ADVERSARIAL DISCRIMINATIVE DOMAIN ADAPTATION FOR DEFORESTATION DETECTION

J. Noa[1,*], P. J. Soto[1], G. A. O. P. Costa[2], Dennis Wittich[3], R. Q. Feitosa[1], Franz Rottensteiner[3]

[1] Dept. of Electrical Engineering, Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Brazil
(jnoat92, psoto, raul)@ele.puc-rio.br
[2] Dept. of Informatics and Computer Science, Rio de Janeiro State University (UERJ), Brazil
gilson.costa@ime.uerj.br
[3] Institute of Photogrammetry and GeoInformation, Leibniz Universität Hannover (LUH), Germany
(wittich, rottensteiner)@ipi.uni-hannover.de

Commission III, WG III/7

KEY WORDS: Change detection, deep learning, domain adaptation, deforestation, margin-based regularization.

ABSTRACT:

Although very efficient in a number of application fields, deep learning based models are known to demand large amounts of labeled data for training. Particularly for remote sensing applications, responding to that demand is generally expensive and time consuming. Moreover, supervised training methods tend to perform poorly when they are tested with a set of samples that does not match the general characteristics of the training set. Domain adaptation methods can be used to mitigate those problems, especially in applications where labeled data is only available for a particular region or epoch, i.e., for a source domain, but not for a target domain on which the model should be tested. In this work we introduce a domain adaptation approach based on representation matching for the deforestation detection task. The approach follows the Adversarial Discriminative Domain Adaptation (ADDA) framework, and we introduce a margin-based regularization constraint in the learning process that promotes a better convergence of the model parameters during training. The approach is evaluated using three different domains, which represent sites in different forest biomes. The experimental results show that the approach is successful in the adaptation of most of the domain combination scenarios, usually with considerable gains in relation to the baselines.

1. INTRODUCTION

Deforestation is an important problem, responsible for the reduction of carbon storage, greenhouse gas emissions, and other serious environmental issues, such as biodiversity losses and climate change (De Sy et al., 2015). Deforestation monitoring has become, therefore, a priority for many public authorities and institutions around the world.

In this respect, many initiatives based on remote sensing (RS) data have been developed for the periodic updating of deforestation maps. A notable example is the Deforestation Monitoring Program (PRODES) developed by the Brazilian National Institute for Space Research (INPE), which produces annual reports about deforestation of native vegetation in Brazilian forest biomes based on the analysis of Landsat images (Valeriano et al., 2004). However, due to the high level of accuracy expected for the official information provided by PRODES and similar initiatives to different stakeholders, such projects rely mostly on visual interpretation and manual operations. There is, therefore, a demand for automatic methods that can support deforestation monitoring applications in ways that can further improve the accuracies obtained and, at the same time, diminish the need for human intervention, so as to shorten their response times (Andrade et al., 2020).

With the rise of deep learning technology, highly accurate models have been developed for image interpretation applications, but under the premise of using a large amount of labeled training data (Krizhevsky et al., 2017). In RS applications, however,

the annotation process depends on expensive field campaigns and human experts, which limits the use of supervised classification methods. Those applications can, therefore, highly benefit from classifiers that are able to properly generalize in the presence of samples with characteristics not seen during training.

Unfortunately, supervised training methods, particularly deep learning models, tend to perform poorly when tested with a set of samples that does not match closely the general characteristics of the training set. Regularization techniques (Kukačka et al., 2017) may help to improve the generalization capacity of supervised classifiers, but only if the test data is already similar to the training data. Supervised transfer learning (Huh et al., 2016) can also help as they provide ways to learn from a few annotated test samples. However, those techniques are of little help when no labeled samples are available for the test set.

Considering the training and test sets as different domains, we can say that the performance of the classification models deteriorates depending on the respective *domain shift* (Wang and Deng, 2018). In RS applications the shift between training (source) and test (target) domains may be due to different acquisition conditions, e.g., data acquired at different epochs or using different sensors; or to data collected from different geographical areas (Soto et al., 2020).

To mitigate the problem, a number of Domain Adaptation (DA) techniques have been proposed (Tuia et al., 2016). Most of such techniques attempt either to align the features extracted from the images of both domains, or to adapt the appearance of those images (Li et al., 2020; Tasar et al., 2020). For example, in

(Deng et al., 2018) a Cycle-Consistent Generative Adversarial Network (CycleGAN) (Zhu et al., 2017) is used to translate images from one domain to the other, preserving the content but transforming their visual appearance. With CycleGANs, two protocols can be adopted for DA. One can use the classifier trained with the labeled source images to classify the adapted target images (i.e., translated to the source domain) or to translate in the opposite direction for training the classifier using the adapted source images (i.e., translated to the target domain) and corresponding labels. The major drawback of this approach is that the CycleGAN tends to generate artifacts during the translation, which may seriously hinder classification accuracy. To overcome that problem, Tasar et al. (2020) propose the so-called ColorMapGAN, another image translation method, which tries to find mappings for all color intensities in the images from the source and target domains. This method, however, has limitations such as noisy outcomes due to the exclusive use of local information, and a computational load that grows exponentially with the spectral and radiometric resolutions, making it virtually impossible to work with Landsat images, for instance.

Techniques based on feature alignment, also denoted as representation matching methods, seek to learn domain-invariant features from both source and target domains. Some methods (Sun and Saenko, 2016; Long et al., 2017) focus on minimizing a divergence metric between source and target features; others (Ganin et al., 2016; Tzeng et al., 2017; Li et al., 2020) try to learn to directly generate domain-invariant features through adversarial training (Goodfellow et al., 2014). The Domain Adversarial Neural Network (DANN) (Ganin et al., 2016), for example, learns a symmetric mapping of both domains to a common feature space using a single feature extractor. Alternatively, the Adversarial Discriminative Domain Adaptation (ADDA) (Tzeng et al., 2017) approach relies on an asymmetric mapping of the domain features using distinct feature generators. Based on the assumption of shared information between source and target representations, Huang et al. (2018) incorporated a regularization term in the ADDA loss function to ensure that the parameters of the source and target generator models do not deviate too much, generally leading to a better adaptation.

Besides the ColorMapGAN (Tasar et al., 2020), some other works have employed the aforementioned DA approaches in the context of semantic segmentation of RS images. For instance, Wittich and Rottensteiner (2019) adapted the ADDA strategy for pixel-wise classification of aerial images and height maps of different urban areas. Specifically for deforestation detection, Soto et al. (2020) evaluated the adaptation of images of the same location acquired at different epochs using a CycleGAN-based approach. Despite the promising results, the latter method is still affected by artifacts generated in the appearance adaptation process.

In this work, we introduce a DA approach based on representation matching for the task of deforestation detection. The proposed approach follows the Adversarial Discriminative Domain Adaptation (ADDA) framework, and includes a margin-based regularization term in the loss function used for training. The new regularization term was devised to better control the divergence between the source and target feature extractor components of the framework, as described in sections 2 and 3. While the proposed regularization term was devised in the context of a specific application, we believe it can help to better tune ADDA-based models for other RS applications. Furthermore, we evaluated the approach using three different domains,

represented by images of different sites in two Brazilian forest biomes, namely, the Amazon and the Brazilian Cerrado.

The remainder of this paper is organized as follows. Section 2 explains the ADDA framework. Section 3 describes and presents the intuition behind the proposed regularization term. Section 4 describes the experimental protocol adopted in the evaluation. Section 5 presents the obtained results, and a discussion about those results. Finally, in Section 6 we present conclusions and directions for further research.

2. ADVERSARIAL DISCRIMINATIVE DOMAIN ADAPTATION

ADDA was originally proposed in (Tzeng et al., 2017), aiming at improving the performance of scene classifiers (for image labeling) trained with labeled images from a source domain, and afterwards applied to images of a target domain, doing without any labeled target samples. In (Huang et al., 2018) the method was extended for semantic segmentation (pixel-wise classification), and in (Wittich and Rottensteiner, 2019) it was successfully employed in semantic segmentation of remote sensing images.

The ADDA domain adaptation strategy enables learning to map features extracted from the images of the source and target domains to a common (source) space, but so that the mapped features remain category discriminative. Let $\{x_n^S, y_n^S\}_{n=1}^N \in S$ and $\{x_m^T\}_{m=1}^M \in T$ be two sets of images belonging to the source (S) and target (T) domains, respectively. We denote as y_n^S the label that corresponds to sample x_n^S, and as N and M the number of images in the two sets, respectively.

First, a deep neural network model is trained to classify the source domain images x^S using their corresponding labels y^S (see Figure 1a). Such a model is composed of a feature extractor E^S with parameter values θ_{E^S} and a label predictor P (a dense label predictor in our case). Next, another feature extractor network E^T, having the same architecture as the source feature extractor E^S, is initialized using the pre-trained parameter values θ_{E^S}, i.e., θ_{E^T} is initialized with the θ_{E^S} values. Finally, using an adversarial training procedure that relies on a domain discriminator D, E^T is trained to produce features for the target domain images that can be properly classified by the pre-trained label predictor P.

The DA procedure is represented in Figure 1b. E^S extracts features from source images, and E^T does so from the target images. During the training process, the parameter values θ_{E^S} are frozen, while E^T learns to produce features that D cannot distinguish from the ones extracted from the source domain. In order to achieve this goal, an adversarial training (Goodfellow et al., 2014) is performed using the loss function described in Equation 1, which is minimized and maximized by updates of the parameters in the feature extractor E^T and the discriminator D, respectively.

$$\begin{aligned}\mathcal{L}(E^T, D) = {} & \mathbb{E}_S[\log D(E^S(x^S))] \\ & + \mathbb{E}_T[\log(1 - D(E^T(x^T)))] + \lambda\mathcal{L}_{reg},\end{aligned} \quad (1)$$

According to Huang et al. (2018), if the domains are similar enough for DA to be feasible, the parameter values θ_{E^T} should
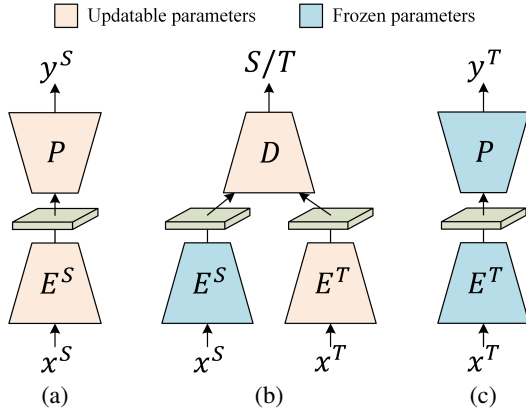
Figure 1. Steps for Adversarial Discriminative Domain Adaptation. (a) Training of the source classifier. (b) Adaptation of the target features. (c) Adapted model able to classify target samples.

not be very different from θ_{E^S}. Therefore, the regularization term $\mathcal{L}_{reg}$ in Equation 1, weighted by the hyperparameter λ, is meant to prevent the drift of the target parameter values θ_{E^T}, keeping them similar to θ_{E^S} according to a distance function such as the L1 norm (see Equation 2).

$$\mathcal{L}_{reg} = \|\theta_{E^S} - \theta_{E^T}\|_1 \qquad (2)$$

Finally, to classify the target samples, a classification model is built using the feature extractor E^T and the label predictor P (see Figure 1c).

3. MARGIN-BASED L1-REGULARIZATION

During the development of this research, we found out that the regularization loss term $\mathcal{L}_{reg}$ in Equation 2 was too restrictive to achieve an adequate adaptation. We recall that the weight λ can be tuned for different applications, which consider domains with different characteristics. In principle, for larger domain shifts, lower λ values should be selected, so that the target feature extractor E^T can have more room to learn proper mappings for the target features to the source feature space.

Figure 2 shows the L1 distances between θ_{E^S} and θ_{E^T} as a function of the training iterations for different values of λ. Note that all curves in the figure begin at a value of zero because θ_{E^T} is initialized with θ_{E^S}. This figure is based on one of the DA scenarios tested in the experiments (source: RO, target: MA; cf. Section 4.1 for the definition of the data).

As it can be observed in Figure 2a, higher λ values better prevent the parameter values θ_{E^T} from drifting away from θ_{E^S} in terms of the L1 distance. On the other hand, lower values of λ introduce a considerable amount of instability in the learning process, which may prevent a successful adaptation. To tackle the problem, we included a margin m in the regularization term, which defines a minimum desirable distance between θ_{E^S} and θ_{E^T}. The proposed regularization term is given by Equation 3. Note that for L1 distance values lower than m, the regularization loss will be zero and, thus, will not influence the target feature extractor E^T parameter updates.

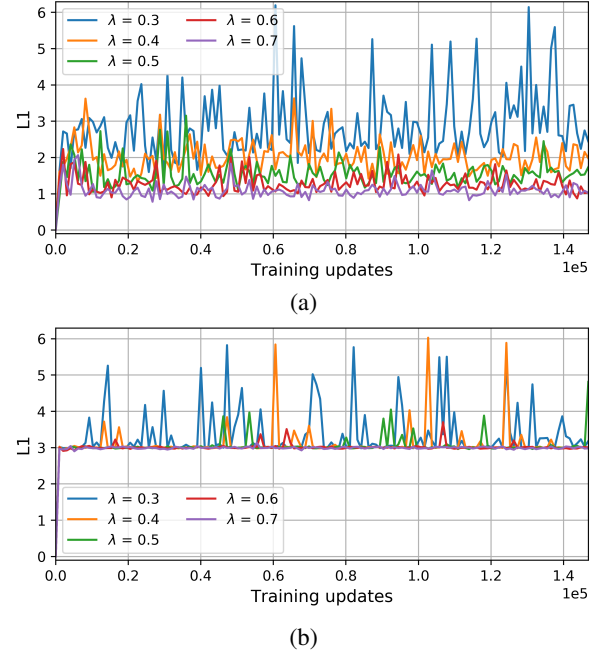$$\mathcal{L}_{reg}^m = \max(0, \mathcal{L}_{reg} - m) \qquad (3)$$

Figure 2. Record of $\|\theta_{E^S} - \theta_{E^T}\|_1$ over 140,000 iterations. (a) Using $\mathcal{L}_{reg}$ for regularization. (b) Using $\mathcal{L}_{reg}^m$ for regularization, with $m = 3$.

Figure 2b shows the L1 distance values obtained using the new regularization term with an arbitrary margin $m = 3$, for the same λ values used in Figure 2a. One can see that the new term enables to set θ_{E^S} and θ_{E^T} parameter values apart, without the need to decrease λ values too much, and thus be subject to unstable weight drifts. Therefore, this modification makes it easier for E^T to learn a proper mapping for the target samples to the common feature space, favoring the convergence of the adaptation process. It is noteworthy that m represents an additional hyperparameter to be tuned.

4. EXPERIMENTS

The proposed DA approach was evaluated using several scenarios in the context of deforestation detection in which different source and target domains are considered. The code developed in this research is publicly available[1].

4.1 Datasets

In the experiments, we considered three remote sensing datasets with particular characteristics as the domains of interest. Each dataset represents pairs of images acquired in consecutive years, covering forested areas located in different Brazilian states: Pará (PA), Rondônia (RO), and Maranhão (MA). The PA and RO sites are located in the Amazon biome, and cover areas characterized as Dense Ombrophyll Forest and Open Ombrophyll Forest, respectively. The MA site is located in a transition zone between the Amazon and the Brazilian Cerrado biomes, covering a Seasonal Deciduous and Semi-Deciduous Forest area. The forest canopy variability in the MA site is the highest, and the lowest in PA. Additionally, the deforestation footprints in MA are more marked, as clearcutting is the usual deforestation practice. Conversely, selective logging is more common

---

[1] https://github.com/jnoat92/Adversarial-Domain-Adaptation-for-Change-Detection

in PA, while deforestation practices are more diverse in RO (Muchagata and Brown, 2003; Marris, 2005).

All images were downloaded from the Earth Explorer web service from the United States Geological Survey (USGS)[2], and were produced with the Landsat-8 OLI sensor system. The images have 7 spectral bands with 30m spatial resolution. The respective deforestation references were provided by the PRODES program, from the Brazilian National Institute of Space Research (INPE) and are freely available at the Terrabrasilis website[3]. The images of the PA site were acquired in August 2016 and July 2017. The RO site images are from July 2016 and July 2017, and the MA images are from August 2017 and August 2018. The selected images are the same ones used in PRODES for deforestation mapping in the respective sites/epochs. Figure 3 shows the image (RGB bands) of the second date for each dataset and the corresponding geographical extents. Table 1 indicates the coordinates of each site and the sizes of the respective images.

4.2 Experimental Setup

The proposed approach was evaluated on six scenarios. In each of these scenarios, one dataset served as the source domain, and one of the two remaining datasets as the target domain. We first trained a pixel-wise classifier using the labeled images from the source domain. Next, we performed the ADDA-based domain adaptation scheme on the target domain and assessed the respective prediction. For each of the six scenarios, we compared the two ADDA variants, i.e., training with the alternative loss terms $\mathcal{L}_{reg}$ and $\mathcal{L}_{reg}^{m}$, in order to evaluate the influence of the proposed regularization term on the results.

The experiments were run five times, and we report the accuracy metrics (mean average precision and F1-score) using the mean of the predicted probabilities. The images of each dataset were split into three disjoint sets of tiles, of which approximately 20% were used for training, 5% for validation and 75% for testing. The source domain training tiles were used in both stages of the adaptation approach, i.e., in the training of the source domain classifier and in the feature alignment procedure. In the latter procedure, the validation tiles from the target domain were only used to track performance during training; they had no influence in the respective training process. The accuracies reported for all experiments were accessed using the test tiles.

In the datasets, the references provide no (change) information in areas that were deforested in previous years. After deforestation is first detected by the PRODES program, the corresponding regions remain marked as deforestation, regardless of any future change. Therefore, both in the source classifiers training and in the evaluation, the pixels that correspond to areas that were deforested prior to the acquisition of the first image in each domain image pair were ignored.

4.3 Classifier Architecture

The deforestation detection classifier follows a fully convolutional encoder-decoder architecture with input size of 128×128 pixels; hence, patches of that size were extracted from the image tiles through a sliding window procedure. For training, the overlap between consecutive patches was of 96% for PA and

MA, and of 94% for RO. The patches used for testing did not overlap each other. The input of the network follows an early fusion configuration, in which the images of a pair are stacked along the spectral dimension. As Table 2 shows, the architecture contains convolutional layers (C) in the encoder, and transposed convolutions (TC) in the decoder; no padding was used. Dropout with $rate = 10\%$ and ReLU activation were employed after all convolutions, except in the last layer before the *softmax*.

In all of the datasets the proportion of deforested areas is very low. To mitigate such class imbalance, we used the weighted cross entropy loss (Panchapagesan et al., 2016) in the training of the source classifier. The weights defined for the classes *deforestation/no-deforestation* were inversely proportional to their representation in the source domain training set. Another strategy used for alleviating the class imbalance was to select patches with at least 2% of deforestation pixels during training. Naturally, that was only done for the source domain patches, since only source labels are available to build the DA model.

For the training procedure of the source classifier (cf. Figure 1a) we stipulated early stopping after 10 epochs without improving the performance on the validation set. The Adam optimizer (Kingma and Ba, 2014) was used with a fixed learning rate of 0.0001, and we used batches of 32 image patches. For data augmentation, the training patches were randomly transformed using anticlockwise 90° rotations and flips.

4.4 Adaptation

Referring to the architecture described in Table 2, the layers up to the one chosen as the *adaptation layer* (marked with an asterisk * in the table) compose the feature extractors E^S and E^T. The subsequent layers compose the label predictor P. Therefore, the activation maps produced by the adaptation layer represent the features of interest for the adaptation process, i.e., the features that comprise the input to the domain discriminator D. Following Wittich and Rottensteiner (2019), we chose a layer prior to the bottleneck of the network as the adaptation layer.

The discriminator architecture D comprises four 1×1 convolutional layers with 512 filters and leaky-ReLU activations, and a final 1×1 convolutional layer with one filter and *sigmoid* activation. As in (Wittich and Rottensteiner, 2019), this discriminator uses padding and $stride = 1$, and returns dense probability predictions.

For the regularization term defined in Equation 3, we set $\lambda = 2$ and $m = 2.5$. The λ value is the same used in (Wittich and Rottensteiner, 2019). The m value was empirically defined in experiments that used the domain combination RO (source) and MA (target). The batch size was set to 1. The set of training image patches from both domains was also augmented using random flips and anticlockwise 90° rotations. We used the Adam optimizer with an initial learning rate of 0.0001. To guarantee a stable model at the end of the training procedure we used learning rate decay after the first 40 epochs. In that sense, we trained the DA phase (cf. Figure 1b) for 150 epochs in total.

5. RESULTS

Figure 4 shows the *precision vs. recall* curves for the class $deforestation$. The curves represent the trade-off between precision and recall for different probability decision thresholds, varying the latter from 1 to 0. The informed *mean average*
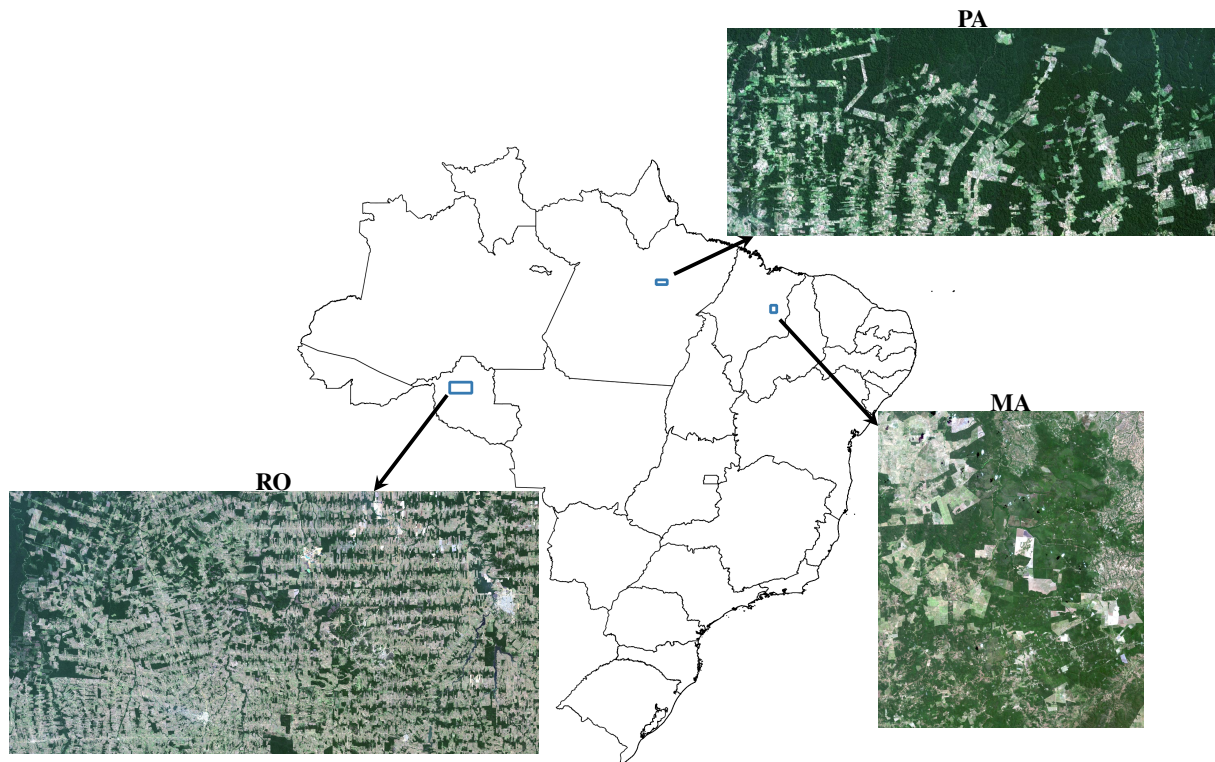

Figure 3. Study areas on Brazilian Biomes; the images are those of the second acquisition epoch in all three test areas.

	PA	RO	MA
Latitude (S)	03°08'21" - 03°26'16"	09°36'51" - 10°18'35"	04°44'52" - 05°12'48"
Longitude (W)	50°34'04" - 51°16'12"	62°56'41" - 64°20'51"	43°37'55" - 44°01'23"
Size (pixels)	1098×2599	2550×5120	1700×1440

Table 1. Locations of the three test areas and size of the corresponding Landsat images.

Encoder	Decoder
$C(5 \times 5, 96, 1)$	$TC(3 \times 3, 512, 1)$
$C(2 \times 2, 96, 2)$	$TC(2 \times 2, 256, 2)$
$C(3 \times 3, 128, 1)^*$	$TC(3 \times 3, 256, 1)$
$C(2 \times 2, 128, 2)$	$TC(2 \times 2, 128, 2)$
$C(3 \times 3, 256, 1)$	$TC(3 \times 3, 128, 1)$
$C(2 \times 2, 256, 2)$	$TC(2 \times 2, 64, 2)$
$C(3 \times 3, 512, 1)$	$TC(5 \times 5, 64, 1)$
	softmax()

* adaptation layer.

Table 2. Network Architecture; $(w \times w, k, s)$ denotes (kernel size, number of kernels, strides)

precision values were obtained by computing the area under the respective curves. For each scenario, four cases were considered: (1) Tr:T,Ts:T, i.e., training and testing on the target domain; (2) Tr:S,Ts:T, i.e., training on the source domain and testing on the target domain; (3) ADDA using $\mathcal{L}_{reg}$, tested on the target domain; (4) ADDA using $\mathcal{L}_{reg}^m$, which we refer to as variant $ADDAm$, also tested on the target domain. Cases (1) and (2) are expected to be upper and lower bounds for the adaptation results, respectively; case (2), which corresponds to a classification of the target domain without DA, is regarded as the baseline against which to compare the DA results.

Figure 5 shows the F1-scores obtained in the experiments, using a probability threshold of 0.5, i.e. choosing the most probable class label for each pixel. Each group of bars in the figure is related to a different DA scenario. The green and orange bars represent the results associated with the two classification cases without DA, as described in the previous paragraph. The remaining bars represent the performance attained using the DA alternatives.

As expected, the mean average precision and F1-score of cases (1) are higher than those achieved in the other cases, because the training and test sets were from the same domain. Conversely, the classifications without adaptation (case (2)) generally show the worst performance. Moreover, the DA procedure brought significant performance improvements in most of the cross domain classification cases.

Analyzing the curves of cases (1), the results obtained with the classifier trained and tested on MA (cf. figures 4b and 4d) and PA (cf. figures 4c and 4f) domains are similar, but significantly better than the ones obtained for the RO domain (cf. figures 4a and 4e). This may be due to the fact that deforestation in MA and PA is mainly driven by agricultural purposes (Marris, 2005), which demand clearcuts of the forest. In RO, however, selective logging, which is harder to detect, is the most common deforestation method (Muchagata and Brown, 2003).

Regarding the domain adaptation results, the largest gaps between curves (1) and (2), i.e., the upper bound and the baseline, occurred when PA was the source domain (figures 4a and 4b). This may be due to a larger variability in the vegetation patterns of the other domains, which makes the classifiers trained on MA and RO more efficient in discerning changes (in the no-
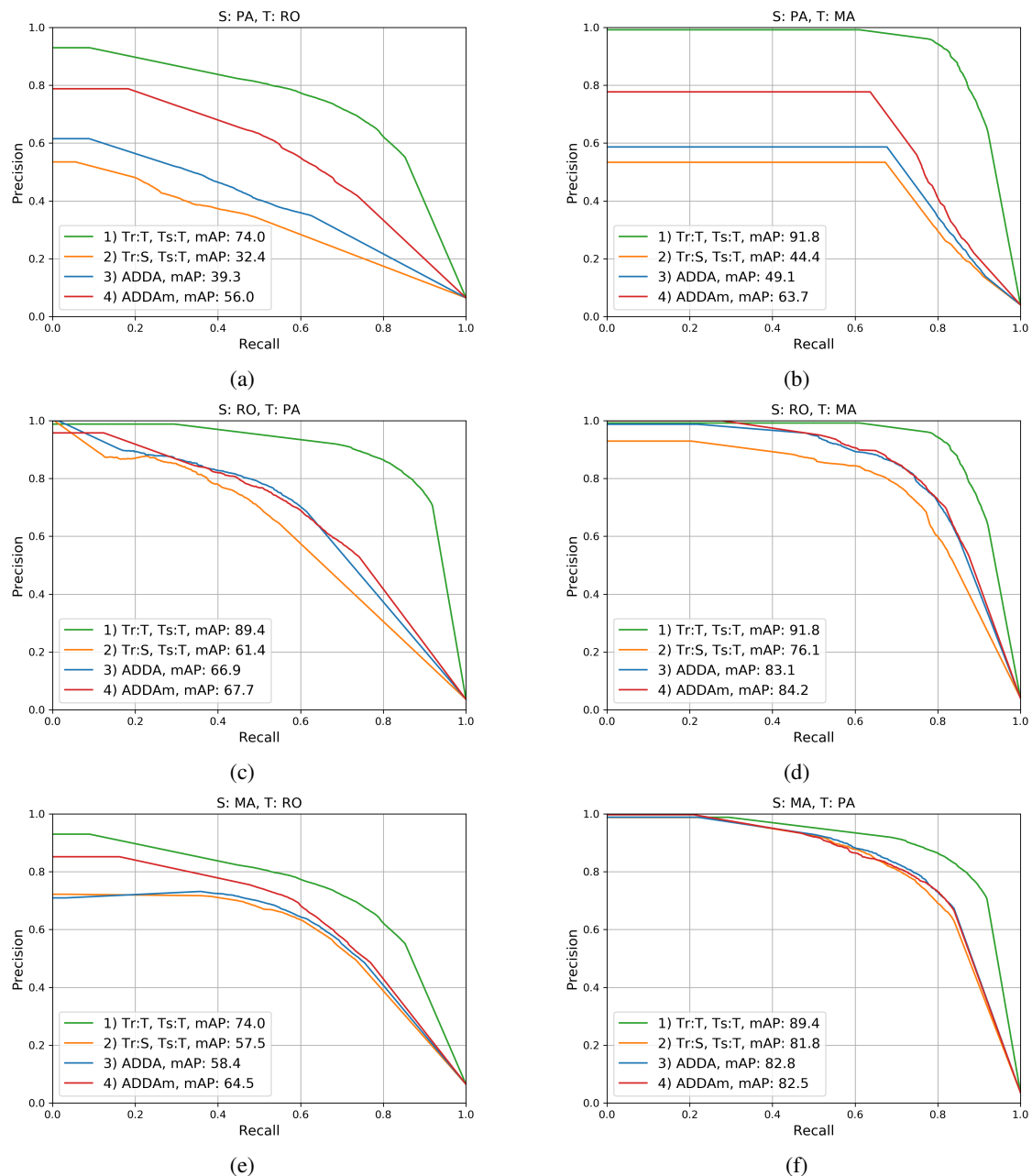
Figure 4. Precision vs. Recall curves and Mean Average Precision (mAP) in [%] for the six scenarios, considering the four variants described in the main text. These metrics are related to the class *deforestation*. S / T: Source / Target domains.

deforestation class) that are not actually associated with deforestation.

Moreover, it can be observed in Figure 4 that the larger the gaps between the upper bound and the baseline curves, the higher the gains brought by the DA approach. Accordingly, in the cases in which the gaps are smaller, e.g., S:MA,T:RO and S:MA,T:PA (refer to figures 4e and 4f, respectively), the approach brought quite small or even negligible improvements. Be that as it may, in most of the scenarios, the variant that uses the proposed margin-based regularization $\mathcal{L}_{reg}^m$ (ADDAm) was superior to the variant based on the raw L1-distance $\mathcal{L}_{reg}$ regularization (ADDA). The only exception was S:MA,T:PA (Figure 4f), but this is the case with the smallest gap between curves (1) and (2), in which, therefore, DA is not of much help.

As for the F1-scores (Figure 5), they show a consistent beha-

viour in relation to the results presented in Figure 4. In the first three scenarios, both adaptation variants significantly outperformed the baselines, with a clear advantage of the ADDAm. In the S:RO,T:MA scenario, the adaptations were also successful, but with almost no difference between them. In the last two scenarios, however, the adaptation approach delivered results that are very similar to the baseline, but as mentioned before, the effect of DA in those cases is restricted, as the gaps between the upper bound and baseline were quite small.

6. CONCLUSIONS

In this work, we introduced a domain adaptation approach for deforestation detection based on representation matching, following the Adversarial Discriminative Domain Adaptation (ADDA) framework. We further introduced a margin-based regulariza-
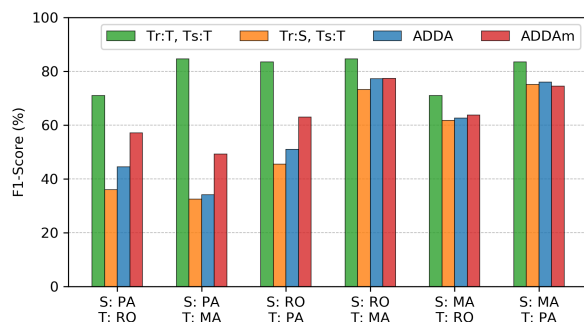
Figure 5. F1-Scores for the class *deforestation* for the six different scenarios and the four classification variants.

tion constraint in the learning process, which promotes a better convergence of the model parameters during training and which is less restrictive than the original term in terms of the feature adaptation process.

We evaluated the approach considering three different domains, which represent sites in the Amazon and Brazilian Cerrado biomes. The results showed that the approach was successful in the adaptation of most of the domain combination scenarios, usually with important gains in relation to the baselines. Unsurprisingly, the larger the shift between domains, the higher the gains brought by the DA approach. Moreover, the approach variant that includes the proposed regularization term delivered better results than the variant using the original regularization loss formulation in most cases.

Finally, we believe that we can further improve the proposed DA approach. For instance, we plan to study ways to select the value of the margin regularization parameter m trough an automatic/adaptive procedure. We also want to explore alternatives to better generalize the cross domain classification, maybe training the classification and adaptation procedures in a single stage, or testing other discriminator architectures that are able to incorporate more contextual information.

ACKNOWLEDGEMENTS

References

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. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Science*, XLIII-B3-2020, 1497–1505.

De Sy, V., Herold, M., Achard, F., Beuchle, R., Clevers, J. G., Lindquist, E., Verchot, L., 2015. Land use patterns and related carbon losses following deforestation in South America. *Environmental Research Letters*, 10(12), 124004.

Deng, W., Zheng, L., Ye, Q., Kang, G., Yang, Y., Jiao, J., 2018. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-

identification. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., Lempitsky, V., 2016. Domain-Adversarial Training of Neural Networks. *J. Mach. Learn. Res.*, 17(1), 2096–2030.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial networks. *arXiv preprint arXiv:1406.2661*.

Huang, H., Huang, Q., Krähenbühl, P., 2018. Domain transfer through deep activation matching. V. Ferrari, M. Hebert, C. Sminchisescu, Y. Weiss (eds), *Computer Vision – ECCV 2018*, Springer International Publishing, Cham, 611–626.

Huh, M., Agrawal, P., Efros, A. A., 2016. What makes ImageNet good for transfer learning? http://arxiv.org/abs/1608.08614.

Kingma, D. P., Ba, J., 2014. Adam: A Method for Stochastic Optimization. http://arxiv.org/abs/1412.6980.

Krizhevsky, A., Sutskever, I., Hinton, G. E., 2017. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84–90.

Kukačka, J., Golkov, V., Cremers, D., 2017. Regularization for deep learning: A taxonomy. *arXiv preprint arXiv:1710.10686*.

Li, J., Chen, E., Ding, Z., Zhu, L., Lu, K., Tao Shen, H., 2020. Maximum density divergence for domain adaptation. *arXiv*, 8828(c).

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.

Marris, E., 2005. The forgotten ecosystem. *Nature*, 437(7061), 944–945. http://www.nature.com/articles/437944a.

Muchagata, M., Brown, K., 2003. Cows, colonists and trees: rethinking cattle and environmental degradation in Brazilian Amazonia. *Agricultural systems*, 76(3), 797–816.

Panchapagesan, S., Sun, M., Khare, A., Matsoukas, S., Mandal, A., Hoffmeister, B., Vitaladevuni, S., 2016. Multi-task learning and weighted cross-entropy for dnn-based keyword spotting. *Interspeech*, 9, 760–764.

Soto, P. J., Costa, G. A. O. P., Feitosa, R. Q., Happ, P. N., Ortega, M. X., Noa, J., Almeida, C. A., Heipke, C., 2020. Domain adaptation with cyclegan for change detection in the amazon forest. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Science*, XLIII-B3-2020, 1635–1643.

Sun, B., Saenko, K., 2016. Deep CORAL: Correlation alignment for deep domain adaptation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9915 LNCS, 443–450.

Tasar, O., Happy, S. L., Tarabalka, Y., Alliez, P., 2020. ColorMapGAN: Unsupervised Domain Adaptation for Semantic Segmentation Using Color Mapping Generative Adversarial Networks. *IEEE Transactions on Geoscience and Remote Sensing*, 58(10), 7178-7193.

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. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 7167–7176.

Valeriano, D. M., Mello, E. M. K., Moreira, J. C., Shimabukuro, Y. E., Duarte, V., Souza, I. M., dos Santos, J. R., Barbosa, C. C. F., de Souza, R. C. M., 2004. Monitoring tropical forest from space: the prodes digital project. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Science*, XXXV-B7, 272–274.

Wang, M., Deng, W., 2018. Deep visual domain adaptation: A survey. *Neurocomputing*, 312, 135–153.

Wittich, D., Rottensteiner, F., 2019. Adversarial domain adaptation for the classification of aerial images and height data using convolutional neural networks. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Science*, IV-2/W7, 197–204.

Zhu, J. Y., Park, T., Isola, P., Efros, A. A., 2017. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. *IEEE International Conference on Computer Vision(ICCV)*, 2242–2251.