Pontifícia Universidade Católica
do Rio de Janeiro

Victor de Almeida Thomaz

Avaliação de Aumento de Dados via Geração de Imagens Sintéticas para Segmentação e Detecção de Pólipos em Imagens de Colonoscopia Utilizando Aprendizado de Máquina

Tese de Doutorado

Tese apresentada como requisito parcial para obtenção do grau de Doutor pelo Programa de Pós–graduação em Informática do Departamento de Informática da PUC-Rio.

Orientador: Prof. Alberto Barbosa Raposo

Rio de Janeiro
Abril de 2020

Pontifícia Universidade Católica
do Rio de Janeiro

Victor de Almeida Thomaz

Avaliação de Aumento de Dados via Geração de Imagens Sintéticas para Segmentação e Detecção de Pólipos em Imagens de Colonoscopia Utilizando Aprendizado de Máquina

Tese apresentada como requisito parcial para obtenção do grau de Doutor pelo Programa de Pós–graduação em Informática do Departamento de Informática da PUC-Rio. Aprovada pela Comissão Examinadora abaixo.

Prof. Alberto Barbosa Raposo
Orientador
Departamento de Informática – PUC-Rio

Prof. Helio Côrtes Vieira Lopes
Departamento de Informática – PUC-Rio

Cesar Augusto Sierra Franco
Departamento de Informática – PUC-Rio

Prof. Aristófanes Corrêa Silva
Departamento de Engenharia de Eletricidade – UFMA

Prof. Jauvane Cavalcante de Oliveira
Coord. de Métodos Matemáticos e Computacionais – LNCC

Rio de Janeiro, 27 de Abril de 2020

Victor de Almeida Thomaz

Graduou-se em Tecnologia da Informação pelo Instituto Superior de Tecnologia de Petrópolis (2009) e obteve o grau de Mestre em Sistemas e Computação pelo Instituto Militar de Engenharia - IME (2012).

Ficha Catalográfica

Thomaz, Victor de Almeida

Avaliação de Aumento de Dados via Geração de Imagens Sintéticas para Segmentação e Detecção de Pólipos em Imagens de Colonoscopia Utilizando Aprendizado de Máquina / Victor de Almeida Thomaz; orientador: Alberto Barbosa Raposo. – 2020.

126 f.: il. color. ; 30 cm

Tese (doutorado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática, 2020.

Inclui bibliografia

1. Informática – Teses. 2. Dados de Treinamento;. 3. Aumento de Dados;. 4. Redes Neurais Convolucionais;. 5. Pólipos;. 6. Colonoscopia;. I. Raposo, Alberto Barbosa. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD: 004

Agradecimentos

Agradeço primeiramente a Deus por me ajudar nesta trajetória.

À minha esposa Eliana pela compreensão, apoio e paciência durante todo o curso do doutorado.

Aos meus pais, familiares e amigos pelo incentivo.

Aos professores da PUC-Rio, em especial meu orientador prof. Alberto Raposo por me atender e direcionar nas etapas desta pesquisa.

Ao prof. Jauvane (Lab. A.C.i.M.A - LNCC) por me ceder o uso dos equipamentos para desenvolvimento e testes.

Ao Cesar Franco pelas valiosas sugestões.

Por último, mas não menos importante, gostaria de agradecer ao CNPq pelo apoio financeiro e à PUC-Rio pela bolsa de isenção de mensalidades do doutorado.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Resumo

Thomaz, Victor de Almeida; Raposo, Alberto Barbosa. **Avaliação de Aumento de Dados via Geração de Imagens Sintéticas para Segmentação e Detecção de Pólipos em Imagens de Colonoscopia Utilizando Aprendizado de Máquina**. Rio de Janeiro, 2020. 126p. Tese de Doutorado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

O câncer de cólon é atualmente a segunda principal causa de morte por câncer no mundo. Nos últimos anos houve um aumento do interesse em pesquisas voltadas para o desenvolvimento de métodos automáticos para detecção de pólipos e os resultados mais relevantes foram alcançados por meio de técnicas de aprendizado profundo. No entanto, o desempenho destas abordagens está fortemente associado ao uso de grandes e variados conjuntos de dados. Amostras de imagens de colonoscopia estão disponíveis publicamente, porém a quantidade e a variação limitada podem ser insuficientes para um treinamento bem-sucedido. O trabalho de pesquisa desta tese propõe uma estratégia para aumentar a quantidade e variação de imagens de colonoscopia, melhorando os resultados de segmentação e detecção de pólipos. Diferentemente de outros trabalhos encontrados na literatura que fazem uso de abordagens tradicionais de aumento de dados (*data augmentation*) e da combinação de imagens de outras modalidades de exame, esta metodologia enfatiza a criação de novas amostras inserindo pólipos em imagens de colonoscopia publicamente disponíveis. A estratégia de inserção faz uso de pólipos gerados sinteticamente e também de pólipos reais, além de aplicar técnicas de processamento para preservar o aspecto realista das imagens, ao mesmo tempo em que cria automaticamente amostras mais diversas com seus rótulos apropriados para fins de treinamento. As redes neurais convolucionais treinadas com estes conjuntos de dados aprimorados apresentaram resultados promissores no contexto de segmentação e detecção. As melhorias obtidas indicam que a implementação de novos métodos para aprimoramento automático de amostras em conjuntos de imagens médicas tem potencial de afetar positivamente o treinamento de redes convolucionais.

Palavras-chave

Dados de Treinamento; Aumento de Dados; Redes Neurais Convolucionais; Pólipos; Colonoscopia;

Abstract

Thomaz, Victor de Almeida; Raposo, Alberto Barbosa (Advisor). **Evaluation of Data Augmentation through Synthetic Images Generation for Segmentation and Detection of Polyps in Colonoscopy Images Using Machine Learning**. Rio de Janeiro, 2020. 126p. Tese de doutorado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Nowadays colorectal cancer is the second-leading cause of cancer death worldwide. In recent years there has been an increase in interest in research aimed at the development of automatic methods for the detection of polyps and the most relevant results have been achieved through deep learning techniques. However, the performance of these approaches is strongly associated with the use of large and varied datasets. Samples of colonoscopy images are publicly available, but the amount and limited variation may be insufficient for successful training. Based on this observation, a new approach is described in this thesis with the objective of increasing the quantity and variation of colonoscopy images, improving the results of segmentation and detection of polyps. Unlike other works found in the literature that use traditional data augmentation approaches and the combination of images from other exam modalities, the proposed methodology emphasizes the creation of new samples by inserting polyps in publicly available colonoscopy images. The insertion strategy makes use of synthetically generated polyps as well as real polyps, in addition to applying processing techniques to preserve the realistic aspect of the images, while automatically creating more diverse samples with their appropriate labels for training purposes. Convolutional neural networks trained with these improved datasets have shown promising results in the context of segmentation and detection. The improvements obtained indicate that the implementation of new methods for the automatic improvement of samples in medical image datasets has the potential to positively affect the training of convolutional networks.

Keywords

Training Data; Data Augmentation; Convolutional Neural Networks; Polyp; Colonoscopy;

Sumário

Lista de figuras

Lista de tabelas

[...] Graças te dou, ó Pai, Senhor do céu e da terra, porque escondeste estas coisas dos sábios e cultos e as revelaste aos pequeninos.

Mateus, *11:25.*

Lista de Abreviaturas

ANN – *Artificial Neural Network* (Rede Neural Artificial)
CLAHE – *Contrast Limited Adaptive Histogram Equalization*
CNN – *Convolutional Neural Network* (Rede Neural Convolucional)
CGAN – *Conditional Generative Adversarial Network* (Rede Adversária Generativa Condicional)
CRF – *Conditional Random Field*
DT – *Decision Tree*
FPR – *False Positive Rate* (Taxa de Falso Positivo)
FPS – Fotogramas Por Segundo
GAN – *Generative Adversarial Network* (Rede Adversária Generativa)
K-NN – *K Nearest Neighbors*
LDA – *Linear Discriminant Analysis*
MFNN – *Multi-layer Feed-forward Neural Network*
MLP – *Multi-layer Perceptron*
Pixel – *Picture element*
PLS – *Partial Least Squares*
RF – *Random Forest*
RM – Ressonância Magnética
ROI – *Region of Interest* (Região de Interesse)
RPN – *Region Proposal Network* (Rede de Proposição de Regiões)
SVM – *Support Vector Machine*
YOLO – *You Only Look Once*

1
Introdução

Este capítulo apresenta as considerações iniciais para o entendimento da pesquisa desenvolvida nesta tese. Primeiramente, o contexto da pesquisa é demonstrado e na sequência os aspectos principais do problema são resumidos. Após, são descritos os objetivos e as contribuições deste estudo. Finalmente, a organização do texto é apresentada com breve descrição dos assuntos em cada capítulo.

1.1
Contexto

Imagens digitais são ferramentas importantes para fornecer informações e, em geral, possuem potencial para contribuir na compreensão e solução de problemas. Na atualidade, recursos computacionais têm sido frequentemente aplicados para obter estas informações em diversas áreas do conhecimento, desde diagnóstico médico, levantamento geológico, defesa antimíssil, controle de qualidade, entre outros. Avanços tecnológicos nos equipamentos médicos como o endoscópio permitem a inspeção de estruturas internas do corpo humano e a consequente captura de imagens. Isto contribuiu para criação de um novo campo de pesquisa chamado de processamento de imagens endoscópicas (*endoscopic image processing*) [84]. Tópicos neste campo de pesquisa incluem, mas não se limitam a: aprimoramento de imagem (*image enhancement*), sistemas de suporte a decisão automáticos (*automated decision support systems*), segmentação de imagens (*image segmentation*) e detecção de pólipos (*polyp detection*) [84]. Esta última está relacionada com os estudos para detecção do câncer de cólon (*colon cancer detecion*).

As lesões conhecidas como pólipos são precursoras do câncer do cólon [97]. Conforme mencionado por Bernal et al. [10], os pólipos são lesões que se manifestam como superfícies salientes na parede do cólon. Atualmente é recomendada a inspeção do cólon, isto é, o procedimento de colonoscopia [60], no qual o médico realiza o exame e observa as imagens (vídeo capturado por endoscópio) do cólon para identificar regiões com lesões e removê-las. O sucesso deste procedimento está baseado na capacidade de percepção das lesões pelo médico, que pode ser comprometida pelo cansaço e pela pouca

experiência. Segundo Leufkens et al. [82] cerca de 22% a 28% dos pólipos não são visualizados durante a colonoscopia. De acordo com dados coletados em 185 países, o câncer de cólon é atualmente o segundo maior em mortalidade [20]. No Brasil, em 2017, ocorreram cerca de 9,12 óbitos por câncer de cólon e reto a cada 100 mil homens e 9,33 no caso das mulheres. Estima-se que nos anos de 2020 até 2022 podem haver 19,63 casos novos a cada 100 mil habitantes do sexo masculino e o equivalente a 19,63 novos casos em mulheres por ano [66].

Com a melhoria na qualidade das imagens retornadas pelos equipamentos, houve um aumento do interesse em pesquisas voltadas para detecção de pólipos [45]. Neste contexto, diversas pesquisas têm sido direcionadas nos últimos anos para o desenvolvimento de métodos automáticos para detecção de pólipos [1–3, 6, 10–12, 23, 63, 64, 68, 71, 73–75, 79, 83, 92, 99, 105, 115, 134–137, 144, 145, 148, 150]. Nos primeiros estudos, métodos baseados em análise de curvatura, cor, textura, intensidade, regiões *Watershed* [94], além de algoritmos sofisticados para modelagem da aparência dos pólipos foram empregados [10–12, 63, 71, 79]. Abordagens de detecção de pólipos mais atuais baseadas em aprendizado de máquina têm apresentado resultados promissores [1–3, 23, 64, 68, 73, 75, 83, 92, 105, 137, 144, 145, 148, 150].

1.2 Problema

Trabalhos recentes como Urban et al. [140], Yu et al. [147], Pogorolev et al. [105], Bardhi et al. [7], Shin et al. [125], Taha et al. [133] tem evidenciado o uso de redes convolucionais profundas (*Deep Convolutional Neural Network*) como um método para detecção de pólipos com altas acurácias. Apesar disso, a eficiência destes métodos está diretamente associada a quantidade e variedade das imagens de colonoscopia presentes nos conjuntos de treinamento.

No contexto médico, dados de treinamento são difíceis de se obter devido ao alto custo da mão de obra especializada e de restrições de privacidade [50]. O processo de rotular a localização de cada lesão deve ser efetuado por especialistas clínicos, sendo um processo demorado e caro [24, 102]. O benefício potencial de aplicações de aprendizado profundo é prejudicado pela falta de disponibilidade de conjuntos de dados de imagens adequadamente gravados (rótulos) em quantidades e variabilidade para atender aos requisitos dos aplicativos de detecção para uso clínico.

Grandes conjuntos de dados são de uso exclusivo de poucos grupos de pesquisa, impactando na produção científica daqueles que não têm acesso a estas informações. Ainda assim, existem imagens médicas disponíveis publi-

camente, e.g. o CVC-ClinicDB, que é composto por 612 imagens extraídas de vídeos de colonoscopia [10]. Entretanto, esta quantidade de imagens pode ser insuficiente para o treinamento de uma rede convolucional profunda, visto que um método com boa acurácia utilizou 8641 imagens [140]. Apesar disso, o estudo de Urban et al. [140] precisou unir imagens de duas modalidades de exames diferentes (*white light endoscopy* e *narrow-band imaging*) para compor um conjunto com esta extensão.

O conjunto de imagens de colonoscopia CVC-ClinicDB é limitado no aspecto de quantidade além da variedade de pólipos capturados. Por exemplo, as imagens de 1 até 25 apresentam a mesma lesão modificando apenas o ponto de vista. Idealmente, um conjunto de dados satisfatório deve ser composto por uma grande quantidade de imagens que apresentem tamanhos e morfologias de pólipos diversas obtidas de exames de diferentes pacientes. Adicionalmente, as imagens devem conter um conjunto de anotações adequadas para seu processamento pelos algoritmos de aprendizado de máquina, neste caso, a localização dos pólipos em cada imagem deve ser devidamente indicada (*annotations* ou *ground truth*) por especialistas da área. Quanto maior e mais variados os dados, melhor será o treinamento e os resultados do sistema que faz uso do aprendizado de máquina.

Mesmo com a evolução dos sistemas de detecção de pólipos que utilizam técnicas de aprendizado profundo, pouco foi feito em relação aos conjuntos de dados no sentido de aprimorar amostras. Baseado nesta observação e na dificuldade de se obter conjuntos de imagens de colonoscopia adequados, surgiu a motivação para criação desta tese.

Assim, é proposto no presente trabalho uma técnica para aumentar a variação de pólipos em imagens de colonoscopia, com o propósito de melhorar a detecção destas lesões em sistemas que utilizam a abordagem de aprendizado de máquina. O conceito de aprimoramento de dados utilizado nesta tese representa a técnica proposta para incrementar a variação de pólipos nas imagens, de modo a fornecer um tipo de aumento de dados (diferente do tradicional) especificamente direcionado a estas lesões.

1.3 Justificativa e Objetivo

Os exames de colonoscopia são amplamente utilizados para prevenir que pólipos se tornem em cânceres, no entanto, o benefício deste procedimento de prevenção depende da taxa de detecção das lesões. Novas abordagens são necessárias para aumentar esta taxa de detecção. Sistemas de diagnóstico assistidos por computador têm potencial para contribuir na tarefa de localização

de pólipos auxiliando médicos endoscopistas durante os exames. Tais sistemas oferecem um benefício no contexto de colonoscopia, pois não há necessidade de nenhuma alteração no colonoscópio ou procedimento.

Apesar dos avanços na área, ainda não há uma solução definitiva devido à variedade de tamanho, forma e localização dos pólipos e a qualidade das imagens obtidas [145]. Um mesmo pólipo pode ter uma aparência muito distinta quando visualizada por diferentes ângulos, intensidade de luz e distância da câmera. Nenhuma das metodologias apresentadas nos estudos citados anteriormente foram adotadas para um tratamento de rotina de pacientes e um sistema de detecção automática de pólipos continua sendo relevante na área médica. O progresso nos resultados das abordagens de detecção com aprendizagem de máquina é frequentemente restrito a quantidade e variedade dos dados disponíveis. Técnicas tradicionais de aumento de dados (*augmentations*) [47] favorecem a variação das imagens, porém, sem acrescentar elementos novos nas amostras. A Figura 1.1 apresenta exemplos de operações comuns utilizadas para o aumento de dados tradicional como rotação, espelhamento e aproximação (*zoom*). Deste modo, a partir da imagem original outras versões modificadas por estas operações são produzidas.

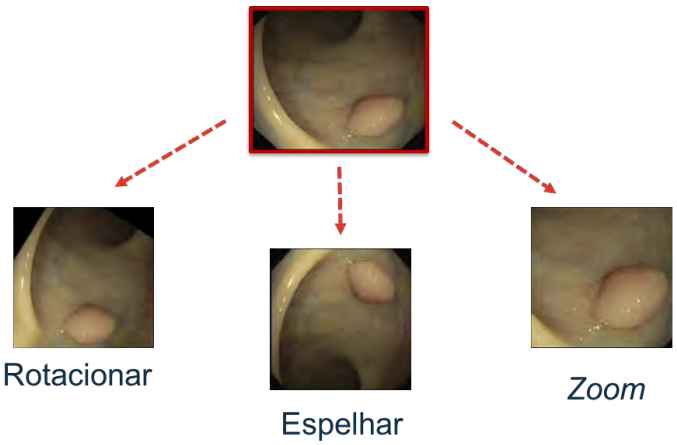

Figura 1.1: Exemplo de técnicas tradicionais de aumento de dados (*data augmentations*).

O objetivo geral desta tese é propor uma estratégia de aprimoramento de dados automática para aumentar a quantidade e variedade de pólipos em conjuntos de imagens de colonoscopia disponíveis publicamente. Com isso, pretende-se melhorar os resultados de segmentação e detecção de pólipos em abordagens baseadas em aprendizado de máquina. Um exemplo deste aprimoramento de dados pode ser visto na Figura 1.2, onde um pólipo (indicado em vermelho) pertencente a imagem origem (a) é acrescentado sobre a imagem destino (b). O resultado é demonstrado na Figura 1.2 (c), que

recebeu este novo pólipo posicionado na região delineada em vermelho. Assim, a proposta de aprimoramento de dados apresentada nesta tese pode beneficiar significativamente o treinamento de redes neurais, eventualmente empregadas para facilitar a localização de pólipos na prática clínica.

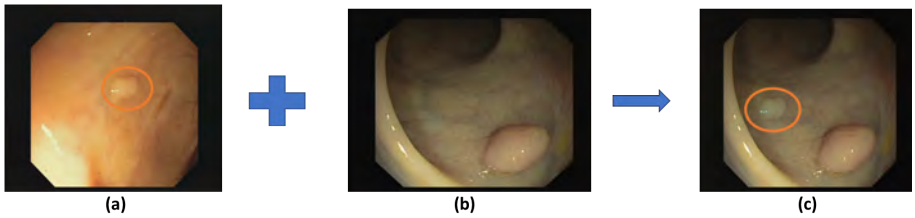

Figura 1.2: Exemplo de aprimoramento de dados proposto nesta tese por meio da inserção de pólipos. (a): imagem origem com área do pólipo destacada em vermelho. (b): imagem destino. (c): imagem aprimorada com cópia do pólipo presente em (a) destacada em vermelho.

Visando atingir o objetivo de aperfeiçoar os dados de treinamento, esta estratégia de aprimoramento está baseada em uma abordagem singular que insere pólipos reais ou sintéticos em imagens de colonoscopia criando novas amostras, que consistem em pares imagens de colonoscopia e *ground truth*. Em função disso são realizados os seguintes passos:

- Desenvolvimento de técnicas para inserção de pólipos em imagens de colonoscopia.

  A inserção de pólipos depende da seleção de regiões adequadas na imagem destino, por isso foi desenvolvida uma metodologia de segmentação de regiões compatíveis com os pólipos a serem inseridos. A imagem aprimorada precisa manter o aspecto realista, portanto, foram estabelecidos meios para manter a coerência visual da imagem de colonoscopia que recebe o novo pólipo. A etapa de treinamento depende da indicação real das lesões para o aprendizado da rede, assim, é preciso implementar um procedimento automático para adaptação da imagem *ground truth* de acordo com a forma e localização adequada do pólipo inserido.

- Desenvolvimento da estratégia para geração de pólipos sintéticos levando em consideração o aspecto visual realista.

  A variedade de formatos e texturas de pólipos pode ser extendida por meio do desenvolvimento um procedimento de geração de pólipos sintéticos baseado em aprendizado de máquina. Neste caso, os dados de treinamento são os pólipos reais disponíveis no conjunto público de imagens de colonoscopia. Por isso, foi desenvolvido um processo de seleção dos pólipos reais e suas respectivas máscaras binárias para treinamento da rede geradora de pólipos sintéticos.

- Desenvolvimento de processo para validação dos conjuntos de dados aprimorados em implementações de aprendizado de máquina para localização de pólipos.

  Para avaliar os efeitos das imagens aprimoradas empregadas no treinamento foram implementadas duas soluções para localização de pólipos. A primeira faz uso de uma rede neural para segmentação de lesões e a segunda está baseada na técnica de detecção. A rede de segmentação foi treinada com imagens aprimoradas de pólipos reais e a rede de detecção foi treinada com uma combinação de imagens que receberam pólipos reais e sintéticos.

1.4 Contribuições

Este estudo traz as seguintes contribuições:

1. **O desenvolvimento de uma estratégia de aprimoramento do conjunto de dados de treinamento para localização de pólipos.**

   Propomos uma estratégia para melhorar resultados de localização de pólipos em aplicações que utilizam aprendizado de máquina. Para isso, imagens de colonoscopia disponíveis publicamente receberam pólipos reais e sintéticos, resultando em novas amostras com pólipos variados. Desta forma, novos conjuntos de imagens aprimorados por esta estratégia foram utilizados no treinamento de redes neurais demonstrando resultados promissores. Esta abordagem difere das técnicas tradicionais de aumento de dados (*augmentations*) pois cria novas amostras ampliando a variedade em conjuntos de dados reduzidos como o CVC-ClinicDB a partir da inserção de pólipos. Apesar da proposta de inserção de pólipos em regiões apropriadamente segmentadas funcionar efetivamente, a composição de duas imagens necessita de ajustes para uma transição suave capaz de reduzir as diferenças de cor e iluminação. A partir desta observação foi aplicada a técnica de *Poisson* como solução para tornar a fusão das imagens mais próxima da aparência real. As imagens aprimoradas contribuíram para o aprendizado da rede demonstrando que a estratégia descrita nesta tese pode ser complementar ao aumento de dados tradicional. Não encontramos nenhum estudo que tenha utilizado esta estratégia de inserção de pólipos para aprimoramento de imagens de treinamento na modalidade de colonoscopia padrão (*white light colonoscopy*), indicando uma provável lacuna que a presente tese pode ser útil para preencher.

2. **O desenvolvimento de uma abordagem de geração de pólipos sintéticos e suas respectivas máscaras binárias a partir de um conjunto com reduzida diversificação de pólipos reais.**

   Esta tese descreve um procedimento para geração de pólipos sintéticos por meio de redes convolucionais. Especificamente com a utilização das redes GAN e CGAN o procedimento mostrou a capacidade de criação de pólipos sintéticos com formato e textura variados. A aparência dos pólipos é semelhante a real incluindo detalhes de reflexão da luz, cores e sombras. Isto permite a criação de uma variedade de pólipos sintéticos a partir da reduzida diversidade de pólipos reais no conjunto de dados CVC-ClinicDB. No melhor de nosso conhecimento nenhum trabalho abordou o aspecto de criação de uma coleção de pólipos sintéticos que podem ser automaticamente selecionados e inseridos em outras imagens de colonoscopia. A aplicação dos pólipos sintéticos refletiu em uma melhora significativa na detecção em relação ao uso das técnicas de aumento de dados tradicionais.

3. **O desenvolvimento de duas abordagens de localização de pólipos baseadas em aprendizado de máquina treinadas com dados aprimorados.**

   Em conformidade com as contribuições anteriores foram implementados dois métodos de localização de pólipos. O primeiro emprega uma rede de segmentação cuja máscara binária resultante indica a posição da lesão. O segundo faz uso de uma rede de detecção onde uma caixa delimitadora indica a localização do pólipo. Ambas as redes foram treinadas com conjuntos de imagens aprimoradas por meio da inserção de pólipos. Estas aplicações demonstram a viabilidade da construção de soluções a partir dos conjuntos de dados aprimorados. Ambas as implementações facilitaram o processo de validação das novas amostras com pólipos reais e sintéticos propostas nesta tese.

1.5 Organização da tese

Esta tese está organizada em seis capítulos:

Capítulo 2 - introduz os principais conceitos e técnicas para melhor compreensão da tese, incluindo processamento de imagens digitais, aprendizado de máquina, métricas de avaliação e imagens de colonoscopia.

Capítulo 3 - fornece uma revisão de estudos relacionados ao contexto desta tese.

Capítulo 4 - descreve todas as etapas da estratégia de inserção de pólipos, a composição dos conjuntos de imagens, validação por meio de uma rede de segmentação e os resultados obtidos.

Capítulo 5 - apresenta a metodologia proposta para a geração de pólipos sintéticos, a aplicação destes em conjuntos de imagens de colonoscopia, experimentos de validação com uma rede de detecção e os resultados alcançados.

Capítulo 6 - conclui esta tese e lista possíveis extensões deste estudo como trabalhos futuros.

2
Fundamentação Teórica

Neste capítulo são apresentados tópicos que descrevem as técnicas mais relevantes que foram utilizadas na abordagem de aprimoramento de dados proposta nesta tese e no treinamento das redes neurais. A finalidade deste capítulo é prover um melhor entendimento sobre os conceitos relacionados ao processamento de imagens, aprendizado de máquina e imagens de colonoscopia.

2.1
Processamento de Imagens Digitais

A área de processamento de imagens digitais consiste na manipulação de imagens por meio de operações que geram outras imagens como resultado [25]. Neste contexto, objetivo do processamento é melhorar o aspecto visual favorecendo a interpretação dos elementos presentes na imagem.

Uma imagem digital pode ser vista como uma função $f(x, y)$ [46]. Considerando uma imagem em tons de cinza, a função $f(x, y)$ representa a variação de níveis de cinza ao longo das coordenadas espaciais x e y [25], onde por convenção $x = [1, 2, ..., M]$ e $y = [1, 2, ..., N]$. Na Figura 2.1 é possível observar melhor os níveis de cinza presentes na região destacada. Em geral, os níveis de cinza em cada *pixel* estão no intervalo inteiro conforme:

$$0 \leq f(x, y) \leq 255 \tag{2-1}$$

A matriz é uma estrutura de dados comumente utilizada para representar uma imagem digitalizada. Cada posição da matriz representa um ponto da imagem (*pixel*). Deste modo, a imagem digital $f(x, y)$ pode ser organizada sob a forma de matriz, sendo m e n valores inteiros, como observado na Equação 2-2.

$$f(x, y) \approx \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \tag{2-2}$$

As imagens são formadas por vários elementos chamados de *pixels*. Normalmente, cada *pixel* é representado por um quadrado, neste sentido uma imagem também pode ser vista como uma grade de *pixels*. Este tipo de

Figura 2.1: Exemplo de imagem digital em tons de cinza com região destacada (17 x 17 *pixels*) [36].

organização faz com que o *pixel* não tenha as mesmas propriedades em todas as direções (i.e. anisotrópico), observando que a distância entre um *pixel* e seus vizinhos na diagonal é de $\sqrt{2}$, enquanto que para vizinhos na vertical e horizontal é de 1 [27]. Então, um *pixel* p possui quatro vizinhos, considerando as direções horizontal e vertical, com as respectivas coordenadas dadas por:

$$(x+1, y), (x-1, y), (x, y+1), (x, y-1) \tag{2-3}$$

Adicionalmente, no caso da vizinhança diagonal do *pixel* p, as coordenadas serão definidas pela Equação 2-4:

$$(x+1, y+1), (x+1, y-1), (x-1, y-1), (x-1, y+1) \tag{2-4}$$

Estas definições são importantes por causa dos algoritmos que precisam preencher áreas e determinar a continuidade de objetos. Devido às características de vizinhança é preciso escolher o tipo conectividade. Isto é, a que considera somente os *pixels* da horizontal e vertical, chamada de $N_4(p)$ ou a que utiliza a vizinhança diagonal junto com a horizontal e vertical, denominada $N_8(p)$. A Matriz 2-5 ilustra a localização dos vizinhos diagonais do *pixel* p (i.e. d_1, d_2, d_3, d_4) e dos vizinhos verticais e horizontais (i.e. v_1, v_2, h_1, h_2). Assim, a distância entre um *pixel* p e seu vizinho irá depender do tipo de conectividade. Se a conectividade for $N_4(p)$, então a distância será 1. No entanto, a distância será $\sqrt{2}$ quando a conectividade $N_8(p)$ for utilizada.

$$\begin{pmatrix} d_1 & v_1 & d_2 \\ h_1 & p & h_2 \\ d_3 & v_2 & d_4 \end{pmatrix} \tag{2-5}$$

Para determinar quais *pixels* estão conectados entre si é preciso que os seguintes critérios sejam satisfeitos [46]:

– Dados dois *pixels* p e q, estes são conectados se:

1. Possuem conectividade $N_4(p)$ ou $N_8(p)$ e;
2. Apresentam níveis de cinza iguais;

Um conjunto de *pixels* conectados em uma imagem recebem o nome de componente conexo. Cada componente conexo pode ser marcado com um rótulo. Isto facilita por exemplo a seleção de um componente conexo do restante da imagem.

2.1.1 Histograma

O histograma é uma técnica que realiza um mapeamento de todos os níveis de cinza presentes em uma imagem digital [25]. Este procedimento fornece informações sobre a distribuição dos *pixels* (percentual ou quantidade), além de ser útil para aplicações de compressão e segmentação de imagens [46]. Um histograma fornece um gráfico da frequência de cada nível de cinza da imagem. Assim, é possível perceber pela aparência do gráfico se, por exemplo, a imagem é predominantemente escura, visto que possui muitos pixels concentrados na parte baixa da escala de intensidade [19]. Segundo Gonzalez e Woods [46], um histograma de uma imagem digital pode ser obtido pela seguinte equação:

$$h(r_k) = \frac{n_k}{MN}, \quad \text{onde } k = 0, 1, 2, \cdots, L-1 \tag{2-6}$$

Como apresentado na Equação 2-6, o total de *pixels* na imagem é MN, sendo M linhas por N colunas. O número de níveis de cinza possíveis é de L ($L = 2^b = 256$, se $b = 8$ bits, por exemplo). Assim, r_k é o valor do nível de cinza de um determinado *pixel* e n_k é a quantidade de ocorrências de r_k na imagem. Usualmente, o histograma é normalizado, por isso, cada componente é dividida pelo total de *pixels* na imagem (i.e. n_k/MN). A Figura 2.2(b) apresenta um exemplo de histograma. O eixo horizontal do gráfico representa os níveis de cinza possíveis e o eixo vertical apresenta a quantidade de *pixels* na imagem com o respectivo nível de cinza. Comparando o histograma com a imagem

original na Figura 2.2(a) é possível perceber que existem muitos *pixels* mais escuros, i.e. maior frequência na parte baixa da escala de intensidade (lado esquerdo no eixo horizontal do gráfico).

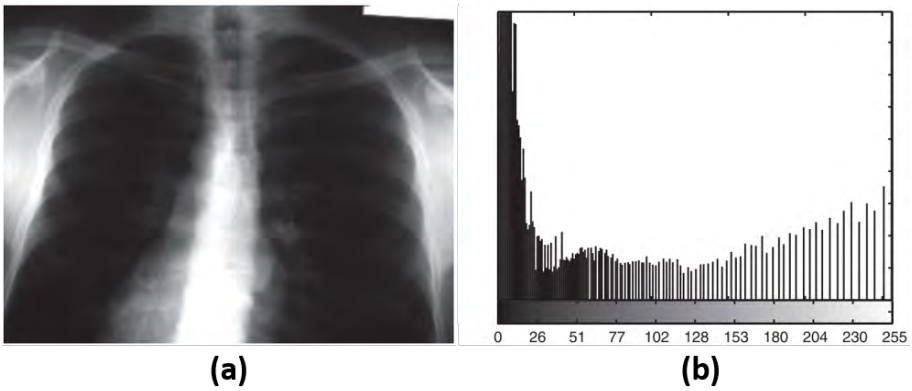

Figura 2.2: Exemplo de histograma de uma imagem (Adaptado de [25]). (a): imagem original. (b): gráfico do histograma indicando que a imagem original possue muitos *pixels* escuros.

O aspecto visual da imagem pode ser melhorado por meio da manipulação dos valores de intensidade retornados no histograma. Por exemplo, é possível aumentar o contraste distribuindo os valores de intensidade mais uniformemente pela escala dos níveis de cinza. Este processo é chamado de equalização do histograma [19]. A Figura 2.3 apresenta o resultado deste processo. O contraste da imagem (Figura 2.3(a)) foi aperfeiçoado devido a equalização, conforme demonstrado pelo gráfico do histograma em Figura 2.3(b).

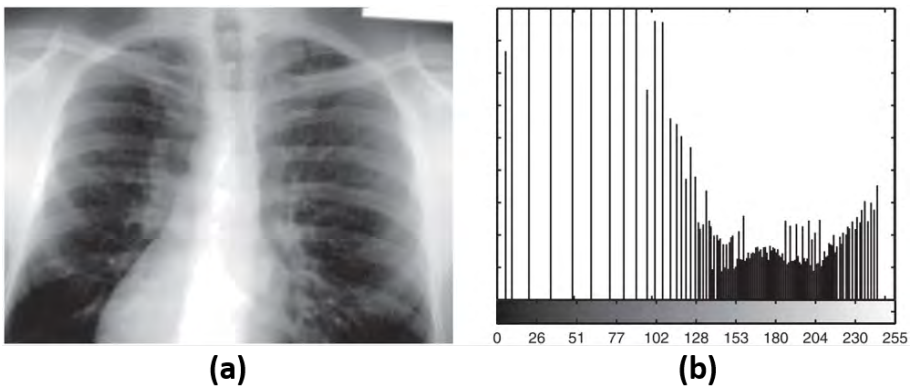

Figura 2.3: Exemplo de equalização do histograma (Adaptado de [25]). (a): imagem (Figura 2.2,a) modificada pela equalização do histograma. (b): histograma que mostra a nova distribuição dos níveis de cinza.

A equalização do histograma favorece imagens que não possuem características mais distintas, como imagens de objetos e pessoas. Porém, seu uso pode amplificar o ruído em áreas relativamente homogêneas da imagem [89].

Para resolver este problema de amplificação de ruído, foi proposta uma versão modificada da equalização de histograma denominada equalização adaptativa de histograma com limitação de contraste (CLAHE[1]) [153]. Esta técnica foi inicialmente aplicada a imagens médicas se mostrando efetiva em diferentes modalidades de exames (e.g. ressonância magnética, tomografia computadorizada, angiograma, raio-x) [104]. O processo de CLAHE é capaz de aprimorar o contraste e reduzir os efeitos do ruído. É aplicado em blocos da imagem ao invés de considerar a imagem como um todo. O aperfeiçoamento do contraste é aplicado em cada bloco, de forma que o contraste de cada seção distinta é utilizado para distribuir melhor os níveis de cinza da imagem inteira resultante do processamento [89]. A Figura 2.4 apresenta uma comparação de duas imagens de colonoscopia, sendo (a) a imagem original em tons de cinza e (b) a mesma imagem, porém processada utilizando a técnica CLAHE. É perceptível que os níveis de cinza estão melhor distribuídos em (d) do que em (b). Este procedimento fez com que os detalhes da mucosa estejam mais nítidos na Figura 2.4(c).

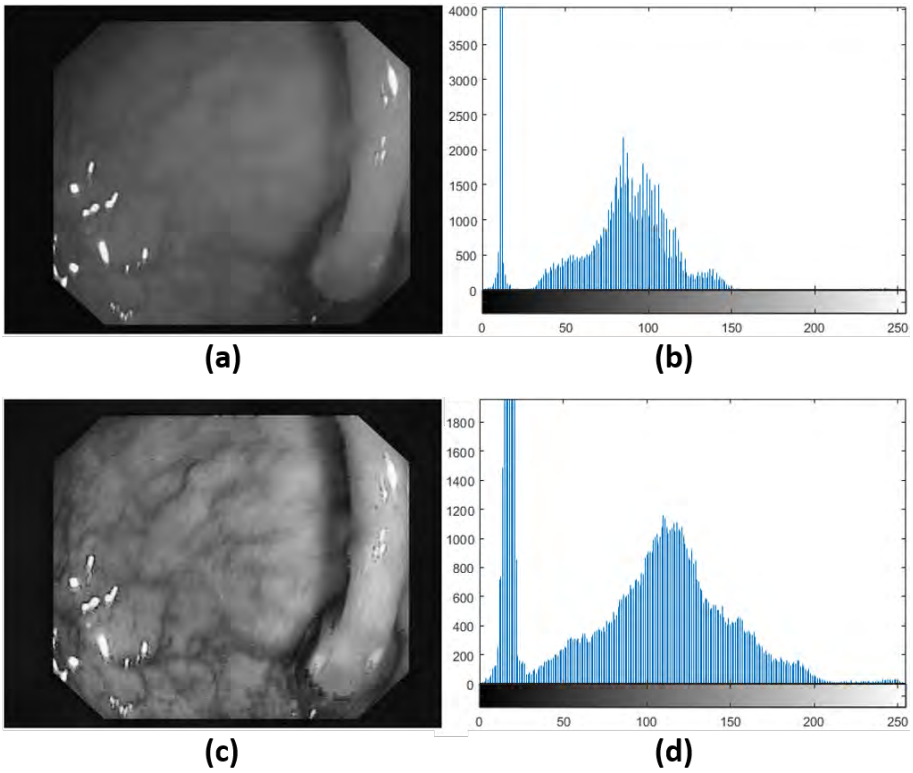

Figura 2.4: Contraste de imagem de colonoscopia aprimorado com o uso da técnica CLAHE. (a): imagem de colonoscopia em tom de cinza. (b): histograma referente a imagem (a). (c): mesma imagem de (a) aprimorada com CLAHE. (d): histograma da imagem (c) obtido pelo processamento CLAHE, que demonstra a melhor distribuição dos níveis de cinza na escala.

[1]Do inglês *Contrast Limited Adaptive Histogram Equalization.*

2.1.2 Limiarização (*Thresholding*)

Para o melhor entendimento e análise das características de uma imagem, é comum o uso da técnica de segmentação. Segmentação é o processo que subdivide uma imagem em suas partes ou objetos constituintes [149]. O processo de segmentação gera uma representação mais significativa da imagem, por meio da marcação de conjuntos de *pixels* (segmentos) localizados em regiões de interesse. Cada região é formada agrupando-se os *pixels* com características semelhantes [8]. Não existe um algoritmo geral que funcione para todos os tipos de imagem [41]. O processo de segmentação é empírico e deve ser adaptado de acordo com as características de cada tipo de imagem. Uma das abordagens mais utilizadas para segmentação é a baseada na similaridade dos *pixels* por meio do método de limiarização [46].

A segmentação por limiarização (*thresholding*) pode ser aplicada a imagens em tons de cinza que apresentem razoável variação entre seus níveis de intensidade e que tenham suas regiões separadas de acordo com a diferença destes valores. Por exemplo, se a diferença de intensidade dos *pixels* permite que elementos sejam caracterizados e separados do fundo da imagem. Neste caso, é utilizado um valor chamado de limiar (*threshold*), que quando comparado a cada valor de intensidade na imagem faz distinção entre *pixels* de maior intensidade do restante, destacando os elementos do fundo.

Dada uma imagem em tons de cinza $f(x, y)$, composta por elementos com valores de cinza mais altos (*pixels* claros) e um fundo de imagem com valores mais baixos (*pixels* escuros), é possível extrair os elementos ("objetos") do fundo da imagem de acordo com a Equação 2-7 [46].

$$g(x,y) = \begin{cases} 1, & \text{se} \quad f(x,y) > T \\ 0, & \text{se} \quad f(x,y) \leq T \end{cases} \tag{2-7}$$

Neste caso, T é um valor limiar e qualquer *pixel* na coordenada (x, y) é pertencente a um objeto quando sua intensidade satisfaz $f(x, y) > T$. Do contrário é um *pixel* de fundo. O resultado deste processo gera uma imagem binária, pois os valores de intensidade dos *pixels* são 1 (objetos) ou 0 (fundo).

No entanto, o nível de qualidade resultante desta técnica de segmentação está vinculada a melhor escolha possível de valor para o limiar T. Uma abordagem que apresenta a melhor separação entre os elementos na imagem e o fundo é chamado de método Otsu [98]. Este método é considerado ótimo no sentido de que encontra um limiar que maximiza a variância dos valores de intensidade entre os objetos e o fundo [46]. A execução do Otsu está baseada no histograma normalizado da imagem e irá gerar automaticamente uma imagem

binária encontrando o melhor limiar por meio de múltiplas iterações. As regiões ("objetos") desta imagem segmentada podem ser individualmente analisadas posteriormente, por exemplo.

2.1.3 Operações Morfológicas

As operações morfológicas têm o objetivo de extrair ou melhorar características das imagens com base em sua forma e estrutura [93]. Estas operações podem ser efetuadas sobre imagens coloridas e tons de cinza, porém, inicialmente foram desenvolvidas para uso com imagens binárias.

Estas técnicas são aplicadas a partir de duas entradas: a imagem original e um elemento estruturante que delimita uma área de comparação próxima a um *pixel* alvo. Este elemento é uma subimagem que conforme movida sobre a imagem original atua como um marcador da área (vizinhança) que será considerada na operação. O *pixel* que é indicado como origem no elemento estruturante é a posição de referência que guia a sobreposição em cada *pixel* da imagem original. Por exemplo, na Figura 2.5(b) e na Figura 2.6(b), o *pixel* central do elemento estruturante está sendo utilizado como origem.

Tais operações são úteis para se obter uma descrição da forma de uma região ou utilizada para pré-processamento, por exemplo. A origem teórica destas operações está na morfologia matemática que utiliza a teoria dos conjuntos. No contexto da imagem digital, os *pixels* pertencentes as regiões e ao fundo da imagem formam subconjuntos, que descrevem a morfologia da imagem. Estes subconjuntos fazem parte de um conjunto maior, que é a imagem como um todo, além do elemento estruturante que é visto como outro conjunto. No contexto das operações morfológicas com imagens binárias, são frequentemente utilizadas as operações primárias de dilatação, erosão, abertura e fechamento [8].

A dilatação é uma operação que adiciona mais *pixels* nas bordas dos objetos, faz com que os objetos fiquem mais visíveis e preenche pequenos buracos. Esta operação faz uso da adição vetorial [8] para efetuar a transformação sobre os conjuntos A (imagem original) e B (elemento estruturante) conforme definido na Equação 2-8.

$$A \oplus B = \{x \mid x = a + b, a \in A, b \in B\} \tag{2-8}$$

A Figura 2.5 representa uma imagem binária onde os quadrados mais escuros indicam o valor 1 e os em branco o valor 0. Nesta ilustração, um *pixel* na imagem resultante (c) será preenchido como valor 1 se alguns dos

pixels vizinhos na mesma localização na imagem (a) (considerando a área do elemento estruturante (b)) tiver o valor 1.

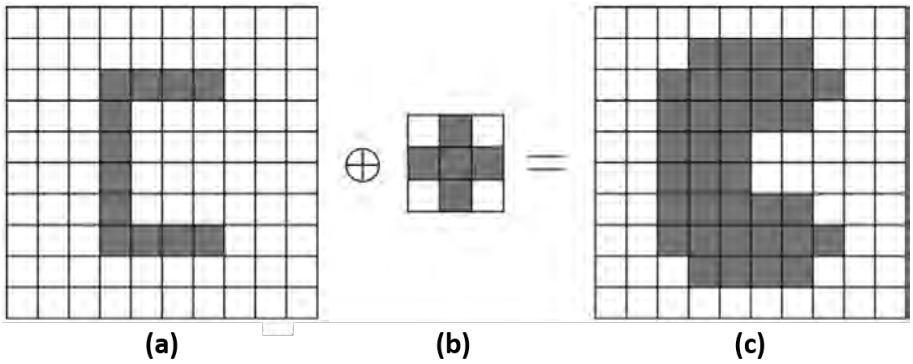

Figura 2.5: Ilustração da operação morfológica dilatação sobre imagem binária (Adaptado de [8]). (a): representação da imagem binária original. (b): elemento estruturante em formato de "cruz". (c): representação da imagem resultante da operação.

A erosão realiza a remoção de objetos pequenos, mantendo os mais substanciais, porém diminuindo as bordas. É uma operação que combina os dois conjuntos (A e B) por meio da subtração vetorial dos elementos deste conjunto [8]. A operação de erosão sobre a imagem A utilizando um elemento estruturante B pode ser representado pela Equação 2-9.

$$A \ominus B = \{x \mid x + b \in A \text{ para todo } b \in B\} \tag{2-9}$$

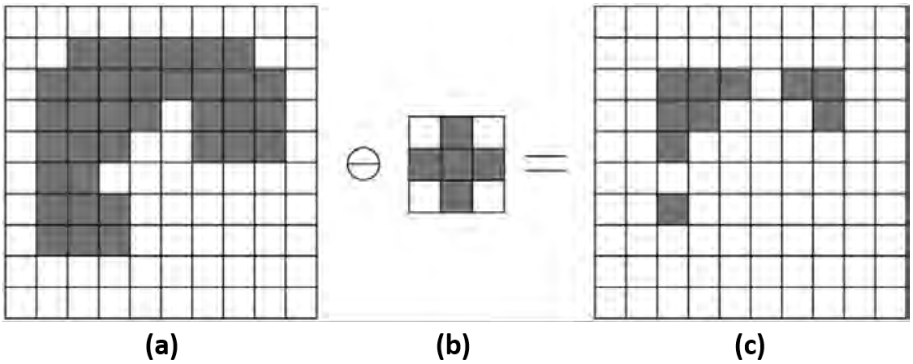

Figura 2.6: Ilustração da operação morfológica erosão sobre imagem binária (Adaptado de [8]). (a): representação da imagem binária original. (b): elemento estruturante em formato de "cruz". (c): representação da imagem resultante da operação.

Conforme visto na Figura 2.6, a operação morfológica de erosão aplica o elemento estruturante (b) sobre todos os *pixels* da imagem (a), gerando o resultado final (c). Neste caso, será atribuído o valor 0 a um *pixel* na imagem

resultante se algum *pixel* coberto pelo elemento estruturante (vizinhança) na imagem original também contiver o valor 0.

A operação morfológica de abertura é uma composição das operações de erosão e dilatação utilizando o mesmo elemento estruturante. Conforme descrito pela Equação 2-10 [8, 57], é executada uma operação de erosão e na sequência a operação de dilatação. O resultado obtido é que os componentes são menos detalhados, com remoção de ruídos e ainda há uma inserção de espaços entre componentes.

$$A \circ B = (A \ominus B) \oplus B \tag{2-10}$$

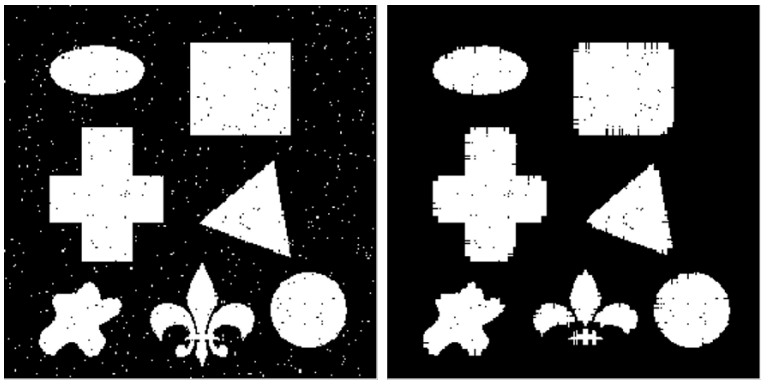

Figura 2.7: Ilustração da operação morfológica abertura sobre imagem binária (Adaptado de [33]). (a): representação da imagem binária original. (b): imagem resultante da operação de abertura morfológica. Foi utilizado o elemento estruturante cruz (9 x 9 *pixels*).

Os efeitos da operação de abertura estão ilustrados na Figura 2.7. Os ruídos presentes no fundo da imagem na Figura 2.7 (a) foram removidos da imagem resultado (b), que também é menos detalhada que a imagem original (a).

A operação de fechamento irá produzir um preenchimento de espaços vazios e tende a ampliar as bordas dos objetos. Se houver buracos formados por *pixels* em preto, estes serão preenchidos. Também é composta pelas operações de dilatação e erosão, conforme definido na Equação 2-11 [8, 57]. No entanto, a ordem das operações é diferente, sendo efetuada a dilatação e depois a erosão, também utilizando o mesmo elemento estruturante para ambas.

$$A \bullet B = (A \oplus B) \ominus B \tag{2-11}$$

A Figura 2.8(b) mostra um exemplo da operação de fechamento sobre uma imagem binária conforme visto em Figura 2.8(a). É possível perceber que

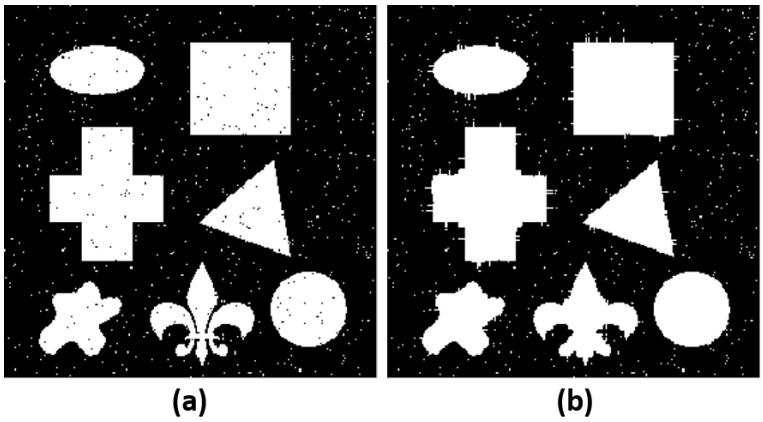

Figura 2.8: Ilustração da operação morfológica fechamento sobre imagem binária (Adaptado de [33]). (a): representação da imagem binária original. (b): imagem resultante da operação de fechamento morfológica. Foi utilizado o elemento estruturante cruz (9 x 9 *pixels*).

as bordas dos objetos foram expandidas e que buracos pequenos localizados no interior dos objetos foram preenchidos.

A reconstrução morfológica possibilita a extração de informações significativas sobre as formas em uma imagem. Operações baseadas nesta abordagem de reconstrução utilizam duas imagens e um elemento estruturante. Uma imagem chamada de marcador é processada com base nas características de outra imagem denominada máscara, que restringe a transformação, com a conectividade definida pelo elemento estruturante [46].

Considerando F como imagem marcador, G como máscara, o elemento estruturante formado por uma matriz de 3 x 3 *pixels* com valor 1 e $F \subseteq G$, a reconstrução de G a partir de F pode ser definida como $R_G(F)$ [46]. E ainda $R_G(F)$ pode ser obtido de acordo com o procedimento a seguir:

1. h_1 é inicializado recebendo a imagem F (marcador);
2. Crie a matriz 3 x 3 preenchida pelo valor 1 (elemento estruturante B);
3. Repita:
   - $h_{k+1} = (h_k \oplus B) \cap G$;
   - Até $h_{k+1} = h_k$;
4. $R_G(F) = h_{k+1}$;

As Figuras 2.9 e 2.10 trazem um exemplo de imagens para efeito de ilustração do procedimento anterior. Neste contexto, as imagens G (Figura 2.9(a)) e F (Figura 2.9(b)) definem as entradas do procedimento junto com o elemento estruturante (não representado na Figura 2.9). Após o término

do processo a imagem $R_G(F)$ será obtida, conforme demonstrado na Figura 2.9(c).

As iterações do procedimento estão representadas por imagens intermediárias, nas quais o resultado final na Figura 2.9(c) está baseado. As imagens referentes a estas etapas podem ser vistas na Figura 2.10. A partir da imagem marcador (Figura 2.9(b)), as áreas em que ocorre interseção com os elementos da imagem máscara (Figura 2.9(a)) vão sendo preenchidas após sucessivas operações de dilatação. As Figuras 2.10 (a-c) apresentam o preenchimento parcial com 100, 200 e 300 iterações, respectivamente.

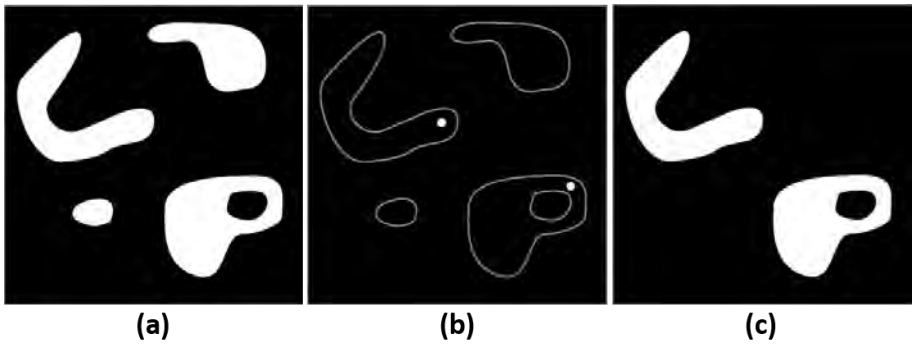

Figura 2.9: Representação do resultado do procedimento de reconstrução morfológica de uma imagem (Adaptado de [46]). Os contornos em (b) são ilustrativos. (a): imagem máscara. (b): imagem marcador. (c): imagem resultado reconstrução morfológica.

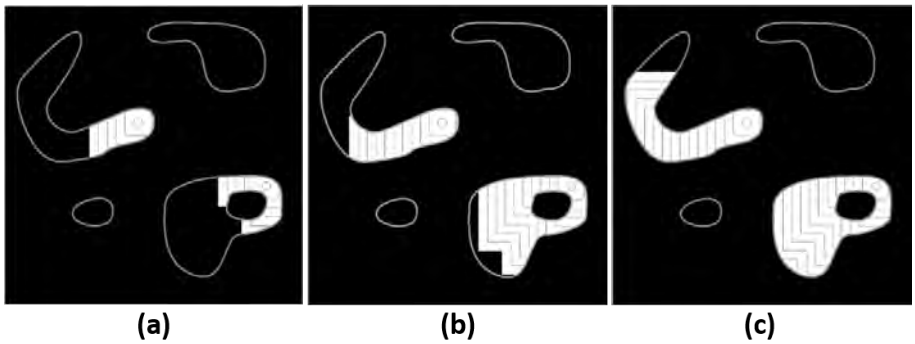

Figura 2.10: Representação das imagens intermediárias durante o processo de reconstrução morfológica de uma imagem (Adaptado de [46]). Os contornos dos objetos na Figura 2.9(a) foram mantidos para auxiliar no entendimento. (a): imagem parcial após 100 iterações. (b): imagem parcial após 200 iterações. (c): imagem parcial após 300 iterações.

As aplicações práticas da reconstrução morfológica são amplas e o que define os resultados obtidos pelas operações é a seleção das imagens marcadoras e máscaras. A combinação das operações morfológicas são úteis para pré-

processamento e extração de características, além de serem utilizadas como base em métodos de segmentação, por exemplo.

2.1.4 Segmentação com *Watershed*

Watershed é uma técnica de segmentação com origem na morfologia matemática [122] amplamente utilizada em imagens médicas [108], incluindo imagens endoscópicas [32]. Já a segmentação é, em linhas gerais, o processo de isolar objetos da imagem, i.e., particionando-a em regiões disjuntas, de forma que cada região seja homogênea em relação a alguma propriedade, como cor, tom de cinza ou textura [56]. Definições do algoritmo de *Watershed* podem ser vistas em [143] e [94][2].

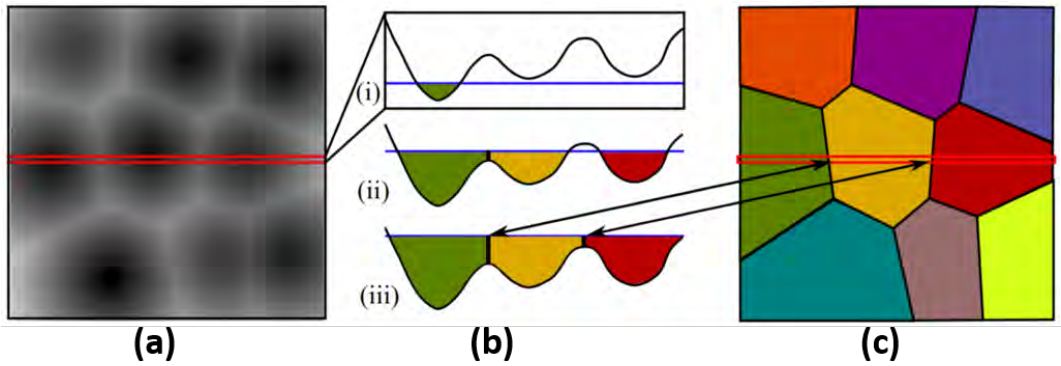

Figura 2.11: Ilustração de segmentação com a técnica de *watershed* [53]. (a): imagem em tons de cinza com seção destacada em vermelho. (b): representação do processo de segmentação referente a seção em vermelho. (c): resultado da segmentação apresentando as barreiras em preto separando as regiões.

O princípio dos algoritmos de segmentação utilizando *watersheds* usualmente consiste em considerar uma imagem como uma superfície topográfica. Os valores dos *pixels* de uma imagem em tons de cinza são considerados como altitudes. Deste modo, o relevo da imagem pode ser representado como regiões de valores baixos produzindo vales (mínimos), valores altos que formam picos (máximos) e regiões com declives (bacias hidrográficas). Além destes, também é considerado o conceito de planalto, que é uma região formada por elementos de mesma altura.

A Figura 2.11 ilustra o processo de segmentação com *watershed*. Na parte (a) está representada uma imagem em tons de cinza, como exemplo. As diferenças dos valores nos *pixels* em (a) formam áreas mais baixas e outras elevadas como ilustrado na Figura 2.11 (b). O algoritmo de *watershed* consiste em preencher as regiões mais baixas do relevo (mínimos), simulando uma

[2]Implementação utilizada nesta tese.

inundação (Figura 2.11 (b,i)). Quando o nível sobe e atinge um novo mínimo (Figura 2.11 (b,ii)) outra região é criada. Cada região alagada recebe um rótulo representado pelas diferentes cores. Para que regiões distintas não se misturem, barreiras são erguidas mantendo uma separação das regiões (linhas em preto na Figura 2.11 (b,iii)). Tais separações é que definem os limites de cada região formando a imagem segmentada, conforme visto na Figura 2.11 (c). Mais detalhes sobre a transformada de *watershed* pode ser visto em [118].

2.1.5
Técnica de *Poisson* para Edição de Imagem

A edição de imagens por meio da técnica de Poisson [101] é uma forma eficiente de sobrepor uma região de interesse pertencente a uma imagem em outra imagem de modo que o resultado seja visualmente convincente. A Figura 2.12 apresenta um exemplo desta composição de imagens que faz uso da técnica de Poisson para suavizar a imagem final. A seleção da região de interesse está representada em amarelo na Figura 2.12(a). Esta região irá sobrepor a área delimitada na Figura 2.12(b). O resultado da inserção da região de (a) (imagem origem) em (b) (imagem destino) pode ser visto na Figura 2.12(c). No entanto, a diferença de textura e brilho entre as partes das diferentes imagens está muito evidente (c). A técnica de Poisson é aplicada sobre a região inserida em (c) de modo a suavizar a descontinuidade das bordas em relação ao fundo da imagem destino, resultando em uma imagem mais visualmente coerente.

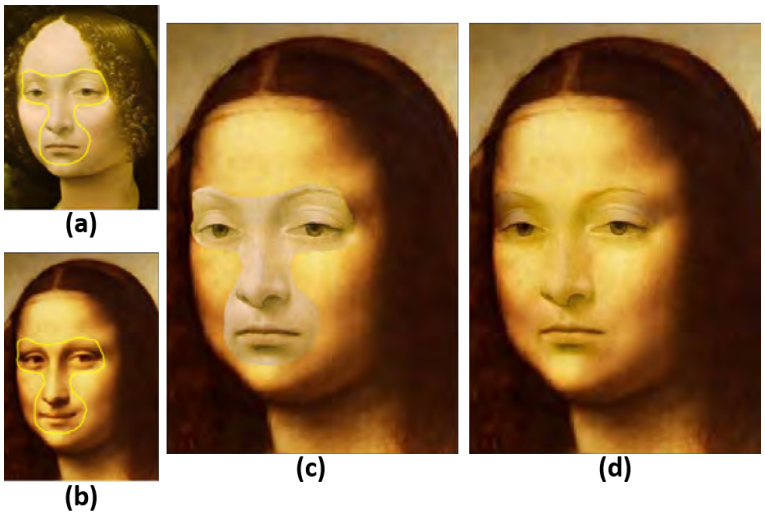

Figura 2.12: Exemplo de composição de imagem (Adaptado de [101]). (a): seleção de região de interesse na imagem fonte. (b): área na imagem destino onde a região será inserida. (c): região de interesse extraída da imagem fonte inserida sobre o fundo da imagem destino. (d): imagem suavizada utilizando o método Poisson.

A teoria básica sobre a técnica de Poisson é fundamentada na execução de uma interpolação da região Ω com a imagem destino. Conforme ilustrado na Figura 2.13, uma região de interesse A (Figura 2.13(a)) é utilizada para compor uma nova imagem (Figura 2.13(b)), de modo que A é inserido na área Ω, sendo $b(\Omega)$ sua área de borda e ainda v sendo o gradiente dos *pixels* em A. Então, os valores dos *pixels* em Ω, i.e. f, são obtidos de acordo com a Equação 2-12 [101]:

$$\min_{f} \int\int_{\Omega} |\nabla f - v|^{2} \text{ com } f|_{b(\Omega)} = f^{*}|_{b(\Omega)} \tag{2-12}$$

Sendo que $\nabla . = \frac{\partial .}{\partial x} + \frac{\partial .}{\partial y}$ é operador gradiente. Assim, os valores dos *pixels* em f são escolhidos de modo que o gradiente em A é mantido em Ω. Além disso, f (valores dos *pixels* em Ω) e f^{*} (valores dos *pixels* na imagem destino) são contínuos na borda de Ω, i.e. $b(\Omega)$ [91].

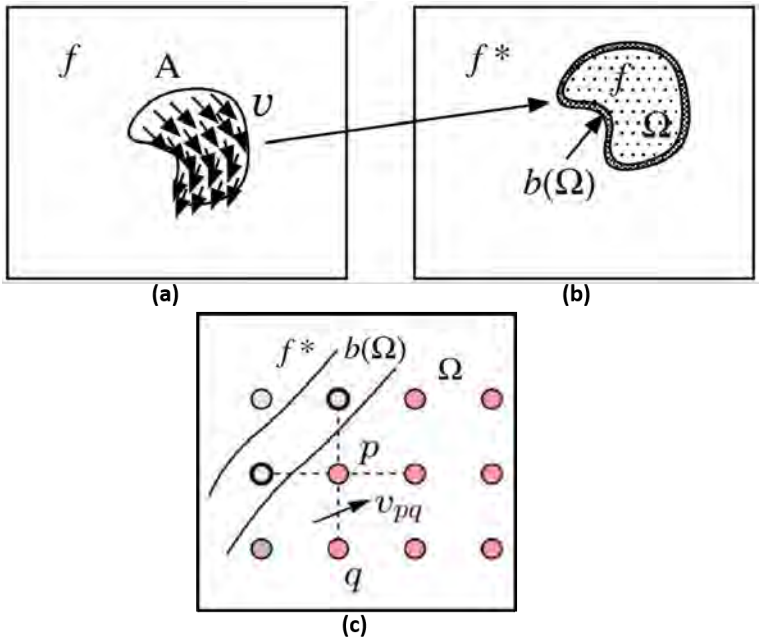

Figura 2.13: Ilustração da técnica de Poisson para suavizar a descontinuidade das bordas da região de interesse na imagem final (Adaptado de [91]). (a): representação da imagem fonte com o gradiente v da região de interesse A. (b): a região de interesse A copiada para área Ω na imagem destino, sendo $b(\Omega)$ a borda de Ω. (c): vizinhança do *pixel* p em Ω e $b(\Omega)$.

A versão discreta pode ser vista na Equação 2-13:

$$|N_p| \cdot f_p - \sum_{q \in N_p \cap \Omega} f_q = \sum_{q \in N_p \cap b(\Omega)} f_q^{*} + \sum_{q \in N_p} v_{pq} \tag{2-13}$$

Onde p é o *pixel* que pertence a Ω, com valor f_p. O conjunto de *pixels* vizinhos de p é representado por N_p (conectividade igual a $N_4(p)$), sendo $|N_p|$

o tamanho do conjunto. Assim como, f_q^* é o valor do *pixel* na imagem destino e v_{pq} é o gradiente do *pixel* p para um vizinho (veja Figura 2.13(c)), conforme Equação 2-14:.

$$v_{pq} = V\left(\frac{p+q}{2}\right) \bullet \vec{pq} \tag{2-14}$$

2.2 Aprendizado de Máquina

O termo aprendizado de máquina se refere a algoritmos que são capazes de aprender uma tarefa específica examinando dados fornecidos como entrada [44]. Uma definição para algoritmos de aprendizado de máquina foi apresentada por Mitchell [96] como:

> "Um programa de computador aprende com a experiência E em relação a alguma classe de tarefas T com desempenho medido por P, se seu desempenho sobre as tarefas T, mensurado por P, melhora de acordo com a experiência E."(tradução nossa).

Uma tarefa T é a previsão que deve ser feita a partir da análise dos dados de entrada. Os dados de entrada definem a experiência E, sendo efetuado um aprendizado de padrões sobre estes dados para melhorar previsão. Os resultados da previsão são avaliados por P, que indica se será necessária uma nova análise dos dados, até que se obtenha o nível considerado adequado de previsão de acordo com o contexto do problema [49].

O processo de aprendizado é geralmente dividido em duas etapas. A primeira é o treinamento, onde os dados de entrada analisados pelo algoritmo na busca de padrões. A segunda etapa é chamada de teste. Neste caso, novos dados de entrada (que não foram usados no treinamento) são recebidos pelo algoritmo que irá utilizar os padrões aprendidos na fase de treinamento para computar uma previsão para cada dado novo.

Quando o aprendizado é chamado de supervisionado [123], significa que os dados usados no treinamento trazem alguma indicação sobre o que se deseja prever. Por exemplo, se o objetivo do sistema é prever a localização de pessoas em imagens, os dados de treinamento serão compostos das imagens de fato e também de respectivas anotações que indiquem a localização de cada pessoa em cada imagem. Deste modo, o sistema tem sempre uma referência (chamado de anotação ou etiqueta ou *ground truth*) para efetuar uma validação do aprendizado antes da etapa de testes, por exemplo.

Por outro lado, existe o método não supervisionado [123] em que a entrada do algoritmo não fornece nenhum indicativo da estrutura interna dos dados de treinamento. Neste caso, o algoritmo avalia as características que tornam os dados mais similares ou menos similares entre si, formando grupos de dados parecidos que representam uma categoria [49]. Considerando o exemplo de localizar pessoas a partir de imagens, o método não supervisionado se empenharia em marcar áreas da imagem com pessoas, formando um grupo por causa das semelhanças apresentadas por estas áreas.

Considerando o aprendizado de máquina supervisionado, as variáveis de entrada (características) do algoritmo podem ser chamadas de $x^{(i)}$. Cada variável $x^{(i)}$ é um vetor com n características, especificado como $x^{(i)} = x_1, x_2, ..., x_n$. A anotação ou etiqueta é representada por $y^{(i)}$, indicando o que deve ser previsto. Uma amostra de treinamento será formado pelo par $(x^{(i)}, y^{(i)})$, e o conjunto completo dos dados de treinamento será uma lista com m amostras de treinamento, i.e. $\left\{(x^{(i)}, y^{(i)}); i = 1, ..., m\right\}$. Os algoritmos de aprendizado supervisionado buscam encontrar a relação entre $\boldsymbol{\chi}$ (Equação 2-15) e $\boldsymbol{\gamma}$ (Equação 2-16).

$$\boldsymbol{\chi} = \{x^{(i)} \mid i = 1, ..., m, x^{(i)} \in \mathbb{R}^d\} \tag{2-15}$$

$$\boldsymbol{\gamma} = \{y^{(i)} \mid i = 1, ..., m, y^{(i)} = h(x^{(i)})\} \tag{2-16}$$

Pode-se dizer que o aprendizado supervisionado se empenha em conseguir estabelecer um modelo preditivo estimando uma função hipótese $h : \boldsymbol{\chi} \rightarrow \boldsymbol{\gamma}$ [76]. Isto é, $h(x)$ é uma "boa" previsão para o respectivo valor de y. A Figura 2.14 ilustra este processo.

De acordo com a Figura 2.14, o conjunto de dados de treinamento está representado pela matriz A, onde cada linha corresponde a uma amostra m composta por n características de interesse. O algoritmo irá analisar as amostras com o objetivo definir um modelo a partir das entradas ($\boldsymbol{\chi}$), ajustando o aprendizado de acordo com semelhança em relação a cada respectiva amostra com a referência correta (anotação ou etiqueta ou *ground truth*) ($y^{(i)}$). No fim do processo, $\boldsymbol{\gamma} = h(\boldsymbol{\chi})$ é a melhor predição de saída encontrada. A partir deste ponto, o modelo está treinado e pode receber um novo conjunto de entradas ($\boldsymbol{\chi}$) e efetuar as respectivas previsões ($\boldsymbol{\gamma}$).

Durante o processo de aprendizado ajustes são realizados de acordo com a proximidade dos valores de previsão em relação a realidade. Tal ajuste é efetuado segundo uma função de custo. Como exemplo, considerando um algoritmo supervisionado de regressão linear, a relação entre variáveis de

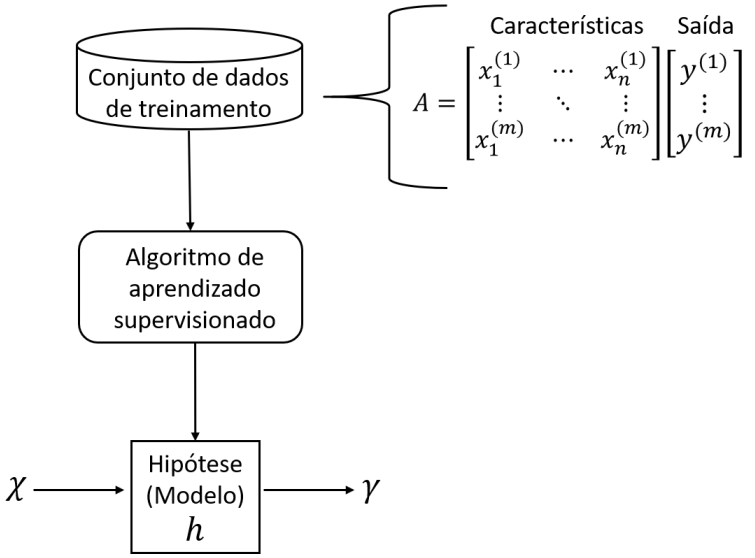

Figura 2.14: Representação das etapas do processo de aprendizado de máquina supervisionado. Adaptado de [40].

entrada e previsão de saída deve ser maximizada levando em conta uma seleção de valores de parâmetros, como descrito na Equação 2-17:

$$h(x^{(i)}) = \theta_0 + \theta_1 x_1^{(i)} + \theta_1 x_2^{(i)} + \cdots + \theta_n x_n^{(i)} = \gamma^{(i)} \tag{2-17}$$

Os parâmetros indicados por θ são os coeficientes da Equação 2-17 e no contexto de aprendizado de máquina são chamados de pesos. Para produzir previsões mais precisas o algoritmo deve selecionar um conjunto de pesos $\{\theta_i, i = 1, \cdots, n\}$, considerando x_n características em cada $x^{(i)}$ amostra, que mais se aproxime de γ.

Para se obter os pesos θ_i é calculada a diferença entre a previsão do algoritmo e os valores reais ($d = y^{(i)} - \gamma^{(i)}$), ou seja, cada previsão $\gamma^{(i)} = h(x^{(i)})$ é comparada com a respectiva referência $y^{(i)}$. Uma forma de medir a proximidade das previsões do algoritmo com a realidade é por meio de uma função de custo que efetua a soma das diferenças entre o valor previsto e o valor real das amostras, conforme visto na Equação 2-18:

$$J(\theta) = \sum_{i=1}^{m} [y^{(i)} - h(x^{(i)})]^2 \tag{2-18}$$

Utilizando m amostras (i.e. quantidade de pontos na regressão linear) a Equação 2-18 mede para cada valor do parâmetro θ o quão perto as previsões de $h(x^{(i)})$ estão dos valores reais $y^{(i)}$. Uma estratégia muito utilizada para as escolhas dos parâmetros θ_i de modo a minimizar $J(\theta)$ é o uso do método de gradiente descendente [47]. A ideia geral de descida do gradiente em

aprendizado de máquina é o ajuste iterativo dos parâmetros para minimizar a função de custo.

2.2.1 Redes Neurais Artificiais

Uma rede neural artificial é um modelo computacional que apresenta características de funcionamento semelhantes às encontradas em redes neurais biológicas [39, 123]. Uma forma de descrever a estrutura de uma rede neural artificial é compará-la a um grafo, onde os nós são chamados de neurônios e as arestas fazem a conexão da saída de um neurônio para a entrada de outro. Deste modo, o elemento mais básico de uma rede neural é o neurônio (*perceptron*) [120], representando um nó do gráfico. O modelo do *perceptron* está descrito na Figura 2.15, onde recebe n entradas x e retorna uma saída y que será conectada a outros neurônios da rede. Neste contexto, o processamento ocorre nos neurônios e sinais são enviados por meio das arestas, sendo que cada aresta possui um valor de peso w associado [39].

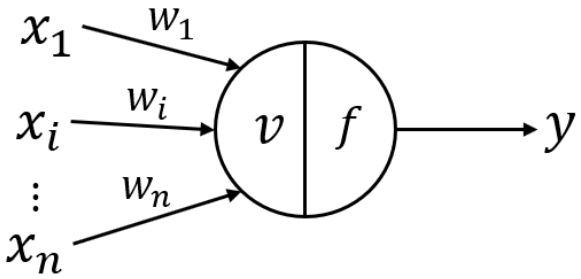

Figura 2.15: Representação do modelo de neurônio artificial (perceptron). Adaptado de [96].

Dados os valores de entrada do neurônio x_1 até x_n (Figura 2.15) e cada peso w_1 até w_n associado as respectivas arestas, a soma ponderada v é obtida pelo produto da entrada x e do peso w conforme a Equação 2-19:

$$v = \sum_{i=1}^{n} \omega_i \cdot x_i \tag{2-19}$$

Além disso, cada neurônio define seu valor (sinal) de saída de acordo com uma função de ativação f (Figura 2.15). Por exemplo, a entrada de um neurônio pode ser obtida a partir da soma ponderada dos pesos das arestas de saída de todos os outros neurônios que estão conectados a ele [123] (veja Figura 2.16). Uma função de ativação que pode ser utilizada como exemplo é a tangente hiperbólica [81], como descrito na Equação 2-20:

$$f(v) = \alpha.tanh(\beta.v) \tag{2-20}$$

Nesta função α e β são constantes[3]. Outras funções de ativação com uso difundido são logística (*sigmoid*) e linear retificada ($ReLU$). Por fim, conforme ilustrado na Figura 2.15, a saída final do neurônio relacionada aos valores de entrada será $y = f(v)$.

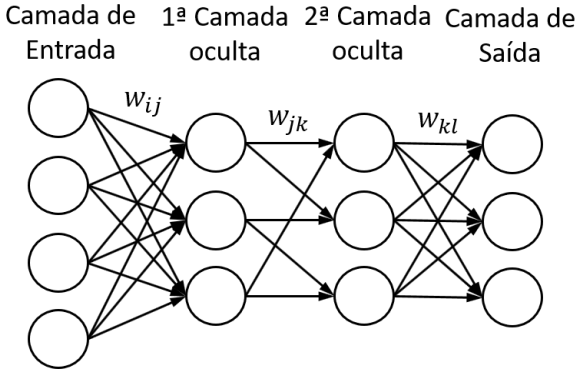

Figura 2.16: Representação de uma rede *perceptron* de múltiplas camadas (*Multilayer Perceptron* (MLP)). Adaptado de [77].

Os modelos de redes neurais artificiais apresentam uma arquitetura em camadas composta por vários neurônios conectados entre si. Esta estrutura é empregada nas redes *Perceptron* de Múltiplas Camadas [120] (*Multilayer Perceptron* (MLP))[4], por exemplo. A ilustração de arquitetura MLP é apresentada pela Figura 2.16, onde são exibidas três tipos camadas: camada de entrada, camada oculta e camada de saída [77]. Nesta arquitetura as informações avançam da camada de entrada para camada de saída (*feed-foward*) passando por uma ou mais camadas ocultas. A Figura 2.16 apresenta somente duas camadas ocultas, deste modo, o sinal em um neurônio i da camada de entrada é transferido para um neurônio j da primeira camada oculta, tendo em conta o peso w_{ij}. Após, o sinal do neurônio j é enviado para o neurônio k e por fim, até o neurônio da camada de saída l.

2.2.1.1 Redes Neurais Convolucionais

Uma rede neural convolucional é semelhante a uma rede neural tradicional no sentido de que sua arquitetura também é composta por neurônios unidos por um padrão de conexões, com respectivos pesos associados e funções de ativação [114]. Estas redes convolucionais são muito utilizadas no contexto de reconhecimento de padrões em imagens, pois esta arquitetura é especialmente projetada para processamento de dados organizados em topologia de grade [47].

[3]LeCun [81] sugere os valores $\alpha = 1.7159$ e $\beta \approx 0.6666$.
[4]Também conhecida como redes neurais *feed-foward*.

O termo convolucional é utilizado devido a operação matemática de convolução que é aplicada em pelo menos uma camada [47]. A estrutura básica desta arquitetura é formada por uma sequência de camadas convolucionais e *pooling*, finalizando com uma camada completamente conectada (*Fully Connected* (FC)).

A operação convolução atua sobre uma imagem de entrada da rede, por exemplo, gerando uma nova imagem onde determinadas características são mais evidentes. Este destaque de características é estabelecido por um filtro (*kernel*) utilizado na convolução. A nova imagem resultante, chamada de mapa de características (*feature map*, *activation map*), desta primeira convolução se torna uma nova entrada para uma próxima operação de convolução.

Os filtros são pequenas matrizes, i.e., em dimensões menores quando comparados a imagem de entrada da rede, onde cada posição da matriz possui uma valor com o peso, sendo o equivalente ao usado na rede neural artificial (veja Seção 2.2.1). Na operação de convolução o filtro se desloca pela imagem de entrada de modo que é executada um somatório das multiplicações de cada posição das matrizes (filtro e imagem). Neste caso, são considerados os valores do filtro e somente os valores da região da imagem sobreposta por este *kernel*.

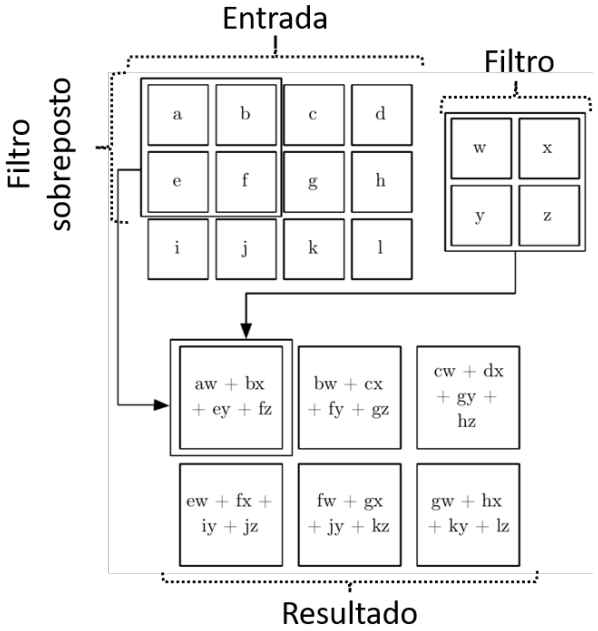

Figura 2.17: Ilustração da operação de convolução: filtro sobreposto na imagem de entrada com resultado obtido a partir do somatório das multiplicações. Adaptado de [47].

Uma ilustração da operação de convolução pode ser vista na Figura 2.17. O filtro está sobreposto na imagem delimitando a região considerada na operação. O somatório das multiplicações é executado com os valores do filtro

e da região sobreposta da entrada gerando parte do resultado. Na sequência, o filtro se desloca pela imagem e a mesma operação é repetida até que se alcance o resultado final.

A convolução discreta sobre uma imagem I de duas dimensões como entrada pode ser definida como na Equação 2-21 [47]:

$$S[i,j] = (I * F)[i,j] = \sum_m \sum_n I[i-m, j-n]F[m,n] \qquad (2\text{-}21)$$

Onde S é o mapa de características e F é o filtro também bidimensional. Conforme a matriz do filtro F se desloca sobre a imagem I, a soma ponderada para cada posição i,j é calculada para a saída $S[i,j]$.

Além das camadas de convolução é comum o uso de camadas de *pooling*. Estas atuam como um redutor de dimensionalidade do mapa de características gerado pela convolução, diminuindo o uso de memória e a carga computacional. A sua forma de atuar também é semelhante a da convolução, porém os utilizam funções de agregação para determinar os valores do resultado ao invés de pesos dos filtros. O método conhecido como *Max Pooling*, por exemplo, gera uma saída com o valor máximo presente na área sobreposta pelo filtro na entrada, ao invés de executar o somatório das multiplicações.

Após várias camadas convolucionais que extraíram sucessivos mapas de características, a rede faz uso de uma camada FC (mesma camada da rede MLP). Devido a natureza da camada FC, todos os neurônios desta estão conectados a todos os neurônios da camada anterior, além dos respectivos valores de pesos. Assim, a camada completamente conectada apresenta os valores de predição de cada classe na saída da rede.

2.2.1.2 Geração de Imagens - GAN

As Redes Adversárias Generativas (*Generative Adversarial Network* (GAN)) são uma subclasse de modelos generativos com impressionante potencial de aprendizado [61, 87]. Um exemplo de aplicação das redes GAN pode ser visto no contexto médico com geração de imagens sintéticas complexas [50, 54]. A rede GAN introduz o conceito de treinamento com duas redes adversárias, uma rede geradora G e outra discriminadora D. A ideia de funcionamento da rede GAN é aprender as características de um conjunto de dados de entrada x e gerar novas amostras sintéticas semelhantes às presentes no conjunto de entrada.

A rede geradora G retorna uma nova amostra $G(z)$ a partir de um vetor ruído z, obtido a partir de uma distribuição normal. Já a rede discriminadora D irá receber como entrada amostras reais x (i.e., dados de treinamento)

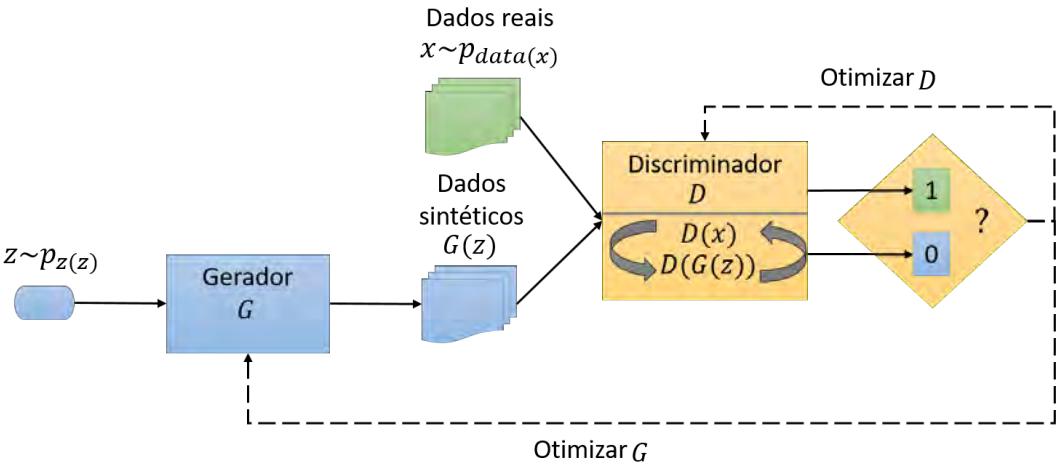

Figura 2.18: Representação do fluxo de treinamento de uma rede neural do tipo GAN. G representa a rede geradora e D a rede discriminadora.

e também amostras sintéticas $G(z)$. A saída da rede D é um valor que indica a probabilidade da amostra que está sendo avaliada ser real, ou seja, o discriminador tenta separar dados reais dos dados falsos.

O gerador G deseja maximizar o erro do discriminador D, enquanto o discriminador D é treinado para minimizar o erro da tarefa de classificar uma imagem como real ou sintética. O gerador G é treinado para capturar a distribuição real dos dados de entrada, pois uma distribuição próxima do real aumentará a probabilidade do discriminador D classificar uma amostra sintética como real. Os valores da função de custo na saída do discriminador são utilizados como métrica de referência para o treinamento do gerador G e também do discriminador D, de modo que a eficiência das redes melhora conforme o treinamento avança.

A Figura 2.18 apresenta uma visão geral do processo que envolve o treinamento da rede GAN, onde $x \sim p_{data}(x)$ é a distribuição de probabilidade dos dados reais, e $z \sim p_z(z)$ indica a distribuição da entrada z. No início do treinamento uma amostra sintética $G(z)$ é gerada a partir de z pelo gerador G. A rede D recebe como entrada $G(z)$ e x alternadamente, i.e., treina com uma amostra real e após treina com uma amostra sintética. A saída de D indica a probabilidade de a amostra ser real, i.e. valor mais próximo de 1 ou mais próximo de 0 no caso de amostra sintética. Ambas as redes são otimizadas durante o treinamento. A Equação 2-22 [48] expressa a relação entre G e D utilizada para treinar ambas as redes:

$$\min_G \max_D V(D,G) = \mathbb{E}_{\boldsymbol{x} \sim p_{data}(\boldsymbol{x})}[logD(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z} \sim p_{\boldsymbol{z}}(\boldsymbol{z})}[log(1 - D(G(\boldsymbol{z})))] \quad (2\text{-}22)$$

A Equação 2-22 pode ser examinada em duas partes[5]. A primeira apresentando a operação de maximizar o discriminador ($\max_D V(D,G)$). Esta estratégia faz sentido pois quando D está no início do treinamento, então $logD(x) \approx -\infty$ e $log(1 - D(G(\boldsymbol{z}))) \approx -\infty$. Porém, se a rede discriminadora D já estiver bem treinada então $logD(x) \approx 0$ e $log(1 - D(G(\boldsymbol{z}))) \approx 0$, que é o seu desempenho ótimo.

A segunda parte considera a rede geradora G, cujo o objetivo é minimizar $\min_G V(D,G)$. O termo $\mathbb{E}_{\boldsymbol{x}\sim p_{data}(\boldsymbol{x})}[logD(\boldsymbol{x})]$ não faz referência ao gerador G, sendo dispensável nesta observação. Então, o objetivo considerando o segundo termo $\mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[log(1-D(G(\boldsymbol{z})))$ é minimizar a probabilidade do discriminador classificar amostras sintéticas como sintéticas, i.e. evitar que $D(G(z)) \approx 0$. Quando a rede G ainda não aprendeu a gerar amostras semelhantes aos dados reais, então a saída de $D(G(z))$ será 0 e por isso $log(1-D(G(z))) \approx 0$. Porém, se o gerador G estiver bem treinado conseguirá enganar a rede discriminadora D, portanto, $D(G(z))$ retornará 1 e $log(1 - D(G(z)))$ será $-\infty$. Ou seja, para o gerador enganar o discriminador é preciso conseguir que $\min_G V(D,G)$, pois é a situação em que D retorna 1 para uma amostra $G(z)$.

Uma das variações do modelo GAN é a chamada Rede Adversária Generativa Condicional (*Conditional Generative Adversarial Network* (CGAN)) [42, 95]. Uma rede CGAN é similar a rede GAN, com o diferencial de permitir que sejam utilizados dados adicionais para condicionar a geração das imagens de acordo com certas características. Para isso, uma informação extra y (e.g. etiquetas de classes ou máscaras binárias) é utilizada como entrada adicional para as redes geradora G e discriminadora D. A distribuição randômica indicada por z é combinada a informação condicional y permitindo um certo nível de controle sobre o que será gerado pela rede. A Equação 2-23 apresenta estas modificações para inclusão do parâmetro adicional y [42]:

$$\min_G \max_D V(D,G) = \mathbb{E}_{\boldsymbol{x},\boldsymbol{y}\sim p_{data}(\boldsymbol{x},\boldsymbol{y})}[logD(\boldsymbol{x},\boldsymbol{y})]+\mathbb{E}_{\boldsymbol{y}\sim p_{\boldsymbol{y}},\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[log(1-D(G(\boldsymbol{z},\boldsymbol{y}),\boldsymbol{y}))] \tag{2-23}$$

Neste caso, $G(z,y)$ e $D(x,y)$ demonstram que as respectivas saídas do gerador e discriminador estão condicionados aos dados em y. A Figura 2.19 apresenta um diagrama do fluxo de treinamento para a rede do tipo CGAN.

O gerador utiliza a distribuição randômica $z \sim p_{z(z)}$ e as características apresentadas por $y \sim p_y$ para criar uma amostra sintética $G(z,y)$. O objetivo de G é produzir uma saída com distribuição mais próxima do real possível que também esteja de acordo as condições y. Já o discriminador recebe

[5]Observe que $log(0) = -\infty$ e $log(1) = 0$.

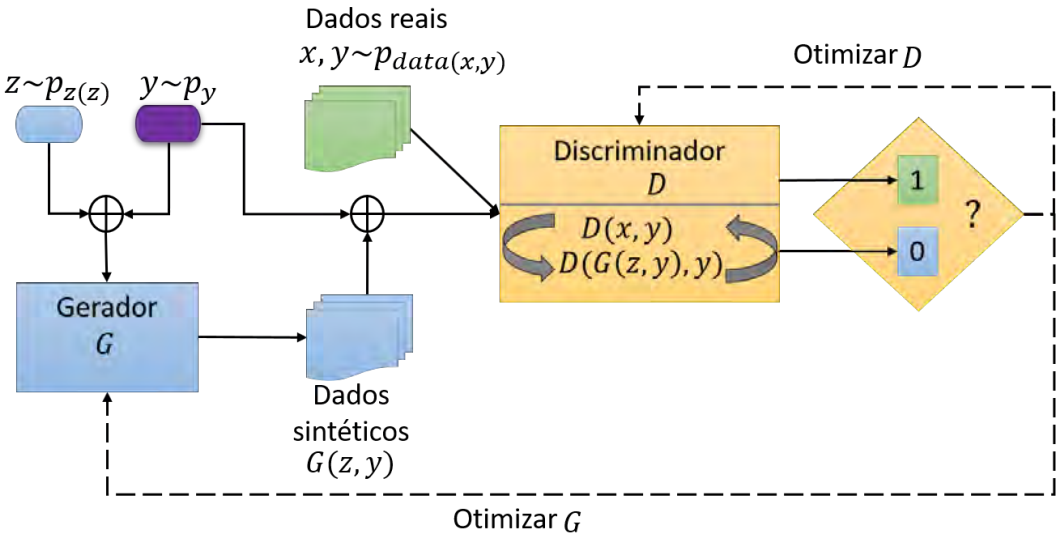

Figura 2.19: Fluxo de treinamento de acordo com o modelo de rede CGAN. G representa a rede geradora e D a rede discriminadora.

alternadamente $D(x, y)$, que são amostras reais com informações condicionais e amostras sintéticas $D(G(z, y), y)$ com o respectivo y usado para gerá-las. Com amostras reais o discriminador aprende a identificar se há distribuição semelhante a real nos dados e se existe correspondência com a informação condicional y. Quando D recebe exemplos criados por G é efetuada uma avaliação de correspondência entre o par amostra sintética e informação condicional, com intuito de diferenciá-los dos casos reais. O discriminador precisa aprender a rejeitar todas as amostras que são sintéticas e que não estão de acordo com y enquanto aceita pares reais de cada amostra com respectivo y. Em G, y funciona como um direcionador para criação de uma amostra com características específicas ao mesmo tempo que em D orienta a distinguir elementos de acordo com os dados em y.

2.2.1.3 Segmentação de Imagens - U-NET

A segmentação é basicamente uma abordagem de processamento de imagem que nos permite separar objetos do restante da imagem. Este processo particiona a imagem em regiões, de modo que *pixels* pertencentes a determinadas áreas de interesse são marcados e com isso é possível destacá-los do fundo da imagem, por exemplo.

A arquitetura U-Net [119] é uma abordagem utilizada para segmentação baseada em redes neurais convolucionais e que apresenta excelente desempenho geral na segmentação de imagens médicas. O modelo U-net está baseado em duas partes denominadas como codificador e decodificador. Características

espaciais são capturadas pelo codificador e utilizadas pelo decodificador para a construção de um mapa de segmentação a partir destas características [65]. Esta rede faz somente uso de camadas de convolução, não utilizando camadas totalmente conectadas permitindo o uso de imagens com tamanho variado como entrada.

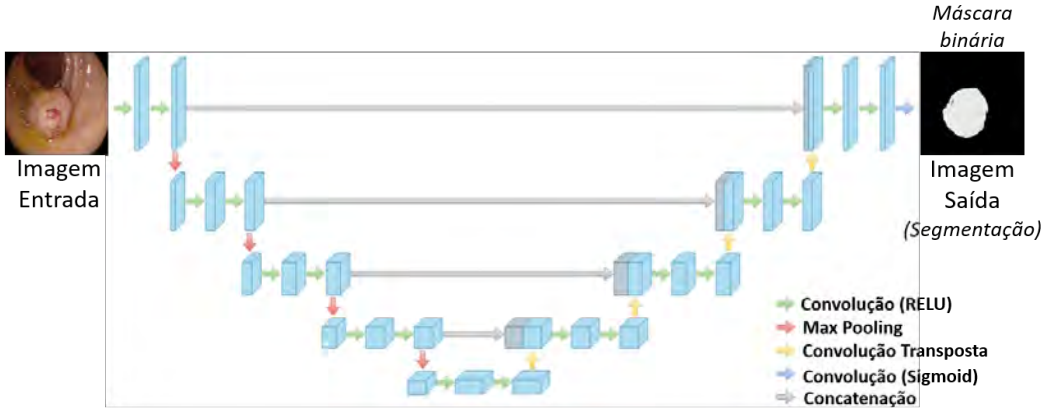

Figura 2.20: Representação da arquitetura U-net. As caixas são os mapas de características e as setas representam as diferentes operações. Adaptado de [65].

A Figura 2.20 ilustra as camadas da rede U-net e as operações. O lado esquerdo representa codificador que recebe a imagem de entrada e aplica sequencialmente operações de convolução e *max pooling* criando representações das características da imagem em diversos níveis. O decodificador é composto pelas camadas e operações no lado direito da ilustração. As saídas das convoluções (mapas de características) de mesmo nível (no codificador) são concatenadas em todas as etapas do decodificador. Isto ajuda a fornecer informações de localização perdidas na operação de *max pooling* [35]. Ao final do processo os *pixels* que pertencem ao objeto de interesse são destacados formando uma máscara binária, por exemplo.

2.2.1.4 Detecção de Objetos - *Faster* R-CNN

A detecção de objetos é uma tarefa essencial para extração de informações em imagens. Métodos recentes baseados em aprendizado profundo, como *Faster R-CNN* [112], aprimoram não somente o desempenho mas também o tempo de processamento neste contexto de detecção. A rede neural convolucional *Faster* R-CNN é um modelo que apresenta bons resultados sobre diversos tipos de imagens em conjuntos de testes [70].

O funcionamento desta rede ocorre em três etapas. Na primeira etapa, uma imagem de entrada é processada por uma rede convolucional criando um

mapa de características, que é retornado como saída. Esta saída é direcionada para a segunda etapa que consiste em uma rede de proposição de regiões (*Region Proposal Network* (RPN)) [112]. Tais regiões são áreas candidatas da imagem que podem conter o objeto de interesse da detecção. Para cada região (*Region Of Interest* (ROI)) é definida uma caixa delimitadora (*bounding box*) que engloba um provável objeto candidato [9]. Um conjunto de regiões propostas é encaminhado para uma camada de *pooling*, na etapa seguinte, que modifica a representação de cada uma para um vetor de características. Estes vetores são utilizados como entrada para o classificador e o regressor, ambos compostos por uma série de camadas FC.

Para cada região proposta são extraídas características usadas na classificação (estima probabilidade) e na regressão, que nesta última determinam as coordenadas da caixa delimitadora. Isto significa que a rede *Faster* R-CNN apresenta duas saídas. A primeira é um valor de probabilidade estimado para cada objeto e a segunda é formada pelas coordenadas das caixas delimitadoras. Assim, uma imagem de entrada é avaliada e como resposta a rede retorna uma lista de caixas delimitadoras (indicando partes da imagem), onde cada uma possui um valor associado, usado para indicar a probabilidade de cada região pertencer uma classe de objeto ou ao fundo da imagem [9].

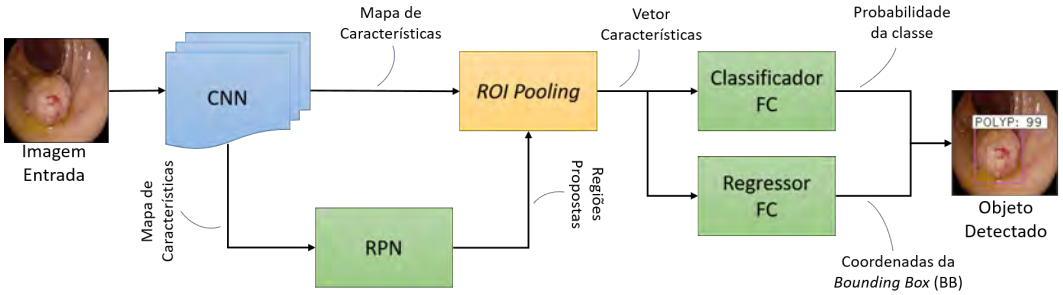

Figura 2.21: Ilustração da arquitetura *Faster* R-CNN.

As camadas convolucionais compartilham parâmetros em comum entre a RPN e o classificador, sendo treinadas em conjunto [131]. A rede de proposição de regiões indica quais são as melhores regiões que devem ser avaliadas pelo classificador para detecção de objetos [38]. A Figura 2.21 apresenta as etapas da rede *Faster* R-CNN. A rede CNN cria um mapa de características a partir da imagem de entrada. A proposição de regiões é efetuada pela rede RPN. Assim, somente as regiões (representadas por caixas delimitadoras) com alta probabilidade de conter um objeto são inseridas em uma lista. A partir do mapa de características e da lista de regiões o objeto é detectado, pois a rede retorna à probabilidade do objeto pertencer a uma classe juntamente com quatro valores de coordenadas da respectiva caixa delimitadora.

2.3 Métricas de Avaliação

O contexto desta tese está relacionado com a detecção de lesões do cólon conhecidas como pólipos. Baseado no estudo de Bernal et al. [14], a detecção desta lesão em uma imagem é a capacidade de determinar a presença de um pólipo em uma imagem de colonoscopia. Uma vez determinada a presença, a localização da lesão dentro da imagem também deve ser indicada. Os experimentos conduzidos nesta tese apresentam a localização das lesões por meio da segmentação (Capítulo 4) e detecção (Capítulo 5). Na segmentação, os *pixels* referentes a região do pólipo são destacados e no caso do experimento de detecção, a localização é determinada por caixa delimitadora que engloba a lesão. Métricas utilizadas para avaliação dos resultados de segmentação e detecção incluem mas não se limitam a IoU (*Intersection Over Union*), precisão e revocação, estas últimas obtidas por meio dos valores de verdadeiro positivo (*True Positive* (TP)), falso positivo (*False Positive* (FP)) e falso negativo (*False Negative* (FN)).

O uso da métrica IoU[6] é geralmente empregado em aplicações de segmentação e detecção de objetos [37, 152]. Considerando dois conjuntos V e P que representam as áreas dos objetos de interesse, onde V representa a área real demarcada[7] na imagem e P a resposta prevista pela abordagem em determinado teste, IoU pode ser definido como [152]:

$$IoU(V, P) = \frac{\|V \cap P\|}{\|V \cup P\|} \tag{2-24}$$

Na Equação 2-24 os operadores $\cup$ e $\cap$ representam união e interseção, respectivamente. Além disso, $\|V\|$ é a norma de V, que indica a quantidade de *pixels* demarcados. O valor de IoU é dado na faixa de 0.0 até 1.0. Uma segmentação ou detecção P é considerada boa se quando comparada a área real V alcança um valor de IoU mais próximo de 1.0.

Esta mesma lógica pode ser aplicada ao caso de detecção, por exemplo, a área de interseção entre as caixas delimitadoras determina o valor do IoU (cor cinza), conforme apresentado na Figura 2.22. A área que representa a resposta da abordagem em relação a localização do objeto de interesse está ilustrada pela caixa em vermelho, enquanto a localização real para comparação é indicada pela caixa em verde. Novamente, o valor de IoU é utilizado como métrica para interseção das áreas sendo, neste caso, 0.41.

Em outras palavras, uma previsão correta gera um verdadeiro positivo (*True Positive* (TP)), enquanto que uma incorreta é considerada um falso

[6] Também conhecida como *Jaccard index* ou *Jaccard similarity coefficient* [69].
[7] Isto é imagem *ground truth*, anotação ou etiqueta.

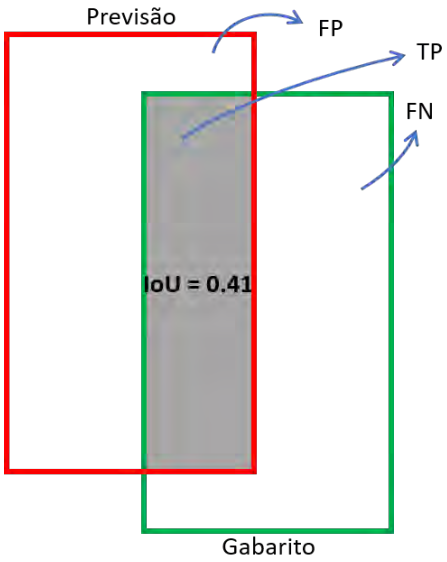

Figura 2.22: Ilustração da métrica IoU no caso de detecção de objeto na imagem.

positivo (*False Positive* (FP)) [52]. Já uma ausência de indicação especifica um falso negativo (*False Negative* (FN)). Além do uso da métrica IoU é preciso definir um limiar L (*threshold*), de forma que o valor de IoU em comparação a este limiar indica a conclusão final para cada caso de detecção. Pode-se dizer que [37, 52, 109, 124]:

- TP = É uma detecção correta onde IoU $\geq L$ (bons e resultados);
- FP = É uma detecção incorreta onde IoU $< L$ (resultados ruins);
- FN = Uma parte ou todo da área imagem *ground truth* não faz parte da previsão (algo relevante não foi encontrado);

A partir disso, é possível calcular outras métricas úteis como, por exemplo, precisão (*precision*) e revocação (*recall*) descritas nas Equações 2-25 e 2-26:

$$\text{Precisão} = \frac{TP}{TP + FP} = \frac{TP}{\text{Todas as previsões (respostas)}} \tag{2-25}$$

$$\text{Revocação} = \frac{TP}{TP + FN} = \frac{TP}{\text{Todos os gabaritos}} \tag{2-26}$$

No caso da segmentação, uma opção para se obter valores de precisão e revocação é a comparação a nível de *pixel* por meio das máscaras binárias, sem a comparação do valor de IoU com o limiar L (*threshold*). Assim, são utilizados os *pixels* de duas máscaras binárias, uma é a segmentação prevista

por um método de localização de objetos e a outra é a própria marcação real da área do objeto. A máscara binária chamada de imagem *ground truth* [10, 14], indica a área correta a ser segmentada na imagem de entrada, i.e., corresponde à região real do objeto. Supondo que determinada abordagem apresente como resposta para tarefa de localizar pólipos a imagem previsão, onde a segmentação é determinada pelos *pixels* em branco, é possível comparar a previsão com a imagem *ground truth*.

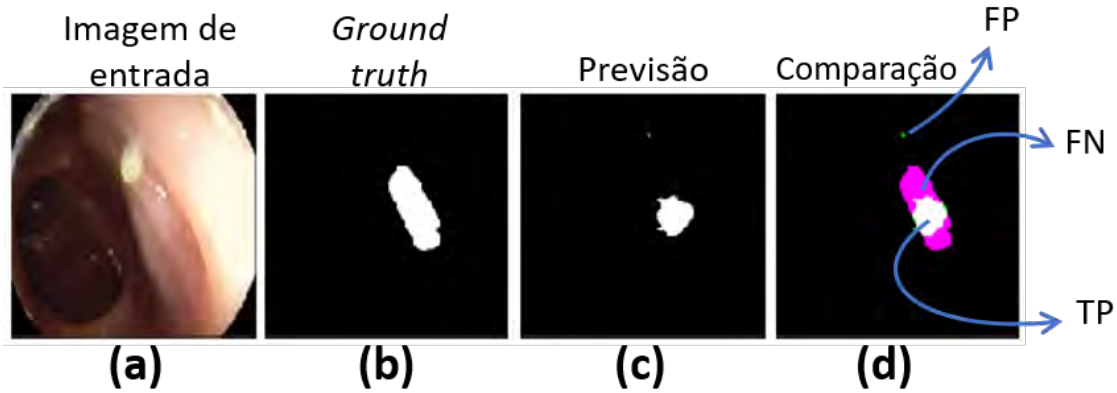

Figura 2.23: Ilustração da comparação da imagem de previsão com a imagem *ground truth*. (a): imagem de entrada para segmentação. (b): segmentação correta da área que corresponde ao objeto de interesse na imagem (a). (c): área segmentada prevista em relação a região do objeto de interesse em (a). (d): sobreposição das áreas do *ground truth* (b) e previsão (c).

Na comparação apresentada na Figura 2.23, as áreas da imagem *ground truth* e da previsão estão sobrepostas de modo que os *pixels* em branco indicam que há uma interseção entre os *pixels* da *ground truth* e da previsão (TP). Os *pixels* em verde representam áreas previstas que não estão na interseção com a imagem *ground truth* (FP). Além disso, os *pixels* em rosa mostram a região em que a imagem de previsão não identificou em relação a imagem *ground truth* (FN).

Assim, a precisão indica a capacidade da abordagem em identificar somente os objetos de interesse (TP), por isso os erros (FP) prejudicam esta métrica. A revocação é a habilidade da abordagem em encontrar todos os objetos de interesse. Neste caso, os FP's não influenciam na métrica. Esta é uma métrica interessante para o contexto médico pois penaliza os FN's, visto que é perigoso não detectar uma doença (reduzir os FN's).

2.4 Imagens de Colonoscopia

No contexto médico, a visualização das lesões no cólon ocorre no momento do exame de colonoscopia. Este é um procedimento no qual o médico introduz um colonoscópio com o objetivo de visualizar o interior do cólon [103]. A câmera presente neste equipamento envia as imagens em tempo real para um monitor de vídeo. Por sua vez, o médico avalia as imagens para verificar se existem lesões como úlceras e pólipos, por exemplo. Atualmente, há variações no tipo de exame (modalidade) de colonoscopia, e.g., cápsula de cólon [107] e colonoscopia virtual [51]. Contudo, nesta tese são empregadas imagens da modalidade de colonoscopia convencional que são obtidas a partir do vídeo gerado pelo colonoscópio [51] e somente serão consideradas as lesões do tipo pólipo.

As imagens de colonoscopia podem ser divididas em dois grupos: as imagens positivas (onde há lesões) e negativas (onde não ocorrem lesões). Estas também podem ser chamadas de imagens normais e anormais como no estudo de Tjoa e Krishnan [139].

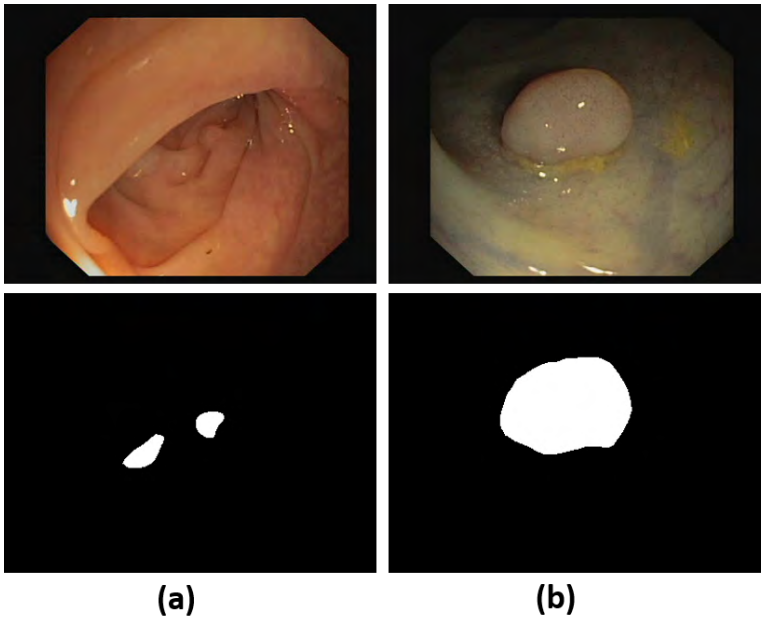

Figura 2.24: Exemplos de imagens de colonoscopia com pólipos pertencentes ao conjunto de imagens CVC-ClinicDB [10]. Colunas (a) e (b): imagens de colonoscopia e suas respectivas imagens *ground truth* (máscaras binárias).

Como exemplo, duas imagens de colonoscopia são apresentadas na Figura 2.24 (linha superior). Estas imagens são quadros extraídos de um vídeo originalmente capturado por colonoscópio. Cada imagem da linha inferior é utilizada como indicador da região na respectiva imagem da linha superior.

Estas imagens indicativas são binárias, i.e., valores dos *pixels* assumem somente dois valores: 0 (preto) e 1 (branco). Os *pixels* em branco indicam o pólipo e a região em preto indica o fundo. As áreas dos pólipos são geralmente demarcadas por médicos especialistas, por isso são utilizadas como imagens de referência (também chamadas de máscaras binárias ou *ground truth*) para comparação com os resultados de uma abordagem de segmentação, por exemplo.

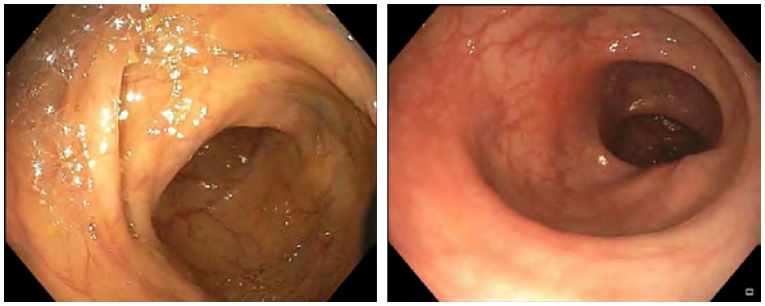

Figura 2.25: Exemplos de imagens de colonoscopia sem pólipos pertencentes ao conjunto de imagens ASU-Mayo [138].

Na Figura 2.25 estão representadas duas imagens sem a presença de pólipos (imagens negativas). Neste caso, as respectivas imagens *ground truth* seriam compostas somente de *pixels* em preto.

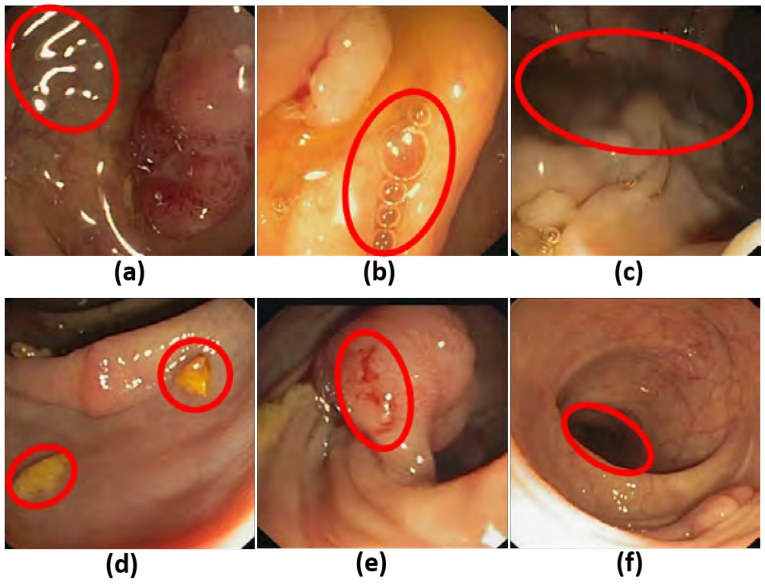

Figura 2.26: Exemplos de elementos presentes nas imagens de colonoscopia [10]. (a): reflexão da luz. (b): bolhas. (c): água. (d): material fecal. (e): sangue. (f): lúmen.

As imagens de colonoscopia apresentam comumente outros elementos como pontos de reflexão da luz, bolhas, água, material fecal, sangue, lúmen [147]. A Figura 2.26 apresenta exemplos destes elementos. Em geral, estas estruturas podem ser confundidas com os pólipos em abordagens de segmentação

ou detecção [13]. Além disso, a presença destes elementos dificulta a visualização de regiões de interesse e ainda prejudica a observação das lesões por parte dos endoscopistas [121].

3
Revisão Bibliográfica

Neste capítulo são apresentados trabalhos encontrados na literatura que tratam da detecção de pólipos por meio da análise de imagens obtidas pela modalidade de colonoscopia convencional[1]. Serão destacadas as características destes estudos como: métodos de detecção, bases de dados de imagens utilizadas para treinamento e métricas de avaliação. Isto permite um melhor entendimento do que está sendo desenvolvido em trabalhos de pesquisa relevantes no contexto desta tese.

Especificamente, os estudos que fazem uso de metodologias de aprendizado de máquina necessitam de uma maior quantidade e variedade de imagens para treinamento. Porém, as bases de dados de disponíveis publicamente podem não ser suficientes para bons resultados, devidos as suas limitações de quantidade e variação de amostras. Assim, tais trabalhos empregam bases de dados combinadas sendo estas públicas ou particulares para favorecer os treinamentos por meio da variação de imagens. A partir desta observação percebeu-se a necessidade de composição de conjuntos de imagens de colonocospia maiores e mais variados, com objetivo de aprimorar os treinamentos em abordagens que fazem uso do aprendizado de máquina.

3.1
Bases de Dados de Imagens de Colonoscopia

Os estudos encontrados na literatura efetuam testes das metodologias empregadas sobre um conjunto de imagens obtidas a partir de endoscópios/colonoscópios (e.g. Fujinon Endoscope, Olympus Colonoscope [23, 74, 99, 144, 145]). Determinados autores fazem uso destas imagens que podem ser reunidas por eles próprios como visto em [68] ou [145], por exemplo.

No entanto, a partir do ano de 2012 foram disponibilizadas bases de dados públicas com imagens de pólipos e ainda as suas respectivas imagens de *ground truth* conforme apresentado na Tabela 3.1. As imagens de *ground truth* são máscaras binárias (*binary masks*) que definem a real posição do pólipo na imagem e são utilizadas para validar os resultados de localização de pólipos nos estudos.

[1]Modalidade de exame que utiliza colonoscópio com luz branca (*white-light colonoscopy*).

Tabela 3.1: Lista com informações das bases de dados de imagens colonoscopia.

Base de dados	**Núm. de imagens**	**Núm. Pólipos Diferentes**	**Resolução (*pixels*)**
CVC-ColonDB[11] [12]	300	15	574 x 500
ETIS-LaribPolypDB[12] [128]	196	44	1225 x 966
CVC-ClinicDB[13] [10]	612	31	384 x 288
ASU-Mayo[14] [138]	19.400	10	Variadas
KVASIR[15] [106]	1000	N/A	720 x 576
CVC-EndoSceneStill[16] [142]	912	46	574 x 500 e 384 x 288

A base de dados CVC-ColonDB [12] não possui imagens que representem a ausência de pólipos (i.e. imagens de uma parte do colon sem lesões). Já a base de dados ASU-Mayo[2] apresenta *frames* sem nenhum pólipo, que pertencem a 10 vídeos curtos. Esta é uma base de dados de vídeos composta por um total de 20 vídeos de colonoscopia com as respectivas imagens de *ground truth* (aproximadamente 19,400 frames) em resoluções variadas (e.g. 712 x 480, 856 x 480, 1920 x 1080). Em 10 vídeos os pólipos estão presentes (aproximadamente 5.200 *frames*) e estão ausentes em outros 10, como citado anteriormente. A CVC-ColonDB é uma base de dados disponível para uso livre com 300 imagens, na qual todas as imagens apresentam pólipos. A base de dados ETIS-Larib Polyp DB [128] também contém somente imagens com lesões, sendo 196 imagens em alta resolução com 1225 x 966 *pixels*.

A base de dados KVASIR [106] apresenta um conjunto de imagens do trato gastrointestinal relacionado as lesões formadas por pólipos e por outras doenças como úlcera, esofagite, síndrome do cólon irritável (colite). São apresentadas também imagens do cólon saudáveis e outras partes gastrointestinais como divisão entre o esôfago e o estômago (*Z-line*), entre o estômago e o duodeno (*pylorus*) e *cecum*. Esta base de dados realiza uma categorização das imagens dividindo-as em classes de lesões distintas. No entanto, a base de dados KVASIR não fornece imagens *ground truth* e na Tabela 3.1 são consideradas somente as informações relativas a classe de imagens dos pólipos.

[2] A base de dados ASU-Mayo possui direitos autorais.
[11] http://mv.cvc.uab.es/projects/colon-qa/cvccolondb
[12] https://polyp.grand-challenge.org/EtisLarib/
[13] https://polyp.grand-challenge.org/CVCClinicDB/
[14] https://polyp.grand-challenge.org/AsuMayo/
[15] http://datasets.simula.no/kvasir/
[16] http://www.cvc.uab.es/CVC-Colon/index.php/databases/cvc-endoscenestill/

Por fim, a base de dados CVC-EndoSceneStill [142] é a composição das bases de dados CVC-ColonDB e CVC-ClinicDB, totalizando 912 imagens extraídas de 44 sequências de vídeo. Foram adicionadas a esta combinação informações (anotações) sobre a localização do lúmen, reflexos da luz, mucosa das paredes do intestino e bordas das imagens.

As imagens pertencentes a cada base de dados são quadros (*frames*) extraídos a partir de vídeos de colonoscopia convencional. As imagens apresentam vários pontos de vista de um mesmo pólipo, assim, a coluna *Número de Pólipos Diferentes* na Tabela 3.1 indica a diversidade de pólipos capturados pelas imagens. É importante destacar que as bases de dados listadas na Tabela 3.1 são relativas à modalidade de exame de colonoscopia convencional que apresentam lesões do tipo pólipo. Existem outros conjuntos de imagens capturados por tipos de modalidades diferentes e que apresentam também outras doenças do trato gastrointestinal. Por exemplo, cápsula de cólon[3], utilizada para gerar a base de dados KID [78].

3.2 Abordagens para Detecção de Pólipos

Nos últimos anos foram publicados vários estudos com o objetivo de detectar pólipos a partir de imagens de colonoscopia. Tais estudos utilizam uma combinação de diferentes metodologias. Uma forma de organizar estes estudos, de acordo com a respectiva abordagem para detecção de pólipos, é dividi-los entre os que utilizam técnicas de aprendizado de máquina e os que fazem uso de outros métodos. A Tabela 3.2 apresenta os estudos selecionados divididos em dois grupos: os que são baseados em aprendizado de máquina e os que não são. Conforme pode ser observado na Tabela 3.2, do total de trabalhos selecionados, 25 deles utilizam o aprendizado de máquina em alguma fase do procedimento de localização dos pólipos[4]. Nove destes empregam arquiteturas *Deep Learning* [72, 73, 75, 90, 92, 105, 137, 148, 150]. Somente 6 artigos não usaram metodologias baseadas em aprendizado de máquina [10–12, 63, 71, 79].

A maior parte dos estudos que utilizam técnicas baseadas em aprendizado de máquina, optaram por *Support Vector Machine* (SVM) como classificador, sendo aplicado em nove trabalhos [1–3, 23, 64, 68, 83, 144, 145]. No entanto, a literatura apresenta diversas metodologias de aprendizado de máquina. Por exemplo, o trabalho de Tajbakhsh et al. [137] apresenta uma abordagem de integração de características dos pólipos como cor, textura, formato e informações temporais de quadros sequenciais em redes neurais convolucionais

[3] *Capsule Endoscopy* (CE).

[4] Alguns estudos fazem uso de mais de uma técnica ou método. Estes foram contabilizados apenas uma vez.

(CNNs). Estas características alimentam três CNNs distintas, onde cada uma é especializada em um tipo de característica. A média das saídas das CNNs são utilizadas para classificar uma área da imagem como pólipo ou não pólipo. Também, o estudo de Maroulis et al. [92] que faz uso de uma rede neural de segmentação em conjunto com extração de características estatísticas e finalmente a classificação destas características com *Multi-Layer Perceptron* (MLP). Karkanis et al. [74] descreve uma abordagem baseada no classificador *Linear Discriminat Analysis* (LDA). Os autores utilizam um modelo estatístico para descrição de texturas, trabalhando para determinar texturas semelhantes.

Estudos mais recentes também apresentam metodologias baseadas em aprendizado de máquina como em [72, 90, 105, 148, 150]. No estudo apresentado por Zhang et al. [150] os autores descrevem uma arquitetura CNN baseada na abordagem YOLO [111] para detecção de pólipos. O processo de detecção considera regiões retornadas pela rede YOLO como pólipos em potencial. Esta abordagem considera o valor de confiabilidade gerado por meio de um filtro de correlação (*Discriminative Correlation Filter* [26]). Este filtro também avalia quadros anteriores para determinar um candidato a pólipo.

Pogorelov et al. [105] apresentam uma abordagem de segmentação baseada no modelo de arquitetura GAN [48]. Mais especificamente no modelo V-GAN [130]. Os autores fizeram uso de vários níveis de avaliação das imagens, i.e., o quadro inteiro, blocos de 128 x 128 *pixels* e também em apenas um *pixel*.

No estudo apresentado por Yuan et al. [148] é descrito um método baseado na detecção de bordas seguido por uma etapa de classificação com um modelo CNN AlexNet [80]. Na etapa de treinamento o método gera um mapa com as bordas a partir de cada quadro. Os autores utilizam uma estratégia para eliminar regiões candidatas baseada no comparativo das distâncias dos pontos centrais das componentes conexas formadas pelas bordas. O comparativo considera os quadros e as imagens *ground truth*. Neste ponto são criados padrões para classes de pólipos e não-pólipos que são utilizados para a classificação nesta etapa de treinamento.

Em Ma et al. [90] foi proposto a utilização do modelo SSD baseado na rede VGG 16 [62]. Os autores modificaram camadas do modelo adicionando estruturas *inception* [132], além do uso de algoritmos de otimização e perda para melhorar eficiência da rede. Esta abordagem aumentou a complexidade da rede sendo capaz de extrair e unir características de variadas dimensões.

O trabalho de Kang e Gwak [72] propõe um método utilizando duas redes Mask R-CNN [58]. Cada rede faz uso de uma estrutura de *backbone* diferente (ResNet50 e ResNet101). Ambos foram pre-treinados com as imagens

do conjunto COCO (*Common Objects in Context*) [85]. Objetivo é combinar o benefício dos *backbones* de maneira que cada rede forneça uma máscara binária indicando a segmentação do objeto, que serão mescladas *pixel* a *pixel* formando a segmentação final.

Os métodos de detecção de pólipos que não estão baseados em aprendizado de máquina procuram diferenciar as regiões de borda entre pólipos, paredes do cólon, reflexos da luz, vasos sanguíneos e outros elementos da imagem. Em geral, estas soluções apresentam uma primeira fase de pré-processamento para reduzir os efeitos que dificultam o processo de análise das bordas. Em uma segunda fase, são definidas as áreas iniciais de interesse, que são as bordas dos elementos presentes na imagem. O uso do detector de bordas *Canny* [21] é bastante difundido, porém outra técnicas também podem ser empregadas. Por exemplo, a que considera os pólipos como uma superfície saliente e que por causa da luz do endoscópio produz sombra ao redor das áreas de borda. Isto possibilita a análise da imagem em termos de diferença de intensidade de regiões (*valley information, intensity valley detection*) [86]. Por fim, cada estudo aplica um ou mais métodos particulares de caracterização de pólipos.

Tabela 3.2: Lista de estudos agrupados por abordagens que usam aprendizado de máquina e que não usam.

MÉTODO/TÉCNICA	**ESTUDOS**
COM APRENDIZADO DE MÁQUINA	Total: 27 Estudos
SVM (Support Vector Machine)	[1–3, 23, 64, 68, 83, 144, 145]
RF (Random Forest)	[134–136]
DT (Decision Tree)	[144, 145]
CRF (Conditional Random Field)	[99]
LDA (Linear Discriminant Analysis)	[74]
PLS (Partial Least Squares)	[6]
K-NN (K Nearest Neighbors)	[115]
MLP (Multi-layer Perceptron)	[92]
MFNN (Multi-layer Feed-forward Neural Network)	[73, 75]
CNN (Convolutional Neural Network)	[72, 90, 137, 148, 150]
GAN (Generative Adversarial Network)	[105]
SEM APRENDIZADO DE MÁQUINA	Total: 6 Estudos
Curvature Analysis	[63, 71, 79]
Appearance of polyps boundaries	[10–12]

Estudos que abordam *Curvature Analysis* como [63, 71, 79] são baseados

na análise de bordas dos elementos presentes nas imagens. Por exemplo, o uso da saída de um detector de bordas *Canny* para diferenciar as bordas nas paredes do intestino de bordas pertencentes a pólipos. Já no trabalho de Hwang et al. [63], os autores utilizam a técnica de *Ellipse Fitting* comparando em cada *frame* as bordas destacadas com a direção da curva e curvatura, por exemplo. Este ainda considera informações em quadros adjacentes por meio do método *Mutual Information* (MI) [22]. No estudo [71] os autores avaliam se os segmentos na imagem pertencem a pólipos, por meio da extração de características (área, forma e cor), para comparação com as particularidades já conhecidas dos formatos das lesões.

Outros trabalhos empregam distintas técnicas para análise das áreas de bordas dos elementos nas imagens, como *ridge and valley detection* [86]. Em Bernal et al. [10], os autores observaram que há nas imagens de colonoscopia uma diferença nos valores de intensidade que destaca as bordas dos pólipos (*valley information*). Esta diferença de intensidade é usada para caracterizar as bordas das lesões de acordo com um *energy map*, onde a alta concentração de valores indica a presença de pólipos. No estudo de Bernal et al. [12], são aplicadas diversas etapas de processamento baseadas em regiões das imagens. Inicialmente a imagem é particionada em regiões aplicando a transformação *watersheds* [16] sobre o gradiente da imagem. As regiões são avaliadas pelo método SA-DOVA (*depth of valleys image* [86]). Este define pontos que provavelmente pertencem à área de um pólipo na imagem baseado em um processo de acumulação de valores sobre estes pontos. Por fim, para cada região são apresentados pontos em que há valores mais altos ao redor, indicando que estes pontos pertencem a um pólipo na imagem. O trabalho de Bernal et al. [11] utiliza o método chamado VO-DOVA, que também está baseado em *ridge and valley detection* [86]. Neste, uma série de setores radiais são utilizados para determinar a orientação dos *valleys* na imagem, produzidos pelo processo de *ridge and valley detection*. Estes devem ter a mesma orientação dos setores radiais ao redor do ponto, indicando que este ponto está contido dentro da região de um pólipo, no caso de as orientações coincidirem.

3.3 Análise Comparativa dos Estudos

Os resultados apresentados pelos estudos selecionados são baseados em testes sobre bases de dados de imagens de colonoscopia. No entanto, os autores fizeram uso de bases de dados e métricas diferentes para avaliação dos resultados. Uma comparação numérica direta dos resultados que utilizam bases de dados tão diversas, não parece ser adequada para encontrar o método mais

eficiente para detecção de pólipos. É importante destacar que a quantidade de imagens no conjunto de dados pode influenciar diretamente, e.g., quando uma metodologia baseada em aprendizado de máquina é utilizada.

Os estudos listados geralmente fazem uso de mais de uma base de dados com intuito de estender a quantidade de amostras. A Tabela 3.3 apresenta a relação dos conjuntos de imagens utilizados por cada trabalho. Somente três estudos fizeram uso exclusivamente da base de dados ASU-Mayo [115, 148, 150]. A mais utilizada foi a base de dados CVC-ColonDB com 8 estudos. Porém, somente 2 utilizaram unicamente esta [12, 135]. O estudo de Bernal et al. [10] fez uso das bases de dados CVC-ColonDB e CVC-ClinicDB, além do estudo de Kang e Gwak [72]. Pogorelov et al. [105] efetuaram um arranjo com experimentos em seis bases de dados de imagens, incluindo CVC-ClinicDB e Kvasir. Outros 4 trabalhos restantes combinaram a base CVC-ColonDB com bases de dados próprias [6, 134, 136, 137]. A maior parte dos estudos, 18 deles, fizeram seus testes exclusivamente em bases de dados próprias [1, 2, 11, 23, 63, 64, 68, 71, 73–75, 79, 83, 90, 92, 99, 144, 145]. O estudo de Amber et al. [3] não apresentou informações sobre as imagens utilizadas nos testes.

Neste cenário, uma comparação mais adequada dos resultados pode ser realizada entre os estudos [135] e [12] que utilizam a mesma base de dados e as mesmas métricas precisão[5] e revocação[6]. Neste caso, a abordagem de aprendizado de máquina utilizando a técnica *Random Forest* no estudo de Tajbakhsh et al. [135] consegue obter melhores resultados. Outra comparação pode ser feita entre os estudos [10] e [105] avaliando a métrica de precisão nos testes efetuados sobre a base de dados CVC-ClinicDB, com melhor resultado obtido por [105].

Pode-se observar que os estudos utilizam métricas em comum para verificar os resultados dos métodos propostos. Alguns dos autores selecionaram pelo menos a dupla de métricas precisão e revocação, que aparecem em 11 estudos [6, 11, 12, 72, 90, 105, 115, 134–136, 150]. Destes apenas [12] e [11] não utilizaram métodos baseados em aprendizado de máquina.

Outra métrica bastante utilizada é acurácia[7]. Esta é aplicada em 9 estudos [1, 3, 11, 64, 68, 83, 90, 92, 148]. Os estudos [3, 64, 83, 92] empregaram unicamente acurácia para aferir os resultados. O trio de métricas acurácia, sensibilidade e especificidade[8] foi utilizado nos trabalhos [1, 68, 90].

[5]Do inglês *precision*.
[6]Do inglês *recall*. O mesmo que sensibilidade (*sensitivity*) e taxa de verdadeiros positivos.
[7]Do inglês *accuracy*.
[8]Do inglês *specificity*.

Tabela 3.3: Estudos e respectivas métricas e bases de dados.

Estudo	**Col.**	**Cli.**	**May.**	**Prop.**	**Métricas**
Riegler et al. [115]			X		Precisão 0.903; Revocação 0.919; F1 Score 0.910
Yuan et al. [148]			X		Acurácia 91.47%; Sensibilidade 91.76%
Zhang et al. [150]			X		Especificidade 97.0%; Precisão 88.6%; Revocação 71.6%; F1 Score 79.2%
Tajbakhsh et al. [135]	X				Precisão 86%; Revocação 86%
Bernal et al. [12]	X				Precisão 54%; Revocação 72%
Bernal et al. [10]	X	X			Precisão 72.33% (CVC-ColonDB); Precisão 70.26% (CVC-ClinicDB)
Kang e Gwak [72]	X	X			Precisão 73.84%; Revocação 74.37%;
Pogorelov et al. [105]		X			Precisão 81.9%; Revocação 61.9%; Especificidade 98.4%; Acurácia 94.6%; F1 Score 70.6%;
Tajbakhsh et al. [136]	X			X	Precisão 94%; Revocação 80%
Bae e Yoon [6]	X			X	Precisão 70.67%; Revocação 70.67%; PR-AUC 65.43%
Tajbakhsh et al. [134]	X			X	Precisão 93%; Revocação 80%
Tajbakhsh et al. [137]	X			X	0.002 falsos positivos por *frame* a 50% sensibilidade
Maroulis et al. [92]				X	Acurácia 95%
Bernal et al. [11]				X	Precisão 89.52%; Revocação 88.33%; Acurácia 89%; Especificidade 89.66%
Iwahori et al. [68]				X	Acurácia 96.8%; Sensibilidade 95.4%; Especificidade 98.2%
Agrahari et al. [1]				X	Acurácia 89.65%; Sensibilidade 90%; Especificidade 89.18%
Ma et al. [90]				X	Acurácia 96.04%; Sensibilidade 93.67%; Especificidade 98.36%
Li et al. [83]				X	Acurácia 83.4%
Iakovidis et al. [64]				X	Acurácia 94%
Wang et al. [144]				X	Revocação 86.3%
Hwang et al. [63]				X	Quadros de pólipos corretamente detectados 26 de 27
Alexandre et al. [2]				X	ROC(AUC) 94.87
Cheng et al. [23]				X	Sensibilidade 86.2; Taxa de falso positivo por imagem 1.26
Kang e Doraiswami [71]				X	N/A
Krishnan et al. [79]				X	N/A
Karkanis et al. [75]				X	Média de reconhecimento 93.3%
Karkanis et al. [73]				X	94 das lesões%
Karkanis et al. [74]				X	Sensibilidade 90%; Especificidade 97%
Wang et al. [145]				X	Revocação 97.7%
Park et al. [99]				X	ROC(AUC) 0.89

Col.: CVC-ColonDB; Cli.: CVC-ClinicDB; May.: ASU-Mayo; Prop.: Base de dados própria;

A revocação (ou sensibilidade) é uma métrica importante no contexto da detecção de pólipos pois é direcionada para redução de falsos negativos, i.e., quando o pólipo não é identificado pelo sistema. Esta métrica é utilizada em 16 estudos [1, 6, 11, 12, 23, 68, 72, 74, 90, 105, 115, 134–137, 144, 145, 148, 150] indicando uma preocupação dos autores em evitar que uma lesão real não seja detectada, já que o sucesso do tratamento do câncer de cólon está relacionado com a detecção precoce.

O estudo de Bernal et al. [10] é baseado em um modelo de aparência dos pólipos e utilizou a métrica precisão sobre as bases de dados CVC-ColonDB e CVC-ClinicDB. Os trabalhos [2] e [99] fazem uso da técnica de visualização *Receiver Operation Characteristic (ROC) curve* [55] que apresenta o desempenho geral relacionando os valores das métricas revocação e precisão. Estes valores são quantificados pela *Area Under the Curve* (AUC) onde valores mais próximos a 1 (ou 100 no caso de [2]) indicam melhores resultados.

O estudo de Bae e Yoon [6] apresenta um gráfico comparativo de precisão e revocação. É uma forma de avaliar a área sob a curva do gráfico[9] (PR-AUC) considerando os Verdadeiros Positivos (TP), onde os valores mais elevados indicam menor quantidade de Falso Negativo (FN) e menor quantidade de Falso Positivo (FP), mas não considera os Verdadeiros Negativos (TN). Este trabalho utiliza técnicas baseadas em aprendizagem de máquina e considera um conjunto de dados de entrada desbalanceado.

Outras métricas são medida-F[10] [141] que combina revocação e precisão utilizando média harmônica, como apresentados no trabalho [115] e os valores da taxa de falso positivo vistos nos estudos [137] e [23] que indicam quando o sistema reconhece um pólipo onde não existe. Em alguns casos a detecção pode confundir alguma parte da parede do cólon com um pólipo, por exemplo. Os estudos [71] e [79] não especificaram nenhuma métrica.

Para que um sistema de detecção seja clinicamente útil e utilizado em exames de rotina, o método de detecção implementado precisa ser capaz de decidir se existe um pólipo na imagem e qual a sua localização a cada quadro do vídeo em tempo real. Segundo Bernal et al. [14], este requisito limita o tempo de análise dos quadros em torno de 40 milissegundos (considerando um vídeo com 25 FPS). Abordagens que apresentam bons resultados de detecção podem não apresentar um bom desempenho relacionado ao tempo de processamento. A resolução das imagens utilizadas em cada estudo também deve ser considerada, o que dificulta uma comparação direta do tempo de processamento apresentado nestes trabalhos. Poucos estudos apresentaram informações relativas ao tempo

[9] *Precision-Recall (PR) curve. Area under the curve (AUC).*
[10] *F-Measure*. Mesmo que *F1 Score*.

de análise das imagens. Os trabalhos que divulgam estes números estão listados na Tabela 3.4.

O estudo de Maroulis et al. [92] é o mais rápido, porém não estão incluídas neste tempo as operações de leitura e escrita e o tempo informado é relativo apenas a etapa de avaliação. Os trabalhos de Zhang et al. [150] e Bae e Yoon [6] são capazes de analisar um quadro a cada 0.153 e 0.6375 segundos, respectivamente. O estudo de Kang e Doraiswami [71] consome o tempo de 1 segundo por quadro, porém os autores informam que metade deste tempo é devido ao algoritmo de detecção de bordas *Canny*. O tempo de processamento por quadro é de 4.8 segundos no estudo [74], no entanto, os autores afirmam que os testes foram feitos em um computador PC de baixo custo, com configuração padrão. Contudo, o algoritmo também pode ser executado paralelamente utilizando *hardware* especializado e assim reduzir o tempo de processamento. O estudo [144] precisa de 7.1 segundos para examinar uma imagem. Cerca de 4 segundos deste tempo é devido a verificação das bordas dos elementos da imagem utilizando a técnica apresentada e a fase de extração de regiões de interesse. O trabalho de Cheng et al. [23] é capaz de processar uma imagem a cada 13 segundos. Por fim, abordagem do estudo [12] efetua uma análise de 19 segundos para cada imagem. Segundo os autores, a paralelização e uma implementação mais eficiente da operação de dilatação pode reduzir este tempo.

Tabela 3.4: Tempo de processamento de imagens apresentado pelos estudos.

Estudo	**Tempo de Proc.**	**Em cada**
Maroulis et al. [92]	0.018s e 0.8s	imagem
Zhang et al. [150]	0.153s	imagem
Bae e Yoon [6]	0.6375s	imagem
Kang e Doraiswami [71]	1.0s	imagem
Karkanis et al. [74]	4.8s	imagem
Wang et al. [144]	7.1s	imagem
Cheng et al. [23]	13s	imagem
Bernal et al. [12]	19s	imagem

Alguns estudos apresentam a quantidade de imagens usadas para obtenção dos resultados. Estes dados estão sintetizados na Tabela 3.5. A quantidade de imagens utilizadas por cada estudo é importante, principalmente nas abordagens de aprendizado de máquina, onde é empregada a etapa de treinamento. A resolução da imagem também pode influenciar no tempo de análise de cada

imagem. Dentre os estudos listados na Tabela 3.5 destacam-se os citados a seguir: Pogorelov et al. [105] utilizam as 21.592 imagens para treinamento e teste em experimentos de localização e detecção. As imagens foram obtidas a partir de seis conjuntos de imagens diferentes. O estudo Iakovidis et al. [64] divide o total de 21.500 quadros[11] em 8.600 quadros para treinamento e 12.900 para testes. Entre os 8.621 quadros no trabalho [63], 826 são quadros em que estão presentes pólipos e 7.806 são registros do cólon sem lesões.

Em Alexandre et al. [2] os autores removeram as áreas em preto na borda de 36 imagens (pré-processamento), resultando em uma imagem de 514x469 pixels. A partir destas 36 imagens os autores geram 4.620 imagens de 40x40 pixels. Os autores do trabalho [136] utilizaram a base de dados CVC-ColonDB composta de 300 imagens e adicionalmente fazem uso de uma base de dados própria consistindo de 1.700 imagens com pólipos e 2.500 sem pólipos. Em Tajbakhsh et al. [134] os autores utilizaram o conjunto CVC-ColonDB como imagens positivas (contém pólipos) e 3000 imagens da base de dados própria como imagens negativas. Os estudos de Bernal et al. [10] e Kang e Gwak [72] utilizam as duas bases de dados CVC-ColonDB e CVC-ClinicDB totalizando 900 imagens. Os estudos [3, 71, 73–75, 92, 99, 115, 137, 150] não disponibilizaram estas informações.

Os trabalhos mais eficientes considerando revocação (ou sensibilidade) foram [145] (97.7%), [68] (95.4%), seguidos de [115] (91.9%). Em termos de acurácia os melhores resultados foram dos estudos [68] (96.8%), [92] (95%) e [105] (94.6%).

Os resultados apresentados por três estudos podem ser destacados [64, 68, 92]. O trabalho de Maroulis et al. [92] (MLP) apresenta destaque no tempo de processamento e nos resultados de acurácia. Porém, os autores não apresentaram outras métricas e não forneceram detalhes sobre as imagens utilizadas nos testes. O estudo de Iwahori et al. [68] (SVM, *Hessian Filter*) se sobressai com valores elevados nas métricas acurácia (96.8%), sensibilidade (95.4%) e especificidade (98.2%), entretanto utiliza um conjunto de imagens bem limitado com apenas 128 imagens e os autores não informaram o tempo de processamento gasto pelo método proposto. Já o estudo de Iakovidis et al. [64] (SVM) conseguiu alcançar uma alta taxa de acurácia (94%) sobre um conjunto composto de 21.500 imagens. Contudo, os autores não forneceram os valores de outras métricas além de acurácia e não indicaram o tempo de processamento. O estudo de Ma et al. [90] baseado em um modelo SSD modificado alcançou acurácia de 96.04%, sensibilidade de 93.67% e especificidade de 98.36%, no entanto precisou aplicar várias técnicas de aumento de dados e

[11]Imagens extraídas dos vídeos (*frames*).

Tabela 3.5: Quantidade de imagens utilizadas por estudo.

Estudo	**Quant.**	**Resolução**
Pogorelov et al. [105]	21.592	Variadas
Iakovidis et al. [64]	21.500	320 x 240
Hwang et al. [63]	8.621	N/A
Alexandre et al. [2]	4.620	40 x 40
Tajbakhsh et al. [136]	4.500	N/A
Tajbakhsh et al. [134]	3.300	N/A
Yuan et al. [148]	2.522	N/A
Ma et al. [90]	1.936	500 x 500
Bae e Yoon [6]	1.642	500 x 574
Wang et al. [144]	1.513	720 x 480
Bernal et al. [10]	912	500 x 574; 384 x 288
Kang e Gwak. [72]	912	500 x 574; 384 x 288
Bernal et al. [12]	300	500 x 574
Tajbakhsh et al. [135]	300	500 x 574
Bernal et al. [11]	300	N/A
Iwahori et al. [68]	128	1000 x 900
Agrahari et al. [1]	87	1000 x 1000
Wang et al. [145]	61	720 x 480
Li et al. [83]	58	256 x 256
Krishnan et al. [79]	6	N/A

a validação foi realizada sobre um conjunto de imagens privadas, dificultando uma comparação direta com outros estudos.

3.4 Aprimoramento dos Dados de Treinamento

Os trabalhos mais recentes de detecção de pólipos que utilizam aprendizado de máquina, em geral, apresentam resultados melhores. Porém, esta metodologia é altamente dependente da quantidade e variedade do conjunto de dados [17, 29, 30, 114]. Neste contexto, os estudos apresentam estratégias para melhorar o conjunto de dados de treinamento. Por exemplo, a utilização de técnicas de aumento de dados (*data augmentations*). Também o uso de várias bases de dados combinadas para aprimorar o treinamento e ainda carregamento de modelos pré-treinados com imagens que não pertencem a área médica como visto em Pogorelov et al. [105]. Há estudos trabalham com bases de dados de imagens próprias, embora estas apresentem alto custo de produção, devido a mão de obra médica especializada para indicar a localização dos

pólipos nas imagens. Por exemplo, em Wittenberg et al. [146] as lesões em 2.484 imagens extraídas de exames de colonoscopia precisaram ser delineadas manualmente e em Ma et al. [90] com 1.936 imagens nas quais dois endoscopistas experientes efetuaram a marcação das áreas das lesões. O estudo de Qadir et al. [110] descreve um método semiautomático para criar as anotações de localização dos pólipos. Isto é, uma tentativa para reduzir o esforço do médico endoscopista na tarefa de marcar as regiões nos pólipos nas imagens.

O uso de vários conjuntos de imagens combinados pode ser visto no trabalho de Urban et al. [140], no qual os autores reuniram 8.641 imagens de colonoscopia para formar um conjunto de dados de treinamento. Importante destacar que os autores fizeram uso de outras modalidades de exame para compor um conjunto maior de imagens, i.e., *Narrow-Band Imaging* (NBI). Este conjunto de dados possui 840 imagens da modalidade NBI e mais 7.801 imagens de colonoscopia convencional. As imagens foram pré-processadas sendo redimensionadas para 224 x 224 *pixels* e também foram aplicadas *data augmentations*. Os autores realizaram testes utilizando três arquiteturas de redes neurais convolucionais profundas VGG16, VGG19 e ResNet50. A VGG19 apresentou os melhores resultados com 96.4% de acurácia. Estes valores se referem ao teste realizado sobre um conjunto de imagens de teste independente com 1330 imagens. Nesta rede foi efetuado um pré-treinamento com o conjunto de imagens ImageNet [31].

Semelhantemente, no estudo de Shin et al. [125] foram utilizadas estratégias de *data augmentations* como rotações, aumento de brilho, desfoque, cisalhamento e aproximação com intuito de aumentar a quantidade de amostras de imagens para treinamento. Os autores também aplicaram a técnica de *transfer learning* de um modelo *Inception Resnet* previamente treinado com o conjunto de imagens COCO (*Common Objects in Context*) [85]. Neste trabalho os autores empregaram o conjunto de imagens de colonoscopia CVC-ClinicDB para treinamento e ETIS-LaribPolypDB para o teste.

O objetivo de formar um conjunto de dados mais variado, apresentado por estes estudos, evidencia a necessidade melhorar a quantidade e a variabilidade das imagens da colonoscopia para abordagens de aprendizado de máquina. Qadir et al. [110] afirmam que a carência de imagens de treinamento e suas respectivas imagens *ground truth* são um dos principais obstáculos para obtenção de melhores resultados em detecção e segmentação de pólipos.

Apesar dos avanços nos sistemas de detecção de pólipos que utilizam técnicas de aprendizado profundo, pouco foi feito em relação aos conjuntos de dados, no sentido de contribuir na variação das amostras. Neste sentido, o estudo de Han et al. [54] utilizou uma estratégia para gerar imagens de

ressonância magnética (RM) diferentes das originais, com foco no aumento de dados e no treinamento médico. No contexto de imagens de colonoscopia, o trabalho de Shin et al. [126] apresenta um método para adicionar mais imagens de colonoscopia ao conjunto de dados, obtendo resultados promissores.

Neste contexto de melhoria de dados de treinamento, estudos aplicados a imagens de tomografia computadorizada do tórax e mamografia apresentam técnicas para criação de amostras artificiais a partir da inserção de lesões [43, 88, 117]. O trabalho de Robins et al. [116] apresenta comparativos entre conjuntos de dados híbridos e dados clínicos reais, onde dados híbridos são criados por meio da introdução lesões artificiais em imagens de tomografias reais. O estudo de Pezeshk et al. [102] descreve um método que utiliza a inserção de lesões como forma de aumento de dados. Segundo os autores, esta composição de imagens via inserção de lesões (nódulos pulmonares) é capaz de melhorar os resultados de classificadores treinados com conjuntos de dados limitados em quantidade.

No melhor do nosso conhecimento, existem poucos estudos na literatura com objetivo de aprimorar a variedade e quantidade de amostras nos conjuntos de imagens de colonoscopia disponibilizados publicamente como por exemplo o CVC-ClinicDB.

4
Aprimoramento de Dados para Segmentação com Pólipos Reais

Este capítulo descreve a metodologia utilizada para o processo de aprimoramento de imagens pertencentes a base de dados CVC-ClinicDB. Inicialmente é introduzida a visão geral do procedimento listando cada etapa (Seção 4.1). Este procedimento forma um novo conjunto de imagens melhorado composto pelas imagens originais CVC-ClinicDB modificadas por meio do processo de inserção de pólipos (Seção 4.2). Tais alterações aumentam a quantidade de amostras e a variabilidade dos dados para uma solução de segmentação de pólipos baseada em aprendizado de máquina. Após, é descrita a organização das imagens aprimoradas em conjuntos de treinamento empregados nos experimentos (Seção 4.3). Em seguida, são detalhados os resultados obtidos nos respectivos testes (Seção 4.4). A estratégia para localização dos pólipos especificada neste capítulo está baseada na arquitetura de segmentação de imagens U-net.

4.1
Visão Geral

O procedimento proposto para criação de um conjunto de dados aprimorado pode ser descrito simplificadamente como uma etapa de extração de regiões com pólipos nas imagens do conjunto de imagens original para uma posterior inserção destas regiões em outras imagens colonoscopia. A execução de forma automática destas tarefas resultará em uma nova base de dados de imagens que será empregada como conjunto de treinamento para uma rede neural.

A ilustração deste procedimento pode ser vista na Figura 4.1(a), onde o pólipo destacado pelo círculo em verde é copiado de uma imagem (origem) pertencente ao conjunto de dados original e é aplicado a outra imagem (destino), formando uma imagem diferente com dois pólipos neste caso. Outro exemplo está representado na Figura 4.1(b) onde outro pólipo é extraído de uma imagem origem diferente e inserido em outra imagem destino a partir do mesmo processo. As imagens destino e origem pertencem a mesma base de dados CVC-ClinicDB, porém as imagens apresentam diferenças de

textura e luminosidade, por exemplo. A localização na qual o novo pólipo será posicionado varia de acordo com a aparência da imagem destino, visto que esta será determinada de acordo com o resultado do processo de segmentação de regiões por *Watershed* (Seção 4.2.2).

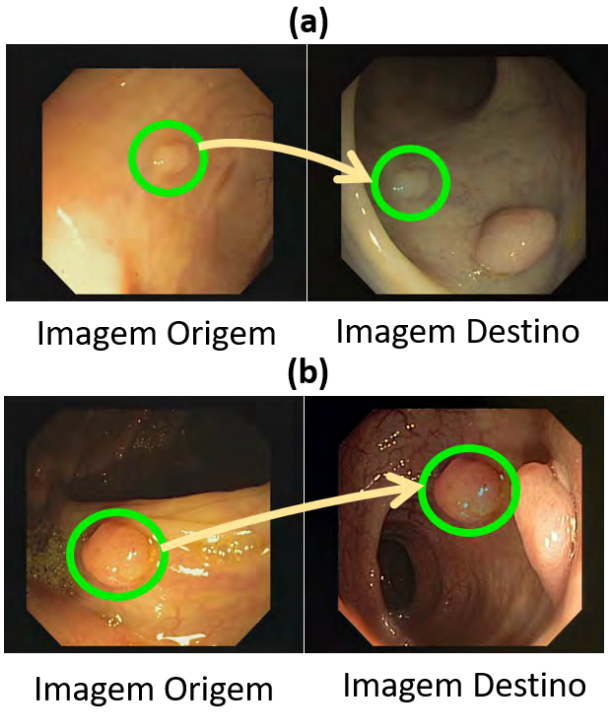

Figura 4.1: Exemplo de pólipo extraído de uma imagem (origem) e inserido em uma outra imagem (destino). Linha (a): exemplo de inserção com pólipo pequeno (em quantidade de *pixels*). Linha (b): exemplo com pólipo médio.

As etapas do procedimento que será descrito neste trabalho podem ser visualizadas na Figura 4.2. A primeira etapa consiste em verificar em todo o conjunto de dados original quais pólipos estão de acordo com um critério (Figura 4.2(a)), neste caso, o critério de tamanho será aplicado. Serão considerados os k-ésimos menores pólipos dentre todas as imagens, onde o tamanho é contabilizado por quantidade de *pixels*. Assim, esta etapa define uma lista de pólipos que atendem o critério. Este critério poderia ser baseado nos tipos de lesões de acordo com a *Paris Classification* [5], por exemplo. No entanto, não há no conjunto de dados CVC-ClinicDB informações correspondentes aos tipos de pólipos apresentados nas imagens.

Em seguida, uma caixa delimitadora (*bounding box*) ao redor da região do pólipo é obtida (Figura 4.2(b)). A partir deste ponto, as coordenadas da caixa delimitadora são utilizadas para se obter uma cópia da região exata do pólipo presente a imagem original (Figura 4.2(c)). Esta região é uma subimagem que será duplicada e adicionada na imagem destino posteriormente.

A próxima etapa determina em qual região na imagem destino o pólipo será adicionado. Um conjunto de regiões de interesse é apresentado de acordo com a saída da operação de *Watershed* [94] sobre a imagem destino (Figura 4.2(d)). A região selecionada para receber o pólipo deve ser a maior em quantidade de *pixels* que não esteja posicionada sobre um outro pólipo já existente. Se alguma região atende este critério, então as restantes são excluídas nesta etapa (Figura 4.2(e)).

Neste ponto, a região que receberá o pólipo já está definida e a cópia do pólipo obtida anteriormente é aplicada dentro desta região (Figura 4.2(f)). A imagem de saída é então obtida por uma composição de uma imagem do conjunto original com o novo pólipo adicionado (Figura 4.2(g)). Os detalhes destas etapas serão apresentados na seção seguinte.

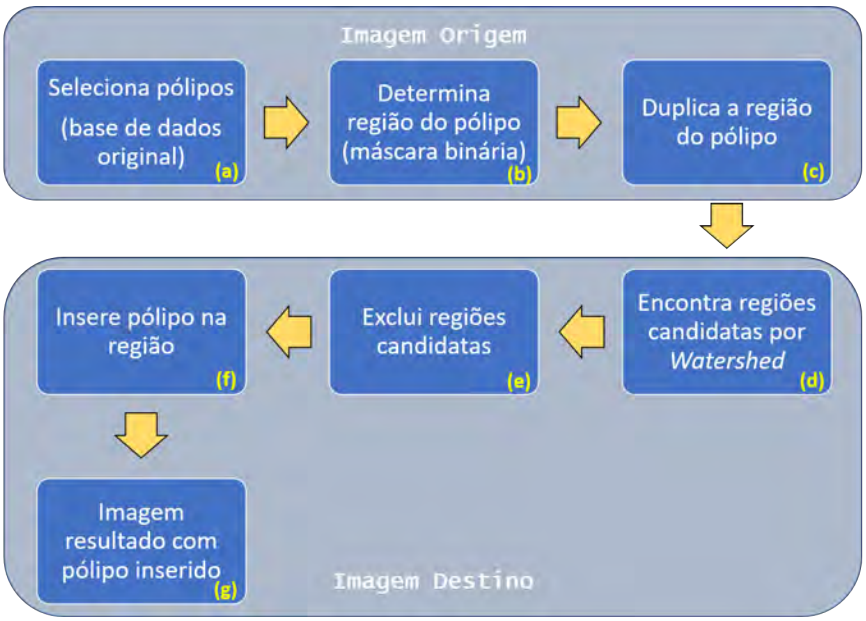

Figura 4.2: Etapas do procedimento para criação de imagens modificadas pela inserção de pólipos.

4.2 Processo de Inserção de Pólipos

Esta seção descreve as etapas do processo de inserção dos pólipos em regiões específicas da imagem destino. Na primeira parte é apresentada a estratégia para seleção dos pólipos considerando as máscaras binárias fornecidas pelo conjunto de imagens CVC-ClinicDB. Na sequência é detalhado o procedimento para a seleção das regiões na imagem destino e a suavização do pólipo inserido para manter a coerência visual.

4.2.1 Seleção do pólipo

Nas imagens de colonoscopia é comum a presença de diversos tipos e tamanhos de pólipos. Porém, a percepção do tamanho das lesões nas imagens pode ser afetada pelo ponto de vista da câmera do colonoscópio. Por exemplo, caso a câmera esteja mais próxima da mucosa o pólipo pode parecer maior do que realmente é. No contexto deste estudo o tamanho do pólipo é definido de acordo com tamanho da área que ele ocupa em *pixels*. Nas imagens CVC-ClinicDB um mesmo pólipo aparece em mais de uma imagem. Assim, o método proposto irá considerar o tamanho deste pólipo em cada imagem separadamente, mesmo que se trate da mesma lesão. A imagem do conjunto de dados que define a posição e a área ocupada pelo pólipo é chamada de imagem *ground truth*. É uma imagem binária, também chamada de máscara, cuja área do pólipo foi delimitada por médicos especialistas. A quantidade de *pixels* em branco na imagem *ground truth* irá determinar o tamanho do pólipo. Isto pode ser visto na Figura 4.3, onde a aparência em relação ao tamanho é diferente mesmo sendo a mesma lesão. O exemplo da Figura 4.3 (a) corresponde a uma determinada imagem e a sua respectiva máscara binária, que são distintas do par de imagens na Figura 4.3 (b), apesar de apresentarem o mesmo pólipo. Neste caso, o tamanho será diferente pois serão considerados os *pixels* em branco das imagens *ground truth* correspondentes na linha inferior.

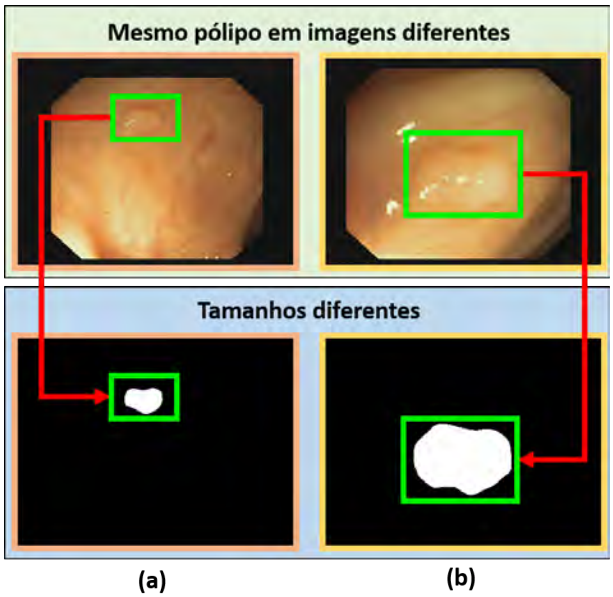

Figura 4.3: Duas imagens do mesmo pólipo. Colunas (a) e (b): Par de imagens contendo a imagem de colonoscopia (acima) e a respectiva imagem *ground truth* (abaixo).

Para cada imagem *ground truth* do conjunto de dados é realizada a

verificação de tamanho, onde são selecionados k-ésimos menores pólipos. Como o objetivo é inserir estes pólipos em regiões da imagem destino é mais adequado a utilização de pólipos menores. A ideia principal é inserir o pólipo cujo tamanho é menor que o da região *Watershed* selecionada (Seção 4.2.2). Como a região tende a ser uma área mais homogênea da imagem, a inserção de um pólipo compatível em tamanho com esta região resulta em uma imagem resultado (última etapa do workflow na Figura 4.2) mais uniforme em relação ao posicionamento do novo pólipo.

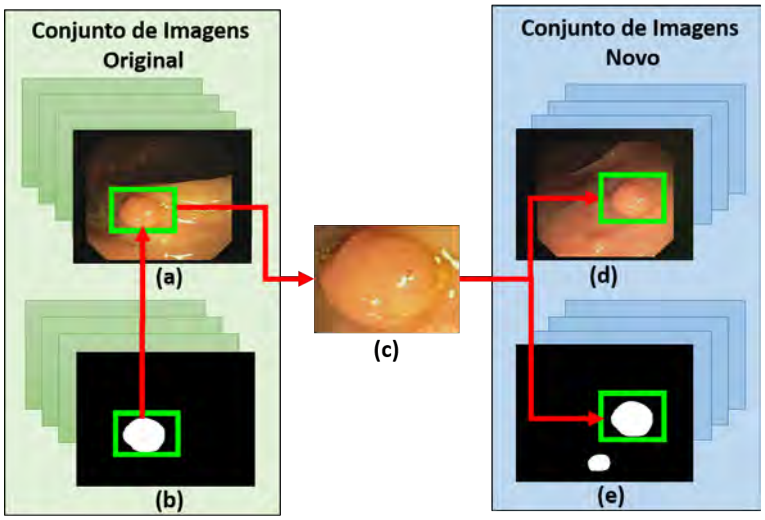

Figura 4.4: Ilustração do processo de seleção e inserção do pólipo. (a) imagem origem com caixa delimitadora em torno do pólipo. (b) caixa delimitadora de seleção do pólipo na imagem *ground truth*. (c) área do pólipo duplicada. (d) pólipo aplicado sobre a região *Watershed* selecionada. (e) imagem *ground truth* criada para conter o pólipo adicionado.

Após estabelecidos os menores pólipos, a menor caixa delimitadora capaz de conter a área do pólipo é definida por meio da imagem *ground truth*, conforme visto na Figura 4.4 (b). Com as informações de tamanho e posição da caixa delimitadora definidas é possível extrair o pólipo da imagem de colonoscopia correspondente. A subimagem apresentada na Figura 4.4 (c) é uma duplicação da área do pólipo na imagem origem (Figura 4.4 (a)) que pertence ao conjunto de dados original. Com a área duplicada, o pólipo pode ser inserido na imagem destino (Figura 4.4 (d)) e na imagem *ground truth* correspondente (Figura 4.4 (e)). A estratégia para a escolha da localização mais adequada para inserção do pólipo é apresentada na Seção 4.2.2.

4.2.2
Escolha da região para inserção do pólipo

Com o pólipo pronto para ser adicionado em outra imagem é preciso decidir em qual região desta imagem ele será posicionado. A escolha da melhor região é importante pois ajuda a manter a coerência na imagem destino no sentido de não prejudicar a aparência, mantendo-a visualmente mais próxima possível de uma imagem de colonoscopia real em que não ocorreu qualquer modificação.

A escolha desta região está baseada na estratégia de *Watershed* (Seção 2.1.4) para segmentação da imagem destino. As etapas deste processo estão representadas na Figura 4.5, onde a imagem destino em (a) é utilizada como base para encontrar um conjunto de regiões ilustradas em (j). O pólipo será adicionado em uma destas regiões, caso alguma seja compatível com pólipo a ser inserido. No caso de incompatibilidade, esta imagem destino é descartada e uma nova é selecionada dentro do conjunto de dados original.

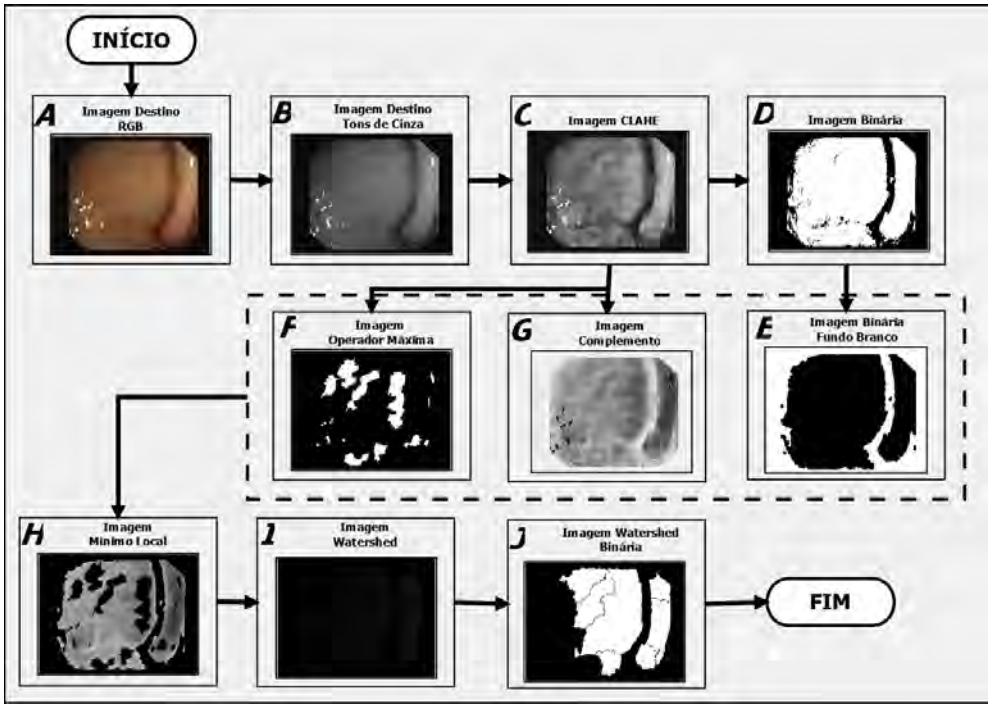

Figura 4.5: Etapas do procedimento de geração de regiões *Watershed* na imagem destino.

As etapas apresentadas têm o objetivo de produzir uma imagem mais adequada para a aplicação do método de *Watershed*, i.e. etapas de pré-processamento[1].

[1]Exemplo das técnicas utilizadas pode ser visto em: `https://blogs.mathworks.com/steve/2006/06/02/cell-segmentation/`.

O processo para geração das regiões é iniciado pela criação de uma versão em escala de cinza da imagem destino e na sequência é aplicado o processo de equalização (Seção 2.1.1) para aumentar contraste e limitar ruídos em áreas homogêneas [113], resultando na imagem equalizada CLAHE vista na Figura 4.5 (c).

A partir da imagem CLAHE são produzidas três versões de imagens resultantes de transformações de binarização (limiar definido pelo método Otsu [98]), operador máxima e complemento (Figura 4.5 (d), (f) e (g) respectivamente). A imagem binária (Figura 4.5 (d)) é utilizada como base para a criação de uma versão modificada, apresentada na Figura 4.5 (e) como imagem binária fundo branco. Esta versão modificada é uma imagem onde áreas com fundo em preto correspondentes na imagem (d), que estão cercadas por áreas em branco são preenchidas para formar uma superfície mais homogênea. Na imagem (e) o fundo foi invertido da cor preto para branco e o primeiro plano de branco para preto.

A imagem (f) também foi obtida a partir de (c), onde se destacam em branco os grupos de *pixels* que foram identificados com valores de intensidade altos e constantes em relação aos outros *pixels* ao redor [129] (Seção 2.1.3). A imagem (g) é o complemento da imagem (c) (áreas escuras tornam-se mais claras e áreas claras tornam-se mais escuras). Isto será útil para aplicar a transformação de *Watershed* em processamento posterior.

A imagem (h) é uma combinação composta pelas imagens (g), (f) e (e). Esta imagem (h) é gerada a partir de uma operação lógica *OR* sobre os valores dos *pixels* das imagens (e) e (f). Sendo que os valores de intensidade da imagem (g) são forçados para serem mínimos locais sempre que os valores correspondes resultantes da operação (e) *OR* (f) são diferentes de zero.

Deste modo é obtida a imagem (h), sendo a entrada para o algoritmo *Watershed* que é capaz de reconhecer regiões de acordo com o nível de intensidade. A imagem (h) fornece os valores de mínimos (áreas escuras) que serão utilizados como sementes para o algoritmo de *Watershed*. O processo *Watershed* retorna regiões encontradas na imagem (i), que é convertida para uma versão binária onde as regiões são áreas claras na imagem (j). Por fim, o pólipo será posicionado eu uma destas regiões da imagem (j).

Com as regiões retornadas pelo processo de *Watershed* já definidas na imagem destino, é preciso escolher em qual região o pólipo será adicionado. Para isso é realizada uma verificação do tamanho da região em termos de quantidade de *pixels*. A maior região encontrada será considerada como candidata para receber o pólipo. Porém, esta região somente receberá o pólipo, caso sua área não esteja sobre um pólipo que já está presente na imagem

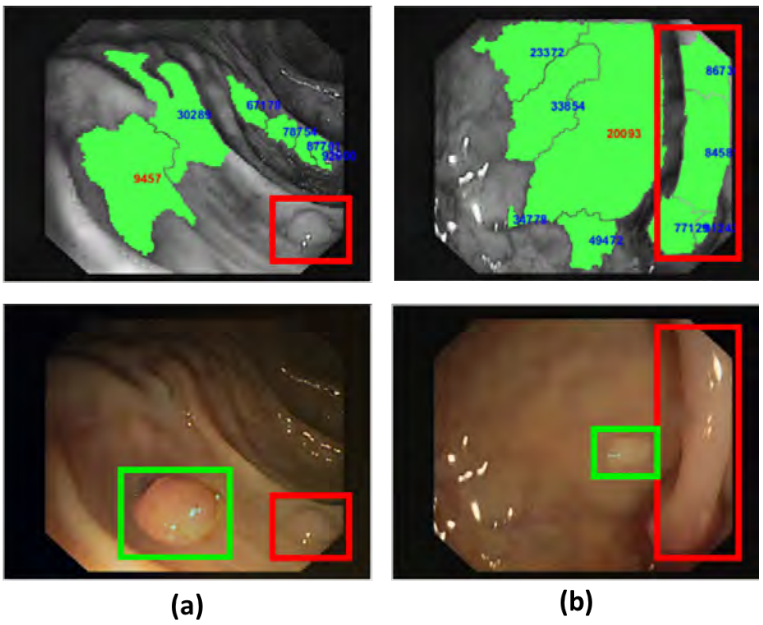

Figura 4.6: Ilustração das regiões *Watershed* geradas e do pólipo adicionado na imagem destino. Coluna (a) imagem acima: regiões *Watershed* e pólipo original delimitado por retângulo vermelho. Coluna (a) imagem abaixo: novo pólipo inserido indicado pelo retângulo verde e pólipo original demarcado em vermelho. Coluna (b) imagem acima: regiões *Watershed* que estão sobre uma área de pólipo já existente indicado pelo retângulo em vermelho. Coluna (b) imagem abaixo: novo pólipo indicado pelo retângulo verde e o pólipo original delimitado pelo retângulo vermelho.

destino.

Uma ilustração deste contexto pode ser vista na Figura 4.6. As regiões estão destacadas em verde, sendo que as maiores apresentam a quantidade de *pixels* indicada com números em vermelho, conforme ilustrado na Figura 4.6. Na coluna (a) o novo pólipo (retângulo verde) foi posicionado sobre a maior região *Watershed* e nenhuma das áreas retornadas estão sobre o pólipo que já estava presente na imagem destino (retângulo vermelho). Já na coluna (b), há algumas regiões (retângulo vermelho) que estão justamente sobre uma área de pólipo. Se alguma destas regiões fosse a maior, o pólipo não seria inserido dentro desta e uma nova região, maior que as restantes, seria selecionada.

Neste processo, podem ocorrer casos em que não existam áreas compatíveis na imagem destino para inserção do pólipo, com isso esta imagem é descartada e uma nova imagem do conjunto de dados original é escolhida como imagem destino.

Por fim, após selecionada uma região *Watershed* compatível, a área contendo pólipo (Figura 4.4 (c)) é inserida na imagem destino (Figura 4.4 (d)).

Porém, existem diferenças de textura e iluminação entre a imagem origem

e a imagem destino. Este problema é minimizado com o uso da técnica *Poisson* [101] (Seção 2.1.5). Esta abordagem permite adequar o aspecto visual do pólipo inserido, que apresenta uma estrutura com contornos complexos, em um novo plano de fundo, i.e., a imagem destino.

Com a aplicação da técnica *Poisson* não há necessidade de delinear precisamente a borda do pólipo. Sendo interessante utilizar uma caixa delimitadora para que regiões próximas ao pólipo na imagem origem também sejam obtidas com o intuito de favorecer a suavização de textura, iluminação e cor na imagem destino.

Esta estratégia de suavização tem por objetivo deixar a imagem mais coerente visualmente de modo a não deixar perceptível que um pólipo pertencente a uma outra imagem foi inserido. A Figura 4.7 apresenta a eficácia da técnica *Poisson* na redução das diferenças de cor e iluminação. O pólipo selecionado a partir da imagem original (a) é inserido na imagem destino (b). Porém, o resultado expõe a diferença evidente entre o pólipo inserido e a imagem destino (b). O emprego da técnica de *Poisson* reduz as diferenças tornando o aspecto visual na imagem final (c) mais coerente. A estratégia de suavização é importante para manter a aparência realista deste novo conjunto de imagens que será utilizado para treinamento da rede neural.

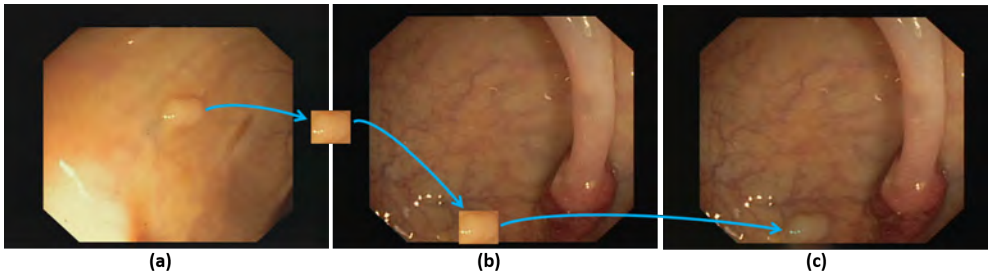

Figura 4.7: Demonstração do efeito da técnica *Poisson* que aplica uma transição suave entre bordas bordas da caixa delimitadora do pólipo e a imagem destino.

Além disso, é fundamental para o treinamento a composição da imagem *ground truth* após o novo pólipo estar presente na imagem destino. Isto pode ser visto na Figura 4.4 (e), onde há uma área com *pixels* em branco menor que corresponde ao pólipo que já estava presente na imagem destino e na área maior representando a forma e posição do novo pólipo inserido.

Para formar uma nova imagem *ground truth* coerente com a imagem que recebeu o novo pólipo, uma caixa delimitadora é posicionada em torno da área do pólipo na imagem origem conforme ilustrado na Figura 4.4 (b). Após, uma cópia desta área é inserida na mesma localização onde o novo pólipo foi

posicionado (Figura 4.4 (d)), porém sobre a imagem *ground truth* vinculada a respectiva imagem destino (Figura 4.4 (e)).

4.3 Experimentos

Nesta seção será descrita a organização das imagens para formação dos conjuntos de treinamentos e testes, assim como os experimentos conduzidos a partir destes conjuntos. Os efeitos do uso de cada conjunto de treinamento estão descritos na Seção 4.4.

O objetivo é avaliar a estratégia de inserção de pólipos no conjunto de dados CVC-ClinicDB para criar novas amostras, conforme proposto nesta tese (Seção 4.2), e a partir disso, comparar o efeito nos resultados da segmentação de pólipos baseada na rede U-net [119], com as técnicas tradicionais de aumento de dados (*augmentations*), normalmente aplicadas a conjuntos de tamanho reduzido. As técnicas de aumento de dados aplicadas nos experimentos foram transformações de rotação (40 graus), *zoom* (20%), deslocamentos horizontais e verticais (20%) e espelhamento horizontal. Os resultados demonstram que a utilização das técnicas tradicionais de aumento de dados (*augmentations*) juntamente com o aprimoramento de imagens por meio da estratégia de inserção pólipos favorece o aprendizado da rede de segmentação, sendo melhor do que somente empregar o aumento de dados tradicional.

Nestes experimentos foram empregados dois conjuntos de dados de imagens de colonoscopia: CVC-ClinicDB [10] que é utilizado para treinamentos e testes, assim como o conjunto ETIS-LaribPolypDB [127] empregado unicamente em testes. O CVC-ClinicDB é composto por 612 imagens, onde cada imagem apresenta um ou mais pólipos, sendo 31 pólipos diferentes (i.e. um mesmo pólipo pode aparecer em várias imagens, porém capturado pela câmera do colonoscópio a partir de pontos de vista diferentes). Para cada imagem de pólipos também são fornecidas as respectivas imagens *ground truth*, que são máscaras binárias que correspondem a localização do pólipo na imagem. Os dados do ETIS-LaribPolypDB formam um conjunto de 196 imagens e suas respectivas imagens *ground truth*, totalizando 44 pólipos diferentes. Em cada imagem deste conjunto há pelo menos um pólipo, totalizando 208 pólipos.

4.3.1 Conjuntos de imagens para treinamento e teste

O conjunto de imagens CVC-ClinicDB foi utilizado para formar conjuntos de treinamento e um conjunto de teste com 100 imagens para validação. As imagens em ETIS-LaribPolypDB foram empregadas exclusivamente para

formar um conjunto de teste (validação). A lista destes conjuntos pode ser vista na Tabela 4.1.

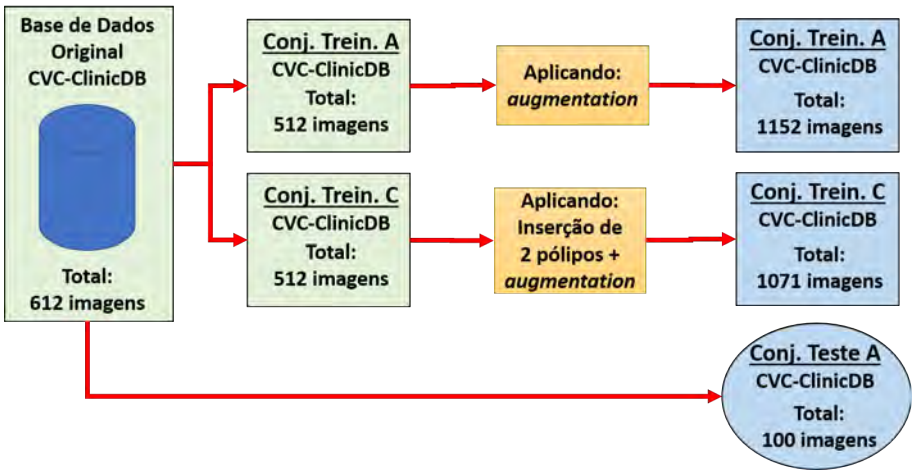

Figura 4.8: Organização dos conjuntos de treinamento e teste empregando as imagens de CVC-ClinicDB para validação das imagens do conjunto de teste A (CVC-ClinicDB).

A Figura 4.8 apresenta a organização dos conjuntos de treinamento utilizados na validação das imagens do conjunto de teste A. O conjunto de treinamento A é formado a partir de 512 imagens do CVC-ClinicDB que são aumentadas (técnicas tradicionais de *data augmentations*), formando o conjunto A usado para treinamento com total de 1152 imagens. Estas mesmas 512 imagens formam inicialmente conjunto de treinamento C, que é aprimorado com dois pólipos inseridos e também aumentado, totalizando 1071 imagens. As 100 imagens restantes do CVC-ClinicDB constituem o conjunto de teste A (validação), conforme representado na Figura 4.8. Este conjunto de teste contém de três a quatro imagens que representam cada um dos 31 pólipos diferentes do conjunto CVC-ClinicDB e nenhuma imagem presente no conjunto de teste A (validação) está presente no conjunto de treinamento A ou C.

Outros conjuntos de treinamento utilizados na validação das imagens do conjunto de teste B (ETIS-LaribPolypDB) podem ser vistos na Figura 4.9. Neste caso, foram utilizadas todas as 612 imagens do conjunto CVC-ClinicDB. Assim, o conjunto de treinamento B contém todas as 612 imagens do CVC-ClinicDB que após aumentado possui 1071 imagens. Outra versão aumentada do conjunto de treinamento B também foi formada a partir das 612 imagens CVC-ClinicDB totalizando 3825 imagens. Finalmente, os conjuntos de treinamento D foram construídos seguindo a mesma metodologia para compor no total 1071 (Conj. Trein. D com dois[2] pólipos inseridos) e 3823 (Conj. Trein.

[2]Referentes aos pólipos das imagens 27.tif e 333.tif do CVC-ClinicDB.

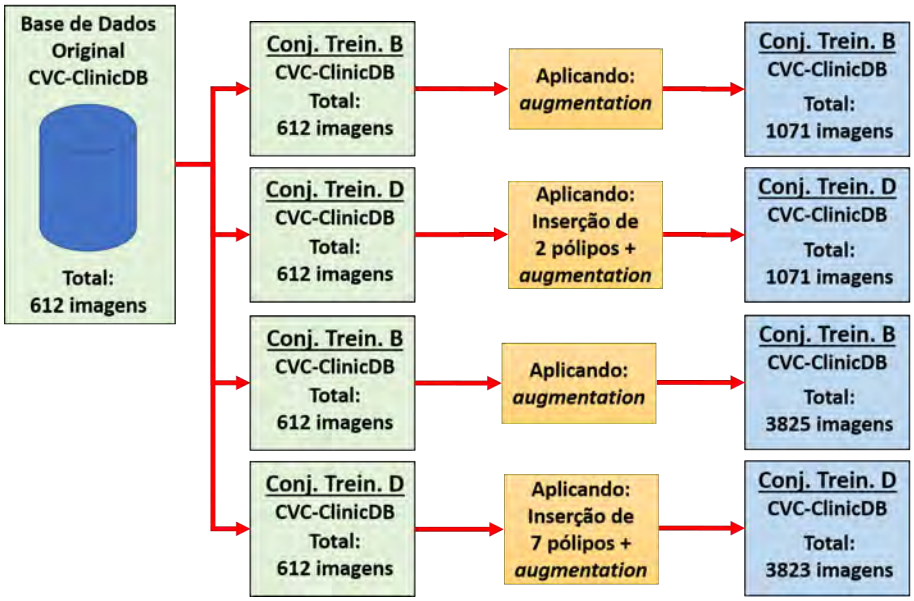

Figura 4.9: Organização dos conjuntos de treinamento empregando as imagens de CVC-ClinicDB para validação das imagens do conjunto de teste B (ETIS-LaribPolypDB).

D com sete[3] pólipos inseridos) imagens, respectivamente, além de aumentados com as técnicas tradicionais. Em todos os conjuntos de treinamento e teste as imagens foram pré-processadas aplicando um redimensionamento das imagens do CVC-ClinicDB e ETIS-LaribPolypDB para 256 x 256 *pixels*.

Exemplos das imagens presentes nos conjuntos CVC-ClinicDB e ETIS-LaribPolypDB podem ser vistos nas colunas (a) e (b) da Figura 4.10 respectivamente. Na coluna (c) está representada uma imagem gerada pela estratégia de inserção de pólipos, que faz parte dos conjuntos de treinamento C e D.

4.3.2 Métricas de avaliação e detalhes dos experimentos

A abordagem escolhida para segmentação dos pólipos está baseada na arquitetura da rede U-net. A saída desta rede é um mapa de segmentação que destaca a localização prevista do pólipo para cada imagem de entrada. Este mapa é transformado para uma versão binária (método Otsu [98]), com finalidade de ser comparado diretamente com a imagem *ground truth* fornecida pelos conjuntos de imagens CVC-ClinicDB e ETIS-LaribPolypDB. Exemplos destas imagens estão representados na Figura 4.14 e na Figura 4.15, onde na coluna (a) estão as imagens de teste, as imagens resultado na coluna (b), além das versões binárias e das imagens *ground truth* nas colunas (c) e (d), respectivamente.

[3]Referentes aos pólipos das imagens 27.tif, 333.tif, 52.tif, 53.tif, 60.tif, 200.tif e 280.tif do CVC-ClinicDB.

Tabela 4.1: Lista dos conjuntos de imagens de treinamento e teste para segmentação.

Nome	**Imagens Utilizadas**	**Quant. Imagens**
Conj. Trein. A	CVC-ClinicDB 512 + augmentation	1152
Conj. Trein. B	CVC-ClinicDB 612 + augmentation	1071
Conj. Trein. B	CVC-ClinicDB 612 + augmentation	3825
Conj. Trein. C	CVC-ClinicDB 512 + (Dois pólipos adicionados + augmentation)	1071
Conj. Trein. D	CVC-ClinicDB 612 + (Dois pólipos adicionados + augmentation)	1071
Conj. Trein. D	CVC-ClinicDB 612 + (Sete pólipos adicionados + augmentation)	3823
Conj. Teste A	CVC-ClinicDB	100
Conj. Teste B	ETIS-LaribPolypDB	196

Para avaliar estas comparações quantitativamente foram utilizadas as métricas de avaliação de precisão, revocação e F1 conforme apresentadas nos estudos [11, 140, 150]. Uma descrição destas métricas pode ser encontrada no estudo de Bernal et al. [14]. Adicionalmente, foi incluída a métrica de taxa de falso positivo (*False Positive Rate* (FPR)) para quantificar a proporção de áreas que não pertencem a pólipos mas foram sinalizadas como pólipos pela rede U-net. As implementações utilizadas estão disponíveis na biblioteca de aprendizado de máquina *Scikit-learn* [100]. Os valores retornados pelas métricas são referentes aos resultados dos experimentos com cada conjunto de treinamento listado na Tabela 4.1.

Na etapa de treinamento da rede U-net foi empregado o otimizador *Adam* [34] com taxa de aprendizado de 0.0001 e tamanho do *batch* igual a 4. O treinamento termina automaticamente (*early stopping*) quando não ocorre melhoria em dez épocas consecutivas, a partir do monitoramento da função

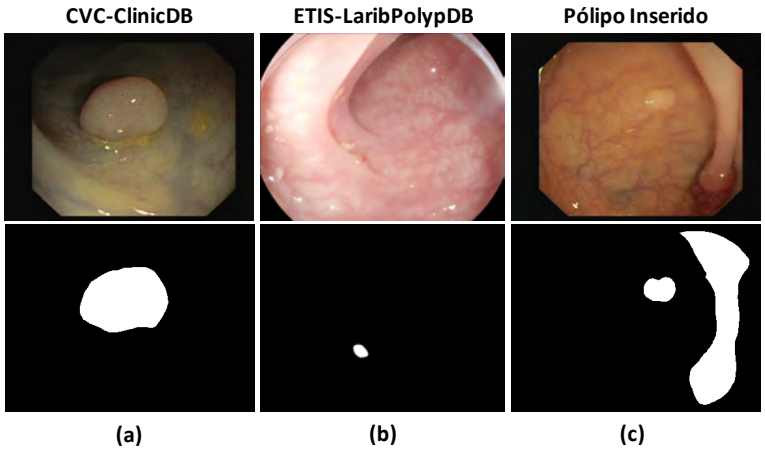

Figura 4.10: Exemplos de imagens de colonoscopia e as respectivas imagens *ground truth* presentes nos conjuntos de dados utilizados nos experimentos. Coluna (a): CVC-ClinicDB. Coluna (b): ETIS-LaribPolypDB. Coluna (c): CVC-ClinicDB com pólipo inserido.

custo[4].

As implementações e execuções dos treinamentos e testes foram efetuadas em um computador PC *desktop* com sistema operacional *Windows* 10, com processador Intel (R) Core (TM) i7-7700HQ 2.80GHz, 16GB de memória RAM e placa gráfica NVidia GeForce GTX 1060 com 6GB de memória dedicada.

4.4 Resultados

Os experimentos de validação demonstram o efeito dos conjuntos de treinamento C e D propostos (i.e., com pólipos inseridos e aumento de dados) em comparação com os conjuntos de treinamento A e B (i.e., somente com técnicas padrão de aumento de dados). As avaliações apresentadas nesta seção foram executadas sobre dois conjuntos de testes diferentes. O primeiro é o conjunto A, composto por 100 imagens selecionadas do CVC-ClinicDB e o segundo é o conjunto B formado pelo ETIS-LaribPolypDB com todas as suas 196 imagens.

A Tabela 4.2 apresenta os resultados do experimento considerando o conjunto de teste A com 100 imagens (CVC-ClinicDB). Os valores demonstram que houve uma melhoria das métricas de precisão, revocação, F1 e taxa de falso positivo (FPR) quando o conjunto de treinamento C (proposto) foi empregado, com destaque FPR que foi reduzida para 0.005, demonstrando o efeito da utilização das imagens aprimoradas por meio do processo de inserção de pólipos. A quantidade total de imagens em ambos os conjuntos de treinamento

[4]Também conhecida como função de perda ou erro (*binary cross-entropy* neste caso).

é semelhante (i.e. 1152 e 1071), para evitar que uma quantidade amostras desbalanceada influencie a comparação.

Tabela 4.2: Experimento de validação sobre o conjunto conjunto de teste A (CVC-ClinicDB). Comparativo entre os conjuntos de treinamento aumentado e com pólipos inseridos (proposto).

Nome	**Prec.**	**Revoc.**	**F1**	**FPR**
Conj. Trein. A (Aumentado) - 1152 imagens	0.918	0.867	0.891	0.014
Conj. Trein. C (Proposto) - 1071 imagens	**0.979**	**0.938**	**0.957**	**0.005**

Os resultados listados na Tabela 4.3 representam os experimentos de validação sobre o conjunto de teste B (ETIS-LaribPolypDB) com 196 imagens. Comparando o conjunto de treinamento B (aumentado, 1071 imagens) com o conjunto D (proposto, 1071 imagens) percebe-se um aumento da precisão, melhora na revocação de 0.125 e de 0.095 para F1 e a taxa de falso positivo (FPR) foi reduzida para 0.0292. Apesar de ambos os conjuntos de imagens possuírem a mesma quantidade de imagens (i.e. 1071), o conjunto proposto D apresentou melhores resultados neste experimento.

A Tabela 4.3 também exibe a avaliação entre o conjunto de treinamento B (aumentado) com 3825 imagens e o conjunto de treinamento D (proposto) com 3823 imagens. Observa-se que há melhorias em termos de revocação de 0.217 e F1 de 0.14, com redução na taxa de falso positivo (FPR) para 0.079. Os valores obtidos usando o conjunto de treinamento D (proposto, 3823) foi superior em todas as métricas, exceto na precisão que manteve um valor próximo do apresentado pelo conjunto B (aumentado, 3825). Também neste segundo experimento de validação, que emprega o conjunto de teste B (ETIS-LaribPolypDB), os resultados mostram que a estratégia de inserção de pólipos proposta é promissora e superior nesta avaliação do que unicamente a aplicação de técnicas tradicionais de aumento de dados.

A Figura 4.11 apresenta mais claramente a comparação entre os conjuntos de treinamento propostos e os conjuntos de dados aumentados (Tabela 4.3), segundo as métricas de precisão e revocação, no contexto do teste sobre conjunto de dados ETIS-LaribDB. No eixo horizontal da Figura 4.11 estão listados os conjuntos de treinamento e no eixo vertical os valores de precisão e revocação alcançados em cada experimento.

Tabela 4.3: Experimento de validação sobre o conjunto conjunto de teste B (ETIS-LaribPolypDB). Comparativo entre os conjuntos de treinamento aumentado e com pólipos inseridos (proposto).

Nome	**Prec.**	**Revoc.**	**F1**	**FPR**
Conj. Trein. B (Aumentado) - 1071 imagens	0.934	0.581	0.697	0.423
Conj. Trein. D (Proposto) - 1071 imagens	**0.939**	0.706	0.792	0.292
Conj. Trein. B (Aumentado) - 3825 imagens	**0.939**	0.681	0.774	0.318
Conj. Trein. D (Proposto) - 3823 imagens	0.936	**0.898**	**0.914**	**0.079**

O número total de imagens foi ajustado para permitir a comparação dos conjuntos de dados de treinamento com uma quantidade de amostras semelhante. Por exemplo, conjunto de treinamento B (aumentado) e o conjunto D (proposto) com 1071 imagens. O mesmo ocorre para os conjuntos B (aumentado) e D (proposto) com 3825 e 3823 imagens, respectivamente.

Especificamente, o uso de 3823 imagens com pólipos inseridos (conjunto de treinamento D proposto) apresenta 0.217 de melhoria em revocação em comparação ao uso do conjunto de treinamento aumentado B com 3825 imagens. No geral, o conjunto de treinamento D, mostra melhorias consideráveis de revocação, mantendo o valor de precisão similar.

Os resultados dos experimentos nas Tabelas 4.2 e 4.3 apresentam uma redução significativa de falsos positivos para o conjunto de treinamento C e D em que os pólipos foram inseridos. Nesta métrica, os valores mais baixos são preferíveis. Especificamente, na Figura 4.12, estão listados os valores de FPR obtidos no experimento sobre o conjunto de teste A com 100 imagens (CVC-ClinicDB), onde apenas o uso da técnica de aumento de dados tradicional (1152 imagens) apresentou FPR de 0.014 e com a adição da estratégia de inserção de pólipos (1071 imagens) alcançou 0.005.

Semelhantemente, no caso dos experimentos no conjunto de dados ETIS-LaribPolypDB (conjunto de teste B), os valores de FPR também tiveram uma redução importante, conforme apresentado na Figura 4.13. Os valores de FPR mostram que quando ambos os conjuntos de dados de treinamento D propostos (1071 e 3823 imagens) são empregados, ocorrem melhorias em relação a utilização do aumento de dados tradicional apenas. Os resultados indicam que

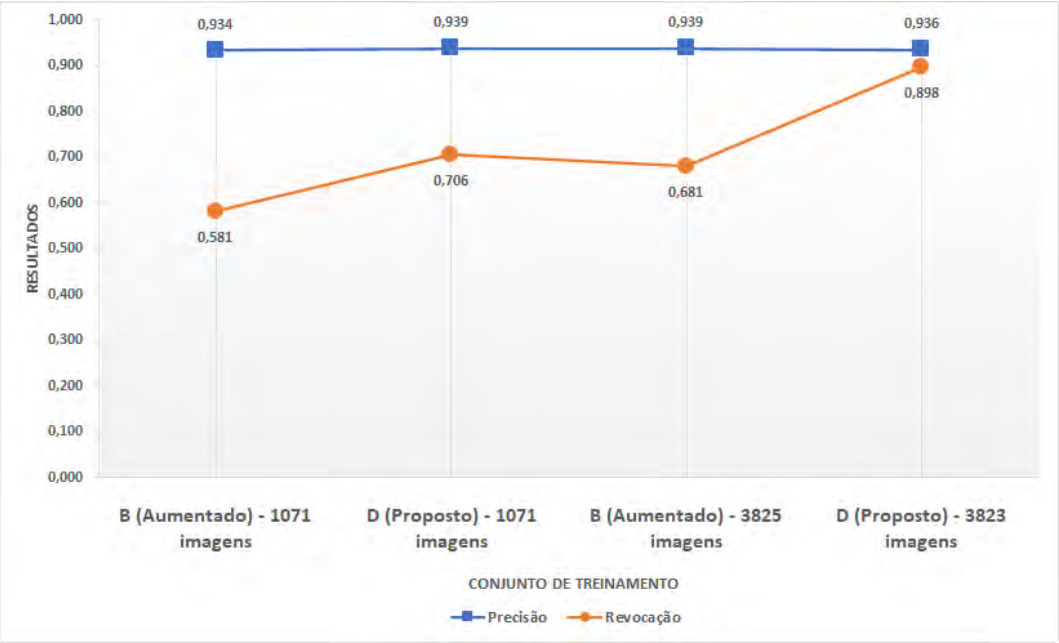

Figura 4.11: Experimento sobre o conjunto de teste B (ETIS-LaribPolypDB). Gráfico comparativo entre os valores de precisão e revocação considerando a quantidade de imagens semelhantes nos conjuntos de treinamento aumentados e propostos.

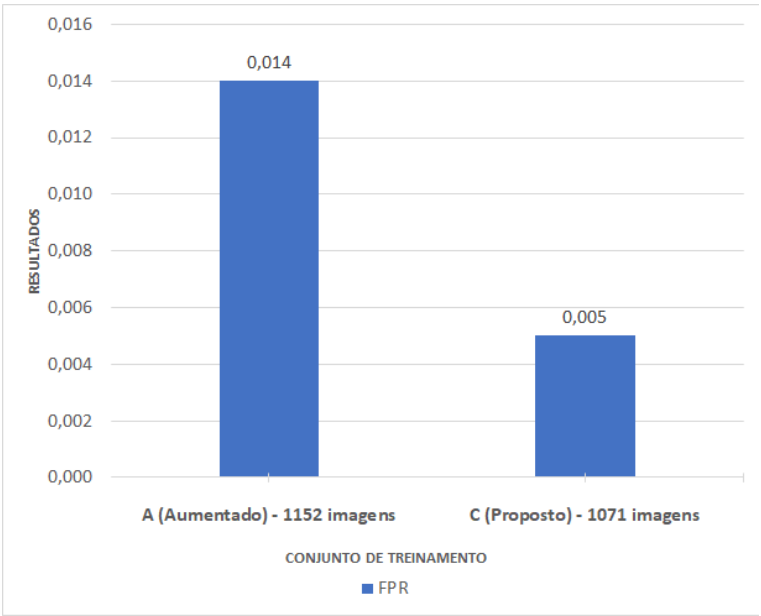

Figura 4.12: Redução da taxa de falso positivo (FPR) no experimento com o conjunto de teste A (CVC-ClinicDB).

o uso do conjunto de treinamento aprimorado D (proposto, com 1071 imagens) reduz os valores de FPR em comparação ao conjunto aumentado B que possui a mesma quantidade de imagens. Com o acréscimo de mais imagens, o conjunto de treinamento aprimorado D (proposto, com 3823 imagens) apresenta valores de FPR bem menores em comparação ao conjunto de dados B (aumentado, com 3825 imagens), mesmo com quantidade de imagens semelhantes, FPR de 0.079 e 0.318 respectivamente.

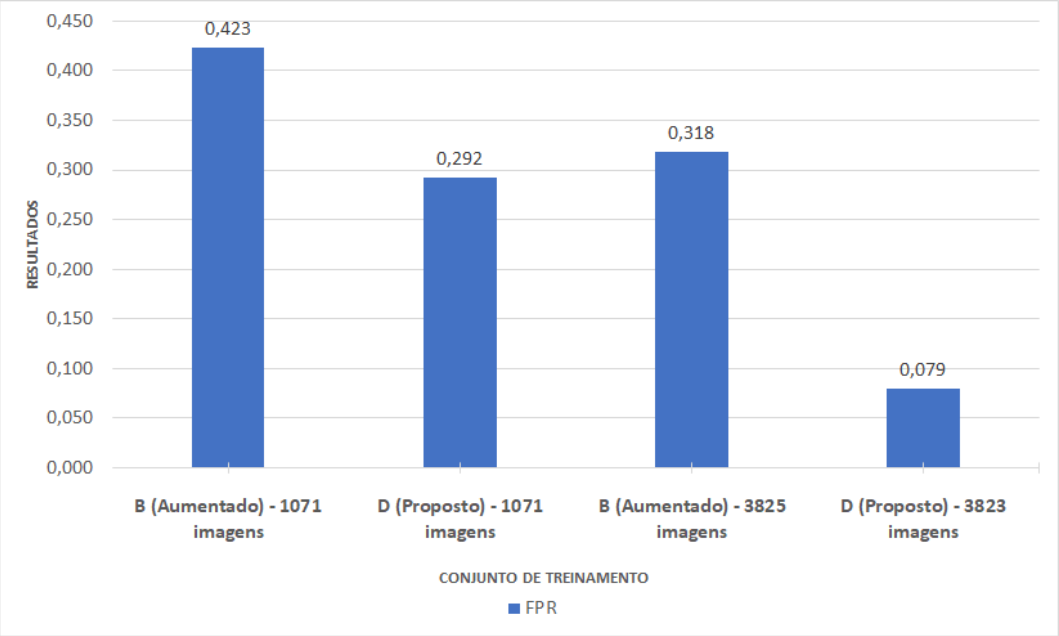

Figura 4.13: Redução da taxa de falso positivo (FPR) no experimento com o conjunto de teste B (ETIS-LaribPolypDB).

Os experimentos de segmentação com a rede U-net sobre o conjunto de teste A (CVC-ClinicDB) utilizando o conjunto de treinamento C (proposto) geraram as imagens apresentadas na Figura 4.14, por exemplo. A coluna (a) representa uma imagem de entrada do conjunto de teste A e a coluna (b) é a respectiva segmentação, onde os *pixels* mais claros representam a área ocupada pelo pólipo, segundo a previsão da rede U-net. O método Otsu [98] é empregado para obter uma versão binária da segmentação da rede, vista na coluna (c). A imagem da coluna (c) pode então ser comparada com a imagem da coluna (d), representa a área real do pólipo (*ground truth*).

Exemplos de imagens segmentadas no experimento sobre o conjunto de teste B (ETIS-LaribPolypDB) operando com o treinamento D (3823), no qual os pólipos foram inseridos, podem ser vistas na Figura 4.15. O resultado da rede U-net (Coluna b) foi pós-processado utilizando o Otsu (Coluna c) e por fim comparado as imagens *ground truth* (Coluna d) fornecidas pelo conjunto de teste ETIS-LaribPolypDB.

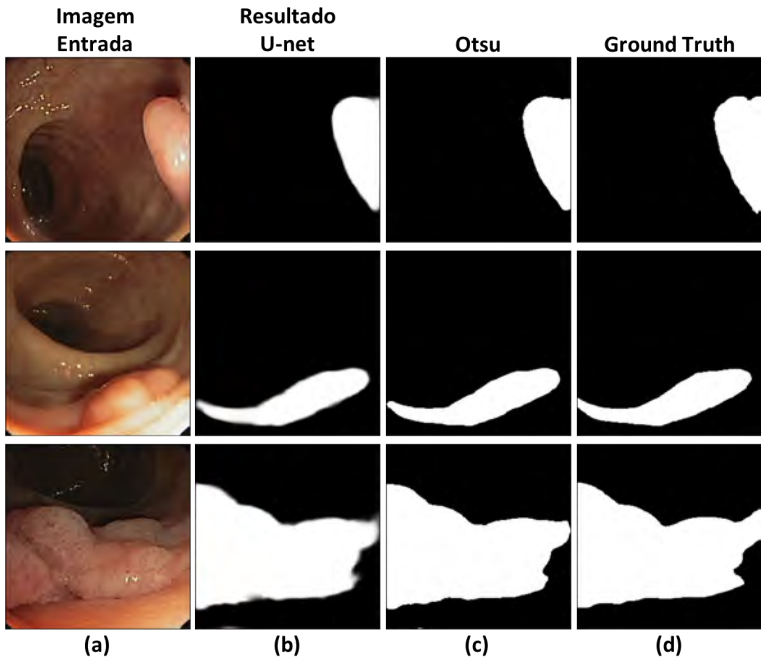

Figura 4.14: Exemplos dos resultados de segmentação usando o conjunto de treinamento C (Tabela 4.2) sobre o conjunto de teste A. Colunas: (a) Imagem de entrada do conjunto de teste A (CVC-ClinicDB). (b) Segmentação resultante da rede U-net. (c) Versão binária da imagem de segmentação. (d) Respectiva imagem *ground truth*.

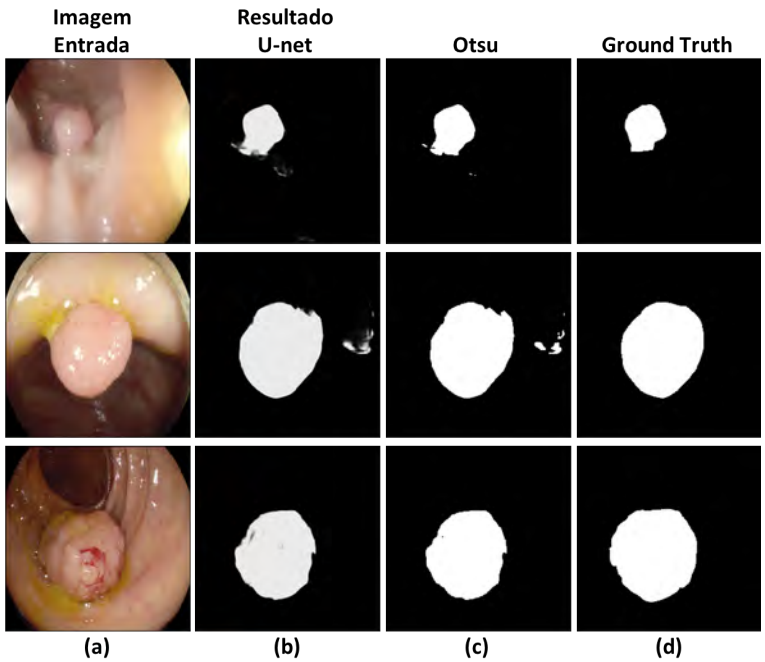

Figura 4.15: Exemplos dos resultados de segmentação usando o conjunto de treinamento D com 3823 amostras (Tabela 4.3) sobre o conjunto de teste B. Colunas: (a) Imagem de entrada do conjunto de teste B (ETIS-LaribPolypDB). (b) Segmentação resultante da rede U-net. (c) Versão binária da imagem de segmentação. (d) Respectiva imagem *ground truth*.

Na Figura 4.16 estão representados casos em que os pólipos não foram corretamente localizados pela rede U-net. Exemplos referentes ao conjunto de teste B (ETIS-LaribPolypDB) estão expostos nas duas linhas superiores, enquanto na linha inferior está a saída referente ao experimento com conjunto de teste A (CVC-ClinicDB).

A coluna (c) mostra os resultados segmentação, já binarizados, que não estão localizados corretamente em comparação às imagens *ground truth* (d). Os pólipos da imagem na primeira linha são muito difíceis de ver. Na segunda linha, é possível observar um exemplo de conteúdo intestinal que favorece uma localização incorreta do pólipo, gerando um falso positivo. A imagem da última linha é muito distinta das demais no conjunto de dados da CVC-ClinicDB com relação à cor e iluminação, que provavelmente contribui para uma segmentação falha.

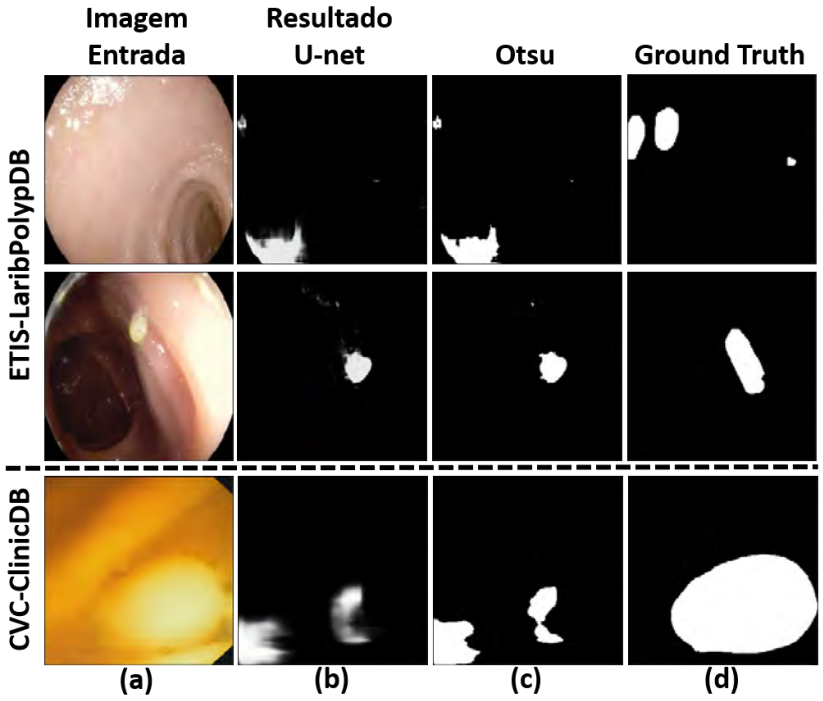

Figura 4.16: Exemplos de casos com segmentação incorreta. Acima: imagens referentes ao conjunto de teste B (ETIS-LaribPolypDB). Abaixo: imagens referentes ao conjunto de teste A (CVC-ClinicDB). Coluna (a): imagem do cojunto de teste. Coluna (b): resultado da segmentação. Coluna (c): resultado da segmentação binarizado. Coluna (d): imagem *ground truth*.

4.4.1 Discussão

Os resultados demonstram que há uma redução significativa na FPR. Isso significa que o processo de inserção de pólipos contribui para o aprendizado

da rede U-net em relação às estruturas que podem ser confundidas com os pólipos, por exemplo, reflexos, bolhas e material fecal.

Os resultados de validação do conjunto de teste A (CVC-ClinicDB) demonstraram que o conjunto de treinamento C proposto alcançou valores significantes em todas as métricas. As imagens deste conjunto de teste A fazem parte da mesma base de dados de imagens utilizadas no treinamento (i.e. CVC-ClinicDB) e podem ter favorecido os valores. Isto ocorre pois imagens que representam quadros sequenciais do vídeo são em geral semelhantes, logo uma imagem no conjunto de teste A pode ser similar (não exatamente a mesma imagem) a outra presente no conjunto treinamento A ou C. No entanto, a estratégia de inserção de pólipos (conjunto de treinamento C) influenciou positivamente os resultados em relação ao conjunto de treinamento A que utiliza somente as técnicas tradicionais de aumento de dados.

O experimento sobre o conjunto de teste B (ETIS-LaribPolypDB) utilizou somente imagens de treinamento do conjunto CVC-ClinicDB, que possui uma distribuição e qualidade de imagem diferentes. Isto é um indicador da capacidade de generalização da rede, pois as imagens do conjunto ETIS-LaribPolypDB nunca foram usadas no treinamento, mas os resultados obtidos são relevantes quando os conjuntos de treinamento aprimorados são empregados.

Existem problemas relacionados à localização do pólipo inserido na imagem de destino. Em certos casos, o pólipo pode ser posicionado em uma região que pode afetar a aparência realista da imagem. Um exemplo é quando um pólipo é posicionado sobre uma dobra das paredes do cólon, conforme visto na Figura 4.17 (a) (imagem superior e inferior).

Outras dificuldades ocorreram por causa da reflexão da luz no pólipo inserido, que não é compatível com a imagem de destino. Nesses casos, o pólipo foi posicionado em uma área com pouca luminosidade (imagem de destino), mas os reflexos da luz presentes no pólipo (devido à imagem de origem) mantêm maior intensidade do que as regiões do em torno, conforme visto na Figura 4.17 (b) (imagem superior). Outro problema é a diferença de foco entre a imagem origem e a imagem destino. O pólipo extraído tem uma textura mais nítida, mas está posicionado sobre uma área na imagem de destino que está embaçada, causando uma discrepância, conforme visto na Figura 4.17 (b) (imagem inferior).

Apesar disso, a estratégia de inserção de pólipos se mostrou promissora nos experimentos de segmentação realizados, em especial sobre o conjunto de teste B (ETIS-LaribPolypDB), melhorando a revocação, F1 e taxa de falso positivo. Este recurso para o aprimoramento do conjunto de imagens pode ser

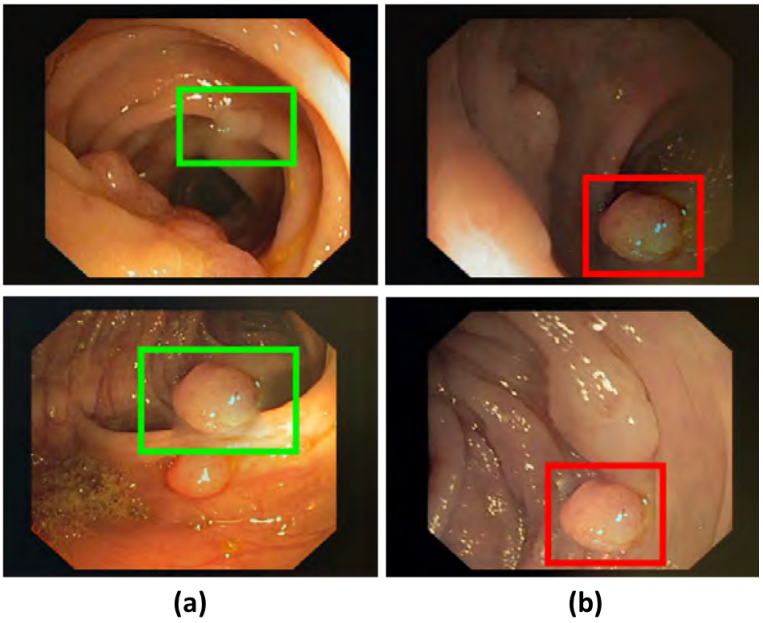

Figura 4.17: Problemas encontrados no processo de inserção de pólipos. Coluna (a): exemplo de inserção de pólipo que afeta a aparência realista da imagem. Coluna (b) imagem acima: discrepância relacionada à luminosidade. Coluna (b) imagem abaixo: pólipo com aparência mais nítida que imagem destino.

proveitoso em um contexto onde a disponibilidade de dados para treinamento é reduzida.

Esta abordagem é capaz de aumentar a variedade das amostras e produzir automaticamente as respectivas imagens *ground truth* mantendo o formato e indicando a nova localização dos pólipos inseridos. As imagens *ground truth* são necessárias para o aprendizado supervisionado e a dependência de especialistas para determinar o formato e localização das lesões é um fator que dificulta a produção de conjuntos de treinamentos maiores. Este problema pode ser atenuado com a estratégia proposta nesta tese.

5
Aprimoramento de Dados para Detecção com Pólipos Reais e Sintéticos

Este capítulo é uma continuação da estratégia de aprimoramento do conjunto de imagens de colonoscopia apresentada anteriormente (Capítulo 4). No entanto, foram adicionados um conjunto de treinamento e outro para testes além dos já descritos. A arquitetura de rede *Faster* R-CNN será empregada para a detecção dos pólipos, ao invés da rede de segmentação U-net. O mesmo processo de inserção de pólipos foi utilizado para aumentar a quantidade e variedade dos conjuntos de dados. Novas amostras de pólipos sintéticos e suas respectivas imagens *ground truth* foram criadas por meio das redes generativas (GAN), para serem adicionadas nas imagens de colonoscopia. Assim, tanto pólipos originais quanto sintéticos serão utilizados para o aprimoramento do conjunto de dados.

5.1
Visão Geral

O procedimento apresentado para o aprimoramento dos dados destaca o uso de pólipos sintéticos. Isto possibilita melhorar variação de textura e forma dos pólipos inseridos. A etapa de criação destes pólipos sintéticos recebe como entrada os pólipos selecionados pertencentes ao conjunto de dados original. A rede generativa produz novos pólipos sintéticos condicionados a uma máscara binária que define a forma destes pólipos. As amostras sintéticas foram inseridas em imagens sem pólipos e com pólipos. Novos conjuntos de dados de treinamento aprimorados foram criados utilizando o processo de inserção de pólipos combinando pólipos originais e sintéticos.

A Figura 5.1 ilustra a utilização dos pólipos originais e sintéticos para formar o conjunto de dados aprimorado. Os pólipos originais são selecionados do conjunto de imagens (a) junto com as respectivas máscaras binárias (b) para o processo de inserção (c). O processo de inserção pode fazer uso de pólipos diretamente extraídos do conjunto de imagens original, como descrito no Capítulo 4 ou utilizar pólipos sintéticos conforme apresentado no presente capítulo.

As etapas destacadas em cinza na Figura 5.1 são exclusivas para a

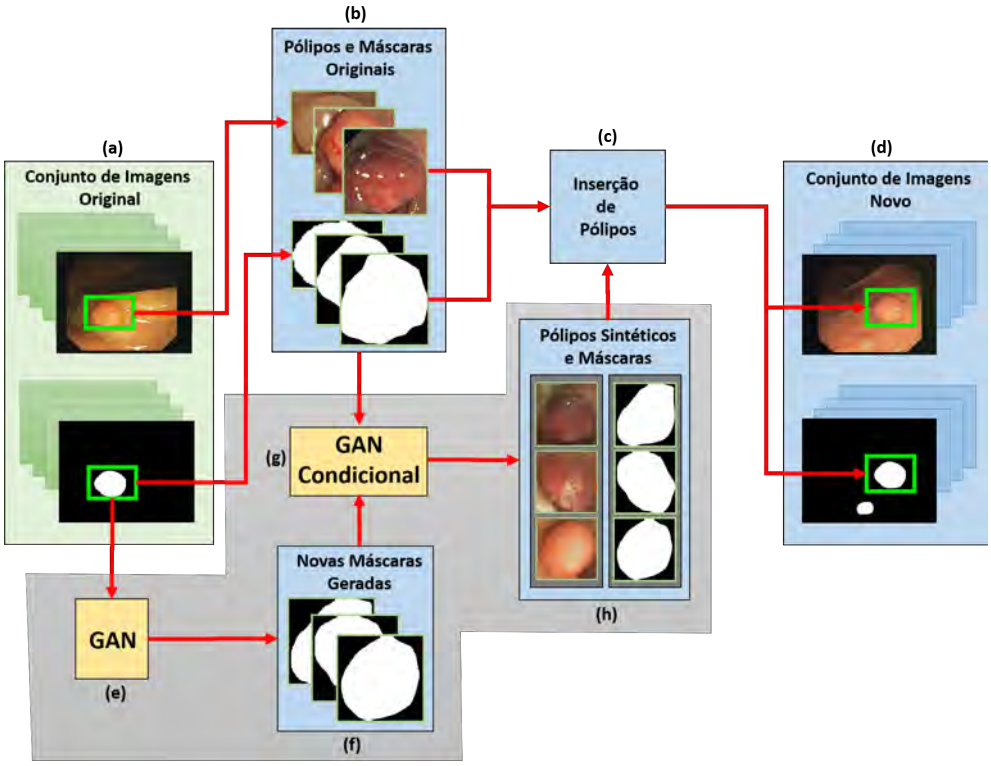

Figura 5.1: Visão geral do processo de inserção com pólipos sintéticos e originais. (a): pares de imagens de colonoscopia no conjunto de dados original. (b): áreas dos pólipos selecionadas para inserção. (c): processo de inserção de pólipos. (d): conjunto de dados aprimorado. (e): rede geradora para criar máscaras binárias. (f): máscaras binárias sintéticas. (g): rede geradora condicional. (h): pares de pólipos e máscaras binárias sintéticas para inserção.

geração dos pólipos sintéticos. A rede generativa (e) cria máscaras binárias com formatos variados (f), usadas para condicionar a forma dos pólipos (h), que serão empregados no processo de inserção (c). A região escolhida para a localização do pólipo na imagem destino segue o mesmo processo de inserção baseado na operação de *Watershed* [94], conforme mencionado no Capítulo 4.

5.2 Geração de Pólipos Sintéticos

A abordagem proposta aumenta a variabilidade dos dados criando novas amostras a partir da inserção de pólipos sintéticos em imagens de colonoscopia. A execução automática das etapas de geração e inserção de pólipos resultam em um conjunto de dados aprimorado para ser usado no treinamento de um modelo *Faster* R-CNN. Especificamente, foi empregado um *backbone* ResNet50 pré-treinado [59] a partir de imagens *ImageNet*. Como mencionado anteriormente, para aumentar a variabilidade dos dados foram definidos dois métodos para se obter os pólipos que serão inseridos em imagens de colonoscopia: 1) por meio

da seleção de pólipos originais do conjunto de dados (Figura 5.1 (a) e (b)) e 2) gerando novos pólipos sintéticos a partir de uma rede Adversária Generativa Condicional (CGAN) [24, 48, 95] (Figura 5.1 (h)).

A ilustração da segunda abordagem baseada em pólipos sintéticos pode ser vista na Figura 5.2. Em (b) as imagens de máscaras binárias com formatos variados são criadas pela rede GAN utilizando o conjunto de dados de imagens que contém as máscaras originais (a). Na próxima etapa, a rede GAN Condicional recebe como entrada as máscaras geradas (b) e os pares de máscaras e imagens de pólipos originais (c). Nesta fase, a rede aprenderá sobre a aparência das imagens de pólipos originais (c) e gerará novas imagens de pólipos (d) condicionadas à forma da máscara (b).

Após a geração do pólipo sintético, a próxima etapa determina a região na imagem de destino ao qual o pólipo será adicionado. O processo de inserção do pólipo sintético segue a mesma estratégia dos pólipos originais. Porém, no caso de pólipos gerados sinteticamente, o respectivo tamanho é definido no momento da inserção nas imagens de destino. O tamanho irá considerar a região de destino retornada pelo processo de *Watershed*, de modo que o pólipo tenha tamanho menor que a região. Além disso, o tamanho do pólipo pode variar entre uma faixa de valores de *pixels* definida empiricamente.

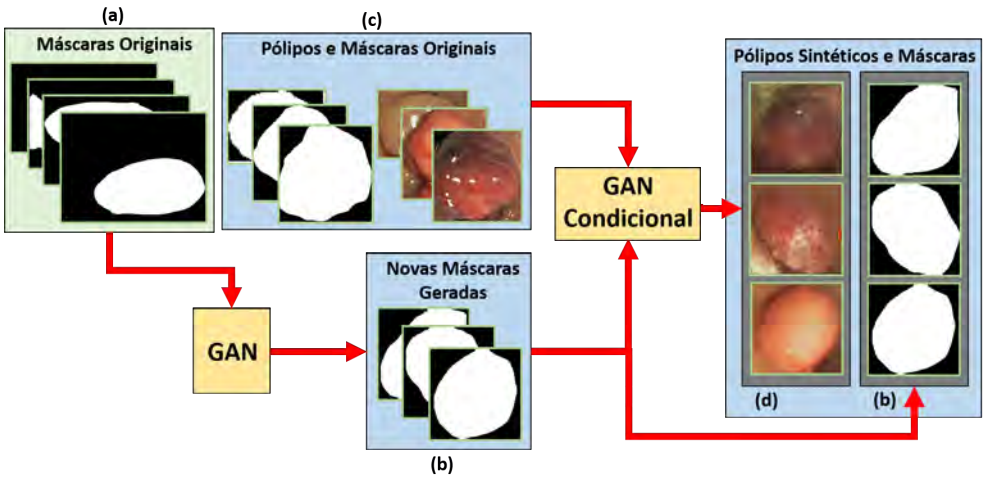

Figura 5.2: Ilustração do procedimento de geração de pólipos sintéticos. (a) Entrada da rede GAN (máscaras originais) e (b) saída (máscaras binárias sintéticas). (c) Imagens dos pólipos originais e as respectivas imagens *ground truth* utilizadas como entrada da rede GAN Condicional juntamente com as (b) máscaras sintéticas. (d) Imagens de pólipos sintéticos gerados a partir do GAN Condicional e respectivas (b) máscaras binárias.

Na etapa final, ilustrada pela Figura 5.3, é formada a nova representação do conjunto de dados na qual os pólipos sintéticos (a) e as máscaras (b) serão adicionadas. Neste caso, também é aplicada a técnica *Poisson* [101] (Capítulo

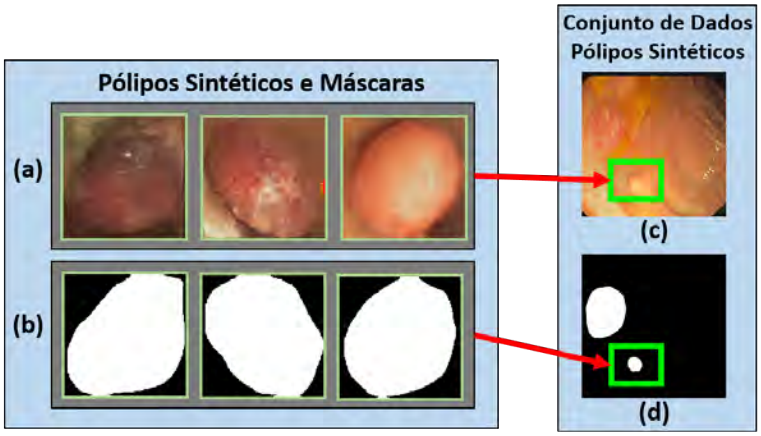

Figura 5.3: Representação da inserção de pólipos sintéticos. (a) pólipo sintético (saída da rede GAN condicional). (b) Máscara binária do pólipo sintético (saída da rede GAN). (c) Imagem destino recebe o pólipo sintético e (d) a imagem *ground truth* recebe a respectiva máscara binária.

2) para minimizar as diferenças de textura e iluminação entre o pólipo sintético (a) e a imagem de destino (c). Essa estratégia de suavização pretende tornar a imagem de destino visualmente coerente, de modo que não fique evidente a inserção do pólipo. Além disso, a respectiva imagem *ground truth* (d) também é produzida recebendo o formato e a localização do pólipo sintético.

A rede GAN foi desenvolvida com base na implementação padrão [48] e a rede GAN condicional em [24, 67]. A execução da rede GAN condicional para geração dos pólipos sintéticos foi realizada na máquina virtual do ambiente *Google Colaboratory* [18]. O treinamento consumiu aproximadamente 33 horas para execução de 15 épocas. O ambiente *Google Colaboratory* limita a execução a no máximo 12 horas[1], então o treinamento foi dividido em três etapas de 5 épocas cada. Esta máquina virtual disponibiliza processadores gráficos (GPU) Nvidia Tesla que variam entre os modelos K80, T4, P4 e P100.

5.3 Experimentos

Nesta seção estão descritos os procedimentos de treinamento e avaliação. Para cada experimento os resultados de desempenho entre o treinamento da rede *Faster* R-CNN com os dados originais e com os conjuntos de dados sinteticamente aprimorados foram comparados.

[1]Limite de tempo do ambiente para uso gratuito.

5.3.1 Conjuntos de imagens para treinamento e teste

Nos experimentos foram utilizados os conjuntos de dados publicamente disponíveis CVC-ClinicDB [10] e ASU-Mayo [138] para fins de treinamento, enquanto para avaliação foram empregados CVC-VideoClinicDB [4, 15] e o ETIS-LaribPolypDB [127]. A Tabela 5.1 apresenta a quantidade de imagens e pólipos presentes nestes conjuntos de imagens de colonoscopia.

As imagens do conjunto ASU-Mayo foram extraídas de vídeos com e sem pólipos, por isso estão divididas em dois grupos: imagens WP são as imagens que apresentam pólipos, e imagens NP, o que significa que não há pólipos na imagem. Além disso, foram selecionadas apenas imagens NP que mostram uma diferença significativa (ou seja, a aparência visual das imagens difere muito) entre todos as imagens NP no conjunto, resultando em apenas 139 imagens. Estas 139 imagens ASU-Mayo NP foram selecionadas para inserção de pólipos.

Os conjuntos de treinamento na Tabela 5.2 são conjuntos combinados com imagens de pólipos sintéticos e originais, exceto o conjunto de treinamento A que emprega somente as imagens do CVC-ClinicDB. Todas as imagens nos conjuntos de dados de treinamento e teste têm pelo menos um pólipo.

Tabela 5.1: Quantidade de imagens e pólipos em cada base de dados.

Base de dados	**Quant. Imagens**	**Quant. Pólipos**
CVC-ClinicDB	612	646
ASU-Mayo WP	3856	3856
ASU-Mayo NP	139	0
ETIS-LaribPolypDB	196	208
CVC-VideoClinicDB	10025	10025

Para estes experimentos foram propostos quatro conjuntos de dados de treinamento (i.e., conjunto de treinamento B, C, D e E na Tabela 5.2) nos quais os pólipos foram inseridos de acordo com o procedimento inserção (Capítulo 4). Existem dois tipos de pólipos, a saber, os pólipos sintéticos e originais (i.e. aqueles que pertencem as imagens originais do conjunto). O conjunto de treinamento B foi constituído por imagens do CVC-ClinicDB que receberam apenas pólipos originais. O conjunto de treinamento C, pólipos sintéticos foram inseridos em cada imagem do CVC-ClinicDB para compor novas amostras.

Da mesma forma, no conjunto de treinamento D na Tabela 5.2, pode-se observar que tanto o CVC-ClinicDB (com pólipos sintéticos adicionados) quanto as imagens com pólipos (ou seja, WP) do conjunto ASU-Mayo foram combinadas para fornecer um conjunto de dados mais diversificado. No caso do conjunto de treinamento E, foram escolhidas imagens sem pólipos ASU-Mayo (isto é, NP). Portanto, serão adicionados pólipos sintéticos nas imagens ASU-Mayo NP e pólipos originais nas imagens CVC-ClinicDB. O conjunto de treinamento A é composto exclusivamente das imagens originais do banco de dados CVC-ClinicDB (i.e., nenhum pólipo foi adicionado).

Outra característica dos conjuntos de treinamento é a quantidade de pólipos presentes em cada imagem. Neste contexto as imagens ASU-Mayo WP apresentam somente um pólipo original por imagem. No caso de ASU-Mayo NP, foram inseridos somente um pólipo por imagem. Logo, as imagens ASU-Mayo nos conjuntos de treinamento representam um pólipo por imagem. No entanto, imagens CVC-ClinicDB já possuem pelo menos um pólipo original por imagem e por meio da estratégia de inserção proposta mais pólipos foram inseridos. As imagens nos conjuntos de treinamento D e E foram combinadas para estabelecer um balanceamento entre imagens com um pólipo (ASU-Mayo) e com mais de um pólipo (CVC-ClinicDB).

Em resumo, as seguintes etapas foram realizadas para a composição dos conjuntos de treinamento listados na Tabela 5.2. Primeiramente, as imagens foram pré-processadas para a resolução de 256 x 256 *pixels*. Em seguida, foi efetuado o procedimento de inserção de pólipos para todos os conjuntos de dados de treinamento, exceto o conjunto de treinamento A que permaneceu com as imagens originais. Para estes experimentos com a rede de detecção *Faster* R-CNN, foram combinadas as abordagens de aprimoramento por meio da inserção de pólipos com estratégias tradicionais de aumento de dados (*augmentations*) em todos os conjuntos de treinamento. Assim, foi possível ampliar ainda mais a variação das amostras e a quantidade total de imagens como mostra a Tabela 5.2.

O conjunto de dados ETIS-LaribPolypDB contém 196 imagens e foi usado apenas para avaliação (conjunto de testes A na Tabela 5.2). O CVC-VideoClinicDB possui um total de 11954 imagens. No entanto, apenas 10025 amostras possuem pelo menos um pólipo, por isso, somente estas imagens foram empregadas para compor o segundo conjunto de dados para avaliar a detecção (conjunto de testes B na Tabela 5.2).

Tabela 5.2: Lista dos conjuntos de imagens de treinamento e teste para detecção.

Nome	**Composição do Conjunto de Imagens**	**Total Imagens**
Conj. Trein. A	CVC-ClinicDB	7356
Conj. Trein. B	CVC-ClinicDB + Pólipos Originais Adicionados	7666
Conj. Trein. C	CVC-ClinicDB + Pólipos Sintéticos Adicionados	7162
Conj. Trein. D	ASU-Mayo WP + (CVC-ClinicDB + Pólipos Sintéticos Adicionados)	6793
Conj. Trein. E	(ASU-Mayo NP + Pólipos Sintéticos Adicionados) + (CVC-ClinicDB + Pólipos Originais Adicionados)	6362
Conj. Teste A	ETIS-LaribPolypDB	196
Conj. Teste B	CVC-VideoClinicDB	10025

5.3.2 Métricas de avaliação e detalhes dos experimentos

Nestes experimentos foram adotadas duas métricas de avaliação dos resultados. No estágio de teste, as saídas da rede de detecção são coordenadas das caixas delimitadoras que fornecem a localização prevista do pólipo para cada imagem de entrada. Para cada experimento (i.e., conjunto de treinamento na Tabela 5.2) foram avaliadas as métricas de precisão e revocação[2]. Uma descrição dessas métricas pode ser encontrada em [14] e estas são definidas como abaixo:

$$\text{Precisão} = \frac{TP}{TP + FP} \tag{5-1}$$

$$\text{Revocação} = \frac{TP}{TP + FN} \tag{5-2}$$

As métricas de precisão e recuperação são definidas a partir dos valores chamados Verdadeiro Positivo (TP), Falso Positivo (FP) e Falso Negativo (FN). Estes valores especificam se a previsão de rede está correta ou não para cada pólipo presente em uma imagem. Isso significa que as caixas delimitadoras previstas e as caixas delimitadoras baseadas na imagem *ground truth* serão comparadas para mensurar o valor de sobreposição chamado *Intersection Over*

[2]Também chamada de sensibilidade.

Union (IoU). A IoU atua como um valor limite que indica se a sobreposição de caixas delimitadoras será considerada como TP, FP ou FN. Caso mais de uma caixa delimitadora prevista esteja sobre uma área de pólipo (*ground truth*), somente uma ocorrência de TP será contabilizada (i.e. aquela com maior valor de IoU). As restantes são consideradas como FP mesmo com IoU acima do limite. Assim estes valores serão estabelecidos conforme a seguir:

- TP: Detecção correta (i.e. maior IoU acima do limite);
- FP: Detecção incorreta (i.e. IoU abaixo do limite e qualquer IoU inferior ao maior IoU acima do limite, se houver);
- FN: Nenhuma detecção foi prevista para uma imagem com pólipo.

A arquitetura utilizada para detecção de pólipos nestes experimentos está baseada na arquitetura *Faster* R-CNN. O valor *non-maximum suppression* (NMS) adotado para o treinamento da rede foi de 0.7 conforme apresentado por Shin et al. [125]. O mesmo valor foi escolhido para o teste. O valor limite de IoU é de 0.2, conforme aplicado em [151]. Na etapa de treinamento, foi utilizada a descida do gradiente estocástico (SGD). O otimizador Adam [34] com taxa de aprendizado de 0.00001 (1e-5) foi escolhido para a rede de proposição de regiões (RPN). Como mencionado em 5.3.1, foi aplicado o aumento de dados tradicional (*augmentations*) que compreende as transformações de rotação (40 graus), *zoom*, deslocamento horizontal e vertical (20%) e espelhamento horizontal. Foram executadas 15 épocas no processo de treinamento com tempo total de aproximadamente 32 horas para cada conjunto de dados de treinamento com quantidade de imagens entre 6362 e 7666 (Tabela 5.2). Os experimentos foram realizados em um computador PC *desktop* com sistema operacional *Windows* 10, com processador Intel (R) Core (TM) i7-7700HQ 2.80GHz, 16GB de memória RAM e placa gráfica NVidia GeForce GTX 1060 com 6GB de memória dedicada.

5.4 Resultados

Nesta seção são apresentados os efeitos de cada conjunto de treinamento proposto B, C, D e E (ou seja, com pólipos inseridos) em comparação com o conjunto de treinamento A (i.e., CVC-ClinicDB original sem pólipos inseridos). Da mesma forma, o desempenho dos conjuntos de dados aprimorados também foi comparado com outros estudos na literatura. Para a avaliação de desempenho da rede foram empregados os conjuntos de testes A (ETIS-LaribPolypDB) e B (somente imagens com pólipos do CVC-VideoClinicDB). Os resultados estão listados nas Tabelas 5.3 e 5.4, respectivamente.

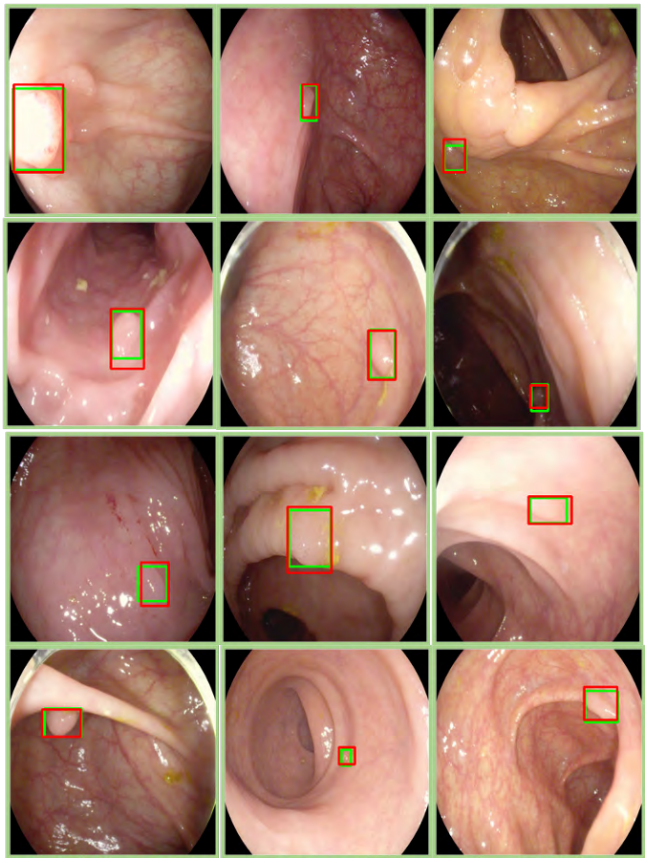

Figura 5.4: Conjunto de testes ETIS-LaribPolypDB: exemplo de imagens do melhor resultado de detecção da rede treinada com o conjunto D (Tabela 5.3). Caixa delimitadora em vermelho: previsão da rede. Caixa delimitadora em verde: localização real do pólipo (*ground truth*).

A avaliação destes experimentos está direcionada para minimizar o número de falsos negativos (FN), enquanto o número de verdadeiros positivos (TP) permanece alto. Considerando o contexto médico, é interessante encontrar todos os pólipos (máxima revocação) no conjunto de dados de teste [140], pois, se o sistema não detectar um pólipo esta lesão pode se tornar um câncer futuramente. Neste caso específico é aceitável uma baixa precisão às custas da maximização da revocação.

A tabela 5.3 mostra os resultados da avaliação de desempenho para o teste realizado no conjunto de dados ETIS-LaribPolypDB (conjunto de testes A). O desempenho de todos os conjuntos de treinamento propostos (B, C, D e E) em termos de TP e FN foram mais favoráveis que os do conjunto de treinamento A (CVC-ClinicDB original). Em outras palavras, os conjuntos de dados aprimorados pela inserção de pólipos promoveram uma melhoria significativa na métrica de revocação. Os resultados na Tabela 5.3 mostram que o conjunto de treinamento D foi superior aos métodos nos estudos [151] e [14] em termos de revocação. No primeiro experimento, a comparação

considera os resultados de CVC-Aug descritos por Zheng et al. [151]. O CVC-Aug é um conjunto de dados empregado no treinamento da rede neural convolucional *You-Only-Look-Once* (YOLO) [111], composto pelas imagens dos conjuntos CVC-ClinicDB e CVC-ColonDB com aumento de dados tradicional (*augmentations*). No segundo experimento o desempenho dos conjuntos de treinamento aprimorados são comparados com os resultados de detecção de pólipos apresentados pela abordagem CNN CUMED no estudo de Bernal et al. [14]. Neste caso, o conjunto de treinamento D alcançou o maior valor de TP e o menor FN em comparação com os dois métodos [151] e [14]. A Figura 5.4 exibe os resultados de detecção da rede *Faster* R-CNN treinada com o conjunto de treinamento D aprimorado. A previsão da rede está representada pela caixa delimitadora em vermelho e a localização real do pólipo está representada em verde.

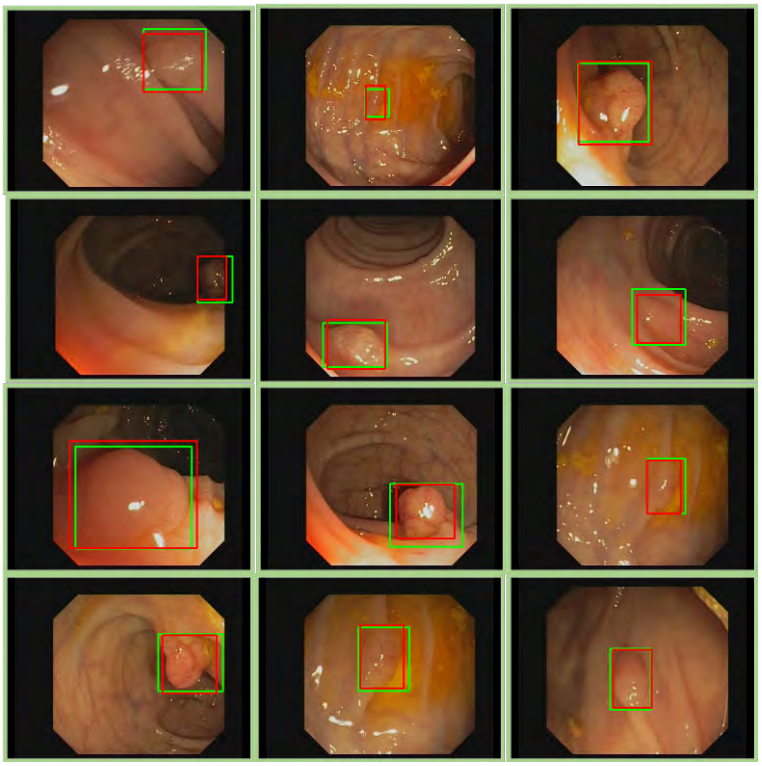

Figura 5.5: Conjunto de testes CVC-VideoClinicDB: exemplo de imagens do melhor resultado de detecção da rede treinada com o conjunto D (Tabela 5.4). Caixa delimitadora em vermelho: previsão da rede. Caixa delimitadora em verde: localização real do pólipo (*ground truth*).

Os resultados listados na Tabela 5.4 mostram a comparação do conjunto de treinamento original A com os conjuntos de treinamento propostos (B, C, D e E) no contexto do conjunto de teste B (CVC-VideoClinicDB). Os conjuntos de treinamento aprimorados apresentaram evolução em termos de TP e FN favorecendo melhores valores de revocação. Na comparação com outro estudo,

Tabela 5.3: Comparação dos resultados da detecção de pólipos sobre o conjunto de testes ETIS-LaribPolypDB.

Nome	**Total Imagens**	**TP**	**FP**	**FN**	**Prec.**	**Rec.**
Zheng et al. [151] CVC-Aug	3968	154	**45**	54	**0.774**	0.740
Bernal et al. [14] CUMED	n/a	144	55	64	0.723	0.692
Conj. Trein. A	7356	110	191	98	0.365	0.528
Conj. Trein. B	7666	134	179	74	0.428	0.644
Conj. Trein. C	7162	144	246	64	0.369	0.692
Conj. Trein. D	6793	**162**	164	**46**	0.496	**0.778**
Conj. Trein. E	6362	133	101	75	0.568	0.639

Tabela 5.4: Comparação dos resultados da detecção de pólipos sobre o conjunto de testes CVC-VideoClinicDB.

Nome	**Total Imagens**	**TP**	**FP**	**FN**	**Prec.**	**Rec.**
Shin et al. [126] Aug-I	5904	7517	1995	2508	0.790	0.750
Shin et al. [126] Aug-II	11808	6831	**1177**	3194	**0.853**	0.681
Conj. Trein. A	7356	7156	4625	2869	0.607	0.713
Conj. Trein. B	7666	7504	5388	2521	0.582	0.748
Conj. Trein. C	7162	7378	6313	2647	0.538	0.735
Conj. Trein. D	6793	**7954**	5904	**2071**	0.573	**0.793**
Conj. Trein. E	6362	7352	3774	2673	0.660	0.733

o conjunto de treinamento D apresenta um desempenho superior de detecção em relação aos valores TP e FN do que nos dois conjuntos de dados Aug-I e Aug-II. Shin et al. [126] descreve estes dois conjuntos que foram utilizados para treinar uma rede *Faster* R-CNN. Considerando a métrica de revocação houve uma melhoria no experimento que utiliza o conjunto de treinamento D em relação a ambos os conjuntos do estudo [126], mesmo usando uma quantidade menor de imagens de treinamento do que Aug-II (6793 amostras contra 11808).

Os resultados sugerem que a estratégia de aprimoramento dos conjuntos de imagens proposta nesta tese é promissora e superior, em termos da métrica de revocação do que aqueles que usam apenas técnicas tradicionais de aumento de dados, como o conjunto de treinamento A, bem como os outros métodos [14, 126, 151] comparados nestes experimentos.

A Figura 5.6 apresenta exemplos dos pólipos sintéticos. A rede generativa condicional foi capaz de produzir imagens com aparência bem próxima da realidade observada no conjunto de dados CVC-ClinicDB, que forneceu os pólipos para o treinamento. A rede conseguiu apresentar as nuances das texturas, reflexões da luz e áreas de sombra próximas as bordas dos pólipos.

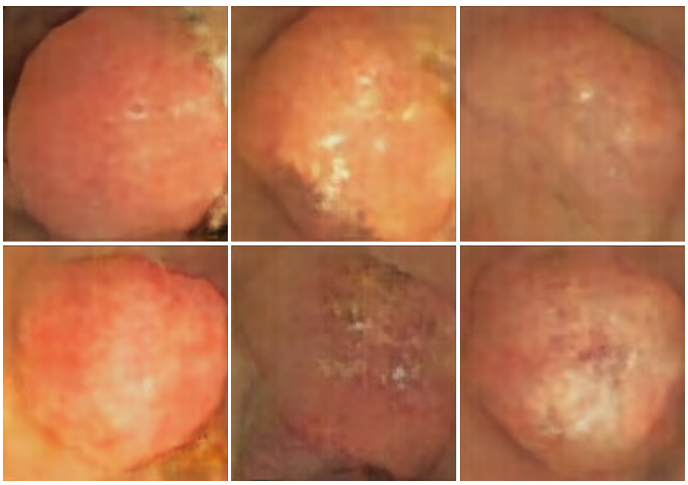

Figura 5.6: Detalhes dos pólipos sintéticos gerados pela rede GAN condicional.

5.4.1 Discussão

A estratégia de inserção de pólipos apresentada nesta tese foi empregada para atenuar a limitada variedade de amostras de pólipos presentes em conjuntos de imagens de colonoscopia disponíveis publicamente. Nesse contexto, foi adotada uma quantidade relativamente próxima de amostras em todos os conjuntos de dados de treinamento propostos para uma comparação justa. A menor quantidade de amostras foi utilizada no conjunto de dados E (6362), enquanto a maior foi empregada no conjunto de dados B (7666).

O conjunto de dados CVC-ClinicDB usado no procedimento de inserção de pólipos possui pelo menos um pólipo por imagem, porém, mais pólipos foram adicionados a estas imagens. No entanto, a maioria das amostras no conjunto de testes A (ETIS-LaribPolypDB) e B (CVC-VideoClinicDB) possui apenas um pólipo. Este pode ser o motivo dos valores de FP mais elevados nos resultados. As imagens de treinamento com dois ou mais pólipos tendem a influenciar a rede no sentido de criar mais *bounding boxes*, que podem estar sobre a área real do pólipo *ground truth* mas somente uma será contabilizada como TP (aquela com valor de IoU mais alto), as restantes serão consideradas FP. O estudo de Qadir et al. [109] apresenta este problema relacionado a precisão baixa quando a revocação é alta devido a mais *bounding boxes* geradas, aumentando os números de FP. Estes autores minimizaram este problema implementando uma etapa de pós-processamento, porém esta etapa não foi aplicada na metodologia da abordagem desta tese.

Considerando a estratégia de inserção, o tamanho dos pólipos adicionados pode ser relevante para a fase de treinamento. Isso pode ocorrer porque um pólipo maior causará uma variação mais significativa na aparência geral da imagem e terá um impacto positivo no treinamento. No entanto, este procedimento precisa considerar o tamanho da região que receberá o pólipo sintético para manter a imagem de destino visualmente coerente.

Algumas das imagens dos pólipos sintéticos criadas pela rede GAN condicional apresentaram falhas conforme apresentado na Figura 5.7. Exemplos destes artefatos podem ser vistos em (a) nas imagens da esquerda e direita formando um padrão em linhas horizontais. Áreas com *pixels* em amarelo e vermelho (b) estão presentes nas três imagens. A imagem central apresentou outras imperfeições como em (c) e (d).

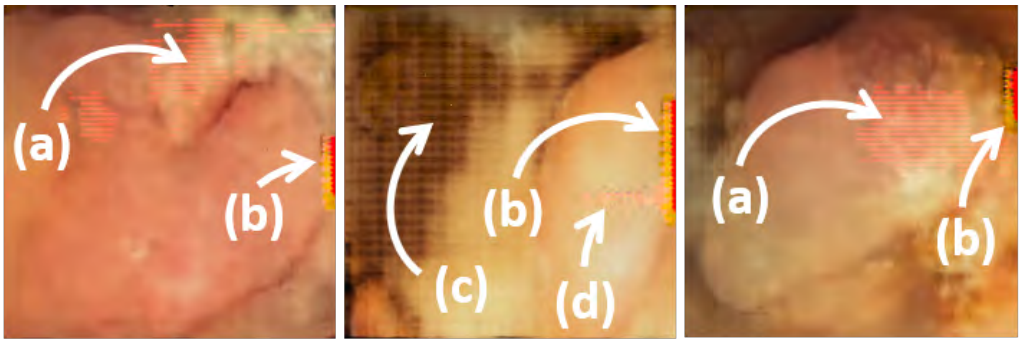

Figura 5.7: Falhas apresentadas por algumas imagens de pólipos sintéticos. (a), (b), (c) e (d): regiões com distorções na aparência.

A rede condicional recebeu como entrada somente as áreas dos pólipos selecionadas das imagens originais. O tamanho da área do pólipo é definido

pela caixa delimitadora *bounding box*. É provável que a seleção de uma área maior ao redor do pólipo (i.e., caixa delimitadora maior) evite as imperfeições nas regiões de borda do pólipo que ao mesmo tempo são bordas da imagem, conforme visto na Figura 5.7(b).

A avaliação da qualidade das imagens sintéticas criadas por redes do tipo GAN necessita de uma inspeção visual a cada etapa. Isto, significa que não há uma quantidade pré-definida de execuções do treinamento que produza a melhor imagem. A realização de treinamentos com uma quantidade de épocas determinada empiricamente pode resultar em imagens sem as imperfeições demonstradas na Figura 5.7.

A geração de uma imagem sintética por meio das redes GAN considerando somente a área do pólipo, ao invés da imagem de colonoscopia completa, apresenta alguns benefícios. O primeiro é a possibilidade de aproveitar uma imagem real por inteiro para receber uma variedade de pólipos disponíveis. Uma imagem de colonoscopia real pode apresentar uma qualidade superior de detalhes, que deve ser de difícil de reproduzir em alto nível por uma rede GAN. Outra vantagem é a combinação direcionada entre pólipos com certas características e imagens onde o ponto de vista está mais próximo ou afastado da mucosa. Outro ponto é que o vídeo de colonoscopia apresenta a maioria dos quadros sem a presença de lesões. Isto possibilita a inserção de pólipos sintéticos, por exemplo, nestas imagens negativas (sem pólipos). Desta forma, a criação da imagem *ground truth* por especialista médico não é necessária, visto que a estratégia de aprimoramento descrita nesta tese cria a respectiva imagem *ground truth* automaticamente. Esta característica é muito útil pois a disponibilidade de conjuntos de dados de imagens de colonoscopia em quantidade e variedade de imagens com as respectivas imagens *ground truth* é limitada.

6
Conclusões e Trabalhos Futuros

Este capítulo apresenta as conclusões, os principais resultados obtidos nos experimentos com as redes de segmentação e detecção e as contribuições. Em seguida são citadas a publicação e submissão de artigos referentes a pesquisa desta tese. Finalmente, são descritos brevemente os trabalhos futuros.

6.1
Conclusões

Nesta tese foi apresentada uma estratégia para o aprimoramento de conjuntos de dados para treinamento de redes neurais. Os dados utilizados consistem em imagens de colonoscopia presentes em conjuntos disponíveis publicamente. No contexto médico, é difícil obter grandes conjuntos de imagens de treinamento devido ao alto custo de mão de obra qualificada e restrições de privacidade, por exemplo. Existem conjuntos de imagens públicos, como o CVC-ClinicDB, porém a variedade de pólipos e a quantidade de imagens podem ser insuficientes para o treinamento bem sucedido de uma rede neural profunda.

A partir desta observação foi proposta uma abordagem que insere pólipos reais e sintéticos em imagens de colonoscopia pertencentes a conjuntos públicos. Inicialmente foi elaborada uma estratégia que seleciona os pólipos reais automaticamente. Tais pólipos precisam ser inseridos em locais coerentes nas imagens destino, por este motivo, foi implementada uma metodologia para seleção de regiões empregando o algoritmo de *Watershed*. A fusão do pólipo com o restante da imagem destino foi ajustada pela técnica de *Poisson* para estabelecer a coerência visual. Para validação destas imagens aprimoradas foram criados conjuntos aplicados ao treinamento de uma rede de segmentação U-net. Outra versão do aprimoramento de imagens foi desenvolvida empregando pólipos sintéticos. Para isso, foi criado um procedimento baseado nas Redes Adversárias Generativas. A rede GAN gerou um conjunto de novas máscaras binárias que foram utilizadas pela rede GAN condicional como referência de formatos sobre os quais texturas variadas foram produzidas formando uma coleção de pólipos sintéticos. As regiões adequadas selecionadas pela segmentação *Watershed* receberam os pólipos sintéticos e foram ajustadas para manter

a coerência visual, formando novas amostras com pólipos sintéticos. Conjuntos de imagens com pólipos reais e sintéticos foram validados no treinamento de uma rede *Faster* R-CNN.

Nos experimentos com a rede de segmentação U-net o conjunto de treinamento D (com 3823 imagens) aprimorado apresentou os melhores resultados em relação aos conjuntos aumentados com técnicas tradicionais. Sobre o conjunto de teste B (ETIS-LaribPolypDB) foram obtidos os valores de precisão 0.936, revocação 0.898, F1 0.914 e taxa de falso positivo (FPR) 0.079. Houve progresso em todas as métricas exceto em precisão, no entanto, com valor muito próximo ao obtido pelo conjunto B aumentado tradicionalmente (com 3825 imagens). A taxa de falso positivo (FPR) foi reduzida significativamente. Os valores de FPR indicam que a rede treinada com os conjuntos de dados com pólipos reais inseridos aprendeu a diferenciar melhor estas lesões de outras estruturas visualmente semelhantes.

A capacidade de localização da rede *Faster* R-CNN foi a avaliada a partir de conjuntos de treinamento aprimorados que incluem pólipos sintéticos e reais. Os testes sobre o conjunto de teste ETIS-LaribPolypDB apresentaram valores superiores de revocação em relação aos conjuntos que fizeram uso de técnicas tradicionais de aumento de dados, além de outros dois estudos (Zheng et al. [151] CVC-Aug, Bernal et al. [14] CUMED). A estratégia de inserção de pólipos sintéticos nas imagens favorece melhores valores de revocação, segundo os testes. O conjunto de treinamento D apresentou o maior valor com revocação de 0.778. Este mesmo conjunto de treinamento D, avaliado sobre o conjunto de teste CVC-VideoClinicDB alcançou revocação de 0.793. Este valor superou duas abordagens (Aug-I e Aug-II) do estudo de Shin et al. [126] em termos de revocação. Maiores valores de revocação demonstram que a rede foi superior em evitar falsos negativos (FN), que é uma característica desejável no contexto médico, pois lesões não encontradas podem se tornar em tumores malignos.

O estudo descrito nesta tese contribui para melhorar resultados de localização de pólipos em aplicações que utilizam aprendizado de máquina com o desenvolvimento de uma estratégia de aprimoramento de dados capaz de aumentar a variação de amostras por meio da inserção de pólipos reais e sintéticos em imagens de colonoscopia. Com a utilização das redes GAN e CGAN este estudo apresentou uma abordagem de geração de pólipos sintéticos e suas respectivas máscaras binárias a partir de um conjunto com reduzida diversificação de pólipos reais. Tal procedimento mostrou a capacidade de criação de pólipos sintéticos com formato e textura variados alcançando detalhes de reflexão da luz, cores e sombras semelhantes às lesões reais. Foi demonstrado que a técnica *Poisson* é suficiente para apresentar resultados satisfatórios em um domínio de

imagens complexas como colonoscopia, sendo útil para o aperfeiçoamento do aspecto visual em casos de mesclagem de imagens para treinamento de redes neurais. Além disso, abordagens de segmentação e detecção implementadas demonstram a viabilidade da construção de soluções de localização de pólipos a partir do treinamento de conjuntos de dados aprimorados. A estratégia de inserção de pólipos proposta é promissora e pode ser empregada juntamente com técnicas tradicionais de aumento de dados, promovendo melhores resultados em conjuntos de dados reduzidos.

6.2 Publicação

A partir da pesquisa descrita nesta tese foi publicado o seguinte artigo [28]:

- DE ALMEIDA THOMAZ, V.; SIERRA-FRANCO, C. A. ; RAPOSO, A. B. **Training data enhancements for robust polyp segmentation in colonoscopy images.** In 2019 IEEE 32nd INTERNATIONAL SYMPOSIUM ON COMPUTER-BASED MEDICAL SYSTEMS (CBMS), p. 192–197, June 2019.

O conteúdo desta publicação é referente aos assuntos do Capítulo 4, sendo premiado como melhor artigo no simpósio internacional IEEE CBMS. Além disso, outro artigo referente ao desenvolvimento apresentado no Capítulo 5 foi submetido a um periódico internacional *Artificial Intelligence in Medicine*.

6.3 Trabalhos Futuros

Como trabalhos futuros propõe-se incluir melhorias na organização dos conjuntos de dados. Isto compreende a obtenção de imagens sem pólipos, que são em geral a maioria dos quadros em vídeos de colonoscopia, para inserção de pólipos sintéticos. Deste modo, pode-se aproveitar a disponibilidade maior das imagens negativas (sem pólipos) para formar novos conjuntos variados com pólipos sintéticos. Outro aspecto é a possibilidade de controle no balanceamento do conjunto em relação a quantidade de pólipos por imagem. O balanceamento neste sentido pode favorecer a redução de falsos positivos, principalmente no caso da rede *Faster* R-CNN. A inserção de pólipos pode ser útil para balancear melhor os conjuntos de dados em relação a certas características da lesão. Por exemplo, a geração sintética e inserção de pólipos pode contribuir para formação de conjuntos que possuam representações de casos raros.

Além disso, outros aspectos não considerados podem ser adicionados como a elaboração de uma estratégia para seleção de quadros do vídeo de colonoscopia que são mais distintos entre si. Os quadros consecutivos costumam apresentar imagens com poucas diferenças afetando negativamente o aprendizado da rede. Preferencialmente, devem ser empregadas imagens mais distintas para receberem os pólipos sintéticos variados.

Outro adicional seria um processo de avaliação do aspecto visual das amostras em que os pólipos são inseridos em relação as imagens reais. O objetivo é analisar o quanto uma nova amostra com certo grau de realismo contribui para o aprendizado da rede. Segundo Han et al. [54] imagens mais realistas nem sempre garantem melhores resultados.

A estratégia de inserção precisa ser ajustada pois em certos casos a localização dos pólipos inseridos não foi adequada. Foi observado que se o pólipo for posicionado sobre uma dobra das paredes do cólon o aspecto realista é prejudicado, mesmo após a aplicação da técnica *Poisson*. Uma possível solução é aplicar um detector de bordas para obter o destaque das dobras da parede do cólon e após implementar uma verificação que indique se a alguma parte da área do pólipo está sobre uma borda, considerando a localização indicada para inserção.

Também serão consideradas melhorias nas redes generativas (GAN e CGAN) para criação de pólipos mais realísticos. Isto inclui ajustes nos parâmetros de treinamento ou alterações na arquitetura da rede. Os resultados também podem ser refinados com o acerto da quantidade de iterações na etapa treinamento, geralmente obtido empiricamente, devido as características destas arquiteturas de redes generativas. É preciso encontrar o ponto no treinamento que produz a melhor imagem isto significa que, em certos momentos, mais iterações podem piorar os resultados.

Por fim, esta proposta de aprimoramento de dados pode eventualmente ser aplicada a outras doenças do trato digestivo como por exemplo colite, esofagite e úlceras. O mesmo pode ser dito para o caso de outro tipo de imagem médica, como ressonância magnética utilizada para identificação de tumores no cérebro, e.g., gliomas. Assim, a investigação desta abordagem de aprimoramento sobre outros domínios pode ser promissora como trabalho futuro.

Referências bibliográficas

[1] Agrahari, H., Iwahori, Y., K Bhuyan, M., Ghorai, S., Kohli, H., J Woodham, R., and Kasugai, K. Automatic polyp detection using dsc edge detector and hog features. In *Proceedings of the 3rd International Conference on Pattern Recognition Applications and Methods* (2014), SCITEPRESS-Science and Technology Publications, Lda, pp. 495–501.

[2] Alexandre, L. A., Nobre, N., and Casteleiro, J. Color and position versus texture features for endoscopic polyp detection. In *BioMedical Engineering and Informatics, 2008. BMEI 2008. International Conference on* (2008), vol. 2, IEEE, pp. 38–42.

[3] Amber, A., Iwahori, Y., Bhuyan, M., Woodham, R. J., and Kasugai, K. Feature point based polyp tracking in endoscopic videos. In *Applied Computing and Information Technology/2nd International Conference on Computational Science and Intelligence (ACIT-CSI), 2015 3rd International Conference on* (2015), IEEE, pp. 299–304.

[4] Angermann, Q., Bernal, J., Sánchez-Montes, C., Hammami, M., Fernández-Esparrach, G., Dray, X., Romain, O., Sánchez, F. J., and Histace, A. Towards real-time polyp detection in colonoscopy videos: Adapting still frame-based methodologies for video sequences analysis. In *Computer Assisted and Robotic Endoscopy and Clinical Image-Based Procedures* (Cham, 2017), M. J. Cardoso, T. Arbel, X. Luo, S. Wesarg, T. Reichl, M. Á. González Ballester, J. McLeod, K. Drechsler, T. Peters, M. Erdt, K. Mori, M. G. Linguraru, A. Uhl, C. Oyarzun Laura, and R. Shekhar, Eds., Springer International Publishing, pp. 29–41.

[5] Axon, A., Diebold, M., Fujino, M., Fujita, R., Genta, R., Gonvers, J., Guelrud, M., Inoue, H., Jung, M., Kashida, H., et al. Update on the paris classification of superficial neoplastic lesions in the digestive tract. *Endoscopy 37*, 6 (2005), 570–578.

[6] Bae, S.-H., and Yoon, K.-J. Polyp detection via imbalanced learning and discriminative feature learning. *IEEE transactions on medical imaging 34*, 11 (2015), 2379–2393.

[7] Bardhi, O., Sierra-Sosa, D., Garcia-Zapirain, B., and Elmaghraby, A. Automatic colon polyp detection using convolutional encoder-decoder model. In *2017 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)* (Dec 2017), pp. 445–448.

[8] Batchelor, B. G. *Machine Vision Handbook*. Springer, 2012.

[9] Benjdira, B., Khursheed, T., Koubaa, A., Ammar, A., and Ouni, K. Car detection using unmanned aerial vehicles: Comparison between faster r-cnn and yolov3. In *2019 1st International Conference on Unmanned Vehicle Systems-Oman (UVS)* (Feb 2019), pp. 1–6.

[10] Bernal, J., Sánchez, F. J., Fernández-Esparrach, G., Gil, D., Rodríguez, C., and Vilariño, F. Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. *Computerized Medical Imaging and Graphics 43* (2015), 99–111.

[11] Bernal, J., Sánchez, J., and Vilarino, F. Integration of valley orientation distribution for polyp region identification in colonoscopy. In *Abdominal Imaging. Computational and Clinical Applications. ABD-MICCAI* (2011), Springer, pp. 76–83.

[12] Bernal, J., Sánchez, J., and Vilarino, F. Towards automatic polyp detection with a polyp appearance model. *Pattern Recognition 45*, 9 (2012), 3166–3182.

[13] Bernal, J., Sánchez, J., and Vilariño, F. Impact of image preprocessing methods on polyp localization in colonoscopy frames. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (July 2013), pp. 7350–7354.

[14] Bernal, J., Tajkbaksh, N., Sanchez, F. J., Matuszewski, B. J., Chen, H., Yu, L., Angermann, Q., Romain, O., Rustad, B., Balasingham, I., Pogorelov, K., Choi, S., Debard, Q., Maier-Hein, L., Speidel, S., Stoyanov, D., Brandao, P., Cordova, H., Sanchez-Montes, C., Gurudu, S. R., Fernandez-Esparrach, G., Dray, X., Liang, J., and Histace, A. Comparative Validation of Polyp Detection Methods in Video Colonoscopy: Results From the MICCAI 2015 Endoscopic Vision Challenge. *IEEE Transactions on Medical Imaging 36*, 6 (jun 2017), 1231–1249.

[15] Bernal, J. J., Histace, A., Masana, M., Angermann, Q., Sánchez-Montes, C., Rodriguez, C., Hammami, M., Garcia-

Rodriguez, A., Córdova, H., Romain, O., Fernández-Esparrach, G., Dray, X., and Sanchez, J. Polyp Detection Benchmark in Colonoscopy Videos using GTCreator: A Novel Fully Configurable Tool for Easy and Fast Annotation of Image Databases. In *Proceedings of 32nd CARS conference* (Berlin, Germany, June 2018).

[16] Beucher, S. Use of watersheds in contour detection. In *Proceedings of the International Workshop on Image Processing* (1979), CCETT.

[17] Billah, M., Waheed, S., and Rahman, M. M. An automatic gastrointestinal polyp detection system in video endoscopy using fusion of color wavelet and convolutional neural network features. *International journal of biomedical imaging 2017* (2017).

[18] Bisong, E. Google colaboratory. In *Building Machine Learning and Deep Learning Models on Google Cloud Platform*. Springer, 2019, pp. 59–64.

[19] Bovik, A. C. *The essential guide to image processing*. Academic Press, 2009.

[20] Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians 68*, 6 (2018), 394–424.

[21] Canny, J. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence* (1986), 679–698.

[22] Chen, H.-m., Varshney, P. K., and Slamani, M.-A. On registration of regions of interest (roi) in video sequences. In *Advanced Video and Signal Based Surveillance, 2003. Proceedings. IEEE Conference on* (2003), IEEE, pp. 313–318.

[23] Cheng, D.-C., Ting, W.-C., Chen, Y.-F., Pu, Q., and Jiang, X. Colorectal polyps detection using texture features and support vector machine. In *International Conference on Mass Data Analysis of Images and Signals in Medicine, Biotechnology, and Chemistry* (2008), Springer, pp. 62–72.

[24] Costa, P., Galdran, A., Meyer, M., Abràmoff, M., Niemejer, M., Mendonca, A., and Campilho, A. Towards adversarial retinal image synthesis. *arxiv* (2017).

[25] da Silva, E. A., and Mendonça, G. V. Digital Image Processing. In *The Electrical Engineering Handbook*. Elsevier, 2005, pp. 891–910.

[26] Danelljan, M., Hager, G., Shahbaz Khan, F., and Felsberg, M. Learning spatially regularized correlation filters for visual tracking. In *Proceedings of the IEEE international conference on computer vision* (2015), pp. 4310–4318.

[27] de Albuquerque, M. P., and de Albuquerque, M. P. Processamento de imagens: métodos e análises. *Centro Brasileiro de Pesquisas Físicas - CBPF. Rio de Janeiro, Brasil 12* (2000).

[28] de Almeida Thomaz, V., Sierra-Franco, C. A., and Raposo, A. B. Training data enhancements for robust polyp segmentation in colonoscopy images. In *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)* (June 2019), pp. 192–197.

[29] de Bruijne, M. Machine learning approaches in medical image analysis: From detection to diagnosis. *Medical Image Analysis 33* (2016), 94 – 97. 20th anniversary of the Medical Image Analysis journal (MedIA).

[30] De Lange, T., Halvorsen, P., and Riegler, M. Methodology to develop machine learning algorithms to improve performance in gastrointestinal endoscopy. *World journal of gastroenterology 24*, 45 (2018), 5057.

[31] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09* (2009).

[32] Dhandra, B. V., Hegadi, R., Hangarge, M., and Malemath, V. S. Analysis of abnormality in endoscopic images using combined hsi color space and watershed segmentation. In *18th International Conference on Pattern Recognition (ICPR'06)* (2006), vol. 4, IEEE, pp. 695–698.

[33] Díaz, J. A. Morphological operations in binary images. Wolfram Demonstrations Project., Dec. 2011.

[34] Diederik, P. K., and Ba, J. L. Adam: A method for stochastic optimization. In *Proc. Int. Conf. Learn. Represent. (ICLR)* (2015).

[35] Drozdzal, M., Vorontsov, E., Chartrand, G., Kadoury, S., and Pal, C. The importance of skip connections in biomedical image segmentation. In *Deep Learning and Data Labeling for Medical Applications*. Springer, 2016, pp. 179–187.

[36] Esquef, I. A., ALBUQUERQUE, M. P. d., and ALBUQUERQUE, M. P. d. Processamento digital de imagens. *Rio de Janeiro 12* (2003).

[37] Everingham, M., Van Gool, L., Williams, C. K., Winn, J., and Zisserman, A. The pascal visual object classes (voc) challenge. *International journal of computer vision 88*, 2 (2010), 303–338.

[38] Fan, Q., Brown, L., and Smith, J. A closer look at faster r-cnn for vehicle detection. In *2016 IEEE Intelligent Vehicles Symposium (IV)* (June 2016), pp. 124–129.

[39] Fausett, L. *Fundamentals of neural networks: architectures, algorithms, and applications*. Prentice-Hall, Inc., 1994.

[40] Franco, C. A. S. *Técnicas de aprendizagem para gerência de recursos em redes móveis heterogêneas e auto-organizáveis*. Tese de doutorado, Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), 2018.

[41] Fu, K. S., and Mui, J. A survey on image segmentation. *Pattern recognition 13*, 1 (1981), 3–16.

[42] Gauthier, J. Conditional generative adversarial nets for convolutional face generation. *Class Project for Stanford CS231N: Convolutional Neural Networks for Visual Recognition, Winter semester 2014*, 5 (2014), 2.

[43] Ghanian, Z., Pezeshk, A., Petrick, N., and Sahiner, B. Computational insertion of microcalcification clusters on mammograms: reader differentiation from native clusters and computer-aided detection comparison. *Journal of medical imaging (Bellingham, Wash.) 5*, 4 (October 2018), 044502.

[44] Giger, M. L. Machine learning in medical imaging. *Journal of the American College of Radiology 15*, 3 (2018), 512–520.

[45] Gil, D., Sánchez, F. J., Fernández-Esparrach, G., and Bernal, J. 3d stable spatio-temporal polyp localization in colonoscopy videos. In *Computer-Assisted and Robotic Endoscopy* (2015), Springer, pp. 140–152.

[46] Gonzalez, R. C., and Woods, R. E. Digital image processing, 2002.

[47] Goodfellow, I., Bengio, Y., and Courville, A. *Deep learning*. MIT press, 2016.

[48] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. Generative adversarial nets. In *Advances in neural information processing systems* (2014), pp. 2672–2680.

[49] Group, R. S. W., et al. Machine learning: the power and promise of computers that learn by example. Tech. rep., Technical report, 2017.

[50] Guibas, J. T., Virdi, T. S., and Li, P. S. Synthetic medical images from dual generative adversarial networks. *arXiv preprint arXiv:1709.01872* (2017).

[51] Guinigundo, A. Is the Virtual Colonoscopy a Replacement for Optical Colonoscopy? *Seminars in Oncology Nursing 34*, 2 (may 2018), 132–136.

[52] Gupta, R., Elamvazuthi, I., Dass, S. C., Faye, I., Vasant, P., George, J., and Izza, F. Curvelet based automatic segmentation of supraspinatus tendon from ultrasound image: a focused assistive diagnostic method. *Biomedical engineering online 13*, 1 (2014), 157.

[53] Guévelou, S. *Characterization of the thermal-radiative properties of foams with numerically controlled structures: towards the design of solar absorbers*. PhD thesis, Université de Nantes, 12 2015.

[54] Han, C., Hayashi, H., Rundo, L., Araki, R., Shimoda, W., Muramatsu, S., Furukawa, Y., Mauri, G., and Nakayama, H. Gan-based synthetic brain mr image generation. In *Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on* (2018), IEEE, pp. 734–738.

[55] Hanley, J. A., and McNeil, B. J. The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology 143*, 1 (1982), 29–36.

[56] Haralick, R. M., and Shapiro, L. G. Image segmentation techniques. *Computer vision, graphics, and image processing 29*, 1 (1985), 100–132.

[57] Haralick, R. M., Sternberg, S. R., and Zhuang, X. Image analysis using mathematical morphology. *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-9*, 4 (July 1987), 532–550.

[58] He, K., Gkioxari, G., Dollár, P., and Girshick, R. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (2017), pp. 2961–2969.

[59] He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016), pp. 770–778.

[60] Hong, D., Tavanapong, W., Wong, J., Oh, J., and De Groen, P. C. 3d reconstruction of virtual colon structures from colonoscopy images. *Computerized Medical Imaging and Graphics 38*, 1 (2014), 22–33.

[61] Hong, Y., Hwang, U., Yoo, J., and Yoon, S. How generative adversarial networks and their variants work: An overview. *ACM Computing Surveys (CSUR) 52*, 1 (2019), 1–43.

[62] Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Fathi, A., Fischer, I., Wojna, Z., Song, Y., Guadarrama, S., et al. Speed/accuracy trade-offs for modern convolutional object detectors. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017), pp. 7310–7311.

[63] Hwang, S., Oh, J., Tavanapong, W., Wong, J., and De Groen, P. C. Polyp detection in colonoscopy video using elliptical shape feature. In *Image Processing, 2007. ICIP 2007. IEEE International Conference on* (2007), vol. 2, IEEE, pp. II–465.

[64] Iakovidis, D. K., Maroulis, D. E., and Karkanis, S. A. An intelligent system for automatic detection of gastrointestinal adenomas in video endoscopy. *Computers in Biology and Medicine 36*, 10 (2006), 1084–1103.

[65] Ibtehaz, N., and Rahman, M. S. Multiresunet: Rethinking the u-net architecture for multimodal biomedical image segmentation. *Neural Networks 121* (2020), 74–87.

[66] Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA). Estimativa 2020: incidência de câncer no brasil.

[67] Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. Image-to-image translation with conditional adversarial networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (July 2017), pp. 5967–5976.

[68] Iwahori, Y., Shinohara, T., Hattori, A., Woodham, R. J., Fukui, S., Bhuyan, M. K., and Kasugai, K. Automatic polyp detection in endoscope images using a hessian filter. In *MVA* (2013), pp. 21–24.

[69] Jaccard, P. Nouvelles recherches sur la distribution florale. *Bull. Soc. Vaud. Sci. Nat. 44* (1908), 223–270.

[70] Jiang, H., and Learned-Miller, E. Face detection with the faster r-cnn. In *2017 12th IEEE International Conference on Automatic Face Gesture Recognition (FG 2017)* (May 2017), pp. 650–657.

[71] Kang, J., and Doraiswami, R. Real-time image processing system for endoscopic applications. In *Electrical and Computer Engineering, 2003. IEEE CCECE 2003. Canadian Conference on* (2003), vol. 3, IEEE, pp. 1469–1472.

[72] Kang, J., and Gwak, J. Ensemble of instance segmentation models for polyp segmentation in colonoscopy images. *IEEE Access 7* (2019), 26440–26447.

[73] Karkanis, S. A., Iakovidis, D. K., Karras, D., and Maroulis, D. Detection of lesions in endoscopic video using textural descriptors on wavelet domain supported by artificial neural network architectures. In *Image Processing, 2001. Proceedings. 2001 International Conference on* (2001), vol. 2, IEEE, pp. 833–836.

[74] Karkanis, S. A., Iakovidis, D. K., Maroulis, D. E., Karras, D. A., and Tzivras, M. Computer-aided tumor detection in endoscopic video using color wavelet features. *IEEE transactions on information technology in biomedicine 7*, 3 (2003), 141–152.

[75] Karkanis, S. A., Iakovidis, D. K., Maroulis, D. E., Magoulas, G. D., and Theofanous, N. Tumor recognition in endoscopic video images using artificial neural network architectures. In *Euromicro Conference, 2000. Proceedings of the 26th* (2000), vol. 2, IEEE, pp. 423–429.

[76] Kotsiantis, S. B. Supervised machine learning: A review of classification techniques. *Informatica 31* (2007), 249–268.

[77] Kotsiantis, S. B., Zaharakis, I. D., and Pintelas, P. E. Machine learning: a review of classification and combining techniques. *Artificial Intelligence Review 26*, 3 (2006), 159–190.

[78] Koulaouzidis, A., Iakovidis, D. K., Yung, D. E., Rondonotti, E., Kopylov, U., Plevris, J. N., Toth, E., Eliakim, A., Johansson, G. W., Marlicz, W., et al. Kid project: an internet-based digital video atlas of capsule endoscopy for research purposes. *Endoscopy international open 5*, 6 (2017), E477–E483.

[79] Krishnan, S., Yang, X., Chan, K., Kumar, S., and Goh, P. Intestinal abnormality detection from endoscopic images. In *Engineering in Medicine and Biology Society, 1998. Proceedings of the 20th Annual International Conference of the IEEE* (1998), vol. 2, IEEE, pp. 895–898.

[80] Krizhevsky, A., Sutskever, I., and Hinton, G. E. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (2012), pp. 1097–1105.

[81] LeCun, Y. Generalization and network design strategies. *Connectionism in perspective 19* (1989), 143–155.

[82] Leufkens, A., Van Oijen, M., Vleggaar, F., and Siersema, P. Factors influencing the miss rate of polyps in a back-to-back colonoscopy study. *Endoscopy 44*, 05 (2012), 470–475.

[83] Li, P., Chan, K. L., Krishnan, S. M., and Gao, Y. Detecting abnormal regions in colonoscopic images by patch-based classifier ensemble. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on* (2004), vol. 3, IEEE, pp. 774–777.

[84] Liedlgruber, M., and Uhl, A. Endoscopic image processing-an overview. In *2009 Proceedings of 6th International Symposium on Image and Signal Processing and Analysis* (2009), IEEE, pp. 707–712.

[85] Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. Microsoft coco: Common objects in context. In *European conference on computer vision* (2014), Springer, pp. 740–755.

[86] López, A. M., Lumbreras, F., Serrat, J., and Villanueva, J. J. Evaluation of methods for ridge and valley detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence 21*, 4 (1999), 327–335.

[87] Lucic, M., Kurach, K., Michalski, M., Gelly, S., and Bousquet, O. Are gans created equal? a large-scale study. In *Advances in neural information processing systems* (2018), pp. 700–709.

[88] Ma, C., Yu, L., Chen, B., Koo, C. W., Takahashi, E. A., Fletcher, J. G., Levin, D. L., Kuzo, R. S., Viers, L. D., Vincent-Sheldon, S. A., Leng, S., and McCollough, C. H. Evaluation of a projection-domain lung nodule insertion technique in thoracic computed tomography. *Journal of medical imaging (Bellingham, Wash.) 4*, 1 (January 2017), 013510.

[89] Ma, J., Fan, X., Yang, S. X., Zhang, X., and Zhu, X. Contrast limited adaptive histogram equalization based fusion for underwater image enhancement. *Preprints* (2017), 127.

[90] Ma, Y., Li, Y., Yao, J., Chen, B., Deng, J., and Yang, X. Polyp location in colonoscopy based on deep learning. In *2019 8th International Symposium on Next Generation Electronics (ISNE)* (2019), IEEE, pp. 1–3.

[91] Maeda, R., and Maruyama, T. An implementation method of poisson image editing on fpga. In *2017 27th International Conference on Field Programmable Logic and Applications (FPL)* (2017), IEEE, pp. 1–6.

[92] Maroulis, D. E., Iakovidis, D. K., Karkanis, S. A., and Karras, D. A. Cold: a versatile detection system for colorectal lesions in endoscopy video-frames. *Computer Methods and Programs in Biomedicine 70*, 2 (2003), 151–166.

[93] Meijster, A. *Efficient sequential and parallel algorithms for morphological image processing*. Phd thesis, University of Groningen, 2004.

[94] Meyer, F. Topographic distance and watershed lines. *Signal processing 38*, 1 (1994), 113–125.

[95] Mirza, M., and Osindero, S. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784* (2014).

[96] Mitchell, T. M. Machine learning, mcgraw-hill higher education. *New York* (1997).

[97] Nadeem, S., and Kaufman, A. Computer-aided detection of polyps in optical colonoscopy images. In *Medical Imaging 2016: Computer-Aided Diagnosis* (2016), vol. 9785, International Society for Optics and Photonics, p. 978525.

[98] Otsu, N. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics 9*, 1 (1979), 62–66.

[99] Park, S. Y., Sargent, D., Spofford, I., Vosburgh, K. G., and A-Rahim, Y. A colon video analysis framework for polyp detection. *IEEE Transactions on Biomedical Engineering 59*, 5 (2012), 1408.

[100] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. Scikit-learn: Machine learning in python. *Journal of machine learning research 12*, Oct (2011), 2825–2830.

[101] Pérez, P., Gangnet, M., and Blake, A. Poisson image editing. *ACM Transactions on graphics (TOG) 22*, 3 (2003), 313–318.

[102] Pezeshk, A., Petrick, N., Chen, W., and Sahiner, B. Seamless lesion insertion for data augmentation in cad training. *IEEE Transactions on Medical Imaging 36*, 4 (2017), 1005–1015.

[103] Physician Data Query (PDQ) Adult Treatment Editorial Board. Colon cancer treatment - patient version. Bethesda, MD: National Cancer Institute., Jan. 2020. Disponível em: `https://www.cancer.gov/types/colorectal/patient/colon-treatment-pdq`. Acesso em: 06 jan. 2020.

[104] Pizer, S. M., Johnston, R. E., Rogers, D., and Beard, D. Effective presentation of medical images on an electronic display station. *Radiographics 7*, 6 (1987), 1267–1274.

[105] Pogorelov, K., Ostroukhova, O., Jeppsson, M., Espeland, H., Griwodz, C., de Lange, T., Johansen, D., Riegler, M., and Halvorsen, P. Deep learning and hand-crafted feature based approaches for polyp detection in medical videos. In *2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS)* (June 2018), pp. 381–386.

[106] Pogorelov, K., Randel, K. R., Griwodz, C., Eskeland, S. L., de Lange, T., Johansen, D., Spampinato, C., Dang-Nguyen, D.-T., Lux, M., Schmidt, P. T., Riegler, M., and Halvorsen, P. Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection. In *Proceedings of the 8th ACM on Multimedia Systems Conference* (New York, NY, USA, 2017), MMSys'17, ACM, pp. 164–169.

[107] Prasath, V. Polyp detection and segmentation from video capsule endoscopy: A review. *Journal of Imaging 3*, 1 (2017), 1.

[108] Preim, B., and Botha, C. Chapter 4 - image analysis for medical visualization. In *Visual Computing for Medicine (Second Edition)*, B. Preim and C. Botha, Eds., second edition ed. Morgan Kaufmann, Boston, 2014, pp. 111 – 175.

[109] Qadir, H. A., Balasingham, I., Solhusvik, J., Bergsland, J., Aabakken, L., and Shin, Y. Improving automatic polyp detection using cnn by exploiting temporal dependency in colonoscopy video. *IEEE Journal of Biomedical and Health Informatics 24*, 1 (2020), 180–193.

[110] Qadir, H. A., Solhusvik, J., Bergsland, J., Aabakken, L., and Balasingham, I. A framework with a fully convolutional neural network for semi-automatic colon polyp annotation. *IEEE Access 7* (2019), 169537–169547.

[111] Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016), pp. 779–788.

[112] Ren, S., He, K., Girshick, R. B., and Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence 39* (2015), 1137–1149.

[113] Reza, A. M. Realization of the contrast limited adaptive histogram equalization (clahe) for real-time image enhancement. *Journal of VLSI signal processing systems for signal, image and video technology 38*, 1 (2004), 35–44.

[114] Ribeiro, E., Uhl, A., Wimmer, G., and Häfner, M. Exploring deep learning and transfer learning for colonic polyp classification. *Computational and mathematical methods in medicine 2016* (2016).

[115] Riegler, M., Pogorelov, K., Markussen, J., Lux, M., Stensland, H. K., de Lange, T., Griwodz, C., Halvorsen, P., Johansen, D., Schmidt, P. T., et al. Computer aided disease detection system for gastrointestinal examinations. In *Proceedings of the 7th International Conference on Multimedia Systems* (2016), ACM, p. 29.

[116] Robins, M., Kalpathy-Cramer, J., Obuchowski, N. A., Buckler, A., Athelogou, M., Jarecha, R., Petrick, N., Pezeshk, A., Sahiner, B., and Samei, E. Evaluation of simulated lesions as surrogates to clinical lesions for thoracic ct volumetry: The results of an international challenge. *Academic Radiology 26*, 7 (2019), e161 – e173.

[117] Robins, M., Solomon, J., Sahbaee, P., Sedlmair, M., Choudhury, K. R., Pezeshk, A., Sahiner, B., and Samei, E. Techniques for virtual lung nodule insertion: volumetric and morphometric comparison of projection-based and image-based methods for quantitative CT. *Physics in Medicine & Biology 62*, 18 (aug 2017), 7280–7299.

[118] Roerdink, J. B., and Meijster, A. The watershed transform: Definitions, algorithms and parallelization strategies. *Fundamenta informaticae 41*, 1, 2 (2001), 187–228.

[119] Ronneberger, O., Fischer, P., and Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (2015), vol. 9351 of *LNCS*, Springer, pp. 234–241.

[120] Rosenblatt, F. *The perceptron, a perceiving and recognizing automaton Project Para*. Cornell Aeronautical Laboratory, 1957.

[121] Sánchez, F. J., Bernal, J., Sánchez-Montes, C., de Miguel, C. R., and Fernández-Esparrach, G. Bright spot regions segmentation and classification for specular highlights detection in colonoscopy videos. *Machine Vision and Applications 28*, 8 (2017), 917–936.

[122] Serra, J. *Image analysis and mathematical morphology*. Academic Press, Inc., 1983.

[123] Shalev-Shwartz, S., and Ben-David, S. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, 2014.

[124] Shin, S. J., Kim, H., and Han, S.-T. Comparison of the performance evaluations in classification. *International Journal of Advanced Research in Computer and Communication Engineering 5*, 8 (2016), 441–4.

[125] Shin, Y., Qadir, H. A., Aabakken, L., Bergsland, J., and Balasingham, I. Automatic colon polyp detection using region based deep cnn and post learning approaches. *IEEE Access 6* (2018), 40950–40962.

[126] Shin, Y., Qadir, H. A., and Balasingham, I. Abnormal colon polyp image synthesis using conditional adversarial networks for improved detection performance. *IEEE Access 6* (2018), 56007–56017.

[127] Silva, J., Aymeric, H., Olivier, R., Xavier, D., and Bertrand, G. Towards embedded detection of polyps in wce images for early diagnosis of colorectal cancer. *International Journal of Computer Assisted Radiology and Surgery, Springer Verlag (Germany)* (2014), 283–293.

[128] Silva, J., Histace, A., Romain, O., Dray, X., and Granado, B. Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer. *International Journal of Computer Assisted Radiology and Surgery 9*, 2 (2014), 283–293.

[129] Soille, P. *Morphological image analysis: principles and applications*. Springer Science & Business Media, 2013.

[130] Son, J., Park, S. J., and Jung, K.-H. Retinal vessel segmentation in fundoscopic images with generative adversarial networks. *arXiv preprint arXiv:1706.09318* (2017).

[131] Sun, X., Wu, P., and Hoi, S. C. Face detection using deep learning: An improved faster rcnn approach. *Neurocomputing 299* (2018), 42 – 50.

[132] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2015), pp. 1–9.

[133] Taha, B., Dias, J., and Werghi, N. Convolutional neural networkasa feature extractor for automatic polyp detection. In *2017 IEEE International Conference on Image Processing (ICIP)* (Sept 2017), pp. 2060–2064.

[134] Tajbakhsh, N., Chi, C., Gurudu, S. R., and Liang, J. Automatic polyp detection from learned boundaries. In *Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on* (2014), IEEE, pp. 97–100.

[135] Tajbakhsh, N., Gurudu, S. R., and Liang, J. A classification-enhanced vote accumulation scheme for detecting colonic polyps. In *Abdominal Imaging. Computation and Clinical Applications. ABD-MICCAI* (2013), Springer, pp. 53–62.

[136] Tajbakhsh, N., Gurudu, S. R., and Liang, J. Automatic polyp detection using global geometric constraints and local intensity variation patterns. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2014), Springer, pp. 179–187.

[137] Tajbakhsh, N., Gurudu, S. R., and Liang, J. Automatic polyp detection in colonoscopy videos using an ensemble of convolutional neural networks. In *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on* (2015), IEEE, pp. 79–83.

[138] Tajbakhsh, N., Gurudu, S. R., and Liang, J. Automated polyp detection in colonoscopy videos using shape and context information. *IEEE transactions on medical imaging 35*, 2 (2016), 630–644.

[139] Tjoa, M. P., and Krishnan, S. M. Feature extraction for the analysis of colon status from the endoscopic images. *BioMedical Engineering OnLine 2*, 1 (2003), 9.

[140] Urban, G., Tripathi, P., Alkayali, T., Mittal, M., Jalali, F., Karnes, W., and Baldi, P. Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy. *Gastroenterology* (2018).

[141] Van Rijsbergen, C. J. *Information retrieval*. Butterworths, 1979.

[142] Vázquez, D., Bernal, J., Sánchez, F. J., Fernández-Esparrach, G., López, A. M., Romero, A., Drozdzal, M., and Courville, A. A benchmark for endoluminal scene segmentation of colonoscopy images. *Journal of healthcare engineering 2017* (2017).

[143] Vincent, L., and Soille, P. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis & Machine Intelligence 13*, 6 (1991), 583–598.

[144] Wang, Y., Tavanapong, W., Wong, J., Oh, J., and De Groen, P. C. Part-based multiderivative edge cross-sectional profiles for polyp detection in colonoscopy. *IEEE Journal of Biomedical and Health Informatics 18*, 4 (2014), 1379–1389.

[145] Wang, Y., Tavanapong, W., Wong, J., Oh, J. H., and De Groen, P. C. Polyp-alert: Near real-time feedback during colonoscopy. *Computer methods and programs in biomedicine 120*, 3 (2015), 164–179.

[146] Wittenberg, T., Zobel, P., Rathke, M., and Mühldorfer, S. Computer aided detection of polyps in whitelight-colonoscopy images using deep neural networks. *Current Directions in Biomedical Engineering 5*, 1 (2019), 231–234.

[147] Yu, L., Chen, H., Dou, Q., Qin, J., and Heng, P. A. Integrating online and offline three-dimensional deep learning for automated polyp detection in colonoscopy videos. *IEEE Journal of Biomedical and Health Informatics 21*, 1 (Jan 2017), 65–75.

[148] Yuan, Z., IzadyYazdanabadi, M., Mokkapati, D., Panvalkar, R., Shin, J. Y., Tajbakhsh, N., Gurudu, S., and Liang, J. Automatic polyp detection in colonoscopy videos. In *Medical Imaging 2017: Image Processing* (2017), vol. 10133, International Society for Optics and Photonics, p. 101332K.

[149] Zaitoun, N. M., and Aqel, M. J. Survey on image segmentation techniques. *Procedia Computer Science 65* (2015), 797–806.

[150] Zhang, R., Zheng, Y., Poon, C. C., Shen, D., and Lau, J. Y. Polyp detection during colonoscopy using a regression-based convolutional neural network with a tracker. *Pattern Recognition 83* (2018), 209 – 219.

[151] Zheng, Y., Zhang, R., Yu, R., Jiang, Y., Mak, T. W. C., Wong, S. H., Lau, J. Y. W., and Poon, C. C. Y. Localisation of colorectal polyps by convolutional neural network features learnt from white light and narrow band endoscopic images of multiple databases. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (July 2018), pp. 4142–4145.

[152] Zhou, D., Fang, J., Song, X., Guan, C., Yin, J., Dai, Y., and Yang, R. Iou loss for 2d/3d object detection. In *2019 International Conference on 3D Vision (3DV)* (2019), IEEE, pp. 85–94.

[153] Zuiderveld, K. Contrast limited adaptive histogram equalization. In *Graphics gems IV* (1994), Academic Press Professional, Inc., pp. 474–485.