Towards Streaming Land Use Classification of Images with Temporal Distribution Shifts

Lorenzo Iovine[1], Giacomo Ziffer[1,2], Andrea Proia[3] and Emanuele Della Valle[1]

1- Politecnico di Milano - DEIB
Via Ponzio, 34/5, Milan - Italy

2- Motus ml
Viale Premuda, 20, Milan - Italy

3- Thales Alenia Space Italia
Via Saccomuro, 24, Rome - Italy

Abstract. In this study, we introduce a new pipeline that integrates Streaming Machine Learning (SML) models and the Momentum Contrastive Learning (MoCo) technique for the streaming classification of satellite images subject to temporal variations in distribution. We present preliminary results of an experimental campaign conducted on the Functional Map of the World-Time dataset, one of the first benchmarks specifically designed to address temporal distribution shifts in satellite imagery. The results demonstrate that the proposed pipeline enhances robustness and generalization over time, surpassing traditional strategies.

1 Introduction

Machine Learning (ML) has made significant progress in recent years, and the field of Streaming Machine Learning (SML) [1] is emerging as an area of growing interest due to its ability to continuously learn in dynamic environments. However, despite the advances in the field of SML, the application of these methodologies to image streams remains underexplored, representing a significant gap. Integrating SML into image stream processing could offer considerable advantages in timely learning from visual data.

Satellite images play a central role in critical applications such as environmental monitoring. The deployment of small satellite constellations with high revisit rates represents a growing trend in the space industry, enabling land-use monitoring at a finer temporal resolution. For instance, a temporal sequence of satellite images might show a road as a simple pathway surrounded by vegetation. In a subsequent image, the same road might appear flooded [2], but distinguishing it from a generic water body is only possible by considering both the spatial and temporal context provided by the sequence of images. However, this progress also introduces significant challenges, requiring advanced algorithms capable of addressing the complexities associated with the high dimensionality of the data, as well as temporal, spatial, and atmospheric variations.

To address the issue of multidimensionality, feature extraction techniques such as Momentum Contrastive Learning (MoCo) have emerged in recent years. These methods allow high-dimensional raw data to be transformed into more compact representations useful for tasks such as classification. The effectiveness of MoCo was recently demonstrated in the context of Continual Learning (CL) [3], where unsupervised pre-training with MoCo significantly improved the performance of continuous learning algorithms. A further challenge in learning from data streams resides in their temporal order [4], where changes in distribution result from the passage of time. These temporal variations in data distribution are formally defined as Temporal Distribution Shifts.

Our work analyzes a subset of satellite imagery from the Functional Map of the World [2] dataset, called FMoW-Time [5]. It explores whether the combination of SML and MoCo can improve the classification of satellite imagery in presence of shifts. To attack this challenge, we pose the following **research question**: *"Can SML models, combined with the MoCo technique, generalize effectively in learning from satellite image streams subject to temporal distribution shifts?"*. Notably, the main **contributions of this work** are: (i) an experimental analysis of the integration of MoCo and SML, exploring how their combination can optimize model performance in contexts characterized by shifts; (ii) a preliminary application of this methodology with satellite image streams, and comparison of the results with CL strategies.

The integration of embedding techniques such as MoCo into SML pipelines for image streams remains largely unexplored. The key innovation of this work lies in the systematic design of a pipeline that aligns MoCo embeddings with the challenges of SML, where models must learn from non-stationary data streams in real-time. We investigate how MoCo-derived representations impact the effectiveness of streaming classifiers and demonstrate that this integration potentially outperforms traditional approaches in handling temporal distribution shifts.

All data streams and algorithms used in this paper are available online[1] as a publicly available benchmark to help other researchers in their work to reproduce the results shown in this study or the development of new algorithms.

The rest of the paper is structured as follows. Section 2 delves into the state of the art. Section 3 describes the approach taken. Section 4 presents the results obtained. Finally, Section 5 analyzes the conclusions and future directions.

2 Related Work and State of the Art

A crucial aspect of **Streaming Machine Learning (SML)** is the ability to handle shifts (named concept drift [6]), i.e., changes in the underlying distribution of data over time. SML includes algorithms designed to learn continuously from streaming data. The Hoeffding tree (HT) [7] is an incremental learning algorithm based on decision trees. The Hoeffding Adaptive Tree [8] is an extension of HT designed to handle shifts. The Extremely Fast Decision Tree [9] is a variant of HT that has theoretical properties that ensure it converges faster. The Naive Bayes is a probabilistic classifier based on Bayes' theorem. The Softmax Regression is an extension of the Logistic Regression for multiclass classification.

[1] `https://github.com/lorenzoiovine99/FMoW-Time`

In SML, **Data Preparation** is a crucial step, as raw data need transformations for efficient model processing. The main tools for streaming data preparation include the **Gaussian Random Projector** [**10**], which reduces dimensionality in a lower-dimensional space using a matrix of random Gaussian values. Other tools widely used in this context are the **Noise Transformer** and the **Normalisation Transformer** to regularize and standardize the data.

Continual Learning (CL) [**11**] is a machine learning paradigm where models learn from data divided into multiple small experiences instead of a single large dataset. Several algorithms have been developed, and these include: Synaptic Intelligence, which preserves learned knowledge by weighting the most relevant parameters for past experiences; A-gem, which uses a subset of past data to adjust the learning gradient; and Fine-Tuning, which applies incremental updating to the model.

Embeddings are used to reduce data from high-dimensional space into lower-dimensional. **Momentum Contrastive Learning (MoCo)** [**12**] is an embedding technique that uses a query encoder for transforming the sample into a compact representation, and a momentum encoder to generate representations for the dictionary samples. The core of MoCo resides in its loss function, which maximizes the similarity between the query and the positive key while minimizing the one with the negative samples.

3 Proposed Approach

The proposed approach, as can be seen from Figure 1, consists of two parts: an embedding technique (MoCo) and a classifier (SML models). The pipeline was then enriched with pre-processing techniques that aimed to improve the performance of the learning process. **MoCo** was chosen as embedding method because of its efficiency and scalability. The implementation used is based on ResNet-18's model [13], replacing its final fully connected layer with a Multi-Layer Perceptron (MLP), which projects learned features into a lower-dimensional space. To ensure good generalization to real-world images, tests were carried out using the pre-trained weights of ResNet-18 on ImageNet1k-V1 [14].

It can be seen from the Figure 1 that our pipeline also includes a Streaming Data Preparation block. The first pre-processing technique used in this paper involves the Gaussian Random Projector. In addition, we used the Noise Transformer and the Normalisation Transformer, to prepare and scale the data.

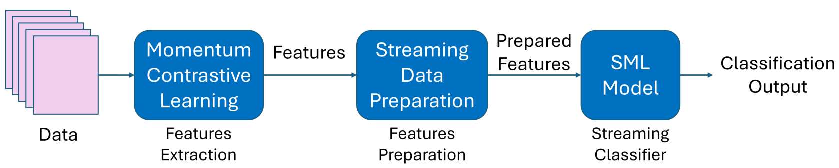

Fig. 1: Visualization of the proposed pipeline.

4 Experimental Evaluation

The following section is divided as follows: Section 4.1 presents the settings for the experimental evaluation; Section 4.2 discusses the results obtained with the SML models; finally, Section 4.3 describes the comparison with the CL strategies.

4.1 Experimental Settings

The dataset used for this experimental campaign is FMoW-Time [5], aimed at classifying the functional purpose of regions in a satellite image. It contains 141,696 RGB 224x224 images labeled into 62 categories of building or land use, including satellite images from 2002 to 2017. The experiments evaluate two metrics taken from Wild-Time [5]: the **In-Distribution (ID)** is the accuracy of the model tested on images with the same distribution as the training ones; the **Out-Of-Distribution (OOD)** is the accuracy of the model tested on images with a distribution different from that of the training. The experiments were carried out using two different evaluation methods: **Eval-Stream (EA)**, which uses 90% of each year's data to train and 10% to calculate the ID Accuracy; **Prequential Evaluation [15]**, in which the model is first tested and then trained on each sample from each year. For both methods, OOD Accuracy was computed using data from the six subsequent years, aligning with Wild-Time's setting. Results include: average ID and OOD accuracies over all years and the worst-performing years (ID and OOD Worst).

4.2 SML Models Results

In the pipeline shown in Figure 1, we have a fixed component, which concerns the feature extraction performed by MoCo, and a variable component, which involves the different SML models used. These models include Naive Bayes (NB), Hoeffding Tree (HT), Hoeffding Adaptive Tree (HAT), Extremely Fast Decision Tree (EFDT), and Softmax Regression (SR). The implementations of these models were taken from the CapyMOA[2] python library, except SR, which was wrapped from the River[3] library. In this subsection we show the results obtained

[2] `https://capymoa.org`
[3] `https://riverml.xyz`

SML Models	ID Accuracy	OOD Accuracy	ID Worst	OOD Worst
HT	19.55%	17.08%	17.66%	**14.99%**
HAT	19.48%	16.96%	17.38%	14.35%
NB	19.30%	17.04%	17.66%	**14.99%**
EFDT	19.62%	15.18%	**18.33%**	8.25%
SR	**26.20%**	**21.68%**	17.03%	14.67%

Table 1: Comparison of SML models and GRP using prequential evaluation. The combination of SR and GRP exhibits the best ID and OOD average accuracies.

using the prequential evaluation with these SML algorithms. We performed different tests using the Gaussian Random Projector (GRP), and we obtained the best results for each model, as shown in Table 1, using 1250 as the value of the hyperparameter *n_components* (dimensionality of the target projection space).

4.3 Results comparison with CL

In this subsection, we present a comparative analysis between the results obtained with our best-performing model and the best models reported in Wild-Time [5]. Those models include Synaptic Intelligence (SI), CORAL, Empirical Risk Minimization (ERM), A-gem, and Fine-Tuning (FT). To have a consistent comparison the tests were carried out using the Eval-Stream setting. From Table 2 and Figure 2 it is noteworthy that our model outperforms all state-of-the-art algorithms in ID Accuracy. The superior ID accuracy can be attributed to the fact that our pipeline enables a better adaptation to temporal shifts in the training data. MoCo embeddings allow our model to benefit from richer feature representations, while the streaming classifiers effectively adapt to the new data. Regarding OOD Accuracy, our model follows the same trend as the results reported in Wild-Time, lagging behind CL techniques. One main reason could be that our approach updates the model incrementally without revisiting past samples, limiting its ability to learn long-term dependencies across time.

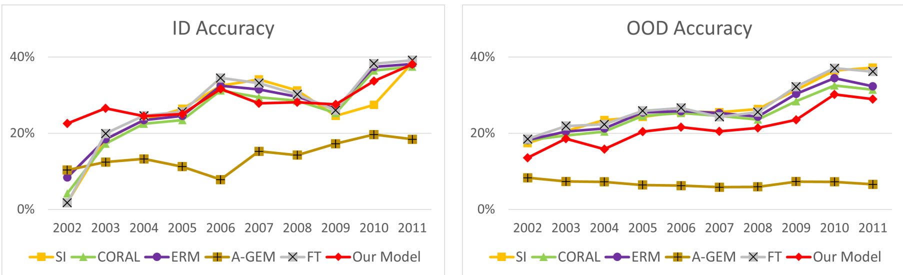

Fig. 2: Graphical comparison between the proposed approach and the best CL techniques tested in Wild-Time benchmark [5].

Models	ID Accuracy	OOD Accuracy	ID Worst	OOD Worst
SI	25.82%	26.81%	1.82%	17.42%
CORAL	25.57%	24.83%	4.23%	18.12%
ERM	27.02%	25.78%	8.42%	18.23%
A-GEM	13.98%	6.84%	7.82%	5.83%
FT	27.31%	**27.05%**	1.76%	**18.49%**
Our Model	**28.55%**	21.44%	**22.52%**	13.57%

Table 2: Comparison between the proposed approach and the best CL techniques tested in Wild-Time benchmark [5].

5 Conclusions and Future Works

We explored the application of SML models combined with MoCo to tackle the classification of a satellite image stream subject to temporal distribution shifts. The results show that the pipeline has achieved promising performance, demonstrating good generalization and robustness over time. This study can thus represent a significant step forward for the space industry, where the deployment of small satellite constellations with high revisit rates is rapidly growing. These advancements enable more frequent and detailed land-use monitoring, so effectively addressing temporal shifts is becoming increasingly important.

This study represents a preliminary investigation focused on a single dataset of satellite images, underscoring the need to extend the evaluation to additional datasets with comparable characteristics. Furthermore, the dataset used in this study is limited to RGB imagery. Expanding the pipeline to incorporate multispectral or hyperspectral data would be an intriguing avenue for future work, as it could enhance temporal change detection and model adaptability.

References

[1] H. M. Gomes, J. Read, A. Bifet, J. P. Barddal, and J. Gama. Machine learning for streaming data: state of the art, challenges, and opportunities. *KDD*, 21:6–22, 2019.

[2] Gordon Christie, Neil Fendley, James Wilson, and Ryan Mukherjee. Functional map of the world. In *Proceedings of the IEEE Conference on CVPR*, June 2018.

[3] Zhiqiu Lin, Jia Shi, Deepak Pathak, and Deva Ramanan. The clear benchmark: Continual learning on real-world imagery. In *NeurIPS 2021 Datasets and Benchmarks Track*.

[4] Giacomo Ziffer, Alessio Bernardo, Emanuele Della Valle, Vitor Cerqueira, and Albert Bifet. Towards time-evolving analytics: Online learning for time-dependent evolving data streams. *Data Science*, 6(1-2):1–16, 2023.

[5] Huaxiu Yao, Caroline Choi, Bochuan Cao, Yoonho Lee, Pang Wei W Koh, and Chelsea Finn. Wild-time: A benchmark of in-the-wild distribution shift over time. *Advances in Neural Information Processing Systems*, 35:10309–10324, 2022.

[6] Alexey Tsymbal. The problem of concept drift: Definitions and related work. 05 2004.

[7] A. Bifet and R. Gavalda. Learning from time-changing data with adaptive windowing. In *SIAM*, pages 443–448. SIAM, 2007.

[8] Albert Bifet and Ricard Gavalda. Adaptive learning from evolving data streams. In *8th International Symposium on IDA, 2009, Lyon, France.*, pages 249–260. Springer.

[9] Chaitanya Manapragada, Geoffrey I. Webb, and Mahsa Salehi. Extremely fast decision tree. In *Proceedings of the 24th ACM SIGKDD*, pages 1953–1962, New York, USA, 2018.

[10] Santosh S Vempala. *The random projection method.* American Mathematical Soc., 2005.

[11] Yen-Chang Hsu, Yen-Cheng Liu, and Zsolt Kira. Re-evaluating continual learning scenarios: A categorization and case for strong baselines. *CoRR*, abs/1810.12488, 2018.

[12] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick. Momentum contrast for unsupervised visual representation learning. In *CVPR*, June 2020.

[13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on CVPR*, June 2016.

[14] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on CVPR*, pages 248–255.

[15] Joao Gama, Raquel Sebastiao, and Pedro Pereira Rodrigues. On evaluating stream learning algorithms. *Machine learning*, 90:317–346, 2013.