

FUSING SENTINEL-1 AND SENTINEL-2 IMAGES WITH TRANSFORMER-BASED NETWORK FOR DEFORESTATION DETECTION IN THE BRAZILIAN AMAZON UNDER DIVERSE CLOUD CONDITIONS

F. Ferrari^{1*}, M. P. Ferreira², R. Q. Feitosa³

^{1,3} Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Rio de Janeiro, RJ, Brazil

¹ferrari@aluno.puc-rio.br, ³raul@ele.puc-rio.br

² Military Institute of Engineering, Rio de Janeiro, RJ, Brazil

²matheus@ime.eb.br

KEY WORDS: Optical Data, SAR Data, Data Fusion, Transformer, Convolution, Deforestation Detection.

ABSTRACT:

Deforestation is an environmental problem that significantly impacts biodiversity and climate change. Deforestation detection is usually performed using optical remote sensing images, limiting the detection capability to the dry season in which images are not comprised of clouds. In this work, we proposed Transformer-based models to fuse bitemporal Sentinel-1 and Sentinel-2 images to identify new deforestation areas in the Brazilian Amazon area under diverse cloud conditions. The models were evaluated considering clear and cloud-covered pixel conditions. The results confirmed previous works in which the fusion of optical and SAR images improved deforestation detection capabilities. We also concluded that the better Transformer-based network reached the F1-Score of 0.92, considering all pixels, outperforming the better Convolution-based which reached the F1-Score of 0.86, without increasing the training and prediction times.

1. INTRODUCTION

The Brazilian Amazon forest is the largest rainforest on the Earth, with a significant role in regulating the Earth's climate, providing habitat for millions of species (Strand et al., 2018). However, deforestation severely threatens the forest, contributing to climate change, biodiversity loss, and social conflicts (Baccini et al., 2017). In 2021, for example, 13,235 km² of the Brazilian Amazon forest were lost (INPE, 2022).

To monitor deforestation, the Brazilian Institute for Space Research (INPE) developed the Program for the Estimation of Deforestation in the Brazilian Amazon (PRODES) in the 1980s. This program uses satellite images to monitor clear-cutting deforestation in the Brazilian Amazon. PRODES relies on optical satellite imagery to track changes in forest cover, providing accurate and reliable data on deforestation rates and trends. However, it still requires a significant amount of visual interpretation, resulting in undesirable implications in terms of time and cost. As a solution, semi-automatic approaches have been explored to minimize accuracy loss to a minimum (INPE, 2022).

Deep Learning (DL) algorithms have become a valuable tool in detecting deforestation from satellite images and analyzing changes in the forest cover by comparing multitemporal satellite images. Usually, the approach involves convolutional neural networks (CNN). Fully convolutional networks (FCNs) have achieved state-of-the-art results in various remote sensing tasks (Ma et al., 2019), including deforestation detection (Adarme et al., 2020). A typical FCN architecture consists of an encoder module for feature extraction and a decoder module that delivers a prediction at the input image resolution. Many FCN variants have been applied for deforestation detection from optical images (Zhang et al., 2018; Torres et al., 2021;

de Bem et al., 2020) and combined optical and SAR data (Ortega et al., 2021; Ferrari et al., 2023).

Recently, an alternative approach called vision transformers (ViT), or simply transformers, emerged as a substitute for FCNs in image-dense prediction tasks (Dosovitskiy et al., 2020). The primary drawback of FCNs is related to their core operation, convolution, which confines the computation of a new pixel representation to a narrow spatial context. Conversely, ViT bypass convolution and can capture the global context of the input. In fact, over the last years, many transformer architectures have outperformed FCNs in dense prediction tasks like semantic segmentation (Liu et al., 2021; Cao et al., 2021).

The primary data sources for monitoring deforestation in tropical areas are images acquired by optical sensors. These images allow for the easy recognition of changes in forest cover resulting from clear-cutting through visual interpretation elements such as tone, form, color, and texture. Many Earth observation satellites also carry optical systems, improving data availability (Belward and Skøien, 2015). However, a significant limitation of optical images is the lack of information in areas covered by clouds, which is very common in tropical environments (Asner, 2001).

When cloud-free optical images are unavailable, synthetic aperture radar (SAR) emerges as an alternative (Silva et al., 2022; Doblus et al., 2020; Bouvet et al., 2018). SAR is sensitive to surface structural properties such as roughness and moisture content, which may aid deforestation monitoring. However, classification results obtained from cloud-free optical images usually outperform those achieved solely with SAR data (e.g., Ortega et al. (2021)).

There are three primary categories of Optical-SAR fusion strategies. The first category, early fusion, involves stacking multispectral and SAR images together into a single tensor to

* Corresponding author

Model Name	Data Source	Base Architecture
CNN-OPT	Optical	Convolution
CNN-SAR	SAR	Convolution
TRA-OPT	Optical	Transformer
TRA-SAR	SAR	Transformer

Table 1. Single-modality models.

Figure 2 presents how the Encoder, Decoder, and Classifier blocks are organized for all single-modality optical (Figure 2a) and SAR (Figure 2b) models.

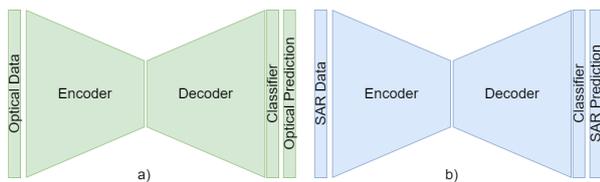


Figure 2. Single-modality models: a) (*)-OPT and b) (*)-SAR.

2.2 Multi-modality models

We considered three strategies for optical and SAR data fusion: early fusion, joint fusion, and late fusion, described by Table 2.

Model Name	Base Architecture	Fusion Strategy
CNN-EF	Convolution	Early Fusion
CNN-JF	Convolution	Joint Fusion
CNN-LF	Convolution	Late Fusion
TRA-EF	Transformer	Early Fusion
TRA-JF	Transformer	Joint Fusion
TRA-LF	Transformer	Late Fusion

Table 2. Single-modality models.

Figure 3 presents how the Encoder, Decoder, and Classifier blocks are organized for all multi-modality models. In the early fusion strategy (Figure 3a), the optical and SAR data are concatenated in the third dimension before the encoder. The encoder, decoder, and classifier setup is the same as in the single-modality models. Just one previous deforestation map was maintained because its information is independent of optical or SAR images. In the joint fusion strategy (Figure 3b), the model architecture has two independent encoder blocks. The encoder outputs (including the skip connections) are concatenated before entering the encoder. Each modal has an independent encoder and decoder in the late fusion strategy (Figure 3c). The decoder outputs are concatenated before entering the classifier block.

3. EXPERIMENTAL ANALYSIS

3.1 Study area and satellite images

We used Sentinel-1 and Sentinel-2 images acquired between 2018 and 2020 in an area of approximately 12,500 km² located in the southeastern portion of Amazonas State, Brazil (Figure 4). We chose this area because of the high deforestation rates and the availability of optical images with diverse cloud coverage conditions. We downloaded all data from the Google Earth Engine (GEE). The GEE Sentinel-1 data includes Level-1 Ground Range Detected (GRD) images with VV and VH polarizations. The image is processed by GEE using the Sentinel-1 Toolbox following the steps: thermal noise removal, radiometric calibration, and terrain correction using SRTM 30. The GEE

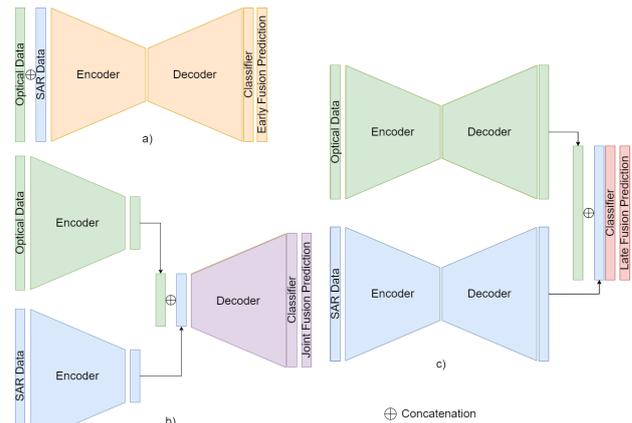


Figure 3. Multi-modality models: a) (*)-EF, b) (*)-JF, and c) (*)-LF.

also provides the Level-1C Sentinel-2 product, in which the images are radiometric and geometric corrected. All 13 bands were resampled to a resolution of 10m when necessary Gorelick et al. (2017).

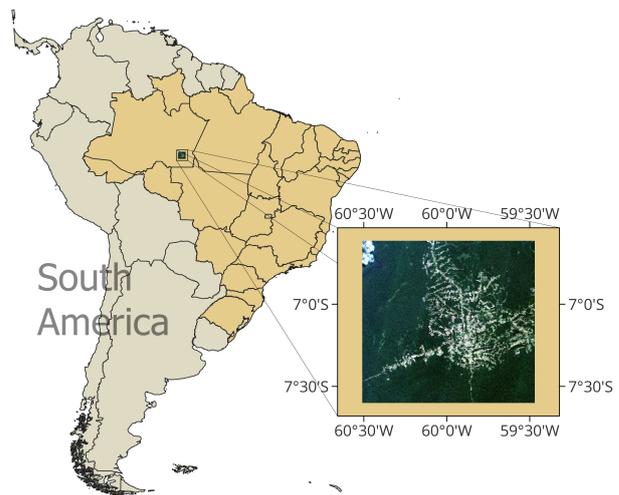


Figure 4. Study Area.

3.2 Data

As reference data, we used the deforestation polygons produced by PRODES, which can be accessed through the TerraBrasilis platform Assis et al. (2019). PRODES employs trained photo-interpreters to manually identify the annual increment of the deforested areas on optical satellite images. Only deforestation areas larger than 6.25 ha where the primary forest was entirely removed are outlined. For visual interpretation, the specialists select images with a reduced cloud cover (usually cloud-free) and acquired within the dry season (July to September) INPE (2022).

We derived the pixel-wise labels for the whole study area for each pair of years (Y_0 - Y_1) from PRODES deforestation polygons. We labeled the image pixels as *No Deforestation*, *Deforestation*, and *Background*. The pixels in which no deforestation was detected by PRODES until Y_1 were classified as *No Deforestation*. The pixels in which new deforestation areas were identified in Y_1 were classified as *Deforestation*. The pixels in which new deforestation areas were identified before Y_0 were

classified as *Background*. Pixels that are classified as "Hydrography" and "No Forest" (or other classes in which deforestation is impossible to occur) by PRODES were also classified as *Background*. Due to differences in the spatial resolution between the images used to generate the PRODES polygons, Landsat images with 30 m spatial resolution, and this work, Sentinel with 10 m spatial resolution, a buffer with 3 pixels was applied to *Deforestation* polygons.

PRODES usually employs one Landsat image per year to detect the deforestation increment areas. Henceforth, the date this image was taken will be called the PRODES reference date. In a ± 15 days window from each PRODES reference date, we selected three optical images with diverse cloud conditions and three SAR images. Figure 5 shows how we chose the images for this work, where R_i , and I_i^j means the PRODES reference date of the year i , and j -th optical-SAR image pair from Sentinel-2 and Sentinel-1, respectively. The images from the first two years ($Y_0 = 2018$ and $Y_1 = 2019$) were used for training purposes, while those from the last two years ($Y_0 = 2019$ and $Y_1 = 2020$) were used for testing.

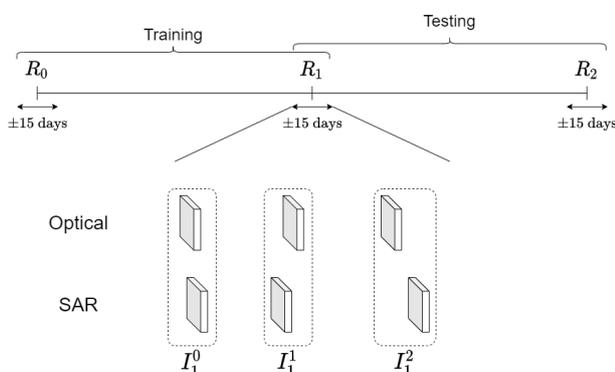


Figure 5. Reference dates and the selected images.

We are already aware of past deforestation through PRODES, and this knowledge can aid in identifying future deforestation areas. One way to represent this previous deforestation information is through a temporal distance map called previous deforestation map, in which the pixels' values vary from 0 to 1. If no previous deforestation occurred in a pixel, its value is 0. However, if a pixel represents a previously deforested area, its value will be based on the year of the deforestation identification. For example, if deforestation occurred in the year Y_0 , the pixel value in the deforestation map is 1. This value decreases linearly as the year of deforestation moves away from Y_0 towards the past.

We estimated the cloud coverage probability of each pixel in every optical image by utilizing the Sen2Cor algorithm (Louis et al., 2016). This method generated a cloud coverage probability map, which provided values ranging from 0 (no cloud) to 1 (cloud). For each pair of optical images used for testing ($Y_0 = 2019$ and $Y_1 = 2020$), we determined the maximum value between the respective cloud coverage probability maps, resulting in the Maximum Cloud Probability (MCP), for each pixel. MCP values will be employed to classify how each pixel was affected by clouds.

3.3 Experimental protocol

All procedures of this work were conducted with an Intel Core i9-10900F processor, with 128 GB of RAM and a GeForce RTX

3090 GPU with 24 GB of dedicated memory and Python language and PyTorch library.

The Transformer-based and Convolution-based architectures have the C and D , respectively, arbitrary values, which affect the number of trainable parameters. Our study set $C = 96$ and $D = 32$, following Cao et al. (2021) and Ferrari et al. (2023) respectively.

3.3.1 Training protocol The entire study area was divided into training and validation tiles, as shown in Figure 6. Subsequently, we extracted patches of 128×128 pixels from these tiles, with a 70% overlap.



Figure 6. Distribution of the training (blue) and validation (green) tiles over the study area.

To reduce the imbalance inherent to the deforestation detection task, we classified the generated patches based on the presence of pixels belonging to the *Deforestation* class. The patches with at least 2% of *Deforestation* pixels were regarded as *Deforestation Presence* patches, while the rest were as *Deforestation Absence*. To train and validate the models, we take the same number of *Deforestation Absence* as the *Deforestation Presence* patches, randomly discarding the excess of *Deforestation Absence* patches.

We trained five models from scratch for each architecture described in Tables 1 and 2. The 9 possible combinations between $\{I_0^0, I_0^1, I_0^2\}$ and $\{I_1^0, I_1^1, I_1^2\}$ were utilized to extract the patches, increasing the diversity of the images saw by the models during the training process. We chose the Adam optimizer with a learning rate $5 \cdot 10^{-5}$. The loss function was the weighted cross entropy, with the weights shown in Table 3. Each model was trained until the validation loss stopped to improve for 10 epochs.

Class	Weight
<i>No Deforestation</i>	0.2
<i>Deforestation</i>	0.8
<i>Background</i>	0.0

Table 3. Classes' weights.

We performed the prediction from 9 possible combinations of the Testing images: $\{I_1^0, I_1^1, I_1^2\}$ and $\{I_2^0, I_2^1, I_2^2\}$. To minimize the patch effect in the outcome, the prediction was generated multiple times for each model, with different overlapping (0.15, 0.2, 0.25, 0.3) between the patches, and the 8 pixels border of all sides were discarded. The final architecture's probability prediction was the average from all models and overlaps. We adopted 0.5 as the probability threshold for assigning the pixel to one of the two classes.

3.3.2 Evaluation protocol To evaluate each architecture in diverse cloud presence conditions, the MCP was estimated from

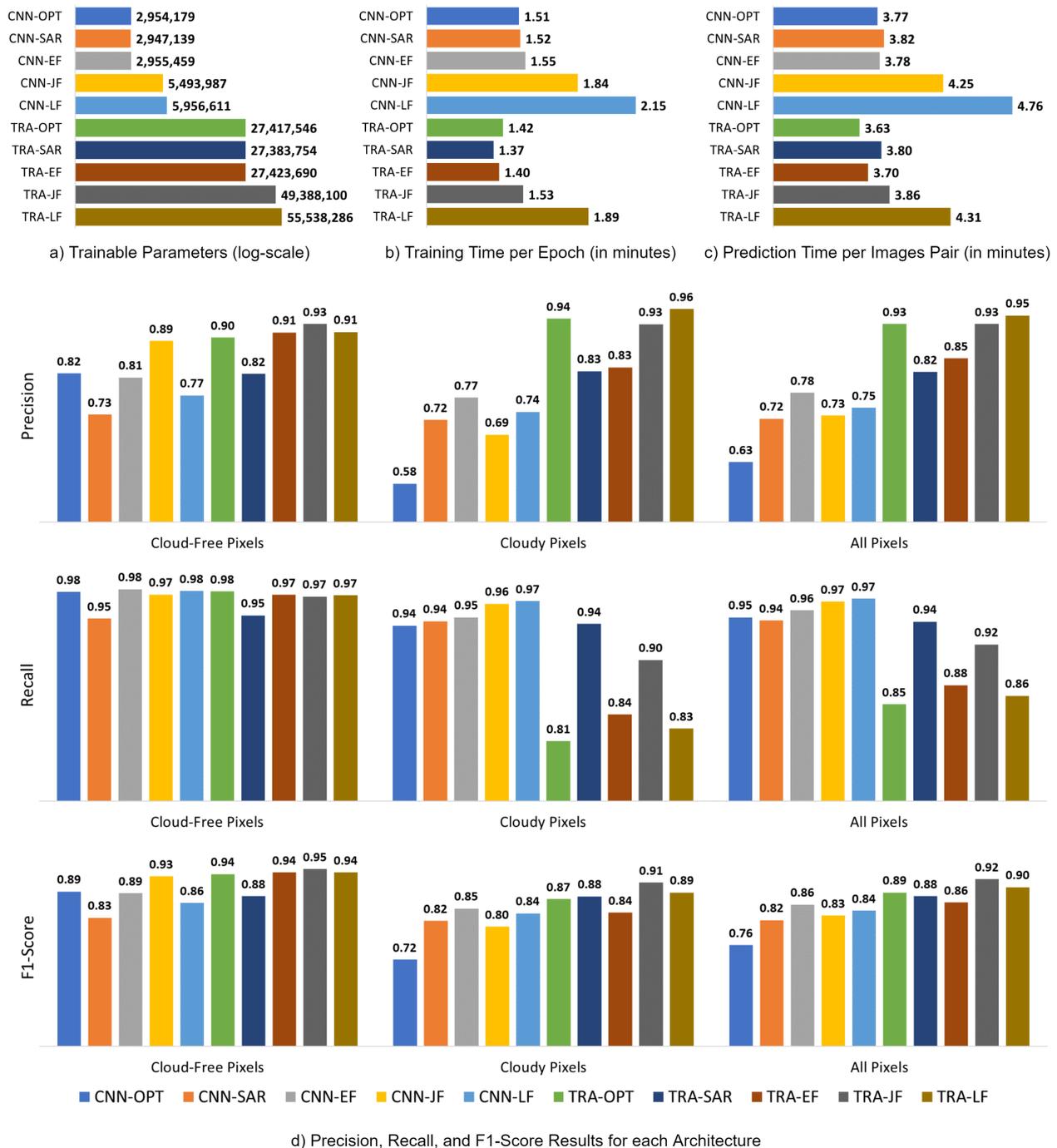


Figure 7. Models parameters (a) and the training (b) and prediction (c) times, with the respective metrics' results (d).

the images $\{I_1^0, I_1^1, I_1^2\}$ and $\{I_2^0, I_2^1, I_2^2\}$. If the MCP of each pixel was ≥ 0.5 , then it was classified as *Cloudy Pixel* because its optical image should be affected by clouds in Y_1 or Y_2 . The remaining pixels were classified as *Cloud-Free Pixels*.

We evaluated the precision, recall, and F1-Score considering each architecture classified prediction, discarding all *Deforestation* predictions with areas smaller than 6.25 hectares to be compatible with labels criteria generated from the PRODES data. All pixels belonging to the class *Background* were also discarded in the accuracy computation.

3.4 Results and Discussion

Figures 7a, 7b, and 7c present the number of trainable parameters of each architecture, and the training and prediction times, respectively. The Convolution-based networks have fewer trainable parameters than the respective Transformer-based. However, this difference didn't manifest in greater training and prediction times.

The number of trainable parameters of the Convolution-based architectures was lower than the Transformer-based. However, this difference in the number of trainable parameters didn't manifest in the training and prediction times, which we expected to be much higher. This result indicates that, despite

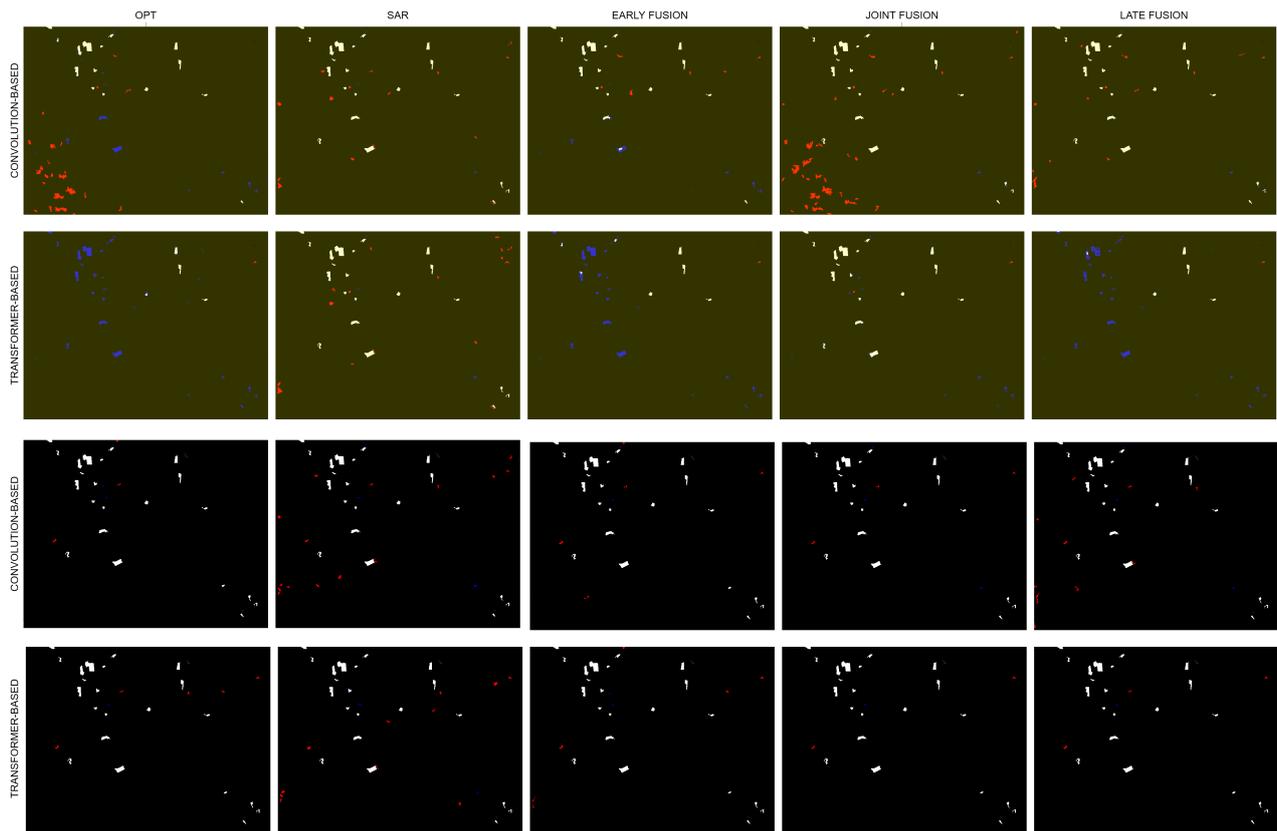


Figure 8. Error maps from the same area in diverse cloud conditions, in which white, black, red, and blue represent the true positive, true negative, false positive, and false negative, respectively. The pixels classified as *Cloudy Pixel* are highlighted in yellow.

the higher number of parameters of the Transformer-based architectures, these networks are more viable to be employed in large areas like the Brazilian Amazon Forest than the respective Convolution-based.

Figure 7d presents the accuracy metrics for each architecture. The results are organized based on the cloud-coverage classification, which classified each pixel in *Cloudy Pixel* or *Cloud-Free Pixels*, whether the pixel is affected by clouds or not, respectively. The results of the set of all pixels were also presented as *All Pixels*.

The Convolution-based models employing optical data (CNN-OPT) presented an expected behavior, in which poor results were observed for *Cloudy Pixels* compared to *Cloud-Free Pixels*, considering all evaluated metrics. However, the Transformer-based optical model (TRA-OPT), which presented fewer positive predictions when affected by clouds, reached slightly better precision results for the *Cloudy Pixels*, but the recall results were highly affected.

All models using SAR data were almost unaffected by cloud conditions. The Transformer-based SAR model (TRA-SAR) outperformed the Convolution-based counterpart (CNN-SAR) for all evaluated metrics.

The early fusion strategy did not improve the results for the Convolution-based model (CNN-EF) in *Cloud-Free Pixels*, in comparison to the respective optical (CNN-OPT). However, it achieved better results in the *Cloudy Pixels* compared to the respective SAR model (CNN-SAR). The Transformer-based early fusion (TRA-EF) presented similar results to the respective Convolution-based in *Cloud-Free Pixels* but presented

worse results than the respective SAR model (TRA-SAR) in *Cloudy Pixels*.

The last two fusion strategies, joint fusion, and late fusion, presented the best F1-Score results, with the respective Transformer-based models outperforming the Convolution-based architecture. Considering diverse pixel cloud conditions, TRA-JF presented better F1-Score results, while TRA-LF presented slightly lower F1-Score. However, the low recall of TRA-LF (0.86) indicates that this model failed in identifying new deforestation areas.

Figure 8 presents examples of error maps from two regions in the study area, where white, black, red, and blue represent the true positive, true negative, false positive, and false negative, respectively. The pixels classified as *Cloudy Pixel* are highlighted in yellow. For each region, the first row represents the Convolution-based models, while the second row represents the Transformer-based models' prediction errors. The first two rows (first region) refer to images fully affected by clouds, while the last two (second region) refer to a cloud-free region.

The error maps corroborate the metrics' results, especially for the models that presented low recall values but high precision results, showing that these models predicted fewer deforestation areas in the presence of clouds.

4. CONCLUSION

We investigated new fusion models, replacing the Convolution-based with Transformer-based models in diverse cloud coverage conditions.

In general, we expected that the cloud presence decreased the quality of the optical models' predictions. However, the TRA-OPT precision result was greater in *Cloudy Pixels* than in *Cloud-Free Pixels*. This behavior may be the result of a few positive predictions generated by TRA-OPT in the locations affected by clouds but may be investigated deeper in the future.

Despite the number of parameters increase, the Transformer-based models did not consume more time to train or generate the predictions than the Convolution-based ones. Very high training times, especially prediction times, could hamper the use of such models.

The results from this work indicates that using the transformer operation, in substitution of the convolution, may improve the deforestation detection capability of the models, independent of the cloud coverage condition. Investigations using other architectures are required to validate these findings.

Further investigations into the role of the information source (optical and SAR) in the fusion models may clarify its behavior in diverse cloud conditions.

The Transformer-based joint fusion strategy presented the best F1-score, raising as a suitable candidate for systems to detect new deforestation areas using optical and SAR images, outside the dry season, increasing its deforestation detection capability.

References

- Adarme, M. O., Feitosa, R. Q., Happ, P. N., Almeida, C. A. D., Gomes, A. R., 2020. Evaluation of Deep Learning Techniques for Deforestation Detection in the Brazilian Amazon and Cerrado Biomes From Remote Sensing Imagery. *Remote Sensing*, 12, 910. <https://www.mdpi.com/2072-4292/12/6/910>.
- Asner, G. P., 2001. Cloud cover in Landsat observations of the Brazilian Amazon. *International Journal of Remote Sensing*, 22, 3855-3862.
- Assis, L. F. F. G., Ferreira, K. R., Vinhas, L., Maurano, L., Almeida, C., Carvalho, A., Rodrigues, J., Maciel, A., Camargo, C., 2019. TerraBrasilis: A Spatial Data Analytics Infrastructure for Large-Scale Thematic Mapping. *ISPRS International Journal of Geo-Information*, 8, 513. <https://www.mdpi.com/2220-9964/8/11/513>.
- Baccini, A., Walker, W., Carvalho, L., Farina, M., Sulla-Menashe, D., Houghton, R., 2017. Tropical forests are a net carbon source based on aboveground measurements of gain and loss. *Science*, 358(6360), 230–234.
- Belward, A. S., Skøien, J. O., 2015. Who launched what, when and why; trends in global land-cover observation capacity from civilian earth observation satellites. *ISPRS Journal of Photogrammetry and Remote Sensing*, 103, 115–128.
- Benedetti, A., Picchiani, M., Frate, F. D., 2018. Sentinel-1 and Sentinel-2 data fusion for urban change detection. *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 1962–1965.
- Bouvet, A., Mermoz, S., Ballère, M., Koleck, T., Toan, T. L., 2018. Use of the SAR Shadowing Effect for Deforestation Detection with Sentinel-1 Time Series. *Remote Sensing*, 10, 1250. <http://www.mdpi.com/2072-4292/10/8/1250>.
- Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M., 2021. Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation.
- de Bem, P. P., de Carvalho, O. A., Guimarães, R. F., Gomes, R. A. T., 2020. Change detection of deforestation in the Brazilian Amazon using landsat data and convolutional neural networks. *Remote Sensing*, 12.
- Doblas, J., Shimabukuro, Y., Sant'Anna, S., Carneiro, A., Aragão, L., Almeida, C., 2020. Optimizing Near Real-Time Detection of Deforestation on Tropical Rainforests Using Sentinel-1 Data. *Remote Sensing*, 12, 3922. <https://www.mdpi.com/2072-4292/12/23/3922>.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N., 2020. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. <http://arxiv.org/abs/2010.11929>.
- Ebel, P., Saha, S., Zhu, X. X., 2021. Fusing Multimodal Data for Supervised Change Detection. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B3-2021, 243–249. <https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLIII-B3-2021/243/2021/>.
- Ferrari, F., Ferreira, M. P., de Almeida, C. A., Feitosa, R. Q., 2023. Fusing Sentinel-1 and Sentinel-2 Images for Deforestation Detection in the Brazilian Amazon Under Diverse Cloud Conditions. *IEEE Geoscience and Remote Sensing Letters*, 20, 1-5.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*. <https://doi.org/10.1016/j.rse.2017.06.031>.
- INPE, 2022. Metodologia utilizada nos sistemas PRODES e DETER.
- Li, Y., Zhou, Y., Zhang, Y., Zhong, L., Wang, J., Chen, J., 2022. DKDFN: Domain Knowledge-Guided deep collaborative fusion network for multimodal unitemporal remote sensing land cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 186, 170-189. <https://linkinghub.elsevier.com/retrieve/pii/S0924271622000557>.
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B., 2021. Swin transformer: Hierarchical vision transformer using shifted windows. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.
- Louis, J., Debaecker, V., Pflug, B., Main-Knorn, M., Bieniarz, J., Mueller-Wilm, U., Cadau, E., Gascon, F., 2016. Sentinel-2 sen2cor: L2a processor for users. *European Space Agency, (Special Publication) ESA SP*, SP-740.
- Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B. A., 2019. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS journal of photogrammetry and remote sensing*, 152, 166–177.
- Ortega, M. X., Feitosa, R. Q., Bermudez, J. D., Happ, P. N., Almeida, C. A. D., 2021. Comparison of optical and SAR data for deforestation mapping in the Amazon rainforest with fully convolutional networks. *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, IEEE, 3769–3772.

- Rosa, L. E. C. L., Oliveira, D. A. B., Feitosa, R. Q., 2021. Investigating fusion strategies on encoder-decoder networks for crop segmentation using SAR and optical image sequences. *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, IEEE, 2405–2408.
- Silva, C. A., Guerrisi, G., Frate, F. D., Sano, E. E., 2022. Near-real time deforestation detection in the Brazilian Amazon with Sentinel-1 and neural networks. *European Journal of Remote Sensing*, 55, 129-149. <https://www.tandfonline.com/doi/full/10.1080/22797254.2021.2025154>.
- Stahlschmidt, S. R., Ulfenborg, B., Synnergren, J., 2022. Multimodal deep learning for biomedical data fusion: a review. *Briefings in Bioinformatics*, 23.
- Strand, J., Soares-Filho, B., Costa, M. H., Oliveira, U., Ribeiro, S. C., Pires, G. F., Oliveira, A., Rajao, R., May, P., van der Hoff, R. et al., 2018. Spatially explicit valuation of the Brazilian Amazon forest's ecosystem services. *Nature Sustainability*, 1(11), 657–664.
- Torres, D. L., Turnes, J. N., Vega, P. J. S., Feitosa, R. Q., Silva, D. E., Junior, J. M., Almeida, C., 2021. Deforestation Detection with Fully Convolutional Networks in the Amazon Forest from Landsat-8 and Sentinel-2 Images. *Remote Sensing*, 13, 5084. <https://www.mdpi.com/2072-4292/13/24/5084>.
- Zhang, Z., Liu, Q., Wang, Y., 2018. Road Extraction by Deep Residual U-Net. *IEEE Geoscience and Remote Sensing Letters*, 15, 749-753. <http://ieeexplore.ieee.org/document/8309343/>.