SEMANTIC SEGMENTATION OF BRAZILIAN SAVANNA VEGETATION USING HIGH SPATIAL RESOLUTION SATELLITE DATA AND U-NET

A. K. Neves [1,2,*], T. S. Körting [1], L. M. G. Fonseca [1], C. D. Girolamo Neto [1], D. Wittich [2], G. A. O. P. Costa[3], C. Heipke[2]

[1] Image Processing Division, National Institute for Space Research (INPE), Brazil
(alana.neves, thales.korting, leila.fonseca, cesare.neto)@inpe.br
[2] Institute of Photogrammetry and GeoInformation, Leibniz Universität Hannover, Germany
(wittich, heipke)@ipi.uni-hannover.de
[3] Institute of Mathematics and Statistics (IME), Rio de Janeiro State University (UERJ), Brazil
gilson.costa@ime.uerj.br

Commission III, WG III/10

KEY WORDS: Cerrado, biome, physiognomies, pixel-wise classification, Remote Sensing, Deep Learning.

ABSTRACT:

Large-scale mapping of the Brazilian Savanna (Cerrado) vegetation using remote sensing images is still a challenge due to the high spatial variability and spectral similarity of the different characteristic vegetation types (physiognomies). In this paper, we report on semantic segmentation of the three major groups of physiognomies in the Cerrado biome (Grasslands, Savannas and Forests) using a fully convolutional neural network approach. The study area, which covers a Brazilian conservation unit, was divided into three regions to enable testing the approach in regions that were not used in the training phase. A WorldView-2 image was used in cross validation experiments, in which the average overall accuracy achieved with the pixel-wise classifications was 87.0%. The F-1 score values obtained with the approach for the classes Grassland, Savanna and Forest were of 0.81, 0.90 and 0.88, respectively. Visual assessment of the semantic segmentation outcomes was also performed and confirmed the quality of the results. It was observed that the confusion among classes occurs mainly in transition areas, where there are adjacent physiognomies if a scale of increasing density is considered, which agrees with previous studies on natural vegetation mapping for the Cerrado biome.

1. INTRODUCTION

The Brazilian Savanna, also known as Cerrado, is the second largest Brazilian biome, covering an area of approximately two million km², which amounts to 24% of the Brazilian territory. The water resources of this biome feed the three largest watersheds in South America: the Amazon, Prata and São Francisco watersheds. Additionally, the Cerrado biome is considered one of the 35 global hotspots for biodiversity conservation (Mittermeier et al., 2011), with a flora containing more than 12,000 species, of which 40% are endemic.

Despite its ecological importance, only 8.6% of the Cerrado natural vegetation belongs to Conservation Units, i.e., specific regions established to protect biodiversity, water bodies and other environmental resources (MMA, 2010). Approximately 47% of the natural vegetation has already been converted to other land use classes, especially pasture (29%) and annual agriculture (9%) (MMA, 2015). Moreover, in the last years, deforestation rates in the Cerrado have been higher than what was observed in the Amazon biome (INPE, 2019). Therefore, accurate mapping of Cerrado vegetation is essential for assessing biodiversity, improving Carbon stock estimation within the biome and guiding conservation policies.

Large-scale mapping of the Cerrado vegetation using Remote Sensing (RS) images is still a challenge, due to the high spatial variability and spectral similarity among its vegetation types (physiognomies). According to the classification system proposed by Ribeiro and Walter (2008), there are 25 physiognomies, which vary in structure, density, biomass and carbon storage. They can be grouped into three major groups of physiognomies, namely: Grasslands, Savannas and Forests.

A large variety of techniques have been employed in vegetation mapping. In recent years, Deep Learning methods based on Convolutional Neural Networks (CNNs) have thrived in the RS field (e.g. Zhu et al., 2017). CNNs are able to perform end-to-end classification, learning from an input dataset features with increasing complexity with the number of layers of the network (LeCun et al., 2015). The results achieved with such methods outperform those obtained with traditional Machine Learning algorithms, such as Random Forest and Support Vector Machine (Kussul et al., 2017; Guirado et al., 2017).

In this paper, we used the U-net CNN architecture (Ronneberger et al., 2015) to perform semantic segmentation (also known as pixel-wise classification) of a high spatial resolution satellite image covering a conservation unit in the Brazilian Cerrado, into the three major groups of physiognomies (Grasslands, Savannas and Forests) according to the Ribeiro and Walter (2008) classification system. We conducted various experiments, training and testing the network in different regions.

To the best of our knowledge, this is the first paper applying Deep Learning techniques for the semantic segmentation of natural vegetation in the Brazilian Savanna.

---

* Corresponding author

1.1 Related work

A few projects have been devoted to mapping the Cerrado physiognomies. The TerraClass Cerrado project employs Landsat images and traditional methods, such as region growing image segmentation followed by visual interpretation, to map land use and land cover in the entire Cerrado biome. Due to the difficulties in class differentiation, natural vegetation was grouped into two classes only: Forest and Non-Forest, the later including Savanna and Grassland. Despite the project's overall accuracy of 80.2%, the accuracies for Forest and Non-Forest classes were only between 60% and 65% (MMA, 2015).

Mapping based mainly on visual interpretation requires a lot of time and can be subjective. The MapBiomas project performs an annual automatic pixel-wise classification of the Cerrado biome using a Random Forest approach and Landsat images. The project runs since 2015, but it produces maps from 1984 onwards. Since the methodology of MapBiomas is constantly being improved, changes in the method generate new collections and the maps of all years are updated. Performing an analysis using data from one collection does not guarantee that the results will be the same when using data from a different collection (MapBiomas, 2020).

An important aspect related to the Cerrado physiognomies is their seasonality. In order to represent the seasonality in the classification, Borges and Sano (2014) and Abade et al. (2015) used time series of vegetation indices derived from MODIS images, and performed the physiognomy classification with Support Vector Machine and Multilayer Perceptron (Abade et al., 2015), and Spectral Angle Mapper (Borges and Sano, 2014). While the revisit time of MODIS is high, the spatial resolution of only 250 meters results in a mixture of physiognomies within single pixels, thus making proper detailing of classes impossible. Also, if Landsat-like images (around 30 meters of spatial resolution) are employed to perform Cerrado vegetation mapping, some mixture of classes is bound to be contained in the result, regardless of the algorithm used – see the works of Jacon et al. (2017) and Girolamo Neto (2018).

Nogueira et al. (2016) was the only work that employed a Deep Learning-based method applied to Cerrado vegetation. They considered the same three classes that are of interest for this work, however, they performed what is called *classification* in computer vision, i.e., patches of Landsat images were entirely designated as Forest, Savanna or Grassland. *Semantic segmentation*, i.e., the assignment of a separate class per pixel, was not performed and a considerable mixture of classes in a single patch could be observed.

Even though no additional applications for Cerrado vegetation mapping can be found, Deep Learning techniques have been increasingly applied in the RS field (Zhu et al., 2017). For example, Yang et al. (2018) used different CNN architectures to not only identify land cover, but also to predict land use classes in digital orthophotos from Germany. Meanwhile, La Rosa et al. (2019) used a time series of radar images for pixel-wise crop recognition through Fully Convolutional Networks (FCN) (Long et al., 2015) in tropical regions.

Recently, some advances related to vegetation mapping and Deep Learning have been achieved. One example is Sothe et al. (2019), that integrated LiDAR (Light Detection and Ranging) and optical data to classify a subtropical forest area and their results presented best accuracies when CNNs were applied. The U-net (Ronneberger et al., 2015), a type of CNN widely used

for semantic segmentation, has also been applied for other vegetation mapping applications, such as forest damage identification (Hamdi et al. 2019) and identification of farmland and woodlands (Zhang et al. 2018).

2. MATERIALS AND METHODS

2.1 Study area

The study area (Figure 1) is the Brasília National Park (BNP) in the Federal District, Brazil, with approximately 300 km² of preserved Cerrado vegetation. The BNP is an important protected area of the Cerrado biome, because it contains several endangered species (e.g., Jaguar – *Panthera onca* and Anteater - *Myrmecophaga tridactyla*) and a dam that is responsible for 25% of the Federal District's water supply.
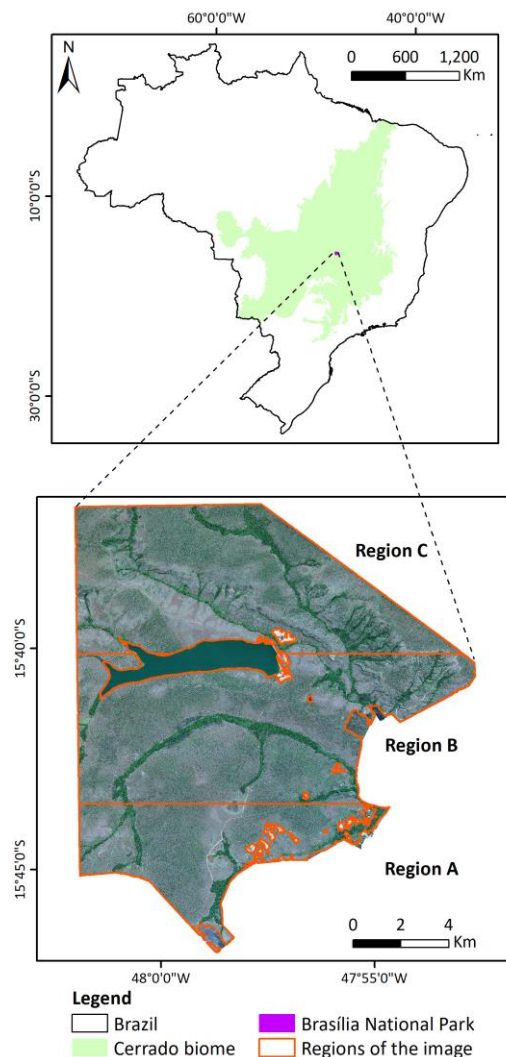

Figure 1. Brasília National Park and its location in the Brazilian Savanna (Cerrado).

In Table 1, the three major groups of physiognomies are described and examples of their patterns, extracted from a WorldView-2 image, are presented. The only Forest physiognomy found in this area is the Gallery Forest (Mata de Galeria). The Savanna physiognomies in BNP are Wooded Savanna (Cerrado Denso), Typical Savanna (Cerrado Típico) and Shrub Savanna (Cerrado Ralo). Shrub Grassland (Campo Sujo), Open Grassland (Campo Limpo) and Rocky Grassland

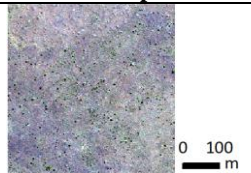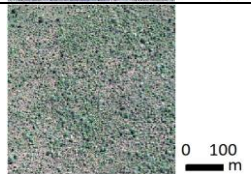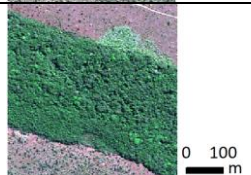(Campo Rupestre) are the Grassland physiognomies present in BNP.

Class	Example	Description
Grassland		Herbaceous vegetation with or without shrubs
Savanna		Trees scattered over shrubs and herbaceous layers, mostly no continuous canopies
Forest		Arboreal species forming continuous or discontinuous canopies

Table 1. Description of the classes and examples of their patterns observed in WorldView-2 image, R(5)G(3)B(2) composition.

2.2 Remote Sensing data

A WorldView-2 image (tile ID 103001003373A600) acquired in July 22, 2014 with a spatial resolution of 2 meters was used in this study. The image has 8 spectral bands: Coastal (400-450nm), Blue (450-510nm), Green (510-580nm), Yellow (585-625nm), Red (630-690nm), Red-Edge (705-745nm), Near Infrared 1 (NIR-1, 770-895nm) and Near Infrared 2 (NIR2, 860-1040nm).

The image was converted from Digital Numbers (DNs) to surface reflectance using the Fast Line-of-sight Atmospheric Analysis of Hypercubes (FLAASH) algorithm (Perkins et al., 2005). Additionally, a mask was created to exclude built-up areas, water bodies, bare soil and burned areas from the analysis.

Reference data was adapted from the "Prevention, Control and Monitoring of Irregular Burnings and Forest Fires in the Cerrado" project (De Brito et al., 2017). In the scope of that project, the entire Cerrado biome was classified in the same three classes considered in this work. However, that was done in a 30-meter spatial resolution from Landsat data acquired in 2000. Therefore, manually editing and adaptation of the reference data was necessary, both to make it suitable for the WorldView-2 spatial resolution and to correct changes that had occurred between 2000 and 2014. The adaptation was based on visual interpretation performed by a human interpreter with experience in mapping Cerrado vegetation.

2.3 Network architecture

In this work, a variant of the well-known U-net architecture (Ronneberger et. al. 2015), proposed by Kumar (2018), was used for the pixel-wise classification. Consisting of convolutional layers only, the network belongs to the group of fully convolutional neural networks (FCNNs, Long et al., 2015). Compared to more traditional CNNs like LeNet (LeCun et al., 1990) and AlexNet (Krizhevsky et al., 2012) that predict a single class for each image patch, FCNNs are tailored to the task of pixel-wise classification. In particular, they take an image patch with an arbitrary number of channels as input and predict a label-map usually of the same spatial size as the input. Ronneberger et. al. (2015) propose to split the network into a multi-layer encoder that successively reduces the spatial resolution and increases the number of filters per kernel and a multi-layer decoder that successively up-scales the features to the original spatial resolution. They further use skip-connections between encoder and decoder layers of the same spatial resolution in order to preserve low-level details, required for the precise prediction of object boundaries.

The architecture used in this work mainly follows the design-choices by Ronneberger et. al. (2015), however, it was modified as follows. While the original version uses unpadded convolutions, a zero-padding was used to preserve the spatial size along the network. As a further modification, the up-sampling is based on transposed convolutions with a stride of two along both spatial dimensions, instead of the originally used up-sampling operation based on interpolation of features. Further network parameters, like the number of layers and filters per layer are depicted in Figure 2.
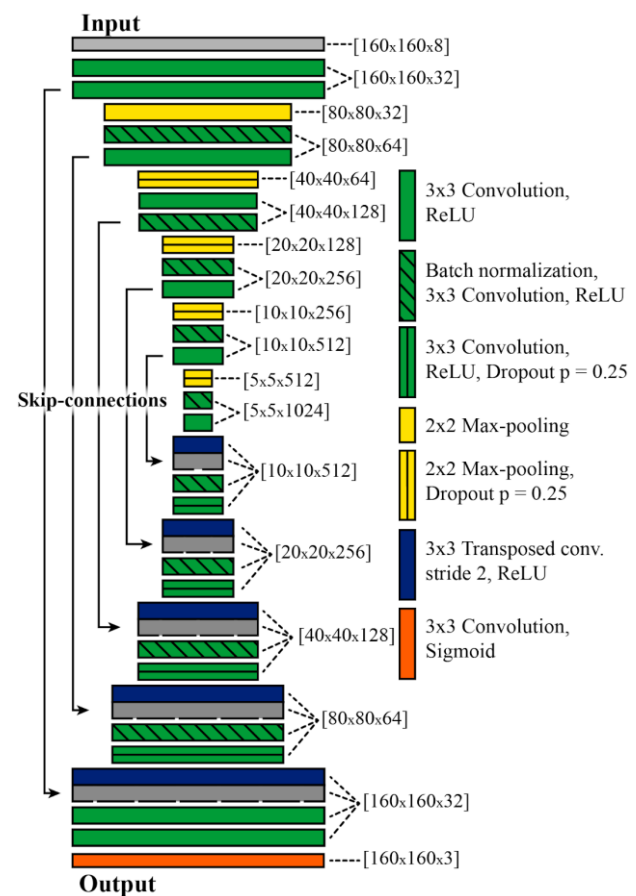

Figure 2. Modified U-net architecture (adapted from Kumar, 2018).

While the last layer of a network for pixel-wise classification of image data is usually modelled by the Softmax function, here the Sigmoid function is used. Using such an output layer is preferred here, since it presented higher accuracies in preliminary tests. This allows the model to predict independent probabilities per class and per pixel. The final class predictions are obtained by choosing the respective classes with the highest

probabilities. The Network is implemented in Keras (Chollet et al., 2015) with TensorFlow as backend (Abadi et al., 2015).

2.4 Training, validation and testing

Both WorldView-2 image and reference data were divided into three regions A, B and C (Figure 1), and were cropped in non-overlapping and adjacent tiles of 160x160 pixel to be used as samples. The selected regions, which were used in the cross-validation procedure explained below, contain roughly similar distributions of the classes of interest. The samples that contained any "no data" value (i.e., pixels originally covering built-up areas, water bodies, bare soil and burned areas) were excluded from further processing.

In each cross-validation experiment, training and validation samples were extracted from two (training) regions (e.g., A and B), and test samples from the other (test) region (e.g., C). From the total of samples from the training regions, 70% were randomly selected to train the network, and 30% for validation. Table 2 shows the numbers of samples used in each experiment. For the training and validation sets, those numbers include samples generated by data augmentation. Six data augmentation techniques were employed: horizontal and vertical flips, transposition and three rotations: by 90, 180 and 270 degrees.

During training, the early stopping criterion (called *patience* in the Keras library) was set to 50, i.e., if after 50 epochs the validation accuracy did not increase, training was stopped. After training, the three networks that resulted from each experiment were tested using the corresponding test regions (see Table 2). The obtained pixel-wise classifications were then compared with the reference data, and a confusion matrix was generated. Based on the confusion matrix, the following evaluation metrics were computed: Overall Accuracy (OA), Precision (P), Recall (R) and F-1 score (F1).

The OA corresponds to the percentage of pixels with the respective labels assigned correctly, considering all classified image. P is the proportion (0 to 1) of pixels that was predicted for a class, and actually belongs to that class; it is the complement of the commission error. R is the proportion (0 to 1) of pixels of a particular class that was successfully identified; it is the complement of the omission error. F1 is the harmonic mean of P and R for each class.

	Training	Validation	Test
Experiment #1	Regions A + B: 70% (5439 samples)	Region A + B: 30% (2331 samples)	Region C (645 samples)
Experiment #2	Regions B + C: 70% (6951 samples)	Regions B + C: 30% (2982 samples)	Region A (336 samples)
Experiment #3	Regions A + C: 70% (4802 samples)	Region A + C: 30% (2065 samples)	Region B (774 samples)

Table 2. Regions and number of samples used in training, validation and test in each cross-validation experiment.

3. RESULTS AND DISCUSSION

3.1 Accuracy assessment

The training, validation and test accuracies of each cross-validation experiment, and the number of epochs needed for the stabilization of the networks' training are presented in Table 3. In training, all accuracies were higher than 90.5% and the difference between training and validation accuracies was not

higher than 2.7%. The average OA, considering the semantic segmentation obtained in all test experiments, was 87.0%.

In the first experiment, the U-net trained with regions A and B was tested in region C. This was the case where the network had most difficulty to generalize the physiognomy class samples. Even though we are dealing with three classes for now, Savannas and Grasslands could be divided into several types of physiognomies in the BNP, according to the classification system proposed by Ribeiro and Walter (2008). The Shrub Savanna, which consists of a less dense type of Savanna (tree cover ranging from 5% to 20% and tree height from 2 to 3 meters), appears more often in region C, whereas the network trained in regions A and B was trained with more samples from Wooded Savanna (tree cover ranging from 20% to 50% and tree height average of 3 to 6 meters). Consequently, the network trained in regions A and B could not always identify Shrub Savanna, and classified some as Grassland.

CV exp.	Train accuracy (%)	Validation accuracy (%)	Epoch	Test Accuracy (%)
1	91.3	88.6	47	83.8
2	90.6	89.0	34	89.6
3	92.1	89.4	77	88.7

Table 3. Epochs and accuracies for each Cross Validation (CV) experiment.

Comparing the predicted semantic segmentation label images with the reference data on a pixel basis, we generated the confusion matrix presented in Table 4. The network had the best performance in the semantic segmentation of Savannas, achieving an F1 of 0.90. The confusion between the classes occurs mainly in the transition areas, where there are adjacent physiognomies if a scale of increasing density is considered, e.g., Wooded Savanna is the densest Savanna physiognomy and was responsible for most of the areas in which Savanna was classified as Forest. This type of error was also reported by Jacon et al. (2017) and Girolamo Neto (2018).

In terms of carbon stock estimation, the worst case of misclassification would be between Grassland (lowest carbon stock) and Forest (highest carbon stock). This case was minimal in the predicted image: only 0.3% of the Grasslands were classified as Forest and only 0.7% of the Forests were classified as Grasslands. These errors occurred mainly in transition areas, in the edges of the Gallery Forests, where often some regions of Humid Open Grassland physiognomy exist, which have wet soil with very dark green vegetation. These two aspects decrease the reflectance and make this type of Grassland more similar to Forest areas, especially because of the shadows among the trees.

Besides the confusion in some transition areas, the Forest class had a high F1 of 0.88. The worst performance occurred in the classification of Grasslands. 17.7% of the Grasslands were classified as Savanna and 9.6% of the Savannas were classified as Grassland. Although the two percentages seem far from each other, the absolute numbers are not so far (2.3 million and 2.8 million pixels, respectively). Savanna occupies a much larger area than Grassland (29.3 million pixels and 13.3 million pixels, respectively) in BNP. Additionally, the confusion between adjacent classes of Grassland and Savanna, considering the classification system of Ribeiro and Walter (2008), is the most common error when classifying the Brazilian Savanna (Abade et al. 2015; Jacon et al. 2017). For this reason, some authors prefer to group Grasslands and Savannas into only one class (MMA, 2015). Also using high spatial resolution images, Silva and Sano (2016) achieved recall of 0.76 for Savannas and

precisions of 0.83 and 0.53 for Forests and Grasslands using a semi-automatic methodology with Euclidean distances. This shows the better performance when using the Deep Learning approach.

		Predict				
		G	S	F	Total	P
Reference	G	10,913,712	2,356,272	40,299	13,310,283	0.82
	S	2,829,064	26,179,248	345,507	29,353,819	0.89
	F	27,124	442,188	3,234,853	3,704,165	0.87
	Total	13,769,900	28,977,708	3,620,659	46,368,267	
	R	0.79	0.90	0.89		
	F1	0.81	0.90	0.88		

Table 4. Confusion matrix, Precision (P), Recall (R) and F1-score (F1) for Grassland (G), Savanna (S) and Forest (F).

3.2 Cerrado physiognomy mapping

In order to analyse spatially the pixel-wise classification, the predicted label image (considering all three experiments) and the reference are presented in Figure 3. As the predicted label image was generated by joining all classified patches, in some regions of the image it is still possible to notice the transition between the patches. This can be seen in the northeastern region of the predicted image in Figure 3.
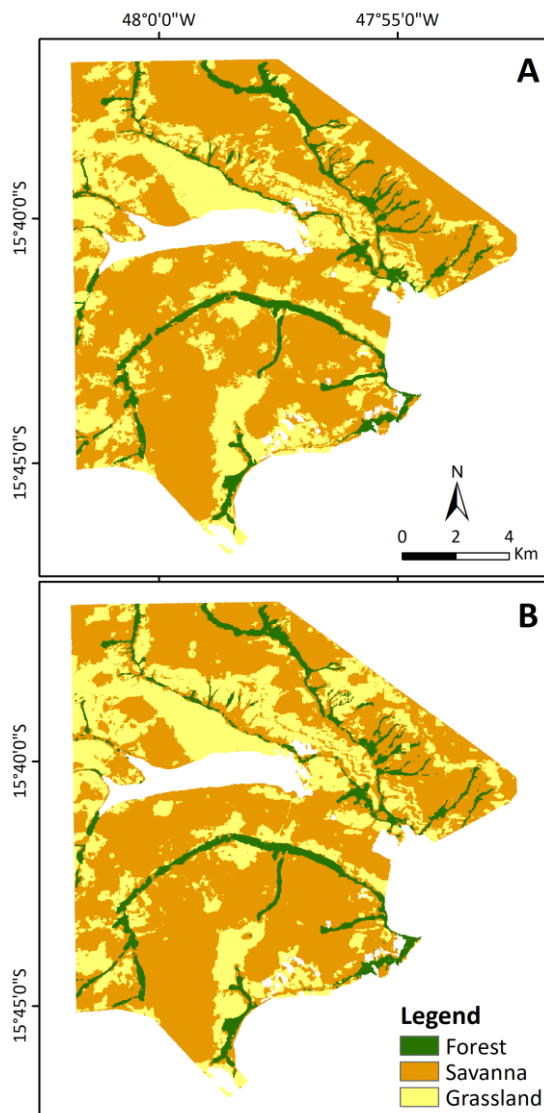

Figure 3. a) Reference and b) Predicted image.

The networks were capable of performing an adequate delineation of Forests (F1 = 0.88). Savannas and Grasslands were classified also coherently, but for these two classes some areas show that our method still needs improvements. In the north of the central lake, in region C, Grassland was overestimated due to the presence of large areas of Shrub Savanna, as mentioned in the previous section.

In Figure 4, two patches were chosen as examples to illustrate closer the obtained classification. The pattern of the spatialized errors corroborates the rates of misclassifications observed in Table 4. The correct delineation of the three studied classes can be confirmed, and it is possible to observe the errors associated with the transition areas. The errors between Forest and Grassland (F x G) can barely be noticed. The largest misclassified areas, in grey (Figure 4), represent transition areas between Grassland and Savanna (G x S). Even in field campaigns, identifying where Grassland turns into Savanna is not an easy task, and this fuzzy aspect is represented in the classification.
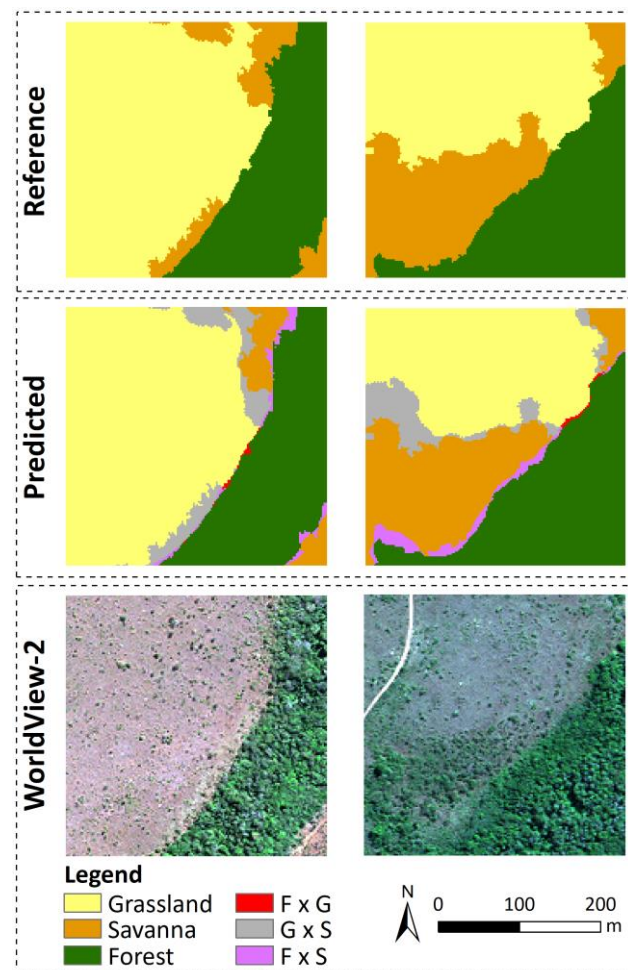

Figure 4. Examples of predicted patches, their references and the corresponding appearance in the WorldView-2 image in R(5)G(3)B(2) composition. FxG refers to confusion between Forest and Grassland; GxS, between Grassland and Savanna; and FxS, between Forest and Savanna.

The other misclassification areas are related to minor aspects of the BNP. There are small regions mainly populated with only one species. This aspect changes the pattern of the physiognomy and can generate some misclassification. Besides that, some

physiognomies, such as Open Grassland with Murundus, occur in very small portions of the image. This physiognomy has a very peculiar pattern, different from usual Open Grassland and Shrub Grassland. Perhaps, in small regions, delineating and excluding these minor areas from the automatic classification can be an option.

4. CONCLUSIONS

Using a modified U-net architecture, we were able to perform a semantic segmentation of the Brasília National Park (BNP) in three major groups of physiognomies: Grasslands, Savannas and Forests. Each of these three classes can be subdivided into several types of physiognomies and, for this reason, one class can be associated to different patterns of vegetation. Despite this variation, the networks trained and tested in different regions of the BNP achieved accuracies close to or higher than 85%. The misclassifications are mainly related to transition areas, where differentiating the edges between classes is a difficult task.

As a future work, we intend to implement a methodology for the more detailed level of the classification system of Ribeiro and Walter (2008), i.e., hierarchically classify types of Grassland, Savanna and Forest based on the result of the first level of the 3-classes delimitation. Additionally, we plan to test different Deep Learning architectures and to integrate the method with Geographic Object-Based Image Analysis (GEOBIA) methods to further improve the accuracy rates of the semantic segmentation.

ACKNOWLEDGEMENTS

REFERENCES

Abade, N., Júnior, O., Guimarães, R., de Oliveira, S. 2015. Comparative analysis of MODIS time-series classification using support vector machines and methods based upon distance and similarity measures in the Brazilian Cerrado-Caatinga boundary. *Remote Sensing*, 7(9), 12160-12191.

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... & Ghemawat, S. 2015. TensorFlow: Large-scale machine learning on heterogeneous systems. Available at: https://www.tensorflow.org/ (08 January 2020).

Borges, E. F., Sano, E. E. 2014. Séries temporais de EVI do MODIS para o mapeamento de uso e cobertura vegetal do oeste da Bahia. *Boletim de Ciências Geodésicas*, 20(3), 526-547.

Chollet, F., others. Keras, 2015. Available at: https://keras.io (17 December 2019).

De Brito, A., Jacon, A., Queiroz, J., Valeriano, D., 2017. Mapping the main vegetation types of Cerrado biome in the year 2000. *PANGAEA*. Available at: https://doi.org/10.1594/PANGAEA.882605

Girolamo Neto, C. D. Identificação de fitofisionomias de Cerrado no Parque Nacional de Brasília utilizando Random Forest aplicado a imagens de alta e média resoluções espaciais. 2018. *Thesis (PhD in Remote Sensing)*, National Institute for Space Research, 188p.

Guirado, E., Tabik, S., Alcaraz-Segura, D., Cabello, J., Herrera, F. 2017. Deep-learning versus OBIA for scattered shrub detection with Google earth imagery: Ziziphus Lotus as case study. *Remote Sensing*, 9(12).

Hamdi, Z. M., Brandmeier, M., Straub, C. 2019. Forest Damage Assessment Using Deep Learning on High Resolution Remote Sensing Data. *Remote Sensing*, 11(17), 1976.

INPE – National Institute for Space Research. OBT – Earth Observation General Coordination. Monitoring program of the Amazon and other biomes. 2019. Available at: http://terrabrasilis.dpi.inpe.br/downloads/ (10 January 2020).

Jacon, A. D., Galvão, L. S., dos Santos, J. R., Sano, E. E. 2017. Seasonal characterization and discrimination of savannah physiognomies in Brazil using hyperspectral metrics from Hyperion/EO-1. *International journal of remote sensing*, 38(15), 4494-4516.

Kumar, S. Deep U-net for satellite image segmentation. Available at: https://github.com/reachsumit/deep-unet-for-satellite-image-segmentation/ (17 December 2019).

Kussul, N., Lavreniuk, M., Skakun, S., Shelestov, A. 2017. Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, 14(5), 778-782.

Krizhevsky, A., Sutskever, I., Hinton, G. E., 2012. ImageNet classification with deep convolutional neural networks. In: *International Conference on Neural Information Processing Systems* (NIPS'12) 25, 1, 1097-1105.

La Rosa, L. E. C., Feitosa, R. Q., Happ, P. N., Sanches, I. D., da Costa, G. A. O. P. 2019. Combining Deep Learning and Prior Knowledge for Crop Mapping in Tropical Regions from Multitemporal SAR Image Sequences. *Remote Sensing*, 11(17).

LeCun, Y.; Bengio, Y.; Hinton, G. 2015. Deep learning. *Nature*, 521(7553), 436-444.

LeCun, Y., Boser, B. E., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. E., Jackel, L. D. 1990. Handwritten digit recognition with a back-propagation network. In *Advances in neural information processing systems*, 396-404.

Long, J., Shelhamer, E., Darrell, T. 2015. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3431-3440.

MapBiomas Project - Collection 4.0 of Brazilian Land Cover & Use Map Series. Available at: http://plataforma.mapbiomas.org/map#coverage (13 January 2020).

Mittermeier, R. A., Turner, W. R., Larsen, F. W., Brooks, T. M., Gascon, C. 2011. Global biodiversity conservation: the critical role of hotspots. *Biodiversity hotspots*, Springer, 3–22.

MMA – Ministério do Meio Ambiente. Cadastro Nacional de Unidades de Conservação – CNUC. 2010. Available at: < http://www.mma.gov.br/areas-protegidas/cadastro-nacional-de-ucs/> (09 January 2020).

MMA – Ministério do Meio Ambiente. Mapeamento do Uso e Cobertura da Terra do Cerrado – Projeto TerraClass Cerrado 2013. 2015. Available at: < http://www.dpi.inpe.br/tccerrado/Metodologia_TCCerrado_201 3.pdf> (10 January 2020).

Nogueira, K., Dos Santos, J. A., Fornazari, T., Silva, T. S. F., Morellato, L. P., Torres, R. D. S. 2016. Towards vegetation species discrimination by using data-driven descriptors. In: *2016 9th IAPR Workshop on Pattern Recogniton in Remote Sensing (PRRS)*, IEEE, 1-6.

Perkins, T., Adler-Golden, S., Matthew, M., Berk, A., Anderson, G., Gardner, J., Felde, G., 2005. Retrieval of atmospheric properties from hyper and multispectral imagery with the flaash atmospheric correction algorithm. *Remote Sensing of Clouds and the Atmosphere X*. International Society for Optics and Photonics, 2005, vol. 5979, p. 59790E. https://doi.org/10.1117/12.626526

Ribeiro, J. F., Walter, B. M. T. 2008. As principais fitofisionomias do Bioma Cerrado. In: Sano, S. M., Almeida, S. P., Ribeiro, J. F. *Cerrado: ecologia e flora*. Brasília: EMBRAPA, 152-212.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 234-241.

Silva, L. R., Sano, E. E. 2016. Analysis of RapidEye Satellite Images for Discrimination of Land Cover of Cerrado Biome. *Revista Brasileira de Cartografia*, 68(7), 1269-1283.

Sothe, C., De Almeida, C. M., Schimalski, M. B., Liesenberg, V., La Rosa, L. E. C., Castro, J. D. B., Feitosa, R. Q. 2019. A comparison of machine and deep-learning algorithms applied to multisource data for a subtropical forest area classification. *International Journal of Remote Sensing*, 41(5), 1943-1969.

Yang, C., Rottensteiner, F., Heipke, C. 2018. Classification of land cover and land use based on convolutional neural networks. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences 4*, Nr. 3, 4(3), 251-258.

Zhang, C., Liu, J., Yu, F., Wan, S., Han, Y., Wang, J., Wang, G. 2018. Segmentation model based on convolutional neural networks for extracting vegetation from Gaofen-2 images. *Journal of Applied Remote Sensing*, 12(4).

Zhu, X. X., Tuia, D., Mou, L., Xia, G. S., Zhang, L., Xu, F., & Fraundorfer, F. 2017. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4), 8-36.