ResGhostU-Net: U-Net compacta para segmentação de eucalipto em imagens multiespectrais da Sentinel-2

Mounif Hassan Tormos

SERVIÇO DE PÓS-GRADUAÇÃO DO FACOM-UFMS

Data de Depósito:

Assinatura:_____

ResGhostU-Net: U-Net compacta para segmentação de eucalipto em imagens multiespectrais da Sentinel-2[1]

Mounif Hassan Tormos

Orientador: *Prof. Dr. Jonathan de Andrade Silva*

Documento de dissertação apresentado a Faculdade de Computação - FACOM - UFMS como parte dos requisitos necessários à obtenção do título de Mestre em Ciência de Computação.

UFMS - Campo Grande
janeiro/2025

[1]Trabalho Realizado com Auxílio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (Capes) - Proc. No: 88887.716106/2022-00

Dedico este trabalho aos meus queridos pais.

Agradecimentos

Agradeço a todos os professores que fizeram parte do meu aprendizado durante este curso de mestrado. Especialmente ao meu orientador, Dr. Jonathan, que me direcionou e ajudou a enfrentar vários problemas, principalmente na pesquisa científica, área em que eu tinha muita dificuldade; ao Dr. Wesley pelas reuniões interessantes sobre os métodos mais avançados de Inteligência Artificial; e ao Dr. José, pelas instruções sobre Sensoriamento Remoto, também agradeço a este pela oportunidade de utilizar a base de dados de eucalipto. Aos meus colegas do LIA e do Laboratório de Geomática, que sempre me ajudaram, compartilhando conhecimentos importantes, especialmente ao Eduardo e ao Mario, pelas ótimas ferramentas e pelas ajudas fornecidas para deste trabalho. Agradeço a todos meus amigos, cujos nomes, apesar de não estarem escritos neste documentos por questões de simplicidade, são dignos de um sincero agradecimento. Ao meu irmão, Ali, à minha irmã, Zahraa, aos meus pais, Amada e Hassan, que sempre me incentivam e me inspiram. Sem o apoio e a inspiração deles, não seria possível trilhar este caminho. A todos os meus parentes, próximos e distantes, pelas mensagens positivas, e aos meus avós maternos, Ramon e Ramona, e paternos, Sokna e Mounif, cujos boas memórias estão guardadas em um lugar especial do meu coração. Também não posso deixar de agradecer a Bela e a Cristal, meus pequenos animais de estimação que fazem parte da família. Agradeço também à Universidade Federal de Mato Grosso do Sul - UFMS, especialmente à FACOM, por toda a estrutura, e ao corpo docente e técnico por todo o suporte fornecido. Agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES, pelo apoio financeiro e tecnológico às pesquisas. E agradeço a Deus, por nos permitir explorar esse universo tão incrível.

Abstract

The mapping of eucalyptus using remote sensing images can be an inaccurate and laborious process, especially when considering the large-scale multitemporal analysis of images. To try to solve this problem, new machine learning approaches have been proposed. In this work, we propose a compact modified U-Net (ResGhostUNet) for the task of semantic segmentation of eucalyptus using Sentinel-2 satellite images. In addition to the simplified architecture that has a reduced number of filters and depth and downsampling convolutions, we introduce the Ghost Residual Block, which allows reducing the computational cost and increasing the training efficiency. This study uses a new dataset that contains images of eucalyptus plantations in different cities in the Brazilian Cerrado biome. The quantitative and qualitative results demonstrate that the proposed method is highly competitive with respect to popular semantic segmentation methods. The ablation study highlights the effectiveness of the proposed component of the method. Furthermore, it demonstrates that using at least four selected bands yields slightly better results compared to utilizing all 13 bands. The proposed method consistently outperforms popular semantic segmentation methods, being simpler in terms of design, lightweight in terms of parameters, and fast in terms of processing. Due to these characteristics, ResGhostU-Net is potentially applicable for large-scale eucalyptus mapping using open-access satellite imagery.

Resumo

O mapeamento do eucalipto utilizando imagens de sensoriamento remoto pode ser um processo impreciso e trabalhoso, especialmente quando se considera a análise multitemporal de imagens em larga escala. Para tentar resolver este problema, novas abordagens de aprendizado de máquina foram propostas. Neste trabalho, propomos uma U-Net modificada compacta (ResGhostU-Net) para a tarefa de segmentação semântica de eucalipto utilizando imagens do satélite Sentinel-2. Além da arquitetura simplificada que possui número reduzido de filtros e convoluções de profundidade e downsampling, introduzimos o Bloco Residual Fantasma, que permite reduzir o custo computacional e aumentar a eficiência do treinamento. Este estudo utiliza um novo conjunto de dados que contém imagens de plantações de eucalipto em diferentes cidades do bioma Cerrado brasileiro. Os resultados quantitativos e qualitativos demonstram que o método proposto é altamente competitivo em relação aos métodos populares de segmentação semântica. O estudo de ablação destaca a eficácia do componente proposto do método. Além disso, demonstra que a utilização de pelo menos quatro bandas selecionadas produz resultados ligeiramente melhores em comparação com a utilização de todas as 13 bandas. O método proposto supera consistentemente os métodos populares de segmentação semântica, sendo mais simples em termos de design, leve em termos de parâmetros e rápido em termos de processamento. Devido a estas características, a ResGhostU-Net é potencialmente aplicável para mapeamento de eucalipto em grande escala usando imagens de satélite de acesso aberto.

Sumário

Lista de Figuras

Lista de Tabelas

Lista de Abreviaturas

AP Aprendizagem Profunda

FA Floresta Aleatória

FN Falso Negativo

FP Falso Positivo

GEE Google Earth Engine

NDVI Normalized Difference Vegetation Index

RGB Red Green Blue

RNC Rede Neural Convolucional

SR Sensoriamento Remoto

VN Verdadeiro Negativo

VP Verdadeiro Positivo

VC Visão Computacional

Lista de Algoritmos

CAPÍTULO

1

Introdução

A produção de eucalipto desempenha um papel fundamental na economia brasileira, posicionando o Brasil como um dos principais produtores mundiais de pasta de celulose (Klein and Luna, 2022). Além de sua importância econômica, as florestas de eucalipto podem temporariamente (enquanto as árvores não são colhidas) atuar como sequestradoras de carbono, absorvendo dióxido de carbono (CO2) e ajudando a reduzir o efeito estufa (Bacani et al., 2024; Teodoro et al., 2024). Neste contexto, o mapeamento da produção de eucalipto é crucial para a compreensão dos seus impactos econômicos e ambientais, apoiando assim a implementação de práticas sustentáveis de uso da terra e contribuindo para os esforços contínuos de inventário florestal.

No Brasil, destaca-se o estado de Mato Grosso do Sul, onde as plantações de eucalipto são o tipo de floresta dominante. É a terceira maior área de plantações de eucalipto e tem a segunda maior taxa de crescimento em expansão de plantações em todo o país (Teodoro et al., 2024). No entanto, o monitoramento da dinâmica dessas plantações, especialmente sua expansão e impactos ambientais, apresenta um desafio significativo devido aos diversos biomas do estado, que incluem o Pantanal, o Cerrado e partes da Mata Atlântica, bem como a variabilidade na densidade das plantações.

As tecnologias de sensoriamento remoto, particularmente os métodos de segmentação semântica baseados em Aprendizagem Profunda (AP), oferecem uma solução eficaz para o mapeamento e monitoramento em larga escala de áreas de plantação (Osco et al., 2021a; Luo et al., 2024). Ao processar imagens de satélite de alta resolução com bandas multiespectrais, estas técnicas permitem a deteção precisa, em tempo real e automatizada de áreas de plantação, apoiando uma análise detalhada e actualizações atempadas. Os produtos

de deteção remota por satélite amplamente utilizados, como o Landsat 8 e o Sentinel, fornecem dados valiosos, com especial destaque para o Sentinel, que oferece uma resolução de 10 metros/pixel (Gazzea et al., 2023; Luo et al., 2024). Isto torna as imagens Sentinel particularmente adequadas para aplicações em grande escala, como a monitorização de povoamentos florestais.

A qualidade da segmentação foi significativamente melhorada por variantes de Rede Neural Convolucional (RNC), como as séries U-Net e Deeplab, que são altamente eficazes na captação de caraterísticas multi-escala com relevância espacial e na sua integração eficiente (Ajibola and Cabral, 2024). No entanto, as RNCs têm dificuldade em captar dependências de longo alcance e, embora a combinação de RNCs com Transformers tenha conduzido a resultados impressionantes no estado da arte da segmentação semântica, esta abordagem tem uma grande desvantagem: elevada complexidade computacional (Xu et al., 2023; Lu et al., 2024). Para além disso, os métodos baseados em Transformers dividem frequentemente as imagens em *patches*, convertendo-as em dados sequenciais unidimensionais, o que pode resultar na perda de informação espacial importante. Consequentemente, embora estes modelos sejam adequados para determinadas tarefas, enfrentam limitações, tais como pequenos lotes de treino e a necessidade de recursos computacionais dispendiosos.

1.1 Objetivos

Motivado por estes desafios, o objetivo principal deste estudo é introduzir uma abordagem de segmentação leve denominada ResGhostU-Net, baseada na arquitetura U-Net, para a segmentação semântica de plantações de eucalipto utilizando imagens Sentinel. Concebida para ser mais eficiente com menos parâmetros, a ResGhostU-Net é computacionalmente mais rápida e mais eficiente em termos de recursos em comparação com outros métodos de segmentação semântica amplamente utilizados na literatura. Ao atingir um equilíbrio ótimo entre desempenho e eficiência de recursos, a ResGhostU-Net é particularmente adequada para ambientes em tempo real ou com recursos limitados, onde os modelos tradicionais podem ter dificuldade em manter a precisão e a velocidade. Esta abordagem procura atender às necessidades específicas do monitoramento de plantações de eucalipto em larga escala no Mato Grosso do Sul, fornecendo uma solução eficiente e escalável para mapeamento automatizado e monitoramento ambiental.

1.2 Contribuições

As principais contribuições deste trabalho são:

- A proposta da ResGhostU-Net, uma modificação da arquitetura da U-Net com a introdução do Bloco Residual Fantasma, que proporciona um equilíbrio eficiente entre desempenho de segmentação, velocidade de processamento e consumo de memória, especificamente projetada para a segmentação de eucaliptos.
- A criação de um conjunto de dados para o treinamento e avaliação do método, que contém imagens e máscaras de plantações de eucalipto em diferentes cidades do bioma Cerrado brasileiro.
- Uma comparação do método proposto com abordagens populares de segmentação semântica, destacando suas vantagens em termos de eficiência e precisão. Testando sua capacidade de trabalhar em diferentes períodos de tempo e áreas geográficas.
- Um estudo de ablação dos componentes propostos e das bandas de entrada, analisando o impacto de cada parte do modelo e a utilização de diferentes bandas espectrais para uma melhor segmentação.

1.3 Organização

Este trabalho está organizado da seguinte forma:

No capítulo 2 apresentamos a fundamentação teórica para ajudar na familiarização com os métodos desenvolvidos. No capítulo 3 apresenta-se os trabalhos relacionados, evidenciando suas desvantagens e o diferencial do trabalho atual. No capítulo 4 apresentamos a U-Net proposta, discutindo suas vantagens em relação a U-Net tradicional. No capítulo 5 apresentamos a metodologia do experimento com detalhes sobre as etapas de geração de base de dados, treinamento e testes do modelos, e métodos aplicados para a avaliação. No capítulo 6 apresenta-se os resultados e discussão do experimento realizado. E finalmente, no capítulo 7 conclusões finais são apresentadas sobre os resultados alcançados, discutindo as suas limitações e trabalhos futuros.

CAPÍTULO

2

Fundamentação Teórica

2.1 Sensoriamento Remoto

O Sensoriamento Remoto (SR) é, segundo Meneses and Almeida (2012):

> "Uma ciência que visa o desenvolvimento da obtenção de imagens da superfície terrestre por meio da detecção e medição quantitativa das respostas das interações da radiação eletromagnética com os materiais terrestres."

Essa abordagem permite a aquisição de informações valiosas em situações onde o acesso direto seria inviável ou limitado, como no monitoramento de áreas de difícil alcance, grandes extensões territoriais ou regiões inóspitas. A detecção e a medição da radiação eletromagnética refletem propriedades físicas e químicas específicas dos materiais terrestres, possibilitando a identificação, classificação e análise de elementos naturais ou artificiais. Essa definição está particularmente associada ao sensoriamento remoto realizado por plataformas orbitais, como satélites, e aéreas, como drones e aviões, que desempenham papéis centrais em aplicações ambientais, agrícolas, urbanas e científicas.

Além disso, o sensoriamento remoto é amplamente utilizado para monitorar mudanças temporais, como o desmatamento, a urbanização e o comportamento de corpos d'água, e para gerar dados essenciais ao planejamento sustentável e ao gerenciamento de recursos naturais. Em ambientes urbanos, ele auxilia no planejamento territorial e na identificação de áreas de risco. Já em contextos agrícolas, permite avaliar a saúde da vegetação, monitorar safras e otimizar o uso de recursos como água e fertilizantes.

O SR integra diferentes áreas do conhecimento, como física, matemática, computação e geociências, para interpretar e analisar as informações adquiridas. Ele envolve desde o entendimento dos princípios da radiação eletromagnética até o uso de técnicas computacionais avançadas, como aprendizado de máquina, para extração e processamento de informações complexas. Isso torna o sensoriamento remoto uma ferramenta interdisciplinar e indispensável para a tomada de decisões em diversos setores, contribuindo significativamente para o desenvolvimento sustentável e a preservação ambiental.

2.1.1 Satélites multiespectrais

As imagens de sensoriamento remoto são capturadas por meio de diferentes tipos de sensores, que podem estar a bordo de plataformas aéreas, como drones e aviões, ou em plataformas orbitais, como satélites. Este trabalho foca nos sensores a bordo de satélites, devido à sua capacidade de monitorar grandes áreas com alta frequência temporal e cobertura global.

Conforme destacado por Gomes and Cubas (2021), o sensoriamento remoto, que obtém informações por meio de sensores a bordo de satélites, tem a grande vantagem de poder coletar grandes áreas da superfícies da Terra em curto tempo, com grande repetitividade e baixo custo para o usuário. Além disso, segundo Meneses and Almeida (2012), essa forma de cobertura repetitiva, obtendo imagens periódicas de qualquer área do planeta, propicia detectar e monitorar mudanças que acontecem na superfície terrestre. Razão pela qual as imagens de satélites passaram a ser a mais eficiente ferramenta para uso nas aplicações que envolvem análises ambientais dos diversos ecossistemas terrestres (Gomes and Cubas, 2021).

As imagens capturadas por satélites são divididas em diferentes bandas espectrais, que representam segmentos do espectro eletromagnético delimitados por dois comprimentos de onda específicos. Cada banda fornece informações complementares sobre os objetos observados, permitindo a análise detalhada de suas propriedades físicas. Por exemplo:

- **Bandas no espectro visível (como o vermelho, verde e azul)**: são úteis para identificar vegetação e superfícies urbanas.
- **Bandas no infravermelho próximo**: são amplamente utilizadas para monitorar o vigor da vegetação e estimar índices como o NDVI (Normalized Difference Vegetation Index).
- **Bandas no infravermelho médio**: auxiliam na identificação de características relacionadas à umidade do solo e detecção de queimadas.

Satélites que operam com múltiplas bandas espectrais são denominados multiespectrais. Eles possuem sensores projetados para capturar diferentes faixas do espectro eletromagnético, cada uma voltada a aplicações específicas, como mapeamento agrícola, análise de corpos d'água, estudo de solos e monitoramento ambiental.

Além disso, os avanços recentes na tecnologia de sensoriamento remoto têm expandido significativamente as capacidades dos sensores multiespectrais, permitindo resoluções espacial, temporal e espectral cada vez mais refinadas. A resolução espacial aprimorada possibilita a detecção de objetos e fenômenos com maior detalhamento, enquanto a alta resolução temporal permite a captura frequente de mudanças dinâmicas na superfície terrestre. Já a resolução espectral estendida viabiliza uma análise mais precisa das propriedades dos materiais, ao cobrir uma ampla gama de comprimentos de onda.

Esses avanços têm ampliado o uso dessas ferramentas em aplicações complexas e críticas, como a detecção de mudanças sutis na cobertura vegetal, onde se analisa a saúde e o vigor da vegetação em função de estresses ambientais, como secas ou pragas. No monitoramento de recursos hídricos, sensores multiespectrais auxiliam na avaliação da qualidade da água, estimando parâmetros como turbidez, concentração de sedimentos e clorofila. Além disso, no contexto de impactos ambientais causados por atividades humanas, essas ferramentas são essenciais para identificar e monitorar desmatamentos, queimadas, expansão urbana e outras intervenções antropogênicas que afetam ecossistemas naturais.

Portanto, os sensores multiespectrais consolidam-se como tecnologias indispensáveis no campo do sensoriamento remoto, devido à sua versatilidade, precisão e capacidade de oferecer dados fundamentais para o monitoramento e análise detalhada dos diversos ecossistemas terrestres.

2.1.2 Sentinel-2

Atualmente, uma constelação de satélites está em operação para atender às demandas de uma ampla gama de usuários. Cada satélite é projetado para oferecer imagens com características geométricas, espectrais e temporais específicas, adaptadas às necessidades de diferentes aplicações. Entre as missões de destaque está a Sentinel-2, desenvolvida e operada pela Agência Espacial Europeia (ESA) como parte do programa Copernicus [1].

A missão Sentinel-2 fornece imagens multiespectrais de alta resolução e ampla cobertura territorial, oferecendo suporte a uma variedade de serviços e aplicações, incluindo monitoramento agrícola, gerenciamento de emergên-

[1] https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-2/overview

Tabela 2.1: Bandas espectrais da Sentinel-2

Nome	Tam. Pixel	Comprimento de Onda	Descrição
B1	60 m	443.9nm (S2A) / 442.3nm (S2B)	Aerossóis
B2	10 m	496.6nm (S2A) / 492.1nm (S2B)	Azul
B3	10 m	560nm (S2A) / 559nm (S2B)	Verde
B4	10 m	664.5nm (S2A) / 665nm (S2B)	Vermelho
B5	20 m	703.9nm (S2A) / 703.8nm (S2B)	Borda Vermelha 1
B6	20 m	740.2nm (S2A) / 739.1nm (S2B)	Borda Vermelha 2
B7	20 m	782.5nm (S2A) / 779.7nm (S2B)	Borda Vermelha 3
B8	10 m	835.1nm (S2A) / 833nm (S2B)	NIR
B8A	20 m	864.8nm (S2A) / 864nm (S2B)	Borda Vermelha 4
B9	60 m	945nm (S2A) / 943.2nm (S2B)	Vapor d'água
B10	60 m	1373.5nm (S2A) / 1376.9nm (S2B)	Cirro
B11	20 m	1613.7nm (S2A) / 1610.4nm (S2B)	SWIR 1
B12	20 m	2202.4nm (S2A) / 2185.7nm (S2B)	SWIR 2

cias, classificação de cobertura do solo e avaliação da qualidade da água. A constelação é composta por dois satélites idênticos, Sentinel-2A e Sentinel-2B, que operam simultaneamente em fases de 180° um do outro, em órbitas sincronizadas com o Sol, a uma altitude média de 786 km. Essa configuração permite uma alta frequência de revisita, cobrindo o mesmo local a cada cinco dias no Equador, o que é ideal para o monitoramento contínuo de mudanças dinâmicas.

Cada satélite da missão Sentinel-2 carrega um instrumento multiespectral (MSI - Multispectral Instrument), projetado para capturar imagens em 13 bandas espectrais, que abrangem o espectro visível (VIS), o infravermelho próximo (VNIR) e o infravermelho de ondas curtas (SWIR). Essas bandas são distribuídas com diferentes resoluções espaciais:

- **Quatro bandas com resolução de 10 metros**: adequadas para mapeamento detalhado de áreas urbanas, vegetação e corpos d'água.
- **Seis bandas com resolução de 20 metros**: ideais para análises agrícolas e estudos ambientais de médio detalhamento.
- **Três bandas com resolução de 60 metros**: projetadas para correções atmosféricas e análise de larga escala.

A Tabela 2.1 apresenta uma descrição detalhada das bandas espectrais disponíveis na missão Sentinel-2, destacando seus comprimentos de onda centrais, resoluções espaciais e aplicações típicas. Essas especificações tornam a missão uma das mais versáteis e amplamente utilizadas na comunidade científica e técnica.

2.2 Aprendizagem profunda

A aprendizagem profunda é uma subárea da aprendizagem de máquina que se concentra no desenvolvimento de modelos compostos por várias camadas de processamento, capazes de aprender representações hierárquicas a partir dos dados. O termo "profunda"refere-se à profundidade dessas arquiteturas, muitas vezes compostas por dezenas ou centenas de camadas interconectadas (Chollet, 2021).

Diferente de abordagens tradicionais, onde as características dos dados precisam ser definidas manualmente, a aprendizagem profunda permite que os modelos aprendam automaticamente representações úteis diretamente dos dados brutos. Essa capacidade tem impulsionado avanços notáveis em áreas como reconhecimento visual, processamento de linguagem natural, saúde, robótica e muito mais, especialmente em problemas que envolvem grandes volumes de dados e padrões complexos (Sarker, 2021).

Entre as vantagens dessa abordagem estão:

- **Escalabilidade**: Modelos profundos podem lidar eficientemente com grandes quantidades de dados, explorando a riqueza de informações disponíveis.

- **Transferência de conhecimento**: Redes profundas podem ser pré-treinadas em tarefas gerais e adaptadas a problemas específicos, reduzindo o esforço computacional e os requisitos de dados.

- **Hierarquias representacionais**: A habilidade de aprender características em diferentes níveis de abstração permite capturar desde padrões simples (bordas, texturas) até conceitos complexos (formas, objetos inteiros).

Essa abordagem é implementada através de diversas arquiteturas, sendo as Redes Neurais Convolucionais (RNCs) uma das mais proeminentes para tarefas de visão computacional(Zhang et al., 2023).

2.3 Redes neurais convolucionais

As Redes Neurais Convolucionais (RNCs) são projetadas especificamente para processar dados estruturados em grades, como imagens. Diferentemente de redes totalmente conectadas, onde cada neurônio é conectado a todos os outros, as RNCs utilizam conexões locais e pesos compartilhados, tornando-as mais eficientes para tarefas visuais.

- **Camadas de convolução**: Essas camadas são responsáveis por aplicar filtros (*kernels*) que capturam padrões locais, como bordas, texturas e formas. O uso de múltiplos filtros em uma única camada permite que o modelo aprenda diversas características da imagem.

- **Camadas de *pooling***: Essas camadas reduzem a resolução espacial dos mapas de características, agregando informações de regiões locais. Isso ajuda a diminuir o custo computacional e torna o modelo mais robusto a pequenas variações nos dados de entrada.

- **Funções de ativação**: Elementos não-lineares, como ReLU (*Rectified Linear Unit*), introduzem não-linearidade nas camadas convolucionais, permitindo que as redes aprendam relações complexas nos dados.

Essas propriedades tornam as RNCs invariantes a translações e capazes de aprender padrões hierárquicos, que são fundamentais para identificar objetos em imagens. Arquiteturas como AlexNet (Krizhevsky et al., 2012), VGG (Simonyan and Zisserman, 2015), ResNet (He et al., 2015) e EfficientNet (Tan and Le, 2019) representam marcos no campo, cada uma introduzindo avanços significativos em termos de precisão e eficiência computacional.

As RNCs têm sido amplamente empregadas em aplicações práticas como:

- **Sensoriamento remoto**: Monitoramento de florestas, detecção de mudanças ambientais e análise agrícola (Kattenborn et al., 2021; Osco et al., 2021b).

- **Saúde**: Diagnósticos assistidos por computador, como detecção de tumores em imagens médicas (Jia et al., 2024).

- **Automação**: Sistemas de visão para veículos autônomos e robótica (Grigorescu et al., 2019; Pierson and Gashler, 2017).

A versatilidade dessas redes, combinada com sua eficiência computacional, continua a impulsionar inovações em diversas áreas.

2.4 Segmentação semântica

A segmentação semântica é uma tarefa fundamental na visão computacional, focada em atribuir um rótulo a cada *pixel* de uma imagem, diferenciando objetos e regiões. Sendo diferente de métodos que apenas classificam objetos em regiões delimitadas (como a detecção de objetos). A segmentação semântica fornece uma visão mais detalhada, fornecendo informações mais precisas

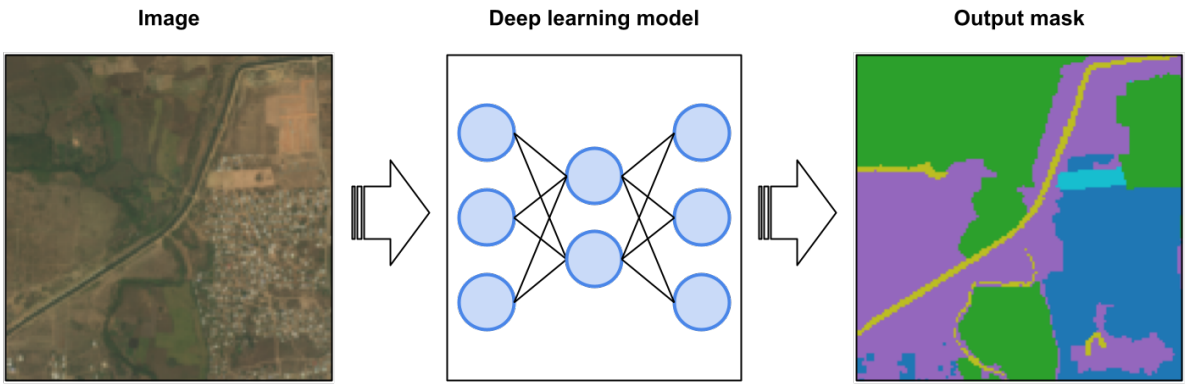

Figura 2.1: Ilustração do processo típico de segmentação semântica através de deep learning, as imagens de satélite e máscara são fornecidas pelo dataset DeepGlobe.

sobre a localização e as bordas dos objetos. Esta tarefa é crítica para aplicações que exigem compreensão detalhada das cenas, como análise urbana e navegação autônoma e mapeamento ambiental.

A Figura 2.1 apresenta um exemplo visual de segmentação semântica aplicada com técnicas de *deep learning*. Como ilustra a imagem, o processo típico envolve a entrada de uma imagem bruta em um modelo treinado, que gera como saída uma máscara de predição. Essa máscara consiste em uma representação pixel a pixel da cena, onde cada pixel está associado a uma categoria predita, como "construções", "pastagens", e "corpos de água".

2.4.1 Evolução dos métodos

Os primeiros métodos de segmentação baseados em redes convolucionais adaptaram arquiteturas de classificação de imagens para realizar predições densas, substituindo as camadas totalmente conectadas por convoluções. Contudo, essas abordagens iniciais enfrentavam desafios significativos, como perda de resolução espacial e limitações na inferência multiescala.

A Fully Convolutional Network (FCN) (Long et al., 2015) marcou uma mudança paradigmática ao propor um modelo fim-a-fim para segmentação semântica. A FCN introduziu (a) conexões de salto (*skip connections*) para combinar informações de diferentes níveis de abstração e (b) predições densas em resolução total, preservando informações espaciais essenciais.

Nos anos subsequentes, outros métodos de segmentação semântica foram desenvolvidos visando aprimorar cada vez mais o estado-da-arte, entre eles, destacam-se:

- **U-Net**: originalmente proposta para segmentação biomédica (Ronneberger et al., 2015), tornou-se um padrão devido à sua arquitetura simétrica de *encoder-decoder* e conexões de salto, que preservam detalhes espa-

ciais enquanto reconstruem saídas de alta resolução. Sua eficácia em problemas de sensoriamento remoto a tornou amplamente utilizada em aplicações como segmentação de cobertura terrestre e corpos d'água.

- **Deeplab**: introduz convoluções dilatadas e módulos de pooling atrous para capturar informações em diferentes escalas sem reduzir a resolução espacial, e incorporam CRFs (*Conditional Random Fields*) para refinar bordas e melhorar a segmentação (Chen et al., 2017).
- **Segformer**: esse método combina a eficiência de transformadores com convoluções leves, permitindo representar contextos globais e locais simultaneamente. Essa abordagem é especialmente útil para cenários complexos e variados, como imagens de alta resolução (Xie et al., 2021).

CAPÍTULO

3

Trabalhos relacionados

Nos últimos anos, um número crescente de trabalhos foram publicados na literatura, que estudam a segmentação semântica de eucalipto utilizando redes neurais profundas. Este crescimento elevado pode ser explicado principalmente pelas novas iniciativas de dados abertos, como o programa Landsat, Sentinel e Google Earth Engine, a crescente disponibilidade de recursos computacionais, as aplicações crescentes da aprendizagem profunda para o sensoriamento remoto, e a evolução dos métodos de segmentação semântica.

Este capítulo dedica-se a pesquisa destes trabalhos, a primeira seção analisa os trabalhos relacionados citando suas características principais e realizando alguns comentários, a segunda seção discute sobre algumas limitações encontradas, enquanto que a última seção apresenta os diferenciais do trabalho atual em relação a estes.

3.1 Redes profundas para segmentação semântica de eucalipto

Nesta seção, listaremos e detalhamos resumidamente os trabalhos relevantes da literatura que visam a *segmentação semântica de florestas de eucalipto utilizando aprendizagem profunda*.

Os autores Wagner et al. (2019) aplicaram a rede convolucional U-Net para o mapeamento de florestas naturais, plantações de eucaliptos, e uma espécie indicadora de distúrbios florestais, Cecropia hololeuca, na região da Mata Atlântica brasileira, utilizando imagens de alta resolução (0.3m) das bandas RGB do World-View-3, eles segmentaram os tipos de floresta com uma acurá-

cia geral acima de 95% e uma interseção sobre a união (IoU) de 96%. Para C. hololeuca, a precisão geral foi de 97% e o IoU foi de 86%. Vale notar que a alta resolução das imagens pode ser benéfica para a segmentação já que permite observar mais detalhes para a diferenciação das classes.

Ferreira et al. (2019) realizaram segmentação de florestas de áreas naturais e plantações de eucaliptos no sudeste do Brasil, ao utilizar imagens da Landsat-TM e dados do projeto MapBiomas, eles comparam RNC e Floresta Aleatória, alcançando 3% de erro. Uma limitação está no fato que a arquitetura de RNC proposta possui uma arquitetura estruturalmente simples, desenhada para prever *patches* de dimensões reduzidas, podendo causar *overhead* computacional ao ser utilizado em larga escala. Vale a pena notar que apenas duas classes foram consideradas no estudo (áreas de floresta natural e plantações de eucalipto), sendo que os pixels de fundo não influenciaram na avaliação da acurácia.

Zhao et al. (2019) avaliaram diversos modelos de AP para a classificação precoce de diversas culturas (arroz, cana-de-açúcar, banana, abacaxi e eucalipto) utilizando séries temporais, situadas nos condados de Suixi e Leizhou na cidade de Zhanjiang, China. De acordo com os autores essa região é caracterizada por possuir dias nublados frequentes, para contornar esta situação, eles utilizaram o radar de abertura sintética (SAR, do inglês Synthetic Aperture Radar) Sentinel-1A, que é menos afetado pelo ruído das nuvens. Os autores avaliaram sistematicamente três métodos de AP, RNCs 1D, LSTM RNNs e GRU RNNs, e os resultados mostram que apesar de todos métodos se provarem efetivos para essa tarefa, as RNCs 1D alcançaram melhor coeficiente Kappa de 0.942.

Dias et al. (2020) segmentaram plantações de eucalipto em vários pontos da América do Sul, utilizando imagens da MODIS, eles codificaram as séries de tempo NDVI para matrizes campo de soma angular de Gramiam (GASF) campo de diferença angular de Gramian (GADF) e campo de transição de Markov (MTF), em seguida extraíram características discriminatórias com RNCs, que foram classificadas por diversos classificadores, apresentando média de acurácia acima de 90%.

Forstmaier et al. (2020) mapearam eucalipto ao longo em Portugal e partes de Espanha com foco em áreas da Natura 2.000 dentro de Portugal, que estão protegidas pelas diretivas europeias sobre aves e habitats. O método permite a detecção de pequenas populações incipientes, bem como de populações mistas fora do ambiente regular de plantação. Eles utilizaram imagens multiespectrais da Sentinel-2 e mapas de *ground truth* para realizar o treinamento das FNNs (Feedforward Neural Networks). Esses modelos prevêem árvores de eucalipto cobertura com sensibilidade de até 75.7% e especificidade

de até 95.8%, alcançando uma acurácia geral de 92.5%.

Firigato et al. (2021) investigaram o potencial da integração de algoritmos de AP com o Google Earth Engine (GEE) no contexto do mapeamento de eucalipto na savana brasileira. Os autores utilizaram a tradicional U-Net obtendo uma acurácia geral de 96.88% e uma IoU de 84%. Confirmando a importância da integração entre o GEE e a aprendizagem profunda.

da Costa et al. (2021) realizaram a segmentação semântica de áreas de arborização de Eucalipto em 3 municípios brasileiros, utilizando imagens multiespectrais da Sentinel-2, eles utilizaram 10 bandas espectrais que originalmente estavam em 10 m e 20 m de resolução espacial, redimensionando-as para 10m de resolução espacial, os caminhos de dimensões (160, 160, 10) foram criados através de uma técnica de janela deslizante com passo, após análise comparativa com combinações entre 6 arquiteturas e 4 codificadores, eles relataram uma melhor intersecção sobre união (IoU) de 76.57% para a arquitetura DeepLabv3+ com o backbone Efficient-Net-b7, e em uma análise posterior demonstraram que os resultados do teste melhoram progressivamente com o redução do tamanho do passo.

Boston et al. (2024) realizaram o mapeamento de vegetações naturais e classes de florestas (incluindo eucalipto) com base imagens compostas anuais de refletância geomediana Landsat no sudeste da Austrália. Este estudo foi focado na aplicação de dois principais métodos identificados na literatura (RNC e Floresta Aleatória). A melhor RNC (U-Net) foi gerada usando seis bandas geomedianas do Landsat como entrada e produziram melhores resultados do que um algoritmo de floresta aleatória baseada em pixels, com maior precisão geral (OA) e pontuação F1 média ponderada para todas as classes de vegetação (93 vs 87% em ambos os casos) e um índice Kappa mais elevado (86 vs. 74%). Ao analisar os dados espectrais, este estudo deu o primeiro passo ao verificar que o eucalipto possui níveis de reflectância distintas das outras classes, concluindo que as características espectrais provavelmente desempenharão um papel papel na detecção pela RNC, além de características espaciais ou texturais para essas classes.

A tabela 3.1 sumariza as principais características de cada trabalho.

Autor	Local	Satélite	Dados de entrada	Melhor modelo	Melhor resultado
(Wagner et al., 2019)	Mata atlântica brasileira	World-View-3	Imagens RGB (0.3m)	U-Net	Acc >95%; IoU: 96%
(Ferreira et al., 2019)	Sudeste do Brasil	Landsat-TM	Imagens RGB	RNC 2D	Erro: 3%
(Zhao et al., 2019)	Suixi/Leizhou - China	Sentinel-1A	Séries temporais SAR	RNC 1D	Kappa: 94.2%
(Dias et al., 2020)	América do Sul	MODIS	Codificação de séries NDVI	DenseNet/ ResNet	Acc >90%
(Forstmaier et al., 2020)	Península Ibérica	Sentinel-2	Imagens multiespectrais	FNN	Acc: 92.5% Specificity: 95.8% Recall: 75.7%
(Firigato et al., 2021)	Savana Brasileira	Sentinel-2	Imagens multiespectrais	U-Net	Acc: 96.88%; IoU: 84%
(da Costa et al., 2021)	Brasil	Sentinel-2	Imagens multiespectrais	Deeplabv3+ (Eff-NetB7)	IoU: 76.57%
(Boston et al., 2024)	Sudeste da Austrália	Landsat	Bandas geomedianas / Rótulos de baixa resolução	U-Net	F-Score: 93%

Tabela 3.1: Principais características dos trabalhos citados.

3.2 Limitações encontradas

Apesar dos progressos registados, o trabalho apresenta desafios significativos em termos de escalabilidade e eficiência computacional. Por exemplo, certos métodos propostos enfrentam dificuldades quando aplicados a áreas maiores ou mais complexas, limitando seu uso prático em operações de larga escala, por exemplo, o trabalho de Wagner et al. (2019) utilizou satélites de alta resolução, que se caracterizam por ter dados mais pesados para processar. Percebemos também que a utilização da U-Net em sua versão tradicional por alguns trabalhos (Firigato et al. (2021), Boston et al. (2024), Wagner et al. (2019)) pode aumentar consideravelmente o custo computacional de treinamento e inferência, resultando em uma alta demanda por recursos computacionais, inviabilizando a implementação em dispositivos com menor capacidade ou em aplicações que necessitem de processamento em tempo real. Além disso, trabalhos como Zhao et al. (2019) e Dias et al. (2020) utilizam dados temporais, essa abordagem tem a desvantagem de ter um alto custo computacional, exigindo um volume considerável de dados temporais rotulados. Estas limitações podem influenciar negativamente a escalabilidade e eficiência computacional dos modelos, o que exige novas soluções inovadoras.

3.3 Diferencial deste trabalho

O presente trabalho procura superar estas limitações, propondo uma abordagem baseada na modificação de uma U-Net, combinada com experiências de teste de generalização espacial e temporal. O objetivo é não só validar a eficácia do modelo numa variedade de condições, mas também melhorar a sua eficiência computacional, aumentando o seu potencial de aplicação prática. Dada a falta de dados publicamente disponíveis na literatura para a segmentação de eucaliptos, criamos um novo conjunto de dados para a segmentação semântica de eucaliptos usando dados do Sentinel-2 em diferentes cidades do bioma Cerrado brasileiro, preenchendo uma lacuna existente na literatura e fornecendo uma base sólida para trabalhos futuros.

Além disso, em um esforço para entender melhor o processo de tomada de decisão do modelo, o presente trabalho explora a adaptação do método de oclusão de sensibilidade introduzido por Zeiler and Fergus (2014) com o objetivo de analisar a importância das bandas. Essas contribuições buscam preencher o gargalo existente na literatura e contribuir para o desenvolvimento de soluções mais robustas e escaláveis para a segmentação semântica de florestas de eucalipto.

CAPÍTULO

4

U-Net compacta para a segmentação semântica de eucalipto

4.1 Arquitetura U-Net

A arquitetura original U-Net, originalmente proposta para o problema de segmentação biomédica (Ronneberger et al., 2015), consiste em um caminho de contração que permite capturar contexto e reduzir a resolução espacial e um caminho de expansão que usa a informação do caminho de contração para gerar a máscara de segmentação, permitindo a localização precisa dos objetos, os dois caminhos são simetricamente conectados lembrando o formato de U (Ronneberger et al., 2015). Esta arquitetura está ilustrada na Figura 4.1.

No caminho de contração, a rede captura o contexto da imagem, ela é formada por camadas de convolução, pooling e funções de ativação ReLU, os mesmos componentes vistos no capítulo 2, seção 2.3.

Já no caminho de expansão, a rede deve aprender a reconstruir os pixels em suas localizações originais utilizando as características do caminho de contração, para gerar a máscara de segmentação. Aqui, as convoluções com ativação ReLU também são aplicadas para refinar a reconstrução. Adicionalmente, vemos os seguintes componentes:

- **Convoluções transpostas**: Responsáveis por aumentar a resolução da imagem de entrada, também chamadas de camadas de *upsampling*.

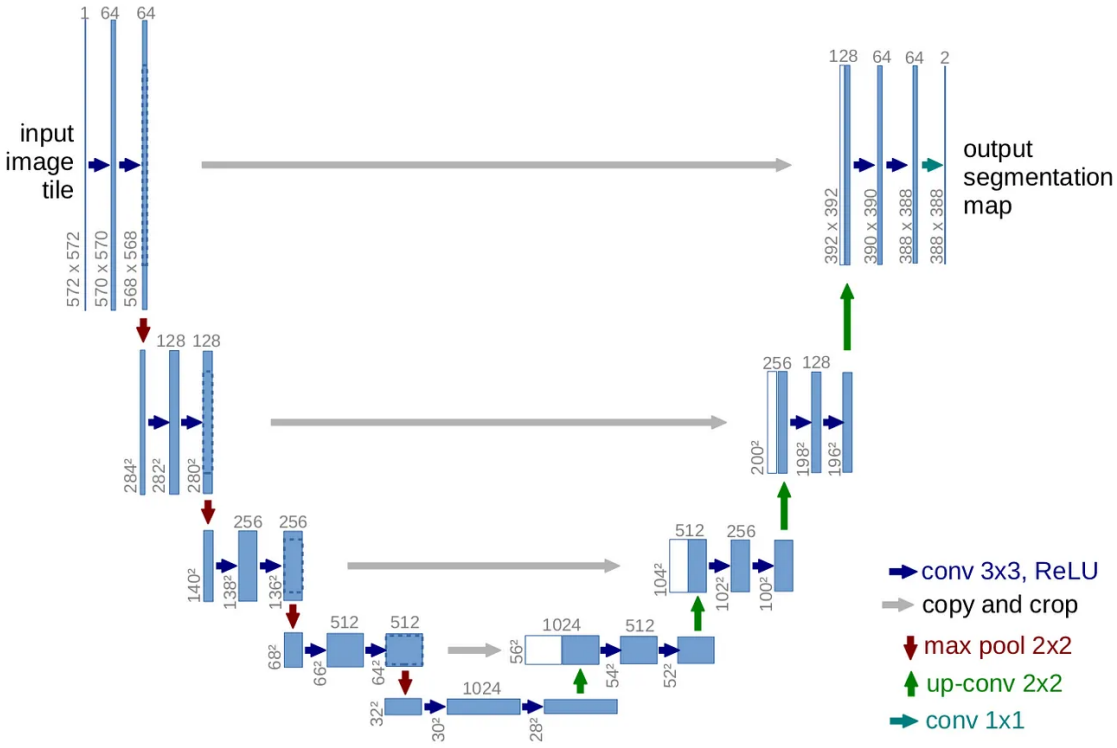

Figura 4.1: Visão geral da arquitetura original de U-Net.

- ***Skip connections***: Responsáveis por conectar os recursos correspondentes do caminho de contração para recuperar detalhes perdidos durante a redução da resolução espacial. Estas conexões consistem em uma das características fundamentais introduzidas na U-Net.

Ao final da rede temos a camada de saída em que é gerada o mapa de segmentação. Onde de há uma convolução 1 x 1 que reduz o número de filtros para a quantidade necessária para a máscara de segmentação (1 para segmentação binária ou n para múltiplas classes).

4.2 Modificações propostas

Com o objetivo de criar uma arquitetura mais simples e leve, e ao mesmo tempo eficiente para o problema de segmentação semântica de eucalipto, nós propomos uma U-Net *lightweight* cuja arquitetura é ilustrada na Figura 4.2.

Como pode ser visto, a rede proposta é mais compacta em relação ao nível de profundidade e número de filtros do que a U-Net original. No codificador, aplicamos camadas de convolução com stride 2 para reduzir a dimensionalidade, o uso de convoluções ao invés do max-pooling original permite reter informações importantes, diferentemente dos max-poolings elas possuem parâmetros treináveis ajudando na preservação dos pequenos detalhes enquanto que ajudam na suavização do fluxo de gradientes.

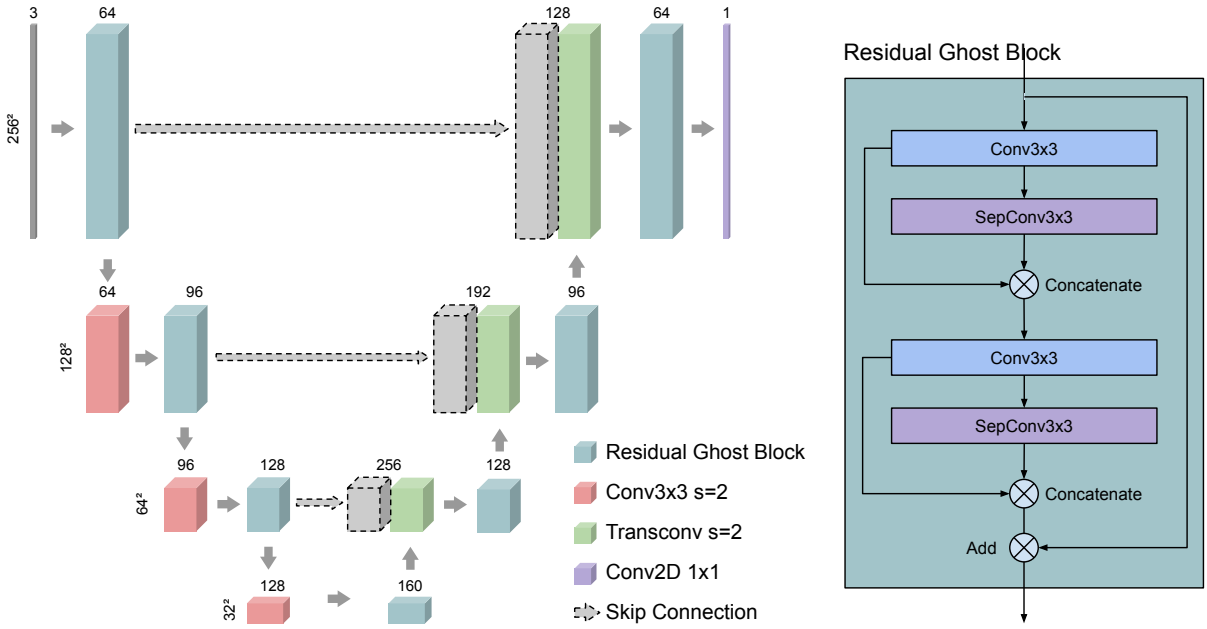

Figura 4.2: Visão geral da arquitetura de U-Net compacta proposta.

A principal modificação na arquitetura consiste nos blocos fantasmas residuais, que permitem reduzir mais parâmetros e computações, além de prevenir o desaparecimento do gradiente e acelerar a convergência do modelo.

O bloco fantasma residual é destacado na Figura 4.2 à direita. Como pode ser visto, as características de entrada passam por dois módulos fantasmas (do inglês, *Ghost Modules*). Cada módulo divide a camada convolucional original em duas partes, a primeira parte consiste em uma operação de convolução tradicional utilizando menos filtros para gerar vários mapas de características intrínsecas. Na segunda parte, um certo número de operações de transformação de baixo custo é aplicado para gerar mapas de características fantasmas. Esta modificação permite evitar redundância nos mapas de características, no sentido que alguns mapas são muito similares em pares, como se um par fosse o "fantasma"de outro. A idéia é obter um mapa através de outro com operações mais baratas computacionalmente. Os módulos *ghost* são uma boa alternativa para camadas de convolução que permitem simplificar as operações e alcançar maior desempenho em tarefas de reconhecimento de imagens (Han et al., 2020).

Adicionalmente, estes blocos possuem uma *skip connection*, onde as características de saída são somadas as de entradas, esta característica permite melhorar o desempenho e a robustez da U-Net, facilitando o fluxo de informação e gradiente sem adicionar nenhum parâmetro extra ou complexidade computacional (He et al., 2015). Apesar de não representado no desenho da arquitetura a fins de simplicidade, cada convolução é seguida por uma normalização de lote (BN) (Ioffe and Szegedy, 2015) e uma tradicional função de

ativação ReLU (*rectified linear unit*), onde também utilizamos a técnica de *padding* para manter o tamanho da entrada em cada camada.

Ao integrar *skip-connections*, módulos fantasmas, e as modificações propostas, a U-Net proposta se torna mais poderosa, eficiente e capaz de fornecer segmentação de alta qualidade com menor custo computacional. Neste trabalho, nós designamos esta rede como ResGhostU-Net (Residual Ghost U-Net).

CAPÍTULO

5

Metodologia

5.1 Região de Interesse

Para este estudo, quatro cidades foram escolhidos, essas cidades possuem grande concentração de plantações de eucalipto e estão situadas no bioma cerrado brasileiro. A Figura 5.1 mostra o mapa do Brasil com as cidades de interesse destacadas. Enquanto que a Figura 5.2 apresenta as imagens de satélite das cidades.

As cidades (a) Àgua Clara, (b) Selvíria e (c) Três Lagoas, estão localizadas no estado de Mato Grosso do Sul (MS), na região Centro-Oeste do Brasil. O clima predominante é tropical, com verões quentes e chuvosos e invernos mais secos, típico da região do Centro-Oeste brasileiro. Enquanto que a cidade (d) Três Marias, é localizada no estado de Minas Gerais, na região Sudeste do Brasil. Ela possui uma localização estratégica, que marca a transição entre a região Sudeste e o bioma do Cerrado. O clima é tropical, com verões quentes e chuvosos, e invernos mais amenos e secos.

5.2 Pré-processamento

Para o pré-processamento das imagens, utilizamos a plataforma Google Earth Engine (GEE), que combina um catálogo de vários petabytes de imagens de satélite e conjuntos de dados geoespaciais com recursos de análise em escala planetária[1].

As imagens foram adquiridas do produto Harmonized Sentinel-2 MSI: Mul-

[1] https://earthengine.google.com/

Figura 5.1: Mapa com as cidades de interesse. As cidades escolhidas estão marcadas no mapa com pinos vermelhos.

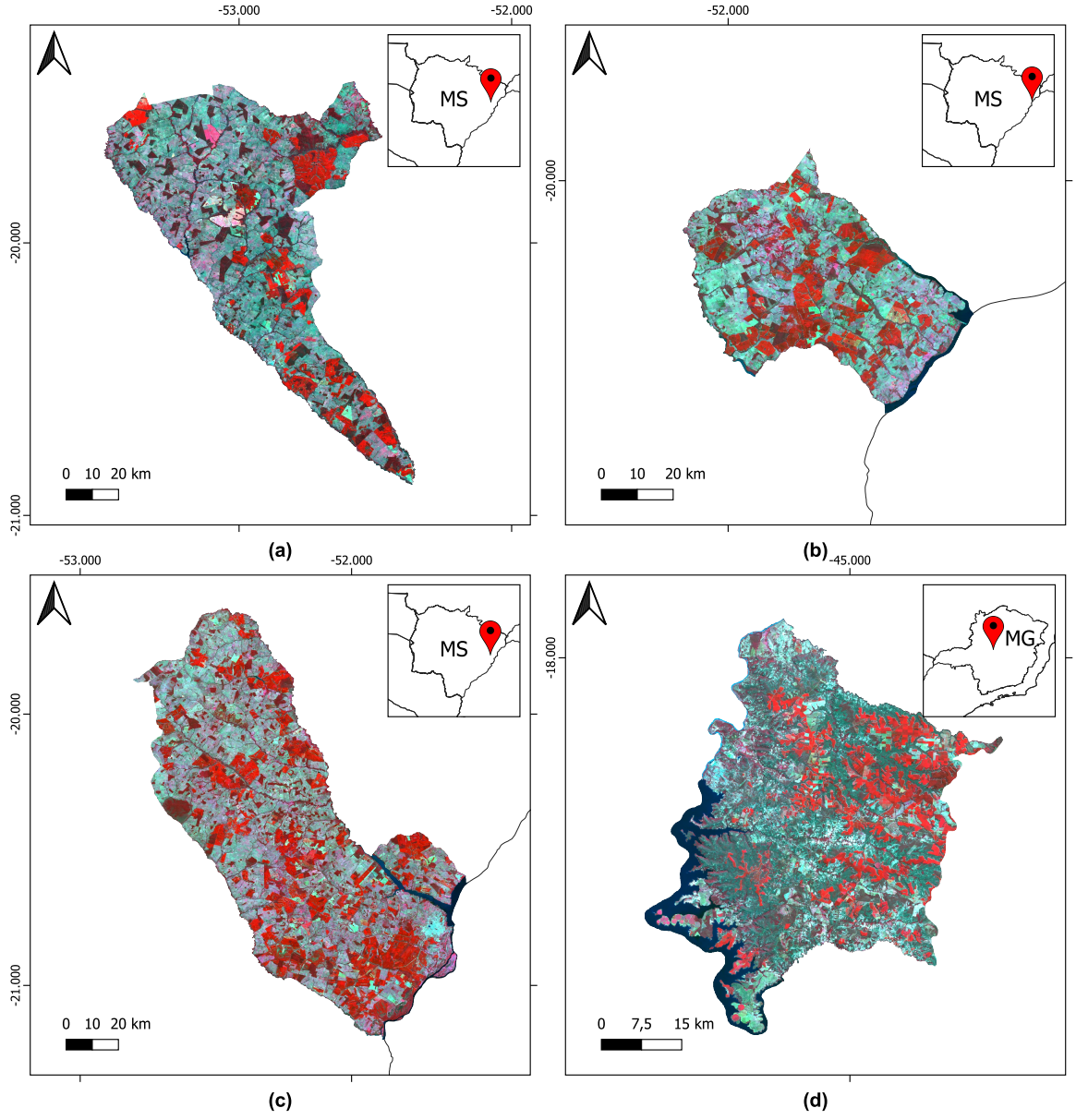

Figura 5.2: Composições de imagens Sentinel-2 em falsa cor da região de Interesse.

tiSpectral Instrument, Level-1C, disponível no catalogo de dados do GEE[2], as imagens contém 13 bandas espectrais representando refletância de superfície, um conjunto de bandas especificas do produto, e 3 bandas de avaliação de qualidade.

As máscaras de *ground-truth* consistem em anotações manuais de polígonos de eucalipto realizadas por especialistas em composições de imagens de satélite Sentinel-2, utilizando a ferramenta QGIS[3].

Primeiramente, na plataforma GEE, filtramos a coleção adquirida de imagens da Sentinel-2 de 3 maneiras:

- Filtragem por região: A coleção de imagens foi filtrada para abranger as cidades de interesse.

- Filtragem por data: A coleção de imagens foi filtrada pelo seguinte intervalo de datas: <YEAR>-03-01 - <YEAR>-07-30, em que YEAR é o ano de 2016 para as cidades e MS e 2023 para a cidade de Três Marias em MG;

- Filtragem por percentagem de nuvens: A coleção de imagens foi filtrada para conter apenas as imagens que têm menos de 20% de percentagem de pixel nebuloso.

Após a filtragem das imagens, para melhorar a análise das imagens, as imagens do Sentinel-2 passaram pelo procedimento de mascaramento de nuvens e cirros. A função utilizada utiliza a banda QA60, uma banda de 16 bits que codifica informações de qualidade, incluindo nuvens e cirros, usando bits específicos[4].

Como última etapa de pré-processamento, usamos a função .median() do GEE para reduzir a coleção de imagens adquiridas. Esta operação permite-nos criar uma imagem composta calculando a mediana de todos os valores de cada pixel na pilha de todas as bandas correspondentes. O uso de uma mediana é particularmente eficaz para lidar com o ruído das nuvens, pois os valores discrepantes causados pela presença de nuvens ou cirros são naturalmente atenuados. Além disso, a função ajuda a suavizar variações temporais indesejadas.

[2] https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_HARMONIZED

[3] http://www.qgis.org

[4] O algoritmo de mascaramento utilizado neste trabalho está disponível no seguinte link: https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_HARMONIZED

City set	Used for	Year	# Imgs	# Pixels		
				Eucalypt	Background	Total
Água Clara (MS)	Training	2016	1402	10 099 224	79 478 070	89 577 294
Selvíria (MS)	Validation	2016	595	8 458 499	30 535 421	38 993 920
Três Lagoas (MS)	Test	2016	1809	19 842 992	95 152 488	114 995 480
Três Marias (MG)	Test	2023	269	4 087 259	13 541 925	17 629 184

Tabela 5.1: Propriedades do dataset

5.3 Geração do dataset

Após o pré-processamento das imagens multiespectrais e máscaras de eucalipto, as imagens foram exportadas e processadas em Python para gerar diferentes subconjuntos para o treinamento e validação dos modelos. Todas as imagens do dataset possuem dimensões 256 x 256 x 13. As 13 bandas foram reamostradas para 10 m de resolução espacial. E para cada imagem, existe uma máscara de *ground-truth* anotada por especialistas. A Tabela 5.1 apresenta algumas propriedades do dataset.

5.4 Métodos para comparação

Neste trabalho, nós comparamos o método proposto com 6 métodos populares de segmentação semântica de imagens (FPN, PSPNet, DeepLabv3+, MANet, UPerNet e SegFormer), descritos a seguir:

- **FPN (Feature Pyramid Network)**: Este método utiliza uma pirâmide de recursos para lidar com objetos de diferentes escalas. Ele é projetado para capturar características multiescala de forma hierárquica, permitindo segmentações mais precisas, especialmente em cenários com variação de tamanho dos objetos (Lin et al., 2017).

- **PSPNet (Pyramid Scene Parsing Network)**: Focado em capturar contextos globais da cena, o PSPNet utiliza pooling piramidal para agregar informações de diferentes escalas. Isso o torna particularmente adequado para tarefas onde o contexto global desempenha um papel crítico na segmentação (Zhao et al., 2017).

- **DeepLabv3+**: Um modelo robusto que combina atrous convolutions (dilated convolutions) e módulos de pooling piramidal espacial para melhorar a captura de detalhes sem comprometer o campo de visão. Ele é amplamente reconhecido por seu desempenho superior em benchmarks de segmentação (Chen et al., 2017).

- **MANet (Multi-scale Attention Network)**: Este método se destaca por integrar mecanismos de atenção multiescala, permitindo que o modelo foque em regiões importantes da imagem em diferentes escalas, o que é crucial para imagens com objetos complexos e heterogêneos (Xu et al., 2021).
- **UPerNet (Unified Perceptual Parsing Network)**: Projetado para tarefas gerais de segmentação, o UPerNet utiliza um backbone com representações multiescala e uma cabeça de segmentação unificada, garantindo bom equilíbrio entre precisão e eficiência (Xiao et al., 2018).
- **SegFormer**: Um modelo moderno que combina eficiência e precisão ao utilizar transformadores para aprender representações globais e convoluções leves para refinar os detalhes. Ele é altamente adaptável a diferentes resoluções de entrada e eficiente em termos de tempo de inferência (Xie et al., 2021).

Esses métodos foram escolhidos para comparação devido à sua popularidade e ao desempenho reconhecido em tarefas de segmentação semântica.

5.5 Backbone

A escolha de um bom *backbone* é de fundamental importância para a extração de características relevantes a serem utilizadas pelos métodos. Neste trabalho, escolhemos a Efficient-Net-B7 (EffNet-B7), especialmente devido às suas características que combinam eficiência computacional, escalabilidade, generalização, capacidade de extração de características e facilidade de integração (Tan and Le, 2019).

Ao utilizar o EfficientNet-B7 como *backbone* em todos os métodos, garante-se uma base comum de extração de características. Isso é fundamental para uma comparação justa, permitindo que as diferenças observadas no desempenho sejam atribuídas principalmente ao método de segmentação, e não a variações no *backbone*. Além disso, o uso de um *backbone* de última geração eleva o padrão geral dos experimentos, destacando o potencial máximo de cada abordagem.

5.6 Treinamento dos modelos

O treinamento de todos os modelos criados foi feito com o otimizador AdamW, a taxa de aprendizagem foi de 1e-2, o $\beta 1$ de 0.9, o $\beta 2$ de 0.999, e o ε de 1e-7. Os modelos foram treinados durante 200 épocas com a função de perda focal binária. Para permitir que o modelo realize o treinamento com atualizações

controladas e mais estáveis, aplicamos uma estratégia de agendamento de taxa de aprendizagem de recozimento de cosseno (*cosine annealing schedule*), reduzindo a taxa de aprendizagem progressivamente desde sua configuração inicial até 1e-5. Durante as 200 épocas, o modelo que alcançou maior IoU (*Intersection Over Union*) de validação durante estas épocas foi salvo em disco para subsequente avaliação. Para o aumento de dados, para cada par de imagem e máscara do lote, realizamos a virada da imagem nos dois sentidos (horizontal e vertical) com uma probabilidade de 0.3, e translação com modificação da escala com probabilidade de 0.5.

Para a aquisição e o treinamento dos modelos utilizamos a biblioteca Segmentation Models PyTorch[5]. Para realizar as transformações nós utilizamos o pacote python Albumentations[6].

5.6.1 Função de perda focal binária

As funções de perda são peças fundamentais para o treinamento de modelos de AP, estas funções são projetadas para medir a dissimilaridade entre a distribuição de probabilidade prevista por um modelo e os verdadeiros rótulos binários de um conjunto de dados.

A função de perda focal tem mostrado resultados promissores no domínio da segmentação semântica de florestas Xia et al. (2021). Foi especificamente concebida para lidar com o desequilíbrio das classes, prestando mais atenção aos exemplos difíceis Lin et al. (2018).

Ao definir a probabilidade de predizer a classe verdadeira, p_t, da seguinte forma:

$$p_t = \begin{cases} p, & \text{se } y = 1 \\ 1-p, & \text{de outra maneira} \end{cases} \tag{5.1}$$

Em que $y \in \{\pm 1\}$ especifica a classe de *ground-truth* e $p \in [0,1]$ é a probabilidade do modelo estimar a classe com $y = 1$.

A perda focal pode ser definida desta maneira:

$$FL(p_t) = -\alpha_t (1-p_t)^{\gamma} log(p_t) \tag{5.2}$$

Em que $FL(p_t)$ é o loss focal, $\alpha \in [0,1]$ é utilizado para lidar com classes imbalanceadas e $\gamma \in [0,5]$ é um parâmetro de foco ajustável que permite controlar o foco nos exemplos difíceis. Neste trabalho, definimos os valores padrão recomendados $\alpha = 0.25$ e $\gamma = 2$.

[5]Um biblioteca Python com redes neurais para segmentação de imagens baseada em PyTorch, disponível em: https://github.com/qubvel-org/segmentation_models.pytorch

[6]https://albumentations.ai

		Condição predita	
		Positivo (P)	Negativo (P)
Condição atual	Positivo (P)	Verdadeiro Positivo (VP)	Falso Negativo (FN)
	Negativo (N)	Falso Positivo (FP)	Verdadeiro Negativo (VN)

Tabela 5.2: Modelo de matriz de confusão

5.7 Avaliação

5.7.1 Matriz de confusão

A matriz de confusão permite entender visualmente as confusões ou erros do modelo, trata-se de uma tabela específica de contingência de duas dimensões (atual e predito), e conjuntos idênticos de "classes"em ambas as dimensões (cada combinação de dimensão e classe é uma variável na tabela de contingência). O modelo para qualquer matriz de confusão binária, isto é, quando há apenas duas classes de interesse, utiliza os quatro tipos de resultados discutidos acima (verdadeiros positivos (VP), falsos negativos (FN), falsos positivos (FP) e verdadeiros negativos (VN)) juntamente com as classificações positivas e negativas. A tabela 5.2 ilustra uma matriz de confusão para um problema binário, com os quatro tipos de resultados sendo formulados.

5.7.2 Métricas de performance

Para avaliar os modelos, nós derivamos algumas métricas da matriz de confusão, a Precisão, o Recall e o F1 Score podem ser definidos pelas Eqs. 5.3 à 5.5.

$$Precision = \frac{VP}{VP+FP} \times 100 \tag{5.3}$$

$$Recall = \frac{VP}{VP+FN} \times 100 \tag{5.4}$$

$$F1-Score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{5.5}$$

Como pode ser visto, a precisão de uma classe é a razão entre o número de predições positivas corretas e o número total de predições positivas, e o recall é definido como a razão entre o número de predições positivas corretas e o número total de segmentos que realmente pertencem à classe positiva. Já o F1-Score é a média harmônica entre precisão e recall.

Já a interseção sobre união (IoU, do inglês *Intersection over Union*) é uma métrica comumente utilizada para a detecção/segmentação de objetos, e mede a sobreposição entre uma predição e a verdade de solo (*ground truth)* ela é

definida pela Equação 5.6.

$$IoU = \frac{\text{Intersecção}}{\text{União}} = \frac{|\text{Predição} \cap \text{Ground Truth}|}{|\text{Predição} \cup \text{Ground Truth}|}. \tag{5.6}$$

Como pode ser observado, ela é a razão da área de intersecção, que é a área comum entre a máscara predita e a área de *ground truth*, pela área de união, que é a soma das áreas da máscara predita e da máscara real, descontando a intersecção.

Adicionalmente, para a segmentação semântica, a IoU pode ser entendida pela equação 5.7:

$$IoU = \frac{TP}{TP + FP + FN}. \tag{5.7}$$

5.8 Avaliação de bandas - algoritmo de oclusão de sensitividade

Algoritmo 1: Análise de Sensibilidade de Oclusão para Importância do Canal

Input: Modelo *model*, tensor de entrada *input_tensor* de tamanho $(tamanho\_do\_lote, canais, altura, largura)$, máscara alvo *target*, função de loss *criterion*, dimensão do canal *channel_dim* (padrão $= 1$)
Output: Lista de canais com pontuações de importância
Definir modelo para modo de avaliação;
Desabilitar cálculos de gradiente;
Calcular saída original: *original_output* = *model*(*input_tensor*);
Calcular perda original: *original_loss* = *criterion*(*original_output*, *target*);
Inicializar uma lista vazia *channel_scores* para armazenar pontuações de importância para cada canal;
for *cada canal c em input_tensor ao longo da dimensão channel_dim* **do**
    Clone o tensor de entrada como *occluded_input* $\leftarrow$ *input_tensor*;
    Zere (oclua) o canal atual *c* em *occluded_input*;
    Calcule a saída ocluída: *occluded_output* = *model*(*occluded_input*);
    Calcule a perda ocluída: *occluded_loss* = *criterion*(*occluded_output*, *target*);
    Calcule a pontuação de importância para o canal *c*: *importance_score* = *occluded_loss* $-$ *original_loss*;
    Acrescente (*c*, *importance_score*) a *channel_scores*;
return *channel_scores*;

O método de oclusão de sensitividade é um método de interpretação de modelos que foi inicialmente introduzido por Zeiler and Fergus (2014) para verificar se o modelo estava realmente identificando a localização do objeto na imagem ou apenas usando o contexto circundante ao mascarar sistematica-

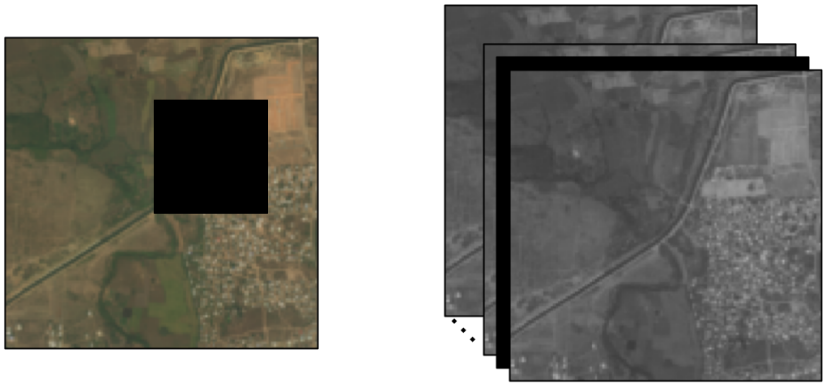

Figura 5.3: Exemplo demonstrando (a) a oclusão de uma parte de uma imagem e (b) a oclusão de uma banda da imagem.

mente partes da imagem de entrada para observar como essa oclusão afeta a previsão da rede. Embora seu trabalho tenha se concentrado principalmente em regiões espaciais, e não em canais ou bandas individuais, ele lançou as bases para adaptações posteriores do método para analisar canais e importância de recursos em vários domínios, incluindo análise de imagens multiespectrais e hiperespectrais.

Neste trabalho, nós adaptamos este método para avaliar a contribuição individual das bandas para a segmentação de eucalipto. Como pode ser visto na comparação da Figura 5.3, nós adaptamos o método ao ocluir sistematicamente as bandas de uma imagem, ao invés de partes da imagem. Ao ocluir sistematicamente (mascarar ou zerar) cada canal de entrada e observar o impacto no desempenho do modelo. Os canais que causam maior degradação no desempenho quando ocluídos são considerados os mais importantes. O algoritmo aplicado está descrito no pseudocódigo 1.

CAPÍTULO

6

Resultados e discussão

6.1 Comparação com métodos populares

Nesta seção, os resultados da comparação do método proposto com os métodos populares de segmentação semântica são apresentados.

6.1.1 Resultados quantitativos

A Tabela 6.1 apresenta uma comparação com outros métodos de segmentação semântica estado-da-arte na qualidade de segmentação medida pela Intersecção sobre União (IoU) para a classe de eucalipto, no conjunto de validação (cidade de Selvíria - MS), no conjunto de teste TL (cidade de Três Lagoas - MS) e no conjunto de teste TM (cidade de Três Marias - MG).

Model	Intersection Over Union (%)			
	Validation (Se.)	Test (TL)	Test (TM)	Mean
FPN	82.31	84.87	72.11	79.76
PSPNet	83.03	86.71	80.92	83.55
DeepLabv3+	84.03	86.62	82.29	84.31
MANet	83.69	**88.55**	81.89	84.71
UPerNet	83.28	87.01	81.23	83.84
SegFormer	83.25	86.46	80.93	83.54
Ours	**84.40**	88.42	**84.16**	**85.66**

Tabela 6.1: Comparação com outros métodos de segmentação semântica estado-da-arte na qualidade de segmentação.
Se. = Selvíria; TL = Três Lagoas; TM = Três Marias.

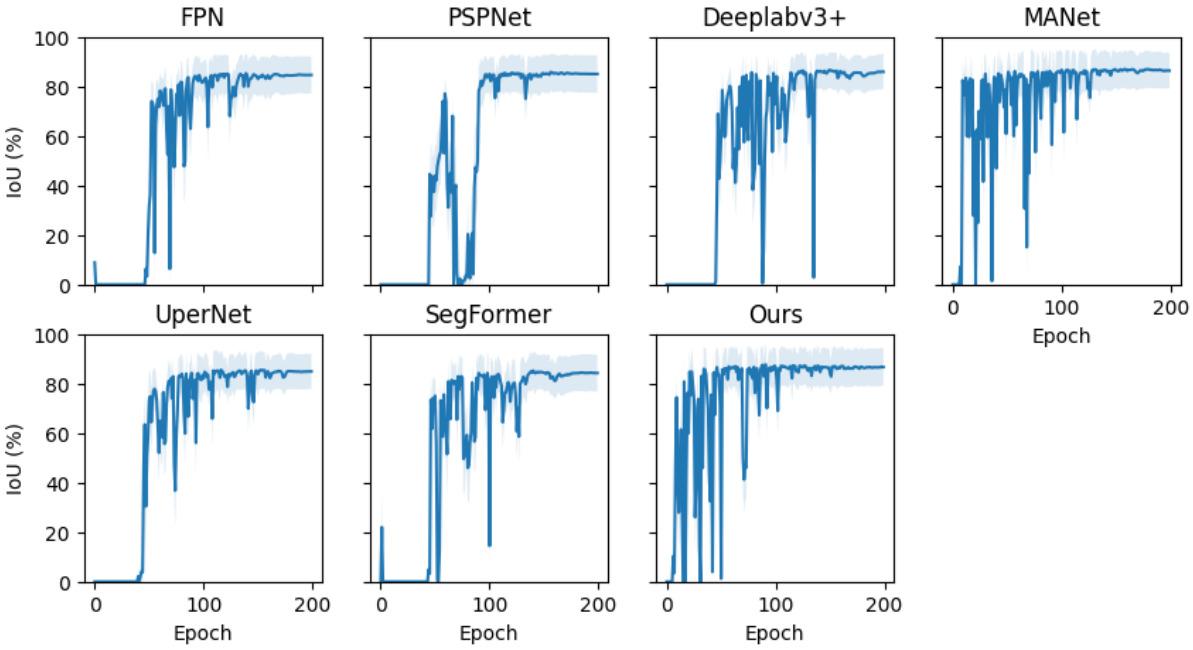

Figura 6.1: IoU de validação por época para cada modelo.

Nota-se que o método proposto consegue melhor desempenho entre os outros métodos, com maior média (85.66) de IoU das três cidades do experimento, com exceção do teste na cidade de Três Lagoas - MS, em que perde por apenas 0.13% para o método MANet, no entanto, o MANet por sua vez, alcança resultado significativamente inferior na cidade de Três Marias - MG, apresentando dessa maneira, performance reduzida de generalização. Através desse experimento, observamos que a ResGhostU-Net alcança resultados consistentes e competitivos em relação aos outros métodos populares tanto na validação como na generalização para outras cidades distantes tanto espacialmente como temporalmente.

6.1.2 Curvaturas de treinamento

A Figura 6.1 apresenta a curvatura de treinamento de IoU para cada modelo treinado durante 200 épocas na cidade de Água Clara, e validado na cidade de Selvíria. Como pode ser visto, o método proposto é capaz de convergir mais rapidamente que os outros métodos comparados, permanecendo relativamente estável já a partir da época 100, demonstrando sua capacidade em termos de aceleração e velocidade de convergência.

6.1.3 Comparação visual de exemplos

Para entender melhor os resultados de avaliação, nesta seção apresentamos uma comparação visual de 4 exemplos com suas respectivas imagens, máscaras e predições para todos os métodos, as imagens estão ilustradas na

Model	Mean IoU (%) ↑	Params (M) ↓	Flops (G) ↓	Time (ms) ↓
FPN	79.76	65.666	8.822	50.073
PSPNet	83.55	63.887	1.923	15.868
DeepLabv3+	84.31	65.106	15.056	75.329
MANet	84.71	78.004	10.434	52.034
UPerNet	83.84	65.382	10.318	44.063
SegFormer	83.25	64.393	8.593	42.492
Ours	**85.66**	**1.406**	**1.853**	**6.189**

Tabela 6.2: Comparação de cada modelo em relação a eIoU (%), Parâmetros (M), Flops (G) e Tempo (ms).

Figura **??**.

A primeira imagem é um exemplo que destaca a capacidade da rede de distinguir com precisão florestas naturais e de eucalipto. O desafio desta imagem é a área florestal que fica muito próxima das plantações de eucalipto, e o modelo consegue segmentar esta imagem de forma satisfatória.

A segunda imagem é desafiadora no sentido de que há bordas confusas de eucaliptos e outros tipos de vegetação. Podemos perceber a capacidade do método de lidar com essas situações desafiadoras ao delinear com precisão as bordas das plantações de eucalipto.

A terceira imagem também testa a capacidade da rede de prever detalhes nas bordas, demonstrando a atenção do método até mesmo aos pequenos detalhes. Observe que além do método proposto, apenas os métodos UperNet e MANet conseguiram prever a pequena área da borda marcada com a seta.

A quarta imagem também testa essa capacidade de prever bordas. Como pode ser observado neste caso, os métodos comparados não conseguem prever com precisão a área indicada e, quando o fazem, predizem erroneamente uma área de eucalipto no canto direito. Porém, observe que o método proposto possui um artefato de erro, uma vez que as arestas não foram delimitadas com precisão.

Apesar dos resultados satisfatórios, ainda podem ser identificados casos em que o modelo falha a generalização da predição, como na última imagem, gerando bordas confusas.

6.1.4 Performance geral

A Tabela 6.2 apresenta uma comparação dos modelos em termos de IoU médio, número de parâmetros, FLOPS (*Floating-point Operations per Second*) e tempo de predição[1].

[1]Para o cálculo das métricas de parâmetros e FLOPs utilizamos entrada de dimensões 1 x 13 x 256 x 256, e para calcular o tempo de predição consideramos o tempo médio de 1000 predições, os cálculos de parâmetros e flops foram realizados utilizando a biblioteca fvcore,

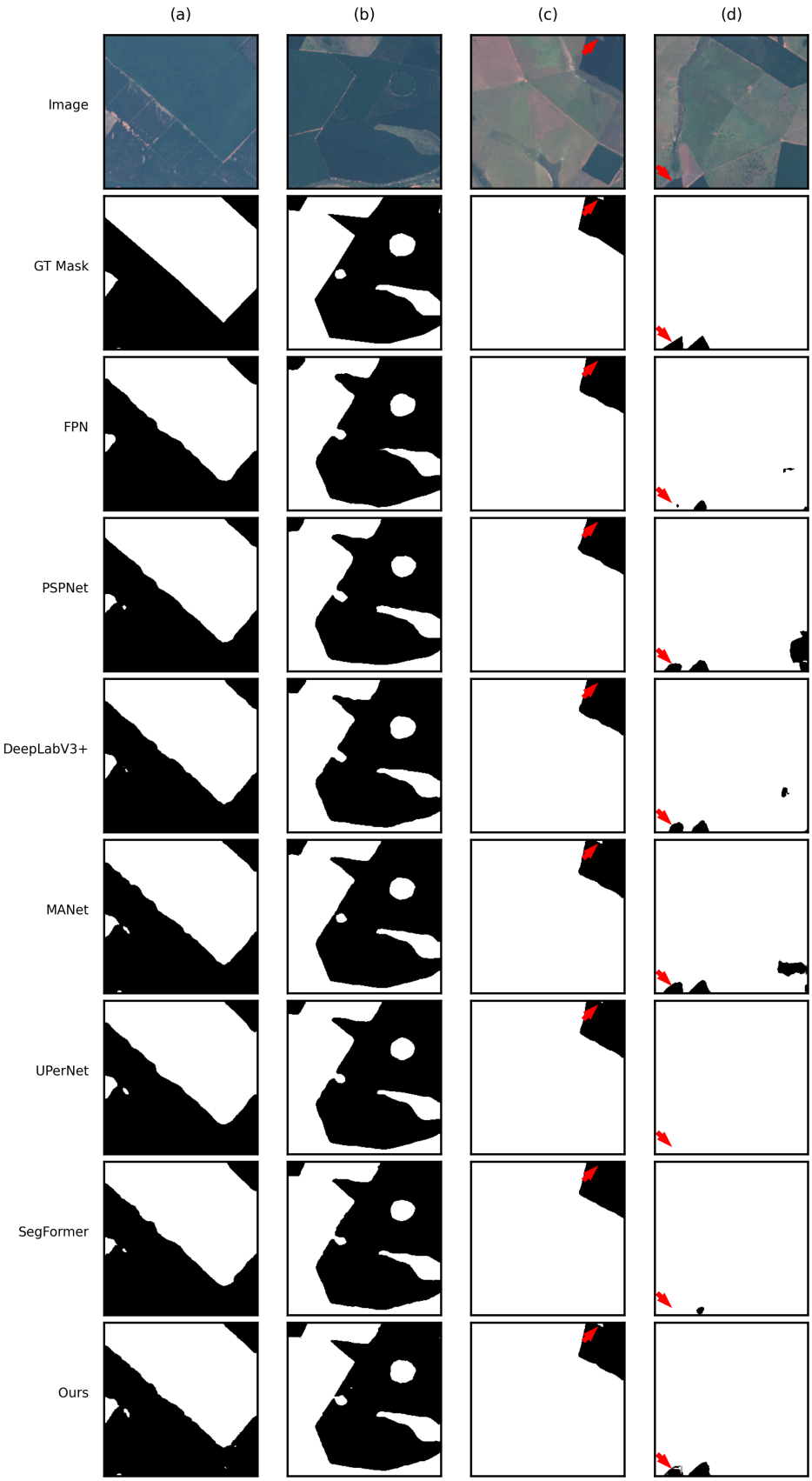

Figura 6.2: Exemplos mostrando suas respectivas imagens, máscaras de verdade fundamental e máscaras de predição. Pontos de referência estão indicados com uma seta vermelha.

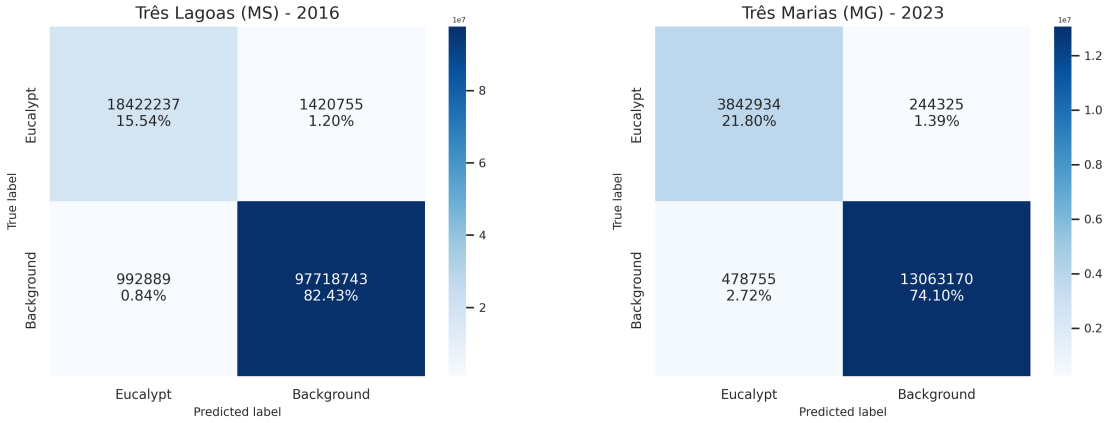

Figura 6.3: Matriz de confusão mostrando os resultados para cada cidade de teste.

Cidade de teste	Metricas (%)			
	Precisão	Recall	F1-Score	IoU
TL	94.89	92.84	93.85	88.42
TM	88.92	94.02	91.40	84.16

Tabela 6.3: Métricas extraídas da matriz de confusão.
TL = Três Lagoas; TM = Três Marias.

Como pode ser visto, a ResGhostU-Net ultrapassar os métodos SOTA tradicionais em termos performance de segmentação, sendo muito mais compacta em termos de parâmetros (1.406 M), mais eficiente em termos de FLOPS (1.853 G), e mais rápida em termos de predição (6.189 ms) do que os outros métodos comparados.

6.2 Matriz de confusão

A Figura 6.3 apresenta as matrizes de confusão gerada pela comparação dos pixels preditos e de verdade fundamental. Nesta matriz, as linhas representam os rótulos verdadeiros e as colunas os rótulos preditos para cada pixel. E a Tabela 6.3 apresenta algumas métricas que foram calculadas a partir da matriz para a classe de eucalipto (precisão, recall, F1-score e IoU).

No primeiro caso de teste, isto é, em Três Lagoas (MS) no ano de 2016, nota-se que apenas 0.84% dos pixels verdadeiramente de background foram preditos como eucalipto (falsos positivos) e 1.20% dos pixels verdadeiramente de eucalipto foram preditos como background, as métricas de precisão e recall foram acima de 90%, resultando em um F1-Score de 93.85, com um IoU de 88.42. Estas métricas são consideradas boa, pois demonstram que o método consegue prever bem uma cidade que está mais próxima das cidades que

disponível no seguinte link: https://github.com/facebookresearch/fvcore.

Bloco Residual Fantasma	Cidade	Metricas (%)			
		Precisão	Recall	F1-Score (TM)	IoU
✗	TL	97.86	79.13	87.51	77.79
✗	TM	96.14	70.41	81.29	68.48
✓	TL	94.89	92.84	93.85	88.42
✓	TM	88.92	94.02	91.40	84.16

Tabela 6.4: Métricas demonstrando a melhora de performance para a classe de eucalipto com e sem a adição dos blocos residuais fantasma.
TL = Três Lagoas; TM = Três Marias.

foram utilizadas para treinamento e no mesmo ano.

No segundo caso de teste, isto é, em Três Marias (MG) para o ano de 2023, nota-se que 1.39% de falsos positivos e 2.72% de falsos negativos, isso resultou em um bom recall mas uma precisão inferior a 90%, resultando em um F1-Score de 91.40 e um IoU de 84.16%. Para este caso de teste nós obtemos métricas boas, mas observamos uma ligeira queda na performance, o que pode ser explicado pela distância espacial e temporal da cidade em relação ao treinamento.

6.3 Estudo de ablação

6.3.1 Bloco residual fantasma

A Tabela 6.4 demonstra a melhora da performance de segmentação ao introduzir os blocos residuais fantasmas a U-Net, utilizando o mesmo número de filtros, sem os blocos o modelo obtém métricas de F1-Score abaixo de 90% e IoU abaixo de 80%, enquanto que com a adição dos blocos o F1-Score fica acima de 90% e o IoU fica acima de 84%, demonstrando que houve melhora para as duas cidades de teste, note também que a alta precisão do modelo tradicional custou muitos falsos negativos, diminuindo consideravelmente o recall, enquanto que com a adição dos blocos alcança-se um bom *trade-off* entre precisão e recall.

6.3.2 Bandas de entrada

A Figura 6.4 apresenta um gráfico que representa a importância de cada canal para a segmentação, essas pontuações são resultantes da aplicação do algoritmo de sensitividade de oclusão em 64 imagens retiradas do dataset de teste. Neste gráfico, quando mais importante é o canal para a segmentação, maior é sua pontuação. Pode-se observar que os canais B7 (Red Edge 3), B6 (Red Edge 2) e B4 (Red) são significativamente mais importantes de acordo

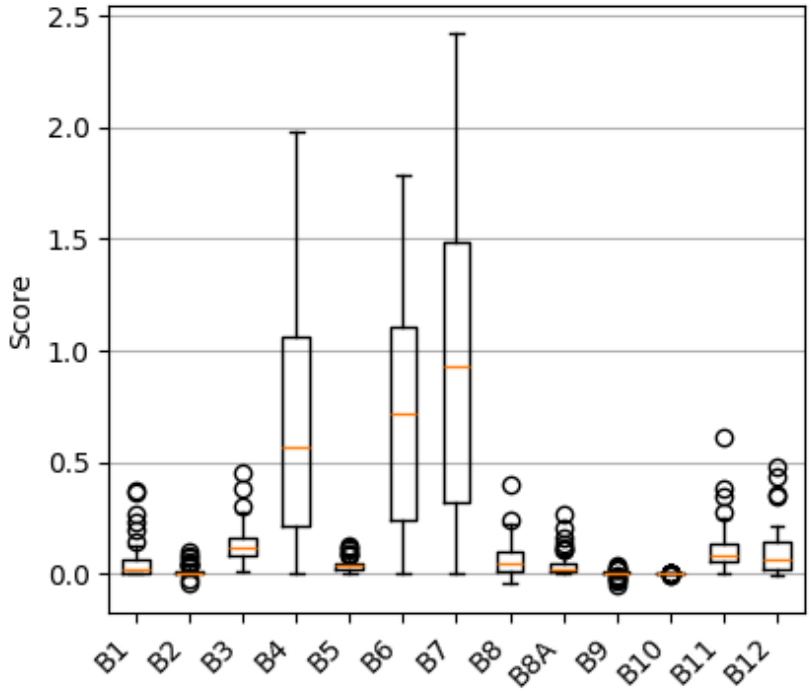

Figura 6.4: Importância de cada canal para a segmentação.

com a média, no entanto, vale a pena notar que estes possuem alta dispersão em torno da média.

A Figura 6.5 apresenta um gráfico que demonstra o efeito da adição cumulativa dos canais com maiores pontuação na qualidade de segmentação. No eixo das abscissas destaca-se em negrito a IoU para cada cidade, utilizando como entrada todas as bandas (**All**) e as bandas vermelho, azul e verde (**RGB**), também encontra-se os resultados com a utilização de bandas selecionadas, por exemplo B(7, 6) significa que a entrada é composta pelas bandas B7 e B6.

Como pode ser visto, o gráfico revela uma relação linear entre o número de bandas adicionadas e o desempenho do método no dataset. Observa-se também que a adição de 3 bandas selecionadas pelo algoritmo (B7 - Borda Vermelha 3, B6 - Borda Vermelha 2 e B4 - Vermelho) alcança um desempenho significativamente melhor do que o uso dos canais RGB. Já com a adição da banda verde (B3), o método alcança um desempenho satisfatório equiparando-se ao desempenho das 13 bandas. Note que a partir da adição destas 4 bandas, o método alcança um platô, com o desempenho variando ligeiramente em torno da média, devido a esse fato, decidimos parar o experimento com 8 bandas.

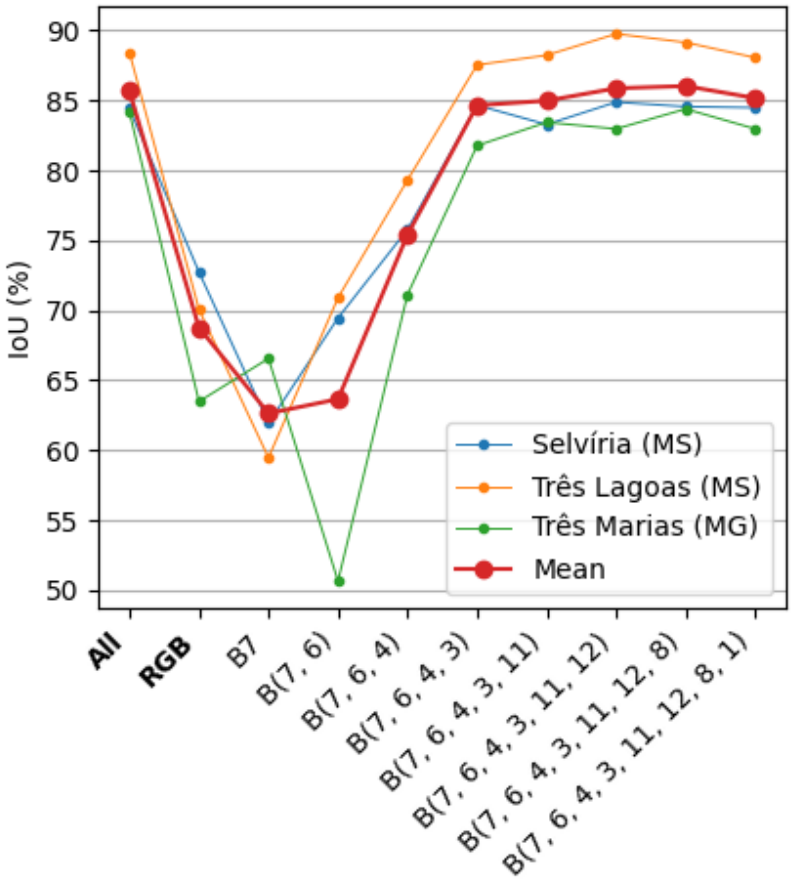

Figura 6.5: Efeito da adição cumulativa de bandas selecionadas na performance do modelo.

6.4 Discussão

Os resultados dos experimentos descritas comprovam a eficácia do método proposto para a segmentação semântica do eucalipto.

Em primeiro lugar, demonstramos que o método proposto é altamente competitivo com outros métodos de segmentação semântica populares em termos de métricas quantitativas. Em seguida, analisando visualmente os exemplos, identificámos casos de sucesso, mesmo em situações em que as classes são muito semelhantes entre si (como florestas naturais e de eucalipto). No entanto, identificamos alguns exemplos em que o modelo ainda não produz resultados ideais, podendo ser necessário um maior ajuste dos parâmetros para atingir tais resultados. Essas constatações deixam espaço para mais estudos futuros. Em seguida, comparando o nosso método em termos de desempenho de segmentação e custo computacional com métodos populares, demonstrámos que o método é capaz de manter um excelente compromisso entre a qualidade da segmentação, o peso, o número de operações e a velocidade de inferência.

Em seguida, estudamos o desempenho do método em diferentes cidades, destacando a capacidade de generalização do método quando lida com a cidade de Três Lagoas - MS no mesmo ano, e até quando lida com imagens de 2023 na cidade de Três Marias - MG, que é uma cidade espacialmente distante, com caraterísticas diferentes daquelas observadas pelo método no treinamento. Além disso, a matriz de confusão mostra uma boa distribuição de falsos negativos e positivos, enquanto as curvas de treinamento demonstram a boa convergência do método em relação aos métodos comparados.

Por fim, o estudo de ablação realizado demonstra a eficácia da adição do bloco residual fantasma em relação aos parâmetros estudados. Além disso, descobrimos que a adição de pelo menos 4 bandas selecionadas pelo algoritmo de oclusão de sensitividade pode alcançar resultados ligeiramente melhores do que a utilização de todas as 13 bandas. Estes resultados corroboram com outros estudos da literatura, que demonstram a importância da utilização de imagens multiespectrais segmentação de eucalipto Forstmaier et al. (2020); Illarionova et al. (2021); Firigato et al. (2021); Dallaqua et al. (2022); da Costa et al. (2021); Boston et al. (2024), salientando a importância da combinação de bandas de infravermelho.

CAPÍTULO

7

Conclusões

Neste trabalho nós propomos um novo método compacto baseado na arquitetura U-Net (ResGhostU-Net) para a segmentação semântica de eucalipto através de imagens de satélite Sentinel-2 em cidades brasileiras. Resultados quantitativos demonstram que o método proposto é altamente competitivo com relação a populares métodos de aprendizagem profunda de segmentação semântica. Enquanto que uma análise visual de exemplos difíceis revela que o modelo produz resultados satisfatórios, conseguindo diferenciar plantações de florestas naturais, lidar com plantações próximas entre si, e generalizar quando há ambiguidade na imagem, no entanto, também identificamos casos onde o modelo falha na generalização, criando bordas confusas, esta limitação deve ser levada em conta em estudos futuros. Além disso, o estudo de ablação demonstra a eficácia do componente proposto, e que o uso de ao menos 4 bandas selecionadas pode alcançar resultados superiores à utilização dos tradicionais canais RGB e ligeriamente superiores ao uso de 13 bandas, salientando a importância da utilização de imagens multiespectrais para a segmentação de eucalipto.

O método proposto, que conta com uma simples e eficiente arquitetura, é capaz de alcançar performance estado da arte em relação a qualidade de segmentação, compactabilidade e velocidade de inferência. Este modelo aprendizagem profunda pode ser futuramente levado a produção e melhor otimizado para o mapeamento, possuindo um vasto campo de aplicações. Mais testes deverão ser realizados futuramente para viabilizar a aplicação do método proposto em larga-escala para a segmentação de eucalipto no Brasil.

Referências Bibliográficas

Ajibola, S. and Cabral, P. (2024). A systematic literature review and bibliometric analysis of semantic segmentation models in land cover mapping. *Remote Sensing*, 16(12). Citado na página 2.

Bacani, V. M., Machado da Silva, B. H., Ayumi de Souza Amede Sato, A., Souza Sampaio, B. D., Rodrigues da Cunha, E., Pereira Vick, E., Ribeiro de Oliveira, V. F., and Decco, H. F. (2024). Carbon storage and sequestration in a eucalyptus productive zone in the brazilian cerrado, using the ca-markov/random forest and invest models. *Journal of Cleaner Production*, 444:141291. Citado na página 1.

Boston, T., Van Dijk, A., and Thackway, R. (2024). U-net convolutional neural network for mapping natural vegetation and forest types from landsat imagery in southeastern australia. *Journal of Imaging*, 10(6). Citado nas páginas 15, 16, 17, e 41.

Chen, L.-C., Papandreou, G., Schroff, F., and Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. Citado nas páginas 12 e 27.

Chollet, F. (2021). *Deep learning with python*. Manning Publications Co. LLC. Citado na página 9.

da Costa, L. B., de Carvalho, O. L., de Albuquerque, A. O., Gomes, R. A., Guimarães, R. F., and de Carvalho Júnior, O. A. (2021). Deep semantic segmentation for detecting eucalyptus planted forests in the brazilian territory using sentinel-2 imagery. *Geocarto International*, 37(22):6538?6550. Citado nas páginas 15, 16, e 41.

Dallaqua, F. B. J. R., Rosa, R. A. S., Schultz, B., Faria, L. R., Rodrigues, T. G., Oliveira, C. G., Kieser, M. E. J., Malhotra, V., Dwyer, T., and Wolfe,

D. S. (2022). Forest plantation detection through deep semantic segmentation. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B3-2022:77–84. Citado na página 41.

Dias, D., Dias, U., Menini, N., Lamparelli, R., Le Maire, G., and Torres, R. d. (2020). Image-based time series representations for pixelwise eucalyptus region classification: A comparative study. *IEEE Geoscience and Remote Sensing Letters*, 17(8):1450?1454. Citado nas páginas 14, 16, e 17.

Ferreira, M. P., La Rosa, L. E. C., Happ, P. N., Theobald, R. B., and Feitosa, R. Q. (2019). Mapping eucalyptus plantations and natural forest areas in landsat-tm images using deep learning. In *Anais do XIX Simpósio Brasileiro de Sensoriamento Remoto. São José dos Campos : INPE. 2019*, volume 19, São José dos Campos, SP, Brasil. INPE. Citado nas páginas 14 e 16.

Firigato, J. O. N., Junior, J. M., Gonçalves, W. N., and Matheus Bacani, V. (2021). Deep learning and google earth engine applied to mapping eucalyptus. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pages 4696–4699. Citado nas páginas 15, 16, 17, e 41.

Forstmaier, A., Shekhar, A., and Chen, J. (2020). Mapping of eucalyptus in natura 2000 areas using sentinel 2 imagery and artificial neural networks. *Remote Sensing*, 12(14). Citado nas páginas 14, 16, e 41.

Gazzea, M., Solheim, A., and Arghandeh, R. (2023). High-resolution mapping of forest structure from integrated sar and optical images using an enhanced u-net method. *Science of Remote Sensing*, 8:100093. Citado na página 2.

Gomes, J. V. P. and Cubas, M. G. (2021). *Fundamentos do Sensoriamento Remoto*. Inter Saberes. Citado na página 6.

Grigorescu, S., Trasnea, B., Cocias, T., and Macesanu, G. (2019). A survey of deep learning techniques for autonomous driving. *Journal of Field Robotics*, 37. Citado na página 10.

Han, K., Wang, Y., Tian, Q., Guo, J., Xu, C., and Xu, C. (2020). Ghostnet: More features from cheap operations. Citado na página 21.

He, K., Zhang, X., Ren, S., and Sun, J. (2015). Deep residual learning for image recognition. Citado nas páginas 10 e 21.

Illarionova, S., Trekin, A., Ignatiev, V., and Oseledets, I. (2021). Tree species mapping on sentinel-2 satellite imagery with weakly supervised classification and object-wise sampling. *Forests*, 12(10). Citado na página 41.

Ioffe, S. and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. Citado na página 21.

Jia, H., Zhang, J., Ma, K., Qiao, X., Ren, L., and Shi, X. (2024). Application of convolutional neural networks in medical images: A bibliometric analysis. *Quantitative Imaging in Medicine and Surgery*, 14(5):3501?3518. Citado na página 10.

Kattenborn, T., Leitloff, J., Schiefer, F., and Hinz, S. (2021). Review on convolutional neural networks (cnn) in vegetation remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 173:24–49. Citado na página 10.

Klein, H. S. and Luna, F. V. (2022). The development of a modern cellulose industry in south america. *Latin American Research Review*, 57(4):753?774. Citado na página 1.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Pereira, F., Burges, C., Bottou, L., and Weinberger, K., editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc. Citado na página 10.

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017). Feature pyramid networks for object detection. Citado na página 27.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2018). Focal loss for dense object detection. Citado na página 29.

Long, J., Shelhamer, E., and Darrell, T. (2015). Fully convolutional networks for semantic segmentation. Citado na página 11.

Lu, W., Zhang, Z., and Nguyen, M. (2024). A lightweight cnn?transformer network with laplacian loss for low-altitude uav imagery semantic segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–20. Citado na página 2.

Luo, Z., Yang, W., Yuan, Y., Gou, R., and Li, X. (2024). Semantic segmentation of agricultural images: A survey. *Information Processing in Agriculture*, 11(2):172–186. Citado nas páginas 1 e 2.

Meneses, P. and Almeida, T. (2012). *Introdução ao Processamento de Imagens de Sensoriamento Remoto*. Citado nas páginas 5 e 6.

Osco, L. P., Marcato Junior, J., Marques Ramos, A. P., de Castro Jorge, L. A., Fatholahi, S. N., de Andrade Silva, J., Matsubara, E. T., Pistori, H., Gonçalves, W. N., and Li, J. (2021a). A review on deep learning in uav remote

sensing. *International Journal of Applied Earth Observation and Geoinformation*, 102:102456. Citado na página 1.

Osco, L. P., Marcato Junior, J., Marques Ramos, A. P., de Castro Jorge, L. A., Fatholahi, S. N., de Andrade Silva, J., Matsubara, E. T., Pistori, H., Gonçalves, W. N., and Li, J. (2021b). A review on deep learning in uav remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 102:102456. Citado na página 10.

Pierson, H. and Gashler, M. (2017). Deep learning in robotics: A review of recent research. *Advanced Robotics*, 31. Citado na página 10.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. Citado nas páginas 11 e 19.

Sarker, I. H. (2021). Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions. *SN Computer Science*, 2. Citado na página 9.

Simonyan, K. and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. Citado na página 10.

Tan, M. and Le, Q. V. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. *CoRR*, abs/1905.11946. Citado nas páginas 10 e 28.

Teodoro, P. E., Rossi, F. S., Teodoro, L. P. R., Santana, D. C., Ratke, R. F., de Oliveira, I. C., Silva, J. L. D., de Oliveira, J. L. G., da Silva, N. P., Baio, F. H. R., Torres, F. E., and da Silva Junior, C. A. (2024). Soil co2 emissions under different land-use managements in mato grosso do sul, brazil. *Journal of Cleaner Production*, 434:139983. Citado na página 1.

Wagner, F. H., Sanchez, A., Tarabalka, Y., Lotte, R. G., Ferreira, M. P., Aidar, M. P., Gloor, E., Phillips, O. L., and Aragão, L. E. (2019). Using the u?net convolutional network to map forest types and disturbance in the atlantic rainforest with very high resolution images. *Remote Sensing in Ecology and Conservation*, 5(4):360?375. Citado nas páginas 13, 16, e 17.

Xia, L., Zhang, R., Chen, L., Li, L., Yi, T., Yao, W., Ding, C., and Xie, C. (2021). Evaluation of deep learning segmentation models for detection of pine wilt disease in unmanned aerial vehicle images. *Remote Sensing*, 13:3594. Citado na página 29.

Xiao, T., Liu, Y., Zhou, B., Jiang, Y., and Sun, J. (2018). Unified perceptual parsing for scene understanding. Citado na página 28.

Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., and Luo, P. (2021). Segformer: Simple and efficient design for semantic segmentation with transformers. Citado nas páginas 12 e 28.

Xu, G., Li, J., Gao, G., Lu, H., Yang, J., and Yue, D. (2023). Lightweight real-time semantic segmentation network with efficient transformer and cnn. *IEEE Transactions on Intelligent Transportation Systems*, 24(12):15897–15906. Citado na página 2.

Xu, Y., Lam, H.-K., and Jia, G. (2021). Manet: A two-stage deep learning method for classification of covid-19 from chest x-ray images. *Neurocomputing*, 443:96–105. Citado na página 28.

Zeiler, M. D. and Fergus, R. (2014). Visualizing and understanding convolutional networks. *Lecture Notes in Computer Science*, page 818?833. Citado nas páginas 17 e 31.

Zhang, A., Lipton, Z. C., Li, M., and Smola, A. J. (2023). *Dive into Deep Learning*. Cambridge University Press. https://D2L.ai. Citado na página 9.

Zhao, H., Chen, Z., Jiang, H., Jing, W., Sun, L., and Feng, M. (2019). Evaluation of three deep learning models for early crop classification using sentinel-1a imagery time series a case study in zhanjiang, china. *Remote Sensing*, 11(22). Citado nas páginas 14, 16, e 17.

Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J. (2017). Pyramid scene parsing network. Citado na página 27.