

Métodos de segmentação em regiões com erosão e solo exposto: análise comparativa entre modelos *Deep Learning*

Jhon J. Majin* Carlos A. Alves* Gustavo L. Mourao*
Jorge E. B. Caceres* Jose A. D. Salazar* Paulo H. M. Piratelo*
Flávio G. O. Barbosa* Carlos A. M. Nascimento**
Lucas M. C. Souza** Antonio Donadon**

* Instituto SENAI de Inovação em Sistemas Embarcados, SC, (e-mail: jhon.erazo@sc.senai.br, carlos.antonio@sc.senai.br, gustavo.leao@sc.senai.br, jorge.caceres@sc.senai.br, jose.salazar@sc.senai.br, paulo.piratelo@sc.senai.br, flavio.barbosa@sc.senai.br).

** Companhia Energética de Minas Gerais (e-mail: caxandre@cemig.com.br, lucas.csouza@cemig.com.br, antonio.donadon@berkan.com.br)

Abstract: Soil erosion is considered one of the greatest natural risks because it impacts various economic sectors worldwide, causing damage especially in the civil and electrical sectors. The high costs of maintaining infrastructures affected by this type of phenomenon, as well as the blocking of different tasks in areas affected by the appearance of these artifacts, are widely documented in the literature. In this sense, it is essential to identify and monitor the appearance of large-scale erosion to minimize economic losses. Recently, there has been the emergence of a range of works that employ the use of convolutional neural networks (CNNs) and satellite images to monitor different instances on earth. However, the literature still lacks work that makes a more in-depth comparison of different CNN architectures in detecting erosion and exposed soil in satellite images. From this perspective, this work aims to evaluate and compare the performance of four semantic segmentation architectures based on CNNs: U-Net, Mask RCNN, RED-CNN and Googlenet. The experimental results carried out on the reference dataset show that the U-Net and Mask RCNN models achieved the best results, with accuracy values of 98% and 92% respectively. In the IoU metric, these models obtained the highest values with an average value of 44%.

Resumo:

A erosão do solo é considerada um dos grandes riscos naturais por afetar diferentes aspectos econômicos em nível mundial, gerando danos especialmente no setor civil e elétrico. Os custos elevados na manutenção de infraestruturas afetadas por esse tipo de fenômeno, bem como o bloqueio de execução de diferentes tarefas em áreas afetadas pelo surgimento desses artefatos são amplamente documentados na literatura. Nesse sentido, é fundamental identificar e monitorar o aparecimento de erosões em grande escala para minimizar perdas econômicas. Recentemente, observa-se o surgimento de uma gama de trabalhos que empregam o uso de redes neurais convolucionais (CNNs) e imagens satélites para o monitoramento de diferentes instâncias na terra. No entanto, a literatura ainda carece de trabalhos que fazem uma comparação mais aprofundada de diferentes arquiteturas de CNNs na detecção de erosão e solo exposto em imagens satelitais. Nesta perspectiva, este trabalho tem como objetivo avaliar e comparar o desempenho de quatro arquiteturas de segmentação semântica baseadas em CNNs: U-Net, Mask RCNN, RED-CNN e Googlenet. Os resultados experimentais realizados no conjunto de dados de referência mostram que os modelos U-Net e Mask RCNN alcançaram os melhores resultados, com valores de acurácia de 98% e 92% respectivamente. Na métrica IoU, estes modelos obtiveram os valores mais elevados com um valor médio de 44%.

Keywords: Semantic Segmentation; Erosion Detection; Convolutional Neural Networks; Satellite Imagery; Remote Sensing.

Palavras-chaves: Segmentação Semântica; Detecção de Erosão; Redes Neurais Convolucionais; Imagem de satélite; Sensoriamento remoto.

1. INTRODUÇÃO

A erosão é um processo natural em que a camada superficial do solo é desgastada e removida por agentes como água, vento ou atividade humana. Esse fenômeno pode ocorrer de forma gradual ou acelerada, dependendo de fatores como a inclinação do terreno, a cobertura vegetal e o tipo de solo Wang et al. (2023). Práticas inadequadas de manejo, como desmatamento, cultivo intensivo e construção desordenada, podem agravar o problema, acelerando a erosão e resultando em assoreamento de rios e desertificação. Além disso, a urbanização e o crescimento populacional exacerbam a perda de vegetação e o deslizamento de terras, obrigando a necessidade de adotar estratégias de gestão do solo a longo prazo e de práticas de gestão eficazes para combater a degradação ambiental e manter a produtividade do solo.

O sensoriamento remoto tem sido amplamente utilizado para o monitoramento da erosão, devido às suas vantagens significativas, como a capacidade de cobrir vastas áreas geográficas e a obtenção periódica de dados. As imagens de satélite, por exemplo, oferecem uma visão abrangente e detalhada das mudanças na superfície terrestre, permitindo a identificação contínua de áreas suscetíveis à erosão. Além disso, os diferentes tipos de sensores utilizados na captura dessas imagens podem fornecer dados sobre a topografia, uso da terra, cobertura vegetal, umidade do solo e outros fatores que influenciam o processo de erosão. A combinação dessas informações possibilita uma análise integrada e uma compreensão mais profunda dos fatores que contribuem para a degradação do solo.

Nos últimos anos, modelos baseados em *Machine Learning* (ML) têm sido amplamente empregados em estudos ambientais e de riscos naturais. Esses modelos se mostraram bastante eficazes na modelagem espacial, permitindo a extração de *insights* e padrões valiosos a partir de dados geoespaciais complexos Rahaman et al. (2023). Em particular, a utilização de modelos baseados em CNNs (do inglês *Convolutional Neural Network*, ou Redes Neurais Convolucionais) tem ganhado destaque em diversos campos devido à sua eficácia na resolução de problemas complexos envolvendo conjuntos de dados geoespaciais. No entanto, apesar do sucesso das CNNs em outras aplicações de modelagem ambiental como uso da terra, dinâmica da cobertura da terra ou dinâmica da cobertura vegetal, há uma carência significativa de pesquisas que avaliem a eficácia das CNNs na detecção ou identificação de erosão do solo por meio de imagens de satélite.

Com base na revisão sistemática da literatura (RSL) realizada para este documento, destacam-se no estado da arte os estudos apresentados por Gafurov and Yermolaev (2020), Gafurov (2022) e Nogueira et al. (2020). Esses estudos implementaram diferentes arquiteturas e abordagens de CNNs diretamente em imagens de satélite para identificar e/ou segmentar erosões. Embora os resultados

obtidos sejam amplamente positivos, é importante destacar a necessidade de uma exploração mais extensa desses tipos de métodos. Desta forma, o principal objetivo deste trabalho é explorar e comparar a eficiência de quatro arquiteturas clássicas de CNNs da literatura no processo de segmentação de erosões utilizando imagens satelitais. Para atingir esse objetivo, foi criada uma base de dados satelitais de alta resolução a qual teve uso igualitário entre todos os modelos. Especificamente, tratamos a identificação de erosão como uma tarefa de segmentação semântica, conhecida como classificação de pixels. Foram escolhidas as arquiteturas: GoogLeNet (Szegedy et al. (2014)), MASK RCNN (He et al. (2017)), RED-CNN (Chen et al. (2017)) e U-Net (Ronneberger et al. (2015)), as quais já possuem amplo uso e aplicação em problemas de segmentação.

A seguir, são apresentadas as contribuições deste estudo:

- Apresentação de um novo conjunto de dados que contempla registros de regiões com erosão de diferentes tamanhos e texturas. O conjunto de dados é um ponto de partida para novas pesquisas na identificação de erosão com imagens satelitais.
- Um novo *pipeline* para a subdivisão (recorte) das imagens satelitais em *patches* com diferentes tamanhos $[H \times W \times C]$.
- Análises e identificação dos modelos baseados em CNNs que apresentam um melhor comportamento no reconhecimento de erosão e solo exposto usando o *dataset* de referência.

Essas contribuições estabelecem um avanço significativo no domínio do monitoramento do solo por meio de imagens satelitais usando técnicas de visão computacional. A contribuição apresentada pode auxiliar no desenvolvimento de aplicações inovadoras de sensoriamento remoto. Esse conhecimento pode ser usado pelos agentes especiais para definir prioridades e propor melhores planos de ação.

O artigo está organizado da seguinte forma: Na Seção 1 é feita a introdução ao tema, na Seção 2 se apresenta alguns dos trabalhos encontrados no estado da arte. A Seção 3 fornece os detalhes dos métodos, implementações e o procedimento experimental. A Seção 4 apresenta os resultados e discussões dos modelos de DL comparados. Por fim, a Seção 5 apresenta as conclusões e considerações finais.

2. TRABALHOS RELACIONADOS

Foi realizada uma RSL dos últimos 10 anos, focada em estudos voltados para a identificação de erosão do solo. Como resultado, observou-se que a detecção de erosão tem sido abordada por quatro principais métodos: métodos estatísticos, ML, mineração de dados e *multi-criteria decision analysis*. Recentemente, os estudos que utilizam imagens de satélite para a detecção de erosão concentraram-se em duas áreas principais: (1) métodos de ML e (2) métodos de DL, como as CNNs. A seguir, são apresentados alguns

dos trabalhos mais recentes do estado da arte que utilizam esses dois enfoques, conforme a RSL¹.

Em Makaya et al. (2019) utilizou-se o algoritmo SVM (*Support Machine Vector*) em imagens do satélite *Sentinel-2* SMI para analisar os possíveis fatores ambientais que contribuem para o início e/ou desenvolvimento da erosão em ravinas (voçorocas). Schellekens and Amani (2022) propuseram uma solução para o monitoramento da erosão em dunas costeiras e aterros através de análises e aplicação do SVM em imagens satelitais para classificar três categorias de *pixels*: areia, terra e água. Foram utilizadas imagens ópticas adquiridas pelo *Landsat-5* e *Landsat-8* para mapear e interpretar a erosão na área costeira.

Ainda em 2022, Liu et al. (2022) desenvolveram uma abordagem em larga escala para o mapeamento de erosão e a degradação do solo em imagens do satélite *Sentinel-2*, as quais foram usadas para treinar o algoritmo *Random Forests* (RF) no processo de detecção de erosão. Posteriormente, utilizou-se o *Digital elevation models* para extrair informação da topografia das áreas analisadas, de forma a minimizar erros na saída da predição. O uso de RF para detecção de erosões não era novidade, visto que em 2019, um trabalho semelhante foi apresentado por Vâgen and Winowiecki (2019) onde o RF foi treinado mediante a fusão de dados obtidos por sensores em *loco* e imagens obtidas pelo satélite *MODIS*.

Na literatura, foram encontrados também diferentes estudos que visavam comparar algoritmos de ML para identificar erosões em solo ou ravinas. Um exemplo pode ser visto em Phinzi et al. (2020), onde foi feita uma comparação entre os algoritmos *Linear Discriminant Analysis*, SVM e RF nas tarefas de detecção de erosão do solo e detecção de ravinas. As imagens para o treinamento dos modelos foram obtidas através do satélite *SPOT-7*, utilizando-se das bandas multi-espectrais RGB, NIR e Pancromática. No trabalho Phinzi et al. (2021) foi realizada uma comparação entre diferentes classificadores de ML para a distribuição e identificação de voçorocas, incluindo: *k-dimensional tree*, *K-Nearest Neighbor*, *Minimum Distance*, *Maximum Likelihood* e RF. Os modelos foram treinados utilizando-se de imagens do *Shuttle Radar Topography Mission* com as bandas RGB. No entanto, o trabalho realça algumas limitações na detecção de erosão em áreas cobertas de vegetação, limitando o funcionamento da abordagem apenas para regiões áridas ou semiáridas.

Importante observar que um trabalho semelhante já havia sido apresentado em 2019 por Bui et al. (2019) onde, assim como o trabalho apresentado em Phinzi et al. (2021), buscava-se mapear erosão em barrancos usando diferentes algoritmos de ML, tais como: RF, SVM, *Logistic Regression* e *Naïve Bayes Multinomial Updatable*.

Saindo um pouco da comparação entre classificadores, encontra-se também alguns trabalhos cuja proposta está ligada ao uso de DL para detecção de erosão do solo. Nesse sentido, os autores Gafurov and Yermolaev (2020) abordaram o problema de detecção automatizada de erosão de ravinas utilizando-se de uma técnica baseada em CNNs. Os

dados usados neste estudo foram obtidos da empresa *DigitalGlobe*, que disponibiliza imagens satelitais em diferentes resoluções e composições espectrais. Os autores usaram o modelo de segmentação semântica U-Net com as bandas RGB do satélite na criação dos conjuntos de dados para treinamento e teste. O conjunto de dados caracterizou-se por conter aproximadamente 1000 *patches* com dimensões de 512×512 *pixels*.

Os autores do trabalho Gafurov and Yermolaev (2020) continuaram desenvolvendo experimentos com CNNs. Em 2022 propuseram uma combinação entre as redes profundas *LinkNet* e *DenseNet* para resolver o problema da detecção de erosão em riachos (*rill erosion*). Os detalhes apresentados por Gafurov (2022) descrevem que o conjunto de dados foi criado com base no satélite *Sentinel-2* MSI sendo utilizada a banda NIR (*Nearinfrared*) com uma resolução de 10 m. As imagens satelitais foram divididas em 10.933 *patches* com dimensões de 256×256 *pixels*. Segundo os autores, a predição de erosão em riachos tem uma alta porcentagem de acerto no reconhecimento, discriminando detecções de ravinas ou caminhos de terra.

Por fim, é novamente de suma importância citar um trabalho desenvolvido visando experimentos no Brasil. Para isso, Nogueira et al. (2020) confeccionaram um *framework* baseado em DL capaz de identificar erosões em linhas ferroviárias no Brasil. Para tal, foram avaliados seis modelos de segmentação, buscando selecionar aquele que apresentasse uma solução mais robusta e confiável. Os modelos avaliados foram: *Pixelwise*, *Fully Convolutional Network* (FCN), *Deconvolution Network*, *DeepLab*, *Mask RCNN*, *Dynamic Dilated ConvNet* (DDCN). Para os experimentos, foi construído um conjunto de dados utilizando as bandas RGB e NIR do satélite *Plêiades*, cuja resolução era de 0,5 m. As imagens foram divididas em *patches* com um tamanho de 448×448 *pixels*, resultando um total 2000 *patches*.

Com base nos trabalhos mencionados anteriormente, observa-se um aumento na utilização de diferentes combinações de algoritmos de DL para detectar erosão. No entanto, é importante destacar que ainda há uma escassez de estudos no estado da arte que empreguem CNNs para a identificação de erosão e solo exposto utilizando imagens de satélite. Isso aponta para uma oportunidade promissora de conduzir novos estudos e realizar descobertas significativas na área.

3. MATERIAIS E MÉTODOS

Esta seção apresenta os materiais e métodos utilizados para a confecção deste trabalho. Inicialmente, na seção 3.1 realiza-se uma contextualização das arquiteturas baseadas em CNNs para a detecção de erosão e solo exposto utilizadas em este estudo. Posteriormente, na Seção 3.2 se apresenta a metodologia empregada para a criação do conjunto de dados e o *pipeline* proposto. Finalmente, a seção 3.3 mostra as métricas utilizadas para a avaliação dos modelos.

3.1 Arquiteturas

Foram escolhidas as arquiteturas U-Net, Mask R-CNN, GoogleNet e RED-CNN devido às suas capacidades superiores em segmentação de objetos em diferentes contex-

¹ É importante destacar que os trabalhos mais recentes voltados para a identificação de erosão utilizando imagens de satélite foram publicados entre os anos de 2019 e 2022.

tos, especialmente na identificação de objetos pequenos, e pela sua habilidade em generalizar adequadamente mesmo quando o conjunto de dados de treinamento é limitado

(1) Googlenet:

A Googlenet é uma arquitetura profunda que extrai características ricas e variadas, essenciais para a segmentação em contextos complexos. Este modelo se destaca por conter diversos módulos de convoluções chamados “módulos *Inception*”. Esses módulos foram projetados para realizar convoluções em várias escalas (diferentes filtros) simultaneamente e ao mesmo tempo concatenar as saídas dessas operações. Esse processo leva a que a rede aumente a robustez no reconhecimento de características complexas nas imagens por permitir uma exploração maior uma vez que múltiplos tamanhos de filtros são utilizados nesses blocos (veja a Figura 1).

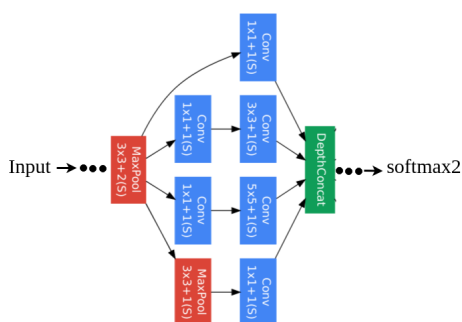


Figura 1. Segmento da arquitetura do modelo GoogleNet Szegedy et al. (2014).

Devido ao seu característico uso dos blocos de *Inception* a arquitetura se tornou referência no uso desse tipo de artefato tornando-se um *baseline* da literatura e um importante modelo para experimentos de comparação de diferentes redes.

(2) RED-CNN (*Residual Encoder-Decoder-CNN*):

A RED-CNN é uma CNN que utiliza os princípios dos blocos residuais utilizados na arquitetura ResNet (Targ et al. (2016)) com uma estrutura de codificador-decodificador para tarefas de segmentação semântica. Estas redes, são formadas por dois segmentos: Codificador - Parte da rede responsável por extrair características da imagem principal e codificá-las em *features* descritores; Decodificador - Parte da rede responsável por decodificar informações anteriormente codificadas pelo segmento codificador. Isso permite a identificação de objetos pequenos, melhorando a capacidade do modelo de aprender representações detalhadas.

A arquitetura clássica conta com 5 camadas convolucionais (codificador) e 5 camadas de deconvolução (decodificador). Nos últimos anos este modelo tem sido utilizado em diversas aplicações, como: classificação de doenças pulmonares e segmentação de imagens cerebrais. Na revisão do estado da arte concluída não foram encontrados experimentos relacionados à segmentação de erosões utilizando da RED-CNN.

Com base na Figura 2, a arquitetura clássica conta com 5 camadas convolucionais (codificador) e 5 camadas de deconvolução (decodificador) e não é utilizado nenhum tipo de *pooling*. A rede ainda tem como caracte-

terística o uso de conexões residuais, que melhoram a convergência do resultado final.

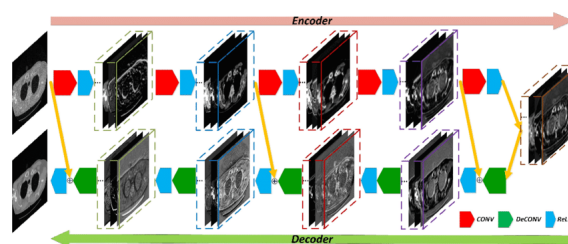


Figura 2. Arquitetura clássica da RED-CNN Chen et al. (2017).

(3) U-Net:

A U-Net é uma CNN que se baseia no conceito do tipo codificador-decodificador, que permite uma segmentação precisa, particularmente eficaz em identificar pequenos objetos com alta resolução. O modelo foi projetado inicialmente para análises de imagens médicas e posteriormente pelos bons resultados tem sido aplicada em outros domínios de visão computacional, especialmente para análises de imagens satelitais. Arquitetura possui um característico formato de “U” devido aos segmentos de codificação e decodificação aliados ao uso de *pooling*, onde recebe seu nome. A Figura 3, apresenta a arquitetura clássica da U-Net

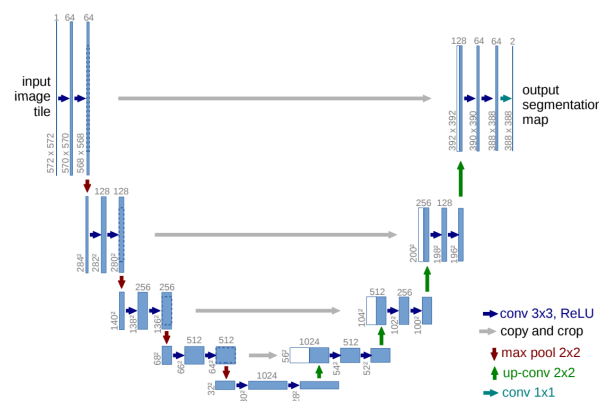


Figura 3. Arquitetura da U-Net Ronneberger et al. (2015).

(4) Mask R-CNN:

Mask R-CNN é um modelo construído para detecção de objetos e tarefas de segmentação semântica em diferentes contextos, sendo ideal para para aplicações onde a precisão na segmentação de pequenos objetos é crucial. O modelo normalmente utiliza um *backbone* pré-treinado, como ResNet para extrair as características das imagens de entrada em diferentes níveis de abstração. O foco principal é a incorporação de uma ramificação adicional na rede que é responsável por gerar máscaras precisas para cada objeto detectado, o que permite separar diferentes instâncias do mesmo tipo de objetos (detecção e segmentação).

A Figura 4 apresenta a arquitetura do modelo Mask R-CNN. O modelo recebe uma imagem (tensor) como entrada e após as camadas convolucionais, são obtidas duas saídas, uma referente à detecção de objetos e outra à segmentação de instâncias.

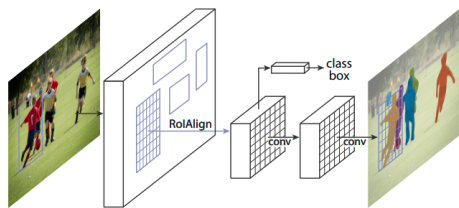


Figura 4. Arquitetura da Mask R-CNN He et al. (2017).

3.2 Descrição do conjunto de dados

Para este estudo foi utilizada uma sequência de imagens multiespectrais capturadas e formadas pelos satélites Planet Scope (BaseMaps)². Essas imagens foram obtidas em diferentes regiões do estado de Minas Gerais no Brasil. Uma característica importante de cada *tiles* é a subdivisão de diversas regiões com geometrias distintas, nas quais há presença de vegetação, lavouras, áreas residenciais, estradas de terra, solo exposto e erosão. No total, a combinação das *tiles* abrange mais de 95.555.600 hectares no estado de Minas Gerais.

As imagens de satélite foram coletadas em diferentes períodos do ano (inverno e verão) entre os anos de 2018 e 2023. O objetivo de utilizar imagens capturadas em diferentes momentos é analisar o impacto da série temporal no aprendizado do modelo considerando diferentes fases da erosão. O conjunto de imagens foi amostrado em regiões e períodos previamente reportados por ocorrência de erosão, em sua maioria. A Figura 5 apresenta a localização dos *Tiles* utilizados no estado de Minas Gerais.

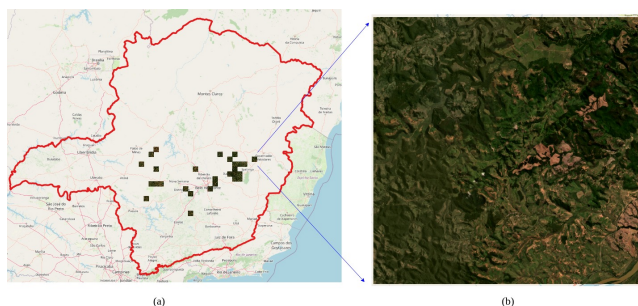


Figura 5. Localização dos *tiles* no estado de Minas Gerais. (a) Representação de Minas Gerais; (b) Imagem RGB do tile.

Foram utilizadas 28 imagens de satélite do Planet Scope com dimensões de 4096×4096 *pixels*. Nesse sentido, devido ao grande volume de dados, as imagens foram transformadas em um conjunto de tensores ou *patches*. Para isso foi proposto um *pipeline* apresentado na Figura 6 composto por três etapas: Inicialmente é criado um arquivo GeoJSON que contém as coordenadas georreferenciadas dos pontos da grade onde as imagens serão recortadas. Posteriormente, aplica-se o algoritmo que relaciona as bandas satelitais (*tiles*) ao arquivo GeoJSON; mantendo as características de localização espacial de cada *patch* em relação às coordenadas específicas orientadas aos *Coordinate Reference System*. Finalmente, é aplicado o algoritmo de recorte, onde são criados tensores das imagens com

² <https://www.planet.com/basemaps>. Acesso em: 25/03/2024

dimensões de $[H, W, N]$; sendo H e W a altura e largura do tensor, e N o número de canais (R,G,B, NIR, etc).

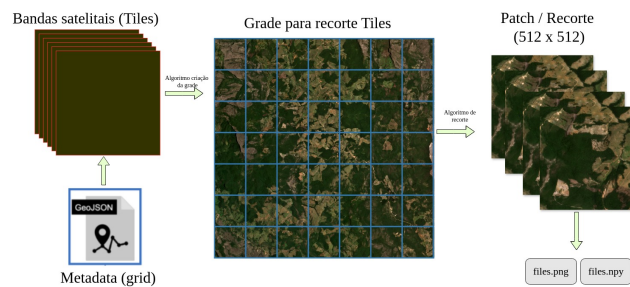


Figura 6. *Pipeline* para processamento e conformação dos *tiles* em *patches*.

A Tabela 1 mostra o conjunto de dados criado a partir do resultado do recorte das imagens de satélite com o *pipeline* proposto. O conjunto de dados consiste em 749 *patches* usando as componentes espectrais R,G,B; cada um com um tamanho de $[512 \times 512 \times 3]$. A resolução das imagens é de 5 metros/pixel normalizada para os valores (0, 255).

Tabela 1. Informações do conjunto de dados criado a partir da constelação PlanetScope.

Bandas	Número tiles	Número patches	Total labels	Tamanho label min/max
RGB	28	749 [512 x 512 x 3]	4000	40/733

Foi realizado um processo de anotação manual inspecionando cada uma das imagens RGB onde havia presença de regiões com erosão³. A Figura 7 apresenta um exemplo de alguns dos *labels* criados para o conjunto de dados, contendo as classes: erosão (em branco) e *background* (preto).

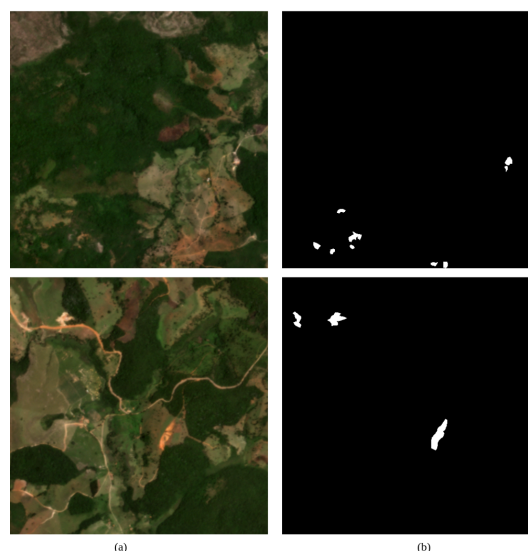


Figura 7. Exemplo dos *labels* criados para identificação de erosão. (a) Imagem RGB;(b) Máscara binária.

³ Este processo foi realizado por um especialista na área de monitoramento de solos.

3.3 Métricas de avaliação

As métricas precisão, acurácia, F1-Score e IoU foram utilizadas para avaliar e comparar o desempenho dos modelos de segmentação semântica. A precisão refere-se à porcentagem de previsões que são instâncias relevantes; ou seja, permite calcular a proporção de *pixels* classificados corretamente dentro das classes em relação ao total de *pixels* da imagem. A acurácia representa a relação de proximidade entre o conjunto de *pixels* de referência considerando a máscara predita. Por outro lado, a métrica F1-Score busca avaliar a relação de similaridade entre as duas máscaras.

O IoU é uma das métricas mais importantes utilizadas na avaliação de modelos de segmentação. O IoU avalia a precisão da segmentação com base na interceptação da máscara predita e a máscara de referência, ao mesmo tempo avalia se as classes estão classificadas corretamente e se a localização espacial das previsões coincide com a localização real dos objetos em a imagem. Matematicamente, a definição de precisão, acurácia, F1-score e IoU são mostrados nas equações (1) a (4), respectivamente.

$$\text{precisão} = \frac{TP}{TP + FP}, \quad (1)$$

$$\text{acurácia} = \frac{TP + TN}{TP + TN + FP + FN}, \quad (2)$$

$$\text{F1-Score} = 2 \times \frac{GT \cap PD}{GT \cup PD}. \quad (3)$$

$$\text{IoU (Intersection of Union)} = \frac{GT \cap PD}{GT \cup PD}, \quad (4)$$

Onde, TP (Verdadeiro Positivo, ou *True Positive*), FP (Falso Positivo, ou *False Positive*) e FN (Falso Negativo, ou *False Negative*) representa detecções corretas, detecções incorretas e detecções perdidas, respectivamente. PD (máscara predita) e GT (máscara de referência, ou *Ground Truth*).

4. RESULTADOS EXPERIMENTAIS

Esta Seção apresenta a análise dos experimentos realizados neste estudo. A seção 4.1 apresenta os resultados alcançados pelos modelos de segmentação utilizando o conjunto de dados de referência. A Seção 4.2 tem como objetivo apresentar uma análise qualitativa dos modelos em imagens de teste. Todos os modelos foram implementados e treinados em um HPC (*High-Performance Computing*) com um sistema operacional Linux Ubuntu 18.04, placa de vídeo DGX Nvidia A100 de 32 GB e 1 TB de RAM. Entre os pacotes de software utilizados estão Python 3.8, TensorFlow 2.1, PyTorch 1.5 e CUDA 11.2.

4.1 Análise quantitativa

Um dos principais objetivos deste documento é a identificação de um modelo de segmentação semântica que apresente um alto desempenho na identificação de erosão em imagens satelitais. Para isso, foram comparados os seguintes modelos: U-Net, Mask RCNN, GoogleNET, RED CNN.

Os modelos de aprendizagem profunda geralmente requerem uma quantidade suficiente de dados de treinamento para alcançar uma boa generalização nas instâncias a serem reconhecidas. Normalmente a técnica de *transfer learning* é usada para treinar os modelos em um novo conjunto de dados; reduzindo assim o tempo de treinamento e evitando *overfitting*. Nesse sentido, os pesos do Imagenet foram utilizados para iniciar o processo de treinamento. Além disso, foi aplicado aumento de dados, incluindo técnicas como inversão vertical, inversão horizontal, variações de escala e alterações de brilho.

Os conjuntos de dados para treinamento dos quatro modelos foram divididos em três blocos, mantendo uma proporção de 70% (964 imagens), 20% (241 imagens), 10% (180 imagens) para treino, validação e teste respectivamente. Cada um dos modelos foi treinado considerando o conjunto de hiperparâmetros apresentados na Tabela 2. Como pode ser visto, há uma variação no número de iterações e no tamanho do *batch*, isso é devido a restrições computacionais inerentes para cada modelo. Além disso, foi aplicado a técnica *early stopping* que visa interromper o treinamento quando o modelo começa apresentar características de *overfitting*.

Tabela 2. Hiperparâmetros empregados em cada modelo

Modelo	Taxa Aprendizagem	Batch size	Épocas	Optimizador
U-Net	0,0001	2	30	Adam
Mask RCNN	0,00025	4	10	Adam
GoogleNet	0,00001	1	100	RMSprop
RED CNN	0,00001	1	100	RMSprop

Os resultados obtidos por cada modelo de segmentação no conjunto de dados de teste (180 imagens) são mostrados em Tabela 3. Como pode ser observado, os modelos Unet e Mask RCNN obtêm os melhores resultados na métrica IoU, com valores de 43% e 42% respectivamente. A Unet foi o melhor modelo na análise das métricas de acurácia e precisão, sendo em média 6% e 80% melhor que Mask RCNN e GoogleNet, além de apresentar a segunda melhor média de tempo de processamento (0,062 s). Por outro lado, o modelo com pior desempenho em todas as métricas de validação foi o RED CNN, onde a precisão na classificação dos pixels não ultrapassa 29% e a sobreposição entre a máscara prevista é inferior a 33% (IoU), apesar disso ela foi a arquitetura com menor tempo de processamento médio das imagens (0,003 s). Este resultado pode ser atribuído a dois motivos principais: 1) à diferença no número de parâmetros utilizados e/ou aprendidos por cada modelo (profundidade da rede) e 2) à variação dos hiperparâmetros utilizados⁴.

Tabela 3. Resultado da segmentação de erosão no conjunto de dados de teste.

Modelo	Acurácia	Precisão	F1-score	IoU	Tempo
U-Net	0,98	0,67	0,58	0,43	0,0625s
Mask R-CNN	0,92	0,62	0,61	0,42	0,82s
GoogleNet	0,10	0,43	0,28	0,41	0,095s
RED-CNN	0,29	0,49	0,47	0,33	0,003s

⁴ É importante notar que a Tabela 3 apresenta o tempo de processamento médio do conjunto de imagens de validação com um tamanho de (512 × 512) pixels.

A partir da Tabela 3, também é possível confirmar que a U-Net produziu um elevado número de verdadeiros positivos e poucos falsos positivos/negativos, alcançando um bom equilíbrio entre a precisão (67%) e F1-score (58%). Esses resultados são confirmados na Figura 8, onde as curvas de IoU na fase de treinamento obtidas por todos os detectores são apresentadas. Como pode ser visto, a curva da U-Net apresenta os valores mais altos⁵.

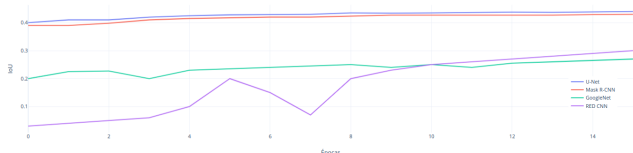


Figura 8. Evolução do IoU ao longo das épocas de treinamento.

Para uma análise mais aprofundada acerca dos custos computacionais e tempo de processamento observa-se que a Mask R-CNN obteve o pior tempo de processamento, isso se dá pela arquitetura complexa e com muitos parâmetros. Em comparação, a RED-CNN obteve o melhor tempo de processamento, uma vez que sua arquitetura é composta de camadas similares e *pipeline* bem definido. Já no caso da U-Net, por se tratar de uma das arquiteturas mais utilizadas e exploradas na literatura, ela acaba sendo altamente eficiente (apesar da quantidade de parâmetros), tornando-se uma das redes com custo mais efetivo, ficando atrás apenas da RED-CNN nos experimentos conduzidos neste trabalho. Por fim temos a GoogleNet que demanda um uso computacional moderado justamente pelo uso de blocos com diferentes tamanhos de filtros. No geral, das 4 arquiteturas utilizadas nesse trabalho, a GoogleNet é a segunda mais custosa.

4.2 Análise qualitativa

Para avaliar qualitativamente os modelos, três regiões de interesse foram selecionadas aleatoriamente. Essas regiões são compostas por imagens ou *patches* representadas por um vetor de tamanho $[512 \times 512 \times 3]$. Essas imagens são caracterizadas por apresentar diferentes tipos de terrenos, como montanhas, áreas planas, vegetação, solo exposto e áreas com erosão. Também é possível observar a escala das imagens de satélite em relação às áreas de interesse (tamanho), o que aumenta a dificuldade em seu reconhecimento e classificação.

A Figura 9 e Figura 10 mostra a predição de segmentação semântica dos modelos nas imagens de teste. Onde são geradas máscaras aceitáveis sobre as regiões de interesse. Esses resultados são correlacionados com as métricas apresentadas na Tabela 3, onde o modelo U-Net e a Mask RCNN apresentam os valores máximos de IoU (43%). Ao fazer uma comparação das predições entre os U-Net e Mask RCNN, é possível observar que as máscaras preditas apresentam alta similaridade em relação ao *ground truth*, mantendo aspectos como geometria e tamanho da erosão. Porém, o modelo Mask RCNN gera mais falsos positivos

⁵ A Figura apresenta a evolução dos modelos ao longo das iterações de treinamento ao que respeita o IoU, considerando até a época 15, onde os modelos atingiram os valores máximos.

que a U-Net, e isso se reflete na acurácia (92%) e precisão (62%) do modelo.

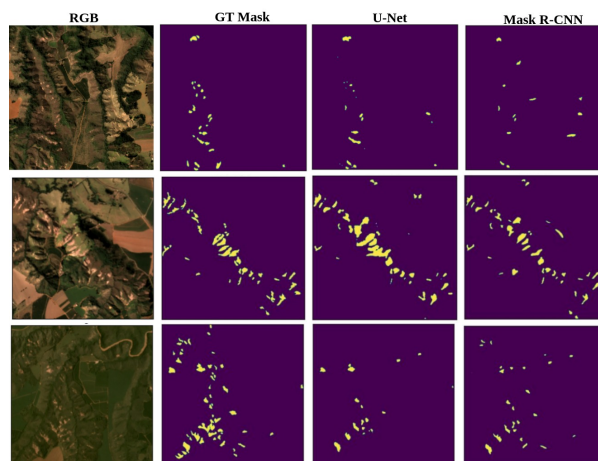


Figura 9. Resultados da segmentação dos modelos U-Net e Mask RCNN nas imagens de teste.

Com base na Figura 10, os modelos GoogleNet e RED CNN apresentam os piores resultados na geração de máscaras para regiões com erosão. O GoogleNet gerou um grande número de falsos positivos e falsos negativos nas três imagens de teste (círculos vermelhos). Por outro lado, a RED CNN não detectou muitas regiões com erosão (falsos negativos), e tal fato mostra que este modelo apresenta dificuldades em identificar regiões pequenas cuja quantidade de pixels é menor que 200.

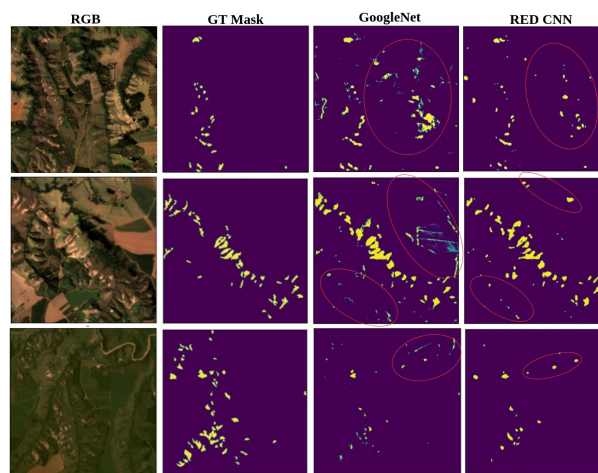


Figura 10. Resultados da segmentação dos modelos GoogleNet e RED CNN nas imagens de teste.

Em geral as máscaras previstas pelos modelos exibem uma geometria que possui alta similaridade com a verdade fundamental particularmente para regiões de grandes erosões. Para regiões pequenas, todos os modelos apresentaram um grau de dificuldade de segmentação, isso se deve principalmente ao fato do conjunto de dados não possuir um número elevado de exemplos de regiões pequenas de erosões para treinamento. Também é importante destacar que devido à arquitetura dos modelos, alguns apresentam melhor aprendizado no reconhecimento de pequenos objetos, aqueles com menos de 200 pixels ou $5000 m^2$.

5. CONCLUSÕES

Neste artigo, o problema de identificação da erosão no estado de Minas Gerais (Brasil) foi abordado com sucesso por meio de imagens de satélite de alta resolução. Um novo conjunto de dados de satélite foi proposto utilizando a combinação de bandas RGB com resolução de 5 metros/pixels. Com o objetivo de identificar a melhor abordagem baseada em DL, foram avaliados quatro modelos de segmentação de última geração existentes, incluindo: U-Net, Mask R-CNN, GoogleNet, RED CNN. Os melhores resultados no reconhecimento de áreas de erosão no conjunto de dados de teste foram obtidos por U-net e Mask R-CNN, com valores de acurácia de 98% e 92% respectivamente. Na métrica IOU, estes modelos obtiveram os valores mais elevados com um valor médio de 44%.

De acordo com a complexidade da identificação da erosão, U-Net e Mask R-CNN demonstraram ter um desempenho notável nas imagens de teste, em que os desafios são elevados, como variação de tamanhos, mudanças na forma e textura das regiões com erosão. Esses resultados se assemelham a outros estudos presentes no estado da arte (como Gafurov and Yermolaev (2020)), que embora a resolução seja alta, os modelos apresentam grande dificuldade em identificar regiões com erosões pequenas. Uma hipótese do porquê da dificuldade em segmentação de erosões pequenas pode ser a anotação das imagens durante a criação das bases de dados, uma vez que a identificação primária das erosões é feita por humanos o que pode levar a erros, falsos positivos ou até falha em identificar alguma erosão devido a resolução das imagens. Além disso, todos os modelos apresentaram tal dificuldade o que leva a uma outra hipótese de que os modelos podem não estar tão preparados para a detecção de objetos tão pequenos em imagens de satélites (lembrando que definimos erosões pequenas aquelas que possuem até 200 pixels).

Como trabalho futuro, será feito um aumento do conjunto de dados proposto incluindo imagens de outras áreas do estado de Minas Gerais, tornando-o mais robusto e representativo. Além disso, objetiva-se realizar a implementação de novas estratégias sobre os modelos de DL com o objetivo de melhorar a identificação de erosões pequenas.

AGRADECIMENTOS

Agradecemos à CEMIG e ANEEL pelo auxílio na publicação e recursos (PD-04950-0664/2023) empregados no trabalho realizado.

REFERÊNCIAS

Bui, D.T., Shirzadi, A., Shahabi, H., Chapi, K., Omidav, E., Pham, B.T., Talebpour Asl, D., Khaledian, H., Pradhan, B., Panahi, M., Bin Ahmad, B., Rahmani, H., Gróf, G., and Lee, S. (2019). A novel ensemble artificial intelligence approach for gully erosion mapping in a semi-arid watershed (iran). *Sensors (Basel, Switzerland)*, 19(11), 2444.

Chen, H., Zhang, Y., Kalra, M.K., Lin, F., Chen, Y., Liao, P., Zhou, J., and Wang, G. (2017). Low-dose ct with a residual encoder-decoder convolutional neural network. *IEEE transactions on medical imaging*, 36(12), 2524–2535.

Gafurov, A. (2022). Mapping of rill erosion of the middle volga (russia) region using deep neural network. *International Journal of Geo-Information*, 11, 1–24.

Gafurov, A. and Yermolaev, O. (2020). Automatic gully detection: Neural networks and computer vision. *Remote Sensing*, 12, 1743–1759.

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, 2961–2969.

Liu, K., Na, J., Fan, C., Huang, Y., Ding, H., Wang, Z., Tang, G., and Song, C. (2022). Large-scale detection of the tableland areas and erosion-vulnerable hotspots on the chinese loess plateau. *Remote Sensing*, 14(8), 1946.

Makaya, N.P., Mutanga, O., Kiala, Z., Dube, T., and Seutloali, K.E. (2019). Assessing the potential of sentinel-2 msi sensor in detecting and mapping the spatial distribution of gullies in a communal grazing landscape. *Physics and Chemistry of the Earth, Parts A/B/C*, 112, 66–74.

Nogueira, K., LS Machado, G., HT Gama, P., CV da Silva, C., Balaniuk, R., and A. dos Santos, J. (2020). Facing erosion identification in railway lines using pixel-wise deep-based approaches. *Remote Sensing*, 12(4), 739.

Phinzi, K., Abriha, D., Bertalan, L., Imre, H., and Szabo, S. (2020). Machine learning for gully feature extraction based on a pan-sharpened multispectral image: Multi-class vs. binary approach. *International Journal of Geo-Information*, 9, 252.

Phinzi, K., Imre, H., and Szabo, S. (2021). Mapping permanent gullies in an agricultural area using satellite images: Efficacy of machine learning algorithms. *Agronomy*, 11, 333.

Rahaman, M.H., Masroor, M., Sajjad, H., et al. (2023). Integrating remote sensing derived indices and machine learning algorithms for precise extraction of small surface water bodies in the lower thoubal river watershed, india. *Journal of Cleaner Production*, 422, 138563.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18*, 234–241. Springer.

Schellekens, J. and Amani, M. (2022). Coastal erosion detection using landsat satellite imagery and support vector machine algorithm. *Journal of Ocean Technology*, 17.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2014). Going deeper with convolutions.

Targ, S., Almeida, D., and Lyman, K. (2016). Resnet in resnet: Generalizing residual architectures. *arXiv preprint arXiv:1603.08029*.

Vågen, T.G. and Winowiecki, L. (2019). Predicting the spatial distribution and severity of soil erosion in the global tropics using satellite remote sensing. *Remote Sensing*, 11(15), 1800.

Wang, Y., Zhang, Y., and Chen, H. (2023). Gully erosion susceptibility prediction in mollisols using machine learning models. *Journal of Soil and Water Conservation*, 78(5), 385–396.