

Gestion du déséquilibre de classe au sein du classifieur de séries d'images astronomiques ConvEntion

Anass BAIROUK¹, Marc CHAUMONT^{1,2},
Dominique FOUCHEZ³ Jérôme PASQUET⁴ Frédéric COMBY¹
LIRMM, Univ Montpellier, CNRS ¹, Univ Nîmes², CPPM, Uni
Aix-Marseille ³, TETIS, Univ Paul Valéry Montpellier⁴

November 13, 2024

Plan

Introduction/Etat de l'art

Notre proposition ICT-ConvEntion

Expériences & Résultats

Conclusions et perspectives

Contexte : Classifier des séries d'images

- ▶ Les astrophysiciens ont besoin de classifier une grosse quantité d'observations,
- ▶ OBJECTIF : Déterminer des constantes astrophysiques et mieux comprendre l'univers.



Figure: Supernovae

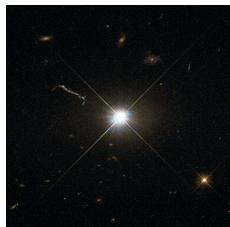


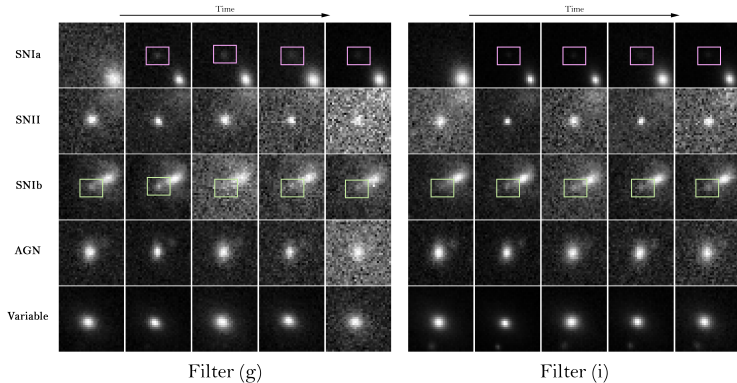
Figure: AGN; Quasar



Figure: Etoile variable

Note : Nous étudions des séries temporelles d'images.

- ▶ 5 bandes : u,g,r,i,z (resp 365, 475, 658, 806, 900 nm),
- ▶ Images 64×64 ,
- ▶ Note : durée SNIa = 100 à 200 nuits.

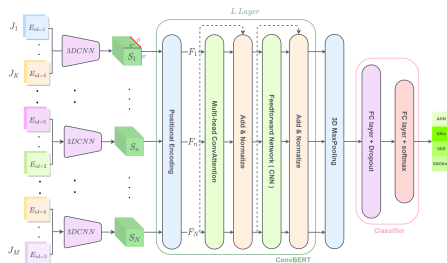


ConvEntion : (1) Un modèle état-de-l'art

Etat de l'art sur SDSS (Sloan Digital Sky Survey; 4 classes).

Model	Bands	Type of data	Accuracy	F1 Score	Num params
ConvEntion[2]	ugriz	Images	79.83	70.62	1.253M
CNN+GRU[4]	ugriz	Images	66.39	63.22	1.993M
CNN+LSTM[11]	ugriz	Images	64.08	60.65	2.190M
SuperNNova _(Bayes)	ugriz	Light curves	65.54	55.40	-
SITS-BERT	ugriz	Light curves	67.43	51.60	0.596M
SCONE (CNN)	ugriz	Light curves	62.57	50.43	22.2K
SuperNNova _(RNN)	ugriz	Light curves	56.30	42.60	-
LSTM	ugriz	Light curves	55.24	40.33	60K

ConvEntion : (2) L'architecture



- ▶ 1 transformer,
- ▶ Formatage des données (absence d'images + bandes),
- ▶ Réduction de la complexité mémoire + calculatoire,
- ▶ Weighted Random Sampling via PyTorch; Paszke et al. 2019,
- ▶ Apprentissage avec base à label non confirmés puis confirmés.

Plan

Introduction/Etat de l'art

Notre proposition ICT-ConvEntion

Expériences & Résultats

Conclusions et perspectives

ICT-ConvEntion

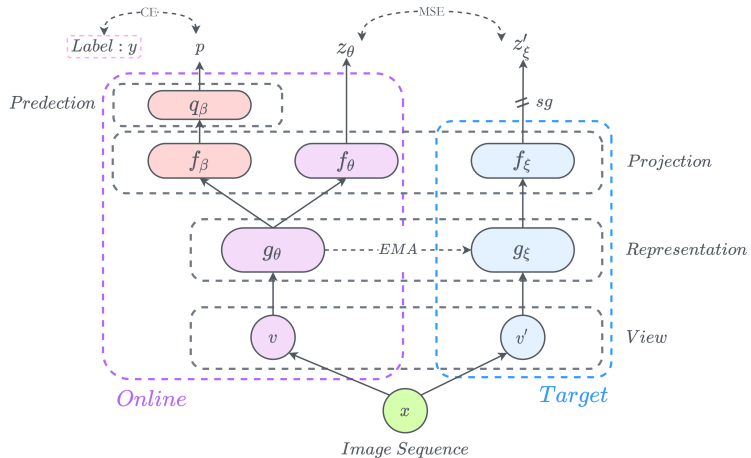
ICT-ConvEntion = Inbalanced Classes Treatment - ConvEntion

- ▶ Apprentissage supervisé,
- ▶ Utilisation d'une technique d'augmentation,
- ▶ Apprentissage avec archi "similaire" à BYOL ou DINO.

Pourquoi une telle approche ?

- ▶ Structurer l'espace (contrastive learning),
- ▶ Archi utilisable en auto-supervisé et/ou supervisé.

ICT-ConvEntion : L'architecture d'apprentissage



$$\mathcal{L} = \mathcal{L}_{MSE} + \mathcal{L}_{CE} + \tilde{\mathcal{L}}_{MSE} + \tilde{\mathcal{L}}_{CE}.$$

Plan

Introduction/Etat de l'art

Notre proposition ICT-ConvEntion

Expériences & Résultats

Conclusions et perspectives

SDSS (Sloan Digital Sky Survey) :

- ▶ 5 bandes : u,g,r,i,z (resp 365, 475, 658, 806, 900 nm),
- ▶ SDSS "Survey" de 2005 à 2008,

Classe	BDD Pre-Train (photométrique)	BDD FineTune (spectrométrique)	BDD Test (spectrométrique)
# AGN	362	362	182
# SNIa	1448	400	99
# Variable	1290	1290	645
# SNOther	2041	72	17

- ▶ → fort déséquilibre,
- ▶ Photométrique : Les labels peuvent être erronés,
- ▶ Spectroscopique : Les labels sont certains.

Résultats

Model	Accuracy (%)	F1 (%)
ICT-ConvEntion	82,18	75,33
ConvEntion [2]	79,83	70,62
CNN+GRU [4]	66,39	63,22
CNN+LSTM [11]	64,08	60,65

- ▶ Accuracy : +2,3% ; F1 : +4,7%,
- ▶ ICT-ConvEntion : Accuracy par classe de 77% à 83% (ecart min-max = 6%).
- ▶ ConvEntion : Accuracy par classe de 67% à 81% (ecart min-max = 14%).
- ▶ MSE ↘ variance inter-classe et Cross-Entropie ↗ séparabilité,
- ▶ Apprentissage 3 × plus long que ConvEntion (= 9h).

Plan

Introduction/Etat de l'art

Notre proposition ICT-ConvEntion

Expériences & Résultats

Conclusions et perspectives

Conclusions et perspectives

ICT-ConvEntion :

- ▶ Classification des séries temporelles d'images astronomiques,
- ▶ Etat de l'art sur SDSS 4 classes,
- ▶ Gestion du déséquilibre (fonctions de perte réduisant les distances intra-classes (MSE) et inter-classes (CE)).

Perspectives :

- ▶ Utiliser d'autres approches pour équilibrer,
- ▶ Exploiter les méta-données ou informations du domaine,
- ▶ Augmenter le nombre de classes,
- ▶ Utiliser les modèles physiques,
- ▶ Faire de la détection précoce, etc.