

Apprendre avec peu d'exemples : Une approche auto-supervisée basée sur les segments avec application à la télédétection

Antoine Saget*, Baptiste Lafabregue*, Antoine Cornuéjols**, Pierre Gançarski*

*ICube, Université de Strasbourg, France
prenom.nom@unistra.fr

**UMR MIA-Paris, AgroParisTech, INRAE - University of Paris-Saclay
antoine.cornuejols@agroparistech.fr

Résumé. Pour tirer profit des données non étiquetées en classification de séries temporelles nous introduisons une nouvelle méthode de sélection des paires positives pour l'apprentissage auto-supervisé contrastif (contrastive self-supervised learning). En utilisant les groupes préexistants (segments) dans les données, cette approche élimine le besoin d'augmenter les données, mécanisme d'habitude indispensable en apprentissage auto-supervisé contrastif. Nous montrons qu'elle permet d'obtenir des performances comparables à celles des modèles de l'état de l'art en nécessitant près de moitié moins de données étiquetées. Nous validons expérimentalement notre méthode sur un nouveau jeu de données à grande échelle de séries temporelles d'images satellites (SITS) Sentinel-2 pour la classification des cultures agricoles. Nous atteignons une accuracy de 80% avec 50 exemples étiquetés par classe, tandis que la meilleure méthode comparée en nécessite 100 pour atteindre le même résultat.

1 Introduction

Les séries d'images satellites (SITS) sont désormais largement disponibles avec des résolutions temporelles et spatiales élevées (Pelletier et al., 2019) et s'avèrent précieuses en observation de la Terre. Bien que les méthodes d'apprentissage automatique actuelles donnent des résultats fiables, elles nécessitent généralement de grandes quantités de données étiquetées (Rußwurm et al., 2019) souvent difficiles à obtenir. En conséquence, la proportion de données étiquetées est très faible par rapport au volume quotidien de données non étiquetées générées par les satellites, laissant ainsi la majorité des données disponibles inexploitées par les méthodes d'apprentissage supervisé classiques.

L'apprentissage auto-supervisé (Self-Supervised Learning, SSL) utilise ces données non étiquetées pour apprendre de nouvelles représentations applicables à diverses tâches en aval. Ces nouvelles représentations offrent une plus grande efficacité en classification que les méthodes supervisées standard (Henaff, 2020). L'apprentissage auto-supervisé apparaît donc comme un candidat naturel pour la classification de SITS. Les méthodes SSL, principalement basées sur des fonctions de coût contrastives ou de reconstruction générative, ont connu un grand

succès en vision par ordinateur (Chen et al., 2020; He et al., 2020; Grill et al., 2020; Chen et He, 2021; Le-Khac et al., 2020). S'appuyant sur ce succès, des méthodes SSL spécifiques aux séries temporelles ont été proposées (Yue et al., 2022; Cheng et al., 2023). Bien que les méthodes SSL basées sur la reconstruction générative et les méthodes SSL contrastives soient toutes les deux prometteuses, dans cet article, nous nous concentrons exclusivement sur les méthodes SSL contrastives.

En apprentissage SSL contrastif, les exemples positifs sont des points de données similaires, généralement créés par augmentation. Cependant, générer des exemples positifs est difficile pour les données SITS, car les augmentations couramment utilisées en traitement d'images ne peuvent pas être directement appliquées aux séries temporelles. Nous proposons donc d'utiliser des groupements préexistants de données (segments) comme exemples positifs pour adapter le SSL contrastif aux SITS.

La suite de cet article est organisée comme suit : dans la section 2, nous introduisons la classification des séries temporelles et l'apprentissage auto-supervisé. Dans la section 3, nous présentons un nouveau jeu de données à grande échelle pour la classification des cultures agricoles à partir de SITS. Dans la section 4, nous détaillons la méthode proposée et montrons que l'utilisation de groupements de données préexistants, dans notre cas les limites des parcelles agricoles, constitue une alternative efficace aux augmentations pour la sélection d'exemples positifs en SSL contrastif. Enfin, dans la section 5, nous présentons les résultats expérimentaux, montrant que la représentation apprise atteint des performances similaires à la meilleure méthode comparée en utilisant presque moitié moins de données d'entraînement étiquetées.

2 Apprentissage auto-supervisé et séries temporelles

Ruiz et al. (2021) comparent les performances de 16 méthodes de classification de séries temporelles sur 26 jeux de données provenant des archives de l'UEA (Bagnall et al., 2018). Ces méthodes couvrent une large gamme d'approches telles que les méthodes basées sur les shapelets, le dynamic time warping, l'extraction de motifs fréquents et les techniques convolutives...

Dans ROCKET (Dempster et al., 2020), des milliers de noyaux convolutifs aléatoires sont appliqués à chaque série temporelle d'entrée pour générer une nouvelle représentation sur laquelle la classification est effectuée. La méthode HIVE-COTE (Lines et al., 2016) est un ensemble de quatre classifieurs différents, l'un d'entre eux étant un ensemble de ROCKET. Bien qu'elle atteigne de très bons résultats, en particulier avec HIVE-COTE 2 (Middlehurst et al., 2021), elle s'avère beaucoup plus lente que les autres méthodes, ce qui la rend quasi inutilisable sur des grands jeux de données. Enfin, la méthode InceptionTime (Ismail Fawaz et al., 2020; Ismail-Fawaz et al., 2023) adapte l'architecture Inception-v4 aux séries temporelles et obtient des résultats similaires à ceux de HIVE-COTE tout en étant significativement plus rapide.

L'apprentissage auto-supervisé (Self-Supervised Learning, SSL) permet d'extraire des caractéristiques utiles de données non étiquetées en apprenant à partir d'étiquettes autogénérées par le biais d'une tâche dédiée (souvent appelée tâche prétexte). La représentation apprise peut ensuite être utilisée pour diverses tâches en aval. Pour cela, l'approche **Generative** implique le masquage de certaines parties des données d'entrée et l'entraînement d'un modèle apprenant à reconstruire ces parties masquées. L'approche **Contrastive** entraîne un modèle minimisant

la distance entre des paires d'exemples similaires (positifs) et maximisant la distance entre des paires dissemblables (négatifs).

L'apprentissage auto-supervisé contrastif a connu un succès notable sur des données image (Chen et al., 2020; He et al., 2020; Grill et al., 2020; Chen et He, 2021; Le-Khac et al., 2020). Toutefois, ces méthodes reposent généralement sur des augmentations de données difficilement transférables aux séries temporelles. Dans SimCLR (Chen et al., 2020), les exemples positifs sont générés en appliquant différentes augmentations au même point de donnée, tandis que les exemples négatifs sont les points restants du batch. Dans TimeMAE, Cheng et al. (2023) utilisent un transformer adapté aux séries temporelles entraîné conjointement sur une tâche de reconstruction et une tâche de classification.

3 Jeu de données massif de classification de séries temporelles Sentinel-2

Nous avons créé un nouveau jeu de données de classification des terres cultivées inspiré par Rußwurm et al. (2019), comprenant plus de 5,8 millions de parcelles étiquetées en France métropolitaine. Nous détaillons sa construction dans la suite de cette section.

Labels. Le Registre Parcellaire Graphique (RPG), fourni annuellement par l'Institut National de l'Information Géographique et Forestière (IGN), répertorie toutes les parcelles cultivées en France avec leurs géométries et leurs types de cultures. Pour notre jeu de données, nous avons utilisé les types de cultures du RPG de 2022 en France métropolitaine comme étiquettes de classe. Il en résulte un jeu de données fortement déséquilibré avec 232 classes comme le montre la Figure 1.

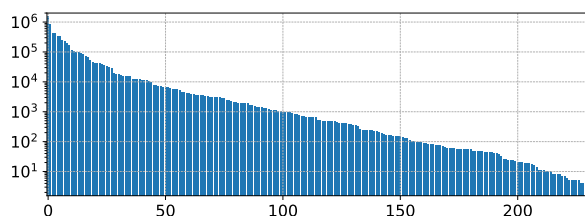


FIG. 1 – Répartition des classes (types de cultures) dans le RPG 2022. Échelle logarithmique en ordonnées.

Features. Pour chaque parcelle agricole du RPG, 100 séries temporelles Sentinel-2 (satellite d'observation de la Terre) à l'intérieur de celle-ci sont extraites au hasard. Les parcelles comportant moins de 100 pixels (moins d'un hectare) sont écartées. Chaque série temporelle s'étend sur 12 bandes radiométriques et sur 60 pas de temps (du 1er février au 30 novembre 2022).

Nous obtenons donc finalement un jeu de données d'environ 5,8 millions de parcelles, chacune constituée de 100 séries temporelles distinctes et étiquetées par l'une des 232 classes possibles.

4 Exploiter les groupes préexistants dans les données

Dans les domaines pour lesquels l’augmentation des données reste un défi, nous pensons que l’identification de groupes préexistants dans les données pour sélectionner des paires positives en SSL contrastif est une alternative viable.

Ainsi, dans cet article, nous utilisons, dans un premier temps, les parcelles agricoles comme groupes préexistants et définissons les exemples positifs à partir de séries temporelles de pixels d’une même parcelle. Nous appelons cette approche “Groupes comme paires positives” (GaPP). Ces groupes sont particulièrement adaptés, car, bien que chaque série temporelle de pixels provenant de la même parcelle appartienne à la même classe, elles sont néanmoins différentes les unes des autres. La reproduction de cette diversité par des augmentations classiques n’est pas triviale et nécessiterait un réglage minutieux des paramètres et des connaissances d’expert.

Dans un second temps, nous utilisons des n -uplets de données d’une même parcelle : n séries temporelles d’une même parcelle sont soumises à l’encodeur. Leurs représentations résultantes sont alors agrégées via une couche de moyennage. Cette opération est nommée “Average Pooling” (AvgP). La représentation sortante de la couche de moyennage peut alors être utilisée pour former les exemples positifs en prenant deux tels n -uplets distincts dans une même parcelle. Nous nommons cette approche (AvgG + GAPP). Ces deux approches sont applicables à toute méthode SSL contrastive. Dans la section suivante, nous présentons nos expérimentations utilisant SimCLR.

5 Expériences et résultats

5.1 Sous-ensembles utilisés et méthodes comparées

Quatre sous-ensembles de données ont été créés à partir du jeu de données complet. Un premier, **contrast**, contient environ 4 millions de séries non étiquetées choisies aléatoirement. Il sera utilisé pour l’apprentissage auto-supervisé contrastif.

Nous avons choisi arbitrairement 20 classes d’intérêt pour simuler un expert fournissant des exemples étiquetés à la main. Nous avons alors extrait trois sous-ensembles ne contenant que des séries étiquetées par l’une de ces 20 classes :

- **train** : sous-ensemble de **contrast** équilibré et de taille variable suivant les expériences.
- **validation** : sous-ensemble équilibré contenant 10 séries par classe.
- **test** : sous-ensemble de 20 000 séries, non équilibré et donc plus proche de la distribution réelle des données.

Les ensembles **train**, **validation** et **test** sont mutuellement exclusifs.

Nous avons comparé notre approche avec quatre méthodes de la littérature. Une régression logistique comme point de référence. Rocket (Dempster et al., 2020) et InceptionTime (Ismail Fawaz et al., 2020), deux méthodes parmi les plus performantes en classification de séries temporelles sur l’archive UEA, et TimeMAE (Cheng et al., 2023), une méthode d’apprentissage auto-supervisé pour les séries temporelles. Pour notre contribution, nous utilisons SimCLR (Chen et He, 2021) avec un encodeur ResNet adapté aux séries temporelles. Tous les paramètres des méthodes sont donnés en Annexe.

5.2 Protocole d'apprentissage et d'évaluation

Les méthodes standard sont entraînées sur *train* et évaluées sur *test*. Les méthodes auto-supervisée sont entraînées sur *contrast*. Une régression logistique est ensuite entraînée avec la nouvelle représentation sur *train* et évaluée sur *test*.

Lors de nos expériences, nous avons observé des cas d'effondrement du modèle lors de l'apprentissage auto-supervisé. Un ensemble réduit de données étiquetées (ici l'ensemble *validation*) permet de mettre en évidence cet effondrement et ainsi d'arrêter l'apprentissage avant qu'il ne se produise. L'utilisation de cet ensemble, bien que très réduit, nous écarte d'un cadre entièrement non supervisé. Cela suggère que nos alternatives d'augmentation ne sont pas assez robustes et nécessitent des améliorations pour un fonctionnement complètement non supervisé.

Les prédictions de chaque modèle sont évaluées sur les classes des parcelles, obtenues pour chacune d'elles par un vote majoritaire sur les prédictions de ses 100 séries temporelles.

5.3 Résultats

Tous les résultats sont des accuracy de classification moyennes sur 20 répétitions avec différents ensembles *train*. Les écarts-type sont inférieurs à 4 (resp. 2) sur les ensembles d'entraînement avec 5 ou 10 (resp. 50 ou 100) échantillons par classe. Nous les avons omis car ils ne diffèrent qu'en fonction de la taille de l'ensemble d'entraînement et non des méthodes. Pour SimSLR, la seule augmentation que nous avons utilisée est un bruit gaussien ($\mu = 0$, $\sigma = 0.1$) sur les données brutes.

Méthode			N exemple/classe dans train			
			5	10	50	100
Régression Logistique			42	49	68	74
MiniRocket			47	58	75	79
InceptionTime			56	65	77	80
TimeMAE			53	61	75	79
SimCLR	<i>Scénario</i>	<i>n-uplet</i>				
	Bruit seul	-	45	53	70	75
	Bruit + AvgP	4	45	53	70	75
	Bruit + GaPP	-	50	60	76	80
	Bruit + AvgP + GaPP	2	56	66	78	82
		4	62	70	80	82
		8	61	69	80	82

TAB. 1 – Performances des méthodes comparées et ablations de nos propositions GaPP et AvgP sur SimCLR

Le Tableau 1 donne les résultats pour les cinq méthodes sur l'ensemble *test* pour différents ensembles *train*. Il montre (en gras) que la meilleure précision est obtenue avec SimCLR, scénario (Bruit+AvgP+GaPP) avec quatre séries temporelles moyennées en sortie de l'encodeur.

Le principal avantage de notre approche est l'efficacité en nombre d'étiquettes. Par exemple, notre méthode avec 50 parcelles étiquetées par classe atteint une précision de 80%, identique à celle d'InceptionTime avec 100 exemples (en bleu dans le tableau).

6 Discussion

Dans cet article, nous avons proposé une alternative aux augmentations en apprentissage autosupervisé contrastif basé sur des groupes préexistants dans les données. Les expériences menées sur un nouvel ensemble de données à grande échelle montrent que l’efficacité des étiquettes sur une tâche de classification en aval a presque doublé par rapport à d’autres méthodes. Néanmoins, plusieurs questions restent ouvertes pour les travaux futurs :

- Un ensemble de données de validation certes réduit est néanmoins nécessaire pour éviter l’effondrement du modèle. Des travaux portant sur la taille minimale (espérée nulle) de cet ensemble de données sont en cours.
- Bien que l’apprentissage auto-supervisé soit réputé pour produire des représentations générales, nous n’avons évalué notre méthode que sur une seule tâche de classification en aval. D’autres recherches devraient tester la polyvalence des représentations apprises dans d’autres tâches.
- Nous avons utilisé les limites des parcelles comme groupes dans nos expériences. Il serait intéressant de tester notre approche sur d’autres ensembles de données avec des groupements différents ou de créer des groupes synthétiques par segmentation ou clustering.
- Nous nous sommes uniquement concentrés sur le SSL contrastif, mais l’utilisation des informations de groupe dans le SSL génératif reste à explorer.

7 Remerciements

Les auteurs remercient l’Agence Nationale de la Recherche (ANR) pour avoir financé ce travail via les projets ArtIC et HERELLES. Les auteurs expriment également leur gratitude à l’Agence spatiale européenne (ESA) et au programme Copernicus pour l’accès gratuit et public aux données Sentinel-2.

Références

- Bagnall, A., H. A. Dau, J. Lines, M. Flynn, J. Large, A. Bostrom, P. Southam, et E. Keogh (2018). The uea multivariate time series classification archive, 2018. *arXiv preprint arXiv :1811.00075*.
- Chen, T., S. Kornblith, M. Norouzi, et G. Hinton (2020). A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pp. 1597–1607. PMLR.
- Chen, X. et K. He (2021). Exploring simple siamese representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 15750–15758.
- Cheng, M., Q. Liu, Z. Liu, H. Zhang, R. Zhang, et E. Chen (2023). Timemae : Self-supervised representations of time series with decoupled masked autoencoders. *arXiv preprint arXiv :2303.00320*.

- Dempster, A., F. Petitjean, et G. I. Webb (2020). Rocket : exceptionally fast and accurate time series classification using random convolutional kernels. *Data Mining and Knowledge Discovery* 34(5), 1454–1495.
- Grill, J.-B., F. Strub, F. Altché, C. Tallec, P. Richemond, E. Buchatskaya, C. Doersch, B. Avila Pires, Z. Guo, M. Gheshlaghi Azar, et al. (2020). Bootstrap your own latent-a new approach to self-supervised learning. *Advances in neural information processing systems* 33, 21271–21284.
- He, K., H. Fan, Y. Wu, S. Xie, et R. Girshick (2020). Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 9729–9738.
- Henaff, O. (2020). Data-efficient image recognition with contrastive predictive coding. In *International conference on machine learning*, pp. 4182–4192. PMLR.
- Ismail-Fawaz, A., M. Devanne, S. Berretti, J. Weber, et G. Forestier (2023). Lite : Light inception with boosting techniques for time series classification. In *2023 IEEE 10th International Conference on Data Science and Advanced Analytics (DSAA)*, pp. 1–10. IEEE.
- Ismail Fawaz, H., B. Lucas, G. Forestier, C. Pelletier, D. F. Schmidt, J. Weber, G. I. Webb, L. Idoumghar, P.-A. Muller, et F. Petitjean (2020). Inceptiontime : Finding alexnet for time series classification. *Data Mining and Knowledge Discovery* 34(6), 1936–1962.
- Le-Khac, P. H., G. Healy, et A. F. Smeaton (2020). Contrastive representation learning : A framework and review. *Ieee Access* 8, 193907–193934.
- Lines, J., S. Taylor, et A. Bagnall (2016). Hive-cote : The hierarchical vote collective of transformation-based ensembles for time series classification. In *2016 IEEE 16th international conference on data mining (ICDM)*, pp. 1041–1046. IEEE.
- Middlehurst, M., J. Large, M. Flynn, J. Lines, A. Bostrom, et A. Bagnall (2021). Hive-cote 2.0 : a new meta ensemble for time series classification. *Machine Learning* 110(11), 3211–3243.
- Pelletier, C., G. I. Webb, et F. Petitjean (2019). Temporal convolutional neural network for the classification of satellite image time series. *Remote Sensing* 11(5), 523.
- Ruiz, A. P., M. Flynn, J. Large, M. Middlehurst, et A. Bagnall (2021). The great multivariate time series classification bake off : a review and experimental evaluation of recent algorithmic advances. *Data Mining and Knowledge Discovery* 35(2), 401–449.
- Rußwurm, M., C. Pelletier, M. Zollner, S. Lefèvre, et M. Körner (2019). Breizhcrops : A time series dataset for crop type mapping. *arXiv preprint arXiv :1905.11893*.
- Yue, Z., Y. Wang, J. Duan, T. Yang, C. Huang, Y. Tong, et B. Xu (2022). Ts2vec : Towards universal representation of time series. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Volume 36, pp. 8980–8987.

Annexe : paramètres et implémentations des méthodes

- Régression logistique : implémentation de `cuml` (1) avec les paramètres par défaut.
- Rocket : implémentation de `aeon` (2) de `MiniRocket` (3) qui s’est avéré plus performante que `ROCKET` (Dempster et al., 2020), `MultiRocket` (4) et `Hydra-MultiRocket` (5) sur notre ensemble de données.

Apprentissage auto-supervisé basé sur les segments

- InceptionTime (Ismail Fawaz et al., 2020) : implémentation de aeon (2) configuré avec batch size 512 et 100 epoch. Ces modifications ont permis d’accélérer l’apprentissage sans affecter l’accuracy de validation.
- TimeMAE (Cheng et al., 2023) : paramètres par défaut, à l’exception de vocab size 64, wave length 6, alpha 1, depth model 512, et batch size 512.
- SimCLR (Chen et He, 2021) : implémentation de Lightly (6) utilisant un encodeur ResNet pour séries temporelles (7) (implémentation de tsai (8)) avec une largeur $\times 4$. Optimisé avec SGD, learning rate 0,02, momentum 0,9, weight decay 5×10^{-4} , et batch size 256.

Références locales :

1. Raschka, S., J. Patterson, et C. Nolet (2020). Machine learning in python : Main developments and technology trends in data science, machine learning, and artificial intelligence. *Information* 11(4), 193
2. aeon - a scikit-learn compatible toolkit for time series (2023)
3. Dempster, A., D. F. Schmidt, et G. I. Webb (2021). Minirocket : A very fast (almost) deterministic transform for time series classification. In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, pp. 248–257.
4. Tan, C. W., A. Dempster, C. Bergmeir, et G. I. Webb (2022). Multirocket : multiple pooling operators and transformations for fast and effective time series classification. *Data Mining and Knowledge Discovery* 36(5), 1623–1646.
5. Dempster, A., D. F. Schmidt, et G. I. Webb (2023). Hydra : Competing convolutional kernels for fast and accurate time series classification. *Data Mining and Knowledge Discovery* 37(5), 1779–1805.
6. Susmelj, I., M. Heller, P. Wirth, J. Prescott, et M. E. et al. (2020). Lightly.
7. Wang, Z., W. Yan, et T. Oates (2017). Time series classification from scratch with deep neural networks : A strong baseline. In *2017 International joint conference on neural networks (IJCNN)*, pp. 1578–1585. IEEE.
8. Oguiza, I. (2023). tsai - a state-of-the-art deep learning library for time series and sequential data. Github

Summary

To take advantage of unlabelled data in time series classification we introduce a method for selecting positive pairs in contrastive self-supervised learning using pre-existing groups (i.e. segments) in the data. We experimentally validate our method on a new large-scale Sentinel-2 satellite image time series dataset for cropland classification.