Deep Learning methods in the era of Cosmology

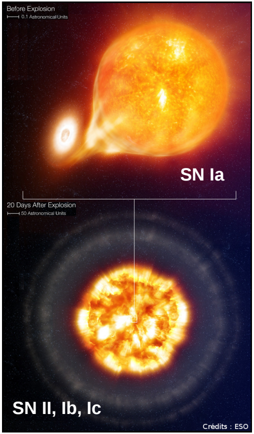
Johanna Pasquet

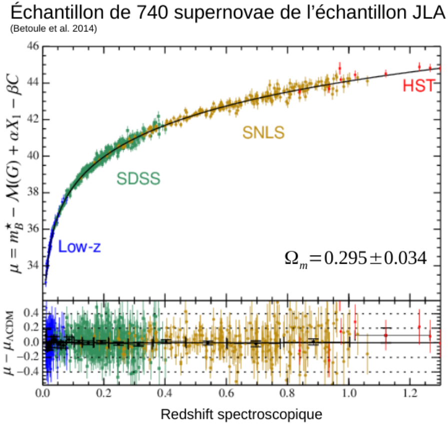
Centre de Physique des Particules de Marseille

Workshop GPU @CC-IN2P3

3 April, 2019

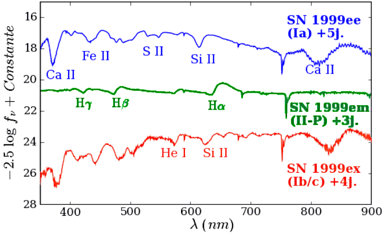
Type Ia supernovae : cosmological probe



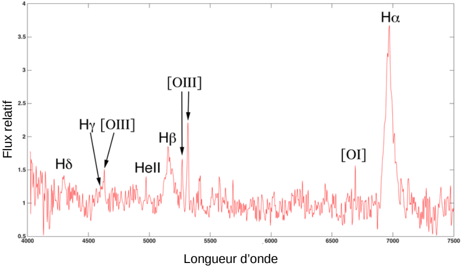


Spectroscopy for the Hubble Diagram

Supernovae identification



Redshift measurement

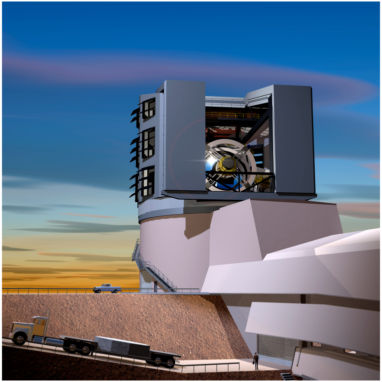


Supernovae Ia

- strong silicon lines
- no H and He lines

A costly method : ∼1h of exposure time for a spectra of a faint supernovae

LSST : the future large photometric survey


Artist view, Credit : Todd Mason, Mason Productions Inc. / LSST Corporation

- billions of observed galaxies
- optimized for supernovae
  - 100 000 supernovae by year

The *Wide Fast Deep program*

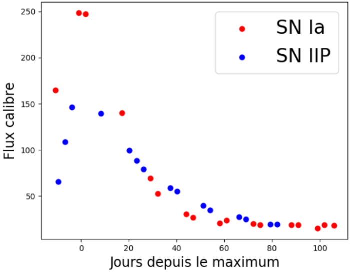
- **large** survey : 18 000 deg^2
- **fast** : 825 visits in 10 years
- **deep** : 27.5 mag in r band

LSST is a technical achievement !

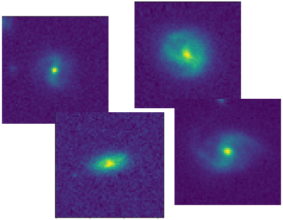
My strategy for the future LSST data

1. Use photometric data

- Identification of SN Ia



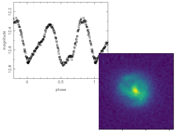
- Photometric redshift measurement



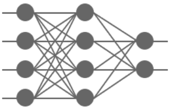
2. My contribution : optimization of *Deep Learning* methods

Deep Learning, a powerful tool to generalize data

Données d'entrée

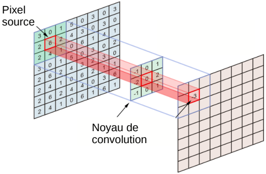


Les caractéristiques sont extraites par l'algorithme

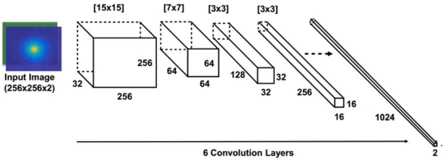


→ Le réseau trouve le meilleur espace de représentativité pour un problème donné

Les convolutions pour extraire des caractéristiques



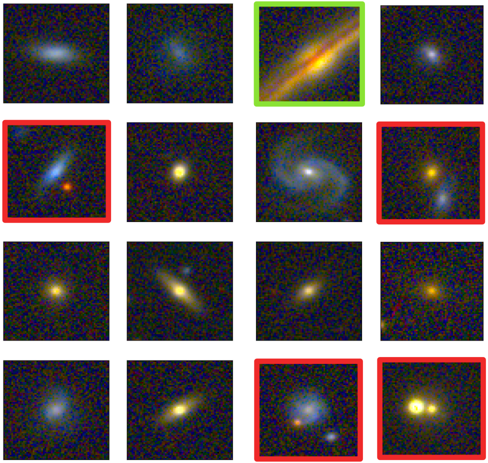
La profondeur pour la généralisation



First application: The estimation of photometric redshift with a deep architecture

J. Pasquet, E. Bertin, M. Treyer, S. Arnouts and D. Fouchez

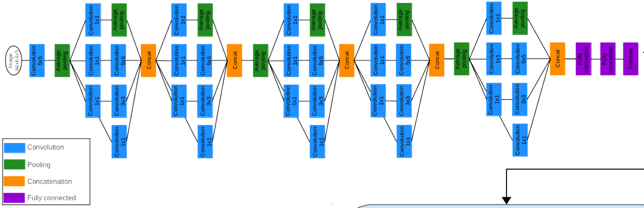
Input SDSS galaxy images transmitted to the CNN



– large galaxies

— crowded images

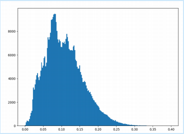
Our architecture



Input images of size 64x64

Output probabilities

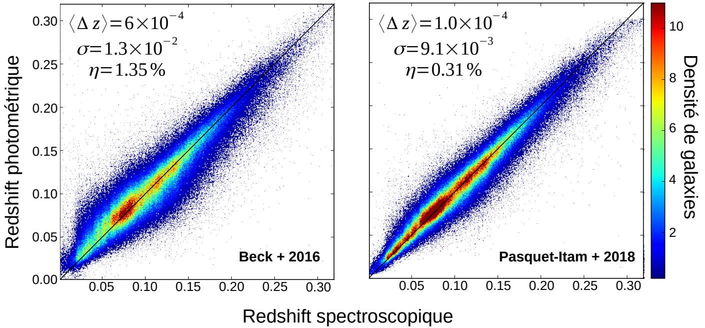
Vector of posterior probabilities for the galaxy redshift to be in a z-bin



The estimator for z-phot is the centroid of the zPDF

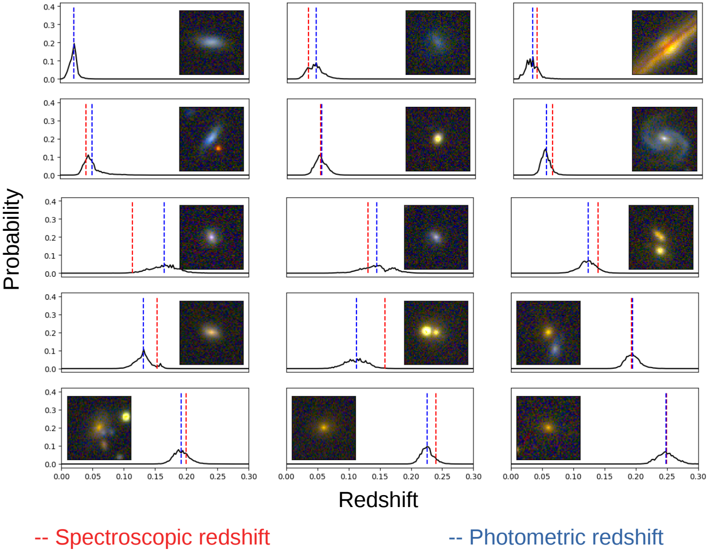
Estimation of galaxy redshifts from SDSS images

- biais Δz is divided by **6**
- dispersion σ is divided by **1.4**
- fraction of catastrophic redshifts η is divided by 4



Photometric redshifts from SDSS images using a Convolutional Neural Network, *Pasquet-Itam et al., A&A, 621 (2019) A26*

Examples of PDFs

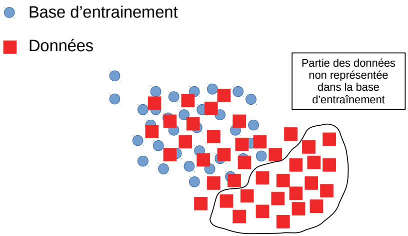


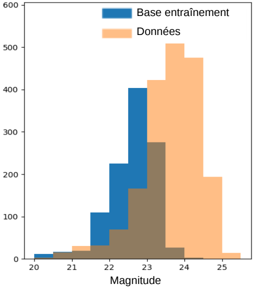
Second application: The classification of light curves of supernovae (SN Ia/ SN Non-Ia)

Johanna Pasquet, Jérôme Pasquet, Marc Chaumont and Dominique Fouchez

Issues for the light curves classification for LSST

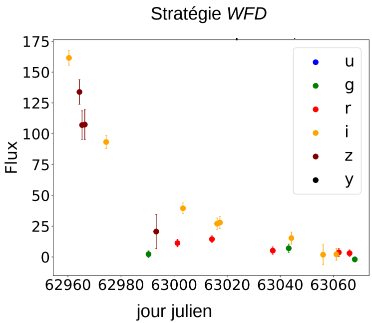
1. The training database is small and limited in flux

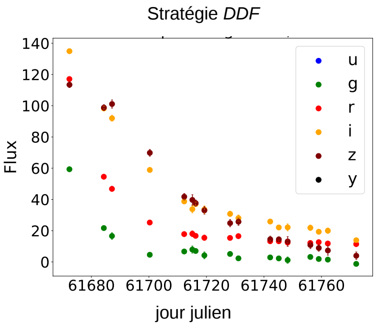




Issues for the light curves classification for LSST

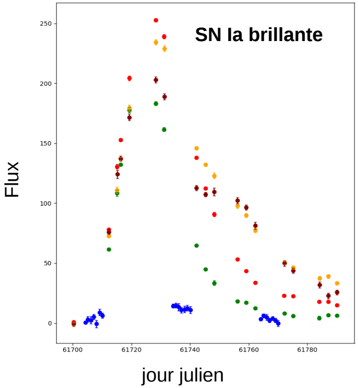
2. Variable sampling due to the observationnal strategy of LSST

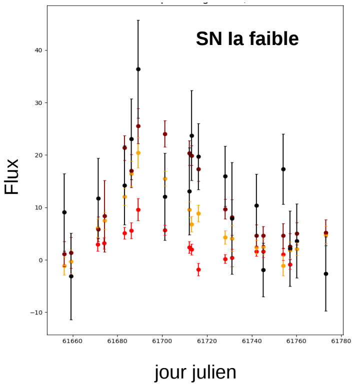




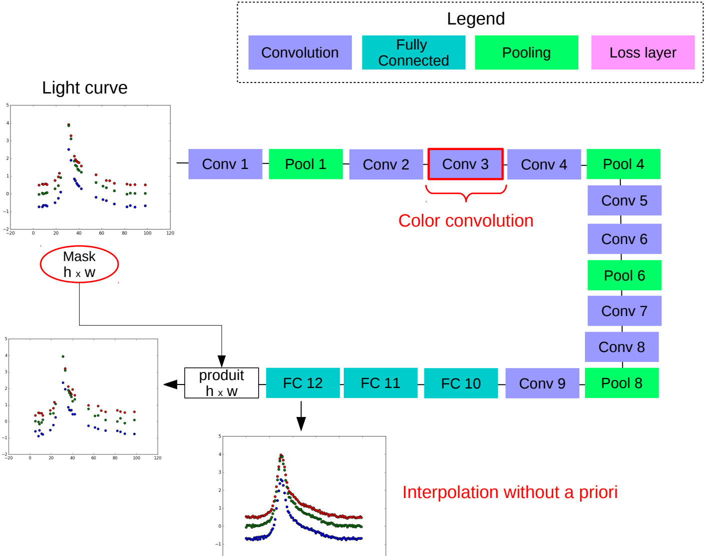
Issues for the light curves classification for LSST

3. Evolution of light curves with redshift

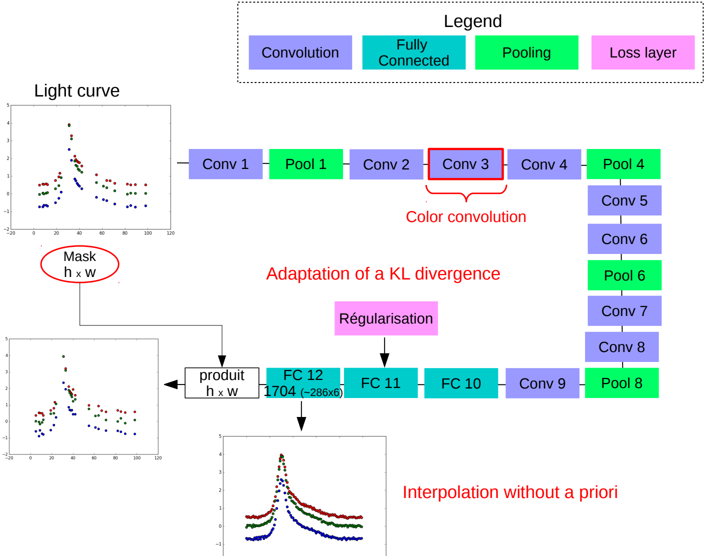




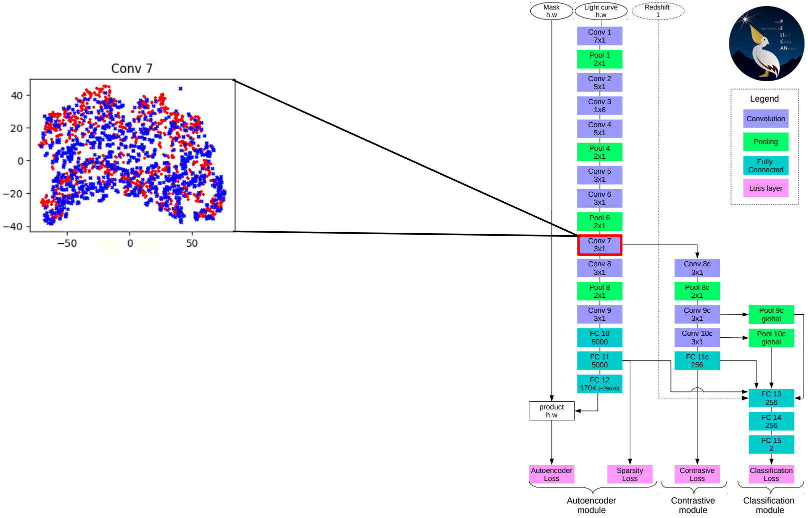
Data interpolation



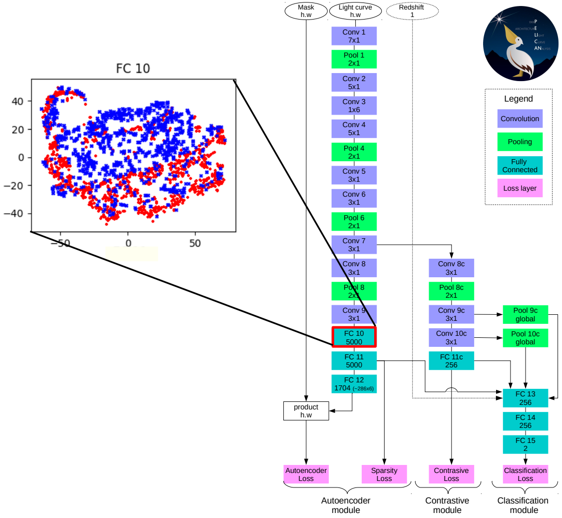
A regularization to manage the irregular sampling



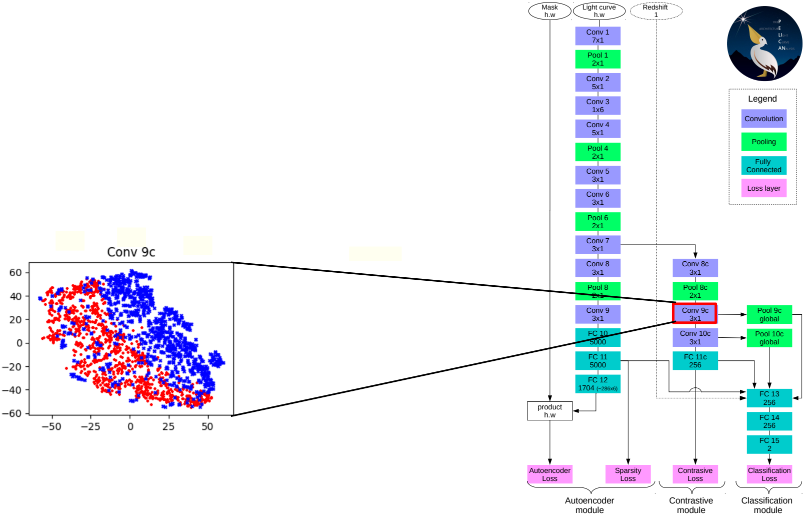
Features visualization



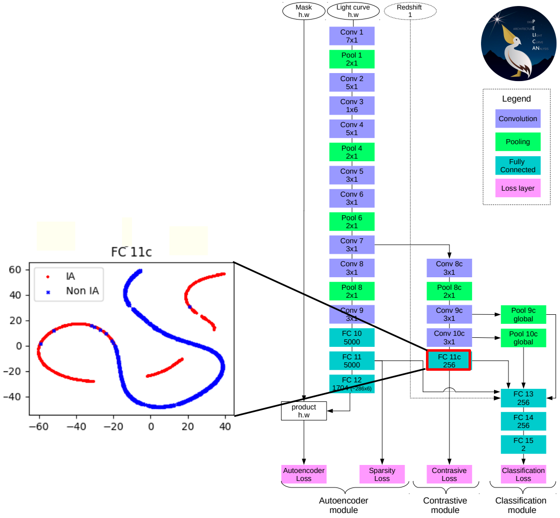
Features visualization



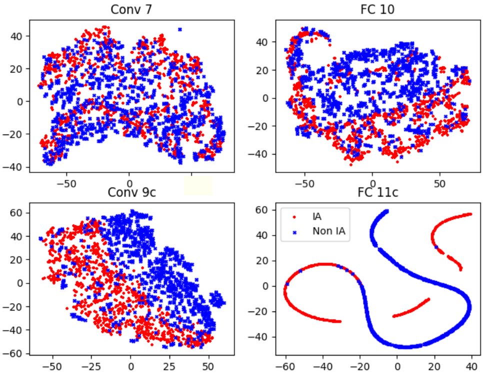
Features visualization

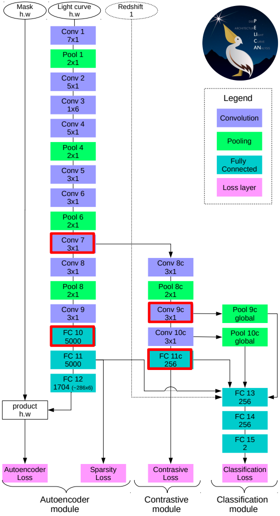


Features visualization



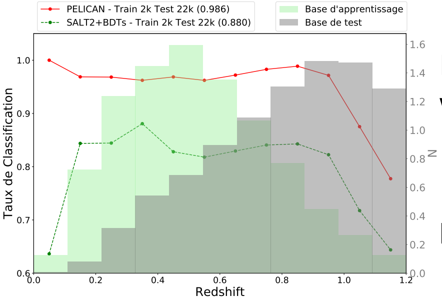
PELICAN


Conv 7
FC 10
Conv 9c
FC 11c
IA
Non IA


Mask h.w
Light curve h.w
Redshift 1
Conv 1 7x1
Pool 1 2x1
Conv 2 5x1
Conv 3 1x6
Conv 4 5x1
Pool 4 2x1
Conv 5 3x1
Conv 6 3x1
Pool 6 2x1
Conv 7 3x1
Conv 8 3x1
Pool 8 2x1
Conv 9 3x1
FC 10 5000
FC 11 5000
FC 12 1704 (~286x6)
product h.w
Conv 8c 3x1
Pool 8c 2x1
Conv 9c 3x1
Conv 10c 3x1
FC 11c 256
Pool 9c global
Pool 10c global
FC 13 256
FC 14 256
FC 15 2
Autoencoder Loss
Sparsity Loss
Contrasive Loss
Classification Loss
Autoencoder module
Contrastive module
Classification module
Legend
Convolution
Pooling
Fully Connected
Loss layer

PELICAN deeP architecturE for the LIght Curve ANalysis

- Classification of light curves from deep fields of LSST



Number of detected SN Ia with a purity > 98%

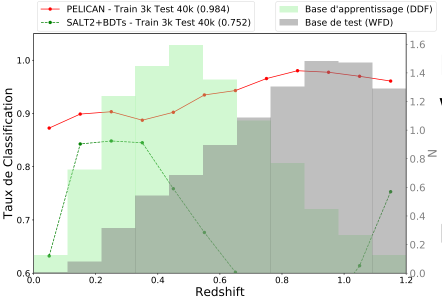
- PELICAN : 85.1%
- LSST baseline : 18.7%

Improvement by a **4.5** factor

PELICAN : deeP architecturE for the LIght Curve ANalysis, Pasquet-Itam et al., accepted to A&A

PELICAN deeP architecturE for the LIght Curve ANalysis

- Classification of light curves from the main survey of LSST



Number of detected SN Ia with a purity > 98%

- PELICAN : 72.9%
- LSST baseline : 0.1%

Major result

PELICAN : deeP architecturE for the LIght Curve ANalysis, Pasquet-Itam et al., accepted to A&A

GPU cards at CPPM and some times

At CPPM:

- 1 machine with a 1080Ti card
- 1 machine with two Titan X cards
- 1 new machine with a Titan V card

Few training times:

- For the redshift application (for **one** network training):
  - 7 hours with a Titan X card
  - We trained 180 networks to produce results for the paper!
- For the light curves classification application (for **one** network training):
  - 5 hours with a Titan X card
  - We trained 140 networks to produce results for the paper!

Conclusion

- The future surveys will deliver multi-band photometry for billions of sources
- Deep Learning methods show performance that outperforms the state of the art results in Cosmology (and other fields of astrophysics)
- With the future amount of data of the future surveys (LSST, Euclid, WFIRST) it is important to develop GPU farm !

Thank you for your attention

Conclusion

- The future surveys will deliver multi-band photometry for billions of sources
- Deep Learning methods show performance that outperforms the state of the art results in Cosmology (and other fields of astrophysics)
- With the future amount of data of the future surveys (LSST, Euclid, WFIRST) it is important to develop GPU farm !

Thank you for your attention

Comparison

For the redshift application (for **one** network training):

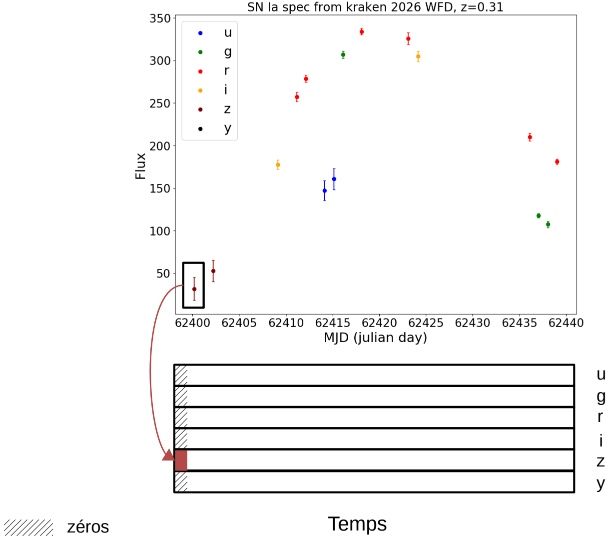
- 7 hours with a Titan X card at CPPM
- 15 hours with a K80 at CC@Lyon (without parallelization)

But we need to train 180 networks to produce results for the paper!
For the light curves classification application (for **one** network training):

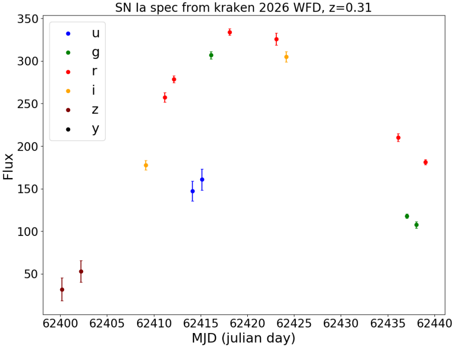
- 1400 light curves per second with a 1080Ti at CPPM (30 hours)
- 2800 light curves per second with two Titan X at CPPM (18 hours)
- 600 light curves per second with two K80 (**with a parallelization!**)

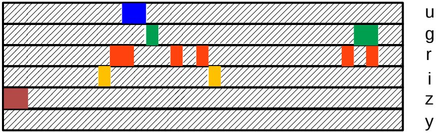
But we need to train 140 networks to produce results for the paper!
In addition these numbers do not take into account the number of experiments needed to test the method in the first stage wich is very large!

Construction LCI



Construction LCI

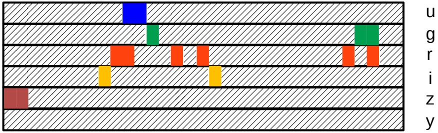




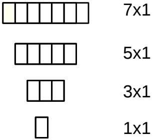
zéros

Temps

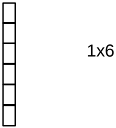
Les convolutions



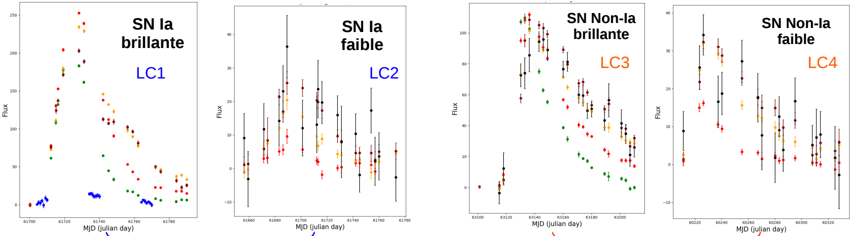
Convolution temporelle Nx1



Convolution par filtre 1xN_{filtre}



Les contrastives



$\|LC1 - LC2\|$

$\|LC3 - LC4\|$

$$max(0, m - \sqrt{\|LC2 - LC3\|})^2$$

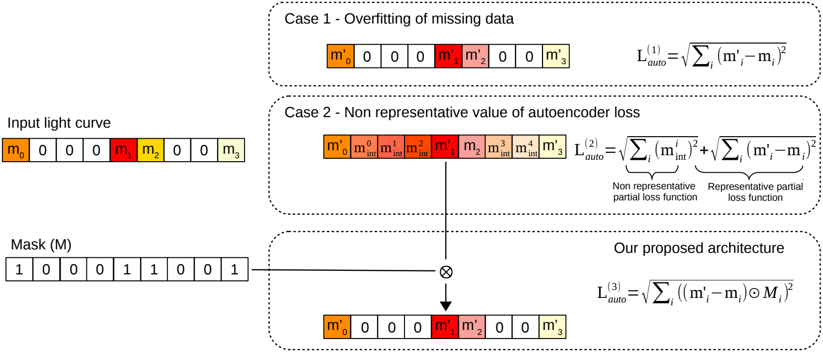
$$max(0, m - \sqrt{\|LC2 - LC4\|})^2$$

$$max(0, m - \sqrt{\|LC1 - LC3\|})^2$$

$$max(0, m - \sqrt{\|LC1 - LC4\|})^2$$

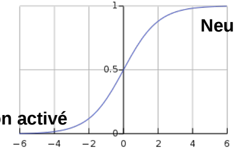
$$\boxed{L = \frac{1}{2} max(0, m - \sqrt{\|LC1 - LC3\|})^2 + \frac{1}{2} max(0, m - \sqrt{\|LC1 - LC4\|})^2 + \frac{1}{2} max(0, m - \sqrt{\|LC2 - LC3\|})^2 + \frac{1}{2} max(0, m - \sqrt{\|LC2 - LC4\|})^2 + \|LC1 - LC2\| + \|LC3 - LC4\|}$$

La fonction de coût de l'autoencoder



La regularisation

1. Utilisation de la fonction d'action Sigmoïde

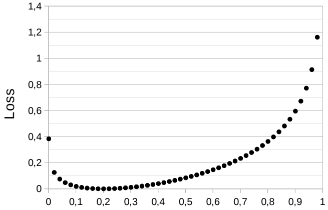


2. Régularisation à l'aide de la divergence de Kullback–Leibler

$$KL(\rho \| \hat{\rho}_j) = \rho \log \frac{\rho}{\hat{\rho}_j} + (1-\rho) \log \left(\frac{1-\rho}{1-\hat{\rho}_j} \right)$$

Activation d'un neurone

Constante

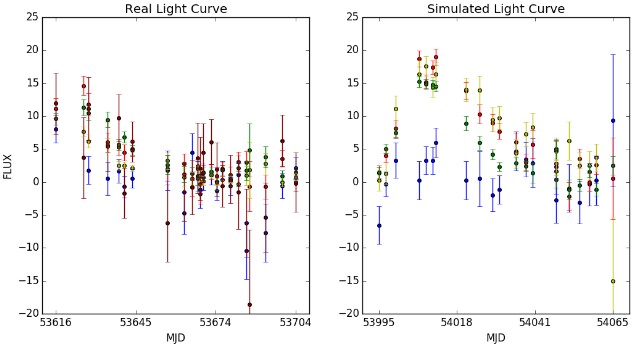


Les comparaisons des résultats LSST

	Training database (spec only)	Test database (phot only)	Accuracy	$Recall_{Ia}$ $Precision_{Ia}$ > 0.95	$Recall_{Ia}$ $Precision_{Ia}$ > 0.98	AUC
DDF	500	1,500	0.849 (0.746)	0.617 (0.309)	0.479 (0.162)	0.937 (0.848)
	2,000	2,000	0.925 (0.783)	0.895 (0.482)	0.818 (0.299)	0.984 (0.882)
	2,000	**22,000**	**0.934 (0.793)**	**0.926 (0.436)**	**0.851 (0.187)**	**0.986 (0.880)**
	10,000	14,000	0.979 (0.888)	0.992 (0.456)	0.978 (0.261)	0.998 (0.899)

	Training database (spec only)	Test database (phot only)	Accuracy	$Recall_{Ia}$ $Precision_{Ia}$ > 0.95	$Recall_{Ia}$ $Precision_{Ia}$ > 0.98	AUC
WFD	DDF Spec : 2, 000	WFD : 15, 000	0.917 (0.650)	0.857 (0.066)	0.485 (0.000)	0.974 (0.765)
	DDF Spec : 3, 000	**WFD : 40, 000**	**0.940 (0.650)**	**0.939 (0.111)**	**0.729 (0.000)**	**0.984 (0.752)**
	DDF Spec : 10, 000	WFD : 80, 000	0.962 (0.651)	0.977 (0.121)	0.889 (0.010)	0.992 (0.760)

Les résultats sur SDSS

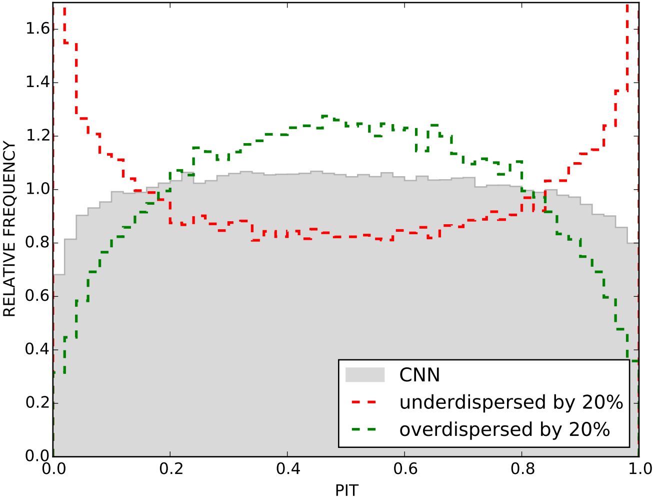


Training database	test database	Accuracy	AUC
SDSS simulations: 219,362	SDSS-II SN confirmed : 582	0.462	0.722
SDSS-II SN confirmed : 80	SDSS-II SN confirmed : 502	0.798	0.586
SDSS simulations : 219,362 SDSS-II SN confirmed : 80	SDSS-II SN confirmed : 502	0.868	0.850

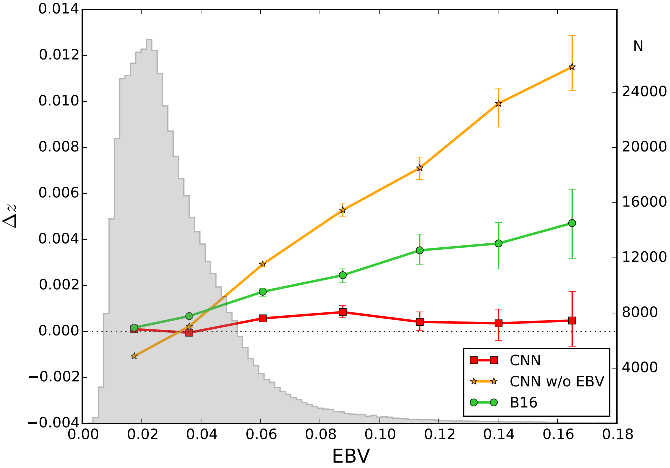
Summary results

Trial	training sample size	bias	σ	η
Training with 80% of the dataset	393,219			
Full test sample		**0.00010**	**0.00912**	**0.31**
(B16)		(0.00062)	(0.01350)	(1.34)
Widest 20% of PDFs		**0.00005**	**0.00789**	**0.06**
Stripe 82 only		-0.00009	0.00727	0.34
Stripe 82 with widest 20% of PDFs removed		**0.00004**	**0.00635**	**0.09**
Training with 50% of the dataset★	250,000	0.00007	0.00910	0.29
Training with 20% of the dataset	**99,001**	**-0.00001**	**0.00914**	0.30
Training with 2% of the dataset	10,100	-0.00017	0.01433	1.26
Training and testing on Stripe 82	**15,771**	-0.00002	**0.00795**	0.38

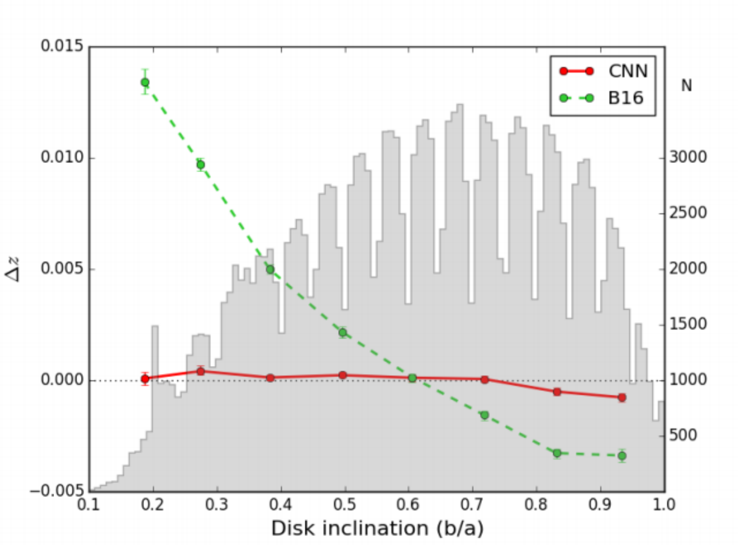
Probability Integral Transform



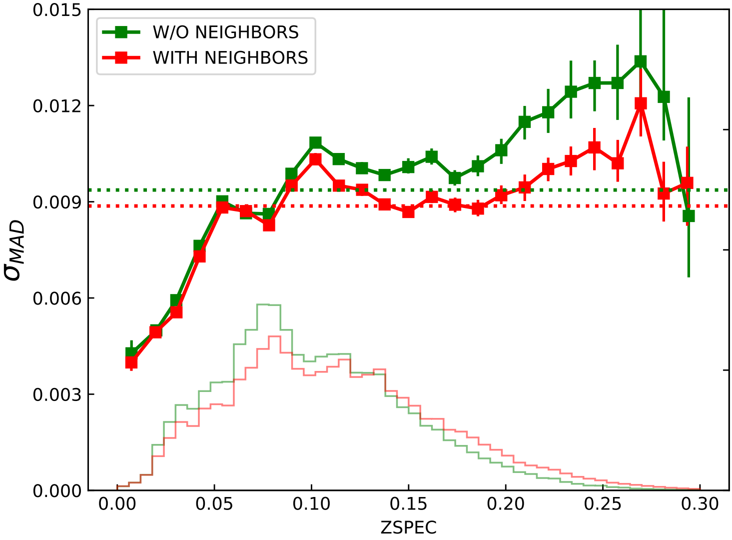
Rougissement



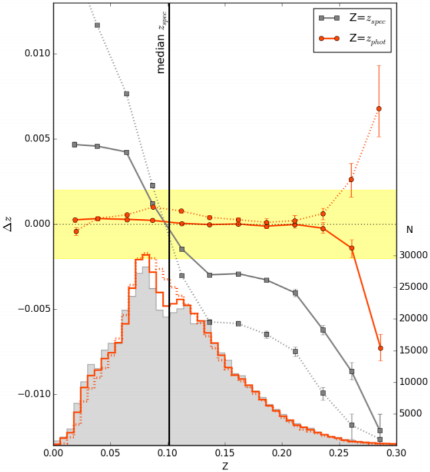
Ellipticité du disque



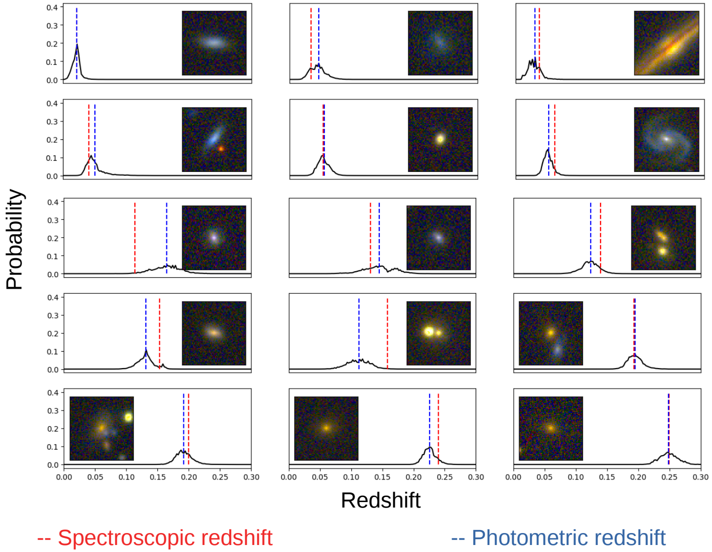
Les voisins


W/O NEIGHBORS
WITH NEIGHBORS
0.015
0.012
0.009
0.006
0.003
0.000
σ_{MAD}
0.00
0.05
0.10
0.15
0.20
0.25
0.30
ZSPEC

Evolution du biais



Exemples de PDFs



Evolution avec le SNR