DEFORESTATION DETECTION WITH WEAK SUPERVISED CONVOLUTIONAL NEURAL NETWORKS IN TROPICAL BIOMES

P. J. Soto[1,*], G. A. O. P. Costa[2], M. X. Ortega[3], J. D. Bermudez[3], R. Q. Feitosa[3]

[1] Ifremer, PDG-REM-EEP-LEP, F-29280 Plouzané, France
pjsotove@ifremer.fr
[2] Dept. of Informatics and Computer Science, Rio de Janeiro State University (UERJ), Brazil
gilson.costa@ime.uerj.br
[3] Dept. of Electrical Engineering, Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Brazil
mortega@aluno.puc-rio.br, (bermudez, raul)@ele.puc-rio.br

Commission III, WG III/7

KEY WORDS: Change detection, deep learning, domain adaptation, deforestation, weak supervision.

ABSTRACT:

Deep learning methods are known to demand large amounts of labeled samples for training. For remote sensing applications such as change detection, coping with that demand is expensive and time-consuming. This work aims at investigating a noisy-label-based weak supervised method in the context of a deforestation mapping application, characterized by a high class imbalance between the classes of interest, i.e., deforestation and no-deforestation. The study sites correspond to different regions in the Amazon and Brazilian Cerrado biomes. To mitigate the lack of ground-truth labeled training samples, we devised an unsupervised pseudo-labeling scheme based on the Change Vector Analysis technique. The experimental results indicate that the proposed approach can improve the accuracy of deforestation detection applications.

1. INTRODUCTION

Deforestation of natural forests is one of the largest sources of greenhouse gas emissions, it is responsible for the reduction of carbon storage and for invaluable loss of biodiversity. Consequently, global concern in this regard has increased considerably, and monitoring Earth's land cover changes in forested areas has become a priority for many authorities around the world. Nevertheless, efficient detection of such changes is a difficult task, which commonly relies on large amounts of remote sensing (RS) data and highly trained professionals in the visual analysis of such data.

Fortunately, the past decades have witnessed an increasing availability of RS data due to improvements in Earth observation systems. Additionally, several techniques for the analysis of single- and multi-date RS data have been already proposed. Specifically for change detection, algebra-based models, which rely on features computed with band-wise mathematical operations, include image difference methods, image ratio and Change Vector Analysis (CVA) (Malila, 1980), which has been the basis for more advanced approaches, e.g., (Thonfeld et al., 2016).

More recently, Deep Learning models have become the dominant trend in most image analysis application fields. Specifically for change-detection, (Daudt et al., 2018b) propose a method based on a Siamese Convolutional Neural Networks (S-CNN), in which co-registered RS images taken at different epochs are used as inputs for each network branch. Alternatively, Daudt et al. (2018a) employs the Early Fusion (EF) scheme, in which the images from the different epochs are concatenated, and used as input of a single network. Concerning deforestation detection, Ortega Adarme et al. (2020) employed an EF-based approach, while Andrade et al. (2020) adapted the Deeplabv3+ model for the task, also using the EF scheme. In addition, Torres et al. (2021) evaluated state-of-the-art deep network architectures for deforestation detection, using Landsat-8 and Sentinel-2 imagery.

Deep Learning models, however, are known to demand large amounts of labeled data for proper training, which is a problem for many RS applications because of the costs involved in the field surveys and visual interpretation required to produce the reference data. Moreover, changes in environmental conditions, geographical variability and different sensor properties typically makes it impossible to employ previously trained classifiers for new data without a significant decrease in classification accuracy. While this phenomenon represents a problem for any supervised classifier, considering the high demand for labeled data, it may seriously impair the operationalization of deep learning-based classification approaches in real-world applications (Vega et al., 2021).

Among the several ways to face the aforementioned problems, semi-supervised learning (Zhu, 2005), weak-supervised learning (Zhou, 2018) and transfer learning (Weiss et al., 2016) approaches have been proposed. Semi-supervised models rely on structural assumptions to leverage from unlabeled data. Transfer learning, relies on models trained on a different task, and weak-supervised learning aims at training classification models using noisy or low-quality labels.

Regarding change detection, a weak supervised approach has been employed in Khan et al. (2016), which uses image-level labels to predict changes between pairs of video frames. Daudt et al. (2021) proposed an edge preserving approach to guide the supervised learning procedure of a change detection classifier. As a specific setting of transfer learning, deep domain adapt-

* Corresponding author

ation (DA) based methods have been recently proposed to improve the generalization of change detection classifiers. Deng et al. (2019) applied DA for change detection in urban areas, and Saha et al. (2019, 2020) used VHR and HR multispectral images for the same purpose. Additionally, Soto et al. (2020), Noa et al. (2021) and Vega et al. (2021) have employed deep semi-supervised DA (Tuia et al., 2016) for the deforestation detection problem.

In this work, we evaluate a noisy-labeled-based, weak-supervised approach for deforestation detection. We aim at investigating if such an unsupervised pseudo-label scheme can be used to effectively mitigate the lack of labeled samples in the training of DL-based models. To verify the effectiveness of the proposed procedure, we conducted experiments considering three different areas associated with images from different sites in the Amazon and Brazilian Cerrado biomes.

The rest of this paper is organized as follows. Section 2 briefly describes the techniques which support the proposed procedure. A detailed description of the proposed method is the subject of Section 3. The experimental protocol is reported in Section 4. Section 5 shows the results obtained in the conducted experiments. Finally, Section 6 presents conclusions and indicates future research directions.

2. FUNDAMENTALS

Weak-supervised learning comprises a set of techniques that aim at mitigating the lack of labeled samples in the training of supervised learning models such as deep neural networks (DNN). Among the various weak-supervised approaches proposed thus far those designed for DNNs constitute the current state-of-the-art (Zhou, 2018).

According to Zhou (2018) weak supervision can be categorized as *incomplete supervision, inexact supervision*, or *inaccurate supervision*. In *incomplete supervision* a small amount of labeled data, insufficient to train a good learner is available, as well as abundant unlabeled data. In *inexact supervision* some supervision is given during training, but not as exact as desired. A typical scenario in that category is when coarse-grained label information is available as, for instance, image level label information in a semantic segmentation problem. Finally, *inaccurate supervision* refers to the scenario in which the supervision information is not always ground-truth; in other words, some label information may suffer from errors/noise. The typical example is learning with noisy-labels (Zhou, 2018). In this work we follow the *inaccurate supervision*-based weak-supervised approach. The noisy-labels are generated using an unsupervised procedure based on Change Vector Analysis (CVA).

Basically, the CVA (Malila, 1980) technique is used to compute the magnitude and direction of change between two coregistered multispectral images acquired at different epochs. Formally, let $\mathbf{x}_{t_0}(i,j)$ and $\mathbf{x}_{t_1}(i,j)$ represent a pixel's spectral vector at a given pixel location (i,j) in a pair of coregistered images acquired at t_0 and t_1, respectively. The magnitude M and direction ϕ of changes are computed as follows:

$$M = ||\mathbf{x}_{t_1}(i,j) - \mathbf{x}_{t_0}(i,j)||_2 \tag{1}$$

$$\cos\phi = \frac{\mathbf{x}_{t_1}(i,j) \cdot \mathbf{x}_{t_0}(i,j)}{||\mathbf{x}_{t_1}(i,j)||_2 ||\mathbf{x}_{t_0}(i,j)||_2} \tag{2}$$

where $||\cdot||_2$ denotes the L_2 norm.

CVA components have been used in several ways, for different change detection purposes. In this work, we binarize each component separately using thresholds calculated with the OTSU (Otsu, 1979) algorithm. Then, we consider the change transition between $\mathbf{x}_{t_1}(i,j)$ and $\mathbf{x}_{t_0}(i,j)$ as positive (with respect to the deforestation change class) if both the corresponding magnitude and phase are greater than the respective thresholds.

3. WEAKLY SUPERVISED LEARNING FOR DEFORESTATION DETECTION

In this work, we follow the Early Fusion approach for deforestation detection, in which RS images from a particular site, taken at different epochs (t_0 to t_1), are concatenated along the spectral dimension forming samples.

Let $x_q^T = [x_{q_{t_0}}^T : x_{q_{t_1}}^T]$ represent the $q-th$ coregistered, concatenated pair of multispectral remote sensing image patch of dimensions $w \times h \times b$ pixels acquired at t_0 and t_1, respectively, where w is the width, h is the height, both in pixels, b the number of spectral bands, and [:] is the concatenation operator. Let T denote the set of target domain samples $\{x_q^T\}$; and $\hat{y}_q^T$ represent the pseudo-label of x_q^T, taking a value from the set $\{0,1\}$, where 1 means *deforestation*, and 0 means *no-deforestation*. The pseudo-label $\hat{y}_q^T$ is predicted by a function $C(\cdot)$. We denote as $Q = |T|$ the number of image pairs in set T .

As described in the algorithms presented in this section, the learning process begins by selecting (weakly) balanced sets of training samples from the target domain. First, pseudo-labels $\hat{y}_q^T$ are created for the target domain samples x_q^T, with q =1,...,Q, (Step 1 of Algorithm 1) by the function $C(\cdot)$, which will be described later. Then, the actual balancing procedure (Step 2 of Algorithm 1) is performed by a function $U(\cdot)$, which employs traditional data augmentation operations, i.e., rotations and reflections, to increase the number of samples of the under-represented positive class (*deforestation*), and which randomly selects an equal number of samples of the over-represented negative class (*no-deforestation*). The resulting labeled set comprises N samples, with $N < Q$.

Since class labels of target domain are not available, function $U(\cdot)$ takes $\{\hat{y}_q^T\}_{q=1}^Q \leftarrow C(\{x_q^T\}_{q=1}^Q)$ as pseudo-labels for the target samples $\{x_q^T\}$. Function $C(\cdot)$ implements the pseudo-label prediction procedure explained in Algorithm 2. The procedure is based on the Change Vector Analysis (CVA) technique (Malila, 1980), and on the OTSU (Otsu, 1979) unsupervised thresholding algorithm.

First, CVA delivers the magnitude M_q and phase ϕ_q values of the difference vector associated with the center pixel location of each patch $x_q^T = [x_{q_{t_0}}^T : x_{q_{t_1}}^T]$. Then, the OTSU procedure computes the thresholds T_{h_M} and T_{h_ϕ} for each set of CVA components $\{M_q\}$ and $\{\phi_q\}$. Finally, a binarization procedure, in which the outcome of magnitude and phase are combined through the AND (&) logic operation, defines the pseudo-label set.

The balanced set of training target samples, represented as T^{tr}, is then used to train the deforestation detection model, EF-CNN with parameters $\{\theta_l\}$, until convergence (Step 3 of Algorithm 1). Once the EF-CNN model have been trained, it is used to classify test samples of the target domain.

Algorithm: Weakly Supervised Learning for Deforestation Detection

Inputs :

- $T = \left\{x_q^T\right\}_{q=1}^{Q}$ Unlabelled target samples.
- N Number of training samples.
- $\{\theta_l\}$ EF-CNN initial set of weights.
- $\{\mu_0, \alpha, \beta\}$ Hyper-parameters

Output: $\{\theta_l\}$ Final set of weights after training.

// Step 1: Compute pseudo-labels:
$\{\hat{y}_q^T\}_{q=1}^{Q} \leftarrow C(T)$
// Step 2: Select N balanced samples with the pseudo-labels:
$T^{tr} = \left\{x_n^T, \hat{y}_n^T\right\}_{n=1}^{N} \leftarrow U(T, N)$ where $N < Q$
// Step 3: Training the EF-CNN model:
while $j <$ # Epochs **do**
  $p \leftarrow \frac{j}{\#Epochs}$
  $\mu \leftarrow \frac{\mu_0}{(1+\alpha p)^{\beta}}$
  // Step 3.1: Building the batch:
  $T_{batch}^{tr} \leftarrow SelectBatch(T^{tr})$
  // Step 3.2: Forward processing:
  $\mathcal{L}_l \leftarrow \frac{1}{|T_{batch}^{tr}|} \sum_{(x_i, \hat{y}_i)} \hat{y}_i \log(G_l(x_i)) + (1 - \hat{y}_i) \log(1 - G_l(x_i))$
  // Step 3.2: Back-propagation:
  $\theta_l \leftarrow \theta_l - \mu \frac{\partial \mathcal{L}_l}{\partial \theta_l}$
  $j \leftarrow j + 1$
end while

Algorithm: Pseudo-labelling and sample balancing functions

Function $C(\{x_q\})$**:**
  // Extract patches‘ central pixel
  $\{\mathbf{x}_{q_{t_0}}\} \leftarrow \{x_{q_{t_0}}\}, \{\mathbf{x}_{q_{t_1}}\} \leftarrow \{x_{q_{t_1}}\}$
  // Compute magnitude and phase of central pixels
  $\{M_q\} \leftarrow \{||\mathbf{x}_{q_{t_1}} - \mathbf{x}_{q_{t_0}}||_2\}$
  $\{\phi_q\} \leftarrow \left\{\arccos \frac{\mathbf{x}_{q_{t_1}} \cdot \mathbf{x}_{q_{t_0}}}{||\mathbf{x}_{q_{t_1}}||_2 ||\mathbf{x}_{q_{t_0}}||_2}\right\}$,
  // Otsu threshold for magnitude and phase
  $T_{h_M} \leftarrow OTSU(\{M_q\}), T_{h_\phi} \leftarrow OTSU(\{\phi_q\})$
  // Generate pseudo-label
  $\{\hat{y}_q\} \leftarrow \left\{ \begin{array}{ll} 1 & \text{if } (\mathbf{M}_q > T_{h_M}) \wedge (\phi_q > T_{h_\phi}) \\ 0 & \text{otherwise} \end{array} \right\}$
  return $\{\hat{y}_q\}$;
Function $U(Samples, N)$**:**
  // Select N negative samples randomly
  $\{x_n^{neg}, y_n^{neg}\}_{n=1}^{N} \leftarrow Select(Samples, N)$
  // Select N positive samples randomly with augmentation
  $\{x_n^{pos}, y_n^{pos}\}_{n=1}^{N} \leftarrow SelectAndAugment(Samples, N)$
  return $\{(x_n^{neg}, y_n^{neg})\} \cup \{(x_n^{pos}, y_n^{pos})\}$;

4. EXPERIMENTAL ANALYSIS

The experiments reported in this section aimed at verifying the effectiveness of the proposed weak supervised approach in the context of deforestation detection in tropical forests.

The selected regions represent forests of different types, affected by various deforestation practices. Two sites are associated with regions in the Brazilian Legal Amazon (BLA) containing Dense Ombrophyll Forest areas, and Open Ombrophyll Forest areas. The third represents a site in the transition between the Brazilian Cerrado and the Amazon Rainforest, which contains Seasonal Deciduous and Semi-Deciduous Forest areas.

Additionally, the results of the proposed alternative were compared with those obtained by a EF-CNN classifier trained with labeled data from one of the above described sites, and evaluated on the remaining sites. Those results are considered as a baseline, and are used to measure the real gain brought with the weak supervised procedure.

4.1 Datasets

The study areas are located in two Brazilian Biomes, namely the Amazon rainforest and the Brazilian Cerrado (Savannah). Two of the sites are located in the Amazon, specifically in the Brazilian states of Rondônia (RO) and Pará (PA). The Cerrado study area is located in Maranhão (MA) state. Table 1 shows detailed information regarding geographical localization, dates, vegetation typology, and class distribution.

As for the forest typologies, the selected regions represent a gradation that goes from a very dense forest, with little variability of the canopy structure (PA), to a seasonal forest with high canopy variability (MA). The open rainforest (RO) is at an intermediate point between these two formations in terms of canopy variability. Figure 1 shows the location of the study areas, as well as a RGB composites of the most recent images of the corresponding image pairs. Table 1 shows the acquisition dates and the forest typologies (IBGE, 2012) in the respective domains, plus the number of pixels in each domain labeled as deforestation, no-deforestation, and previous deforestation. As can be seen in the table, the class distribution is highly imbalanced.

The images were acquired by the Landsat 8-OLI sensor with 30m resolution and 7 spectral bands. The images were acquired with minimum cloud cover at dates in the dry season, which occurs from late May to late August. They have the following dimensions: 2550×5120 pixels (RO); 1100×2600 pixels (PA); and 1700×1440 pixels (MA). All images underwent Level-1 data processing and were downloaded from the Earth Explorer web service from the United States Geological Survey (USGS)[1]. In all experiments, the individual image bands were normalized to zero mean and variance equal to one.

The ground truth for deforestation was produced by the PRODES Deforestation Mapping project of the Brazilian National Institute for Space Research (INPE), which, according to (Pinheiro Maurano et al., 2019; Parente et al., 2021), have accuracies of approximately 93% in both the Amazon and Cerrado biomes. The data is freely available at the Terrabrasilis website[2]. We observe that the images that compose the datasets for this study were also used in PRODES for deforestation mapping, for the respective sites and epochs (Almeida et al., 2021). As shown in Table 1, all images were acquired in the months of July and August, in which the acquisition conditions are optimum with respect to cloud coverage.

Figure 2 shows the deforestation reference for the respective image pairs (dark orange), which represent the deforestation that happened between the acquisition of the two images. The

[1] https://earthexplorer.usgs.gov/
[2] http://terrabrasilis.dpi.inpe.br/map/deforestation

Study areas	RO	PA	MA
Coordinates	09°36'51" S - 10°18'35" S 62°56'41" W - 64°20'51" W	03°08'21" S - 03°26'16" S 50°34'04" W - 51°16'12" W	04°44'52" S - 05°12'48" S 043°37'58" W - 044°01'23" W
Vegetation	Open Ombrophyll Forest	Dense Ombrophyll Forest	Seasonal Deciduous and Semi-Decidous Forest
Date 1	July 18, 2016	August 2, 2016	August 18, 2017
Date 2	July 21, 2017	July 20, 2017	August 21, 2018
deforestation pixels	225 635 (3%)	82 970 (3%)	71 265 (3%)
no-deforestation pixels	3 816 981 (29%)	1 867 929 (65%)	1 389 844 (57%)
previous deforestation pixels	9 013 384 (69%)	903 901 (32%)	986 891 (40%)

Table 1. Detailed information of each domain: image acquisition dates, coordinates, class distribution, and vegetation typology.

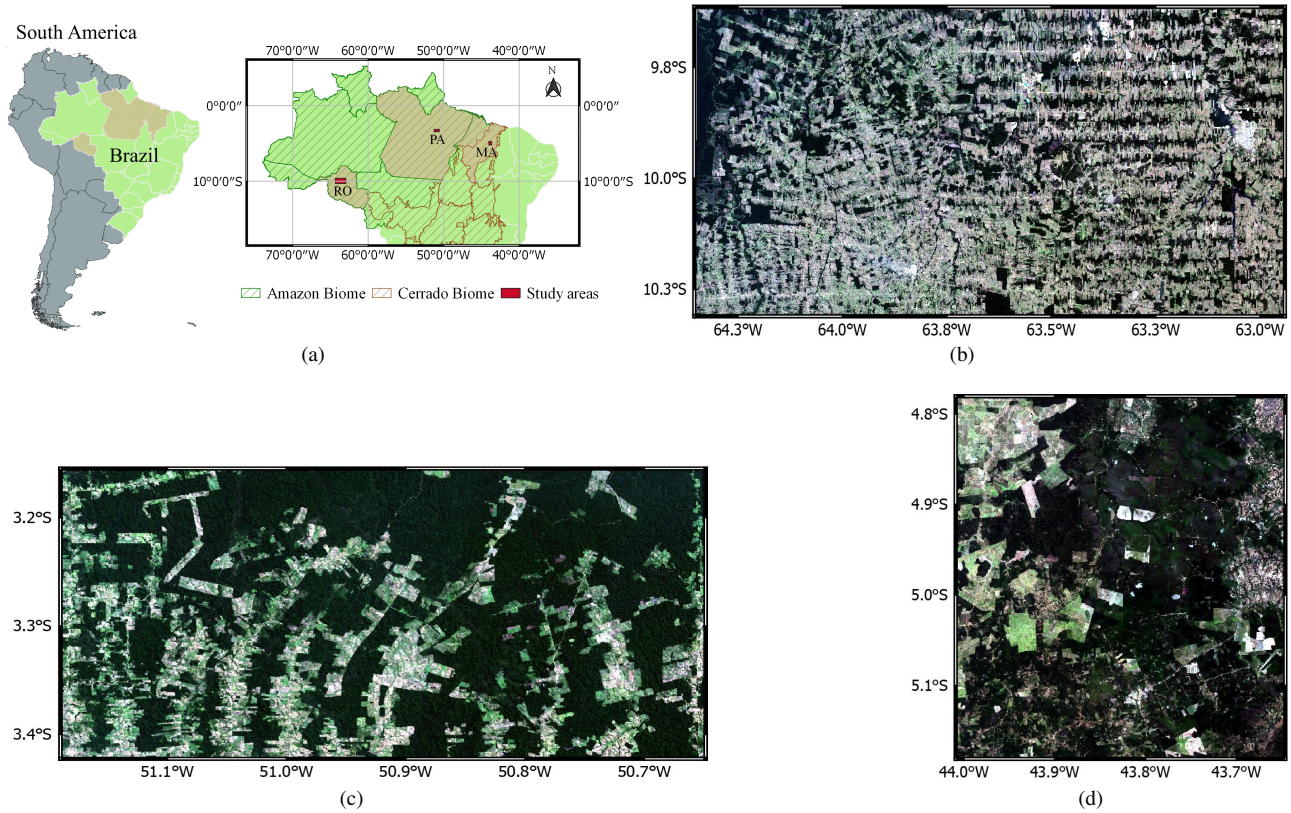

Figure 1. Visual representation and localization of each study area used in the experiments carried out in this work. (a) Geographical localization of the respective sites. True color composites of the images covering the sites, corresponding to the acquisition date 2017 (b) Rondônia (RO), (c) Pará (PA), and (d) Maranhão (MA). Figure taken from Vega et al. (2021).

figure also shows the accumulated deforestation (light gray), which occurred between 1988 and the acquisition year of the first image of the pair.

4.2 Classifier Training Setup

For the deforestation detection accuracy assessment, we used the Early Fusion (EF-CNN) classifier proposed in (Ortega Adarme et al., 2020). Additionally, also following (Ortega Adarme et al., 2020; Noa et al., 2021; Vega et al., 2021), the image space was divided into 100 tiles. Approximately 20% of the tiles were used to extract training samples/patches, 5% to extract validation patches, and the remaining 75% to extract the patches used for the evaluation of the classifier. Figure 2 shows the training, validation and test configuration tiles.

The input to EF-CNN is a tensor of size $29\times29\times14$. The respective image patches were extracted following an overlapping sliding windows procedure with a stride of 3, as in Ortega Adarme et al. (2020). During the training and evaluation steps, patches with central pixels having the following characteristics were avoided: belonging to polygons that have been deforested in previous years; lying inside a buffer around the deforestation reference polygons; and lying inside deforestation polygons smaller than 6.25 ha, which corresponds to 69 pixels.

Regarding the first and third conditions, we simply adopted the same procedure employed in the PRODES project. The second condition aims at avoiding the impact of inaccuracies in the ground truth produced by the rasterization process. Based on visual inspection of the correspondence between the ground truth and the deforested areas in the images, the width of the buffer was set to 6 pixels: 4 outside the polygons, and 2 inside them for RO while 2 pixels outside the polygons were adopted for PA and MA.

Data augmentation has been applied only to patches which the central pixel is labeled as deforestation (positive samples). A 90° rotation, as well as vertical and horizontal flips, were the data augmentation transformations. Additionally, only part of the no-deforestation patches (negative samples) in the training

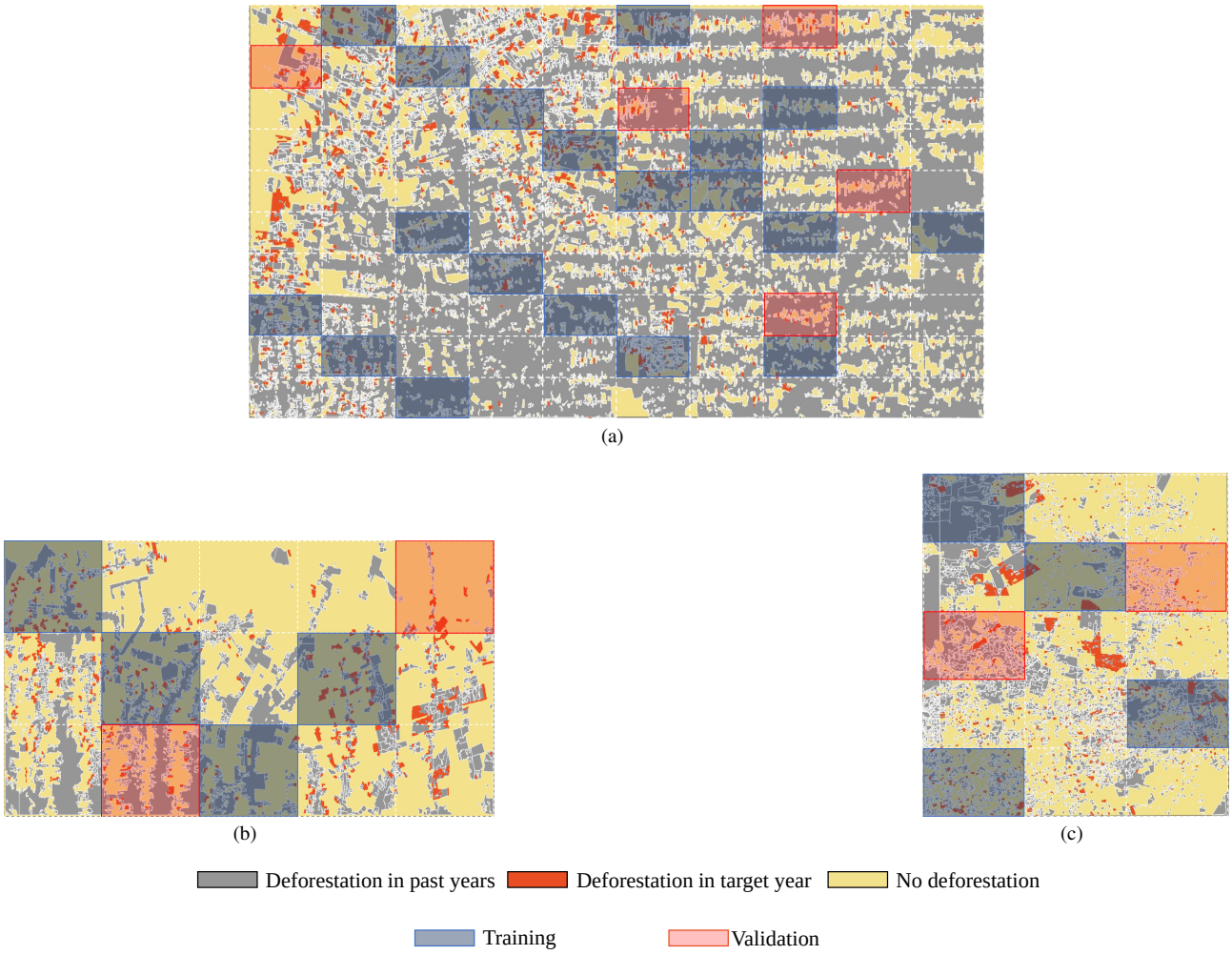

Figure 2. Distribution of image tiles for training, validation and testing in the respective study areas: (a) Rondônia (RO); (b) Pará (PA); and (c) Maranhão (MA). Please note that the tiles that are not shaded correspond to the ones selected for testing. The figure also shows the polygons associated with the deforestation that occurred during the image acquisition dates of the respective domains, the polygons associated with the deforestation that occurred prior to the date of the first image of the respective image pairs, and the areas labeled as not deforested. Figure taken from Vega et al. (2021).

and validation tiles were (randomly) selected (the same number as the positive samples after data augmentation), in order to balance the number of training samples per class. In the end, as in the patch-wise classification employed in (Ortega Adarme et al., 2020), the classification outcome of an input patch was assigned to its central pixel.

During training, the binary cross entropy objective function was minimized using the Momentum optimizer and learning rate decay. We set the initial learning rate μ_0 and momentum β_1 equal to 0.01 and 0.9 respectively. The batch size was 32, and the early stopping procedure was used to avoid overfitting. The patience parameter, which controls the number of epochs without improvements in the validation loss, was set to 10. The classifier was executed 10 times, each time with a different (random) initialization of the trainable parameters, and with a different set of randomly selected negative samples/patches.

4.3 Network Architecture

The network architecture of the EF-CNN classifier is described in Table 2 and Table 3. In those tables, the symbols identifies the operations for each layer: convolution (C), ReLU (R_e), MaxPooling (M_P), and Batch normalization (B_n). The number of filters, filters' dimensions and the convolution stride are indicated in parenthesis. In the case of MaxPooling, the values in parenthesis refer to the kernel dimension and stride. In the case of Reflection Padding, the value in parenthesis refer to the padding height/width.

Layer	Output shape
Input	(29, 29, 14)
$CR_e(128, 3, 1)$	(29, 29, 128)
$M_P(2, 2)$	(14, 14, 256)
$CR_e(256, 3, 1)$	(14, 14, 256)
$M_P(2, 2)$	(7, 7, 256)
$CR_e(512, 3, 1)$	(7, 7, 512)
ResNet block	(7, 7, 512)
ResNet block	(7, 7, 512)
ResNet block	(7, 7, 512)
Global Average Pooling	(512, 1)
Dropout	(512, 1)
Softmax	(2, 1)

Table 2. EF-CNN classifier architecture.

Layer
Reflection Padding (1)
$\mathrm{CB}_n R_e(64, 3, 1)$
Reflection Padding (1)
$\mathrm{CB}_n(64, 3, 1)$

Table 3. ResNet block architecture.

5. RESULTS

Table 4 shows the F1-scores obtained with the plain CVA+OTSU classification; with the EF-CNN model trained with ground truth labels; and with the EF-CNN model trained with pseudo-labels produced with the CVA+OTSU procedure, hereinafter denoted EF-CNN-PS. Note that the scores in Table 4 refer to the classifiers trained and tested with samples from the same domains.

Table 4 shows that the EF-CNN model trained with pseudo-labels produced with the CVA+OTSU procedure (EF-CNN-PS) obtained better results compared with the sole use of the CVA+OTSU procedure for classification. We believe that those results have to do with the capacity of the network to consider context in the classification of a single pixel.

Methods / Domains	PA	RO	MA
EF-CNN	83.2	67.0	85.5
EF-CNN-PS	56.1	66.2	73.0
CVA+OTSU	55.0	42.0	73.0

Table 4. F1-scores of the different approaches.

In order to enrich the analysis of the different classification approaches, we compared their results with the ones obtained with the EF-CNN model in a cross-domain classification setup.

Figures 3 and 4 show the accuracies obtained with the different approaches under single and cross-domain scenarios, in terms of F1-Score and mean average precision (mAP), respectively. Each group of bars represents a scenario: S refers to the source domain, and T to the target domain. The blue bars represent the accuracy of the EF-CNN model trained with ground truth labels from the target domain and tested on target domain samples. The orange bars represent the accuracy of the EF-CNN model trained with pseudo-labels from the target domain and tested on target domain samples (EF-CNN-PS). The yellow bars correspond to the EF-CNN model trained ground truth labels from a source domain and evaluated with samples from a target domain (cross-domain classification).Finally, the gray bar shows the results of the plain CVA+OTSU method for the corresponding target domain. Figure 4 does not show mAP accuracy values for CVA+OTSU because its output is not a score value a threshold can be applied to, which prevents the mAP computation.

As expected, the EF-CNN model trained and tested with ground truth labels from the same domain (blue bars) produced the best F1-Score and mAP values. Surprisingly, when used in the cross-domain classification (yellow bars) for the combination [Train(MA)|Test(PA)], the EF-CNN classifier delivered an accuracy in terms of mAP that is better than the one obtained in the intra-domain case, a result that needs further investigation.

It is interesting to see that the intra-domain classifications with the the EF-CNN model trained with pseudo-labels (EF-CNN-PS, orange bars) delivered results that are not so distant from the

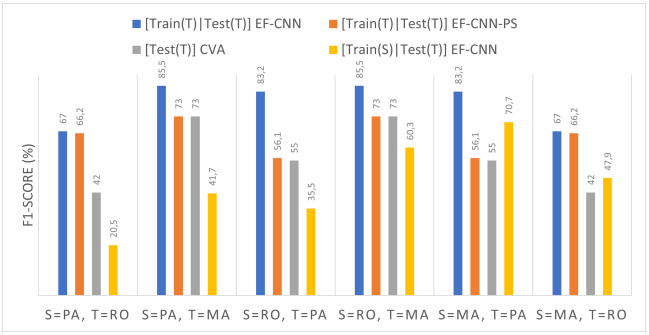

Figure 3. F1-Score of the different techniques.

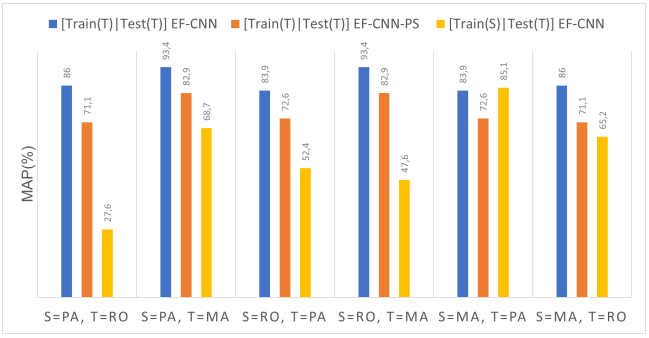

Figure 4. Mean Average Precision (mAP) of the different techniques.

ones obtained with EF-CNN model trained with ground truth labels.

Additionally, in all but one cross-domain scenario ([Train(MA)|Test(PA)]), the EF-CNN-PS approach was significantly superior to the EF-CNN model trained with ground truth labels from one (source) domain and tested with samples of another (target) domain. We believe those results indicate the appropriateness of using the pseudo-labels produced with the CVA+OTSU procedure to aid domain adaptation approaches.

6. CONCLUSIONS

In this work we proposed a noisy-label-based, weak supervised approach for change detection, applied to deforestation detection in tropical biomes. The weak supervision relies on an unsupervised pseudo-labeling procedure based on Change Vector Analysis, and aims at mitigating the lack, or low availability, of labeled samples for training deep learning-based supervised models.

We compared the performance of the proposed weak supervised approach with those obtained with the same underlying classification model (EF-CNN) in cross-domain combination scenarios. The results showed that in intra-domain classification scenarios, the proposed approach delivered results not so distant from the ones obtained with with ground truth labels.

Also, in most cross-domain scenarios, the proposed approach was significantly superior to the EF-CNN model trained with ground truth labels from a source domain and tested with samples of a target domain.

Based on the cross-domain classification results, we hypothesize that the pseudo-labels produced with the devised procedure

can be used to improve the accuracy of unsupervised domain adaptation approaches.

ACKNOWLEDGEMENTS

The authors would like to thank the founding provided by CAPES, CNPq, FAPERJ, IFREMER and NVIDIA corporation.

References

Almeida, C. A., Maurano, L. E. P., Valeriano, D. D. M., Camara, G., Vinhas, L., Gomes, A. R., Monteiro, A. M. V., Souza, A. A. A., Renno, C. D., Silva, D. E., Adami, M., Escada, M. I. S., Mota, M., Kampel, S. A., 2021. Methodology for Forest Monitoring used in PRODES and DETER projects. Technical report, INPE, São José dos Campos.

Andrade, R., Costa, G., Mota, G., Ortega, M., Feitosa, R., Soto, P., Heipke, C., 2020. EVALUATION OF SEMANTIC SEGMENTATION METHODS FOR DEFORESTATION DETECTION IN THE AMAZON. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, 1497–1505.

Daudt, R. C., Le Saux, B., Boulch, A., 2018a. Fully convolutional siamese networks for change detection. *2018 25th IEEE International Conference on Image Processing (ICIP)*, IEEE, 4063–4067.

Daudt, R. C., Le Saux, B., Boulch, A., Gousseau, Y., 2018b. Urban change detection for multispectral earth observation using convolutional neural networks. *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 2115–2118.

Daudt, R. C., Le Saux, B., Boulch, A., Gousseau, Y., 2021. Weakly supervised change detection using guided anisotropic diffusion. *Machine Learning*, 1–27.

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.

IBGE, 2012. Manual Técnico da Vegetação Brasileira.

Khan, S. H., He, X., Porikli, F., Bennamoun, M., Sohel, F., Togneri, R., 2016. Learning deep structured network for weakly supervised change detection. *arXiv preprint arXiv:1606.02009*.

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

Noa, J., Soto, 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 Adarme, M., Queiroz Feitosa, R., Nigri Happ, P., Aparecido De Almeida, C., Rodrigues Gomes, A., 2020. Evaluation of Deep Learning Techniques for Deforestation Detection in the Brazilian Amazon and Cerrado Biomes From Remote Sensing Imagery. *Remote Sensing*, 12(6), 910.

Otsu, N., 1979. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1), 62–66.

Parente, L., Nogueira, S., Baumann, L., Almeida, C., Maurano, L., Affonso, A. G., Ferreira, L., 2021. Quality assessment of the PRODES Cerrado deforestation data. *Remote Sensing Applications: Society and Environment*, 21, 100444.

Pinheiro Maurano, L. E., Sobral Escada, M. I., Daleles Renno, C., 2019. Padrões espaciais de desmatamento e a estimativa da exatidão dos mapas do PRODES para Amazônia Legal Brasileira. *Ciência Florestal (01039954)*, 29(4).

Saha, S., Bovolo, F., Bruzzone, L., 2019. Unsupervised deep change vector analysis for multiple-change detection in VHR images. *IEEE Transactions on Geoscience and Remote Sensing*, 57(6), 3677–3693.

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*.

Soto, P., Costa, G., Feitosa, R., Happ, P., Ortega, M., Noa, J., Almeida, C., Heipke, C., 2020. DOMAIN ADAPTATION WITH CYCLEGAN FOR CHANGE DETECTION IN THE AMAZON FOREST. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, 1635–1643.

Thonfeld, F., Feilhauer, H., Braun, M., Menz, G., 2016. Robust Change Vector Analysis (RCVA) for multi-sensor very high resolution optical satellite data. *International Journal of Applied Earth Observation and Geoinformation*, 50, 131–140.

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). https://www.mdpi.com/2072-4292/13/24/5084.

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.

Vega, P. J. S., da Costa, G. A. O. P., Feitosa, R. Q., Adarme, M. X. O., 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.

Weiss, K., Khoshgoftaar, T. M., Wang, D., 2016. A survey of transfer learning. *Journal of Big Data*, 3(1), 9.

Zhou, Z.-H., 2018. A brief introduction to weakly supervised learning. *National science review*, 5(1), 44–53.

Zhu, X. J., 2005. Semi-supervised learning literature survey.