Evaluation of Tiling and SAHI Strategies for Symptomatic Region Detection in UAV Images of Coffee Plantations Using YOLO

Luan Pinto[1], **Guilherme Silva**[2], **Arthur Negrão**[1], **Mateus Bueno**[3], **Pedro Silva**[4]

[1] Postgraduate Program in Computer Science, Federal University of Ouro Preto
Ouro Preto – MG – Brazil
[2] Computing Department – Federal University of São João del-Rei
São João del-Rei – MG – Brasil
[3] Federal Institute of Education, Science and Technology of Southern Minas Gerais
Campus Muzambinho – Muzambinho – MG – Brazil
[4] Computing Department – Federal University of Ouro Preto
Ouro Preto – MG – Brazil
luan.patrik@aluno.ufop.edu.br

Abstract. *Automatic detection of symptomatic regions in coffee plantations from aerial imagery can support large-scale phytosanitary monitoring and reduce agricultural production losses. This work evaluates YOLO-based strategies for detecting symptomatic regions in RGB UAV images. A dataset composed of 162 images and 641 annotations distributed across three symptom severity levels was used. Three approaches were compared: conventional YOLO, YOLO with Slicing Aided Hyper Inference (SAHI), and YOLO trained with image tiling combined with SAHI inference. The YOLO + SAHI strategy achieved the best performance, reaching 53.9% precision, 53.9% mAP@50, and 35.7% mAP@50-95. The findings indicate that SAHI improves the detection of small symptomatic regions by preserving local image information, highlighting its potential for automated agricultural monitoring applications.*

Resumo. *A detecção automática de regiões sintomáticas em cafezais por meio de imagens aéreas pode auxiliar o monitoramento fitossanitário em larga escala e reduzir perdas na produção agrícola. Este trabalho avalia estratégias baseadas na arquitetura YOLO para detectar regiões sintomáticas em imagens RGB obtidas por VANTs. Foram utilizadas 162 imagens contendo 641 anotações distribuídas em três níveis de severidade dos sintomas. Três abordagens foram comparadas: YOLO convencional, YOLO com Slicing Aided Hyper Inference (SAHI) e YOLO treinado com tiling combinado ao SAHI. A estratégia YOLO + SAHI apresentou o melhor desempenho, alcançando 53,9% de precisão, 53,9% de mAP@50 e 35,7% de mAP@50-95. Os resultados indicam que o uso de SAHI melhora a detecção de pequenas regiões sintomáticas ao preservar informações locais da imagem, demonstrando potencial para aplicações em monitoramento agrícola automatizado.*

1. Introdução

A cafeicultura possui elevada relevância econômica e social, especialmente em países produtores como o Brasil, que figura entre os maiores produtores e exportadores de café do mundo [Companhia Nacional de Abastecimento 2023,

Food and Agriculture Organization of the United Nations 2023]. A produtividade e a qualidade dos grãos dependem diretamente das condições fitossanitárias das lavouras, o que torna o monitoramento contínuo das plantas uma etapa central do manejo agrícola [Mesquita et al. 2016]. Nesse contexto, a identificação precoce de sintomas associados a estresses bióticos e abióticos pode reduzir perdas produtivas e apoiar estratégias de manejo mais eficientes [Albuquerque and Guedes 2024, Mesquita et al. 2016].

A inspeção de lavouras é tradicionalmente realizada por vistorias em campo, um processo que demanda tempo, mão de obra especializada e oferece cobertura espacial limitada. Com a expansão das áreas cultivadas, esses métodos tornam-se menos eficientes para a detecção rápida de problemas localizados. Como alternativa, a Agricultura de Precisão tem incorporado tecnologias capazes de aumentar a escala, a resolução e a frequência do monitoramento agrícola [Lamparelli 2016, Bernardi et al. 2017, Inamasu et al. 2014].

Entre essas tecnologias, destacam-se os veículos aéreos não tripulados (VANTs), que permitem adquirir imagens de alta resolução de forma rápida e com baixo custo operacional [Pereira et al. 2023, Goltz 2011]. Imagens aéreas obtidas por VANTs possibilitam a observação detalhada da lavoura e a extração de informações visuais relacionadas ao vigor vegetativo, falhas de plantio, estresse hídrico e padrões visuais associados à ocorrência de pragas e doenças. Entretanto, o grande volume de imagens gerado exige métodos automatizados capazes de extrair informações relevantes de maneira eficiente.

Nesse cenário, técnicas de visão computacional e aprendizado profundo têm sido amplamente empregadas na análise automática de imagens agrícolas [Marengoni and Stringhini 2009, Ludermir 2021, Yang and Guo 2017]. Entre os modelos de detecção de objetos, a família *You Only Look Once* (YOLO) destaca-se pela alta velocidade de inferência e pelo desempenho competitivo em diferentes aplicações [Redmon et al. 2016]. Trabalhos anteriores demonstraram a viabilidade do uso de YOLO em imagens obtidas por drones para tarefas como contagem automática de mudas de café [Santana et al. 2019], detecção de gado [Takano et al. 2021, Manfrim 2022].

Apesar desses avanços, a detecção de regiões sintomáticas em cafezais a partir de imagens aéreas ainda apresenta desafios relevantes, sobretudo quando os sintomas ocupam pequenas porções da imagem. Em detectores como YOLO, a imagem de entrada é usualmente redimensionada para uma resolução fixa, por exemplo, 1.024×1.024 *pixels*. Embora esse procedimento favoreça a eficiência computacional, ele pode reduzir a representação espacial de alvos pequenos e eliminar detalhes discriminativos necessários à detecção. Além disso, fatores como altura de voo, resolução espacial e variação nas condições de iluminação podem afetar significativamente o desempenho dos modelos [Pinto et al. 2026].

Diante desse contexto, este trabalho investiga se estratégias voltadas à detecção de pequenos alvos em imagens aéreas podem melhorar a identificação automática de regiões sintomáticas em cafezais. Para isso, avalia-se o uso da arquitetura YOLO combinada com *tiling* e *Slicing Aided Hyper Inference* (SAHI) [Akyon et al. 2022]. São comparadas três abordagens: (i) inferência convencional com YOLO em imagens completas; (ii) inferência com YOLO assistida por SAHI; e (iii) treinamento baseado em *tiling* combinado com inferência assistida por SAHI. O *tiling* consiste na divisão de imagens de alta resolução em blocos menores para o treinamento, enquanto o SAHI sistematiza esse processo durante a

inferência ao recortar a imagem, aplicar o detector em cada recorte, reprojetar as caixas delimitadoras para o sistema de coordenadas original, agregar as predições e remover duplicatas. Essa estratégia busca preservar detalhes locais de alvos pequenos que poderiam ser degradados pelo redimensionamento da imagem completa.

As principais contribuições deste trabalho são: (i) avaliar estratégias de recorte para a detecção de regiões sintomáticas em imagens aéreas de cafezais; (ii) comparar a inferência convencional com YOLO, a inferência assistida por SAHI e o treinamento com *tiling* associado ao SAHI; e (iii) analisar o impacto dessas estratégias na detecção de alvos pequenos em imagens de alta resolução. Os resultados indicam que a combinação do YOLO com a inferência com SAHI melhora a detecção de regiões sintomáticas em relação à inferência convencional, conforme discutido na Seção 4.

O restante deste trabalho está organizado da seguinte forma: a Seção 2 apresenta os trabalhos relacionados. A Seção 3 descreve a metodologia utilizada, e os resultados na Seção 4. Por fim, a Seção 5 apresenta as conclusões e os trabalhos futuros.

2. Trabalhos Relacionados

Em seu artigo, [Chaurasia and Patro 2023] treinaram modelos YOLOv7 em conjunto com o *framework* SAHI para a detecção de pássaros em tempo real para vigilância de fazendas. Observou-se que o uso de SAHI aumentou o número de detecções em 136,6% para modelos RGB e 82% para modelos infravermelhos, demonstrando a eficácia do fatiamento de imagens para identificar alvos pequenos e distantes.

De forma semelhante, [Muzammul et al. 2024] integraram o modelo Real-Time Detection Transformers (RT-DETR-X) com a metodologia SAHI utilizando o *dataset* VisDrone, alcançando uma precisão média (AP) de 54,8% a 74 FPS. Este trabalho reforçou que a combinação de detectores de tempo real com técnicas de fatiamento supera modelos tradicionais como Faster R-CNN e versões base de YOLO em cenários de alta resolução com objetos em pequena escala.

No contexto específico de pragas e culturas agrícolas, [Ye et al. 2024] propuseram um modelo leve baseado em YOLOv8 para a detecção de pragas de chá, incorporando uma rede de agregação de camadas eficiente generalizada (GELAN) e o mecanismo de atenção BiFormer. O modelo alcançou uma precisão de 96,32% e um mAP@50 de 98,17%, superando modelos convencionais e equilibrando alta velocidade de processamento com precisão em alvos microscópicos.

No âmbito do monitoramento aéreo, [Liang et al. 2025] desenvolveram um algoritmo nomeado de SAHI-Improved-YOLOv8 para a detecção de covas de plantio de árvores em imagens de alta resolução capturadas por drones. Para superar desafios como a distribuição densa e o tamanho reduzido dos alvos, os autores incorporaram ao modelo as técnicas de Focal loss, Spatial Pyramid Pooling Concurrent Spatial Pyramid Convolution (SPPCSPC) e o Convolutional Block Attention Module (CBAM), visando mitigar o desequilíbrio de dados e fortalecer características semânticas fundamentais. Os resultados demonstraram que esta abordagem superou versões do YOLOv3 ao YOLOv11, atingindo uma precisão de 85,17% e um F1-score de 85,12%, provendo uma solução robusta para o gerenciamento florestal em larga escala.

Especificamente na avaliação de qualidade do café,[Adiwijaya et al. 2024] aplica-

ram o YOLOv8 em conjunto com o *framework* SAHI para detectar imperfeições em grãos de café Robusta. O estudo destacou que, enquanto pesquisas anteriores focavam em grãos isolados, a integração com SAHI permite lidar com cenários reais onde os grãos estão densamente compactados e sobrepostos, alcançando valores de mAP superiores a 99%.

Apesar dos trabalhos citados acima, observa-se que a maioria das pesquisas concentra-se na detecção de objetos bem definidos, como aves, pragas, covas ou imperfeições em produtos agrícolas. Diferentemente, este trabalho aborda um problema que lida com a variação sutil de coloração nas folhas do cafeeiro, sendo essa uma mudança que ocorre de maneira gradual conforme a idade da folha e as condições hídricas, o que torna a tarefa mais complexa do que a detecção de objetos com forma e aparência bem definidas.

Além disso, as regiões sintomáticas ocupam apenas pequenas porções das imagens aéreas, o que pode dificultar sua detecção após o redimensionamento aplicado por modelos como o YOLO. Nesse contexto, técnicas como SAHI e *tiling* podem contribuir para aumentar a representatividade desses alvos durante a inferência e o treinamento. Assim, este trabalho avalia o impacto dessas estratégias na detecção de regiões sintomáticas em cafezais.

3. Metodologia Proposta

Esta seção descreve a metodologia adotada para avaliar estratégias de detecção de regiões sintomáticas em imagens aéreas de cafezais. Inicialmente, apresenta-se a base de dados utilizada, incluindo o processo de aquisição das imagens, as classes de detecção e o particionamento experimental. Em seguida, são descritas as estratégias de treinamento e inferência avaliadas. Por fim, são definidas as métricas empregadas para comparar o desempenho dos modelos.

3.1. Base de Dados

Os experimentos utilizam a base de dados de imagens aéreas de cafezais apresentada em [Pinto et al. 2026]. O conjunto foi construído a partir de imagens RGB adquiridas por veículos aéreos não tripulados (VANTs) em diferentes alturas de voo: 10, 15 e 25 metros; e ângulos de aquisição: 30°, 45° e 90°, contemplando variações de perspectiva e escala dos alvos de interesse.

As imagens contêm anotações de regiões sintomáticas em plantas de café, organizadas em três classes: estágio inicial, intermediário e avançado. Essas classes representam diferentes níveis de severidade dos sintomas visualmente observáveis nas plantas. O processo de aquisição das imagens, a caracterização da área experimental, a definição das classes e o protocolo de anotação são descritos em detalhes em [Pinto et al. 2026].

Neste trabalho, foram utilizadas as mesmas imagens e anotações disponibilizadas na base original, totalizando 162 imagens RGB com resolução de 5.472×3.648 *pixels* e 641 anotações. As anotações seguem o formato YOLO, no qual cada objeto é representado pelas coordenadas normalizadas (x, y) do centro da caixa delimitadora, bem como por sua largura e altura normalizadas.

Seguindo a metodologia adotada em [Pinto et al. 2026], o conjunto de dados foi particionado em 126 imagens para treinamento e 36 imagens para teste. O conjunto de

treinamento foi utilizado para ajustar os parâmetros dos modelos avaliados, enquanto o conjunto de teste foi reservado para a avaliação de desempenho e comparação entre as estratégias investigadas. A Figura 1 apresenta exemplos de imagens do conjunto de dados.



Figura 1. Exemplo das imagens utilizadas durante o treinamento.

Pré-processamento: No fluxo convencional de treinamento, as imagens completas, originalmente com resolução de 5.472×3.648 *pixels*, foram redimensionadas para 1.024×1.024 *pixels*. Esse procedimento preserva o fluxo padrão de entrada do detector, mas reduz a escala relativa das pequenas regiões sintomáticas, o que pode dificultar sua detecção.

3.2. Estratégias de treinamento

Os experimentos foram estruturados para analisar o impacto de diferentes estratégias de treinamento e inferência na detecção de regiões sintomáticas em cafeeiros. Foram avaliadas três configurações: (i) inferência convencional com YOLO em imagens completas; (ii) inferência com YOLO assistida por *Slicing Aided Hyper Inference* (SAHI); e (iii) treinamento com *tiling* combinado com inferência assistida por SAHI.

As Estratégias 1 e 2 utilizam o mesmo procedimento de treinamento e diferem apenas na etapa de inferência. As Estratégias 2 e 3 utilizam a mesma inferência assistida por SAHI, mas diferem no procedimento de treinamento. Dessa forma, o desenho experimental permite isolar o efeito do SAHI durante a inferência e o efeito adicional do *tiling* durante o treinamento.

O detector YOLO versão 26 na variante *nano* (YOLO26n) foi adotado por sua compatibilidade com fluxos eficientes de detecção de objetos e por seu suporte a cenários com alvos em diferentes escalas [Jocher et al. 2026]. Essas características são relevantes para o problema investigado, uma vez que regiões sintomáticas em imagens aéreas frequentemente ocupam pequenas porções da imagem e podem perder detalhes discriminativos após o redimensionamento da entrada. A Tabela 1 resume as estratégias avaliadas.

Tabela 1. Resumo das estratégias avaliadas.

Estratégia	Treinamento	Inferência
YOLO padrão	Imagem completa	Imagem completa
YOLO + SAHI	Imagem completa	SAHI
YOLO + *tiling* + SAHI	*Tiling*	SAHI

3.2.1. Estratégia 1 – YOLO Padrão

A primeira estratégia corresponde ao *baseline* experimental. Nessa configuração, o treinamento foi realizado com as imagens originais, sem subdivisão adicional dos dados.

Durante a inferência, as imagens de teste foram processadas integralmente pelo detector, seguindo o fluxo convencional de detecção de objetos. A Figura 2 resume essa estratégia, em que as imagens originais são utilizadas no treinamento do modelo YOLO e, na etapa de inferência, cada imagem de teste é processada como uma única entrada.

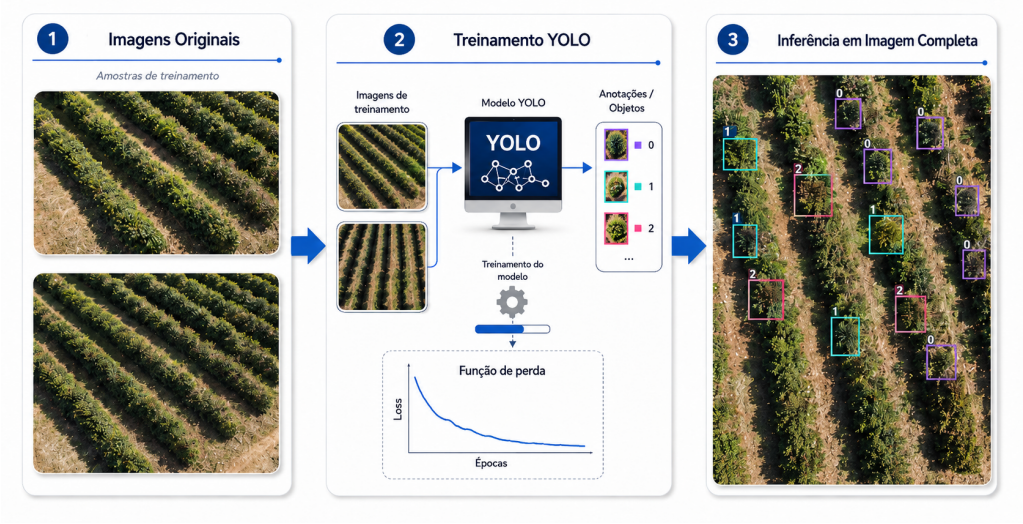


Figura 2. Estratégia *baseline*.

3.2.2. Estratégia 2 – YOLO + SAHI

Na segunda estratégia, o treinamento foi realizado com a mesma configuração da Estratégia 1. A diferença ocorre apenas na etapa de inferência, na qual foi aplicada a técnica SAHI. O SAHI divide cada imagem em múltiplos recortes menores, aplica o detector individualmente em cada recorte e, em seguida, combina as predições no sistema de coordenadas da imagem original [Akyon et al. 2022].

A estratégia SAHI busca mitigar a perda de detalhes associada ao redimensionamento de imagens de alta resolução. Ao processar recortes menores, os objetos pequenos passam a ocupar uma área proporcionalmente maior na entrada do modelo, o que pode favorecer sua detecção [Akyon et al. 2022]. Além disso, a sobreposição entre recortes reduz a probabilidade de perda de objetos localizados nas bordas das janelas. Após a detecção em cada recorte, as caixas delimitadoras são reprojetadas para a imagem original e pós-processadas por *Non-Maximum Suppression* (NMS), a fim de remover predições duplicadas em regiões sobrepostas. Uma vantagem dessa estratégia é permitir a aplicação de SAHI na inferência sem necessidade de retreinamento do detector. A Figura 3 resume essa configuração, na qual o treinamento permanece idêntico ao da estratégia *baseline*, enquanto a inferência é realizada por meio de recortes processados com SAHI.

Os parâmetros utilizados no SAHI incluem um limiar de confiança de 0,25, janelas de 512×512 *pixels* e uma taxa de sobreposição de 0,2 nas direções horizontal e vertical. As detecções obtidas em cada janela foram combinadas por NMS para eliminar predições duplicadas em regiões sobrepostas. Esses parâmetros foram definidos com base em [Akyon et al. 2021].

Com uma janela de tamanho w e taxa de sobreposição r (neste trabalho, r = 0,2), o

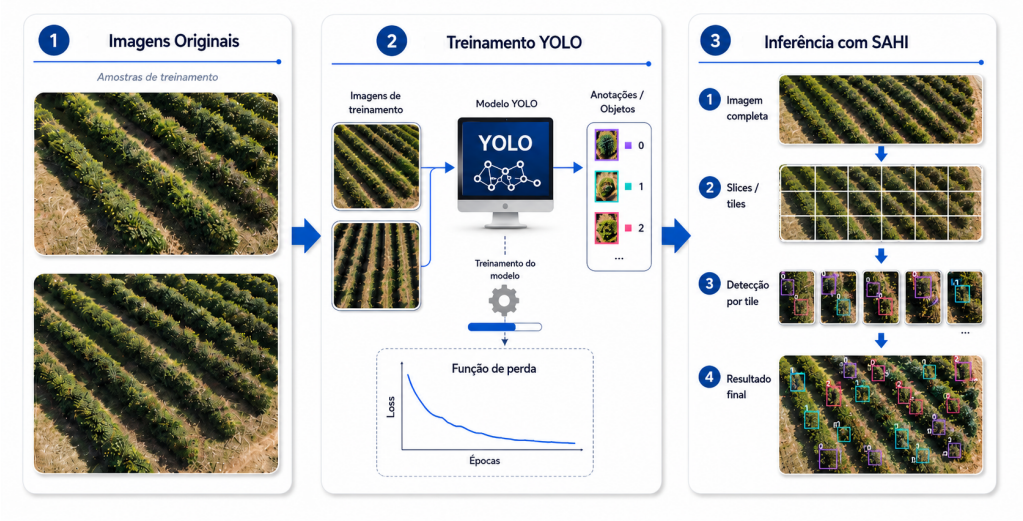


Figura 3. Estratégia utilizando SAHI na inferência.

deslocamento entre janelas adjacentes é dado por: $s = w(1 - r)$. Assim, para $w = 512$ e $r = 0,2$, tem-se $s = 409,6 \approx 410$ *pixels*. Com essa configuração, cada imagem original é dividida em 14 recortes na largura e nove recortes na altura, totalizando 126 recortes por imagem.

3.2.3. Estratégia 3 – YOLO Treinado com Tiling + SAHI

A terceira estratégia combina *tiling* durante o treinamento e SAHI durante a inferência. Inicialmente, as imagens do conjunto de treinamento foram subdivididas em blocos menores, de modo que as regiões sintomáticas ocupassem uma área proporcionalmente maior em cada recorte. Diferentemente do SAHI, aplicado na inferência, o *tiling* é utilizado para gerar amostras de treinamento a partir de imagens de alta resolução.

O uso de *tiling* busca reduzir os efeitos da perda de escala causada pelo redimensionamento das imagens durante o treinamento. Como consequência, regiões sintomáticas pequenas tendem a ocupar uma fração maior dos blocos gerados, favorecendo a aprendizagem de características visuais associadas a alvos pequenos e potencialmente aumentando a capacidade de detecção do modelo.

A inferência segue o mesmo procedimento da Estratégia 2, com a aplicação de SAHI nas imagens de teste. A diferença está no conjunto de treinamento: cada imagem foi particionada em blocos de 1.024×1.024 *pixels*, com sobreposição de 20% entre blocos adjacentes. Durante a geração dos blocos, apenas objetos com pelo menos 40% de sua área original visível dentro do recorte foram mantidos nas anotações correspondentes. Além disso, foi adotado um limite máximo de 25% de blocos sem objetos anotados, a fim de controlar o desbalanceamento entre amostras positivas e negativas. Esses parâmetros foram definidos com base em [Akyon et al. 2021].

Essa configuração permite avaliar o efeito combinado do aumento relativo da escala dos objetos durante o treinamento e da análise local promovida pelo SAHI durante a inferência. A Figura 4 resume essa estratégia, na qual o modelo YOLO é treinado com blocos extraídos por *tiling* e avaliado com inferência assistida por SAHI.

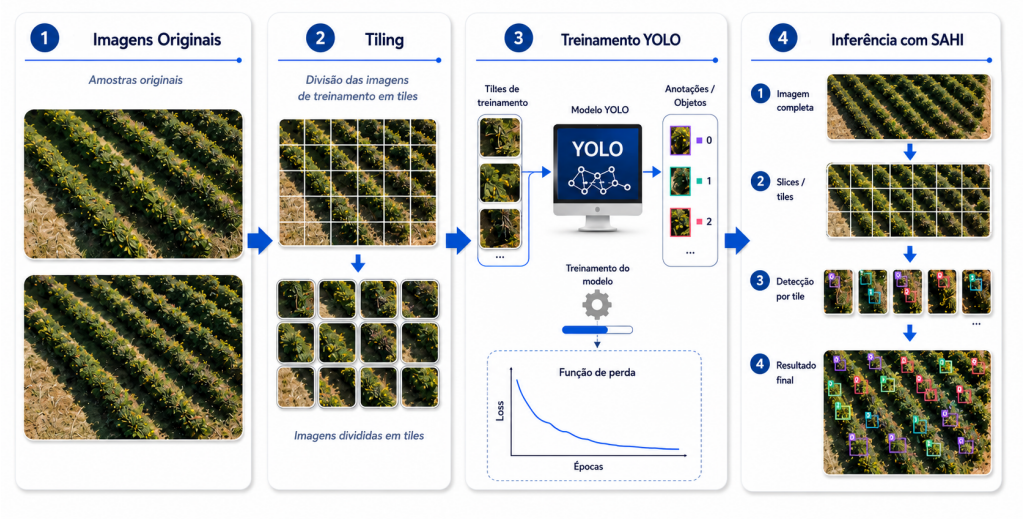


Figura 4. Estratégia utilizando SAHI na inferência e *tiling* no treinamento.

Na etapa de *tiling*, a resolução dos blocos foi definida em 1024×1024 *pixels*, com sobreposição de 20% entre blocos adjacentes. Usando $s = w(1 - r)$, para $w = 1.024$ e $r = 0,2$, obtém-se $s = 819,2 \approx 819$ *pixels*. Essa configuração resulta em até sete recortes na largura e cinco recortes na altura para cada imagem original, ou 35 blocos candidatos por imagem. Nem todos os blocos candidatos são necessariamente preservados: blocos sem anotações são limitados a, no máximo, 25% do conjunto gerado.

Na etapa de inferência, manteve-se a mesma configuração de SAHI da Estratégia 2, com janelas de 512×512 *pixels*, sobreposição de 20% e NMS para remoção de duplicatas. Dessa forma, as Estratégias 2 e 3 diferem apenas no procedimento de treinamento, permitindo isolar o efeito do *tiling* na aprendizagem do detector.

3.3. Avaliação do Modelo

O desempenho dos modelos foi avaliado por quatro métricas amplamente utilizadas em detecção de objetos: precisão (*Precision*), revocação (*Recall*), *Mean Average Precision* com limiar de IoU de 50% (mAP@50) e *F1-score*.

A precisão mede a proporção de predições positivas corretas entre todas as predições positivas feitas pelo modelo, refletindo sua confiabilidade. A revocação mede a proporção de instâncias relevantes corretamente detectadas, indicando a capacidade do modelo de recuperar os objetos de interesse. O mAP@50 avalia a precisão média considerando uma sobreposição mínima de 50%, medida pela métrica *Intersection over Union* (IoU), entre a caixa predita e a caixa de referência. Por fim, o *F1-score* corresponde à média harmônica entre precisão e revocação, fornecendo uma medida única que equilibra falsos positivos e falsos negativos.

4. Experimentos e Resultados

Esta seção apresenta os resultados experimentais obtidos na comparação entre as três estratégias descritas na Seção 3.

4.1. *Setup* de Experimentos

O modelo YOLO26 foi treinado por 150 épocas, com decaimento de peso (*weight decay*) de 5×10^{-4}, taxa de aprendizado inicial (lr_0) de 1×10^{-2}, momento (*momentum*) de 0,937

e tamanho de lote (*batch size*) igual a 8. Esses hiperparâmetros foram definidos empiricamente a partir do desempenho observado em experimentos preliminares, considerando as características do conjunto de dados utilizado. O último modelo de cada estratégia foi selecionado para relatar os resultados, e as estratégias de inferência foram avaliadas sob as mesmas configurações de entrada e pós-processamento e o mesmo modelo YOLO padrão.

O treinamento foi conduzido em Python, utilizando PyTorch v1.8 e Ultralytics v8.3.146. Os experimentos foram executados em uma máquina equipada com processador Intel i9-10900, com 10 núcleos físicos, 20 *threads* e frequência base de 2,80 GHz, 128 GB de memória RAM DDR4 e uma GPU NVIDIA RTX 3090 com 24 GB de memória GDDR6X.

4.2. Resultados

A Tabela 2 compara o desempenho das estratégias avaliadas. A combinação YOLO + SAHI obteve os maiores valores de precisão $(0,539)$, F1-score $(0,308)$, mAP50 $(0,539)$ e mAP50-95 $(0,357)$. Esses resultados sugerem que a inferência assistida por SAHI pode beneficiar a detecção de regiões sintomáticas pequenas e pouco contrastantes ao preservar informações locais que tendem a ser atenuadas pelo redimensionamento global da imagem.

Tabela 2. Comparação das métricas obtidas por diferentes métodos de inferência.

Método	Precisão	*Recall*	F1-Score	mAP50	mAP50-95
YOLO	0,466	0,222	0,300	0,466	0,315
YOLO + SAHI	**0,539**	0,215	**0,308**	**0,539**	**0,357**
Tiling + YOLO + SAHI	0,279	**0,275**	0,277	0,279	0,183

O desempenho superior de YOLO + SAHI em precisão e mAP é consistente com as características visuais do problema: os sintomas foliares apresentam contornos pouco definidos e se manifestam por variações sutis de coloração e textura. Nesse cenário, o uso de SAHI favorece a análise de regiões locais da imagem, aumentando a resolução efetiva dos alvos durante a inferência e, consequentemente, a capacidade do detector de capturar padrões discretos associados ao estresse vegetal.

Apesar do melhor desempenho em precisão e mAP, a estratégia YOLO + SAHI apresentou *recall* ligeiramente inferior ao YOLO padrão. Esse resultado pode estar associado ao processo de filtragem e fusão de detecções realizado pelo SAHI, que pode favorecer predições mais confiáveis em detrimento de uma cobertura mais ampla dos objetos detectáveis. Essa hipótese deve ser confirmada por uma análise explícita de falsos positivos e falsos negativos.

A abordagem *Tiling* + YOLO + SAHI apresentou o menor desempenho em precisão $(0,279)$ e mAP50-95 $(0,183)$, embora tenha obtido o maior *recall* $(0,275)$. Como as Estratégias 2 e 3 compartilham a mesma etapa de inferência, diferindo apenas no treinamento, essa queda de desempenho é atribuível ao procedimento de *tiling* adotado no treinamento, e não totalmente ao SAHI. Isso se deve pela diferença de tamanho entre os recortes do treino (1.024×1.024 *pixels*) e da inferência (512×512 *pixels*), que faz os sintomas parecerem proporcionalmente maiores na inferência. Também contribuem para essa queda os sintomas cortados nas bordas dos recortes de treino e a proporção de recortes sem sintomas anotados, diferente da observada nas imagens completas.

O YOLO padrão apresentou desempenho intermediário, com precisão de $0,466$ e mAP50-95 de $0,315$. Embora preserve um equilíbrio razoável entre precisão e *recall*, seu desempenho é limitado em regiões pequenas e de baixo contraste, possivelmente devido ao redimensionamento das imagens de entrada para dimensões fixas, o que reduz a representatividade visual de sintomas sutis.

De modo geral, os resultados indicam que o desempenho em detecção depende da forma como a imagem é apresentada ao modelo tanto na inferência quanto no treinamento, sendo este último responsável pela maior queda de desempenho observada. Estratégias que preservam informações locais na inferência, como YOLO + SAHI, mostram-se mais adequadas para a detecção de sintomas foliares em cafezais, enquanto o treinamento com blocos fragmentados (*tiling*) tende a aumentar a sensibilidade às custas da precisão. Ainda assim, conclusões mais fortes dependem da consolidação do protocolo experimental, da validação da tabela de métricas e da inclusão de medidas de variabilidade entre execuções.

A principal limitação da estratégia YOLO + SAHI é o aumento do custo computacional durante a inferência. Enquanto o YOLO padrão processa a imagem completa em uma única passagem, a abordagem com SAHI subdivide a imagem em aproximadamente 126 recortes, exigindo múltiplas inferências para produzir o resultado final. Esse procedimento aumenta o tempo de processamento e pode limitar sua aplicação em dispositivos embarcados com restrições de memória, energia e capacidade computacional.

Esse custo, entretanto, pode ser mitigado por meio do ajuste dos parâmetros de recorte, como o tamanho das janelas e o grau de sobreposição, reduzindo o número total de regiões processadas. Além disso, a adoção de uma variante mais leve do YOLO pode diminuir a carga computacional da etapa de inferência. Assim, embora o uso de SAHI possa ser inviável para processamento embarcado em tempo real, ele permanece uma alternativa viável em fluxos de trabalho nos quais o VANT realiza apenas a aquisição das imagens e o processamento é executado posteriormente em um servidor com GPU dedicada.

5. Conclusões

Este trabalho avaliou estratégias de inferência voltadas à detecção de pequenos alvos para a identificação automática de regiões sintomáticas em cafezais a partir de imagens aéreas adquiridas por VANTs. Os resultados indicam que métodos baseados em recortes podem superar a inferência convencional sobre a imagem completa, desde que preservem detalhes locais e utilizem mecanismos adequados de agregação e filtragem das predições.

A estratégia YOLO + SAHI obteve melhor desempenho em precisão, F1-score, mAP50 e mAP50-95, sugerindo maior adequação para sintomas pequenos, pouco contrastantes e visualmente sutis. Por outro lado, o treinamento com *tiling* aumentou o *recall*, mas reduziu a precisão, indicando maior sensibilidade às custas de mais falsos positivos. Como a Estratégia 3 difere da Estratégia 2 apenas no treinamento, esse resultado indica que o procedimento de *tiling* adotado nessa etapa teve impacto relevante e negativo sobre o desempenho do modelo, reforçando que o desempenho depende tanto do treinamento quanto da inferência.

Como trabalhos futuros, pretende-se avaliar sistematicamente os parâmetros de *tiling* e SAHI, incluindo tamanho dos recortes, sobreposição e critérios de fusão das

detecções. Também será investigada uma estratégia de *adaptive color-aware tiling* combinada à normalização de iluminação, com o objetivo de definir recortes de forma adaptativa a partir de características como cor, contraste, textura e brilho, priorizando regiões mais informativas e reduzindo os efeitos de sombras, reflexos e variações de exposição.

Agradecimentos

Os autores gostariam de agradecer a Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG, subsídio APQ-01768-24), à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à Universidade Federal de Ouro Preto (UFOP/PROPPI) por apoiar o desenvolvimento do presente estudo.

Declaração sobre uso de Inteligência Artificial

Reconhecemos o uso de IA Generativa (ChatGPT da empresa OpenAI) para auxiliar na revisão gramatical e refinamento deste artigo, que foi devidamente revisado, e assumimos total responsabilidade por seu conteúdo.

Referências

Adiwijaya, N. O., Sarno, R., and Wijaya, D. R. (2024). Real time detection of coffee bean defects using yolo method and sahi (slicing aided hyper inference) framework. In 2024 Beyond Technology Summit on Informatics International Conference (BTS-I2C), pages 286–291.

Akyon, F. C., Altinuc, S. O., and Temizel, A. (2022). Slicing aided hyper inference and fine-tuning for small object detection. In 2022 IEEE international conference on image processing (ICIP), pages 966–970. IEEE.

Akyon, F. C., Cengiz, C., Altinuc, S. O., Cavusoglu, D., Sahin, K., and Eryuksel, O. (2021). SAHI: A lightweight vision library for performing large scale object detection and instance segmentation.

Albuquerque, L. D. and Guedes, E. B. (2024). Coffee plant leaf disease detection for digital agriculture. Journal on Interactive Systems, 15(1):220–233.

Bernardi, A. C. d. C., Bettiol, G. M., Grego, C. R., Andrade, R. G., Rabello, L. M., and Inamasu, R. Y. (2017). Ferramentas de agricultura de precisão como auxílio ao manejo da fertilidade do solo. Cadernos de Ciência & Tecnologia, 32(1/2):211–227.

Chaurasia, D. and Patro, B. (2023). Real-time detection of birds for farm surveillance using yolov7 and sahi. In 2023 3rd International Conference on Computing and Information Technology (ICCIT), pages 442–450.

Companhia Nacional de Abastecimento (2023). 4º Levantamento da Safra de Café. Accessed: 2024-06-06. ISSN: 2318-7913.

Food and Agriculture Organization of the United Nations (2023). Coffee. Accessed: 2024-06-03.

Goltz, G. A. M. (2011). Redes neurais artificiais em imagens para estimação da posição de um vant. Dissertação (mestrado em computação aplicada), Instituto Nacional de Pesquisas Espaciais, São José dos Campos, SP, Brasil.

Inamasu, R. Y., Bernardi, A. d. C., et al. (2014). Agricultura de precisão. BERNARDI, ACC; NAIME, JM; RESENDE, AV; BASSOI, LH, pages 21–33.

Jocher, G., Qiu, J., Liu, M., Lyu, S., Akyon, F. C., and Kalfaoglu, M. E. (2026). Ultralytics yolo26: Unified real-time end-to-end vision models.

Lamparelli, R. A. C. (2016). Agricultura de precisão. Agência Embrapa de Informação Tecnológica, Brasília.

Liang, X., Xiang, J., Qin, S., Xiao, Y., Chen, L., Zou, D., Ma, H., Huang, D., Huang, Y., and Wei, W. (2025). Small target detection algorithm based on sahi-improved-yolov8 for uav imagery: A case study of tree pit detection. Smart Agricultural Technology, 12:101181.

Ludermir, T. B. (2021). Inteligência artificial e aprendizado de máquina: estado atual e tendências. Estudos Avançados, 35:85–94.

Manfrim, A. C. (2022). Estudo e desenvolvimento de um sistema de detecção e contagem de gado em imagens aéreas utilizando redes neurais. Dissertação de mestrado, Universidade Estadual de Londrina.

Marengoni, M. and Stringhini, S. (2009). Tutorial: Introdução à visão computacional usando opencv. Revista de Informática Teórica e Aplicada, 16(1):125–160.

Mesquita, C. d., Rezende, J. E., Carvalho, J. S., Júnior, M., MORAES, N. C., Tavares, P., Carvalho, A., and Araújo, W. G. (2016). Manual do café: colheita e preparo (coffea arábica l.). Belo Horizonte: EMATER-MG, pages 1–56.

Muzammul, M., Algarni, A., Ghadi, Y. Y., and Assam, M. (2024). Enhancing uav aerial image analysis: Integrating advanced sahi techniques with real-time detection models on the visdrone dataset. IEEE Access, 12:21621–21633.

Pereira, D. C., Fernandes, L. F. S., and Fernandes, D. (2023). As vantagens da utilização de drones no agronegócio e na agricultura de precisão. Revista Alomorfia, 7(2):705–716.

Pinto, L., Silva, G., Negrão, A., Bueno, M., Freitas, V., and Silva, P. (2026). Detection of affected areas in coffee plantations using aerial drone imagery. In Brazilian Conference on Intelligent Systems. Springer.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 779–788.

Santana, L. S., do Santos, G. H. R., Araújo, G., Ferraz, S., dos Santos, L. M., Barbosa, B. D. S., and Bento, N. L. (2019). Treinamento de algoritmo para identificação de mudas de café por meio de imagens aéreas. X Simpósio de Pesquisa dos Cafés do Brasil.

Takano, W. H., Almeida, L. L., and da Silva, F. A. (2021). Detecção de animais bovinos utilizando imagens aéreas por meio de redes neurais. Colloquium Exactarum, pages 47–56.

Yang, X. and Guo, T. (2017). Machine learning in plant disease research. March, 31:1.

Ye, R., Gao, Q., Qian, Y., Sun, J., and Li, T. (2024). Improved yolov8 and sahi model for the collaborative detection of small targets at the micro scale: A case study of pest detection in tea. Agronomy, 14(5).