

ALGORITMO DE DELINEAÇÃO DE COPAS BASEADO EM DEEP LEARNING

José R. G. Braga ^{1,a}, Haroldo F. Campos Velho ^{1,b}, Elcio H. Shiguemori ^{2,d}, Fabien H. Wagner ^{1,c}

¹ Instituto Nacional de Pesquisas Espaciais (INPE), Caixa Postal 515, São José dos Campos, São Paulo, jgarciabraga@gmail.com ^a, haroldo.camposvelho@inpe.br ^b, wagner.h.fabien@gmail.com ^c, ² Instituto de Estudos Avançados (IEAv), Caixa Postal 6044, São José dos Campos, São Paulo, elcio@ieav.cta.br ^d

RESUMO

O sensoriamento remoto é fonte de informações que têm sido aplicadas com sucesso no monitoramento e mapeamento de florestas. Os inventários florestais são uma importante ferramenta para auxiliar na exploração e acompanhamento de espécies. Quando comparadas as técnicas de campo para a realização de um inventário florestal, as técnicas de obtenção de dados por sensoriamento remoto podem baratear os custos de produção e ampliar a área de cobertura. Para auxiliar na exploração sustentável e monitoramento, o inventário florestal deve considerar cada árvore individualmente, o que exige que algoritmos de delimitação de copas sejam desenvolvidos. Atualmente, técnicas de aprendizagem profunda têm obtido resultados consideráveis no processamento de imagens de alta resolução espacial. Desta forma, esta pesquisa apresenta uma nova metodologia baseada em aprendizagem profunda, para realizar a localização e delimitação individual das copas de árvores. Os resultados obtidos são promissores e discutidos no artigo.

Palavras-chave – delimitação de copas, aprendizagem profunda, rede neural convolucional.

ABSTRACT

Remote sensing data have been applied to mapping and forest monitoring. Forest inventories are important to watch the exploration and to monitor plantations and trees species in natural forest. When compared to the field technique used to product forest inventory, remote sensing techniques can reduce production costs and expand the coverage area. The forest inventory should consider each tree individually to ensure a sustainable exploration and high monitoring. Thus, for this purpose, tree delineation algorithms with great precision should be developed. Recently, deep learning techniques have achieved considerable results in high spatial resolution image processing. Therefore, this research presents a new methodology based on deep learning to perform the location and individual tree crown delineation. The results obtained are promising and they are discussed in the article.

Key words – tree crown delineation, deep learning, convolutional neural network.

1. INTRODUÇÃO

Inventários florestais são fundamentais para a conservação e para a tomada de decisão no manejo sustentável de florestas [1]. Os algoritmos para a Delimitação Individual de Copas de Árvores (ITCD, do inglês, Individual Tree Crown Delineation) são uma importante ferramenta para auxiliar o desenvolvimento de um inventário florestal. Nas últimas três décadas, diversos algoritmos foram desenvolvidos com este propósito, os quais foram aplicados como etapas anteriores a processos de obtenção de métricas das árvores ou, até mesmo, para a identificação da espécie. A grande maioria dos algoritmos de ITCD foram desenvolvidos para regiões de clima temperado e sua aplicação em regiões tropicais é muito mais desafiadora, mas não impossível [1, 2].

A grande disponibilidade de dados florestais de alta resolução, oriundos principalmente de tecnologias de sensoriamento remoto, exige que novas técnicas computacionais sejam aplicadas para analisá-los. Recentemente, as técnicas de Aprendizagem Profunda, normalmente designada por seu termo em inglês *Deep Learning*, têm sido aplicadas com sucesso para a análise de dados de sensoriamento remoto [3].

A Aprendizagem Profunda é um conjunto de técnicas que aplicam as Redes Neurais Artificiais (RNAs) profundas, que possuem um grande número de camadas intermediárias (ou escondidas), para a solução de problemas complexos que requerem alto nível de abstração [4]. As RNAs profundas apresentam as mesmas características das RNAs tradicionais, ou seja, sua arquitetura é composta de neurônios e pesos.

A Aprendizagem Profunda tem sido aplicada em diversas áreas do conhecimento, entre elas, a análise de dados de sensoriamento remoto, principalmente para a estudo e acompanhamento da ocupação do solo [3]. Atualmente, as Redes Neurais Convolucionais (CNN, do inglês, Convolutional Neural Networks), que são uma categoria de RNA profunda e alimentada adiante, têm ganhado destaque por sua aplicação em visão computacional, principalmente, para a solução de problemas de segmentação de instância de objetos em uma cena [4]. A segmentação de instância de objetos, uma das mais complexas tarefas de visão computacional, é a tarefa de identificar corretamente o contorno de um objeto no nível do seus pixels [4]. Uma nova categoria de CNN é a *Mask R-CNN*, que possui arquitetura residual. O principal avanço obtido por esta rede é em relação a classificação dos objetos. Com a rede *Mask R-CNN* o objeto é identificado no nível de seus pixels com grande precisão, o que permite sua completa delimitação [4].

Neste contexto, este artigo apresenta o resultado da aplicação da rede *Mask R-CNN* para localizar e delimitar as

copas das árvores presentes em uma imagem de satélite de alta resolução, com o objetivo de auxiliar na produção de inventários florestais.

2. MATERIAIS E MÉTODOS

2.1. Local de Estudo

O local de estudo compreende o município de Nova Xavantina, Mato Grosso (MT), município que localiza-se dentro do Cerrado. Este bioma é o segundo maior em extensão no Brasil, cobrindo aproximadamente 25% de seu território. A Figura 1 apresenta o local de estudo.

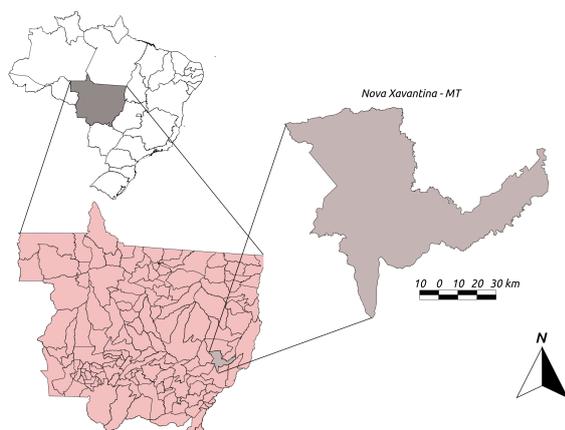


Figura 1: Local de estudo: município de Nova Xavantina, MT

Devido sua importância, há grande interesse no desenvolvimento de um inventário florestal do Cerrado brasileiro. Por este motivo, que uma região deste bioma foi selecionada para o desenvolvimento desta pesquisa.

2.2. Materiais

2.2.1. Conjunto de Dados

O conjunto de dados é formado por uma imagem de uma região do município de Nova Xavantina, Mato Grosso, obtida pelo satélite WorldView-2, e por um conjunto de 1300 exemplos de copas de árvores coletados desta imagem. A coleta dos exemplos de copas foi realizada por meio de uma análise visual (amostragem visual). A imagem da região de interesse utilizada nesta pesquisa foi obtida por meio da composição das bandas R (*Red*, vermelho), G (*Green*, verde) e B (*Blue*, azul), seguido pela re-projeção e fusão com a banda pancromática. Assim, obteve-se uma imagem multi-espectral (imagem colorida) com resolução espacial de 0,5 m. Este pré-processamento foi realizado com auxílio da biblioteca *Orfeo Toolbox*. Da imagem pré-processada foram coletados 1300 exemplares de copas de árvores, cada exemplar é representado por um polígono, que foram obtidos com auxílio do software *Quantum Gis (QGis)*. A média da área das copas de árvores coletadas é igual a 23,7 m² e a reflectância espectral média, em relação as bandas R, G e B é igual a 0.093. A Figura 2 apresenta alguns exemplos de copas de árvores coletados para a composição do conjunto de dados.

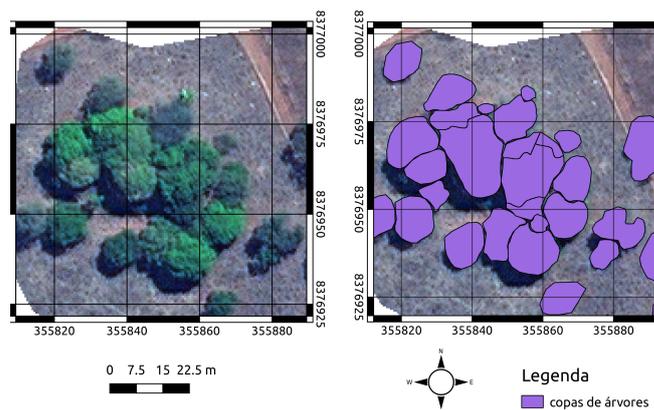


Figura 2: Exemplos de copas de árvore coletadas para a formação do conjunto de dados.

2.3. Métodos

Nesta pesquisa, a rede convolucional *Mask R-CNN* é aplicada para localizar e delinear as copas de árvores. A rede *Mask R-CNN* é treinada para localizar cada copa de árvore presente em uma imagem e determinar quais pixels na região da árvore pertencem ou não a sua copa, ou seja, realizar a segmentação de instância das copas em relação aos outros objetos que compõem a cena. A arquitetura da rede residual ResNet-101 foi utilizada para a configuração da arquitetura da rede *Mask R-CNN*. A arquitetura da ResNet-101 foi selecionada, pois esta rede já foi treinada com um conjunto de dados bastante extenso. Desta forma, não é necessário um conjunto de exemplos de copas de árvores com grande número de elementos para que a rede aprenda a detectar estes padrões [4].

Para o treinamento da *Mask R-CNN*, o conjunto de 1300 exemplares de copas de árvores foi dividido em três: o primeiro conjunto com 750 exemplares para treinamento da rede convolucional, o segundo com 250 exemplares para a validação durante o treinamento e os 300 padrões restantes foram utilizados para a avaliação da rede convolucional na delimitação das copas das árvores. O algoritmo de retro-propagação do erro foi aplicado para o treinamento da rede, ou seja, realizar o ajuste dos parâmetros livre da rede convolucional (valores de peso e viés da rede).

O algoritmo de retro-propagação de erro busca por um conjunto de valores de pesos e viés que minimizam a função de custo $J(w)$, exibida na Equação 1. Os valores de pesos e viés são modificados por uma função de otimização, Equação 2, que calcula o gradiente de cada neurônio da rede, ou seja, a derivada parcial em relação aos pesos e viés do neurônio. Os valores dos pesos e viés são atualizados na direção oposta do valor do gradiente calculado [4].

$$J(w)_k = d_k - o_k \quad (1)$$

Na Equação 1, d_k é a resposta desejada da rede para o padrão apresentado na k -ésima iteração e o_k é a saída da rede para o mesmo padrão.

$$\Delta_{k+1} = w_{k+1} - w_k = -(\eta \frac{\partial}{\partial w_k} J(w)_k) \quad (2)$$

Na Equação 2, w representa os valores dos pesos do neurônio na k -ésima iteração e η a taxa de aprendizado da rede convolucional

A rede *Mask R-CNN* foi treinada apenas para a localização e delimitação de copas de árvores, todos os outros elementos que compõem a cena são considerados plano de fundo (*background*). O treinamento foi realizado em 30 épocas. Após o treinamento da rede convolucional, cada uma das imagens, que contém os padrões utilizados para a avaliação do algoritmo de ITCD, foram apresentadas à rede *Mask R-CNN*, na qual sua resposta é uma imagem com todos os pixels pertencentes a uma mesma copa de árvore destacados por uma cor. Destaca-se que a rede *Mask R-CNN* também classifica cada instância de copa de árvore, ou seja, distingue cada árvore que compõe a cena. A implementação da rede convolucional *Mask R-CNN* foi realizada na linguagem Python, com auxílio da biblioteca TensorFlow [4].

Para a avaliação dos resultados obtidos pelo algoritmo de ITCD proposto, duas métricas foram aplicadas. A primeira métrica, (A_g), verifica a acurácia global do algoritmo, sua formulação é apresentada na Equação 3.

$$A_g = \frac{P_c}{P_t} \cdot 100 \quad (3)$$

sendo P_t a quantidade de pixels que compõem o polígono do padrão da copa de árvore. Os polígonos foram definidos visualmente durante a coleta de dados e representam a verdade-terreno. A variável P_c é a quantidade de pixels classificados corretamente, pela rede convolucional, em relação à verdade-terreno.

A segunda métrica de avaliação aplicada é denominada de Intersecção Sobre União (*IoU*, Intersection Over Union), sua formulação é apresentada na Equação 4. A métrica *IoU* verifica a quantidade de sobreposição entre o *bounding box* do objeto na imagem de verdade-terreno e o *bounding box* do objeto na imagem classificada pelo algoritmo de ITCD, quanto maior seu valor, em uma escala de $0 \dots 1$, melhor a performance do algoritmo na localização e delimitação do objeto.

$$IoU = \frac{A_s}{A_t} \quad (4)$$

Na Equação 4, A_s é o valor da área de intersecção entre o *bounding box* do objeto na imagem de verdade-terreno e do *bounding box* na imagem classificada e A_t é a soma das áreas dos dois *bounding box* menos a área de intersecção.

3. RESULTADOS

A Figura 3 apresenta uma das imagens utilizadas para a avaliação da rede *Mask R-CNN* como algoritmo de ITCD. Nesta região, a reflectância espectral média das copas de árvores para as bandas R, G e B da imagem foi de 0,09, e área média de cobertura é de $15,2 \text{ m}^2$.

A Figura 4 apresenta o resultado da localização e delimitação das copas pela rede *Mask R-CNN* sobre a Figura 3. Na Figura 4, cada cor representa uma copa de árvore diferente detectada pelo algoritmo. Por meio da análise da Figura 4, pode-se observar que o algoritmo detectou com precisão a



Figura 3: Exemplo de imagem de região do município de Nova Xavantina utilizada para avaliação do algoritmo de ITCD

maioria das árvores que estão presentes na imagem mesmo quando agrupadas, mas o algoritmo apresentou dificuldades na detecção de copas muito pequenas ou com pouca cobertura de folhas.



Figura 4: Localização e delimitação das copas de árvores pelo rede *Mask R-CNN*.

A Figura 5 apresenta o resultado da sobreposição dos *bounding boxes* das copas da verdade-terreno e do estimado pelo algoritmo, ou seja, o resultado visual da métrica *IoU*. Nesta figura, é possível observar que o algoritmo proposto obtém uma alta taxa de acerto na localização e delimitação de copas em relação à verdade-terreno. As copas delimitadas apenas com a cor vermelha são aquelas que não foram definidas na obtenção da verdade-terreno mas foram detectadas pelo algoritmo proposto. Em verde apresenta-se a verdade-terreno e em marrom o resultado obtido pelo algoritmo. Para a região exibida o valor da métrica *IoU* é igual a 0,803.

A Figura 6 apresenta em vermelho o erro de delimitação obtido pelo algoritmo proposto, ou seja, a diferença entre a delimitação da verdade-terreno e a delimitação obtida pelo método proposto. Para a região exibida na Figura 6 a métrica



Figura 5: Resultado da sobreposição do bounding box da verdade-terreno com o delineado pelo algoritmo proposto.

A_g é igual a 0,653.



Figura 6: Resultado do erro de delineação.

Para os 300 padrões aplicados para a avaliação do algoritmo de ITCD a média da variável IoU ficou em 0,812 e média da variável A_g ficou em 0,638.

4. DISCUSSÃO

A análise dos resultados demonstram que o algoritmo proposto apresenta resultados superiores as técnicas tradicionais de ITCD, como: algoritmo de máximo e mínimo local, segmentação por crescimento de região e segmentação *watershed*. Em um estudo realizado por [1], o autor fez um comparativo entre as essas três técnicas de ITCD, e seu estudo chegou a um resultado de acurácia global (A_g) próximo a 30 %, quando as técnicas foram aplicadas em imagens de alta resolução do satélite *Quickbird*. O algoritmo proposto nesta pesquisa alcançou um resultado de acurácia global de 63,8%. Em relação a métrica IOU , segundo [4], o resultado obtido é significativo, pois seu valor é superior a 0,8, ou seja, a média da sobreposição correta entre o padrão estimado e a verdade-terreno é superior a 80%

Os erros de delineação do algoritmo proposto, apresentados na Figura 6, são principalmente devido às copas de árvores de tamanho pequeno e as regiões de fronteira das copas das árvores, onde há a presença de sombras e outros elementos que possam confundir o algoritmo. Entretanto, a análise da mesma Figura 5 e das Figuras 5 e 4 e os valores das métricas A_g e IOU permitem uma avaliação positiva do algoritmo proposto.

5. CONCLUSÕES

Algoritmos de ITCD aplicados em imagens de alta resolução espacial têm recebido atenção de diversos grupos de pesquisas. Diversos métodos têm sido desenvolvidos e aplicados em diferentes áreas para auxílio na obtenção de informações, que sejam úteis para a conservação do meio ambiente e forneçam dados para a criação de inventários florestais mais abrangentes, precisos e baratos. Por estes motivos, o trabalho apresentou uma nova metodologia, baseada em *deep learning*, para realizar a delineação de copas de árvores. O algoritmo apresentado tanto detecta diferentes instâncias de árvores quanto faz a delineação de cada uma das copas. A principal inovação apresentada no algoritmo proposto é a aplicação de redes convolucionais com o objetivo de delinear as copas das árvores. Os resultados obtidos demonstram um grande potencial do método proposto, pois supera os resultados obtidos por algoritmos tradicionais que realizam a mesma tarefa. O método proposto obteve uma acurácia global de 63,8% e o valor da métrica IOU ficou em 0,812. Apesar do desempenho obtido, o algoritmo apresentou falhas na delineação de copas de árvores pequenas e nas regiões próximas as fronteira das copas, devido à presença de sombra. Estas falhas podem ser contornadas caso sejam acrescentado ao conjunto de treinamento padrões de copas de árvores pequenas e, possivelmente, pelo pré-processamento da imagem, para detectar e remover regiões sombreadas. Testes futuros serão realizados sobre regiões de florestas tropicais com alta densidade de copas.

6. AGRADECIMENTOS

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo apoio financeiro ao processo 2015/50484-0

7. REFERÊNCIAS

- [1] KE, Y.; QUACKENBUSH, L. J. A comparison of three methods for automatic tree crown detection and delineation from high spatial resolution imagery. *International Journal of Remote Sensing*, v. 32, n. 13, p. 3625-3647, 2011.
- [2] WAGNER, F. H.; et al. Individual tree crown delineation in a highly diverse tropical forest using very high resolution satellite images. *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 1, p. 1-16, 2018.
- [3] SUN, Y. et al. Developing a multi-filter convolutional neural network for semantic segmentation using high-resolution aerial imagery and lidar data. *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 143, p. 3-14, 2018.
- [4] HE, K. et al. Mask R-CNN. *Cornell University - Computer Research Repository*, v. 1, p. 1-12, 2017.