Assessing the Generalization Capacity of Convolutional Neural Networks and Vision Transformers for Deforestation Detection in Tropical Biomes

Pedro J Soto Vega, Daliana Lobo Torres, Gustavo X Andrade-Miranda, Gilson a O P da Costa, Raul Queiroz Feitosa

Assessing the Generalization Capacity of Convolutional Neural Networks and Vision Transformers for Deforestation Detection in Tropical Biomes

Pedro J. Soto Vega[1], Daliana Lobo Torres[2], Gustavo X. Andrade-Miranda[3], Gilson A. O. P. da Costa[4], Raul Queiroz Feitosa[2]

[1] L@bISEN, Vision-AD and Auto-ROB, ISEN Yncréa Ouest, 20 rue Cuirassé Bretagne, 29200 Brest, France - pedro-juan.soto-vega@isen-ouest.yncrea.fr
[3] University Brest, LaTIM, INSERM UMR 1101, Brest, France -andradema@univ-brest.fr
[2] Dept. of Electrical Engineering, Computer Vision Laboratory, Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, Brazil - daliana91@aluno.puc-rio.br, rqfeitosa@ele.puc-rio.br
[4] Institute of Mathematics and Statistics, State University of Rio de Janeiro (UERJ), Rio de Janeiro, Brazil – gilson.costa@ime.uerj.br

Keywords: Deforestation Detection, Deep Learning, Convolutions, Transformers, Domain Shift

Abstract

Deep Learning (DL) models, such as Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs), have become popular for change detection tasks, including the deforestation mapping application. However, not enough attention has been paid to the domain shift issue, which affects classification performance when pre-trained models are used in areas with different forest covers and deforestation practices. This study compares DL methods for deforestation detection, focusing on assessing how well CNNs and ViTs can adapt to the domain shift. Two different models, namely, DeepLabv3+ and UNETR, were trained using remote sensing images and references from a specific location and then tested in other sites to simulate real-world scenarios. The results showed that the ViT-based architecture achieved better performance when trained and tested in the same region but showed lower generalization capacity in cross-domain scenarios. We consider this a work in progress that needs further research to confirm its findings, with the evaluation of additional architectures on a wider range of domains.

1. Introduction

Deforestation is a critical environmental issue, contributing significantly to greenhouse gas emissions and impacting carbon storage and biodiversity. The ability to accurately monitor changes in natural forests is crucial for public and private organizations aiming to mitigate these effects.

With the technological advancements in Earth observation technologies in the last couple of decades, remote sensing (RS) data availability has increased substantially, enabling the development of sophisticated automatic change detection techniques (Soto Vega et al., 2022; Andrade et al., 2022).

Among the variety of methods proposed recently, deep learning (DL) models, especially convolutional neural networks (CNNs), have gained prominence in image analysis tasks, including change detection in RS data. Techniques utilizing Siamese CNNs for urban area change detection and Early Fusion (EF) schemes for deforestation detection have shown promising results, demonstrating the effectiveness of CNNs in change detection applications (Daudt et al., 2018; Ortega Adarme et al., 2020). Moreover, exploiting different deep network architectures across various sensors has underscored the adaptability of DL methods to the challenges of deforestation detection (De Bem et al., 2020; Torres et al., 2021).

Vision Transformers (ViTs) have recently emerged as a novel paradigm in computer vision, demonstrating superior performance over traditional CNNs by effectively capturing global image features. Their employment in change detection tasks, including deforestation mapping, has revealed their potential to revolutionize this application field (Bandara and Patel, 2022; Ferrari and Feitosa, 2023).

However, an important problem that remains inadequately addressed in change detection is domain shift, as the performance of DL-based models degrades when tested in areas with data distributions different in relation to the data they were trained with. This issue is particularly pertinent in deforestation detection, where the diversity of forest types and deforestation practices can vary widely across different geographical regions (Vega et al., 2021; Tuia et al., 2016).

This study aims to address this gap by comparing the generalization capacity of CNNs and ViTs in the context of deforestation detection, with a specific focus on their resilience to domain shift. By training models on RS data from distinct Brazilian regions and evaluating their performance across different domains, we intend to assess the robustness and adaptability of these models to the variations in forests and deforestation patterns.

The main contributions of this work are the following:

- A comprehensive examination of the impact of domain shift on the generalization abilities of CNNs and ViTs in deforestation detection, addressing a critical gap in current research.
- A comparative analysis considering two representative DL models (i.e., DeepLabv3+, and UNETR) in the deforestation detection task, evaluating their performances and generalization capacity across different domains.

The remainder of this paper is organized as follows. Section 2 presents the DL-based architectures evaluated in this work. Section 3 describes the datasets, the experimental setup, the network implementations and the adopted performance metrics. Section 4 presents the obtained results, and finally, Section 5 presents conclusions and directions for future research.

2. Methods

In the next sections, we will briefly describe the neural network models evaluated in this work namely DeepLabv3+ (Chen et al., 2018) and UNETR (Hatamizadeh et al., 2022).

2.1 DeepLabv3+

The DeepLab series includes fully convolutional CNN architectures that have advanced semantic image segmentation through a number of innovations. Version 1 introduced atrous convolution, which expanded the filters' receptive fields to capture a broader context. Version 2 featured Atrous Spatial Pyramidal Pooling (ASPP) to capture multi-scale information. Version 3 incorporated image pooling to include global context.

DeepLabv3+ (Chen et al., 2018) builds upon DeepLabv3 (Gao, 2023) by adding a decoder to improve segmentation quality, especially at object boundaries. It upsamples the encoder output, i.e., high-level features, and combines it with low-level features from the backbone network to preserve spatial details, as shown in figure 1. The model allows using different architectures as encoders, such as Xception (Chollet, 2017). In this work, we used the ResNet (He et al., 2016) as the model's backbone.

2.2 UNETR

The UNETR(Hatamizadeh et al., 2022) is a hybrid network initially proposed for medical image applications. It adopts a U-Net (Ronneberger et al., 2015) style architecture for segmenting 3D images. The network uses ViT blocks as an encoder and has a CNN as a decoder. The encoder is designed to learn features from input image patches, allowing it to capture global information and long-range spatial dependencies. Additionally, the encoder communicates directly with the decoder through skip connections to extract multi-scale information and integrate it for pixel-wise predictions.

The input of the encoding architecture encompasses a linear layer for patch embeddings. This layer is responsible for projecting the patches and their corresponding position embeddings to yield 1D flattened patches. Additionally, a 1D learnable positional embedding is appended to the end of the feature vector to retain spatial information. Subsequently, the patches undergo multiple multi-head self-attention blocks before progressing to the decoding stage. The transition to the segmentation mask space is accomplished through CNN up-sampling combined with multi-level feature aggregation. Figure 2 provides comprehensive details about the architecture components.

3. Experimental Analysis

3.1 Datasets

The selected sites represent different forest types and are affected by various deforestation practices. Two sites are located in the Brazilian Amazon and correspond to Dense Ombrophyll and Open Ombrophyll forest areas. The third site is located in the transition zone between the Brazilian Cerrado and the Amazon Rainforest and contains Seasonal Deciduous and Semi-Deciduous forest cover.

Detailed information about the geographical location, dates, vegetation types, and class distribution can be found in Table 1. In this study, each site is considered a separate domain.

To address the generalization capacity of the DL models within a domain shift context, six different domain combinations were evaluated, considering each of the three areas as training and evaluating domains in turn. In each experiment, the models were trained with data from one site/domain and evaluated on the other sites/domains.

The RS images used in the experiments were acquired by the Landsat 8-OLI sensor, with 30m resolution and seven spectral bands, covering regions within the Brazilian states of Rondônia (RO), Pará (PA) and Maranhão (MA). The images have the following dimensions: 2550×5120 pixels for RO; 1100×2600 pixels for PA; and 1700×1440 pixels for 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]. The deforestation ground truth was produced by the PRODES Deforestation Monitoring Project (Almeida et al., 2021), operated by the Brazilian National Institute for Space Research (INPE). The data is freely available at the Terrabrasilis website [2]. Figure 3 shows the location of the study areas, as well as RGB composites of the most recent images of the corresponding image pairs.

We observe that the RS images used in this study were the same used in PRODES for deforestation mapping for the respective sites and epochs. As shown in Table 1, all images were acquired in July and August, in which the acquisition conditions are optimum concerning cloud coverage.

Figure 4 shows the deforestation reference for the respective image pairs (dark orange), representing the deforestation between the acquisition dates of the two images. The figure also shows the total deforestation, which occurred between 1988 and the acquisition year of the first image of the pair.

3.2 Classifiers Training Setup

We used the Early Fusion (EF) strategy proposed in (Daudt et al., 2018; Ortega Adarme et al., 2020) for the deforestation detection task. Additionally, following (Ortega Adarme et al., 2020; Noa et al., 2021; Vega et al., 2021; Soto Vega et al., 2022; Soto et al., 2022; Vega et al., 2023), the image space was divided into 100 tiles for RO and 15 for PA and PA. Approximately 20% of the tiles were used to extract training patches, 5% to extract validation patches, and the remaining 75% to extract the patches used for the evaluation of the classifier. Figure 4 depicts the deforestation reference as well as the training, validation, and testing tiles.

The patches forwarded through to the networks were tensors of size $64\times64\times14$. The respective patches were extracted following a sliding windows procedure with an overlap of 90%, as in Vega et al. (2023). Additionally, to be selected for training and validation, at least 2% of the total number of pixel positions in a patch had to be labeled as deforestation. Following Ortega Adarme et al. (2020); Vega et al. (2021); Noa et al. (2021); Soto Vega et al. (2022); Soto et al. (2022), and Vega et al. (2023), during training and testing, the pixels covering image areas with the following characteristics were masked out (therefore not considered in loss or accuracy computation): (i) belonging to regions that were subjected to deforestation before the date of the first image in the pair; (ii) situated within a buffer of a width of two pixels outside the deforestation polygons; (iii)

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

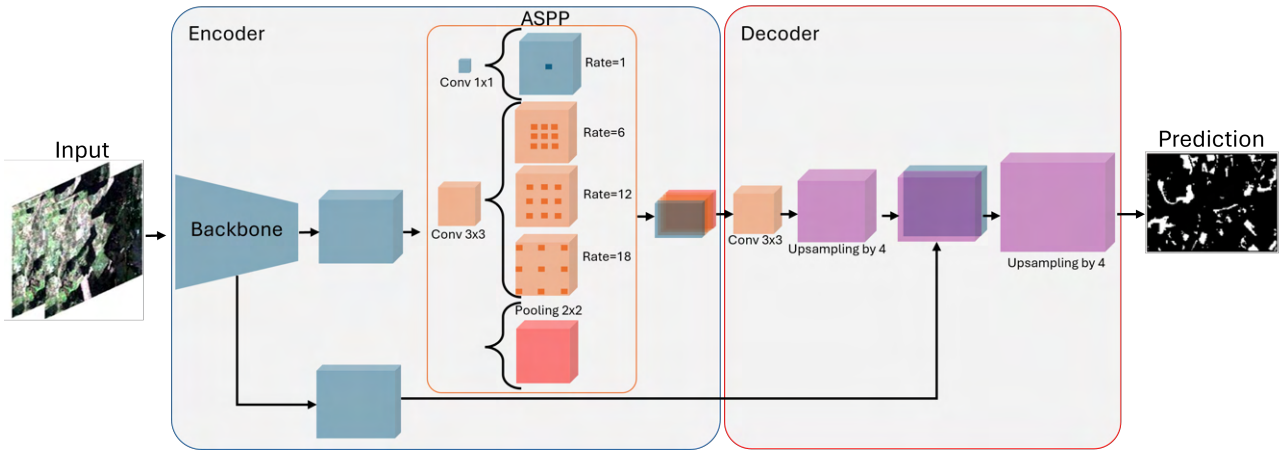

Figure 1. Overview of DeepLabv3+ architecture. In the figure, ASPP stands for Atrous Spatial Pyramidal Pooling.

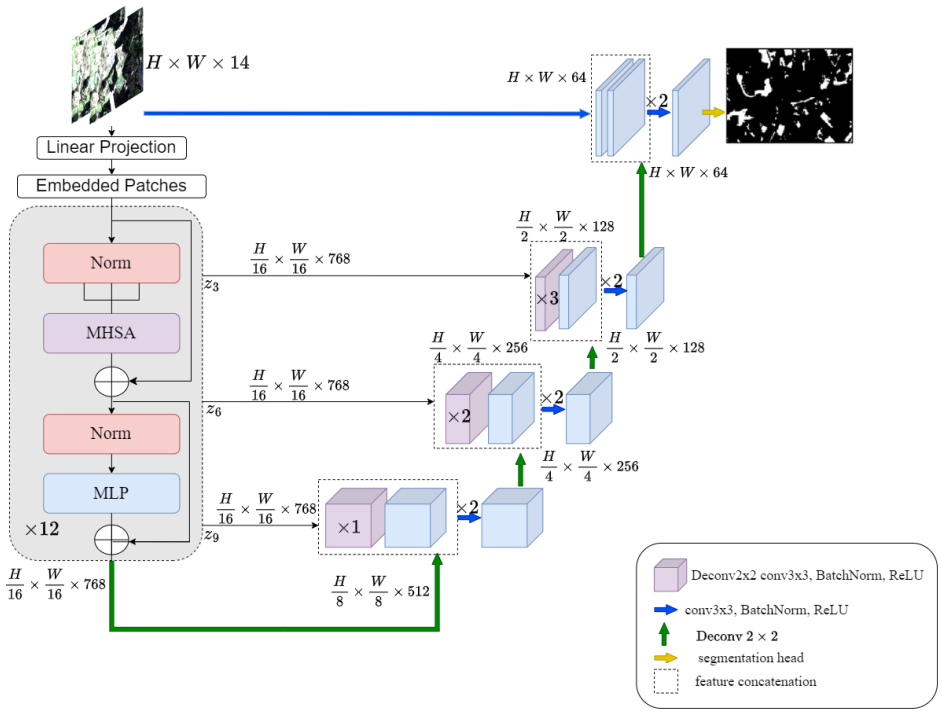

Figure 2. Overview of UNETR architecture (Hatamizadeh et al., 2022)

Domains	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 domain information: image acquisition dates, coordinates, class distribution, and vegetation typology. Data taken from (Vega et al., 2021; Soto Vega et al., 2022)

inside deforestation polygons smaller than 6.25 ha (equivalent to 69 pixels) in the Amazon sites and deforestation polygons smaller than 1 ha (equivalent to 11 pixels) for the Cerrado site. The first restriction is due to the absence of data in the reference dataset regarding changes in regions that were deforested in previous years; once PRODES identifies deforestation, the corresponding areas remain classified as deforested, irrespective of any future changes. The second condition is designed to mitigate the impact of minor inaccuracies in the deforestation reference polygons resulting from the rasterization process. In

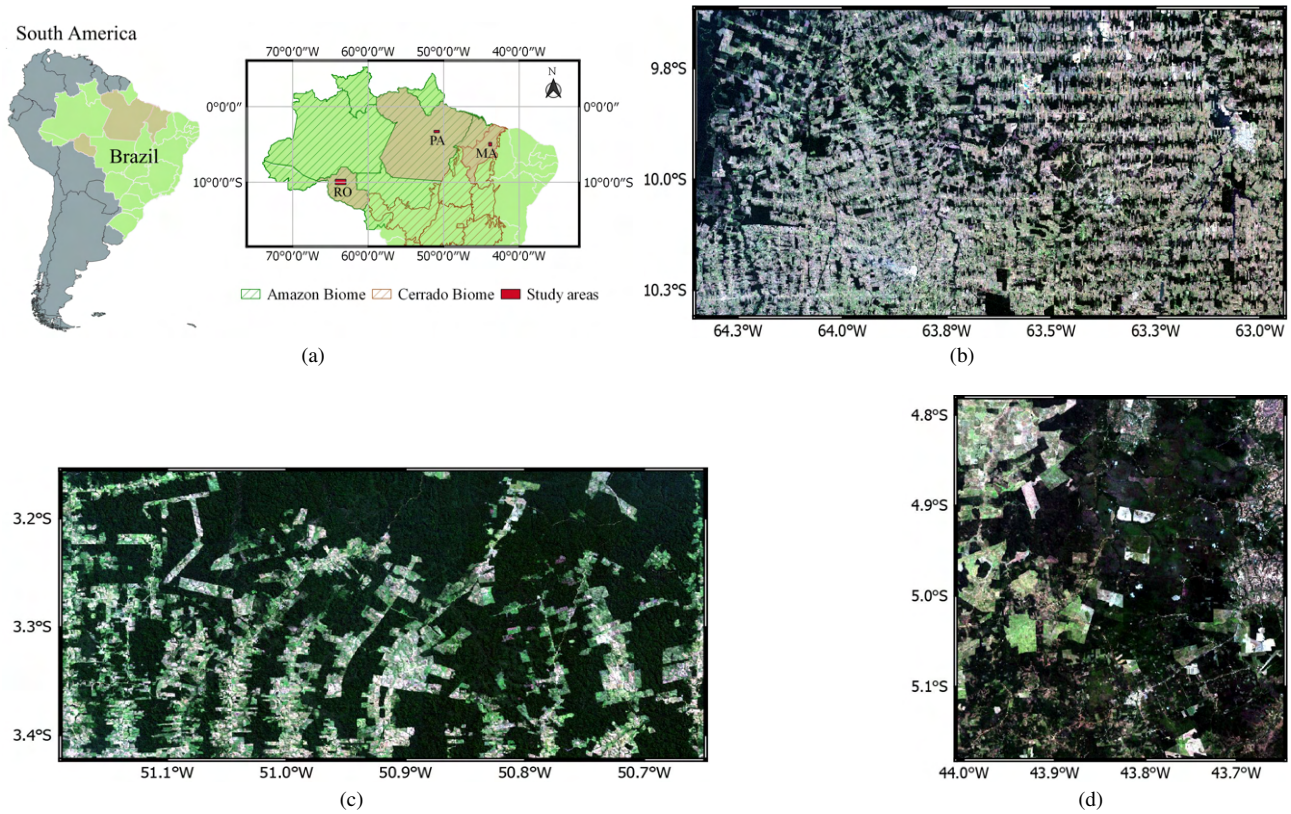

Figure 3. Visual representation and localization of each study area used in the experiments carried out in this work. (a) Geographical localization of the respective sites. True color composites of the images covering the sites, corresponding to the acquisition date 2017 (b) Rondônia (RO), (c) Pará (PA), and (d) Maranhão (MA). Figure taken from Vega et al. (2021).

regard to the third constraint, we have adopted the same criteria as PRODES, considering its minimum mapping units (Vega et al., 2021).

The data augmentation transformations were a 90° rotation and vertical and horizontal flips. To alleviate the imbalance in the dataset, we adopted focal loss as a cost function with γ equal to 2, which was minimized using the Adam optimizer with learning rate decay. We set the initial learning rate μ_0 and momentum β_1 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 20. For each evaluated architecture, the training and testing procedures were executed three times, each time with a different (random) initialization of the trainable parameters. Regarding the network architectures, figures 1 and 2 provide detailed information about the structure and composition of each model.

At test time, the trained DNN assigns probability values for all pixels of patches extracted from the test regions using a sliding window procedure with a 25% overlap. A heat map mosaic is created by stitching together only the center portions of the patches' predictions. Following Soto et al. (2022), such a procedure aims at reducing artifacts by removing weak predictions close to the borders of the patches.

3.3 Metrics

The performance of the classifications in all scenarios is expressed in terms of the average F1-score (F1) considering the positive class (deforestation). Specifically, the F1-score is the harmonic mean between Precision and Recall, as follows:

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall}, \tag{1}$$

where,

$$Precision = \frac{t_p}{t_p + f_p}, \tag{2}$$

$$Recall = \frac{t_p}{t_p + f_n} \tag{3}$$

In equations 2 and 3, t_p is the number of pixels correctly assigned to the deforestation class (true positives), f_p represents the number of pixels erroneously classified as deforestation (false positives). Similarly, f_n corresponds to the number of pixels incorrectly classified as non-deforestation (false negatives).

4. Results and Discussion

Tables 2, 3, and 4 show the average F1 scores, precision, and recall of the models delivered after three training and testing runs. Scores on the table's diagonal represent the accuracies observed when the models are trained and tested on data from the same domain. The scores off-diagonal indicate the accuracies when the models were trained on data from one domain and tested on a different domain. In all tables, values highlighted in bold represent the higher accuracies achieved when the models were trained and tested with data from the same domain. The values in italics correspond to the best results in the cross-domain scenarios.

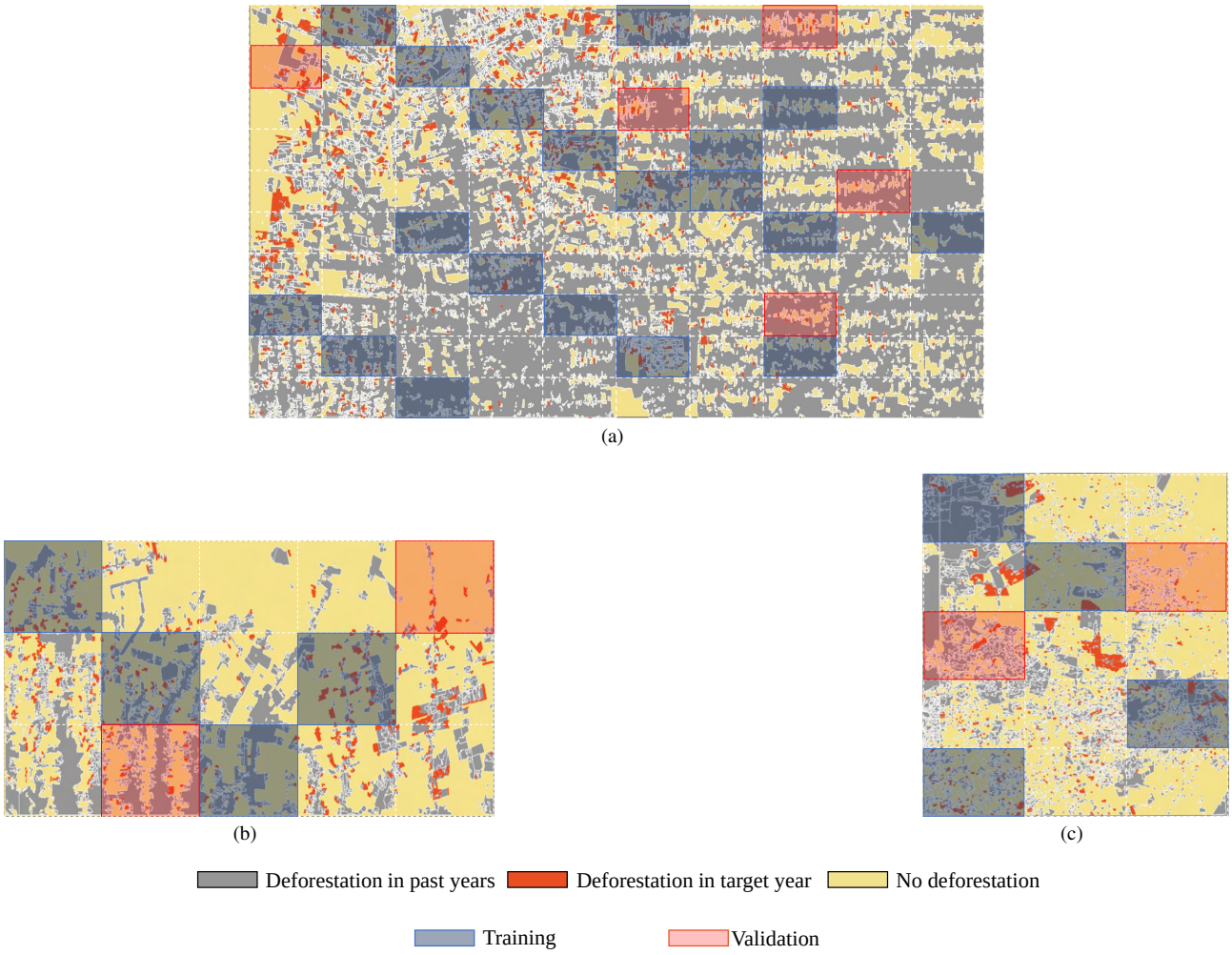

Figure 4. Distribution of image tiles for training, validation and testing in the respective study areas: (a) Rondônia (RO); (b) Pará (PA); and (c) Maranhão (MA). Please note that the tiles that are not shaded correspond to the ones selected for testing. The figure also shows the polygons associated 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).

Figure 5 shows the ground truth and deforestation detection results (in red) overlaid on an RGB composition of the input Landsat-8 image. The first line represents the deforestation ground truth (figures 5(a)(b)(c)). Figures 5(d)-(l) represent the result obtained with DeepLabv3+, and figures 5(m)-(u) represent the result obtained with UNETR.

The results show that in all but one case the transformer-based model outperformed the convolutional one when trained and tested with data from the same domain. Moreover, training and testing the models on the same domains produced higher accuracies for both models, as expected. Conversely, the F1 scores associated with the cross-domain combinations were significantly lower. In those cases, however, the DeepLabv3+ fully convolutional model obtained higher accuracies in terms of F1 score in all but two cases (when models were evaluated on the PA site).

By carefully inspecting Table 2 it can be observed that the highest absolute accuracies in the cross-domain evaluations were obtained when the models were trained on MA and tested in PA. Moreover, the results reveal that the classifiers trained on the RO domain generalize worse when tested on the PA or MA domains. The reason for such behavior was suggested in Vega et al. (2021), where the authors analyzed the complexity of the deforestation and non-deforestation classes in each domain aiming at explaining the behavior of a U-Net classifier in the same task and over the same datasets.

Succinctly, Vega et al. (2021) used the number of clusters in difference images to indicate class complexity. They found that forest complexity is lowest in PA and highest in MA, while the opposite is true for deforested areas. In RO, the complexity falls between that of the PA and MA sites.

Based on those findings, we believe that the classifiers trained on MA are better at recognizing changes not linked to deforestation. As seen in tables 3 and 4, in general, both architectures achieved higher precision and recall when trained in MA compared to other domains (in the cross-domain scenarios). In other words, such a higher variability in MA forested regions leads to a classifier with a lower false positive rate.

However, the higher diversity of deforestation in PA could cause the classifier trained with data from that site to incorrectly classify changes that are not related to deforestation in different

Methods	Training on:	Evaluating on:		
	Domains	RO	PA	MA
DeepLabv3+ (ResNet)	RO	64.9	6.1	*45.3*
	PA	*38.5*	73.3	*54.3*
	MA	*50.4*	64.7	68.7
	Domains	RO	PA	MA
UNETR + ViT small	RO	**66.4**	*6.4*	29.3
	PA	25.5	**80.3**	35.8
	MA	43.4	*71.3*	**75.9**

Table 2. F1 scores (%) of DeepLabv3+ and UNETR semantic segmentation DL-based architectures.

Methods	Training on:	Evaluating on:		
	Domains	RO	PA	MA
DeepLabv3+ (ResNet)	RO	**64.1**	*4.4*	*43.4*
	PA	*50.8*	81.5	*38.6*
	MA	*74.1*	58.6	55
	Domains	RO	PA	MA
UNETR + ViT small	RO	63.6	4.3	22.3
	PA	30.8	**84.3**	22.3
	MA	47.8	*77.6*	**62.7**

Table 3. Precision (%) of DeepLabv3+ and UNETR semantic segmentation DL-based architectures.

sites. This could result in more false positives, as indicated by the lower precision values shown in Table 3. Additionally, both models delivered low precision and recall when trained in RO and tested in PA (see also figures 5(e) and (n)). We believe this is due to the poor representativeness of changes in forested and deforested regions of RO. The situation in PA is in extremes, with the highest change variability in deforested regions and the lowest in forested ones, which we believe leads to poor precision and recall in the respective domain combinations.

Notwithstanding, the ViT-based architecture showed a lower generalization capacity compared to its convolutional counterpart in most cross-domain scenarios. We understand that the result deserves further investigation, however, it is worth noting that the lower performance is mainly caused by the lower precision of UNETR, as shown in Table 3, in the cross-domain scenarios. As for the recall metric, Table 4, the UNETR achieved better results than DeepLabv3+. The former demonstrates a lack of robustness in avoiding false positives in forested regions when compared with the performance of DeepLabv3+ in similar scenarios, which can also be observed in Figure 5(n)(o)(r)(s), where several regions predicted as deforested are not actually deforested, i.e., false positives when compared against the ground truth.

5. Conclusions

In this research, we examined different deep learning-based models, namely convolutional and transformer architectures, by assessing their capacity for change detection in the context of deforestation detection within tropical forests. Our study involved a comparative analysis of the models’ performances in cross-domain combination scenarios, focusing on the F1-score, Precision, and Recall metrics.

The results showed that the highest performances were obtained for the intra-domain classification scenarios, for both models. In this context, the transformer-based architecture achieved higher deforestation detection rates, which can be considered evidence of the superiority of transformer approaches for semantic segmentation and classification tasks.

Methods	Training on:	Evaluating on:		
	Domains	RO	PA	MA
DeepLabv3+ (ResNet)	RO	65.8	12.6	48.9
	PA	*31.4*	66.8	91.6
	MA	*57.8*	44.6	92.1
	Domains	RO	PA	MA
UNETR + ViT small	RO	**72.4**	*24.3*	*53.4*
	PA	21.7	**77**	*96.9*
	MA	41.7	*66.7*	**97.4**

Table 4. Recall (%) of DeepLabv3+ and UNETR semantic segmentation DL-based architectures.

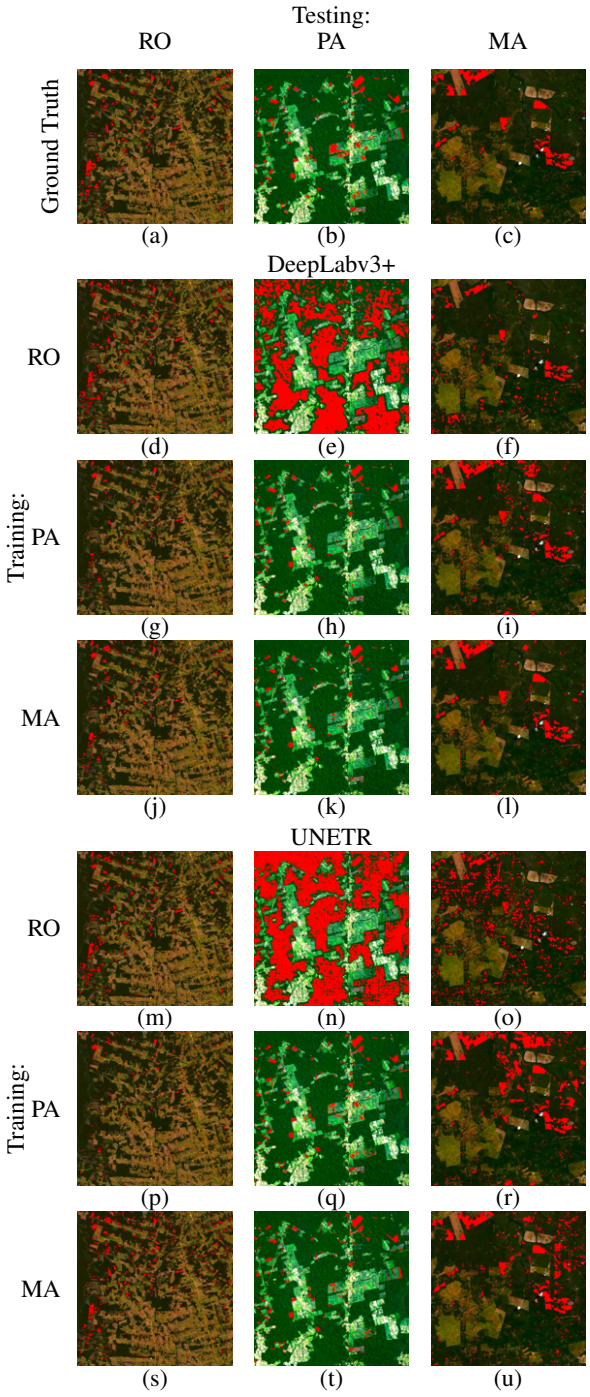

Figure 5. Deforestation ground truth and semantic segmentation results (red areas) obtained with DeepLabv3+ and UNETR.

The performance of both models decreased substantially in the cross-domain evaluation scenarios. We attribute this behavior to the previously studied variability in change patterns within forested and deforested regions. Moreover, in the majority of cross-domain combinations, the convolutional-based model consistently outperformed the transformer-based model. While we understand that lower acuracy of the transformer-based model in these cases may be influenced by various problems, such as overfitting, further investigation must be done to determine their causes.

Finally, we consider it is important to conduct further research to confirm our findings. Such research should involve evaluating additional architectures transformer, convolutional and hybrid architectures, and a wider range of domains and aplications.

6. Acknowledgments

The authors would like to thank the funding provided by CAPES, CNPq, FAPERJ, 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. B., 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). https://www.mdpi.com/2072-4292/14/19/4694.

Bandara, W. G. C., Patel, V. M., 2022. A transformer-based siamese network for change detection. *IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 207–210.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. *Proceedings of the European conference on computer vision (ECCV)*, 801–818.

Chollet, F., 2017. Xception: Deep Learning with Depthwise Separable Convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1251–1258.

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

De Bem, P. P., de Carvalho Junior, O. A., Fontes Guimarães, R., Trancoso Gomes, R. A., 2020. Change detection of deforestation in the Brazilian Amazon using landsat data and convolutional neural networks. *Remote Sensing*, 12(6), 901.

Ferrari, F., Feitosa, M. P. F. R. Q., 2023. A Transformer-based network for deforestation detection from bitemporal optical satellite images.

Gao, R., 2023. Rethinking dilated convolution for real-time semantic segmentation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4675–4684.

Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H. R., Xu, D., 2022. Unetr: Transformers for 3d medical image segmentation. *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 574–584.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016-Decem, 770–778.

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.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*, Springer, 234–241.

Soto, P. J., Costa, G. A., 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., Costa, G., Ortega, M., Bermudez, J., Feitosa, R., 2022. Deforestation detection with weak supervised convolutional neural networks in tropical biomes. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, 713–719.

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., Adarme, M. X. O., Castro, J. D. B., Feitosa, R. Q., 2023. Weakly Supervised Domain Adversarial Neural Network for Deforestation Detection in Tropical Forests. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16, 10264–10278.

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.