Towards Explainable AI4EO: An Explainable Deep Learning Approach for Crop Type Mapping using Satellite Images Time Series

Adel Abbas, Michele Linardi, Etienne Vareille, Vassillis Christophides, Claudia Paris

TOWARDS EXPLAINABLE AI4EO: AN EXPLAINABLE DEEP LEARNING APPROACH FOR CROP TYPE MAPPING USING SATELLITE IMAGES TIME SERIES

Adel Abbas[1]*, Michele Linardi*[1]*, Etienne Vareille*[1]*, Vassillis Christophides*[3]*, Claudia Paris*[2]

[1]ETIS UMR-8051, CY Cergy Paris Université
[2]Department of Natural Resources, ITC, University of Twente, Enschede, The Netherlands
[3]ETIS UMR-8051, ENSEA Cergy

ABSTRACT

Deep Learning (DL) models are extremely effective for crop-type mapping. However, they generalize poorly when there is a temporal shift between the Satellite Image Time Series (SITS) acquired in the source domain (where the model is trained) and the target domain (never seen by the network). To address this challenge, this paper proposes an Explainable Artificial Intelligence (xAI) approach that leverages the interpretability of the inner workings of transformer encoders to automatically capture and mitigate the temporal shift between SITS acquired in different regions. The Positional Encoding (PE) output computed on the source SITS is used as a proxy to quantify the temporal shift with respect to the PE output obtained on the target SITS. This condition allows us to re-align the latter to the representation that the model natively adopts to discriminate crop types through a Dynamic Time Warping (DTW) approach. Compared to the baseline architecture, the proposed method increases the Overall Accuracy (OA) up to 8% on the TimeMatch benchmark dataset.

Index Terms— Transformers, AI4EO, Dynamic Time Warping (DTW), Crop Type Mapping, Sentinel-2 data.

1. INTRODUCTION

Large-scale crop type mapping is crucial for agricultural planning and management. Recently, Deep Learning (DL) models based on transformer encoders have proven their effectiveness in classifying long and dense Satellite Image Time Series (SITS) to generate accurate crop type maps [1]. Although these models have outperformed shallow machine learning techniques, they pose challenges in terms of explainability and confidence in the results, as their input-output mapping function is usually estimated without revealing any information about the internal structure of the model. For these reasons, recently, efforts have been made to produce Explainable Artificial Intelligence (xAI) approaches able to provide additional information to end users, thus improving their understanding of the DL model results [2]. Although xAI methods shed light on how DL models work, very little has been done to develop xAI approaches tailored to SITS [3] and to understand why these models are unable to handle domain shifting. Indeed, DL models can be highly effective in capturing and processing temporal patterns for the training region (source domain) while they typically generalize poorly to regions never seen by the network (target domain) [1]. This is mainly due to different climate conditions and management decisions, which determines different timing of phenological phases for the same crop type in different regions [4]. Several approaches have recently been proposed to focus explicitly on the temporal dimensions of SITS acquired in different areas for large-scale crop type mapping [4, 5]. However, they focus on the data distribution shift, by completely neglecting the role played by the considered DL model.

This paper presents a novel xAI approach that aims to improve the generalization capability of the existing object-based crop classifier based on Pixel Set Encoder (PSE) and Temporal Attention Encoder (TAE) widely used to perform crop type mapping [5]. Leveraging the interpretability of the model's inner workings, the proposed method automatically estimates the temporal shift of the SITS acquired in the source and the target domains by comparing their Positional Encoding (PE) outputs. To mitigate the detected temporal shift, the Dynamic Time Warping (DTW) technique is used to re-align the target PE output to the representation that the model natively adopts to discriminate crop types. By adopting this correction, the proposed approach outperforms the classification accuracy obtained by the standard network architecture without requiring any labeled data in the target domain.

2. PRELIMINARIES AND FORMULATION

This section summarizes the problem formulation while recalling the fundamentals of the considered DL architecture. Let $\mathbf{TS}^S = \{\mathbf{X}_t^S\}_1^b$ be the SITS acquired in the source domain (i.e, where the labeled data available are used to train the DL model), made up of b images, where $t \in [1, b]$ represents a generic time image acquisition. Let $\mathbf{TS}^T = \{\mathbf{X}_t^T\}_1^u$ be the SITS acquired in the target domain (where no labeled data are available), made up of u images. In a realistic scenario, it is reasonable to assume that $u \neq b$ since SITS acquired

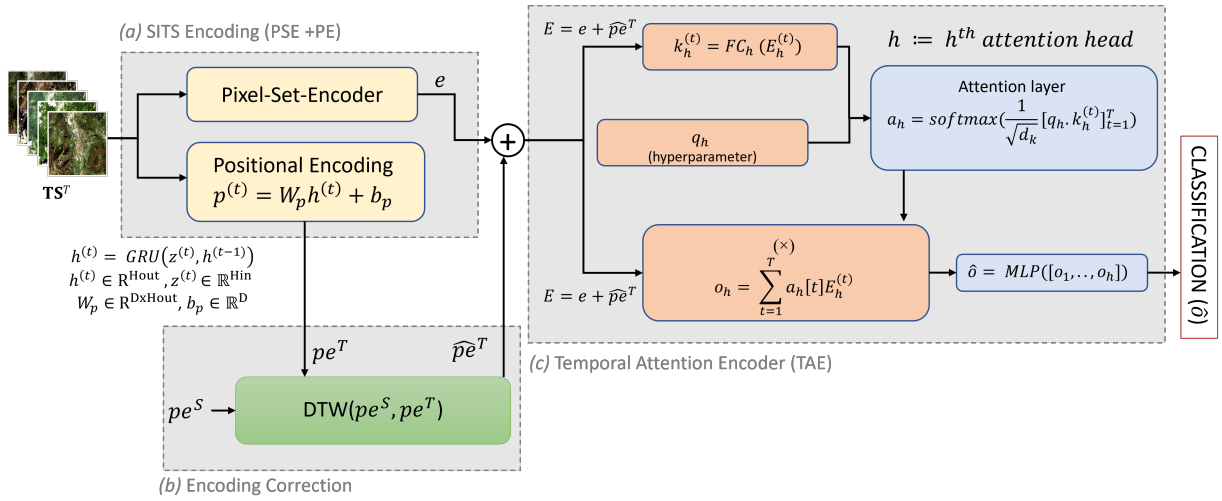

Fig. 1. Schematic representation of the proposed xAI approach based on the PSE+TAE model. The method is based on 3 main components: *(a)* SITS *Encoding*, (b) *Positional Encoding Correction*, and (c) *Temporal Attention Encoding*. First, the PSE and PE are applied to $\mathbf{TS}^T$. Then, the PE outputs generated on source and target SITS (i.e., pe^S and pe^T, respectively) are compared to re-align pe^T to the representation that the PSE+TAE model adopts to discriminate crop types. Finally, the TAE generates the crop type classification for $\mathbf{TS}^T$ considering the re-aligned target PE output, i.e., $\hat{pe}^T$.

over different regions are typically characterized by unequal lengths and different temporal sampling rates [6].

To generate the crop type maps in the source and target domains, we considered the object-based crop classifier PSE + TAE [5]. This network has been widely used for crop-type mapping due to its ability to process the spatial, temporal and spectral information provided by SITS. The standard network architecture consists of three different encoders: (1) PSE that extracts descriptors from the spectral distribution of each input observation, (2) PE that captures the periodic relationships between elements in the input sequence, and (3) TAE that handles the temporal information of the SITS. The TAE takes as input the combination of the PSE and PE outputs and computes the attention scores (a_h) for each head h (independent input slice) as the similarity (dot product) between all keys $k_h^{(t)}$ (for all time instant t) and the query q_h, re-scaled by a softmax layer. The k_h is learned by a fully connected layer, whereas the query is chosen to be a model hyperparameter. As suggested by Nyborg et al. [5], we use a learnable PE that considers vectorized positions of sine-cosine encoding $z(t)$ as input of a Gated Recurrent Units (GRU) layer that captures sequential dependencies of crop temporal evolution, providing better classification performance.

3. PROPOSED AI4EO APPROACH FOR LARGE SCALE CROP TYPE MAPPING

Figure 1 shows the schematic representation of the proposed xAI approach based on three main components: (a) SITS *Encoding*, (b) *Positional Encoding Correction*, and (c) *Temporal Attention Encoding*. Due to the variability in climate and cultivars, the same crop type can display different timing of its phenological phases in different areas, thus leading to a temporal shift between $\mathbf{TS}^S$ and $\mathbf{TS}^T$. Since the classification model learns how to discriminate crop types by optimizing the PE exclusively in the source domain, it is plausible to expect a performance decrease when such encoding shifts in the target domain. To address this challenge, in the *Positional Encoding Correction* step, we propose a two-fold approach, which first explains the temporal shift from a model perspective (i.e., source-target PE misalignment), and then, leverages this information to adjust such temporal shift on the SITS acquired in the target domain. Let pe^S and pe^T be the representations learned by the PE in the source and target domains, where $pe^S \in \mathbb{R}^{b \times D}$ and $pe^T \in \mathbb{R}^{u \times D}$. To preserve the model generalization capability, the temporal misalignment is estimated and mitigated through the DTW approach. This method has been widely used to compute the alignment (a.k.a warping path) that minimizes the distance between two series A and B, where $A, B \in \mathbb{R}^n$. The DTW requires computing the matrix $M \in \mathbb{R}^{n \times n}$, where each $M_{i,j}$ element corresponds to the square Euclidean distance $d(A_i, B_j)$. A warping path $P \in \mathbb{R}^n$ denotes the n elements in M such that:

$$min \sum_{i=1}^{n} P_t = DTW(A, B) \tag{1}$$

When computing DTW between the two multi-dimensional PE outputs, namely pe^S and pe^T, we still consider a unique warping P, which draws values from a matrix M', where $M'_{i,j} = [\sum_{m=1}^{b} dist(pe^S_{m,i}, pe^T_{m,j})]$ and $dist(pe^S_{m,i}, pe^T_{m,j})$ is the squared Euclidean distance between the pe^S, pe^T on the dimension m, at indexes i and j respectively [7]. In the considered implementation of the method, to represent the PE learned in the source, we compute the DTW Barycenter [8]

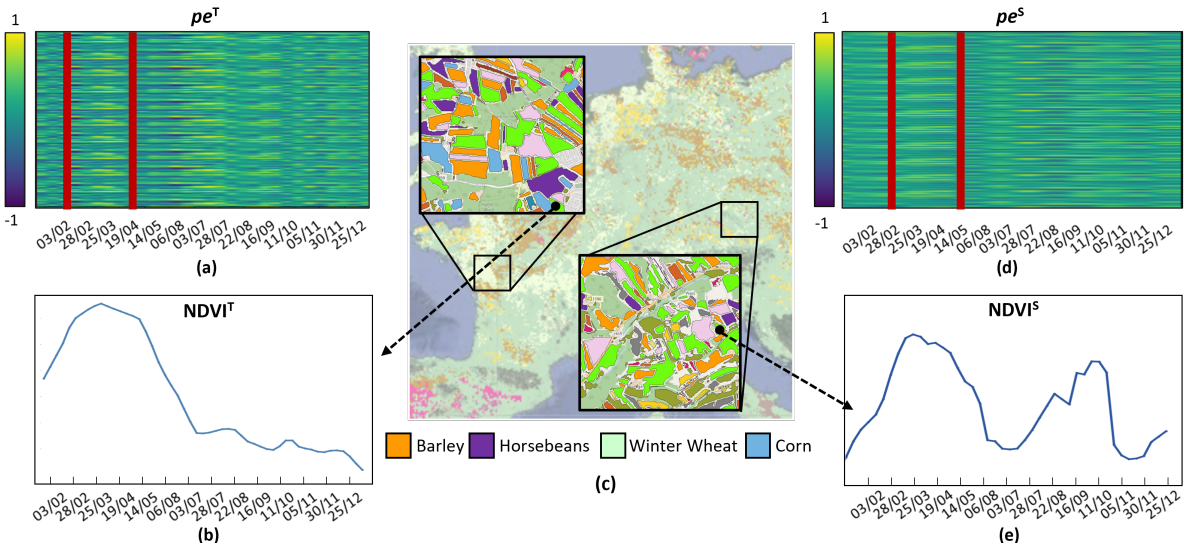

Fig. 2. Qualitative example of PE outputs generated in the source and target domains for the same crop type (i.e., winter wheat): (a) pe^T, (b) time series of NDVI on the $\mathbf{TS}^T$, (c) crops locations, (d) pe^S, and (e) time series of NDVI computed on the $\mathbf{TS}^S$.

of the SITS acquired in the source domain. At the inference stage, when the target SITS is classified by the trained DL model, we compute the DTW distance between the PE of the instance to classify and the source Barycenter. The resulting warping path is then considered to correct the positional encoding, i.e., $\hat{pe}^T$. To that extent, we replace the values of pe^T with those of pe^S along the indexes contained in the warping path in order to maximize the similarity of pe^T and pe^S. Figure 2 shows a qualitative example of PE outputs generated for the same crop type when ingesting $\mathbf{TS}^T$ (Figure 2a) and $\mathbf{TS}^S$ (Figure 2d). The temporal evolutions of the Normalized Difference Vegetation Index (NDVI) computed on $\mathbf{TS}^T$ and $\mathbf{TS}^S$ in depicted in Figure 2b and Figure 2e, respectively. Although both samples belong to the same crop type (i.e., winter wheat), the pattern shift visible in pe^T and pe^S (highlighted with the vertical bars) explains and confirms the temporal shifts observed on the two NDVIs. By correcting this misalignment, it is possible to increase the accuracy in the target domain.

Table 1. Number of available labeled samples. While the labeled data of Denmark (DK), Austria (AT) and southern France (FR1) were used to train the models, the data in mid-west France (FR2) were used for performance evaluation.

		Training Data			Test Data
	Crop Type	**DK**	**AT**	**FR1**	**FR2**
	Corn	7252	2790	4726	2450
	Horsebeans	306	237	421	143
	Meadow	11434	9002	9746	4528
	Spring barley	199	10219	169	65
	Unknown	2804	3169	10880	955
	Winter barley	3089	2671	2662	1366
	Winter rapeseed	763	2170	1868	230
	Winter triticale	1022	239	593	386
	Winter wheat	4747	7263	17959	1487

4. DATASET DESCRIPTION AND RESULTS

To test the performance of the proposed approach, we considered the public TimeMatch benchmark dataset [5], which has been released for evaluating cross-region crop type mapping models. For the sake of reproducibility, the source code of our solution and other relevant material are available online[1]. The SITS, acquired in 2017 over 4 regions in Europe, are made up of Sentinel-2 images having cloud cover lower than 80%. This leads to an unequal temporal sampling and variations between the SITS acquisition time. Table 1 shows the number of labeled samples per crop type available per country. In the considered experimental setup, we used the labeled data available in Austria (tile 33UVP), Denmark (tile 32VNH) and southern France (tile 31TJC) to train the DL models, while the labeled data available in mid-west France (tile 30TXT) was used only for performance evaluation.

Table 2 shows the F1 score (F1%) and the Overall Accuracy (OA%) obtained on the test set by applying the Standard and the Proposed PSE+TAE architecture when using the training data from: (i) Denmark (DK), (ii) Austria (AT), (iii) southern France (FR1), and (iv) the three countries (AT+DK+FR1). As expected the best results are achieved

[1] https://github.com/adelabbs/XAI4EO/

Table 2. F1% and OA% obtained on the test set located in mid-west France (Sentinel-2 tile 30TXT) with the Standard and the Proposed PSE+TAE architecture when training with the data from: (i) Denmark (DK), (ii) Austria (AT), (iii) southern France (FR1), and (iv) the three countries (AT+DK+FR1). The best results are highlighted in bold per crop type.

Crop Type	Standard PSE+TAE (F1%)				Proposed PSE+TAE			
	DK	AT	FR1	AT+DK+FR1	DK	AT	FR1	AT+DK+FR1
Corn	89.62	89.02	67.82	**95.63**	84.33	81.58	36.86	92.87
Horsebeans	0.00	21.21	53.24	**79.09**	0.00	52.78	7.62	78.95
Meadow	44.00	95.57	88.25	**97.22**	54.53	93.81	91.41	96.69
Spring barley	11.95	30.77	0.72	**41.23**	10.47	27.37	2.62	21.95
Unknown	21.55	48.99	38.90	**68.50**	23.14	43.31	37.83	62.91
Winter barley	18.36	43.33	73.18	54.79	24.97	78.75	60.12	**80.00**
Winter rapeseed	91.43	92.53	71.63	**98.76**	83.77	62.26	70.78	98.33
Winter triticale	0.00	29.95	22.90	40.00	0.35	15.57	8.93	**63.37**
Winter wheat	7.19	60.16	68.57	90.27	48.36	85.04	73.44	**94.41**
OA%	38.18	78.97	68.94	89.63	46.26	82.33	69.17	**90.42**
Macro avg. F1%	31.57	56.84	53.91	73.95	36.66	60.05	43.29	**76.60**
Weighted avg. F1%	41.34	79.84	73.87	89.98	53.15	81.78	70.51	**90.87**

when using the largest training dataset (AT+DK+FR1). In particular, the Proposed PSE+TAE is able to increase the OA%, macro average F1% and weighted F1% of almost 1%, 3% and 1% compared to the Standard PSE+TAE architecture. This is due to the fact that it was able to better handle the most critical classes such as "Winter barley" and "Winter triticale" (increasing their F1% of 26% and 23%) while achieving similar accuracy on the other crop types except for "Spring barley" (the F1% decreases of 20%). Moreover, the Proposed PSE+TAE is able to improve the classification results obtained also when considering a small training set located in only one country. Compared to the standard PSE+TAE architecture, the proposed method slightly increases the OA% when training the model with the French training data, while decreasing the macro average F1% and weighted F1% of 10% and 3%, respectively. However, when training the model with the Danish and the Austrian training data, it significantly increases the OA%, macro average F1% and weighted F1% of 8% and 2%, 5% and 4%, and 12% and 2%, respectively.

5. CONCLUSION

This paper presented a new xAI approach designed for large-scale crop mapping. The method, based on the PSE+TAE architecture, aims to improve its generalization capability by automatically comparing and re-aligning the PE outputs computed in the source and target domains without the need for any target labeled data. Preliminary results demonstrate the proposed approach outperformed the standard architecture for almost all experiments. As future developments, we aim to further study the effectiveness of the proposed approach in different regions. Moreover, we intend to further investigate the best strategy to mitigate the temporal shift detected between the PE outputs computed in the source and target domains.

6. REFERENCES

[1] Joachim Nyborg, Charlotte Pelletier, and Ira Assent, "Generalized classification of satellite image time series with thermal positional encoding," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 1392–1402.

[2] Caroline M Gevaert, "Explainable ai for earth observation: A review including societal and regulatory perspectives," *International Journal of Applied Earth Observation and Geoinformation*, vol. 112, pp. 102869, 2022.

[3] Ivica Obadic, Ribana Roscher, Dario Augusto Borges Oliveira, and Xiao Xiang Zhu, "Exploring self-attention for crop-type classification explainability," *arXiv preprint arXiv:2210.13167*, 2022.

[4] Ziqiao Wang, Hongyan Zhang, Wei He, and Liangpei Zhang, "Phenology alignment network: A novel framework for cross-regional time series crop classification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 2940–2949.

[5] Joachim Nyborg, Charlotte Pelletier, Sébastien Lefèvre, and Ira Assent, "Timematch: Unsupervised cross-region adaptation by temporal shift estimation," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 188, pp. 301–313, 2022.

[6] Giulio Weikmann, Claudia Paris, and Lorenzo Bruzzone, "Timesen2crop: A million labeled samples dataset of sentinel 2 image time series for crop-type classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 4699–4708, 2021.

[7] Mohammad Shokoohi-Yekta, Bing Hu, Hongxia Jin, Jun Wang, and Eamonn J. Keogh, "Generalizing DTW to the multi-dimensional case requires an adaptive approach," *Data Min. Knowl. Discov.*, 2017.

[8] François Petitjean, Alain Ketterlin, and Pierre Gançarski, "A global averaging method for dynamic time warping, with applications to clustering," *Pattern Recognit.*, vol. 44, no. 3, pp. 678–693, 2011.