

Detecção de Copas de Árvores Urbanas Utilizando YOLOv11: Uma Análise Comparativa e Aplicações em Sensoriamento Remoto

Ana Luiza Campos Souza
Universidade de Brasília
Brasília, Brasil
211055441@aluno.unb.br

Célio Júnio de Freitas Eduardo
Universidade de Brasília
Brasília, Brasil
211010350@aluno.unb.br

Davi de Araujo Garcez Bueno
Universidade de Brasília
Brasília, Brasil
211060586@aluno.unb.br

Abstract—Este trabalho apresenta uma investigação sobre o desempenho do modelo YOLOv11 na tarefa de detecção individual de copas de árvores em imagens RGB de alta resolução, com foco em ambientes urbanos. O estudo descreve a metodologia de treinamento e avaliação do YOLOv11, destacando sua eficácia em termos de precisão (AP_{50}), complexidade computacional (parâmetros e FLOPs) e velocidade de inferência. Os resultados obtidos são comparados rigorosamente com o estado da arte, referenciando métodos avaliados no trabalho de Zamboni et al. (2021), demonstrando que o YOLOv11 atinge um desempenho competitivo com modelos de ponta, mantendo alta eficiência. Abordamos também as possibilidades futuras para a extensão desta pesquisa, incluindo a exploração de cenários desafiadores e a integração com outras fontes de dados.

Index Terms—Detecção de Objetos, Aprendizado Profundo, Visão Computacional, YOLOv11, Sensoriamento Remoto, Copas de Árvores Urbanas.

I. INTRODUÇÃO

A detecção individual de copas de árvores a partir de imagens de sensoriamento remoto de alta resolução é uma tarefa fundamental, fornecendo dados cruciais para inventários florestais, planejamento urbano e análise ambiental [1].

Este trabalho explora a aplicação do YOLOv11 (You Only Look Once v11), uma das arquiteturas mais recentes na família de detectores de objetos de estágio único, para a detecção de copas de árvores urbanas. Os modelos YOLO são reconhecidos por seu notável equilíbrio entre alta precisão e velocidade de inferência. O objetivo principal é avaliar o desempenho do YOLOv11 neste domínio específico e compará-lo com os resultados do estudo referência de Zamboni.

Este relatório descreve a metodologia empregada, os resultados quantitativos obtidos e posiciona o YOLOv11 em relação a outros detectores, conforme os benchmarks existentes.

II. TRABALHOS RELACIONADOS

A evolução da detecção de objetos em imagens de sensoriamento remoto tem sido drasticamente acelerada pelas técnicas de Aprendizado Profundo. As Redes Neurais Convolucionais (CNNs) tem permitido a extração automática de características hierárquicas e a construção de modelos mais adaptáveis.

Um estudo importante, e base para a comparação aqui, é o de Zamboni [1]. Nele, foi realizado uma avaliação exaustiva

de 21 métodos de Aprendizado Profundo para a detecção individual de copas de árvores urbanas em imagens RGB de alta resolução. Os métodos foram categorizados como baseados em âncoras (de um ou dois estágios) e sem âncoras (anchor-free), e seu desempenho foi minuciosamente analisado em termos de Average Precision (AP_{50}). Este estudo fornece um *benchmark* valioso, identificando modelos de ponta como *Double Heads*, *ATSS*, *CARAFE*, *FSAF* e *FoveaBox*, e destacando desafios persistentes como a detecção em áreas de alta densidade e a necessidade de conjuntos de dados mais diversos para generalização.

A família de modelos You Only Look Once (YOLO) [2] tem se consolidado como uma das soluções mais influentes para detecção de objetos em tempo real. Os modelos YOLO priorizam a velocidade de inferência sem comprometer significativamente a precisão, tornando-os ideais para aplicações que demandam processamento rápido. O YOLOv11 da Ultralytics [2], uma arquitetura recente de alto desempenho, incorpora otimizações arquiteturais e melhorias nos algoritmos de treinamento, sendo inserida em um framework de desenvolvimento ativo, robusto e com amplo suporte da comunidade.

III. METODOLOGIA

A. Justificativa para a Escolha do Modelo

O escopo original deste projeto, conforme definido nas especificações, previa a implementação e avaliação do modelo YOLO-MS. A equipe iniciou os trabalhos tentando configurar o ambiente de desenvolvimento utilizando o código-fonte disponibilizado publicamente.

No entanto, foram encontrados desafios técnicos significativos relacionados a dependências de bibliotecas descontinuídas e conflitos de versão, que impediram a replicação do ambiente e a execução do treinamento de forma estável e reproduzível. Diante desses obstáculos, e para garantir a viabilidade do projeto dentro do cronograma, tomou-se a decisão estratégica de substituir o modelo.

A escolha recaiu sobre o modelo YOLOv11 da Ultralytics [2], já apresentado na seção anterior, devido à sua robustez e suporte ativo. Essa escolha, embora represente um desvio do plano inicial, mantém o objetivo central do projeto: avaliar

um detector de objetos moderno e eficiente, comparando seu desempenho com o benchmark estabelecido por Zamboni.

B. Conjunto de Dados

Visando garantir a reprodutibilidade e a comparabilidade dos resultados com o estudo base [1], a metodologia foi estruturada da seguinte forma. Para assegurar a comparabilidade, foi utilizado o mesmo conjunto de dados de imagens RGB de alta resolução do trabalho de referência. O *dataset* é publicamente acessível via GitHub [3] e é composto por imagens com *bounding boxes* para cada árvore individualmente identificada. A uniformidade do conjunto de dados é fundamental para uma análise comparativa justa das métricas de desempenho.

C. Configuração e Treinamento do YOLOv11

O modelo de detecção de objetos empregado foi o **YOLOv11s** (versão *small*). O treinamento foi realizado utilizando a biblioteca Ultralytics YOLO (versão 8.3.162) em um ambiente com aceleração por GPU (Tesla T4). A escolha da versão ‘s’ (*small*) se baseou em seu conhecido *trade-off* favorável entre precisão e eficiência computacional. As principais configurações de treinamento incluem:

- **Épocas de Treinamento:** 40 épocas.
- **Resolução das Imagens (*imgsz*):** 512x512 pixels.
- **Tamanho do Lote (*Batch Size*)** 16.
- **Otimizador:** AdamW.
- **Modelo Base:** *yolov11s.pt* (modelo pré-treinado na base COCO).

Parâmetros adicionais, como taxas de aprendizado e estratégias de aumento de dados, seguiram as configurações padrão recomendadas pela Ultralytics para otimizar a performance.

D. Métricas de Desempenho e Coleta de Dados

Para a avaliação do modelo e a comparação com a literatura, as seguintes métricas foram coletadas e analisadas:

- **Average Precision (AP_{50}):** Métrica principal para medir a precisão da detecção, conforme o critério de avaliação de Zamboni [1].
- **Average Precision (AP_{50-95}):** Média da AP em múltiplos limiares de IoU (0.50 a 0.95), para uma visão mais robusta da detecção.
- **Número de Parâmetros (Params):** Quantifica a complexidade do modelo.
- **FLOPs:** Mede o custo computacional do modelo.
- **Velocidade de Inferência (ms/imagem):** Tempo médio para processar uma imagem.

Os valores foram extraídos dos logs e do sumário do modelo gerados pela biblioteca Ultralytics.

IV. RESULTADOS

Apresentaremos os resultados quantitativos e qualitativos obtidos pelo modelo YOLOv11s, comparando seu desempenho com os modelos do artigo de Zamboni e analisando sua eficiência.

A. Desempenho de Precisão (AP_{50})

O treinamento do YOLOv11s por 40 épocas resultou em um desempenho de validação de AP_{50} que atingiu o pico de **0.73009**. Este valor representa a capacidade do modelo em detectar e localizar as copas das árvores com alta acurácia no conjunto de validação. A Tabela I compara este resultado com os métodos destacados por Zamboni [1].

Analisando a Tabela I, vemos que o YOLOv11s alcançou um nível de precisão notável. Seu AP_{50} supera as médias de desempenho de todos os grupos de detectores avaliados por Zamboni e se posiciona em um patamar altamente competitivo com os melhores modelos daquele estudo, como o Double Heads (0.732). Essa competitividade pode ser atribuída a fatores como o design eficiente de sua rede de extração de características (backbone) e o uso da *Distribution Focal Loss* (DFL). A DFL aprimora a regressão das caixas delimitadoras, permitindo uma localização mais precisa dos objetos, o que é fundamental em cenários com alvos pequenos e aglomerados como copas de árvores.

TABLE I
COMPARAÇÃO DO AP_{50} DO YOLOv11s COM MÉTODOS DE ZAMBONI ET AL. (2021)

Método/Grupo	Tipo de Detector	AP_{50}
YOLOv11s (Este Trabalho)	Anchor-Free / Single-Stage	0.73009
Zamboni et al. (2021) - Médias dos Grupos		
AB-OS Média	Anchor-Based (Um Estágio)	0.657 ± 0.032
AB-TS/MS Média	Anchor-Based (Dois/Multi-Estágios)	0.669 ± 0.023
AF Média	Anchor-Free	0.686 ± 0.014
Zamboni et al. (2021) - Melhores Métodos		
Double Heads	AB-TS	0.732
ATSS	AF	0.728
CARAFE	AB-TS	0.724
FSAF	AF	0.720
FoveaBox	AF	0.719

B. Complexidade e Eficiência Computacional

O modelo YOLOv11s demonstrou ser altamente eficiente, com 9.41M de parâmetros, 21.3 GFLOPs e uma velocidade de inferência de 3.4ms por imagem. A Tabela II contextualiza esses números. A combinação de alta precisão com baixa complexidade e alta velocidade é o aspecto mais notável do modelo, qualificando-o como uma ferramenta extremamente eficaz para aplicações práticas.

TABLE II
COMPARAÇÃO DE COMPLEXIDADE E VELOCIDADE DE INFERÊNCIA

Modelo	AP_{50} (COCO)	Parâmetros (M)	FLOPs (G)	Velocidade (ms/img)
YOLOv11s (Este Trabalho)	0.73009	9.41	21.3	3.4
RTMDet-tiny	0.505	4.9	8.1	7.2
YOLOv7-tiny	0.506	6.2	6.9	5.2
YOLOv5s	0.507	7.2	8.3	5.4
YOLO-MS-S	0.518	8.13	15.58	—
RTMDet-s	0.528	8.9	14.8	—
YOLOv6s	0.535	17.2	21.9	8.9

C. Desafios Remanescentes e Análise Qualitativa

Apesar do desempenho quantitativo de destaque, o YOLOv11s enfrenta os mesmos desafios de outros detec-

tores baseados em *bounding boxes*: a dificuldade em individualizar copas de árvores em ***alta densidade ou com sobreposição significativa***. Isso leva a fusões de instâncias, omissões de árvores menores e localização imprecisa. Outros desafios persistentes incluem a grande variação de escala das árvores, oclusões por estruturas urbanas e condições ambientais variáveis.

DISPONIBILIDADE DO CÓDIGO

O código-fonte do projeto, os scripts de treinamento, avaliação e os resultados detalhados estão disponíveis publicamente no repositório do GitHub [4].



Fig. 1. Mosaico de imagens do conjunto de treinamento. As anotações (caixas azuis) ilustram a alta densidade e sobreposição das copas, um dos principais desafios para a detecção individual.

V. TRABALHOS FUTUROS

A partir dos resultados e desafios identificados, diversas linhas de pesquisa podem ser exploradas:

- **Da Detecção para a Segmentação de Instâncias:** Evoluir para modelos como Mask R-CNN ou versões de segmentação do YOLO para delimitar a área exata de cada copa, possibilitando análises mais ricas, como estimativa de biomassa.
- **Integração de Dados Multimodais:** Fundir dados RGB com LiDAR (para informações de altura que separam copas sobrepostas) ou imagens **hiperespectrais** (para diferenciar espécies) para superar oclusões e agregar valor ecológico.
- **Análise Temporal e Monitoramento Contínuo:** Aplicar o modelo em imagens de diferentes anos para monitorar o crescimento da floresta urbana, desmatamento ou o sucesso de políticas de plantio.
- **Otimização para Aplicações em Tempo Real:** Otimizar o modelo com técnicas como quantização



Fig. 2. Visualização das predições do modelo em um lote de validação. Cada detecção exibe a caixa delimitadora, a classe ('tree') e a confiança, evidenciando a capacidade do modelo em identificar múltiplas instâncias.

para implantação em drones, permitindo a contagem de árvores em tempo real durante o voo.

VI. CONCLUSÃO

Este trabalho avaliou a eficácia do modelo YOLOv11s na tarefa de detecção individual de copas de árvores em imagens aéreas, demonstrando resultados expressivos. O treinamento e a validação, utilizando o conjunto de dados de referência de Zamboni et al. [1], resultaram em uma precisão média (AP_{50}) de **0.73009**. Este valor, que não apenas supera com folga as médias de desempenho dos detectores avaliados no trabalho base, mas também se mostra altamente competitivo com métodos de ponta mais complexos como o *Double Heads*, aliado a uma altíssima eficiência computacional (3.4ms/imagem), valida o uso de arquiteturas de estágio único e eficientes para aplicações de sensoriamento remoto em larga escala.

Apesar do sucesso, o estudo reafirmou que o principal desafio metodológico reside na individualização de árvores em áreas de alta densidade, uma limitação inerente às caixas delimitadoras. Conclui-se, portanto, que o YOLOv11 é uma ferramenta poderosa e viável para inventários florestais urbanos automatizados. As futuras pesquisas devem se concentrar em superar as limitações atuais através da segmentação de instâncias e da fusão de dados multimodais, visando um mapeamento ainda mais preciso e detalhado da vegetação urbana.

REFERENCES

- [1] P. Zamboni, F. A. F. Junior, J. E. M. S. Vidal, D. B. Lobell, and R. d. S. Torres, "Individual tree crown detection in urban environments using high-resolution RGB imagery: A deep learning-based comparison," *Remote Sensing*, vol. 13, no. 13, p. 2482, 2021.
- [2] G. Jocher, A. Chaurasia, and J. Qiu, "YOLO by Ultralytics," <https://github.com/ultralytics/ultralytics>, 2023.

- [3] P. Zamboni, "Individual urban tree crown detection dataset," https://github.com/pedrozamboni/individual_urban_tree_crown_detection, 2021.
- [4] A. L. C. Souza, C. J. F. Eduardo, and D. A. G. Bueno, "YOLOV11-Detector-de-Arvores: Código-fonte do projeto," GitHub repository, 2024. [Online]. Available: <https://github.com/HomiRuim/YOLOV11-Detector-de-Arvores/tree/main>