

# DETECÇÃO DE RODOVIAS RURAIS EM IMAGENS ORBITAIS ATRAVÉS DO EMPREGO DE REDES NEURAIAS CONVOLUCIONAIS

Gustavo Rota Collegio<sup>1</sup>, Antonio Gaudencio Guimarães Filho<sup>2</sup>, Aluir Porfírio Dal Poz<sup>3</sup>

<sup>1,2,3</sup> Universidade Estadual Paulista (UNESP), Faculdade de Ciências e Tecnologia, campus de Presidente Prudente – SP - FCT/UNESP, gustavo.collegio@unesp.br, guimares.filho@unesp.br, aluir.dal-poz@unesp.br.

## RESUMO

A detecção de rodovias por meio de imagens orbitais apresenta relevância significativa na comunidade científica em função das diversas aplicações que as concerne, tais como: planejamento urbano, atualização de banco de dados cartográficos etc. O método proposto se baseia em uma Rede Neural Convolucional (RNC), daqui em diante identificada como RNC/U-Net, que visa a detecção de rodovias em regiões rurais, por meio de um processo denominado segmentação semântica. A área teste usada para avaliar o método se localiza no estado do Mato Grosso. A RNC/U-Net alcançou 58,44% de *recall* e 49,65% de *precision*, com 36,26% de *Intersection-Over-Union*. Os resultados obtidos mostraram que a arquitetura é eficiente na detecção de rodovias rurais; no entanto para aquelas de caráter radiométrico e geométrico similar com outros alvos, a RNC/U-Net ainda é passível de aperfeiçoamentos e adaptações, visando contribuição direta na segmentação das rodovias.

**Palavras-chave** — Detecção de rodovias, RNC/U-Net, segmentação semântica.

## ABSTRACT

*Road detection through orbital images is extremely relevant in the scientific community due to the various applications that concern them, such as urban planning, cartographic databases updating etc. The proposed method is based on a Convolutional Neural Network (CNN), from here on identified as CNN/U-Net, that aims at detecting roads in rural regions, through a process that is known as semantic segmentation. The test area used to evaluate the proposed method is localized in Mato Grosso state. The RNC/U-Net reached 58.44% of recall and 49.65% of precision, with 36.26% of Intersection-Over-Union. The results obtained showed that the architecture was efficient in detecting rural roads, however, for those of radiometric and geometric character similar to other targets, the RNC/U-Net is still subject to improvements and adaptations, aiming at a direct contribution to the segmentation of roads.*

**Key words** — Road detection, CNN/U-Net, semantic segmentation.

## 1. INTRODUÇÃO

Em virtude das frequentes modificações presentes na malha viária, mantê-la atualizada é fundamental para diversas aplicações, tais como: sistema de navegação, planejamento urbano, atualização de informações espaciais geográficas etc. Contudo, quando são usados métodos manuais para obtenção de informações espaciais das rodovias, dados já obsoletos podem ser obtidos com alta demanda de tempo, o que requer o desenvolvimento de metodologias automáticas para tal tarefa.

Desde a década de 70 [1], uma variedade de algoritmos para detecção de rodovias surgiram. Atualmente, contextualizam-se as abordagens automáticas de detecção, visto que o fácil acesso aos dados vetoriais espaciais e metadados disponíveis, implica em adaptar algoritmos em função das evoluções tecnológicas [2]. Para isto, destacam-se as técnicas de aprendizado profundo (*Deep Learning* – DL), as quais exploram, automaticamente, as características geométricas, radiométricas, topológicas e contextuais das rodovias (padrões).

Com eficientes resultados atingidos em tarefas de classificação de objetos em diversas áreas do conhecimento [3], a inteligência artificial é, no momento presente, um dos principais tópicos de pesquisa da comunidade científica. Conseqüentemente, as Redes Neurais Convolucionais (RNCs), pertencentes à subárea de DL, são o estado da arte na tarefa de detecção de rodovias no processo de segmentação semântica [2], devido à minimização de intervenção humana e facilidade de trabalho com *big data*.

As primeiras tentativas no cenário de detecção de rodovias com a utilização de dados ópticos por meio de RNCs podem ser observadas no trabalho de [4]. Entretanto, das arquiteturas pioneiras até as desenvolvidas atualmente, estas minimamente priorizam atenção para pesquisa em rodovias rurais. Ademais, em se tratando de regiões rurais, os principais banco de dados vetoriais são limitados no tocante à atualização de informação espaciais, principalmente por grandes modificações em pequenos intervalos de tempo.

Por conseqüência, de modo a investigar a performance de RNCs em regiões rurais para detecção de rodovias, a motivação deste projeto se contextualiza. É proposto extrair informações espaciais das rodovias para detecção, de forma rápida e acurada, por meio da RNC/U-Net (processo de segmentação semântica usando a U-Net, arquitetura específica de RNC, daí a denominação RNC/U-Net). Para

isto, um conjunto de dados (imagens orbitais) é utilizado, cuja fonte permeia as características de fácil acesso e grande quantidade de informações disponíveis.

## 2. MATERIAL E MÉTODOS

### 2.1. Área de estudo e conjunto de dados

O conjunto de dados utilizado refere-se às imagens orbitais obtidas pelo satélite PlanetScope da região do estado do Mato Grosso – MT (Brasil), de resolução espacial de 3 m. Quanto à implementação e treinamento da RNC/U-Net, as IDEs Jupyter Notebook e Google Colab Pro foram utilizadas com a biblioteca gratuita e de código aberto TensorFlow/Keras. A Figura 1 apresenta a área de estudo.

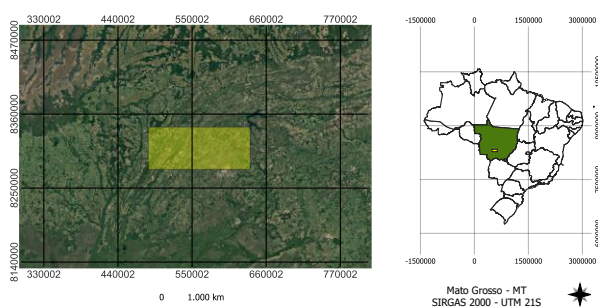


Figura 1 – Área de estudo.

Cada ortoimagem possui dimensão de 13248x9393 pixels. Para cobertura da área de estudo, 36 ortoimagens foram necessárias. A geração de verdade terrestre, por sua vez, baseou-se na base de dados vetoriais do *Open Street Map* (OSM), com rotulações manuais em áreas desatualizadas.

O pré-processamento deste conjunto de dados objetivou organizar os dados em termos de treinamento, validação e teste, isto é, com a imagem original RGB e o respectivo rótulo. Após a ortomosaicagem, as imagens foram recortadas em dimensão de 2048x2048 pixels, definidas em 70% para treinamento, 20% para validação e 10% para teste. Contudo, para introdução à RNC/U-Net, os dados foram recortados em dimensão de 512x512 pixels. Definiu-se, portanto, 1009 amostras para treinamento, 303 destinadas a validação e 151 para teste. A fim de aumentar artificialmente os dados de treinamento, por meio de rotações, espelhamento e recortes aleatórios, os recortes foram preenchidos (*mirror padding*) de modo que a dimensão de 600x600 pixels fosse obtida.

### 2.2. Arquitetura U-Net

Uma RNC, de modo geral, possui quatro princípios [3]: conexões locais, compartilhamento de pesos, sub-amostragem e o uso de muitas camadas convolucionais. Isto implica em um problema de otimização, cujos parâmetros da arquitetura são treináveis a partir de pesos e vieses.

A RNC proposta, isto é, a RNC/U-Net [5], foi desenvolvida para propósitos biomédicos de segmentação de

imagens. Trata-se de uma RNC do tipo *encoder e decoder*. Na Figura 2 são apresentadas as camadas que estruturam a RNC/U-Net.

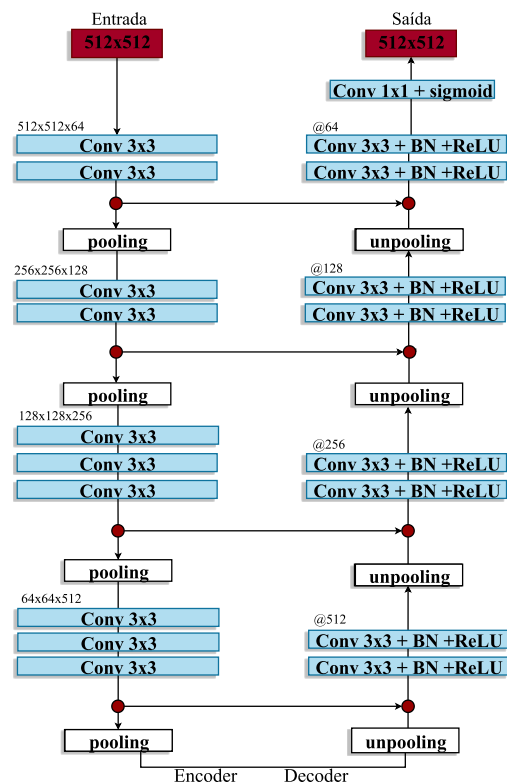


Figura 2 – RNC/U-Net para detecção de rodovias.

Por padrão, a RNC/U-Net utiliza como classificador o VGG-16 [6]. O termo *encoder* define-se como o processo de extração de informações, cujos mapas de características são obtidos por meio das camadas convolucionais e de *max pooling*, o qual reduz a dimensão espacial desses mapas gradativamente, ampliando assim, o campo receptivo local da rede. Para cada sub-amostragem, o número de filtros do mapa de característica é duplicado.

Já o processo de *decoder*, define-se por meio de camadas que realizam a super-amostragem (*unpooling*). Realiza-se, portanto, convoluções com *batch normalization* (BN) [7] e função de ativação ReLU [8], necessitando de um método de interpolação para a reamostragem. Deste modo, os mapas de características obtidos são concatenados e classificados por meio da função de ativação *sigmoid*.

Os processos mencionados acima se resumem na extração de características das vias (informações semânticas), para que tais pixels sejam detectados e agrupados em uma mesma classe (classificação binária). Neste sentido, para aprendizado da RNC/U-Net, os dados de treinamento são utilizados. Para supervisão do modelo e atualização interna de parâmetros, empregam-se os dados de validação. Para análise de qualidade da inferência via RNC/U-Net, o conjunto de dados teste é usado.

### 2.2.1. Função perda

O treinamento de uma RNC consiste basicamente na otimização global de uma função perda do modelo, a fim de que pesos e vieses sejam iterativamente atualizados via algoritmo de *backpropagation*. Duas convencionais funções perdas são utilizadas: a *dice loss* [9] e a *focal tversky loss* [10]. Conforme reportam os autores, a primeira proposita maximizar a sobreposição de regiões de referência com a de predição; a segunda contextualiza-se no desbalanceamento de classes, penalizando pixels que ocorrem com maior frequência.

### 2.3. Análise de qualidade da inferência

As métricas de qualidade, definidas a seguir, são utilizadas para avaliar os resultados da inferência. As métricas *recall* (*R*), *precision* (*P*), *F1-Score* e *Intersection-Over-Union* (*IoU*) são dadas pelas Equações 1,2,3 e 4, respectivamente.

$$R = \frac{TP}{TP + FN} \quad (1)$$

$$P = \frac{TP}{TP + FP} \quad (2)$$

$$F1_{Score} = \frac{2TP}{2TP + FP + FN} \quad (3)$$

$$IoU = \frac{TP}{TP + FP + FN} \quad (4)$$

Onde, TP são verdadeiros positivos, TN são verdadeiros negativos, FP são falsos positivos e FN, falsos negativos.

## 3. RESULTADOS E DISCUSSÃO

Para realização dos experimentos com os dados de teste, combinações de hiperparâmetros para o treinamento da RNC/U-Net foram testadas. Os principais são: determinação da taxa de aprendizado, quantidade de épocas e *batch*, otimizador e *dropout*.

Na Tabela 1 apresentam-se duas comparações. A primeira refere-se ao treinamento com taxa de aprendizado de  $10^{-2}$  e função perda *dice loss*; a segunda com a mesma taxa de aprendizado, no entanto com a função perda *focal tversky loss*. Para ambas, 150 épocas foram definidas com subamostras igual a 4 (*batch*), o otimizador *Stochastic Gradient Descendent* (SGD), e o hiperparâmetro de interrupção *early stopping* para 40 épocas.

| Modelo*        | Recall       | Precision    | F1           | IoU          |
|----------------|--------------|--------------|--------------|--------------|
| RNC/U-Net-ftl  | <b>61,68</b> | 35,56        | 41,59        | 25,09        |
| RNC/U-Net-dice | 52,14        | <b>37,41</b> | <b>42,44</b> | <b>25,59</b> |

\*ftl: *focal tversky loss* e dice: *dice loss*.

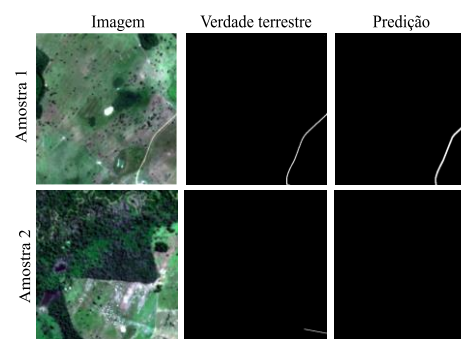
**Tabela 1. Métrica da classificação.**

Evidencia-se por meio das métricas obtidas que a escolha

de uma função perda impacta diretamente na predição e análise de qualidade de um classificador. A função perda *focal tversky loss*, cujo objetivo é o de penalizar falsos positivos, atinge maior valor para a métrica *recall*. A função *dice loss*, em contrapartida, atinge menor valor para esta mesma métrica, entretanto, para *precision*, *F1-Score* e *IoU*, a análise de qualidade é similar.

| Modelo    | Recall | Precision | F1    | IoU   |
|-----------|--------|-----------|-------|-------|
| Amostra 1 | 98,63  | 54,01     | 69,88 | 53,61 |
| Amostra 2 | 0      | 100       | 0     | 0     |

**Tabela 2. Análise quantitativa de amostras individuais.**



**Figura 3 – Amostras analisadas individualmente.**

Ao verificar predições individuais, como apresentam a Tabela 2 e Figura 3, investigou-se que o desbalanceamento de classes influencia na métrica da predição. Como exemplo, a amostra 1 enfatiza a predição de melhor qualidade, enquanto a amostra 2 expressa que para as cenas com o mínimo de informação de rodovias, o modelo computa 100% de *precision* e valores nulos para as demais métricas. Por consequência, a métrica é diretamente afetada. Eliminar amostras desbalanceadas de um conjunto de dados é uma alternativa para obter-se análises mais efetivas da classificação.

Sob este aspecto, procurou-se eliminar as amostras com o mínimo de informação de rodovias do conjunto de dados de teste e adotar uma tolerância em pixels de borda das rodovias para computação das métricas [11]. Isto se dá em função das verdades terrestres (geradas via OSM ou manualmente) estarem ligeiramente deslocadas em relação as rodovias da imagem original. Portanto, em concordância com [11], 3 pixels de tolerância a partir das bordas foram considerados como verdadeiro positivo. O resultado é apresentado na Tabela 3 e Figura 4.

| Modelo*        | Recall       | Precision    | F1           | IoU          |
|----------------|--------------|--------------|--------------|--------------|
| RNC/U-Net-ftl  | <b>58,44</b> | 49,65        | <b>51,01</b> | <b>36,26</b> |
| RNC/U-Net-dice | 47,45        | <b>54,58</b> | 47,90        | 33,74        |

\*ftl: *focal tversky loss* e dice: *dice loss*.

**Tabela 3. Análise quantitativa final.**

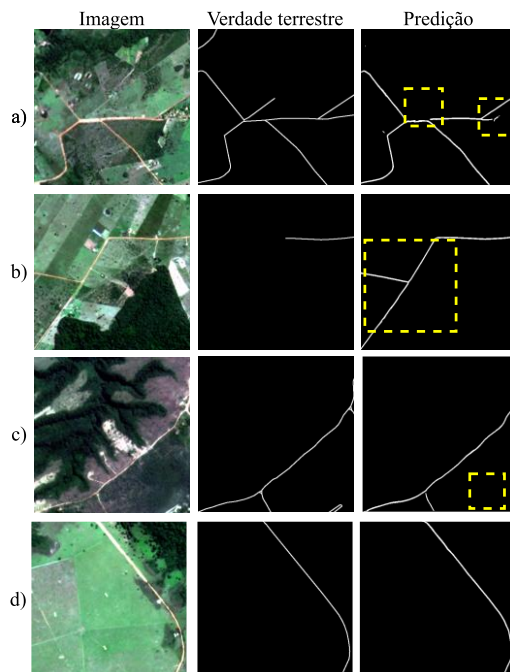


Figura 4 – Resultados da predição com a RNC/U-Net.

Conforme apresenta a Tabela 3, a métrica IoU passa agora a ter um incremento de aproximadamente 10% em termos de segmentação semântica. Os valores de *precision* são mais robustos, atingindo cerca de 50% (aumento de 12 a 15%) e significando que os pixels detectados estão mais coerentes em relação a sua verdade terrestre. Quanto ao *recall*, mantém-se a perspectiva de boa classificação na classe dos positivos.

A Figura 4 também destaca em amarelo os falsos positivos da classificação, além de erros grosseiro de rotulação das amostras (Figura 4 (b)). Por conseguinte, observa-se que o modelo mantém a continuidade e delineamento das rodovias na detecção, entretanto, eventualmente ativando características de outros alvos de aspectos geométricos e/ou radiométricos similares na classificação das rodovias.

## 5. CONCLUSÕES

O trabalho apresentou um experimento para detecção automática de rodovias rurais a partir da abordagem de RNCs por meio de imagens orbitais.

A implementação e treinamento da RNC/U-Net mostram que a detecção de rodovias depende do contexto na cena. Enfatiza-se robustez, portanto, às rodovias pavimentadas ou aquelas não pavimentadas cuja identificação é ressaltada no contexto rural. Isto implica em considerar que a RNC/U-Net ainda é vulnerável na classificação com alvos de características similares ou por complexos contextos. Além disso, o desbalanceamento de classes é capaz de influenciar negativamente no controle de qualidade da segmentação ou comportar-se de diferentes maneiras.

Ao passo que a U-Net é uma arquitetura de fácil adaptação, a escolha de outro classificador na RNC (*backbone*), a fim de se obter maior aprendizado de características para segmentação das rodovias são recomendações futuras para continuidade desta pesquisa.

## 6. AGRADECIMENTOS

Ao CNPq/PIBIC pela bolsa de Iniciação Científica (IC) concedida, e à FAPESP (Processo no. 2021/03586-2) pelos recursos computacionais, os quais foram de grande valia para o desenvolvimento desta pesquisa de IC.

## 7. REFERÊNCIAS

- [1] R. Bajcsy and M. Tavakoli, "Computer Recognition of Roads from Satellite Pictures," *IEEE Trans. Syst. Man. Cybern.*, vol. SMC-6, no. 9, pp. 623–637, Sep. 1976, doi: 10.1109/TSMC.1976.4309568.
- [2] A. G. G. Filho, "Deep Convolutional Neural Network (DCNN) para extração automática de rede viária a partir da FUSÃO de dados varredura LASER aerotransportado e imagens de altíssima resolução em ambiente urbano," Programa de Pós-Graduação em Ciências Cartográficas - FCT/UNESP - Campus de Presidente Prudente, 2022. doi: 10.17616/R31N39.
- [3] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nat.* 2015 5217553, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.
- [4] V. Mnih and G. E. Hinton, "Learning to Detect Roads in High-Resolution Aerial Images," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 6316 LNCS, no. PART 6, pp. 210–223, 2010, doi: 10.1007/978-3-642-15567-3\_16.
- [5] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *arXiv1505.04597 [cs]*, Sep. 2015, [Online]. Available: <http://arxiv.org/abs/1505.04597>
- [6] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, Sep. 2014, doi: 10.48550/arxiv.1409.1556.
- [7] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," *32nd Int. Conf. Mach. Learn. ICML 2015*, vol. 1, pp. 448–456, Feb. 2015, doi: 10.48550/arxiv.1502.03167.
- [8] X. Glorot, A. Bordes, and Y. Bengio, "Deep Sparse Rectifier Neural Networks," *undefined*, 2011.
- [9] F. Milletari, N. Navab, and S. A. Ahmadi, "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," *Proc. - 2016 4th Int. Conf. 3D Vision, 3DV 2016*, pp. 565–571, Jun. 2016, doi: 10.48550/arxiv.1606.04797.
- [10] N. Abraham and N. M. Khan, "A Novel Focal Tversky loss function with improved Attention U-Net for lesion segmentation," *Proc. - Int. Symp. Biomed. Imaging*, vol. 2019-April, pp. 683–687, Oct. 2018, doi: 10.48550/arxiv.1810.07842.
- [11] V. Mnih, "Machine Learning for Aerial Image Labeling," 2013.