

SoYolo: Detecção automática de vagens e grãos de soja

Víctor Vinícius Welter¹, Larissa Guder^{1,2}

¹ Laboratório de Pesquisas Avançadas para Computação em Nuvem (LARCC), Sociedade Educacional Três de Maio (SETREM), Três de Maio, RS, Brasil.

²Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), Porto Alegre, Brasil.

{victorwelter2003, lariguder10}@gmail.com

Abstract. *Soybean yield estimation is a traditionally manual, slow, and imprecise task. We propose SoYOLO, a computer vision model for the automatic detection of soybean pods and grains using smartphone images. Based on the YOLOv8 architecture, the model was trained and validated on 943 and 236 publicly available images, respectively. A final evaluation was conducted on a distinct set of 601 locally captured images. The best-performing model, YOLOv8m, achieved a mean Average Precision (mAP50-95) of 0.484. Our results demonstrate the potential of this method as an accessible tool for agricultural analysis, identifying occlusion and shading as the primary challenges affecting detection accuracy.*

Resumo. *A estimativa de produtividade em lavouras de soja é um processo manual, lento e impreciso. Este artigo propõe o SoYOLO, um modelo de visão computacional para a detecção automática de vagens e grãos de soja a partir de imagens de smartphone. Utilizando a arquitetura YOLOv8, o modelo foi treinado com um dataset de 943 imagens e validado em outras 236, ambas oriundas de fontes públicas. O teste final foi conduzido em um conjunto distinto, com 601 imagens capturadas localmente. A configuração que apresentou melhor desempenho, YOLOv8m, alcançou uma Precisão Média (mAP50-95) de 0.484. Os resultados demonstram a viabilidade da abordagem como uma ferramenta de baixo custo para análises agrícolas, ao mesmo tempo em que revelam os desafios de oclusão e sombreamento como principais barreiras para a precisão da detecção.*

1. Introdução

A soja é uma das culturas agrícolas mais significativas do mundo, sendo uma das principais produzidas e exportadas no Brasil. Segundo dados da Embrapa, na safra 2023/24, o Brasil produziu 147,35 milhões de toneladas, o que representa mais de 37% da produção global [Embrapa 2025]. Essa cultura desempenha um papel estratégico na economia nacional, sendo amplamente utilizada na fabricação de alimentos, rações e biocombustíveis.

Apesar da sua relevância, os métodos tradicionalmente utilizados para estimar a produtividade em lavouras de soja ainda dependem de processos manuais, como a coleta de amostras, a trilha de grãos e os cálculos baseados em médias. Esses procedimentos, além de demandarem tempo e mão de obra, estão sujeitos a erros humanos e baixa escalaabilidade [Ferreira 2024].

Para superar essas limitações da contagem manual, abordagens baseadas em aprendizado profundo têm se destacado na literatura, especialmente a família de algoritmos YOLO (*You Only Look Once*). Abordagens como o YOLO POD, por exemplo, já demonstravam alta correlação com a contagem manual (R^2 de 0,967) [Xiang et al. 2023], enquanto outras focaram em criar modelos leves e eficientes, como o SP-YOLO, que atingiu 94,6% de mAP@0.5 com apenas 12,5 MB [Yu et al. 2024]. A arquitetura YOLOv8, em particular, emergiu como o estado da arte. Em um estudo de grande relevância, [Jia et al. 2025] aprimorou o YOLOv8 para a detecção de vagens pequenas e oclusas, elevando sua performance para 97,3% de mAP@0.5. O sucesso e a performance do YOLOv8, validados por esses trabalhos, fundamentam sua escolha como modelo para o presente estudo.

Este trabalho propõe o desenvolvimento e a avaliação de um modelo de visão computacional, baseado na arquitetura YOLOv8, capaz de realizar a detecção automática de vagens e grãos de soja em imagens reais. O objetivo central é validar a eficácia desta abordagem utilizando imagens de *smartphones*, visando oferecer um método de baixo custo e alta acessibilidade para análises agrícolas.

2. Metodologia

Para garantir a generalização do modelo, os dados foram divididos em três conjuntos: treino, validação e teste. Os conjuntos de treino e validação foram construídos a partir dos datasets descritos em [Jiang et al. 2025]. O dataset Chongzhou contém 570 imagens (Figura 1(a)) e o dataset Renshou contém 878 imagens (Figura 1(b)). Devido à heterogeneidade das fontes, foi realizado um processo de padronização, unificando os diversos formatos de anotação e remapeando as classes para um esquema multiclasse baseado no número de sementes por vagem (ex: '0spp', '1spp', '2spp', etc.). Após, randomicamente os dados foram divididos em 80% para o treino e 20% para validação. O conjunto de teste denominado noroeste (Figura 1(c)), foi construído através de imagens capturadas com smartphone no município de Três de Maio - Rio Grande do Sul. Ao todo, 601 imagens foram capturadas e anotadas para este processo.

O algoritmo de detecção de objetos escolhido foi o YOLOv8, da Ultralytics, pois a literatura recente tem mostrado que ele obtém ótimos resultados na detecção de vagens pequenas e oclusas [Jia et al. 2025]. Foram avaliadas comparativamente as arquiteturas YOLOv8s (*Small*) e YOLOv8m (*Medium*). Os experimentos foram conduzidos utilizando o tamanho da imagem de entrada de 640x640 e aplicando um conjunto avançado de técnicas de aumento de dados (*Data Augmentation*) para simular diferentes condições de iluminação e captura. Os modelos foram treinados por até 150 épocas, utilizando *early stopping*. O ambiente de execução foi: GPU NVIDIA RTX 4060 Ti, processador AMD Ryzen 7 5700X, 16gb de RAM, e Windows 11. A performance foi avaliada quantitativamente utilizando as métricas padrão para detecção de objetos: Precisão (*Precision*), Revocação (*Recall*) e, principalmente, a média da Precisão Média (mAP50-95), o coeficiente de determinação (R^2), o Erro Médio Absoluto (MAE), que serviu como critério principal para a seleção da melhor configuração do modelo.

2.1. Seleção do Melhor Modelo

Os modelos YOLOv8s e YOLOv8m foram treinados e avaliados com resoluções de imagem de 640x640 com diferentes batches. As métricas de Precisão (*Precision*), Revocação



(a) Dataset Chongzhou

(b) Dataset Renshou

(c) Dataset Noroeste

Figura 1. Exemplos de imagens presentes em cada um dos datasets utilizados no processo de treinamento do modelo

(*Recall*) e, principalmente, a média da Precisão Média (*mAP50-95*), foram utilizadas para a comparação de desempenho. Os resultados estão consolidados na Tabela 1, e mostram que a configuração **yolov8m de 8 batch** alcançou o melhor desempenho geral, com um valor de *mAP50-95* de **0.484**.

Tabela 1. Resultados comparativos dos modelos YOLOv8 no conjunto de validação.

Modelo	Resolução	Batch	Precisão	Recall	<i>mAP50-95</i>	R ²	MAE
yolov8s	640x640	8	0.699	0.534	0.442	0.4380	8.4761
yolov8m	640x640	8	0.711	0.607	0.484	0.4905	7.8715
yolov8s	640x640	16	0.625	0.586	0.423	0.1461	10.5743
yolov8m	640x640	16	0.597	0.623	0.479	0.3981	8.5214
yolov8s	640x640	32	0.705	0.540	0.449	0.2100	10.0302
yolov8m	640x640	32	0.486	0.339	0.163	-0.6233	16.4433

A análise visual dos resultados, ilustrada na Figura 1, ajuda a explicar o desempenho numérico do modelo. A observação de que o sistema consegue identificar corretamente as vagens em condições ideais, por exemplo, justifica seu nível de *precision* **0.711**, pois as detecções que ele efetivamente realiza são, em sua maioria, acertadas. No entanto, algumas vagens não são identificadas, especialmente quando estão muito agrupadas, sobrepostas ou em partes da imagem com sombra, o que explica diretamente por que a *Recall* **0.607** é mais limitada. São justamente esses erros, em que o modelo falha em detectar vagens existentes, que impedem a métrica final de *mAP50-95* de ser mais alta, apontando onde o modelo precisa ser aprimorado.

A validação final da capacidade de detecção do modelo foi realizada utilizando o conjunto de teste "Noroeste"(601 imagens). O modelo alcançou um R² de **0.4905**,

indicando uma correlação positiva entre os valores preditos e os reais. Adicionalmente, o MAE foi de **7.8715** vagens, quantificando a margem de erro média do sistema. Em conjunto, estes resultados, que complementam a análise de detecção da Tabela 1, validam a eficácia do modelo para a tarefa de detecção de vagens e grãos.

3. Considerações Finais

Este trabalho demonstrou a viabilidade de um modelo de visão computacional para detecção de vagens de soja. Dentre as configurações avaliadas, a YOLOv8m (batch size 8) alcançou o melhor desempenho de mAP50-95 com **0.484**. A análise dos resultados também revelou as atuais limitações do modelo. Observou-se que a performance do modelo diminui em cenários com vagens sobrepostas e sombreamento, mesmo em imagens de ambiente controlado. Esta constatação é um resultado importante, pois indica que o principal desafio não reside apenas na variabilidade do ambiente de campo, mas na complexidade de individualizar objetos próximos e similares. Em complemento aos trabalhos existentes na literatura, este trabalho realizou o processo de testes em um novo conjunto de dados, para testar principalmente o comportamento do modelo em relação a diferentes culturas de soja.

Essa é uma pesquisa em andamento, e os próximos passos incluem: (1) ampliar dataset de imagens, com foco em imagens com vagens agrupadas e sobrepostas. (2) investigar novas técnicas de pré-processamento de imagem para melhorar a detecção das vagens. (3) testar com maior variação de hiperparâmetros conforme disponibilizado pelo modelo. (4) realizar a contagem das vagens e grãos para realizar a estimativa de produção de soja. (5) avaliar a versão mais recente do modelo, o YOLO11.

Referências

- Embrapa (2025). Dados econômicos - portal embrapa. <https://www.embrapa.br/soja/cultivos/soja1/dados-economicos>. Acesso em: 01 maio 2025.
- Ferreira, J. A. (2024). O uso da inteligência artificial na agricultura. *Revista Multidisciplinar do Nordeste Mineiro*, 9. Acesso em: 24 maio 2025.
- Jia, X., Hua, Z., Shi, H., Zhu, D., Han, Z., Wu, G., and Deng, L. (2025). A soybean pod accuracy detection and counting model based on improved yolov8. *Agriculture*, 15(6):617.
- Jiang, T., Shao, M., Zhang, T., Liu, X., and Yu, Q. (2025). Soybean pod and seed counting in both outdoor fields and indoor laboratories using unions of deep neural networks. *arXiv preprint arXiv:2502.15286*.
- Xiang, S., Wang, S., Xu, M., Wang, W., and Liu, W. (2023). YOLO POD: a fast and accurate multi-task model for dense Soybean Pod counting. *Plant Methods*, 19(1):8.
- Yu, Z., Wang, Y., Ye, J., Liufu, S., Lu, D., Zhu, X., Yang, Z., and Tan, Q. (2024). Accurate and fast implementation of soybean pod counting and localization from high-resolution image. *Frontiers in Plant Science*, 15.