A Deep Learning approach for the estimation of photometric redshifts of galaxies from the Main galaxy Sample of SDSS

Johanna Pasquet

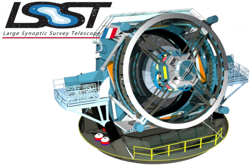
Centre de Physique des Particules de Marseille

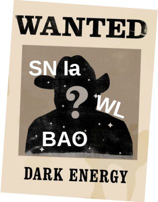
LSST Webinaire

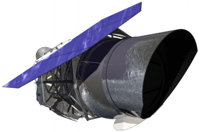
22 May, 2018

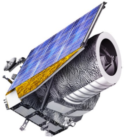
Outline

The future image surveys

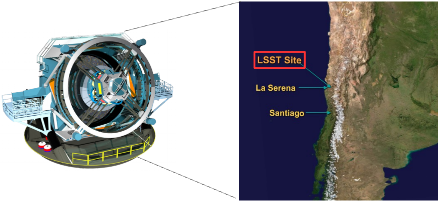








The Large Synoptic Survey Telescope



Large Synoptic Survey Telescope

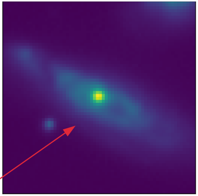
- LSST will produce the deepest, widest, image of the Universe :
  - 37 billion stars and galaxies
  - 10 year survey of the sky
  - 15 Terabytes of data ...every night !

Issues :

- LSST will discovery hundreds of thousands of type Ia supernovae
- Be able to automatically identify Sne Ia among all the supernovae with only the photometric information

Redshifts of galaxies

Step 2 : Identify and measure redshift of the galaxy



galaxy

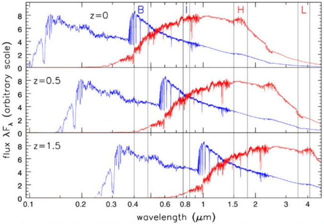


Fig 8.12 (S. Charlot) 'Galaxies in the Universe' Sparke/Gallagher CUP 2007

Spectroscopy

z-phot : a very useful tool

- Host galaxy of SN Ia

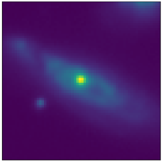
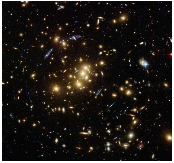


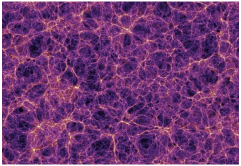
Image from SDSS DR9

- Weak lensing



Strong gravitational lensing around galaxy cluster CL0024+17
Credit : NASA/ESA/M.J. Jee (John Hopkins University)

- Cosmic web



Results of a digital simulation showing the large-scale distribution of matter, with filaments and knots.
Credit: V.Springel, Max-Planck Institut für Astrophysik, Garching bei München

Why use Deep Learning methods?

- Context of LSST
  - spectroscopic follow up limited → deal with a huge quantity of photometric data
  - Large amount of images
  - Feature extractions introduce information loss and is computationally expensive
- Deep Learning
  - Very efficient with large quantities of data
  - CNN won a lot of competitions for the classification of images (ImageNet)
  - Deep Learning methods do not involve a feature extraction step and can be fast to produce inferences

Why use Deep Learning methods?

- Context of LSST
  - spectroscopic follow up limited → deal with a huge quantity of photometric data
  - Large amount of images
  - Feature extractions introduce information loss and is computationally expensive
- Deep Learning
  - Very efficient with large quantities of data
  - CNN won a lot of competitions for the classification of images (ImageNet)
  - Deep Learning methods do not involve a feature extraction step and can be fast to produce inferences

Why use Deep Learning methods?

- Context of LSST
  - spectroscopic follow up limited → deal with a huge quantity of photometric data
  - Large amount of images
  - Feature extractions introduce information loss and is computationally expensive

- Deep Learning
  - Very efficient with large quantities of data
  - CNN won a lot of competitions for the classification of images (ImageNet)
  - Deep Learning methods do not involve a feature extraction step and can be fast to produce inferences

Why use Deep Learning methods?

- Context of LSST
  - spectroscopic follow up limited $\rightarrow$ deal with a huge quantity of photometric data
  - Large amount of images
  - Feature extractions introduce information loss and is computationally expensive
- Deep Learning
  - Very efficient with large quantities of data
  - CNN won a lot of competitions for the classification of images (ImageNet)
  - Deep Learning methods do not involve a feature extraction step and can be fast to produce inferences

Why use Deep Learning methods?

- Context of LSST
  - spectroscopic follow up limited → deal with a huge quantity of photometric data
  - Large amount of images
  - Feature extractions introduce information loss and is computationally expensive
- Deep Learning
  - Very efficient with large quantities of data
  - CNN won a lot of competitions for the classification of images (ImageNet)
  - Deep Learning methods do not involve a feature extraction step and can be fast to produce inferences

Why use Deep Learning methods?

- Context of LSST
  - spectroscopic follow up limited → deal with a huge quantity of photometric data
  - Large amount of images
  - Feature extractions introduce information loss and is computationally expensive

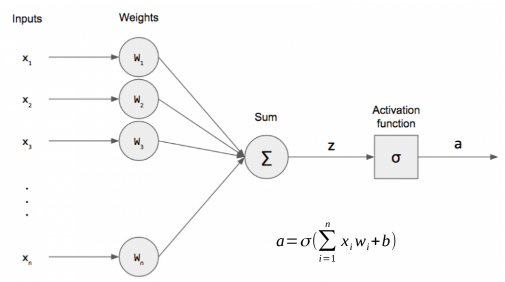
- Deep Learning
  - Very efficient with large quantities of data
  - CNN won a lot of competitions for the classification of images (ImageNet)
  - Deep Learning methods do not involve a feature extraction step and can be fast to produce inferences

Outline

Outline

History

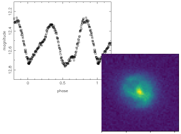
1957 Perceptron (Rosenblatt)
1986 MLP (Rumelhart et al.)
1998 LeNet (LeCun et al.)
2012 A CNN won ImageNet (Alexnet, Krizhevsky et al.)



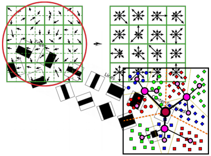
The main property of deep learning

Classical methods

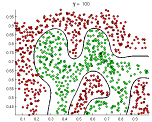
Input data



Feature crafting

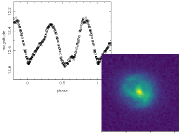


Separation with a classifier

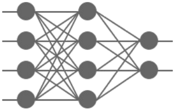


Deep learning

Input data



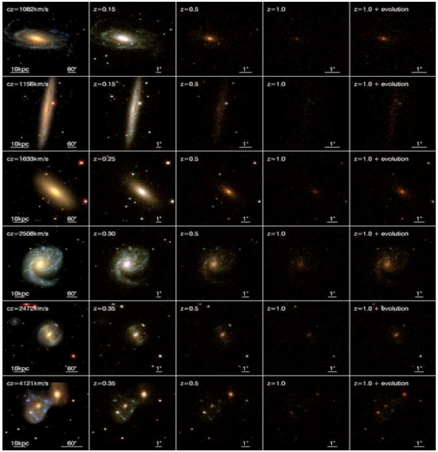
Feature learning



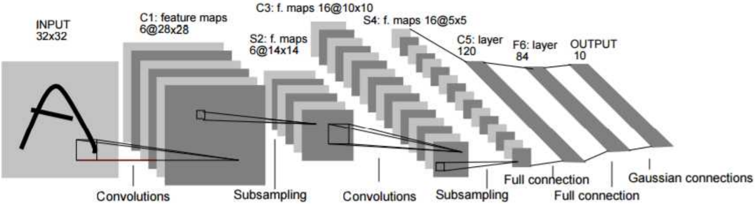
→ The best feature space representation is found by the network

The convolutional neural network in astronomy

Kaggle challenge with the goal to build an algorithm to classify the different morphologies of galaxies from JPEG images : a CNN won the challenge (Dieleman et al. 2015)



LeNet5

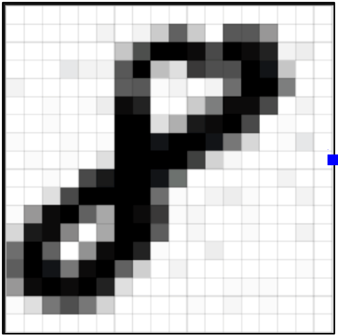


Lecun et al. 1998

3 operations:

- Convolution + non linearity (feature extraction)
- Pooling
- Fully Connected (classification)

Convolutions



0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	12	0	11	39	137	37	0	152	147	84	0	0	0
0	0	1	0	0	0	41	160	250	255	235	162	255	238	206	11	13	0
0	0	0	16	9	9	150	251	45	21	184	159	154	255	233	40	0	0
10	0	0	0	0	0	145	146	3	10	0	11	124	253	255	107	0	0
0	0	3	0	4	15	236	216	0	0	38	109	247	240	169	0	11	0
1	0	2	0	0	0	253	253	23	62	224	241	255	164	0	5	0	0
6	0	0	4	0	3	252	250	228	255	255	234	112	28	0	2	17	0
0	2	1	4	0	21	255	253	251	255	172	31	8	0	1	0	0	0
0	0	4	0	163	225	251	255	229	120	0	0	0	0	0	11	0	0
0	0	21	162	255	255	254	255	126	6	0	10	14	6	0	0	9	0
3	79	242	255	141	66	255	245	189	7	8	0	0	5	0	0	0	0
26	221	237	98	0	67	251	255	144	0	8	0	0	7	0	0	11	0
125	255	141	0	87	244	255	208	3	0	0	13	0	1	0	1	0	0
145	248	228	116	235	255	141	34	0	11	0	1	0	0	0	1	3	0
85	237	253	246	255	210	21	1	0	1	0	0	6	2	4	0	0	0
6	23	112	157	114	32	0	0	0	0	2	0	8	0	7	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Every image is a matrix of pixel values.

Convolutions

An image

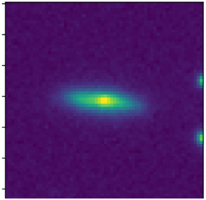
1	1	1	0	0
0	1	1	1	0
0	0	1	1	1
0	0	1	1	0
0	1	1	0	0

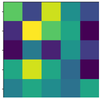
A kernel

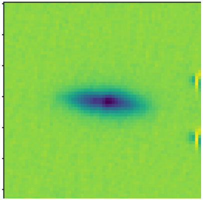
1	1	1
0	1	1
0	0	1

A convolved image

4	3	4
2	4	3
2	3	4

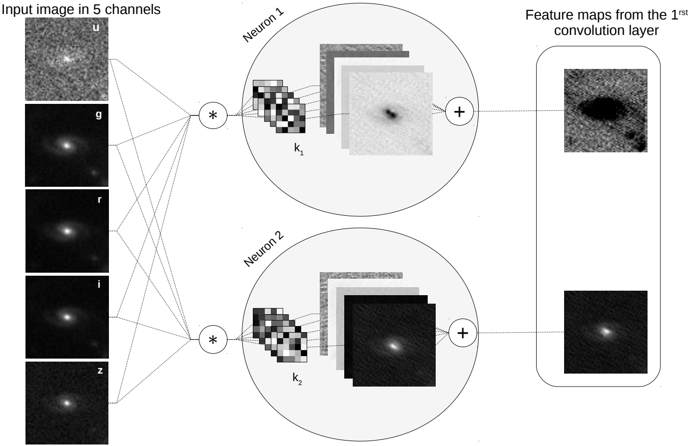






Then introduce non-linearity (tanh, ReLu...)

Convolutions



Pooling

A feature map

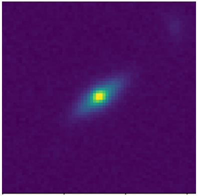
5	1	3	0
0	1	2	7
2	1	1	4
3	1	1	2

Pooling operation

Max in a 2x2 sliding window with a stride of 2

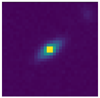
A subsampled feature map

5	7
3	4



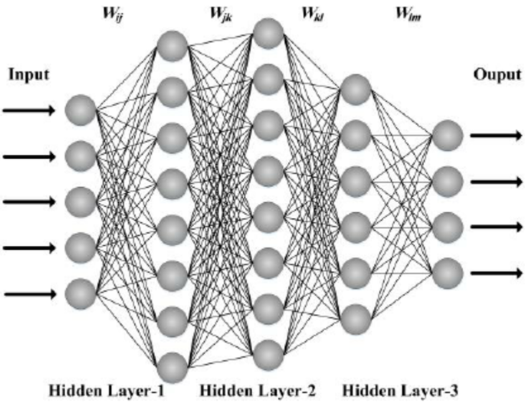
64x64

Max in a 2x2 sliding window with a stride of 2



32x32

Fully connected



Outline

Existing methods

Template fitting:

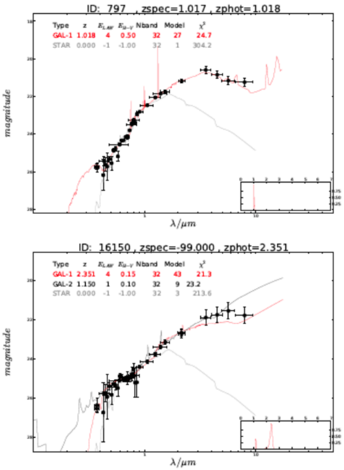
- template fitting

  ⇓

  match the photometry of a galaxy to a suite of templates across a large redshift interval.

  ⇓

  computationally intensive, degeneracies in colour — redshift space can occur



from Hsu et al. (2014)

Existing methods

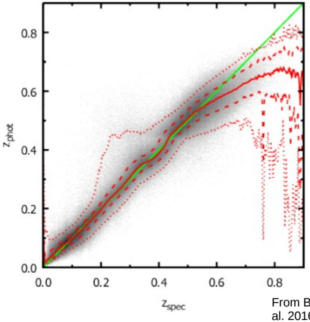
- machine learning

  ⇓

  Use a training database and a set of predefined features (KNN, RF...)

  ⇓

  limited by the training set and the features chosen



From Beck et al. 2016

Our Deep Learning approach

A collaborative work (will be submitted soon):

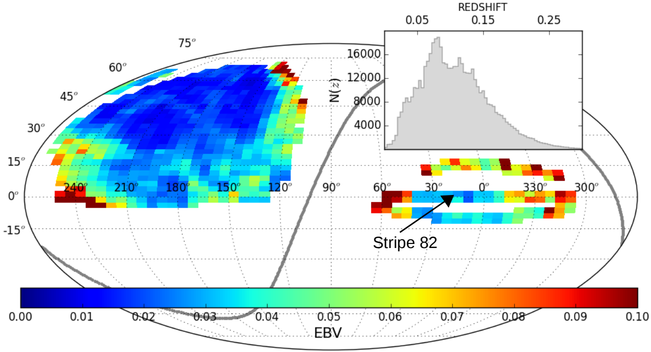
Johanna Pasquet(CPPM), Emmanuel Bertin (IAP), Marie Treyer (LAM), Stephane Arnouts (LAM) and Dominique Fouchez (CPPM)

What we want to improve :

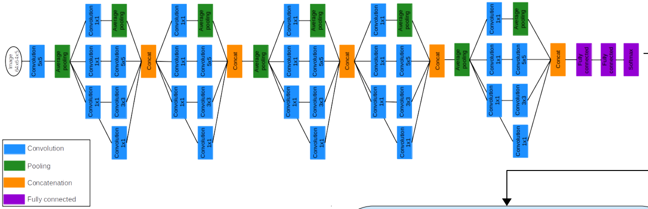
1. Have a well representative and a complete training database with r-band magnitude < 17.8
2. Deliver not only single photoz values but also PDFs
3. Make photo-z estimates immune to IQ variations and contamination by neighbours
4. Optimize our own architecture and not use an existing model
5. Obtain the best performance compared to existing methods!

Outline

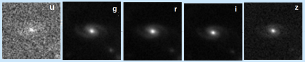
Main Galaxy Sample SDSS



Our architecture

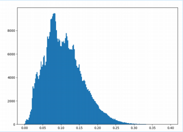


Input images of size 64x64



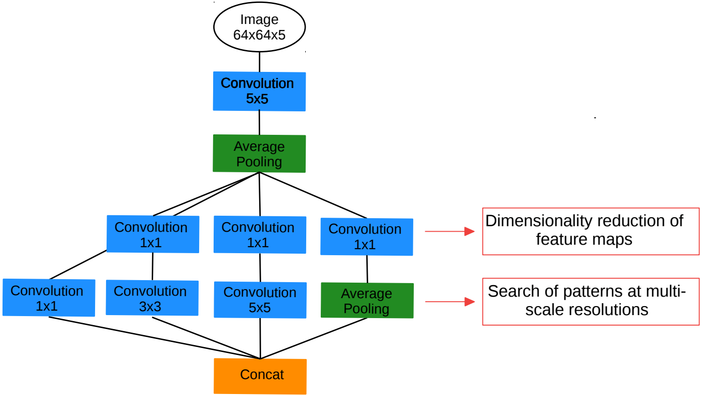
Output probabilities

Vector composed of the probabilities of the photometric redshift to be inside each z-bin

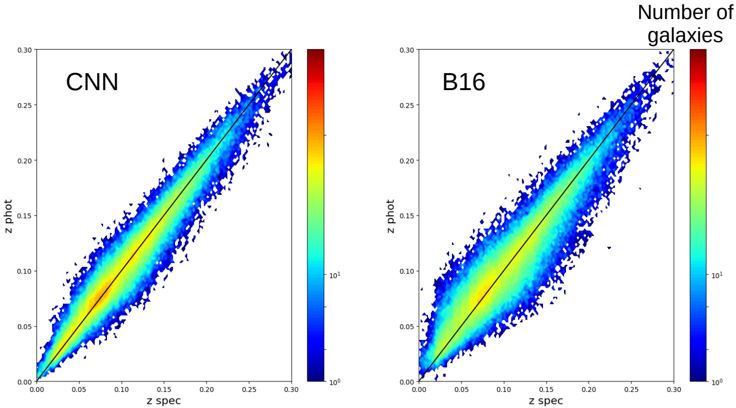


Obtain the z-PDFs
Best prediction of z-phot is the value of the centroid

An Inception block

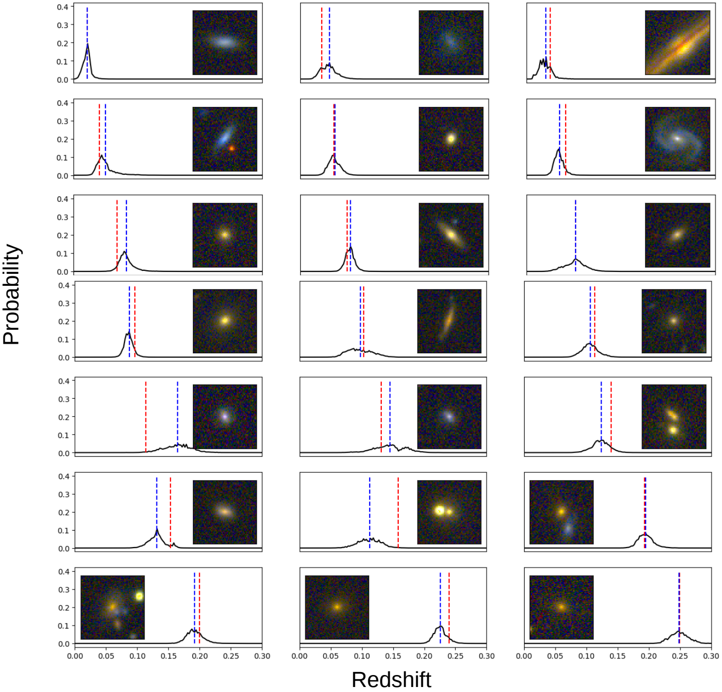


Performance of the method



CNN		B16
$\langle \Delta z \rangle = 1.0 \times 10^{-4}$	← 6 factor improvement	$\langle \Delta z \rangle = 6 \times 10^{-4}$
$\sigma = 9.1 \times 10^{-3}$	← 30 % improvement	$\sigma = 1.3 \times 10^{-2}$
$\eta = 0.31\,\%$	← 4 factor improvement	$\eta = 1.35\,\%$

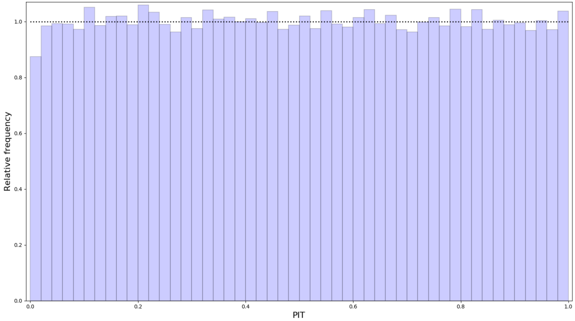
$$\Delta z = (z_{phot} - z_{spec}) / (1 + z_{spec})$$
$$\sigma = 1.4826 \times MAD$$
$$MAD = Median(|\Delta z - Median(\Delta z)|)$$
$$\eta = |\Delta z| > 0.05$$


Probability
0.0
0.1
0.2
0.3
0.4
0.00
0.05
0.10
0.15
0.20
0.25
0.30
Redshift

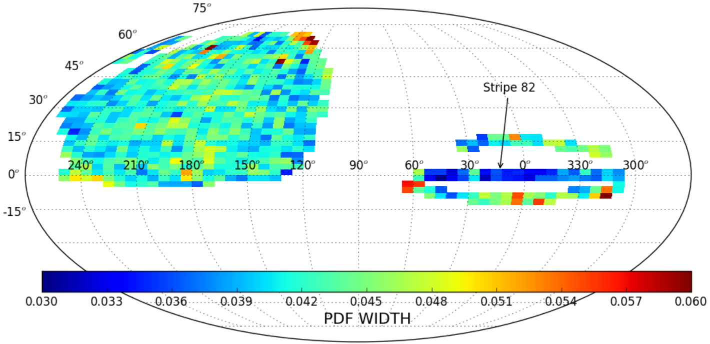
Evaluation of the PDFs

The PIT statistic (Dawid 1984) is based on the histogram of the cumulative probabilities at the true value. For galaxy i with spectroscopic redshift z_i in the test sample :

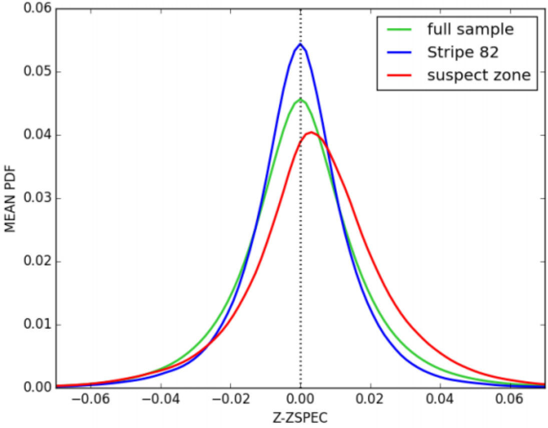
$$\mathrm{PIT}_i = \int_{-\infty}^{z_i} PDF_i(z)dz$$



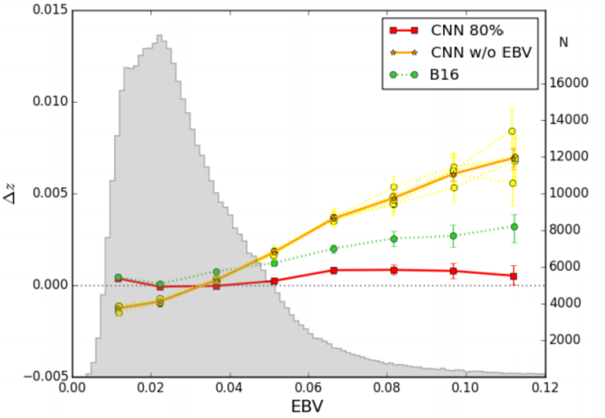
PDF widths


75°
60°
45°
30°
15°
0°
-15°
240°
210°
180°
150°
120°
90°
60°
30°
0°
330°
300°
Stripe 82
0.030
0.033
0.036
0.039
0.042
0.045
0.048
0.051
0.054
0.057
0.060
PDF WIDTH

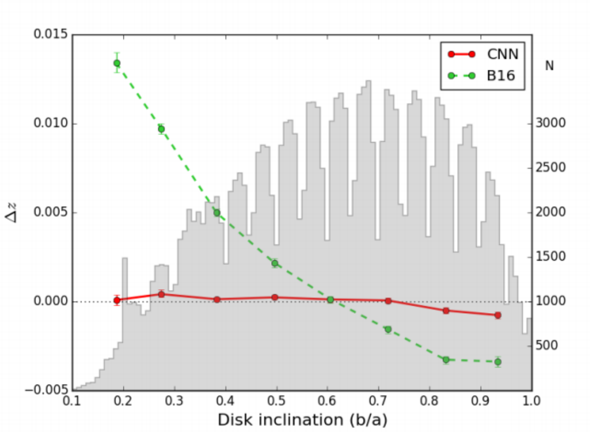
PDF distributions



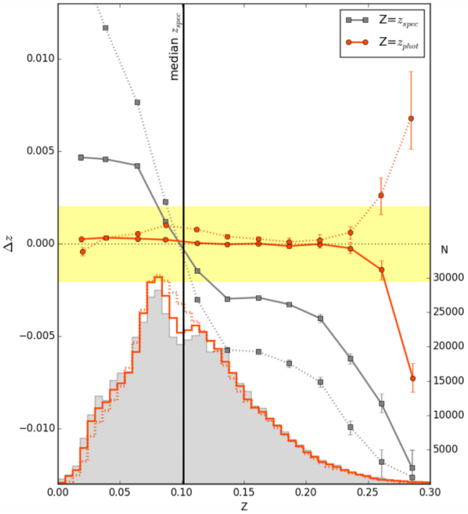
Integrate the reddening into the training



The disk inclination



The bias with spectroscopic redshifts



Summary results

Trial		CNN training sample size	CNN bias	CNN σ	CNN outliers %	CRPS
Training using 80% of the dataset	Full test sample (B16)	393,219	0.00010 (0.00064)	0.00912 (0.01350)	0.31 (1.34)	0.00674
	Widest 20% of PDFs removed	393,219	0.00011	0.00792	0.06	0.00557
	Testing on Stripe 82	393,219	-0.00009	0.00727	0.34	0.00574
	Testing on Stripe 82, widest 20% of PDFs removed	393,219	0.00005	0.00669	0.10	0.00502
Training using 20% of the dataset		99,001	0.00005	0.00917	0.30	0.00679
Training using 2% of the dataset		10,100	-0.00001	0.01440	1.29	0.01013
Training w/o Stripe 82, testing on Stripe 82		486,560	-0.00077	0.00760	0.41	0.00606
Training and testing on Stripe 82		15,771	-0.00002	0.00795	0.38	0.00622

Summary

- Deep Learning methods have emerged in Astronomy for classification tasks from images and light-curves
- In the context of large surveys like **LSST** we need to develop this kind of tool to deal with the **huge** quantity of data
- Our work shows **significant improvements** for:
  - the prediction of photometric redshifts
  - the **zPDFs** that are **well calibrated**
  - no measurable bias with the reddening and the inclination of galaxies

Thank you!

Supported by the OCEVU Labex (ANR-11-LABX-0060) funded by the "Investissements d'Avenir" French government program