

# MULTI-ATTENTION GHOSTNET FOR DEFORESTATION DETECTION IN THE AMAZON RAINFOREST

M. X. Ortega Adarme<sup>1,\*</sup>, G. A. O. P. da Costa<sup>2</sup>, R. Q. Feitosa<sup>1</sup>

<sup>1</sup> Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Brazil  
(mortega, raul)@ele.puc-rio.br

<sup>2</sup> State University of Rio de Janeiro (UERJ), Rio de Janeiro, Brazil  
gilson.costa@ime.uerj.br

Commission III, WG III/7

**KEY WORDS:** Deforestation Detection, Amazon Rainforest, GhostNet, Spatial Attention Mechanisms, Fully Convolutional Networks.

## ABSTRACT:

Efficient deforestation detection techniques are essential to monitor and control illegal logging, thus reducing forest loss and carbon emissions in the Amazon rainforest. Recent works based on Deep Learning (DL) models have been proposed for that purpose. DL-based methods, however, are known to require large amounts of training data to be properly trained. Moreover, the deforestation detection application is characterized by a high class imbalance, as recent deforestation areas usually represent a small fraction of the geographic extents being monitored. In order to produce a lightweight architecture in terms of the number of learnable parameters and address the high class imbalance of the deforestation detection application, we propose a DL model based on the GhostNet architecture, which combines Ghost modules in a fully convolutional architecture. The proposed architecture also includes Spatial Attention Mechanisms attached to the skip connections of the GhostNet in order to better capture the spatial relationships among class features. Experiments were carried out using Sentinel-2 images of a region in the Pará state, Brazil, in the Amazon rainforest. The results obtained show that the proposed model achieves accuracy levels that are superior to those delivered by state-of-the-art DL architectures, with a lower computational cost due to the smaller number of learnable parameters.

## 1. INTRODUCTION

Comprising 5.5 million km<sup>2</sup>, 60% of which located in Brazilian territory, the Amazon is the largest rainforest, and one of the most biodiverse ecosystems on Earth (Lorenz et al., 2021). Unfortunately, in the past decades the biome has experienced sustained threats caused by human intervention. The large scale use of land for agricultural activities have induced biodiversity loss and degradation, carbon emissions, water pollution and deforestation (Morris, 2010). In particular, deforestation in the Brazilian Legal Amazon (BLA) has increased considerably and alarmingly annually (Lorenz et al., 2021). According to (Pereira et al., 2020), if deforestation reaches around 40% of the total forest area it can produce an increase in global temperatures of up to 4° Celsius, and that catastrophic scenario can happen in this century if the current rates are kept. Therefore, the implementation of wide-reaching monitoring systems is of paramount importance to understand the nature and effects of the underlying processes, and to efficiently combat illegal logging.

The Brazilian National Institute for Space Research (INPE) monitors the annual deforestation rates in the BLA since 1988 through the Program for Deforestation Monitoring in the Brazilian Legal Amazon (PRODES) (INPE, 2021). As classification accuracy is extremely important in that application since the reported figures are official ones, PRODES's methodology still involves significant human intervention. In this regard, automatic solutions are necessary in order to reduce human effort as well as the time taken to perform the underlying tasks.

\* Corresponding author

In recent years, DL-based methods have become the state-of-the-art in many computer vision fields, including image classification, object detection, and semantic segmentation (Alam et al., 2021). With Deep Neural Networks (DNNs) it is possible to learn robust representations that can improve prediction accuracies (Srivastava and Biswas, 2020). However, to achieve suitable performances, conventional DNNs demands large volumes of training samples, and are characterized by large numbers of parameters and high computational costs (Paoletti et al., 2021).

Indeed, one of the recent trends in DL is the design of high performance DNNs with portable and efficient architectures (Han et al., 2020). For instance, MobileNet (Howard et al., 2017), ShuffleNet (Zhang et al., 2018) and GhostNet (Han et al., 2020) employ depthwise and pointwise convolutions to replace traditional convolution layers, and drastically reduce the number of weights to be learned during training. Those structures allow creating efficient architectures, with fewer learnable parameters and state-of-the-art performance.

In this work we propose a new Fully Convolutional Network (FCN) architecture, inspired by the GhostNet and containing attention modules. The architecture was evaluated in a deforestation detection task, in a particular region of the Amazon forest. The model was designed to comprise a reduced number of parameters and calculations, as well as to deal with high class imbalanced scenarios, as it is the case of the target application, since recent deforestation areas usually represent a small fraction of the geographic extents under study. We further compared the outcome of the proposed method with

those delivered by alternative architectures, and analyzed the uncertainty in the predictions generated by all the methods.

The remainder of this paper is structured as follows: Section 2 reviews related works about deforestation detection and attention mechanisms. Section 3 introduces the proposed method and the uncertainty metrics used in this work. Section 4 describes the experimental protocol and discusses the obtained results. Finally, the main conclusions drawn from the experimental analysis are presented in Section 5.

## 2. RELATED WORKS

### 2.1 DL-Based Deforestation Detection

Some recently published works have demonstrated promising results for deforestation mapping using FCN architectures. Specifically considering deforestation detection in Amazon sites, Bem and co-authors (de Bem et al., 2020) evaluated different FCN architectures including SharpMask, U-Net and ResNet. They also compared those techniques with the traditional machine learning (ML) algorithms, i.e., Random Forest and Multilayer Perceptron, showing that the DL models significantly outperformed the ML algorithms.

Similarly, (Ortega et al., 2021) compared three established FCN architectures, namely U-Net, ResU-net, and a Siamese Network, employed for deforestation detection in the BLA using data from different sensors: Landsat-8, Sentinel-2, and Sentinel-1. In most cases, the ResU-Net architecture delivered the best results and identified the deforested areas with higher precision.

Additionally, a number of similar works employed the U-Net model for deforestation mapping, and reported interesting results, e.g., (Wagner et al., 2019, Bragagnolo et al., 2021b, Bragagnolo et al., 2021a). All the previously mentioned works, however, are based on deep CNN architectures characterized by a large number of trainable parameters, and thus dependant of large sets of labeled training samples in order to avoid over-fitting and produce good performances, especially considering the class imbalance in deforestation detection application.

### 2.2 Deep Attention Mechanisms

In the context of DL, attention mechanisms (AM) attempt to mimic the way the human brain processes information (Ghaffarian et al., 2021). One of the key characteristics of human perception is that we tend not to process all available information at once. Indeed, humans have a tendency to selectively focus on one piece of information when and where it is needed, while ignoring other perceptual information (Niu et al., 2021).

AM was proposed by Bahdanau and co-authors in the context of neural machine translation (Bahdanau et al., 2014). Subsequently, the underlying structures and adaptations were employed in other applications, including computer vision, image processing and remote sensing (Koščević et al., 2019, Zeng et al., 2020, Zhang et al., 2020, Niu et al., 2021). Recent works have demonstrated the potential of AM to improve the DL approaches, e.g., (Gao et al., 2020, Li et al., 2020, Qing and Liu, 2021, Xue et al., 2021). The fundamental idea of an AM is to assign different weights to different pieces of information,

thus allowing DL models to focus on and identify relevant features for particular tasks (Hu et al., 2018).

Recent literature show encouraging results of AM employed in remote sensing, and in particular in change detection applications (Chen et al., 2020, Jiang et al., 2020, Lu et al., 2021, Guo et al., 2021). In that context, AM is used to enhance feature representation of image information, improving discrimination between changed and unchanged regions. Those studies seem to indicate that the inclusion of AM can improve the performance of state-of-the-art DL approaches applied to deforestation detection. Indeed, in (Tovar et al., 2021), a siamese network with Spatial Attention Mechanism (SAM) and Channel Attention Mechanism (CAM) was evaluated for detecting deforestation in a region of the Amazon rainforest. The results showed that these dual AM improved the performance of the network. In addition, the authors reported that the spatial information is more relevant for AM than the channel information. However, as it is a conventional network, it requires a large number of learnable parameters, which can be optimized with more efficient networks.

## 3. METHODOLOGY

In this section, we explain the proposed method, starting with the description of the proposed Multi-attention GhostNet architecture. In sequence, we describe the metrics used in the experimental analysis to measure the uncertainty of the method's predictions.

### 3.1 Multi-attention GhostNet

Inspired by the GhostNet architecture, introduced in (Han et al., 2020), we propose a fully connected architecture that includes a Spatial Attention Mechanism (SAM), for the deforestation detection task. As we are concerned with detecting changes, the network receives as input two co-registered images acquired at different dates, represented as  $I_{T_0}$  and  $I_{T_1}$ . The images are stacked along the spectral dimension, producing a tensor  $\mathbf{I} \in \mathbb{R}^{H \times W \times C}$ , where  $H$  and  $W$  denote the spatial dimensions, and  $C$  the number of image channels.

The proposed model follows a symmetric encoder-decoder architecture with skip connections, as can be observed in Figure 1(a). The encoder network is composed of several *Ghost blocks* with residual mappings. The structure of a Ghost block is illustrated in Figure 1(b)), it consists of two stacked *Ghost modules*, with a conventional (Conv) and a depthwise convolution (DWConv). The structure of a Ghost module is shown in Figure 1(c). This module starts by applying a primary convolution to produce the intrinsic feature map. Then, a series of cheap linear operations ( $\Phi_i$ ) are applied to obtain the final *ghost feature map*, through a depthwise convolution. During this process, an identity mapping is employed to preserve the intrinsic feature maps. The Spatial Attention Mechanisms (SAM) included in the skip connections help to combine low- and high-level feature maps (Section 3.2). The decoder network is composed of sequence of bilinear up-sampling and convolution operations, and its output is a tensor with the posterior class probabilities for all spatial locations.

### 3.2 Spatial Attention Mechanism (SAM)

The Spatial Attention Mechanism (SAM) (Woo et al., 2018) leverages from inter-spatial relationships of features to produce

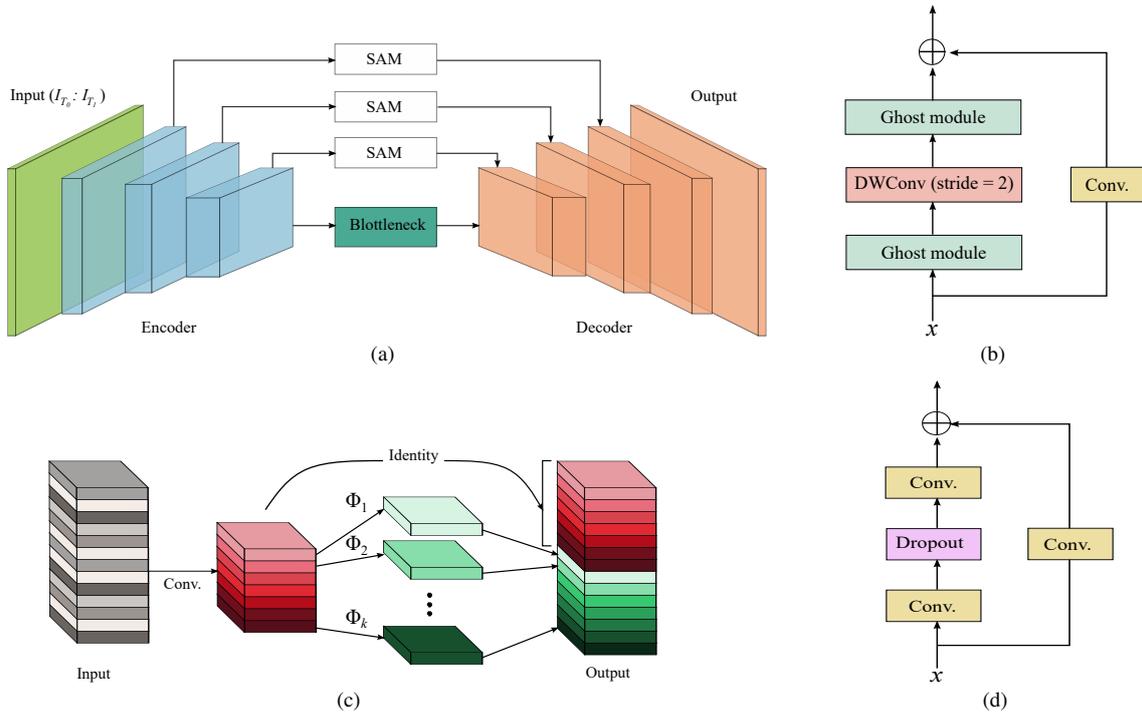


Figure 1. (a) Multi-attention DL architecture. (b) Ghost block, (c) Ghost module,  $\Phi$  represents the cheap operations. (d) Residual block.

a spatial attention map. Its structure is presented in Figure 2. To highlight informative regions, first average-pooling and max-pooling operations are computed and their results concatenated. That is followed by a convolution and a sigmoid layer, resulting on spatial attention  $\mathbf{M}$ s features, which are multiplied by the input  $\mathbf{F}$  to generate the output feature tensor  $\mathbf{F}'$ .

### 3.3 Uncertainty

In this section we present the three metrics used to quantify the predictive uncertainty: predictive variance, predictive entropy, and mutual information. In this work, we employed deep ensemble modeling, which is an effective strategy to measure uncertainty performance of supervised learners (Abdar et al., 2021). We trained several models, with the same architecture, but with different random weight initializations and random batch selections.

#### 3.3.1 Predictive Variance

Let's denote  $\{y^{(i)}(\mathbf{p})\}_{i=1}^n$  as the set of  $n$  different predictions (softmax) at pixel coordinate  $\mathbf{p}$  for all  $K$  classes. Also,  $y_k^{(i)}(\mathbf{p})$  stands for the  $i$ -th element of  $y^i(\mathbf{p})$  corresponding to the prediction for class  $k$  at pixel coordinate  $\mathbf{p}$ . Then, the final prediction  $\mu_k(\mathbf{p})$  for pixel  $\mathbf{p}$  and class  $k$  is the average over all  $n$  predictions  $y_k^i(\mathbf{p})$  in a pixel-wise fashion (Seeböck et al., 2019):

$$\mu(\mathbf{p}) = \frac{1}{n} \sum_{i=1}^n y_k^{(i)}(\mathbf{p}) \quad (1)$$

The variance for each class  $k$  is computed by:

$$\sigma_k^2(\mathbf{p}) = \frac{1}{n} \sum_{i=1}^n \left( y_k^{(i)}(\mathbf{p}) - \mu_k(\mathbf{p}) \right)^2 \quad (2)$$

The final predictive variance for pixel  $\mathbf{p}$  is obtained by averaging all estimate over the  $k$  class-specific variances:

$$u(\mathbf{p}) = \frac{1}{K} \sum_{k=1}^K \sigma_k^2(\mathbf{p}) \quad (3)$$

#### 3.3.2 Predictive Entropy

Entropy provides a measure of the average level of information or uncertainty inherent to the probable outcomes of a random variable (Shannon, 2001). Predictive Entropy can be defined as follows:

$$H(y(\mathbf{p})|\mathbf{x}(\mathbf{p})) \approx -\frac{1}{K} \sum_{k=1}^K \mu_k(\mathbf{p}) \log(\mu_k(\mathbf{p})) \quad (4)$$

#### 3.3.3 Mutual Information

Mutual information measures non-linear relations between two random variables. It expresses how much information can be obtained from a random variable by observing another random variable. The mutual information for a pixel  $\mathbf{p}$  is the difference between the predicted entropy computed on the final prediction and the average of the entropies of each prediction:

$$MI(y(\mathbf{p})|\mathbf{x}(\mathbf{p})) = H(y(\mathbf{p})|\mathbf{x}(\mathbf{p})) - \frac{1}{n} \sum_{i=1}^n H^{(i)}(y^{(i)}(\mathbf{p})|\mathbf{x}^{(i)}(\mathbf{p})) \quad (5)$$

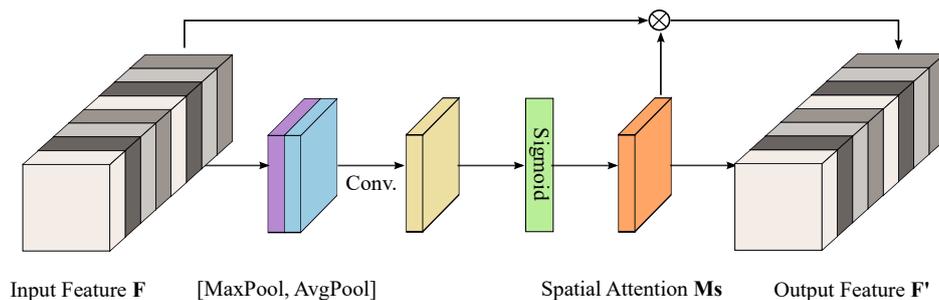


Figure 2. Spatial Attention Module.

#### 4. EXPERIMENTS

In the following sections we present the design and results of the detect deforestation experiments carried out in this work. We start by describing the dataset used to train and evaluate the proposed method. Next, we detail the experimental setup, and finally, we analyze the effects of including the spatial attention mechanism (SAM) in the network architecture. For that purpose, we compared the accuracy obtained with the proposed Multi-attention GhostNet with that of the fully convolutional GhostNet without SAM. To serve as baselines, we also trained and evaluated two variants of a compact ResU-Net model (with the same number of layers of the implemented GhostNet), with and without SAM.

##### 4.1 Study Area

The study area corresponds to a region of the BLA, located in Pará State, Brazil. The site is centered on coordinates of  $06^{\circ} 54' 16''$  South and  $055^{\circ} 11' 52''$  West (see Figure 3).

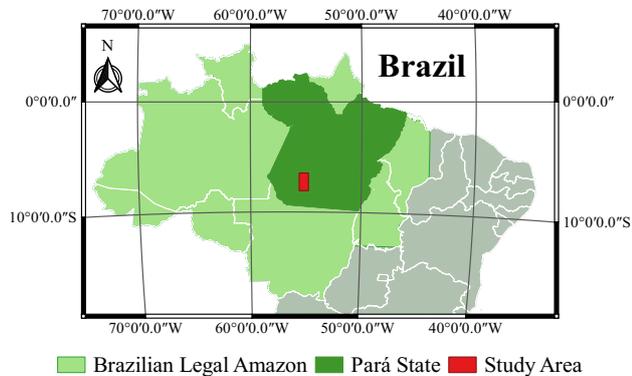


Figure 3. Geographical localization of the study area.

The dataset comprise two coregistered Sentinel-2 images, downloaded and preprocessed using the Google Earth Engine (GEE) platform (Gorelick et al., 2017). The images were processed to Level-1C, which means they are orthorectified, map-projected and contain top-of-atmosphere reflectance data. The input  $I$  to the networks tested here was a stack of image bands of size  $9200 \times 17730 \times 20$ , as we only considered the bands with 10m and 20m of spatial resolutions (we applied the nearest-neighbor interpolation to the 20m resolution bands). Moreover, each band was individually normalized to zero mean and unit variance.

Table 1 shows the acquisition dates of the images. We observe that as the study area is quite large, so for each epoch the inputs are mosaics of two Sentinel-2 scenes. The reference

data was obtained from the PRODES database (INPE, 2021). It represents the deforestation that occurred between the 21 July 2018 and the 24 July 2019.

Image	Date <sub>T0</sub>	Date <sub>T1</sub>
PRODES Ref.	[07-21-2018]	[07-24-2019]
Sentinel-2	[07-21-2018], [07-26-2018]	[07-21-2019], [07-26-2019]

Table 1. Acquisition dates of the Sentinel-2 images of the Pará state, and dates of the PRODES reference data.

We observe that the dataset is very unbalanced. According to the PRODES reference, only about 1.13% of the area is associated with the *deforestation* class, that is, with the deforestation that occurred during the selected period; 61.71% belong to the *no-deforestation* class; and 37.16% correspond to *past deforestation*, areas deforested prior to the selected period.

##### 4.2 Experimental Setup

The dataset was divided into 20 tiles, each one with a size of  $2300 \times 3546$  pixels, maintaining a distribution of 40%, 10%, and 50% for training, validation, and test, respectively. The network was trained on patches. In all experiments, patches of size  $128 \times 128$  pixels were extracted from the input image, with stride equal to 32.

Table 2 shows the GhostNet and ResU-Net architectures, with detailed information about the network layers. What the table does not show are the skip-connections between corresponding encoder and decoder layers (please refer to Figure 1(a)). Additionally, the following parameter values were used in all experiments: batch size equal to 32; Adam optimizer with learning rate equal to  $1e^{-3}$ , and  $\beta$  equal to 0.9. In order to prevent over-fitting, the early stopping strategy was used.

Considering that the dataset is highly unbalanced, we set the *weighted cross entropy* as a loss function with a vector of weights equals to  $[0.2, 0.8]$  for class *no-deforestation* and *deforestation*, respectively. Furthermore, to ensure that all the patches contain samples from both classes, only patches with at least 2% of pixels from the *deforestation* class were used for training. Data augmentation operations were employed for the training patches: rotation ( $90^{\circ}$ ), and flipping (horizontal, vertical) transformations.

In accordance with the PRODES methodology, we ignored pixels within a two pixel wide buffer at the inner and outer edges of all polygons identified as *deforestation* in the reference data. Those pixels were ignored for training, validation, and test. The same was done for areas (pixel clusters) smaller than 625 pixels (6,25 ha), and for all *past-deforestation* pixels.

Method	Encoder	Bottleneck	Decoder	Output	Number of Parameters	
					Without SAM	With SAM
ResU-Net	MP(RB(3 × 3, 16)) MP(RB(3 × 3, 32)) MP(RB(3 × 3, 64)) MP(RB(3 × 3, 128))	3 × RB(3 × 3, 128)	US(C(3 × 3, 128)) US(C(3 × 3, 64)) US(C(3 × 3, 32)) US(C(3 × 3, 16))	Softmax (C(1 × 1, #CL))	1.050.659	1.050.956
GhostNet	MP(GB(3 × 3, 16)) MP(GB(3 × 3, 32)) MP(GB(3 × 3, 64)) MP(GB(3 × 3, 128))	3 × GB(3 × 3, 128)	US(C(3 × 3, 128)) US(C(3 × 3, 64)) US(C(3 × 3, 32)) US(C(3 × 3, 16))	Softmax (C(1 × 1, #CL))	611.971	612.268

Table 2. ResU-Net and GhostNet Architectures. Symbols: C (strided convolution), MP (max-pooling), RB (residual block), GB (ghost block), US (bilinear up-sampling), CL (Output classes). The values in parenthesis refer to (Kernel Height x Kernel Height, Number of filters).

Finally, we observe that, as we employed a deep ensemble strategy, we trained independently 10 networks for each architecture, each with a different random initialization and random batch selections. Moreover, the classification results shown in the next section are based on the average probability, considering all networks in an ensemble, for each pixel position.

### 4.3 Results and Discussion

In this section, we present and analyze the results obtained using four methods: a standard ResU-Net; a Multi-attention ResU-Net; the fully convolutional GhostNet; and the Multi-attention GhostNet. The results are summarized quantitatively in terms of classification accuracy metrics values, and qualitatively, though deforestation probability maps and uncertainty maps.

The experimental results in terms of Precision vs. Recall curves, computed on the average prediction map, are shown in Figure 4. Additionally, Table 3 reports the Recall, Precision, F1-Score and Mean Average Precision (mAP) for the four methods. Observing the curves in Figure 4 it is possible to notice that the four architectures presented a similar tendency, close to ideal case (upper-right axis), with a slightly better performance for the GhostNet and the Multi-attention GhostNet, which is reflected in the mAP metrics in Table 3. It can be also observed in Table 3, that the GhostNet variants are consistently superior to the ResU-Net ones in terms of Recall, Precision and F1-score; and the Multi-attention GhostNet produced the best overall results.

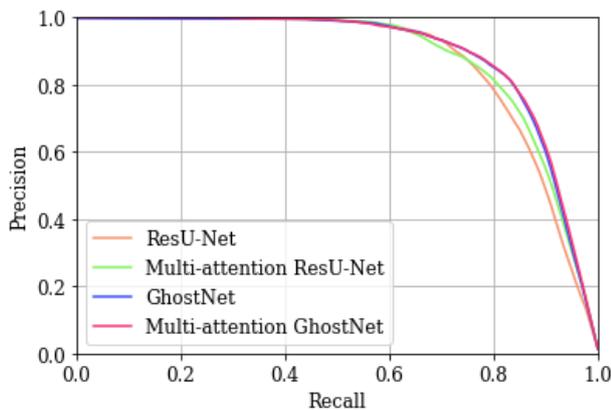


Figure 4. Recall vs. Precision curves of the deep learning architectures.

Method	Rec.	Prec.	F1	mAP
ResU-Net	71.64	89.64	79.63	87.31
Multi-attention ResU-Net	72.29	90.39	80.33	88.26
GhostNet	73.86	90.87	81.48	89.10
Multi-attention GhostNet	<b>73.98</b>	<b>91.30</b>	<b>81.73</b>	<b>89.45</b>

Table 3. Metrics results, in [%], for the four methods evaluated.

Figure 5 shows the deforestation probability maps produced with the four methods. Those maps also represent the average prediction map obtained from each model. It presents three different snips of test tiles. The first two columns on the left represent the co-registered pair of images from the different epochs (RGB composition), i.e.,  $T_0$  and  $T_1$ . The third column shows the reference mask, in which the blue color represents *no-deforestation*; the red color represents *deforestation*, and the black color, *past deforestation*. The four columns on the right contain the probability maps produced by each method. In the prediction maps for snips *a*) and *b*) one can notice that the Multi-attention GhostNet provided more confident values, as well as better defined polygons. In addition, considering snip *c*), it is possible to observe that part of the  $T_1$  image is covered by clouds. The ResU-Net variants identified some cloud parts as deforestation. The GhostNet variants were able to better classify those areas, demonstrating their superior robustness in that type of scenario.

Finally, we analyze the uncertainty associated with the predictions of the network ensembles in terms of Predictive Variance, Predictive Entropy and Mutual Information. Table 4 presents the uncertainty scores that correspond to the classification of pixel positions with each ensemble, by setting a threshold equal to 0.5 for the averaged output probabilities. The values in the table are averaged uncertainty values computed for pixels that correspond to: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

Although expected, it is interesting to observe that for all architectures the uncertainty associated with the correctly classified pixels (TP and TN) is much lower than for the incorrectly classified ones (FP and FN) – the ensembles usually fail when its components disagree to a larger extent. Additionally, in the case of the correctly classified pixels, lower uncertainties occur in the classification of *no-deforestation* (TN), that is, the ensembles are less confident in the classification of the *deforestation* class (TP). Also referring to Table 4: in most cases the architectures with SAM showed lower uncertainties than their counterparts; the GhostNet-based architectures provided lower uncertainties than the ResU-Net variants; and the proposed Multi-attention GhostNet consistently outperformed all other architectures in terms of the uncertainty of its predictions.

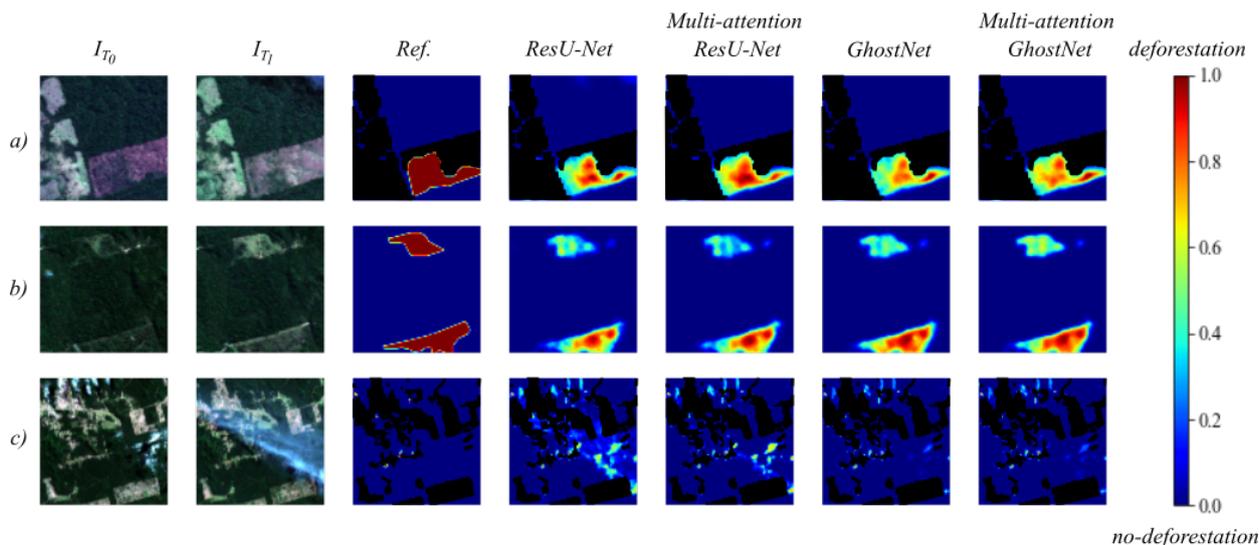


Figure 5. Deforestation probability maps of three different snips from the test tiles. From left to right: input snip of the 2018 image; input snip of the 2019 image; reference; probability maps (ResU-Net, Multi-attention ResU-Net, GhostNet, Multi-attention GhostNet). Blue and red colors represent lower and higher probability of belonging to the *deforestation* class, respectively. Black regions correspond to *past-deforestation*.

For a visual interpretation, Figure 6 shows the uncertainty maps obtained for each method over snip *c*). It is easy to notice that the plain ResU-Net architecture provided the highest uncertainty values. When SAM is included in the ResU-Net, however, uncertainty was reduced. Moreover, the GhostNet variants delivered more confident maps, i.e., with lower uncertainty values, the best of which are associated with the proposed Multi-attention GhostNet.

## 5. CONCLUSIONS

This work introduced a novel fully-convolutional architecture based on the GhostNet, which includes Spatial Attention Mechanisms (SAM). The model was employed in deforestation detection in a particular site in the Amazon region, a problem that is characterized by a high-class imbalance.

The experimental results demonstrated that the proposed Multi-attention GhostNet consistently outperformed the baseline approaches: ResU-Net, Multi-attention ResU-Net, and GhostNet without SAM, considering all classification accuracy metrics evaluated. Furthermore, the results also showed that the inclusion of SAM in both ResU-Net and GhostNet led to improvements in classification accuracy.

Additionally, according to three uncertainty metrics, we investigated the predictions of all methods in terms of their uncertainties. We found that the inclusion of SAM in both the ResU-Net and GhostNet architectures led to less uncertainty in the predictions. The GhostNet-based architectures were superior to the ResU-Net-based ones in that respect. In conclusion, the Multi-attention GhostNet model produced the most accurate and less uncertain classification maps in the deforestation detection task, at least for the study area considered in this work.

In the future, we plan to enrich the experimental analysis by considering different sites, with varying types of forest, in the Amazon and the Brazilian Cerrado (Savannah) biome. We also want to assess the generalization capacity of the proposed

model and possibly use it as a backbone for Domain Adaptation solutions. Finally, we plan to investigate ways to employ uncertainty information to improve semantic segmentation results in change detection tasks.

## 6. ACKNOWLEDGEMENTS

Authors would like to thank Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), and Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) for the financial support.

## REFERENCES

- Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U. R. et al., 2021. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information Fusion*.
- Alam, M., Wang, J.-F., Guangpei, C., Yunrong, L., Chen, Y., 2021. Convolutional Neural Network for the Semantic Segmentation of Remote Sensing Images. *Mobile Networks and Applications*, 26(1), 200–215.
- Bahdanau, D., Cho, K., Bengio, Y., 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.
- Bragagnolo, L., da Silva, R., Grzybowski, J., 2021a. Towards the automatic monitoring of deforestation in Brazilian rainforest. *Ecological Informatics*, 66, 101454.
- Bragagnolo, L., da Silva, R. V., Grzybowski, J. M. V., 2021b. Amazon forest cover change mapping based on semantic segmentation by U-Nets. *Ecological Informatics*, 62, 101279.
- Chen, L., Zhang, D., Li, P., Lv, P., 2020. Change Detection of Remote Sensing Images Based on Attention Mechanism. *Computational Intelligence and Neuroscience*, 2020.

Uncertainty metric	Method	TP	TN	FP	FN
Predictive Variance	ResU-Net	0.0120	0.0012	0.0597	0.0156
	Multi-attention ResU-Net	0.0089	0.0006	0.0477	0.0151
	GhostNet	0.0082	0.0004	0.0236	0.0125
	Multi-attention GhostNet	<b>0.0068</b>	<b>0.0003</b>	<b>0.0231</b>	<b>0.0119</b>
Predictive Entropy	ResU-Net	0.1220	0.0188	0.2914	0.2789
	Multi-attention ResU-Net	0.1151	0.0259	0.2801	0.2749
	GhostNet	0.1120	0.0160	0.2574	0.2697
	Multi-attention GhostNet	<b>0.1053</b>	<b>0.0153</b>	<b>0.2539</b>	<b>0.2669</b>
Mutual information	ResU-Net	0.0073	0.0056	0.0516	0.0319
	Multi-attention ResU-Net	0.0107	0.0027	0.0633	0.0288
	GhostNet	0.0062	0.0026	0.0219	0.0252
	Multi-attention GhostNet	<b>0.0051</b>	<b>0.0020</b>	<b>0.0211</b>	<b>0.0238</b>

Table 4. Uncertainty metrics for True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) computed with the average predictions of the four methods.

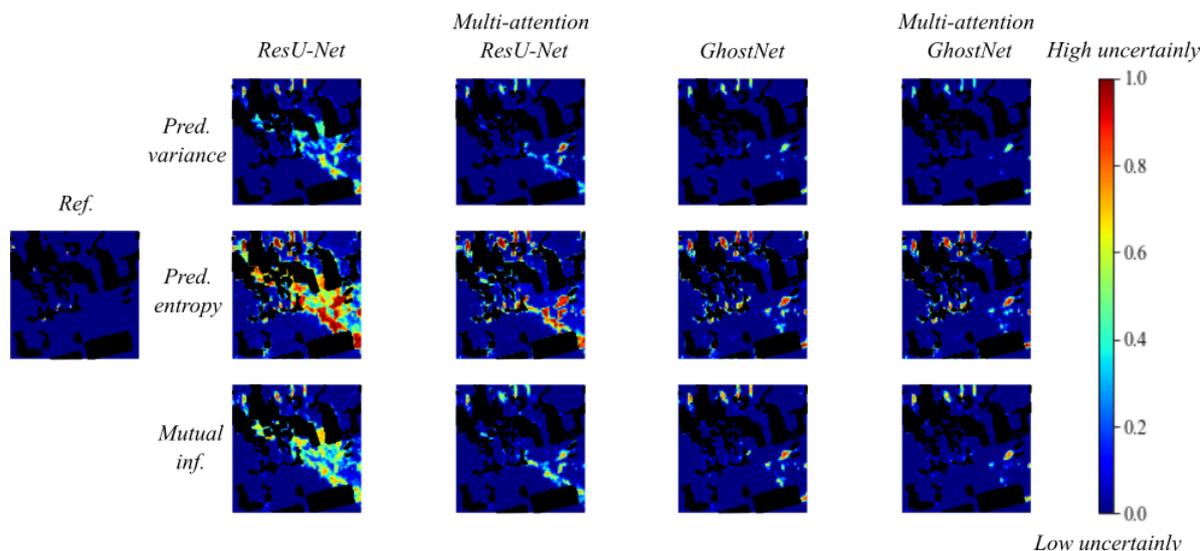


Figure 6. Uncertainty maps of snip c), From left to right: reference; uncertainty maps (ResU-Net, Multi-attention ResU-Net, GhostNet, Multi-attention GhostNet). The rows correspond to: predictive variance, entropy and mutual information. Blue and red colors represent lower and higher uncertainty scores, black regions correspond to *past-deforestation*.

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).

Gao, F., He, Y., Wang, J., Hussain, A., Zhou, H., 2020. Anchor-free convolutional network with dense attention feature aggregation for ship detection in SAR images. *Remote Sensing*, 12(16), 2619.

Ghaffarian, S., Valente, J., Van Der Voort, M., Tekinerdogan, B., 2021. Effect of attention mechanism in deep learning-based remote sensing image processing: A systematic literature review. *Remote Sensing*, 13(15), 2965.

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*, 202, 18–27.

Guo, Q., Zhang, J., Zhu, S., Zhong, C., Zhang, Y., 2021. Deep Multiscale Siamese Network with Parallel Convolutional Structure and Self-Attention for Change Detection. *IEEE Transactions on Geoscience and Remote Sensing*.

Han, K., Wang, Y., Tian, Q., Guo, J., Xu, C., Xu, C., 2020. Ghostnet: More features from cheap operations. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1580–1589.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H., 2017. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.

Hu, J., Shen, L., Sun, G., 2018. Squeeze-and-excitation networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7132–7141.

INPE, 2021. National institute for space research. general coordination of earth observation. monitoring program of the amazon and other biomes. deforestation - legal amazon -. <http://terrabrasilis.dpi.inpe.br>.

Jiang, H., Hu, X., Li, K., Zhang, J., Gong, J., Zhang, M., 2020. Pga-siamnet: Pyramid feature-based attention-guided siamese network for remote sensing orthoimagery building change detection. *Remote Sensing*, 12(3), 484.

Koščević, K., Subašić, M., Lončarić, S., 2019. Attention-based convolutional neural network for computer vision color constancy. *2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA)*, IEEE, 372–377.

- Li, W., Liu, K., Zhang, L., Cheng, F., 2020. Object detection based on an adaptive attention mechanism. *Scientific Reports*, 10(1), 1–13.
- Lorenz, C., de Oliveira Lage, M., Chiaravalloti-Neto, F., 2021. Deforestation hotspots, climate crisis, and the perfect scenario for the next epidemic: The Amazon time bomb. *Science of the Total Environment*, 783, 147090.
- Lu, D., Wang, L., Cheng, S., Li, Y., Du, A., 2021. CANet: A Combined Attention Network for Remote Sensing Image Change Detection. *Information*, 12(9), 364.
- Morris, R. J., 2010. Anthropogenic impacts on tropical forest biodiversity: a network structure and ecosystem functioning perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1558), 3709–3718.
- Niu, Z., Zhong, G., Yu, H., 2021. A review on the attention mechanism of deep learning. *Neurocomputing*, 452, 48–62.
- Ortega, M. X., Feitosa, R. Q., Bermudez, J. D., Happ, P. N., De Almeida, C. A., 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.
- Paoletti, M. E., Haut, J. M., Pereira, N. S., Plaza, J., Plaza, A., 2021. Ghostnet for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*.
- Pereira, E. J. d. A. L., de Santana Ribeiro, L. C., da Silva Freitas, L. F., de Barros Pereira, H. B., 2020. Brazilian policy and agribusiness damage the Amazon rainforest. *Land Use Policy*, 92, 104491.
- Qing, Y., Liu, W., 2021. Hyperspectral Image Classification Based on Multi-Scale Residual Network with Attention Mechanism. *Remote Sensing*, 13(3), 335.
- Seeböck, P., Orlando, J. I., Schlegl, T., Waldstein, S. M., Bogunović, H., Klimescha, S., Langs, G., Schmidt-Erfurth, U., 2019. Exploiting epistemic uncertainty of anatomy segmentation for anomaly detection in retinal OCT. *IEEE transactions on medical imaging*, 39(1), 87–98.
- Shannon, C. E., 2001. A mathematical theory of communication. *ACM SIGMOBILE mobile computing and communications review*, 5(1), 3–55.
- Srivastava, V., Biswas, B., 2020. CNN-based salient features in HSI image semantic target prediction. *Connection Science*, 32(2), 113–131.
- Tovar, P., Adarme, M., Feitosa, R., 2021. Deforestation Detection in the Amazon Rainforest with Spatial and Channel Attention Mechanisms. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, 851–858.
- Wagner, F. H., Sanchez, A., Tarabalka, Y., Lotte, R. G., Ferreira, M. P., Aidar, M. P., Gloor, E., Phillips, O. L., Aragao, L. E., 2019. Using the U-net convolutional network to map forest types and disturbance in the Atlantic rainforest with very high resolution images. *Remote Sensing in Ecology and Conservation*, 5(4), 360–375.
- Woo, S., Park, J., Lee, J.-Y., Kweon, I. S., 2018. Cbam: Convolutional block attention module. *Proceedings of the European conference on computer vision (ECCV)*, 3–19.
- Xue, Z., Yu, X., Liu, B., Tan, X., Wei, X., 2021. HResNetAM: Hierarchical Residual Network With Attention Mechanism for Hyperspectral Image Classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 3566–3580.
- Zeng, Y., Ritz, C., Zhao, J., Lan, J., 2020. Attention-based residual network with scattering transform features for hyperspectral unmixing with limited training samples. *Remote Sensing*, 12(3), 400.
- Zhang, J., Zhou, Q., Wu, J., Wang, Y., Wang, H., Li, Y., Chai, Y., Liu, Y., 2020. A Cloud Detection Method Using Convolutional Neural Network Based on Gabor Transform and Attention Mechanism with Dark Channel Subnet for Remote Sensing Image. *Remote Sensing*, 12(19), 3261.
- Zhang, X., Zhou, X., Lin, M., Sun, J., 2018. Shufflenet: An extremely efficient convolutional neural network for mobile devices. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 6848–6856.