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

Anass BAIROUK[1], Marc CHAUMONT[1,2], Dominique FOUCHEZ[3],
Jérome PASQUET[4], Frédéric COMBY[1]
[1] *LIRMM, Équipe ICAR, Université de Montpellier, CNRS, Montpellier, France*
[2] *Université de Nîmes, Nîmes, France*
[3] *CPPM, Université d'Aix-Marseille, Marseille, France*
[4] *TETIS, Université Paul Valéry Montpellier 3, Montpellier, France*
{anass.bairouk, marc.chaumont, frederic.comby}@lirmm.fr, dominique.fouchez@cppm.in2p3.fr, jerome.pasquet@univ-montp3.fr

Résumé

En classification de séquences d'images astronomiques, l'approche état-de-l'art (ConvEntion) repose sur l'utilisation d'une architecture basé sur une convolution 3D et un transformer. Cette architecture ConvEntion ne gère pas parfaitement le déséquilibre de classes. Dans cet article, nous proposons d'améliorer cela en s'appuyant sur les méthodologies auto-supervisées. Nous réduisons la variance intra-classe en passant par une architecture à deux branches. Chacune des deux branches traite une version augmentée de la donnée d'entrée. Dans le même temps, nous conservons la contrainte de classification ce qui nous permet de faire l'apprentissage sur un petit ensemble de données labellisées. Les résultats de notre modèle ICT-ConvEntion nous ont permis d'obtenir une amélioration de l'exactitude (accuracy) de 2.3% et du score F1 de 4.7% sur la base SDSS Supernova Survey.

Mots clefs

Astrophysique, Séquence d'images, Deep Learning, Classification, Transformer, DINO, BYOL, ConvEntion.

1 Introduction

Afin de déterminer les constantes astrophysiques et mieux comprendre l'univers, les astrophysiciens ont notamment besoin de détecter les supernovae Ia [1]. La détection consiste à pointer un télescope vers une zone du ciel, à suivre l'événement durant un certain nombre de nuits [2], puis à classer l'évenement. Historiquement, la séquence d'imagettes centrée sur l'événement était transformée en plusieurs séries de scalaires (une par bande) et appelée courbe de lumière. Ces courbes de lumière étaient alors utilisées pour classer le phénomène dans une des classes d'intérêt comme par exemple une supernova Ia.

1. Les supernovae de type Ia sont particulièrement intéressantes en raison de leur mécanisme d'explosion standard produisant une énergie de flux presque identique pour chaque supernova, ce qui permet de déduire la distance lumineuse de la supernova qui a explosé.
2. Une supernova est observable durant 100 à 200 nuits.

Nous avons récemment proposé l'approche « ConvEntion » qui permet de traiter directement la série d'imagettes issue du télescope. L'approche repose entre autre sur l'utilisation d'un "Transformer", le formatage des données afin de gérer l'absence d'imagettes et de bandes de fréquence, la réduction de la complexité via l'utilisation d'un 3DCNN, etc. Nous avons publié une version courte dans la conférence française GRETSI [1] et une version journal dans Astronomy & Astrophysics [2].
ConvEntion dépasse largement l'état de l'art sur la base Sloan Digital Sky Survey (SDSS) [3] sur une tache de classification à 4 classes difficiles. Nous améliorons l'exactitude ("accuracy") de 13% par rapport à l'approaches préliminaire de [4] et 14% par rapport à la meilleur approche basée courbe de lumière [5].
Dans cet article, nous proposons de mieux gérer le déséquilibre des classes en forçant les représentations latentes de différentes versions d'une même séquence d'image à être proches. Pour cela, nous avons repris le principe d'apprentissage de type « teacher-student » présents dans les approches auto-supervisé comme BYOL [6] et DINO [7]. Ainsi, en plus d'avoir un terme de perte (i.e. loss) pour la classification supervisée, nous ajoutons un terme de perte non-supervisé en conjonction d'une architecture « à la manière » de BYOL/DINO.
Dans la section 2 nous présentons notre proposition appelée ICT-ConvEntion (pour *Inbalanced Classes Treatment* ConvEntion) qui est une amélioration de ConvEntion. Dans la section 3 nous présentons brièvement les résultats [3].

2 Architecture utilisée pour l'apprentissage de ICT-ConvEntion

ICT-ConvEntion reprend la structure d'apprentissage des approches auto-supervisées récentes comme [6, 7] etc. Cependant, ICT-ConvEntion est une architecture qui peut apprendre avec des petits ensembles de données. Par ailleurs comme les méthodes contrastives [9, 10], notre approche

3. Ces résultats sont plus amplement détaillés dans la thèse d'Anass Bairouk [8].

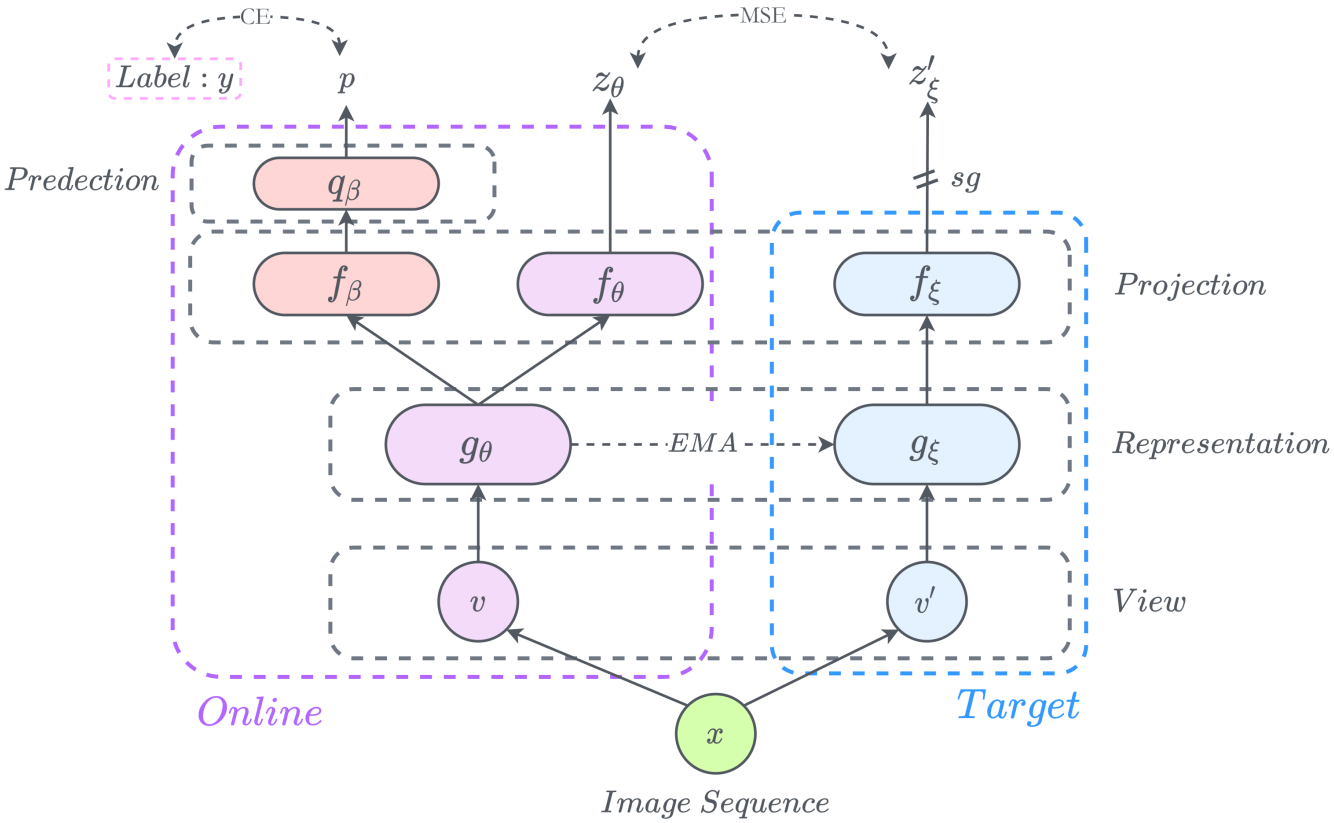

FIGURE 1 – *L'architecture générale d'ICT-ConvEntion. L'apprentissage d'ICT-Convention consiste à minimiser la différence entre z_θ et z'_ξ en utilisant l'erreur quadratique moyenne, ainsi qu'à minimiser l'entropie croisée entre y et p , où θ représente les poids, ξ représente les poids mis à jour par moyenne mobile exponentielle (EMA) de θ, et sg signifie "stop-gradient" (le sg empêche la rétropropagation sur le réseau cible).*

repose sur l'augmentation des données pour l'apprentissage de la représentation. C'est en partie grâce cette augmentation que nous gérons mieux le déséquilibre des classes.

La figure 1 illustre la structure d'apprentissage d'ICT-ConvEntion. Dans un premier temps, nous créons deux séquences d'images augmentées (des vues) v et v', à partir de la séquence originale. Plusieurs techniques d'augmentation ont été utilisées, notamment la suppression aléatoire d'images de la séquence pour en créer une nouvelle, la rotation des images de la séquence, le retournement horizontal et vertical, et le décalage temporel pour créer une séquence plus courte que l'original.

De la même manière que pour BYOL/DINO, nous utilisons deux réseaux neuronaux dont l'architecture est identique (nous utilisons notre architecture ConvEntion [2]), appelés réseau *en ligne* et réseau *cible*. Le réseau en ligne (resp. cible) est défini par un ensemble de poids θ, β (resp. ξ), et est composé de trois (resp. deux) étapes. A la première étape, se trouve un encodeur g_θ (resp. g_ξ) qui reprend l'architecture de ConvEntion. L'objectif de l'encodeur est l'apprentissage de la représentation. La deuxième étape contient deux parties pour le réseau en ligne, f_β et f_θ afin de projeter la représentation issue de g_θ. Le réseau cible ne contient qu'une seul brique de projection f_ξ. La projection f_β est utilisée par la troisième étape (pour la classification supervisée). Les projections f_θ et f_ξ sont utilisées pour minimiser la distance entre les représentations des réseaux en ligne et cible. La dernière étape est spécifique au réseau en ligne et à la partie classification. Cette étape contient une couche de prédiction q_β qui est utilisée pour minimiser l'entropie croisée avec le label vérité terrain. A noter que g_θ et f_θ partagent la même architecture respectivement avec g_ξ et f_ξ.

Après avoir obtenu les deux vues augmentées, v et v', de la séquence d'images, nous transmettons la première vue v', au réseau cible pour qu'elle passe par l'encodeur et le projecteur afin d'obtenir la sortie z'_ξ. Pendant ce temps, nous faisons passer v par le réseau en ligne pour obtenir la projection z_θ et la prédiction de classe p. Nous normalisons ensuite les projections des deux réseaux $\tilde{z}_\theta = z_\theta / ||z_\theta||_2$ et $\tilde{z'}_\xi = z'_\xi / ||z'_\xi||_2$.

La fonction de perte inspirée de l'apprentissage auto-supervisé, est appliquée sur les deux projections normalisées comme suit :

$$\mathcal{L}_{MSE}=||\tilde{z}_\theta - \tilde{z}'_\xi||_2^2 = 2 - 2 \cdot \frac{\langle z_\theta, z'_\xi \rangle}{||z_\theta||_2 \cdot ||z'_\xi||_2}. \quad (1)$$

La fonction de perte de classification est calculée entre le label, y, et la prédiction du réseau en ligne, p, comme suit :

$$\mathcal{L}_{CE} = -\sum_{c=1}^{M} y_{x,c} \log(p_{v,c}), \quad (2)$$

où M est le nombre de classes, c la classe, tandis que x et v sont respectivement la séquence d'entrée et la nouvelle vue générée à partir de la séquence d'entrée. Nous symétrisons les deux pertes $\mathcal{L}_{MSE}$ et $\mathcal{L}_{CE}$, en envoyant v' au réseau en ligne, et v au réseau cible, afin d'obtenir deux autres fonction de perte $\tilde{\mathcal{L}}_{MSE}$ et $\tilde{\mathcal{L}}_{CE}$. Nous calculons ensuite la perte totale comme suit :

$$\mathcal{L} = \mathcal{L}_{MSE} + \mathcal{L}_{CE} + \tilde{\mathcal{L}}_{MSE} + \tilde{\mathcal{L}}_{CE}. \quad (3)$$

A noter que les paramètres du réseau cible sont mis à jour à l'aide d'une moyenne mobile exponentielle (EMA) des paramètres du réseau en ligne. Les poids ξ sont mis à jour en suivant la relation $\xi \leftarrow \tau\xi + (1-\tau)\theta$ où $\tau \in [0, 1]$ est le taux de déclin (decay). Les projections f_β, f_θ, f_ξ consistent en deux couches entièrement connectées avec une activation ReLU et un "dropout" entre les deux. Pour le réseau en ligne, la couche de prédiction q_β consiste en une couche entièrement connectée qui prend en entrée la sortie de f_β et produit une sortie avec une dimension qui correspond au nombre de classes dans l'ensemble de données.

3 Détails expérimentaux et résultats/discussions

3.1 La base de données issue du SDSS

Le Sloan Digital Sky Survey (SDSS) [3] est un programme d'étude qui recueille des images, des spectres et des informations descriptives de millions d'objets célestes à l'aide d'un télescope dédié équipé d'instruments photométriques et spectroscopiques. Nous utilisons dans cet article le SDSS Supernova Survey, une composante de l'extension du SDSS-II portant sur la période 2005 à 2008. L'étude des supernovae consistait à imager la même région du ciel tous les deux soirs, en utilisant cinq filtres à large bande pour construire une vaste base de données d'images afin de découvrir de nouveaux objets célestes.

La base de données issue de cette étude comprend des séquences d'images d'étoiles variables galactiques, de noyaux actifs de galaxie (AGN), de supernovae (SNe) et d'autres transitoires astronomiques. Certaines de ces séquences sont également complétées du résultat spectroscopique, pour notamment identifier avec certitude les SNe et mesurer leur décalage vers le rouge. Cette base de données comporte un fort déséquilibre des classes, qui peut constituer un obstacle important pour l'apprentissage profond ainsi que la présence de classes dont les caractéristiques sont très similaires.

Nous avons séparé le jeu de données en deux ensembles de données : l'un contenant des données de typage photométrique (dont la vérité terrain peut être erronée) et l'autre des données confirmées par spectroscopie (la vérité terrain est certaine). Nous avons d'abord entraîné le modèle sur les données photométriques, puis nous avons utilisé l'apprentissage par transfert pour affiner le modèle sur les données confirmées par spectroscopie. Le tableau 1 résume la partition et les classes des données que nous avons utilisées dans ce travail.

Class	Pre-Train	FineTune	Test
AGN	362	362	182
SNIa	1448	400	99
Variable	1290	1290	645
SNOther	2041	72	17

TABLEAU 1 – *Nombre d'objets dans chaque ensemble de données. Le Pre-Train ne contient que des données photométriques, le FineTune et le Test ne contiennent que des données confirmées par spectroscopie.*

3.2 Hypers-paramètres et détails d'implémentation

Le modèle a été entraîné en utilisant l'optimiseur adamw, une taille de lot de 128 répartie sur 4 GPU. Le taux d'apprentissage a été fixé à 2×10^{-3} et un « dropout » de 0,3. Nous diminuons le taux d'apprentissage avec un cosinus « schedule ». Le taux de décroissance pour l'EMA est fixé à $\tau = 0,99$. Pour la brique « ConvEntion » nous avons utilisé les mêmes paramètres que ceux utilisés dans l'article original [2] avec $K = 3$ et $M = 99$. Le nombre de couches de ConvEntion est fixé à $L = 2$ et le nombre de têtes d'attention à $T = 4$. Les modèles sont entraînés à l'aide d'une validation croisée de cinq plis. Toutes les architectures présentées dans cet article suivent ce même processus et sont implémentées en utilisant PyTorch.

3.3 Résultats et discussions

Le tableau 2 contient les évaluations sur les quelques approches état-de-l'art utilisant des séries d'images astronimiques. Notre solution obtient une exactitude (accuracy) de 82,18% et un score F1 [4] de 75,33%. Notre proposition ICT-ConvEntion permet d'obtenir un gain en score F1 d'environ 5% par rapport à l'approche ConvEntion. Par ailleurs, les exactitudes (accuracy) par classe sont bien plus équilibrées pour ICT-ConvEntion (cela va de 77% à 83%) que pour ConvEntion (cela va de 67% à 81%) [5]. Ces résultats confirment l'intérêt à introduire des fonctions de perte

4. $F1 = 2\frac{presicion.rappel}{precision+rappel}$.
5. Voir les matrices de confusion dans la thèse d'Anass Bairouk[8].

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

TABLEAU 2 – *Comparaison des performances en termes de score F1 moyen et de précision moyenne sur 5 plis de validation croisée.*

contraignant l'espace latent. Cela permet, entre autres, de mieux prendre en compte le déséquilibre des classes, mais également d'augmenter l'exactitude (accuracy) de la classification. La fonction de perte MSE (Eq. 1) réduit efficacement la variance intra-classe, ce qui permet à la fonction de perte de classification (Eq. 2) d'augmenter la séparabilité des classes. Notons tout de même que le temps d'apprentissage de la méthode ICT-ConvEntion est assez coûteux, puisqu'il prend trois fois plus de temps que l'apprentissage du modèle ConvEntion. La réduction du temps d'apprentissage est donc une des possibles pistes de recherches futures.

4 Conclusion

En conclusion, l'approche ICT-ConvEntion proposée pour traiter la classification des séries temporelles d'images astronomiques s'est avérée efficace pour augmenter l'exactitude (accuracy) de la classification. Les résultats montrent une augmentation de l'exactitude de 2,3% et du score F1 de 4,7% sur la base le SDSS Supernova Survey par rapport à l'état de l'art ConvEntion. Cette amélioration est grandement due à une meilleure gestion du déséquilibre des classes grâce à une fonction de perte réduisant les distances inter-classes, qui est obtenue via une architecture « à la BYOL/DINO », tout en gardant la fonction de perte pour la classification.

Remerciement

Ce travail a été réalisé grâce au soutien du projet ANR DEEPDIP (ANR-19-CE31-0023). Ce travail utilise les données du Sloan Digital Sky Survey (SDSS).

Références

[1] Anass Bairouk, Marc Chaumont, Dominique Fouchez, Jérome Pasquet, et Frédéric Comby. ConvEntion : Classification des séries chronologiques d'images astronomiques à l'aide d'attention convolutive. Dans *GRETSI 2022 - 28e Colloque Francophone de Traitement du Signal et des Images*, numéro 001-034 dans 28, pages 137–140, Nancy, France, Septembre 2022.

[2] Anass Bairouk, Marc Chaumont, Dominique Fouchez, Frédéric Comby, Jerome Pasquet, et Julian Bautista. Astronomical image time series classification using CONVolutional attENTION (ConvEntion). *Astronomy and Astrophysics - A&A*, 673 :A141, 2023.

[3] Jon A. Holtzman et al.. The sloan digital sky survey-ii : Photometry and supernova ia light curves from the 2005 data. *The Astronomical Journal*, 136(6) :2306, nov 2008.

[4] Catalina Gómez, Mauricio Neira, Marcela Hernández Hoyos, Pablo Arbeláez, et Jaime E Forero-Romero. Classifying image sequences of astronomical transients with deep neural networks. *Monthly Notices of the Royal Astronomical Society*, 499(3) :3130–3138, 10 2020.

[5] A Möller et T de Boissière. SuperNNova : an open-source framework for Bayesian, neural network-based supernova classification. *Monthly Notices of the Royal Astronomical Society*, 491(3) :4277–4293, 12 2019.

[6] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre H. Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Daniel Guo, Mohammad Gheshlaghi Azar, Bilal Piot, Koray Kavukcuoglu, Rémi Munos, et Michal Valko. Bootstrap your own latent a new approach to self-supervised learning. Dans *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20, Red Hook, NY, USA, 2020. Curran Associates Inc.

[7] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, et Armand Joulin. Emerging properties in self-supervised vision transformers. Dans *2021 IEEE/CVF International Conference on Computer Vision, ICCV 2021, Montreal, QC, Canada, October 10-17, 2021*, pages 9630–9640. IEEE, 2021.

[8] Anass Bairouk. *Astronomical Image Time-series Classification Using Deep Learning*. Theses, Université de Montpellier, Octobre 2023.

[9] Ting Chen, Simon Kornblith, Mohammad Norouzi, et Geoffrey Hinton. A simple framework for contrastive learning of visual representations. Dans *Proceedings of the 37th International Conference on Machine Learning*, ICML'20. JMLR.org, 2020.

[10] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, et Ross Girshick. Momentum contrast for unsupervised visual representation learning. Dans *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9726–9735, 2020.

[11] Rodrigo Carrasco-Davis, Guillermo Cabrera-Vives, Francisco Förster, Pablo A. Estévez, Pablo Huijse, Pavlos Protopapas, Ignacio Reyes, Jorge Martínez-Palomera, et Cristóbal Donoso. Deep learning for image sequence classification of astronomical events. *Publications of the Astronomical Society of the Pacific*, 131(1004) :108006, sep 2019.