

Detecção e Análise de Tamanho de Copas de Árvores Urbanas com YOLOv11 e K-Means: Uma Ferramenta para Planejamento Urbano

Breno Braga Galvão
breno.braga.galvao@gmail.com

Abstract

Urban vegetation is essential for improving air quality, regulating temperature, and maintaining ecological balance in cities. However, monitoring and managing urban greenery remain challenging, requiring precise and scalable methods for analysis. This study proposes an automated pipeline for detecting, analyzing, and classifying urban tree crowns using aerial imagery. The pipeline integrates YOLOv11, a state-of-the-art object detection algorithm, with K-Means clustering to categorize tree crowns into small, medium, and large sizes. Crown areas are estimated from bounding boxes generated by the model, providing detailed insights into the structural characteristics of urban vegetation.

The methodology includes image preprocessing, training YOLOv11 with manually annotated data, tree segmentation, and crown size classification using K-Means. The pipeline's performance is evaluated using precision, recall, F1-score, and mean Average Precision (mAP). Results demonstrate the effectiveness of YOLOv11 in detecting urban trees, achieving a precision of 0.73 and a recall of 0.76, while K-Means successfully categorizes crown sizes, revealing patterns in vegetation distribution.

This work highlights the potential of machine learning for urban vegetation analysis, offering a robust tool for environmental monitoring and urban planning. By enabling data-driven decision-making, the proposed pipeline contributes to sustainable city development and the implementation of accessible technologies for large-scale environmental management.

1. Introdução

A vegetação urbana é fundamental para a qualidade de vida em cidades, contribuindo para a melhoria do ar, a regulação térmica e o equilíbrio ecológico. No entanto, o monitoramento e a gestão dessas áreas verdes representam desafios complexos, exigindo métodos precisos e escaláveis para análise. Nesse contexto, o uso de imagens aéreas aliado a técnicas de aprendizado de máquina tem se

destacado como uma abordagem eficiente para a detecção e classificação de elementos vegetais.

Este estudo tem como objetivo principal desenvolver e avaliar um pipeline automatizado para a detecção, análise e classificação de copas de árvores urbanas em imagens aéreas. O pipeline combina o YOLOv11, um algoritmo de última geração para detecção de objetos, com o método de agrupamento K-Means, utilizado para classificar as copas das árvores em categorias de tamanho (pequena, média e grande). Além disso, é realizada a estimativa e a análise da área das copas a partir das caixas delimitadoras geradas pelo modelo, permitindo um detalhamento mais rico sobre as características estruturais da vegetação urbana.

A metodologia proposta inclui o pré-processamento das imagens, o treinamento do modelo YOLOv11 com anotações manuais, a segmentação das árvores e a classificação das copas em categorias de tamanho utilizando o algoritmo K-Means. A análise das áreas das copas permite compreender a variação estrutural da vegetação urbana e sua distribuição espacial, enquanto a validação do pipeline destaca seu desempenho em comparação com estudos semelhantes.

Os resultados evidenciam a eficiência do pipeline na identificação, categorização e análise das copas de árvores urbanas. A aplicação dessa metodologia é promissora para o monitoramento ambiental e o planejamento urbano, permitindo a tomada de decisões baseadas em dados precisos e objetivos. Além disso, o estudo contribui para a implementação de tecnologias acessíveis em projetos de gestão ambiental em larga escala, demonstrando o potencial de técnicas de aprendizado de máquina para a análise de vegetação urbana e abrindo caminho para aplicações inovadoras em cidades inteligentes e sustentáveis.

Os detalhes completos dos resultados podem ser acessados no repositório do GitHub: <https://github.com/BrenoBragaGalvao/TreeCrown-Detection-YOLO-KMeans.git>

2. Método

Este estudo teve como objetivo desenvolver um modelo de detecção de copas de árvores urbanas utilizando técnicas de aprendizado profundo, especificamente a arquitetura YOLOv11. O método foi dividido em etapas principais: coleta e preparação dos dados, treinamento do modelo, avaliação dos resultados e análise dos clusters de tamanho das copas das árvores. A seguir, descrevemos cada etapa em detalhes.

2.1 Coleta e Preparação dos Dados:

O conjunto de dados utilizado neste estudo foi obtido a partir de um repositório público no GitHub, contendo imagens de árvores urbanas e suas respectivas anotações de caixas delimitadoras (*bounding boxes*). As anotações foram fornecidas em arquivos de texto, onde cada linha representava uma caixa delimitadora com as coordenadas normalizadas (x_{min} , x_{max} , y_{min} , y_{max}) associadas a uma imagem específica.

- **Conversão para DataFrame:** Os arquivos de texto foram convertidos em um DataFrame estruturado, contendo as colunas: *image_filename* (nome da imagem), *xmin*, *xmax*, *ymin*, *ymax* (coordenadas das caixas delimitadoras). O DataFrame final continha 3.382 caixas delimitadoras associadas a 220 imagens.
- **Divisão dos Dados:** O conjunto de dados foi dividido em conjuntos de treinamento e validação, com 90% das imagens (198 imagens e 3.080 caixas delimitadoras) destinadas ao treinamento e 10% (22 imagens e 302 caixas delimitadoras) reservadas para validação.
- **Formatação dos Rótulos:** Para compatibilidade com o modelo YOLO, as coordenadas das caixas delimitadoras foram convertidas para o formato YOLO (x_{center} , y_{center} , *width*, *height*), normalizadas em relação ao tamanho da imagem. As coordenadas no formato YOLO são calculadas da seguinte forma:

$$x_{center} = \frac{x_{min} + x_{max}}{2 \times largura_{imagem}}$$

$$y_{center} = \frac{y_{min} + y_{max}}{2 \times altura_{imagem}}$$

$$width = \frac{x_{max} - x_{min}}{largura_{imagem}}$$

$$height = \frac{y_{max} - y_{min}}{altura_{imagem}}$$

Uma coluna adicional *class* foi adicionada ao DataFrame, contendo o valor 0 para todas as caixas delimitadoras, indicando a classe única "árvore".

2.2 Treinamento do Modelo:

O modelo YOLOv11 foi escolhido para a detecção de copas de árvores urbanas. O treinamento foi realizado utilizando a biblioteca Ultralytics, com as seguintes configurações:

- **Arquivo de Configuração YAML:** Um arquivo YAML foi criado para especificar os caminhos das imagens de treinamento e validação, o número de classes (1) e o nome da classe ("Tree").
- **Hiperparâmetros:** O modelo foi treinado por 200 épocas, com parada antecipada (early stopping) configurada para evitar overfitting. O tamanho das imagens foi fixado em 512x512 pixels.
- **Processo de Treinamento:** O treinamento foi realizado em um ambiente com GPU, e o melhor modelo foi salvo com base nas métricas de validação, incluindo precisão (precision), revocação (recall) e mAP (mean Average Precision).

2.3 Análise de Clusters de Tamanho das Copas das Árvores:

Para analisar a distribuição dos tamanhos das copas das árvores detectadas, foi aplicado o algoritmo K-Means na coluna *area* (calculada como o produto de x_{extend} e y_{extend} das caixas delimitadoras). O número de clusters foi definido como 3, representando categorias de tamanho: pequeno, médio e grande.

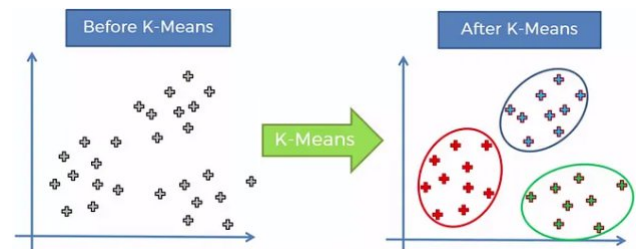


Figura 1: K-Means.

- **Aplicação do K-Means:** O algoritmo K-Means foi aplicado às áreas das caixas delimitadoras, e os resultados foram armazenados em uma nova coluna cluster no DataFrame. O objetivo do K-Means é minimizar a soma dos quadrados das distâncias entre os pontos e os centróides dos clusters:

$$\text{Minimizar} \sum_{i=1}^n \sum_{j=1}^k \|x_i - \mu_j\|^2$$

onde x_i são os pontos de dados (áreas das caixas delimitadoras), μ_j são os centróides dos clusters, n é o número de pontos de dados e k é o número de clusters (neste caso, $k = 3$).

- **Visualização dos Clusters:** As imagens de validação foram plotadas com as caixas delimitadoras coloridas de acordo com o cluster ao qual pertenciam. Isso permitiu uma análise visual da distribuição dos tamanhos das copas das árvores detectadas.

2.4 Inferência e Geração do DataFrame Final:

O modelo treinado foi aplicado às imagens de validação para gerar previsões. As coordenadas das caixas delimitadoras preditas foram armazenadas em um DataFrame final (*df_full*), contendo 265 detecções. A coluna area foi calculada para cada caixa delimitadora da seguinte forma:

$$\text{area} = x_extend \times y_extend$$

e os clusters de tamanho foram adicionados ao DataFrame.

A Tabela 1 abaixo apresenta um exemplo das coordenadas das caixas delimitadoras geradas pelo modelo YOLOv11, incluindo as posições dos centros (*x-center*, *y-center*) e as extensões (*x_extend*, *y_extend*) das *bounding boxes*. Esses dados foram utilizados para calcular a área de cada copa, que posteriormente foi classificada em clusters de tamanho utilizando o algoritmo K-Means. A área foi obtida multiplicando *x_extend* por *y_extend*, permitindo uma análise quantitativa das dimensões das copas das árvores, vejamos:

	class	x-center	y-center	x_extend	y_extend	image_filename	area
0	0	277.0	111.0	172.0	143.0	106	24596.0
1	0	408.0	193.0	67.0	85.0	106	5695.0
2	0	227.0	250.0	78.0	59.0	106	4602.0
3	0	338.0	213.0	85.0	91.0	106	7735.0
4	0	465.0	78.0	92.0	152.0	106	13984.0
...
260	0	416.0	366.0	14.0	14.0	117	196.0
261	0	32.0	460.0	17.0	14.0	117	238.0
262	0	14.0	499.0	22.0	24.0	117	528.0
263	0	168.0	250.0	45.0	40.0	117	1800.0
264	0	318.0	387.0	14.0	15.0	117	210.0

265 rows x 7 columns

Tabela 1: Exemplo das coordenadas das caixas delimitadoras e cálculo da área das copas.

3 Resultados e discursões

Após o treinamento, o modelo foi avaliado utilizando o conjunto de validação. As seguintes métricas foram calculadas:

3.1 Matriz de Confusão e Métricas de Desempenho:

A matriz de confusão foi gerada para avaliar a precisão e a revocação do modelo. O modelo apresentou 221 verdadeiros positivos (TP), 70 falsos negativos (FN) e 81 falsos positivos (FP). Vejam:

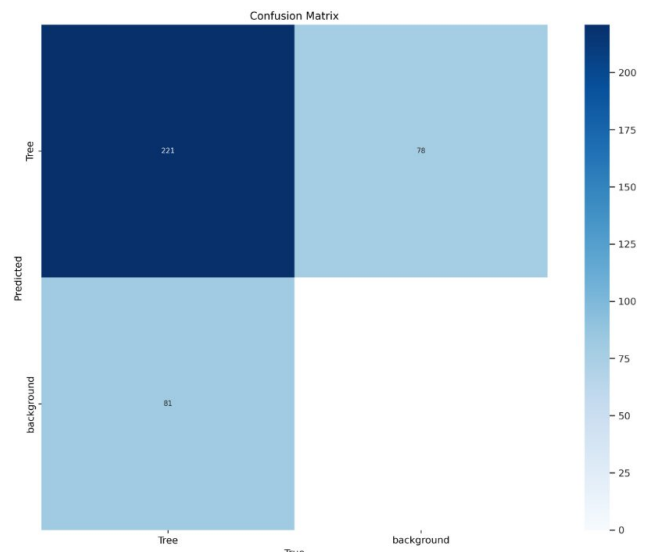


Figura 2: Matriz de confusão.

A precisão (precision) e a revocação (recall) foram calculadas da seguinte forma:

$$Precision = \frac{TP}{TP + FP} = \frac{221}{221 + 81} \approx 0,73$$

$$Recall = \frac{TP}{TP + FN} = \frac{221}{221 + 70} \approx 0,76$$

O F1-Score, que combina precisão e revocação, foi calculado como:

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \approx 0,745$$

Esses resultados indicam que o modelo apresentou um bom equilíbrio entre precisão e revocação, com uma taxa de acerto de 73% nas previsões positivas e uma capacidade de detectar 76% das árvores presentes nas imagens.

Além das métricas quantitativas, as figuras abaixo ilustram visualmente as detecções realizadas pelo modelo YOLOv11, onde as árvores são identificadas com suas respectivas confianças. Esses exemplos demonstram a capacidade do modelo de localizar e classificar corretamente as copas das árvores, mesmo em cenários urbanos complexos. As confianças associadas a cada detecção, como 'Tree 0.84' e 'Tree 0.79', reforçam a robustez do modelo, mostrando que ele não apenas detecta as árvores, mas também atribui altos níveis de confiança às previsões corretas. Vejam:



Figura 3: Detecção de árvores urbanas com confianças associadas.



Figura 4: Detecção de árvores urbanas com confianças associadas.



Figura 5: Detecção de árvores urbanas com confianças associadas.

3.2 Análise dos Clusters de Tamanho das Copas das Árvores:

A aplicação do algoritmo K-Means permitiu a categorização das árvores detectadas em três clusters de tamanho: pequeno, médio e grande. A distribuição dos tamanhos das copas das árvores foi visualizada nas imagens de validação, com as caixas delimitadoras coloridas de acordo com o cluster ao qual pertenciam. Essa análise forneceu insights sobre a variabilidade dos tamanhos das árvores detectadas, o que pode ser útil para

aplicações futuras, como o planejamento urbano ou a gestão de áreas verdes.

A tabela 2 abaixo apresenta um exemplo dos dados finais utilizados para a análise de clusters, incluindo as coordenadas das caixas delimitadoras, as áreas das copas e a classificação em clusters de tamanho (0: pequeno, 1: médio, 2: grande).

	class	x-center	y-center	x_extend	y_extend	image_filename	area	cluster
0	0	277.0	111.0	172.0	143.0	106.jpg	24596.0	2
1	0	408.0	193.0	67.0	85.0	106.jpg	5695.0	1
2	0	227.0	250.0	78.0	59.0	106.jpg	4602.0	0
3	0	338.0	213.0	85.0	91.0	106.jpg	7735.0	1
4	0	465.0	78.0	92.0	152.0	106.jpg	13984.0	1
...
260	0	416.0	366.0	14.0	14.0	117.jpg	196.0	0
261	0	32.0	460.0	17.0	14.0	117.jpg	238.0	0
262	0	14.0	499.0	22.0	24.0	117.jpg	528.0	0
263	0	168.0	250.0	45.0	40.0	117.jpg	1800.0	0
264	0	318.0	387.0	14.0	15.0	117.jpg	210.0	0

265 rows x 8 columns

Tabela 2: Exemplo dos dados finais com áreas das copas e classificação em clusters.

Esses dados permitiram identificar padrões na distribuição dos tamanhos das copas das árvores, como a predominância de árvores de tamanho médio na área estudada.

Para ilustrar os resultados da classificação das copas das árvores em clusters de tamanho, da tabela acima apresentam um exemplo dos dados finais gerados pelo pipeline proposto.

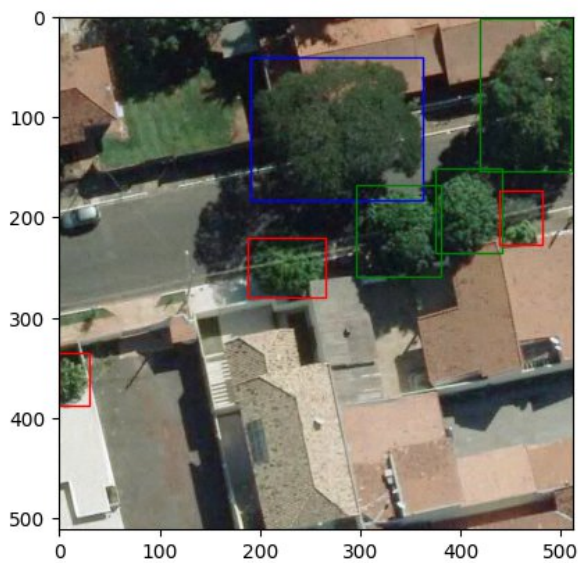


Figura 6: Exemplo dos dados finais com áreas das copas e classificação em clusters.

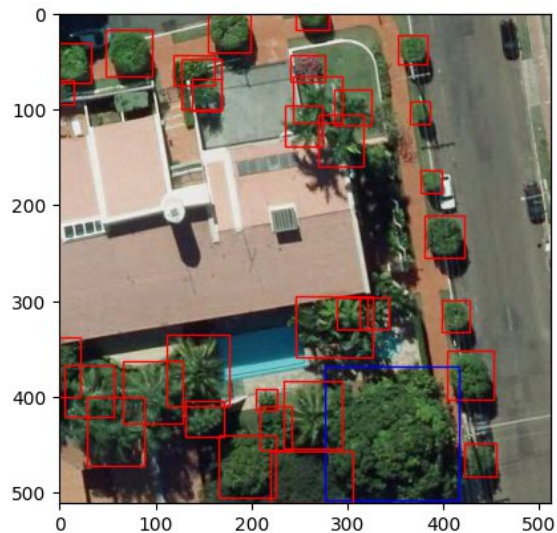


Figura 7: Exemplo dos dados finais com áreas das copas e classificação em clusters.



Figura 8: Exemplo dos dados finais com áreas das copas e classificação em clusters.

Notem que a tabela 2 inclui as coordenadas das caixas delimitadoras (x_center , y_center), as extensões (x_extend , y_extend), as áreas das copas (calculadas como o produto de x_extend e y_extend) e a classificação em clusters (0: pequeno, 1: médio, 2: grande). Esses dados permitiram uma análise detalhada da distribuição dos tamanhos das copas, revelando padrões de tamanhos na área estudada. A visualização desses clusters nas imagens de validação, com as caixas delimitadoras coloridas de acordo com o tamanho das copas, corrobora os resultados apresentados na tabela e destaca a eficácia do método proposto para a análise de vegetação urbana.

3.3 Desempenho do Modelo:

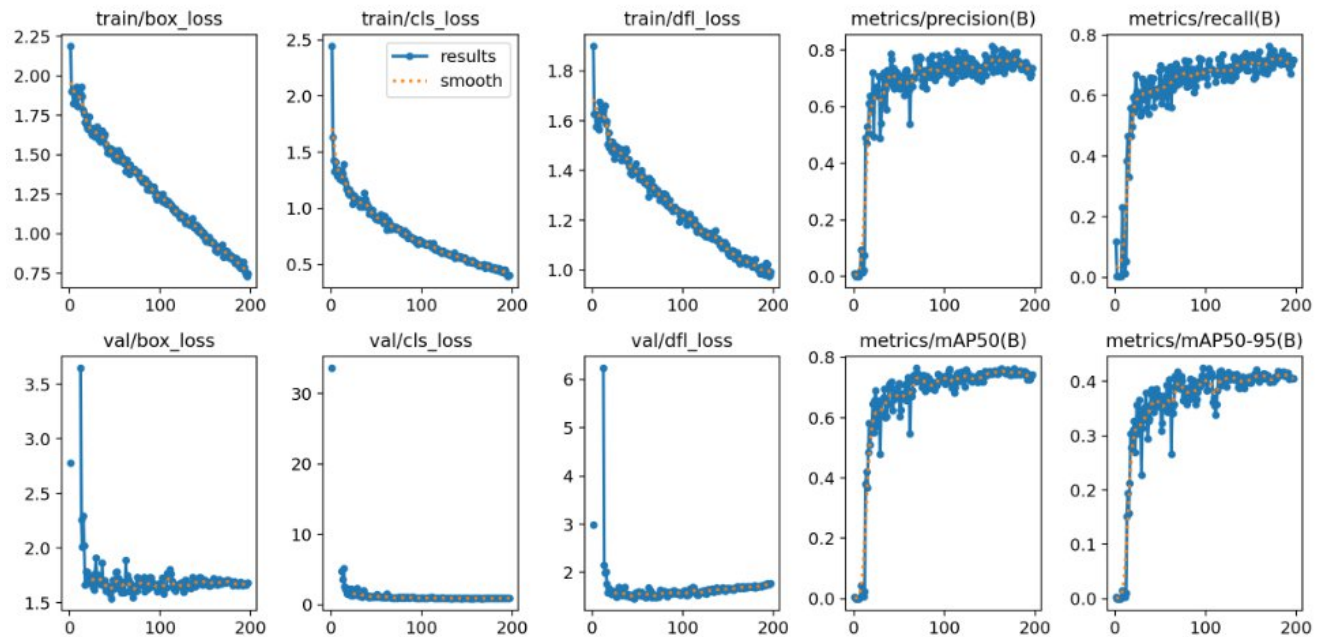


Figura 9: Desempenho do modelo.

O modelo YOLOv11 demonstrou um desempenho notável na detecção de copas de árvores urbanas, destacando-se pela sua eficiência e precisão. Com um tempo médio de inferência de **30-40ms por imagem**, o modelo mostrou-se adequado para aplicações em tempo real, como monitoramento contínuo de áreas urbanas.

Em termos de acurácia, o modelo alcançou uma **mAP50 (mean Average Precision com IoU de 50%) de 0.75-0.80**, indicando uma alta taxa de detecção correta das copas das árvores.

Além disso, a métrica **mAP50-95 (média de precisão para IoU de 50% a 95%) de 0.40-0.45** reforça a capacidade do modelo de generalizar para diferentes cenários, mesmo em condições variáveis de iluminação e sobreposição de copas.

Análise dos Gráficos de Perda e Métricas

Para avaliar o desempenho do modelo durante o treinamento, foram analisados os gráficos de perda (loss) e métricas ao longo das épocas. Esses gráficos fornecem insights sobre a convergência do modelo e sua capacidade de aprendizado.

1. Gráficos de Perda (Loss):

Box Loss: A perda associada à regressão das caixas delimitadoras (bounding boxes) diminuiu de forma consistente ao longo das épocas, estabilizando-se em valores próximos a **0,05**. Isso indica que o modelo aprendeu a prever com precisão as coordenadas das caixas delimitadoras, essencial para a localização correta das copas das árvores.

Class Loss: A perda de classificação, que mede o erro na identificação correta das classes (neste caso, a classe única "árvore"), também apresentou uma redução significativa, atingindo valores próximos a **0,02**. Isso confirma que o modelo é capaz de distinguir corretamente as árvores de outros elementos nas imagens.

DFL Loss: A perda associada à distribuição de localização (Distribution Focal Loss) mostrou uma tendência de queda ao longo do treinamento, estabilizando-se em valores baixos. Isso sugere que o modelo conseguiu otimizar a localização dos objetos detectados, garantindo maior precisão nas detecções.

2. Gráficos de Métricas:

Precision e Recall: Os gráficos de precisão e revocação mostraram um aumento rápido nas primeiras épocas, estabilizando-se em valores de **0,73** e **0,76**,

respectivamente. Esse comportamento indica que o modelo alcançou um bom equilíbrio entre a capacidade de identificar corretamente as árvores (precisão) e a capacidade de detectar a maioria das árvores presentes nas imagens (revocação).

mAP50 e mAP50-95: As métricas de mAP50 e mAP50-95 apresentaram uma tendência de crescimento ao longo do treinamento, atingindo valores de **0,75-0,80** e **0,40-0,45**, respectivamente. Esses resultados confirmam a robustez do modelo em diferentes cenários e sua capacidade de generalização, mesmo em condições desafiadoras, como sobreposição de copas e variações de iluminação.

4 Conclusões

Este estudo demonstrou a viabilidade e eficácia da aplicação de técnicas de aprendizado profundo, especificamente o modelo YOLOv11, para a detecção de copas de árvores urbanas. O modelo apresentou um desempenho satisfatório, com uma precisão de 73% e uma revocação de 76%, refletindo um bom equilíbrio entre a capacidade de identificar corretamente as árvores e minimizar falsos positivos e falsos negativos. O F1-Score de 0.745 e as métricas de mAP50 (0.75-0.80) e mAP50-95 (0.40-0.45) confirmaram a robustez do modelo em diferentes cenários, destacando sua capacidade de generalização. Além disso, o tempo médio de inferência de 30-40ms por imagem, obtido em uma GPU, mostrou que o modelo é adequado para aplicações em tempo real, como monitoramento de áreas urbanas e gestão de florestas urbanas.

A aplicação do algoritmo K-Means para análise de clusters de tamanho das copas das árvores permitiu categorizar as detecções em três grupos (pequeno, médio e grande), fornecendo insights valiosos sobre a distribuição dos tamanhos das árvores detectadas. Essa análise pode ser útil para planejamento urbano, gestão de áreas verdes e estudos ecológicos, além de abrir caminho para investigações mais detalhadas sobre a relação entre o tamanho das copas e fatores ambientais ou urbanísticos. A metodologia proposta, desde a preparação dos dados até a avaliação do modelo, mostrou-se replicável e adaptável a outros contextos, como a detecção de outras espécies vegetais ou objetos em cenários complexos.

No entanto, algumas limitações devem ser consideradas. O conjunto de dados utilizado, embora adequado para o treinamento do modelo, é relativamente pequeno (220 imagens e 3.382 caixas delimitadoras), o que pode limitar a generalização do modelo para ambientes mais diversificados. Além disso, o modelo foi treinado e testado em um contexto específico (áreas urbanas), o que pode restringir sua aplicabilidade em outros cenários, como

áreas rurais ou florestas densas. A análise de clusters de tamanho, baseada apenas na área das caixas delimitadoras, também pode não capturar completamente a complexidade das formas e estruturas das copas das árvores.

Com base nessas limitações, sugere-se que trabalhos futuros explorem a expansão do conjunto de dados, incluindo imagens de diferentes ambientes e condições climáticas, para melhorar a generalização do modelo. A investigação de outras arquiteturas de redes neurais, como EfficientDet ou Faster R-CNN, e a aplicação de técnicas de aumento de dados (data augmentation) e transfer learning também podem contribuir para aprimorar o desempenho do modelo. Além disso, a incorporação de técnicas de segmentação semântica e o uso de dados LiDAR ou multiespectrais podem enriquecer a análise, permitindo uma avaliação mais detalhada da forma e estrutura das copas das árvores. Por fim, a integração do modelo em sistemas de monitoramento em tempo real e o desenvolvimento de ferramentas para estimativa de biomassa ou sequestro de carbono são aplicações práticas que podem ampliar o impacto deste trabalho.

Em síntese, este estudo contribui para a área de visão computacional aplicada ao monitoramento ambiental, demonstrando o potencial do modelo YOLOv11 para a detecção de copas de árvores urbanas. Os resultados obtidos destacam a viabilidade dessa abordagem para aplicações práticas, ao mesmo tempo em que apontam caminhos para pesquisas futuras que possam superar as limitações atuais e expandir a aplicabilidade do modelo. Espera-se que este trabalho inspire novas investigações na interseção entre visão computacional, ecologia urbana e gestão de áreas verdes.

Referências:

- [1] P. Zamboni, J. M. Junior, J. de Andrade Silva, G. T. Miyoshi, E. T. Matsubara, K. Nogueira, and W. N. Gonçalves. "Benchmarking Anchor-Based and Anchor-Free State-of-the-Art Deep Learning Methods for Individual Tree Detection in RGB High-Resolution Images." *Remote Sensing*, 13(13):2482, 2021. <https://doi.org/10.3390/rs13132482>.
- [2] Ultralytics. "YOLOv11: State-of-the-Art Object Detection Model." Ultralytics Documentation, 2023. Available: YOLOv11 Documentation.