



Land Use and Occupation Detection in Satellite Images Using Deep Learning: Evaluation of LinkNet and D-LinkNet Architectures

Álisson de Oliveira Alves, Luisa Christina de Souza,
Luiz Eduardo Nunes Cho Luck,
Raniere Rodrigues Melo de Lima,
Carlos Augusto Teixeira de Moura,
Wesley José dos Santos Marinho,
Rafael de Medeiros Mariz Capuano,
Bruno Cesar Pereira da Costa, Marina de Siqueira,
Arthur Diniz Flor Torquato Fernandes,
Jesaias Carvalho Pereira Silva and Pablo Javier Alsina

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

January 6, 2025

Detecção de Uso e Ocupação do Solo em Imagens de Satélite utilizando *Deep Learning*: Avaliação das Arquiteturas LinkNet e D-LinkNet

Álisson de Oliveira Alves* ** Luisa Christina de Souza* **
Luiz Eduardo Nunes Cho Luck** Raniera Rodrigues Melo de Lima**
Carlos Augusto Teixeira** Wesley José dos Santos Marinho**
Rafael de Medeiros Mariz Capuano** Bruno Cesar Pereira da Costa**
Marina de Siqueira** Arthur Diniz Flor Torquato Fernandes***
Jesaias Carvalho Pereira Silva** Pablo Javier Alsina*

* Departamento de Engenharia de Computação e Automação,
Universidade Federal do Rio Grande do Norte, Natal, Brasil

** Instituto SENAI de Inovação em Energias Renováveis, Natal, Brasil

*** Departamento de Engenharia Industrial, Universidade de Nápoles
Federico II, Nápoles, Itália

Abstract: Detection of land cover in remote sensing images can contribute to various scientific and social applications. The development of techniques for this detection has been driven by advancements in deep learning algorithms, such as convolutional neural networks. Thus, this study employed two neural network architectures, LinkNet and D-LinkNet, using different *backbones* to perform land cover segmentation in high-resolution satellite images. The LinkNetB7 architecture, with EfficientNet-B7 as *backbone*, demonstrated higher sensitivity and precision, providing a more refined segmentation of land cover compared to other models, achieving an accuracy of 0,91, a sensitivity of 0,93 and a IoU of 0,84.

Resumo: A detecção da ocupação do solo em imagens provenientes de sensoriamento remoto pode contribuir para várias aplicações nas esferas científica e social. O desenvolvimento de técnicas para realizar essa detecção foi impulsionado pelo aprimoramento de algoritmos de aprendizado profundo, como redes neurais convolucionais. Dessa forma, o presente trabalho utilizou duas arquiteturas de redes neurais, a LinkNet e a D-LinkNet, aplicando *backbones* distintos para realizar a segmentação da ocupação do solo em imagens de satélite de alta resolução. A arquitetura LinkNetB7, com o *backbone* EfficientNet-B7, demonstrou maior sensibilidade e precisão, proporcionando uma segmentação mais refinada da ocupação do solo em comparação com os demais modelos, alcançando uma acurácia de 0,91, sensibilidade de 0,93 e IoU de 0,84.

Keywords: Deep Learning, Convolutional Neural Networks, Remote sensing, Satellite Image Segmentation, Backbones.

Palavras-chaves: Aprendizado Profundo, Redes Neurais Convolucionais, Sensoriamento Remoto, Segmentação de Imagens de satélite, *Backbones*.

1. INTRODUÇÃO

A detecção de uso e ocupação do solo para diferentes tipos de construções é crucial em estudos e aplicações como planejamento urbano, gestão de riscos, infraestrutura, estudos socioeconômicos e modelagem ambiental (de Sousa Coelho et al., 2024). O desenvolvimento de técnicas precisas para detecção de construções são essenciais para compreender as cidades e tomar decisões para seu desenvolvimento sustentável (Robinson et al., 2022).

As redes neurais convolucionais (*Convolutional Neural Networks* - *CNN*) destacam-se como ferramentas poderosas para análise de imagens devido à sua habilidade em aprender representações complexas diretamente dos dados (Géron, 2019). Além disso, avanços na análise e seg-

mentação de imagens de satélite têm sido facilitados pela disponibilidade de grandes conjuntos de dados, melhorias na arquitetura das redes e técnicas de treinamento mais eficazes, apesar de desafios como variações na iluminação, oclusões e mudanças sazonais, particularmente em áreas com cobertura vegetal densa ou ambientes urbanos complexos (Neupane et al., 2021).

Nesse contexto, (Ohleyer, 2018) comparou três metodologias de segmentação de imagens de sensoriamento remoto, a *Multilayer Perceptron* (MLP), *Fully Convolutional Network* (FCN) e um algoritmo baseado nas regiões da máscara, intitulado *Mask R-CNN*. Os resultados são analisados utilizando imagens de validação de cinco cidades americanas. O MLP obteve o melhor desempenho geral,

com um Índice de *Intersection of Union* (IoU) de aproximadamente 0,68.

A adaptação de redes neurais convolucionais (CNNs) pré-treinadas, como ResNet (Targ et al., 2016), AlexNet (Krizhevsky et al., 2017) e VGG (Simonyan and Zisserman, 2014), para a análise de imagens de sensoriamento remoto tem se mostrado promissora. Embora desenvolvidas para visão computacional com imagens convencionais, essas arquiteturas podem ser ajustadas para as características específicas das imagens de satélite (Chaurasia and Culurciello, 2017). Frequentemente utilizadas como *backbones* em modelos de segmentação mais complexos, essas CNNs permitem a integração de informações espaciais e contextuais, aproveitando o conhecimento adquirido em grandes conjuntos de dados para melhorar o desempenho na análise de imagens de satélite (Neupane et al., 2021).

O trabalho de (Jiwani et al., 2021) propõe um módulo modificado *DeeplabV3+* com *Dilated ResNet* como *backbone* para segmentação de edifícios em imagens de satélite RGB. Utilizando uma medida F-Beta na função objetivo para enfrentar desequilíbrios de classe e treinamento cruzado de conjuntos de dados, o estudo obteve um IoU variando de 0,70 a 0,92, conforme o conjunto de dados. Da mesma forma, (Li et al., 2019) apresentam uma arquitetura baseada na U-Net para segmentar áreas de uso e ocupação do solo em imagens de satélite de alta resolução, empregando conjuntos de dados do SpaceNet (Demir et al., 2018) e mapas GIS públicos, como *OpenStreetMap* e *Google Maps*. As estratégias adotadas, incluindo aumento dos dados, processamento posterior e integração de dados GIS e imagens de satélite, resultaram em um F1-score total de 0,70.

A pesquisa de (Chaurasia and Culurciello, 2017) propõe a LinkNet, uma estrutura de codificador-decodificador que faz uso da ResNet34, uma arquitetura pré-treinada em imagens do conjunto de dados ImageNet. O objetivo do trabalho é realizar segmentação semântica de forma eficiente, em termos de computação e memória, com recursos escassos. Baseando-se nesse trabalho, (Zhou et al., 2018) implementou uma abordagem utilizando a LinkNet com adição de convoluções dilatadas na parte central da arquitetura. No Desafio de Extração de Estradas DeepGlobe 2018 da CVPR (Demir et al., 2018), os melhores *scores* de IoU no conjunto de validação e no conjunto de teste atingidos pelo modelo foram, respectivamente, 0,64 e 0,63.

Este estudo investiga a detecção de uso e ocupação do solo através de algoritmos de segmentação utilizando as arquiteturas LinkNet e D-LinkNet com diversos *backbones*. Estes *backbones* são cruciais, pois fornecem as estruturas básicas para redes neurais convolucionais, impactando diretamente a capacidade de aprendizado e generalização do modelo. A análise visa entender o desempenho dessas arquiteturas na detecção em imagens de satélite de alta resolução, contribuindo para o desenvolvimento de métodos mais eficazes no monitoramento e análise de áreas urbanas e rurais, com aplicações em planejamento urbano, gestão ambiental e desenvolvimento sustentável.

2. MATERIAIS E MÉTODOS

A Figura 1 apresenta o fluxo de atividades do estudo, que inicia com a obtenção das imagens de satélite, se-

guidas pelo pré-processamento necessário antes de serem utilizadas nos algoritmos de segmentação com diferentes *backbones*. Além disso, são descritas as configurações de treinamento, que variam conforme as funções de perda utilizadas. Por fim, as métricas de desempenho escolhidas são avaliadas para a comparação quantitativa dos modelos, visando a seleção do modelo de segmentação mais adequado.

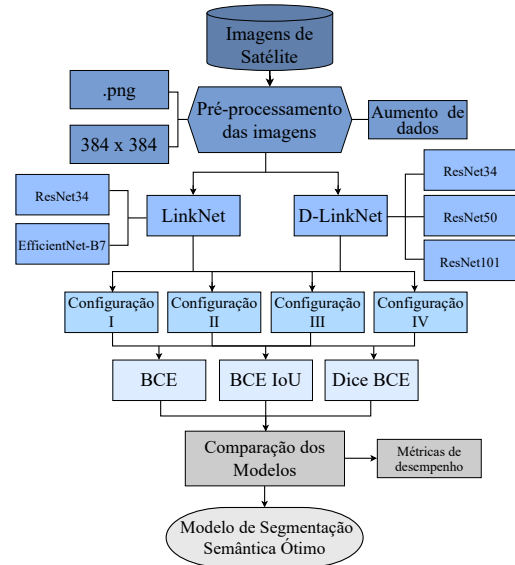


Figura 1. Fluxograma de atividades para a comparação dos modelos de segmentação semântica de imagens de satélite de alta resolução para a detecção de uso e ocupação do solo. Fonte: Elaboração própria

2.1 Obtenção e pré-processamento das imagens de satélite

A etapa inicial do trabalho consiste na aquisição de imagens de satélite de alta resolução, provenientes do *Google Earth*, que serão utilizadas para alimentar os algoritmos de segmentação. As imagens foram escolhidas devido à sua alta resolução, cerca de 60 cm por *pixel*, e pelo processo de aquisição simplificado, dispensando a composição de bandas espectrais.

Uma vantagem adicional dessas imagens é a possibilidade de georreferenciamento, permitindo a obtenção de máscaras binárias, ou *groundtruth*, através do *Open Buildings V3 Polygons* (Sirko et al., 2021), um conjunto de dados cartográficos que representa edifícios e estruturas urbanas. No entanto, o *Open Buildings V3 Polygons* só fornece máscaras a partir de 2023, limitando a disponibilidade de dados históricos. Assim, optou-se por utilizar apenas imagens de 2023 para compor a base de dados de treinamento.

A etapa de pré-processamento das imagens envolve a adequação do par de imagem, composto por uma imagem de satélite e uma máscara binária correspondente, aos requisitos pré-estabelecidos pelas arquiteturas de rede neural. Os pares originais, que anteriormente apresentavam extensão *.TIFF* e apresentavam a dimensão de cerca de 8000 x 3000 *pixels* por imagem, passaram então pela conversão para o formato *.PNG* e foram divididas em *patches* de comprimento 384 x 384 *pixels*. O resultado dos processamentos

gerou no total 1029 pares de imagens. A Figura 2 exibe um par de exemplo utilizado no treinamento.

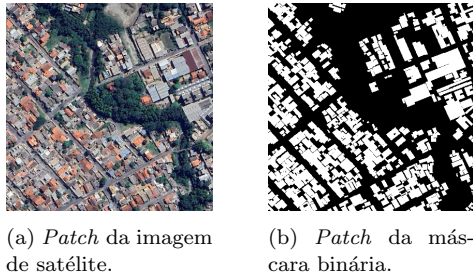


Figura 2. Par de imagens de exemplo de uma região do Brasil. (a) Recorte de uma imagem de satélite em sua composição original (RGB). (b) Máscara correspondente obtida através do *Open Buildings V3 Polygons*. Fonte: Elaboração própria

Do total de imagens, 80% foram usadas para o treinamento do modelo, enquanto os 20% restantes foram reservados para validação, que é essencial para ajustar os hiperparâmetros e evitar o sobreajuste. Para aprimorar a generalização do modelo, foi realizada a técnica de aumento de dados. Aplicando transformações como rotação, espelhamento e deslocamento, o conjunto de treinamento foi expandido de 720 para 8.640 pares de imagens. Essa diversificação artificial introduziu novas variabilidades, permitindo ao modelo aprender com maior robustez e melhorar a generalização para novas imagens.

2.2 Arquiteturas de Redes Neurais Artificiais

A segmentação semântica, fundamental na visão computacional, classifica cada pixel de uma imagem em categorias como estradas, edifícios, vegetação e água. Modelos baseados em CNNs têm demonstrado superioridade em precisão e eficiência computacional em relação a abordagens tradicionais (Jiwani et al., 2021). Este estudo utiliza duas arquiteturas de redes neurais artificiais, LinkNet e D-LinkNet, com variações nos modelos finais conforme os diferentes *backbones* adotados.

Modelos Baseados na LinkNet: A rede neural de segmentação semântica LinkNet, descrita por (Chaurasia and Culurciello, 2017), tem sua arquitetura ilustrada na Figura 3. Consiste em uma composição de blocos de codificação (*encoder*) para decompor a imagem reduzindo sua dimensionalidade, seguidos por blocos de decodificação (*decoder*) encarregados de reconstruir o dado. Trata-se de um algoritmo simples e eficiente, requerendo menos dados e tempo de treinamento e mantendo o resultado da segmentação comparável a outras metodologias similares. Para a aplicação atual, foram utilizadas duas redes neurais convolucionais baseadas na arquitetura da LinkNet:

- **LinkNet34:** Utiliza a ResNet34 pré-treinada no conjunto de dados ImageNet como seu codificador (Zhou et al., 2018).
- **LinkNetB7:** Faz uso do EfficientNet-B7 pré-treinado como codificador, o que possibilita o aprimoramento da extração de características e da generalização do modelo. Utiliza escalonamento composto para ajuste dos parâmetros da rede neural, resultando em um

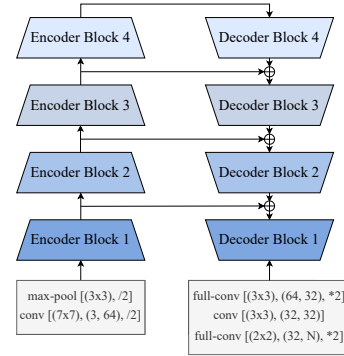


Figura 3. Arquitetura de Rede Neural Profunda LinkNet. Adaptada de: (Chaurasia and Culurciello, 2017).

modelo de segmentação mais eficiente (Akyel and Arici, 2022).

Modelos Baseados na D-LinkNet: A segunda CNN base é a *D-LinkNet*, proposta por (Zhou et al., 2018), para realizar a extração de estradas em imagens de satélite de alta resolução. Trata-se de uma arquitetura similar a LinkNet, porém faz uso de convoluções dilatadas, o que torna a extração de detalhes finos nas imagens de satélite mais eficiente. A Figura 4 exibe a arquitetura do algoritmo de segmentação *D-LinkNet*.

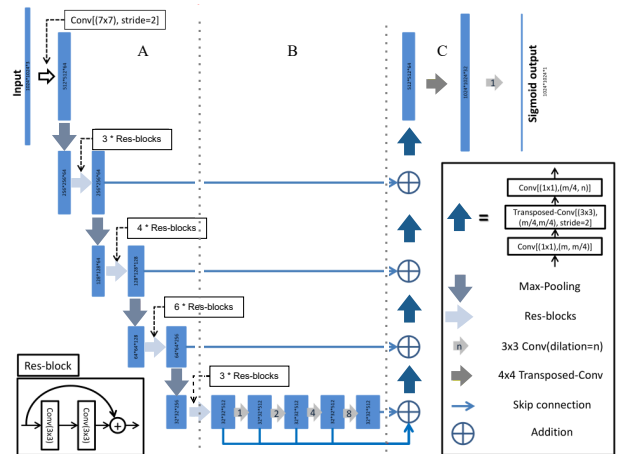


Figura 4. Arquitetura da Rede Neural Profunda D-LinkNet. Adaptada de: (Zhou et al., 2018).

O bloco A, ou bloco codificador, é responsável pela captura e extração dos padrões de características dos dados de entrada, como bordas, textura, formas além de características hierárquicas e abstratas. O bloco B, ou *bottleneck*, é a parte central do modelo, que utiliza convolução dilatada nas modalidades cascata e paralela. Em cada modalidade, o campo receptivo das camadas convolucionais é distinto, permitindo que a rede capture e processe informações de diferentes escalas da imagem. O bloco C é a unidade *decoder* que utilizará poucos parâmetros, pois compartilha conhecimento com o codificador (Zhou et al., 2018). Para a aplicação atual, foram empregadas três redes neurais convolucionais baseadas na arquitetura da D-LinkNet:

- **D-LinkNet34:** Seu codificador é o ResNet34, uma arquitetura com blocos residuais de 34 camadas, ideal

para tarefas menos complexas devido à sua eficiência computacional em comparação com arquiteturas ResNet mais profundas (Meng et al., 2023).

- **D-LinkNet50:** Utiliza o ResNet50 como codificador, uma rede neural com 50 camadas, adequada para tarefas mais complexas devido à profundidade da arquitetura (Li and Liu, 2022).
- **D-LinkNet101:** O *backbone* deste modelo é o ResNet101, uma arquitetura com 101 camadas, ideal para tarefas mais complexas que exigem alta precisão. É semelhante ao ResNet50, mas possui mais blocos residuais, resultando em uma capacidade de representação mais detalhada (Seo et al., 2022).

2.3 Configurações dos parâmetros de treinamento

Com o objetivo de encontrar o modelo ideal para a segmentação de imagens de satélite, quatro cenários de configuração dos parâmetros de treinamento foram delineados, com base na literatura (Zhou et al., 2018; Akyel and Arıcı, 2022). Os cenários variam em épocas e tamanho de lote (*batchsize*). Hiperparâmetros como taxa de aprendizado inicial e *threshold* de binarização foram mantidos conforme trabalhos relacionados. A Tabela 1 resume os cenários de treinamento.

Tabela 1. Configurações dos parâmetros de treinamento das Redes Neural Convolucionais.

Configuração	Épocas	Tamanho de <i>Batch</i>
I	20	4
II	40	8
III	20	8
IV	40	4

Além de variar os cenários de treinamento, foram aplicadas três funções de perda: *Binary Cross Entropy* (BCE), que mede a entropia entre a distribuição de probabilidade prevista e a máscara binária (Ruby and Yendapalli, 2020); BCE IoU, que combina BCE e a métrica IoU, que pune as discrepâncias e as sobreposições entre a máscara de predição e a máscara binária (Batchkala and Ali, 2021); e *Dice* BCE, que equilibra precisão local e global na segmentação (Montazerolghaem et al., 2023). Cada uma das quatro configurações de parâmetros pré-definidos foram testadas com todas as funções de perda. Desta maneira, para cada arquitetura de rede neural, foram conduzidos 12 treinamentos com hiperparâmetros distintos.

2.4 Ambiente Experimental

Os algoritmos foram implementados em linguagem de programação Python com a biblioteca PyTorch, em uma máquina com CPU AMD Ryzen Threadripper 3970X x64. Foi utilizado o *toolkit* CUDA para agilizar o treinamento dos modelos, aproveitando o poder de processamento da GPU NVIDIA GeForce RTX 4090 com 24 GB, e memória RAM de 256 GB.

2.5 Métricas de avaliação

As Equações 1, 2 e 3 descrevem métricas objetivas utilizadas para a avaliação dos resultados associados aos modelos de segmentação.

$$Acurácia(ACC) = \frac{TP + TN}{TP + TN + FP + FN}, \quad (1)$$

$$Precisão(PRE) = \frac{TP}{TP + FP}, \quad (2)$$

$$Sensibilidade(SEN) = \frac{TP}{TP + FN}, \quad (3)$$

TP , TN , FP e FN representam o total de *pixels* associados à taxa de verdadeiro positivo, verdadeiro negativo, falso positivo e falso negativo, respectivamente.

$$Intersection\ of\ Union(IOU) = \frac{GT \cap PD}{GT \cup PD}. \quad (4)$$

e

$$F1 - Score = 2 \times \frac{PRE * SEN}{PRE + SEN}. \quad (5)$$

GT e PD representam a relação de *pixels* associados à máscara de referência (*groundtruth*) e a máscara predita, respectivamente.

A acurácia mede a proximidade entre o conjunto de *pixels* de referência e a máscara predita. A precisão é a proporção de exemplos positivos corretamente classificados entre todos os exemplos considerados positivos, enquanto a sensibilidade é a proporção de exemplos positivos identificados pelo modelo entre todos os exemplos reais. A métrica *IoU* (ou *Jaccard*) avalia a sobreposição entre a máscara predita e a de referência, e a *F1-Score* é a média harmônica entre precisão e sensibilidade (Huyen, 2022).

3. ANÁLISE DOS RESULTADOS E DISCUSSÃO

Após o treinamento e validação dos modelos de segmentação de imagens de satélite para detecção de uso e ocupação do solo, as métricas de desempenho da validação foram avaliadas para cada configuração. Cada cenário, considerando diferentes funções de perda, foi analisado em todas as métricas, sendo atribuído um ponto para o melhor desempenho em cada métrica. Em caso de empate, o tempo de treinamento foi utilizado como critério de desempate. Dessa forma, o cenário de melhor desempenho global foi identificado, priorizando modelos com tempos de treinamento mais eficientes em casos de empate. Os modelos mais satisfatórios foram listados na Tabela 2.

Após a análise quantitativa dos melhores desempenhos, foi realizada uma avaliação qualitativa dos modelos de segmentação de uso e ocupação do solo. Essa avaliação baseou-se na observação das predições das arquiteturas mencionadas em uma imagem de satélite de uma cena em Minas Gerais. A Figura 5 ilustra os resultados das inferências, com o item (a) mostrando a imagem original e os demais itens apresentando a sobreposição das predições na imagem original. A Figura 6 exhibe, no item (a), a máscara binária (*groundtruth*) obtida pelo *Open Buildings V3 Polygons*, enquanto os demais itens mostram a máscara de predição (PD).

Com base na Tabela 2, a rede neural LinkNetB7, na configuração III com a função de perda BCE IoU, demonstrou os melhores resultados nas cinco métricas de desempenho e no tempo de treinamento, justificando sua escolha como o modelo mais adequado para o cenário. Embora não

Tabela 2. Melhores Performances de Validação das Rede Neural de Segmentação.

Modelo	Configuração	Função de Perda	Acurácia	F1-Score	IoU	Precisão	Sensibilidade	Tempo de Treinamento
LinkNet34	III	BCE IoU	0,90	0,76	0,83	0,66	0,91	11 h 06 min
LinkNetB7	III	BCE IoU	0,91	0,78	0,84	0,68	0,93	9 h 54 min
D-LinkNet34	I	Dice BCE	0,93	0,71	0,88	0,65	0,83	14 h 52 min
D-LinkNet50	III	BCE IoU	0,90	0,77	0,83	0,66	0,91	9 h 49 min
D-LinkNet101	IV	Dice BCE	0,93	0,72	0,88	0,66	0,81	17 h 42 min

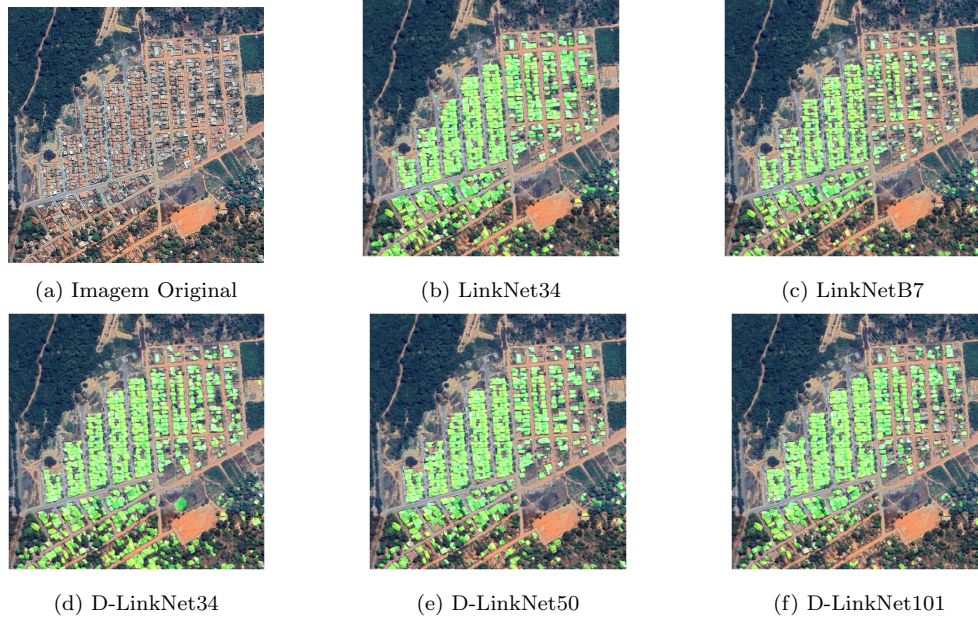


Figura 5. Comparação das representações visuais da segmentação da ocupação do solo sobreposta a imagem original.

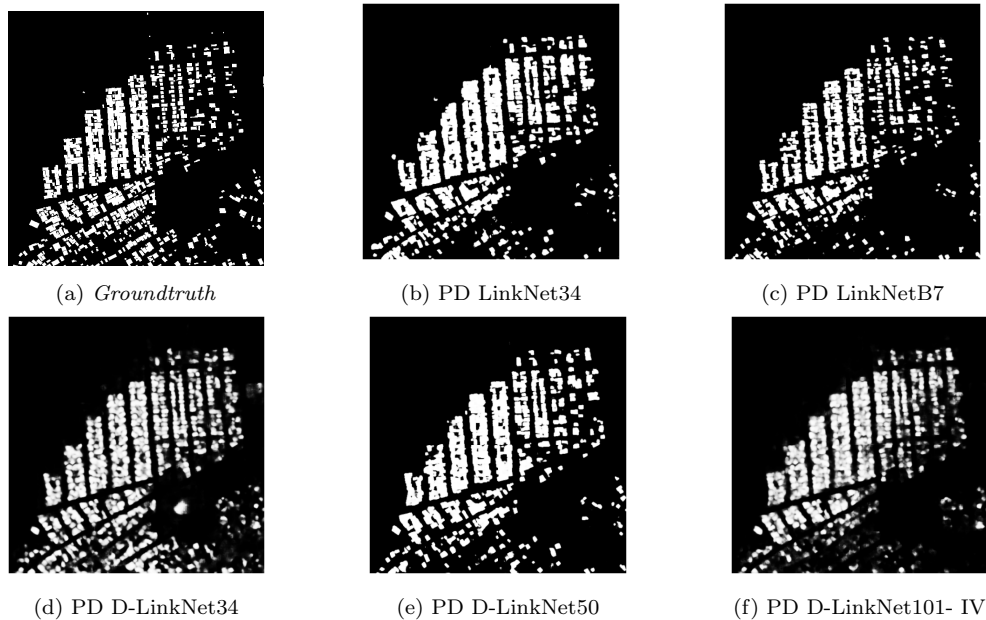


Figura 6. Comparação das máscaras preditas (PD) por diferentes modelos de rede neural.

tenha a maior acurácia, com 0,91 comparado a 0,93 dos modelos D-LinkNet34 e D-LinkNet34, o LinkNetB7 com o *backbone* EfficientNet-B7 obteve os maiores valores de sensibilidade e precisão, essenciais para a segmentação de imagens pela sua influência na qualidade e confiabilidade das predições. Adicionalmente, apesar de sua maior

complexidade, o LinkNetB7 apresentou o segundo menor tempo de treinamento, com 9 horas e 54 minutos para todas as iterações.

O desempenho superior do LinkNetB7 é facilmente perceptível na análise qualitativa. Essa superioridade é evidenciada na Figura 6, onde a segmentação do uso e ocupação

do solo pelo LinkNetB7 se destaca por um refinamento notável, em contraste com as detecções em forma de grandes blocos observadas nos demais modelos. Esta precisão aprimorada se traduz em uma representação mais detalhada e precisa da tarefa de segmentação, proporcionando bases mais robustas para análises espaciais e fornecendo informações eficientes para diversos setores.

Em comparação com estudos consolidados na literatura, os modelos propostos neste trabalho demonstraram um IoU similar ou superior aos trabalhos de referência, como (Ohleyer, 2018) e (Jiwani et al., 2021), e um *F1-score* superior à aplicação desenvolvida por (Li et al., 2019).

4. CONCLUSÃO

Este estudo compara as arquiteturas de rede neural LinkNet e D-LinkNet, utilizando diferentes *backbones*, para detectar ocupação do solo em imagens de sensoriamento remoto. A configuração LinkNetB7, com função de perda BCE IoU, destacou-se como a mais eficaz em termos de precisão, sensibilidade e tempo de treinamento aceitável. Apesar de não ter a mais alta acurácia, sua capacidade de refinamento na segmentação proporcionou uma representação mais detalhada e precisa do ambiente. Os modelos desenvolvidos neste estudo apresentaram resultados comparáveis ou superiores aos trabalhos consolidados na literatura. Essa pesquisa contribui substancialmente para o monitoramento de áreas urbanas e rurais, fornecendo métodos mais eficazes para a detecção de ocupação do solo em imagens de sensoriamento remoto, com implicações importantes para o planejamento urbano, gestão ambiental e desenvolvimento sustentável.

AGRADECIMENTOS

Agradecemos à CEMIG, ANEEL (PD-04950-0664/2023), CNPq e CAPES pelo apoio financeiro e colaboração, essenciais para este estudo e para o avanço da pesquisa, desenvolvimento e inovação (PD&I) no setor elétrico.

REFERÊNCIAS

- Akyel, C. and Arıcı, N. (2022). Linknet-b7: noise removal and lesion segmentation in images of skin cancer. *Mathematics*, 10(5), 736.
- Batchkala, G. and Ali, S. (2021). Real-time polyp segmentation using u-net with iou loss.
- Chaurasia, A. and Culurciello, E. (2017). Linknet: Exploiting encoder representations for efficient semantic segmentation. In *2017 IEEE visual communications and image processing (VCIP)*, 1–4. IEEE.
- de Sousa Coelho, G., Liá, S.d.L.S., Liá, F.S., de Almeida, L.R., de Brito, L.G., Santos, I.B., da Silva, W.A., and da Silva, C.M. (2024). Análise do uso e ocupação do solo em diferentes períodos no município de cidelândia-ma. *CONTRIBUCIONES A LAS CIENCIAS SOCIALES*, 17(1), 8077–8093.
- Demir, I., Koperski, K., Lindenbaum, D., Pang, G., Huang, J., Basu, S., Hughes, F., Tuia, D., and Raskar, R. (2018). Deepglobe 2018: A challenge to parse the earth through satellite images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 172–181.
- Géron, A. (2019). *Mãos à Obra: Aprendizado de Máquina com Scikit-Learn e TensorFlow*. Alta Books, 2 edition.
- Huyen, C. (2022). *Designing machine learning systems*. O’Reilly Media, Inc.”.
- Jiwani, A., Ganguly, S., Ding, C., Zhou, N., and Chan, D.M. (2021). A semantic segmentation network for urban-scale building footprint extraction using rgb satellite imagery. *arXiv preprint arXiv:2104.01263*.
- Krizhevsky, A., Sutskever, I., and Hinton, G.E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84–90.
- Li, S. and Liu, X. (2022). Multi-type road extraction and analysis of high-resolution images with d-linknet50. In *2022 3rd International Conference on Geology, Mapping and Remote Sensing (ICGMRS)*, 244–248. IEEE.
- Li, W., He, C., Fang, J., Zheng, J., Fu, H., and Yu, L. (2019). Semantic segmentation-based building footprint extraction using very high-resolution satellite images and multi-source gis data. *Remote Sensing*, 11(4), 403.
- Meng, S., Di, Z., Yang, S., and Wang, Y. (2023). Large-scale weakly supervised learning for road extraction from satellite imagery. *arXiv preprint arXiv:2309.07823*.
- Montazerolghaem, M., Sun, Y., Sasso, G., and Haworth, A. (2023). U-net architecture for prostate segmentation: the impact of loss function on system performance. *Bioengineering*, 10(4), 412.
- Neupane, B., Horanont, T., and Aryal, J. (2021). Deep learning-based semantic segmentation of urban features in satellite images: A review and meta-analysis. *Remote Sensing*, 13(4), 808.
- Ohleyer, S. (2018). Building segmentation on satellite images. *Web: https://project.inria.fr/aerialimagelabeling/files/2018/01/fp_ohleyer_compressed.pdf*.
- Robinson, C., Ortiz, A., Park, H., Lozano, N., Kaw, J.K., Sederholm, T., Dodhia, R., and Ferres, J.M.L. (2022). Fast building segmentation from satellite imagery and few local labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1463–1471.
- Ruby, U. and Yendapalli, V. (2020). Binary cross entropy with deep learning technique for image classification. *Int. J. Adv. Trends Comput. Sci. Eng*, 9(10).
- Seo, B., Jang, S., Bong, J., Park, K., Lee, I., and Jeong, M. (2022). Application of deep learning to the production of sub-divided land cover maps. *HUMAN-CENTRIC COMPUTING AND INFORMATION SCIENCES*, 12.
- Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Sirko, W., Kashubin, S., Ritter, M., Annkah, A., Bouchareb, Y.S.E., Dauphin, Y., Keyzers, D., Neumann, M., Cisse, M., and Quinn, J. (2021). Continental-scale building detection from high resolution satellite imagery. *arXiv preprint arXiv:2107.12283*.
- Targ, S., Almeida, D., and Lyman, K. (2016). Resnet in resnet: Generalizing residual architectures. *arXiv preprint arXiv:1603.08029*.
- Zhou, L., Zhang, C., and Wu, M. (2018). D-linknet: Linknet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 182–186.