U-Net to predict segmentation in raster images

Otaviano da Cruz Neto[1]

INPE - Instituto Nacional de Pesquisas Espaciais, São José Dos Campos, Brazil
otavianocruz@id.uff.br

Abstract. This project use the U-Net architecture to predict segmentation of deforestation in the Amazon Rain forest, located in Jamari National Forest in Rondônia State. Two images 657 by 1196 pixels and spatial resolution of 3m by 3m were collected by COSMO-Sky-Med platform, a system composed by 4 satellite receiving information in X-Band. The captures were one on June 5th, 2018 and the other on October 8th, 2018. For this, it was necessary to run a grid search with the number of feature maps of the U-Net. The winner model was the model with 128 feature maps and producing 0.02 on the validation error. This model did a great work locating the deforestation in most cases but the model performance can be increased by using a autoencoder data augmentation and compare the effect of this approach to the static data augmentation (flip and rotation).

Keywords: Segmentation · U-Net · Amazon Deforestation.

1 Introduction

The data-driven learning applied in Deep Learning start to navigate between different areas. In computational vision, machine learning models brought new technologies to automate object tracking[], people identification[], and so on. The Convolutional Neural Networks (CNN) are one of the most successful architectures that use convolutional operations to create feature maps that is more significant than the raw pixels space.

The Remote Sensing raw material is images that can easily feed a CNN model to identify different behavior using RGB images, radar images and others spectral information. Knowing that, researchers are using this approach to segment images in order to identify deforestation in protected area[4].

With that understanding, we structured this work as follows. Section 2 compile informations about study area, data collection and the methodology used in this project. The results and descriptions about the performed processes were presented in Section 3. Finally, the conclusions and next steps of the work were exposed in Section 4.

2 Materials and Methods

2.1 Study Area

The studies are directed to the location of the Jamari National Forest, in the State of Rondônia. This region is the scene of several conflicts due to the high rate of deforestation in the region.

2.2 Data Collection

The area was registered by a geo referenced image of 657 pixels by 1196 pixels obtained by the COSMO-Sky-Med platform. This platform is composed of 4 satellites that capture spatial resolution information of 3 meters by 3 meters and X-band. In this work, two images will be used, one captured on June 5th, 2018 and the other on October 8th, 2018. Due to the large proportion of the images and the need to create training, testing and validation sets, cutouts for the X band were created. These cutouts have dimensions of 64 by 64, totaling 180 images. That is, the methodology, Figure 1 will be a supervised learning approach whose input will be two images and the model output will be compared with the segmentation generated from classical algorithms described by Tahisa Neitzel kuck et. al in the article entitled 'A Comparative Assessment of Machine-Learning Techniques for Forest Degradation Caused by Selective Logging in an Amazon Region Using Multitemporal X-Band SAR Images' [3]. The image set was separated into training, validation and testing subsets with proportions 68%, 12% and 20%, respectively.

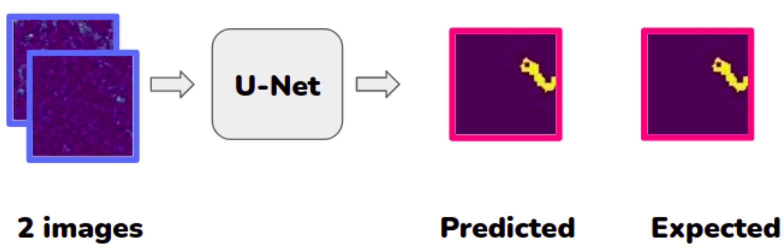

Fig. 1. The model training methodology is based on the input of two images of 64x64 pixels and only one band, that is, the input format is 64x64x2. These images are presented to the network whose output is compared with the expected result from the *Root Mean Square Error* metric. With this result, the error backpropagation is performed in the form of updating the degrees of freedom present in the U-Net[5]

2.3 U-Net

The network architecture is described by Olaf Ronneberger et. al. at 'U-Net: Convolutional Networks for Biomedical Image Segmentation' is described as contracting path (left side) and an expansive path (right side). The contracting path follows the typical architecture of a convolutional network. It consists of the repeated application of two 3x3 convolutions (unpadded convolutions), each followed by a rectified linear unit (ReLU) and a 2x2 max pooling operation with stride 2 for downsampling. At each downsampling step we double the number of feature channels. Every step in the expansive path consists of an upsampling of the feature map followed by a 2x2 convolution ("up-convolution") that halves the number of feature channels, a concatenation with the correspondingly cropped feature map from the contracting path, and two 3x3 convolutions, each followed by a ReLU. The cropping is necessary due to the loss of border pixels in every convolution. At the final layer a 1x1 convolution is used to map each 64-component feature vector to the desired number of classes. In total the network has 23 convolutional layers. The U-Net model will be built from the Tensorflow [1] library. The model can be illustrated by the Figure 2.

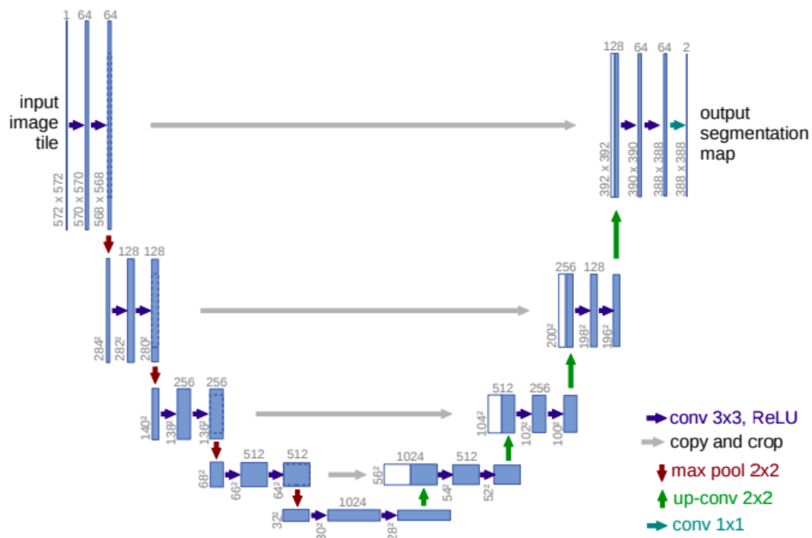

Fig. 2. The U-Net model architecture [5]

2.4 Hyper parameter optimization

To generalize the behavior of the data, it was necessary to evaluate which combinations of hyper parameters fit best to describe them. The evaluated hyper parameters were the number of features maps in the U-Net topology, ranging between 32 and 128. For this case, around 1800 models will be evaluated. This type of hyperparameter optimization via brute force has an extensive evaluation time due to the large number of models, so we use the *Optuna* [2] package for the Python programming language as a way to parallelize the process model training, in addition to selecting the best models and recording all the information generated. The metric used for evaluation is the combination of three factors (Runtime, Validation Error and Training Error), those models that manage to minimize these three parameters will be selected as promising models. To compile in a single panel all the results generated as the learning curve and hyperparameters of the model, the *Tensorboard* [1] was used. In addition to the tools, a device with an Intel(R) Xeon(R) Gold 5118 processor @ 2.3GHz was used, with 755GB of RAM memory and equipped with 4 NVIDIA Tesla V100 16GB GDDR5 video cards.

3 Results

The resultant model of the grid search was the model with 128 map features. The metric used to evaluate the model was the root mean square error in order to try to predict the pixel values in gray scale. By training the model it was possible to build the learning curve, Figure 3,

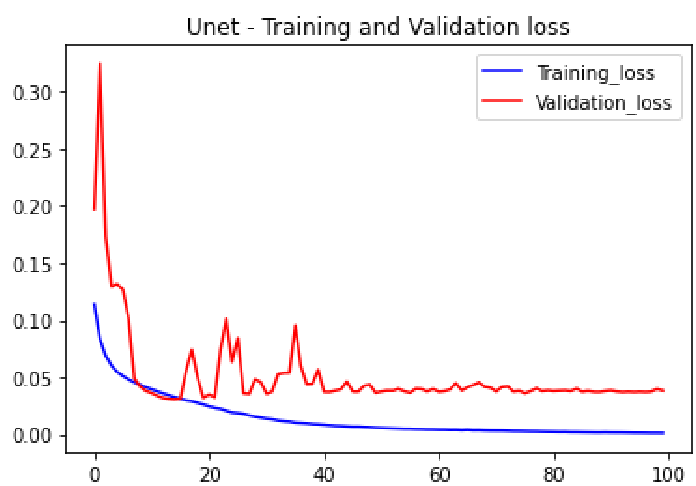

Fig. 3. Learning curve containing loss function values for different training epochs.

Figure 3 shows us that the optimization behavior is being performed and that, through the validation curve, it is able to obtain an error measure close to those obtained within the training set. In the test phase, a mean square error of

0.02 was obtained and thus we can select a sample to observe the performance of the model.

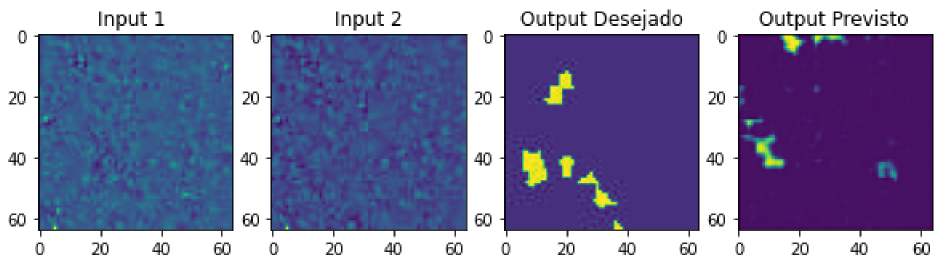

Fig. 4. Sample with good result in the generalization of learning in the test set.

The test sample shown in Figure 4 shows us an image segmentation error but it is visible that there is a visible learning of the segmentation positioning. Below, Figure 5, we can see a bad example of segmentation performed by the model,

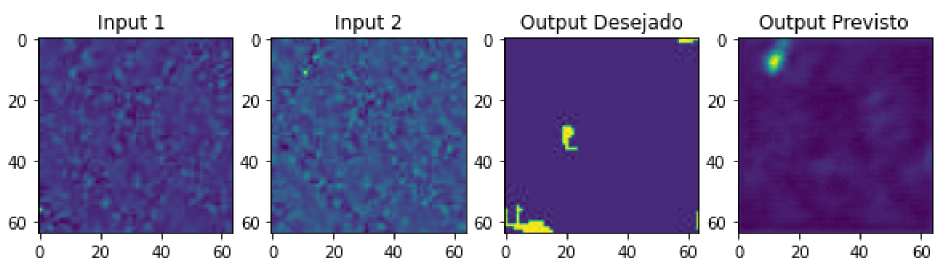

Fig. 5. Sample with poor result in the generalization of learning in the test set.

In the next section we will compile the main results and possible modifications that can generate better results.

4 Conclusions

After the training, validation and testing phases, we can conclude that U-Net has good generalization capacity, considering the learning curves. A possible change that can improve the model's segmentation results is to perform a binary classification of the pixels instead of calculating the probability of extracting wood or not. Furthermore, increasing the amount of images available to use the model can positively affect the results, as exposing the model to new different conformations induces it to extract new information. As a future work we can use autoencoder as a data augmentation and compare with a static data augmentation (flip and rotation).

Acknowledgments

We would like to thank Marcos Quiles Gonçalvez and Juarez Lopes Ferreira da Silva of the Center for Innovation on New Energy (CINE) to provide the hardware to run this grid search.

Resources

The code to run the grid search and create the dashboard is disponible on https://github.com/HardProxy/seg_unet.

References

1. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X.: TensorFlow: Large-scale machine learning on heterogeneous systems (2015), https://www.tensorflow.org/, software available from tensorflow.org
2. Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M.: Optuna: A next-generation hyperparameter optimization framework. CoRR **abs/1907.10902** (2019), http://arxiv.org/abs/1907.10902
3. Kuck, T.N., Sano, E.E., Bispo, P.d.C., Shiguemori, E.H., Silva Filho, P.F.F., Matricardi, E.A.T.: A comparative assessment of machine-learning techniques for forest degradation caused by selective logging in an amazon region using multitemporal x-band sar images. Remote Sensing **13**(17) (2021). https://doi.org/10.3390/rs13173341, https://www.mdpi.com/2072-4292/13/17/3341
4. Lee, S.H., Han, K.J., Lee, K., Lee, K.J., Oh, K.Y., Lee, M.J.: Classification of landscape affected by deforestation using high-resolution remote sensing data and deep-learning techniques. Remote Sensing **12**(20) (2020). https://doi.org/10.3390/rs12203372, https://www.mdpi.com/2072-4292/12/20/3372
5. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. CoRR **abs/1505.04597** (2015), http://arxiv.org/abs/1505.04597