

Segmentação Semântica de Tipos de Uso de Solo na Amazônia Utilizando Aprendizado Profundo

Joel P. de Oliveira¹, Marly G. F. Costa² e Cícero F. F. Costa Filho²

¹Centro Gestor e Operacional do Sistema de Proteção da Amazônia (CENSIPAM)
Manaus – AM – Brasil

²Universidade Federal do Amazonas (UFAM)
Manaus – AM – Brasil

joelparente@gmail.com, marlygfcosta@gmail.com, cffccfilho@gmail.com

Abstract. *This work proposes a deep learning methodology to perform the semantic segmentation of LANDSAT-8 images of the following types of uses and land cover: forest, pasture and agriculture. The field of study is the Amazon region. The reference data were extracted from the TerraClass project in 2014. A CNN architecture was evaluated along with three optimization methods: SGDM, ADAM and RMSProp and the dropout and L_2 regularization methods, as methods for generalization improvement. The best results were obtained using the RMSProp optimization method. The accuracy values obtained for the evaluated images were above 93%.*

Resumo. *Este trabalho propõe uma metodologia de aprendizado profundo para realizar a segmentação semântica de imagens LANDSAT-8 dos seguintes usos e cobertura de solos: floresta, pasto e agricultura. O campo de estudo é a região Amazônica. Os dados de referência foram extraídos do projeto TerraClass do ano de 2014. Foi avaliada uma arquitetura CNN juntamente com três métodos de otimização: SGDM, ADAM e RMSProp e os métodos dropout e regularização L_2 , como métodos para a melhoria de generalização. Os melhores resultados foram obtidos utilizando o método de otimização RMSProp. Os valores de acurácia obtidos para as imagens avaliadas ficaram acima de 93%.*

1. Introdução

Sensoriamento remoto é a utilização de diversas tecnologias com o objetivo de estudar os fenômenos que ocorrem na superfície da Terra. Essas tecnologias compreendem sensores, equipamentos instalados a bordo de aeronaves, espaçonaves e outras plataformas. Os dados gerados a partir de sistemas de sensoriamento remoto são de grande utilidade para várias aplicações, dentre as quais podemos citar: planejamento urbano, agrícolas, geológicas, monitoramento de desmatamento [Novo 2008].

Com respeito ao monitoramento do desmatamento na Amazônia através de sensoriamento remoto, o Instituto Nacional de Pesquisas Espaciais (INPE) é uma referência mundial. Dentre os vários projetos desenvolvidos pelo INPE para esse monitoramento, destacam-se o Programa de Monitoramento do Desflorestamento na Amazônia Legal (PRODES) e o TerraClass. O PRODES fornece dados por meio de mapas anuais de desmatamento na região Amazônica. O projeto TerraClass utiliza os dados gerados pelo

PRODES para realizar uma classificação de uso e cobertura de solo nas seguintes classes: floresta, pasto, agricultura, áreas urbanas, mineração e outros. Esse tipo de informação ajuda os órgãos do governo a desenvolverem políticas públicas de prevenção para conter o avanço do desmatamento [Noma et al. 2013]. Embora o PRODES e o TerraClass sejam grandes projetos e forneçam dados bastante confiáveis, eles ainda contam com uma significativa parcela do trabalho realizada com intervenção humana.

Para tornar mais eficiente o trabalho que envolva a análise de imagens de sensoriamento remoto, trabalhos têm sido propostos na área de aprendizado de máquina. [Bem et al. 2020] utilizaram aprendizado profundo para mapear o desmatamento entre 2017 e 2018 e entre 2018 e 2019. Foram utilizadas três cenas do LANDSAT-8 de regiões do estado do Pará e do Amazonas. Os dados de referência utilizados foram os dados PRODES dos anos de 2017 a 2019. Os melhores resultados obtidos com a CNN ResUnet foram acurácia e F1-Score de 99,93% e 94,65%, respectivamente. [Adarme et al. 2020] utilizaram aprendizado profundo para detecção automática de desmatamento por meio de duas imagens LANDSAT-8. As áreas de estudo foram duas regiões com diferentes padrões de desmatamento: os biomas Amazônia e Cerrado no Brasil. Os dados de referências utilizados foram os dados PRODES dos anos de 2017 e 2018. Os autores obtiveram como melhores resultados uma acurácia e um F1-Score de 95% e 63%, respectivamente, na Amazônia, e de 97% e 78%, respectivamente, no Cerrado.

Segundo [Aggarwal 2018], a partir da primeira década desse século, a rede neural renasceu sob o novo rótulo chamado de aprendizado profundo. Aprendizado profundo aborda o uso de modelos computacionais com arquiteturas hierárquicas compostas por múltiplas camadas de processamento a fim de "aprender" determinadas representações de dados nos mais diferentes formatos: áudio, imagens e texto [Lecun et al. 2015]. Quatro modelos principais compõem o aprendizado profundo: Redes Neurais Convolucionais (*Convolutional Neural Networks* - CNN), Máquinas de Boltzmann, *Autoencoders* e *Sparse Coding*. Segundo [Aggarwal 2018], a grande quantidade de dados disponíveis nos últimos anos, juntamente com o aumento do poder computacional, permitiram a utilização de arquiteturas mais profundas em relação ao que era possível anteriormente.

Os trabalhos de [Bem et al. 2020] e [Adarme et al. 2020] utilizaram aprendizado profundo para detectar áreas desmatadas em regiões da Amazônia brasileira. Os resultados alcançados por esses autores foram bastante satisfatórios. No entanto, os autores avaliaram uma região específica da Amazônia. Outra crítica que fazemos aos trabalhos previamente publicados é que os mesmos não se preocupam em disponibilizar o conjunto de dados utilizado para *benchmark*.

Um grande desafio para se treinar CNNs com imagens de sensoriamento remoto é que, normalmente, as classes são desbalanceadas. Em outras palavras, em uma região capturada pela imagem, existe um grande desbalanceamento em termos de área dos diversos tipos de solos. Esse problema pode levar, no treinamento da CNN, aos métodos de otimização terem um melhor desempenho nas classes mais frequentes. Para contornar esse problema, nesse trabalho utiliza-se a técnica de imagem-mosaico. Essa técnica consiste de extrair, a partir da imagem de satélite, pequenas retalhos retangulares uniformes, ou seja, com apenas um tipo de solo. A partir desses retalhos monta-se uma imagem maior, denominada de imagem-mosaico.

Portanto, considerando a ausência de um modelo de rede em aprendizado profundo adequado para classificação dos diversos tipos de solo para a região amazônica, e a dificuldade de treinamento de uma CNN com bom desempenho em todas as classes, esse trabalho pretende, através da técnica de imagem mosaico, utilizar uma CNN para segmentação e classificação de diversas regiões da Amazônia nos seguintes tipos de uso e cobertura de solo: agricultura, pasto e floresta.

2. Material e métodos

2.1. Conjunto de dados utilizados

Neste trabalho foram utilizadas imagens LANDSAT-8/OLI. As imagens estão disponíveis gratuitamente no site: <http://earthexplorer.usgs.gov/>. Foram utilizadas as seguintes bandas: Vermelho (B4); Infravermelho próximo (B5); Infravermelho de ondas curtas 1 (B6). Segundo [Yu et al. 2019], B4, B5 e B6 é a melhor combinação de três bandas para aplicações de sensoriamento remoto cujo o objetivo é realizar a classificação do solo. As imagens utilizadas abrangem os estados do Amazonas, Mato Grosso, Pará e Rondônia. Para a geração do padrão-ouro do classificador, foram utilizados os resultados do projeto TerraClass, do ano de 2014. Os dados foram adquiridos gratuitamente no site do INPE [Inpe 2019]. Foram extraídas informações sobre áreas de floresta, pasto e agricultura, correspondendo a um problema de reconhecimento de três classes.

Utilizando os dados de referências do projeto TerraClass, foram gerados retalhos de imagens de tamanho 40x40 *pixels* de cada uma das classes de solo. Foram gerados 4.000, 225.000 e 6.000 retalhos de agricultura, floresta e pasto, respectivamente. Utilizando os conjuntos de retalhos de imagens, foram geradas imagens que definiremos como imagens-mosaico. Cada uma dessas imagens-mosaico tem dimensão 400x400 *pixels* e são geradas selecionando-se aleatoriamente retalhos de agricultura, floresta ou pasto. Para cada imagem-mosaico gerada, gera-se também uma imagem-mosaico correspondente ao padrão-ouro da mesma. Para a construção dessas imagens que constituíram o padrão-ouro, os *pixels* que correspondem a região de floresta, pasto e agricultura foram marcados com o valor 255, 100 e 1, respectivamente. O conjunto de dados gerado possui um total de 3.600 imagens-mosaico, sendo que 3000 foram destinadas para o conjunto de treinamento e 600 para o conjunto de validação. Na Figura 1 mostra-se um exemplo de imagem-mosaico e o padrão-ouro correspondente. Pode ser observado que, dos 100 retalhos, 34 são de agricultura, 35 de floresta e 31 de pasto.

2.2. Métodos

Neste trabalho, foi avaliada uma arquitetura de CNN, três métodos de otimização e três métodos para a melhoria da generalização. Os métodos de otimização avaliados foram o Gradiente Descendente Estocástico com Momento (SGDM), Propagação da Raiz Média Quadrática (RMSProp) e Estimativa de Dinâmica Adaptativa (ADAM). Para cada um desses métodos foram empregados as seguintes técnicas para a melhoria de generalização: nenhuma técnica, camada de *dropout*, regularização L2 e camada de *dropout* com regularização L2. Os dados de entrada utilizados foram as imagens-mosaico. Foram realizadas 12 simulações (1 arquitetura x 3 métodos de otimização x 4 métodos para melhoria de generalização). Em seguida, o modelo com o melhor desempenho no conjunto de validação foi selecionado para classificar algumas imagens da Amazônia.

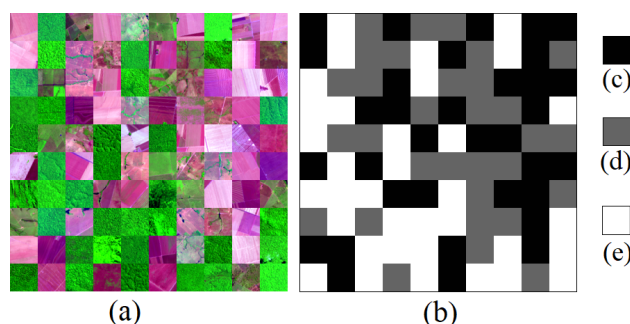


Figura 1. Exemplo de imagem-mosaico. Em (a) mostra-se uma imagem mosaico composta por retalhos de agricultura, floresta e pasto, em (b) a imagem padrão-ouro, e em (c), (d) e (e) temos o padrão-ouro para cada retalho de imagem, correspondente a área de agricultura, pasto e floresta, respectivamente.

2.2.1. Arquitetura CNN, Métricas e Parâmetros de Treinamento

A arquitetura CNN utilizada neste trabalho foi baseada nas arquiteturas propostas no trabalho de [Miyagawa et al. 2018]. Neste trabalho, os autores realizaram a segmentação do lúmen em imagens de tomografia por coerência ótica intravascular (IVOCT). Os melhores resultados para acurácia, valor de Dice e de Jaccard ficaram acima de 99%, 98% e 97%, respectivamente. No trabalho ora apresentado, a CNN utilizada possui duas etapas de subamostragem (*maxpooling*) e duas de sobreamostragem. Antes de cada subamostragem, existem três sequências de camadas convolucionais 3 x 3, camada *batch normalization* e ReLU.

Neste trabalho foram calculadas a Acurácia Global (ACCG), Acurácia Média (ACCM), coeficiente de similaridade de Jaccard (J), coeficiente de similaridade Jaccard ponderado (JP) e Score F1 (F1) [MathWorks 2017]. Uma estação de trabalho com *Windows 10*, *Matlab 2019a* e com *NVIDIA Quadro GV100 32GB* e 5120 núcleos CUDA foi utilizado nos experimentos. Em relação aos parâmetros de treinamento da CNN, foi utilizada a taxa de aprendizado inicial = 0,001, fator de queda de taxa de aprendizado = 0,5, número de épocas = 200, tamanho do lote = 2, parâmetro da camada *dropout* = 0,3, fator de regularização L_2 = 0,001. Esses valores foram ajustados de maneira experimental.

3. Resultados

A Tabela 1 apresenta o desempenho obtido para a CNN quando combinada com os método de otimização e com os métodos para melhoria da generalização. A Tabela 2 apresenta a matriz de confusão para o modelo com o melhor desempenho no conjunto de validação, que neste caso foi aquele no qual foi empregado o método RMS-Prop como método de otimização e sem utilizar nenhuma técnica para melhoria da generalização. Esse modelo foi utilizado para segmentar/classificar algumas imagens da região Amazônica. Na Figura 2 mostramos três imagens LANDSAT-8 de regiões da Amazônia com seus respectivos padrão-ouro, e a imagem classificada pelo modelo CNN. As imagens 1 e 2 correspondem regiões da cena 001/66, enquanto que a imagem 3, a uma área da cena 224/68. Os valores de acurácia obtidos para as imagens 1, 2 e 3 foram 98,91%, 96,72% e 93,10% respectivamente.

	Experimento	ACCG (%)	ACCM (%)	J (%)	JP (%)	FI (%)
1	SGDM	95,984	95,995	92,406	92,389	86,854
2	SGDM/Dropout	94,900	94,919	90,487	90,463	87,111
3	SGDM/L2	96,370	96,381	93,091	93,075	87,981
4	SGDM/Dropout/L2	95,947	95,960	92,328	92,311	89,076
5	ADAM	97,158	97,171	94,537	94,524	89,433
6	ADAM/Dropout	97,265	97,272	94,740	94,727	89,827
7	ADAM/L2	97,095	97,102	94,420	94,408	89,292
8	ADAM/Dropout/L2	96,824	96,838	93,906	93,892	90,229
9	RMSProp	97,467	97,477	95,109	95,097	89,517
10	RMSProp/Dropout	96,807	96,810	93,877	93,865	88,308
11	RMSProp/L2	97,260	97,268	94,729	94,717	90,253
12	RMSProp/Dropout/L2	96,533	96,546	93,387	93,371	90,273

Tabela 1. Desempenho da CNN.

		Classes Preditas		
		Agricultura	Pasto	Floresta
Classes Reais	Agricultura	0,9799	1,96E-02	4,11E-04
	Pasto	4,87E-02	0,9466	4,67E-03
	Floresta	2,56E-04	1,99E-03	0,9978

Tabela 2. Matriz de Confusão para o modelo CNN/RMSProp.

4. Discussão

A partir da Tabela 1, com respeito aos métodos otimização, conclui-se que os resultados para a acurácia obtidos utilizando-se os métodos ADAM e RMSProp ficaram muito próximos, em média 97,08% e 97,017%, respectivamente. Por outro lado, os resultados obtidos com o método SGDM, foram inferiores, com média de 95,80%. A partir da tabela de confusão, verificou-se um maior erro de classificação em regiões em que o padrão-ouro apontava áreas como sendo de pasto mas que foram classificadas como agricultura e vice-versa. O modelo com melhor desempenho foi o modelo CNN com o método RMSProp, sem utilizar nenhuma técnica para melhoria da generalização. Esse modelo foi empregado para avaliar a classificação de algumas regiões na Amazônia apresentadas na Figura 2. A Acurácia obtida para essas três regiões variaram significativamente, entre 98,91% e 93,10%.

5. Conclusões

Este trabalho propôs uma metodologia para segmentar o uso do solo para a região amazônica para as classes de pasto, agricultura e floresta. A metodologia consistiu em avaliar uma arquitetura de CNN e treiná-la utilizando um banco de dados de imagens-mosaico. Foram utilizadas imagens óticas LANDSAT-8 da região amazônica. Diante dos resultados apresentados, pode-se concluir que a metodologia proposta neste trabalho mostrou-se promissora para realizar a tarefa de segmentação/classificação de imagens de sensoriamento para regiões da Amazônia. Com o objetivo de melhorar a acurácia da segmentação das imagens, pretende-se, em trabalhos futuros, avaliar arquiteturas de CNNs mais profundas, utilizar transferência de conhecimento e utilizar mais bandas da imagem LANDSAT-8. Por fim, pretende-se também avaliar a metodologia proposta com um maior número de regiões da Amazônia.

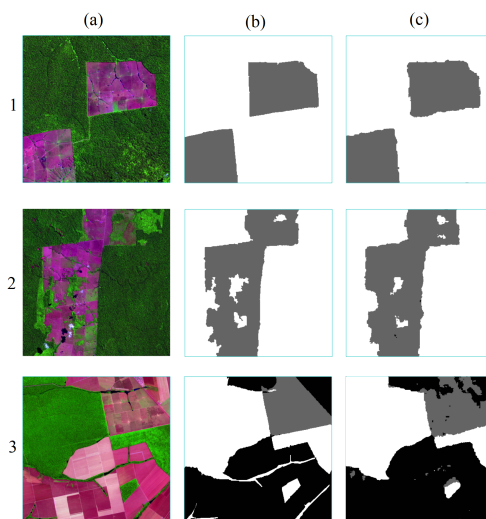


Figura 2. Imagens LANDSAT-8 de regiões da Amazônia. Na coluna (a) temos as imagens originais. Em (b) temos o padrão-ouro. Em (c) temos as imagens classificadas pelo modelo CNN/RMSProp. A acurácia obtida para as imagens 1,2 e 3 foi de 98,91%, 96,72% e 93,10% respectivamente.

Referências

- Adarme, M. O., Feitosa, R. Q., Happ, P. N., A., A. C., and Gomes, A. R. (2020). Evaluation of deep learning techniques for deforestation detection in the brazilian amazon and cerrado biomes from remote sensing imagery. *Remote Sensing*, 12:910.
- Aggarwal, C. C. (2018). *Neural Networks and Deep Learning*. Springer, New York.
- Bem, P. P., Carvalho Junior, O. A., Guimarães, R. F., and Gomes, R. A. T. (2020). Change detection of deforestation in the brazilian amazon using landsat data and convolutional neural networks. *Remote Sensing*, 12:901.
- Inpe (2019). Terraclass. http://www.inpe.br/cra/projetos_pesquisas/dados_terraclass.php.
- Lecun, Y., Bengio, Y., and Hilton, G. (2015). Deep learning. *Nature*, 521:436–444.
- MathWorks (2017). Evaluate semantic segmentation data set against ground truth. <https://in.mathworks.com/help/vision/ref/evaluatesemanticsegmentation.html>.
- Miyagawa, M., Costa, M. G. F., Gutierrez, M. A., Costa, J. P. G. F., and Costa Filho, C. F. F. (2018). Lumen segmentation in optical coherence tomography images using convolutional neural network. *40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 1:1–4.
- Noma, A., Körting, T. S., and Fonseca, L. M. G. (2013). Uma comparação entre classificadores usando regiões e perfis evi para agricultura. *Anais XVI Simpósio Brasileiro de Sensoriamento Remoto*, 1:2250–2257.
- Novo, E. M. L. M. (2008). *Sensoriamento Remoto: Princípios e Aplicações*. Edgard Blücher Ltda, São Paulo.
- Yu, Z., L. D., Yang, R., and Tang, J. (2019). Selection of landsat 8 oli band combinations for land use and land cover classification. *8th International Conference on Agro-Geoinformatics*, 1:1–5.