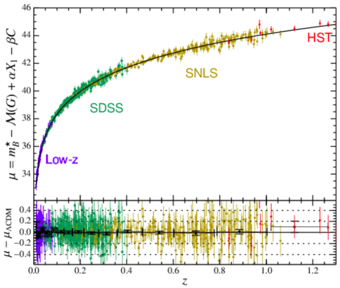
Photometric classification of type Ia Supernovae with machine learning

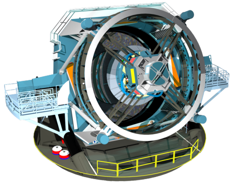
Johanna PASQUET

LSST France

Introduction



Betoule et al., 2014



SDSS-II/SNLS3 Joint Light-curve Analysis

- **740** spectroscopically confirmed type Ia supernovae with high quality light curves

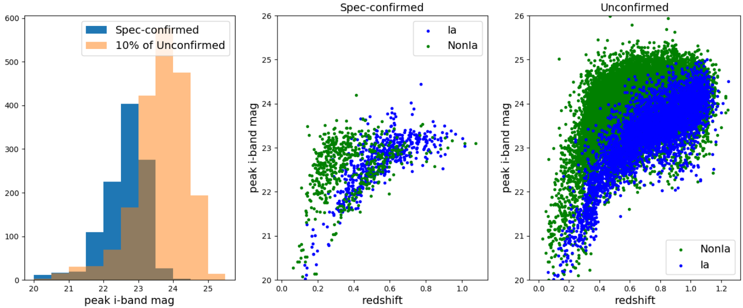
LSST Supernovae Ia Analysis :

- LSST will discover **hundreds of thousands** of type Ia supernovae
- Spectroscopy of host galaxies for a subsample of SN
- Be able to automatically identify SNe Ia among all the supernovae with the photometric light curves

Prepare the data analysis : the SPCC

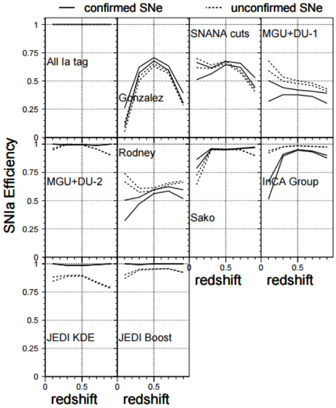
The supernovae Photometric Classification Challenge (SPCC, Kessler 2010)

- $\sim$ 20 000 SNe light curves, generated using the SuperNova ANAlysis (SNANA) light curves simulator (Kessler et al., 2009),
- Designed to mimich data from the DES,
- The spectroscopically confirmed sub-sample is composed of $\sim$ 1000 light curves.

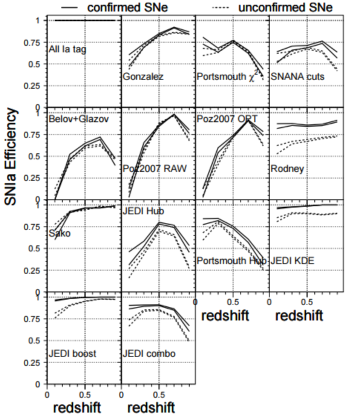


Classification methods for the SPCC

no host z information



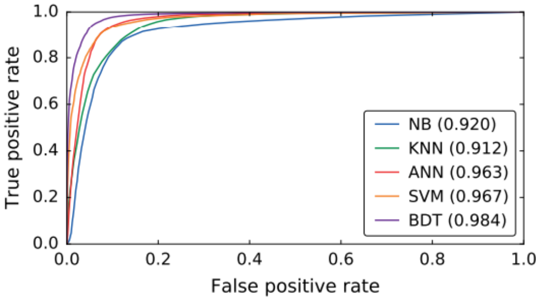
with host z information



After the Challenge : post-SPCC, an updated version of the data, including all labels, bug fixes and other improvements

Machine learning methods for the SPCC

- Use of representative training set,
- SALT 2 features (t_0, x_0, x_1 and c) with redshift,



(b) SALT2 model, with redshift

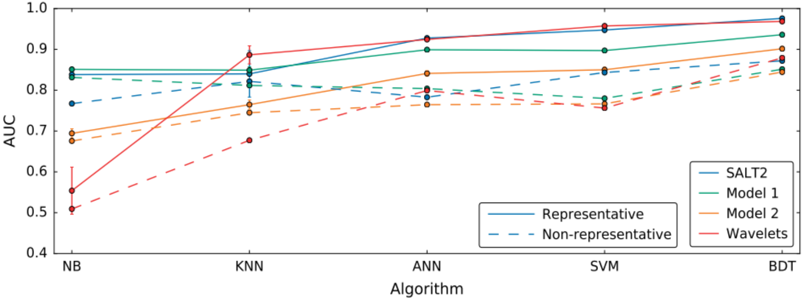
From Lochner et al. (2016)

$$\mathrm{TPR} = \frac{TP}{TP+FN}$$

$$\mathrm{FPR} = \frac{FP}{FP+TN}$$

Non-representative SPPC training set

- Real problem of mismatch between the training set and the testing set

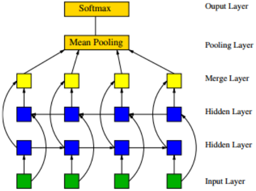


From Lochner et al. (2016)

Recent work with Deep Learning

Charnock et al., 2017

Method	Training size	AUC	Accuracy (%)	F_1	Purity (%)	Completeness (%)	Host z
A	10,660	0.986 ± 0.001	94.7 ± 0.2	0.64 ± 0.01	87.3 ± 0.8	91.4 ± 1.1	True
A	10,660	0.981 ± 0.001	93.6 ± 0.3	0.60 ± 0.02	87.4 ± 1.7	85.4 ± 2.6	False
A	5,330	0.975 ± 0.003	92.9 ± 0.6	0.57 ± 0.03	86.6 ± 2.0	83.4 ± 3.4	True
A	5,330	0.973 ± 0.002	92.3 ± 0.4	0.55 ± 0.02	86.2 ± 2.4	80.8 ± 3.8	False
B	1,103	0.910 ± 0.012	85.9 ± 0.9	0.31 ± 0.03	72.4 ± 0.4	66.1 ± 6.0	True
B	1,103	0.901 ± 0.016	84.6 ± 1.7	0.28 ± 0.05	68.2 ± 3.4	66.3 ± 5.5	False
C	~10,660	-	-	0.58	85	88	True
C	~10,660	-	-	0.51	82	85	False
C	1,045	-	-	0.33	70	75	True
C	1,045	-	-	0.29	67	71	False
D	~8,000	-	-	0.55	-	-	True
D	~2,000	-	-	0.45	-	-	True
E	1,103	0.94 ± 0.03	-	-	-	-	True
E	1,103	0.89 ± 0.53	-	-	-	-	False
E	1,103	-	-	-	90	85	True
E	1,103	-	-	-	87	90	True

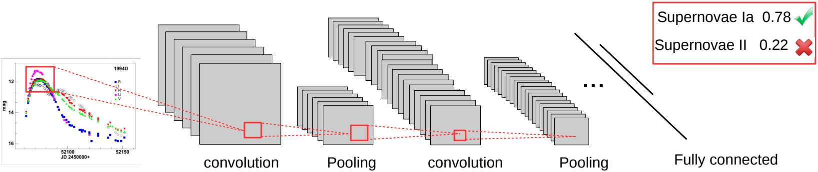


Example of bidirectional LTSM

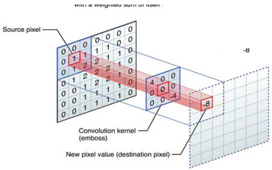
Method A : unidirectional LTSM [16,16]
Method B : unidirectional LTSM [4]
Method C : Karpenka et al. (2013)
Method D : Newling et al. (2011)
Method E : Lochner et al. (2016)

For a training size of 50% of the representational SPCC dataset (around 10^4 supernovae) a type-Ia vs. non-type-Ia classification accuracy of 94.7% and an area under the Receiver Operating Characteristic curve AUC of 0.986

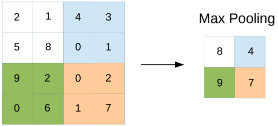
CNN approach



Convolution



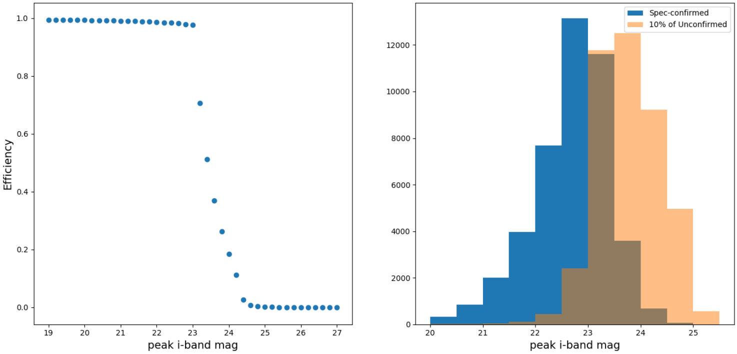
Pooling



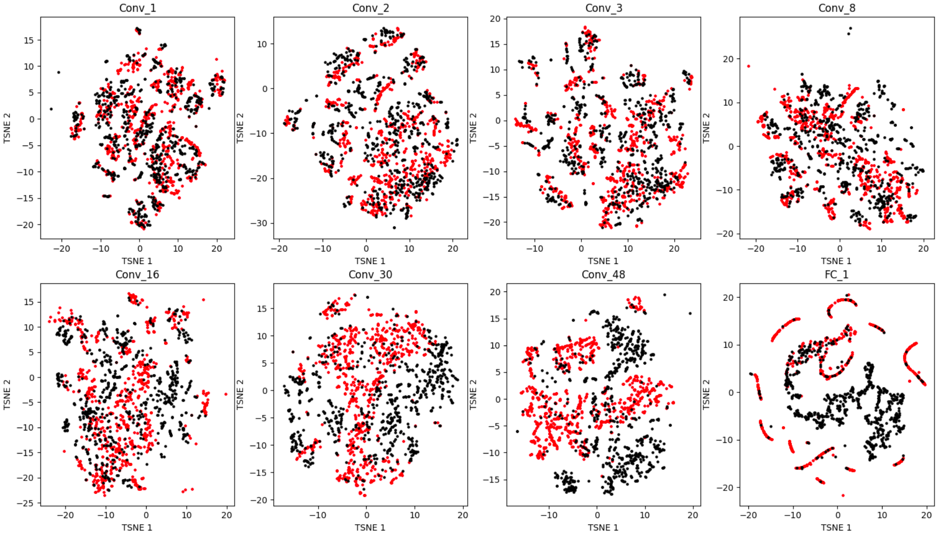
A promising approach : J. Pasquet Itam et al., 2017 (arXiv:1712.02777)
Mahabal et al., 2017 (arXiv:1709.06257)

Test a CNN on augmented data

- Simulation of light curves with SNANA to mimich data from the DES,
- Training set (spectroscopically confirmed sub-sample) composed of $\sim 43\,000$ Ia and Non Ia,

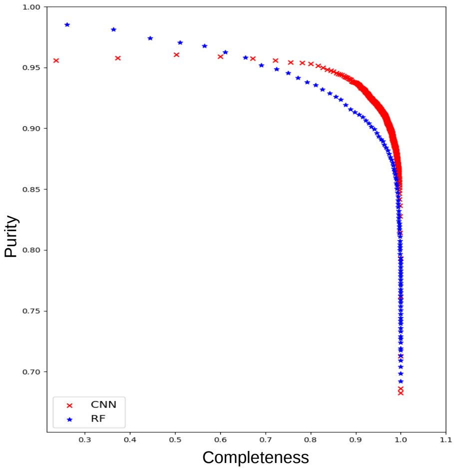


t-SNE Along the network



First preliminary results

- Computation of the purity and the completeness by varying the probability threshold,



To conclude

- Since the SPCC, lots of classification methods were proposed to deal with the problem of photometric classification of SNe
- The machine learning methods showed promising results including or not the information of Host z
- Deep learning methods are emerging as they are able to identify a set of features that is context specific
- As of now Deep learning methods suffer from the lack of training data and the sparsity of data
- The future photometric challenge PLAsTiCC (Photometric LSST Astronomical Time-series Classification Challenge) could provide solutions : Emille's talk !

Thank you for your attention !

To conclude

- Since the SPCC, lots of classification methods were proposed to deal with the problem of photometric classification of SNe
- The machine learning methods showed promising results including or not the information of Host z
- Deep learning methods are emerging as they are able to identify a set of features that is context specific
- As of now Deep learning methods suffer from the lack of training data and the sparsity of data
- The future photometric challenge PLAsTiCC (Photometric LSST Astronomical Time-series Classification Challenge) could provide solutions : Emille's talk !

Thank you for your attention !