

Classificação e Detecção Inteligentes de Grãos para Agricultura Digital na Cultura do Milho

Arthur Uguen de Mendonça, Elloá B. Guedes

¹Laboratório de Sistemas Inteligentes
Escola Superior de Tecnologia
Universidade do Estado do Amazonas
Av. Darcy Vargas, 1200 – Manaus – Amazonas
{audm.snf21, ebgcosta}@uea.edu.br

Abstract. *This work addresses the tasks of corn kernel classification and detection in images using state-of-the-art Deep Learning-based Computer Vision models. The goal is to enable the development of solutions for Digital Agriculture. In the classification task, it was observed that the imbalance of the classes had a low impact on the good performance of the models. In the detection task, it was possible to surpass results from the literature (a percentage increase of 16,33 %) and also to assess the generalization in other scenarios. The results obtained contribute to the estimation of productivity and decision making in the context of corn cultivation.*

Resumo. *Este trabalho teve por objetivo abordar tarefas de classificação e detecção de grãos de milhos em imagens a partir de modelos do estado da arte de Visão Computacional com Deep Learning, viabilizando o desenvolvimento de soluções para Agricultura Digital. Na tarefa de classificação, observou-se que o desbalanceamento das classes contempladas teve baixo impacto no bom desempenho dos modelos, ao passo que na tarefa de detecção foi possível superar resultados de trabalhos da literatura (aumento percentual de 16,33 %) e também aferir a generalização em outros cenários. Os resultados obtidos colaboram para a estimativa de produtividade e tomada de decisão no contexto da cultura do milho.*

1. Introdução

O milho (*Zea mays*) é um cereal amplamente cultivado em todo o mundo, sendo extensivamente empregado na alimentação humana e na produção de ração animal em virtude de suas propriedades nutricionais [Cruz et al. 2008]. No Brasil, é o segundo grão mais cultivado e exportado, com duas safras anuais. Esse cultivo vem se beneficiando de novas tecnologias, da expansão das áreas de plantio e do aumento da produtividade, permitindo assim uma maior participação do País no mercado internacional com esta *commodity* agrícola [Souza et al. 2018].

Segundo o monitoramento da série histórica da produção de grãos no Brasil feito pela Companhia Nacional de Abastecimento (CONAB), em 2023 o País produziu mais de 110 Mi t de milho, um aumento percentual da ordem de 450 % nos últimos 40 anos [CONAB 2024]. A área plantada teve uma expansão de 34,97 % nesse mesmo período, o que ressalta as novas tecnologias como tendo um papel significativo para alcançar tal patamar de produtividade. Neste sentido, citam-se o uso de cultivares de alto potencial genético, o espaçamento reduzido associado à uma maior densidade no plantio, a melhoria na qualidade das sementes, o controle químico de doenças e a correção do solo [Souza et al. 2018]. O conhecimento, novas tecnologias, políticas públicas e empreendedorismo permitirão ao Brasil produzir sistematicamente

mais de 100 Mt de milho, promovendo a segurança alimentar da população e consolidando o País como um grande celeiro mundial de alimentos [Miranda et al. 2019].

Na contemporaneidade, a Agricultura Digital consiste no aprimoramento da produção agrícola em termos de eficiência, qualidade e sustentabilidade por intermédio de novas tecnologias, especialmente na esfera da Computação [Tang et al. 2002]. No contexto brasileiro, em particular, técnicas como Processamento de Imagens e a Visão Computacional (VC) são elencadas como componentes essenciais nos eixos estratégicos a serem desenvolvidos e implementados ao longo da próxima década, conforme indicado pela Embrapa [2014].

Nesse contexto, visando colaborar para o desenvolvimento de soluções de Agricultura Digital que potencializem a cultura do milho, especialmente no Brasil, este trabalho tem como objetivo apresentar resultados experimentais de problemas de VC aplicados à classificação e à detecção de grãos de milho. Os problemas de VC abordados neste trabalho incluem a categorização de grãos entre instâncias saudáveis ou defeituosas (doente ou incompleto, por exemplo) e a delimitação de grãos em uma espiga para fins de contagem. Estes resultados foram obtidos com modelos de *Deep Learning* do estado da arte e apresentam avanços em relação às contribuições existentes na literatura, com potencial para serem utilizados em sistemas de automação para classificação e contagem de grãos de milho, auxiliando na estimativa de qualidade e produtividade da cultura.

Para apresentar o que se propõe, este trabalho está organizado como segue: os trabalhos relacionados são apresentados na Seção 2; o material e os métodos utilizados encontram-se descritos na Seção 3; os resultados obtidos e a discussão sobre os mesmos estão dispostos na Seção 4; por fim, as considerações finais e perspectivas futuras são listadas na Seção 5.

2. Trabalhos Relacionados

A estimativa da produtividade da cultura do milho pode ser uma ferramenta valiosa para a tomada de decisões no manejo agrícola, como a determinação do momento ideal de colheita e o planejamento das estratégias de comercialização [Sasseron 2013]. A contagem de grãos de milho é feita tipicamente de maneira amostral, em que são contabilizados o número de fileiras de grãos por espiga de milho, o número de grãos por fileira e o peso médio de grãos por espiga, para obter estimativas da produtividade (em toneladas por hectare) de grandes áreas de plantio. Nesses métodos também são consideradas outras variáveis, tais como o espaçamento médio entre linhas e dados climáticos, mas cabe a ressalva de que não são exatos, visto que há influência de inúmeros fatores, inclusive de ordem subjetiva [Rodrigues et al. 2005].

Com vistas a obter uma compreensão abrangente do problema em questão, elucidar as metodologias das soluções do estado da arte, bem como identificar tendências e lacunas nas soluções propostas, foi efetuada uma análise dos resultados de pesquisas no Google Scholar entre os anos de 2020 a 2024 em que os critérios de inclusão focaram em trabalhos baseados em *Deep Learning* para a contagem automática de grãos de milho. A partir dos resultados preliminarmente obtidos, foi possível perceber que a literatura já contemplava soluções no espectro considerado.

O trabalho de Khaki *et al.* [2020] propôs uma solução baseada no uso de Redes Neurais Convolucionais (CNNs, do inglês *Convolutional Neural Networks*) combinadas ao algoritmo de supressão não-maxima para obtenção das caixas delimitadoras dos grãos. Os autores treinaram experimentalmente o modelo proposto a partir em uma base de dados contendo 6.978 amostras de grãos e 9.413 amostras de outros objetos. A etapa de testes foi efetuada em cinco imagens contendo uma ou mais espigas. A aferição dos resultados, por sua vez, foi feita usando uma métrica de regressão, verificando se a contagem de grãos proposta pelo modelo era

próxima ao quantitativo esperado [Khaki et al. 2020]. Essa abordagem de aferição de desempenho, porém, não contempla o grau de qualidade nas detecções positivas e pode beneficiar-se indevidamente das detecções falso positivas. O modelo proposto demanda exemplos que não são da classe de interesse do problema, o que pode ser de difícil obtenção e sujeito à subjetividade. Ademais, os autores reportaram um tempo de inferência de 5,79 s, valor considerado alto para aplicações de tempo real.

Gillette *et al.* [2023] propuseram um método portátil, acessível e de baixo custo para quantificar grãos de milho. O funcionamento da solução se dá a partir da captura de uma imagem panorâmica por um *smartphone* acoplado à uma plataforma giratória. A imagem adquirida é analisada por um software de código aberto e com Inteligência Artificial proposto pelos autores. A solução proposta demonstrou ser precisa (93,7 %) na contagem de grãos de milho com padronagem homogênea e também obteve bons resultados (de 84,8 % a 61,8 %) em cenários com diferentes padrões. Além disso, o método possibilitou a estimativa rápida da produção de grãos e a categorização de diferentes tipos de grãos, como amido, doce e glutinoso, demonstrando potencial para estudos de genes que controlam a cor e a textura [Gillette et al. 2023]. Embora seja portátil e economicamente acessível, algumas das limitações da solução proposta por estes autores estão associadas ao hardware para captura das imagens panorâmicas, desde sua aquisição, *expertise* para manipulação correta e limitação de análise de uma espiga por vez.

A partir de ponderações na literatura sobre a ausência de dados rotulados sobre grãos de milho para experimentos com soluções do estado da arte de VC [Khaki et al. 2020], um conjunto de dados abrangente e com o propósito de padronizar e ampliar pesquisas no tema foi disponibilizado por Hobbs e outros [2021], o qual contemplava imagens diversificadas e rotuladas de grãos de milho. Além da proposição da base de dados, os autores também efetuaram comparações preliminares entre abordagens baseadas em *Deep Learning* para detecção, incluindo Faster R-CNN e YOLO, e em estimativa de densidade. A partir dos resultados obtidos, foi possível destacar o modelo YOLOv5 como tendo melhor desempenho no problema, o qual se caracteriza por ser arquitetura leve e viável para dispositivos embarcados e que não apresentou perda de eficiência com o aumento do número de grãos por imagem [Hobbs et al. 2021]. Porém, desde a apresentação desses resultados, novas versões da YOLO foram propostas na literatura, lacuna esta que também foi investigada no escopo deste trabalho almejando ganhos na tarefa, como será proposto e apresentado nas seções a seguir.

3. Material e Métodos

Com foco na cultura do milho, este trabalho propõe a aplicação de técnicas de VC com *Deep Learning*, sob o paradigma de Aprendizado Supervisionado, para realizar tarefas de avaliação da qualidade e contagem de grãos. Nas seções a seguir encontram-se os dados experimentais, a descrição das tarefas elencadas e dos modelos considerados para as mesmas, bem como as métricas de avaliação de desempenho.

3.1. Dados Experimentais

A base de dados utilizada no escopo deste trabalho é denominada *Corn Kernel Counting* [Hobbs et al. 2021] e contempla 402 imagens de espigas de milho, as quais foram rotuladas por especialistas. O *dataset* se subdivide nas seguintes partes: (i) *Base*, com imagens de diversas aparências e condições de captura; (ii) *Narrow*, com imagens de espigas de tipo único, em que as imagens favorecem condições ideais para detecção automática; e (iii) *Many*, a qual contempla imagens com múltiplas espigas. Em virtude dos exemplos em *Base* serem mais

abrangentes, esta subdivisão foi considerada no escopo experimental deste trabalho, exceto quando explicitamente mencionado.

Uma primeira preocupação prática no contexto da cultura do milho é saber se um determinado grão está em boas condições ou não. Para tanto, a partir das coordenadas disponíveis em *Base*, as imagens ali disponíveis foram recortadas e cada parte foi rotulada conforme as seguintes categorias, como ilustrado na Figura 1: (1) grão saudável; (2) grão doente; (3) grão incompleto, no tocante à maturação; (4) ponta estéril, em que o topo da espiga não produz grãos; e (5) “área da espiga”, em que observou-se a presença de grãos na espiga, mas cujas condições de captura da imagem impediram uma anotação mais precisa. As imagens obtidas e rotuladas nesse contexto serão sujeitas à uma tarefa de classificação.

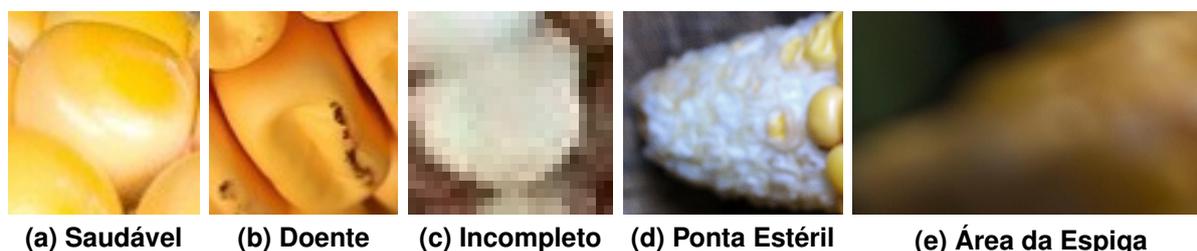


Figura 1: Exemplos aleatoriamente selecionados de cada classe da base de dados.
Fonte: [Hobbs et al. 2021].

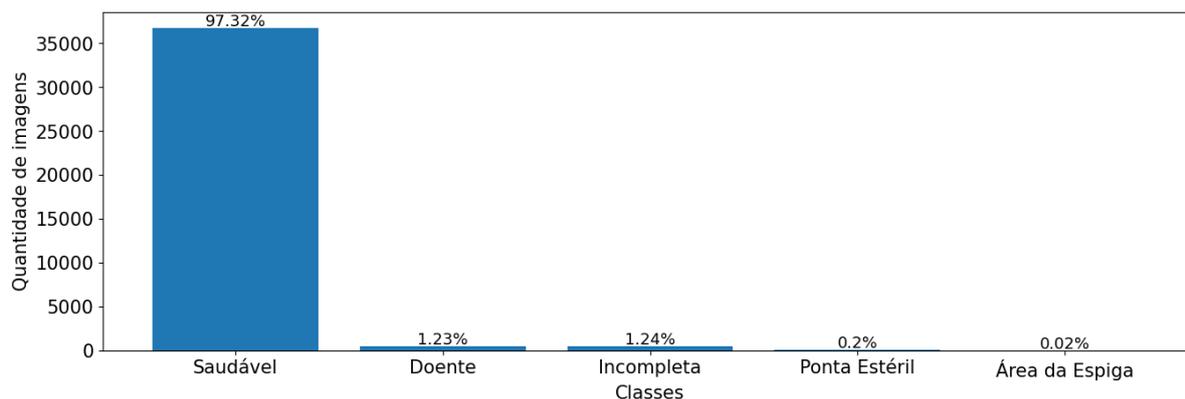
O histograma da Figura 2a ilustra o quantitativo e a distribuição dos exemplos para a tarefa de classificação segundo uma abordagem tradicional. A partir de uma inspeção visual, observou-se que as classes não estavam distribuídas de maneira equitativa, o que ensejou a adoção de uma técnica de subamostragem da classe predominante de grãos saudáveis, conforme Figura 2b, em que descartou-se aleatoriamente exemplos da classe mais prevalente (saudável) até que este quantitativo fosse igual ao da segunda classe mais frequente (incompleta). Como resultado, a classe de exemplos saudáveis teve uma redução percentual de 98,72 %, com um total de 468 exemplos, e consolidou-se uma tarefa de classificação segundo uma abordagem de subamostragem.

Para o cenário de detecção, contabilizou-se o número total de caixas delimitadoras por imagem e sua distribuição, conforme Figura 3. Percebe-se uma prevalência de imagens com até 356 caixas delimitadoras e alguns exemplos esporádicos com quantitativo superior. Em virtude do alto número de caixas por imagem e do tamanho dos grãos nas mesmas, ocupando poucos pixels, tem-se que esta tarefa é de natureza difícil por modelos de *Deep Learning*, como já reportado na literatura [Lin et al. 2014].

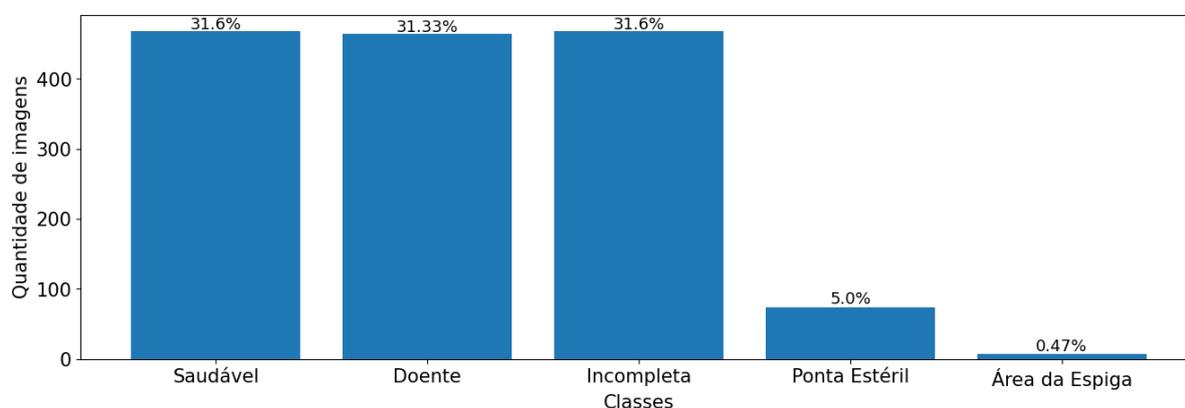
3.2. Tarefa de Classificação: Modelos e Parametrização

Para a tarefa de classificação, adotou-se como referência as CNNs, um tipo especializado de rede neural artificial projetado para processar dados de alta dimensionalidade e que utilizam filtros convolucionais para integrar o contexto espacial, facilitando a extração de características discriminativas [Goodfellow et al. 2016]. As seguintes arquiteturas de CNNs foram escolhidas:

1. **MobileNetV2.** Projetada para alta performance em dispositivos móveis, possui arquitetura similar à MobileNet original, diferenciando-se pela utilização de blocos residuais invertidos com características *bottleneck* e pela notável redução da quantidade de parâmetros quando comparada ao modelo anterior [Howard and Chen 2018];



(a) Abordagem Tradicional



(b) Abordagem de Subamostragem

Figura 2: Histograma dos exemplos para tarefa de classificação.

2. **InceptionV3**. Tem como principal característica o uso de módulos “Inception” para processamento de camadas convolucionais em paralelo. Aliada à incorporação da normalização de *batches* e convoluções fatorizadas, permitindo melhor eficiência durante o treinamento com número de parâmetros reduzidos [Szegedy et al. 2015];
3. **EfficientNetV2B0**. Com emprego de escalonamento composto para balancear a profundidade, largura e resolução da rede, é projetada para ser computacionalmente eficiente sem perder a acurácia competitiva [Tan and Le 2019];

Para todos os modelos, considerou-se a estratégia de Transferência de Aprendizado com a inicialização de pesos oriundos da base de dados ImageNet [Deng et al. 2009]; o treinamento considerou um número máximo de 300 épocas; adotou-se a técnica de *Early Stopping* com uma paciência de 30 épocas para evitar *overfitting*, em que foram monitoradas métricas no conjunto de validação; a técnica de *Model Checkpoint* para monitorar a acurácia no conjunto de validação e persistir em disco o conjunto de pesos que provia melhor generalização; taxa de aprendizado inicial igual a 10^{-3} ; uso da função de ativação ReLU e do otimizador Adam; e escolha do hiperparâmetro relativo ao tamanho dos *batches* deu-se de forma empírica para cada arquitetura, balanceando a quantidade de parâmetros e a utilização dos recursos computacionais disponíveis, a exemplo da memória principal.

3.3. Tarefa de Detecção

As arquiteturas de Redes Neurais Convolucionais Baseadas em Região (R-CNNs, do inglês *Region-Based Convolutional Neural Networks*) da Família YOLO (acrônimo para *You Only*

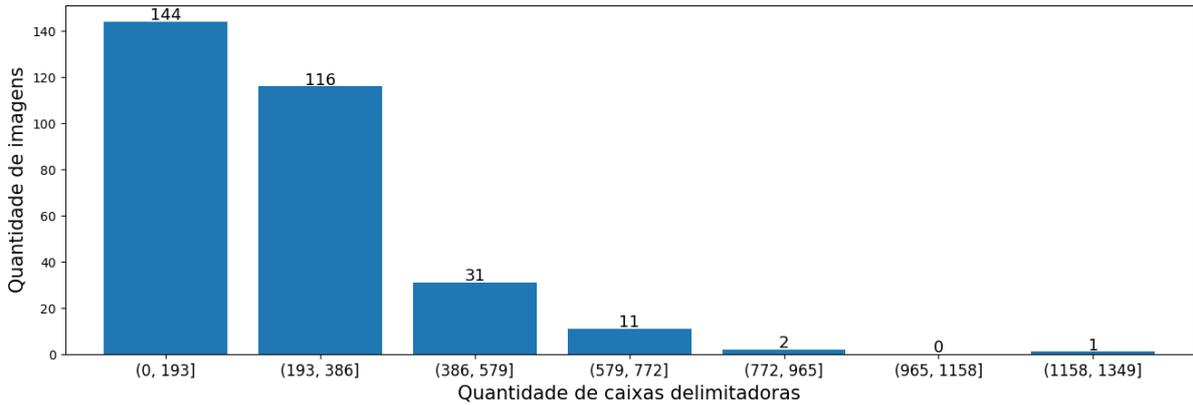


Figura 3: Número de caixas delimitadoras por imagem no conjunto *Base*.

Look Once) foram propostas originalmente em 2015 com o objetivo de unificar as tarefas necessárias para detecção (localizar e classificar os objetos) em uma única passagem (*single-shot*) da entrada pelo modelo. Ademais, reformularam o problema de detecção como uma tarefa de regressão, em que a rede prevê as coordenadas das caixas delimitadoras, seus coeficientes de confiança e também uma distribuição de probabilidades das classes à elas associados [Redmon et al. 2016]. Em virtude dessas características, os modelos da Família YOLO são considerados rápidos e acurados, de baixo custo computacional e atualmente constituem o estado da arte para este tipo de problema, com aplicações em diversos contextos [Michelucci 2019, Diwan et al. 2023].

Ao longo do tempo houve diversas melhorias na proposição seminal da YOLO que ensejaram versões sucessoras. A YOLOv8, iteração mais recente, baseia-se nos sucessos de seus antecessores enquanto introduz vários avanços importantes. Notavelmente, incorpora a arquitetura *Patch-based Transformer*, que utiliza mecanismos de autoatenção para capturar efetivamente dependências de longo alcance dentro de características. Essa inovação aprimora a capacidade do modelo de lidar com interações complexas de objetos e oclusões, levando a um melhor desempenho de detecção. Além disso, emprega uma nova função de perda diversificada que aborda a questão do desequilíbrio de classes prevalente em conjuntos de dados de detecção de objetos, função esta que prioriza o aprendizado de classes sub-representadas, resultando em modelos mais robustos e generalizáveis. De maneira geral, a YOLOv8 mostra-se adequado para aplicações em tempo real em várias plataformas e as melhorias implementadas contribuem para seu desempenho superior e aplicabilidade mais ampla [Jocher et al. 2023].

3.4. Avaliação de Desempenho

Para avaliar os modelos selecionados perante o problema de classificação, adotou-se a estratégia de validação cruzada do tipo *holdout* com estratificação, com 53,6% dos dados disponíveis para treino, 10% para validação e 26,4% para testes, em que esta última partição foi utilizada para aferir o desempenho conforme as seguintes métricas, em que C denota o conjunto das classes do problema:

$$\text{Acurácia} = \frac{1}{|C|} \sum_{c \in C} \left(\frac{\text{TP}_c + \text{FP}_c}{\text{TP}_c + \text{FP}_c + \text{FN}_c + \text{FP}_c} \right) \quad (1)$$

$$\text{Precisão} = \frac{1}{|C|} \sum_{c \in C} \left(\frac{\text{TP}_c}{\text{TP}_c + \text{FP}_c} \right) \quad (2)$$

$$\text{Revocação} = \frac{1}{|C|} \sum_{c \in C} \left(\frac{\text{TP}_c}{\text{TP}_c + \text{FN}_c} \right) \quad (3)$$

$$F_1\text{-Score} = 2 \left(\frac{\text{Precisão} \cdot \text{Revocação}}{\text{Precisão} + \text{Revocação}} \right) \quad (4)$$

Nas Eqs. (1)-(4), as siglas indicam os quatro resultados possíveis de uma tarefa de classificação binária, a citar: TP (do inglês, *True Positive*) demonstra o quantitativo das classificações corretas para a classe positiva; TN (do inglês, *True Negative*), as classificações corretas para a classe negativa; FP (do inglês, *False Positive*), os Erros Tipo I; e FN (do inglês, *False Negative*), os Erros Tipo II.

Para a tarefa de detecção foram consideradas as partições propostas pelos autores do *dataset* e a métrica mAP com *threshold* $t \geq 0.5$ para o IoU (*Intersection over Union*), a qual será denotada como mAP@0.5. Uma explicação detalhada do cálculo e semântica de tais métricas no contexto da detecção de objetos em VC encontra-se disponível no *survey* de Padilla *et al.* [2020]. Ademais, considerou-se também a métrica de regressão MAPE (do inglês, *Mean Absolute Percentage Error*) adaptada para o contexto da contagem de grãos, definida como segue:

$$\text{Count MAPE} = \frac{1}{J} \sum_{j=1}^J \frac{\text{Observado}_j - \text{Previsto}_j}{\text{Observado}_j}, \quad (5)$$

em que J denota o total de imagens no conjunto de testes de detecção, e Observado_j e Previsto_j são os números de grãos rotulados e previstos pelo modelo na j -ésima imagem, respectivamente.

4. Resultados e Discussão

A linguagem de programação Python, juntamente com os *frameworks* Tensorflow, Keras e Pytorch, foram as ferramentas primárias para treino e avaliação dos modelos. As implementações foram executadas em um servidor computacional equipado com processador Intel® Core™ i9-10900K com clock de 3,7 GHz, memória principal de 64 GB, 7 TB de memória secundária e 3 placas gráficas NVIDIA GTX 3060 com 12 GB VRAM cada, para promover a aceleração em hardware do treino dos modelos.

Para a tarefa de classificação, cujos resultados encontram-se sintetizados na Tabela 1, as métricas observadas na abordagem tradicional mostram-se ligeiramente superiores àquelas obtidas na abordagem de subamostragem, uma consequência direta do menor número de exemplos disponíveis para treinamento. Apesar dessa diferença, sua natureza discreta não permite constatar diferenças estatisticamente significativas, sendo os modelos equivalente entre si no tocante à classificação de grãos de milho em imagens realísticas nas cinco categorias propostas. A medida de tempo é feita a partir do início do treinamento até a parada com o Early Stopping. Para fins de utilização prática, recomenda-se o modelo MobileNetV2 por possuir o menor número de parâmetros dentre as arquiteturas consideradas, além de ser voltado para dispositivos embarcados.

No tocante ao treinamento e teste dos modelos YOLOv8, considerou-se a mesma metodologia de detectar apenas grãos saudáveis proposta por Hobbs e outros [2021]. Os resultados obtidos encontram-se dispostos na Tabela 2, juntamente com o *baseline* proposto na literatura. Optou-se por realizar três repetições experimentais de cada modelo para mitigar os efeitos da flutuação estocástica existente, ressaltando o quantidade de épocas necessárias para o treinamento em cada repetição.

Tabela 1: Resultados experimentais para a tarefa de classificação.

Arquitetura	Abordagem	Precisão	Revocação	F ₁ -Score	Acurácia	Épocas	Tempo
MobileNetV2	Tradicional	0,919	0,931	0,924	0,931	214	1.487 min
InceptionV3	Tradicional	0,920	0,952	0,934	0,952	203	1.358 min
EfficientNetV2B0	Tradicional	0,921	0,945	0,932	0,945	190	663 min
MobileNetV2	Subamostragem	0,919	0,932	0,925	0,932	126	160 min
InceptionV3	Subamostragem	0,918	0,909	0,913	0,909	108	141 min
EfficientNetV2B0	Subamostragem	0,919	0,849	0,882	0,849	83	55 min

Tabela 2: Resultados experimentais para a tarefa de detecção.

Arquiteturas	mAP@0.5	Count MAPE	Narrow Count MAPE	Parâmetros (M)	Épocas
YOLOv8 (Extra Large)	0.8235 ± 0.0080	0.22 ± 0.007	0.05 ± 0.0020	68.2	191 ± 74
YOLOv8 (Large)	0.8245 ± 0.0008	0.20 ± 0.014	0.05 ± 0.0040	43.7	209 ± 18
YOLOv8 (Medium)	0.8160 ± 0.0035	0.20 ± 0.007	0.05 ± 0.0004	25.9	200 ± 39
YOLOv8 (Small)	0.8040 ± 0.0032	0.22 ± 0.002	0.05 ± 0.0020	11.2	234 ± 52
YOLOv8 (Nano)	0.7690 ± 0.0039	0.24 ± 0.002	0.05 ± 0.0090	3.2	203 ± 35
YOLOv5 [Hobbs et al. 2021]	0.7000	0.24	0.17	n/c	n/c

A YOLOv8 *Large* apresentou o maior valor de mAP@0.5, sendo considerada a solução de referência para a tarefa de detecção. A Figura 4 mostra exemplos de detecção em testes, com múltiplos grãos detectados em condições não controladas de iluminação e fundo. Algumas detecções falso positivo também são observadas, evidenciando a dificuldade da tarefa devido à similaridade da base de dados com o contexto real do problema.

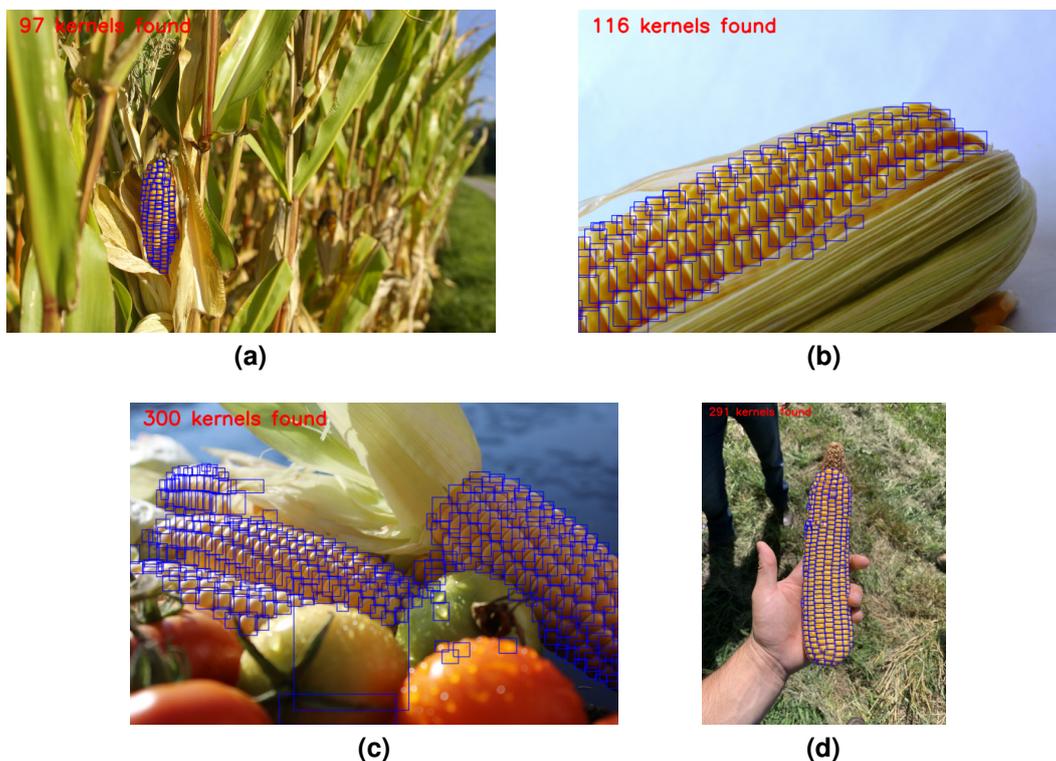


Figura 4: Exemplos de previsões do modelo no conjunto de testes.

É possível constatar que todos os modelos YOLOv8 superaram o *baseline* proposto no tocante ao mAP@0.5 e foram iguais ou melhores na métrica *Count MAPE* para o conjunto *Base*. Além disso, quando houve generalização para os exemplos do conjunto *Narrow*, também verificou-se melhorias significativas nesta última métrica, o que é sugestivo de boas capacidades de generalização dos modelos propostos.

5. Considerações Finais

Este trabalho teve por objetivo abordar problemas de VC associados à cultura do milho, em particular, a classificação do tipo de grão e a detecção e contagem de grãos saudáveis. Em ambas as tarefas foram utilizados modelos do estado da arte de *Deep Learning* e, em particular, na tarefa de detecção foi possível observar melhorias em relação ao disposto por Hobbs e outros [2021], o que contribui positivamente na obtenção de estimativas de produtividade desta *commodity* agrícola no tocante à contagem de grãos, colaborando com os objetivos da Agricultura Digital. Além disso, os modelos propostos podem ser embarcados em dispositivos móveis, vindo a compor soluções de Computação de Borda, em Névoa ou Nuvem no contexto da Agricultura 4.0 [Pachouri et al. 2023].

Em trabalhos futuros almeja-se realizar a detecção multi-classe, favorecendo a identificação de espigas não-conformes, além da implementação da solução proposta em um dispositivo computacional de baixo custo para testes em campo com a colaboração de um especialista da área agrícola, visando avaliar o desempenho da solução em condições práticas.

Referências

- CONAB (2024). Planilhas de custo de produção – milho. Disponível em <https://www.conab.gov.br/info-agro/custos-de-producao/planilhas-de-custo-de-producao/itemlist/category/821-milho>. Acesso em 25 de maio de 2024.
- Cruz, J. C., Karam, D., Magalhães, P. C., and Monteiro, M. A. R. (2008). *A Cultura do Milho*. Embrapa.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, Miami, FL, USA. IEEE.
- Diwan, T., Anirudh, G., and Tembhurne, J. V. (2023). Object Detection Using YOLO: Challenges, Architectural Successors, Datasets and Applications. *Multimed. Tools Appl.*, 82(6):9243–9275.
- Embrapa (2014). *Visão 2014–2034 – O Futuro do Desenvolvimento Tecnológico da Agricultura Brasileira (Síntese)*. Embrapa, Distrito Federal, Brasil.
- Gillette, S., Yin, L., Kianian, P. M. A., Pawlowski, W. P., and Chen, C. (2023). Corn360: a method for quantification of corn kernels. *Plant Methods*, 19(1):23.
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. The MIT Press, Cambridge.
- Hobbs, J., Khachatryan, V., Anandan, B. S., Hovhannisyan, H., and Wilson, D. (2021). Broad dataset and methods for counting and localization of on-ear corn kernels. *Front. Robot. AI*, 8:627009.
- Howard, A., S. M. Z. M. Z. A. and Chen, L.-C. . (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, page 4510–4520.

- Jocher, G., Chaurasia, A., and Qiu, J. (2023). YOLOv8 – real-time object detection. Disponível em <https://docs.ultralytics.com/models/yolov8/>. Acesso em 25 de maio de 2024.
- Khaki, S., Pham, H., Han, Y., Kuhl, A., Kent, W., and Wang, L. (2020). Convolutional neural networks for image-based corn kernel detection and counting. *Sensors (Basel)*, 20(9):2721.
- Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft COCO: Common Objects in Context. In *Proceedings of the 13th European Conference on Computer Vision (ECVV 2014)*, pages 740–755, Suíça. Springer.
- Michelucci, U. (2019). *Advanced Applied Deep Learning: Convolutional Neural Networks and Object Detection*. Apress, Suíça.
- Miranda, R. A. d., Durães, F. O. M., Garcia, J. C., Parentoni, S., Santana, D. P., Purcino, A. C., and Alves, E. (2019). Supersafra de milho e o papel da tecnologia no aumento da produção. *Revista de Política Agrícola*, 28(2):149.
- Pachouri, V., Pandey, S., Gehlot, A., Negi, P., Chhabra, G., and Joshi, K. (2023). Agriculture 4.0: Inculcation of big data and internet of things in sustainable farming. In *2023 IEEE International Conference on Contemporary Computing and Communications (InC4)*, volume 1, pages 1–4, Bangalore, India. IEEE.
- Padilla, R., Netto, S. L., and da Silva, E. A. B. (2020). A Survey on Performance Metrics for Object-Detection Algorithms. In *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, pages 237–242, Niterói, Brasil.
- Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, Estados Unidos. IEEE.
- Rodrigues, V. d. N., Von Pinho, R. G., Paglis, C. M., Bueno Filho, J. S. d. S., and Brito, A. H. d. (2005). Comparação entre métodos para estimar a produtividade de grãos de milho. *Ciênc. Agrotecnologia*, 29(1):34–42.
- Sasseron, J. C. (2013). Modelos matemáticos para previsão da produtividade do milho em dois sistemas de cultivo. Disponível em <http://tede2.unifenas.br:8080/jspui/handle/jspui/15>. Acesso em 25 de maio de 2024.
- Souza, A. E. d., Dos Reis, J. G. M., Raymundo, J. C., and Pinto, R. S. (2018). Estudo da produção do Milho no Brasil. *S. Am. Dev. Soc. J.*, 4(11):182.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 1–9, Massachusetts, Estados Unidos. IEEE.
- Tan, M. and Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97, pages 6105–6114.
- Tang, S., Zhu, Q., Zhou, X., Liu, S., and Wu, M. (2002). A Conception of Digital Agriculture. In *IEEE International Geoscience and Remote Sensing Symposium*, pages 3026–3028, Canada. IEEE.