

Segmentando Desmatamento em Imagens de Satélite Usando U-Net

Danilo L. Cardoso¹
EMAp/FGV, Rio de Janeiro, RJ

Com o desenvolvimento da computação e de algoritmos de inteligência artificial, a área de Visão Computacional vem a cada dia se revelando útil para uma série de aplicações, como diagnósticos médicos, carros autônomos e sensoriamento remoto. Esta área se beneficiou bastante com o desenvolvimento de algoritmos de *deep-learning* [2], que utilizam grandes conjuntos de dados (anotados ou não) para automatizar etapas que antes precisavam de bastante conhecimento e interferência humana, como a detecção de *features* na imagem e a escolha de algoritmos de classificação.

Aplicações de Visão Computacional em imagens de satélite (sensoriamento remoto) são muito pesquisadas, pois oferecem uma maneira automática de se analisar a grande quantidade de dados que um satélite pode coletar durante sua operação. Essas aplicações incluem: previsão do tempo, detecção de queimadas, análise de biodiversidade, entre outras. No entanto, apesar desta vasta quantidade de dados de imagem disponíveis, há uma dificuldade em se criar anotações, muitas vezes necessitando de interferência humana especializada. Algoritmos não supervisionados lidam com esse problema em certas aplicações.

Este trabalho será focado na detecção de desmatamento usando sensoriamento remoto. Como os satélites disponibilizam dados semanalmente, esse tipo de aplicação permite uma rápida ação de governantes para atuar contra possíveis ações criminosas em florestas, além de oferecer uma estimativa de quanto a floresta foi degradada ao longo dos anos.

A base de dados utilizada para fazer a inferência foi disponibilizada pelo *MultiEarth 2023* [1], um *workshop* online de sensoriamento remoto focado na floresta amazônica. O *workshop* disponibilizou uma série de imagens de satélite e uma série de máscaras das áreas desmatadas anotadas manualmente por especialistas. Uma vantagem de utilizar os dados do *workshop* é que podemos comparar os resultados do nosso método com o de outras pessoas.

Para o trabalho, foram escolhidas as imagens do satélite *Landsat 8* por serem mais leves que as imagens do *Sentinel* e por abrangerem mais bandas que o *Landsat 5*. As imagens têm tamanho 85×85 *pixels*, resolução de $30m$ e, diferentemente das imagens de três bandas visuais (*RGB*), elas consistem de 7 bandas espalhadas pelo espectro de radiofrequências. Essas bandas adicionais são uma das vantagens de utilizar imagens de satélite, pois assim o modelo pode capturar informações que não seriam percebidas por uma câmera *RGB*. Um dos principais problemas ao se trabalhar com imagens de satélite é a presença de aerossóis (como nuvens e fumaça), que atrapalham a captura da informação do solo. O banco de dados escolhido adiciona uma camada de “qualidade” às imagens para indicar a presença dos aerossóis.

O modelo utilizado no trabalho é a *U-Net*[4], um tipo de rede convolucional desenvolvida originalmente para a segmentação de imagens médicas. Para contornar o problema dos aerossóis nas imagens, foram removidas as imagens que possuem mais de 60% de sua área coberta, e em seguida foram selecionadas as três imagens anteriores à região que se deseja segmentar. Cada dessas imagens passa por um *encoding* (composto de duas convoluções e um ASPP [5]) de pesos compartilhados antes de entrarem na *U-Net*. A rede recebe as três imagens codificadas e retorna

¹danilo.leмос.cardoso@hotmail.com

um valor entre 0 e 1 para cada pixel, que será interpretado como a certeza que o modelo tem de que aquela região da imagem é uma área desmatada. Para rodar a descida de gradiente foram utilizadas duas perdas: a “entropia cruzada binária”, e a perda SSIM [3], que mede a similaridade entre a máscara anotada e a máscara retornada pelo modelo. As perdas utilizadas para avaliar a segmentação foram a medida F1 e a Interseção Sobre União. Todo o código pode ser encontrado em: <https://github.com/danilolc/deforestation-segmentation>.

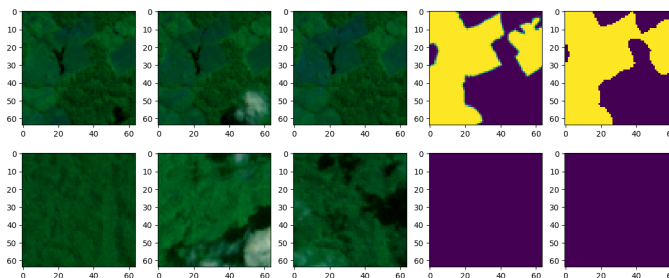


Figura 1: Primeiras três colunas: as três imagens de entrada na rede. Quarta e quinta colunas: *ground-truth* e resultado da segmentação respectivamente.

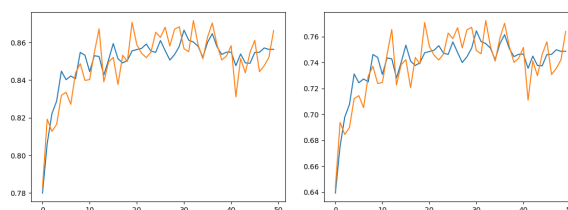


Figura 2: Gráfico com a medida F1 e Interseção Sobre União em cada *epoch* do treinamento. A linha laranja é no conjunto de treino e a linha azul é no conjunto de validação. A medida F1 máxima no conjunto de validação foi de 0.866.

Referências

- [1] M. Cha, G. Angelides, M. Hamilton, A. Soszynski, B. Swenson, N. Maidel, P. Isola, T. Perron e B. Freeman. **MultiEarth 2023 – Multimodal Learning for Earth and Environment Workshop and Challenge**. 2023. arXiv: 2306.04738 [cs.CV]. URL: <https://arxiv.org/abs/2306.04738>.
- [2] I. J. Goodfellow, Y. Bengio e A. Courville. **Deep Learning**. <http://www.deeplearningbook.org>. Cambridge, MA, USA: MIT Press, 2016.
- [3] J. Nilsson e T. Akenine-Möller. **Understanding SSIM**. 2020. arXiv: 2006.13846 [eess.IV]. URL: <https://arxiv.org/abs/2006.13846>.
- [4] R. Olaf, F. Philipp e B. Thomas. **U-Net: Convolutional Networks for Biomedical Image Segmentation**. 2015. arXiv: 1505.04597 [cs.CV]. URL: <https://arxiv.org/abs/1505.04597>.
- [5] Y. Wang, B. Liang, M. Ding e J. Li. “Dense Semantic Labeling with Atrous Spatial Pyramid Pooling and Decoder for High-Resolution Remote Sensing Imagery”. Em: **Remote Sensing** 11.1 (2019), p. 20. URL: <https://www.mdpi.com/2072-4292/11/1/20>.