tecnia revista de educação, ciência e tecnologia do ifg

v. 10, Edição Especial 1 | 2025

Submetido 09/06/2024. Aprovado 08/03/2025
Avaliação: revisão duplo-anônimo

Modelos de aprendizado profundo aplicados à detecção de pólipo colorretal

THE APPLICATION OF DEEP LEARNING MODELS IN THE DETECTION OF COLORECTAL POLYPN

MODELOS DE APRENDIZAJE PROFUNDO APLICADOS A LA DETECCIÓN DE PÓLIPOS COLORRECTALES

Álisson Assis Cardoso
Universidade Federal de Goiás (UFG)
alsnac@ufg.br

Diene Xie
Universidade Federal de Goiás (UFG)
diene12374@gmail.com

Larissa Silva Xavier Rosa
Universidade Federal de Goiás (UFG)
larissarosa@discente.ufg.br

Ricardo Augusto Pereira Franco
Universidade Federal de Goiás (UFG)
ricardo@inf.ufg.br

Vilmar Cardoso Prestes Filho
Universidade Federal de Goiás (UFG)
vilmarfilho@discente.ufg.br

Resumo

O câncer colorretal, uma das principais causas de mortalidade no mundo, pode ser prevenido com a detecção precoce de pólipos. Este artigo propõe realizar um estudo de detecção de pólipos em imagens de colonoscopia utilizando modelos de redes neurais profundas de detecção de objetos. Para tanto, uma revisão de trabalhos na literatura é realizada com o objetivo de selecionar modelos avançados de detecção de objetos para serem utilizados. Além disso, são apresentados conjuntos de dados públicos de imagens de exames de colonoscopia, usados nos experimentos para detecção de pólipos. Também são apresentadas análises utilizando a técnica de pré-processamento de equalização de histograma, a fim de melhorar o contraste das imagens e, consequentemente, espera-se melhora no desempenho dos modelos. Resultados em termos de precisão, recall e mean average precision (mAP) são apresentados e usados para fins de comparação entre os modelos implementados. Os resultados obtidos indicam que os modelos treinados para a detecção de pólipos apresentaram resultados superiores em relação aos relatados na literatura, evidenciando que esses modelos podem ser um poderoso aliado da medicina no auxílio à detecção de pólipos e na prevenção do câncer colorretal.

Palavras-chave: detecção de objetos; pólipos; aprendizado profundo; câncer colorretal.

Abstract
The following text presents a synopsis of the abstract. Colorectal cancer, a leading cause of mortality worldwide, is preventable. The primary objective is to facilitate the timely identification of polyps. The present article puts forth a proposal for a study on the detection of polyps in colonoscopy images. The utilization of deep neural network models for the purpose of object detection is a subject of considerable interest. In order to this end, a literature review will identify and select advanced object detection models. The object is to be utilized. Furthermore, public datasets of colonoscopy images will be utilized. The following presentation will outline the utilization of various methodologies in experimental settings aimed at detecting polyps. Analyses employing the histogram equalization preprocessing technique are also presented with the objective of enhancing the quality of the results. The contrast of the images has been demonstrated to enhance the performance of the models. The following section presents the results of the study. Precision, recall, and mean average precision (mAP) are presented and utilized for the purpose of a comparison of the implemented models. The findings suggest that the models were successfully trained. The results obtained from the polyp detection process exhibited a higher degree of precision and accuracy when compared with the results reported in the existing literature. It has been demonstrated that these models have the capacity to serve as a valuable instrument in the field of medicine, particularly in the context of polyp detection. The following text will provide a comprehensive overview of colorectal cancer prevention.

Keywords: Object detection is a process that involves identifying the presence of an object in an image or video. polyps. deep learning. colorectal cancer.

Resumen
El cáncer colorrectal, una de las principales causas de mortalidad en el mundo, se puede prevenir com la detección temprana de pólipos. Este artículo propone realizar un estudio de detección de pólipos em imágenes de colonoscopia utilizando modelos de redes neuronales profundas para la detección de objetos. Para esto, se realiza una revisión de trabajos en la literatura con el objetivo de seleccionar modelos avanzados de detección de objetos a utilizar. Además, se presentan conjuntos de datos públicos de imágenes de exámenes de colonoscopia, utilizados en experimentos para detectar pólipos. También se presentan análisis utilizando la técnica de preprocesamiento de ecualización de histogramas, con el fin de mejorar el contraste de las imágenes y, en consecuencia, se espera una mejora en el rendimiento de los modelos. Se presentan los resultados en términos de precisión, recuperación y precisión promedio media (mAP) y se utiliza para fines de comparación entre los modelos implementados. Los resultados obtenidos indican que los modelos entrenados para la detección de pólipos presentaron resultados superiores en relación a los obtenidos en la literatura, demostrando que estos modelos pueden ser un poderoso aliado en medicina para ayudar en la detección de pólipos y en la prevención del cáncer colorrectal.

Palabras clave: Detección de objetos. Pólipos. Aprendizaje profundo. Cáncer colonrectal .

Introdução

O câncer ainda é uma doença complexa e que representa uma ameaça significativa à saúde global. De acordo com a Agência Internacional de Pesquisa sobre o Câncer (IARC), em 2020, foram estimados cerca de 19,3 milhões de novos casos de câncer e 10 milhões de mortes relacionadas ao câncer em todo o mundo (IARC Global Cancer Observatory, 2020). São vários os fatores de risco associados a essa doença, tais como genética, estilo de vida e exposição ambiental (ACS, 2020).

O câncer é uma doença que pode se manifestar de diferentes formas em diversas partes do corpo humano. Uma dessas formas de câncer que merece destaque

é o câncer de intestino, também conhecido como câncer de cólon e reto ou câncer colorretal. Ele é o terceiro câncer mais comum em homens e o segundo câncer mais comum em mulheres. De acordo com o Instituto Nacional de Câncer, no Brasil, o número estimado de novos casos para cada ano entre 2023 e 2025 é de 45.630 novos casos (Inca, 2022). Já em termos de mortalidade, de acordo com o Instituto Nacional de Câncer José Alencar Gomes da Silva, em 2020, ocorreram 20.245 óbitos por câncer colorretal, sendo 9.889 óbitos entre homens e 10.356 entre mulheres (Inca, 2022).

O câncer colorretal é um crescimento desordenado de células no cólon (intestino grosso) ou no reto. A primeira etapa observada geralmente é a formação de pólipos no revestimento interno do cólon ou reto. Os pólipos são pequenos crescimentos de tecido anormal que se projetam para dentro do lúmen intestinal. Somente certos tipos de pólipos, chamados de adenomas, transformam-se em câncer no decorrer do tempo. Com o crescimento, os pólipos adenomatosos podem eventualmente invadir camadas mais profundas e, se entrarem na corrente sanguínea ou no sistema linfático, podem se espalhar em outras partes do corpo, podendo gerar tumores secundários (Fearon; Vogelstein, 1990).

A evolução para o câncer colorretal é um processo lento, e a detecção precoce e a intervenção médica podem salvar vidas, pois aumentam significativamente as chances de um tratamento bem-sucedido (Krishnendu; Geetha; Gopakumar, 2020). Por essa razão, é importante fazer o rastreamento dos pólipos por meio da realização constante de consultas médicas e exames, permitindo a remoção deles antes que se tornem malignos. Um dos exames é o de colonoscopia, que ajuda no diagnóstico e na detecção precoce do câncer colorretal, bem como na avaliação de pólipos e outras condições do cólon e reto. Durante o exame, um tubo flexível, chamado colonoscópio, com uma câmera na sua extremidade, é inserido no reto e avança lentamente pelo cólon. A câmera transmite imagens em tempo real para um monitor, permitindo que o médico examine o revestimento interno do cólon em busca de anomalias, como os pólipos (Levin; Winawer, 2008).

Detecção automática de pólipos utilizando Inteligência Artificial

A análise dos resultados de exames de colonoscopia é demorada e pode apresentar falhas de detecção de pólipos ou identificação de falsos pólipos, resultando em procedimentos incorretos de tratamento. A detecção automática dos pólipos com Inteligência Artificial (IA) surge como uma das abordagens para auxiliar nessa tarefa (Krishnendu; Geetha; Gopakumar, 2020).

A utilização da IA tem se destacado na área da saúde como uma ferramenta de apoio à tomada de decisões. Ela permite analisar grandes quantidades de dados médicos e identificar padrões complexos, fornecendo observações valiosas aos profissionais de saúde (Obermeyer; Emanuel, 2016). A aprendizagem profunda (*deep learning*) é um ramo da IA que se baseia em redes neurais profundas e tem a capacidade de extrair automaticamente características relevantes em imagens médicas. Dessa forma, estudos recentes têm surgido sobre o uso de modelos de redes neurais convolucionais para a detecção de pólipos em imagens de exames de colonoscopia (Elkarazle *et al.*, 2023).

A detecção de pólipo colorretal no contexto da IA corresponde ao processo de treinamento de um modelo de aprendizado de máquina (machine learning) para aprender um conjunto de características específicas que representem os pólipos e, assim, detectá-los por meio de uma caixa delimitadora (bounding box), a partir de uma imagem ou vídeo de entrada (Elkarazle *et al.*, 2023).

São várias as etapas envolvidas na construção de um sistema de detecção de pólipos usando IA. Primeiramente, a seleção de um conjunto de dados para treinamento e validação, em seguida um pré-processamento dos dados, depois aplicação de técnicas de aumento de dados para aumentar a variedade de imagens, seguido do treinamento do modelo e finalizando com a sua validação. Existem diversas formas de diversificar cada etapa, desde a escolha de parâmetros até a escolha do modelo a ser treinado (Lecun; Bengio; Hinton, 2015).

Além disso, há diversos modelos de detecção de objetos, de arquiteturas diferentes com parâmetros distintos e aplicados com objetivos variados, em conjunto de dados diferentes. Um modelo pode não ter o mesmo valor de resultados obtidos ao alterar o tipo de objeto a ser detectado. Dessa forma, o objetivo deste artigo é apresentar uma análise de modelos existentes, selecionando os mais atuais (estado da arte) para realizar suas implementações práticas, avaliar e validar a eficácia na detecção de pólipos em imagens de colonoscopia.

Equalização de Histograma

Histograma de imagem é uma representação gráfica de como estão distribuídas as intensidades dos pixels da imagem (Queiroz; Gomes, 2011). Por exemplo, o histograma de uma imagem na escala cinza mostra a quantidade de pixels de cada valor possível, ou seja, valores variando entre 0 (preto) e 255 (branco).

Em imagens coloridas,tem-se um histograma para cada canal de cor. Por exemplo, no espaço RGB (espaço de cores Red, Green e Blue), teria um histograma para vermelho (red), um para verde (green) e um para azul (blue). Outro tipo de espaço bastante conhecido é o YCrCb, sendo Y o brilho e Cr e Cb, os componentes cromáticos de diferença de azul e de diferença de vermelho (Gonzalez; Woods, 2007).

O histograma é uma ferramenta importante e muito utilizada na área de processamento de imagens, já que ele pode mostrar diversas informações, a exemplo do contraste da imagem obtida a partir da largura do histograma. Quanto mais ampla for a distribuição, maior será o contraste. Outra informação importante que se obtém analisando o histograma é a exposição: muitos pixels com intensidades próximas a 255 representam uma imagem superexposta, enquanto muitos pixels com intensidades próximas a 0 representam uma imagem subexposta (Thomaz, 2020).

A equalização do histograma é uma técnica bastante usada e consiste em redistribuir as intensidades dos pixels de forma mais uniforme durante o intervalo de valores disponíveis, com o objetivo de melhorar o contraste da imagem (Souza; Correia, 2007). Um exemplo é apresentado na Figura 1: a imagem à esquerda possui pixels com valores concentrados em uma pequena faixa, enquanto a imagem à direita apresenta um histograma com valores distribuídos ao longo de toda a faixa possível e, consequentemente, exibe um contraste mais bem distribuído no intervalo considerado.

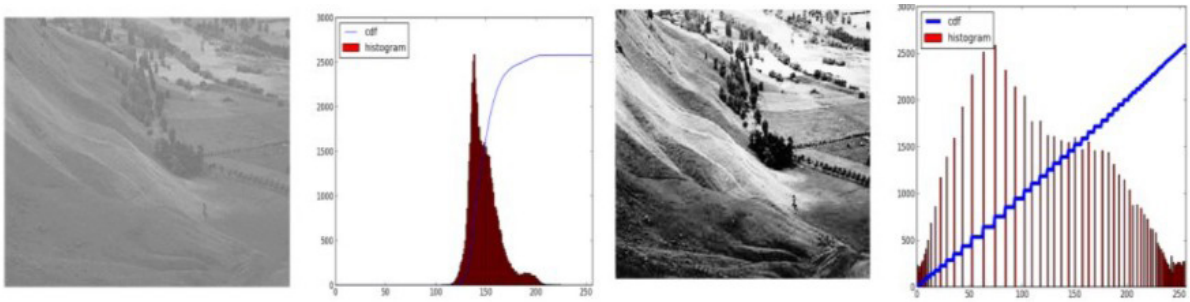

Figura 1 – À esquerda, tem-se uma imagem e seu histograma, com valores concentrados em uma pequena faixa do eixo x. À direita, observa-se a mesma imagem com histograma equalizado e, portanto, com melhor contraste, já que os valores no eixo x estão mais distribuídos no intervalo analisado
Fonte: Bradski e Kaehler (2000).

Organização do artigo

O artigo está organizado da seguinte forma: a Seção 2 apresenta a revisão bibliográfica de trabalhos relacionados; a metodologia incluindo coleta de dados, modelos de treinamento e inferências, equalização de histograma de imagens e as métricas utilizadas para análises é apresentada na Seção 3; os resultados obtidos são descritos na Seção 4; e a conclusão é apresentada na Seção 5.

Trabalhos relacionados

Grande parte dos estudos recentes sobre o uso de técnicas de inteligência artificial na área da saúde baseiam-se em métodos de aprendizagem profunda. Algoritmos de aprendizado profundo têm sido aplicados com sucesso em muitos campos da saúde, especialmente na análise de imagens médicas (Krishnendu; Geetha; Gopakumar, 2020). Dessa forma, esta seção apresenta alguns trabalhos desenvolvidos na literatura para realizar a tarefa de detecção de pólipos utilizando imagens de colonoscopia e algoritmos de detecção de objetos.

Region-based Convolutional Neural Network (R-CNN) é um algoritmo de detecção de objetos desenvolvido em 2014 (Zhang *et al.*, 2023).O R-CNN usa redes neurais convolucionais para localizar objetos de interesse e realizar a extração de características de forma independente de cada região de interesse para processamento posterior. Elkarazle *et al.* (2023) apresentaram bons resultados na detecção de pólipos utilizando modelos baseados no modelo R-CNN, detectando pólipos de formas e tamanhos distintos.

O modelo Single-Shot Detector (SSD) depende de uma única rede neural profunda para detectar um objeto (Zhang *et al.*, 2023). O modelo SSD discretiza o espaço de saída das caixas delimitadoras em várias caixas com diferentes proporções. Depois que as caixas delimitadoras são discretizadas, a rede é dimensionada por localização do mapa de características. A localização de um objeto de interesse é prevista usando vários mapas de características com diferentes resoluções. Os trabalhos em (Zhang *et al.*, 2023) propuseram utilizar o modelo SSD para realizar a detecção de pólipos colorretais.

A primeira versão do You Only Look Once (YOLO) foi apresentada em 2016 e é um dos algoritmos mais utilizados recentemente para a tarefa de detecção de objetos. O algoritmo YOLO foi modificado diversas vezes, gerando novas versões no decorrer

dos últimos anos. O YOLO também é um dos modelos de detecção de objetos que vem sendo utilizado na detecção de pólipos (Redmon; Farhadi, 2016).

O modelo Real-Time Detection Transformer (RT-DETR) foi apresentado em julho de 2023 (Zhang et al., 2023). Uma de suas características é o uso de transformers e consiste em usar um backbone, um codificador híbrido e um decodificador transformador com cabeçotes auxiliares de predição. Os autores apresentaram ainda os resultados que superaram os resultados com outros modelos, como a YOLO e suas diversas versões, utilizando o conjunto de dados da COCO 2017.

Zhang *et al.* (2023) também projetaram e treinaram uma Convolutional Neural Network (CNN) profunda para realizar a detecção de pólipos usando um conjunto diversificado e representativo de imagens de colonoscopias de triagem coletadas de mais de 2.000 pacientes. Foram treinadas diferentes arquiteturas CNN nesse estudo.

Além da escolha de um modelo para treinamento, é de suma importância considerar quais dados de treinamento serão utilizados. Como os bancos de dados usados na maioria dos trabalhos são conjuntos de dados limitados em tamanho e diversidade, Ma *et al.* (2021) propuseram um conjunto de dados em grande escala,chamado de LDPolypVideo, com uma variedade de pólipos e ambientes intestinais. Os testes realizados apresentaram quedas tanto na métrica recall quanto na precisão. De acordo com os autores, as quedas demonstram os desafios em trabalhar com conjuntos de dados de grande escala e com imagens de pólipos diversificados.

Nesse contexto, Tanwar *et al.* (2022) apresentaram um trabalho envolvendo técnicas de pré-processamento aplicadas em imagens de colonoscopia, tais como uso de filtros e de equalização de histograma, antes de realizarem um treinamento usando o modelo Single-Shot Detector (SSD). O modelo proposto pelos autores pode detectar e classificar pólipos colorretais com 92% de precisão.

Metodologia

Esta seção tem como objetivo descrever os métodos adotados na implementação prática dos modelos de detectores analisados. Os modelos implementados foram versões do YOLO e do modelo RT-DETR. Os experimentos foram conduzidos com três conjuntos de dados, realizando testes com e sem equalização de histograma.

A coleta de dados desempenha um papel fundamental em qualquer pesquisa baseada em aprendizado profundo, pois influencia diretamente na qualidade de resultados, mas sabe-se que há uma certa dificuldade para disponibilização de imagens médicas, por questões éticas, e em virtude da Lei Geral de Proteção de Dados Pessoais. Diante disso, foi realizado um levantamento de conjuntos de dados públicos (*datasets*) contendo imagens de colonoscopia.

Conjunto de dados

Após o levantamento, foi selecionado o conjunto de dados Kvasir-SEG, contendo 1.000 imagens de pólipos e suas anotações. A resolução das imagens varia de 332x487 a 1920x1072 pixels, e a divisão para treinamento e teste foi feita na proporção de 95/5 (Jha *et al.*, 2020).

Foi utilizado também o conjunto de dados CVC-ClinicDB (Bernal *et al.*, 2015), com 612 imagens com resolução de 384×288 pixels. Para o treinamento, validação e teste a divisão foi feita na proporção de 80/10/10.

Por fim, utilizou-se o conjunto de dados LDPolyp, que possui 40.186 imagens com resolução de 560×480 pixels, obtidas por meio de frames de vídeos (Ma *et al.*, 2021). Este último teve a divisão baseada na separação dos vídeos, sendo 100 vídeos para o conjunto de treinamento e 60 vídeos para o de teste e a validação.

Entre os conjuntos de dados escolhidos, os datasets Kvasir e o CVC-ClinicDB continham originalmente anotações de segmentação. Dessa forma, foi necessário realizar a conversão das anotações para o formato de detecção de objetos.

Modelos

Em sua primeira versão, a YOLO apresentou uma abordagem inovadora com mais velocidade em relação a outros modelos da época, como a Fast-RCNN, Faster-RCNN e DPM. Além disso, a velocidade de inferência era consideravelmente mais rápida, sendo possível realizar a inferência em tempo real (Redmon *et al.*, 2016).

A YOLO versão 3 apresenta como uma de suas principais melhorias: novo modelo de backbone (DarkNet53), com mais camadas convolucionais; adição de conexões residuais; previsões em múltiplas escalas, melhorando a detecção de objetos pequenos; e a possibilidade de múltiplas classes em uma mesma caixa delimitadora, em razão da troca da função softmax para entropia binária cruzada(Redmon; Farhadi, 2018).

A YOLOv5 foi o primeiro modelo a utilizar a biblioteca Pytorch, sendo disponibilizado pela Ultralytics em 2020 (Terven; Cordova-Esparza, 2023; Jocher *et al.*, 2022). Suas melhorias mais relevantes foram: um algoritmo de âncora automática, que aplica um algoritmo k-means nos dados e evolui as posições das caixas predefinidas para serem usadas nas predições; o modelo de backbone utilizado foi a CSPDarkNet, uma versão modificada que começa com uma camada de convolução ampliada, reduzindo custos computacionais; uma série de transformações para aumento de dados, que melhora o aprendizado; há também uma camada SPPF (Spatial Pyramid Pooling Fast), que acelera o cálculo agrupando características de diferentes escalas em um mapa de características de tamanho fixo.

A versão YOLOv6 foi publicada em 2022 pelo Meituan Vision AI Department, e superou os modelos anteriores em métricas de precisão e velocidade (Terven; Cordova-Esparza, 2023; Li *et al.*, 2022). As novidades destacadas dessa versão foram: um *backbone* baseado em RepVGG, chamado EfficientRep, que usa maior paralelismo; e uma estratégia de auto-destilação (*self-distillation*), que permite transferir o aprendizado de camadas profundas para camadas superficiais por meio de módulos de atenção (Ding *et al.*, 2021).

Assim como a versão 6, a YOLOv7 superou os detectores da época em precisão e velocidade na faixa de 5 FPS a 160 FPS (Terven; Cordova-Esparza, 2023; Wang; Bochkovskiy; Liao, 2022). Sua precisão foi incrementada sem afetar a velocidade de inferência, sobretudo em virtude das suas inovações: o uso da estratégia E-ELAN (Extended Efficient Layer Aggregation Network), que permite que um modelo profundo aprenda e convirja de forma mais eficiente, controlando o caminho de gradiente; normalização dos lotes (*batch normalization*) durante a convolução, que regula as ativações intermediárias, melhorando a estabilidade e o desempenho geral da rede (Terven; Cordova-Esparza, 2023).

A versão YOLOv8 foi lançada em janeiro de 2023, é considerada um modelo de última geração para tarefas de detecção de objetos (Terven; Cordova-Esparza, 2023). Ela deriva diretamente da YOLOv5 e também foi disponibilizada pela Ultralytics. Quanto a suas melhorias, citam-se: alteração no backbone (CSPDarkNet), permitindo combinar recursos na convolução, melhorando, consequentemente, a precisão;

funções de ativação separadas para objetividade, probabilidade de a caixa delimitadora conter um objeto, e classificação, probabilidade de o objeto ser de determinada classe, utilizando a função sigmóide para a primeira tarefa e a função softmax para a segunda tarefa; as funções de perda (*loss function*) de caixa delimitadora e de classificação também foram separadas, o que auxilia na detecção de objetos pequenos (Terven; Cordova-Esparza, 2023).

O modelo RT-DETR foi lançado em julho de 2023, e apresenta como principal característica o uso de *transformers*. Mais detalhadamente, o modelo utiliza um codificador híbrido eficiente para processar recursos em várias escalas e propõe uma seleção de consultas com reconhecimento de IoU (Intersection over Union), fornecendo consultas de objetos iniciais de maior qualidade ao decodificador (Zhao *et al.*, 2023).

Para fins de comparação de resultados, os seguintes modelos pré-treinados, realizando a transferência de aprendizado (*transfer learning*) no conjunto de dados Kvasir- SEG: YOLOv5 (modelos Nano, Small, Medium, Large e Extra-large); YOLOv6 (modelos Nano, Small, Medium e Large); YOLOv7 (modelos Tiny e X); YOLOv8 (modelos Nano, Small, Medium, Large e Extra-large); e RT-DETR (modelos Large e Extra-large).

Todos os modelos foram treinados no Google Colab, utilizando o GPU (Graphic Processing Unit) NVIDIA Tesla T4, exceto o treinamento, aplicando o modelo RT-DETRX. Em razão das restrições de memória RAM da GPU do Google Colab, foi utilizada uma GPU Tesla V100 da DGX-1 NVIDIA para realizar o treinamento com o dataset LDPolyp completo com o modelo RT-DETR-X. Os parâmetros dos treinamentos utilizados foram de 300 épocas, utilizando um batch size, com variação de 8 a 32. Os resultados obtidos serão discutidos na Seção 4.

Equalização de histograma

No contexto desta pesquisa, é comum que as imagens de colonoscopia obtidas em exames apresentem baixo contraste em virtude das condições de iluminação, como sombras e reflexos. Isso ocorre porque o colonoscópio, o equipamento utilizado, é inserido no intestino, o que pode dificultar a obtenção de imagens de alta qualidade.

A equalização de histograma aplicada em imagens de colonoscopia pode ajudar a realçar o contraste, consequentemente destacando melhor as características da mucosa intestinal, como vasos sanguíneos e pólipos. As imagens resultantes da aplicação da equalização de histograma podem, então, ser utilizadas para aprimorar a detecção de pólipos (Thomaz, 2020).

Os testes de equalização de histograma foram conduzidos utilizando imagens do dataset LDPolyp. Foi realizada a equalização de histograma no espaço YCrCb, sendo cada imagem foi convertida do espaço RGB para YCrCb, e a equalização do histograma foi aplicada apenas na banda Y – que representa o brilho da imagem – e, finalmente, a imagem foi convertida de volta para o espaço RGB. Na Figura 2, é possível ver a imagem antes da aplicação da equalização do histograma e após a aplicação da técnica.

Com o objetivo de analisar a equalização de histograma, utilizou-se uma amostra do dataset LDPolyp (1.000 imagens avaliadas) e o dataset completo Kvasir-SEG. Para o dataset LDPolyp, a análise foi realizada com o treinamento do modelo RT-DETR-X e para o dataset Kvasir-SEG a análise foi realizada com o treinamento dos modelos YOLOv8-N, YOLOv8-S, YOLOv8-M, YOLOv8-L, YOLOv8-X e RT-DETR-X.

Métricas de Avaliação

As métricas utilizadas para comparar os resultados entre os modelos de treinamento RT-DETR e YOLO nas versões 5 a 8 foram: Precisão, Recall, IoU, mAP@50 e mAP@50-95.

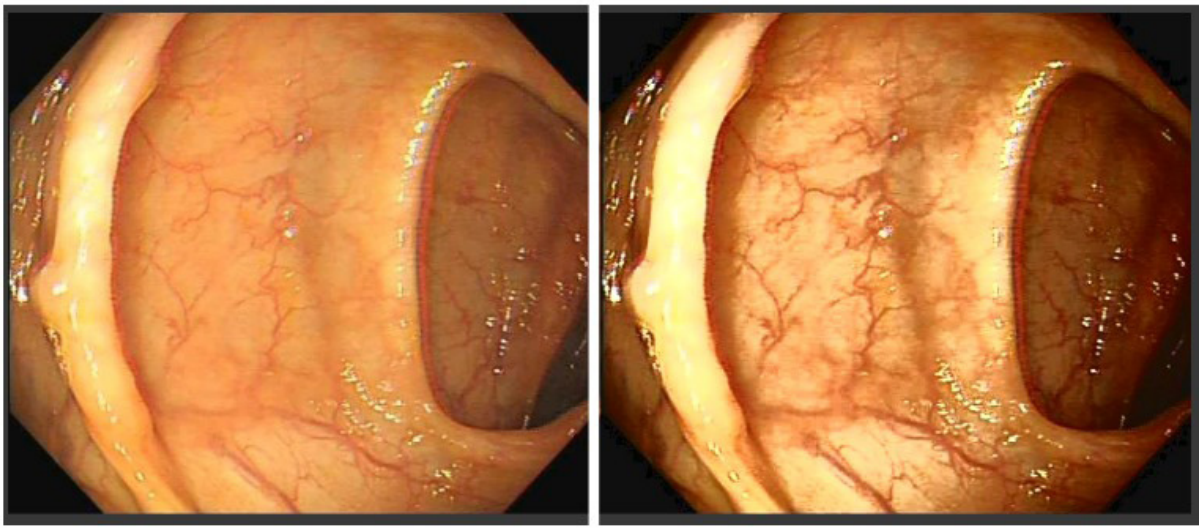

Figura 2 – A imagem à esquerda é a imagem original do dataset LDPolyp, e à direita observa-se a mesma imagem após a equalização do histograma da banda Y (brilho da imagem) no espaço YCrCb
Fonte: Elaborado pelos(as) autores(as).

IoU (Intersection over Union) é a métrica que representa a interseção sobre a união, comparando as caixas delimitadoras anotadas e previstas. A equação 1 representa o IoU, sendo A a caixa delimitadora verdadeira e B a caixa delimitadora prevista.

$$IoU(A,B) = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

Dado um valor limite de IoU, um resultado de previsão de modelo pode ser considerado como: Verdadeiro Positivo (VP) - classificação certa da classe positivo; Falso Negativo (FN) - modelo previu como classe negativo, sendo o valor real a classe positivo; Falso Positivo (FP) - modelo previu como classe positivo, sendo o valor real a classe negativo; ou Verdadeiro Negativo (VN) - classificação certa da classe negativo.

A Precisão corresponde à métrica que avaliará, dentre todas as classificações de classe positivo que o modelo fez, quantas de fato têm esse valor real. Ela é representada pela equação 2.

$$Preciso = \frac{VP}{VP + FP} \quad (2)$$

Recall é a métrica que verificará quantas classificações corretas foram feitas da classe positivo em relação ao total esperado, ou seja, em relação a todas as situações de classe positivo como valor esperado. Essa métrica está representada na equação 3.

$$Recall = \frac{VP}{VP + FN} \quad (3)$$

A métrica Average Precision (AP) representa a área debaixo da curva de um gráfico das métricas Precisão e Recall, conhecida também como curva AP.

O mAP@50 (Mean Average Precision) é a média da AP dado um limiar de sobreposição de pelo menos 50% de IoU e o mAP@50-95 é a média dos valores de AP para cada limiar de sobreposição no intervalo de 50% a 95%, com incrementos de 5%.

Resultados

Foram realizados primeiramente experimentos sem equalização de histograma. Após isso, realizaram-se experimentos com os modelos que obtiveram melhores valores de métricas utilizando a técnica de equalização de histograma, visando proporcionar um ganho no desempenho dos modelos. A seguir, são apresentados os resultados obtidos em ambos os experimentos, seguindo a ordem em que foram realizados.

Resultados sem aplicação de Equalização de Histograma

No processo de treinamento no conjunto de dados Kvasir-SEG (Jha *et al.*, 2020), foram registradas as métricas exibidas na Tabela 1 para os modelos YOLO nas versões 5 a 8 e RT-DETR.

Modelo	Precisão	Recall	mAP@50	mAP@50-95
YOLOv5-N	0,884	0,830	0,910	0,726
YOLOv5-S	0,923	0,871	0,913	0,743
YOLOv5-M	0,889	0,872	0,916	0,744
YOLOv5-L	0,889	0,927	0,932	0,755
YOLOv5-X	0,918	0,815	0,906	0,725
YOLOv6-N	0,872	0,866	0,917	0,707
YOLOv6-S	0,840	0,873	0,897	0,718
YOLOv6-M	0,854	0,850	0,887	0,673
YOLOv6-M	0,815	0,855	0,877	0,650
YOLOv7-Tiny	0,896	0,782	0877	0,604
YOLOv7	0,877	0,909	0,917	0,720
YOLOv7-X	0,938	0,836	0,915	0,692
YOLOv8-N	0,873	0,875	0,923	0,723
YOLOv8-S	0,876	0,903	0,931	0,751
YOLOv8-M	0,922	0,855	0,931	0,757
YOLOv8-L	0,920	0,873	0,930	0,774
YOLOv8-X	0,887	0,891	0,915	0,734
RT-DETR L	0,882	0,855	0,885	0,736
RT-DETR X	0,905	0,855	0,872	0,737

Tabela 1 - Métricas de desempenho dos modelos treinados para o conjunto de dados Kvasir (os melhores resultados de cada métrica estão sublinhados)
Fonte: Elaborado pelos(as) autores(as).

Os modelos com melhores métricas estão reunidos na Tabela 2. O critério de seleção foi o map@50-95. Considerando esse critério, destaca-se o modelo YOLOv8L.

Modelo	Precisão	Recall	mAP@50	mAP@50-95
YOLOv8-N	0,873	0,875	0,923	0,723
YOLOv8-S	0,876	0,903	0,931	0,751
YOLOv8-M	0,922	0,855	0,931	0,757
YOLOv8-L	0,920	0,873	0,930	0,774
YOLOv8-X	0,887	0,891	0,915	0,734

Tabela 2 - Desempenho dos melhores modelos treinados para o conjunto de dados Kvasir (os melhores resultados de cada métrica estão sublinhados)
Fonte: Elaborado pelos(as) autores(as).

A Figura 3 exibe algumas imagens de pólipos detectados pelo modelo que obteve as melhores métricas, isto é, o modelo Yolov8L, bem como os níveis de precisão associados a cada detecção. Pode-se observar uma compreensão visual e quantitativa do quão preciso e confiável é o modelo na tarefa de detecção de pólipos em imagens de colonoscopia.

Após os resultados obtidos, também foi realizada uma análise comparativa, investigando na literatura os resultados de outros modelos que utilizaram o mesmo conjunto de dados. Essa análise foi realizada considerando a metodologia apresentada em Jha *et al.* (2020). A comparação está ilustrada na Tabela 3, e a métrica avaliada foi o Mean Average Precision 50 (mAP@50).

Método	Rede Principal	mAP@50
EfficientDet-D0	EfficientNet-b0, biFPN	0,5047
Faster R-CNN	ResNet50	0,8418
RetinaNet	ResNet50	0,9095
RetinaNet	ResNet101	0,9095
YOLOv3+spp	Darknet53	0,8532
YOLOv4	Darknet53, CSP	0,8234
ColonSegNet	-	0,8166
YOLOv8-L (proposto)	-	0,9300

Tabela 3 - Métricas de desempenho dos modelos obtidos na literatura e do melhor modelo treinado (YOLOv8-L) para o conjunto de dados Kvasir
Fonte: Elaborado pelos(as) autores(as).

Além do conjunto de dados Kvasir, também foram feitos treinamentos com os melhores modelos para o conjunto de dados CVC Clinic, cujos resultados são exibidos na Tabela 4.

Para o CVC Clinic, observa-se que os modelos YOLOv8-N e YOLOv8-M exibem boas pontuações de Precisão, Recall e mAP@50, aproximando-se de 1, o que evidencia uma capacidade de identificação precisa dos objetos de interesse nas imagens. Esses resultados sustentam a confiabilidade e robustez desses modelos na tarefa de detecção de objetos.

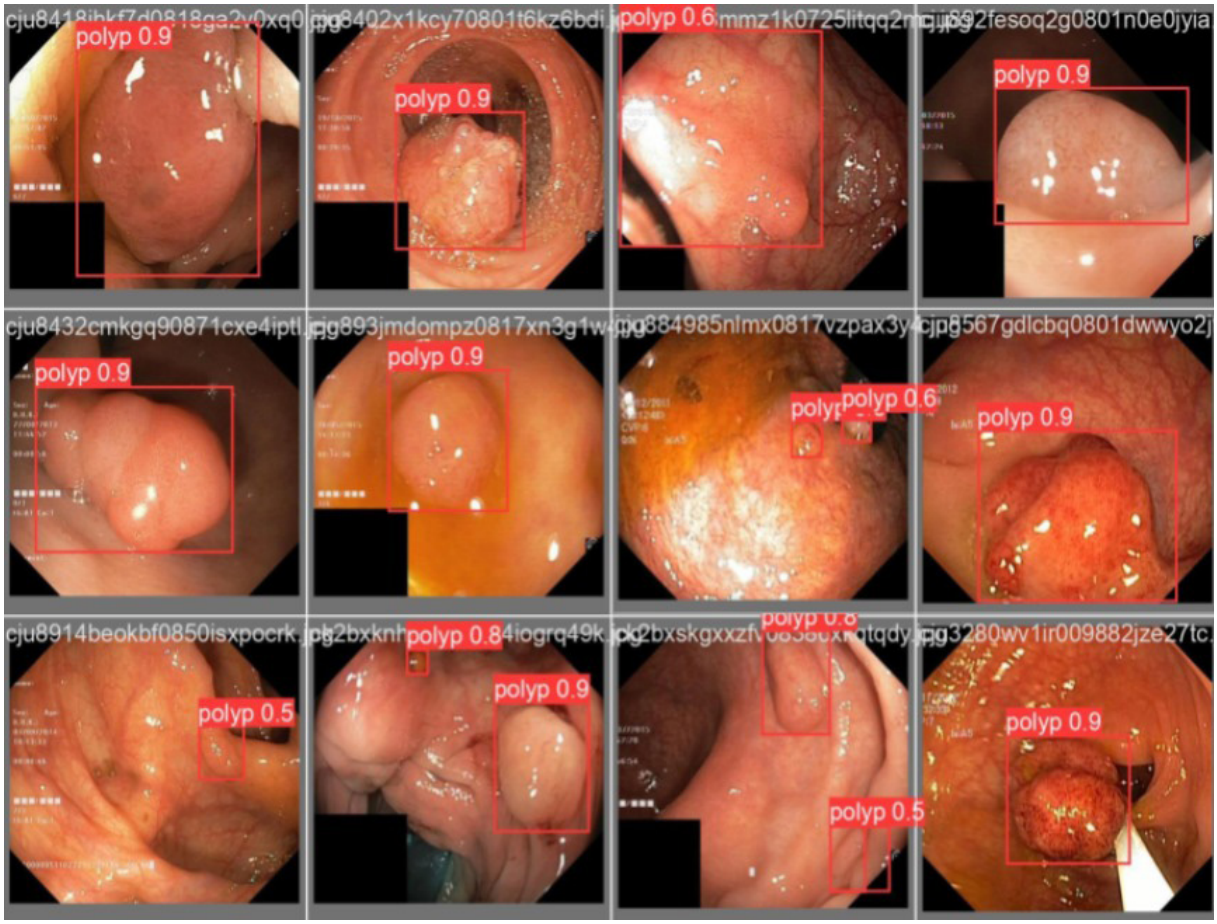

Figura 3 – Pólipos detectados na validação do modelo YOLOv8L
Fonte: Elaborado pelos(as) autores(as).

Modelo	Precisão	Recall	mAP@50	mAP@50-95
YOLOv8-N	0,996	1,000	0,995	0,857
YOLOv8-M	0,983	1,000	0,995	0,856
YOLOv8-L	0,984	0,998	0,995	0,842
YOLOv9-C	0,969	1,000	0,993	0,848
RT-DETR-L	0,998	1,000	0,995	0,845

Tabela 4 - Métricas de desempenho dos modelos validados para o CVC Clinic (os melhores resultados de cada métrica estão sublinhados)
Fonte: Elaborado pelos(as) autores(as).

Em contrapartida, o desempenho do modelo YOLOv8-L também se destaca, embora se posicione abaixo dos modelos YOLOv8-N e YOLOv8-M em termos de precisão e mAP@50-95. No entanto, sua pontuação de Recall, próximo a 1, sugere uma alta habilidade em recuperar todos os objetos relevantes nas imagens, tornando-o uma escolha sólida para aplicações que demandam mais sensibilidade na detecção.

O modelo YOLOv9-C apresenta resultados competitivos, apesar de uma precisão inferior em comparação com os modelos YOLOv8. No entanto, suas métricas de Recall e mAP@50-95 indicam que ainda é capaz de identificar a maioria dos objetos com uma precisão aceitável.

Por fim, pode-se observar que o modelo RT-DETR-X se destaca por sua precisão excepcionalmente alta, indicando uma habilidade quase perfeita em evitar falsos positivos

para esse conjunto de dados. Contudo, sua pontuação de mAP@50-95 é ligeiramente menor em comparação com os modelos YOLOv8, sugerindo uma precisão um pouco reduzida na detecção de objetos mais desafiadores de localizar.

Para o conjunto de dados LDPolyp, em virtude da sua grande quantidade de imagens, o treinamento foi realizado apenas do modelo RT-DETR, cujos resultados estão apresentados na Tabela 5.

Modelo	Precisão	Recall	mAP@50	mAP@50-95
RT-DETR-X	0,1339	0,7428	0,3875	0,4163

Tabela 5 - Desempenho do RTDETR-X treinado com o conjunto de dados LDPolyp
Fonte: Elaborado pelos(as) autores(as).

Resultados com aplicação de Equalização de Histograma

Em seguida, foram realizados os testes com equalização de histograma, a fim de verificar o ganho nas métricas de desempenho dos modelos ao utilizar essa técnica. Na Tabela 6, pode-se observar a comparação das métricas dos melhores modelos antes e depois da aplicação da técnica de equalização do histograma no conjunto de dados Kvasir.

A análise comparativa dos modelos treinados para o conjunto de dados Kvasir, tanto antes quanto após a equalização de histograma, revela descobertas que enfatizam os benefícios dessa técnica de pré-processamento para alguns dos modelos.

Um dos aspectos mais notáveis é a melhoria na métrica de precisão do modelo RTDETR- X após a aplicação da equalização. Isso significa que o modelo está conseguindo detectar corretamente as áreas de interesse nas imagens, reduzindo, assim, o número de falsos positivos. Essa precisão aprimorada é vital, especialmente em tarefas críticas como a detecção de anomalias, onde cada erro pode ter consequências significativas.

No entanto, é importante ressaltar que a implementação da equalização de histograma não é uma solução universal. Embora tenha apresentado melhorias significativas em termos de precisão, em alguns casos, pode haver uma leve redução no Recall ou em outras métricas de desempenho. Portanto, é necessário realizar uma avaliação abrangente e equilibrada dos resultados, considerando não apenas métricas isoladas, mas também o impacto geral no desempenho do modelo.

Modelo	Equalização de Histograma	Precisão	Recall	mAP@50	mAP@50-95
YOLOv8-N	Não	0,873	0,875	0,923	0,723
YOLOv8-N	SIm	0,873	0,855	0,915	0,747
YOLOv8-S	Não	0,876	0,903	0,931	0,751
YOLOv8-S	Sim	0,888	0,863	0,896	0,736
YOLOv8-M	Não	0,922	0,855	0,931	0,757
YOLOv8-M	Sim	0,848	0,909	0,910	0,724
YOLOv8-L	Não	0,920	0,873	0,930	0,774
YOLOv8-L	Sim	0,854	0,873	0,895	0,700
YOLOv8-X	Não	0,887	0,891	0,915	0,734
YOLOv8-X	Sim	0,884	0,830	0,904	0,720

RT-DETR-X	Não	0,905	0,855	0,872	0,737
RT-DETR-X	Sim	0,958	0,873	0,917	0,762

Tabela 6 - Desempenho dos modelos treinados para o conjunto de dados Kvasir antes e após a equalização de histograma (os melhores resultados de cada métrica estão sublinhados)
Fonte: Elaborado pelos(as) autores(as).

Com o objetivo de avaliar a utilização da técnica de equalização de histogramas em imagens de vídeos, utilizou-se uma parcela de 1.000 amostras do conjunto de dados LDPolyp. A Tabela 7 apresenta a comparação das métricas antes e depois da aplicação da referida técnica em uma amostra de 1.000 imagens do conjunto de dados LDPolyp.

Modelo	Equalização de Histograma	Precisão	Recall	mAP@50	mAP@50-95
RT-DETR-X	Não	0,757	0,417	0,425	0,193
RT-DETR-X	Sim	0,857	0,500	0,516	0,238

Tabela 7 - Desempenho do RTDETR-X treinado com o conjunto de dados LDPolyp (os melhores resultados de cada métrica estão sublinhados)
Fonte: Elaborado pelos(as) autores(as).

Nota-se, observando a Tabela 7, a melhoria em todas as métricas avaliadas do modelo RT-DETR-X após a equalização de histograma na amostra de 1.000 imagens do conjunto de dados LDPolyp. A precisão registrou um aumento de aproximadamente 20%, saltando de 0,757 para 0,857. Da mesma forma o Recall, com cerca de 19%, passando de 0,417 para 0,500. O mAP@50 e o mAP@50-95 registraram melhorias de aproximadamente 21,4% e 23,3% respectivamente, refletindo um aumento na precisão média e na detecção em uma variedade mais ampla de cenários e condições. Isso sugere um progresso na capacidade do RT-DETR-X de identificar objetos com maior precisão e confiança após a aplicação da equalização de histograma.

Conclusão

Este estudo abordou a detecção precoce de pólipos colorretais por meio do uso de IA, com um foco significativo na metodologia adotada. A seleção criteriosa de conjuntos de dados públicos, compostos por uma variedade de imagens de colonoscopia, foi essencial para garantir a representatividade da amostra.

Os treinamentos dos modelos YOLO (versões 5 até a 8) e RT-DETR (Large e Extra-Large) foram conduzidos de maneira sistemática e controlada, com ajustes de hiperparâmetros para otimizar o desempenho. A análise comparativa com outros modelos, como o RetinaNet, destaca a superioridade dos modelos YOLO em termos de detecção de pólipos (Lin *et al.*, 2017).

Além disso, foram feitos testes com a aplicação da equalização de histograma para melhorar o contraste das imagens e verificar se, consequentemente, haveria melhora no desempenho dos modelos. Os resultados dos testes indicaram que a equalização de histograma teve um impacto positivo no desempenho dos modelos, melhorando ligeiramente as métricas. Contudo, essa técnica não foi um fator decisivo para o desempenho final dos modelos, sugerindo que outras técnicas de pré-processamento podem ser exploradas em futuros trabalhos.

Referências

AMERICAN CANCER SOCIETY. *Cancer Facts & Figures 2020*. [*S. l.*]: ACS, 2020.

BERNAL, J.; SÁNCHEZ, F. J.; FERNÁNDEZ-ESPARRACH, G.; GIL, D.; RODRÍGUEZ, C.; VILARIñO, F. Wm-dova maps for accurate polyp highlighting in colonoscopy: validation vs. saliency maps from physicians. *Computerized Medical Imaging and Graphics*, [*s. l.*], v. 43, p. 99-111, 2015.

BRADSKI, G.; KAEHLER, A. OpenCV. *Dr. Dobb's journal of software tools*, [*s. l.*], v. 3, n. 2, 2000.

DING, X.; ZHANG, X.; MA, N.; HAN, J.; DING, G.; SUN, J. Repvgg: Making vgg- style convnets great again. *IEEE*, [*s. l.*], 2021. Disponível em: https://ieeexplore.ieee.org/document/9577516. Acesso em: 9 set. 2025.

ELKARAZLE, K.; RAMAN, V.; THEN, P.; CHUA, C. Detection of colorectal polyps from colonoscopy using machine learning: A survey on modern techniques. *National Library of Medicine*, [*s. l.*], 2023. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9953705/. Acesso em: 9 set. 2025.

FEARON, E. R.; VOGELSTEIN, B. A genetic model for colorectal tumorigenesis. *National Library of Medicine*, [*s. l.*], 1990. Disponível em: https://pubmed.ncbi.nlm.nih.gov/2188735/. Acesso em: 9 set. 2025.

GONZALEZ, R. C.; WOODS, R. E. *Processamento Digital de Imagens*. 5. ed. São Paulo: Pearson Education do Brasil, 2007.

IARC GLOBAL CANCER OBSERVATORY. *Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries*. [*S. l.*]: IARC Global Cancer Observatory, 2020. Disponível em: https://pubmed.ncbi.nlm.nih.gov/33538338/. Acesso em: 9 set. 2025.

INSTITUTO NACIONAL DE CÂNCER. *Estimativa 2023*: incidência de câncer no Brasil. Rio de Janeiro: Inca, 2022. Disponível em: https://www.inca.gov.br/publicacoes/livros/estimativa-2023-incidencia-de-cancer-nobrasil. Acesso em: 9 set. 2025.

JHA, D.; SMEDSRUD, P. H.; RIEGLER, M. A.; HALVORSEN, P.; LANGE, T.; JOHANSEN, D.; JOHANSEN, H. D. Kvasir-seg: A segmented polyp dataset. *arXiv*, [*s. l.*], 2020. Disponível em: https://arxiv.org/abs/1911.07069. Acesso em: 9 set. 2025.

JOCHER, G.; CHAURASIA, A.; STOKEN, A.; BOROVEC, J. ultralytics/yolov5: v7.0 yolov5 sota realtime instance segmentation. *Zenodo*, [*s. l.*], 2022. DOI: https://doi.org/10.5281/zenodo.7347926.

KRISHNENDU, S.; GEETHA, S.; GOPAKUMAR, G. A review on polyp detection and segmentation in colonoscopy images using deep learning. *International Journal of Engineering Research & Technology (IJERT)*, [*s. l.*], v. 9, n. 10, 2020. Disponível

em: https://www.ijert.org/a-review-on-polyp-detection-and-segmentation-in-colonoscopyimages-using-deep-learning. Acesso em: 9 set. 2025.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *nature*, [*s. l.*], 2015. DOI: https://doi.org/10.1038/nature14539.

LEVIN, B.; LIEBERMAN, D. A.; MCFARLAND, B.; ANDREWS, K. S. Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the american cancer society, the us multi-society task force on colorectal cancer, and the american college of radiology. *National Library of Medicine*, [*s. l.*], 2008. Disponível em: https://pubmed.ncbi.nlm.nih.gov/18384785/. Acesso em: 9 set. 2025.

LI, C.; LI, L.; JIANG, H.; WENG, K.; GENG, Y.; LI, L.; KE, Z.; LI, Q.; CHENG, M.; NIE, W.; LI, Y.; ZHANG, B.; LIANG, Y.; ZHOU, L.; XU, X.; CHU, X.; WEI, X.; WEI, X. Yolov6: A single-stage object detection framework for industrial applications. *arXiv*, [*s. l.*], 2022. Disponível em: https://arxiv.org/abs/2209.02976. Acesso em: 9 set. 2025.

LIN, T.-Y.; GOYAL, P.; GIRSHICK, R.; HE, K.; DOLLÁR, P. Focal loss for dense object detection. *In*: PROCEEDINGS of the IEEE International Conference on Computer Vision. [*S. l.: s. n.*], 2017. p. 2980-2988.

MA, Y.; CHEN, X.; CHENG, K.; LI, Y.; SUN, B. Ldpolypvideo benchmark: A largescale colonoscopy video dataset of diverse polyps. *Medical Image Computing and Computer Assisted Intervention Society*, [*s. l.*], 2021, p. 387-396.

OBERMEYER, Z.; EMANUEL, E. J. Predicting the future - big data, machine learning, and clinical medicine. *National Library of Medicine*, [*s. l.*], 2016. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5070532/. Acesso em: 9 set. 2025.

QUEIROZ, J. E. R.; GOMES, H. M. *Introdução ao Processamento Digital de Imagens*. Campina Grande: UFCG, 2011. Disponível em: http://www.dsc.ufcg.edu.br/~hmg/disciplinas/graduacao/vc-2011.2/RitaTutorialPDI.pdf. Acesso em: 9 set. 2025.

REDMON, J.; DIVVALA, S.; GIRSHICK, R.; FARHADI, A. You only look once: Unified, real-time object detection. *arXiv*, [*s. l.*], 2016. Disponível em: https://arxiv.org/abs/1506.02640. Acesso em: 9 set. 2025.

REDMON, J.; FARHADI, A. Yolov3: An incremental improvement. *arXiv*, [*s. l.*], 2018. Disponível em: https://arxiv.org/abs/1804.02767. Acesso em: 9 set. 2025.

SOUZA, T.; CORREIA, S. Estudo de técnicas de realce de imagens digitais e suas aplicações. *In*: II CONGRESSO DE PESQUISA E INOVAÇÃO DA REDE NORTE NORDESTE DE EDUCAÇÃO TECNOLÓGICA, 2007, Paraíba. *Anais* […]. Paraíba: Connepi, 2007.

TANWAR, S.; VIJAYALAKSHMI, S.; SABHARWAL, M.; KAUR, M.; ALZUBI, A. A.; LEE, H.-N. Detecção e classificação de pólipo colorretal usando aprendizado profundo. *BioMed Research International*, [*s. l.*], v. 2024, n. 1, 2022. Disponível em: https://onlinelibrary.wiley.com/doi/full/10.1155/2022/2805607. Acesso em: 9 set. 2025.

TERVEN, J.; CORDOVA-ESPARZA, D. A comprehensive review of yolo: From yolov1 and beyond. *arXiv*, [*s. l.*], 2023. Disponível em: https://arxiv.org/abs/2304.00501. Acesso em: 9 set. 2025.

THOMAZ, V. A. *Avaliação de Aumento de Dados via Geração de Imagens Sintéticas para Segmentação e Detecção de Pólipos em Imagens de Colonoscopia Utilizando Aprendizado de Máquina*. 2020. Tese (Doutorado em Ciências da Computação) — Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 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. *arXiv*, [*s. l.*], 2022. Disponível em: https://arxiv.org/abs/2207.02696. Acesso em: 9 set. 2025.

ZHANG, A.; LIPTON, Z. C.; LI, M.; SMOLA, A. J. Dive into Deep Learning. *GitHub*, [*s. l.*], 2023. Disponível em: http://d2l.ai/index.html. Acesso em: 10 set. 2025.

ZHAO, Y.; LV, W.; XU, S.; WEI, J.; WANG, G.; DANG, Q.; LIU, Y.; CHEN, J. Detrs beat yolos on real-time object detection. *arXiv*, [*s. l.*], 2023. Disponível em: https://arxiv.org/abs/2304.08069. Acesso em: 9 set. 2025.