



Comparando a Eficiência de Redes Neurais Convolutivas para Detecção de Frutos em Pomares de Laranja

Gabriel B. Moura¹, Kleber X. S. de Souza², João C. Neto³, Sônia Ternes⁴

¹ Autor, Bolsista CNPq (PIBIC): Graduação em Engenharia de Computação, Unicamp
Campinas, SP

² Orientador: Pesquisador da Embrapa Agricultura Digital - Campinas, SP

³ Analista da Embrapa Agricultura Digital - Campinas, SP

⁴ Pesquisadora da Embrapa Agricultura Digital - Campinas, SP

gbmoura0606@gmail.com, {kleber.sampaio, joao.camargo, sonia.ternes}
@embrapa.br

Abstract. *Given the importance of citrus farming in Brazil's economy, one of the global powers in this sector, it is essential to constantly renew cultivation techniques to make production more efficient. This study tests the efficiency of the YOLOv8 neural network for object detection, with the purpose of obtaining a more accurate orange crop estimate. The overall performance presented by the network was satisfactory, but it did not allow us to demonstrate that it is better than its previous version YOLOv5 in all metrics.*

Resumo. *Dada a importância da citricultura na economia do Brasil, uma das potências mundiais nesse setor, é essencial estar sempre renovando as técnicas de cultivo para tornar a produção mais eficiente. Este trabalho testa a eficiência da rede neural YOLOv8 para a detecção de objetos, com o propósito de obter uma estimativa de safra de laranjas mais precisa. O desempenho geral apresentado pela rede foi satisfatório, porém não nos permitiu demonstrar que ela supera sua versão anterior YOLOv5 em todas as métricas.*

1. Introdução

A citricultura desempenha um papel significativo na economia brasileira, sendo uma das atividades agrícolas mais relevantes do país, devido à sua dimensão e potencial de geração de emprego e renda. De acordo com dados da Organização das Nações Unidas para a Alimentação e a Agricultura [FAO 2021], o Brasil é o maior produtor de laranjas do mundo, responsável por mais de 20% da produção. Dada a relevância e competitividade da agricultura citrícola nos cenários nacional e global, busca-se cada vez mais aprimorar as técnicas de cultivo, a fim de otimizar a cadeia produtiva e garantir um sistema eficiente e sustentável.

Um dos principais desafios enfrentados pelos produtores de laranja é a estimativa da safra anual, um fator fundamental para o planejamento e tomada de

decisões no setor. Atualmente, o método tradicional de estimativa de safra é a derriça, que consiste na retirada dos frutos das árvores para contagem manual e avaliação da qualidade da produção. No entanto, essa abordagem demanda um alto custo, por ser intensiva em tempo e mão de obra.

Diante desse cenário, torna-se cada vez mais essencial buscar alternativas que possibilitem a previsão da safra de forma precisa, eficiente e não-destrutiva. Uma das abordagens promissoras é a utilização de técnicas de visão computacional baseadas em redes neurais de aprendizado profundo, que têm sido aplicadas com sucesso em diferentes áreas, principalmente em problemas de detecção de objetos e reconhecimento de padrões.

Existem diversas arquiteturas e redes disponíveis para auxiliar neste problema, bem como trabalhos na literatura científica que exploram seu potencial, como a SSD [Sousa et. al. 2021] e a Faster-RCNN [Cerqueira et. al. 2020]. Neste trabalho, será explorada a última versão da rede neural YOLO, a YOLOv8 [Ultralytics 2023], cuja eficiência será comparada a versões anteriores [Ribeiro et. al. 2022], a fim de avaliar seu desempenho e prever sua viabilidade para a detecção de frutos em pomares de laranja.

2. Material e Métodos

2.1. A rede neural YOLOv8

YOLO (*You Only Look Once*) é um algoritmo de detecção de objetos baseado em redes neurais convolucionais, reconhecido por seu desempenho na abordagem desse tipo de problema. O termo (em tradução livre do inglês, “Você Só Olha Uma Vez”) se refere ao fato de que é um detector de apenas um estágio, isto é, a imagem é processada apenas uma vez ao longo do método, proporcionando um aumento da velocidade e eficiência da detecção.

O formato da arquitetura utilizada nas redes YOLO consiste em três partes principais: o *backbone* (espinha dorsal), que obtém características e captura informações relevantes das imagens de entrada através de camadas de convolução; o *neck* (pescoço), parte intermediária da arquitetura que recebe as características extraídas pelo *backbone* e aplica transformações para melhorar a representação dos dados e refinar as informações; e a *head* (cabeça), que recebe as características aperfeiçoadas pelo *neck* e toma as decisões para gerar as saídas finais da rede, como as caixas delimitadoras e probabilidades de classe dos objetos detectados.

Atualmente, a versão YOLOv8 é considerada o estado da arte para o problema de detecção de objetos. A principal diferença em relação à YOLOv5 é a otimização de alguns filtros convolucionais [King 2023], tanto no *backbone* quanto na *head*, na qual também é implementado um sistema de divisão sem caixas delimitadoras pré-definidas (livre de ancoragens) [Ultralytics 2023], que aumenta a eficiência da arquitetura como um todo. Assim, por manter um equilíbrio entre precisão e velocidade de inferência, a rede apresenta um desempenho geral promissor para o problema de detecção de laranjas, também por sua alta adaptabilidade a diferentes plataformas de *hardware* [Ultralytics 2023].

2.2. Conjunto de dados

O conjunto de dados utilizado, disponibilizado pelo Fundecitrus, é composto por 3064 fotografias de laranjeiras, em que foram feitas marcações, no formato YOLO, com as posições de cada fruto presente nas imagens, todas elas de tamanho 416x416 pixels.

A divisão do conjunto de dados foi realizada visando evitar o sobreajuste do modelo. Um modelo sobreajustado opera bem sobre dados conhecidos, porém fornece previsões imprecisas para novos dados por estar enviesado, e isso pode ocorrer devido a pouca quantidade ou variedade do conjunto de treinamento. Para tanto, foram separadas 2053 imagens para treino, 200 para validação e 811 para teste. As imagens, obtidas a partir de uma propriedade produtora de laranja, contam com variações de posição, ângulo, iluminação e grau de maturação dos frutos, a fim de não enviesar o modelo.

2.3. Treinamento

Na etapa de treinamento, foram utilizados os modelos com os pesos pré-treinados disponibilizados pela Ultralytics, para todos os seus tamanhos disponíveis: YOLOv8n (nano), YOLOv8s (*small*, pequena), YOLOv8m (*medium*, média), YOLOv8l (*large*, grande) e YOLOv8x (*extra large*, extra grande).

Por ser uma versão mais recente da YOLO, a execução do treinamento foi realizada diretamente na linha de comando através do comando `yolo`, sem a necessidade de rodar Python explicitamente, embora a aplicação tenha sido elaborada nessa linguagem. Os parâmetros do treinamento alterados em relação ao padrão automático foram: o tamanho da imagem, reduzido de 640 a 416; o dispositivo (NVIDIA GeForce RTX 3070); a paciência, para que o treinamento finalizasse as épocas independente da melhora; e o número de épocas, aumentado para 900, com mil iterações cada.

3. Resultados e Discussão

3.1. Resultados da YOLOv8

Para avaliar o desempenho dos modelos no problema de detecção de laranjas, foram utilizadas as métricas: mAP_{0.5} (*mean average precision*, precisão média), calculado com a medida interseção sobre união (do inglês, *intersection over union*, IoU) fixa em 0,5; precisão, que representa, dentre os objectos identificados, quantos eram realmente laranjas; revocação, que representa quantas laranjas foram detectadas dentre todas as presentes; e F1, índice que relaciona a precisão e revocação, calculado como a média harmônica dessas duas métricas.

Uma vez que a precisão média leva em conta a sobreposição das caixas delimitadoras, ela é importante ao avaliar a capacidade do modelo de identificar as laranjas em relação à sua localização atual, enquanto o índice F1, ao ser mais sensível a desequilíbrios entre falsos positivos e falsos negativos, avalia o desempenho global dos modelos de detecção de objetos.

Durante a análise dessas métricas ao longo do treinamento, percebe-se que o melhor modelo não necessariamente é o último gerado, conforme Figura 1, pois o pico do índice F1 para a YOLOv8m, por exemplo, ocorre por volta da iteração 480, decaindo em seguida.

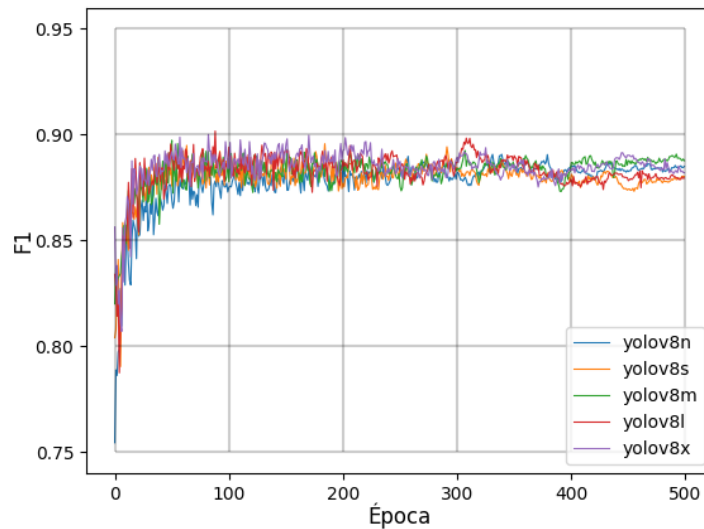


Figura 1. Variação do índice F1 durante o treinamento das redes, com mil iterações por época

Assim, para a validação, foram escolhidos apenas os modelos de melhor F1, melhor mAP_0.5, o gerado na última época do treinamento e o de melhor desempenho geral. Dentre eles, seguem os resultados apresentados nas Tabelas 1 e 2.

Tabela 1. Validação do treinamento no conjunto de validação

| | YOLOV8n | YOLOV8s | YOLOV8m | YOLOV8l | YOLOV8x |
|----------------|---------|---------|---------|---------|---------|
| Melhor P | 0,953 | 0,949 | 0,967 | 0,975 | 0,977 |
| Melhor R | 0,853 | 0,879 | 0,834 | 0,867 | 0,846 |
| Melhor F1 | 0,8855 | 0,8864 | 0,8838 | 0,9075 | 0,8959 |
| Melhor mAP_0.5 | 0,917 | 0,923 | 0,906 | 0,922 | 0,908 |

Tabela 2. Validação do treinamento no conjunto de teste

| | YOLOV8n | YOLOV8s | YOLOV8m | YOLOV8l | YOLOV8x |
|----------------|---------|---------|---------|---------|---------|
| Melhor P | 0,956 | 0,962 | 0,966 | 0,972 | 0,985 |
| Melhor R | 0,898 | 0,894 | 0,883 | 0,876 | 0,875 |
| Melhor F1 | 0,9101 | 0,9150 | 0,9081 | 0,8754 | 0,9119 |
| Melhor mAP_0.5 | 0,94 | 0,945 | 0,932 | 0,929 | 0,930 |

A princípio, os resultados apresentam um bom desempenho com suas métricas, porém tratando-se do caso específico de contagem de laranjas, é interessante também compará-los com os de uma versão anterior da YOLO, a exemplo da YOLOv5 [Ribeiro 2022].



Figura 2. Imagem do conjunto de teste processada pelo algoritmo, com o grau de confiança de cada caixa delimitadora.

3.2. Comparação com a YOLOv5

Com o propósito de se ter uma maior dimensão do desempenho da rede YOLOv8, suas métricas foram comparadas aos testes realizados com sua versão anterior YOLOv5 [Ribeiro et. al. 2022], gerando a Tabela 3.

Tabela 3. Comparação das redes utilizando o conjunto de teste

| | YOLOn | | YOLOs | | YOLOm | | YOLOl | | YOLOx | |
|----------------|--------------|---------------|---------------|--------------|--------------|---------------|---------------|--------------|---------------|--------------|
| | V5 | V8 | V5 | V8 | V5 | V8 | V5 | V8 | V5 | V8 |
| Melhor P | 0,942 | 0,956 | 0,945 | 0,962 | 0,932 | 0,966 | 0,934 | 0,972 | 0,937 | 0,985 |
| Melhor R | 0,878 | 0,898 | 0,875 | 0,894 | 0,894 | 0,883 | 0,893 | 0,876 | 0,906 | 0,875 |
| Melhor F1 | 0,9090 | 0,9101 | 0,9200 | 0,9150 | 0,9030 | 0,9081 | 0,9070 | 0,8754 | 0,9210 | 0,9119 |
| Melhor mAP_0.5 | 0,951 | 0,940 | 0,950 | 0,945 | 0,955 | 0,932 | 0,956 | 0,929 | 0,958 | 0,930 |

Embora todos os valores da métrica mAP_0.5 permaneçam superiores na versão menos recente da rede, nota-se uma melhora significativa da precisão para a YOLOv8, bem como da revocação para os modelos menores.

Comparando-se os índices F1, não há uma rede com absoluta vantagem, mas o que se pode inferir de ambas as versões é que não necessariamente os modelos com mais parâmetros geram as melhores métricas (a exemplo do índice F1 da YOLOv8s, que superou o da YOLOv8x). Isso é interessante pois os modelos podem apresentar bom desempenho mesmo com a limitação de poder computacional enfrentada no contexto do problema.

4. Conclusão

Neste trabalho, foi testado o desempenho da rede neural YOLOv8 no problema de detecção de laranjas, com o objetivo de realizar uma estimativa precisa da safra fotografada, essencial para melhorar a cadeia produtiva como um todo. A rede avaliada

apresentou bons resultados, porém a maioria de suas métricas de desempenho permaneceu inferior às de uma versão anterior a sua, a YOLOv5.

5. Agradecimentos

Devemos agradecimentos ao PIBIC/CNPq, pela concessão da bolsa de Iniciação Científica ao aluno Gabriel Batista Moura, e à equipe do PES/Fundecitrus pela disponibilização das fotos de laranjeiras.

6. Referências

- Cerqueira, L. M.; Souza, K. X. S. de; Ternes, S. e Camargo NETO, J. (2020) “Usando a rede neural FasterRCNN para identificar frutos verdes em pomares de laranja, In: CONGRESSO INTERINSTITUCIONAL DE INICIAÇÃO CIENTÍFICA, 14. Anais... Campinas: Embrapa Informática Agropecuária. p. 1-9. Evento online. CIIC 2020. No 20605. Disponível em: <<https://www.alice.cnptia.embrapa.br/handle/doc/1127731>>. Acesso em: jun. 2023.
- FAO. “FAOStat”. Disponível em: <<https://www.fao.org/faostat/en/#compare>>. Acesso em: jul. 2023.
- Fundecitrus. (2023) “Estimativa da safra de laranja 2023/24 do cinturão citrícola de São Paulo e Triângulo/Sudoeste Mineiro: sumário executivo”, Araraquara. 16 p. Disponível em: <https://www.fundecitrus.com.br/pdf/pes_relatorios/2023_05_10_Sum%C3%A1rio_Executivo_da_Estimativa_da_Safra_de_Laranja_2023-2024.pdf> . Acesso em: jul. 2023.
- King, R. (2023) “Brief summary of YOLOv8 model structure”. Disponível em: <<https://github.com/ultralytics/ultralytics/issues/189>>. Acesso em: jul. 2023.
- Ribeiro, A. F.; Souza, K. X. S. de; Camargo Neto, J.; Ternes, S. e Yano, I. H. (2022) “Testando a rede neural YOLOv5 para detecção de frutos em pomares de laranja”, In: CONGRESSO INTERINSTITUCIONAL DE INICIAÇÃO CIENTÍFICA, 16., Campinas. Anais... Campinas: Instituto Agrônômico. p. 1-11. Evento online. CIIC 2022. No 22602. Disponível em: <<http://www.alice.cnptia.embrapa.br/alice/handle/doc/1146123>>. Acesso em: fev. 2023.
- Sousa, M. A. de; Souza, K. X. S. de; Camargo Neto, J.; Ternes, S. e Yano, I. H. (2021) “Usando a rede neural SSD para identificar frutos verdes em pomares de laranja”, In: CONGRESSO INTERINSTITUCIONAL DE INICIAÇÃO CIENTÍFICA, 15. Anais... Campinas: Instituto de Zootecnia. p. 1-10. Evento online. CIIC 2021. No 21610. Disponível em: <<https://www.alice.cnptia.embrapa.br/handle/doc/1135159>>. Acesso em: fev. 2023.
- Ultralytics. (2023) “YOLOv8”. Disponível em: <<https://docs.ultralytics.com/models/yolov8>>. Acesso em: abr. 2023.