



HAL
open science

Reuse out-of-year data to enhance land cover mapping via feature disentanglement and contrastive learning

Cassio F. Dantas, Raffaele Gaetano, Claudia Paris, Dino Ienco

► To cite this version:

Cassio F. Dantas, Raffaele Gaetano, Claudia Paris, Dino Ienco. Reuse out-of-year data to enhance land cover mapping via feature disentanglement and contrastive learning. 2024. hal-04547452

HAL Id: hal-04547452

<https://hal.science/hal-04547452>

Preprint submitted on 16 Apr 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Reuse out-of-year data to enhance land cover mapping via feature disentanglement and contrastive learning

Cássio F. Dantas^{a,c,*}, Raffaele Gaetano^{b,c}, Claudia Paris^d, Dino Ienco^{a,c}

^a*INRAE, UMR TETIS, Univ. Montpellier, Montpellier, France*

^b*CIRAD, UMR TETIS, Univ. Montpellier, Montpellier, France*

^c*INRIA, Montpellier, France*

^d*Department of Natural Resources, ITC, University of Twente, Enschede, Netherlands*

Abstract

Timely up-to-date land use/land cover (LULC) maps play a pivotal role in supporting agricultural territory management, environmental monitoring and facilitating well-informed and sustainable decision-making. Typically, when creating a land cover (LC) map, precise ground truth data is collected through time-consuming and expensive field campaigns. This data is then utilized in conjunction with satellite image time series (SITS) through advanced machine learning algorithms to get the final map. Unfortunately, each time this process is repeated (e.g., annually over a region to estimate agricultural production or potential biodiversity loss), new ground truth data must be collected, leading to the complete disregard of previously gathered reference data despite the substantial financial and time investment they have required. How to make value of historical data, from the same or similar study sites, to enhance the current LULC mapping process constitutes a significant challenge that could enable the financial and human-resource efforts invested

*Corresponding author

Email addresses: cassio.fraga-dantas@inrae.fr (Cássio F. Dantas), raffaele.gaetano@cirad.fr (Raffaele Gaetano), c.paris@utwente.nl (Claudia Paris), dino.ienco@inrae.fr (Dino Ienco)

in previous data campaigns to be valued again. Aiming to tackle this important challenge, we here propose a deep learning framework based on recent advances in domain adaptation and generalization to combine remote sensing and reference data coming from two different domains (e.g. historical data and fresh ones) to ameliorate the current LC mapping process. Our approach, namely *REFeD* (data Reuse with Effective Feature Disentanglement for land cover mapping), leverages a disentanglement strategy, based on contrastive learning, where invariant and specific per-domain features are derived to recover the intrinsic information related to the downstream LC mapping task and alleviate possible distribution shifts between domains. Additionally, *REFeD* is equipped with an effective supervision scheme where feature disentanglement is further enforced via multiple levels of supervision at different granularities. The experimental assessment over two study areas covering extremely diverse and contrasted landscapes, namely *Koumbia* (located in the West-Africa region, in Burkina Faso) and *Centre Val de Loire* (located in centre Europe, France), underlines the quality of our framework and the obtained findings demonstrate that out-of-year information coming from the same (or similar) study site, at different periods of time, can constitute a valuable additional source of information to enhance the LC mapping process.

Keywords: Satellite Image Time Series (SITS), Land Cover (LC) Mapping, Domain Adaptation, Contrastive Learning, Data-Centric Artificial Intelligence (AI).

1. Introduction

The unprecedented availability of Earth observation (EO) information regularly acquired through modern public and private EO Programmes and Missions

(e.g. ESA Copernicus, NASA Landsat and PlanetScope to cite a few) opens the opportunity to collect satellite image time series (SITS) over the same study area to characterize and study the underlying spatio-temporal dynamics. Such rich information has been demonstrated to be largely beneficial in a variety of different fields, such as ecology (Kolecka et al., 2018), agriculture (Ienco et al., 2017), forestry (Wulder et al., 2012), environmental monitoring (Huang et al., 2023) and facilitating well-informed and sustainable decision-making policies (Kavvada et al., 2020). SITS are commonly exploited in conjunction with in situ ground truth data, acquired via costly and labor-intensive field campaigns, in order to derive timely up-to-date land cover (LC) maps over a specific region (Gómez et al., 2016).

Typically, for the creation of LC maps over a region at a certain period of time, ground truth data are collected at a particular moment through expensive and time-demanding field campaigns. These data are then utilized in conjunction with SITS information through advanced machine learning algorithms (Zhong et al., 2019) to get the final LC map. While the access to high resolution EO data is no longer a major constraint, collecting up-to-date ground truth data constitutes a consumable (neither enduring nor lasting) effort. Once served its purpose, ground truth data will be disregarded losing any further relevance. Furthermore, when the process is repeated (e.g., estimate agricultural production or potential biodiversity loss for a new year for the same or a related study site), new field campaigns must be afforded again with, in general, no way to profit from previous efforts.

Recently, endeavors related to the systematic and effective exploitation of available high-quality data have been increasing in both machine learning and computer vision communities. To this end, research actions in this direction have

been proposed under the umbrella of data-centric Artificial Intelligence (AI) (Zha et al., 2023). Under this movement, the attention of researchers and practitioners is gradually shifting from advancing model design (model-centric AI) to enhancing the quality and quantity of the data (data-centric AI).

Considering geospatial and EO data, the data-centric AI perspective is even more important since it can steer the community towards developing methodologies to provide further improvements related to the generalization ability with impact on real-world relevant problems and applications (Roscher et al., 2023). Nevertheless, the two perspectives (model-centric and data-centric AI) play a complementary role in the larger machine learning deployment cycle since standard approaches still struggle to manage and exploit valuable data coming from different and heterogeneous distributions like, for instance, in the case of combining historical and up-to-date reference data for the downstream task of LC mapping (Gao et al., 2023) where distribution shifts can be related to the different environmental and/or climatic factors that determine the acquisition conditions.

When data from different distributions are combined together under the same learning framework, challenges related to distribution shifts can impede the effective training of machine learning models (Ben-David et al., 2010). To cope with this issue, the most widely studied setting is Unsupervised Domain Adaptation (UDA) (Wilson and Cook, 2020) where the main goal is to learn a model over a labelled source domain and transfer it to an unlabelled target one. Recent UDA advances predominantly concentrate on deriving features that are domain-invariant. This is achieved through either aligning domains via data transformation (Sun and Saenko, 2016) or employing adversarial training (Long et al., 2018), aiming to minimize the distribution gap between the source and target domains. In the

context of remote sensing, early research have focused on devising UDA strategies tailored for high spatial resolution imagery (Tuia et al., 2016). Only recently, strategies have begun to emerge to analyse SITS data for spatial (Nyborg et al., 2022) and temporal transfer tasks (Capliez et al., 2023).

Despite the efforts invested in designing and implementing UDA techniques, these approaches typically assume the complete absence of reference data for the target domain. However, for the LC mapping task, it is often reasonable to assume access to a certain amount of ground truth target data. Moreover, the success of UDA depends largely on the discrepancy between the source and target distributions, making these methods susceptible to potential pitfalls and limited generalizability. In scenarios where a limited amount of labeled data is available for the target domain, the paradigm shifts to Semi-Supervised Domain Adaptation (SSDA). Current SSDA approaches generally aim to align the target data with the labeled source data with feature space mapping and self-training assignments using pseudo-labels (Yu and Lin, 2023). Although the SSDA setting holds potential for various real-world problems, it remains largely unexplored when dealing with SITS data (Lucas et al., 2023) in the context of LC mapping. Another related setting is Domain Generalization (DG) (Jo and Yoon, 2023; Chen et al., 2023) wherein the objective is to learn classification models over a set of diverse labelled source domains that can generalize over new unseen target data. To the best of our literature survey, no DG frameworks have been proposed for EO data analysis. While DG and domain adaptation settings are closely related, DG operates inductively, assuming that target data are not available during the training stage. Conversely, in our problem setting, where the goal is to exploit together both historical and recent EO data along with reference data to enhance LC map-

ping, target data and their associated labels are already accessible.

In addressing the significant challenge outlined above, we present a novel framework, namely *REFeD*, rooted on recent advances in the field of domain adaptation/generalization. *REFeD* adopts a model-centric AI perspective, aiming to fulfill a data-centric AI objective related to the effective exploitation of EO and reference data coming from two different domains (e.g. historical data and recent ones) with the aim to give value again to historical and/or overlooked reference data, and enhance the accuracy of the recent LC mapping result.

More precisely, *REFeD* is a deep learning framework built upon a pseudo-siamese network with unshared parameters that, given a sample, extracts simultaneously domain-invariant and domain-specific features, using the former to make the final decision. The objective is to disentangle useful information for the downstream LC mapping task while isolating and discarding domain specific features that can hinder the learning process in the presence of data belonging to multiple domains with unaligned data distributions. The disentanglement process is achieved by shaping a representation manifold, via contrastive learning, that jointly structures both domain-invariant and domain-specific features. In addition, *REFeD* integrates an effective supervision strategy (Mohammadi et al., 2023) that further enforces the disentanglement process via multiple levels of supervision at different granularities.

Extensive experimental evaluations are carried out to assess the behavior of *REFeD* considering both baseline and domain adaptation/generalization approaches. To assess the behaviour of our method we perform both quantitative and qualitative evaluations considering two study sites covering extremely diverse and contrasted landscapes, namely *Koumbia* (located in the West-Africa region, in Burk-

ina Faso) and *Centre Val de Loire* (located in centre Europe, France). For the former study site we consider a LC mapping task where data covering exactly the same geographical area are available for two different years (2020 and 2021) while, for the latter study site, we consider a crop type mapping task where data covering two closely related areas are available for the agricultural seasons 2018 and 2021.

This manuscript is organized as follows: Section 2 introduces the proposed framework based on feature disentanglement and contrastive learning to enhance LC mapping combining multiple reference data. Study sites and the associated information are described in Section 3. The experimental evaluation and the related findings are reported and discussed in Section 4, while Section 5 draws the conclusions of this paper.

2. Proposed Deep Learning Framework

2.1. Problem Formulation and Notations

The proposed deep learning framework aims at improving the accuracy of LC mapping results obtained on recently acquired satellite data, i.e., target domain, by using pre-existing reference data coming from a different yet correlated domain, i.e., source domain. Typically, the source domain could consist of some readily-available historical or out-of-year data, for instance. Differences in climate, weather and other environmental conditions can lead to non-negligible distribution shifts within SITS data from the different domains. These shifts may prevent the full exploitation of the source data as a naive direct enrichment of the target data in a classic supervised learning setup (Bruzzone and Marconcini, 2009). Moreover, in case the source data is more abundant, the learned classifier

may likely be biased toward the source domain. Differently from the literature, the ultimate goal of REFeD is to maximize the classification performance in the target domain while taking full advantage of all the reference data available.

In this work, we suppose we are given a set of N_t labeled samples from a target domain $\mathcal{D}^t = \{(x_i^t, y_i^t)\}_{i=1}^{N_t}$, for which we want to train a classifier. Moreover, we dispose of additional labelled data (say N_s samples) from a source domain $\mathcal{D}^s = \{(x_i^s, y_i^s)\}_{i=1}^{N_s}$ that we aim to exploit in order to improve the performance of our classifier on the target domain.

In our case, each sample $x_i \in \mathbb{R}^{T \times C}$ is the content of a pixel’s C spectral bands from a SITS defined over T timestamps. In the considered experimental setup, we assume that the source and target SITS have the same number of T timestamps. The corresponding label $y_i \in \{1, \dots, K\}$ is given by one of K existing classes, shared between source and target domains — i.e., a closed-set scenario (Kundu et al., 2020). Depending on the considered classification task, the classes could be, for instance, different crop and/or LC types. Let us also define as $y'_i \in \{s, t\}$ the binary label associated to all the available labeled samples $\{(x_i, y'_i)\}_{i=1}^{N_s+N_t}$, which specify from which domain each spectral samples x_i belongs, i.e., \mathcal{D}^t or \mathcal{D}^s .

2.2. REFeD: Overview

Figure 1 shows an overview of the proposed deep learning framework, by depicting the data sources needed during the training and inference stages. In the first stage, the supervised classifier is trained using the reference data from both \mathcal{D}^s and \mathcal{D}^t . To take full advantage of reference data coming from distinct domains, we propose to disentangle the information carried by the labeled input data into two parts: 1) domain-specific information, and 2) domain-invariant information (i.e., useful discriminative information for the subsequent classification task). The

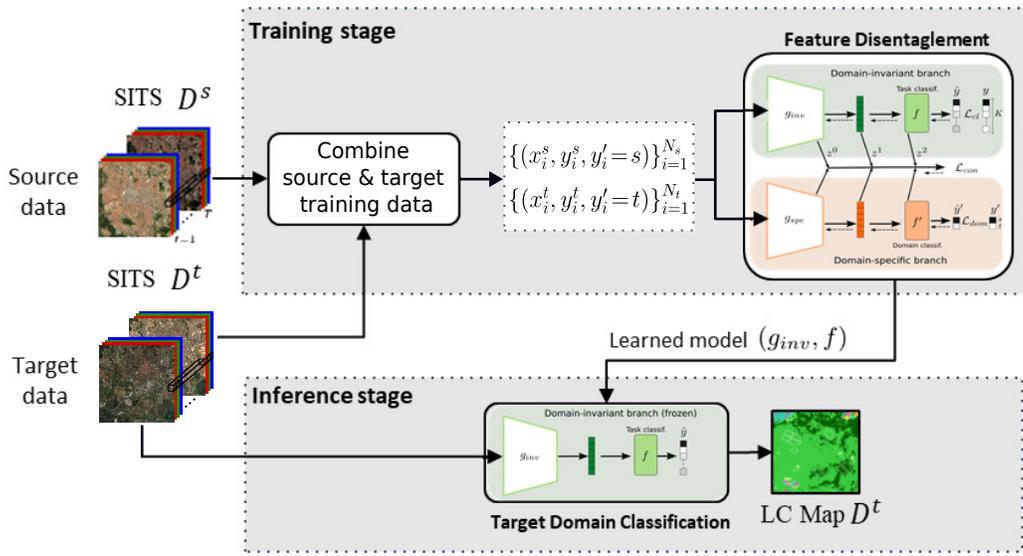


Figure 1: Overview of the proposed framework. Training and inference stages are distinguished: while the former is performed on data coming from both domains, the latter is done exclusively on target data and uses only the domain-invariant branch (g_{inv}, f) of the learned model.

former is closely related to the domain to which the data belong, thus potentially hindering the learning model’s ability to generalize. The second contains semantic information associated with the underlying classes, thus usable knowledge that can be exploited later for the classification process.

To this end, we take inspiration from current literature in domain generalization (Chen et al., 2023), even though in their setting the target domain is considered to be unseen, i.e., they consider an inductive setting while in our case we have a transductive setup. In particular, in the training stage we leverage two branches with separate encoders to generate the feature vectors, i.e., g_{spe} and g_{inv} . Dedicated losses, \mathcal{L}_{cl} , \mathcal{L}_{dom} and \mathcal{L}_{con} (detailed in section 2.3) are employed to effectively disentangle domain-specific from domain-invariant information in each of the two obtained embeddings, by training a domain classifier $f'(\cdot)$ and task classifier $f(\cdot)$, respectively. This condition allows us to benefit from the labeled samples available in all domains, taking into account the domain to which each labeled sample belongs.

To maximize the classification results obtained in the target domain, in the inference stage, the domain-specific encoder g_{spe} is discarded and only the domain-invariant encoder g_{inv} is considered. In particular, we generate the LC map of the target domain using g_{inv} for generating the feature representation to be classified along with the task classifier $f(\cdot)$ trained in the previous stage on the whole set of reference data. In the following, details are given.

2.3. Feature disentanglement

Figure 2 depicts the pseudo-siamese network with unshared parameters used for feature disentanglement. The two encoders, denoted g_{spe} and $g_{inv} : \mathbb{R}^{T \times C} \rightarrow \mathbb{R}^D$ for domain-specific and domain-invariant, respectively, share the same ar-

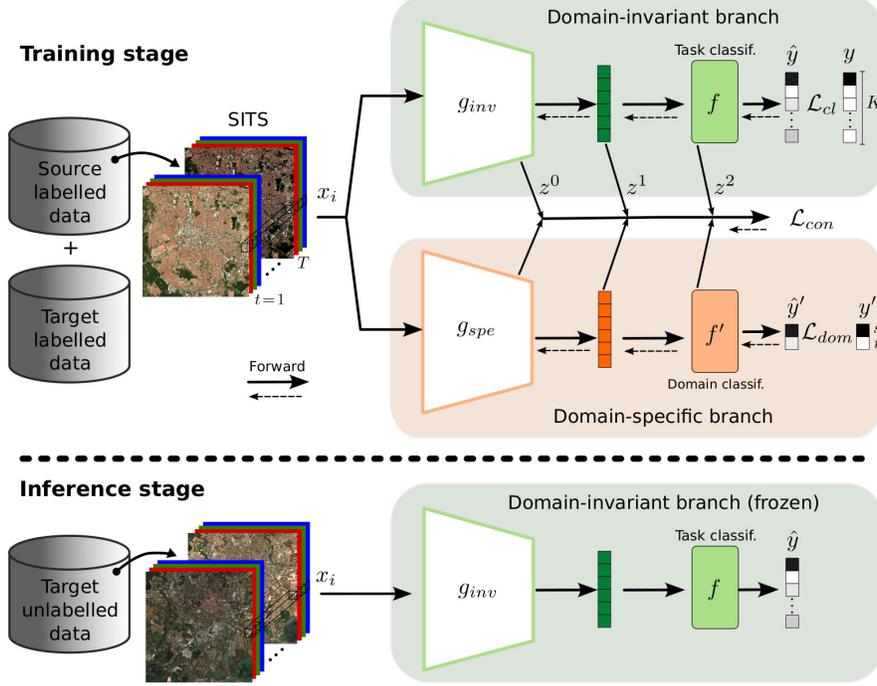


Figure 2: Architecture of the proposed pseudo-siamese network used in the training stage and composed of two independent branches which disentangle the domain-invariant information (top branch) from domain-specific information (bottom branch). Class (\mathcal{L}_{cl}) and domain (\mathcal{L}_{dom}) discrimination losses used respectively on the top and bottom branches, while a multi-level contrastive loss (\mathcal{L}_{con}) is used to intermediate features at different depths from both branches. At inference time, only the domain-invariant encoder is used for classifying the target domain.

architecture but are learned independently with unshared weights via different loss functions.

In the domain-invariant branch, a task classifier $f(\cdot)$ is applied to the domain-invariant features extracted by g_{inv} . In the domain-specific branch, the domain-specific features obtained via g_{spe} are fed to a domain classifier $f'(\cdot)$ which encourages the domain-discriminant information to be channeled to this branch.

Domain classifier. The domain classifier aims to accurately predict domain labels $y'_i \in \{s, t\}$, i.e., determine if each sample x_i belongs either to the source or the target domain, by minimizing a cross-entropy loss ℓ_{ce} as follows:

$$\mathcal{L}_{dom} = \frac{1}{N_s + N_t} \sum_{i=1}^{N_s+N_t} \ell_{ce}(f' \circ g_{spe}(x_i), y'_i) \quad (1)$$

Task classifier. In its turn, the task classifier $f : \mathbb{R}^D \rightarrow \{1, \dots, K\}$ maps the domain-invariant features onto one of the K classes of interest guided by the following cross-entropy loss:

$$\mathcal{L}_{cl} = \frac{1}{N_s + N_t} \sum_{i=1}^{N_s+N_t} \ell_{ce}(f \circ g_{inv}(x_i), y_i) \quad (2)$$

Contrastive learning. To further decouple of the two separate branches, we employ contrastive loss which has shown promising results for feature disentanglement in previous works (Chen et al., 2023). In our case, since we have access to both the class labels and domain labels, we can employ the supervised version of the InfoNCE contrastive loss proposed by Khosla et al. (2020), where the positive pairs are given by all samples sharing the same label. However, here, the application is not straightforward since we have two separate label spaces (class and domain labels). To address this issue, we adopt a mixed label space \mathcal{Y}_{mix} composed of $3K$ classes, where the domain-invariant features are mapped onto the K first labels while the last $2K$ are reserved to the domain-specific features – more specifically, K for the source domain and K for target domain. This leads to $\mathcal{Y}_{mix} = \{1, \dots, K, s1, \dots, sK, t1, \dots, tK\}$.¹

Denoting z the extracted embeddings (features) $g_{inv}(x)$ and $g_{spe}(x)$, we consider an augmented batch I of size $2B$ containing both $g_{inv}(x_i)$ and $g_{spe}(x_i)$ fea-

¹More generally, we define $\mathcal{Y}_{mix} = \mathcal{Y} \cup (\mathcal{Y} \times \mathcal{Y}')$ with a total of $|\mathcal{Y}_{mix}| = (|\mathcal{Y}'|+1)|\mathcal{Y}|$ classes.

tures for each $i \in \{1, \dots, B\}$ in the original batch. The resulting supervised contrastive loss is defined as follows:

$$\mathcal{L}_{con} = - \sum_{i \in I} \frac{1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(z_i \cdot z_p / \tau)}{\sum_{a \in I \setminus \{i\}} \exp(z_i \cdot z_a) / \tau} \quad (3)$$

where $P(i) := \{p \in I \setminus \{i\} : y_p = y_i\}$ with cardinality $|P(i)|$ is the set of *positive* examples w.r.t. the current *anchor* $i \in I := \{1, \dots, 2B\}$ and $\tau \in \mathbb{R}^+$ is a scalar temperature parameter. Therefore, the goal of this loss is to push together, in the feature space, the embeddings corresponding to the same category (positive examples) while repelling them from the rest (negative examples). The positive examples here correspond to those that share simultaneously: 1) the same class, among the K existing ones and 2) same domain type, among three options: source or target (for domain-specific features), or domain-invariant.

2.4. Multi-level supervision

To further enforce the feature disentanglement, we propose to perform contrastive learning not only at the level of the encoder’s output, but at multiple depths within the network architecture.

Then, the loss function described in (3) is actually also applied to intermediate features at different depths of the network. For that matter, we denote \mathcal{L}_{con}^l the contrastive loss (3) applied to the intermediate features z^l at depth l , as depicted in Figure 2. Specifically, in our case, we use three levels of supervision with: $l = 0$ for the encoder’s last internal layer; $l = 1$ for the encoder’s output features; $l = 2$ for the output of the classifier’s first fully-connected layer. Note that \mathcal{L}_{con}^l applies exclusively to features at depth l , which share the same space dimension, and, thus, features at different depths are never mixed together.

2.5. Model summary and training

The resulting loss is given by:

$$\mathcal{L} = \mathcal{L}_{ce} + \mathcal{L}_{dom} + \sum_l \mathcal{L}_{con}^l$$

Empirically, we observed that weighting the different losses did not have a conclusive impact on the final model performance. For this reason, we used a simple unweighted sum of the three terms as described above.

The proposed approach is completely agnostic to the encoder (g_{inv} , g_{spe}) and classifier modules architecture, but in this particular work we leverage the Temporal Convolutional Neural Network architecture (TempCNN) proposed by Pelletier et al. (2019) for pixel-based SITS classification tasks. The encoder part consists on three 1D-convolutional layers with 64 channels each. The classifier is composed of a fully-connected layer with 256 hidden units, batch normalisation and ReLU activation, followed by a linear output layer with Softmax activation.

3. Satellite Data and Ground Truth

The first study site covers an area around the town of *Koumbia*, in the Province of Tuy, *Hauts-Bassins* region, in the south-west of Burkina Faso. This area has a surface of about 2 338 km^2 , and is situated in the sub-humid sudanian zone. The surface is covered mainly by natural savannah (herbaceous and shrubby) and forests, interleaved with a large portion of land (around 35%) used for rainfed agricultural production (mostly smallholder farming). The main crops are cereals (maize, sorghum and millet) and cotton, followed by oleaginous and leguminous. Several temporary watercourses constitute the hydrographic network around the city of Koumbia.

Figure 3 presents the study site with the 2020 reference data (ground truth) superposed on a Sentinel-2 image. A more detailed view corresponding to the red box in the overview is also depicted on the bottom right of the figure. A specific analysis of the ground truth is provided in Section 3.2.

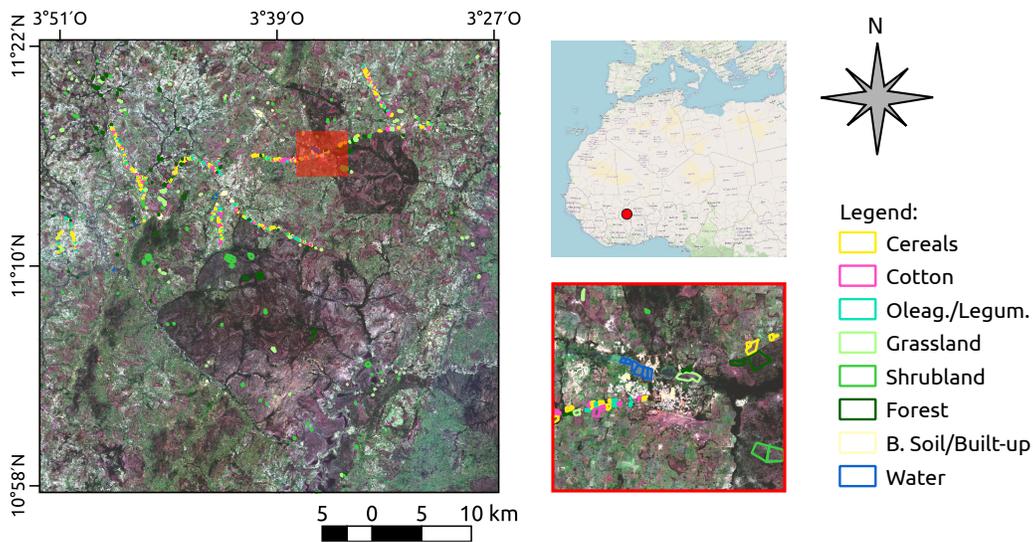


Figure 3: View and location of Koumbia study site. The ground truth data coming from the 2020 year is superposed to a Sentinel-2 image covering the whole area. In the red box (bottom right) a more detailed view of the study site is depicted.

The second study site covers two areas in the *Centre Val de Loire* region located in the center of France. This region of France is characterized by an intensive agricultural activity with agricultural surfaces representing around 70% of the whole region with cereals and oleaginous as major crops. The two areas have a cumulative surface of about 840 km^2 .

Figure 4 presents the two areas related to the *Centre Val de Loire* study site depicting reference (ground truth) data for year 2018 and 2021 superposed on a Sentinel-2 image. On the right of the figure, a detail for each of the areas is

proposed, in red for 2018 and in blue for 2021. A specific analysis of the ground truth is provided in the Section 3.2.

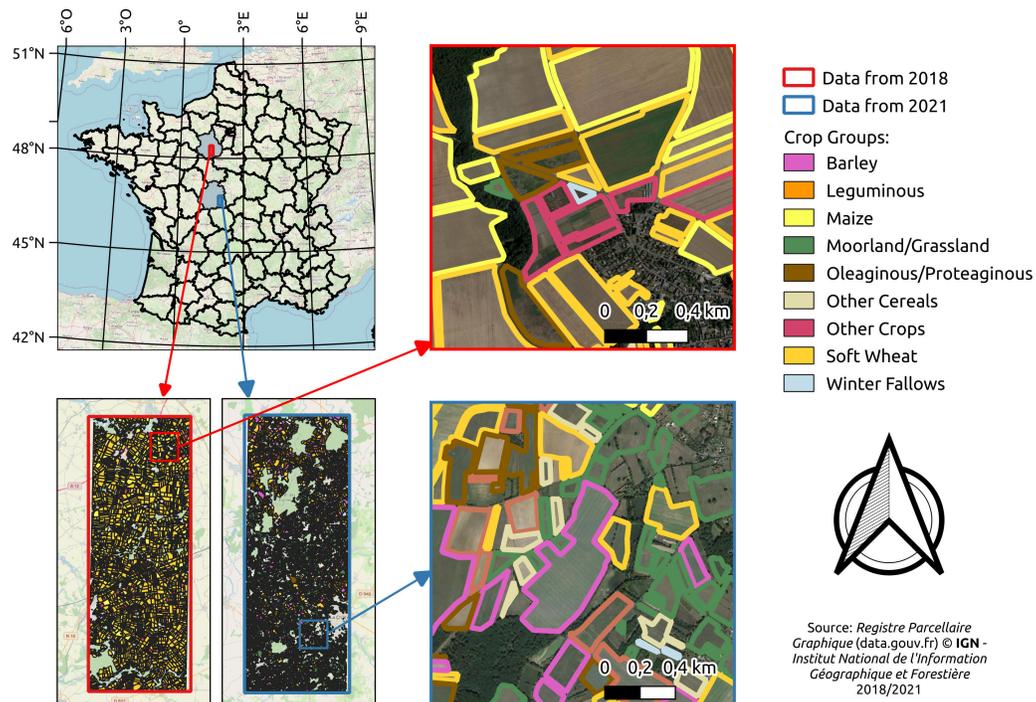


Figure 4: View and location of the *Centre Val de Loire* study site. The ground truth data coming from the 2018 and 2021 year area superposed to a Sentinel-2 image. On the right, a detail for each of the areas is proposed, in red for 2018 and in blue for 2021.

3.1. Satellite Image Time Series

For each study area and each considered year, we collected satellite image time series of Sentinel-2 imagery spanning an entire year via the Microsoft Planetary Computer platform² that allows to access level-2A Sentinel-2 products. We consider all bands at 10 m and 20 m of spatial resolution for a total of 10 bands

²<https://planetarycomputer.microsoft.com/>

per image. We have conducted resampling of the SWIR 20 m bands to 10 m resolution, as well as image time series gap filling of cloudy pixels using multi-temporal linear interpolation as explained in Inglada et al. (2015) and gap-filled images were generated at a regular 10-day frequency resulting in a sequence of 72 images for each study area and year.

3.2. Ground truth data

For the *Koumbia* study site, the ground truth data for 2020 and 2021 has been derived from a large agricultural LULC data set available online (Jolivot et al., 2021), mainly consisting of field data collected by local experts on several sites all over the tropics. For the *Koumbia* study site, the field surveys were conducted around the growing peak of the cropping season. The ground truth data cover the exact same surface for the two reference years.

For the *Centre Val de Loire* study site, the ground truth data for the first area, related to 2018, was obtained through the EuroCrop dataset (Schneider et al., 2023) while the ground truth data for the second area, related to 2021, are gathered from the RPG (*Registre Parcellaire Graphique*), the French land parcel identification system. The data covers only agricultural areas in order to set up a crop type mapping task. This second dataset covers a problem that implies both spatial and temporal transfer at the same time.

For both study sites, ground truth has been assembled in a Geographic Information System (GIS) vector file, containing a collection of polygons, each attributed with a LC or crop type category based on information reported in the original database – see Tables 1 and 2 for statistics about ground truth data distribution.

Table 1: Ground truth statistics for year 2020 and 2021 on the *Koumbia* study site.

Class Name	Class ID.	2020		2021	
		# Polygons	# Pixels	# Polygons	# Pixels
Cereals	1	230	9731	268	11435
Cotton	2	139	6971	121	6575
Oleaginous / Leguminous	3	281	7950	263	7316
Grassland	4	122	12998	113	11100
Shrubland	5	83	22546	90	24324
Forest	6	82	17435	82	16984
Bare soil / Built-up	7	51	1125	51	1022
Water	8	10	1205	10	1205
	Total	998	79961	998	79961

4. Experiments

In this section, we report and discuss the experimental evaluation carried out on the study sites presented in Section 3. Our objective is to evaluate the performance of *REFeD* across various dimensions. First, we undertake a quantitative assessment comparing the performance of *REFeD* against baselines and competing approaches. Second, we conduct a qualitative examination of the land cover maps generated by *REFeD*. Lastly, we inspect the internal representations learned by our model, visually comparing them to those of some of the top-performing competitors.

Table 2: Ground truth statistics for year 2018 and 2021 on the *Centre Val de Loire* study site.

Class Name	Class ID.	2018		2021	
		# Polygons	# Pixels	# Polygons	# Pixels
Soft wheat	1	1048	9388	1341	9268
Maize	2	198	9442	264	9437
Barley	3	674	9345	527	9222
Other cereals	4	421	9288	360	9473
Oleaginous / Proteaginous	5	842	9222	775	9168
Winter Fallows	6	413	8390	611	7396
Leguminous	7	5	3418	12	4753
Fodder	8	255	9350	105	7422
Meadow	9	4132	8254	127	6030
Other crops	10	184	8183	355	9681
	Total	8172	84280	12649	81850

4.1. Competing methods

With the goal to assess the performance of *REFeD* w.r.t. baselines and strategies coming from SSSA and DG literature, we consider:

- *Only Source*: this strategy trains a model only considering source data and, then, the obtained classifier is directly deployed on the target data. The main purpose of this method is to have an empirical estimate about the distribution shift between source and target domains. We implement this baseline considering both Random Forest and TempCNN (Pelletier et al., 2019) as classifiers.

- *Only Target*: this strategy trains a model only considering the target labelled data and, then, the obtained classifier is employed to classify the remaining target samples. The main purpose of this method is to provide the reference performances without the use of historical or out-of-year data. We implement this baseline considering both Random Forest and TempCNN (Pelletier et al., 2019) as classifiers.
- *Source+Target*: this strategy trains a model over both source and target labelled data. Then, the obtained classifier is employed to classify the remaining target samples. Here, the data coming from different distributions are mixed together constituting a multi-domain training dataset. We implement this baseline considering both Random Forest and TempCNN (Pelletier et al., 2019) as classifiers.
- *Fine Tuning*: this strategy trains a model over the labelled source domain and, then, the resulting model is fine tuned on target labelled samples. This is an alternative way to combine both source and target data. We implement this baseline considering only the TempCNN (Pelletier et al., 2019) approach.
- *Sourcerer* (Lucas et al., 2023): this recent SSDA approach has been proposed to cope with the analysis of SITS data for the downstream task of LULC mapping. Sourcerer is a bayesian-inspired, deep learning-based framework, that internally exploits the TempCNN model as backbone, similarly to *REFeD*. The technique leverages a deep learning model trained on a source domain and then fine-tunes the model on the available target domain via a regularizing term that automatically adjusts the degree to which the model

weights are modified to fit the target data.

- *POEM* (Jo and Yoon, 2023): this recent DG approach learns domain-invariant and domain-specific representations via a pseudo-siamese network with un-shared weights and it enforces polarization via orthogonality constraints. This approach was primarily introduced for image classification. In order to transfer it on SITS data, also for this competitor we employ the TempCNN model as a backbone.

4.2. Experimental settings

For all the competing approaches, labelled source data are entirely employed while, for the target domain, data are split into three parts: training, validation and test sets following a proportion of 50%, 20% and 30% of the original target data set, respectively. Regarding the *Only Source* baseline, the model is trained considering only the labelled source data. Furthermore, with the aim to avoid possible spatial bias in the evaluation procedure (Karasiak et al., 2021), we impose that all the pixels belonging to the same object will be exclusively associated to one of the data partitions (training, validation or test). The splitting procedure is repeated five times and the average results are reported.

Concerning the evaluation tasks, according to the data presented in Section 3, we set up two transfer tasks per benchmark. Each transfer task is denoted as $(\mathcal{D}_s + \mathcal{D}_t \rightarrow \mathcal{D}_t)$ where the right arrow indicates the transfer direction from the combined source/target labelled training dataset $(\mathcal{D}_s + \mathcal{D}_t)$ to the test target (\mathcal{D}_t) dataset. For the *Koumbia* study site, we consider as transfer tasks $(2020 + 2021 \rightarrow 2021)$ and $(2021 + 2020 \rightarrow 2020)$ and for the *Centre Val de Loire* study site, we consider the transfer tasks $(2018 + 2021 \rightarrow 2021)$ and $(2021 + 2018 \rightarrow 2018)$.

The values of the SITS benchmarks were scaled per year and per band considering the 2nd and 98th percentile of the data distribution as minimum and maximum values. The assessment of the model performances was done considering the following metrics: *Weighted F1-score* (simply indicated with *F1-score*) and *Accuracy* (global precision).

Implementation details: For the neural network approaches, the training stage has been conducted for 200 epochs. For methods based on fine-tuning, we used 100 epochs for the initial training and 100 epochs for the fine-tuning stage. For all methods, we adopt a learning rate of 10^{-4} , the AdamW (Loshchilov and Hutter, 2019) optimizer and a batch size of 256. Regarding *REFeD*, based on recent literature on contrastive learning (Chen et al., 2022), we set the temperature hyperparameter τ to 0.07 and we consider a batch size of 512 since it has been noted that contrastive loss benefits from larger batch sizes. The drop out value is set to 50%. Considering Random Forest classifiers, we optimize the model via the tuning of one parameter: the number of trees in the forest. We vary this parameter in the range {100, 200, 300, 400, 500}. The optimization of this parameter is based on the validation set.

Experiments are carried out on a workstation with a dual Intel (R) Xeon (R) CPU E5-2667v4 (@3.20GHz) with 256 GB of RAM and four TITAN X (Pascal) GPU. All the deep learning methods are implemented using the Pytorch deep learning library. All the models run on a single GPU. The Random Forest is implemented using the Python *Scikit-learn* library (Pedregosa et al., 2011).

4.3. Quantitative results

Tables 3 and 4 summarize the results obtained for the two study areas, *Koumbia* and *Centre Val de Loire* respectively by reporting the average F1-score and the

Strategy	Method	2020 + 2021 → 2021		2021 + 2020 → 2020	
		F1-Score	Accuracy	F1-Score	Accuracy
<i>Only Source</i>	RF	72.40 ± 3.94	72.25 ± 4.08	71.98 ± 3.54	71.81 ± 3.63
	TempCNN	64.54 ± 4.32	65.25 ± 4.14	69.54 ± 4.88	69.95 ± 5.00
<i>Target domain</i>	RF	77.56 ± 2.82	77.45 ± 2.90	76.78 ± 3.53	76.65 ± 3.54
	TempCNN	76.59 ± 2.94	76.63 ± 2.91	75.54 ± 5.04	75.47 ± 5.09
<i>Source+Target</i>	RF	77.49 ± 3.90	77.30 ± 3.93	78.42 ± 4.02	78.32 ± 4.03
	TempCNN	<u>78.60</u> ± 2.94	<u>78.48</u> ± 3.03	<u>78.95</u> ± 4.27	<u>78.92</u> ± 4.27
<i>Fine Tuning</i>	TempCNN	78.04 ± 2.75	78.00 ± 2.78	78.56 ± 3.63	78.55 ± 3.64
	Sourcerer	76.72 ± 2.63	76.54 ± 2.64	76.34 ± 3.94	76.24 ± 3.99
	POEM	78.35 ± 3.43	78.27 ± 3.41	78.41 ± 4.63	78.35 ± 4.66
	<i>REFeD</i>	79.23 ± 3.33	79.17 ± 3.27	82.15 ± 3.65	82.09 ± 3.67

Table 3: Overall performances (F1 scores) on Koumbia

Strategy	Method	2018 + 2021 → 2021		2021 + 2018 → 2018	
		F1-Score	Accuracy	F1-Score	Accuracy
<i>Only Source</i>	RF	69.66 ± 1.64	72.35 ± 1.07	59.31 ± 3.19	62.36 ± 2.56
	TempCNN	47.96 ± 1.23	46.26 ± 0.76	43.10 ± 2.46	41.68 ± 3.31
<i>Target domain</i>	RF	79.95 ± 2.45	80.70 ± 1.78	70.63 ± 3.05	71.86 ± 2.55
	TempCNN	82.60 ± 1.15	80.61 ± 1.41	73.39 ± 3.12	71.46 ± 3.46
<i>Source+Target</i>	RF	79.00 ± 2.24	79.83 ± 1.60	68.94 ± 2.66	70.45 ± 2.31
	TempCNN	<u>83.46</u> ± 1.40	81.54 ± 2.07	<u>75.17</u> ± 3.61	74.17 ± 3.71
<i>Fine Tuning</i>	TempCNN	82.94 ± 2.04	<u>83.51</u> ± 1.75	74.43 ± 3.78	75.70 ± 3.29
	Sourcerer	82.76 ± 1.69	83.49 ± 1.29	74.02 ± 3.36	74.93 ± 2.90
	POEM	82.89 ± 1.71	<u>83.51</u> ± 1.31	74.28 ± 3.11	<u>75.14</u> ± 2.80
	<i>REFeD</i>	84.45 ± 1.61	84.86 ± 1.35	77.60 ± 3.15	78.02 ± 2.90

Table 4: Overall performances (F1 scores) on *Centre Val de Loire* (CVL)

Accuracy considering the different combination of methods and strategies. As expected, regardless of the dataset, for both the RF and TempCNN classifiers, the lowest accuracy is obtained when only source-labeled data are considered, while the highest classification results are obtained when training data from both domains are used.

By focusing our attention on the TempCNN architecture, the F1-scores obtained using only the source-labeled data are 64.54 and 69.54 in *Koumbia* on EO data acquired in 2021 and 2020, respectively, and 47.96 and 43.10 in *Centre Val de Loire* on EO data acquired in 2021 and 2018, respectively. Although valuable, the historical reference data may not be completely representative of the recently acquired EO data. Moreover, the class statistical distributions of SITS acquired over different years can severely shift. Using only the target-labeled data the obtained accuracy increases, i.e., F1 scores of 76.59 and 75.54 in *Koumbia* on EO data acquired in 2021 and 2020, respectively, and 82.60 and 73.39 in *Centre Val de Loire* on EO data acquired in 2021 and 2018, respectively. It is worth noting that the importance of using labeled data from the target domain is even more visible in the *Centre Val de Loire* dataset since the source and target domains are different from both the spatial and temporal viewpoints. The joint use of source and target labeled data has a positive impact on the classification performances, leading to F1 scores of 78.60 and 78.95 in *Koumbia* on EO data acquired in 2021 and 2020, respectively, and 83.46 and 75.17 in *Centre Val de Loire* on EO data acquired in 2021 and 2018, respectively.

These results further improve when using SSDA and DG methods, which better combine source and target information. However, the highest classification accuracies are obtained by the proposed approach *REFeD* which achieves F1 scores

of 79.23 and 82.15 in *Koumbia* on EO data acquired in 2021 and 2020, respectively, and 84.45 and 77.60 in *Centre Val de Loire* on EO data acquired in 2021 and 2018, respectively.

4.3.1. Per-class analysis

Per-class performances are detailed in Tables 5 and 6 for the *Koumbia* study site and in Tables 7 and 8 and *Centre Val de Loire* respectively.

Concerning the *Koumbia* benchmark, Tables 5 and 6, we can observe that, no matter the transfer task, *REFeD* achieves the best performances in terms of F1-Score on all the agricultural land cover classes. This is of particular interest since such classes are characterized by strong shifts from one cultural year to another one due to crop rotations related to the underlying agricultural practices. A notable improvement, related to the proposed method, can also be noted on the *Forest* land cover class, especially for the transfer task (2021 + 2020 \rightarrow 2020). Regarding all the other land cover classes, *REFeD* achieves comparable results w.r.t. to all the other competing methods with, for all these cases, less than a point of difference in terms of F1-score.

Regarding the *Centre Val de Loire* benchmark, Tables 7 and 8, we can note that, *REFeD* obtains a systematic improvement, in terms of F1-Score, for the family of cereal classes (*Soft wheat*, *Maize*, *Barley* and *Other cereals*), regardless of the transfer task. Notably, the most significant enhancement is observed in the *Leguminous* crop class, where *REFeD* attains an improvement between 6 and 10 points of F1-Score compared to the best competing approach. This is even interesting since the *Leguminous* class is the most underrepresented crop type in the considered benchmark in terms of number of samples. This point further underscores the quality of the proposed approach demonstrating its ability to handle

Class	TempCNN (Source+Target)	Sourcerer	POEM	REFeD
Cereals	74.06	70.79	72.48	75.73
Cotton	75.77	72.79	76.04	77.87
Oleaginous	64.18	60.22	64.24	66.96
Grassland	78.58	75.62	78.11	77.61
Shrubland	78.90	76.47	78.32	77.90
Forest	84.27	85.06	84.78	86.05
Bare soil	83.45	80.11	82.83	83.04
Water	100.0	100.0	100.0	100.0

Table 5: Per-class average F1 scores for *Koumbia* scenario (2020 + 2021 → 2021).

scenarios characterized by significant class imbalances, a common situation in real-world applications. Finally, still regarding the *Centre Val de Loire* benchmark, we can underline that for the transfer task (2021 + 2018 → 2018), *REFeD* achieves the best performances for all the crop type classes.

4.4. Visual analysis

In this part of the experimental assessment, we provide qualitative analyses to further evaluate the behaviour of *REFeD* considering the *Koumbia* site, in the transfer task (2021 + 2020 → 2020). To this end, in addition to our framework, we also consider the top-performing competitors: *POEM*, *Sourcerer* and *TempCNN* (Source + Target). We first inspect some extracts from the obtained land cover maps and, then we visually examine the internal representations learned by the different methods by means of the t-SNE (van der Maaten and Hinton, 2008)

Class	TempCNN (Source+Target)	Sourcerer	POEM	REFeD
Cereals	77.48	74.25	76.39	79.38
Cotton	78.10	75.48	77.18	82.96
Oleaginous	72.91	69.93	71.69	75.14
Grassland	85.22	82.68	84.14	86.38
Shrubland	77.31	74.64	77.29	81.38
Forest	77.03	74.97	77.33	81.94
Bare soil	79.55	77.61	76.78	78.71
Water	100.0	100.0	100.0	100.0

Table 6: Per-class average F1 scores for *Koumbia* scenario (2021 + 2020 → 2020).

Class	TempCNN (Source+Target)	Sourcerer	POEM	REFeD
Soft wheat	91.35	91.53	90.52	91.91
Maize	94.91	92.72	94.73	96.14
Barley	94.94	94.71	94.75	95.58
Other cereals	85.55	85.84	84.36	87.54
Oleaginous	86.76	88.56	86.87	86.43
Winter Fallows	73.82	71.98	73.22	73.87
Leguminous	55.26	52.26	54.93	61.63
Fodder	74.31	73.04	72.11	75.82
Meadow	74.35	71.33	73.92	73.48
Other crops	84.16	83.88	84.06	83.92

Table 7: Per-class average F1 scores for *Centre Val de Loire* scenario (2018 + 2021 → 2021).

Class	TempCNN (Source+Target)	Sourcerer	POEM	REFeD
Soft wheat	85.55	85.23	84.21	85.80
Maize	85.65	85.48	86.23	87.04
Barley	94.28	94.34	93.63	94.63
Other cereals	73.35	73.87	71.35	73.79
Oleaginous	72.49	70.01	73.02	76.98
Winter Fallows	74.79	74.24	73.78	75.13
Leguminous	54.77	46.51	53.05	65.15
Fodder	70.47	68.24	66.91	72.07
Meadow	72.99	71.62	71.34	73.04
Other crops	57.40	57.76	57.66	64.06

Table 8: Per-class average F1 scores for *Centre Val de Loire* scenario (2021 + 2018 → 2018).

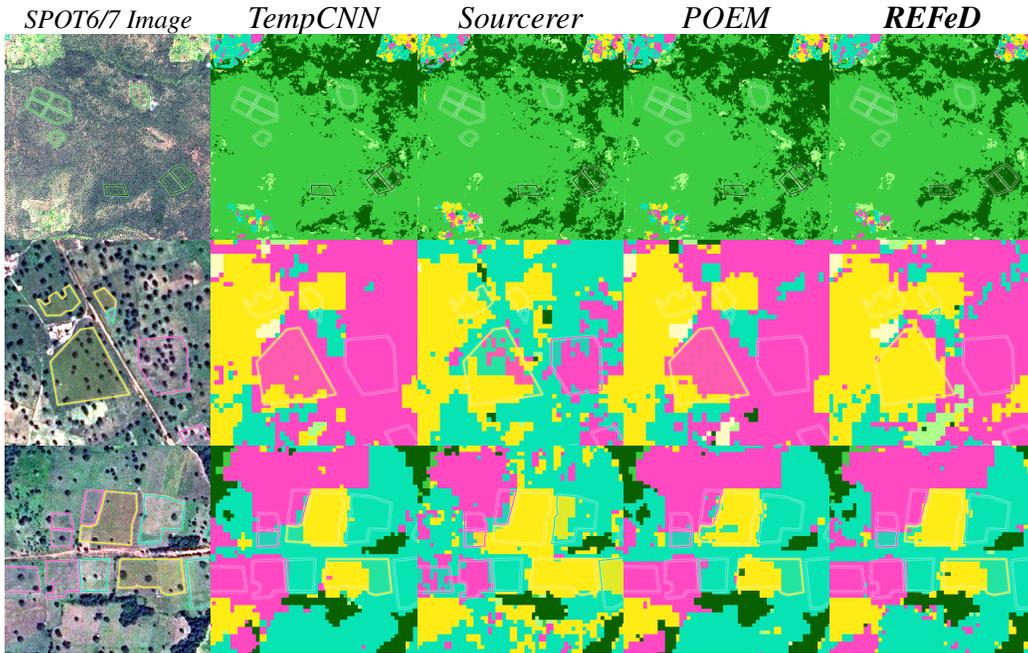


Figure 5: Extracts from the provided land cover maps per method. Ground truth areas outlined over the extracts using the same color codes of Fig. 3.

dimensionality reduction technique.

4.4.1. Land cover maps

In order to give a further insight into the performances of the proposed method, we performed a qualitative analysis of the land cover maps provided by each of the competing methods. In Fig. 5 we report some examples for the (2021 + 2020 \rightarrow 2020) transfer scenario over the *Koumbia* study site. As easily observable, the *Sourcerer* method tend to generate much noisier maps with respect to the competitors, which seems to be the main factor limiting its global performances. The other methods provide maps of comparable spatial characteristics, with *REFeD* significantly outperforming the competitors on classes related to natural vegeta-

tion with different densities (like the *Forest* class on the first row - bottom/right of the clip). More occasionally, *REFeD* also seem to retrieve the correct crop class to whole fields within the cropland which are entirely misclassified by other methods (e.g. second row, for the big *Cereal* field in the middle). Otherwise, the systematic improvement of *REFeD* over its best competitors mainly occurs at the finest scales, like for the fields in the example on the last row of Fig. 5, where most of the “holes” generated by *TempCNN* and *POEM* appear as properly filled.

4.4.2. Visualisation of internal model representations

In this last stage of our experimental evaluation, we provide a visual inspection of the internal feature representation learned by *REFeD*, *POEM*, *Sourcerer* and *TempCNN* (Source + Target). To this end, we randomly chose 50 samples per land cover class from the target domain and we extracted the corresponding feature representation per method. Subsequently, we applied t-SNE (van der Maaten and Hinton, 2008) to reduce the feature dimensionality for visualisation purposes. Results are depicted in Figure 6. We can note that all the methods well separate samples coming from the *Water* and *Baresoil* classes from the rest of the data. However, while competing approaches clearly mix samples from all other land cover classes together, *REFeD* partially alleviates clutter issues on the remaining classes providing a better visual behaviour in terms of cluster structure, on the considered subset of target data. This can be noted, for instance, regarding both the agricultural (*Cereals*, *Cotton* and *Oleaginous/Leguminous*) and natural vegetation (*Grassland*, *Shrubland* and *Forest*) classes. Overall, the visualisation of internal features representation is coherent with the quantitative as well as qualitative findings we previously discussed.

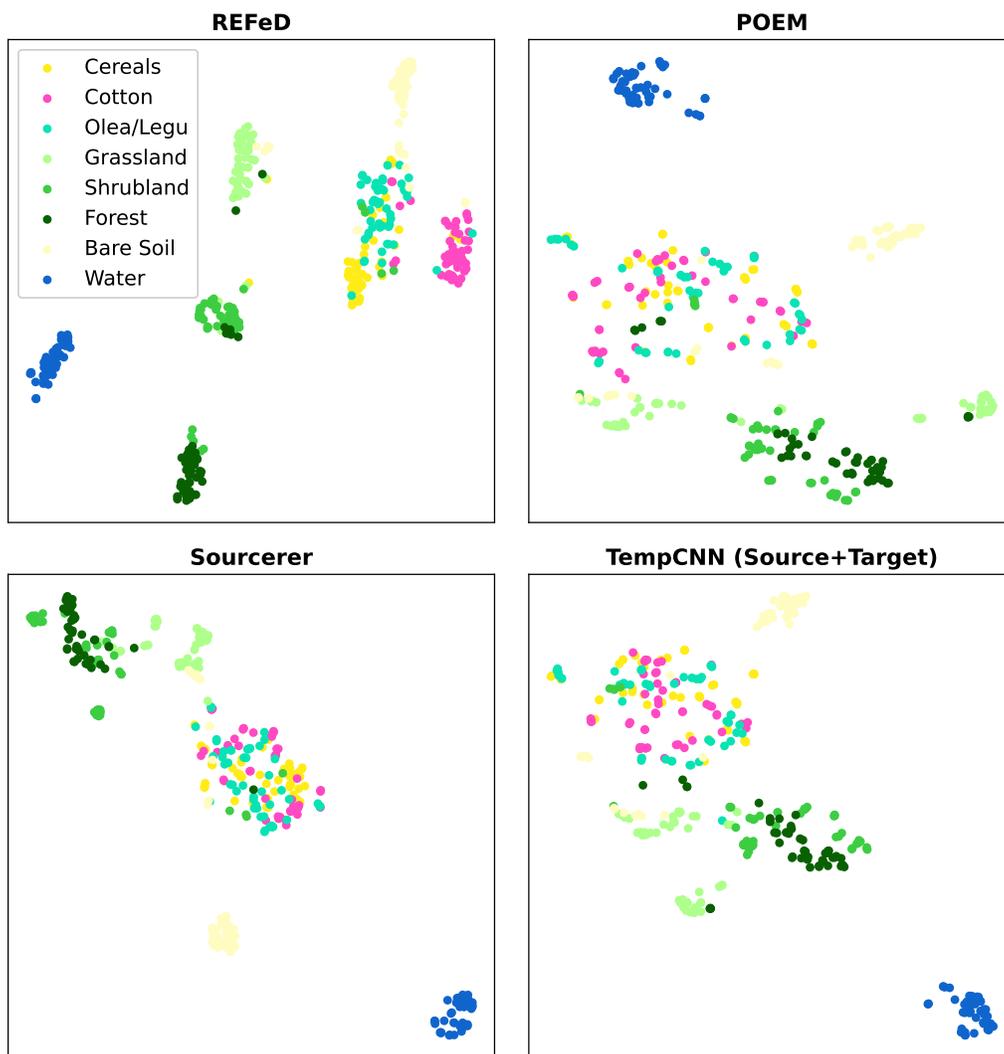


Figure 6: t-SNE results for the proposed approach (top left) and three different baselines: POEM (top right), Sourcerer (bottom left) and TempCNN (Source+Target) (bottom right).

5. Conclusion

In this work we have presented *REFeD*, a novel deep learning framework that enhances the accuracy of the current LC mapping process by combining together EO and reference data coming from two different domains (e.g. historical data and recent ones) with the aim to give value again to historical and/or overlooked reference data under a data-centric perspective.

REFeD is based on a pseudo-siamese architecture with unshared weights and it relies on contrastive learning to disentangle invariant and specific per-domain features to recover the intrinsic information related to the downstream LULC mapping task. Furthermore, *REFeD* is equipped with an effective supervision scheme where feature disentanglement is further enforced via multiple levels of supervision.

The obtained results on two study areas covering extremely diverse and contrasted landscapes have highlighted the quality of our framework regarding both quantitative and qualitative analyses with respect to all considered competitors. Most importantly, *REFeD* systematically outperforms models that only exploit target data paving the way to the reuse of historical and/or overlooked reference data for the LULC mapping task taking as input satellite image time series data.

Potential future extensions related to our framework may include extending *REFeD* to encompass a multi-source remote sensing setting where the two domains are described by different sensors (e.g. Sentinel-2 and Landsat). Additionally, we could contemplate a scenario where multiple historical or out-of-year datasets are available further advancing the effective reuse of overlooked and/or neglected efforts related to past resource-intensive field campaigns conducted on the same or related study areas.

Acknowledgements. This research was funded by the French Space Study Center (CNES, TOSCA 2024 project) and the French National Research Agency under the Investments for the Future Program, referred as ANR-16-CONV-0004 (DigitAg). Dino Ienco acknowledges support for this work from the *Agence Nationale de la Recherche* (ANR) under the ANR-23-IAS1-0002 (GEO-ReSeT) project.

References

- Ben-David, S., Blitzer, J., Crammer, K., Kulesza, A., Pereira, F., and Vaughan, J. (2010). A theory of learning from different domains. *Machine Learning*, 79:151–175.
- Bruzzone, L. and Marconcini, M. (2009). Domain adaptation problems: A dasvm classification technique and a circular validation strategy. *IEEE transactions on pattern analysis and machine intelligence*, 32(5):770–787.
- Capliez, E., Ienco, D., Gaetano, R., Baghdadi, N. N., and Hadj-Salah, A. (2023). Temporal-domain adaptation for satellite image time-series land-cover mapping with adversarial learning and spatially aware self-training. *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.*, 16:3645–3675.
- Chen, C., Zhang, J., Xu, Y., Chen, L., Duan, J., Chen, Y., Tran, S., Zeng, B., and Chilimbi, T. (2022). Why do we need large batchsizes in contrastive learning? A gradient-bias perspective. In *NeurIPS*.
- Chen, H., Zhang, Q., Huang, Z., Wang, H., and Zhao, J. (2023). Towards domain-specific features disentanglement for domain generalization. *CoRR*, abs/2310.03007.

- Gao, K., Yu, A., You, X., Qiu, C., Liu, B., and Zhang, F. (2023). Cross-domain multi-prototypes with contradictory structure learning for semi-supervised domain adaptation segmentation of remote sensing images. *Remote. Sens.*, 15(13):3398.
- Gómez, C., White, J. C., and Wulder, M. A. (2016). Optical remotely sensed time series data for land cover classification: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 116:55–72.
- Huang, Z., Zhong, L., Zhao, F., Wu, J., Tang, H., Lv, Z., Xu, B., Zhou, L., Sun, R., and Meng, R. (2023). A spectral-temporal constrained deep learning method for tree species mapping of plantation forests using time series sentinel-2 imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 204:397–420.
- Ienco, D., Gaetano, R., Dupaquier, C., and Maurel, P. (2017). Land cover classification via multitemporal spatial data by deep recurrent neural networks. *IEEE Geosc. and Rem. Sens. Letters*, 14(10):1685–1689.
- Inglada, J., Arias, M., Tardy, B., Hagolle, O., Valero, S., Morin, D., Dedieu, G., Sepulcre, G., Bontemps, S., Defourny, P., and Koetz, B. (2015). Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite optical imagery. *Remote Sensing*, 7(9):12356–12379.
- Jo, S. and Yoon, S. W. (2023). POEM: polarization of embeddings for domain-invariant representations. In Williams, B., Chen, Y., and Neville, J., editors, *AAAI*, pages 8150–8158.

- Jolivot, A., Lebourgeois, V., Leroux, L., Ameline, M., Andriamanga, V., Bellón, B., Castets, M., Crespín-Boucaud, A., Defourny, P., Diaz, S., Dieye, M., Dupuy, S., Ferraz, R., Gaetano, R., Gely, M., Jahel, C., Kabore, B., Lelong, C., le Maire, G., Lo Seen, D., Muthoni, M., Ndao, B., Newby, T., de Oliveira Santos, C. L. M., Rasoamalala, E., Simoes, M., Thiaw, I., Timmermans, A., Tran, A., and Bégué, A. (2021). Harmonized in situ datasets for agricultural land use mapping and monitoring in tropical countries. *Earth System Science Data*, 13(12):5951–5967.
- Karasiak, N., Dejoux, J.-F., Monteil, C., and Sheeren, D. (2021). Spatial dependence between training and test sets: another pitfall of classification accuracy assessment in remote sensing. *Machine Learning*.
- Kavvada, A., Metternicht, G., Kerblat, F., Mudau, N., Haldorson, M., Laldaparsad, S., Friedl, L., Held, A., and Chuvieco, E. (2020). Towards delivering on the sustainable development goals using earth observations. *Remote Sensing of Environment*, 247:111930.
- Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C., and Krishnan, D. (2020). In *NeurIPS*.
- Kolecka, N., Ginzler, C., Pazur, R., Price, B., and Verburg, P. H. (2018). Regional scale mapping of grassland mowing frequency with sentinel-2 time series. *Remote Sensing*, 10(8):1221.
- Kundu, J. N., Venkat, N., Revanur, A., Babu, R. V., et al. (2020). Towards inheritable models for open-set domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12376–12385.

- Long, M., CAO, Z., Wang, J., and Jordan, M. I. (2018). Conditional adversarial domain adaptation. In Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc.
- Loshchilov, I. and Hutter, F. (2019). Decoupled weight decay regularization. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net.
- Lucas, B., Pelletier, C., Schmidt, D. F., Webb, G. I., and Petitjean, F. (2023). A bayesian-inspired, deep learning-based, semi-supervised domain adaptation technique for land cover mapping. *Mach. Learn.*, 112(6):1941–1973.
- Mohammadi, S., Belgiu, M., and Stein, A. (2023). Improvement in crop mapping from satellite image time series by effectively supervising deep neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 198:272–283.
- Nyborg, J., Pelletier, C., Lefèvre, S., and Assent, I. (2022). Timematch: Unsupervised cross-region adaptation by temporal shift estimation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 188:301–313.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

- Pelletier, C., Webb, G. I., and Petitjean, F. (2019). Temporal convolutional neural network for the classification of satellite image time series. *Remote. Sens.*, 11(5):523.
- Roscher, R., Rußwurm, M., Gevaert, C., Kampffmeyer, M., dos Santos, J. A., Vakalopoulou, M., Hänsch, R., Hansen, S., Nogueira, K., Prexl, J., and Tuia, D. (2023). Data-centric machine learning for geospatial remote sensing data. *CoRR*, abs/2312.05327.
- Schneider, M., Schelte, T., Schmitz, F., and Körner, M. (2023). EuroCrops: The Largest Harmonized Open Crop Dataset Across the European Union. *Scientific Data*, 10(1):612.
- Sun, B. and Saenko, K. (2016). Deep CORAL: correlation alignment for deep domain adaptation. In Hua, G. and Jégou, H., editors, *Computer Vision - ECCV 2016 Workshops - Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III*, volume 9915, pages 443–450.
- Tuia, D., Persello, C., and Bruzzone, L. (2016). Recent advances in domain adaptation for the classification of remote sensing data. *Geoscience and Remote Sensing Magazine*, 4(2):41–57.
- van der Maaten, L. and Hinton, G. E. (2008). Visualizing data using t-sne. *Journal of Machine Learning Research*, 9:2579–2605.
- Wilson, G. and Cook, D. J. (2020). A survey of unsupervised deep domain adaptation. *ACM Trans. Intell. Syst. Technol.*, 11(5):51:1–51:46.
- Wulder, M. A., Masek, J. G., Cohen, W. B., Loveland, T. R., and Woodcock, C. E. (2012). Opening the archive: How free data has enabled the science

and monitoring promise of landsat author links open overlay panel. *Remote Sensing of Environment*, 122:2–10.

Yu, Y. and Lin, H. (2023). Semi-supervised domain adaptation with source label adaptation. In *CVPR*, pages 24100–24109. IEEE.

Zha, D., Bhat, Z. P., Lai, K., Yang, F., Jiang, Z., Zhong, S., and Hu, X. (2023). Data-centric artificial intelligence: A survey. *CoRR*, abs/2303.10158.

Zhong, L., Hu, L., and Zhou, H. (2019). Deep learning based multi-temporal crop classification. *Remote Sensing of Environment*, 221:430–443.