tidal, and contaminant stellar populations. By extending this approach to include the effects of dark matter mini-halos on the kinematics of stars in GC outskirts, we can test the hypothesis that dark matter plays a significant role in the formation and evolution of GCs. Furthermore, this study would provide a better understanding of the distribution and properties of dark matter in the Milky Way, contributing to the broader field of near-field cosmology. Integrating this information with the existing knowledge of the hierarchical assembly of the Milky Way (Viswanathan et al. 2023) and the role of rapid gas accretion in the inner Galactic disc (Snaith et al. 2021) would help paint a more comprehensive picture of our Galaxy's formation history and its underlying dark matter distribution.

Ciuca et al. 2023

What is ML in Astronomy?

Supervised vs. Unsupervised

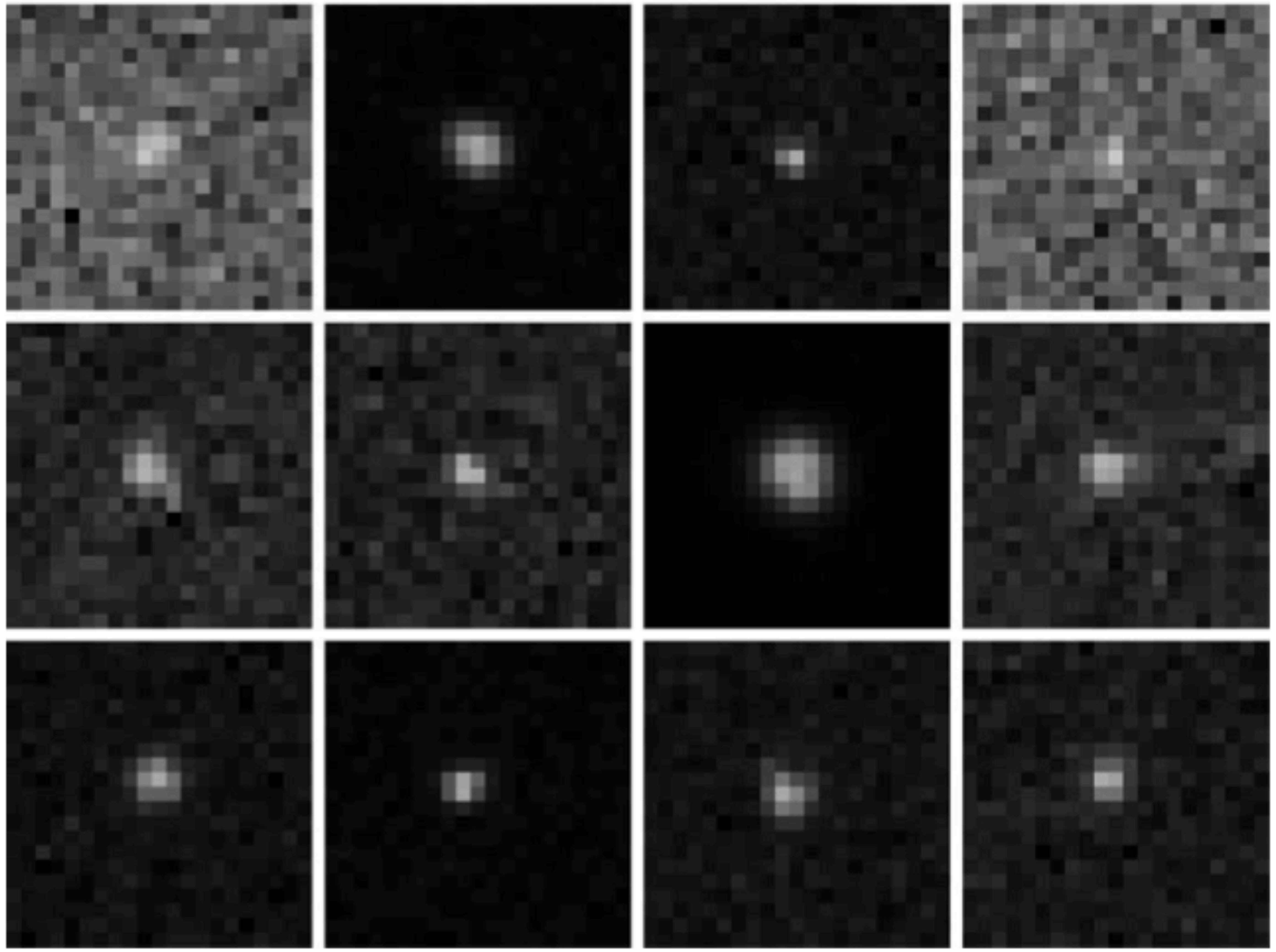
Supervised Learning in Astronomy

Supervised Learning: *Galaxy Images*

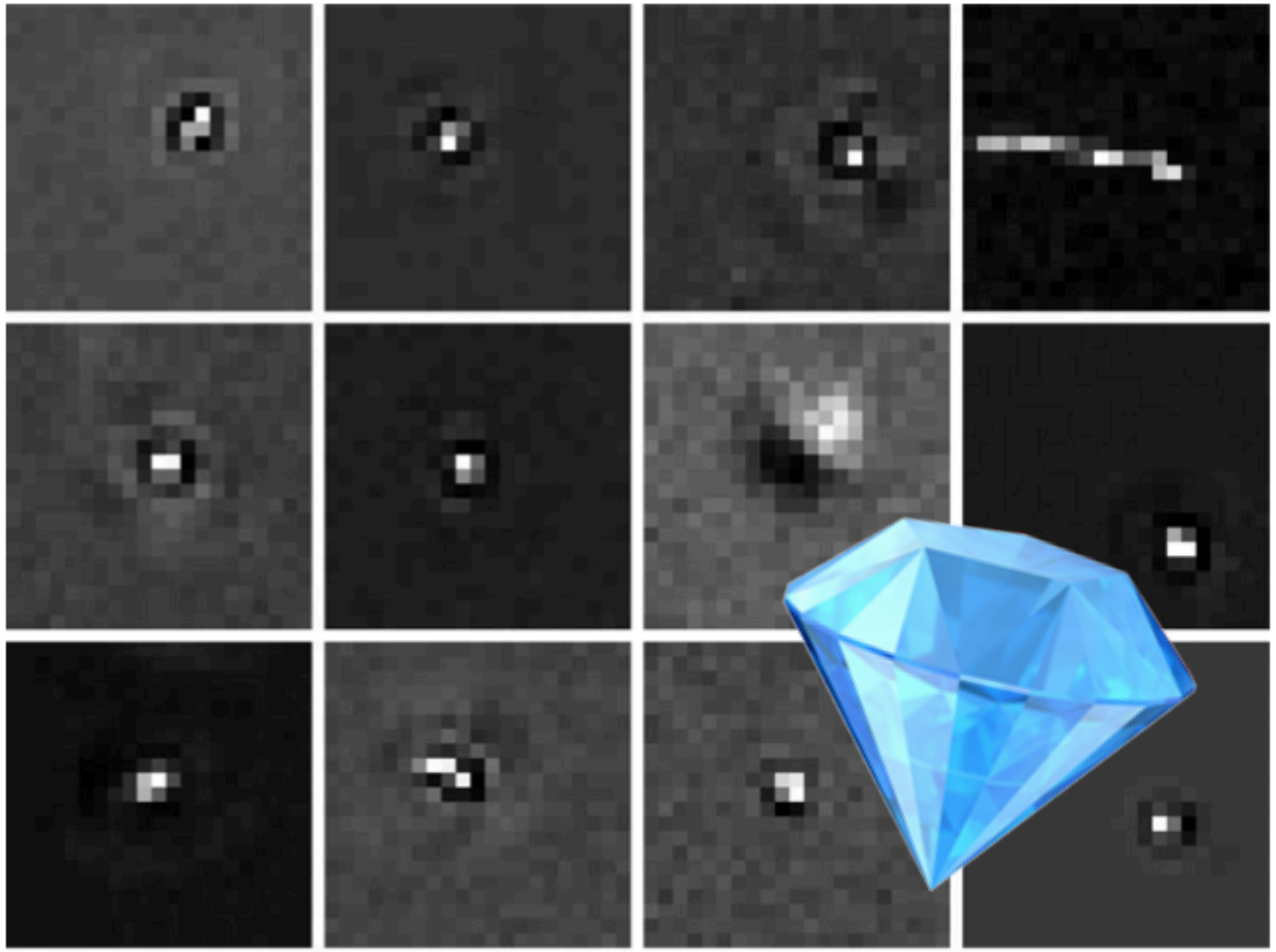
Unsupervised Learning in Astronomy

Unsupervised Learning: *Search By Image*

Real-Bogus: fake vs. real transients with ML



Real



Bogus

Brink + 2013

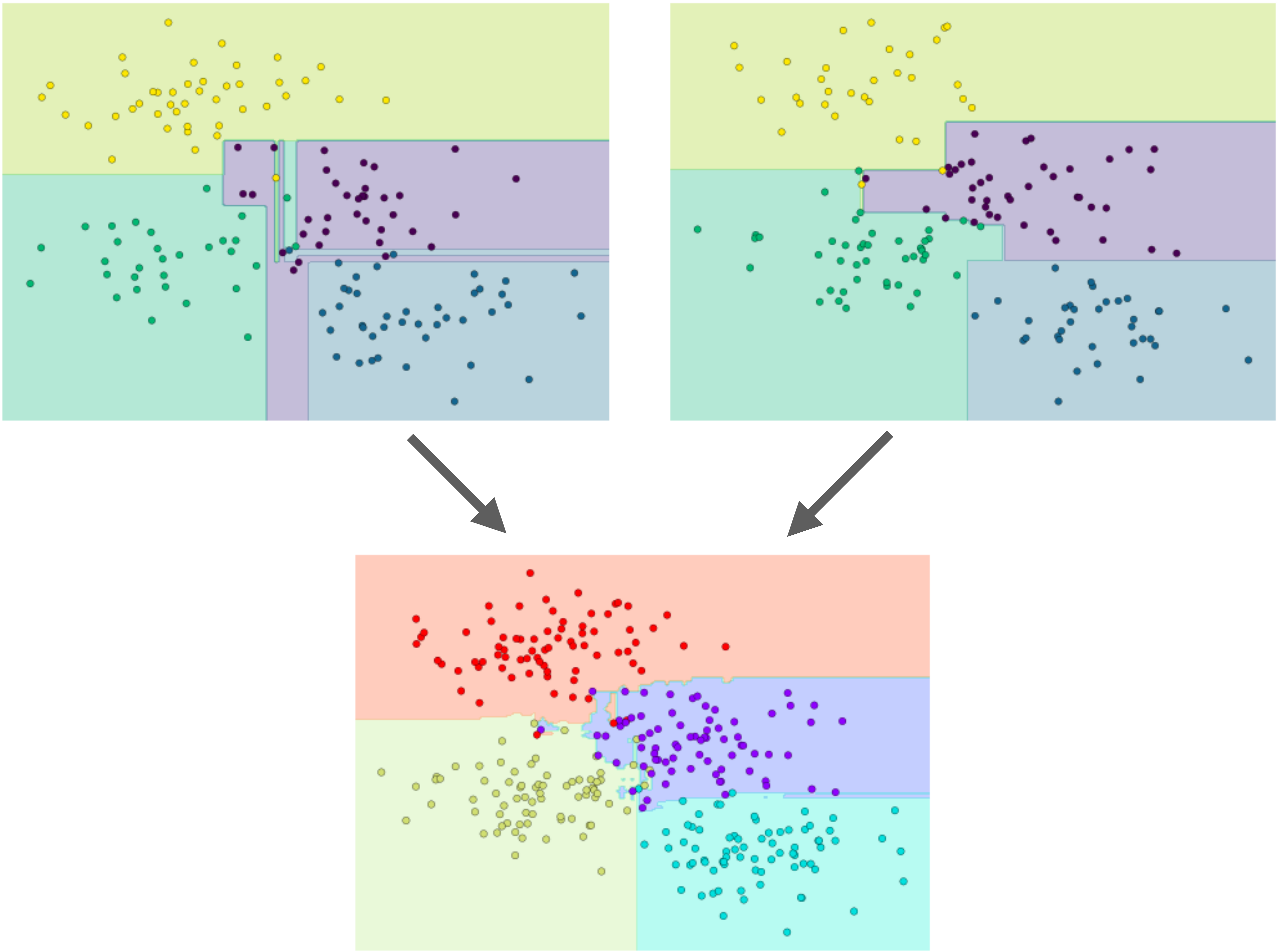
Real-Bogus: Based on hand-built 30+ dimensional feature vectors

RB1		mag	USNO-B1.0 derived magnitude of the candidate on the difference image
		mag_err	Estimated uncertainty on mag
		a_image	Semimajor axis of the candidate
	✓	b_image	Semiminor axis of the candidate
		fwhm	Full width at half-maximum (FWHM) of the candidate
	✓	flag	Numerical representation of the SExtractor extraction flags
	✓	mag_ref	Magnitude of the nearest object in the reference image if less than 5 arcsec from the candidate
	✓	mag_ref_err	Estimated uncertainty on mag_ref
	✓	a_ref	Semimajor axis of the reference source
	✓	b_ref	Semiminor axis of the reference source
		n2sig3	Number of at least negative 2σ pixels in a 5×5 box centred on the candidate
		n3sig3	Number of at least negative 3σ pixels in a 5×5 box centred on the candidate
		n2sig5	Number of at least negative 2σ pixels in a 7×7 box centred on the candidate
	✓	n3sig5	Number of at least
	✓	flux_ratio	Ratio of the apertur of the reference sou
		ellipticity	Ellipticity of the ca
	✓	ellipticity_ref	Ellipticity of the ref
	✓	nn_dist_renorm	Distance in arcseco
		magdiff	When a reference s magnitude and the Else, the difference and the limiting ma
	✓	maglim	True if there is no r
		sigflux	Significance of the estimated uncertain
		seeing_ratio	Ratio of the FWHM of the seeing on the
	✓	mag_from_limit	Limiting magnitude
		normalized_fwhm	Ratio of the FWHM
	✓	normalized_fwhm_ref	Ratio of the FWHM of the reference source to the seeing in the reference image
	✓	good_cand_density	Ratio of the number of candidates in that subtraction to the total usable area on that array
	✓	min_distance_to_edge_in_new	Distance in pixels to the nearest edge of the array on the new image

RB2	✓	**ccdid**	Numerical ID of the specific camera detector (1–12)
		sym	Measure of symmetry, based on dividing the object into quadrants
	✓	seeingnew	FWHM of the seeing on the new image
	✓	extracted	Number of candidates on that exposure found by SExtractor
	✓	obsaved	Number of candidates on that exposure saved to the data base (a subset of extracted)
		pos	True for a positive (i.e. brighter) residual, False for a negative (fading) one
	✓	gauss	Gaussian best-fitting sqaured difference value
		corr	Gaussian best-fitting correlation value
		scale	Gaussian scale value
	✓	amp	Gaussian amplitude value
	✓	l1	Sum of absolute pixel values
		smooth1	Filter 1 output
		smooth2	Filter 2 output
		pca1	First principal component
		pca2	Second principal component
Test		empty	Zero for all candidates (i.e. no information)
		random	A random number generated for every candidate (i.e. pure noise)

Brink + 2013

Real-Bogus: Feed these vectors to a Random Forest classifier

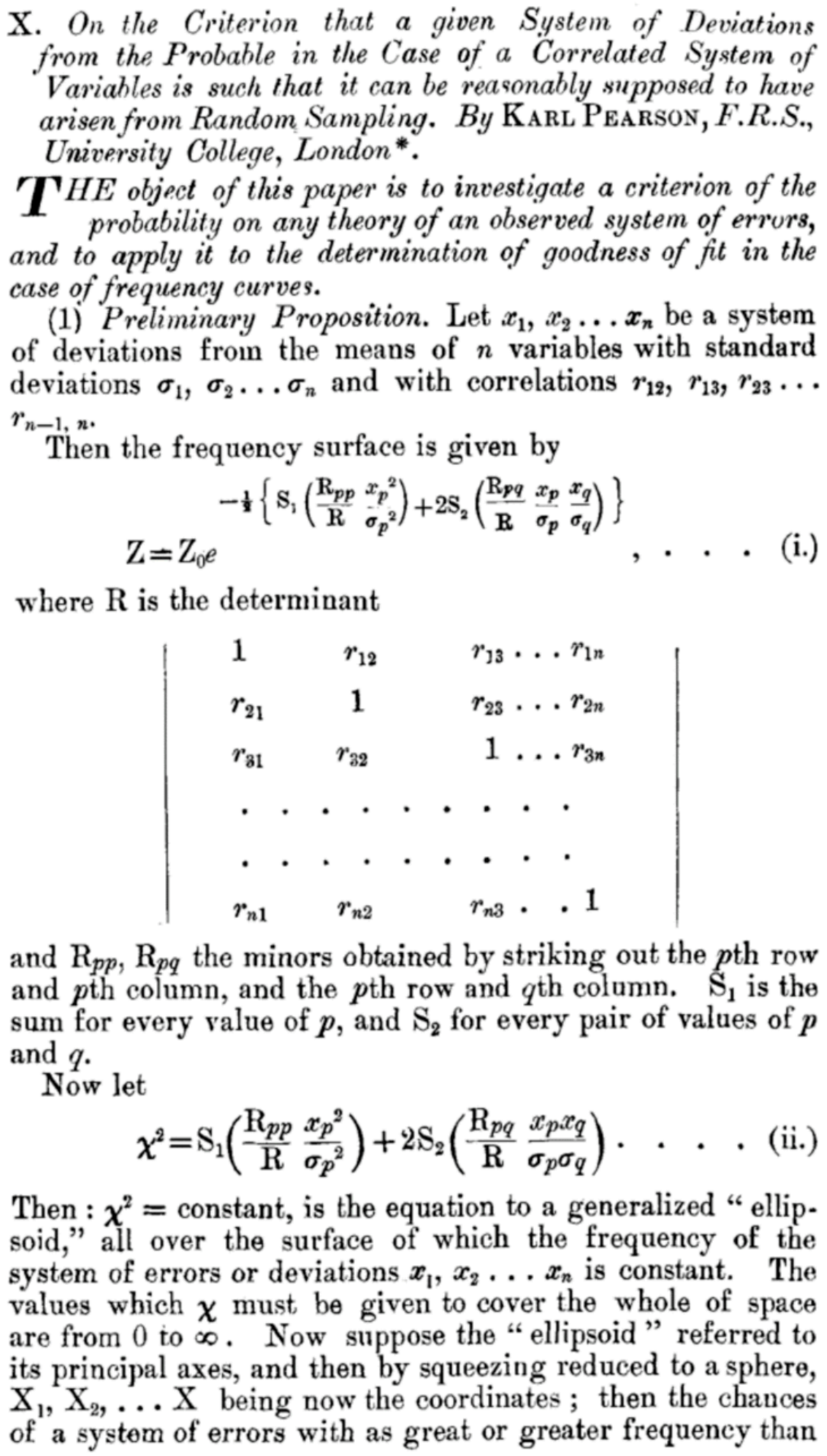


Images from jakevdp

A Brief Interlude on the Hegemony of Homoscedasticity

Homoscedasticity in ML and Stats: the astronomer's bane

$\chi 2$: errors in 1D



X. *On the Criterion that a given System of Deviations from the Probable in the Case of a Correlated System of Variables is such that it can be reasonably supposed to have arisen from Random Sampling.* By KARL PEARSON, *F.R.S., University College, London**.

THE object of this paper is to investigate a criterion of the probability on any theory of an observed system of errors, and to apply it to the determination of goodness of fit in the case of frequency curves.

(1) *Preliminary Proposition.* Let $x_1, x_2 \dots x_n$ be a system of deviations from the means of n variables with standard deviations $\sigma_1, \sigma_2 \dots \sigma_n$ and with correlations $r_{12}, r_{13}, r_{23} \dots r_{n-1,\,n}$.

Then the frequency surface is given by

$$Z = Z_0 e^{-\frac{1}{2}\left\{S_1\left(\frac{R_{pp}}{R}\frac{x_p^2}{\sigma_p^2}\right) + 2S_2\left(\frac{R_{pq}}{R}\frac{x_p}{\sigma_p}\frac{x_q}{\sigma_q}\right)\right\}} \quad , \quad . \quad . \quad . \quad \text{(i.)}$$

where R is the determinant

$$\begin{vmatrix} 1 & r_{12} & r_{13} \dots r_{1n} \\ r_{21} & 1 & r_{23} \dots r_{2n} \\ r_{31} & r_{32} & 1 \dots r_{3n} \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ r_{n1} & r_{n2} & r_{n3} \dots 1 \end{vmatrix}$$

and R_{pp}, R_{pq} the minors obtained by striking out the pth row and pth column, and the pth row and qth column. S_1 is the sum for every value of p, and S_2 for every pair of values of p and q.

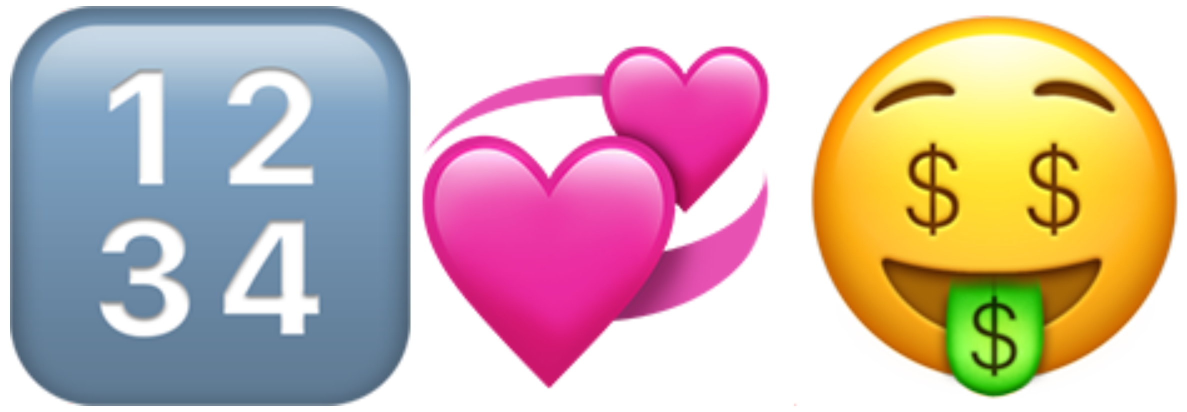
Now let

$$\chi^2 = S_1\left(\frac{R_{pp}}{R}\frac{x_p^2}{\sigma_p^2}\right) + 2S_2\left(\frac{R_{pq}}{R}\frac{x_p x_q}{\sigma_p \sigma_q}\right) . \quad . \quad . \quad . \quad \text{(ii.)}$$

Then: χ^2 = constant, is the equation to a generalized "ellipsoid," all over the surface of which the frequency of the system of errors or deviations $x_1, x_2 \dots x_n$ is constant. The values which χ must be given to cover the whole of space are from 0 to ∞. Now suppose the "ellipsoid" referred to its principal axes, and then by squeezing reduced to a sphere, $X_1, X_2, \dots X$ being now the coordinates; then the chances of a system of errors with as great or greater frequency than

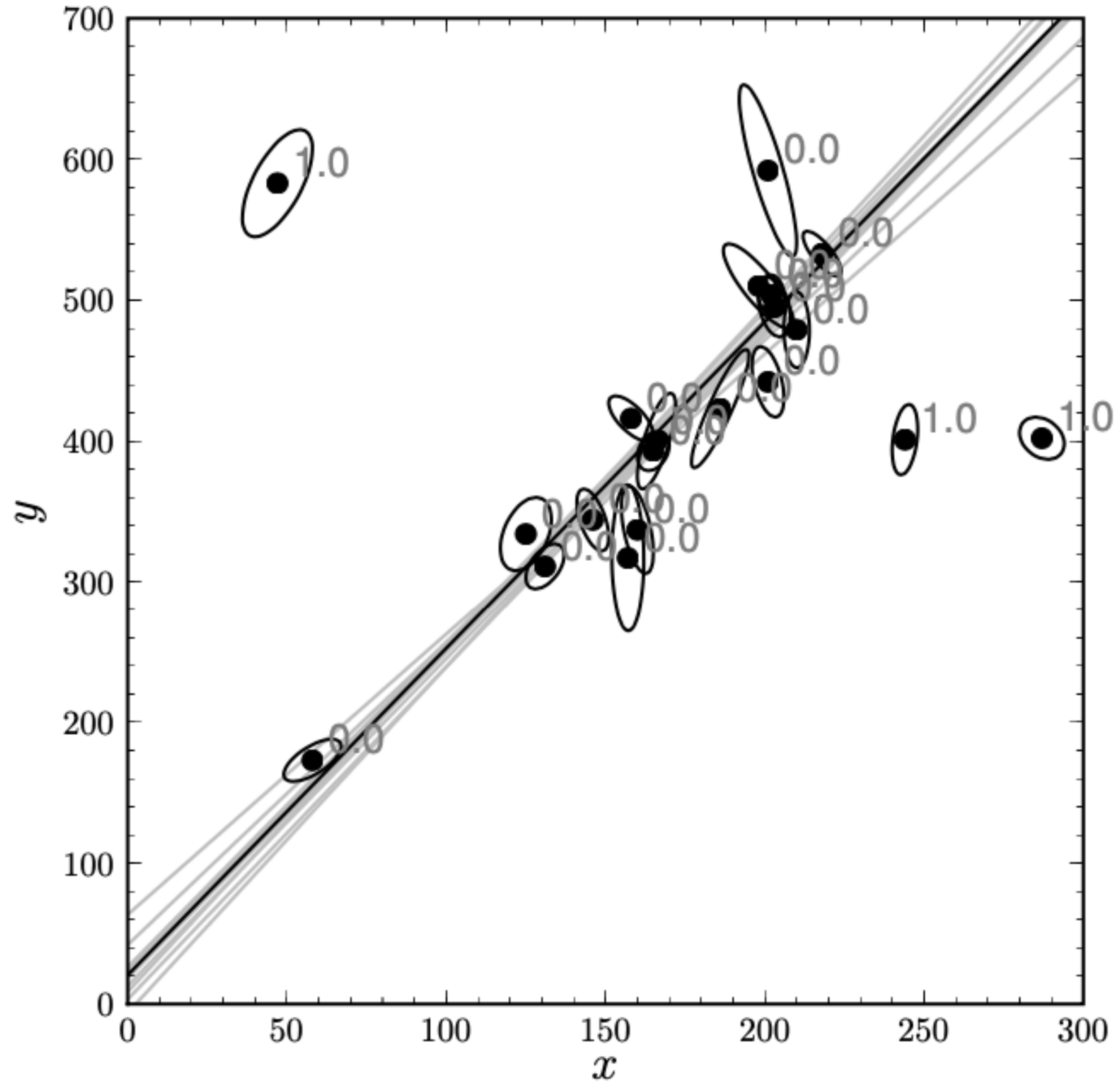
Pearson 1900

Economics



Everyone, entirety of 20th c.

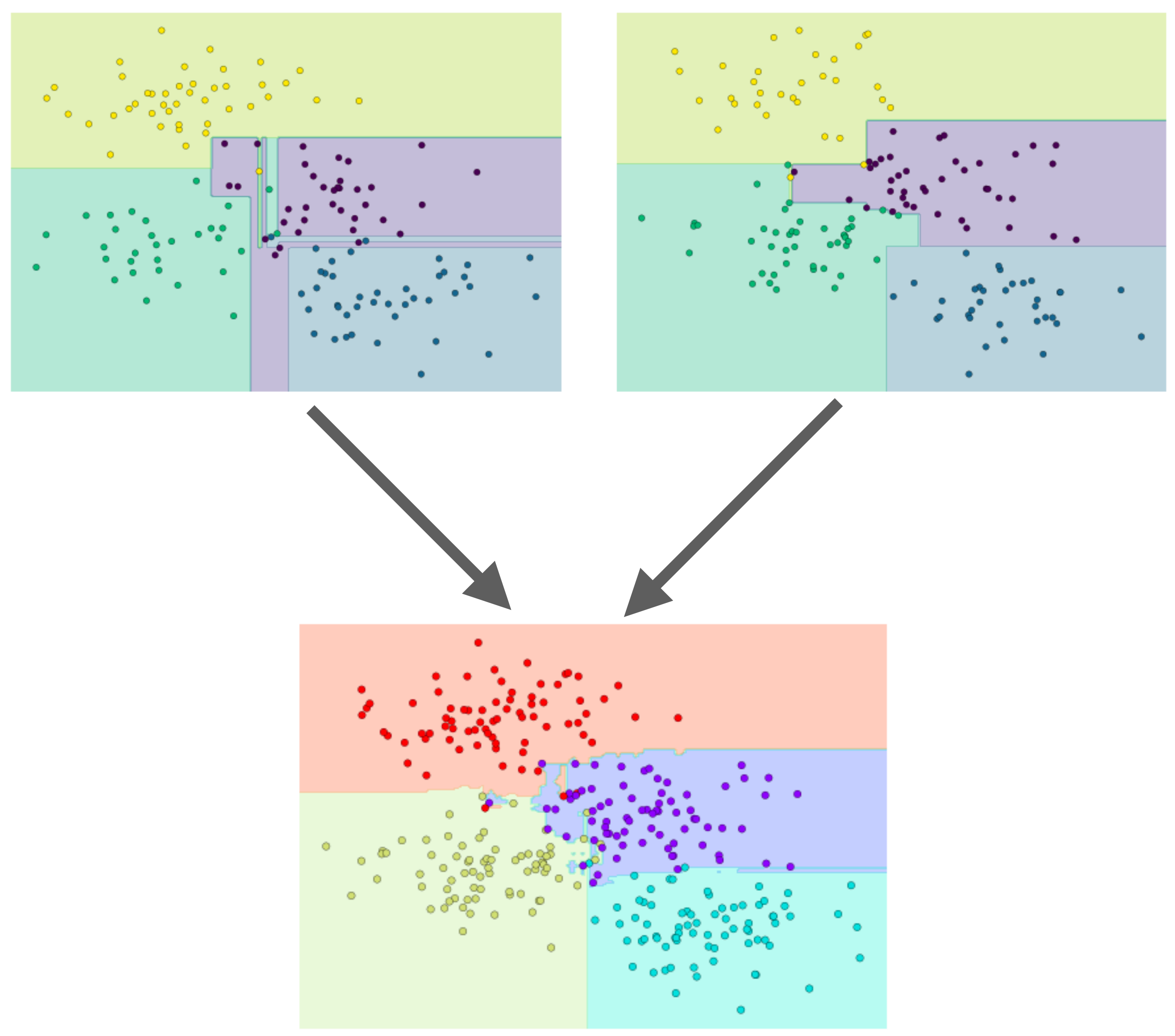
errors in 2D



Hogg, Bovy, & Lang 2010

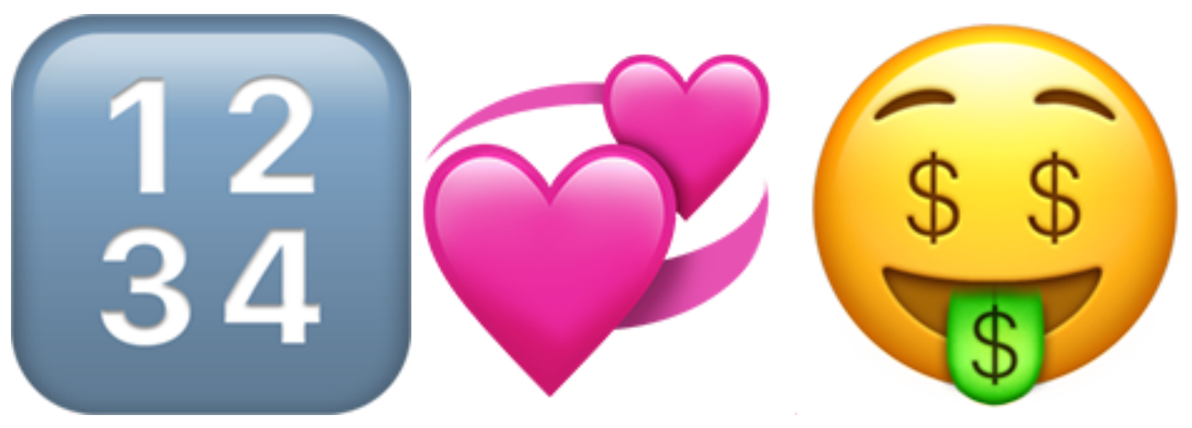
Homoscedasticity in ML and Stats: the astronomer's bane

Random Forest



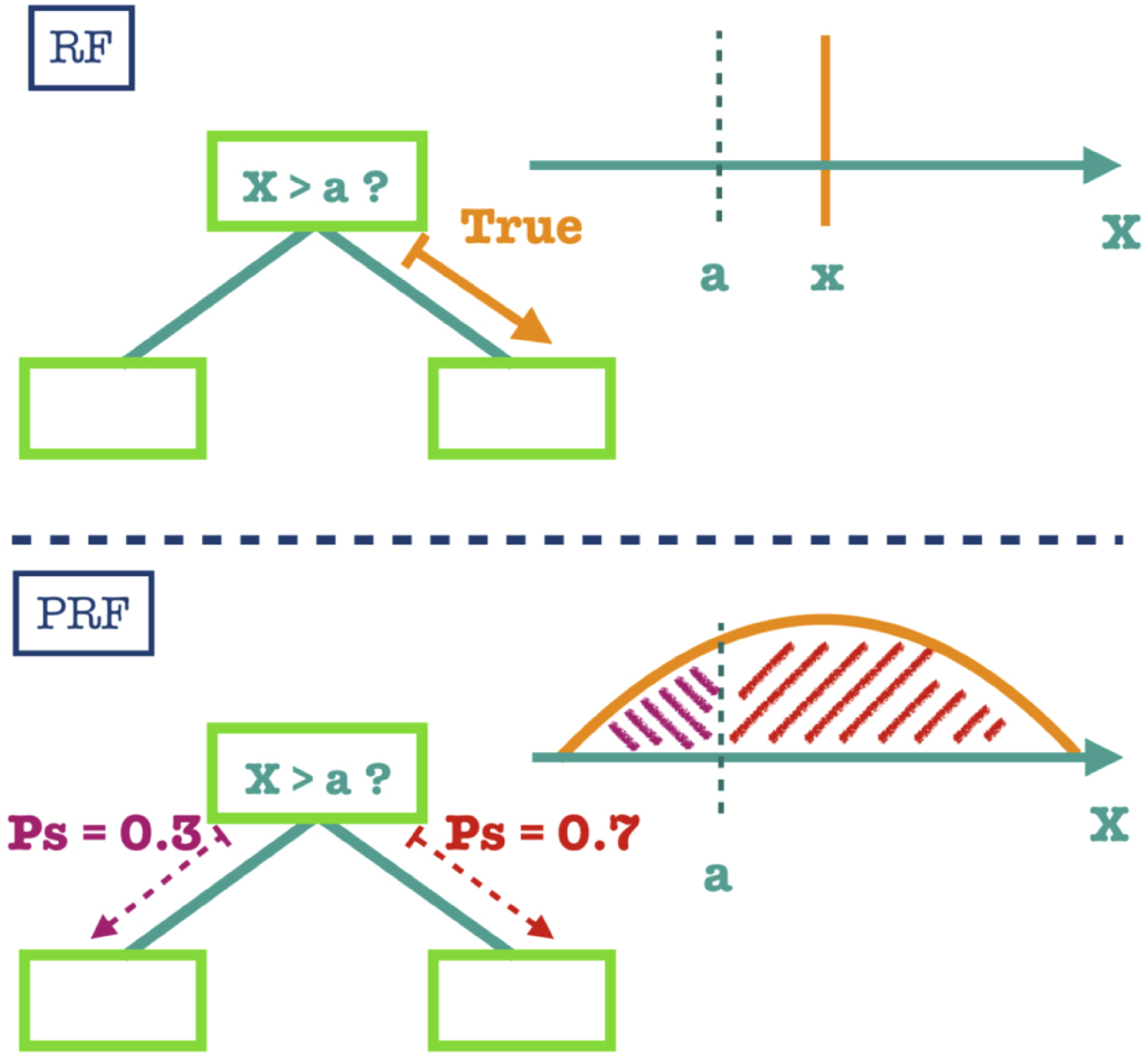
Ho 1995

Economics



Everyone, forever, apparently

Probabilistic RF

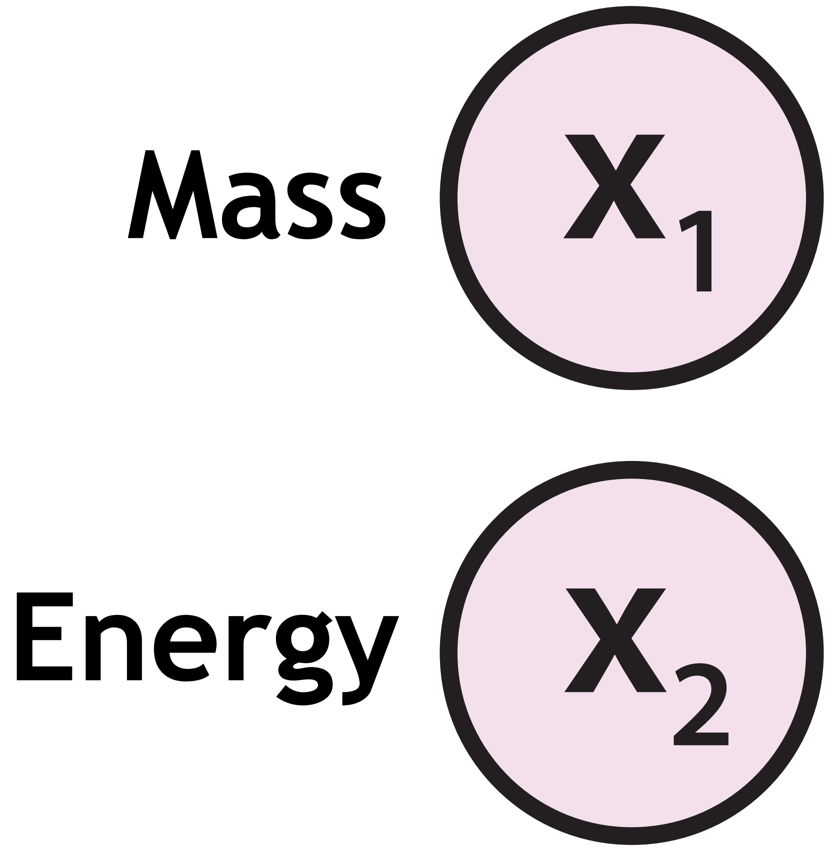


Reis, Baron, & Shahaf 2017

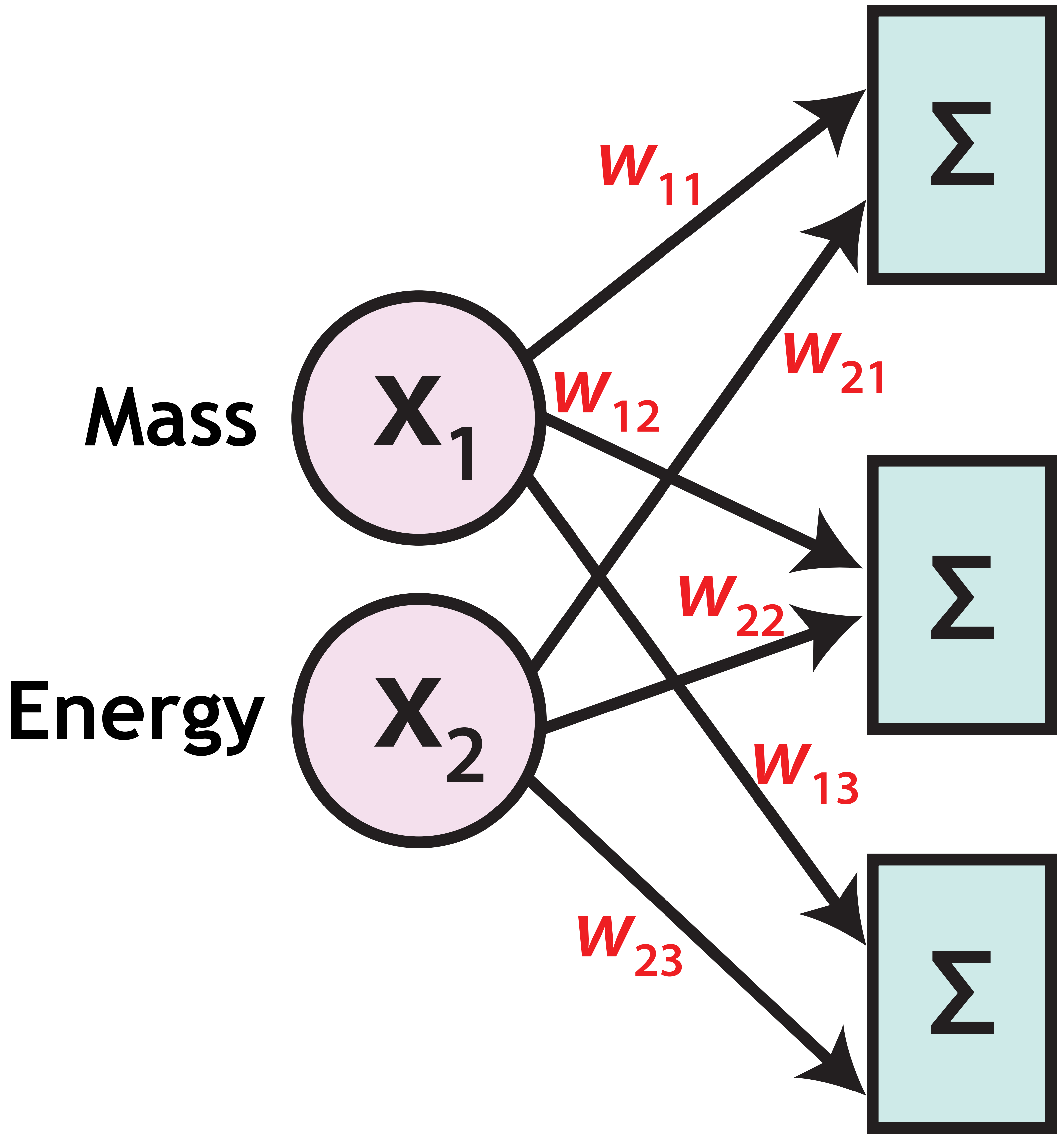
A Brief Interlude on the Hegemony of Homoscedasticity

Astronomers cannot rely on the ML community, we must join it!

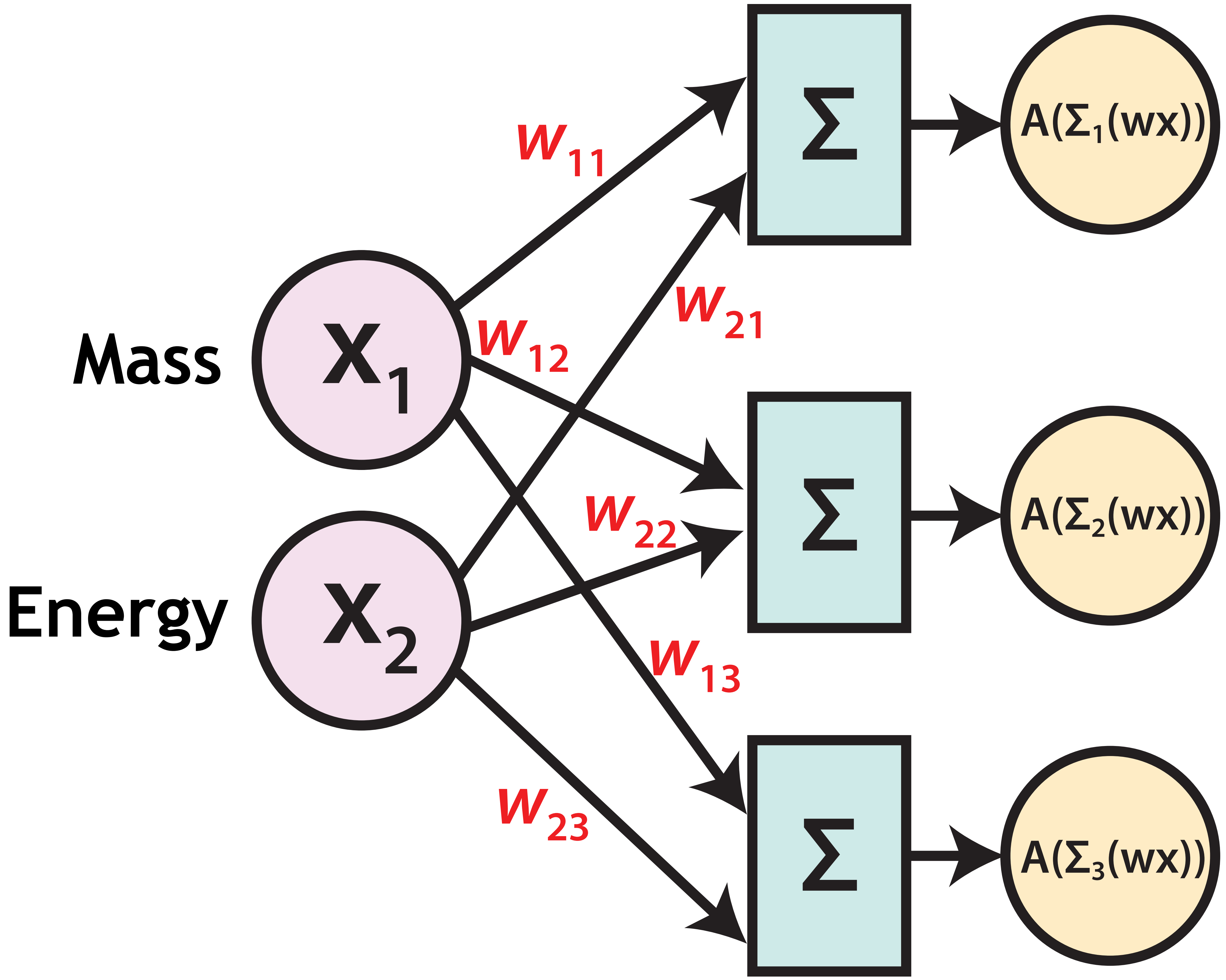
Neural Networks are actually pretty simple*



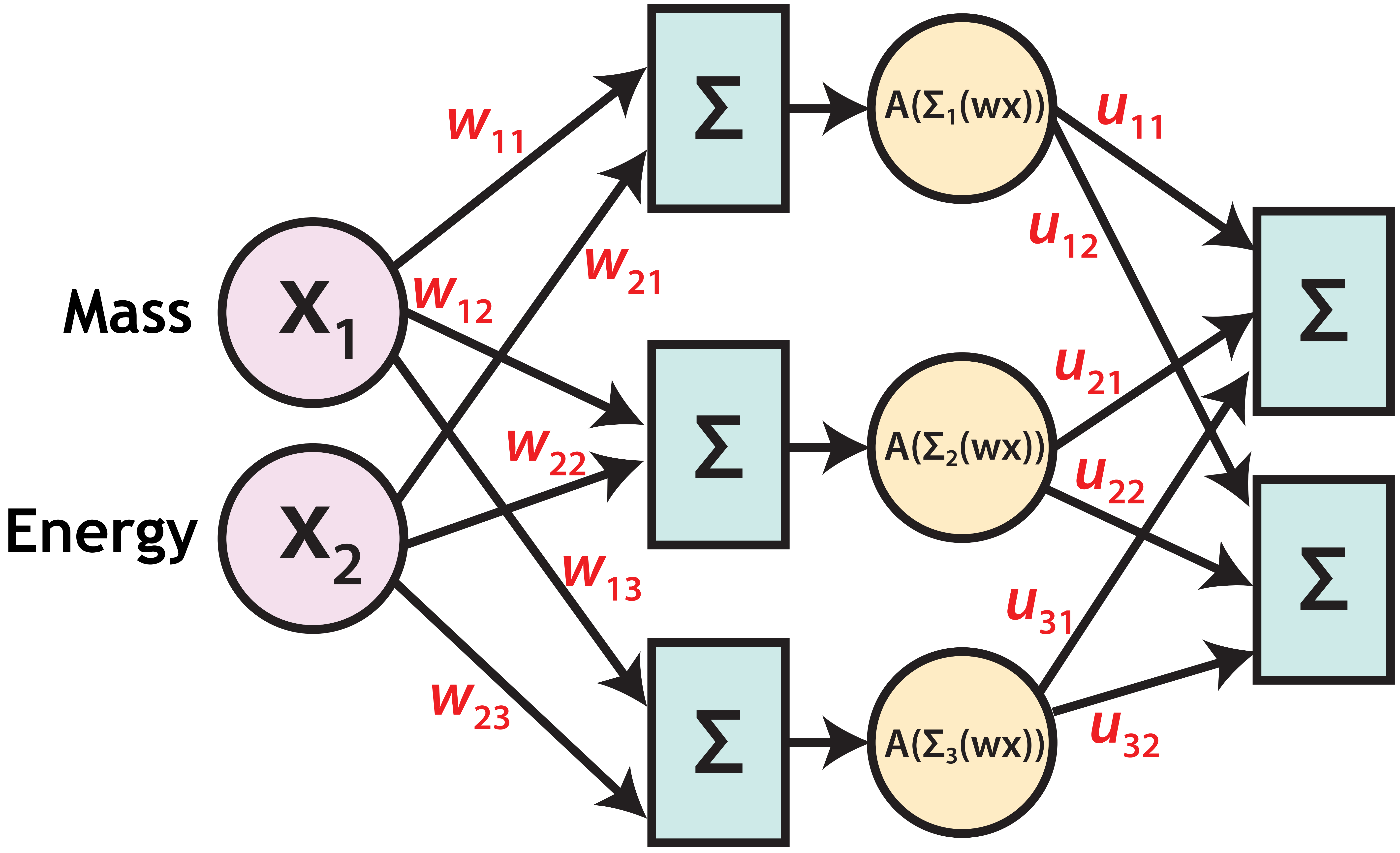
Neural Networks are actually pretty simple



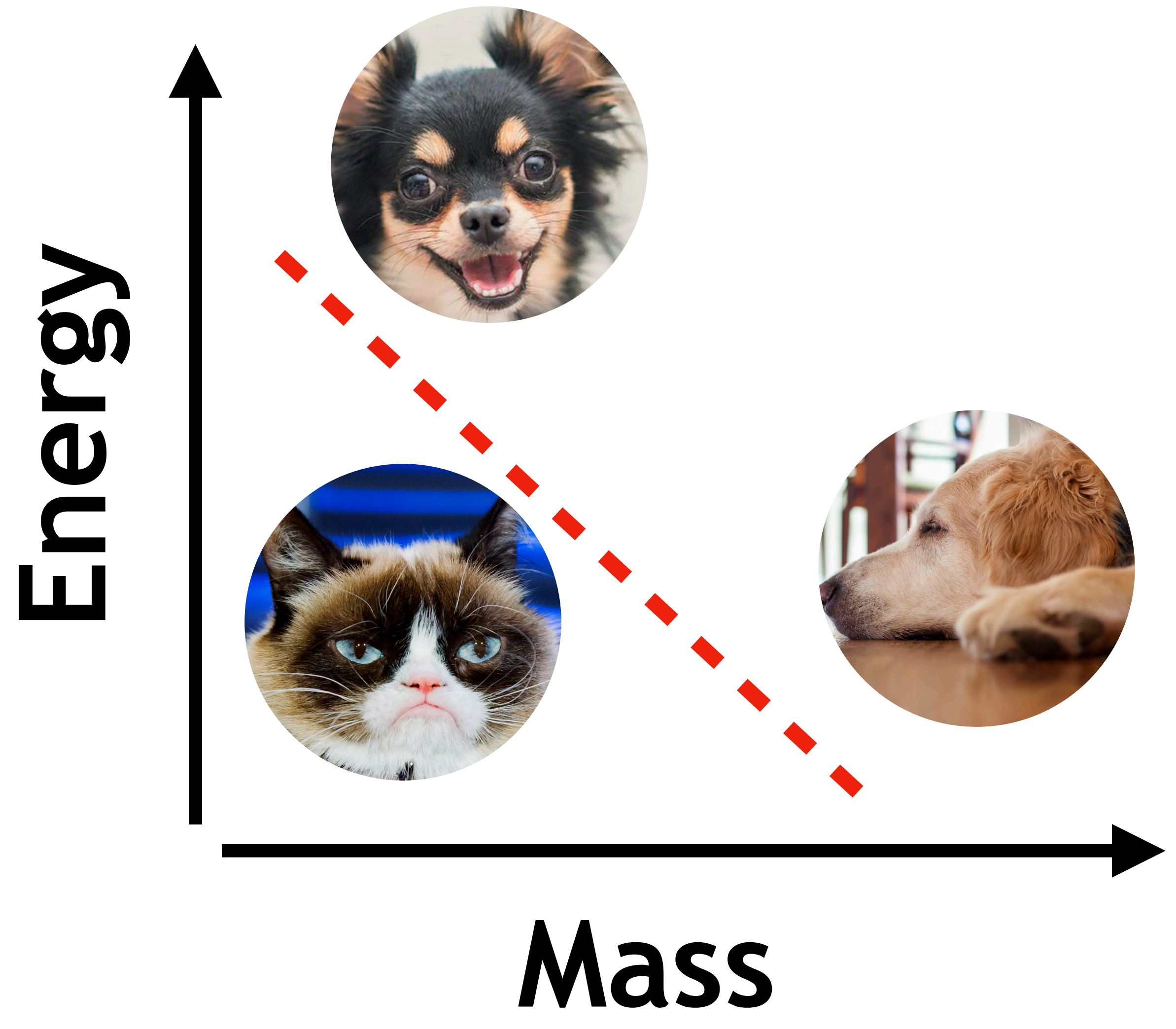
Neural Networks are actually pretty simple

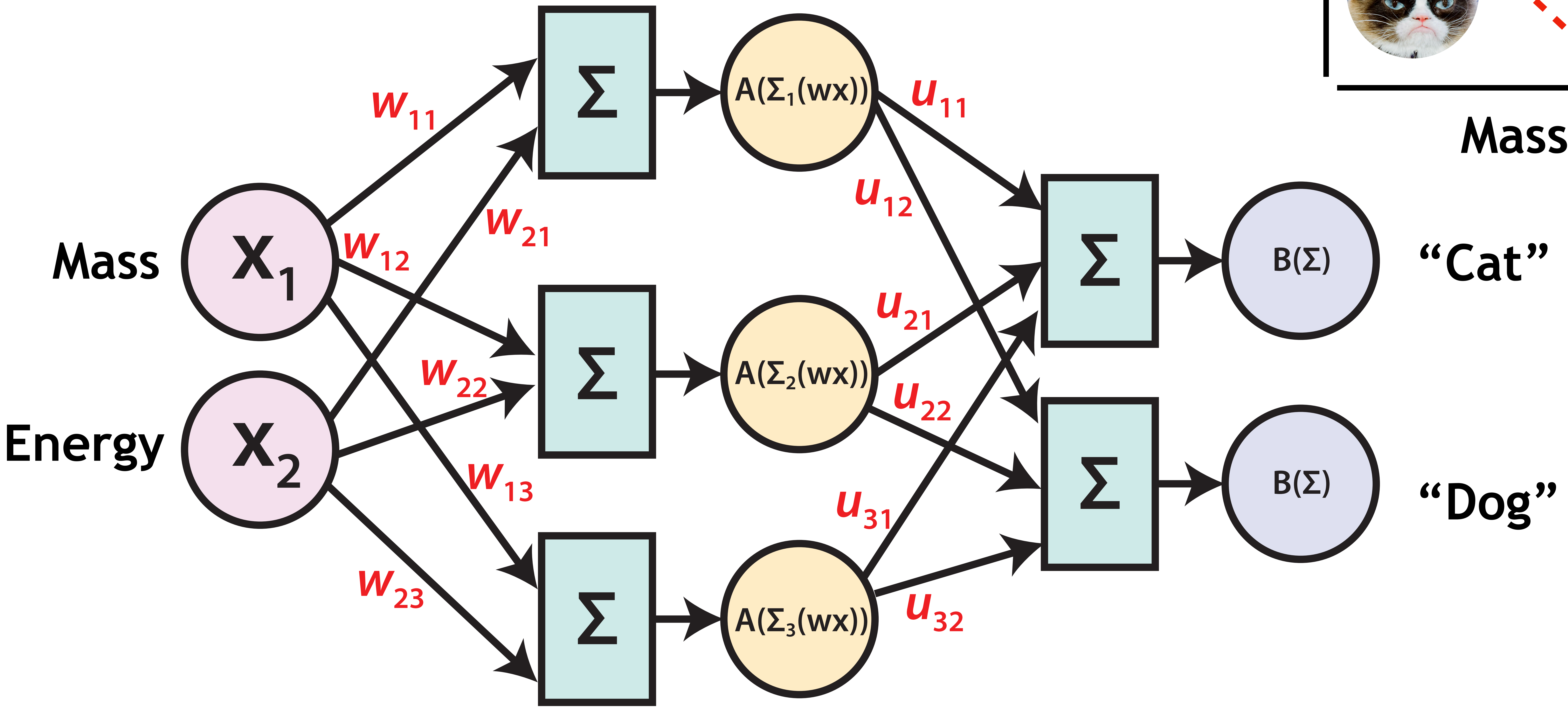


Neural Networks are actually pretty simple

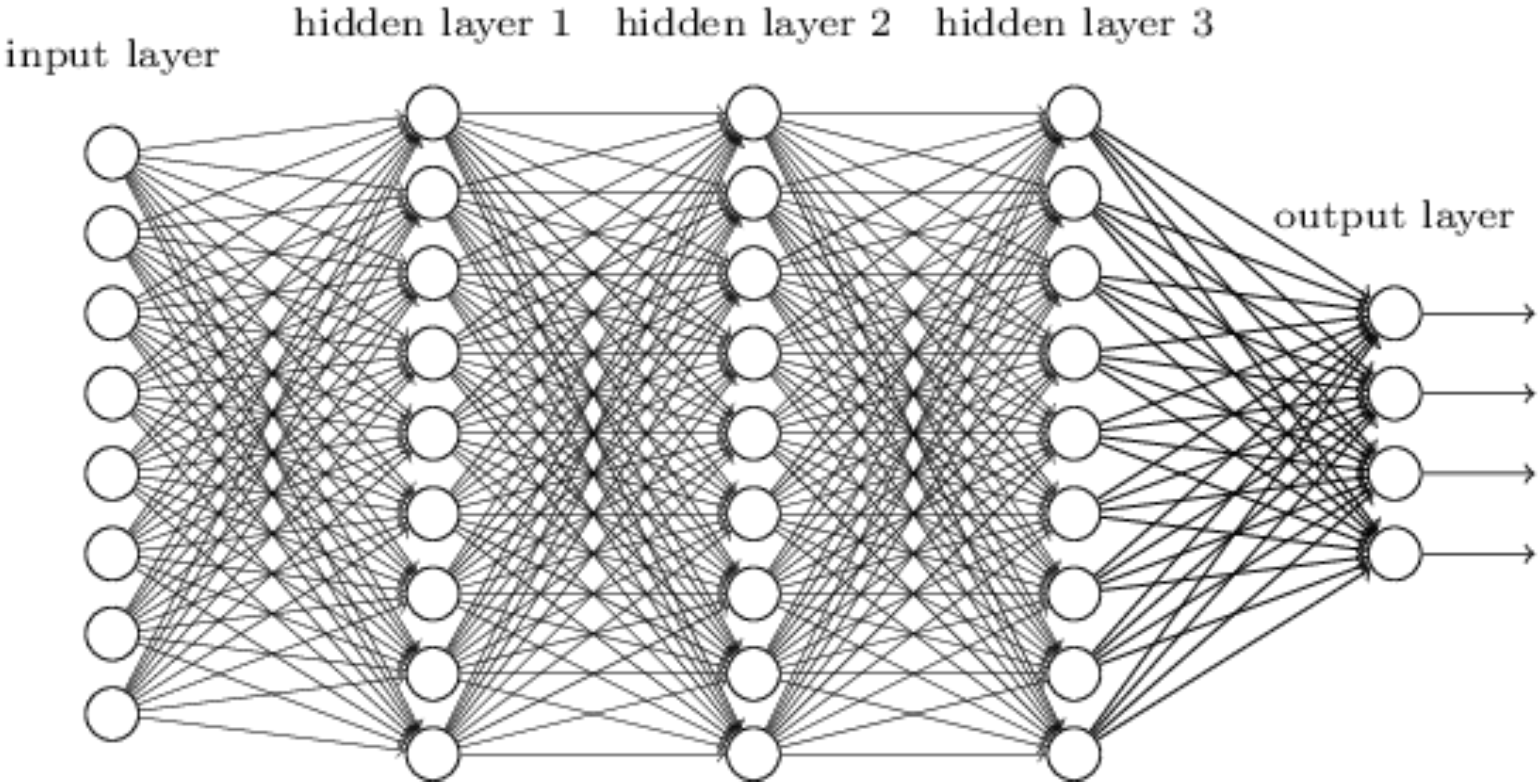


Neural Networks are actually pretty simple



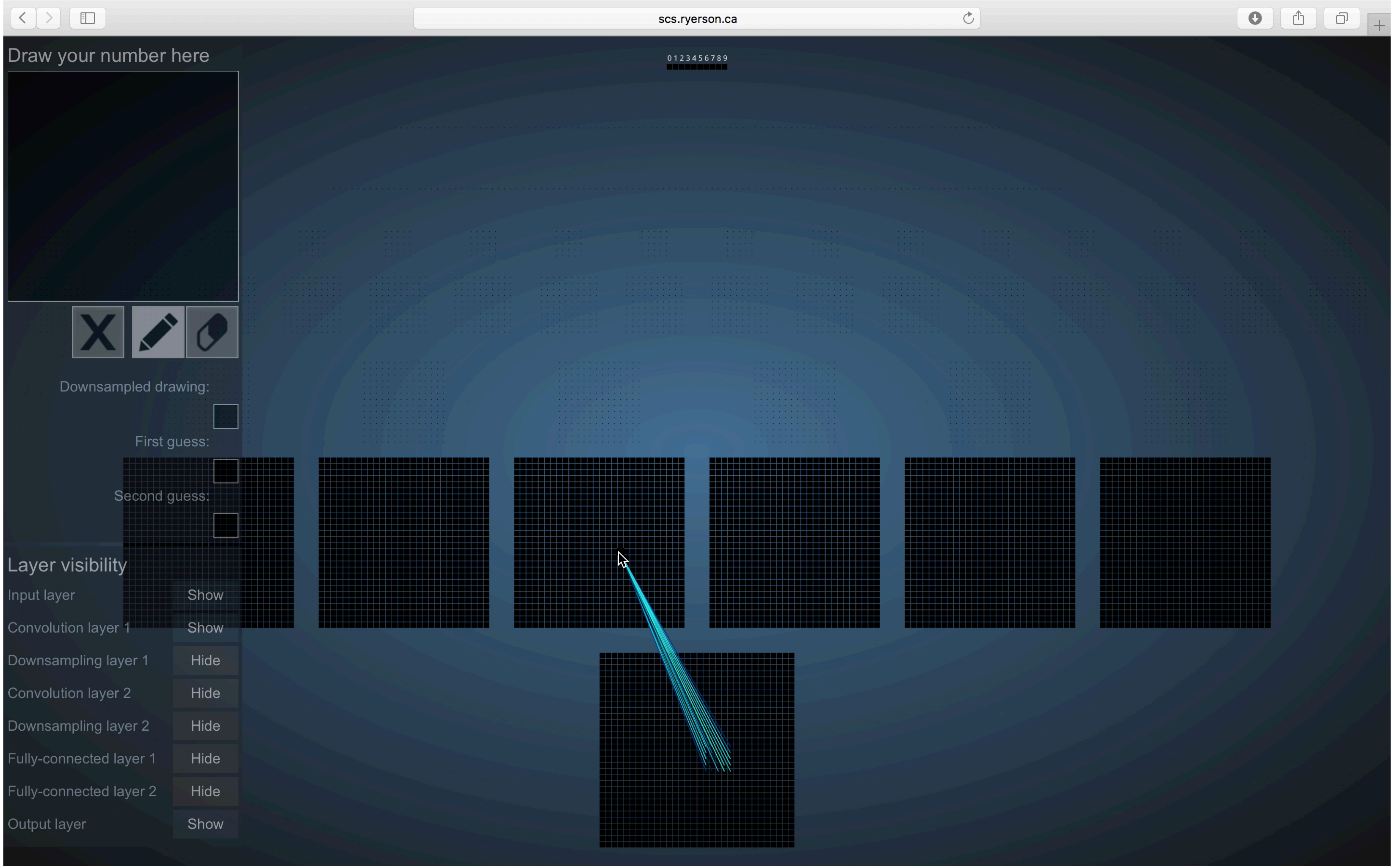


Deep Neural Networks use Many Hidden Layers; hard for many inputs



Weights $\propto$ # neuron^2 x # layers 😱😱😱

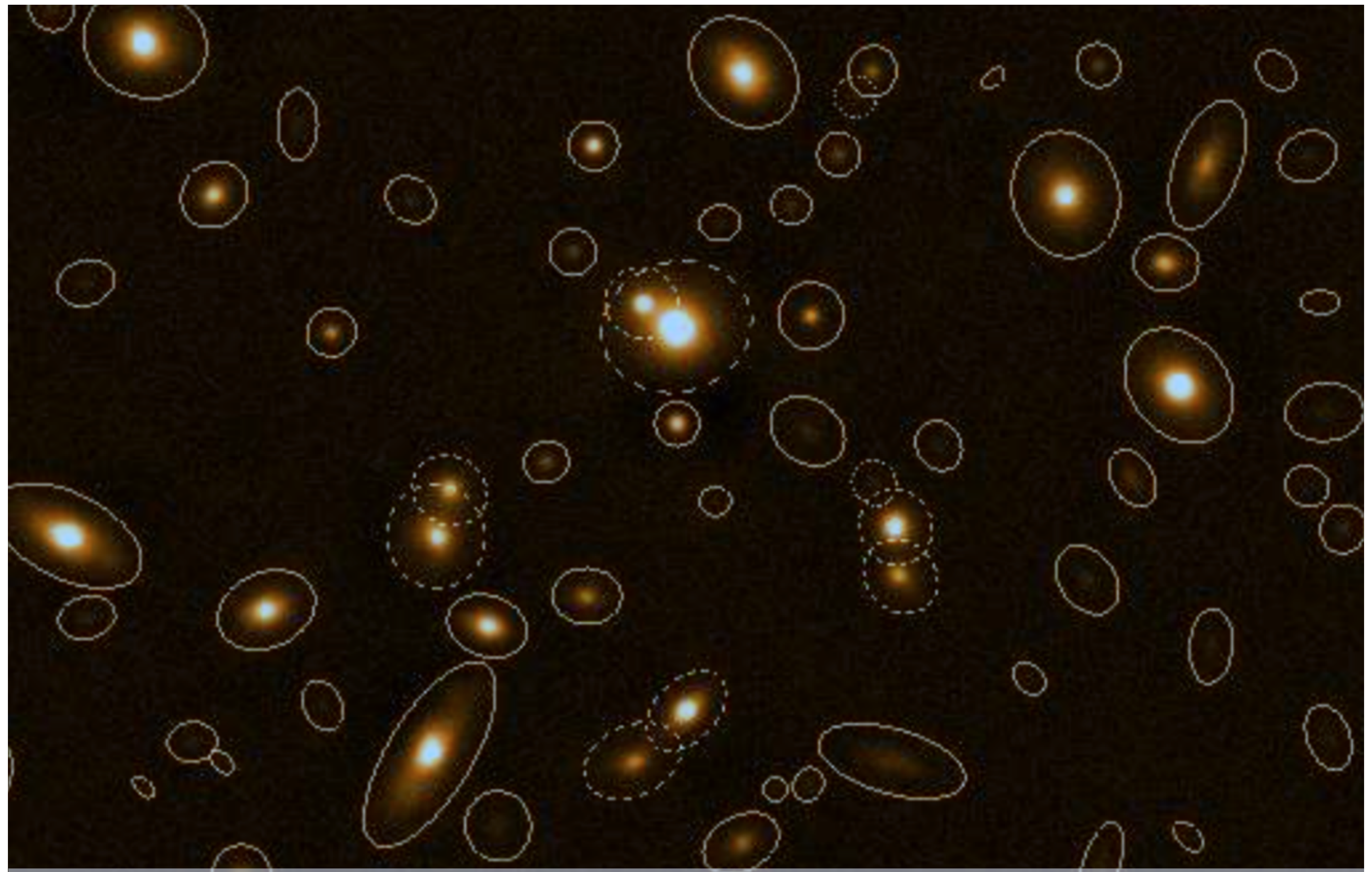
Convolutional Neural Networks (CNNs) are designed for images



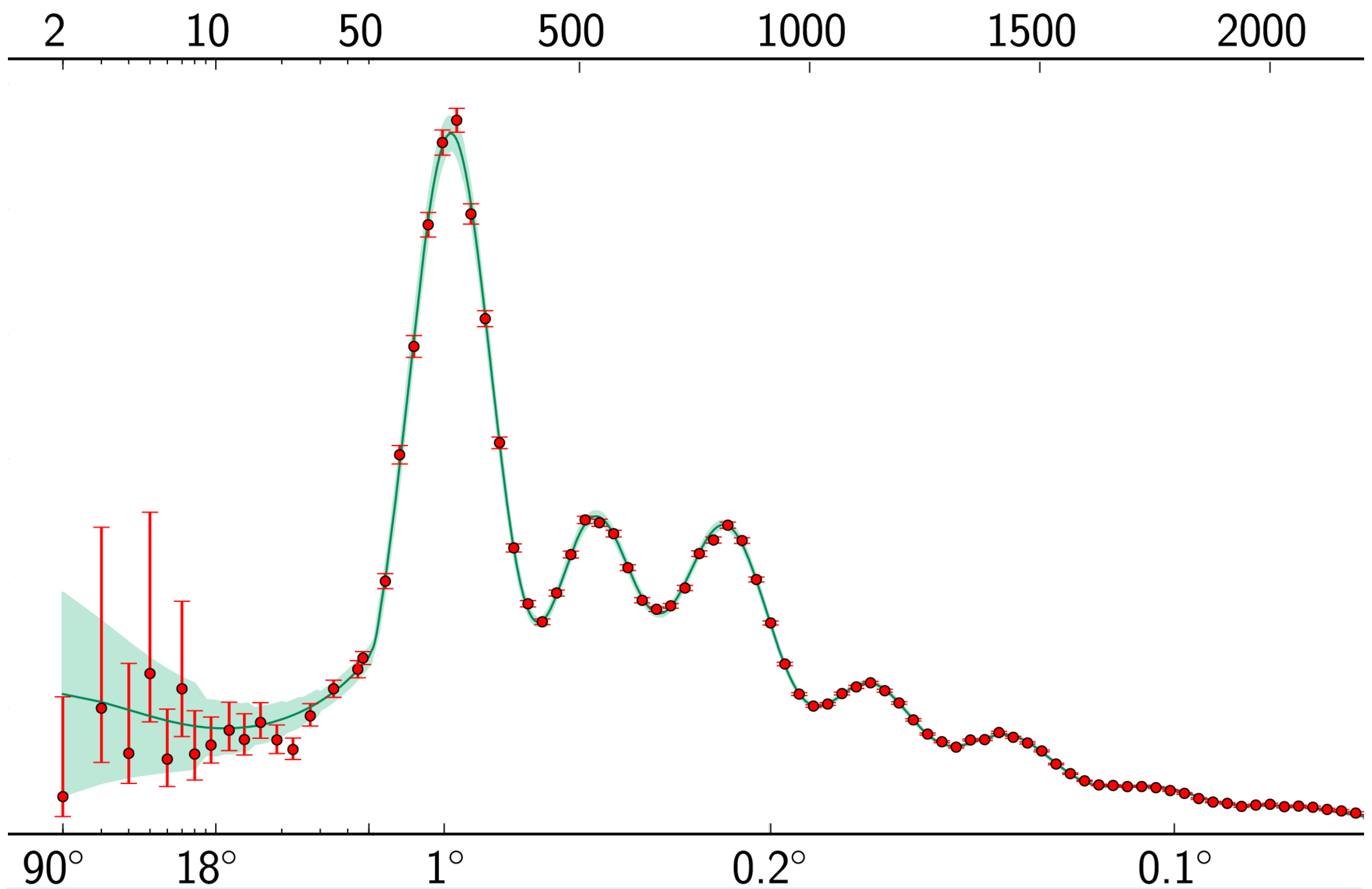
A Brief Interlude on image information:

Pixel-driven ML (e.g. neural nets) impose **many fewer** *assumptions about where the information is*

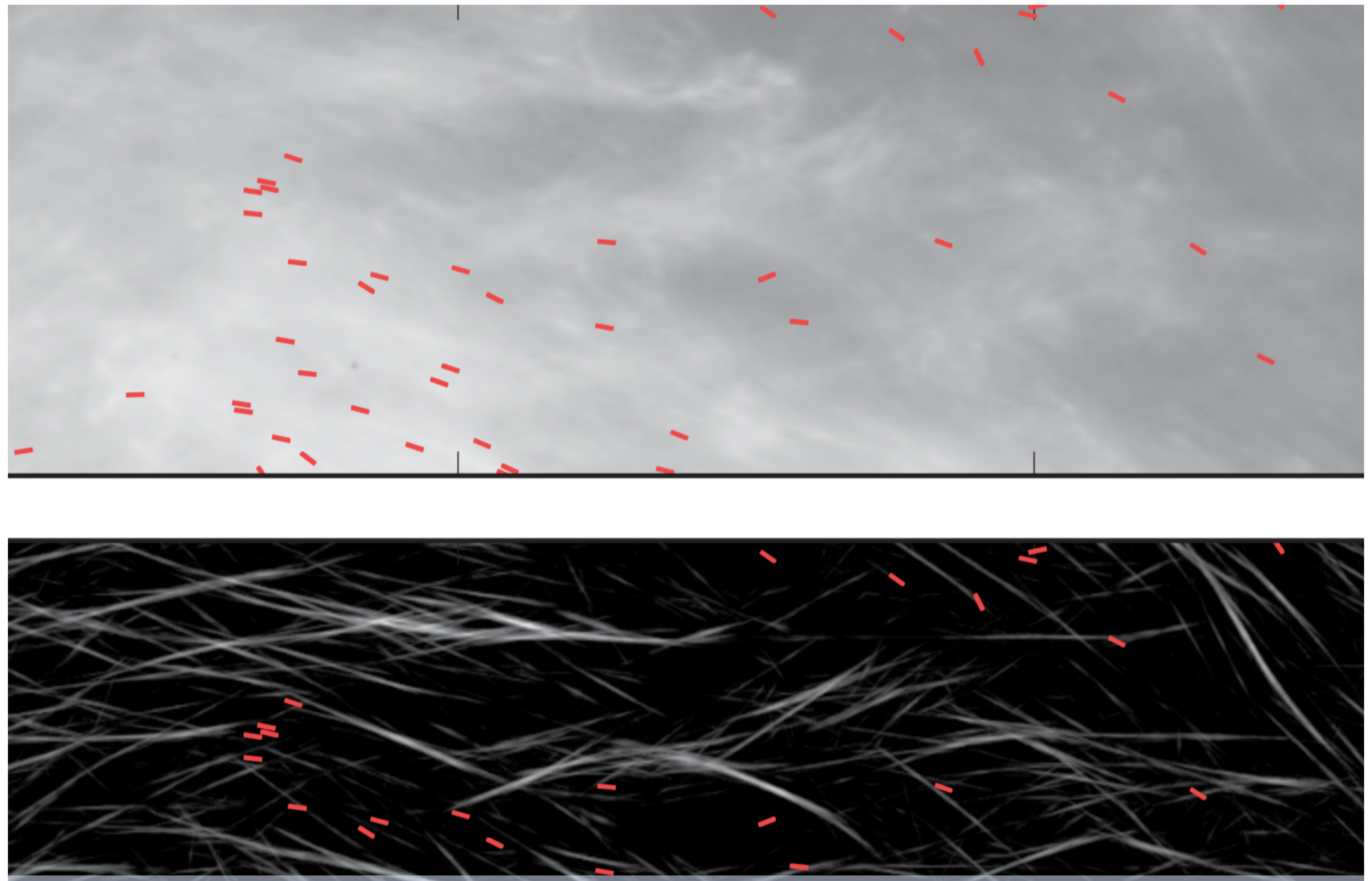
Astronomers know how to look at the sky in 3 ways



Black-Sky Segmentation



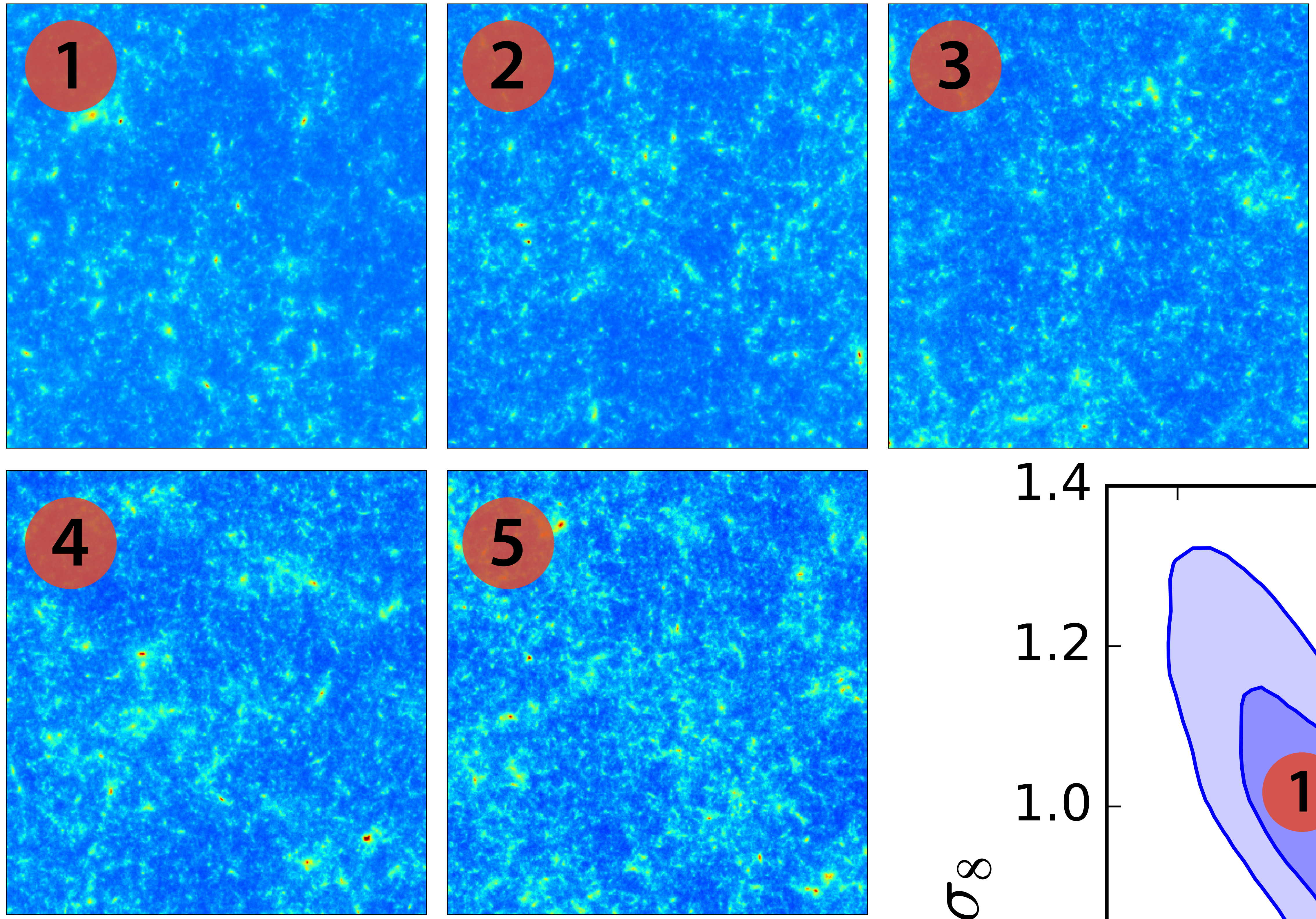
Power Spectra

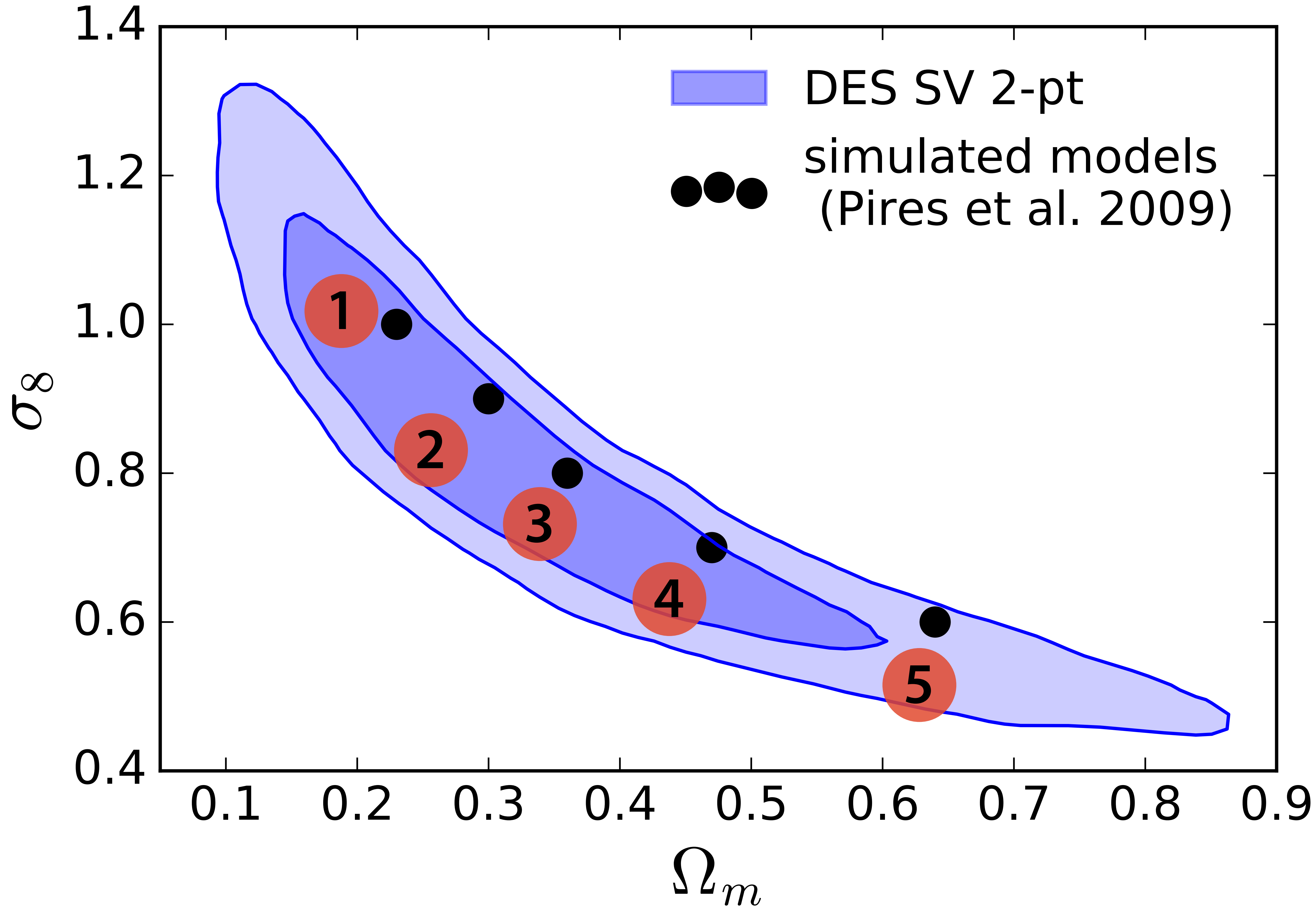


Bespoke Algorithms

Astronomers know how to look at the sky in 3 ways

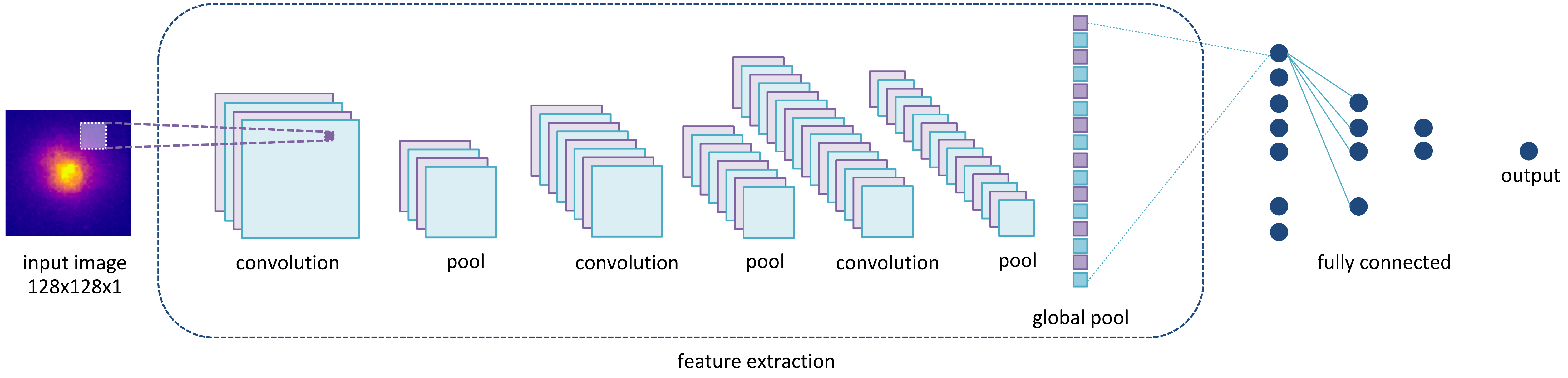
CNNs can extract shape information from Weak Lensing maps





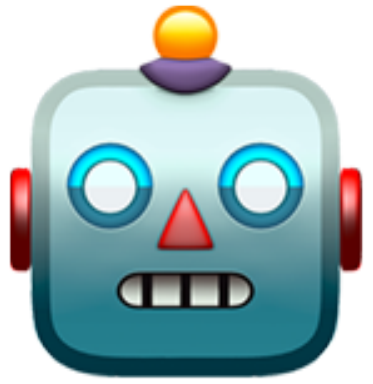
Schmelzle+2017

CNNs can extract masses from galaxy clusters by excising cores



Ntampaka +2019

Networks don’t have to be in the abstract



The Astrophysical Journal Supplement Series, 234:39 (19pp), 2018 February
https://doi.org/10.3847/1538-4365/aaa3e2

CrossMark

The DECam Plane Survey: Optical Photometry of Two Billion Objects in the Southern Galactic Plane

E. F. Schlafly[1,8], G. M. Green[2], D. Lang[3,4], T. Daylan[5], D. P. Finkbeiner[5,6], A. Lee[5], A. M. Meisner[1], D. Schlegel[1], and F. Valdes[7]
[1] Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, USA
[2] Kavli Institute for Particle Astrophysics and Cosmology, Physics and Astrophysics Building, 452 Lomita Mall, Stanford, CA 94305, USA
[3] Department of Astronomy & Astrophysics and Dunlap Institute, University of Toronto, 50 Saint George Street, Toronto, ON, M5S 3H4, Canada
[4] Department of Physics & Astronomy, University of Waterloo, 200 University Avenue West, Waterloo, ON, N2L 3G1, Canada
[5] Department of Physics, Harvard University, 17 Oxford Street, Cambridge, MA 02138, USA
[6] Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA
[7] National Optical Astronomy Observatories, P.O. Box 26732, Tucson, AZ 85719, USA
Received 2017 October 3; revised 2017 December 12; accepted 2017 December 18; published 2018 February 9

Abstract

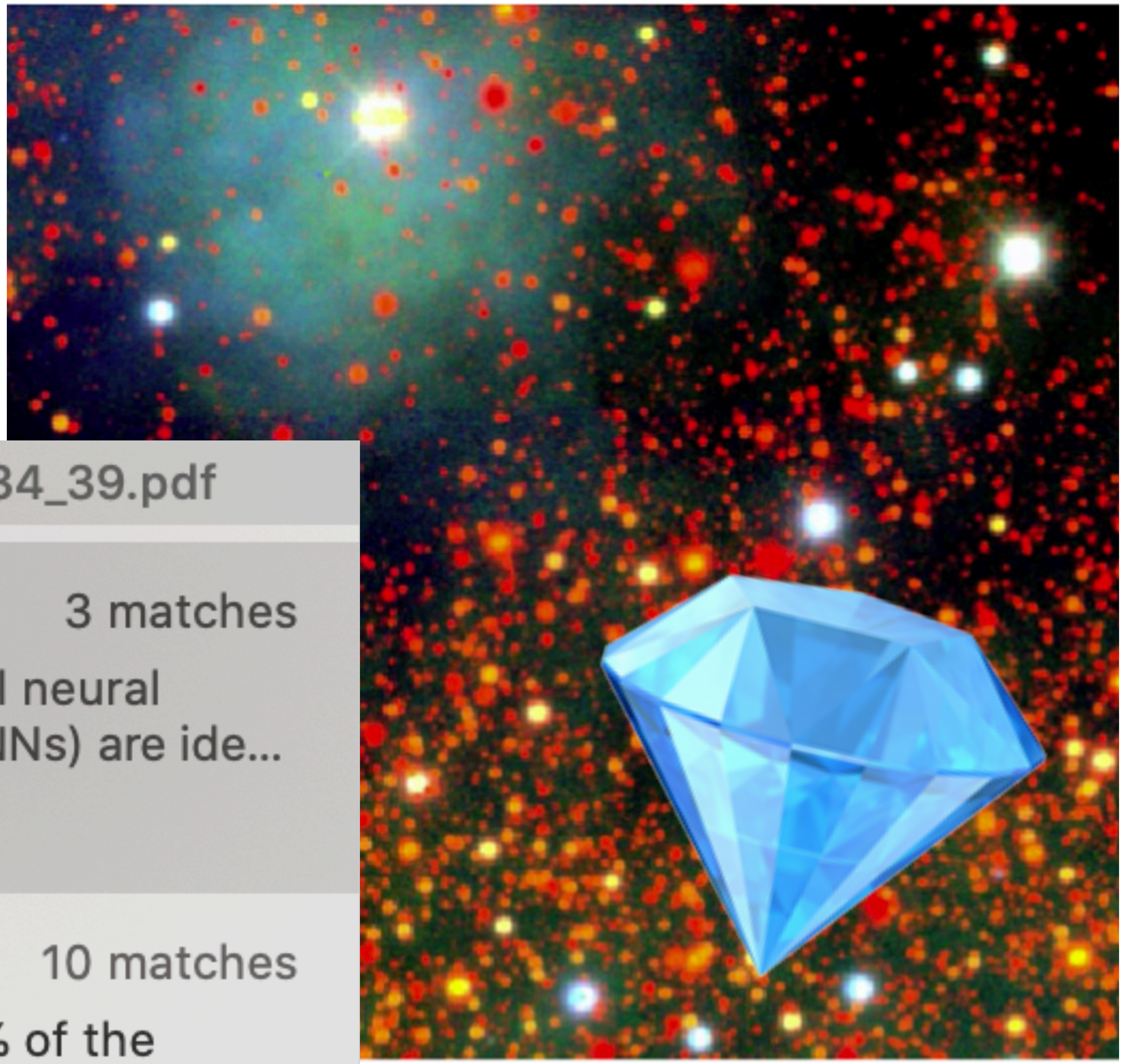
The DECam Plane Survey is a five-band optical and near-infrared survey of the southern
Dark Energy Camera at Cerro Tololo. The survey is designed to reach past the main-
populations at the distance of the Galactic center through a reddening $E(B-V)$ of
exposure depths are 23.7, 22.8, 22.3, 21.9, and 21.0 mag (AB) in the *grizY* bands, wit
footprint covers the Galactic plane with $|b| \lesssim 4°$, $5° > l > -120°$. The survey pipeline s
the positions and fluxes of tens of thousands of sources in each image, delivering positio
two billion stars with better than 10 mmag precision. Most of these objects are highly r
Galactic disk, probing the structure and properties of the Milky Way and its interste
processed images and derived catalogs are publicly available.

Key words: catalogs – surveys – techniques: photometric

1. Introduction

Many of the Milky Way stars and much of its gas and dust reside in a disk. Accordingly, observations of the Milky Way disk are critical to understanding the Milky Way—particularly observations toward the inner Galaxy, where most of the mass lies. At optical wavelengths, however, the interpretation of observations of the Milky Way disk can be challenging due to the tremendous number of stars and due to extinction by dust, motivating surveys of the disk at infrared wavelengths where

photometric measurements o
of data for understanding the
DECaPS occupies a specia
targeting the Milky Way. Th
PS1 survey (Chambers et al.
very similar set of filters and
is roughly 1 mag deeper in i
covers the entire sky abov
epochs than DECaPS does
valuable survey for understa



Schlafly_2018_ApJS_234_39.pdf

Page 4 3 matches
Convolutional neural networks (CNNs) are ide...

Page 5 10 matches
We used 80% of the images to train the netw...

Schlafly + 2018

What is ML in Astronomy?

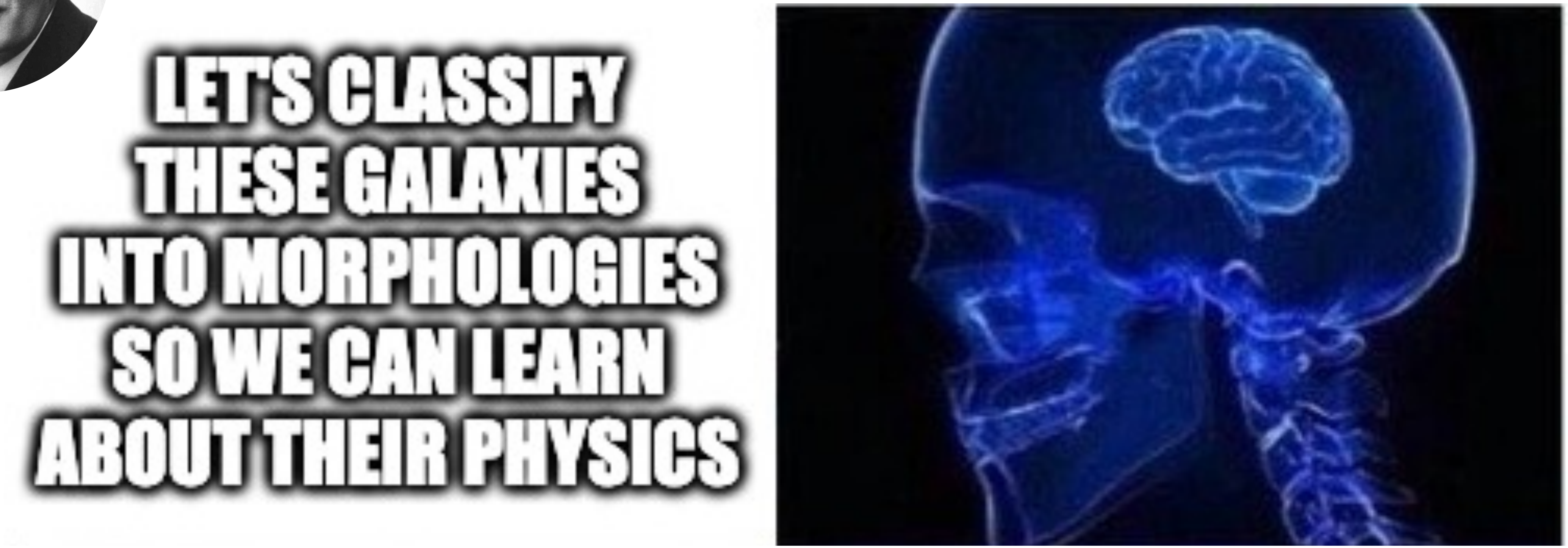
Supervised vs. Unsupervised

Supervised Learning in Astronomy

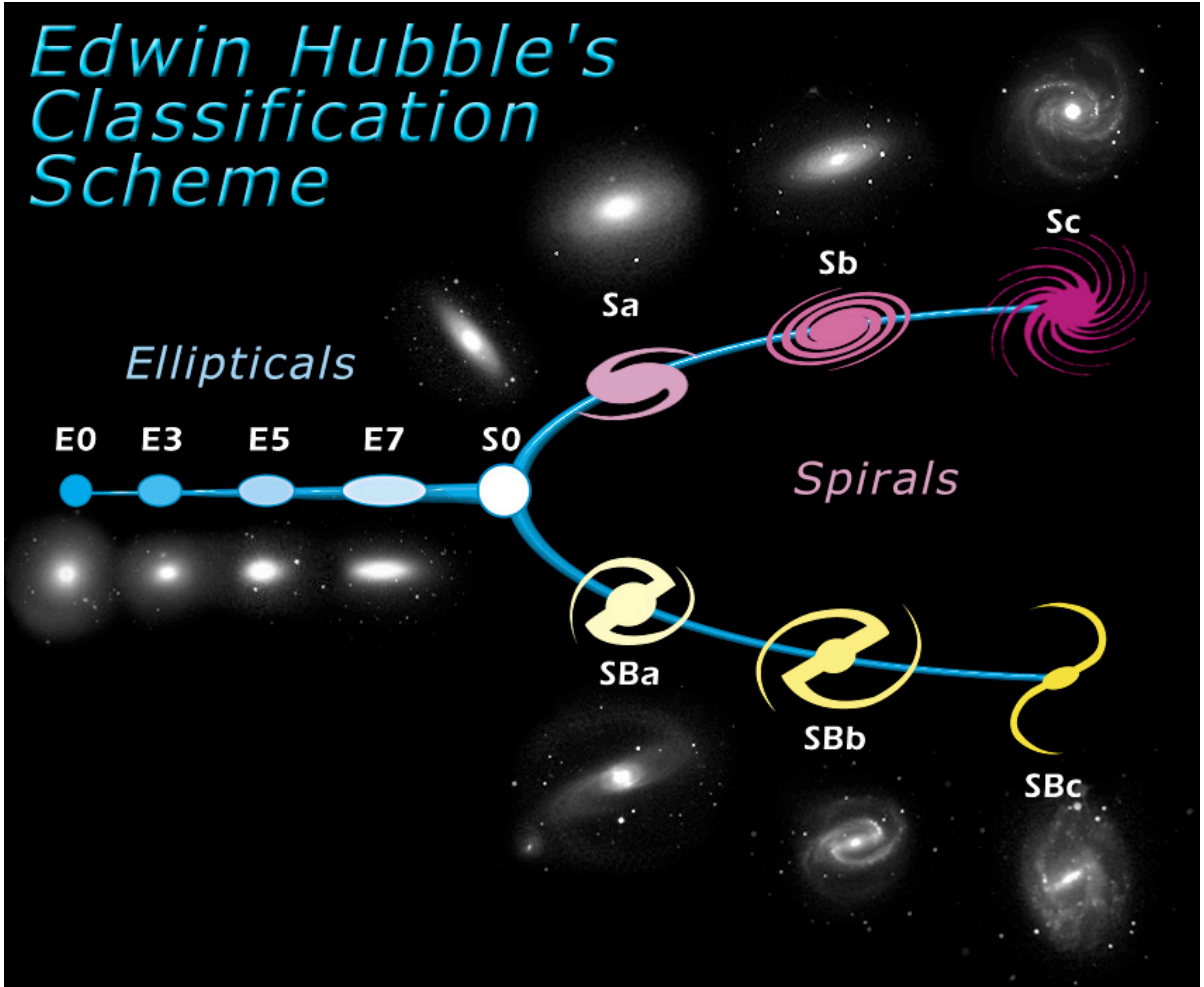
Supervised Learning: *Galaxy Images*

Unsupervised Learning in Astronomy

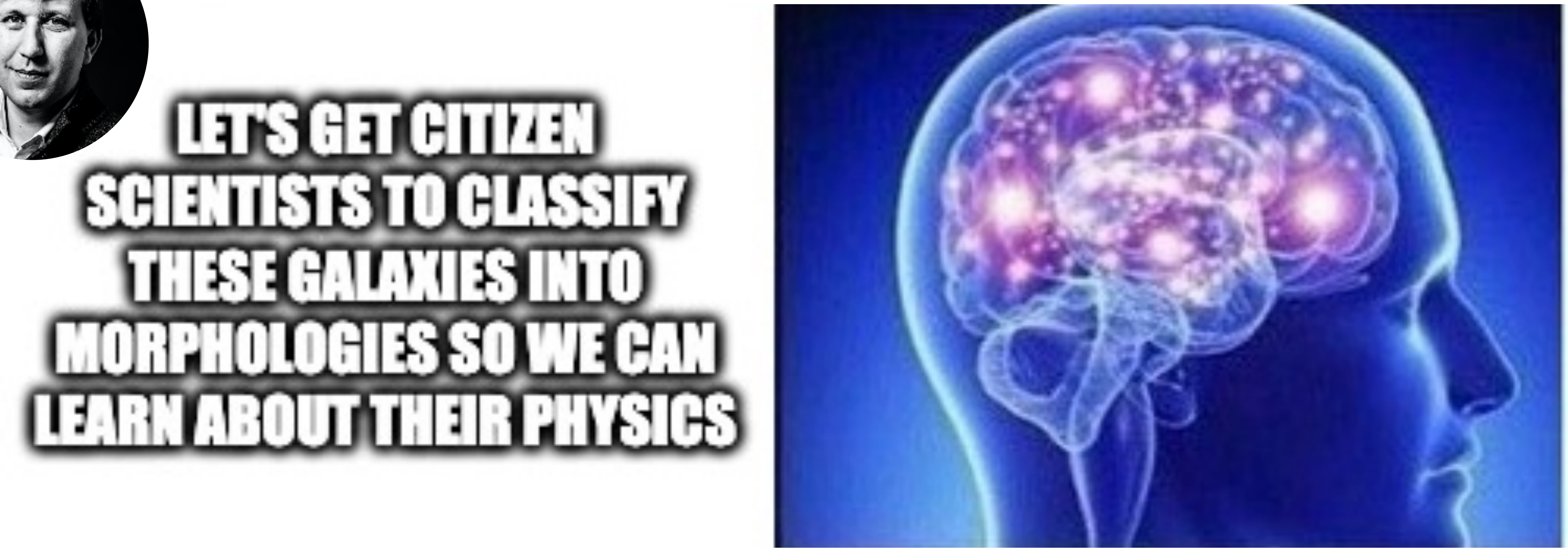
Unsupervised Learning: *Search By Image*


LET'S CLASSIFY
THESE GALAXIES
INTO MORPHOLOGIES
SO WE CAN LEARN
ABOUT THEIR PHYSICS

Morphology is a key weapon in the physicist's arsenal

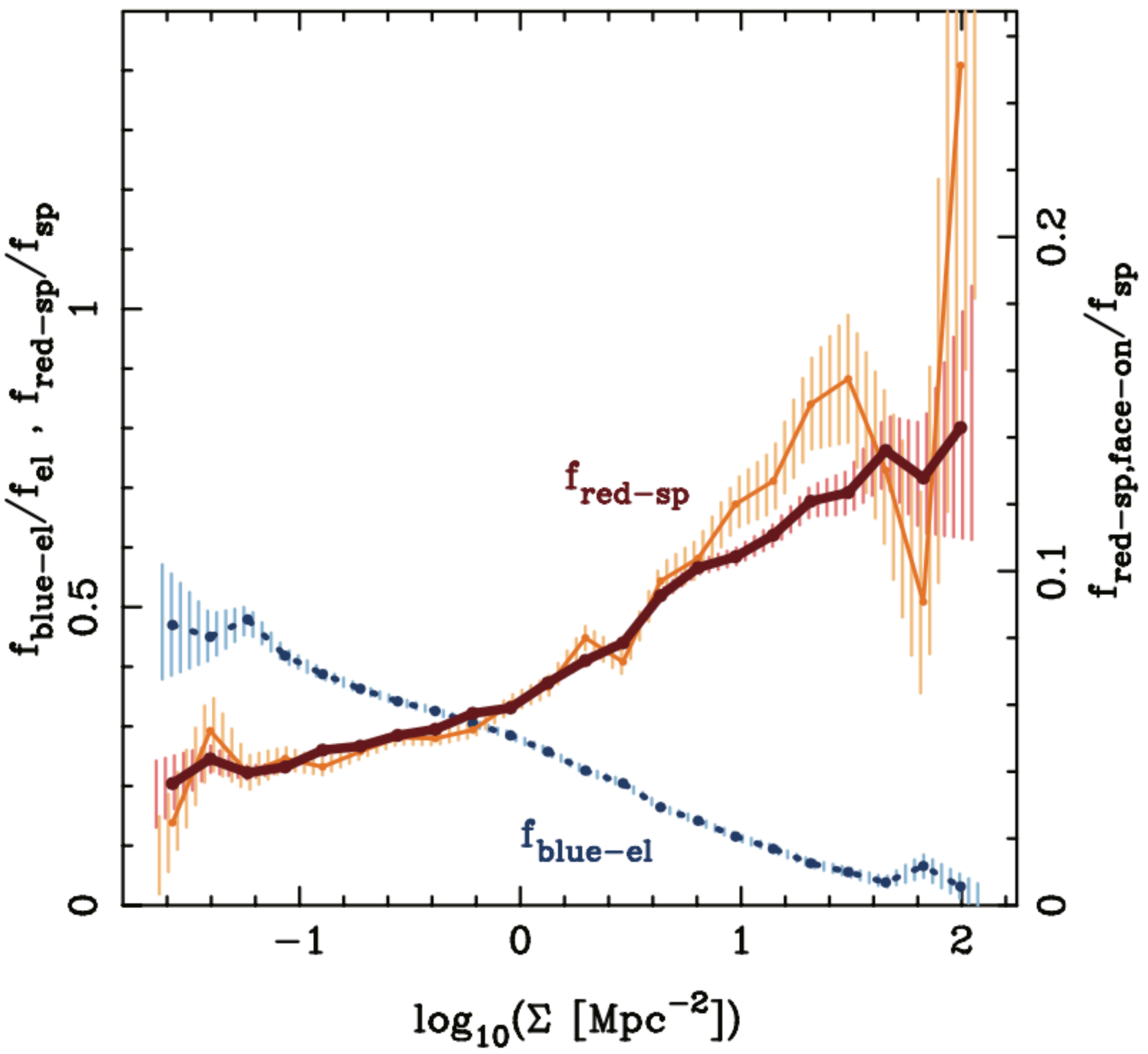


Galaxy Zoo took morphology to the big data era



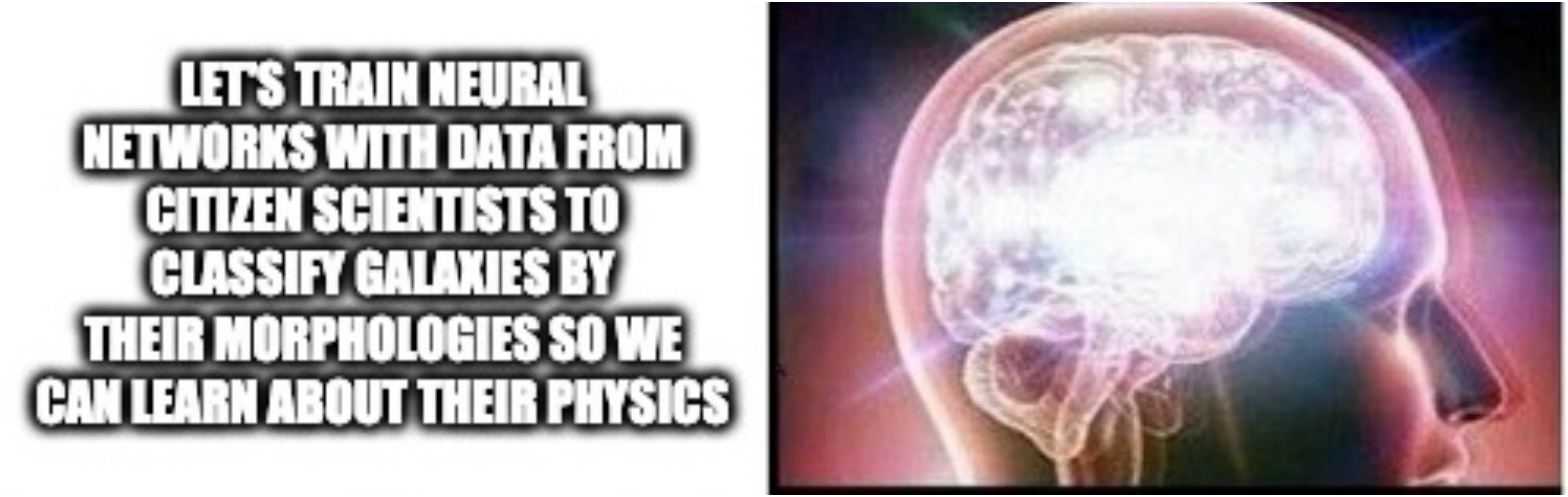

Lintott+ 2008

Now we can study morphology like color

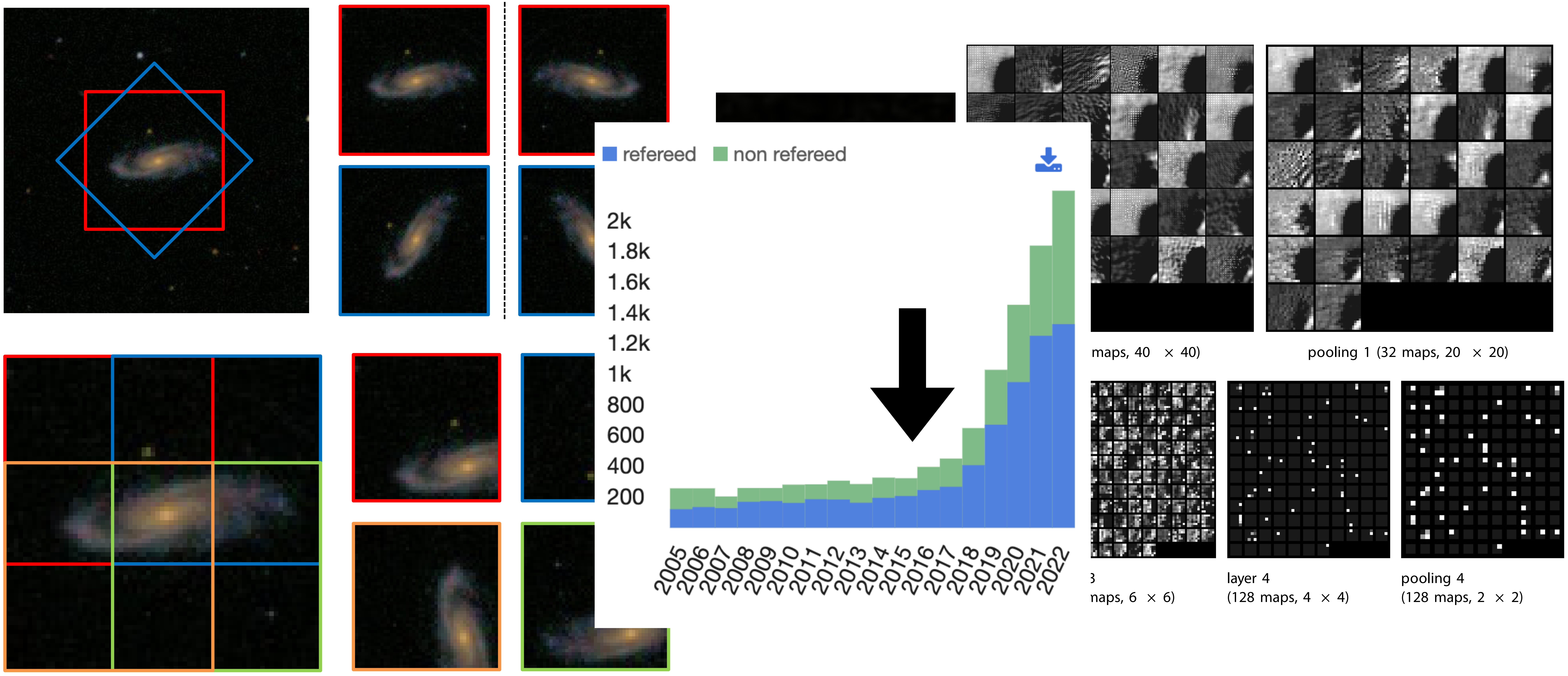


Bamford + 2009

Machine Learning let us take the next jump



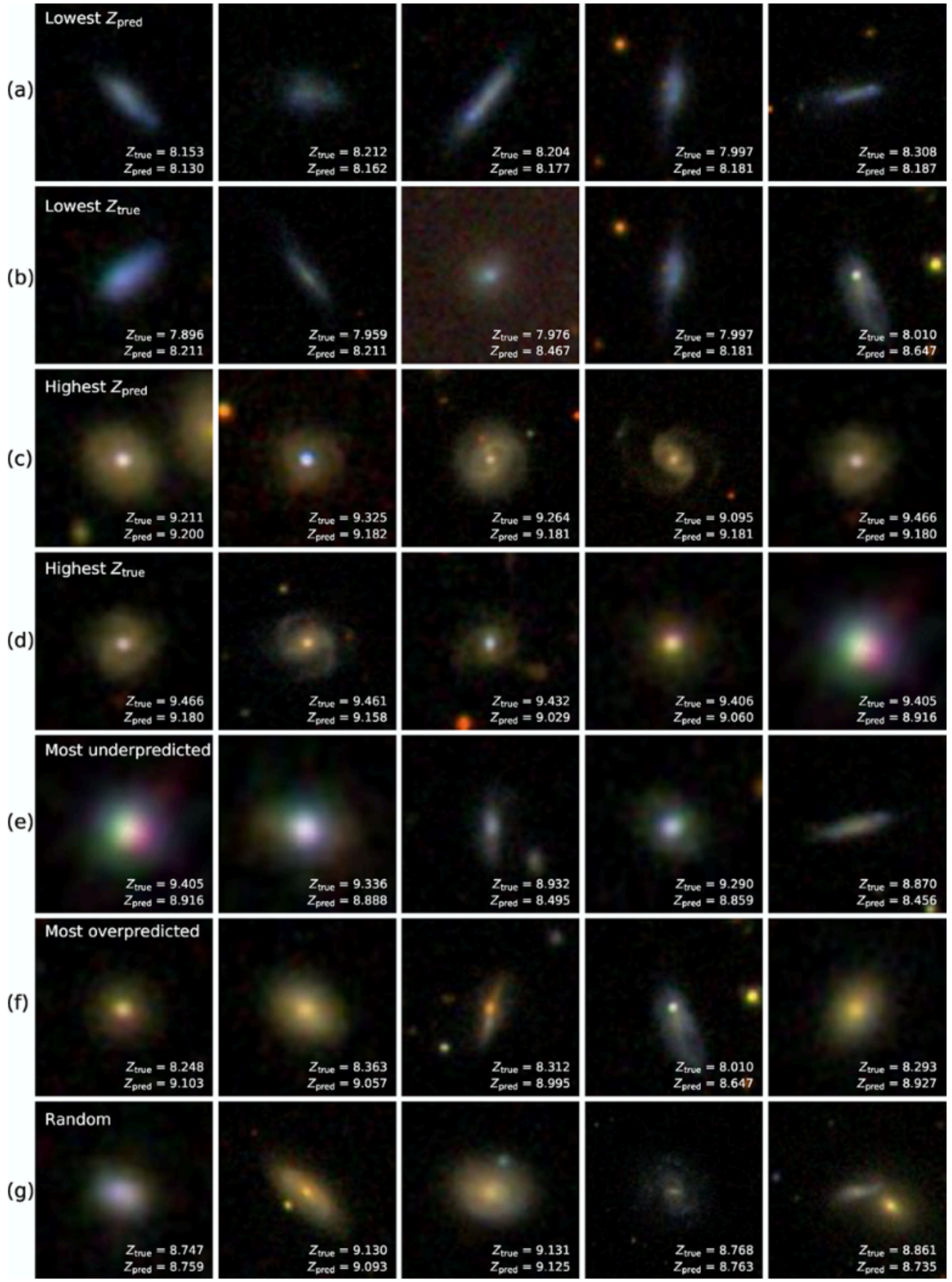
This is the starting gun of deep learning in astro

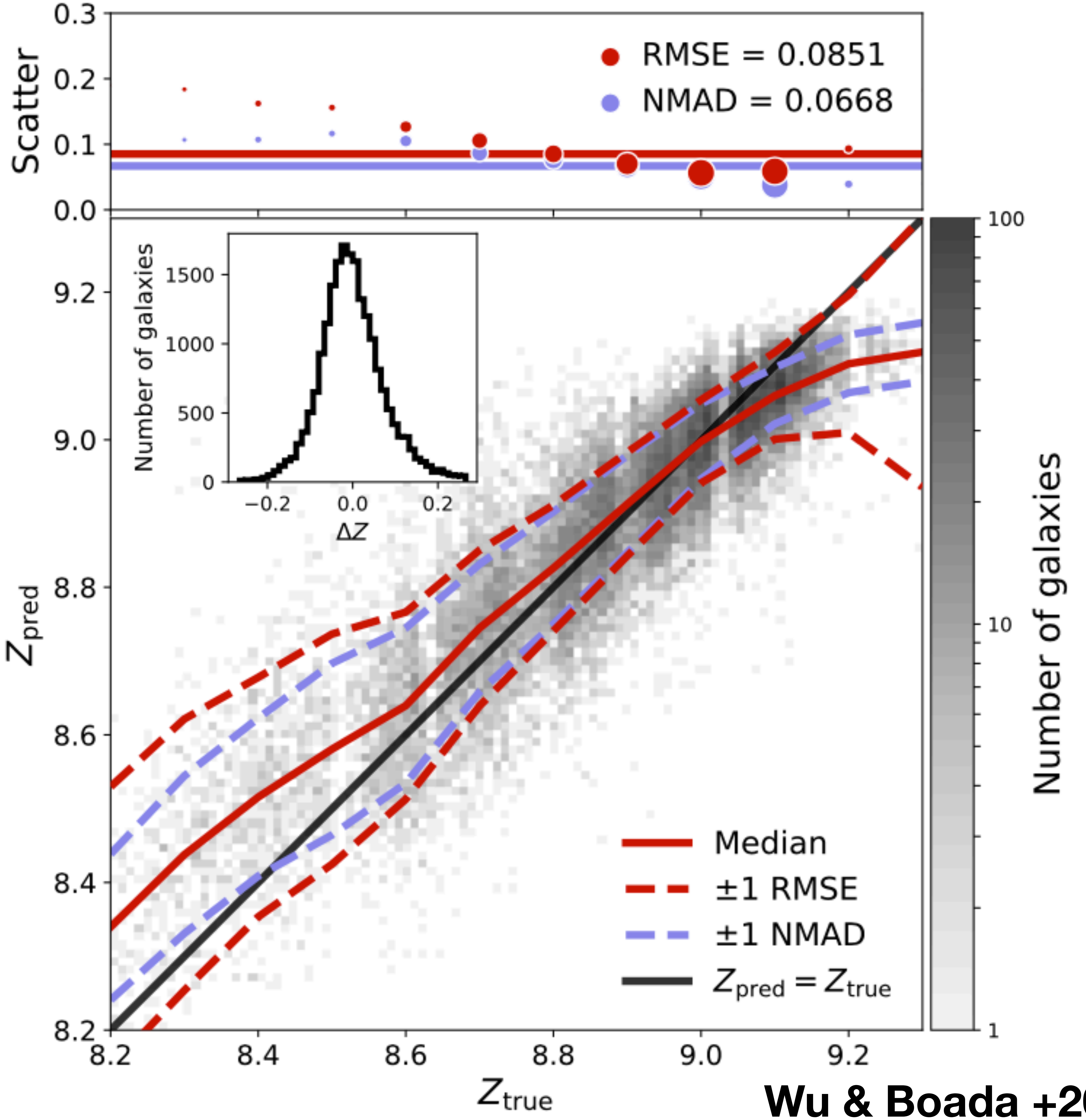


Dieleman+2015

LET'S TRAIN
NEURAL NETS TO LEARN
ABOUT PHYSICS DIRECTLY
FROM GALAXY IMAGES

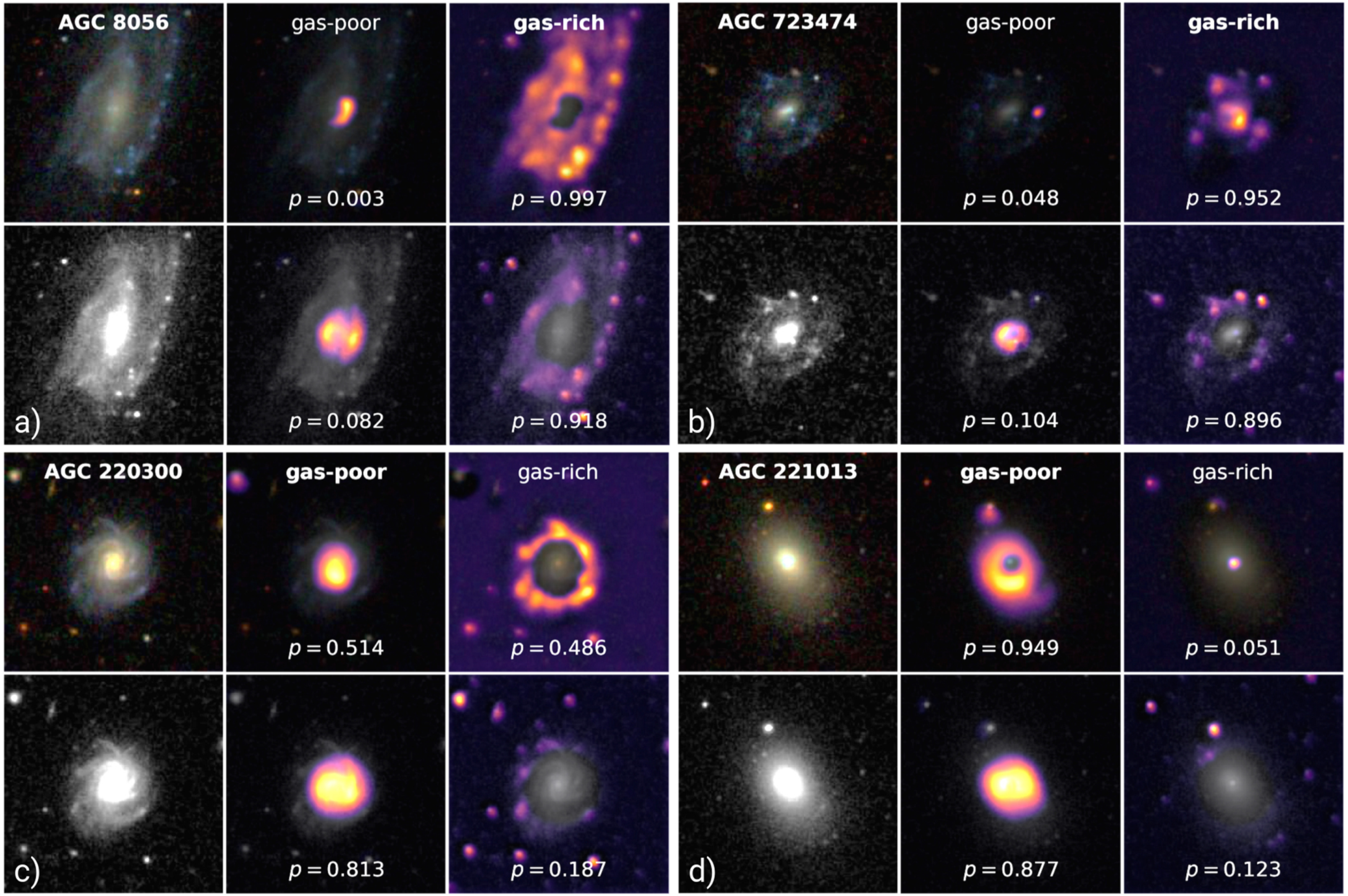
Images contain information about metallicity





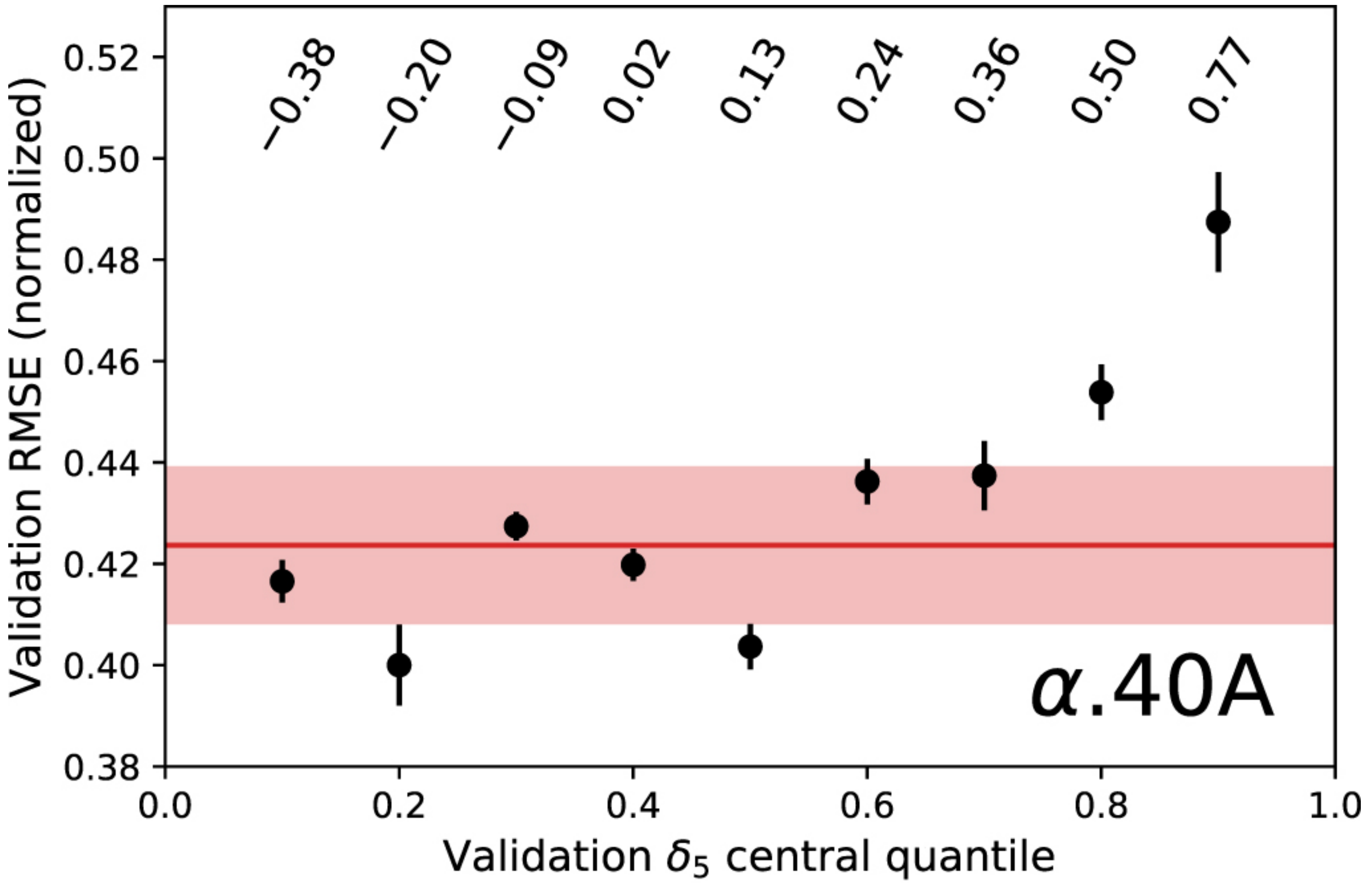
Wu & Boada +2019

Networks can tell us *how* they know something physical



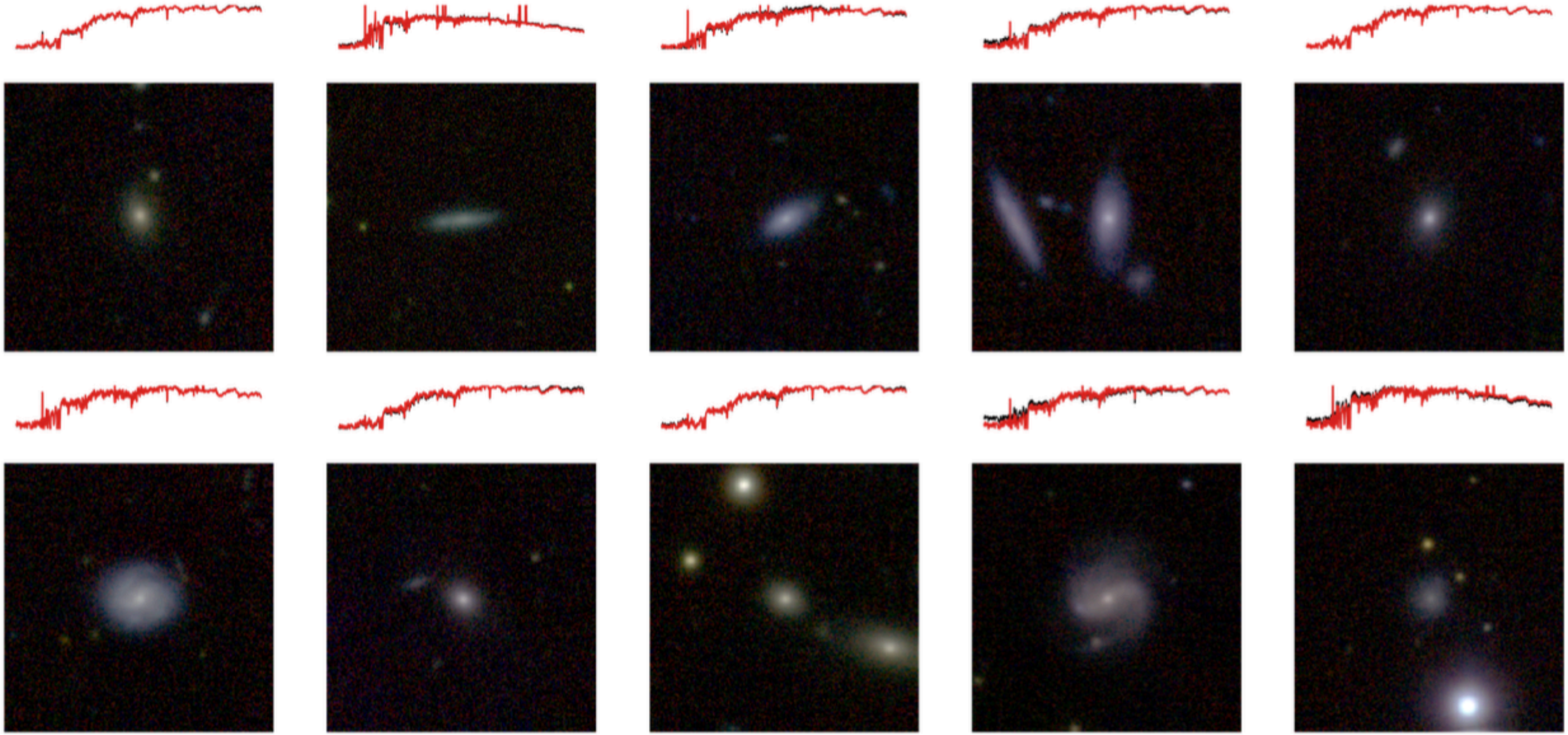
Wu 2020

Networks can tell us *how* they know something physical



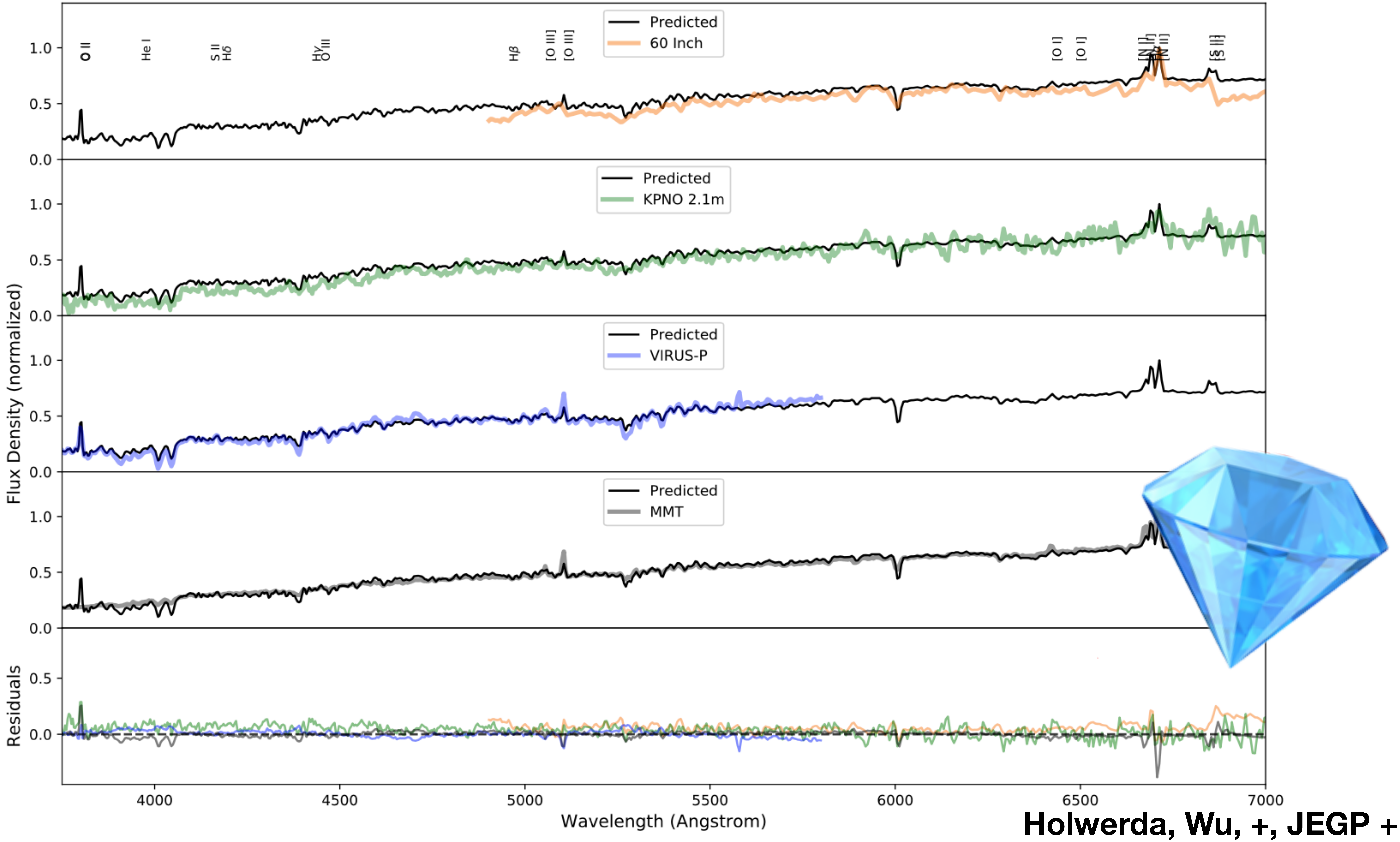
Wu 2020

Cut to the chase: the entire spectrum from the image



Wu & JEGP 2020

Cut to the chase: the entire spectrum from the image



Holwerda, Wu, +, JEGP + 2021

A practical example: finding low-z galaxies

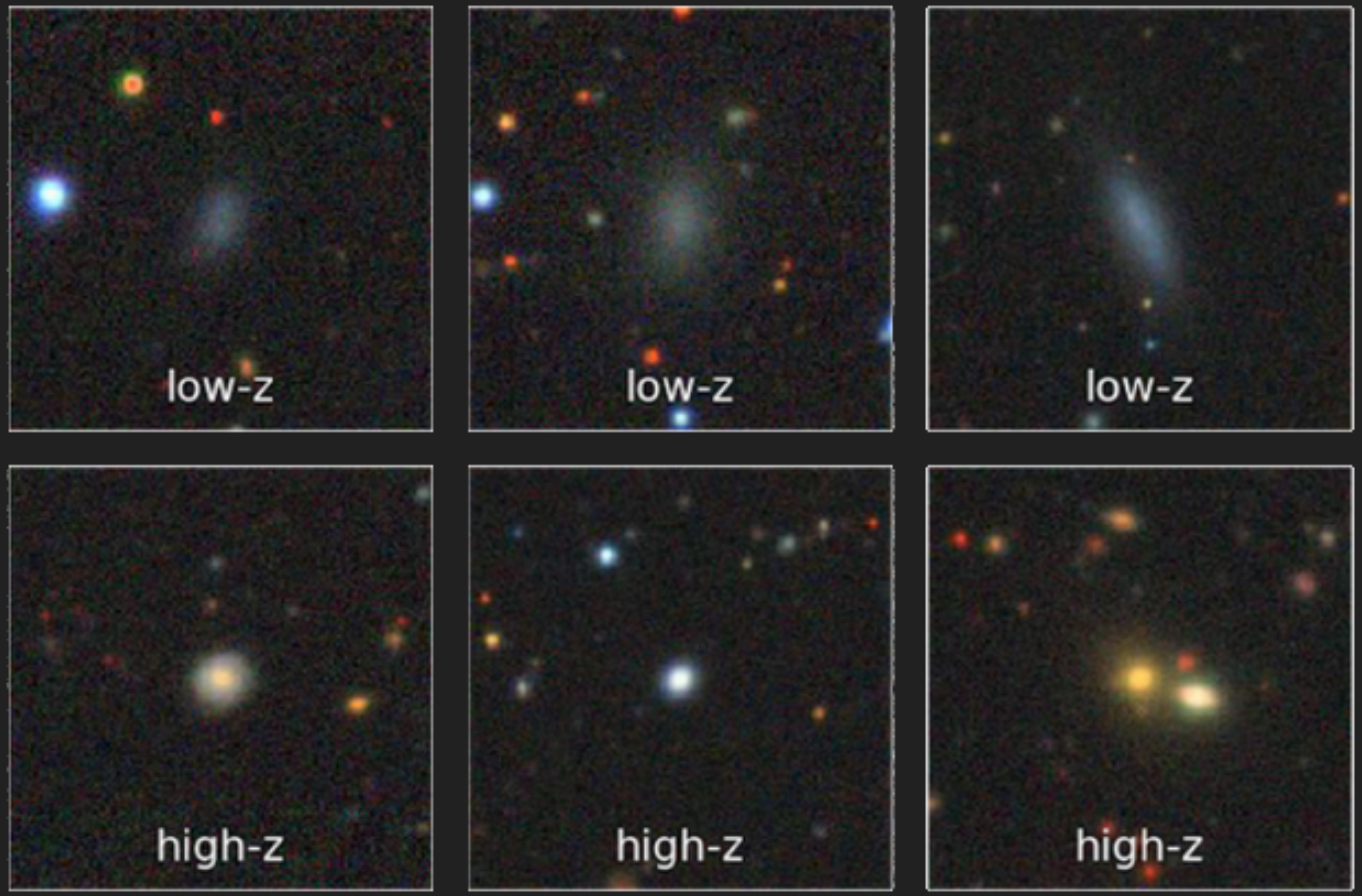
SAGA is the premier spectroscopic survey of low-z satellites

378+ new satellites around 101 hosts, using > 75,000 spectra

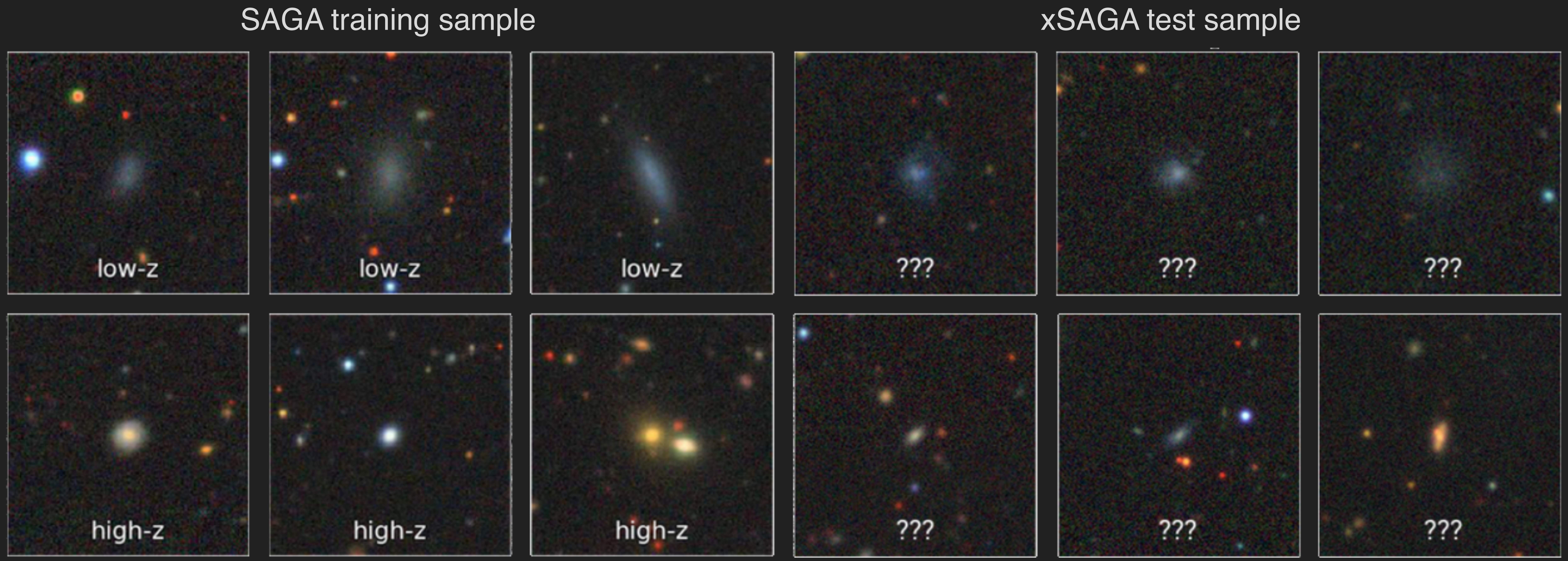
Geha+ 17, Mao+21, Mao+24

A CNN robustly selects low-z galaxies

SAGA training sample

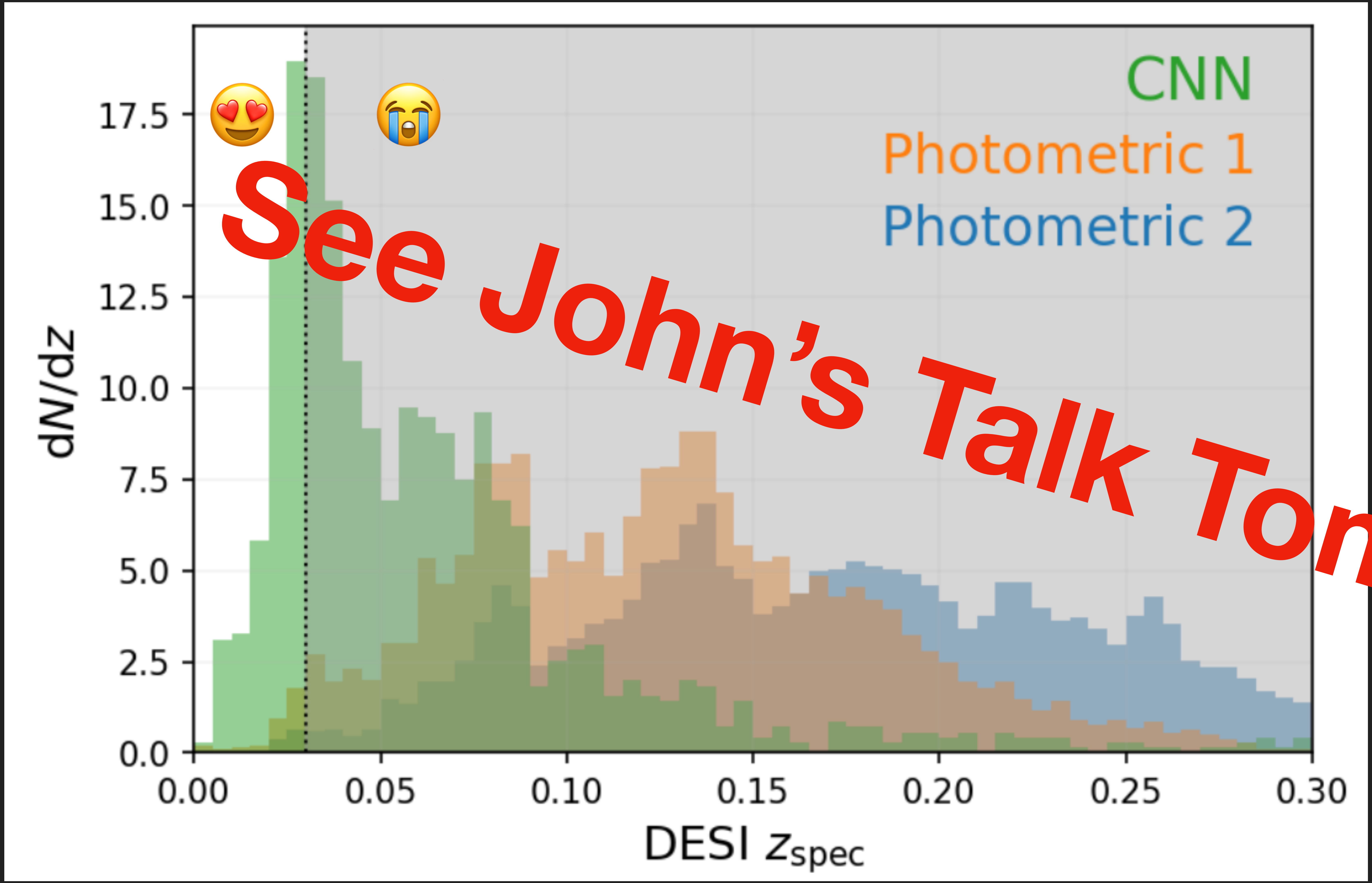


A CNN robustly selects low-z galaxies



Wu, JEGP+ 22

CNN is ~15x better than photo selection tested nightly on DESI



See John's Talk Tomorrow!

20.1%
1.5%
0.9%

Percent of **DESI LOWZ** targets at **z < 0.03**

Darragh-Ford, **Wu, + JEGP +** 22


SDSS gri


Legacy Survey *grz*


HSC Survey *grz*
(similar to LSST)


Hubble F606W/F814W
(similar to *Roman*)
We are just getting started…

What is ML in Astronomy?

Supervised vs. Unsupervised

Supervised Learning in Astronomy

Supervised Learning: *Galaxy Images*

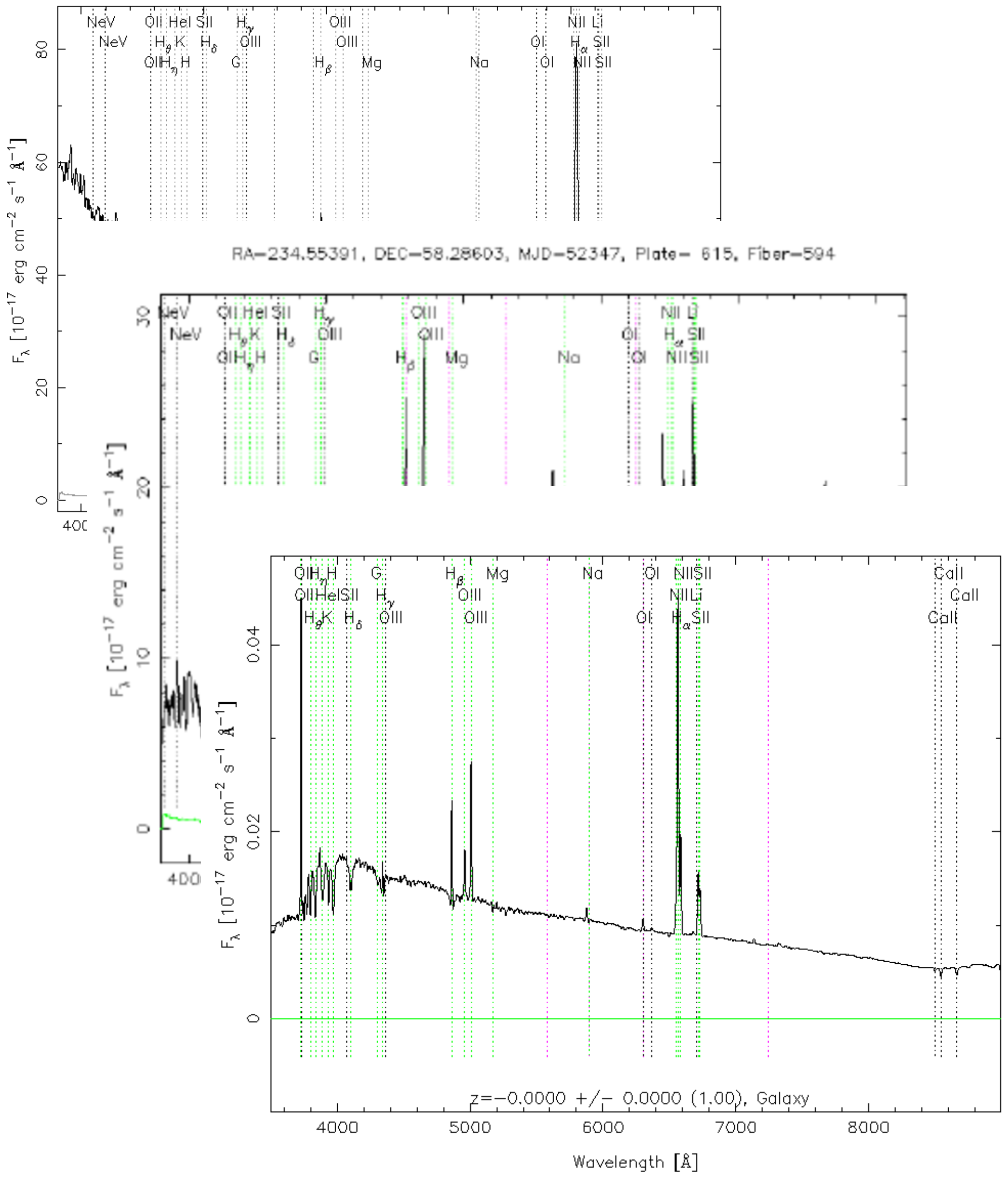
Unsupervised Learning in Astronomy

Unsupervised Learning: *Search By Image*

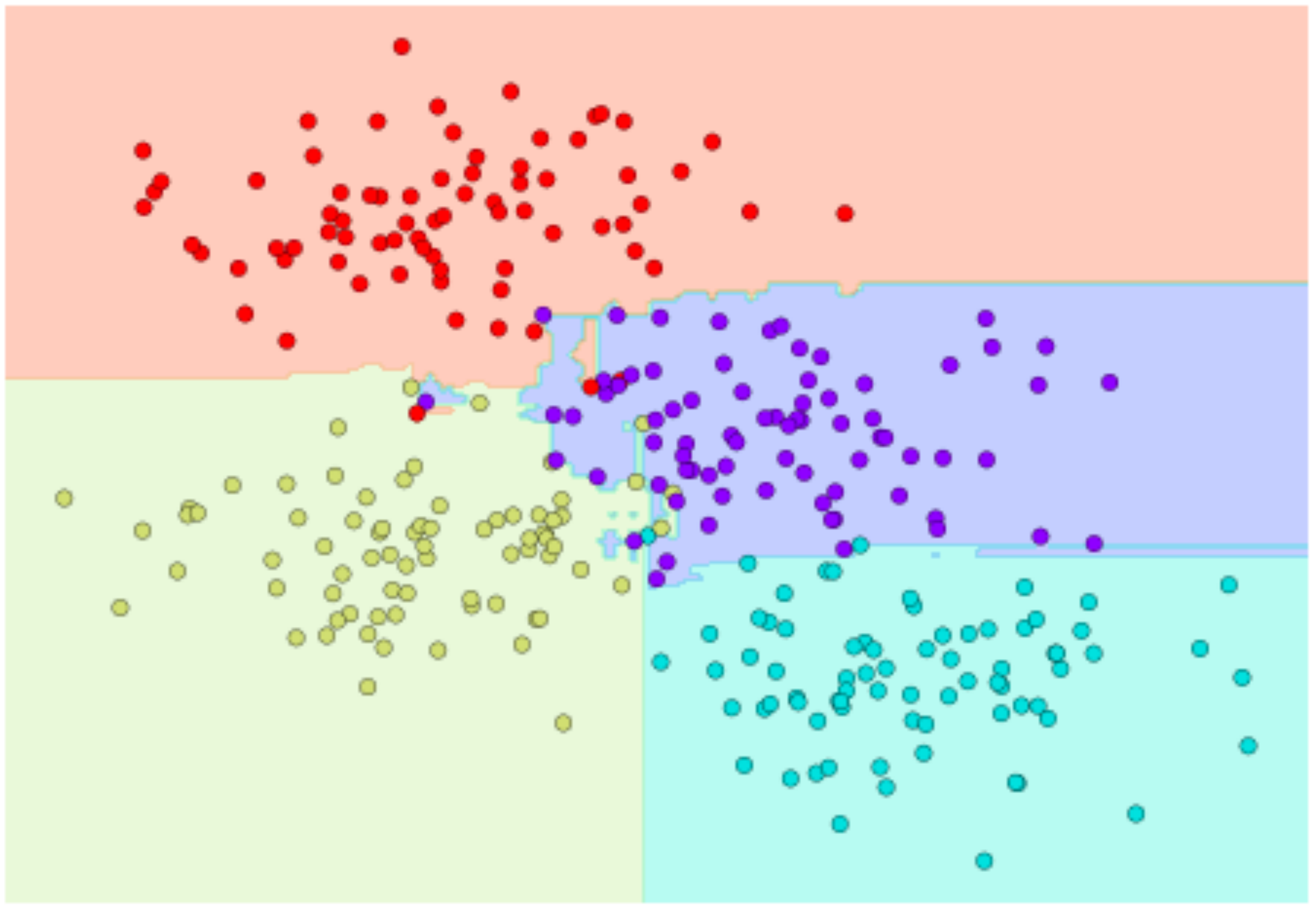
Why Unsupervised (or Self-Supervised) Learning?

hypothesis generation
in the big data era is
an unsolved problem

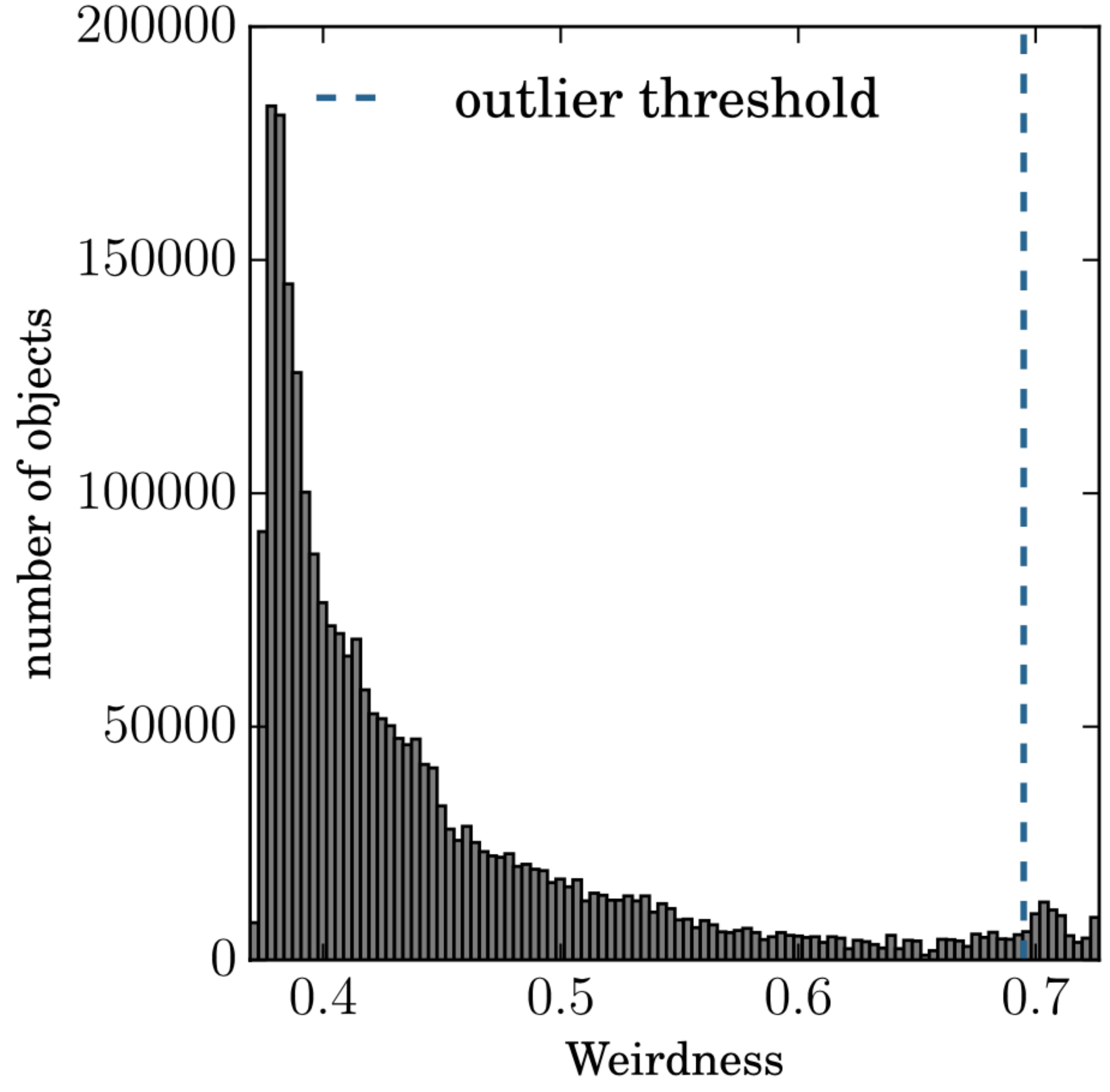
Finding the *weirdest* Galaxies in SDSS spectra



2e6 de-redshifted spectra with 15000 "features"



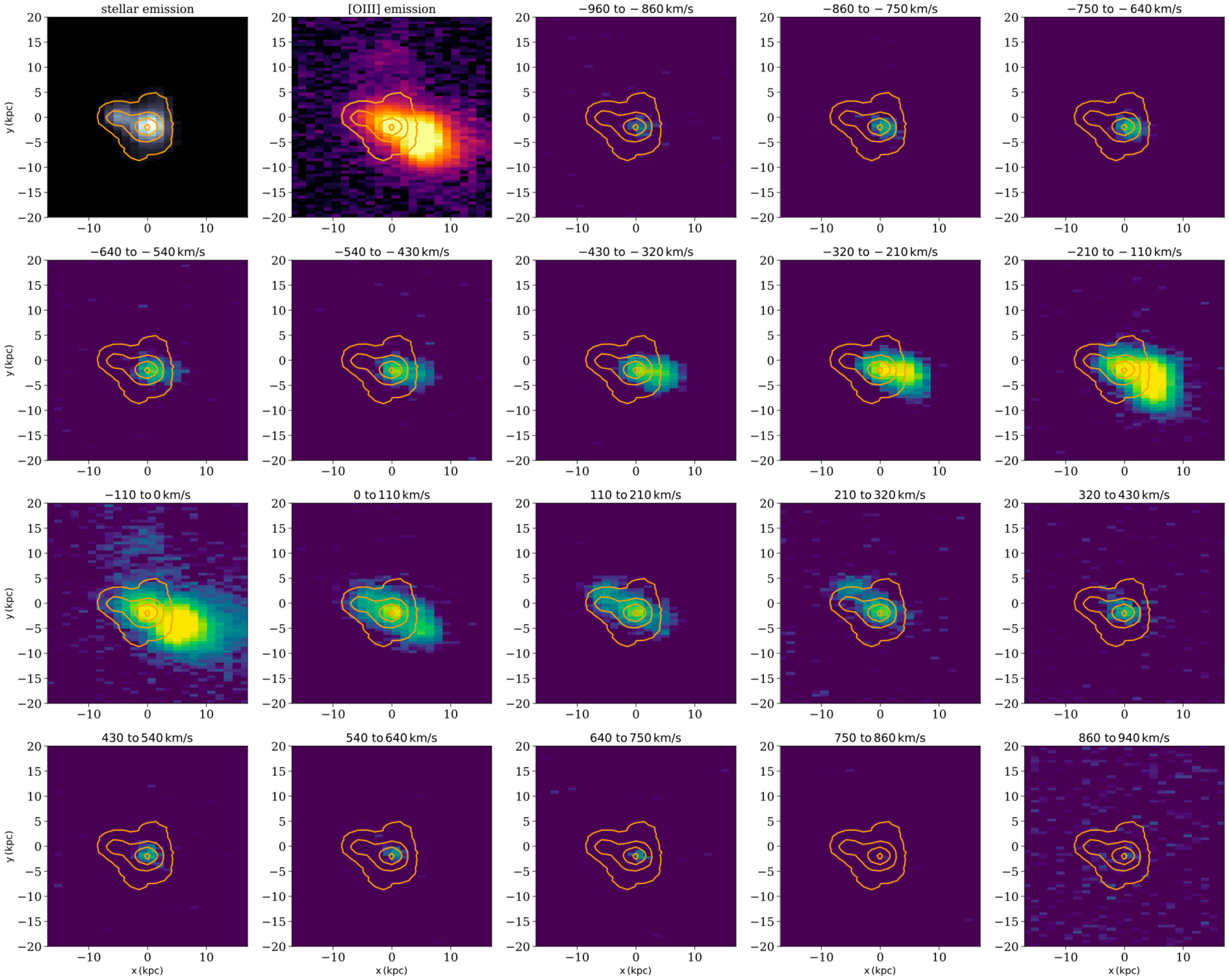
Random Forest to find similarity: how often in the same leaf?



Weirdness score: on average how far from every other galaxy?

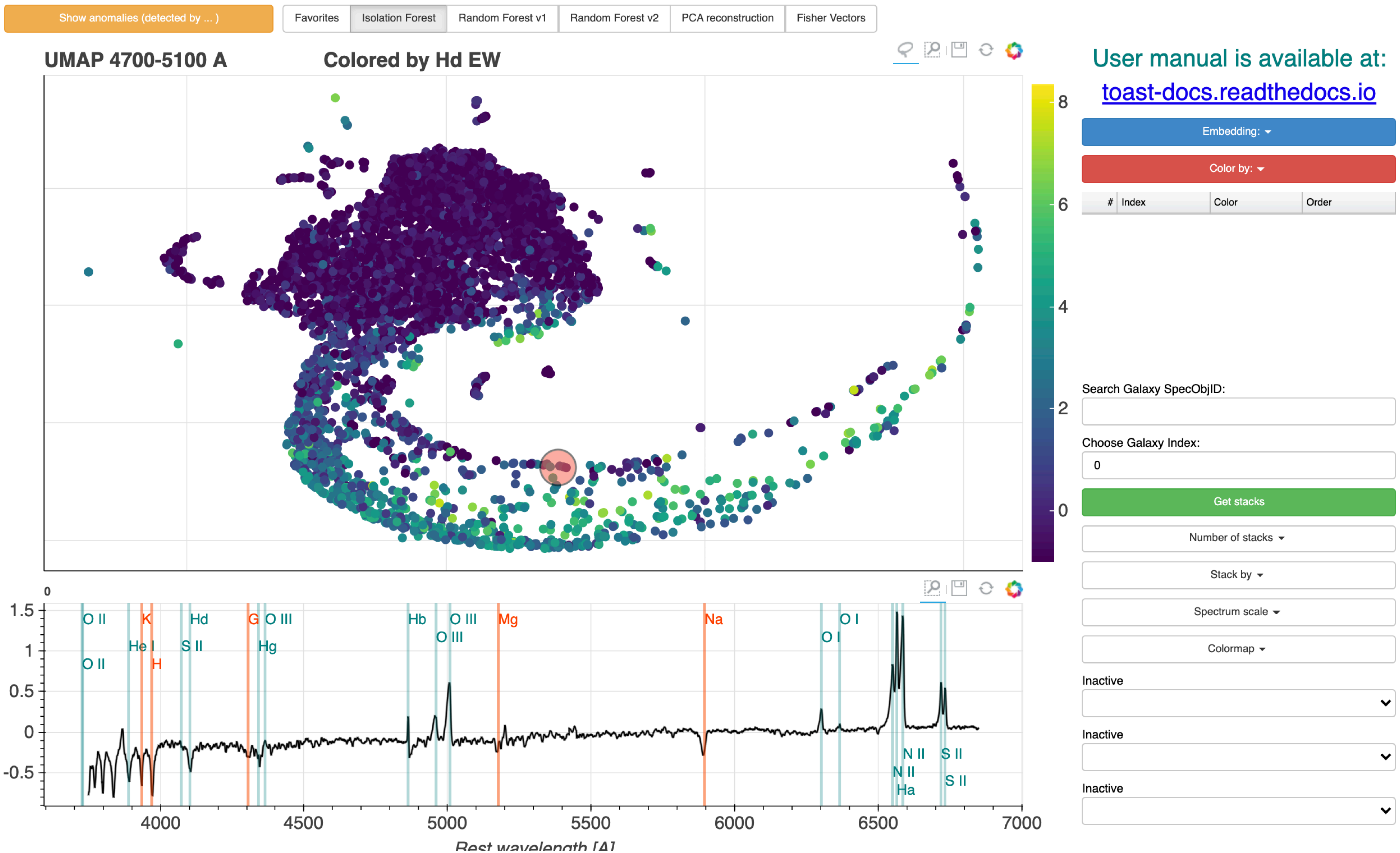
Baron & Poznanski 2017

Detailed studies with KCWI show an E+A galaxy with AGN winds



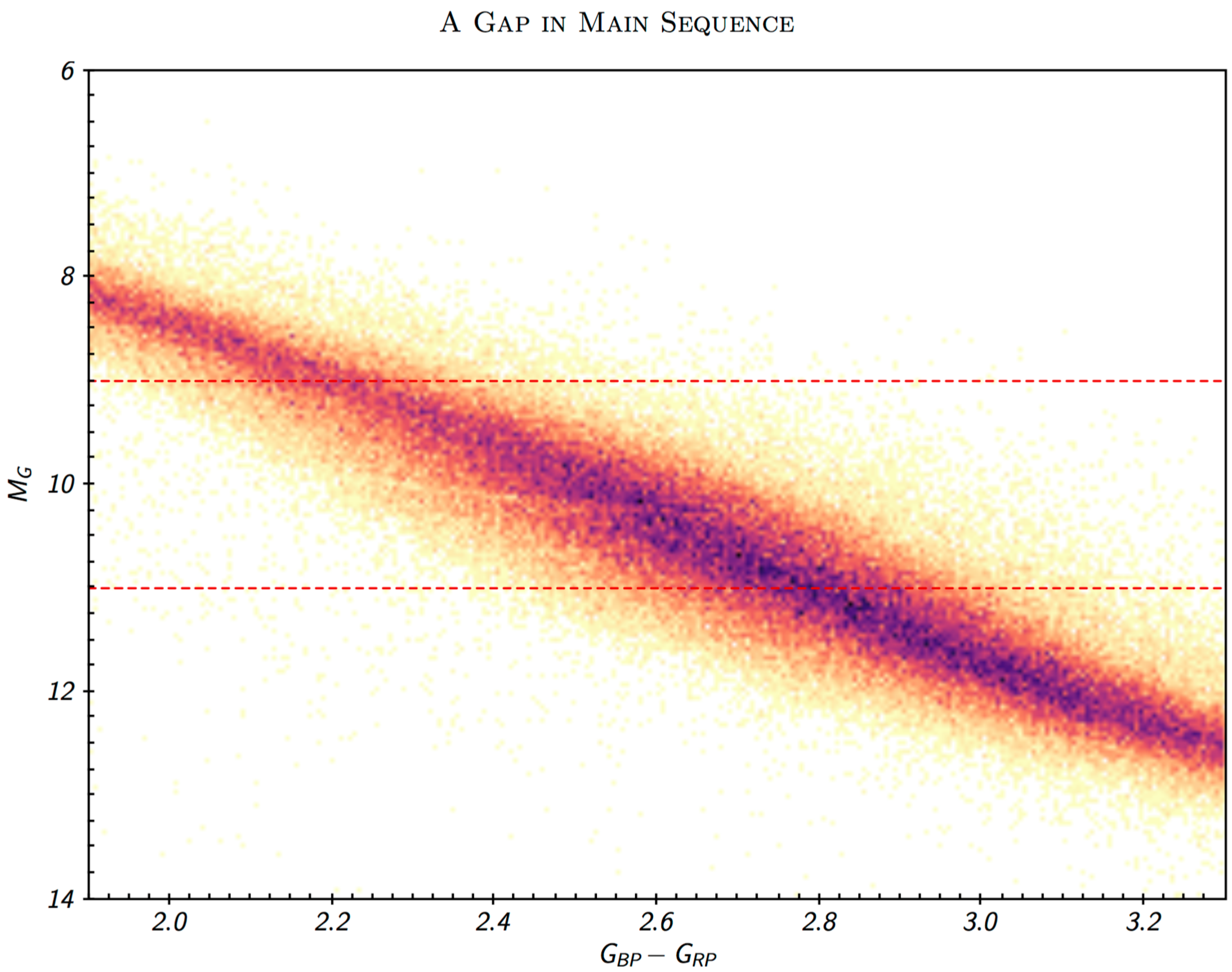
Baron+ 2018

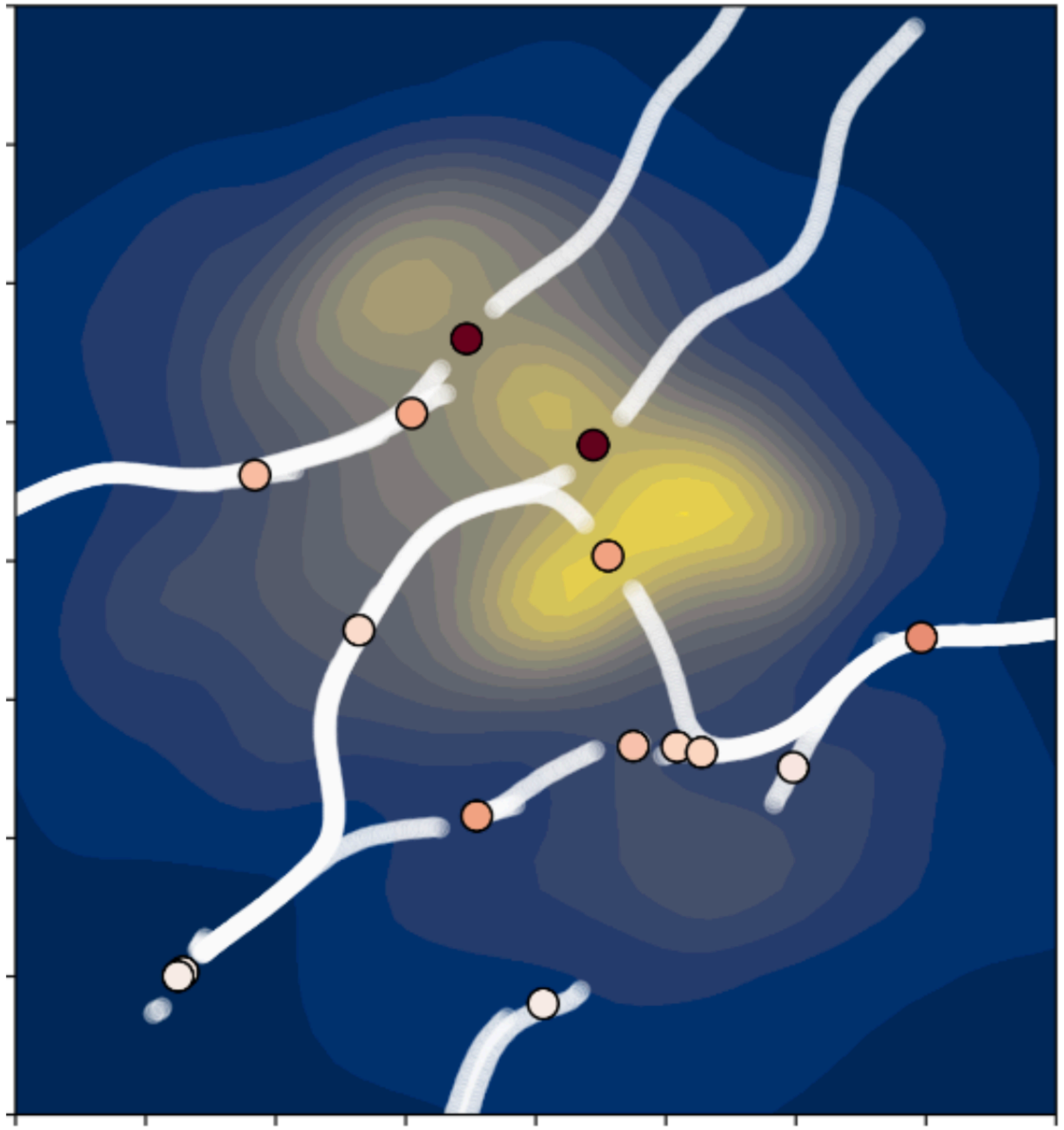
All galaxies are public, but *few have been explored*



http://galaxyportal.space, *Reis+ 2018*

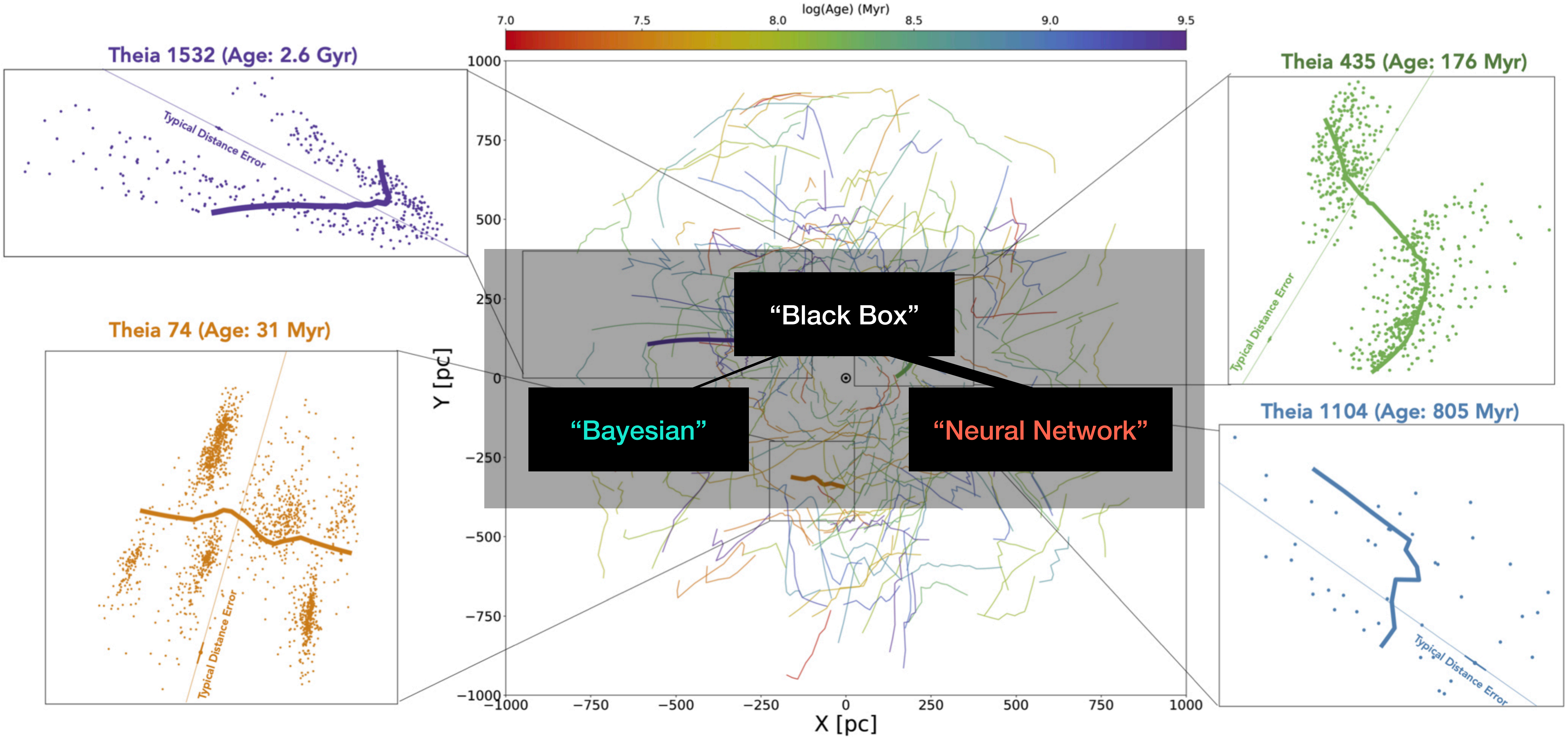
We also need big data discovery beyond outliers: gaps





*Jao+ 2018, **Contardo**, Hogg, Hunt, JEGP, Chen ,2022*

But: Machine Learning is also a stumbling block



*Zucker, **JEGP**, & Loebman 2022; Powered by SciX*

What is ML in Astronomy?

Supervised vs. Unsupervised

Supervised Learning in Astronomy

Supervised Learning: *Galaxy Images*

Unsupervised Learning in Astronomy

Unsupervised Learning: *Search By Image*

How do you find data in an archive?

NASA science image archives allow users to find images by metadata:
What camera, filter, exposure time, PI, position?

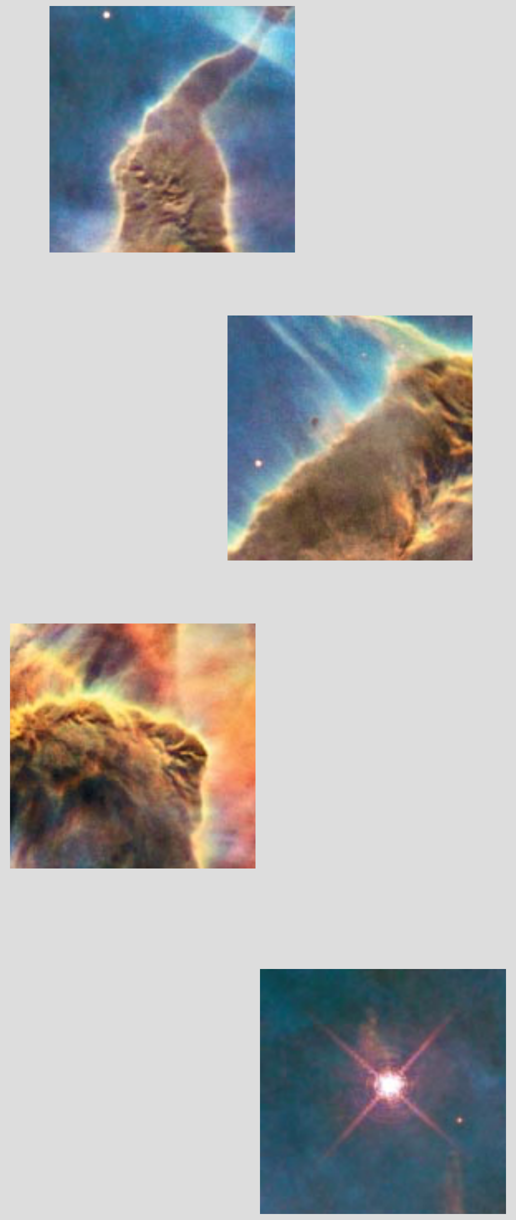
Catalogs of the objects contained in those images provide a limited search of the data itself:
Find images containing objects of a given brightness, position, basic shape …

We are developing tools to answer a much harder question:

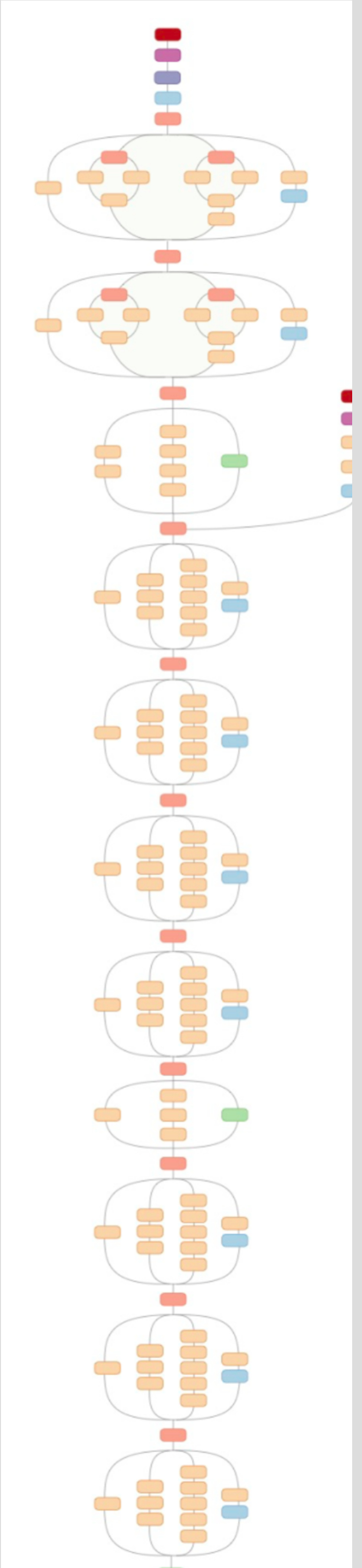
If I have a complex image, how can I find all the images in the archive that look like it?

The Prototype "Search by Image" Method

Image



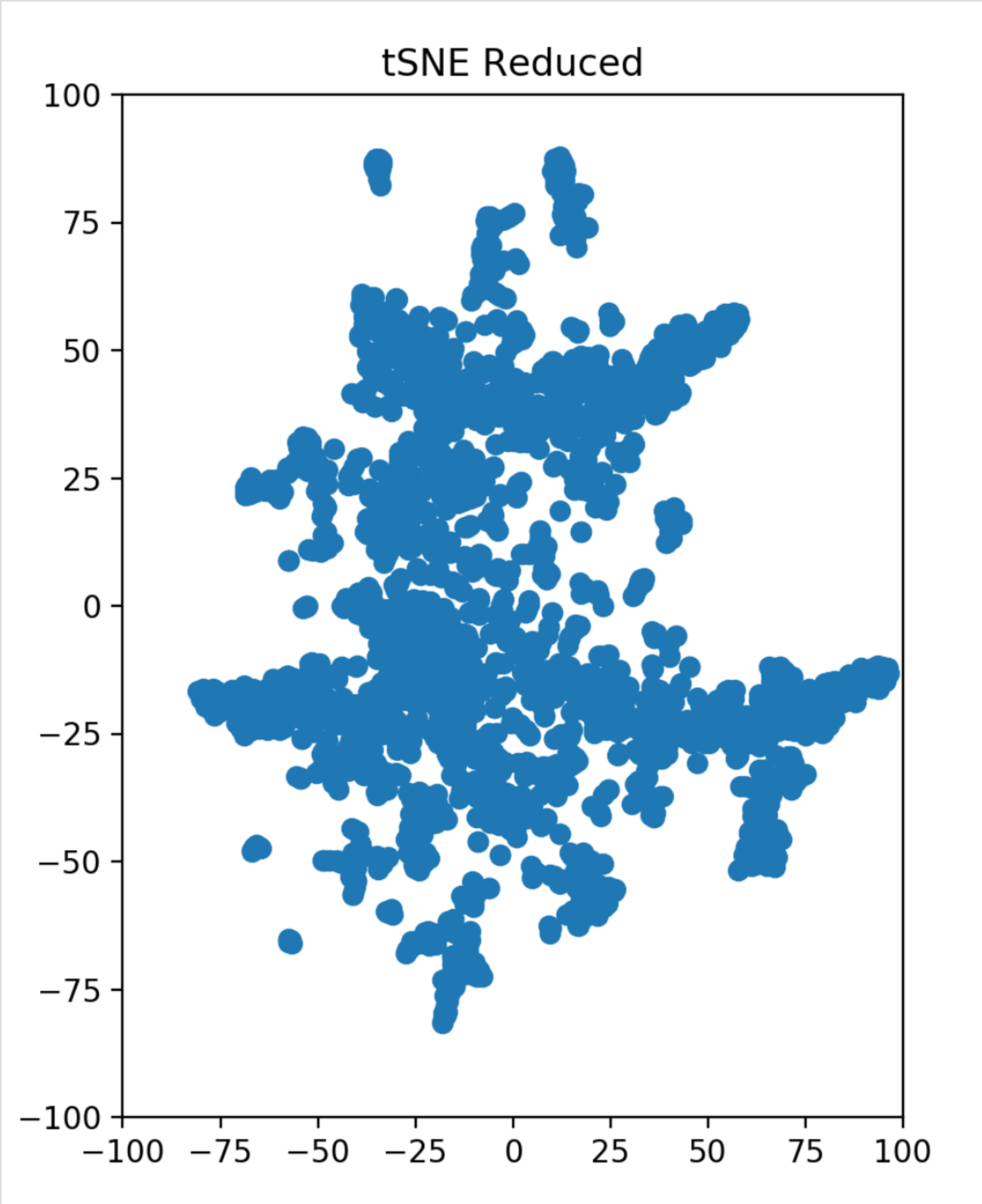
Cutouts



Existing Network

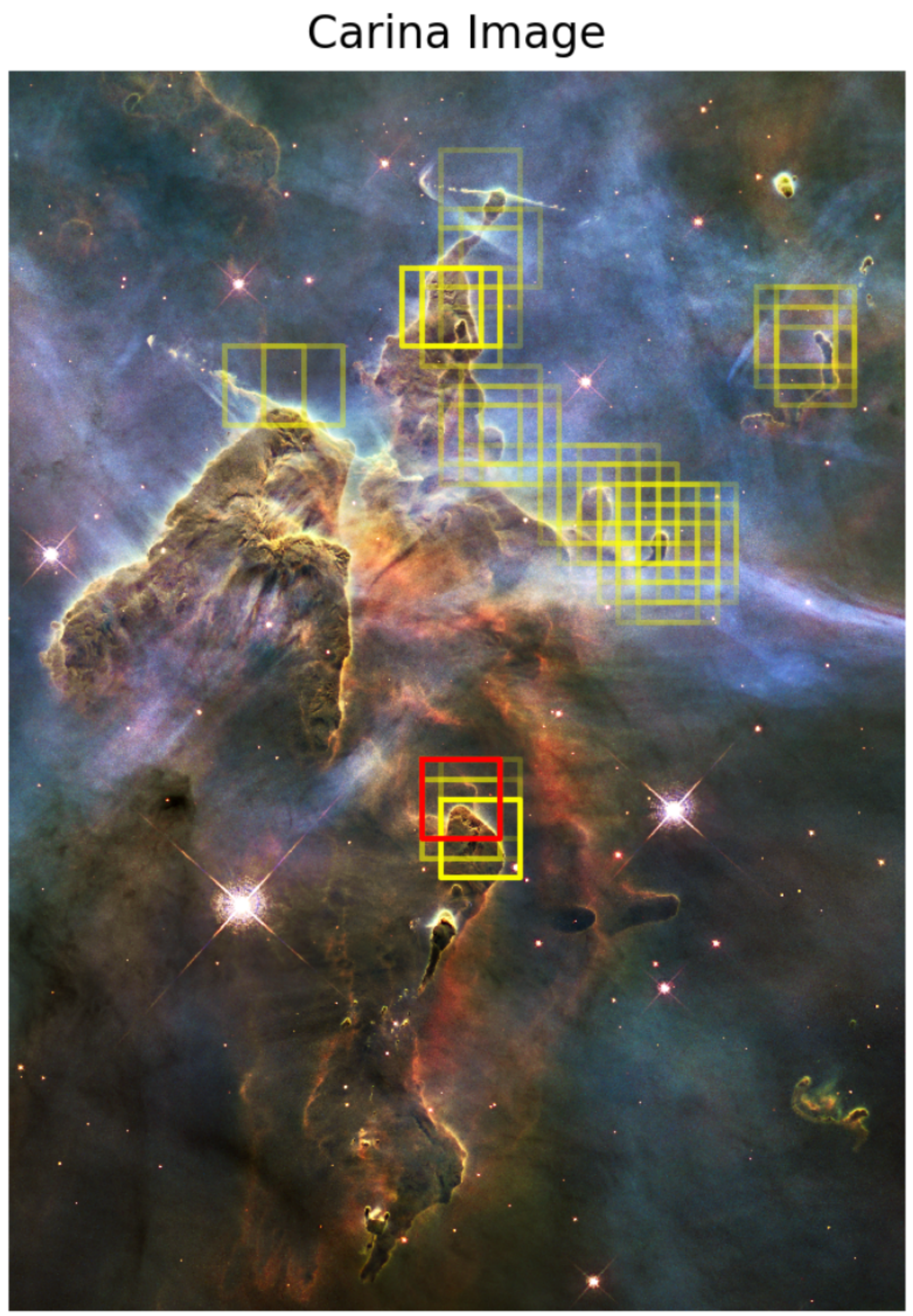
```
velvet:           10%
binder:          8.2%
jean:            6.4%
stingray:        4.5%
hammerhead       3.8%
book jacket:     3.1%
handkerchief:    2.1%
prayer rug:      1.5%
tiger shark:     1.4%
…
```

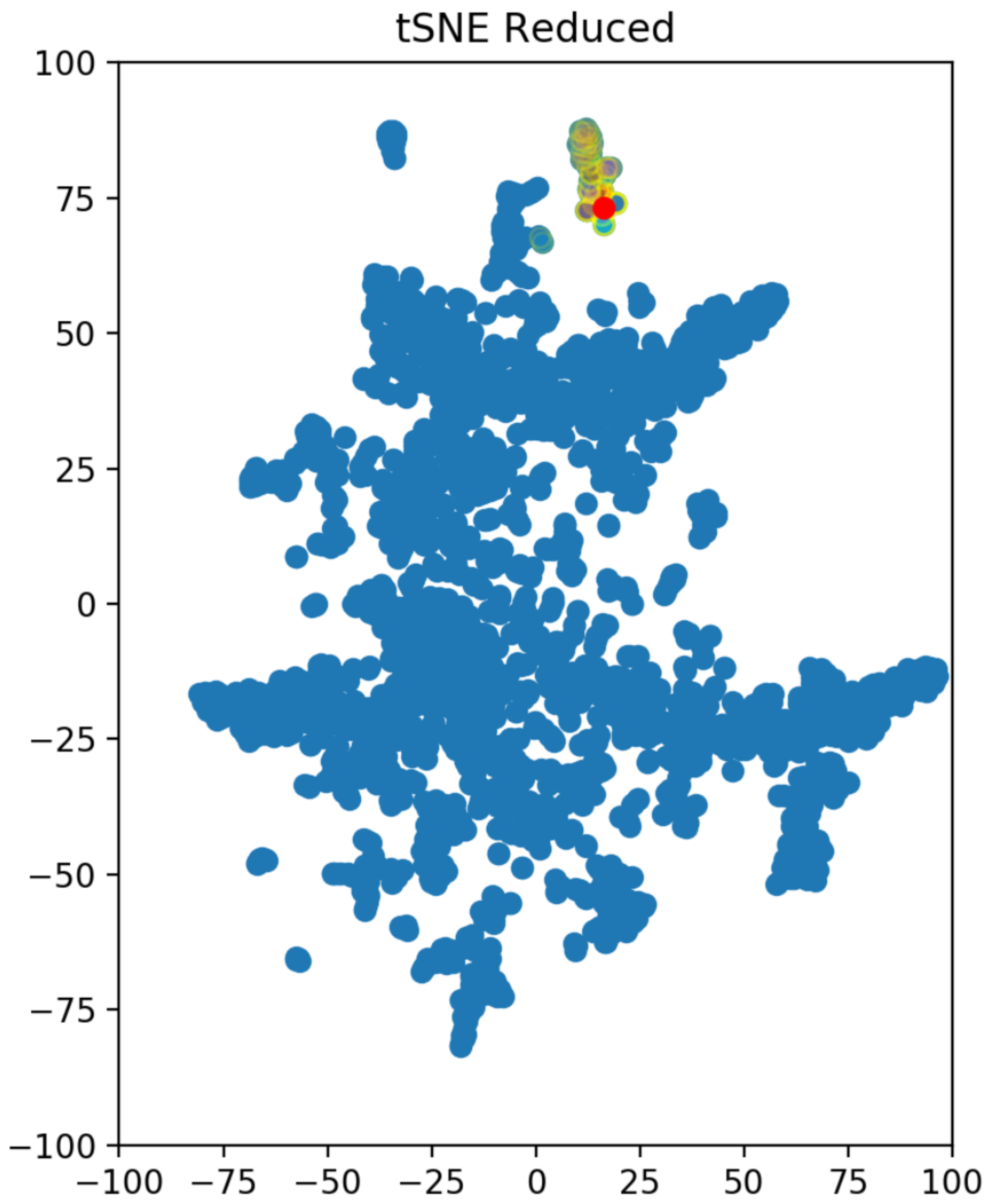
Feature Vector



Clustered data

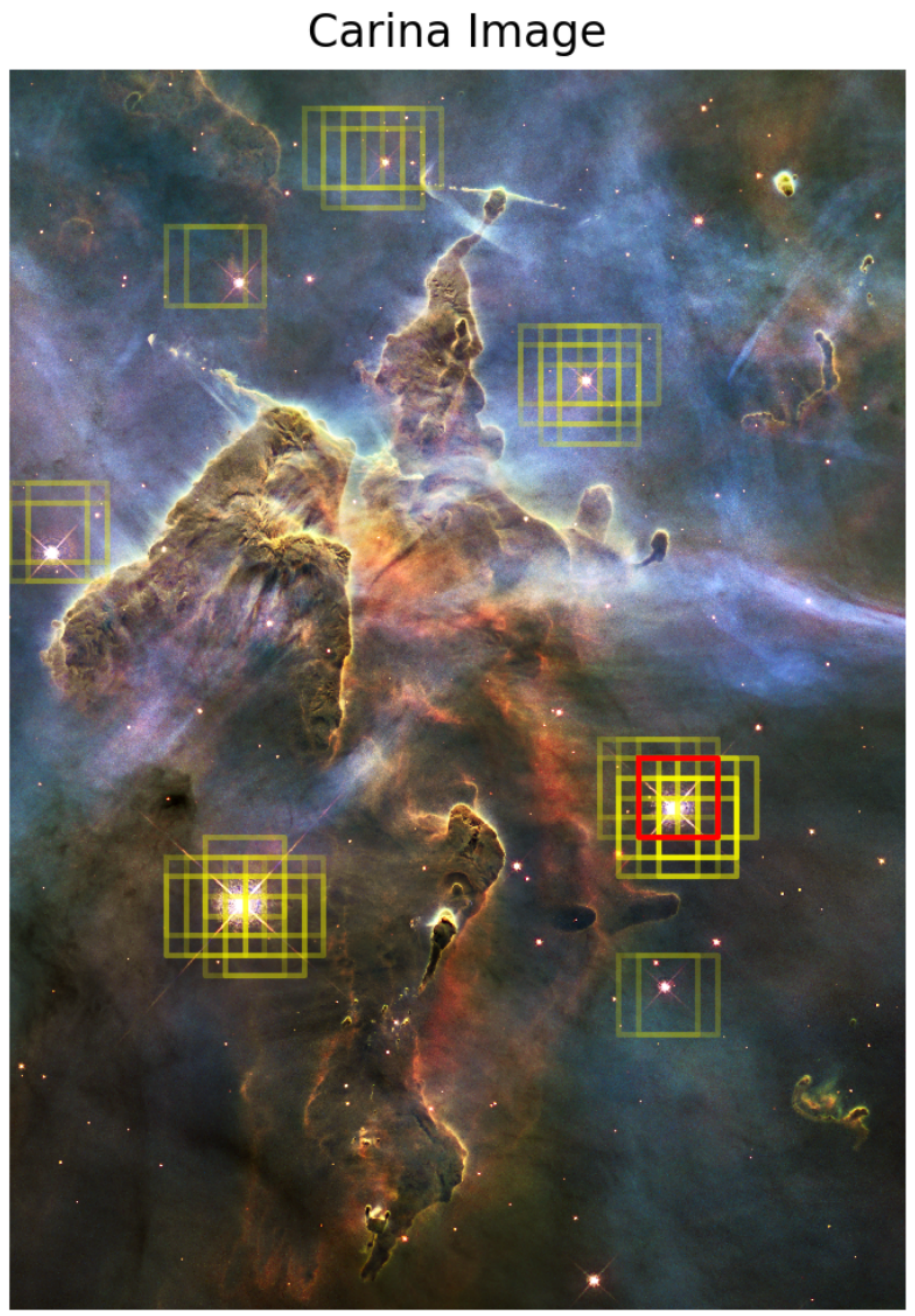
The Prototype “Search by Image” Method

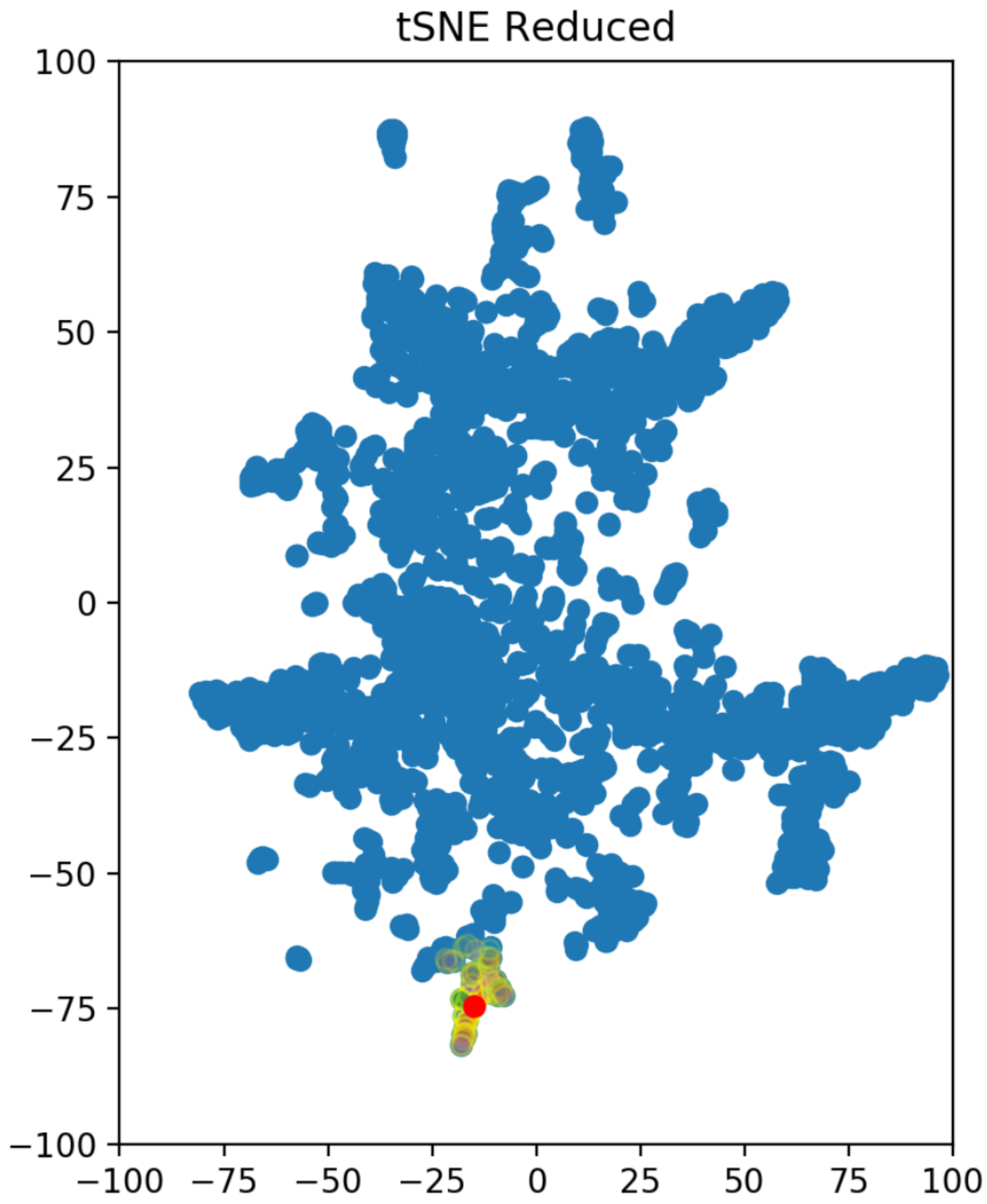




JEGP+ 2020

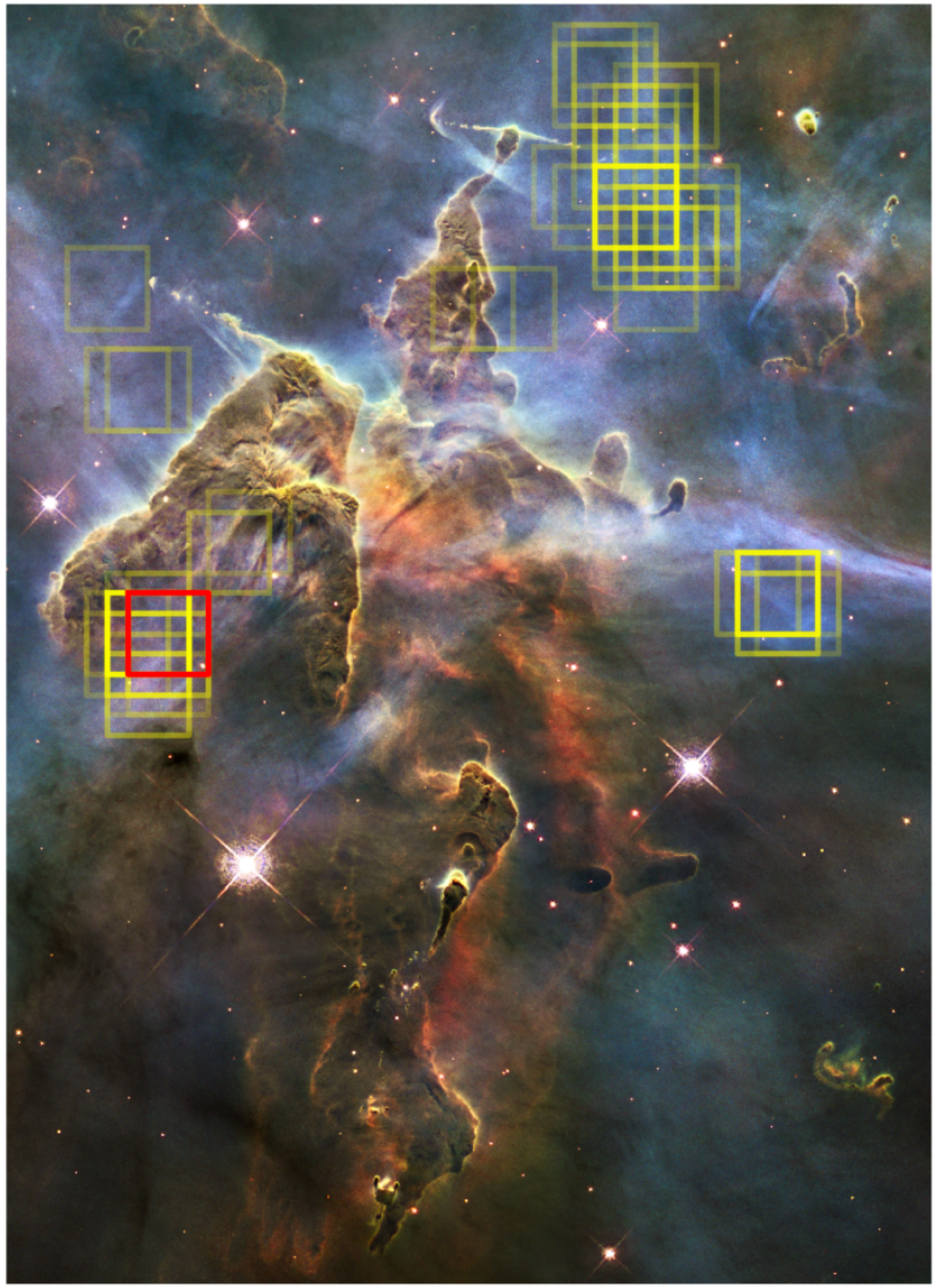
The Prototype “Search by Image” Method

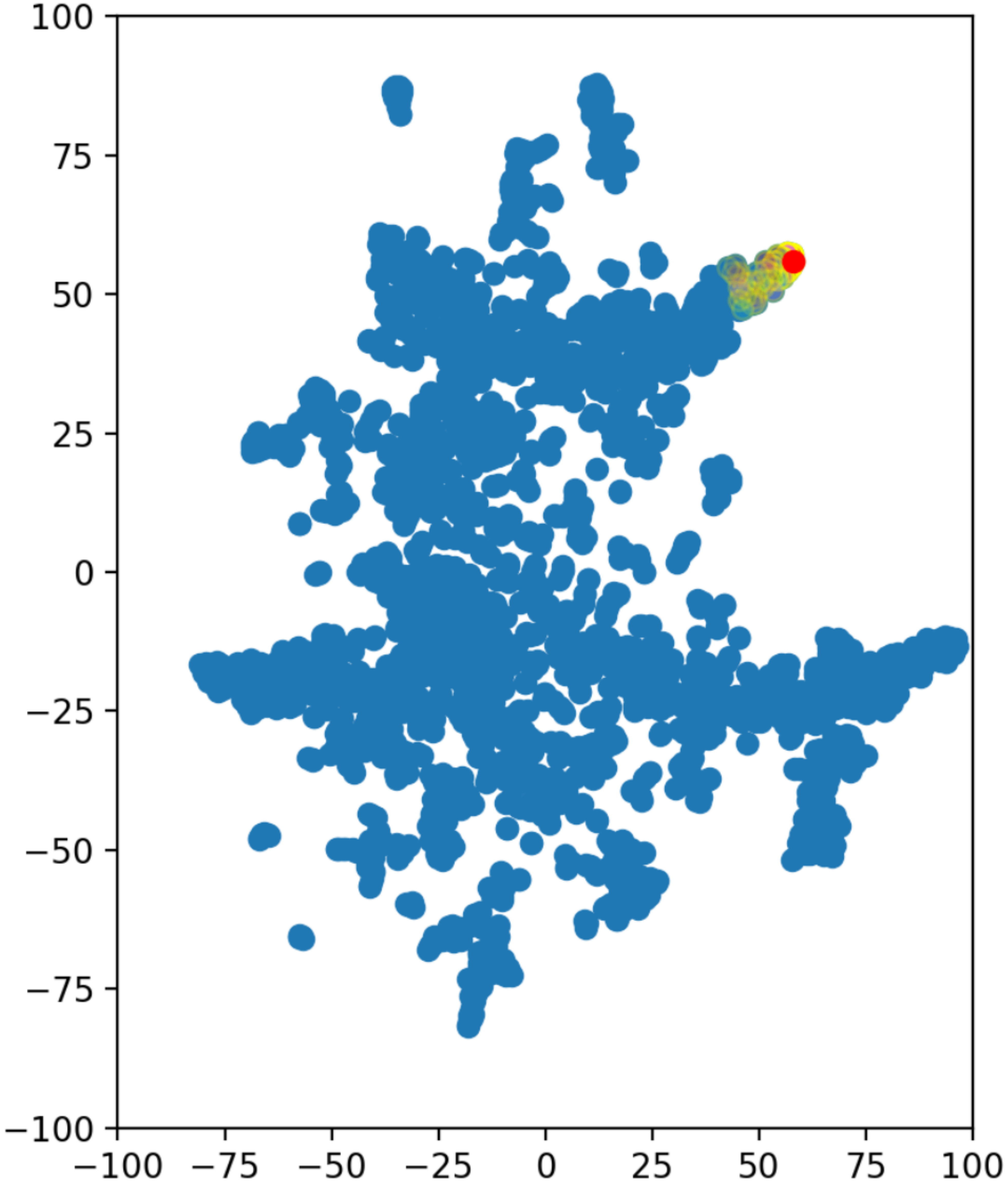




JEGP+ 2020

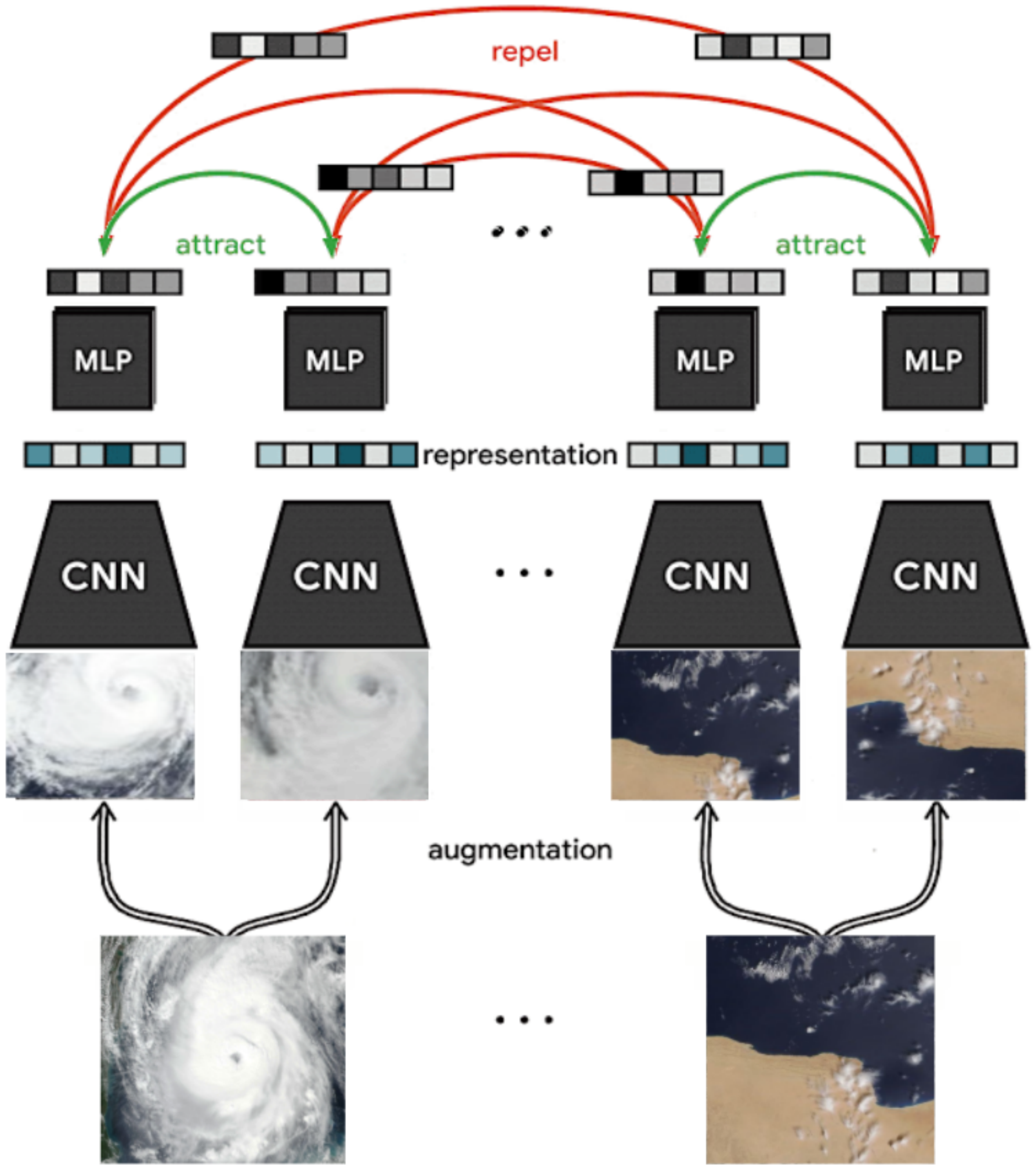
The Prototype “Search by Image” Method





JEGP+ 2020

Search by Image: a problem with many approaches for many sciences



Self-Supervised Learning is a generic approach to unlabeled data:

Modify images through rotation, stretch, cropping, and associate them to each other.

Earth Science is interested in similar problems with unlabeled data:

Finding hurricanes, wildfires, ice movement, etc

SpaceML is a team of industry professionals working with Earth Science on DL approaches

We teamed up with SpaceML to test SSL approaches and provide verification data sets

Walker Stevens for the SpaceML team

How do we optimize this approach? Which way is best?

We need a sample of

images with known similarities

to test, improve, and compare our

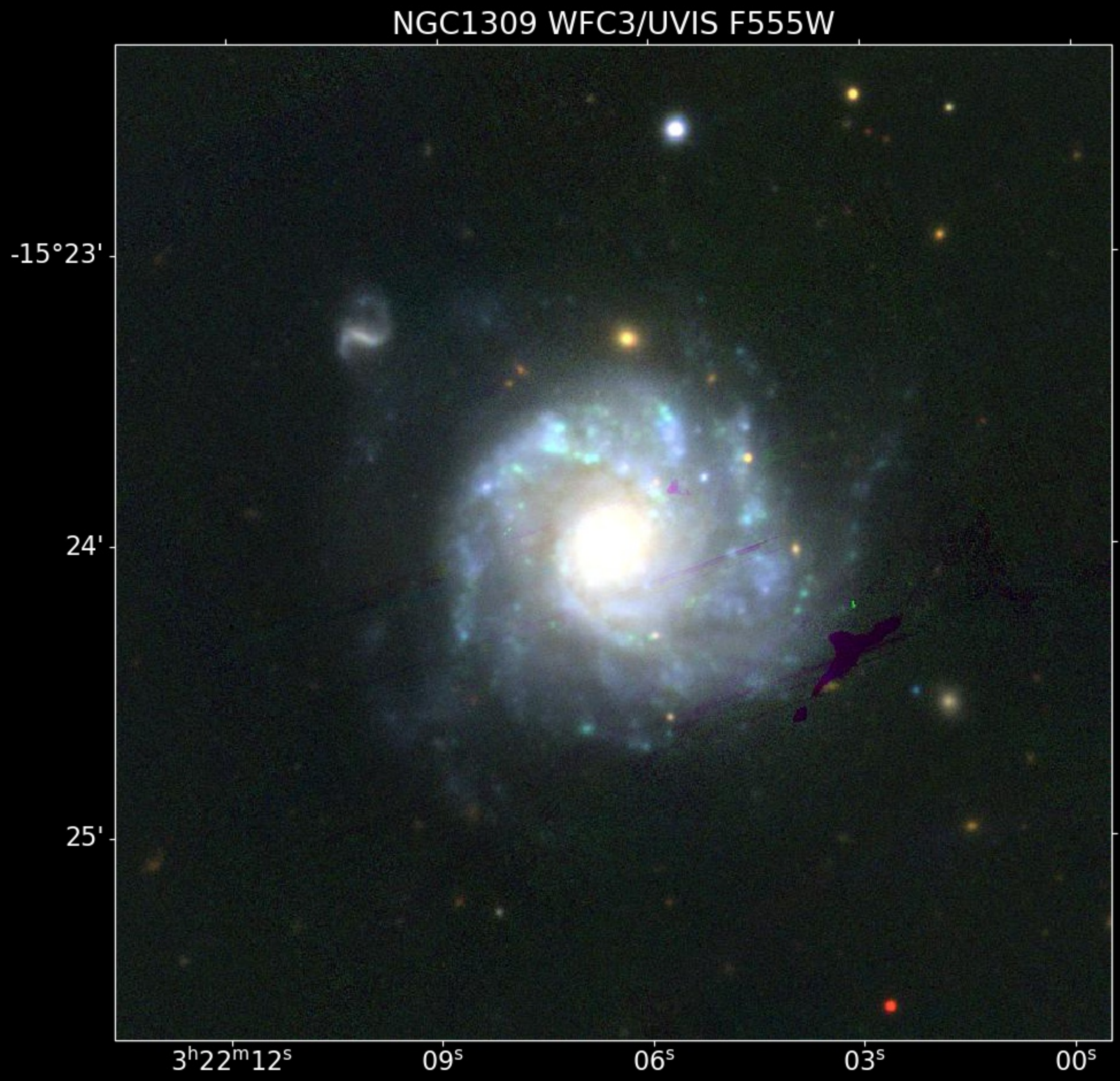
algorithms!

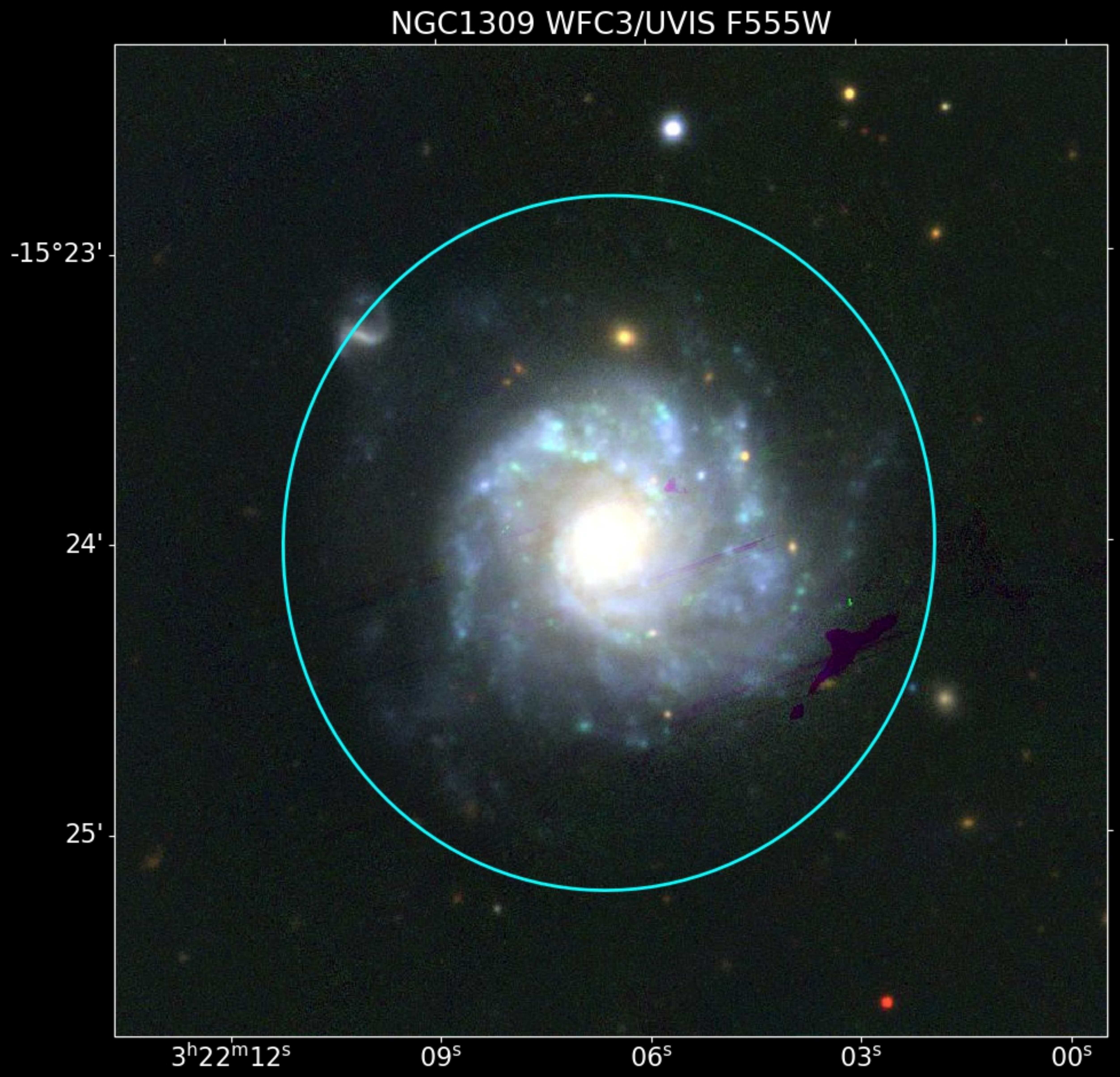
The Hubble and Planetary Image Similarity Projects (HISP/PISP)

HISP and PISP aims to create a large database of similarity information between segments of Hubble images (ACS & WFC3) and between segments of Mars Reconnaissance Orbiter (CTX) .

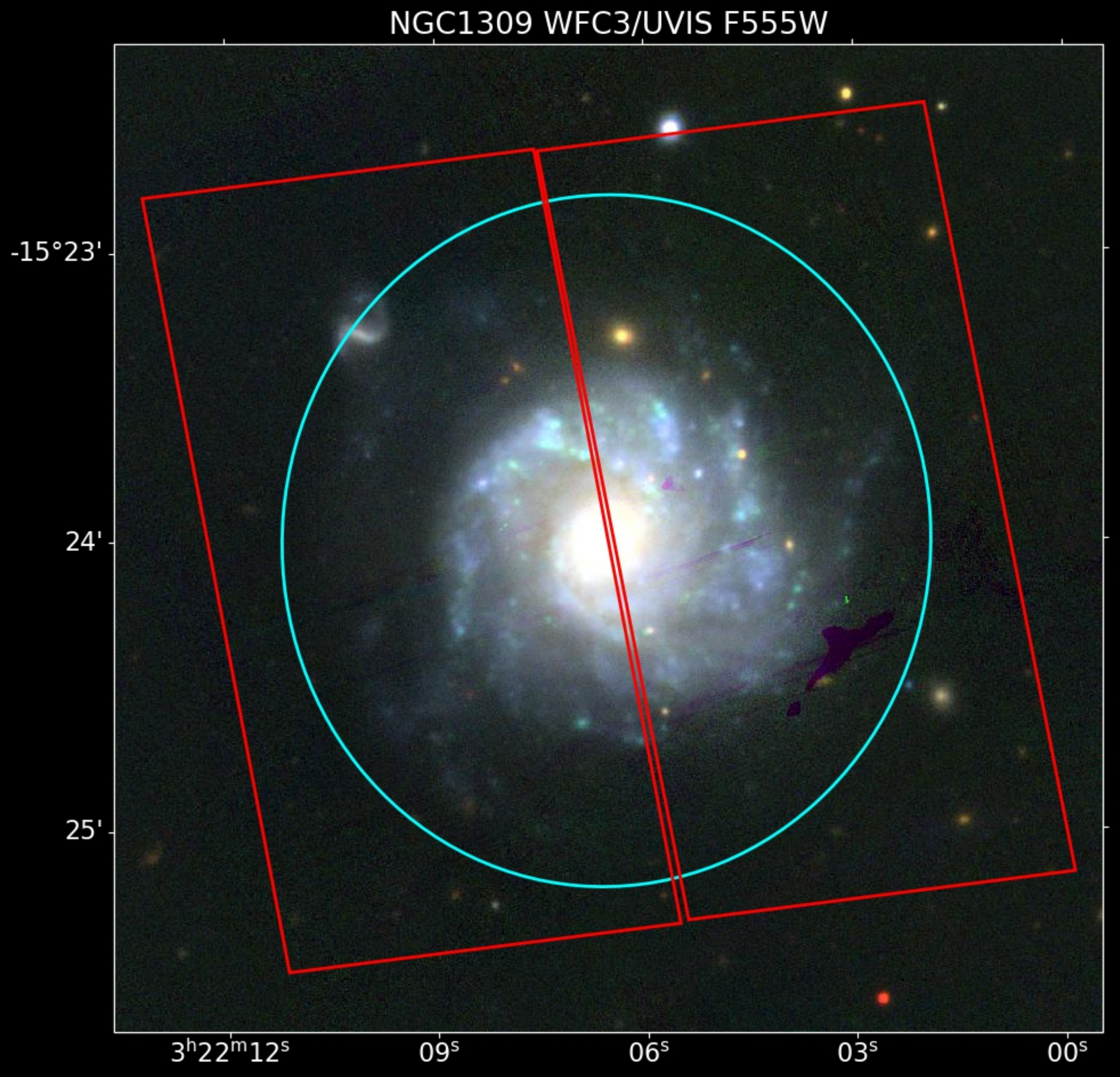
- The images are compared by humans in a citizen science project.
- We also designed the project for community impact:
  - We employ service-industry professionals from the local area near STScI in Baltimore who were impacted by the Covid-19 pandemic.
  - They are paid a fair wage for their work through the Amazon Mechanical Turk (AMT) system.

White & JEGP in prep


NGC1309 WFC3/UVIS F555W
-15°23'
24'
25'
3h22m12s
09s
06s
03s
00s

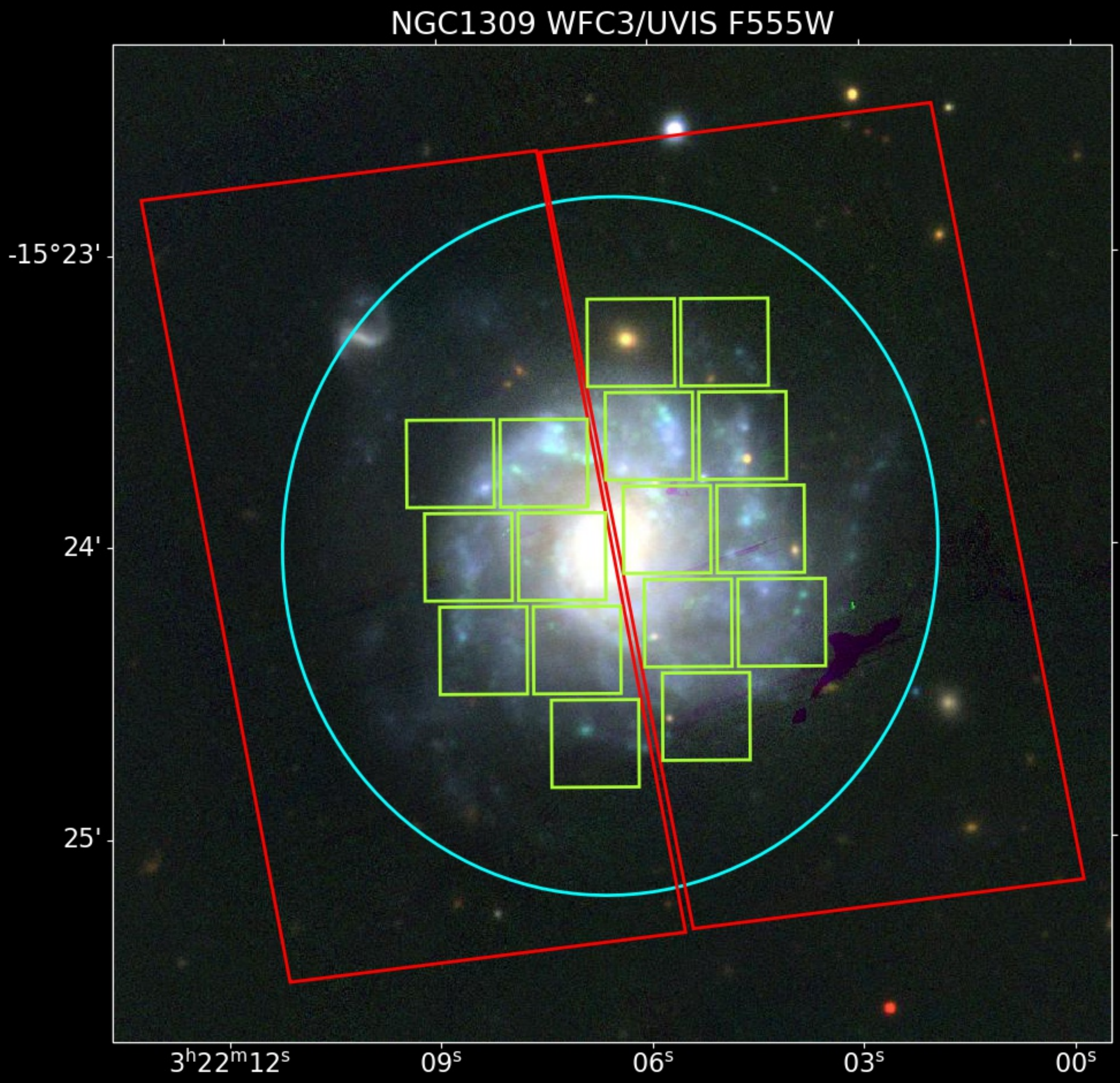


Elliptical NGC object

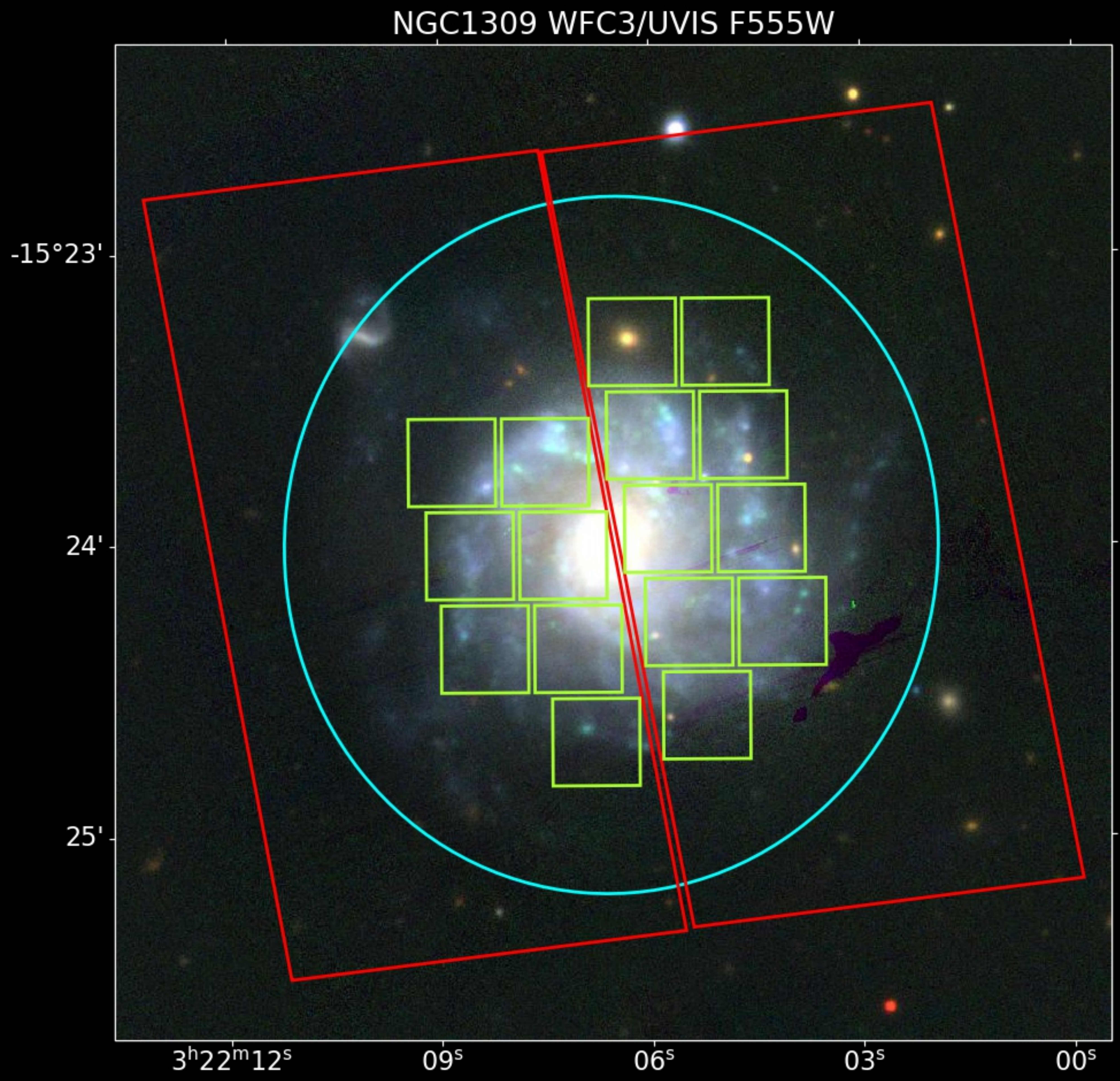


Elliptical NGC object

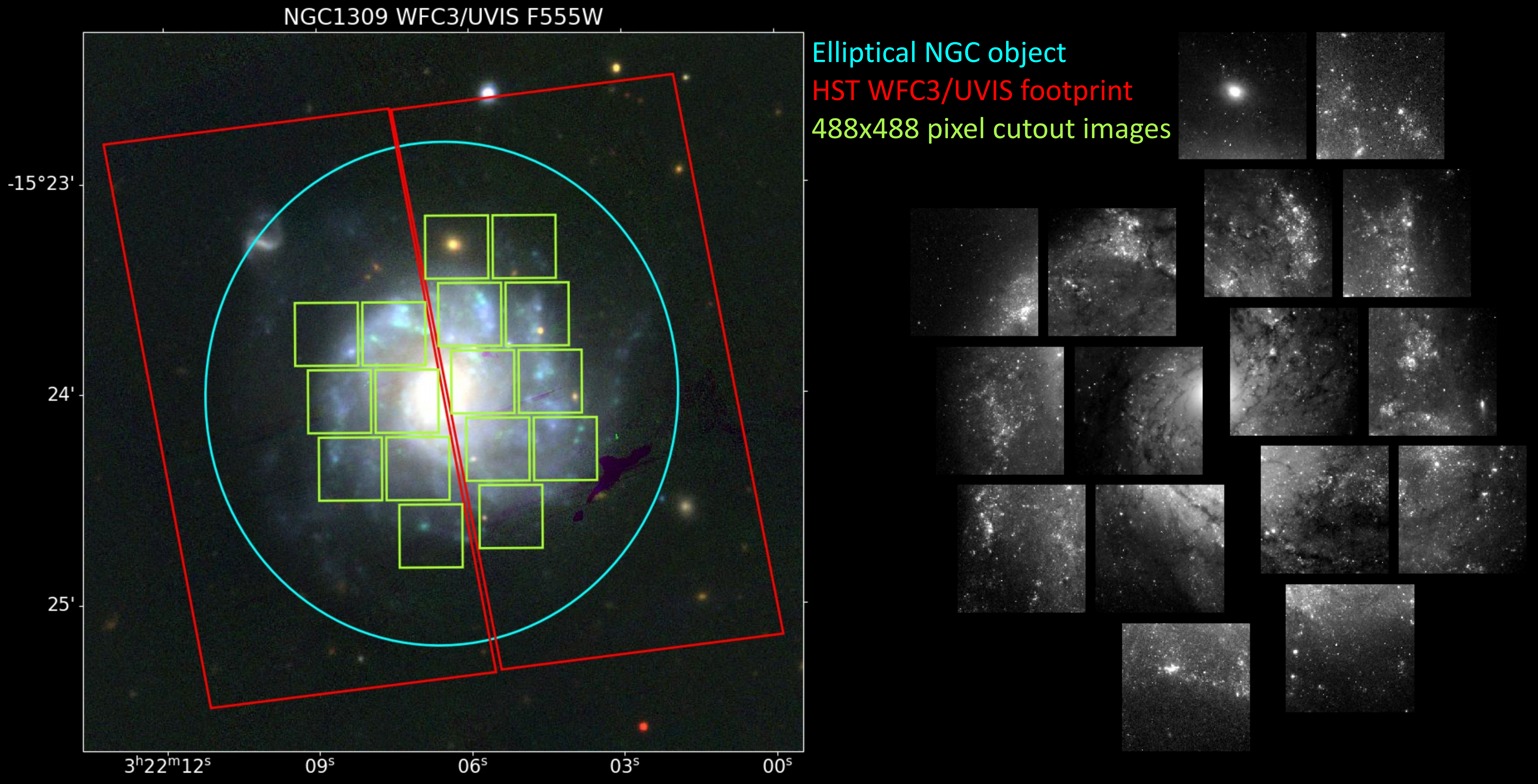
HST WFC3/UVIS footprint



Elliptical NGC object
HST WFC3/UVIS footprint
488x488 pixel cutout images



Elliptical NGC object
HST WFC3/UVIS footprint
488x488 pixel cutout images


NGC1309 WFC3/UVIS F555W
-15°23'
24'
25'
3h22m12s
09s
06s
03s
00s
Elliptical NGC object
HST WFC3/UVIS footprint
488x488 pixel cutout images

Sample includes nebulae, clusters, star formation regions

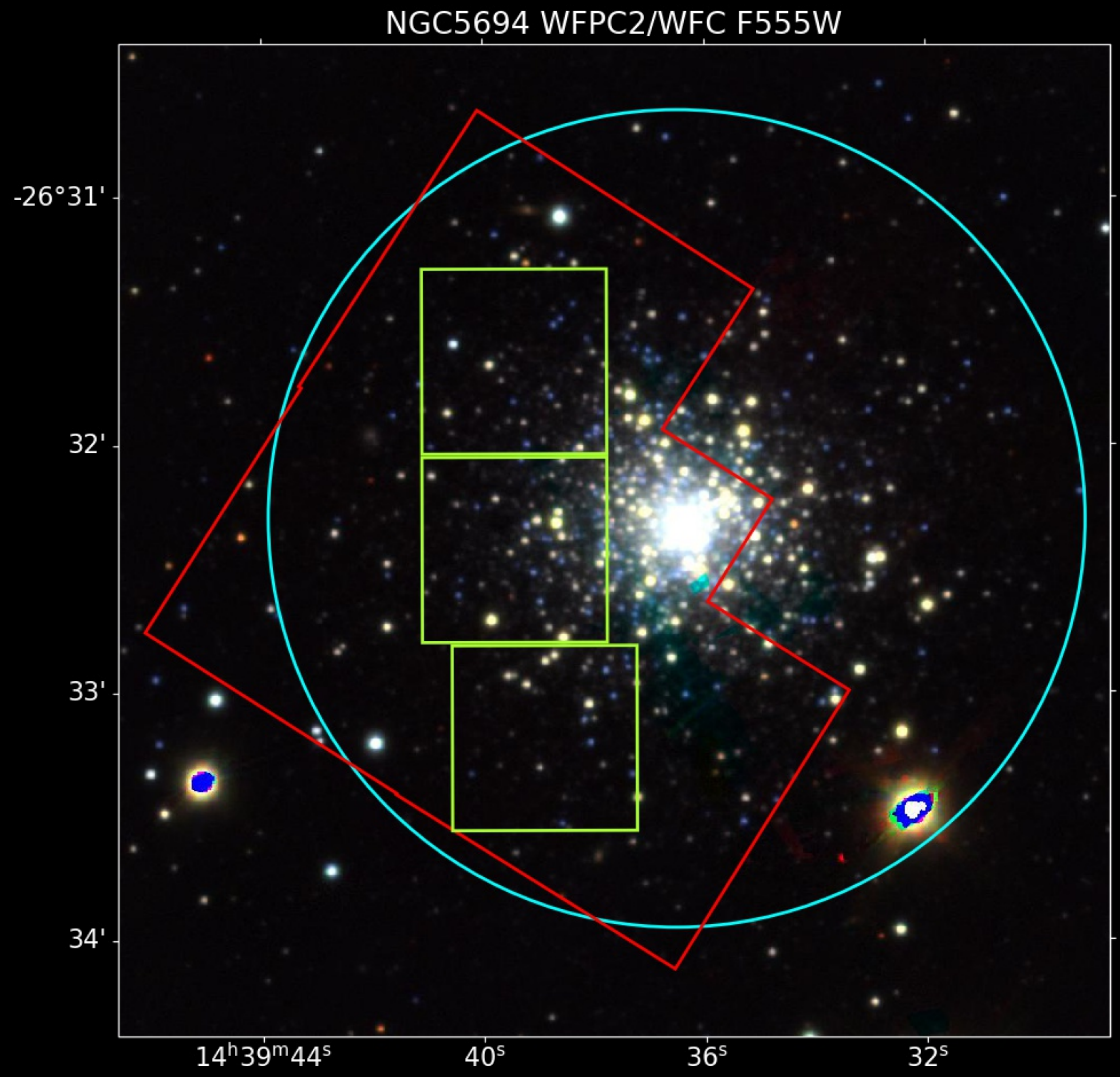
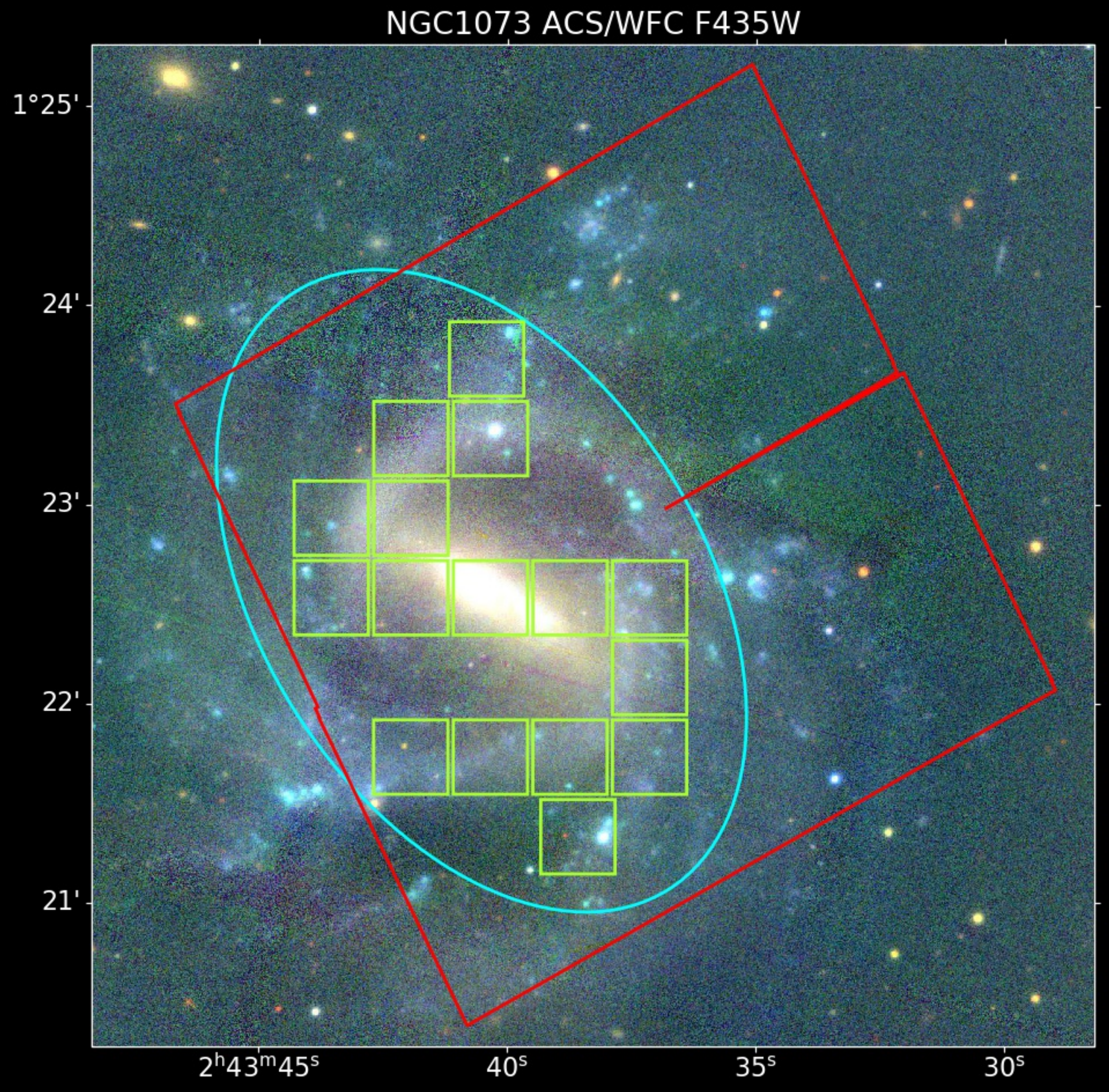
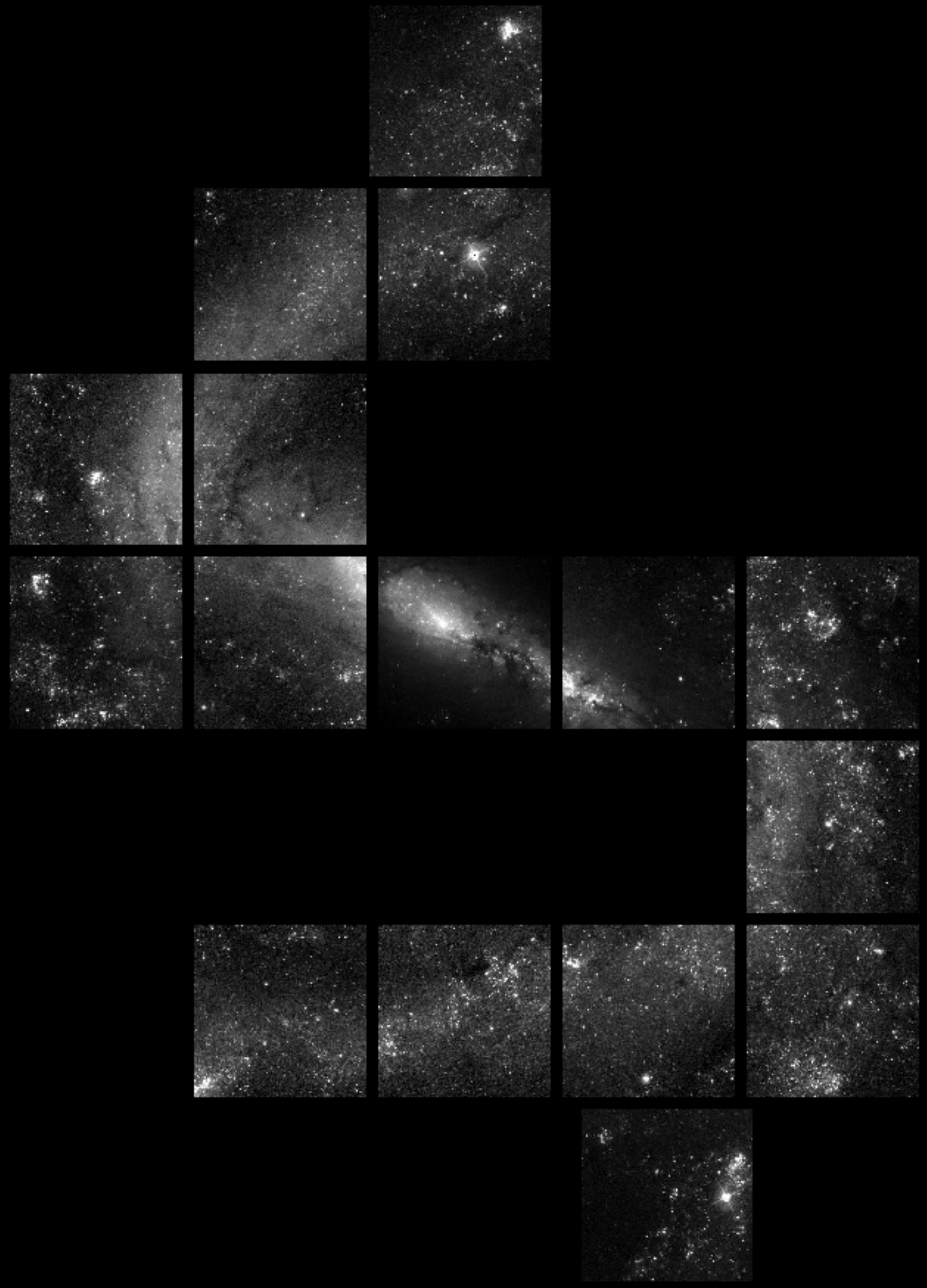


Image entropy selects cutouts with good contrast






Select similar images from the Hubble Space Telescope
Comparison image
Which of these 15 images are similar to the Comparison Image at left?
No images are similar
Submit
Select similar images from the Hubble Space Telescope
Comparison image
Pick the image that looks the most similar, ignoring the rotation of the image.
Submit
Submit
Submit
Submit
Submit

Select similar images from the Mars Reconnaissance Orbiter

Comparison image

Which of these 15 images *are similar* to the Comparison Image at left?

☐ No images are similar

Submit

Submit

Submit


Click to show closest images from any cluster
Cl 16 (28 members)
40
20
0
−20
−40
−40
−30
−20
−10
0
10
20
30
40
x=-2.578 y=-33.481
IC4641 (-2.5,-33.9) 0.000
NGC3021 (-1.7,-33.4) 0.897
NGC3278 (-1.8,-33.1) 1.058
NGC0278 (-2.8,-32.8) 1.105
NGC3982 (-1.4,-34.5) 1.213
NGC7823 (-3.2,-34.9) 1.243
NGC2255 (-1.7,-34.9) 1.259
NGC6181 (-1.2,-33.9) 1.294
NGC3949 (-1.8,-32.8) 1.300
NGC3310 (-3.8,-34.3) 1.376
NGC4940 (-1.1,-34.3) 1.417
NGC2882 (-1.4,-32.9) 1.450
NGC4625 (-3.2,-32.5) 1.573
NGC4800 (-0.9,-34.4) 1.600
NGC4790 (-3.1,-35.4) 1.671
NGC2601 (-1.5,-35.4) 1.803

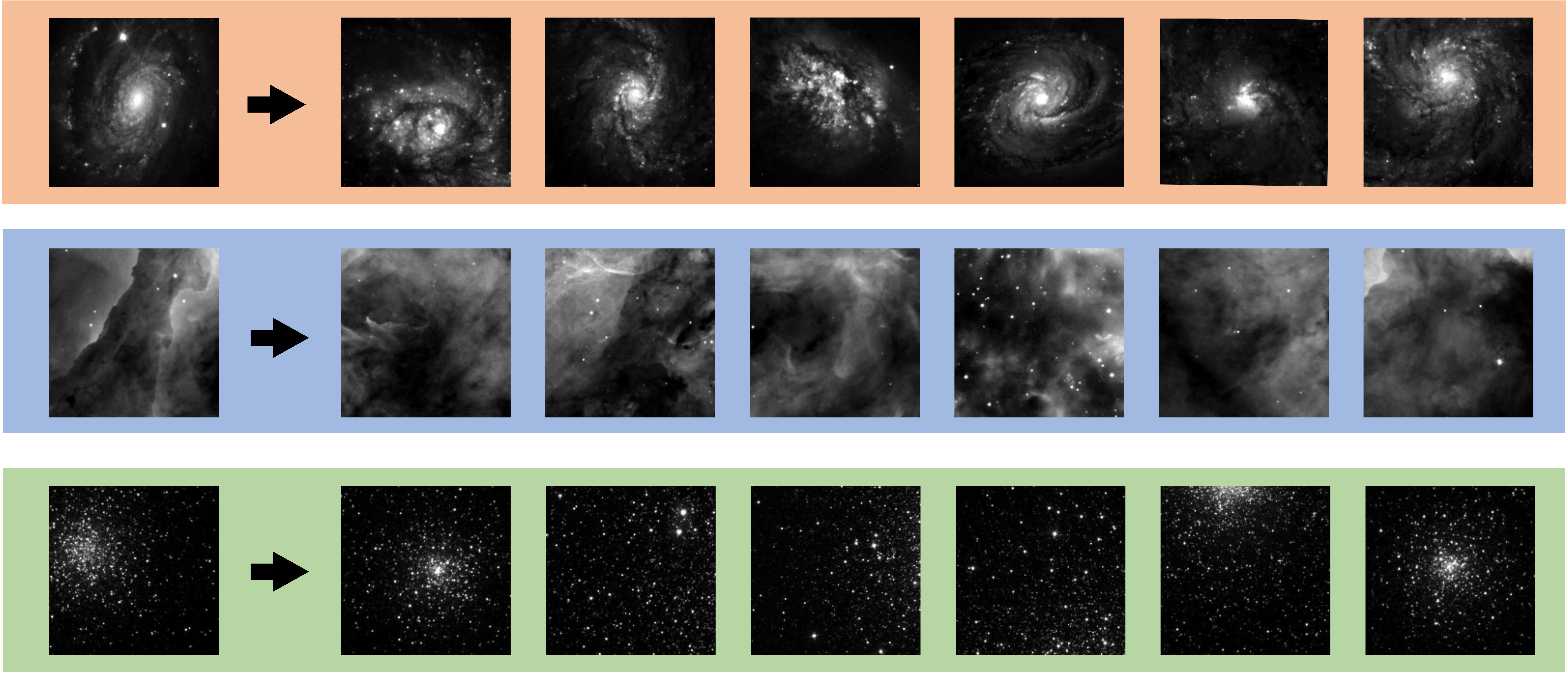
Click to show closest images from any cluster
Cl 9 (33 members)
40
20
0
−20
−40
−40
−30
−20
−10
0
10
20
30
40
x=-1.114 y=-6.377
NGC3433 (-0.8,-5.8) 0.000
NGC7407 (-0.8,-5.4) 0.378
NGC0289 (-0.4,-6.6) 0.908
NGC6323 (0.4,-6.2) 1.288
NGC6557 (-2.1,-5.4) 1.344
NGC5064 (-2.1,-6.3) 1.419
NGC3145 (-2.2,-6.9) 1.804
NGC3054 (-2.0,-4.3) 1.830
NGC0266 (-0.1,-4.0) 1.868
NGC0986 (-2.2,-7.1) 1.935
NGC0634 (-0.1,-7.8) 2.159
NGC0192 (-0.7,-7.9) 2.190
NGC5448 (-2.6,-4.3) 2.348
NGC0634 (-1.3,-8.3) 2.552
NGC5985 (-1.9,-3.4) 2.584
NGC6862 (-3.3,-4.8) 2.634

Click to show closest images from any cluster
Cl 15 (34 members)
x=-8.556 y=-20.164
NGC7541 (-8.5,-20.7) 0.000
NGC5775 (-7.4,-20.7) 1.084
NGC3370 (-9.5,-19.5) 1.516
NGC4634 (-8.0,-19.2) 1.543
NGC5010 (-8.7,-22.7) 2.032
NGC2748 (-7.4,-18.7) 2.285
NGC4634 (-6.5,-19.7) 2.288
NGC4522 (-6.7,-19.2) 2.360
NGC7130 (-10.1,-18.7) 2.499
NGC2748 (-7.5,-18.3) 2.559
NGC3287 (-8.3,-18.0) 2.666
IC1024 (-10.5,-22.7) 2.799
NGC4496A (-10.9,-19.0) 2.862
NGC5714 (-8.2,-23.6) 2.911
IC1127 (-6.0,-19.2) 2.971
IC5283 (-10.7,-22.8) 2.992

Click to show closest images from any cluster
Cl 31 (38 members)
x=3.548 y=16.215
NGC3603 (3.5,16.2) 0.000
NGC3603 (3.2,16.3) 0.342
NGC3603 (3.9,15.7) 0.639
NGC3603 (2.8,16.1) 0.802
NGC3603 (3.3,15.4) 0.887
NGC6357 (4.3,16.8) 0.995
NGC3603 (2.5,15.9) 1.124
NGC2083 (3.7,17.6) 1.394
NGC3603 (4.6,15.3) 1.411
NGC2078 (2.8,15.0) 1.455
NGC3603 (3.7,14.7) 1.523
NGC2078 (3.3,14.7) 1.542
NGC2060 (2.6,17.5) 1.593
NGC3603 (4.1,14.6) 1.723
NGC2078 (2.9,14.6) 1.725
NGC3603 (1.9,16.9) 1.781

Click to show closest images from any cluster
Cl 46 (8 members)
40
20
0
−20
−40
−40
−30
−20
−10
0
10
20
30
40
x=37.196 y=-9.504
NGC4452 (37.5,-9.6) 0.000
NGC4452 (37.6,-9.7) 0.179
IC0335 (36.6,-10.5) 1.290
NGC0705 (37.0,-8.2) 1.450
IC0335 (36.9,-8.2) 1.450
IC0335 (37.1,-7.9) 1.734
NGC4111 (37.3,-7.2) 2.352
NGC1380A (35.6,-7.8) 2.582
NGC1381 (35.3,-8.0) 2.661
IC3773 (35.0,-8.1) 2.911
NGC1381 (36.4,-6.2) 3.505
NGC1381 (35.1,-6.7) 3.756
NGC4350 (34.9,-6.6) 3.947
NGC4623 (35.4,-6.1) 4.010
NGC1380A (36.2,-5.1) 4.660
NGC4350 (35.9,-5.1) 4.777

The Self-Supervised Learning approach works!



Walker Stevens for the SpaceML team

Conclusions:

Conclusions:

Machine Learning is a popular and powerful tool for astronomy, but not a replacement for statistics

Conclusions:

Machine Learning is a popular and powerful tool for astronomy, but not a replacement for statistics

Astronomers need to join with ML experts to develop tools appropriate for astronomical problems

Conclusions:

Machine Learning is a popular and powerful tool for astronomy, but not a replacement for statistics

Astronomers need to join with ML experts to develop tools appropriate for astronomical problems

We have gone from morphological classification to true physical analyses with deep learning on galaxy images and we are just getting started

Conclusions:

Machine Learning is a popular and powerful tool for astronomy, but not a replacement for statistics

Astronomers need to join with ML experts to develop tools appropriate for astronomical problems

We have gone from morphological classification to true physical analyses with deep learning on galaxy images and we are just getting started

Hypothesis generation in the big data era is still an unsolved problem in astronomy

Conclusions:

Machine Learning is a popular and powerful tool for astronomy, but not a replacement for statistics

Astronomers need to join with ML experts to develop tools appropriate for astronomical problems

We have gone from morphological classification to true physical analyses with deep learning on galaxy images and we are just getting started

Hypothesis generation in the big data era is still an unsolved problem in astronomy

Cross-industry and cross-science deep learning approaches will create new tools for astroinformatics in the 2020s

Conclusions:

Machine Learning is a popular and powerful tool for astronomy, but not a replacement for statistics

Astronomers need to join with ML experts to develop tools appropriate for astronomical problems

We have gone from morphological classification to true physical analyses with deep learning on galaxy images and we are just getting started

Hypothesis generation in the big data era is still an unsolved problem in astronomy

Cross-industry and cross-science deep learning approaches will create new tools for astroinformatics in the 2020s