DEFORESTATION DETECTION IN THE AMAZON RAINFOREST WITH SPATIAL AND CHANNEL ATTENTION MECHANISMS

P. Tovar[1,*], M. O. Adarme[1], R. Q. Feitosa[1]

[1] Department of Electrical Engineering, Pontifical Catholic University of Rio de Janeiro,
22451-900 Rio de Janeiro, Brazil - (ptovar@opto.cetuc.puc-rio.br, mortega@ele.puc-rio.br, raul@ele.puc-rio.br)

KEY WORDS: Remote Sensing, Deforestation Detection, Attention Mechanisms, Convolutional Neural Networks.

ABSTRACT:

Deforestation in the Amazon rainforest is an alarming problem of global interest. Environmental impacts of this process are countless, but probably the most significant concerns regard the increase in CO_2 emissions and global temperature rise. Currently, the assessment of deforested areas in the Amazon region is a manual task, where people analyse multiple satellite images to quantify the deforestation. We propose a method for automatic deforestation detection based on Deep Learning Neural Networks with dual-attention mechanisms. We employed a siamese architecture to detect deforestation changes between optical images in 2018 and 2019. Experiments were performed to evaluate the relevance and sensitivity of hyperparameter tuning of the loss function and the effects of dual-attention mechanisms (spatial and channel) in predicting deforestation. Experimental results suggest that a proper tuning of the loss function might bring benefits in terms of generalisation. We also show that the spatial attention mechanism is more relevant for deforestation detection than the channel attention mechanism. When both mechanisms are combined, the greatest improvements are found, and we reported an increase of 1.06% in the mean average precision over a baseline.

1. INTRODUCTION

The environmental impacts of deforestation in the Amazon rainforest have been attracting research interest for many decades (Lean, Warrilow, 1989, Shukla et al., 1990). As this region is the largest tropical forest globally, it is paramount to assess its deforestation consequences. Countless works report predictions on climate changes due to deforestation in the Amazon rainforest. For instance, Swann and co-workers developed a regional ecosystem model to investigate the future consequences of climate changes caused by deforestation in South America (Swann et al., 2015). Medvigy simulated the effects of Amazon deforestation in the northern hemisphere climate using the Ocean-Land-Atmosphere Model, where significant temperature and precipitation changes were predicted (Medvigy et al., 2013). However, the effects of deforestation are not restricted to changes in temperature and precipitation. Deforestation for land use is dual-severe in terms of greenhouse gases because at the same time that it is one of the largest causes of anthropogenic CO_2 emissions; it also ends up reducing the natural capacity of terrestrial carbon storage (Sy et al., 2015). Furthermore, deforestation has an enormous impact on biodiversity loss, being even more harmful to species of high conservation and functional value (Barlow et al., 2016). Besides these more evident effects, there are also indirect and hard to predict consequences of deforestation. An example is a recent study that relates the increase of malaria transmission in Brazil due to Amazon deforestation (MacDonald, Mordecai, 2019).

According to National Institute for Space Research, the deforested areas in Amazon decreased almost steadily from 2004 to 2014 (INPE, 2020). However, in the last years, the deforestation processes have significantly increased, reaching in 2019 the largest deforested area in the past ten years. Therefore, this forest's supervision is essential to control man-made hazards and natural disasters and enforce public policies to avoid irregular activities and contribute to climate change mitigation. In this regard, Remote Sensing (RS) has demonstrated to be a cost-effective solution to monitor these regions. The advances of this technology and the easy access to high-resolution images throughout the globe propelled remote sensing Change Detection (CD) techniques.

Many approaches have been proposed to enhance CD performance. Mostly, we can group them into four main categories: image algebra; transformation; time-series analysis and classification (Hecheltjen et al., 2014). Image algebra methods are based on spectral values analysis, and they range from simple image differencing to change vector analysis (CVA), which calculates both the magnitude and direction of changes. Methods that convert the input data into another dimensional-space for change analysis are part of the transformation category. Examples of transformation methods are Fast Fourier Transform (FFT) and Principal Component (PC) based change detection (Wiemker, 1997). Time-series analysis encompasses the methods using more than two images of the same location taken at different times to detect the changing trends. The last category covers classification methods, which relies on a great quantity and quality of classified images to produce change detection outputs (Lu et al., 2004).

Classification methods for CD started to stand out due to the remarkable results of Deep Learning (DL) applied to image classification. Since then, DL classification methods have been widely used for various CD applications, including land use and land cover change, urban settlements, changes caused by natural catastrophes, and deforestation changes (Asokan, Anitha, 2019). Given the importance of deforestation detection in the Amazon rainforest, methods based on DL have been applied to this problem aiming to provide a robust and automatic way to monitor Amazon deforested areas (Ortega Adarme et al., 2020, de Bem et al., 2020).

In this work, a recently proposed deep-learning method for

---

* Corresponding author

CD (Chen et al., 2020) was applied to deforestation detection in the Amazon rainforest. The methodology's main idea is to use dual-attention mechanism (spatial and channel) to improve the robustness against pseudo-changes in remote sensing applications, i.e., efficiently distinguishing between relevant changes and circumstantial ones, such as noise and context. This concept, which will be described in details in the following sections, has great potential for deforestation detection, and in this case, the relevant change is current deforestation itself. Furthermore, Chen's approach (Chen et al., 2020) uses a new loss function, which intends not only to compensate for the imbalance problem often found in training samples but also to optimise the training process.

This paper is organised as follows. Section 1 describes the relevance of deforestation detection in the Amazon rainforest and how it has been approached with recent DL methods. Section 2 covers the related works in the literature. In section 3 it is detailed the methodology used in this work, where the network architecture, attention mechanisms and loss function are explained. Section 4 presents the experiments assessed and their results. Finally, Section 5 concludes the work.

2. RELATED WORKS

2.1 Deforestation Detection

The number of papers studying deforestation detection using DL techniques increased considerably in the past few years. Recently, a Convolutional Neural Network (CNN) was employed by Grings to monitor deforestation events based on an 18-years time-series of Enhanced Vegetation Index (EVI) in the Chaco Forest (Grings et al., 2020). The EVI is a vegetation index based on the spectral content of satellite images. It oscillates according to seasonal changes, but the oscillation exhibits a break-point in a deforestation event. Grings' model was trained to learn the break-points' patterns, estimating the probability of deforestation for a given EVI sequence. The model showed to be sensitive to mislabeled deforestation events, which limited its performance. Also, it requires long EVI time-series for training, which are often unavailable. And since the CNN is trained with samples of EIV time-series, the spatial relation of pixels is not considered in the learning process. A similar work reported by Adarme and co-workers (Ortega Adarme et al., 2020) evaluated the performance of three DL algorithms applied to the deforestation detection problem in two different Brazilian biomes: Amazon and Cerrado. The assessed strategies were a CNN, Siamese Network, Convolutional Support Vector Machine and Support Vector Machine. The obtained results were promising. However, all these methods rely on a patch-wise classification, which can be computationally intensive.

An analysis considering a Fully Convolutional Networks (FCN) was successfully applied and reported in (de Bem et al., 2020). The dataset consists of images from the Amazon territory acquired one year apart. Training samples were formed by stacking two images from the same location at different years. It was evaluated the performance of five different architectures on the task of deforestation classification: SharpMask, U-Net, ResNet, Random Forest and Multilayer Perceptron. The firsts three architectures produced better classification results than the last two, showing that DL approaches are better suited for RS applications than classic machine-learning algorithms.

Other works with similar approaches also reported remarkable results using DL methods for deforestation detection (Lee et al.,

2020, Maretto et al., 2020). However, despite the positive results, current methods are still sensitive to pseudo-changes. For instance, some areas might be misclassified as deforested because they have a different luminosity, shadow, noise, or any other factor that might mislead the neural network. Yet, this problem is not restricted to deforestation detection but rather common to most CD applications. Aiming to mitigate this problem for change detection between images taken at different seasons, Chen proposed the usage of extended self-attention mechanism (Chen et al., 2020).

2.2 Attention Mechanisms

When reading a sentence, humans naturally focus their attention on the most relevant words. Although each word has its importance grammatically, some words give much more information than others. The same happens with images. We pay more attention to regions in images providing more context information. For instance, by looking at an image of a cat playing with a ball at a backyard, we rapidly focus on the cat's primary features - pointy ears, paws, feline nose and mouth - and on the ball's spherical shape, not paying much attention to the background. This natural capability of focusing on more relevant features is key for image classification, as only the most significant aspects are learned, making us exceptional in generalization tasks. Recently, this attention concept has been brought to deep learning algorithms.

Attention mechanisms (AM) were first proposed in (Bahdanau et al., 2015) to solve the long-range dependence issue of sequence-to-sequence models. Since then, AM have been explored in several applications (Xu et al., 2019, Mahayossanunt et al., 2019, Li et al., 2020). The principal idea of AM is to allow to the network focus on more meaningful features by weighting them according to their importance. Hu et al. reported an interesting configuration of AM by defining Squeeze-and-Excitation (SE) blocks (Hu et al., 2018), which were designed to model the interdependencies between image channels explicitly.

Other researches extended the concept of Channel Attention Mechanism to the spatial dimension (Guo et al., 2018). For instance, in (Roy et al., 2019), the performance of SE attention blocks was assessed separately for channel and spatial dimensions and in a joint channel-spatial attention configuration. Tests with these three configurations were performed with medical imagery for semantic segmentation, and most results outperformed configurations without AM.

For change detection applications, a similar idea was proposed by Chen and co-workers. Dual-attention mechanisms, acting both on the channel and spatial dimensions, were included in a siamese network architecture to improve robustness against pseudo-changes, primarily seasonal. Results showed consistent improvement in all metrics evaluated (Chen et al., 2020). These results suggest that the performance of state-of-the-art deep-learning architectures used for deforestation detection could be improved with dual-attention mechanisms. In our work, we implemented the concepts of dual-attention mechanisms to deforestation detection in the Amazon rainforest.

3. METHODOLOGY

3.1 Network Architecture

The network architecture is shown in Figure 1. Its input is composed of two co-registered images of the same area acquired

at different times, shown as I_{T_0} and I_{T_1}. Both images undergo the same transformations throughout the network, i.e., Net_0 and Net_1 share all weights, denoting a siamese relation. Each input image is transferred to an embedding space through a sequence of convolutions, non-linear activations (ReLU) and max-pooling operations ($convs$ in Figure 1), leading to features f_{0_1} and f_{1_1}. Then, each feature is sent through an *attention block*, where two attention mechanisms take place. Details of the attention block operations will be described in section 3.2. Still, it can be understood as a transformation block that generates a feature embedding of the same size of its input but with channel and spatial features weighted by their own importance.

The outputs of the attention blocks, f_{0_2} and f_{1_2}, are used in a distance function $d(\cdot, \cdot)$ to compute the pixel-wise distance between feature embeddings. Last, a bi-linear upsample operation is applied to the distance function output, generating a change map of the same size of the input images. The upsampled change map and deforestation label are used to calculate the loss, which is a contrastive function with two margins, as will be explained in section 3.3.
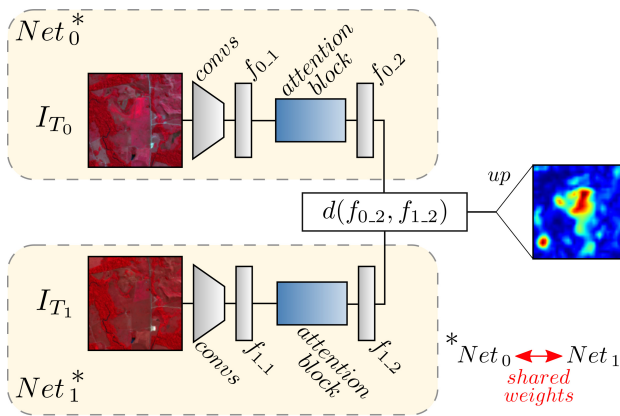

Figure 1. Network architecture.

3.2 Attention Mechanisms

The type of attention mechanism implemented in this work is defined as self-attention. Its goal is to generate a better representation of a given feature embedding by focusing on more relevant elements of the embedding itself, i.e., without additional information. Since feature embeddings of images are usually represented as tensors with at least three dimensions, it is possible to force an attention mechanism to focus on a specific dimension. In this section, we detail the two attention mechanisms that rely within the *attention block* shown in Figure 1, one focusing on the spatial dimension and the other on the channel dimension. Henceforth, these will be referred to as Spatial Attention Mechanism (SAM) and Channel Attention Mechanism (CAM).

Both SAM and CAM uses the concept of query, key and value to implement self-attention, which has been used in a similar way in (Ramachandran et al., 2019). This concept is borrowed from retrieval systems, where it is realised a similarity measurement between a query and keys to return a match with the best value. We start by analysing the operations for the SAM. Let us represent query, key and value for SAM as q_{SAM}, k_{SAM} and

v_{SAM}, respectively. These are defined as follows,

$$
\begin{aligned}
q_{SAM} &= reshape(CONV_1(f_{X\_1}))^T \\
k_{SAM} &= reshape(CONV_1(f_{X\_1})) \\
v_{SAM} &= reshape(CONV_2(f_{X\_1})),
\end{aligned}
\tag{1}
$$

where $f_{X\_1} \in \mathbb{R}^{C_{IN} \times H \times W}$ is the input of the attention block, such that X can assume values $\{0, 1\}$, depending on the network analysed (Net_0 or Net_1). The operations $CONV_1$ and $CONV_2$ are 1×1 convolutions with output channels N and C_{IN}, respectively. Applying $CONV_1$ to f_{X_1} gives a tensor in the space $\mathbb{R}^{N \times H \times W}$. k_{SAM} is obtained by reshaping the output of this convolution to the tensorial space $\mathbb{R}^{N \times H * W}$. The query, q_{SAM}, is calculated by transposing a reshaped convolved tensor, giving an element in $\mathbb{R}^{H * W \times N}$. The value, v_{SAM}, is obtained by reshaping the output of a 1×1 convolution with C_{IN} channels, resulting in a tensor in space $\mathbb{R}^{C_{IN} \times H * W}$.

The central element of the SAM is the calculus of the spatial attention tensor A_{SAM}, which is given by

$$
A_{SAM} = \sigma(q_{SAM} \cdot k_{SAM}),
\tag{2}
$$

where σ is the softmax operation. Note that A denotes a multiplication between matrices of dimensions $(H * W, N)$ and $(N, H * W)$, such that the output resides in space $\mathbb{R}^{H * W \times H * W}$. We can think of this operation as a contraction in the channel dimension, resulting in a tensor that provides spatial-context relations. The softmax operation guarantees that the values of A lies in the interval $[0, 1]$. Next, we apply the attention tensor to the value v_{SAM} and reshape the output to have the same dimension of the input f_{X_1}, i.e.

$$
s_{SAM} = reshape(v_{SAM} \cdot A_{SAM}),
\tag{3}
$$

with $s_{SAM} \in \mathbb{R}^{C_{IN} \times H \times W}$. Last, we write the output of the SAM as a sum between the input f_{X_1} and s_{SAM} weighted by a learnable parameter η, such that the neural network will learn the relevance of the SAM. Thus, the output of SAM can be written as:

$$
out_{SAM} = f_{X\_1} + \eta\, s_{SAM}
\tag{4}
$$

For the CAM we have similar key, query and value tensors, but no convolutions are used (Chen et al., 2020). These are defined as follows:

$$
\begin{aligned}
q_{CAM} &= reshape(f_{X\_1}) \\
k_{CAM} &= reshape(f_{X\_1})^T \\
v_{CAM} &= reshape(f_{X\_1}).
\end{aligned}
\tag{5}
$$

The paramount difference between Equations 1 and 5 is the position of the transpose operation. It the former, it is applied to the query, while in the latter it is applied to the key. Thus, q_{CAM} resides in the space $\mathbb{R}^{C_{IN} \times H * W}$ and k_{CAM} in $\mathbb{R}^{H * W \times C_{IN}}$. With this modification, when multiplying these elements as done in Equation 2 for the SAM case, we obtain a channel attention tensor with dimension (C_{IN}, C_{IN}), given by

$$
A_{CAM} = \sigma(q_{CAM} \cdot k_{CAM}).
\tag{6}
$$

In a similar way as done before for the SAM case, we calculate

s_{CAM} and out_{CAM} as follows:

$$s_{CAM} = reshape(A_{CAM} \cdot v_{CAM})$$
$$out_{CAM} = f_{X\_1} + \gamma\, s_{CAM}, \tag{7}$$

where γ is a learnable parameter. Unlike Equation 2, in the channel attention tensor A_{CAM} the contraction is in the spatial dimension, providing channel-context relations.

Last, the output of the *attention block* is a summation between SAM and CAM outputs, resulting in a feature embedding with enhanced spatial and channel information, which is written as

$$f_{X\_2} = out_{SAM} + out_{CAM}. \tag{8}$$

3.3 Loss Function

Similar to (Chen et al., 2020), we use the *Weighted Double Margin Contrastive (WDMC) Loss* as a loss function. It is an extension of the Contrastive Loss function introduced in (Hadsell et al., 2006), which can be seen as a special case of the WDMC Loss. These functions are designed to assign a high loss to dissimilar pairs, and low loss to similar ones. Let $\{x_i^{(0)}, x_i^{(1)}\}_{i \in [1,M]}$ be a set of paired elements in a batch with size M. Then, the WDMC loss is defined as

$$\mathcal{L} = \sum_{i=1}^{M} w_1\, y_i\, max(d(x_i^{(0)}, x_i^{(1)}) - m_1, 0)^2$$
$$+ w_2(1 - y_i)max(m_2 - d(x_i^{(0)}, x_i^{(1)}), 0)^2, \tag{9}$$

where y_i is the label associated with the pair $(x_i^{(0)}, x_i^{(1)})$, assuming value of 1 for similar pairs or corresponding to the same class, and 0 for dissimilar pairs or different class. The parameters w_1 and w_2 are weights, which are selected to mitigate the class imbalance problem. The parameters m_1 and m_2 are margins, which are designed to repel dissimilar pairs for at least a margin m_2 and approximate similar pairs by at most a margin m_1.

As mentioned before, the Contrastive Loss is a special case of the *WDMC Loss*, where w_1 and w_2 assume unit value, and m_1 is set to zero. Note that, in this case, even very similar pairs (small L^2 distance) would still imply a positive loss, such that the algorithm is slightly more prone to overfit to the training set. The inclusion of a margin m_1 greater than zero alleviates the similarity requirement, which might help in the process's generalisation.

4. EXPERIMENTS

A series of experiments were investigated to evaluate the performance of both *WDMC loss* and AM applied to deforestation detection in a region of the Amazon rainforest. We start this section by describing the study area to evaluate the algorithm in section 4.1. Next, in section 4.3, we examine the importance of selecting proper margins in the WDMC Loss function and how they affect the predictions of deforestation. The relevance of weights in the loss function is evaluated in section 4.4. Last, we analyse the effects of AM in section 4.5.

4.1 Study Area

The study area corresponds to a region of the Amazon rainforest located in the Pará State, Brazil. This region is centered on co-
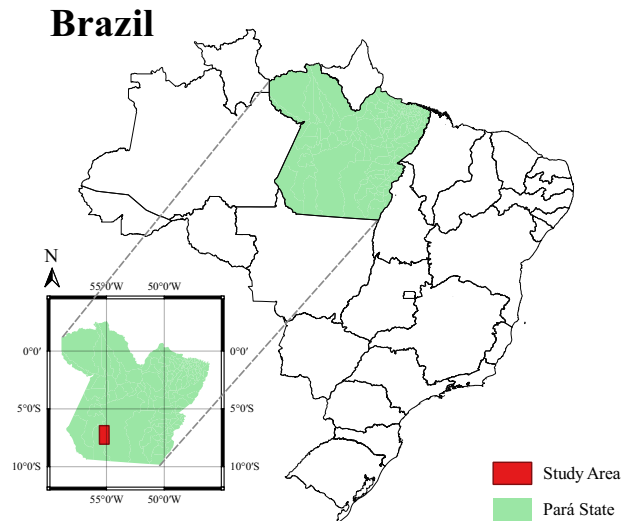

Figure 2. Study area corresponding to a region of the Pará State, Brazil.

ordinates of 06° 54' 16" South and 055° 11' 52" West (see Figure 2). This state reported one of the highest deforestation rates in 2019, which represented more than 40% of the total forest loss in Brazilian Legal Amazon (Assis et al., 2019). The database comprises two optical images acquired from the Landsat8-OLI sensor, with a resolution of 30 m. These are co-registered images, with dimension of 5905×3064 pixels, and they were downloaded from the United States Geological Survey [1]. Each image contains seven spectral bands: Coastal/Aerosol, Blue, Green, Red, NIR, SWIR-1 and SWIR-2. The reference deforestation map was downloaded from the INPE web site, which is publicly available (INPE, 2020).

This reference refers to the deforestation that occurred between July 2018 and July 2019. For this study, we defined three classes: 1) non-deforested areas, 2) deforested areas from July-2018 to July-2019, and 3) past deforestation and borders of current deforestation. Pixels annotated with this last class shall not be considered in the loss function, as they represent an overall unknown class for deforestation. These classes are shown in Figure 3 a), where blue regions denote non-deforested areas, white regions correspond to deforested areas, and red regions represent past deforestation or current deforestation borders.

4.2 Experimental Setup

To build the training, validation and test sets, we split the database images into 18 tiles, as shown in Figure 3 a). Given the lack of available databases with annotation for deforestation detection, we partition these tiles as follows. Seven tiles were randomly selected for training, two tiles for validation, and the remaining nine tiles for testing, resulting in a proportion 7 : 2 : 9. We selected patches of size (128×128) for each training tile, provided that each patch contains at least 2% of deforestation (class 2). This condition was required to build a training set with a significant portion of deforested areas. Four examples of training patches are shown in Figure 3 b), where the first column contains the reference, and the second/third columns show patches from 2018/2019 images, respectively. Note that

---

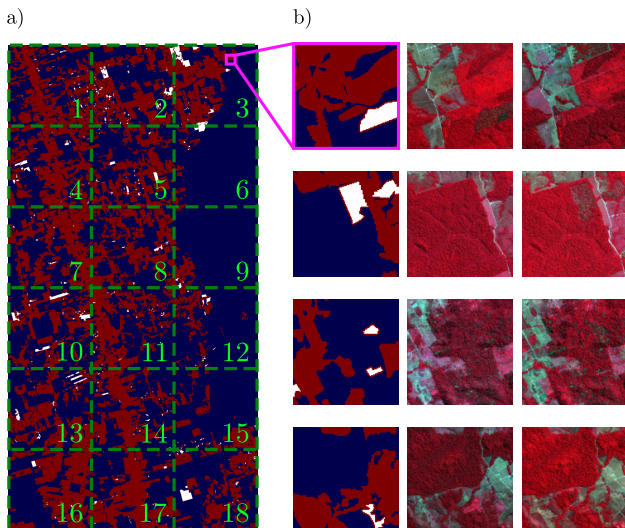[1] https://earthexplorer.usgs.gov/

Figure 3. Overview of dataset used in the experiments. In a), Reference of Amazon deforestation between 2018 and 2019. Blue color represents non-deforested areas, white color represents deforested areas, and red color indicates past deforested areas and current deforestation borders. In b), four examples of patches used in training phase; from left to right: label, image in 2018 and image in 2019, both represented in the *NIR-G-B* composition.

although image patches are presented in Figure 3 b) *NIR-G-B* composition, all seven bands were used in the training, validation and test phases.

For the selected training patches, a data augmentation procedure was applied, including random rotations and horizontal/vertical flip, resulting in a total of 1704 patches for training.

4.3 Single Margin vs Double Margin

To assess the relevance of margins selection in the *WDMC loss function*, the following experiment was set. We trained a network without attention mechanisms for four values of m_2, 2.0, 2.5, 3.0 and 3.5, and for each of these values, we evaluated the mean Average Precision (mAP) on the test set for six different m_1's. One of these six values was zero, denoting a single margin case. The weights w_1 and w_2 were maintained fixed at 0.18 and 0.82, respectively, for all scenarios. These values represent the proportion of deforested (18%) and not-deforested (82%) areas in the training patches. For the distance function $d(\cdot, \cdot)$, we chose to use the L^2 norm, as it provided the best results in (Chen et al., 2020). Figure 4 displays the experimental results.

In the four different cases evaluated in Figure 4, values of m_1 closer to 0 resulted in an mAP very similar to the Single Margin case, i.e., when m_1 is set to zero. This result indicates that small variations of margins have little impact on the prediction result. However, it is interesting to note that there was a value of $m_1 > 0$ that outperformed the Single Margin condition in all four cases. After reaching this best-case scenario, any increase in m_1 tends to decrease the mAP. Although the mAP improvement reached a maximum value less than 1% above the Single Margin case for $m_2 = 2.5$ and $m_1 = 0.3$, this behaviour indicates that when using a single margin, the network might be overfitting when trying to approximate similar pairs
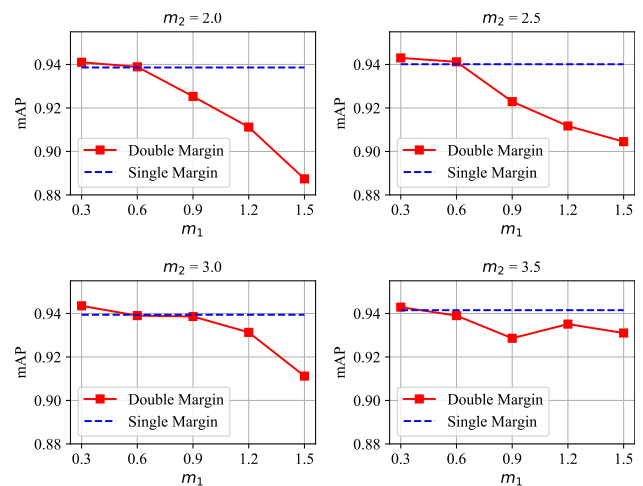

Figure 4. Analysis of mean Average Precision for different margins used in the *WDMC loss function*. Dashed lines indicate a reference for $m_1 = 0$.

to the point where the distance is equal to zero. Also, it is evident that whenever m_1 starts getting closer to m_2, it becomes difficult to distinguish between changed and unchanged pairs. For instance, when $m_2 = 2.0$ and $m_1 = 1.5$, the gap between margins narrows down to 0.5, and we obtained the worst mAP, 88.74%.

Given the results obtained with this experiment, in the following sections we kept margins m_1 and m_2 fixed at 0.3 and 3.0, respectively, as they provided the best mAP of 94.43%.

4.4 Dataset Imbalance Compensation

An assessment of the relevance of weights in the WDMC Loss function is done in a similar fashion. Ten cases were evaluated in this experiment, where values of w_1 ranged from 0.05 to 0.50 in steps of 0.05, while w_2 is simply its compliment, i.e., $w_2 = 1 - w_1$. The used distance function was again the L^2 norm, no attention mechanisms were applied, and margins remain fixed for all cases. Figure 5 displays experimental results, where the mAP obtained in the previous section for $w_1 = 0.18$ and $w_2 = 0.82$ is shown in blue color.

For relations w_2/w_1 closer to unity, the predictions showed a worse mAP. In contrast, relations closer to the optimal theoretical value (Chen et al., 2020), which weights changed and
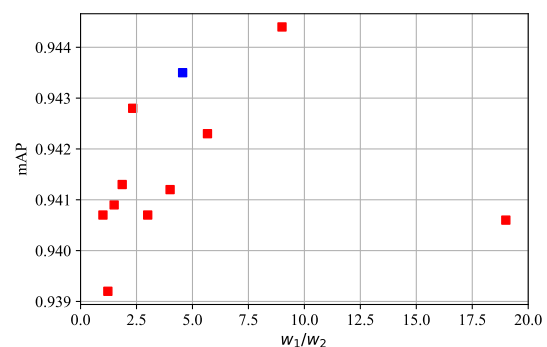

Figure 5. Analysis of mean Average Precision for different weights relations used in the *WDMC loss function*. Blue marker indicates the weights relation calculated based on the proportion of changed and unchanged pixels.

unchanged pixels according to their proportion in the dataset, provided the best results. We can also observe that by increasing the relevance of deforested areas even further, using a relation much greater than the theoretical optimal, the mAP decreases. Last, it is relevant to highlight the difference in mAP between a poorly set pair of weights and the best-case scenario, which was found to be 0.52% according to Figure 5.

4.5 Effects of Attention Mechanisms

An analysis of the effects of AM on deforestation detection is provided in this section. In contrast to the previous sections, where the experiments considered modifications only in the loss function, here, we implemented modifications in the network architecture to evaluate the contribution of each AM separately. Four methods were considered: (I) a baseline (BL) architecture without AM, with the distance function applied directly to the feature embeddings f_{0_1} and f_{1_1}; (II) an architecture with only a CAM present in the *attention block*; (III) another with only a SAM present; and (IV) with the dual-attention mechanisms combined. Weights w_1 and w_2 of WDMC loss function were kept fixed and equal to 0.1 and 0.9, respectively, as these provided the best mAP in the previous section, equal to 94.44%.

Visual results for deforestation predictions of networks trained with these four strategies are shown in Figure 6. There, five examples are provided, each presented in a single row. The first three columns show an input test patch from the 2018 image, its co-registered pair from 2019, and the ground truth (GT). The last four columns present the change map results from the four methods evaluated (I)-(IV), respectively. Blue colour represents the zero probability of deforestation class and red the maximum probability. For instance, the first row shows that architectures using the dual-attention modules distinguish the deforested regions accurately, providing more confident outputs and resolving complex geometries in a better way. On the other hand, the BL architecture presented more false-positive regions, classifying the class past deforestation as deforestation. The CAM and SAM outputs deliver fewer false-positive regions, but they produced less confident values than the dual-attention mechanisms.

Any further analysis requires a comparison of metrics to evaluate the performance of each method. For this reason, we provide a plot related to precision and recall scores for the four evaluated methods (see Figure 7). Furthermore, Table 1 summarises the numerical results in terms of Recall, Precision, F1-score (F1), and mAP. By carefully analysing the result in Figure 7, we can see an increase in the prediction performance when SAM is employed. The SAM method and dual-attention method curves lie above the CAM and baseline curves for all recall values.

Method	Rec.	Prec.	F1	mAP
Baseline	93.00	81.44	86.83	94.44
CAM	94.91	80.16	86.91	94.49
SAM	94.28	80.80	87.02	95.15
SAM + CAM	93.99	81.62	87.36	95.50

Table 1. Metrics results, in [%], for the four methods evaluated.

Since there are a few crosses between the curves of CAM and baseline methods and between SAM and dual-attention methods, we must analyse metrics results numerically to further conclusions. Table 1 shows that methods using spatial attention mechanisms indeed outperformed CAM and baseline methods.

We believe one of the following factors is the dominant reason for this result: the usage of convolution operations in the SAM method might better resolve complex features, or spatial relations are indeed more relevant than channels' for deforestation detection. Further analysis needs to be done before asserting which is the prevailing factor.

Also, according to Table 1, the CAM method did not provide many improvements compared to the baseline method. However, when combining CAM with SAM, nearly all metrics showed an improvement compared to the SAM method isolated. The best mAP obtained was for the dual-attention method, with an increase of 1.06% compared to the baseline method.

5. CONCLUSIONS

This paper reported the application of a deep-learning network, equipped with a dual-attention mechanism, to the task of deforestation detection in the Amazon rainforest. A recently proposed loss function, the *WDMC loss*, was used throughout the work. A set of experiments were implemented to analyse the effects of both margins and weights of the *WDMC loss* function in the prediction of deforestation. Results suggest that adding a second margin to the classic contrastive loss function might bring benefits in terms of generalisation, as it can help avoid overfitting. Also, it was shown that the weights of the loss function could impact the mean average precision of deforestation detection in up to 0.52%.

Further experiments were developed to investigate the effects of self-attention mechanisms in deforestation detection. Results showed that the spatial content of images was more relevant for attention mechanism than the channel content. The usage of a spatial attention mechanism resulted in an improvement in mean average precision of 0.71%, and when combined with channel attention mechanism, the improvement increased to 1.06%. To our best knowledge, this is the first work that reports an analysis of dual-attention mechanisms with WDMC loss function applied to deforestation detection. Thus, given the consistent improvements in predicting deforestation, we believe this work might encourage other researchers in the challenging task of deforestation detection.

6. ACKNOWLEDGEMENTS

REFERENCES

Asokan, A., Anitha, J., 2019. Change detection techniques for remote sensing applications: a survey. *Earth Science Informatics*, 12, 143–160.

Assis, F., Fernando, L., 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), 513.
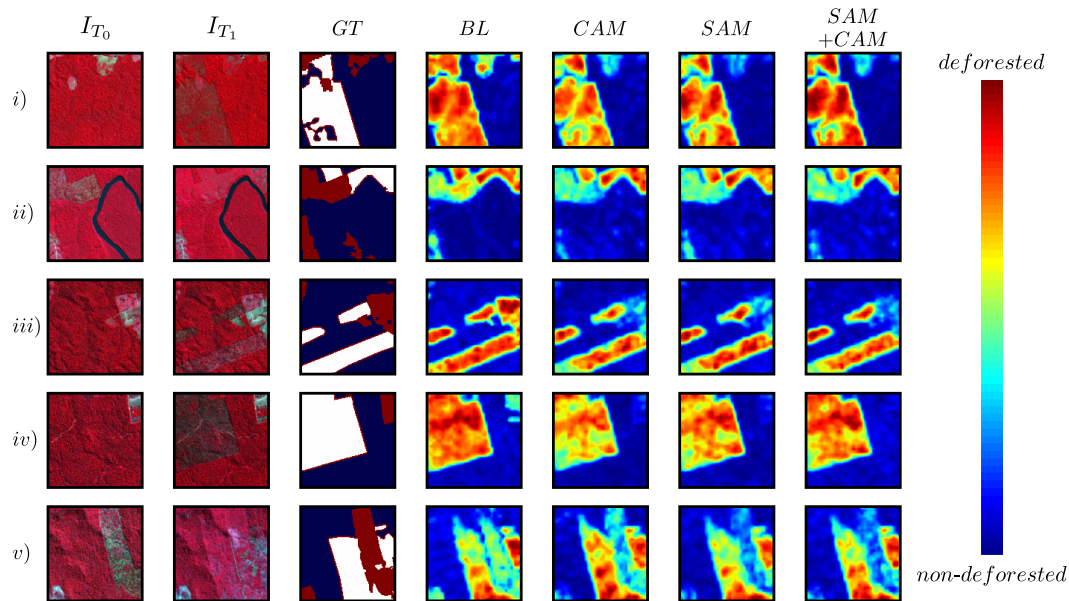
Figure 6. Visual results for the four methods evaluated. Each row presents an example of test patch assessed, and columns, from left to right, are: input image patch taken in 2018; input image patch taken in 2019; ground truth; and the change maps of baseline method; CAM; SAM; and dual-attention method.
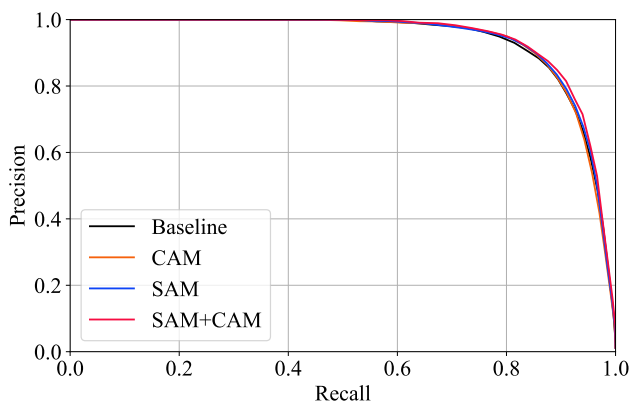

Figure 7. Relation between precision and recall for the four methods evaluated.

Bahdanau, D., Cho, K., Bengio, Y., 2015. Neural Machine Translation by Jointly Learning to Align and Translate. *CoRR*, abs/1409.0473.

Barlow, J., Lennox, G. D., Gardner, T. A., 2016. Anthropogenic disturbance in tropical forests can double biodiversity loss from deforestation. *Nature*, 535, 144-147.

Chen, J., Yuan, Z., Peng, J., Chen, L., Haozhe, H., Zhu, J., Liu, Y., Li, H., 2020. DASNet: Dual attentive fully convolutional siamese networks for change detection of high resolution satellite images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 1-1.

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).

Grings, F., Roitberg, E., Barraza, V., 2020. EVI Time-Series Breakpoint Detection Using Convolutional Networks for On-line Deforestation Monitoring in Chaco Forest. *IEEE Transactions on Geoscience and Remote Sensing*, 58(2), 1303-1312.

Guo, T., Wang, D., Jiang, Z., Men, A., Zhou, Y., 2018. Deep network with spatial and channel attention for person re-identification. *2018 IEEE Visual Communications and Image Processing (VCIP)*, 1–4.

Hadsell, R., Chopra, S., LeCun, Y., 2006. Dimensionality reduction by learning an invariant mapping. *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, 2, 1735–1742.

Hecheltjen, A., Thonfeld, F., Menz, G., 2014. *Recent Advances in Remote Sensing Change Detection – A Review*. Springer Netherlands, 145–178.

Hu, J., Shen, L., Sun, G., 2018. Squeeze-and-excitation networks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7132–7141.

INPE, 2020. Monitoring of the brazilian amazonian forest by satellite. Available online: http://www.obt.inpe.br/OBT/assuntos/programas/amazonia/prodes (accessed on 29 November 2020).

Lean, J., Warrilow, D. A., 1989. Simulation of the regional climatic impact of Amazon deforestation. *Nature*, 342, 411–413.

Lee, S.-H., Han, K.-J., Lee, K., Lee, K.-J., Oh, K.-Y., Lee, M.-J., 2020. Classification of Landscape Affected by Deforestation Using High-Resolution Remote Sensing Data and Deep-Learning Techniques. *Remote Sensing*, 12(20).

Li, F., Bai, H., Zhao, Y., 2020. Learning a Deep Dual Attention Network for Video Super-Resolution. *IEEE Transactions on Image Processing*, 29, 4474-4488.

Lu, D., Mausel, P., Brondízio, E., Moran, E., 2004. Change detection techniques. *International Journal of Remote Sensing*, 25(12), 2365-2401.

MacDonald, A. J., Mordecai, E. A., 2019. Amazon deforestation drives malaria transmission, and malaria burden reduces forest clearing. *Proceedings of the National Academy of Sciences*, 116(44), 22212–22218.

Mahayossanunt, Y., Thannamitsomboon, T., Keatmanee, C., 2019. Convolutional neural network and attention mechanism for bone age prediction. *2019 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)*, 249–252.

Maretto, R. V., Fonseca, L. M. G., Jacobs, N., Körting, T. S., Bendini, H. N., Parente, L. L., 2020. Spatio-Temporal Deep Learning Approach to Map Deforestation in Amazon Rainforest. *IEEE Geoscience and Remote Sensing Letters*, 1-5.

Medvigy, D., Walko, R. L., Otte, M. J., Avissar, R., 2013. Simulated Changes in Northwest U.S. Climate in Response to Amazon Deforestation*. *Journal of Climate*, 26(22), 9115-9136.

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).

Ramachandran, P., Parmar, N., Vaswani, A., Bello, I., Levskaya, A., Shlens, J., 2019. Stand-alone self-attention in vision models. H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, R. Garnett (eds), *Advances in Neural Information Processing Systems*, 32, Curran Associates, Inc., 68–80.

Roy, A. G., Navab, N., Wachinger, C., 2019. Recalibrating Fully Convolutional Networks With Spatial and Channel "Squeeze and Excitation" Blocks. *IEEE Transactions on Medical Imaging*, 38(2), 540-549.

Shukla, J., Nobre, C., Sellers, P., 1990. Amazon Deforestation and Climate Change. *Science*, 247(4948), 1322–1325.

Swann, A. L., Longo, M., Knox, R. G., Lee, E., Moorcroft, P. R., 2015. Future deforestation in the Amazon and consequences for South American climate. *Agricultural and Forest Meteorology*, 214-215, 12 - 24.

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

Wiemker, R., 1997. An iterative spectral-spatial bayesian labeling approach for unsupervised robust change detection on remotely sensed multispectral imagery. G. Sommer, K. Daniilidis, J. Pauli (eds), *Computer Analysis of Images and Patterns*, Springer Berlin Heidelberg, Berlin, Heidelberg, 263–270.

Xu, B., Liu, J., Hou, X., Liu, B., Garibaldi, J., Ellis, I. O., Green, A., Shen, L., Qiu, G., 2019. Look, investigate, and classify: A deep hybrid attention method for breast cancer classification. *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 914–918.