

DESEMPENHO DE VARIANTES DA YOLOv8 PARA DETECÇÃO DE PLANTAS DANINHAS EM TEMPO REAL EM CULTURAS AGRÍCOLAS E IMPACTOS NA AGRICULTURA PRECISÃO

William Augusto Ellert Kroth¹; Estéfani Sulzbach²; Ismael Scheeren³, Mauricio Cagliari Tosin⁴,
Catarine Markus⁵

Palavras-chave: You Only Look Once, detecção em tempo real, redes neurais, aplicação localizada de herbicidas

Introdução

A cultura da soja tem ganhado destaque em áreas de terras baixas, através do sistema “ping-pong”, uma rotação estratégica com a cultura do arroz, que contribui para o manejo sustentável, melhora o controle fitossanitário e otimiza o uso das áreas. Com o avanço da inteligência artificial (IA), especialmente por meio de técnicas de visão computacional e *deep learning*, surgiram soluções automatizadas capazes de detectar e classificar plantas daninhas com alta precisão e que tem gerado possibilidades de realizar a aplicação localizada de herbicidas e em tempo real (GERHARDS et al., 2022). No entanto, essas tecnologias ainda não estão disponíveis para os sistemas de arroz e soja. A aplicação localizada de herbicidas nesses cultivos possibilitará a redução dos custos com herbicidas, aumento de sustentabilidade das áreas e menor pressão de seleção para a resistência aos herbicidas.

Entre os algoritmos mais eficientes para essa finalidade, destaca-se a arquitetura YOLO (*You Only Look Once*), que se consolidou como referência devido à sua rapidez, capacidade de lidar com objetos sobrepostos e alta precisão (BOCHKOVSKIY et al., 2020). Essas características são fundamentais para sistemas de aplicação localizada de herbicidas, eliminando a necessidade de mapeamentos prévios (HASAN et al., 2024). Este trabalho teve como objetivo avaliar a eficiência da YOLOv8 no reconhecimento de plantas daninhas no cultivo da soja, de forma a testar diferentes configurações de treinamento (com e sem pré-treinamento), e otimizar o modelo por meio da redução de parâmetros, visando aumentar a velocidade de processamento sem comprometer a acurácia. A primeira fase deste estudo tem como objetivo o estabelecimento da tecnologia na cultura da soja. A segunda fase (em fase de implementação) contempla o treinamento do sistema para a cultura do arroz, a qual representa um desafio adicional, uma vez que as principais espécies daninhas pertencem à mesma família da cultura (Poaceae). A implementação dessa abordagem poderá viabilizar sua integração a equipamentos agrícolas, permitindo o controle automatizado e preciso de plantas daninhas, o que contribuirá para a redução do uso de herbicidas e para a promoção da sustentabilidade na oricultura.

Material e Métodos

A aquisição das imagens para compor o banco de dados foi realizada em diversas lavouras do Rio Grande do Sul, durante os períodos de novembro a março nas safras de 2020 a 2023. Para o registro das imagens, foi empregada uma câmera digital da marca Sony, posicionada a aproximadamente um metro de altura em relação ao solo. As imagens foram obtidas em diferentes condições de luminosidade (dias ensolarados, parcialmente nublado, até condições de céu completamente encoberto), diferentes estágios fenológicos das plantas daninhas (plântulas até plantas adultas em fase reprodutiva, com presença de inflorescências) e da cultura da soja (estágio vegetativo V3 até o reprodutivo R7) e níveis distintos de palha no sistema. As plantas daninhas presentes foram categorizadas em dois grupos morfológicos: dicotiledôneas (folhas largas) e monocotiledôneas (folhas estreitas).

¹ Aluno de graduação em Agronomia, Universidade Federal do Rio Grande do Sul, UFRGS, Porto Alegre – RS, [williamkroth64@gmail.com](mailto:wiliamkroth64@gmail.com)

² Engenheira agrônoma, Doutoranda em fitotecnia, Universidade Federal do Rio Grande do Sul, UFRGS, Porto Alegre – RS, estefanisulzbach@gmail.com

³ Engenheiro mecânico, Me, Accore Automação & Controle Industrial, Porto Alegre, ismael.scheeren@accore.com.br

⁴ Doutor em eng eletrica, Professor Adjunta, Universidade Federal do Rio Grande do Sul, UFRGS, Porto Alegre – RS, mauricio.ctosin@gmail.com

⁵ Doutora em fitotecnia, Professora Adjunta, Universidade Federal do Rio Grande do Sul, UFRGS, Porto Alegre – RS, catarine.markus@ufrgs.br

A arquitetura YOLO requer um aprendizado supervisionado, assim foi necessário realizar um processo de anotação manual. Para essa etapa utilizou-se o software LabelImg, que permitiu a criação de caixas delimitadoras ao redor de cada objeto (planta presente nas imagens). O processo resultou em 21.092 anotações (instâncias) para a cultura da soja, 47.075 para plantas daninhas dicotiledôneas e 26.242 para monocotiledôneas (Tabela 1). A distribuição do conjunto de dados seguiu uma proporção de 80% imagens para treinamento (1.563 imagens), 10% para teste (196 imagens) e 10% para validação (195 imagens).

Tabela 1. Número total de instâncias anotadas para cada classe nas fases de treinamento, teste e validação.

Classes	Treinamento (1.563 imagens)	Teste (196 imagens)	Validação (195 imagens)	Instâncias	
Soja	17.415	1.839	1.838		
Dicotiledôneas	37.736	3.941	5.398		
Monocotiledôneas	20.923	2.429	2.890		

Para a avaliação de desempenho foram testadas nove variações da arquitetura YOLOv8. Desse total, quatro correspondem às versões originais do modelo, que foram testadas com e sem pré-treinamento (YOLOv8n (nano), YOLOv8s (small), YOLOv8l (large), e YOLOv8x (extra-large)). Enquanto as cinco redes restantes foram geradas com otimização do número de parâmetros (YOLOv8 femto, YOLOv8 atto, YOLOv8 pico e YOLOv8 zepto). Adicionalmente, realizou-se a validação cruzada ($k=5$) nos modelos que apresentaram melhor desempenho inicial (Raschka, 2020), com o objetivo de aumentar a confiabilidade dos resultados. Nesse processo, o conjunto de dados foi dividido aleatoriamente em cinco subconjuntos, sendo utilizado alternadamente como conjunto de validação durante o treinamento.

As métricas selecionadas para avaliação dos modelos incluíram o mAP (*Mean Average Precision*) e os GFLOPs (*Giga Floating Point Operations Per Second*). O mAP foi calculado considerando um limiar de IoU (*Intersection over Union*) de 0.5 (50%), conforme estabelecido por Henderson & Ferrari (2017), que mede a sobreposição entre as caixas delimitadoras geradas pelo modelo e as anotações manuais de referência. Desta forma o mAP indica a assertividade do modelo de detectar e classificar objetos, considerando tanto a exatidão quanto a cobertura. Por sua vez, os GFLOPs, que representam bilhões de operações de ponto flutuante por segundo, foram utilizados como indicador de eficiência computacional, sendo valores mais altos associados a melhor desempenho de processamento. Essa abordagem metodológica permitiu uma análise abrangente do desempenho dos diferentes modelos, considerando a precisão na detecção quanto à eficiência operacional.

Resultados e Discussão

A partir dos resultados obtidos com o treinamento dos modelos originais da rede neural YOLOv8 (YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, YOLOv8x) foi verificado um platô no desempenho tanto para arquiteturas sem pré-treino (*Ground truth*) quanto para aquelas com pré-treino (*Pre-trained*) (Figura 1). As redes sem pré-treino apresentaram desempenhos similares às redes pré-treinadas, ambas com valores de mAP50 entre 0,70 e 0,73. Além disso, foi verificado que a YOLOv8n alcançou mAP50 de 0,70 e 0,71 com número menor de parâmetros 3.157.200 e 3.006.233; para treinamentos sem e com pré-treino, respectivamente. Em contraste, as arquiteturas YOLOv8l e YOLOv8x, apresentaram um incremento no mAP50 (de 0,71 para 0,73) quando pré-treinadas, no entanto, apresentaram aumento no número de parâmetros (43.691.520 e 68.229.648) e GFLOPs de 165,7 e 258,5; respectivamente. Essa diferença ressalta a importância de se buscar modelos mais eficientes, especialmente quando consideramos aplicações em sistemas embarcados, onde a velocidade de processamento de dados são críticos para o funcionamento em tempo real. A capacidade da YOLOv8n de manter um alto desempenho com um número reduzido de parâmetros sugere que essa arquitetura pode ser ideal para implementação em equipamentos agrícolas autônomos ou sistemas de pulverização inteligente.

Além disso, a similaridade de desempenho entre modelos com e sem pré-treinamento indica que, em certos cenários, o pré-treinamento pode não ser necessário.

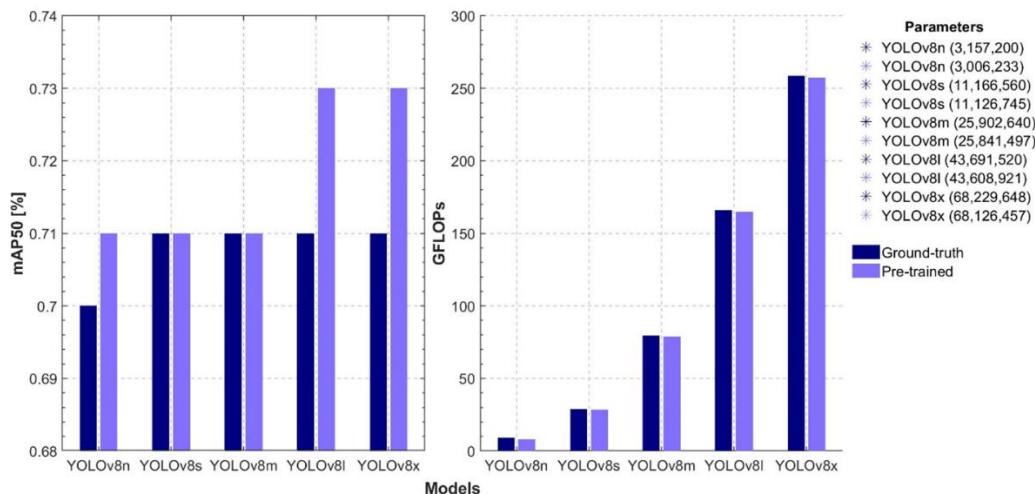


Figura 1. Comparação dos modelos originais da rede neural YOLOv8 gerados sem e com pré-treinamento.

Conforme os resultados encontrados na Figura 1 foi possível investigar qual arquitetura causa uma redução gradual de desempenho com a redução do número de parâmetros. Ao analisar as quatro redes YOLOv8 fêmto, YOLOv8 atto, YOLOv8 pico, YOLOv8 zepto com número menor de parâmetros <1.000.000, verificou-se maiores valores de mAP50: 0,70; 0,67; 0,70; e 0,58; e menores de GFLOPs: 5,2; 2,6; 7,1; e 1,4; respectivamente. Já as versões originais YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, YOLOv8x propostas pelos desenvolvedores da YOLOv8 com número de parâmetros >3.000.000 mantiveram o mAP50 entre 0,70 e 0,71 e apresentaram valores maiores de GFLOPs: 8,9; 28,8; 79,3; 165,7; 258,5; respectivamente. Esses resultados indicam que as quatro redes com otimização de parâmetros apresentaram perda mínima de performance. O número de parâmetros está diretamente relacionado a mensuração da complexidade do modelo, além de exigir uma quantidade maior de memória para implantação e afeta o custo e tempo de inferência (DANG et al., 2023). Nessa linha, foi verificado que A YOLOv8 fêmto com 904.633 parâmetros e 5,2 GFLOPs, manteve um mAP50 de 0,70, ou seja, é possível reduzir significativamente a complexidade do modelo sem comprometer o desempenho.

A confiabilidade dos resultados da Figura 2 foram confirmados pela validação cruzada realizada nas arquiteturas YOLOv8n e YOLOv8 fêmto (Tabela 1). Ambas mantiveram valores consistentes de mAP50 (0,70 a 0,73) e GFLOPs ao longo das cinco iterações, reforçando a robustez dos modelos. Essa consistência é crucial para aplicações práticas no campo, onde variações no desempenho podem impactar diretamente a eficácia do controle de plantas daninhas. A YOLOv8 fêmto mostrou-se uma alternativa para estudos de validação a campo e para treinamento com um banco de imagens de áreas de arroz para detecção de plantas daninhas importantes para a cultura do arroz.

Os resultados deste estudo destacam o potencial da YOLOv8, especialmente em suas versões com menor número de parâmetros, para o manejo de plantas daninhas na cultura da soja. A capacidade de detectar espécies invasoras com alta precisão e baixo custo computacional abre caminho para a adoção em larga escala de tecnologias de agricultura de precisão, contribuindo para o aumento da produtividade e a sustentabilidade. Além disso, a descoberta de que o pré-treinamento pode não ser essencial em certos casos simplifica o desenvolvimento de soluções personalizadas, tornando a tecnologia mais acessível a diferentes realidades agrícolas.

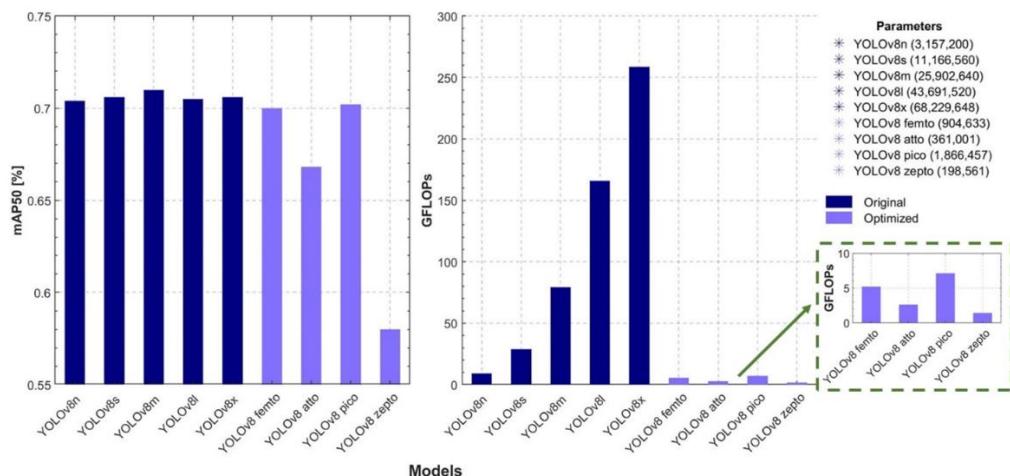


Figura 2. Desempenho das cinco arquiteturas originais da YOLOv8 sem pré-treino (do zero) versus as quatro redes com otimização dos parâmetros.

Tabela 1. Parâmetros após a realização do *Cross validation* para as arquiteturas YOLOv8n e YOLOv8 femto.

Modelo	Interações	Parâmetros	GFLOPs	mAP50	Modelo	Interações	Parâmetros	GFLOPs	mAP50
YOLOv8n	1	3.006.233	8,1	0,70	YOLOv8 femto	1	904.633	5,2	0,70
	2	3.006.233	8,1	0,69		2	904.633	5,2	0,70
	3	3.006.233	8,1	0,71		3	904.633	5,2	0,70
	4	3.006.233	8,1	0,71		4	904.633	5,2	0,71
	5	3.006.233	8,1	0,73		5	904.633	5,2	0,72

Conclusões

Este é o primeiro estudo comparado diferentes variantes da última versão da YOLO lançada em 2023 (YOLOv8) e com otimização de parâmetros para detecção de plantas daninhas verificadas no Brasil para a cultura da soja. As variantes da YOLOv8 sem pré-treino apresentam desempenho similar às redes pré-treinadas. A redução do número de parâmetros pode ser uma importante ferramenta para reduzir a complexidade e processamento de dados sem afetar o desempenho das arquiteturas. Esses resultados serão importantes para a geração de banco de dados de plantas daninhas em áreas de arroz e posteriormente para a seleção de modelos com menor complexidade e sem perder a performance na detecção de plantas daninhas em tempo real.

Agradecimentos

À Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS) pelo financiamento desta pesquisa (processos nº 22/2551-0000394-0 e nº 24/2551-0001163-3).

Referências

- BOCHKOVSKIY, A.; WANG, CY; LIAO, HYM. YOLOv4: Optimal Speed and Accuracy of Object Detection. Arxiv, v.2004, p.10934, 2020.
- DANG, F. et al. YOLOWeeds: A novel benchmark of YOLO object detectors for multi-class weed detection in cotton production systems. **Computers and Electronics in Agriculture**, v. 205, p. 107655, 2023.
- GUO, Y. et al. Deep learning for visual understanding: A review. **Neurocomputing**, v. 187, p. 27-48, 2016.
- HASAN, A. S. M. M. et al. Object-level benchmark for deep learning-based detection and classification of weed species. **Crop Protection**, v. 177, p. 106561, 2024.