



Galaxy Morphology classification using CNN

Astronomy Institute, UNAM
Dr. J Antonio Vázquez-Mata

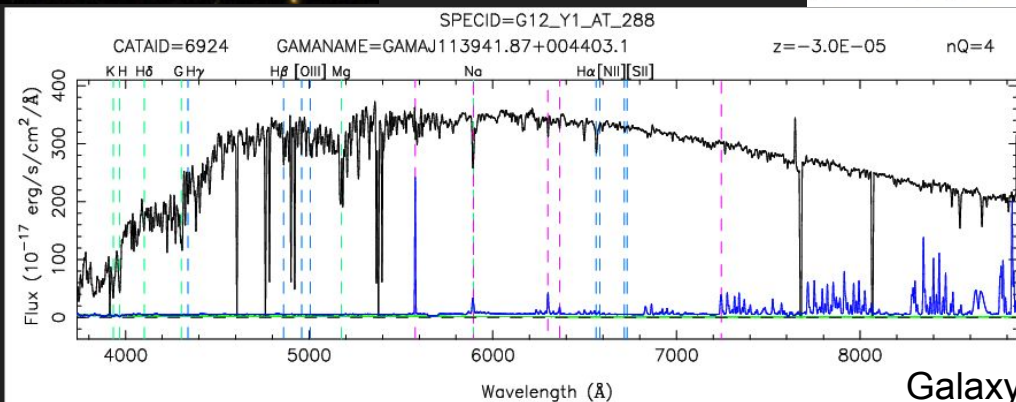
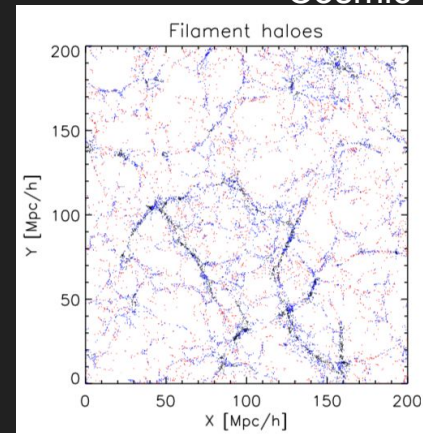
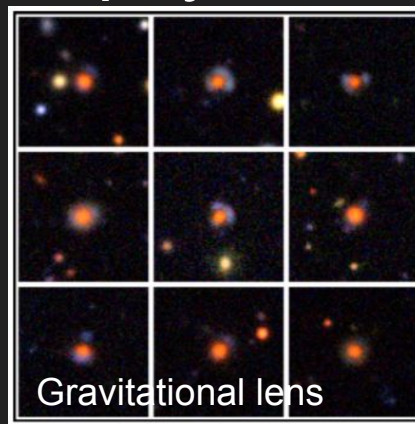
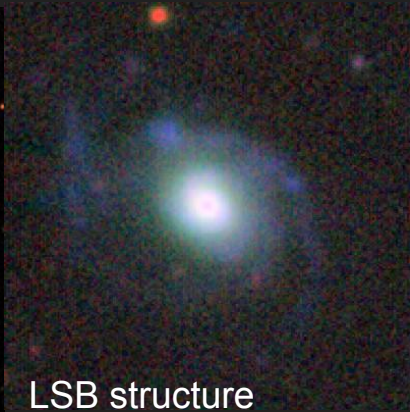
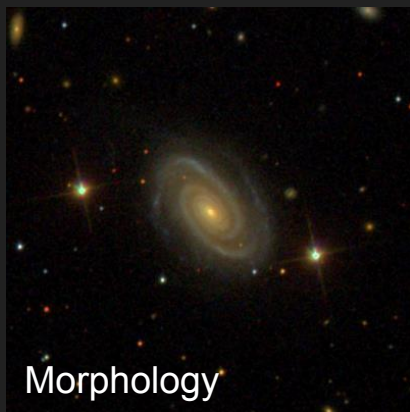
In collaboration with: Dr. H. Hernández-Toledo
Luis Carlos Mascherpa

Symposium Artificial Intelligence for Science, Industry and Society
UNAM
21st October 2019

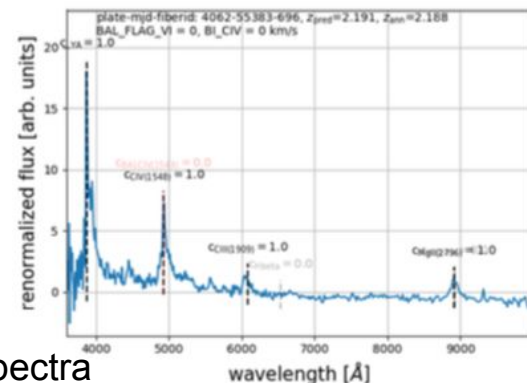


Why do we need ML in astrophysics?

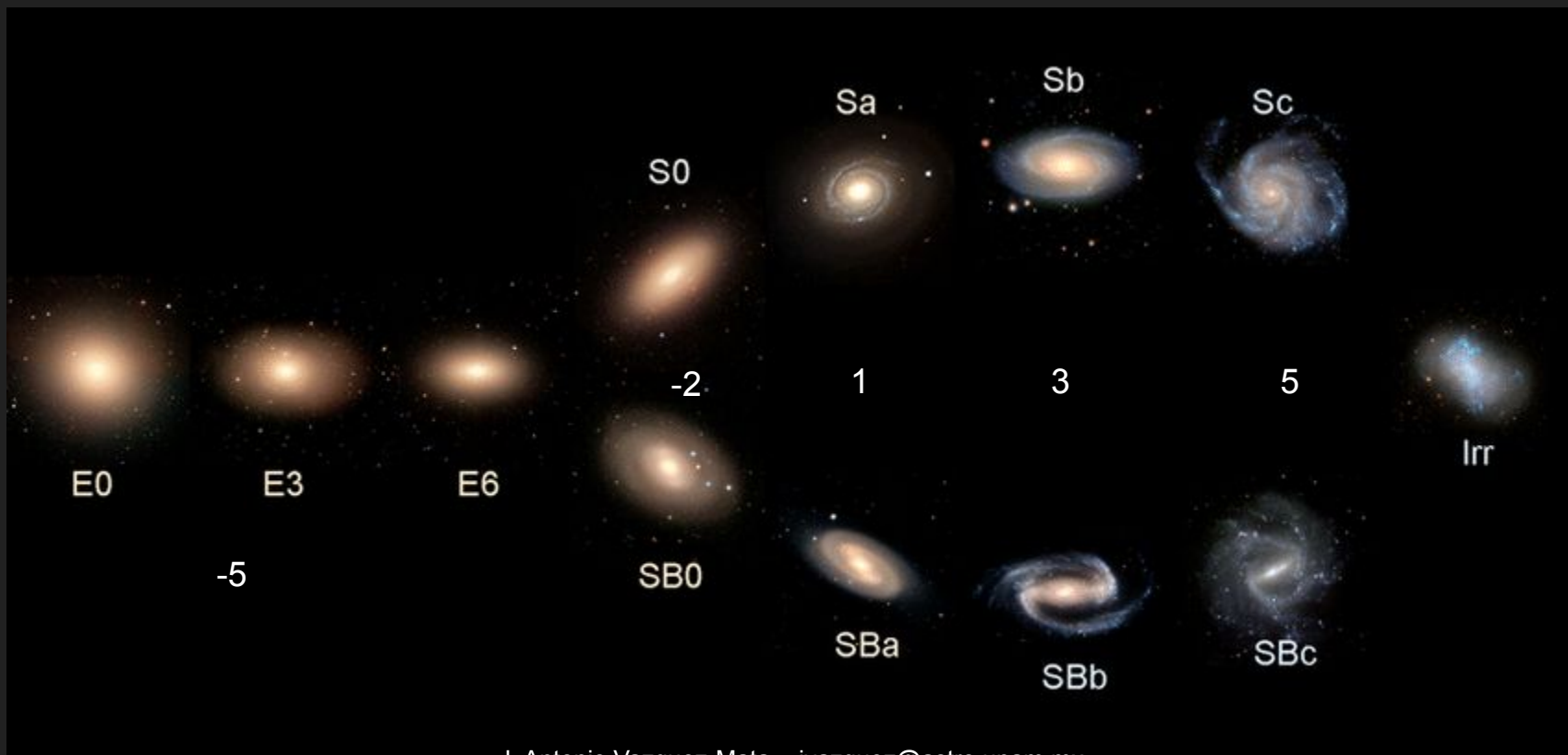
Cosmic Web

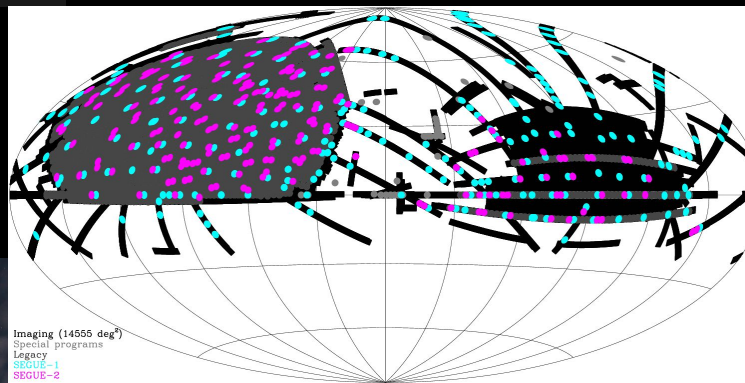


Galaxy Spectra

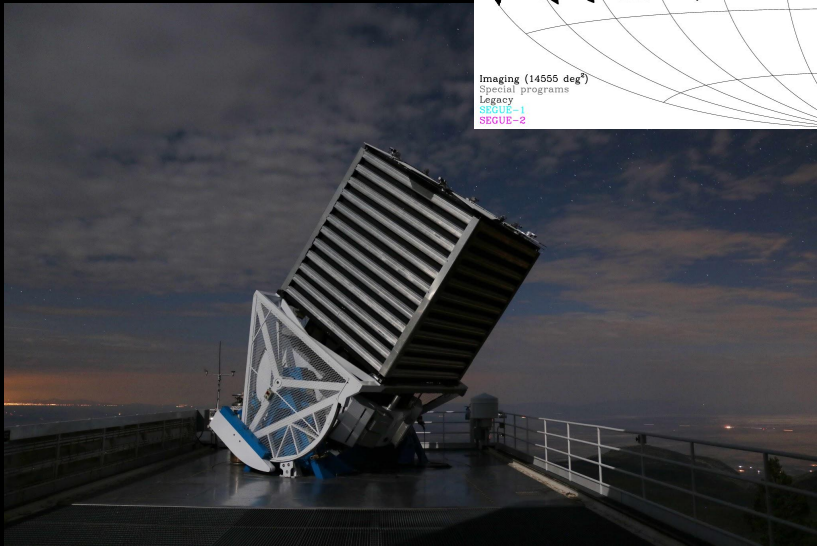
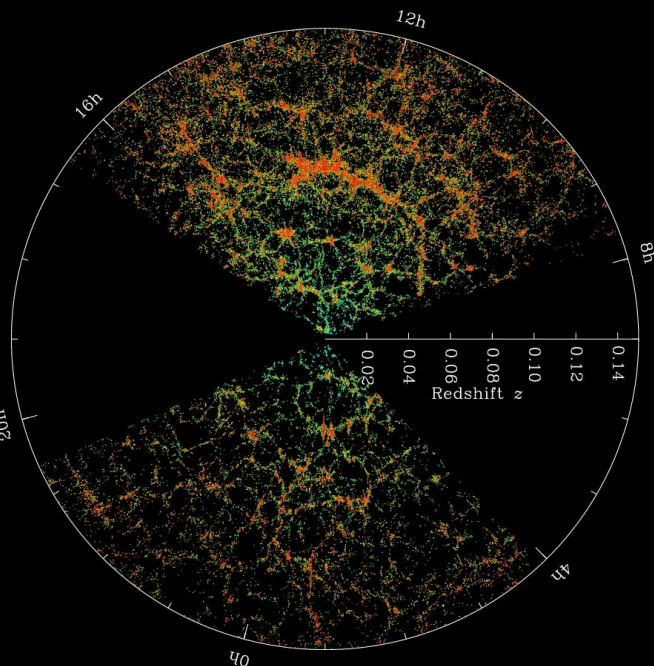


Morphology (Hubble Sequence)





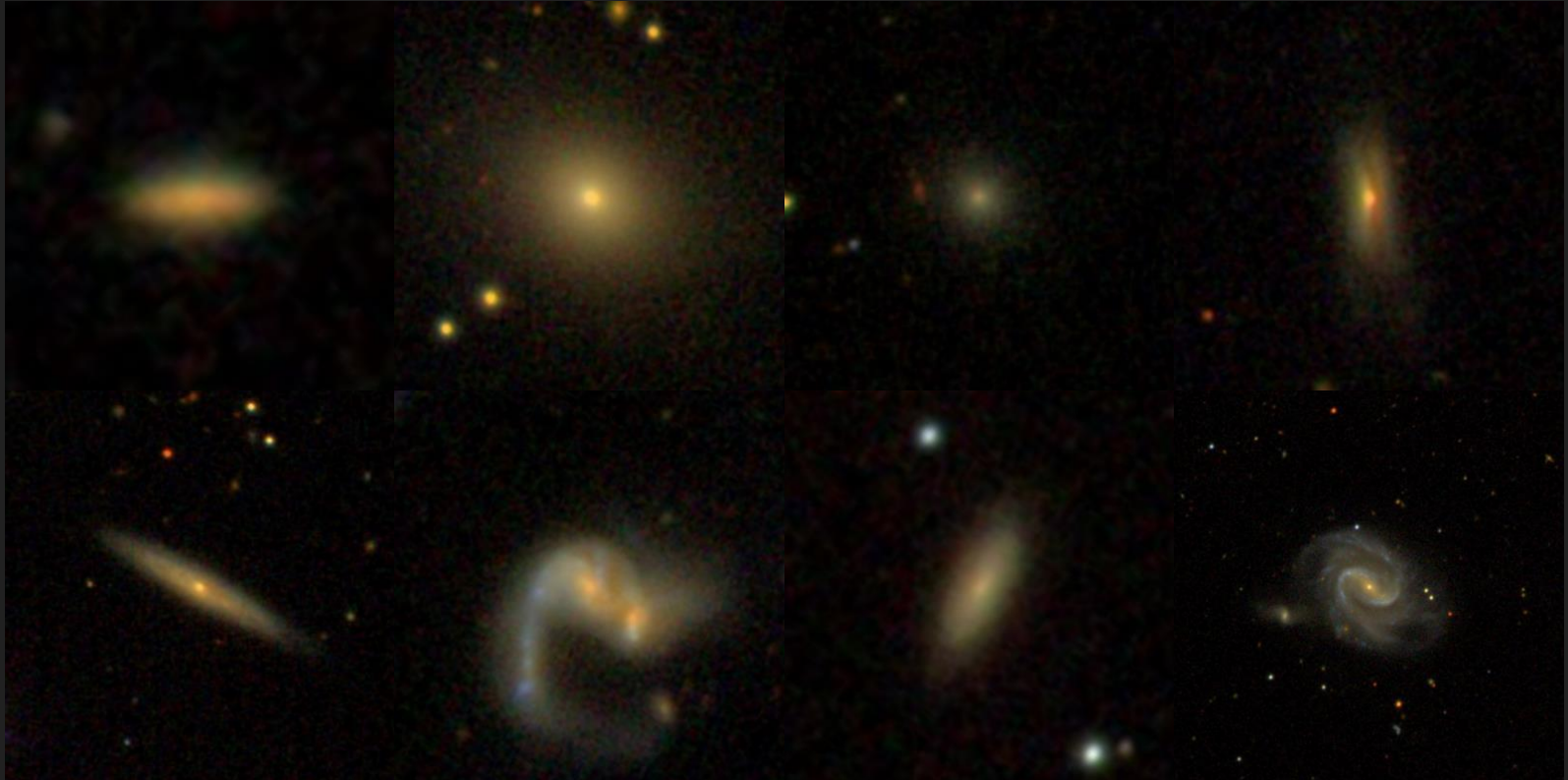
Imaging (14555 deg²)
Special programs
Legacy
SEGUE-1
SEGUE-2



Telescope 2.5 m, New Mexico
 $\frac{1}{3}$ of sky covered
 ~ 1.6 million galaxies with spectra
Se observa la estructura a gran escala

What kind of objects do we have?

Sloan Digital Sky Survey (SDSS) images



Hubble Classification using SVM

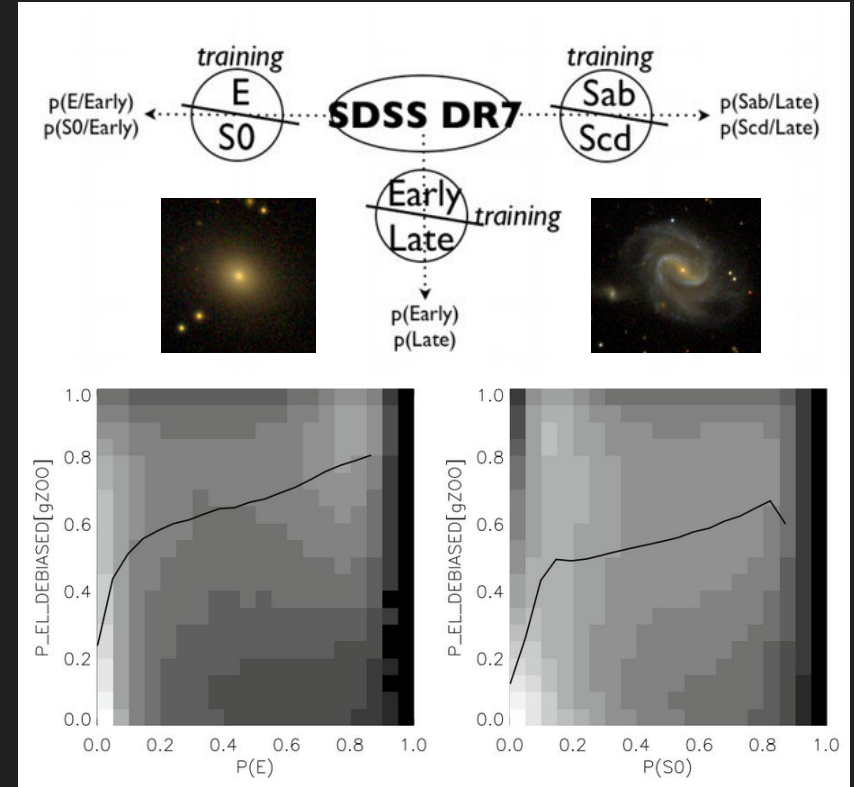
Huertas-Company et al. (2008, 2009, 2011)

SDSS galaxies considering colour, axis ratio and light concentration.

First, it was used to classify galaxies at high redshifts i.e. very far, faint, difficult to classify by eye.

~2000 classified galaxies as training sample

Good correlation with visual classification.



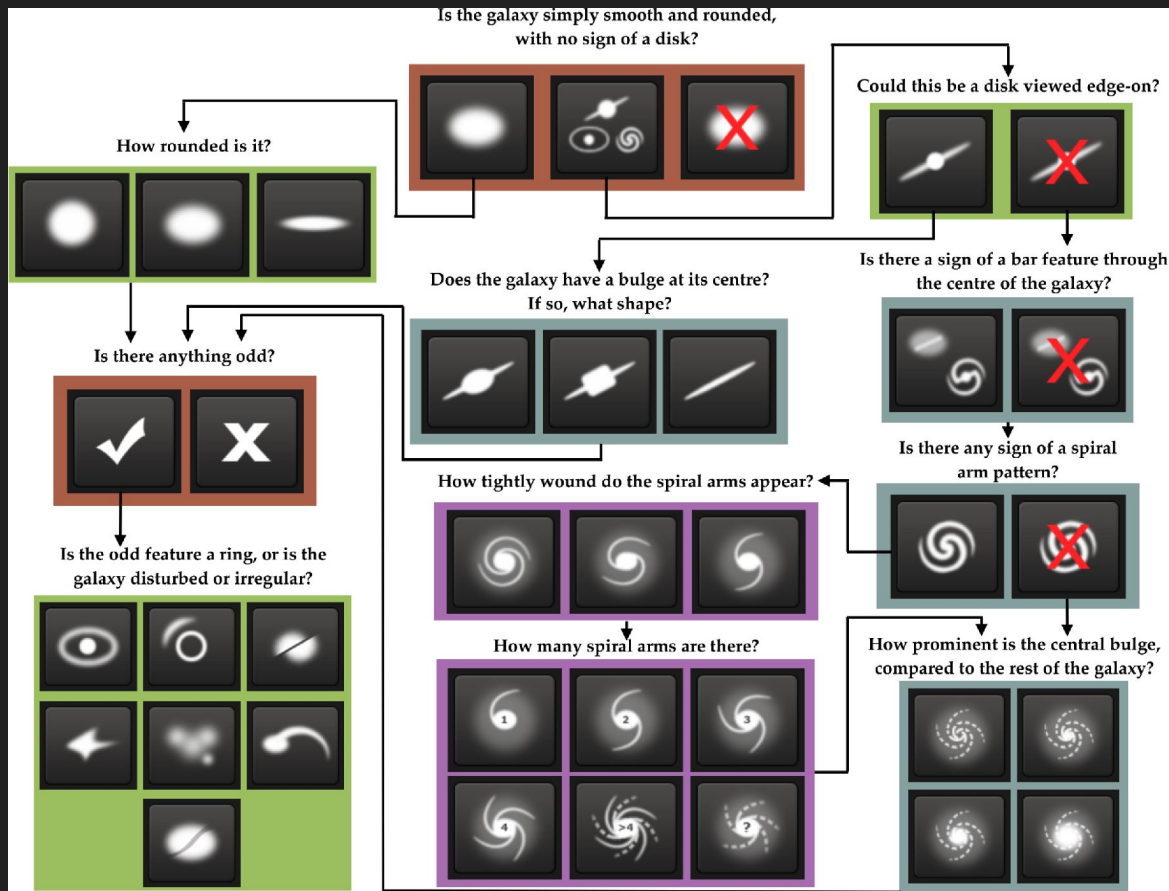


SDSS classified galaxies:

~ 300 000

carried out by thousands of “normal” people, and then a probability is given based on the number of votes.

<https://zoo4.galaxyzoo.org>



Galaxy Zoo (Kaggle competition)

Galaxy Zoo - The Galaxy Challenge

Classify the morphologies of distant galaxies in our Universe

sponsored by Intel

\$16,000 · 326 teams · 6 years ago

Overview Data Notebooks Discussion Leaderboard Rules [Join Competition](#)

Overview

Description	Understanding how and why we are here is one of the fundamental questions for the human race. Part of the answer to this question lies in the origins of galaxies, such as our own Milky Way. Yet questions remain about how the Milky Way (or any of the other -100 billion galaxies in our Universe) was formed and has evolved. Galaxies come in all shapes, sizes and colors: from beautiful spirals to huge ellipticals. Understanding the distribution, location and types of galaxies as a function of shape, size, and color are critical pieces for solving this puzzle.
Evaluation	
The-galaxy-zoo-decision-tree	
About-winton-capital	
Prizes	Galaxy
About-galaxy-zoo	The Whirlpool Galaxy (M51). Credit: NASA and European Space Agency

With each passing day telescopes around and above the Earth capture more and more images of distant

61 578 training images

79 975 test images

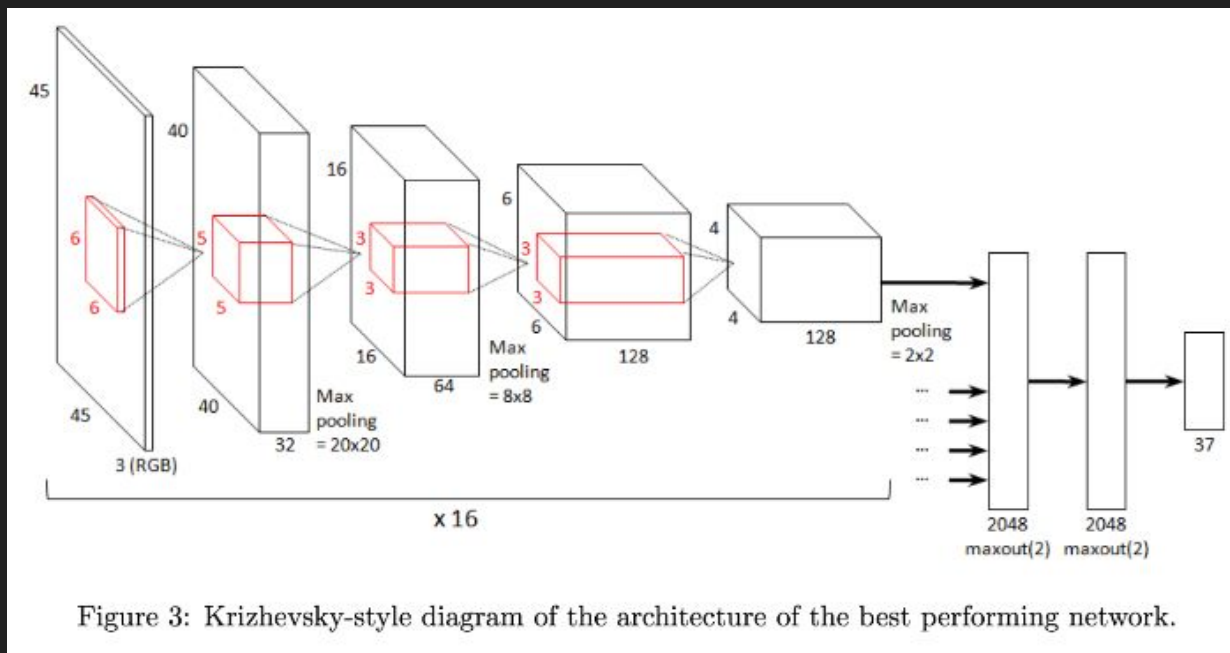
SDSS images
45x45 pixels

WINNER:

Sander Dieleman in 2014

Dieleman et al. (2015)

Galaxy Zoo (Kaggle competition)



Accuracy > 90%

For galaxies with high classification probabilities -> ~99%

Dieleman et al. (2015)

Improving galaxy morphologies for SDSS with Deep Learning (Dominguez-Sanchez et al. 2018)

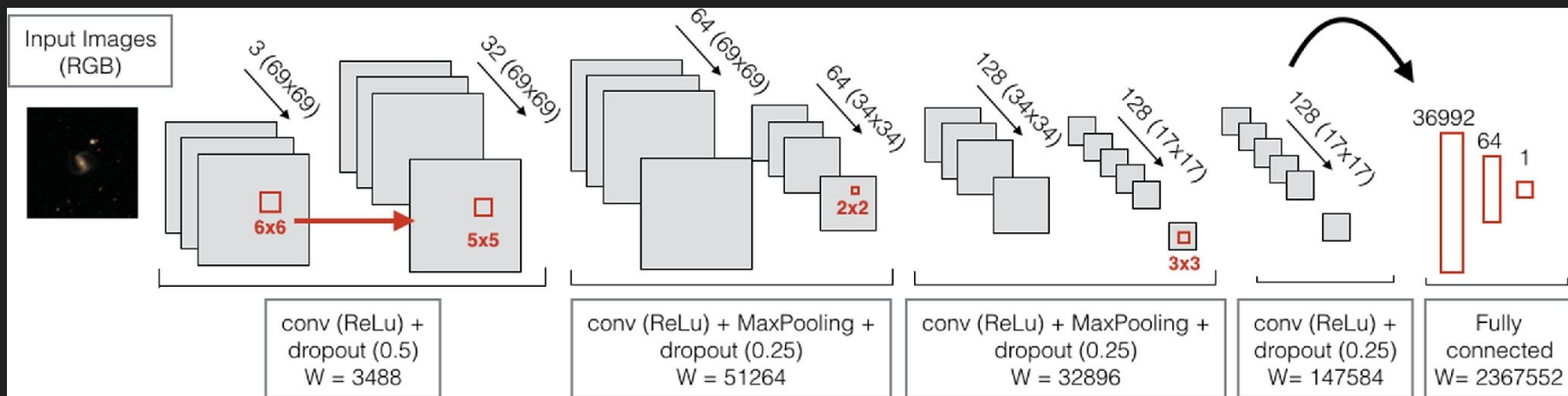
Catalogues:

- ~ 240 000 SDSS images from Galaxy Zoo (Decision tree)
- ~ 14 000 SDSS images from Nair et al. 2010 (Actual Hubble type)

Results:

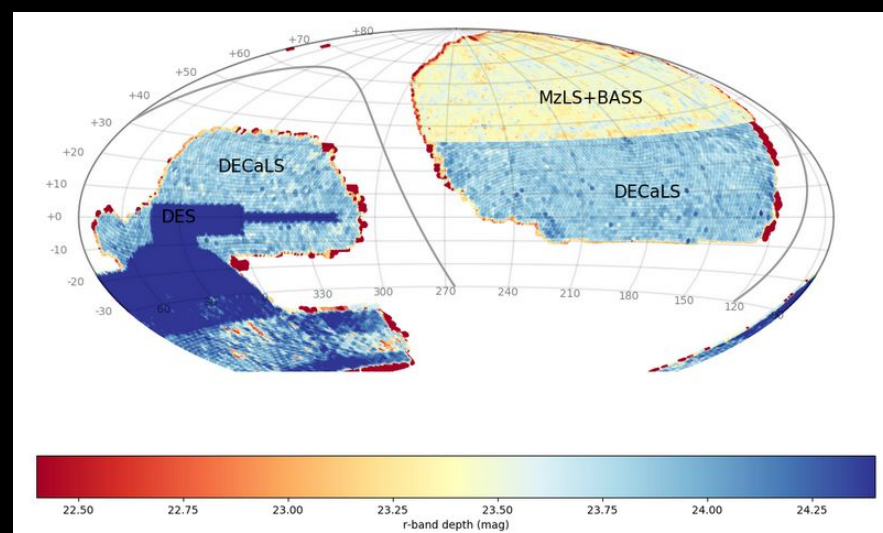
- ~ 670 000 galaxies classified using SDSS images
- acc > 0.9 using GZoo
- for multiclass, authors do not report accuracy

Improving galaxy morphologies for SDSS with Deep Learning (Dominguez-Sanchez et al. 2018)



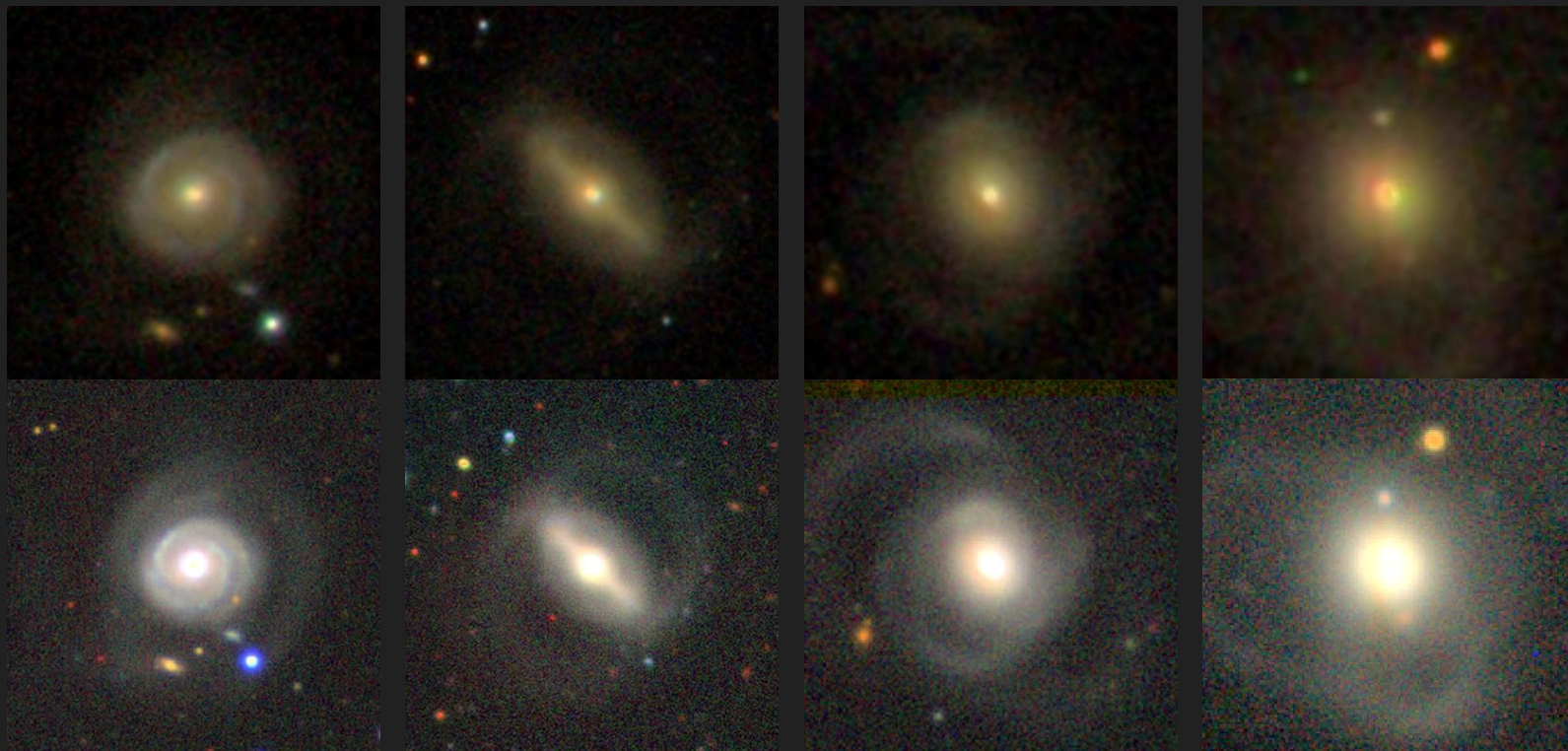


DARK ENERGY SPECTROSCOPIC INSTRUMENT



Mayall Telescope 4 m, Arizona.
40 million galaxies with spectra

SDSS and DESI



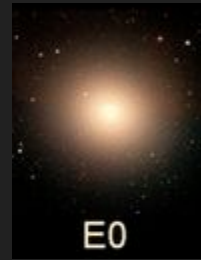
Tested methods (Cheng et al. 2019)

Labels	Machine Learning Algorithms
1	K-Nearest Neighbour (KNN)
2	Support Vector Machine (SVM)
3	SVM + Restricted Boltzmann Machine (SVM+RBM)
4	Logistic Regression (LR)
5	LR + Restricted Boltzmann Machine (LR+RBM)
6	Multi-Layer Perceptron Classifier (MLPC)
7	Convolutional Neural Network (CNN)

Using the Galaxy Zoo sample

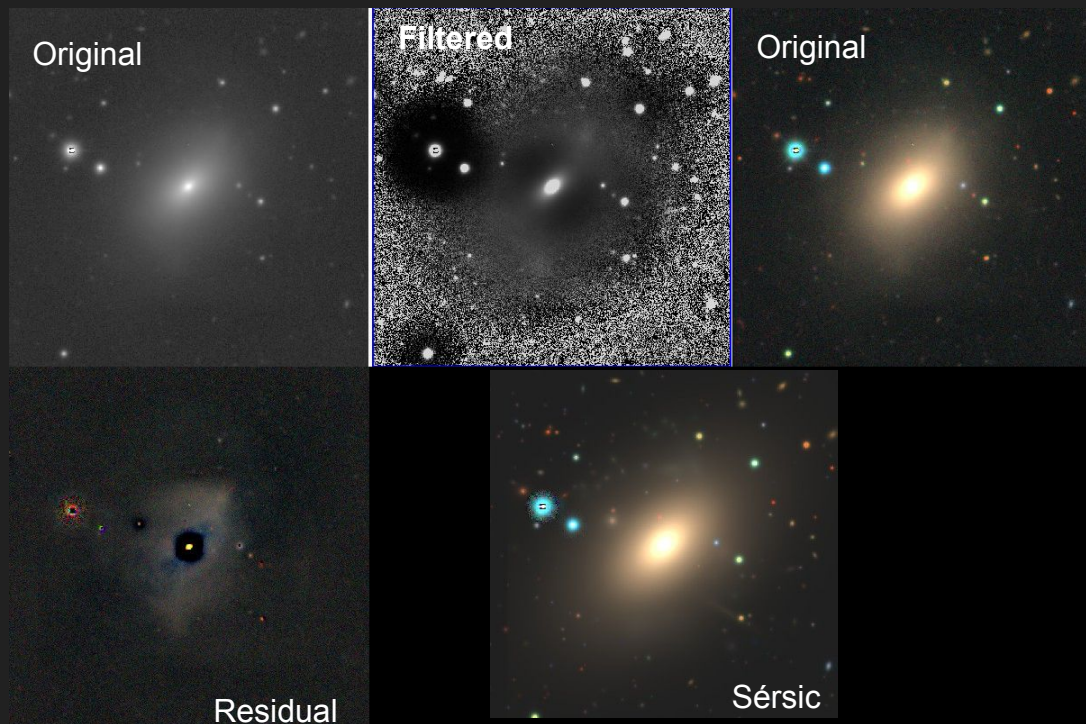
DESI images

Binary classification:
Early / Late types



Winner: Convolutional Neural Network

Current work



Gaussian filter applied to DESI images to enhance internal structures (Hernández-Toledo et al. 2010)

Tidal debris detection

Residual image after subtracting a Sérsic model.

~ 6500 MaNGA galaxies

-> classification improvement

New challenge for the CNN

Our CNN results

(Luis C. Mascherpa thesis)

~10 000 galaxies from Nair et al. (2010) and MaNGA catalogues to train

Accuracy using Inception

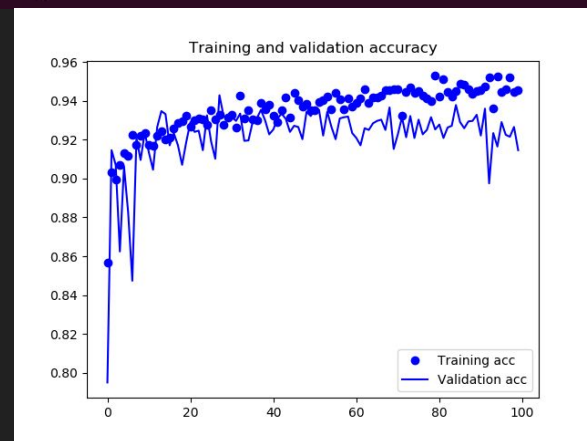
~ 0.94 in training without overfitting
~ 0.92 when applying a test set (~1000 galaxies)

multiclass:

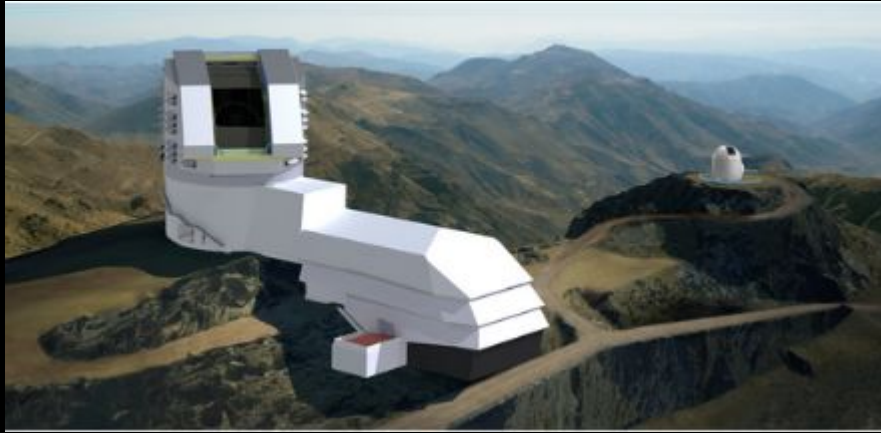
~ 0.6 due to the number of galaxies in each class

Layer (type)	Output Shape	Param #
inception_v3 (Model)	(None, 3, 3, 2048)	21802784
flatten_1 (Flatten)	(None, 18432)	0
dropout_1 (Dropout)	(None, 18432)	0
dense_1 (Dense)	(None, 64)	1179712
dense_2 (Dense)	(None, 1)	65

=====
Total params: 22,982,561
Trainable params: 22,948,129
Non-trainable params: 34,432

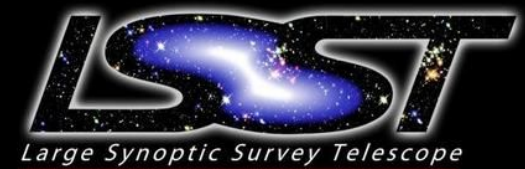


Future



Operations will start in 2023

- 8.4 meters Telescope, Cerro Pachón, Chile
- Scanning the whole sky every 3 nights
- Every region will be observed in 6 bands for 10 years
- 3 000 Mpix camera, 20 Tbytes per night
- Millions of galaxies



New strategies, unsupervised classification?

Martin et al. 2019

Using deep data from the Subaru Telescope
(deeper than DESI)

~50 000 objects detected

160 clusters identified at higher redshifts

Results are comparable with those
obtained using the Hubble morphology

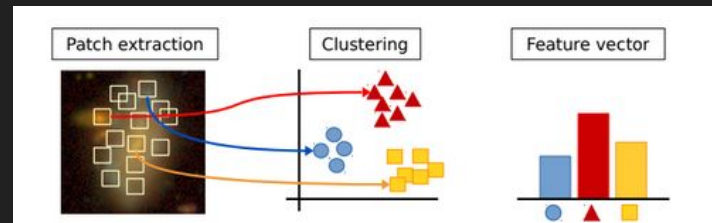
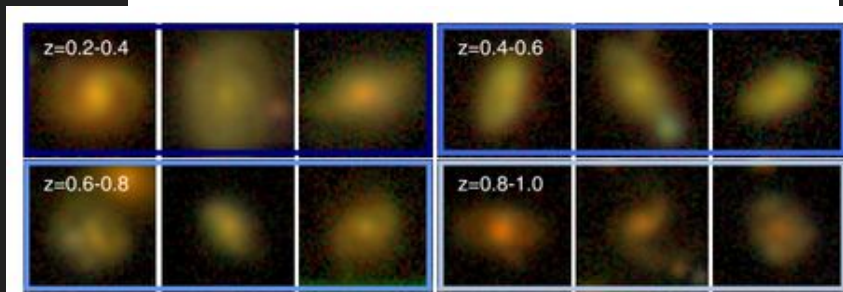


Figure 1. A schematic view of the morphological classification process. Patches are extracted around detected pixels in survey images and clustering methods are used to group these patches into a library of patch types. Galaxy feature vectors can then be constructed by creating a histogram for each object which describes the frequency of each patch 'type'.



(a) *Spiral galaxies.*

Summary

- Morphology of galaxies is fundamental to study the evolution of the universe.
- Machine Learning is playing an important rôle to classify very large samples of galaxies.
- Limitations: large samples well classified, image resolution/processed and number of galaxies in each class.
- Can unsupervised morphology be the real way to face the new/future large galaxy samples?

