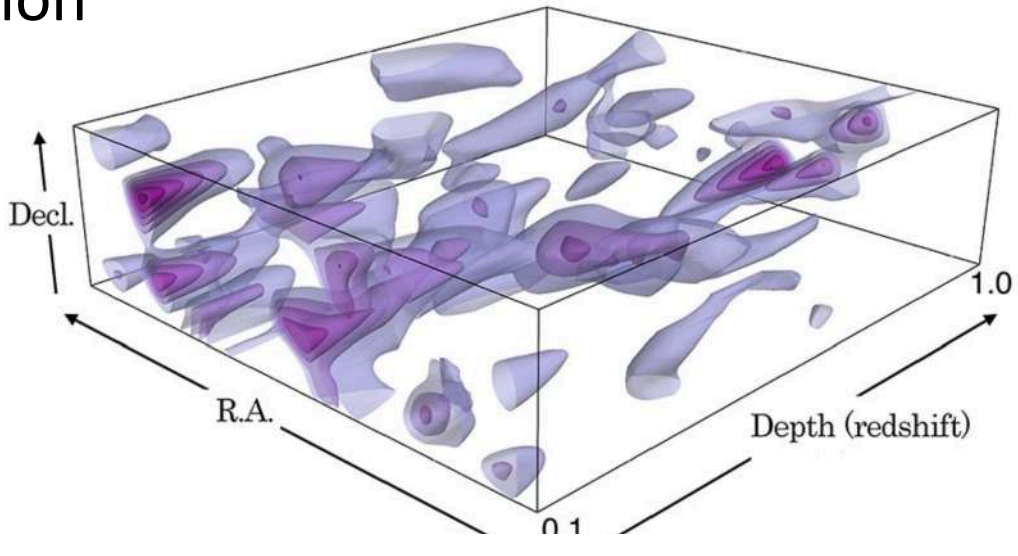
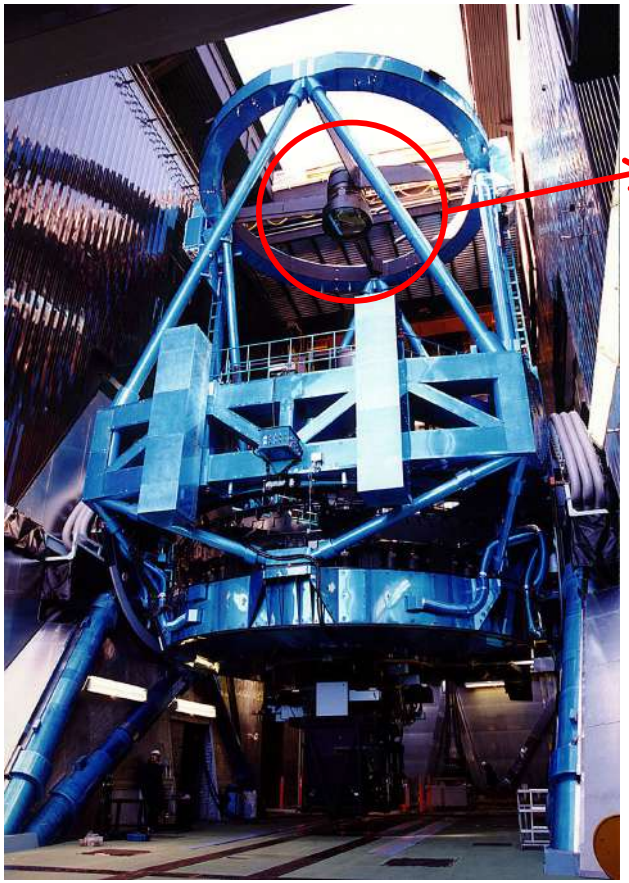
Cosmology and Fundamental Physics with Big Astronomical Data

Naoki Yoshida, Nao Suzuki, Ichiro Takahashi (U-Tokyo)
Naonori Ueda, Akisato Kimura (NTT)
Shiro Ikeda, Mikio Morii (Institute of Statistical Mathematics)
Hideyuki Kawashima, Osamu Tatebe (University of Tsukuba)
JST CREST Big Data Application




Night Sky

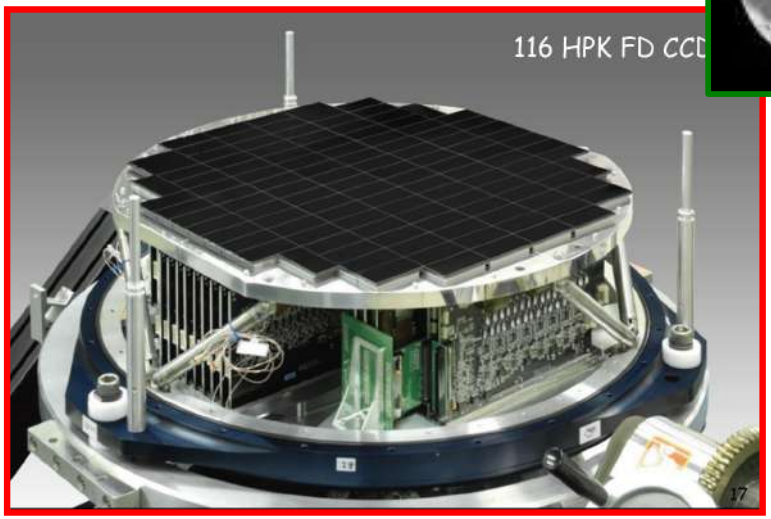
Subaru Hyper Suprime-Cam



First light in 2013

8.2 m diameter
Subaru
Telescope

Largest (3ton) digital camera
1.5 deg. FoV = 10 times
1000 times Hubble ST



104 CCDs produce a 1Giga pixel
image *per snapshot*

The 5-year survey from March 2014 to 2019, spending 300 nights.

1 PB data will be delivered. → 10 times larger than the largest survey (SDSS).

We need a much faster application system to reduce the images and analyze the data
to produce scientific outputs such as the 3D distribution of dark matter
and determination of the cosmological parameters.

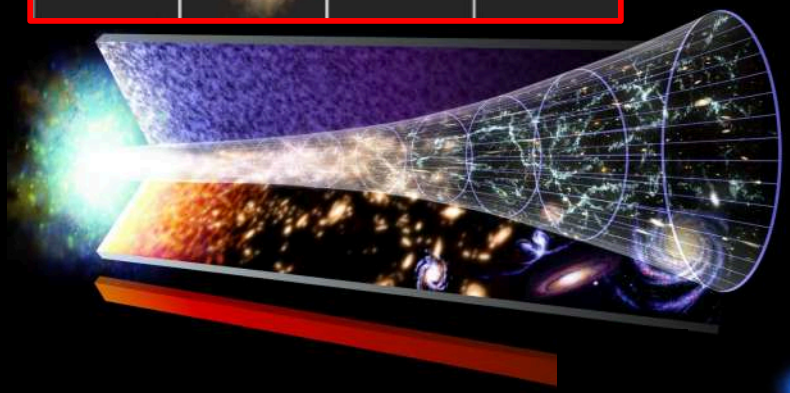
500,000,000 objects (galaxies/stars) will be catalogued on a high-speed database.

BIG DATA

BILLION GALAXIES

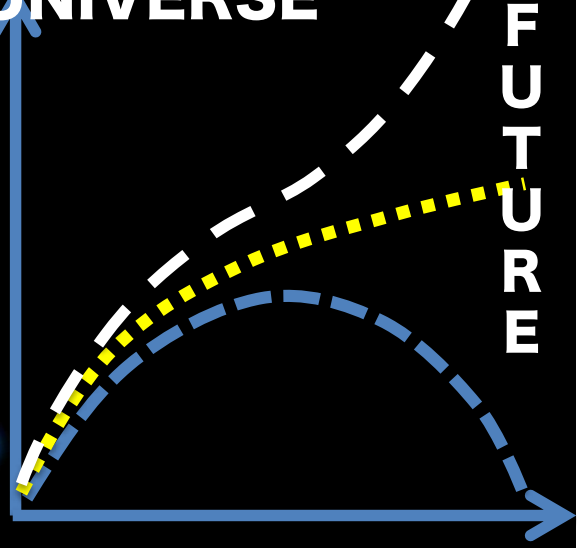
Subaru Telescope

DISTANT SUPERNOVAE

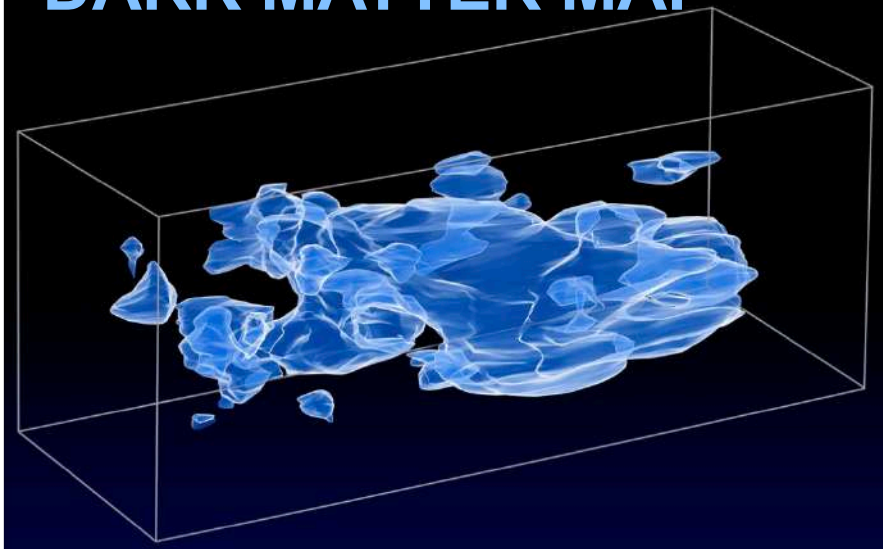


DATA ANALYSIS

EVOLUTION OF THE UNIVERSE



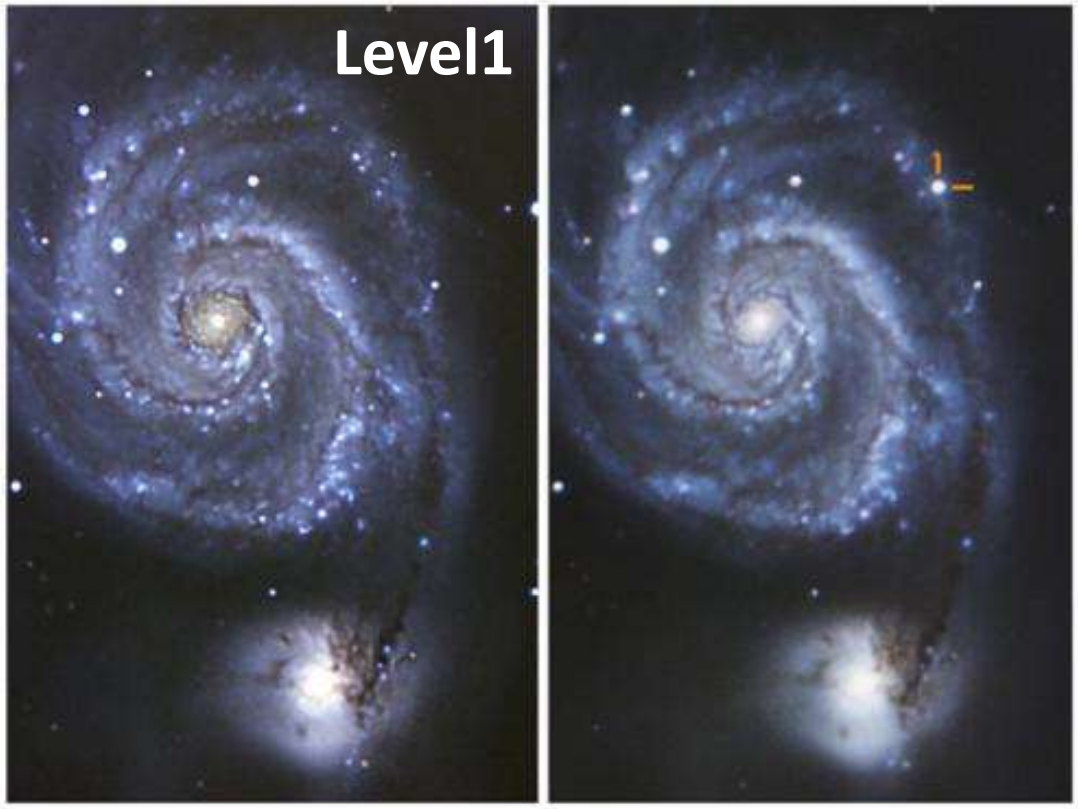
DARK MATTER MAP



Statistics
Data science
Machine-learning
Super-computing
Database eng.

Supernovae detection and classification

The Need for Real-Time Data Analysis



HSC detects **100 objects per night.**
Automatic classification is necessary.

They can be detected only through time-differencing multiple images.

The variation time-scale is from years to days, even over minutes.
Classification and rapid follow-up within a few days are desired.

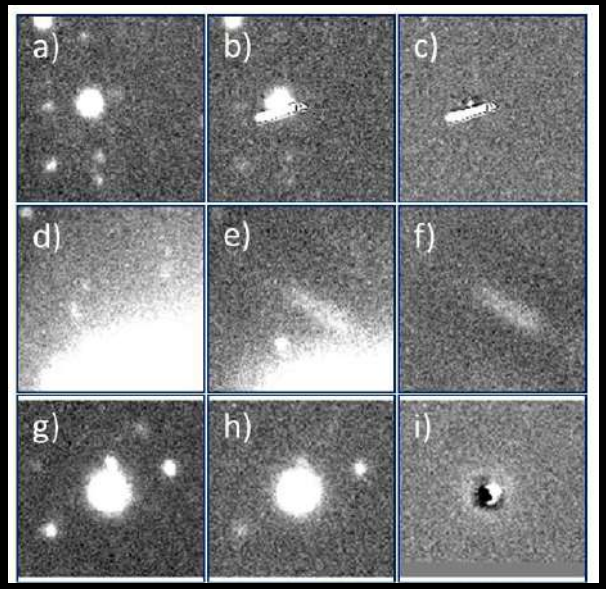
Type Ia supernovae, among others, can be used as a precise ruler, providing a unique way to measure the distance and the time.
→ the evolution and the future of the universe.

2014Apl

Level 99

Difference of
1 Giga-byte images

2015May



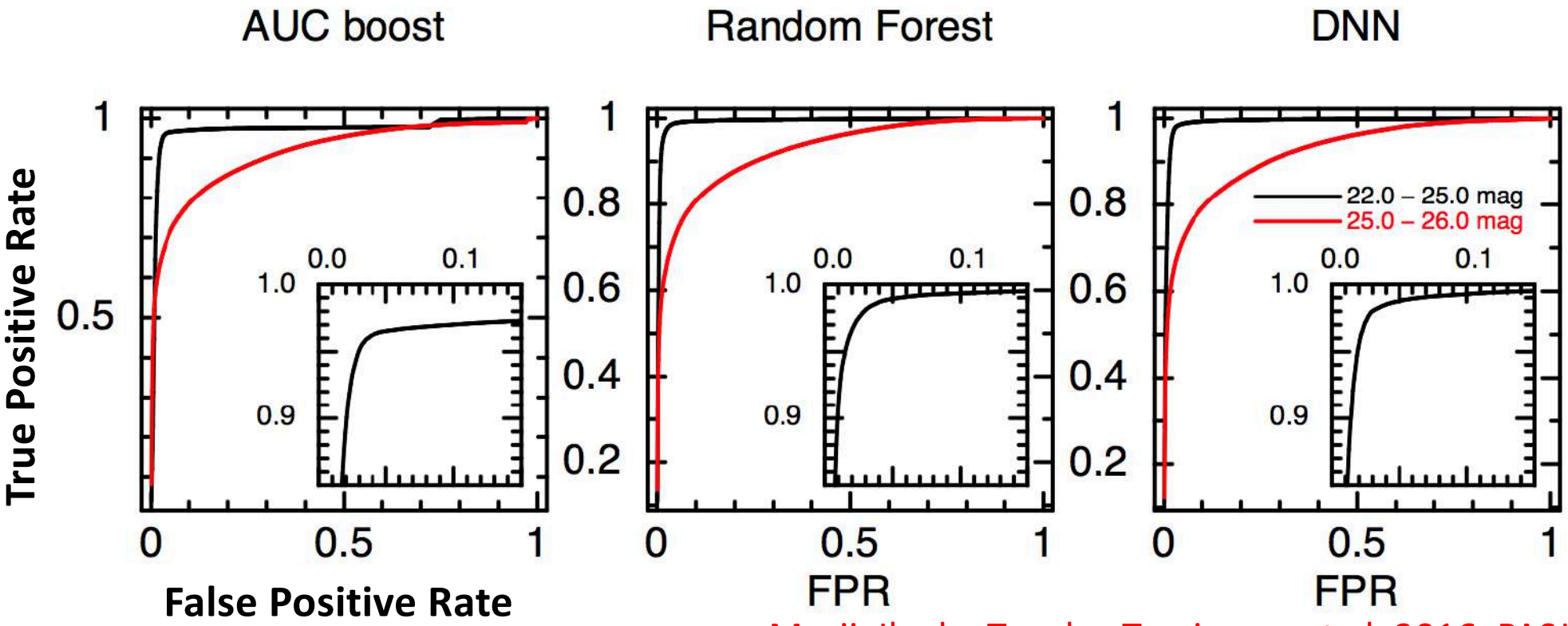
Object detection: Machine performance

- **Training data: 24000 transients including supernovae.**
  Real and artificial sources and use data augmentation

↑ Peculiar feature of our task: "1 positive out of 1000 negatives"

→ New machines: Random Forest, DNN, Boosting by AUC

- **23 features and/or 2-D images**



Morii, Ikeda, Tanaka, Tominaga, et al. 2016, PASJ

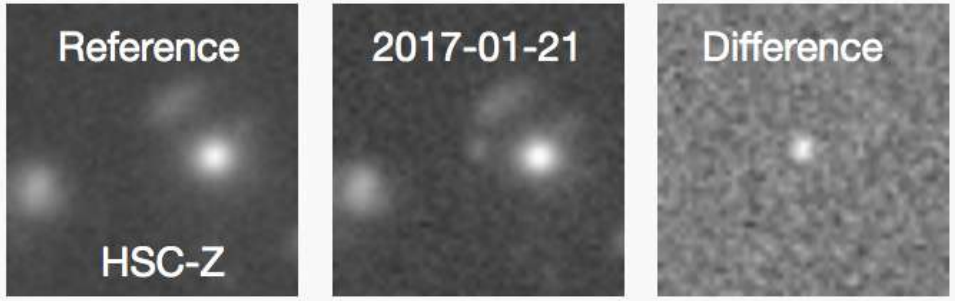
magnitude 21.7

magnitude 22.7

magnitude 23.1

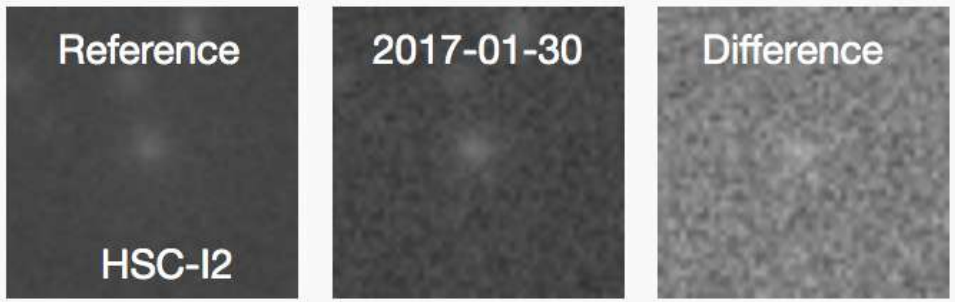
The power of Subaru

magnitude 25.1

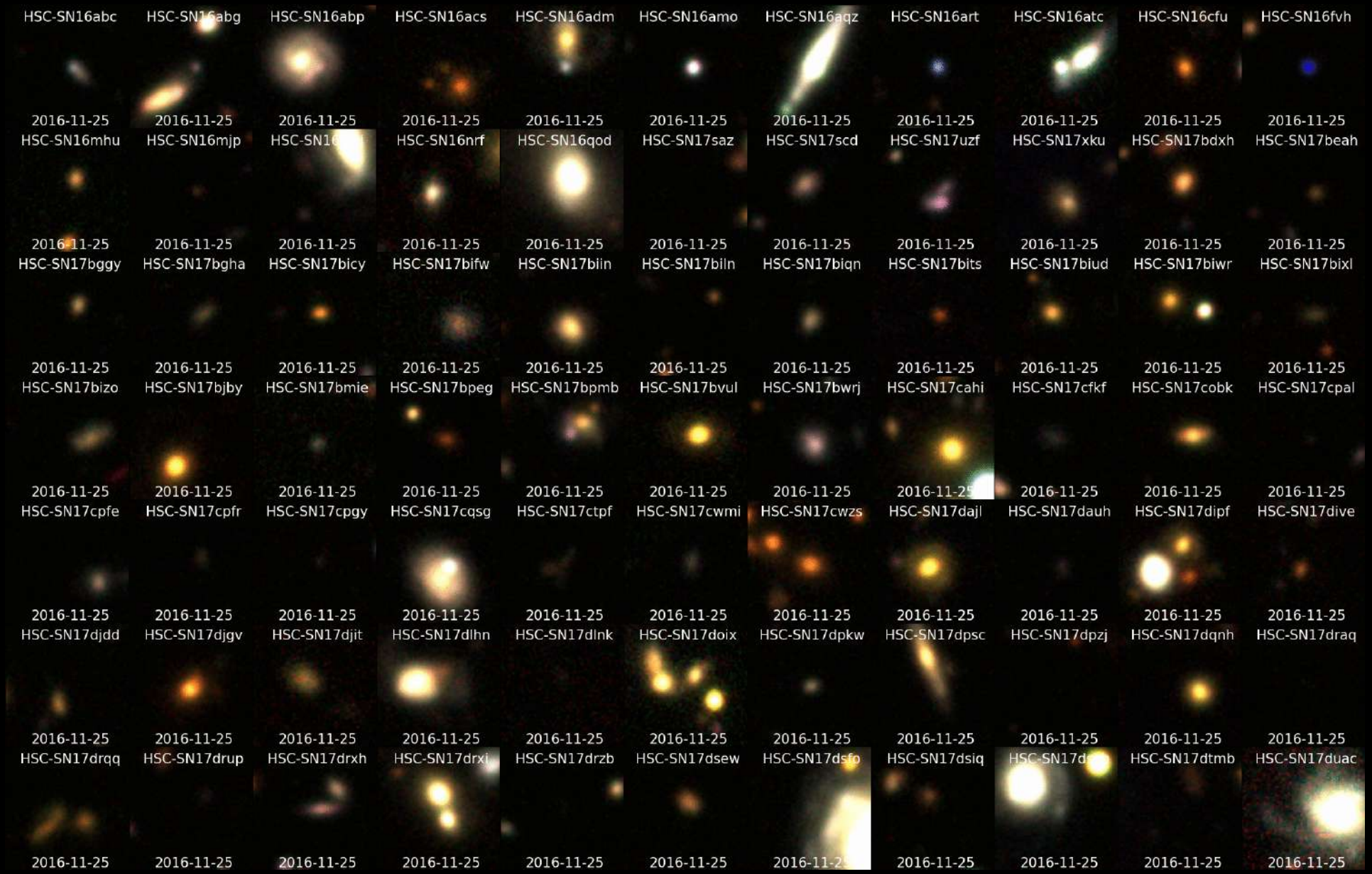


magnitude 25.3

magnitude 25.7



A gallery of discovered supernovae of many different types



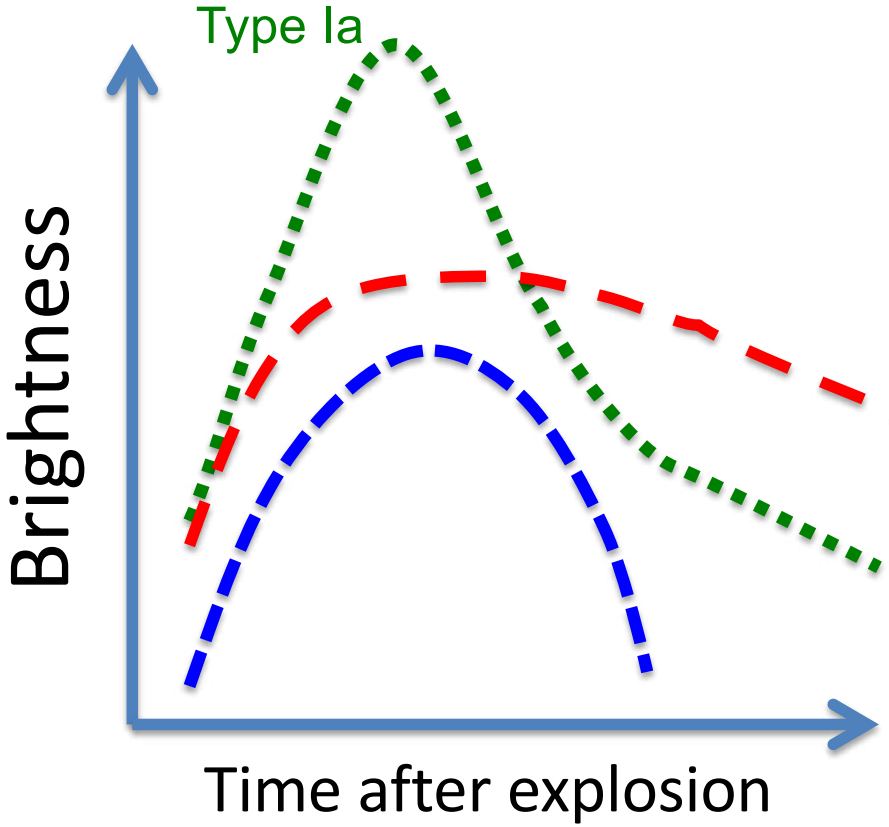
The machines are used for 52-nights observation in 2016/17

Runs	Nov/Dec	Dec/Jan	Jan/Feb	Feb / Mar	Mar / Apr	Apr / May
Transients	3597	9282	21232	29720	34538	35625
SN candidates	162	366	727	1025	1219	1293
Sent for classification	116	224	371	718	565	566
SN Ia (prediction > 0.9)	23	20	48	160	79	66
SN Ia (prediction > 0.8)	27	30	63	183	103	87

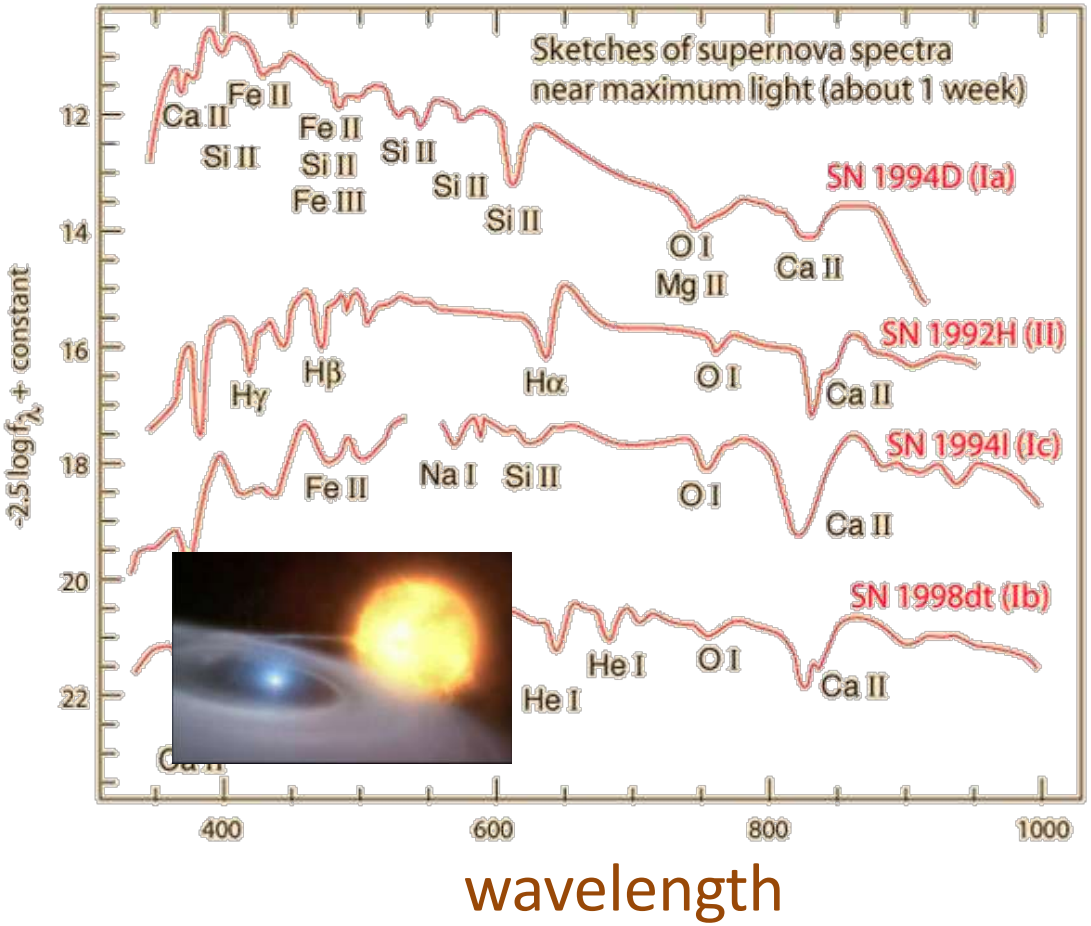
Best 20 are sent for space-telescope follow-up.

NASA Hubble Space Telescope

Supernova type and light curves



Spectroscopic classification

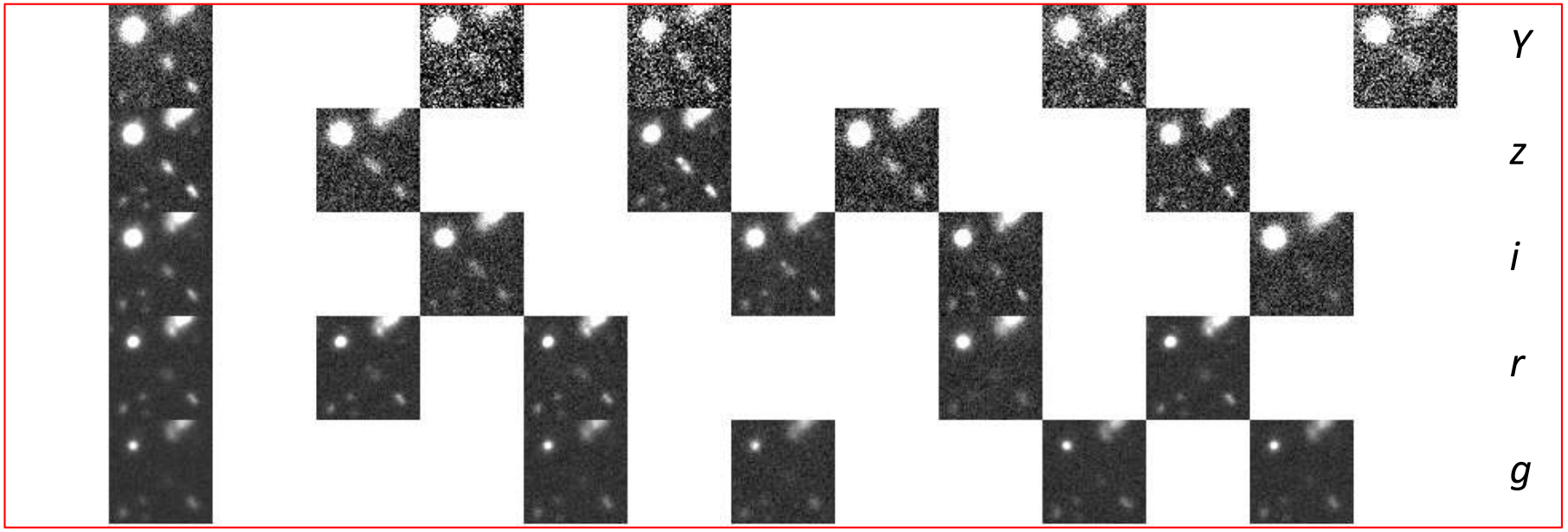


Spectroscopy needs many more photons (time!).

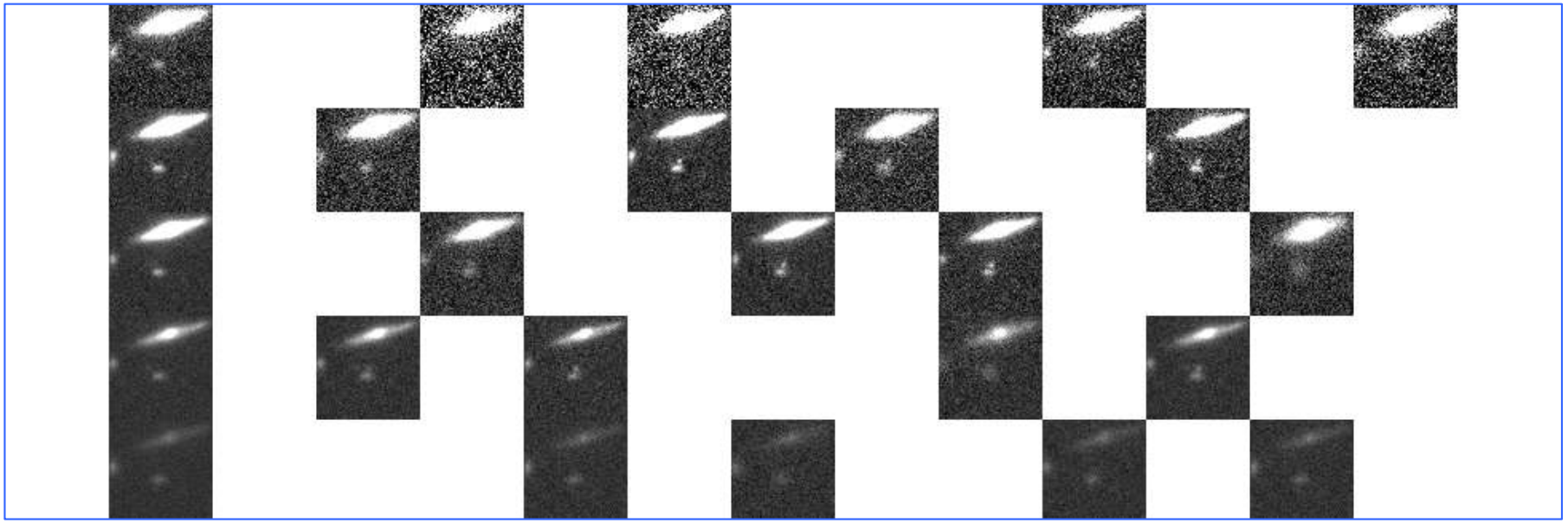
It is an accurate method, but very expensive.

Classify using only imaging data

Type Ia supernova, z (distance) 1.41

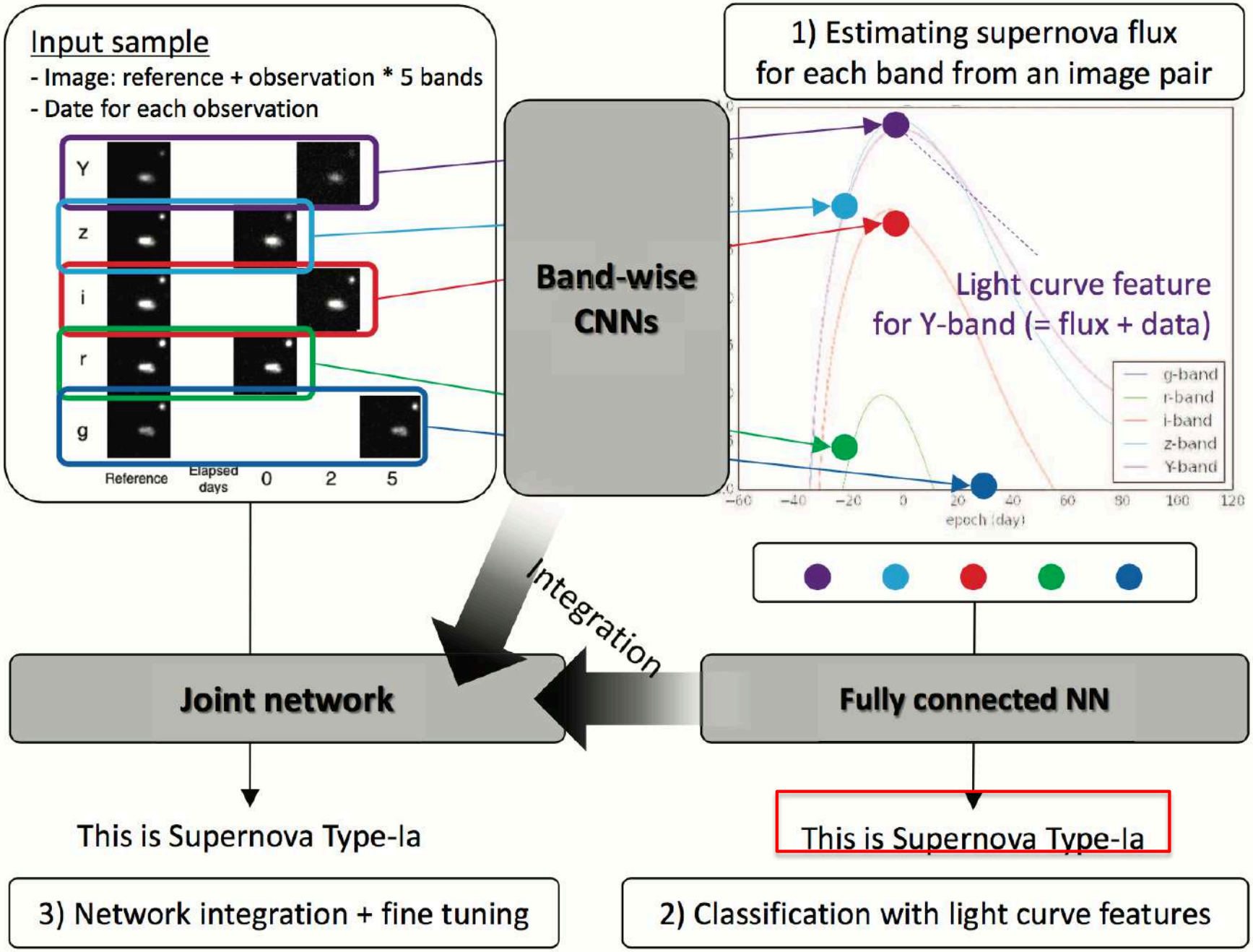


Type II supernova, z (distance) 0.87



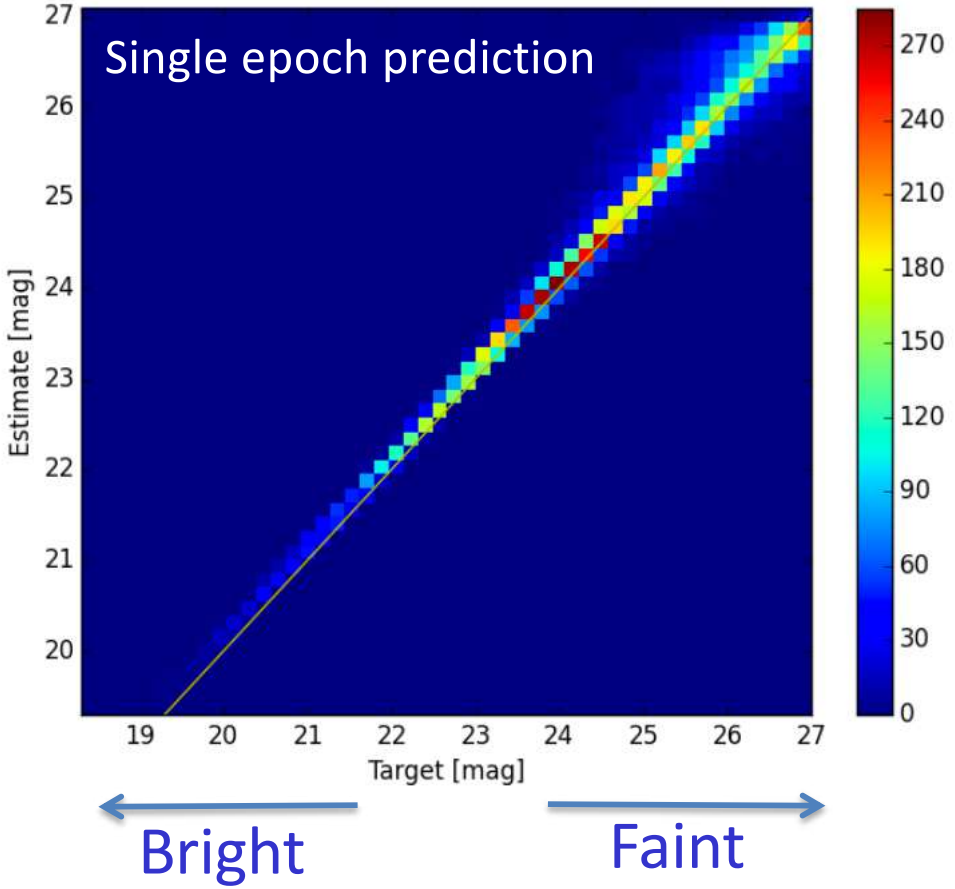
Classification by Convolutional NN

A. Kimura et al. arXiv:1711.11526

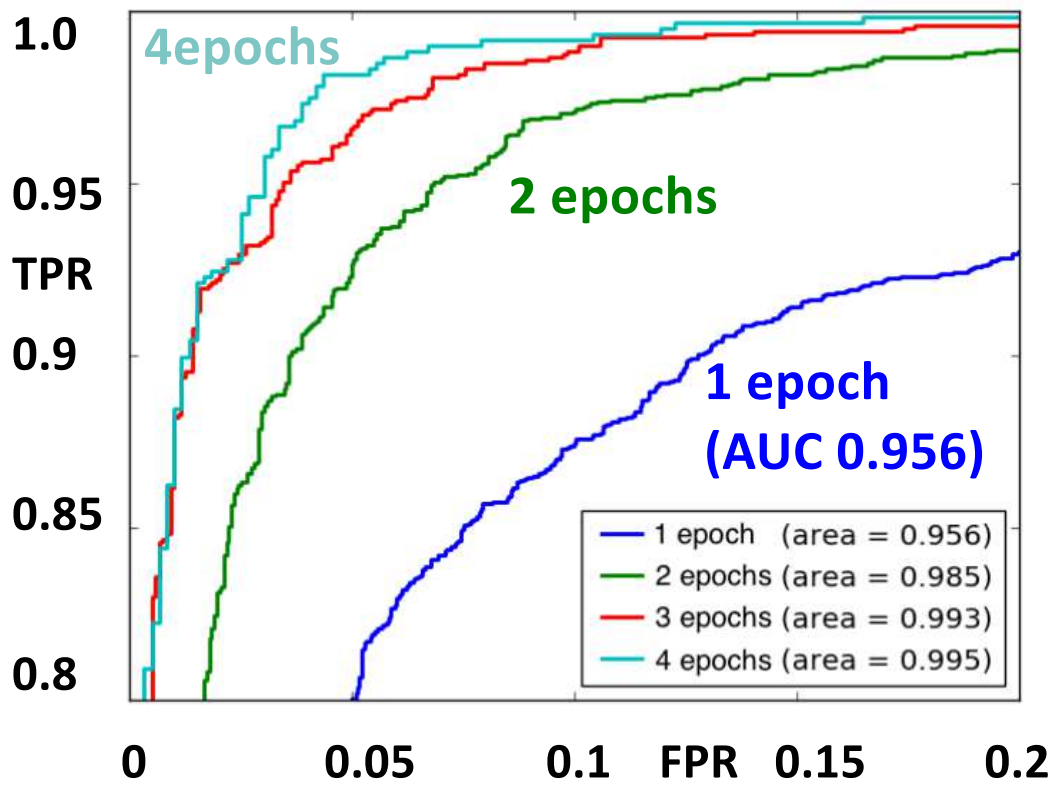


Binary classification result (Type Ia)

Magnitude comparison

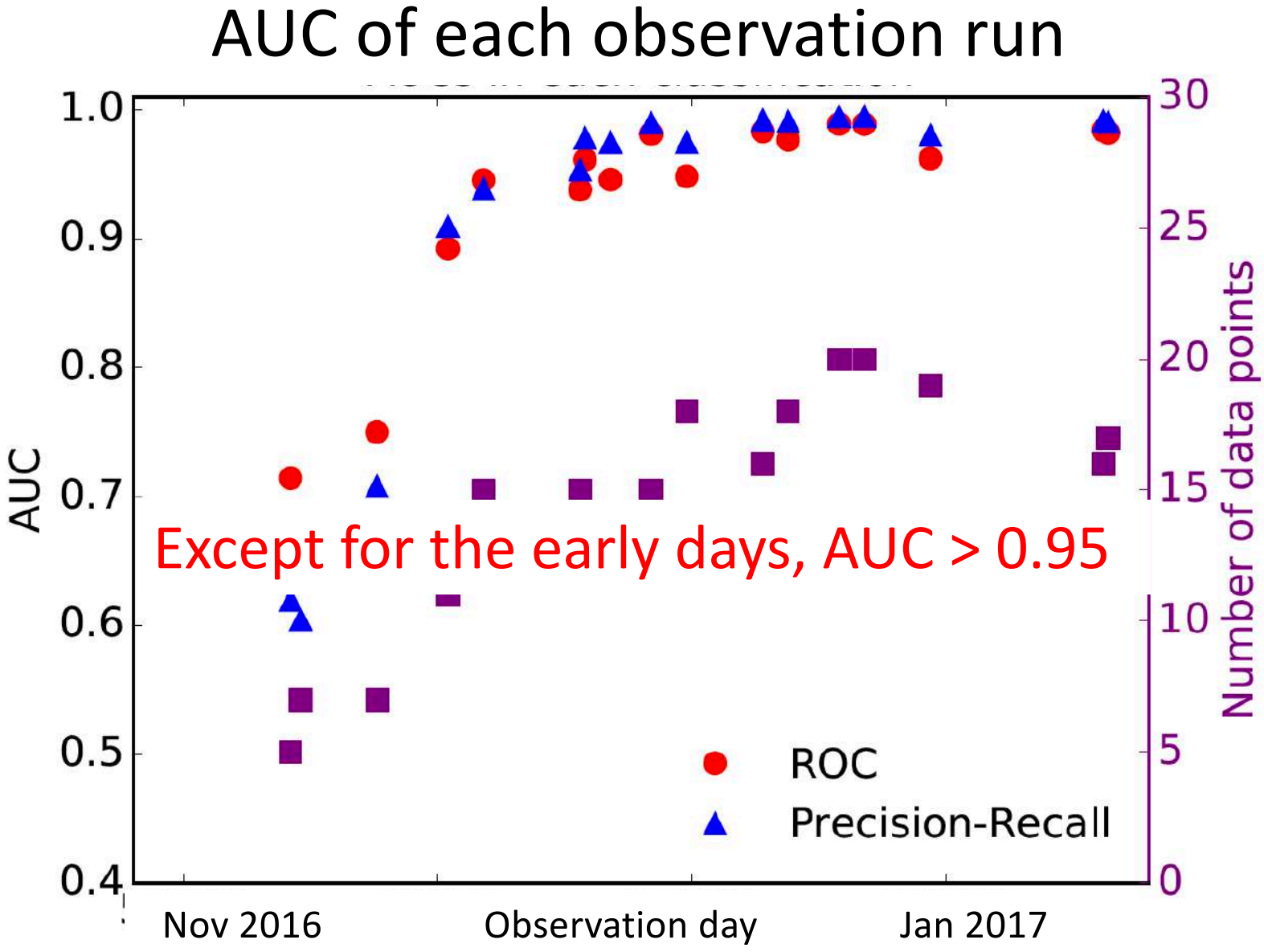


ROC Curve for type selection



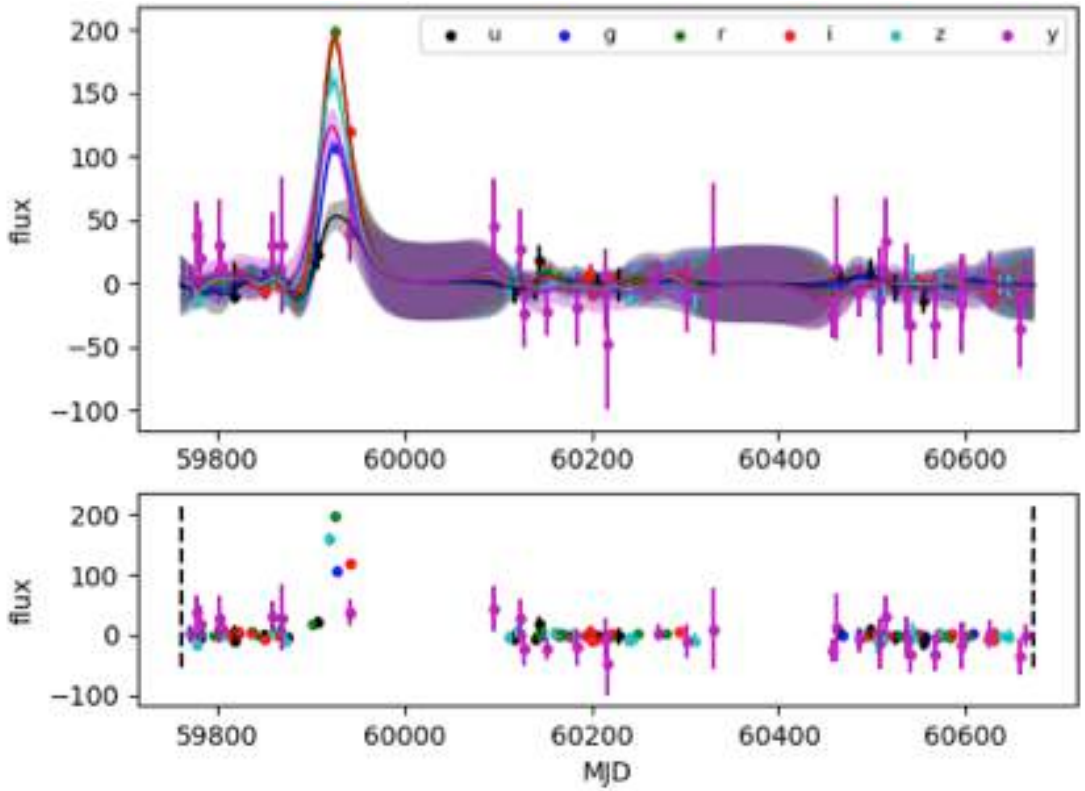
A. Kimura et al. arXiv:1711.11526

Performance with the actual data

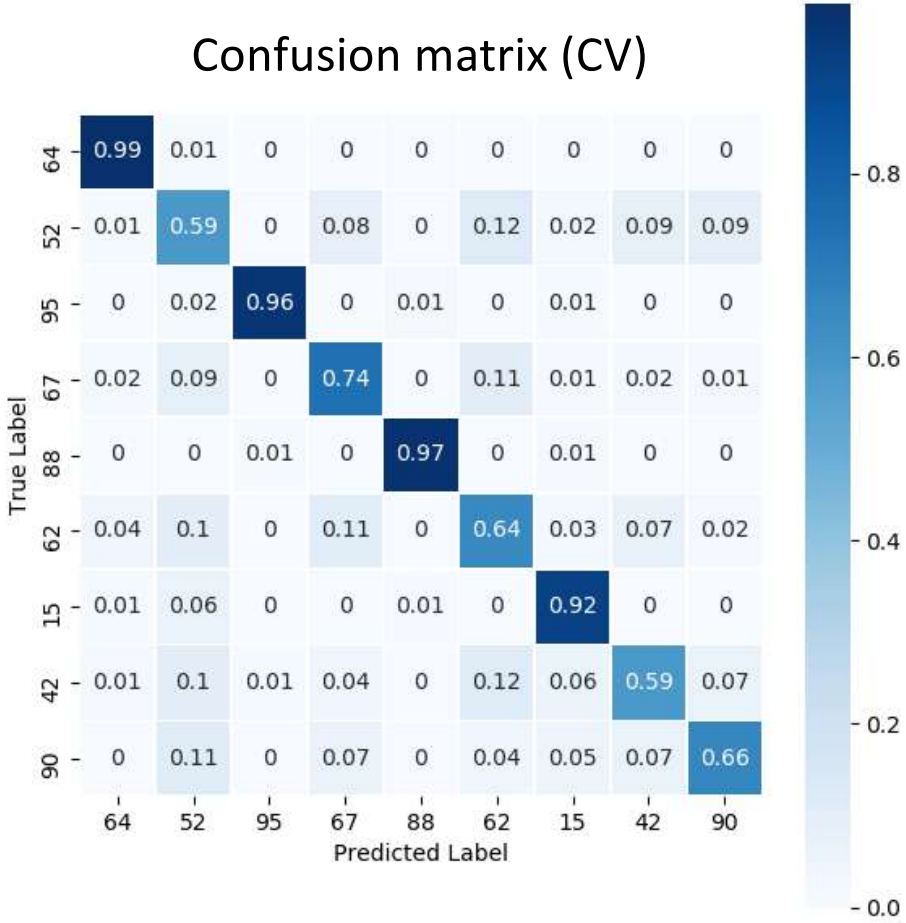


Classifier performance

Extract feature from Gaussian-process interpolated light curve data



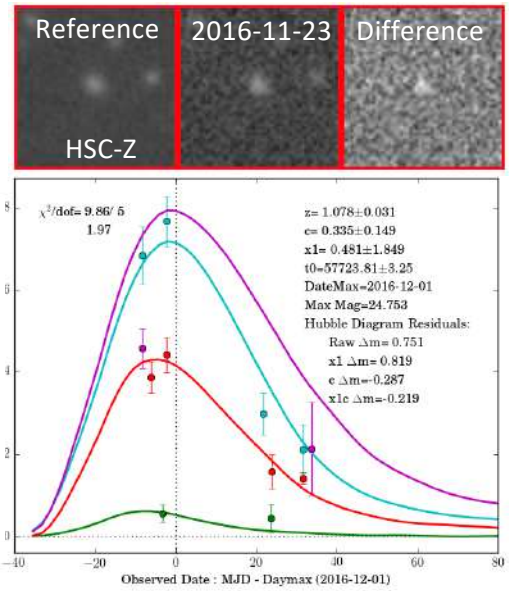
Confusion matrix (CV)



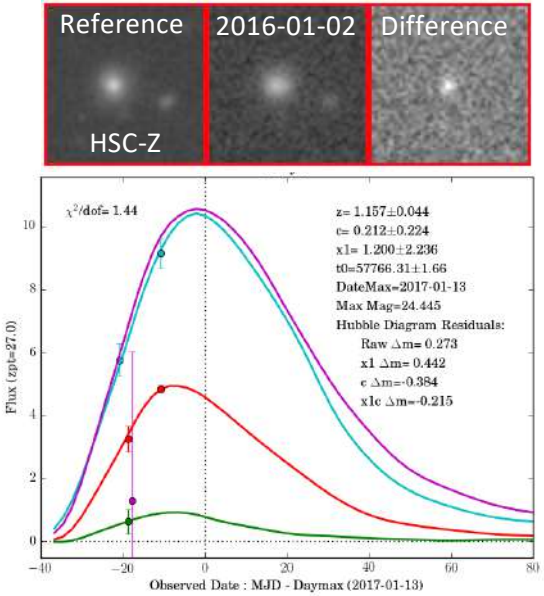
Selected for space telescope observation

Brightness

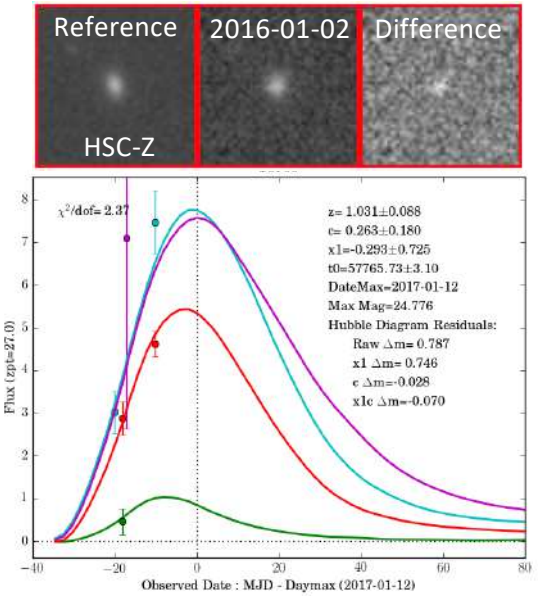
caa (Ia pred. : 0.93)



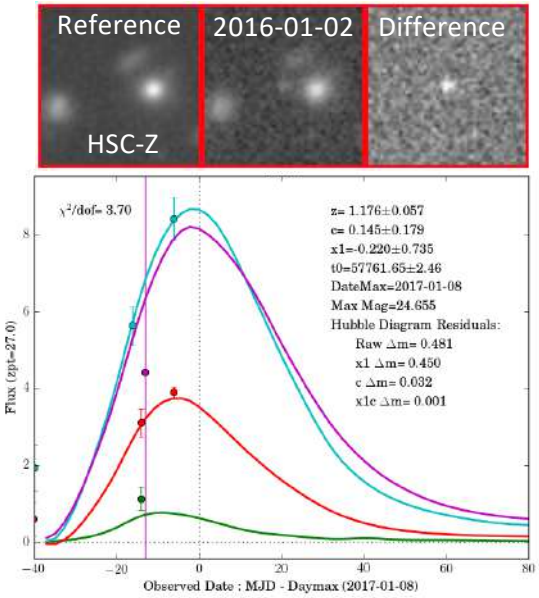
ryz (Ia pred. : 0.95)



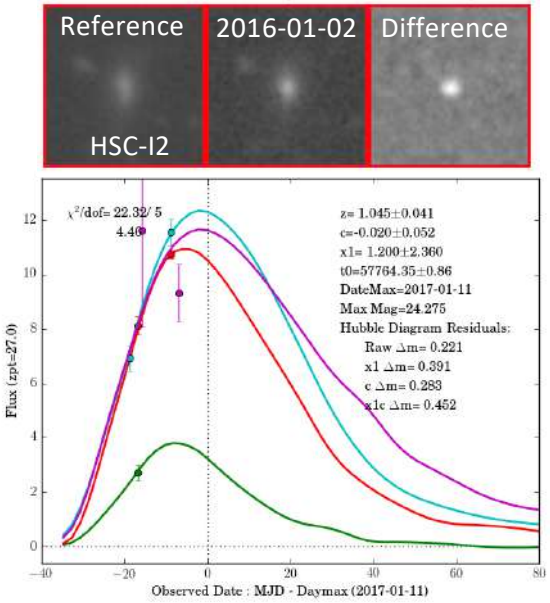
roo (Ia pred. : 0.85)



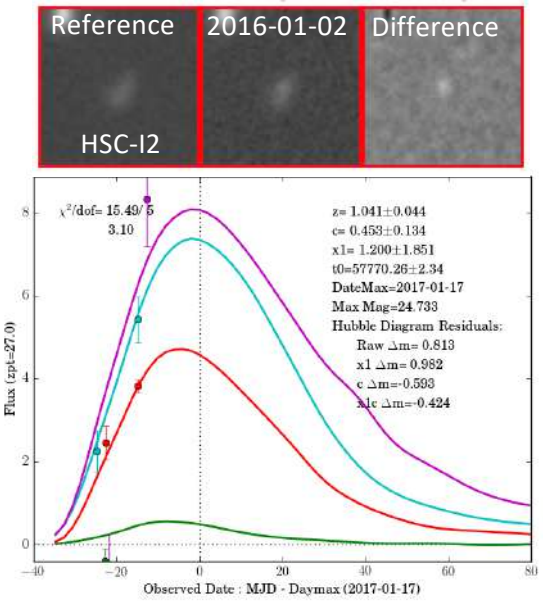
siv (Ia pred. : 0.94)



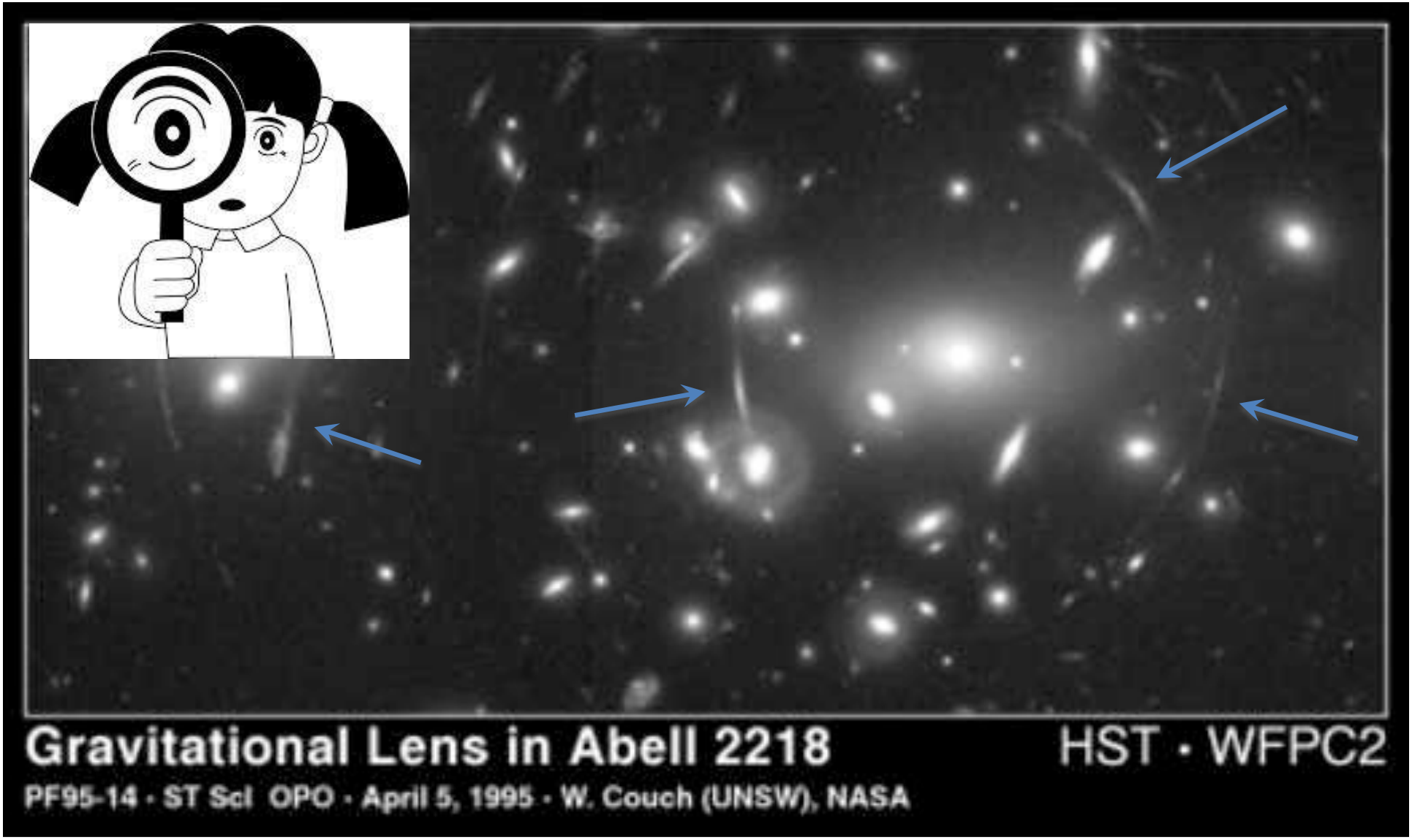
qgj (Ia pred. : 0.79)



vlr (Ia pred. : 0.89)

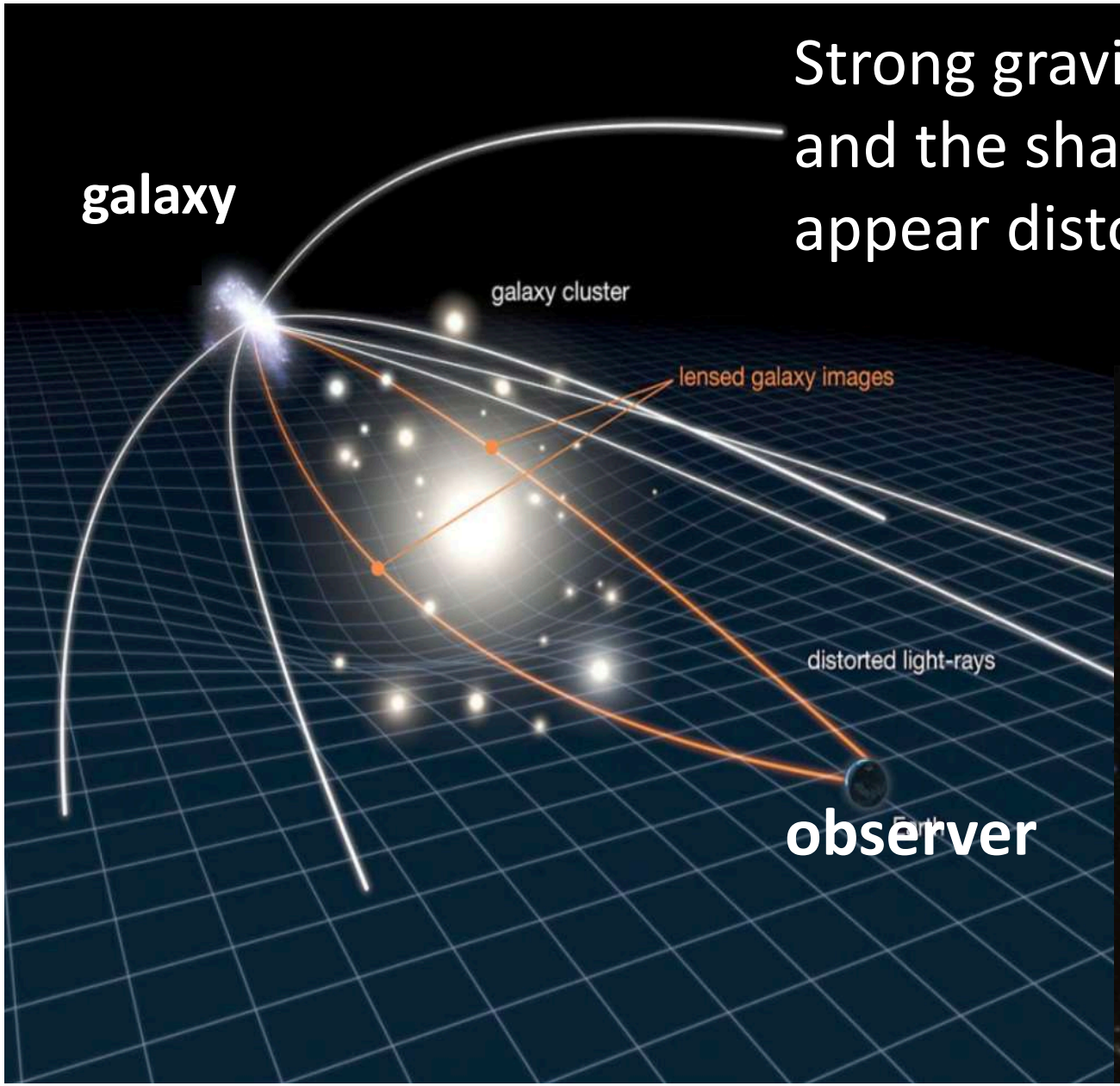


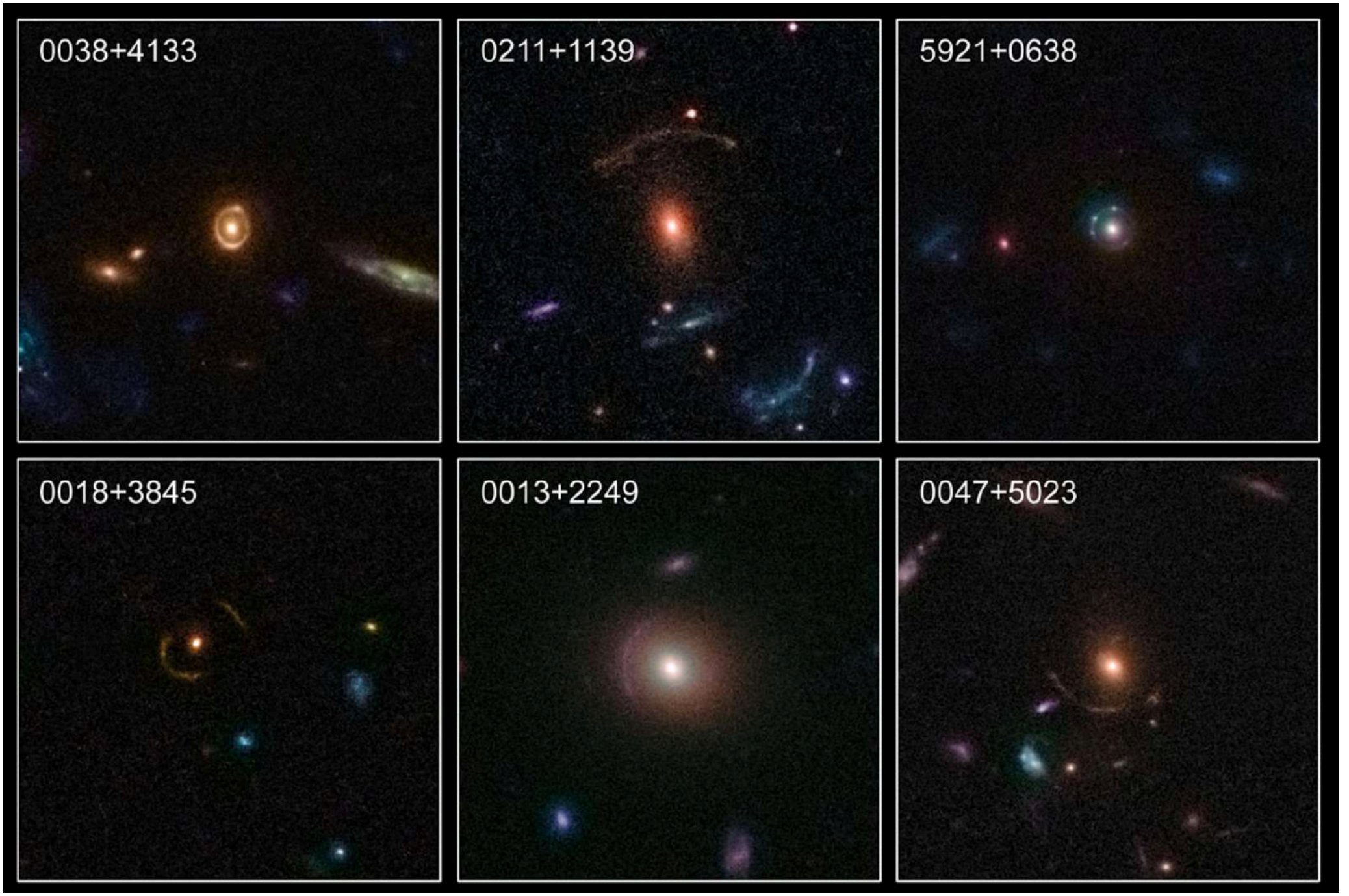
Cosmic Lens: Galaxy cluster Abell 2218



GRAVITATIONAL LENSING MAPS

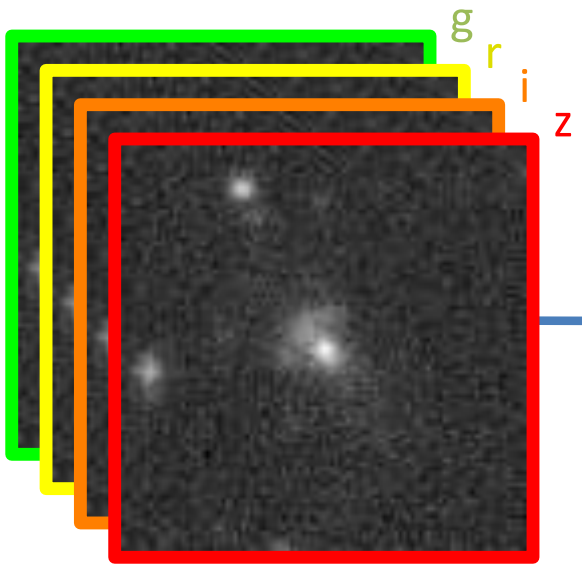
Strong gravity bends light-rays, and the shapes of distant galaxies appear distorted to us (observer).




0038+4133
0211+1139
5921+0638
0018+3845
0013+2249
0047+5023

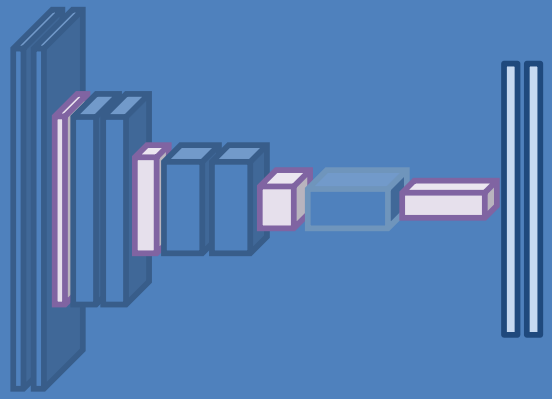
CNN “mass” calculator

4-band images



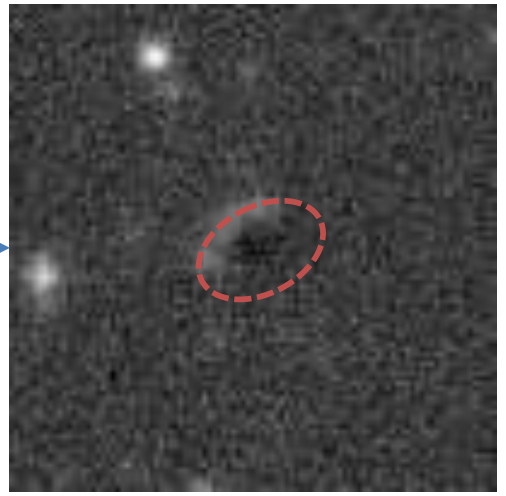
Model

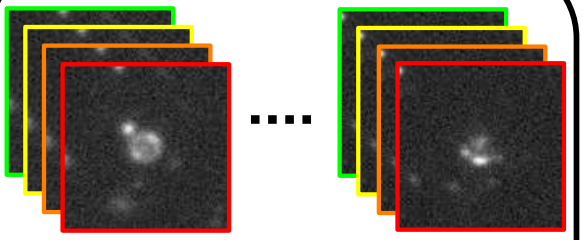
- Subtracting foreground galaxies
- Estimating gravitational lensing parameters



Lens parameters

Einstein radius
Ellipticity & angle

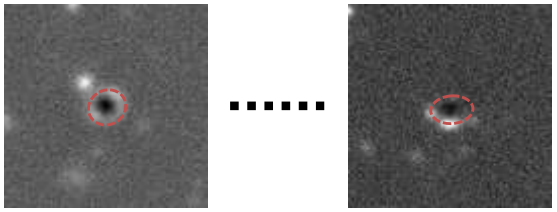




Training samples

350,000 images

Model training

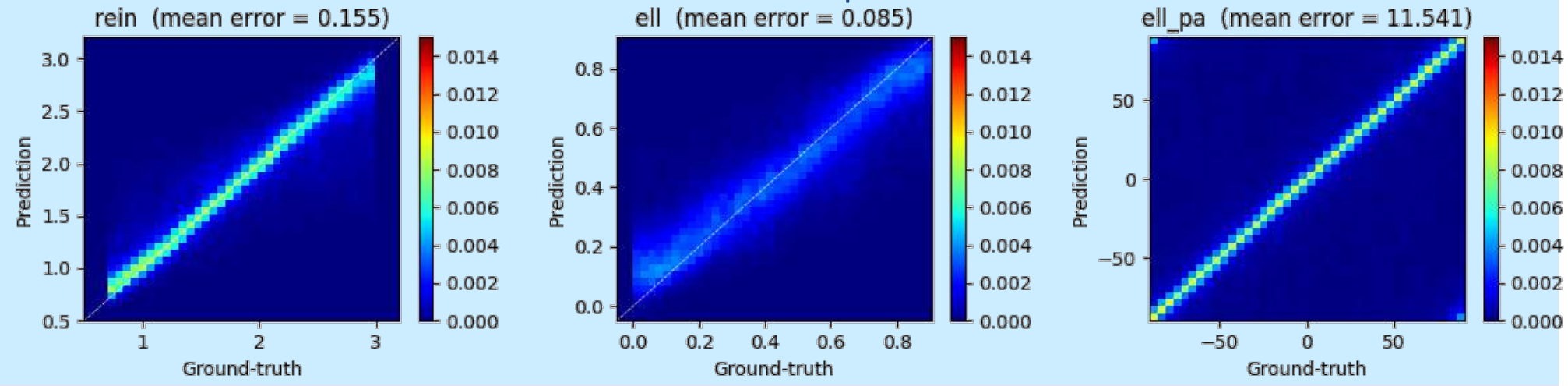


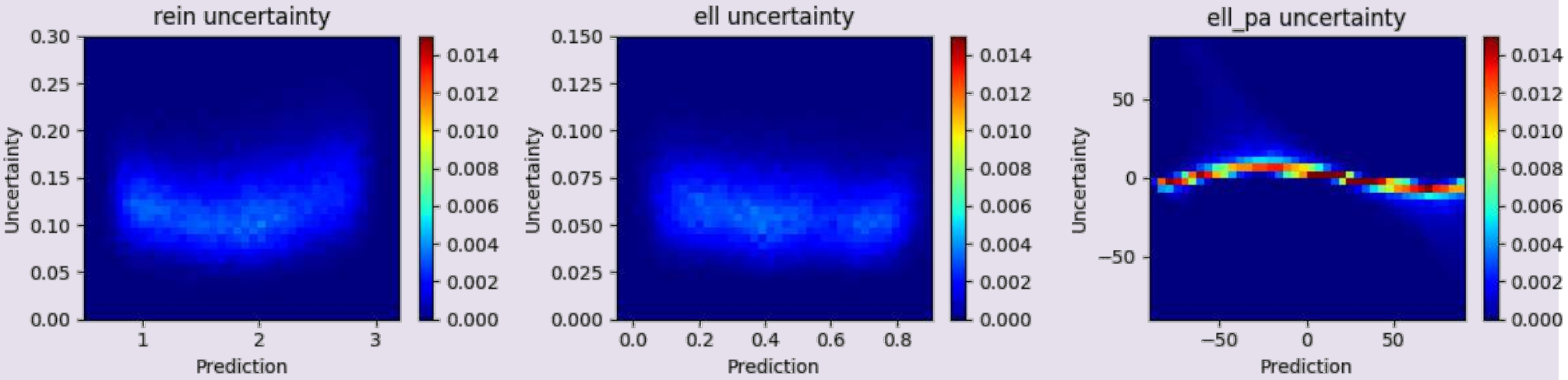
Training samples

350,000 parameter sets

Machine performance

Ground-truths vs mean predictions

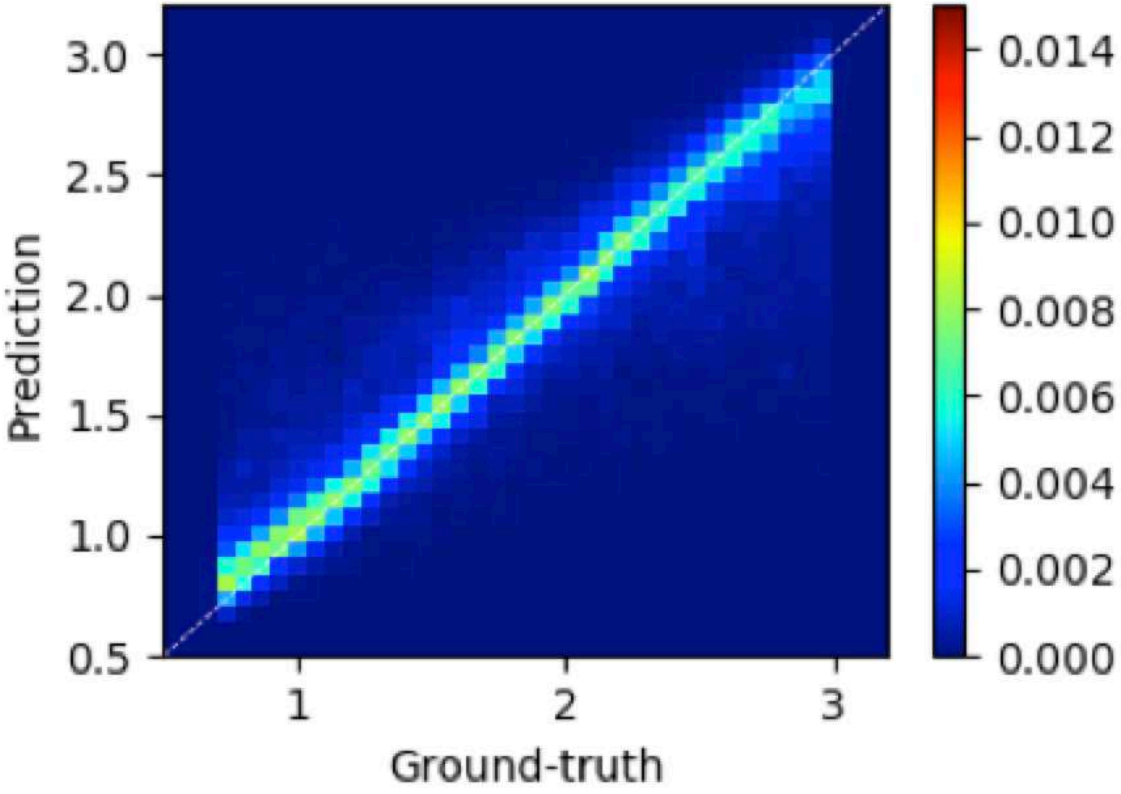




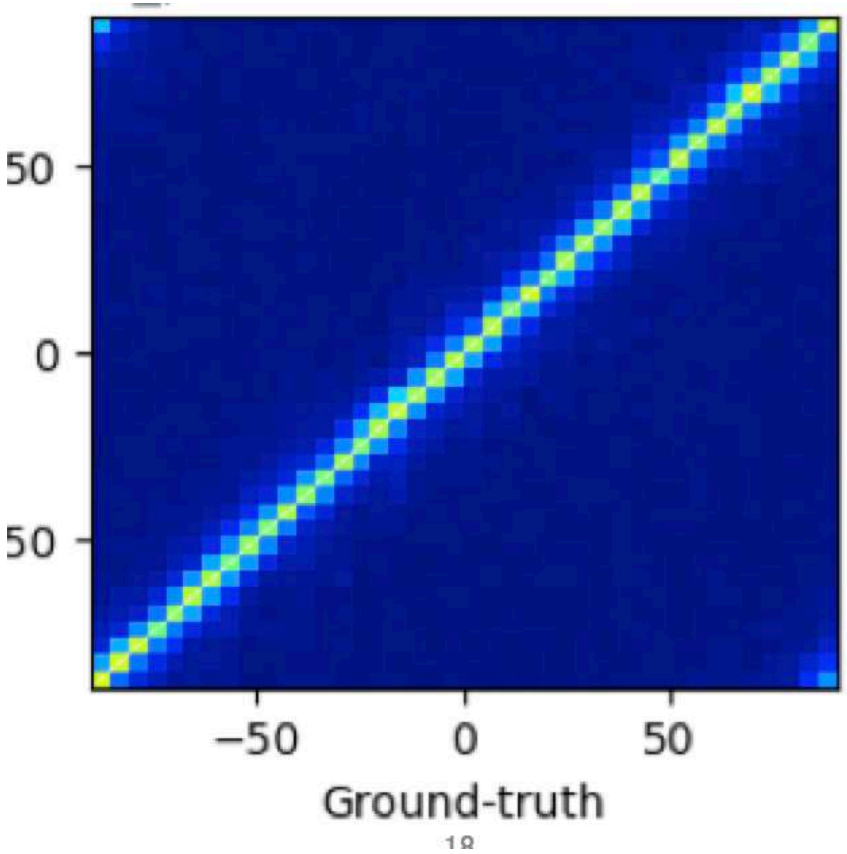
Mean predictions vs standard deviation of predictions

Machine performance

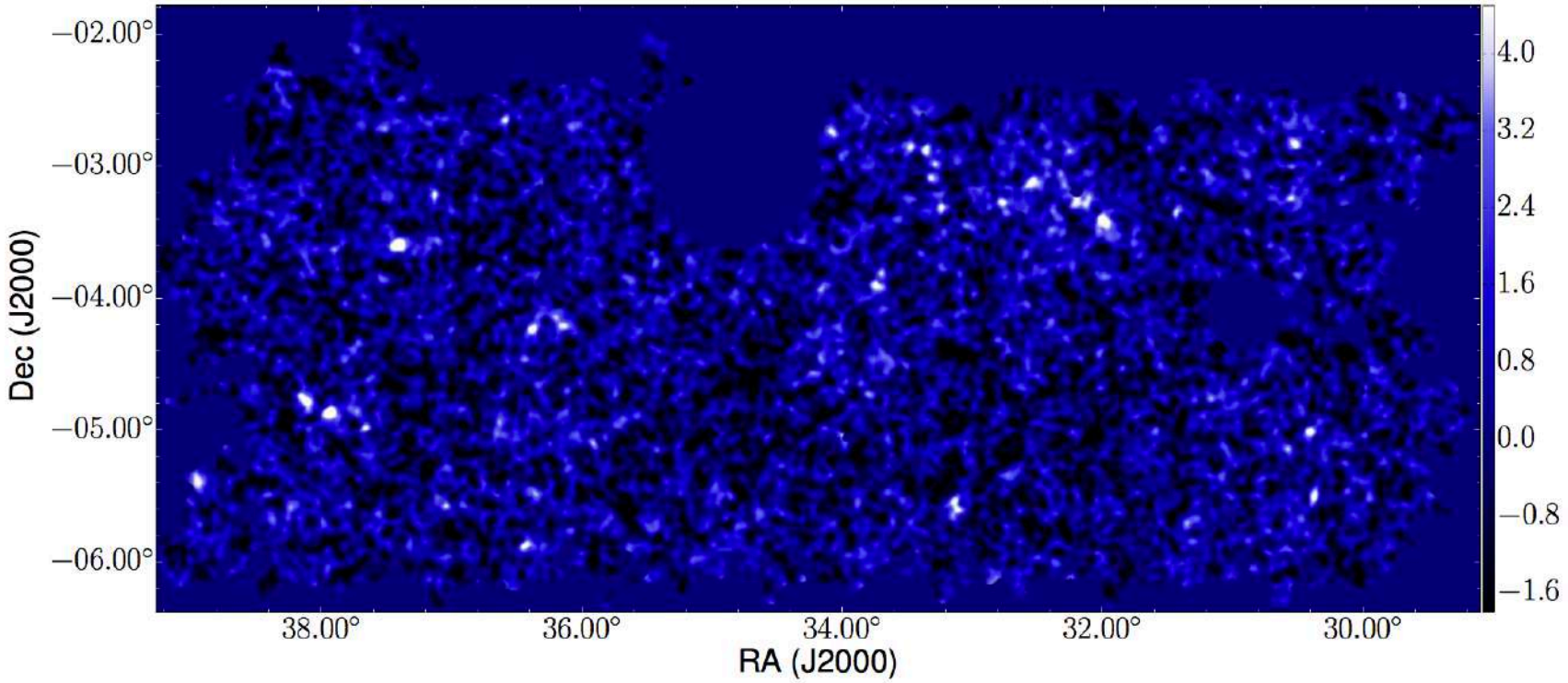
The mass of the central dark object

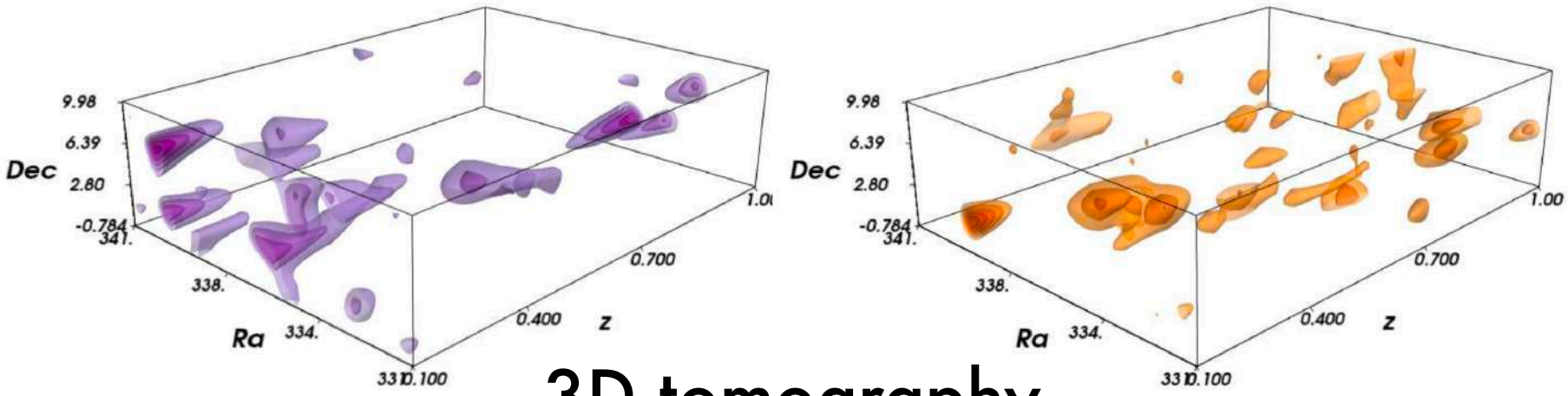


The ellipse orientation



DENSITY FIELD FROM HSC DATA

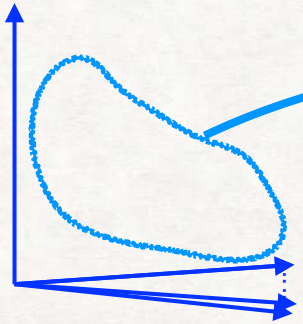




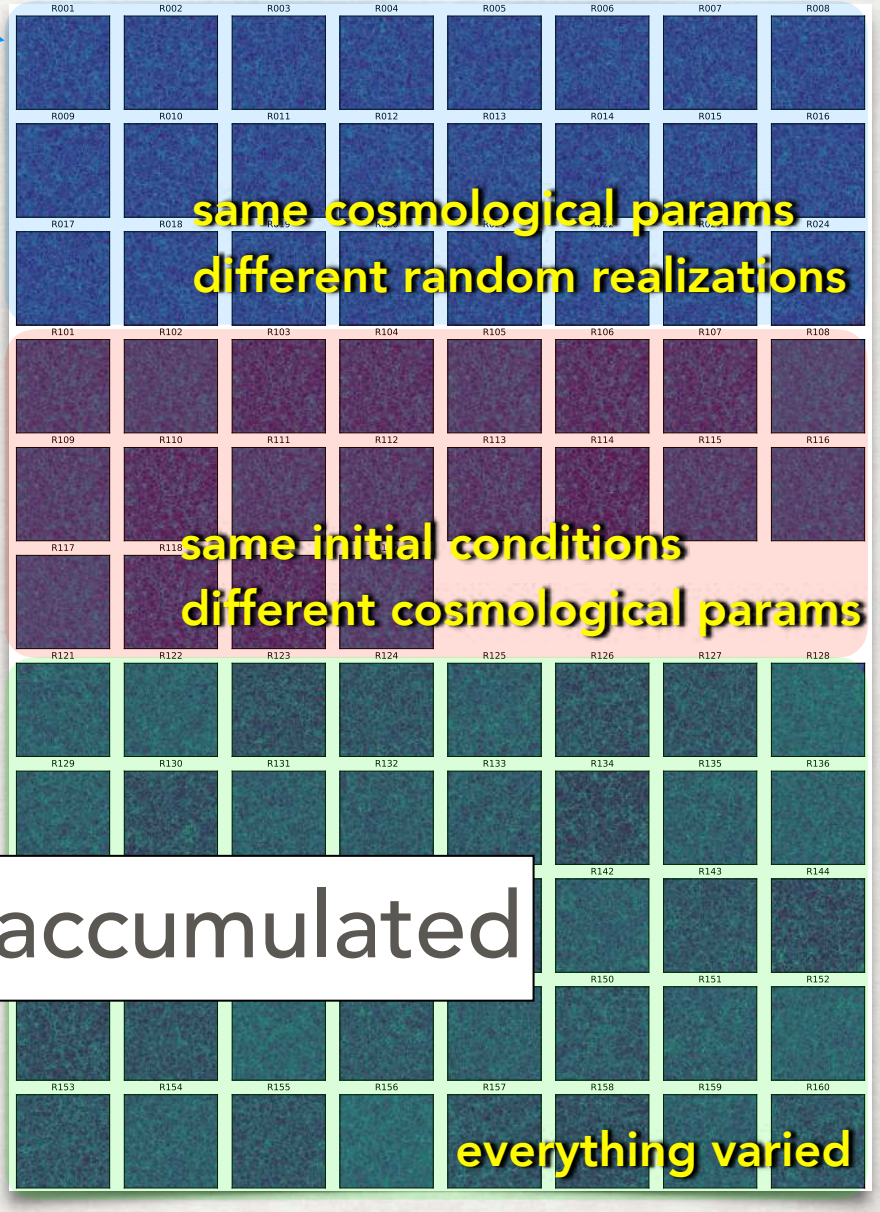
3D tomography

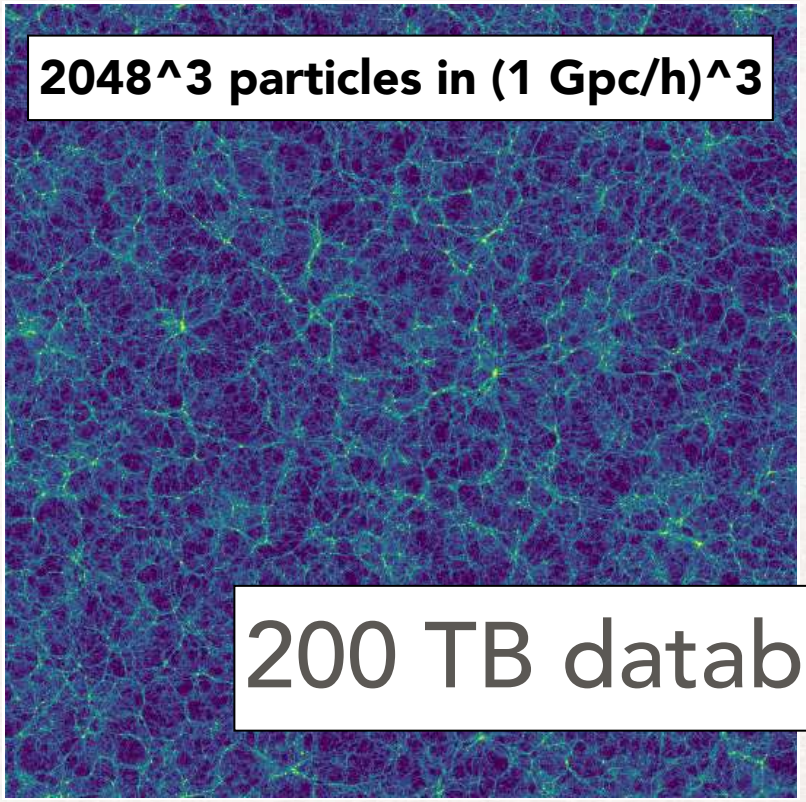
VIRTUAL UNIVERSES

SIMULATION ENSEMBLE IN 6-D PARAMETER SPACE



Efficient sampling scheme
to cover multi-D space





Density distribution over a billion lightyears
(color contour level represents the mass density)

200 TB database accumulated

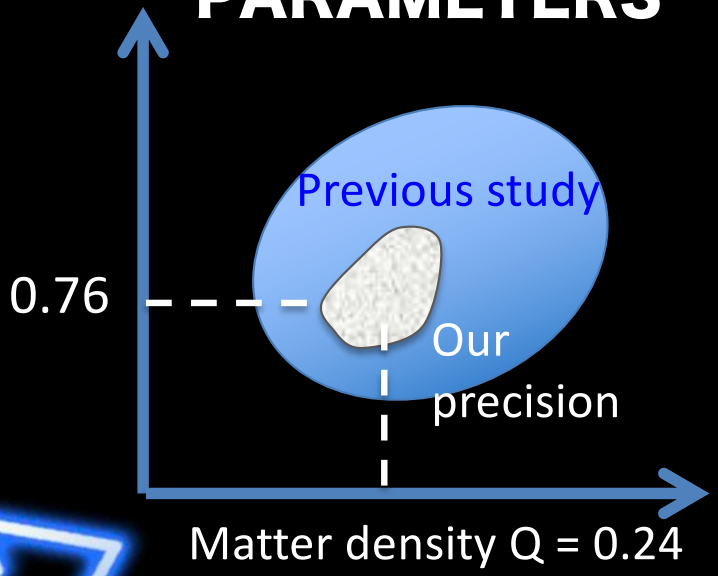
Simulation database (UTokyo)

Statistics (ISM&NTT)

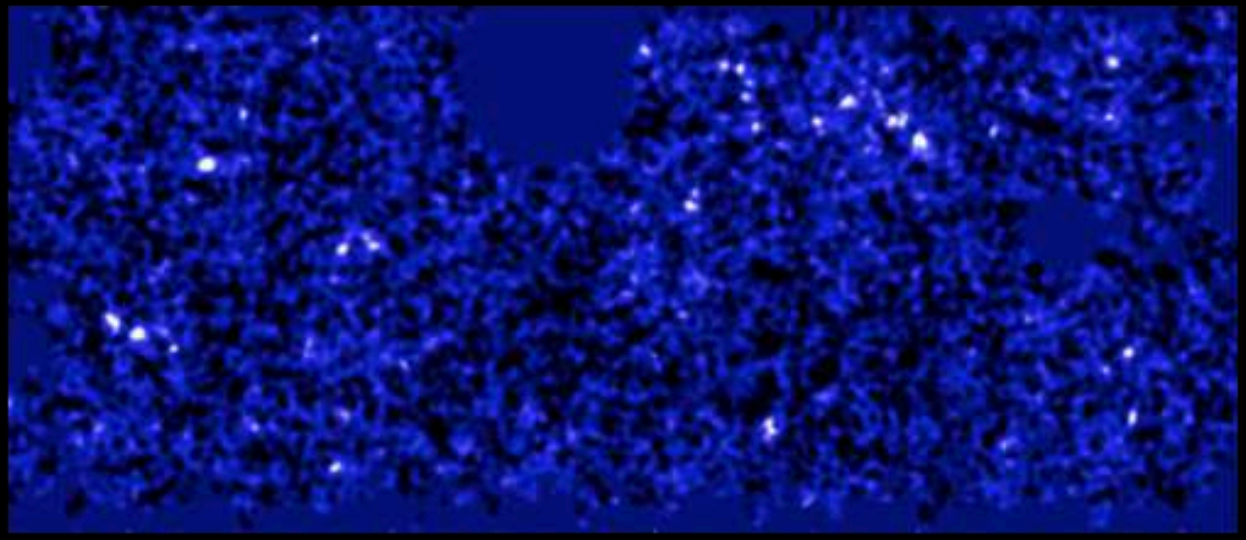
Parallel analysis tools (Tsukuba)

EMULATOR

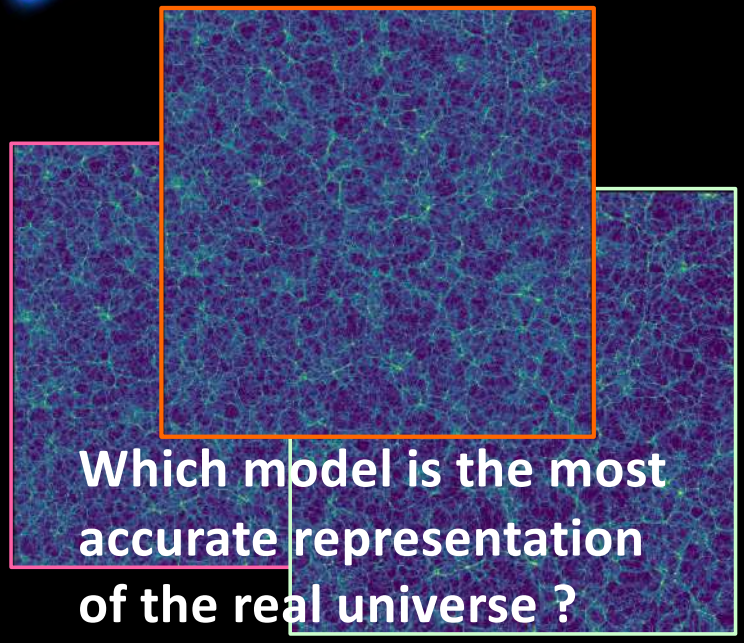
FUNDAMENTAL PARAMETERS



DARK MATTER MAP



Theory models

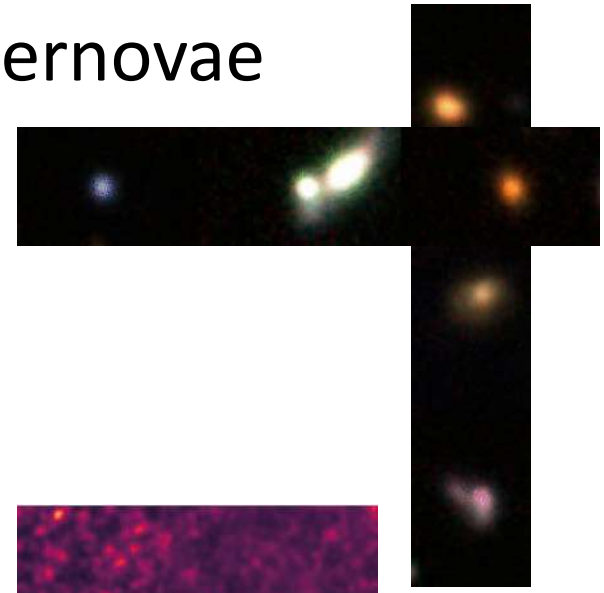


Which model is the most accurate representation of the real universe ?

Achievement summary

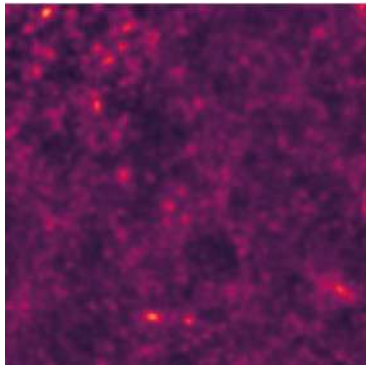
IMAGE ANALYSIS TOOL

1. 65000 variable objects and 1800 supernovae from the first 55 terabyte data
   - A record breaking rate of detection
   - All classified and web-catalogued



STATISTICAL METHOD

2. Extremely fast statistical tools for cosmic “map”
   - 2-days supercomputer simulation in 1 second “effectively”



DATA ANALYSIS TOOL

3. Parallel data reduction pipeline
   - One-night data processed in 4 hours!
   - Simulation data analysis 30 times faster