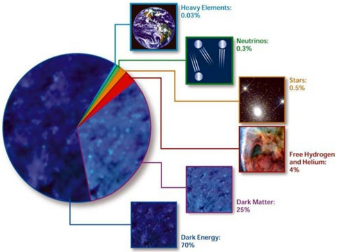
A deep learning approach to observational cosmology with Supernovae

Johanna Pasquet

Centre de Physique des Particules de Marseille

February 18, 2019

Current cosmology questions



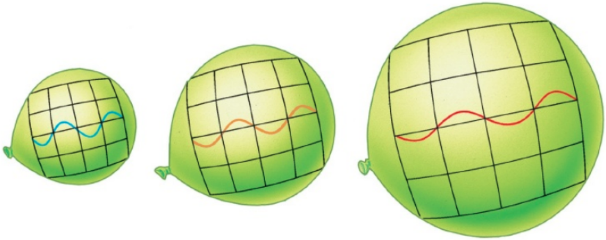
Credit : NASA

- What is the nature of dark matter ?
- What is the nature of dark energy ?
- Is it "dark energy" arising from quantum fluctuations in the vacuum, or is it new gravitational physics ?

Need accurate redshits for cosmology

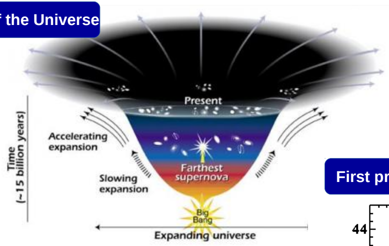
As the universe expands, the radiation is stretched in wavelength

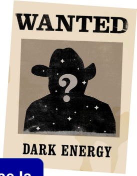
$1 + z = \frac{\lambda_{obs}}{\lambda_{emit}}$



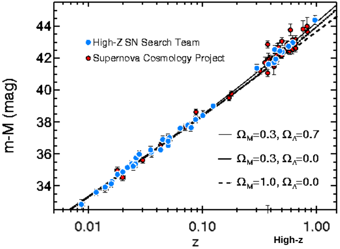
Supernovae Ia as cosmological probe

History of the Universe





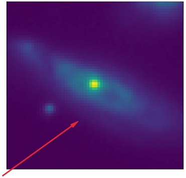
First proof with supernovae Ia



- Dark energy causes the universal expansion to accelerate
- Recent observations of supernovae have produced a value for an acceleration that implies a universe that is about 70 % dark energy

The spectroscopic follow-up to identify SN Ia and measure redshift

Identify and measure the redshift of a galaxy



galaxy

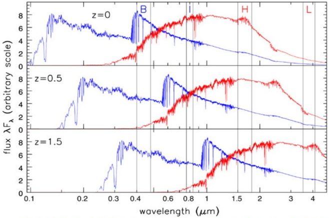
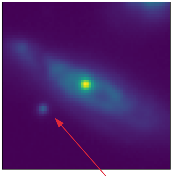
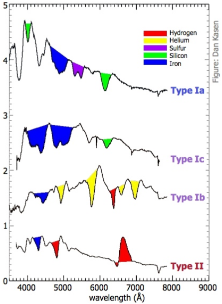


Fig 8.12 (S. Charlot) 'Galaxies in the Universe' Sparke/Gallagher CUP 2007

Determine the nature of an observed object

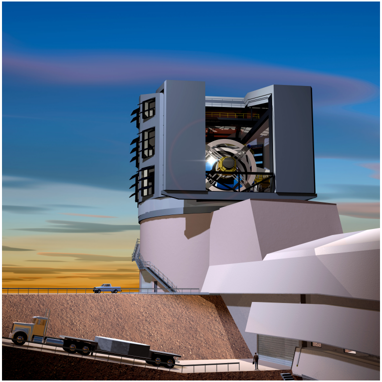


Supernovae



The future image surveys

1. The Large Synoptic Survey Telescope (LSST)

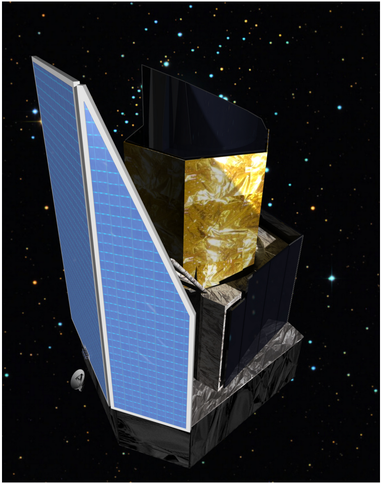


Artist view, Credit : Todd Mason, Mason Productions Inc. / LSST Corporation

- a 10-year survey of the sky
- first light in 2020
- a 8.4-meter special three-mirror design, creating an exceptionally wide field of view, and has the ability to survey the entire sky in only three nights.
- 200 petabyte set of images and data products !

The future image surveys

2. Euclid



- understanding the nature of the source responsible for this acceleration
- slitless spectroscopy
- launch is planned for 2021
- a 6-year survey
- 10 billion sources will be observed !

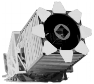
Artist view of the Euclid Satellite -©ESA

The era of Big Data

1924 Henry Draper Catalog (0.2 Million)

1989 Guide Star Catalog (20 Million)

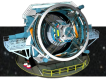
2008 **SDSS (230 Million)**



2018 **Dark Energy Survey (400 Million)**

2027 **Euclid (10 billion)**

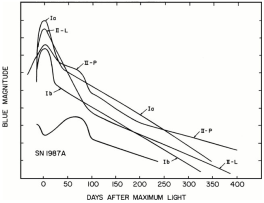
2032 **Large Synoptic Survey Telescope (37 billion)**



A full photometric analysis

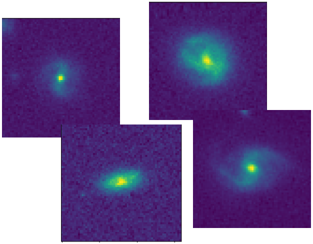
Use all the photometric information in several photometric bands

Light curves of Supernovae



In the LSST context, full photometric SN analyses become crucial

Galaxy images

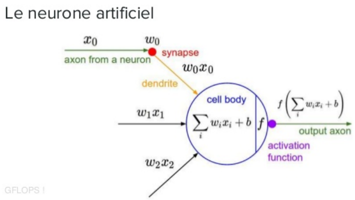


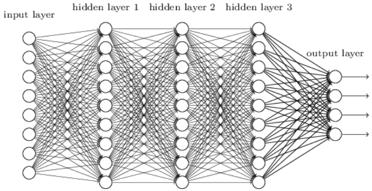
Outline

1. Deep Learning
2. Photometric redshifts
3. Classification of light curves
4. Conclusion

History

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

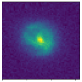




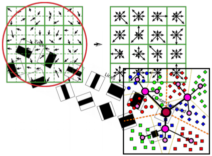
The main property of deep learning

Classical methods

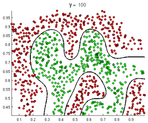
Input data



Feature crafting

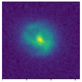


Separation with a classifier

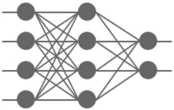


Deep learning

Input data

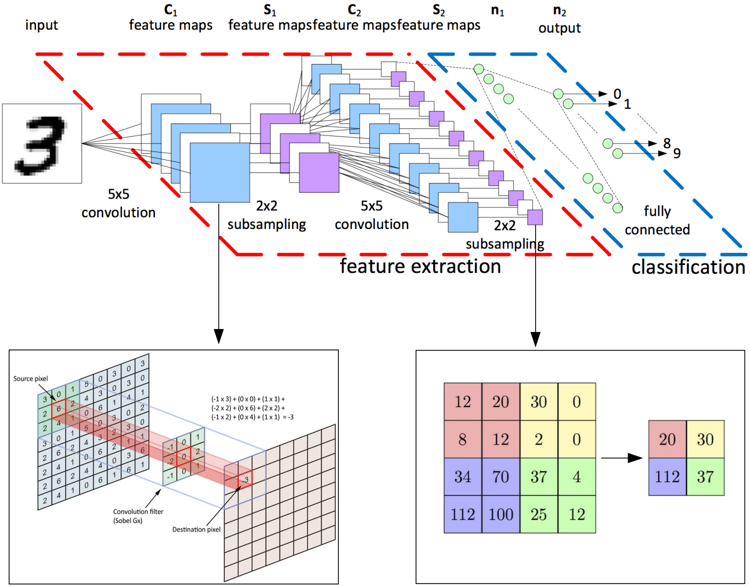


Feature learning



The best feature space representation is found by the network

Typical CNN architecture



Convolutions

An image

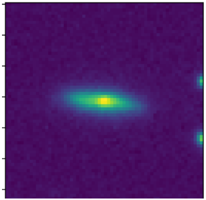
1	1	1	0	0
0	1	1	1	0
0	0	1	1	1
0	0	1	1	0
0	1	1	0	0

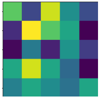
A kernel

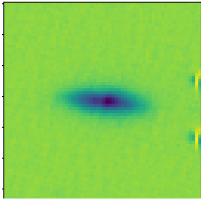
1	1	1
0	1	1
0	0	1

A convolved image

6	5	3
4	6	4
3	4	4

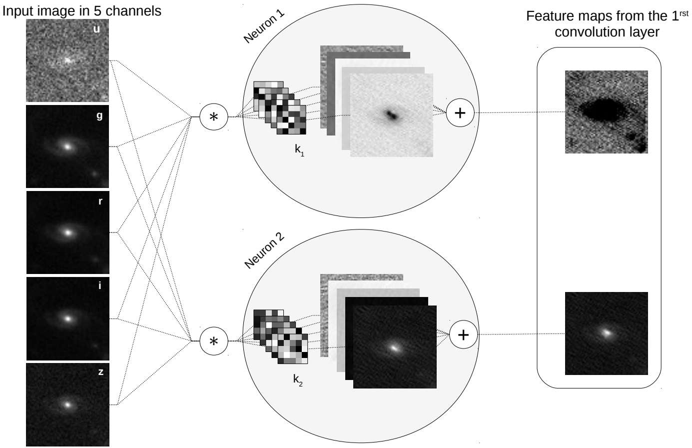






Convolution operation is followed by a non linear function (tanh, ReLu...)

Convolutions

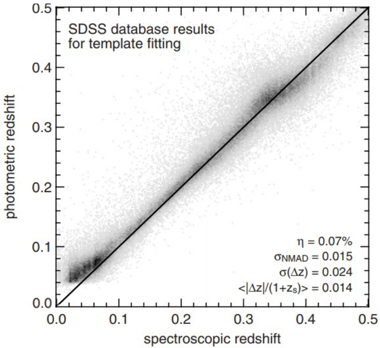


First step: The estimation of photometric redshift with a deep architecture

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

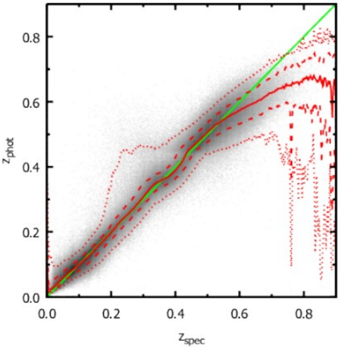
Existing methods

A template fitting method



From Greisel et al. 2013

A machine learning method (KNN)



From Beck et al. 2016

Preliminary results with Deep Learning methods (Hoyle 2016, D'Isanto 2018)

Photometric redshifts from SDSS images using a Convolutional Neural Network (**J. Pasquet**, E. Bertin, M. Treyer, S. Arnouts and D. Fouchez)
A&A, 611 :A97, 2018, **arxiv**: 1806.06607, **code available at**: https://github.com/jpasquet/Photoz

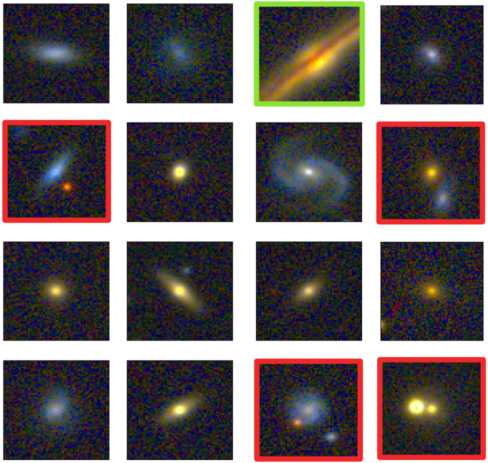
Key elements :

1. A representative and a complete training database with r-band magnitude ≤ 17.8 and redshift, $z \leq 0.4$ (516,525 galaxies)
2. Photoz values + associated Probability Distribution Functions
3. Photoz immune to IQ variations and neighbours contamination
4. A dedicated Neural Network architecture

Results obtained :

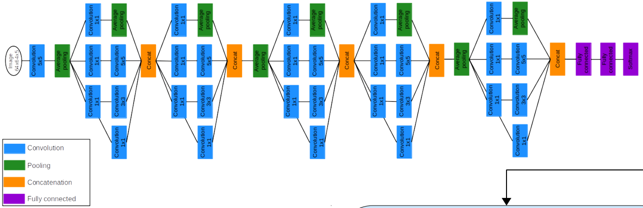
Clear improvements compared to other methods!

Input SDSS galaxy images transmitted to the CNN



– large galaxies — crowded images

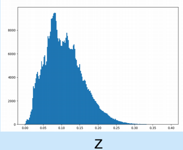
Our architecture



Input images of size 64x64

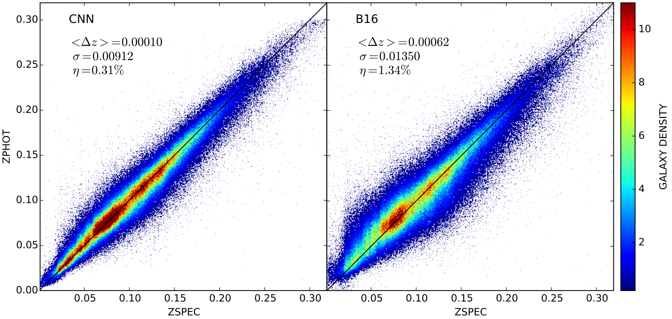
Output probabilities

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

Performance never achieved before!



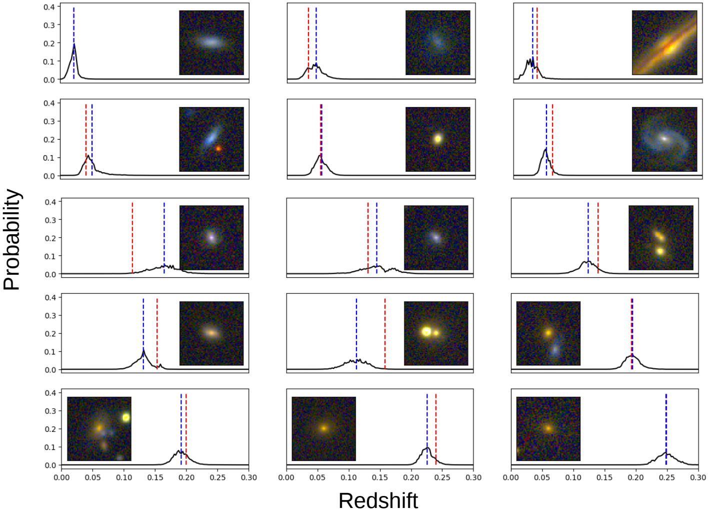
$\langle \Delta z \rangle = 1.0 \times 10^{-4}$ ← Factor of 6 improvement $\langle \Delta z \rangle = 6 \times 10^{-4}$

$\sigma = 9.1 \times 10^{-3}$ ← 30 % improvement $\sigma = 1.3 \times 10^{-2}$

$\eta = 0.31\,\%$ ← Factor of 4 improvement $\eta = 1.35\,\%$

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

Examples of PDFs



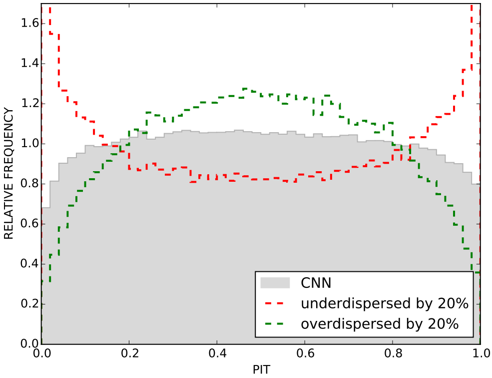
-- Spectroscopic redshift

-- Photometric redshift

Assess the prediction quality of our PDFs

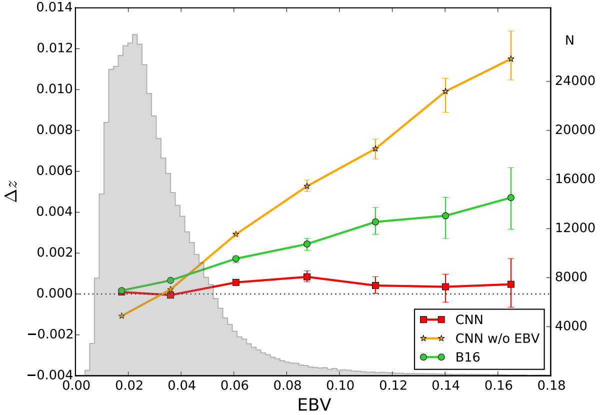
The PIT statistic (Dawid 1984) is based on the histogram of the cumulative probabilities at the true value. For galaxy i with spectroscopic redshift z_i in the test sample :

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



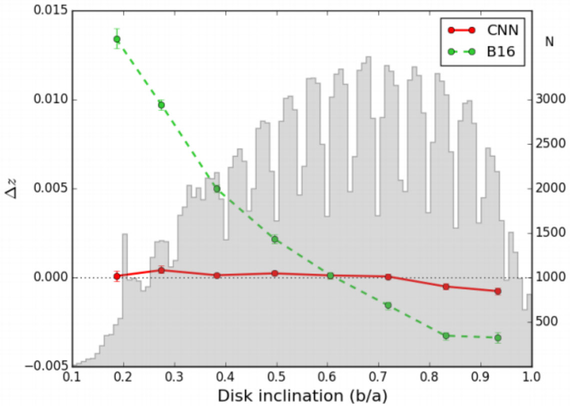
Impact of the extinction of our Galaxy on photometric redshifts

Our method tends to overestimate redshifts in obscured regions (confusing galactic dust attenuation with redshift dimming), unless $E_{(B-V)}$ is used for training



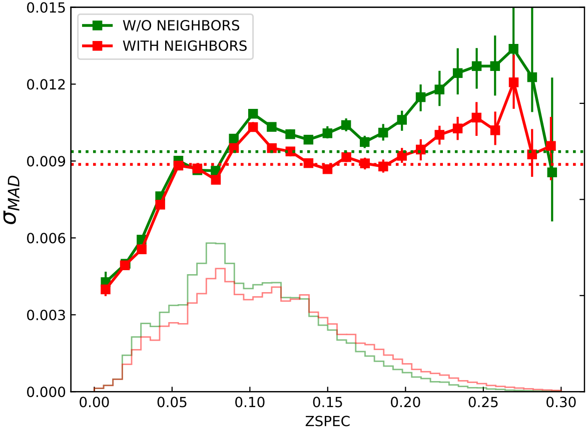
Impact of the disk inclination of galaxies on photometric redshifts

Our method automatically corrects for galactic dust reddening which increases with disk inclination



Neighboring galaxies

The MAD deviation is significantly improved for galaxies with fainter neighbors (43%) compared to those without



Summary results

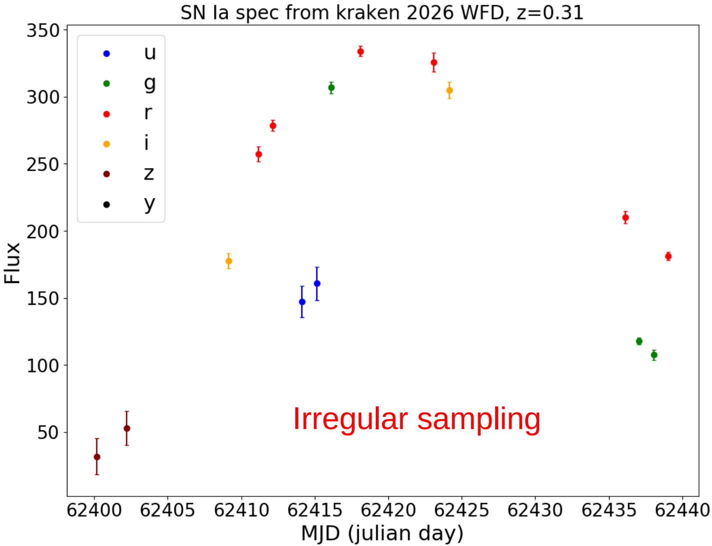
Trial	training sample size	bias	σ	η
Training with 80% of the dataset	393,219			
Full test sample		**0.00010**	**0.00912**	**0.31**
(B16)		(0.00062)	(0.01350)	(1.34)
Widest 20% of PDFs		**0.00005**	**0.00789**	**0.06**
Stripe 82 only		-0.00009	0.00727	0.34
Stripe 82 with widest 20% of PDFs removed		**0.00004**	**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

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

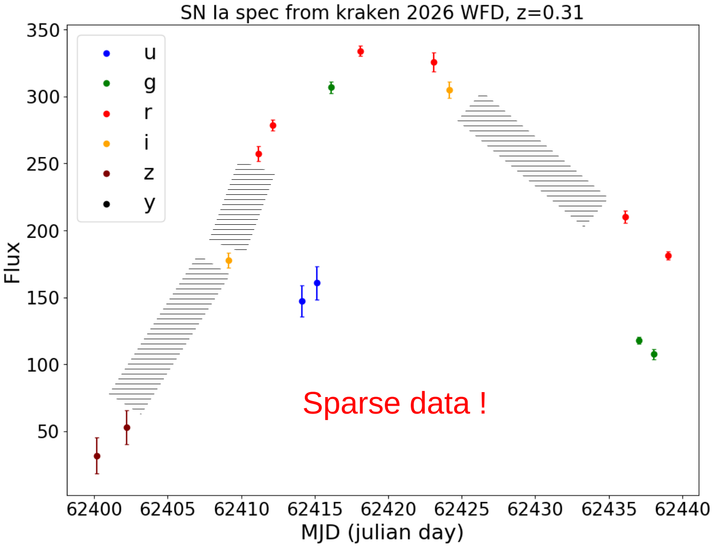
Johanna Pasquet, Jérôme Pasquet, Marc Chaumont and Dominique Fouchez

DEEP
ARCHITECTURE
LIGHT
CURVE
ANALYSIS

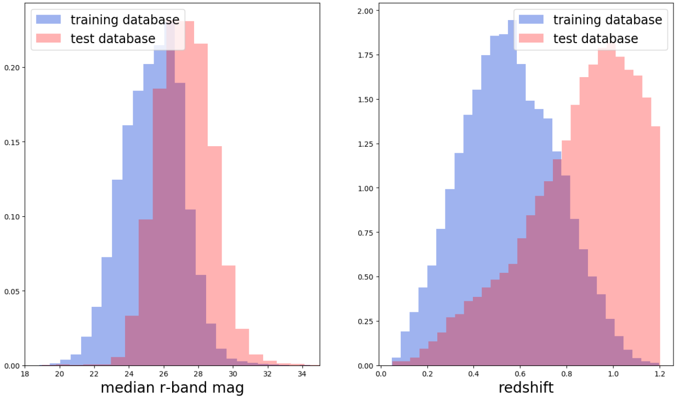
Sparse data


SN Ia spec from kraken 2026 WFD, z=0.31
u
g
r
i
z
y
Flux
MJD (julian day)
350
300
250
200
150
100
50
62400
62405
62410
62415
62420
62425
62430
62435
62440
Irregular sampling

Sparse data


SN Ia spec from kraken 2026 WFD, z=0.31
u
g
r
i
z
y
Flux
350
300
250
200
150
100
50
62400
62405
62410
62415
62420
62425
62430
62435
62440
MJD (julian day)
Sparse data !

A testing database not representative in flux

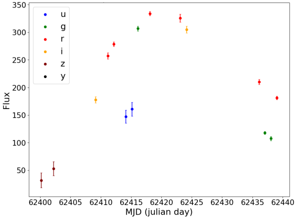


The non-representativeness of the databases, which is a problem of mismatch, is critical for machine learning process.

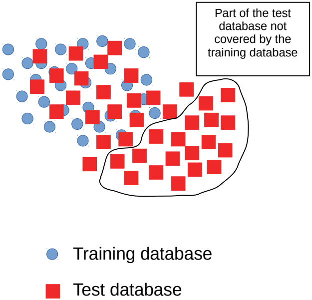
To sum up

Small training databases

Data can be sparse with an irregular sampling

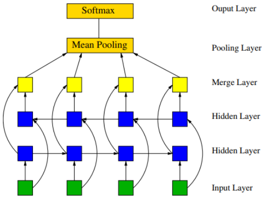


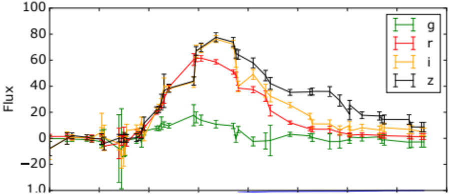
Non-representativeness between the training and the test databases



What deep learning method should we adopt?

- Recurrent neural network: suited to time series



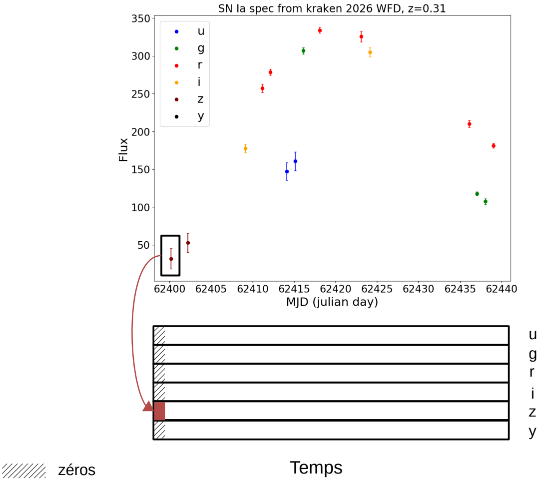


Charnock + 2017

$\Rightarrow$ Interpolation of data can biais the learning
$\Rightarrow$ Performance comparable to classical method

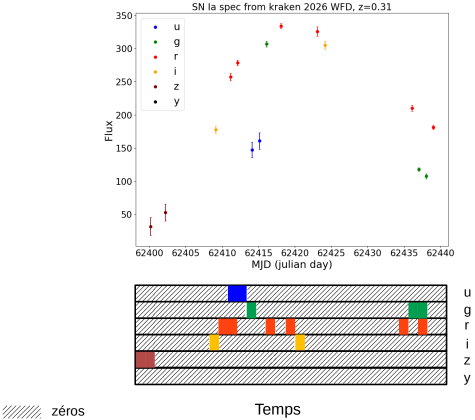
Convolutional neural network

1. Transform input lignt curves into images : Light Curve Images (LCI)



Convolutional neural network

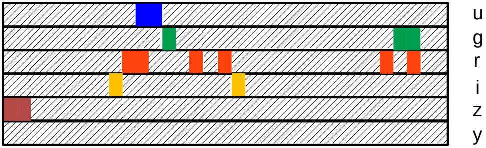
1. Transform input light curves into images : Light Curve Images (LCI)

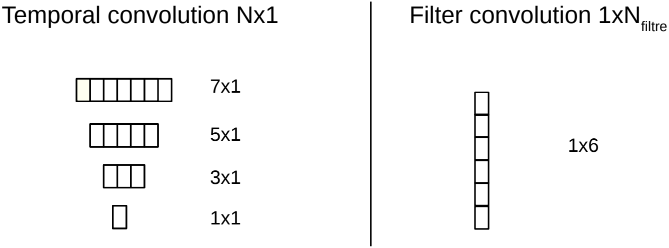


Overfitting of missing data (zero values)

Convolutional neural network

1. Transformer les courbes de lumière en image: les Light Curve Images (LCI)
2. Adapt convolution operations





PELICAN: a deeP architecturE for the LIght Curve ANalysis

J Pasquet et al. accepted to A&A after minor corrections, arXiv:1901.01298

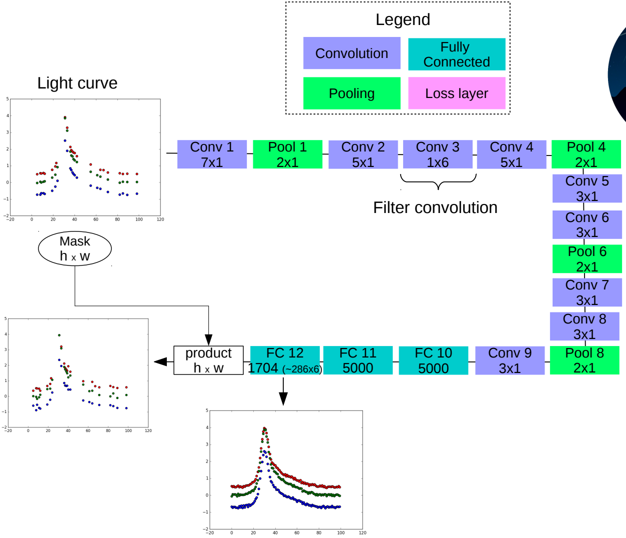
Key elements :

1. a complex Deep Learning architecture to classify light curves of supernovae
2. trained on a small and biased training database
3. overcome the problem of non-representativeness between the training and the test databases
4. deal with the sparsity of data and the difference of sampling and noise

The ability of PELICAN to deal with the different causes of non-representativeness between the training and test databases, and its robustness against survey properties and observational conditions, put it on the forefront of the light curves classification tools for the LSST era.

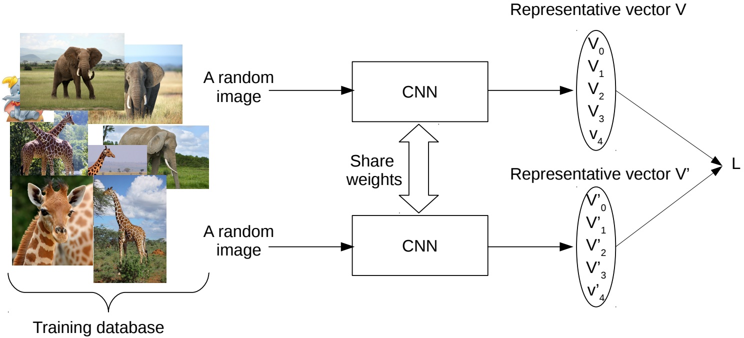
How can we obtain representative features from the test database?

An autoencoder for light curves



How can we make the algorithm understand that a bright and a faint SN is the same object?

The contrastive loss



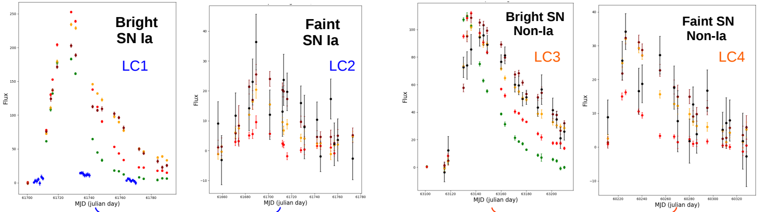
L = (1-Y) x D² + Y x max(0, m – D)²

$$L = (1-Y) \times D^2 + Y \times \max(0, m - D)^2$$

Y : 0, if two images have the same label (elephant-elephant ou giraffe-giraffe)

Y : 1, if two images have different label (elephant-giraffe ou giraffe-elephant)

The contrastive loss applied to light curves



$$\|LC1 - LC2\| \qquad \|LC3 - LC4\|$$

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

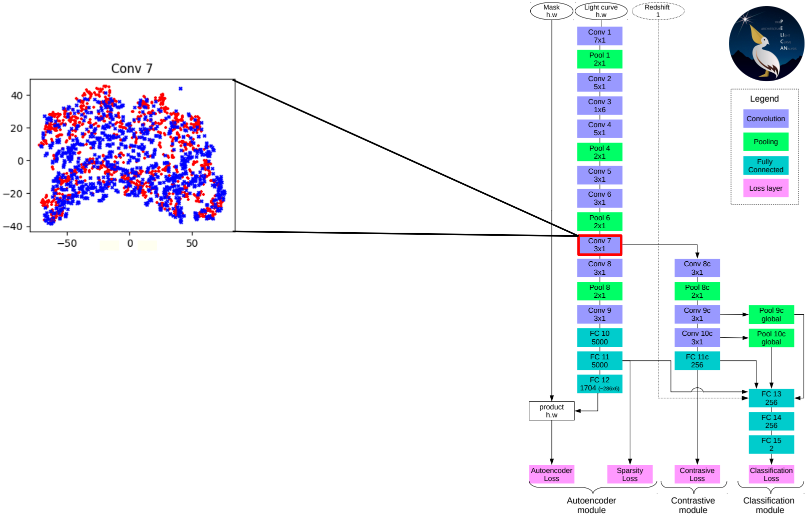
$$max\left(0, m-\sqrt{\|LC2-LC4\|}\right)^2$$

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

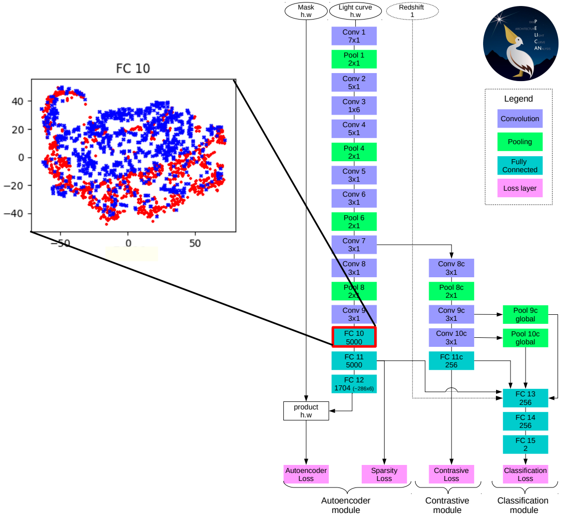
$$max\left(0, m-\sqrt{\|LC1-LC4\|}\right)^2$$

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

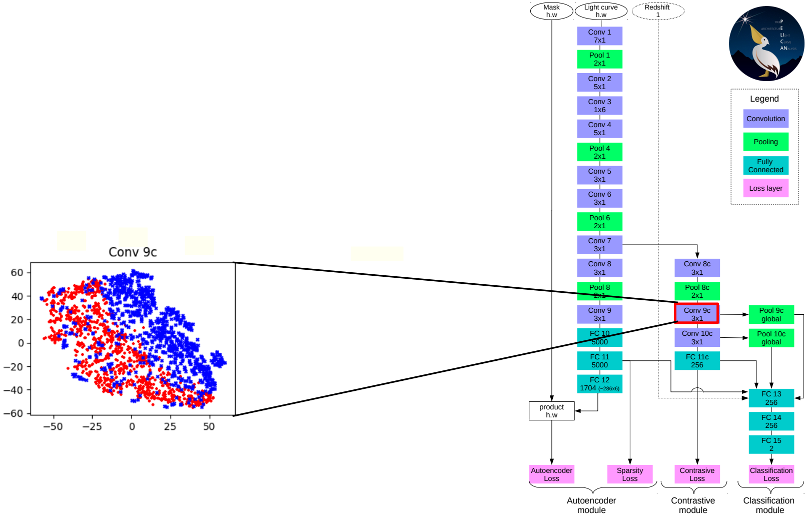
PELICAN


Conv 7
40
20
0
−20
−40
−50
0
50
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

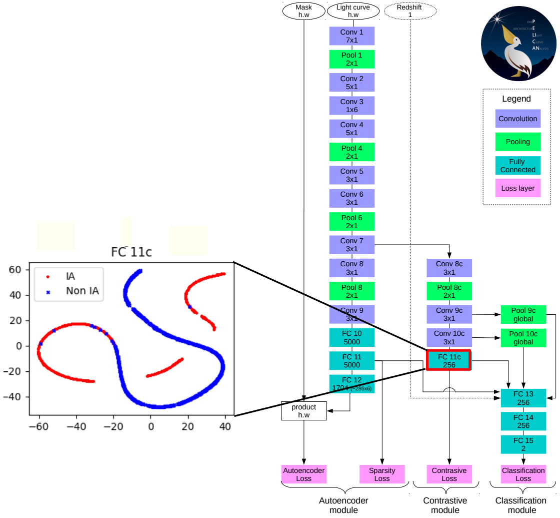
PELICAN


FC 10
40
20
0
−20
−40
−50
0
50
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

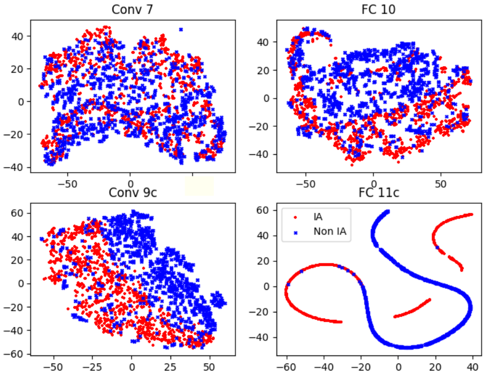
PELICAN

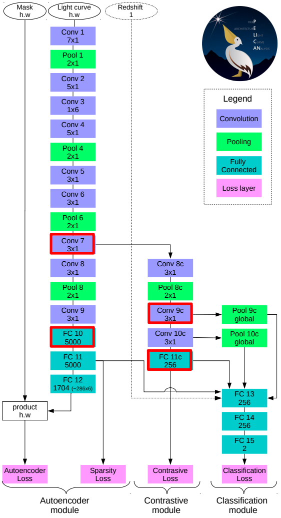

Conv 9c
Mask h.w
Light curve h.w
Redshift 1
Conv 1 7x1
Pool 1 2x1
Conv 2 5x1
Conv 3 1x6
Conv 4 5x1
Pool 4 2x1
Conv 5 3x1
Conv 6 3x1
Pool 6 2x1
Conv 7 3x1
Conv 8 3x1
Pool 8 2x1
Conv 9 3x1
FC 10 5000
FC 11 5000
FC 12 1704 (~286x6)
product h.w
Conv 8c 3x1
Pool 8c 2x1
Conv 9c 3x1
Conv 10c 3x1
FC 11c 256
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

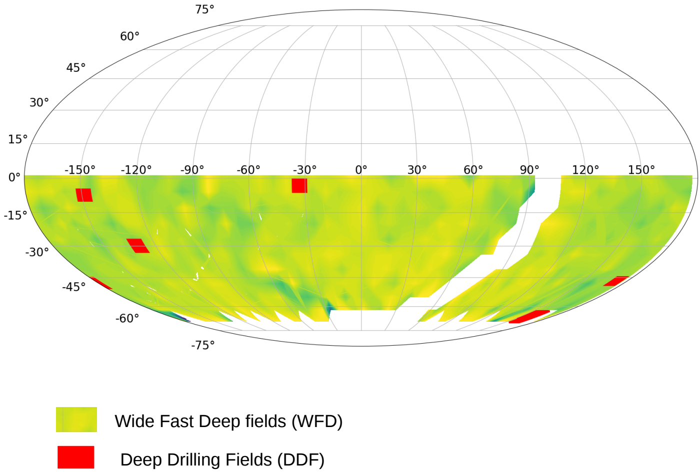

FC 11c
IA
Non IA
60
40
20
0
−20
−40
−60
−40
−20
0
20
40
Mask
h.w
Light curve
h.w
Redshift
1
Conv 1
7x1
Pool 1
2x1
Conv 2
5x1
Conv 3
1x6
Conv 4
5x1
Pool 4
2x1
Conv 5
3x1
Conv 6
3x1
Pool 6
2x1
Conv 7
3x1
Conv 8
3x1
Pool 8
2x1
Conv 9
3x1
FC 10
5000
FC 11
5000
FC 12
1704 (=286x6)
product
h.w
Conv 8c
3x1
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


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

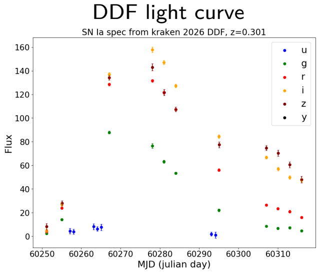
The main survey and the deep fields of LSST

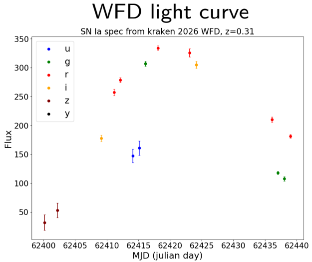


LSST simulated data

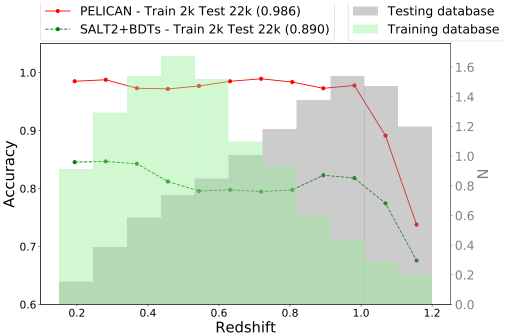
Two methodologies:

1. A training and a test on deep fields (DDF)
2. A training on deep fields and a test on the main survey (WFD)



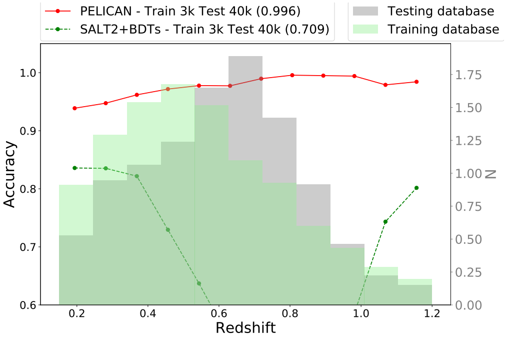


Results on DDF



	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)

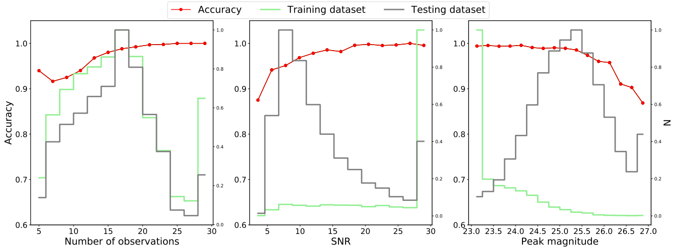
Results on WFD



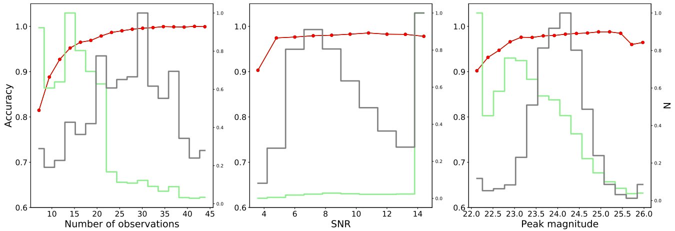
	Training database (spec only)	Test database (phot only)	Accuracy	Recall_{Ia} $\text{Precision}_{Ia} > 0.95$	Recall_{Ia} $\text{Precision}_{Ia} > 0.98$	AUC
WFD	DDF Spec : 2, 000	WFD : 15, 000	0.917 (0.650)	0.857 (0.066)	0.485 (0.000)	0.974 (0.765)
	DDF Spec : 3, 000	**WFD : 40, 000**	**0.940 (0.650)**	**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)

Further analysis of the behaviour of PELICAN

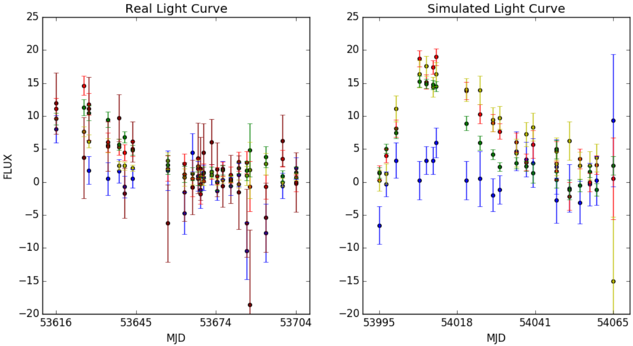
DDF



WFD



SDSS data



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

- The future surveys will deliver multi-band photometry for billions of sources
- Many issues for the classification algorithms
- Promising results for the estimation of photometric redshifts
- Performance never achived for the classification of light curves by considering a non-representative training database

Perspectives

- Estimate photometric redshift from light curves with PELICAN
- Propagation of uncertainties due to the photometric redshift of the supernovae and the host galaxy and the classification errors in the Hubble diagram

Thank you for your attention!

Summary

- The future surveys will deliver multi-band photometry for billions of sources
- Many issues for the classification algorithms
- Promising results for the estimation of photometric redshifts
- Performance never achived for the classification of light curves by considering a non-representative training database

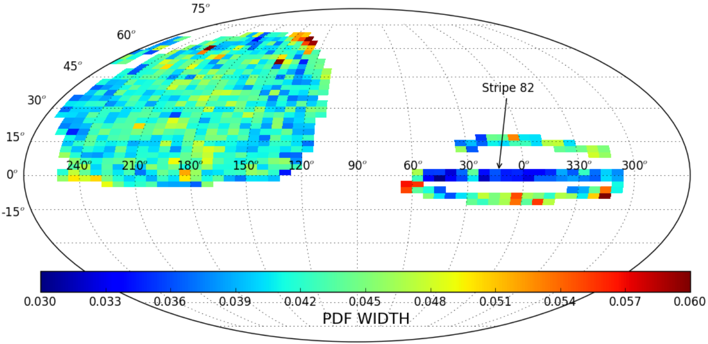
Perspectives

- Estimate photometric redshift from light curves with PELICAN
- Propagation of uncertainties due to the photometric redshift of the supernovae and the host galaxy and the classification errors in the Hubble diagram

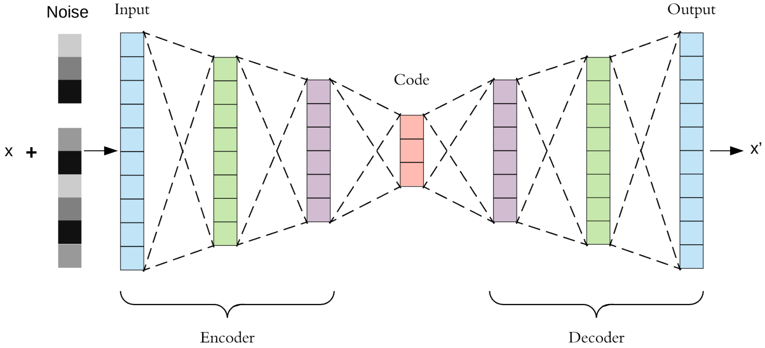
Thank you for your attention!

pact of Signal-to-Noise Ratio (SNR) on widths of PDFs

The Stripe 82 region, which combines repeated observations of the same part of the sky, gives us the opportunity to look into the impact of SNR



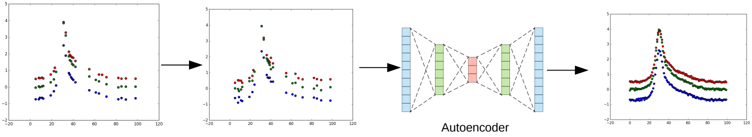
Autoencodder



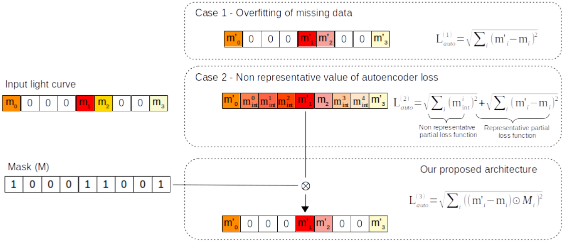
Fonction de perte = $|| x - x' ||_2$

Autoencodder

Courbe de lumière — Courbe de lumière bruitée — Courbe de lumière reconstruite



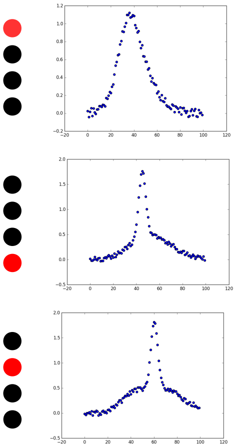
Comment calculer une fonction de perte cohérente ?

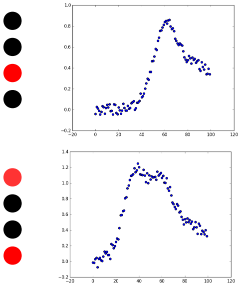


Autoencodder

FC 11
5000

Neurone non activé (=0)
Neurone activé (=1)

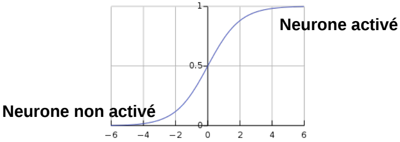




Comment forcer le réseau à activer / désactiver certains neurones ?

Autoencodder

1. Utilisation de la fonction d'action Sigmoïde

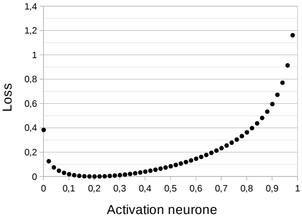


2. Régularisation à l'aide de la divergence de Kullback–Leibler

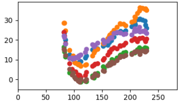
$$KL(\rho \| \hat{\rho}_j) = \rho \log \frac{\rho}{\hat{\rho}_j} + (1-\rho) \log \left(\frac{1-\rho}{1-\hat{\rho}_j} \right)$$

Activation d'un neurone

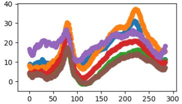
Constante



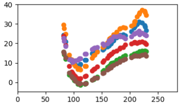
Autoencodder



Courbe de lumière
en entrée

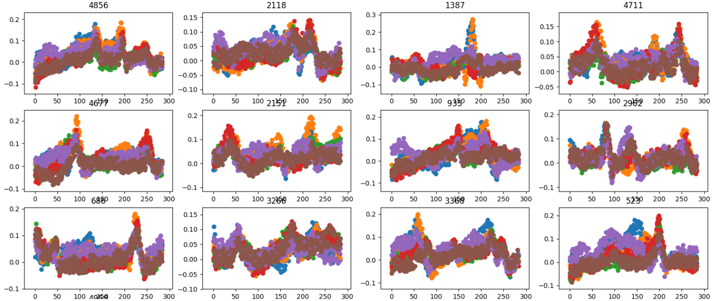


Courbe de lumière
reconstruite



Courbe de lumière
reconstruite multipliée
par le masque

Activation de 12 neurones :



=> Sur 5 000 neurones un nombre très réduit (entre 10 et 30) s'active avec un score supérieur à 0.2