Application of convolutional neural network to pixel-wise classification in deforestation detection using PRODES data

Felipe Ferraz Martins[1], Matheus Cavassan Zaglia[2]

[1]Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio)
Rua Marquês de São Vicente, 225 – 22451-900 – Rio de Janeiro – RJ – Brasil

[2]Instituto Nacional de Pesquisas Espaciais (INPE)
Av. dos Astronautas, 1758 – 12227-010 – São José dos Campos – SP – Brasil

fferraz@ele.puc-rio.br, matheus.zaglia@inpe.com

Abstract. *The INPE PRODES project, which since the 1980s maps and quantifies deforestation in the Brazilian Legal Amazon, can be considered the main systematic monitoring project for tropical forests in the world. Over the time, the project has gone through several stages, and today its methodology is the visual interpretation of images by remote sensing experts. This paper aims to evaluate the use of neural networks to automate this process, improving accuracy and minimizing the time required for interpretation. Results will be compared to official PRODES data.*

1. Introduction

Deforestation in the Brazilian Amazon rainforest gained momentum in the 1970s and 1980s due to tax incentives and subsidized credit to large ranchers, having a disastrous impact on local biodiversity and climate [Fearnside 2005]. Since then, deforestation rates have varied according to the economy and government policies [Fearnside 1987]. In 1978, the National Institute for Space Research (INPE) conducted a survey using digital imagery of the LANDSAT satellite to measure the amount of deforested areas of the legal Amazon [Tardin et al. 1979]. In a total of 552000km^2 of analyzed area, approximately 7.4% - an area of 41000km^2 - were interpreted as deforested area. Recent estimates indicate 19.3% of total deforested area[1]. It is important to highlight that those rates are only related to native forest, and do not include reforested areas.

The deforestation and burning of the legal Amazon has been in the worldwide news recently, and the environmental policies of the Brazilian government have been questioned. In this context, research for assisting deforestation detection is becoming increasingly relevant. New approaches can promote a better understanding of the drivers of deforestation, thus enabling the generation of future scenarios, prediction of risk sites, and the foundation of public policies for deforestation prevention and response [Lambin et al. 1994].

Since 1988, INPE has been producing annual reports on the deforestation rate of the Amazon rainforest through the PRODES[2] project. Over the years, the methodology used by PRODES to identify deforestation has changed, and today is based on a visual

[1]https://rainforests.mongabay.com/amazon/deforestation_calculations.html
[2]http://www.obt.inpe.br/OBT/assuntos/programa/amazonia/prodes

interpretation of digital images by remote sensing experts. The current methodology has shown great results [Kintisch 2007]. However, a huge amount of time is spent due to the manual process of interpreting satellite images in order to detect deforestation. To advance the knowledge on methodologies for automatic deforestation detection, computer vision, machine learning and neural networks are current topics.

This work presents an evaluation of the use of a deep learning technique applied to Amazon deforestation detection in satellite images. As noted by [Shoham et al. 2017], artificial intelligence techniques are rapidly gaining more attention in the computer science community, where "machine learning" and "deep learning" are becoming more common. The number of articles published with the keyword "Artificial Intelligence" since 1996 has increased more than ninefold, and errors in image labeling have gone from 28.5% to less than 2.5% since 2010.

2. Literature Review

In this section, work related to the use of machine learning techniques in the field of computer vision will be discussed.

As noted by [Ronneberger et al. 2015], convolutional networks have improved the state of the art in visual recognition tasks. Their limitation is linked to the large amount of data required for training, but this problem has become less relevant due to the higher availability and data processing capacity today.

Studies using recent AI techniques such as [Iglovikov et al. 2017] and [Iglovikov and Shvets 2018] present great results using pixel-wise classification models through fully convolutional neural networks using satellite images. The first one applies those models to the Dstl Satellite Imagery Feature Detection competition database proposed by Kaggle[3], featuring city images with ten labels such as large vehicle, small vehicle and building, while the second, to Aerial Image Labeling Dataset[4], which features city images with labels of building and not-building.

3. Model

3.1. U-Net

U-Net is a semantic segmentation fully convolutional neural network model proposed by [Ronneberger et al. 2015], modified and extended to work with fewer images during training in order to generate precise segmentation. Fully convolutional networks have an encoder-decoder architecture, in which features are learned by the encoder through convolution layers, while the decoder converts these features into a pixel-level classification through up-sampling layers. The model used stands out for using intermediate outputs of the encoder concatenated to the decoder. This reduces the negative effects of dimensionality reduction performed by max-pooling functions during the encoder, improving segmentation. The original architecture is presented in Figure 1.

In order to improve the result and shorten the time required for training convolutional layer models, the use of Residual Blocks is state of the art [He et al. 2016]. Such blocks are composed of a chain of convolutional layers, where the output of this chain is

[3]https://www.kaggle.com/c/dstl-satellite-imagery-feature-detection
[4]http://project.inria.fr/aerialimagelabeling/

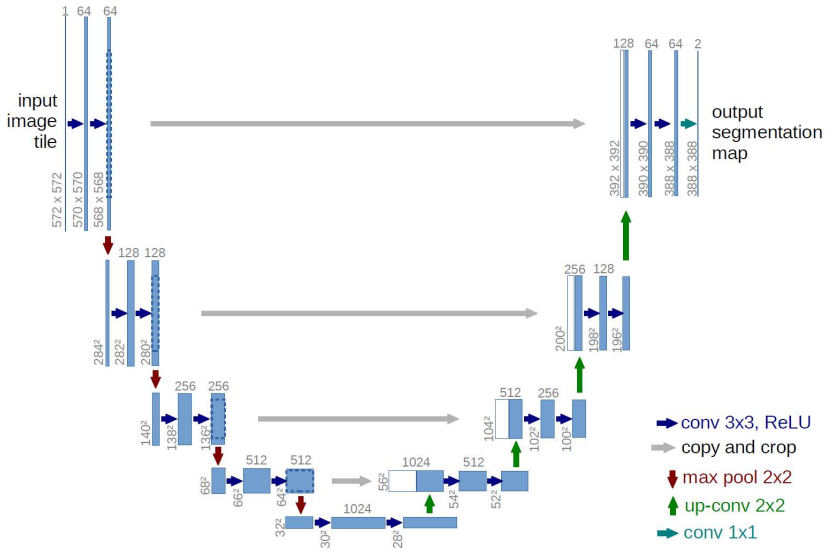

Figure 1. U-Net architecture [Ronneberger et al. 2015]

added to its input. The advantage of this structure is that it creates a free path to the gradient, which, in a structure without the residual block, can become derisory after subsequent derivatives and activation functions in the backpropagation step. A visual representation of this block is shown in Figure 2.

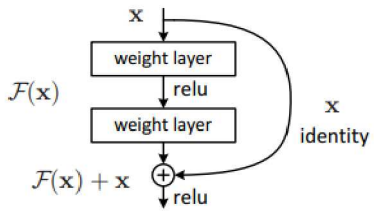

Figure 2. Residual block [He et al. 2016]

In the present work, it was chosen to fine-tune the U-Net model architecture with Residual Blocks to get a better accuracy for the proposed task changing the network's hyperparameters, but not creating a deeper network than the original U-Net [Ronneberger et al. 2015]. Using ADAM optimizer [Kingma and Ba 2014] with a decaying learning rate of 0.001, our best network with these assumptions was as follows (Table 1). ReLU activation function was used in every convolution layer, and Softmax, on the output layer.

3.2. Evaluation Metrics

To evaluate the trained neural network model, several metrics can be used. They are also important at the training phase: if the metric is differentiable, it can be used as a loss function. In the present work, three metrics were monitored. Dice Coefficient was used for training the model.

Table 1. Network Architecture. "+" refers to a Sum layer, "||" refers to a Concatenate layer

Layer	Size	Layer	Size
Input	256x256x3	Dropout	32x32x512
Convolution 1	256x256x32	Convolution 15	32x32x256
Convolution 2	256x256x32	Convolution 16	32x32x256
MaxPooling	128x128x32	Convolution 17	32x32x256
Dropout	128x128x32	Conv 15 + Conv 17	32x32x256
Convolution 3	128x128x64	Deconvolution 2	64x64x128
Convolution 4	128x128x64	Deconv 2 || Conv 6 + Conv 8	64x64x256
Convolution 5	128x128x64	Dropout	64x64x256
Conv 3 + Conv 5	128x128x64	Convolution 18	64x64x128
MaxPooling	64x64x64	Convolution 19	64x64x128
Dropout	64x64x64	Convolution 20	64x64x128
Convolution 6	64x64x128	Conv 18 + Conv 20	64x64x128
Convolution 7	64x64x128	Deconvolution 3	128x128x64
Convolution 8	64x64x128	Deconv 3 || Conv 3 + Conv 5	128x128x128
Conv 6 + Conv 8	64x64x128	Dropout	128x128x128
MaxPooling	32x32x128	Convolution 21	128x128x64
Dropout	32x32x128	Convolution 22	128x128x64
Convolution 9	32x32x256	Convolution 23	128x128x64
Convolution 10	32x32x256	Conv 21 + Conv 23	128x128x64
Convolution 11	32x32x256	Deconvolution 4	256x256x32
Conv 9 + Conv 11	32x32x256	Deconv 4 || Convolution 2	256x256x64
MaxPooling	16x16x256	Dropout	256x256x64
Dropout	16x16x256	Convolution 24	256x256x32
Convolution 12	16x16x512	Convolution 25	256x256x32
Convolution 13	16x16x512	Convolution 26	256x256x32
Convolution 14	16x16x512	Conv 24 + Conv 26	256x256x32
Conv 12 + Conv 14	16x16x512	Convolution 27	256x256x3
Deconvolution 1	32x32x256	Output	256x256x3
Deconv 1 || Conv 9 + Conv 11	32x32x512		

- Cross-Entropy: Entropy is a measure of uncertainty associated with a $q(y)$ distribution. Such uncertainty is given by Equation 1.

$$H(X) = -\sum_{i=1}^{n} P(x_i) log_b P(x_i) \tag{1}$$

For the context of a classifier, one work with two different distributions: the labeled distribution and the one predicted by the classifier. Thus, the loss function for Binary Cross Entropy is given by Equation 2, while for Categorical Cross Entropy, Equation 3.

$$BCE(y, \hat{y}) = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot log(\hat{y}_i) + (1 - y_i) \cdot log(1 - \hat{y}_i) \tag{2}$$

$$CCE(y, \hat{y}) = -\sum_{j=0}^{M}\sum_{i=0}^{N} y_{ij} \cdot log(\hat{y}_{ij}) \tag{3}$$

where N represents the number of examples, M represents the number of classes, y_i represents the labeled data, and $\hat{y}_i$, the output of the classifier.

- Dice Coefficient: Initially proposed by [Dice 1945], it is used in computer vision to represent the similarity between two images. This similarity is quantified as:

$$DSC(X,Y) = \frac{2\,|X \cap Y|}{|X| + |Y|}. \tag{4}$$

Thus, the loss function used in the training step can be written as

$$J(y,\hat{y}) = -\sum_{i=1}^{n} \frac{y_i \cdot \hat{y}_i}{y_i + \hat{y}_i}, \tag{5}$$

where y_i represents the ground truth and $\hat{y}_i$, the network output.

- Jaccard Index ou Intersection over Union (IoU): a term coined by Paul Jaccard, it is a statistic used for gauging the similarity and diversity of sample sets. It is defined as the size of the intersection divided by the size of the union of two sets, as presented in Equation 6.

$$IoU(X,Y) = \frac{|X \cap Y|}{|X \cup Y|} = \frac{|X \cap Y|}{|X| + |Y| - |X \cap Y|}. \tag{6}$$

4. Dataset

For the development of the dataset for the present work, a set of LANDSAT 8 scenes referring to the Xingu Indigenous Park were chosen, all referring to the year 2018. From those scenes, only three spectral bands were included in the dataset: red, near-infrared and short-wavelength infrared. The path row of the scenes used can be seen in Table 2. The labels were obtained in the PRODES database[5].

Table 2. Scenes in the dataset

Path	Row
224	68
225	67, 68

The scenes and masks were cut into small 256x256-pixel images (Figure 3), resulting in a total of 1256 clippings. These were divided into 1017 for training, 113 for validation and 126 for testing. Preprocessing was done to mask all non-forest and non-deforestation areas by prior knowledge, giving it a "Background" label. Images with high presence of this label were discarded for the training phase.

It is important to note that this dataset is expanded annually, and the area classified as non-forest in one year is maintained in the next year's classification, even if this area has been reforested.

[5]http://www.obt.inpe.br/OBT/assuntos/programas/amazonia/prodes

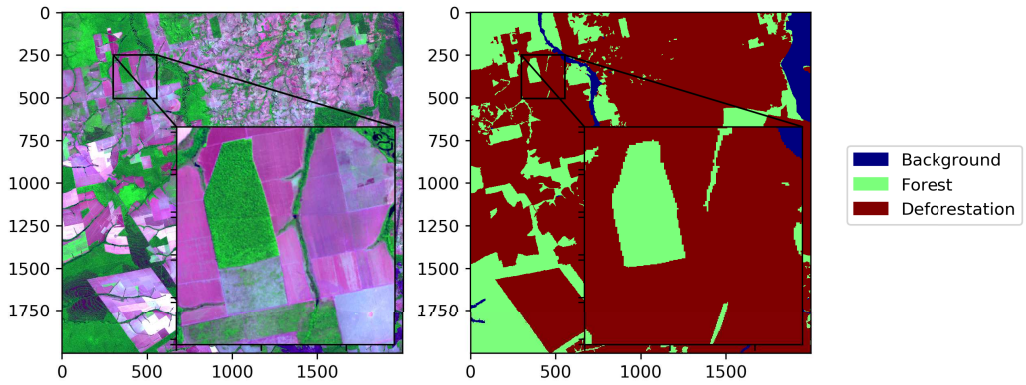

Figure 3. 256x256-pixel image and label

5. Results

The results presented in this section were achieved after an average of 40 training epochs.

The Dice coefficient of the best trained model, in addition to the cross entropy and Jaccard Index values, can be found in Table 3. Figure 4 presents the evolution of those metrics during the training phase.

Table 3. Metrics in training and validation sets

Metric	Training set	Validation set
Dice Coefficient	0.9696	0.9650
Jaccard Index	0.9014	0.8910
Binary Cross-Entropy	0.0993	0.1238
Categorical Cross-Entropy	0.1505	0.1871

In addition to these metrics, Precision, Recall, and F1-score related to each class in the test set were calculated, as well as the overall accuracy. The values are presented in Table 4.

Table 4. Metrics in the Test Set. Global accuracy: 0.95171.

Class	Precision	Recall	F1-Score	Support
Forest	0.93900	0.97459	0.95646	4,490,972
Deforestation	0.94757	0.87877	0.91187	2,346,132

Figure 5 compares model prediction results with the ground truth. It is noticed that the neural network managed to generalize well for different situations.

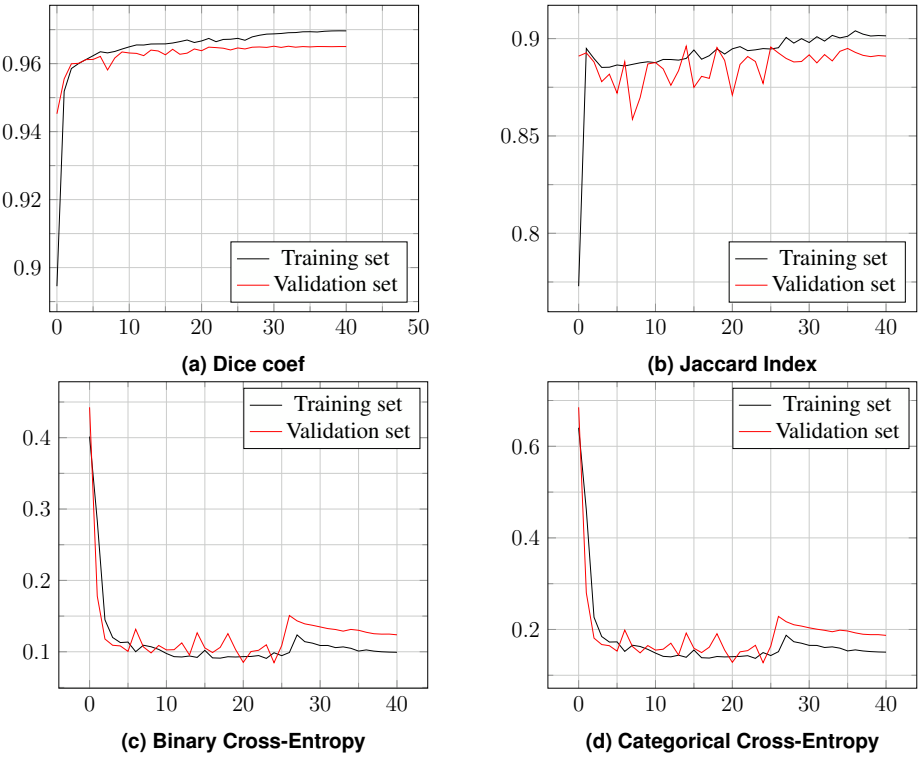

Figure 4. Metrics during training

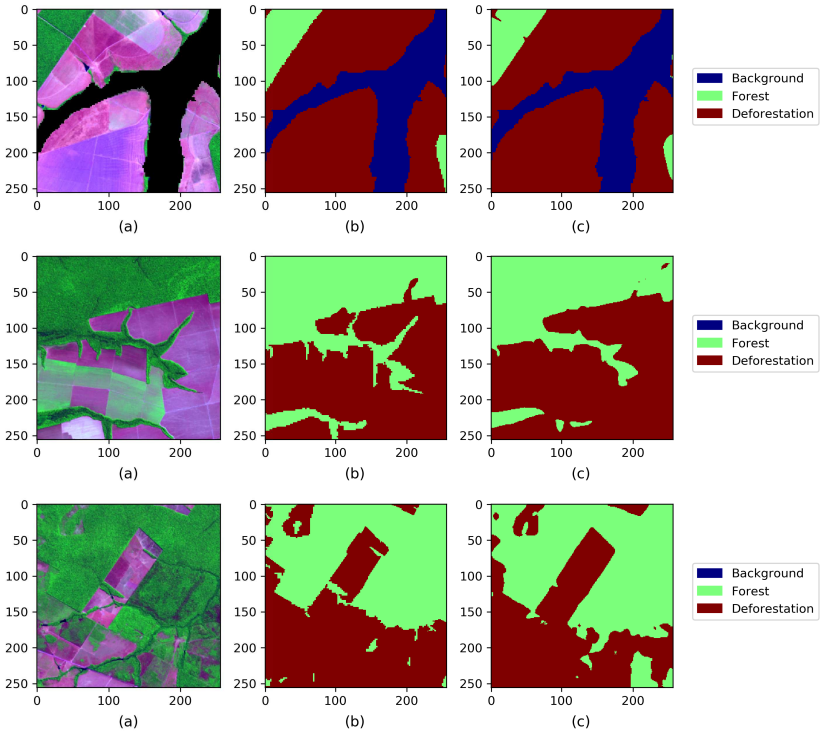

Figure 5. Comparison between model prediction (c) and ground truth (b), related to image (a).

6. Conclusion

The model used in this work obtained a good accuracy rate when performing pixel-wise classification of satellite images from LANDSAT-8, using labels from PRODES data for training. PRODES global accuracy of the mapping of deforestation for the state of Mato Grosso for the year 2014 was 94.5% $\pm$ 2.05 [Adami et al. 2017]. Our model got a 95.2% accuracy rate. Despite a good accuracy, the model does not distinguish native forest from reforestated areas. Most of the prediction errors of the network are from reforested areas. As the model used has no information about previous years, it is difficult to distinguish areas with such characteristics. In addition, other bands of the multispectral image may retain this information.

7. Future Work

To improve what was discussed in the present work, we will evaluate the use of multispectral images and compare with the results presented here. In addition, minor changes to the loss function recently proposed, such as the boundary error presented in [Bokhovkin and Burnaev 2019], may improve results. One can also use recurrent models so that the system has memory from previous years for the current year prediction, as in [Jia et al. 2017].

References

Adami, M., Gomes, A. R., Beluzzo, A., et al. (2017). A confiabilidade do prodes: estimativa da acurácia do mapeamento do desmatamento no estado mato grosso. In *Embrapa Amazônia Oriental-Artigo em anais de congresso (ALICE)*. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 18., 2017, Santos. Anais ....

Bokhovkin, A. and Burnaev, E. (2019). Boundary loss for remote sensing imagery semantic segmentation. *CoRR*, abs/1905.07852.

Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302.

Fearnside, P. M. (1987). Causes of deforestation in the Brazilian Amazon. *The geophysiology of Amazonia: vegetation and climate interactions*, pages 37–61.

Fearnside, P. M. (2005). Deforestation in brazilian amazonia: history, rates, and consequences. *Conservation biology*, 19(3):680–688.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Iglovikov, V., Mushinskiy, S., and Osin, V. (2017). Satellite imagery feature detection using deep convolutional neural network: A kaggle competition. *arXiv preprint arXiv:1706.06169*.

Iglovikov, V. and Shvets, A. (2018). Ternausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation. *arXiv preprint arXiv:1801.05746*.

Jia, X., Khandelwal, A., Nayak, G., et al. (2017). Incremental dual-memory lstm in land cover prediction. In *Proceedings of the 23rd ACM SIGKDD International Conference*

on Knowledge Discovery and Data Mining, KDD '17, pages 867–876, New York, NY, USA. ACM.

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kintisch, E. (2007). Improved monitoring of rainforests helps pierce haze of deforestation. *Science*, 316(5824):536–537.

Lambin, E., for Remote Sensing Applications, I., and Office, E. S. A. E. E. P. (1994). *Modelling Deforestation Processes: A Review*. EUR / European Commission. Office for Official Publications of the European Community.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.

Shoham, Y., Perrault, R., Brynjolfsson, E., et al. (2017). Artificial intelligence index 2017 annual report.

Tardin, A. T., Santos, A. P. d., and Lee, D. C. L. (1979). Levantamento de áreas de desmantamento na amazonia legal atraves de imagens do satélite landsat. In *Levantamento de áreas de desmantamento na Amazonia Legal atraves de imagens do satélite LANDSAT*. Inpe.