Galaxy survey data reduction with deep learning

ACAT 2022

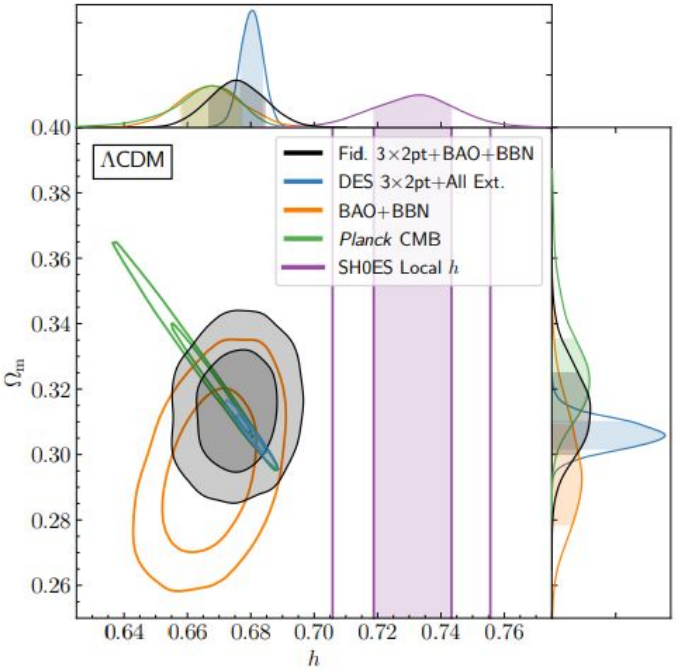
26/10/2022

Martin B. Eriksen

Laura Cabayol García

Cosmology

- What is causing the accelerated expansion of the universe?
- What is dark matter and dark energy?
- Plot: Cosmological analysis of DES Y3 combined with Planck.



Dark matter fraction

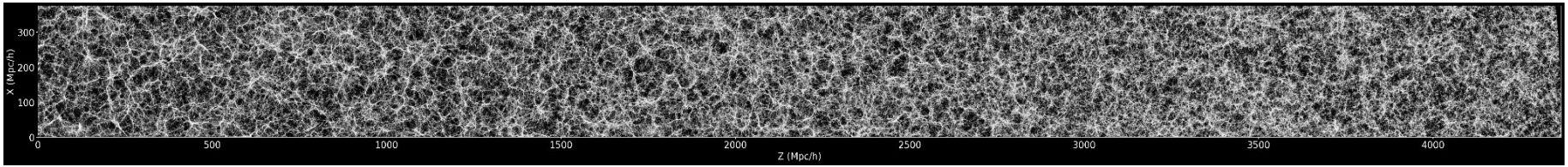
Abbot et.al. 2021

Expansion speed

Large scale structure

Several cosmological probes use the large scale structure of the universe. Under is the slice of the Flagship 2 galaxy simulations (3.6 billion galaxies in total)

Space



Redshift (z) / time

Today

Early Universe

Credit: J.Carretero

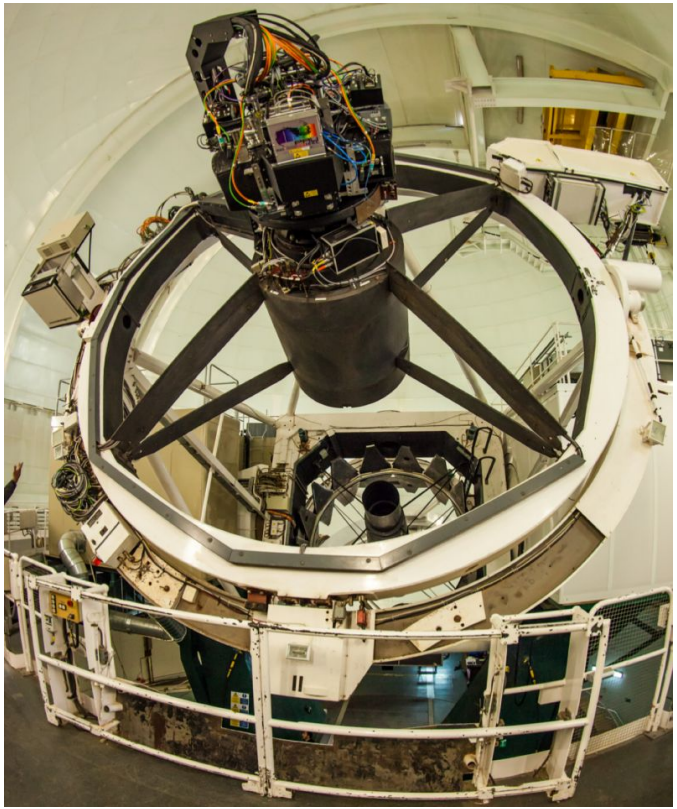
Galaxy imaging surveys

- DES and Kids, current generation surveys.
- LSST and Euclid, soon to start.
- LSST will observe 20 billion galaxies and produce 20TB of raw data per night.

Euclid satellite, Credit: ESA

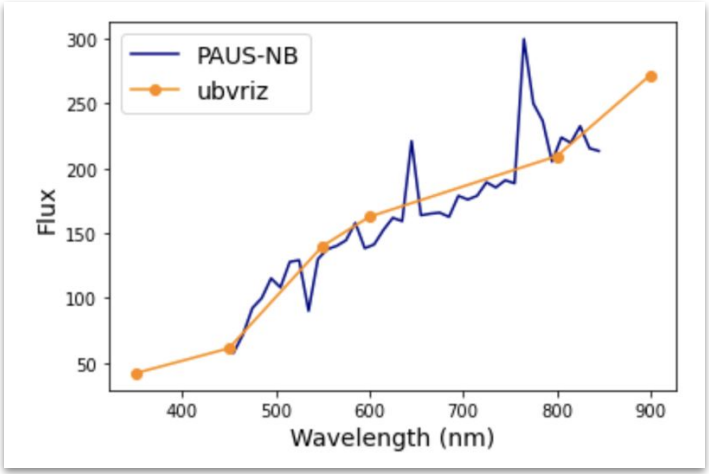
PAU survey

- Imaging survey with a 40 narrow band photometric filters camera (PAUCam) (Padilla et al 2019).
- The camera is installed in the 4.2m - William Herschel Telescope, in La Palma.
- Covers a wavelength range from 450nm to 850nm.
- It effectively measures high resolution photometric spectra (R ~ 50).



PAUS photo-z with classical algorithms

- Photoz is the measure of shift in wavelength from the expanding Universe.
- PAUS reaches a photo-*z* **precision of 0.0037(1+*z*) with a template-fitting algorithm** for the best selected 50% of the sample with i_{AB}<22.5 (Eriksen et al 2019).
- About 14 times better compared with broad bands

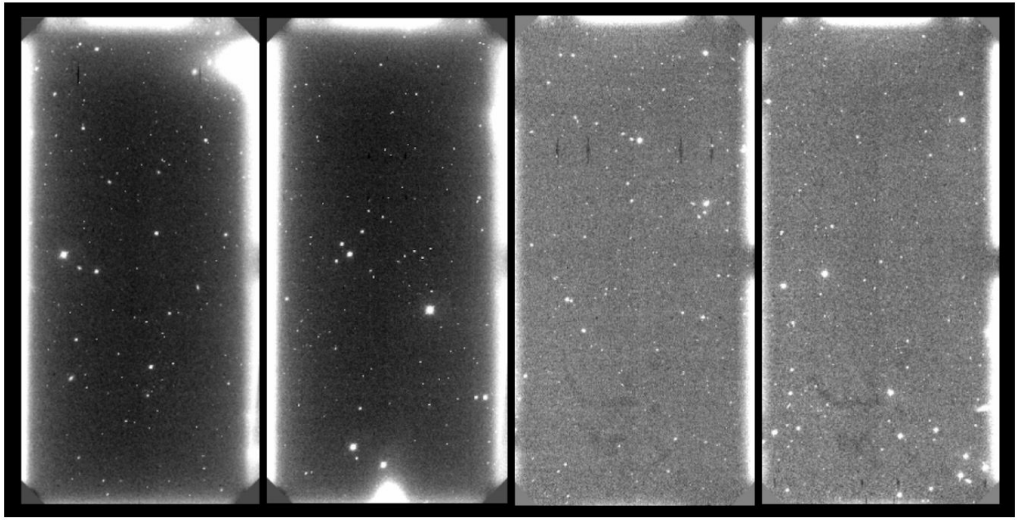


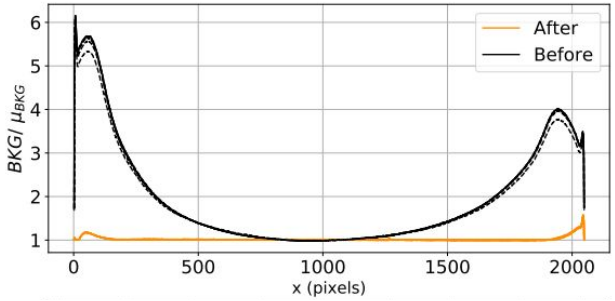
Example galaxy

PAUS Scattered light pattern

Before intervention

After intervention



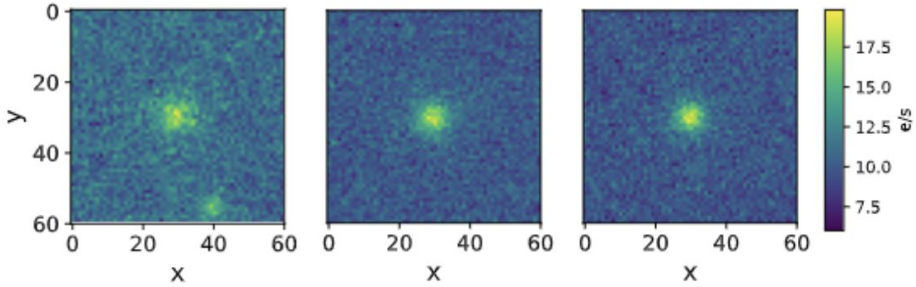


Cabayol, Eriksen, et al. 2020, MNRAS, Volume 491, Issue 4, p.5392-5405

Optimized background estimation using a CNN.

Galaxy photometry - Training sample

Instead of only predicting the background, could we predict the background subtracted signal?

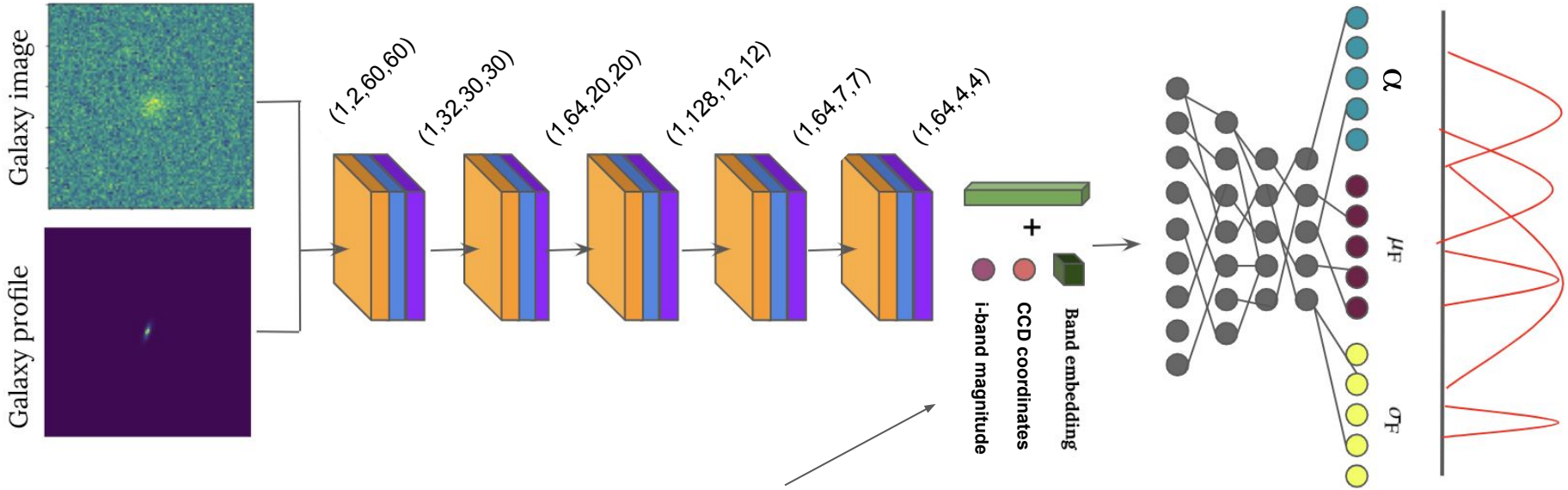


These are **60x60 bulge+disk galaxy cutouts** mimicking the PAUS photometry distributions drawn with astropy in a high resolution grid (600x600)

The galaxies are stacked on 60x60 PAUCam cutouts randomly selected that simulate the background noise and include artifacts

Cabayol, Eriksen, et al. 2021, MNRAS, Volume 506, Issue 3, pp.4048-4069

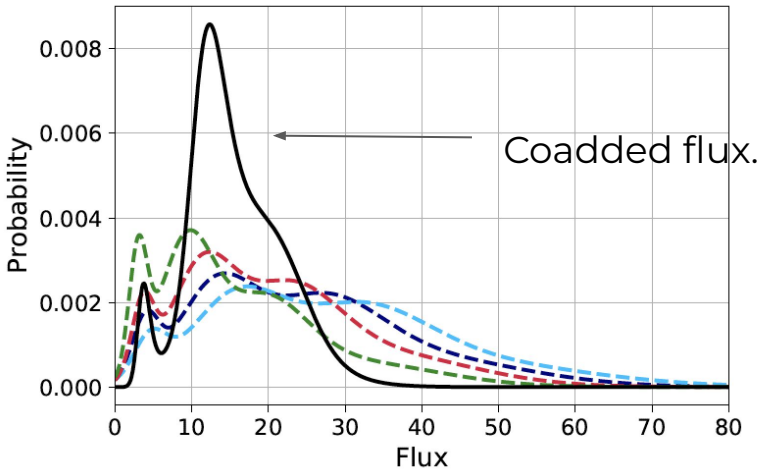
Galaxy photometry - Network (Lumus)



Additional “physics” information

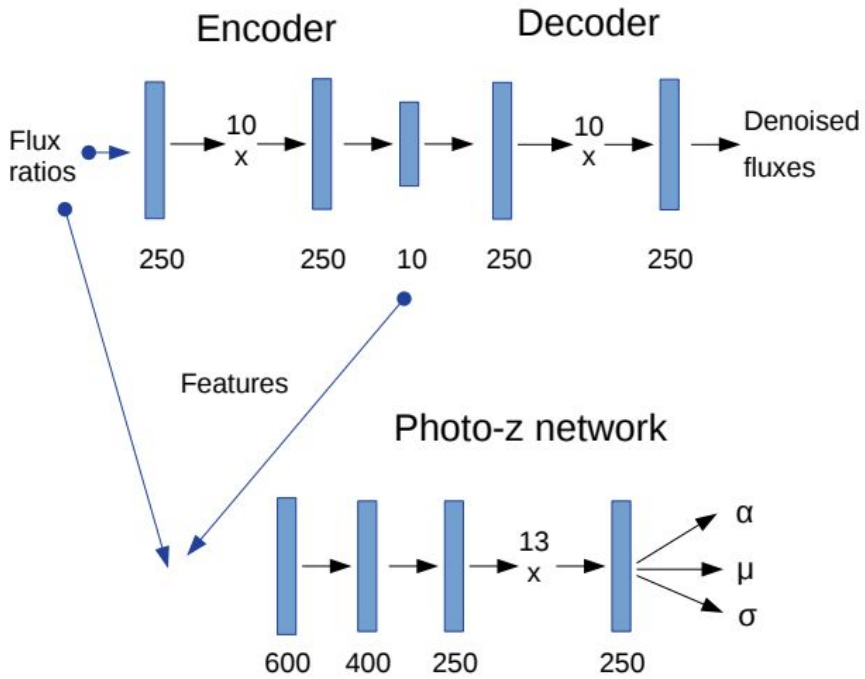
Galaxy photometry - Combining exposures

- PAUS has taken at least five observation per galaxy per band.
- Combining the measurements is called “coadding”.
- Instead of combining individual flux measurements, Lumos combines the probability distribution of the measurements.
- Better results than combining the peaks, which is more typical.



Galaxy distance determination (Deepz)

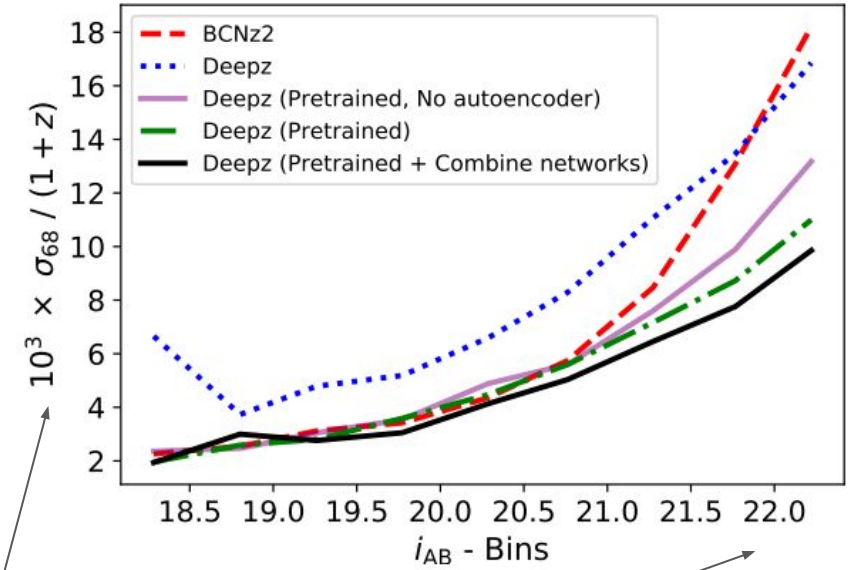
- 40 PAUS narrow bands + 6 broad bands fluxes as input.
- Also create (10) features using an auto-encoder.
- Estimate the p(z) using a mixture density network.



Eriksen et al. 2020, MNRAS, Volume 497, Issue 4, pp.4565-4579

Network - Impact of different parts

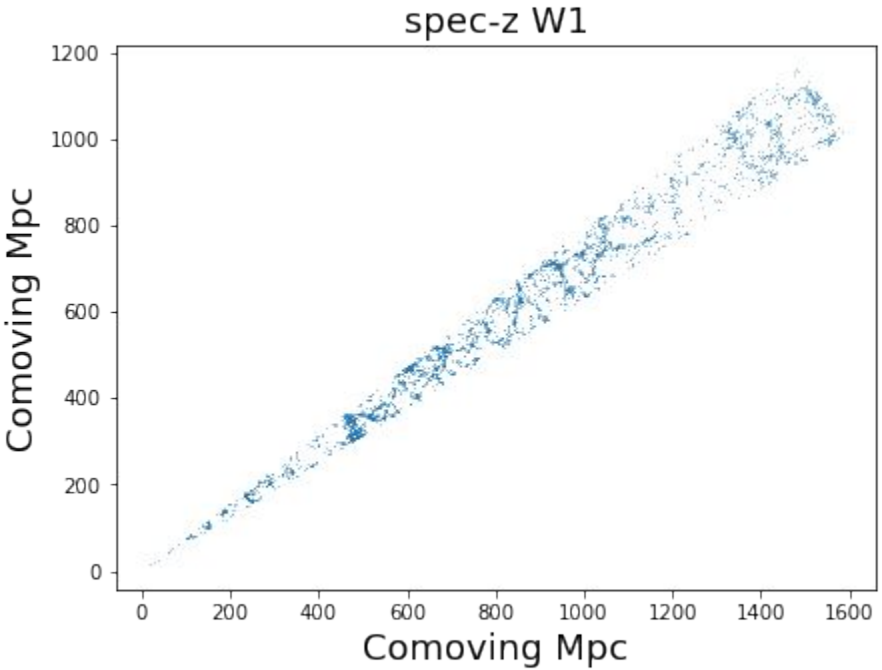
- BCNZ2 - PAUS template fitting code, eriksen 2019
- Deepz with no pretraining has worse results (blue).
- Pretraining has a large effect on the precision (green)
- Moderate effect of an auto-encoder (green vs purple)
- Combining multiple networks helps

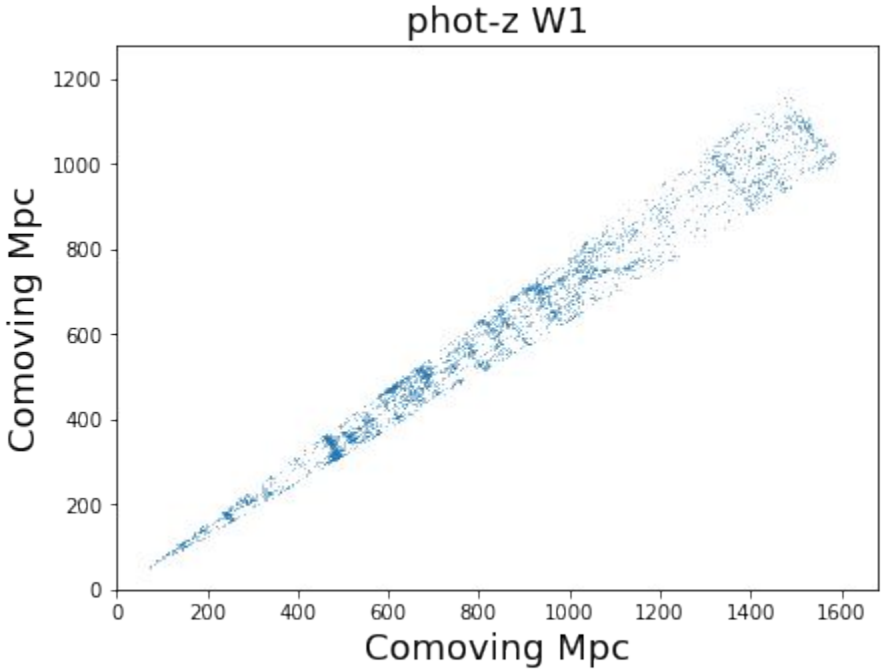


Precision on distance determination

Fainter galaxy

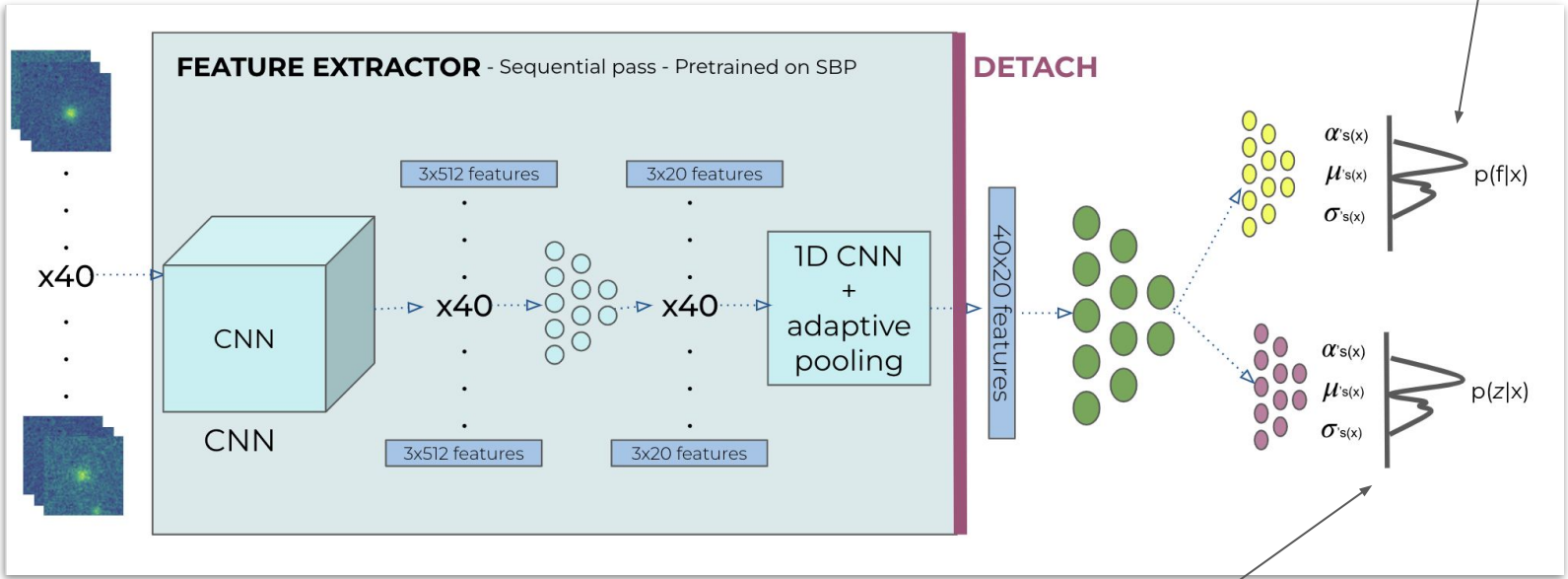
Resolving LSS with PAUS





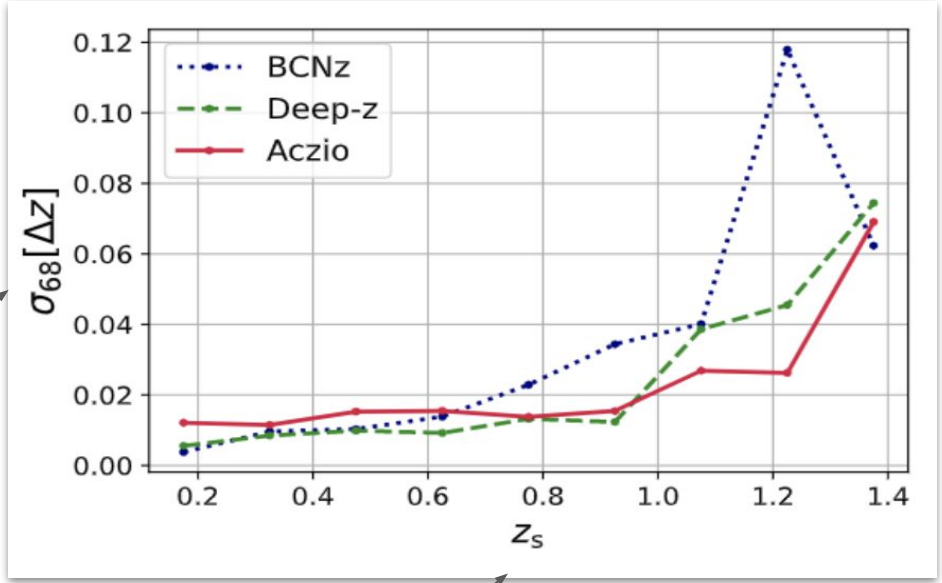
Credit: Ingrid Vanessa Daza

End-to-end pipeline (Aczio)



Aczio redshift determination

- Better results with Aczio for more distant galaxies, which are the challenging ones.



Precision of distance determination

Redshift

Conclusions

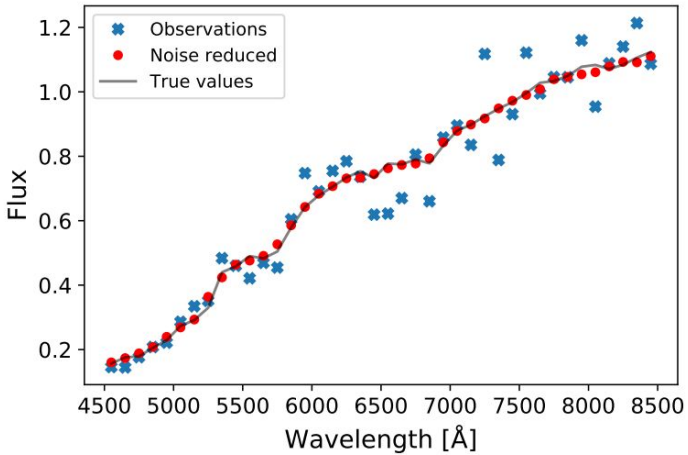
- Imaging surveys are an essential part of modern astronomy.
- Developed end-to-end deep learning network for predicting both the galaxy fluxes and distrances.
- Software available on Github under the GPL-3.0 license.

BCNz: https://github.com/PAU-survey/bcnz
BKGnet: https://github.com/PAU-survey/bkgnet
Lumus: https://github.com/PAU-survey/lumos
Deepz: https://github.com/PAU-survey/deepz

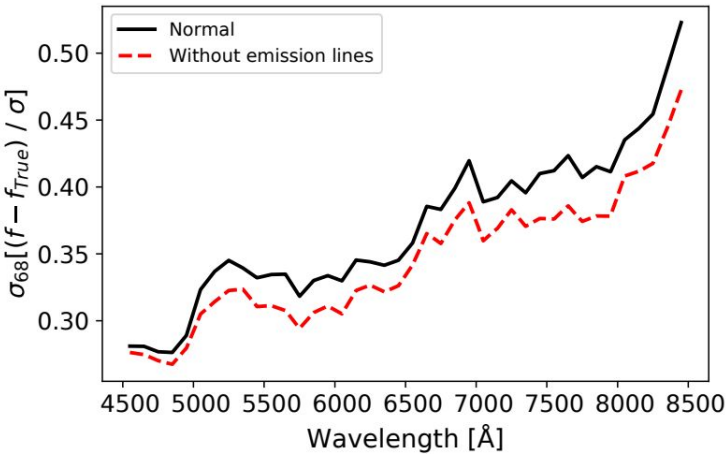
Additional slides

Auto-encoder properties

Test on a single elliptical galaxy template, SNR=10 in NB, SNR = 35 in BB



Noise reduction on a simulation (FSPS).



MULTI-BAND PHOTOMETRY

- Photometry band-by-band. The network only uses information from a given image to predict its photometry
- Multi-band photometry. The network only uses information from all bands to predict its photometry in a single band

Using data from all bands have several potential benefits:

1. Full-SED information
2. Morphology from several bands
3. Easier detection of spurious observations