



UNIVERSIDADE FEDERAL DO MARANHÃO
Programa de Pós-Graduação em Ciência da Computação

Matheus Levy de Lima Bessa

**Detecção de Ovos de Parasito em Imagens de
Microscópio via Ensemble de Detectores**

São Luís - MA

2025

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).
Diretoria Integrada de Bibliotecas/UFMA

de Lima Bessa, Matheus Levy.

Detecção de Ovos de Parasito em Imagens de Microscópio
via Ensemble de Detectores / Matheus Levy de Lima Bessa. -
2024.

79 f.

Corientador(a) 1: João Dallyson Sousa de Almeida.

Orientador(a): Geraldo Braz Junior.

Dissertação (Mestrado) - Programa de Pós-graduação em
Ciência da Computação/ccet, Universidade Federal do
Maranhão, Ufma, 2024.

1. Detecção. 2. Ovos de Parasito. 3. Imagens de
Microscópio. 4. Ensemble. 5. Rede Neural Convolutacional.
I. Braz Junior, Geraldo. II. Sousa de Almeida, João
Dallyson. III. Título.

Matheus Levy de Lima Bessa

Detecção de Ovos de Parasito em Imagens de Microscópio via Ensemble de Detectores

Dissertação apresentada como requisito parcial para obtenção do título de Mestre em Ciência da Computação, ao Programa de Pós-Graduação em Ciência da Computação, da Universidade Federal do Maranhão.

Programa de Pós-Graduação em Ciência da Computação

Universidade Federal do Maranhão

Orientador: Prof. Dr. Geraldo Braz Junior

Coorientador: Prof. Dr. João Dallyson Sousa de Almeida

São Luís - MA

2025

Matheus Levy de Lima Bessa

Deteccção de Ovos de Parasito em Imagens de Microscópio via Ensemble de Detectores

Dissertação apresentada como requisito parcial para obtenção do título de Mestre em Ciência da Computação, ao Programa de Pós-Graduação em Ciência da Computação, da Universidade Federal do Maranhão.

Trabalho aprovado São Luís - MA, 11 de março de 2025:

Prof. Dr. Geraldo Braz Junior
Orientador
Universidade Federal do Maranhão

Prof. Dr. João Dallyson Sousa de Almeida
Coorientador
Universidade Federal do Maranhão

Prof. Dr. Anselmo Cardoso de Paiva
Examinador Interno
Universidade Federal do Maranhão

Prof. Dr. António Cunha
Examinador Externo
Universidade de Trás-os-Montes e Alto Douro

São Luís - MA
2025

Dedico a todos que conseguiram chegar nesta página.

Agradecimentos

Agradeço ao meu orientador, Geraldo Braz Júnior, que me ajudou ao longo da minha jornada acadêmica até este momento. Também agradeço a meu coorientador João Dallyson Sousa de Almeida, que também me auxiliou durante minha graduação e durante o mestrado. Agradeço também aos meus colegas de laboratório, do VIPLab, que estiveram sempre compartilhando ideias, me apoiando e me fazendo rir durante meus longos anos na UFMA. Não menos importante, agradeço à Universidade Federal do Maranhão e ao PPGCC por me conceder os recursos necessários para tornar esse trabalho realidade.

"Houve um tempo em que o homem enfrentou o universo sozinho e sem amigos. Agora ele tem criaturas para ajudá-lo; criaturas mais fortes que ele próprio, mais fiéis, mais úteis e totalmente devotadas a ele. A humanidade não está mais sozinha."

(Isaac Asimov)

Resumo

As infecções parasitárias intestinais representam um problema de saúde global, especialmente em regiões com condições sanitárias precárias e zonas tropicais. Segundo a Organização Mundial da Saúde (OMS), estima-se que 1,5 bilhão de pessoas sejam afetadas por essas infecções. O diagnóstico manual, realizado por profissionais que analisam amostras microscópicas para identificar e classificar ovos parasitários, é demorado e suscetível a erros, principalmente devido à variação na experiência dos profissionais e à presença de estruturas parasitárias de pequenas dimensões ou parcialmente ocultas por impurezas. Para enfrentar esse desafio, este trabalho propõe uma metodologia baseada em um ensemble que combina uma arquitetura de detecção especializada com detectores genéricos refinados em um conjunto de imagens específico. A arquitetura personalizada, desenvolvida a partir da HRNet-W32, incorpora blocos residuais e composição de perdas intermediárias. Detectores como *Cascade R-CNN*, TOOD e *CenterNet* foram ajustados utilizando o conjunto de dados *Chula-ParasitEgg11*. A fusão das detecções foi realizada por meio da técnica *Weighted Box Fusion*, que integra resultados sem descartar potenciais identificações relevantes. A abordagem alcançou desempenho significativo, com F1 Score de 0,945 e mAP[0.5:0.95] de 0,868, demonstrando potencial para automatizar o diagnóstico com precisão e eficiência.

Palavras-chave: Detecção, Ovos de Parasito, Imagens de Microscópio, Ensemble, Rede Neural Convolutiva.

Abstract

Intestinal parasitic infections represent a global health problem, especially in regions with precarious sanitary conditions and tropical areas. According to the World Health Organization (WHO), it is estimated that 1.5 billion people are affected by these infections. The diagnostic manual, carried out by professionals who analyze microscopic samples to identify and classify parasitic eggs, is time-consuming and susceptible to errors, mainly due to variation in professionals' experience and the presence of small parasitic structures or those partially hidden by impurities. To address this challenge, this work proposes an ensemble-based methodology that combines a specialized detection architecture with generic detectors refined on a specific image set. The custom architecture, developed from HRNet-W32, incorporates residual blocks and loss composition. Detectors such as Cascade R-CNN, TOOD and CenterNet were fine-tuned using the Chula-ParasitEgg11 dataset. The fusion of detections was carried out using the Weighted Box Fusion technique, which integrates results without discarding possible relevant identifications. The approach achieved significant performance, with an F1-Score of 0.945 and mAP[0.5:0.95] of 0.868, demonstrating the potential to automate diagnosis accurately and efficiently.

Keywords: Detection, Parasite Eggs, Microscope Images, Ensemble, Convolutional Neural Network.

Lista de ilustrações

Figura 1 – Amostras por microscópio de ovos de helmintos.	18
Figura 2 – Estados de um ovo de <i>Ascaris lumbricoides</i>	19
Figura 3 – Estágios de desenvolvimento do ovo de <i>Ascaris lumbricoides</i>	24
Figura 4 – Ovo de Tênia.	24
Figura 5 – Ovo de <i>Enterobius vermiculares</i>	25
Figura 6 – Ovo de <i>Capillaria philippinensis</i>	26
Figura 7 – Ovo de <i>Fasciola hepatica</i>	27
Figura 8 – Espécies de helmintos	28
Figura 9 – Amostra em microscópio de <i>Enterobius vermicularis</i>	29
Figura 10 – Tópicos de Detecção de Objetos	29
Figura 11 – Exemplo de relação entre característica e a imagem original	32
Figura 12 – Relação entre centros de âncora no Mapa de Características e a Imagem Original	33
Figura 13 – Relação entre âncora no Mapa de Características e a Imagem Original	34
Figura 14 – Predição de Detecções a partir de um mapa de características	35
Figura 15 – Diferença entre cabeças de detecção acopladas e desacopladas	36
Figura 16 – Arquitetura da <i>Faster R-CNN</i>	38
Figura 17 – Arquitetura da SSD	39
Figura 18 – Arquitetura da YOLO	39
Figura 19 – Arquitetura da <i>Cascade R-CNN</i>	41
Figura 20 – Arquitetura da HRNet	41
Figura 21 – Bloco <i>Bottleneck</i> da <i>ResNet</i>	42
Figura 22 – Exemplo de Curva Precisão x Sensibilidade	43
Figura 23 – Exemplo de Curvas Precisão x Sensibilidade para multi classe	44
Figura 24 – Etapas da Metodologia Proposta.	52
Figura 25 – Amostras das classes da Chula-Parasite-Egg11	53
Figura 26 – Modelo de Otimização da HRNet.	55
Figura 27 – Exemplo de <i>ensemble</i> com o WBF.	58
Figura 28 – Contagem de <i>Bounding Boxes</i> Pequenas, Médias e Grandes.	61
Figura 29 – Exemplos de amostras com artefato.	61
Figura 30 – Amostras com ponto vermelho	62
Figura 31 – Resultados da Validação e Teste para HRNet utilizando bloco <i>Bottleneck</i> da <i>ResNet</i>	63
Figura 32 – Plotagem dos Ovos no Conjunto de Teste em Verde e das Detecções em Vermelho.	67
Figura 33 – Amostras de <i>Hymenolepis d.</i>	69

Figura 34 – Estudo de Caso entre *Ground Truth* (Verde) e Predições (Vermelho)
em Detecções do Conjunto de Teste. 69

Lista de tabelas

Tabela 1 – Sumarização dos Trabalhos Relacionados a Base Chula-ParasiteEgg-11	50
Tabela 2 – Resultados Otimização Utilizando Perda Composta por Média.	63
Tabela 3 – Resultados Otimização Utilizando Perda Composta por Média Ponderada.	64
Tabela 4 – Resultados Otimização Sem Composição de Perda.	65
Tabela 5 – Resultados da Otimização de Cabeças de Detecção.	65
Tabela 6 – Resultados de Modelos Genéricos da Literatura.	66
Tabela 7 – Resultado dos modelos Individuais e Ensemble no Teste.	67
Tabela 8 – Resultados por Classe dos Modelos Individuais e Ensemble (Conjunto de Teste)	68
Tabela 9 – Comparação com Trabalhos Relacionados.	70

Lista de abreviaturas e siglas

PI	<i>Parasitose Intestinal</i>
OMS	<i>Organização Mundial da Saúde</i>
CDC	<i>Centro de Controle e Prevenção de Doenças</i>
HPJ	<i>Hoffman, Pons e Janer</i>
AP	<i>Average Precision</i>
mAP	<i>mean Average Precision</i>
HRNet	<i>High-resolution networks</i>
IoU	<i>Intersection over Union</i>
mIoU	<i>mean Intersection over Union</i>
FPN	<i>Feature Pyramid Network Pyramid Network</i>
SSD	<i>Single Shot MultiBox Detector</i>
DETR	<i>Detection Transformer</i>
SAHI	<i>Slicing Aided Hyper Inference</i>
TOOD	<i>Task Aligned One-Stage Object Detection</i>
WBF	<i>Weighted Box Fusion</i>
OPG	<i>Ovos por Grama</i>
CNN	<i>Convolutional Neural Network</i>
RPN	<i>Region Proposal Network</i>
COCO	<i>Common Objects in Context</i>
SGD	<i>Stochastic Gradient Descent</i>

Sumário

1	INTRODUÇÃO	15
1.1	Definição do Problema	17
1.2	Objetivos	19
1.2.1	Questões de Pesquisa	20
1.2.2	Contribuições	20
1.3	Organização do Trabalho	20
2	FUNDAMENTAÇÃO TEÓRICA	22
2.1	Parasitoses	22
2.1.1	Helmintos	22
2.1.2	Desafios do diagnóstico por microscópio	27
2.2	Detecção de Objetos com Aprendizado Profundo	28
2.2.1	Extração de Características usando Redes Convolucionais para Detecção de Objetos	31
2.2.2	Âncoras e Mapas de Características	32
2.2.3	Cabeça de Detecção	33
2.2.4	Detectores de Objetos	36
2.2.4.1	Faster R-CNN	37
2.2.4.2	SSD	37
2.2.4.3	YOLO	38
2.2.4.4	RetinaNet	40
2.2.4.5	Cascade R-CNN	40
2.2.5	HRNet	41
2.3	Métricas de Avaliação	43
2.4	Considerações Finais	46
3	TRABALHOS RELACIONADOS	47
3.1	Considerações finais	51
4	MATERIAIS E MÉTODO	52
4.1	Aquisição de Imagens	52
4.2	Construção de Arquitetura Específica	53
4.2.1	Construção de <i>Backbone</i>	54
4.3	Fine Tuning de Modelos Genéricos	56
4.4	Ensemble	56
4.5	Avaliação	58

4.6	Considerações Finais	59
5	RESULTADOS E DISCUSSÕES	60
5.1	Aquisição de Imagens	60
5.2	Construção de Arquitetura Específica	62
5.2.1	Perda Composta por Média	62
5.2.2	Perda Composta por Média Ponderada	63
5.2.3	Sem perda Composta	64
5.2.4	Otimização da Cabeça de Detecção	65
5.3	<i>Fine Tuning</i> de Modelos Genéricos	66
5.4	Construção do Ensemble	66
5.5	Discussão	67
6	CONCLUSÃO	72
	REFERÊNCIAS	74

1 Introdução

A Parasitose Intestinal (PI) é uma doença causada por parasitos que se alojam no intestino e se reproduzem através da disposição de ovos (FOCACCIA, 2005). As PI podem ser causadas por diversos micro-organismos e parasitos, podendo ser amebas, bactérias e helmintos ou geo-helmintos (*Ascaris lumbricoides*, *Trichuris trichiura*, *Ancylostoma duodenale* e *Necator americanus*) (FOCACCIA, 2005; PINTO, 2011). A contaminação ocorre por ingestão de água ou alimentos contaminados com o parasito, ou ovos do parasito (PINTO, 2011). Portanto, regiões com baixas condições sanitárias tendem a apresentar maiores índices de infecção. Além disso, é mais prevalente em regiões tropicais e subtropicais (HABIB et al., 2021; HOTEZ, 2021).

Isto torna as PIs um problema de saúde global. A Organização Mundial da Saúde (OMS), estimou cerca de 1,5 bilhão de pessoas infectadas por geo-helmintos em todo o mundo, isto equivale a 24% da população mundial (OMS, 2023). Conforme o relatório da Organização Mundial de Saúde, em 2018, a malária, causada pelo *Plasmodium*, pode provocar 400 mil mortes (ORGANIZATION et al., 2018). Especificamente sobre a *Ascaris lumbricoides*, o Centro de Controle e Prevenção de Doenças (CDC) estima que entre 807 milhões e 1,2 bilhão de pessoas em todo o mundo estejam infectadas (CDC, 2024).

O Brasil configura-se como uma região de risco para doenças parasitárias devido à conjunção de fatores geográficos e socioeconômicos. Sua localização em zona tropical favorece a proliferação de patógenos, cenário agravado por deficiências estruturais no saneamento básico, especialmente em áreas rurais. Dados do Censo Demográfico de 2022, realizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE), revelam que somente 64,69% dos domicílios nacionais possuem acesso à rede de esgoto, com disparidades regionais críticas: na região Norte, esse índice cai para 24,51%. Essa precariedade na cobertura sanitária, associada a condições ambientais propícias, evidencia a vulnerabilidade epidemiológica do país, exigindo abordagens intersetoriais para mitigação de riscos à saúde pública.

Estudos epidemiológicos reforçam essa realidade. Em 2001, Prado et al. (2001) identificou uma taxa de infecção de 66,1% em crianças em Salvador, Bahia. Uma década depois, Belo et al. (2012) reportou 29% de prevalência em amostras de crianças de São João del-Rei, Minas Gerais, com ênfase em infecções em escolas de zonas rurais. Mais recentemente, Alves et al. (2023) evidenciaram altos índices de contaminação por geo-helmintos em Porto Velho, Rondônia, corroborando a persistência do problema. Em uma análise abrangente, Mesquita et al. (2021) destacaram, por meio de revisão sistemática, a disseminação nacional de infecções por geo-helmintos, reforçando a necessidade de

estratégias integradas de controle.

Embora seja uma capital, São Luís do Maranhão apresenta altas taxas de PI em crianças. O estudo [Silva \(2018\)](#), conduzido em três escolas da rede pública de ensino fundamental da cidade, identificou taxas de infecção de 65% por *Ascaris lumbricoides*, 30% por *Trichuris trichiura* e 80% por *Ancylostoma duodenale*.

Estes dados são mais impactantes ao se considerar que as PI em crianças apresentam maiores danos do que em adultos, sendo possível atrasar tanto o desenvolvimento físico como o cognitivo, além de aumentar o risco de mortalidade infantil devido à imunodeficiência ([PAPIER et al., 2014](#)). A progressão da doença pode causar anemia, diarreia, dor abdominal e, em casos mais avançados, obstrução intestinal, que requer intervenção cirúrgica.

O diagnóstico é essencial para garantir uma melhor qualidade de vida. Este é feito via exame parasitológico de fezes utilizando um microscópio para observação direta. A técnica utilizada no preparo da amostra pode variar, porém, todas se baseiam na análise microscópica das amostras em busca de ovos de parasito ([PINTO, 2011](#); [BRASIL et al., 2018](#)). Isto é feito, pois os helmintos depositam uma alta quantidade de ovos diariamente. A *Ascaris lumbricoide* fêmea produz diariamente cerca de 200.000 ovos ([PINTO, 2011](#)).

As técnicas mais comuns de preparo de amostras de diagnósticos de PI são: Método Direto ([BRASIL et al., 2018](#)), Sedimentação ([BRASIL et al., 2018](#)) e Método de Kato-Katz ([KATZ; CHAVES; PELLEGRINO, 1972](#)).

O Método Direto se baseia no preparo de um esfregaço de uma pequena quantidade de fezes em uma lâmina com salina. Para uma melhor visualização, pode-se adicionar lugol (combinação de Iodo Metalóide e Iodeto de Potássio). Apesar de ser uma técnica econômica e pouco trabalhosa, exige maior experiência do profissional, pois não aplica nenhum enriquecimento de contraste para destacar os ovos de parasito ([BRASIL et al., 2018](#)).

Já a Sedimentação ou Método de HPJ (Hoffman, Pons e Janer) dilui 3 gramas de fezes em água, e executa uma filtração utilizando uma gaze dobrado. Esta filtração é repetida até que o líquido filtrado esteja claro. Então, uma amostra do sedimento filtrado na gaze é coletada utilizando uma pipeta e colocada entre a lâmina e a lamínula ([BRASIL et al., 2018](#)). Apesar de utilizar mais material e tempo, esta técnica apresenta sensibilidade superior ao método direto, sendo ideal para ovos de helmintos e cistos de protozoários.

Por último, o Método de Kato-Katz possui sensibilidade igual ou superior aos métodos anteriores. A técnica envolve a seguinte sequência de etapas: uma amostra fecal é colocada sobre papel absorvente, sobre a qual é pressionada uma malha de náilon para filtrar as fezes. Em seguida, utilizando uma espátula, as fezes que passam pela malha são coletadas e depositadas em uma placa quantificadora, formando um pequeno cilindro. Esse cilindro é então coberto com uma lamínula de papel celofane previamente embebida

em solução diafanizadora (FERREIRA; CARVALHO, 1972). Posteriormente, a lâmina é invertida e pressionada contra papel filtro. Por fim, a lâmina é colocada em uma estufa por 4 horas para secagem ((SEVIR); (COVAT); (CEVET), 2022). Neste método, o número de ovos contados é multiplicado pelo fator 24, resultando no número de ovos por grama. Os métodos descritos preparam lâminas que são analisadas em um microscópio óptico.

O preparo da amostra é importante para uma boa análise. Porém, o bom preparo não impede que artefatos indesejados estejam presentes na amostra. Na Figura 1 nota-se a presença de manchas decorrentes do processo de preparo. Observa-se que na amostra de *Trichuris trichiura* as manchas estão sobre o ovo. Essa sobreposição dificulta a detecção da PI pelo técnico laboratorial, causando falsos negativos no diagnóstico. Isso pode ser mais propenso a infecções com escassez ou inconstância da eliminação de ovos e em infecções leves (GOMES, 2022). Adicionalmente, os ovos de um mesmo parasito podem apresentar diferentes formas e texturas a depender do seu estado, como pode ser observado na Figura 2.

Estes fatores exigem que o técnico laboratorial tenha memorizado as características morfológicas e suas variações de diversas espécies de parasitos. O que torna o diagnóstico preciso uma tarefa desafiadora (ANANTRASIRICHAJ et al., 2022). Se tratando assim de uma tarefa de memorização e repetição. Deste modo, esforços na área de visão computacional são feitos para automatizar essa tarefa.

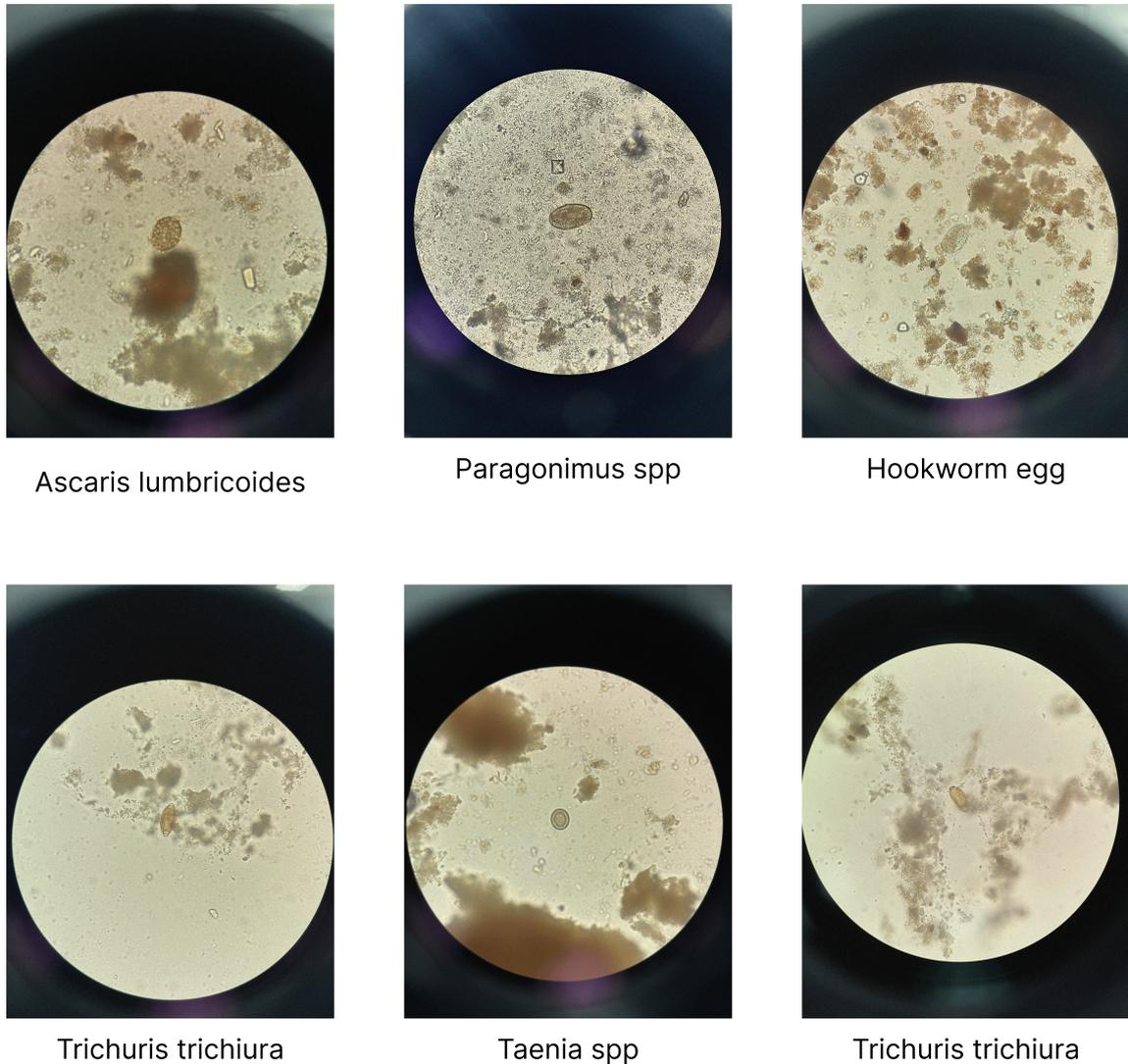
1.1 Definição do Problema

Métodos baseados em aprendizado de máquina com extração manual de características e em aprendizagem profunda foram propostos e obtiveram resultados promissores.

A localização automática de ovos de parasitos em imagens de microscópio foi inicialmente, entre 2001 e 2015, tratada na literatura como um problema de classificação. Nessa abordagem, técnicas de segmentação eram utilizadas para extrair os ovos, seguidas por modelos de redes neurais para classificá-los. Com os avanços no aprendizado profundo e nas redes neurais convolucionais, essa abordagem evoluiu para um problema de detecção de objetos. Além disso, o desenvolvimento de redes de detecção baseadas em aprendizado profundo consolidou o paradigma de detecção de objetos como o padrão predominante. Uma revisão da literatura (KUMAR et al., 2023) destaca como detectores baseados em aprendizado profundo superaram as técnicas tradicionais de aprendizado de máquina.

Apesar dos avanços e esforços para automatizar a detecção de ovos de parasitos, ainda existem lacunas importantes a serem exploradas. Uma das principais é a ausência de uma métrica consolidada para avaliar o desempenho dos modelos. Embora o *mean Average Precision* (mAP) seja amplamente utilizado na detecção de objetos, trabalhos recentes adotam o *F1 Score*, uma métrica tradicionalmente empregada na classificação.

Figura 1 – Amostras por microscópio de ovos de helmintos.



Fonte: [Anantrasirichai et al. \(2022\)](#)

No entanto, o *F1 Score* apresenta limitações significativas, pois não considera a confiança das detecções. Dessa forma, erros com baixa confiança recebem a mesma importância que erros com alta confiança, dificultando uma avaliação do desempenho. O destaque dado ao *F1 Score* na literatura recente está principalmente associado ao desafio [Anantrasirichai et al. \(2022\)](#), realizado em 2022 que o utilizou como uma das métricas.

A segunda lacuna encontra-se na ausência de arquiteturas específicas para detecção de ovos de parasito. Pode-se notar, na literatura recente, o uso de modelos de detecção de propósito geral para a automação da tarefa. Isso indica uma dificuldade em superar o atual estado-da-arte. Portanto, foca-se na combinação de múltiplos detectores a fim de suprimir erros específicos de um modelo. Destacam-se os modelos de características em múltiplas escalas ([AUNG; SRITHAWORN; ACHAKULVISUT, 2022](#)) e modelos

Figura 2 – Estados de um ovo de *Ascaris lumbricoides*.

Fonte: [Pinto \(2011\)](#)

baseados em *transformer* ([RUIZ-SANTAQUITERIA et al., 2022](#); [PEDRAZA et al., 2022](#)).

1.2 Objetivos

O objetivo geral deste trabalho consiste em construir uma arquitetura específica baseada em características de alta resolução para detecção automática de ovos de parasito em imagens de microscópio. Dessa forma, acelerando o processo de diagnóstico e auxiliando profissionais menos experientes.

Especificamente, este trabalho busca os seguintes objetivos específicos aplicados ao problema de detecção de ovos de parasito em amostras fecais usando imagens de microscópio:

- Analisar as dificuldades específicas das imagens de ovos de parasito em microscópio a fim de desenvolver modelos de aprendizado profundo;
- Utilizar técnicas de redes convolucionais para extração de características de alta resolução;
- Propor estratégias para construção e treinamento de arquiteturas robustas utilizando transferência de aprendizado;
- Analisar o impacto de preprocessamentos na detecção de ovos de parasito;
- Validar os métodos propostos por experimentos em imagens microscópicas;
- Combinar os resultados de múltiplos detectores a fim de reduzir falsos negativos;
- Comparar os resultados dos métodos propostos com a literatura conhecida utilizando as principais métricas de detecção de objetos;

Esta dissertação se concentra na investigação de características em alta resolução, visando superar as limitações dos detectores tradicionais na identificação de objetos menores, além de possibilitar o treinamento de uma rede robusta e eficiente baseada nessas características.

1.2.1 Questões de Pesquisa

Para reforçar o cumprimento dos objetivos estabelecidos para esta dissertação, busca-se responder às seguintes questões de pesquisa:

- **Q1:** É mais eficaz adicionar estágios extras na *High-resolution networks* (HRNet) do que simplesmente aumentar o número de canais nos estágios originais da rede?
- **Q2:** A utilização de perdas intermediárias na composição da perda total contribui para um treinamento mais estável em redes profundas?
- **Q3:** A detecção de ovos de parasito em imagens de microscópio pode ser aprimorada pela combinação de múltiplos modelos em um *ensemble*?
- **Q4:** A utilização de características em alta resolução proporciona melhor sensibilidade de ovos menores?

1.2.2 Contribuições

Destacam-se como principais contribuições desta dissertação:

- Construção de uma arquitetura convolucional de alta resolução para detecção de ovos de parasito pequenos, especificamente:
 - Adaptação da arquitetura HRNet adicionando múltiplos estágios seguidos após o fim da rede;
 - Composição da função de perda somada ao longo dos estágios da rede para garantir um treinamento uniforme;

Esta pesquisa resultou na publicação do trabalho [Bessa, Junior e de Almeida \(2025\)](#) como primeiro autor na conferência HCist – *International Conference on Health and Social Care Information Systems and Technologies*.

1.3 Organização do Trabalho

Este trabalho está estruturado da seguinte forma:

- O Capítulo 2 trata da fundamentação teórica das técnicas utilizadas. São abordados conceitos fundamentais de processamento de imagens médicas, processamento de imagens de microscópio, teorias de Aprendizado Profundo e Redes Convolucionais e Detecção de Objetos.
- O Capítulo 3 descreve trabalhos relacionados ao tema, apresentando seus métodos, métricas, base de imagens e resultados para detecção de ovos de parasito em imagens de microscópio.
- O Capítulo 4 apresenta as etapas adotadas que compõem a metodologia proposta para este trabalho.
- O Capítulo 5 trata sobre os resultados obtidos e discussões em relação aos experimentos realizados.
- O Capítulo 6 apresenta as considerações finais sobre os resultados e trabalhos futuros e os artigos científicos resultantes da pesquisa.

2 Fundamentação Teórica

Este capítulo aborda conceitos e fundamentos essenciais em visão computacional, matemática e medicina. Ele se inicia com um estudo na área médica de parasitologia, com foco específico em ovos de parasitos. Em seguida, apresenta os fundamentos matemáticos do aprendizado de máquina e, por fim, conclui com uma análise aprofundada da aplicação da aprendizagem profunda em visão computacional para detecção de objetos.

2.1 Parasitoses

O parasitismo pode ser definido da seguinte forma:

O parasitismo é a relação desarmônica entre espécies diferentes, sendo que um (parasito) se beneficia retirando os meios para sua sobrevivência, podendo prejudicar o outro (hospedeiro) (PINTO, 2011)

Os ovos de parasitos são conhecidos como a forma infectante do parasito, sendo responsáveis por sua penetração no organismo. Após eclodirem, eles dão origem aos parasitos propriamente ditos, que, por sua vez, depositam novos ovos no hospedeiro. Dessa forma, a contaminação por ovos de parasitos pode ser caracterizada como uma doença infecciosa, uma vez que envolve a penetração, o desenvolvimento e a multiplicação de um agente no organismo (PINTO, 2011).

Embora a parasitologia seja um campo vasto de estudo, esta seção concentra-se no estudo das parasitoses causadas por helmintos. Essas formas demandam diagnóstico por meio de imagens microscópicas, abrindo possibilidades para a aplicação de técnicas de visão computacional.

2.1.1 Helmintos

Os helmintos são organismos pertencentes a dois filos distintos: os Platelmintos (vermes achatados) e os Nematelmintos (vermes cilíndricos). Portanto, o termo não se limita a um único filo, mas abrange ambos (DOLABELLA; BARBOSA, 2011).

Na parasitologia humana, as classes de Platelmintos mais relevantes são *Trematoda* e *Cestoda*. No Brasil, os principais representantes da classe *Trematoda* são os parasitos *Fasciola hepatica* e *Schistosoma mansoni*, presentes amplamente por todo o território nacional (DOLABELLA; BARBOSA, 2011). Na classe *Cestoda*, destaca-se *Taenia solium*, cuja infecção está associada ao consumo de carne bovina contaminada. Já no filo dos

Nematelmintos, existem outros parasitos importantes para a parasitologia humana, como *Ascaris lumbricoides* e *Ancylostoma duodenale*.

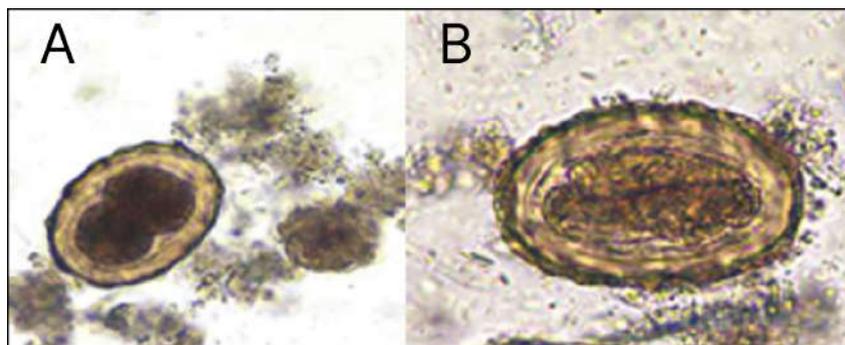
Os helmintos são amplamente conhecidos por utilizarem o corpo humano como hospedeiro para completar seu ciclo de vida, causando diversas doenças. Esses parasitas se instalam, predominantemente, no intestino humano, onde obtêm nutrientes, reproduzem-se e produzem ovos. Esses ovos são eliminados com as fezes do hospedeiro, atuando como agentes de dispersão, facilitando a contaminação de novos hospedeiros e a perpetuação do ciclo infeccioso.

A ascaridíase é uma doença provocada pela infecção com o nematoide *Ascaris lumbricoides*, popularmente conhecido como lombriga. Por ser um geo-helminto, a transmissão ocorre por meio do contato com solo, água ou alimentos contaminados, onde os ovos vão para o intestino delgado e eclodem. Essa infecção é considerada perigosa devido à elevada capacidade de locomoção do parasito e ao seu porte significativo, podendo atingir até 40 cm de comprimento e 6 mm de diâmetro. Essas características aumentam o risco de obstrução intestinal no hospedeiro, uma condição grave que, em muitos casos, requer intervenção cirúrgica.

Em média, estima-se que cada hospedeiro abrigue entre 4 e 16 áscaris, embora casos extremos possam apresentar uma carga parasitária de 500 a 700 vermes (FOCACCIA, 2005). Há uma relação direta entre a quantidade de parasitos e a gravidade da doença. As fêmeas da espécie possuem uma impressionante capacidade reprodutiva, com cerca de 27 milhões de ovos ao longo da vida, liberando diariamente entre 200 mil e 240 mil ovos.

Os ovos de *Ascaris lumbricoides* apresentam três camadas: interna, média e externa. A camada externa é de coloração marrom, resultado da absorção de pigmentos provenientes das fezes. Os ovos férteis possuem forma oval, com dimensões de 60×45 µm, enquanto os ovos inférteis são mais alongados e possuem uma casca mais fina, medindo entre 80 e 90 µm de comprimento. A presença de ovos inférteis está relacionada ao sexo do verme, geralmente indicando infecção por vermes fêmeas. Além do fato de fertilidade, é necessário atentar-se ao estágio de desenvolvimento do ovo, podendo ser inicial ou maduro, como mostra a Figura 3, destacando-se a percepção visual da larva no ovo no estágio maduro. A formação dos ovos ocorre a partir do quarto ou último estágio do verme, aproximadamente 65 dias após a infecção.

O diagnóstico da ascaridíase é realizado por meio da análise microscópica de amostras fecais, devido à deposição diária de ovos, com cerca de 200 mil ovos por fêmea a cada dia. Esses ovos possuem uma característica de serem pesados, facilitando sua extração pelo método de segmentação. Além disso, o método de Kato-Katz também apresenta boa sensibilidade, permitindo a contagem dos ovos e, assim, a determinação do grau de infecção (FOCACCIA, 2005).

Figura 3 – Estágios de desenvolvimento do ovo de *Ascaris lumbricoides*

(A) Estágio Inicial. (B) Estágio Maduro. Adaptado de (FOCACCIA, 2005)

A teníase é uma parasitose intestinal causada pela infecção com *Taenia solium*, *Taenia saginata* ou *Taenia asiatica*, sendo popularmente conhecida como "solitária". A infecção ocorre por meio da ingestão de carne contaminada e mal cozida de animais infectados. A *Taenia solium* pode parasitar seres humanos, suínos, cães, gatos e macacos, enquanto a *Taenia saginata* infecta bovinos, girafas e lhamas. Os principais sintomas da teníase incluem desconforto abdominal, náuseas, vômitos, aumento ou diminuição do apetite (bulimia) e desnutrição (PINTO, 2011).

A *Taenia solium*, principal parasita de seres humanos, pode atingir até 9 metros de comprimento. Seu corpo é composto por segmentos anulares chamados proglotes, entre os quais se destacam as proglotes maduras ou grávidas. Essas proglotes desprendem-se do corpo principal, sendo eliminadas com as fezes. No ambiente, elas liberam milhares de ovos, garantindo a continuidade do ciclo de vida do parasita (FOCACCIA, 2005).

Os ovos da tênia apresentam formato oval e são envoltos por uma membrana externa espessa e estriada, conforme ilustrado na Figura 4. Seu tamanho varia entre 30 e 40 μm de diâmetro.

Figura 4 – Ovo de Tênia.



Adaptado de (PINTO, 2011)

Em alguns casos, o ser humano pode ser infectado pelas formas larvais das tênias, desenvolvendo uma condição conhecida como cisticercose. A forma mais grave ocorre quando o cisticerco se aloja no sistema nervoso central, resultando na neurocisticercose, que pode levar a complicações severas e, em casos extremos, à morte.

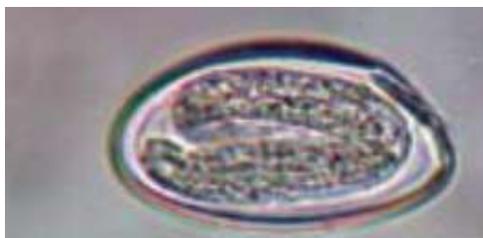
Semelhante à ascaridíase, o diagnóstico da teníase é realizado por meio da análise microscópica de amostras fecais, já que o parasita libera ovos ou proglotes nas fezes. A sensibilidade do método é difícil de determinar, ao depender de diversos fatores, como o número de ovos, o tamanho e a qualidade da amostra fecal. É importante destacar que o exame de fezes não permite a identificação da espécie de tênia, uma vez que os ovos são morfológicamente indistinguíveis entre as diferentes espécies (FOCACCIA, 2005).

Outra doença comum causada por helmintos é a enterobiíase, provocada pelo *Enterobius vermicularis* ou *Oxyurus vermicularis*, sendo o primeiro de maior relevância médica. Esse parasita destaca-se por ser exclusivo do ser humano. Diferentemente de outros helmintos, não é robusto; pelo contrário, apresenta um corpo delicado, pequeno e de coloração branca, medindo até 12 mm de comprimento. As fêmeas possuem dois úteros, capazes de produzir, em média, 11 mil ovos ao longo de 20 a 60 dias.

Diferentemente dos vermes anteriormente mencionados, os ovos de *Enterobius vermicularis* são raramente encontrados nas fezes. Isso ocorre porque, quando grávida, a fêmea migra do ceco (região inicial do intestino grosso) para a área perianal, onde deposita os ovos. Por essa razão, a infecção é transmitida principalmente por meio das mãos que entram em contato com a região anal, favorecendo a disseminação dos ovos. (FOCACCIA, 2005)

Os ovos apresentam características de destaque como transparência ou coloração esbranquiçada e formato oval e alongado, como pode ser visto na Figura 5.

Figura 5 – Ovo de *Enterobius vermiculares*



Adaptado de (PINTO, 2011)

Os sintomas gastrointestinais mais comuns incluem náuseas, vômitos e dores abdominais, decorrentes da fixação dos vermes na mucosa intestinal. Entretanto, a infecção por *Enterobius vermicularis* apresenta manifestações específicas, como prurido anal e vulvar. Esses sintomas tornam-se mais intensos à noite, devido ao aumento da temperatura corporal, que estimula a movimentação dos parasitas e agrava a coceira.

Como mencionado anteriormente, os ovos são depositados na região anal, resultando em uma sensibilidade de apenas 5% para o diagnóstico por análise microscópica de amostras fecais (FOCACCIA, 2005). Dessa forma, a técnica de diagnóstico mais utilizada é o *swab anal*, que consiste na aplicação de uma fita adesiva na região anal para coletar os ovos presentes no local.

A capilaríase é uma doença causada pelo nematoide *Capillaria philippinensis*. Este é um muito pequeno, medindo entre 2,3 e 3,2 mm. Ele se localiza na mucosa do intestino grosso. A principal forma de contaminação é através do consumo de carne de peixe crua infectada pela larva do verme.

Este parasito apresenta uma especificidade no seu ciclo de vida. Os ovos das fêmeas possuem dois estágios: embrionários e não embrionários. A fêmea despeja os ovos não embrionários no intestino, excretados junto às fezes. No ambiente, esse ovo se torna em um ovo embrionário que, quando é consumido por um peixe, eclode e se torna uma larva (CDC, 2020).

O diagnóstico é feito através da análise de amostras fecais no microscópio em busca de ovos não embrionários. Estes ovos medem entre 35 e 45 µm de comprimento e 20 a 25 µm de largura. Possuem duas proeminências polares planas e uma concha estriada, como pode ser visto na Figura 6.

Figura 6 – Ovo de *Capillaria philippinensis*

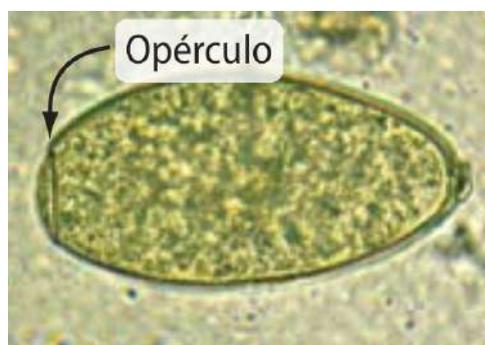


Adaptado de (CDC, 2020)

A fasciolíase é uma doença predominante em animais herbívoros de pastagem e causada por parasitos *Fasciolopsis*. Porém, passou a ter importância médica devido ao crescente aumento nos casos de infecções em humanos. O verme adulto possui formato alongado e corpo recoberto por espinhos. Seu tamanho é de aproximadamente 3 cm de comprimento e 1,5 cm de largura.

A infecção ocorre através do caramujo *Lymnaea* que é o hospedeiro intermediário. Os caramujos de dispersão metacercárias na água, a infecção, se dá pelo consumo da água contaminada. Ao chegar no intestino delgado, ocorre a desincrustação e o estágio inicial do verme migra para os canais biliares, onde se desenvolve até atingir a maturidade sexual. Em outros casos, as larvas podem ir para o sistema sanguíneo e se distribuírem por diferentes localidades do corpo. Os adultos que alcançam os canais biliares põem ovos que descem com a bile para o intestino, sendo dispostos com as fezes (PINTO, 2011). Seus ovos possuem formato oval e medem entre 140 µm de comprimento e 60 µm de largura. Estes ovos possuem um opérculo, sendo uma abertura de onde sai a larva após formada, como pode ser visto na Figura 7.

Figura 7 – Ovo de *Fasciola hepatica*



Adaptado de (PINTO, 2011)

O diagnóstico é feito por análise de amostras de fezes no microscópio. Recomenda-se a utilização do método HPJ ou sedimentação para separação dos ovos.

Embora não mencionadas, existem outras espécies de helmintos que parasitam humanos, como apresentado na Figura 8.

2.1.2 Desafios do diagnóstico por microscópio

Como observado na seção 1, o método de diagnóstico amplamente utilizado é a análise de amostras fecais por microscopia. Embora existam outros métodos disponíveis, o exame microscópico se destaca por sua acessibilidade e baixo custo, uma vez que requer apenas o preparo da amostra e o uso de um microscópio. Além disso, já existem kits específicos, como os utilizados no método Kato-Katz, que facilitam o preparo das amostras e são amplamente empregados ((SEVIR); (COVAT); (CEVET), 2022). Adicionalmente, microscópios USB acessíveis também estão sendo utilizados em análise de ovos de parasito (SUWANNAPHONG et al., 2023).

Apesar dos avanços, a localização, identificação e contagem de ovos de parasitos em amostras continuam integralmente dependentes da experiência do técnico de laboratório, acarretando dois problemas principais: o diagnóstico demorado e a incerteza nos resultados.

Figura 8 – Espécies de helmintos

Filo	Classe	Família	Gênero	Espécies
Platyhelminthes	Trematoda	Schistosomatidae	Schistosomatidae	S. mansoni
		Fasciolidae	Fasciola	S. japonicum S. haematobium F. hepatica
	Cestoda	Taeniidae	Taenia Echinococcus	T. solium T. saginata E. granulosus
		Hymenolepididae	Hymenolepis	H. nana H. diminuta
		Ascarididae	Ascaris Toxocara	A. lumbricoides T. canis
Nematoda	Secernentea	Oxyuridae	Enterobius	E. vermicularis
		Strongyloididae	Strongyloides	S. stercoralis
		Ancylostomidae	Ancylostoma Necator	A. duodenale A. braziliense N. americanus
		Trichuridae	Trichuris	T. trichiura
		Onchocercidae	Wuchereria Onchocerca	W. bancrofti O. volvulus

Adaptado de (PINTO, 2011)

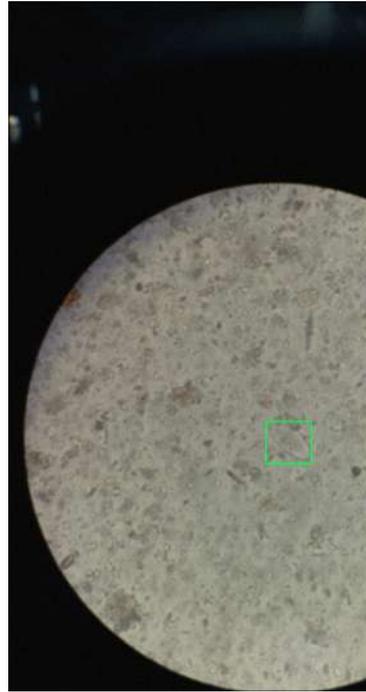
O primeiro diz respeito ao tempo necessário para a análise cuidadosa de cada amostra. Esse processo pode levar até 30 minutos por amostra, conforme destacado por [Anantrasirichai et al. \(2022\)](#).

O segundo problema está relacionado à possível falha na identificação de ovos, frequentemente causada pela inexperiência ou pela sobrecarga de trabalho do especialista. Essa limitação pode reduzir a precisão na detecção de ovos menores e mais raros, como os de *Enterobius vermicularis*, como pode ser observado na Figura 9, sendo mais difíceis de identificar. Em contraste, ovos maiores e mais pesados, como os de *Ascaris lumbricoides*, são detectados com maior facilidade e sensibilidade. Isto compromete a contagem do número de ovos por grama de fezes (OPG) utilizado para estimar o grau da infecção (GOMES, 2022).

2.2 Detecção de Objetos com Aprendizado Profundo

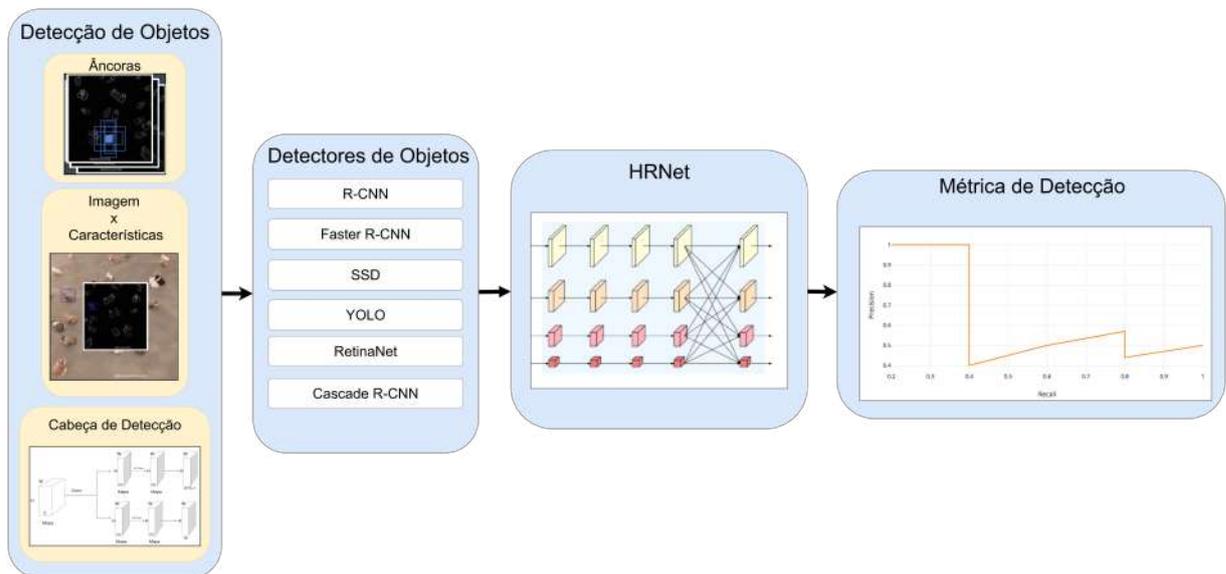
Com o avanço das técnicas de aprendizado profundo, especialmente as redes neurais convolucionais, tarefas de detecção em imagens médicas têm sido automatizadas. Arquiteturas específicas foram desenvolvidas para solucionar a tarefa de detecção de objetos.

Esta seção visa apresentar os fundamentos da detecção de objetos com técnicas de aprendizado profundo. Para isso, está organizada em tópicos que abordam inicialmente

Figura 9 – Amostra em microscópio de *Enterobius vermicularis*

Fonte: [Anantrasirichai et al. \(2022\)](#)

Figura 10 – Tópicos de Detecção de Objetos



a relação entre imagem e características, âncoras, e cabeças de detecção. Em seguida, explora a evolução das arquiteturas ao longo do tempo, os avanços recentes com a rede HRNet e, por fim, discute as métricas utilizadas para avaliar a detecção de objetos. Esse fluxo pode ser visualizado na Figura 10.

A detecção de objetos por meio de arquiteturas de aprendizado profundo baseia-se em princípios fundamentais, como a relação espacial entre os mapas de características e a

imagem original, o uso de centros de âncoras para localizar os objetos e a aplicação de cabeças de detecção para realizar a identificação de múltiplos objetos.

A detecção de objetos por meio de aprendizado profundo fundamenta-se na relação entre os mapas de características produzidos por uma *backbone* e a imagem original.

Essa relação é utilizada para localizar pixels demarcados, chamados âncoras, na imagem original. Deste modo, posicionar uma âncora na imagem original é correspondente a colocar uma âncora no mapa de característica. Esta relação será detalhada na Seção 2.2.1.

As âncoras são o segundo fundamento da detecção de objetos em aprendizado profundo. Elas são caixas delimitadoras com diferentes formatos, que são utilizadas para localizar objetos. O centro de uma âncora é posicionada em um pixel do mapa de característica, de modo que a rede deve aprender a prever se uma âncora possui ou não um objeto nela e também como melhor ajustar a âncora ao objeto se houver. As especificidades das âncoras serão abordadas na Seção 2.2.2.

Essa predição e ajustes são feitos através da cabeça de detecção que recebe um mapa de característica, posiciona as âncoras e gera a predição para todas as âncoras. A fundamentação computacional e matemática das cabeças de detecção será explorada na Seção 2.2.3.

Pode-se compreender que a detecção de objetos com aprendizado profundo possui três fundamentos: relação entre característica e a imagem original, âncoras e cabeça de detecção. As seções seguintes irão abordar os detalhes de cada um desses fundamentos.

Os modelos de detecção desenvolvidos até o momento seguem esses fundamentos. As duas principais abordagens utilizadas são: a detecção em um estágio, na qual as predições são feitas diretamente a partir dos mapas de características, e a detecção em dois estágios, onde os objetos são inicialmente localizados e, em seguida, classificados. As diferentes abordagens para construção de modelos de detecção serão exploradas em ordem cronológica na seção 2.2.4.

Recentemente, características de alta resolução têm sido incorporadas aos detectores de dois estágios, resultando em avanços significativos. Isso se deve à melhoria na detecção de objetos menores, pois características de alta resolução proporcionam um campo receptivo mais amplo e detalhado. A fundamentação computacional será descrita na Seção 2.2.5.

A detecção de objetos adotou uma métrica específica para avaliação: o mAP. Essa métrica se popularizou por considerar diferentes limiares de IoU e níveis de confiança das predições. No entanto, métricas de classificação também podem ser adaptadas para a detecção de objetos ao utilizar limiares de confiança. A definição de cada uma dessas métricas é detalhada na Seção 2.3.

2.2.1 Extração de Características usando Redes Convolucionais para Detecção de Objetos

Os detectores de objetos baseados em aprendizado profundo fundamentam-se na relação entre o mapa de características e a imagem original. Isso ocorre porque cada região no mapa contém informações correspondentes a uma área específica da imagem original.

A convolução é uma operação que mapeia uma área da imagem para uma região correspondente em outra imagem, definida pela Equação 2.1. Nesse processo, o valor do mapa de características M na posição (i,j) é obtido pela multiplicação do *kernel* K na posição (m,n) com a imagem I na posição $(i+m,j+n)$, resultando no mapeamento de um valor de I para M . Se a matriz K , que representa o *kernel*, for maior que 1×1 , o mapa M terá uma resolução inferior à da imagem I . Para manter a resolução original, pode-se aplicar *padding*. Além disso, quando o *stride* for maior que 1, o *kernel* avança mais de um pixel por vez, reduzindo ainda mais a resolução da saída. Nesse caso, a equação é modificada, como mostrado na Equação 2.2.

$$M(i, j) = (I * K)(i, j) = \sum_m \sum_n I(i + m, j + n)K(m, n) \quad (2.1)$$

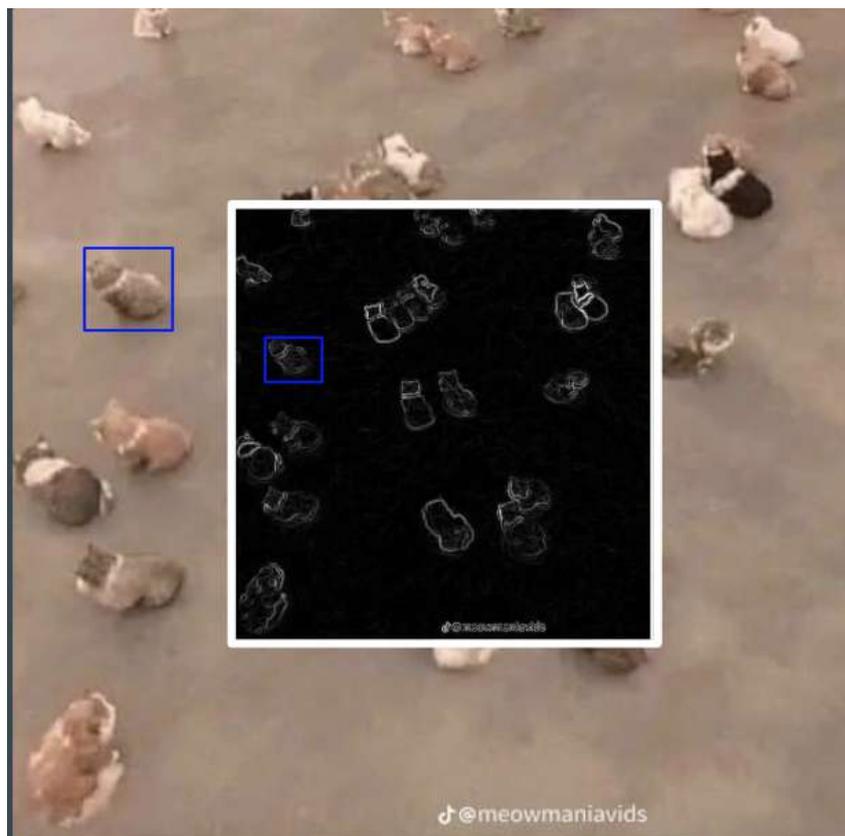
$$M(i, j) = (I * K)(i, j) = \sum_m \sum_n I(i \cdot s + m, j \cdot s + n)K(m, n) \quad (2.2)$$

Nas redes convolucionais, devido aos requisitos de memória, as convoluções reduzem a resolução da imagem original para gerar um mapa de características. Dessa forma, o mapa M terá uma resolução menor que a da imagem I . Isso significa que um pixel (i, j) em M corresponderá a uma área maior da imagem I , resultando em regiões menores que representam características de áreas maiores. Esse conceito pode ser observado no exemplo da Figura 11.

Portanto, as características de um determinado objeto na imagem original ocuparão uma região menor no mapa de características após a convolução. Para objetos menores, isto implica que terão cada vez menos área nos mapas de características à medida que a rede se aprofunda e extrai mapas de menor resolução.

Considerando uma imagem I de dimensões 512×512 , um *kernel* K de dimensões 3×3 e *stride* de 1, e um objeto em I com dimensões 12×12 , após a primeira convolução, o mapa de características M terá dimensões 510×510 e o objeto será reduzido para 10×10 . Após a segunda convolução, o mapa de características M' terá dimensões 508×508 e o objeto será reduzido para 8×8 .

Figura 11 – Exemplo de relação entre característica e a imagem original



Elaborado pelo Autor

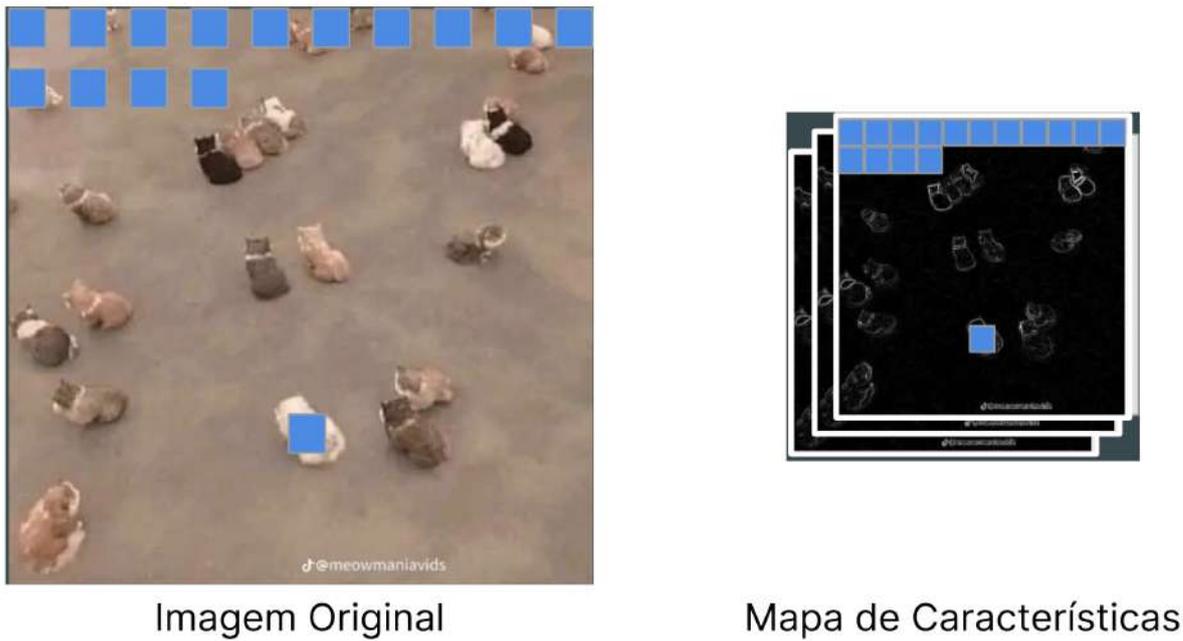
2.2.2 Âncoras e Mapas de Características

A partir dessa relação, os detectores de objetos baseados em aprendizado profundo utilizam centros de âncoras no mapa de características para corresponder a pixels na imagem original. Um centro de âncora é uma marcação que indica a posição onde âncoras serão atribuídas. Inicialmente, o mapa de características é dividido em uma malha, usando suas próprias dimensões como divisores. Assim, um mapa de características de 32×32 pixels corresponde a uma malha de 32×32 centros de âncoras. Em cada pixel, é colocado um centro de âncora, como ilustrado na Figura 12.

Em cada centro de âncora são colocadas k âncoras e as âncoras possuem diferentes formatos, de modo a cobrir uma variedade de objetos de diferentes formatos, como pode ser visto na Figura 13. Portanto, quanto maior a resolução do mapa de características, mais âncoras serão colocadas.

Essas âncoras podem ser ajustadas por meio de deltas (Δ) que representam variações nas coordenadas x e y , além da largura (W) e altura (H) da âncora. Esses deslocamentos são aplicados nas Equações 2.3, 2.4, 2.5 e 2.6, permitindo o cálculo das novas coordenadas x' , y' , $altura'$ e $largura'$.

Figura 12 – Relação entre centros de âncora no Mapa de Características e a Imagem Original



Elaborado pelo Autor.

Isso permite que as âncoras se ajustem de forma mais precisa aos objetos. Nos detectores baseados em aprendizado profundo, esses deltas são calculados pela cabeça de detecção do modelo.

$$x' = x + \Delta x \quad (2.3)$$

$$y' = y + \Delta y \quad (2.4)$$

$$largura' = largura + \Delta largura \quad (2.5)$$

$$altura' = altura + \Delta altura \quad (2.6)$$

2.2.3 Cabeça de Detecção

Após a extração de características, os dados são enviados para a cabeça de detecção. Essa etapa é responsável pela regressão dos deltas e deslocamentos, além da classificação de cada âncora. Cada âncora pode ser classificada como não objeto ou em uma classe alvo, e cada âncora possui a regressão de 4 deltas de deslocamento. Para cada pixel temos um vetor de tamanho $(n * k) + (4 * k) + 1$, onde n é o número de classes e k é o número

Figura 13 – Relação entre âncora no Mapa de Características e a Imagem Original

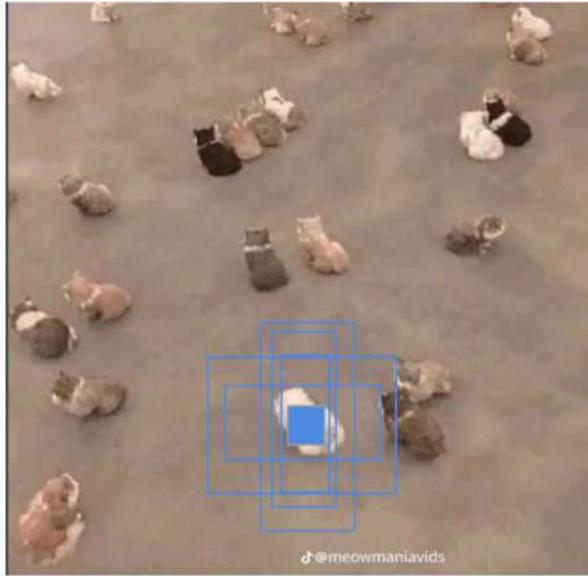
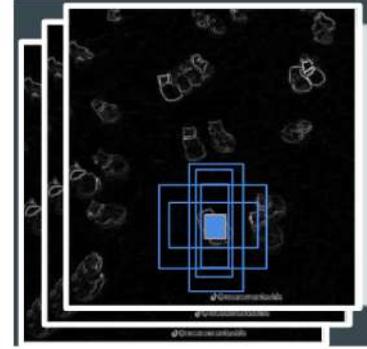


Imagem Original



Mapa de Características

Elaborado pelo Autor.

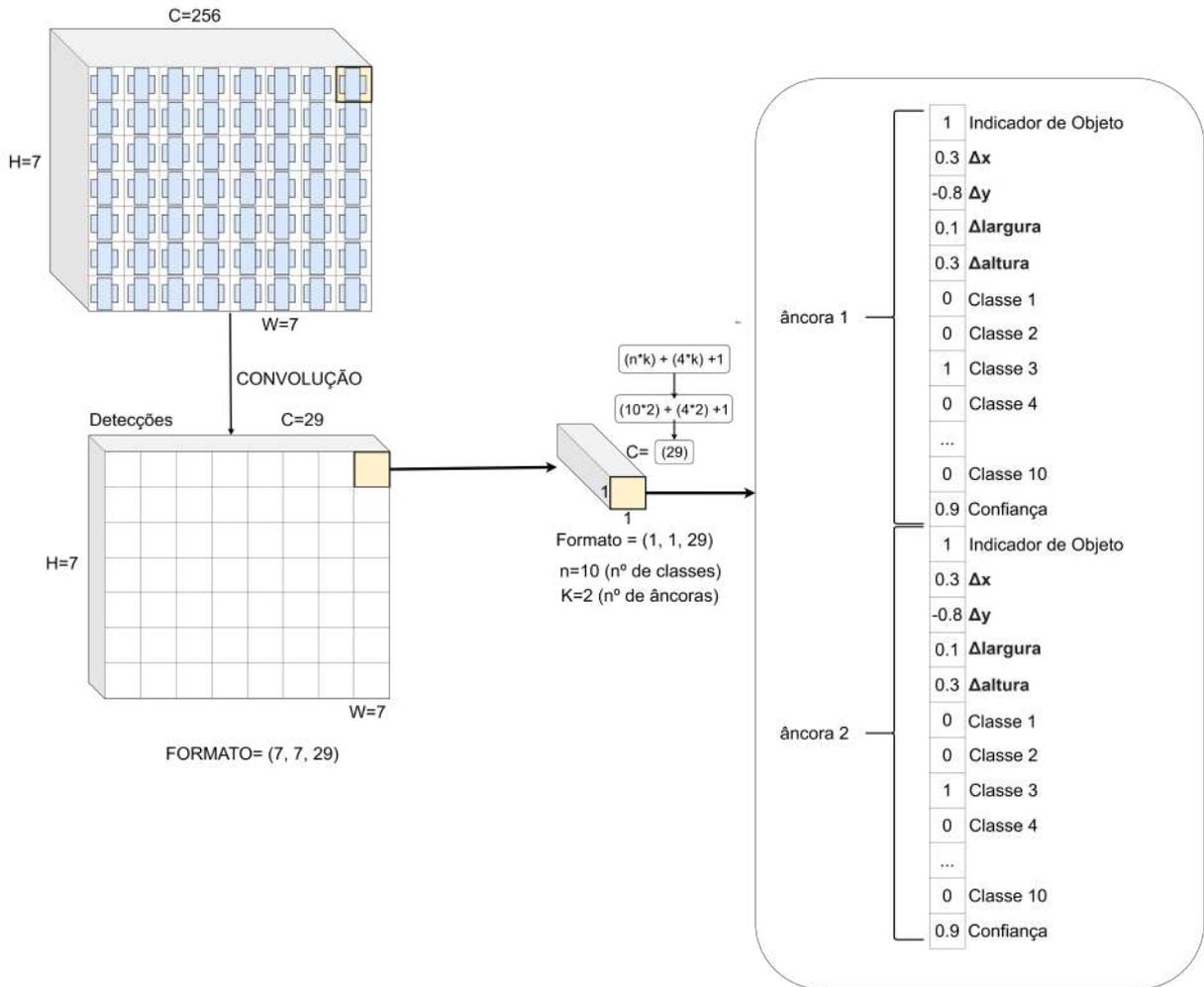
de âncoras. Ou seja, temos para cada âncora a classificação em uma das n classes, e a regressão dos 4 deltas, e adicionalmente temos 1 valor que corresponde à pontuação de confiança do modelo. Como a saída da cabeça de detecção possui um formato fixo, ela é modelada como convoluções. Essa modelagem pode ser visualizada na Figura 14 e é descrita de forma geral na Equação 2.7. No entanto, ao utilizar um filtro 1×1 , essa operação pode ser simplificada para a Equação 2.8. Nesse caso, cada valor de saída em um pixel é apenas uma combinação linear dos valores de entrada no mesmo pixel, utilizando os pesos aprendidos para cada canal.

$$Y(i, j, z) = \sum_{m=1}^C \sum_{u=-P}^P \sum_{v=-P}^P W_z(u, v, m) * X(i + u, j + v, m) + b_z, \quad (2.7)$$

Onde,

- $Y(i, j, z)$ é o valor de saída para o pixel (i, j) no canal z ;
- $W_k(u, v, m)$ é o peso do filtro associado a canal c , para o mapa de característica m , com deslocamento espacial u e v ;
- $X(i, j, m)$ é o valor do tensor de entrada para o pixel (i, j) no canal m ;
- b_k é o viés (*bias*) associado ao canal z ;
- P é o tamanho do *padding* utilizado.

Figura 14 – Predição de Detecções a partir de um mapa de características



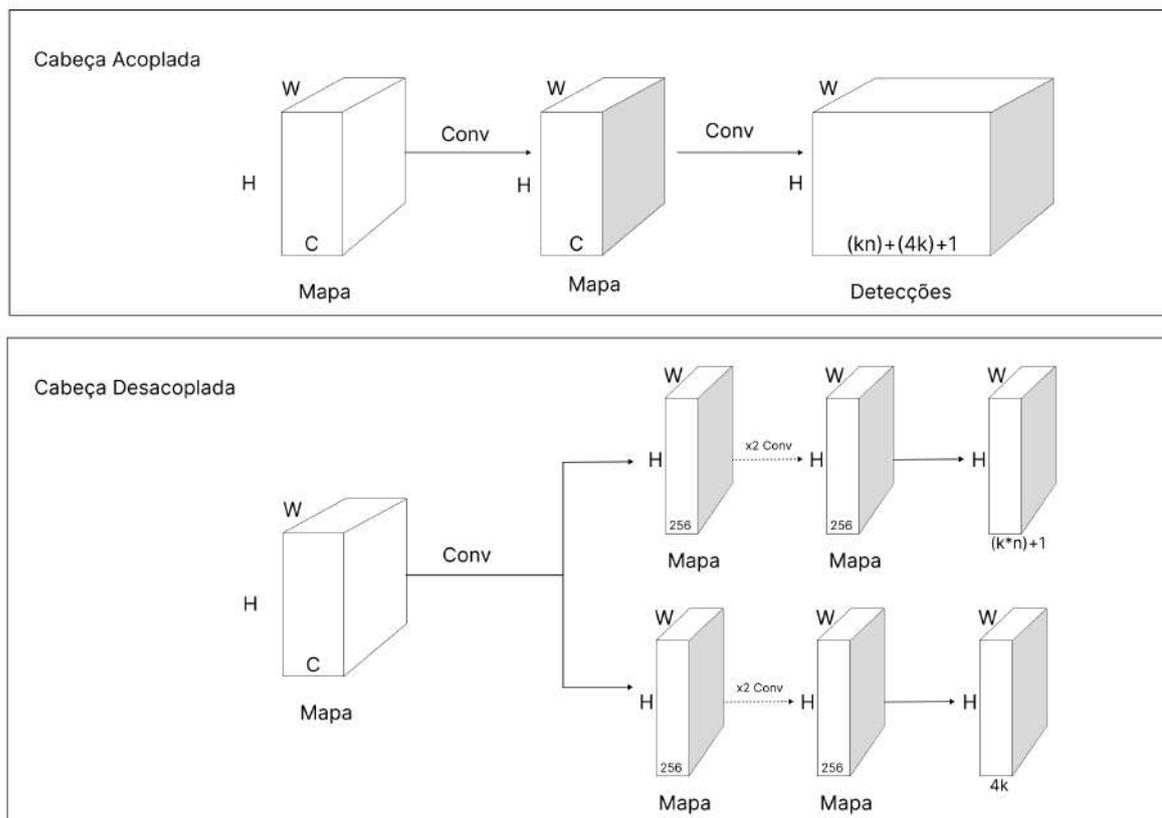
$$Y(i, j, z) = \sum_{m=1}^{256} W_z(0, 0, m) * X(i, j, m) + b_z \quad (2.8)$$

Onde,

- $Y(i,j,z)$ é o valor de saída para o pixel (i,j) no canal z ;
- $W_z(u, v, m)$ é o peso do filtro associado ao canal c , para o mapa de características m , com deslocamento espacial u e v ;
- $X(i, j, m)$ é o valor do tensor de entrada para o pixel (i,j) no canal m ;
- b_k é o viés (*bias*) associado ao canal z ;

A classificação e a regressão podem ser feitas de forma unificada utilizando um único ramo que executará ambas as tarefas, chamadas cabeças de detecção acopladas, ou utilizando dois ramos distintos para cada tarefa. A Figura 15 exemplifica a diferença entre os dois formatos de cabeça de detecção.

Figura 15 – Diferença entre cabeças de detecção acopladas e desacopladas



Elaborado pelo Autor

2.2.4 Detectores de Objetos

A integração do aprendizado profundo com a detecção de objetos resulta em modelos capazes de localizar e classificar um ou mais objetos em uma imagem. Esses modelos utilizam técnicas avançadas de aprendizado profundo e possuem a vantagem de serem soluções ponta-a-ponta, possibilitando um treinamento mais eficiente.

Atualmente, os detectores de objetos são classificados em duas categorias principais: de um estágio e de dois estágios. Em 2014, foi introduzida a *Regions with CNN* (RCNN) (GIRSHICK et al., 2014), um modelo de detecção de dois estágios que, primeiramente, propunha regiões potenciais contendo objetos e, em seguida, realizava a classificação dessas regiões. No entanto, devido ao alto custo computacional e ao longo tempo de processamento desses modelos, detectores mais rápidos foram desenvolvidos, como o SSD (LIU et al., 2016). Este modelo inovou ao unificar a proposta de regiões e a classificação em um único estágio, proporcionando maior eficiência.

2.2.4.1 Faster R-CNN

Uma evolução da RCNN, a *Faster R-CNN* (REN et al., 2016) introduz o uso de uma rede totalmente convolucional para gerar propostas de regiões. Com isso, é possível treinar uma Rede de Proposta de Regiões (RPN).

Na *Faster R-CNN*, ocorre a divisão dos mapas de característica em uma malha e são posicionadas âncoras em cada célula dessa malha. Essas âncoras possuem diferentes proporções e tamanhos, permitindo que se adaptem melhor a objetos de diferentes formatos. Para cada âncora, é avaliado se ela contém um objeto ou não, além de calcular uma regressão de quatro deslocamentos (deltas) que ajustam as coordenadas da âncora. Esses ajustes permitem deformar a âncora para que ela se alinhe melhor ao objeto. Essa abordagem é implementada utilizando dois ramos convolucionais na RPN: um para classificação e outro para regressão.

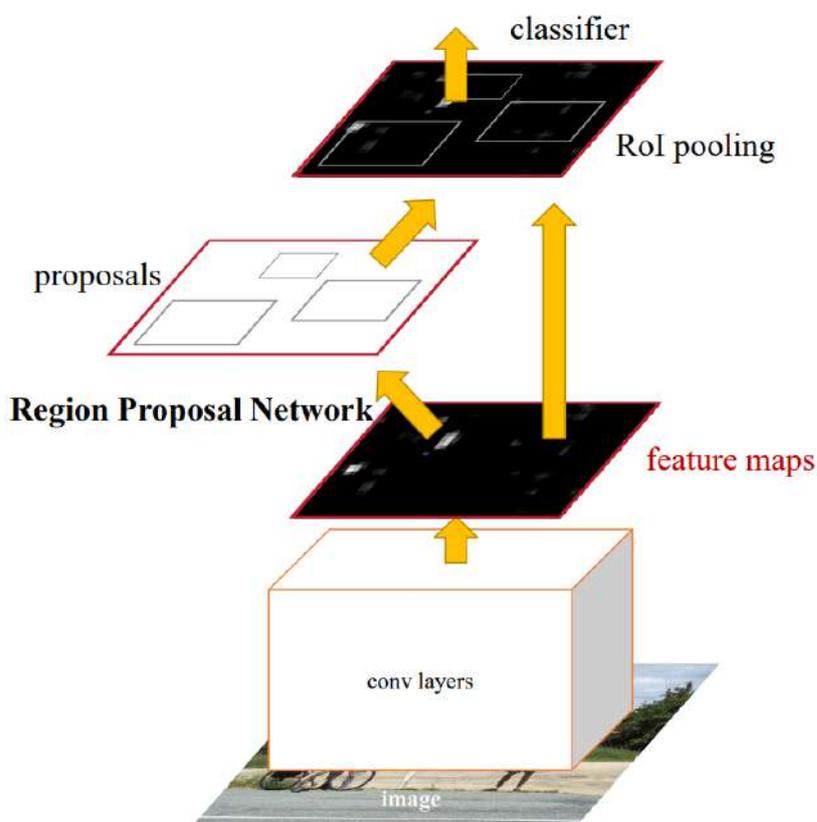
Ao final, temos uma rede totalmente convolucional capaz de propor regiões de interesse que provavelmente contêm objetos. A RPN substitui o método de *selective search* (UIJLINGS et al., 2013) na geração de propostas de regiões, como ilustrado na Figura 16, tornando o modelo mais generalista e flexível para diferentes domínios.

Apesar do avanço em eficiência trazido pelo *Faster R-CNN*, o tempo de inferência ainda não era satisfatório para aplicações em tempo real. Para resolver esse problema, surgiram os modelos de detecção de um único estágio. Esses modelos buscam unificar as etapas de proposta de regiões e classificação. Assim, durante a geração das propostas de regiões, a classificação de cada âncora é realizada simultaneamente, eliminando a necessidade de uma etapa adicional de classificação.

2.2.4.2 SSD

Com base nessa ideia, foi desenvolvido a *Single Shot MultiBox Detector* (SSD) (LIU et al., 2016), que implementa essa abordagem por meio de cabeças de detecção, modeladas como convoluções. No entanto, o ramo de classificação no SSD possui $n*k$ características, onde n é o número de classes e k é o número de âncoras. Isso ocorre porque cada âncora é classificada como pertencente a uma das n classes ou como um “não objeto”. Em contrapartida, a RPN possui $2*k$ características, já que considera k âncoras e apenas duas classes possíveis: “objeto” e “não objeto”.

O SSD utiliza cabeças de detecção acopladas, ou seja, uma única convolução é responsável por gerar o volume correspondente à classificação e à regressão dos deltas de deslocamento das âncoras. Essa abordagem permite uma execução mais rápida, mas compromete a precisão do modelo. Isso ocorre porque um único ramo de convoluções é usado para realizar ambas as tarefas, diferentemente da RPN, que utiliza ramos desacoplados. Na RPN, há um ramo dedicado para regressão e outro para classificação, permitindo a

Figura 16 – Arquitetura da *Faster R-CNN*

Fonte: Ren et al. (2016)

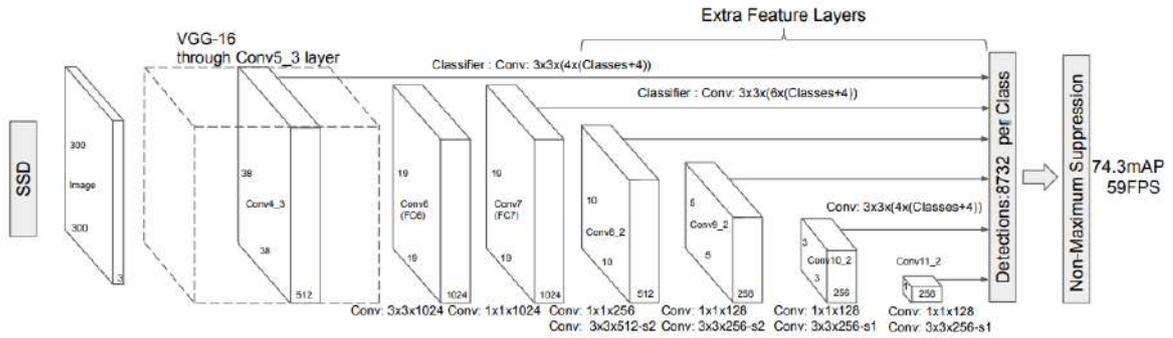
extração de características específicas para cada tarefa de forma independente.

A arquitetura do SSD foi projetada para aproveitar características em múltiplas resoluções, extraídas de diferentes níveis de profundidade do *backbone*, para realizar a detecção, como pode ser visto na Figura 17. Essa abordagem permite combinar desde características de alta resolução com baixa profundidade até características de baixa resolução com alta profundidade. Essa estratégia visa capturar informações espaciais detalhadas para objetos pequenos, garantindo que eles não sejam perdidos no posicionamento das âncoras, enquanto utiliza informações de textura e contexto para detectar objetos maiores eficazmente.

2.2.4.3 YOLO

A YOLO surge com uma abordagem totalmente convolucional. Ao contrário da SSD que utiliza características em múltiplas resoluções de camadas intermediárias do *backbone*, a YOLO aplica uma Perceptron Multicamadas (MLP) no último mapa de características para gerar as predições. Isso resultava em dificuldades para detectar objetos pequenos,

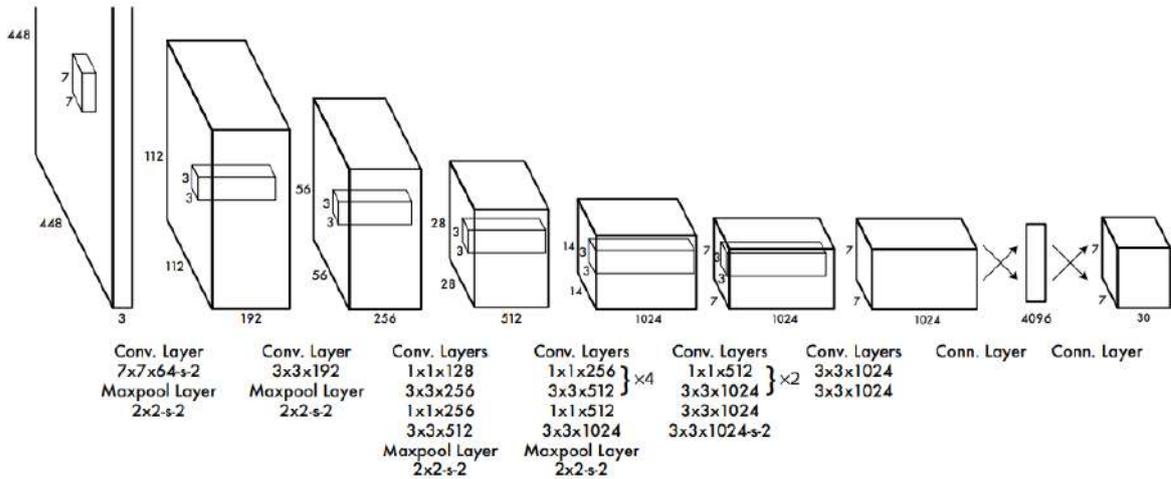
Figura 17 – Arquitetura da SSD



Fonte: Liu et al. (2016)

pois, como ilustrado na Figura 18, a saída do *backbone* possuía uma resolução de apenas 7x7 pixels. Considerando uma entrada de 448 pixels, cada pixel do mapa de características representava uma área de 64 pixels na imagem original, tornando a detecção de objetos pequenos praticamente inviável.

Figura 18 – Arquitetura da YOLO



Fonte: Redmon (2016)

Apesar dessa limitação, a rede YOLO foi um marco importante, apresentando um menor erro na classificação de fundo em comparação com a *Fast R-CNN*, uma arquitetura de dois estágios. Isso demonstra que os detectores de único estágio, ao utilizarem informações de toda a imagem para classificação, conseguem uma distinção mais eficaz entre objeto e fundo. Em contraste, a *Fast R-CNN* se baseia apenas em informações locais das regiões propostas, o que pode limitar sua capacidade de diferenciação.

2.2.4.4 RetinaNet

A *RetinaNet* (ROSS; DOLLÁR, 2017), uma arquitetura que, estruturalmente, não apresenta grandes inovações, se utiliza da *Feature Pyramid Networks* para detecção de objetos (LIN et al., 2017). Ela utiliza um *backbone* para extração de características, uma FPN para fusão dessas características e cabeças de detecção desacopladas. No entanto, a principal contribuição da *RetinaNet* está na introdução de uma nova função de custo para detectores de um estágio: a *Focal Loss*.

Essa função adiciona fatores de supressão à clássica entropia cruzada, como mostrado na Equação 2.9. Em detectores de um estágio, há um desbalanceamento significativo entre âncoras negativas (sem objetos) e âncoras positivas (com objetos). Esse desequilíbrio dificulta o treinamento, pois os erros gerados pelas âncoras negativas superam consideravelmente os erros das âncoras positivas, fazendo com que o modelo tenha dificuldade em aprender a detectar objetos.

Para resolver esse problema, a *RetinaNet* propõe a redução do impacto do excesso de erros negativos, ajustando-os para que sejam menos relevantes do que os erros positivos. Isso permite um treinamento mais equilibrado e eficaz para detectores de um estágio.

$$FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t) \quad (2.9)$$

Onde,

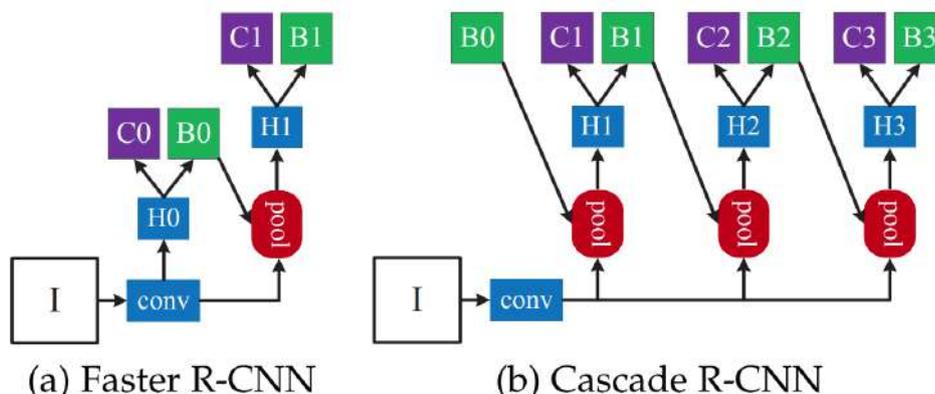
- α_t é o fator de ponderação para balanceamento de classes;
- p_t é a probabilidade estimada pelo modelo para a classe verdadeira;
- γ é fator de supressão;

2.2.4.5 Cascade R-CNN

Em 2019, foi proposta a *Cascade R-CNN* (CAI; VASCONCELOS, 2019), uma extensão da *Faster R-CNN* que introduz uma estrutura em cascata. Essa abordagem permite que cada estágio da cascata refine as predições recebidas do estágio anterior, filtrando progressivamente as melhores detecções. Assim, se um estágio descarta uma detecção, essa não será passada para os estágios subsequentes.

A arquitetura da *Cascade R-CNN* pode ser entendida como múltiplas instâncias da *Faster R-CNN* organizadas em sequência, como ilustrado na Figura 19. Essa organização em cascata aprimora a qualidade das predições, proporcionando uma detecção mais precisa a cada estágio.

Figura 19 – Arquitetura da *Cascade R-CNN*

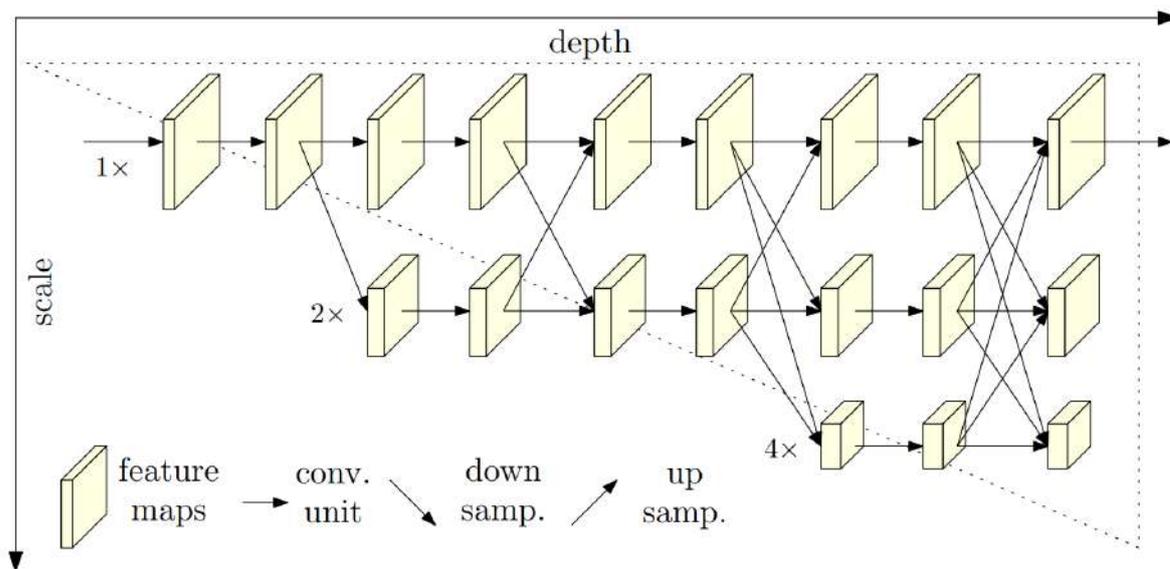


Fonte: Cai e Vasconcelos (2019)

2.2.5 HRNet

A rede HRNet foi inicialmente concebida para a tarefa de estimar poses humanas (SUN et al., 2019). Essa rede destaca-se por sua capacidade de extrair e preservar mapas de características em alta resolução, desenvolvendo múltiplos ramos paralelos para diferentes resoluções, como ilustrado na Figura 20. A manutenção de mapas de alta resolução provou ser eficaz, e a HRNet foi posteriormente incorporada em detectores de objetos.

Figura 20 – Arquitetura da HRNet



Fonte: Sun et al. (2019)

No entanto, como a saída da HRNet consiste em quatro mapas de características em diferentes resoluções, não há necessidade de utilizar a FPN para fusão de características, uma vez que a própria rede já desempenha essa função. Para integrar as saídas da HRNet em detectores eficientemente, foi proposto o *High-resolution Feature Pyramid Network*

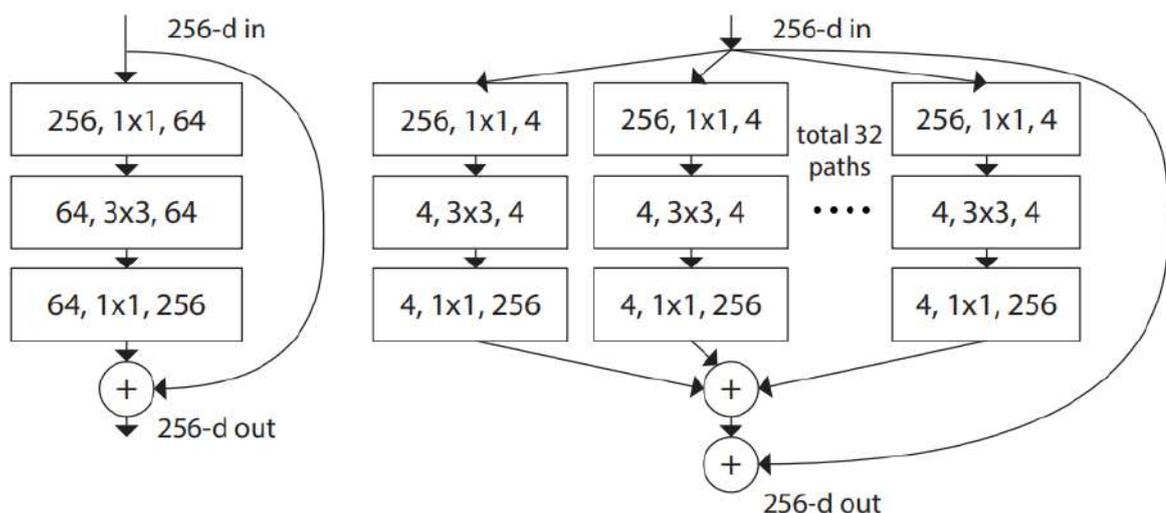
(HRFPN), que realiza a fusão das saídas por meio de *upsampling* e concatenação. Com o uso do HRFPN, tornou-se possível aproveitar a HRNet de maneira eficiente em detectores de objetos, maximizando seu potencial para tarefas de detecção.

Para evitar um aumento excessivo no número de parâmetros da rede em função da resolução das características, foram utilizados blocos *bottleneck*, conforme proposto nas arquiteturas *ResNet* e *ResNeXt*, para a extração de características.

O bloco *bottleneck* clássico é composto por três convoluções sequenciais e uma conexão de atalho (*skip connection*), conforme ilustrado na Figura 21. Esse bloco é mais eficiente do que as convoluções tradicionais, ao iniciar com uma convolução 1×1 para reduzir a dimensionalidade, seguida por uma convolução 3×3 em um volume menor, e, por fim, expande a dimensionalidade novamente. Esse processo reduz consideravelmente o número de parâmetros, prevenindo a explosão do gradiente. Além disso, a *skip connection* facilita a propagação do gradiente pelo modelo, proporcionando um fluxo suave e atenuando o problema do desaparecimento do gradiente.

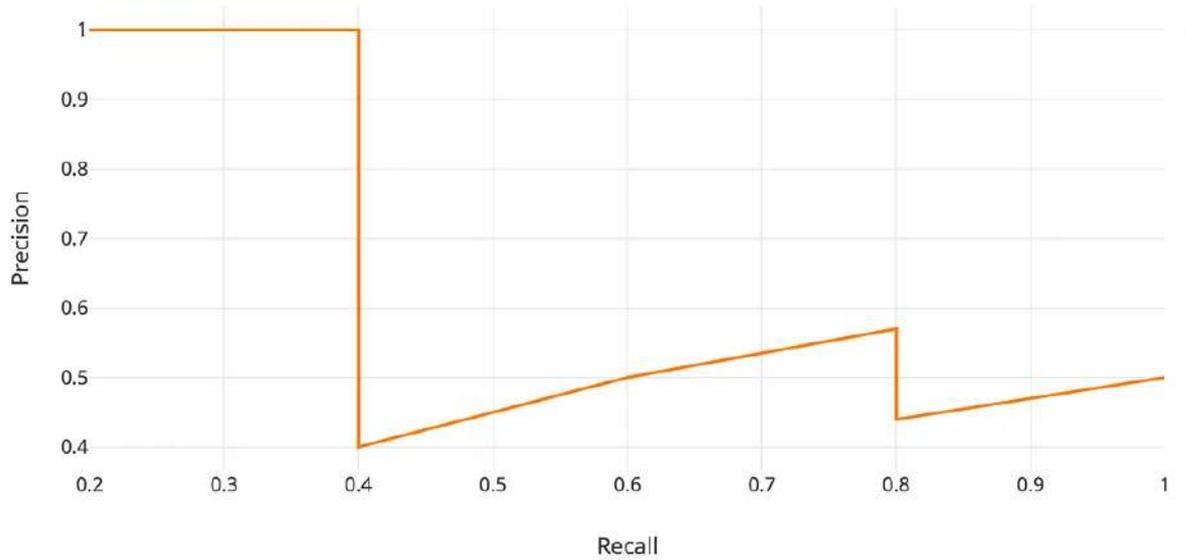
Em seguida, propôs-se a utilização de agregação no *bottleneck* da *ResNet*, o que possibilita a execução de convoluções em paralelo e a agregação por soma, como ilustrado na Figura 21. Ao empregar mais grupos de convoluções, em vez de aumentar o número de filtros, foi possível obter uma maior diversidade de características, resultando em uma melhor generalização do modelo.

Figura 21 – Bloco *Bottleneck* da *ResNet*



A esquerda o *bottleneck* da *ResNet* e a direita o da *ResNeXt*. Adaptado de Xie et al. (2017)

Figura 22 – Exemplo de Curva Precisão x Sensibilidade



Fonte: Hui (2018)

2.3 Métricas de Avaliação

A métrica mais popular em detecção de objetos é o mAP. Essa métrica foi proposta em 2007 no desafio [Everingham et al. \(\)](#) e se popularizou nos desafios seguintes na base *Common Objects in Context* (COCO) ([LIN et al., 2014](#)). Esta métrica se tornou o padrão de ouro para avaliação de modelos de detecção de objetos, pois avalia tanto a precisão quanto a sensibilidade considerando diferentes limiares de IoU e limiares de confiança.

Sendo o *Average Precision* (AP) definido como área sob a curva de sensibilidade x precisão, dada pela Equação 2.10 para cada classe e visualizado na Figura 22.

$$AP = \int_0^1 p(r) dr \quad (2.10)$$

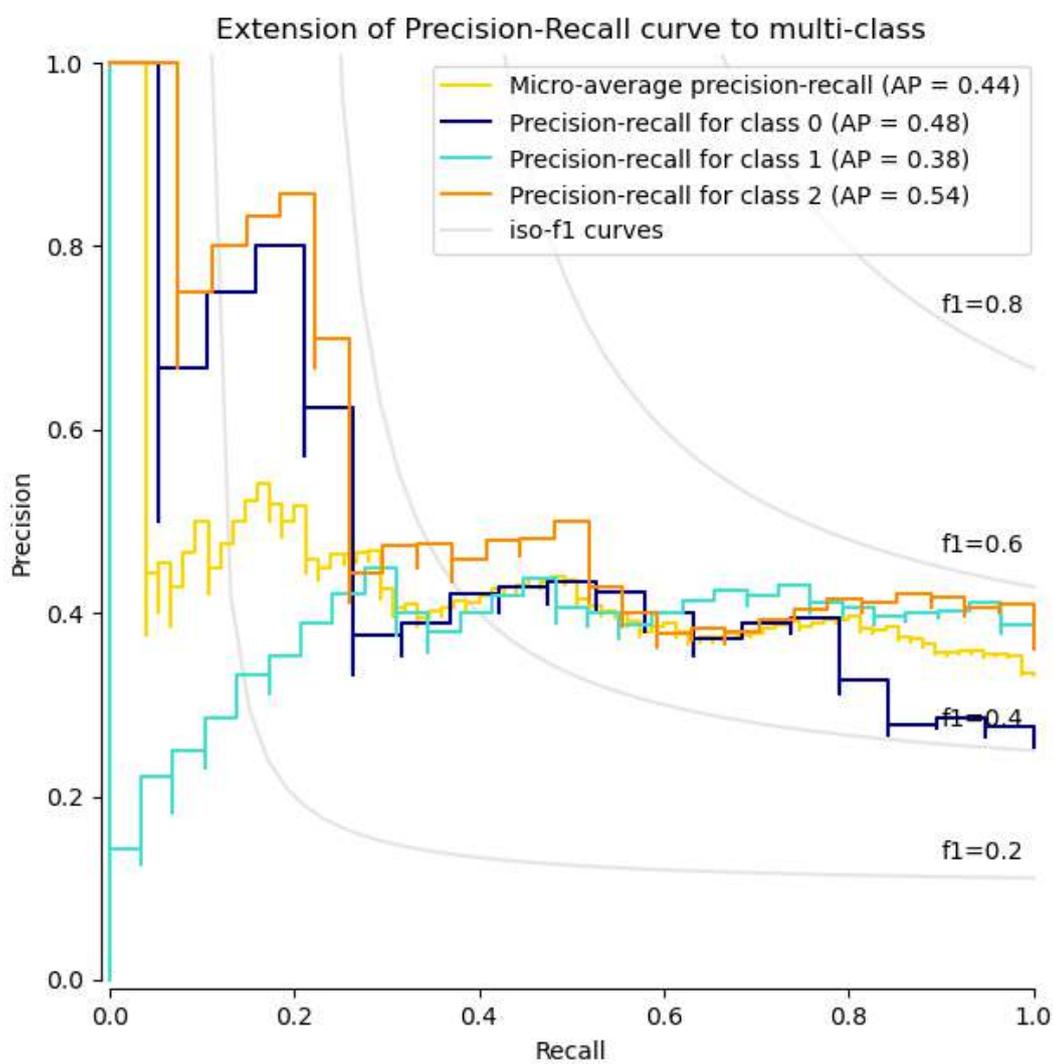
Portanto, o mAP é definido como a média dos AP sobre todas as classes e expresso na Equação 2.11 e podemos visualizar as múltiplas curvas na Figura 23;

$$mAP = \frac{1}{n} \sum_{i=1}^{i=n} AP_i, \quad (2.11)$$

onde AP_k é o AP da i -ésima classe e n é número de classes.

Por sua vez, para medir somente o ajuste das caixas delimitadoras da detecção, se utiliza o *mean Intersection Over Union* (mIoU), pois este desconsidera a classe da detecção e mede somente o ajuste das caixas. O mIoU é definido como a média da Intersecção sobre a União definida pela Equação para cada classe.

Figura 23 – Exemplo de Curvas Precisão x Sensibilidade para multi classe



Fonte: Hui (2018)

$$IoU = \frac{\text{Intersecção}}{\text{União}} \quad (2.12)$$

Métricas de classificação podem ser aplicadas na detecção de objetos, mas com adaptações necessárias. Métricas como acurácia, sensibilidade, precisão e *F1 score* avaliam apenas se a predição corresponde à classe correta ou não. Isso pode ser problemático para tarefas de detecção de objetos, ao ser preciso considerar fatores adicionais, como os limiares de IoU e a confiança do modelo nas predições.

Como observado na Figura 22 quando maior a precisão, menor a sensibilidade e vice-versa. Para evitar isso, é necessário estipular um limiar de confiança do modelo. Isto permite evitar que detecções de baixa confiança interfiram nas métricas. Utilizando essa abordagem as métricas de classificação podem utilizar a mesma definição para a detecção de objetos.

A acurácia é definida como previsões corretas sobre o total de precisões, sendo definida pela Equação 2.13.

$$\text{Acurácia} = \frac{VP + VN}{VP + VN + FP + FN} \quad (2.13)$$

A sensibilidade é definida como as observações corretas em relação a todas as observações, sendo definida pela Equação 2.14. Esta métrica permite avaliar se o modelo está deixando passar detecções.

$$\text{Sensibilidade} = \frac{VP}{VP + FN} \quad (2.14)$$

A precisão é definida como observações positivas corretas sobre o total de observações positivas, sendo definida pela Equação 2.15. Esta métrica permite avaliar se o modelo está classificando as detecções corretamente.

$$\text{Precisão} = \frac{VP}{VP + FP} \quad (2.15)$$

Por fim, o *F1 Score* é definido como a média harmônica da precisão e sensibilidade. Sendo assim, uma forma de avaliar ambas as métricas em conjunto. Esta é definida pela Equação 2.16.

$$F1 = 2 \frac{\text{Precisão} * \text{Sensibilidade}}{\text{Precisão} + \text{Sensibilidade}} \quad (2.16)$$

2.4 Considerações Finais

Como mencionado na seção anterior, detectores de objetos em dois estágios demonstraram melhor desempenho na detecção de objetos. Isso ocorreu devido à divisão mais clara das tarefas: a RPN, uma rede específica e treinável, foi responsável pela localização inicial, enquanto o refinamento das regiões e a classificação foram realizados por cabeças de detecção desacopladas. Essa abordagem proporcionou maior precisão nas detecções, embora com um custo de maior tempo de inferência.

Em aplicações médicas, com exceção dos atendimentos de emergência — como na maioria dos processos laboratoriais —, redes com maior desempenho foram preferidas. Nesses casos, a perda de tempo de inferência foi menos relevante em comparação aos riscos de um falso negativo em um exame.

A detecção de ovos de parasitas em imagens de microscópio apresentou um dos maiores desafios na detecção de objetos: lidar com grandes variações de escala. Isso ocorreu porque houve ovos extremamente pequenos, que ocuparam apenas alguns pixel da imagem, e ovos maiores, que ocuparam uma porção significativa. Objetos pequenos, em particular, foram difíceis de detectar devido à baixa resolução dos mapas de características.

Para enfrentar esse problema, empregou-se a HRNet em conjunto com o HRFPN para melhorar a detecção de objetos. Como discutido no Capítulo 3, a combinação da *Cascade R-CNN* com a HRNet resultou em uma arquitetura de alto desempenho para a detecção de ovos de parasitas.

O aprimoramento de um detector em dois estágios com mapas de características em alta resolução tornou-o mais eficiente na detecção de ovos de parasitas que apresentaram grandes variações de escala. A utilização combinada dos resultados de múltiplos detectores pôde gerar uma detecção final mais confiável.

3 Trabalhos Relacionados

Este capítulo apresenta uma revisão da literatura relacionada à aplicação de técnicas de aprendizado profundo na análise de imagens de ovos de parasito humanos obtidas por microscópio. O objetivo central é mapear métodos e sistemas automatizados capazes de detectar ovos de parasito em imagens de microscópio via aprendizado profundo.

O aprendizado Profundo, um ramo do aprendizado de máquina, surgiu visando eliminar a necessidade de extração manual de características. Esse método consegue aprender automaticamente as representações mais relevantes dos dados ao minimizar uma função de perda, permitindo a identificação de características significativas sem intervenção humana.

Devido à sua capacidade de aprender diretamente dos dados de entrada, o aprendizado profundo tem sido amplamente utilizado para automatizar diversas tarefas. Na área médica, ele tem se destacado especialmente no processamento de imagens de exames, oferecendo suporte aos profissionais de saúde no diagnóstico de doenças.

No contexto de imagens microscópicas, especialmente na detecção de ovos de parasitos, diversos estudos têm sido realizados visando automatizar essa tarefa de maneira eficiente e precisa. Esses trabalhos utilizam modelos de aprendizado profundo para detecção, integrando técnicas de pré-processamento e pós-processamento que aprimoram os resultados obtidos pelos modelos.

O estudo de [Viet, ThanhTuyen e Hoang \(2019\)](#) teve como objetivo a detecção automática de ovos de parasitos em imagens de microscopia por meio da *Faster R-CNN*. Utilizando uma base de 246 imagens da *Kansas State University*, abrangendo 8 classes de parasitos, o modelo empregou a *VGG16* como extrator de características e um limiar de Intersecção sobre a União (IoU) de 0,7. Diversos experimentos foram conduzidos para otimizar parâmetros como taxa de aprendizado, número de iterações e configurações de âncoras. Destaca-se a adaptação de um detector complexo a um contexto de baixo volume de dados.

Diferentemente do estudo anterior, [Yang et al. \(2019\)](#) propõem o uso do modelo *Faster R-CNN* para detectar parasitos de malária em 2.967 amostras de sangue analisadas por meio de *smartphones*. Os autores desenvolveram uma arquitetura simplificada de extração de características, composta por 4 camadas convolucionais e 2 camadas de *max-pooling*. Para facilitar o processamento, as imagens foram segmentadas em *patches* de 252×189 pixels.

[Pho, Amin e Yoshitaka \(2019\)](#) propõem uma abordagem inovadora para a detecção

e segmentação de protozoários em 129 imagens de microscópio, combinando a *RetinaNet* com uma rede codificador-decodificador. O método é composto em múltiplas etapas: inicialmente, a *RetinaNet* detecta os protozoários por meio de caixas delimitadoras. Em seguida, o *Feature Pyramid Network* (FPN) da *RetinaNet* é invertido para formar uma rede *encoder-decoder*, que gera a máscara de segmentação correspondente. Essa máscara é então aplicada à imagem original por multiplicação elemento a elemento, eliminando o fundo indesejado. Por fim, a imagem segmentada é processada por uma segunda *RetinaNet*, treinada separadamente, para a classificação dos protozoários identificados. A principal contribuição deste trabalho está na proposição do modelo de duas etapas sequenciais no contexto do problema.

Um *Single Shot MultiBox Detector* (SSD) com *InceptionV2* como *backbone* para extração de características foi proposto em (MATHISON et al., 2020) para a detecção de parasitos intestinais em imagens microscópicas. Este trabalho contribuiu aplicando um modelo de detecção de um estágio para a detecção ampla de múltiplas classes de parasito em um grande conjunto de imagens.

Mais recentemente, Anantrasirichai et al. (2022) apresentaram um catálogo dos resultados dos métodos submetidos ao desafio *ICIP 2022 Challenge on Parasitic Egg Detection and Classification in Microscopic Images*, destacando o conjunto de dados, as abordagens utilizadas e os resultados obtidos. Este desafio torna pública a base de imagens *Chula-ParasiteEgg11*, que conta com 11.000 imagens de treino e 2.200 imagens teste e 11 classes de ovos de parasito em imagens de microscópio. O desafio impulsionou múltiplas publicações relacionadas à aplicação de aprendizado profundo na tarefa de detecção de ovos de parasito. Alguns trabalhos, embora citados pelos organizadores, não foram publicados abertamente para uma revisão de seus métodos.

Pedraza et al. (2022) investigam o uso de *transformers* na detecção de parasitos na base *Chula-ParasiteEgg-11*, testando diversas arquiteturas, incluindo *Detection Transformer* (DETR), *Mask R-CNN* e *Cascade Mask R-CNN* com diferentes *backbones* baseados em *transformers*. O estudo emprega normalização, redimensionamento e aumentos de dados padrão, ajustando os detectores para treinar sem segmentação devido à ausência de máscaras no conjunto de dados.

Já Tureckova, Turecek e Oplatkova (2022) investigam a detecção de pequenos ovos de parasito no mesmo *dataset* aplicando a técnica *Slicing Aided Hyper Inference* (SAHI). O método divide as imagens em fragmentos menores para treinar um detector, combinando as detecções na inferência com supressão não máxima para eliminar sobreposições. Os autores aplicam a rede *Task Aligned One-Stage Object Detection* (TOOD) com *ResNet-101*, treinada com validação cruzada de cinco *folds*. A previsão final é gerada por um ensemble dos modelos, empregando uma abordagem semelhante ao *Weighted Box Fusion* (WBF), resultando em maior precisão na detecção de parasitos em imagens microscópicas.

Na linha de uso de ensembles, [Ruiz-Santaquiteria et al. \(2022\)](#) exploram o uso de ensembles no mesmo dataset. O método combina predições de cinco modelos distintos, incluindo *Faster R-CNN* com *ResNet-50*, *TOOD* com *ResNet-101*, *Cascade Mask R-CNN* com *Swin-Transformer* e *ConvNeXt-Large*, além de *YOLOX* com *Darknet53*. A técnica de WBF foi também empregada para otimizar a fusão das detecções.

O trabalho [Pratama et al. \(2022\)](#) explora a utilização da *YOLOv5x* para detectar ovos de parasitos no desafio *ICIP 2022*, com ênfase na ampliação do conjunto de dados. O conjunto foi expandido em 7 vezes, totalizando 78.000 imagens, incluindo 77.000 geradas por aumento de dados e 1.000 sintéticas. As técnicas de aumento de dados aplicadas incluíram desfoque médio, inversões, rotação, ajuste de matiz e restauração de imagens usando um método de espaço latente profundo. Para gerar as imagens sintéticas, ovos de parasitos foram recortados com o algoritmo *GrabCut* e inseridos em fundos microscópicos com o uso de *Poisson Image Editing*. A principal contribuição deste trabalho está na melhoria da detecção de ovos de parasitos por meio da ampliação significativa da base de dados, tornando o modelo mais robusto e preciso.

Numa linha similar, [Aung, Srithaworn e Achakulvisut \(2022\)](#) propõem um método focado em aprimorar a detecção de ovos de parasitos por meio do treinamento com pseudo-máscaras de segmentação. Para isso, os autores utilizaram uma rede de segmentação profunda pré-treinada para gerar segmentações dos conjuntos de treino e teste, que foram então empregadas no treinamento de modelos para detecção e segmentação. Entre os modelos testados, destacam-se a *Hybrid Task Cascade* com HRNet e *ResNet101*, além de três modelos voltados exclusivamente para a detecção.

A pesquisa desenvolvida por [Pho et al. \(2022\)](#) apresenta como hipótese a detecção de ovos de parasitos no desafio por meio da segmentação guiada por atenção. Para isso, os autores utilizaram uma rede codificador-decodificador para gerar máscaras de segmentação, aplicando transferência de aprendizagem a partir de um modelo previamente treinado para segmentação de protozoários. Como as máscaras geradas continham ruído devido à mudança de domínio, foi incorporado um bloco de atenção com atenção guiada e auto-atenção para refinamento. A máscara refinada foi combinada com a entrada e utilizada como um mecanismo de realce antes de alimentar um modelo *RetinaNet*, responsável pela detecção. Contudo, os testes foram conduzidos em um subconjunto do conjunto de treinamento, e não no conjunto oficial de testes. Este trabalho se destaca pela utilização de segmentação como filtro de ruídos para imagens de microscópio.

Além dos modelos de aprendizagem profunda, destaca-se o trabalho de [Lee et al. \(2022\)](#) que propõe a *Helminth Egg Analysis Platform (HEAP)*, uma plataforma para a detecção automatizada de ovos de parasitos em imagens de microscópio. A plataforma dispõe de modelos pré-treinados, como *Faster R-CNN*, *U-Net* e *SSD*, mas também permite que os usuários integrem seus próprios modelos treinados. A base de imagens inicial para

Tabela 1 – Sumarização dos Trabalhos Relacionados a Base Chula-ParasiteEgg-11

Trabalho	Objetivo	Método	Contribuição	Resultado
Yang et al. (2019)	Deteção de Parasitos de Malaria	Faster R-CNN	Extrator de características leve	0,96 Sensibilidade
Pho, Amin e Yoshitaka (2019)	Deteção de Protozoários	RetinaNet	Utilizar segmentação como filtro para deteção	0,94 mAP
Mathison et al. (2020)	Deteção de Ovos de Parasito	SSD	Deteção de Múltiplas classes de parasito	0,98 PPA 0,98 NPA
Lee et al. (2022)	Plataforma de Deteção de Ovos de Parasito	HEAP	Framework comunitário para desenvolvimento de modelos	-
Pedraza et al. (2022)	Deteção de Ovos de Parasito	Cascade R-CNN	Utilização de transformer	0,875 mIoU 0,955 F1 Score
Tureckova, Turecek e Oplatkova (2022)	Deteção de Ovos de Parasito	TOOD	Utilização da técnica SAHI e ensemble	0,887 mIoU 0,964 F1 Score
Ruiz-Santaquiteria et al. (2022)	Deteção de Ovos de Parasito	Ensemble de Modelos	Utilização de 5 modelos para ensemble	0,915 mIoU 0,974 F1 Score
Pratama et al. (2022)	Deteção de Ovos de Parasito	YOLOv5	Extensivo aumento de dados	0,717 mIoU 0,864 F1 Score
Aung, Srithaworn e Achakulvisut (2022)	Deteção de Ovos de Parasito	Ensemble de Modelos	Utilização de pseudo máscaras de segmentação	0,933 mIoU 0,988 F1 Score
Pho et al. (2022)	Deteção de Ovos de Parasito	RetinaNet	Utilização de auto atenção para refinar segmentação	0,82 mAP 0,99 Precisão 0,87 Sensibilidade
Suwannaphong et al. (2023)	Deteção de Ovos de Parasito	ResNet50	Deteção através de probabilidades de patches	0,98 Acurácia 0,96 Precisão

treino é composta por 8.482 amostras de ovos de parasitos em 17 classes. Os modelos fornecidos foram desenvolvidos utilizando imagens de alta resolução com diferentes ajustes de foco, e a plataforma oferece a possibilidade de os usuários carregarem seus próprios conjuntos de dados para análise. No entanto, os autores não disponibilizam informações sobre o desempenho dos modelos em bases de referência.

A Tabela 3 apresenta resumidamente algumas informações dos trabalhos relacionados. Observamos melhoria nos resultados a partir da publicação do desafio ICIP 2022. Os modelos tornam-se cada vez mais robustos com a disponibilização pública da base *Chula-ParasiteEgg11*, permitindo o desenvolvimento de ensembles de detectores mais sofisticados, como a *Cascade R-CNN*, além da adoção de arquiteturas baseadas em *transformers*.

Os estudos iniciais, como os de Viet, ThanhTuyen e Hoang (2019) e Yang et al. (2019), adotaram detectores de dois estágios, priorizando precisão em detrimento da eficiência computacional. Posteriormente, houve uma migração para modelos de estágio único e abordagens híbridas, como a combinação de RetinaNet com redes codificador-decodificador em Pho, Amin e Yoshitaka (2019). Recentemente, o uso de transformers e ensembles emergiu como estratégia para aprimorar a robustez, embora muitas vezes à custa de complexidade e custo computacional. Nota-se, porém, que a maioria das arquiteturas empregadas são adaptações de modelos genéricos, sem inovações específicas para as peculiaridades do domínio.

A introdução do dataset Chula-ParasiteEgg11 no desafio ICIP 2022 representou um divisor de águas, permitindo a comparação direta de métodos e o desenvolvimento de técnicas mais sofisticadas. Trabalhos como os de Pratama et al. (2022) e Tureckova, Turecek e Oplatkova (2022) exploraram estratégias de aumento de dados e fragmentação de imagens (SAHI), respectivamente, para mitigar limitações de volume e resolução. Contudo,

a dependência excessiva de ampliação sintética levanta questões sobre a generalização em cenários clínicos reais, onde a variabilidade morfológica e a qualidade das imagens podem divergir significativamente.

Embora os estudos tenham avançado no uso da base *Chula-ParasiteEgg11*, ainda existem lacunas inexploradas. São empregadas arquiteturas genéricas para a tarefa, combinadas por meio de ensembles. Embora essa abordagem melhore as métricas e sejam o estado da arte, não geram contribuições significativas para a área, pois não investiga o aprimoramento de arquiteturas específicas para resolver o problema. Essa limitação está diretamente ligada às questões de pesquisa Q1, Q2 e Q4 na Seção 1.2.1.

3.1 Considerações finais

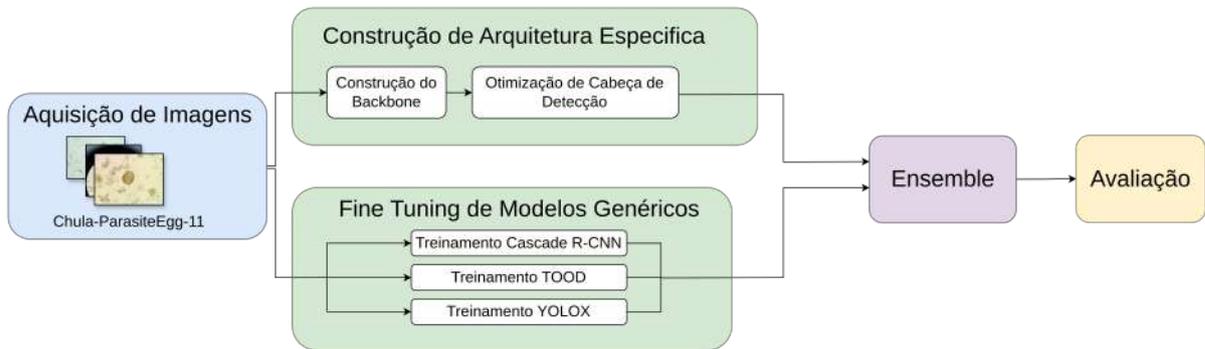
Este capítulo sintetizou o estado da arte em detecção automatizada de ovos de parasitos via aprendizado profundo, destacando a maturação do campo a partir da disponibilização de bases públicas e da adoção de arquiteturas modernas. Contudo, a análise crítica revela que o progresso recente tem sido incremental, centrado em ajustes metodológicos (e.g., ensembles, aumento de dados) em vez de inovações paradigmáticas.

As principais lacunas identificadas — como a carência de modelos especializados, a subexploração de técnicas de eficiência computacional e a ausência de avaliação clínica integrada — orientam as contribuições desta dissertação. Ao propor uma arquitetura otimizada para características microscópicas.

4 Materiais e Método

Este capítulo apresenta as etapas da metodologia desenvolvida. O objetivo é construir um modelo ensemble de detectores que combine uma arquitetura projetada especificamente para o problema e detectores genéricos. A metodologia é estruturada nas seguintes etapas: aquisição de imagens, construção da arquitetura específica, finetuning de modelos genéricos, ensemble e avaliação. O fluxo do processo metodológico é ilustrado na Figura 24.

Figura 24 – Etapas da Metodologia Proposta.



Elaborado pelo Autor

4.1 Aquisição de Imagens

Esta etapa pretende a obtenção de imagens de microscópio de ovos de parasito. Para atingir este objetivo, utiliza-se a base de imagens Chula-ParasiteEgg-11 ([ANANTRASIRICHAI et al., 2022](#)) publicamente disponibilizada em 2022 no desafio "*ICIP 2022 Challenge: Parasitic Egg Detection and Classification in Microscopic Images*".

Esta é a maior base pública atualmente de ovos de parasito disponível, contendo 11 tipos de parasitos e um total de 13.750 imagens. As classes estão balanceadas, com 1.000 imagens por classe no conjunto de treinamento e 250 imagens por classe no conjunto de teste. Vale destacar que o conjunto de teste é fornecido separadamente, garantindo consistência ao comparar os métodos desenvolvidos nesta base de dados.

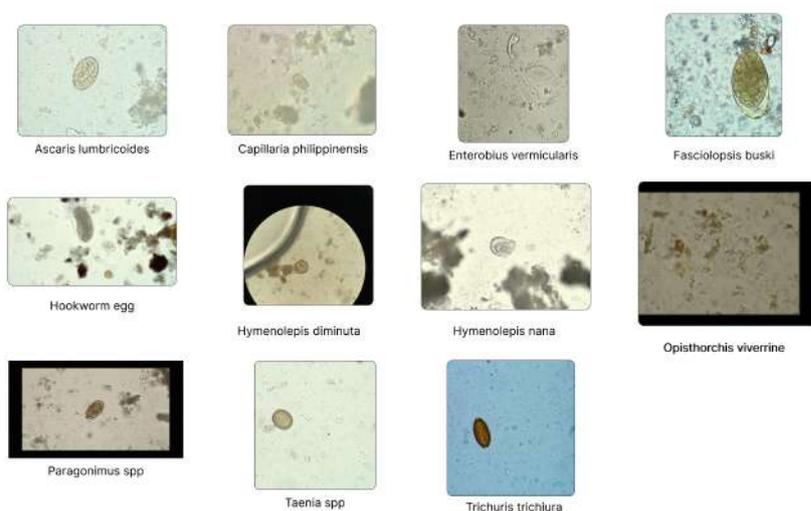
As amostras fecais foram preparadas em laboratório pelo *Department of Clinical Microscopy, Faculty of Allied Health Sciences, Chulalongkorn University*. As observações foram realizadas diretamente em um microscópio óptico com ampliação de 40x. Três especialistas conduziram a detecção, localizando os ovos de parasitos com caixas delimitadoras e classificando a espécie. A anotação final foi definida por votação.

Durante o processo de análise e aquisição das imagens, foram utilizados diversos microscópios e câmeras, o que resultou em variações de resolução, foco, brilho e na presença de artefatos. Adicionalmente, os autores do desafio adicionaram degradações às imagens para tornar o conjunto de imagens mais desafiador. As técnicas utilizadas foram:

- Recorte aleatório de 0 a 30% em um dos lados da imagem;
- *Gaussian Blur*;
- *Motion Blur*;
- Ruído Gaussiano;
- Ruído *Poisson*;
- Variações de Saturação;
- Variações de Contraste.

A aplicação destes efeitos gera variações para imagens de câmera e microscópio iguais, como pode ser visto na Figura 25.

Figura 25 – Amostras das classes da Chula-Parasite-Egg11



Fonte: [Anantrasirichai et al. \(2022\)](#)

4.2 Construção de Arquitetura Específica

Esta etapa abrange a construção de uma arquitetura específica para detecção e classificação de parasitos.

A arquitetura HRNet (Seção 2.2.5) foi selecionada como *backbone* principal do método. Essa escolha é fundamentada pela capacidade da rede de manter mapas de alta resolução em todas as etapas de processamento, contrastando com abordagens que reduzem a resolução espacial e tentam recuperar detalhes posteriormente via *upsampling*. A preservação contínua de resolução é particularmente relevante para detecção de objetos menores, uma vez que mapas de alta resolução retêm mais pixels representando o objeto, capturando características refinadas (como bordas e texturas) essenciais para distinguir instâncias pequenas ou parcialmente ocluídas.

Como cabeça de detecção (Seção 2.2.3), adotou-se a cabeça da Cascade R-CNN. A fixação da cabeça de detecção assegura consistência nos resultados, evitando variações não controladas que poderiam surgir utilizando diferentes cabeças de detecção.

O objetivo final desta etapa é construir um detector composto por *backbone* e cabeça de detecção específico para detecção de ovos de parasito em imagens de microscópio.

4.2.1 Construção de *Backbone*

Nesta subetapa, são otimizadas diferentes variações de blocos e da função de perda. Para gerar as predições, a cabeça de detecção da Cascade R-CNN (Seção 2.2.4.5) é mantida fixa. Cada arquitetura desenvolvida é avaliada com base na métrica mAP (Seção 2.3). O objetivo desta etapa é projetar uma versão otimizada da HRNet para a detecção de ovos de parasito.

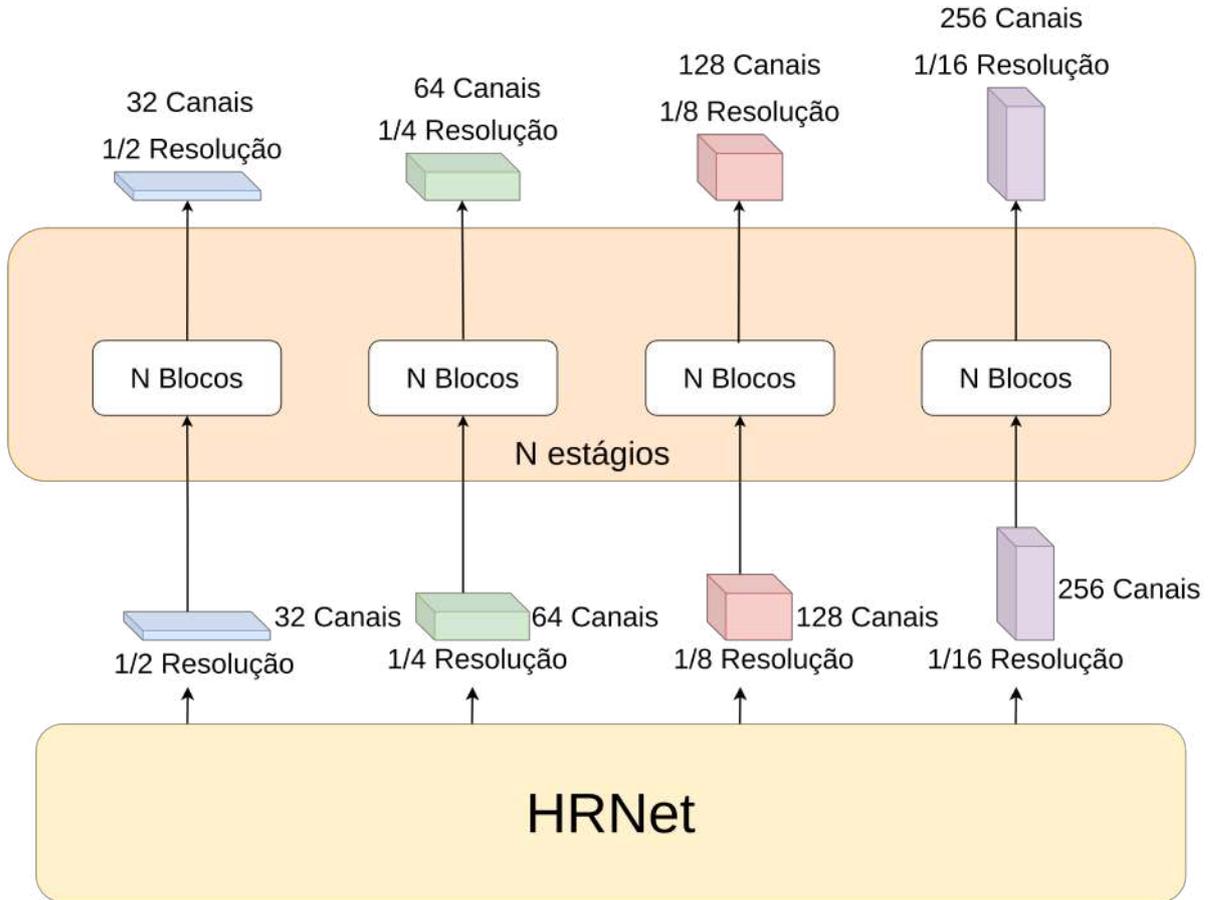
A estrutura original da HRNet é mantida, com a adição de blocos ao final da arquitetura. Essa abordagem permite o aproveitamento dos pesos pré-treinados na ImageNet, acelerando o processo de treinamento. Cada conjunto de blocos inseridos é denominado estágio. Essa forma de extensão de estágios se relaciona à questão de pesquisa Q1 (Seção 1.2.1). O modelo de otimização de blocos está ilustrado na Figura 26.

Os blocos escolhidos para otimização nesta etapa foram os *bottlenecks* das arquiteturas ResNet (HE et al., 2016) e ResNeXt (XIE et al., 2017). O bloco da ResNet foi selecionado por ser utilizado na arquitetura original da HRNet, enquanto o da ResNeXt apresenta avanços na agregação de características em relação ao *bottleneck* tradicional.

Adicionalmente, foi utilizado o Bloco Residual Inverso, introduzido na arquitetura da MobileNet (HOWARD et al., 2017). Esse bloco pode ser visto como uma versão mais eficiente do *bottleneck*, na qual os canais são inicialmente expandidos para capturar mais informações e, em seguida, comprimidos para reduzir a complexidade computacional. Este bloco foi escolhido por seguir a estrutura residual do bloco de *bottleneck* e por apresentar menor custo computacional.

Também é avaliada a composição da função de perda durante o treinamento. Para isso, são capturadas tanto a última saída da arquitetura original da HRNet quanto as

Figura 26 – Modelo de Otimização da HRNet.



Elaborado pelo Autor

saídas de todos os estágios adicionais subsequentes. Assim, para n estágios extras, obtemos um total de $n + 1$ saídas. Como tanto a HRNet quanto cada estágio adicional geram 4 mapas de características, isso resulta em $n+1$ conjuntos, cada um contendo 4 mapas de características.

Considerando as resoluções e números de canais da arquitetura HRNet utilizada, vista na Figura 26, podemos expressar todas as saídas da arquitetura na Equação 4.1.

$$S_i = \{T_1^i, T_2^i, T_3^i, T_4^i\}, \forall i \in \{0, 1, \dots, n + 1\} \quad (4.1)$$

Durante a composição da função de perda, cada saída S_i será utilizada para calcular uma perda correspondente. Para isso, a cabeça de detecção do detector será empregada. Cada saída será processada individualmente pela cabeça de detecção, e as detecções resultantes serão utilizadas no cálculo da perda para essa saída. Portanto, cada saída S_i irá gerar uma perda L_i correspondente.

A perda final pode ser obtida dos L_i por média ou média ponderada, onde a perda

de cada saída S_i possui um peso W_i associado. Essa metodologia relaciona-se à questão de pesquisa Q2 (Seção 1.2.1).

A perda composta por média descrita na Equação 4.2 e a perda por média ponderada está descrita na Equação 4.3.

$$L_{\text{final}} = \frac{1}{n+1} \sum_{i=1}^{n+1} L_i \quad (4.2)$$

$$L_{\text{final}} = \frac{1}{n+1} \sum_{i=1}^{n+1} W_i * L_i \quad (4.3)$$

4.3 Fine Tuning de Modelos Genéricos

Esta etapa pretende construir e treinar três modelos para a detecção de ovos de parasitos em imagens de microscópio. Para isso, além do desenvolvimento de uma arquitetura específica, são selecionados modelos genéricos amplamente utilizados na literatura. Esses modelos, em conjunto com a arquitetura específica, comporão o ensemble abordado na próxima seção.

A seleção dos modelos genéricos foi baseada em sua relevância nos trabalhos relacionados. O primeiro modelo escolhido foi a Cascade R-CNN, devido à sua ampla adoção e desempenho consistente na literatura. O segundo modelo, TOOD, foi selecionado com base nos resultados apresentados por Tureckova, Turecek e Oplatkova (2022). Por fim, a CenterNet é escolhida pela sua inovadora forma de calcular as coordenadas das caixas delimitadores. Esta utiliza regressão para calcular a coordenada do centro do objeto, o que é útil para objetos que possuem formas bem determinadas, como os ovos de parasito.

Os modelos são treinados com o *backbone* ResNet-101, conforme descrito em Tureckova, Turecek e Oplatkova (2022). A avaliação é feita individualmente utilizando a métrica mAP 2.3.

Como resultado, esta etapa entrega os três melhores modelos genéricos da literatura, treinados na base de imagens. Espera-se que esses modelos contribuam para melhorar as detecções no ensemble, em conjunto com a arquitetura específica.

4.4 Ensemble

O objetivo desta etapa é combinar as detecções da Arquitetura Específica com as dos Modelos Genéricos, visando gerar um resultado aprimorado. A expectativa é consolidar as acurácias e minimizar os erros individuais de cada modelo. Para essa integração, utiliza-se a técnica de Weighted Box Fusion (WBF) (SOLOVYEV; WANG; GABRUSEVA, 2021).

O WBF é conhecido por gerar uma caixa delimitadora média a partir das caixas individuais de cada modelo. A confiança da nova caixa é obtida calculando a média das confianças das caixas. As novas coordenadas X_1 , Y_1 , X_2 e Y_2 são calculadas via uma média ponderada. Inicialmente é montado um grupo onde temos todas as caixas referentes ao mesmo objeto. E para este grupo calcula-se a confiança média, utilizando a Equação 4.4, onde T é o número de caixas do grupo. As novas coordenadas são calculadas mediante uma média ponderada nas equações 4.5, 4.6, 4.7 e 4.8. Por fim, a pontuação de confiança final é ajustada utilizando a Equação 4.9.

Isto permite que as caixas delimitadores sejam combinadas considerando a confiança de cada modelo em sua predição. Deste modo, caixas de menor confiança terão menor influência na caixa combinada final.

$$C_{avg} = \frac{C_1 + C_2 + \dots + C_T}{T}, \quad (4.4)$$

onde C_{avg} é a média de confiança, C_i é a confiança da i -ésima detecção e T é o número de detecções.

$$X_1 = \frac{C_1 * X1_1 + C_2 * X1_2 + \dots + C_T * X1_T}{C_1 + C_2 + \dots + C_T}, \quad (4.5)$$

onde X_1 é a nova coordenada, C_i é a confiança da i -ésima detecção, $X1_i$ é a coordenada $X1$ da i -ésima detecção e T é o número de detecções.

$$X_2 = \frac{C_1 * X2_1 + C_2 * X2_2 + \dots + C_T * X2_T}{C_1 + C_2 + \dots + C_T}, \quad (4.6)$$

onde X_2 é a nova coordenada, C_i é a confiança da i -ésima detecção, $X2_i$ é a coordenada $X2$ da i -ésima detecção e T é o número de detecções.

$$Y_1 = \frac{C_1 * Y1_1 + C_2 * Y1_2 + \dots + C_T * Y1_T}{C_1 + C_2 + \dots + C_T}, \quad (4.7)$$

onde Y_1 é a nova coordenada, C_i é a confiança da i -ésima detecção, $Y1_i$ é a coordenada $Y1$ da i -ésima detecção e T é o número de detecções.

$$Y_2 = \frac{C_1 * Y2_1 + C_2 * Y2_2 + \dots + C_T * Y2_T}{C_1 + C_2 + \dots + C_T}, \quad (4.8)$$

onde Y_2 é a nova coordenada, C_i é a confiança da i -ésima detecção, $Y2_i$ é a coordenada $Y2$ da i -ésima detecção e T é o número de detecções.

$$C = C_{avg} * \frac{T}{N}, \quad (4.9)$$

onde C é a confiança final, C_{avg} é a média de confiança, T é o número de detecções e N é o número de modelos utilizados.

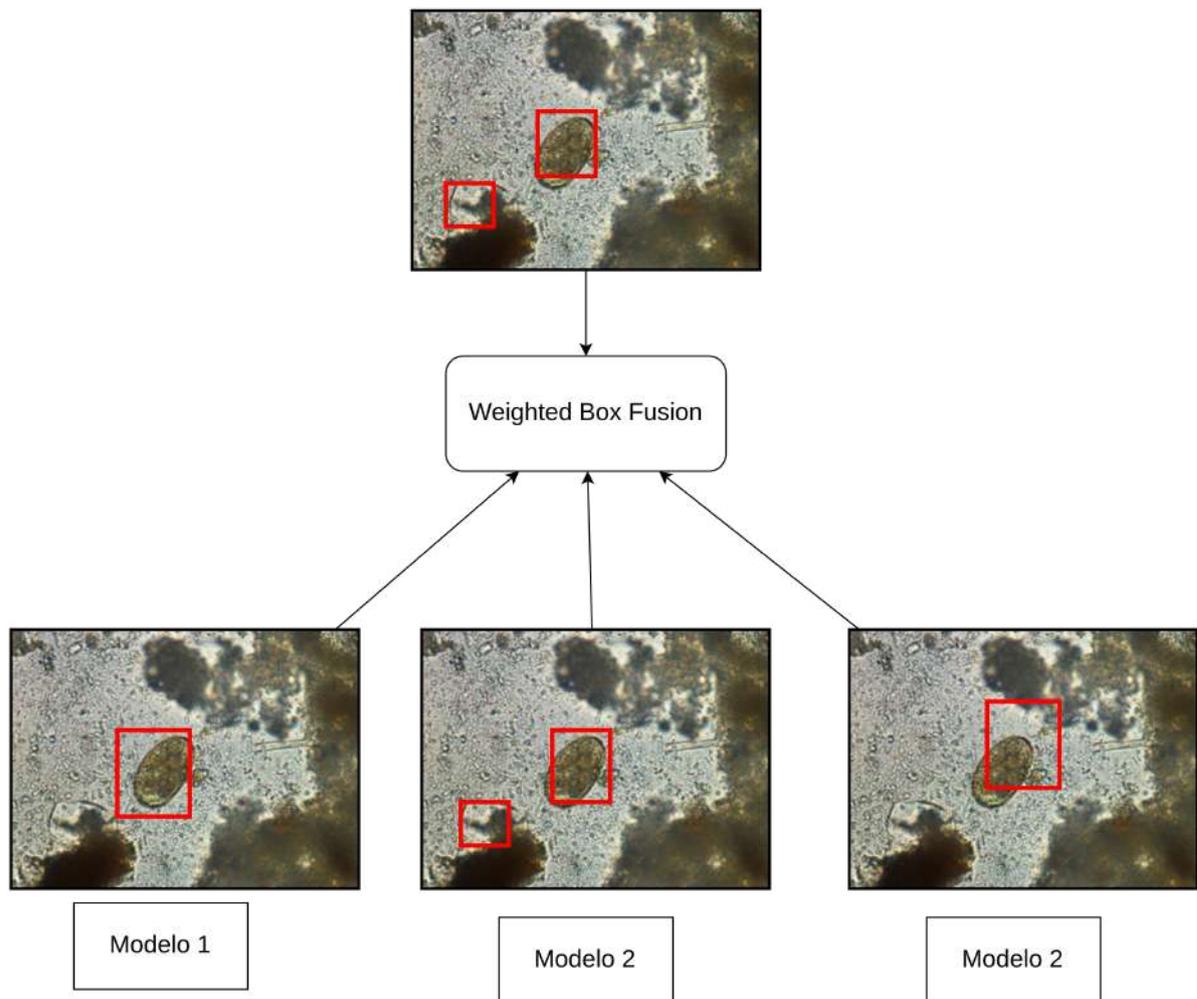


Figura 27 – Exemplo de *ensemble* com o WBF.

Esta técnica se utiliza de todas as caixas delimitadoras, assim não descartando a contribuição de nenhum dos modelos. Isto permite uma melhor formação de uma detecção final, pois modelos com elevado número de parâmetros tendem a se ajustar fortemente ao conjunto de treino. Este processo está ilustrado na Figura 4.4

A aplicação de um ensemble como etapa final permite reduzir os falsos negativos ao combinar as predições de múltiplos modelos. Essa abordagem possibilita responder à questão de pesquisa Q3 (Seção 1.2.1).

4.5 Avaliação

Esta etapa procura fixar métricas e formas de avaliar quantitativa e qualitativamente as detecções produzidas pelo ensemble.

Em problemas de detecção, a métrica mais comumente empregada é o mAP. No

entanto, em estudos voltados à detecção de ovos de parasitos, observa-se uma variação significativa nas métricas utilizadas. Por essa razão, é importante adotar também as métricas aplicadas nesses trabalhos para permitir uma comparação consistente.

Além disso, para uma análise mais detalhada dos resultados, é indispensável o uso de uma matriz de confusão, que oferece uma visão mais abrangente sobre o desempenho do modelo nas diferentes classes.

Portanto, para uma avaliação ampla do resultado do modelo final, utilizam-se não somente a métrica de mAP, como também as métricas de classificação, como descritas na Seção 2.3. Assim, serão computadas as seguintes métricas para avaliação do modelo: mAP, mIoU, acurácia, sensibilidade, precisão e F1 Score.

4.6 Considerações Finais

Este capítulo descreveu as etapas da metodologia, especificando a abordagem adotada em cada fase. Inicialmente, abordou-se o desenvolvimento de modelos genéricos e de uma arquitetura específica para a detecção de ovos de parasitas em imagens de microscopia. Em seguida, apresentou-se a etapa de ensemble utilizando a técnica WBF e, por fim, foram detalhadas as métricas empregadas para avaliar o desempenho do ensemble. A análise e discussão dos resultados obtidos serão apresentadas no capítulo seguinte.

5 Resultados e Discussões

Este capítulo apresenta as configurações dos experimentos realizados e seus respectivos resultados. Os experimentos foram estruturados com base nas questões de pesquisa e objetivos estabelecidos, garantindo que cada etapa da metodologia tenha uma seção correspondente neste capítulo. Inicialmente, é feita a exploração da base de imagens, seguida pelos resultados da Construção da Arquitetura Específica. Em seguida, aborda-se a construção do *backbone* e a otimização da Cabeça de Detecção. O capítulo prossegue com o treinamento dos modelos genéricos *Cascade R-CNN*, *TOOD* e *CenterNet*, culminando na avaliação do ensemble por meio das métricas descritas na Seção 4.5. Para todos os experimentos, foram utilizados os mesmos conjuntos de treino, validação e teste, conforme detalhado na seção seguinte.

5.1 Aquisição de Imagens

Na primeira etapa metodológica, adotou-se a base *Chula-ParasiteEgg11*, divulgada publicamente em um desafio. Estudos anteriores já aplicaram técnicas de aprendizado profundo nessa base, o que permite comparação direta, uma vez que a maioria utiliza o mesmo conjunto de imagens de teste.

Esta base possui um conjunto de teste específico com 2.200 imagens, tendo 11.000 imagens para o treino, as quais foram divididas aleatoriamente estratificadamente em 8.800 (80%) imagens de treino e 2.200 (20%) imagens de validação.

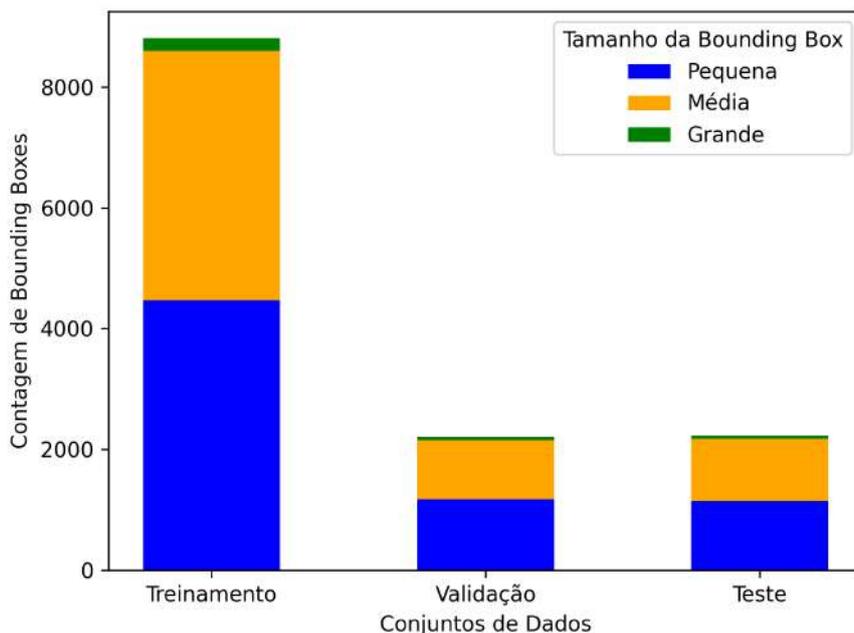
O conjunto de treino (8.800 imagens) possui 8.814 marcações de ovos, o que indica aproximadamente um ovo de parasito por imagem. Destas, 4.473 (50,7%) correspondem a ovos pequenos ($\leq 1\%$ da área da imagem), 4.132 (46,8%) correspondem a ovos médios ($1\% < x \leq 10\%$ da área da imagem), e 209 (2,5%) são grandes ($> 10\%$ da área da imagem).

o conjunto de validação (2.220 imagens) possui 2.208 marcações de ovos, também aproximadamente 1 ovo/imagem. Destas, 1.173 (53,12%) são ovos pequenos, 976 (44,2%) são ovos médios e 59 (2,6%) são ovos grandes.

O conjunto de teste (2.200 imagens) possui 2.208 marcações, das quais 1.137 (51,49%) são de ovos pequenos, 1.039 (47,05%) são ovos médios, e 52 (2,3%) são ovos grandes.

As anotações concentram-se em médias e pequenas, como apresentado na Figura 28, reforçando a necessidade de focar na detecção destes ovos de parasito menores e médios.

Algumas amostras da base de dados contêm um artefato que pode causar *data*

Figura 28 – Contagem de *Bounding Boxes* Pequenas, Médias e Grandes.

Elaborado pelo Autor

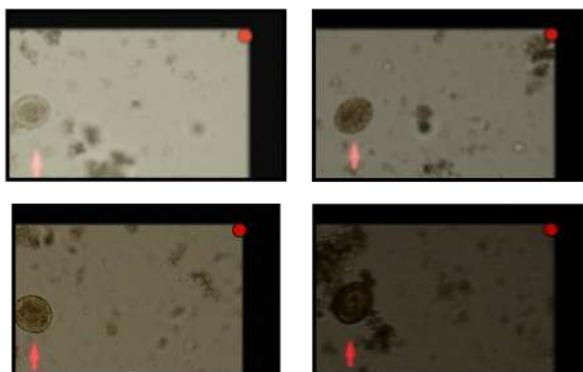
Figura 29 – Exemplos de amostras com artefato.



leakage: uma seta vermelha que aponta para o ovo do parasito na imagem, conforme ilustrado na Figura 29. Embora esse artefato não permita deduzir diretamente a classe do ovo, ele pode servir como um indicativo de localização para o modelo.

Outro artefato identificado na base de dados é um pequeno ponto vermelho no canto superior direito da imagem, presente em algumas amostras com *padding*. Embora esse artefato não esteja diretamente relacionado aos rótulos, sua recorrência pode influenciar o modelo. A Figura 30 ilustra esse padrão.

Figura 30 – Amostras com ponto vermelho



5.2 Construção de Arquitetura Específica

Nesta etapa, foram experimentadas diversas configurações de blocos e estágios adicionais na arquitetura da HRNet (Seção 2.2.5). Utilizando os blocos *BottleNeck* da *ResNet* (HE et al., 2016) e *ResNeXt* (XIE et al., 2017), além do Bloco Residual Inverso. Definiu-se um limite de 2 estágios extras e até 8 blocos extras por estágio. Para facilitar a análise das questões de pesquisa Q2 e Q1 (Seção 1.2.1), a seção foi subdividida conforme os diferentes métodos de composição de perda. Na composição de perda, utilizam-se as perdas intermediárias dos estágios extras, sendo que a saída da HRNet e a saída de cada estágio extra subsequente contribuem para a formação da perda final. Para consistência dos experimentos, todos os modelos foram treinados utilizando precisão mista (*mixed precision*) (MICIKEVICIUS et al., 2017), e *Stochastic Gradient Descent* (SGD) como otimizador, imagens em resolução de 1024x1024 pixels e 50 épocas. Nesta etapa, as arquiteturas são avaliadas com base no conjunto de validação, enquanto as métricas no teste são utilizadas para análise. Esse cuidado evita a introdução de viés na seleção do melhor modelo, garantindo que a escolha seja baseada no desempenho no conjunto de validação e não no de teste.

5.2.1 Perda Composta por Média

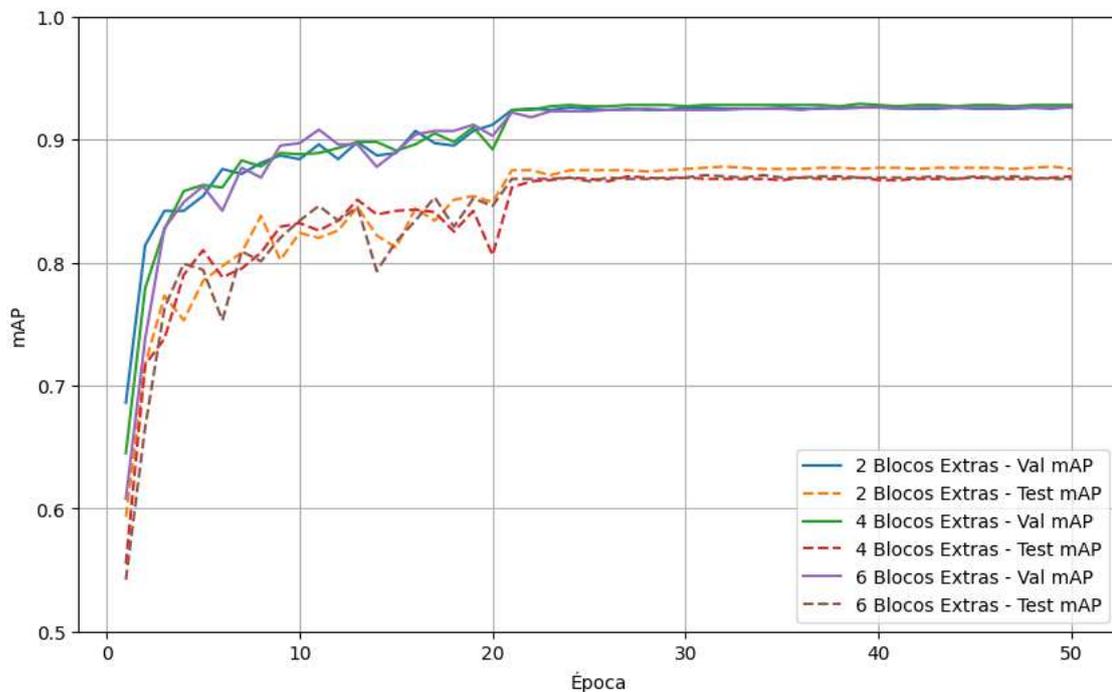
A primeira etapa dos experimentos envolve a otimização dos estágios e blocos extras por meio de uma composição de perda calculada pela média, conforme a Equação 4.2. Conforme demonstrado na Tabela 2, a adição de mais blocos ou estágios não se reflete em uma melhora da métrica de mAP, mas na maioria dos casos resulta em piora. Ao passar para a utilização de dois estágios extras, encontramos dissipação do gradiente, não sendo mais possível calcular a perda. Esses resultados indicam que o uso de uma média pura para compor a perda gera problemas de gradiente em arquiteturas mais profundas, dificultando o treinamento.

Embora os modelos utilizando o bloco de *bottleneck* tenham performado bem

Tabela 2 – Resultados Otimização Utilizando Perda Composta por Média.

Modelo Base	Estágios Extras	Tipo de Bloco	Blocos Extras	mAP Validação
HRNet-W32	1	ResNeXt	2	0,927
HRNet-W32	1	ResNeXt	4	0,924
HRNet-W32	1	ResNeXt	8	0,907
HRNet-W32	2	ResNeXt	2	-
HRNet-W32	2	ResNeXt	4	-
HRNet-W32	1	ResNet	2	0,927
HRNet-W32	1	ResNet	4	0,929
HRNet-W32	1	ResNet	8	0,926
HRNet-W32	2	ResNet	2	-
HRNet-W32	2	ResNet	4	-
HRNet-W32	1	Bloco Residual Inverso	2	0,928
HRNet-W32	1	Bloco Residual Inverso	4	0,925
HRNet-W32	1	Bloco Residual Inverso	8	0,926
HRNet-W32	2	Bloco Residual Inverso	2	-
HRNet-W32	2	Bloco Residual Inverso	4	-

durante sua validação, como indica o gráfico na Figura 31, o mesmo não se replica para o conjunto de teste. O que indica que os blocos adicionados ao final da rede podem estar causando *overfitting*.

Figura 31 – Resultados da Validação e Teste para HRNet utilizando bloco *Bottleneck* da *ResNet*

5.2.2 Perda Composta por Média Ponderada

A ponderação da perda em cada estágio para calcular a média final utiliza pesos de 1, 0,5 e 0,25, respectivamente, para a perda da última camada e das camadas extras

subsequentes. Esses valores foram escolhidos de forma empírica. Essa estratégia mitigou a dissipação do gradiente em arquiteturas mais profundas. Conforme observado anteriormente, ao utilizar uma média simples, modelos com mais de um estágio apresentaram dissipação do gradiente. Conforme apresentado na Tabela 3, somente um modelo sofreu esse problema, permitindo alcançar o melhor resultado de mAP no conjunto de validação e teste em comparação com os demais.

Tabela 3 – Resultados Otimização Utilizando Perda Composta por Média Ponderada.

Modelo Base	Estágios Extras	Tipo de Bloco	Blocos Extras	mAP Validação
HRNet-W32	1	ResNeXt	2	0,932
HRNet-W32	1	ResNeXt	4	0,932
HRNet-W32	1	ResNeXt	8	0,930
HRNet-W32	2	ResNeXt	2	0,929
HRNet-W32	2	ResNeXt	4	0,931
HRNet-W32	1	ResNet	2	0,929
HRNet-W32	1	ResNet	4	0,929
HRNet-W32	1	ResNet	8	0,932
HRNet-W32	2	ResNet	2	-
HRNet-W32	2	ResNet	4	-
HRNet-W32	1	Bloco Residual Inverso	2	0,933
HRNet-W32	1	Bloco Residual Inverso	4	0,929
HRNet-W32	1	Bloco Residual Inverso	8	0,928
HRNet-W32	2	Bloco Residual Inverso	2	0,929
HRNet-W32	2	Bloco Residual Inverso	4	-

No entanto, a diferença entre as métricas no conjunto de validação e teste ainda persiste, o que indica que a ponderação das perdas intermediárias não proporciona regularização suficiente para eliminar os efeitos de *overfitting* durante o treinamento.

5.2.3 Sem perda Composta

A ausência de composição de perda resulta em desempenho superior em comparação com o uso de uma perda composta pela média, como mostrado na Tabela 4. No entanto, ainda apresenta resultados inferiores à composição baseada em média ponderada.

Isso sugere que a composição por média simples é menos eficaz do que a ausência de composição de perda. Dessa forma, podemos responder à questão de pesquisa Q2, concluindo que a eficácia da composição de perdas intermediárias depende da abordagem adotada. Enquanto a média simples pode desestabilizar o treinamento, a média ponderada tende a estabilizá-lo, proporcionando melhores resultados e evitando a dissipação do gradiente.

Além disso, foram treinadas variações da HRNet com um número maior de canais por ramo. A primeira variação, denominada HRNet-W40, utiliza 40, 80, 160 e 320 canais para os respectivos ramos, enquanto a segunda, HRNet-W48, emprega 48, 120, 200 e 512 canais. Ambas as variações não alcançaram sucesso no treinamento devido à dissipação do

Tabela 4 – Resultados Otimização Sem Composição de Perda.

Modelo Base	Estágios Extras	Tipo de Bloco	Blocos Extras	mAP Validação
HRNet-W32	1	ResNeXt	2	0,929
HRNet-W32	1	ResNeXt	4	0,928
HRNet-W32	1	ResNeXt	8	-
HRNet-W32	2	ResNeXt	2	0,929
HRNet-W32	2	ResNeXt	4	-
HRNet-W32	1	ResNet	2	0,931
HRNet-W32	1	ResNet	4	0,926
HRNet-W32	1	ResNet	8	-
HRNet-W32	2	ResNet	2	0,928
HRNet-W32	2	ResNet	4	-
HRNet-W32	1	Bloco Residual Inverso	2	0,929
HRNet-W32	1	Bloco Residual Inverso	4	0,929
HRNet-W32	1	Bloco Residual Inverso	8	-
HRNet-W32	2	Bloco Residual Inverso	2	0,925
HRNet-W32	2	Bloco Residual Inverso	4	-

gradiente. Assim, em relação à questão de pesquisa Q1 (1.2.1), aumentar ingenuamente o número de canais na HRNet resulta em dissipação do gradiente; entretanto, a adição de estágios extras, aliada à composição de perda por média ponderada, pode favorecer um treinamento mais eficaz.

Finalizando, esta etapa define tanto a arquitetura quanto o processo de treinamento. A arquitetura resultante utiliza a HRNet-W32 como modelo base, incorporando um estágio extra com dois blocos residuais invertidos e perda composta por média ponderada. No entanto, foi empregada a cabeça de detecção da *Cascade R-CNN*, que pode não ser a escolha mais adequada. Para abordar essa questão, a próxima seção se concentrará na otimização da cabeça de detecção mais eficiente.

5.2.4 Otimização da Cabeça de Detecção

Nesta seção, são apresentados os experimentos realizados com as cabeças de detecção aplicadas ao *backbone* previamente obtido (Tabela 3). Para tanto, o *backbone* foi treinado utilizando as cabeças de detecção dos detectores TOOD, *RetinaNet* e *Faster R-CNN*. Conforme evidenciado na Tabela 5, a configuração que incorporou a cabeça da *Cascade R-CNN* demonstrou desempenho superior.

Tabela 5 – Resultados da Otimização de Cabeças de Detecção.

Cabeça de Detecção	mAP Validação
Cascade R-CNN	0,933
Faster R-CNN	0,922

Observou-se que detectores de estágio único, TOOD e RetinaNet, apresentaram dificuldades durante o treinamento com o *backbone*, pois a dissipação do gradiente ocorreu

antes do término da primeira época. Em contrapartida, modelos de dois estágios permitiram um treinamento eficaz ao empregar uma função de perda composta por média ponderada. Portanto, o modelo final que apresentou os melhores resultados utiliza o *backbone* HRNet-W32 como base, acrescido de um estágio extra e de dois blocos residuais inverso adicionais em cada estágio, treinado com uma função de perda composta por média ponderada.

5.3 *Fine Tuning* de Modelos Genéricos

A etapa subsequente consistiu no treinamento de modelos genéricos de detecção, previamente ajustados para o conjunto de dados de ovos parasitários. Os modelos estabelecidos na Seção 4.3 foram treinados com arquitetura *ResNet-101*, utilizando os seguintes hiperparâmetros: tamanho do lote (*batch size*) de 4, taxa de aprendizagem de 0,01, otimizador SGD e imagens com resolução de 1536×1536 pixels.

Tabela 6 – Resultados de Modelos Genéricos da Literatura.

Modelo	mAP Validação
Cascade R-CNN	0,935
TOOD	0,938
CenterNet	0,903

Conforme demonstrado na Tabela 6, a rede TOOD obteve o melhor desempenho entre os modelos avaliados da literatura no conjunto de validação, destacando-se no contexto específico da detecção de ovos parasitários.

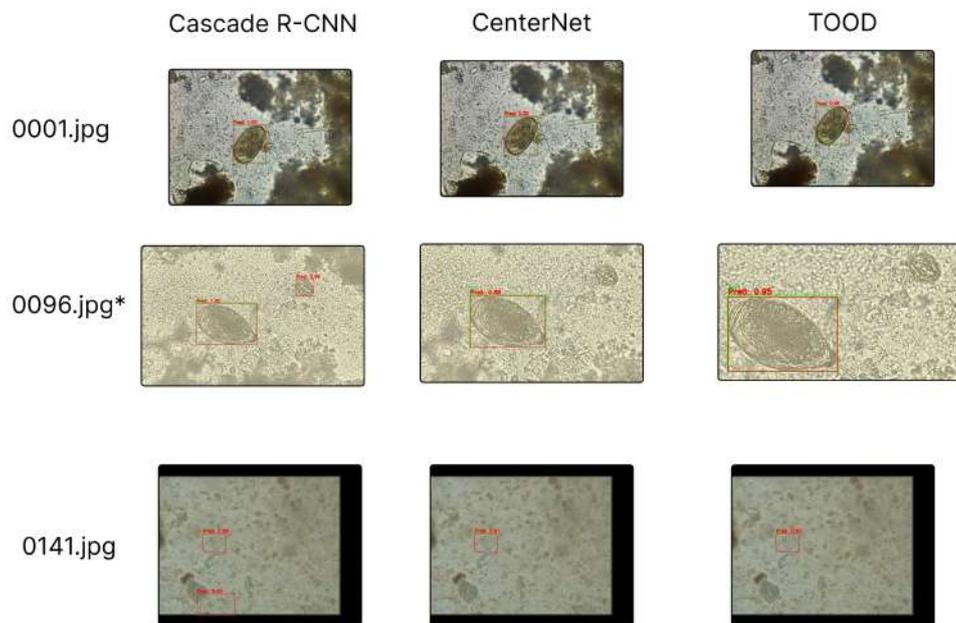
Como pode ser visto na Figura 32, a rede *Cascade R-CNN* apresenta um comportamento de maior ajuste aos dados de treinamento, pois durante o teste apresenta falsos positivos sobre artefatos que se assemelham a ovos de parasito. Outro fator que indica sobre ajuste é a confiança do modelo, no caso da imagem 0096.jpg a *Cascade* obteve um falso positivo com confiança de 0,99, o que indica que o modelo possui certeza de que o artefato é um ovo de parasito. Os demais modelos obtêm menor confiança, porém menos falsos positivos, indicando uma melhor generalização do problema.

5.4 Construção do Ensemble

Nesta etapa, realizou-se o ensemble dos modelos genéricos *Cascade R-CNN*, *CenterNet* e TOOD com a Arquitetura Específica. Utilizaram-se pesos iguais para os 4 modelos, um limite de IoU de 0,5 e limite de confiança de 0,25 para o ensemble.

Conforme evidenciado na Tabela 7, o método de ensemble superou todos os modelos individuais em 5 das 7 métricas avaliadas, destacando-se como a abordagem mais eficaz. Esse resultado comprova a capacidade do ensemble de reduzir os erros individuais e

Figura 32 – Plotagem dos Ovos no Conjunto de Teste em Verde e das Detecções em Vermelho.



* Recorte de Imagem maior

potencializar os acertos coletivos, reforçando sua superioridade em relação ao uso de modelos isolados.

Tabela 7 – Resultado dos modelos Individuais e Ensemble no Teste.

Modelo	mIoU	mAP[0.5:0.95]	mAP[0.5]	Precisão	Sensibilidade	F1 Score	Acurácia
Cascade R-CNN	0,918	0,858	0,917	0,942	0,915	0,927	0,88
CenterNet	0,912	0,807	0,884	0,961	0,882	0,912	0,86
TOOD	0,926	0,866	0,931	0,950	0,930	0,939	0,90
Arquitetura Específica	0,919	0,85	0,923	0,931	0,923	0,926	0,89
Ensemble	0,931	0,868	0,924	0,972	0,921	0,945	0,91

Conforme observado na Tabela 8 a Arquitetura específica obteve maior sensibilidade na maioria das classes. Isto se deve a sua construção ser focada em características que permitam a melhor localização dos ovos de parasito. O Ensemble foi capaz de superar os modelos individuais em precisão em 10 das 11 classes, falhando somente em superar a *CenterNet* na classe *Capillaria P.*

5.5 Discussão

Analizamos nessa seção os resultados obtidos através de estudo de caso e comparação com trabalhos relacionados. A classe *Ascaris l.* é uma das mais difíceis de serem detectadas, somente a arquitetura específica foi capaz de alcançar uma alta sensibilidade nela, em

Tabela 8 – Resultados por Classe dos Modelos Individuais e Ensemble (Conjunto de Teste)

Classe/Métricas	Modelos				Ensemble
	Cascade R-CNN	TOOD	CenterNet	Específica	
Ascaris l.					
Precisão	0,950	0,955	0,950	0,764	0,977
Sensibilidade	0,840	0,826	0,748	0,956	0,821
Capillaria p.					
Precisão	0,898	0,959	0,988	0,690	0,974
Sensibilidade	0,876	0,931	0,881	0,965	0,931
Enterobius v.					
Precisão	0,978	0,963	0,982	0,913	1,0
Sensibilidade	0,895	0,925	0,860	0,950	0,905
Fasciolopsis b.					
Precisão	0,989	0,984	0,983	0,894	1,0
Sensibilidade	0,949	0,979	0,889	0,979	0,954
Hookworm egg					
Precisão	0,980	0,985	0,979	0,980	0,990
Sensibilidade	0,970	0,980	0,955	0,975	0,975
Hymenolepis d.					
Precisão	0,884	0,854	0,934	0,787	0,938
Sensibilidade	1,0	1,0	0,995	1,0	1,0
Hymenolepis n.					
Precisão	0,925	0,912	0,916	0,810	0,930
Sensibilidade	0,861	0,826	0,866	0,930	0,866
Opisthorchis v.					
Precisão	0,912	0,989	0,966	0,945	0,989
Sensibilidade	0,940	0,955	0,860	0,96	0,93
Paragonimus spp					
Precisão	0,869	0,880	0,904	0,773	0,914
Sensibilidade	0,923	0,913	0,861	0,947	0,923
Taenia spp egg					
Precisão	0,98	0,984	0,979	0,970	0,984
Sensibilidade	0,960	0,955	0,921	0,960	0,950
Trichuris t.					
Precisão	1,0	0,989	0,994	1,0	1,0
Sensibilidade	0,855	0,945	0,865	0,955	0,88

contrapartida ficou com baixa precisão. Isto se da, devido as diversas variações de formato dos ovos de *Ascaris l.* (Seção 2.1.1. Por outro lado, a classe *Hymenolepis d.* foi a mais fácil de ser detectada, com a maioria dos modelos obtendo 100% de sensibilidade, embora não acompanhado por uma alta precisão. O que indica que esta classe é facilmente identificada como um ovo de parasito, porém nem sempre classificada corretamente. A fácil identificação se dá pelo formato circular bem definido e texturas internas bem marcas, como pode ser

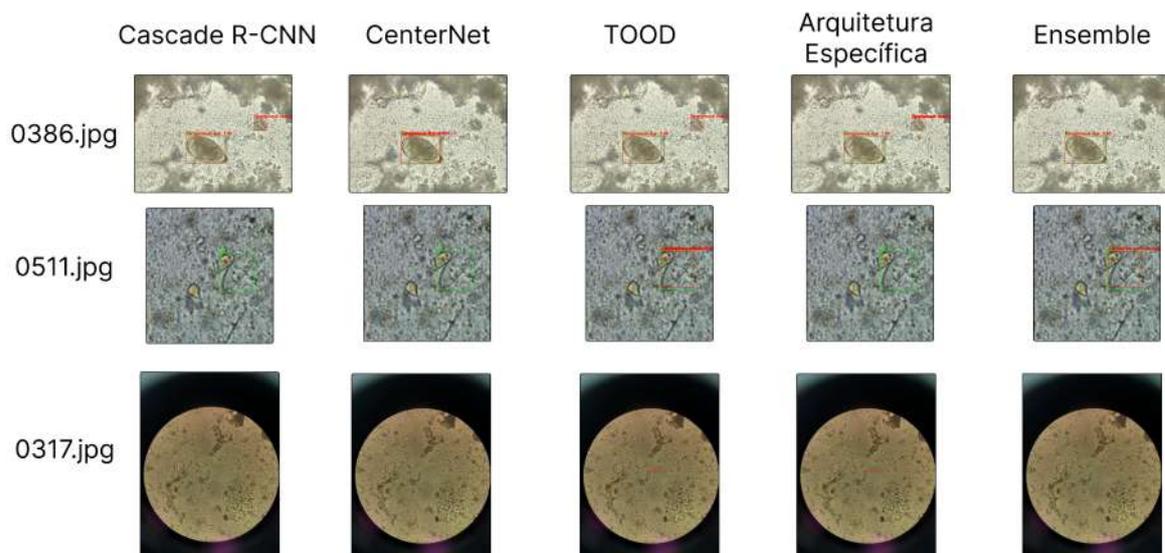
Figura 33 – Amostras de *Hymenolepis d.*



visto na Figura 33.

Conforme ilustrado na Figura 34, na amostra 0386.jpg do conjunto de teste, três dos modelos analisados apresentaram um falso positivo. A exceção foi a *CenterNet*, que, apesar de ser o modelo com o menor desempenho geral conforme a Tabela 7, não gerou detecções incorretas nesse caso. O método de ensemble, por sua vez, demonstrou eficácia ao reduzir o score da detecção falsa abaixo do limiar de confiança de 0,25, descartando-a corretamente. Esse resultado evidencia a capacidade do ensemble em mitigar falsos positivos, reforçando sua robustez em comparação aos modelos individuais.

Figura 34 – Estudo de Caso entre *Ground Truth* (Verde) e Predições (Vermelho) em Detecções do Conjunto de Teste.



Na amostra 0511.jpg, observa-se um caso em que todos os detectores apresentaram um falso negativo, exceto o TOOD, que identificou duas detecções para o ovo de parasito.

Durante o processo de ensemble, a detecção com menor confiança foi descartada, mantendo-se somente a de maior confiança. Esse comportamento demonstra a capacidade do ensemble de reduzir falsos negativos, aprimorando a precisão das detecções.

A arquitetura específica destaca-se por sua estrutura aprofundada de alta resolução, projetada para melhorar a detecção de ovos de parasito de menores dimensões. Esse efeito é evidenciado na amostra 0317.jpg, na qual somente os detectores TOOD e a Arquitetura Específica conseguiram identificar o ovo, com a Arquitetura Específica alcançando um nível de confiança superior ao do TOOD. Esse caso evidência à questão de pesquisa Q4 (Seção 1.2.1), confirmando que o uso de características de alta resolução aumenta a sensibilidade na detecção de ovos menores.

A análise desses casos, em conjunto com as métricas obtidas, responde à questão Q3 (1.2.1), comprovando a eficácia da utilização de um ensemble que combina múltiplos modelos de detecção. Essa abordagem não somente melhora a robustez da arquitetura, mas também supera as limitações individuais de cada modelo, garantindo maior precisão e confiabilidade nos resultados.

Em comparação com os trabalhos da literatura (Tabela 9), o ensemble proposto neste trabalho superou o desempenho de diversas metodologias. No entanto, em algumas métricas, os resultados obtidos permanecem inferiores a determinados trabalhos, como o de Aung, Srithaworn e Achakulvisut (2022), que emprega um ciclo de re-treinamento com *pseudo-labels* para otimizar um ensemble de modelos diretamente no conjunto de teste.

Tabela 9 – Comparação com Trabalhos Relacionados.

Trabalho	mIoU	mAP[0.5:0.95]	mAP[0.5]	Precisão	Sensibilidade	F1 Score	Acurácia
Pedraza et al. (2022)	0,875	-	-	-	-	0,955	-
Tureckova, Turecek e Oplatkova (2022)	0,887	-	-	-	-	0,964	-
Ruiz-Santaquiteria et al. (2022)	0,915	-	-	-	-	0,974	-
Pratama et al. (2022)	0,717	-	-	-	-	0,864	-
Aung, Srithaworn e Achakulvisut (2022)	0,933	-	-	-	-	0,988	-
Pho et al. (2022)	-	0,82	-	0,99	0,87	0,926	-
Ensemble	0,931	0,868	0,924	0,972	0,921	0,945	0,91

Em termos de contribuição, este trabalho apresenta avanços para a área de pesquisa em comparação com a literatura. Primeiramente, pode-se destacar a ampla comparação utilizando a métrica do mAP, sendo uma métrica própria para detecção de objetos. Além da utilização de métricas específicas de sensibilidade e precisão, que permitem uma melhor análise das detecções do modelo. Em contraste, os trabalhos relacionados divulgam o *F1 Score* e mIoU, que respectivamente são métricas de classificação e segmentação. Portanto, não conseguimos avaliar precisamente o desempenho em relação as detecções.

Em sequência, os trabalhos da literatura não utilizam conjunto de validação. Assim, treinando os modelos em todo o conjunto de teste. As métricas sobre o conjunto de teste foram obtidas via submissões. O que pode ser especialmente útil para otimizar um modelo para vencer um desafio, é prejudicial na construção de um modelo genérico que é capaz de

performar em um cenário real. Pois, ao não utilizar um conjunto de validação, estamos transformando o conjunto de teste em validação, portanto ficando sem um conjunto de teste.

Finalmente, a utilização de uma rede de alta resolução otimizada para o problema demonstrou a capacidade de detectar objetos menores que não conseguiram ser detectados pelos modelos genéricos frequentemente empregados na literatura. O que mostra que explorar características de alta resolução neste problema pode levar a resultados melhores para ovos pequenos que, como visto na Seção 5.1 representam metade de cada conjunto de treino, validação e teste. Além de que ovos pequenos são mais difíceis de serem detectados (Figura 34).

A metodologia empregada neste trabalho permitiu otimizar uma arquitetura de alta resolução e refinar modelos genéricos e posteriormente combiná-los em um ensemble. Este produziu resultados melhores que qualquer modelo individual. Porém, ainda existem possibilidades de melhorias a serem exploradas em trabalhos futuros. Estas futuras melhorias serão discutidas na seção seguinte.

6 Conclusão

As características morfológicas de um mesmo ovo de parasito podem variar conforme seu estágio de vida ou estado de fertilidade, dificultando a distinção entre espécies com base na análise de ovos. Adicionalmente, a presença de ovos de dimensões reduzidas e a similaridade visual com artefatos representam desafios adicionais para a identificação precisa. Neste estudo, propõe-se uma abordagem abrangente baseada em ensemble de múltiplos detectores, visando ampliar a capacidade de detecção e aumentar a sensibilidade na análise de imagens microscópicas.

Este trabalho desenvolveu uma metodologia eficiente para problemas de detecção em larga escala, composta por três etapas principais: (i) otimização de *backbones* baseados na HRNet-W32 para construção de uma arquitetura especializada em alta resolução, (ii) refinamento de modelos de detecção genéricos no *dataset*, e (iii) integração da arquitetura otimizada com os modelos refinados em uma arquitetura de ensemble.

No Capítulo 5, cada questão de pesquisa foi abordada por meio de uma configuração experimental específica. A questão Q1 (Seção 1.2.1) foi investigada na Seção 5.2, com a otimização de estágios da arquitetura. Os resultados demonstraram que o aumento do número de canais na arquitetura base acarreta problemas como a dissipação de gradiente. Por outro lado, a adição de estágios e a incorporação de perdas intermediárias na composição da perda total reduziram a dissipação de gradiente e permitiram aprofundar a rede sem comprometer seu desempenho, o que responde à questão Q2 (Seção 1.2.1). A combinação de detectores com diferentes níveis de generalização mostrou-se eficaz para reduzir falsos negativos e aumentar a sensibilidade do sistema. A *Cascade R-CNN*, embora produza detecções de alta confiança, pode gerar falsos positivos com elevada certeza. Esses erros são mitigados pelo uso de ensemble, que equilibra as previsões dos modelos, aproveitando suas vantagens complementares, respondendo assim à questão Q3 (Seção 1.2.1). Durante o ensemble, notou-se que a arquitetura específica possui maior facilidade para detectar ovos menores, indicando a eficácia das características de alta resolução para detecção de objetos menores, assim respondendo à questão Q4 (Seção 1.2.1).

Embora este trabalho tenha explorado lacunas e otimizações para o problema de detecção de ovos de parasito em imagens microscópicas, sua proposta final permite melhorias que serão endereçadas em trabalhos futuros.

Enumeram-se as seguintes abordagens para trabalhos futuros:

- Utilizar *Generative Adversarial Networks* (GAN) para remover as setas vermelhas de indicação da posição do ovo de parasito;

- Aplicação de *Knowledge Distillation* (KD) para reduzir o custo computacional associado à utilização de ensembles de modelos, viabilizando sua implementação em ambientes de produção;
- Produzir máscaras de segmentação para os ovos de parasito para treinar um ramo de segmentação na *Cascade R-CNN*;
- Utilizar técnicas de aumento de dados sintéticos recortando segmentações dos ovos de parasito para inserir em uma nova imagem de microscópio, possibilitando assim uma maior variação de ovos de parasito por imagem;

Apesar destas limitações, este trabalho contribuiu com um estudo de modificações na arquitetura da HRNet, como também fornecer um método eficaz de ensemble.

Referências

- ALVES, N. N. N.; MAGALHÃES, T. N.; SANTOS, Y. R. M. dos; RAMOS, R. A.; MESQUITA, E. A. de. Prevalência de enteroparasitos em espaço público destinado o lazer, esporte e turismo em porto velho, rondônia, brasil. *Natureza Online*, v. 21, n. 1, 2023. Disponível em: <<https://www.naturezaonline.com.br/revista/article/download/485/466>>. Citado na página 15.
- ANANTRASIRICHAI, N.; CHALIDABHONGSE, T. H.; PALASUWAN, D.; NARUENATTHANASET, K.; KOBCHAISAWAT, T.; NUNTHANASUP, N.; BOONPENG, K.; MA, X.; ACHIM, A. Icip 2022 challenge on parasitic egg detection and classification in microscopic images: Dataset, methods and results. In: IEEE. *2022 IEEE International Conference on Image Processing (ICIP)*. 2022. p. 4306–4310. Disponível em: <[10.1109/ICIP46576.2022.9897267](https://doi.org/10.1109/ICIP46576.2022.9897267)>. Citado 7 vezes nas páginas 17, 18, 28, 29, 48, 52 e 53.
- AUNG, Z. H.; SRITHAWORN, K.; ACHAKULVISUT, T. Multitask learning via pseudo-label generation and ensemble prediction for parasitic egg cell detection: Ieee icip challenge 2022. In: IEEE. *2022 IEEE International Conference on Image Processing (ICIP)*. 2022. p. 4273–4277. Disponível em: <[10.1109/ICIP46576.2022.9897464](https://doi.org/10.1109/ICIP46576.2022.9897464)>. Citado 4 vezes nas páginas 18, 49, 50 e 70.
- BELO, V. S.; OLIVEIRA, R. B. d.; FERNANDES, P. C.; NASCIMENTO, B. W. L.; FERNANDES, F. V.; CASTRO, C. L. F.; SANTOS, W. B. d.; SILVA, E. S. d. Fatores associados à ocorrência de parasitoses intestinais em uma população de crianças e adolescentes. *Revista Paulista de Pediatria*, SciELO Brasil, v. 30, p. 195–201, 2012. Disponível em: <<https://www.scielo.br/j/rpp/a/3KC98Fgr6WvZDnNQZrxwN5w/?format=pdf&lang=pt>>. Citado na página 15.
- BESSA, M. L.; JUNIOR, G. B.; de Almeida, J. D. S. Neural network ensemble for detecting parasite eggs in microscopic images. *Procedia Computer Science*, v. 256, p. 739–746, 2025. ISSN 1877-0509. CENTERIS - International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1877050925005319>>. Citado na página 20.
- BRASIL; SAÚDE, M. da; SAÚDE, S. de Vigilância em; TRANSMISSÍVEIS, D. de Vigilância das D. *Guia prático para o controle das geo-helminthiases*. Ministério da Saúde Brasília, 2018. Disponível em: <https://bvsmms.saude.gov.br/bvs/publicacoes/guia_pratico_controle_geohelminthiases.pdf>. Citado na página 16.
- CAI, Z.; VASCONCELOS, N. Cascade r-cnn: High quality object detection and instance segmentation. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 43, n. 5, p. 1483–1498, 2019. Citado 2 vezes nas páginas 40 e 41.
- CDC. *Intestinal Capillariasis*. 2020. Laboratory Identification of Parasites of Public Health Concern. Acessado em: 05-12-2024. Disponível em: <<https://www.cdc.gov/dpdx/intestinalcapillariasis/>>. Citado na página 26.

CDC. *Diagnosis of Parasitic Diseases*. 2024. Centers for Disease Control and Prevention Site. Acessado em : 14-02-2024. Disponível em: <<https://www.cdc.gov/parasites/testing-diagnosis/index.html>>. Citado na página 15.

DOLABELLA, S. S.; BARBOSA, L. Fundamentos de parasitologia. *CESAD*, 2011. Citado na página 22.

EVERINGHAM, M.; GOOL, L. V.; WILLIAMS, C. K. I.; WINN, J.; ZISSERMAN, A. *The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results*. [Http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html](http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html). Citado na página 43.

FERREIRA, C. S.; CARVALHO, M. E. d. Diafanização de esfregaços de fezes. *Revista de Saúde Pública*, Faculdade de Saúde Pública da Universidade de São Paulo, v. 6, n. 1, p. 19–23, Mar 1972. ISSN 0034-8910. Disponível em: <<https://doi.org/10.1590/S0034-89101972000100003>>. Citado na página 17.

FOCACCIA, V. *Tratado de Infectologia*. [S.l.]: Atheneu, 2005. Citado 5 vezes nas páginas 15, 23, 24, 25 e 26.

GIRSHICK, R.; DONAHUE, J.; DARRELL, T.; MALIK, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2014. p. 580–587. Citado na página 36.

GOMES, C. V. F. V. da S. *Guia para pesquisa de ovos de helmintos em fezes através do método Kato-Katz*. Secretaria de Saúde do Ceará, 2022. Disponível em: <https://www.saude.ce.gov.br/wp-content/uploads/sites/9/2022/02/guia_pesquisa_ovos_helmintos_fezes_metodo_KatoKatz_20220402.pdf>. Citado 2 vezes nas páginas 17 e 28.

HABIB, A.; ANDRIANONIMIADANA, L.; RAKOTONDRAINIPIANA, M.; ANDRIANTSALAMA, P.; RANDRIAMPARANY, R.; RANDREMANANA, R. V.; RAKOTOARISON, R.; VIGAN-WOMAS, I.; RAFALIMANANTSOA, A.; VONAESCH, P. et al. High prevalence of intestinal parasite infestations among stunted and control children aged 2 to 5 years old in two neighborhoods of antananarivo, madagascar. *PLoS neglected tropical diseases*, Public Library of Science San Francisco, CA USA, v. 15, n. 4, p. e0009333, 2021. Disponível em: <<https://journals.plos.org/plosntds/article/file?id=10.1371/journal.pntd.0009333&type=printable>>. Citado na página 15.

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. Citado 2 vezes nas páginas 54 e 62.

HOTEZ, P. J. *Forgotten people, forgotten diseases: the neglected tropical diseases and their impact on global health and development*. John Wiley & Sons, 2021. Disponível em: <https://www.researchgate.net/profile/Christopher-Rensing-2/publication/304588160_Pseudomonas_Genomics_and_Molecular_Biology/links/592c3257458515e3d4715f05/Pseudomonas-Genomics-and-Molecular-Biology.pdf>. Citado na página 15.

HOWARD, A. G.; ZHU, M.; CHEN, B.; KALENICHENKO, D.; WANG, W.; WEYAND, T.; ANDREETTO, M.; ADAM, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017. Citado na página 54.

- HUI, J. *mAP (mean Average Precision) for Object Detection*. 2018. <<https://jonathan-hui.medium.com/map-mean-average-precision-for-object-detection-45c121a31173>>. Accessed: 01/02/2025. Citado 2 vezes nas páginas 43 e 44.
- KATZ, N.; CHAVES, A.; PELLEGRINO, J. A simple, device for quantitative stool thick-smear technique in schistosomiasis mansoni. *Revista do instituto de medicina tropical de São Paulo*, v. 14, n. 6, p. 397–400, 1972. Citado na página 16.
- KUMAR, S.; ARIF, T.; ALOTAIBI, A. S.; MALIK, M. B.; MANHAS, J. Advances towards automatic detection and classification of parasites microscopic images using deep convolutional neural network: methods, models and research directions. *Archives of Computational Methods in Engineering*, Springer, v. 30, n. 3, p. 2013–2039, 2023. Disponível em: <<https://link.springer.com/article/10.1007/s11831-022-09858-w>>. Citado na página 17.
- LEE, C.-C.; HUANG, P.-J.; YEH, Y.-M.; LI, P.-H.; CHIU, C.-H.; CHENG, W.-H.; TANG, P. Helminth egg analysis platform (heap): An opened platform for microscopic helminth egg identification and quantification based on the integration of deep learning architectures. *Journal of Microbiology, Immunology and Infection*, Elsevier, v. 55, n. 3, p. 395–404, 2022. Citado 2 vezes nas páginas 49 e 50.
- 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. Citado na página 40.
- 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: SPRINGER. *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part V 13*. [S.l.], 2014. p. 740–755. Citado na página 43.
- 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. Citado 3 vezes nas páginas 36, 37 e 39.
- MATHISON, B. A.; KOHAN, J. L.; WALKER, J. F.; SMITH, R. B.; ARDON, O.; COUTURIER, M. R. Detection of intestinal protozoa in trichrome-stained stool specimens by use of a deep convolutional neural network. *Journal of clinical microbiology*, Am Soc Microbiol, v. 58, n. 6, p. 10–1128, 2020. Citado 2 vezes nas páginas 48 e 50.
- MESQUITA, C. A. M.; TORQUETE, M. A.; BARÇANTE, J. M. P.; BARÇANTE, T. A.; GUIMARÃES, A. M.; ROCHA, C. Prevalência de infecções por geo-helminthos em comunidades brasileiras: A systematic review. *Spei Domus*, v. 17, n. 2, p. 1–18, 2021. Disponível em: <<https://revistas.ucc.edu.co/index.php/sp/article/download/3931/3275>>. Citado na página 15.
- MICIKEVICIUS, P.; NARANG, S.; ALBEN, J.; DIAMOS, G.; ELSEN, E.; GARCIA, D.; GINSBURG, B.; HOUSTON, M.; KUCHARIEV, O.; VENKATESH, G. et al. Mixed precision training. *arXiv preprint arXiv:1710.03740*, 2017. Disponível em: <<https://arxiv.org/pdf/1710.03740>>. Citado na página 62.

OMS. *Soil-transmitted helminth infections*. 2023. World Health Organization Newsroom. Acessado em : 14-02-2024. Disponível em: <<https://www.who.int/news-room/fact-sheets/detail/soil-transmitted-helminth-infections>>. Citado na página 15.

ORGANIZATION, W. H. et al. *High burden to high impact: a targeted malaria response*. [S.l.], 2018. Disponível em: <<https://www.who.int/publications/i/item/WHO-CDS-GMP-2018.25>>. Citado na página 15.

PAPIER, K.; WILLIAMS, G. M.; LUCERES-CATUBIG, R.; AHMED, F.; OLVEDA, R. M.; MCMANUS, D. P.; CHY, D.; CHAU, T. N.; GRAY, D. J.; ROSS, A. G. Childhood malnutrition and parasitic helminth interactions. *Clinical Infectious Diseases*, Oxford University Press, v. 59, n. 2, p. 234–243, 2014. Disponível em: <<https://doi.org/10.1093/cid/ciu211>>. Citado na página 16.

PEDRAZA, A.; RUIZ-SANTAQUITERIA, J.; DENIZ, O.; BUENO, G. Parasitic egg detection and classification with transformer-based architectures. In: IEEE. *2022 IEEE International Conference on Image Processing (ICIP)*. 2022. p. 4301–4305. Disponível em: <[10.1109/ICIP46576.2022.9897846](https://doi.org/10.1109/ICIP46576.2022.9897846)>. Citado 4 vezes nas páginas 19, 48, 50 e 70.

PHO, K.; AMIN, M. K. M.; YOSHITAKA, A. Segmentation-driven hierarchical retinanet for detecting protozoa in micrograph. *International Journal of Semantic Computing*, World Scientific, v. 13, n. 03, p. 393–413, 2019. Disponível em: <<https://doi.org/10.1142/S1793351X19400178>>. Citado 2 vezes nas páginas 47 e 50.

PHO, K.; LAM, H.; LE, T.; NGUYEN, H. T.; YOSHITAKA, A. Attention-driven retinanet for parasitic egg detection. In: IEEE. *2022 IEEE International Symposium on Multimedia (ISM)*. 2022. p. 265–272. Disponível em: <[10.1109/ISM55400.2022.00060](https://doi.org/10.1109/ISM55400.2022.00060)>. Citado 3 vezes nas páginas 49, 50 e 70.

PINTO, E. C. G. e. M. M. I. I. Carlos José de C. *Tratado de Infectologia*. [S.l.]: Universidade Federal de Santa Catarina, 2011. Citado 8 vezes nas páginas 15, 16, 19, 22, 24, 25, 27 e 28.

PRADO, M. d. S.; BARRETO, M. L.; STRINA, A.; FARIA, J. A. S.; NOBRE, A. A.; JESUS, S. R. Prevalência e intensidade da infecção por parasitas intestinais em crianças na idade escolar na cidade de salvador (bahia, brasil). *Revista da Sociedade Brasileira de Medicina Tropical*, SciELO Brasil, v. 34, p. 99–101, 2001. Disponível em: <<https://doi.org/10.1590/S0037-86822001000100016>>. Citado na página 15.

PRATAMA, Y.; FUJIMURA, Y.; FUNATOMI, T.; MUKAIGAWA, Y. Parasitic egg detection and classification by utilizing the yolo algorithm with deep latent space image restoration and grabcut augmentation. In: IEEE. *2022 IEEE International Conference on Image Processing (ICIP)*. [S.l.], 2022. p. 4311–4315. Citado 3 vezes nas páginas 49, 50 e 70.

REDMON, J. You only look once: Unified, real-time object detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. Citado na página 39.

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. Citado 2 vezes nas páginas 37 e 38.

- 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. Citado na página 40.
- RUIZ-SANTAQUITERIA, J.; PEDRAZA, A.; VALLEZ, N.; VELASCO, A. Parasitic egg detection with a deep learning ensemble. In: IEEE. *2022 IEEE International Conference on Image Processing (ICIP)*. 2022. p. 4283–4286. Disponível em: <[10.1109/ICIP46576.2022.9897858](https://doi.org/10.1109/ICIP46576.2022.9897858)>. Citado 4 vezes nas páginas 19, 49, 50 e 70.
- (SEVIR), S. E. de Vigilância em Saúde e R.; (COVAT), C. da Vigilância Ambiental e Saúde do Trabalhador e da T.; (CEVET), C. de Vigilância Entomológica e Controle de V. *Guia para pesquisa de ovos de helmintos em fezes através do método Kato-Katz*. Secretaria Executiva de Vigilância em Saúde e Regulação, 2022. Acessado em 20 de agosto de 2024. Disponível em: <https://www.saude.ce.gov.br/wp-content/uploads/sites/9/2022/02/guia_pesquisa_ovos_helmintos_fezes_metodo_KatoKatz_20220402.pdf>. Citado 2 vezes nas páginas 17 e 27.
- SILVA, M. B. M. *A Ocorrência de Parasitoses Intestinais em alunos do ensino fundamental na zona rural rede Municipal de São Luís-Maranhão e sua relação com os hábitos alimentares*. Dissertação (Mestrado) — Escola Superior de Tecnologia da Saúde de Coimbra, 2018. Disponível em: <<https://comun.rcaap.pt/entities/publication/f50ddd42-0ca3-492d-a955-cdfed7d89d83>>. Citado na página 16.
- SOLOVYEV, R.; WANG, W.; GABRUSEVA, T. Weighted boxes fusion: Ensembling boxes from different object detection models. *Image and Vision Computing*, Elsevier, v. 107, p. 104117, 2021. Citado na página 56.
- SUN, K.; XIAO, B.; LIU, D.; WANG, J. Deep high-resolution representation learning for human pose estimation. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. [S.l.: s.n.], 2019. p. 5693–5703. Citado na página 41.
- SUWANNAPHONG, T.; CHAVANA, S.; TONGSOM, S.; PALASUWAN, D.; CHALIDABHONGSE, T. H.; ANANTRASIRICHA, N. Parasitic egg detection and classification in low-cost microscopic images using transfer learning. *SN Computer Science*, Springer, v. 5, n. 1, p. 82, 2023. Citado 2 vezes nas páginas 27 e 50.
- TURECKOVA, A.; TURECEK, T.; OPLATKOVA, Z. K. Icip 2022 challenge: Pedcmi, tood enhanced by slicing-aided fine-tuning and inference. In: IEEE. *2022 IEEE International Conference on Image Processing (ICIP)*. 2022. p. 4292–4295. Disponível em: <[10.1109/ICIP46576.2022.9897826](https://doi.org/10.1109/ICIP46576.2022.9897826)>. Citado 4 vezes nas páginas 48, 50, 56 e 70.
- UIJLINGS, J. R.; SANDE, K. E. V. D.; GEVERS, T.; SMEULDERS, A. W. Selective search for object recognition. *International journal of computer vision*, Springer, v. 104, p. 154–171, 2013. Citado na página 37.
- VIET, N. Q.; THANHTUYEN, D. T.; HOANG, T. H. Parasite worm egg automatic detection in microscopy stool image based on faster r-cnn. In: *Proceedings of the 3rd international conference on machine learning and soft computing*. [S.l.: s.n.], 2019. p. 197–202. Citado 2 vezes nas páginas 47 e 50.
- XIE, S.; GIRSHICK, R.; DOLLÁR, P.; TU, Z.; HE, K. Aggregated residual transformations for deep neural networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2017. p. 1492–1500. Citado 3 vezes nas páginas 42, 54 e 62.

YANG, F.; YU, H.; SILAMUT, K.; MAUDE, R. J.; JAEGER, S.; ANTANI, S. Parasite detection in thick blood smears based on customized faster-rcnn on smartphones. In: IEEE. *2019 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*. [S.l.], 2019. p. 1–4. Citado 2 vezes nas páginas [47](#) e [50](#).