EVALUATION OF SEMANTIC SEGMENTATION METHODS FOR DEFORESTATION DETECTION IN THE AMAZON

R. B. Andrade[1], G. A. O. P. Costa[1,*], G. L. A. Mota[1], M. X. Ortega[2],
R. Q. Feitosa[2], P. J. Soto[2], C. Heipke[3]

[1] Dept. of Informatics and Computer Science, Rio de Janeiro State University (UERJ), Brazil
renanbides@gmail.com, (gilson.costa, guimota)@ime.uerj.br
[2] Dept. of Electrical Engineering, Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Brazil
(mortega, raul, psoto)@ele.puc-rio.br
[3] Institute of Photogrammetry and GeoInformation, Leibniz Universität Hannover (LUH), Germany
heipke@ipi.uni-hannover.de

Commission III, WG III/7

KEY WORDS: Amazon Forest, Deforestation, Semantic Segmentation, Change Detection, Deep Learning, DeepLabv3+

ABSTRACT:

Deforestation is a wide-reaching problem, responsible for serious environmental issues, such as biodiversity loss and global climate change. Containing approximately ten percent of all biomass on the planet and home to one tenth of the known species, the Amazon biome has faced important deforestation pressure in the last decades. Devising efficient deforestation detection methods is, therefore, key to combat illegal deforestation and to aid in the conception of public policies directed to promote sustainable development in the Amazon. In this work, we implement and evaluate a deforestation detection approach which is based on a Fully Convolutional, Deep Learning (DL) model: the DeepLabv3+. We compare the results obtained with the devised approach to those obtained with previously proposed DL-based methods (Early Fusion and Siamese Convolutional Network) using Landsat OLI-8 images acquired at different dates, covering a region of the Amazon forest. In order to evaluate the sensitivity of the methods to the amount of training data, we also evaluate them using varying training sample set sizes. The results show that all tested variants of the proposed method significantly outperform the other DL-based methods in terms of overall accuracy and F1-score. The gains in performance were even more substantial when limited amounts of samples were used in training the evaluated methods.

1. INTRODUCTION

Covering an area of approximately 5.5 million km^2, which is equivalent to approximately one third the size of the South American continent, the Amazon rainforest encompasses half of the remaining tropical forest area on the planet (World Wildlife Fund, 2020a). Home to the largest collection of plants and animal species on the planet, the Amazon biome contains unparalleled biodiversity: it is the natural habitat of one tenth of the known species in the world (The Worldwatch Institute, 2015).

The forest covers most of the Amazon river basin, source of 20% of all free-flowing fresh water on Earth (Assunção, Rocha, 2019). Additionally, the Amazon forest produces vast quantities of water for most of South America's extents. The so-called "flying rivers", formed by masses of air loaded with water vapor generated through evapotranspiration, carry moisture to most of Brazil and regulate rainfall regimes in the central, southeast and southern regions of South America (Lovejoy, Nobre, 2018). The induced rain is responsible for irrigating crops and for filling the rivers and dams used to generate electrical energy by a large number of hydropower plants.

Moreover, tropical rainforests store from 90 to 140 billion metric tons of carbon, and are known to help stabilize the worldwide climate. The Amazon forest alone contains 10% of all biomass on the planet (De Sy et al., 2015). Unfortunately, for decades the Amazon biome has faced several threats as a result of unsustainable economic development, primarily caused by the extension of agricultural activities at industrial scale, such as soybean cropping and cattle farming, forest fires, illegal mining and logging and expansion of informal settlements (Goodman et al., 2019, Malingreau et al., 2012, Nogueron et al., 2006). All these factors are directly associated with deforestation.

According to the National Institute for Space Research (INPE) (Shimabukuro et al., 2013), deforestation accelerated significantly in the Brazilian Legal Amazon area during the 1990's and early 2000's. Likewise, the World Wildlife Fund (World Wildlife Fund, 2020b) estimates that more than a quarter of the rainforest will vanish by 2030, if the current rate of deforestation continues.

Deforestation is one of the largest sources of CO_2 emissions related to anthropogenic activities. It is a wide-reaching problem, responsible for the reduction of carbon storage, greenhouse gas emissions, and other serious environmental issues such as biodiversity losses and climate change (De Sy et al., 2015).

The above mentioned facts indicate the importance of the preservation of the Amazon biome, and Remote Sensing (RS) data provide key capability to monitor this environment. It can be used not only in the combat of illegal activities, but also in the planning and development of public policies to promote sustainable development in the region (Sathler et al., 2018).

Since the late 1980's, the Brazilian National Institute for Space Research (INPE) has been using RS data to monitor the Brazi-

* Corresponding author

lian Legal Amazon area (BLA). Since 1988, the Amazon Deforestation Monitoring Project (PRODES) has produced annual reports about deforestation of native vegetation in the BLA, delivering deforestation maps derived from Landsat images (Valeriano et al., 2004). Relying on MODIS data the Near Real-Time Deforestation Detection project (DETER-A) started in 2004 to support actions from governmental agencies against illegal deforestation (Shimabukuro et al., 2006). With the change in the deforestation patterns observed in the last decade, in which most deforestation polygons started to show areas of less than 25 ha, a new version of the project, the DETER-B, was launched in 2015, in order to monitor, on a daily basis, changes in vegetation cover of as small as 1 ha, from the WFI/CBERS-4 and AWiFS/IRS sensor systems (Diniz et al., 2015).

All the above mentioned projects, however, rely mostly on visual interpretation and manual operations. This is due basically to the high level of accuracy expected for the official information provided by those projects to different stakeholders. There is, therefore, a demand for automatic methods that can support such projects in ways that can further improve the accuracies obtained and, at the same time, diminish the need for human intervention, so as to improve their response times.

In this work, we evaluate an approach based on a specific Deep Learning (DL) Fully Convolutional Network (FCN) architecture, the DeepLabv3+ (Chen et al., 2018b), which we adapted to deforestation change detection. We also compare the results obtained with this approach to the ones reported in a previous work (Ortega et al., 2019) over the same study area. Additionally, we investigate the different methods' demands for training samples in relation to the delivered accuracies. In short, the major contributions in this work are:

- We adapted the previously proposed semantic segmentation method DeepLabv3+ to deforestation change detection, and evaluated the method's performance over an area of the Amazon forest.
- We evaluated the impact of varying hyperparameter values of the proposed method in the deforestation detection accuracy.
- We evaluated the sensitivity of the proposed method to the amount of labeled samples used in training stage.
- We compared the performance of the proposed method to those delivered by deep learning-based methods previously employed in deforestation detection in the Amazon region.

The remainder of this article is organized as follows. In the next section we review related works. Section 3 describes the DeepLabv3+ architecture, while section 4 presents the adaptations of the original module carried out in this research. In Section 4 we also describe the other change detection methods, to which the proposed method is compared. Section 5 is dedicated to the description of the experiments. In Section 6 we present and analyze the results of the experiments, and in Section 7 we present the conclusions and directions for further work.

2. RELATED WORK

This section presents some DL patch-wise classification and semantic segmentation approaches.

2.1 Patch-wise Classification

Patch-wise classification change detection produces a global decision by considering two distinct patches of the same object acquired at distinct time instances. Among the outstanding methods that can be found in the literature, (Chu et al., 2016) proposes a CD method that uses a pair of Deep Belief Networks (DBN), one for each patch. A modified backpropagation algorithm minimizes the DBNs' outputs distances for non-changed examples and maximizes them for the changed ones. The DBN outputs are submitted to PCA/k-means clustering, which produces the final result. In experiments using very high resolution (VHR) images of urban areas, the method outperformed traditional approaches.

A similar idea underlies the Siamese Convolutional Neural Networks (S-CNN) applied in (Daudt et al., 2018), which corresponds to a pair of convolutional nets with shared weights. Convolutional outputs are concatenated and a fully connected network delivers the decision. An alternative approach to S-CNN is also presented by the authors: the so-called Early Fusion (EF), which consists of concatenating two image pairs as the input of the convolutional network. In the experiments, Sentinel-2 RS images of urban areas were employed to compare the performance of EF and S-CNN to some baseline methods. The authors reported that the EF and S-CNN delivered the best results, and that the EF method was slightly superior to the S-CNN method.

2.2 Semantic Segmentation

Semantic segmentation is characterized by producing pixel level decisions, in contrast to global patch-wise classification. A successful method for the semantic segmentation of VHR images is presented in (Wang et al., 2019). It employs an ensemble of several multiscale multiconnection ResNets and a class-specific attention model. Experiments compared it to six state-of-the-art models, including the DeepLabv3+, in two urban benchmark datasets (Mnih, 2013, ISPRS, 2020). In order to reduce the loss of spatial features and strengthen object boundaries, the so-called dense-coordconv network (DCCN) was proposed in (Yao et al., 2019). In the experiments, the authors compared DCCN with other deep convolutional neural networks (U-net, SegNet, DeepLabv3), showing that DCCN delivered higher accuracies. Aiming at increasing the robustness of segmentation in blurred or partially damaged VHR RS images, (Peng et al., 2019) proposes the RobustDenseNet and the use of multimodal data (NIR, RGB and DSM). Experiments compared the proposed model with DeepLabv3+ on the ISPRS Postdam 2D dataset (ISPRS, 2020), with randomly added motion blur to spectral data, and randomly deleted colors of small areas. The results show the superiority of the proposed model over DeepLabv3+. (Guo et al., 2020) used a pre-trained modified aligned Xception (Chollet, 2016) and the DeepLabv3+ model combined with transfer learning strategies for extraction of snow cover from high spatial resolution RS images.

An FCN architecture inspired by U-Net is presented in (de Jong, Bosman, 2019). One of its benefits is that it can take advantage of previously trained U-Nets. Another important aspect is its multi-scale structure, which is able to generate multiple scales difference images. This method was applied to high resolution RS images of the ISPRS Vaihingen dataset (ISPRS, 2020), achieving over 90% overall accuracy. A Fully Atrous convolutional neural network (FACNN) architecture for semantic segmentation and CD was introduced in (Zhang et al., 2019). Test-

ing results using VHR images showed that the FACNN significantly outperforms several recent FCN models (FCN-16, U-Net, Dense-Det, DeepLabv3 and SR-FCN (Ji et al., 2019)) in land cover classification.

2.3 Deforestation Mapping

The work (Ortega et al., 2019) presents an evaluation of methods for automatic deforestation detection, an Early Fusion (EF) Convolutional Network and a Siamese Convolutional Network (S-CNN), taking a Support Vector Machine (SVM) as baseline. These patch classification methods were evaluated on a dataset covering the Brazilian Legal Amazon. The DL-based approaches outperformed the SVM baseline, both in terms of F1-score and Overall Accuracy, with a superiority of S-CNN over EF.

To the best of our knowledge, (Ortega et al., 2019) is the only research reported in the literature dedicated to monitoring Amazon deforestation based on DL. The research presented herein uses the same dataset. But, instead of employing patch-wise classification methods, this work is dedicated to the investigation of the benefits of using a particular semantic segmentation architecture, the DeepLabv3+ in deforestation monitoring.

3. THE DEEPLAB V3+ MODEL

The first DeepLab module was proposed in (Chen et al., 2014). Its major novelty is the particular implementation of the 'hole algorithm', which has been previously proposed for improving efficiency on the computation of the undecimated wavelet transform (Mallat, 2018), and came to be known in the DL field by the terms atrous or dilated convolution. Dilated convolution is designed to enlarge the field-of-view of traditional filters, thus incorporating larger image contexts, without increasing the number of parameters or the amount of computations.

The second DeepLab version introduced atrous spatial pyramid pooling (ASPP). ASPP were designed to probe a feature layer with filters with different fields-of-view, by performing a sequence of dilated convolutions with different sampling rates, thus capturing image context at multiple scales (Chen et al., 2018a).

The first two versions of DeepLab relied on a fully connected Conditional Random Field (CRF) (Krähenbühl, Koltun, 2011) to enhance the level of detail (e.g., of object boundaries) of the outcome of the convolutional networks. The CRF component was droped in the third DeepLab module (Chen et al., 2017), in which the ASPP component was augmented by using image-level features, or image pooling (Liu et al., 2015), which encode global image context.

Finally, the DeepLabv3+ model (Chen et al., 2018b) adopts an encoder-decoder structure using the original DeepLabv3 as an encoder. The simple decoder module was devised to enhance segmentation results especially along object boundaries. Specifically, the last feature map before the logits layer in the original DeepLabv3 model is used as the encoder output.

The encoder features from DeepLabv3 are usually computed with an output stride of 16, the output stride being the ratio of the input image spatial resolution to the output feature map spatial resolution. In order to recover object segmentation details which are not recovered through a simple bilinear upsampling by a factor 16, the encoder output feature map in DeepLabv3+ is first bilinearly upsampled by a factor of 4 and then concatenated with the corresponding low-level feature map from the network backbone that has the same spatial resolution. Before concatenation, a 1×1 convolution is applied on the low-level features to reduce the number of channels associated with those features, so that they do not outweigh the importance of the encoder's output. After concatenation, 3×3 convolutions are applied to refine the features, finally followed by another simple bilinear upsampling by a factor of 4 (Chen et al., 2018b).

In (Chen et al., 2018b) the authors also used the Xception model (Chollet, 2016), adapted for the task of semantic segmentation, as the backbone of the encoder network. A deeper model, as in (Dai et al., 2017), was used. The max pooling operations were replaced by depthwise separable convolution with striding, and batch normalization (Ioffe, Szegedy, 2015) and ReLU activation were added after each 3×3 depthwise convolution, as in the MobileNet design (Howard et al., 2017).

4. DEFORESTATION DETECTION METHODS

In this section, we shortly describe the methods evaluated in this work for deforestation detection, the previously proposed Early Fusion (EF) and Siamese Convolutional Network (S-CNN), and the one introduced in this work, based on fully convolutional semantic segmentation with the DeepLabv3+ model (DL).

4.1 Early Fusion (EF)

The EF method is based on the CNN model proposed in (Daudt et al., 2018), originally employed for change detection in urban areas. It is composed of a number of convolutions and pooling layers, followed by a fully connected layer and a softmax layer, to carry out the final classification.

The term Early Fusion is related to the concatenation of co-registered images from two different epochs, before further processing. The images are stacked along the spectral dimension to generate a unique (synthetic) input image for subsequent patch extraction.

The image patches are defined and extracted through a sliding window procedure. Each patch is submitted to classification, and the corresponding class label is assigned to the central pixel of each patch.

The EF model evaluated in this work is composed of three convolutional layers (Conv) with ReLU as activation function, two max-pooling (MaxPool) layers and two fully connected layers, the last one being a softmax with two outputs, associated with the deforestation and no-deforestation classes. The input tensor has dimensions $15\times15\times16$ ($H\times W\times C$), the first Conv layer has 96 $7\times7\times16$ filters, using padding. It is followed by a MaxPool (2×2) layer, which generates a $7\times7\times96$ tensor. The second Conv layer has 192 $5\times5\times96$ filters, it is also followed by a MaxPool (2×2) layer, resulting in a $3\times3\times192$ tensor. The last Conv layer has 256 $3\times3\times192$ filters, generating a $3\times3\times256$ tensor which is flattened (amounting to a 2304 feature vector) and connected to the softmax layer.

4.2 Siamese Convolutional Network (S-CNN)

A Siamese convolutional network comprises two identical CNN branches that share the same hyperparameters and weight values (Zhang et al., 2018). The model evaluated in this work is also based in the work of (Daudt et al., 2018).

Corresponding patches from co-registered images of two different epochs are processed individually on each branch of the network, generating feature vectors that are concatenated and associated to a fully connected layer followed by a softmax layer with two outputs (Zhang et al., 2018). Similar to the EF approach, a class label is assigned to the central pixel of each patch.

Each branch of the network is similar to the architecture described in the previous section, with the difference that the input tensor has dimensions $15\times15\times8$, and the filters in the first Conv layer have sizes $7\times7\times8$. Also, the final feature vectors are concatenated into a 4608 sized one, and connected to the softmax layer.

4.3 DeepLab-based Change Detection (DLCD)

As in the EF approach, the proposed DeepLab-based deforestation detection (DLCD) technique takes as input a synthetic image, created through stacking along the spectral dimension two co-registered images from different epochs. But different from the previously described methods, which were devised for patch, or image classification in the computer vision terminology, DLCD delivers dense labeling of input patches through a FCN.

Since the size of the deforestation areas (or objects) in the dataset evaluated in this work are much smaller than the objects present in the images of the datasets tested in (Chen et al., 2018b) (i.e., PASCAL VOC 2012 and Cityscapes), and the vast majority of the samples/pixels are of the no-deforestation (or background) class, we experimented with patches of smaller sizes than in (Chen et al., 2018b) and obtained better and more consistent results with a patch size of 64×64 pixels.

In order to adjust the network architecture to the selected input patch size and to an output stride of 8, we changed the rates of the convolutions in the atrous spatial pyramid pooling to 3 and 6 (originally, rates 12 and 24 were employed), and removed the convolution with the highest rate because it would degenerate into a 1×1 convolution. We also used rate 1 in the middle block, and rates 1 and 2 in the exit blocks convolutions of the adapted Xception backbone (originally, rates 2 and 4 were employed in the exit blocks). Figure 1 shows the architecture of the DLCD model.

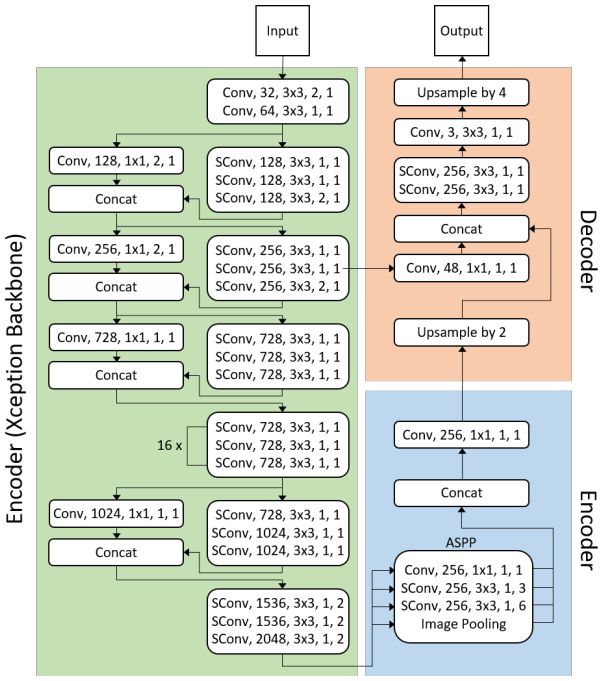


Figure 1. DLCD model. Layer descriptions contain: convolution type (Conv for regular convolution; SConv for depthwise separable convolution), number of filters, filter size, stride, dilation rate.

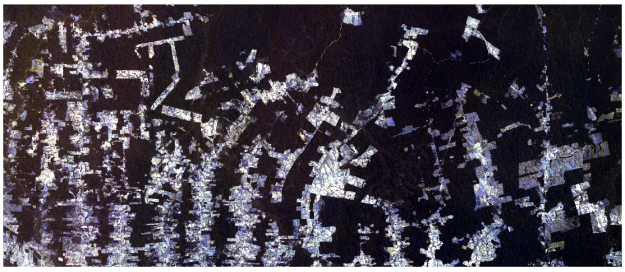


Figure 2. T1: August, 2016.

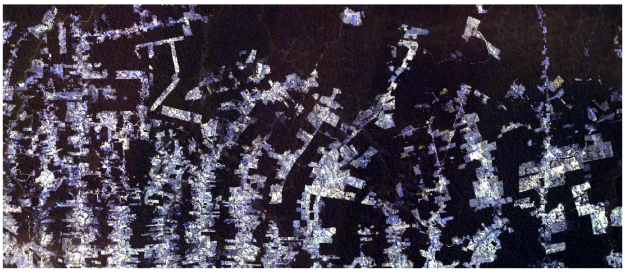


Figure 3. T2: July, 2017.

5. EXPERIMENTS

5.1 Data Set Description

The study area is located in the Brazilian Legal Amazon, more specifically in Pará State, Brazil, centered on coordinates of 03° 17' 23" S and 050° 55' 08" W. This area has faced a significant deforestation process in the period tracked and monitored by PRODES (Valeriano et al., 2004).

Figure 2 shows the study area on August 2nd, 2016 and Figure 3 shows the same area on July 20th, 2017. These dates were chosen due to the lower presence of clouds, a common problem over all the Brazilian Legal Amazon region.

Figure 4 shows the reference change map of deforestation that occurred between December 2016 and December 2017. This data is freely available at the PRODES database (http://terrabrasilis.dpi.inpe.br/map/deforestation). However, some polygons of the reference were not considered in the analysis, because they had been deforested in previous years.

The reference deforestation polygons represent transitions from forest to no-forest. In this work image pixels that intersect such polygons are considered as samples of the deforestation class. The other pixels are associated with the no-deforestation class, which includes areas where the forest cover remains unchanged and areas where deforestation has previously occurred.

The dataset comprises a pair of Landsat 8-OLI images, with

30m spatial resolution. We applied an atmospheric correction to each scene, and clipped them to the target area. The final images have 1100×2600 pixels and seven spectral bands (Coastal/Aerosol, Blue, Green, Red, NIR, SWIR-1, and SWIR-2). Following (Ortega et al., 2019), we also included an additional band in those images, which corresponds to the Normalized Difference Vegetation Index (NDVI) (Carlson, Ripley, 1997) calculated for every pixel using the Red and the NIR bands.

The dataset is extremely imbalanced considering the ratio of the deforested area in the studied period to the area in which deforestation did not occur. Table 1 shows the proportions of the deforestation area in relation to the total area in the study region. The training, validation and test set rows in Table 1 show the proportions in relation to the total area covered by the tiles considered in the respective sets (see next section).

We observe that highly imbalanced datasets pose a challenge for predictive modeling as the learning process of most classification algorithms is often biased toward the majority class examples, so that minority ones are not well modeled into the final system (Guo, Viktor, 2004). As described in the next section, in this work we explore different techniques to deal with class imbalance in the training process of the proposed method.

Table 1. Deforestation area in the study region.

Deforestation	Area (pixels)	Proportion (%)
Total	72298	2.6
Training set	24438	3.3
Validation set	8807	2.3
Test set	39053	2.3

5.2 Experimental Setup

Following (Ortega et al., 2019) the experiments relied on the two optical images mentioned in the previous section. As an NDVI band was stacked along the spectral dimension of the corresponding images, the resulting input images for the deforestation detection methods comprise eight bands, which were normalized to zero mean and unit variance.

We divided the input images into tiles of the same size and obtained a total of 15 tiles. Tiles 1, 7, 9 and 13 were used for training, tiles 5 and 12 for validation and tiles 2, 3, 4, 6, 8, 10, 11, 14 and 15 for test. Figure 4 shows the image tile locations and the corresponding reference deforestation areas (in blue color).

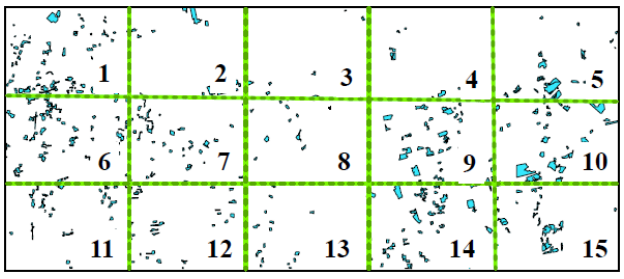


Figure 4. Tile references and deforestation polygons.

In order to evaluate the sensitivity of the methods to the amount of training data, we ran experiments considering four different scenarios: using training samples from a single tile (13); from two tiles (1 and 13); from three tiles (1, 7 and 13); and from four tiles (1, 7, 9 and 13). Table 2 shows the proportions of the deforestation area in relation to the total area of the tiles considered in the different training scenarios.

Table 2. Deforestation area in the training scenarios.

Training tiles	Area (pixels)	Proportion (%)
1 tile	2137	1.1
2 tiles	12112	3.3
3 tiles	16376	2.9
4 tiles	24438	3.3

The patch sizes for the EF and S-CNN methods were set to 15×15. During the training procedure of both methods, data augmentation was performed only on patches associated to the deforestation class, i.e., for which the central pixel belongs to the deforestation class. Each training patch was rotated by 90°, and flipped in the horizontal and vertical axis.

Additionally, under-sampling was employed for the no-deforestation class to balance the number of training patches for both classes. In this way, 8,118 training pairs of patches were obtained for each class. The validation set was composed by a total of 40,642 pairs of patches, 963 of the deforestation class, and 39,679 of no-deforestation class, which corresponds to the class distribution in the test set, which comprises 1,716,000 pairs, of which 40,392 were deforestation pairs and 1,675,608 no-deforestation pairs.

For training the EF and S-CNN methods the batch size was set to 32. We used early stopping to break after 10 epochs without improvement and a dropout rate of 0.2 was set for the last fully connected layer. We employed the Adam optimizer, with the learning rate of 0.001 and weight decay of 0.9.

For the DLCD method we used patches of 64×64, with an overlap of 48×48 pixels. The selected training patches were subjected to data augmentation, they were rotated by 90°, 180°, 270°; and the original and rotated versions of the patches were flipped vertically. Patches with no deforestation pixels were not used in the training procedure.

The batch size was set to 16, and the number of epochs was set to 100. We also used early stopping, to halt the procedure after 10 epochs without improvement. Adam optimizer with a learning rate of 0.001 was used in training.

As mentioned before, in order to deal with the high class imbalance between the majority class (no-deforestation) and the minority class (deforestation), in the case of the EF and S-CNN methods, we balanced the training set so that it contains the same number of deforestation and no-deforestation patches. This was possible because those techniques perform patch-wise classification, so that a patch can be labeled as belonging to a particular class (that of its center pixel).

In the case of the proposed DLCD method, which assigns a label to each pixel within a patch, it is not so simple to balance the training set. We could have decided to select for training patches with a higher proportion of deforestation pixels, say 10%, but this would result in a much smaller training set. Therefore, in an attempt to further deal with class imbalance we employed the weighted focal loss function (Lin et al., 2017) in the training of the DLCD model.

The weighted focal loss function was proposed for object recognition problems with extreme foreground-background class

imbalance. It is described in Equation 1, where $y \in \{\pm 1\}$ specifies the ground-truth class; $p \in [0, 1]$ is the model's estimated probability for class $y = 1$ (the deforestation class); and α represents the weight associated to the deforestation class.

$$WFL(p_t) = -\alpha_t(1 - p_t)^{\gamma} log(p_t) \quad (1)$$

$$\text{where} \quad p_t = \begin{cases} p & \text{if } y = 1 \\ 1 - p & \text{otherwise} \end{cases}$$

$$\text{and} \quad \alpha_t = \begin{cases} \alpha & \text{if } y = 1 \\ 1 - \alpha & \text{otherwise} \end{cases}$$

In order to investigate the influence of the parameters α and γ of the weighted focal loss function in the classification accuracy of the DLCD method, we performed a grid search. Accordingly, in the experiments we considered the following values of α: 0.9, 0.8, 0.7, 0.6 and 0.5. Additionally, for each different α value, we varied the γ values from 0 to 5 (integer values), which amounts to a total of 30 combinations of α and γ values.

6. RESULTS

Figures 5 and 6 summarize the results of our experiments in terms of F1-score for the deforestation class and Overall Accuracy achieved by the Early Fusion (EF) and Siamese Networks (S-CNN) methods used in (Ortega et al., 2019) and by the DeepLabv3+ change detection (DLCD) model implemented in this work. The figures show the performance obtained by each method for different number of tiles used for training. For the sake of clarity, we show the results of only four combinations of the weights/γ values of the weighted focal loss function, namely: DLCD-1, for weights 0.1/0.9 and γ equal to 0; DLCD-2, for weights 0.5/0.5 and γ equal to 0; DLCD-3, for weights 0.5/0.5 and γ equal to 1; DLCD-4, for weights 0.5/0.5 and γ equal to 2. Those combinations were the ones that produced the best results, considering all the 30 possible variations.

The results show that the DeepLabv3+ based models (DLCD) significantly outperformed all the other methods in terms of overall accuracy and F1-score, in all the training scenarios (i.e., number of training tiles).

In the best result for the F1-score using 4 tiles for training, 71.8%, was obtained with the DLCD model, using a weight of 0.1 for the no-deforestation class and a weight of 0.9 for the deforestation class, and a γ equal to 0 (DLCD-1). In this scenario, the DLCD-1 variant outperformed both Early Fusion and Siamese Networks by 8.6% and 8.9%, respectively.

When using one, two and three tiles for training, the DLCD-1 variant outperformed both Early Fusion and Siamese Networks methods by, respectively, 20.5% and 16.3% for one training tile; 14.4% and 13.1% for two training tiles; and 9.3% and 8.6% for three training tiles.

In terms of Overall Accuracy, the best result, when using 4 tiles for training, was obtained using weight of 0.5 for both deforestation and no-deforestation classes, with γ equal to 1 (DLCD-3), with a score of 98.8%. In this scenario, the DLCD-3 variant outperformed the Early Fusion and Siamese Networks methods by 0.97% and 0.77%, respectively. The same weight configuration but with a γ of 2 (DLCD-4) obtained a similar score of 98.8%. When using one, two and three tiles for training, the DLCD-4 variant outperformed both Early Fusion and Siamese Networks methods by, respectively, 3.66% and 2.83% for one training tile, 1.81% and 1.64% for two training tiles and 1.46% and 1.05% for three training tiles.

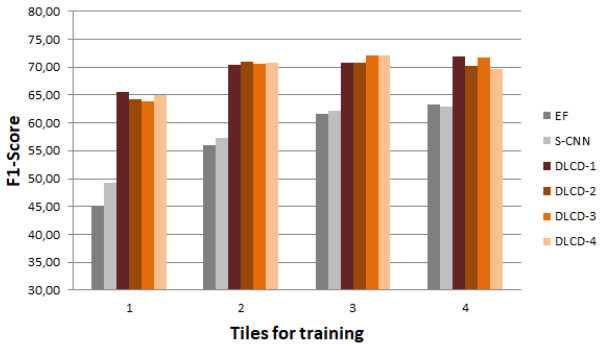


Figure 5. F1-score×number of tiles for training.

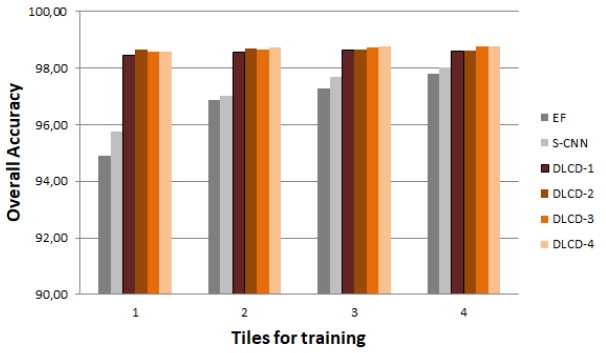


Figure 6. Overall accuracy×number of tiles for training.

Figures 8 and 10 show visual examples of the outcomes of the Early Fusion, Siamese Networks and DeepLabv3+ based methods (DLCD-1 variant), over the test tiles 6 and 14.

As can be seen in figures the method implemented in this work produced a notably lower number of false deforestation areas, which is particularly important for operational reasons, considering the effort and costs involved in the reconnaissance of the actual deforestation by the local authorities, involved in penalizing the perpetrators or in mitigating the effects of illegal deforestation.

7. CONCLUSION

In this work we evaluated three deep learning-based methods employed for the task of deforestation detection in the Amazon rainforest. We compared the performances of two previously presented methods: Early Fusion (EF), Siamese Convolutional Neural Network (S-CNN), with the one of a method based on the DeepLabv3+ model, implemented in this work.

We evaluated the methods over Landsat OLI-8 images, acquired in 2016 and 2017 over the same region of the Brazilian Legal Amazon, and used deforestation polygons produced by the PRODES deforestation monitoring project, of the Brazilian National Space Research Institute (INPE), as references.

We also evaluated the methods with varying sizes of training sample sets, and, in the case of the proposed DeepLabv3+ method, we tested various combinations of parameters of the focal loss function, used in the training procedure. The results showed that all variants of the proposed method significantly outperformed the EF and S-CNN methods in terms of F1-score, and also yielded better Overall Accuracy results.

The gains in performance were even more significant when limited amounts of samples were used in training the deep learning models, which indicates that the proposed method has a better generalization capacity than the evaluated counterparts. While no important gains were noted by varying the parameters of the selected loss function, this seems to indicate that proposed method deals satisfactory with high class imbalance.

A natural path for further investigation is to evaluate the proposed method using data from other sensors, especially from SAR systems, since cloud coverage is a critical problem for forest monitoring in tropical regions.

Additionally, in this work we evaluated the proposed method considering a specific site in the Brazilian Legal Amazon. Further studies should be carried out to investigate the transferability potential of the method, e.g., evaluating its performance when training with samples from a particular site, and testing on images covering different sites in the Amazon.

ACKNOWLEDGEMENTS

This work is supported by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), and FAPERJ (Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro).

REFERENCES

Assunção, J., Rocha, R., 2019. Getting greener by going black: the effect of blacklisting municipalities on Amazon deforestation. *Environment and Development Economics*, 24, 115-137.

Carlson, T. N., Ripley, D. A., 1997. On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote Sensing of Environment*, 62(3), 241 - 252. http://www.sciencedirect.com/science/article/pii/S0034425797001041.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A. L., 2014. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs. *CoRR*, abs/1412.7062.

Chen, L., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A. L., 2018a. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4), 834-848.

Chen, L., Papandreou, G., Schroff, F., Adam, H., 2017. Rethinking Atrous Convolution for Semantic Image Segmentation. *CoRR*, abs/1706.05587. http://arxiv.org/abs/1706.05587.

Chen, L., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018b. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. *CoRR*, abs/1802.02611. http://arxiv.org/abs/1802.02611.

Chollet, F., 2016. Xception: Deep Learning with Depthwise Separable Convolutions. *CoRR*, abs/1610.02357. http://arxiv.org/abs/1610.02357.

Chu, Y., Cao, G., Hayat, H., 2016. Change detection of remote sensing image based on deep neural networks. *2016 2nd International Conference on Artificial Intelligence and Industrial Engineering (AIIE 2016)*, Atlantis Press, 262–267.

Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., Wei, Y., 2017. Deformable Convolutional Networks. *2017 IEEE International Conference on Computer Vision (ICCV)*, 764-773.

Daudt, R. C., Le Saux, B., Boulch, A., Gousseau, Y., 2018. Urban change detection for multispectral earth observation using convolutional neural networks. *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, 2115–2118.

de Jong, K. L., Bosman, A. S., 2019. Unsupervised change detection in satellite images using convolutional neural networks. *2019 International Joint Conference on Neural Networks (IJCNN)*, 1–8.

De Sy, V., Herold, M., Achard, F., Beuchle, R., Clevers, J., Lindquist, E., Verchot, L., 2015. Land use patterns and related carbon losses following deforestation in south america. *Environmental Research Letters*, 10(12).

Diniz, C. G., d. A. Souza, A. A., Santos, D. C., Dias, M. C., d. Luz, N. C., d. Moraes, D. R. V., Maia, J. S., Gomes, A. R., d. S. Narvaes, I., Valeriano, D. M., Maurano, L. E. P., Adami, M., 2015. DETER-B: The New Amazon Near Real-Time Deforestation Detection System. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(7), 3619-3628.

Goodman, R., Aramburu, M., Gopalakrishna, T., Putz, F., Gutiérrez, N., Alvarez, J., Aguilar-Amuchastegui, N., Ellis, P., 2019. Carbon emissions and potential emissions reductions from low-intensity selective logging in southwestern Amazonia. *Forest Ecology and Management*, 439, 18-27.

Guo, H., Viktor, H. L., 2004. Learning from Imbalanced Data Sets with Boosting and Data Generation: The DataBoost-IM Approach. *SIGKDD Explor. Newsl.*, 6(1), 30–39. https://doi.org/10.1145/1007730.1007736.

Guo, X., Chen, Y., Liu, X., Zhao, Y., 2020. Extraction of snow cover from high-resolution remote sensing imagery using deep learning on a small dataset. *Remote Sensing Letters*, 11(1), 66-75. https://doi.org/10.1080/2150704X.2019.1686548.

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*, abs/1704.04861.

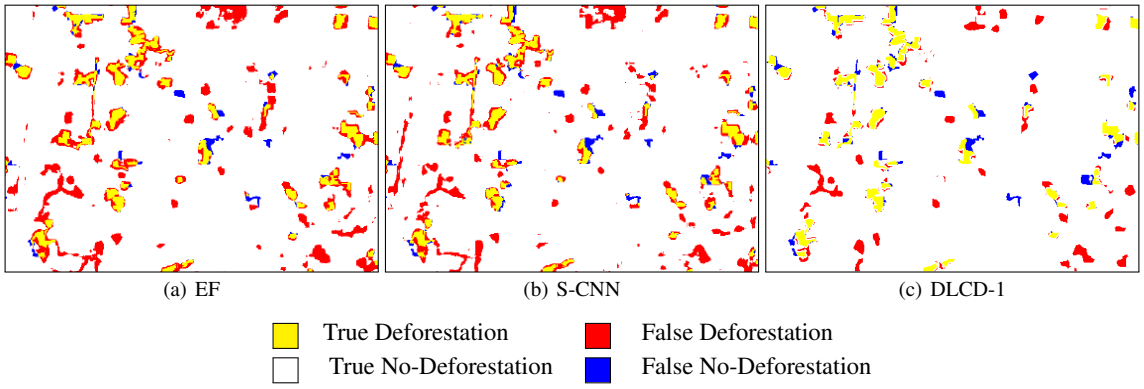
Ioffe, S., Szegedy, C., 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *ArXiv*, abs/1502.03167.

ISPRS, 2020. ISPRS 2d semantic labeling challenge. http://www2.isprs.org/commissions/comm3/wg4/semantic-labeling.html. Accessed: 2020-02-03.

Ji, S., Wei, S., Lu, M., 2019. A scale robust convolutional neural network for automatic building extraction from aerial and satellite imagery. *International Journal of Remote Sensing*, 40(9), 3308-3322.

(a) T1 - 2016 (b) T2 - 2017

Figure 7. Test tile number 6.

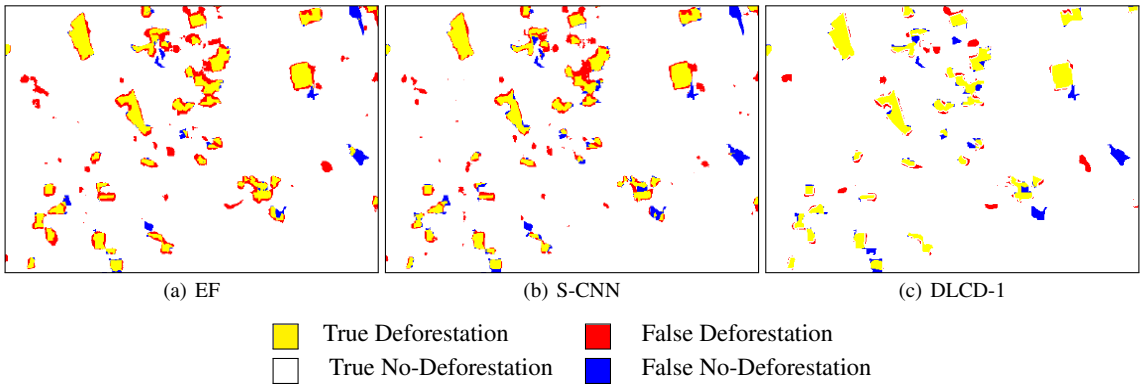


(a) EF (b) S-CNN (c) DLCD-1

Figure 8. Change maps predicted by EF, S-CNN and DLCD-1 on test tile number 6.

(a) T1 - 2016 (b) T2 - 2017

Figure 9. Test tile number 14.



(a) EF (b) S-CNN (c) DLCD-1

Figure 10. Change maps predicted by EF, S-CNN and DLCD-1 on test tile number 14.

Krähenbühl, P., Koltun, V., 2011. Efficient inference in fully connected CRFs with Gaussian edge potentials. *Advances in Neural Information Processing Systems*, 24, 109-117.

Lin, T., Goyal, P., Girshick, R. B., He, K., Dollár, P., 2017. Focal Loss for Dense Object Detection. *CoRR*, abs/1708.02002. http://arxiv.org/abs/1708.02002.

Liu, W., Rabinovich, A., Berg, A. C., 2015. ParseNet: Looking Wider to See Better. *ArXiv*, abs/1506.04579.

Lovejoy, T. E., Nobre, C., 2018. Amazon Tipping Point. *Science Advances*, 4(2).

Malingreau, J., Eva, H., de Miranda, E., 2012. Brazilian Amazon: a significant five year drop in deforestation rates but figures are on the rise again. *Ambio*, 41(3), 309-314.

Mallat, S., 2018. *A Wavelet Tour of Signal Processing*. 3 edn, Academic Press, Cambridge, MA, 832.

Mnih, V., 2013. Machine Learning for Aerial Image Labeling. PhD thesis, University of Toronto.

Nogueron, R., Barreto, P., Souza Jr, C., Anderson, A., Salomão, R., 2006. *Human pressure on the Brazilian Amazon forests*. World Resources Institute, Washington, DC.

Ortega, M., Castro, J., Nigri Happ, P., Gomes, A., Feitosa, R., 2019. Evaluation of Deep Learning Techniques for Deforestation Detection in the Amazon Forest. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2/W7, 121-128.

Peng, Y., Sun, S., Pan, Y., Li, R., 2019. Robust Semantic Segmentation By Dense Fusion Network On Blurred VHR Remote Sensing Images. *arXiv e-prints*, arXiv:1903.02702.

Sathler, D., Adamo, S., Lima, E., 2018. Deforestation and local sustainable development in Brazilian Legal Amazonia: an exploratory analysis. *Ecology and Society*, 23(2).

Shimabukuro, Y., dos Santos, J., Formaggio, A., Duarte, V., Rudorff, B., 2013. The brazilian amazon monitoring program: Prodes and deter projects. F. Archard, M. Hansen (eds), *Global Forest Monitoring from Earth Observation*, CRC Press, Boca Raton, chapter 9, 167–184.

Shimabukuro, Y., Duarte, V., Anderson, L., Valeriano, D., Arai, E., Freitas, R., Rudorff, B. F., Moreira, M., 2006. Near real time detection of deforestation in the Brazilian Amazon using MODIS imagery. *Ambiente e Agua - An Interdisciplinary Journal of Applied Science*, 1(1), 37-47.

The Worldwatch Institute, 2015. *Vital Signs Volume 22 - The Trends That Are Shaping Our Future*. Island Press.

Valeriano, D. M., Mello, E. M. K., Moreira, J. C., Shimabukuro, Y. E., Duarte, V., Souza, I. M., dos Santos, J. R., Barbosa, C. C. F., de Souza, R. C. M., 2004. Monitoring tropical forest from space: the PRODES digital project. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXV(B7), 272–274.

Wang, J., Shen, L., Qiao, W., Dai, Y., Li, Z., 2019. Deep Feature Fusion with Integration of Residual Connection and Attention Model for Classification of VHR Remote Sensing Images. *Remote Sensing*, 11(13). https://www.mdpi.com/2072-4292/11/13/1617.

World Wildlife Fund, 2020a. Places: Amazon. `https://www.worldwildlife.org/places/amazon`. Accessed: 2020-02-20.

World Wildlife Fund, 2020b. Amazon deforestation. `https://wwf.panda.org/our_work/forests/deforestation_fronts2/deforestation_in_the_amazon`. Accessed: 2020-02-20.

Yao, X., yang, h., Wu, Y., Penghai, W., Wang, B., Zhou, X., Wang, S., 2019. Land Use Classification of the Deep Convolutional Neural Network Method Reducing the Loss of Spatial Features. *Sensors*, 19, 2792.

Zhang, C., Wei, S., Ji, S., Lu, M., 2019. Detecting Large-Scale Urban Land Cover Changes from Very High Resolution Remote Sensing Images Using CNN-Based Classification. *ISPRS International Journal of Geo-Information*, 8(4), 189. http://dx.doi.org/10.3390/ijgi8040189.

Zhang, Z., Vosselman, G., Gerke, M., Tuia, D., Yang, M. Y., 2018. Change detection between multimodal remote sensing data using siamese cnn.