

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE INFORMÁTICA  
CURSO DE CIÊNCIA DA COMPUTAÇÃO

IZAIAS SATURNINO DE LIMA NETO

**Detecção de Objetos Aplicada a  
Nanopartículas com YOLO**

Monografia apresentada como requisito parcial  
para a obtenção do grau de Bacharel em Ciência  
da Computação

Orientador: Prof. Dr. Anderson Rocha Tavares

Porto Alegre  
2025

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitora: Prof<sup>a</sup>. Marcia Cristina Bernardes Barbosa

Vice-Reitor: Prof. Pedro de Almeida Costa

Pró-Reitora de Graduação: Prof<sup>a</sup>. Nádyá Pesce da Silveira

Diretor do Instituto de Informática: Prof. Luciano Paschoal Gaspary

Coordenador do Curso de Ciência de Computação: Prof. Marcelo Walter

Bibliotecário-chefe do Instituto de Informática: Alexsander Borges Ribeiro

## RESUMO

Nanopartículas possuem propriedades físicas, químicas e biológicas únicas, decorrentes de sua estrutura em escala nanométrica. Essas propriedades têm amplas aplicações em campos científicos e industriais. A determinação do tamanho das nanopartículas é crucial para a compreensão de suas propriedades e comportamentos, tornando necessárias técnicas de medição precisas. A medição manual é demorada, frequentemente exigindo horas de trabalho, o que destaca a demanda por automação para melhorar sua eficiência.

Este trabalho tem como objetivo desenvolver um modelo de visão computacional, especificamente YOLOv8, para automatizar parcialmente a detecção e medição dos tamanhos de nanopartículas de bismuto em imagens de microscopia eletrônica de transmissão. O modelo trabalha com dados limitados, identifica caixas delimitadoras ao redor das nanopartículas, inferindo seus diâmetros, e responde às limitações da contagem manual, garantindo precisão em dados com qualidades de imagem variadas.

Diferentes modelos de detecção de objetos foram comparados, incluindo o *simple blob detector* disponibilizado pelo OpenCV e versões do modelo YOLOv8, com e sem aumento de dados offline. O modelo YOLOv8 com caixas delimitadoras orientadas, treinado sem aumento de dados adicional, apresentou o melhor desempenho geral, sendo o mais eficaz na medição e alcançando a segunda melhor performance na detecção de nanopartículas dentre os modelos baseados em *deep learning*.

A solução proposta é de código aberto, o que pode democratizar o acesso a ferramentas de medição automatizada, possibilitando que pequenos grupos de pesquisa e instituições com menos recursos realizem estudos de nanopartículas de forma mais eficiente.

Além disso, a metodologia desenvolvida tem o potencial de ser adaptada para analisar outras estruturas microscópicas, como células, microplásticos ou materiais nanoestruturados, ampliando sua aplicabilidade para outras investigações científicas.

**Palavras-chave:** Detecção de objetos. Rede neural convolucional. Nanopartícula. Microscopia eletrônica de transmissão. Visão computacional.

# Nanoparticle Object Detection with YOLO

## ABSTRACT

Nanoparticles possess unique physical, chemical, and biological properties that stem from their nanoscale structure. These properties have broad applications in scientific and industrial fields. Determining the size of nanoparticles is crucial for understanding their properties and behaviors, making precise measurement techniques essential. Manual size measurement is time-intensive and often requires hours of effort. This highlights the need for automation to improve efficiency.

This study aims to develop a computer vision model, specifically YOLOv8, for partially automating the detection and measurement of bismuth nanoparticle sizes in transmission electron microscopy images. The model works with limited data, identifies bounding boxes around nanoparticles, inferring their diameters and addresses the challenges of manual counting, ensuring accuracy across datasets with varying image qualities.

Different object detection models were compared, including the simple blob detector provided by OpenCV and versions of the YOLOv8 model, with and without offline data augmentation. The YOLOv8 model with oriented bounding boxes, trained without additional data augmentation, showed the best overall performance, being the most effective in measurement and achieving the second-best performance in nanoparticle detection of the deep learning based models.

The proposed solution is open-source and low-cost, which can democratize access to automated measurement tools, possibly enabling smaller research groups and institutions with less resources to conduct nanoparticle studies more effectively.

Furthermore, the methodology developed has the potential to be adapted for analyzing other microscopic structures, such as cells, microplastics, or nanostructured materials. This could expand its applicability to a broader range of scientific investigations.

**Keywords:** Object detection. Convolutional neural network. Nanoparticle. Transmission electron microscopy. Computer vision.

## LISTA DE FIGURAS

Figura 1.1	Imagem de medição de nanopartículas.....	9
Figura 2.1	Imagem parcial de um microscópio eletrônico de transmissão. ....	13
Figura 2.2	Diagrama do funcionamento de um microscópio eletrônico de varredura....	14
Figura 2.3	Convolução de um <i>kernel</i> em uma imagem.....	16
Figura 2.4	Diagrama de uma rede neural simples.....	17
Figura 2.5	Diagrama de uma rede convolucional simples. ....	20
Figura 2.6	Execução de um <i>maxpooling</i> . ....	21
Figura 2.7	Achatamento de um <i>feature map</i> de uma rede convolucional. ....	21
Figura 2.8	Processo de detecção YOLO. ....	25
Figura 4.1	Fluxograma da avaliação dos modelos. ....	32
Figura 4.2	Redimensionamento independente da altura e largura. ....	33
Figura 4.3	Exemplos das imagens de nanopartículas de bismuto obtidas. ....	33
Figura 4.4	Processo de extração de caixas delimitadoras. ....	34
Figura 4.5	Histograma dos tamanhos de diâmetro das nanopartículas. ....	35
Figura 4.6	Histograma da quantidade de nanopartículas. ....	36
Figura 4.7	Identificação do diâmetro a partir da <i>bounding box</i> em partículas achatadas.	36
Figura 4.8	Identificação de <i>bounding box</i> adequada. ....	38
Figura 5.1	Predições do modelo YOLO OBB após a segunda fase. ....	43
Figura 5.2	Predições do melhor modelo. ....	44
Figura 5.3	Predições do <i>simple blob detector</i> . ....	45

## LISTA DE TABELAS

Tabela 5.1	Número de épocas executado com cada modelo. ....	40
Tabela 5.2	Avaliação dos modelos na detecção de nanopartículas. ....	41
Tabela 5.3	Avaliação dos modelos na medição de diâmetros de nanopartículas. ....	43

## LISTA DE ABREVIATURAS E SIGLAS

SEM	<i>Scanning electron microscope</i>
TEM	<i>Transmission electron microscope</i>
ANN	<i>Artificial neural network</i>
DNN	<i>Deep neural network</i>
CNN	<i>Convolutional neural network</i>
YOLO	You Only Look Once
NMS	<i>Non-maximum suppression</i>
HSV	<i>Hue, saturation, value</i>
R-CNN	Region-based Convolutional Neural Network
SSD	Single Shot MultiBox Detector
DETR	DEtection TRansformer
CIoU	<i>Complete intersection over union</i>
DFL	<i>Distribution focal loss</i>
VFL	<i>Varied focal loss</i>

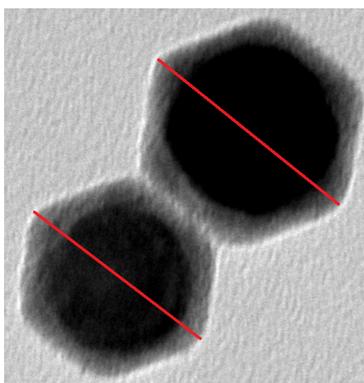
## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>9</b>
<b>2 FUNDAMENTAÇÃO TEÓRICA</b> .....	<b>12</b>
<b>2.1 Imagens de Microscopia Eletrônica</b> .....	<b>12</b>
<b>2.2 Processamento de Imagens</b> .....	<b>15</b>
<b>2.3 Redes Neurais Artificiais</b> .....	<b>16</b>
2.3.1 Redes Neurais Profundas .....	18
2.3.2 Redes Neurais Convolucionais .....	19
2.3.3 Detecção de Objetos .....	21
2.3.4 Aprendizagem por Transferência e <i>Fine-Tuning</i> .....	23
<b>2.4 YOLO - You Only Look Once</b> .....	<b>24</b>
<b>2.5 Evolução das Técnicas de Detecção de Objetos</b> .....	<b>26</b>
<b>3 TRABALHOS RELACIONADOS</b> .....	<b>28</b>
<b>3.1 Visão Geral de Detecção de Objetos</b> .....	<b>28</b>
<b>3.2 Detecção em Microscopia</b> .....	<b>30</b>
<b>4 METODOLOGIA</b> .....	<b>32</b>
<b>4.1 Construção do Dataset</b> .....	<b>32</b>
4.1.1 Construção da Anotação Padronizada .....	34
<b>4.2 Treinamento do Modelo YOLOv8</b> .....	<b>37</b>
<b>5 EXPERIMENTOS</b> .....	<b>40</b>
<b>5.1 Avaliação da Detecção de Objetos</b> .....	<b>41</b>
<b>5.2 Avaliação da Medição de Nanopartículas</b> .....	<b>42</b>
<b>5.3 Avaliação Geral</b> .....	<b>42</b>
<b>6 CONCLUSÃO</b> .....	<b>46</b>
<b>6.1 Desafios da Pesquisa</b> .....	<b>46</b>
<b>6.2 Trabalhos Futuros</b> .....	<b>48</b>
<b>REFERÊNCIAS</b> .....	<b>50</b>

## 1 INTRODUÇÃO

As nanopartículas são materiais de escala nanométrica. Devido a essa escala, essas partículas exibem propriedades físicas, químicas e biológicas singulares, que diferem substancialmente das observadas em materiais em escala macroscópica. Tais propriedades são influenciadas pela elevada relação entre a área superficial e o volume, pelos efeitos quânticos e pela predominância de átomos em sua superfície, tornando as nanopartículas de grande relevância para investigações em diversos campos científicos e industriais. Algumas de suas aplicações incluem criação de materiais de construção, como concreto (LI; ZHANG; OU, 2006) e (TIWARI; CHOWDHURY, 2013), e vetorização de terapêuticos (LOCKMAN et al., 2002). Um exemplo de medição de nanopartícula está presente na Figura 1.1.

Figura 1.1: Imagem de nanopartículas de bismuto gerada a partir de um microscópio eletrônico de transmissão e a medição dos seus maiores diâmetros.



Fonte: NanoPUCRS, adaptado

Entretanto, embora as propriedades singulares associadas ao tamanho das nanopartículas abram diversas possibilidades, elas também impõem desafios consideráveis. A determinação precisa do tamanho das nanopartículas é crucial para a compreensão de suas propriedades e comportamentos (HORNYAK et al., 2008, Cap. 5). A variabilidade na distribuição de tamanhos pode resultar em inconsistências no comportamento das nanopartículas, limitando a reprodutibilidade e a escalabilidade das suas aplicações. Dessa forma, técnicas de medição precisas são essenciais tanto para a pesquisa quanto para as aplicações práticas das nanopartículas. No contexto das nanopartículas, esse processo é frequentemente realizado manualmente e pode exigir centenas de medições. As ferramentas atualmente empregadas demandam várias horas para a contagem manual, enquanto a automação poderia reduzir significativamente esse tempo, permitindo que os pesquisado-

res dediquem-se a outras atividades.

O objetivo deste estudo é desenvolver uma solução para a medição do tamanho de nanopartículas de bismuto em imagens de microscopia eletrônica de transmissão. Para solucionar esse problema, existem várias abordagens possíveis para tratamento de imagens e detecção de características (XING et al., 2018; EDE, 2021). Alguns desses métodos visam aprimorar o contraste das imagens, definir os limites dos objetos e diminuir o ruído das imagens. Outros métodos visam aumentar a precisão de detecção, como a segmentação semântica, que realiza a segmentação e rotulação precisa dos objetos em uma imagem. O aprendizado por transferência também contribui para a melhoria do desempenho do modelo ao adaptar redes previamente treinadas em conjuntos de dados de microscopia menores (EDE, 2021).

Na abordagem adotada neste trabalho, o problema de medição foi formulado como uma tarefa de detecção de objetos, na qual o diâmetro de nanopartículas circulares é inferido a partir da caixa delimitadora do objeto, ou *bounding box*. As imagens utilizadas no conjunto de dados deste trabalho foram cedidas pelo Centro Interdisciplinar de Nanociência e Micro-Nanotecnologia da Pontifícia Universidade Católica do Rio Grande do Sul (NanoPUCRS). A partir disso, um modelo de visão computacional, YOLOv8 (ULTRALYTICS, 2023), foi explorado para automatizar, ao menos parcialmente, a detecção e medição do tamanho das nanopartículas em imagens de microscopia, garantindo precisão e eficiência em dados com diferentes qualidades de imagem e superando as limitações da contagem manual. Esse incremento na eficiência permitirá que os pesquisadores processem conjuntos de dados maiores em prazos mais curtos, facilitando estudos envolvendo um número elevado de partículas.

YOLOv8 é um modelo de detecção de objetos pré-treinado com dados de imagem cotidianas. Especificamente, é baseado em redes neurais convolucionais e treinado com o conjunto de dados COCO (LIN et al., 2014). Para realizar essa tarefa de detecção, o modelo YOLOv8 foi investigado tanto para detecção de *bounding boxes* quanto de *bounding boxes* orientadas, incorporando os pesos pré-treinados do conjunto de dados COCO, várias técnicas de aumento de dados e duas fases de treinamento.

Os experimentos foram realizados para o propósito de uma comparação entre diferentes modelos de detecção: o *simple blob detector* disponível no OpenCV, um modelo YOLOv8 padrão, um modelo YOLOv8 para *bounding boxes orientadas* e um modelo YOLOv8 para *bounding boxes orientadas*, com o uso de aumento de dados *offline* adicional. Entre os resultados obtidos, o YOLOv8 com caixas delimitadoras orientadas, somente

uma etapa de treinamento e sem aumento de dados *offline* adicional foi selecionado como o modelo de melhor desempenho, alcançando o melhor resultado na medição e o segundo melhor na detecção das nanopartículas entre os modelos baseados em *deep learning*.

Ao democratizar o acesso a ferramentas automatizadas de medição, essa implementação de código publicamente disponível <sup>1</sup> e baixo custo poderia contribuir para a criação de condições mais equilibradas entre grupos de pesquisa menores e instituições com recursos limitados, promovendo uma análise mais ampla e acessível para estudos sobre nanopartículas. Ademais, a metodologia desenvolvida neste estudo pode, no futuro, ser adaptada para a análise de outras estruturas microscópicas, como células, microplásticos ou materiais nanoestruturados, ampliando gradualmente as possibilidades de aplicação desta pesquisa.

Este trabalho está organizado da seguinte maneira: o Capítulo 2 aborda conceitos fundamentais para o entendimento deste trabalho; Capítulo 3 discorre sobre trabalhos relacionados; Capítulo 4 explica a metodologia utilizada na padronização dos dados e no treinamento do modelo de detecção de objetos; Capítulo 5 descreve os experimentos baseados na metodologia explicada na capítulo anterior; Capítulo 6 discute sobre os resultados obtidos nos experimentos e conclui o trabalho.

---

<sup>1</sup>Endereço do repositório para o código fonte do detector: <<https://github.com/ultralytics/ultralytics>>. Endereço do repositório para o projeto do detector: <<https://github.com/izaias-saturnino/yolo-nanoparticle>>

## 2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, são discutidos os conceitos e as teorias fundamentais que embasam as metodologias e abordagens adotadas neste estudo. As seções apresentadas fornecem, em conjunto, a base teórica que sustenta a compreensão das técnicas e tecnologias empregadas neste trabalho.

Inicialmente, são abordadas as características das imagens obtidas por microscopia eletrônica na Seção 2.1, as quais constituem o conjunto de dados utilizado nesta pesquisa. Em seguida, na Seção 2.2, é explorado o processamento de imagens, com ênfase nas técnicas empregadas para manipular e extrair informações relevantes a partir dessas imagens.

Na Seção 2.3, são apresentados os conceitos fundamentais sobre redes neurais artificiais, detalhando seus princípios básicos e as diferentes arquiteturas existentes, incluindo redes neurais profundas e redes neurais convolucionais. Também é discutida a tarefa de detecção de objetos.

Por fim, na Seção 2.4, é introduzido o modelo You Only Look Once (YOLO), um algoritmo de última geração para detecção de objetos em tempo real, utilizado na metodologia para a análise das imagens de microscopia eletrônica.

### 2.1 Imagens de Microscopia Eletrônica

Imagens de microscopia eletrônica são geradas por um microscópio eletrônico e possuem certas características relativas à forma de operação do equipamento. A geração dessas imagens é um processo de caracterização usando sonda de elétrons, que interage com o material da amostra para produzir imagens de alta resolução. Os métodos de caracterização por sonda de elétrons são especialmente úteis em nanociência devido à sua capacidade de analisar a estrutura de partículas extremamente pequenas, permitindo a visualização de objetos na escala nanométrica (HORNYAK et al., 2008, Cap. 3). O alto nível de ampliação desses microscópios é essencial para visualizar nanopartículas, permitindo aos pesquisadores observar estruturas que não são visíveis com microscopia óptica.

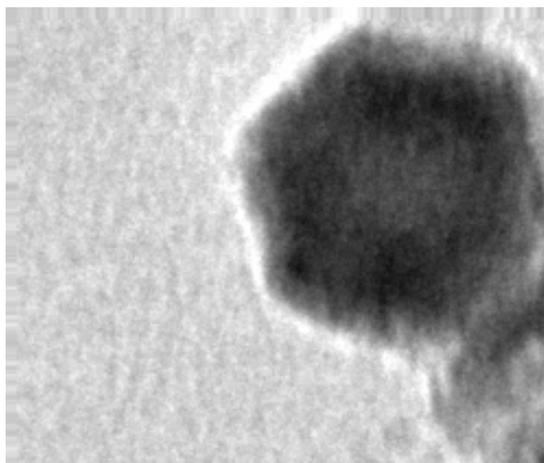
Existem diferentes tipos de microscópios eletrônicos usados na detecção de nanopartículas, como o microscópio eletrônico de varredura (*scanning electron microscope*, SEM) e o microscópio eletrônico de transmissão (*transmission electron microscope*, TEM). O SEM é geralmente utilizado para estudos da morfologia de superfície, enquanto

o TEM pode revelar estruturas internas ao transmitir elétrons através de amostras ultrafinais (HORNYAK et al., 2008, Cap. 3). Cada tipo tem vantagens específicas, dependendo da amostra e do resultado de imagem desejado.

Os elétrons interagem com a matéria de várias maneiras, influenciados pela estrutura do material com o qual o elétron interage. Essas interações são utilizadas para gerar contraste e intensidade na imagem. A intensidade das interações é analisada em cada ponto da superfície ou do volume do material, e, quando combinados, esses pontos formam uma imagem (HORNYAK et al., 2008, Cap. 3).

Mesmo com essa capacidade de magnificação, existem desafios na microscopia eletrônica, como artefatos e ruídos que podem afetar a qualidade da imagem. Esses problemas podem resultar da preparação da amostra, das interações com o feixe de elétrons, entre outros (HORNYAK et al., 2008, Cap. 3). Um exemplo de imagem ruidosa obtida através de um TEM está apresentado na Figura 2.1.

Figura 2.1: Imagem parcial de um microscópio eletrônico de transmissão.



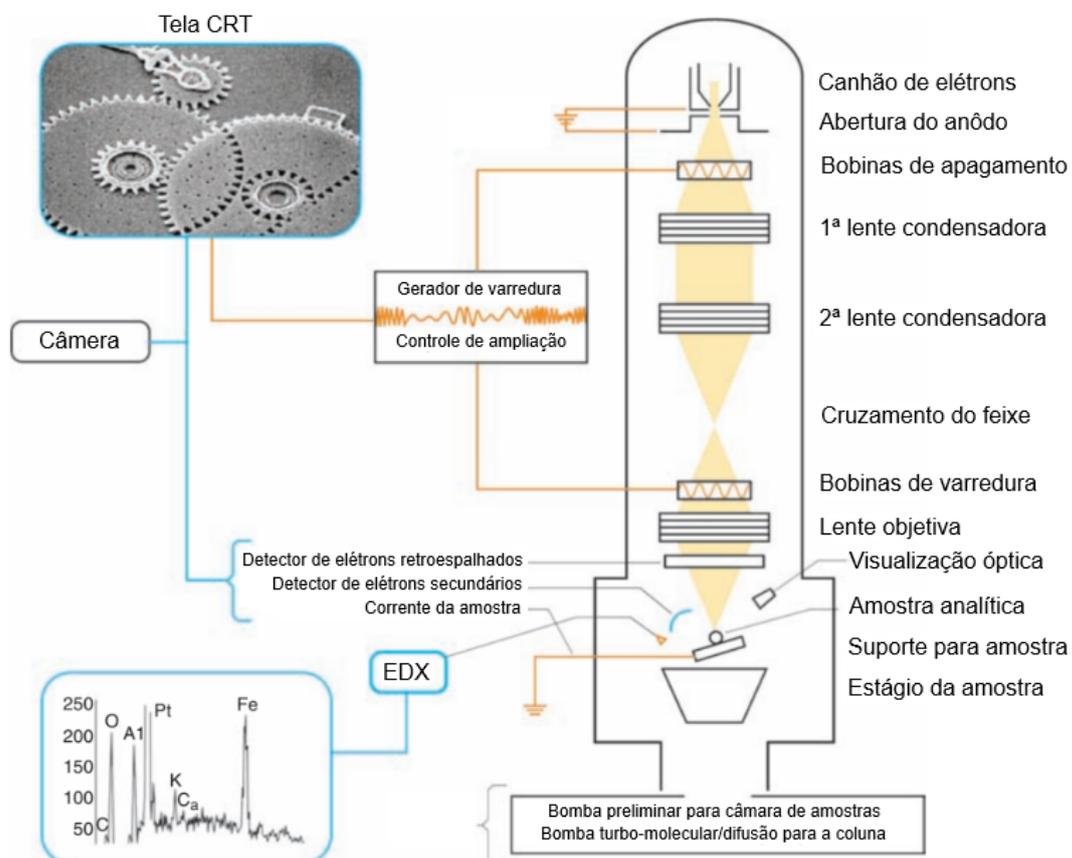
Fonte: NanoPUCRS, adaptado

O processo de aquisição de imagens geralmente envolve a varredura em grade no SEM, onde o feixe de elétrons é direcionado em um padrão de grade na amostra. Esse processo está ilustrado na Figura 2.2.

No TEM, um feixe transmitido passa pela amostra para criar uma imagem com base nas variações de densidade de elétrons (HORNYAK et al., 2008, Cap. 3). Essas imagens podem ser processadas posteriormente para realçar características de interesse, como contornos de nanopartículas ou defeitos.

Além da análise qualitativa, as imagens de microscopia eletrônica também permitem medições quantitativas, como determinação do tamanho, forma e distribuição de

Figura 2.2: Diagrama do funcionamento de um microscópio eletrônico de varredura (SEM). Um feixe de elétrons de alta potência, originado do canhão de elétrons, é acelerado na direção da amostra. Detectores próximos e ao redor de uma amostra coletam e assimilam sinais secundários. Se o SEM possuir capacidade de espectroscopia de raios X por dispersão de energia (EDX), essa análise também é realizada.



Fonte: (HORNYAK et al., 2008, Fig. 3.6, adaptado)

nanopartículas. Essas medições são vitais para compreender as propriedades dos nanomateriais e seu comportamento (HORNYAK et al., 2008, Cap. 5). Essa diversidade de comportamentos possibilita sua utilização em diversas aplicações, nas quais o tipo do material e o tamanho das partículas impactam a compatibilidade (LOCKMAN et al., 2002), flexibilidade (LI; ZHANG; OU, 2006) e resistência (TIWARI; CHOWDHURY, 2013), entre outras características do produto final. A microscopia eletrônica é, portanto, uma ferramenta indispensável na caracterização de nanopartículas e desempenha um papel crucial na nanociência e na nanotecnologia.

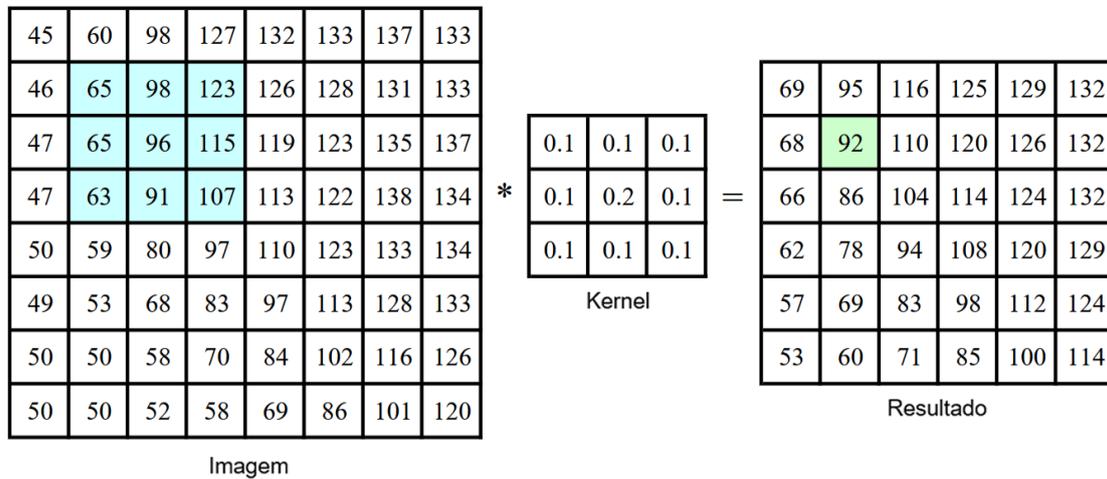
## 2.2 Processamento de Imagens

O processamento de imagens é indispensável para a visão computacional e concentra-se na manipulação e análise de dados visuais para extrair informações, elevar a qualidade da imagem ou desempenhar tarefas específicas (SZELISKI, 2010, Cap. 3). Nessa área, existem algoritmos fundamentais que sustentam o desenvolvimento de algoritmos de detecção de objetos. Processos básicos em processamento de imagens, como convolução e detecção de bordas são essenciais para a identificação e localização de objetos (SZELISKI, 2010, Cap. 5).

A convolução é uma operação fundamental no processamento de imagens, onde um filtro, ou *kernel*, é aplicado à imagem para extrair suas características. Este processo envolve percorrer o filtro sobre a imagem, destacando certos padrões ou características correspondentes à composição específica do filtro, como bordas, texturas ou formas. À medida que um filtro extrai uma característica, essa operação também reduz a dimensionalidade da imagem, simplificando sua representação (SZELISKI, 2010, Cap. 5). Como a operação atua sobre grupos de *pixels* e gera um único valor para cada grupo, a imagem resultante possui menos *pixels*. Por meio de convoluções, estruturas relevantes de uma imagem são destacadas, auxiliando na diferenciação de objetos. Convoluções para entradas com múltiplos canais, ou com profundidade, são adaptadas para essas entradas ao empregar um *kernel* que possua o mesmo número de canais da entrada. A Figura 2.3 apresenta um exemplo de convolução em uma imagem com somente um canal.

Embora técnicas de processamento de imagens sejam eficazes em determinados cenários, elas enfrentam limitações ao lidar com imagens complexas e objetos parcialmente ocultos por sobreposição. Essas limitações se devem à criação manual de filtros para detectar características da imagem. Esse problema não ocorre com redes neurais

Figura 2.3: Convolução de um *kernel* em uma imagem de um canal, produzindo um resultado.



Fonte: (SZELISKI, 2010, Fig. 3.10, adaptado)

convolucionais, detalhadas na Subseção 2.3.2, devido a capacidade de aprender filtros que aumentem o poder de predição do modelo.

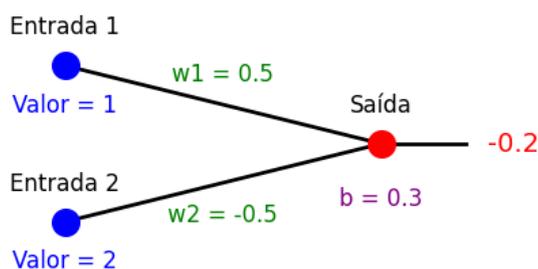
## 2.3 Redes Neurais Artificiais

As redes neurais artificiais, ou *artificial neural networks* (ANNs), são modelos computacionais inspirados pelos mecanismos de processamento de informações do cérebro humano (HAYKIN, 2001). Elas são utilizadas no aprendizado de máquina para resolver tarefas como classificação, regressão e reconhecimento de padrões. O conceito fundamental por trás das redes neurais é a capacidade de aprender com os dados, ajustando iterativamente os parâmetros do modelo durante o treinamento para otimizar a precisão das previsões.

Redes neurais artificiais são compostas por neurônios artificiais, que servem como unidades computacionais básicas (HAYKIN, 2001). Cada neurônio recebe valores de entrada, processa-os e gera um valor de saída. O somatório de todas as entradas multiplicadas pelos seus respectivos pesos e somado ao viés, ou *bias*, é equivalente ao potencial de ativação do neurônio. Esse potencial de ativação é o estímulo da função de ativação, que determina o valor de saída do neurônio com base no potencial recebido. O diagrama da Figura 2.4 exemplifica o seu funcionamento.

A ativação final do neurônio é determinada pelo potencial de ativação do neurônio

Figura 2.4: Diagrama de uma rede neural simples, composta por uma camada de entrada e uma camada de saída com apenas um neurônio. Cada neurônio de entrada é rotulado como “Entrada”, o neurônio de saída é rotulado como “Saída” e os valores das entradas são rotulados como “Valor”. A rede opera com base em conexões ponderadas entre os neurônios de entrada e saída, nas quais o peso de cada conexão ( $w$ ) é representado e associado ao respectivo neurônio de entrada. Além disso, um termo de viés ( $b$ ) é incluído no modelo, que modifica a soma ponderada antes que a função de ativação calcule seu valor final. A saída da rede é o resultado da aplicação de uma função de ativação, linear neste caso, à soma ponderada das entradas e ao viés.



Fonte: O autor

e pela função de ativação. Formalmente,  $z = b + \sum_i w_i x_i$ , onde  $b$  representa o viés,  $w_i$  o peso da entrada  $i$ ,  $x_i$  o valor da entrada  $i$  e  $z$  o potencial de ativação. A saída é então representada por  $f(z)$ , onde  $f$  representa a função de ativação do neurônio (HAYKIN, 2001).

O processo de treinamento de uma rede neural envolve a modificação dos pesos e vieses para minimizar uma métrica de erro definida, conhecida como função de perda, ou *loss function*. A função de perda agrega os erros individuais da rede para cada entrada, considerando a discrepância entre as previsões da rede e os valores reais esperados. Uma técnica de otimização utilizada para treinamento é o gradiente descendente, onde os gradientes da função de perda em relação aos pesos são calculados e os pesos são atualizados na direção que reduz o erro (HAYKIN, 2001). Essa redução do erro por meio do gradiente é multiplicada por uma taxa de aprendizado, ou *learning rate*, que é utilizada para garantir que o modelo siga na direção que minimiza a função de perda, sem que haja risco de comprometer sua trajetória no processo de otimização. A *learning rate* pode ser adaptativa, e o parâmetro que regula sua velocidade de adaptação é denominado *momentum* (PRINCE, 2023, Cap. 6).

No treinamento de ANNs é possível incorporar técnicas de aumento de dados. Aumento de dados consiste em criar novos dados artificialmente a partir dos dados originais, onde aumento *online* refere-se a criação automatizada de novos dados a cada época de treinamento e aumento *offline* a criação de um conjunto adicional de dados fixo que é

reutilizado.

Definições relevantes para a discussão de ANNs incluem as camadas totalmente conectadas, as camadas ocultas e o *multilayer perceptron* (MLP). ANNs apresentam camadas totalmente conectadas quando todos os neurônios de uma camada subsequente estão interconectados com todos os neurônios da camada anterior. As camadas intermediárias, também denominadas camadas ocultas, são aquelas que não pertencem à camada de entrada ou à camada de saída. MLP é uma arquitetura específica de rede neural que possui, no mínimo, uma camada oculta, sendo composta exclusivamente por camadas totalmente conectadas.

### 2.3.1 Redes Neurais Profundas

Uma rede neural profunda, ou *deep neural network* (DNN), é um tipo de rede neural caracterizada pela presença de várias camadas ocultas entre a camada de entrada e a camada de saída (PRINCE, 2023, Cap. 4). Essas camadas permitem que as DNNs capturem e modelem padrões complexos de dados. Ao incorporar camadas adicionais, uma DNN pode aprender características mais abstratas e sofisticadas. De forma geral, aumentar o número de camadas melhora a capacidade da rede de representar relacionamentos mais complexos.

A inferência, ou *forward propagation*, é o processo pelo qual os dados de entrada são passados sequencialmente por cada camada da rede, produzindo uma saída. A retropropagação, ou *back propagation*, sugere um processo oposto. Esse processo é utilizado para treinar redes neurais profundas e envolve calcular o erro na saída da rede e usar esse erro para atualizar os pesos e os vieses da rede (PRINCE, 2023, Cap. 7). *back propagation* emprega a regra da cadeia do cálculo diferencial para determinar o gradiente da função de perda em relação a cada peso e viés. Esses gradientes indicam os ajustes necessários nos pesos para minimizar o erro.

O *overfitting* refere-se a uma situação em que um modelo apresenta um desempenho elevado nos dados de treinamento, mas demonstra uma capacidade reduzida de generalização quando aplicado a dados de teste não observados (PRINCE, 2023, Cap. 2). Isso ocorre quando o modelo aprende uma ativação que é particularmente eficaz para os dados de treinamento, mas não aprende padrões que podem ser generalizados para novas instâncias. O *overfitting* ocorre notavelmente em modelos de aprendizado de máquina que são excessivamente complexos em relação aos dados disponíveis. Redes com menos

camadas são menos propensas ao *overfitting*, pois não têm a capacidade de adequar-se perfeitamente aos detalhes dos dados de treinamento. No entanto, redes menores podem ter dificuldades em capturar relações complexas. Para mitigar o *overfitting* enquanto permite-se uma maior complexidade do modelo, técnicas de regularização, como aumento de dados, são comumente utilizadas (PRINCE, 2023, Cap. 9).

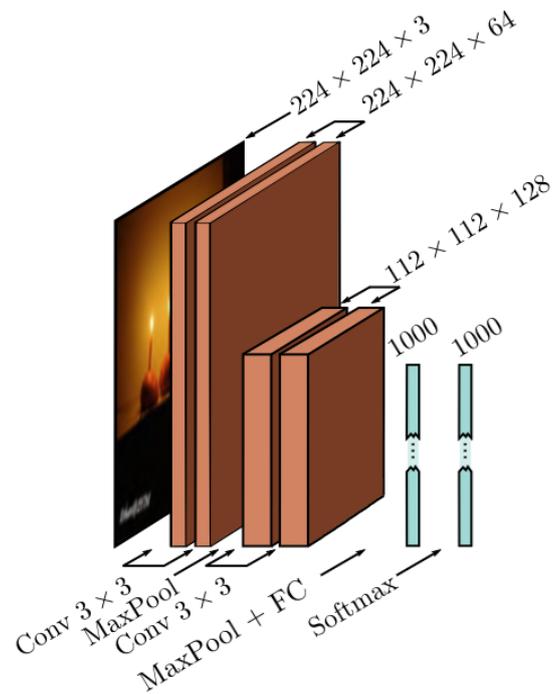
### 2.3.2 Redes Neurais Convolucionais

Imagens apresentam uma estrutura espacial local, ou seja, pontos próximos exercem influência mútua. As redes neurais convolucionais, ou *convolutional neural networks* (CNNs), foram projetadas especificamente para aproveitar a estrutura espacial local presente nas imagens, oferecendo uma abordagem mais eficiente e eficaz em comparação com DNNs tradicionais em tarefas visuais (PRINCE, 2023, Cap. 10). Diferentemente de DNNs clássicas, que conectam cada pixel da imagem a cada neurônio na rede, as CNNs utilizam um filtro de convolução, ou *kernel*. Ao efetuar a convolução, esse filtro extrai características da imagem e essas informações são registradas como saída, assim como discutido na Seção 2.2. Isso significa que o mesmo filtro ou *kernel* é aplicado repetidamente a diferentes partes da imagem, reduzindo significativamente o número de parâmetros que precisam ser aprendidos. A Figura 2.5 ilustra a arquitetura de uma CNN simples.

As CNNs aprendem estruturas hierárquicas de características, nas quais as camadas iniciais capturam padrões elementares, evoluindo progressivamente para representações mais complexas nas camadas mais profundas (PRINCE, 2023, Cap. 4). As primeiras camadas convolucionais da rede normalmente detectam características simples, como bordas, cantos e texturas, enquanto camadas mais profundas combinam essas características em abstrações de nível superior, como partes de objetos ou objetos inteiros. Portanto, a saída de uma camada em uma CNN é um mapa de características gerado a partir da interação dessas camadas, o qual possui informações para a execução da tarefa aprendida.

Além das camadas convolucionais, CNNs possuem camadas de subamostragem, ou *pooling*, que são responsáveis por reduzir a resolução espacial dos mapas de características gerados nas etapas anteriores (PRINCE, 2023, Cap. 10). Dessa forma, *pooling* retém informações agregadas resumindo as regiões locais do mapa de características, preservando padrões importantes e reduzindo as dimensões espaciais e a complexidade computacional. Um exemplo de *max pooling*, um tipo de *pooling* onde se mantém o maior

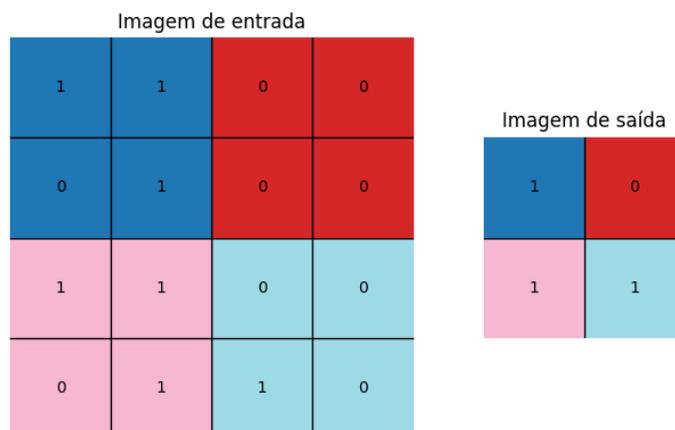
Figura 2.5: Diagrama de uma rede convolucional simples, composta por uma camada de entrada, camadas de convolução e *maxpooling*, achatamento através de uma camada totalmente conectada e uma camada de saída obtida após a aplicação da função de ativação *softmax*. Cada camada apresenta suas dimensões indicadas acima de sua representação, enquanto os processos que conectam as camadas estão ilustrados abaixo delas.



Fonte: (PRINCE, 2023, Fig. 10.17, adaptado)

valor, está ilustrado na Figura 2.6.

Figura 2.6: Execução de um *maxpooling* de *stride* dois em uma imagem simplificada.



Fonte: O autor

Após a aplicação de múltiplas camadas convolucionais e de *pooling*, as características de alto nível extraídas da imagem são achatadas e submetidas a uma ou mais camadas totalmente conectadas, como descrito em um MLP no início da Seção 2.3. Essas camadas são responsáveis pelo cálculo final da previsão, com base nas características aprendidas nas etapas anteriores. Um exemplo de achatamento está presente na Figura 2.7.

Figura 2.7: Achatamento de um *feature map* de uma rede convolucional.



Fonte: O autor

### 2.3.3 Detecção de Objetos

O tipo mais fundamental de tarefa envolvendo redes neurais consiste em realizar previsões com base em dados tabulares. Um exemplo simples é a previsão de preços de

imóveis, levando em consideração variáveis como área em metros quadrados, localização, número de cômodos, entre outras. Este tipo de tarefa é frequentemente abordado por meio de redes neurais totalmente conectadas, nas quais cada neurônio de uma camada está interconectado com todos os neurônios da camada anterior. O modelo, então, aprende as relações entre as variáveis de entrada e as variáveis de saída (PRINCE, 2023, Cap. 1).

Uma possível abordagem para aplicar uma rede neural a uma imagem consiste em achatá-la em um vetor unidimensional. Após ser achatada, a imagem pode ser alimentada em uma rede neural. O modelo aprenderia a classificar a imagem com base nos valores dos *pixels*.

Embora as DNNs possam ser aplicadas a tarefas relacionadas a imagens, as CNNs são mais apropriadas para a detecção de objetos. A principal distinção entre as duas reside na maneira como a rede lida com a estrutura espacial das imagens. As CNNs são projetadas com camadas convolucionais, que permitem aprender hierarquias espaciais de características, tornando-as particularmente adequadas para tarefas relacionadas a imagens (PRINCE, 2023, Cap. 10).

A classificação de imagens é uma tarefa relativamente simples, pois seu objetivo é gerar apenas um rótulo por imagem. Nessa tarefa, a predição consiste em gerar um valor de confiança, ou *score*, para cada classe que possa ser prevista pelo classificador, sendo que o *score* da predição corresponde ao maior valor entre todos os valores de confiança atribuídos (PRINCE, 2023, Cap. 1). Em contraste, a detecção de objetos é uma tarefa mais complexa, pois exige que a rede trate múltiplos objetos, com diferentes tamanhos e posições na imagem (PRINCE, 2023, Cap. 10).

Em algumas aplicações, os objetos podem estar orientados em direções variadas. Nesses casos, a regressão de *bounding boxes* orientadas é utilizada. Nesse caso, além de prever a posição, largura e altura, o modelo também prevê o ângulo de rotação da *bounding box*. Essa abordagem permite que a *bounding box* se ajuste adequadamente a objetos que não estão alinhados com os eixos da imagem.

Outro desenvolvimento nas técnicas de detecção de objetos é o uso de heurísticas para remoção de *bounding boxes* redundantes e de baixa confiança (PRINCE, 2023, Cap. 10), ou *non-maximum suppression* (NMS). Quando várias *bounding boxes* são previstas para o mesmo objeto, elas frequentemente se sobrepõem. Para evitar isso, seleciona-se a caixa com o maior *score* e suprime-se todas as outras caixas que possuem uma grande sobreposição com a caixa selecionada. Esse processo garante que as caixas com maior confiança permaneçam, melhorando a precisão da detecção de objetos e evitando a dupli-

cação de detecções (HOIEM; CHODPATHUMWAN; DAI, 2012). O parâmetro utilizado para determinar se a sobreposição é suficientemente significativa é o limiar de interseção sobre união, ou *intersection over union (IoU) threshold*. Para prevenir a detecção de objetos com *scores* baixos, o limiar de confiança, ou *confidence threshold*, é um parâmetro que define se um objeto detectado deve ser incluído nos resultados finais ou descartado. Esse parâmetro corresponde à pontuação mínima necessária para que uma detecção seja considerada válida. O limite é aplicado à saída de um modelo após a realização das previsões (HOIEM; CHODPATHUMWAN; DAI, 2012).

### 2.3.4 Aprendizagem por Transferência e *Fine-Tuning*

A aprendizagem por transferência é uma técnica de aprendizagem de máquina que envolve o aproveitamento do conhecimento adquirido de uma tarefa ou domínio (SZELISKI, 2010, Cap. 5). Essa técnica baseia-se na ideia de que certos padrões ou recursos aprendidos de um conjunto de dados podem ser úteis quando aplicados a outro, mesmo que as tarefas ou domínios sejam diferentes em alguns aspectos (PRINCE, 2023, Cap. 9). Essa é diferente de um simples treinamento, onde um modelo é tipicamente treinado do zero em um conjunto de dados novo.

Essa técnica é particularmente vantajosa em situações em que os dados para a tarefa de destino são limitados, mas dados abundantes estão disponíveis para uma tarefa relacionada (PRINCE, 2023, Cap. 9). Ao transferir conhecimento da tarefa de origem, a aprendizagem por transferência pode melhorar o desempenho do modelo na tarefa de destino.

O *fine-tuning* é o processo de adaptação de um modelo pré-treinado para uma tarefa ou domínio específico (PRINCE, 2023, Cap. 9). Inicialmente, o modelo é treinado em um conjunto de dados amplo e genérico, e o *fine-tuning* visa ajustá-lo a um conjunto de dados menor e mais especializado. Durante esse processo, os parâmetros do modelo são atualizados de maneira a refletir de forma mais precisa as características da nova tarefa ou domínio. *Fine-tuning* pode envolver a modificação das camadas finais do modelo ou até mesmo um retreinamento mais profundo, caso o novo domínio apresente diferenças substanciais em relação ao original (SZELISKI, 2010, Cap. 5).

## 2.4 YOLO - You Only Look Once

O modelo You Only Look Once (YOLO) (REDMON et al., 2016) é uma CNN utilizada para detecção de objetos em tempo real. Ele se destaca especialmente por sua velocidade e eficiência em tarefas como reconhecimento de imagens, rastreamento de objetos e detecção. O princípio fundamental dessa rede é analisar a imagem uma única vez, realizando previsões simultâneas para várias *bounding boxes* e seus respectivos rótulos de classe, correspondentes a cada objeto presente.

A entrada da imagem é dividida em uma grade de dimensões  $S \times S$ . Cada célula da grade é responsável por detectar objetos cujo centro esteja localizado dentro da sua área. Para isso, a célula faz previsões para um número fixo de *bounding boxes* e suas respectivas pontuações de confiança para cada classe.

Para cada *bounding box*, o modelo também calcula as probabilidades de classe, ou seja, a probabilidade de o objeto detectado pertencer a uma classe específica. Para obter a previsão final é realizada uma heurística de remoção de *bounding boxes* redundantes e de baixa confiança, ou NMS, assim como descrito na Subseção 2.3.3. O processo de detecção é ilustrado na Figura 2.8.

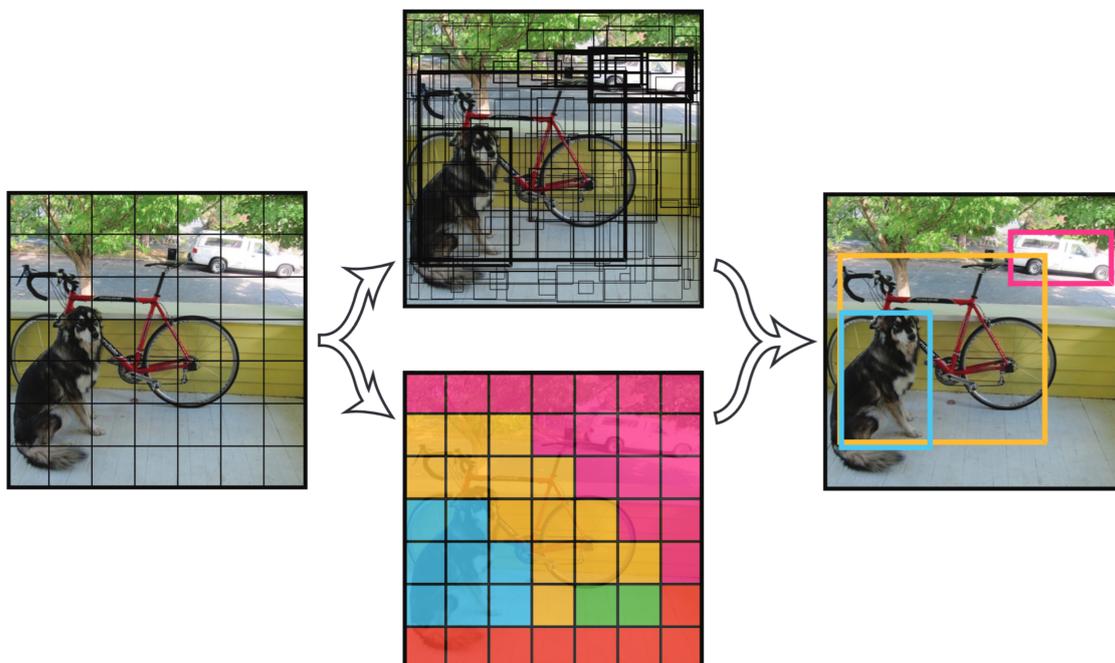
Existem múltiplas versões da rede YOLO, sendo que as versões subsequentes, publicadas posteriormente, permitiram o aprimoramento de sua velocidade e precisão, por meio da implementação de melhorias incrementais em cada nova versão. Neste trabalho, o modelo selecionado para o treinamento foi o YOLOv8 (ULTRALYTICS, 2023), o qual foi especificamente avaliado com o conjunto de dados COCO (LIN et al., 2014). Ao ser lançado, esse modelo alcançou o estado da arte na detecção de objetos aéreos em tempo real (Reis et al., 2023). Esse resultado decorre das modificações na sua composição e no processo de treinamento, incluindo a utilização de um número maior de épocas de treinamento, a adoção de uma nova arquitetura que incorpora variações validadas experimentalmente, bem como sua capacidade aprimorada de aprender o formato das *bounding boxes* (SELÇUK; SERIF, 2023). YOLOv8 possui várias opções para aumento de dados. A configuração padrão do aumento de dados do YOLOv8 consiste em aumento de dados *online*. Os aumentos de dados disponíveis são de variação de canais HSV<sup>1</sup>, translação, redimensionamento, inversão horizontal da imagem e aumento em mosaico<sup>2</sup>.

---

<sup>1</sup>Como HSV se refere a *hue* (matiz), *saturation* (saturação) e *value* (valor), aumento de dados em relação ao contraste está contemplado.

<sup>2</sup>O aumento de dados em mosaico combina quatro imagens diferentes em uma única imagem durante o treinamento.

Figura 2.8: Processo de detecção YOLO. A imagem é segmentada em células de uma grade. Cada célula tem a tarefa de identificar objetos cujo centro esteja localizado dentro de sua área. Para isso, cada célula faz previsões sobre um número pré-determinado de *bounding boxes* e suas respectivas pontuações de confiança para cada classe. Após a previsão das *bounding boxes*, aplica-se uma heurística de *non-maximum suppression* para determinar a previsão final.



Fonte: (REDMON et al., 2016, Fig. 2)

A *loss* do modelo YOLOv8 é composta por três componentes principais: a *complete intersection over union (CIoU) loss*, a *distribution focal loss (DFL)* e a *varied focal loss (VFL)*. A *CIoU loss* é uma função de perda associada à regressão das *bounding boxes*, a *DFL* está relacionada à tarefa de classificação, e a *VFL* ajusta dinamicamente o peso da classificação, com o objetivo de melhorar a detecção de instâncias mais complexas.

A representação dos rótulos no modelo YOLOv8 segue um formato predefinido, no qual cada objeto é descrito pelo seu identificador de classe, seguido pelas coordenadas  $x$  e  $y$  de seu centro, além das dimensões correspondentes à sua largura e altura. Em contrapartida, a representação dos rótulos no modelo YOLOv8 para *bounding boxes* orientadas adota um formato distinto. Nesse caso, cada objeto é caracterizado pelo seu identificador de classe, seguido pelas coordenadas  $x$  e  $y$  de cada um dos quatro vértices que definem a sua *bounding box* orientada.

Embora versões subsequentes tenham sido disponibilizadas, a utilização do YOLOv8 permanece pertinente para o desenvolvimento da metodologia deste trabalho, sendo possível a sua substituição por versões mais recentes, caso seja adequado. O YOLOv8, assim como algumas de suas versões anteriores, oferece a funcionalidade de regressão de *bounding boxes* orientadas, a qual foi explorada neste trabalho como uma alternativa para a detecção de nanopartículas.

## 2.5 Evolução das Técnicas de Detecção de Objetos

No Capítulo 2, foram revisados os fundamentos teóricos das técnicas contemporâneas de detecção de objetos, abordando sua evolução desde os primeiros métodos de processamento de imagens até as avançadas arquiteturas baseadas em aprendizado profundo. Este capítulo abordou inicialmente imagens de microscopia eletrônica e o processamento essencial de imagens, os quais estabeleceram as bases para a compreensão de conceitos mais complexos. As abordagens pioneiras de detecção, como a extração de características e o aprimoramento de imagens, desempenharam um papel crucial na preparação dos dados visuais, enquanto os métodos modernos utilizam algoritmos sofisticados para a identificação de objetos nas imagens.

A exploração das ANNs, especialmente das DNNs e das CNNs, demonstrou como essas arquiteturas revolucionaram a análise de imagens, possibilitando a extração automática de características e superando as limitações das abordagens tradicionais. A transição das técnicas manuais para a aprendizagem de características diretamente nas CNNs repre-

sentou um marco significativo no avanço da detecção de objetos. Entre as metodologias atuais, abordou-se o modelo YOLO.

Por fim, as tecnologias discutidas ao longo deste capítulo evidenciam a evolução da detecção de objetos, desde os métodos tradicionais até os modernos modelos de aprendizado profundo, como o YOLO, que se caracterizam por sua eficiência e por serem soluções *end-to-end*. Esses avanços refletem um esforço contínuo para superar o desafio de identificar objetos em ambientes variados e dinâmicos. No contexto da detecção de nanopartículas, as metodologias aqui apresentadas constituem a base para a aplicação do YOLOv8 em um domínio especializado.

### 3 TRABALHOS RELACIONADOS

Este trabalho tem como objetivo investigar e aplicar métodos de detecção de objetos em nanopartículas, utilizando o modelo YOLO. A detecção precisa e eficiente de nanopartículas representa um desafio considerável, em virtude de suas características microscópicas e da necessidade de alta resolução nas imagens. Nesse sentido, torna-se fundamental analisar o estado da arte e os avanços recentes em técnicas de detecção de objetos, além de explorar soluções específicas voltadas para a microscopia, considerando as variações nas condições e escalas das imagens.

Este capítulo tem como propósito contextualizar a pesquisa, abordando dois temas centrais: a detecção de objetos, na Seção 3.1, que oferece uma visão geral das investigações realizadas na área, e a detecção em microscopia, na Seção 3.2, que discute diretamente aplicações que abordam as particularidades das imagens em alta resolução e em escalas reduzidas. Ao examinar esses dois tópicos, será possível obter uma compreensão mais aprofundada dos desafios envolvidos e das abordagens existentes.

#### 3.1 Visão Geral de Detecção de Objetos

Existem diversas abordagens e arquiteturas de Redes Neurais Convolucionais (CNNs) aplicadas à detecção de objetos. A Region-based Convolutional Neural Network (GIRSHICK et al., 2014) (R-CNN) combinou a capacidade das CNNs com a utilização de propostas de regiões, permitindo uma localização precisa dos objetos dentro de uma imagem. As versões subsequentes da R-CNN, como o Fast R-CNN (GIRSHICK, 2015) e o Faster R-CNN (REN et al., 2016), introduziram melhorias significativas em termos de precisão e eficiência. Essas arquiteturas reduziram o número de predições necessárias, e no caso do Faster R-CNN, incorporaram um método interno para a proposição das regiões dentro da imagem. O Mask R-CNN (HE et al., 2017), uma extensão do Faster R-CNN, foi projetado para realizar segmentação semântica.

Outro modelo relevante é o Single Shot MultiBox Detector (LIU et al., 2016) (SSD). O SSD é denominado *single shot* devido à sua capacidade de localizar e classificar objetos em uma única passagem pela rede, o que o torna mais eficiente em termos de tempo em comparação com os modelos baseados em R-CNN. Esse modelo emprega uma técnica chamada *multibox*, que permite a previsão de múltiplas caixas delimitadoras para cada objeto detectado na imagem.

O SSD opera aplicando filtros convolucionais em mapas de características gerados a partir da imagem de entrada em diferentes escalas. Isso permite que o modelo faça previsões de caixas delimitadoras com níveis variados de granularidade. A rede neural também prevê deslocamentos para um conjunto de caixas delimitadoras padrão, denominadas âncoras, que são predefinidas com diferentes proporções e escalas. Uma das principais vantagens do SSD é sua capacidade de realizar detecção de objetos em tempo real. Uma desvantagem da utilização de âncoras é o elevado erro relativo para objetos de dimensões reduzidas na imagem, uma vez que erros de poucos *pixels* resultam em discrepâncias proporcionalmente significativas.

Além das CNNs, a aplicação de *transformers* na detecção de objetos também é explorada. O DEtection TRansformer (CARION et al., 2020) (DETR) se destaca por sua capacidade de realizar uma análise global da imagem. Ao converter a imagem em *tokens*, o modelo *transformer* prediz as posições e as classes dos objetos com base no contexto global e nas informações derivadas das diferentes partes da imagem. No entanto, esse tipo de modelo apresenta desafios, como a necessidade de grandes volumes de dados, alto custo computacional e longos tempos de treinamento, principalmente devido ao mecanismo de atenção característico dos *transformers*. Algumas variações desse modelo incluem o Deformable DETR (ZHU et al., 2020) e o Swin Transformer (LIU et al., 2021). O Deformable DETR, ao incorporar mecanismos de atenção deformáveis, permite que o modelo direcione sua atenção de maneira mais eficiente para as partes relevantes da imagem. Em contrapartida, o Swin Transformer segmenta a imagem em janelas não sobrepostas, aplicando o mecanismo de atenção apenas dentro de cada janela, o que melhora a eficiência sem comprometer o desempenho.

A detecção de objetos pequenos, especialmente em imagens médicas, apresenta desafios significativos. Entre esses desafios estão a escassez de informações contextuais e a dificuldade em distinguir objetos pequenos do ruído de fundo. Esses problemas surgem devido ao menor número de *pixels* ocupados por tais objetos, dificultando a obtenção do contexto necessário para uma detecção precisa. Além disso, fatores como baixo contraste, oclusões e alta taxa de ruído tornam a tarefa ainda mais complexa. Erros na localização e regressão das caixas delimitadoras também podem comprometer a precisão da detecção. A falta de dados anotados para objetos pequenos é outro obstáculo significativo.

Para enfrentar esses desafios, foram desenvolvidas técnicas como a detecção em múltiplas escalas (LIN et al., 2017; ROSS; DOLLÁR, 2017), modelos sem âncoras (DUAN et al., 2019; LAW; DENG, 2018) e mecanismos de atenção (ZHU et al., 2020). Além

disso, melhorias nas redes *backbone* (HE et al., 2016; TAN; LE, 2019) têm contribuído para melhorar a detecção de objetos pequenos.

O YOLO e suas diversas versões foram aplicados a uma variedade de tarefas. O YOLOv1 foi utilizado na detecção e classificação de resíduos em tempo real (YE et al., 2020) e em um sistema inteligente de binóculos (ZHENG; GE, 2021). Além disso, foi empregado na detecção de glândulas mamárias (RASTOGI; RYUH, 2019). O YOLOv2 foi aplicado na detecção de veículos, com ajustes específicos para o fluxo de tráfego (LI et al., 2018), e também na identificação de imagens e cores de luzes (ZHANG et al., 2018). Por fim, o YOLOv4 foi utilizado na detecção de seres humanos em imagens térmicas (KANNADAGULI, 2020) e na detecção de navios (JIANG et al., 2021).

### 3.2 Detecção em Microscopia

Diversos estudos na área de detecção de objetos em microscopia concentram-se principalmente na identificação de células e doenças (XING et al., 2018). Essas pesquisas empregam diversas técnicas de imagem e abordagens para desenvolver redes neurais voltadas para a detecção de objetos de interesse. A aplicação de detecção de objetos em microscopia por sonda eletrônica, como a microscopia eletrônica de varredura (SEM) ou a microscopia eletrônica de transmissão (TEM), apresenta desafios únicos, especialmente devido às características específicas das imagens. A pesquisa nesse domínio frequentemente depende de conjuntos de dados pequenos e específicos, o que limita a generalização dos modelos (EDE, 2021).

Em pesquisas voltadas à detecção de nanopartículas, um dos métodos explorados utiliza exclusivamente técnicas de processamento de imagem. Esses métodos incluem a detecção de bordas, limiarização e filtragem, como apresentado em (WRIGHT; YOUNG; ARAYA, 2023). Embora eficazes em cenários específicos, esses procedimentos frequentemente enfrentam desafios em condições de imagem complexas, como a sobreposição de partículas ou variações de contraste.

Um estudo relacionado a este trabalho é apresentado em (BADEROT et al., 2022), no qual o uso de *deep learning* é explorado para a detecção de objetos em imagens de microscopia eletrônica. A análise é baseada em um resumo, uma vez que o acesso completo ao artigo está restrito por *paywall*.

A segmentação não supervisionada de nanopartículas também é abordada em (WANG et al., 2021), embora o nível de ruído e oclusão nas imagens analisadas seja

consideravelmente inferior ao das imagens deste estudo. Em (LEE et al., 2020), uma abordagem baseada em algoritmos genéticos é utilizada para otimizar parâmetros de análise de imagem, combinada com a técnica de *stacking*, visando separar objetos do fundo. É importante destacar que, mais uma vez, as imagens desse estudo apresentam níveis de ruído e oclusão mais baixos do que os encontrados neste trabalho.

Além disso, (KOLENOV; PEREIRA, 2020) explora a aplicação de técnicas de aprendizado de máquina não supervisionado em dados de dispersão de Fourier, possibilitando a detecção de nanopartículas sem a necessidade de criação de imagens.

Outros modelos foram empregados para detectar nanopartículas em ambientes com alto nível de ruído através de dados sintéticos (FREI; KRUIS, 2018). Geralmente conjuntos de dados de microscopia eletrônica de transmissão são pequenos (EDE, 2020). Essa escassez também indica o aprendizado por transferência e a adaptação de domínio como técnicas promissoras para soluções com *deep learning*.

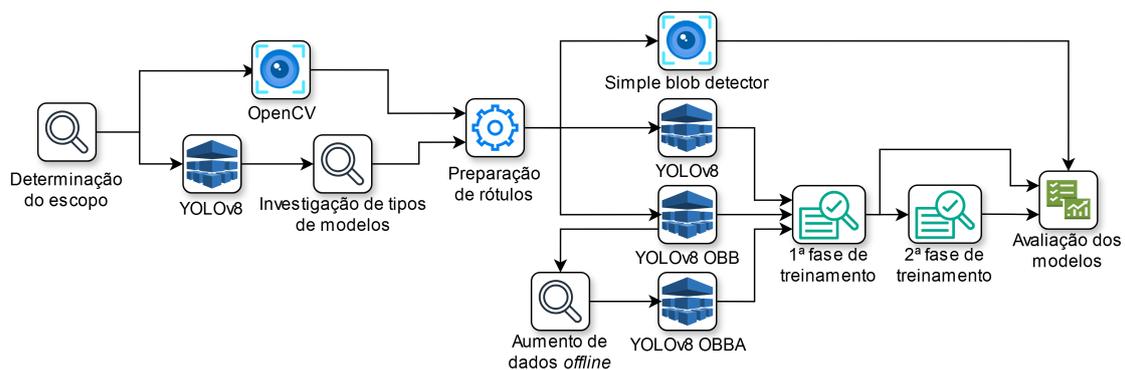
Além dos desafios técnicos de desenvolver modelos de detecção robustos, etapas de pré-processamento, como redução de ruído e remoção de artefatos, são de considerável importância para alcançar resultados confiáveis em microscopia eletrônica. Métodos para aperfeiçoar a qualidade da imagem, como redução de ruído com aplicação de redes neurais para restauração de imagens, foram explorados (BUCHHOLZ et al., 2019).

Este estudo tem como objetivo abordar alguns dos desafios mencionados, desenvolvendo um modelo especificamente adaptado para a detecção de nanopartículas de bismuto em imagens obtidas por TEM. Métodos tradicionais de processamento de imagem apresentam limitações significativas quando confrontados com os efeitos de transparência e oclusões típicos dessa técnica. Abordagens como as R-CNN e suas variantes são caracterizadas por um elevado custo computacional. SSD é útil para detectar objetos em diferentes escalas, mas YOLOv8 utiliza técnicas de múltiplas escalas no seu treinamento. Abordagens como o DETR, embora promissoras, também apresentam alto custo computacional e demandam grandes volumes de dados para treinamento. Ao considerar as restrições únicas dos dados de microscopia eletrônica e utilizar uma arquitetura de redes neurais promissora, este trabalho busca contribuir para o corpo de pesquisa na detecção de objetos em escala nanométrica.

## 4 METODOLOGIA

Este capítulo descreve a metodologia empregada para detecção de nanopartículas usando YOLOv8, detalhando a criação do conjunto de dados, o procedimento de treinamento e as métricas de avaliação. Primeiro, descreve-se o processo de coleta e padronização da anotação de imagens de nanopartículas. Em seguida, discute-se o *pipeline* de treinamento, incluindo técnicas de aumento de dados e as estratégias de treinamento do modelo YOLOv8. Esta metodologia visa fornecer uma solução robusta e eficiente para detectar nanopartículas em imagens complexas, mesmo com poucos dados. Um fluxograma geral da avaliação dos modelos está presente na Figura 4.1. Nessa figura, YOLO OBB corresponde ao modelo que usa *bounding boxes* orientadas e YOLO OBBA corresponde ao modelo que usa um aumento de dados *offline* de redimensionamento independente da altura e largura nas imagens do seu conjunto de dados em adição às características do YOLO OBB. Imagens exemplificando redimensionamentos independentes da altura e largura estão presentes na Figura 4.2.

Figura 4.1: Fluxograma da avaliação dos modelos. Todos os modelos YOLO são modelos de versão 8 (YOLOv8). Simple blob detector refere-se ao detector de blobs simples disponível no OpenCV. YOLO OBB corresponde ao modelo que usa *bounding boxes* orientadas. YOLO OBBA corresponde ao modelo que usa um aumento de dados de redimensionamento independente da altura e largura nas imagens do seu conjunto de dados em adição às características do YOLO OBB.

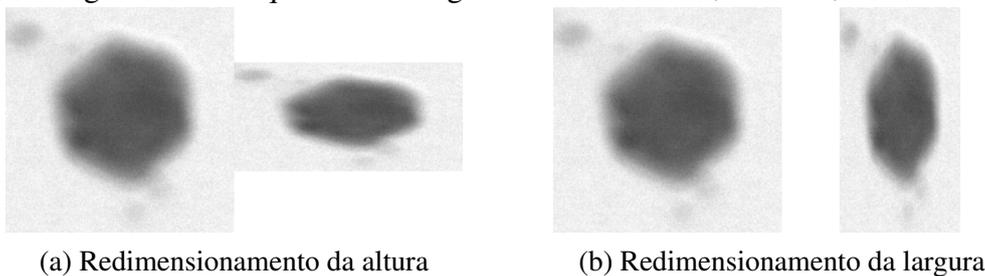


Fonte: O autor

### 4.1 Construção do Dataset

Esta seção dedica-se à descrição da metodologia utilizada na construção do conjunto de dados, detalhando suas etapas. Nesta seção, abordam-se as características dos

Figura 4.2: Exemplos de redimensionamento independente da altura e da largura. A imagem original está à esquerda e a imagem redimensionada, à direita, em cada exemplo.

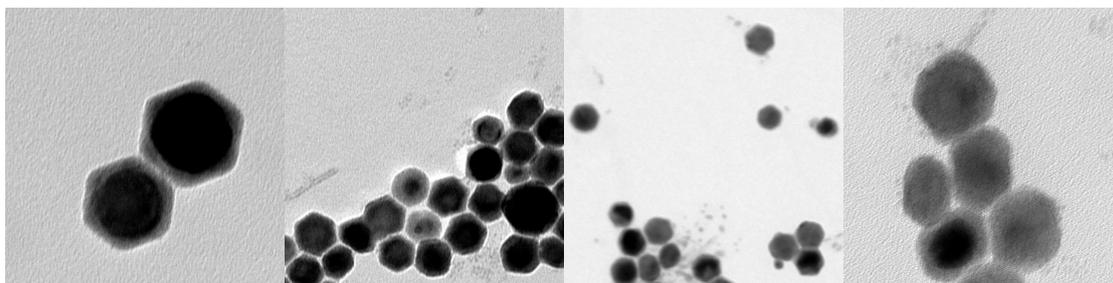


atributos do conjunto de dados. Além disso, a Subseção 4.1.1 trata do processo de anotação dos dados, enfatizando a necessidade de uma abordagem consistente e padronizada. Após a construção criteriosa do conjunto de dados, ele será utilizado no treinamento e na avaliação do modelo YOLOv8, conforme detalhado nas seções seguintes.

As imagens de nanopartículas utilizadas para a criação do conjunto de dados, *dataset*, foram cedidas pelo NanoPUCRS e são resultados de experimentos de síntese de nanopartículas de bismuto.

Em virtude da obtenção das imagens por meio do microscópio de transmissão, as imagens originais possuem formato *grayscale*. Também observa-se a sobreposição de objetos resultante dos efeitos de transparência. Os objetos, ou nanopartículas, presentes nas imagens originais, predominantemente, apresentam formas circulares ou hexagonais e, eventualmente, formas achatadas. Exemplos das imagens obtidas estão presentes na Figura 4.3.

Figura 4.3: Exemplos das imagens de nanopartículas de bismuto obtidas através de um microscópio eletrônico de transmissão. A figura é composta de partes de quatro imagens distintas.

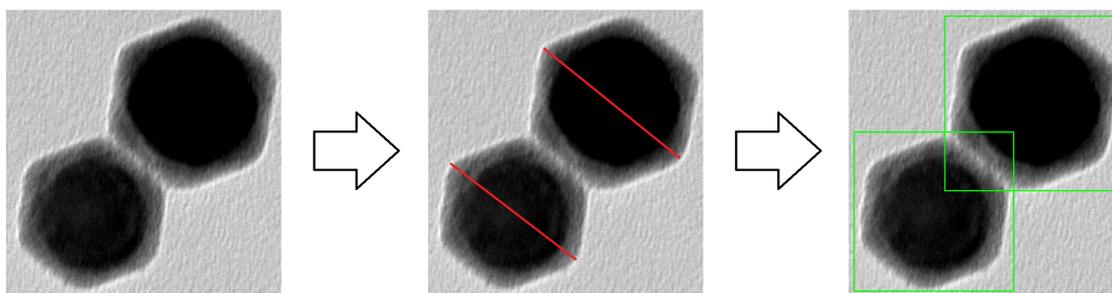


Fonte: NanoPUCRS, adaptado

### 4.1.1 Construção da Anotação Padronizada

As imagens em questão foram acompanhadas de outras imagens com anotações não padronizadas. Essas anotações consistem em segmentos de linha reta coloridos, os quais representam os maiores diâmetros das nanopartículas. A partir das imagens originais e das imagens de anotação obtidas foi criado um *dataset* padronizado. Como a maioria das nanopartículas é circular, a transformação da representação dos segmentos de linha coloridos em uma *bounding box* foi um processo simples. Primeiramente, foram criadas novas imagens contendo regiões das imagens originais, nas quais todos os objetos eram circulares e estavam devidamente anotados. Em seguida, foi aplicado processamento de imagens envolvendo filtragem por cor e detecção de linhas para adquirir as *bounding boxes* a partir das anotações anteriores. Esse processo é ilustrado na Figura 4.4.

Figura 4.4: Processo de extração de caixas delimitadoras. A partir das regiões de imagem selecionadas, é aplicado um processamento de imagens que inclui etapas de filtragem por cor e detecção de linhas, com o objetivo de obter as caixas delimitadoras com base nas anotações previamente realizadas. Após isso, a anotação extraída é pareada com a imagem sem a anotação de linhas coloridas, para fins de treinamento. A anotação das linhas coloridas foi ressaltada para facilitar a visualização.



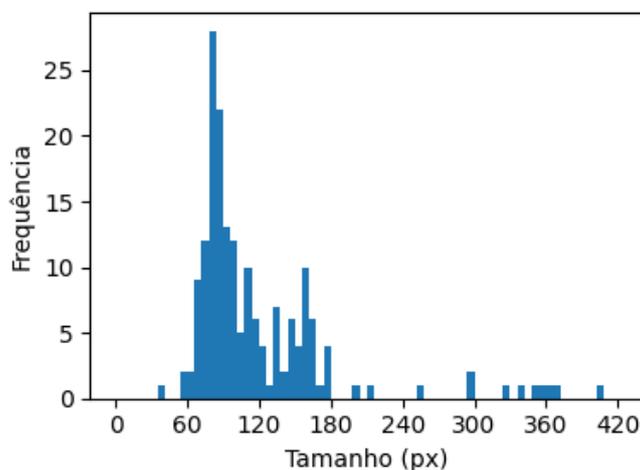
Fonte: O autor

As novas anotações foram padronizadas para o formato YOLO, descrito na Seção 2.4. Após isso essas foram pareadas com imagens das regiões sem segmentos de linha reta coloridos, garantindo que a rede neural aprendesse a detectar as nanopartículas, e não as anotações. Duas anotações foram criadas para os dados obtidos, tanto em formato para *bounding boxes* orientadas quanto para *bounding boxes*. As anotações no formato de *bounding boxes* orientadas consistem apenas em uma conversão de formato, mantendo a orientação original da *bounding box*.

O *dataset* contém apenas nove imagens, o que pode limitar a capacidade de ge-

neralização do modelo. Essa quantidade reduzida deve-se a dificuldades enfrentadas na obtenção de um número maior de imagens. Com isso, diversas técnicas de aumento de dados foram empregadas. As metodologias adotadas para o aumento de dados são discutidas detalhadamente na Seção 4.2. O tamanho total das imagens é de 27,4 MB. As imagens originais possuem dimensões de 2048x2048 *pixels* e as imagens pertencentes ao *dataset* são regiões com anotações completas nessas imagens. O número total de nanopartículas presente nas imagens é 179. A divisão dos dados nos conjuntos de treino, validação e teste foi efetuada de maneira a alocar três imagens por subconjunto, considerando a quantidade de objetos presentes em cada imagem. A distribuição foi efetuada com o objetivo de equilibrar a combinação das características das imagens originais e proporcionar um número maior de objetos nas imagens de treinamento, especialmente, e de validação. Um histograma do tamanho das partículas em todas as imagens está disponível na Figura 4.5. Além disso, um histograma referente à quantidade de partículas presentes em cada imagem pode ser visualizado na Figura 4.6.

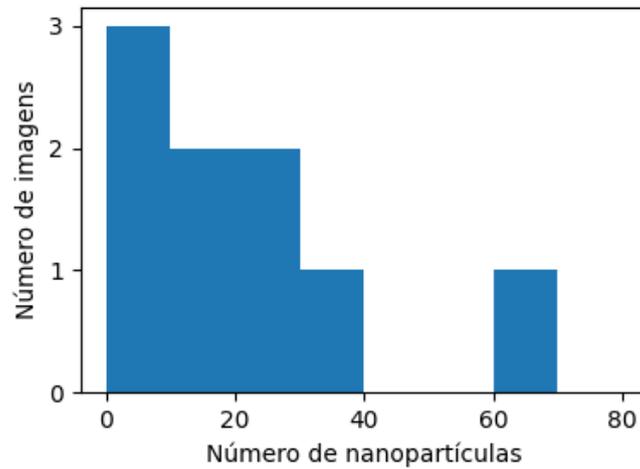
Figura 4.5: Histograma dos tamanhos de diâmetro das nanopartículas em todas as imagens.



Fonte: O autor

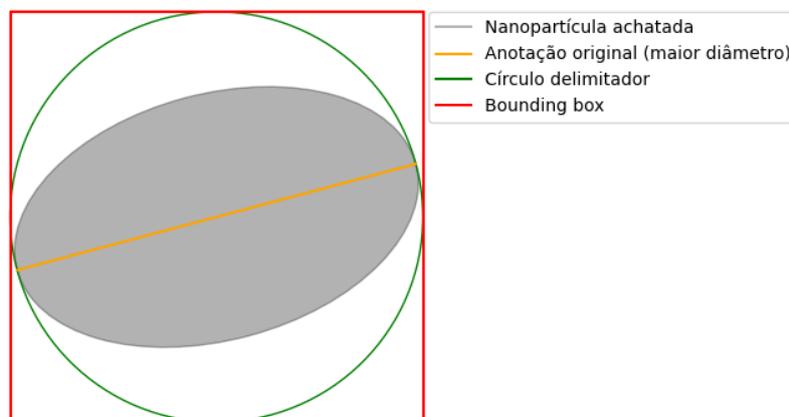
Não foi possível utilizar partículas achatadas, ou elipsoides, no conjunto de dados pois a anotação refletia somente o maior diâmetro, não sendo possível diferenciar o tamanho do outro diâmetro. Como a extração das *bounding boxes* a partir das anotações originais assume que as partículas são circulares, uma *bounding box* apropriada para essas partículas não pôde ser extraída a partir da anotação. Esse comportamento está representado na Figura 4.7.

Figura 4.6: Histograma da quantidade de nanopartículas em cada imagem.



Fonte: O autor

Figura 4.7: Diagrama representativo de uma nanopartícula achatada. Observa-se que não é possível identificar o tamanho do maior diâmetro com a informação obtida a partir da *bounding box* extraída pelo processo de padronização da anotação. É evidente que a informação do outro diâmetro também é necessária para determinar a *bounding box* correta.



Fonte: O autor

Imagens geradas por outros equipamentos também foram consideradas, mas foram descartadas por problemas relacionados à presença de objetos parcialmente oclusos nessas imagens, o que dificultaria o treinamento do modelo de detecção de objetos.

## 4.2 Treinamento do Modelo YOLOv8

No processo de desenvolvimento da metodologia de treinamento, inicialmente foi efetuado o treinamento do modelo com todo o aumento de dados disponível e apropriado. Porém esse método não foi utilizado na metodologia apresentada neste trabalho. Isso porque a utilização de várias técnicas de aumento de dados simultaneamente impossibilitou a aplicação de uma parada antecipada, ou *early stopping*, que evita o treinamento do modelo por várias épocas sem diminuição da *loss*. Essa limitação decorreu da complexidade do aumento de dados escolhido, o qual dificultou a progressão do modelo nas primeiras épocas de treinamento. Para viabilizar o uso de *early stopping*, o processo de treinamento utilizado na metodologia deste trabalho foi organizado em duas fases distintas.

Na primeira fase, o modelo foi inicializado com pesos treinados no *dataset* COCO (LIN et al., 2014). Foi utilizado o modelo YOLOv8 com configuração de treinamento padrão. Ou seja, com aumento de dados padrão, *loss* padrão e otimizador AdamW (LOSHCHILOV; HUTTER, 2019) com *learning rate* de 0,002, *momentum* de 0,9. O treinamento foi conduzido ao longo de 1000 épocas, com a interrupção antecipada após 100 épocas, seguida pela restauração dos melhores pesos encontrados, com o objetivo de evitar a perda de generalização devido a *overfitting*. Na segunda fase, foi aplicado um aumento de dados *online* adicional, o qual incluiu uma rotação de até um grau, a técnica *mixup* com fator 0.5<sup>1</sup> e a inversão vertical com fator 0.5 nas imagens, com o objetivo de aprimorar a robustez do modelo. Os demais hiperparâmetros permaneceram inalterados.

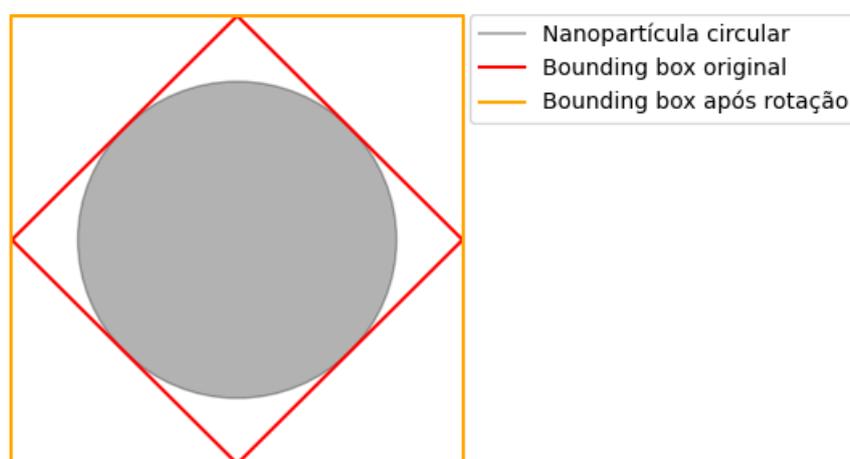
O mesmo treinamento foi repetido em duas ocasiões subsequentes, com modificações. Após o treinamento em duas fases do primeiro modelo, foi realizado o mesmo treinamento em duas fases empregando o modelo YOLOv8 para detecção de *bounding boxes* orientadas juntamente com a padronização dos dados apropriada para *bounding boxes* orientadas. A utilização de *bounding boxes* orientadas proporciona a vantagem de aumentar o limite de rotação das imagens para 180 graus, permitindo realizar qualquer rotação possível. A rotação não foi utilizada em modelos sem *bounding box* orientada com

---

<sup>1</sup>O aumento de dados *mixup* consiste na combinação de duas imagens e seus respectivos rótulos, gerando maior variabilidade visual.

esse grau de liberdade devido ao impacto no formato da *bounding box*, o que reduziria a precisão na medição do diâmetro. Esse impacto é demonstrado visualmente pela Figura 4.8. Observa-se que, apesar disso, o tamanho da *bounding box* verdadeira é invariável em relação à rotação, tratando-se de uma questão meramente de implementação.

Figura 4.8: Diagrama representativo de uma nanopartícula circular. Percebe-se que, ao rotacionar a *bounding box*, uma nova *bounding box* maior é necessária para cobrir toda a área da *bounding box* anterior, mesmo quando uma *bounding box* menor poderia ser utilizada.



Fonte: O autor

Em seguida, foi realizado um experimento adicional seguindo a mesma metodologia, possuindo um conjunto de dados adicional para o modelo orientado. Esse conjunto de dados adicional incorpora aumento de dados *offline* de redimensionamento independente da altura e largura, com o intuito de incrementar o poder preditivo para partículas achatadas. No entanto, essa abordagem adicional resultou em um desempenho inferior na detecção de partículas circulares. Consequentemente, experimentos adicionais que incluíssem conjuntos de dados com partículas achatadas em conjunto com esse aumento de dados foram descartados.

Além disso, constatou-se uma correlação entre a função de perda padrão e o erro relativo médio do diâmetro. Essa correlação foi observada ao constatar que o *overfitting* em uma métrica ocorre concomitantemente com o *overfitting* na outra.

De maneira complementar, utilizou-se um detector de blobs simples do OpenCV (BRADSKI, 2000), ou *simple blob detector*. Esse detector opera em quatro etapas distintas: limiarização, agrupamento, fusão e cálculo de centros e raios. A etapa de limiarização consiste na transformação da imagem original em múltiplas imagens por meio de diversos processos de limiarização. O agrupamento conecta os *pixels* de mesma cor após a

aplicação da limiarização. A fusão refere-se à combinação de *blobs* próximos. Por fim, é realizado o cálculo dos centros e raios dos objetos identificados. Os parâmetros utilizados para a detecção de nanopartículas no conjunto de teste com o uso do *simple blob detector* foram os seguintes: área mínima de 2000, área máxima de 40000, circularidade mínima de 0,2, circularidade máxima de 1, convexidade mínima de 0,1, convexidade máxima de 1, razão mínima de inércia de 0,1 e razão máxima de inércia de 1. Para melhorar o desempenho do detector de *blobs*, foi aplicado um pré-processamento para o aumento do contraste das imagens do *dataset*.

## 5 EXPERIMENTOS

Neste capítulo, apresenta-se a avaliação experimental do modelo YOLOv8 para detecção e medição de objetos de nanopartículas. Os experimentos são organizados em três subseções principais: avaliação de detecção de objetos, avaliação de medição de nanopartículas e avaliação geral do desempenho do modelo.

Na Seção 5.1, avalia-se a capacidade de alguns modelos na detecção de nanopartículas em imagens de microscopia eletrônica, usando métricas padrão como precisão, sensibilidade e *F1-score*. Esta avaliação fornece entendimento sobre a capacidade dos modelos de localizar nanopartículas sob condições variadas.

Na Seção 5.2, concentra-se na capacidade do modelo de medir com precisão o tamanho de nanopartículas. Avalia-se os modelos em relação ao erro de predição do seu diâmetro em relação ao seu tamanho verdadeiro.

Por fim, a Seção 5.3 consolida os resultados das avaliações anteriores, proporcionando uma análise abrangente do desempenho. Trata-se de uma análise comparativa que considera diversos aspectos dos resultados, visando identificar o modelo mais adequado com base nos critérios estabelecidos nas avaliações anteriores, bem como em observações gerais. Ao realizar uma avaliação rigorosa do desempenho dos modelos nas tarefas propostas, este capítulo tem como objetivo evidenciar a eficácia do modelo selecionado.

Todos os experimentos foram conduzidos em um processador convencional, mais especificamente em um computador equipado de 16 GB de memória RAM e um processador Ryzen 5 X, com frequência de 3,6 GHz, 6 núcleos e 12 threads. O número de épocas relativo a cada treinamento está disponível na Tabela 5.1.

Tabela 5.1: Número de épocas executado com cada modelo em cada fase de treinamento.

Detector	Fase	Épocas
YOLO	1	387
YOLO OBB	1	267
YOLO OBBA	1	487
YOLO	2	375
YOLO OBB	2	171
YOLO OBBA	2	134

Fonte: O autor

## 5.1 Avaliação da Detecção de Objetos

Para a avaliação da detecção de objetos dos modelos, foram adotadas as métricas de precisão, sensibilidade e o  $F1$ -score e utilizado o limiar de confiança de 0,5. A precisão mede quantos dos itens que o modelo classificou como positivos são realmente positivos, enquanto a sensibilidade mede quantas instâncias positivas o modelo identificou corretamente. O  $F1$ -score é a média harmônica entre a precisão e a sensibilidade, definido formalmente por  $F_1 = 2 \times \frac{\text{Precisão} \times \text{Sensibilidade}}{\text{Precisão} + \text{Sensibilidade}}$ . O  $F1$ -score equilibra precisão e sensibilidade, fornecendo uma única pontuação que considera tanto falsos positivos quanto falsos negativos. Assim, o melhor modelo para detecção de nanopartículas é aquele que apresenta um  $F1$ -score elevado. A Tabela 5.2 apresenta as métricas obtidas no conjunto de teste referentes a detecção de objetos após cada fase de treinamento, onde as detecções consideradas corretas são aquelas que apresentam o maior valor de IoU e um IoU superior a 0,5.

Tabela 5.2: Avaliação de desempenho dos modelos de detecção de objetos para detecção de nanopartículas após cada fase de treinamento. Todos os modelos YOLO são modelos YOLO de versão 8 (YOLOv8). “SBD” refere-se ao *simple blob detector* disponível no OpenCV. YOLO OBB corresponde ao modelo que usa *bounding boxes* orientadas. YOLO OBBA corresponde ao modelo que usa um aumento de dados de redimensionamento independente da altura e largura nas imagens do seu conjunto de dados em adição às características do YOLO OBB. A tabela apresenta métricas de detecção de objetos, precisão, sensibilidade e  $F1$ -score para cada modelo juntamente com sua respectiva fase de treinamento. As detecções consideradas corretas são aquelas que apresentam o maior valor de IoU e um IoU superior a 0,5. Os melhores valores para cada métrica estão destacados em negrito.

Detector	Fase	Precisão	Sensibilidade	$F1$ -score
SBD	-	<b>100%</b>	12.8%	22.7%
YOLO	1	88.4%	97.4%	92.7%
YOLO OBB	1	90.7%	<b>100%</b>	95.1%
YOLO OBBA	1	86.9%	96.6%	91.5%
YOLO	2	82.6%	97.4%	89.4%
YOLO OBB	2	95%	97.4%	<b>96.2%</b>
YOLO OBBA	2	88.2%	89.7%	89%

Fonte: O autor

## 5.2 Avaliação da Medição de Nanopartículas

Para a avaliação da medição correta dos objetos pelos modelos foram adotadas as métricas do erro relativo médio de predição do diâmetro (E) e seu intervalo de confiança de 99% (IC). O erro relativo de predição do diâmetro foi calculado com base na definição de um diâmetro verdadeiro, ou *ground truth*, para cada anotação. Considerando o método de criação das *bounding boxes*, onde essas são *bounding boxes* para círculos, o diâmetro verdadeiro corresponde ao tamanho de tanto a largura quanto a altura. Após a predição, escolhe-se a maior dimensão entre a altura e a largura da *bounding box predita*, devido à importância da maior dimensão no comportamento físico da partícula, como o diâmetro de predição. Posteriormente, tanto o diâmetro original como o predito são normalizados pelo tamanho do diâmetro original, o que possibilita o cálculo do erro relativo de predição. Este processo é repetido para cada instância. Em seguida, foi realizado o cálculo da média e do intervalo de confiança para a predição do diâmetro, o qual é formalmente definido por  $IC = \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$ , onde  $z_{\alpha/2}$  representa o valor crítico da distribuição normal padrão correspondente ao nível de confiança desejado,  $\sigma$  é o desvio padrão populacional,  $n$  denota o tamanho da amostra, e  $\alpha$  é o nível de significância adotado.

Assim, um modelo com melhor desempenho é aquele que apresenta um erro médio e intervalo de confiança menor. Para a aplicação do modelo em um cenário real, deve-se considerar o erro aceitável. Na metodologia empregado na medição manual, o erro relativo de  $\pm 5\%$  é considerado um erro aceitável. A Tabela 5.3 apresenta as métricas obtidas no conjunto de teste referentes a medição de nanopartículas após cada fase de treinamento.

## 5.3 Avaliação Geral

Apesar da seleção criteriosa das técnicas de aumento de dados aplicadas na segunda fase do treinamento, selecionando somente aquelas que não distorcem o tamanho do diâmetro das nanopartículas, as métricas dessa fase não demonstraram melhorias significativas, como mostrado nas Tabelas 5.2 e 5.3. Os desempenhos permaneceram semelhantes aos obtidos na primeira fase ou até mesmo apresentaram deterioração, o que sugere a ocorrência de *overfitting*. Para visualizar esse comportamento, a Figura 5.1, com predições do modelo YOLO OBB após a segunda fase de treinamento, é apresentada.

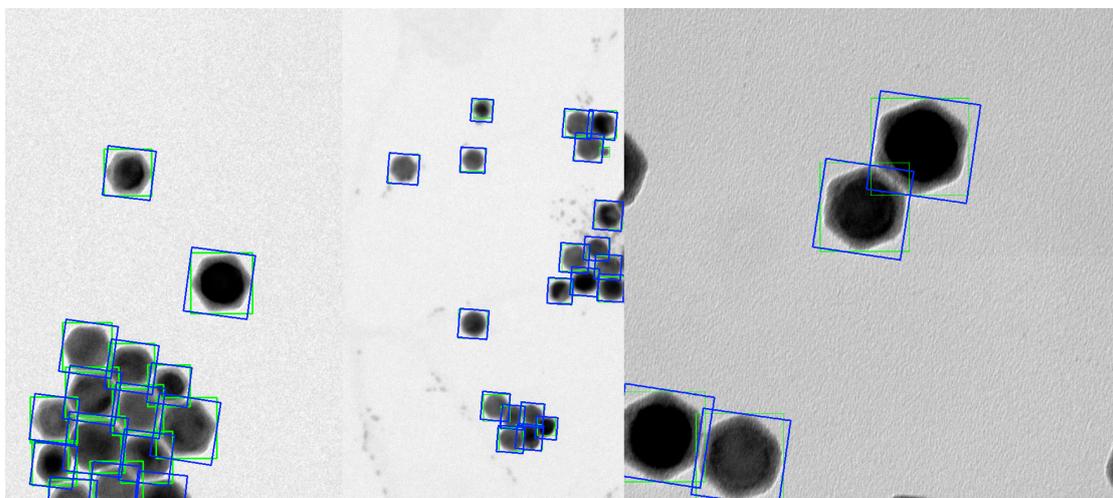
Ao considerar a possibilidade de *overfitting* pelos modelos após a segunda fase de

Tabela 5.3: Avaliação de desempenho dos modelos de detecção de objetos para medição de nanopartículas após cada fase de treinamento. Todos os modelos YOLO são modelos YOLO de versão 8 (YOLOv8). “SBD” refere-se ao *simple blob detector* disponível no OpenCV e está presente para fins de comparação, não possuindo treinamento. YOLO OBB corresponde ao modelo que usa *bounding boxes* orientadas. YOLO OBBA corresponde ao modelo que usa um aumento de dados de redimensionamento independente da altura e largura nas imagens do seu conjunto de dados em adição às características do YOLO OBB. A tabela apresenta o erro relativo médio do diâmetro das predições de diâmetro de nanopartículas (E) e o intervalo de confiança (IC) de 99% para a predição relativa média do diâmetro juntamente com sua respectiva fase de treinamento. Os melhores valores para cada métrica estão destacados em negrito.

Detector	Fase	E	IC
SBD	-	<b>+0.07%</b>	<b>±0.10%</b>
YOLO	1	+5.8%	±10.8%
YOLO OBB	1	+0.5%	±2.6%
YOLO OBBA	1	+3.3%	±5.5%
YOLO	2	+5.5%	±10%
YOLO OBB	2	+3.7%	±9%
YOLO OBBA	2	+3.2%	±7.9%

Fonte: O autor

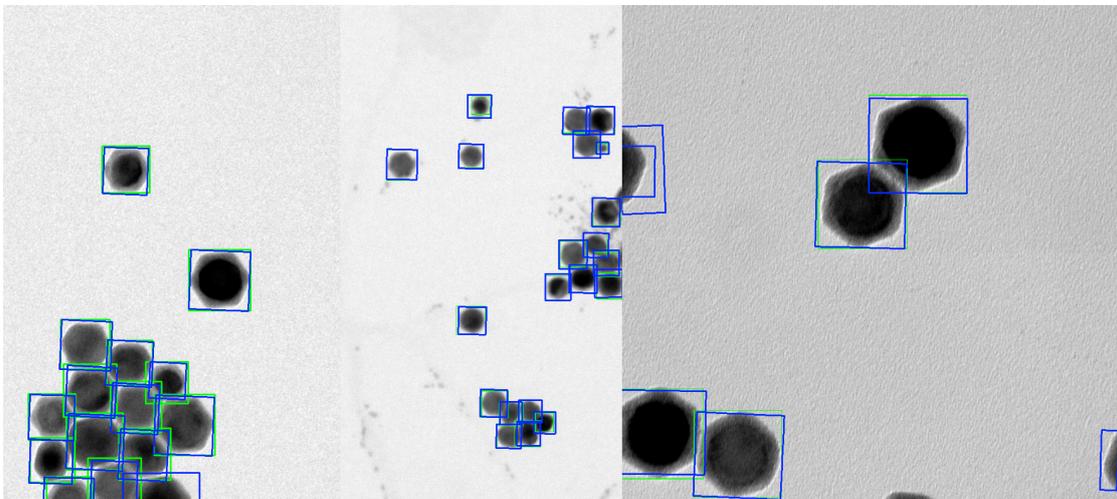
Figura 5.1: Predição de caixas delimitadoras do modelo YOLO OBB após a segunda fase de treinamento em todo o conjunto de teste. As caixas delimitadoras verdadeiras estão em verde enquanto as predições estão em azul.



Fonte: O autor

treinamento e os desempenhos nas métricas obtidas, o modelo YOLO OBB (com *bounding boxes* orientadas, sem o aumento de dados *offline* de redimensionamento independente da altura e largura) foi designado como o modelo de melhor desempenho. Esse modelo também atingiu um erro aceitável de medição ao considerar o erro aceitável da medição manual. Os resultados visuais de suas previsões podem ser observados na Figura 5.2.

Figura 5.2: Predição de caixas delimitadoras do melhor modelo em todo o conjunto de teste. As caixas delimitadoras verdadeiras estão em verde enquanto as previsões estão em azul.



Fonte: O autor

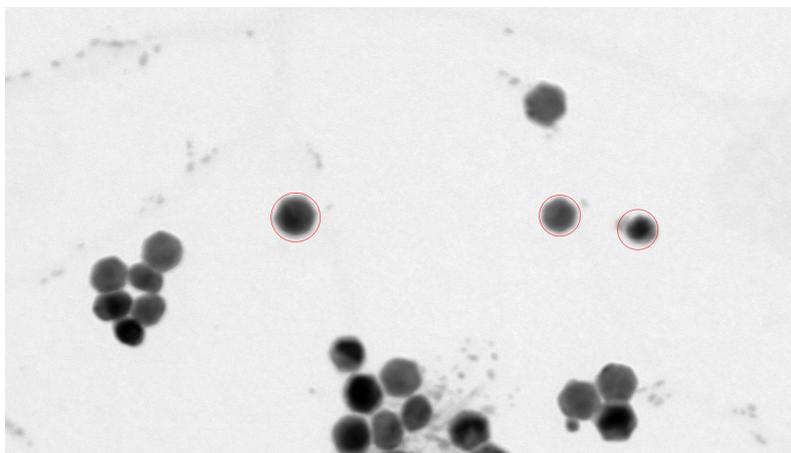
Os resultados numéricos e qualitativos sugerem que o modelo com *bounding boxes* orientadas favorece uma maior capacidade de generalização, possivelmente em razão da maior possibilidade de imagens através da aplicação do aumento de dados. Em contrapartida, o aumento de dados de redimensionamento independente da altura e largura não proporcionou melhorias adicionais. Esse fato pode ser atribuído à natureza do aumento de dados *offline*, que é uma abordagem fixa e carece da variabilidade inerente aos métodos de aumento *online*. Vale destacar que o método de redimensionamento empregado, que ajusta a largura e a altura de forma independente, não é atualmente implementado como uma técnica de aumento de dados *online* no YOLOv8. Em vez disso, o YOLOv8 suporta transformações afins<sup>1</sup> aleatórias de imagens e redimensionamento proporcional de altura

<sup>1</sup>Transformações afins são um tipo de transformação geométrica e envolvem operações de translação, rotação, redimensionamento e cisalhamento em um espaço. Uma transformação afim pode ser expressa pela equação  $y = Ax + b$  onde  $x$  é um vetor de coordenadas no espaço de entrada,  $A$  é uma matriz que representa a parte linear da transformação (como rotação, escala ou cisalhamento),  $b$  é um vetor de translação que desloca os pontos e  $y$  é o vetor de coordenadas resultante após a transformação.

e largura.

Ao analisar os resultados, observa-se que o modelo de melhor desempenho foi eficaz na detecção e medição das nanopartículas nas imagens obtidas por microscopia, embora o tamanho limitado do conjunto de dados possa representar uma restrição à capacidade de generalização do modelo. Em uma análise comparativa com o *simple blob detector*, os modelos YOLO demonstram ser inferiores em cenários onde a precisão, tanto na medição quanto na detecção, é mais crítica do que a sensibilidade, como demonstrado pelos resultados numéricos. No entanto, a sensibilidade evidencia que os modelos YOLO foram mais eficazes na tarefa de detecção, devido à sua capacidade de lidar com um maior número de instâncias e instâncias mais complexas, o que não foi viável com o *simple blob detector*. Um exemplo visual desse comportamento do *simple blob detector* está disponível na Figura 5.3.

Figura 5.3: Predição de círculos delimitadores do *simple blob detector*, disponível no OpenCV, em uma imagem do conjunto de teste. Detecções de nanopartículas estão destacadas em vermelho. Como pode ser observado, somente três nanopartículas foram detectadas nessa imagem.



Fonte: O autor

Devido a limitações de tempo, recursos ou capacidade computacional, não foi possível realizar repetições dos experimentos para avaliar a significância estatística dos resultados obtidos durante o treinamento dos modelos. A ausência de replicações impede uma análise mais robusta da variabilidade dos resultados.

## 6 CONCLUSÃO

No presente estudo, foi realizada a investigação da detecção de nanopartículas de bismuto utilizando o modelo YOLOv8. Durante o desenvolvimento da pesquisa, um dos principais desafios enfrentados foi a escassez de dados, tanto em termos quantitativos quanto na qualidade das anotações. Para contornar essas limitações, foram minuciosamente analisadas e selecionadas diversas técnicas de aumento de dados, além da elaboração de uma ferramenta automatizada para a criação de anotações padronizadas e remoção de nanopartículas achatadas do conjunto de dados.

Os resultados indicam que o YOLOv8, quando treinado com *bounding boxes* orientadas e uma única fase de treinamento, se mostra uma solução promissora, possibilitando a semi-automatização do processo de contagem e medição de nanopartículas. Essa automação pode resultar em uma significativa redução do tempo gasto em tarefas repetitivas e contribuir para um aumento na taxa de análise das nanopartículas. No entanto, algumas limitações ainda persistem, como a falta de significância estatística nos resultados e a ausência de uma avaliação mais robusta sobre o impacto do aumento de dados, o que exigiria múltiplas sessões de treinamento.

### 6.1 Desafios da Pesquisa

Neste trabalho, diversos obstáculos foram encontrados, tanto no processo de pesquisa quanto na execução e avaliação dos resultados. Somente um modelo demonstrou ser capaz de alcançar um erro aceitável para a tarefa de medição, o que ressalta a complexidade ao longo do desenvolvimento do trabalho. Além disso, foram identificados problemas específicos e intrínsecos à natureza da tarefa, que exigem abordagens especializadas para sua resolução, o que não foi antecipado. Tais desafios resultaram em dificuldades adicionais, agravadas por limitações de tempo.

A escolha de imagens obtidas por microscopia eletrônica de varredura foi adotada como a principal fonte de dados para a análise. O processo de padronização dos dados revelou-se relativamente simples ao empregar filtragem por cor e detecção de linhas, resultando na geração das *bounding boxes* correspondentes.

Inicialmente, foi utilizado o detector de blobs simples do OpenCV, o qual foi minuciosamente estudado, com ênfase na análise de todos os seus parâmetros. A investigação experimental do modelo incluiu a seleção dos valores mais adequados para cada

parâmetro, a fim de alcançar resultados satisfatórios nas imagens do conjunto de dados utilizado. Embora o desempenho desse detector fosse razoável, com taxas de sensibilidade e precisão aceitáveis em imagens mais simples, ele não foi capaz de lidar de forma eficaz com instâncias mais complexas. Um dos principais desafios enfrentados foi a oclusão entre as nanopartículas, o que dificultou a produção de bons resultados. Essa limitação estava diretamente relacionada ao funcionamento do algoritmo do detector, o que motivou a exploração de modelos baseados em aprendizado de máquina.

O modelo YOLOv8 foi escolhido para a etapa subsequente da pesquisa, exigindo uma análise detalhada de suas configurações de treinamento. Esse processo envolveu diversas modificações, como o ajuste do número máximo de épocas, a utilização de early stopping e alterações nas configurações de aumento de dados. A falta de conhecimento prévio sobre as melhores escolhas para os hiperparâmetros tornou o processo mais desafiador, demandando uma abordagem qualitativa e experimental. Sempre que possível, realizou-se uma análise qualitativa. No entanto, em muitas situações em que a abordagem qualitativa não era viável, foram realizadas avaliações experimentais para testar o impacto das diferentes configurações nos resultados obtidos. Além disso, para garantir a aplicação adequada do early stopping, o treinamento foi dividido em duas fases, uma vez que a utilização do aumento de dados em uma única fase de treinamento causava um aumento precoce da função de perda, interrompendo o treinamento antes que o modelo tivesse oportunidade de aprender de forma satisfatória.

Durante o processo de otimização, os parâmetros de aumento de dados foram inicialmente mantidos nos valores padrão, sendo ajustados conforme necessário, com base na análise qualitativa das imagens geradas. Algumas transformações de aumento de dados, como a transformação afim e o cisalhamento, foram descartadas por se mostrarem ineficazes para a tarefa em questão. A rotação também passou por ajustes, com a análise qualitativa indicando a necessidade de reduzir o valor máximo de rotação permitido. Com o avanço no entendimento do aumento de dados, foi incorporada uma nova transformação, a inversão vertical da imagem.

Ao alterar o conjunto de dados para imagens provenientes de microscopia eletrônica de transmissão, surgiram novos desafios. A principal vantagem desse novo conjunto era a melhor visualização das oclusões entre as nanopartículas, algo não evidenciado nas imagens anteriores. No entanto, essa mudança demandou uma nova rodada de ajustes experimentais nos parâmetros, como o número máximo de épocas e o early stopping, uma vez que os resultados obtidos nas fases anteriores não eram diretamente aplicáveis ao

novo conjunto de imagens. A análise do desempenho dos modelos revelou que o detector de blobs simples do OpenCV, previamente avaliado, apresentou desempenho inferior com o novo conjunto de dados, com sensibilidade significativamente mais baixa.

A seguir, o modelo YOLOv8 foi novamente explorado, mas desta vez com a utilização de uma versão que incorporava bounding boxes orientadas. Essa alteração exigiu uma nova avaliação experimental dos parâmetros de treinamento, como o número máximo de épocas e o early stopping, para adequar-se às particularidades do novo modelo. Foi identificada a possibilidade de realizar o aumento de dados por rotação com maior flexibilidade, sem comprometer a generalização do modelo, algo que não era viável na versão do YOLOv8 sem orientação. Além disso, foi testado um aumento de dados offline, que consistia no redimensionamento independente das dimensões de altura e largura. No entanto, esse tipo de aumento de dados não resultou em benefícios substanciais, não promovendo uma melhoria significativa nos resultados.

Portanto, a configuração dos modelos, com suas múltiplas tentativas e ajustes, representou um período de considerável complexidade, dada a quantidade de variáveis envolvidas e a necessidade de análises detalhadas para alcançar configurações que se mostrassem eficazes.

## **6.2 Trabalhos Futuros**

Apesar da escassez de dados, os resultados sugerem que o modelo YOLOv8 consegue aprender a tarefa proposta, uma vez que, mesmo com imagens significativamente diferentes no conjunto de teste, o desempenho do modelo foi promissor. No futuro, várias abordagens podem ser exploradas para melhorar a qualidade da detecção de nanopartículas. A disponibilidade de conjuntos de dados bem anotados e diversificados para nanopartículas ainda é limitada, o que pode restringir a capacidade do modelo de detectar imagens de nanopartículas em diversas condições. Para superar essa limitação, esforços serão feitos para coletar dados de várias fontes, abrangendo uma diversidade de formas, tamanhos e materiais de nanopartículas.

Outra possibilidade a ser considerada é a integração do treinamento do modelo com uma ferramenta automatizada ou semi-automatizada de anotação. A anotação manual de grandes conjuntos de dados é uma tarefa demorada, e desenvolver uma ferramenta para automatizar ou auxiliar parcialmente esse processo representaria um avanço significativo. Um sistema semi-automatizado permitiria uma anotação mais rápida e precisa

das imagens de nanopartículas, melhorando a eficiência na criação dos conjuntos de dados. Nessa abordagem, o próprio modelo poderia ser usado para pré-annotar as imagens, com revisores humanos validando e corrigindo as anotações, o que ajudaria a incrementar tanto a quantidade quanto a qualidade dos dados de treinamento e manteria o pipeline de treinamento do modelo sempre atualizado com os dados mais recentes.

Além disso, outra estratégia para trabalhos futuros é explorar o uso de outros modelos para diferentes aplicações do estado da arte, como segmentação semântica, detecção em duas etapas, entre outros. A implementação dessas diferentes abordagens pode gerar delimitações mais precisas, permitindo uma análise detalhada da morfologia das partículas e melhorando o desempenho nas tarefas de análise de tamanho (WANG et al., 2021). Outra linha de pesquisa seria o treinamento do modelo para localizar diferentes tipos de células ou anomalias celulares (VETA; DIEST; PLUIM, 2016). A automação da detecção celular pode acelerar consideravelmente a análise de grandes volumes de imagens microscópicas, oferecendo uma alternativa mais eficiente à análise manual.

Esses esforços visam não apenas ampliar o conjunto de dados para um treinamento mais eficaz, mas também melhorar o desempenho do modelo em aplicações reais, como a caracterização de nanopartículas. Com a detecção mais precisa de nanopartículas em diferentes modalidades de imagem, o modelo pode ter um impacto significativo em áreas como a nanomedicina e a ciência dos materiais, nas quais a detecção de nanopartículas é crucial para a compreensão e manipulação de nanomateriais.

## REFERÊNCIAS

BADEROT, J.; GROULD, M.; MISRA, D.; CLÉMENT, N.; HALLAL, A.; MARTINEZ, S.; FOUCHER, J. Application of deep-learning based techniques for automatic metrology on scanning and transmission electron microscopy images. **Journal of Vacuum Science & Technology B**, AIP Publishing, v. 40, n. 5, 2022.

BRADSKI, G. **The Opencv Library**. [S.l.: s.n.], 2000.

BUCHHOLZ, T.-O.; JORDAN, M.; PIGINO, G.; JUG, F. Cryo-care: Content-aware image restoration for cryo-transmission electron microscopy data. In: IEEE. **2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)**. [S.l.], 2019. p. 502–506.

CARION, N.; MASSA, F.; SYNNAEVE, G.; USUNIER, N.; KIRILLOV, A.; ZAGORUYKO, S. End-to-end object detection with transformers. In: SPRINGER. **European conference on computer vision**. [S.l.], 2020. p. 213–229.

DIWAN, T.; ANIRUDH, G.; TEMBHURNE, J. V. Object detection using yolo: Challenges, architectural successors, datasets and applications. **multimedia Tools and Applications**, Springer, v. 82, n. 6, p. 9243–9275, 2023.

DUAN, K.; BAI, S.; XIE, L.; QI, H.; HUANG, Q.; TIAN, Q. Centernet: Keypoint triplets for object detection. In: **Proceedings of the IEEE/CVF international conference on computer vision**. [S.l.: s.n.], 2019. p. 6569–6578.

EDE, J. M. Warwick electron microscopy datasets. **Machine Learning: Science and Technology**, IOP Publishing, v. 1, n. 4, p. 045003, 2020.

EDE, J. M. Deep learning in electron microscopy. **Machine Learning: Science and Technology**, IOP Publishing, v. 2, n. 1, p. 011004, mar. 2021. ISSN 2632-2153. Disponível em: <<http://dx.doi.org/10.1088/2632-2153/abd614>>.

FREI, M.; KRUIS, F. E. Fully automated primary particle size analysis of agglomerates on transmission electron microscopy images via artificial neural networks. **Powder Technology**, v. 332, p. 120–130, 2018. ISSN 0032-5910. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0032591018302249>>.

GIRSHICK, R. Fast r-cnn. In: **2015 IEEE International Conference on Computer Vision (ICCV)**. [S.l.: s.n.], 2015. p. 1440–1448.

GIRSHICK, R.; DONAHUE, J.; DARRELL, T.; MALIK, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2014. p. 580–587.

HAYKIN, S. **Redes Neurais: Princípios e Prática**. [S.l.]: Bookman Editora, 2001. 27-339 p. ISBN 9788577800865.

HE, K.; GKIOXARI, G.; DOLLÁR, P.; GIRSHICK, R. Mask r-cnn. In: **Proceedings of the IEEE international conference on computer vision**. [S.l.: s.n.], 2017. p. 2961–2969.

- HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2016. p. 770–778.
- HOIEM, D.; CHODPATHUMWAN, Y.; DAI, Q. Diagnosing error in object detectors. In: FITZGIBBON, A.; LAZEBNIK, S.; PERONA, P.; SATO, Y.; SCHMID, C. (Ed.). **Computer Vision – ECCV 2012**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. p. 340–353. ISBN 978-3-642-33712-3.
- HORNYAK, G.; DUTTA, J.; TIBBALS, H.; RAO, A. **Introduction to Nanoscience**. [S.l.]: CRC Press, 2008. 123-140, 237-260 p. ISBN 9781420048063.
- JIANG, J.; FU, X.; QIN, R.; WANG, X.; MA, Z. High-speed lightweight ship detection algorithm based on yolo-v4 for three-channels rgb sar image. **Remote Sensing**, MDPI, v. 13, n. 10, p. 1909, 2021.
- KANNADAGULI, P. Yolo v4 based human detection system using aerial thermal imaging for uav based surveillance applications. In: IEEE. **2020 International Conference on Decision Aid Sciences and Application (DASA)**. [S.l.], 2020. p. 1213–1219.
- KOLENOV, D.; PEREIRA, S. Machine learning techniques applied for the detection of nanoparticles on surfaces using coherent fourier scatterometry. **Optics Express**, Optica Publishing Group, v. 28, n. 13, p. 19163–19186, 2020.
- LAW, H.; DENG, J. Cornernet: Detecting objects as paired keypoints. In: **Proceedings of the European conference on computer vision (ECCV)**. [S.l.: s.n.], 2018. p. 734–750.
- LEE, B.; YOON, S.; LEE, J. W.; KIM, Y.; CHANG, J.; YUN, J.; RO, J. C.; LEE, J.-S.; LEE, J. H. Statistical characterization of the morphologies of nanoparticles through machine learning based electron microscopy image analysis. **ACS nano**, ACS Publications, v. 14, n. 12, p. 17125–17133, 2020.
- LI, H.; ZHANG, M. hua; OU, J. ping. Abrasion resistance of concrete containing nano-particles for pavement. **Wear**, v. 260, n. 11, p. 1262–1266, 2006. ISSN 0043-1648. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0043164805004308>>.
- LI, X.; LIU, Y.; ZHAO, Z.; ZHANG, Y.; HE, L. A deep learning approach of vehicle multitarget detection from traffic video. **Journal of Advanced Transportation**, Wiley Online Library, v. 2018, n. 1, p. 7075814, 2018.
- LIN, T.-Y.; DOLLÁR, P.; GIRSHICK, R.; HE, K.; HARIHARAN, B.; BELONGIE, S. Feature pyramid networks for object detection. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2017. p. 2117–2125.
- LIN, T.-Y.; MAIRE, M.; BELONGIE, S.; HAYS, J.; PERONA, P.; RAMANAN, D.; DOLLÁR, P.; ZITNICK, C. L. Microsoft coco: Common objects in context. In: FLEET, D.; PAJDLA, T.; SCHIELE, B.; TUYTELAARS, T. (Ed.). **Computer Vision – ECCV 2014**. Cham: Springer International Publishing, 2014. p. 740–755. ISBN 978-3-319-10602-1.

- LIU, W.; ANGUELOV, D.; ERHAN, D.; SZEGEDY, C.; REED, S.; FU, C.-Y.; BERG, A. C. Ssd: Single shot multibox detector. In: SPRINGER. **Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14**. [S.l.], 2016. p. 21–37.
- LIU, Z.; LIN, Y.; CAO, Y.; HU, H.; WEI, Y.; ZHANG, Z.; LIN, S.; GUO, B. Swin transformer: Hierarchical vision transformer using shifted windows. In: **Proceedings of the IEEE/CVF international conference on computer vision**. [S.l.: s.n.], 2021. p. 10012–10022.
- LOCKMAN, P.; MUMPER, R.; KHAN, M.; ALLEN, D. Nanoparticle technology for drug delivery across the blood-brain barrier. **Drug Development and Industrial Pharmacy**, v. 28, n. 1, p. 1–13, 2002.
- LOSHCHILOV, I.; HUTTER, F. **Decoupled Weight Decay Regularization**. 2019. Disponível em: <<https://arxiv.org/abs/1711.05101>>.
- PRINCE, S. J. **Understanding Deep Learning**. The MIT Press, 2023. Disponível em: <<http://udlbook.com>>.
- RASTOGI, A.; RYUH, B. S. Teat detection algorithm: Yolo vs. haar-cascade. **Journal of Mechanical Science and Technology**, Springer, v. 33, p. 1869–1874, 2019.
- REDMON, J.; DIVVALA, S.; GIRSHICK, R.; FARHADI, A. You only look once: Unified, real-time object detection. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2016.
- Reis, D.; Kupec, J.; Hong, J.; Daoudi, A. Real-Time Flying Object Detection with YOLOv8. **arXiv e-prints**, p. arXiv:2305.09972, 2023.
- REN, S.; HE, K.; GIRSHICK, R.; SUN, J. Faster r-cnn: Towards real-time object detection with region proposal networks. **IEEE transactions on pattern analysis and machine intelligence**, IEEE, v. 39, n. 6, p. 1137–1149, 2016.
- ROSS, T.-Y.; DOLLÁR, G. Focal loss for dense object detection. In: **proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2017. p. 2980–2988.
- SELÇUK, B.; SERIF, T. A comparison of yolov5 and yolov8 in the context of mobile ui detection. In: \_\_\_\_\_. [S.l.: s.n.], 2023. p. 161–174. ISBN 978-3-031-39763-9.
- SZELISKI, R. **Computer Vision: Algorithms and Applications**. [S.l.]: Springer London, 2010. (Texts in Computer Science). ISBN 9781848829350.
- TAN, M.; LE, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In: PMLR. **International conference on machine learning**. [S.l.], 2019. p. 6105–6114.
- TIWARI, A. K.; CHOWDHURY, S. An overview of application of nanotechnology in construction materials. In: CHAKRABORTY, S.; BHATTACHARYA, G. (Ed.). **Proceedings of the International Symposium on Engineering under Uncertainty: Safety Assessment and Management (ISEUSAM - 2012)**. India: Springer India, 2013. p. 483–492. ISBN 978-81-322-0757-3.

TREDER, K. P.; HUANG, C.; KIM, J. S.; KIRKLAND, A. I. Applications of deep learning in electron microscopy. **Microscopy**, v. 71, p. i100–i115, 2022. ISSN 2050-5701. Disponível em: <<https://doi.org/10.1093/jmicro/dfab043>>.

ULTRALYTICS. **Ultralytics YOLOv8**. 2023. Acesso em: 23 dez. 2024. Disponível em: <<https://docs.ultralytics.com/pt/models/yolov8>>.

VETA, M.; DIEST, P. J. van; PLUIM, J. P. W. Cutting out the middleman: Measuring nuclear area in histopathology slides without segmentation. In: OURSELIN, S.; JOSKOWICZ, L.; SABUNCU, M. R.; UNAL, G.; WELLS, W. (Ed.). **Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016**. Cham: Springer International Publishing, 2016. p. 632–639. ISBN 978-3-319-46723-8.

WANG, X.; LI, J.; HA, H. D.; DAHL, J. C.; ONDRY, J. C.; MORENO-HERNANDEZ, I.; HEAD-GORDON, T.; ALIVISATOS, A. P. Autodetect-mnp: an unsupervised machine learning algorithm for automated analysis of transmission electron microscope images of metal nanoparticles. **Jacs Au**, ACS Publications, v. 1, n. 3, p. 316–327, 2021.

WRIGHT, A. S.; YOUMANS, N. P.; ARAYA, E. F. V. **Recursive Detection and Analysis of Nanoparticles in Scanning Electron Microscopy Images**. 2023. Disponível em: <<https://arxiv.org/abs/2308.08732>>.

XING, F.; XIE, Y.; SU, H.; LIU, F.; YANG, L. Deep learning in microscopy image analysis: A survey. **IEEE Transactions on Neural Networks and Learning Systems**, v. 29, n. 10, p. 4550–4568, 2018.

YE, A.; PANG, B.; JIN, Y.; CUI, J. A yolo-based neural network with vae for intelligent garbage detection and classification. In: **Proceedings of the 2020 3rd International Conference on Algorithms, Computing and Artificial Intelligence**. [S.l.: s.n.], 2020. p. 1–7.

ZHANG, X.; QIU, Z.; HUANG, P.; HU, J.; LUO, J. Application research of yolo v2 combined with color identification. In: IEEE. **2018 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)**. [S.l.], 2018. p. 138–1383.

ZHENG, Y.; GE, J. Binocular intelligent following robot based on yolo-lite. In: EDP SCIENCES. **MATEC web of conferences**. [S.l.], 2021. v. 336, p. 03002.

ZHU, X.; SU, W.; LU, L.; LI, B.; WANG, X.; DAI, J. Deformable detr: Deformable transformers for end-to-end object detection. **arXiv preprint arXiv:2010.04159**, 2020.

ZOU, Z.; CHEN, K.; SHI, Z.; GUO, Y.; YE, J. Object detection in 20 years: A survey. **Proceedings of the IEEE**, IEEE, v. 111, n. 3, p. 257–276, 2023.