Distant galaxies analysis with Deep Neural Networks

Raúl Cacho Martínez

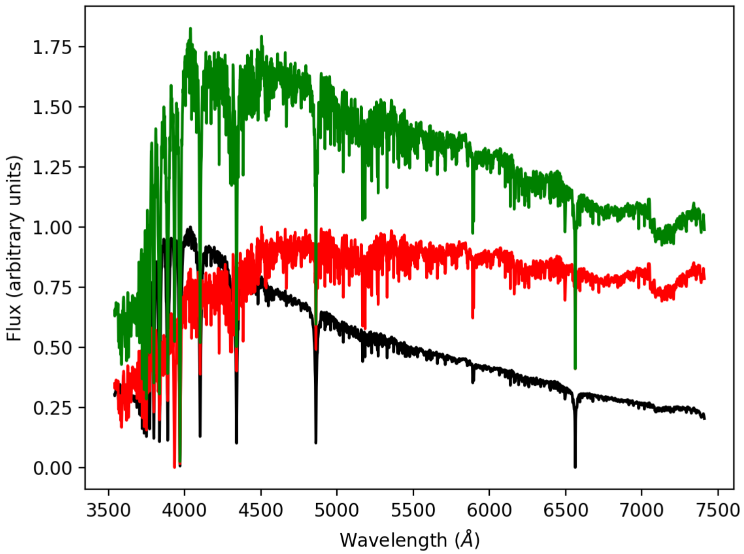
Máster Universitario en Ciencia de Datos
Supervisor: Anna Bosch Rué

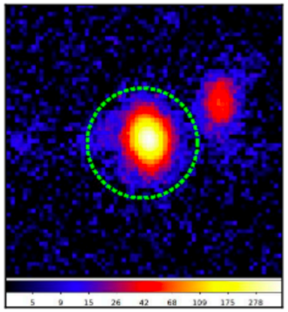
Contents

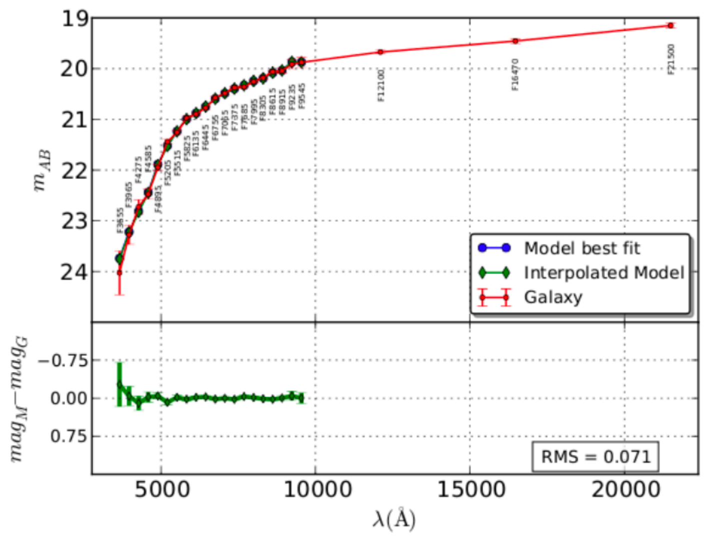
1. Introduction

1. Introduction: Description

- Spectral Synthesis and SED fitting



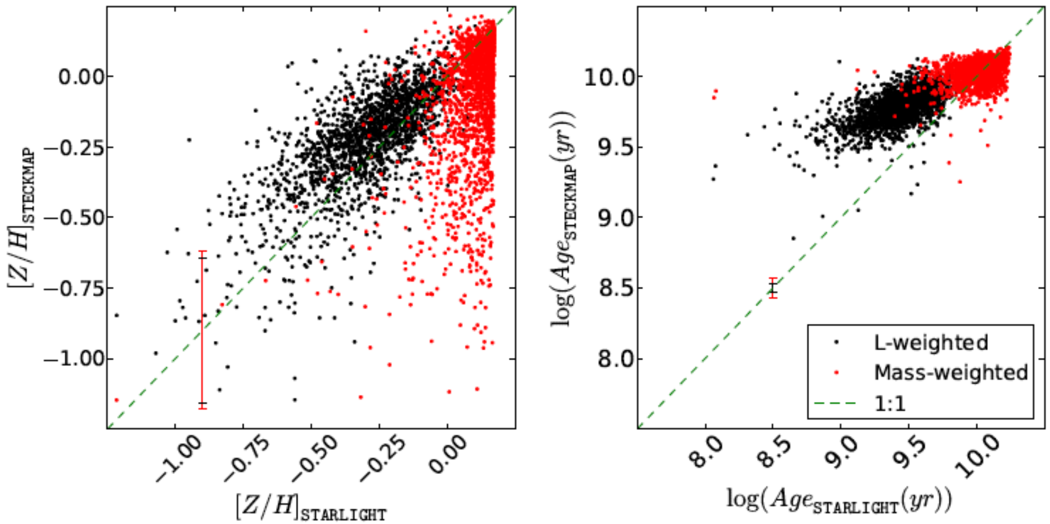




http://www.sedfitting.org/Fitting.html

1. Introduction: Motivation

- Discrepancies among SED Fitting tools



- Machine Learning in Astrophysics (1% of papers in NASA ADS Abstract Service with keyword Machine Learning)

1. Introduction: Goals

- Learn the whole life cycle of data, in particular the capture, the cleaning, the analysis and the visualization.

- Design a tool to calculate photometric redshift and estimate different variables related to distant galaxies (redshift and stellar mass, among others).

- To design, train and test a neural network, capable of accepting the emission of the galaxies in different predefined filters as the input, and returning a reliable array as close as possible to the parameters previously obtained for the galaxy using other techniques

1. Introduction: Methodology

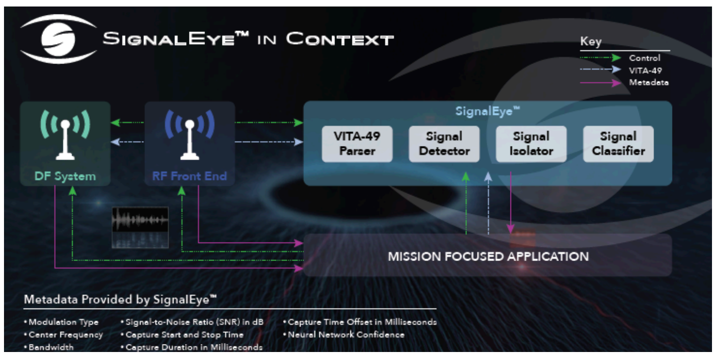
- Deep Neural Networks.

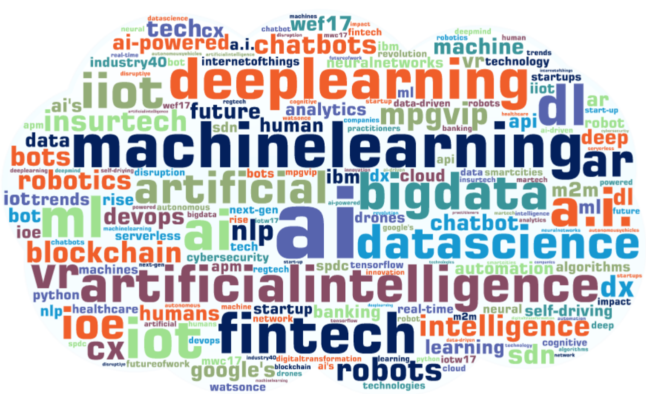
- Python 3.6.6, with the following frameworks and modules installed
    - AstroPY, version 3.2.1
    - iPython, version 6.5.0
    - Keras, version 2.2.4 using TensorFlow backend
    - Matplotlib, version 2.2.2
    - Numpy, version 1.15.2
    - PyAstronomy, version 0.13.0
    - Scipy, version 1.1.0
    - Spectres
    - TensorFlow, version 1.12.0
- ProjectLibre and MSProject, for the planning
- SublimeText, for coding
- TeXworks, using MikTeX backend for writing the report

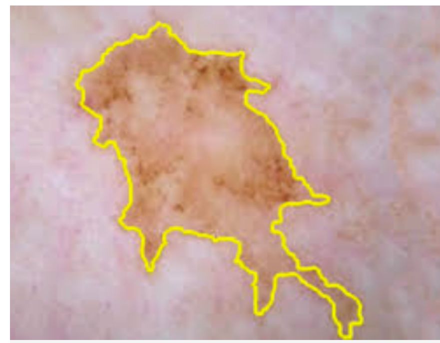
2. State of the art

2. State of the Art: Signal Processing

- Computer Vision —> Trend technology
- Signal Processing

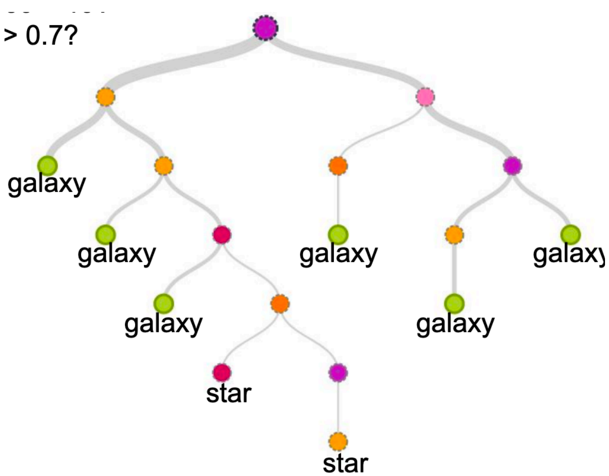


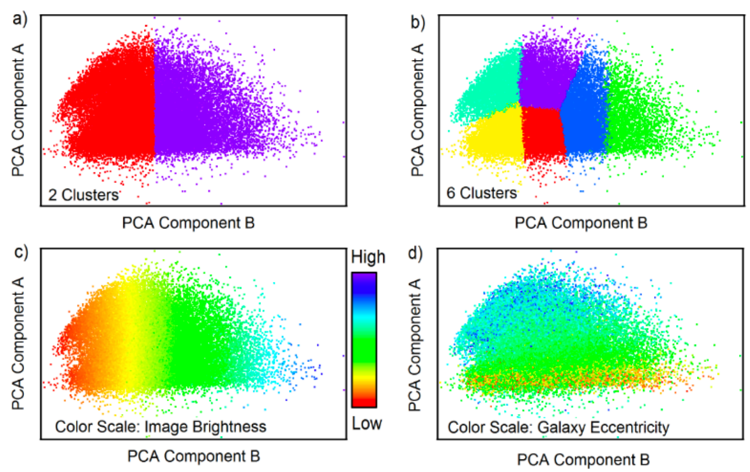


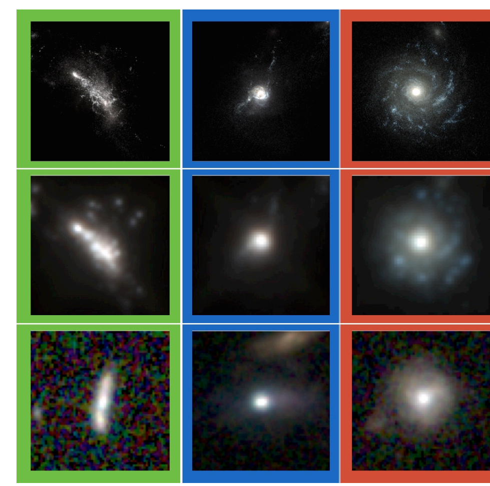


2. State of the Art: ML in Astrophysics

- Star-galaxy classification with random forest (Costa-Duarte *et al.* 2018)
- Galaxy morphology Gauthier *et al.* 2016)
- Data augmentation (Huertas-Company *et al.* 2018)

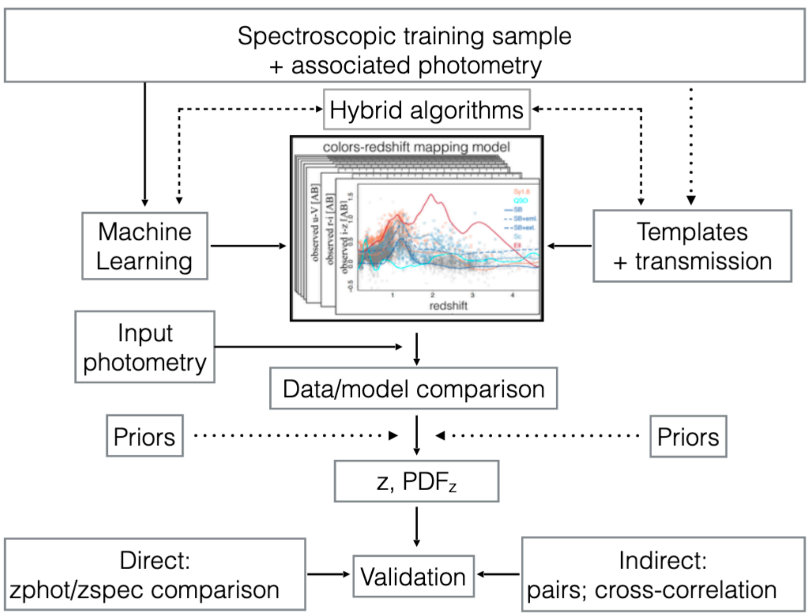


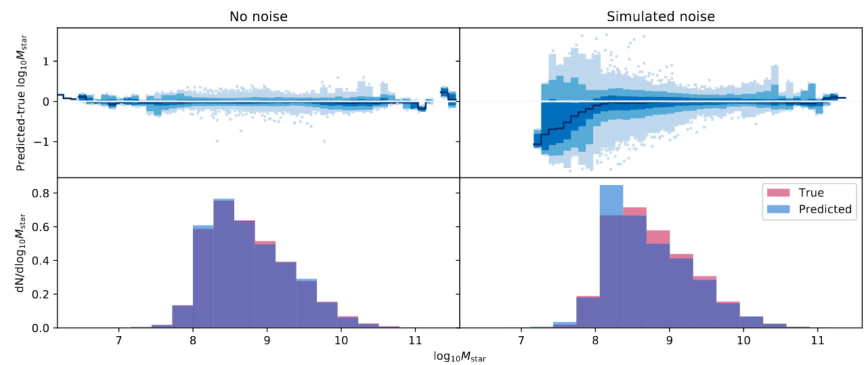


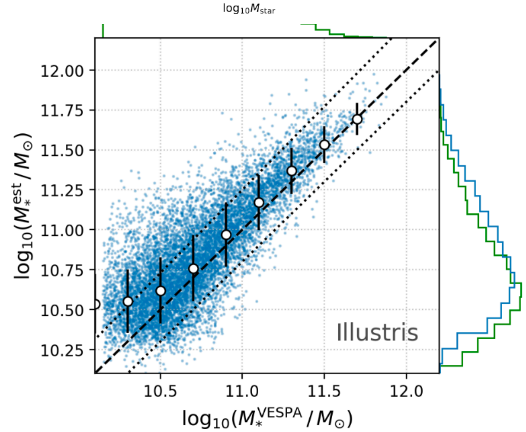


2. State of the Art: SED Fitting

- Photometric redshift (Salvato *et al.* 2018).
- Stellar mass, stellar metallicity, and average star formation rate (Simet *et al.* 2019, Lovell *et al.* 2019).







2. State of the Art: Narrow Band Filter Surveys

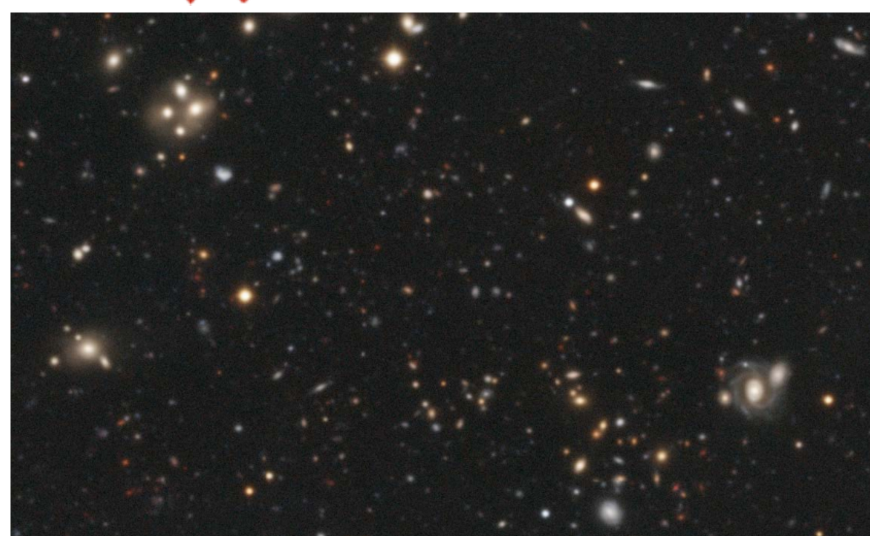
- ALHAMBRA, SHARDS, J-PAS/J-PLUS…

3. Implementation

3. Implementation: Data

- ALHAMBRA Survey
  - Csv file (‘,’ as delimiter) with 446,343 Objects
  - Integers: ID, Field, Pointing, CCD number
  - Float: Sky coordinates, CCD coordinates, *stellarity,* Fluxes, redshift, spectral type
  - String: F814W_Image
- Input: Fluxes in filters
- Outputs:
  - Redshift
  - Stellarity
  - Spectral Type
  - Stellar Mass

ALHAMBRA SURVEY



3. Implementation: Preprocessing

- Pandas dataframe
  - No NaNs
  - Fluxes: -99 if image is saturated; 99 if no detection —> clipping to [0, 27]
  - Stellarity —> categorical value
  - No need for normalization

3. Implementation: Modeling

- 4 different NN
  - 3 for regression: Redshift, Spectral Type, Stellar Mass
  - 1 for classification: Stellarity

	Input	Hidden	Output layer	Optimizer	Loss	Metrics
Redshift	75 Linear	120 Linear	1 neuron Linear	Adam	MSE	R Square
Stellarity	75 ReLu	2, 3 ReLu	2 SoftMax	Adam	Categorical Crossentropy	Accuracy
Spectral Type	75 Linear	20, 10 Linear	1 neuron Linear	RMSprop	MSE	R Square
Stellar Mass	75 Linear	120, 200, 50 Linear	1 neuron Linear	Adam	MSE	R Square

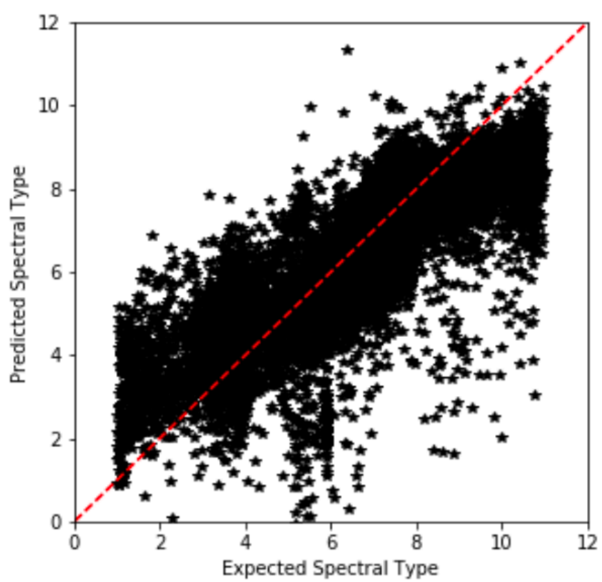
3. Implementation: Training

- Training epochs
  - Subset of data (10% for training, 1% for validation)
  - Training for 500 epochs
  - Loss *vs* epoch and Metrics *vs* epoch

- Optimal training epochs
  - Redshift: 80 epochs
  - Stellarity: 20 epochs
  - Stellar Mass: 7 epochs
  - Spectral Type: 10 epochs

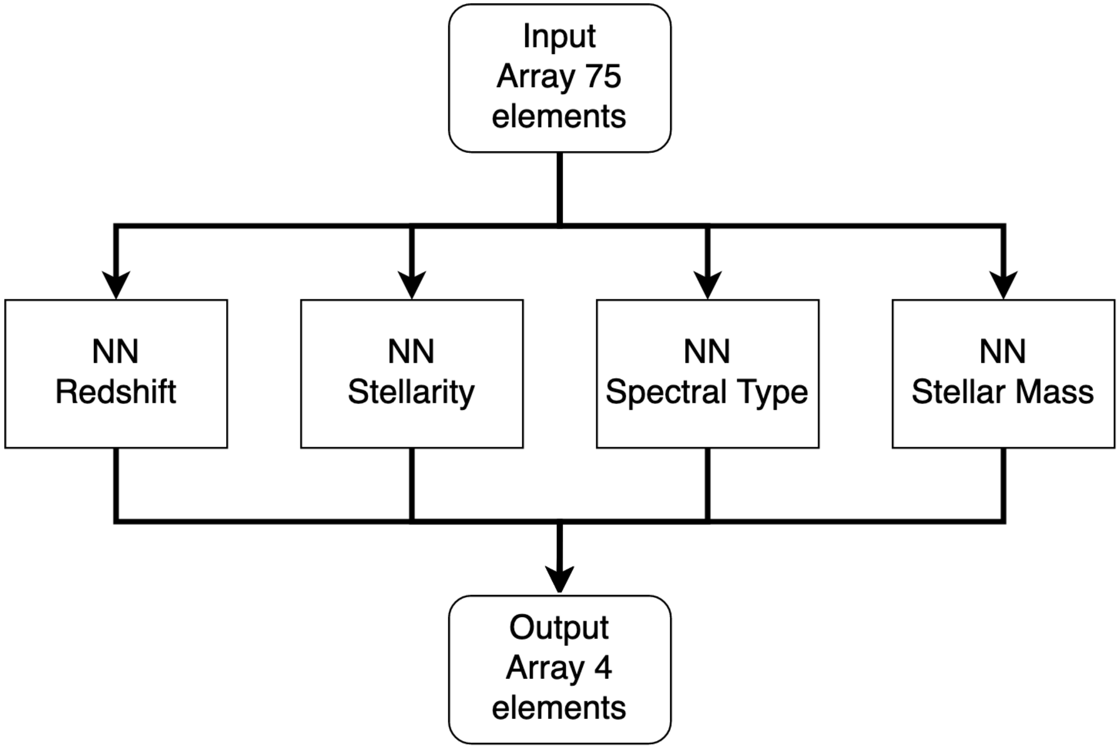
3. Implementation: Postprocessing

- No normalization —> no need for “un-normalization”
- Output *vs* Predicted values:
  - Correlation, but slope is not 1
  - Need for correction
    - Modification of Neural Network (see Future Work)
    - Polynomial Fitting



3. Implementation: Ensemble

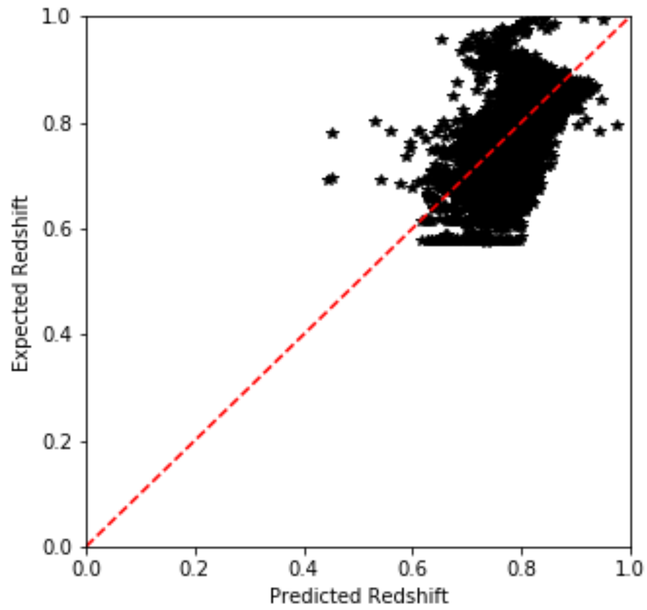
- 4 independent NN



4. Results

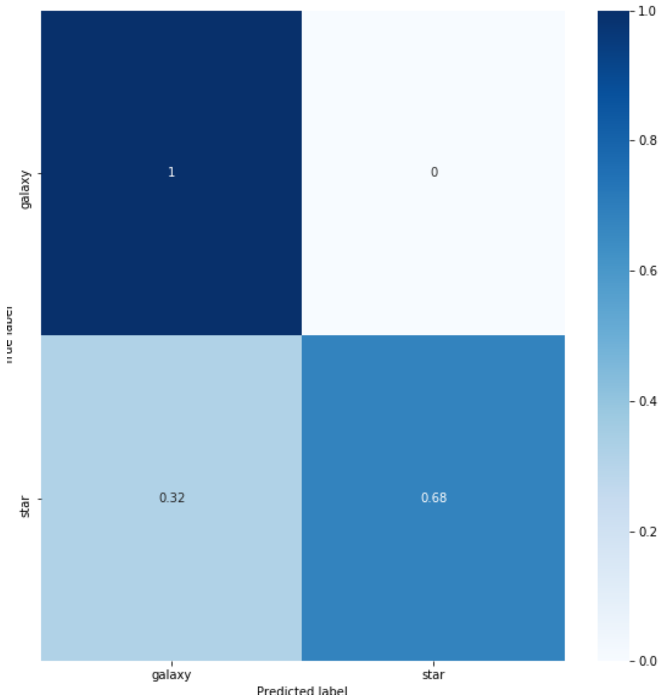
4. Results: Redshift

	Training dataset	Validation dataset
Initial Loss	0.4815	0.1607
Final Loss	0.0027	0.0021
Initial r2	-152.4	-48.8
Final r2	0.2019	0.3760



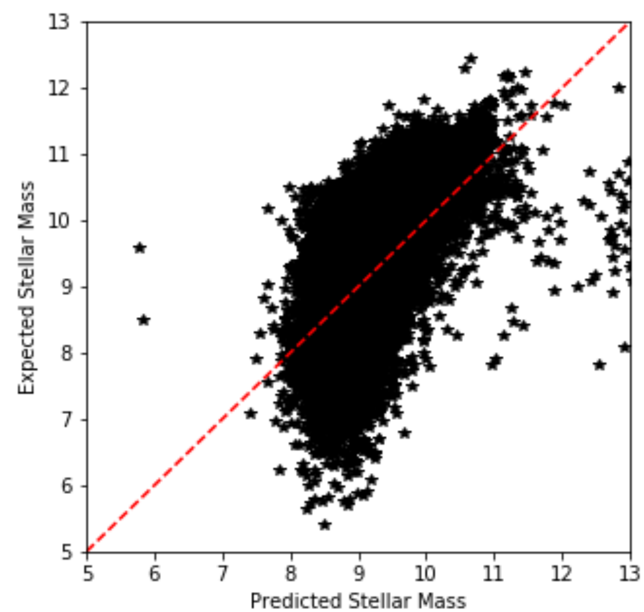
4. Results: Stellarity

	Training dataset	Validation dataset
Initial Loss	0.1111	0.1084
Final Loss	0.0446	0.0494
Initial accuracy	97.40%	97.24%
Final accuracy	98.33%	98.32%



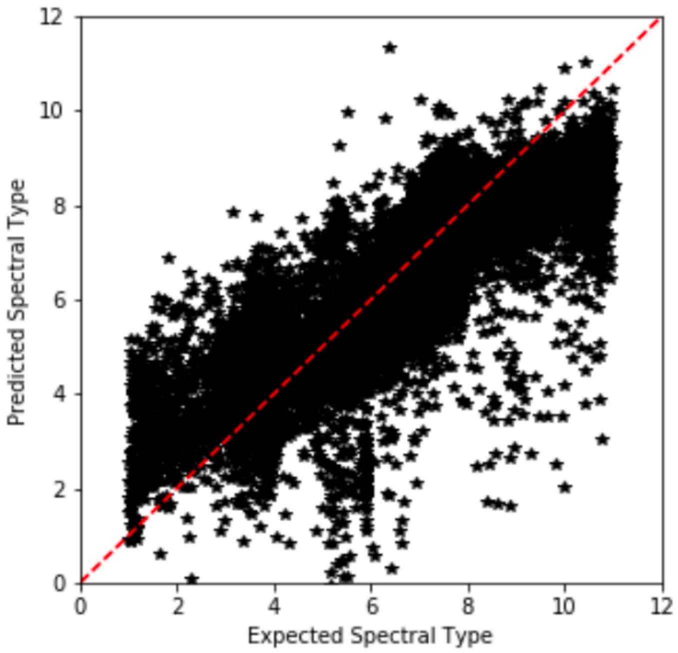
4. Results: Stellar Mass

	Training dataset	Validation dataset
Initial Loss	1.2458	0.5330
Final Loss	0.5310	0.5124
Initial r2	-0.5766	0.3251
Final r2	0.3240	0.3552



4. Results: Spectral Type

	Training dataset	Validation dataset
Initial Loss	2.6171	1.3255
Final Loss	1.1477	1.2242
Initial r2	0.1708	0.5957
Final r2	0.6399	0.6246



5. Discusion

5. Discussion: Summary

- Design and training of 4 different neural networks
- Prediction of Redshift, Stellarity, Stellar Mass and Spectral Type of galaxies
- Data from Alhambra Survey
- Comparison of Predicted and Expected output

5. Discusion: Conclusions

- Life cycle of data: Capture, exploratory analysis, cleaning, analysis and visualization
- Desing and training of 4 models to predict different types of varibles in observed objects
- Solution for a complex problem in Astrophysics
- Star-Galaxy Classification
- High accurate results:
  - $\frac{\Delta z}{1+z} = 0.03 \rightarrow \frac{\Delta z}{1+z} \leq 0.006$
  - Star-Galaxy precision = 98%
  - Stellar Mass: $\Delta M_{stellar} = 0.14$
  - Spectral Type: $\Delta ST = 0.2$

5. Discussion: Future Work

- Improve Neural networks: Include slope correction for predicted values
- Scalability of model for other photometric surveys
- Prediction of the spectrum of the object
- Improve model to accept spectra instead of photmetric data
- Implement calculation of uncertainties using MC simulations

Bibliography

Bibliography

Cacho, R. (2015). Formation and evolution of galactic bulges: the importance of secular processes. Tesis Doctoral.

Cappellari, M. and Emsellem, E. (2004). Parametric Recovery of Line-of-Sight Velocity Distributions from Absorption-Line Spectra of Galaxies via Penalized Likelihood. Publications of the Astronomical Society of the Pacific, 116(816): 138-147}.

Chen, S., et al. (2019). Deep learning for large-scale real-world ACARS and ADS-B radio signal classification. CoRR, abs/1904.09425.

Cid Fernandes, et al. (2005). Semiempirical analysis of Sloan Digital Sky Survey galaxies - I. Spectral synthesis method. Monthly Notices of the Royal Astronomical Society, 358:363-378.

Falcón-Barroso, J., et al. (2011). An updated MILES stellar library and stellar population models. Astronomy & Astrophysics, 532:A95.

Fussell, L. and Moews, B. (2019). Forging new worlds: high-resolution synthetic galaxies with chained generative adversarial networks. Monthly Notices of the Royal Astronomical Society, 485(3):3203-3214.

Gillet, N., et al. (2019). Deep learning from 21-cm tomography of the cosmic dawn and reionization. Monthly Notices of the Royal Astronomical Society.

Herbel, J., et al. (2018). Fast point spread function modeling with deep learning. Journal of Cosmology and Astroparticle Physics, 2018(07):054-054.

Hon, M., et al. (2018). Detecting solar-like oscillations in red giants with deep learning. The Astrophysical Journal, 859(1):64.

Katebi, R., et al. (2019). Galaxy morphology prediction using capsule networks. Monthly Notices of the Royal Astronomical Society, 486(2):1539-1547.

Khan, A., et al. (2019). Deep learning at scale for the construction of galaxy catalogs in the dark energy survey. Physics Letters B, 795:248-258.

Kim, E. J. and Brunner, R. J. (2016). Star-galaxy classification using deep convolutional neural networks. Monthly Notices of the Royal Astronomical Society, 464(4):4463-4475.

Li, X., et al. (2019). A survey on deep learning techniques in wireless signal recognition. Wireless Communications and Mobile Computing, 2019.

Lovell, C. C., et al. (2019). Learning the relationship between galaxies spectra and their star formation histories using convolutional neural networks and cosmological simulations.

Ma, Z., et al. (2018). Radio Galaxy Morphology Generation Using DNN Autoencoder and Gaussian Mixture Models. arXiv e-prints, page arXiv: 1806.00398.

Ocvirk, P., Pichon, C., Lançon, A., and Thiébaut, E. (2006a). STEllar Content and Kinematics from high resolution galactic spectra via Maximum A Posteriori. Monthly Notices of the Royal Astronomical Society, 365:74-84.

O'Shea, T. J., Corgan, J., and Clancy, T. C. (2016). Convolutional Radio Modulation Recognition Networks. In Engineering Applications of Neural Networks, pages 213-226. Springer International Publishing.

Regier, J., et al. (2018). Approximate Inference for Constructing Astronomical Catalogs from Images. arXiv e-prints, page arXiv:1803.00113.

Reiman, D. M. and Göhre, B. E. (2019). Deblending galaxy superpositions with branched generative adversarial networks. Monthly Notices of the Royal Astronomical Society, 485(2):2617-2627.

Salvato, M., Ilbert, O., and Hoyle, B. (2019). The many flavours of photometric redshifts. Nature Astronomy, 3(3):212-222.

Silburt, A., et al. (2019). Lunar crater identification via deep learning. Icarus, 317:27-38.

Simet, M., et al. (2019). Comparison of observed galaxy properties with semianalytic model predictions using machine learning.

Tuccillo, D., et al. (2016). Deep learning for studies of galaxy morphology. Proceedings of the International Astronomical Union, 12(S325):191-196.

Walmsley, M., et al. (2019). Galaxy zoo: Probabilistic morphology through bayesian cnns and active learning. ArXiv, abs/1905.07424.

West, N. E. and O'Shea, T. J. (2017). Deep architectures for modulation recognition. CoRR, abs/1703.09197.

Wilkinson, D. M., et al. (2017). FIREFLY (Fitting IteRativEly For Likelihood analYsis): a full spectral fitting code. Monthly Notices of the Royal Astronomical Society, 472(4):4297-4326.

Universitat Oberta
de Catalunya

Thank you for your time

UOC.universitat
@UOCuniversitat
UOCuniversitat