DEEP LEARNING APPLIED TO REMOTE SENSING: AN APPROACH FOR THE DETECTION OF CATTLE DRINKING FOUNTAINS USING PLANET IMAGES

Leandro Leal Parente[1], Ana Paula Matos e Silva[2], Evandro Taquary[3] e Laerte Guimarães Ferreira[4]

[1]leal.parente@gmail.com; [2]annapaulamatos1@gmail.com; [3]evandro.taquary@gmail.com; [4]laerte@ufg.br;
Image Processing and GIS Lab at the Federal University of Goiás (LAPIG/UFG),
CAMPUS II Samambaia, Goiânia – GO

ABSTRACT

The recent increscent of orbital sensors is producing an unprecedented amount of global surface data in the history of mankind, which has a tremendous potential for application of deep learning (DL) approaches. In this direction, this work implemented a process of object detection (*i.e.* cattle drinking fountains) over Planet images, acquired in two distinct months (*i.e.* nov/2017 and mar/2018), and identified more than 24,000 drinking fountains in an area of 18,000 km^2, located in the state of Goiás. This approach, based on the neural network architecture of U-Net and trained with data obtained by visual inspection, produced results that are spatially consistent and compatible with high resolution images. The results indicate a potential of application in other satellite data (*e.g.* Worldview, Sentinel) and in the identification of other objects (*e.g.* treetop, pivot, road).

Key words — deep-learning, machine learning, planet images, cattle drinking fountain, pasturelands

1. INTRODUCTION

The increasing availability of orbital sensors with better spectral, spatial, and temporal resolutions is producing an unprecedented volume of satellite data in recent years [1]. In this perspective, CubeSat constellations, with units measuring 10 × 10 × 11.35 cm and weighing <1.33 kg [2], are making possible a full coverage of the global surface with a daily frequency and high spatial resolution (*e.g.* 3-5m), putting Planet, with more than 130 CubeSats in orbit, as one of the major provider of satellite data nowadays [3].

In this direction, approaches based on deep learning (DL) can enable a better information extraction of this growing data volume. Using modern and computationally efficient techniques of machine learning (ML), these approaches are being used to solve a wide range of problems (e.g. computer vision, signal processing, natural language), analyzing only raw data and eliminating the need of feature engineering, a fundamental step for classical ML algorithms (e.g. Random Forest, SVM) [4].

Considering the application of DL in satellite data, approaches based on semantic segmentation are a promising alternative for classification and object detection in high spatial resolution images [5]. Thus, this work implemented a semantic segmentation approach (*i.e.* U-Net [6]) for object detection (*i.e.* cattle drinking fountains) over Planet images, obtained in two distinct months (i.e. nov/2017 and mar/2018), and evaluated the results and the main challenges related to the usage of this technique in remote sensing.

2. DATA AND METHODS

Our study area was the IBGE topographic chart SD-22-ZC, comprising an area of 18,000 km^2[7], located in the state of Goiás. For this area, we analysed all Planet data available for two months, November/2017 (beginning of the wet season) and March/2018 (end of the wet season), respectively 2,020 and 3,141 images with four spectral bands (*i.e.* blue, green, red, and near infrared; 4-band PlanetScope Scene). Only images with less than 20% of cloud coverage were selected to compose the two monthly mosaics for the entire study area.

These mosaics were produced using a cloud screening approach, based on a specific threshold of spectral values; an equalization of histograms, for each band, using a reference image for each month; and a best pixel approach, based on the median of all monthly pixel values. This entire routine was implemented in Google Earth Engine, a cloud-computing platform that allows the analysis of remote sensing data on a planetary scale [8], resulting in the exportation of two monthly mosaics, through Google Drive, to a local workstation.

Considering these mosaics, a semantic segmentation approach was implemented to detect cattle drinking fountains. We chose this specific object in light of the following criteria:

- Characteristics of land use and occupation: The chart SD-22-ZC comprises a relevant set of municipalities in Goiás (*e.g.* Aruanã, Britânia, Jussara, Itapirapuã, Novo Brasil) with extensive pasture areas, whose the main economic activity is cattle raising.

- Potential applications: Once detected, drinking fountains can indicate better managed pastures and/or be correlated with the carrying capacity of a specific pasture area.
- Potential of detection: Generally the size of cattle drinking fountains varies from 100 to 500 m^2, an area potentially detectable with Planet images.

The implementation of this approach required a set of training data compatible with the spatial resolution of the Planet images, which was elaborated through vector editing, through GIS software, and visual inspection of the monthly mosaics and Google Earth images. Only small soil excavations with the purpose of accumulating water for animal watering, located in pasture areas, were considered as cattle drinking fountains, thus ignoring small lakes for irrigation and residential supply, resulting in a total of 2,634 polygons.

The semantic segmentation, implemented through TensorFlow [10], took as reference U-Net [6], a neural network architecture capable to consider the spatial context in the process of object detection. Its training occurred through the use of several chips (*i.e.* a set of pixels with regular squared size) of input and training (figure 1), produced from planet mosaics, which were stacked and normalized according to the median and standard deviation, and with the polygons of drinking fountains converted to raster format. The ratio between the size of the input and training chip, respectively 286x286 and 100x100 pixels, was established according to the amount of convolutional and pooling layers, originally proposed by the U-net architecture [6].

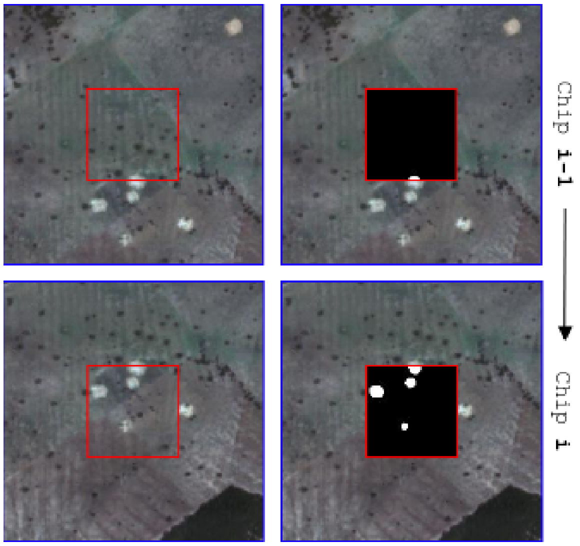

Figure 1. The ratio between input chips, highlighted in blue, and training chips, highlighted in red, indicating that to classify an area of 100x100 pixels the model considered 286x286 pixels, taking into account the spatial context, in a more comprehensive way, in the process of object detection

The generation of chips occurred by a sequential process for the entire training region, an equivalent of 10% of the study area. However, to avoid a extreme unbalanced training scenario, only chips with at least 1 pixel of drinking fountain and located specifically in areas of confusion, previously known (*i.e.* rivers, roads, urban area), were considered. At the end of this process 11,144 chips were produced using two data augmentation strategies: the horizontal flip of chip and the overlap between neighboring chips, an equivalent of 25%.

The detection of cattle drinking fountains was treated as a problem of binary classification, which allowed the use of the Intersection-Over-Union, a cost function more suitable for this category problems [11]. Other modifications of U-Net architecture were implemented as described below:

- Batch normalization [12] and *l2_regularizer* equal to 0.3 in all convolutional layers of neural network, and dropout [13] equal to 0.5, only in the last convolutional layer, seeking a reduction in the possibility of overfitting during the training step;
- Sigmoid activation function in the output layer, guaranteeing a result with pixel values between 0 and 1;
- Utilization of Nadam optimizer [14], which incorporates nesterov momentum in the Adam optimizer, increasing, in general, the speed of convergence and the possibility of finding a better local minimum for the cost function;

The training of U-Net was performed with 80% of chips (20% of chips were used for testing), a mini-batch equal to 32 and by 100 epochs on a multi-GPU server, with an NVIDIA Titan-X and an NVIDIA 1080ti. The resulting model was used to estimate, on a pixel basis, a producer and user accuracy of test chips and to identify all cattle drinking fountains throughout the chart SD-22-ZC. The source code used by this study is available at https://github.com/NexGenMap/dl-semantic-segmentation.

3. RESULTS AND DISCUSSIONS

The training process took 6 hours to complete and presented cost curves, for training and test data, with similar trends (figure 2). From epoch 20, cost curves began to drift apart, a indication of overfitting, however the trend of downward remained until the last epoch. The user and producer accuracy, calculated according to the test data, were respectively 95% and 91%, indicating a prevalence of omission errors. Despite of relatively high values of accuracy, the prevalence of omission errors, on a pixel based estimation, suggests that the model was not able to delimit the correct format of the cattle drinking fountains.

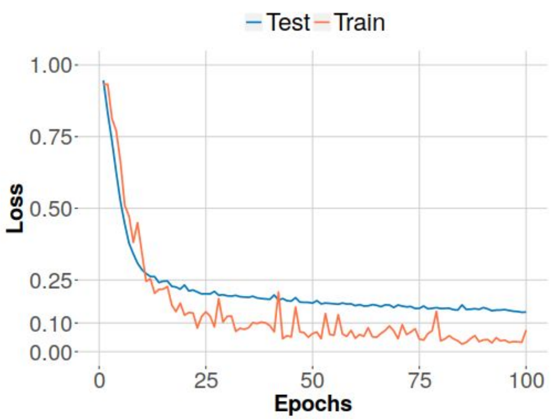

Figura 2. The chart of cost function Intersection-Over-Union, for training and test data, showing a similar trend for both and a possible generalization capacity of the model.

The prediction for the entire chart SD-22-ZC detected 24,766 drinking fountains, which presented a spatial consistency with the pasture areas of the region (figure 3). The model was able to generalize the criteria of vector delimitation used in the visual inspection, which considered only pixels with presence of water in at least one of the Planet mosaics. It avoided the confusion with areas of bare soil and roads, which present the same spectral behavior in both mosaics; on the other hand, underestimated the area of some large drinking fountains. The model has confused pixels contaminated by clouds and cloud shadows, an effect that can be corrected by using a better cloud screening approach in the process of mosaic generation. There was confusion in regions located in watercourses, which, at the beginning of the wet season were sandbank. These confusions could be minimized by including chips with such aspects and conditions in the training process, which would consequently increase the generalization capacity of the model.

Despite the promising results achieved, the generalization capability of the cattle drinking fountain detection, beyond the study area, is unknown. An assessment onto another geographic region located, for example, in the Amazon biome could reveal aspects of the landscape that were not considered in our training approach, increasing the uncertainty of object detection process. This uncertainty could be better quantified and comprehended with very high resolution images (*i.e.* <1m), obtained in the same months evaluated by this study, and/or with a field validation, allowing an accuracy assessment with a set of data completely independent.

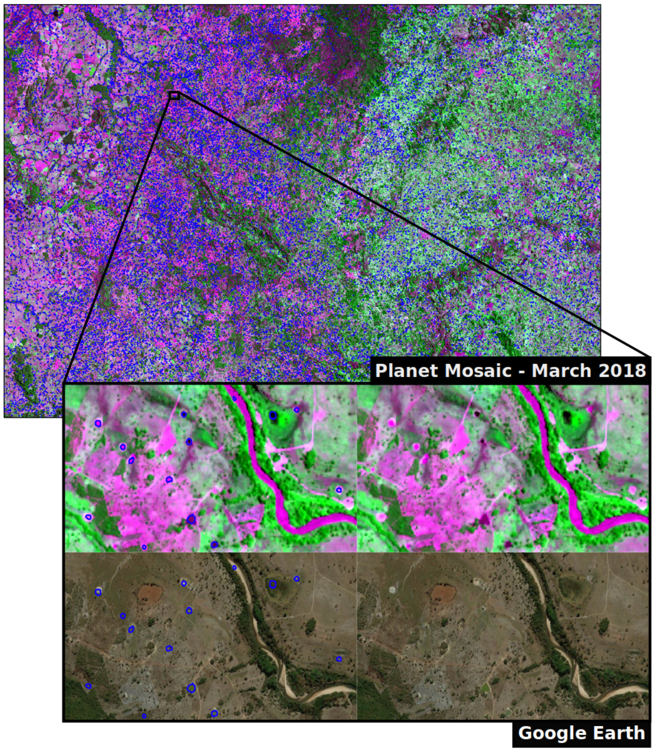

Figure 3. Cattle drinking fountains (highlighted in blue), a total of 24,766 objects, identified throughout the chart SD-22-ZC, over the Planet mosaic of March 2018 (false color RED / NIR / Green). Despite the spatial consistency of drinking fountain with pasture areas, even in Google Earth images, there were confusions in regions located in watercourses, which, at the beginning of the wet season were sandbank.

4. CONCLUDING REMARKS

This study processed a few thousand of Planet images, through Google Earth Engine, to produce two monthly mosaics and implemented a deep learning approach applied to object detection, which identified more than 24,000 cattle drinking fountains in an area of 18,000 km^2. Our results presented a spatial consistency and will be used, in future studies, to promote a better understanding of pasturelands of the evaluated region. As any semantic segmentation, this approach requires spatially explicit training data, generated through vector editing and visual inspection. With this kind of training data, the U-Net neural network has the potential to detect other objects (*e.g.* tree canopy, pivot, road) in Planet images, as well as in other high resolution images (*e.g.* Worldview, Sentinel).

ACKNOWLEDGMENTS

This work, under the NexGenMap and MapBiomas initiatives (http://mapbiomas.org), was supported by the Gordon and Betty Moore Foundation (GBMF), The Nature Conservancy (TNC), WWF Brazil, the State of Goiás

Research Foundation (FAPEG) and the Brazilian National Council for the Scientific and Technological Development (CNPq).

5. REFERENCES

[1] Yan, J.; Ma, Y.; Wang, L.; Choo, K.K.R. & Jie, W., A cloud-based remote sensing data production system. Future Generation Computer Systems, 86 (v.), 1154-1166 (pp.), 2018.

[2] Puig-Suari, J.; Turner, C. & Ahlgren, W., Development of the standard CubeSat deployer and a CubeSat class PicoSatellite. In:IEEE Aerospace Conference Proceedings, 1 (v.), 347 - 353 (pp.), 2001.

[3] Houborg, R. & McCabe, M.F., Cubesat enabled Spatio-Temporal Enhancement Method (CESTEM) utilizing Planet, Landsat and MODIS data. Remote Sensing of Environment, 209 (v.), 211-226 (pp.), 2018.

[4] LeCun Y., Bengio Y. & Hinton G., Deep learning. Nature, 521 (v.), 436-444 (pp.), 2015.

[5] Kemker, R., Salvaggio, C. & Kanan, C., Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning. ISPRS Journal of Photogrammetry and Remote Sensing, in prees, 2018.

[6] Ronneberger, O., Fischer, P. & Brox, t., U-net: Convolutional networks for biomedical image segmentation. In:International Conference on Medical image computing and computer-assisted intervention. Munich:Springer, 234-241 (pp.), 2015.

[7] IBGE. Plano de Trabalho do Departamento de Cartografia. Available in: <https://biblioteca.ibge.gov.br/visualizacao/periodicos/206/dgc_1971.pdf>. Accessed in: 18 junho 2018.

[8] Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D. & Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment, 202 (v.), 18-27 (pp.), 2017.

[9] Vieira, P.; Ferreira, N. & Ferreira L, Análise da vulnerabilidade natural da paisagem em relação aos diferentes níveis de ocupação da bacia hidrográfica do rio vermelho, Estado de Goiás. Sociedade & Natureza, 26 (v.), 385-400 (pp.), 2014.

[10] Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; Kudlur, M.; Levenberg, J.; Monga, R.; Moore, S.; G. Murray, D.; Steiner, B.; Tucker, P.; Vasudevan, V.; Warden, P.; Wicke, M.; Yu, Y. & Zheng, X., Tensorflow: a system for large-scale machine learning. In: OSDI., 265-283 (pp.), 2016

[11] Rahman, A. & Wang, Y., Optimizing intersection-over-union in deep neural networks for image segmentation. In: International Symposium on Visual Computing, 234-244 (pp.), 2016.

[12] Ioffe, S. & Szegedy, C., Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167, 2015.

[13] Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I. & Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research, 15 (v.), 1 (n.), 1929-1958 (pp.), 2014.

[14] DOZAT, T. Incorporating nesterov momentum into adam. Available in : <https://openreview.net/pdf?id=OM0jvwB8jIp57ZJjtNEZ>. Accessed in: 01 outubro 2018.