

**UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ**

**DARLAN ALAOR SELZLEIN**

**APLICAÇÃO DO PROCESSAMENTO DE IMAGENS E DE RECONHECIMENTO  
DE PADRÕES COM REDE NEURAL NA ESTIMAÇÃO DA PRODUÇÃO DE  
MACIEIRAS**

**MEDIANEIRA**

**2022**

**DARLAN ALAOR SELZLEIN**

**APLICAÇÃO DO PROCESSAMENTO DE IMAGENS E DE RECONHECIMENTO  
DE PADRÕES COM REDE NEURAL NA ESTIMAÇÃO DA PRODUÇÃO DE  
MACIEIRAS**

**Application of image processing and pattern recognition with neural  
network in apple production estimation**

Trabalho de Conclusão de Curso de Graduação apresentado como requisito para obtenção do título de Bacharel em Ciência da Computação do Curso de Bacharelado em Ciência da Computação da Universidade Tecnológica Federal do Paraná.

Orientador: Prof. Dr. Jorge Aikes Junior

Coorientador: Prof. Dr. Pedro Luiz de Paula Filho

**MEDIANEIRA**

**2022**



[4.0 Internacional](https://creativecommons.org/licenses/by/4.0/)

Esta licença permite compartilhamento, remixe, adaptação e criação a partir do trabalho, mesmo para fins comerciais, desde que sejam atribuídos créditos ao(s) autor(es). Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.

**DARLAN ALAOR SELZLEIN**

**APLICAÇÃO DO PROCESSAMENTO DE IMAGENS E DE RECONHECIMENTO  
DE PADRÕES COM REDE NEURAL NA ESTIMAÇÃO DA PRODUÇÃO DE  
MACIEIRAS**

Trabalho de Conclusão de Curso de Graduação  
apresentado como requisito para obtenção do  
título de Bacharel em Ciência da Computação  
do Curso de Bacharelado em Ciência da  
Computação da Universidade Tecnológica  
Federal do Paraná.

Data de aprovação: 20/junho/2022

---

Jorge Aikes Junior  
Doutor  
Universidade Tecnológica Federal do Paraná

---

Arnaldo Candido Junior  
Doutor  
Universidade Tecnológica Federal do Paraná

---

Nelson Miguel Betzek  
Doutor  
Universidade Tecnológica Federal do Paraná

**MEDIANEIRA**

**2022**

## **AGRADECIMENTOS**

Agradeço, primeiramente, à minha família. Aos meus pais, pela exigência e confiança na finalização deste trabalho, e aos meus irmãos, pelos momentos de conversa e aconselhamento.

Agradeço também à todos meus amigos que de alguma forma contribuíram ao longo do trabalho com ideias, conselhos, conversas e suporte emocional para os momentos mais difíceis, e de maneira especial ao João, à Maria e ao Michel, que me acompanharam durante uma longa caminhada, estando ao meu lado tanto nos momentos bons quanto nos momentos ruins.

À minha namorada, Darah, pela confiança, pensamento positivo e capacidade de trazer alegria àqueles que estão ao seu redor: obrigado por ter sido luz nos dias mais escuros.

Também agradeço à todos aqueles que fizeram parte diretamente desse trabalho. Aos meus orientadores, Prof. Dr. Pedro Luiz de Paula Filho e Prof. Dr. Jorge Aikes Junior, agradeço pela paciência, pelo companheirismo, pelos conselhos, pelos ensinamentos e por todo o suporte ao longo do projeto. Agradeço também à Tassiane Perico Anzolin, por ter dado início ao projeto que originou este trabalho, e ao Carlos Moysés Auzi pelo auxílio na coleta das imagens. Sou grato, também, à EMBRAPA Uva e Vinho, ao Engenheiro Agrônomo Luciano Gebler, ao Prof. Herculano e ao Claudemir, pela oportunidade de realizar este trabalho e pelo auxílio com a confecção e montagem da estrutura.

Por fim, agradeço de modo geral todos aqueles que fizeram parte dessa etapa de minha vida. Muito obrigado.

## RESUMO

A tecnologia é uma ferramenta importante no meio agrícola mundial, sendo responsável por melhorias na produtividade, com técnicas como a agricultura de precisão, e pela automação de processos antes manuais, como a colheita. Entre as diversas áreas da agricultura, a produção frutífera possui grande destaque global e um dos cultivos mais comuns desse campo é a maçã. Como todos os campos da agricultura, a maçã é alvo de investimentos tecnológicos que buscam melhorias de produtividade e de qualidade de cultivo e de vida dos agricultores. Diante disso, este trabalho teve como objetivo fazer a coleta de imagens de pomares de maçã e por meio delas estimar a quantidade de maçãs produzidas. Para tal, uma ferramenta de coleta de fotos foi elaborada e utilizada para a obtenção dessas imagens, que posteriormente foram unificadas em diferentes imagens panorâmicas. Foram, então, treinados modelos de detecção de objetos para identificar as maçãs nas fotos e, dessa forma, gerar a contagem. A técnica capaz de alcançar uma contagem de maçãs equivalente à 103,15% da contagem real no teste inicial, porém não foi capaz de manter esse resultado para testes subsequentes.

**Palavras-chave:** processamento de imagem; redes neurais artificiais; reconhecimento de padrões; produção de maçãs.

## ABSTRACT

Technology is an important tool in the global agricultural environment, being responsible for improvements in productivity, with techniques such as precision agriculture, and for the automation of previously manual processes, such as harvesting. Among the various areas of agriculture, fruit production has great global prominence and one of the most common crops in this field is apple. Like all fields of agriculture, the apple is the target of technological investments that seek to improve productivity and the quality of cultivation and life of farmers. Therefore, this work aimed to collect images of apple orchards and, through them, estimate the amount of apples produced. To this end, a photo collection tool was developed and used to obtain these images, which were later unified into different panoramic images. Object detection models were then trained to identify the apples in the photos and, in this way, generate the count. The technique was able to achieve an apple count equivalent to 103.15% of the actual count in the initial test, but was not able to maintain this result for subsequent tests.

**Keywords:** image processing; artificial neural networks; pattern recognition; apple production.

## LISTA DE FIGURAS

Figura 1 – Pomar de maçãs . . . . .	16
Figura 2 – Exemplo de <i>stitching</i> de imagens . . . . .	18
Figura 3 – Representação de um neurônio artificial . . . . .	21
Figura 4 – Rede Neural Artificial Multicamada . . . . .	22
Figura 5 – Aplicação de filtro em camada de convolução em uma das regiões dos dados de entrada . . . . .	24
Figura 6 – Aplicação de filtro em camada de convolução com <i>stride 2</i> . . . . .	25
Figura 7 – Aplicação de <i>max pooling</i> com <i>stride 2</i> . . . . .	25
Figura 8 – Fluxograma da CNN <i>Faster R-CNN</i> . . . . .	26
Figura 9 – Detecção de pontos chaves entre duas fotos obtidas pelo VANT . . . . .	27
Figura 10 – Detecção de maçãs utilizando o modelo treinado por Gené-Mola <i>et al.</i> (2019) . . . . .	28
Figura 11 – Detecção de maçãs utilizando o modelo treinado por Tian <i>et al.</i> (2019) . . . . .	29
Figura 12 – Área de pomares no campo experimental da Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA) Uva e Vinho - Vacaria/RS . . . . .	30
Figura 13 – Trator entre linhas de pomar da EMBRAPA Uva e Vinho - Vacaria/RS . . . . .	31
Figura 14 – Processos da metodologia . . . . .	32
Figura 15 – Estrutura embarcada em trator . . . . .	33
Figura 16 – Raspberry Pi 3B (RasPi) e carregador portátil fixados à estrutura . . . . .	34
Figura 17 – Imagem coletada no período da manhã (esquerda) e imagem coletada no período da tarde (direita) . . . . .	35
Figura 18 – Colagem vertical de imagens utilizando taxa de sob. vertical maior que zero . . . . .	36
Figura 19 – Rotulação de maçãs na ferramenta <i>Computer Vision Annotation Tool</i> (CVAT) . . . . .	37
Figura 20 – Imagens obtidas ao mesmo tempo pelas câmeras 1, 2 e 3, da esquerda para a direita, com destaque para o mesmo grupo de maçãs presentes em todas as fotos . . . . .	40
Figura 21 – Sob. de captura entre duas fotos da câmera 1 em tempo vizinhos, com destaque para o mesmo grupo de maçãs aparecendo em ambas as imagens . . . . .	40
Figura 22 – Imagem borrada (esquerda) e imagem ofuscada (direita) . . . . .	41
Figura 23 – <i>Stitching</i> com aplicação do método <i>inRange</i> nas imagens para a localização dos pontos chaves . . . . .	42

Figura 24 – <i>Stitching</i> com a manipulação da camada de cores nas imagens para a localização dos pontos chaves . . . . .	42
Figura 25 – <i>Stitching</i> com aplicação do método <i>inRange</i> e manipulação da camada de cores nas imagens para a localização dos pontos chaves . . . . .	43
Figura 26 – Panorama feito através da colagem de imagens . . . . .	43



## LISTA DE TABELAS

Tabela 1 – Distribuição de imagens entre os <i>datasets</i> . . . . .	37
Tabela 2 – Total de imagens coletadas . . . . .	39
Tabela 3 – Resultados da primeira seção de treinos . . . . .	44
Tabela 4 – Resultados da segunda seção de treinos . . . . .	44
Tabela 5 – Quantidade de maçãs coletadas no campo experimental . . . . .	45
Tabela 6 – Resultados da detecção de maçãs na fila 1 . . . . .	45
Tabela 7 – Resultados da detecção de maçãs nas filas 02 e 03 . . . . .	46

## LISTA DE ABREVIATURAS E SIGLAS

### Abreviaturas

sob. Sobreposição

### Siglas

AM	Aprendizado de Máquina
AP	<i>Average Precision</i>
CNN	<i>Convolutional Neural Network</i>
CPU	<i>Central Processing Unit</i>
CVAT	<i>Computer Vision Annotation Tool</i>
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
FAST	<i>Features from Accelerated Segment Test</i>
GPU	<i>Graphics Processing Unit</i>
IA	Inteligência Artificial
mAP	<i>Mean Average Precision</i>
RasPi	Raspberry Pi 3B
RNA	Rede Neural Artificial
ROI	<i>Region of Interest</i>
SIFT	<i>Scale Invariant Feature Transform</i>
SSD	<i>Solid State Drives</i>
UTFPR	Universidade Tecnológica Federal do Paraná
VANT	Veículo Aéreo Não Tripulado
VGG	<i>Visual Geometry Group</i>
YOLO	<i>You Only Look Once</i>

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>13</b>
1.1	Objetivos geral e específicos	13
<b>2</b>	<b>REFERENCIAL TEÓRICO</b>	<b>15</b>
2.1	Maçãs	15
2.2	Processamento de imagens	17
2.3	Inteligência Artificial e Aprendizado de Máquina	19
2.4	Redes Neurais Artificiais	20
2.4.1	Neurônios	21
2.4.2	Redes multicamadas	22
2.4.3	Treinamento	23
2.4.4	Métricas de desempenho	23
2.4.5	Redes Neurais Convolucionais	23
2.4.6	Modelos	25
2.5	Trabalhos correlatos	26
2.6	Considerações finais	28
<b>3</b>	<b>MATERIAIS E MÉTODOS</b>	<b>30</b>
3.1	Materiais	30
3.2	Métodos	32
3.2.1	Montagem da estrutura	32
3.2.2	Coleta de imagens	34
3.2.3	Geração de panoramas	35
3.2.4	Preparação dos <i>datasets</i>	36
3.2.5	Treinamento de modelos	37
3.2.6	Aplicação de modelo aos panoramas	38
<b>4</b>	<b>RESULTADOS E DISCUSSÕES</b>	<b>39</b>
4.1	Coleta de Imagens	39
4.2	Geração de Panoramas	41
4.3	Treinamento de Modelos	42
4.4	Aplicação de modelo aos panoramas	44
<b>5</b>	<b>CONSIDERAÇÕES FINAIS</b>	<b>47</b>

<b>5.1</b>	<b>Conclusão . . . . .</b>	<b>47</b>
<b>5.2</b>	<b>Trabalhos futuros . . . . .</b>	<b>48</b>
	<b>REFERÊNCIAS . . . . .</b>	<b>49</b>

## 1 INTRODUÇÃO

A agricultura é uma prática ancestral que acompanha as civilizações humanas desde tempos antigos, sendo o seu surgimento um evento de grande impacto na humanidade, permitindo um rápido crescimento na população da espécie (ALFEN, 2014). Ao longo da história da humanidade, novas tecnologias foram desenvolvidas e aplicadas em diversas áreas, sendo uma delas a agricultura, proporcionando aumento de produtividade e maior conforto no trabalho rural, no plantio, manutenção e colheita da lavoura. Além disso, a coleta e processamento de dados por meio de tecnologias permite fornecer aos agricultores maior segurança de planejamentos de gestão de recursos (FOUNDATION, 2021).

No Brasil, a agricultura é uma atividade econômica de grande importância, tendo um papel importante nas exportações do país, permitindo o equilíbrio das contas públicas, além de empregar milhões de trabalhadores. Por ser um país de dimensões continentais, possuindo um território que engloba áreas desde acima da linha do Equador até abaixo do Trópico de Capricórnio, engloba uma diversidade muito grande de climas, e por consequência abriga uma grande variedade de cultivos, produzindo não somente grãos, como soja e milho, como também cana-de-açúcar, café, algodão e frutas (EMBRAPA, 2018).

Dentre todos os cultivos existentes no Brasil, está o cultivo de maçãs, que ocorre majoritariamente no sul do país, gerando produtividade anual que varia de 15 à 30 toneladas por hectare (SEBRAE, 2013). Assim como todos os demais cultivos, as macieiras também recebem aplicações tecnológicas para melhoria de produtividade, de qualidade da produção e de conforto/segurança dos agricultores. Um exemplo disso é o estudo da manipulação genética para criar uma variedade de macieiras com menor exigência de frio (FIALHO, 2019).

Um dos principais órgãos de investimento em pesquisas tecnológicas na agricultura brasileira é a EMBRAPA, que desde seu surgimento em 26 de abril de 1973 busca desenvolver a tecnologia da agricultura do país (EMBRAPA, 2021).

Com intuito de trazer maior segurança no planejamento do uso de recursos para produtores de maçãs, a Universidade Tecnológica Federal do Paraná (UTFPR), em parceria com a EMBRAPA, desenvolveu uma estrutura física embarcada em veículo agrícola (trator) que é capaz coletar imagens de pomares que possam ser utilizadas para realizar um cálculo de previsão da produção de maçãs. Tal ferramenta possibilitará, não só um melhor planejamento por parte do agricultor, como também resultará em mais dados estatísticos para estudos posteriores.

### 1.1 Objetivos geral e específicos

Este trabalho tem como objetivo utilizar ferramentas de processamento de imagens e redes neurais para realizar a contagem de maçãs a partir de imagens coletadas de pomares brasileiros. Para alcançar tal objetivo, os seguintes objetivos específicos devem ser atingidos:

- Montar estrutura para captura de imagens;
- Coletar imagens dos pomares que sejam adequadas para a criação de imagens panorâmicas;
- Processar as imagens obtidas para criação de imagens panorâmicas;
- Treinar e utilizar Rede Neural Artificial (RNA) para contagem de maçãs nas imagens geradas.

## 2 REFERENCIAL TEÓRICO

Neste capítulo serão apresentadas informações científicas sobre temas que tenham relevância para a realização e compreensão deste trabalho.

Por início, será apresentada uma descrição sobre a maçã, suas características científicas, sua relevância na agricultura global e nacional, além de uma descrição técnica sobre a maneira de cultivo aplicada pela EMBRAPA.

Ainda neste tópico será abordado o processamento de imagens, bem como técnicas de processamento utilizadas para a realização do trabalho. Além disso, será descrito o que é uma RNA e tecnologias necessárias para a realização da detecção de objetos.

Por fim, serão apresentados trabalhos correlatos a fim de demonstrar relevância, viabilidade e técnicas usadas do tema proposto.

### 2.1 Maçãs

A maçã cultivada é um complexo híbrido interespecífico geralmente representado pela espécie *Malus domestica* Borkh, que pertence à sub-família Maloideae da família Rosaceae, que reúne ao redor do mundo em torno de 25 à 30 espécies reconhecidas (BYRNE; BADENES, 2012).

A sua origem, de acordo com análises genéticas, ocorreu na Ásia junto com uma grande contribuição secundária de espécies silvestres de maçã da Europa (CORNILLE *et al.*, 2014). Dada sua origem em regiões de clima temperado, a árvore de maçã é adaptada à ambientes mais frios (PETRI *et al.*, 2019), sendo necessário para as espécies mais comuns temperaturas abaixo de 7°C para desencadeamento da floração (BYRNE; BADENES, 2012).

De acordo com Cornille *et al.* (2014), as maçãs domésticas são de grande importância para a regiões de clima temperado por representarem uma parcela importante da produtividade de frutas dessas regiões, além de estarem presentes na cultura e culinária de diversos povos ao redor do mundo, já que há séculos os humanos exploram as suas características, antes mesmo delas serem domesticadas.

No ano de 2018 a maçã foi a segunda fruta mais produzida do ano no mundo, acumulando um total de 86 milhões de toneladas, perdendo apenas para a produtividade de bananas, que alcançou a marca de 116 milhões de toneladas (VASYLIEVA; JAMES, 2021). Desde o ano 2000 até 2018, a produção de maçã cresceu mais de 50%, podendo ser encontrada em 96 países, dentre os quais os cinco maiores produtores em 2019 eram: China, Estados Unidos da América, Polônia, Turquia e Irã (VASYLIEVA; JAMES, 2021).

As primeiras ocorrências de produção de maçãs registradas no Brasil ocorreram na cidade de Valinhos, estado de São Paulo, no ano de 1926, mas a expansão dos pomares como produção comercial somente ocorreu na década de 70, quando foi percebido o potencial regional para suprir uma demanda crescente (PETRI *et al.*, 2019).

Desde o início, a produção de maçãs no Brasil sofreu com obstáculos para expandir-se. O fato da fruta ser adaptada à climas temperados faz diminuir as regiões capazes de produzi-la diante do clima tropical dominante no país. Além disso, problemas relacionados à suscetibilidade à doenças também levantaram preocupações e custos extras para a manutenção das safras. Diante de tais problemas, desde a década de 70, pesquisas começaram a ser feitas sobre o cruzamento de espécies, buscando criar mutações de maçãs Gala e Fuji que fossem menos exigentes ao frio e mais resistentes à doenças (DENARDI; KVITSCHAL; HAWERROTH, 2019).

Por razões climáticas, a produção brasileira de maçãs encontra-se concentrada na região sul do país, que em 2004 abrigava 98% de toda a produção nacional. Em números de toneladas, a produção obteve grandes aumentos ao decorrer da história, começando com uma produção de 1528 toneladas em 1974 e alcançando a marca de 800 mil toneladas no ano 2000 (MELLO, 2004). Na safra de 2011/2012 a produção brasileira alcançou um novo pico com 1,4 milhões de toneladas produzidas (BLANKE, 2013). Já na safra de 2019/2020, de acordo com Mendes e Barbieri (2020), foram colhidas apenas 939 mil toneladas da fruta.

De acordo com Anzolin *et al.* (2021), a produção de maçãs realizada em campos da EMBRAPA se dá em pomares nos quais as árvores ficam organizadas em fileiras. Cada árvore possui em média uma altura de quatro metros, sendo 3,5 metros de copa e meio metro de tronco. Além disso, as fileiras de árvores possuem uma distância entre si de quatro metros e uma distância entre cada árvore da fileira que varia de 0,9 até 1,7 metros, permitindo circulação de pessoas e maquinários para o manuseio da cultura. A Figura 1 mostra um pomar de maçãs organizado em fileiras de árvores.

**Figura 1 – Pomar de maçãs**



Fonte: Adaptado de Sousa (2021).



## 2.2 Processamento de imagens

A utilização do processamento de imagens é fundamental para a realização deste trabalho, sendo um dos tópicos fundamentais abordados aqui. A seguir é discutido sobre a definição, aplicação e ferramentas de processamento de imagem.

De acordo com Gonzalez e Woods (2018), uma imagem digital pode ser definida como uma função  $f(x, y)$  em que  $x$  e  $y$  representam coordenadas de um plano e  $f$  a intensidade de cinza dadas as posições de  $x$  e  $y$ . As imagens digitais são planos de pontos finitos, nos quais os valores de  $f$ ,  $x$  e  $y$  são números discretos, e uma posição de  $x$  e  $y$  com um valor de intensidade  $f$  é chamado de *pixel*.

Ainda de acordo com Gonzalez e Woods (2018), não há consenso científico na definição de uma fronteira entre o que é processamento de imagem e o que é visão computacional. Há autores que definem o processamento de imagem como apenas os processos que possuem tanto entrada quanto saída de dados em forma de imagem, ao passo que a visão computacional tenha imagens como dados de entrada, mas outros tipo de dados de saída, como descrições, ambientes de três dimensões, etc. Os autores do livro discordam de tal definição, visto que uma tarefa de processamento de imagens simples como calcular a intensidade média dos pixels de uma imagem teria uma saída numérica.

Eles dividem os processos de imagens em três níveis: processos de baixo, médio e alto nível. Os processos de baixo nível são os mais simples, sendo processos responsáveis pelo pré-processamento das imagens, reduzindo ruído, aumentando o contraste, etc. Os processos de nível médio são aqueles que buscam retirar informações da imagem, encontrando bordas, contornos, etc. Por último, os processos de alto nível buscam encontrar “sentido” nas imagens, por exemplo, reconhecendo objetos e os descrevendo.

Diante do exposto, os autores consideram como processamento de imagem não apenas os processos que possuam imagens como dados de entrada e saída, mas também aqueles que extraem delas atributos e reconheçam nelas objetos individuais.

Uma das técnicas de processamento de imagens é a filtragem de cor, capaz de identificar e extrair de uma imagem regiões de pixels pertencentes à uma faixa de cor (RASHID *et al.*, 2019). Já entre os processos de manipulação de imagens, existe o *stitching*, tendo aplicação prática como a geração de mapas por fotos de satélite, por exemplo (SZELISKI, 2011).

O termo *stitching* pode ser traduzido do inglês como “costura” e é utilizado para descrever algoritmos que realizam o alinhamento e colagem de imagens através da Sobreposição (sob.) de áreas iguais, permitindo, dessa forma, a criação imagens panorâmicas (ou mosaicos de imagens) em alta resolução (WANG; YANG, 2020).

Para o processo de *stitching* é fundamental a realização da sob. das imagens de maneira correta, e para realizar tal tarefa diversos métodos foram desenvolvidos. Os primeiros e mais simples são os métodos baseados em área, que podem ser resumidos como técnicas de comparação de pixels, que buscam encontrar nas imagens de entrada pontos ou áreas seme-

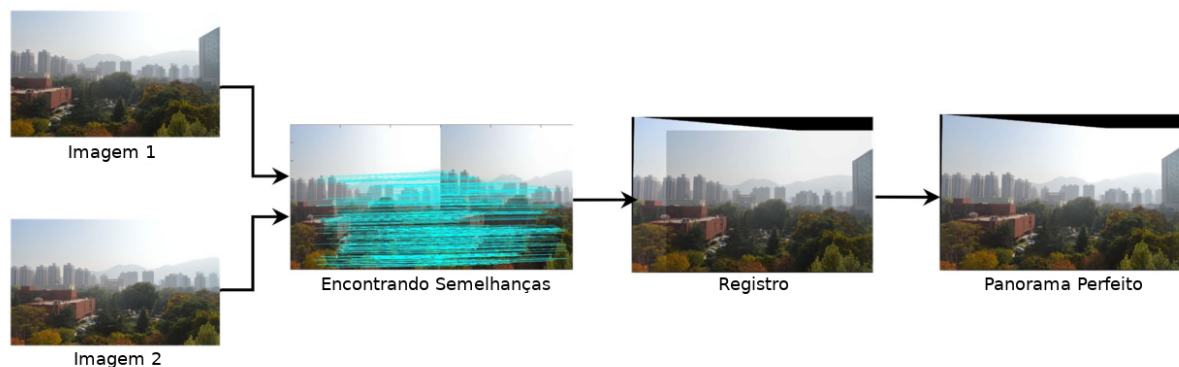
lhantes para então sobrepô-las, formando o mosaico. Esses algoritmos, embora mais simples de implementar, necessitam de uma alta taxa de sob. entre as imagens de entrada, além de terem alto custo computacional (GHOSH; KAABOUCH, 2016).

Outros tipos de algoritmo que são mais utilizados para criar os mosaicos de imagens são conhecidos como *feature-based*, ou, em português, “baseados em características”. Esses algoritmos, ao invés de buscarem pontos ou regiões semelhantes entre imagens, buscam encontrar elementos que se assemelhem, como cantos, bordas e cores. Dentre tais algoritmos, dois de destaque são o detector de bordas *Features from Accelerated Segment Test* (FAST) e o detector de características *Scale Invariant Feature Transform* (SIFT). O algoritmo FAST é conhecido por ser mais rápido que a maioria dos demais algoritmos, porém é menos robusto, sofrendo com imagens que possuam uma maior variação entre si. Por outro lado, o algoritmo SIFT, embora computacionalmente exigente, é robusto e capaz de realizar a colagem até de imagens com pequena variação de perspectiva (GHOSH; KAABOUCH, 2016).

Existem duas maneiras de tirar fotos para a geração da imagem panorâmica. A primeira delas e mais comum é fixando a posição da câmera girando-a em torno de si mesma. A outra maneira é fixando a angulação da câmera e movendo-a horizontalmente para a lateral (WANG; YANG, 2020). Um problema enfrentado na criação de mosaicos de imagens é o efeito de paralaxe, que de acordo com Ghosh e Kaabouch (2016) pode ser evitado se a obtenção das imagens for feita sem o translado da câmera ou se o objeto fotografado for relativamente planar.

Na Figura 2 pode ser visto um exemplo de *stitching* feito a partir da junção de duas imagens. Após serem reconhecidas as regiões em comum, as imagens são sobrepostas na etapa de registro utilizando de um sistema único de coordenadas, e então corrigidas para formar uma única imagem panorâmica sem rastros de borda das imagens originais.

**Figura 2 – Exemplo de *stitching* de imagens**



**Fonte: Adaptado de Wang e Yang (2020).**

### 2.3 Inteligência Artificial e Aprendizado de Máquina

De acordo com Heaton (2013), a Inteligência Artificial (IA) é uma técnica de algoritmos de computadores que aborda a solução de problemas de maneira semelhante à um cérebro humano, porém, sem necessariamente emular um cérebro em todas suas funções.

O Aprendizado de Máquina (AM), por sua vez, é um campo da Ciência da Computação que estuda a capacidade de um computador em realizar uma tarefa sem ter sido explicitamente programado para tal e é considerado o primeiro grande passo para a criação de inteligências artificiais (TAYLOR, 2017). De acordo com Shalev-Shwartz e Ben-David (2014), AM é a detecção automática de padrões significativos em conjuntos de dados, ou seja, é a capacidade de um computador encontrar padrões lógicos dada uma entrada de dados, sem a necessidade de que um humano explicitamente ensine à máquina tais padrões.

Esse campo pode ser considerado interdisciplinar, envolvendo várias outras áreas da ciência. Entretanto, uma das áreas com a que o AM mais é relacionado é a IA. O AM pode ser visto como um braço da IA, visto que a capacidade de aprender baseado em experiência é uma característica do cérebro humano. Apesar disso é importante ressaltar que o AM não busca replicar a capacidade de um cérebro humano, mas de explorar recursos computacionais para reconhecer padrões e realizar tarefas que são complexas para o ser humano (SHALEV-SHWARTZ; BEN-DAVID, 2014).

Existem diversas aplicações do AM presentes nas atividades diárias das pessoas e muitas vezes podem passar despercebidas. Entre as aplicações mais conhecidas podem ser citados os algoritmos para anúncios do *YouTube*, que tentam exibir anúncios mais relevantes, e os algoritmos de filtragem de e-mail spam, utilizados por provedores de e-mail como o *Yahoo!* e o *Gmail* (TAYLOR, 2017).

Para que tudo isso seja possível, o AM trabalha com algoritmos de treinamento, responsáveis por ensinar à máquina as regras que ela deve seguir para realizar previsões sobre novas entradas de dados. Entretanto, não há um algoritmo genérico para a aprendizagem, pois cada situação exige características diferentes no algoritmo para um bom funcionamento, e tais características tornam possível separá-los de acordo com quatro parâmetros descritos da seguinte maneira de acordo com Shalev-Shwartz e Ben-David (2014):

- Aprendizagem supervisionada e não supervisionada: esse parâmetro diz respeito à quantidade de informações dispostas para o treinamento da máquina. Na aprendizagem supervisionada o grupo de dados para treinamento traz junto de si qual deve ser o resultado obtido pela máquina, permitindo que ela identifique a lógica que resolva todos os dados de entrada de maneira correta. Já na aprendizagem não supervisionada, os dados para treinamento não possuem nenhuma informação sobre a resposta correta. Dessa maneira a máquina deve comparar os dados entre si para determinar características que os diferenciem uns dos outros.

- **Aprendizes passivos e ativos:** os algoritmos também podem ser classificados de acordo com sua atuação no ambiente de aprendizagem. Os algoritmos passivos são aqueles que não interagem durante o período de aprendizagem, apenas observam, captando as informações oferecidas pela fonte de dados. O algoritmos ativos, por outro lado, podem interagir com o ambiente durante esse período, realizando experimentos ou fazendo perguntas, por exemplo.
- **Utilidade do professor:** diz respeito à maneira que os dados são gerados para a aprendizagem. Os dados utilizados no treinamento podem ser tratados de maneira à exaltar as características mais importantes para que a máquina aprenda, descrevendo assim um professor útil. Também existe o professor passivo, que simplesmente oferece a informação sem fazer peso de importância. Por último há o professor adversário, que busca dificultar o processo de aprendizagem. Como exemplo do último caso, é possível citar os algoritmos para filtro de e-mail spam que precisam aprender a reconhecer como spam mensagens de e-mail que se esforçam para não se parecerem com um spam.
- **Online e aprendizado por lote:** é a diferenciação de quando o algoritmo deve começar a dar respostas. No aprendizado *online*, o aprendiz vai tomando decisões ao mesmo tempo que aprende, fazendo com que cometa diversos erros antes de tornar-se capaz de realizar a atividade com boa acurácia. Já no aprendizado por lote, o algoritmo só deve tomar ação depois de analisar todo o conjunto de informações fornecidas à ele.

É possível, também, separar um algoritmo de AM baseado no tipo de problema que ele resolve. Goodfellow, Bengio e Courville (2016) citam alguns desses tipos de problema, dentre os quais estão a classificação e a regressão. Na tarefa de classificação, o algoritmo é responsável por classificar a entrada, ou seja, dizer à qual classe, dado um conjunto de classes possíveis, o conjunto de dados de entrada pertence. A regressão, por outro lado, tem a característica de prever um resultado numérico baseado nos valores de entrada.

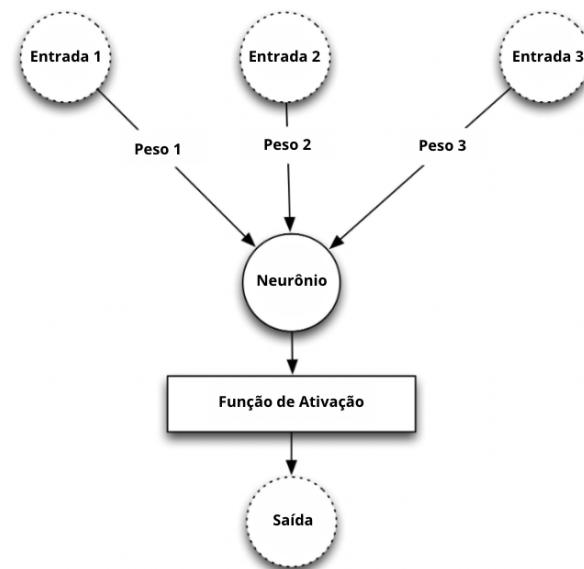
## 2.4 Redes Neurais Artificiais

De acordo com Shalev-Shwartz e Ben-David (2014), RNAs são modelos de IA inspirados nas redes neurais biológicas. São também um dos algoritmos mais populares no mundo de AM e têm o objetivo de resolver problemas de maneira semelhante à um cérebro humano (TAYLOR, 2017).

### 2.4.1 Neurônios

As RNAs são compostas por pequenas unidades de computação (neurônios) interligadas, sendo cada unidade responsável por processar uma informação simples recebida de outros neurônios e enviar o resultado como sinal para outros neurônios, produzindo, através da interação entre muitos neurônios, uma informação complexa (SHALEV-SHWARTZ; BEN-DAVID, 2014). O processamento da informação que ocorre em cada neurônio é feito por uma função de ativação que recebe como parâmetro a somatória de todas as saídas dos neurônios ligados à ele multiplicadas por um respectivo peso (HEATON, 2015) e somadas à um *bias*, que é um peso do próprio neurônio (VASILEV *et al.*, 2019). Uma representação de um neurônio pode ser observada na Figura 3.

**Figura 3 – Representação de um neurônio artificial**



Fonte: Adaptado de Heaton (2015).

A saída  $y$  de um neurônio está representada na Equação (1), na qual  $f$  representa a função de ativação,  $x_i$  uma entrada do neurônio,  $w_i$  o peso da entrada e  $b$  o *bias* do neurônio.

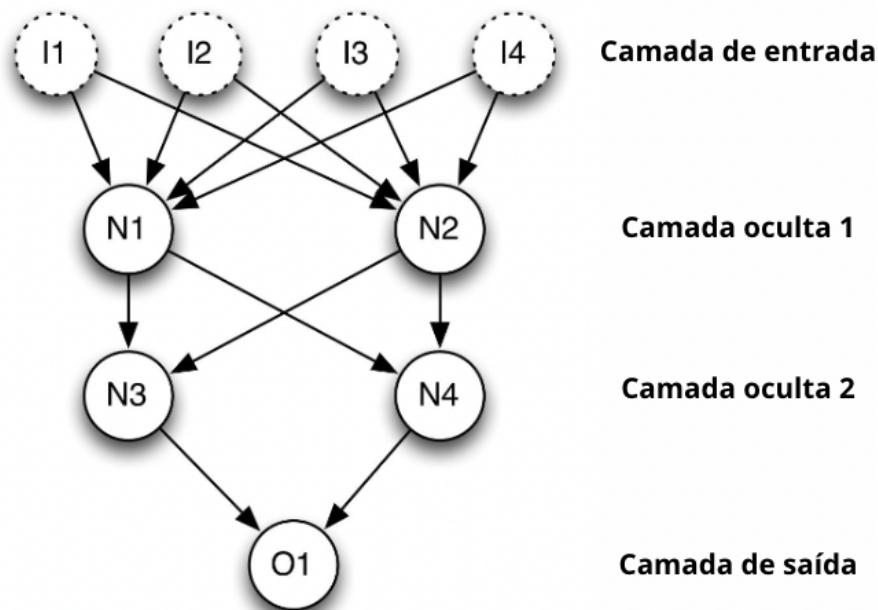
$$y = f\left(\sum_i x_i w_i + b\right) \quad (1)$$

Estes tipos de IA abordam a solução de problemas de uma maneira bastante diferente em comparação a programas de computador habituais. Em programas normais existe uma sequência lógica de instruções que devem ser seguidas pelo computador. Já em uma RNA são simulados neurônios interligados que trabalham paralelamente para a solução do problema (TAYLOR, 2017).

### 2.4.2 Redes multicamadas

Uma descrição visual de uma RNA pode ser dada por um grafo, no qual cada nó é um neurônio e cada aresta uma ligação entre eles (SHALEV-SHWARTZ; BEN-DAVID, 2014). Uma representação visual pode ser vista na Figura 4. Nela é possível observar os neurônios representados pelos círculos e as ligações entre eles representadas pelas setas.

**Figura 4 – Rede Neural Artificial Multicamada**



Fonte: Adaptado de Heaton (2015).

As RNAs possuem três camadas ou regiões em sua estrutura: a camada de entrada, a camada oculta e a camada de saída. A camada de entrada é por onde a rede recebe as informações com as quais vai trabalhar. A camada oculta pode conter várias subcamadas de neurônios e é o caminho pelo qual a informação trafega até chegar na camada de saída, que é onde o resultado da operação é exposto (TAYLOR, 2017).

De acordo com Taylor (2017), as RNAs podem ser construídas de duas maneiras: *feed-forward* e *feedback*. Na construção *feedforward* a informação trafega apenas para a frente (da entrada para a saída). Esse tipo de rede é amplamente utilizado no reconhecimento de padrões. A construção *feedback*, por outro lado, permite que a informação trafegue em qualquer sentido, tornando possível a existência de laços de repetição. Essa construção se assemelha mais com a maneira que o cérebro humano funciona, entretanto é menos utilizada por ser mais complexa e possuir algoritmos de aprendizagem menos eficazes.

### 2.4.3 Treinamento

O treinamento de uma RNA é o processo responsável por ajustar as medidas de peso e *bias* da rede de maneira que o resultado da rede se aproxime com o resultado esperado. Para RNAs do tipo *feedforward* um dos algoritmos de treinamento existentes é o *backpropagation* (HEATON, 2015).

De acordo com Aggarwal (2018), o *backpropagation* pode ser dividido em duas fases: *forward* (para frente) e *backward* (para trás). Na fase *forward* a RNA é alimentada com dados de treino, fazendo com que o modelo calcule uma resposta que é então comparada com a resposta desejada. Após isso, na fase *backward* os pesos da rede são corrigidos de trás para frente, da última camada em direção a primeira, pelo gradiente de cada peso de acordo com o erro encontrado na primeira fase.

### 2.4.4 Métricas de desempenho

Com a finalidade de validar um modelo de RNA é necessário estabelecer métricas que descrevam o desempenho dele. Tais métricas podem ser de acurácia, tempo de treinamento, requisitos de armazenamento, entre outros. Entre as métricas de acurácia podem ser citadas a precisão, que representa a proporção de classificações corretas de todas as classificações feitas pelo modelo, e a cobertura, que representa a proporção entre classificações corretas e o total de classes positivas existentes (FACELI *et al.*, 2011).

Outras métricas são a precisão média, ou *Average Precision* (AP), que é a média da precisão de classificação de uma categoria em diferentes entradas, e a *Mean Average Precision* (mAP), que é a média das médias de precisão de diferentes categorias. A mAP é uma medida que demonstra o poder de um modelo em identificar as classes nos dados de entrada (XIAO *et al.*, 2020).

### 2.4.5 Redes Neurais Convolucionais

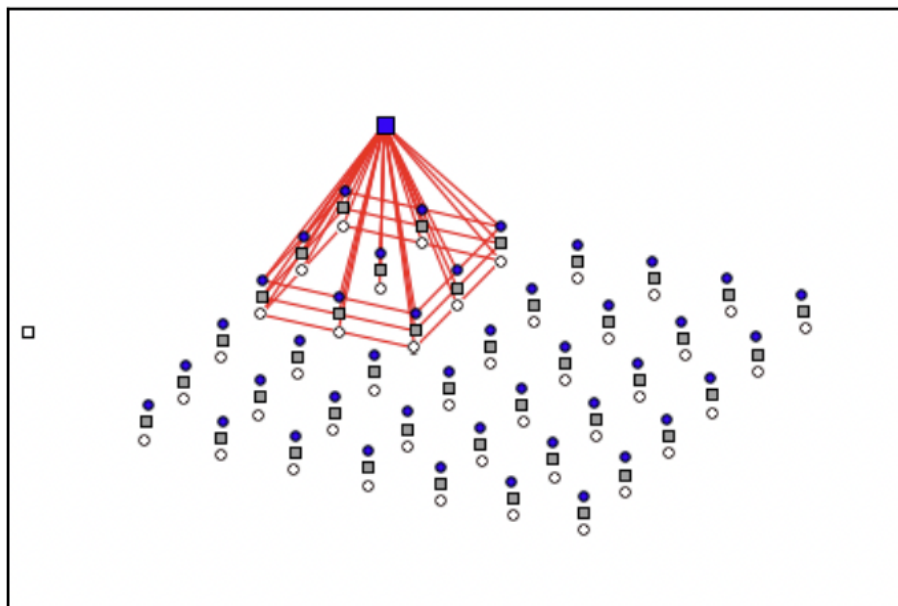
Uma Rede Neural Convolucional, ou *Convolutional Neural Network* (CNN), é uma RNA do tipo *feedforward* projetada para trabalhar com entradas de dados em formato de grade em que a disposição dos dados seja um fator relevante para sua análise. Dessa forma, essas redes apresentam bons resultados quando se trata da análise de imagens, pois uma imagem é uma grade de pixels em que a disposição desses pixels é responsável por dar sentido à ela, e, por isso, são amplamente utilizadas para realizar a detecção de objetos em imagens e vídeos (AGGARWAL, 2018).

Uma característica muito importante da CNN e que a diferencia de outras RNAs é a camada de convolução, que consiste em um grupo de filtros, também conhecidos como *ker-*

*nels*, que são aplicados por todas as áreas dos dados de entrada (VASILEV *et al.*, 2019). O objetivo dessa camada é buscar reconhecer padrões de ocorrência de dados em diferentes partes geográficas da entrada. A convolução geralmente busca encontrar pequenos padrões que vão sendo unidos até formar uma informação complexa. Um exemplo disso é o processo de identificar número escritos à mão: a primeira operação de convolução pode ser responsável por encontrar bordas verticais e horizontais na imagem, que unidas formassem formas mais complexas, como linhas ou círculos, que novamente unidas formassem números completos, permitindo que a rede classifique a entrada de acordo com qual número as formas detectadas mais se assemelham (AGGARWAL, 2018).

De acordo com VASILEV *et al.* (2019), um filtro é definido por um conjunto de pesos configuráveis que pode ter qualquer tamanho:  $2 \times 2$ ,  $3 \times 3$ ,  $4 \times 4$ , entre outros. Esse filtro pode ser aplicado em qualquer camada da CNN e possui pesos compartilhados, ou seja, os pesos utilizados na aplicação do filtro são os mesmos para todas as regiões dos dados de entrada. Na Figura 5 está representada a aplicação de um filtro de tamanho  $3 \times 3$  à uma das regiões de um dado de entrada com três camadas, como seria para uma imagem com três camadas de cores: vermelho, verde e azul.

**Figura 5 – Aplicação de filtro em camada de convolução em uma das regiões dos dados de entrada**

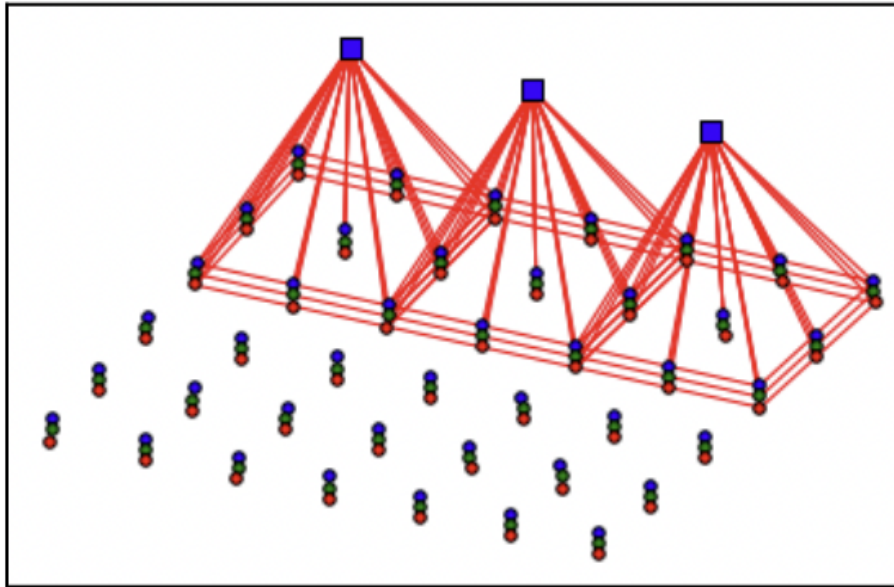


Fonte: VASILEV *et al.* (2019).

Esse filtro, para que seja aplicado à toda a entrada de dados, pode ser “deslizado” de uma em uma posição vertical e horizontal, gerando uma saída que tenha um tamanho de altura e largura menor em duas unidades referentes às bordas horizontais e verticais. Porém é possível “deslizar” o filtro um número maior de posições, diminuindo ainda mais o tamanho dos dados de saída da convolução. Essa técnica é conhecida como *stride* e pode ser observado na Figura 6, que representa a aplicação de um filtro com *stride* de 2 (VASILEV *et al.*, 2019).



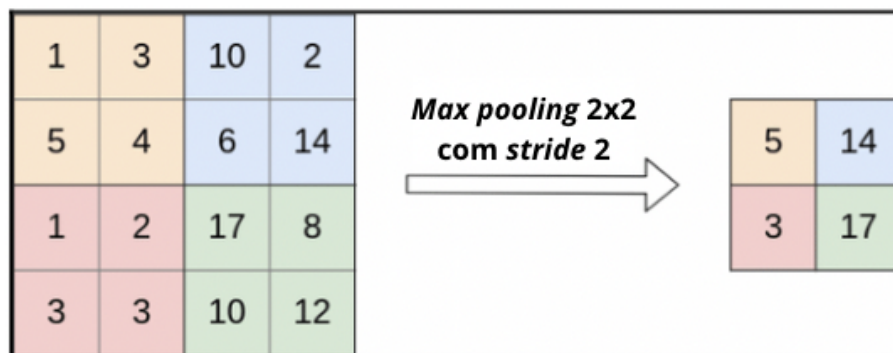
Figura 6 – Aplicação de filtro em camada de convolução com *stride* 2



Fonte: VASILEV *et al.* (2019).

Outra técnica utilizada em CNNs é o *pooling*, que separa os dados de entrada em diferentes regiões, e através de uma operação busca resumir os valores daquela região em um único valor. Existem diferentes tipos de *pooling*, porém o mais famoso deles é o *max pooling*, que consiste em selecionar o maior valor de cada região para ser o seu “representante” (VASILEV *et al.*, 2019). A operação de *max pooling* em regiões de tamanho  $2 \times 2$  está representada na Figura 7.

Figura 7 – Aplicação de *max pooling* com *stride* 2

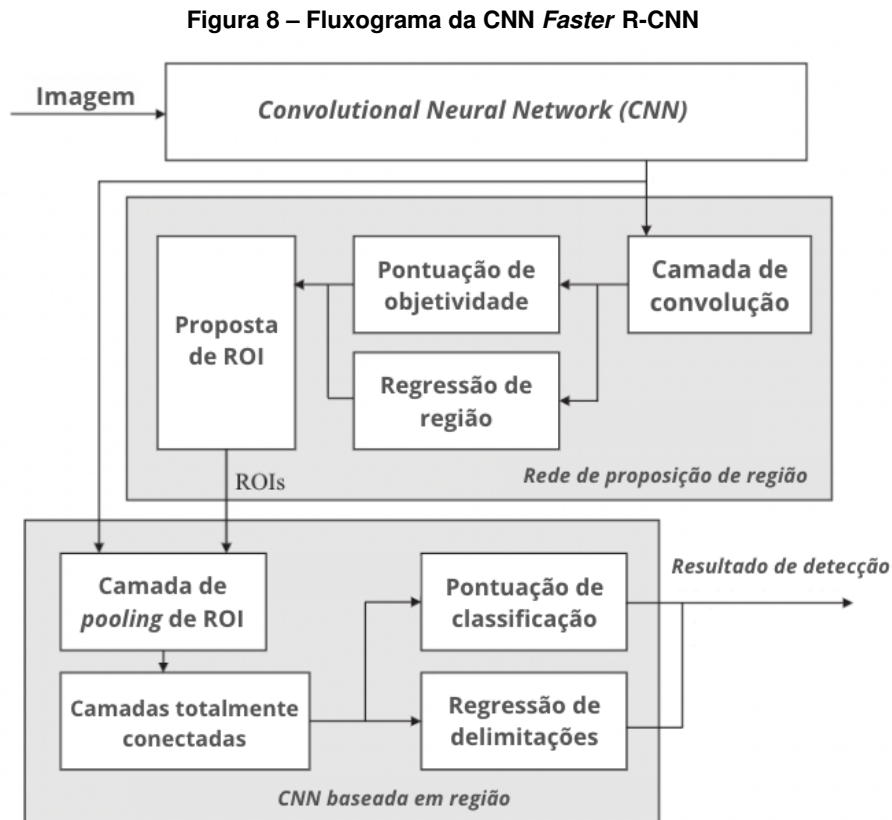


Fonte: Adaptado de VASILEV *et al.* (2019).

#### 2.4.6 Modelos

Uma das arquiteturas de CNN mais famosas é a *Visual Geometry Group* (VGG), introduzida em 2014 sendo comumente utilizada para comparação com arquiteturas mais modernas. Redes dessa arquitetura geralmente possuem camadas alternadas de convolução e de *pooling*, sendo as duas redes mais famosas a VGG16 e a VGG19 (VASILEV *et al.*, 2019).

Um dos estados da arte na área de detecção de objetos com uso de CNNs é a RNA *Faster R-CNN*. De acordo com Lee, Kim e Oh (2016), essa RNA pode ser dividida em três etapas: uma rede convolucional responsável por identificar características na imagem original; uma rede de proposição de região, responsável por identificar regiões de interesse, ou *Region of Interest* (ROI), que possam conter objetos; e uma rede de classificação, responsável por classificar cada região encontrada. A Figura 8 mostra um fluxograma do funcionamento da CNN *Faster R-CNN*.



Fonte: Adaptado de Lee, Kim e Oh (2016).

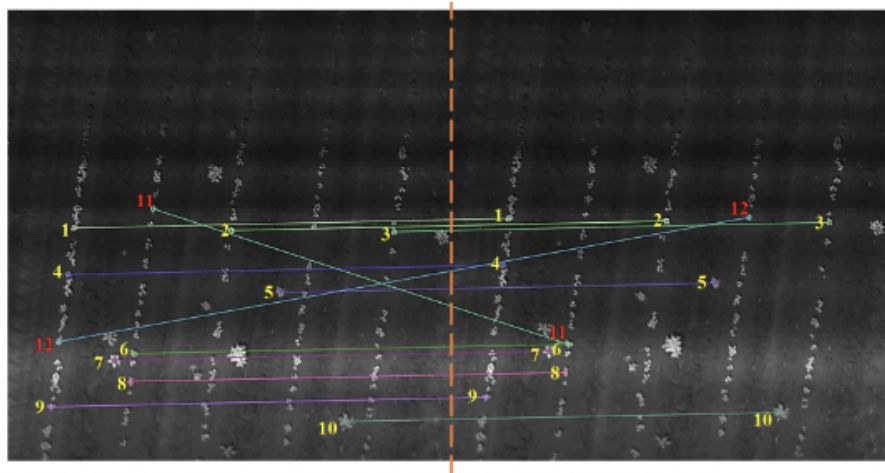
Outro estado da arte é a CNN *You Only Look Once* (YOLO)v3, que é a terceira versão da família de modelos de CNN de detecção de objetos YOLO. Essa RNA se diferencia da *Faster R-CNN* por tornar o processo de detecção de objetos em um problema de regressão ao invés de um problema de classificação. Além disso, a YOLO não necessita da identificação de regiões de interesse, o que faz com que ela seja muito mais rápida (CHENG, 2020).

## 2.5 Trabalhos correlatos

O *stitching* de imagens possui aplicações práticas presentes no dia-a-dia das pessoas, como ocorre nos mapas gerados por satélites. Porém, essa técnica também se mostra útil como uma ferramenta a produtores rurais, podendo auxiliar em diferentes atividades relacionadas ao campo. Um exemplo disso pode ser visto no trabalho de Feng *et al.* (2020), no qual o uso dessa

tecnologia permite a elaboração de um mosaico de imagens da lavoura de algodão que facilita analisar a emergência das plantas em duas semanas após o plantio. Para gerar esse mosaico, foi utilizado um Veículo Aéreo Não Tripulado (VANT) equipado com câmeras e voando a uma altitude de 50 metros acima do área de plantio. Com as imagens obtidas pelo VANT, foram aplicados algoritmos de alinhamento e de *stitching* customizados, gerando mosaicos de imagem que permitiram a contagem de mudas com uma acurácia de 84,1%, o que demonstrou a viabilidade do método empregado, além da utilidade dos métodos de *stitching*. Na Figura 9 é possível observar a identificação de pontos chaves entre fotos coletadas pelo VANT, etapa necessária para a realização do *stitching*.

**Figura 9 – Detecção de pontos chaves entre duas fotos obtidas pelo VANT**



Fonte: Feng *et al.* (2020).

Utilizar o processamento de imagens junto com visão computacional para realizar detecção de maçãs em pomares não é uma atividade inédita deste trabalho, como pode ser visto em Gené-Mola *et al.* (2019) e Tian *et al.* (2019).

No trabalho de Gené-Mola *et al.* (2019) a CNN para reconhecimento de objetos *Faster R-CNN* com RNA VGG16 foi modificada para a realização de reconhecimento de maçãs em imagens com cinco canais: três de cores (RGB), um de profundidade (D) e outro de faixa corrigida de intensidade (S). As imagens foram obtidas em um pomar comercial da Espanha através de duas câmeras RGB-D formando o banco de dados KFuji RGB-DS. O resultado do experimento foi que a detecção de maçãs em uma imagem com cinco canais teve os melhores resultados, alcançando uma pontuação F1 de 0,898 e uma precisão média de 94,8%, demonstrando a validade da utilização dessas imagens para a detecção de objetos com a rede R-CNN com VGG16. Na Figura 10 é possível observar a detecção de maçãs realizada por esse modelo proposto.

Em Tian *et al.* (2019), foi realizada uma análise do desempenho do modelo de detecção YOLO V3 melhorado com uma incorporação de uma RNA *DenseNet* para a detecção de maçãs em pomares em diferentes estágios de crescimento. Para realizar o trabalho, a CNN proposta foi treinada e testada com um *dataset* de 4800 imagens de um pomar, que foram obtidas por

**Figura 10 – Detecção de maçãs utilizando o modelo treinado por Gené-Mola *et al.* (2019)**



Fonte: Gené-Mola *et al.* (2019).

meio de um processo de aumento de dados, conhecido como *data augmentation*, realizado em um conjunto de 480 imagens originais. Os processos utilizados para aumentar o *dataset* foram: giro e espelhamento da imagem, alteração da cor, alteração do brilho e aplicação de filtro *blur*. Tendo feito o treinamento, a RNA foi testada e os resultados foram comparados com o de outros modelos, como o YOLO-V3 puro e a *Faster R-CNN* com rede VGG16, demonstrando uma maior acurácia e capacidade de reconhecer frutas parcialmente cobertas por folhas, galhos ou outras frutas em comparação com os outros modelos. Além disso, é destacado o tempo médio utilizado pelo modelo YOLOV3-dense para reconhecer as frutas que foi calculado em 0,304 segundos, enquanto a *Faster R-CNN* com rede VGG16, como melhor concorrente da YOLOV3-dense em relação à qualidade dos resultados, fez média de tempo de 2,42 segundos, somando um diferença de tempo média maior que dois segundos. Na Figura 11 é possível observar a detecção de maçãs realizada pelo modelo YOLOV3-dense.

## 2.6 Considerações finais

Neste capítulo foram apresentados conhecimentos gerais acerca das áreas envolvidas no desenvolvimento deste trabalho. Foi apresentada uma visão geral sobre a cultura das maçãs

**Figura 11 – Detecção de maçãs utilizando o modelo treinado por Tian *et al.* (2019)**



**Fonte: Adaptado de Tian *et al.* (2019).**

no Brasil e no mundo, além de uma breve retratação do modelo de produção aplicado em campos da EMBRAPA.

Também foram trazidos conhecimentos básicos sobre processamento de imagens e sobre RNAs, de maneira a introduzir conceitos utilizados no esforço de atingir o objetivo geral proposto, como a ferramenta de *stitching* e as CNNs, utilizadas para a detecção de objetos.

Por último foram apresentados trabalhos de outros autores que demonstram a verossimilidade das ideias aqui propostas.

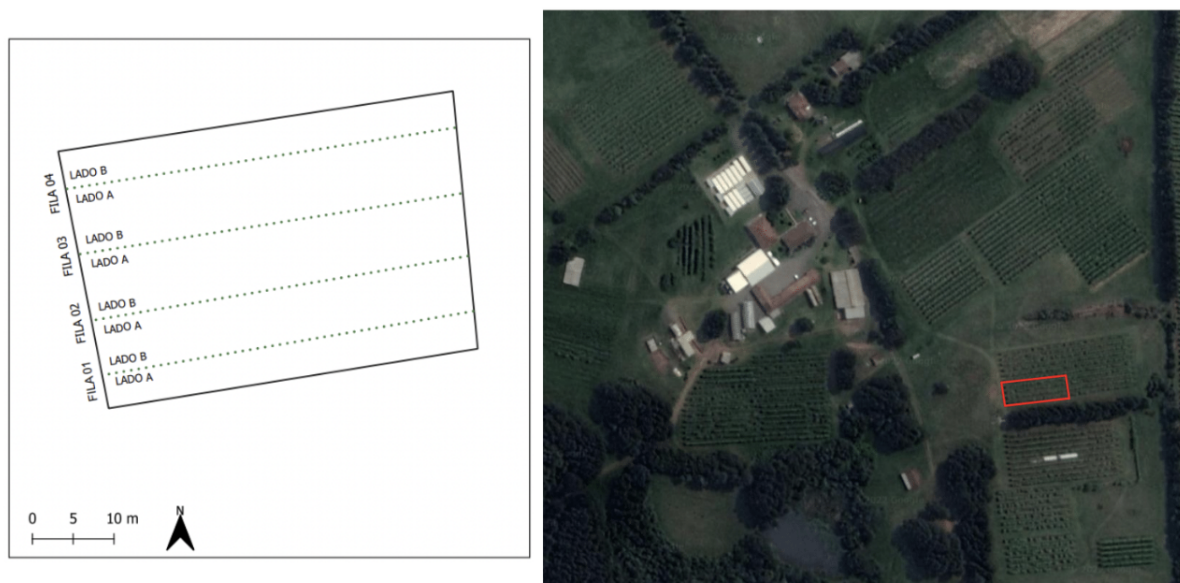
### 3 MATERIAIS E MÉTODOS

Neste capítulo serão apresentados os materiais usados para a realização do projeto, bem como os métodos empregados para que os objetivos propostos fossem alcançados.

#### 3.1 Materiais

Para a realização desta pesquisa foi utilizado o campo experimental de maçãs da EMBRAPA Uva e Vinho localizado na cidade de Vacaria, no estado do Rio Grande do Sul. Esse campo experimental é composto por diversas áreas de pomares de maçã e possui estrutura e equipamentos próprios para o manejo da cultura, como tratores, que também fazem parte dos materiais utilizados. Na Figura 12 é possível observar uma vista de cima do campo experimental, bem como um *layout* da estrutura do pomar utilizado para os experimentos.

**Figura 12 – Área de pomares no campo experimental da EMBRAPA Uva e Vinho - Vacaria/RS**



**Fonte: Autoria própria (2022).**

Os pomares do campo experimental são organizados em fileiras de macieiras, representadas pelas linhas pontilhadas no *layout* da Figura 12 e numeradas de um à quatro, por entre as quais os tratores se locomovem durante o manejo, como pode ser observado na Figura 13.

Outros materiais físicos também foram utilizados: um micro-computador RasPi, com 1GB de memória RAM e processador de quatro núcleos; três câmeras de 720p com conexão USB 2.0; um carregador portátil de 10000 mAh; duas barras de alumínio de 2,5 metros e de perfil “U”, perfuradas; segmentos de madeira de *pinus*; e ferramentas de marcenaria em geral, como serra, furadeira e parafusos.

Figura 13 – Trator entre linhas de pomar da EMBRAPA Uva e Vinho - Vacaria/RS



Fonte: Autoria própria (2022).

Três diferentes ambientes de desenvolvimento foram utilizados ao longo do projeto: um ambiente para elaboração de algoritmos de coleta de fotos, formado por um computador *Dell* modelo *Inspiron 15* com uma *Central Processing Unit* (CPU) Intel I5 de sétima geração, com *Graphics Processing Unit* (GPU) integrada, 8GB de memória RAM e armazenamento de 960GB de *Solid State Drives* (SSD); um ambiente para execução de algoritmos de *stitching*, composto por um computador *Apple Macbook Pro M1 2020* com *chip M1*, que é formado por uma CPU e uma GPU, ambas de 8 núcleos, 8 GB de memória RAM e 256 GB de armazenamento SSD; e um ambiente para treino e execução de modelos de rede neural, composto pelo plano *pro*, com custo de R\$ 58,00 mensais, do *Google Colaboratory*, ou simplesmente *Colab*, que conta com uma CPU de dois núcleos *Xeon 2,2GHz*, uma GPU *Tesla P100-PCIE 16 GB*, 25,46GB de RAM, 166,83GB de armazenamento em disco.

Em todas as etapas de execução do trabalho foi utilizada a linguagem de programação multi-paradigma e de alto nível *Python*<sup>1</sup> (PYTHON, 2021), na versão 2.7. Também foi utilizada, no decorrer do projeto, a biblioteca *OpenCV* para *Python*, que conta com diversas ferramentas de manipulação de imagens, como um algoritmo para a realização do *stitching* e métodos como o *inRange*, que permite realizar o processo de filtragem de cores (OPENCV, 2021).

No que tange a área de redes neurais, foram utilizados diferentes modelos de redes, com pesos pré-treinados com o *dataset* COCO, da quinta versão da família YOLO, ou YOLOv5<sup>2</sup>,

<sup>1</sup> <https://www.python.org/>

<sup>2</sup> <https://docs.ultralytics.com/>

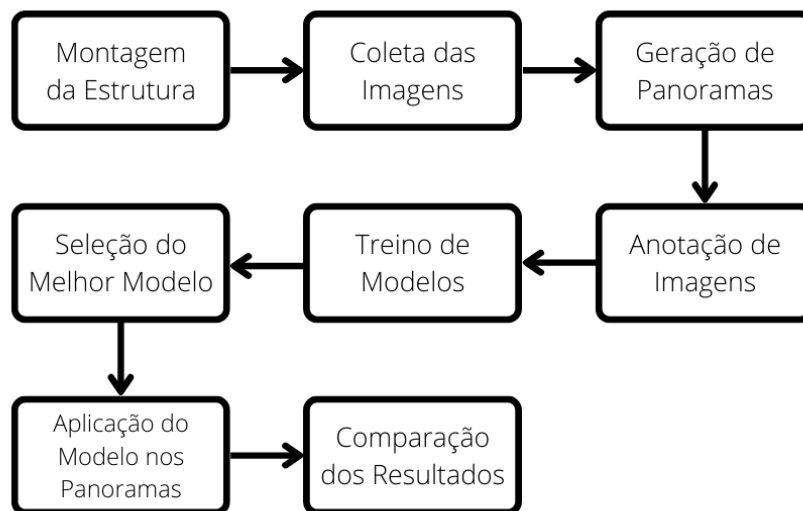
implementada em *Python*. Também foi utilizada a biblioteca SAHI<sup>3</sup>, que permite aplicar com maior efetividade modelos da YOLOv5 em imagens de larga escala através da segmentação da imagem (SAHI, 2022).

O *dataset* utilizado neste trabalho para treino dos modelos de CNN foi criado durante a realização do trabalho através da coleta de imagens do pomar da EMBRAPA e a rotulação delas foi realizada com o auxílio da ferramenta CVAT (CVAT, 2021).

### 3.2 Métodos

Nesta seção estão descritos os métodos utilizados para que os objetivos propostos fossem alcançados. Tais métodos podem ser divididos em oito etapas sequenciais: montagem da estrutura; coleta de imagens; geração de panoramas; anotação de imagens; treino de modelos de rede neural; seleção do melhor modelo; aplicação do modelo nos panoramas; e comparação dos resultados. Um diagrama de todos os processos envolvidos na realização deste trabalho pode ser visto na Figura 14.

Figura 14 – Processos da metodologia



Fonte: Autoria própria (2022).

#### 3.2.1 Montagem da estrutura

A coleta das imagens ocorreu nos dias três e quatro de Março de 2022 nos pomares do campo experimental da EMBRAPA com uso de uma ferramenta composta por uma estrutura com três câmeras embarcada em um trator que se locomoveu por entre as linhas de macieiras coletando imagens da lateral de quatro fileiras de árvores, como pode ser observado na

<sup>3</sup> <https://github.com/obss/sahi>



Figura 13. Essa ferramenta foi inicialmente proposta por Anzolin *et al.* (2021) e tem o esqueleto de sua estrutura formado pelas duas barras de alumínio, fixadas entre si e ao contrapeso do trator por segmentos de tábua de *pinus*. Além disso, o topo da estrutura de 2,5 metros é fixado à cabine do veículo por duas hastes de *pinus*, responsáveis por dar maior estabilidade à ferramenta. A estrutura montada bem como um simples diagrama mostrando os elementos principais dela podem ser vistos na Figura 15.

**Figura 15 – Estrutura embarcada em trator**

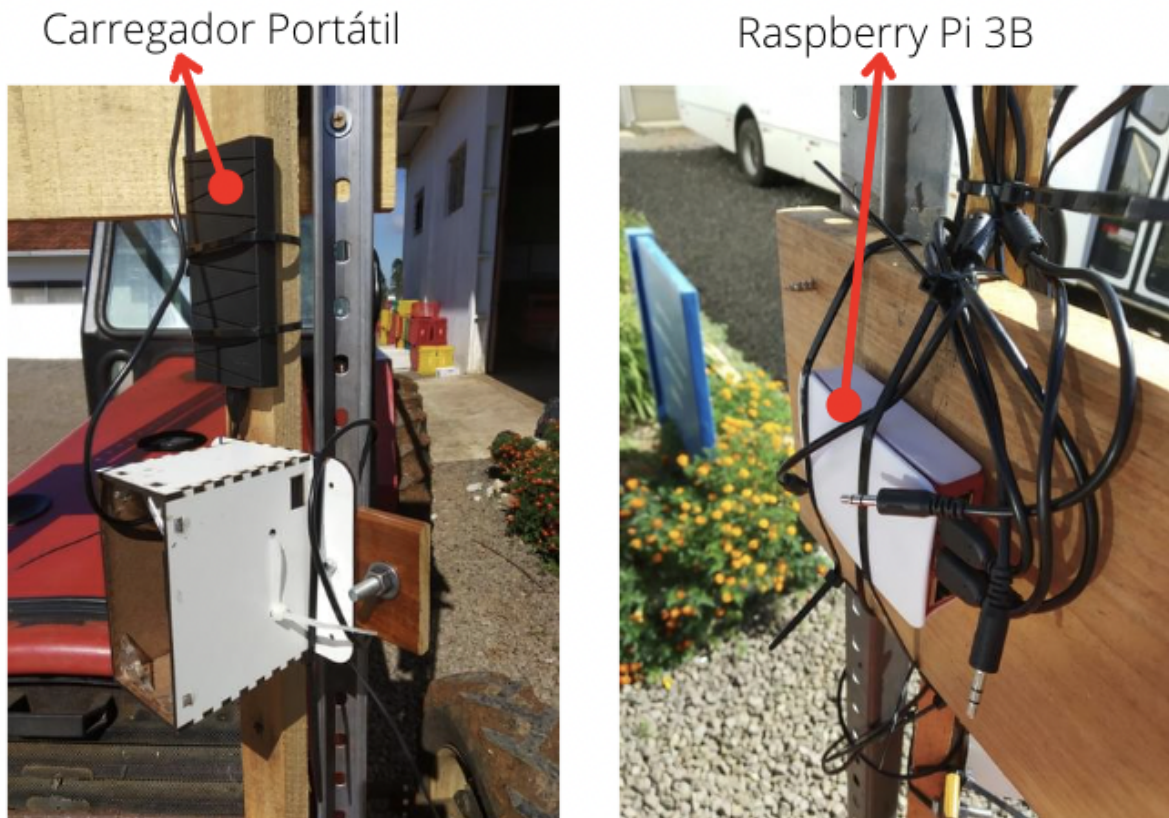


**Fonte: Autoria própria (2022).**

Como pode ser observado na Figura 15, a estrutura deu suporte às três câmeras, que foram posicionadas enfileiradas verticalmente em relação ao solo, com uma distância entre elas grande o suficiente para que as árvores fossem fotografadas desde o tronco ao topo da copa, mas também pequena o suficiente para que houvesse intersecção entre as áreas fotografadas por câmeras adjacentes. As câmeras foram dispostas ao lado oposto do trator em relação à linha de árvores que elas fotografaram com o objetivo de aumentar a área das árvores fotografadas por cada câmera. Por conta desse posicionamento invertido, elas foram giradas 45 graus para frente, fazendo com que a própria estrutura ficasse de fora de sua área de captura. Além disso, a estrutura foi elaborada de maneira a suportar seis ou mais câmeras, permitindo que as árvores aos dois lados do trator fossem fotografadas simultaneamente, porém para este experimento apenas um lado foi utilizado.

A estrutura também foi responsável por suportar o microcomputador RasPi que controlou as câmeras e armazenou as fotos, e uma bateria portátil que foi responsável por alimentar o RasPi e o conjunto de câmeras. Para controlar o microcomputador, iniciando o programa de coleta de fotos e separando e nomeando os conjuntos de imagens coletadas, ele foi conectado ao computador *Dell* através de uma conexão SSH, feita por meio de uma rede local transmitida por um dispositivo celular. Os componentes fixados à estrutura podem ser vistos com mais detalhes na Figura 16.

**Figura 16 – RasPi e carregador portátil fixados à estrutura**



**Fonte: Aatoria própria (2022).**

### 3.2.2 Coleta de imagens

No dia quatro de Março de 2022, entre as 11 e as 12 horas da manhã, o trator com a estrutura se deslocou entre as fileiras de um dos pomares coletando fotos dos dois lados de quatro fileiras de macieiras. O sistema coletou fotos de todas as câmeras ao mesmo tempo em intervalos de meio segundo, de maneira a garantir a sob. lateral das imagens tiradas por uma mesma câmera. As fotos foram separadas em conjuntos distintos para cada uma das quatro filas. As fotos de cada fila também foram separadas entre lado A e lado B e as fotos de cada lado separadas para cada câmera (1, 2 e 3).

No mesmo dia, às 16 horas, o trator novamente se deslocou entre as linhas do pomar coletando fotos das filas 2 e 3 com a finalidade de coletar imagens com uma iluminação diferente.

Ao todo, foram coletadas 2626 imagens no período da manhã e 1436 imagens no período da tarde. Na Figura 17 é possível observar exemplos de imagens coletadas no período da manhã (esquerda) e no período da tarde (direita).

Figura 17 – Imagem coletada no período da manhã (esquerda) e imagem coletada no período da tarde (direita)



Fonte: Autoria própria (2022).

### 3.2.3 Geração de panoramas

O objetivo nesta etapa foi a geração de imagens panorâmicas através de técnica de *stitching*. A principal razão para a criação desse tipo de imagem é evitar que maçãs que apareçam em mais de uma imagem sejam contadas mais de uma vez.

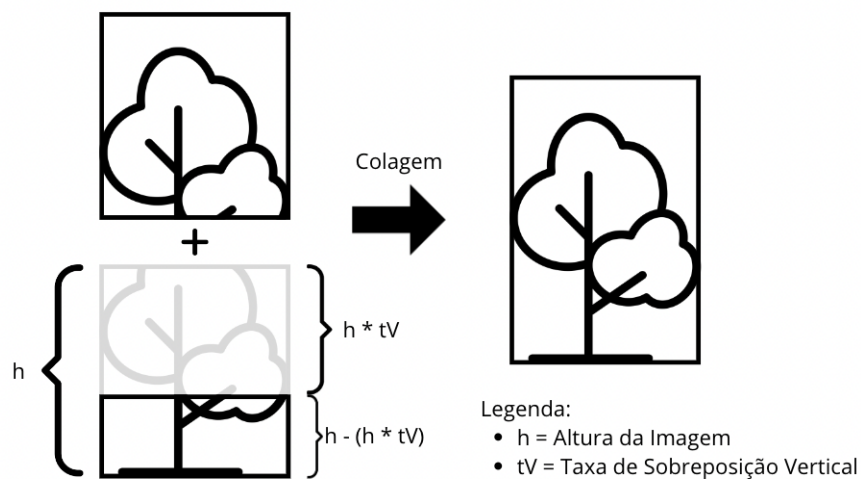
Para a geração do panorama, diversas combinações das imagens coletadas foram testadas em um algoritmo de *stitching* implementado pela biblioteca de manipulação de imagens *OpenCV*. Foram também utilizadas técnicas de tratamento de imagens com a mesma biblioteca, juntamente com algoritmos *stitching* personalizados, com a finalidade de conseguir uma colagem de imagens mais eficaz. As técnicas utilizadas foram a manipulação das cores e a aplicação do método *inRange* do *OpenCV* nas imagens utilizadas para a detecção de pontos chaves, de maneira a destacar as maçãs, que são os principais elementos que devem ser sobrepostos corretamente na etapa de *stitching*.

Também foi implementado um algoritmo simples para permitir a colagem das imagens lateralmente, sem fazer nenhum cálculo de sob. ou de detecção de pontos chave nas imagens, como é feito pelo *stitching*: o algoritmo apenas realizou a colagem das imagens horizontalmente e verticalmente, de maneira a formar uma única imagem extremamente larga, chegando à uma largura de 25920 pixels para uma colagem de 54 imagens com 480 pixels de largura cada, representando toda a lateral de uma fila de árvores. O algoritmo contou com duas variáveis

configuráveis que podem receber valores de 0 à 1 e que referem-se às porcentagens de sob. previstas verticalmente e horizontalmente entre as fotos.

Nesse processo, a colagem das imagens foi feita ordenadamente, começando pela imagem mais à esquerda e mais alta do pomar, que sempre é utilizada por inteira. A partir dela novas imagens passaram a ser coladas abaixo e à direita. A variável de sob. vertical define quanto de altura da imagem colada deve ser descartada: se for 0, a imagem será colada abaixo inteiramente, mas se for de 0,5, a metade superior da imagem será considerada como já presente na imagem de cima e portanto é descartada. O processo de colagem de imagens à direita é análogo. Na Figura 18 pode ser observada a técnica de colagem vertical aplicada com uma taxa de sob. vertical maior que 0.

**Figura 18 – Colagem vertical de imagens utilizando taxa de sob. vertical maior que zero**



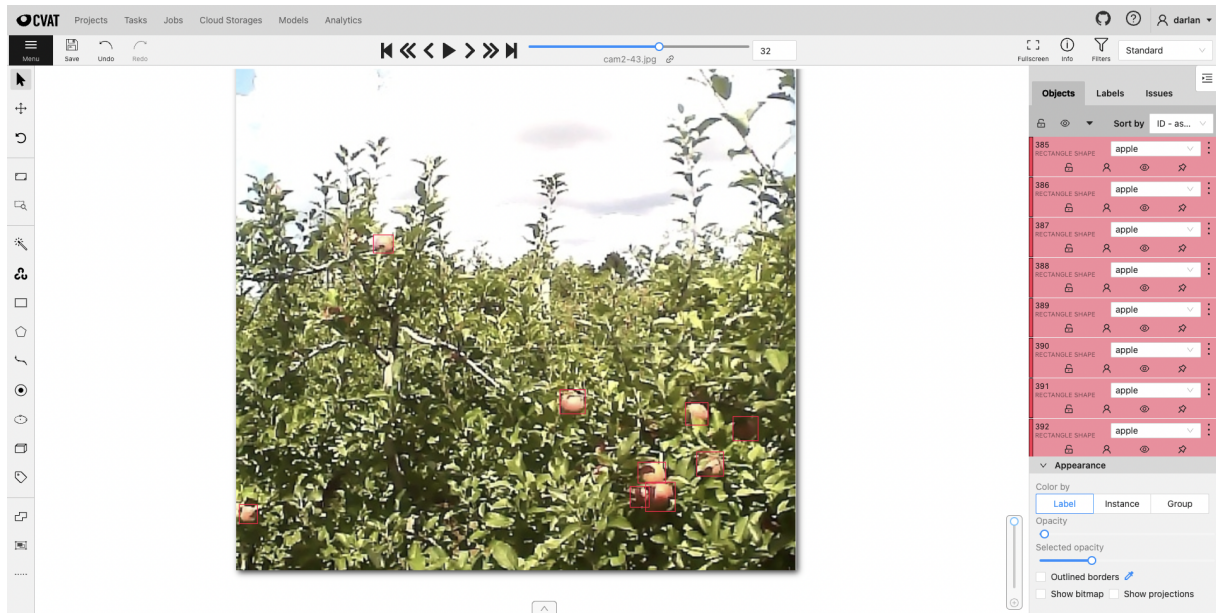
**Fonte: Autoria própria (2022).**

Através desse processo, foram criados diferentes panoramas para a fila 1, utilizando todas as combinações de taxas de sob. horizontal e vertical para os seguintes valores: 0, 0,2, 0,4, 0,6 e 0,8. Portanto, foram criados 50 panoramas, 25 para cada lado da fila 1.

### 3.2.4 Preparação dos *datasets*

Utilizando a ferramenta CVAT, as imagens das quatro filas coletadas durante o período da manhã foram rotuladas com a única classe relevante para o problema discutido: maçã. As imagens rotuladas da fila 1 foram separadas para teste, pois foi a fila escolhida para ser utilizada na contagem das frutas. As imagens das filas 2, 3 e 4 foram separadas para o *dataset* de treino, com exceção das imagens da câmera 2 do lado B e câmera 1 do lado A da fila 4, que foram separadas para o conjunto de validação. Essa escolha foi feita de forma a permitir que o conjunto de validação fosse composto por imagens de câmeras com ângulo e iluminação diferentes. Uma imagem da ferramenta CVAT com a rotulação das frutas pode ser vista na Figura 19. O tamanho de cada *dataset* pode ser observado na Tabela 1.

Figura 19 – Rotulação de maçãs na ferramenta CVAT



Fonte: Autoria própria (2022).

Tabela 1 – Distribuição de imagens entre os *datasets*

<i>Dataset</i>	Fila de Origem	Total de Imagens
Treino	02, 03 e 04	836
Validação	04	105
Teste	01	322

Fonte: Autoria própria (2022).

### 3.2.5 Treinamento de modelos

Nesta etapa, diferentes treinos foram realizados no ambiente do *Colab* com diferentes modelos de CNN da arquitetura YOLOv5 com o *dataset* gerado na etapa anterior. Os resultados dos modelos foram analisados a fim de escolher a rede com melhor desempenho na detecção de maçãs nas imagens do *dataset*.

A primeira seção de treinos realizados teve como objetivo identificar quais parâmetros de treinamento poderiam dar os melhores resultados. Para tanto, foram selecionados os modelos base com pesos pré-treinados YOLOv5n e YOLOv5n6, que são os modelos mais simples da família, sendo eles pré-treinados, respectivamente, com imagens de resolução de  $640 \times 640$  e  $1280 \times 1280$  pixels. A partir desses modelos, foram executados treinos alterando os parâmetros de treino *batch size*, entre 8 e 16, e resolução da imagem de entrada, entre  $640 \times 640$  e  $1280 \times 1280$  pixels. Após o fim de cada treino, o modelo foi utilizado para a detecção de maçãs no *dataset* de teste para obter seus resultados de precisão, cobertura e mAP.

A partir dos resultados da primeira sessão de treinos, foram escolhidos os parâmetros e a resolução de pré-treino que entregaram melhores resultados de precisão e cobertura. A partir de tais parâmetros, foi executada outra seção de treinos, dessa vez alterando os modelos da família entre os modelos pré-treinados com a resolução de pré-treino que obteve os melhores

resultados. Assim como na primeira seção de treinos, os modelos foram testados com o *dataset* de teste.

Por fim, com o objetivo de escolher o modelo que demonstrasse maior capacidade de identificar todas as maçãs presentes nas fotos, o modelo que obteve os melhores resultados de cobertura foi escolhido para ser utilizado nas etapas seguintes do projeto.

### 3.2.6 Aplicação de modelo aos panoramas

Com o melhor modelo treinado em mãos, foi utilizada a biblioteca SAHI para detecção de maçãs nas imagens panorâmicas da fila 1 e os resultados obtidos foram comparados com os dados reais, fornecidos pela EMBRAPA Uva e Vinho. As taxas de sob. dos panoramas com resultados mais próximos do real foram então escolhidas e utilizadas na geração de panoramas para as fotos das filas 2 e 3 coletadas às 16 horas, com o objetivo de checar se os resultados se mantinham próximos aos reais.

## 4 RESULTADOS E DISCUSSÕES

Neste capítulo, é descrito o resultado do processo de coleta das imagens no campo experimental da EMBRAPA, além de ser descrito o processo de rotulação das imagens e a separação delas em diferentes *datasets*.

Também são descritos aqui os resultados do processo de *stitching*, os resultados dos treinos dos modelos de RNA, bem como a contagem das maçãs nas imagens panorâmicas

### 4.1 Coleta de Imagens

O processo de coleta das imagens resultou em 4062 imagens coletadas, todas de tamanho  $640 \times 480$ . Essas fotos, porém, contavam com partes do pomar que já haviam sido colhidas e não possuíam mais maçãs para serem contadas. Foi feita, então, uma separação dessas fotos em um novo conjunto que só possuísse imagens da parte do pomar que ainda não tivesse sido colhido. Nesta etapa, elas também foram separadas em lado A e B da fila de árvores. O resultado desse processo foi um novo conjunto de 1908 imagens distribuídas de acordo com a Tabela 2.

Tabela 2 – Total de imagens coletadas

Fila - Hr. Coleta	Câmera	Total	Lado A	Lado B	Relevantes
01 - 11h	01	212	54	54	108
01 - 11h	02	212	54	54	108
01 - 11h	03	212	54	54	108
02 - 11h	01	220	49	49	98
02 - 11h	02	220	49	49	98
02 - 11h	03	220	49	49	98
03 - 11h	01	226	55	55	110
03 - 11h	02	225	55	55	110
03 - 11h	03	225	55	55	110
04 - 11h	01	218	54	53	107
04 - 11h	02	218	54	53	107
04 - 11h	03	218	54	53	107
02 - 16h	01	256	52	52	104
02 - 16h	02	256	52	52	104
02 - 16h	03	256	52	52	104
03 - 16h	01	223	54	55	109
03 - 16h	02	223	54	55	109
03 - 16h	03	222	54	55	109
<b>Total</b>	-	4062	-	-	1908

Fonte: Autoria própria (2022).

Quanto à sob. entre as imagens, na Figura 20 pode-se observar a sob. vertical entre diferentes câmeras através da identificação de um mesmo grupo de maçãs presentes nas três

imagens. Já na Figura 21 pode-se observar a sob. horizontal entre fotos tiradas em diferentes tempos, dado a presença, também, de um mesmo grupo de maçãs nas duas imagens.

**Figura 20 – Imagens obtidas ao mesmo tempo pelas câmeras 1, 2 e 3, da esquerda para a direita, com destaque para o mesmo grupo de maçãs presentes em todas as fotos**



Fonte: Aatoria própria (2022).

**Figura 21 – Sob. de captura entre duas fotos da câmera 1 em tempo vizinhos, com destaque para o mesmo grupo de maçãs aparecendo em ambas as imagens**



Fonte: Aatoria própria (2022).

A qualidade das imagens obtidas obteve variações consideráveis. O reflexo do sol, bem como o movimento da estrutura com as câmeras fez com que diversas fotos ficassem ofuscadas e/ou borradas. Isso pode ser observado na Figura 22.



**Figura 22 – Imagem borrada (esquerda) e imagem ofuscada (direita)**



**Fonte: Aatoria própria (2022).**

## 4.2 Geração de Panoramas

Na etapa de geração de panoramas foram selecionadas diversas duplas de imagens com sob. bastante clara e que não estivessem borradas ou escuras por conta do reflexo do sol. Entretanto, apesar dessa escolha de imagem específica, a ferramenta de *stitching* do *OpenCV* não foi capaz de criar um panorama, mesmo entre apenas duas imagens, tornando impossível a geração do panorama de toda a lateral da linha do pomar através desta técnica.

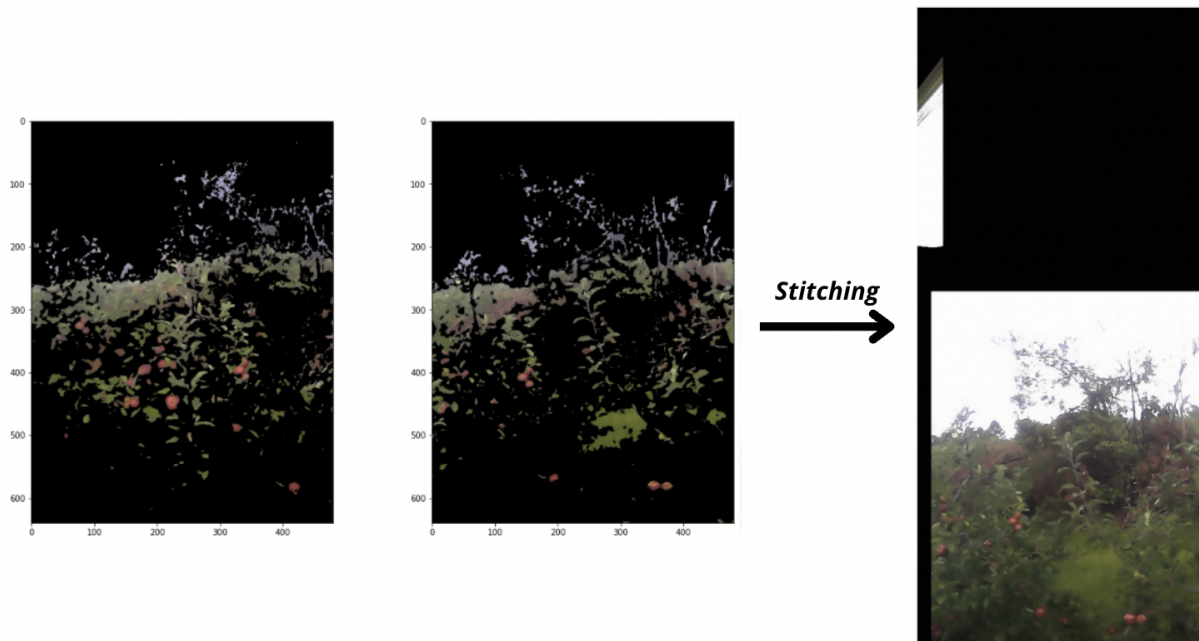
Para o *stitching* personalizado para a detecção de pontos chaves em imagens pré-processadas, os resultados obtidos também não foram promissores. Na Figura 23 pode ser observado a aplicação do método *inRange* para a remoção de céu, folhas e galhos das imagens utilizadas para a detecção de pontos chaves, de maneira a destacar a presença das maçãs. Apesar desse pré-processamento, os pontos chaves encontrados não permitiram uma colagem adequada das imagens.

Resultados semelhantes podem ser observados na Figura 24, na qual as imagens foram pré-processadas com uma manipulação das camadas de cores para ressaltar a cor vermelha das frutas, e na Figura 25, que mostra a junção dos dois pré-processamentos nas imagens.

Para que o trabalho pudesse ter continuidade, o processo de colagem de imagens descrito na seção 3.2.2 foi executado, resultando em um montante de 50 imagens panorâmicas para a fila 1.

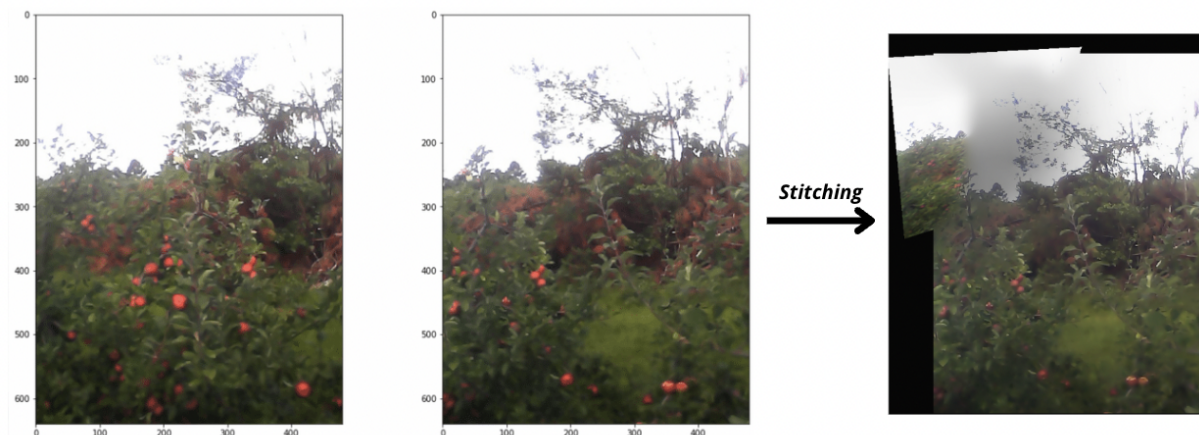
Também deve-se ressaltar que foram utilizadas apenas as imagens das câmeras 1 e 3, dado que existe uma área de intersecção entre a captura dessas duas câmeras, como visto na Figura 20, tornando desnecessário o uso da câmera 2 nesta situação. Um exemplo de resultado desse processo para uma das laterais da fila 1 pode ser observado na Figura 26, no qual foram utilizadas taxas de sob. vertical e horizontal iguais a 0,6 e 0,2, respectivamente.

Figura 23 – *Stitching* com aplicação do método *inRange* nas imagens para a localização dos pontos chaves



Fonte: Autoria própria (2022).

Figura 24 – *Stitching* com a manipulação da camada de cores nas imagens para a localização dos pontos chaves



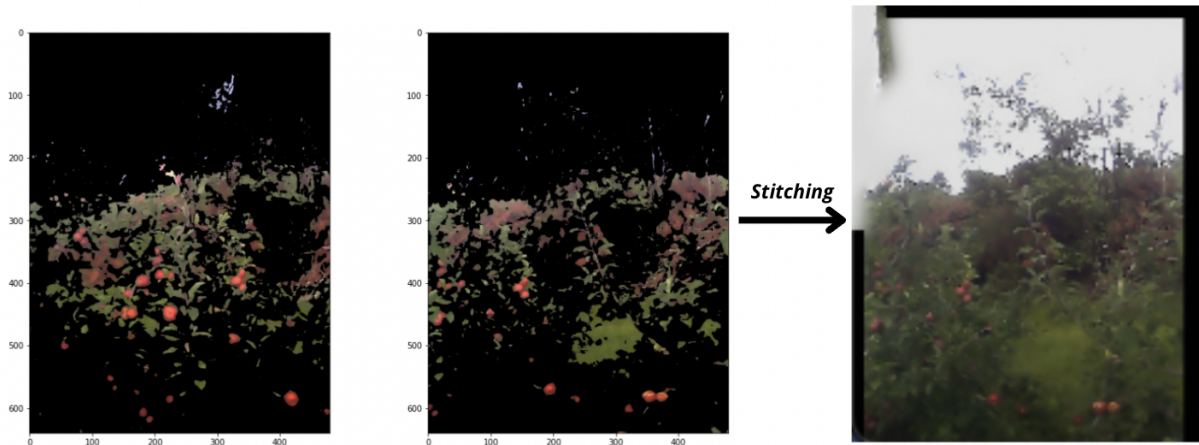
Fonte: Autoria própria (2022).

### 4.3 Treinamento de Modelos

Nesta seção estão descritos os resultados dos treinamentos dos modelos de CNN executados de acordo com a metodologia. Os resultados da primeira seção de treinamento pode ser observado na Tabela 3.

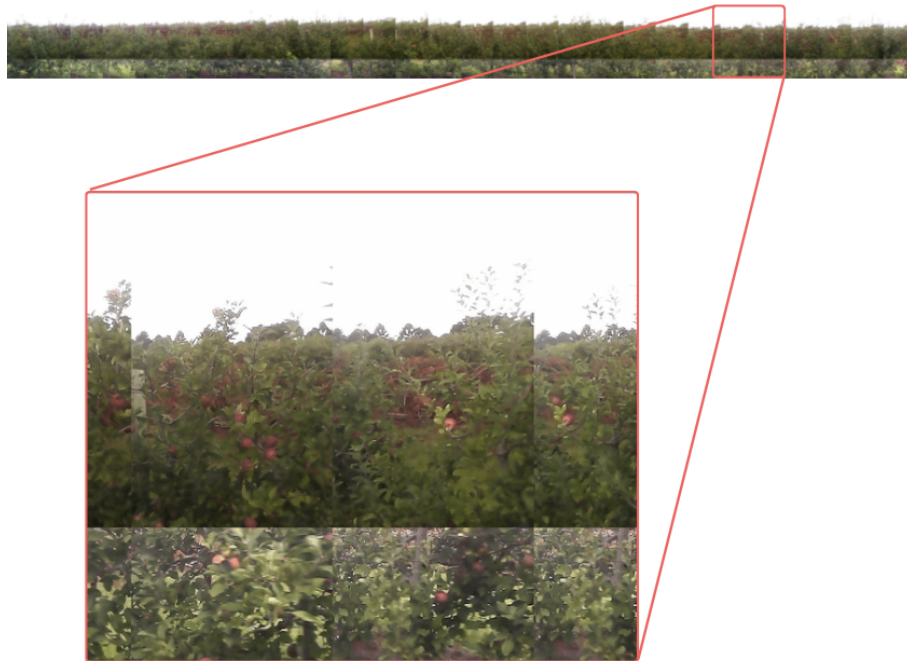
Os resultados de todos os modelos treinados ficaram bastante próximos, tendo a precisão variado apenas de 79,7% à 82% e a cobertura tendo variado entre 70,3% e 73,5%. Entre as redes treinadas com *batch size* de tamanho 8, foi obtida uma precisão variando de 79,7% à 82,0% e uma cobertura variando de 70,7% à 73,1%. Já entre os modelos treinados com *batch size* de tamanho 16, as variações foram entre 80,4% e 82,0% para precisão e entre 70,3% e

**Figura 25 – *Stitching* com aplicação do método *inRange* e manipulação da camada de cores nas imagens para a localização dos pontos chaves**



Fonte: Autoria própria (2022).

**Figura 26 – Panorama feito através da colagem de imagens**



Fonte: Autoria própria (2022).

73,5% para cobertura, o que mostra que o *batch size* maior pôde obter melhores resultados para a cobertura, porém sem apresentar melhorias para a precisão.

Também é possível comparar os resultados para as diferentes resoluções de entrada escolhidas. Para a resolução de  $640 \times 640$  pixels, a precisão variou de 79,7% até 82,0% e a cobertura de 70,3% até 70,8%. Para a resolução de  $1280 \times 1280$  pixels a variação foi entre 81,1% e 82,0% para a precisão e entre 73,0% e 73,5% para a cobertura, mostrando que a resolução de entrada maior permite ao modelo obter melhores resultados de cobertura.

Os parâmetros escolhidos para serem utilizados no treinamento seguinte foram os do treinamento que utilizou os pesos pré-treinados com imagens de resolução de  $640 \times 640$  pixels

Tabela 3 – Resultados da primeira seção de treinos

Modelo	Batch size	Resolução (px)	Precisão (%)	Cobertura (%)	mAP:0.5	Tempo (h:m:s)
YOLOv5n	8	640	81,8	70,7	0,786	00:27:29
YOLOv5n	8	1280	81,1	73,0	0,798	01:02:32
YOLOv5n	16	640	82,0	70,3	0,778	00:18:59
<b>YOLOv5n</b>	<b>16</b>	<b>1280</b>	<b>81,5</b>	<b>73,5</b>	<b>0,809</b>	<b>00:54:43</b>
YOLOv5n6	8	640	79,7	70,8	0,772	00:29:48
YOLOv5n6	8	1280	82,0	73,1	0,809	01:05:30
YOLOv5n6	16	640	80,4	70,6	0,773	00:20:07
YOLOv5n6	16	1280	81,9	73,1	0,810	00:58:00

Fonte: A autoria própria (2022).

(YOLOv5n), *batch size* de 16 e resolução de  $1280 \times 1280$  pixels para as imagens de entrada, pois foi o modelo com a melhor cobertura (73,5%) e precisão apenas 0,5% abaixo do melhor resultado (82%). O *batch size* foi alterado de acordo com cada modelo treinado, para que o ambiente tivesse condições de executar o treinamento, visto que, para treinamento com *batches* de tamanho 16, os modelos YOLOv5m e YOLOv5l necessitavam mais memória RAM do que a GPU tinha a oferecer, visto que esses modelos são computacionalmente mais exigentes por conterem uma quantidade maior de pesos a serem treinados. A partir destes parâmetros, três novos treinos foram feitos, utilizando agora os seguintes modelos com pesos pré treinados: YOLOv5s, YOLOv5m e YOLOv5l. Os resultados desses treinos estão dispostos na Tabela 4.

Tabela 4 – Resultados da segunda seção de treinos

Modelo	Batch size	Precisão (%)	Cobertura (%)	mAP:0.5	Tempo (h:m:s)
YOLOv5s	16	82,8	75,9	0,836	01:51:09
<b>YOLOv5m</b>	<b>12</b>	<b>82,5</b>	<b>77,8</b>	<b>0,846</b>	<b>04:54:10</b>
YOLOv5l	6	82,5	75,4	0,833	07:14:38

Fonte: A autoria própria (2022).

Diante desses resultados, o segundo modelo da Tabela 4 que utiliza o modelo com pesos pré-treinados YOLOv5m foi selecionado para o processo de identificação e contagem de maçãs nas imagens panorâmicas, pois é o modelo que apresentou maior cobertura, ou seja, que identificou o maior número de maçãs corretamente, mantendo uma precisão semelhante aos demais modelos analisados.

#### 4.4 Aplicação de modelo aos panoramas

Para que o desempenho do modelo aplicado aos panoramas possa ser medido, é necessário a comparação com a quantidade real de frutas presentes nas filas do pomar. Portanto, os dados reais da colheita de maçãs da área analisada, obtidos pela EMBRAPA Uva e Vinho, podem ser vistos na Tabela 5.

Utilizando o melhor modelo obtido na etapa anterior, foi então realizada a detecção de maçãs no panorama gerado na seção 4.2 para a fila 1. Os resultados da detecção para cada

Tabela 5 – Quantidade de maçãs coletadas no campo experimental

Fila	Lado A	Lado B	Total
01	769	1230	1999
02	754	952	1706
03	1110	1510	2620
04	1004	1322	2326

Fonte: Autoria própria (2022).

panorama gerado podem ser vistos na Tabela 6 e comparados com o valor da contagem real, que foi de 1999 maçãs para essa fila. A tabela também acompanha um dado de cobertura, que refere-se à proporção entre maçãs contadas pelo modelos e o valor real de maçãs presente na fila.

Tabela 6 – Resultados da detecção de maçãs na fila 1

Sob. Vertical	Sob. Horizontal	Lado A	Lado B	Total	Cobertura (%)
0,8	0,8	86	77	163	8,15
0,8	0,6	197	185	382	19,11
0,8	0,4	286	278	564	28,21
0,8	0,3	387	406	793	39,67
0,8	0,0	452	513	965	48,27
0,6	0,8	102	91	193	9,65
0,6	0,6	237	216	453	22,66
0,6	0,4	345	355	700	35,02
0,6	0,3	486	523	1009	50,48
0,6	0,0	583	673	1256	62,83
0,4	0,8	134	136	270	13,51
0,4	0,6	297	295	592	29,61
0,4	0,4	439	467	906	45,32
0,4	0,3	637	666	1303	65,18
0,4	0,0	778	846	1624	81,24
0,2	0,8	156	151	307	15,36
0,2	0,6	342	331	673	33,67
0,2	0,4	506	521	1027	51,38
0,2	0,3	739	755	1494	74,74
0,2	0,0	897	979	1876	93,85
0,0	0,8	181	186	367	18,36
0,0	0,6	381	378	759	37,97
0,0	0,4	551	603	1154	57,73
0,0	0,3	787	860	1647	82,39
<b>0,0</b>	<b>0,0</b>	<b>962</b>	<b>1100</b>	<b>2062</b>	<b>103,15</b>

Fonte: Autoria própria (2022).

Analisando os resultados, é possível observar que o panorama no qual o modelo contou uma quantidade de maçãs mais próxima da real, contanto uma quantidade equivalente a 103,15% da quantidade real, foi no panorama gerado com uma taxa de sob. vertical e horizontal igual a 0. Considerando que, com essa taxa, as fotos são utilizadas inteiramente e, por isso, há a repetição de maçãs ao longo da imagem, acredita-se que o método para obter esse resultado

se “escora” na múltipla contagem de uma mesma maçã para suprir a dificuldade do modelo de CNN treinado em identificar todas as frutas na imagem.

Por fim, para avaliar a constância do sistema, foi realizada a colagem das imagens coletadas à tarde para as filas 2 e 3, com taxa de sob. horizontal e vertical igual à 0. O modelo de CNN treinado foi então aplicado às imagens e os resultados obtidos para a contagem de frutas pode ser observado na Tabela 7.

**Tabela 7 – Resultados da detecção de maçãs nas filas 02 e 03**

<b>Fila</b>	<b>Lado A</b>	<b>Lado B</b>	<b>Total Real</b>	<b>Total Contado</b>	<b>Cobertura (%)</b>
02 - 16h	532	687	1706	1219	71,45
03 - 16h	596	1033	2620	1629	62,18

**Fonte: Autoria própria (2022).**

Dados os resultados da Tabela 7, é possível observar que o sistema não teve constância na proporção entre maçãs contadas por ele e a quantidade real de maçãs colhidas, visto que os valores de cobertura obtidos foram de 71,45% e 62,18%, valores bem abaixo da cobertura de 103,15% obtida no teste inicial.

## 5 CONSIDERAÇÕES FINAIS

Neste capítulo será abordada a conclusão de todo o trabalho, além de discutir possíveis trabalhos futuros realizados em continuidade ou em paralelo ao trabalho aqui disposto.

### 5.1 Conclusão

Este trabalho teve como objetivo a montagem e a utilização de um equipamento capaz de coletar imagens de macieiras em pomares, bem como a utilização dessas imagens para a contagem de maçãs produzidas. Para a montagem da estrutura foram utilizados materiais como tábuas de madeira e barras de alumínio, que deram suporte à um RasPi, responsável por controlar a coleta das imagens obtidas por três câmeras conectadas à ele. Para realizar a estimativa da produção, foi adotada a ideia de fazer colagem das fotos obtidas através da biblioteca *OpenCV*, de maneira a evitar a aparição da mesma maçã em fotos diferentes, além de ter sido treinada uma CNN da família YOLOv5 para realizar a contagem propriamente dita.

A montagem da estrutura e a coleta das imagens ocorreu de maneira bem sucedida, resultando em mais de 4000 imagens coletadas. O processo de colagem das imagens utilizando o algoritmo de *stitching* disponibilizado pela biblioteca *OpenCV*, bem como o uso de algoritmo *stitching* personalizado com o pré-processamento das imagens para obtenção dos pontos-chaves, não foi possível, e necessita de estudos mais aprofundados para que possa ser realizado com sucesso. Entretanto, uma alternativa mais simples foi utilizada para a geração de imagens panorâmicas através da colagem de imagens lateralmente sem a realização da sob. delas, que embora não tenha garantido a unicidade de cada maçã presente na foto, tornou possível a continuidade do trabalho.

Para o processo de treinamento da RNA foi necessário primeiro a preparação de *datasets* a partir das imagens coletadas através da ferramenta CVAT. Todos os *datasets* necessários foram preparados com sucesso e diversos modelos de RNA foram treinados. A melhor rede treinada teve uma precisão e uma cobertura de 82,5% e 77,8%, respectivamente, e foi então aplicada na contagem de maçãs em diferentes modelos de panoramas gerados, tendo o melhor resultado alcançado uma contagem equivalente a 103,15% da contagem real. Tal proporção de contagem, entretanto, não se manteve em contagens realizadas para outras filas do pomar, demonstrando instabilidade no sistema proposto.

Embora tenha sido possível realizar uma contagem bastante próxima à real, o que prova a viabilidade do conceito proposto, o método ainda pode ser aprimorado para obter resultados mais estáveis, de maneira que possa obter uma proporção entre frutas contadas e colhidas constante em todas as aplicações, visto que a média de cobertura para todos os testes realizados foi de 78,92%.

## 5.2 Trabalhos futuros

Os estudos apresentados neste trabalho podem ser aprofundados e levados adiante com a finalidade de obter melhores resultados. O processo de *stitching* pode ser aprofundado, a fim de encontrar maneira de realizar a sobreposição das frutas e a colagem das imagens de maneira bem sucedida e gerar um panorama mais adequado para a contagem das frutas. Também é possível o estudo com diferentes parâmetros, modelos e arquiteturas de CNNs para obter melhores resultados da identificação de maçãs nas imagens do *dataset* criado.

Além disso, a estrutura montada para a coleta das imagens pode ser utilizada para a montagem de outros *datasets*, podendo coletar fotos de outros cultivos ou de outros estágios dos pomares de maçãs, com a finalidade de realizar diferentes estudos.

Por fim, o *dataset* coletado pode ser estudado para diferentes problemas, como a identificação da qualidade das frutas do pomar e a identificação de doenças nas plantas.



## REFERÊNCIAS

- AGGARWAL, C. C. **Neural Networks and Deep Learning**. [S.l.]: Springer, 2018. ISBN 978-3-319-94463-0.
- ALFEN, N. K. V. Encyclopedia of agriculture and food systems. In: \_\_\_\_\_. [S.l.]: Elsevier Inc., 2014. v. 1, p. 406–416.
- ANZOLIN, T. B. P. *et al.* Desenvolvimento de mosaico de imagens para automatização de processos na fruticultura de precisão. In: **18º Congresso Latino-americano de Software Livre e Tecnologias Abertas**. Foz do Iguaçu: [s.n.], 2021.
- BLANKE, M. ‘gala’ und ‘fuji’ – brasiliens apfelfanbau mit 2 sorten im einklang mit der natur zwischen alten araucaria-bäumen und agroforestry. **Erwerbs-Obstbau**, v. 55, p. 87–93, 12 2013. ISSN 0014-0309. Disponível em: <http://link.springer.com/10.1007/s10341-013-0190-2>.
- BYRNE, D. H.; BADENES, M. L. **Fruit Breeding**. [S.l.]: Springer-Verlag New York, 2012. 875 p.
- CHENG, R. A survey: Comparison between convolutional neural network and yolo in image identification. **Journal of Physics: Conference Series**, Institute of Physics Publishing, v. 1453, p. 012139, 1 2020. ISSN 1742-6588. Disponível em: <https://iopscience.iop.org/article/10.1088/1742-6596/1453/1/012139>.
- CORNILLE, A. *et al.* **The domestication and evolutionary ecology of apples**. [S.l.]: Elsevier Ltd, 2014. 57-65 p.
- CVAT. **About Us**. 2021. Disponível em: <https://openvinotoolkit.github.io/cvat/about/>.
- DENARDI, F.; KVITSCHAL, M. V.; HAWERROTH, M. C. A brief history of the forty-five years of the epagri apple breeding program in brazil. **Crop Breeding and Applied Biotechnology**, v. 19, p. 347–355, 2019. Disponível em: <http://dx.doi.org/10.1590/1984-70332019v19n3p47>.
- EMBRAPA. Visão 2030: O futuro da agricultura brasileira. 2018.
- EMBRAPA. **Quem somos**. 2021. Portal Embrapa. Disponível em: <https://www.embrapa.br/quem-somos>.
- FACELI, K. *et al.* **Inteligência Artificial: uma abordagem de aprendizado de máquina**. [S.l.]: LTC, 2011. ISBN 978-85-216-1880-5.
- FENG, A. *et al.* Evaluation of cotton emergence using uav-based narrow-band spectral imagery with customized image alignment and stitching algorithms. **Remote Sensing**, MDPI AG, v. 12, p. 1764, 5 2020. ISSN 2072-4292. Disponível em: <https://www.mdpi.com/2072-4292/12/11/1764>.
- FIALHO, V. M. Z. B. **Cientistas desvendam mecanismo de percepção do frio na macieira**. 2019. Portal Embrapa. Disponível em: <https://www.embrapa.br/busca-de-noticias/-/noticia/43442887/cientistas-desvendam-mecanismo-de-percepcao-do-frio-na-macieira>.
- FOUNDATION, S. M. S. **Role of Modern Technology in Agriculture**. 2021. Sehgal Foundation Blog. Disponível em: <https://www.smsfoundation.org/role-of-modern-technology-in-agriculture/>.
- GENÉ-MOLA, J. *et al.* Multi-modal deep learning for fuji apple detection using rgb-d cameras and their radiometric capabilities. **Computers and Electronics in Agriculture**, Elsevier B.V., v. 162, p. 689–698, 7 2019. ISSN 01681699. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0168169919301413>.

- GHOSH, D.; KAABOUC, N. A survey on image mosaicing techniques. **Journal of Visual Communication and Image Representation**, Academic Press Inc., v. 34, p. 1–11, 1 2016. ISSN 10473203. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1047320315002059>.
- GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing**. [S.l.]: Pearson, 2018. 1168 p.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. [S.l.]: Massachusetts Institute of Technology, 2016. ISBN 9780262035613.
- HEATON, J. **Artificial intelligence for humans**. 1. ed. [S.l.]: Heaton Research, 2013. v. 1. ISBN 978-1493682225.
- HEATON, J. **Artificial intelligence for humans**. 1. ed. [S.l.]: Heaton Research, 2015. v. 3. ISBN 978-1505714340.
- LEE, C.; KIM, H. J.; OH, K. W. Comparison of faster r-cnn models for object detection. In: . IEEE, 2016. p. 107–110. ISBN 978-89-93215-11-3. Disponível em: <http://ieeexplore.ieee.org/document/7832305/>.
- MELLO, L. M. R. d. Produção e mercado brasileiro de maçã. **Embrapa Uva e Vinho**, 2004.
- MENDES, A. R.; BARBIERI, M. **MAÇÃ/CEPEA: Volume colhido recua quase 20% na campanha 2019/20**. 2020. Portal HF Brasil. Disponível em: <https://www.hfbrasil.org.br/br/maca-cepea-volume-colhido-recua-quase-20-na-campanha-2019-20.aspx>.
- OPENCV. **About**. 2021. Disponível em: <https://opencv.org/about/>.
- PETRI, J. L. *et al.* Advances in fruit crop propagation in brazil and worldwide – apple trees. **Revista Brasileira de Fruticultura**, Sociedade Brasileira de Fruticultura, v. 41, 2019. ISSN 01002945.
- PYTHON. **About Python**. 2021. Disponível em: <https://www.python.org/about/>.
- RASHID, M. *et al.* Novel image processing technique for feature detection of wheat crops using python opencv. In: **2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)**. [S.l.: s.n.], 2019. p. 559–563.
- SAHI. **SAHI: Slicing Aided Hyper Inference**. 2022. Disponível em: <https://github.com/obss/sahi>.
- SEBRAE. **O cultivo e o mercado da maçã**. 2013. Portal Sebrae. Disponível em: <https://www.sebrae.com.br/sites/PortalSebrae/artigos/o-cultivo-e-o-mercado-da-maca,ea7a9e665b182410VgnVCM100000b272010aRCRD>.
- SHALEV-SHWARTZ, S.; BEN-DAVID, S. **Understanding Machine Learning**. [S.l.]: Cambridge University, 2014. ISBN 978-1-107-05713-5.
- SOUSA, M. L. de. Manual de boas práticas de fruticultura. 2021. Disponível em: [https://www.inia.pt/images/Noticias/manual\\_de\\_fruticultura\\_macieira.pdf](https://www.inia.pt/images/Noticias/manual_de_fruticultura_macieira.pdf).
- SZELISKI, R. **Computer Vision: Algorithms and Applications**. [S.l.]: Springer, 2011. 833 p.
- TAYLOR, M. **Neural Networks A Visual Introduction for Beginners**. [S.l.]: Blue Windmill Media, 2017. ISBN 9781549869136.
- TIAN, Y. *et al.* Apple detection during different growth stages in orchards using the improved yolo-v3 model. **Computers and Electronics in Agriculture**, Elsevier B.V., v. 157, p. 417–426, 2 2019. ISSN 01681699.

VASILEV, I. *et al.* **Python deep learning: exploring deep learning techniques and neural network architectures with PyTorch, Keras, and TensorFlow**. 2. ed. [S.l.]: Packt Publishing, 2019. ISBN 978-1-78934-846-0.

VASYLIEVA, N.; JAMES, H. Production and trade patterns in the world apple market. **Innovative Marketing**, LLC CPC Business Perspectives, v. 17, p. 16–25, 1 2021. ISSN 18142427. Disponível em: <https://businessperspectives.org/journals/innovative-marketing/issue-374/production-and-trade-patterns-in-the-world-apple-market>.

WANG, Z.; YANG, Z. Review on image-stitching techniques. **Multimedia Systems**, Springer, v. 26, p. 413–430, 8 2020. ISSN 0942-4962. Disponível em: <http://link.springer.com/10.1007/s00530-020-00651-y>.

XIAO, Y. *et al.* A review of object detection based on deep learning. **Multimedia Tools and Applications**, Springer, v. 79, p. 23729–23791, 9 2020. ISSN 15737721.