INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Sistema de Monitorização Visual para a Otimização da Produção de Uva de Mesa

Rodrigo Miguel Belchiorinho Alves

Mestrado em Engenharia Informática

Orientador:
Doutor Tomás Gomes da Silva Serpa Brandão, Professor Auxiliar,
Iscte - Instituto Universitário de Lisboa

Setembro, 2025

TECNOLOGIAS
E ARQUITETURA

Departamento de Ciência e Tecnologia da Informação

Sistema de Monitorização Visual para a Otimização da Produção de Uva de Mesa

Rodrigo Miguel Belchiorinho Alves

Mestrado em Engenharia Informática

Orientador:
Doutor Tomás Gomes da Silva Serpa Brandão, Professor Auxiliar,
Iscte - Instituto Universitário de Lisboa

Setembro, 2025

Agradecimentos

Este projeto representa muito mais do que a conclusão de um percurso académico, que não só reflete todos os desafios, aprendizagens e conquistas onde posso dizer que só foi possível por todo o esforço, dedicação, e sem esquecer, de todo apoio que tenho recebido desde sempre por diferentes pessoas.

Em primeiro lugar, deixo o mais profundo agradecimento ao meu orientador Professor Doutor Tomás Brandão, que foi o meu orientador neste projeto onde foi incansável a todos os níveis, desde a disponibilidade constante, a partilha do seu conhecimento em diversas áreas que foram essenciais para o desenvolvimento deste projeto.

De seguida, quero expressar a minha sincera gratidão, à pessoa que permitiu que fosse possível iniciar este projeto, ao Engenheiro Hugo Piteira e respetiva colaboração da empresa Prazer dos Aromas Unipessoal, LDA por abrir as portas da sua exploração para a realização dos testes, recolha de dados, pela partilha de todo o conhecimento, do qual posso dizer que me ajudou a perceber melhor tanto a área como o desafio em estudo.

Desta forma, agradeço principalmente por terem aceitado e apoiado este projeto, onde consegui unir duas áreas em que tenho um profundo interesse, a agricultura e a tecnologia, num desafio que transformou a forma como vejo o futuro.

Aos meus amigos, que sempre mostraram interesse pelo meu trabalho, agradeço de coração a todos por todo apoio recebido, que me motivou sempre a continuar o que estava a ser feito, mesmo que fosse um longo e duro caminho a ser percorrido.

À minha família, que é muito especial e importante neste percurso, onde sempre esteve comigo, em todos os momentos, estarei eternamente grato por tudo o que fizeram por mim. Tudo o que fiz até hoje, bem como a pessoa em que me tornei, foi graças a todos.

Assim, quero deixar um reconhecimento muito especial aos meus avós, que são para mim referências de vida e pessoas de enorme importância. Mesmo ausentes, continuarão sempre presentes em espírito, nos valores que me transmitiram e nas memórias que guardo com carinho.

Por fim, e de forma muito especial, agradeço profundamente ao meu pai, à minha mãe e à minha irmã, por terem sido sempre incansáveis e muito importantes durante todo este tempo em que estive a desenvolver este projeto, por estarem sempre ao meu lado com apoio incondicional, sempre prontos a ajudar, independentemente do que fosse necessário, e por sempre acreditarem em mim.

Resumo

Nos últimos anos, o setor agrícola tem enfrentado diversos desafios, entre os quais se destaca a monitorização precisa do estado de maturação da produção, um problema particularmente relevante para os produtores de uva de mesa. Este indicador é definido a partir da contabilização dos cachos, do número de bagos e da cor predominante. Desta forma, identificou-se uma oportunidade de melhoria na extração deste indicador, permitindo o aperfeiçoamento da gestão agrícola da produção e, refletindo diretamente na qualidade final do cacho.

O presente estudo propõe o desenvolvimento de um algoritmo capaz de detetar e analisar o estado de maturação com base no processamento de vídeos adquiridos no terreno. Assim, foi estruturada uma metodologia que integra técnicas avançadas de visão computacional, de modo a otimizar o desempenho da tarefa de monitorização.

Assim, o algoritmo utiliza o modelo YOLOv12s, que foi treinado com base num conjunto de dados específico para a deteção dos cachos, onde obteve um F1-score de 0.935 e um mAP de 0.98. Além destes resultados obtidos, o algoritmo apresentou valores muito próximos das contagens manuais realizadas em campo e uma análise de estimativas consistentes da cor predominante associada ao grau de maturação.

No que se refere à contagem de bagos, o sistema utilizou o modelo SAM2.1b+ para a segmentação de cachos e bagos, seguido de uma estimativa do número de bagos que compensa as oclusões, por meio de uma função polinomial. O método obteve uma média de 50 bagos por cacho, demonstrando resultados promissores.

Palavras-chave: Monitorização da produção; Deteção de cachos; Segmentação; Estado de maturação da uva; Aprendizagem Profunda; Visão Computacional

Abstract

In recent years, the agricultural sector has faced several challenges, among which the accurate monitoring of the ripeness of produce stands out, a particularly relevant problem for table grape producers. This indicator is defined based on the number of bunches, the number of berries, and the predominant colour. This has identified an opportunity for improvement in the extraction of this indicator, allowing for the refinement of agricultural production management and directly reflecting on the final quality of the bunch.

This study proposes the development of an algorithm capable of detecting and analysing the state of ripeness based on the processing of videos acquired in the field. Therefore, the methodology was structured, integrates advanced computer vision techniques to optimise the performance of the monitoring task.

Therefore, the algorithm uses the YOLOv12s model, which was trained based on a specific dataset for detecting bunches, where it got an F1-score of 0.935 and a mAP of 0.98. In addition to these results, the algorithm presented values very close to the manual counts performed in the field and a consistent analysis of estimates of the predominant colour associated with the degree of ripeness.

About berry counting, the system used the SAM2.1b+ model for cluster and berry segmentation, followed by a conversion from 2D to 3D counting using a polynomial function. The method obtained an average of 50 berries per cluster, demonstrating promising results.

Keywords: Production monitoring; Bunch detection; Segmentation; Grape ripeness; Deep learning; Computer vision

Índice

Lista de Figuras

Lista de Tabelas

Glossário

ASFF	*Adaptative Spatial Feature Fusion*
Bot-SORT	*Bag-of-Tricks for* SORT
CMY	*Cyan, Magenta and Yellow*
CNN	*Convolutional Neural Network*
YOLO	*You Only Look Once*
CRF	*Conditional Random Field*
DSRM	*Design Science Research Methodology*
PRISMA	*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*
HEVC	High Efficiency Video Coding
HSB	*Hue, Saturation and Brightness*
HSV	*Hue, Saturation, Value*
HTC	*Hybrid Task Cascade*
INE	Instituto Nacional de Estatística
IoU	*Intersection over Union*
LSTM	*Long Short-Term Memory*
mAP	*mean Average Precision*
MOV	*QuickTime File Format*
MOT	*Multi-Object Tracking*
R-CNN	*Region-based* CNN
ReLU	*Rectified Linear Unit*
RGB	*Red, Green and Blue*
RHS	*Royal Horticultural Society*
RMSE	*Root Mean Square Error*
ROI	*Region Of Interest*
RPN	*Region Proposal Network*
RT-DETR	*Real-Time Detection Transformer*
SAM	*Segmentation Anything Model*
SORT	*Simple Online and Realtime Tracking*
SSD	*Single Shot MultiBox Detector*
SVM	*Support Vector Machine*
UAV	*Unmanned Aerial Vehicle*

CAPÍTULO 1

Introdução

No contexto de transição digital, o setor agrícola é inquestionavelmente um setor que carece de desenvolvimento, sendo a escassez de mão de obra uma realidade que impõe muitos obstáculos a ultrapassar, tanto a nível nacional quanto a nível internacional.

Além disso, a exigência da qualidade do produto produzido tem vindo aumentar, representando um desafio diretamente relacionado com todas as fases de desenvolvimento do produto, como acontece no caso da uva destinada ao consumo direto, designada como uva de mesa.

Desta forma, é necessária uma monitorização da produção, com o intuito de melhorar todas as operações. Contudo, para realização desta tarefa é necessária muita mão de obra para que seja feito um controlo da produção.

Neste enquadramento, identifica-se uma lacuna nos contextos de investigação e de produção, nomeadamente a necessidade do desenvolvimento de novas metodologias automáticas, menos dependentes de mão de obra, que permitam realizar uma monitorização eficiente e precisa durante todas as fases de maturação da uva, e em particular, uva de mesa.

1.1. Contexto

Em Portugal, o setor agrícola tem evoluído ao longo dos anos, diversificando-se em diferentes atividades de produção para responder às exigências do mercado e às novas tendências de consumo.

Entre as diversas atividades que impulsionam a economia agrícola nacional, uma das mais importantes deste setor é a produção de uva [1], que por sua vez se divide na uva para produção de vinho e de uva para consumo direto, a designada uva de mesa. Esta última tipologia distingue-se não só pelas necessidades diferenciadas relativas à sua qualidade que são valorizadas pelos consumidores [2], como também gera valor económico relevante para os produtores [1].

A partir dos dados divulgados pelo Instituto Nacional de Estatística (INE) [3], apresentados na Figura 1, constata-se que em 2023 a produção de uva de mesa em Portugal apresenta uma superfície cultivada de 584 hectares no Alentejo, o que representa cerca de 25,9% da área total dedicada a esta cultura, onde a superfície nacional atinge os 2.255 hectares. Contudo, ao comparar a superfície nacional com a região do Alentejo, esta região apesar de ocupar uma área proporcionalmente menor, destaca-se pela sua elevada produtividade, atingindo 14.345 kg/ha, quase o dobro da média nacional.

Com base nestes dados, é possível perceber que a produção de uva de mesa no Alentejo apresenta um bom índice de produtividade. Contudo, esta produção exige um controlo rigoroso durante todas as fases do seu desenvolvimento, de forma a cumprir os requisitos do mercado nacional e internacional, que são cada vez mais exigentes.

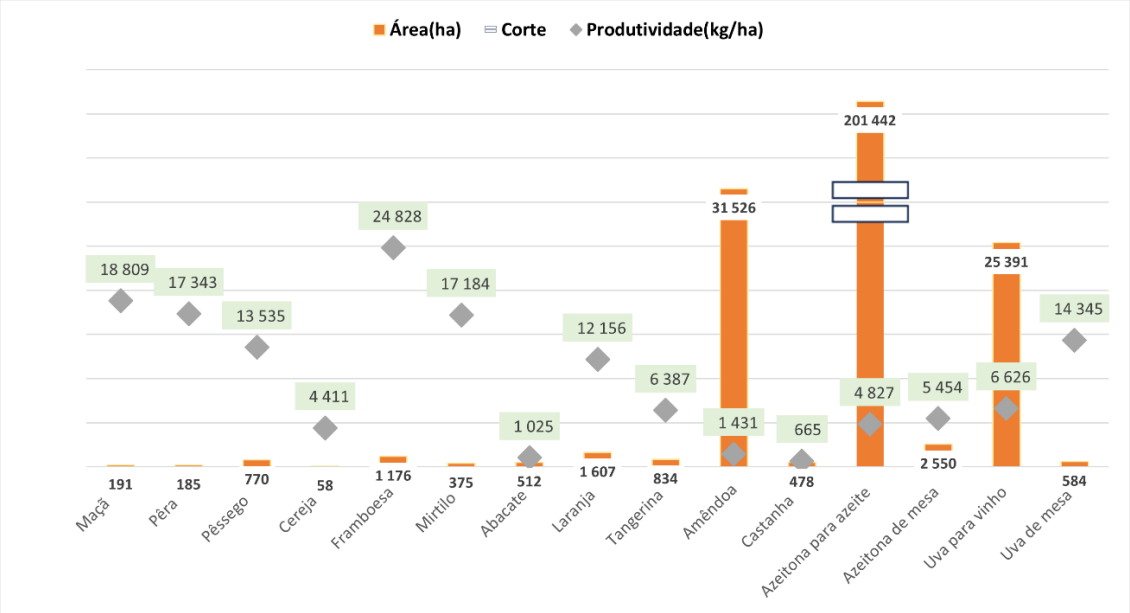

Figura 1. Produção das principais culturas agrícolas no Alentejo em 2023

De forma a garantir a qualidade do produto final, é necessário ter uma avaliação precisa do estado de maturação da produção pois este indicador influencia não só o sabor e a textura da uva, mas também o valor comercial do produto [4], [5], [6]. Este indicador está correlacionado com a quantidade de cachos contabilizados, a coloração predominante dos cachos e o número médio de bagos por cacho. Assim, podemos aferir uma previsão da produtividade. Desta forma, uma gestão eficiente e precisa do grau de maturação é essencial para garantir a qualidade da uva, assegurando um equilíbrio adequado entre açúcares, acidez e consistência do produto que é entregue ao consumidor.

No âmbito do trabalho desta dissertação, foi realizada uma recolha de dados num produtor nacional com 33 hectares no Alentejo "com peso no mercado nacional e no mercado de exportação" [1, p. 139], Prazer dos Aromas Unipessoal LDA [7], onde houve oportunidade de tomar conhecimento dos métodos de monitorização utilizados na produção, nomeadamente para avaliar o estado de maturação da variedade *Midnight Beauty*.

Esta variedade de uva sem grainha híbrida destaca-se pela sua colheita precoce, textura crocante e elevada relação entre o açúcar e o ácido, tornando-a uma opção apreciada tanto pelos produtores como pelos consumidores, foi resultado de um cruzamento entre duas variedades, pelo investigador David W. Cain em 1990 , mas que foi só patenteada em 1998 em conjunto com a *Sun World international* [8], uma empresa muito reconhecida pelo desenvolvimento de variedades inovadoras de frutas, inclusive uva de mesa.

A partir da Figura 2, podemos perceber que os bagos da *Midnight Beauty* possuem um formato médio a grande, ligeiramente alongado, com uma coloração preta densa e, por vezes, com um leve tom avermelhado. A pele, embora resistente, não interfere negativamente na experiência de consumo, pois é impercetível ao mastigar e não confere um travo amargo. A polpa, crocante e de sabor suave, apresenta bons níveis de doçura, tornando esta variedade bastante atrativa para o mercado de uvas de mesa [2].

Figura 2. Variedade Midnight Beauty

Para esta variedade, bem como para todas as outras variedades de uva de mesa produzidas neste produtor, os métodos de monitorização aplicados a esta variedade, baseiam-se atualmente em contagens realizadas manualmente por colaboradores da empresa, numa pequena área da produção que seja considerada representativa da área global plantada com a respetiva variedade de uva a ser monitorizada.

Assim foi possível constatar que esta abordagem apresenta elevadas oportunidades de melhoria, pois a metodologia atual utilizada, não é suficientemente robusta. As contagens realizadas através desta metodologia, são realizadas em áreas consideradas representativas, contudo, a comprovação desta representatividade necessita de evidências consistentes.

A ausência desta validação compromete a representatividade dos dados recolhidos, conduzindo assim a decisões menos precisas que podem afetar tanto o processo de desenvolvimento da cultura como a sua comercialização e consequente rentabilidade.

1.2. Motivação

A monitorização ineficiente do estado de maturação das uvas representa um problema significativo para os produtores, independentemente da sua dimensão. Assim, todos os produtores enfrentam dificuldades em ter dados realistas da produção, que permitam auxiliar na otimização das operações agrícolas, que são determinantes na influência direta da qualidade do produto final [6], [9].

A monitorização da maturação da uva de mesa é um processo complexo, influenciado por diversos fatores ambientais e pelo processo de gestão agrícola das parcelas. Em paralelo, os métodos tradicionais de monitorização atuais baseiam-se em avaliações e contagens visuais, limitando potencialmente a precisão e a representatividade dos dados recolhidos.

Desta forma, é possível concluir que os métodos atuais têm um potencial de melhoria para otimizar operações agrícolas bem como melhorar todo o processo de desenvolvimento do produto até à sua colheita. Caso contrário, o resultado pode influenciar por exemplo, uvas com um sabor excessivamente ácido e uma textura inadequada, ou, pelo contrário, perda firmeza e ter um teor de açúcar desajustado. Estes cenários afetam não só a perceção sensorial do consumidor, mas também a vida útil das uvas, o seu processamento logístico e o seu valor comercial.

Além do impacto na qualidade do produto, a monitorização tradicional também pode gerar consequências económicas, o uso ineficiente de recursos, de mão-de-obra e de equipamentos durante todo o processo de produção e distribuição, podendo assim aumentar os custos operacionais e reduzir a rentabilidade da produção [1], [3].

Desta forma, é identificada uma oportunidade de resolução de uma metodologia pouco eficiente no sentido de se desenvolver um sistema com um algoritmo de análise de imagens capaz de realizar uma monitorização mais precisa do estado de maturação, tornando-se uma oportunidade para fortalecer o setor e garantir o seu crescimento e sustentabilidade a longo prazo e simultaneamente, potenciando outras abordagens nas vertentes da produção e comercialização, alavancando assim ainda mais a rentabilidade desta cultura em particular.

1.3. Questões de Investigação

O objetivo desta dissertação é responder às questões de investigação apresentadas. A primeira questão de investigação (RQ1) é a principal. A segunda questão de investigação (RQ2) aprofunda

aspetos técnicos relacionados com os modelos de visão computacional utilizados e respetivos processos de treino.

- RQ1 – É possível recolher automaticamente indicadores representativos da exploração, como o número de cachos, o número médio de bagos por cacho e a respetiva coloração predominante, através de um sistema de monitorização baseado em vídeos?
- RQ2 – De entre os modelos de visão computacional mais recentes para a deteção e segmentação de objetos, quais as variantes às suas arquiteturas e aos processos de treino que conduzem a melhores resultados na deteção e segmentação de cachos e bagos de uva?

1.4. Objetivos

O principal objetivo desta dissertação é desenvolver um sistema de monitorização visual para a otimização da produção de um conjunto limitado de variedades de uva de mesa, garantindo uma avaliação mais precisa e eficiente do estado de maturação das uvas.

Desta forma, este sistema será preparado para classificar, rastrear e fornecer uma estimativa representativa do nível de maturação da produção de uva, permitindo uma tomada de decisão mais informada por parte dos produtores. Para atingir este objetivo principal, destacam-se os seguintes objetivos específicos:

- Conceptualizar e implementar um sistema capaz de identificar e contabilizar automaticamente o número de cachos presentes nas imagens capturadas, fornecendo uma estimativa mais precisa da produção.
- Implementar um sistema que classifique o estado de maturação de cada cacho detetado, garantindo uma monitorização da produção.
- Desenvolver uma solução que registe dados essenciais sobre cada cacho, incluindo as cores predominantes, o número de cachos e de bagos, permitindo uma análise complementar da produção.

Com este sistema, é expectável contribuir para um processo produtivo mais eficiente e tecnologicamente avançado, permitindo que os produtores tomem decisões mais informadas, minimizem perdas e garantam um produto final de maior qualidade, alinhado com as exigências do mercado nacional e internacional e cumulativamente com maior rentabilidade.

1.5. Metodologia

Com o intuito de alcançar o objetivo mencionado na secção anterior, foi seguida uma metodologia que permitiu estruturar o desenvolvimento e avaliação do sistema proposto, a *Design Science Research Methodology* (DSRM) [10], conforme apresentado na Figura 3.

A primeira etapa do DSRM corresponde numa abordagem centrada no "*Identify Problem & Motivate*", dado que o problema que esta dissertação procura resolver foi previamente identificado. Desta forma, a contextualização, a motivação, os objetivos e os problemas identificados, foram mencionados anteriormente ao longo deste capítulo. Assim, este ponto de entrada de investigação envolve a definição clara da necessidade de um sistema automatizado capaz de avaliar de forma precisa o estado de maturação das uvas, contabilizar a produção e fornecer informações relevantes para os produtores.

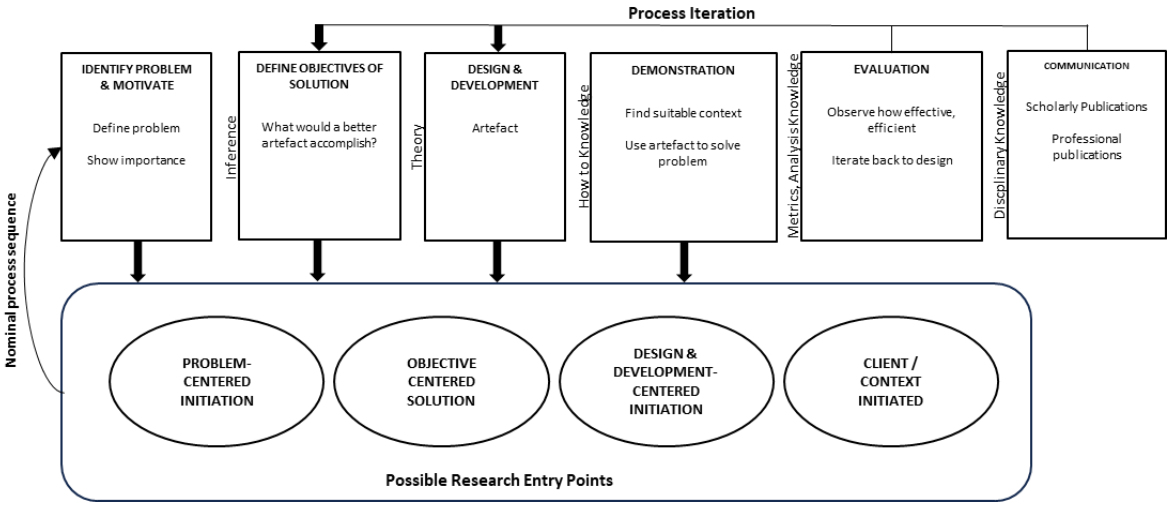

Figura 3. Modelo DSRM

A segunda etapa do DSRM, "*Define Objectives of Solution*", foi também abordada no Capítulo 1, onde são definidos os objetivos principais e específicos do projeto a ser desenvolvido. De seguida, a etapa "*Design & Development*" aborda o desenvolvimento da solução proposta, que neste trabalho consiste em desenvolver um sistema de visão computacional capaz de analisar o estado de maturação dos cachos de uva detetados numa determinada área de exploração.

Na quarta etapa do DSRM, "*Demonstration*", a solução desenvolvida no processo de pesquisa é aplicada ao problema real identificado, com um foco específico no contexto do cliente. Assim, para além da implementação da solução proposta também é realizada uma adaptação e personalização da solução às necessidades concretas e ao ambiente em que será utilizada do produtor onde foram recolhidos os dados.

A quinta etapa, "*Evaluation*", corresponde à fase de testes e validação do sistema, garantindo a sua eficácia e precisão. Por fim, a etapa "*Communication*" assegura a disseminação dos resultados da Dissertação, documentando as conclusões e disponibilizando-as à comunidade científica e a potenciais interessados, incentivando a continuidade do estudo e futuras melhorias.

1.6. Estrutura do Documento

O desenvolvimento do sistema proposto foi resultado de um planeamento detalhado de cada uma das etapas deste projeto, no qual se definiu uma metodologia capaz de garantir o cumprimento dos objetivos estabelecidos dentro do período previsto. Assim, a organização deste documento reflete todo o planeamento seguido, apresentando-se de forma ordenada e alinhada com as diferentes fases de desenvolvimento do projeto, apresentadas na Figura 4.

Durante todo o processo, houve sempre um contacto frequente com o orientador através de reuniões regulares, o que possibilitou a discussão contínua sobre os desafios encontrados, a definição das próximas tarefas e a realização da escrita deste documento. Além disso, também foi estabelecida comunicação direta com a Administração da área de uva de mesa da empresa Prazer dos Aromas Unipessoal Lda [7].

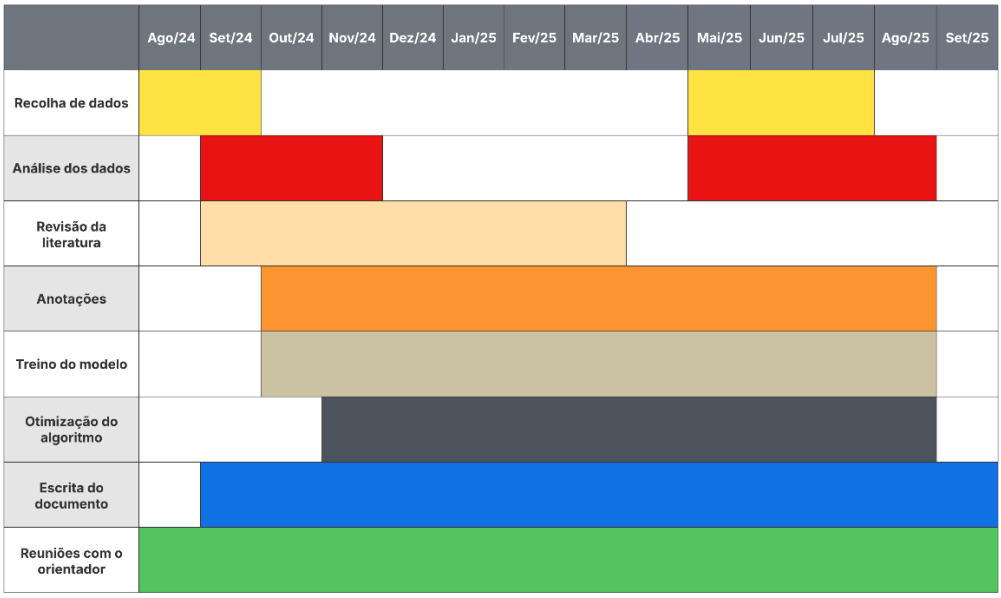

Figura 4. Duração de cada uma das fases de desenvolvimento do sistema proposto

A primeira fase deste projeto, está descrita no Capítulo 1, onde é apresentada uma introdução detalhada do tema desta Dissertação. Esta etapa permite entender melhor o contexto do estudo, a identificação das questões de investigação especificas.

Paralelamente, foi necessário entender o estado da arte perante o contexto em estudo, com o intuito de definir a abordagem mais adequada para alcançar o objetivo definido. Desta forma, o Capítulo 2 descreve toda a revisão da literatura existente sobre os métodos de monitorização do estado de maturação de frutas, deteção de objetos e métodos de visão computacional aplicados para uva de mesa. A partir da revisão foi delineado o desenvolvimento de um modelo de deteção, e a utilização de um modelo de segmentação.

Em seguida, o Capítulo 3 apresenta de forma detalhada os procedimentos adotados para o desenvolvimento do sistema, incluindo o método de recolha dos conjuntos de dados, a análise dos dados, a anotação dos dados, o treinamento do modelo e o processo de melhoria e otimização do algoritmo, conduzido por meio de diversos testes realizados ao longo de toda a fase do seu desenvolvimento. Todas estas fases do projeto.

No Capítulo 4 são analisados os resultados obtidos através da aplicação do sistema desenvolvido. Desta forma, esta fase permitiu entender melhor o desempenho do sistema, bem como a identificação dos principais desafios observados durante a execução dos testes.

Por último, o Capítulo 5 tem como objetivo apresentar uma visão geral do sistema desenvolvido, respondendo às questões de investigação definidas, bem como evidenciar as limitações identificadas no trabalho e possíveis direções para trabalhos futuros.

CAPÍTULO 2

Revisão da Literatura

Este capítulo apresenta e contextualiza os conceitos de base para o desenvolvimento desta dissertação, das metodologias e técnicas utilizadas e uma análise dos trabalhos relacionados. Além disso, esta secção estabelece também a base teórica necessária para a abordagem proposta, facilitando a compreensão das escolhas metodológicas adotadas ao longo do estudo.

2.1. Redes Neuronais Convolucionais

As Redes Neuronais Convolucionais (CNN) são um tipo de rede neuronal artificial amplamente utilizada para o processamento e análise de imagens. Deste modo, diferenciam-se das redes neuronais clássicas pela sua arquitetura, que aproveita a estrutura espacial das imagens para extrair características relevantes em diferentes níveis de abstração.

As CNN organizam os pixéis em blocos de pequenas dimensões, permitindo que os neurónios interajam apenas com regiões específicas da imagem. Assim, esta estrutura facilita a deteção de padrões e a redução da complexidade computacional [11].

A arquitetura das CNN é composta por diferentes tipos de camadas, sendo as principais:

- Camadas convolucionais: Estas aplicam filtros sobre a imagem para extrair características, como contornos, texturas e formas. Os coeficientes dos filtros são ajustados durante o treino da rede e constituem os pesos aprendidos pelo modelo.
- Função de ativação *ReLU* [12]: É uma função de ativação não linear aplicada após a utilização das camadas convolucionais, que substitui valores negativos por zero, que consequentemente melhora a eficiência computacional e reduz o risco do problema do *vanishing gradient* [13].
- Camadas de subamostragem: Reduzem a dimensionalidade dos mapas de características, tornando a rede mais eficiente e menos sensível a pequenas variações na posição dos objetos.
- Camadas densas: Estas são semelhantes às camadas das redes neuronais clássicas, sendo responsáveis pela tomada de decisão final, combinando as características extraídas nas camadas anteriores para realizar classificações e determinadas tarefas específicas, como a deteção de objetos e a segmentação de imagens.

2.2. Modelos de Deteção de Objetos

A deteção de objetos consiste na identificação e localização de elementos visuais específicos numa imagem ou num vídeo. Atualmente, os métodos de deteção baseiam-se sobretudo em CNN e podem ser classificados em duas categorias principais: modelos de etapa única e modelos de duas etapas.

Os modelos de etapa única eliminam a necessidade da fase de proposta de regiões, realizando a deteção e a classificação simultaneamente numa única passagem pela rede neuronal. Este processo é geralmente mais rápido, sendo ideal para aplicações de que exigem alto desempenho em tempo real.

Atualmente, estes são os modelos de etapa única mais utilizados:

- YOLO (*You Only Look Once*) [14]: Este é um dos modelos mais eficientes na deteção de objetos em tempo real, pois consegue processar uma imagem completa de uma só vez e prevendo múltiplas *bounding boxes* simultaneamente.
- SSD (*Single Shot MultiBox Detector*) [15]: É um modelo de deteção de objetos semelhante ao YOLO [14], diferencia-se por utilizar *feature maps* de múltiplas camadas da rede para realizar predições em diferentes escalas, ou seja, possibilita a deteção de objetos com diferentes tamanhos de forma mais consistente, ao mesmo tempo em que mantém elevada eficiência computacional e baixo tempo de inferência.
- RetinaNet [16]: É um modelo que introduz a *Focal Loss*, que melhora a deteção de objetos pequenos e raros em comparação com YOLO [14] e SSD [15].
- EfficientDet [17]: Este modelo é mais recente e têm a capacidade de otimizar a eficiência computacional mantendo uma elevada precisão de deteção.
- RT-DETR (Real-Time Detection Transformer) [18]: É um modelo de deteção de objetos que utiliza uma arquitetura baseada em *Transformers*. Este modelo, ao contrário dos outros, que dependem de etapas separadas para geração de propostas e classificação, este adota um design *end-to-end*, simplificando o processo de deteção e aumentando sua eficiência. Além disso, o modelo processa de forma eficaz características em múltiplas escalas, separando a interação intra-escala da fusão inter-escala para melhorar a qualidade das predições.

Além destes modelos, existem os modelos de duas etapas, que são caracterizados por identificar em primeiro lugar as regiões de interesse onde os objetos podem estar presentes, posteriormente estas regiões que são refinadas, ou seja, as localizações das regiões são ajustadas com melhor precisão e posteriormente são classificadas. Embora geralmente este método se mostre mais preciso, é mais lento devido ao processamento adicional necessário.

Atualmente, estes são alguns modelos de duas etapas mais utilizados nos sistemas:

- R-CNN (*Region-based* CNN) [19]: Este modelo extrai as regiões de interesse (ROIs) da imagem e aplica uma CNN para classificar os objetos presentes.
- *Fast* R-CNN [20]: Otimiza a R-CNN [19], reduzindo o tempo de inferência ao processar toda a imagem em uma única passagem pela rede.
- *Faster* R-CNN [21]: Introduz uma Rede de Propostas de Região (RPN) [22], acelerando o processo de deteção.
- *Mask* R-CNN[11]: É uma extensão da *Faster* R-CNN [21] que adiciona um mecanismo de segmentação semântica para gerar máscaras dos objetos identificados.

2.3. Modelos de Segmentação de Objetos

Os modelos de segmentação de objetos desempenham um papel fundamental no domínio da visão computacional, cuja finalidade é delimitar e classificar regiões específicas de uma imagem que correspondem a diferentes objetos. A segmentação tem como objetivo atribuir uma etiqueta a cada pixel da imagem, resultando numa representação mais precisa do objeto identificado.

A segmentação de objetos pode ser realizada de duas formas diferentes:

- Segmentação semântica - Atribui uma categoria a cada pixel da imagem, mas sem distinguir instâncias individuais de objetos da mesma classe. Por exemplo, todos os pixels correspondentes a "cachos de uvas" seriam rotulados como tal, mas sem diferenciar cada cacho.
- Segmentação por instâncias - Além de classificar os pixels, distingue cada ocorrência individual de um objeto. Neste caso, por exemplo, cada cacho de uva seria identificado separadamente, permitindo análises mais detalhadas.

Ao longo dos últimos anos, diversos modelos de segmentação têm sido propostos, alguns tornando-se referência devido ao seu desempenho, robustez e ampla adoção em diferentes áreas estudadas pela comunidade:

- Mask R-CNN [11] - Mencionado anteriormente, além de realizar a deteção de objetos também gera máscaras de segmentação para cada instância. Este modelo é considerado um dos modelos mais importantes para segmentação por instâncias.
- DeepLab [23] - Desenvolvido pela Google, utiliza convoluções Atrous (*dilated convolutions*) [24][25] e mecanismos de *Conditional Random Field* (CRF) [26] para obter resultados mais precisos em segmentação semântica.

- *Segment Anything Model* (SAM) [27] - Desenvolvido pela Meta AI, é um modelo generalista de segmentação capaz de identificar qualquer objeto em imagens com base em *prompts* (pontos, caixas ou máscaras iniciais). Destaca-se pela sua generalização e pela capacidade de ser aplicado em diferentes domínios sem necessidade de treino extenso.

2.4. Revisão Sistemática

Para garantir um conhecimento atualizado e relevante na área de estudo abordada nesta dissertação, foi conduzida uma pesquisa aprofundada em fontes de informação científicas e académicas. A partir desta pesquisa, foram identificados e analisados estudos relacionados que tratam de temas essenciais que estão enquadrados com a área de investigação e que sustentam a solução proposta.

Para a realização desta pesquisa, foram selecionadas duas fontes de informação de renome no contexto científico, a *Web of Science* [28] e o Scopus [29]. A primeira fonte é reconhecida pela seleção rigorosa de periódicos de maior prestígio e pelo histórico extenso de citações, enquanto a segunda oferece uma cobertura mais ampla e interdisciplinar. Desta forma, a integração destas duas fontes garante uma base sólida e atualizada para sustentar as decisões tomadas na solução proposta nesta dissertação.

Com o intuito de recolher artigos relevantes para o trabalho proposto, foi realizada uma pesquisa avançada utilizando uma *query* específica. Com base nesta *query*, foi possível obter um conjunto de artigos diretamente relacionados com os objetivos do trabalho.

A *query* utilizada, conforme definida na Tabela 1, descreve de forma detalhada todas as etapas envolvidas na definição da *query* utilizada para o estudo. A tabela organiza as diversas palavras-chave empregadas, bem como a aplicação dos campos definidos em cada uma das colunas, permitindo um entendimento claro da forma como foi possível restringir o grupo de artigos relevantes.

Tabela 1. Método de definição da query

CONCEITOS	TÉCNICAS	CONTEXTO	RESTRIÇÕES
"*deep learning*" "*neural network*" "*computer vision*"	"*detection*" "*tracking*" "*classification*"	"*grape* AND (*bunch* OR *berries*)"	"English AND (Review Article OR Article)

Para a definição da *query*, foram selecionados conceitos e técnicas fundamentais, como "*deep learning*", "*neural network*", e "*computer vision*", associando-os com temas de "*detection*", "*tracking*" e "*classification*". A combinação destes conceitos e técnicas com o contexto relacionado à produção de uvas ("*grape* AND (*bunch* OR *berries*)"). A formulação da *query* seguiu uma lógica booleana, aplicando "AND" entre as diferentes colunas de termos e "OR" entre os elementos de cada coluna, conforme apresentado na Tabela 1.

Este método de definição da *query*, reflete a intenção de focar em artigos que abordem esses tópicos no contexto da viticultura, especificamente em relação ao uso de técnicas de visão computacional para deteção e análise de uvas. Além disso, foi aplicado um filtro específico que restringiu os resultados a artigos em inglês e classificados como "*Review Article*" ou "*Article*".

De forma a garantir um processo sistemático e transparente na seleção dos artigos a serem utilizados na secção 2.5, foi aplicado o método PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*) [30]. Esta metodologia permitiu organizar e documentar as diferentes fases do processo de revisão, desde a identificação inicial dos estudos até à inclusão final, assegurando a reprodutibilidade e a qualidade da análise.

A Figura 5 ilustra o diagrama de fluxo PRISMA gerado, que detalha cada etapa do ajuste do conjunto de artigos, evidenciando o número de artigos identificados, excluídos e, finalmente, selecionados para análise.

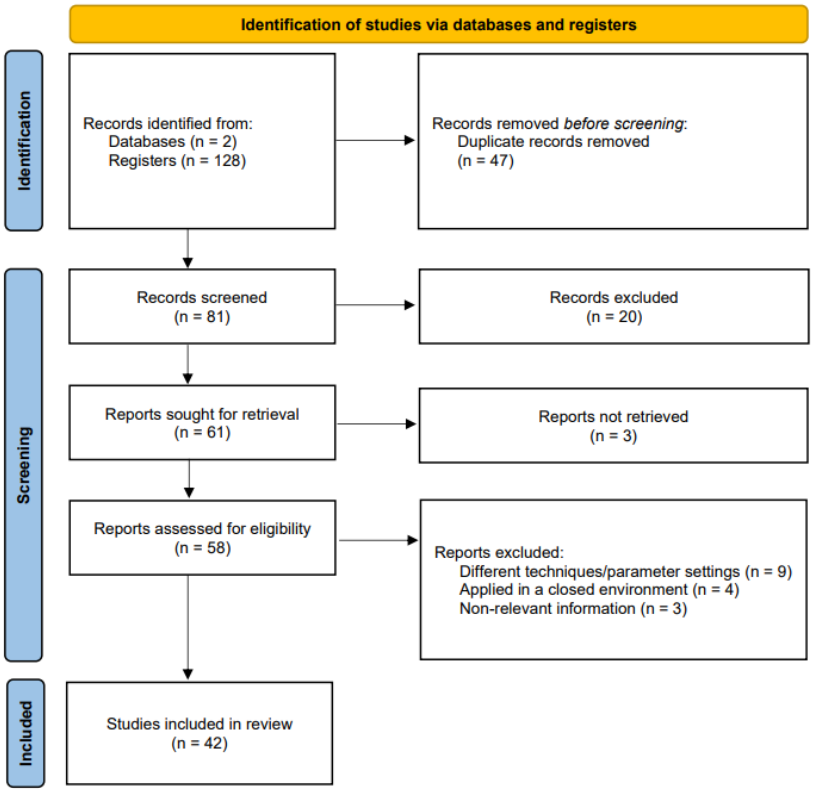

Figura 5. Diagrama de fluxo PRISMA

Na fase de identificação, foi realizada uma seleção inicial de artigos não duplicados provenientes das duas fontes de informação mencionadas. Desta forma, foram identificados 128 artigos no total, mas após a identificação de 47 duplicados, o número de artigos únicos foi reduzido para 81.

De seguida, os artigos únicos que foram armazenados, foram avaliados com base nos seus títulos e resumos, tendo como referência os critérios de inclusão e exclusão previamente definidos. Nesta fase, 20 artigos foram removidos do conjunto de artigos definido anteriormente, pois não atendiam aos requisitos estabelecidos. Desta forma, o conjunto de artigos passou a ser igual a 61 no total.

Assim, foi pretendido o acesso de cada artigo do conjunto atualizado, exigindo a requisição completa dos documentos. No entanto, 3 artigos não puderam ser recuperados, restando 58 artigos disponíveis na fase de avaliação de elegibilidade.

Cada um destes artigos foi submetido a uma análise aprofundada, com o intuito de excluir artigos com informações não relevantes e desenquadrados em relação à solução proposta dissertação. Os critérios de exclusão definidos para esta revisão foram os seguintes:

- Técnicas e configurações diferentes: 9 artigos utilizavam metodologias ou configurações de parâmetros distintos dos adotados nesta dissertação. Além disso, os equipamentos utilizados nestes estudos permitiam a recolha de dados que não são viáveis na solução definida, focando-se em análises mais aprofundadas da uva, o que não se alinha com os objetivos do trabalho.
- Ambiente de estudo controlado: 4 artigos reportavam estudos conduzidos em ambientes fechados, onde as condições eram rigorosamente controladas pelos investigadores. No entanto, nesta dissertação, tanto a recolha de dados bem como a aplicação da solução definida ocorrem no terreno, sob condições reais e variáveis.
- Falta de informações relevantes: 3 artigos não apresentavam dados significativos para o objetivo desta dissertação, não contribuindo de forma relevante para a análise pretendida.

No final deste processo, 42 estudos foram considerados adequados e incluídos na revisão sistemática, pois cumpriam todos os critérios estabelecidos e forneciam informações relevantes para o desenvolvimento desta dissertação.

2.5. Trabalho Relacionado

Com base nas análises realizadas durante todo o processo definido na secção anterior, verifica-se um crescimento significativo no desenvolvimento de métodos computacionais aplicados à viticultura de precisão, com o objetivo de otimizar diversos processos na produção de uva.

De forma geral, as técnicas de deteção, segmentação e classificação têm sido amplamente exploradas para resolver desafios como a identificação automática de cachos e bagos e a monitorização do estado das vinhas.

Com o objetivo de compreender melhor os temas mais abordados nos estudos recolhidos, foi gerada uma *Wordcloud*, apresentada na Figura 6. A *Wordcloud* revela diversos conceitos-chave que refletem as principais tendências e desafios da pesquisa na área. Entre os termos mais proeminentes, destacam-se aqueles relacionados à "*detection*", "*segmentation*", "*model*" e "images", evidenciando uma forte ênfase no uso de visão computacional e para análise de "*grape*".

A presença de termos como "*bunch*", "*berries*" e "*vineyard*" sugere que o foco das investigações está na caracterização detalhada da fruta e das suas unidades de análise. Além disso, palavras como "*counting*", "*yield*", "*quality*", "*estimation*", "*maturity*" e "*color*" indicam que um dos principais objetivos destes estudos é quantificar e prever a produtividade e qualidade das uvas com base em imagens.

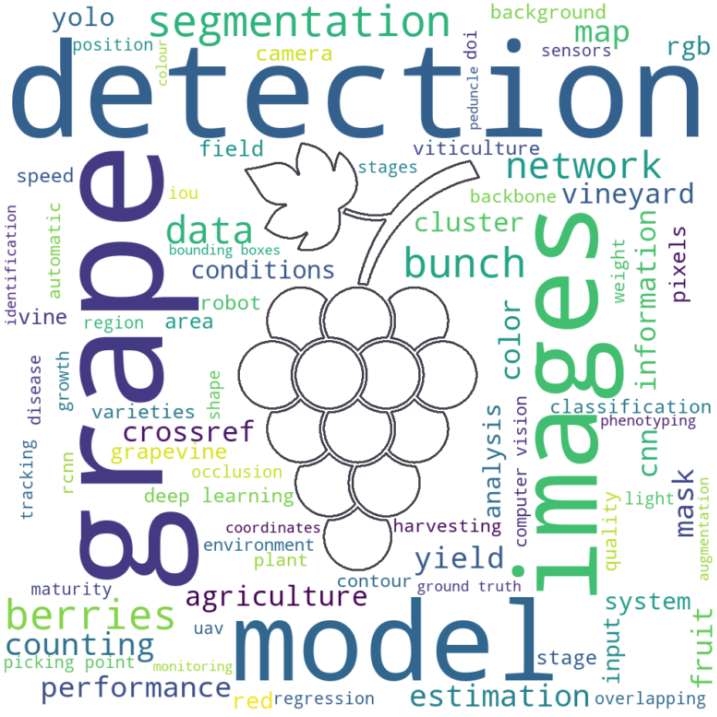

Figura 6. Wordcloud

Outro ponto relevante é a recorrência de termos técnicos associados à inteligência artificial, como "*network*", "*cnn*", "*deep learning*" e "*model*", evidenciando a predominância de abordagens baseadas em redes neurais convolucionais para analisar o estado das uvas através de elementos visuais. Além disso, palavras como "*segmentation*", "*detection*", "*tracking*", "*yolo*" e "*mask*" apontam para a utilização de técnicas avançadas de processamento de imagens para identificação de cachos e bagas.

A *Wordcloud* também destaca um interesse crescente na automação agrícola, com a presença de termos como "*robot*", "*sensors*", "*camera*" e "*field*", indicando o uso de sensores óticos e sistemas robóticos para monitorização e colheita da uva, com o intuito de otimizar os processos produtivos e reduzir custos operacionais.

Finalmente, palavras como "*disease*", "*occlusion*", "*overlapping*" e "*conditions*" sugerem desafios enfrentados na análise visual das uvas, tais como a deteção de doenças e as dificuldades associadas a condições adversas de iluminação e sobreposição de cachos, o que exigem técnicas avançadas para garantir um processamento robusto e preciso.

Desta forma, é possível perceber que todas as palavras em destaque na *Wordcloud* e que foram mencionadas anteriormente, comprovam que nos últimos anos a viticultura de precisão tem passado por uma transformação significativa, impulsionada pela integração de tecnologias avançadas com o objetivo de alcançar uma gestão mais eficiente e sustentável da produção de uva. A estimativa precisa é crucial na viticultura, pois o número de cachos por videira é responsável por 60% da variabilidade no rendimento do campo, destacando-se por isso a importância de uma contagem precisa [31].

Durante esta transformação foi possível perceber que existiu uma evolução das técnicas de visão computacional aplicadas, pois têm sido continuamente aperfeiçoadas através das novas metodologias e avanços científicos. Inicialmente, tinham como base a análise de imagens, mas estas técnicas evoluíram para abordagens mais complexas, incorporando modelos de aprendizagem automática, em particular os que utilizam aprendizagem profunda [32].

Desta forma, vários estudos têm explorado diferentes tipos de aplicações na viticultura como a segmentação e contagem de cachos e bagos de uva, a identificação de características fenotípicas, a deteção automática de variedades de uvas, doenças e pragas presentes nas mesmas, bem como a otimização dos processos de desbaste de cachos.

Com base nos estudos selecionados, foi possível entender que existe um amplo conjunto de materiais e de estratégias utilizadas tanto no processo de extração como também na interpretação das informações relevantes para as técnicas mencionadas anteriormente. De forma a apresentar todas as abordagens relevantes no âmbito desta dissertação, foi realizada uma seleção por temas de modo a organizar toda a informação recolhida nos artigos analisados.

2.5.1. Estratégias de Deteção de Cachos

A deteção e contagem de cachos de uva de forma precisa, desempenham um papel fundamental na análise do estado das vinhas, permitindo uma gestão mais eficiente da produção. A recolha destas informações possibilita uma avaliação aprofundada da qualidade e desenvolvimento da uva, contribuindo para decisões mais informadas ao longo do ciclo produtivo.

Para tal, têm sido aplicadas técnicas baseadas em modelos de deteção de objetos, como o YOLO [14]. Um exemplo é o trabalho apresentado em [33], onde se realiza a deteção em tempo real de cachos de uvas destinadas a vinho de alta qualidade. O modelo desenvolvido foi testado nas variedades *Chardonnay* e *Merlot*, em diferentes cenários, onde obteve resultados muito interessantes, alcançando uma sensibilidade de 0.95, que representa a proporção de casos positivos reais que foram corretamente identificados pelo modelo.[25]

Além da deteção, é também essencial classificar os cachos de uva de acordo com o seu estado de saúde, distinguindo entre uvas saudáveis e danificadas com base na presença de lesões biofísicas nos bagos [34]. Neste contexto, relacionado na deteção de características do cacho, foi desenvolvido um sistema capaz de estimar, em tempo real, o peso da colheita, contribuindo para uma previsão mais precisa da produção [35].

Para garantir a fiabilidade destes sistemas de deteção, é necessário selecionar cuidadosamente a versão do YOLO [14] a utilizar. Desta forma, é necessário avaliar diferentes versões que foram treinadas com conjuntos de dados heterogéneos, compostos por imagens com diferentes características.

Num estudo analisado em [36], foi testado particularmente em variedades de uvas brancas, devido à maior dificuldade na deteção de bagos claros sobre um fundo foliar. Os resultados evidenciaram a elevada precisão da versão YOLOv5x na identificação de cachos, enquanto a versão YOLOv4-tiny demonstrou um equilíbrio eficaz entre precisão e velocidade. Contudo, recomenda-se a melhoria do desempenho dos modelos em situações de oclusão, com o objetivo de aumentar a precisão na contagem dos cachos e reduzir falsos positivos e negativos.

A fim de desenvolver uma metodologia não invasiva e robusta em diferentes condições de oclusão, é crucial a recolha de imagens em ambiente de campo [31]. Assim, a construção de um conjunto de dados abrangente, que inclua os cenários mais complexos em que os cachos de uva possam ser encontrados, torna-se essencial. A deteção em tempo real, a precisão e o tamanho dos modelos são fatores determinantes para o sucesso destas soluções [37].

Um conjunto de dados bem estruturado não só orienta o modelo na aprendizagem, como também facilita todo o processo de desenvolvimento do sistema. Desta forma, a otimização dos dados e da arquitetura de um modelo como a do YOLOv5 [38], influencia diretamente o desempenho final. Estas melhorias incluíram por exemplo a substituição da camada convolucional inicial, a introdução de operações de atenção convolucional e a reformulação de módulos como o *bottleneck*, etc. Estas alterações visam reduzir o tamanho do modelo, simplificar o seu funcionamento e melhorar o desempenho.

Desenvolver modelos mais leves permite a sua aplicação em dispositivos com capacidade computacional reduzida. Neste sentido, foi projetada uma nova arquitetura *backbone* leve [39], que reduziu significativamente o número de parâmetros em comparação com o modelo YOLOv4. Adicionalmente, foi incorporado um mecanismo de fusão de *Adaptative Spatial Feature Fusion* (ASFF), concebido para lidar com os desafios impostos por sinais de elevada densidade e oclusão. Estas inovações resultaram numa melhoria do mAP e numa redução de recursos computacionais, apresentando-se como uma solução promissora para aplicações reais.

Através da incorporação de outras melhorias ao nível da arquitetura, tornou-se possível identificar de forma rápida e precisa cachos de uva com crescimento denso e ocluídos em ambientes não estruturados [40]. Os resultados mostraram ganhos de desempenho significativos, com aumentos até 11,47% na precisão média e até 23,33% no recall, especialmente para uvas vermelhas e verdes em cenários desafiadores.

Uma solução inovadora [4] foi a criação de óculos inteligentes que para além de fazerem a deteção de cada bago de uva presente no cacho, tem como objetivo a determinação da adequação das uvas para colheita com base na cor apresentada, utilizando um *autoencoder-based anomaly detection model* e um modelo de estimativa de cor.

Para além da avaliação do estado das uvas, é também essencial monitorizar o crescimento e a saúde das videiras. Neste âmbito, foi desenvolvido um sistema de monitorização inteligente, capaz de identificar oito patologias comuns através de um algoritmo de busca seletiva, contribuindo para um acompanhamento mais eficiente da vinha [41].

Adicionalmente, uma nova abordagem foi proposta para a contagem automática de bagos de uva utilizando câmaras de *smartphones* [42]. O funcionamento deste sistema consiste no processamento de imagens recolhidas pelo *smartphone* pelo modelo YOLOv7, que conseguiu alcançar uma precisão média de 0.97, demostrando um alto desempenho nesta tarefa de deteção. Contudo, foram observadas limitações, como a perda de bagas nas extremidades dos cachos e a deteção errada de elementos de fundo.

2.5.2. Estratégias de Segmentação e de Contagem

O uso de modelos de *Deep Learning*, em particular das CNN, tem se consolidado como uma ferramenta fundamental no processamento de imagens aplicadas à viticultura, pois estas têm demonstrando um elevado desempenho na deteção e contagem de flores e bagos de uva em tempo real. Assim, modelos como o *Mask* R-CNN [11] e o YOLOv4 destacam-se por sua elevada taxa de sucesso, atingindo até 99% de precisão, e a sua combinação com técnicas de reconstrução SLAM 3D [43], o que permite melhorar substancialmente a precisão da contagem em campo [32].

Para se obter melhores resultados, é essencial avaliar comparativamente diferentes arquiteturas de redes neurais, de forma a identificar a solução mais eficaz para tarefas específicas. Esta análise permite ainda perceber quais aspetos podem ser otimizados dentro de cada arquitetura para maximizar o desempenho [44].

A previsão precisa do rendimento da vinha é um fator chave para a gestão eficiente da produção, por isso foi proposto um sistema que recorre à regressão do peso dos cachos para estimar a produção. Este sistema integra diferentes componentes, como o *Mask* R-CNN [11] com *Swin Transformer* [45] para segmentação e identificação dos cachos, um módulo de rastreamento baseado no SIAMFC [46], e uma abordagem baseada em densidade para contagem de bagas. Os resultados obtidos revelam erros inferiores a 5% em dois dos três painéis analisados, com um painel a apresentar um erro de aproximadamente 15% devido a sobreposição de frutos [9].

No entanto, a análise detalhada da produção, com base na segmentação e contagem precisa dos bagos nos cachos de uva, através da utilização da *Mask* R-CNN [11] revelou-se superior face a outras alternativas. Contudo, para contornar as limitações associadas à oclusão dos cachos, foi implementado um método de pós-processamento com ponderação linear que aumentou a precisão da deteção [47].

Assim, foi demonstrado que é possível utilizar uma única rede neuronal para gerir todo o fluxo de análise, desde a deteção dos objetos até à contagem de suas partes internas. Esta abordagem centralizada permite integrar múltiplos módulos com funções distintas numa única rede, facilitando o processamento e a eficiência do sistema [48].

No domínio da segmentação de cachos de uvas, foi testada a utilização de modelos pré-treinados [49], que através da seleção otimizada dos mapas de probabilidade, tentaram superar os desafios típicos da segmentação em ambientes naturais. Contudo, os resultados revelam que, apesar da sua utilidade, estes nem sempre oferecem a melhor precisão para este tipo de aplicação.

Entre as diferentes arquiteturas avaliadas na segmentação dos cachos, a que apresentou melhor desempenho foi o *Mask* R-CNN [11] com *backbone* ResNet101 [50], o qual além de segmentar com alta precisão, também é capaz de estimar o nível de maturação dos cachos, apresentando um mAP de 0.93 e precisão geral de 0.94 [5].

O uso de câmaras RGB-D de baixo custo, que permitem recolher imagens no espetro de cores tradicional (*Red, Green and Blue*) e gerar um mapa de profundidade que indica a distância de cada ponto até à câmara. Deste modo, estas câmaras ao serem instaladas em tratores agrícolas, tem possibilitado a recolha de conjunto de imagens para segmentação semântica em viticultura, representando uma alternativa prática e económica para aplicações em campo [51].

Além disso, também têm sido exploradas técnicas de *Deep Learning* para a classificação automática de variedades de uvas, com base em características ampelográficas, ou seja, características morfológicas visíveis da videira, como a forma e o perfil das folhas, assim como a densidade dos cachos. Um modelo CNN de 15 camadas, desenvolvido especificamente para esta tarefa, alcançou uma precisão de 0.94 na classificação de folhas e 0.97 na de frutos, superando modelos pré-treinados [52].

Outro sistema proposto, com base na integração entre *Mask* R-CNN [11] e calibração quadriculada, permite detetar automaticamente o início da fase de latência em bagos, com base na estimativa do seu diâmetro e crescimento ao longo do tempo [6].

No contexto do desbaste dos cachos de uva, foi desenvolvido um sistema baseado numa versão otimizada do DeepLabV3+ [23], resultando em melhorias na segmentação dos pedúnculos. Assim, este sistema permite identificar os bagos a serem removidos de um dado cacho, como também possibilita a extração de características fenotípicas relevantes [53].

A incorporação da tecnologia de realidade aumentada em sistemas que utilizam redes neurais, permitiu aos agricultores realizar o desbaste com instruções visuais em tempo real. Este sistema, baseado numa arquitetura híbrida que combina ResNet18 [54] e LSTM [55], mostrou-se eficaz mesmo com utilizadores sem experiência, melhorando a qualidade do desbaste em mais de 8% relativamente a especialistas [56].

Outra abordagem inovadora [57] propõe o uso de pseudo-rótulos gerados automaticamente para deteção e segmentação de uvas de mesa. Este método recorre à análise de movimento 3D entre *frames* e à aplicação posterior de *Mask* R-CNN [11] para gerar máscaras de segmentação com base nas caixas delimitadoras pseudo-rotuladas [58]. Além disso, a evolução do modelo *Mask* R-CNN [11], permitiu desenvolver um sistema que realiza a extração de características fenotípicas, que contribui para uma análise mais precisa.

Foi ainda proposto um modelo CNN orientado por perceção para deteção de cachos [59], que incorpora fundamentos como a teoria tricromática, a teoria das cores oponentes e a lei do contraste de Weber. A combinação de diferentes espaços de cor e componentes visuais permitiu alcançar precisões superiores a 96% para uvas azuis e 89% para uvas amarelas.

Adicionalmente, modelos de deteção *multibox* quantizados têm sido aplicados com sucesso na identificação de cachos em diferentes estágios de crescimento, demonstrando potencial para aplicação prática em condições reais [60].

Finalmente, ao considerar as características da copa da videira, modelos de *Deep Learning* têm mostrado elevada precisão na estimativa do número real de bagas até 60 dias antes da colheita, evidenciando o enorme potencial dessas técnicas para a agricultura de precisão [61].

Com base nas abordagens discutidas anteriormente, observa-se que existem diversas técnicas desenvolvidas para resolver desafios na viticultura, evidenciando uma base comum sobre a qual diferentes arquiteturas podem ser construídas.

No caso da segmentação de cachos de uva, que é uma das abordagens mais utilizadas pelos sistemas, existe uma proposta interessante [62] que utiliza a abordagem *DepthSeg* [63], incorporando um método de agrupamento baseado em profundidade e aplicando o modelo *Segment Anything Model* (SAM) [27] para melhorar a segmentação. Em comparação com o algoritmo original, a proposta descrita apresenta uma leve melhoria na contagem de objetos e um avanço considerável na segmentação a nível de pixel.

A segmentação mais precisa de um cacho de uva permite extrair informações mais detalhadas, como a contagem do número de bagos, o tamanho e a disposição dos bagos que influenciam diretamente a classificação de um cacho de uva. Desta forma, foi desenvolvido um sistema automatizado [64] que através da introdução de nova técnica de aumento de dados, adaptação do modelo de segmentação *Hybrid Task Cascade* (HTC) [65] tornando mais sensível à localização e a utilização de modelos de regressão, permitiu que este sistema conseguisse detetar e contabilizar bagos, realizar extração de características geométricas dos mesmos, em cachos de uva de mesa.

A luminância é fator que em certos casos compromete o desempenho dos modelos ao nível da segmentação, e por isso foi desenvolvida uma estratégia [66] que converte as imagens RGB para um espaço de cores CMY (*Cyan, Magenta and Yellow*) com o intuito de realçar o contraste entre os cachos e o fundo. De seguida, estas novas imagens são usadas para treinar um classificador *Support Vector Machine* (SVM) [67], gerando um mapa preliminar de segmentação. Este mapa é posteriormente refinado através da análise dos componentes da matriz, saturação e brilho no espaço de cores HSB (*Hue, Saturation and Brightness*), que vai ser utilizado por um segundo classificador SVM [67] com o objetivo de segmentar os pixels correspondentes aos cachos de forma mais precisa.

No que diz respeito à deteção de bagos individualmente, uma abordagem interessante [68] propõe o uso de algoritmos de busca de contornos e deteção de cantos, com o objetivo de identificar pontos côncavos com precisão, de forma a ser possível segmentar as bordas dos bagos e estimar tanto o número quanto o tamanho dos mesmos num dado cacho. Embora tenham sido alcançados bons resultados, foram identificadas limitações na deteção de bagos mais delgadas, bem como na velocidade do processo, que se mostrou relativamente lenta.

Além disso, a tarefa de deteção não se restringe apenas a cachos e bagos, podendo também abranger doenças que afetam a videira, como é o caso da Flavescência Dourada (FD). Esta doença de alto impacto na Europa, incentivou o desenvolvimento de um sistema para seu diagnóstico [69]. Em primeiro lugar, são identificadas e classificadas folhas não saudáveis em três categorias, sendo que após esta análise, estas deteções são processadas por uma rede que deteta brotos e cachos sintomáticos. Por fim, um classificador Random Forest [70] integra os sintomas para realizar o diagnóstico final. O estudo revelou que a análise isolada das folhas não é suficiente devido à semelhança dos sintomas, sendo crucial a combinação de sinais presentes em folhas, rebentos e cachos para um diagnóstico mais fiável.

Neste contexto, torna-se valioso o desenvolvimento de sistemas capazes de integrar diferentes funcionalidades num único processo, como é o caso dos mapas de probabilidade [71]. A partir do uso destes mapas é possível representar as relações espaciais entre os cachos e os bagos, permitindo a execução simultânea da deteção e contagem tanto dos cachos, contagem e bagos por cacho. Este sistema demonstrou um bom desempenho em relação à contagem de cachos, mas apresentou um erro médio de 14,2% na contagem de bagos por cacho.

2.5.4. Veículos Aéreos não Tripulados e Robots

A utilização de veículos aéreos não tripulados (UAVs) associada à inteligência artificial tem se destacado como uma abordagem inovadora na viticultura de precisão, pois esta combinação permite a recolha e análise de imagens em diferentes estágios fenológicos das uvas e videiras, sendo particularmente úteis em vinhedos situados em encostas e vales.

Os dados recolhidos pelos UAVs são aplicados em modelos de deteção e segmentação baseados no Índice de Vegetação por Diferença Normalizada (NDVI) e na composição nutricional das folhas, o que demonstra a relevância da automação na delimitação de zonas de gestão. Assim, esta tecnologia [72] possibilita a identificação precisa e rápida de videiras saudáveis e comprometidas, otimizando a monitorização e a tomada de decisão no campo.

Além disso, a combinação de UAVs com algoritmos de *Deep Learning* permite aos agricultores desenvolver soluções de recolha de imagens das videiras que posteriormente são processadas pelos algoritmos em plataformas web intuitivas para a gestão do vinhedo [73].

Contudo, a recolha de imagens por câmaras RGB introduzidas em UAVs, apresentam desafios devido à complexidade do ambiente do vinhedo, fatores como forte incidência solar, oclusão causada pela folhagem e a uniformidade dos cachos dificultam a segmentação e identificação precisa de objetos relevantes. Além disso, o tamanho elevado das imagens em relação aos objetos de interesse impõe limitações aos algoritmos de visão computacional, exigindo o desenvolvimento de técnicas otimizadas para um processamento eficiente [74].

Para superar as dificuldades mencionadas, torna-se essencial a criação de conjuntos de dados extensos baseados em imagens aéreas de baixa altitude. A deteção precoce de doenças na videira é um dos principais motivos para este avanço, pois os métodos tradicionais, baseados em imagens de satélite ou capturas em laboratório, frequentemente não conseguem fornecer informações detalhadas sobre vinhedos reais. Assim, é de realçar a importância da criação de um conjunto de dados com diferentes variedades de uvas afetadas por doenças anotadas. Este processo de anotação pode ser realizado a partir de uma ferramenta como LabelIMG [75], com o intuito de contribuir para um avanço significativo na automação da identificação de doenças na viticultura, bem como no controlo fitossanitário mais preciso e eficiente [76].

Para além dos veículos aéreos não tripulados (UAVs), os robots agrícolas têm-se mostrado uma alternativa eficaz para tarefas de controlo na viticultura, complementando ou até substituindo o uso dos UAVs em contagens automáticas de cachos de uva e na estimativa do seu volume. Estes dados são recolhidos através da combinação da segmentação semântica de imagens, do agrupamento baseado em profundidade e da reconstrução 3D para estimar o peso das uvas [77].

Além da estimativa de produtividade, os sistemas robóticos também desempenham um papel crucial na automação da colheita, pois a eficiência deste processo depende tanto da precisão do corte do cacho quanto da velocidade de execução. Desta forma, foi desenvolvido um método inovador [78] para determinar rapidamente pontos de colheita, que são calculados através da captura de imagens RGB e de profundidade com o intuito de criar uma região tridimensional de interesse (ROI), que foi projetada conforme a estrutura do robô. Este sistema robótico obteve uma taxa média de sucesso de reconhecimento de 97,1% e uma taxa média de sucesso de posicionamento de 93,5%. Além disso, a velocidade média de colheita atingiu 6,18 segundos por cacho, demonstrando a viabilidade desse sistema na automação da colheita.

Além disso, para tornar o método de corte ainda mais precisa e robusto, novas abordagens foram propostas, incluindo a localização do pedúnculo da uva em [79], que é fator essencial para determinar o ponto exato de corte do cacho. Esta metodologia integra imagens RGB com mapas de profundidade monocular, que permitem melhorar significativamente a precisão na segmentação de cachos e pedúnculos de uva em comparação com abordagens que utilizam apenas imagens RGB.

2.5.5. Decisões Tomadas

A revisão de literatura permitiu identificar diferentes abordagens para realizar a recolha dos dados sobre a produção, bem como nas tarefas de deteção, segmentação, contagem de cachos de uvas. No entanto, tornou-se necessário selecionar as técnicas mais adequadas e alinhadas aos objetivos definidos para o sistema proposto.

Um dos pontos centrais deste trabalho foi a definição de uma forma prática e eficiente de recolha de dados em campo. Embora a literatura apresente diferentes tipos de equipamentos para aquisição de imagens e vídeos, optou-se por uma solução que não dependesse de dispositivos de elevado custo ou de utilização complexa, de modo a facilitar a sua adoção em campo. Assim, o sistema foi concebido para permitir a captura de vídeos diretamente a partir de um *smartphone*, garantindo maior acessibilidade, portabilidade e simplificando a etapa inicial de recolha de dados.

O desenvolvimento do sistema proposto foi concebido de forma modular, sendo composto por três módulos principais (deteção, segmentação e contagem), que correspondem às tarefas necessárias para alcançar o objetivo definido. Esta abordagem modular proporciona uma separação clara das responsabilidades de cada componente, facilitando o desenvolvimento individual de cada tarefa, bem como a sua validação e otimização. Além disso, a estrutura modular permite uma análise mais rigorosa dos resultados de cada etapa e da integração entre módulos, garantindo que o sistema, como um todo, produza indicadores fiáveis e representativos da produção.

Neste contexto, no primeiro módulo, optou-se pela utilização do modelo de deteção o YOLO, que a partir dos estudos analisados, demonstraram que este modelo apresenta resultados interessantes mesmo em condições desafiadoras, como sobreposições e oclusões parciais e cenários complexos utilizando versões deste modelo antigas. No entanto, para que este desempenho seja atingido, é essencial o desenvolvimento de um conjunto de dados representativo e de elevada qualidade de anotação, garantindo uma capacidade de generalização para diferentes condições de campo.

Em seguida, no segundo módulo, que representa a etapa de segmentação dos cachos e dos bagos, foi definido a utilização do SAM, uma vez que este possui um extenso conjunto de dados, que é um dos maiores e mais diversos conjuntos de dados de segmentação de vídeo disponíveis. Assim o modelo SAM possui uma elevada capacidade de generalização para diferentes domínios e cenários, mesmo sem necessidade de treino adicional, logo a utilização deste modelo tornou-se particularmente atrativo para este trabalho, uma vez que reduz o esforço de preparação de dados anotados.

Após a etapa de segmentação, para inicializar um processo de classificação do cacho, foi definido que a abordagem mais adequada para recolher a cor predominante de um dado cacho seria através do processamento no espaço de cores HSV, que permite identificar a cor representativa sem que esta seja influenciada por fatores como a iluminação do ambiente onde foi capturada a imagem do cacho.

Além disso, outro parâmetro importante é o número de bagos presentes em cada cacho, e para estimar este número podia ser feita uma conversão da contagem obtida em 2D para uma estimativa em 3D, o que, em princípio, poderia ser realizado por meio de uma fórmula baseada em parâmetros físicos da estrutura do cacho. No entanto, constatou-se que alguns desses parâmetros não poderiam ser obtidos com os equipamentos disponíveis, pois exigiriam instrumentos de medição mais avançados para tornar a aplicação da abordagem fiável e robusta. Diante desta limitação, foi desenvolvida para este terceiro e último módulo do sistema, uma função polinomial ajustada a partir de dados empíricos, permitindo aproximar a relação entre a contagem bidimensional e a quantidade real de bagos.

A integração de todas estas metodologias num único sistema não apenas possibilita alcançar o objetivo proposto, como também representa um avanço para a área de estudo. Enquanto trabalhos anteriores utilizaram versões mais antigas das técnicas de deteção, segmentação e análise, este estudo inova ao explorar e avaliar o desempenho das versões mais recentes destes modelos. Além disso, a integração dos três módulos num só sistema constitui uma solução única, uma vez que, até ao momento, não existe uma abordagem que combine todas estas etapas de forma unificada. Assim, esta proposta representa um contributo original para a área em estudo.

CAPÍTULO 3

Funcionamento do Sistema Proposto

Nesta secção, pretende-se apresentar um enquadramento abrangente do sistema desenvolvido, detalhando os componentes utilizados, as técnicas aplicadas, os princípios de funcionamento e o modo de utilização previsto pelo algoritmo proposto nesta Dissertação.

Adicionalmente, são descritas e justificadas as principais decisões metodológicas e tecnológicas tomadas ao longo do processo de desenvolvimento, de modo a assegurar a implementação de um sistema robusto, fiável e alinhado com os objetivos previamente definidos para este trabalho.

3.1. Arquitetura Geral do Sistema

Para a implementação do sistema, foi necessário estudar de forma detalhada a documentação de cada uma das ferramentas selecionadas, tanto de hardware como de software, com o intuito de compreender qual seria a abordagem mais adequada de integrar cada um dos módulos do sistema.

A arquitetura geral do sistema proposto é composta por uma sequência de módulos interligados que asseguram o processamento dos dados, desde a sua aquisição até à geração dos resultados com base nas análises realizadas pelo sistema. Na Figura 7 apresenta-se um diagrama geral da arquitetura do sistema proposto composto pelos seguintes módulos:

- O primeiro módulo inicia-se através da aquisição de vídeos em campo, realizada com um smartphone e um estabilizador compatível, permitindo registar ambos os lados de cada carreira de plantas de forma prática e acessível. Em seguida, os vídeos são transferidos para o computador, onde são processados pelo sistema, dando início ao procedimento de deteção de cachos. Para esta tarefa, foi selecionado o modelo de deteção YOLO, integrado com o algoritmo *Bag-of-Tricks for* SORT (BoT-SORT), que permite fazer o seguimento dos cachos ao longo dos diferentes *frames* e a sua contagem de forma consistente. Após a obtenção do conjunto de deteções retornados pelo modelo, foram selecionados e armazenados os *frames* mais representativos de cada cacho identificado.
- A partir dos *frames* mais representativos de cada cacho, inicia-se o segundo módulo do sistema, responsável pela segmentação dos cachos detetados, utilizando o modelo SAM. O objetivo é isolar cada cacho, com o intuito de reduzir o ruído visual exterior, garantindo que apenas as regiões de interesse sejam consideradas nas análises subsequentes. A partir das máscaras segmentadas, é realizada a extração da cor predominante, recorrendo ao espaço de cores HSV, que permite uma representação mais robusta das tonalidades.

- De seguida, inicia-se o terceiro módulo, que é responsável pela segmentação dos bagos presentes nos cachos previamente segmentados. O modelo SAM é novamente utilizado com o objetivo de contabilizar o número de segmentos (bagos) presentes no cacho identificado na imagem 2D. Como por cada cacho foram selecionadas duas imagens (*frames*) consideradas mais representativas, é realizada a média do número de bagos segmentados em ambas as imagens. Ao ter esta estimativa de uma imagem bidimensional, esta é posteriormente ajustada utilizando um polinómio desenvolvido para converter os valores obtidos numa aproximação tridimensional, com o intuito de possuir uma estimativa próxima do real, relativamente ao número médio total de bagos por cacho.

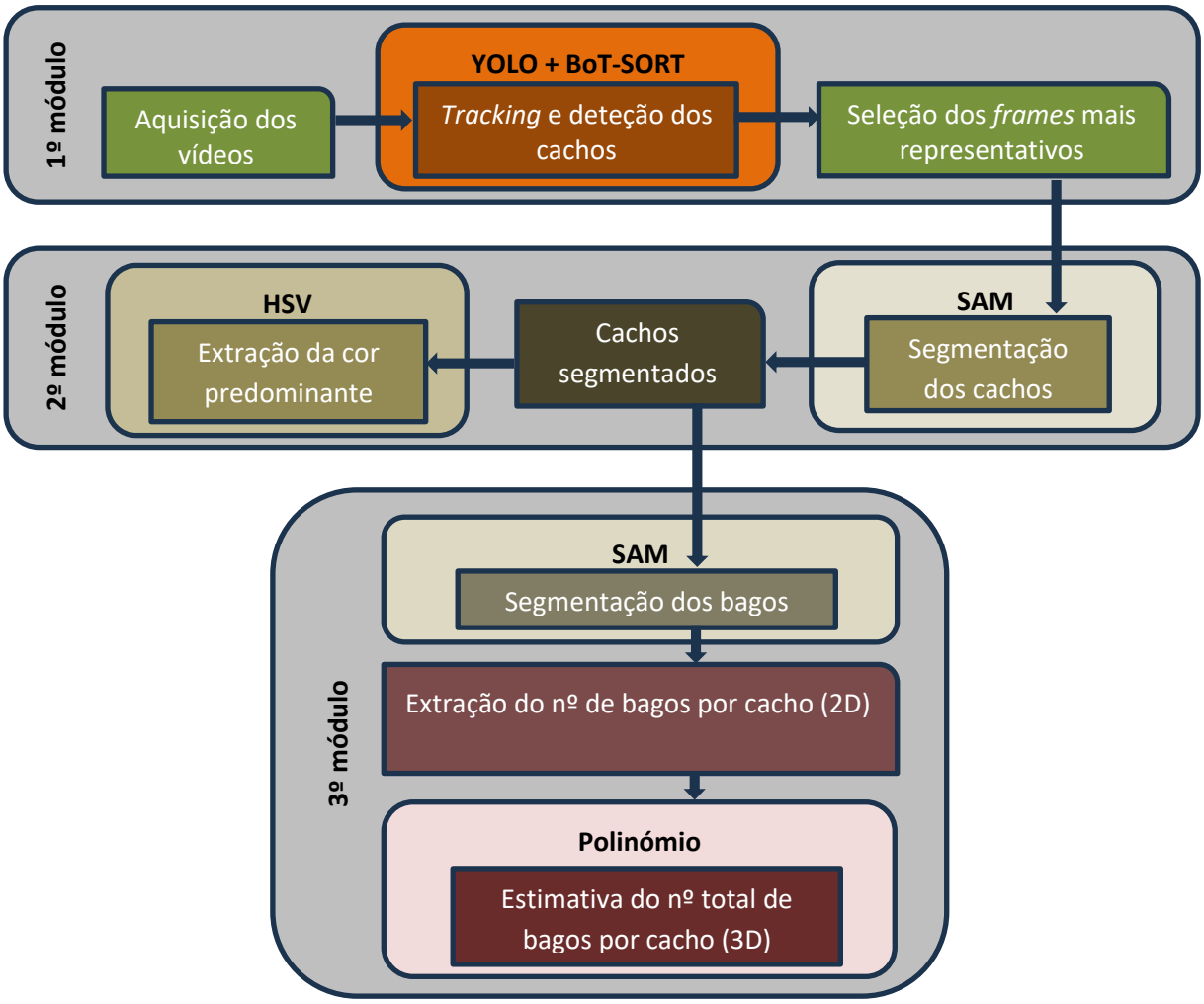

Figura 7. Representação dos módulos do sistema desenvolvido

3.2. Aquisição de Vídeo

O sistema desenvolvido foi concebido para processar vídeos de uma da área de produção, em que este devem ser capturados segundo um formato de gravação padronizado, garantindo consistência e qualidade dos dados recolhidos.

A definição de uma carreira corresponde a um conjunto de plantas alinhadas, como apresentado na Figura 8, que ocupam, em média, uma área de cerca de 300 m², com aproximadamente 100 metros de comprimento e 3 metros de largura, sendo que esta largura incorpora os dois lados de ramificação de cada uma das 33 videiras, em média, presentes na carreira.

Figura 8. Aquisição de dados num dos lados da carreira

No que diz respeito ao método de aquisição de dados de uma carreira, foram testadas diferentes abordagens de gravação, com o objetivo de otimizar tanto o tempo quanto os recursos necessários para a recolha dos dados. Como resultado, a estratégia adotada consistiu na gravação contínua de vídeo, de forma a contemplar ambos os lados da carreira num único percurso.

Assim, cada gravação referente a uma carreira possui uma duração média que varia entre 9 e 12 minutos, pois esta variação está diretamente relacionada não só com o comprimento, com a velocidade de deslocação do utilizador como também com as condições do terreno, durante a aquisição dos dados. Durante a gravação, como apresentado na Figura 9, a pessoa deve posicionar-se de forma a manter o ombro direito posicionado à esquerda da planta, garantindo que a filmagem seja realizada sempre no mesmo enquadramento.

Figura 9. Posicionamento e orientação do smartphone

Para além do correto posicionamento, é fundamental destacar que a orientação do *smartphone* deve ser ligeiramente inclinada para cima, de modo a permitir que os cachos sejam visualizados forma frontal, mas a partir de um ângulo que amplia o campo de visão, melhorando significativamente a perceção e a deteção em condições reais de campo, como ilustrado na Figura 8.

Esta perspetiva inclinada contribui também para maximizar a visibilidade dos cachos, reduzindo o impacto de oclusões, sobreposições entre cachos ou folhas garantindo uma representação mais representativa e fiável dos cachos recolhidos.

Além disso para assegurar a estabilidade e consistência das gravações, foi utilizado um estabilizador compatível com o *smartphone*, o *Zhiyun Smooth* Q3. Este componente auxiliar de gravação de vídeo, desempenha um papel essencial na padronização do formato de gravação, garantindo assim que os vídeos estejam de acordo com o formato de gravação, evitando desvios que possam comprometer a qualidade da análise.

Neste estudo, foi utilizado um *smartphone* iPhone 14, que apesar de possuir diferentes modos de resolução de vídeo, foi definido que todas as gravações adquiridas teriam uma resolução de 720p, mesmo existindo opções de maior qualidade, como 1080p ou 4K. A escolha desta resolução é suportada pelo facto deste tipo de resolução ser igual ou próxima da resolução utilizada pela maioria dos *smartphones* disponíveis no mercado, inclusive modelos mais antigos ou de gamas intermédias.

Desta forma, garante-se que o sistema desenvolvido seja acessível e replicável por qualquer utilizador, sem a necessidade de equipamentos de elevado custo ou especificações avançadas. Além disso, a resolução de 720p representa um equilíbrio entre qualidade de imagem suficiente para a realização das diferentes tarefas do sistema e tamanho dos ficheiros gerados, permitindo gravações mais longas, processamento mais rápido e menor consumo de espaço de armazenamento.

Os vídeos recolhidos na resolução mencionada, foram armazenados no formato MOV (*QuickTime File Format*) com um *codec* HEVC (High Efficiency Video Coding), que são parâmetros padrão de gravação de vídeos dos iPhones. Esta combinação garante elevada compressão sem perda percetível de qualidade, facilitando tanto no armazenamento como também na transferência dos ficheiros para processamento posterior.

Apesar de ter sido utilizado o formato MOV, o formato MP4 (MPEG-4 Parte 14), que geralmente é mais usado em dispositivos Android, suporta o mesmo tipo de *codec* apresentado e utilizado neste estudo. Desta forma, os vídeos podem ser gravados e processados de forma consistente em diferentes plataformas, garantindo compatibilidade entre dispositivos iOS e Android sem comprometer a qualidade ou a eficiência do processamento.

3.3. Preparação do Conjunto de Dados

A aplicação do sistema proposto em contexto real de campo representa um desafio, em diversas vertentes, pois este não é um ambiente controlado. Assim, foi necessária a preparação de um conjunto de dados específico, com o objetivo de adquirir vídeos que contemplem as várias situações que podem ocorrer no terreno. O conjunto deste trabalho foi composto por 3485 anotações, realizadas manualmente através da plataforma LabelIMG [75], a qual permite a marcação detalhada das regiões de interesse nas imagens.

A decisão de criar um conjunto de dados do zero deve-se a dois fatores principais. Em primeiro lugar, o ambiente de captura e as condições particulares de recolha das imagens dos cachos de uva diferem significativamente de cenários genéricos, tornando inadequado o recurso a bases de dados já existentes. Em segundo lugar, até ao momento em que foi terminada a revisão da literatura, não foi encontrado um conjunto de dados público disponível que esteja de acordo com os requisitos específicos deste estudo, nomeadamente na forma como é realizada a deteção dos cachos em contexto real de campo.

O conjunto de dados resultante é composto por imagens com características variadas, incluindo diferenças de iluminação, distâncias e condições de oclusão e de sobreposição entre vários cachos. Desta forma, esta diversidade confere ao conjunto uma maior representatividade, tornando-o uma fonte de conhecimento robusta para o treino do modelo.

Além disso é importante destacar que a anotação manual, apesar de exigir muito tempo e dedicação, constitui uma etapa fundamental para assegurar a qualidade do conjunto de dados, pois o desempenho do modelo também depende da qualidade das anotações. Uma anotação manual cuidada permite minimizar erros e ambiguidades, fornecendo uma base sólida e fidedigna para o treino e a avaliação dos algoritmos desenvolvidos.

3.4. Modelo de Deteção

A partir da análise dos estudos apresentados no estado da arte, é possível constatar que diferentes abordagens demonstraram a viabilidade da deteção de objetos a partir de conjuntos de dados previamente definidos.

Estes trabalhos foram conduzidos tanto em ambientes controlados, com condições estáveis de iluminação e captura, como em ambientes não controlados, sujeitos a uma maior variabilidade e complexidade. Entre as abordagens analisadas, destacou-se a utilização do modelo YOLO (You Only Look Once) [14], que se revelou relevante devido ao seu elevado desempenho em termos de precisão.

No entanto, verificou-se que grande parte dos estudos disponíveis utilizou versões antigas do YOLO [14], e nesse sentido, seria interessante a aplicação de versões mais recentes deste modelo de deteção. Estas novas versões disponíveis para serem utilizadas pela comunidade, apresentam melhorias significativas tanto na arquitetura bem como no desempenho do modelo.

Assim, para selecionar a versão mais adequada ao presente contexto de aplicação, foram analisados não só os indicadores de desempenho dos modelos (precisão, recall, mAP50 e mAP50-95), mas também os seus tempos de inferência, determinantes para garantir a eficiência do sistema desde o processamento dos dados recolhidos até ao retorno de informação relevante sobre a produção.

Neste trabalho, foram comparadas as versões mais recentes do Yolo, nomeadamente o YOLOv11 e o YOLOv12, onde ambas foram submetidas a um processo de *data augmentation*, com o objetivo de ampliar de forma sintética o conjunto de dados e, consequentemente, aumentar a sua diversidade e robustez. Este procedimento é essencial para melhorar a capacidade de generalização do modelo, uma vez que permite expô-lo a variações adicionais que simulam condições reais, tais como mudanças de dimensões dos objetos, ângulo de visão e oclusões parciais dos objetos.

3.4.1. *Data* Augmentation

A estratégia de *data augmentation*, pode ser realizada através da aplicação de diferentes parametrizações [80], em que cada uma delas permite realizar um determinado tipo de operação sobre os dados. No contexto desta Dissertação, privilegiaram-se aquelas que melhor se adequavam ao ambiente e ao tipo de dados utilizados neste estudo, assegurando que as transformações aplicadas fossem realistas e pertinentes ao domínio em análise.

Desta forma, foram definidos e aplicados os seguintes tipos de *data augmentation*:

- *translate* - deslocamento da imagem e das *bounding boxes* em eixos X/Y.
- *scale* - alteração da escala da imagem (zoom in/out) com ajuste das *bounding boxes*.
- *fliplr* - espelhamento horizontal da imagem e das *bounding boxes*.
- *mosaic* - combinação de 4 imagens numa única amostra, com variação de escala e posição.
- *copy_paste* - recorte e inserção de objetos de uma imagem em outra.
- *mixup* - interpolação linear entre duas imagens e *labels* correspondentes.

Em primeiro lugar, foram conduzidos testes individuais para cada um dos tipos referidos, variando as suas parametrizações de forma a identificar aqueles que mais contribuíam para o aumento do desempenho do modelo de deteção. Concluída esta etapa, procedeu-se à análise de combinações, aplicando dois ou mais tipos de *data augmentation* em simultâneo, com o objetivo de avaliar se a sua interação poderia gerar resultados superiores aos obtidos de forma isolada. Após uma série de experiências, constatou-se que a combinação que apresentou a melhoria mais expressiva correspondeu à aplicação conjunta de *scale* e *translate*.

De seguida, procedeu-se à análise do desempenho de cada modelo sobre o conjunto de validação, considerando a estratégia de *data augmentation* definida neste estudo, com o intuito de comparar de forma sistemática os resultados obtidos pelas diferentes versões do YOLO, apresentados na Tabela 2.

Tabela 2. Resultados dos modelos de deteção

Modelo	***Precision***	***Recall***	**mAP50**	**mAP50-95**
YOLOv11n	0.867	0.855	0.942	0.752
YOLOv11s	0.816	0.886	0.937	0.763
YOLOv12n	0.824	0.882	0.922	0.747
YOLOv12s	0.844	0.88	0.941	0.78
YOLOv11n + (*data augmentation)*	0.828	0.912	0.93	0.755
YOLOv11s + (*data augmentation)*	0.856	0.886	0.94	0.775
YOLOv12n + (*data augmentation)*	0.853	0.867	0.934	0.747
YOLOv12s + (*data augmentation)*	**0.939**	**0.931**	**0.981**	**0.82**

As métricas utilizadas para avaliar o desempenho dos modelos foram a *Precision*, *Recall*, mAP50 e mAP50-95. A *Precision* e o *Recall* medem, respetivamente, a proporção de deteções corretas e a capacidade do modelo de identificar todos os cachos presentes. As métricas mAP50 e mAP50-95 representam a média da precisão em diferentes limiares de *Intersection over Union* (IoU), sendo que a segunda é mais exigente e reflete melhor o desempenho global do modelo em termos de localização e classificação.

A partir desta análise, observa-se que a versão YOLOv12 [14] demonstrou um desempenho superior em relação à versão anterior, evidenciando uma evolução consistente entre gerações do modelo. Entre as diferentes versões testadas, destaca-se a YOLOv12s, que alcançou os melhores resultados globais, confirmando-se como a configuração mais eficiente no contexto deste trabalho.

3.4.2. Limiar de Confiança

Os resultados apresentados na Tabela 2 não consideram a influência do limiar de confiança (*conf*) aplicado durante a validação. Este parâmetro desempenha um papel fundamental, uma vez que define a probabilidade mínima para que uma predição seja considerada válida, afetando diretamente o equilíbrio entre as falsas deteções e os objetos não detetados.

Assim, ainda que as métricas apresentem uma visão geral do desempenho, tornou-se essencial complementar a análise com uma avaliação qualitativa. Para tal, foram realizadas inspeções visuais sobre as deteções produzidas pelo modelo desenvolvido, como por exemplo as que se encontram ilustradas na Figura 10.

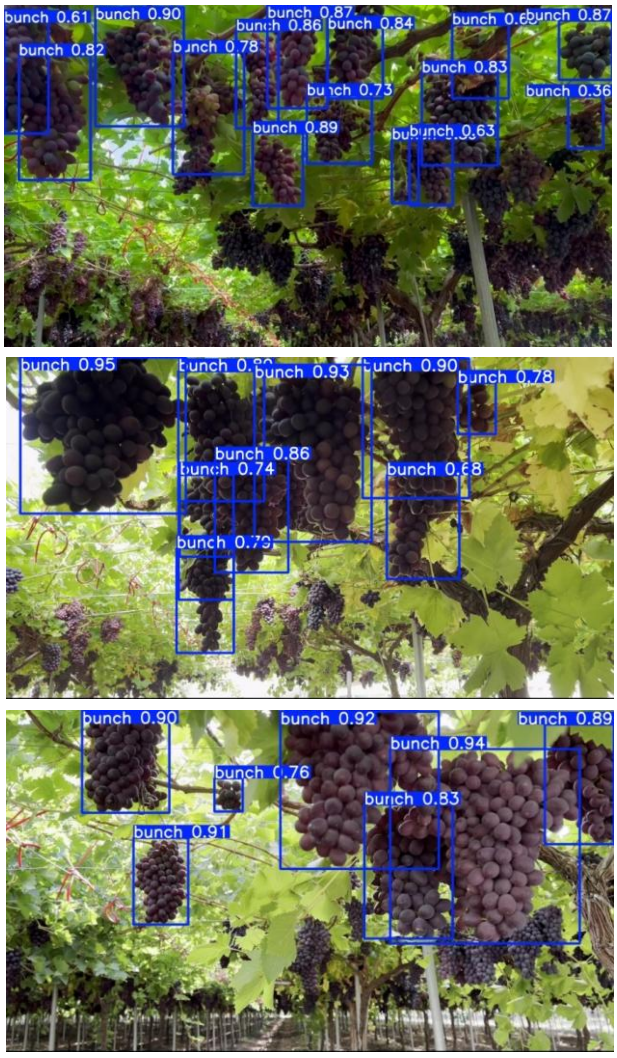

Figura 10. Exemplos de cachos detetados pelo modelo desenvolvido

A partir da análise dos diferentes cenários apresentados anteriormente observou-se que, de uma forma geral, quanto maior a aproximação do cacho à câmara, maior é o limiar de confiança obtido associado à deteção, uma vez que os detalhes visuais se tornam mais nítidos e favorecem a identificação correta do cacho.

No entanto, em determinadas situações do vídeo, verificou-se que alguns cachos, devido ao seu posicionamento, à sua orientação ou ao ângulo de captura da imagem, apresentam limiares de confiança na ordem dos 0.6 de confiança. Embora estes cachos possuam uma menor confiabilidade associada, representam elementos reais da produção e, portanto, devem ser incluídos no processo de análise. Assim, para garantir que este valor do limiar de confiança não comprometesse o desempenho do sistema desenvolvido, foi realizado um estudo para avaliar o impacto de diferentes valores de limiar de confiança no desempenho do modelo, como apresentado na Tabela 3.

Tabela 3. Resultados do modelo para diferentes limiares de confiança

Modelo	*conf*	*Precision*	*Recall*	mAP50	mAP50-95
YOLOv12s + (*data augmentation)*	**0.6**	**0.941**	**0.936**	**0.975**	**0.842**
YOLOv12s + (*data augmentation)*	0.7	0.957	0.92	0.953	0.829
YOLOv12s + (*data augmentation)*	0.8	0.992	0.714	0.854	0.759

A partir dos resultados obtidos, é possível perceber que o valor de 0.6 proporciona o melhor equilíbrio entre *precision* e *recall*, resultando num valor de mAP50 de 0.975, o valor mais elevado entre os limiares testados. Embora limiares de confiança mais altos apresentem um ligeiro aumento da *precision*, estes conduzem a uma redução significativa do *recall*, o que implica que um número maior de cachos reais deixa de ser identificado. Assim, a escolha do limiar de confiança igual a 0.6 revela-se a mais adequada, assegurando uma deteção de cachos de forma mais consistente.

3.4.3. Seguimento e Contagem dos Cachos

Ao serem definidas todas as parametrizações do modelo de deteção desenvolvido, foi igualmente definida uma estratégia de *tracking* a ser utilizada pelo mesmo no contexto em estudo. Esta estratégia consiste na integração de um método de *multi-object tracking* (MOT), com o intuito de assegurar a identificação consistente e o seguimento dos cachos ao longo dos diferentes *frames* do vídeo.

Desta forma, foi selecionado o algoritmo *Bag-of-Tricks for SORT* (BoT-SORT) [81], reconhecido pela sua robustez e elevado desempenho em cenários de seguimento complexos. Este método combina três componentes fundamentais que atuam de forma complementar.

O primeiro componente é a predição de movimento, que é responsável por estimar a posição futura de cada cacho nos *frames* seguintes, assegurando a continuidade do rastreamento mesmo em situações de falhas temporárias de deteção. O segundo componente é a associação espacial baseada em *Intersection over Union*, que estabelece correspondências entre as novas deteções e as anteriores com base na sua sobreposição espacial. Por fim, o algoritmo recorre a descritores visuais (Re-ID *embeddings*), que permitem distinguir objetos visualmente semelhantes, evitando trocas de identidade em situações de oclusão parcial ou proximidade entre objetos.

A integração dos componentes do algoritmo selecionado permite manter uma identificação consistente dos cachos ao longo do tempo, bem como preservar os dados das deteções retornadas pelo modelo. No entanto, para este projeto decidiu-se que, entre todas as deteções de cada cacho, apenas seriam utilizados os *frames* mais representativos, de forma a garantir maior fiabilidade na análise subsequente dos mesmos e na otimização do sistema no processo de monitorização.

Deste modo, o processo de identificação dos *frames* mais representativos de cada um dos cachos é estruturado por diferentes etapas. Na primeira etapa deste processo, é realizado uma verificação se cada cacho detetado é identificado num número mínimo de *frames*, assegurando que a sua deteção é consistente ao longo de um período do vídeo. Em seguida, analisam-se os intervalos de tempo durante os quais o cacho foi identificado, de modo a garantir que as representações selecionadas pertencem a um período contínuo mínimo. De seguida, entre os períodos contínuos válidos, é selecionado o intervalo que possui um valor de limiar médio de confiança mais elevado com a intenção de, dentro do mesmo, serem escolhidos os *frames* mais representativos para utilização nas etapas subsequentes de segmentação e contagem, assegurando a fiabilidade e consistência dos dados analisados.

3.5. Modelo de Segmentação

Após a deteção dos cachos, e em conformidade com o objetivo do sistema proposto, torna-se necessário extrair informações detalhadas acerca de cada cacho identificado. Desta forma, é indispensável aplicar uma etapa adicional de segmentação sobre os cachos detetados.

Enquanto a deteção permite localizar a posição geral do objeto na imagem, a segmentação proporciona uma representação mais rigorosa, delimitando os contornos exatos do cacho, facilitando posteriormente a análise individual dos bagos que o constituem.

Assim, esta abordagem possibilita não apenas a contagem mais precisa dos bagos, mas também a obtenção da cor predominante, fatores que são de grande importância em contextos de monitorização da produção e apoio à tomada de decisão na viticultura.

3.5.1. Seleção do Modelo de Segmentação

De acordo com os estudos analisados no estado da arte, a maioria das abordagens recorreu a modelos de segmentação que necessitam de ser treinados com conjuntos de dados desenvolvidos, como é o caso dos modelos Mask R-CNN [11] ou DeepLab [23]. Embora estes modelos apresentem bons resultados, a sua aplicação exige a criação ou adaptação de conjuntos de dados dedicados, processo que é altamente trabalhoso e demoroso.

Neste contexto, o modelo SAM [27], desenvolvido pela Meta AI, destacou-se por apresentar resultados muito promissores. A principal vantagem do SAM [27] reside no facto de não necessitar de treino adicional, uma vez que foi previamente treinado sobre um dos maiores conjuntos de dados de segmentação já compilados, o que lhe confere uma capacidade de generalização muito superior, permitindo que seja aplicado em diferentes domínios e cenários sem necessidade de ajustamentos extensivos.

Adicionalmente, a criação de um conjunto de dados específico para esta etapa implicaria um investimento significativo de tempo e recursos, o que poderia comprometer o avanço global do desenvolvimento do sistema. Assim, a adoção do SAM [27] mostrou-se uma solução prática e eficiente, conciliando desempenho e viabilidade.

Além disso, foi disponibilizada a versão SAM2.1, que introduz melhorias substanciais face à versão anterior, que se traduzem não apenas em maior precisão de segmentação, mas também em redução significativa do tempo de inferência, fator crítico quando se trata de uma tarefa computacionalmente intensiva como a segmentação de objetos. Desta forma, a utilização do SAM2.1 representa uma escolha estratégica e uma aposta para o sistema proposto, assegurando um compromisso entre robustez, eficiência e aplicabilidade prática no domínio em estudo.

Com o objetivo de avaliar se a versão do SAM2.1, incorporada com os novos checkpoints disponibilizados, apresenta melhorias em relação às versões anteriores, foram conduzidos diversos testes focados na segmentação integral dos cachos de uva detetados.

Esta avaliação contemplou dois critérios principais; o primeiro é a qualidade das máscaras geradas, analisada em termos de precisão e consistência dos contornos segmentados; e segundo e não menos importante que o anterior, o tempo de inferência necessário para a execução da segmentação, que é uma métrica particularmente relevante dada a elevada complexidade computacional desta tarefa computacional.

3.5.2. Qualidade de Segmentação do Cacho

Conforme discutido anteriormente, o tempo de inferência constitui um fator determinante no desempenho de sistemas de segmentação, sobretudo em aplicações que requerem eficiência operacional. Nesse contexto, tornou-se imprescindível analisar qual a versão do modelo de segmentação que seria mais apropriada para a tarefa proposta.

Com esse propósito, foi conduzido um estudo comparativo envolvendo as diferentes versões do SAM [27], no qual foram avaliados características como número de parâmetros, capacidade de generalização, qualidade da segmentação e velocidade de processamento.

A definição do conjunto de modelos analisados, apresentados na Tabela 4, teve como base de seleção a procura de um equilíbrio entre complexidade arquitetural e eficiência computacional, de forma a selecionar aquelas versões mais promissoras para a tarefa de segmentação de cachos de uva.

A partir dos testes realizados, tornou-se possível comparar não apenas o tempo de inferência necessário para a execução da tarefa de segmentação, mas também, e de forma igualmente relevante, a qualidade das máscaras resultantes. Esta análise da qualidade foi conduzida a partir de dois critérios principais: a avaliação visual direta das máscaras geradas e o nível de confiança atribuído a cada um dos cachos de uva segmentados, indicador este disponibilizado pelo modelo.

Com o objetivo de evidenciar o desempenho típico das diferentes versões do modelo em estudo, este documento apresenta uma imagem de teste contendo dois cachos de uva com características distintas. A seleção desta imagem foi cuidadosamente realizada, de modo a contemplar cenários que representam desafios recorrentes no ambiente real, conforme é apresentado na Figura 11.

Figura 11. Imagem de teste

O primeiro cacho encontra-se parcialmente ocluso, devido à presença de um ramo que atravessa a região entre diversos bagos, o que introduz uma maior complexidade no processo de segmentação do cacho. O segundo cacho, que se encontra à direita do cacho descrito anteriormente, caracteriza-se por apresentar um elevado grau de sobreposição entre os bagos, condição que corresponde ao padrão mais frequentemente observado durante a aquisição das imagens no campo.

Neste cenário, a versão SAM_base demonstrou um tempo de inferência inferior em comparação com as restantes versões, contudo, esta vantagem em eficiência temporal não se refletiu na qualidade das segmentações obtidas, uma vez que as máscaras resultantes apresentaram diversos erros, incluindo elevado nível de ruído e inconsistências na representação dos cachos de uva.

Desta forma, foi testada a versão SAM_large, que possui uma arquitetura mais robusta e um número significativamente maior de parâmetros em comparação com a versão SAM_base. Apesar desta maior capacidade de representação, o modelo não conseguiu eliminar completamente os problemas observados na versão SAM_base, apresentando ainda ruído, falhas na delimitação precisa dos bagos e um tempo de inferência mais elevado.

Em seguida, foram avaliados os modelos SAM2.1_tiny e SAM2.1_small. Ambos apresentaram tempos de inferência semelhantes, demonstrando eficiência computacional comparável. Em termos de desempenho de segmentação, os modelos mostraram avanços significativos em relação às versões anteriores, evidenciando maior precisão e robustez na delimitação das instâncias. No entanto, ainda foram observados níveis problemáticos de ruído, que comprometiam a segmentação dos cachos, especialmente em regiões com sobreposição de bagos ou contraste reduzido.

Tabela 4. Tempo de inferência de cada modelo

Modelo	Tempo de inferência
SAM_base	0.22s
SAM_large	0.49s
SAM2.1_tiny	0.64s
SAM2.1_small	0.70s
SAM2.1_b+	1.09s
SAM2.1_large	2.85s

No entanto, a aplicação das versões SAM2.1_b+ e SAM2.1_large demonstraram um desempenho significativamente superior em relação às versões analisadas anteriormente, destacando-se na capacidade de remover com precisão o ruído presente ao redor dos cachos segmentados, que comprometia a qualidade da segmentação.

A versão SAM2.1_large forneceu uma segmentação ligeiramente mais precisa, refletindo seu maior poder de representação, como é possível de visualizar a partir da Figura 12. No entanto, esta precisão adicional vem acompanhada de um aumento considerável no tempo de inferência, quase três vezes maior que o observado para o SAM2.1_b+.

Figura 12. Segmentação realizado pela versão SAM2.1_large

Por outro lado, o SAM2.1_b+ destacou-se como o modelo que apresentou o melhor equilíbrio entre rapidez de execução e qualidade das segmentações em todos os testes realizados. De forma geral, as máscaras geradas foram bastante fiéis à estrutura real do cacho, apresentando contornos bem definidos e com imperfeições mínimas ou praticamente inexistentes. A análise da Figura 13 evidencia que este modelo mantém um desempenho consistente mesmo em cenários de maior complexidade de segmentação, conseguindo lidar com sobreposições e variações de iluminação. Embora no segundo cacho se observe um leve ruído, este não compromete de forma significativa a qualidade global da segmentação.

Figura 13. Segmentação realizada pela versão SAM2.1_b+

Quando se comparam os resultados apresentados nas Figuras 12 e 13, que representam os modelos SAM2.1_large e SAM2.1_b+, respetivamente, é percetível que os resultados são bastante semelhantes, sendo as diferenças observadas marginais e pouco relevantes para justificar a adoção da versão SAM2.1_large.

Além disso, deve-se considerar que o modelo SAM2.1_large exige maior capacidade computacional, tornando o SAM2.1_b+ a opção mais vantajosa por oferecer um desempenho praticamente equivalente, aliado a uma menor necessidade de recursos.

3.5.3. Segmentação dos Cachos

Desta forma, após a seleção do modelo de segmentação e a deteção do cacho pelo modelo de deteção desenvolvido, procedeu-se à aplicação da primeira segmentação realizada pelo sistema, que é direcionada sobre a região correspondente ao cacho detetado.

O objetivo principal desta operação consiste em reduzir o ruído visual presente na imagem, eliminando elementos irrelevantes no enquadramento, como folhas, ramos ou fundo do ambiente de gravação, de modo a delimitar com precisão a região de interesse (*Region Of Interest* - ROI).

Esta etapa assume particular relevância no contexto do projeto, uma vez que garante que o processamento subsequente incide unicamente sobre a área de análise pertinente, assegurando a extração de informação mais fidedigna relativamente ao cacho detetado, como é possível de visualizar no exemplo da Figura 14 apresentada abaixo.

Figura 14. Segmentação do cacho detetado

Do ponto de vista técnico, a definição da ROI apresenta diversas vantagens, pois esta região contribui para a otimização dos recursos computacionais, reduzindo a quantidade de informação a processar e, consequentemente, o tempo de inferência, além de diminuir os requisitos de memória gráfica. Em segundo lugar, permite aumentar a robustez da análise, na medida em que minimiza a probabilidade de ocorrência de falsos positivos ou de interferências resultantes da sobreposição de elementos externos, como folhas parcialmente oclusivas.

Em terceiro lugar, possibilita a melhoria da qualidade dos dados extraídos, uma vez que a segmentação focada, promove uma representação mais rigorosa das características intrínsecas do cacho, como a cor predominante, a distribuição espacial dos bagos e a estimativa do seu número total.

Assim, a implementação desta abordagem conduz a uma representação mais limpa, precisa e consistente do cacho de uva, permitindo a recolha de informação relevante de forma sistemática e confiável e eficiente.

Esta estratégia reforça não apenas a eficácia do sistema desenvolvido, mas também a sua capacidade de generalização e aplicabilidade prática no domínio da viticultura de precisão, constituindo um passo metodológico fundamental para a integração de técnicas de visão computacional em processos de monitorização e avaliação da produção.

3.5.4. Extração da Cor Predominante

Uma vez realizada a segmentação do cacho, um dos parâmetros extraídos pelo sistema corresponde à identificação da cor predominante de cada cacho detetado pelo modelo. A obtenção desta informação relevante baseia-se exclusivamente na região de interesse (ROI) previamente delimitada, assegurando que a análise cromática incide unicamente sobre os pixéis que representam o cacho, eliminando assim potenciais interferências visuais provenientes de folhas, ramos ou do fundo da imagem.

No entanto, a cor definida pelo smartphone, encontra-se fortemente condicionada pelas condições de iluminação no momento da gravação. Fatores como excesso de luminosidade, presença de sombras projetadas ou variações na intensidade e direção da luz ao longo da carreira podem alterar significativamente a perceção cromática, reduzindo a fiabilidade dos resultados, como é apresentado na primeira imagem presente na Figura 15.

Figura 15. Comparação entre a cor do cacho na imagem original e a cor predominante calculada

Assim, foi implementada uma transformação do espaço de cor RGB para o espaço HSV (*Hue*, *Saturation*, *Value*), que é uma conversão amplamente utilizada em visão computacional devido à sua capacidade de separar a informação cromática da intensidade luminosa. O componente *Hue* (matriz) permite descrever de forma mais robusta a tonalidade da cor, relativamente independente de variações de iluminação, enquanto os componentes *Saturation* e *Value* fornecem informação sobre a intensidade e brilho.

Assim, esta abordagem possibilita uma extração mais consistente e realista da cor predominante do cacho, aproximando-se da sua tonalidade efetiva observada em campo, como apresentado pela na segunda imagem da Figura. Para além de garantir maior fiabilidade na análise, a informação cromática extraída possui elevada relevância, pois permite avaliar o estado de maturação das uvas, estando diretamente relacionada com parâmetros de qualidade, como teor de açúcares, acidez e compostos fenólicos.

A monitorização automática desta característica permite, portanto, identificar variações no estado de maturação entre diferentes cachos, contribuindo para a tomada de decisão tanto no processo de desenvolvimento do cacho bem como na colheita.

3.5.5. Segmentação dos Bagos

Após a segmentação dos cachos, tornou-se necessário estudar as melhores estratégias para extrair informação relevante a partir das máscaras obtidas, de modo a alcançar os objetivos do sistema. Duas informações são consideradas como prioritárias para os produtores:

- O estado de maturação, que pode ser inferido a partir da coloração predominante dos cachos segmentados;
- O número de bagos por cacho, informação de interesse direto para a estimativa da produção.

No caso da cor, a segmentação permitiu isolar a região correspondente ao cacho, possibilitando a análise da sua coloração média e predominante sem interferência do fundo da imagem. No entanto, a contagem dos bagos revelou-se uma tarefa mais complexa, exigindo o estudo e teste de diferentes abordagens analisadas e compatíveis com a metodologia do sistema proposto.

Uma das abordagens testadas neste estudo, apresentada no artigo [68], consistiu na aplicação de uma filtragem bilateral para suavizar a imagem segmentada do cacho, preservando as bordas, que posteriormente foi convertida para tons de cinzento e submetida ao algoritmo *Canny* [82], com o objetivo de identificar os contornos dos bagos. Para lidar com contornos incompletos, foram aplicados algoritmos de deteção de concavidades, de forma a reconstruir regiões ausentes. A partir dos contornos, procurou-se extrair os centroides correspondentes a cada bago. Contudo, esta abordagem testada revelou várias limitações em situações com elevada oclusão ou forte sobreposição de bagos, pois os contornos extraídos não correspondiam fielmente à realidade, resultando em contagens incorretas.

Desta forma, foi possível comprovar que esta metodologia de deteção dos bagos apresenta limitações relativamente a condições de iluminação, distância da câmara e qualidade de resolução variáveis como é apresentado na Figura 16.

Figura 16. Cenário que demonstra o ambiente complexo em estudo

Esta dependência tornava-os pouco robustos em ambientes não controlados, comuns em cenários agrícolas. Por estas razões, verificou-se que tais métodos de processamento clássico, embora úteis em casos simples, não eram adequados para lidar com a complexidade do problema em estudo.

Outra estratégia considerada foi o recurso a redes neuronais convolucionais (CNNs), treinadas especificamente para identificar e segmentar bagos de uva. Apesar de o potencial desta abordagem ser reconhecido na literatura, a sua aplicação no presente estudo mostrou-se inviável, uma vez que exigiria a criação de um conjunto de dados extenso e anotado ao nível dos bagos.

Como alternativa, foi explorada uma solução inovadora, que foi a aplicação do SAM2.1 também à segmentação dos bagos, após a segmentação prévia do cacho. No entanto, para isso utilizou-se o módulo *SAM2AutomaticMaskGenerator* desenvolvido para este modelo de segmentação, com o checkpoint do SAM2.1_b+, que permite gerar segmentações automáticas sem necessidade de *prompts* adicionais, ajustando-se automaticamente às regiões de interesse.

Nos testes realizados, verificou-se que a aplicação deste módulo, sem qualquer tipo de alteração dos seus parâmetros, demostrou que os resultados da sua aplicação no cenário em estudo, não eram totalmente satisfatórios, sobretudo em casos com elevada densidade e oclusão. Estes fatores dificultavam a segmentação de maior parte dos bagos presentes num cacho.

Desta forma, foi necessário perceber o motivo deste desempenho do módulo perante diversos casos. Após diferentes experimentações e ajustes, foi possível perceber que ao aumentar o número de pontos em cada uma das partes constituintes da imagem, permitiu refinar a técnica de forma a obter segmentações suficientemente precisas dos bagos, permitindo a estimativa do seu número com um nível de confiança adequado.

Assim, a combinação da segmentação inicial do cacho com o checkpoint do modelo SAM2.1_b+, e a aplicação subsequente do *SAM2AutomaticMaskGenerator* com a utilização do mesmo *checkpoint* para segmentação de bagos, apresentado na Figura 17.

Figura 17. Segmentação de bagos utilizando o modelo SAM2.1b+

Apesar do método proposto apresentar uma fiabilidade interessante na deteção e segmentação dos bagos, verificam-se limitações em condições menos favoráveis, como em situações de iluminação inadequada ou de sobreposição de bagos. Nestas circunstâncias, o modelo de segmentação pode não conseguir identificar todos os elementos presentes, resultando numa subestimação na contagem total de bagos. Este efeito pode ser observado na comparação entre as duas imagens apresentadas na Figura 18, em que a primeira representa o resultado da segmentação do modelo e a segunda o resultado esperado.

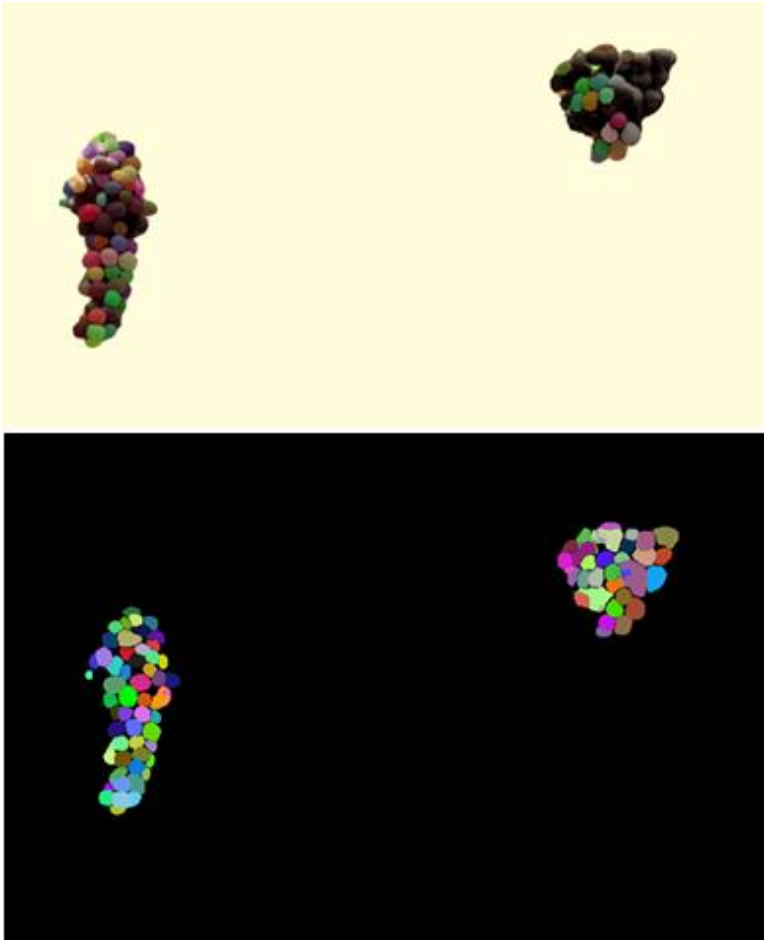

Figura 18. Comparação entre o resultado produzido pelo modelo e o esperado

Com o intuito de compreender de forma mais detalhada o impacto das perdas de bagos não segmentados em diferentes imagens, foi realizado um estudo específico num conjunto de 50 casos, em que a sua maioria são considerados complexos em termos de segmentação dos bagos. As imagens que constituem o conjunto utilizado nesta avaliação, apresentam diferentes condições adversas, como variações de iluminação, sobreposição de bagos e oclusões parciais, que representam alguns dos maiores desafios para os algoritmos de visão computacional. A avaliação teve como base a metodologia apresentada na Figura 19, que é realizada uma comparação entre o número de bagos segmentados pelo sistema e o valor real apresentado.

Os resultados obtidos evidenciam que, sob estas condições complexas, o modelo apresentou dificuldades significativas em aproximar-se dos valores reais esperados. Com base nos resultados obtidos, o sistema em média foi capaz de contabilizar apenas 15 bagos por cacho, enquanto o número real era de aproximadamente 62 bagos, o que revela uma discrepância considerável entre a predição e a realidade em condições adversas e instáveis de segmentação de bagos.

3.6. Estimativa da Quantidade de Bagos num Cacho

Após a contagem inicial dos bagos presentes na imagem segmentada, é importante salientar que este valor representa apenas uma estimativa parcial, uma vez que a captura do vídeo é realizada apenas de um lado do cacho. Consequentemente, o número obtido não corresponde ao total real de bagos, mas sim uma aproximação do número de bagos visíveis.

Além disso, o sistema proposto não foi concebido para realizar a filmagem da mesma carreira em sentidos opostos, o que impossibilitaria a reconstrução tridimensional do cacho. Esta limitação deve-se, sobretudo, ao facto de não ser possível determinar com precisão a localização no espaço 3D do cacho durante a captura do vídeo.

A partir desta limitação, tornou-se necessário investigar abordagens que permitissem estimar o número total de bagos a partir dos dados de uma imagem 2D. O artigo [64] propôs um método que considerava múltiplos fatores, como o número de bagos visíveis, diâmetro, circularidade, densidade e distribuição homogénea dos bagos no cacho. Embora esta abordagem fosse bastante robusta, a sua aplicação prática revelou-se inviável no presente trabalho, dado que a determinação do diâmetro real dos bagos exigiria informações adicionais sobre a distância entre a câmara e o cacho.

Deste modo, optou-se por uma estratégia alternativa baseada em modelagem polinomial, uma vez que esta abordagem permite criar um polinómio a partir dos dados recolhidos de diferentes cachos, com o objetivo de estimar o número real de bagos presentes em cada um deles. Assim, foi constituído um conjunto de dados de referência, composto por 50 cachos de uva *Midnight Beauty* que foram cuidadosamente selecionados e por sua vez contabilizados os bagos presentes nos mesmo, manualmente. A seleção dos cachos foi realizada de forma criteriosa, procurando garantir diversidade nas características morfológicas, como o tamanho, a forma e a densidade dos cachos. Além disso, foram incluídos cachos de diferentes estágios de maturação, com o intuito de tornar o polinómio resultante mais robusto e representativo da variabilidade encontrada em condições reais de campo.

Foram testadas diferentes possibilidades de regressão polinomial, variando o grau dos polinómios entre grau 1 (regressão linear) e grau 5. A regressão polinomial de grau 3, apresentada na figura 19, revelou ser a possibilidade que conduziu aos melhores resultados na estimativa do número real de bagos com base no número de bagos detetados nas imagens 2D. A derivada do polinómio ajustado, calculada para valores entre 20 e 100, apresenta variações entre aproximadamente 1.43 e 1.9, refletindo a taxa de crescimento do número de bagos detetados em função do número total de bagos reais.

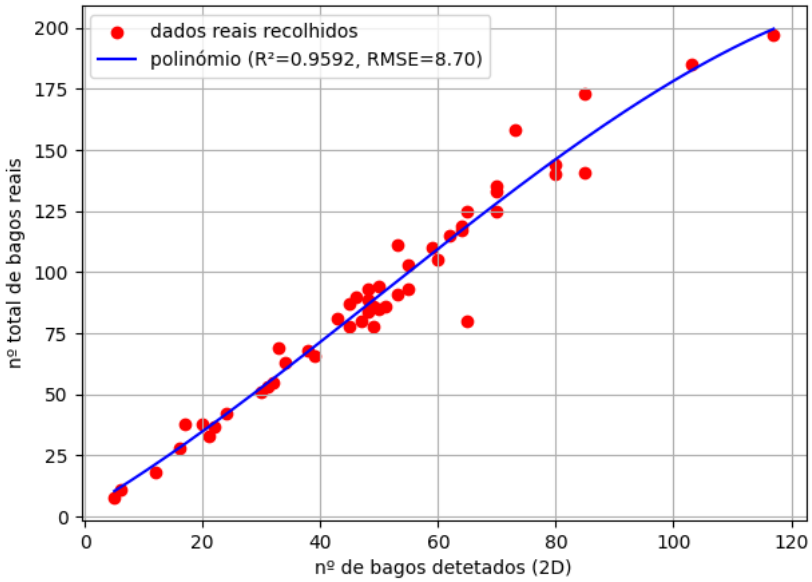

Figura 19. Gráfico do polinómio desenvolvido

O polinómio ajustado apresenta uma elevada proximidade com os valores reais observados, evidenciando a adequação do polinómio gerado para este estudo. Esta adequação é reforçada pelo indicador de coeficiente de determinação (R^2), que apresenta um valor aproximado de 0.96, que significa que o polinómio explica cerca de 96% da variabilidade dos dados reais. Além disso, o *Root Mean Square Error* (RMSE) foi igual a 8.7, demonstrando que a discrepância média entre o número de bagos estimado pelo polinômio e o número de bagos observado é relativamente baixo, reforçando a confiabilidade do polinómio.

CAPÍTULO 4

Análise de Resultados

Após a definição e implementação das abordagens consideradas mais adequadas ao caso em estudo, foi conduzida uma série de testes experimentais com o objetivo de avaliar o desempenho real do sistema em condições práticas.

Estes testes tiveram como finalidade compreender não apenas a eficácia das metodologias propostas em termos de deteção, segmentação e extração de informação, mas também a sua aplicabilidade no contexto operacional da viticultura.

A avaliação prática permitiu simular a utilização do sistema por um produtor, analisando a sua capacidade de fornecer, de forma mais precisa, rápida e acessível, informações relevantes sobre o estado da produção.

4.1. Análise de Resultados do Sistema

Para avaliar de forma rigorosa o desempenho do modelo desenvolvido, procedeu-se a uma análise detalhada dos resultados gerados por cada um dos módulos constituintes do sistema proposto, a deteção, a segmentação e classificação. Esta avaliação permite identificar o contributo individual de cada etapa para o desempenho global.

4.1.1. Método de Validação

O número de cachos detetados, a cor predominante média e o número médio de bagos por cacho são indicadores que ajudam a avaliar o estado da produção. Para que estes indicadores tenham valor prático e forneçam informação útil ao produtor, o seu cálculo refletir de maneira realista as condições observadas em campo.

Para garantir a fiabilidade do sistema, foi realizado um conjunto de testes específicos para cada técnica implementada no algoritmo. Em geral, estes testes basearam-se na comparação direta entre os resultados produzidos pelo sistema e os valores obtidos manualmente por um observador humano, considerados como referência.

Assim, para aferir o desempenho do sistema, foi capturado um vídeo de uma carreira que não integra o conjunto de treino utilizado em nenhuma das partes do algoritmo desenvolvido. O vídeo tem um tempo de gravação de nove minutos e sete segundos, com uma resolução de gravação de 720p, com o intuito de permitir analisar o comportamento do modelo e o seu desempenho na aquisição de dados.

Na carreira utilizada como referência, foi realizada uma contagem manual, na qual foram contabilizados 963 cachos de uva. Estes cachos encontravam-se, na sua maioria, em fase final de maturação, apresentando como cor predominante uma coloração preta densa, característica desta etapa de desenvolvimento do cacho.

No que respeita ao número médio de bagos por cacho, foi adotado o valor utilizado na prática agrícola de poda de bagos, que tem como objetivo uniformizar os cachos, garantindo um equilíbrio entre a dimensão e qualidade do cacho. Na operação agrícola a que se refere este trabalho, procurou-se que a poda conduzisse a que cada cacho contivesse aproximadamente 90 bagos, seguindo-se uma estratégia que promove um desenvolvimento mais eficiente do cacho sem comprometer a qualidade final do produto.

Todos estes dados que caracterizam a carreira, nomeadamente o número total de cachos, o estado de maturação, a cor predominante média e o número médio de bagos por cacho, foram definidos como valores de referência. Desta forma, estes valores servem como padrão de comparação para avaliar o desempenho do sistema proposto, permitindo validar se os resultados produzidos pelo sistema estão de acordo com as condições reais observadas em campo.

4.1.2. Número de Cachos Detetados

A partir do primeiro módulo do sistema proposto, é possível calcular automaticamente o número de cachos identificados em cada vídeo processado. Assim, a avaliação deste módulo tem como base a comparação do valor de referência mencionado com o valor devolvido pelo modelo. Além desta avaliação, também foi realizada uma análise ao tempo de inferência que é necessário para que o sistema conclua o processamento.

Na Tabela 5 são apresentados os principais parâmetros de avaliação referentes a este módulo do sistema, que incluem o número de cachos detetados pelo modelo, o número de cachos de referência obtidos manualmente e o tempo de inferência necessário para que o modelo processe as imagens e devolva os resultados.

Tabela 5. Resultados do desempenho da deteção do sistema

Nº de cachos de referência	Nº de cachos detetados (Sistema)	Tempo de Inferência
963	**984**	**15 minutos**

Os resultados apresentados têm como base o processo completo de deteção dos cachos presentes no vídeo adquirido, que envolve a leitura dos mesmos, a análise individual de cada *frame* e o registo de todas as informações sobre a deteção do cacho identificado numa estrutura de dados previamente definida. A execução deste processamento foi realizada na plataforma Google Colab [83], recorrendo a uma GPU NVIDIA A100.

Com base nos resultados obtidos, o modelo identificou 984 cachos, enquanto o valor de referência foi de 963 cachos, ou seja, esta proximidade entre os valores demonstra um desempenho bastante interessante do sistema. Além disso, o processamento dos 16421 *frames* do vídeo, foi concluído em 14 minutos e 8 segundos, resultando em um tempo médio de processamento por *frame* de aproximadamente 51.6 ms/*frame*. Assim, este desempenho representa um resultado bastante relevante, tendo em conta que uma contagem manual realizada por um humano demora, em média, uma hora e meia por carreira.

4.1.3. Cor Predominante e Número de Bagos por Cacho

De seguida, foi realizada uma análise da cor predominante retornada pelo modelo, que consistiu no cálculo de similaridade entre a cor predominante das colorações recolhidas com a cor de referência característica desta variedade. Após esta comparação, procedeu-se ao cálculo do número médio de bagos por cacho.

Com base na Figura 20, observa-se que a cor predominante identificada pelo sistema na subfigura a), mostra uma boa similaridade com a cor de referência da variedade em b), que foi convertida do padrão *Royal Horticultural Society* (RHS) *Black* 202A para o espaço de cor RGB. Para avaliar esta similaridade, foi calculada a distância euclidiana no espaço de cor perceptualmente uniforme LAB, que é um espaço projetado para refletir a perceção humana da cor, de modo que distâncias maiores correspondem a diferenças visíveis para o observador.

a) b)

Figura 20. Comparação entre a cor predominante detetada a) e a cor de referência b)

Neste caso, o valor desta métrica foi igual a 9.27, indicando uma diferença percetível entre as cores, mas ainda dentro de limites aceitáveis para a variedade em questão, por isso o sistema consegue estimar de forma confiável a cor predominante dos cachos, fornecendo informações coerentes com as características visuais esperadas para o ponto de maturação avaliado.

Após a extração da cor predominante da produção, procedeu-se à estimativa do número médio de bagos por cacho. A partir dos cachos previamente segmentados e armazenados num conjunto de imagens, estas são processadas com o objetivo de realizar a contagem dos bagos. Deste forma, cada imagem é lida e só depois é que o modelo de segmentação selecionado é aplicado sobre os cachos identificados. Após a segmentação, procede-se à recolha do número de segmentos gerados, sendo cada um deles correspondente a um bago identificado no cacho de uva.

A Tabela 6 apresenta os resultados obtidos, incluindo o número médio de bagos detetado, bem como o tempo de inferência necessário para a execução do processo de segmentação dos bagos. O tempo de inferência de uma hora e cinco minutos inclui todos os processos descritos anteriormente, mas se forem excluídos todos menos a parte da aplicação do modelo de segmentação, verifica-se que o modelo demorou na realidade, em média, 1.93 segundos a segmentar cada imagem que pode conter um ou mais cachos de uva.

Tabela 6. Resultados do desempenho da segmentação do sistema

Nº de bagos de referência	**Nº médio de bagos detetados**	**Tempo de inferência**
90	**50**	**1 hora e 5 minutos**

Os resultados obtidos após a aplicação da estratégia de regressão polinomial descrita na secção 3.6 indicam que, em média, cada cacho contém aproximadamente 50 bagos, valor que, embora relevante para a caracterização da produção, apresenta uma diferença considerável quando comparado com o valor de referência de inferência de 90 bagos por cacho.

Esta discrepância sugere que o sistema ainda apresenta limitações na etapa de segmentação interna dos bagos, possivelmente devido a fatores como sobreposição de bagos, variação de iluminação ou oclusões parciais nos cachos, que dificultam a deteção completa de todos os elementos.

CAPÍTULO 5

Conclusões

O principal objetivo desta dissertação consistiu no desenvolvimento de um sistema de visão computacional para a monitorização da produção de uva de mesa, capaz de extrair informação relevante sobre o estado da produção a partir de dados recolhidos em campo.

O sistema proposto foi concebido de modo a responder às questões de investigação formuladas no início deste trabalho, cumprindo assim a finalidade delineada para o estudo. Em relação à primeira e segunda questão de investigação, o sistema desenvolvido demonstra que a integração dos diferentes módulos, baseados nos modelos YOLOv12s e SAM2.1_b+ nas tarefas de deteção e segmentação, possibilita a análise automática e detalhada dos vídeos recolhidos em campo. Assim, a partir do algoritmo incorporado no sistema desenvolvido, é possível extrair um conjunto de indicadores quantitativos e qualitativos sobre o estado da produção, tais como o número de cachos, o número médio de bagos por cacho e a cor predominante.

Estes assumem um papel fundamental na avaliação do estado da produção, uma vez que permitem de forma prática e acessível conhecer melhor as características da produção num dado momento. Com base nestas informações, é possível realizar diferentes tipos de análise, como o cálculo de estimativas de rendimento mais próximas da realidade, apoiando decisões estratégicas no planeamento agrícola e na gestão dos recursos disponíveis. No caso específico da coloração, o sistema disponibiliza uma métrica objetiva que pode ser utilizada como referência na comparação com a cor ideal ou esperada, auxiliando na determinação do momento ótimo de colheita e na garantia da qualidade do produto final.

O trabalho desenvolvido demonstrou que a integração de técnicas de visão computacional no contexto da viticultura, permite desenvolver uma nova ferramenta promissora no auxílio de uma monitorização mais eficiente, contribuindo para uma gestão mais rigorosa e sustentável da produção.

Assim, esta dissertação representa um contributo significativo para o avanço do conhecimento na área da agricultura de precisão aplicada à viticultura, abrindo caminho para investigações futuras e para a aplicação prática do sistema em diferentes realidades produtivas.

5.1. Limitações

Apesar dos resultados alcançados, o sistema desenvolvido apresenta algumas limitações que importa considerar no seu estado atual, como é o caso da iluminação quando é realizada as captura dos vídeos. É importante referir que tanto as situações de luminosidade excessiva como as de luminosidade

reduzida comprometem a precisão dos processos de deteção e segmentação dos cachos, provocando a perda de informação nestas situações.

Assim, recomenda-se que as gravações sejam realizadas em condições de luz equilibradas, preferencialmente em períodos do dia com iluminação uniforme e difusa, de modo a reduzir a presença de sombras intensas e reflexos que possam introduzir ruído no processamento.

Relativamente à exigência computacional do sistema, este foi concebido para ser de fácil utilização por diferentes utilizadores. A sua execução eficiente requer o uso de uma unidade de processamento gráfico (GPU) recente, dado o elevado custo computacional dos modelos de deteção e segmentação envolvidos.

Esta necessidade pode representar um entrave para produtores ou utilizadores sem acesso a hardware de elevado desempenho. Contudo, esta limitação pode ser contornada recorrendo a serviços de computação em nuvem, como o Google Colab [83] ou outras plataformas equivalentes, que permitem alugar recursos de processamento a custos acessíveis e sem necessidade de investimento em equipamentos dedicados.

Além disso, importa referir que o sistema desenvolvido requer de uma validação mais extensiva, de modo a abranger um número superior de casos de teste e cenários de aplicação, com o intuito de avaliar o seu desempenho de forma mais abrangente e robusta, identificando eventuais limitações em contextos distintos daqueles utilizados durante a fase experimental.

De forma complementar, seria igualmente relevante dispor de um conjunto de dados de campo mais representativo, que pudesse ser utilizado como referência para a avaliação quantitativa do desempenho. A utilização de dados reais permitiria substituir os atuais valores de referência teóricos, como é o caso do número de bagos de referência utilizado na avaliação do sistema.

5.2. Trabalho Futuro

Relativamente ao trabalho futuro, podem ser realizadas algumas melhorias de modo a ampliar o impacto e a aplicabilidade do sistema desenvolvido, que não foram possíveis de serem desenvolvidas e testadas no tempo útil deste projeto proposto.

Uma das principais direções consiste na criação de uma plataforma *user-friendly*, que permita aos utilizadores interagir com o sistema de forma simples e intuitiva, reduzindo a necessidade de conhecimentos técnicos de software, promovendo a sua adoção por produtores e técnicos agrícolas. Assim, esta plataforma poderá incluir funcionalidades como upload direto de vídeos, visualização imediata dos resultados e exportação automática dos relatórios obtidos, sem ter de visualizar qualquer tipo de informação sobre o código por detrás do sistema desenvolvido.

Outra melhoria, passa pelo aumento e diversificação do conjunto de dados através de outras variedades. Embora o conjunto de dados atual tenha permitido alcançar resultados interessantes, o aumento do número de dados, a partir da aquisição de mais imagens de diferentes variedades de uva de mesa, em condições de iluminação, estágios de maturação em ângulos de visão diferentes, permitirá aumentar a robustez do modelo, garantindo melhor capacidade de generalização e sobretudo em melhorias no seu desempenho a nível da deteção e segmentação. Adicionalmente, seria importante complementar a *ground-truth* do novo conjunto de dados com mais informações acerca dos números reais de cachos por carreira e de bagos por cacho, com o intuito de avaliar melhor os resultados produzidos pelo sistema.

Numa perspetiva mais avançada, destaca-se a possibilidade de integrar com o sistema implementado câmaras RGB-D, que permitiria ter a localização 3D dos cachos. A utilização deste tipo de ferramentas, possibilitaria o mapeamento espacial da produção e uma melhor perceção do tamanho dos cachos e bagos, criando oportunidades de análise mais eficiente, como a correlação com cartas de solos, sistemas de rega ou outros indicadores fundamentais para a gestão da produção.

No que diz respeito à contagem do número total de bagos por cacho, o método atual demonstrou bons resultados, mas o seu desempenho depende da qualidade do vídeo capturado e das condições de iluminação. Assim, torna-se relevante investigar sobre metodologias mais robustas, capazes de considerar diferentes características sobre uma dada deteção realizada pelo modelo.

Por fim, uma vertente adicional de melhoria passa pelo estudo aprofundado da otimização do sistema implementado relativamente à gestão de memória e o desempenho do modelo em diferentes unidades de processamento gráfico (GPU), nomeadamente através da utilização de dispositivos com maior capacidade computacional. Este tipo de análise seria relevante não apenas para compreender de forma mais detalhada a relação entre o poder de processamento e o tempo de inferência necessários.

Referências Bibliográficas

[1] Empresa de Desenvolvimento e Infraestruturas do Alqueva (EDIA), "Anuário Agrícola de Alqueva 2024," 2024. Accessed: Apr. 29, 2025. [Online]. Available: https://www.edia.pt/pt/o-que-fazemos/apoio-ao-agricultor/anuario-agricola/

[2] Good Fruit Guide, "Midnight Beauty (Sugra13) - Good Fruit Guide." Accessed: May 02, 2025. [Online]. Available: https://goodfruitguide.co.uk/product/midnight-beauty/

[3] Instituto Nacional de Estatística (INE), "AER2023_III_05," 2023. Accessed: Apr. 29, 2025. [Online]. Available: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&indOcorrCod=0013080&contexto=bd&selTab=tab2

[4] T. Amemiya, C. S. Leow, P. Buayai, K. Makino, X. Mao, and H. Nishizaki, "Appropriate grape color estimation based on metric learning for judging harvest timing," *Visual Computer*, vol. 38, no. 12, pp. 4083–4094, Dec. 2022, doi: 10.1007/s00371-022-02666-0.

[5] Y. Li, Y. Wang, D. Xu, J. Zhang, and J. Wen, "An Improved Mask RCNN Model for Segmentation of 'Kyoho' (Vitis labruscana) Grape Bunch and Detection of Its Maturity Level," *Agriculture (Switzerland)*, vol. 13, no. 4, Apr. 2023, doi: 10.3390/agriculture13040914.

[6] P. Upadhyaya, M. Karkee, S. Kshetri, and A. Paudel, "Automated lag-phase detection in wine grapes using a mobile vision system," *Smart Agricultural Technology*, vol. 7, Mar. 2024, doi: 10.1016/j.atech.2023.100381.

[7] Iberinform, "PRAZER DOS AROMAS - UNIPESSOAL, LDA." Accessed: Apr. 29, 2025. [Online]. Available: https://www.iberinform.pt/empresa/24695857/prazer-dos-aromas-unipessoal-lda

[8] Sun World International, "Sun World - Global Fruit Variety Development & Licensing." Accessed: May 02, 2025. [Online]. Available: https://www.sun-world.com/

[9] D. Ahmedt-Aristizabal *et al.*, "An In-Field Dynamic Vision-Based Analysis for Vineyard Yield Estimation," *IEEE Access*, vol. 12, pp. 102146–102166, 2024, doi: 10.1109/ACCESS.2024.3431244.

[10] C. Lawrence, T. Tuunanen, and M. D. Myers, "Extending design science research methodology for a multicultural world," *IFIP Adv Inf Commun Technol*, vol. 318, pp. 108–121, 2010, doi: 10.1007/978-3-642-12113-5_7.

[11] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," *IEEE Trans Pattern Anal Mach Intell*, vol. 42, no. 2, pp. 386–397, Mar. 2017, doi: 10.1109/TPAMI.2018.2844175.

[12] "A Gentle Introduction to the Rectified Linear Unit (ReLU) - MachineLearningMastery.com." Accessed: Sep. 28, 2025. [Online]. Available: https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks/

[13] S. Basodi, C. Ji, H. Zhang, and Y. Pan, “Gradient amplification: An efficient way to train deep neural networks,” *Big Data Mining and Analytics*, vol. 3, no. 3, pp. 196–207, Sep. 2020, doi: 10.26599/BDMA.2020.9020004.

[14] “YOLO12: Detecção de Objetos Centrada na Atenção - Documentos Ultralytics YOLO.” Accessed: Aug. 23, 2025. [Online]. Available: https://docs.ultralytics.com/pt/models/yolo12/

[15] W. Liu *et al.*, “SSD: Single Shot MultiBox Detector,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9905 LNCS, pp. 21–37, Dec. 2016, doi: 10.1007/978-3-319-46448-0_2.

[16] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, “Focal Loss for Dense Object Detection,” *IEEE Trans Pattern Anal Mach Intell*, vol. 42, no. 2, pp. 318–327, Aug. 2017, doi: 10.1109/TPAMI.2018.2858826.

[17] M. Tan, R. Pang, and Q. V. Le, “EfficientDet: Scalable and Efficient Object Detection,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 10778–10787, Nov. 2019, doi: 10.1109/CVPR42600.2020.01079.

[18] “RT-DETR da Baidu: Um Detector de Objetos em Tempo Real Baseado em Vision Transformer.” Accessed: Aug. 23, 2025. [Online]. Available: https://docs.ultralytics.com/pt/models/rtdetr/

[19] “Region-Based Convolutional Networks for Accurate Object Detection and Segmentation | IEEE Journals & Magazine | IEEE Xplore.” Accessed: Aug. 23, 2025. [Online]. Available: https://ieeexplore.ieee.org/document/7112511

[20] “[1504.08083] Fast R-CNN.” Accessed: Aug. 23, 2025. [Online]. Available: https://arxiv.org/abs/1504.08083

[21] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” *IEEE Trans Pattern Anal Mach Intell*, vol. 39, no. 6, pp. 1137–1149, Jun. 2015, doi: 10.1109/TPAMI.2016.2577031.

[22] “Explicação sobre os detectores de objectos de duas fases | Ultralytics.” Accessed: Aug. 28, 2025. [Online]. Available: https://www.ultralytics.com/pt/glossary/two-stage-object-detectors

[23] “Semantic Image Segmentation with DeepLab in TensorFlow.” Accessed: Aug. 24, 2025. [Online]. Available: https://research.google/blog/semantic-image-segmentation-with-deeplab-in-tensorflow/

[24] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs,” *IEEE Trans Pattern Anal Mach Intell*, vol. 40, no. 4, pp. 834–848, Jun. 2016, doi: 10.1109/TPAMI.2017.2699184.

[25] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, “Rethinking Atrous Convolution for Semantic Image Segmentation,” Jun. 2017, Accessed: Sep. 15, 2025. [Online]. Available: https://arxiv.org/pdf/1706.05587

[26] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs,” pp. 1–12, Dec. 2014, Accessed: Sep. 15, 2025. [Online]. Available: https://arxiv.org/pdf/1412.7062

[27] “Segment Anything | Meta AI.” Accessed: Aug. 24, 2025. [Online]. Available: https://segment-anything.com/

[28] Web of Science, “Web of Science | Clarivate.” Accessed: May 02, 2025. [Online]. Available: https://clarivate.com/academia-government/scientific-and-academic-research/research-discovery-and-referencing/web-of-science/

[29] Scopus, “Scopus preview - Scopus - Welcome to Scopus.” Accessed: May 02, 2025. [Online]. Available: https://www.scopus.com/

[30] PRISMA statement, “PRISMA 2020 flow diagram — PRISMA statement.” Accessed: May 02, 2025. [Online]. Available: https://www.prisma-statement.org/prisma-2020-flow-diagram

[31] R. Íñiguez, S. Gutiérrez, C. Poblete-Echeverría, I. Hernández, I. Barrio, and J. Tardáguila, “Deep learning modelling for non-invasive grape bunch detection under diverse occlusion conditions,” *Comput Electron Agric*, vol. 226, Nov. 2024, doi: 10.1016/j.compag.2024.109421.

[32] L. Mohimont, F. Alin, M. Rondeau, N. Gaveau, and L. A. Steffenel, “Computer Vision and Deep Learning for Precision Viticulture,” Oct. 01, 2022, *MDPI*. doi: 10.3390/agronomy12102463.

[33] S. Lu, X. Liu, Z. He, X. Zhang, W. Liu, and M. Karkee, “Swin-Transformer-YOLOv5 for Real-Time Wine Grape Bunch Detection,” *Remote Sens (Basel)*, vol. 14, no. 22, Nov. 2022, doi: 10.3390/rs14225853.

[34] I. Pinheiro *et al.*, “Deep Learning YOLO-Based Solution for Grape Bunch Detection and Assessment of Biophysical Lesions,” *Agronomy*, vol. 13, no. 4, Apr. 2023, doi: 10.3390/agronomy13041120.

[35] N. Sneha, M. Sundaram, and R. Ranjan, “Acre-Scale Grape Bunch Detection and Predict Grape Harvest Using YOLO Deep Learning Network,” *SN Comput Sci*, vol. 5, no. 2, Feb. 2024, doi: 10.1007/s42979-023-02572-9.

[36] M. Sozzi, S. Cantalamessa, A. Cogato, A. Kayad, and F. Marinello, “Automatic Bunch Detection in White Grape Varieties Using YOLOv3, YOLOv4, and YOLOv5 Deep Learning Algorithms,” *Agronomy*, vol. 12, no. 2, Feb. 2022, doi: 10.3390/agronomy12020319.

[37] C. Zhang, H. Ding, Q. Shi, and Y. Wang, “Grape Cluster Real-Time Detection in Complex Natural Scenes Based on YOLOv5s Deep Learning Network,” *Agriculture (Switzerland)*, vol. 12, no. 8, Aug. 2022, doi: 10.3390/agriculture12081242.

[38] T. Zhang, F. Wu, M. Wang, Z. Chen, L. Li, and X. Zou, “Grape-Bunch Identification and Location of Picking Points on Occluded Fruit Axis Based on YOLOv5-GAP,” *Horticulturae*, vol. 9, no. 4, Apr. 2023, doi: 10.3390/horticulturae9040498.

[39] J. Chen *et al.*, “GA-YOLO: A Lightweight YOLO Model for Dense and Occluded Grape Target Detection,” *Horticulturae*, vol. 9, no. 4, Apr. 2023, doi: 10.3390/horticulturae9040443.

[40] B. Liu *et al.*, “An improved lightweight network based on deep learning for grape recognition in unstructured environments,” *Information Processing in Agriculture*, vol. 11, no. 2, pp. 202–216, Jun. 2024, doi: 10.1016/j.inpa.2023.02.003.

[41] M. Rudenko *et al.*, “Intelligent Monitoring System to Assess Plant Development State Based on Computer Vision in Viticulture,” *Computation*, vol. 11, no. 9, Sep. 2023, doi: 10.3390/computation11090171.

[42] B. Parr, M. Legg, and F. Alam, “Grape yield estimation with a smartphone’s colour and depth cameras using machine learning and computer vision techniques,” *Comput Electron Agric*, vol. 213, Oct. 2023, doi: 10.1016/j.compag.2023.108174.

[43] “What Is SLAM (Simultaneous Localization and Mapping)? - MATLAB & Simulink.” Accessed: Aug. 24, 2025. [Online]. Available: https://www.mathworks.com/discovery/slam.html

[44] M. R. González, M. E. Martínez-Rosas, and C. A. Brizuela, “Comparison of CNN architectures for single grape detection,” *Comput Electron Agric*, vol. 231, Apr. 2025, doi: 10.1016/j.compag.2025.109930.

[45] Z. Liu *et al.*, “Swin Transformer: Hierarchical Vision Transformer using Shifted Windows”, Accessed: Aug. 24, 2025. [Online]. Available: https://github.

[46] L. Bertinetto, J. Valmadre, J. F. Henriques, A. Vedaldi, and P. H. S. Torr, “Fully-Convolutional Siamese Networks for Object Tracking,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9914 LNCS, pp. 850–865, Jun. 2016, doi: 10.1007/978-3-319-48881-3_56.

[47] Y. Chen, X. Li, M. Jia, J. Li, T. Hu, and J. Luo, “Instance Segmentation and Number Counting of Grape Berry Images Based on Deep Learning,” *Applied Sciences (Switzerland)*, vol. 13, no. 11, Jun. 2023, doi: 10.3390/app13116751.

[48] F. Khoroshevsky, S. Khoroshevsky, and A. Bar-Hillel, “Parts-per-object count in agricultural images: Solving phenotyping problems via a single deep neural network,” *Remote Sens (Basel)*, vol. 13, no. 13, Jul. 2021, doi: 10.3390/rs13132496.

[49] R. Marani, A. Milella, A. Petitti, and G. Reina, “Deep neural networks for grape bunch segmentation in natural images from a consumer-grade camera,” *Precis Agric*, vol. 22, no. 2, pp. 387–413, Apr. 2021, doi: 10.1007/s11119-020-09736-0.

[50] “resnet101 — Torchvision main documentation.” Accessed: Aug. 24, 2025. [Online]. Available: https://docs.pytorch.org/vision/main/models/generated/torchvision.models.resnet101.html

[51] A. Casado-García, J. Heras, A. Milella, and R. Marani, “Semi-supervised deep learning and low-cost cameras for the semantic segmentation of natural images in viticulture,” *Precis Agric*, vol. 23, no. 6, pp. 2001–2026, Dec. 2022, doi: 10.1007/s11119-022-09929-9.

[52] I. Terzi, M. M. Ozguven, and A. Yagci, “Automatic detection of grape varieties with the newly proposed CNN model using ampelographic characteristics,” *Sci Hortic*, vol. 334, Aug. 2024, doi: 10.1016/j.scienta.2024.113340.

[53] W. Du, X. Cui, Y. Zhu, and P. Liu, “Detection of table grape berries need to be removed before thinning based on deep learning,” *Comput Electron Agric*, vol. 231, Apr. 2025, doi: 10.1016/j.compag.2025.110043.

[54] “resnet18 — Torchvision main documentation.” Accessed: Aug. 24, 2025. [Online]. Available: https://docs.pytorch.org/vision/main/models/generated/torchvision.models.resnet18

[55] “(PDF) Long Short-Term Memory.” Accessed: Aug. 24, 2025. [Online]. Available: https://www.researchgate.net/publication/13853244_Long_Short-Term_Memory

[56] P. Buayai, K. Yok-In, D. Inoue, H. Nishizaki, K. Makino, and X. Mao, “Supporting table grape berry thinning with deep neural network and augmented reality technologies,” *Comput Electron Agric*, vol. 213, Oct. 2023, doi: 10.1016/j.compag.2023.108194.

[57] L. Zabawa, A. Kicherer, L. Klingbeil, R. Töpfer, H. Kuhlmann, and R. Roscher, “Counting of grapevine berries in images via semantic segmentation using convolutional neural networks,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 164, pp. 73–83, Jun. 2020, doi: 10.1016/j.isprsjprs.2020.04.002.

[58] T. A. Ciarfuglia, I. M. Motoi, L. Saraceni, M. Fawakherji, A. Sanfeliu, and D. Nardi, “Weakly and semi-supervised detection, segmentation and tracking of table grapes with limited and noisy data,” *Comput Electron Agric*, vol. 205, Feb. 2023, doi: 10.1016/j.compag.2023.107624.

[59] V. Bruni, G. Dominijanni, D. Vitulano, and G. Ramella, “A perception-guided CNN for grape bunch detection,” *Math Comput Simul*, vol. 230, pp. 111–130, Apr. 2025, doi: 10.1016/j.matcom.2024.11.004.

[60] A. S. Aguiar *et al.*, “Grape bunch detection at different growth stages using deep learning quantized models,” *Agronomy*, vol. 11, no. 9, Sep. 2021, doi: 10.3390/agronomy11091890.

[61] F. Palacios, P. Melo-Pinto, M. P. Diago, and J. Tardaguila, “Deep learning and computer vision for assessing the number of actual berries in commercial vineyards,” *Biosyst Eng*, vol. 218, pp. 175–188, Jun. 2022, doi: 10.1016/j.biosystemseng.2022.04.015.

[62] R. P. Devanna, G. Reina, F. A. Cheein, and A. Milella, “Boosting grape bunch detection in RGB-D images using zero-shot annotation with Segment Anything and GroundingDINO,” *Comput Electron Agric*, vol. 229, Feb. 2025, doi: 10.1016/j.compag.2024.109611.

[63] N. Zhou *et al.*, “DepthSeg: Depth prompting in remote sensing semantic segmentation,” Jun. 2025, Accessed: Aug. 25, 2025. [Online]. Available: https://arxiv.org/pdf/2506.14382

[64] P. Buayai, K. R. Saikaew, and X. Mao, “End-to-End Automatic Berry Counting for Table Grape Thinning,” *IEEE Access*, vol. 9, pp. 4829–4842, 2021, doi: 10.1109/ACCESS.2020.3048374.

[65] K. Chen *et al.*, “Hybrid Task Cascade for Instance Segmentation,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2019-June, pp. 4969–4978, Jan. 2019, doi: 10.1109/CVPR.2019.00511.

[66] V. Bruni, G. Dominijanni, and D. Vitulano, “A Machine-Learning Approach for Automatic Grape-Bunch Detection Based on Opponent Colors,” *Sustainability (Switzerland)*, vol. 15, no. 5, Mar. 2023, doi: 10.3390/su15054341.

[67] “1.4. Support Vector Machines — scikit-learn 1.7.1 documentation.” Accessed: Aug. 25, 2025. [Online]. Available: https://scikit-learn.org/stable/modules/svm.html

[68] L. Luo *et al.*, “Grape berry detection and size measurement based on edge image processing and geometric morphology,” *Machines*, vol. 9, no. 10, Oct. 2021, doi: 10.3390/machines9100233.

[69] M. Tardif *et al.*, “Two-stage automatic diagnosis of Flavescence Dorée based on proximal imaging and artificial intelligence: a multi-year and multi-variety experimental study,” *Oeno One*, vol. 56, no. 3, pp. 371–384, Sep. 2022, doi: 10.20870/oeno-one.2022.56.3.5460.

[70] “RandomForestClassifier — scikit-learn 1.7.1 documentation.” Accessed: Aug. 28, 2025. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

[71] C. Yang, T. Geng, J. Peng, and Z. Song, “Probability map-based grape detection and counting,” *Comput Electron Agric*, vol. 224, Sep. 2024, doi: 10.1016/j.compag.2024.109175.

[72] M. Gavrilović, D. Jovanović, P. Božović, P. Benka, and M. Govedarica, “Vineyard Zoning and Vine Detection Using Machine Learning in Unmanned Aerial Vehicle Imagery,” *Remote Sens (Basel)*, vol. 16, no. 3, Feb. 2024, doi: 10.3390/rs16030584.

[73] S. Vélez, M. Ariza-Sentís, M. Triviño, A. C. Cob-Parro, M. Mila, and J. Valente, “Framework for smartphone-based grape detection and vineyard management using UAV-trained AI,” *Heliyon*, vol. 11, no. 4, Feb. 2025, doi: 10.1016/j.heliyon.2025.e42525.

[74] M. Ariza-Sentís, H. Baja, S. Vélez, and J. Valente, “Object detection and tracking on UAV RGB videos for early extraction of grape phenotypic traits,” *Comput Electron Agric*, vol. 211, Aug. 2023, doi: 10.1016/j.compag.2023.108051.

[75] “labelImg · PyPI.” Accessed: Aug. 29, 2025. [Online]. Available: https://pypi.org/project/labelImg/

[76] D. E. Székely *et al.*, “Bacterial-fungicidal vine disease detection with proximal aerial images,” *Heliyon*, vol. 10, no. 14, Jul. 2024, doi: 10.1016/j.heliyon.2024.e34017.

[77] R. P. Devanna, L. Romeo, G. Reina, and A. Milella, “Yield estimation in precision viticulture by combining deep segmentation and depth-based clustering,” *Comput Electron Agric*, vol. 232, May 2025, doi: 10.1016/j.compag.2025.110025.

[78] Z. Xu *et al.*, “Realtime Picking Point Decision Algorithm of Trellis Grape for High-Speed Robotic Cut-and-Catch Harvesting,” *Agronomy*, vol. 13, no. 6, Jun. 2023, doi: 10.3390/agronomy13061618.

[79] G. Coll-Ribes, I. J. Torres-Rodríguez, A. Grau, E. Guerra, and A. Sanfeliu, “Accurate detection and depth estimation of table grapes and peduncles for robot harvesting, combining monocular depth estimation and CNN methods,” *Comput Electron Agric*, vol. 215, Dec. 2023, doi: 10.1016/j.compag.2023.108362.

[80] “Aumento de Dados usando Ultralytics YOLO - Documentos Ultralytics YOLO.” Accessed: Aug. 29, 2025. [Online]. Available: https://docs.ultralytics.com/pt/guides/yolo-data-augmentation/#saturation-adjustment-hsv_s

[81] N. Aharon, R. Orfaig, and B.-Z. Bobrovsky, “BoT-SORT: Robust Associations Multi-Pedestrian Tracking,” Jun. 2022, Accessed: Sep. 28, 2025. [Online]. Available: https://arxiv.org/pdf/2206.14651

[82] “OpenCV: Canny Edge Detection.” Accessed: Aug. 29, 2025. [Online]. Available: https://docs.opencv.org/4.x/da/d22/tutorial_py_canny.html

[83] “colab.google.” Accessed: Aug. 30, 2025. [Online]. Available: https://colab.google/