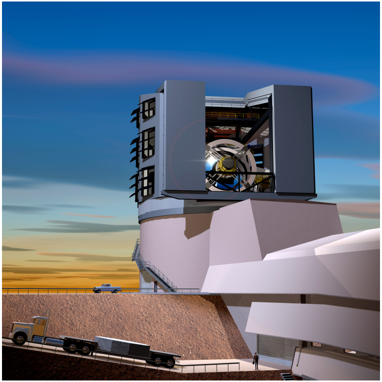
PELICAN: a deeP architecturE for the LIght Curve ANalysis

Johanna Pasquet, Jérôme Pasquet, Marc Chaumont and Dominique Fouchez

Centre de Physique des Particules de Marseille

March 15, 2019

The Large Synoptic Survey Telescope (LSST)



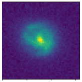
Artist view, Credit : Todd Mason, Mason Productions Inc. / LSST Corporation

- 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 !

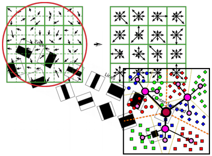
The main property of deep learning

Classical methods

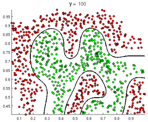
Input data



Feature crafting

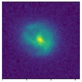


Separation with a classifier

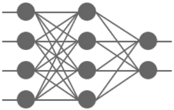


Deep learning

Input data

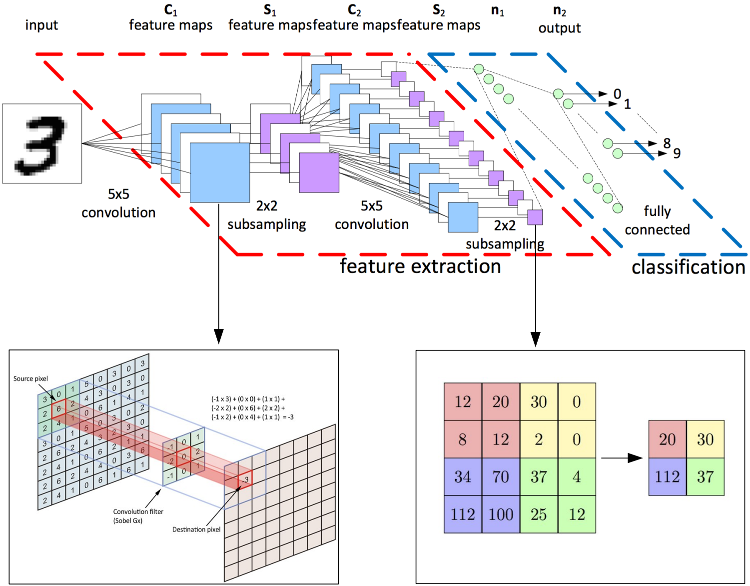


Feature learning

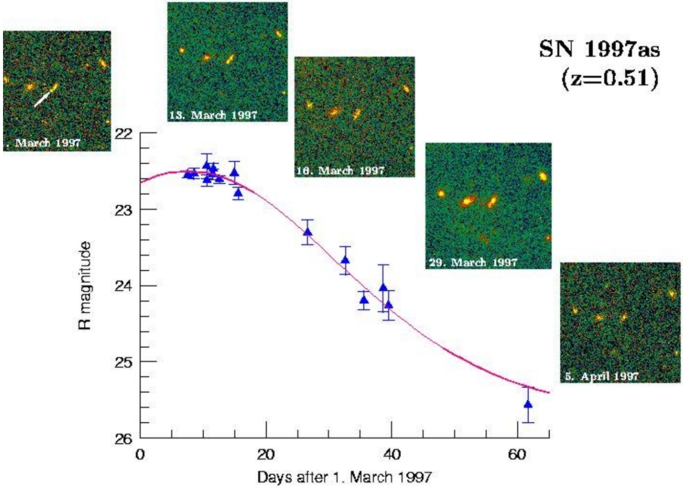


The best feature space representation is found by the network

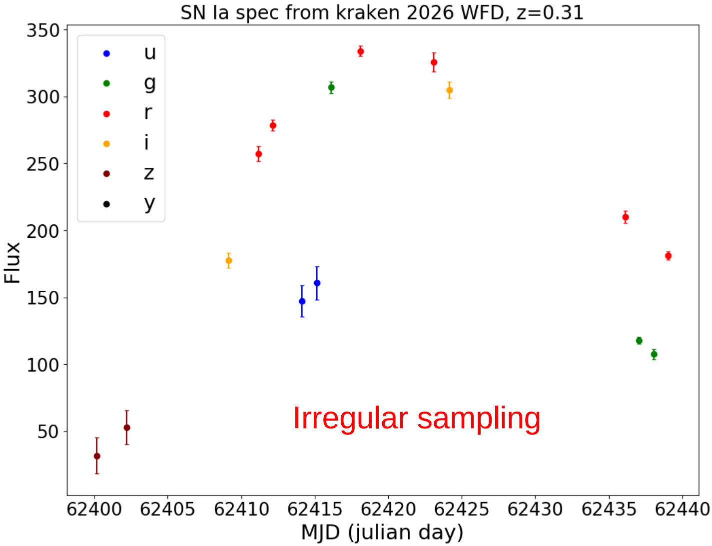
Typical CNN architecture



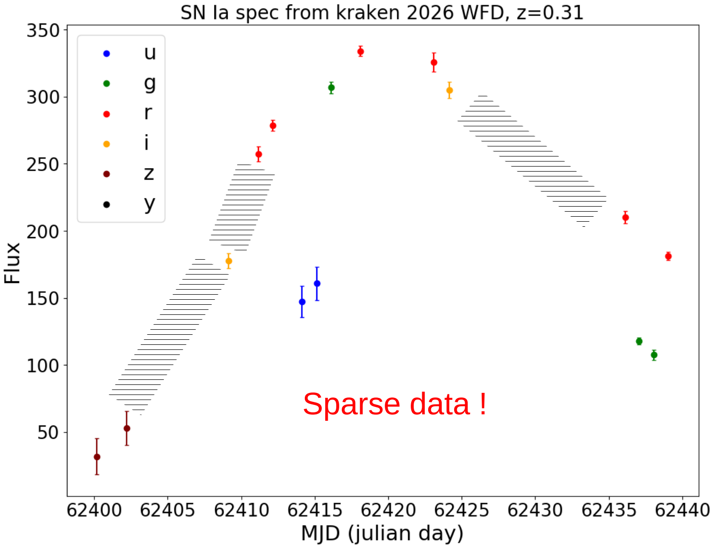
The light curves



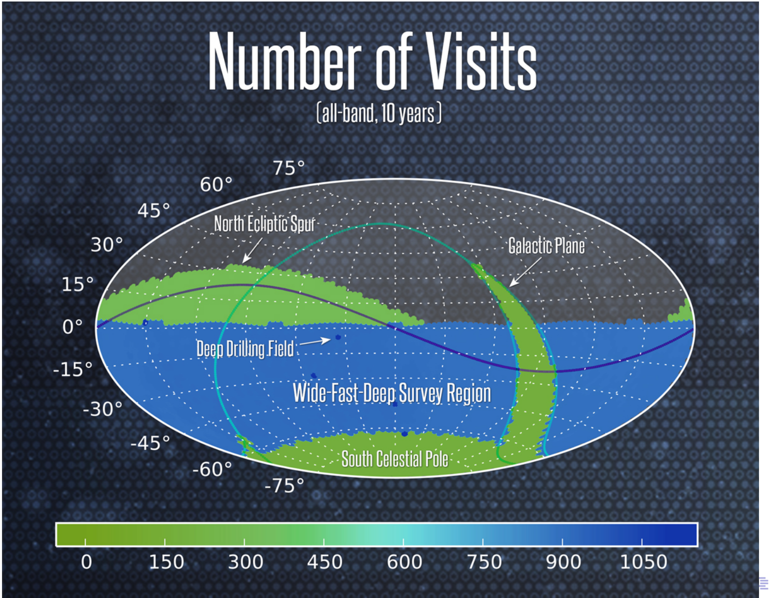
Irregular sampling



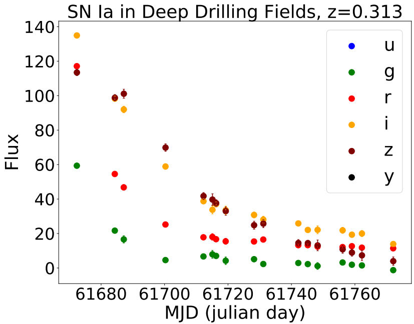
Sparse data

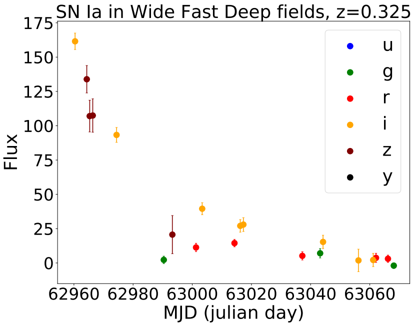

SN Ia spec from kraken 2026 WFD, z=0.31
u
g
r
i
z
y
Flux
350
300
250
200
150
100
50
62400
62405
62410
62415
62420
62425
62430
62435
62440
MJD (julian day)
Sparse data !

The observational strategy



Two different sampling

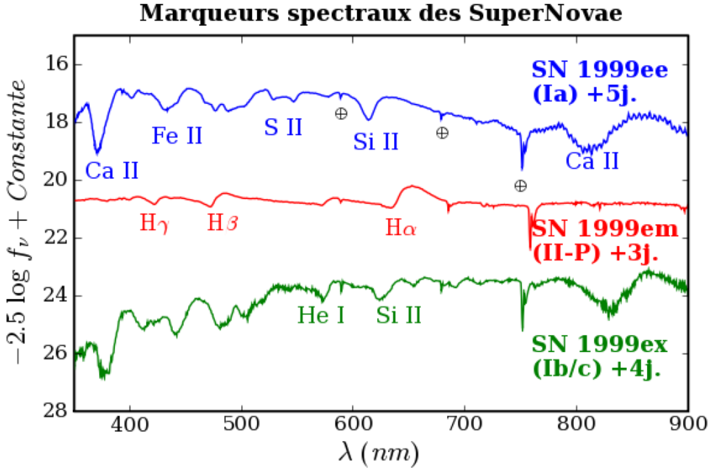




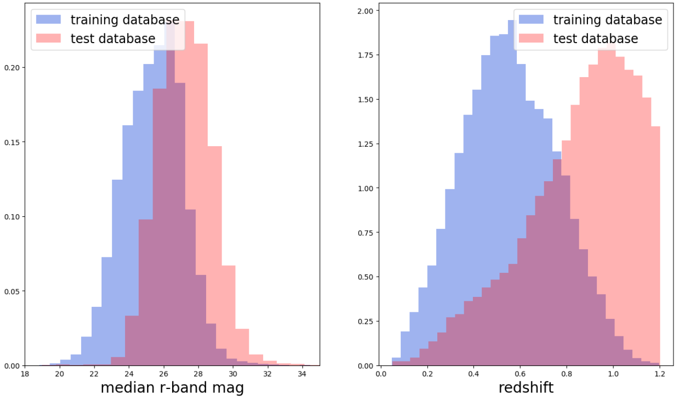
How to label light curves?

The spectrocopy technique

Study of the decomposition of the light by a dispersive element (prism, optical fibres) to analyze the composition of an astrophysical object



A testing database not representative in flux



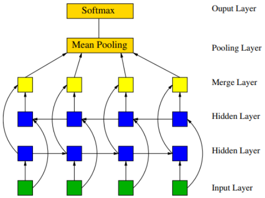
The non-representativeness of the databases, which is a problem of mismatch, is critical for machine learning process.

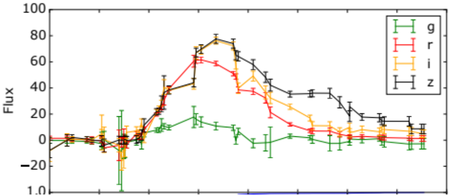
The classification of light curves of supernovae (SN Ia/ SN Non-Ia)

Johanna Pasquet, Jérôme Pasquet, Marc Chaumont and Dominique Fouchez
(arXiv:1901.01298)

What deep learning method should we adopt?

- Recurrent neural network: suited to time series





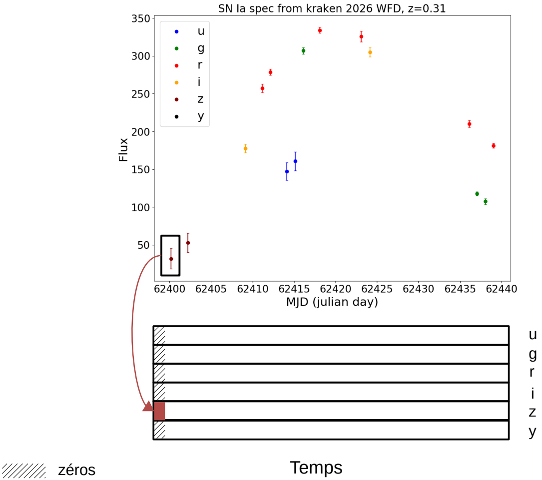
Charnock + 2017

⇒ Interpolation of data can biais the learning

⇒ Performance comparable to classical method

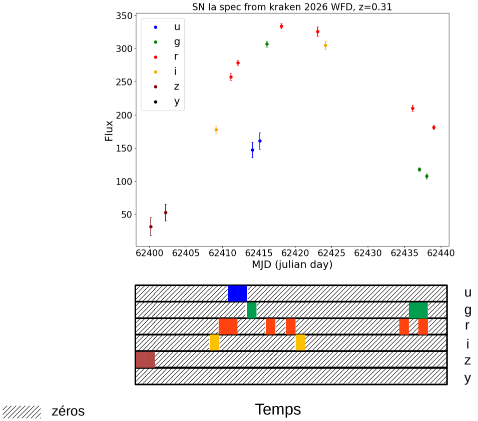
Convolutional neural network

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



Convolutional neural network

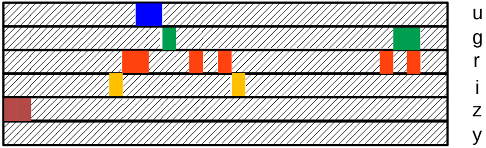
1. Transform input light curves into images : Light Curve Images (LCI)

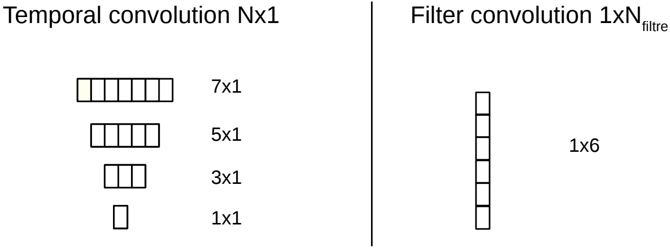


Overfitting of missing data (zero values)

Convolutional neural network

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





Problem n°1

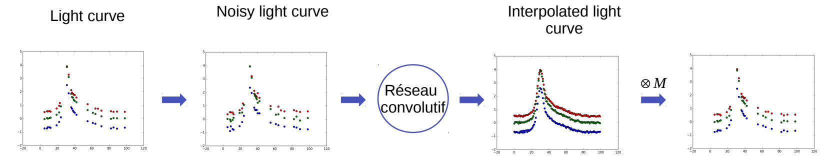
Non-representativeness bewteen the training and the test databases

Problem n°1

Non-representativeness bewteen the training and the test databases

Our solution

Non-supervised learning to extract features from the test light curves



m^o_0	m^o_0	0	m^o_1	0	m^o_2	0	m^o_3

Input data in a given filter

1	1	0	1	0	1	0	1

Mask M

m'^o_0	$m^o_{int,0}$	$m^o_{int,1}$	m'^o_1	$m^o_{int,2}$	m'^o_2	$m^o_{int,3}$	m'^o_3

Output interpolated data

m'^o_0	0	0	m'^o_1	0	m'^o_2	0	m'^o_3

Reconstructed light curve

Cost function : $\sqrt{\sum_i ((m_i - m_i') \otimes M_i)^2}$

Problem n°2

Variable sampling depending on the observational strategy of LSST

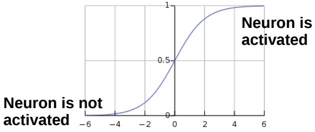
Problem n°2

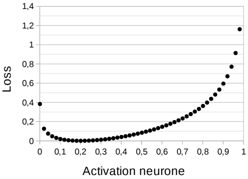
Variable sampling depending on the observational strategy of LSST

Our solution

Add a regularization term inside the network

1. Use of a Sigmoid function





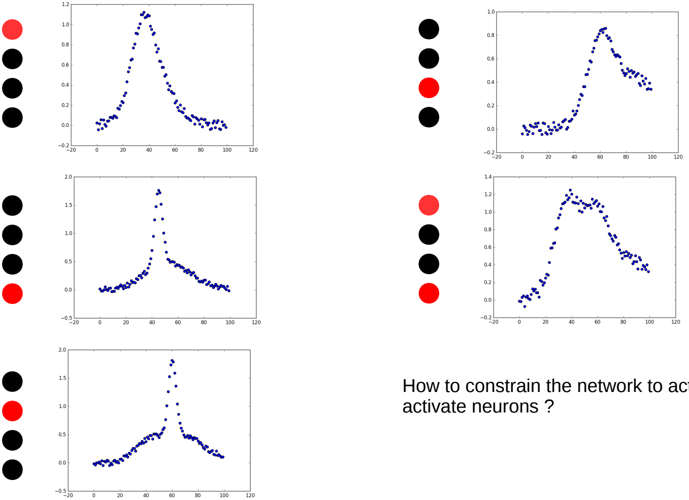
2. Regularization with a Kullback–Leibler divergence

$$\underbrace{KL(\rho \| \hat{\rho}_j)}_{\text{Activation of a neuron}} = \underbrace{\rho}_{\text{Constant}} \log \frac{\rho}{\hat{\rho}_j} + (1-\rho) \log \left(\frac{1-\rho}{1-\hat{\rho}_j} \right)$$

A regularization term

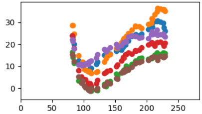
FC 11
5000

Neuron is not activated (=0)
Neuron is activated (=1)

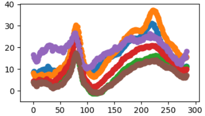


How to constrain the network to activate/not activate neurons ?

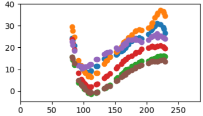
Visualization



Input light curve

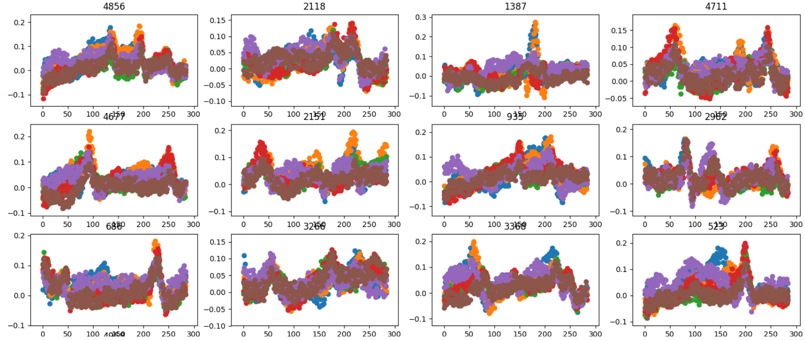


Interpolated light curve



Reconstructed light curve

Activation of 12 neurons :



=> Among 5 000 neurons only a restricted number of them are activated (between 10 and 30) with a score above 0.2

Problem n°3

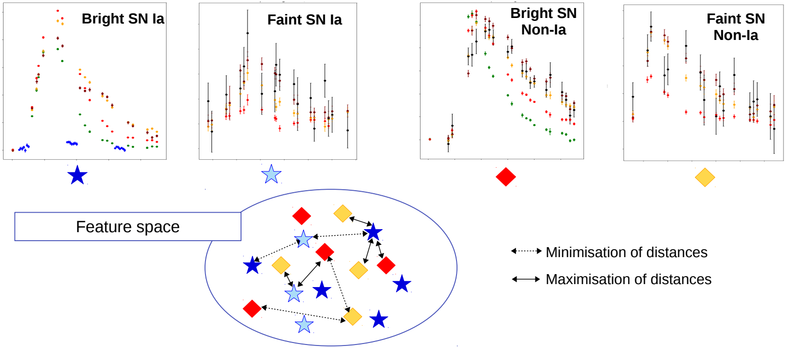
Evolution of light curves with distances in the Universe

Problem n°3

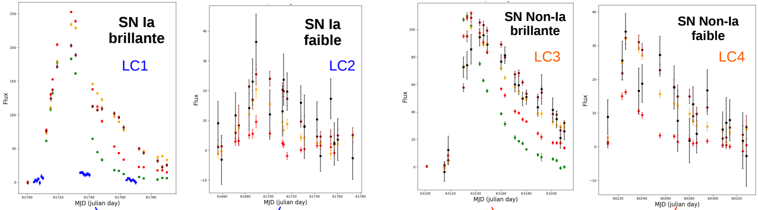
Evolution of light curves with distances in the Universe

Our solution

Semi-supervised learning by minimizing distances in the feature space



The contrastive loss applied to light curves



||LC1 - LC2||

||LC3 - LC4||

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

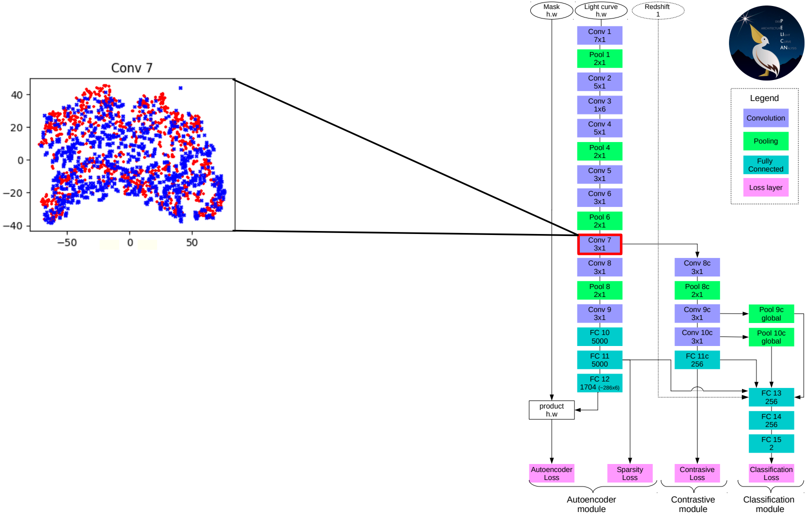
$$max\left(0,m-\sqrt{\|LC2-LC4\|}\right)^2$$

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

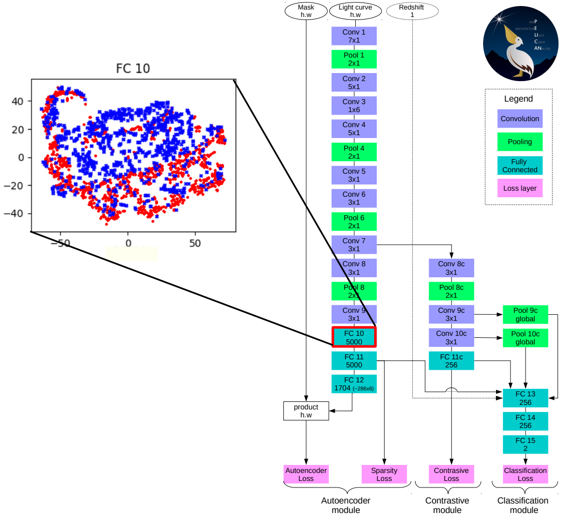
$$max\left(0,m-\sqrt{\|LC1-LC4\|}\right)^2$$

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

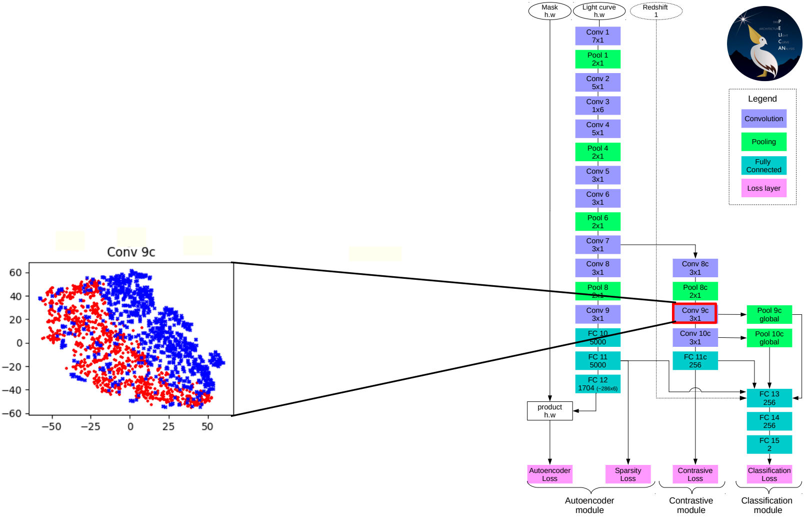
PELICAN


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

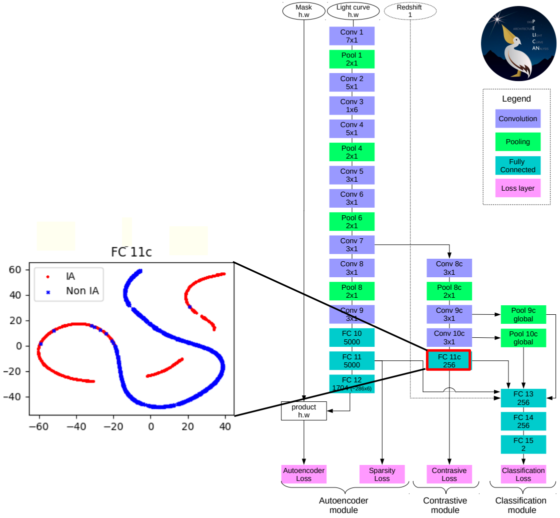
PELICAN


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

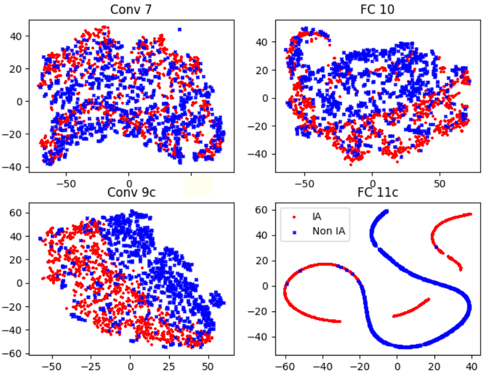
PELICAN

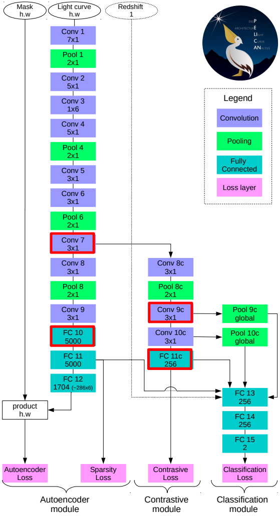

Conv 9c
60
40
20
0
−20
−40
−60
−50
−25
0
25
50
Mask
h.w
Light curve
h.w
Redshift
1
Conv 1
7x1
Pool 1
2x1
Conv 2
5x1
Conv 3
1x6
Conv 4
5x1
Pool 4
2x1
Conv 5
3x1
Conv 6
3x1
Pool 6
2x1
Conv 7
3x1
Conv 8
3x1
Pool 8
2x1
Conv 9
3x1
FC 10
5000
FC 11
5000
FC 12
1704 (~286x6)
product
h.w
Conv 8c
3x1
Pool 8c
2x1
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

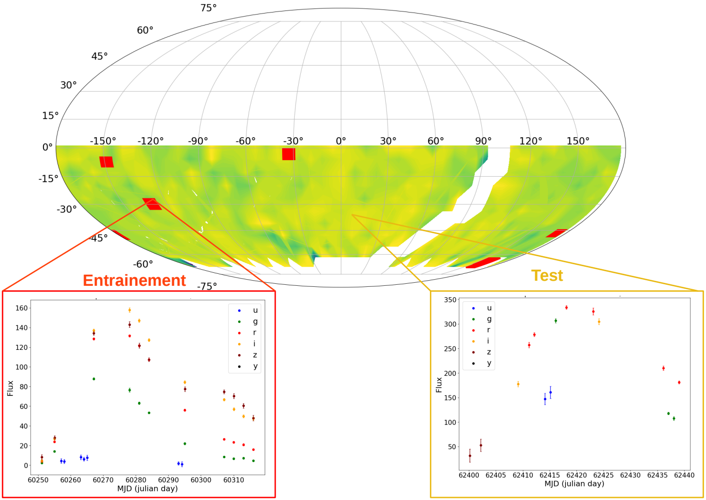

FC 11c
IA
Non IA
60
40
20
0
−20
−40
−60
−40
−20
0
20
40
Mask
h.w
Light curve
h.w
Redshift
1
Conv 1
7x1
Pool 1
2x1
Conv 2
5x1
Conv 3
1x6
Conv 4
5x1
Pool 4
2x1
Conv 5
3x1
Conv 6
3x1
Pool 6
2x1
Conv 7
3x1
Conv 8
3x1
Pool 8
2x1
Conv 9
3x1
FC 10
5000
FC 11
5000
FC 12
1704 (=286x6)
product
h.w
Conv 8c
3x1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

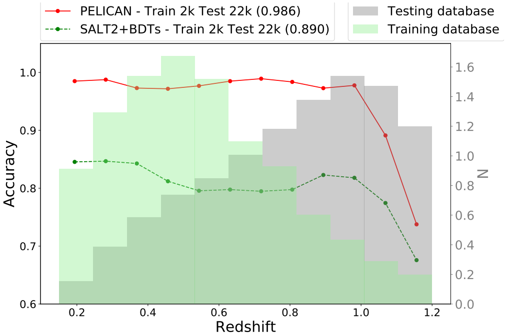

Conv 7
FC 10
Conv 9c
FC 11c
IA
Non IA


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)
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
product h.w
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

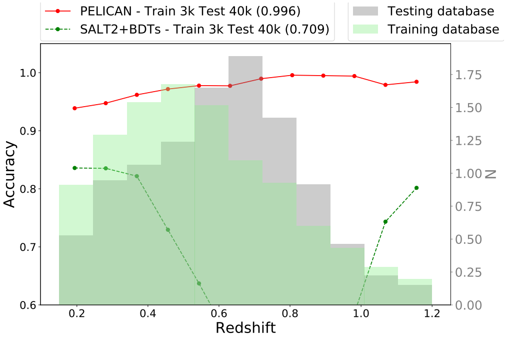


Results on DDF



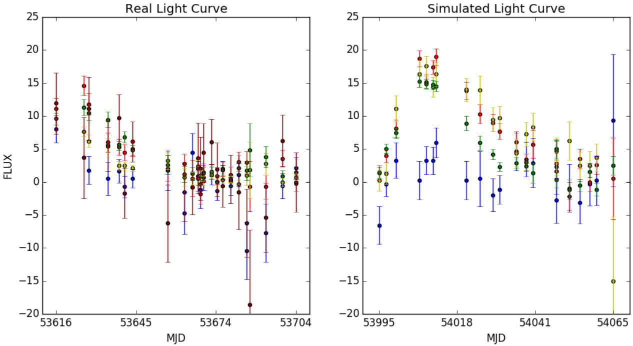
	Training database (spec only)	Test database (phot only)	Accuracy	$\text{Recall}_{\text{Ia}}$ $\text{Precision}_{\text{Ia}}>0.95$	$\text{Recall}_{\text{Ia}}$ $\text{Precision}_{\text{Ia}}>0.98$	AUC
DDF	500	1,500	0.849 (0.746)	0.617 (0.309)	0.479 (0.162)	0.937 (0.848)
	2,000	2,000	0.925 (0.783)	0.895 (0.482)	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	$\text{Recall}_{\text{Ia}}$ $\text{Precision}_{\text{Ia}} > 0.95$	$\text{Recall}_{\text{Ia}}$ $\text{Precision}_{\text{Ia}} > 0.98$	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)

Validate PELICAN on real 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 astrophysical surveys will deliver multi-band photometry for billions of sources
- Many issues for the classification algorithms
- Performance never achived for the classification of light curves by considering a non-representative training database

Perspectives

- The method can be used for different kind of noisy images as sonar images

Thank you for your attention!

Summary

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

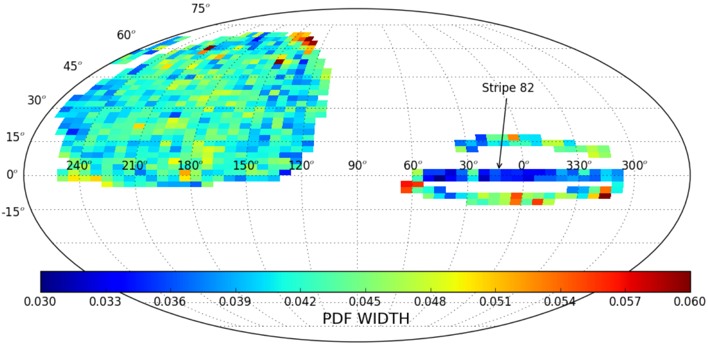
Perspectives

- The method can be used for different kind of noisy images as sonar images

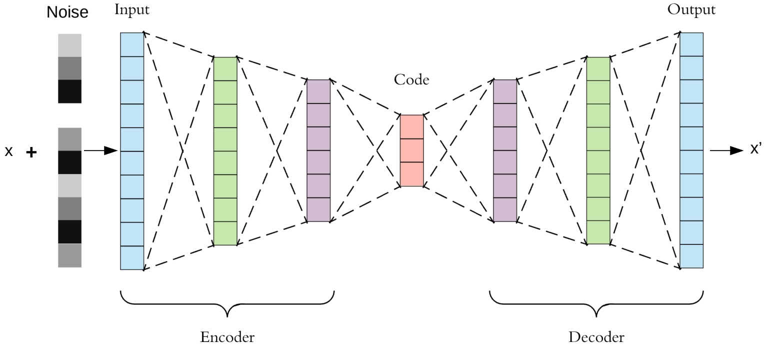
Thank you for your attention!

act 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

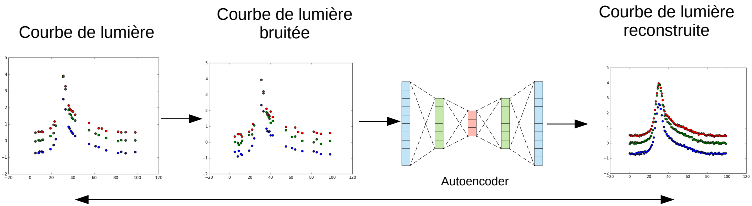


Autoencodder

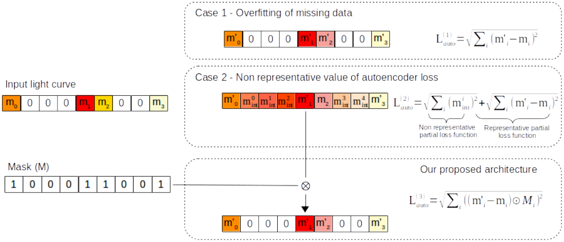


Fonction de perte = $|| x - x' ||_2$

Autoencodder



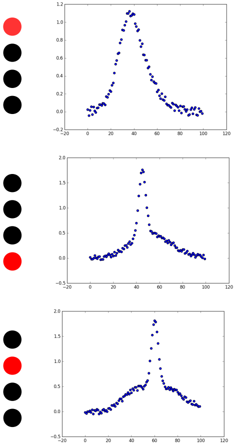
Comment calculer une fonction de perte cohérente ?

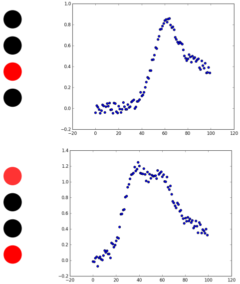


Autoencodder

FC 11
5000

Neuron is not activated (=0)
Neuron is activated (=1)

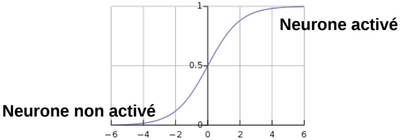




How to constrain the network to activate/not activate neurons ?

Autoencodder

1. Utilisation de la fonction d'action Sigmoïde

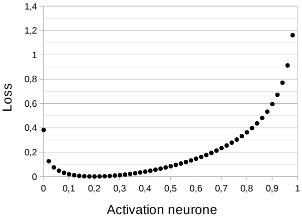


2. Régularisation à l'aide de la divergence de Kullback–Leibler

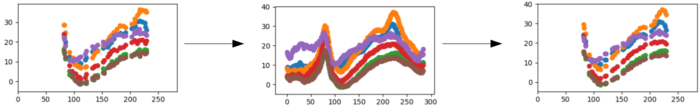
$$KL(\rho \| \hat{\rho}_j) = \rho \log \frac{\rho}{\hat{\rho}_j} + (1-\rho) \log \left(\frac{1-\rho}{1-\hat{\rho}_j} \right)$$

Activation d'un neurone

Constante

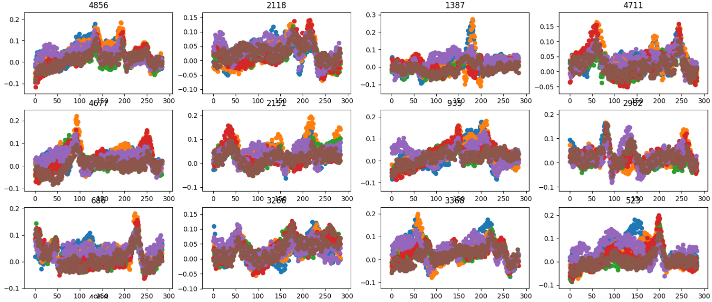


Autoencodder



Input light curve → Interpolated light curve → Reconstructed light curve

Activation of 12 neurons :



=> Among 5 000 neurons only a restricted number of them are activated (between 10 and 30) with a score above 0.2