DEFORESTATION DETECTION IN THE AMAZON WITH SENTINEL-1 SAR IMAGE TIME SERIES

Kaan Karaman*, Vivien Sainte Fare Garnot, Jan Dirk Wegner

Institute for Computational Science, University of Zurich, CH–8057 Zurich, Switzerland
Emails: kaan.karaman@uzh.ch, vsaint@ics.uzh.ch, jandirk.wegner@uzh.ch

KEY WORDS: Attention Mechanism, Deforestation Detection, Multi-Temporal Dataset, Synthetic Aperture Radar.

ABSTRACT:

Deforestation has a significant impact on the environment, accelerating global warming and causing irreversible damage to ecosystems. Large-scale deforestation monitoring techniques still mostly rely on statistical approaches and traditional machine learning models applied to multi-spectral, optical satellite imagery and meta-data like land cover maps. However, clouds often obstruct observations of land in optical satellite imagery, especially in the tropics, which limits their effectiveness. Moreover, statistical approaches and traditional machine learning methods may not capture the wide range of underlying distributions in deforestation data due to limited model capacity. To overcome these drawbacks, we apply an attention-based neural network architecture that learns to detect deforestation end-to-end from time series of synthetic aperture radar (SAR) images. Sentinel-1 C-Band SAR data are mostly independent of the weather conditions and our trained neural network model generalizes across a wide range of deforestation patterns of Amazon forests. We curate a new dataset, called BraDD-S1TS, comprising approximately 25,000 image sequences for deforested and unchanged land throughout the Brazilian Amazon. We experimentally evaluate our method[†] on this dataset and compare it to state-of-the-art approaches. We find it outperforms still-in-use methods by 13.7 percentage points in intersection over union (IoU). We make BraDD-S1TS[‡] publicly available along with this publication to serve as a novel testbed for comparing different deforestation detection methods in future studies.

1. INTRODUCTION

Deforestation is one of the most significant causes of climate change (Shukla et al., 1990, Heidari et al., 2021). Human activities such as farming, mining, and logging in forest regions are responsible for a majority of deforestation, causing even more harm than natural disasters. The rate of deforestation is alarming and a large portion is caused by the enormous demand for arable land by the food industry, especially in the Brazilian Amazon (Nepstad et al., 2008). To monitor deforestation, the Brazilian government implements a few deforestation alert systems based on remote sensing imagery, including *PRODES* and *DETER* (Assis et al., 2019). The existing systems primarily rely on optical remote sensing imagery acquired by airborne or space-borne platforms like Landsat. Their effectiveness is, however, limited by heavy cloud coverage during the wet season in the Amazon. An alternative that is largely independent of cloud coverage is Synthetic Aperture Radar (SAR) data. Especially for rapid-alert systems that strive for detecting deforestation as early as possible, SAR data acquired by satellite missions with high-revisit rates like the Sentinel-1 C-Band sensor (Torres et al., 2012) provide a promising source of evidence. In this paper, we aim to take an initial step toward developing a deforestation detection system that can ultimately assist governmental organizations and other stakeholders in safeguarding forested areas. By providing timely alerts before deforestation reaches an irreversible level, our system can aid in preventing irreparable damage to the environment.

The major challenge of this task is that SAR data is susceptible to the speckle effect and exhibits effects like shadowing, layover, and foreshortening (Moreira et al., 2013). It is, thus,

often easier for human annotators to work with optical satellite imagery for mapping deforestation as in the case of *PRODES* and *DETER*. While SAR images are available for the entire Brazilian Amazon (and beyond), existing label masks annotated by experts are usually based on optical images. This does not only introduce slight geometric differences due to the different imaging principles of optical and SAR sensors but also a time gap between the manual annotation of the ground truth label masks and the acquisition of the SAR image. A deforested region can only be labeled manually in an optical satellite image under cloud-free conditions whereas a C-Band SAR image can collect evidence immediately through clouds. This misalignment between often late labels from optical imagery and early SAR image acquisition can cause *label noise*, which negatively impacts model optimization and detection performance (Kendall and Gal, 2017). Another challenge is that the time gap between consecutive Sentinel-1 acquisitions does vary for different locations. To cope with these challenges, we propose an attention-based neural network approach for detecting deforestation from the time series of Sentinel-1 SAR images. Our method builds on the U-TAE work of (Garnot and Landrieu, 2021), and adapts it to deforestation detection with SAR imagery. The U-TAE model incorporates an attention mechanism to encode the non-uniformly sampled temporal information and combines it with a U-Net architecture to extract semantic information in the images. This allows us to simultaneously learn deforestation patterns over space and time. Although our approach does not have the ability to detect deforestation in near-real time, our study takes a first step in this direction. It explores how longer time series of more than 50 images per sequence can benefit deforestation detection.

To the best of our knowledge, no publicly available deforestation detection dataset exists in the literature that covers a large area (i.e., the whole Amazon), is multi-temporal, and consists

---

* Corresponding author
† https://github.com/kaankaramanofficial/BraDD-S1TS
‡ https://zenodo.org/record/8060250

of Sentinel-1 SAR data. We thus collect the first multi-temporal Sentinel-1 SAR dataset based on *PRODES* alerts (Assis et al., 2019) for deforestation detection to serve as a testbed for our method and to compare it to related work. We publicly release it to serve as a new benchmark dataset to facilitate comparisons between different methods. Our experimental results demonstrate that the attention-based U-TAE architecture outperforms previous methods and sets a new state-of-the-art. Additional experiments with *temporal dropout* that reduces temporal information during the training phase indicate the crucial role of longer satellite image time series for accurate deforestation detection. In summary, our contributions are the following:

- Building on the attention-based U-TAE neural network architecture of (Garnot and Landrieu, 2021), we propose a deforestation detection approach that relies on Sentinel-1 SAR image time series. Experimental evaluation shows that our U-TAE approach outperforms existing ones for detecting deforestation.

- We introduce a new multi-temporal dataset that consists of Sentinel-1 SAR images and ground truth masks derived from the *PRODES* project. This dataset called Brazilian Amazon Deforestation Dataset with Sentinel-1 Time Series (BraDD-S1TS) includes $\sim 25,000$ image sequences. Each image sequence consists of ~ 50 images acquired over a period of 1 year and the corresponding binary deforestation masks. Images in the sequence have size 48×48 pixels, covering 480×480 m^2 area. We make this dataset publicly available such that it can serve as a new testbed for future research on deforestation detection using SAR imagery.

The structure of this paper is as follows: We review related work in Section 2 and define the task and explain our proposed method in Section 3. Section 4 describes the properties of the newly collected BraDD-S1TS dataset in detail. Experiments are presented in Section 5 and we finally draw conclusions in Section 6.

2. RELATED WORK

A large body of literature exists for change detection in remote sensing images and more specifically, for deforestation detection and mapping. While a full review of all existing research is beyond the scope of this paper, we provide examples of the most relevant research related to ours below.

Deforestation detection with traditional machine learning
Two of the most prominent and widely used approaches in practice for deforestation detection at a very large scale using sequences of satellite images are GLobal Analysis and Discovery (GLAD) (Hansen et al., 2016) and RAdar for Detecting Deforestation (RADD) (Reiche et al., 2021). Both methods use traditional machine learning techniques and compare changes in forests over multiple consecutive images. GLAD works with optical data from Landsat, whereas, RADD is based on Sentinel-1 SAR images as input, like our approach. Because RADD is based on SAR images as input, it works independently of cloud coverage and is thus capable of extracting evidence over denser time series per region of interest. This is a significant improvement over the Landsat-based GLAD alerts, especially in the Amazon forest, which is cloudy very frequently (Reiche et al., 2021). Both GLAD and RADD alerts are probabilistic models based on Bayesian Rule, as described in (Reiche

et al., 2015). A Gaussian mixture model is computed for each pixel, and the time information is taken into account by moving step-by-step along the temporal dimension. In addition to the core probabilistic model, there are extensive pre- and post-processing steps and seasonality removal techniques that significantly improve performance (see Subsection 5.3). Both alert systems are commonly used for worldwide deforestation detection and up-to-date alerts can be found on Global Forest Watch.

Deep learning-based deforestation detection Modern deep learning approaches for bi-temporal change comparing one remote sensing observation before the deforestation event, and one afterward, are compared in (De Bem et al., 2020). The authors train several deep-learning-based models on a dataset of optical satellite images and achieved better performance compared to classical approaches. In another, similar study (Ortega Adarme et al., 2020) several approaches using multispectral Landsat images are compared for deforestation detection in the Brazilian Amazon. The models in this paper also use bi-temporal samples as inputs to predict binary deforestation masks. It draws a similar conclusion as (De Bem et al., 2020), which deep-learning-based models significantly exceed the performance of the standard machine learning approaches. Recently, (Cherif et al., 2022) proposes to use longer image sequences and multi-class data. The authors apply multiple common deep learning approaches such as U-Net and DeepLab on combined optical and SAR data of the Sentinel-1 and Sentinel-2 missions. They propose two merging techniques for combining multi-modal time series and obtain better outcomes for rare classes in their dataset. These findings highlight the potential benefits of utilizing longer image time series to leverage the performance of classification on remote sensing data.

Except (Cherif et al., 2022) (which mostly investigates the optical-SAR fusion methods), there is only limited research in deforestation detection using deep learning models that use more than two images according to surveys such as (Gong et al., 2016, You et al., 2020, Khelifi and Mignotte, 2020, Shi et al., 2020, Jiang et al., 2022). The primary reasons are the large dataset size when using more than two images per scene and the computational costs associated with managing longer time series of satellite images. The acquisition of more evidence over time of the same location, however, does potentially enable the reduction of noise and ultimately more accurate results. We explore this idea in this paper and build on modern deep-learning approaches for satellite image sequence analysis.

According to a survey (Shi et al., 2020), there is still a lack of openly available large SAR datasets in the literature. Although there are a few numbers of works on forest monitoring, the datasets exploited for training the models vary from study to study. It is therefore difficult to make a fair comparison between different studies. To address this issue, we create a new SAR dataset for the deforestation detection task whose samples are selected from the whole Brazilian Amazon region. It should be highlighted that this new dataset also fills the gap in publicly available datasets for multi-temporal change detection tasks from SAR images in remote sensing.

Satellite time series analysis via deep learning Our research presented in this paper is inspired by recent advances in crop classification using multi-temporal remote sensing data. One of the first studies that proposed modern deep learning for crop classification from satellite image sequences is (Rußwurm and Körner, 2017). The authors design an LSTM to learn temporal

evidence from optical, Sentinel-2 data. The same authors extend their approach in (Rußwurm and Körner, 2018) and introduce a convolutional-LSTM and as well as GRU directly learn to filter out redundant, cloudy images from the input data. An alternative approach (Pelletier et al., 2019) proposes Temporal Convolutional Neural Networks (Temp-CNNs) to learn temporal information through convolutions from Formosat-2 multi-spectral optical image series and get better performance than RNN-based models. Another more recent approach (Turkoglu et al., 2022), however, with an RNN at its core proposes the STAckable Recurrent cell (STAR) architecture to reduce the number of trainable parameters in deep learning-based models compared to standard LSTM and GRU designs, and improves the performance. Modern self-attention methods are explored in the works of (Rußwurm and Körner, 2020, Garnot et al., 2020). In (Rußwurm and Körner, 2020), various mechanisms are compared on the same optical dataset for crop type identification, and the transformer-based method outperforms RNN-based approaches. Furthermore, (Garnot et al., 2020) merges the point-set encoder with the transformer and used it for crop classification. This approach achieves a new state-of-the-art result over the standard implementation of the transformer.

Recent studies propose solutions such as using Ordinary Differential Equations (ODEs) with RNNs to address the issue of irregular temporal spacing (Metzger et al., 2021). Another promising solution to this problem is the U-Net-inspired Temporal Attention Encoder (U-TAE) proposed by (Garnot and Landrieu, 2021), which employs a multi-head attention mechanism to process temporal information and U-Net architecture (Ronneberger et al., 2015) for spatial encoding. By leveraging transformers, the proposed U-TAE model is able to extract rich temporal features, while being computationally more efficient than RNNs that require sequential processing. Furthermore, its architecture inspired by U-Net can efficiently encode spatial information image-wise, making it possible to achieve state-of-the-art performance on the public PASTIS dataset while using fewer computational resources compared to RNN-based approaches. Overall, the combination of transformers and U-Net architecture in U-TAE results in both improved performance and resource efficiency.

In this work, we build on the recently proposed U-TAE approach of (Garnot and Landrieu, 2021) and evaluate the model's performance on a new task of deforestation detection using a newly collected Sentinel-1 SAR image time series dataset with a different data distribution than the optical one that U-TAE was tested previously. Our primary motivation for utilizing U-TAE is to leverage its temporal and spatial encoding capabilities while minimizing memory consumption, which is particularly critical when designing a model for the entire Amazon region or for the whole globe.

3. METHOD

We frame the problem of deforestation detection as a semantic segmentation task applied to a time series of SAR images. Our model takes a sequence of Sentinel-1 SAR satellite images from a specific area as input and produces a binary spatial mask, with each output pixel value indicating whether the region is deforested or not. Time intervals between consecutive images are irregularly sampled because the revisit frequency of the Sentinel-1 satellites varies between 6 to 12 days depending on the location.

More formally, let $D = \{(x_i, t_i, y_i)\}_{i=1}^{N}$ be a dataset which contains N samples. Each $x_i \in \mathbb{R}^{T \times C \times H \times W}$ denotes pixel values of the SAR image time series and $y_i \in \{0, 1\}^{H \times W}$ denotes the label mask of the sample i. T, C, H, and W are the number of time steps, channels (VV and VH polarization for SAR images), height, and width of an image at a time step, respectively. For evaluation on the patch level (image-level), the patch label mask $y \in \{0, 1\}$ takes on values 1 for a deforested (changed) region and 0 for undisturbed (unchanged) land cover.

In addition to the images, we utilize the temporal coordinates of the images to compute the positional encoding of the attention mechanism. Following (Garnot and Landrieu, 2021), we express the dates as $t_i \in \mathcal{Z}_{>0}^{T}$, which represents the relative day difference between the acquisition dates of the images in x_i. Note that the acquisition intervals vary due to different revisit rates of the Sentinel-1 mission, as explained before.

We use the U-TAE architecture as introduced by (Garnot and Landrieu, 2021) as the model defined by $f_\theta(\cdot, \cdot)$. The U-TAE is based on the U-Net architecture (Ronneberger et al., 2015), which consists of an encoder and a decoder for spatial processing. Each encoder block reduces the feature maps by a factor of two, and each decoder block doubles them. For our deforestation detection task, we adopt the U-TAE structure of (Garnot and Landrieu, 2021), where the network has four decoder and encoder blocks. The output feature map is predicted in the same spatial size as the input image. Multi-head self-attention blocks are employed to extract temporal information from the unevenly sampled satellite image series in addition to the spatial encoding.

Consequently, we obtain the output probability map $\hat{y} \in [0, 1]^{H \times W}$ along with the attention masks $A \in \mathbb{R}^{T \times m \times h \times w}$, where h and w are the height and width of the encoded feature space. A pixel at (i, j) has a probability $\hat{y}_{(i,j)}$ of belonging to the deforestation region. In our setup, $h = H/2^4$ and $w = W/2^4$, since we have four blocks for the encoder and m represents the number of heads in the temporal encoder.

We train our model $f_\theta(\cdot, \cdot)$ using a loss function $\mathcal{L}(\cdot, \cdot)$. This loss function takes the ground truth labels y and predicted labels $\hat{y}$ as inputs. Instead of a standard cross-entropy loss, we opt for a focal loss to deal with the imbalanced class distribution, i.e., our dataset is dominated by unchanged land cover and only a small portion of all pixels covers deforestation events. The formula of α-balanced focal loss (Lin et al., 2017) is shown in Equation 1, where α and γ are hyperparameters.

$$
\begin{aligned}
\mathcal{L}_{FL}(y, \hat{y}) = & -y \cdot \alpha \cdot (1 - \hat{y})^{\gamma} \cdot \log(\hat{y}) \\
& -(1 - y) \cdot (1 - \alpha) \cdot \hat{y}^{\gamma} \cdot \log(1 - \hat{y}).
\end{aligned}
\tag{1}
$$

4. DATASET COLLECTION

As shown in Section 2, multi-temporal deforestation detection from radar time series lacks a large-scale dataset benchmark. We introduce BraDD-S1TS to fill this gap. Our dataset contains $25,988$ Sentinel-1 image time series and their associated binary deforestation mask (see Figure 1). We use Ground Range Detected (GRD) observations at a 10×10 meters per pixel resolution. Each time series covers a patch of 48×48 pixels and contains 50 observations on average, all with the same descending orbit. Depending on the availability of Sentinel-1 platforms the actual number of observations varies between 19 and 63 dates (see Figure 2 for the distribution). BraDD-S1TS covers an
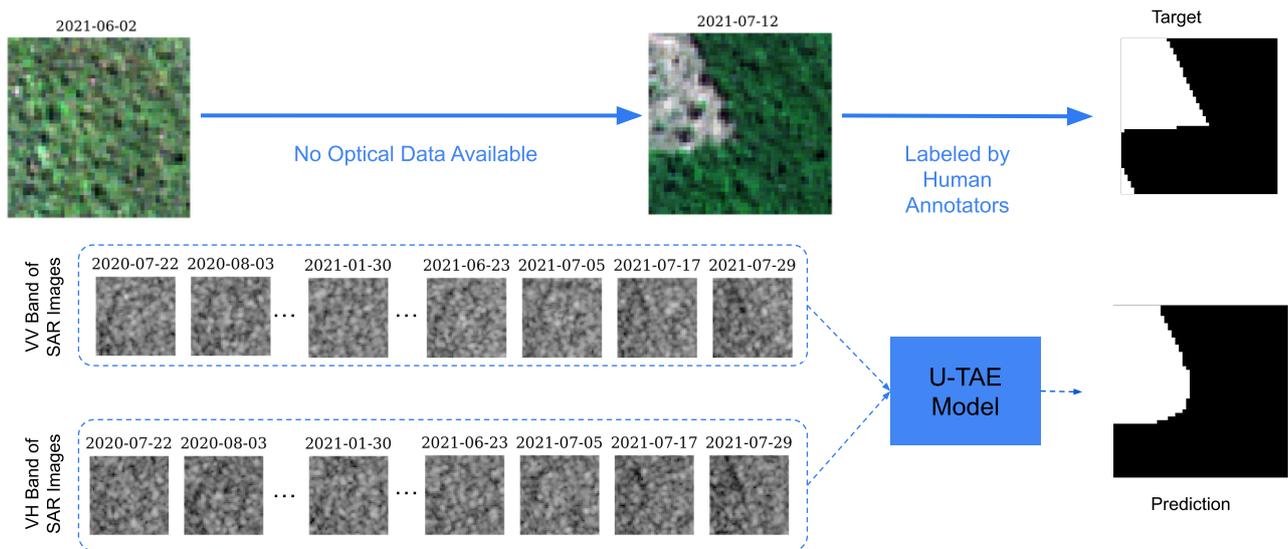
Figure 1. A random sample in the dataset is used for testing the models during the experiments. The optical data is only for visualization, it is not provided in BraDD-S1TS.

area of 57.5 km^2 in total in the Amazon rainforest, and contains $14,373$ different deforestation events from *PRODES* (Assis et al., 2019). The total size of the dataset is 17.7 GB and is freely available at the link. We further describe how we construct the dataset in the following paragraphs.
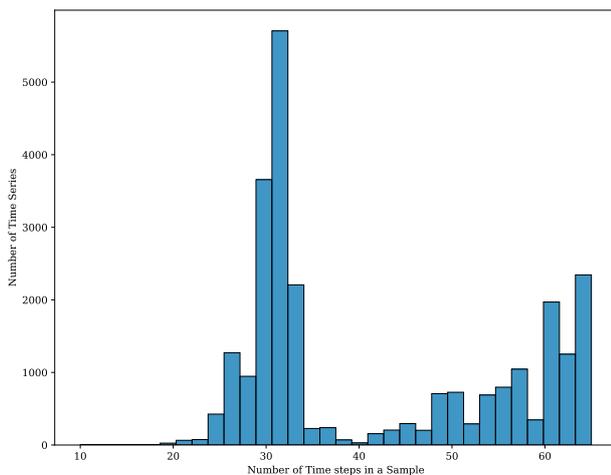

Figure 2. The distribution of the number of time steps in a sample.

Deforestation Labels We use the PRODES alerts as ground truth labels[1]. These alerts are annotated by human experts on optical satellite observations and range from 2008 to 2021. We focus on alerts raised between July 17, 2020, and November 12, 2021. Each alert consists of a geo-referenced polygon showing the extent of the observed deforestation, and the date at which the deforestation was *observed*. We show the distribution of the alert dates in Figure 3. Note that the alert dates cluster during the dry months of July and August. This is because cloud obstruction prevents labeling during the rainy season. As a result, a deforestation event happening during the rainy season is typically flagged with a long delay. We refer to this problem as the *temporal uncertainty* of the labels. The lack of precise temporal

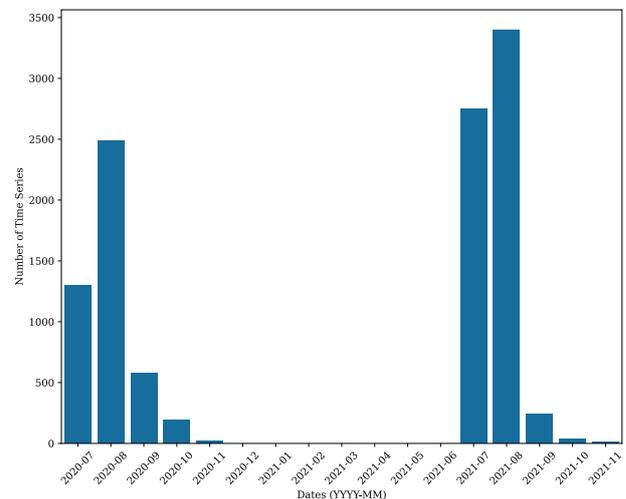[1] http://terrabrasilis.dpi.inpe.br/en/home-page/

Figure 3. The temporal distribution of the alert dates.

labels motivates us to frame the problem as a time series classification instead of deforestation date prediction, and we leave this more challenging setting for further research. Here, since the annotation campaigns of the *PRODES* mission are done on a yearly basis, we consider that an alert implies that deforestation happened at some point during the year preceding the alert.

Positive samples We randomly select positive patches of two types: First by randomly selecting a point within a *PRODES* polygon, and second by choosing a random point on the boundary of the polygon. These points then serve as centers of the 48×48 patches. This ensures that our dataset contains patches of completely deforested pixels, as well as mixed patches with both changed and non-changed pixels. In total, we select $8,712$ inner and $2,294$ boundary positive samples. This increases the diversity of the positive samples. In addition, the boundary patches may be affected by the shadowing effect of SAR which can be used by a detector. For all selected points, we make sure that the corresponding patch does not overlap with any other *PRODES* polygon to avoid confusing samples. For each positive patch, we collect the available Sentinel-1 observations start-

ing from 1 year and 2 weeks before the alert date until 2 weeks after it, with a 2-week error margin. This ensures that the deforestation event happened sometime between the first and last observation.

Negative samples We select three types of negative samples to make for a challenging benchmark:

- *Non-forest regions*: We select $1,953$ patches in locations corresponding to non-forest vegetation land cover. We use *Copernicus Land Cover Dataset* (Buchhorn et al., 2020) and select patches in the herbaceous, agricultural, and shrub land cover types.

- *Already-deforested regions*: Since the aim is to detect a change from forest to deforestation, we also include in our negative samples patches of already deforested areas. Specifically, we select $7,557$ patches within *PRODES* polygons and collect satellite observations starting at least 6 months after the corresponding alert.

- *Unchanged forest*: Lastly we include $5,472$ patches from undisturbed forest areas as identified by (Assis et al., 2019).

When selecting starting and ending dates for negative samples, we ensure that they are chosen randomly to maintain a similar distribution as for the positive samples, and thus avoid a dataset bias. For each negative sample, the time difference between the starting and ending date is always the same as that of the positive samples.

SAR data We use Sentinel-1 data in GRD format with 10-meter ground sampling distance. We utilize the data without any additional pre-processing beyond the standard ones applied by Google Earth Engine (GEE) (Gorelick et al., 2017), which includes four primary steps: (1) removal of border noise in GRD data, (2) removal of thermal noise, (3) radiometric calibration, and (4) terrain correction. The pixel values are downloaded in dB scale.

For each time series, we first identify all available SAR images from Sentinel-1 between the given start and end date. Then, we determine the most frequent relative orbit number in this sequence and collect the images only from this relative orbit
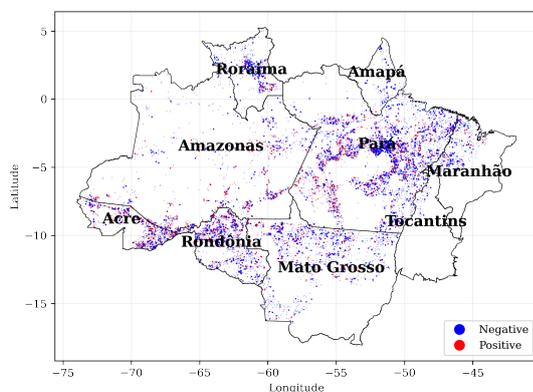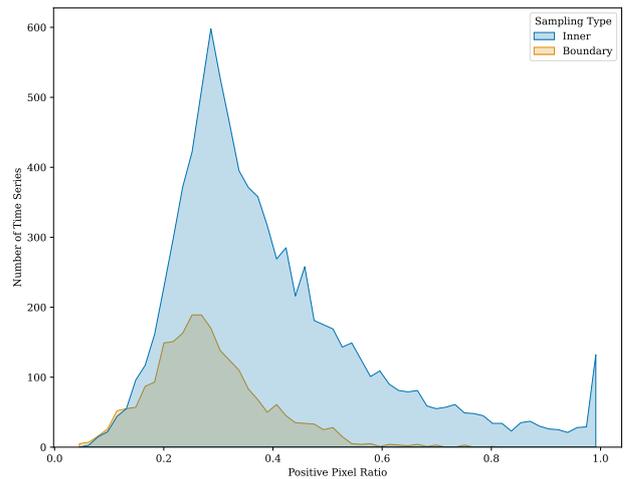

Figure 5. The positive pixel distribution in positive samples.

number in order to avoid potential problems resulting from different orbits. However, for some regions in our dataset, only one platform can acquire images, while for others, both platforms can obtain images. Furthermore, due to technical issues, the time intervals between the acquisition dates of the images are not even. Figure 2 illustrates the non-uniform distribution in our dataset.

In terms of spatial dimensions, we opt to use patches in the dataset that cover 480 meters by 480 meters, with each patch containing $H = W = 48$ pixels. Our primary concerns when choosing these numbers are to ensure that they are small enough to enable memory-efficient processing and large enough to utilize the surrounding pixels' information. In this context, we consider the channel as the polarization of SAR images, with the first channel denoting co-polarization (VV), and the second one representing cross-polarization (VH).

Dataset summary In summary, we collect a dataset on the Brazilian Amazon region, which covers 9 federal units in Brazil, as shown in Figure 4. The dataset comprises a total of $25,988$ samples, with $11,006$ positives and $14,982$ negatives, resulting in a positive sample ratio of 42.35%. The imbalance ratio at the pixel level, on the other hand, is 15.86%. The distribution of the ratio of positive pixels in positive samples is shown in Figure 5. Further information on the dataset can be found from the metadata in the link.

5. EXPERIMENTS

5.1 Implementation details

We split the BraDD-S1TS dataset into train ($15,625$ samples), validation ($5,251$ samples), and test ($5,112$ samples) sets, and keep it fixed across all our experiments. We build our model based on the official PyTorch implementation of U-TAE (Garnot and Landrieu, 2021) from `https://github.com/VSainteuf/utae-paps`, keeping the initial learning rate fixed at 10^{-3} and a weight decay of 10^{-6}. The model is trained using the *AdamW* optimizer (Loshchilov and Hutter, 2017), and we employ a scheduler of *ReduceLROnPlateau* over the *IoU* score on the validation set. We use a batch size of 4. The source code of our implementation has been made publicly available in this GitHub repository.

Figure 4. Distribution of negative (blue) and positive (red) BraDD-S1TS samples across the nine Brazilian states.

Following this, we implement the core *RADD* model based on the paper (Reiche et al., 2021) from scratch. This approach enables us to evaluate the results of *RADD* without pre- and post-processing explained in (Reiche et al., 2021), which we compare with other models that do not utilize such operations.

To further evaluate *RADD*, we also obtain results with pre- and post-processing for the geo-locations in our dataset from https://data.globalforestwatch.org/datasets/ gfw::deforestation-alerts-radd/about. These results are noted as *RADD+*. To ensure the fairness of our results, we only take into account the *RADD+* alerts for each sample by considering the start and end dates of the image sequence in that region. That is, we filter the downloaded alerts by date in the same range given to the other models. Because the other model can only process the time sequences between these dates.

5.2 Performance Metrics

We report Intersection of Union (IoU), Precision, and Recall of the positive (*deforestation*) class. We compute the metrics at pixel *and* patch level (image-wise). Indeed, the spatial imprecision of the *PRODES* polygons (mentioned in Section 1) can negatively impact the reported pixel-based performance metrics. We consider a patch to be positive when it contains at least one deforested pixel, and negative otherwise. Similarly, a predicted patch is positive if at least one pixel is predicted as positive. This way, the patch-based metric reflects how well the method detects a deforestation event, regardless of the precision of the pixel-level delineation of its extent.

5.3 Experimental results

Comparison to non-deep-learning baselines First, we benchmark the performance of U-TAE against existing approaches. Here, we train the default U-TAE configuration with different numbers of attention heads and with cross-entropy loss. To the best of our knowledge, the state-of-the-art for multi-temporal deforestation classification from radar data is based on the statistical approach of RADD alert. We also test the models of Conv-GRU, Conv-LSTM, and 3D-UNet (Rustowicz et al., 2019) from the crop-type mapping literature on our dataset.

	Method	IoU (%) ↑	P (%) ↑	R (%) ↑
Patch	RADD	27.9	39.6	48.5
	RADD+	57.0	72.4	76.1
	Conv-GRU	61.2	65.0	**91.3**
	Conv-LSTM	63.7	71.6	85.3
	3D-UNet	68.1	80.5	81.5
	U-TAE (1 head)	67.2	76.1	85.1
	U-TAE (2 head)	**70.7**	**82.4**	83.2
	U-TAE (8 head)	67.7	73.1	90.2
	U-TAE (16 head)	69.5	76.1	88.9
Pixel	RADD	1.7	6.1	2.3
	RADD+	19.7	75.1	21.1
	Conv-GRU	41.8	66.5	53.0
	Conv-LSTM	44.6	68.4	56.2
	3D-UNet	45.8	69.3	57.4
	U-TAE (1 head)	43.8	67.0	55.8
	U-TAE (2 head)	44.0	**75.4**	51.4
	U-TAE (8 head)	**47.3**	70.9	**58.6**
	U-TAE (16 head)	46.1	72.9	55.6

Table 1. Test scores of the methods. P and R stand for precision and recall, respectively.

We compare the performance of all these deep learning methods to this baseline in Table 1. The top nine rows show patch-level scores, while the bottom rows show pixel-level scores. *RADD* refers to the method applied to raw images without pre-processing, while *RADD+* refers to the final results from the RADD deforestation alert system with pre- and post-processing steps mentioned in (Reiche et al., 2021). Results in the table show that the U-TAE, even with only a single head, achieves a pixel IoU of 43.8, roughly twice that of the *RADD+* baseline and 10 times that of the *RADD* alert. Our findings demonstrate that deep learning modeling can greatly benefit deforestation classification from radar series. Despite the significant improvement over existing statistical-based approaches, the level of performance remains moderate for a binary classification problem. This highlights the fact that deforestation monitoring from the SAR time series is a challenging problem that should be further explored in future research.

At the patch level, U-TAE with a single head achieves a better IoU score of 67.2, significantly higher than the two non-deep-learning baselines. These results indicate that the model can be used to flag deforested regions on a coarse grid of 480×480 meters. Table 1 also displays results of the U-TAE with varying numbers of attention heads. For the sake of simplicity, we use only the U-TAE model with eight heads and pixel-level IoU scores in our further investigations.

Importance of the temporal dimension In this paper, we argue that leveraging the full-time series of SAR observations is crucial for better deforestation detection, as opposed to mono-temporal or bi-temporal change detection approaches. To investigate this hypothesis experimentally, we re-train the U-TAE model with varying numbers of input image patches over time. We start with training on only one image patch, i.e., the last available date in the time series. Second, we check for the bi-temporal case and train on only two image patches, i.e., the first and the last available observations. We then gradually increase the number of image patches per time series until all available observations are used. We report the pixel-level test IoU of these different models in Figure 6. This experiment shows that correctly predicting deforestation from a single SAR observation is extremely challenging (IoU= 3.3%). Bi-temporal deforestation detection reaches a significantly better performance of 36.0% IoU. The performance is then still increased significantly with longer time series, and seems to saturate around 10 dates.
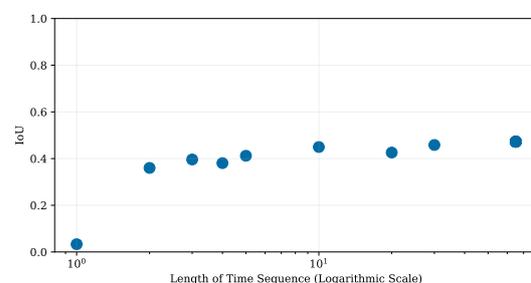

Figure 6. The graph of pixel-level scores versus different settings for temporal dropout. Note that the maximum length of time-series in BraDD-S1TS is 65.

Dealing with data imbalance Our dataset contains a slight imbalance of 15.86% (at pixel level), which might negatively impact the optimization. Here, we explore how focal loss helps

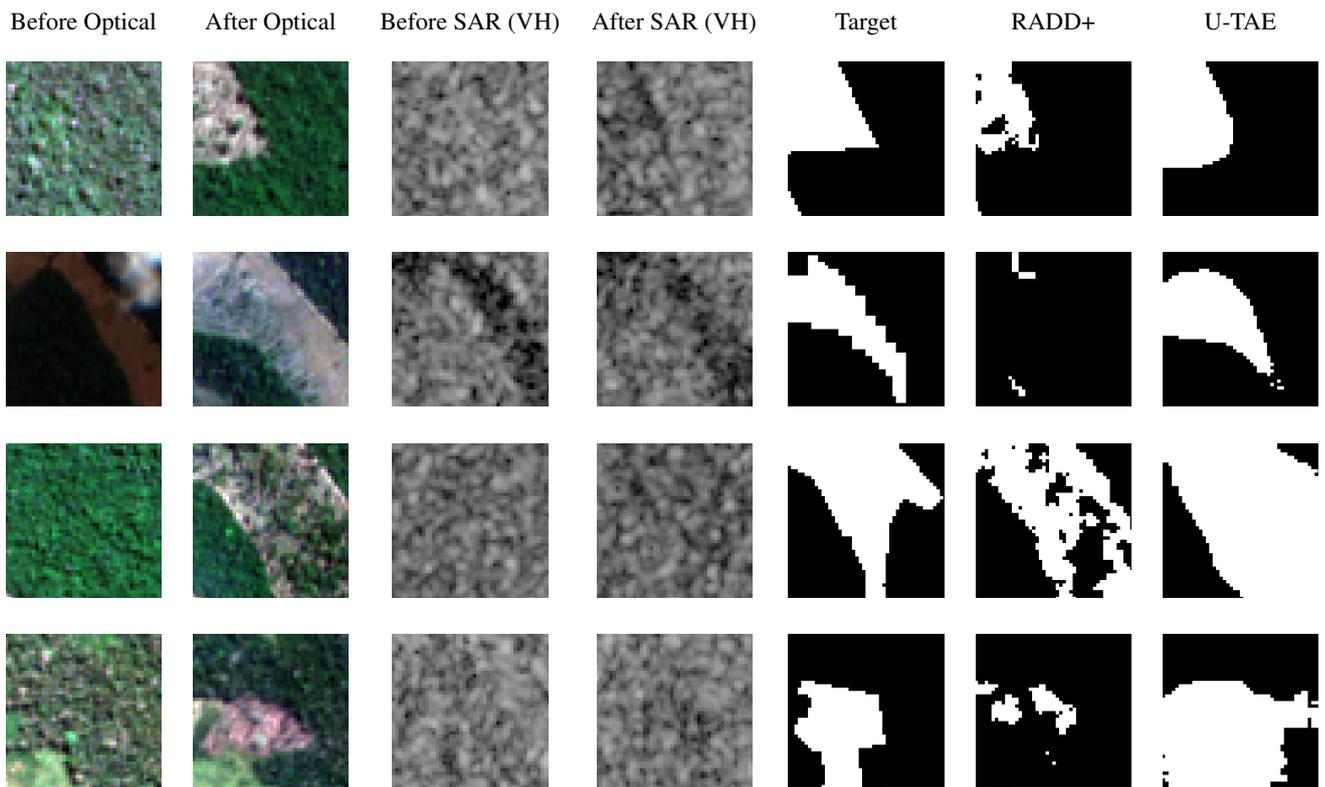| Before Optical | After Optical | Before SAR (VH) | After SAR (VH) | Target | RADD+ | U-TAE |

Figure 7. Sample image patches of BraDD-S1TS. The first and second columns taken from the optical data (Sentinel-2) are for visualization only, they are not included in the dataset.

	Alpha		
Gamma	0.25	0.5	0.75
1.0	35.9	47.2	**48.6**
2.0	35.0	47.0	48.2

Table 2. Test IoU scores of the models at pixel level with different hyper-parameters of focal loss.

alleviate this issue. We show in Table 2 the test performance of the U-TAE at pixel-level trained with different hyper-parameters α and γ. We observe that $\alpha = 0.75$ and $\gamma = 1$ lead to an increase of 1.3% IoU.

5.4 Qualitative Results

We complement our quantitative analysis with qualitative results shown in Figure 7. This figure shows the prediction masks obtained for four randomly chosen test patches. Our results demonstrate that pixel-level accuracy is adequate for deforestation monitoring in real-life systems.

6. CONCLUSION

In this study, we have introduced BraDD-S1TS, a novel, publicly available dataset of multi-temporal SAR observations for deforestation detection in the Amazon rainforest. We evaluated existing state-of-the-art methods and proposed to use recent advances in attention-based deep learning models on this task. Our experiments showed that such deep learning methods significantly improve overall performance. We have also shown that leveraging the full time series of available observations during one year almost doubles the performance compared to bi-temporal change detection. We hope our encouraging results,

as well as the open access dataset, will foster further research into this problem, as the achieved performance leaves room for improvement. Finally, we argue that designing methods that are able to predict the exact date of the deforestation event is a challenging and exciting venue for further research.

REFERENCES

Assis, L. F., 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(11).

Buchhorn, M., Smets, B., Bertels, L., Roo, B. D., Lesiv, M., Tsendbazar, N.-E., Herold, M., Fritz, S., 2020. Copernicus Global Land Service: Land Cover 100m: collection 3: epoch 2019: Globe.

Cherif, E., Hell, M., Brandmeier, M., 2022. DeepForest: Novel Deep Learning Models for Land Use and Land Cover Classification Using Multi-Temporal and -Modal Sentinel Data of the Amazon Basin. *Remote Sensing*, 14(19). https://www.mdpi.com/2072-4292/14/19/5000.

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.

Garnot, V. S. F., Landrieu, L., 2021. Panoptic segmentation of satellite image time series with convolutional temporal attention networks. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 4872–4881.

Garnot, V. S. F., Landrieu, L., Giordano, S., Chehata, N., 2020. Satellite image time series classification with pixel-set encoders and temporal self-attention. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12325–12334.

Gong, M., Zhao, J., Liu, J., Miao, Q., Jiao, L., 2016. Change Detection in Synthetic Aperture Radar Images Based on Deep Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems*, 27(1), 125-138.

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.

Hansen, M. C., Krylov, A., Tyukavina, A., Potapov, P. V., Turubanova, S., Zutta, B., Ifo, S., Margono, B., Stolle, F., Moore, R., 2016. Humid tropical forest disturbance alerts using Landsat data. *Environmental Research Letters*, 11(3), 034008.

Heidari, H., Warziniack, T., Brown, T. C., Arabi, M., 2021. Impacts of Climate Change on Hydroclimatic Conditions of U.S. National Forests and Grasslands. *Forests*, 12(2). https://www.mdpi.com/1999-4907/12/2/139.

Jiang, H., Peng, M., Zhong, Y., Xie, H., Hao, Z., Lin, J., Ma, X., Hu, X., 2022. A Survey on Deep Learning-Based Change Detection from High-Resolution Remote Sensing Images. *Remote Sensing*, 14(7), 1552.

Kendall, A., Gal, Y., 2017. What uncertainties do we need in bayesian deep learning for computer vision? *Advances in neural information processing systems*, 30.

Khelifi, L., Mignotte, M., 2020. Deep learning for change detection in remote sensing images: Comprehensive review and meta-analysis. *Ieee Access*, 8, 126385–126400.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2017. Focal loss for dense object detection. *Proceedings of the IEEE international conference on computer vision*, 2980–2988.

Loshchilov, I., Hutter, F., 2017. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.

Metzger, N., Turkoglu, M. O., D'Aronco, S., Wegner, J. D., Schindler, K., 2021. Crop classification under varying cloud cover with neural ordinary differential equations. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–12.

Moreira, A., Prats-Iraola, P., Younis, M., Krieger, G., Hajnsek, I., Papathanassiou, K. P., 2013. A tutorial on synthetic aperture radar. *IEEE Geoscience and remote sensing magazine*, 1(1), 6–43.

Nepstad, D. C., Stickler, C. M., Filho, B. S., Merry, F., 2008. Interactions among Amazon land use, forests and climate: prospects for a near-term forest tipping point. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1498), 1737-1746. https://royalsocietypublishing.org/doi/abs/10.1098/rstb.2007.0036.

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).

Pelletier, C., Webb, G. I., Petitjean, F., 2019. Temporal Convolutional Neural Network for the Classification of Satellite Image Time Series. *Remote Sensing*, 11(5). https://www.mdpi.com/2072-4292/11/5/523.

Reiche, J., De Bruin, S., Hoekman, D., Verbesselt, J., Herold, M., 2015. A Bayesian Approach to Combine Landsat and ALOS PALSAR Time Series for Near Real-Time Deforestation Detection. *Remote Sensing*, 7(5), 4973–4996. https://www.mdpi.com/2072-4292/7/5/4973.

Reiche, J., Mullissa, A., Slagter, B., Gou, Y., Tsendbazar, N.-E., Odongo-Braun, C., Vollrath, A., Weisse, M. J., Stolle, F., Pickens, A. et al., 2021. Forest disturbance alerts for the Congo Basin using Sentinel-1. *Environmental Research Letters*, 16(2), 024005.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, Springer, 234–241.

Rustowicz, R., Cheong, R., Wang, L., Ermon, S., Burke, M., Lobell, D., 2019. Semantic segmentation of crop type in africa: A novel dataset and analysis of deep learning methods. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

Rußwurm, M., Körner, M., 2017. Temporal vegetation modelling using long short-term memory networks for crop identification from medium-resolution multi-spectral satellite images. *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1496–1504.

Rußwurm, M., Körner, M., 2018. Multi-Temporal Land Cover Classification with Sequential Recurrent Encoders. *ISPRS International Journal of Geo-Information*, 7(4). https://www.mdpi.com/2220-9964/7/4/129.

Rußwurm, M., Körner, M., 2020. Self-attention for raw optical Satellite Time Series Classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 169, 421-435. https://www.sciencedirect.com/science/article/pii/S0924271620301647.

Shi, W., Zhang, M., Zhang, R., Chen, S., Zhan, Z., 2020. Change Detection Based on Artificial Intelligence: State-of-the-Art and Challenges. *Remote Sensing*, 12(10). https://www.mdpi.com/2072-4292/12/10/1688.

Shukla, J., Nobre, C., Sellers, P., 1990. Amazon deforestation and climate change. *Science*, 247(4948), 1322–1325.

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B., Floury, N., Brown, M. et al., 2012. GMES Sentinel-1 mission. *Remote sensing of environment*, 120, 9–24.

Turkoglu, M. O., D'Aronco, S., Wegner, J. D., Schindler, K., 2022. Gating Revisited: Deep Multi-Layer RNNs That can be Trained. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(8), 4081-4092.

You, Y., Cao, J., Zhou, W., 2020. A survey of change detection methods based on remote sensing images for multi-source and multi-objective scenarios. *Remote Sensing*, 12(15), 2460.