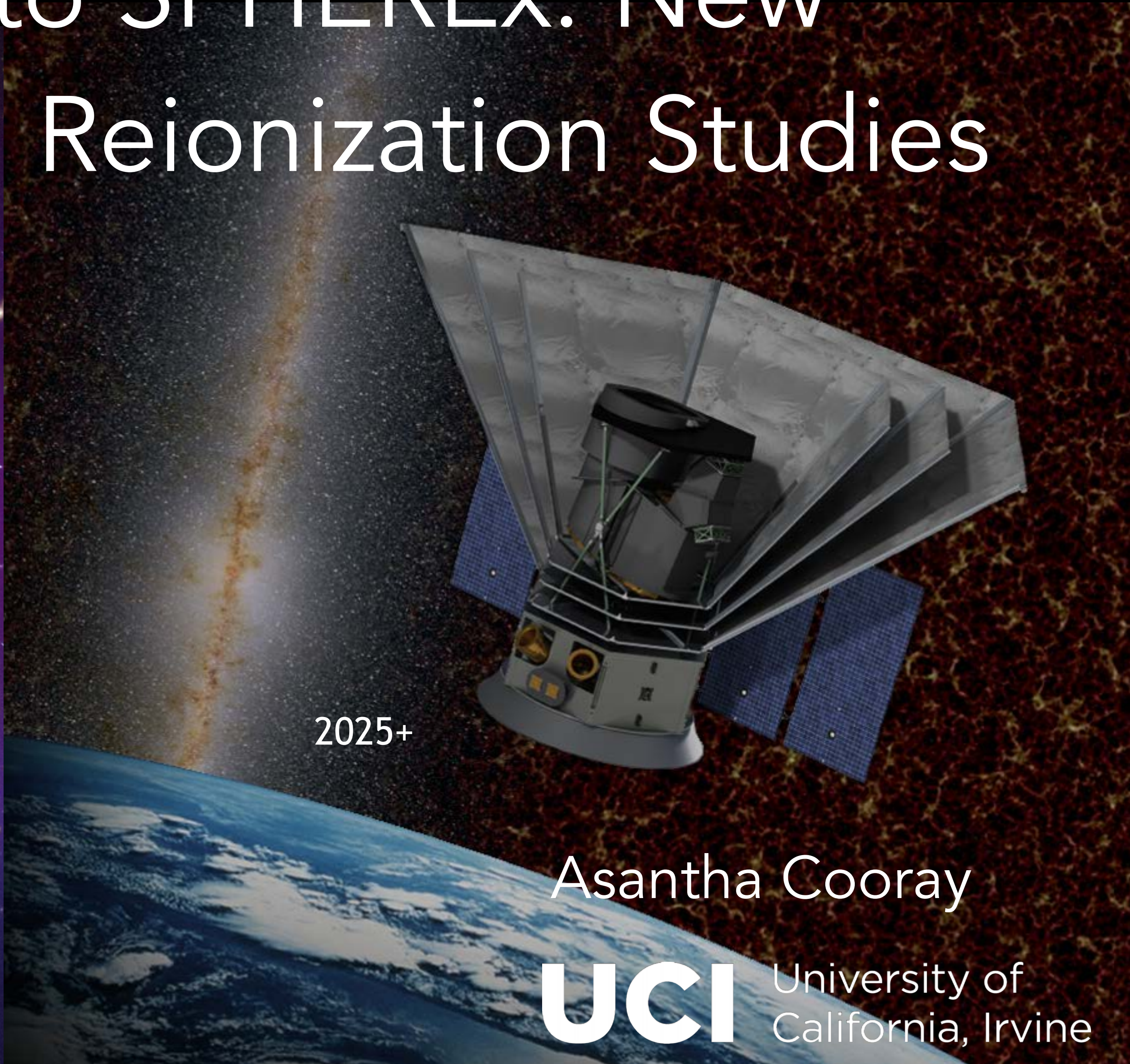
From Euclid to SPHEREx: New Opportunities in Reionization Studies



Asantha Cooray

UCI University of California, Irvine

Outline

Three separate sub-projects led by students.

1. Identifying highest-z's lensed quasars and galaxies with Euclid
2. Quasar identification at $z > 6$ with SPHEREx and Euclid
3. Intensity mapping opportunities to study reionization and structure formation with SPHEREx and Euclid (and other datasets in 2027-2030)

Euclid Strong Lens Detection with CNNs

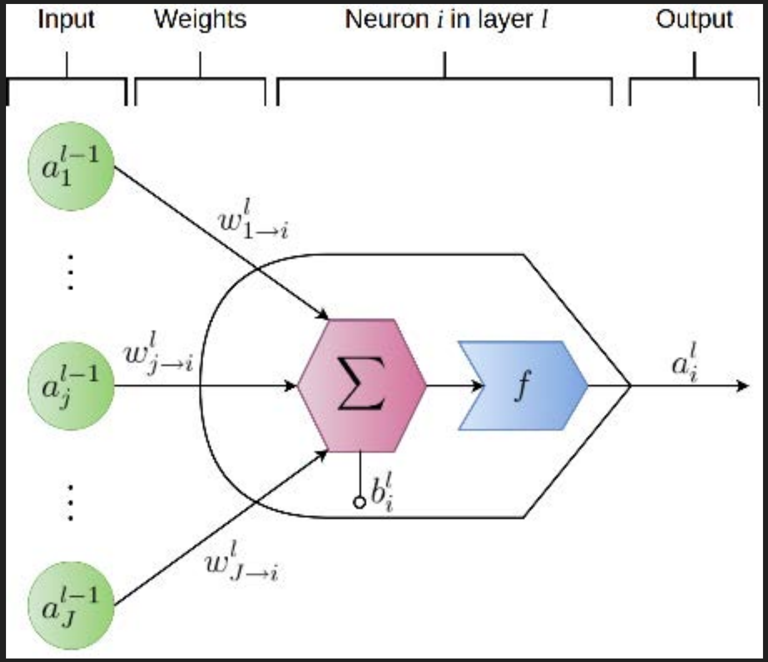
- **Goal: identify and separate (VIS & NISP) z > 6 lensed quasars and LGBs.**
- Objective: Automate the detection of strong lensing events from Euclid's survey with deep learning, aided by human classification or downsample for a citizen science project.

Methods

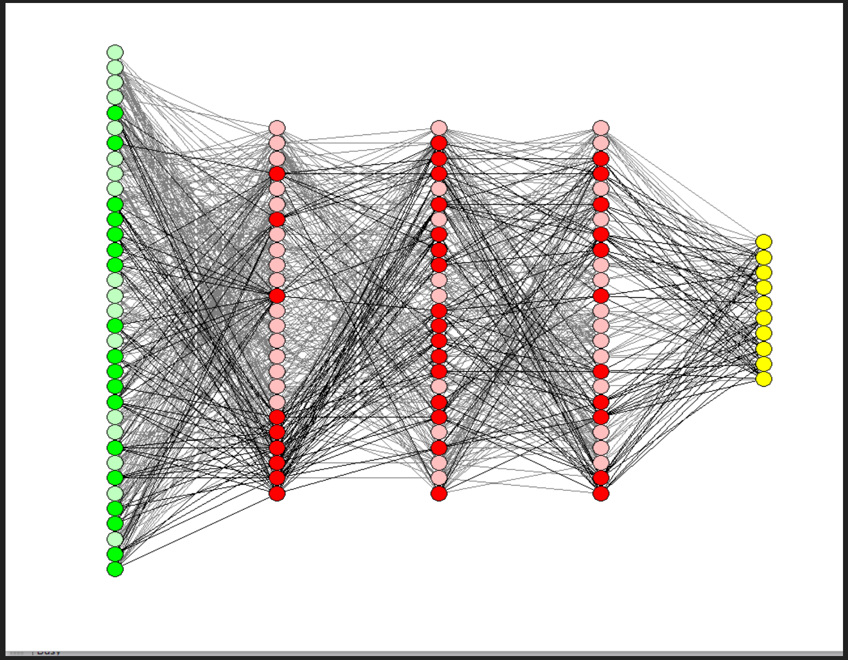
- Sample galaxies from CANDELS survey to generate mock images of strong lensing events observed by Euclid
- Use the images to train a Convolutional Neural Network (CNN)
- Work in progress: Accurate Euclid data simulations to find strong lenses and in the future post-launch apply to survey data.

Thomas Li
(aided by Milad Pourrahmani, Nima Chartab)

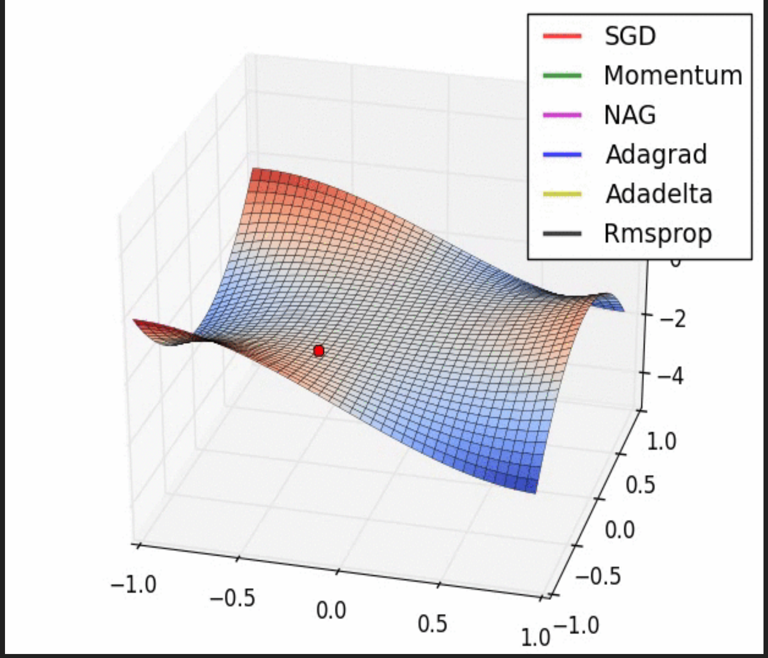
What are artificial neural nets and how do they learn?



I. An Artificial Neuron

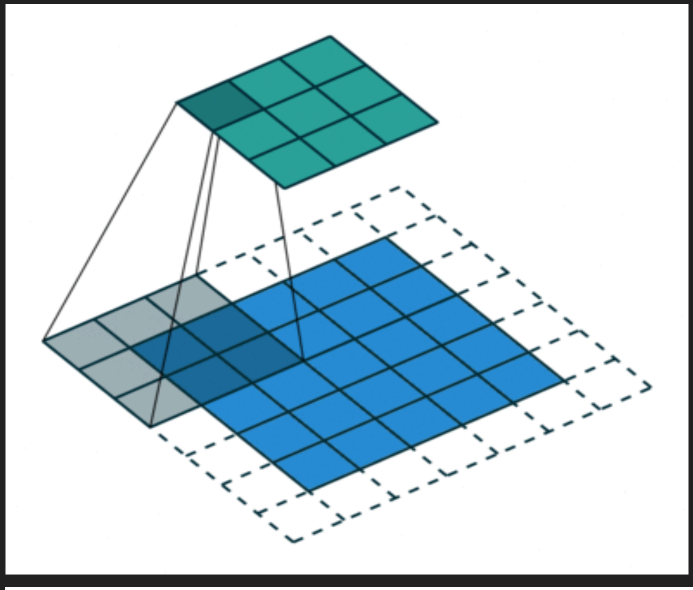


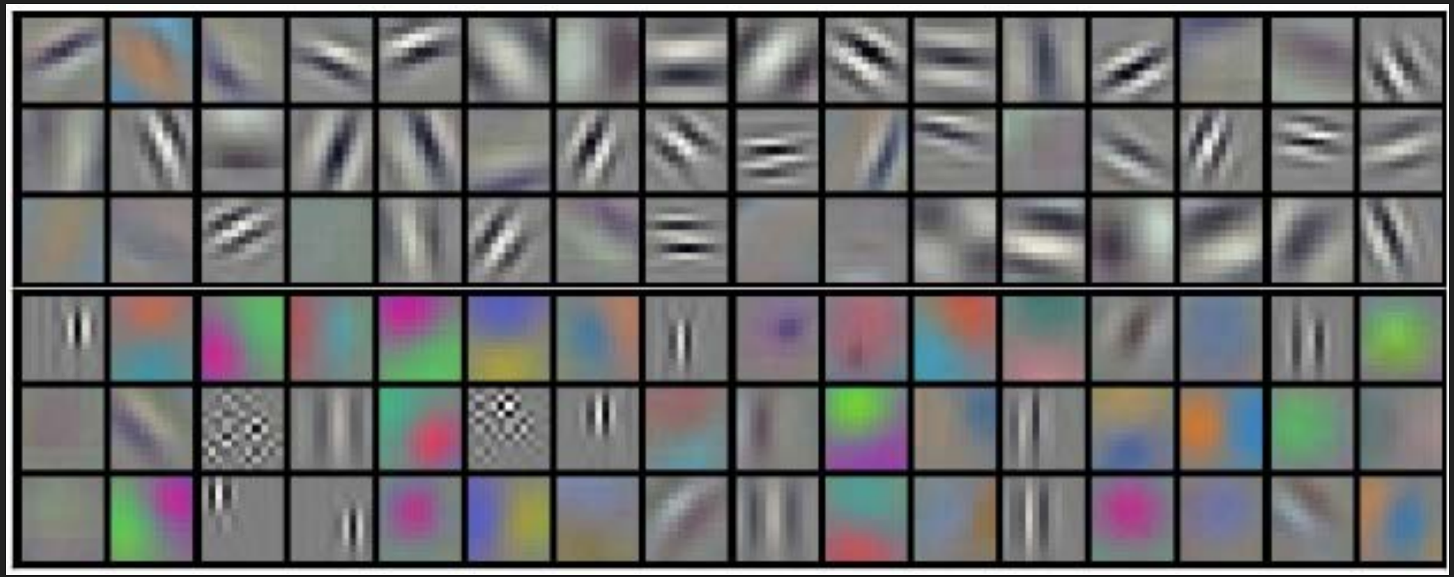
II. A Neural Network

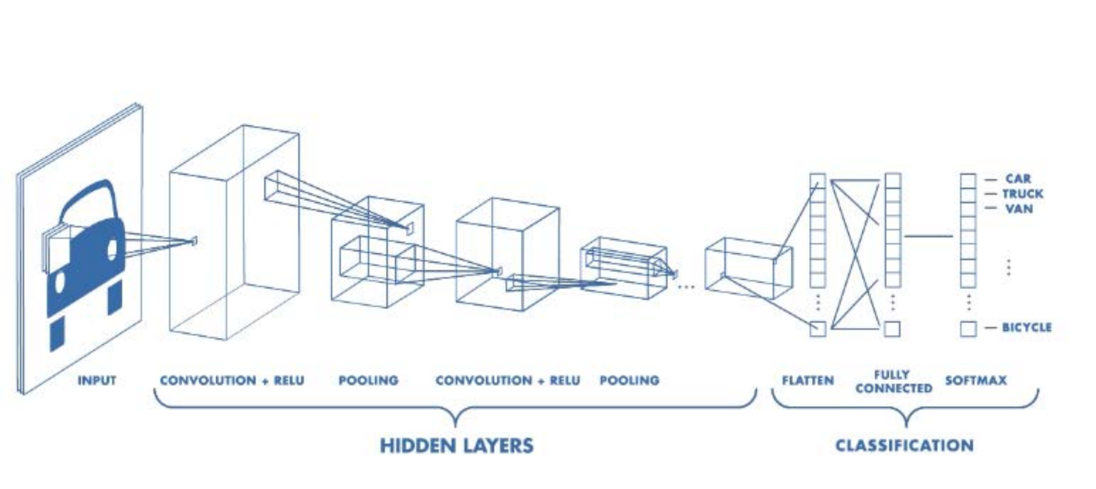


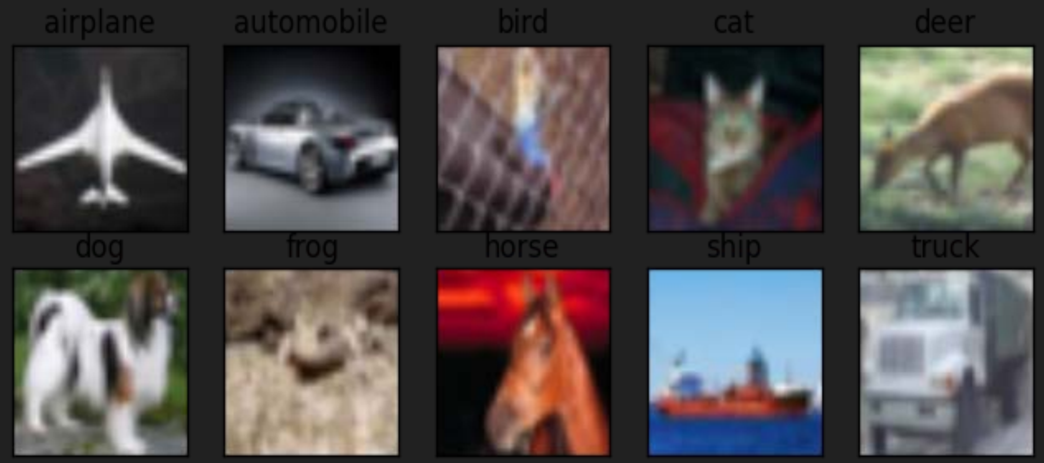
III. Gradient Descent

Deep ConvNets









Identifying lenses in HST/ACS i-band in COSMOS 2 deg2.

- ➔ Phase 0: Pre-train on 2 classes of CIFAR
- ➔ Phase 1: Train on lenses and all non-lenses
  - ◆ Balanced batch selection: n_class x 64
- ➔ Phase 2: Train on lenses and high rank non-lenses
- ➔ <u>We identified ~90 lenses half of which were new</u>

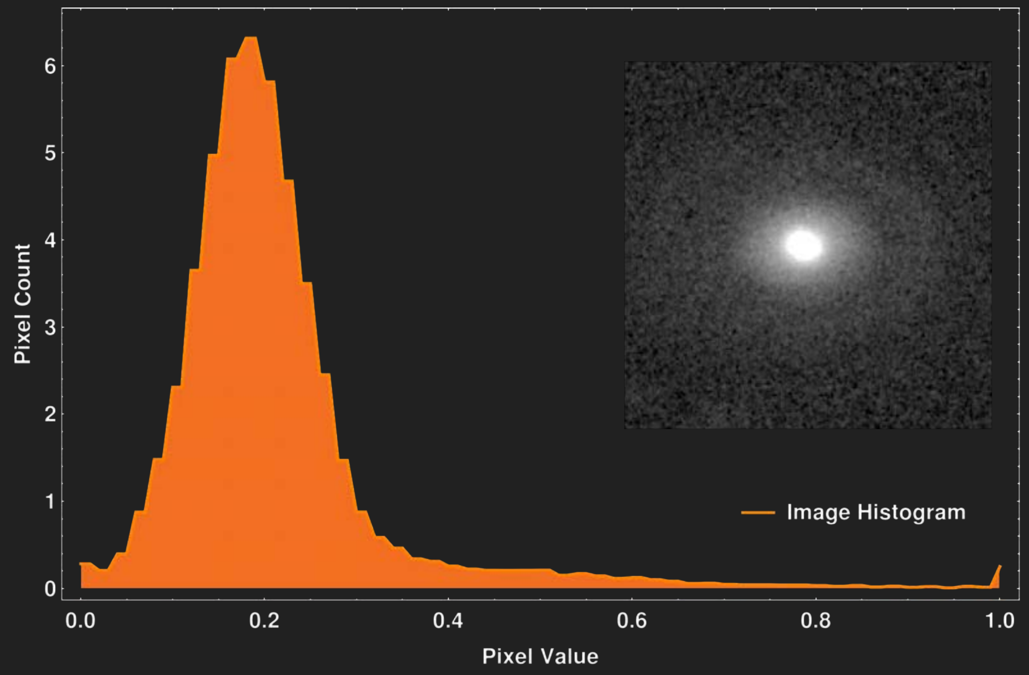
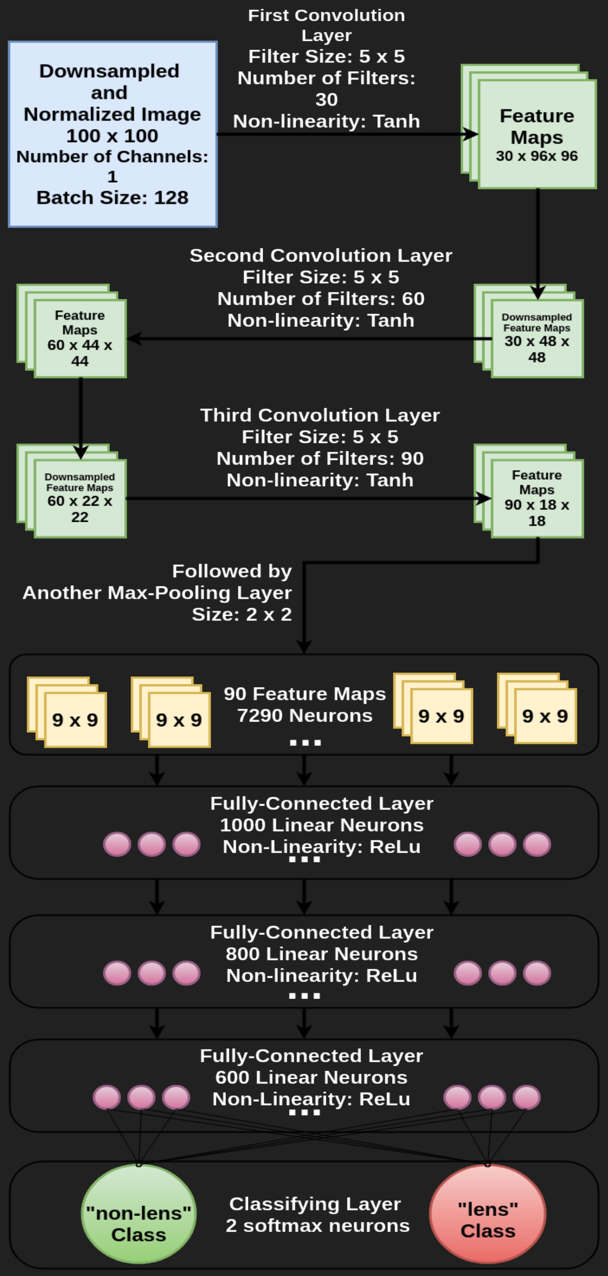


Table 1
Tabulated Architecture of the LENSFLOW ConvNet

Layer	Type	Data Dimensionality
	input	$1 \times 100 \times 100$
1	convolution	$30 \times 96 \times 96$
2	tanh	$30 \times 96 \times 96$
3	pooling	$30 \times 48 \times 48$
4	convolution	$60 \times 44 \times 44$
5	tanh	$60 \times 44 \times 44$
6	pooling	$60 \times 22 \times 22$
7	convolution	$90 \times 18 \times 18$
8	tanh	$90 \times 18 \times 18$
9	pooling	$90 \times 9 \times 9$
10	flatten	7290
11	linear	1000
12	ReLU	1000
13	dropout	1000
14	linear	800
15	ReLU	800
16	dropout	800
17	linear	600
18	ReLU	600
19	dropout	600
20	linear	2
21	softmax	2
	output	2

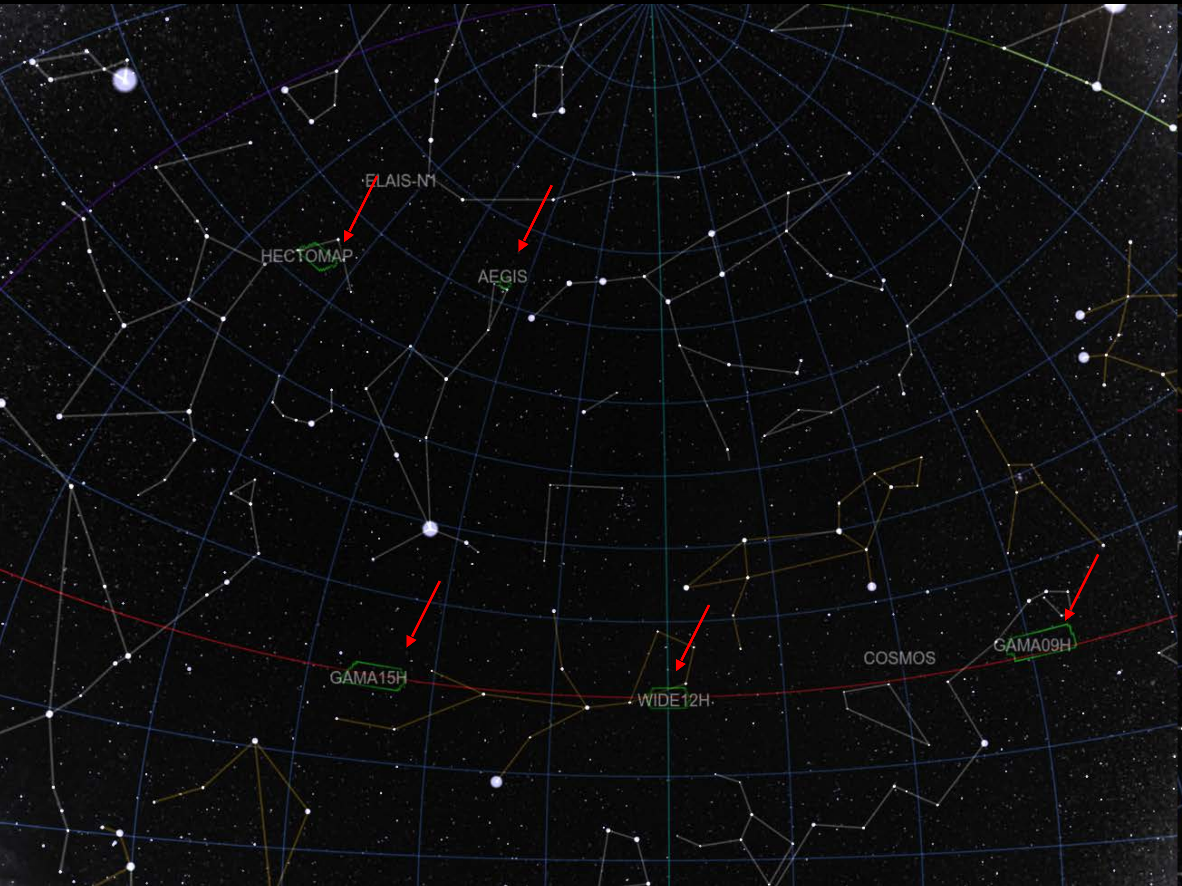


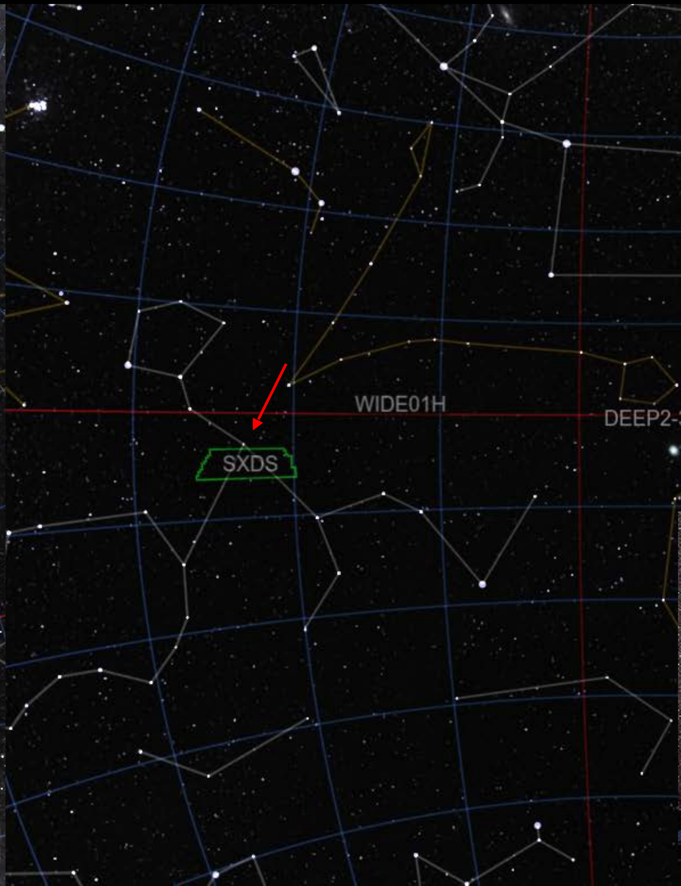
Pourrahmani et al. 2019

Subaru HSC Survey lens identification

Pourrahmani et al. 2020

~1000 deg^2 wide in *grizy* bands

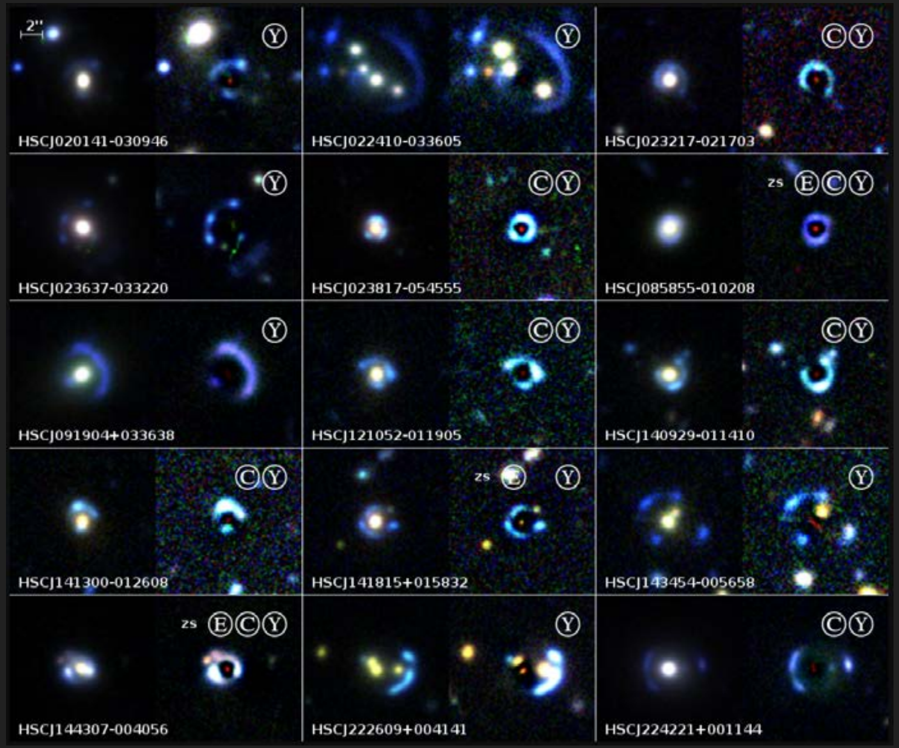




Problem! Not enough lenses to create a training set.

SuGOHI Project found only 50 grade A and B lenses in HSC-PDR1

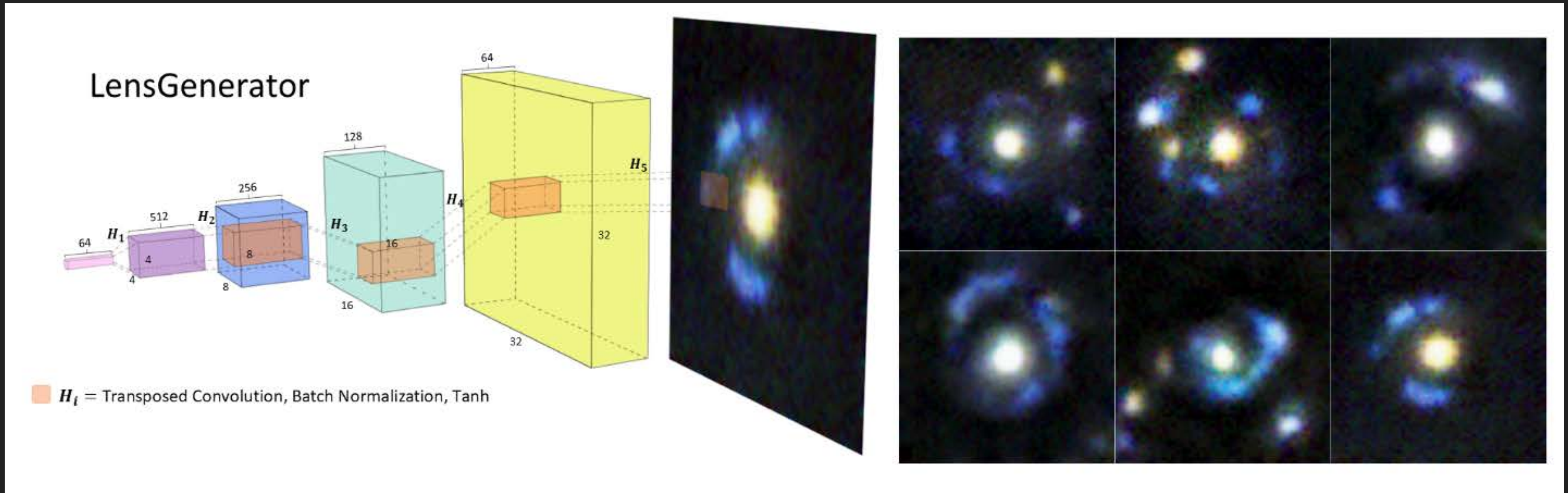
Option 1: Simulate
Option 2: GAN Augment



Sonnenfeld et al. 2017

Augmentation with Generative Adversarial Network (GANs)

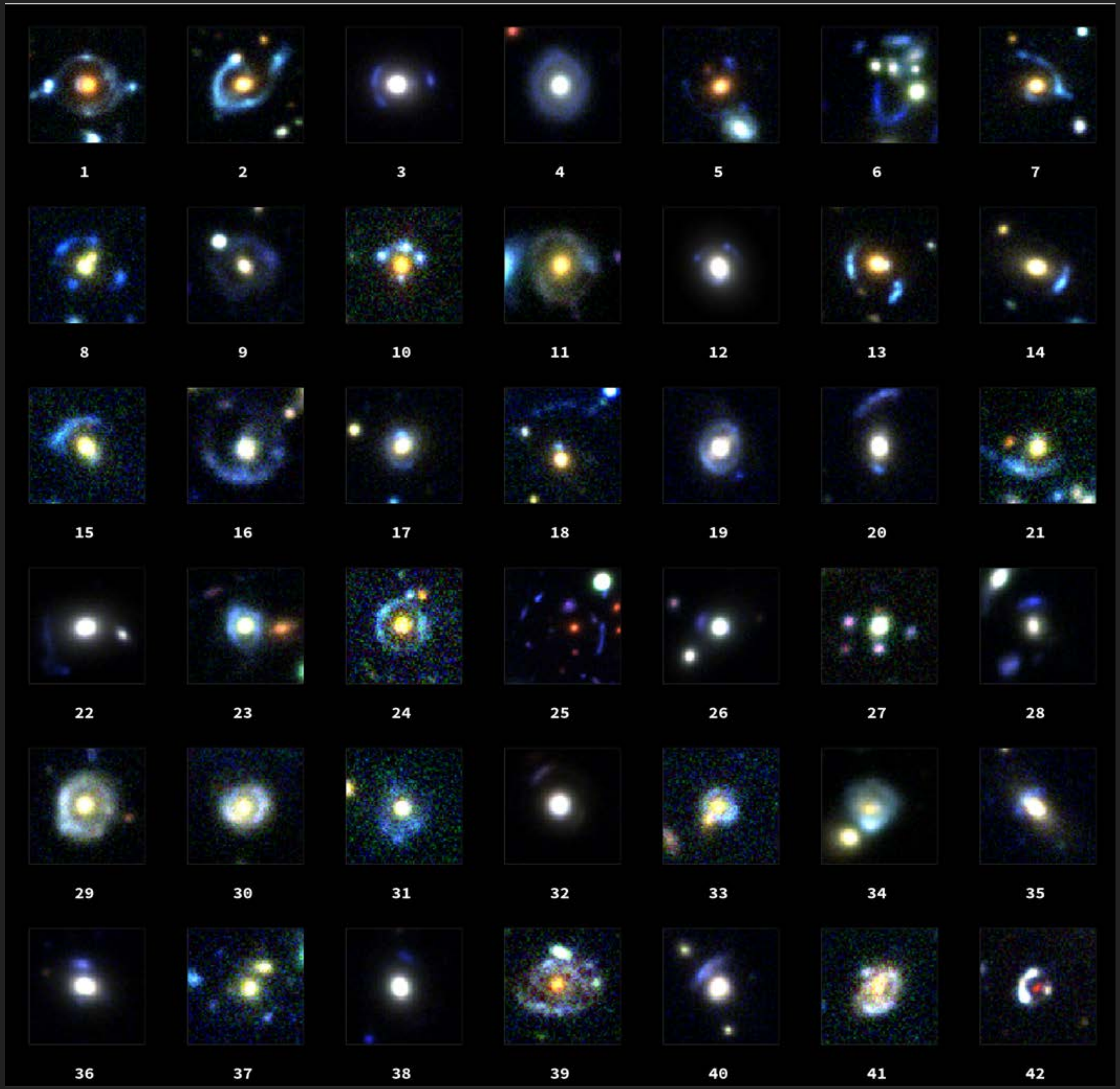
- Generator vs Discriminator
- Pre-train on negative images
- Train on real lenses
- Mode Collapse Problem
    - Retrain and resample 200 images for 10 cycles → 2k generated lenses total



Pourrahmani et al. 2020

HSC Results

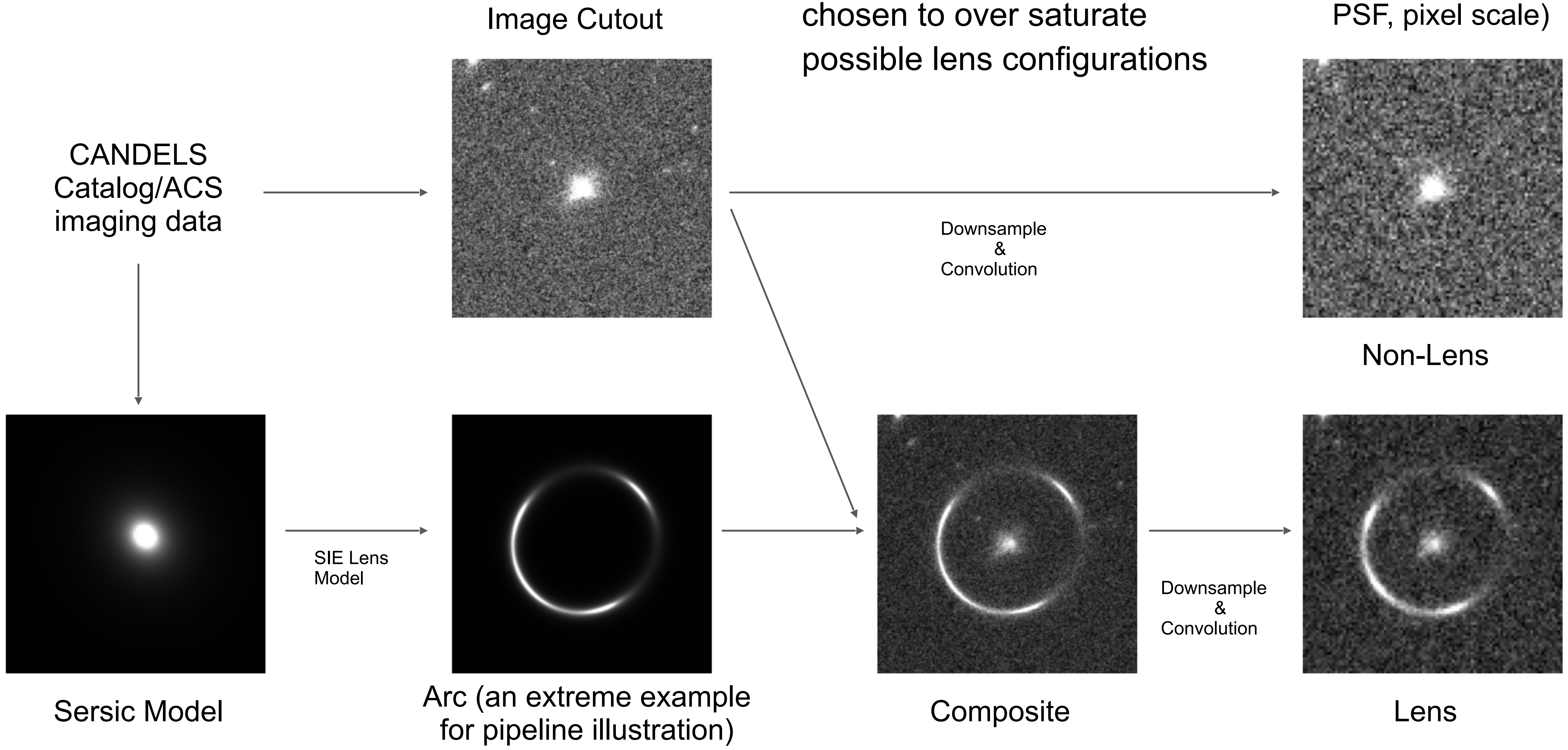
Lens	Right Ascension (deg)	Declination (deg)	Einstein Radius (arcsec)	Grade A/B/C
1	+216.2043	−00.8894	3.25	*A*
2	+215.2654	+00.3720	2.20	*A*
3	+340.5898	+00.1957	2.44	*A*
4	+219.4564	+00.4944	2.45	*A*
5	+335.1725	+00.8201	2.99	*A*
6	+333.5790	+01.1769	2.54	*A*
7	+217.8079	−00.1037	2.26	*A*
8	+218.7266	−00.9496	2.71	*A*
9	+333.1764	+00.1892	2.38	*A*
10	+132.6942	+00.6515	1.59	*A*
11	+035.3941	−04.4112	2.80	*A*
12	+181.9302	−01.0653	2.21	*A*
13	+216.9515	+00.1663	2.73	*A*
14	+215.0377	−00.2429	2.88	*A*
15	+218.8296	−01.1101	2.68	*A*
16	+135.7236	+00.8684	3.31	*B*
17	+178.9889	−00.2107	1.75	*B*
18	+215.0287	+00.3112	4.09	*B*
19	+215.7069	−00.5532	1.95	*B*
20	+214.2848	+00.9822	2.35	*B*
21	+036.0038	−03.7738	2.65	*B*
22	+178.8366	+00.8846	3.94	*B*
23	+333.8429	+01.0912	1.27	*B*
24	+338.1610	−00.4261	2.29	*B*
25	+136.0186	+01.4212	7.49	*B*
26	+218.5273	−00.4848	1.83	*B*
27	+132.0146	+02.0579	3.13	*B*
28	+334.5879	−00.0316	2.41	*C*
29	+220.1407	−00.6051	2.30	*C*
30	+132.0751	+02.0876	1.79	*C*
31	+035.2166	−04.5833	2.65	*C*
32	+215.1524	−00.1260	3.88	*C*
33	+136.6667	+01.0630	1.25	*C*
34	+131.9165	+01.9710	2.06	*C*
35	+032.1952	−03.4577	1.45	*C*
36	+337.4952	+00.1038	2.18	*C*
37	+135.8591	−00.1686	3.18	*C*
38	+333.6634	+01.2666	2.89	*C*
39	+178.7919	−01.3283	2.52	*C*
40	+216.4902	+00.9385	2.19	*C*
41	+218.2470	−00.5127	1.86	*C*
42	+337.1984	+01.0535	1.35	*C*



Pourrahmani et al. 2020

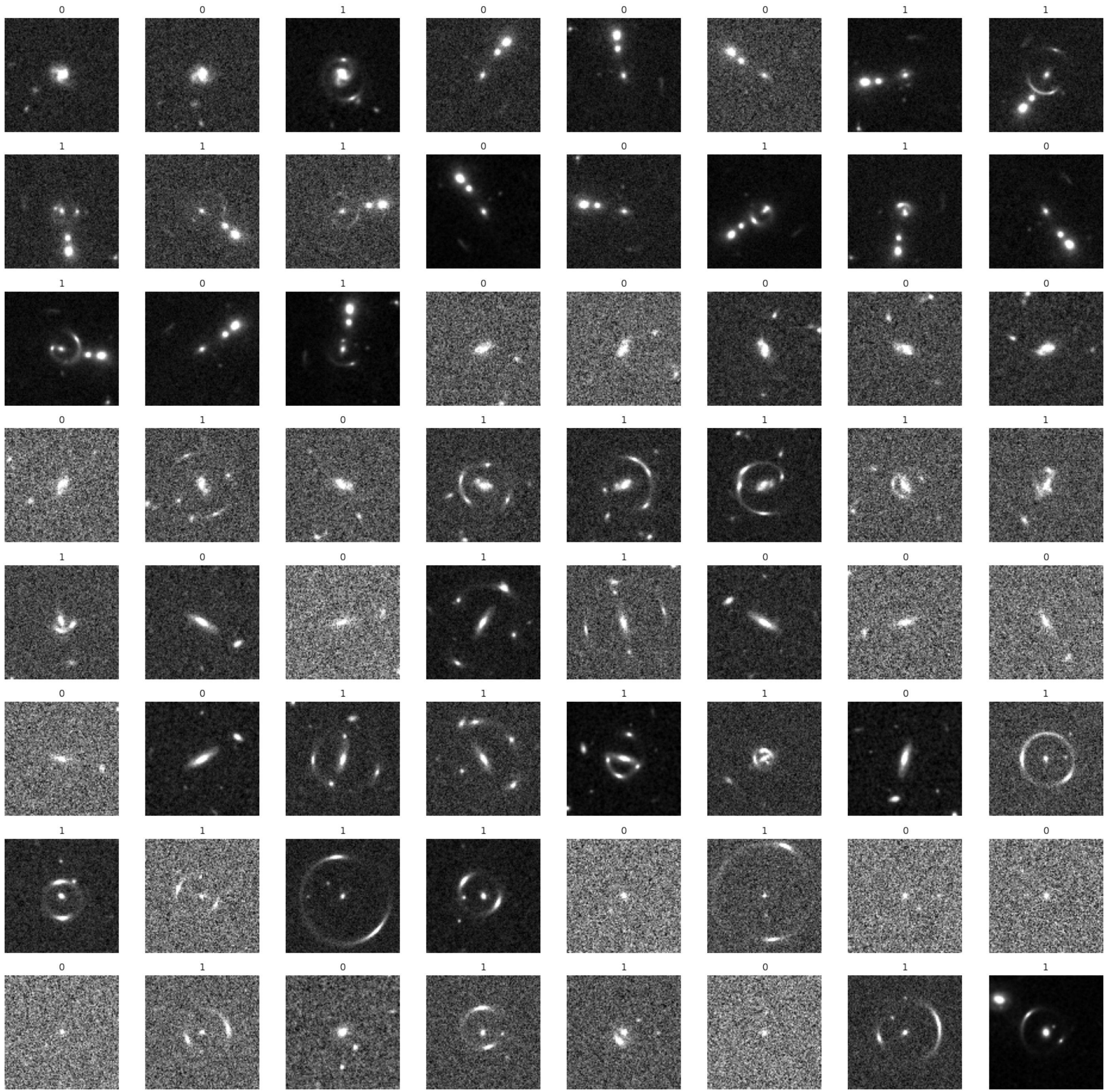
Euclid high-z
Training Data Generation

~60,000 images generated (half with lensing event & half without), parameters are chosen to over saturate possible lens configurations



Examples
(PSF and noise matched to Euclid VIS all-sky)

0 - non-lens
1 - lens



CNN Architecture

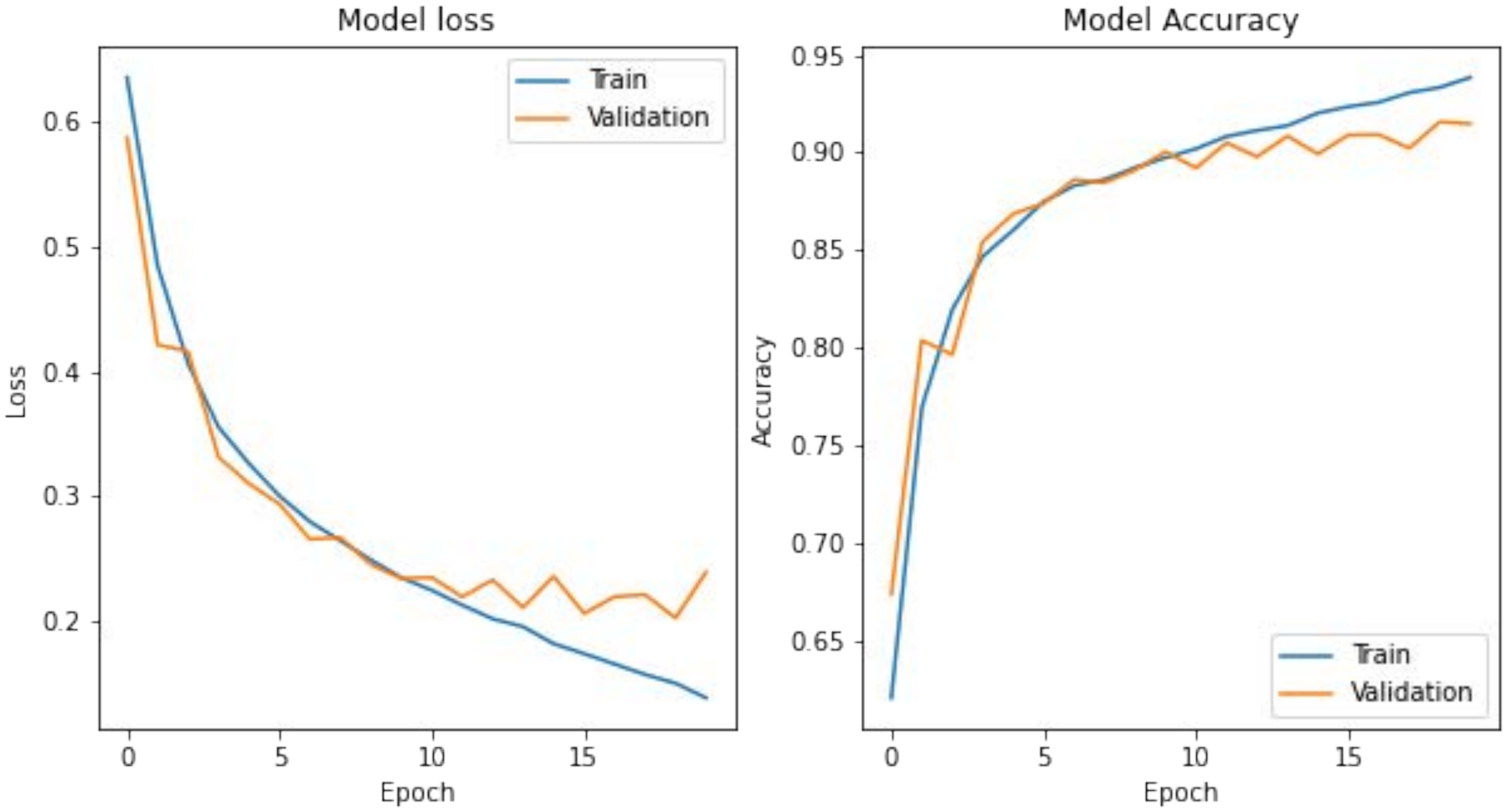
#	Layers	Output Size
1	Input	96 x 96 x 1
2	Convolution (7 x 7)	96 x 96 x 64
3	Convolution (1 x 1)	96 x 96 x 32
4	Convolution (7 x 7)	96 x 96 x 64
	Max Pooling (3 x 3)	32 x 32 x 64
5	Convolution (5 x 5)	32 x 32 x 128
6	Convolution (1 x 1)	32 x 32 x 64
7	Convolution (5 x 5)	32 x 32 x 128
	Max Pooling (2 x 2)	16 x 16 x 128
8	Convolution (3 x 3)	16 x 16 x 256
9	Convolution (1 x 1)	16 x 16 x 128
10	Convolution (3 x 3)	16 x 16 x 256
	Max Pooling (2 x 2)	8 x 8 x 256

#	Layers	Output Size
11	Convolution (3 x 3)	8 x 8 x 256
12	Convolution (1 x 1)	8 x 8 x 128
13	Convolution (3 x 3)	8 x 8 x 256
	Max Pooling (2 x 2)	4 x 4 x 256
14	Dense	256
	Dropout (0.5)	
15	Dense	256
	Dropout (0.5	
16	Dense	1

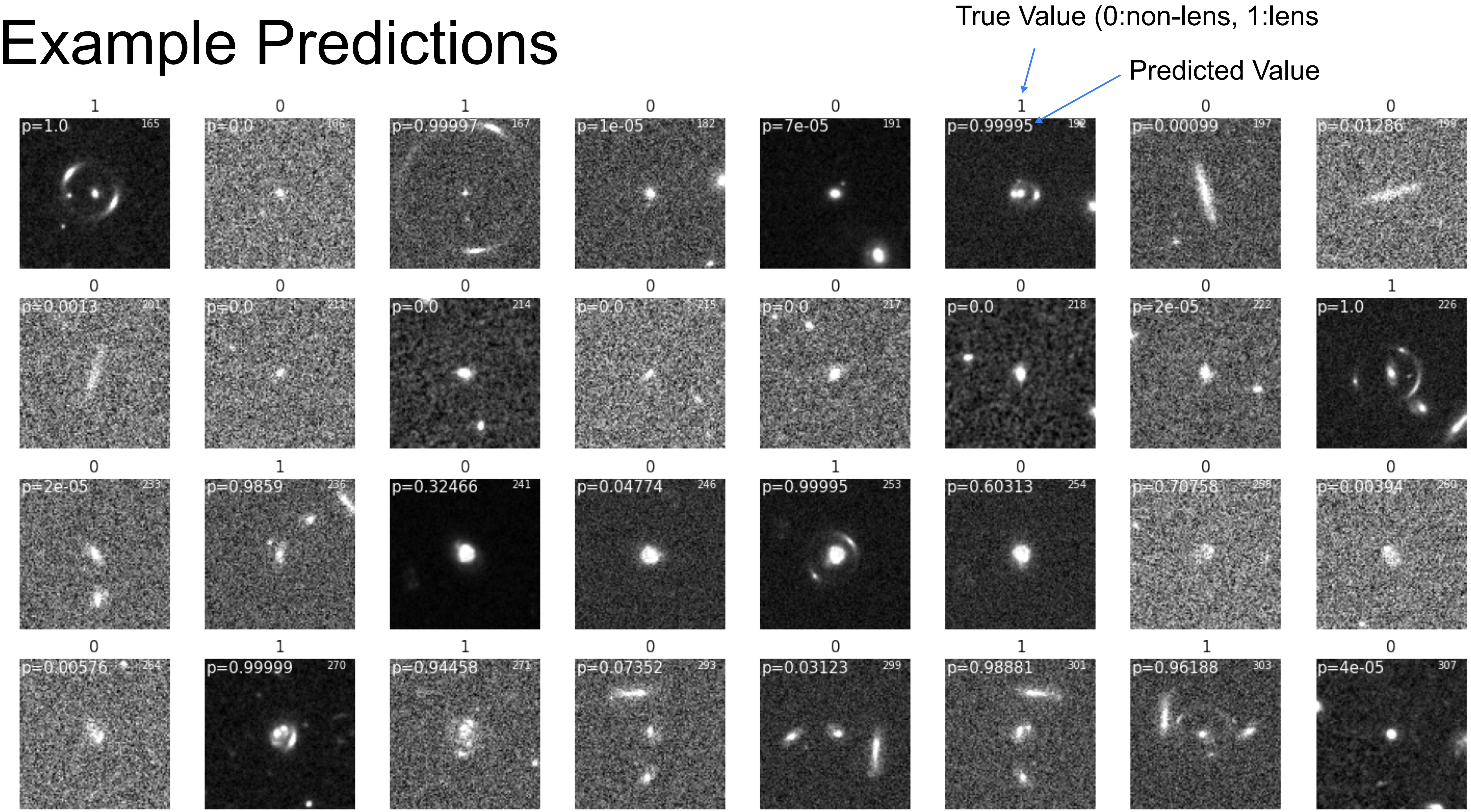
Trainable Parameters:

3,180,065

Training Curves

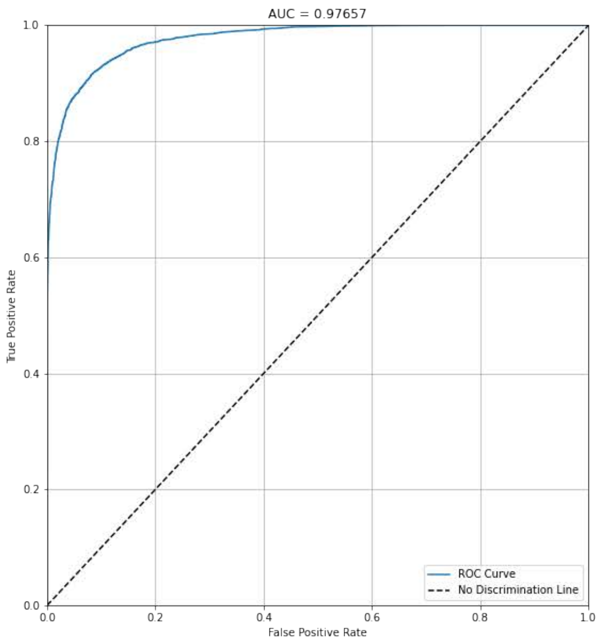


Example Predictions



CNN Prediction

The ROC curve shows the True Positive Rate (TPR) and False Positive Rate (FPR) of the predictions at various thresholds



Predicted Performance

(assume 1 strong lens in 10,000 galaxies)

Optimum Threshold (balanced in maximizing TPR and minimizing FPR):

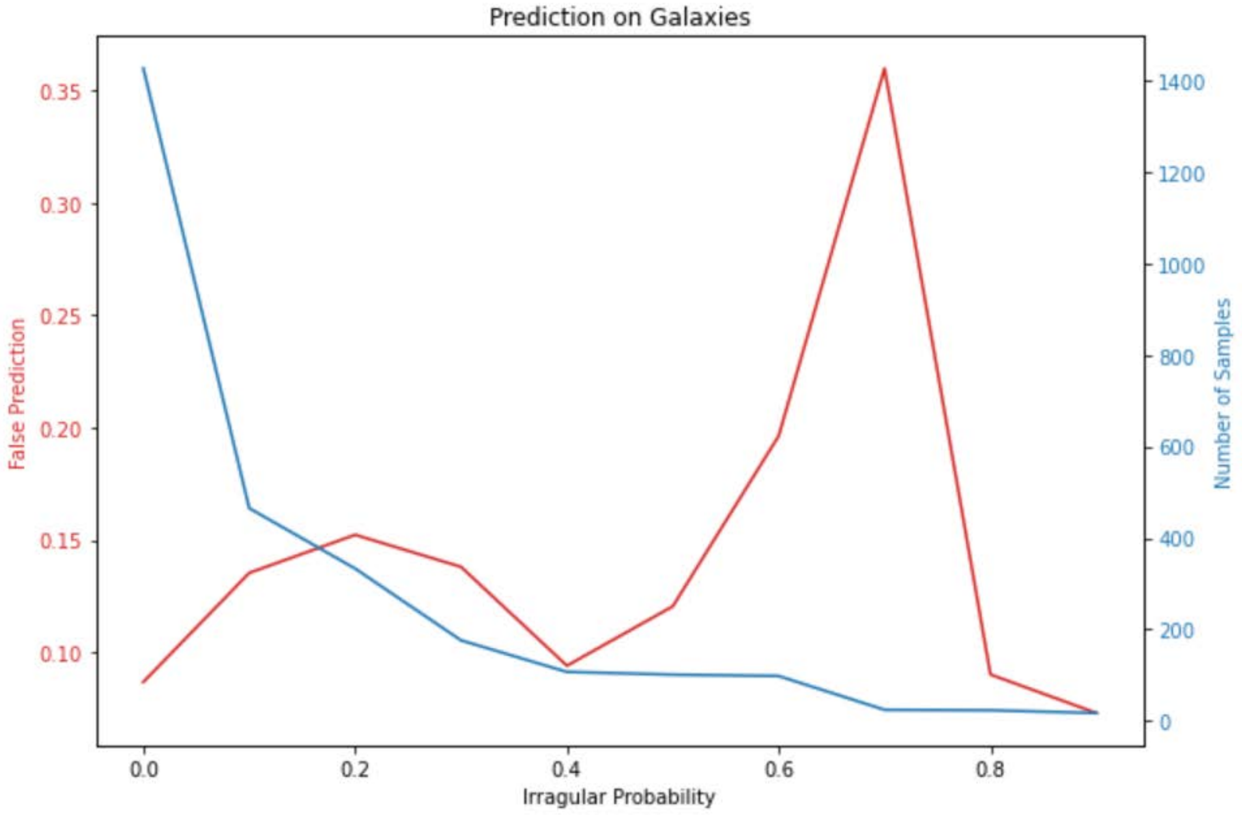
- Completeness: 92% , Purity: 0.11% (1 true positive in 900 positive detection)

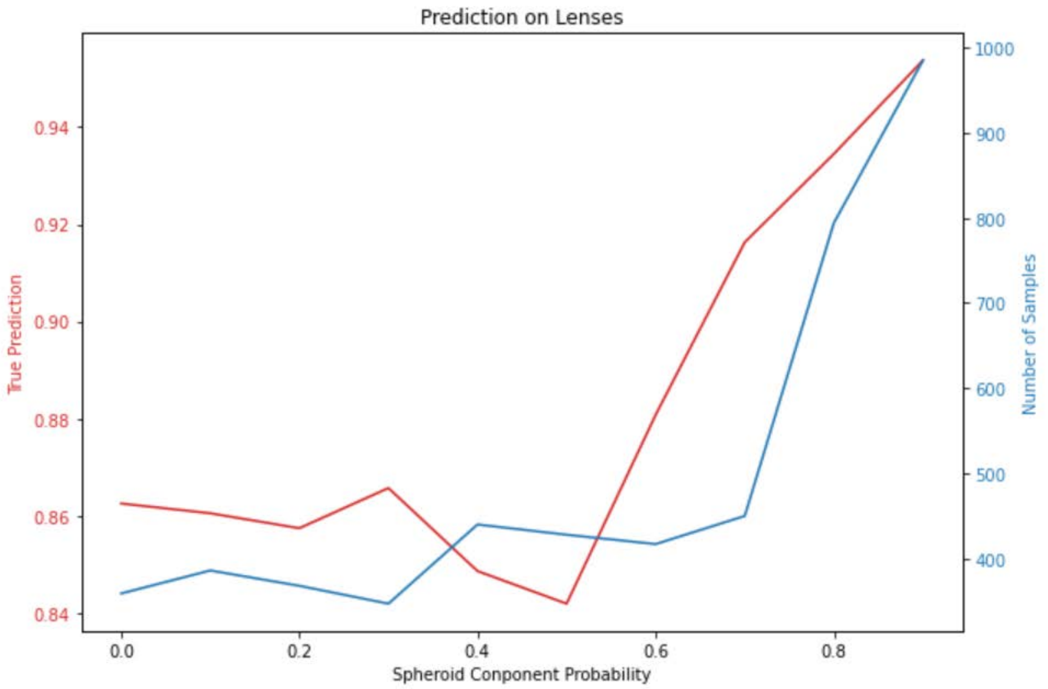
Purity Favored Threshold (sacrifice TPR to get even lower FPR):

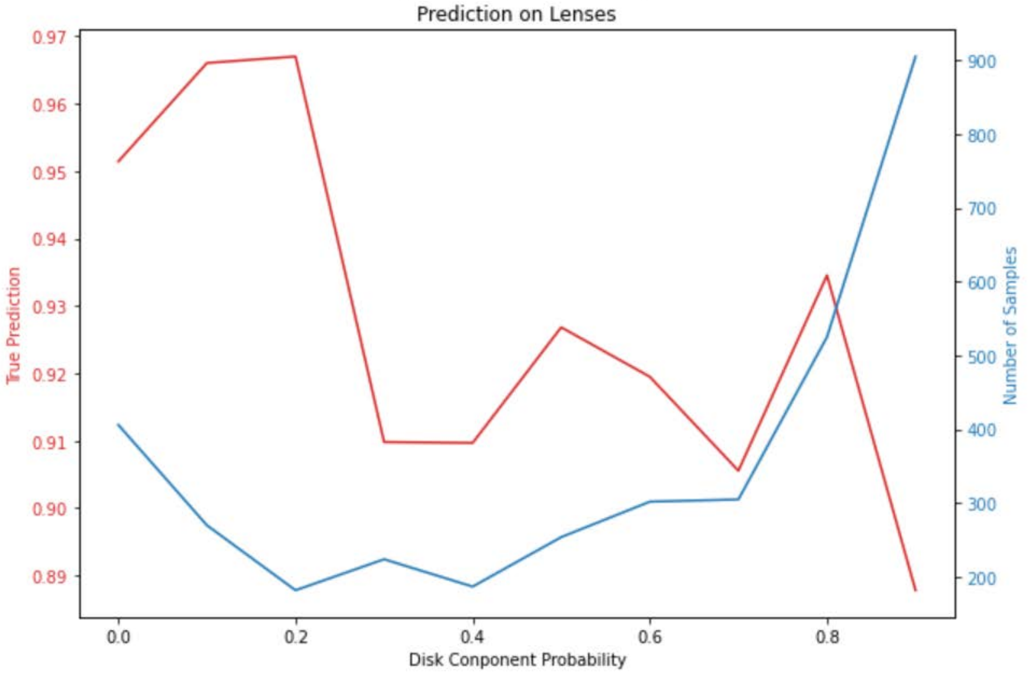
- Completeness: 66% , Purity: 1.6% (1 true positive in 65 positive detection)

Machine learning alone is not going to be enough; will need human supervision or some activity as a citizen type project.

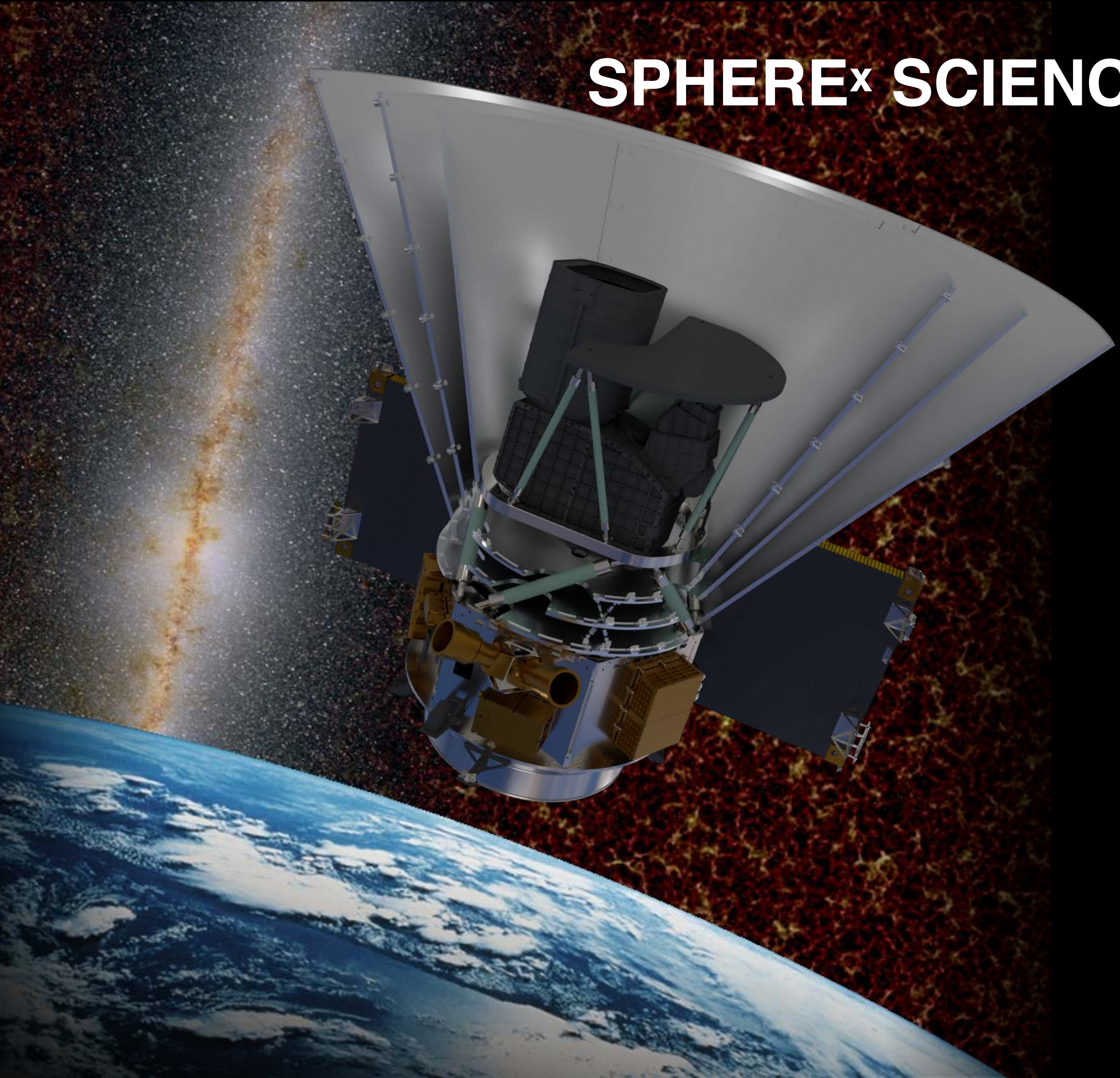
Effect of Morphological Features







SPHEREx SCIENCE OPPORTUNITIES



Designed to Explore

- Origin of the Universe
- Origin and History of Galaxies
- Origin of Water in Planetary Systems

First All-Sky Near-IR Spectral Survey

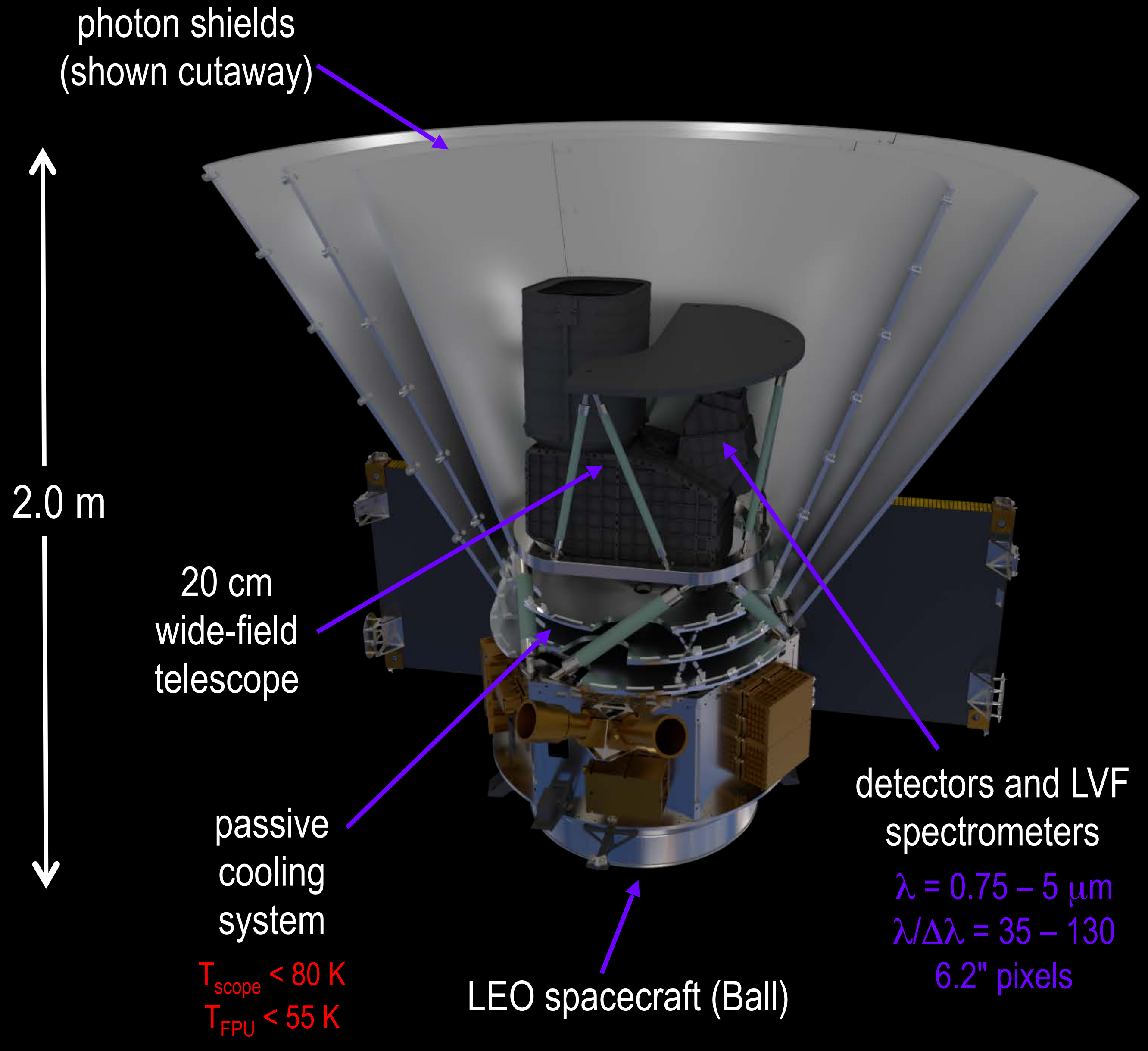
- A Rich Legacy Archive for Astronomy with 100s of Millions of Stars and Galaxies
- Many opportunities outside of the planned core science program

Testing and Integration Imminent!

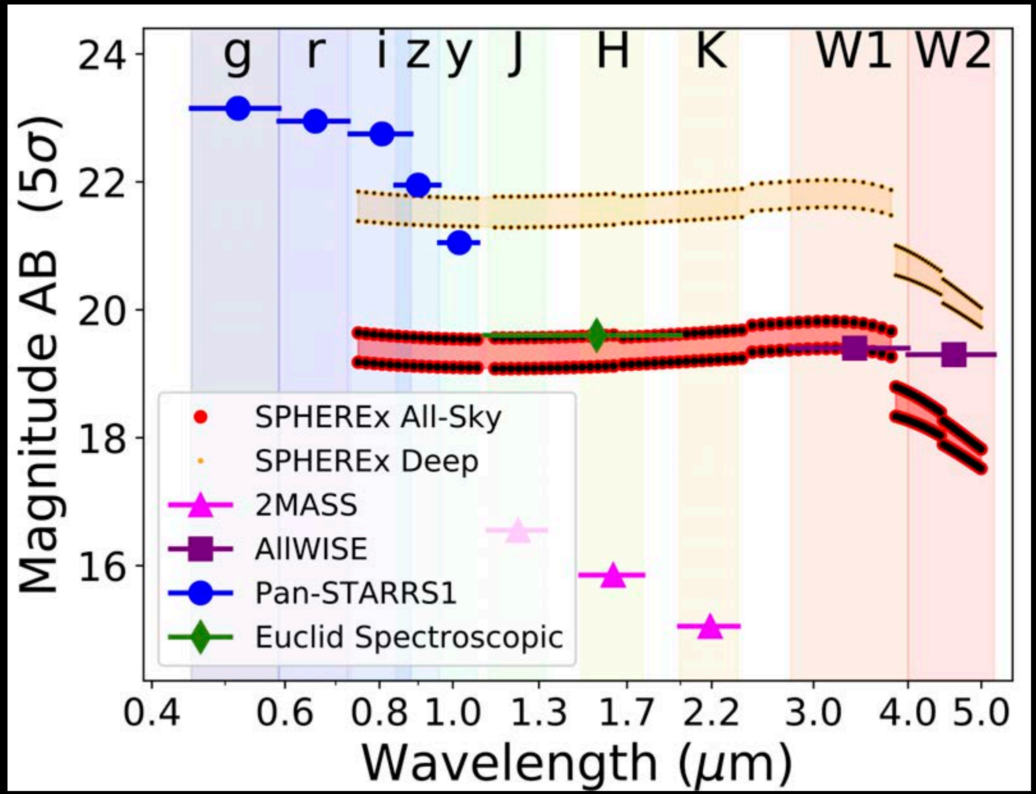
- Design Stable and Mature
- Critical Hardware Arriving
- Healthy Scientific Performance Margins

Anticipated launch readiness review Feb 2025

SPHEREx IN A NUTSHELL



Sensitivity Per Spectral Channel

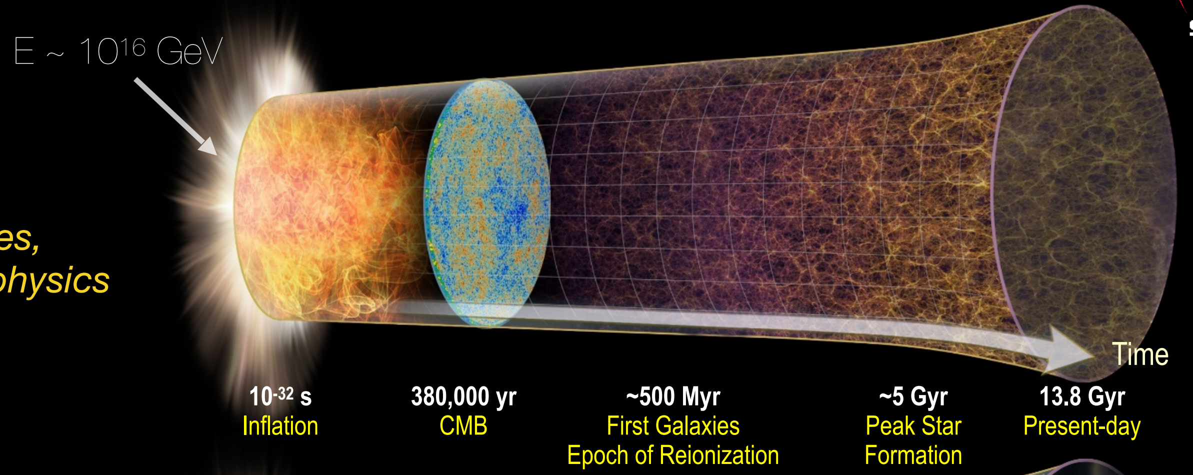


SPHEREx has 102 channels (Δλ ≡ λ/R)

SPHEREx CORE SCIENCE PROGRAM

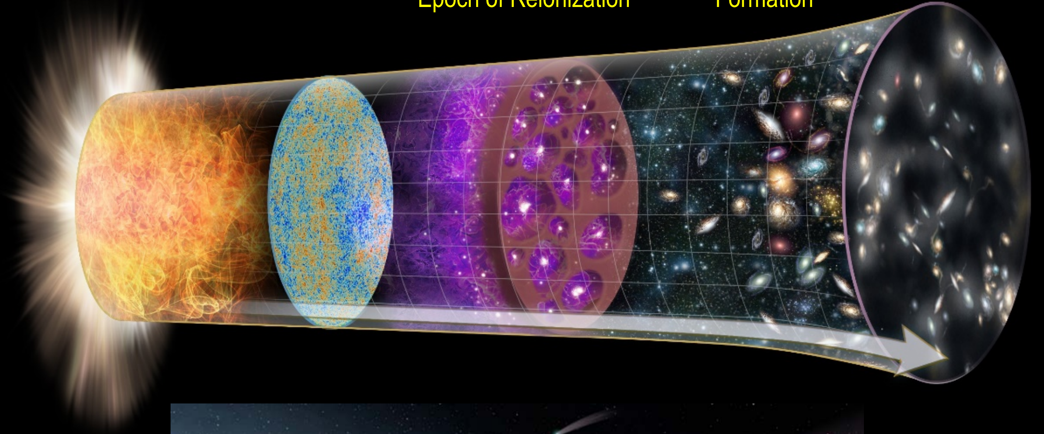
How did the Universe Begin?

SPHEREx observes the 3D distribution of galaxies, measuring ‘Non-Gaussianity’ to probe inflation physics



How did Galaxies Begin?

SPHEREx extragalactic background measurements determine the total light emitted by galaxies



What are the Conditions for Life?

SPHEREx will measure the H_2O, CO, CO_2, CH_3OH ice content in clouds and disks, determining how ices are inherited from parent could vs. processed in disks

Interstellar ices sourced Earth’s water

CATALOG-BASED OPPORTUNITIES

SPHEREx

Improving Exo-Planet Characterization

→ SPHEREx provides precise near-IR spectra of 100s of millions of stars

→ Exo-planet characterization is often limited by knowledge of the host star!

Combining SPHEREx and Gaia data and improve the uncertainty of exo-planet radii by a factor of ~10

Low-Mass Star & Brown Dwarf Survey

→ SPHEREx provides spectra of the nearest 10,000 brown dwarfs and low-mass stars

→ Near-IR spectra are ideal for studying temperatures and atmospheric features

New understanding of composition, age and evolution

Galaxy Cluster Survey

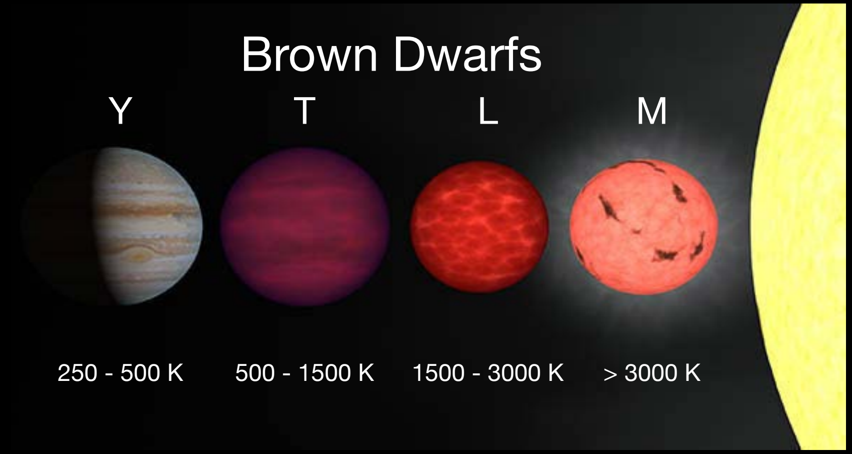
→ SPHEREx spectra will provide redshifts for a large fraction of the ~100,000 clusters expected from eROSITA, Euclid, Rubin, Simon's Observatory, SPT and CMB-S4

Cluster abundance, mass, and redshifts enable sensitive tests of cosmological parameters

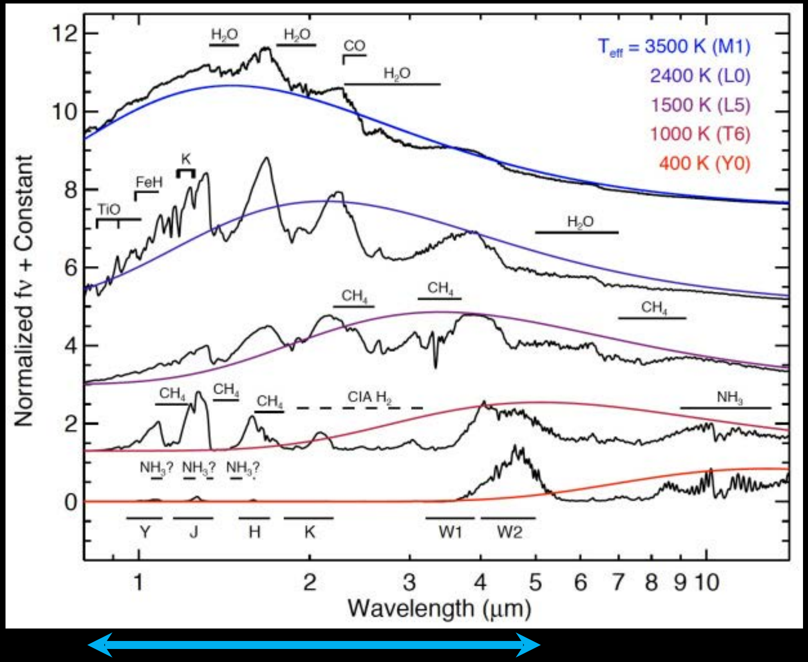
Solar System Objects Survey

→ SPHEREx provides an unbiased spectral survey of 100,000 asteroids and comets (Zeljko et al. 2022)

Vital information on origin and migration of solar system bodies

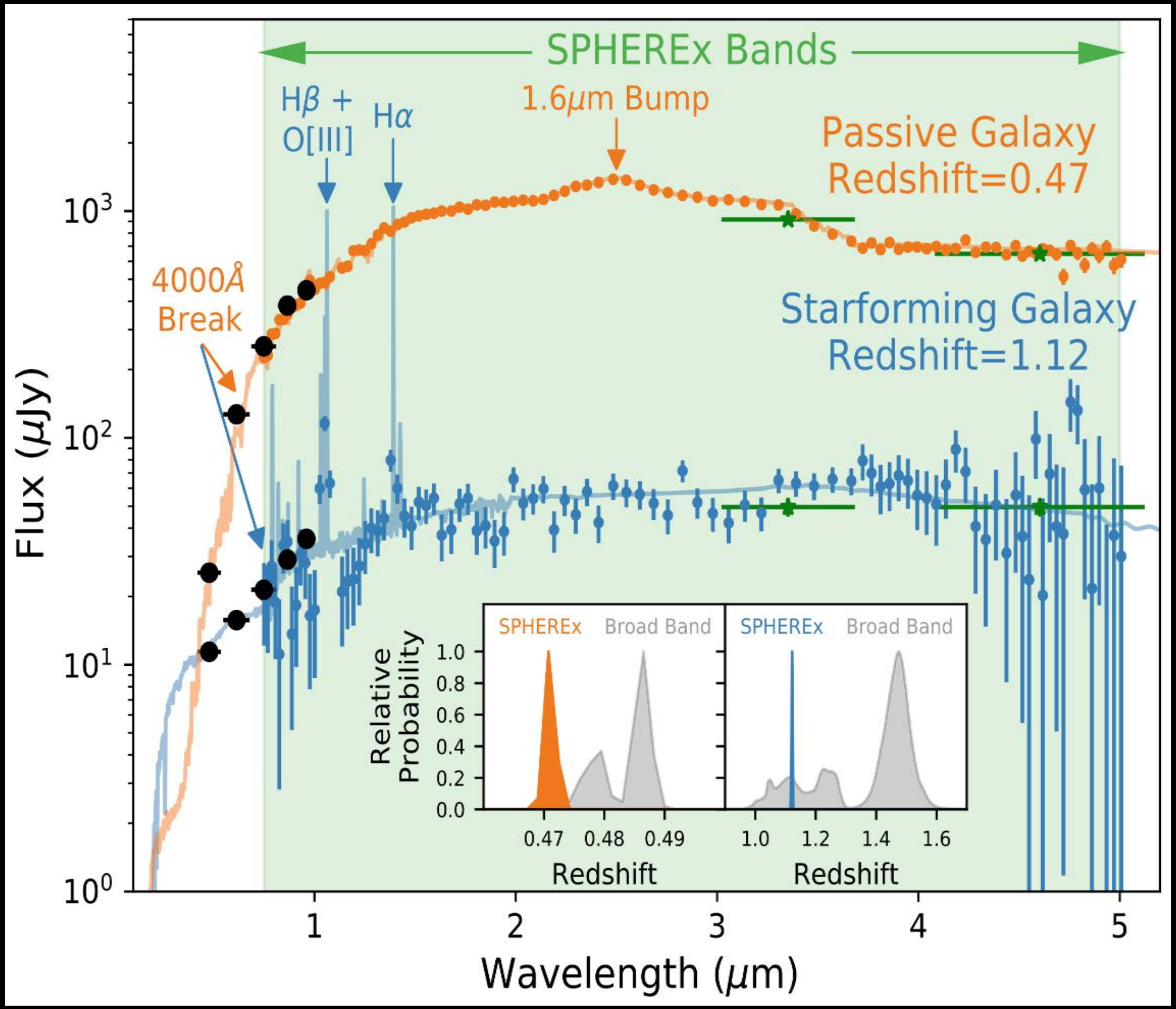


Brown Dwarf Atmospheric Spectra



SPHEREx Wavelength Coverage

How SPHEREx Determines z



Detected galaxies		**> 1 billion**
Galaxies	**$\Delta z/1+z$ < 10 %**	**> 450 million**
Galaxies	**$\Delta z/1+z$ < 0.3%**	**> 10 million**

- We extract the spectra of *known* sources using the full-sky catalogs from PanSTARRS/DES.

➡ Controls blending and confusion

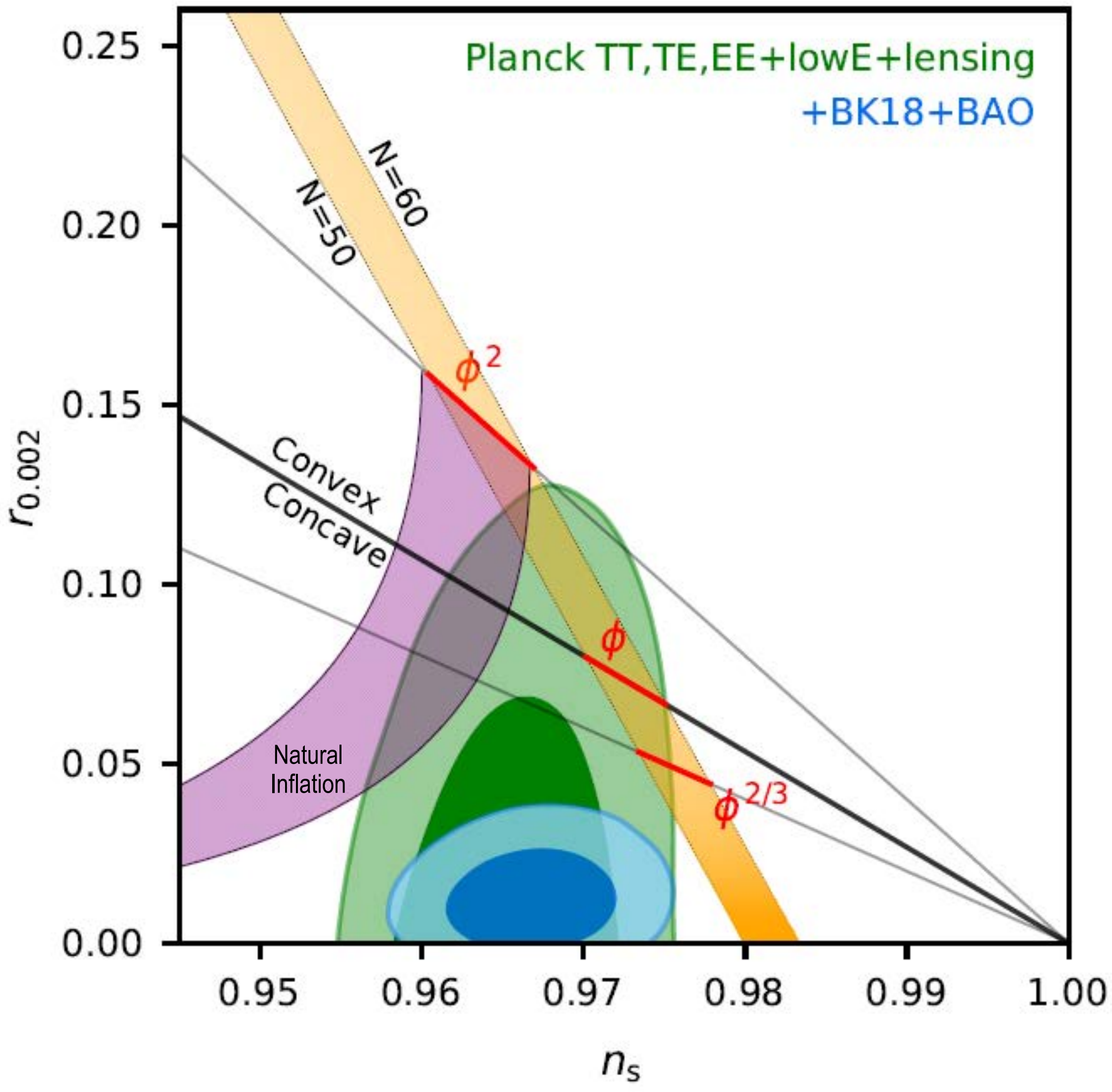
- We compare this spectra to a template library (robust for z < 1.5 sources).

➡ For each galaxy: redshift & type

- The 1.6 μm bump is a well established universal photometric indicator, see Simpson & Eisenhardt 99.

- We simulate this process using the COSMOS data set (similar to Euclid/WFIRST assessments; Stickley et al.)

CURRENT CMB POLARIZATION CONSTRAINTS



BICEP-Keck Continues to Lead the Field

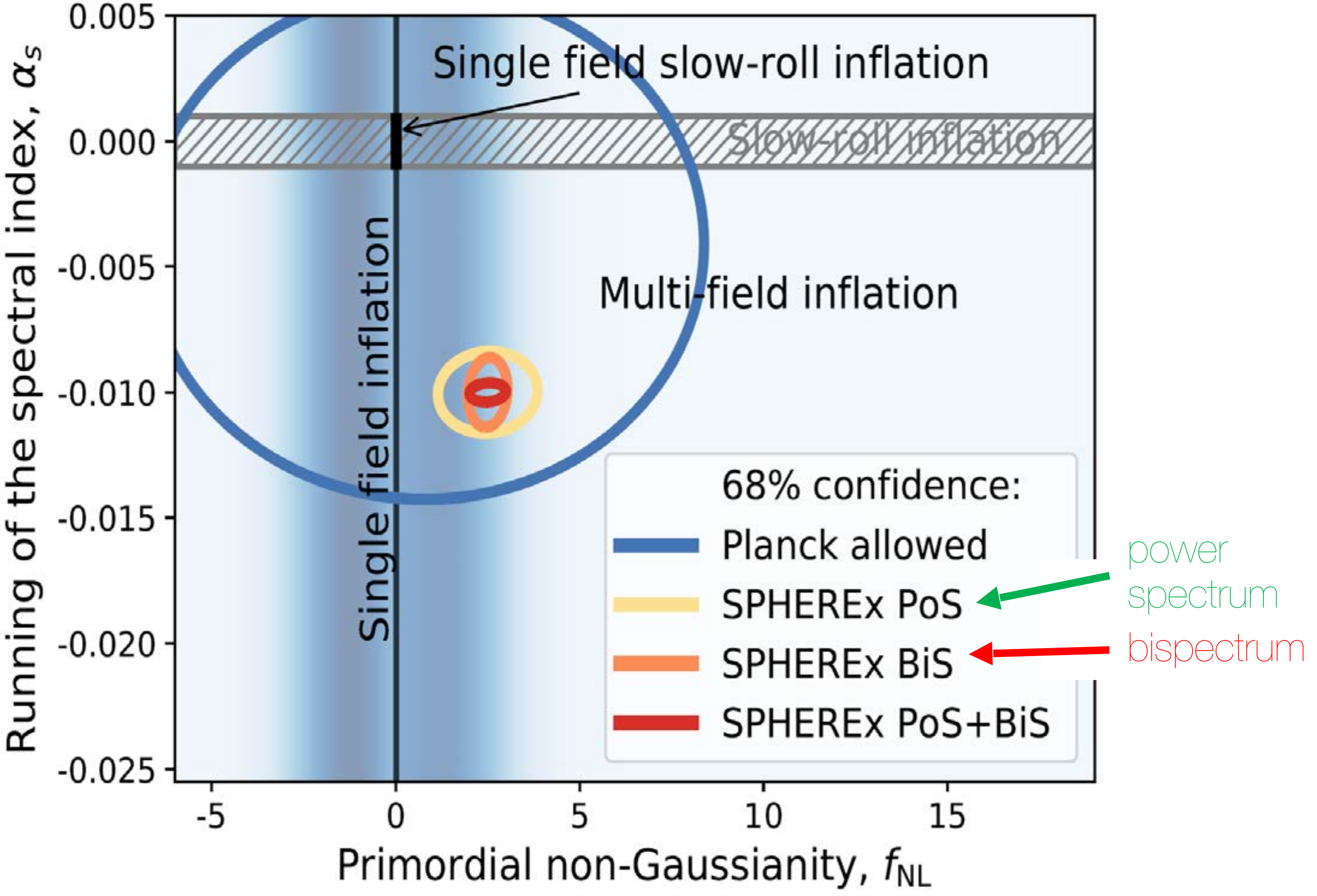
	arXiv	$\sigma(r)$ BB
DASI	0409357	7.5
BICEP1 2yr	0906.1181	0.28
WMAP 7yr	1001.4538	1.1
QUIET-Q	1012.3191	0.97
QUIET-W	1207.5034	0.85
BICEP1 3yr	1310.1422	0.25
BICEP2	1403.3985	0.10
BK13/*Planck*	1502.00612	0.034
BK14/WP	1510.09217	0.024
ABS	1801.01218	0.7
Planck	1807.06209	~ 0.2
BK15	1810.05216	0.020
POLARBEAR	1910.02608	0.3
SPTPOL	1910.05748	0.22
Planck/Tristram	2010.01139	0.07
Spider	2103.13334	0.13
BK18	2110.00483	0.009

BK + SPT	c. 2024	0.006
BICEP Array	c. 2029	0.003
CMB-S4	c. 2038	0.0005

SPHERE^{X} TESTS INFLATIONARY NON-GAUSSIANITY

- Single-field models predict $f_{NL} < 0.01$
- Multi-field models predict $f_{NL} > \sim 1$
- Non-inflationary models (Steinhardt *et al.*) predict $f_{NL} \sim 1$

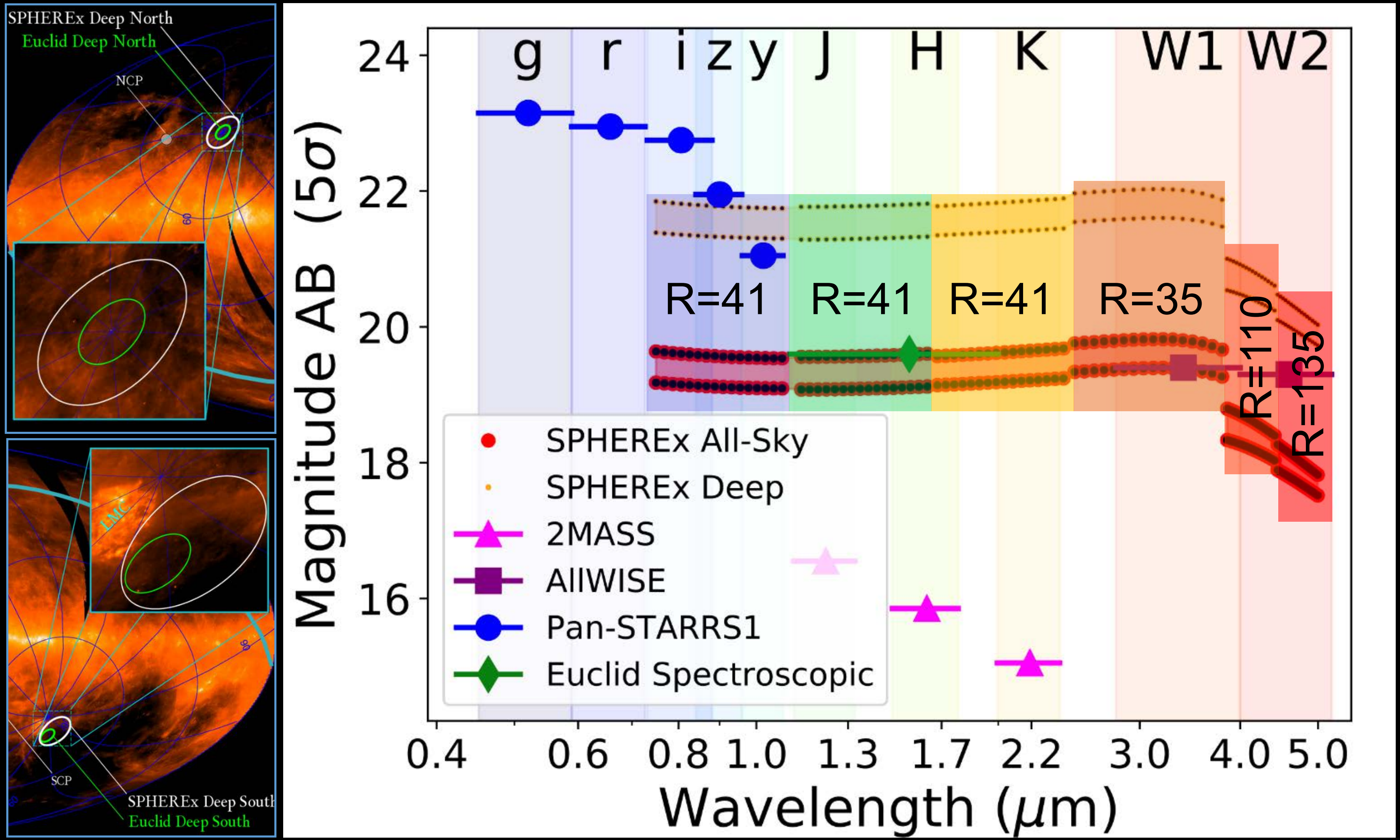
Did Multiple Fields Drive Inflation?



SPHEREx improves accuracy to σ(fNL) < 0.5

- >10x improvement on cosmic-variance-limited CMB f_{NL} measurements

SPHEREx Deep Fields + Euclid Deep



Predicting High Redshift Quasar Detection with SPHEREx (+ Euclid)

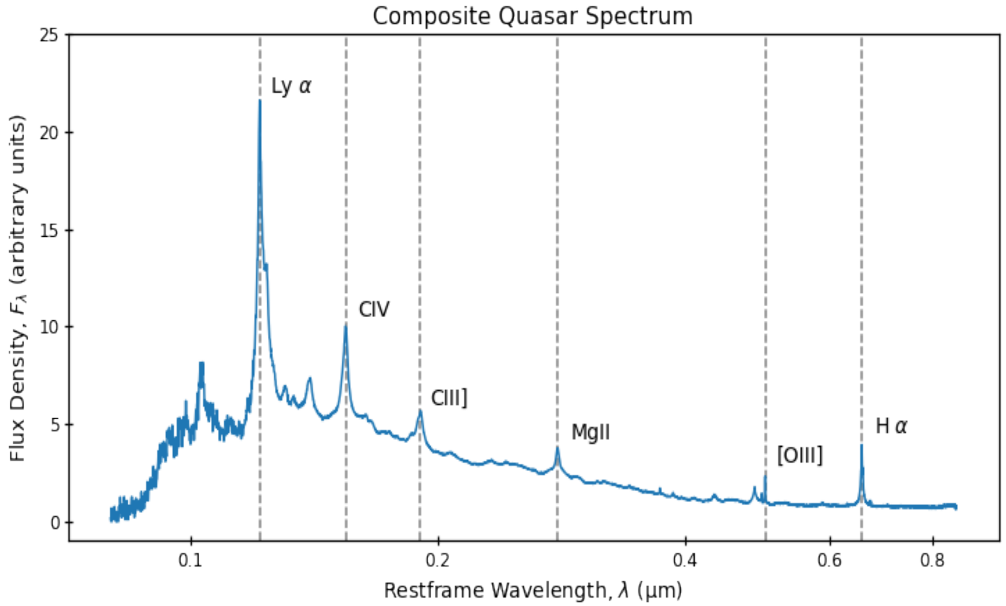
Brandan Buschmann

- Objective: Simulate the number of high redshift quasars detected by SPHEREx, aided by Euclid photometry.

Methods

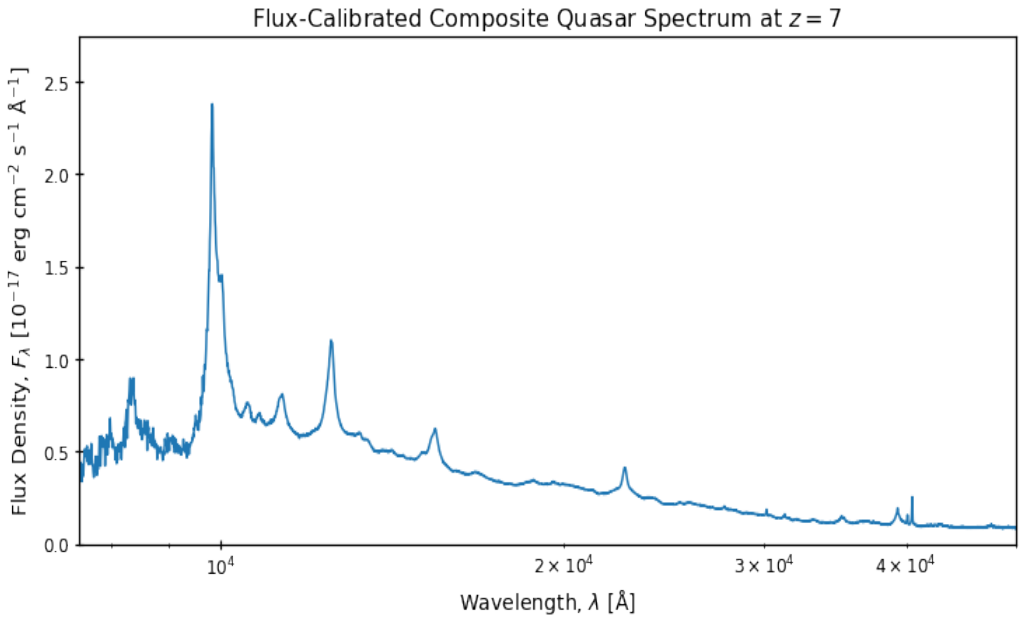
- Generate a synthetic catalog of high redshift quasars.
- Use catalog to inject quasars into SPHEREx Sky Simulator and generate a mosaic for each band and channel.
- Use photometric detection algorithms to recover quasars in mosaics.
- Future: Compare with actual SPHEREx/Euclid data.

Generating composite high redshift quasar spectra at z = 5 - 8



Composite quasar spectrum from Sloan Digital Sky Survey [1] in rest-frame, labeled with emission lines.

Used spectrum of ULAS J112001.48+064124.3 [2] to flux-calibrate original spectrum.

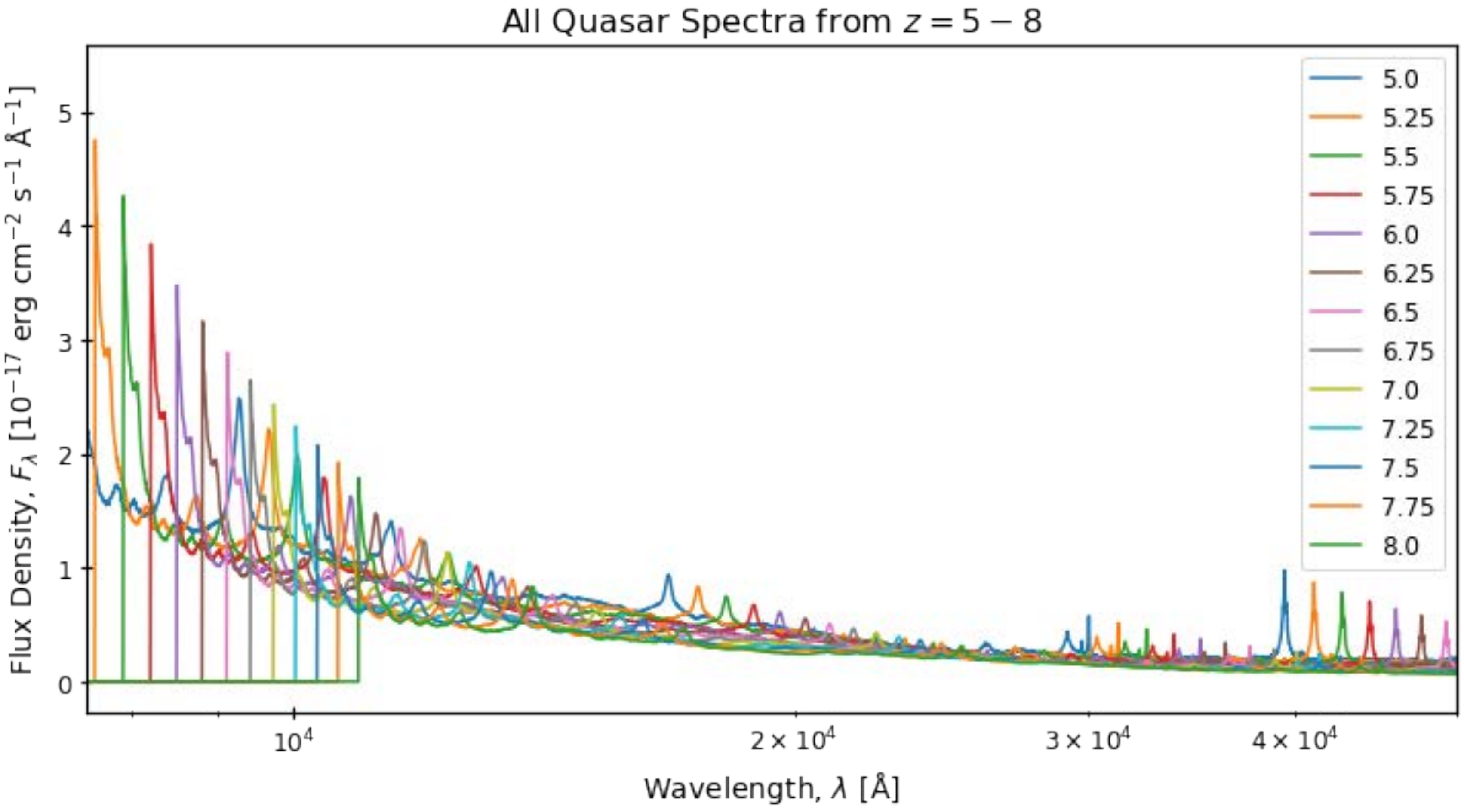


Flux-calibrated spectrum between wavelengths 0.75 – 5 microns.

[1] D.E. Vanden Berk, et al., Astron. J. **122**, 549 (2001).
[2] D. Mortlock, et al., Nature **474** (2011).

Generating composite high redshift quasar spectra at z = 5 – 8 continued

- The flux-calibrated spectrum was adjusted for redshift to generate 13 quasar spectra between z = 5 – 8 with Lyman break.



Creating a synthetic high redshift quasar catalog

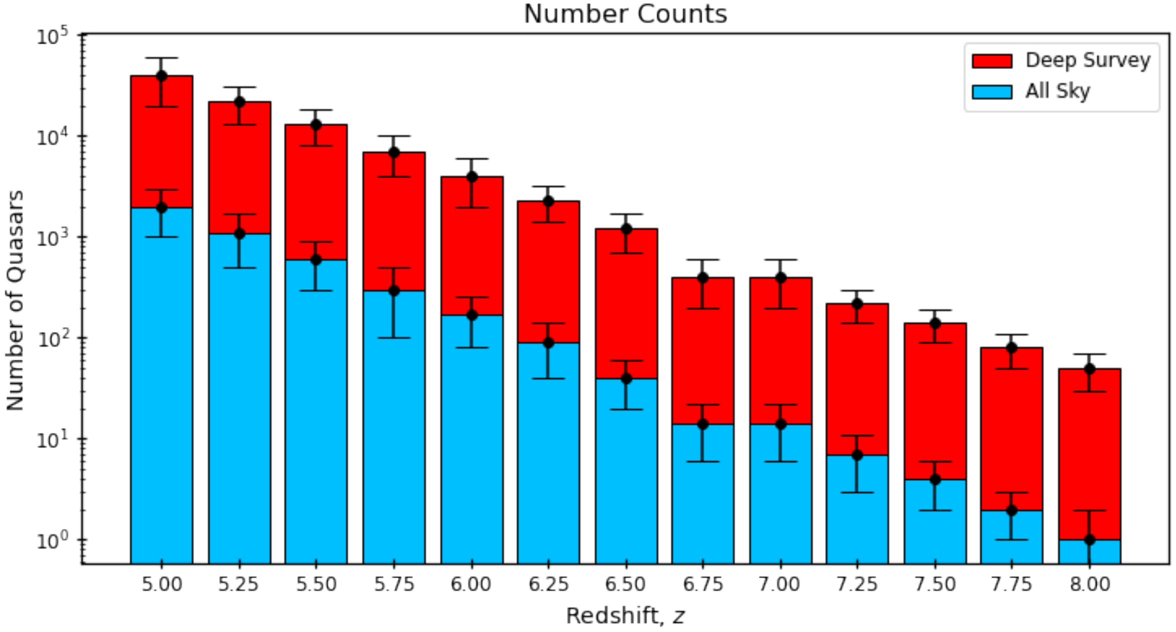
- Integrated a double power law quasar luminosity function [3] to calculate number counts at each redshift.

$$\Phi(M, z) = \frac{\Phi^*(z)}{10^{0.4(\alpha+1)(M-M^*)} + 10^{0.4(\beta+1)(M-M^*)}}$$

Table 5
MLE Fit Parameters

$\log \Phi^*(z=6)$[a]	M^*_{1450}	α	β
$-8.94^{+0.20}_{-0.24}$	$-27.21^{+0.27}_{-0.33}$	$-2.03^{+0.15}_{-0.14}$	-4.00
$-8.10^{+0.30}_{-0.25}$	$-25.88^{+0.60}_{-0.49}$	-1.50	$-3.12^{+0.28}_{-0.41}$
$-8.40^{+0.03}_{-0.03}$	-26.39	-1.80	-3.26

Notes. Parameters without uncertainty ranges are fixed during the maximum likelihood fitting.
[a] $\log \Phi^*(z) = \log \Phi^*(z=6) + k(z-6)$, with $k = -0.47$.

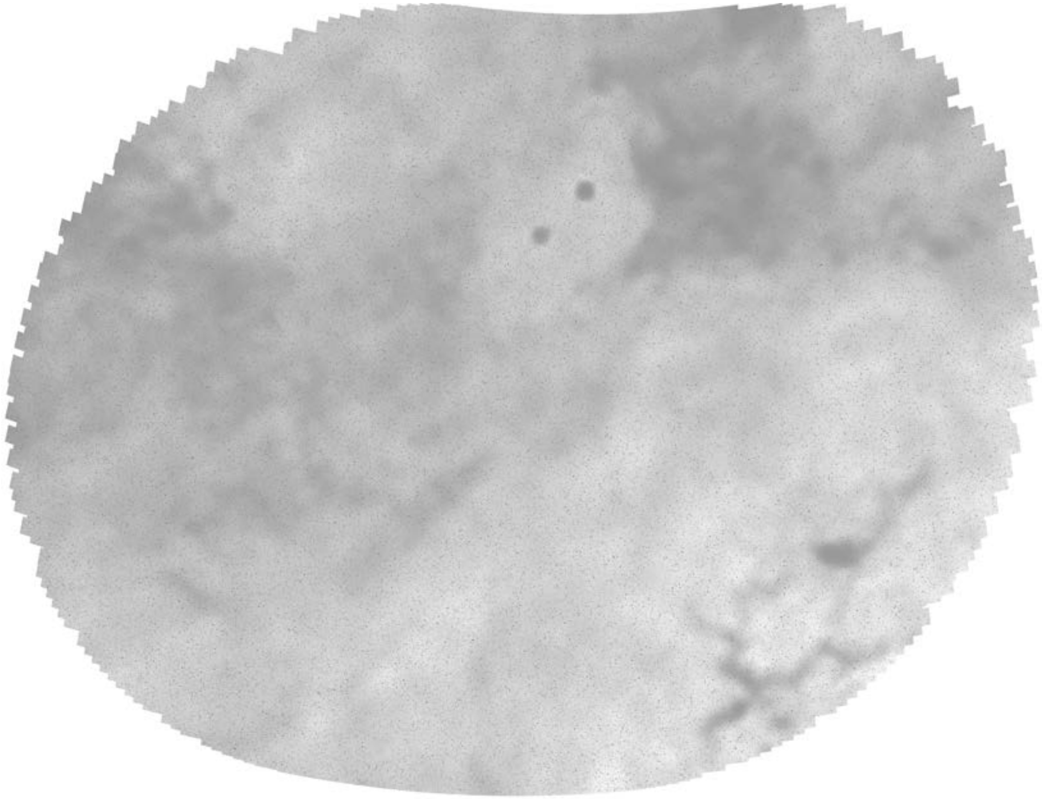


- Error bars were generated by 1000 trials of random sampling the QLF parameters (α, β, M^*, etc.) within their uncertainty ranges.
- By randomly assigning an RA and Dec to each quasar, these locations were combined with the spectra on the previous slide to generate a synthetic catalog of 90790 high redshift quasars.

[3] I.D. McGreer, et al., Astrophys. J. **768**, 105 (2013).

Injecting quasars into the SPHEREx Sky Simulator to generate mosaics

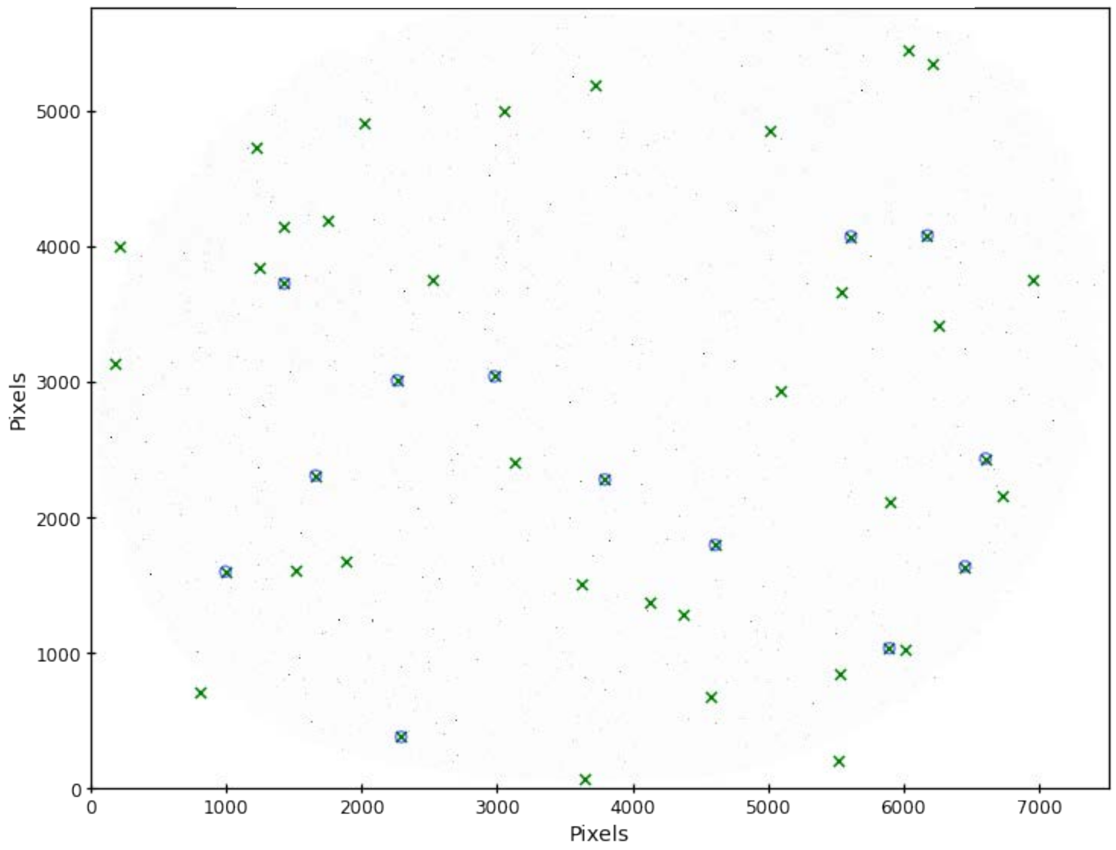
- The catalog was used to inject high redshift quasars into the Sky Simulator, which generates mosaics of the sky.
- SPHEREx has 6 bands with 17 channels in each, for a total of 6 x 17 = 102 mosaics.



Mosaic for array 1, channel 6 (0.88 – 0.9 microns).

Using photometric detection algorithms to recover quasars in the mosaics

- Currently using the DAOStarFinder algorithm from the Python package Photutils to locate the quasars in the mosaics.
- Quasar recovery in the mosaic for array 1, channel 6 is shown to the right.
  - 13 of the 44 quasars at z = 7 were recovered within 1 pixel of their known locations, for a recovery rate of 30%.
- Currently refining the algorithm's parameters, such as the detection threshold, to increase the recovery rate.
- May resort to machine learning if this photometric detection method is ineffective at recovering most of the quasars.
- To improve detection will be adding Euclid photometry in overlapping SEP/NEP deep fields.



Detection of z = 7 quasars in the mosaic for array 1, channel 6. The green X's are the known positions of quasars, and the blue circles are the recovered quasars.

SPHERE^{X} LINE INTENSITY MAPPING

How Does Line Intensity Mapping Work?

→ Maps large scale-structure using the collective light from galaxies -- not from individual galaxy detections

→ Line emission uniquely gives the redshift

A powerful 3-D map of galaxy and star formation!

What Does SPHEREx Provide?

→ The core SPHERE^{X} program is 2-D and does not use line spectroscopy

→ However all of the deep field maps will be ready-built for spectroscopic analysis

→ Line emission maps can detect Hα, Hβ, OII and OIII lines with high SNR

→ Lyα line accessible at high redshifts z > 5.2

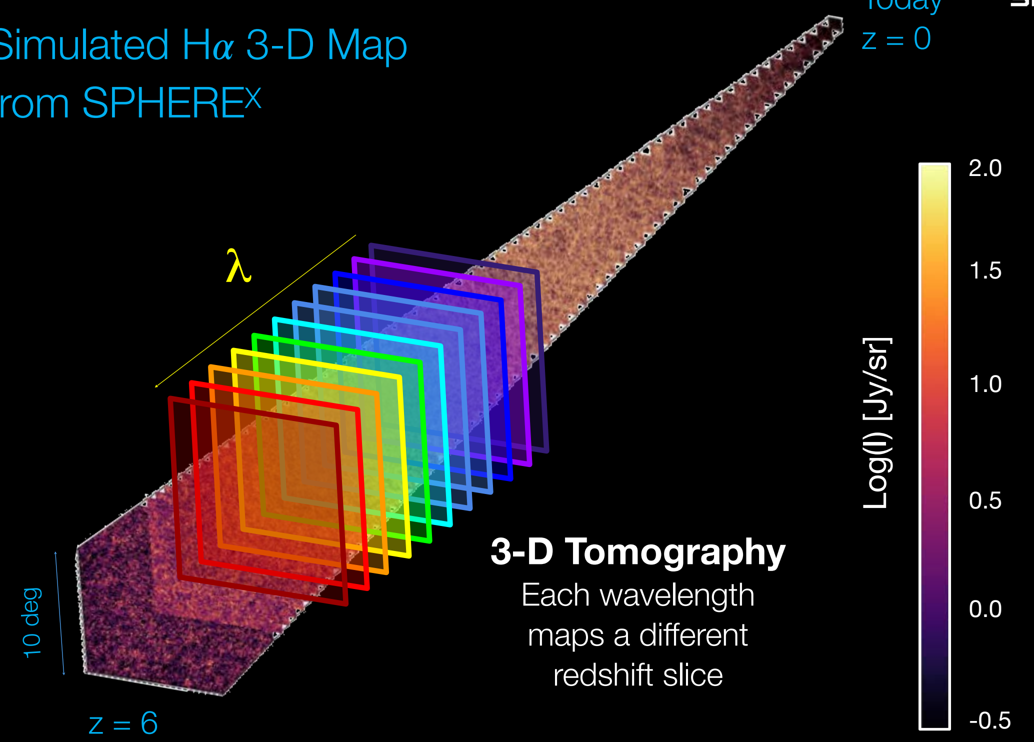
Scientific Opportunity

→ Map the entire History of Galaxy and Star Formation in multiple lines (Hα, Hβ, OII and OIII)

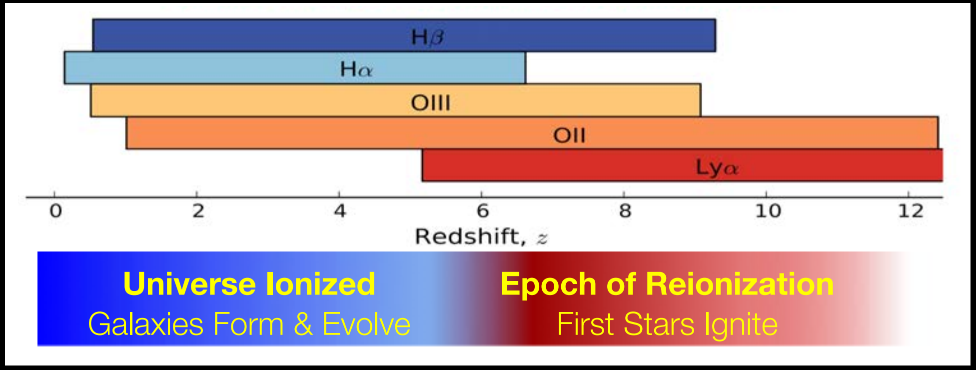
→ Offer new insights on the Epoch of Reionization through Lyα and OIII

→ Provide unique measurements on the Geometry of the Universe at High Redshift (z ~ 4-5)

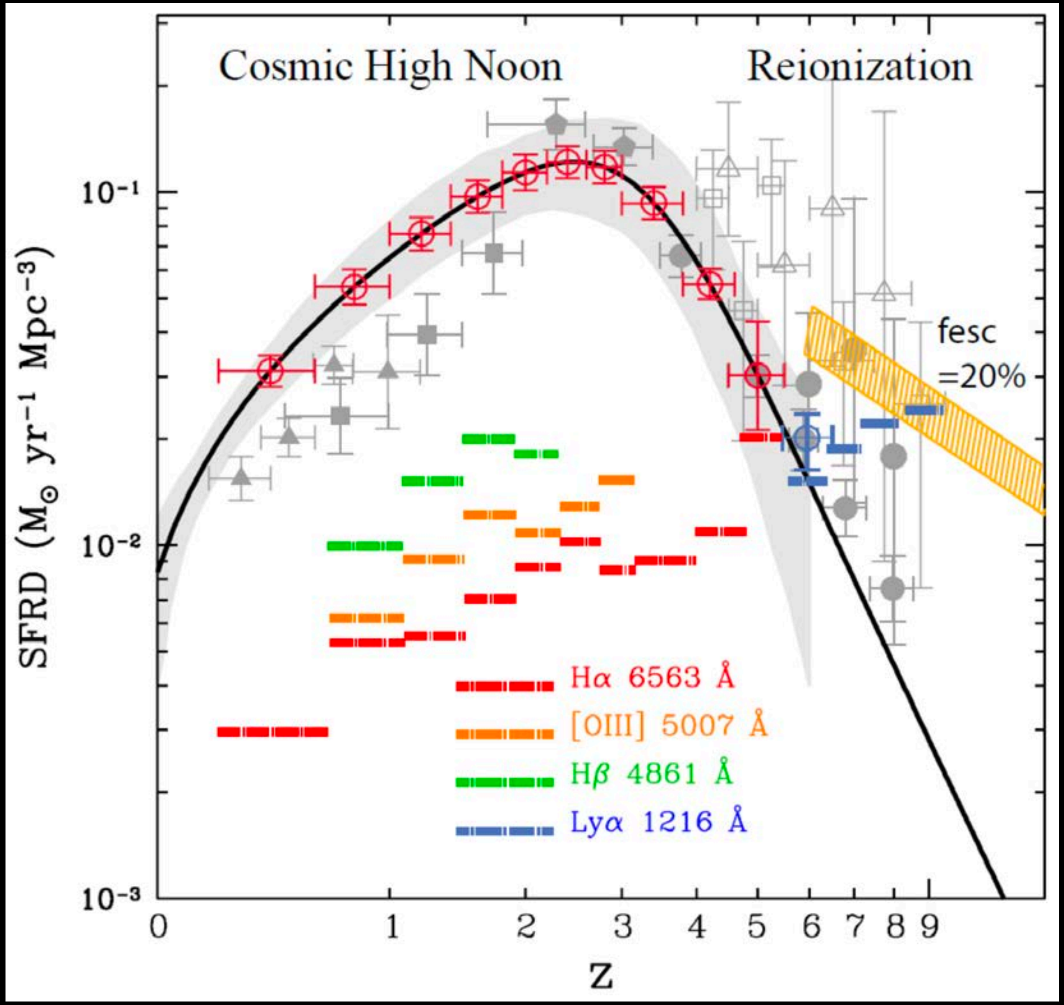
Simulated Hα 3-D Map from SPHERE^{X}

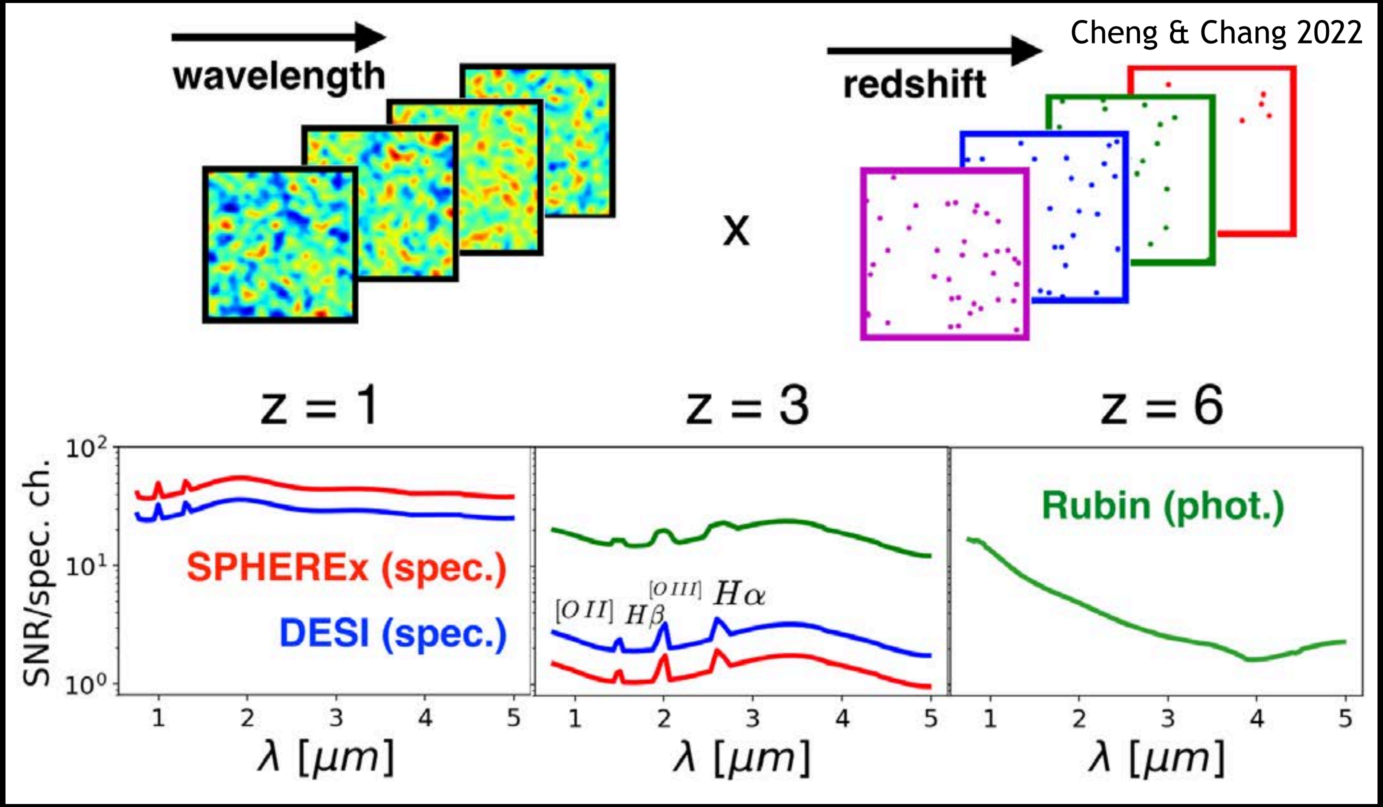


Emission Lines Observable by SPHERE^{X}



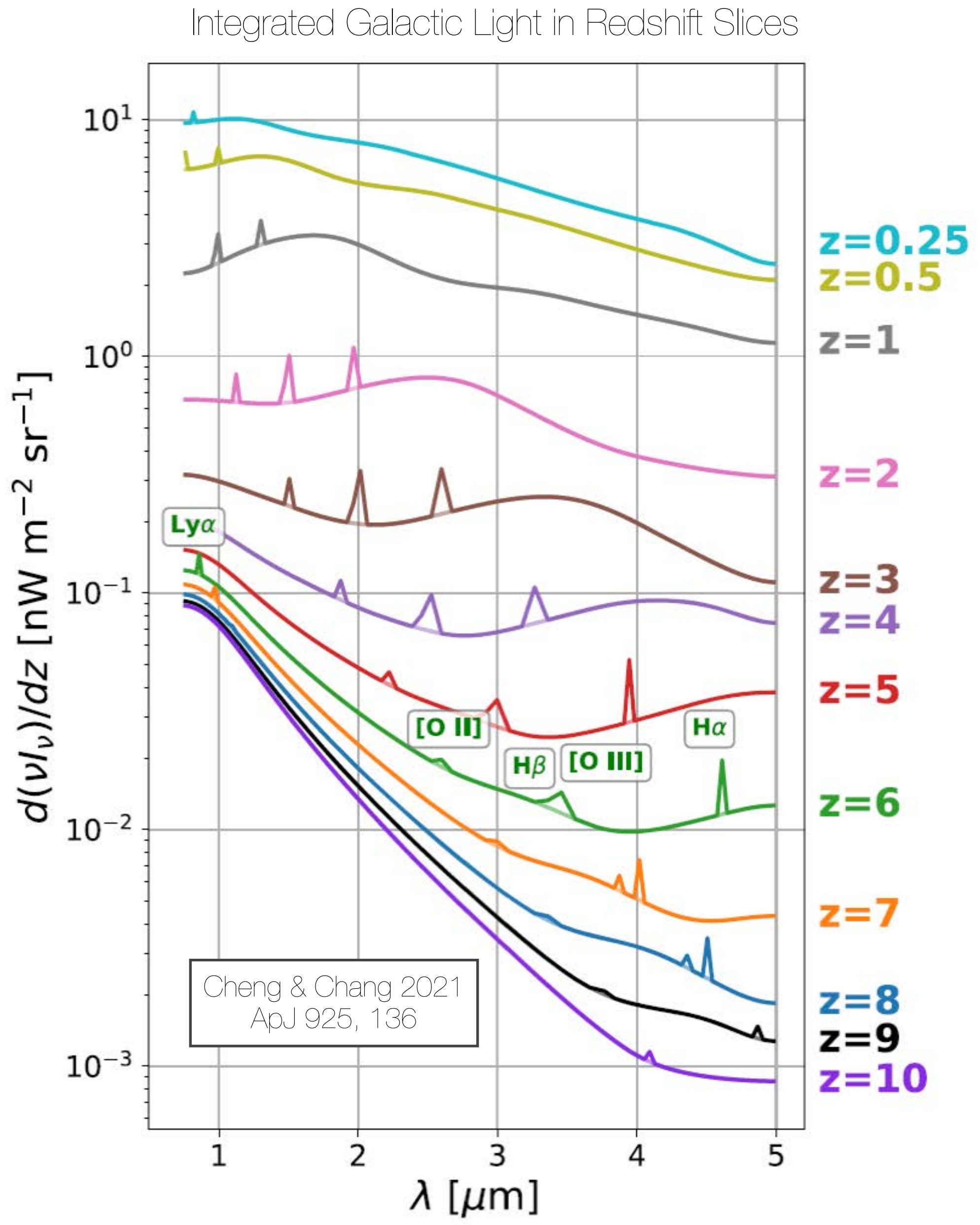
Intensity Mapping with SPHEREx

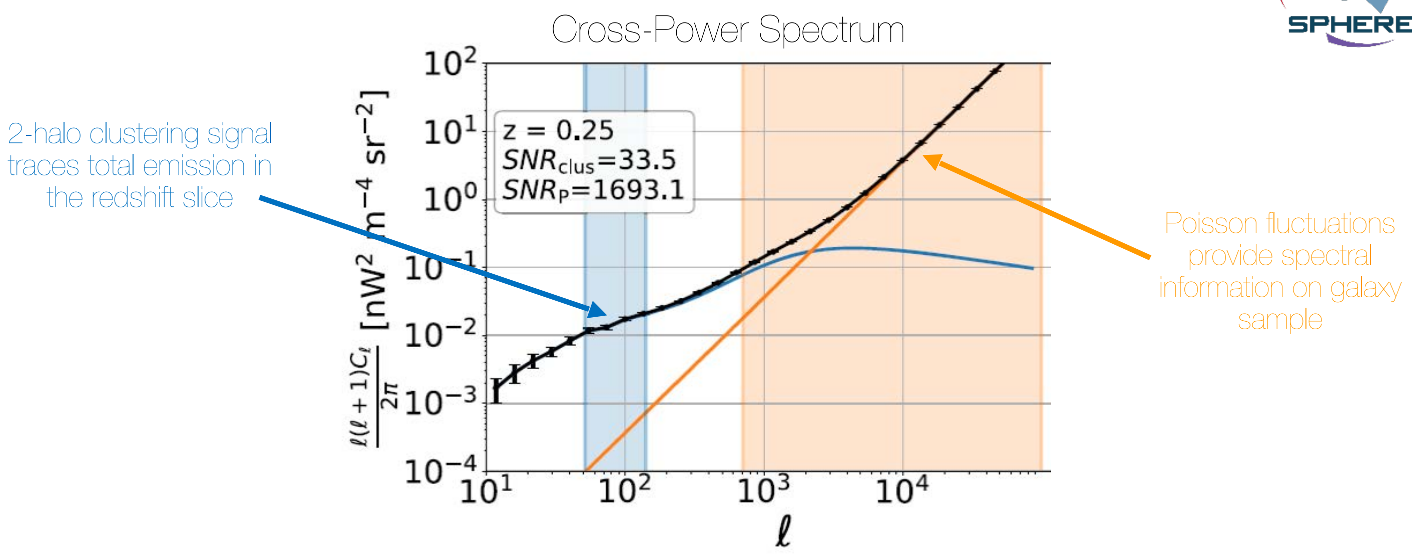




Auto-correlations in R=41.5 bands internal to SPHEREx.
Cross-correlations w/ galaxy catalogs (e.g., Euclid-Deep, Rubin/LSST etc) map history of emission in galaxies.

LINE-INTENSITY MAPPING WITH GALAXY CROSS-CORRELATIONS

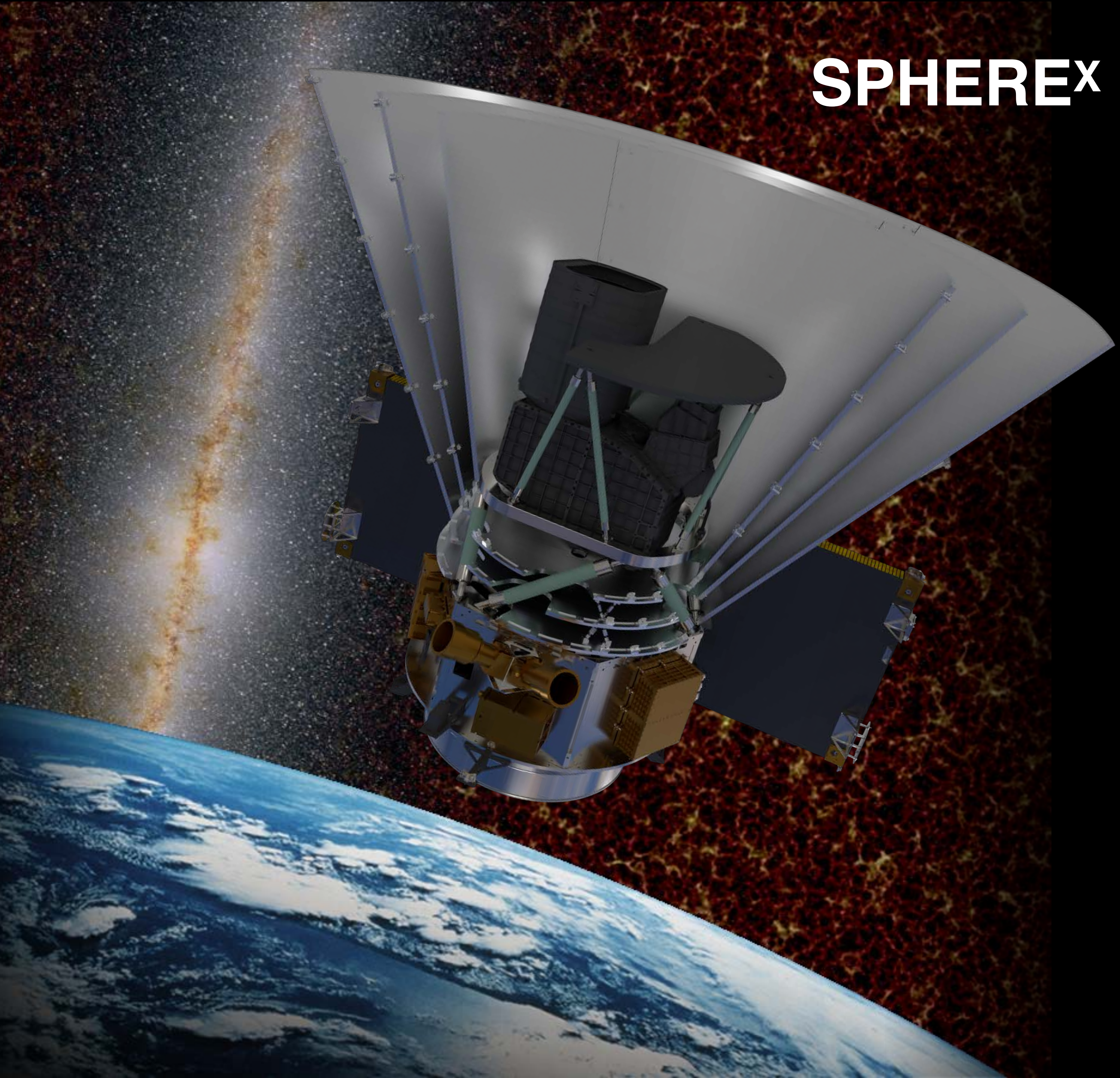




Redshift	Δz slice	SNR per channel*	Galaxy Redshifts	Lines in range
0.25	0.1	50 – 65	SPHEREx	Hα
0.5	0.2	100 – 120	SPHEREx	Hα, OIII, Hβ
1	0.4	100 – 150	SPHEREx	Hα, OIII, Hβ
2	1	15 – 25	SPHEREx	Hα, OIII, Hβ, OII
3	1	6 – 9	DESI	Hα, OIII, Hβ, OII
4	1	15 – 30	Rubin/Euclid	Hα, OIII, Hβ, OII
5	1	5 – 20	Rubin/Euclid	Hα, OIII, Hβ, OII
6	1	3 – 10	Rubin/Euclid	Hα, OIII, Hβ, OII, Lyα
7	1	0.2 – 2	Roman/Euclid	OIII, Hβ, OII, Lyα

*There are 102 spectral channels in total

SPHERE^X IN SUMMARY



SPHEREX Core Science

- Origin of the Universe
- Origin and History of Galaxies
- Origin of Water in Planetary Systems

Exciting Potential Opportunities

- Exo-Planet Host Stars
- Low-Mass Star & Brown Dwarf Survey
- Galaxy Cluster Survey
- Solar System Objects Survey
- Line Intensity Mapping (and X-Correlations)
- Joint Cosmology with CMB-S4
- *You May Have More Ideas!*

Project is Entering a Critical Phase

- Starting Hardware Test and Integration