

JESUS DOURADO DE ALBUQUERQUE

**UMA APLICAÇÃO DE REDES NEURAIIS CONVOLUCIONAIS
REGIONAIS PARA DETECÇÃO DE MALÁRIA**

Trabalho de Conclusão de Curso apresentado
à banca avaliadora do Curso de Engenharia
de Computação, da Escola Superior de
Tecnologia, da Universidade do Estado do
Amazonas, como pré-requisito para obtenção
do título de Engenheiro de Computação.

Orientador(a): Prof. Dra. Elloá Barreto Guedes da Costa

Manaus – Março – 2023

Ficha Catalográfica

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).
Sistema Integrado de Bibliotecas da Universidade do Estado do Amazonas.

J58ua Albuquerque, Jesus Dourado de
 Uma Aplicação de Redes Neurais Convolucionais
 Regionais para Detecção de Malária / Jesus Dourado de
 Albuquerque. Manaus : [s.n], 2023.
 42 f.: color.; 29 cm.

 TCC - Graduação em Engenharia de Computação -
 Universidade do Estado do Amazonas, Manaus, 2023.
 Inclui bibliografia
 Orientador: Elloá Barreto Guedes da Costa

 1. Deep Learning. 2. YOLO. 3. Malária. I. Elloá
 Barreto Guedes da Costa (Orient.). II. Universidade do
 Estado do Amazonas. III. Uma Aplicação de Redes Neurais
 Convolucionais Regionais para Detecção de Malária

Elaborado por Jeane Macelino Galves - CRB-11/463



FOLHA DE APROVAÇÃO

Uma Aplicação de Redes Neurais Convolucionais Regionais para Detecção de Malária

Jesus Dourado de Albuquerque

Trabalho de Conclusão de Curso defendido e aprovado pela banca avaliadora constituída pelos professores:

Profª Drª Elloá Barreto Guedes da Costa
Presidente

Prof. Dr. Sergio Cleger Tamayo
Avaliador(a)

Profª Drª Áurea Hiléia da Silva Melo
Avaliador(a)

Manaus, 28 de Março de 2023.

Resumo

Este trabalho aborda a doença da malária no âmbito global, nacional e regional, pontuando as características do modo de transmissão, vetores e sintomas, além de mencionar o diagnóstico precoce utilizando imagens microscópicas para, então, aplicar o tratamento eficaz. Neste contexto, o trabalho apresenta a seguinte solução: por meio de *Deep Learning* aplicada à Visão Computacional, foi desenvolvido um modelo, baseado na arquitetura YOLO, capaz de facilitar o processo de diagnóstico microscópico da malária realizado por profissionais de saúde. Os resultados obtidos foram finalmente avaliados, levando em consideração quais modelos alcançaram melhores métricas de desempenho em relação ao custo computacional de espaço e tempo, mostrando que a solução proposta é viável para o problema postulado.

Palavras-Chave: Deep Learning; YOLO; Malária.

Abstract

This work aims the disease of malaria at the global, national and regional levels, pointing out the characteristics of the mode of transmission, vectors and symptoms, in addition to mentioning the early diagnosis using microscopic images to then apply effective treatment. In this context, the work presents the following solution: through Deep Learning applied to Computer Vision, a model was developed, based on the YOLO architecture, capable of facilitating the process of microscopic diagnosis of malaria carried out by health professionals. The results obtained were finally evaluated, taking into account which models achieved better performance metrics in relation to the computational cost of space and time, showing that the algorithm is viable for the problem raised.

Keywords: *Deep Learning; YOLO; Malaria.*

Sumário

Lista de Tabelas	vi
Lista de Figuras	viii
1 Introdução	1
1.1 Objetivos	3
1.2 Justificativa	4
1.3 Metodologia	4
1.4 Organização do Documento	5
2 Fundamentação Teórica	6
2.1 <i>Deep Learning</i> com CNNs	6
2.2 Detecção de Objetos com Redes Neurais Convolucionais	10
2.3 Família YOLO	13
2.4 Trabalhos Relacionados	16
3 Materiais e Métodos	19
3.1 Dados Experimentais	19
3.2 Modelos, Parâmetros e Hiper-parâmetros	22
3.3 Avaliação de Desempenho	23
4 Resultados e Discussão	24
5 Considerações Finais e Trabalhos Futuros	29

Lista de Tabelas

3.1	Descrição estatística do tamanho das caixas delimitadoras.	22
3.2	Modelos da Família YOLO considerados e suas características.	22
4.1	Síntese dos resultados experimentais.	24

Lista de Figuras

2.1	Modelo abstrato de um neurônio artificial. Fonte: (BRAGA; CARVALHO; LUDERMIR, 2012).	8
2.2	Papel desempenhado pelos neurônios em cada camada de uma RNA MLP. Fonte: (FACELI et al., 2015).	8
2.3	Tarefa de detecção de objetos sendo realizada. Fonte: (REDMON et al., 2015)	10
2.4	Pesquisa seletiva destacando regiões de interesse. Fonte: (MICHELUCCI, 2019)	12
2.5	Abordagem da YOLO para detecção de objetos. Fonte: (REDMON et al., 2016).	13
2.6	Arquitetura da CNN utilizada na primeira versão da YOLO. Fonte: (REDMON et al., 2016).	14
2.7	Desempenho dos modelos YOLOv5 perante o MS COCO. Fonte: (JOCHER et al., 2022).	15
2.8	Comparação entre as últimas versões da YOLO propostas na literatura. O eixo vertical denota a eficiência do modelo e o eixo horizontal denota o tempo de treinamento. Fonte: (LI et al., 2022).	16
3.1	Exemplo com a imagem original, acrescida de centroides e das caixas delimitadoras quadradas.	20
3.2	Histograma da quantidade de caixas delimitadoras por imagens da base de dados.	21
3.3	Mapa de calor da disposição das caixas delimitadoras nos exemplos.	21
4.1	Gráficos de monitoramento de métricas da YOLOv7 X perante os conjuntos de treino e de validação.	26
4.2	Exemplo com poucos protozoários e suas respectivas detecções pela YOLOv7 X .	27

4.3	Exemplo com muitos protozoários e suas respectivas detecções pela YOLOv7 <i>X</i> .	28
-----	---	----

Capítulo 1

Introdução

A malária é uma doença causada pelo protozoário do gênero *Plasmodium* e transmitida pela fêmea do mosquito do gênero *Anopheles*. No Brasil, existem três espécies do *Plasmodium* que transmitem malária para seres humanos: *P. vivax*, *P. malariae* e *P. falciparum*, sendo esta última considerada a mais grave por causar alterações estruturais nos glóbulos vermelhos (GOMES et al., 2011; LOIOLA; SILVA; TAUIL, 2002).

Quanto à enfermidade, de acordo com a Fundação Oswaldo Cruz (FIOCRUZ), trata-se de uma doença infecciosa. Em um paciente portador do protozoário, o seu estado clínico é variável a partir do início da infecção, mas listam-se os seguintes sintomas como os mais comuns: febre alta e intermitente, dores musculares, cefaleia e delírios. Quando o tratamento é negligenciado, a doença pode se agravar e manifestar sintomas mais severos, tais como desorientação, convulsões, vômitos, sonolência ou excitação e, eventualmente, levar o paciente a óbito (FIOCRUZ, 2022).

A malária ainda é considerada um problema grave de saúde no mundo, sendo uma das principais doenças letais em regiões tropicais e subtropicais ao redor do globo. De acordo com o Relatório Mundial de Malária de 2019 produzido pela Organização Mundial de Saúde (OMS), foram registrados 228 milhões de casos e 405 mil mortes em âmbito global, em que a grande maioria ocorreu na África (OMS, 2019).

No Brasil, a malária é uma doença de notificação compulsória e, portanto, todos os casos suspeitos ou confirmados devem ser, obrigatoriamente, notificados às autoridades de saúde (BRASIL, 1976). A série histórica dos casos de malária no Brasil foi iniciada em 1959 e até o

o ano de 2005 possuía forte tendência crescente, vindo a diminuir significativamente até o ano de 2013 (BRASIL, 2013). Entre os anos de 2017 a 2019 há uma queda percentual de 19,3% no número de casos, mas há suspeita de subnotificação levantada pelas autoridades de saúde (BRASIL, 2020). Mais recentemente, os anos de 2020 e 2021 registraram 141 mil e 135 mil casos, respectivamente (BRASIL, 2021). Ressalta-se que, em todo o período observado, a maioria dos casos da doença ocorreu na região Amazônica (Acre, Amapá, Amazonas, Maranhão, Mato Grosso, Pará, Rondônia, Roraima e Tocantins), área endêmica para a doença (BRASIL, 2019).

O diagnóstico da malária é feito atualmente por meio da análise microscópica de lâminas contendo amostra de sangue coletadas pelo método da gota espessa ou ainda por testes rápidos de diagnóstico (RDTs, do inglês *Rapid diagnostic tests*) (OMS, 2019). Segundo um estudo recente, os RDTs possuem uma variação de desempenho de detecção da doença a depender da localização geográfica considerada e da idade do paciente, o que ressalta sua limitação. Apesar disso, os autores recomendam-os para uso em áreas remotas e reforçam a análise microscópica como técnica de referência para o diagnóstico (BERZOSA et al., 2018). A análise de amostras sanguíneas para detecção de malária requer não apenas equipamentos de boa qualidade, mas também habilidade em microscopia, o que é dificultada pela má disposição de treinamento para os profissionais da área, o que enseja diagnósticos incorretos (PAZ; SANTIAGO, 2015).

Uma estratégia para contornar as dificuldades enfrentadas para o diagnóstico da malária é considerar o papel relevante das soluções computacionais baseadas em métodos e técnicas de *Deep Learning* para problemas de Visão Computacional na área de Saúde, tais como em Imagiologia Médica e Microscopia (SHEN; WU; SUK, 2017; LEE et al., 2017; XING et al., 2017). Essa sub-área emergente da Inteligência Artificial baseia-se principalmente no uso de Redes Neurais Convolucionais Profundas (CNNs, do inglês *Convolutional Neural Networks* para extração sucessiva de características hierárquicas em dados de alta dimensionalidade (GOOD-FELLOW; BENGIO; COURVILLE, 2016), permitindo aplicações de classificação, detecção e segmentação em imagens mediante Aprendizado Supervisado a partir de conjuntos massivos de dados (CHOLLET, 2018; KHAN et al., 2018).

É intuitivo, assim, conjecturar acerca do potencial de aplicação de métodos de *Deep Le-*

arning em imagens microscópicas para colaborar no diagnóstico de malária. Esta questão já foi explorada em outros trabalhos da literatura, os quais consideraram a tarefa de classificação binária quanto à presença ou ausência do protozoário em uma dada célula sanguínea, previamente segmentada manualmente ou conforme algum método de Visão Computacional tradicional (GUIMARÃES; GUEDES, 2022; YANG et al., 2020). Embora os resultados obtidos por estes autores denotem bom desempenho das CNNs nesta tarefa, dependem crucialmente de uma etapa de pré-processamento para isolamento das múltiplas células contidas em uma imagem microscópica.

Considerando as limitações observadas na literatura, este trabalho de conclusão visou investigar a viabilidade no uso de métodos de *Deep Learning* para detecção de protozoários causadores de malária em uma dada imagem microscópica fornecida como entrada. A vantagem de uma solução desta natureza é a delimitação e localização de cada parasita na imagem, permitindo a posterior supervisão por um especialista humano; favorecendo a contagem dos protozoários, o que é crucial para determinar a estratégia de tratamento; e colaborando para mitigar os erros de diagnóstico em face das habilidades requeridas pelos microscopistas humanos nesta tarefa.

1.1 Objetivos

O objetivo geral deste trabalho consistiu em avaliar modelos de *Deep Learning* para detecção de malária em imagens microscópicas. Para alcançar este objetivo, foi necessário contemplar as seguintes metas:

1. Identificar e preparar uma base de dados da literatura no domínio da detecção de malária;
2. Elaborar e conduzir um estudo de caso experimental para treinamento e avaliação comparativa das arquiteturas de CNNs para detecção de objetos;
3. Avaliar e analisar os resultados obtidos de maneira qualitativa e quantitativa.

1.2 Justificativa

Mesmo tendo sido descoberta no final do Século XIX, a malária ainda é uma doença endêmica em todo o mundo, com um grande quantitativo de casos e óbitos, em que a principal justificativa para a perda de vidas é o diagnóstico tardio combinado com a falta de acesso aos recursos para tratamento (OMS, 2019). No Brasil, embora o tratamento seja gratuito, mais de 70 % dos casos ocorreram em áreas rurais ou regiões indígenas (BRASIL, 2021), o que se mostra um desafio em um país de dimensões continentais. Em todo o caso, o diagnóstico de referência da malária é feito via análise microscópica, o que depende de profissionais altamente especializados (PAZ; SANTIAGO, 2015).

Colaborar para o desenvolvimento de métodos automáticos e inteligentes para o diagnóstico de malária pode mitigar o fardo dessa doença, antecipando o diagnóstico e, consequentemente, o tratamento. Isto pode ser especialmente crucial para as populações mais vulneráveis e aquelas que residem em regiões remotas, em que não há disponibilidade de serviços de saúde e tampouco infraestrutura de alta qualidade para microscopia. Ademais, é possível também ampliar a quantidade de testes realizados, permitindo um melhor monitoramento de populações residentes em áreas endêmicas. Ressalta-se, entretanto, que no escopo deste trabalho de conclusão de curso não haverá coleta de dados diretamente com sujeitos e que tampouco os modelos propostos e avaliados serão utilizados ou testados em cenários reais. Se viáveis, estes passos podem ocorrer em momentos posteriores no contexto da proposição de outros projetos de pesquisa e com a devida supervisão de Comitê de Ética.

Por fim, deve-se mencionar a importância da realização deste trabalho com vistas a colaborar com as atividades desenvolvidas pelo LSI, uma iniciativa do Grupo de Pesquisas em Sistemas Inteligentes da Escola Superior de Tecnologia (EST) da UEA.

1.3 Metodologia

A metodologia utilizada para o desenvolvimento das atividades presentes neste trabalho, com o intuito de atingir os objetivos especificados, é descrita a seguir:

1. Estudo dos conceitos relativos às CNNs perante tarefas de detecção (R-CNNs);
2. Levantamento do ferramental tecnológico para implementação das R-CNNs;
3. Identificação e análise exploratória de uma base de dados disponível na literatura voltada para a tarefa de Visão Computacional de detecção de malária;
4. Identificação de arquiteturas de R-CNNs do estado da arte para problemas de detecção, especialmente da Família YOLO;
5. Conceber um cenário experimental para avaliação dos modelos de R-CNNs para detecção de malária;
6. Treinamento dos modelos;
7. Aferição das métricas de desempenho no teste dos modelos;
8. Escrita da proposta de Trabalho de Conclusão de Curso;
9. Defesa da proposta de Trabalho de Conclusão de Curso;
10. Escrita do Trabalho de Conclusão de Curso;
11. Defesa do Trabalho de Conclusão de Curso.

1.4 Organização do Documento

Este documento está organizado como segue: os fundamentos teóricos que embasam o trabalho, incluindo conceitos de *Deep Learning* com CNNs e sua aplicação na detecção de objetos encontram-se dispostos no Capítulo 2, incluindo também uma apresentação de contribuições relacionadas ao problema já existentes na literatura. Os materiais e métodos adotados para a solução proposta encontram-se discriminados no Capítulo 3. Os resultados obtidos são apresentados e discutidos no Capítulo 4. Por fim, as considerações finais podem ser consultadas no Capítulo 5.

Capítulo 2

Fundamentação Teórica

Nas seções a seguir encontra-se uma breve apresentação dos conceitos que fundamentaram a realização deste trabalho. Uma visão geral sobre *Deep Learning* com CNNs encontra-se disponível na Seção 2.1; a aplicação de CNNs na tarefa de detecção de objetos é detalhada na Seção 2.2; as características das arquiteturas das R-CNNs da Família YOLO são apresentadas na Seção 2.3; por fim, os trabalhos correlatos da literatura são apresentados e discutidos na Seção 2.4.

2.1 *Deep Learning* com CNNs

O Aprendizado de Máquina (AM) é uma sub-área da Inteligência Artificial que lida com algoritmos que são melhorados de forma implícita a partir de exemplos do domínio considerado (FACELI et al., 2015). A maioria dos problemas abordados com modelos, métodos e técnicas dessa área compreendem tarefas complexas cuja solução analítica por meio dos algoritmos tradicionais é inviável ou até mesmo impraticável (BRINK; RICHARDS; FETHEROLF, 2017).

No contexto de AM, “aprender” é um termo que denota o processo de busca automática pela melhor representação da informação que viabiliza a transformação dos dados de entrada com o intuito de facilitar a resolução de uma dada tarefa considerada, a exemplo de uma classificação binária (CHOLLET, 2018). Para efetuar essa busca automática, entretanto, faz-se necessária a intervenção humana altamente especializada para escolha dos melhores modelos, parâmetros

e hiper-parâmetros, os quais são diretamente relacionados ao sucesso na resolução da tarefa considerada (KHAN et al., 2018).

Com o intuito de tornar o processo de aprendizado mais independente, *Deep Learning* (DL), uma sub-área de AM, têm se destacado em diversos contextos práticos, pois objetiva aprender automaticamente, por meio de diversas camadas hierárquicas e sucessivas, múltiplas representações complexas de dados de entrada multimodais (imagens, sons, vídeo, etc.), aumentando assim o domínio de problemas que podem ser abordados (BUDUMA, 2017). Para o domínio de DL, as CNNs são o modelo de referência para múltiplas tarefas. As mesmas são um tipo recente de Rede Neural Artificial *Feedforward Multilayer Perceptron* (RNA MLP), as quais compõem o paradigma conexionista da Inteligência Artificial (RUSSEL; NORVIG, 2010). As RNAs MLP são compostas por um conjunto de unidades básicas de processamento, denominadas neurônios artificiais, que são fortemente interconectados e organizados segundo camadas, operando nos dados de entrada com o objetivo de inferir uma função que aprenda as respectivas saídas, conforme o paradigma de Aprendizado Supervisionado (KHAN et al., 2018).

Os neurônios artificiais são inspirados nos neurônios biológicos e funcionam de forma independente como unidades computacionais simples que recebem sinais de entrada e atribui-lhes pesos, produzindo um sinal de saída usando uma função de ativação. A Figura 2.1 ilustra a estrutura de um neurônio abstrato com n entradas, em que cada canal de entrada é um valor $x_i \in R$. Cada entrada possui um peso associado $w_i \in R$, e o corpo do neurônio calcula a soma ponderada das entradas e sujeita esse valor a uma função de ativação $f(\cdot)$, que define o limite no qual o neurônio produz um sinal de saída ativado (BRAGA; CARVALHO; LUDERMIR, 2012).

O poder computacional de um neurônio individual é limitado às funções linearmente separáveis. Ao dispor os neurônios em camadas e conectá-los a todos os neurônios da camada seguinte, conforme estabelecido nas RNAs MLP, torna-se possível uma transformação sucessiva nos dados de entrada que os tornam linearmente separáveis à medida que se aproximam da camada de saída (FACELI et al., 2015), conforme ilustrado na Figura 2.2.

O aprendizado das RNAs se dá por meio de um processo iterativo de ajustes aplicado a

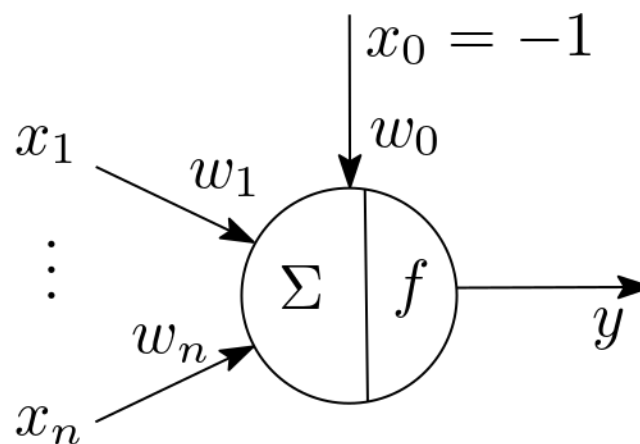


Figura 2.1: Modelo abstrato de um neurônio artificial. Fonte: (BRAGA; CARVALHO; LUDERMIR, 2012).

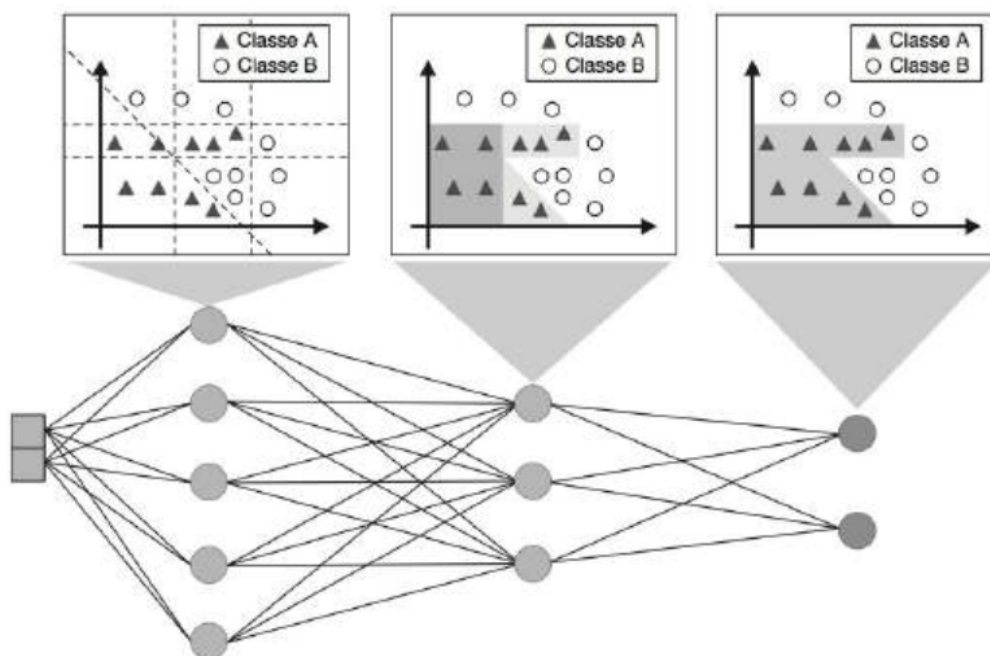


Figura 2.2: Papel desempenhado pelos neurônios em cada camada de uma RNA MLP. Fonte: (FACELI et al., 2015).

seus pesos, fazendo com que o modelo obtenha melhor desempenho, normalmente medido pela acurácia preditiva. Esses algoritmos, são formados por um conjunto de regras bem definidas que determinam quando e como dever ser alterado o valor de cada peso (FACELI et al., 2015). Um método popular para o aprendizado supervisionado de MLPs é o algoritmo de *backpropagation*, que consiste em duas fases: uma fase pra frente (*forward*) e outra pra trás (*backwards*). Na fase *forward*, os pesos sinápticos da rede são fixos e o sinal de entrada é propagado pela rede, camada

por camada, até atingir a saída. Na segunda fase *backwards*, um sinal de erro é produzido comparando a saída da rede com uma resposta desejada. O sinal de erro resultante é propagado através da rede, camada por camada, porém a propagação é realizada pra trás (FACELI et al., 2015; HAYKIN, 2009). Nesta segunda fase, ajustes sucessivos são feitos nos pesos da rede utilizando o algoritmo do gradiente descendente (BRAGA; CARVALHO; LUDERMIR, 2012).

As CNNs são uma categoria de RNAs especialmente voltadas para lidar com dados de alta dimensionalidade, tais como imagens, áudio, vídeo, etc. Elas operam de maneira muito semelhante às RNAs MLP, mas com algumas diferenças fundamentais no tocante aos neurônios e aos tipos de camada. Os neurônios artificiais convolucionais são caracterizados por um filtro que é convoluído com a entrada, permitindo uma extração apropriada de características frente ao tipo de dado de entrada, pois são capazes de abstrair características como texturas, contornos, cores, etc. As camadas compostas de neurônios convolucionais são chamadas de camadas convolucionais (KHAN et al., 2018).

As CNNs também fazem uso de camadas de *pooling*, tipicamente dispostas após as camadas convolucionais, com vistas a promover uma redução de dimensionalidade, produzindo representações condensadas das entradas. As diversas camadas convolucionais e de *pooling* na arquitetura de uma CNN são responsáveis pela extração automática de características da entrada, as quais passam por camadas densas, tipicamente dispostas ao final da CNN, com vistas a produzir a saída desejada para a tarefa em questão (classificação, regressão, etc.) (BUDUMA, 2017; CHOLLET, 2018).

Como a extração de características nas CNNs é automática e massiva, é importante utilizar alguma estratégia para descartar valores que podem estar associados à ruído ou que possuem pouca influência na saída do modelo. Isso é feito com camadas de *Dropout*, as quais são compostas de neurônios com limiar para descartar aleatoriamente certas conexões ao longo do treino, colaborando para a regularização e para a diminuição no número de parâmetros ajustáveis das CNNs (BUDUMA, 2017; CHOLLET, 2018).

As CNNs são tipicamente utilizadas perante o paradigma de Aprendizado Supervisionado com especial destaque para tarefas de classificação em Visão Computacional, representando o

estado da arte em diversos problemas nesse domínio (KHAN et al., 2018). Citam-se algumas arquiteturas canônicas de CNNs, tais como, LeNet, AlexNet, Inception e VGG-16, as quais obtiveram desempenho notável em diversas edições do *ImageNet Large Scale Visual Recognition Challenge* (LSVRC) (RUSSAKOVSKY et al., 2015).

2.2 Detecção de Objetos com Redes Neurais Convolucionais

A detecção e o reconhecimento de objetos são componentes chave para a maioria dos sistemas de visão computacional e determinam o desempenho de muitos aplicativos, como rastreamento, recuperação, vigilância por vídeo e legendagem de imagens. O desempenho da detecção e reconhecimento de objetos depende muito da qualidade das características extraídas e da robustez dos modelos, uma vez que a aparência das imagens pode ser influenciada por diversos fatores, como condições de iluminação, pose, refletância dos objetos e características intrínsecas das câmeras (JIANG et al., 2019). Por isso que, nos últimos anos, as redes neurais profundas ganharam espaço nos problemas mais complexos da atualidade como processamento de imagens e visão computacional. A Figura 2.3 ilustra a detecção de várias classes para uma mesma entrada.



Figura 2.3: Tarefa de detecção de objetos sendo realizada. Fonte: (REDMON et al., 2015)

É importante salientar que a classificação de uma imagem, isto é, um modelo classificar o conteúdo de uma foto, por exemplo, é diferente de detectar exatamente em que área da foto o objeto encontra-se. Com isso, é possível perceber que os modelos que abordam essa tarefa precisam ser mais robustos pois existe agora uma nova etapa na qual o modelo precisa realizar: definir na imagem quais os limites da entidade em questão. Em resumo, o modelo, para detecção de objetos numa imagem, deve ser capaz de identificar a localização de uma ou mais entidades (as classes) além de emitir qual classe ela(s) pertence(m) (MICHELUCCI, 2019).

A métrica necessária para verificar a qualidade dessa tarefa é justamente a interseção da área correspondente ao objeto (resultado esperado) em relação à área detectada sobre a união dessas áreas em questão, conhecida também como Interseção sobre União (IoU) (CHOLLET, 2018). O desafio está em detectar as áreas das entidades na imagem. Uma das primeiras estratégias que solucionaram esse problema é a Aproximação da Janela Deslizante: uma certa área (a janela deslizante) igual ou menor que a imagem total é inicialmente definida e percorre todas as combinações possíveis com a imagem original, calculando o IoU. Em seguida, outra janela deslizante é definida com outro tamanho e o processo é refeito. No fim, verifica-se qual IoU foi o mais satisfatório para o resultado final da detecção da entidade. Essa estratégia é bastante custosa pois podem existir milhares de combinações possíveis em cada teste, razão pela qual não é muito utilizada (MICHELUCCI, 2019).

Outra estratégia para a detecção é conhecida como Region-Based CNN (R-CNN), ou CNN Baseada em Regiões. Como a estratégia das janelas deslizantes necessita de muitas combinações, a baseada em regiões propõe inicialmente 2000 regiões na imagem. Em seguida as regiões adjacentes são unidas baseadas em características similares como cor, textura e forma, sendo possível visualizar a técnica na Figura 2.4. Então, com menos regiões para serem analisadas, o cálculo de IoU é também realizado e a melhor métrica acaba escolhendo também a melhor região para o resultado (MICHELUCCI, 2019).



Figura 2.4: Pesquisa seletiva destacando regiões de interesse. Fonte: (MICHELUCCI, 2019)

Como as R-CNN são modelos mais complexos devido ao janelamento, uma solução para torná-las mais rápidas foi desenvolvida: a chamada Fast R-CNN. A razão pela qual esse algoritmo é mais rápido que o R-CNN está no fato de que torna-se desnecessário alimentar 2000 propostas de região para a rede neural convolucional, o processo é realizado uma única vez com a extração de mapa de características da imagem, similar às características semelhantes das regiões inicializadas nas R-CNN (MICHELUCCI, 2019).

Mesmo as Fast R-CNN tendo se mostrado mais rápidas que as R-CNN convencionais, o custo para treinar o modelo ainda é alto visto que a pesquisa para selecionar as regiões propostas é ainda gargalo do algoritmo. Com isso, uma outra solução para essa etapa foi desenvolvida: utilizar uma outra rede neural para aprender as regiões de dados rotulados, removendo então a etapa demorada de procura seletiva das regiões. Assim, esse modelo foi chamado de Faster R-CNN por ser significativamente mais rápido que R-CNN e Fast R-CNN (MICHELUCCI, 2019).

2.3 Família YOLO

Com a popularização das R-CNNs e o crescente interesse prático nas tarefas de detecção de objetos, em 2016 foi proposta a R-CNN YOLO (acrônimo para *You Only Look Once*), a qual se propõe a realizar as tarefas de localização e classificação de múltiplos objetos em uma única etapa (*single-shot*) (REDMON et al., 2016). Essa R-CNN mostrou-se significativamente mais rápida para detecção do que os modelos anteriores do mesmo tipo disponíveis na literatura (MICHELUCCI, 2019).

Conforme ilustrado na Fig. 2.5, a abordagem de detecção de objetos por R-CNNs YOLO consiste nos seguintes passos: primeiramente há a divisão da imagem de entrada em uma grade de dimensões $S \times S$; em seguida, para cada célula, há uma verificação se há um objeto cujo centro é englobado pela mesma; para as células em que tal verificação foi afirmativa, calcula-se a quantidade B de caixas delimitadores com seus respectivos coeficientes de confiança – obtidos por meio da multiplicação da métrica IoU (do inglês, *Intersection Over Union*) com a probabilidade da célula conter um certo objeto de uma dada classe – refletindo numericamente o quão acurada está uma caixa delimitadora que contém o objeto; por fim, células adjacentes com alta confiança para um mesmo tipo de objeto são unificadas, culminando na previsão das coordenadas das caixas e seus respectivos rótulos (MICHELUCCI, 2019).

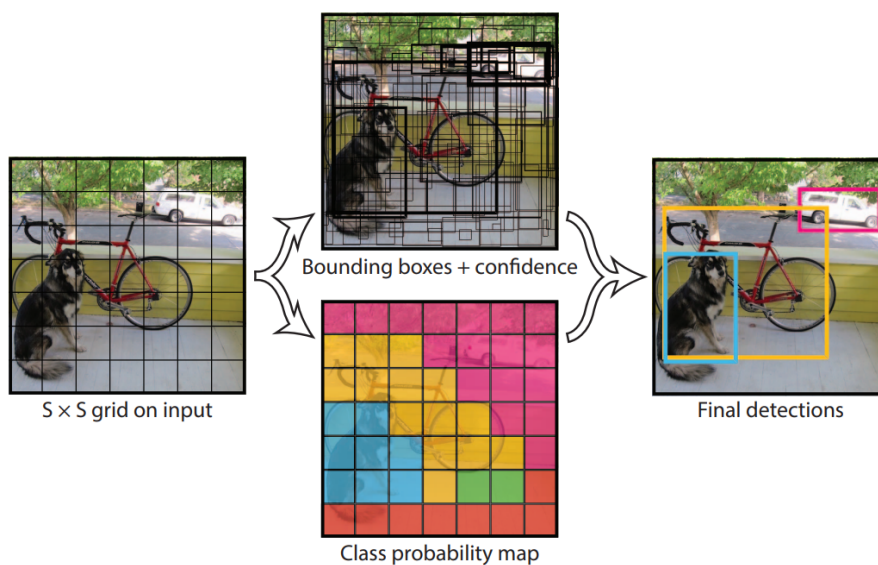


Figura 2.5: Abordagem da YOLO para detecção de objetos. Fonte: (REDMON et al., 2016).

A arquitetura utilizada na YOLOv1 (primeira versão) foi inspirada na CNN Inception pré-treinada com pesos oriundos da base de dados ImageNet (DENG et al., 2009). Esta arquitetura é constituída por 20 camadas convolucionais seguidas por 2 camadas totalmente conectadas. Ao invés de usar os módulos Inception originais, os autores da YOLO optaram por utilizar camadas de *pooling* com filtros de dimensões 1×1 para redução de dimensionalidade seguidos de camadas convolucionais com filtros 3×3 , conforme ilustrado na Figura 2.6.

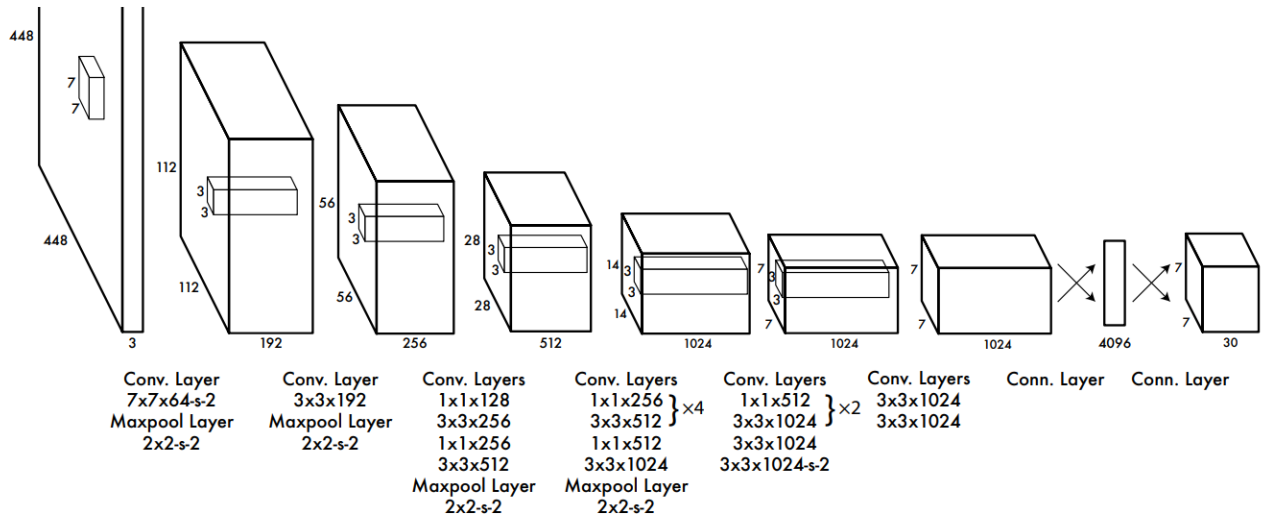


Figura 2.6: Arquitetura da CNN utilizada na primeira versão da YOLO.

Fonte: (REDMON et al., 2016).

Ao longos dos anos, novas técnicas e melhorias foram incorporadas em versões atualizadas de modelos da Família YOLO. Para a YOLOv4, por exemplo, utilizou-se o *backbone* CSPDarkNet53, tornando-a duas vezes mais rápida que outras arquiteturas de referência na época do seu lançamento. Ainda, comparando com a sua versão anterior (YOLOv3) as métricas AP e FPS foram melhoradas em 10 % e 12 %, respectivamente (BOCHKOVSKIY; WANG; LIAO, 2020).

A YOLOv5 decorreu logo após a proposição da YOLOv4 tendo como maior diferença a implementação do *framework* DarkNet de forma nativa com o Pytorch, o que facilitou drasticamente o treinamento e o *deployment* desse modelo. A principal inovação da YOLOv5 em relação às suas antecessoras foi a introdução do *Mosaic Augmentation*, uma técnica de regularização por meio do aumento artificial de dados que combina quatro imagens em quatro blocos de proporção aleatória. Conforme argumentam seus proponentes, a principal vantagem dessa

inovação é a melhoria de desempenho na detecção de objetos pequenos. No sétimo *release*, diferentes versões da YOLOv5 foram propostas para atender domínios específicos, citadas a seguir por ordem crescente de parâmetros e robustez: *Nano*, *Small*, *Medium*, *Large* e *XLarge*. Os modelos mais simples visam aplicações embarcadas por serem mais leves, enquanto os mais complexos foram projetados para problemas mais difíceis, com imagens de maior dimensão, grande número de classes e tamanho de caixas delimitadoras bastante variados (JOCHER et al., 2022). A Figura 2.7 ilustra o desempenho das diferentes versões da YOLOv5 perante a base de dados MS COCO (LIN et al., 2014), um *benchmark* da literatura para detecção de objetos.

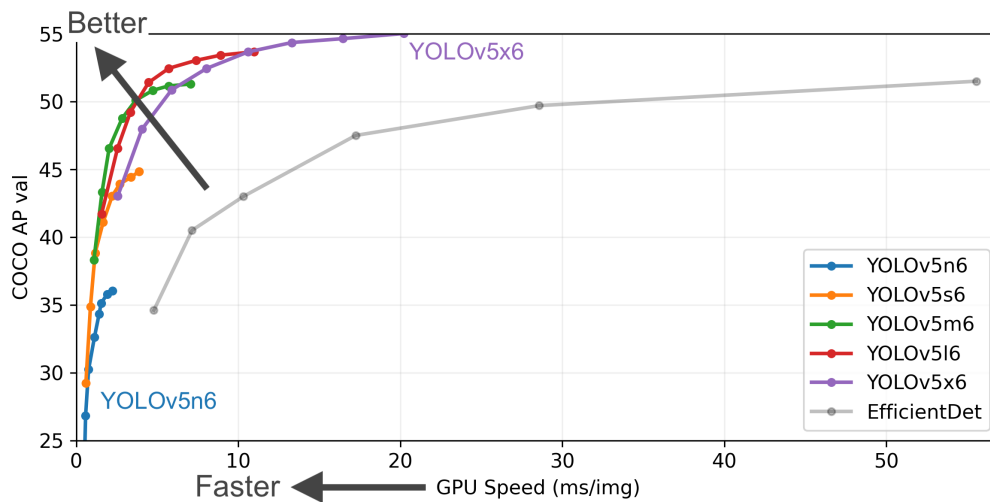


Figura 2.7: Desempenho dos modelos YOLOv5 perante o MS COCO.
Fonte: (JOCHER et al., 2022).

Outras versões da YOLO foram propostas por diferentes autores, mas nem todas se consolidaram perante a comunidade técnico-científica pois, apesar de apresentarem um bom desempenho no *benchmark* MS COCO, não reproduziam tal eficiência em problemas de outros domínios. Ademais, era notável a dificuldade no treinamento em face da carência de documentação e de uma comunidade pequena de usuários. Como exemplo, cita-se o caso da YOLOv6 (LI et al., 2022).

A YOLOv7 deu continuidade ao aprimoramento dos modelos da Família YOLO ao incorporar um processo denominado reparametrização, com o intuito de diminuir o tamanho do modelo para *deployment*, o que é especialmente útil para sistemas embarcados. Também faz uso de *Feature Pyramid Networks* em sua arquitetura, as quais empilham camadas convoluci-

onais e produzem previsões em diferentes escalas, colaborando para um melhor desempenho. Diferentemente das versões anteriores, propôs uma arquitetura padrão e uma estratégia de dimensionamento em escala para outras quantidades de parâmetros, sejam elas maiores ou menores (WANG; BOCHKOVSKIY; LIAO, 2022). A Figura 2.8 apresenta uma análise comparativa entre os diferentes modelos da Família YOLO recém propostos na literatura perante o MS COCO.

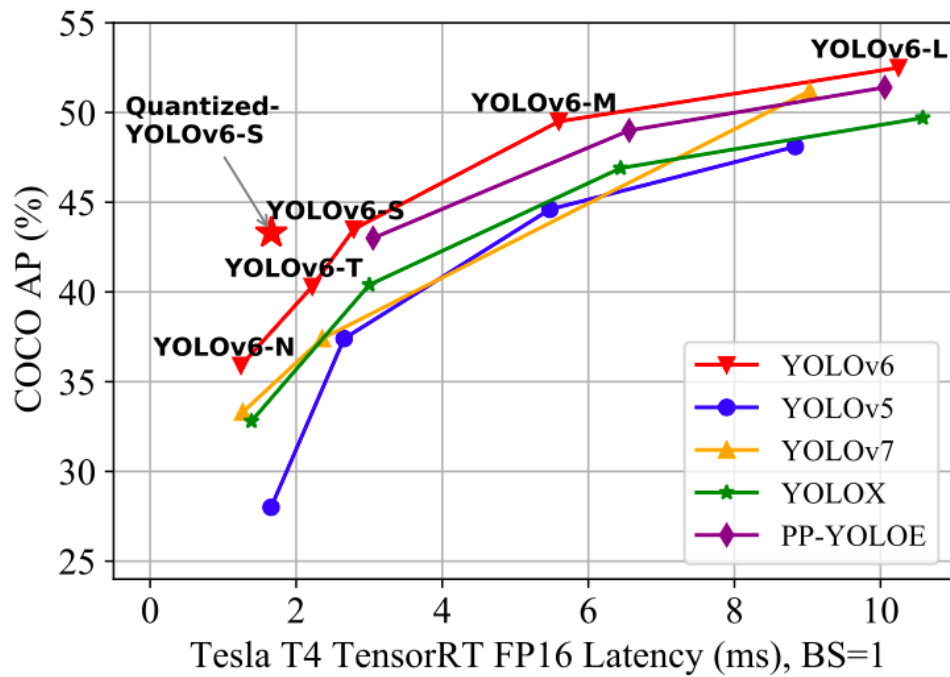


Figura 2.8: Comparação entre as últimas versões da YOLO propostas na literatura. O eixo vertical denota a eficiência do modelo e o eixo horizontal denota o tempo de treinamento. Fonte: (LI et al., 2022).

2.4 Trabalhos Relacionados

A malária ainda é um problema mundial e o padrão ouro para o diagnóstico da doença utiliza a análise de imagens microscópicas do sangue do paciente, um processo considerado lento e que depende de especialistas. Quando feito tardiamente compromete drasticamente o desfecho do tratamento, essencial para uma boa recuperação do paciente (DALRYMPLE; MAPPIN; GETTING, 2015). Assim, soluções que empreguem Visão Computacional para mitigar esse problema são desejadas, pois além de acelerarem o diagnóstico podem ajudar com *insights* e

prognóstico acerca tanto da doença como o estado do paciente (SOUZA et al., 2019).

Observando o estado da arte sobre *Deep Learning* para o diagnóstico de Malária, ressalta-se o trabalho de Guimarães e Guedes (2022). Os autores utilizaram uma base de dados contendo 27.588 imagens de células do sangue humano oriunda de um pré-processamento em imagens de lâminas de exame de sangue. Os modelos de CNNs abordados pelos autores que realizam a tarefa de classificação foram: VGG-16, InceptionV3, ResNet-50 e EfficientNetB0, cada um com o treinamento utilizando 250 épocas. Foram obtidas ótimas métricas de desempenho, destacando-se a Precisão e F-Score (GUIMARÃES; GUEDES, 2022).

Um trabalho correlato de importante menção foi realizado por Yang e outros, os quais fizeram uso da mesma base de dados considerada no escopo desse trabalho. Os autores consideraram duas tarefas de Aprendizado de Máquina para abordar o problema: a localização dos protozoários e sua posterior classificação. Nessa última tarefa, utilizaram os modelos VGG-19, ResNet-50, AlexNet e uma arquitetura própria de CNN. Como os modelos mencionados eram voltados apenas para a tarefa de classificação, as imagens originais sofreram um pré-processamento para destacar as regiões candidatas por meio de um algoritmo de janelamento. Por fim, ao aferir o desempenho, ressaltam-se métricas satisfatórias no que tange à classificação, com especial destaque para AUC, Acurácia e Precisão, inclusive para o modelo proposto (YANG et al., 2020).

Vale ressaltar que, dos trabalhos citados, a tarefa principal dos modelos de *Deep Learning* utilizando CNNs foi a classificação, ou seja, as imagens de lâminas de sangue para o exame da malária não poderiam ser usadas tal como foram capturadas, sendo necessária uma etapa de pré-processamento, quer feita por um especialista ou por um algoritmo de localização. No trabalho de Guimarães e Guedes, a base de dados possuía todas as células do sangue e protozoários devidamente individualizadas (GUIMARÃES; GUEDES, 2022). Em relação ao trabalho de Yang e outros, um algoritmo para destacar as células candidatas precisou ser aplicado à base de dados para se extrair as imagens que, posteriormente, foram fornecidas aos modelos de classificação (YANG et al., 2020). A fim de contornar essas limitações observadas na literatura, o presente trabalho utilizou as imagens microscópicas originais obtidas nos exames como entrada

para um modelo de *Deep Learning* para detecção de objetos *single-shot* da Família YOLO, o qual localiza e classifica todos os protozoários na imagem original em uma única passada da imagem de entrada pela rede, sem qualquer demanda ou ônus de pré-processamento.

Capítulo 3

Materiais e Métodos

Este trabalho aborda o problema de Visão Computacional de detecção de protozoários causadores de malária em imagens de lâminas de sangue capturadas por microscópio. Para tanto, considera a tarefa de detecção de objetos de uma única classe perante o paradigma de Aprendizado Supervisionado utilizando R-CNNs da Família YOLO. Para elaborar a solução proposta, os dados experimentais utilizados para aquisição de experiência são apresentados na Seção 3.1; o detalhamento dos modelos YOLO, suas versões e parametrizações são apresentados na Seção 3.2; e, por fim, a apresentação da estratégia de validação experimental e da coleta de métricas de desempenho encontram-se na Seção 3.3.

3.1 Dados Experimentais

Os dados experimentais utilizados no escopo desse trabalho são oriundos de uma base de dados pública do Hospital Universitário Chittagong em Bangladesh com 150 pacientes e 1830 imagens de lâminas de sangue (YANG et al., 2020), capturadas por uma câmera de celular conectada a um microscópio. As imagens obtidas possuem a resolução de 4032×3024 px e, para cada uma, existe um arquivo com os rótulos contendo informações dos tamanhos circulares e localizações dos protozoários da malária ou das células brancas do sangue. As anotações relativas às células brancas foram descartadas por não serem do escopo de interesse do problema em questão. No total, havia 84.509 rótulos para a classe de interesse nos exemplos da base de dados.

Como os modelos da Família YOLO utilizam caixas retangulares para detecção dos objetos, foi necessário realizar uma etapa de pré-processamento para substituir os círculos originalmente disponíveis. Para tanto, calculou-se o raio do círculo e obteve-se então a caixa delimitadora de formato quadrado circunscrita ao círculo, conforme exemplificado na Figura 3.1. Ademais, observou-se que 0,55 % das imagens não possuíam rótulos.

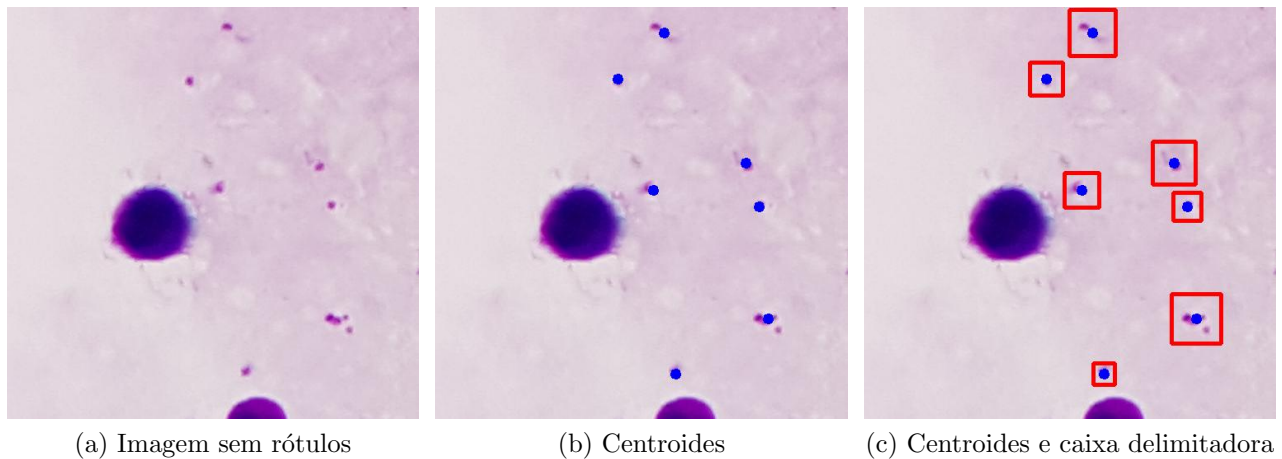


Figura 3.1: Exemplo com a imagem original, acrescida de centroides e das caixas delimitadoras quadradas.

Ao efetuar uma análise exploratória das imagens e dos rótulos disponíveis nas mesmas, conforme histograma da Figura 3.2, foi possível perceber que a maioria das imagens (967) possuía até 25 caixas delimitadoras. Do total de imagens, somente 17 não possuíam nenhum rótulo (0,55 % dos exemplos da base de dados), indicando pacientes saudáveis. Embora não sejam relevantes para o aprendizado de padrões a respeito dos protozoários, colaboram no aprendizado do *background* pelos modelos de detecção com *Deep Learning*. Observou-se também que poucas imagens possuíam mais de 200 protozoários com caixas delimitadoras, correspondendo a 2,1 % dos exemplos da base de dados.

Um mapa de calor foi elaborado para ilustrar a disposição espacial dos rótulos das caixas delimitadoras nas imagens da base de dados, conforme mostrado na Figura 3.3. Nota-se uma distribuição uniforme de tais caixas sobre toda a região visível pelo microscópio, o que reforça a dificuldade na tarefa de detecção por não haver boas regiões candidatas na distribuição *a priori* dos exemplos.

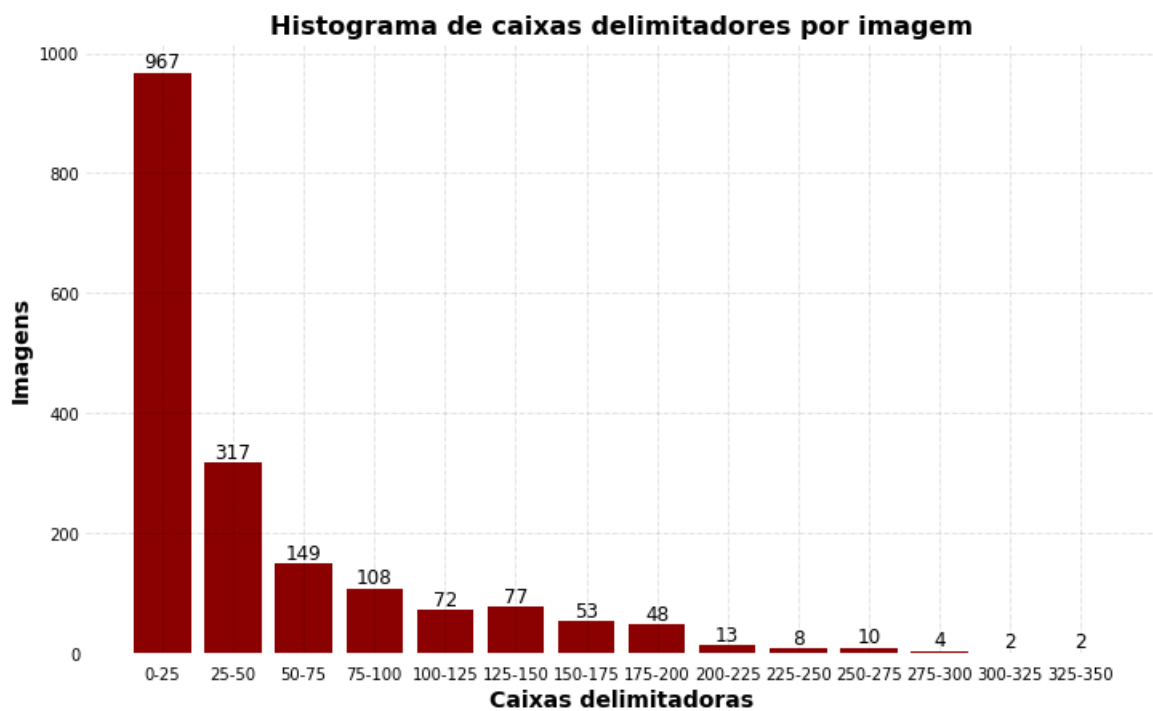


Figura 3.2: Histograma da quantidade de caixas delimitadoras por imagens da base de dados.

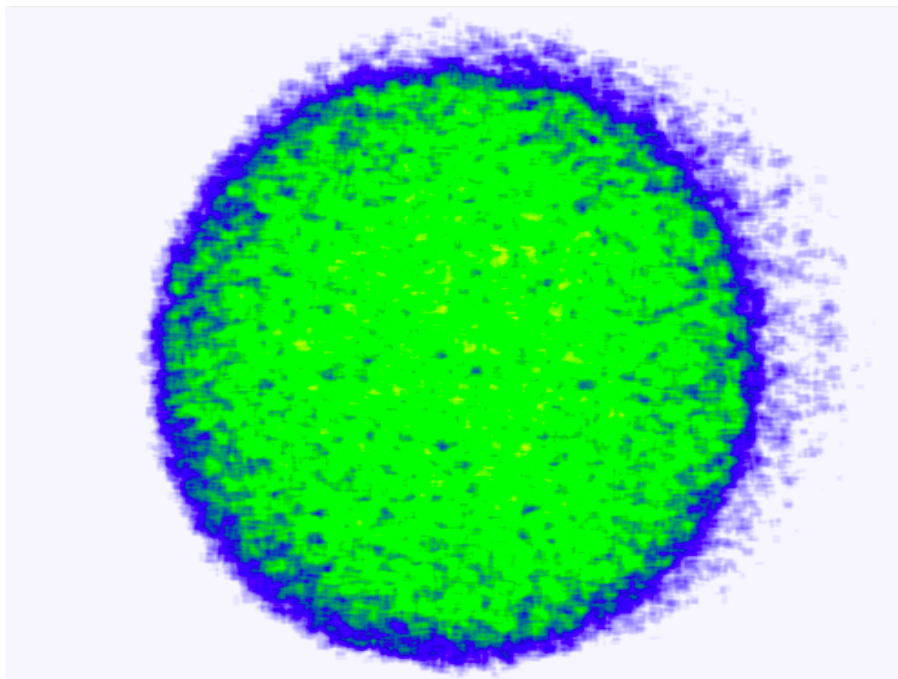


Figura 3.3: Mapa de calor da disposição das caixas delimitadoras nos exemplos.

Analisando o tamanho das caixas delimitadoras, conforme disposto na Tabela 3.1, é possível afirmar que sua área é de cerca de 1,5% da área total da imagem. Para a detecção de objetos com R-CNNs da família YOLO, em particular, objetos com percentual menor que 12,5% da

área total são considerados de difícil detecção (XIAO, 2021). Esse fato reforça as dificuldades em abordar um problema realístico no âmbito da detecção automática de objetos

Tabela 3.1: Descrição estatística do tamanho das caixas delimitadoras.

	Lado			Caixas Delimitadoras		
	Média	Máx	Mín	Média	Máx	Mín
Protozoário	41.84 ± 9.71	191	3	46.17 ± 56.35	341	0

3.2 Modelos, Parâmetros e Hiper-parâmetros

No escopo deste trabalho foram considerados R-CNNs da Família YOLO dos modelos YOLOv5 (JOCHER et al., 2022) e YOLOv7 (WANG; BOCHKOVSKIY; LIAO, 2022), considerando as versões dispostas na Tabela 3.2.

Tabela 3.2: Modelos da Família YOLO considerados e suas características.

Modelo	Versão	Camadas	Parâmetros (Mi)
YOLOv5	<i>Nano</i>	10	1,7
YOLOv5	<i>Small</i>	20	7
YOLOv5	<i>Medium</i>	29	20,8
YOLOv7	<i>Tiny</i>	-	6
YOLOv7	<i>Normal</i>	-	36,4
YOLOv7	<i>X</i>	-	70,7

Como o número de parâmetros na YOLOv7 em todas as suas versões é substancialmente superior ao número relacionado de parâmetros na YOLOv5, foi necessário levar em consideração o ônus computacional de tempo e memória para o planejamento dos experimentos. No caso das YOLOv5, optou-se por utilizar 300 ou 500 épocas e *batches* de tamanho 16. Para as YOLOv7, considerou-se apenas 500 épocas, pois supõe-se que mais parâmetros requeiram uma maior quantidade de ajustes. Nas versões *Tiny* e *Normal* utilizou-se *batches* de tamanho 8, mas para a versão *X* apenas foi viável utilizar *batches* de tamanho 4. Em todos os casos, considerou-se a regularização com *Early Stopping* e paciência igual a 100, pesos pré-treinados e demais

hiperparâmetros com valores padrão.

3.3 Avaliação de Desempenho

Para avaliação de desempenho, foi utilizada a técnica de validação cruzada *holdout*, com 70% dos dados destinados ao treinamento, 10% para validação e 20% para testes. As métricas de desempenho foram aferidas na partição de testes e contemplam: Precisão, Revocação, *F-Score* e mAP com *threshold* $t \geq 0.5$ para o IoU (*Intersection over Union*), a qual será denotada como mAP@0.5. Uma explicação detalhada do cálculo e semântica de tais métricas no contexto da detecção de objetos em Visão Computacional encontra-se disponível no *survey* de Padilla *et al.* (2020).

Capítulo 4

Resultados e Discussão

Para a execução dos *scripts* de treinamento foi utilizado um servidor com a seguinte configuração: processador Intel(R) Core(TM) i7-8700 CPU @ 3.20GHz, 32 GB de memória principal, 960 GB de memória secundária e 2 placas de vídeo NVIDIA GTX 1080 Ti com 11 GB de VRAM para aceleração em *hardware* do treinamento. Os resultados obtidos da aferição do desempenho para as diferentes configurações propostas encontram-se dispostos na Tabela 4.1.

Tabela 4.1: Síntese dos resultados experimentais.

Modelo	Versão	Épocas	Precisão	Revocação	F-Score	mAP@0.5	Parâmetros	Tempo
YOLOv5	<i>Nano</i>	293/300	79,5 %	77,5 %	78,48 %	78,3 %	1.760.518	10 h19 min
YOLOv5	<i>Small</i>	173/300	81,5 %	79,7 %	80,58 %	80,7 %	7.012.822	7 h26 min
YOLOv5	<i>Medium</i>	300/300	59,5 %	63,1 %	61,20 %	55,5 %	20.852.934	9 h7 min
YOLOv5	<i>Nano</i>	500/500	79,9 %	78,2 %	79,04 %	79,0 %	1.760.518	12 h38 min
YOLOv5	<i>Small</i>	194/500	81,4 %	79,4 %	80,38 %	80,4 %	7.012.822	6 h
YOLOv5	<i>Medium</i>	264/500	81,7 %	78,9 %	80,27 %	79,9 %	20.852.934	9 h58 min
YOLOv7	<i>Tiny</i>	500/500	75,2 %	77,9 %	76,52 %	77,0 %	6.007.596	23 h1 min
YOLOv7	<i>Normal</i>	500/500	79,2 %	78,8 %	78,99 %	79,8 %	36.481.772	26 h34 min
YOLOv7	<i>X</i>	500/500	80,6 %	79,9 %	80,24 %	81,4 %	70.782.444	52 h5 min

Ao tomar o F-Score aferido em todos os experimentos, tem-se que a escolha de detectores de objetos baseados em *Deep Learning* da Família YOLO mostrou-se uma escolha acertada, pois em 90 % dos cenários essa métrica mostrou-se acima de 78 %. Ao tomar o mAP@0.5, uma métrica de referência no âmbito da detecção de objetos, verifica-se que esta também foi alta, sendo superior a 77 % no mesmo quantitativo de experimentos. Este argumento é corroborado pelo fato da Família YOLO ser estado da arte em diversas tarefas de detecção de objetos com

Deep Learning (SUMIT et al., 2020).

O modelo YOLOv5 *Medium* obteve resultados contraintuitivos pois, quando treinado com 300 épocas não incorreu em *early stopping*, mas obteve o menor desempenho geral em termos de mAP@0.5. Esse resultado é indicativo de que uma grande quantidade de parâmetros não necessariamente implica em melhor desempenho. Apesar disso, no experimento em que foram consideradas 500 épocas, visando favorecer um melhor ajuste de parâmetros, o modelo conseguiu melhor ajuste utilizando apenas 264 épocas, com desempenho 30,53 % superior ao cenário anterior. Percebe-se, portanto, que o modelo em questão é de difícil ajuste fino e o seu desempenho é delicado em razão da flutuação estocástica, ou seja, não demonstra robustez.

O modelo YOLOv5 *Nano* apresentou um bom desempenho no tocante ao mAP@0.5, embora tenha demandado mais tempo de treinamento dentre todos os YOLOv5 avaliados. Apesar disso, esse é um dos modelos com maior potencial de uso prático em dispositivos embarcados, pois possui menor tempo de inferência dentre todos os avaliados em consequência do menor número de parâmetros.

Em relação aos modelos YOLOv7, as métricas obtidas foram parecidas com aquelas das YOLOv5, apesar de ser necessário ressaltar o aumento significativo no número de parâmetros, no tempo de treino e no número de épocas necessárias. Ressalta-se o modelo YOLOv7 *X* como tendo obtido maior mAP@0.5 nos experimentos realizados. Os gráficos do monitoramento da perda e da acurácia desse modelo nos conjuntos de treinamento e validação encontram-se ilustrados na Figura 4.1. As Figuras 4.2 e 4.3 ilustram detecções de exemplos do conjunto de testes efetuadas por esse modelo, denotadas em azul, enquanto as caixas delimitadoras desejadas são denotadas em vermelho. É perceptível um bom grau de sobreposição entre os objetos desejados e os previstos pelo modelo, embora estes últimos venham a ter uma área menor que o esperado.

Apesar dos resultados superiores da YOLOv7 *X* no tocante ao mAP@0.5 se mostrarem melhores que todos os demais modelos avaliados, não é possível afirmar a supremacia dessa solução perante as demais, haja vista, por exemplo, a menor precisão desse modelo frente à YOLOv5 *Small*, a qual possui 90,09 % menos parâmetros. Seriam necessários mais experimentos

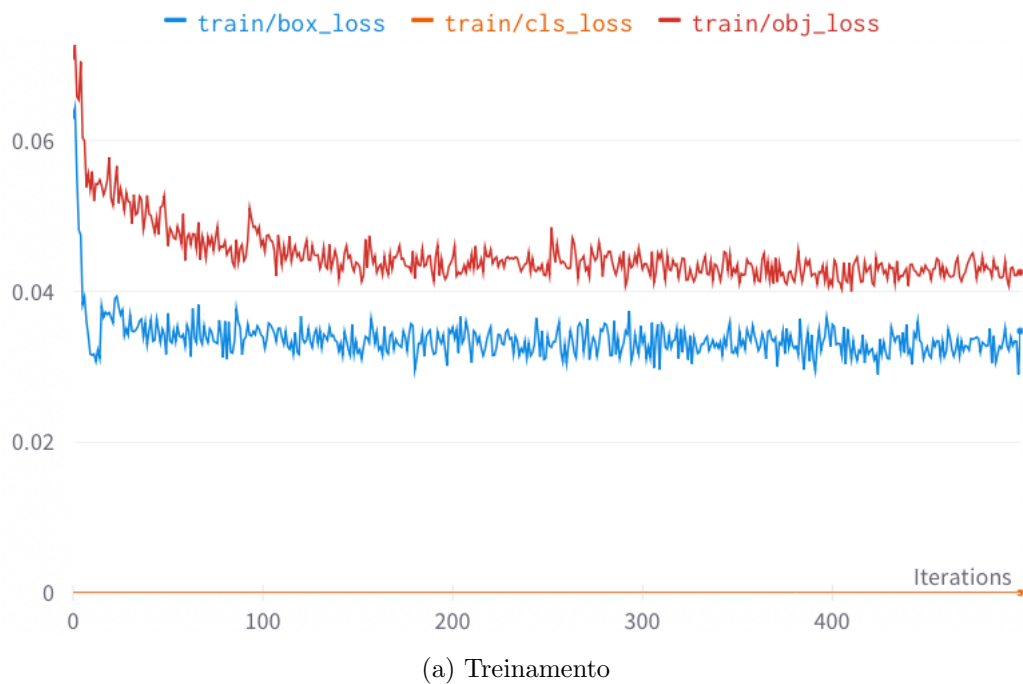


Figura 4.1: Gráficos de monitoramento de métricas da YOLOv7 X perante os conjuntos de treino e de validação.

e uma análise estatística cuidadosa para derivar conclusões nesse sentido. De maneira geral, todos os modelos mostraram-se competitivos para a tarefa, de natureza realística e com um grau de dificuldade acentuado pela área das caixas delimitadoras e pela distribuição uniforme de sua disposição nas imagens.

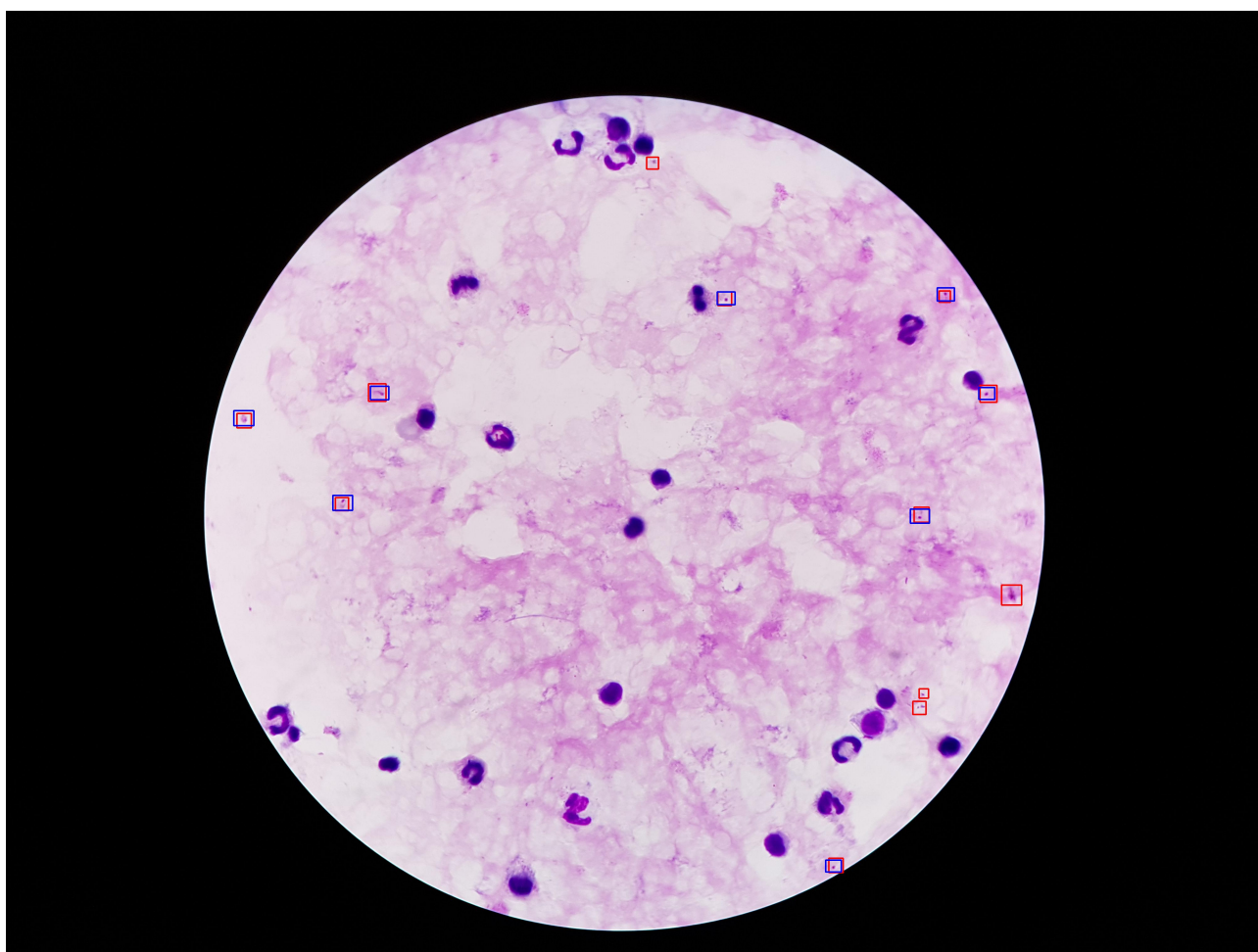


Figura 4.2: Exemplo com poucos protozoários e suas respectivas detecções pela YOLOv7 X.

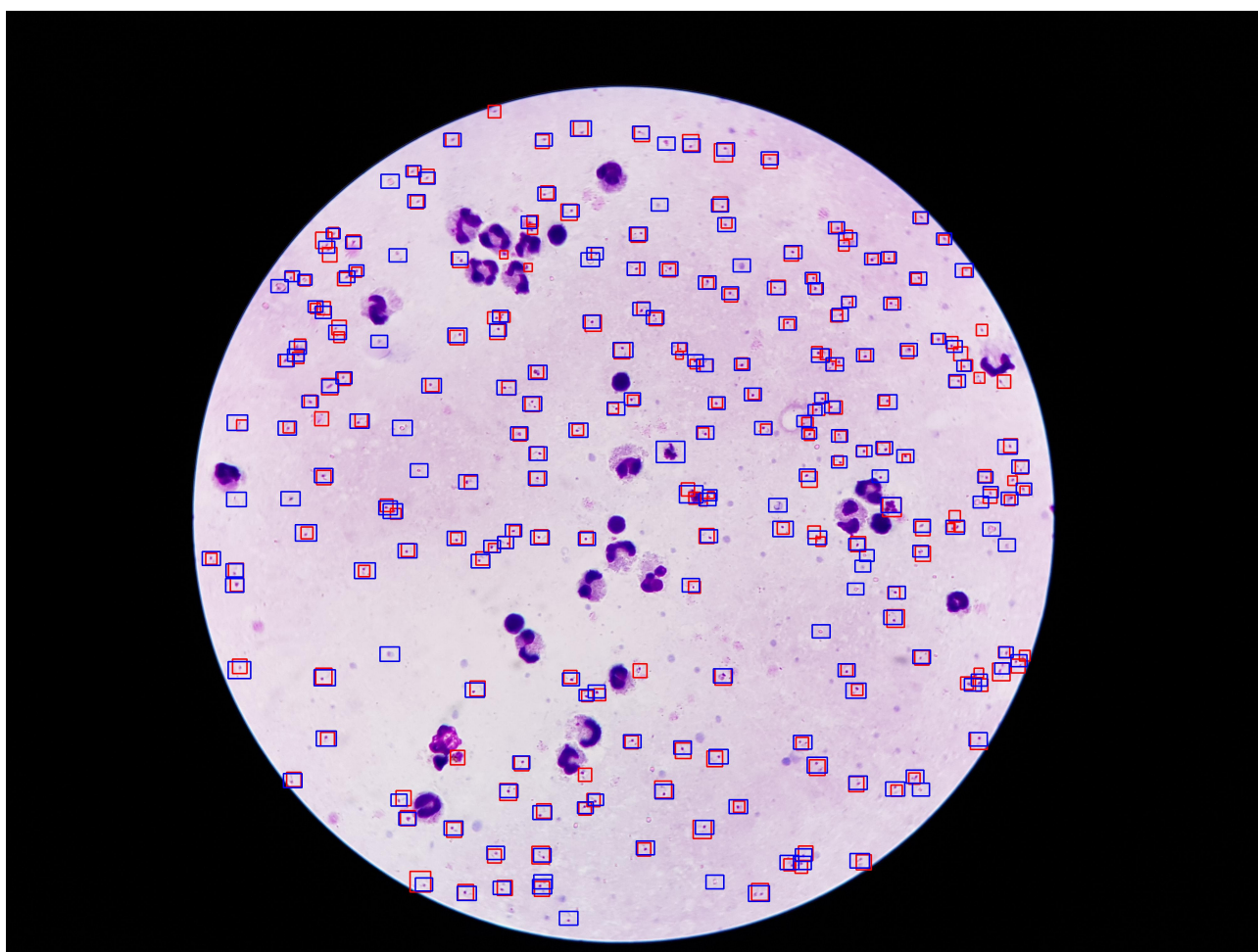


Figura 4.3: Exemplo com muitos protozoários e suas respectivas detecções pela YOLOv7 *X*.

Capítulo 5

Considerações Finais e Trabalhos Futuros

O presente trabalho de conclusão de curso mostrou a viabilidade dos modelos de detecção de objetos baseados em *Deep Learning* da Família YOLO no tocante à tarefa de detecção de protozoários causadoras de malária em imagens de lâminas de sangue oriundas de microscópio. Esses resultados foram observados a partir de experimentos computacionais com modelos YOLOv5 e YOLOv7 em uma base de dados realística da literatura contendo objetos pequenos e uniformemente distribuídos na imagem, o que a tornou especialmente desafiadora. As entradas dos modelos não necessitaram de pré-processamento computacional ou de intervenção humana para sua preparação para os modelos, o que se mostrou um avanço no estado da arte frente à outras contribuições identificadas na literatura. Os resultados evidenciaram a YOLOv7 *X* como tendo melhor desempenho nos experimentos realizados, mas não foi possível assegurar a superioridade desse modelo frente aos demais, sendo necessários mais experimentos e também a ponderação quanto aos recursos computacionais disponível no âmbito de utilização da solução. Nesse sentido, ressalta-se que os resultados obtidos não se propõem em hipótese alguma a substituir especialistas humanos na tarefa, sendo essencial aprofundar as análises para verificar tal viabilidade.

Em trabalhos futuros almeja-se avaliar outras versões de modelos da Família YOLO recém propostos na literatura, tais como a YOLOv8 (JOCHER; CHAURASIA; QIU, 2023). Além disso, é importante validar o desempenho dos modelos em mais exemplos oriundos de mais localidades, especialmente da Região Amazônica onde tal doença é endêmica. Para tanto,

encoraja-se pesquisadores de outras áreas a contribuir com bases de dados para esta tarefa de Visão Computacional.

Referências Bibliográficas

- BERZOSA, P. et al. Comparison of three diagnostic methods (microscopy, RDT, and PCR) for the detection of malaria parasites in representative samples from equatorial guinea. *Malar. J.*, Springer Science and Business Media LLC, v. 17, n. 1, p. 333, set. 2018.
- BOCHKOVSKIY, A.; WANG, C.-Y.; LIAO, H.-Y. M. *YOLOv4: Optimal Speed and Accuracy of Object Detection*. 2020. Disponível em <<https://arxiv.org/abs/2004.10934>>. Acesso em 5 de abril de 2023.
- BRAGA, A. de P.; CARVALHO, A. P. de Leon F. de; LUDERMIR, T. B. *Redes Neurais Artificiais: Teoria e Aplicações*. 2. ed. Rio de Janeiro, Brasil: LTC, 2012.
- BRASIL. *Decreto 78.231, de 12 de agosto de 1976*. Brasil: Ministério da Saúde do Brasil, 1976. Disponível em <http://www.planalto.gov.br/ccivil_03/decreto/1970-1979/d78231.htm>. Acesso em 5 de abril de 2023.
- BRASIL. *Boletim Epidemiológico 2013*. 1. ed. Brasil: Ministério da Saúde do Brasil, 2013. 16 p. Disponível em <https://bvsms.saude.gov.br/bvs/periodicos/boletim_epidemiologico_numero_1_2013.pdf>. Acesso em 5 de abril de 2023.
- BRASIL. *Malária*. Brasil: Sistema de Informação de Agravos de Notificação (SINAN), 2019. Disponível em <<http://portalsinan.saude.gov.br/malaria>>. Acesso em 5 de abril de 2023.
- BRASIL. *Boletim Epidemiológico 2020 - Malária*. Especial. Brasil: Ministério da Saúde do Brasil, 2020. 116 p. Disponível em <https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/boletins/boletins-epidemiologicos/especiais/2020/boletim_especial_malaria_1dez20_final.pdf>. Acesso em 5 de abril de 2023.
- BRASIL. *Boletim Epidemiológico 2021*. 17. ed. Brasil: Ministério da Saúde do Brasil, 2021. 29 p. Disponível em <<https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/boletins/boletins-epidemiologicos/edicoes/2022/boletim-epidemiologico-vol-53-no17.pdf>>. Acesso em 5 de abril de 2023.
- BRINK, H.; RICHARDS, J. W.; FETHEROLF, M. *Real-World Machine Learning*. Estados Unidos: Manning Publications, 2017.
- BUDUMA, N. L. N. *Deep Learning with Python*. 1. ed. [S.l.]: O'Reilly Media, 2017.
- CHOLLET, F. *Fundamentals of Deep Learning*. 1. ed. [S.l.]: Manning Shelter Island, 2018.
- DALRYMPLE, U.; MAPPIN, B.; GETTING, P. W. Malaria mapping: understanding the global endemicity of falciparum and vivax malaria. *BMC medicine*, BioMed Central, v. 13, n. 1, p. 1–8, 2015.

- DENG, J. et al. ImageNet: A large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. Miami, FL: IEEE, 2009.
- FACELI, K. et al. *Inteligência Artificial – Uma abordagem de aprendizado de máquina*. Rio de Janeiro: Editora LTC, 2015.
- FIOCRUZ. *Fundação Oswaldo Cruz – Malária*. 2022. Disponível em <<https://portal.fiocruz.br/taxonomia-geral-7-doencas-relacionadas/malaria>>. Acesso em 5 de abril de 2023.
- GOMES, A. P. et al. Malária grave por plasmodium falciparum. *Revista Brasileira de Terapia Intensiva*, GN1 Genesis Network, v. 23, n. 3, p. 358–369, 2011.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. Cambridge: The MIT Press, 2016. v. 1.
- GUIMARÃES, F. B.; GUEDES, E. B. Estudo de caso de transferência de aprendizado com redes neurais convolucionais para classificação de malária. *Trabalho de Conclusão de Curso*, Universidade do Estado do Amazonas, v. 1, n. 1, p. 12, 2022.
- HAYKIN, S. *Neural Networks and Learning Machines*. 3. ed. Estados Unidos: Pearson Prentice-Hall, 2009.
- JIANG, X. et al. (Ed.). *Deep learning in object detection and recognition*. 1. ed. Singapore, Singapore: Springer, 2019.
- JOCHER, G.; CHAURASIA, A.; QIU, J. *YOLOv8 by Ultralytics*. 2023. Disponível em <<https://github.com/ultralytics/ultralytics>>. Acesso em 5 de abril de 2023.
- JOCHER, G. et al. *ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation*. [S.l.]: Zenodo, 2022. Disponível em <<http://dx.doi.org/10.5281/ZENODO.7347926>>. Acesso em 5 de abril de 2023.
- KHAN, S. et al. *A Guide to Convolutional Neural Network for Computer Vision*. 1. ed. [S.l.]: Morgan and Claypool, 2018.
- LEE, J.-G. et al. Deep learning in medical imaging: general overview. *Korean journal of radiology*, The Korean Society of Radiology, v. 18, n. 4, p. 570–584, 2017.
- LI, C. et al. *YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications*. arXiv, 2022. Disponível em <<https://arxiv.org/abs/2209.02976>>. Acesso em 5 de abril de 2023. Disponível em: <<https://arxiv.org/abs/2209.02976>>.
- LIN, T.-Y. et al. Microsoft COCO: Common Objects in Context. In: *Proceedings of the 13th European Conference on Computer Vision (ECCV 2014)*. Suíça: Springer, 2014. p. 740–755.
- LOIOLA, C. C. P.; SILVA, C. M. d.; TAUIL, P. L. Controle da malária no brasil: 1965 a 2001. *Revista Panamericana de Salud Pública*, SciELO Public Health, v. 11, n. 4, p. 235–244, 2002.
- MICHELUCCI, U. *Advanced Applied Deep Learning. Convolutional Neural Network and Object Detection*. 1. ed. [S.l.]: APress, 2019.

- OMS. *World malaria report 2019*. 1. ed. França: Organização Mundial de Saúde, 2019. 232 p. Disponível em <<https://www.who.int/publications/i/item/9789241565721>>. Acesso em 5 de abril de 2023.
- PADILLA, R.; NETTO, S. L.; SILVA, E. A. B. da. A Survey on Performance Metrics for Object-Detection Algorithms. In: *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*. Niterói, Brasil: [s.n.], 2020. p. 237–242.
- PAZ Érika R. S.; SANTIAGO, S. B. Diagnóstico de malária – a importância da habilidade em microscopia. *SAÚDE & CIÊNCIA EM AÇÃO*, UNIFAN, v. 1, n. 1, p. 12, 2015.
- REDMON, J. et al. *You Only Look Once: Unified, Real-Time Object Detection*. arXiv, 2015. Disponível em <<https://arxiv.org/abs/1506.02640>>. Acesso em 5 de abril de 2023. Disponível em: <<https://arxiv.org/abs/1506.02640>>.
- REDMON, J. et al. You only look once: Unified, real-time object detection. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, Estados Unidos: IEEE, 2016.
- RUSSAKOVSKY, O. et al. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, v. 115, n. 3, p. 211–252, 2015.
- RUSSEL, S.; NORVIG, P. *Artificial Intelligence – A Modern Approach*. 3. ed. Nova Jersey: Prentice Hall, 2010.
- SHEN, D.; WU, G.; SUK, H.-I. Deep learning in medical image analysis. *Annu. Rev. Biomed. Eng.*, v. 19, p. 221–248, jun. 2017.
- SOUZA, E. P. de et al. Aplicações do deep learning para diagnóstico de doenças e identificação de insetos vetores. *Saúde em Debate*, Centro Brasileiro de Estudos de Saúde, v. 57, n. 3, p. 126–159, 2019. Disponível em <<https://www.redalyc.org/journal/4063/406369189012/html/>>. Acesso em 5 de abril de 2023.
- SUMIT, S. S. et al. In Object Detection Deep Learning Methods, YOLO Shows Supremum to Mask R-CNN. *Journal of Physics: Conference Series*, IOP Publishing, v. 1529, n. 4, p. 042086, apr 2020.
- WANG, C.-Y.; BOCHKOVSKIY, A.; LIAO, H.-Y. M. *YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors*. 2022. Disponível em <<https://arxiv.org/abs/2207.02696>>. Acesso em 5 de abril de 2023.
- XIAO, J. exYOLO: A small object detector based on YOLOv3 object detector. *Procedia Comput. Sci.*, Elsevier BV, v. 188, p. 18–25, 2021. Disponível em <<https://www.sciencedirect.com/science/article/pii/S1877050921011339>>. Acesso em 5 de abril de 2023.
- XING, F. et al. Deep learning in microscopy image analysis: A survey. *IEEE transactions on neural networks and learning systems*, IEEE, v. 29, n. 10, p. 4550–4568, 2017.

YANG, F. et al. Deep learning for smartphone-based malaria parasite detection in thick blood smears. *IEEE J. Biomed. Health Inform.*, Institute of Electrical and Electronics Engineers (IEEE), v. 24, n. 5, p. 1427–1438, 2020. Disponível em <<https://ieeexplore.ieee.org/document/8846750>>. Acesso em 5 de abril de 2023.