Combinando métodos de detecção de objetos com sistema de localização e mapeamento simultâneos

Rodrigo de Almeida Silva

SERVIÇO DE PÓS-GRADUAÇÃO DA FAENG-UFMS

Data de Depósito:

Assinatura:_____

Rodrigo de Almeida Silva

Orientador: *Prof. Dr. Wesley Nunes Gonçalves*

Dissertação apresentada à Universidade Federal de Mato Grosso do Sul na Faculdade de Engenharias, Arquitetura e Urbanismo e Geografia, como requisito para a obtenção do grau de Mestre em Engenharia Elétrica.

UFMS - Campo Grande
Março/2024

Aos meus pais,
Paulo e Simone,

À minha amada,
Lúria,

E à minha querida sobrinha,
Ana Carolina,

Agradecimentos

Agradeço de coração ao meu orientador, Prof. Dr. Wesley Nunes Gonçalves, e ao meu coorientador, Prof. Dr. José Marcato Júnior, por todo o apoio, orientação e principalmente oportunidades que me proporcionaram ao longo desta trajetória. Sem a ajuda de vocês, certamente não teria chegado tão longe e não teria alcançado os resultados que alcancei. Vocês foram fundamentais tanto na minha jornada acadêmica quanto profissional, sou imensamente grato por tudo que fizeram por mim.

À minha família, que sempre esteve ao meu lado, agradeço por todo o incentivo, amor e compreensão. Vocês foram a minha base, o meu porto seguro em todos os momentos. Agradeço por acreditarem em mim e me apoiarem em todas as minhas decisões.

E à minha querida namorada, não tenho palavras para expressar o quanto sou grato por ter você ao meu lado. Sua compreensão e paciência foram verdadeiras inspirações, e cada momento compartilhado com você tornou esta jornada ainda mais especial. Agradeço por ser a minha maior motivação e por tornar cada passo dessa jornada mais significativo.

Não posso deixar de agradecer à CAPES por me conceder uma bolsa de mestrado, o que possibilitou que eu me dedicasse integralmente aos estudos e à pesquisa, sem preocupações financeiras. Sua contribuição foi essencial para o sucesso deste trabalho.

E por fim, agradeço à UFMS por proporcionar a estrutura e os recursos necessários para que eu pudesse realizar o meu mestrado. O ambiente acadêmico e os recursos disponibilizados pela universidade foram de grande importância para o desenvolvimento deste trabalho.

Enfim, agradeço a todos que fizeram parte desta caminhada, direta ou indiretamente. Sou grato por cada aprendizado, por cada desafio superado e por cada conquista alcançada. Estou animado com o que o futuro reserva, e tudo isso só foi possível graças ao apoio e amor de vocês. Muito obrigado!

Abstract

This work explored the combination of Deep Learning (DL) and Simultaneous Localization and Mapping (SLAM) to enhance precision agriculture, with a focus on detecting and estimating the distance to apples in orchards. A thorough literature review was conducted, analyzing approaches that integrate deep neural networks with traditional SLAM methods, identifying promising applications in various fields, including agriculture.

Data collection was performed using a stereoscopic camera, capturing images with depth information. Bounding boxes were manually annotated around visible apples, and the MinneApple dataset was added to enhance model generalization.

We trained 12 variations of the YOLO architecture, with YOLOv5x achieving the best performance, reaching 0.861 in F1-Score on the validation set. An algorithm was developed to estimate the distance to each detected apple, integrating it into the YOLO detection pipeline.

The results demonstrated the accuracy and real-time viability of the system, allowing efficient detection and distance estimation of apples. This work contributes to the evolution of the combination of DL and SLAM, opening new research prospects for automation and robotics, particularly in fruit orchard monitoring and harvesting applications.

Resumo

Este trabalho explorou a combinação entre Deep Learning (DL) e Simultaneous Localization and Mapping (SLAM) para melhorar a agricultura de precisão, com ênfase na detecção e estimativa de distância de maçãs em pomares. Realizamos uma revisão detalhada da literatura, analisando abordagens que unem redes neurais profundas com métodos tradicionais de SLAM, identificando aplicações promissoras em várias áreas, incluindo a agricultura.

A coleta de dados foi feita usando a câmera estereoscópica, que captura imagens com informações de profundidade. Anotamos manualmente bounding boxes nas maçãs visíveis e adicionamos o conjunto MinneApple ao nosso dataset para aprimorar a generalização dos modelos.

Treinamos 12 variações da arquitetura YOLO e a YOLOv5x alcançou o melhor desempenho, atingindo 0.861 em F1-Score no conjunto de validação. Desenvolvemos um algoritmo para estimar a distância até cada maçã detectada, integrando-o ao fluxo de detecção da YOLO.

Os resultados demonstraram a precisão e viabilidade do sistema em tempo real, permitindo a detecção e estimativa de distância das maçãs de maneira eficiente. O trabalho contribui para a evolução da combinação entre DL e SLAM, abrindo novas perspectivas de pesquisa para a automação e robótica, especialmente em aplicações de monitoramento e colheita de frutas em pomares.

Sumário

Lista de Figuras

Lista de Tabelas

Lista de Abreviaturas

DL Deep Learning

SLAM Simultaneous Localization and Mapping

V-SLAM Visual Simultaneous Localization and Mapping

F1-Score F1 Score

RMSE Root Mean Square Error

LIDAR Light Detection and Ranging

RGB-D Red, Green, Blue, Depth

NIR Near-Infrared

YOLO You Only Look Once

CAPÍTULO

1

Introdução

A detecção e contagem de frutas em pomares são tarefas cruciais para a automação agrícola. Elas podem ser usadas para reduzir atividades rotineiras de cultivo e criação, além de fornecer estimativas importantes para a colheita e as próximas estações de crescimento. Além disso, a detecção precisa de frutas possibilita a oportunidade de colheita robótica, que tem o potencial de eliminar um dos processos mais intensos em mão de obra para os agricultores. James et al. (2023)

Além disso, a detecção precisa de objetos em imagens e vídeos tem desempenhado um papel fundamental em diversas aplicações, desde a segurança até a automação industrial e a condução autônoma. Entre os objetos de interesse, as frutas desempenham um papel crucial na indústria agrícola e no setor de alimentos. Segundo Bargoti and Underwood (2017), a detecção e rastreamento preciso de frutas, como maçãs, podem auxiliar no controle de qualidade, na estimativa de colheita e no monitoramento do crescimento das culturas.

No contexto atual da agricultura de precisão, a adoção de tecnologias avançadas tem se mostrado cada vez mais relevante. O avanço da visão computacional e do aprendizado profundo tem proporcionado métodos eficazes para a detecção de objetos em imagens e vídeos, permitindo que sistemas autônomos realizem tarefas complexas de forma eficiente e precisa. Nesse cenário, este trabalho de mestrado se concentra na detecção de maçãs em vídeos e na estimativa de distância de cada maçã até a câmera, com potenciais aplicações na agricultura de precisão.

De acordo com Dandekar et al. (2021), a detecção de maçãs é um desafio devido à sua variedade de formas, tamanhos e cores, bem como às variações

de iluminação e oclusões presentes no ambiente agrícola. Além disso, a estimativa de distância é uma informação crítica para aplicações como sistemas de colheita automatizada e monitoramento de culturas. O conhecimento da distância das maçãs em relação à câmera pode fornecer insights importantes sobre o crescimento das frutas, permitindo tomadas de decisões mais precisas e otimizadas para os produtores.

Para abordar esse desafio, utilizamos um conjunto de dados coletado na região de Trentino-Alto Ádige, na Itália, com a colaboração valiosa de alunos da Università degli Studi di Trento e produtores locais. O conjunto de dados foi composto por vídeos de maçãs, capturados com uma câmera estereoscópica Stereolabs ZED 2, considerando variações de iluminação e distâncias estabelecidas de 20, 40, 60 e 80 cm. Essa abordagem permitiu a obtenção de dados representativos do ambiente agrícola e das condições reais de cultivo de maçãs.

Durante o processo experimental, realizamos a anotação das bounding boxes nas imagens usando a plataforma RoboFlow. Em seguida, exploramos diferentes versões da arquitetura YOLO (You Only Look Once), incluindo as versões YOLOv5, YOLOv7 e YOLOv8. O objetivo foi encontrar um método de detecção rápido e eficiente, capaz de ser implantado em dispositivos embarcados para aplicações em tempo real no campo.

Os resultados obtidos após a análise experimental foram fundamentais para a seleção do modelo mais adequado. Observamos que a versão YOLOv5x se destacou, alcançando o melhor desempenho em termos de F1-Score, com uma capacidade impressionante de detecção precisa das maçãs nas imagens, mesmo em condições desafiadoras de iluminação e oclusões.

Este trabalho representa uma contribuição para o avanço da detecção de maçãs em vídeos e da estimativa de distância, fornecendo informações valiosas para a indústria agrícola e permitindo a otimização de processos relacionados à produção e ao monitoramento de culturas. Além disso, as técnicas e metodologias exploradas neste estudo podem ser estendidas para outras aplicações na agricultura de precisão e em diversas áreas que requerem a detecção precisa e a estimativa de distância de objetos.

Ao unir o poder do aprendizado profundo e da visão computacional com a precisão das câmeras estereoscópicas, abre-se um horizonte de possibilidades para sistemas autônomos e robótica, possibilitando a automação de tarefas complexas e a geração de informações valiosas para tomadas de decisões inteligentes em diferentes setores. Com isso, esperamos que este trabalho inspire pesquisas futuras e estimule a adoção de soluções inovadoras para enfrentar os desafios da agricultura moderna e das tecnologias assistivas, proporcionando avanços significativos em direção a um futuro mais sustentável e

eficiente.

1.1 Motivação

A automação e otimização da colheita de maçãs são desafios enfrentados pela agricultura. A detecção de objetos em vídeos e a estimativa de distância podem fornecer soluções eficientes e econômicas. Neste trabalho, utilizamos a câmera Stereolabs ZED 2, com sua capacidade estereoscópica, para detectar maçãs em vídeos e estimar a distância até cada maçã. Essa abordagem visa melhorar a eficiência da colheita e permitir o monitoramento agrícola com maior precisão, contribuindo para a agricultura de precisão e práticas sustentáveis.

1.2 Objetivos

A detecção precisa de objetos em imagens e vídeos, combinada com a estimativa de distância, desempenha um papel fundamental em diversas aplicações, especialmente na agricultura de precisão e no setor de tecnologias agrícolas inovadoras. Nesse contexto, o objetivo principal deste trabalho é desenvolver um sistema avançado capaz de detectar maçãs em vídeos e estimar a distância de cada maçã até a câmera, utilizando a tecnologia da câmera Stereolabs ZED 2. Para alcançar esse objetivo, foram estabelecidos os seguintes objetivos específicos:

1. Coletar um conjunto diversificado de imagens contendo maçãs em diferentes cenários e condições de iluminação, na região de Trentino-Alto Ádige, Itália. Essa coleta de dados é fundamental para representar a variabilidade de situações encontradas em ambientes agrícolas reais, permitindo o treinamento e a validação eficaz dos modelos de detecção e estimativa de distância.

2. Realizar a anotação manual das bounding boxes em cada imagem, delimitando as regiões de interesse que contêm as maçãs. A anotação precisa e detalhada é essencial para o treinamento supervisionado dos algoritmos de detecção e para o desenvolvimento do algoritmo de estimativa de distância.

3. Desenvolver um algoritmo personalizado para estimar a distância até o centro de cada bounding box detectada, ou seja, até cada maçã presente nas imagens, utilizando as informações de profundidade fornecidas pela câmera Stereolabs ZED 2. A estimativa de distância precisa é um componente-chave para aplicações como sistemas de colheita automatizada e monitoramento de culturas, pois fornece informações essenciais sobre o crescimento das frutas e otimiza o processo de colheita.

4. Treinar diferentes modelos de detecção de objetos baseados na arquite-

tura YOLO. A seleção e treinamento adequados dos modelos de detecção são fundamentais para obter um sistema de detecção de maçãs eficaz e robusto, capaz de lidar com as variações de iluminação, formas, tamanhos e cores das maçãs encontradas nos ambientes agrícolas.

5. Comparar o desempenho dos modelos de detecção e avaliar sua capacidade de identificar e localizar maçãs em vídeos, utilizando métricas como o F1-Score. A análise comparativa permitirá selecionar o modelo mais adequado para a tarefa específica de detecção de maçãs em vídeos, garantindo a precisão e confiabilidade do sistema.

6. Aplicar a estimativa de distância desenvolvida em cenários práticos, explorando possíveis aplicações em sistemas de colheita automatizada e monitoramento de culturas. A aplicação da estimativa de distância em cenários reais demonstrará a relevância e o potencial do sistema para otimizar operações agrícolas e oferecer soluções tecnológicas inovadoras para o setor.

7. Fornecer conclusões relevantes com base nos resultados obtidos, contribuindo para o avanço da detecção de maçãs e estimativa de distância na agricultura de precisão e tecnologias agrícolas inovadoras.

CAPÍTULO

2

Revisão de Literatura

Neste capítulo, apresentamos uma revisão da interseção entre o Simultaneous Localization and Mapping (SLAM) e o Deep Learning, destacando as principais abordagens que têm sido propostas na área da robótica e da visão computacional. Nós exploramos as recentes tendências de pesquisa que integram redes neurais profundas com métodos tradicionais de SLAM, destacando as aplicações na agricultura de precisão.

O SLAM consistem em mapear um ambiente desconhecido e ao mesmo tempo se localizar nesse ambiente usando dados de sensores, Qian et al. (2023). Embora diferentes sensores possam contribuir para a formação de mapas, o SLAM visual está se tornando cada vez mais popular por ser capaz de produzir informações de mapeamento detalhadas que são úteis para muitas aplicações, como robótica, transporte, busca e resgate, construções e muitas outras. O V-SLAM (Visual SLAM) depende principalmente de câmeras de vários tipos, incluindo monoculares, estéreo e RGBD, devido à sua capacidade de compreender a cena em comparação com outros sensores, como lasers. Krishna et al. (2023)

Os algoritmos de V-SLAM sofreram avanço significativo e encontraram uma ampla gama de aplicações em diversos cenários, incluindo robôs de serviço indoor, veículos autônomos urbanos e dispositivos de realidade aumentada, como descrito por Xu et al. (2023). Na agricultura de precisão não é diferente, o V-SLAM tem se mostrado cada vez mais presente, principalmente quando combinado com técnicas de Deep Learning.

A combinação entre DL e informações espaciais tem sido aplicada em diversas áreas e contextos. Desde aplicações no campo como fora dele, essas técnicas têm se mostrado promissoras. Como apresentado em Qureshi et al.

(2023), onde é realizado um trabalho com maçãs na Nova Zelândia, um mercado em franca expansão com perspectivas de exportação atingindo a marca de US$ 2 bilhões até 2030. Os autores abordam a crescente necessidade de aumentar a força de trabalho sazonal, bem como capacitá-la para tarefas especializadas, como a seleção de frutas. Nesse contexto, os pesquisadores desenvolveram uma plataforma robótica inovadora, utilizando um braço robótico UR5 equipado com câmeras estéreo e empregando técnicas de aprendizado profundo para detecção das maçãs. Os resultados obtidos demonstram que a seleção de maçãs pode ser realizada com precisão por meio dessa abordagem.

Além da seleção de frutas, Freeman et al. (2022) apresenta uma abordagem alternativa para medir o tamanho e acompanhar o crescimento de maçãs. A proposta utiliza visão computacional com estereoscopia para obter resultados comparáveis aos métodos atuais baseados em calibradores. Os resultados mostram que o sistema computacional é capaz de prever taxas de crescimento, sem a necessidade de esforço humano para medir ou rotular as frutas. Embora os resultados sejam promissores, ainda há desafios técnicos a serem superados para tornar o sistema totalmente autônomo, incluindo o uso de câmeras estéreo que possam operar em ambientes com variação de luz e a investigação de algoritmos de correspondência estéreo mais rápidos e leves.

O mapeamento tem ganhando bastante foco, Kang et al. (2022) e Kang and Wang (2023) abordam o uso de um LIDAR juntamente com uma câmera para conseguirem um sistema de percepção robótica precisa em pomares de maçãs. Eles aplicaram segmentação semântica na nuvem de pontos e com isso conseguem a localização precisa das frutas. Por conta do uso de um LIDAR, temos um sistema de alto custo, que poderia se tornar mais barato utilizando uma camera stereoscópica.

Além disso, a etapa de colheira tem ganhado destaque. Costanzo et al. (2023) propõem uma pipeline de percepção e controle para a colheita robótica de frutas, onde é utilizado um sistema de câmera RGB-D (assim como a Zed). O pipeline começa com uma etapa de estimativa de pose 6D usando a rede neural DOPE, que é treinada com dados sintéticos de frutas. Em seguida, é aplicado um algoritmo de otimização baseado na informação de profundidade da câmera RGB-D para refinar a estimativa de pose e dimensionar as frutas em tempo real. A informação de profundidade é fundamental para o ajuste preciso da pose estimada para diferentes tamanhos de frutas. O sistema é capaz de realizar colheitas bem-sucedidas mesmo em frutas com dimensões que diferem significativamente do modelo de treinamento da rede neural, graças ao uso da câmera RGB-D.

Do mesmo modo, temos ainda operações com frutas mais sensíveis. Qiu et al. (2023) apresentam uma garra robótica, capaz de detectar o amadure-

cimento das amoras e colher frutas delicadas com visão computacional. Os testes mostraram sucesso na colheita de amoras maduras sem danos significativos, utilizando uma câmera NIR. O trabalho exige alto grau de precisão na garra e poderia ser melhorado utilizando uma câmera estereoscopica, fornecendo informações espaciais precisas à garra.

As aplicações vão para além do campo, como mostrado por Dikshit et al. (2023), os autores propõem um sistema robótico autônomo para cortar diversos tipos de frutas e legumes em uma tábua de corte. O sistema utiliza abordagem baseada em visão, usando os modelos YOLO e SAM para detecção e segmentação dos objetos. O sistema foi avaliado em experimentos, alcançando uma taxa geral de sucesso em torno de 77-80%. No entanto, há algumas limitações, como a suposição de que as ações de corte sempre são bem-sucedidas. Os autores sugerem melhorias futuras, mas o sistema mostra resultados promissores no desenvolvimento de robôs autônomos para tarefas de corte complexas. Com o uso de uma estereocamera, o robô poderia ter noções espaciais que o ajudariam na tarefa.

Além do contexto de frutas, aplicações em outras áreas tem se tornado cada vez mais comuns. Adhikari and Bhandari (2023) abordam o uso de câmeras estereoscópicas para obter informações físicas precisas de objetos do mundo real, mas com foco na estimativa de velocidade de veículos em sistemas de transporte inteligente. Eles utilizam o modelo SiamMask para realizar o rastreamento progressivo dos veículos e estimam a distância entre eles a partir do mapa de disparidade. Em seguida, empregam modelos de regressão, como LightGBM, para estimar a velocidade dos veículos. Os resultados mostraram que essa abordagem supera o os resultados anteriores, alcançando uma RMSE de 0.416.

Por outro lado, Ekanayake et al. (2023) abordam a aplicação de mapeamento do ambiente utilizando visão computacional estereoscópica e um sensor de laser (LRF - Laser Range Finder) de baixo custo. O mapeamento é útil para extrair informações sobre objetos, como comprimento e altura, tipo de material, entre outros. A proposta consiste em utilizar duas câmeras web de baixo custo para a visão estereoscópica e um sensor LRF para obter informações em 2D do ambiente. Combinando esses dados, foi possível construir um mapa 3D do ambiente. A técnica proposta é muito econômica, em termos de hardware. O resultado é um mapeamento do ambiente que pode ser usado para identificar objetos, calcular distâncias e até mesmo determinar o material dos objetos presentes no ambiente.

Outra área muito relevante e que tem ganhado destaque, é a área da saúde e tecnologias assistivas, como mostrado por Thakurdesai et al. (2019). Neste trabalho os autores propõem um sistema de assistência à locomoção para pes-

soas com deficiência visual, combinando a detecção de objetos usando YOLO com a estimativa de profundidade através de visão monocular. O sistema utiliza uma rede neural convolucional para detectar objetos em imagens em tempo real, fornecendo informações auditivas sobre a distância e a localização desses objetos para orientar o usuário. A estimativa de profundidade é realizada gerando uma segunda imagem sintética da cena e aplicando uma correspondência estéreo para criar um mapa de disparidade. O sistema alcança uma boa precisão na detecção de objetos, no entanto, a estimativa de distância poderia ser melhorada ao usar uma estereocamera.

Além dessas, um grande número de outras áreas tem ganhado atenção e tem se beneficiado das técnicas de DL combinados com SLAM, como pode ser visto em Fu and Kong (2023) e também em Chen et al. (2020), onde é abordado SLAM para ambientes dinâmicos.

Percebe-se que a utilização da combinação entre DL e SLAM tem impulsionado avanços significativos em diversas aplicações, tanto na agricultura de precisão quanto em outras áreas, permitindo a automatização de tarefas complexas e oferecendo soluções promissoras para o futuro da robótica e da visão computacional.

CAPÍTULO

3

Materiais e Métodos

Neste capítulo, descrevemos os materiais utilizados e os métodos empregados ao longo do desenvolvimento deste trabalho. A Seção 3.1 aborda os detalhes dos materiais utilizados. Em seguida, na Seção 3.2.1, apresentamos os métodos empregados para a coleta de imagens. Na Seção 3.2.2 damos detalhes da anotação das áreas de interesse. Na sequência, na Seção 3.2.3, discorremos sobre o treinamento dos modelos de detecção. Logo após, na Seção 3.2.4, descrevemos o funcionamento da obtenção de distância até cada maçã detectada. Por fim, na Seção 3.2.5, esquematizamos o fluxo completo de operação.

3.1 Materiais

Utilizamos uma câmera estereoscópica Stereolabs ZED 2 para a coleta de imagens. Essa câmera é equipada com dois sensores de imagem, permitindo capturar informações de profundidade essenciais para a estimativa de distância até as maçãs nas imagens. As gravações foram feitas na resolução 1280x720 pixels, a 30 fps.

Figura 3.1: Câmera Stereolabs ZED 2 utilizada.

Juntamente, um laptop para poder armazenar as gravações. Para manter

a distância da linha de plantio fixa, utilizamos uma trena comum. Na etapa de treinamento, fizemos o uso de um computador equipado com duas placas de vídeo RTX 3090.

3.2 Métodos

3.2.1 Coleta de Imagens

Na fase de coleta de imagens, nos deslocamos até Mezzolombardo, Itália e procuramos caminhar paralelamente às linhas de plantio, com distâncias fixas de 20, 40, 60 e 80cm. Para cada distância, capturamos vídeos de 10 a 30 segundos de duração, variando posições a favor e contra o sol. Também realizamos gravações estáticas nas mesmas distâncias, a fim de avaliar a precisão da câmera ao fornecer dados de distância, totalizando 12 vídeos curtos. Nas figuras 3.2, 3.3, 3.4 e 3.5 temos exemplos de frames capturados pela zed a cada distância e com variações de iluminação.

Figura 3.2: Exemplo de imagem capturada pela ZED, a 20cm e a favor do sol.

3.2.2 Anotação de Bounding Boxes

Montamos um dataset separando aleatóriamente 20% dos frames capturados em campo. Não utilizamos os vídeos estáticos, apenas os vídeos contendo movimento. A anotação das bounding boxes, que representam as regiões de interesse contendo as maçãs nas imagens, foi realizada utilizando a plataforma RoboFlow. Essa ferramenta proporcionou uma anotação precisa e eficiente, simplificando o processo de marcação manual. Procuramos ajustar cada bounding box de forma precisa a todas as maçãs visíveis nos frames.

Para obter maior variabilidade e modelos mais genéricos, adicionamos aos nossos dados o dataset MinneApple, apresentado por Hani et al. (2020), com

Figura 3.3: Exemplo de imagem capturada pela ZED, a 40cm e contra o sol.

Figura 3.4: Exemplo de imagem capturada pela ZED, a 60cm e contra o sol.

isso conseguimos melhorar significativamente a generalização dos modelos.

No total, o dataset completo possui 3621 imagens, variando entre 2 e 171 maçãs por imagens, o que nos dá uma média de 36 maçãs por imagem, como pode ser visto pela distribuição apresentada na Figura 3.6. Separamos 95% para treino, 4% para validação e 1% para teste.

3.2.3 Treinamento dos Modelos

Com o conjunto de dados anotado, procedemos ao treinamento dos modelos de detecção. Todos os modelos foram treinados em um computador equipado com duas placas de vídeo NVIDIA RTX 3090. Utilizamos a arquitetura YOLO (You Only Look Once), onde exploramos diferentes versões, incluindo YOLOv5, YOLOv7 e YOLOv8. Todos os treinamentos foram feitos com os hyperparâmetros e configurações padrões de cada versão.

Figura 3.5: Exemplo de imagem capturada pela ZED, a 80cm e a favor do sol.

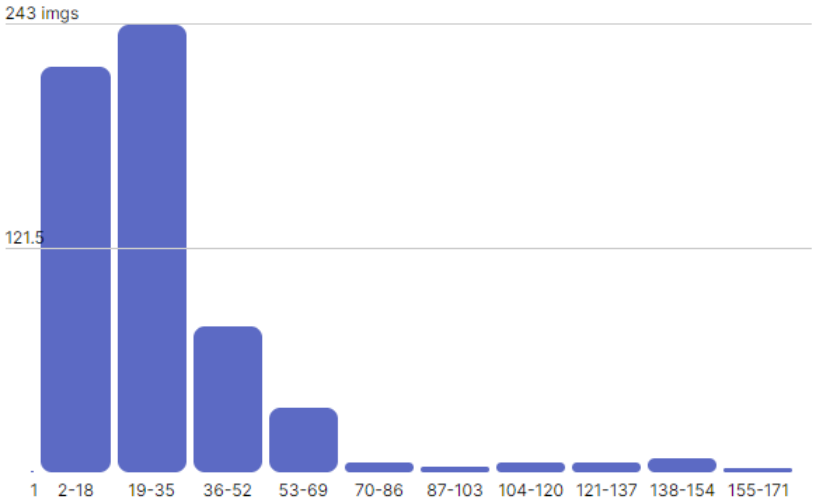

Figura 3.6: Distribuição de maçãs por imagem.

3.2.4 Estimativa de Distância

Considerando que a câmera ZED utilizada possui a capacidade de fornecer uma estimativa de distância em cada pixel da imagem, desenvolvemos um algoritmo personalizado para obter a distância até o centro de cada bounding box detectada, ou seja, até cada maçã. Esse processo permite que cada maçã tenha uma distância estimada correspondente, o que é fundamental para aplicações que requerem informações de distância, como sistemas de colheita automatizada e monitoramento de culturas. Para esse propósito, consideramos o pixel central de cada bounding box como uma representação da distância até cada maçã detectada, de modo que, dado uma bounding box B_1 com coordenadas (x_1, y_1) para o vértice superior esquerdo e (x_2, y_2) para o vértice inferior direito, temos que o centro de B_1 é (cx_1, cy_1), tal que $cx_1 = x_1 + \lfloor (x_2 - x_1)/2 \rceil$ e $cy_1 = y_1 + \lfloor (y_2 - y_1)/2 \rceil$. Então dado uma maçã M_1, detectada por B_1, a distância da câmera até M_1 será a distância da câmera até o centro de B_1, em (cx_1, cy_1).

3.2.5 Fluxo completo

Dessa forma, podemos resumir a operação como mostrado na Figura 3.7. Para cada instante de tempo, a ZED captura um frame (RGB) em cada lente e juntamente, adiciona dados para os demais sensores (acelerômetro, giroscópio, barômetro, magnetômetro e termômetro), adicionalmente ela calcula um mapa de profundidade (D), baseado nas imagens registradas pelas duas lentes, dessa forma temos para cada instante de tempo imagens RGB-D. Sendo assim, separamos o mapa de profundidade da imagem, ficando portanto com dois dados separados, o frame RGB e o mapa de profundidade. O frame é enviado ao modelo, para obtermos as predições que por sua vez são combinadas com o mapa de profundidade a fim de obtermos a distância até cada maçã, conforme descrito em 3.2.4.

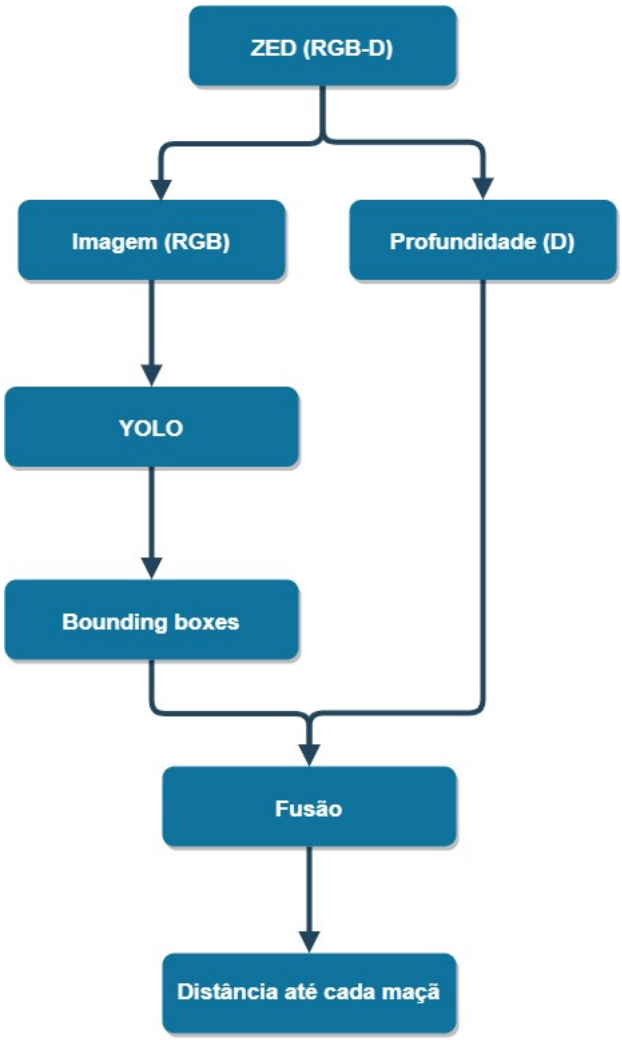

Figura 3.7: Fluxograma para obtenção das distâncias até cada maçã detectada.

CAPÍTULO

4

Resultados

4.0.1 Resultados Quantitativos

Após o treinamento, todos os modelos foram avaliados e comparados utilizando métricas de desempenho, como o F1-Score, que avalia a capacidade dos modelos de detecção em termos de precisão e revocação.

Os resultados foram analisados, e verificou-se que a versão YOLOv5x se destacou, alcançando o melhor desempenho em termos de F1-Score, como pode ser visto na Tabela 4.1.

Modelo	MAP 50	MAP 95	P	R	F1
Yolov5n	0.740	0.302	0.780	0.671	0.721
Yolov5s	0.774	0.334	0.808	0.702	0.751
Yolov5m	0.833	0.376	0.834	0.785	0.808
Yolov5l	0.866	0.413	0.859	0.820	0.839
Yolov5x	**0.886**	0.439	**0.876**	**0.847**	**0.861**
Yolov7	0.725	0.291	0.772	0.639	0.699
Yolov7x	0.621	0.245	0.648	0.575	0.609
Yolov8n	0.722	0.325	0.757	0.653	0.701
Yolov8s	0.800	0.374	0.809	0.725	0.764
Yolov8m	0.862	0.417	0.837	0.806	0.821
Yolov8l	0.863	0.428	0.832	0.819	0.825
Yolov8x	0.878	**0.447**	0.840	0.837	0.838

Tabela 4.1: Desempenho dos modelos de detecção de maçãs.

Para avaliar a capacidade da câmera em fornecer a distância das maçãs, montamos um experimento, onde posicionamos maçãs à distâncias fixas de

20, 40, 60 e 80cm, como pode ser visto nas Figuras 4.1, 4.2, 4.3 e 4.4, respectivamente. Neste experimento a ZED não conseguiu computar a distância para maçãs distantes até 40cm da câmera. Por outro lado, para as maçãs que estavam a 60cm e 80cm, a câmera obteve uma performance muito boa, com erro inferior a 2cm. A distância é informada na bounding box, juntamente com a predição "dist x.xm".

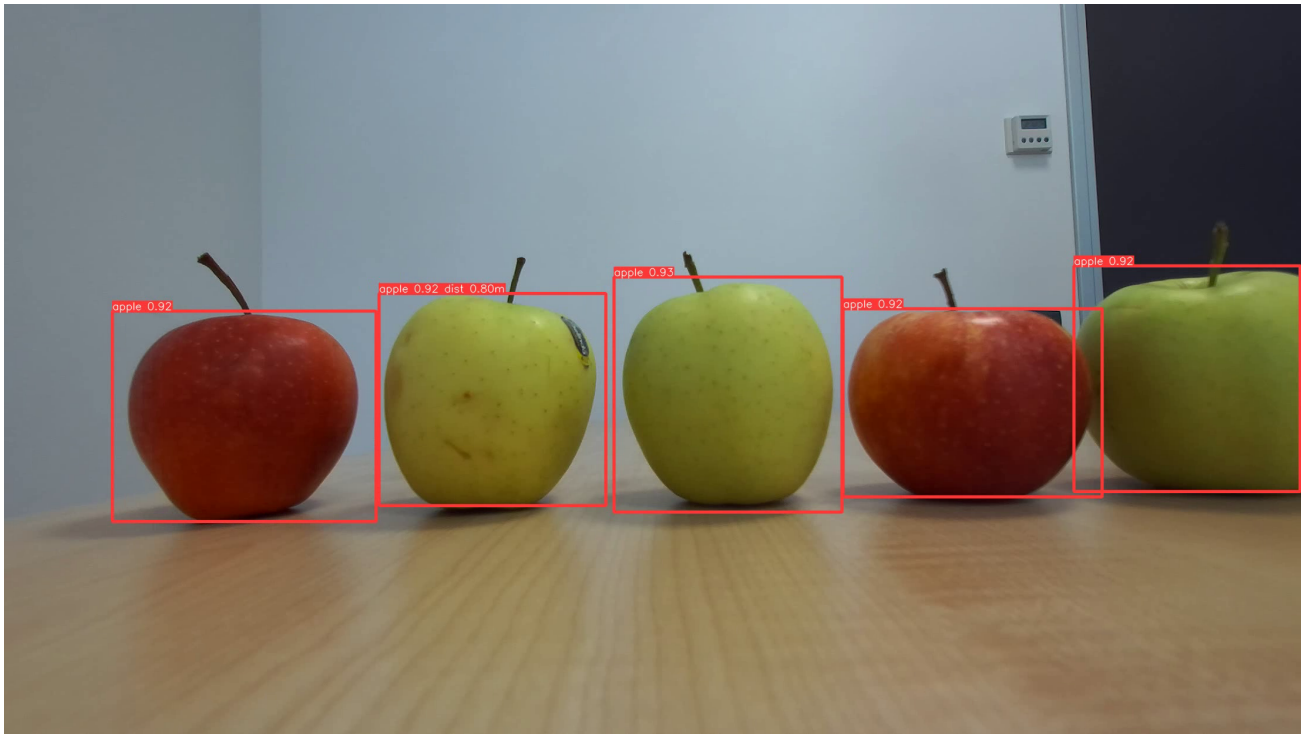

Figura 4.1: Experimento a 20cm.

4.0.2 Resultados Qualitativos

Analisando de forma qualitativa, o modelo demonstrou ter generalizado bem, pois realizamos testes em dados não utilizados no treinamento e inclusive gravados com outros tipos de câmeras e a detecção manteve-se consistente, como pode ser visto na Figura 4.5. Nas Figuras 4.6, e 4.7 podemos observar um frame, gravado a 60cm da linha de plantio, contra a luz do sol (cenário mais desafiador) que acaba provocando bastante sombras nas regiões de interesse e mesmo assim o modelo consegue performar bem, com uma taxa de acerto alta.

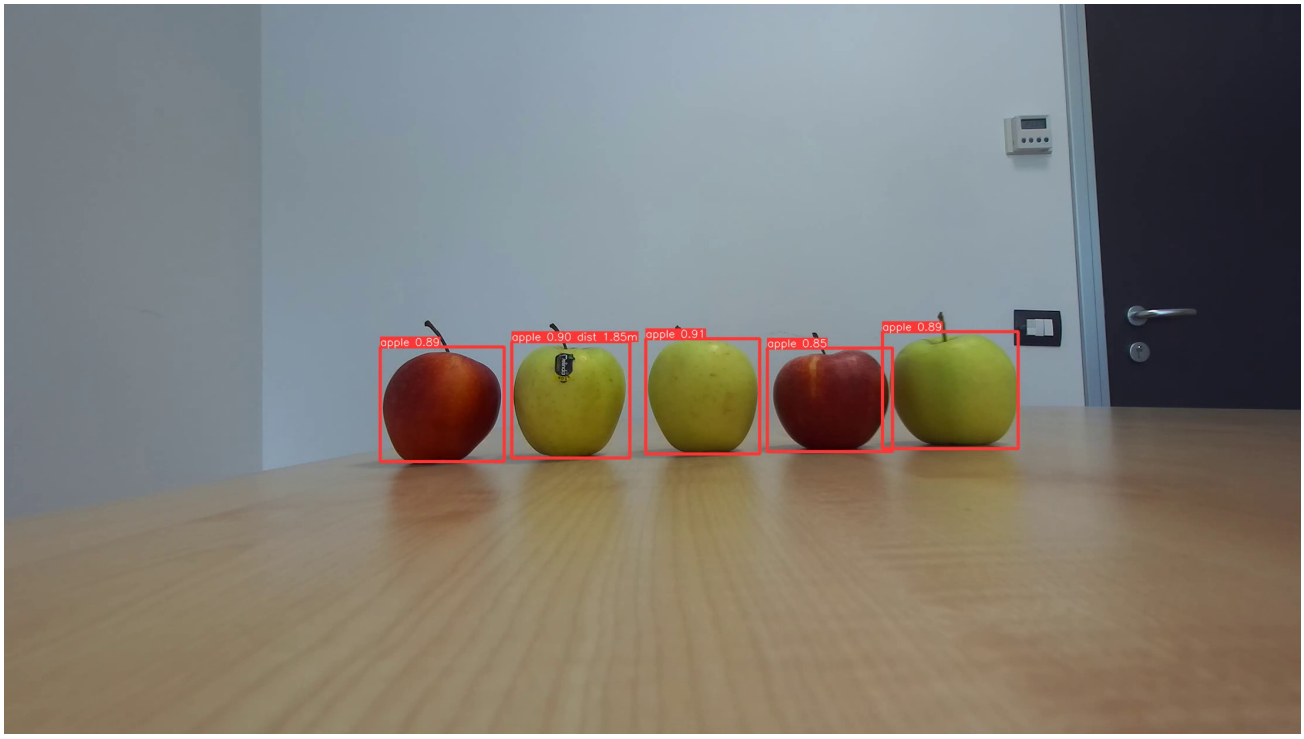

Figura 4.2: Experimento a 40cm.

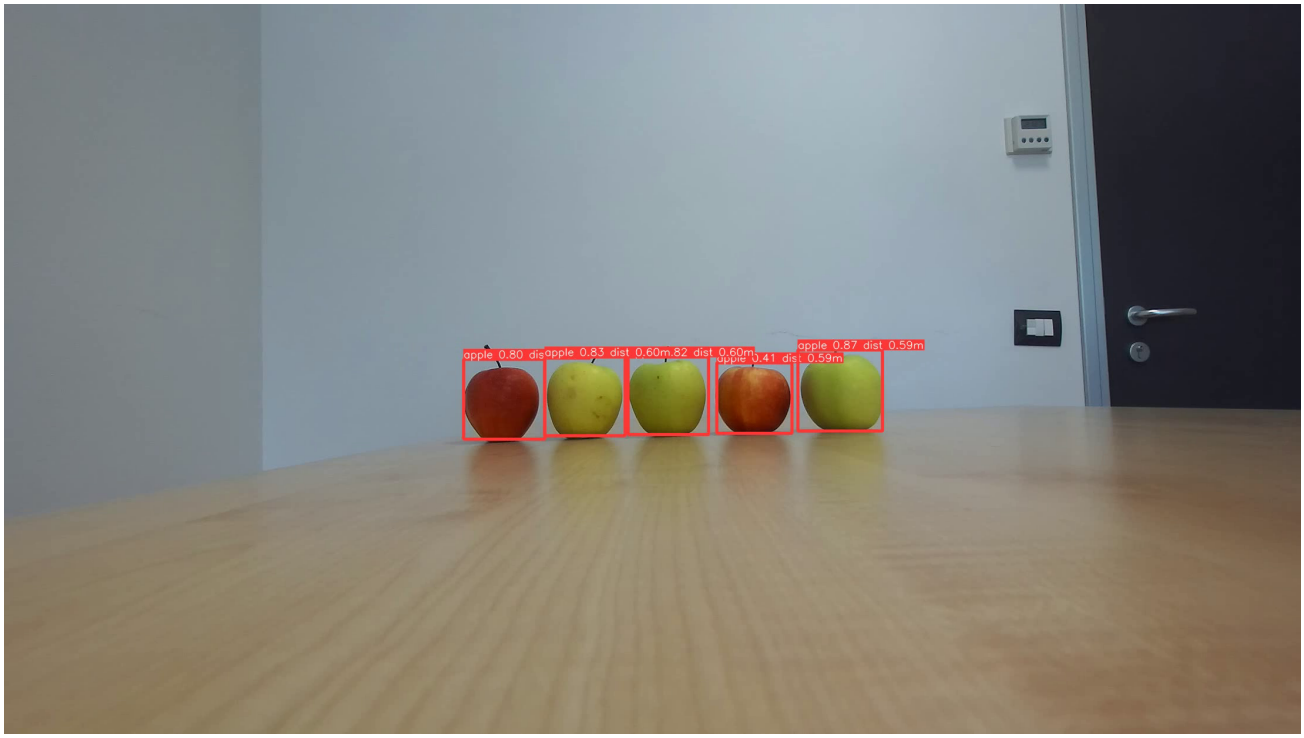

Figura 4.3: Experimento a 60cm.

Figura 4.4: Experimento a 80cm.

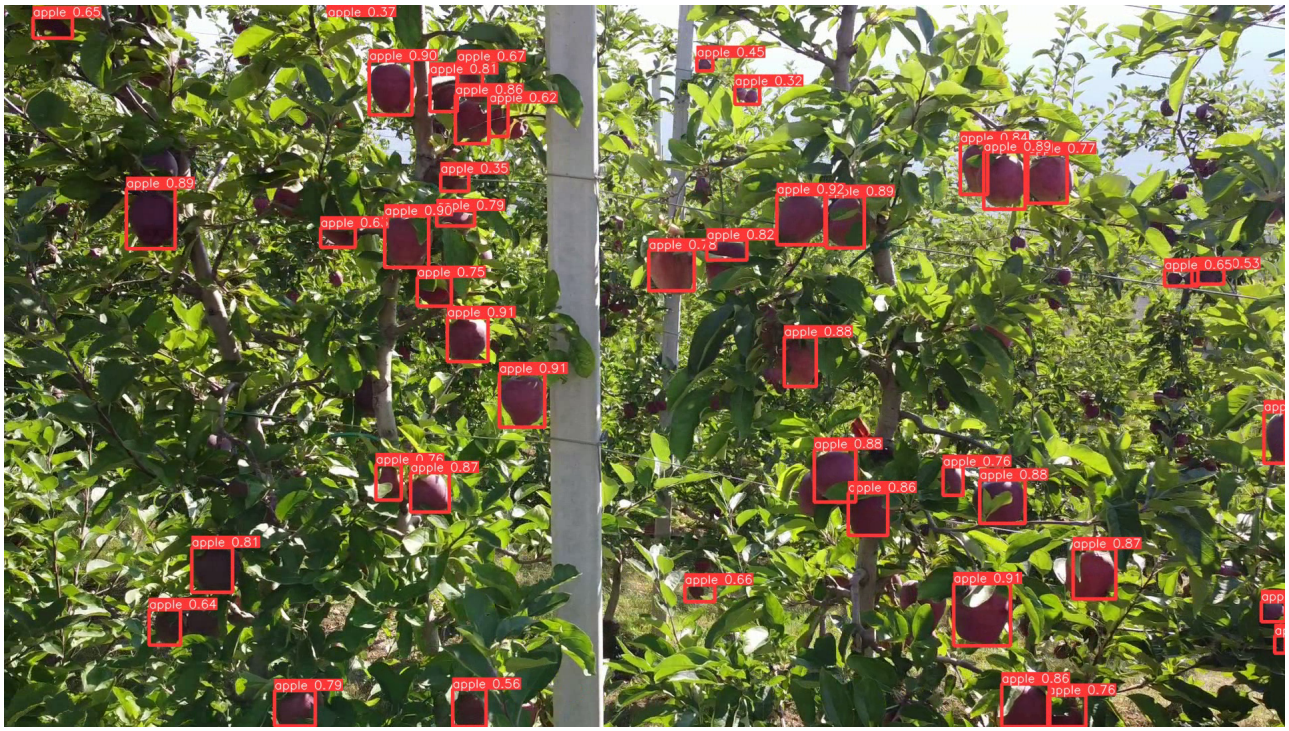

Figura 4.5: Teste realizado com dados não utilizados em treinamento.

Figura 4.6: Frame capturado contra a luz do sol, a 60cm.

Figura 4.7: Detecção no frame capturado contra a luz do sol, a 60cm.

CAPÍTULO

5

Conclusões

Neste capítulo são apresentadas as conclusões deste trabalho. Na Seção 5.1 é realizado um paralelo entre os objetivos desta tese e os resultados obtidos. Na Seção 5.2 são discutidas algumas limitações das soluções propostas e na Seção 5.3 são apresentadas algumas direções de trabalhos futuros.

5.1 Resumo dos Objetivos e Principais Resultados

Neste trabalho, exploramos a combinação entre DL e SLAM, com foco em agricultura de precisão. Por meio de uma revisão detalhada dos trabalhos relacionados, pudemos constatar o crescente interesse e sucesso dessa abordagem em aplicações como a colheita de frutas, bem como em áreas não relacionadas à agricultura de precisão.

O trabalho contou com a ajuda de alunos da Università degli Studi di Trento e de pomares de maçãs da mesma região para obtenção dos dados inicias. Realizamos a anotação manual de bounding boxes nas maçãs visíveis em 20% dos frames capturados em campo. Adicionamos ao nosso dataset, o conjunto MinneApple, a fim de diversificar e proporcionar melhor generalização aos modelos.

Treinamos 12 variações da arquitetura YOLO e conseguimos melhores resultados com a YOLOv5x, que atingiu 0.861 de F1-Score no conjunto de validação.

Desenvolvemos um script para encontrar a distância até cada maçã detectada e o integramos no fluxo de detecção da YOLO.

Os resultados obtidos demonstraram que o sistema é preciso e viável em tempo real, graças ao poder das câmeras utilizadas e à velocidade das arqui-

teturas YOLO. Todos os objetivos propostos foram alcançados com sucesso.

5.2 Limitações

No entanto, é importante destacar algumas limitações. A incompletude do mapa de profundidade calculado pela câmera pode prejudicar a precisão das estimativas de distância devido a suas flutuações. Além disso, a detecção de maçãs parcialmente oclusas, quando um objeto está em sua frente, pode afetar a confiabilidade da estimativa de distância.

5.3 Trabalhos Futuros

Como trabalhos futuros, podemos destacar a técnica de tracking, a fim de diferenciar uma fruta de outra, proporcionando assim informações mais detalhadas para um possível robô que fará a colheita. Adicionalmente, é possível melhorar a estimativa de distância, como demonstrado por CARVALHO (2023), onde é feito um pós processamento no mapa de profundidade calculado pela ZED, corrigindo valores faltantes e diminuindo ruídos.

Concluímos que a combinação entre DL e SLAM possui um potencial promissor para revolucionar a agricultura de precisão e outras áreas da robótica e visão computacional. O trabalho apresentado e as aplicações descritas nos trabalhos revisados mostraram resultados encorajadores, estimulando pesquisas futuras para aprimorar ainda mais essa abordagem, tornando-a cada vez mais precisa, eficiente e acessível para diversas aplicações no campo da automação e robótica.

Referências Bibliográficas

Adhikari, B. e Bhandari, P. (2023). Estimation of vehicular velocity based on non-intrusive stereo camera. Citado na página 7.

Bargoti, S. e Underwood, J. (2017). Deep fruit detection in orchards. Citado na página 1.

CARVALHO, M. D. A. (2023). Deep learning approaches to segment eucalyptus tree images. Citado na página 22.

Chen, X., Xue, J., Fang, J., Pan, Y., e Zheng, N. (2020). Using detection, tracking and prediction in visual slam to achieve real-time semantic mapping of dynamic scenarios. In *2020 IEEE Intelligent Vehicles Symposium (IV)*, páginas 666–671. Citado na página 8.

Costanzo, M., Simone, M. D., Federico, S., Natale, C., e Pirozzi, S. (2023). Enhanced 6d pose estimation for robotic fruit picking. Citado na página 6.

Dandekar, M., Punn, N. S., Sonbhadra, S. K., Agarwal, S., e Kiran, R. U. (2021). Fruit classification using deep feature maps in the presence of deceptive similar classes. In *2021 International Joint Conference on Neural Networks (IJCNN)*. IEEE. Citado na página 1.

Dikshit, A., Bartsch, A., George, A., e Farimani, A. B. (2023). Robochop: Autonomous framework for fruit and vegetable chopping leveraging foundational models. Citado na página 7.

Ekanayake, E. M. S. P., Thelasingha, T. H. M. N. C., Udugama, U. V. B. L., Godaliyadda, G. M. R. I., Ekanayake, M. P. B., Samaranayake, B. G. L. T., e Wijayakulasooriya, J. V. (2023). Object dimension extraction for environment mapping with low cost cameras fused with laser ranging. Citado na página 7.

Freeman, H., Qadri, M., Silwal, A., O'Connor, P., Rubinstein, Z., Cooley, D., e Kantor, G. (2022). Autonomous apple fruitlet sizing and growth rate tracking using computer vision. Citado na página 6.

Fu, A. e Kong, L. (2023). Real-time slam pipeline in dynamics environment. Citado na página 8.

Hani, N., Roy, P., e Isler, V. (2020). MinneApple: A benchmark dataset for apple detection and segmentation. *IEEE Robotics and Automation Letters*, 5(2):852–858. Citado na página 10.

James, J. A., Manching, H. K., Mattia, M. R., Bowman, K. D., Hulse-Kemp, A. M., e Beksi, W. J. (2023). Citdet: A benchmark dataset for citrus fruit detection. Citado na página 1.

Kang, H. e Wang, X. (2023). Semantic segmentation of fruits on multi-sensor fused data in natural orchards. *Computers and Electronics in Agriculture*, 204:107569. Citado na página 6.

Kang, H., Wang, X., e Chen, C. (2022). Accurate fruit localisation using high resolution LiDAR-camera fusion and instance segmentation. *Computers and Electronics in Agriculture*, 203:107450. Citado na página 6.

Krishna, G. S., Supriya, K., e Baidya, S. (2023). 3ds-slam: A 3d object detection based semantic slam towards dynamic indoor environments. Citado na página 5.

Qian, J., Chatrath, V., Servos, J., Mavrinac, A., Burgard, W., Waslander, S. L., e Schoellig, A. P. (2023). Pov-slam: Probabilistic object-aware variational slam in semi-static environments. Citado na página 5.

Qiu, A., Young, C., Gunderman, A., Azizkhani, M., Chen, Y., e Hu, A.-P. (2023). Tendon-driven soft robotic gripper with integrated ripeness sensing for blackberry harvesting. Citado na página 6.

Qureshi, A., Loh, N., Kwon, Y. M., Smith, D., Gee, T., Bachelor, O., McCulloch, J., Nejati, M., Lim, J., Green, R., Ahn, H. S., MacDonald, B., e Williams, H. (2023). Seeing the fruit for the leaves: Towards automated apple fruitlet thinning. Citado na página 5.

Thakurdesai, N., Tripathi, A., Butani, D., e Sankhe, S. (2019). Vision: A deep learning approach to provide walking assistance to the visually impaired. Citado na página 7.

Xu, C., Bonetto, E., e Ahmad, A. (2023). Dynapix slam: A pixel-based dynamic slam approach. Citado na página 5.