BACHAREL EM CIÊNCIA DA COMPUTAÇÃO

CLASSIFICAÇÃO AUTOMÁTICA DE ESTÁGIOS DE MATURAÇÃO DO TOMATE USANDO REDES NEURAIS CONVOLUCIONAIS YOLO

AMANDA BEATRIZ MARTINS ARAUJO

Rio Verde, GO

2025

INSTITUTO FEDERAL GOIANO - CAMPUS RIO VERDE
BACHAREL EM CIÊNCIA DA COMPUTAÇÃO

CLASSIFICAÇÃO AUTOMÁTICA DE ESTÁGIOS DE MATURAÇÃO DO TOMATE USANDO REDES NEURAIS CONVOLUCIONAIS YOLO

AMANDA BEATRIZ MARTINS ARAUJO

Trabalho de Conclusão de Curso apresentado ao Instituto Federal Goiano - Campus Rio Verde, como requisito parcial para a obtenção do Grau de Bacharel em Ciência da Computação.

Orientador: Prof. Dra. Heyde Francielle do Carmo França

Rio Verde, GO

3, 2025

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO

PARA DISPONIBILIZAR PRODUÇÕES TÉCNICO-CIENTÍFICAS NO REPOSITÓRIO INSTITUCIONAL DO IF GOIANO

Com base no disposto na Lei Federal nº 9.610, de 19 de fevereiro de 1998, AUTORIZO o Instituto Federal de Educação, Ciência e Tecnologia Goiano a disponibilizar gratuitamente o documento em formato digital no Repositório Institucional do IF Goiano (RIIF Goiano), sem ressarcimento de direitos autorais, conforme permissão assinada abaixo, para fins de leitura, download e impressão, a título de divulgação da produção técnico-científica no IF Goiano.

IDENTIFICAÇÃO DA PRODUÇÃO TÉCNICO-CIENTÍFICA

- ☐ Tese (doutorado)
- ☐ Dissertação (mestrado)
- ☐ Monografia (especialização)
- ☑ TCC (graduação)
- ☐ Artigo científico
- ☐ Capítulo de livro
- ☐ Livro
- ☐ Trabalho apresentado em evento
- ☐ Produto técnico e educacional - Tipo:

Nome completo do autor:
Amanda Beatriz Martins Araujo

Matrícula:
2018102201940376

Título do trabalho:
Classificação Automática de Estágios de Maturação do Tomate Usando Redes Neurais Convolucionais YOLO

RESTRIÇÕES DE ACESSO AO DOCUMENTO

Documento confidencial: ☑ Não ☐ Sim, justifique:

Informe a data que poderá ser disponibilizado no RIIF Goiano: 24 / 02 / 2025

O documento está sujeito a registro de patente? ☐ Sim ☑ Não

O documento pode vir a ser publicado como livro? ☐ Sim ☑ Não

DECLARAÇÃO DE DISTRIBUIÇÃO NÃO-EXCLUSIVA

O(a) referido(a) autor(a) declara:

• Que o documento é seu trabalho original, detém os direitos autorais da produção técnico-científica e não infringe os direitos de qualquer outra pessoa ou entidade;

• Que obteve autorização de quaisquer materiais inclusos no documento do qual não detém os direitos de autoria, para conceder ao Instituto Federal de Educação, Ciência e Tecnologia Goiano os direitos requeridos e que este material cujos direitos autorais são de terceiros, estão claramente identificados e reconhecidos no texto ou conteúdo do documento entregue;

• Que cumpriu quaisquer obrigações exigidas por contrato ou acordo, caso o documento entregue seja baseado em trabalho financiado ou apoiado por outra instituição que não o Instituto Federal de Educação, Ciência e Tecnologia Goiano.

Rio Verde, GO — Local

24 / 02 / 2024 — Data

gov.br Documento assinado digitalmente
AMANDA BEATRIZ MARTINS ARAUJO
Data: 06/03/2025 11:31:08-0300
Verifique em https://validar.iti.gov.br

Assinatura do autor e/ou detentor dos direitos autorais

Ciente e de acordo:

gov.br Documento assinado digitalmente
HEYDE FRANCIELLE DO CARMO FRANCA
Data: 24/02/2025 15:31:57-0300
Verifique em https://validar.iti.gov.br

Ficha de identificação da obra elaborada pelo autor, através do
Programa de Geração Automática do Sistema Integrado de Bibliotecas do IF Goiano - SIBi

Araujo, Amanda Beatriz Martins
A663c Classificação Automática de Estágios de Maturação do Tomate Usando Redes Neurais Convolucionais YOLO / Amanda Beatriz Martins Araujo. Rio Verde 2025.

17f. il.

Orientadora: Profª. Dra. Heyde Francielle do Carmo França.
Tcc (Bacharel) - Instituto Federal Goiano, curso de 0219201 - Bacharelado em Ciência da Computação - Integral - Rio Verde (Campus Rio Verde).
1. YOLO. 2. Classificação. 3. Segmentação. I. Título.

RESUMO

ARAUJO, Amanda. **Classificação Automática de Estágios de Maturação do Tomate Usando Redes Neurais Convolucionais YOLO.** 3, 2025. 17 f. Monografia – (Curso de Bacharel em Ciência da Computação), Instituto Federal Goiano - Campus Rio Verde. Rio Verde, GO.

RESUMO DO TRABALHO

O tomate é uma das frutas mais consumidas globalmente, mas sua colheita e classificação manual por cor ainda são processos trabalhosos e sujeitos a falhas humanas, como subjetividade e inconsistência entre avaliadores. Com o avanço da automação agrícola, métodos de visão computacional vêm sendo aplicados para melhorar a eficiência e a padronização na detecção de frutos. Este estudo utilizou redes neurais convolucionais YOLO para classificar tomates cereja em três estágios de maturação (verde, intermediário e maduro), a partir de uma base de imagens rotuladas manualmente. Nesse trabalho, foram testadas as versões YOLO v8, v9, v10 e v11, comparando desempenho em precisão, recall, mAP50 e tempo de treinamento. O YOLO v9 obteve a maior precisão, enquanto o YOLO v11 destacou-se pela velocidade de treinamento e bom recall. Além disso, a segmentação mostrou-se mais eficaz do que as caixas delimitadoras para anotação das imagens. Considerando todas as métricas avaliadas, o YOLO v11 foi apontado como a melhor opção para a classificação automatizada de tomates, oferecendo um equilíbrio entre eficiência e desempenho.

Palavras-chave: tomate cereja, YOLO, Segmentação .

ABSTRACT

ARAUJO, Amanda. **Automatic Classification of Tomato Ripening Stages Using YOLO Convolutional Neural Networks.** 3, 2025. 17 f. Trabalho de Conclusão de Curso – Bacharel em Ciência da Computação, Instituto Federal Goiano - Campus Rio Verde. Rio Verde, GO, 3, 2025.

The tomato is one of the most widely consumed fruits globally, but its manual harvesting and color-based classification remain labor-intensive processes prone to human errors, such as subjectivity and inconsistency among evaluators. With advancements in agricultural automation, computer vision methods have been applied to enhance efficiency and standardization in fruit detection. This study utilized YOLO convolutional neural networks to classify cherry tomatoes into three ripening stages (green, intermediate, and ripe) based on a manually labeled image dataset. YOLO versions v8, v9, v10, and v11 were tested, comparing performance in terms of precision, recall, mAP50, and training time. YOLO v9 achieved the highest precision, while YOLO v11 stood out for its training speed and strong recall. Additionally, segmentation proved to be more effective than bounding boxes for image annotation. Considering all evaluated metrics, YOLO v11 was identified as the best option for automated tomato classification, offering a balance between efficiency and performance.

Keywords:Cherry tomato. YOLO. Segmentation.

SERVIÇO PÚBLICO FEDERAL
MINISTÉRIO DA EDUCAÇÃO
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA GOIANO

Ata nº 2/2025 - GEPTNM-RV/DE-RV/CMPRV/IFGOIANO

ATA DE DEFESA DE TRABALHO DE CURSO

Ao(s) **06** dia(s) do mês de fevereiro de 2025, às 14 horas e 00 minutos, reuniu-se a banca examinadora composta pelos docentes: Heyde Francielle do Carmo (orientadora), Fábio Montanha Ramos(membro), Fábia Barbosa da Silva(membro), para examinar o Trabalho de Curso intitulado "*Classificação Automática de Estágios de Maturação do Tomate Usando Redes Neurais Convolucionais YOLO*" do(a) estudante Amanda Beatriz Martins Araujo, Matrícula nº 2018102201940376 do Curso de Bacharelado em Ciência da Computação do IF Goiano – Campus Rio Verde. A palavra foi concedida ao(a) estudante para a apresentação oral do TC, houve arguição do(a) candidato pelos membros da banca examinadora. Após tal etapa, a banca examinadora decidiu pela APROVAÇÃO da estudante. Ao final da sessão pública de defesa foi lavrada a presente ata que segue assinada pelos membros da Banca Examinadora.

(Assinado Eletronicamente)

Heyde Francielle do Carmo

Orientador(a)

(Assinado Eletronicamente)

Fábio Montanha Ramos

Membro

(Assinado Eletronicamente)

Fábia Barbosa da Silva

Membro

Observação:

( ) O(a) estudante não compareceu à defesa do TC.

Documento assinado eletronicamente por:

- **Heyde Francielle do Carmo Franca**, **PROFESSOR ENS BASICO TECN TECNOLOGICO**, em 06/02/2025 15:25:47.
- **Fabia Barbosa da Silva**, **2020202341360019 - Discente**, em 07/02/2025 07:32:32.
- **Fabio Montanha Ramos**, **PROFESSOR ENS BASICO TECN TECNOLOGICO**, em 10/02/2025 17:03:40.

Este documento foi emitido pelo SUAP em 06/02/2025. Para comprovar sua autenticidade, faça a leitura do QRCode ao lado ou acesse https://suap.ifgoiano.edu.br/autenticar-documento/ e forneça os dados abaixo:

Código Verificador: 672896
Código de Autenticação: ad8969b9df

INSTITUTO FEDERAL GOIANO
Campus Rio Verde
Rodovia Sul Goiana, Km 01, Zona Rural, 01, Zona Rural, RIO VERDE / GO, CEP 75901-970
(64) 3624-1000

SUMÁRIO

1 – INTRODUÇÃO **1**

2 – TRABALHOS RELACIONADOS **3**

3 – MATERIAL E MÉTODOS **4**

4 – RESULTADOS E DISCUSSÕES **9**

Referências **15**

1 INTRODUÇÃO

A Organização das Nações Unidas para Agricultura e Alimentação (FAO) destaca a importância do tomate, que é a segunda fruta mais consumida globalmente, com mais de 180 milhões de toneladas produzidas anualmente. O consumo varia muito, e em algumas regiões excede 100 kg per capita ((PERALTA; SPOONER, 2004)). Apesar da grande demanda, a colheita do tomate é uma tarefa que necessita de muita mão de obra, assim como a medição manual de informações fenotípicas ((MINERVINI; SCHARR; TSAFTARIS, 2015)). Nos últimos anos, tem havido interesse na automação de processos agrícolas, como a colheita ((BAC et al., 2014)), poda ((PAULIN et al., 2015)), e pulverização localizada((OBERTI et al., 2016)). Especialmente, na colheita, esse interesse é evidenciado pelo crescente desenvolvimento de métodos de análise de imagens e visão computacional, particularmente na detecção de frutos ((AFONSO et al., 2020)).

Nessa fase, a precisão na classificação dos estágios de maturação é fundamental, pois impacta diretamente no Controle de Qualidade. No processo, a cor da casca da fruta é o principal indicador de maturação. Os tomates, na fase de maturidade de mesa, caracterizados pela cor vermelha (índice $a^*/b^* \geq 0{,}85$), exibem maior valor nutricional, incluindo uma maior concentração de antioxidantes e de ácido ascórbico ((WEINGERL; UNUK, 2015)).

A detecção de frutos por meio da visão computacional é uma forma rápida e não destrutiva de medição, sendo útil para classificar o fruto segundo seu estágio de maturação, bem como analisar outras características da planta ((CHOUDHURY; SAMAL; AWADA, 2019)). Considerando essa vantagem, neste trabalho foi criada uma base de dados de imagem de tomates do tipo grape para treinamento de uma rede neural convolucional YOLO, afim de classificar os tomates em três estágios de maturação. As imagens que compoem a base são classificadas em maduro, verde e intermediário, sendo essas três classes, os estágios de maturação do fruto, as classes estão ilustradas na Figura 1. Observe que na imagem os tomates presentes são de outra espécie, não sendo tomate Cereja. Na imagem, a primeira classe maduro (caixa delimitadora com borda roxa), a segunda classe intermediário (caixa delimitadora com borda vermelha) e a última classe, verde (caixa delimitadora com borda amarela). Os frutos imaturos possuem uma coloração esverdeada semelhante aos galhos e folhas, enquanto os maduros exibem uma cor vermelha vibrante. Durante o processo de maturação, a cor dos tomates muda gradualmente de esverdeada para vermelha vibrante, com o grau de coloração se intensificando e a área colorida se expandindo de pequena para grande.

Este trabalho propõe o desenvolvimento de um modelo baseado na arquitetura YOLO (You Only Look Once) para detecção e classificação automática dos estágios de maturação do tomate grape. O objetivo é oferecer uma solução computacionalmente eficiente e precisa, capaz de realizar análises em tempo real, o que é fundamental para aplicações agrícolas como colheita automatizada e monitoramento de qualidade

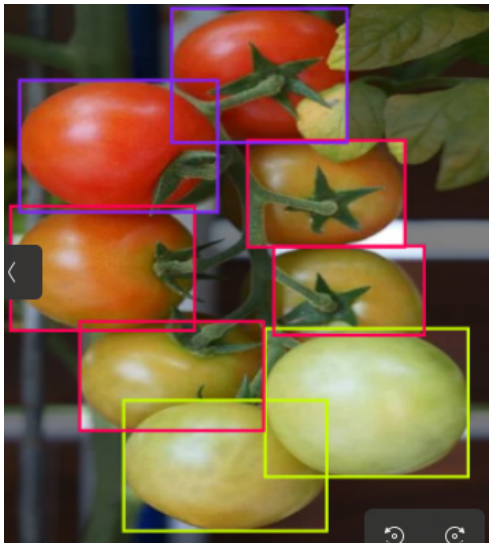

Figura 1 – As três classes de maturação de frutos, foto proveniente da base de dados pública, com nova rotulação.Fonte: (GUILHERMEHC, 2024)

2 TRABALHOS RELACIONADOS

Modelos desenvolvidos com aprendizado profundo e redes neurais convolucionais (CNN), alcançaram alta precisão, algumas vezes superando o nível humano em tarefas bem conhecidas, como classificação de imagens, detecção de objetos e segmentação de instâncias. Se tratando de tomate cereja, há trabalhos que utilizam a Yolov5m para detecção de frutos ((WANG et al., 2023)) e identificação de anomalias nos frutos ((WANG; LIU, 2021)).

Há também a pesquisa de (CAI et al., 2024), que utiliza uma versão aprimorada da rede Yolov7-tiny, aprimorando a precisão da detecção de tomates cereja em estufas. Este modelo alcançou taxas de coleta bem-sucedidas de mais de 80% durante os testes. No trabalho de foi utilizado a arquitetura Yolov4-tiny com FENB. A abordagem categoriza os tomates cereja em várias classes com base na oclusão, atingindo uma precisão média (mAP) de 94,72% durante a detecção noturna (ZHANG et al., 2022).

Também existe aplicações de detecção de tomates cereja em tempo real, como no trabalho de (ZENG et al., 2023), utilizando YOLOv5s. Contudo, os tomates cereja brasileiros possuem características agronômicas diferentes, refletindo a diversidade e adaptabilidade dessa cultura no país. Há destaque comercial para as cultivares SCS375 Kaiçara, Cereja Vermelho e Flavor Top ((PEIL; NETO; ROMBALDI, 2014)). Nesse trabalho, o objeto de estudo é o tomate Cereja Vermelho.

3 MATERIAL E MÉTODOS

Ter uma base de dados de alta qualidade é fundamental para treinar modelos eficazes de classificação. A baixa qualidade dos dados pode levar a previsões tendenciosas ou imprecisas do modelo ((MAO; XU; SUAREZ, 2023)). Para construir a base de dados, inicialmente feita uma curadoria de dados sobre um banco de imagens público ((GUILHERMEHC, 2024)) de tomate. Inicialmente, o conjunto de dados público continha um total de 345 imagens, com uma única classe denominada ”tomato”e 3.415 tomates devidamente anotados, com anotações do tipo Caixa Delimitadora. Essa base pública continha tomates de diferentes espécies, não sendo apenas de tomate cereja.

Nesses dados, através da técnica de data augmentation, foi acrescido à base novas imagens. A captura das imagens foi feita em um sistema de agricultura vertical controlado de tomates do tipo grape, no Instituto Federal Goiano - Campus Rio Verde. Essas imagens foram capturadas através de um smartphone, em dois dias distintos, no início da manhã e no final da tarde. A base de dados formada pelas novas imagens coletadas, juntamente com a base de dados pública, nesse trabalho, foi administrada através do Roboflow((ROBOFLOW, 2024)).

O Roboflow serve como uma ferramenta para rotular bancos de dados, particularmente no contexto de detecção de objetos e aprendizado de máquina. Ele facilita a criação de diversos conjuntos de dados por meio de técnicas de rotulagem automatizadas e semiautomáticas, aumentando a eficiência e a precisão da preparação de dados para o treinamento de modelos ((IWAMURA; TSUKADA; KISE, 2013)). Nesse trabalho, foram comparadas duas formas de marcação: Caixa Delimitadora e Segmentação. A Figura 2 ilustra as diferenças visuais dos dois tipos de marcação. A Caixa Delimitadora é o tipo de anotação, que fornece as coordenadas que delimitam a área retangular onde o objeto está presente ((RAJCHL et al., 2016)). Enquanto que a anotação do tipo Segmentação, ou Máscara de Segmentação é delimitado as magens em pixels, em que cada pixel recebe um rótulo correspondente a um objeto ou plano de fundo específico ((RHEE; KOO; CHO, 2022)). Há casos em que o uso de segmentação permite que os modelos se concentrem

(a) Caixa Delimitadora

(b) Segmentação

Figura 2 – Tipos de anotações no banco de dados de imagens. Fonte: próprio autor

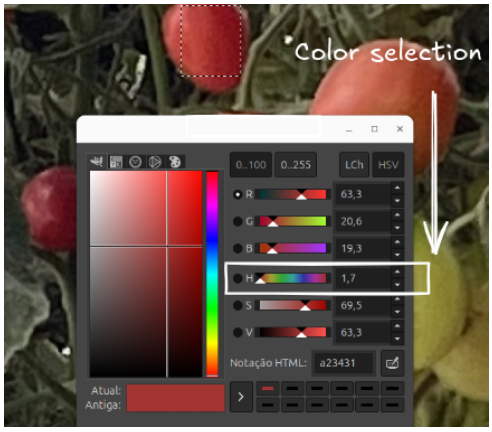

Figura 3 – Seleção da cor média do tomate usando software GIMP. Fonte: próprio autor

em áreas relevantes, alcançando alta precisão mesmo com dados limitados ((KARIRYAA, 2020)).

Para rotular as imagens foi utilizado o software GIMP, para aferir com precisão, a cor média do fruto. Na marcação, foi eliminando as tonalidades escuras, a Figura 3 ilustra o uso do software GIMP para identificação da classe do tomate em evidência. Nessa análise, foi levado em conta, na escala no sistema de cores HSV, considerando apenas a matiz (hue), desconsiderando o brilho e a saturação. A medição da matiz é exibida no componente H da seleção de cores. Como mostrado na tabela 1, a matiz(H) de 0° a 25°, a classe delimitada foi a classe madura; de 25° a 45°, foi escolhida a classe intermediária; e a faixa de 45° a 120° refere-se a tomates que estão verdes. Na Figura 3, a medição do H é 1.7 °, sendo portanto: maduro.

Matiz	**Classe**
0° a 25°	red_tomato
25° a 45°	intermediary_tomato
45° a 120°	green_tomato

Tabela 1 – Relação entre Matiz e classe do tomate

A Figura 4 indica a quantidade absoluta de cada classe em cada um dos conjunto de dados construídos, demonstrando a evolução na distribuição das classes. A primeira rotulação feita, a partir da base de dados pública ((GUILHERMEHC, 2024)) constitui 256 imagens que foram rotuladas. Já a segunda base rotulada (de 258 imagens), foi classificada utilizando o software GIMP e a matiz. Nessa base, ocorre a diminuição da frequência nas classes intermediário e verde. Nas bases de dados de 256 e 258 imagens, as classes de tomate intermediário e verde estavam subrepresentadas, o que poderia causar risco de underfitting no treinamento dos modelos. Porém durante a construção das novas bases de dados, com mais imagens, essas classes foram mais representativas nos treinamentos.

Após construir o banco de dados de imagem, e também rotular segundo as três classes analisadas, o treinamento da rede foi feito no Google Colab. O Google Colab é uma ferramenta baseada em nuvem que permite a execução de código Python diretamente no navegador. É amplamente utilizado para análise de dados, aprendizado de máquina e outras tarefas computacionais intensivas ((BASSETT; KUMARAN, 2008)). O tipo de rede neural escolhido foi Rede Neural Convolucional, um tipo de rede neural que divide a imagem em células visuais, evitando problemas como o overfitting, onde ocorre falta

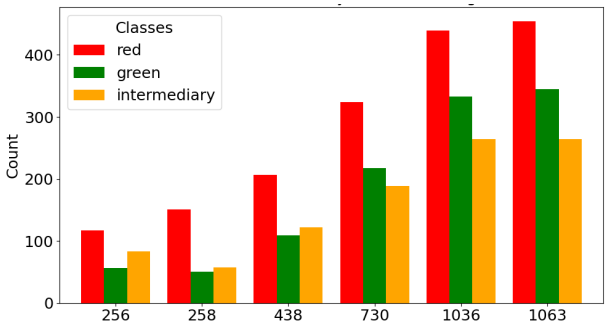

Figura 4 – Distribuição de classes. Fonte: próprio autor

de generalização no aprendizado ((KRIZHEVSKY; SUTSKEVER; HINTON, 2012)), e a arquitetura escolhida foi a YOLO ((ULTRALYTICS, 2023)).

A escolha da arquitetura YOLO se deu por conta de ser o estado da arte, e pela sua popularidade em problemas de detecção de objetos, como em detecção de tomates salada ((YANG et al., 2023)), detecção em imagens de sensoriamento remoto((SHEN; LANG; SONG, 2023)), e até mesmo na detecção de comportamento de estudantes em sala de aula ((CHEN; ZHOU; JIANG, 2023)). De acordo com o criador da rede YOLO ((REDMON et al., 2016)), a YOLO se destaca por sua simplicidade, pois utiliza apenas uma rede convolucional para prever simultaneamente várias caixas delimitadoras e as probabilidades de classe correspondentes. Diferentemente dos métodos tradicionais de detecção de objetos, a YOLO treina em imagens completas e otimiza diretamente o desempenho da detecção. Essa abordagem unificada apresenta várias vantagens, sendo a velocidade um dos principais benefícios. Ao formular a detecção como um problema de regressão, a YOLO evita a necessidade de pipelines complexos. Durante o teste, a rede neural é aplicada diretamente a uma nova imagem para realizar as previsões de detecção.

A versão 8 utiliza uma rede totalmente convolucional projetada para prever caixas delimitadoras e probabilidades de classe de objetos em tempo real ((REDMON et al., 2016)). Já o YOLOv9 introduziu métodos inovadores, como as Informações de Gradiente Programáveis (PGI) e a Rede de Agregação de Camadas Eficientes Generalizadas (GELAN), para resolver de maneira eficaz problemas relacionados à perda de informações e à eficiência computacional ((WANG; YEH; LIAO, 2024)). O YOLOv10 apresenta duas melhorias principais em relação às versões anteriores: um sistema de Atribuições Duplas Consistentes para Treinamento Sem NMS e um Design de Modelo Focado em Eficiência e Precisão, visando otimizar o desempenho geral ((WANG et al., 2024)). O YOLOv11, a versão mais recente da série de algoritmos YOLO, aprimora a extração de características com uma arquitetura avançada de backbone e neck, enquanto melhora a velocidade e eficiência por meio de designs e pipelines de treinamento otimizados, equilibrando precisão e desempenho ((ULTRALYTICS,)).

Parâmetro	**Valor**
Learning rate	0.01
Batch size	8
Número de iterações(Épocas)	1000
Momentum	0.937
Weight decay	0.0005

Tabela 2 – Parâmetros de treinamento dos modelos YOLO. Fonte: próprio autor

Para o treinamento dessas diferentes versões da YOLO, foram utilizados os parâmetros listados na Tabela 2. Esses parâmetros foram escolhidos para otimizar o desempenho do modelo. A taxa de aprendizado foi definida como 0,01, permitindo atualizações moderadas nos pesos do modelo durante o treinamento, garantindo a convergência e evitando o overshooting. O tamanho do batch foi ajustado para 8, equilibrando a carga computacional e proporcionando generalização. O modelo foi treinado por 1000 épocas, garantindo que a rede tivesse tempo suficiente para aprender com os dados. Além disso, foi utilizado um valor de momentum de 0,937 para acelerar a convergência, incorporando uma fração da atualização anterior dos pesos, ajudando o modelo a escapar de mínimos locais. Para prevenir overfitting, foi aplicada uma decadência de peso de 0,0005, penalizando pesos grandes e incentivando modelos mais simples e generalizáveis.

Nesse trabalho, são comparados as diferentes versões da YOLO. e para avaliar o desempenho da rede, foi utilizado a precisão, Recall e mAp50. A seguir, são apresentadas as definições e explicações dessas métricas.

A precisão é uma métrica que avalia a capacidade do modelo de fazer previsões corretas entre todas as previsões realizadas ((PADILLA; NETTO; SILVA, 2020)). Em outras palavras, mede a proporção de verdadeiros positivos (*True Positives*, TP) em relação ao total de objetos classificados como positivos pelo modelo (verdadeiros positivos + falsos positivos). A fórmula para calcular a precisão é dada por:

$$\text{Precisão} = \frac{TP}{TP + FP}$$

Onde *TP* (True Positives) representa objetos corretamente identificados como positivos e *FP* (False Positives) representa objetos incorretamente identificados como positivos. Uma precisão mais alta indica que o modelo está fazendo menos previsões erradas como positivas, o que reflete maior confiança nas suas detecções.

O recall, também conhecido como sensibilidade, é uma métrica que avalia a capacidade do modelo de identificar todos os objetos relevantes em um conjunto de dados ((PADILLA; NETTO; SILVA, 2020)). Em outras palavras, mede a proporção de verdadeiros positivos em relação ao total de objetos presentes no conjunto de dados, incluindo tanto os verdadeiros positivos quanto os falsos negativos. A fórmula para calcular o recall é dada por:

$$\text{Recall} = \frac{TP}{TP + FN}$$

Onde *TP* (True Positives) representa objetos corretamente identificados como positivos e *FN* (False Negatives) representa objetos que deveriam ser identificados como positivos, mas foram classificados como negativos. Um valor de recall mais alto indica que o modelo está identificando corretamente a maioria dos objetos, minimizando as falhas de detecção. A métrica mAP, ou *mean Average Precision* a um *Intersection over Union* (IoU) de 0.5, fornece uma medida global de desempenho do modelo de detecção de objetos ((HENDERSON; FERRARI, 2017)). O mAP é a média da precisão para várias classes de objetos, calculada a partir da curva precisão-recall. O valor de mAP é calculado para diferentes limiares de confiança, e mAP50 refere-se ao valor médio da precisão considerando um limiar de 0.5 para o IoU, ou seja, se a sobreposição entre a caixa predita e a caixa real for superior a 50%, a predição é considerada correta.A fórmula para o cálculo de mAP50 é:

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^{N} \text{AP}_i$$

Onde AP_i é a precisão média para a classe i e N é o número total de classes. Quanto maior o valor de mAP50, melhor o desempenho global do modelo, com maior capacidade de detectar e classificar objetos com precisão, considerando um limiar de IoU de 0.5.

O F1-score, que é a média harmônica de precisão e recall, fornece uma métrica balanceada que combina tanto precisão quanto recall. Essa métrica é útil para otimizar o desempenho de detecção onde ambas as métricas são importantes((JUNIOR et al., 2022)). A Curva de Confiança F1 ilustra a relação entre a pontuação F1 e o limite de confiança do modelo.

$$F1 = \frac{2 \cdot \text{precisão} \cdot \text{recall}}{\text{precisão} + \text{recall}}$$

4 RESULTADOS E DISCUSSÕES

Para todos os treinamentos que serão apresentados os resultados nesta sessão, foram utilizados a Segmentação e não Caixas Delimitadoras. São avaliados tempo de treinamento e métricas precisão, recall, F1-score e mAP50, essas métricas estão dispostas na Tabela 3 . O tempo de treinamento em aprendizado de máquina é um fator crítico que influencia a eficiência e a eficácia da implantação de modelos, especialmente em ambientes com recursos limitados. O gráfico da Figura 5 dispõe o tempo de treinamento para cada versão da YOLO, em todos os modelos que foram construídos. Obviamente, à medida que o número de imagens aumenta, o tempo necessário para treinar o modelo também aumenta. Isso é esperado, pois o modelo precisa processar mais dados para aprender as características relevantes das imagens.

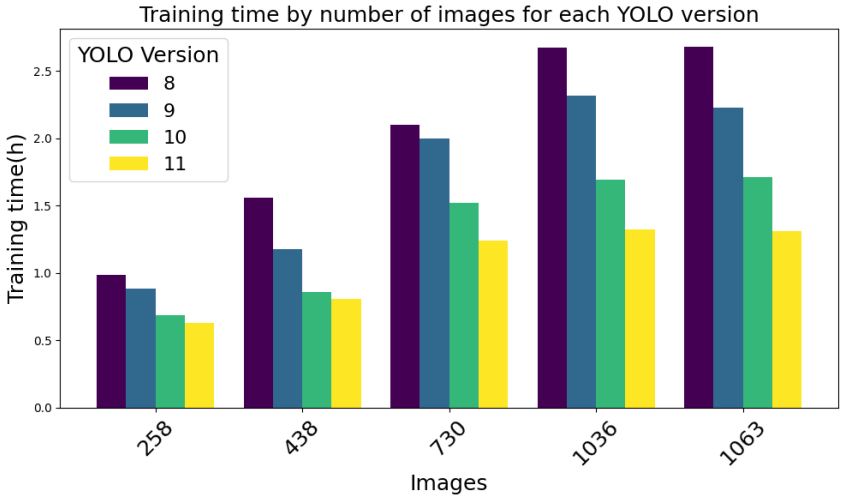

Figura 5 – Tempo de treinamento para cada versão da YOLO. Fonte: próprio autor

A versão 8 da YOLO apresenta o tempo de treinamento mais longo, pois não implementa as otimizações das novas versões da arquitetura. A versão YOLO v11, apresenta um tempo de treinamento mais rápido em comparação com as outras versões, especialmente para conjuntos de dados maiores. As versões YOLO v9 e v11 têm tempos de treinamento semelhantes, um pouco mais lentos que o YOLOv11, mas ainda competitivos.

Com relação às métricas que avaliam o desempenho do modelo, a precisão é uma métrica que indica a capacidade do modelo de classificar corretamente os objetos nas imagens. O desempenho dos modelos nessa métrica está ilustrado na Figura 6. A analise mostra uma tendência de aumento da precisão à medida que o número de imagens presentes na base de dados aumenta para todas as versões do YOLO. Isso é esperado, pois modelos de aprendizado de máquina geralmente se beneficiam de mais dados para aprender padrões mais complexos

A versão 8 apresenta uma alta precisão inicial e mantém um desempenho consistente ao longo do aumento do número de objetos de treinamento, porém no maior conjunto de dados, decai significativamente a precisão. A YOLO v9 começa com uma precisão um pouco menor que o YOLOv8, mas apresenta um crescimento mais acentuado com o aumento do número de objetos, alcançando um desempenho similar ao YOLOv8 para a base de dados de 730 imagens, tendo desempenho competitivo no conjunto de dados com

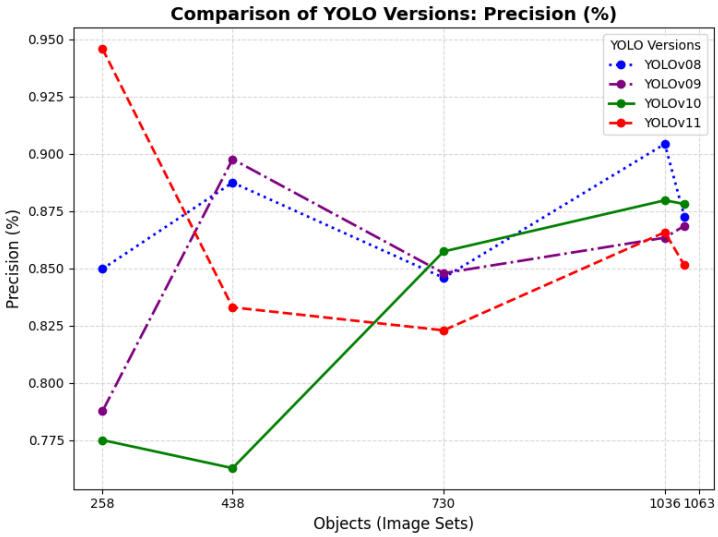

Figura 6 – Gráfico de Precisão. Fonte: próprio autor

maior quantidade de imagens. Considerando a versão v10, o modelo tem uma curva de aprendizado mais suave, com um aumento gradual da precisão à medida que o número de objetos aumenta, e no conjunto de dados com maior quantidade de imagens, o modelo YOLOv10 se saiu melhor na precisão. Na YOLOv11 ocorre uma precisão inicial mais baixa e um crescimento mais lento em comparação com as outras versões, e desempenho consideravelmente pior. Essa análise indica que, para a detecção de tomate cereja com segmentação, a versão do YOLOv10 produziu resultados mais precisos que as demais versões analisadas da YOLO.

Considerando a capacidade dos modelos em identificar todos os objetos relevantes em uma imagem, esse conceito é medido através do Recall. Para todos as versões da YOLO analisadas, a Figura 7 dispõe graficamente essa métrica. Assim como ocorre na precisão, com o aumento no número de imagens de treinamento, o Recall também aumenta, diminuindo o risco de underfitting, pois com mais dados de treinamento, os modelos são capazes de identificar um número maior de objetos. A versão 11 tem uma recall inicial mais baixa e um crescimento mais lento em comparação com as outras versões. No entanto, para o maior conjunto de dados, alcança um recall competitiva, melhorando muito para a base de dados com maior quantidade de imagens. A versão 9 começa com uma recall mais baixa, mas apresenta um crescimento mais acentuado com o aumento do número de objetos, alcançando o melhor desempenho em Recall na maior base de dados, melhor do que os demais modelos. Para a base com mais imagens, as versões 8 e 10 tiveram pior desempenho.

Após analisar essas duas métricas, precisão e Recall, será analisada a mAP50, que é uma métrica que combina precisão e recall, fornecendo uma avaliação mais completa do desempenho de um modelo de detecção de objetos. Assim como nas demais métricas apresentadas, o conjunto de dados maior oferece menos risco de underfitting. Nessa métrica a versão YOLOv8 e YOLOv10 não tiveram bom desempenho, se comparados com as demais versões. Para a base de dados com 1064 imagens de tomates, as demais versões da YOLO tiveram desempenho semelhante em mAP50, contudo a YOLO v09 é a que possui aprendizado mais suave conforme novas imagens foram acrescentadas ao modelo.

A YOLO v10 tem uma mAP50 inicial mais baixa e um crescimento mais lento em comparação com as outras versões, mas na base de dados com mais imagens não consegue um bom desempenho. Como mAP50 é uma métrica que resume o desempenho

Tabela 3 – Tabela com as métricas de cada versão da YOLO

Versão	**Instâncias**	**Precision**	**Recall**	**mAP50**	**F1-Score**	
					% acerto	**confiança**
8	258	0.84976	0.84852	0.91367	0.82	0.75
9	258	0.78769	0.62335	0.76450	0.72	0.22
10	258	0.77509	0.84394	0.80749	0.88	0.83
11	258	0.94582	0.86289	0.92315	0.90	0.78
8	438	0.88745	0.70489	0.81739	0.85	0.77
9	438	0.89764	0.75476	0.89569	0.88	0.62
10	438	0.76285	0.80047	0.80697	0.83	0.54
11	438	0.83301	0.75647	0.86173	0.83	0.63
8	730	0.84588	0.72709	0.82100	0.81	0.58
9	730	0.84790	0.78943	0.79324	0.84	0.58
10	730	0.85739	0.79155	0.81445	0.80	0.30
11	730	0.82296	0.83700	0.85835	0.84	0.52
8	1036	0.90430	0.82579	0.91070	0.86	0.76
9	1036	0.86339	0.83390	0.85478	0.84	0.21
10	1036	0.87969	0.81353	0.85935	0.83	0.34
11	1036	0.87487	0.83056	0.86365	0.85	0.53
8	1063	0.87265	0.77307	0.83057	0.85	0.51
9	1063	0.86854	0.83859	0.85760	0.87	0.24
10	1063	0.87809	0.77055	0.81878	0.88	0.83
11	1063	0.85142	0.82768	0.85364	0.84	0.72

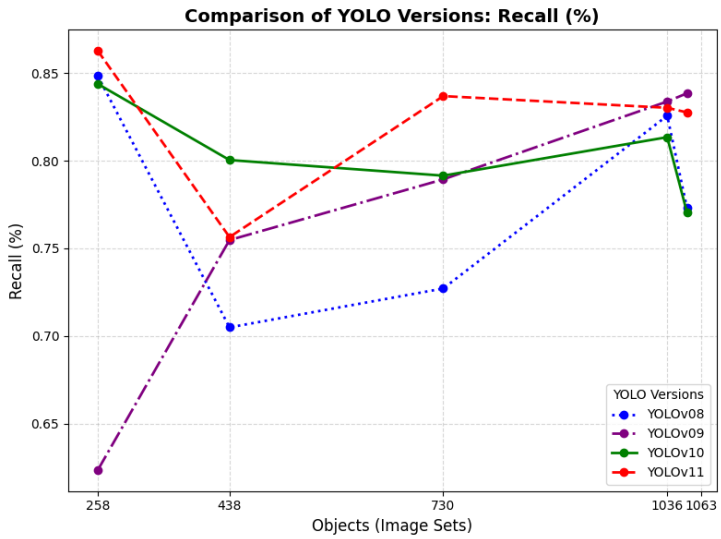

Figura 7 – Gráfico de Recall. Fonte: próprio autor

global dos modelos, de forma geral, o modelo que saiu melhor para as bases de dados com mais imagens, é a YOLO v11. Cada versão do YOLO possui uma arquitetura diferente, com diferentes números de camadas, filtros e outras características que podem influenciar o desempenho, e a YOLOv11 foi a que teve melhor tempo de treinamento e melhor desempenho global. A figura 9 apresenta a comparação dos resultados de detecção dos modelos treinados utilizando a YOLOv11 com 258, 438, 730, 1036 e 1063 instâncias.

Considerando o modelo com YOLO v11 para 1063 imagens, o modelo que foi

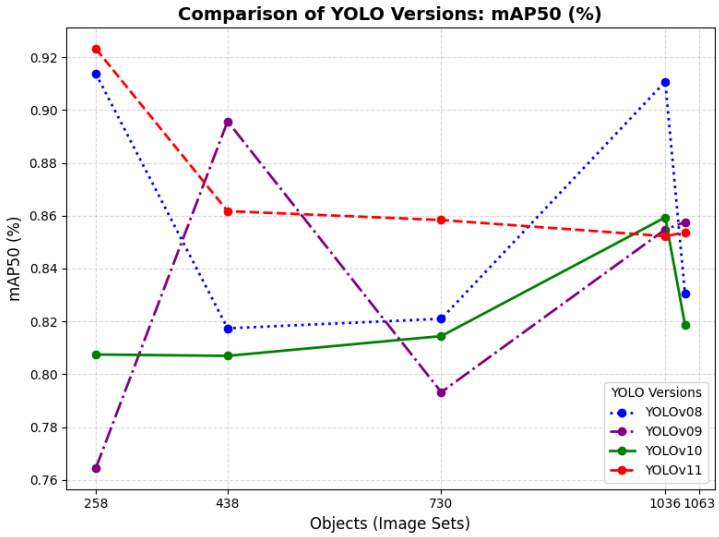

Figura 8 – Gráfico mAP50. Fonte: próprio autor

Figura 9 – Comparação da YOLOv11 na detecção de tomates com modelos treinados em 258, 438, 730, 1036 e 1063 instâncias.

melhor avaliado. A Figura 10 ilustra todas as métricas com relação ao treinamento do modelo. As grandezas train/box_loss, train/cls_loss, train/dfl_loss representam as perdas (losses) associadas à detecção do polígono de segmentação, classificação da classe e perda focal, respectivamente.A tendência geral para essas métricas é de decréscimo, isso indica que o modelo está aprendendo a realizar essas tarefas com mais precisão ao longo do treinamento. Flutuações podem ocorrer, especialmente no início do treinamento, à medida que o modelo se ajusta aos dados. Já as métricas. Já a mAP50 (mean Average Precision) é uma métrica que combina precisão e recall, fornecendo uma avaliação mais completa do desempenho do modelo. A mAP50-95 considera um intervalo de IOUs de 0.5 a 0.95, oferecendo uma visão mais abrangente da performance do modelo em diferentes níveis de

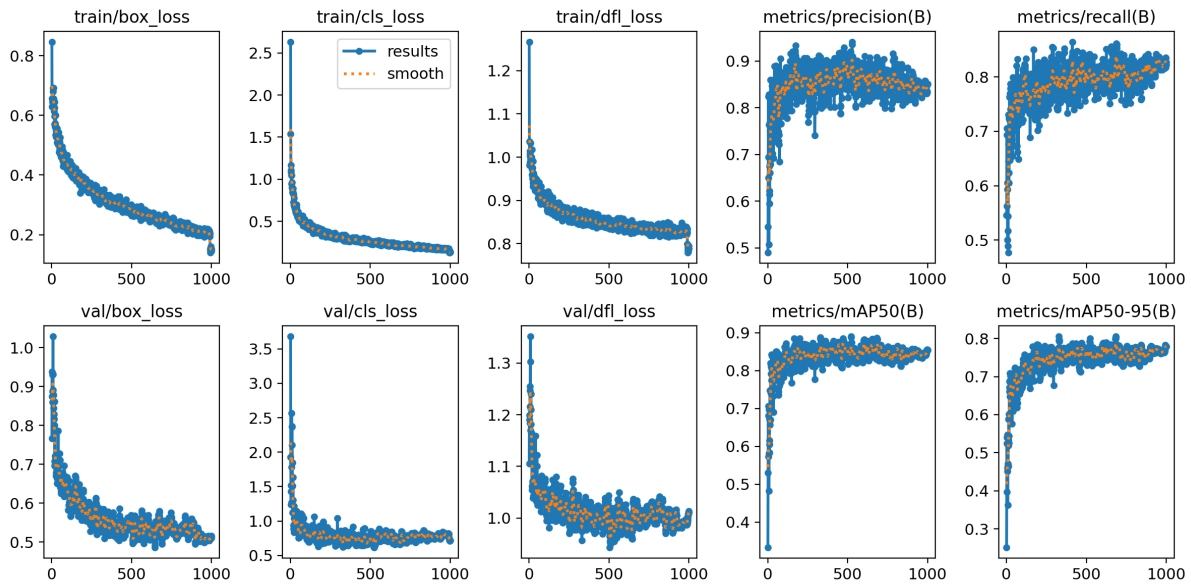

Figura 10 – Métricas para YOLO v11 1063 imagens. Fonte: próprio autor

sobreposição entre as caixas previstas e as verdadeiras. O aumento da mAP50 indica que o modelo está melhorando na detecção de objetos.

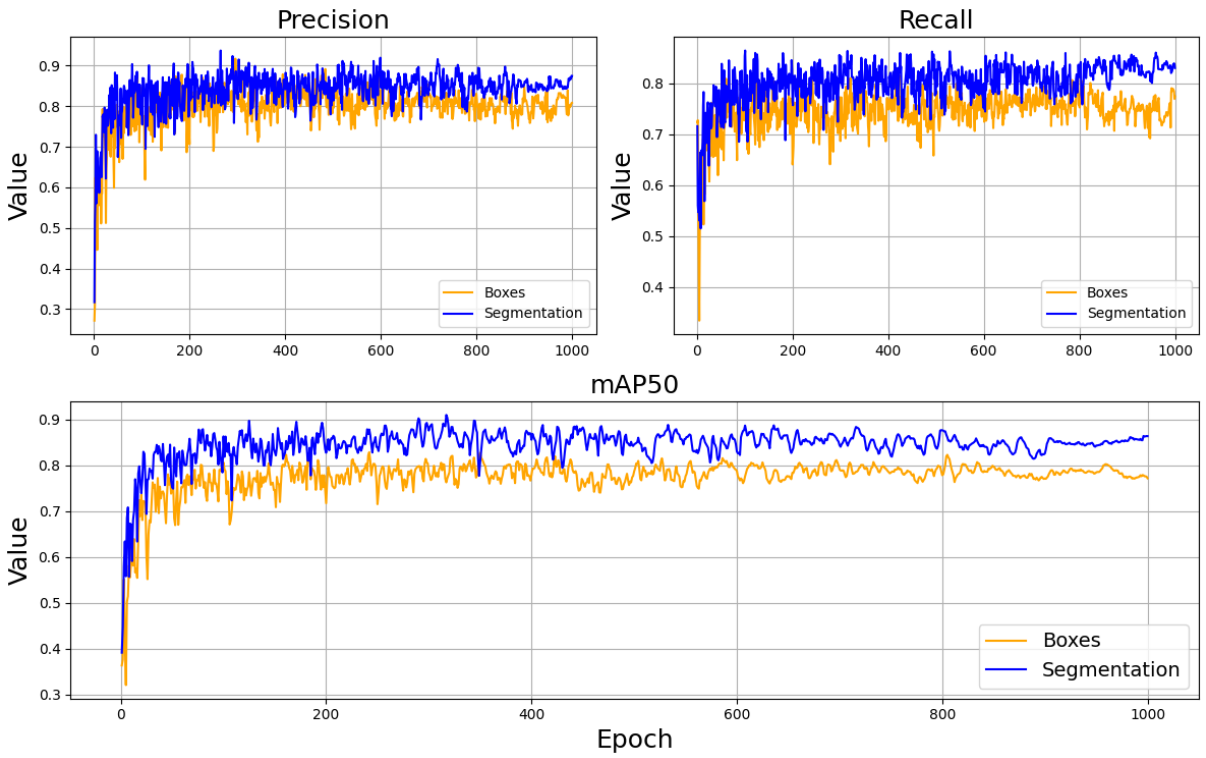

Figura 11 – Gráfico comparativo de Boxes e Segmentação

Após avaliar o modelo que possui melhor desempenho quando o tipo de rotulação é segmentação, será agora discutido o uso de segmentação versus caixa delimitadora. Será utilizado um modelo com YOLO v11 e 1036 imagens para treinamento para duas bases de dados diferentes, a fim de comparar o uso desses dois tipos de rotulação. Essa base de dados é uma base com quantidade média de imagens, e cada objeto possui o tamanho de

600 pixels de largura e 600 pixels de comprimento.

Analisando a Figura 11, que denota a convergência dos modelos nas métricas precisão, recall e mAP50, o tipo de anotação é superior a anotação usando caixas, pois convergiu para valores absolutos maiores em cada uma das métricas. Se tratando da precisão, ambo os os modelos iniciam com pouca precisão, e conforme o modelo é treinado. a precisão aumenta. O modelo de caixas delimitadoras é menos competitivo na precisão que o modelo que utilizou segmentação. Porém, essa diferença não é tão grande, o modelo de caixas teve precisão de 0.808, enquanto que o uso de segmentação teve 0.875, oferecento cerca de 8% de acréscimo nessa métrica.

Considerando o Recall, que é uma métrica que procura identificar corretamente os exemplos positivos de um conjunto de dados, o tipo de anotação do tipo segmentação foi muito mais superior, principalmente ao treinar com mais de 800 épocas. Ao final do treinamento, atingindo 1000 épocas, enquanto o modelo que utiliza caixas teve recall de 0.7744, o modelo de segmentação teve 0.8312. Essa diferença de 0.0576 pode parecer pequena, mas representa uma melhora significativa na capacidade do modelo de segmentação em detectar objetos que o modelo de caixas delimitadoras não conseguiu identificar.

A métrica mAP50 é uma métrica referente ao desempenho global do modelo, medindo a precisão do modelo, levando em consideração se a área de interseção entre a anotação que foi prevista pelo modelo e a anotação real é de pelo menos 50%. Ambos os modelos mostram um aumento no mAP50 à medida que o treinamento avança, indicando que estão aprendendo a detectar melhor os objetos. O modelo YOLO v11, treinado com anotações de segmentação, apresentou um desempenho superior em termos de mAP50 em comparação com o modelo treinado com anotações de boxes. Isso denota que as anotações de segmentação fornecem informações mais precisas para o treinamento do modelo, e consequentemente uma melhor capacidade de detecção de objetos. Enquanto que o modelo de boxes teve mAP50 de 0.7723, o modelo de segmentação teve 0.8631. De todas as métricas, essa é a que denota maior diferença entre os modelos.

Apesar da anotação de segmentação ter maior dificuldade no processo de marcação, sendo necessário construír um polígono, essa característica de forma dos objetos ajuda na identificação. Contudo, a anotação de caixa delimitadora é mais fácil de rotular, e pode ser utilizada em ambientes com fundo parecido com as imagens de treinamento. Grande parte do erro que ocorre no uso de caixas está no aprendizado da área que não denota o fruto, as bordas da caixa. Considerando os resultados de precisão de 0.85142, recall 0.82768 e mAP50 0.85364 obtidos pela YOLOv11, é possível concluir que o modelo se destaca como a melhor opção devido à sua eficiência no tempo de treinamento e na performance geral, principalmente quando se leva em conta o equilíbrio entre a pontuação F1 e o limite de confiança. Esses fatores tornam a YOLOv11 uma escolha mais adequada para cenários onde o tempo de processamento e a eficiência são relevantes, sem comprometer a precisão do modelo.

Acknowledgments: The authors would like to thank the Research Foundation of Brazil (National Council for Scientific and Technological Development (CNPq), the Coordination for Upgrading Higher Institution Personnel (CAPES); the Research Support Foundation of the State of Goias

(FAPEG); the Financier of Studies and Projects (FINEP); Center of Excellence in Agro Exponen- tial (CEAGRE) and the Federal Institute Goiano for their financial and logistical support.

Referências

AFONSO, M. et al. Tomato fruit detection and counting in greenhouses using deep learning. *Frontiers in Plant Science*, v. 11, p. 571299, November 2020. Citado na página 1.

BAC, C. W. et al. Harvesting robots for high-value crops: State-of-the-art review and challenges ahead. *Journal of Field Robotics*, v. 31, 07 2014. Citado na página 1.

BASSETT, D.; KUMARAN, M. Libraries and google co-op. *Journal of Library Administration*, v. 46, p. 181 – 189, 2008. Citado na página 5.

CAI, Y. et al. 1. cherry tomato detection for harvesting using multimodal perception and an improved yolov7-tiny neural network. *Agronomy*, 2024. Citado na página 3.

CHEN, H.; ZHOU, G.; JIANG, H.-S. Student behavior detection in the classroom based on improved yolov8. *Sensors (Basel, Switzerland)*, v. 23, 2023. Citado na página 6.

CHOUDHURY, S. D.; SAMAL, A.; AWADA, T. Leveraging image analysis for high-throughput plant phenotyping. *Frontiers in Plant Science*, v. 10, April 2019. Citado na página 1.

GUILHERMEHC. *Open Source Dataset, detect-and-counting-tomato-finish Dataset*. [S.l.]: Roboflow, 2024. <https://universe.roboflow.com/guilhermehc/detect-and-counting-tomato-finish>. Visited on 2024-05-08. Citado 3 vezes nas páginas 2, 4 e 5.

HENDERSON, P.; FERRARI, V. End-to-end training of object class detectors for mean average precision. In: LAI, S. et al. (Ed.). *Computer Vision – ACCV 2016*. Springer International Publishing, 2017. p. 198–213. Disponível em: <https://doi.org/10.1007/978-3-319-54193-8_13>. Citado na página 7.

IWAMURA, M.; TSUKADA, M.; KISE, K. 3. automatic labeling for scene text database. In: . [S.l.: s.n.], 2013. Citado na página 4.

JUNIOR, G. D. B. V. et al. Métricas utilizadas para avaliar a eficiência de classificadores em algoritmos inteligentes. *Centro de Pesquisas Avançadas em Qualidade de Vida*, v. 14, n. 2, p. 1, 2022. Citado na página 8.

KARIRYAA, A. *4. MaskIt: Masking for efficient utilization of incomplete public datasets for training deep learning models.* 2020. Citado na página 5.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: CURRAN ASSOCIATES, INC. *Advances in neural information processing systems*. [S.l.], 2012. v. 25, p. 1097–1105. Citado na página 6.

MAO, Z.; XU, Y.; SUAREZ, E. 4. dataset management platform for machine learning. *arXiv.org*, 2023. Citado na página 4.

MINERVINI, M.; SCHARR, H.; TSAFTARIS, S. Image analysis: The new bottleneck in plant phenotyping [applications corner]. *IEEE Signal Processing Magazine*, v. 32, p. 126–131, 07 2015. Citado na página 1.

OBERTI, R. et al. Selective spraying of grapevines for disease control using a modular agricultural robot. *Biosystems Engineering*, v. 146, p. 1–12, January 2016. Citado na página 1.

PADILLA, R.; NETTO, S.; SILVA, E. da. A survey on performance metrics for object-detection algorithms. In: *Proceedings of the 2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*. [s.n.], 2020. p. 237–242. Disponível em: <https://doi.org/10.1109/IWSSIP48289.2020.9145130>. Citado na página 7.

PAULIN, S. et al. A comparison of sampling-based path planners for a grape vine pruning robot arm. *ICARA 2015 - Proceedings of the 2015 6th International Conference on Automation, Robotics and Applications*, p. 98–103, 04 2015. Citado na página 1.

PEIL, R. M.; NETO, A. A. de A.; ROMBALDI, C. V. 4. densidade de plantio e genótipos de tomateiro cereja em sistema fechado de cultivo em substrato. *Horticultura Brasileira*, 2014. Citado na página 3.

PERALTA, I. E.; SPOONER, D. M. *History, origin and early cultivation of tomato (Solanaceae)*. [S.l.]: Enfield, [NH] : Science Publishers, c2005., 2004. 1-24 p. Citado na página 1.

RAJCHL, M. et al. Deepcut: Object segmentation from bounding box annotations using convolutional neural networks. *IEEE Transactions on Medical Imaging*, v. 36, p. 674–683, 2016. Citado na página 4.

REDMON, J. et al. You only look once: Unified, real-time object detection. In: . [S.l.: s.n.], 2016. p. 779–788. Citado na página 6.

RHEE, H.; KOO, H. I.; CHO, N. I. 5. object segmentation using parametric representation. In: . [S.l.: s.n.], 2022. Citado na página 4.

ROBOFLOW. *Roboflow: Computer vision tools for developers and enterprises*. 2024. <https://roboflow.com/>. Acesso em 7 de novembro de 2024. Citado na página 4.

SHEN, L.; LANG, B.; SONG, Z. Ds-yolov8-based object detection method for remote sensing images. *IEEE Access*, v. 11, p. 125122–125137, 2023. Citado na página 6.

ULTRALYTICS. *YOLO11 NEW*. Accessed October 21, 2024. Disponível em: <https://docs.ultralytics.com/models/yolo11>. Citado na página 6.

ULTRALYTICS. *YOLOv8: Real-Time Object Detection and Segmentation*. 2023. Accessed: YYYY-MM-DD. Disponível em: <https://github.com/ultralytics/ultralytics>. Citado na página 6.

WANG, A. et al. Yolov10: real-time end-to-end object detection. *arXiv*, v. 2405, n. 14458, 2024. Citado na página 6.

WANG, C. et al. A lightweight cherry tomato maturity real-time detection algorithm based on improved yolov5n. *Agronomy*, v. 13, p. 2106, 08 2023. Citado na página 3.

WANG, C.-Y.; YEH, I.-H.; LIAO, H.-Y. Yolov9: learning what you want to learn using programmable gradient information. *arXiv*, v. 2402, n. 13616, 2024. Citado na página 6.

WANG, X.; LIU, J. Tomato anomalies detection in greenhouse scenarios based on yolo-dense. *Frontiers in Plant Science*, Frontiers Media SA, v. 12, p. 634103–634103, 2021. Citado na página 3.

WEINGERL, V.; UNUK, T. 1. chemical and fruit skin colour markers for simple quality control of tomato fruits. *Croatian journal of food science and technology*, 2015. Citado na página 1.

YANG, G. et al. A lightweight yolov8 tomato detection algorithm combining feature enhancement and attention. *Agronomy*, 2023. Citado na página 6.

ZENG, T. et al. 14. lightweight tomato real-time detection method based on improved yolo and mobile deployment. *Computers and Electronics in Agriculture*, 2023. Citado na página 3.

ZHANG, F. et al. 3. multi-class detection of cherry tomatoes using improved yolov4-tiny. *International Journal of Agricultural and Biological Engineering*, 2022. Citado na página 3.