Unsupervised multi-target domain adaptation for deforestation detection in tropical rainforest

Mabel X. Ortega Adarme[1,2], Gilson A. O. P. da Costa[3], Pedro J. Soto Vega[4], Christian Heipke[2], Raul Q. Feitosa[1]

[1] Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Brazil - mortega@aluno.puc-rio.br, raul@ele.puc-rio.br
[2] Institute of Photogrammetry and GeoInformation, Leibniz Universität Hannover, Germany - heipke@ipi.uni-hannover.de
[3] State University of Rio de Janeiro, Brazil - gilson.costa@ime.uerj.br
[4] L@bISEN, Vision-AD and Auto-ROB, France - pedro-juan.soto-vega@isen-ouest.yncrea.fr

Keywords: Deforestation detection, domain adaptation, multi-target.

Abstract

Geographic variability of the classes of interest, differences in sensor characteristics and changes in atmospheric conditions during image acquisition, among other factors, make it challenging to use a pre-trained deep learning classifier on new remote sensing data without a substantial drop in classification accuracy. This phenomenon occurs due to the so-called domain shift problem. Deep domain adaptation techniques have been used to mitigate the problem and thus avoid the time-consuming and costly collection of new labeled samples. Most recent domain adaptation approaches rely on single-source and single-target domains, refraining from exploiting other data distributions that are usually available. This work introduces a new unsupervised multi-target domain adaptation in the context of a change detection application, namely deforestation detection. The proposed approach addresses the substantial class imbalance typical of such application by applying unsupervised algorithms for selecting pseudo-labels in the target domain that will later serve as additional training references. We report results of experiments to evaluate the proposed method in four distinct sites of two Brazilian biomes using Sentinel-2 images. The results indicate that the proposed unsupervised domain adaptation method is a promising solution to reduce the effects of domain shift and to deal with the scarcity of labeled training data.

1. Introduction

In the last decades, significant technological progress has facilitated access to a vast amount of remote sensing (RS) data. Such an increase in data availability has allowed to better explore a number of applications concerned with human activities, ecosystems, and their interactions. In the context of RS, change detection (CD) has emerged as a prominent area of interest due to its capacity to identify changes in land cover, land use, and environmental conditions.

The target application of this work is bi-temporal deforestation detection within the Brazilian tropical forests, a problem that is of paramount importance since deforestation is one of the primary contributors to climate change (Cabral et al., 2024). Specifically, in this work we automatically analyse pairs of co-registered satellite images covering sites of different Brazilian biomes to identify the occurrence of deforestation between two epochs, at the pixel level.

In that regard, the National Institute for Space Research (INPE) conducts systematic annual assessments to detect deforestation through the Program for Deforestation Monitoring in the Brazilian Amazon and other Biomes (PRODES)[1]. Despite delivering highly accurate deforestation mapping, the methodology employed in PRODES relies on visual photo-interpretation. Considering the total cost and time involved in such a human intensive procedure, automating it would offer obvious advantages.

Several deep learning (DL) methods have been proposed for CD, delivering high success rates in many RS tasks, including deforestation detection (Ramachandran et al., 2024, Dalagnol et al., 2023). Even though such methods produce good results when the training data is characterized by a distribution that is similar to that of the test data, their performances decrease substantially when there is a significant difference between the training and test data distributions (Tasar et al., 2020, Soto Vega et al., 2021). That phenomenon is regarded as domain shift (DS). In the context of deforestation detection, a domain may correspond to a dataset from a particular geographic location. In this case, DS is mostly related with inherent inter- and intra-class variability, which may be associated with particular forest types and deforestation practices. A number of domain adaptation (DA) techniques have been proposed to mitigate the DS problem (Tuia et al., 2016, Wittich and Rottensteiner, 2021) in various applications.

Existing DA approaches typically address adaptation issues in one of two scenarios: (semi) supervised DA and unsupervised DA. The first considers that besides having labeled instances from the source domain, there is a small portion of labeled instances from the target domain available for training or adapting a classifier. On the contrary, unsupervised DA (uDA) does not use any labels from the target domain, but instead exploits the similarity of the domain data distributions. In the present work, we pursue the uDA scheme, which is more demanding due to the lack of label correspondences between source and target (Gholami et al., 2020, Yang et al., 2020).

Most uDA research focuses on single-source and single-target domain strategies. However, in several real-world applications, and specially in RS, unlabeled data may come from different domains with different properties, but with content related to a common task. In that regard, an ideal uDA strategy for large-scale classification should be able to learn from multiple labeled source domains and properly process multiple unlabeled target domains, even when all domains have different data distributions (Tasar et al., 2020).

Another important factor to be considered in the deforestation

[1] http://terrabrasilis.dpi.inpe.br/map/deforestation

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-3-2024-293-2024 | © Author(s) 2024. CC BY 4.0 License.

detection application is the high level of class imbalance, since usually a minor portion of the areas of interest correspond to deforestation. In the uDA methods, this is an important problem, since labeled training samples are only available for the source domain. Therefore, the absence of labels in the target domains can cause DA methods to be biased towards the majority class (i.e., no-deforestation). That would be the effect of randomly selecting target samples to take part in the DA procedure. An alternative is to use some pseudo-labeling strategy to assign (semi) labels to unlabeled data in an unsupervised way. However, this can result in many incorrect (noisy) pseudo-labels and low-confidence outcomes.

In order to tackle the above mentioned issues, this work proposes an unsupervised, multi-target domain adaptation method for deforestation mapping. Our method adapts the so-called Multi-Target DA Information-Theoretic-Approach (MTDA-ITA) (Gholami et al., 2020), which was initially proposed for image classification, to perform change detection through semantic segmentation. The main contributions of this work are the following:

- An unsupervised multi-target deep domain adaptation method based on fully convolutional neural networks, which exploits the spatial context in the image data to improve classification accuracy.
- An unsupervised strategy for balancing target domain samples in the domain adaptation training procedure to deal with severe class imbalance.
- An evaluation of the proposed method on a deforestation mapping application, in which the domains consist of pairs of remote sensing images covering four distinct sites in two biomes of the Brazilian rain forest.

2. Related work

In this section, we first present an overview of recent work addressing the task of deforestation detection. Afterwards, we summarize publications which describe domain adaptation in the context of multi-target scenarios.

Several DL-based approaches to quantify deforestation processes with bi-temporal images have been proposed, most of them following an encoder-decoder architectures (Md Jelas et al., 2024). For instance, various Fully Convolutional Networks (FCN) were evaluated in (Ortega et al., 2021, Torres et al., 2021). The authors conducted a comparative analysis of various FCN architectures, assessing their performance across multiple sites within the Amazon region. For the experiments data from PRODES project and image pairs from different RS sensors including Sentinel-1, -2, and Landsat-8 were used. Similarly, (Andrade et al., 2022) evaluated a DeepLabv3+ model variant to detect deforestation in the Amazon forest using Landsat-8 image pairs. Although, the reported results showed satisfactory classification performance, the methods were evaluated under optimal conditions where the training and test data distribution present similar characteristics, potentially limiting their generalization capability to other regions or datasets (Md Jelas et al., 2024). Considering the complexity of the Amazon forest, when these models are used on datasets with temporal and spatial variations, a substantial decrease in classification performance may be observed, as is reported in (Soto Vega et al., 2021). Therefore, to successfully apply DL approaches in real-world applications, the Domain Shift (DS) issue must be addressed (Sun et al., 2016).

According to the recent literature, Domain Adaptation (DA) represents the set of techniques used to mitigate the DS problem (Tuia et al., 2016), and these techniques have been extensively explored in the RS field. Regarding uDA methods, they are commonly divided into discrepancy-based approaches and adversarial-based techniques (Peng et al., 2022). The former group tries to find domain-invariant features to minimize the domain shift by using statistical measures (e.g., MMD-based metric) (Long et al., 2015, Sun and Saenko, 2016, Long et al., 2017, Zhu et al., 2020) The latter is inspired by generative adversarial nets (GAN) (Goodfellow et al., 2014), and learns domain-invariant features through adversarial learning. Following the structure of GANs, these methods comprise a generator and a discriminator networks. Here the task of the generator is to produce samples similar to the source domain to encourage domain confusion to the discriminator network (Wittich and Rottensteiner, 2021, Tzeng et al., 2017, Ganin et al., 2016). Although the methods discussed above reported promising results, they are primarily concentrate on employing a single source domain and rarely address the transfer of information across multiple distributions (Li et al., 2023).

Regarding multi-target DA methods, Multi-teacher Multi-target DA (MT-MTDA) (Nguyen-Meidine et al., 2021) was proposed to transfer information via knowledge distillation. In this approach, each target domain is associated to a special so called teacher network, and the information is transferred to a single common network called student. Instead of learning domain-adapted features directly, each teacher instructs the student on how to perform uDA for a particular target. In (Gholami et al., 2020), an information-theoretic approach for DA was also introduced. The method tries to separate shared and private information across domains and learns to maximize mutual information between domain labels and domain-specific features while minimizing mutual information between domain labels and shared features. Although the previous methods report satisfactory results, all of them are focused on the image classification task. In the context of semantic segmentation, Isobe et al. (Isobe et al., 2021), introduced an unsupervised multi-target DA relying on a collaborative learning framework. In this scheme, different expert models are trained with data available from all domains, and the individual results are exchanged following a collaborative learning strategy by optimizing the Kullback-Leibler (KL)-divergence among multiple expert distributions. The result is used to teach a student model, which is able to perform suitably well across multiple-target domains. Similarly, Saporta et al. (Saporta et al., 2021) proposed a method based on two adversarial frameworks, a multi-discriminator to align the target domains, and multi-target knowledge transfer employing a multi-teacher/single-student distillation scheme. In a more complex scenario, Tasar et al. (Tasar et al., 2020) introduced DAugNet, which is an unsupervised, multi-source, multi-target, life-long DA method. This approach is composed of a data augmentor and a classifier. During training, the data augmentor applies a style transfer to all domains to generate more diversified data, and therefore train a more robust classifier. However, one drawback of this method is that the pre-processing and the DA stages have to be trained in a sequential order, preventing an end-to-end training.

Few DA methods have been proposed in the context of

CD (Deng et al., 2019, Chen et al., 2020, Saha et al., 2020), most of them focusing on urban changes. In the particular case of deforestation detection, works based on the DANN (Soto Vega et al., 2022), ADDA (Noa et al., 2021) and CycleGAN (Soto Vega et al., 2021) approaches are available, but all of them for single-target DA only. In CD applications such as deforestation detection, in which a high-class imbalance usually occurs, DA methods tend to be biased towards the majority class, disregarding minority classes, thus presenting poor adaptation performance (Zou et al., 2018). In that regard, Soto, et al. (Soto Vega et al., 2022) introduced an unsupervised technique for balancing the target domain samples during the adaptation procedure. The technique relies on pseudo-labels produced through unsupervised Change Vector Analysis (CVA). An important aspect of that attempt is that the pseudo-labels are noisy in the sense that they may be wrong at some pixels.

3. Methodology

The method presented in this work is designed to tackle the problem of deforestation detection in a multi-domain context, where source and target domains may significantly differ with respect to environmental conditions and deforestation practices. The main goal of our approach is to provide robust unsupervised domain adaptation in the context of deforestation detection, operating on a source domain with enough labeled training samples and multiple target domains without any labeled training samples.

Following (Gholami et al., 2020), let’s consider a labeled source domain $\mathcal{D}_S = \left\{\left(\mathbf{x}_i^S, \mathbf{y}_i^S\right)\right\}_{i=1}^{n_S}$, where $\mathbf{x}_i^s$ indicates the i-th sample and y_i^s is its corresponding label, n_S is the number of labeled samples in the source domain, and m unlabeled target domains $\mathcal{D}_{T_m} = \left\{\left\{\mathbf{x}_i^T\right\}_{i=1}^{n_T}\right\}_{m=1}^{M}$, where $\mathbf{x}_i^{T_m}$ is the i-th sample, and n_{T_m} is the number of unlabeled samples in the m-th target domain. For our application, $\mathbf{x}_i$ corresponds to two co-registered images denoted as $\mathbf{x}_{i_{t_0}}$ and $\mathbf{x}_{i_{t_1}}$, acquired on dates t_0 and t_1 that define the time interval within which we want to detect deforested regions. The images are concatenated along the spectral dimension, producing a tensor $\mathbf{x}_i \in \mathbb{R}^{H\times W\times 2B}$, where H and W denote the spatial dimensions, and B the number spectral bands from each image. $\mathbf{y}_i^S$ denotes the class label map of $\mathbf{x}_i^S$, in which each pixel location takes a value from the set $\{0, 1\}$, where 1 means *Deforestation (DF)*, and 0 means *No-deforestation (NDF)*. $\mathbf{d}_i$ represents the one-hot representations of the domain labels for $\mathbf{x}_i$. The latent space representation of an input sample $\mathbf{x}_i$, either from the source or from a target domain, is denoted as $\mathbf{z}_{s_i}$ in the shared latent space, and $\mathbf{z}_{p_i}$ in the private latent space. In addition, $\hat{\mathbf{x}}_i$ represents the reconstruction of the input $\mathbf{x}_i$.

Figure 1 shows the method’s components. They comprise

– a shared encoder E_s, with parameters θ_s, that captures the common features $\mathbf{z}_{s_i}$ across domains, formally,

$$\mathbf{z}_{s_i} = E_s\left(\mathbf{x}_i; \theta_s\right) \tag{1}$$

– a private encoder E_p, with parameters θ_p, for learning domain-specific features $\mathbf{z}_{p_i}$, formally,

$$\mathbf{z}_{p_i} = E_p\left(\mathbf{x}_i; \theta_p\right) \tag{2}$$

– a decoder F, with parameters ϕ, which produces a reconstruction $\hat{\mathbf{x}}_i$ of the input $\mathbf{x}_i$ from $\mathbf{z}_{s_i}$ and $\mathbf{z}_{p_i}$, formally,

$$\hat{\mathbf{x}}_i = F\left(\mathbf{z}_{s_i}, \mathbf{z}_{p_i}; \phi\right) \tag{3}$$

– a domain classifier D, with parameters ψ that aims to predict at its output $\hat{\mathbf{d}}_i$ the domain label from $\mathbf{z}_i$, where $\mathbf{z}_i$ corresponds either to $\mathbf{z}_{s_i}$ or $\mathbf{z}_{p_i}$, formally,

$$\hat{\mathbf{d}}_i = D\left(\mathbf{z}_i; \psi\right) \tag{4}$$

– a classifier C, with parameters γ, whose task is to infer at its output the class label map relying only on $\mathbf{z}_{s_i}$, formally,

$$\hat{\mathbf{y}}_i = C\left(\mathbf{z}_{s_i}; \gamma\right) \tag{5}$$

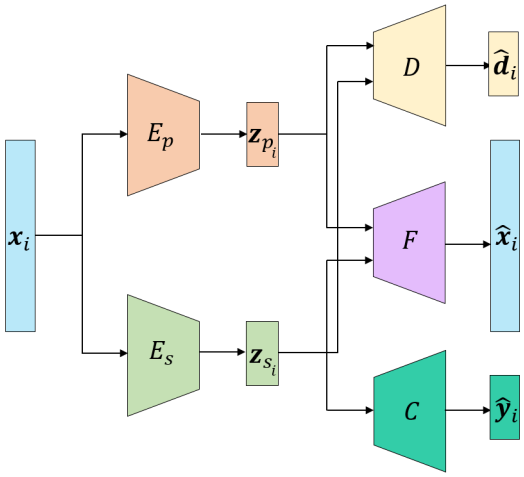


Figure 1. Structure of the proposed multi-target, deforestation detection DA method. The encoders shared E_s and private E_p capture the common and domain-specific features, respectively. The decoder F attempts to recreate the input sample from the shared and private features. The domain classifier D learns to predict the domain labels whether from shared or private features. The classifier C learns to predict the class label from the shared features.

For training, the model relies on a loss function that combines the following terms:

– **Decoder loss $\mathcal{L}_F$:** refers to the difference between the input $\mathbf{x}_i$ and its reconstruction $\hat{\mathbf{x}}_i$ at the output, formally:

$$\mathcal{L}_F = \frac{\lambda_r}{N}\sum_{i=1}^{N}\left\|\mathbf{x}_i - \hat{\mathbf{x}}_i\right\|_1 \tag{6}$$

where λ_r denotes the weight of the reconstruction loss.

– **Domain classifier loss $\mathcal{L}_D$:** is composed of the sum of the cross entropy of the domain classifier output having as input the shared and the private features, formally:

$$\mathcal{L}_D = -\frac{\lambda_{ds}}{N}\sum_{i=1}^{N}\mathbf{d}_i^\top ln\left(\hat{\mathbf{d}}_{s_i}\right) - \frac{\lambda_{dp}}{N}\sum_{i=1}^{N}\mathbf{d}_i^\top ln\left(\hat{\mathbf{d}}_{p_i}\right) \tag{7}$$

where λ_{ds} and λ_{dp} denote the weight of the multi-domain separation loss using the shared and private features, respectively.

– **Label classifier loss $\mathcal{L}_C$:** refers to the cross entropy of the classifier outcome computed only upon the source domain samples, formally:

$$\mathcal{L}_C = -\frac{1}{N_s}\sum_{i=1}^{N_s}\mathbf{y}_i^\top ln\left(\hat{\mathbf{y}}_i\right) \tag{8}$$

– **Shared encoder loss $\mathcal{L}_S$** : is made up of three terms: the decoder loss, the classifier loss, and the part of the domain classifier loss referring to shared features, formally:

$$\mathcal{L}_S = \frac{\lambda_r}{N}\sum_{i=1}^{N}\left\|\mathbf{x}_i - \hat{\mathbf{x}}_i\right\|_1 - \frac{\lambda_c}{N_s}\sum_{i=1}^{N_s}\mathbf{y}_i^\top ln\left(\hat{\mathbf{y}}_i\right) + \frac{\lambda_{ds}}{N}\sum_{i=1}^{N}\mathbf{d}_i^\top ln\left(\hat{\mathbf{d}_{s_i}}\right) \tag{9}$$

where λ_c denotes the weight of the classification loss. Note, that the domain classifier output occurs in $\mathcal{L}_S$ with opposite sign, implementing adversarial training

– **Private encoder loss $\mathcal{L}_P$:** is composed of two terms, the decoder loss and the domain classifier loss using the privated features, formally:

$$\mathcal{L}_P = \frac{\lambda_r}{N}\sum_{i=1}^{N}\left\|\mathbf{x}_i - \hat{\mathbf{x}}_i\right\|_1 - \frac{\lambda_{dp}}{N}\sum_{i=1}^{N}\mathbf{d}_i^\top ln\left(\hat{\mathbf{d}_{p_i}}\right) \tag{10}$$

3.1 Training

The training process estimates the parameter values $\hat{\theta}_s, \hat{\theta}_p, \hat{\gamma}, \hat{\psi}$, and $\hat{\phi}$ by iteratively updating each component of the method. Agorithm 1 summarizes the DA process. Here, batches ($\mathcal{B}^S$ and $\mathcal{B}_m^T$) are sampled from the labeled training data of the source domain denoted as (L^S) and of the m unlabeled target domains denoted as U_m^T. To ensure that training samples of the target domains have pixels belonging to classes *DF* and *NDF*, we generated pseudo-labels with unsupervised algorithms, denoted as $\tilde{\mathbf{y}}_j^T$ containing a value for each pixel location either 0 or 1 for classes *DF* and *NDF*, respectively. It is worth mentioning that only the parameters γ of C are determined by conventional supervised training using L^S.

3.2 Pseudo-label generation

One key idea refers to the strategy for training sample selection from the target domains $\mathcal{D}_T$. We extended the strategy presented in (Li et al., 2021), where an ensemble of algorithms is employed instead of relying on a single unsupervised algorithm. Therefore, for a sample to qualify for the training set of the target domains in the DA models, the pseudo-labels assigned to that sample have to meet some consistency criterion.

Different consistency criteria can be considered. One possibility, which we explored in the experimental analysis, imposes unanimity among all ensemble members. We used two unsupervised algorithms to build up the ensemble whose outcomes were further subjected to a consistency criterion to obtain the pseudo-label maps for the target domains.

The first algorithm is CVA (Malila, 1980). It computes the magnitude $\mathcal{M}(h,w)$ and direction α of change between the image pair $\mathbf{x}_{t_0}^T$ and $\mathbf{x}_{t_1}^T$ at each pixel position (h,w). To obtain the binary changed maps we used the Otsu algorithm to find the optimal thresholds T_{mg} and T_{ph} using the normalized histograms from $\mathcal{M}$ and α.

Algorithm 1: Unsupervised multi-target domain adaptation algorithm

Input:
- L^S // `labelled training data from source domain`
- U_M^T // `unlabelled training data from` M`-target domains`
- $\{\theta_s, \theta_p, \gamma, \phi, \psi\}$ // `initial random weights of` $\hat{\theta_s}$, $\hat{\theta_p}$, $\hat{\gamma}$, $\hat{\phi}$, `and` $\hat{\psi}$
- $\{\lambda_r, \lambda_{ds}, \lambda_{dp}, \lambda_c\}$ // `model hyper-parameters`
- β // `set of optimizer parameters`

Output: $\{\theta_s, \theta_p, \gamma, \phi, \psi\}$ // `optimal model parameters`

while $StopCriterion ==$ ***False*** **do**
- $\mathcal{B}^S \leftarrow$ batch from L^S // `labeled data from` $\mathcal{D}_S$
- $\mathcal{B}_M^T \leftarrow$ batch from U_M^T // `unlabeled data from` $\mathcal{D}_{T_M}$
- $\theta_s \leftarrow$ by minimizing $\mathcal{L}_S$ using B^S and B_M^T // `Eq. 9`
- $\theta_p \leftarrow$ by minimizing $\mathcal{L}_P$ using B^S and B_M^T // `Eq. 10`
- $\phi \leftarrow$ by minimizing $\mathcal{L}_F$ using B^S and B_M^T // `Eq. 6`
- $\psi \leftarrow$ by minimizing $\mathcal{L}_D$ using B^S and B_M^T// `Eq. 7`
- $\gamma \leftarrow$ by minimizing $\mathcal{L}_C$ using B^S // `Eq. 8`

End while

Figure 2. Training process of the unsupervised multi-target domain adaptation method. The parameter values $\hat{\theta}_s, \hat{\theta}_p, \hat{\gamma}, \hat{\psi}$, and $\hat{\phi}$ are iteratively updated to optimize the adaptation across multiple target domains.

Next, we formed a set $\tilde{\mathbf{y}}_{cva}$ with the pseudo labels at each pixel position (h,w) that meet one of the following conditions,

$$\tilde{\mathbf{y}}_{cva} = \begin{cases} DF, & \text{if } (\mathcal{M}(h,w) \geq T_{mg}) \text{ and } (\alpha(h,w) \geq T_{ph}) \\ NDF, & \text{otherwise} \end{cases} \tag{11}$$

The second algorithm is SSIM (Wang et al., 2004). This measure was initially introduced for assessing image similarity. However, it can estimate whether a pair of pixels has changed or not in CD tasks, through statistical similarity measures for a image pair $\mathbf{x}_{t_0}^T$ and $\mathbf{x}_{t_1}^T$ by computing:

$$SSIM_{dif}(h,w) = 1 - SSIM(h,w)$$

for all pixel positions. Similar to *CVA*, the threshold $Tssim$ was computed by the Otsu algorithm using the normalized histogram from $SSIM_{dif}$: Again, we formed a set $\tilde{\mathbf{y}}_{ssim}$ with the pseudo labels at each pixel position (h,w) that met one of the following conditions,

$$\tilde{\mathbf{y}}_{ssim} = \begin{cases} DF, & \text{if } (SSIM_{dif}(h,w) \geq T_{ssim}) \\ NDF, & \text{otherwise} \end{cases} \tag{12}$$

Lastly, the final pseudo-label map $\tilde{\mathbf{y}}^T$ is produced by applying the consistency criterion, which was defined as the unanimity among all ensemble outputs and follows the criterion expressed below:

$$\tilde{\mathbf{y}}^T = \tilde{\mathbf{y}}_{cva} \text{ and } \tilde{\mathbf{y}}_{ssim}$$

3.3 Network architectures

Tables 2, 3, 4, and 5 present detailed architectures of the shared E_s, private E_p encoders, decoder F, domain classifier D, and the FCN classifier C, respectively. The input for E_s and E_p comprises patches with dimensions $H \times W \times 2B$, where H, W, and $2B$ represent the height, width, and number of bands of each sample, respectively, and the outputs are tensors with dimension $32 \times 32 \times 64$. These tensors are concatenated and fed into F, which produces an output matching the dimensions of the input training samples $H \times W \times 2B$. Additionally, the outputs of E_s and E_p are input to D, which generates domain predictions with dimensions $1 \times ND$, where ND is the number of domains. Lastly, C takes the output of E_s as input and generates label map predictions with dimensions $128 \times 128 \times CL$, where CL corresponds to the number of classes, specifically, DF and NDF. In addition, Table 6 describes the network architecture used for the classifier selected as a baseline, where the DA method was not applied, which follows an encoder-decoder structure.

Layer	Layer type	H/W	Depth
1	Input layer	128	2B
2	Conv(7), stride 1, IN, ReLu	128	16
3	Conv(3), stride 2, IN, ReLu	64	32
4	Conv(3), stride 2, IN, ReLu	32	64
5-10	Residual block, IN, ReLu	32	64

Table 1. Architecture of shared E_s and private E_p encoders.

Layer	Residual block	H/W	Depth
1	Input layer	32	64
2	Conv(3), stride 1	32	64
3	Conv(3), stride 1	32	64
4	Add(3, 1)	32	64

Table 2. Architecture of residual block of E_s and E_p

Layer	Layer type	H/W	Depth
1	Input layer	32	128
2	Conv(3), stride 1, IN, ReLu	32	64
3	Residual block2, IN, ReLu	32	64
4	Upsample	64	32
5	Upsample	128	16
6	Conv(1), stride 1, TanH	128	2B

Table 3. Architecture of the decoder F.

Layer	Layer type	H/W	Depth
1	Input layer	32	64
2	Conv(3), stride 1, IN, ReLu	32	4
3	Conv(3), stride 1, IN, ReLu	32	8
4	Conv(3), stride 1, IN, ReLu	32	16
5	Conv(3), stride 1, IN, ReLu	32	32
6	Flatten		
7	Dense, Softmax	1/ND	

Table 4. Architecture of the domain classifier D.

	Layer	Layer type	H/W	Depth
Encoder	1	Input layer	32	64
	2	Conv(3), stride 1, ReLu	32	64
	3	Conv(3), stride 2, ReLu	16	128
	4	Conv(3), stride 2, ReLu	8	128
Decoder	5	Upsample,	16	256
	6	Upsample,	32	128
	7	Upsample	64	32
	8	Upsample	128	16
	9	Conv(1), Softmax	128	CL

Table 5. Architecture of the FCN classifier C.

	Layer	Layer type	H/W	Depth
Encoder	1	Input layer	128	2CH
	2	Conv(7), stride 1, ReLu	128	16
	3	Conv(3), stride 2, ReLu	64	32
	4	Conv(3), stride 2, ReLu	32	64
	5	Conv(3), stride 2, ReLu	16	128
	6	Conv(3), stride 2, ReLu	8	128
Decoder	7	Upsample	16	128
	8	Upsample	32	64
	9	Upsample	64	32
	10	Upsample	128	16
	11	Conv(1), Softmax	128	CL

Table 6. Architecture of the FCN classifier (baseline).

4. Experiments

In this section, we provide a detailed overview of the experiments conducted to illustrate the potential of the developed method and subsequently an analysis of the results. Initially, we introduce the dataset used for training and evaluation of the classifier. Next, we describe the experimental setup, and finally, we report and discuss the obtained results.

4.1 Study areas

This study relied on Sentinel-2 data from four sites within the Amazon and Cerrado Brazilian Biomes. The sites are located in the Brazilian states Pará (PA), Mato Grosso (MT), Rondônia (RO), and Maranhão (MA). The images were downloaded and preprocessed to Level-1C using the Google Earth Engine (GEE) platform (Gorelick et al., 2017). We used all bands with spatial resolutions of 10m and 20m. The 20m bands were resampled to 10m using the nearest neighbor technique.

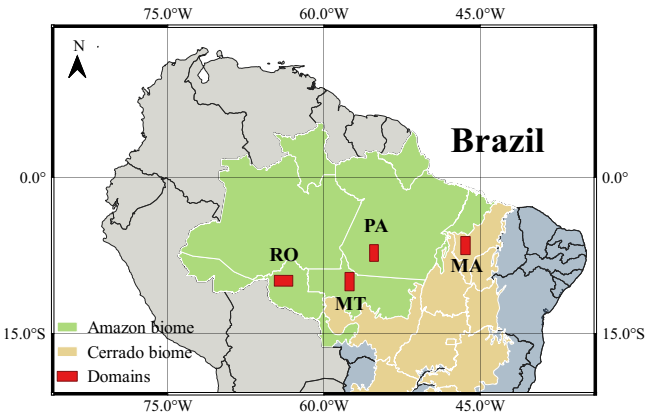


Figure 3. Geographical location of the study areas.

The reference change maps used in the experiments consider the deforestation that occurred between the epochs t_0 and t_1, in our case corresponding to the years 2020 and 2021, which are reported in the PRODES deforestation mapping program (INPE, 2021). That information is freely available and was downloaded from thethe PRODES database of the Brazilian National Institute for Space Research (INPE) site. This project measures the annual deforestation rate from August 1st of each year, during the dry season (June to September), when the cloud cover is at a minimum. It is worth mentioning that PRODES only contains information regarding primary deforestation, i.e. areas previously identified as deforested are omitted from the annual manual labeling process. Since there is uncertainty about these regions which may or may not face another regrowth and deforestation cycle, those

Domains	PA	MT	RO	MA
Biome	Amazon	Amazon	Amazon	Cerrado
Vegetation	Dense ombrophyll	Dense and open ombrophyll	Open ombrophyll	Seasonal deciduous and semidecidous
Dimensions (px)	9200×17730	9544×19430	11384×19365	10000×19295
Date t_0	July 15, 2020	August 2, 2020	July 29, 2020 August 01, 2020	August 2, 2020 August 10, 2020
Date t_1	July 25, 2021 August 4, 2021	July 23, 2021	July 19, 2021 July 22, 2021	August 20, 2021
DF (%)	1.86	0.95	1.30	1.30
ND (%)	56.40	59.62	58.13	58.58
PD (%)	41.74	39.43	40.57	40.12

Table 7. Detailed information of each domain: vegetation pattern, image acquisition dates, and class distributions

areas cannot be used to train a model for deforestation detection as the reference labels are unknown. To address this issue, a third label, *Past Deforestation* (*PD*), is assigned to areas in a label map that were labelled as *DF* at any point in time earlier than t_0. Such areas do not contain significant information for bi-temporal deforestation classification between epochs t_0 and t_1 and they are disregarded in the training procedure (INPE, 2021). Table 7 shows details of each domain including vegetation pattern, dimensions, acquisition dates of the images, and percentages of class distribution. Note that image data for some epochs are mosaics of two Sentinel-2 scenes. It is important to observe the high imbalance in class distribution across all domains, particularly for the Deforestation (*DF*) class, where percentages are lower than 2%. This significant class imbalance poses challenges for model training and highlights the necessity for techniques that can handle such disparities effectively.

4.2 Experimental setup

First, the images from the different sites, acquired at t_0 and t_1, were stacked along the spectral dimension. Then, the values in each pixel position of each band were normalized, in the range of $[-1, 1]$. Each image pair was divided into twenty (20) tiles, 40% (8 tiles), were used for training, 10% (2 tiles) for validation, and 50% (10 tiles) for testing. To compose the training and validation sets, patches with dimension 128×128 pixels, and stride equal to 64, were extracted from each image tile and used as the input to the encoder networks. In each experiment, we selected one $\mathcal{D}_S$ and three target $\mathcal{D}_T$.

As mentioned, all pixels labeled as past-deforestation according to PRODES were ignored both in training and in testing. Simple data augmentation operations were employed for the training patches: rotation (30°, 90°), and flipping (horizontal, vertical) transformations. During training, the Adam optimizer was used with learning rate γ and momentum β_1 equal to 0.0002 and 0.5, respectively. The batch size was 16. An early stopping strategy with patience equal to 10 was employed. The loss function weights λ_r and λ_c were both set to 0.5, and λ_{ds} was set to 1 in all experiments. For λ_{dp} we used two different values $[1, 0]$, as reported later.

In accordance with the PRODES methodology, we ignored pixels within a two pixel wide buffer at the inner and outer borders of all polygons identified as deforestation in the reference data. Those pixels were ignored for training, validation, and test. The same was done for all past-deforestation pixels, and areas (pixel clusters) smaller than 625 pixels (6,25 ha) for the Amazon regions and 100 pixels (1 ha) for the Cerrado region.

4.3 Evaluation of the cross-classification baselines

Train on	Test on PA	MT	RO	MA
PA	**77.51**	54.64	83.70	15.33
MT	73.51	**63.01**	83.24	32.23
RO	63.48	53.18	**84.88**	18.42
MA	62.73	52.16	78.43	**75.77**

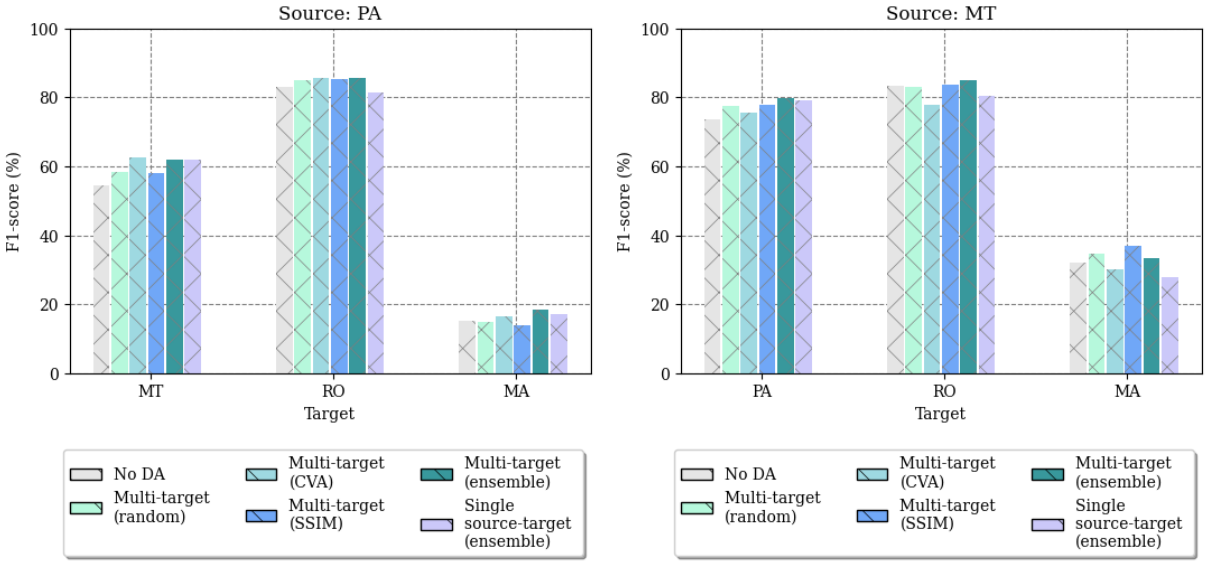
Table 8. F1-scores [%] for the class *DF* for intra- and cross-domain scenarios, without any adaptation procedure. Bold values along the diagonal represent the classification accuracy of the model trained and evaluated on the same domain, while values outside the diagonal report the evaluation results on different domains.

Before presenting the DA results, we report the cross-classification baselines (without DA). It is important to show the accuracies (in terms of F1-Scores) obtained by using solely the classification branch of the proposed DA model, trained and tested on the same domain, and tested on different domains. These results are presented in Table 8, which also can give an insight into the the domain discrepancies presented in each domain combination. Notice that the best scores for the baseline were obtained when the MA site was the source. However, when the MA site was a target domain, the scores were low, indicating that these were the most challenging among the tested scenarios. In particular, the MA region presents higher complexity in the forested areas due to the more prominent inter-annual variability (Soto Vega et al., 2021).

4.4 Evaluation of the multi-target adaptation

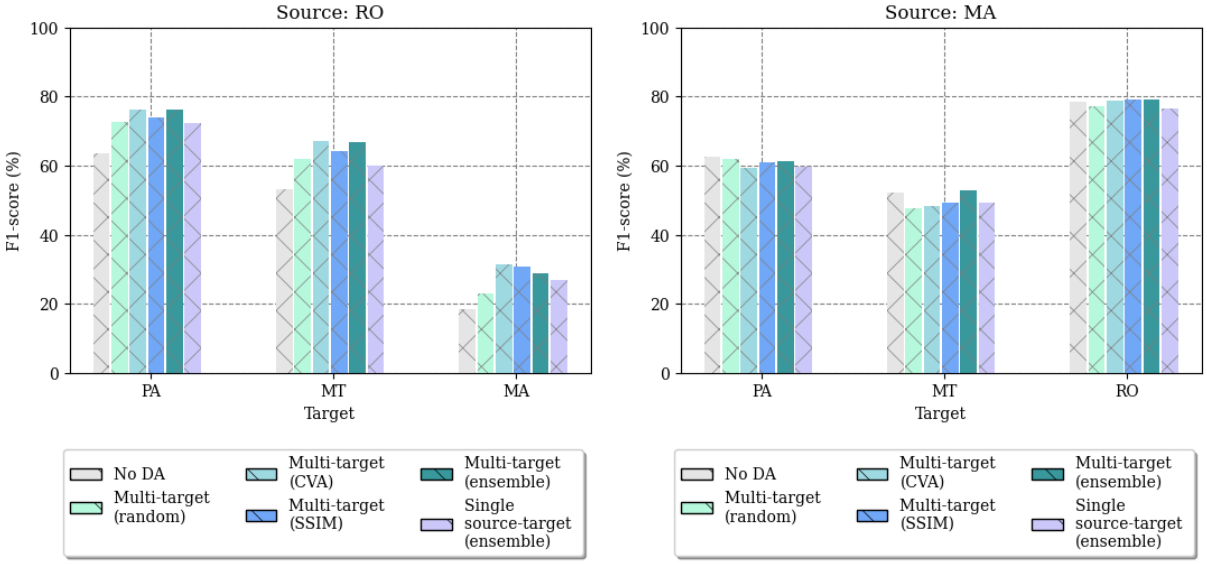
In the following experiments, we defined one source domain and three target domains. We again used the F1-score as a metric for classification accuracy. For selecting the training samples of the source domain, we used the references from PRODES, and for the target domains, we evaluated four different sample selection strategies:

- *random*, using 1000 patches from the training and validation image tiles randomly selected.
- *CVA*, using pseudo-labels determined with the CVA algorithm.
- *SSIM*, using pseudo-labels determined with the SSIM algorithm.
- *ensemble*, using the combination of the CVA and SSIM outcomes, following the consistency criterion described in Section 3.2.



(a) F1-scores when PA is defined as $\mathcal{D}_S$

(b) F1-scores when MT is defined as $\mathcal{D}_S$



(c) F1-scores when RO is defined as $\mathcal{D}_S$

(d) F1-scores when MA is defined as $\mathcal{D}_S$

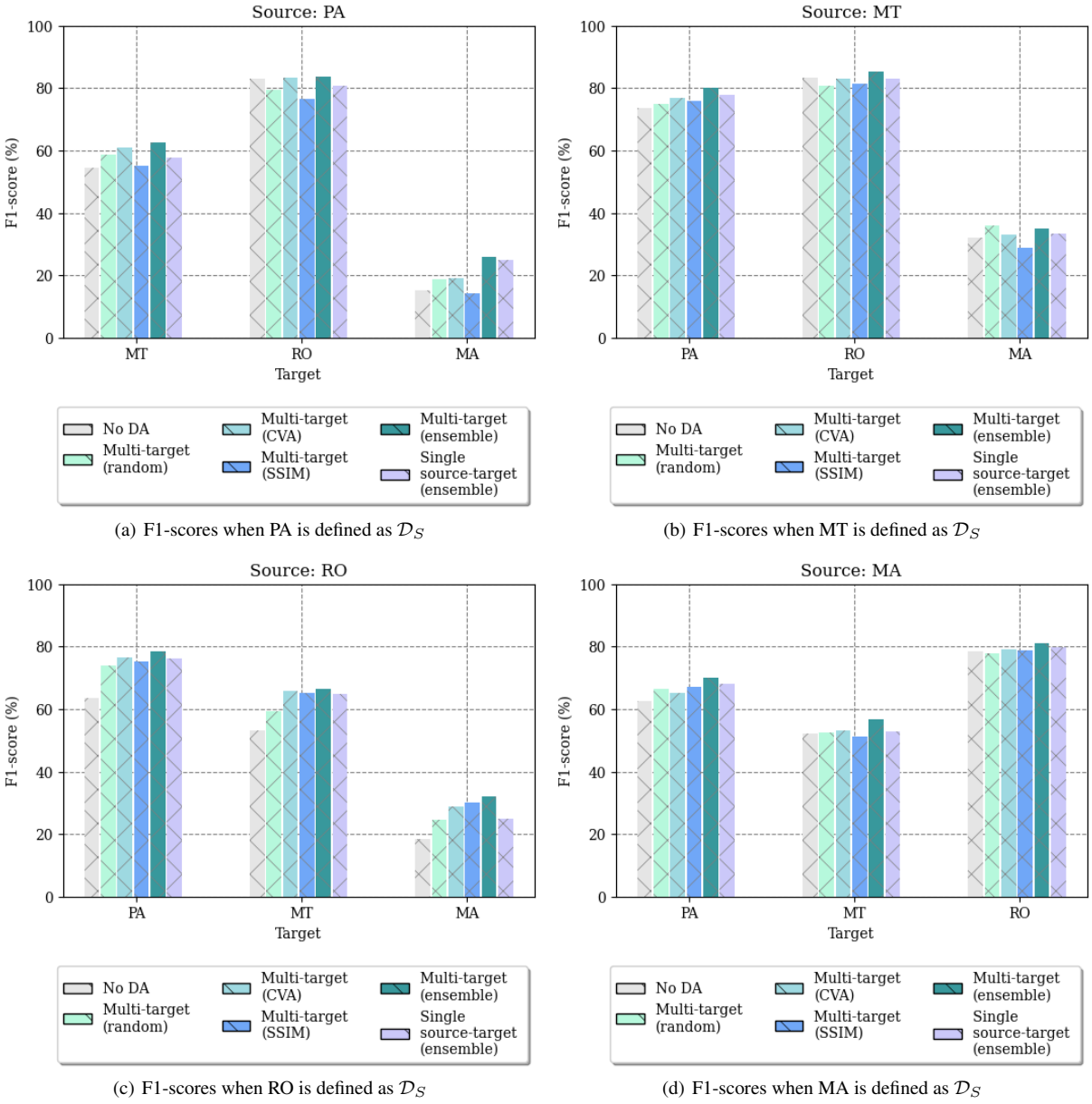
Figure 4. F1-scores [%] for the class *DF* on $\mathcal{D}_T$ using the multi-target DA method when $\lambda_{dp} = 1$.

For the last three strategies, we selected training and validation patches containing at least 2% of deforestation pixels. The same selection criterion was used for the source domain, but relying on ground-truth (PRODES) labels. Furthermore, the aforementioned scenarios were evaluated with two values of λ_{dp} in Equations 7 and 10, i.e., 1, and 0, in order to investigate the effect of considering private features in domain classifier during training.

Figures 4 and 5 present the results in terms of the average F1-scores after five runs using the DA method defining $\lambda_{dp} = 1$ and $\lambda_{dp} = 0$, respectively. Each figure reports the classification accuracy for the class *DF* for all domain combinations using one source and three target domains. The first bar corresponds to our baseline, where no adaptation procedure was applied (see Table 8). These results correspond to the worst scenarios since the training and testing sets come from different distributions. The next four greenish bars represent the results of the multi-DA method with the four unsupervised strategies to select the training samples from the target domains. Furthermore, for a more comprehensive analysis and comparison of the multi-target results, we evaluated the DA method within the single-source-single target framework using the pseudo-labels from the *ensemble* strategy , represented by the final purplish bar.

When we set λ_{dp} to 1, the domain classifier D weights equally the shared and private features in trying to discriminate between domains. It can be seen that when DA was employed, the scores improved in almost all cases, especially when using the *ensemble* strategy. Here, the DA method tends to yield the best overall performances across different scenarios, suggesting that a combination of techniques for selecting the target samples is beneficial for effective adaptation procedure. In addition, it is possible to notice that he multi-target DA outperformed the single-source results, meaning that the use of multiple targets improved the generalization capability of the model. However, in the domain combinations where MA was defined as a source

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-3-2024-293-2024 | © Author(s) 2024. CC BY 4.0 License.



(a) F1-scores when PA is defined as $\mathcal{D}_S$

(b) F1-scores when MT is defined as $\mathcal{D}_S$

(c) F1-scores when RO is defined as $\mathcal{D}_S$

(d) F1-scores when MA is defined as $\mathcal{D}_S$

Figure 5. F1-scores [%] for the class *DF* on $\mathcal{D}_T$ using the multi-target DA method when $\lambda_{dp} = 0$.

domain, the model produced lower classification metrics. Only with the *ensemble* strategy, metrics close to the baseline were achieved.

When we set λ_{dp} to 0, the domain classifier disregarded the private features and relied only on the shared features for domain classification. Notice that, in principle, the task of the domain classifier becomes more challenging, as private features should convey specific information from the different domains. Again we notice that we obtained the best results using the pseudo-label map from the *ensemble* strategy in all cases, again outperforming the method in the single-source-target scheme. Interestingly, in the most challenging scenarios in which MA was a target domain, the performance increased substantially compared to the previous settings. Those results seem to indicate that by not considering the private features in the domain classifier, the DA method found a better common representation shared between domains, improving the overall performance of the model.

Finally, Figure 6 shows the average deforestation probability maps generated with the DA method using the four different strategies to balance the target samples and varying the λ_{dp} values. Those maps correspond to the MA→PA domain combination. The first row shows the RBG composition of a subset of the target domain images; the respective ground-truth; and the baseline result. The last two rows show the outputs associated with the different λ_{dp} configurations and balancing strategies. Blue, red and black colors in the reference mask represent no-deforestation, deforestation and past-deforestation, respectively. The figure shows that when no DA was employed, the output was a noisy probability map with many false positives and negatives. The same can be said when the random target sample selection strategy was used. By reducing the λ_{dp} we obtained more confident outcomes and better defined polygons. In the best case, when λ_{dp} was set to zero and the ensemble strategy was employed to balance the target samples, the DA method presented the best segmentation output.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-3-2024-293-2024 | © Author(s) 2024. CC BY 4.0 License.

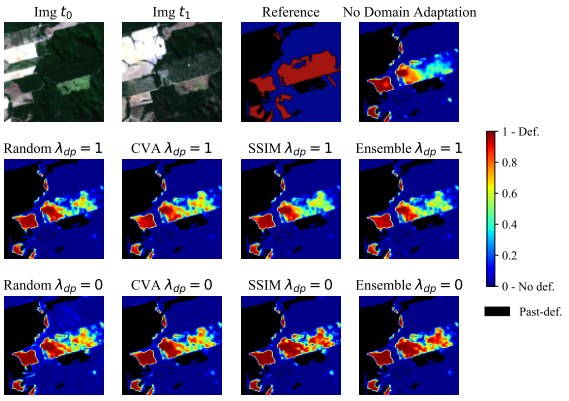


Figure 6. Deforestation probability maps of a snip from a test tile of MA→PA domain combination. The first row shows the RGB compositions of the target domain images; the reference map; and the baseline result. The next two rows represent the output from the four unsupervised strategies to balance the target samples with two different values of λ_{dp}. Blue and red colors represent lower and higher probability of belonging to the deforestation class, respectively. Black regions correspond to past-deforestation.

5. Conclusion

In this paper we introduced an unsupervised multi-target Domain Adaptation (DA) method based on fully convolutional models for deforestation detection. Two encoders map the input into two latent feature spaces; a shared one, which captures the common features across domains, and a private one, which learns domain-specific features. Three decoders account for the pixel-wise classification, the input reconstruction, and the discrimination between source and target domains, respectively. The whole scheme adopts adversarial learning for training. We also explored unsupervised algorithms to produce pseudo-labels for the target domains, aiming to mitigate the problems brought by the high class imbalance, characteristic of that application, to DA approaches. Additionally, we evaluated the performance of the proposed method and its variants in four test sites/domains with different forest covers and deforestation patterns. In most cases, the ensemble balancing scheme outperformed the baselines which did not employ any DA. Furthermore, experimental results demonstrated that domain discrimination improved when it relied only on shared features. This means that at least in our experiments, private features disturbed adversarial learning and prevented finding proper common feature representations. In future work, we plan to extend the method to a multi-source scheme, and to investigate the potential of the method in applications in which the label space is not binary, e.g., land cover and land use classification.

References

Andrade, R., Mota, G. L. A., da Costa, G. A. O. P., 2022. Deforestation Detection in the Amazon Using DeepLabv3+ Semantic Segmentation Model Variants. *Remote Sensing*, 14(19), 4694.

Cabral, B. F., Yanai, A. M., de Alencastro Graça, P. M. L., Escada, M. I. S., de Almeida, C. M., Fearnside, P. M., 2024. Amazon deforestation: A dangerous future indicated by patterns and trajectories in a hotspot of forest destruction in Brazil. *Journal of Environmental Management*, 354, 120354.

Chen, H., Wu, C., Du, B., Zhang, L., 2020. Dsdanet: Deep siamese domain adaptation convolutional neural network for cross-domain change detection. *arXiv preprint arXiv:2006.09225*.

Dalagnol, R., Wagner, F. H., Galvão, L. S., Braga, D., Osborn, F., da Conceição Bispo, P., Payne, M., Junior, C. S., Favrichon, S., Silgueiro, V. et al., 2023. Mapping tropical forest degradation with deep learning and Planet NICFI data. *Remote Sensing of Environment*, 298, 113798.

Deng, X., Yang, H. L., Makkar, N., Lunga, D., 2019. Large scale unsupervised domain adaptation of segmentation networks with adversarial learning. *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 4955–4958.

Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., March, M., Lempitsky, V., 2016. Domain-adversarial training of neural networks. *Journal of machine learning research*, 17(59), 1–35.

Gholami, B., Sahu, P., Rudovic, O., Bousmalis, K., Pavlovic, V., 2020. Unsupervised multi-target domain adaptation: An information theoretic approach. *IEEE Transactions on Image Processing*, 29, 3993–4002.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. *Advances in neural information processing systems*, 27.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote sensing of Environment*, 202, 18–27.

INPE, 2021. National institute for space research. general coordination of earth observation. monitoring program of the amazon and other biomes. deforestation - legal amazon -. `http://terrabrasilis.dpi.inpe.br`.

Isobe, T., Jia, X., Chen, S., He, J., Shi, Y., Liu, J., Lu, H., Wang, S., 2021. Multi-target domain adaptation with collaborative consistency learning. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8187–8196.

Li, K., Lu, J., Zuo, H., Zhang, G., 2023. Multidomain adaptation with sample and source distillation. *IEEE Transactions on Cybernetics*.

Li, Q., Gong, H., Dai, H., Li, C., He, Z., Wang, W., Feng, Y., Han, F., Tuniyazi, A., Li, H. et al., 2021. Unsupervised Hyperspectral Image Change Detection via Deep Learning Self-Generated Credible Labels. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 9012–9024.

Long, M., Cao, Y., Wang, J., Jordan, M., 2015. Learning transferable features with deep adaptation networks. *International conference on machine learning*, PMLR, 97–105.

Long, M., Zhu, H., Wang, J., Jordan, M. I., 2017. Deep transfer learning with joint adaptation networks. *International conference on machine learning*, PMLR, 2208–2217.

Malila, W. A., 1980. Change vector analysis: An approach for detecting forest changes with landsat. *LARS symposia*, 385.

Md Jelas, I., Zulkifley, M. A., Abdullah, M., Spraggon, M., 2024. Deforestation detection using deep learning-based semantic segmentation techniques: a systematic review. *Frontiers in Forests and Global Change*, 7, 1300060.

Nguyen-Meidine, L. T., Belal, A., Kiran, M., Dolz, J., Blais-Morin, L.-A., Granger, E., 2021. Unsupervised multi-target domain adaptation through knowledge distillation. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 1339–1347.

Noa, J., Soto Vega, P., Costa, G., Wittich, D., Feitosa, R., Rottensteiner, F., 2021. Adversarial Discriminative Domain Adaptation for Deforestation Detection. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 3, 151–158.

Ortega, M. X., Feitosa, R. Q., Bermudez, J. D., Happ, P. N., De Almeida, C. A., 2021. Comparison of optical and sar data for deforestation mapping in the amazon rainforest with fully convolutional networks. *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, IEEE, 3769–3772.

Peng, J., Huang, Y., Sun, W., Chen, N., Ning, Y., Du, Q., 2022. Domain adaptation in remote sensing image classification: A survey. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15, 9842–9859.

Ramachandran, N., Irvin, J., Sheng, H., Johnson-Yu, S., Story, K., Rustowicz, R., Ng, A. Y., Austin, K., 2024. Automatic deforestation driver attribution using deep learning on satellite imagery. *Global Environmental Change*, 86, 102843.

Saha, S., Solano-Correa, Y. T., Bovolo, F., Bruzzone, L., 2020. Unsupervised deep transfer learning-based change detection for HR multispectral images. *IEEE Geoscience and Remote Sensing Letters*, 18(5), 856–860.

Saporta, A., Vu, T.-H., Cord, M., Pérez, P., 2021. Multi-target adversarial frameworks for domain adaptation in semantic segmentation. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 9072–9081.

Soto Vega, P. J., da Costa, G. A. O. P., Feitosa, R. Q., Ortega, M. X., Bermudez, J. D., Turnes, J. N., 2022. Domain-Adversarial Neural Networks for Deforestation Detection in Tropical Forests. *IEEE Geoscience and Remote Sensing Letters*, 19, 1–5.

Soto Vega, P. J., da Costa, G. A. O. P., Feitosa, R. Q., Ortega, M. X., de Almeida, C. A., Heipke, C., Rottensteiner, F., 2021. An unsupervised domain adaptation approach for change detection and its application to deforestation mapping in tropical biomes. *ISPRS Journal of Photogrammetry and Remote Sensing*, 181, 113–128.

Sun, B., Feng, J., Saenko, K., 2016. Return of frustratingly easy domain adaptation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 30number 1.

Sun, B., Saenko, K., 2016. Deep coral: Correlation alignment for deep domain adaptation. *Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14*, Springer, 443–450.

Tasar, O., Giros, A., Tarabalka, Y., Alliez, P., Clerc, S., 2020. Daugnet: Unsupervised, multisource, multitarget, and life-long domain adaptation for semantic segmentation of satellite images. *IEEE Transactions on Geoscience and Remote Sensing*, 59(2), 1067–1081.

Torres, D. L., Turnes, J. N., Soto Vega, P. J., Feitosa, R. Q., Silva, D. E., Marcato Junior, J., Almeida, C., 2021. Deforestation detection with fully convolutional networks in the Amazon Forest from Landsat-8 and Sentinel-2 images. *Remote Sensing*, 13(24), 5084.

Tuia, D., Persello, C., Bruzzone, L., 2016. Domain adaptation for the classification of remote sensing data: An overview of recent advances. *IEEE geoscience and remote sensing magazine*, 4(2), 41–57.

Tzeng, E., Hoffman, J., Saenko, K., Darrell, T., 2017. Adversarial discriminative domain adaptation. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7167–7176.

Wang, Z., Bovik, A. C., Sheikh, H. R., Simoncelli, E. P., 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4), 600–612.

Wittich, D., Rottensteiner, F., 2021. Appearance based deep domain adaptation for the classification of aerial images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 180, 82–102.

Yang, X., Deng, C., Liu, T., Tao, D., 2020. Heterogeneous graph attention network for unsupervised multiple-target domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Zhu, Y., Zhuang, F., Wang, J., Ke, G., Chen, J., Bian, J., Xiong, H., He, Q., 2020. Deep subdomain adaptation network for image classification. *IEEE transactions on neural networks and learning systems*, 32(4), 1713–1722.

Zou, Y., Yu, Z., Kumar, B., Wang, J., 2018. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. *Proceedings of the European conference on computer vision (ECCV)*, 289–305.