

# Classification de Supernovae par CNN

Anthony BRUNEL<sup>1,2</sup>, Johanna PASQUET<sup>3</sup>, Jérôme PASQUET<sup>1,2</sup>  
Nancy RODRIGUEZ<sup>1,2</sup>, Frédéric COMBY<sup>1,2</sup>, Dominique FOUCHÉZ<sup>3</sup>,  
Marc CHAUMONT<sup>1,4</sup>  
LIRMM<sup>1</sup>, Univ Montpellier<sup>2</sup>, CPPM<sup>3</sup>, Univ Nîmes<sup>4</sup>

Réunion du GdR ISIS (Apprentissage pour l'analyse des images de télédétection),  
18 octobre 2018, CNES Paris-Daumesnil, Paris.

# Plan

1 Présentation du domaine

2 Les données

3 Le réseau CNN

4 Résultats

5 Conclusion

6 Annexe

# Introduction

## La cosmologie

Science qui étudie la structure, l'origine et l'évolution de l'univers

## Le challenge

Analyser/Détecter des quantités énormes de données



Figure: Simulation d'image du ciel tel que le verra LSST

## Les Supernovæ

Ensemble de phénomènes résultant de **l'explosion d'une étoile** (divers types : Ia, Ib, Ic, II..)

# Supernova

## Supernova Ia : La chandelle standard

- Une naine blanche accrète de la matière d'une étoile compagnon, et finit par exploser.

*Vidéo supernova Ia* Video credit: ESO; Downloaded from: <http://www.eso.org>



- Explosion toujours identique (luminosité intrinsèque)  
Luminosité observée sur Terre  $\Rightarrow$  déduction distance

## Supernova non Ia

- Explosion dite à effondrement de cœur

# Identification d'objets célestes

## Méthodes d'identification

- Identification par spectroscopie
  - ▶ Étudie le spectre de l'objet
- Identification par photométrie
  - ▶ Identifie les astres à l'aide de différents filtres

## Photométrie vs spectroscopie

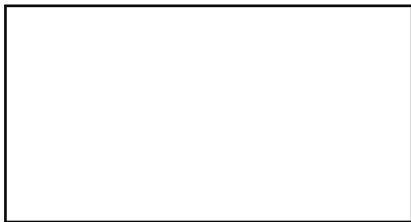
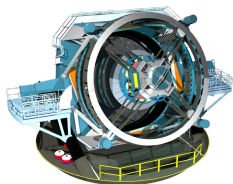
- La spectroscopie = précis, objet doit être lumineux, coûteux
- La photométrie = moins précis mais moins coûteux

⇒ **LSST fournira des données photométriques**

# Large Synoptic Survey Telescope (LSST)

## LSST

- Fonctionnel à partir de 2022
- Projet d'observation 10 ans
- Sondera le ciel visible, deux fois par semaine
- Fournira 15 Tera de données chaque nuit
- Plus de 10 millions de supernovæ seront découvertes



*Vidéo LSST* Video credit: Guillaume Doyen; Downloaded from:

<http://astro-space-page.blogspot.com>

# Courbes de lumière

## Courbes de lumière

- Évolution du flux lumineux au cours du temps

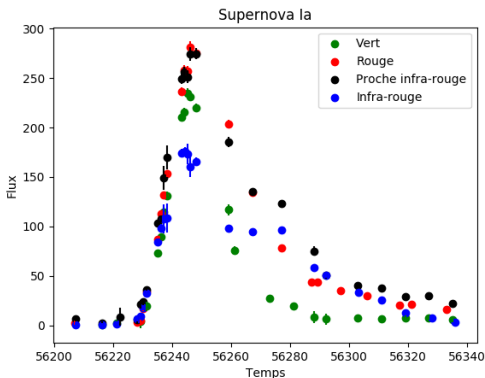


Figure: Exemple de supernova Ia simulée

# Courbes de lumière

## Problèmes

- Échantillonnage temporel irrégulier
- Durée variable
- Extrêmement *sparse*

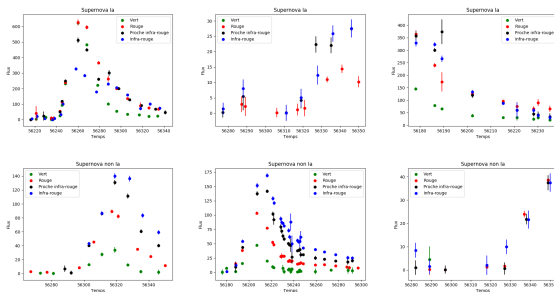


Figure: Exemple de courbes de lumière de supernovae Ia et non Ia



# Plan

1 Présentation du domaine

**2 Les données**

3 Le réseau CNN

4 Résultats

5 Conclusion

6 Annexe

# Les données

## Données simulées

- Données simulées avec le logiciel SNANA et correspondant à celle du Dark Energy Survey
- Simulations générées de façon à être les plus réalistes possibles

## Quelques chiffres

- Courbes de lumière : 5 000
- Supernovæ Ia : 2 500
- Supernovæ non Ia : 2 500
- Sparse : Plus de 70% de zéros

# Représentation des données d'entrées

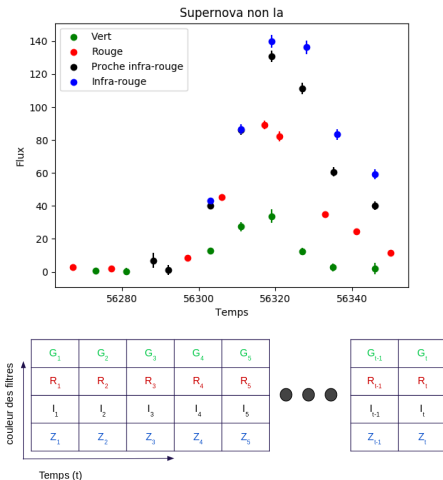


Figure: Représentation des courbes de lumière sous forme de matrice

# Plan

1 Présentation du domaine

2 Les données

**3 Le réseau CNN**

4 Résultats

5 Conclusion

6 Annexe

# Architecture du réseau CNN

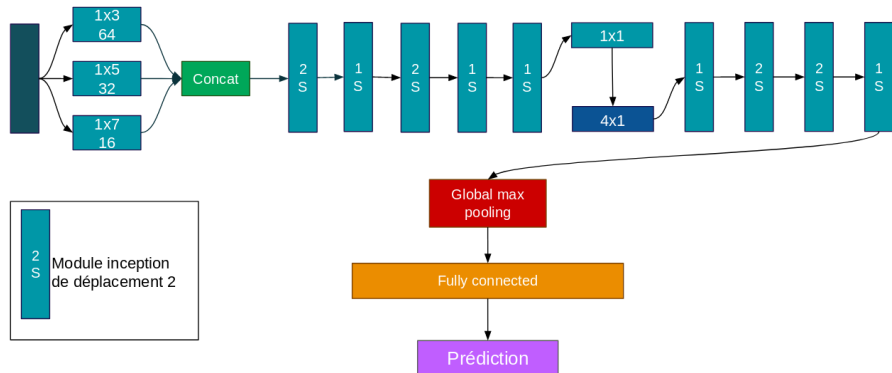


Figure: Réseau de neurones convolutifs

Anthony Brunel, Johanna Pasquet, Jérôme Pasquet, Nancy Rodriguez, Frédéric Comby, Dominique Fouchez, Marc Chaumont, "A CNN adapted to time series for the classification of Supernovae", EI'2019, in Proceedings of Color Imaging XXIV: Displaying, Processing, Hardcopy, and Applications - joint sessions dealing with color in astronomy / astrophysics, Part of IS&T International Symposium on Electronic Imaging, Burlingame (suburb of San Francisco), California USA, 13 - 17 January, 2019.

The CNN is downloadable there: <https://github.com/Anzzy30/SupernovaeClassification>

# Plan

1 Présentation du domaine

2 Les données

3 Le réseau CNN

**4 Résultats**

5 Conclusion

6 Annexe

# Base et Hyper-paramètres

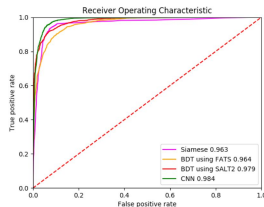
## Base d'apprentissage / Base de test

- **4-Fold** cross-validation
  - **3 750 courbes** de lumière pour l'**apprentissage**
  - **1 250 courbes** de lumière pour le **test**
  - **Augmentation virtuelle** par crop de 40% à 80% des courbes de lumière, (réduit l'over-fitting et améliore les résultats)
- 
- *4500 itérations, dropout de 0.4 sur le fully connected, learning rate avec exponential decay de  $10^{-2}$  à  $5 \times 10^{-4}$ , Adam optimizer, cross entropy loss, poids initialisés via approche Xavier, batch de taille 128...*
  - Apprentissage sur NVIDIA GTX 1080

# Résultats

Model	Training set	AUC	Accuracy
CNN	3750	<b>0.984</b>	<b>94.6</b>
SALT2	3750	0.979	92.3
FATS	3750	0.964	90.1
Siamese	3750	0.963	93.0

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$



[Lochner et al. 2016] SALT2 : Boosted decision tree (BDT) using SALT2 features (Spectral Adaptive Light curve Template 2)

[Nun et al. 2015] FATS : Boosted decision tree (BDT) using FATS (Feature Analysis for Time Series)



# Analyse

- Meilleurs que l'état de l'art "feature+classifieur"
- Performances réseaux deep peuvent facilement être améliorées
- Augmenter le nombre d'exemples (pas montré ici) améliore les résultats de deep
- Résultat de RNN du même ordre de grandeur (pas montré ici)

# Améliorations

- Augmentation virtuelle par bruitage,
- Amélioration possible des performances par ensemble, transfert learning, multi-class, utilisation du redshift
- Extension : Mieux gérer la sparsité
- Extension : Gérer le faible nombre d'échantillons
- Extension : Gérer le mismatch entre objets célestes proches et loins
- Extension : Gérer le mismatch de cadence (sparsité)

En cours de soumission : "PELICAN: deeP architecturE for the LIght Curve ANalysis" Johanna Pasquet, Jérôme Pasquet, Marc Chaumont and Dominique Fouchez. Astronomy & Astrophysics 2018

# Plan

- 1 Présentation du domaine
- 2 Les données
- 3 Le réseau CNN
- 4 Résultats
- 5 Conclusion**
- 6 Annexe

# Conclusion et Perspectives

## Objectifs

- Classification binaire des supernovæ Ia et non Ia
- Gérer les verrous dus aux données de LSST

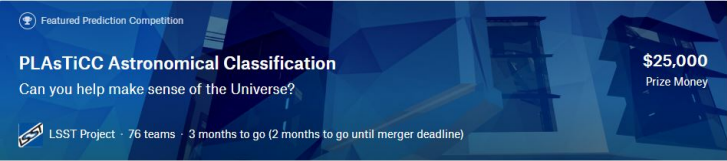
## Contributions

- Deux réseaux à l'état de l'art
- Les approches CNN ou par triplet loss sont envisageables

## Perspectives

- En cours de soumission : "PELICAN: deeP architecturE for the LIght Curve ANalysis" Johanna Pasquet, Jérôme Pasquet, Marc Chaumont and Dominique Fouchez. Astronomy & Astrophysics 2018
- Kaggle

# Kaggle Competition: Astronomical Classification Challenge



Featured Prediction Competition

**PLAsTiCC Astronomical Classification**

Can you help make sense of the Universe?

LSST Project · 76 teams · 3 months to go (2 months to go until merger deadline)

**\$25,000**  
Prize Money

- The Large Synoptic Survey Telescope (LSST) is about to **revolutionize the field**, discovering 10 to 100 times more astronomical sources that vary in the night sky than we've ever known.
- The **Photometric LSST Astronomical Time-Series Classification Challenge (PLAsTiCC)** asks Kagglers to help **prepare to classify** the data from this new survey.
- \$25,000 in total prizes
- Timeline:
  - ▶ September 28, 2018 - Launch
  - ▶ December 10, 2018 - Entry deadline and Team Merger deadline
  - ▶ December 17, 2018 - Final submission deadline.
  - ▶ January 15, 2019 - LSST Workshop entry deadline.
  - ▶ February 15, 2019 - LSST Workshop announcement.

<https://www.kaggle.com/c/PLAsTiCC-2018>

# Plan

- 1 Présentation du domaine
- 2 Les données
- 3 Le réseau CNN
- 4 Résultats
- 5 Conclusion
- 6 Annexe**

# Analyse photométrique

## Les filtres

Chaque filtre est conçu pour laisser entrer la lumière dans une longueur d'onde spécifique.

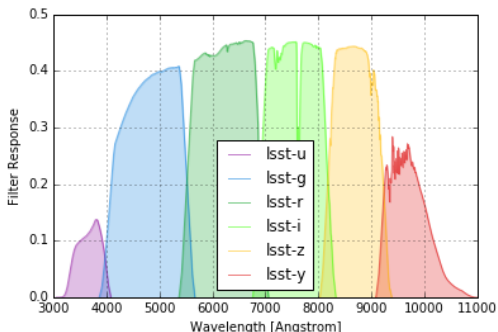
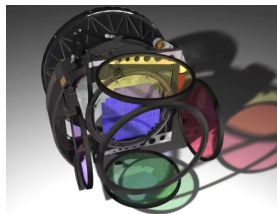
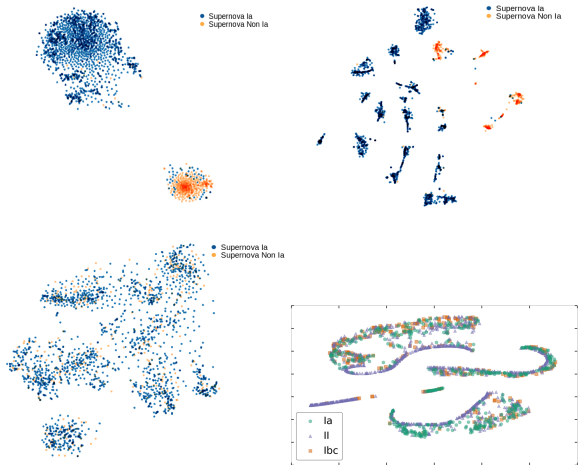


Figure: Filtres de LSST



**Figure:** t-SNE: Triplet loss, Triplet loss modifiée, CNN, SALT2 extrait de [Lochner]



# Augmentation de données

## Découpage aléatoire

- Chaque courbe a une chance d'être modifiée à chaque itération
- Seulement une portion de la courbe est sélectionnée

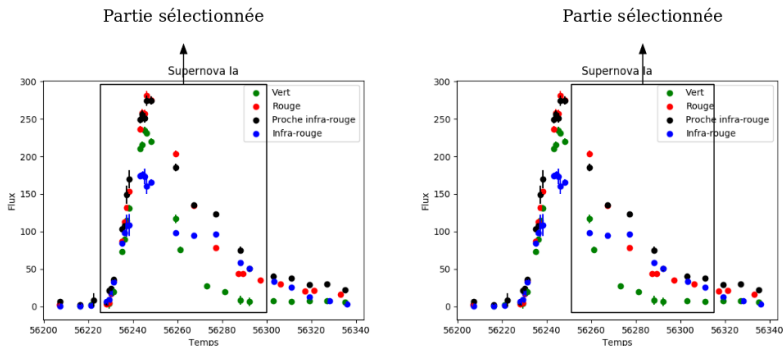


Figure: Illustration du découpage aléatoire

# Augmentation de données

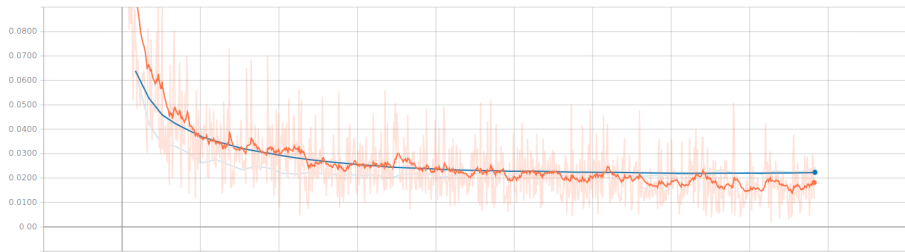
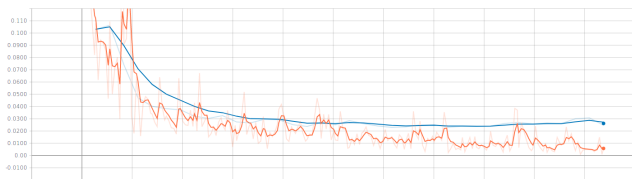


Figure: Fonction de perte avec découpage (100 époques, courbe orange entraînement, courbe bleue validation)

# Architecture du réseau CNN

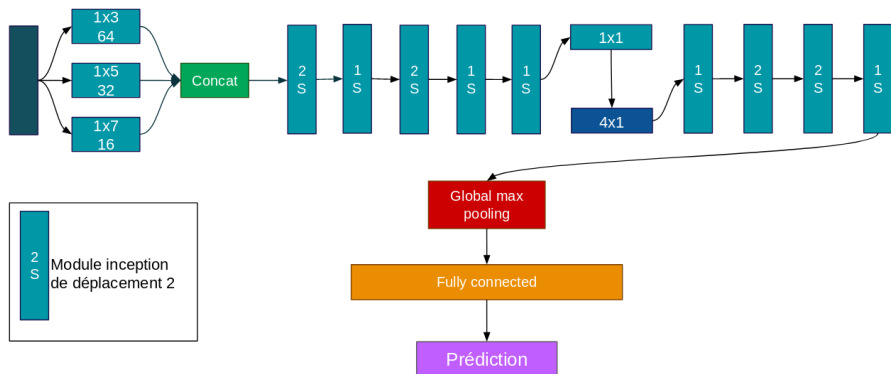


Figure: Réseau de neurones convolutifs

# Convolution et pooling 1D

## Convolution et pooling temporels

- Noyaux 1x3, 1x5, 1x7
- Extraction des informations temporelles

## Convolution couleur

- Noyau 4x1
- Réunit les informations photométriques
- Transforme la matrice d'entrée en un tableau 1D
- Permet un gain de 5% de taux de classification correcte

# Convolution et pooling 1D

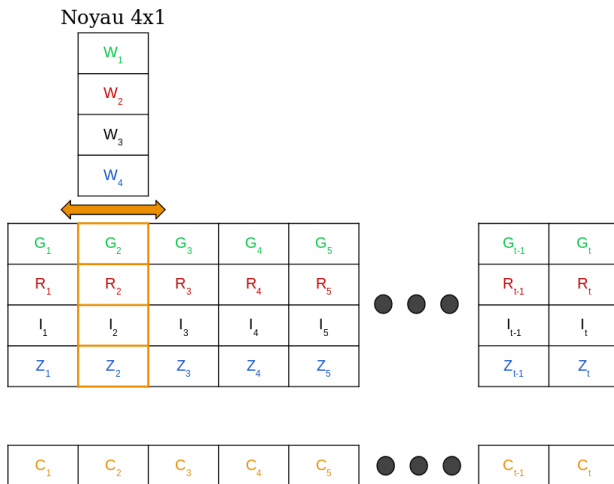


Figure: Illustration de la convolution 1D couleur

# Module inception

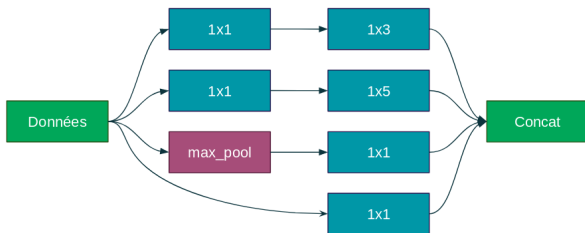


Figure: Module inception

## Principe

- Extraction de caractéristiques à plusieurs échelles
- Réduction de dimensionnalité, réduction du nombre d'opérations, ajoute de la non linéarité

# Global max pooling

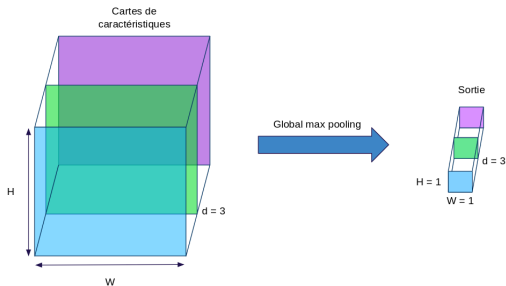


Figure: Global max pooling

## Global max pooling

- Transforme l'ensemble des cartes de caractéristiques en un vecteur 1D
- Permet de passer des courbes (matrices) de durée (largeur) différente en entrée