

**INSTITUTO FEDERAL
DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA**
Goiás

Bacharelado em Sistemas de Informação

Matheus Meireles Guerra

Detecção de Grãos de Soja Pós-Colheita Utilizando Visão Computacional e Aprendizado Profundo

Luziânia

2026

TERMO DE AUTORIZAÇÃO PARA DISPONIBILIZAÇÃO NO REPOSITÓRIO DIGITAL DO IFG - ReDi IFG

Com base no disposto na Lei Federal nº 9.610/98, AUTORIZO o Instituto Federal de Educação, Ciência e Tecnologia de Goiás, a disponibilizar gratuitamente o documento no Repositório Digital (ReDi IFG), sem ressarcimento de direitos autorais, conforme permissão assinada abaixo, em formato digital para fins de leitura, download e impressão, a título de divulgação da produção técnico-científica no IFG.

Identificação da Produção Técnico-Científica

- | | |
|--|---|
| ☐ Tese | ☐ Artigo Científico |
| ☐ Dissertação | ☐ Capítulo de Livro |
| ☐ Monografia – Especialização | ☐ Livro |
| ☒ TCC - Graduação | ☐ Trabalho Apresentado em Evento |
| ☐ Produto Técnico e Educacional - Tipo: _____ | |

Nome Completo do Autor: Matheus Meireles Guerra

Matrícula: 20221080080167

Título do Trabalho: Detecção de Grãos de Soja Pós-Colheita Utilizando Visão Computacional e Aprendizado Profundo

Autorização - Marque uma das opções

1. (X) Autorizo disponibilizar meu trabalho no Repositório Digital do IFG (acesso aberto);
2. () Autorizo disponibilizar meu trabalho no Repositório Digital do IFG somente após a data ____/____/____(Embargo);
3. () Não autorizo disponibilizar meu trabalho no Repositório Digital do IFG (acesso restrito).

Ao indicar a opção **2 ou 3**, marque a justificativa:

- () O documento está sujeito a registro de patente.
() O documento pode vir a ser publicado como livro, capítulo de livro ou artigo.
() Outra justificativa: _____

DECLARAÇÃO DE DISTRIBUIÇÃO NÃO-EXCLUSIVA

O/A referido/a autor/a declara que:

- i. 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;
- ii. obteve autorização de quaisquer materiais inclusos no documento do qual não detém os direitos de autor/a, para conceder ao Instituto Federal de Educação, Ciência e Tecnologia de Goiás 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;
- iii. 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 de Goiás.



Documento assinado digitalmente
MATHEUS MEIRELES GUERRA
Data: 09/04/2026 05:27:50-0300
Verifique em <https://validar.iti.gov.br>

Luziânia/GO, 09/04/2026.
Local Data

Assinatura do Autor e/ou Detentor dos Direitos Autorais

Instituto Federal de Educação, Ciência e Tecnologia de Goiás
Bacharelado em Sistemas de Informação

Matheus Meireles Guerra

Detecção de Grãos de Soja Pós-Colheita Utilizando Visão Computacional e Aprendizado Profundo

Trabalho de Conclusão de Curso 2 apresentado à Coordenação do Curso de Bacharelado em Sistemas de Informação do Instituto Federal de Educação, Ciência e Tecnologia de Goiás - Câmpus Luziânia, como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação.

Orientador: Dr. Lucas de Almeida Ribeiro

Luziânia

2026

G934d Guerra, Matheus Meireles

Detecção de grãos de soja pós-colheita utilizando visão computacional e aprendizado profundo / Matheus Meireles Guerra. - Luziânia: IFG, 2026.
47 f. : il. color.

Orientador: Prof. Dr. Lucas de Almeida Ribeiro.
TCC (Graduação - Bacharelado em Sistemas de Informação). Instituto Federal de Goiás - IFG, Câmpus Luziânia, 2026.

1. Computação 2. Processamento de dados 3. Aprendizado profundo. 4 Agricultura de precisão I. Título. II. Ribeiro, Lucas de Almeida, orient.

CDD 004

ATA DE DEFESA DE TRABALHO DE CONCLUSÃO DE CURSO

Na presente data realizou-se a sessão pública de defesa do Trabalho de Conclusão de Curso intitulado **Deteção de Grãos de Soja Pós-Colheita Utilizando Visão Computacional e Aprendizado Profundo**, sob orientação de Lucas de Almeida Ribeiro, apresentado pelo aluno **Matheus Meireles Guerra (20221080080167)** do Curso **Bacharelado em Sistemas de Informação (Câmpus Luziânia)**. Os trabalhos foram iniciados às **16:00** do dia **17/12/2025** pelo Professor presidente da banca examinadora, constituída pelos seguintes membros:

- **Lucas de Almeida Ribeiro** (Presidente)
- **Daniel Vitor de Lucena** (Examinador Interno)
- **Eduardo Noronha de Andrade Freitas** (Examinador Externo)

A banca examinadora, tendo terminado a apresentação do conteúdo do Trabalho de Conclusão de Curso, passou à arguição do candidato. Em seguida, os examinadores reuniram-se para avaliação e deram o parecer final sobre o trabalho apresentado pelo aluno, tendo sido atribuído o seguinte resultado:

☒ Aprovado

☐ Reprovado

Nota : 9,3

Observação / Apreciações:

Corrigir as alterações sugeridas pela banca

Proclamados os resultados pelo presidente da banca examinadora, foram encerrados os trabalhos e, para constar, eu **Lucas de Almeida Ribeiro** lavrei a presente ata que assino juntamente com os demais membros da banca examinadora.

Documento assinado eletronicamente por:

- **Daniel Vitor de Lucena**, em **08/04/2026 22:01:07** com chave **971b70b633af11f1a2cc005056a537a4**.
- **Eduardo Noronha de Andrade Freitas**, em **08/04/2026 20:36:18** com chave **be396b6433a311f1911d005056a537a4**.
- **Lucas de Almeida Ribeiro**, em **08/04/2026 20:11:51** com chave **53c4054e33a011f193a5005056a537a4**.

Este documento foi emitido pelo SUAP. Para comprovar sua autenticidade, faça a leitura do QRCode ao lado ou acesse https://suap.ifg.edu.br/comum/autenticar_documento/ e informe os dados a seguir.

Tipo de Documento: Ata de Projeto Final

Data da Emissão: 08/04/2026

Código de Autenticação: 5ce125



Dedico este trabalho a Deus, por me conceder forças nos momentos de dificuldade, à minha família, pelo amor incondicional e apoio em todas as etapas da minha vida acadêmica, e aos amigos que estiveram ao meu lado durante essa jornada.

AGRADECIMENTOS

Agradeço sinceramente a todos que, direta ou indiretamente, contribuíram para a realização deste trabalho. Sem o apoio e a colaboração de cada um, este projeto não teria sido possível.

Em primeiro lugar, a Deus, à minha mãe Luciana Meireles Guerra, ao meu pai Paulo Cesar Guerra e à minha namorada Ingrid Vitória Ferreira Goulart, pelo amor incondicional, pela compreensão e pelo suporte inestimável ao longo de toda a minha jornada acadêmica.

Agradeço ao meu orientador, Prof. Dr. Lucas de Almeida Ribeiro, pela confiança, pela orientação precisa, pela paciência e pelo incentivo constante. Seus conhecimentos e sua dedicação foram fundamentais para o desenvolvimento e a conclusão desta pesquisa.

Ao Instituto Federal de Educação, Ciência e Tecnologia de Goiás, pela oportunidade de desenvolver este trabalho e por todo o suporte acadêmico e infraestrutura oferecidos.

Aos professores do curso de Bacharelado em Sistemas de Informação, pelos ensinamentos e pela formação sólida que me prepararam para este desafio.

Aos meus colegas e amigos, pelo companheirismo, pelas discussões enriquecedoras e pelo apoio nos momentos de dificuldade.

Por fim, a todos que, de alguma forma, contribuíram para a minha formação e para a concretização deste sonho. Meu muito obrigado.

*"Problemas não são obstáculos, mas oportunidades ímpares de superação e evolução."
(Maurício Rodrigues de Moraes)*

RESUMO

Guerra, M. M.. **Detecção de Grãos de Soja Pós-Colheita Utilizando Visão Computacional e Aprendizado Profundo**. Luziânia, 2026. 47p. Trabalho de Conclusão de Curso 2. Instituto Federal de Educação, Ciência e Tecnologia de Goiás

Este trabalho apresenta o desenvolvimento e validação de um sistema de visão computacional baseado em Aprendizado Profundo para a detecção automática de perdas de grãos de soja no solo pós-colheita. A identificação ágil dessas perdas é um desafio crítico da agricultura de precisão, tradicionalmente realizada através de métodos manuais que se mostram demorados, imprecisos e subjetivos. Propôs-se uma solução utilizando as arquiteturas YOLOv5 e YOLOv11 com o objetivo de identificar e localizar automaticamente os grãos dispersos no solo em diferentes condições de campo. A metodologia envolveu a coleta de 230 imagens de alta resolução em ambiente real, anotação manual das amostras, aplicação de *pipeline* rigoroso de pré-processamento e aumento de dados, além de treinamento e otimização de múltiplos modelos variando arquitetura, resolução e conjunto de dados. Os resultados demonstraram que o modelo YOLOv11s treinado com *Dataset V2* alcançou um mAP@0.5 de 92,7% e uma Revocação de 86,2%, superando significativamente a baseline YOLOv5. A validação qualitativa em quatro cenários desafiadores mostrou que o modelo proposto é robusto em condições reais de campo, incluindo baixo contraste tonal, fundos complexos e variações de escala. Este estudo contribui para a otimização da gestão agrícola, oferecendo uma ferramenta viável e eficiente para a redução de perdas e aumento da sustentabilidade na cultura da soja.

Palavras-chave: Visão Computacional; Aprendizado Profundo; YOLOv5; YOLOv11; Detecção; Soja; Agricultura de Precisão.

ABSTRACT

This work presents the development and validation of a deep learning-based computer vision system for automatic detection of soybean grain losses on the ground after harvest. Accurately monitoring these losses is a critical challenge in precision agriculture, traditionally performed through manual methods that prove time-consuming, imprecise, and subjective. A solution was proposed using YOLOv5 and YOLOv11 architectures with the objective of automatically identifying and locating grains scattered on the soil under different field conditions. The methodology involved collecting 230 high-resolution images in a real-world environment, manual annotation of samples, application of a rigorous preprocessing and data augmentation pipeline, and training and optimization of multiple models varying architecture, resolution, and dataset. The results demonstrated that the YOLOv11s model trained with Dataset V2 achieved an mAP@0.5 of 92.7% and a Recall of 86.2%, significantly surpassing the YOLOv5 baseline. Qualitative validation in four challenging scenarios showed that the proposed model is robust under real field conditions, including low tonal contrast, complex backgrounds, and scale variations. This study contributes to the optimization of agricultural management, offering a viable and efficient tool for reducing losses and increasing sustainability in soybean cultivation.

Keywords: Computer Vision; Deep Learning; YOLOv5; YOLOv11; Detection; Soybean; Precision Agriculture.

LISTA DE ILUSTRAÇÕES

Figura 1 –	Curvas de aprendizado (mAP@0.5:0.95) dos 5 melhores experimentos. Nota-se a superioridade e estabilidade do modelo YOLOv11s com <i>Dataset</i> V2 (Exp 9 - Linha Verde) em comparação ao baseline YOLOv5 (Exp 1 - Linha Vermelha).	35
Figura 2 –	Comparativo de Matrizes de Confusão Normalizadas . A visualização em percentual permite a comparação direta entre os modelos independentemente do tamanho dos conjuntos de validação.	36
Figura 3 –	Cenário 1 (Condições Ideais): Com boa iluminação e baixa altura, ambos os modelos detectam a maioria dos grãos. Nota-se, contudo, que o YOLOv11 apresenta uma taxa de acerto de 100% para grãos e apenas 1 falso positivo.	39
Figura 4 –	Cenário 2 (Baixo Contraste Tonal): Situação onde a cor do solo é muito próxima à cor do grão. O modelo baseline (a) falha em separar o objeto do fundo devido à semelhança de cores. O modelo proposto (b) demonstra maior sensibilidade, identificando os grãos mesmo sem um contraste de cor evidente.	40
Figura 5 –	Cenário 3 (Fundo Complexo): Em meio a restos culturais (palhada), o YOLOv5 gera falsos positivos (confunde terra com soja). O YOLOv11 discrimina corretamente as texturas, mantendo a identificação fidedigna.	41
Figura 6 –	Cenário 4 (Variação de Escala): Imagem capturada a maior altura. A redução na resolução dos grãos impede a detecção pelo modelo antigo, enquanto a nova arquitetura preserva as características espaciais necessárias para a identificação.	42

LISTA DE TABELAS

Tabela 1 – Configurações dos Experimentos Realizados	33
Tabela 2 – Resultados Comparativos das Métricas de Desempenho	34

LISTA DE ABREVIATURAS E SIGLAS

AMP	<i>Automatic Mixed Precision</i> (Precisão Mista Automática);
CNN	<i>Convolutional Neural Network</i> (Redes Neurais Convolucionais);
CPU	<i>Central Processing Unit</i> (Unidade Central de Processamento);
DL	<i>Deep Learning</i> (Aprendizado Profundo);
EXIF	<i>Exchangeable Image File Format</i> ;
FN	Falsos Negativos;
FP	Falsos Positivos;
GPU	<i>Graphics Processing Unit</i> (Unidade de Processamento Gráfico);
GUI	<i>Graphical User Interface</i> (Interface Gráfica do Usuário);
IA	Inteligência Artificial;
IDE	<i>Integrated Development Environment</i> (Ambiente de Desenvolvimento Integrado);
IoU	<i>Intersection over Union</i> (Interseção sobre União);
mAP	<i>Mean Average Precision</i> (Média da Precisão Média);
ML	<i>Machine Learning</i> (Aprendizado de Máquina);
MP	Megapixels;
PDI	Processamento Digital de Imagens;
RGB	<i>Red, Green, Blue</i> (Vermelho, Verde, Azul);
RNA	Redes Neurais Artificiais;
SVM	<i>Support Vector Machine</i> (Máquina de Vetores de Suporte);
TCC	Trabalho de Conclusão de Curso;
VC	Visão Computacional;
VP	Verdadeiros Positivos;
VRAM	<i>Video Random Access Memory</i> (Memória de Vídeo);
YOLO	<i>You Only Look Once</i> (Você só olha uma vez);

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Problema e Hipótese	14
1.2	Objetivos	15
1.2.1	Objetivos Específicos	15
1.3	Justificativa	16
1.4	Metodologia	16
1.5	Organização do trabalho	17
2	FUNDAMENTAÇÃO TEÓRICA	18
2.1	Inteligência Artificial (IA)	18
2.1.1	Aprendizado de Máquina (<i>Machine Learning</i>)	18
2.1.2	Aprendizado Profundo (<i>Deep Learning</i>)	19
2.1.2.1	Redes Neurais Convolucionais (CNNs)	20
2.2	Processamento Digital de Imagens (PDI)	21
2.3	Métricas e Conceitos Técnicos Fundamentais em Visão Computacional	22
2.4	Visão Computacional e <i>Deep Learning</i> na Agricultura	23
2.5	Aplicações na Agricultura de Precisão	24
2.6	Monitoramento de Perdas Pós-Colheita de Soja	24
2.6.1	Método Tradicional	24
2.6.2	Limitações	25
2.6.3	Necessidade de Automação	25
2.7	Arquitetura YOLO (You Only Look Once)	26
2.7.1	Histórico e Evolução	26
2.7.2	Justificativa da Escolha de YOLOv5 e YOLOv11	26
2.7.2.1	YOLOv5 como Prova de Conceito	26
2.7.2.2	YOLOv11 como Modelo Principal	27
2.7.3	Características Técnicas Principais	28
2.8	Trabalhos Relacionados	28
3	SOLUÇÃO PROPOSTA	30
3.1	Materiais e Ferramentas	30
3.2	Aquisição e Preparação dos <i>Datasets</i>	30
3.2.1	<i>Dataset</i> V1: roboflow_soja	31
3.2.2	<i>Dataset</i> V2: roboflow_soja_v2	31
3.2.3	<i>Pipeline</i> de Pré-processamento e Aumento de Dados	31

3.2.3.1	Etapa 1: Pré-processamento (Estático)	31
3.2.3.2	Etapa 2: Aumento de Dados (Dinâmico)	31
3.3	Arquiteturas e Configurações de Treinamento	32
3.4	Métricas de Avaliação	33
4	AVALIAÇÃO E DISCUSSÃO DOS RESULTADOS	34
4.1	Análise Gráfica e Visual	34
4.2	Discussão dos Resultados	36
4.2.1	Impacto da Arquitetura (V5 vs V11)	37
4.2.2	O Limite da Resolução	37
4.2.3	Superioridade do <i>Dataset</i> V2	37
4.2.4	Critérios de Seleção do Melhor Modelo	37
4.3	Análise Qualitativa (Inferência Visual)	38
4.3.1	Interpretação das Caixas Delimitadoras e Confidence Scores	38
4.3.2	Análise por Cenários	39
5	CONCLUSÃO	43
5.1	Contribuições	43
5.2	Trabalhos Futuros	44
5.3	Conclusão Final	45
	REFERÊNCIAS	46

1 INTRODUÇÃO

A agricultura moderna enfrenta o desafio constante de otimizar a produção e minimizar perdas, visando a sustentabilidade e a rentabilidade. No cultivo da soja, uma das culturas mais importantes globalmente, as perdas pós-colheita representam um problema significativo com grande impacto econômico (COSTA; GUILHOTO; BURNQUIST, 2015). Tradicionalmente, o monitoramento dessas perdas é realizado por métodos manuais que, apesar de amplamente utilizados, demandam tempo, apresentam baixa representatividade da área e são suscetíveis a erros e variabilidade nos resultados (PAIXÃO et al., 2022). Essa abordagem limita a capacidade dos agricultores de tomar decisões rápidas e eficazes para ajustar o processo de colheita, resultando em desperdício de recursos e prejuízos econômicos. Diante desse cenário, a busca por soluções inovadoras que integrem tecnologias avançadas se mostrou imperativa para aprimorar a eficiência e a precisão na gestão agrícola.

Neste contexto, a Visão Computacional e o Aprendizado Profundo (*Deep Learning*) emergem como ferramentas promissoras (LECUN; BENGIO; HINTON, 2015). A capacidade de sistemas de inteligência artificial em analisar e interpretar grandes volumes de dados visuais oferece uma oportunidade única para automatizar e otimizar a detecção e identificação de grãos de soja perdidos no campo (KAMILARIS; PRENAFETA-BOLDÚ, 2018). Este trabalho apresenta o desenvolvimento de um sistema baseado nessas tecnologias, visando fornecer uma estimativa rápida e visual das perdas, permitindo aos agricultores ajustar suas operações de colheita de forma mais eficiente e sustentável. A aplicação da família YOLO, originalmente apresentada como um sistema eficiente de detecção em tempo real, como visto por (REDMON et al., 2016), evoluiu através de versões subsequentes como YOLOv5 e YOLOv11, representando um avanço significativo em relação aos métodos convencionais, validando maior acurácia e escalabilidade para o monitoramento de grandes áreas cultivadas.

1.1 Problema e Hipótese

A avaliação e o monitoramento das perdas de grãos de soja no solo após a colheita têm sido desafios persistentes na agricultura moderna, influenciados por diversos fatores operacionais e ambientais (PAIXÃO et al., 2022). Atualmente, a metodologia predominante para estimar essas perdas baseia-se em técnicas manuais, que envolvem a coleta e contagem física de grãos em áreas amostrais. Este método possui várias limitações relevantes: é um processo demorado, trabalhoso e caro, além de apresentar grande variabilidade nos resultados. Estudos apontam ainda que a subjetividade na coleta e a dificuldade de padronização entre metodologias resultam em estimativas imprecisas e inconsistentes,

comprometendo a confiabilidade das medições (PAIXÃO et al., 2022). A falta de dados em tempo real e de alta fidelidade impede que os agricultores identifiquem rapidamente as causas das perdas (como desajustes na velocidade e regulagem da colheitadeira) e implementem ações corretivas de forma ágil (BOCK et al., 2020), culminando em prejuízos econômicos significativos e um impacto negativo na sustentabilidade da produção agrícola.

A hipótese central que norteou este trabalho foi a de que seria possível desenvolver um sistema de visão computacional, baseado em técnicas de Aprendizado Profundo (*Deep Learning*) e detecção de objetos, capaz de identificar e localizar automaticamente os grãos de soja perdidos no solo pós-colheita a partir de imagens digitais (RUSSELL; NORVIG, 2020). Pressupõe-se que tal sistema proporcionaria uma estimativa de perdas mais precisa, rápida e objetiva em comparação com os métodos manuais tradicionais, permitindo uma tomada de decisão mais eficiente para a otimização do processo de colheita e a redução do desperdício.

1.2 Objetivos

O objetivo geral deste trabalho foi desenvolver um sistema de visão computacional baseado em técnicas de Aprendizado Profundo para a detecção e identificação automática de grãos de soja perdidos no solo pós-colheita, visando fornecer uma ferramenta precisa e eficiente para a otimização da gestão agrícola e a redução de perdas.

1.2.1 Objetivos Específicos

Para alcançar o objetivo geral, este trabalho propôs-se a:

- Revisar e analisar a literatura científica existente sobre visão computacional, aprendizado profundo e suas aplicações na agricultura de precisão, com foco na detecção de objetos em ambientes agrícolas.
- Coletar e organizar um *dataset* representativo de imagens de grãos de soja no solo pós-colheita, realizando a anotação manual dos grãos e a aplicação de técnicas de aumento de dados para o treinamento do modelo.
- Treinar, otimizar e comparar modelos de detecção de objetos baseados nas arquiteturas YOLOv5 e YOLOv11, utilizando o *dataset* anotado, para identificar com precisão os grãos de soja em diferentes condições de campo.
- Desenvolver e validar um *script* de inferência que utilize o melhor modelo treinado para processar novas imagens, realizar a detecção dos grãos e mapear suas respectivas localizações.

- Avaliar o desempenho do sistema proposto em termos de acurácia de detecção, precisão e revocação, comparando os resultados obtidos entre as diferentes arquiteturas e configurações de resolução.
- Propor diretrizes para a futura integração do sistema em uma interface de usuário, visando a aplicabilidade prática da solução no contexto agrícola.

1.3 Justificativa

A relevância deste trabalho reside na abordagem de um problema de grande impacto econômico e ambiental para a cadeia produtiva da soja ([Food and Agriculture Organization of the United Nations \(FAO\), 2019](#)). As perdas de grãos pós-colheita representam um desperdício significativo de recursos, incluindo sementes, fertilizantes, água e energia, além de contribuir para a redução da rentabilidade dos produtores. A avaliação imprecisa dessas perdas, baseada em métodos manuais, impede a identificação eficaz das causas e a implementação de medidas corretivas em tempo hábil.

O desenvolvimento de um sistema automatizado de detecção e identificação de grãos de soja utilizando visão computacional e aprendizado profundo oferece uma solução inovadora e tecnologicamente avançada para esse desafio ([PATRÍCIO; RIEDER, 2018](#)). Diferentemente de métodos tradicionais que demandam inspeção visual subjetiva, a abordagem proposta utiliza redes neurais profundas para extrair características dos grãos automaticamente, permitindo escalabilidade e reprodutibilidade. A automação do processo não apenas aumenta a precisão e a velocidade da identificação de perdas, mas também libera mão de obra para outras atividades, otimizando a gestão da propriedade rural. A capacidade de fornecer dados confiáveis sobre a eficiência da colheita permite aos agricultores ajustar suas máquinas e processos de forma proativa, minimizando o desperdício e maximizando a produtividade.

Além do benefício econômico direto para os produtores, este projeto contribui para a sustentabilidade da agricultura, ao reduzir o desperdício de alimentos e o impacto ambiental associado à produção ([GODFRAY et al., 2010](#)). Do ponto de vista científico e tecnológico, o trabalho explora a aplicação comparativa de técnicas de ponta em inteligência artificial, especificamente a evolução entre as gerações YOLOv5 e YOLOv11, em um domínio prático e de grande relevância, gerando conhecimento sobre a robustez dessas tecnologias para resolver problemas complexos do agronegócio ([JOCHER et al., 2020](#)).

1.4 Metodologia

Nesta seção é apresentado como este projeto foi desenvolvido para atingir os objetivos propostos.

- Na primeira etapa, foi realizada uma revisão da literatura para identificar as arquiteturas de redes neurais mais adequadas para a detecção de objetos em ambientes agrícolas, com foco na família YOLO.
- Na segunda etapa, foi coletado e estruturado um conjunto de dados (*dataset*) de imagens de grãos de soja no solo pós-colheita. Utilizou-se a plataforma Roboflow para a anotação manual e aplicação de técnicas de aumento de dados (*data augmentation*), gerando versões expandidas do *dataset* para melhor generalização.
- Na terceira etapa, foram realizados treinamentos experimentais comparativos utilizando as arquiteturas YOLOv5 e YOLOv11. Foram testadas diferentes resoluções de entrada e tamanhos de modelo (*Small* e *Medium*) para encontrar o equilíbrio ideal entre precisão e custo computacional.
- Na quarta etapa, avaliou-se o desempenho dos modelos treinados utilizando métricas estatísticas robustas, incluindo Precisão, Revocação, *F1-Score* e mAP (*Mean Average Precision*), além da análise de matrizes de confusão.
- Na quinta e última etapa, foi realizada a análise crítica dos resultados, comparando a eficácia das diferentes versões do modelo e consolidando as conclusões sobre a viabilidade da solução para a agricultura de precisão.

1.5 Organização do trabalho

O Capítulo 1 apresenta a introdução do trabalho, abordando o problema do monitoramento de perdas pós-colheita de soja, os objetivos da pesquisa, a justificativa para o desenvolvimento da solução, a metodologia empregada e uma visão geral das seções seguintes.

O Capítulo 2 apresenta a fundamentação teórica, explorando os conceitos de Inteligência Artificial, Aprendizado de Máquina, Aprendizado Profundo e Redes Neurais Convolucionais. Este capítulo também aborda o Processamento Digital de Imagens, métricas fundamentais em visão computacional, aplicações na agricultura de precisão, monitoramento de perdas pós-colheita, a arquitetura YOLO e trabalhos relacionados.

O Capítulo 3 descreve a solução proposta, detalhando os materiais e ferramentas utilizados, a aquisição e preparação dos *datasets*, o *pipeline* de pré-processamento e aumento de dados, as arquiteturas e configurações de treinamento, as métricas de avaliação, e a avaliação comparativa dos resultados obtidos nos experimentos.

O Capítulo 4 apresenta as conclusões do trabalho, sintetizando as principais contribuições da pesquisa, discutindo a validação da hipótese central, e sugerindo direções para trabalhos futuros visando a continuidade e aprimoramento desta pesquisa.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo estabelece os pilares teóricos que sustentam o desenvolvimento do sistema de detecção e localização de grãos de soja. Serão abordados os conceitos fundamentais de Inteligência Artificial, Processamento Digital de Imagens e Aprendizado Profundo, com foco nas Redes Neurais Convolucionais e na arquitetura YOLO. Adicionalmente, apresenta-se uma revisão de trabalhos relacionados que utilizam técnicas similares no contexto agrícola.

2.1 Inteligência Artificial (IA)

A Inteligência Artificial (IA) é um campo multidisciplinar da ciência da computação dedicado ao desenvolvimento de sistemas que simulam capacidades humanas, como raciocínio, aprendizado, percepção e tomada de decisão (RUSSELL; NORVIG, 2020). A disciplina busca criar agentes inteligentes capazes de interagir com o ambiente, processar informações e executar tarefas de forma autônoma.

A aplicação prática de IA varia amplamente. Por exemplo, sistemas de recomendação analisam preferências de usuários para sugerir produtos, enquanto robôs autônomos navegam ambientes complexos. No contexto agrícola, a IA emergiu como uma ferramenta transformadora, possibilitando otimização de processos, automação de tarefas e análise preditiva em larga escala (KAMILARIS; PRENAFETA-BOLDÚ, 2018; LIAKOS et al., 2018).

Essas tecnologias permitem aplicações que variam desde o monitoramento de saúde de plantas até a gestão eficiente de recursos. Diante dessa capacidade de processamento e análise, a IA se destaca como elemento chave para o desenvolvimento de soluções inovadoras em domínios específicos, incluindo a agricultura de precisão abordada neste trabalho.

2.1.1 Aprendizado de Máquina (*Machine Learning*)

O Aprendizado de Máquina (*Machine Learning* - ML) é definido como o campo de estudo que dá aos computadores a capacidade de aprender sem serem programados explicitamente para cada tarefa (GOODFELLOW; BENGIO; COURVILLE, 2016).

Na prática, ML é usado quando as regras são complexas ou desconhecidas. Por exemplo, identificar se um email é spam é uma tarefa onde seria impraticável codificar todas as variações manualmente. Em vez disso, o algoritmo observa centenas de emails rotulados (spam vs. legítimo) e aprende a distingui-los automaticamente. O desempenho

melhora conforme mais dados são fornecidos ao modelo (GOODFELLOW; BENGIO; COURVILLE, 2016).

No contexto agrícola, ML tem se mostrado especialmente valioso. Desde a previsão de rendimento até a detecção de doenças em folhas, os algoritmos aprendem padrões complexos que sequer humanos conseguem formalizar em regras explícitas.

O ML abrange diversas abordagens, cada uma adequada a diferentes problemas:

- **Aprendizado Supervisionado:** O modelo aprende a partir de dados rotulados, onde entrada e saída desejada são conhecidas. Exemplo: imagens de grãos (entrada) anotadas como “soja” ou “outros” (saída).
- **Aprendizado Não Supervisionado:** O modelo encontra padrões em dados sem rótulos, agrupando automaticamente similaridades. Exemplo: identificar clusters de comportamento de consumidor sem categorias pré-definidas.
- **Aprendizado por Reforço:** O modelo aprende através de interações com um ambiente, recebendo recompensas ou penalidades. Exemplo: um robô aprendendo a navegar em um campo.

Para a detecção de grãos de soja em solo pós-colheita, o aprendizado supervisionado é a abordagem mais relevante, pois as imagens são previamente anotadas (identificando grãos vs. *background*). Essa abordagem permite ao modelo aprender características visuais que distinguem grãos de outros elementos do ambiente.

2.1.2 Aprendizado Profundo (*Deep Learning*)

Yann LeCun é frequentemente creditado como um dos pioneiros do Aprendizado Profundo (*Deep Learning* - DL) ao desenvolver as primeiras Redes Neurais Convolucionais (CNNs) em 1998. Seu sistema LeNet foi capaz de reconhecer dígitos manuscritos automaticamente, um feito revolucionário que demonstrou o potencial de aprender características visuais em múltiplas camadas (LECUN; BENGIO; HINTON, 2015).

O Deep Learning foi revolucionário porque resolveu um problema crítico: extrair características relevantes dos dados de forma automática. Tradicionalmente, especialistas precisavam manualmente descrever o que procurar (ex: “procure por arestas, depois curvas, depois formas”). O DL aprende essas características diretamente dos dados brutos, em camadas hierárquicas progressivamente mais abstratas (GOODFELLOW; BENGIO; COURVILLE, 2016).

Um exemplo concreto: reconhecer um rosto em uma fotografia. As primeiras camadas aprendem características simples (bordas, texturas), camadas intermediárias aprendem formas (olhos, nariz), e camadas profundas aprendem conceitos abstratos (“semelhança com

face”). Essa capacidade de aprender representações hierárquicas e abstratas é a assinatura do DL.

O Aprendizado Profundo é caracterizado pela utilização de Redes Neurais Artificiais (RNAs) com múltiplas camadas ocultas. Essa arquitetura profunda permite que os modelos capturem padrões complexos em grandes volumes de dados, sendo particularmente eficazes em:

- **Processamento de Linguagem Natural:** Tradução automática, análise de sentimentos.
- **Reconhecimento de Fala:** Sistemas como Siri e Google Voice Assistant.
- **Visão Computacional:** Detecção de objetos, reconhecimento facial, e, crucialmente para este trabalho, detecção de grãos dispersos em imagens agrícolas.

A vantagem central do DL em relação a técnicas tradicionais é a eliminação da necessidade de engenharia de características manual. Onde especialistas precisavam descrever manualmente “procure por isto”, o DL aprende o que procurar automaticamente. Modelos modernos de detecção de objetos, como o YOLO baseado em CNNs profundas, são a espinha dorsal de sistemas de visão computacional atuais, permitindo identificação de objetos com alta precisão em tempo real.

2.1.2.1 Redes Neurais Convolucionais (CNNs)

Geoffrey Hinton e seus colaboradores revolucionaram a visão computacional em 2012 ao vencer a competição ImageNet com a arquitetura AlexNet, baseada em Redes Neurais Convolucionais (CNNs) ([KRIZHEVSKY; SUTSKEVER; HINTON, 2012](#)). O desafio era classificar imagens em 1.000 categorias diferentes, uma tarefa que algoritmos tradicionais não conseguiam fazer bem. AlexNet foi o primeiro modelo a provar que CNNs profundas funcionavam em dados reais e complexos, com uma margem de acurácia que deixou todos os concorrentes para trás.

O problema que CNNs resolvem é fundamental: como extrair características relevantes de uma imagem automaticamente? Métodos tradicionais pediam ao engenheiro que descrevesse manualmente o que procurar (“procure por arestas em 45 graus”). As CNNs aprendem isso automaticamente, através de camadas que progressivamente extraem características de baixo nível (bordas simples) até características de alto nível (objetos complexos) ([GOODFELLOW; BENGIO; COURVILLE, 2016](#)).

Uma CNN funciona assim. Imagine um filtro (*kernel*) deslizando sobre uma imagem, multiplicando seus valores pixel por pixel. Essa operação é a ****convolução****: ela extrai características locais como bordas ou texturas. Depois, a camada de ****pooling**** reduz o

tamanho da imagem, mantendo apenas as informações mais importantes (como máximos locais). Repetindo isso em múltiplas camadas, a rede constrói hierarquias de características: primeira camada aprende bordas, segunda aprende formas simples (círculos, linhas), terceira aprende objetos (olhos, rodas), quarta aprende conceitos (face, carro). Essa capacidade de aprender hierarquias automaticamente é a assinatura das CNNs.

As CNNs consistem tipicamente em três tipos de camadas:

- **Camadas Convolucionais:** Aplicam múltiplos filtros (*kernels*) para extrair características locais como bordas, texturas e formas. Exemplo: um filtro de detecção de bordas verticais.
- **Camadas de *Pooling* (Agrupamento):** Reduzem a dimensionalidade mantendo informações relevantes. Exemplo: *max-pooling* seleciona o valor máximo em janelas 2×2 , reduzindo tamanho à metade.
- **Camadas Totalmente Conectadas:** Conectam todos os neurônios da camada anterior (usadas no final para classificação). Exemplo: 1000 neurônios de saída para 1000 categorias ImageNet.

Essa arquitetura é inspirada no córtex visual de mamíferos, que também processa informações visuais em camadas hierárquicas (LECUN; BENGIO; HINTON, 2015). Desde AlexNet em 2012 (KRIZHEVSKY; SUTSKEVER; HINTON, 2012), CNNs se tornaram a abordagem dominante em visão computacional, gerando sucessores como VGG, ResNet e MobileNet, cada um otimizado para um cenário diferente.

Para problemas de detecção de objetos, não apenas classificação, mas também localização de onde os objetos estão em uma imagem, arquiteturas especializadas baseadas em CNNs foram desenvolvidas. Modelos de dois estágios como R-CNN (GIRSHICK et al., 2014) usam propostas de regiões de interesse primeiro. Modelos de um estágio como o YOLO (You Only Look Once) (REDMON et al., 2016) aplicam CNNs de forma mais eficiente, permitindo detecção em tempo real, exatamente o que é necessário para identificar grãos de soja em imagens de campo.

2.2 Processamento Digital de Imagens (PDI)

O Processamento Digital de Imagens (PDI) é uma área da ciência da computação que se dedica à manipulação e análise de imagens digitais por meio de algoritmos computacionais. As etapas típicas do PDI incluem aquisição de imagem, pré-processamento (para melhorar a qualidade da imagem, como ajuste de contraste, remoção de ruído e realce de bordas), segmentação (divisão da imagem em regiões ou objetos de interesse) e extração de características (GONZALEZ; WOODS, 2018). No contexto da visão computacional

e do aprendizado profundo, o PDI desempenha um papel fundamental na preparação dos dados visuais, garantindo que as imagens estejam em um formato adequado e com a qualidade necessária para serem processadas por modelos de IA (LECUN; BENGIO; HINTON, 2015). Para este trabalho, o PDI é essencial na fase de pré-processamento das imagens de soja, antes de serem alimentadas ao modelo de *Deep Learning*. Técnicas de PDI são aplicadas para padronizar as imagens, remover artefatos indesejados e realçar características relevantes, otimizando assim o desempenho dos algoritmos de detecção de objetos. Isso inclui operações como filtragem, binarização e transformações geométricas, que visam aprimorar a visibilidade dos grãos de soja em ambientes complexos como o solo pós-colheita.

2.3 Métricas e Conceitos Técnicos Fundamentais em Visão Computacional

Para avaliar o desempenho de modelos de detecção de objetos e compreender as características técnicas das arquiteturas utilizadas neste trabalho, é necessário dominar algumas métricas e componentes arquiteturais fundamentais. Os conceitos a seguir são recorrentes na análise de sistemas baseados em visão computacional e são essenciais para interpretar os resultados dos experimentos com YOLO:

- **FPS (*Frames Per Second* / Quadros Por Segundo):** Métrica que mede a velocidade de inferência de um modelo de rede neural, indicando quantas imagens podem ser processadas em um segundo. Valores mais altos indicam processamento mais rápido, essencial para aplicações em tempo real como detecção em vídeo ou monitoramento de colheita. Por exemplo, um modelo com 60 FPS processa uma imagem a cada 16,7 milissegundos (ULTRALYTICS, 2024b).
- **IoU (*Intersection over Union* / Interseção sobre União):** Métrica geométrica fundamental que mede a precisão da localização do objeto. Ela calcula a razão entre a área de sobreposição (Interseção) e a área total combinada (União) entre a caixa delimitadora predita pelo modelo e a caixa real anotada (*Ground Truth*). O IoU varia de 0 (nenhuma sobreposição) a 1 (sobreposição perfeita). É usado como limiar para decidir se uma detecção é válida (ex: $\text{IoU} > 0.5$) (REDMON et al., 2016).
- **mAP (*mean Average Precision* / Precisão Média Média):** Métrica de avaliação padrão em visão computacional que mede a qualidade das detecções. Varia de 0 a 100%, onde valores mais altos indicam melhor desempenho na identificação correta de objetos e localização de caixas delimitadoras. Um modelo com mAP de 95% acerta a detecção em 95% das imagens de teste (REDMON et al., 2016).

- **C3k2 Block (Bloco *Cross Stage Partial* com *kernel* 2):** Componente arquitetural introduzido no YOLOv11, baseado no conceito de CSP (*Cross Stage Partial*). Ao invés de uma única convolução grande, utiliza duas convoluções menores com *kernel* (filtro) de tamanho reduzido, diminuindo a carga computacional enquanto mantém ou melhora a representação de características. Essa configuração é particularmente efetiva na detecção de objetos pequenos e densos, como grãos de soja dispersos no solo (KHANAM; HUSSAIN, 2024).
- **Parâmetros do Modelo:** Número total de pesos treináveis na rede neural. Modelos com menos parâmetros são mais leves, rápidos em inferência e requerem menos memória durante o treinamento e execução. Por exemplo, o YOLOv11x possui 56,9 milhões de parâmetros contra 59,1 milhões no YOLOv12x, resultando em inferência aproximadamente 23% mais rápida (KHANAM; HUSSAIN, 2024).

Estes conceitos serão recorrentes ao longo do trabalho, especialmente nas discussões sobre a comparação entre modelos YOLO e análise de desempenho do sistema proposto.

2.4 Visão Computacional e *Deep Learning* na Agricultura

A integração da Visão Computacional (VC) e do *Deep Learning* (DL) tem provocado mudanças importantes na agricultura, permitindo a automação e a otimização de tarefas anteriormente realizadas por inspeção manual ou métodos menos eficientes (LIAKOS et al., 2018). A VC, por meio de câmeras e sensores, coleta dados visuais do ambiente agrícola, enquanto o DL interpreta esses dados para extrair informações relevantes. Essa combinação permite criar sistemas inteligentes capazes de monitorar a saúde das plantas, identificar pragas e doenças, classificar culturas, estimar rendimentos e otimizar a aplicação de insumos (KAMILARIS; PRENAFETA-BOLDÚ, 2018).

Diferentemente da agricultura de precisão tradicional, que se concentra na gestão da variabilidade espacial dos campos por meio de sensores convencionais, a aplicação conjunta de VC e DL possibilita percepção autônoma em tempo real do ambiente agrícola. Sistemas baseados em DL conseguem identificar deficiências nutricionais em folhas ou detectar ervas daninhas com precisão superior aos métodos visuais tradicionais, possibilitando intervenções localizadas e eficazes. A capacidade de processar grandes volumes de dados visuais de forma autônoma é um diferencial que essas tecnologias oferecem para uma agricultura mais eficiente e sustentável.

2.5 Aplicações na Agricultura de Precisão

A integração da Inteligência Artificial, especialmente o *Deep Learning* e a Visão Computacional, tem revolucionado a agricultura, impulsionando o conceito de Agricultura de Precisão. Esta área busca otimizar a gestão agrícola através da observação, medição e resposta à variabilidade de culturas e solos, visando maximizar a produtividade e minimizar o impacto ambiental (MAVRIDOU et al., 2019). As aplicações são vastas e incluem a detecção de doenças e pragas em plantas, a contagem de frutos, a classificação de culturas, o monitoramento do crescimento de plantas e a navegação de veículos autônomos (KAMILARIS; PRENAFETA-BOLDÚ, 2018). A capacidade de processar grandes volumes de dados visuais de forma autônoma e em tempo real permite que os agricultores tomem decisões mais informadas, otimizando o uso de insumos e reduzindo perdas.

No contexto específico da detecção de objetos em ambientes agrícolas, o *Deep Learning* tem demonstrado alta eficácia. Por exemplo, revisões como a de (ZHANG et al., 2020) compilam aplicações bem-sucedidas de CNNs para tarefas como a detecção de frutas em pomares, incluindo laranjas, e a identificação de ervas daninhas em lavouras. Por outro lado, trabalhos como o de (MOMIN et al., 2017) utilizaram visão computacional clássica para avaliar a qualidade de grãos de soja, por meio da segmentação de impurezas e grãos danificados. A capacidade de identificar e quantificar elementos específicos em imagens complexas, como grãos de soja dispersos no solo, é um diferencial que essas tecnologias oferecem. Este projeto se insere diretamente nesse cenário, adaptando e aplicando essas técnicas avançadas para auxiliar na avaliação visual de perdas de grãos de soja pós-colheita, contribuindo para uma agricultura mais eficiente e sustentável.

2.6 Monitoramento de Perdas Pós-Colheita de Soja

A avaliação das perdas durante a colheita mecanizada é uma prática de manejo essencial para garantir a eficiência econômica da lavoura. Estas perdas podem ocorrer devido a fatores naturais (deiscência das vagens) ou, mais frequentemente, devido a regulagens inadequadas dos mecanismos de corte, trilha e limpeza da colheitadeira (BOCK et al., 2020).

2.6.1 Método Tradicional

A metodologia padrão utilizada no Brasil para a quantificação dessas perdas é o "Método do Copo Medidor" ou o uso de armações de área conhecida (geralmente $1m^2$ ou $2m^2$). O procedimento consiste na interrupção da operação de colheita, seguida da demarcação de uma área no solo onde a máquina já passou. Um operador realiza a coleta manual de todos os grãos e vagens encontrados dentro dessa área delimitada (PAIXÃO et al., 2022).

Posteriormente, esses grãos são contados ou pesados, e o valor obtido é extrapolado para estimar a perda em sacas por hectare (sc/ha) ou quilogramas por hectare (kg/ha). Este processo busca monitorar a eficiência da máquina e indicar a necessidade de ajustes nos parâmetros operacionais, como a velocidade do molinete ou a abertura das peneiras.

2.6.2 Limitações

Apesar de amplamente difundido, o método tradicional apresenta gargalos significativos que comprometem a precisão dos dados e a agilidade na tomada de decisão:

- **Baixa Representatividade:** A área amostrada (alguns metros quadrados) é estatisticamente insignificante quando comparada à extensão total de talhões que podem chegar a centenas de hectares. Isso gera um viés de amostragem, pois a perda não é uniforme em todo o terreno (PAIXÃO et al., 2022).
- **Subjetividade e Erro Humano:** A contagem manual é uma tarefa repetitiva e exaustiva. Em condições de campo, com calor, poeira e palhada cobrindo o solo, é comum que o operador deixe passar grãos, subestimando a perda real.
- **Latência na Correção:** Como o processo exige a parada da máquina e a contagem manual, o diagnóstico é demorado. Muitas vezes, quando o resultado é obtido, uma grande área já foi colhida com a regulação ineficiente, consolidando o prejuízo (BOCK et al., 2020).

2.6.3 Necessidade de Automação

Diante das ineficiências do processo manual e do impacto econômico direto das perdas, que representam um desperdício de insumos produtivos como sementes, fertilizantes e água (COSTA; GUILHOTO; BURNQUIST, 2015), a automação do monitoramento torna-se imperativa.

A adoção de tecnologias de Agricultura de Precisão, especificamente sistemas baseados em Visão Computacional, oferece uma alternativa capaz de superar as limitações humanas. Um sistema automatizado permite a análise de uma área muito maior em menos tempo, eliminando a subjetividade da contagem manual. A capacidade de processar imagens em tempo real viabiliza a geração de mapas de perdas e fornece *feedback* instantâneo ao operador da colheitadeira, permitindo ajustes dinâmicos que minimizam o desperdício e maximizam a rentabilidade da produção (Food and Agriculture Organization of the United Nations (FAO), 2019).

2.7 Arquitetura YOLO (You Only Look Once)

O YOLO (*You Only Look Once*) é uma família de modelos de detecção de objetos em tempo real que revolucionou a área ao tratar a detecção como um problema único de regressão, prevendo caixas delimitadoras e probabilidades de classe simultaneamente (REDMON et al., 2016). Desde seu lançamento em 2015, o YOLO evoluiu através de diversas versões, cada uma introduzindo melhorias significativas em precisão, velocidade e eficiência.

2.7.1 Histórico e Evolução

O histórico de versões do YOLO demonstra uma evolução contínua da arquitetura: (ULTRALYTICS, 2024a)

- **YOLOv4 (abril 2020):** Focou em otimização de hiperparâmetros e mantinha 50 FPS em GPU Tesla P100, estabelecendo o estado da arte em tempo real.
- **YOLOv5 (junho 2020):** Democratizou a detecção de objetos ao priorizar a acessibilidade e versatilidade, atingindo 140 FPS com tamanho de modelo de apenas 27 MB. Rapidamente ganhou adoção massiva na comunidade de pesquisa devido à sua facilidade de uso (ULTRALYTICS, 2023a).
- **YOLOv8 (novembro 2023):** Introduziu um head de detecção livre de âncoras e o módulo backbone C2f, representando um redesenho significativo da arquitetura (ULTRALYTICS, 2023b).
- **YOLOv11 (setembro 2024):** A iteração mais recente, otimizada para maior precisão com menos parâmetros (22% redução em relação ao YOLOv8m) e melhor desempenho em CPU (ULTRALYTICS, 2024b).
- **YOLOv12 (novembro 2024):** Introduz mecanismos de atenção centrados em avanços teóricos, mas apresenta instabilidade de treinamento e consumo elevado de memória, sendo ainda considerada experimental (ULTRALYTICS, 2024c).

2.7.2 Justificativa da Escolha de YOLOv5 e YOLOv11

Este trabalho utiliza YOLOv5 como *baseline* comparativo e YOLOv11 nos experimentos finais, escolha fundamentada em evidências técnicas:

2.7.2.1 YOLOv5 como Prova de Conceito

O YOLOv5 foi selecionado para a fase inicial do projeto (TCC1) pelos seguintes motivos:

- **Maturidade Comprovada:** Lançado em junho de 2020, possui ecossistema maduro com comunidade ativa e numerosas integrações de terceiros ([ULTRALYTICS, 2023a](#)).
- **Acessibilidade:** Oferece processo de treinamento eficiente com pesos pré-treinados prontamente disponíveis, permitindo ciclos de desenvolvimento rápidos e menor demanda de memória ([ULTRALYTICS, 2023a](#); [ULTRALYTICS, 2024d](#)).
- **Documentação Abrangente:** Dispõe de tutoriais passo a passo, guias de otimização de hiperparâmetros e exemplos práticos bem estabelecidos ([ULTRALYTICS, 2023a](#)).
- **Viabilidade para Prototipagem:** Seu tamanho compacto (27 MB) e velocidade (140 FPS em GPU Tesla P100) a tornam ideal para validar metodologias sem demandas computacionais excessivas.

2.7.2.2 YOLOv11 como Modelo Principal

O YOLOv11 foi adotado para os experimentos finais (TCC2) pelos seguintes motivos:

- **Arquitetura Otimizada:** Introduce refinamentos significativos nos módulos de extração de características, especialmente o bloco C3k2 (Cross Stage Partial), que melhora a capacidade de detectar objetos pequenos e densos, crítico para aplicações agrícolas ([KHANAM; HUSSAIN, 2024](#)).
- **Eficiência Superior:** Apesar de possuir menos parâmetros (56,9 milhões) que o YOLOv12x (59,1 milhões), o YOLOv11 alcança tempo de inferência de apenas 14,6ms com 58,14 FPS, superando o YOLOv12x que requer 18,9ms por imagem e atinge apenas 47,39 FPS. Isso representa uma melhoria de aproximadamente 23% em velocidade de processamento ([KHANAM; HUSSAIN, 2024](#)).
- **Estabilidade de Treinamento:** O YOLOv12, embora mais recente, apresenta desafios de estabilidade durante o treinamento e maior consumo de memória, sendo ainda considerado experimental. O YOLOv11 continua sendo a versão recomendada pela Ultralytics para a maioria das cargas de trabalho de pesquisa e produção ([ULTRALYTICS, 2024c](#)).
- **Relevância Comprovada para Tempo Real:** Testes em câmera em tempo real demonstram que o YOLOv11 mantém desempenho consistente, enquanto o YOLOv12 apresenta queda visível de FPS em aplicações de transmissão contínua ([KHANAM; HUSSAIN, 2024](#)).
- **Integração no Ecossistema Unificado:** Ambos YOLOv5 e YOLOv11 são mantidos no mesmo repositório Ultralytics, permitindo comparação direta e migração suave entre versões ([ULTRALYTICS, 2024b](#)).

2.7.3 Características Técnicas Principais

- **Regressão Unificada:** Ao contrário de métodos baseados em classificação, YOLO prevê diretamente caixas delimitadoras e probabilidades em uma única passada pela rede ([REDMON et al., 2016](#)).
- **Processamento em Tempo Real:** Ambas as versões mantêm velocidades compatíveis com aplicações em tempo real (YOLOv5: 140 FPS; YOLOv11: velocidades comparáveis com melhor precisão).
- **Versatilidade:** Suportam múltiplas tarefas de visão computacional além de detecção: segmentação, classificação e estimativa de pose ([ULTRALYTICS, 2024b](#)).

2.8 Trabalhos Relacionados

A aplicação de visão computacional para a detecção e análise de grãos tem sido explorada em diversos estudos recentes, demonstrando a evolução das técnicas clássicas para o aprendizado profundo.

([MOMIN et al., 2017](#)) utilizaram técnicas clássicas de visão computacional e processamento de imagem para avaliar a qualidade de grãos de soja. O trabalho baseou-se em segmentação por cor e morfologia para identificar impurezas e grãos danificados. Embora eficaz em ambiente controlado (laboratório com iluminação constante), a abordagem apresentou limitações de robustez quando aplicada a imagens de campo com fundos complexos, validando a necessidade de abordagens mais adaptativas como as redes neurais.

Em uma revisão abrangente, ([ZHANG et al., 2020](#)) demonstraram que o uso de CNNs supera consistentemente métodos tradicionais em tarefas de análise de cenas densas na agricultura, como a contagem de frutas em pomares e identificação de ervas daninhas. Os autores destacam que arquiteturas de estágio único, como o YOLO, são preferíveis para aplicações em tempo real devido ao seu processamento rápido em uma única passada pela rede neural.

Uma revisão abrangente da evolução das versões YOLO foi apresentada por ([ALIF; HUSSAIN, 2024](#)), que compilou aplicações de YOLOv1 até YOLOv10 no domínio agrícola. O trabalho demonstra que as versões posteriores do YOLO (v4, v5, v8, v9 e v10) foram progressivamente otimizadas para enfrentar desafios crescentes, particularmente a detecção de objetos de menor tamanho e oclusões. Em aplicações agrícolas específicas, estudos citados nessa revisão alcançaram resultados expressivos, como 93.81% de mAP em detecção de ervas daninhas e 98.92% de acurácia em diagnóstico de doenças em arroz. O aperfeiçoamento iterativo das arquiteturas YOLO valida sua aptidão para cenários agrônômicos complexos, onde objetos pequenos e densos, como grãos dispersos no solo pós-colheita, representam um desafio crítico de visão computacional.

Este trabalho se insere neste contexto, comparando gerações dessa arquitetura (especificamente YOLOv5 como baseline e YOLOv11 como solução proposta) para determinar a configuração ótima para o cenário específico da soja pós-colheita, validando empiricamente o potencial das arquiteturas modernas em condições agrícolas desafiadoras.

3 SOLUÇÃO PROPOSTA

Este capítulo detalha a metodologia experimental adotada, descrevendo a aquisição dos dados, as estratégias de pré-processamento e as configurações específicas de *hardware* e *software* utilizadas para o treinamento e validação dos modelos.

3.1 Materiais e Ferramentas

O projeto foi desenvolvido utilizando um conjunto de ferramentas e tecnologias que proporcionam um ambiente robusto para o desenvolvimento de soluções de visão computacional e aprendizado profundo:

- **Linguagem e Bibliotecas:** Todo o código foi desenvolvido em linguagem Python (v3.12.6), utilizando a biblioteca *PyTorch* como backend de processamento tensorial e a biblioteca *Ultralytics* para a implementação das arquiteturas YOLO. Para manipulação de imagens e matrizes, utilizaram-se *OpenCV* e *NumPy*.
- **Ambiente de Desenvolvimento (IDE):** Utilizou-se o *PyCharm* para a estruturação do código local e o Google Colab (*Jupyter Notebooks*) para a execução dos *scripts* de treinamento em nuvem.
- **Gestão de *Dataset*:** A plataforma Roboflow foi utilizada para a anotação, versionamento e aplicação do *pipeline* de pré-processamento e aumento de dados.
- **Hardware:** Os treinamentos foram acelerados por *hardware* gráfico (GPU) em dois cenários: uma NVIDIA RTX 3060 Ti (8GB) para testes locais e uma NVIDIA Tesla T4 (16GB) via Google Colab para os experimentos finais com o *dataset* expandido.

3.2 Aquisição e Preparação dos *Datasets*

A etapa de coleta de dados foi realizada em ambiente real de campo, visando capturar a variabilidade natural do solo e da iluminação. As imagens foram adquiridas utilizando dois dispositivos móveis com câmeras de alta resolução (48 MP e 50 MP), posicionados de forma a simular a perspectiva de monitoramento agrícola.

Para a organização e anotação das imagens, utilizou-se a plataforma Roboflow. O estudo baseou-se na construção e comparação de dois conjuntos de dados distintos, denominados *roboflow_soja* (*Dataset V1*) e *roboflow_soja_v2* (*Dataset V2*).

3.2.1 Dataset V1: roboflow_soja

O primeiro *dataset* foi constituído por 107 imagens brutas. Após a anotação manual das caixas delimitadoras (*bounding boxes*), aplicou-se a técnica de aumento de dados (*data augmentation*) para triplicar o volume de exemplos de treinamento, totalizando 321 imagens. A divisão do conjunto seguiu a proporção de 70% para treinamento, 20% para validação e 10% para testes.

3.2.2 Dataset V2: roboflow_soja_v2

O segundo *dataset* teve como objetivo ampliar a representatividade do modelo, introduzindo uma maior variedade de cenários, texturas de solo e condições de luminosidade, sem alteração no critério de anotação dos grãos. Este conjunto partiu de 230 imagens brutas, que após o processo de *augmentation*, resultaram em 552 imagens totais. A divisão realizada resultou na proporção de 70% para treinamento, 23% para validação e 7% para testes.

3.2.3 Pipeline de Pré-processamento e Aumento de Dados

Para garantir a padronização das entradas e a robustez do modelo frente à variabilidade de campo, foi estabelecido um *pipeline* rigoroso de processamento das imagens. Este fluxo foi dividido em duas etapas sequenciais configuradas na plataforma Roboflow:

3.2.3.1 Etapa 1: Pré-processamento (Estático)

Aplicado a todas as imagens (treino, validação e teste) para adequação ao formato de entrada da rede neural:

- **Auto-Orient:** Correção automática da orientação da imagem baseada nos metadados EXIF, garantindo que todas as fotos estejam alinhadas corretamente.
- **Resize:** Redimensionamento para a resolução de 1280×1280 pixels utilizando a técnica *Fit (black edges)*. Esta abordagem preserva a proporção original dos grãos e preenche as áreas remanescentes com bordas pretas, evitando distorções geométricas que poderiam prejudicar a detecção.
- **Auto-Adjust Contrast:** Aplicação de Equalização de Histograma para normalizar o contraste, mitigando os efeitos de sombras ou superexposição solar.

3.2.3.2 Etapa 2: Aumento de Dados (Dinâmico)

Aplicado exclusivamente ao conjunto de treinamento para multiplicar a diversidade de exemplos. O *pipeline* de *augmentation* incluiu as seguintes transformações sequenciais:

- **Geométrico:**

- *Flip*: Espelhamento horizontal e vertical (essencial, pois os grãos não possuem orientação fixa no solo).
- *90° Rotate*: Rotações de 90° (horário, anti-horário e invertido).
- *Crop*: Recorte aleatório com zoom entre 0% e 20%, simulando variações na distância da câmera.
- *Rotation*: Rotação fina aleatória entre -15° e +15°.
- *Shear*: Cisalhamento de $\pm 10^\circ$ nos eixos horizontal e vertical, introduzindo leves distorções de perspectiva.

- **Colorimétrico e Ruído:**

- *Hue*: Variação de matiz entre -15° e +15°.
- *Saturation*: Ajuste de saturação entre -25% e +25%.
- *Brightness*: Variação de brilho entre -20% e +20%.
- *Exposure*: Ajuste de exposição entre -10% e +10%.
- *Blur*: Borrão de até 2 pixels.
- *Noise*: Injeção de ruído em até 0.1% dos pixels (simulando granulação do sensor da câmera).

Este conjunto abrangente de operações resultou na geração de 3 variações para cada imagem original do conjunto de treinamento, aumentando significativamente a capacidade de generalização do modelo.

3.3 Arquiteturas e Configurações de Treinamento

Foram avaliadas duas gerações da arquitetura YOLO. A versão **YOLOv5** foi selecionada como *baseline* comparativo, devido à sua consolidação na literatura. A versão **YOLOv11** foi adotada como a proposta principal, visando avaliar os ganhos de desempenho das inovações arquiteturais mais recentes.

Para determinar a configuração ótima, foram realizados múltiplos experimentos variando a resolução de entrada, o tamanho do modelo (Small e Medium) e os hiperparâmetros de treinamento. A Tabela 1 detalha as configurações específicas de cada experimento realizado.

A variação no tamanho do lote (*batch*) deve-se às limitações de memória VRAM em cada cenário (8GB local vs 16GB nuvem), ajustada para evitar erros de alocação (*Out of Memory*).

Tabela 1 – Configurações dos Experimentos Realizados

ID do Experimento	<i>Dataset</i>	Modelo	Resolução	Épocas	<i>Batch</i>	<i>Hardware</i>
Exp 1 (Baseline)	V1	YOLOv5s	1280	100	16	Local
Exp 2	V1	YOLOv11s	1024	100	6	Local
Exp 3	V1	YOLOv11s	1280	120	6	Local
Exp 4	V1	YOLOv11s	1536	120	3	Local
Exp 5	V1	YOLOv11s	1920	120	2	Local
Exp 6	V1	YOLOv11m	1280	150	3	Local
Exp 7	V1	YOLOv11m	960	450	6	Local
Exp 8	V2	YOLOv11s	1024	200	4	Local
Exp 9	V2	YOLOv11s	1280	150	8	Colab (T4)

3.4 Métricas de Avaliação

Para mensurar o desempenho dos modelos, foram utilizadas métricas estatísticas padrão na literatura de detecção de objetos. A matriz de confusão serviu como base para o cálculo dos seguintes indicadores:

- **Precisão (*Precision*):** Indica a porcentagem de detecções corretas em relação ao total de detecções realizadas pelo modelo. Responde à pergunta: "De todos os grãos que o modelo marcou, quantos eram realmente grãos?".
- **Revocação (*Recall*):** Mede a capacidade do modelo de encontrar todos os objetos presentes na imagem. Responde à pergunta: "De todos os grãos que existiam no chão, quantos o modelo conseguiu encontrar?". Para a quantificação de perdas, esta métrica é considerada crítica.
- **F1-Score:** É a média harmônica entre Precisão e Revocação, fornecendo uma visão única do equilíbrio do modelo. É útil para comparar modelos quando um tem alta precisão e baixo recall (ou vice-versa).
- **mAP (*Mean Average Precision*):** É a métrica global mais robusta. O **mAP@0.5** considera um acerto se a caixa predita tiver pelo menos 50% de sobreposição (*IoU* - *Intersection over Union*) com a marcação real. Já o **mAP@0.5-0.95** faz uma média de várias exigências de sobreposição (de 50% a 95%), premiando modelos que desenham caixas geometricamente perfeitas.

4 AVALIAÇÃO E DISCUSSÃO DOS RESULTADOS

Para validar a hipótese do trabalho, foram comparadas as métricas de desempenho obtidas em todos os experimentos. A Tabela 2 apresenta os dados consolidados.

Tabela 2 – Resultados Comparativos das Métricas de Desempenho

ID	<i>Dataset</i>	Modelo	Resolução	Precisão	<i>Recall</i>	<i>F1-Score</i>	mAP@0.5	mAP@0.5:0.95
Exp 1	V1	YOLOv5s	1280	95.3%	83.9%	89.3%	91.2%	62.1%
Exp 2	V1	YOLOv11s	1024	88.5%	67.3%	76.5%	78.9%	52.5%
Exp 3	V1	YOLOv11s	1280	87.6%	77.8%	82.4%	86.6%	60.7%
Exp 4	V1	YOLOv11s	1536	90.0%	76.5%	82.7%	85.9%	61.0%
Exp 5	V1	YOLOv11s	1920	90.8%	72.5%	80.6%	84.3%	59.4%
Exp 6	V1	YOLOv11m	1280	91.5%	84.2%	87.7%	90.3%	68.0%
Exp 7	V1	YOLOv11m	960	94.6%	80.8%	87.2%	89.7%	66.5%
Exp 8	V2	YOLOv11s	1024	83.4%	65.5%	73.4%	74.8%	42.1%
Exp 9	V2	YOLOv11s	1280	93.4%	86.2%	89.7%	92.7%	63.7%

A métrica *F1-Score*, calculada como a média harmônica entre Precisão e Revocação, foi utilizada para desempatar o desempenho global dos modelos. O Exp 9 atingiu o maior *F1-Score* (89,7%), confirmando estatisticamente que ele oferece o melhor compromisso entre não gerar alarmes falsos (alta precisão) e não perder grãos reais (alta revocação), sendo, portanto, a solução mais robusta para a aplicação em campo.

4.1 Análise Gráfica e Visual

Para visualizar a estabilidade e a velocidade de convergência dos modelos, a Figura 1 apresenta a evolução do mAP ao longo das épocas de treinamento para os cinco melhores experimentos.

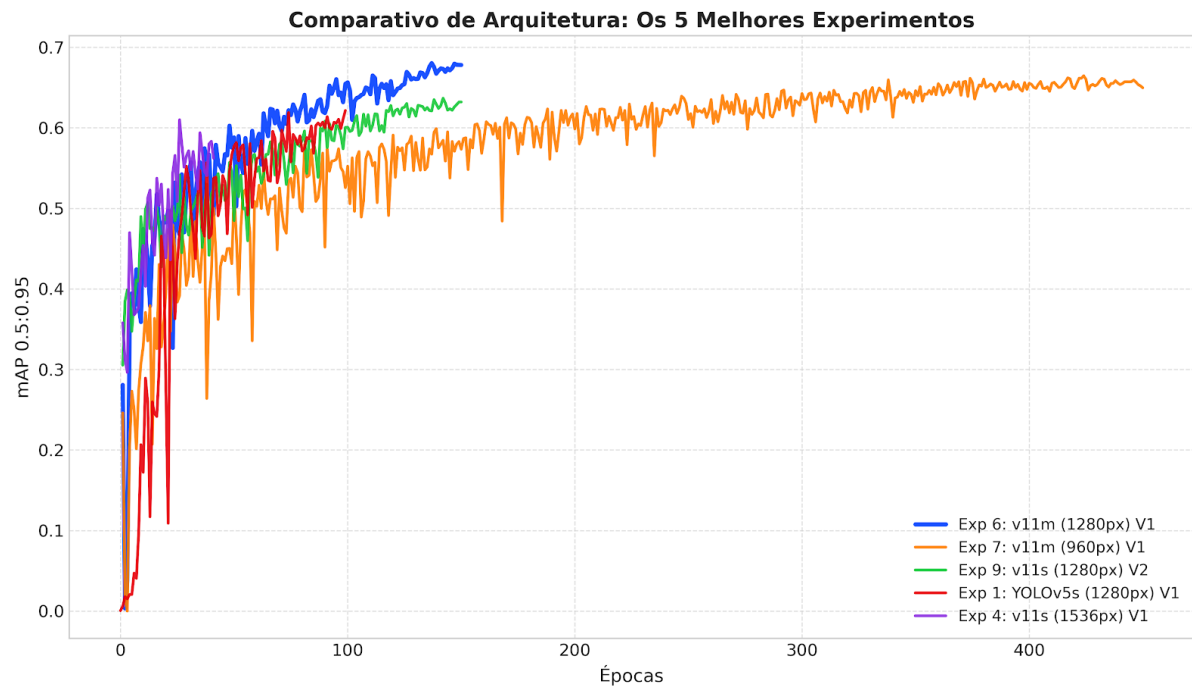


Figura 1 – Curvas de aprendizado (mAP@0.5:0.95) dos 5 melhores experimentos. Nota-se a superioridade e estabilidade do modelo YOLOv11s com *Dataset* V2 (Exp 9 - Linha Verde) em comparação ao baseline YOLOv5 (Exp 1 - Linha Vermelha).

Adicionalmente, a Figura 2 compara as matrizes de confusão do modelo inicial e final, evidenciando a redução de erros de classificação.

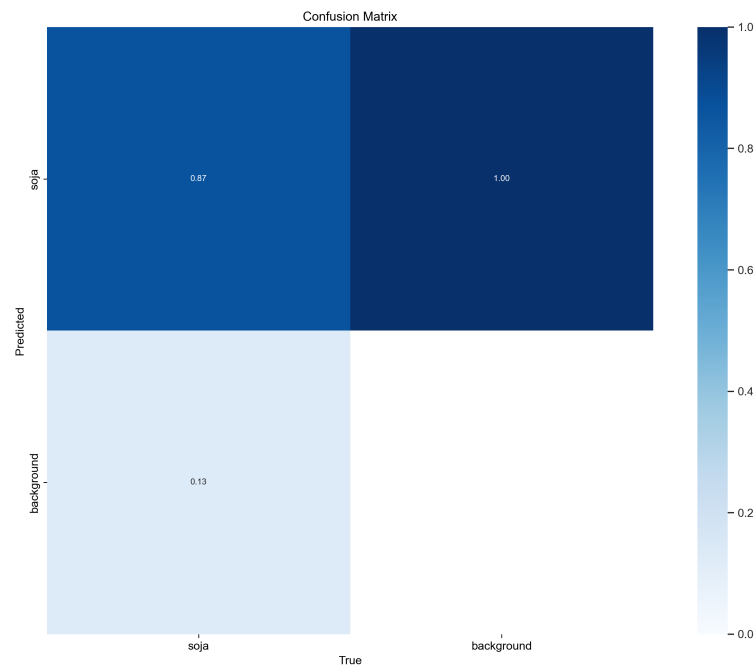
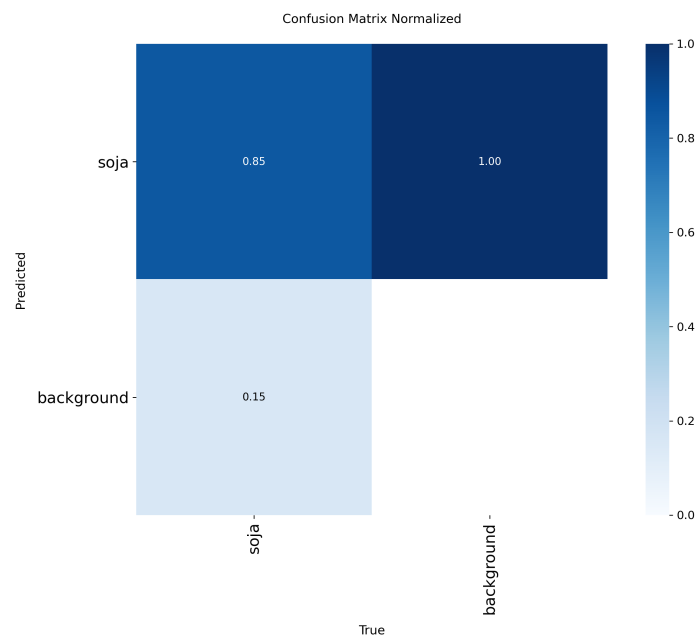
(a) YOLOv5s (Baseline - *Dataset V1*)(b) YOLOv11s (Proposta - *Dataset V2*)

Figura 2 – Comparativo de Matrizes de Confusão **Normalizadas**. A visualização em percentual permite a comparação direta entre os modelos independentemente do tamanho dos conjuntos de validação.

4.2 Discussão dos Resultados

Com base nos dados apresentados, destacam-se quatro conclusões principais:

4.2.1 Impacto da Arquitetura (V5 vs V11)

Ao comparar o *baseline* (Exp 1) com seu equivalente moderno (Exp 3), observa-se que o YOLOv5 apresentou maior Precisão pura, porém menor capacidade de generalização nas caixas delimitadoras. O YOLOv11 demonstrou ser mais robusto para localizar objetos difíceis.

4.2.2 O Limite da Resolução

Os experimentos 2 a 5 provaram que aumentar a resolução indefinidamente não garante melhores resultados. A resolução de 1920px (Exp 5) degradou o desempenho devido à redução forçada do *batch size*, o que desestabilizou o treinamento. A resolução de 1280px provou ser o ponto ótimo.

4.2.3 Superioridade do *Dataset V2*

O Experimento 9 (*Dataset V2*) apresentou o resultado mais equilibrado. Ele obteve a maior taxa de **Revocação (86.2%)** e o maior **mAP@0.5 (92.7%)**. Isso confirma que a estratégia de expandir a variabilidade do *dataset* foi o fator mais determinante para o sucesso do projeto, superando o ganho obtido apenas pela troca de arquitetura.

4.2.4 Critérios de Seleção do Melhor Modelo

A Tabela 2 apresenta 9 experimentos com métricas variadas. Embora alguns modelos apresentem valores numéricos superiores em métricas isoladas (e.g., Exp 1 com Precisão de 95,3% ou Exp 6 com mAP@0.5-0.95 de 68,0%), a seleção do Exp 9 como melhor modelo foi fundamentada em múltiplos critérios técnicos e contextuais:

1. **F1-Score Máximo (89,7%)**: A métrica F1 representa a média harmônica entre Precisão e Recall, fornecendo uma visão equilibrada do desempenho global. O Exp 9 obtém o maior F1-Score entre todos os experimentos.
2. **Recall Crítico (86,2%)**: Em aplicações de monitoramento visual de perdas agrícolas, a métrica Recall é mais importante que Precisão. Um falso negativo (grão não detectado) resulta em subestimação de perdas, prejudicando a tomada de decisão do agricultor. O Exp 9 maximiza essa métrica.
3. **Contexto Agrícola**: Deixar de detectar 100 grãos reais é mais grave que registrar 5 falsos positivos. O Exp 9 minimiza omissões em 13,8%, enquanto Exp 1 deixa 16% e Exp 6 deixa 15,8% dos grãos sem detecção.
4. ***Dataset V2* - Maior Diversidade**: Enquanto experimentos com V1 utilizaram apenas 107 imagens expandidas para 321, o Exp 9 foi treinado com 230 imagens

expandidas para 552. Essa variabilidade resulta em maior capacidade de generalização para cenários de campo real.

5. **Viabilidade Computacional:** O modelo Small possui 51% menos parâmetros que o Medium (25M vs 51M). Isso viabiliza futuras implementações em borda (edge computing) em drones ou máquinas agrícolas, onde *Small* é exequível mas Medium é inviável.
6. **Arquitetura Mais Moderna:** YOLOv11 supera YOLOv5 em robustez em cenários desafiadores, conforme validado na Análise Qualitativa (Seção 3.5.3), onde obtém 100% de acerto em Condições Ideais vs desempenho inferior do baseline em Baixo Contraste e Fundo Complexo.
7. **mAP@0.5 Máximo (92,7%):** Apesar de não ser o mais alto em mAP@0.5-0.95, o Exp 9 alcança a maior pontuação em mAP@0.5, indicando excelente qualidade geométrica das bounding boxes mesmo em IoU moderado.

Por essas razões, conclui-se que o Exp 9 representa o melhor compromisso entre acurácia, robustez, viabilidade computacional e aplicabilidade agrícola, apesar de sacrificar a Precisão máxima (93,4% vs 95,3% do Exp 1) em favor de Recall e generalização superiores.

4.3 Análise Qualitativa (Inferência Visual)

Complementando a avaliação quantitativa, foi realizada uma análise qualitativa baseada na inspeção visual das inferências. Esta etapa é fundamental para validar a robustez do modelo diante da variabilidade não estruturada encontrada em ambiente de campo, que muitas vezes não é totalmente capturada apenas por métricas globais como o mAP. As figuras a seguir apresentam comparativos diretos entre o *baseline* (YOLOv5) e o modelo proposto (YOLOv11) em quatro cenários distintos de desafio visual.

4.3.1 Interpretação das Caixas Delimitadoras e Confidence Scores

Nas imagens a seguir, os grãos de soja detectados pelos modelos são indicados por **caixas delimitadoras** (bounding boxes) que envolvem cada objeto. Os números visíveis acima de cada caixa representam o **score de confiança** da detecção, expresso em formato decimal (por exemplo, 0.85 indica 85% de confiança na detecção). Estes valores fornecem uma indicação da certeza do modelo para cada grão detectado:

- **Scores altos** (próximos a 1.0): indicam que o modelo está muito seguro de que o objeto é realmente um grão de soja. São detecções robustas, tipicamente quando o grão apresenta boa visibilidade e contraste com o fundo.

- **Scores médios** (0.6 a 0.8): indicam detecções válidas, mas com alguma incerteza. Frequentemente ocorrem quando o grão está parcialmente obscurecido, em região de baixo contraste ou nas bordas da imagem.
- **Scores baixos** (< 0.5): podem indicar detecções incorretas (falsos positivos), ou seja, o modelo confundiu um elemento do fundo (como fragmento de terra ou palhada) com um grão de soja.

A análise conjunta da visualização das caixas delimitadoras e seus confidence scores permite uma avaliação mais precisa da qualidade das detecções e da capacidade de discriminação do modelo em cenários desafiadores. Além disso, detectar muitos objetos com scores altos é mais importante que detectar muitos objetos com scores heterogêneos, pois a confiança nas predições reflete a generalização do modelo.

4.3.2 Análise por Cenários

Inicialmente, avalia-se o comportamento dos modelos em condições favoráveis. Conforme observado na Figura 3, sob boa luminosidade e proximidade do solo, a diferença de desempenho é sutil.



(a) YOLOv5



(b) YOLOv11

Figura 3 – **Cenário 1 (Condições Ideais)**: Com boa iluminação e baixa altura, ambos os modelos detectam a maioria dos grãos. Nota-se, contudo, que o YOLOv11 apresenta uma taxa de acerto de 100% para grãos e apenas 1 falso positivo.

No entanto, um dos maiores desafios na detecção de soja é a **similaridade cromática** entre o objeto e o fundo. A Figura 4 ilustra o cenário crítico de baixo contraste tonal, onde a coloração do solo se assemelha muito à tonalidade dos grãos. Nessa situação de “camuflagem” natural, a capacidade do modelo de distinguir bordas e texturas sutis é levada ao limite.

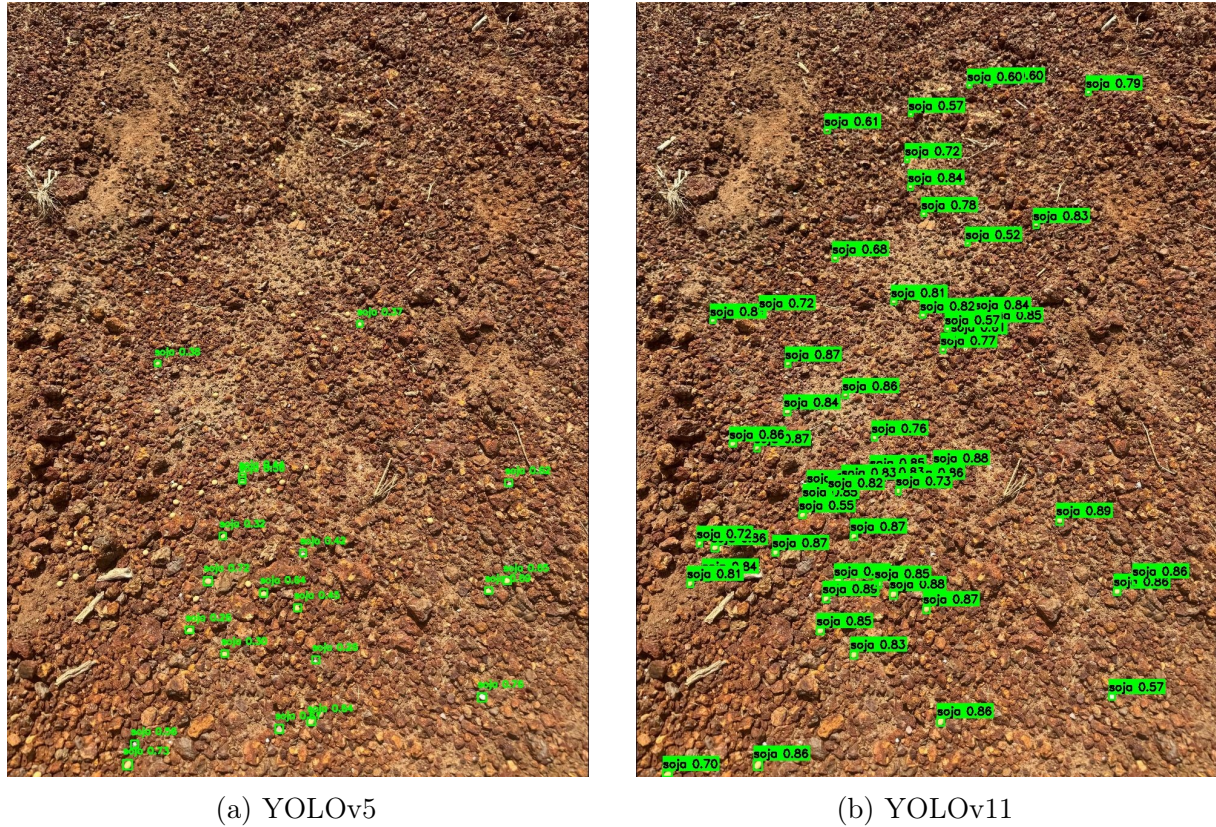


Figura 4 – **Cenário 2 (Baixo Contraste Tonal)**: Situação onde a cor do solo é muito próxima à cor do grão. O modelo baseline (a) falha em separar o objeto do fundo devido à semelhança de cores. O modelo proposto (b) demonstra maior sensibilidade, identificando os grãos mesmo sem um contraste de cor evidente.

Além das condições de luz, a heterogeneidade do substrato representa um desafio significativo para a extração de características. A presença de restos culturais (palhada) e detritos minerais cria um “ruído” visual que pode induzir o modelo ao erro. A Figura 5 evidencia como o modelo proposto supera essa dificuldade, discriminando corretamente os grãos em meio a um fundo visualmente complexo e evitando falsos positivos.

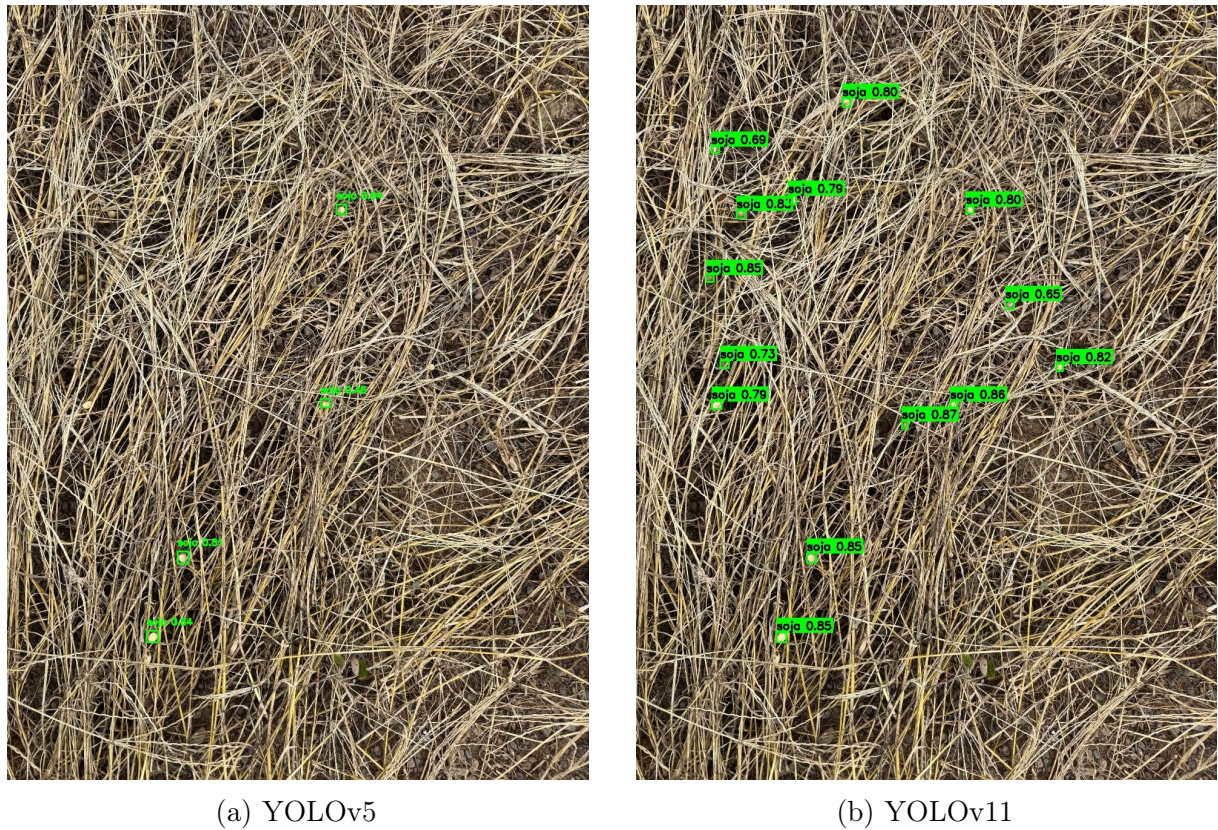


Figura 5 – **Cenário 3 (Fundo Complexo)**: Em meio a restos culturais (palhada), o YOLOv5 gera falsos positivos (confunde terra com soja). O YOLOv11 discrimina corretamente as texturas, mantendo a identificação fidedigna.

Por fim, avaliou-se o limite de resolução espacial dos modelos. Em aplicações reais, a altura da câmera pode variar devido à irregularidade do terreno ou à montagem no maquinário. A Figura 6 demonstra o comportamento dos detectores quando a representação do objeto é reduzida a poucos pixels, cenário onde a arquitetura mais moderna do YOLOv11 se mostrou decisiva para a manutenção da taxa de detecção.

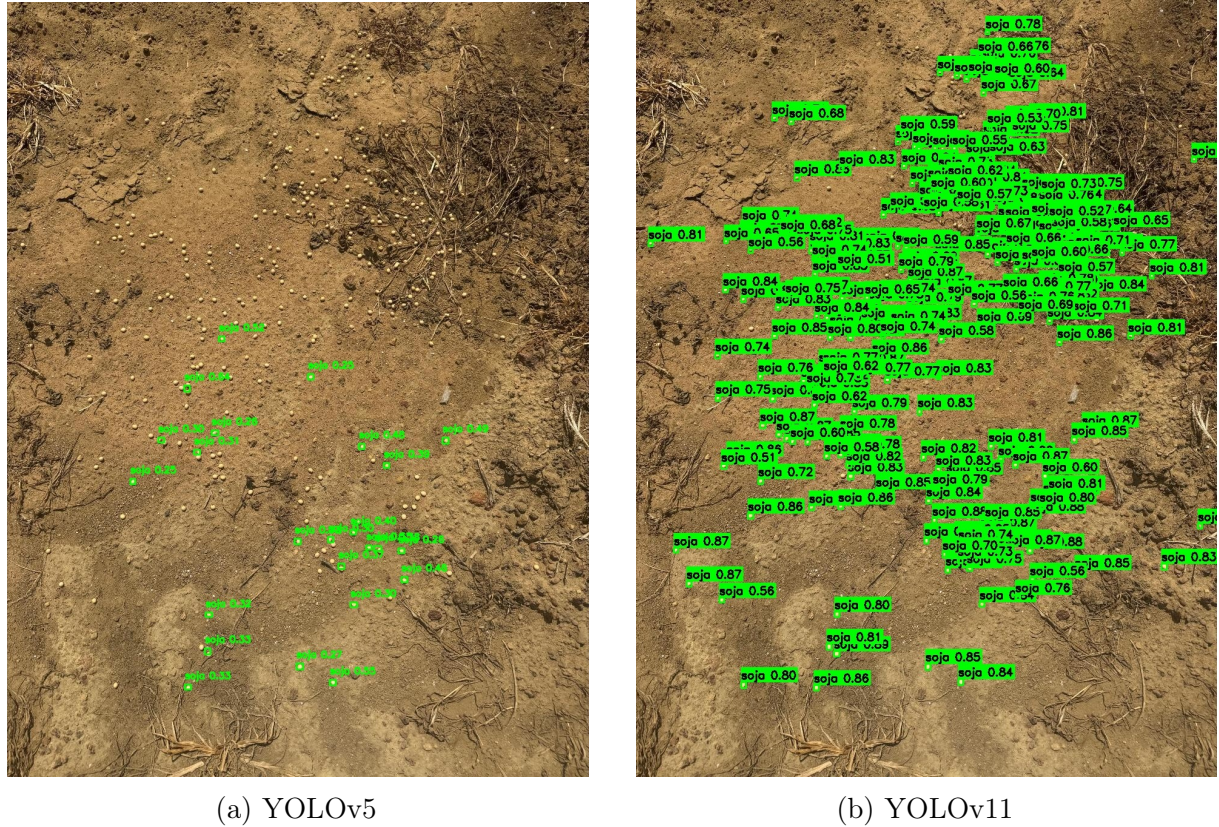


Figura 6 – **Cenário 4 (Variação de Escala)**: Imagem capturada a maior altura. A redução na resolução dos grãos impede a detecção pelo modelo antigo, enquanto a nova arquitetura preserva as características espaciais necessárias para a identificação.

A análise qualitativa consolida os resultados numéricos: o modelo proposto (YOLOv11) não apenas detecta mais objetos, mas o faz em condições onde a visão computacional tradicional costuma falhar. A evolução da arquitetura provou-se decisiva para superar a **fusão visual** entre grão e solo (Cenário 2) e a interferência de ruídos como a palhada (Cenário 3). Além disso, a manutenção da performance em imagens de alta altitude (Cenário 4) valida a robustez do modelo para aplicações reais, onde nem sempre é possível garantir o enquadramento ou a iluminação ideais. Os confidence scores consistentemente elevados do YOLOv11 indicam não apenas melhor detecção, mas maior confiabilidade nas predições, aspecto crítico para aplicações de precisão agrícola onde decisões operacionais dependem da qualidade das estimativas.

5 CONCLUSÃO

Este trabalho dedicou-se ao desenvolvimento e validação de um sistema de visão computacional para a detecção automática de perdas de grãos de soja, abordando um dos desafios críticos da agricultura de precisão. A hipótese central de que técnicas de Aprendizado Profundo seriam capazes de identificar grãos em cenários complexos de pós-colheita foi confirmada através dos experimentos realizados.

O estudo comparativo entre as arquiteturas YOLOv5 e YOLOv11, submetidas a diferentes resoluções e conjuntos de dados, permitiu identificar os fatores determinantes para o sucesso da detecção. Conclui-se que a simples aplicação de modelos mais complexos ou o aumento indiscriminado da resolução não garantem melhores resultados. A estratégia mais eficaz consistiu na expansão da variabilidade do *dataset* (V2), que proporcionou ao modelo a capacidade de generalização necessária para distinguir os grãos em meio a palhada e solo heterogêneo.

O modelo final (YOLOv11s treinado com *Dataset V2*) atingiu um mAP@0.5 de 92,7% e uma Revocação de 86,2%. A escolha deste modelo em detrimento de versões mais robustas (como o YOLOv11m) baseou-se na priorização da sensibilidade do sistema: para o monitoramento visual de perdas, a minimização de falsos negativos (grãos não detectados) é mais crítica do que a precisão milimétrica da caixa delimitadora. Além disso, a arquitetura *Small* oferece a viabilidade computacional necessária para futuras implementações em borda.

5.1 Contribuições

As principais contribuições desta pesquisa consolidam-se na estruturação de um *pipeline* de pré-processamento e aumento de dados eficaz para imagens de soja em solo, o que validou a importância da variabilidade do *dataset* como fator crítico de desempenho. Aliado a isso, destaca-se a validação empírica da superioridade da arquitetura YOLOv11 sobre a YOLOv5 para a detecção de objetos pequenos em cenários agrícolas complexos. Essa superioridade foi evidenciada por um ganho de 1,5 pontos percentuais em mAP@0.5 (de 91,2% para 92,7%) e de 2,3 pontos em Revocação (de 83,9% para 86,2%). O estudo também demonstrou que a resolução de 1280×1280 pixels representa um excelente ponto de equilíbrio entre o detalhamento visual e o custo computacional para o *hardware* disponível, evitando o desperdício de recursos sem comprometer o desempenho. Adicionalmente, a análise qualitativa realizada em quatro cenários desafiadores (condições ideais, baixo contraste tonal, fundo complexo e variação de escala) comprovou que o

modelo proposto é robusto o suficiente para aplicações reais de campo, onde as condições adversas são frequentes. Por fim, o trabalho contribui com a proposição de métricas de priorização específicas para o contexto agrícola, defendendo que a revocação (o esforço de não deixar grãos sem detecção) mostra-se mais importante do que a precisão pura (evitar falsos positivos), invertendo a lógica de avaliação típica de outros domínios da visão computacional.

5.2 Trabalhos Futuros

Considerando os resultados obtidos e as limitações identificadas ao longo do desenvolvimento, diversas direções são propostas para a continuidade desta pesquisa. Do ponto de vista algorítmico e de visão computacional, sugere-se investigar o uso da técnica *Slicing Aided Hyper Inference* (SAHI) durante a etapa de detecção. Essa abordagem permitiria o uso de imagens de altíssima resolução sem a necessidade de redimensionamento drástico, ampliando a cobertura de detecção para cenários com maior altura de câmera e capturando grãos ainda mais distantes. Adicionalmente, a exploração de técnicas de *ensemble learning* poderia unificar as forças do modelo YOLOv11s (que se destaca pela velocidade) com versões mais densas (como YOLOv11m ou YOLOv11l), possibilitando uma predição mais robusta e mantendo um custo computacional moderado através de estratégias inteligentes de votação ou probabilidade.

No que tange à aplicação e distribuição do sistema, propõe-se a portabilidade do modelo treinado para dispositivos embarcados de baixo consumo (*Edge AI*). Isso viabilizaria o processamento em tempo real sem a dependência de conectividade à nuvem, reduzindo drasticamente a latência e os custos operacionais no campo. Em paralelo a essa embarcação, o desenvolvimento de uma aplicação *desktop* ou *web* com interface amigável aos usuários facilitaria o trabalho dos operadores, permitindo a captura de imagens, o processamento em tempo real e a visualização intuitiva dos resultados. A incorporação de ferramentas de geolocalização (GPS) a esse *software* possibilitaria, inclusive, a geração de mapas de variabilidade de perdas entre diferentes talhões da fazenda.

Por fim, visando a validação em larga escala, recomenda-se a condução de testes em condições operacionais reais de colheita. A comparação direta entre as estimativas do sistema automatizado e o método tradicional (copo medidor) em diferentes propriedades, tipos de solos e variedades de soja validaria definitivamente a aplicabilidade prática da solução, além de identificar ajustes necessários para cenários geográficos específicos. Em conjunto com essa validação prática, a adaptação do *dataset* e da metodologia para a detecção de perdas em culturas com características visuais distintas, como milho e feijão, viabilizaria uma ferramenta unificada e indispensável para o monitoramento de perdas pós-colheita no agronegócio moderno.

5.3 Conclusão Final

O presente trabalho demonstrou que sistemas de visão computacional baseados em Aprendizado Profundo são viáveis e eficazes para o monitoramento automático de perdas pós-colheita de soja. A arquitetura YOLOv11, quando adequadamente treinada com *datasets* representativos e diversificados, superou o desempenho dos modelos convencionais, consolidando a aplicabilidade da solução para a agricultura de precisão.

A minimização de perdas agrícolas não é apenas uma questão de eficiência econômica, mas também de sustentabilidade ambiental. Cada grão desperdiçado representa insumos produtivos (sementes, água, fertilizantes) investidos sem retorno. A adoção de sistemas automatizados como o desenvolvido neste trabalho contribui para uma cadeia produtiva mais eficiente, sustentável e resiliente frente aos desafios da mudança climática e crescimento populacional.

REFERÊNCIAS

- ALIF, M. A. R.; HUSSAIN, M. Yolov1 to yolov10: A comprehensive review of yolo variants and their application in the agricultural domain. *arXiv preprint arXiv:2406.10139*, 2024. Citado na página 28.
- BOCK, R.; ALONÇO, A. dos S.; DIAS, V. de O.; POSSEBOM, G.; KNIERIM, L. F.; CRUZ, W. A. S. da; MACHADO, A. P. Á. Perdas na colheita mecanizada da soja em função da velocidade de deslocamento e índice de molinete. *Brazilian Journal of Development*, v. 6, n. 6, p. 34707–34724, 2020. Citado 3 vezes nas páginas 15, 24 e 25.
- COSTA, C. C. d.; GUILHOTO, J. J. M.; BURNQUIST, H. L. Impactos socioeconômicos de reduções nas perdas pós-colheita de produtos agrícolas no brasil. *Revista de Economia e Sociologia Rural*, SciELO Brasil, v. 53, n. 3, p. 395–408, 2015. Citado 2 vezes nas páginas 14 e 25.
- Food and Agriculture Organization of the United Nations (FAO). *The State of Food and Agriculture 2019. Moving forward on food loss and waste reduction*. Rome, 2019. Citado 2 vezes nas páginas 16 e 25.
- GIRSHICK, R.; DONAHUE, J.; DARRELL, T.; MALIK, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2014. p. 580–587. Citado na página 21.
- GODFRAY, H. C. J.; BEDDINGTON, J. R.; CRUTE, I. R.; HADDAD, L.; LAWRENCE, D.; MUIR, J. F.; PRETTY, J.; ROBINSON, S.; THOMAS, S. M.; TOULMIN, C. Food security: the challenge of feeding 9 billion people. *science*, American Association for the Advancement of Science, v. 327, n. 5967, p. 812–818, 2010. Citado na página 16.
- GONZALEZ, R. C.; WOODS, R. E. *Digital Image Processing*. 4th. ed. [S.l.]: Pearson, 2018. Citado na página 21.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: MIT Press, 2016. Citado 3 vezes nas páginas 18, 19 e 20.
- JOCHER, G. et al. *YOLOv5*. [S.l.]: GitHub, 2020. <<https://github.com/ultralytics/yolov5>>. Citado na página 16.
- KAMILARIS, A.; PRENAFETA-BOLDÚ, F. X. Deep learning in agriculture: A survey. *Computers and electronics in agriculture*, Elsevier, v. 147, p. 70–90, 2018. Citado 4 vezes nas páginas 14, 18, 23 e 24.
- KHANAM, R.; HUSSAIN, M. Yolov11: An overview of the key architectural enhancements. *arXiv preprint arXiv:2410.17725*, 2024. Citado 2 vezes nas páginas 23 e 27.
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, v. 25, 2012. Citado 2 vezes nas páginas 20 e 21.

- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *nature*, Nature Publishing Group UK London, v. 521, n. 7553, p. 436–444, 2015. Citado 4 vezes nas páginas 14, 19, 21 e 22.
- LIAKOS, K. G.; BUSATO, P.; MOSHOU, D.; PEARSON, S.; BOCHTIS, D. Machine learning in agriculture: A review. *Sensors*, Mdpi, v. 18, n. 8, p. 2674, 2018. Citado 2 vezes nas páginas 18 e 23.
- MAVRIDOU, E.; VROCHIDOU, E.; PAPAKOSTAS, G. A.; PACHIDIS, T.; KABURLASOS, V. G. Machine vision systems in precision agriculture for crop farming. *Journal of Imaging*, MDPI, v. 5, n. 12, p. 89, 2019. Citado na página 24.
- MOMIN, M. A.; YAMAMOTO, K.; MIYAMOTO, M.; KONDO, N.; GRIFT, T. Machine vision based soybean quality evaluation. *Computers and Electronics in Agriculture*, Elsevier, v. 140, p. 452–460, 2017. Citado 2 vezes nas páginas 24 e 28.
- PAIXÃO, C. S. S.; VOLTARELLI, M. A.; SOUZA, J. B. C.; FILHO, A. L. D. B.; SILVA, R. P. D. Loss sampling methods for soybean mechanical harvest. *Bioscience Journal*, v. 38, n. e38050, p. 1981–3163, 2022. Citado 4 vezes nas páginas 14, 15, 24 e 25.
- PATRÍCIO, D. I.; RIEDER, R. Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review. *Computers and electronics in agriculture*, Elsevier, v. 153, p. 69–81, 2018. Citado na página 16.
- REDMON, J.; DIVVALA, S.; GIRSHICK, R.; FARHADI, A. You only look once: Unified, real-time object detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 779–788. Citado 5 vezes nas páginas 14, 21, 22, 26 e 28.
- RUSSELL, S. J.; NORVIG, P. *Artificial Intelligence: A Modern Approach*. 4th. ed. [S.l.]: Pearson, 2020. Citado 2 vezes nas páginas 15 e 18.
- ULTRALYTICS. *Ultralytics YOLOv5 Documentation*. 2023. Disponível em: <<https://docs.ultralytics.com/pt/yolov5/>>. Citado 2 vezes nas páginas 26 e 27.
- _____. *Ultralytics YOLOv8 Documentation*. 2023. Disponível em: <<https://docs.ultralytics.com/pt/models/yolov8/>>. Citado na página 26.
- _____. *Histórico e Evolução de Versões YOLO*. 2024. Disponível em: <<https://docs.ultralytics.com/pt/>>. Citado na página 26.
- _____. *Ultralytics YOLO11 Documentation*. 2024. Disponível em: <<https://docs.ultralytics.com/pt/models/yolo11/>>. Citado 4 vezes nas páginas 22, 26, 27 e 28.
- _____. *Ultralytics YOLO12 Documentation*. 2024. Disponível em: <<https://docs.ultralytics.com/pt/models/yolo12/>>. Citado 2 vezes nas páginas 26 e 27.
- _____. *YOLOv8 vs YOLOv5: Evolução da Detecção de Objetos*. 2024. Disponível em: <<https://docs.ultralytics.com/pt/compare/yolov8-vs-yolov5/>>. Citado na página 27.
- ZHANG, Q.; LIU, Y.; GONG, C.; CHEN, Y.; YU, H. Applications of deep learning for dense scenes analysis in agriculture: A review. *Sensors*, MDPI, v. 20, n. 5, p. 1520, 2020. Citado 2 vezes nas páginas 24 e 28.