

# 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 - Université Paris-Saclay  
antoine.cornuejols@agroparistech.fr

## 1 Introduction

Cet article est une traduction raccourcie de Saget et al. (2024). De vastes quantités d'images satellites sont capturées chaque jour, mais leur exploitation par apprentissage supervisé classique nécessite un étiquetage préalable coûteux et difficile à obtenir, laissant la majorité des données disponibles inexploitées. L'apprentissage auto-supervisé (Self-Supervised Learning, SSL) utilise les données non étiquetées pour apprendre de nouvelles représentations nécessitant moins de données étiquetées que les méthodes supervisées standard pour atteindre une même performance. En apprentissage SSL contrastif, les exemples positifs sont des points de données similaires, généralement créés par augmentation. Cependant, 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 se passer d'augmentations et adapter le SSL contrastif aux SITS.

## 2 Données et Méthode

**Données.** Notre jeu de données, inspiré par Rußwurm et al. (2020), comprend plus de 5,8 millions de parcelles agricoles étiquetées en France métropolitaine. Les labels proviennent du Registre Parcellaire Graphique 2022 (232 classes). Pour chaque parcelle, 100 séries temporelles Sentinel-2 sont extraites, chacune comportant 12 bandes radiométriques sur 60 pas de temps (du 1er février au 30 novembre 2022).

**Méthode.** Notre approche "Groups as Positive Pairs" (GaPP) définit comme exemples positifs toutes paires de séries temporelles de pixels provenant d'une même parcelle agricole. Aussi, plutôt que de soumettre des paires de séries temporelles individuelles à l'encodeur, nous lui soumettons des  $n$ -uplets de séries temporelles d'une même parcelle, puis les agrégeons avec une couche de moyennage ("Average Pooling", AvgP).