Identificação Automática de Áreas de Cana-de-Açúcar e Árvores Isoladas Utilizando Segmentação Semântica

Thiago Edgar Bauce Venancio

PROGRAMA DE PÓS-GRADUAÇÃO DE CIÊNCIA DA COMPUTAÇÃO DA FACOM-UFMS

Data de Depósito:

Assinatura:_____

Thiago Edgar Bauce Venancio

Orientador: *Prof. Dr. Wesley Nunes Gonçalves*

Dissertação apresentada à Faculdade de Computação Facom-UFMS como parte dos requisitos necessários à obtenção para do título de Mestre em Ciência de Computação.

UFMS - Campo Grande
Novembro/2025

Aos meus pais,
À minha irmã e sobrinhos,
À minha esposa e nossa família,
Aos meus filhos de 4 patas,
Aos meus amigos,
Ao meu orientador, amigos e colegas de laboratório.

Agradecimentos

A jornada para a conclusão desta dissertação foi um caminho de aprendizado e crescimento, e seria impossível percorrê-lo sem o apoio das pessoas. A cada uma delas, expresso minha mais profunda gratidão.

Aos meus pais, pelo amor incondicional, pelos valores que me ensinaram e por serem meu alicerce em todos os momentos. À minha irmã e sobrinhos, por serem fonte de alegria e inspiração. À minha esposa e nossa família, por toda a paciência, incentivo e compreensão, especialmente nos momentos de maior dedicação a este trabalho. Aos meus filhos de 4 patas, por me lembrarem da importância de uma pausa e de um carinho. Aos meus amigos, por estarem sempre presentes.

Agradeço à minha chefe, Liana, por toda a confiança e flexibilidade que me permitiram conciliar as responsabilidades profissionais com as acadêmicas. Aos meus colegas do suporte, pela compreensão e parceria, especialmente nos momentos em que precisei me ausentar para ir ao laboratório.

Um agradecimento especial ao meu orientador, Prof. Wesley, por sua orientação sem igual, por compartilhar seu vasto conhecimento e por me desafiar e incentivar constantemente a ir além. Sua paixão e comprometimento foram fundamentais para a qualidade deste trabalho.

Ao Prof. Marcato, minha gratidão por me incluir nos projetos e por acreditar em meu potencial, tornando esta dissertação uma realidade.

Por fim, agradeço ao Programa de Pós-Graduação em Ciências da Computação (PGCC) por me proporcionar esta oportunidade de crescimento acadêmico e profissional, e a todos os professores e colegas de laboratório, com quem compartilhei momentos de aprendizado e colaboração. A cada um de vocês, meu sincero muito obrigado.

Abstract

Manual orthophoto segmentation has limitations such as high costs, low scalability, and susceptibility to errors. Semantic segmentation is a crucial task for agricultural monitoring but faces challenges like class imbalance.The main objective of this work is to investigate semantic segmentation methodologies for the automatic identification of sugarcane areas and the mapping of isolated trees in orthophotos.One of the most significant contributions was the comparative evaluation of state-of-the-art models, such as SegFormer and InternImage, against more traditional and widely used convolutional architectures, like U-Net and DeepLabV3+. This analysis provided a comprehensive overview of the performance of different approaches in agricultural image segmentation.Furthermore, the study delved into the impact of class imbalance, a significant technical challenge in agricultural domains. To mitigate this issue, pre-processing strategies were proposed and investigated, including Crop-Focused sampling and Boundary-Zone sampling.The research also included an analysis of the effect of patch size on segmentation performance, addressing the trade-off between capturing global context and preserving local details. This demonstrated the viability of automating essential mappings, such as identifying planting areas and isolated trees, which shows direct gains in productive efficiency and a reduction in operational costs.These contributions, collectively, validate the capability of semantic segmentation solutions for automated monitoring, offering a path toward the implementation of more efficient and sustainable agricultural practices.

Keywords: Sugarcane; Isolated Trees; Class Imbalance; Precision Agriculture; SegFormer; InternImage; Semantic Segmentation.

Resumo

A segmentação manual de ortofotos tem limitações como alto custo, baixa escalabilidade e suscetibilidade a erros. A tarefa de segmentação semântica é fundamental para o monitoramento agrícola, mas enfrenta desafios como o desequilíbrio de classes. O objetivo central deste trabalho é investigar metodologias de segmentação semântica para a identificação automática de áreas de cana-de-açúcar e o mapeamento de árvores isoladas em ortofotos. Uma das contribuições mais importantes foi a avaliação comparativa de modelos do estado da arte, como o SegFormer e o InternImage, em relação a arquiteturas convolucionais mais tradicionais e amplamente utilizadas, como a U-Net e a DeepLabV3+. Essa análise forneceu um panorama abrangente do desempenho de diferentes abordagens na segmentação de imagens agrícolas. Além disso, o estudo aprofundou-se no impacto do desequilíbrio de classes, um desafio técnico significativo em domínios agrícolas. Para mitigar esse problema, foram propostas e investigadas estratégias de pré-processamento, como a amostragem focada em classes (Crop-Focused) e a amostragem por zona de borda (Boundary-Zone). A pesquisa também incluiu uma análise do efeito do tamanho dos patches na performance da segmentação, abordando a relação de compromisso entre a captura de contexto global e a preservação de detalhes locais. Com isso, foi possível demonstrar a viabilidade da automação de mapeamentos essenciais, como a identificação de áreas de plantio e árvores isoladas, o que evidencia ganhos diretos em eficiência produtiva e redução de custos operacionais. Essas contribuições, em conjunto, validam a capacidade de soluções de segmentação semântica para o monitoramento automatizado, oferecendo um caminho para a implementação de práticas agrícolas mais eficientes e sustentáveis.

Palavras-chave: Cana-de-açúcar; Árvores Isoladas; Desiquilíbrio de Classes; Agricultura de Precisão; SegFormer; InternImage; Segmentação Semântica.

Sumário

Lista de Figuras

Lista de Tabelas

Lista de Abreviaturas

AD Árvore de Decisão

AM Aprendizado de Máquina

AP Agricultura de Precisão

ASPP *Atrous Spatial Pyramid Pooling*

CNN *Convolutional Neural Network*

DCNv3 *Deformable Convolution v3*

FCN *Fully Convolutional Network*

FN Falso Negativo

FP Falso Positivo

IA Inteligência Artificial

IoU *Intersection over Union*

MiT *Mix Transformer*

MLP *Multi-Layer Perceptron*

NLP *Natural Language Processing (Processamento de Linguagem Natural)*

SAM *Segment Anything Model*

ViT *Vision Transformer*

VN Verdadeiro Negativo

VP Verdadeiro Positivo

VANT Veículo Aéreo Não Tripulado

Lista de Algoritmos

CAPÍTULO

1

Introdução

1.1 Contextualização e Motivação

A cana-de-açúcar ocupa posição central no agronegócio brasileiro, consolidando o país como maior produtor mundial. Na safra 2024/2025, o Brasil processou cerca de 680 milhões de toneladas, sendo que a região Centro-Sul respondeu por 91,5% desse total [11]. Esse desempenho é impulsionado pela crescente demanda por etanol, pela produção de açúcar e pela adoção de tecnologias avançadas de cultivo, resultando em um crescimento de 19,29% em relação à safra anterior [12]. Entre os derivados, o etanol se destaca como alternativa renovável aos combustíveis fósseis, contribuindo para a redução de emissões e para a sustentabilidade ambiental [13].

Nesse cenário, técnicas de monitoramento agrícola tornam-se essenciais para sustentar ganhos de eficiência e produtividade. O sensoriamento remoto, especialmente por meio de ortofotos obtidas com Veículos Aéreos Não Tripulados (VANTs), tem revolucionado essa prática [8]. Diferentemente de fotografias aéreas convencionais, as ortofotos passam por correções geométricas que garantem precisão cartográfica e permitem medições exatas da superfície.

Satélites fornecem informações relevantes para o monitoramento agrícola, porém enfrentam limitações como baixa resolução, cobertura por nuvens e menor frequência de aquisição [42]. VANTs surgem como solução eficaz, oferecendo resoluções centimétricas, flexibilidade operacional e custos reduzidos [26, 18]. Essa alta resolução possibilita identificar elementos críticos no campo — como árvores isoladas, erosões, estradas, rios ou manchas de plantas daninhas [16] — fundamentais para o planejamento da colheita e a

sustentabilidade da produção.

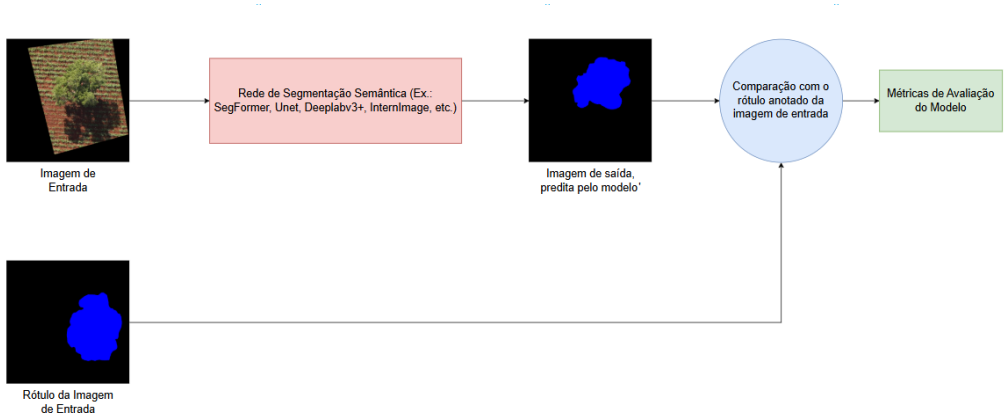
Entretanto, a análise manual de ortofotos é inviável quando se considera a escala operacional do setor sucroenergético: ela é custosa, lenta e sujeita a erros humanos [38]. Isso torna imprescindível o desenvolvimento de métodos de *segmentação semântica* automatizada, capazes de identificar elementos agrícolas com rapidez, precisão e escalabilidade.

1.2 Segmentação Semântica

A segmentação semântica consiste na atribuição de rótulos a cada pixel de uma imagem, separando regiões que correspondem a diferentes classes de interesse — como cana-de-açúcar, árvores isoladas e fundo. O pipeline, mostrado na Figura 1.1, típico dessa tarefa envolve:

1. Aquisição das imagens (ortofotos ou imagens de satélite);
2. Pré-processamento, incluindo normalização, ortorretificação e recorte em *patches*;
3. Extração de características realizada por CNNs ou Transformers;
4. Classificação por pixel e reconstrução do mapa segmentado;
5. Pós-processamento, como filtragens e ajustes morfológicos;
6. Avaliação por métricas como IoU, F1-Score e acurácia por classe.

Figura 1.1: Fluxo de processamento de imagens comum às tarefas de segmentação semântica.



Historicamente, redes neurais convolucionais (CNNs) dominaram essa área. A U-Net, amplamente utilizada em aplicações agrícolas [4, 49, 53, 40], e a DeepLabV3+, reconhecida por seu desempenho em ambientes complexos [2, 47, 55], tornaram-se arquiteturas de referência.

No entanto, CNNs possuem limitações na modelagem de dependências espaciais de longo alcance. Modelos mais recentes baseados em *Transformers*, como o SegFormer [51], e em convoluções deformáveis, como o InternImage [50], vêm superando essas restrições e avançando o estado da arte em segmentação de imagens.

1.3 Objetivos

Diante da relevância do monitoramento automatizado de áreas agrícolas, este trabalho tem como objetivo investigar modelos modernos de segmentação semântica aplicados ao mapeamento de cana-de-açúcar e árvores isoladas em ortofotos.

1.3.1 Objetivo Geral

Desenvolver e avaliar metodologias de segmentação semântica para a identificação automática de áreas de cana-de-açúcar e de árvores isoladas em ortofotos de alta resolução.

1.3.2 Objetivos Específicos

- Avaliar comparativamente arquiteturas clássicas e modernas de segmentação (U-Net, DeepLabV3+, SegFormer e InternImage);
- Analisar o impacto do desbalanceamento de classes e propor estratégias de pré-processamento para mitigá-lo;
- Investigar o efeito do tamanho dos *patches* na capacidade do modelo de capturar contexto global e detalhes locais;
- Demonstrar a viabilidade da automação do mapeamento agrícola, evidenciando benefícios produtivos e ambientais.

1.4 Contribuições

As principais contribuições deste trabalho incluem:

- Uma análise comparativa abrangente entre modelos clássicos e arquiteturas recentes baseadas em Transformers e convoluções deformáveis;
- Estratégias práticas de pré-processamento e amostragem para lidar com desbalanceamento e preservar detalhes estruturais em ortofotos;

- Estudo sistemático do efeito do tamanho de *patches* na qualidade da segmentação para diferentes modelos;
- Demonstração da aplicabilidade das soluções propostas ao monitoramento agrícola automatizado.

1.5 Estrutura da Dissertação

Esta dissertação está organizada da seguinte forma:

- **Capítulo 2 – Trabalhos Relacionados**: apresenta a evolução das técnicas de segmentação e suas aplicações na agricultura.
- **Capítulo 3 – Materiais e Métodos**: descreve o conjunto de dados, o pré-processamento, as arquiteturas avaliadas e o protocolo experimental.
- **Capítulo 4 – Resultados**: discute as métricas obtidas, as análises qualitativas e as comparações entre modelos.
- **Capítulo 5 – Conclusões**: resume os principais achados, limitações e perspectivas futuras.

CAPÍTULO

2

Trabalhos Relacionados

2.1 Introdução

A segmentação semântica de imagens consiste na tarefa de atribuir uma etiqueta semântica a cada pixel, facilitando a compreensão detalhada da estrutura visual presente em imagens digitais. Esta tarefa é crucial para diversas áreas, tais como análise médica, sensoriamento remoto e, particularmente, agricultura de precisão [10, 24].

Nas últimas duas décadas, avanços em aprendizado profundo transformaram este cenário. Redes Neurais Convolucionais (CNNs) possibilitaram a extração hierárquica de padrões locais e texturais, enquanto arquiteturas baseadas em Transformers expandiram a capacidade de capturar relações globais e contextuais [15, 30]. Mais recentemente, surgiram modelos fundacionais capazes de operar em regime *zero-shot* ou *few-shot*, oferecendo adaptabilidade inédita em ambientes com pouca anotação [3].

Este capítulo apresenta a evolução histórica e tecnológica dos modelos de segmentação, desde CNNs clássicas até arquiteturas híbridas e Transformers modernos, culminando nos modelos fundacionais. Paralelamente, destaca-se o impacto dessas tecnologias no domínio agrícola e ambiental.

2.2 Modelos Convencionais Baseados em CNNs

O sucesso das redes convolucionais deve-se à sua capacidade de extrair características estruturais e texturais por meio de filtros aplicados sequencialmente, possibilitando a construção de representações hierarquizadas [28, 46].

A Fully Convolutional Network (FCN) marcou um avanço ao realizar segmentação por meio de convoluções sem camadas densas, permitindo a geração de mapas de segmentação com tamanho adaptável [22]. Apesar disso, sofria com detalhes perdidos e limitações no contexto espacial.

Modelos encoder-decoder, como a U-Net, superaram parte dessas restrições por meio de skip connections, que integravam informações de diferentes níveis de abstração, melhorando a resolução espacial das predições [41, 45]. Extensões como U-Net++ [56] e U-Net3+ [21] ampliaram esse conceito ao incluir conexões mais densas e multiescala, favorecendo a segmentação de detalhes finos.

No escopo de segmentação urbana, natural e agrícola, modelos DeepLab empregaram convoluções dilatadas (*atrous convolutions*) e o módulo ASPP, enriquecendo a captura de contexto em diversas escalas sem perda de resolução [6, 5]. O DeepLabV3+ ampliou ainda mais essas capacidades com decoder refinado, tornando-se referência em benchmarks internacionais e aplicações reais.

2.3 Mecanismos de Atenção e Arquiteturas Híbridas

No intuito de superar limitação de campo receptivo e incorporar relacionamentos contextuais, mecanismos de atenção foram agregados tanto em blocos isolados quanto em arquiteturas híbridas CNN-transformer. Modelos Attention U-Net e variantes aplicaram atenção espacial e de canal para fomentar a discriminação de regiões relevantes [35, 44].

Híbridos como a TransUnet combinam múltiplas conexões densas entre encoder e decoder, além de aprimorados mecanismos de fusão de informações, proporcionando maior flexibilidade e precisão na segmentação de regiões complexas [56, 45]. Embora originalmente desenvolvida para imagens médicas de alta complexidade, a arquitetura tem demonstrado potencial significativo para aplicação em cenários agrícolas, contribuindo para o sensoriamento remoto. Na agricultura de precisão, modelos têm sido empregados para segmentação detalhada de culturas, diferenciação de plantas e remoção eficiente de ervas daninhas, permitindo monitoramento e manejo sustentáveis [36]. Tais abordagens contribuem para aumentar a produtividade e reduzir o uso indiscriminado de insumos, refletindo a versatilidade dessa arquitetura em domínios múltiplos.

2.4 Aplicações em Agricultura e Monitoramento Ambiental

A segmentação semântica possibilita avanços significativos no monitoramento agrícola, desde o mapeamento de talhões até a identificação precisa de árvores isoladas e diferenciação de espécies [31, 54]. A integração de imagens RGB, multiespectrais e dados LiDAR com algoritmos robustos tem melhorado a resolução e a confiabilidade das análises.

Modelos baseados em CNNs são especialmente eficazes no detalhamento de cultivos e detecção de ervas daninhas, sendo crucial para manejo sustentável e economia de insumos [17, 37]. A emergência de modelos híbridos e Transformers amplia as capacidades para grandes áreas e condições ambientais adversas [48].

Estudos recentes têm explorado o uso de aprendizado profundo em agricultura de precisão para tarefas como segmentação da maturidade de frutos, previsão de rendimento de culturas e análise de estresse hídrico [33, 23]. Por exemplo, modelos baseados em imagens multiespectrais adquiridas por drones têm sido aplicados para identificar o estágio de maturidade de tomates [33], enquanto abordagens integradas de sensoriamento remoto e aprendizado profundo têm sido utilizadas para prever o rendimento de arroz [23]. Além disso, muitos desses modelos também permitem avaliar condições ambientais relacionadas ao estresse hídrico [25], ampliando o escopo das análises agronômicas.

Uma das propostas relevantes nessa linha é o AgriSegNet, que combina UAVs com o modelo DeepLabV3+ e mecanismos de atenção para detectar anomalias como água parada e ervas daninhas [1]. Apesar dos avanços, o modelo apresentou limitações em cenários onde diferentes culturas compartilham características visuais semelhantes, evidenciando a necessidade de refinamento.

Outro avanço foi a integração de FCNs e U-Net com shapefiles, permitindo aprimorar a segmentação de talhões e exportar predições para plataformas de agricultura de precisão [43]. Para a classificação de culturas, abordagens baseadas em CNNs alcançaram resultados expressivos, com precisão de até 92.64% em tarefas que incluíram a cana-de-açúcar entre as classes analisadas [39].

No caso específico da cana-de-açúcar, métodos que combinam UAVs com arquiteturas como U-Net, LinkNet e PSPNet mostraram-se promissores na segmentação de linhas de plantio, com destaque para o desempenho da U-Net. Além disso, a utilização da Transformada de Radon como etapa de refinamento contribuiu para maior uniformidade na segmentação [40]. Abordagens posteriores avançaram ao detectar não apenas linhas retas, mas também cur-

vas e falhas no plantio, enfrentando desafios muitas vezes negligenciados em métodos convencionais [14].

A segmentação de árvores individuais em ambientes florestais complexos também tem recebido atenção. Técnicas baseadas em dados UAV-LiDAR, utilizando algoritmos adaptativos em forma de coroa para detecção de pontos-semente, demonstraram eficácia na redução de problemas de supersegmentação e subsegmentação, atingindo taxas de acerto de até 87.7% [54]. Esses resultados têm relevância direta para o mapeamento preciso de árvores isoladas em áreas agrícolas e florestais.

Por fim, redes baseadas em Transformers e em convoluções com múltiplas escalas têm se mostrado eficazes na detecção de ervas daninhas em cana-de-açúcar. A combinação de convoluções e mecanismos de atenção aumentou a eficiência computacional e a precisão da segmentação, com potencial aplicação em diferentes contextos agrícolas e florestais [48].

2.5 Considerações Finais

A Tabela 2.1 consolidou uma análise quantitativa e qualitativa entre diferentes modelos de segmentação, abrangendo arquiteturas clássicas baseadas em convoluções, modelos híbridos e os recentes paradigmas fundamentados em Transformers e modelos fundacionais. Esses modelos foram avaliados a partir de critérios como acurácia, eficiência computacional, escalabilidade, interpretabilidade e adaptabilidade a diferentes domínios. Tal síntese permite compreender não apenas as capacidades individuais de cada abordagem, mas também as direções emergentes na área de visão computacional aplicada ao sensoriamento remoto e à agricultura de precisão.

A análise evidencia que, embora as CNNs clássicas — como U-Net e DeepLabV3+ — mantenham relevância pela simplicidade, robustez e grande disponibilidade de implementações, seu desempenho tende a saturar em cenários de alta variabilidade espacial e espectral, como é comum em ambientes agrícolas. Além disso, sua limitação em capturar dependências espaciais de longo alcance torna-se um fator crítico quando se trabalha com ortofotos de grande abrangência geográfica, que exigem compreensão contextual ampla.

Modelos baseados em Transformers, por sua vez, apresentam capacidade superior em aprender relações espaciais globais, o que tem impulsionado sua adoção em tarefas que envolvem heterogeneidade estrutural e múltiplas escalas. Arquiteturas eficientes como o SegFormer oferecem excelente compromisso entre desempenho e custo computacional, tornando-se particularmente adequadas para cenários agrícolas em que grandes áreas precisam ser processadas de forma rápida e escalável. Abordagens hierárquicas, como a Swin-

Tabela 2.1: Comparativo detalhado de modelos de segmentação semântica

Modelo	**Tipo / Arquitetura**	**Domínio / Foco**	**Métrica**	**Limitações**
U-Net	CNN encoder-decoder	Médica/Agrícola	Dice	Contexto local limitado, sensível a ruído
DeepLab V3+	CNN atrous/ASPP	Urbano, agrícola	mIoU	Complexidade, tuning preciso
Vision Transformer (ViT)	Transformer puro	Geral	Accuracy	Forte dependência de grandes datasets; alto custo
SegFormer	Transformer eficiente	Multidomínio	mIoU	Trade-off acurácia/velocidade; menor interpretabilidade
Swin-Unet	Transformer hierárquico	Alta resolução	DSC e HD	Ótimo detalhamento; maior demanda computacional
TransUNet	CNN + Transformer híbrido	Biomédico, geral	Dice	Robusto a ruído; treinamento complexo
InternImage	Convoluções deformáveis	Fundacional, geral	mIoU, Accuracy, mAP	SOTA e alto custo de hardware
ViT-CoMer	Modelo híbrido	Multiescala, geral	AP	Preciso e eficiente; arquitetura complexa
SAM	Prompt-based, zero-shot	Geral, remoto	mIoU	Excelente adaptação; dependência de prompts
CLIP	Multimodal (visão+texto)	Zero-shot / open-vocabulary	Accuracy	Não gera máscaras nativas; exige adaptações

Unet, demonstram especial habilidade para preservar detalhes finos sem comprometer a representação global.

A ascensão de modelos fundacionais — como o SAM — representa uma mudança de paradigma ao permitir segmentação zero-shot ou guiada por prompts, minimizando a dependência de anotações extensivas. Esses modelos são particularmente relevantes para agricultura, onde a coleta e rotulagem de dados é cara e demorada. No entanto, sua integração prática ainda enfrenta desafios, como a necessidade de engenharia de prompts, o ajuste a ambientes altamente heterogêneos e limitações na segmentação de objetos muito pequenos ou confusos com o fundo.

As tendências mais recentes da literatura indicam uma convergência de linhas metodológicas, unindo Transformers, convoluções deformáveis, aprendizado com poucos dados e algoritmos multimodais [29, 52, 19]. Essa convergência dialoga com a necessidade crescente de modelos capazes de operar em múltiplos domínios, generalizar bem entre órgãos agrícolas, variedades de culturas e diferentes condições climáticas.

Outro ponto crucial refere-se ao desenvolvimento de modelos leves e eficientes, aptos para execução em dispositivos embarcados e plataformas edge, como drones, robôs agrícolas ou estações de campo. A demanda por processamento em tempo real cresce continuamente, uma vez que decisões operacionais — como detecção de falhas de plantio, mapeamento de mato-competição ou identificação de árvores isoladas — dependem de respostas rápidas [32, 19].

Adicionalmente, destaca-se como tendência a integração de dados provenientes de múltiplos sensores (RGB, multiespectral, LiDAR, SAR), uma abordagem que amplia a robustez e reduz ambiguidades em cenários complexos [27]. O uso de modelos multimodais e estratégias de adaptação zero-shot ou few-shot tende a crescer, especialmente em aplicações que exigem generalização espacial e temporal — uma característica central no monitoramento agrícola.

Em síntese, ao longo deste capítulo foi apresentada uma visão abrangente da evolução dos modelos de segmentação semântica, desde as CNNs clássicas até os mais recentes modelos fundacionais e multimodais. O cenário atual aponta para um futuro em que abordagens multiparadigma, multimodais e computacionalmente eficientes serão essenciais para lidar com a diversidade ambiental, a escassez de dados anotados e a necessidade de análises rápidas e escaláveis no campo. Esse avanço tecnológico será determinante para consolidar a agricultura de precisão como pilar da sustentabilidade, produtividade e tomada de decisão baseada em dados.

CAPÍTULO

3

Materiais e Métodos

3.1 Visão Geral da Proposta

O objetivo deste trabalho é desenvolver e avaliar uma metodologia inovadora para a segmentação automática de ortofotos agrícolas, focada na identificação precisa de *talhões* e *árvores isoladas* em ambientes heterogêneos. A proposta central reside na investigação sistemática do impacto do tamanho dos *patches* de entrada e das estratégias de amostragem sobre a qualidade da segmentação, frente ao forte desbalanceamento de classes inerente aos cenários agrícolas.

Diferentemente de abordagens tradicionais que utilizam a simples repartição das imagens ou amostragem aleatória, esta pesquisa propõe estratégias sistemáticas de amostragem de patches projetadas para atacar diferentes desafios: a representação adequada dos elementos minoritários e o refinamento dos contornos de segmentação. A nossa ideia é que a escolha combinada entre tamanho dos patches e estratégia de amostragem pode mitigar o efeito do desbalanceamento de classes e maximizar a eficácia dos modelos supervisionados.

3.2 Conjunto de Dados

O conjunto de dados é composto por 92 ortofotos, cada uma representando uma propriedade agrícola diferente com um ou mais talhões e árvores. As ortofotos foram divididas aleatoriamente em 70% para treinamento (65 ortofotos), 15% para validação (13 ortofotos) e 15% para teste (14 ortofotos). A

distribuição está apresentada na Tabela 3.1, incluindo a relação aproximada de quilômetros quadrados e a quantidade de árvores por fazenda para cada conjunto de dados.

Tabela 3.1: Distribuicao dos dados.

Conjunto de Dados	**Área total aproximada (km^2)**	**Quantidade de Árvores**
Treino	92.02	640
Validação	28.14	187
Teste	23.92	207
Total	144.08	1034

Todas as ortofotos, utilizadas neste estudo, foram anotadas manualmente utilizando o software *QGIS*, em sua versão 3.22, delimitando cuidadosamente as regiões correspondentes aos *talhões* (em vermelho) e às *árvores* (em azul) conforme Figura 3.1. Esse processo manual foi essencial para garantir consistência e qualidade ao conjunto de dados. Para aumentar a representatividade e variabilidade do conjunto de dados, foram incluídos talhões em diferentes estágios de crescimento, desde recém-plantados com baixa cobertura do solo (como na Figura 3.2a) até estágios avançados, com dossel denso e homogêneo (Figura 3.2b).

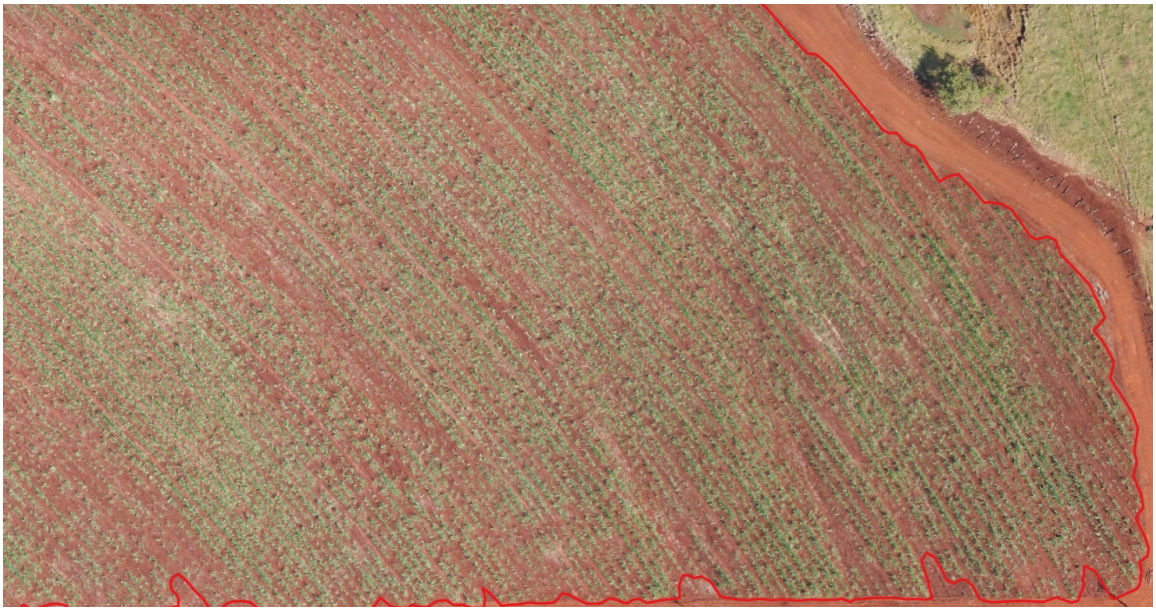
Figura 3.1: Detalhe das anotações, onde a classe *talhões* está em vermelho e *árvores* em azul. A resolução da imagem é de 3 cm/pixel.

A diversidade geográfica e fenológica dos talhões (Figura 3.2), somada à variedade de características das árvores, como espécie, estrutura da copa, contorno e presença de galhos secos (Figura 3.3), foi de grande importância para aumentar a representatividade e variabilidade do conjunto de dados, o que possibilitou a generalização para a tarefa de segmentação.

A Figura 3.4 apresenta uma das ortofotos utilizadas no conjunto de treino,

Figura 3.2: Diferentes fazendas de cultivo de cana de açúcar contendo anotações de talhões (com contorno vermelho).

(a)



(b)

Figura 3.3: Outros exemplos de árvores presentes no conjunto de treino.

(a) (b)

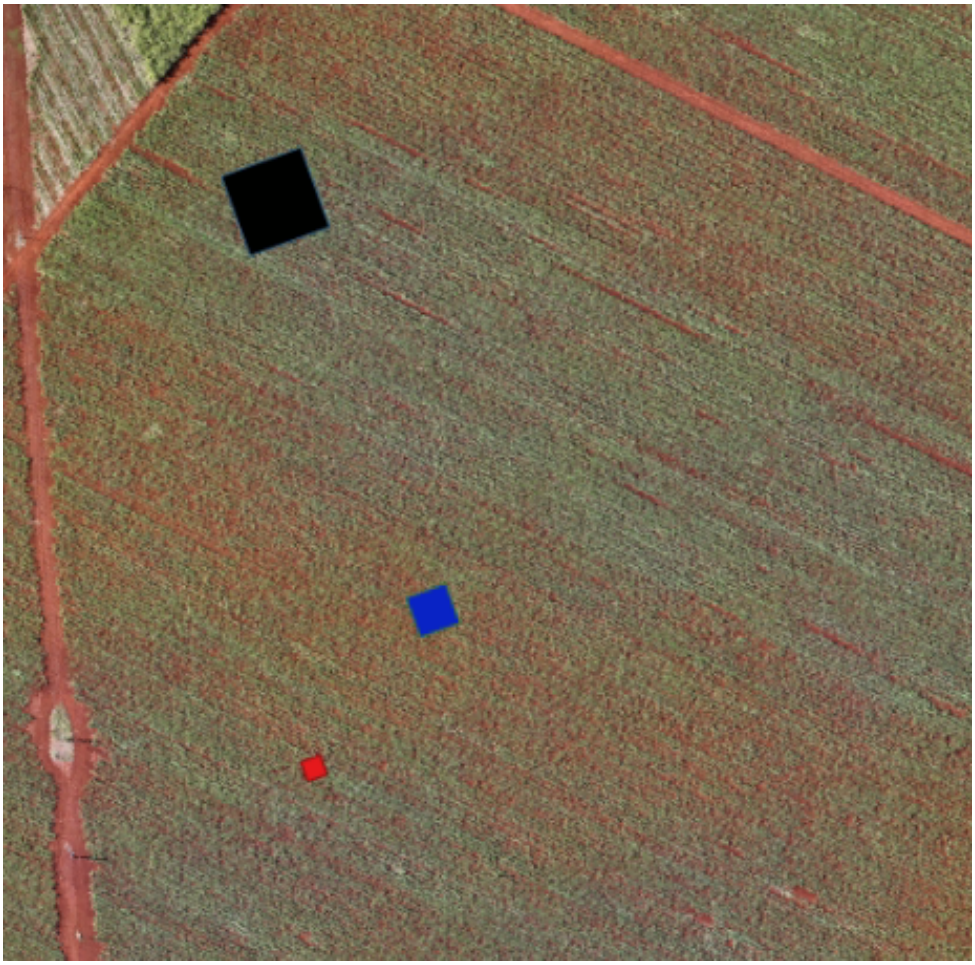
com aproximadamente 1,7 km^2 de área e resolução espacial de 3 cm/pixel. Essa elevada resolução gera imagens com milhões de pixels, tornando inviável o seu processamento direto por redes neurais profundas. Além disso, as extensas áreas correspondentes aos talhões, contrastando com a baixa frequência de árvores e outros elementos, acentuam o desequilíbrio entre classes no conjunto de dados.

As Figuras 3.4 e 3.5 ilustram, de forma comparativa, a relação entre a ortofoto e as dimensões reais dos *patches* empregados no processo de segmentação. Na Figura 3.4, a ortofoto é exibida na escala aproximada de 1:4000, com os talhões destacados em vermelho e as árvores em azul. Já a Figura 3.5 mostra uma ampliação da mesma região, na escala aproximada de 1:2000, evidenciando a proporção entre os *patches* e a área original. Nessa representação, o quadrado vermelho corresponde ao *patch* de 256×256 pixels, o azul ao de 512×512 pixels e o preto ao de 1024×1024 pixels.

Figura 3.4: Exemplo de uma ortofoto de uma fazenda presente no conjunto de dados de treino, onde os talhões são contornados em vermelho e as árvores em azul.

Com o intuito de contornar os desafios impostos pelo grande tamanho das

Figura 3.5: Retrato da real dimensão do tamanho do pacth em relação ao tamanho da ortofoto. O quadrado em vermelho corresponde ao *patch* de 256×256, o quadrado em azul ao de 512×512 pixels e o quadrado em preto ao de 1024×1024 pixels.



imagens e pelo desequilíbrio entre classes, as ortofotos foram particionadas em *patches* menores. Além disso, diferentes estratégias de amostragem foram empregadas para garantir que as classes menos representadas fossem adequadamente contempladas durante o treinamento.

3.3 Estratégias de Amostragem Propostas

3.3.1 Motivação e Hipótese

A escolha do tamanho dos patches em segmentação de ortofotos agrícolas de alta resolução é um fator muito importante porque influencia diretamente a qualidade dos resultados obtidos. Essa decisão impacta a capacidade do modelo de representar tanto padrões locais quanto estruturas globais presentes nas imagens, especialmente em cenários onde coexistem objetos de diferentes escalas, como áreas amplas e elementos isolados. Assim, justificar criteriosamente o tamanho dos patches e a amostragem é fundamental para evitar

possíveis perdas de informação e garantir que a segmentação seja adequada à heterogeneidade do ambiente mapeado.

A motivação deste trabalho nasce da observação de que a escolha do tamanho dos patches e das estratégias de amostragem não é trivial e pode influenciar diretamente a capacidade do modelo em captar detalhes relevantes e contexto. Isso provoca inconsistências nos resultados e dificulta a generalização dos modelos para diferentes tipos de elementos presentes nas ortofotos, impactando negativamente a aplicação prática na agricultura de precisão.

A hipótese que norteia esta pesquisa é que existe um tamanho ideal de patches para a segmentação eficiente das ortofotos agrícolas, que combina a preservação de detalhes finos com o contexto necessário para a correta delimitação das classes. Ademais, a aplicação de estratégias de amostragem direcionadas é fundamental para mitigar o desbalanceamento entre classes presentes nas imagens, como áreas homogêneas e elementos minoritários. Assim, a combinação apropriada de tamanho de patch e amostragem pode maximizar a qualidade e robustez da segmentação automática.

Vale destacar que a definição do tamanho dos patches está intrinsecamente relacionada à resolução espacial das ortofotos utilizadas neste trabalho, de 3cm/pixel. Essa resolução implica que patches de 256×256, 512×512 e 1024×1024 pixels correspondem a áreas aproximadas de 7.68m, 15.36m e 30.72m de lado, respectivamente. Assim, o tamanho ideal de patch não é uma escolha arbitrária, mas sim uma função direta da escala dos elementos presentes nas imagens. Patches muito pequenos podem limitar o contexto necessário para a correta delimitação dos talhões, enquanto patches muito grandes podem diluir detalhes importantes, como a forma e a assinatura visual das árvores isoladas. Dessa forma, a busca pelo tamanho ótimo de patch deve considerar simultaneamente a resolução da imagem e a escala espacial dos alvos de interesse.

3.3.2 Descrição das Estratégias

Para contornar a limitação de processamento imposta pela alta resolução das imagens, as ortofotos foram divididas em patches menores, através do Algoritmo 1, e diferentes estratégias de amostragem foram aplicadas para garantir que as classes menos frequentes fossem representadas adequadamente durante o treinamento. A escolha dos tamanhos dos patches foi guiada pelo objetivo de equilibrar o contexto espacial e a preservação dos detalhes finos. Por isso, selecionamos três tamanhos para análise: 256x256, 512x512 e 1024x1024 pixels. Patches menores, como os de 256x256, são eficazes para a segmentação de elementos pequenos, como árvores isoladas e bordas finas de talhões, por preservarem detalhes importantes. Em contrapartida, patches

maiores, como os de 1024x1024, fornecem mais contexto espacial, o que é útil para a segmentação de áreas extensas, mas podem diluir detalhes finos e, com isso, prejudicar a precisão em elementos menores.

Uma janela de sobreposição de 50% foi adotada para mitigar o efeito de borda e garantir que os objetos, como árvores e as bordas de talhões, não fossem perdidos na divisão, o que é uma prática comum em arquiteturas de segmentação baseadas em patches [7, 20, 34].

Entrada: Diretório raiz com ortofotos e shapefiles.
Saída : Patches de imagens RGB e seus respectivos rótulos.
Carregar a configuração do experimento, que especifica o diretório raiz contendo as ortofotos e shapefiles;
for *cada diretório* **do**
  Ler a ortofoto (.tif file) e seus shapefiles associados (.shp files);
  Gerar uma máscara binária para a área a ser processada, incluindo máscaras separadas para cada classe;
  Extrair os canais vermelho, verde e azul da imagem da ortofoto;
  for *Cada combinação de tamanho do patch* **do**
    Calcular o tamanho do passo como 50% do tamanho do patch;
    Deslizar uma janela por toda a imagem com o tamanho e passo especificados;
    for *Cada patch da janela* **do**
      if *o patch contém informação relevante* **then**
        Extrair o patch RGB correspondente e o rótulo;
        Salvar o patch RGB e a imagem rotulada em um diretório;
        Aplicar mapeamento de cores à imagem de rótulo;
      end
    end
  end
end

Algorithm 1: Algoritmo de Geração de Patches

Para lidar com o significativo desequilíbrio de classes presente nos dados, onde as áreas de *talhõ* são muito maiores que as de *árvores* e outras classes, exploramos três estratégias de amostragem de patches. Essas estratégias foram aplicadas a cada uma das tarefas de segmentação de forma independente (segmentação dos *talhões* e segmentação das *árvores*).

1. **Estratégia 1: Amostragem de todos os patches (*All-Patches*)**: Todos os patches gerados a partir das ortofotos foram utilizados no treinamento, sem qualquer critério de seleção prévio. Esta abordagem serve como uma linha de base, avaliando o desempenho do modelo quando exposto a um conjunto de dados com o desequilíbrio de classes original, onde a classe majoritária domina o treinamento. Essa estratégia inclui patches

de fundo puro (ver Figura 3.6) e também patches com fundo predominante (ver Figura 3.6).

2. **Estratégia 2: Amostragem focada na classe (*Crop-Focused*)**: Para garantir que o modelo aprendesse a segmentar as classes de interesse de forma eficaz, esta estratégia focou na seleção de patches que continham as classes alvo. Foram utilizados apenas patches que continham pelo menos 1 pixel de *talhões* ou *árvores*. O objetivo foi aumentar a proporção de amostras relevantes no conjunto de treinamento e, assim, forçar o modelo a aprender as características visuais dessas classes. Exemplos incluem patches com classe de interesse pura (Figura 3.6) e patches com pequena presença de classe de interesse (Figura 3.6).

3. **Estratégia 3: Amostragem por zona de borda (*Boundary Zone*)**: Esta estratégia foi projetada para aprimorar a capacidade do modelo de identificar os limites das classes. Para isso, o treinamento foi focado em patches que continham as bordas, ou seja, onde a classe de interesse fazia transição para o *fundo*. Foram utilizados apenas patches que continham pelo menos 1 pixel de *talhões* ou *árvores* e ao mesmo tempo pelo menos 1 pixel de *fundo*. Ao priorizar o treinamento nessas áreas de transição, espera-se uma melhoria na precisão da segmentação dos contornos. Exemplos são os patches de transição com predominância da classe de interesse (Figura 3.6) e patches com mistura de classe de interesse e fundo (Figura 3.6).

A Tabela 3.2 mostra a quantidade de patches gerados para cada tamanho e estratégia em cada conjunto de dados.

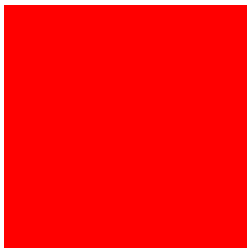
Tabela 3.2: Quantidade de patches por tamanho, estratégia de amostragem e conjunto de dados para as classes *Talhões* e *Árvores*. Todas as estratégias foram validadas e testadas utilizando a estratégia All-Patches, por conta disso, os respectivos campos na tabela estão com --".

Estratégia	**Tamanho do Patch**	**Talhões**			**Árvores**		
		Treino	**Validação**	**Teste**	**Treino**	**Validação**	**Teste**
All-Patches	256x256	2,443,885	448,451	646,263	63,978	13,248	23,870
	512x512	628,310	125,520	166,315	22,148	4,878	8,110
	1024x1024	165,746	33,008	43,968	9,034	2,084	3,206
Crop-Focused	256x256	1,869,472	–	–	17,516	–	–
	512x512	489,438	–	–	8,071	–	–
	1024x1024	132,675	–	–	4,428	–	–
Boundary Zone	256x256	535,484	–	–	16,412	–	–
	512x512	123,923	–	–	8,050	–	–
	1024x1024	56,446	–	–	4,428	–	–

Figura 3.6: Exemplos de patches utilizados nas diferentes estratégias de amostragem.

(a) Patch com 100% fundo — Estratégia 1 (All-Patches).

(b) Patch com 100% de classe de interesse — Estratégia 2 (Crop-Focused).



(c) Patch de transição — Estratégia 3 (Boundary Zone).

(d) Patch classe + fundo misturados — Estratégia 3.

(e) Patch pequena presença da classe — Estratégia 2.

(f) Patch de fundo predominante — Estratégia 1.

3.3.3 Fluxo Metodológico Geral

O processo metodológico proposto inicia-se com a divisão das ortofotos em patches de tamanhos distintos — 256x256, 512x512 e 1024x1024 pixels —, buscando equilibrar o contexto espacial e a preservação de detalhes finos relevantes para a segmentação. Após essa divisão, foram aplicadas as estratégias de amostragem projetadas para lidar com o desbalanceamento de classes, garantindo que as classes menos frequentes, como árvores isoladas, estivessem adequadamente representadas no conjunto de treinamento.

Os patches gerados foram então utilizados no treinamento e validação dos modelos de segmentação, abrangendo arquiteturas clássicas como U-Net e DeepLabv3, bem como modelos modernos baseados em transformers (SegFormer) e convoluções deformáveis (InternImage). Cada modelo é avaliado quanto à sua capacidade de segmentar, de forma independente, os talhões e as árvores isoladas, considerando as diferenças morfológicas e estatísticas entre essas classes.

Para garantir a eficácia do treinamento, os experimentos são organizados em duas etapas principais: inicialmente, a definição do tamanho de patch ideal, seguida da comparação das estratégias de amostragem mais promissoras. O fluxo completo da metodologia inclui o pré-processamento das imagens, a geração dos patches, a aplicação das amostragens, o treinamento dos modelos e a avaliação final, buscando a melhor combinação para maximizar a

qualidade da segmentação nas ortofotos agrícolas.

3.4 Modelos de Segmentação

Neste trabalho, foram considerados quatro modelos representativos de diferentes gerações de abordagens: U-Net, DeepLabv3+, SegFormer e InternImage. Cada um deles ilustra um marco na evolução dos métodos de segmentação.

- **U-Net**: Desenvolvida originalmente para aplicações biomédicas, a U-Net consolidou-se como uma das arquiteturas mais populares para segmentação. Sua principal característica é a estrutura em formato de "U", com um caminho de codificação (encoder) que extrai as características (*features*) e um caminho de decodificação (decoder) que reconstrói a máscara, incluindo conexões de pulo *skip connections* que preservam detalhes espaciais. É simples, eficiente e serve como um forte ponto de comparação inicial.

- **DeepLabv3+**: É uma evolução da família DeepLab, que introduz a técnica de *Atrous Spatial Pyramid Pooling* (ASPP) para capturar informações em múltiplas escalas. A versão "+" aprimora o decoder, tornando a reconstrução espacial mais precisa. Esse modelo é reconhecido por seu bom equilíbrio entre detalhamento de bordas e robustez na detecção de objetos de diferentes tamanhos, amplamente utilizada em tarefas de segmentação semântica de modo geral.

- **SegFormer**: Representa a transição das CNNs para arquiteturas baseadas em *transformers*. Seu backbone MiT (*Mix Transformer*) combina extração hierárquica de características (*features*) com o mecanismo de auto-atenção (*self-attention*), permitindo capturar dependências globais. O *decoder*, por outro lado, é implementado com um conjunto de MLPs projetados de forma simples e eficiente, o que resulta em baixo custo computacional. Esse modelo foi escolhido pelo seu desempenho competitivo aliado a uma arquitetura enxuta.

- **InternImage**: É uma arquitetura mais recente, baseada em convoluções deformáveis (DConv) que se ajustam de forma adaptativa ao conteúdo da imagem. Essa abordagem combina a eficiência das convoluções com a capacidade de modelar relações mais complexas, alcançando resultados de ponta em benchmarks recentes de segmentação. Sua inclusão visa avaliar se essas melhorias de última geração trazem ganhos significativos para o domínio específico deste estudo.

Assim, a escolha desses quatro métodos permite analisar desde arquiteturas clássicas baseadas em CNNs até propostas modernas baseadas em *transformers* e convoluções dinâmicas, oferecendo um panorama abrangente da evolução dos modelos de segmentação.

3.5 Métricas de Avaliação

Para comparar o desempenho dos modelos, utilizamos diversas métricas de segmentação. Todas elas são calculadas a partir de quatro valores fundamentais:

- **VP (Verdadeiro Positivo):** Pixels corretamente identificados como pertencentes à classe de interesse.
- **FP (Falso Positivo):** Pixels incorretamente classificados como sendo da classe de interesse.
- **VN (Verdadeiro Negativo):** Pixels corretamente identificados como pertencentes ao fundo.
- **FN (Falso Negativo):** Pixels que deveriam ser classificados como da classe de interesse, mas foram identificados como fundo.

A partir desses valores, derivam-se diferentes métricas que permitem avaliar tanto a qualidade da segmentação quanto o equilíbrio entre erros e acertos do modelo. As principais utilizadas neste trabalho são:

- **IoU (Intersection over Union)**
  Também conhecido como Coeficiente de *Jaccard*, mede a sobreposição entre a máscara predita e a máscara real.

$$IoU = \frac{VP}{VP + FP + FN} \tag{3.1}$$

- **Coeficiente de Dice (Dice Coefficient)**
  Muito utilizado em segmentação biomédica, é uma métrica semelhante ao IoU, mas dá maior peso aos acertos.

$$Dice = \frac{2 \times VP}{2 \times VP + FP + FN} \tag{3.2}$$

- **Precision (Precisão)**
  Mede a proporção de pixels corretamente classificados como a classe de interesse entre todos os que foram preditos como tal.

$$Precision = \frac{VP}{VP + FP} \tag{3.3}$$

- **Recall (Revocação ou Sensibilidade)**
  Mede a proporção de pixels da classe de interesse que foram corretamente identificados pelo modelo.

$$Recall = \frac{VP}{VP + FN} \tag{3.4}$$

- **F1-Score**
  Representa a média harmônica entre Precision e Recall, oferecendo um equilíbrio entre as duas métricas. É matematicamente equivalente ao Coeficiente de Dice.

$$FScore = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3.5}$$

- **Accuracy (Acurácia)**
  Mede a proporção total de pixels classificados corretamente (tanto positivos quanto negativos). Contudo, pode ser enganosa em bases desbalanceadas.

$$Accuracy = \frac{VP + VN}{VP + VN + FP + FN} \tag{3.6}$$

3.6 Configurações Experimentais

Os experimentos foram realizados em uma *WorkStation* equipada com GPU NVIDIA GeForce RTX 3090, CPU AMD Ryzen 9 5950X e 96 GB de RAM. O ambiente de software foi configurado em Ubuntu 22.04, utilizando Python 3.8.19, CUDA 12.1, PyTorch 2.3.0+cu121, TorchVision 0.18.0, OpenCV 4.9.0 e MMEngine 0.10.4, além do framework MMSegmentation da OpenMMLab [9].

Foram avaliados quatro modelos de segmentação semântica, todos configurados para tarefas binárias (duas classes): **SegFormer** (backbone MiT-B5), **DeepLabv3+** (backbone ResNet-101-D8), **UNet** (S5-D16 com FCN) e **InternImage** (DCNv3 integrado ao *head* do SegFormer). Os pesos pré-treinados foram empregados em todos os casos, e o treinamento utilizou o otimizador AdamW com taxa de aprendizado inicial de 6×10^{-5}, *decaimento de peso* de 0.01 e duração de 80 mil *iterações*.

As imagens de entrada foram divididas em *patches* de três tamanhos distintos: 256×256, 512×512 e 1024×1024 pixels. O tamanho do *batch* foi definido de forma proporcional à resolução, respeitando a capacidade da GPU: 32 para 256×256, 8 para 512×512 e 1 para 1024×1024.

Durante o treinamento, a cada 8 mil iterações os modelos eram avaliados com base nas métricas **mIoU** e **mFscore**, sendo o melhor resultado salvo de acordo com o valor de mIoU. Além disso, foi desenvolvido um script especí-

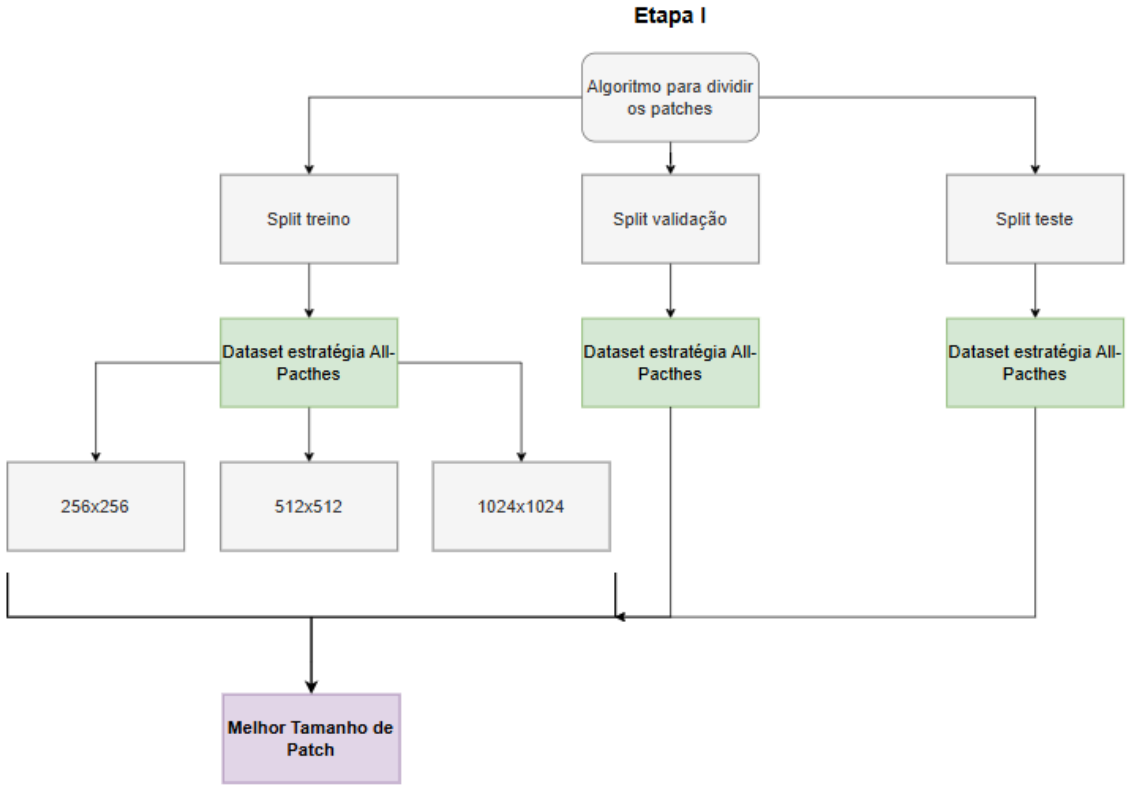
fico para a avaliação em larga escala, que permitiu realizar a predição sobre ortofotos completas (Algoritmo 2).

O processo experimental foi organizado em duas etapas principais. Na Etapa 1, a definição do melhor tamanho de patch foi realizada exclusivamente com o modelo *SegFormer*, escolhido como arquitetura de referência devido ao seu bom desempenho e custo computacional equilibrado. Assim, todos os experimentos dessa etapa foram conduzidos mantendo o *SegFormer* fixo e variando apenas o tamanho dos patches. Somente após essa definição os demais modelos foram avaliados na Etapa 2.

Em seguida, na Etapa 2, com o melhor tamanho de patch já definido, passamos a comparar as diferentes estratégias de amostragem. Avaliamos as estratégias All-Patches, Crop-Focused e Boundary-Zone, aplicando cada uma delas em todas as arquiteturas de segmentação consideradas: U-Net, DeepLabv3+, SegFormer e InternImage. Essa abordagem multifacetada permitiu uma melhor análise do impacto da amostragem e da arquitetura no desempenho final da segmentação, conforme o fluxo da Figura 3.8.

A fim de tornar a análise mais precisa, as tarefas de segmentação foram tratadas de forma independente, resultando em dois modelos distintos para cada arquitetura avaliada: um dedicado à segmentação de talhões e outro voltado à segmentação de árvores isoladas. Essa separação foi necessária devido às diferenças morfológicas e estatísticas entre as classes — enquanto os talhões apresentam grandes áreas contínuas e textura homogênea, as árvores são estruturas pequenas, esparsas e com alto contraste local. Assim, cada modelo pôde ser otimizado para lidar com a natureza específica de sua respectiva classe, evitando que o desbalanceamento severo entre categorias prejudicasse o aprendizado. Essa decisão metodológica também facilitou a análise comparativa dos resultados, permitindo avaliar de forma mais clara o impacto de cada arquitetura e estratégia de amostragem sobre problemas com diferentes níveis de granularidade espacial.

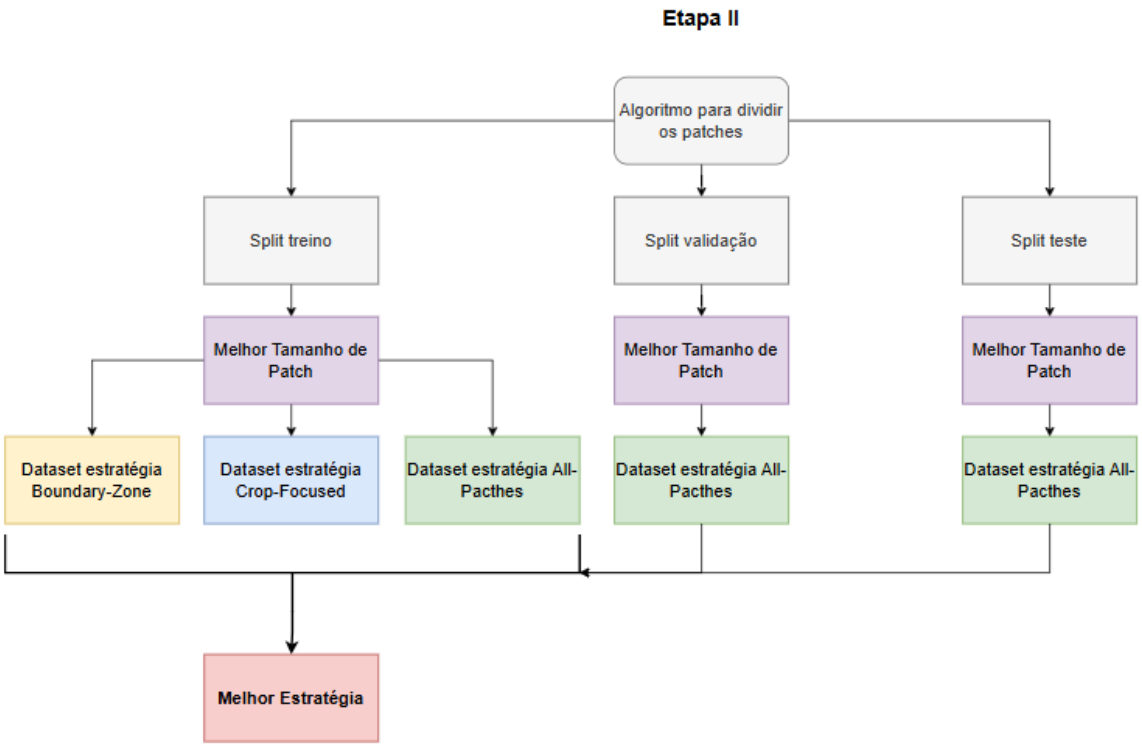
Figura 3.7: Fluxo de treinamento para escolha do melhor tamanho de patch e posteriormente a melhor estratégia.



```
Entrada: Dados de entrada do algoritmo
Saída  : Resultados de saída do algoritmo
Carregar configuração do modelo;
Definir parâmetros de inferência (patch size, passo, dispositivo);
for cada ortofoto no diretório do
    if não existir .tif ou .shp then
        Ignorar diretório e continuar;
    end
    else
        Inicializar matriz de probabilidades;
        for cada talhão no shapefile do
            for cada posição (x,y) na região do
                Extrair patch;
                if patch contém pixels da classe alvo: then
                    Adicionar à lista;
                    if lista atingir batch size then
                        Inferir com modelo;
                        Atualizar matriz;
                        Limpar lista;
                    end
                end
            end
            if restarem patches na lista then
                Inferir e atualizar matriz;
            end
        end
        Calcular mapa final;
        Aplicar máscara e salvar shapefile;
    end
end
```

Algorithm 2: Algoritmo de Predição em Ortofotos

Figura 3.8: Fluxo de treinamento para escolha da melhor estratégia a partir do melhor tamanho de patch, definido na Figura 3.7.



CAPÍTULO

4

Resultados

4.1 Resultados Quantitativos

Nesta seção são apresentados e analisados os resultados quantitativos obtidos nos experimentos de segmentação semântica realizados. O objetivo é compreender o impacto de três fatores principais no desempenho dos modelos: (i) o tamanho dos patches utilizados no treinamento, (ii) as diferentes estratégias de amostragem — *All-Patches*, *Crop-Focused* e *Boundary-Zone* — e (iii) a arquitetura das redes neurais empregadas, incluindo U-Net, DeepLabv3+, SegFormer e InternImage.

Os resultados são expressos nas métricas mais utilizadas em segmentação de imagens — IoU (*Intersection over Union*), *Accuracy* (Acc), *Fscore*, *Precision* e *Recall* — que permitem avaliar o desempenho global e a capacidade dos modelos em identificar corretamente as classes.

Tabela 4.1: Resultados da estratégia All-Pacthes com diferentes tamanhos de patch com o SegFormer.

Patch	Class	IoU	Acc	Fscore	Precision	Recall
1024	**Background**	75.06	92.22	85.75	79.24	92.22
	Talhões	84.83	88.23	91.79	95.65	88.23
512	**Background**	**81.60**	87.44	89.87	92.43	87.44
	Talhões	**91.90**	96.90	95.78	94.68	96.90
256	**Background**	73.03	92.06	84.41	77.94	92.06
	Talhões	87.05	89.76	93.08	76.64	89.76

A análise dos resultados evidencia de forma clara a influência combinada do tamanho dos patches, da estratégia de amostragem e da arquitetura dos

Tabela 4.2: Resultados comparativos dos modelos UNet, SegFormer, InternImage e DeepLabv3+ na estratégia *All-Patches* com tamanho de patch 512x512.

Modelo	Classe	IoU	Acc	Fscore	Precision	Recall
DeepLabv3+	**Background**	80.86	86.07	89.41	93.03	86.07
	Talhões	91.67	97.21	95.65	94.15	97.21
UNet	**Background**	74.08	88.51	85.11	81.96	88.51
	Talhões	87.22	91.56	93.17	94.84	91.56
SegFormer	**Background**	**81.60**	87.44	89.87	92.43	87.44
	Talhões	**91.90**	96.90	95.78	94.68	96.90
InternImage	**Background**	**82.08**	89.17	90.16	91.16	89.17
	Talhões	**91.94**	96.25	95.80	95.35	96.25

Tabela 4.3: Resultados comparativos dos modelos UNet, SegFormer, InternImage e DeepLabv3+ na estratégia *Crop-Focused* com tamanho de patch 512x512.

Modelo	Classe	IoU	Acc	Fscore	Precision	Recall
DeepLabv3+	**Background**	81.46	85.35	89.78	94.70	85.35
	Talhões	92.09	97.93	95.88	93.91	97.93
UNet	**Background**	80.55	86.83	89.23	91.77	86.83
	Talhões	91.41	96.63	95.51	94.42	96.63
SegFormer	**Background**	**83.57**	88.85	91.05	93.36	88.85
	Talhões	**92.78**	97.26	96.25	95.27	97.26
InternImage	**Background**	**85.20**	88.34	92.01	95.99	88.34
	Talhões	**93.67**	98.40	96.73	95.12	98.40

Tabela 4.4: Resultados comparativos dos modelos UNet, SegFormer, InternImage e DeepLabv3+ na estratégia *Boundary-Zone* com tamanho de patch 512x512.

Modelo	Classe	IoU	Acc	Fscore	Precision	Recall
DeepLabv3+	**Background**	78.66	90.20	88.05	86.01	90.20
	Talhões	89.82	93.64	94.64	95.66	93.64
UNet	**Background**	76.80	84.95	86.88	88.90	84.95
	Talhões	89.56	95.40	94.49	93.60	95.40
SegFormer	**Background**	**79.39**	83.85	88.51	93.72	83.85
	Talhões	**91.18**	97.56	95.39	93.31	97.56
InternImage	**Background**	**83.08**	88.63	90.76	93.00	88.63
	Talhões	**92.55**	97.11	96.13	95.17	97.11

Tabela 4.5: Resultados do modelo com diferentes tamanhos de patch para a classe árvores com o SegFormer.

Patch	Class	IoU	Acc	Fscore	Precision	Recall
1024	**Background**	97.48	98.59	98.73	98.86	98.59
	Árvores	58.43	75.91	73.76	71.73	75.91
512	**Background**	**97.16**	98.48	98.56	98.64	98.48
	Árvores	**68.69**	82.32	81.44	80.58	82.32
256	**Background**	95.91	97.39	97.91	98.44	97.39
	Árvores	68.83	85.57	81.54	77.87	85.57

Tabela 4.6: Resultados comparativos dos modelos UNet, SegFormer, InternImage e DeepLabv3+ na estratégia *All-Patches* com tamanho de patch 512x512.

Patch	Class	IoU	Acc	Fscore	Precision	Recall
Deeplabv3+	**Background**	96.98	98.31	98.47	98.62	98.31
	Árvores	67.26	82.08	80.43	78.84	82.08
Unet	**Background**	96.48	97.36	98.21	99.07	97.36
	Árvores	65.50	88.02	79.15	71.91	88.02
SegFormer	**Background**	**97.16**	98.48	98.56	98.64	98.48
	Árvores	**68.69**	82.32	81.44	80.58	82.32
InternImage	**Background**	**98.08**	98.82	99.03	99.25	98.82
	Árvores	**78.19**	90.27	87.76	85.39	90.27

Tabela 4.7: Resultados comparativos dos modelos UNet, SegFormer, InternImage e DeepLabv3+ na estratégia *Crop-Focused* com tamanho de patch 512x512.

Patch	Class	IoU	Acc	Fscore	Precision	Recall
Deeplabv3+	**Background**	97.57	98.45	98.77	99.09	98.45
	Árvores	73.42	88.27	84.67	81.36	88.27
Unet	**Background**	97.42	98.46	98.69	98.93	98.46
	Árvores	71.68	86.05	83.50	81.11	86.05
SegFormer	**Background**	**98.15**	98.95	99.07	99.19	98.95
	Árvores	**78.62**	89.41	88.03	86.69	89.41
InternImage	**Background**	**98.24**	98.97	99.11	99.25	98.97
	Árvores	**79.59**	90.31	88.64	87.03	90.31

Tabela 4.8: Resultados comparativos dos modelos UNet, SegFormer, InternImage e DeepLabv3+ na estratégia *Boundary-Zone* com tamanho de patch 512x512.

Patch	Class	IoU	Acc	Fscore	Precision	Recall
Deeplabv3+	**Background**	97.51	98.48	98.74	99.0	98.48
	Árvores	72.63	87.04	84.15	81.44	87.04
Unet	**Background**	96.72	97.36	98.33	99.32	97.36
	Árvores	67.95	91.31	80.92	72.65	91.31
SegFormer	**Background**	**97.56**	98.45	98.76	99.08	98.45
	Árvores	**73.27**	88.10	84.58	81.32	88.10
InternImage	**Background**	**98.19**	98.98	99.08	99.19	98.98
	Árvores	**78.93**	89.40	88.23	87.08	89.40

modelos sobre a qualidade da segmentação, tanto para os talhões quanto para as árvores isoladas.

Nos experimentos com diferentes tamanhos de patch (Tabela 4.1), observa-se que o tamanho de 512×512 pixels apresentou o melhor desempenho geral para ambas as classes. Esse valor maximizou as métricas de IoU e Fscore, especialmente na segmentação de talhões (**IoU**: 91.90; **Fscore**: 95.78). Patches menores (256×256) mostraram perda de contexto espacial, enquanto patches maiores (1024×1024) diluíram os detalhes de borda, reduzindo a precisão em áreas mais finas. Assim, o patch de 512×512 representou um equilíbrio ideal entre contexto e detalhamento.

Ao comparar as estratégias de amostragem — *All-Patches*, *Crop-Focused* e *Boundary-Zone* — nota-se um ganho expressivo ao priorizar regiões contendo a classe de interesse. Com a estratégia *Crop-Focused* (Tabela 4.3), as métricas de IoU e Fscore aumentaram de forma consistente para todas as arquiteturas, principalmente nas classes minoritárias. No caso das árvores, a melhora foi ainda mais marcante: a IoU subiu de 68.69 (na *All-Patches*) para 78.62 (SegFormer) e 79.59 (InternImage), conforme mostram as Tabelas 4.6 e 4.7. Esses resultados indicam que estratégias direcionadas ajudam a mitigar o desbalanceamento e tornam o treinamento mais sensível às regiões relevantes.

Entre as arquiteturas avaliadas, o SegFormer e o InternImage destacaram-se com os melhores resultados, superando a U-Net e o DeepLabv3+ em praticamente todas as métricas. O InternImage obteve ligeira vantagem em métricas como IoU para árvores (geralmente 1–2 p.p. acima do SegFormer), mas com custo computacional bem mais alto, devido à complexidade de seu backbone baseado em convoluções deformáveis profundas. Já o SegFormer atingiu resultados muito próximos, com diferença inferior a 2 pontos percentuais na maioria dos casos, oferecendo uma estrutura mais leve e eficiente.

Considerando o desempenho e o custo computacional, o SegFormer se mostra a opção mais equilibrada para aplicações práticas. Ele combina excelente desempenho, menor tempo de inferência e menor consumo de memória, sem comprometer significativamente a precisão. O InternImage, por outro lado, é indicado apenas para cenários em que cada ponto percentual adicional de acurácia seja realmente crítico — especialmente na detecção de classes mais desafiadoras, como árvores isoladas.

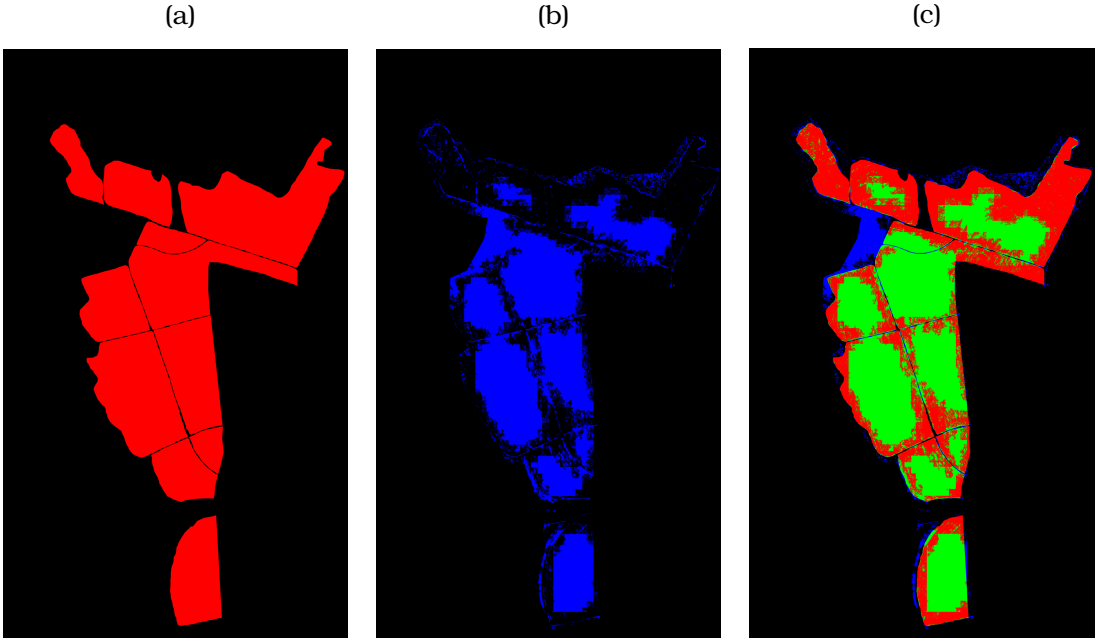
4.2 Resultados Qualitativos e Discussão

A análise visual complementa os resultados numéricos apresentados nas Tabelas 4.1 a 4.8, oferecendo uma compreensão mais visual e interpretativa do desempenho dos modelos de segmentação. As Figuras 4.2 a 4.5 ilustram

exemplos representativos das predições realizadas pelos modelos U-Net, DeepLabv3+, SegFormer e InternImage sobre ortofotos do conjunto de teste, considerando diferentes tamanhos de patch e estratégias de amostragem.

A Figura 4.1 apresenta o esquema de avaliação visual utilizado. Nela, observa-se (a) o *Ground Truth* (GT) em vermelho, (b) a predição gerada pelos modelos em azul e (c) o contraste entre acertos e erros, onde áreas em verde correspondem aos Verdadeiros Positivos (VP), em azul aos Falsos Positivos (FP) e em vermelho aos Falsos Negativos (FN). Esse formato de visualização foi adotado em todas as análises subsequentes, por facilitar a interpretação das regiões corretamente ou incorretamente segmentadas.

Figura 4.1: (a)Avaliação visual dos modelos com Ground Truth (GT) em vermelho; (b) Predição em azul, e; (c) O contraste de VP em verde, FP em azul e FN em vermelho.



Nas Figuras 4.2 e 4.3, que retratam a segmentação da classe talhões, é possível notar diferenças perceptíveis entre as arquiteturas. O modelo U-Net demonstrou boa coerência nas regiões centrais dos talhões, preservando o formato geral das áreas de plantio; contudo, apresentou tendência à supersegmentação em regiões de sombra e subsegmentação nas bordas, o que reduziu a precisão em áreas de transição entre solo e vegetação.

O DeepLabv3+ mostrou desempenho mais robusto em grandes extensões contínuas, mas suavizou os contornos, resultando em fronteiras menos nítidas e pequenas perdas de detalhes estruturais. Já o SegFormer obteve segmentações mais uniformes e contornos bem definidos, preservando tanto as regiões internas quanto as extremidades dos talhões. Sua arquitetura baseada em atenção global permitiu capturar o contexto da paisagem de maneira equilibrada, reduzindo erros em zonas de sombra e heterogeneidade.

Por fim, o InternImage apresentou os melhores resultados visuais para talhões, combinando contornos precisos e boa continuidade espacial das regiões segmentadas. O modelo manteve a integridade das fronteiras mesmo em áreas com variação de textura, evitando falhas comuns de suavização observadas em arquiteturas mais simples. Essa robustez visual reflete a capacidade das convoluções deformáveis do InternImage em se adaptar a diferentes padrões espaciais, o que resulta em máscaras mais uniformes e próximas ao Ground Truth.

As Figuras 4.4 e 4.5 mostram a segmentação da classe árvores isoladas, tarefa notoriamente mais desafiadora devido à escala reduzida e à dispersão espacial desses elementos.

O modelo U-Net teve dificuldade em distinguir árvores de pequenas manchas de vegetação, frequentemente omitindo árvores próximas às bordas dos talhões. O DeepLabv3+ apresentou leve melhora, mas ainda produziu máscaras fragmentadas, com predições descontínuas em regiões densas.

O SegFormer, em contraste, exibiu maior estabilidade na detecção de árvores individuais, identificando corretamente copas sobrepostas e reduzindo falsos positivos em áreas de vegetação rasteira.

O InternImage destacou-se como o modelo mais sensível a estruturas pequenas, preservando o formato e o contorno das árvores com alta fidelidade visual. Em alguns casos, porém, gerou pequenas supersegmentações em áreas de brilho especular ou sombra intensa, indicando que, embora altamente preciso, o modelo ainda é influenciado por variações de iluminação.

De forma geral, os resultados qualitativos confirmam as tendências observadas nas análises quantitativas. A estratégia *Crop-Focused* produziu máscaras mais detalhadas e equilibradas, especialmente nas classes minoritárias; o uso de patches de 512×512 pixels mostrou-se o melhor compromisso entre detalhamento local e contexto global; e os modelos SegFormer e InternImage geraram segmentações visualmente mais coerentes e próximas ao *Ground Truth*.

Essa superioridade qualitativa dos modelos mais avançados também se refletiu nos resultados quantitativos. O InternImage alcançou o melhor IoU médio (89,435), superando o SegFormer em 1,26 p.p., o DeepLabV3+ em 2,66 p.p. e a U-Net em 3,455 p.p. Entretanto, esse ganho vem acompanhado de um aumento significativo no custo computacional. O InternImage possui 329.973.983 parâmetros, isto é, 248 milhões a mais que o SegFormer (81.970.370), exigindo maior memória e processamento. Essa diferença se reflete diretamente no tempo de inferência: no conjunto de teste, o InternImage levou 2911 segundos (0.085 segundos por patch aproximadamente), tornando-se 1132 segundos mais lento que o SegFormer (1779 segundos - 0.052 segundos por patch).

Figura 4.2: Resultados Ilustrativos dos Modelos de Segmentação Para a Classe de Talhões em uma fazenda do conjunto de teste.

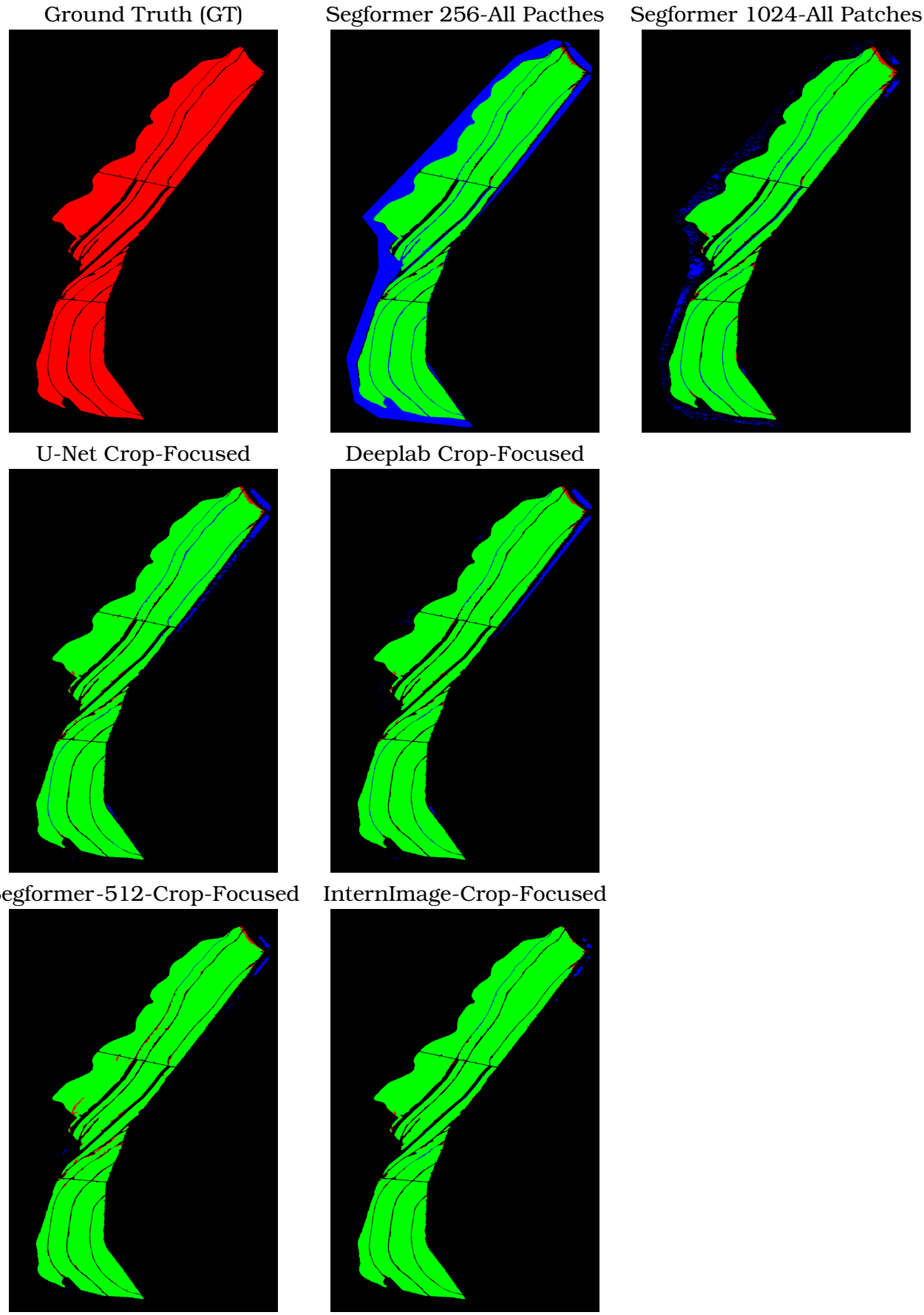
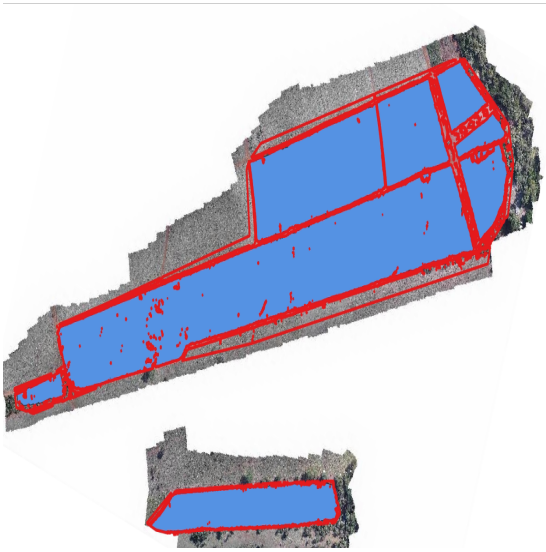
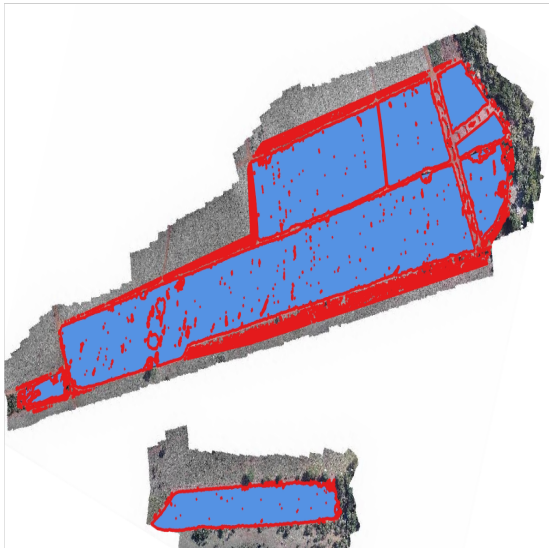


Figura 4.3: Resultados dos Modelos de Segmentação Para a Classe de Talhões em shapefile em uma outra fazenda do conjunto de teste.

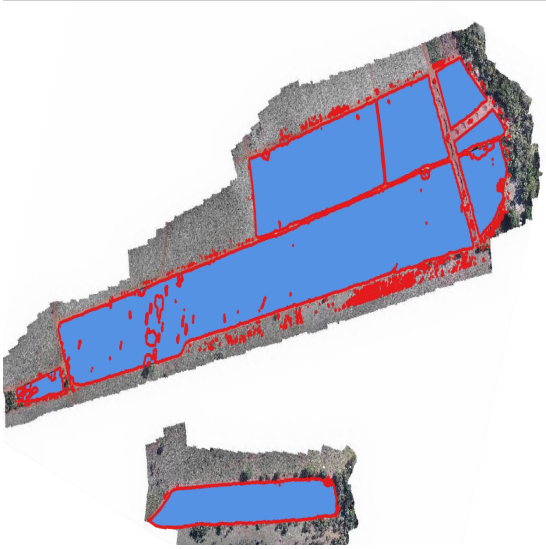
Segformer 256-All Pacthes



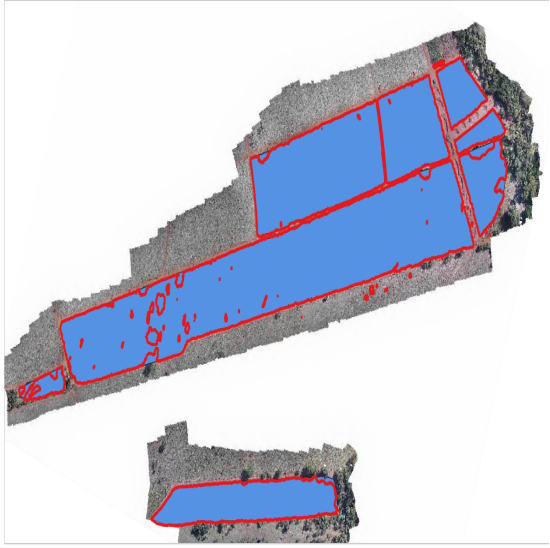
Segformer 1024-All Patches



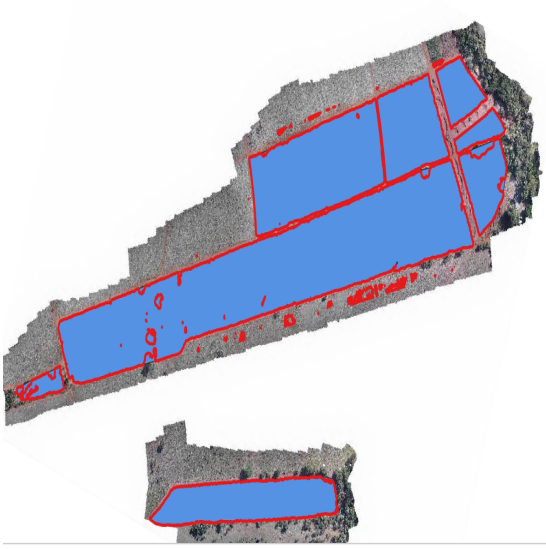
Unet 512 Crop-Focused



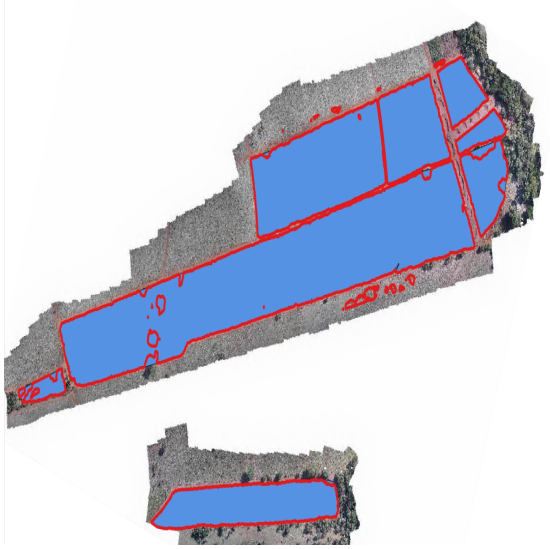
Deeplab 512-Crop-Focused



Segformer-512-Crop-Focused



InternImage-Crop-Focused



Assim, embora o InternImage ofereça o melhor desempenho em termos de acurácia, sua adoção precisa considerar o custo computacional associado. Em aplicações onde cada ponto percentual de IoU tem impacto operacional significativo — como mapeamentos de alta precisão, monitoramentos críticos ou estudos científicos — o investimento adicional pode ser plenamente justificado. Por outro lado, em cenários operacionais, embarcados, ou com restrições de tempo e hardware, o SegFormer representa uma alternativa mais equilibrada, oferecendo desempenho competitivo com menor complexidade e inferências mais rápidas. Dessa forma, o SegFormer tende a apresentar a melhor relação custo-benefício para a maioria dos usos práticos, enquanto o InternImage se torna a escolha ideal quando a prioridade é maximizar a precisão, independentemente do custo computacional.

Figura 4.4: Resultados Ilustrativos dos Modelos de Segmentação Para a Classe de Árvores em uma fazenda do conjunto de teste.

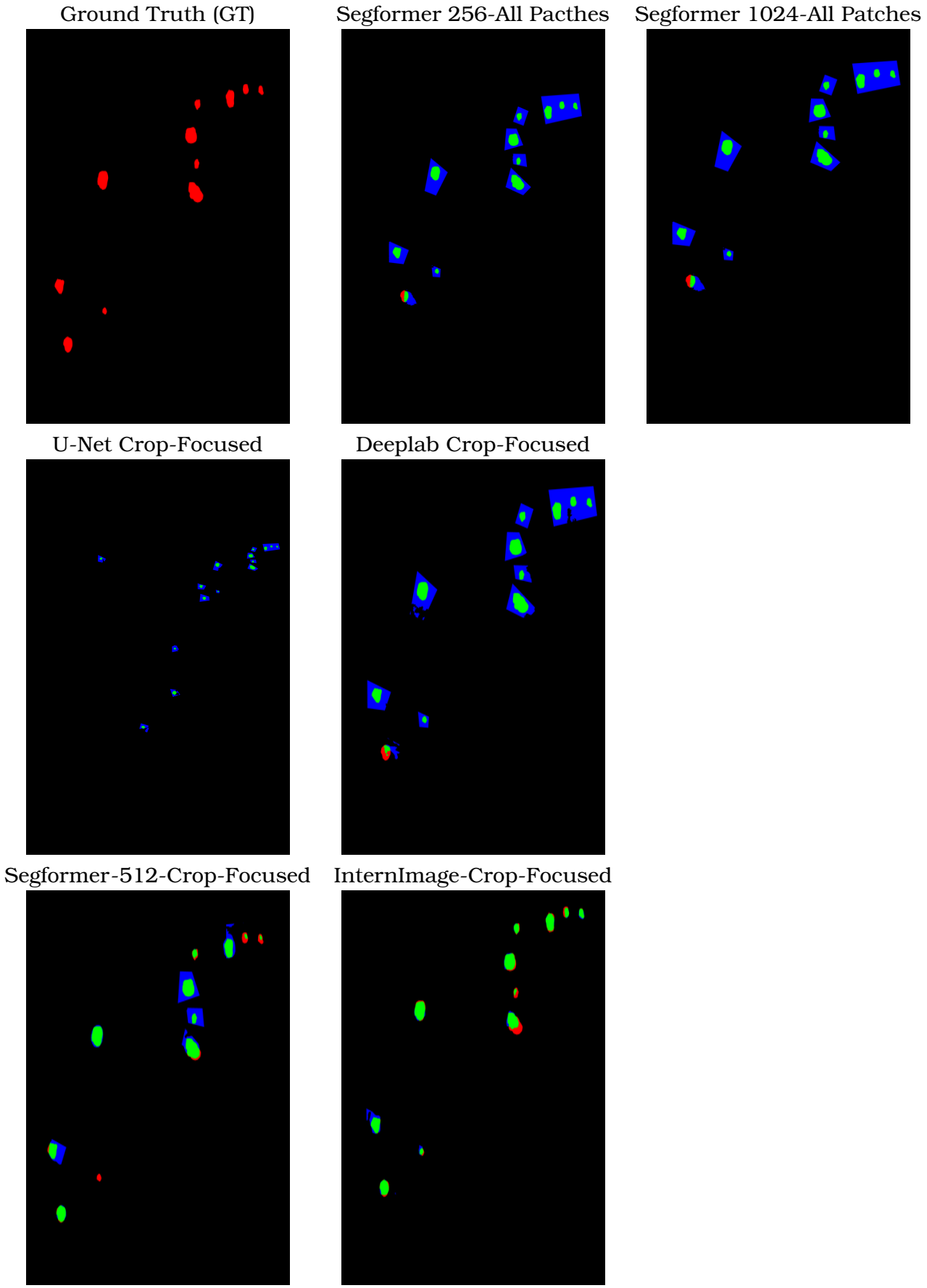
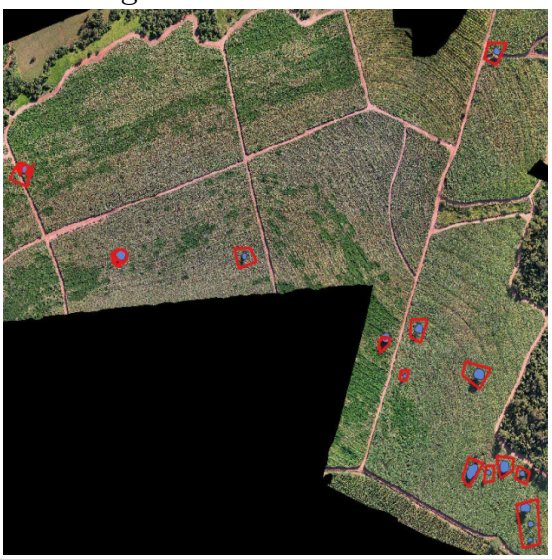


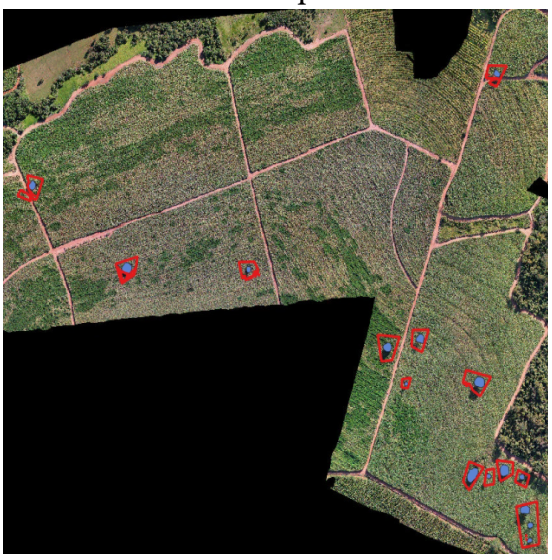
Figura 4.5: Resultados dos Modelos de Segmentação Para a Classe de Árvores em shapefile em uma outra fazenda do conjunto de teste.

Segformer 256-All Pacthes

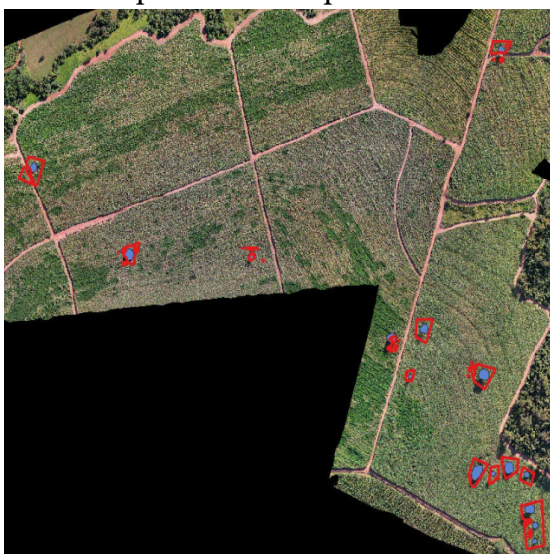


Segformer 1024-All Patches

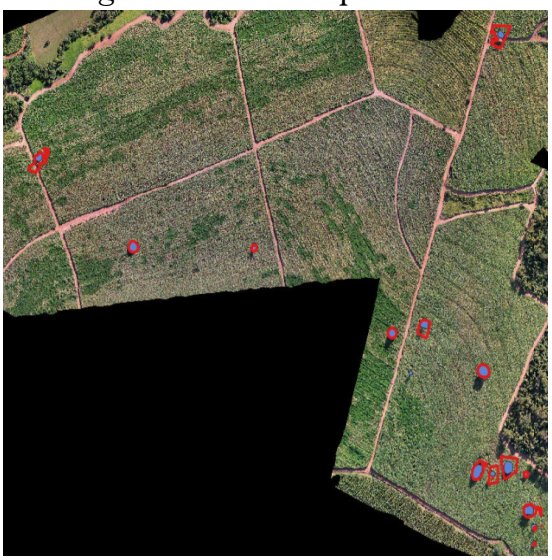
Unet 512 Crop-Focused



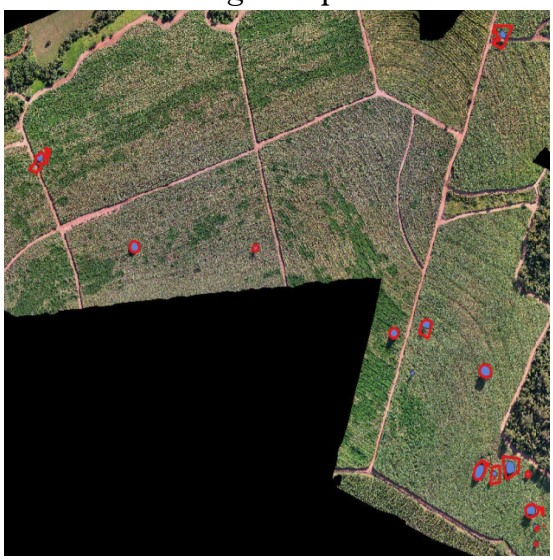
Deeplab 512-Crop-Focused



Segformer-512-Crop-Focused



InternImage-Crop-Focused



CAPÍTULO

5

Conclusões

Neste trabalho, exploramos e avaliamos metodologias de segmentação semântica para a identificação automatizada de áreas de plantio de cana-de-açúcar e árvores isoladas em ortofotos de alta resolução. A pesquisa foi guiada pelo objetivo de aprimorar o monitoramento agrícola, superando os desafios impostos por imagens de grandes dimensões, o desbalanceamento de classes e a necessidade de equilibrar o contexto global e os detalhes locais.

5.1 Resumo dos Objetivos e Principais Resultados

Nossa análise comparativa incluiu arquiteturas amplamenta utilizadas na literatura, como U-Net e DeepLabv3+, e modelos do estado da arte, como SegFormer e InternImage, representando a evolução de CNNs para Vision Transformers e convoluções deformáveis. Os resultados quantitativos demonstram que o tamanho de patch ideal é de 512x512 pixels, pois ele otimiza o equilíbrio entre o contexto espacial e a preservação de detalhes finos, maximizando o desempenho das métricas para ambas as classes.

Descobrimos também que as estratégias de amostragem direcionadas são cruciais para mitigar o problema de desbalanceamento de classes. A abordagem Crop-Focused superou significativamente a estratégia *All-Patches* ao priorizar patches que contêm a classe de interesse, o que foi especialmente evidente na segmentação de classes minoritárias, como as árvores e vimos que ao descartar muitos patches, como na estratégia *Boundary zone*, também perdemos informação relevante que diminuiu a eficácia do modelo.

Entre os modelos avaliados, o InternImage e o SegFormer consistentemente entregaram os melhores resultados, superando as arquiteturas mais antigas.

No entanto, ao considerarmos a eficiência computacional, o SegFormer se destaca por oferecer desempenho similar ao InternImage, mas com menor consumo de memória e tempo de inferência, tornando-o a escolha mais prática para aplicações reais no agronegócio.

5.2 Limitações

Apesar dos resultados promissores, o estudo enfrentou algumas limitações. A principal delas é a vulnerabilidade dos modelos a variações nos dados de entrada. Como observado em estudos anteriores, modelos de segmentação podem ter desempenho limitado em novos cenários que apresentam culturas visualmente semelhantes ou sob condições ambientais distintas. A generalização para diferentes culturas ou para áreas com características visuais variadas pode ser um desafio e necessitaria de reajustes ou de treinamento adicional. Além disso, a segmentação das bordas não é ideal, como ocorre na identificação de árvores isoladas.

5.3 Trabalhos Futuros

Com base nos resultados e nas limitações deste estudo, as seguintes direções de pesquisa são propostas:

Validação da Generalização: Expandir a pesquisa para outras culturas agrícolas além da cana-de-açúcar. Isso envolve a coleta de novos conjuntos de dados e a avaliação da capacidade do modelo de se adaptar a diferentes cenários sem a necessidade de um novo treinamento completo.

Melhoria na Segmentação de Bordas: Investigar a implementação do *F1-Boundary Score* para uma avaliação mais precisa da qualidade das bordas segmentadas. Esta métrica pode fornecer uma visão mais clara do desempenho dos modelos em áreas de transição, que são críticas para o mapeamento detalhado de talhões.

Adoção de Modelos Fundacionais: Avaliar o desempenho de modelos de fundação, como o Segment Anything Model (SAM), que podem operar em um regime de zero-shot ou few-shot, reduzindo a necessidade de grandes conjuntos de dados anotados. A viabilidade de tais modelos para o contexto agrícola precisa ser verificada, considerando a sensibilidade a prompts e a capacidade de gerar máscaras precisas.

Desenvolvimento de Arquitetura Multiclasse e Multiescala: Implementar uma rede siamesa robusta que combine uma rede principal para capturar o contexto de toda a imagem com uma rede auxiliar focada em recortes menores, ricos em detalhes. Essa abordagem pode permitir que o modelo realize a

segmentação de múltiplas classes de forma simultânea e mais eficiente, capturando tanto a visão geral dos talhões quanto os detalhes finos das árvores.

Disclaimer

Os autores utilizaram inteligência artificial generativa para auxiliar na melhoria da linguagem e legibilidade deste manuscrito. Os autores revisaram e editaram o texto conforme necessário e assumem total responsabilidade pelo conteúdo desta publicação.

Referências Bibliográficas

[1] T. Anand, S. Sinha, M. Mandal, V. Chamola, and F. R. Yu. Agrisegnet: Deep aerial semantic segmentation framework for iot-assisted precision agriculture. *IEEE Sensors Journal*, 21(16):17581–17590, 2021. Citado na página 7.

[2] A. Berka, Y. Es-saady, M. El Hajji, R. Canals, and A. Hafiane. Enhancing deeplabv3+ for aerial image semantic segmentation using weighted upsampling. pages 1–6, 05 2024. Citado na página 2.

[3] R. Bommasani, D. A. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx, M. S. Bernstein, J. Bohg, A. Bosselut, E. Brunskill, E. Brynjolfsson, S. Buch, D. Card, R. Castellon, N. S. Chatterji, A. S. Chen, K. A. Creel, J. Davis, D. Demszky, C. Donahue, M. Doumbouya, E. Durmus, S. Ermon, J. Etchemendy, K. Ethayarajh, L. Fei-Fei, C. Finn, T. Gale, L. E. Gillespie, K. Goel, N. D. Goodman, S. Grossman, N. Guha, T. Hashimoto, P. Henderson, J. Hewitt, D. E. Ho, J. Hong, K. Hsu, J. Huang, T. F. Icard, S. Jain, D. Jurafsky, P. Kalluri, S. Karamcheti, G. Keeling, F. Khani, O. Khattab, P. W. Koh, M. S. Krass, R. Krishna, R. Kuditipudi, A. Kumar, F. Ladhak, M. Lee, T. Lee, J. Leskovec, I. Levent, X. L. Li, X. Li, T. Ma, A. Malik, C. D. Manning, S. P. Mirchandani, E. Mitchell, Z. Munyikwa, S. Nair, A. Narayan, D. Narayanan, B. Newman, A. Nie, J. C. Niebles, H. Nilforoshan, J. F. Nyarko, G. Ogut, L. Orr, I. Papadimitriou, J. S. Park, C. Piech, E. Portelance, C. Potts, A. Raghunathan, R. Reich, H. Ren, F. Rong, Y. H. Roohani, C. Ruiz, J. Ryan, C. R'e, D. Sadigh, S. Sagawa, K. Santhanam, A. Shih, K. P. Srinivasan, A. Tamkin, R. Taori, A. W. Thomas, F. Tramèr, R. E. Wang, W. Wang, B. Wu, J. Wu, Y. Wu, S. M. Xie, M. Yasunaga, J. You, M. A. Zaharia, M. Zhang, T. Zhang, X. Zhang, Y. Zhang, L. Zheng, K. Zhou, and P. Liang. On the opportunities and risks of foundation models. *ArXiv*, 2021. Citado na página 5.

[4] W. Boonpook, Y. Tan, A. Nardkulpat, K. Torsri, P. Torteeka, P. Kamsing,

U. Sawangwit, J. Pena, and M. Jainaen. Deep learning semantic segmentation for land use and land cover types using landsat 8 imagery. *ISPRS International Journal of Geo-Information*, 12(1), 2023. Citado na página 2.

[5] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam. Rethinking atrous convolution for semantic image segmentation. *ArXiv*, abs/1706.05587, 2017. Citado na página 6.

[6] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *ECCV*, 2018. Citado na página 6.

[7] C.-I. Cira, M.- Manso-Callejo, R. Alcarria, T. Iturrioz, and J.-J. Arranz-Justel. Insights into the effects of tile size and tile overlap levels on semantic segmentation models trained for road surface area extraction from aerial orthophotography. *Remote Sensing*, 16(16), 2024. Citado na página 17.

[8] I. Colomina and P. Molina. Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 92:79–97, 2014. Citado na página 1.

[9] M. Contributors. MMSegmentation: Openmmlab semantic segmentation toolbox and benchmark. https://github.com/open-mmlab/mmsegmentation, 2020. Citado na página 22.

[10] G. Csurka, R. Volpi, and B. Chidlovskii. Semantic image segmentation: Two decades of research. *Found. Trends Comput. Graph. Vis.*, 14:1–162, 2023. Citado na página 5.

[11] M. da Agricultura e Pecuária. Fotografia do setor sucroenergético no brasil e os benefícios econômicos, ambientais e sociais gerados, 2024. Citado na página 1.

[12] U. da Indústria de Cana-de-Açúcar e Bioenergia (Unica). Safra 2023/2024 termina como a maior da história, 2024. Citado na página 1.

[13] E. E. de Miranda and M. F. Fonseca. Chapter 4 - sugarcane: food production, energy, and environment. In F. Santos, S. C. Rabelo, M. De Matos, and P. Eichler, editors, *Sugarcane Biorefinery, Technology and Perspectives*, pages 67–88. Academic Press, 2020. Citado na página 1.

[14] S. Dhariwal and A. Sharma. Aerial images were used to detect curved-crop rows and failures in sugarcane production. In *2022 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, pages 1–6. IEEE, 2022. Citado na página 8.

[15] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*. ICLR, 2021. Citado na página 5.

[16] F. Fahmi, Sawaluddin, U. Andayani, and B. Siregar. Sensing of vegetation status using ortophotos image generated with uav. *ABDIMAS TALENTA: Jurnal Pengabdian Kepada Masyarakat*, 2018. Citado na página 1.

[17] A. d. S. Ferreira, D. M. Freitas, G. G. d. Silva, H. Pistori, and M. T. Folhes. Weed detection in soybean crops using convnets. *Computers and Electronics in Agriculture*, 143:314–324, 2017. Citado na página 7.

[18] J. C. Franchini, A. A. Balbinot Junior, L. A. d. C. Jorge, H. Debiasi, W. P. Dias, C. V. Godoy, A. d. Oliveira Junior, F. B. Correa, and M. C. N. d. Oliveira. Uso de imagens aéreas obtidas com drones em sistemas de produção de soja. Technical report, Empresa Brasileira de Pesquisa Agropecuária - Embrapa, 2018. Citado na página 1.

[19] N. Ganatra and A. Patel. Deep learning methods and applications for precision agriculture. 09 2020. Citado na página 10.

[20] B. Huang, D. Reichman, L. M. Collins, K. Bradbury, and J. M. Malof. Tiling and stitching segmentation output for remote sensing: Basic challenges and recommendations. *CoRR*, abs/1805.12219, 2018. Citado na página 17.

[21] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y. Chen, and J. Wu. Unet 3+: A full-scale connected unet for medical image segmentation. *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1055–1059, 2020. Citado na página 6.

[22] E. S. J. Long and T. Darrell. Fully convolutional networks for semantic segmentation. In *The IEEE Conference on Computer Vision and Pattern Recognition*, pages 3431–3440. IEEE, 2015. Citado na página 6.

[23] S. Jeong, J. Ko, J. oh Ban, T. Shin, and J. min Yeom. Deep learning-enhanced remote sensing-integrated crop modeling for rice yield prediction. *Ecological Informatics*, 84:102886, 2024. Citado na página 7.

[24] J. Jia, J. Song, Q. Kong, H. Yang, Y. Teng, and X. Song. Multi-attention-based semantic segmentation network for land cover remote sensing images. *Electronics*, 12(6):1347, 2023. Citado na página 5.

[25] K. Jin, J. Zhang, Z. Wang, J. Zhang, N. Liu, M. Li, and Z. Ma. Application of deep learning based on thermal images to identify the water stress in cotton under film-mulched drip irrigation. *Agricultural Water Management*, 299:108901, 2024. Citado na página 7.

[26] A. King. Technology: The future of agriculture. *Nature*, 544:S21–S23, 2017. Citado na página 1.

[27] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár, and R. Girshick. Segment anything. *arXiv:2304.02643*, 2023. Citado na página 10.

[28] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. Burges, L. Bottou, and K. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25, pages 1097 – 1105. Curran Associates, Inc., 2012. Citado na página 5.

[29] X. Li, M. Cai, X. Tan, C. Yin, W. Chen, Z. Liu, J. Wen, and Y. Han. An efficient transformer network for detecting multi-scale chicken in complex free-range farming environments via improved rt-detr. *Computers and Electronics in Agriculture*, 224:109160, 2024. Citado na página 10.

[30] X. Li, H. Ding, H. Yuan, W. Zhang, J. Pang, G. Cheng, K. Chen, Z. Liu, and C. C. Loy. Transformer-Based Visual Segmentation: A Survey . *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 46(12):10138–10163, 2024. Citado na página 5.

[31] Z. Luo, W. Yang, Y. Yuan, R. Gou, and X. Li. Semantic segmentation of agricultural images: A survey. *Information Processing in Agriculture*, 11(2):172–186, 2024. Citado na página 7.

[32] S. Mehta and M. Rastegari. Mobilevit: Light-weight, general-purpose, and mobile-friendly vision transformer. In *International Conference on Learning Representations*, 2022. Citado na página 10.

[33] M. Narimani, A. Pourreza, A. Moghimi, M. Mesgaran, P. Farajpoor, and H. Jafarbiglu. Drone-based multispectral imaging and deep learning for timely detection of branched broomrape in tomato farms. 13053:1305304, 2024. Citado na página 7.

[34] B. Neupane, T. Horanont, and J. Aryal. Deep learning-based semantic segmentation of urban features in satellite images: A review and meta-analysis. *Remote Sensing*, 13(4), 2021. Citado na página 17.

[35] O. Oktay, J. Schlemper, L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Hammerla, B. Kainz, B. Glocker, and D. Rueckert. Attention u-net: Learning where to look for the pancreas. 04 2018. Citado na página 6.

[36] L. P. Osco, J. Marcato Junior, A. P. Marques Ramos, L. A. de Castro Jorge, S. N. Fatholahi, J. de Andrade Silva, E. T. Matsubara, H. Pistori, W. N. Gonçalves, and J. Li. A review on deep learning in uav remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 102:102456, 2021. Citado na página 6.

[37] P. C. Pereira Júnior, A. Monteiro, R. d. L. Ribeiro, A. C. Sobieranski, and A. von Wangenheim. Comparison of classical computer vision vs. convolutional neural networks for weed mapping in aerial images. *Revista de Informática Teórica e Aplicada*, 27(4):20–33, 2020. Citado na página 7.

[38] L. Qi, D. Zuo, Y. Wang, Y. Tao, R. Tang, J. Shi, J. Gong, and B. Li. Convolutional neural network-based method for agriculture plot segmentation in remote sensing images. *Remote. Sens.*, 16:346, 2024. Citado na página 2.

[39] M. T. R, A. S. Chandel, H. Gupta, R. Jain, and S. Garg. A novel approach for detection of agricultural crops using cnn. In *Proceedings of the 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pages 1–8, 2023. Citado na página 7.

[40] J. B. Ribeiro, R. R. da Silva, J. D. Dias, M. C. Escarpinati, and A. R. Backes. Automated detection of sugarcane crop lines from uav images using deep learning. *Information Processing in Agriculture*, 2023. Citado nas páginas 2 e 7.

[41] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241. Springer International Publishing, 2015. Citado na página 6.

[42] G. Samseemoung, P. Soni, H. Jayasuriya, and V. Salokhe. Application of low altitude remote sensing (lars) platform for monitoring crop growth and weed infestation in a soybean plantation. *Precision Agriculture*, 13, 12 2012. Citado na página 1.

[43] A. Sayed, V. Kumbhar, S. Jadhav, and S. K. Shah. Agricultural field boundary delineation using deep learning techniques. In *2023 Interna-*

tional Conference on Emerging Smart Computing and Informatics (ESCI), pages 1–6. IEEE, 2023. Citado na página 7.

[44] J. Schlemper, O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, and D. Rueckert. Attention gated networks: Learning to leverage salient regions in medical images. *Medical image analysis*, 53:197–207, 2019. Citado na página 6.

[45] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni. U-net and its variants for medical image segmentation: A review of theory and applications. *IEEE Access*, 9:82031–82057, 2021. Citado na página 6.

[46] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. pages 1–14. Computational and Biological Learning Society, 2015. Citado na página 5.

[47] Z. Su, K. Chen, and M. Liu. Farmland parcel extraction and area calculation from uav images based on semantic segmentation. *Remote Sensing Applications: Society and Environment*, 40:101734, 2025. Citado na página 2.

[48] C. Sun, M. Zhang, M. Zhou, and X. Zhou. An improved transformer network with multi-scale convolution for weed identification in sugarcane field. *IEEE Access*, 12:31168–31181, 2024. Citado nas páginas 7 e 8.

[49] F. Tong and Y. Zhang. Individual tree crown delineation in high resolution aerial rgb imagery using stardist-based model. *Remote Sensing of Environment*, 319:114618, 2025. Citado na página 2.

[50] W. Wang, J. Dai, Z. Chen, Z. Huang, Z. Li, X. Zhu, X. Hu, T. Lu, L. Lu, H. Li, X. Wang, and Y. Qiao. InternImage: Exploring Large-Scale Vision Foundation Models with Deformable Convolutions . In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14408–14419, Los Alamitos, CA, USA, June 2023. IEEE Computer Society. Citado na página 3.

[51] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo. Segformer: Simple and efficient design for semantic segmentation with transformers. In *Neural Information Processing Systems (NeurIPS)*, 2021. Citado na página 3.

[52] W. Xie, M. Zhao, Y. Liu, D. Yang, K. Huang, C. Fan, and Z. Wang. Recent advances in transformer technology for agriculture: A comprehensive survey. *Engineering Applications of Artificial Intelligence*, 138:109412, 2024. Citado na página 10.

[53] H. Yan, G. Liu, Z. Li, Z. Li, and J. He. Sceca u-net crop classification for uav remote sensing image. *Cluster Computing*, 28, 10 2024. Citado na página 2.

[54] J. Yu, L. Lei, and Z. Li. Individual tree segmentation based on seed points detected by an adaptive crown shaped algorithm using uav-lidar data. *Remote Sensing*, 16(5):825, 2024. Citado nas páginas 7 e 8.

[55] H. Zhao and C. Zhang. Applications of remote sensing in agricultural soil and crop mapping. *Agriculture*, 15(21), 2025. Citado na página 2.

[56] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 3–11. Springer International Publishing, 2018. Citado na página 6.