



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO - CTC
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Marcus Venicius Lau Branch

**Metodologia para Segmentação de Pólipos em Imagens de Colonoscopia com o auxílio
de Redes Neurais Convolucionais**

Florianópolis
2023

Marcus Venicius Lau Branch

**Metodologia para Segmentação de Pólipos em Imagens de Colonoscopia com o auxílio
de Redes Neurais Convolucionais**

Dissertação submetida ao Programa de Pós Graduação
em Engenharia Elétrica da Universidade Federal de
Santa Catarina para a obtenção do título de em Mestre
em Engenharia Elétrica.

Orientador: Prof. Raimes Moraes, Dr.

Florianópolis

2023

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Branch, Marcus Venicius Lau
Metodologia para Segmentação de Pólipos em Imagens de
Colonoscopia com o auxílio de Redes Neurais Convolucionais
/ Marcus Venicius Lau Branch ; orientador, Raimés Moraes,
2023.
85 p.

Dissertação (mestrado) - Universidade Federal de Santa
Catarina, Centro Tecnológico, Programa de Pós-Graduação em
Engenharia Elétrica, Florianópolis, 2023.

Inclui referências.

1. Engenharia Elétrica. 2. Segmentação. Pólipos. Redes
Neurais Convolucionais. Colonoscopia.. I. Moraes, Raimés.
II. Universidade Federal de Santa Catarina. Programa de Pós
Graduação em Engenharia Elétrica. III. Título.

Marcus Venicius Lau Branch

Metodologia para Segmentação de Pólipos em Imagens de Colonoscopia com o auxílio de Redes Neurais Convolucionais

O presente trabalho em nível de mestrado foi avaliado e aprovado em 01 de junho de 2023, por banca examinadora composta pelos seguintes membros:

Fernanda Regina Smith Neves Corrêa, Dra.
Universidade Federal do Amapá - UNIFAP

Roberto Célio Limão e Oliveira, Dr.
Universidade Federal do Pará - UFPA

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de mestre em Engenharia Elétrica.

Prof. Telles Brunelli Lazzarin, Dr.

Coordenador do Programa de Pós-Graduação em Engenharia Elétrica

Prof.(a) Raimes Moraes, Dr.

Orientador(a)

AGRADECIMENTOS

Agradeço a Deus por ter me proporcionado chegar até aqui para alcançar e finalizar esta etapa tão importante em minha vida.

À minha família, principalmente aos meus pais Mauro Branch e Dulcirene Lau e ao meu irmão Mauricio Branch, que nunca deixaram de me apoiar e sempre me ajudaram em tudo da melhor forma possível e indicaram o melhor caminho.

A todos os grandes amigos que fiz durante o mestrado que fizeram ser um pouco mais fácil a chegada em uma cidade desconhecida, em especial, Victor Rocha, Raul Nascimento e Adriele Carvalho e pelos ótimos momentos, mas também pelos momentos de aperto nas disciplinas que renderam grandes aprendizados.

À minha namorada e companheira Cintia Amaral por sempre me apoiar e acreditar em mim e tornar tudo mais fácil.

Ao orientador deste projeto, Prof. Dr. Raimes Moraes, por ter acreditado na pesquisa e pela grande ajuda na realização desta. E claro, a todos os demais professores do programa de mestrado por todo conhecimento repassado durante esses anos.

“O que todos devemos fazer é ter certeza de que estamos usando a inteligência artificial de uma maneira que beneficie a humanidade, e não que a deteriore.”

Tim Cook

RESUMO

Nesta pesquisa, o objetivo principal é utilizar técnicas avançadas de aprendizado profundo para a segmentação precisa de pólipos em imagens de colonoscopia. A intenção é superar os resultados alcançados por abordagens anteriores. Para isso, foram empregadas as redes neurais convolucionais MobileNetV2 e EfficientNetB0, disponíveis na biblioteca Keras do Python, como componentes fundamentais da rede U-Net. Os modelos de redes selecionados para este trabalho foram pré-treinados no conjunto de dados ImageNet, disponível na biblioteca Keras, e a abordagem de aprendizado por transferência foi utilizada para adaptá-los ao nosso propósito específico. Os bancos de dados Kvasir-SEG, PICCOLO, CVC-ClinicDB e ETIS-LaribPolypDB foram empregados nesta pesquisa. Esses bancos de dados contêm diversas imagens de pólipos extraídas de gravações de exames de colonoscopia, juntamente com suas segmentações realizadas manualmente por profissionais médicos. As imagens desses bancos de dados foram diretamente utilizadas para treinamento e avaliação das redes neurais. Os resultados dos oito testes executados com as redes treinadas revelaram métricas de desempenho muito satisfatórias, com uma média de acerto de 85,91% para a métrica DSC e 77,42% para a métrica IoU. Esses números demonstram que o método proposto é eficiente na tarefa de segmentação de pólipos. Ao compararmos com outras abordagens publicadas, fica evidente que o método descrito neste trabalho apresenta um desempenho superior.

Palavras-chave: Segmentação. Pólipos. Redes Neurais Convolucionais. Colonoscopia.

ABSTRACT

This research aims to utilize deep learning techniques for polyp segmentation in colonoscopy exam images. The goal is to achieve segmentation metrics superior to previously proposed methods. The MobileNetV2 and EfficientNetB0 convolutional neural networks available in the Keras library in Python were used as integral parts of the U-Net architecture. The chosen network models for this work are pre-trained on the ImageNet database, also available in the Keras library, and will be implemented using transfer learning. The databases used in this research include Kvasir-SEG, PICCOLO, CVC-ClinicDB, and ETIS-LaribPolypDB, all of which consist of various colonoscopy exam images with manually annotated polyp segmentations by medical professionals. The images from these databases were directly used to train and evaluate the neural network. The results of the eight tests conducted on the trained networks in this research showed satisfactory performance metrics with an average accuracy rate of 85.91% for the DSC metric and 77.42% for the IoU metric. Therefore, the proposed method demonstrated good segmentation performance. A comparison with other published approaches revealed superior performance of the method described in this work.

Keywords: Segmentation. Polyps. Convolutional Neural Networks. Colonoscopy.

LISTA DE FIGURAS

Figura 1 – Colonoscópio	16
Figura 2 – Ilustração do exame de colonoscopia.....	16
Figura 3 – Rede neural.....	20
Figura 4 – Arquitetura de um <i>perceptron</i>	21
Figura 5 – Rede neural <i>feedforward</i>	22
Figura 6 – Rede neural convolucional.....	24
Figura 7 – Camada convolucional com passo igual a 1	25
Figura 8 – Exemplo de operação de <i>max-pooling</i> para uma matriz 2x2.....	25
Figura 9 – Exemplo de operação de <i>average-pooling</i> para uma matriz 2x2.....	26
Figura 10 – Aprendizado por transferência	27
Figura 11 – Limiarização.....	29
Figura 12 – Imagem original e máscara correspondente.....	34
Figura 13 – Diagrama de blocos do treinamento das redes.....	35
Figura 14 – Técnica de redimensionamento.....	37
Figura 15 – Região de interesse.....	38
Figura 16 – Aumento de dados em imagens e máscaras	39
Figura 17 – Arquitetura U-Net	41
Figura 18 – Comparação do desempenho do modelo com codificador da MobileNetV2 a) com pesos aleatórios e b) com pesos pré-treinados.....	47
Figura 19 – Comparação do desempenho do modelo com codificador da EfficientNetB0 a) com pesos aleatórios e b) com pesos pré-treinados.....	48
Figura 20 – MobileNetV2 + U-Net	49
Figura 21 – EfficientNetB0 + U-Net	50
Figura 22 – Curvas das métricas DSC e IoU no KvasirSEG com a rede MobileNetV2.....	52
Figura 23 – Curva da função perda de dados no treino e validação no Kvasir-SEG com a rede MobileNetV2+U-Net.....	53
Figura 24 – Comparação dos resultados das segmentações: a) imagens originais; b) máscaras originais; c) máscara gerada pela rede MobileNetV2+U-Net com o banco de dados Kvasir-SEG.....	54
Figura 25 – Curvas das métricas DSC e IoU no PICCOLO com a rede MobileNetV2.....	55

Figura 26 – Curva da função perda de dados no treino e validação no PICCOLO com a rede MobileNetV2+U-Net.....	55
Figura 27 – Comparação dos resultados das segmentações: a) imagens originais; b) máscaras originais; c) máscara gerada pela rede MobileNetV2+U-Net com o banco de dados PICCOLO.....	56
Figura 28 – Curvas das métricas DSC e IoU no CVC-ClinicDB com a rede MobileNetV2 ...	57
Figura 29 – Curva da função perda de dados no treino e validação no CVC-ClinicDB com a rede MobileNetV2+U-Net.....	58
Figura 30 – Comparação dos resultados das segmentações: a) imagens originais; b) máscaras originais; c) máscara gerada pela rede MobileNetV2+U-Net com o banco de dados CVC-ClinicDB.....	59
Figura 31 – Curvas das métricas DSC e IoU no ETIS-LaribPolypDB com a rede MobileNetV2+U-Net.....	60
Figura 32 – Curva da função perda de dados no treino e validação no ETIS-LaribPolypDB com a rede MobileNetV2+U-Net.	60
Figura 33 – Comparação dos resultados das segmentações: a) imagens originais; b) máscaras originais; c) máscara gerada pela rede MobileNetV2+U-Net com o banco de dados ETIS-LaribPolypDB.....	61
Figura 34 – Curvas das métricas DSC e IoU no KvasirSEG com a rede EfficientNetB0+U-Net.	62
Figura 35 – Curva da função perda de dados no treino e validação no Kvasir-SEG com a rede EfficientNetB0+U-Net.	62
Figura 36 – Comparação dos resultados das segmentações: a) imagens originais; b) máscaras originais; c) máscara gerada pela rede EfficientNetB0+U-Net com o banco de dados Kvasir-SEG.....	63
Figura 37 – Curvas das métricas DSC e IoU no PICCOLO com a rede EfficientNetB0+U-Net.	64
Figura 38 – Curva da função perda de dados no treino e validação no PICCOLO com a rede EfficientNetB0+U-Net.	64
Figura 39 – Comparação dos resultados das segmentações: a) imagens originais; b) máscaras originais; c) máscara gerada pela rede EfficientNetB0+U-Net com o banco de dados PICCOLO.....	65
Figura 40 – Curvas das métricas DSC e IoU no CVC-ClinicDB com a rede EfficientNetB0+U-Net.	66

Figura 41 – Curva da função perda de dados no treino e validação no CVC-ClinicDB com a rede EfficientNetB0+U-Net.....	66
Figura 42 – Comparação dos resultados das segmentações: a) imagens originais; b) máscaras originais; c) máscara gerada pela rede EfficientNetB0+U-Net com o banco de dados CVC-ClinicDB.....	67
Figura 43 – Curvas das métricas DSC e IoU no ETIS-LaribPolypDB com a rede EfficientNetB0+U-Net.	68
Figura 44 – Curva da função perda de dados no treino e validação no ETIS-LaribPolypDB com a rede EfficientNetB0+U-Net.	68
Figura 45 – Comparação dos resultados das segmentações: a) imagens originais; b) máscaras originais; c) máscara gerada pela rede EfficientNetB0+U-Net com o banco de dados ETIS-LaribPolyDB.....	69

LISTA DE QUADROS

Quadro 1 – Informações dos bancos de dados	35
Quadro 2 – Valores dos hiperparâmetros	45

LISTA DE TABELAS

Tabela 1 – Métricas da arquitetura MobileNetV2+U-Net no banco de dados KvasirSEG.....	53
Tabela 2 – Métricas da arquitetura MobileNetV2+U-Net no banco de dados PICCOLO	56
Tabela 3 – Métricas da arquitetura MobileNetV2+U-Net no banco de dados CVC-ClinicDB.	58
Tabela 4 – Métricas da arquitetura MobileNet+U-Net no banco de dados ETIS-LaribPolypDB.....	61
Tabela 5 – Métricas da arquitetura EfficientNet+U-Net no banco de dados Kvasir-SEG	63
Tabela 6 – Métricas da arquitetura EfficientNet+U-Net no banco de dados PICCOLO	65
Tabela 7 – Métricas da arquitetura EfficientNet+U-Net no banco de dados CVC-ClinicDB..	67
Tabela 8 – Métricas da arquitetura EfficientNet+U-Net no banco de dados ETIS-LaribPolypDB.....	69
Tabela 9 – Comparação das métricas obtidas no CVC-ClinicDB com a rede EfficientNet+U-Net com outras pesquisas..	71
Tabela 10 – Comparação das métricas obtidas no Kvasir-SEG com a rede EfficientNet+U-Net com outras pesquisas...	71

LISTA DE ABREVIATURAS E SIGLAS

OPAS Organização Panamericana de Saúde

SAC Sociedade Americana do Câncer

CNN *Convolutional Neural Network*

IoU *Intersection over Union*

DSC *Dice Similarity Coefficient*

IA Inteligência Artificial

NLP *Natural Language Processing*

RNA Redes Neurais Artificiais

MLP *Multi-Layer Perceptron*

TP *True Positive*

FP *False Positive*

FN *False Negative*

GPU *Graphics Processing Unit*

RAM *Random Access Memory*

CPU *Central Processing Unit*

LISTA DE SÍMBOLOS

θ	Vetores com informações
σ	Função de ativação e desvio padrão
$*$	Convolução
μ	Média dos pixels

SUMÁRIO

1	INTRODUÇÃO	15
1.1	O EXAME DE COLONOSCOPIA	16
1.2	OBJETIVO	17
1.2.1	Objetivos Específicos	18
1.3	JUSTIFICATIVA	18
1.4	ESTRUTURA DA DISSERTAÇÃO	18
2	FUNDAMENTAÇÃO TEÓRICA.....	20
2.1	REDES NEURAIS ARTIFICIAIS	20
2.1.1	Redes Neurais Convolucionais	23
2.1.1.1	<i>Camada Convolucional</i>	<i>24</i>
2.1.1.2	<i>Camada de Agrupamento</i>	<i>25</i>
2.1.1.3	<i>Camada Totalmente Conectada (Fully Connected)</i>	<i>26</i>
2.1.2	Redes com Transferência de Aprendizado.....	27
2.1.3	Aprendizado Profundo.....	28
2.2	SEGMENTAÇÃO SEMÂNTICA.....	28
2.2.1	Segmentação semântica com métodos clássicos.....	29
2.2.2	Segmentação semântica com aprendizado profundo	30
2.3	TRABALHOS RELACIONADOS	31
3	METODOLOGIA.....	34
3.1	BANCOS DE DADOS UTILIZADOS	34
3.2	PROCEDIMENTO PROPOSTO	35
3.3	PRÉ-PROCESSAMENTO	36
3.3.1	Redimensionamento	36
3.3.2	Normalização.....	37
3.3.3	Expansão das imagens do banco de dados	38
3.3.4	Divisão dos dados.....	39

3.4	EXTRAÇÃO DE CARACTERÍSTICAS.....	39
3.4.1	U-Net.....	40
3.4.2	EfficientNetB0.....	41
3.4.3	MobileNetV2.....	42
3.5	TREINAMENTO.....	43
3.5.1	Hiperparâmetros.....	43
3.5.2	Métricas de Avaliação.....	45
3.5.2.1	<i>Dice Similarity Coefficient (DSC)</i>	45
3.5.2.2	<i>Intersection over Union (IoU)</i>	46
3.5.2.3	<i>Soft Dice (Perda de Dados)</i>	47
3.5.3	Comparação dos métodos de inicialização dos pesos.....	47
3.5.4	U-Net com MobileNetV2 pré-treinada.....	49
3.5.5	U-Net com EfficientNetB0 pré-treinada.....	50
3.5.6	Recursos Computacionais.....	51
4	RESULTADOS.....	52
4.1	ABORDAGEM UTILIZANDO A MOBILENETV2.....	52
4.1.1	Kvasir-SEG.....	52
4.1.2	PICCOLO.....	54
4.1.3	CVC-ClinicDB.....	57
4.1.4	ETIS-LaribPolypDB.....	59
4.2	ABORDAGEM UTILIZANDO A EFFICIENTNETB0.....	61
4.2.1	Kvasir-SEG.....	61
4.2.2	Piccolo.....	64
4.2.3	CVC-ClinicDB.....	66
4.2.4	ETIS-LaribPolypDB.....	68
4.3	COMPARAÇÃO ENTRE AS ABORDAGENS.....	70
4.4	COMPARAÇÃO com a literatura.....	70

5	CONCLUSÃO.....	73
	REFERÊNCIAS.....	74

1 INTRODUÇÃO

Segundo a Organização Pan-Americana de Saúde (OPAS, 2019), o câncer do tipo colorretal apresenta uma significativa incidência, podendo ocasionar óbitos, principalmente em países desenvolvidos (AGÊNCIA INTERNACIONAL PARA PESQUISA EM CÂNCER, 2018). Uma estimativa realizada por Bray et al. (2018) aponta que houve mais de um milhão de casos e cerca de 550 mil mortes no ano de 2018. Há projeção de que, até o final desta década, o número de casos será de cerca de 2,2 milhões e que as mortes associadas alcançarão o número de 1,1 milhão. Este tipo de câncer inicia com o surgimento de pólipos que se consistem em uma projeção saliente do tecido da parede intestinal ou retal. A priori, estes pólipos são benignos; porém, com o tempo, podem se tornar malignos e trazer sérios riscos à saúde e à qualidade de vida do paciente.

De acordo com a Sociedade Americana do Câncer (SAC, 2021), o câncer colorretal tem cinco estágios de evolução com diferentes opções de tratamento e diferentes taxas de sobrevivência. O estágio inicial é chamado de pré-clínico ou carcinoma intramucoso, pois o pólipo ainda não cresceu além da camada interna do cólon ou reto (TRESKA, 2021). No segundo estágio, já é possível observar um crescimento do pólipo e sua progressão para um tumor além da camada interna; porém, ainda é local e não se espalhou por outras partes próximas. Neste caso, a possibilidade de sobrevivência de cinco anos após o diagnóstico é de 95%. O terceiro estágio do câncer é observado quando o câncer já se espalhou para além do tecido inicial em que surgiu e passa a oferecer riscos maiores para o paciente; após tal diagnóstico, cerca de 40% dos pacientes falecem dentro de cinco anos (HARI et al., 2013). No estágio quatro, o câncer já se espalhou para fora do cólon e também para as áreas próximas e começa a comprometer os linfonodos (LEE, S. B. et al., 2021), os quais são pequenas estruturas existentes no organismo que auxiliam no combate a infecções e células cancerígenas. Neste estágio, o tratamento passa a ser mais agressivo que os anteriores, visto que a taxa de sobrevivência após o diagnóstico fica entre 35% e 60%. No último estágio, há a ocorrência de metástase, ou seja, quando a doença atinge outros órgãos do corpo, tornando muito difícil o tratamento e com, aproximadamente, 14% de chance de sobrevivência após a descoberta. Visto que cada estágio apresenta diferentes gravidades, há também a necessidade de diferentes tratamentos de acordo com o avanço da doença, podendo ser desde uma simples cirurgia de remoção do pólipo ou da área comprometida (quando nos estágios iniciais) até a necessidade de realizar quimioterapia nos casos mais avançados.

1.1 O EXAME DE COLONOSCOPIA

Dada a gravidade e velocidade do avanço da doença, é de extrema importância que estes pólipos sejam detectados nos estágios iniciais pelos médicos; o diagnóstico ocorre por meio do exame de colonoscopia (HASSINGER et al., 2010). Este exame é o padrão ouro para detecção de lesões colorretais, consistindo-se de um procedimento invasivo que captura imagens em tempo real do intestino grosso e de parte do íleo terminal (a porção final do intestino delgado). Para realização do exame, é utilizado um aparelho chamado colonoscópio (Figura 1).

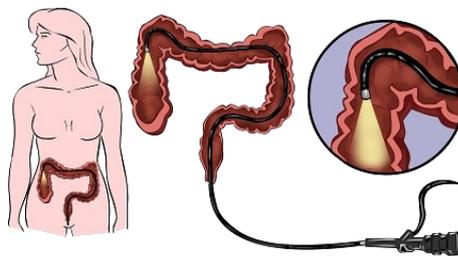
Figura 1 – Colonoscópio.



Fonte: (PENTAX MEDICAL, 2018).

O aparelho controlado por um gastroenterologista possui um tubo fino e flexível com uma câmera na sua extremidade para filmar o interior do intestino. O objetivo é identificar a presença de câncer colorretal, pólipos e doenças inflamatórias intestinais. Uma ilustração de como ocorre o exame é apresentada na Figura 2.

Figura 2 – Ilustração do exame de colonoscopia.



Fonte: (GASTROCLÍNICA FLORIANÓPOLIS, 2018).

De acordo com Sociedade Americana de Câncer, o exame de colonoscopia é um procedimento que deve ser realizado visando o diagnóstico nos estágios precoces da doença. A qualidade do exame de colonoscopia faz com que este exame ainda seja considerado o

padrão ouro, ou principal forma para identificar pólipos, inflamações ou câncer, reduzindo a taxa de mortalidade (LEE et al., 2016).

Além da colonoscopia, existem outras abordagens não invasivas para examinar o cólon, como ultrassom, Raio X e tomografia computadorizada (OBARO; BURLING; PLUMB, 2018). Esses outros métodos são capazes de produzir imagens em 2D e 3D, com riqueza de detalhes que facilitam a análise. Contudo, como são não invasivos, não permitem realizar ressecção ou biópsia do pólipo durante o exame. Devido a isso, segundo Doubeni et al. (2018), o exame de colonoscopia é a abordagem mais utilizada entre os profissionais, pois permite intervir e amostrar ou remover os pólipos.

Sabe-se da grande importância de detectar os pólipos em seus estágios iniciais para que sejam rapidamente examinados. Porém, de acordo com Van Rijn et al. (2006), muitos pólipos não são detectados durante os exames, visto que depende diretamente da experiência do profissional que o realiza. Portanto, um método que seja capaz de detectar e segmentar estes pólipos automaticamente auxiliaria bastante o diagnóstico, reduzindo a chance de que alguns pólipos passem despercebidos e, conseqüentemente, diagnósticos errados sejam realizados.

Para lidar com esse problema, algoritmos de visão computacional e técnicas de aprendizado de máquina foram estudados e implementados no estado da arte, visando automatizar a análise das imagens provenientes dos exames de colonoscopia.

A aplicação de técnicas de *Machine Learning* e *Deep Learning* na segmentação de imagens médicas tem estado em grande evidência e crescido atualmente. A abordagem para realizar a segmentação difere de acordo com os resultados que se deseja obter. A segmentação pode ser por meio de classificação binária (em que parte da imagem será segmentada, provavelmente recebendo o rótulo 1, e o resto será o fundo da imagem, recebendo o rótulo 0) ou multiclasse, quando o objetivo é segmentar várias texturas na imagem (BARROWCLOUGH et al., 2021) (MONTEIRO et al., 2020).

1.2 OBJETIVO

Esta pesquisa tem o objetivo de propor modelo para analisar imagens de pólipos provenientes de exames de colonoscopia baseado em redes neurais convolucionais e avaliar a capacidade do modelo proposto para segmentar estes pólipos através da comparação dos resultados com segmentação realizada por especialistas médicos.

1.2.1 Objetivos Específicos

Para que o objetivo geral seja alcançado, esta pesquisa tem como objetivos específicos:

- Realizar revisão bibliográfica do estado da arte para avaliar os avanços na área de segmentação de imagens.
- Levantamento de bancos de dados abertos e anotados que possam ser utilizados nesta pesquisa.
- Após a importação dos dados, avaliar o pré-processamento das imagens.
- A partir dos resultados obtidos, avaliar o desempenho das arquiteturas, hiperparâmetros, técnicas adotadas e a influência destes fatores no desempenho final, como métricas e tempo de treinamento.

1.3 JUSTIFICATIVA

A revisão da literatura especializada aponta que a segmentação de imagens médicas tem se tornado cada vez mais importante para auxiliar os profissionais da área a entregarem diagnósticos cada vez mais completos para os pacientes. Sendo assim, a comunidade científica vem investigando técnicas de processamento de imagens e diversas outras técnicas para possibilitar a implementação de um método automático para segmentação de pólipos em exames de colonoscopia. A razão deste interesse deve-se ao aumento da capacidade computacional e ao desenvolvimento de técnicas de processamento que se mostram mais confiáveis e com maior taxa de sucesso nos resultados.

Este trabalho visa contribuir com estes esforços, avaliando o desempenho de diferentes implementações para alcançar este propósito.

1.4 ESTRUTURA DA DISSERTAÇÃO

Este capítulo introduz e contextualiza a proposta de pesquisa e descreve seus objetivos. O próximo capítulo apresenta a fundamentação teórica, abordando pontos relevantes do tema dessa dissertação e que auxiliarão na sua melhor compreensão, bem como apresenta trabalhos relacionados existentes na literatura. O Capítulo 3 expõe a metodologia

utilizada para aplicação dos conhecimentos e as técnicas para a obtenção dos resultados. O Capítulo 4 contém a apresentação dos resultados obtidos com as curvas, métricas e predições geradas. No penúltimo capítulo, realiza-se uma comparação entre as abordagens realizadas com duas diferentes redes; os resultados são também comparados com resultados de outras pesquisas existentes no estado da arte. Por fim, o Capítulo 5 contém as conclusões acerca desta pesquisa com a validação dos resultados alcançados.

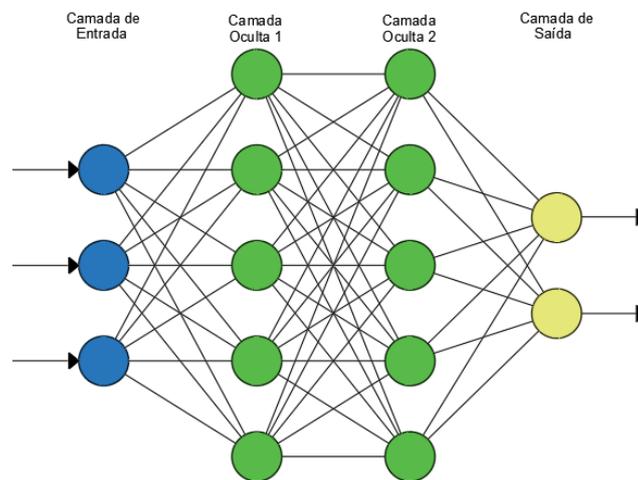
2 FUNDAMENTAÇÃO TEÓRICA

Esta seção apresenta a fundamentação teórica dessa dissertação. Aborda-se aspectos gerais das redes neurais, descrevendo importantes técnicas para o processamento e segmentação de imagens. Apresenta também, métricas propostas para a avaliação do desempenho, bem como, discutem bases teóricas relevantes para esta pesquisa.

2.1 REDES NEURAI ARTIFICIAIS

As Redes Neurais Artificiais (RNAs) são sistemas geralmente utilizados para classificação ou reconhecimento de padrões. Um modelo de rede neural artificial é capaz de processar dados mimetizando o que ocorre no cérebro humano com o objetivo de aprender e criar conhecimento a respeito de algo (AGGARWAL, 2018). Esses modelos possuem uma estrutura neuronal inspirada no córtex cerebral, mas em escala bastante reduzida, sendo por isso, denominada de rede neural (Figura 3).

Figura 3 – Rede neural.



Fonte: O autor.

A estrutura é composta por um grande número de unidades de processamento (denominadas neurônios) conectadas entre si e organizadas em camadas. Os neurônios trabalham em conjunto para resolver um problema e transmitir informações através de sinapses, ou seja, suas conexões. A camada de entrada recebe os dados, enquanto a camada de

saída gera o resultado final. Entre a primeira e a última camada, existem uma ou mais camadas também interligadas e conhecidas como ocultas.

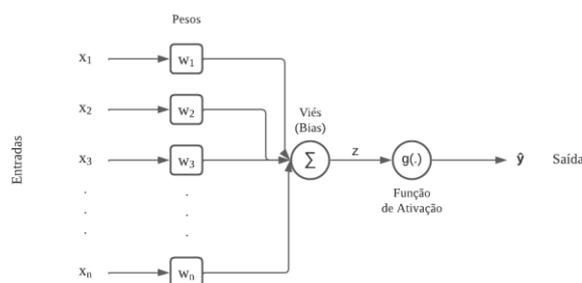
A proposta mais simples de uma rede neural é chamada de *perceptron*. Esse modelo menos robusto foi desenvolvido pelo matemático Walter Pitts e o neurofisiologista Warren McCulloch na década de 40. Eles criaram um modelo computacional fazendo analogia entre as células nervosas e o processo eletrônico de aprendizado que abriu caminho para a pesquisa das redes neurais. De acordo com Haykin (2009), o *perceptron* possui limitações, sendo aplicado em tarefas linearmente separáveis, visto que possui apenas um neurônio ou unidade. O modelo matemático do *perceptron* recebe entradas binárias x_i (x_1, x_2, \dots, x_n) que são relacionados a valores numéricos de saída por meio de pesos w_i (w_1, w_2, \dots, w_n). A adequada ponderação das entradas pelos pesos estabelece a influência destas na geração da saída (LIU et al., 2018). Segundo Gerstner et al. (2014), cada conexão tem um peso e cada neurônio tem um viés (ou *bias*) e uma função de ativação; a saída (com valores 0 ou 1) é determinada a partir da soma dos valores de entradas ponderadas pelos pesos por uma função de ativação com um viés aplicado de forma externa. A representação matemática está apresentada nas Equações 1 e 2:

$$z = \mathbf{w}^T \mathbf{x} + b \quad (1)$$

$$h = g(\mathbf{w}^T \mathbf{x} + b) = g(z) \quad (2)$$

Nas equações, \mathbf{x} representa o vetor de entrada, \mathbf{w} é o vetor peso e b corresponde ao viés. Na segunda equação, há a função de ativação $g(\cdot)$, a qual permite que a saída de cada conexão seja um valor escalar. A estrutura de um *perceptron* é apresentada na Figura 4.

Figura 4 – Arquitetura de um *perceptron*.

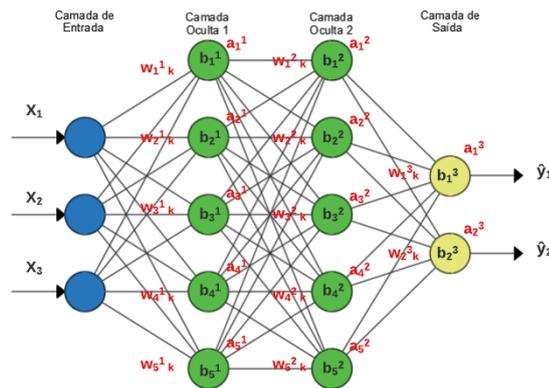


Fonte: O autor.

De acordo com Haykin (2009), para que este modelo funcione da forma esperada, as duas classes da tarefa devem ser linearmente separáveis. Para trabalhar com tarefas que não podem ser separadas linearmente, foram desenvolvidas as redes neurais artificiais, também chamadas de *feedforward* ou *Multi-Layer Perceptron* (MLP) por possuir vários neurônios.

Uma rede neural *feedforward* (Figura 5) é composta por camadas organizadas como neurônios humanos; cada unidade em uma camada se relaciona com todas as outras unidades das camadas vizinhas. Nesse modelo, as informações são processadas e transmitidas apenas em uma direção; ou seja, a partir da entrada de dados nos nós e camadas de entrada, para os nós e camadas ocultas e, por fim, para a camada de saída (HERTZ; KROGH; PALMER, 2019).

Figura 5 – Rede neural *feedforward*.



Fonte: O autor.

Conforme Goodfellow, Bengio e Courville (2016), a implementação desse modelo de rede pretende estimar o valor da saída representado por \hat{y} (Equação 3) com base na saída esperada y ,

$$\hat{y} = f(x|\theta) \quad (3)$$

onde θ representa os vetores que contêm os parâmetros de pesos da rede e auxilia na qualidade da predição da rede. A partir da Figura 5, pode-se observar que a interação entre as camadas pode ser expressa matematicamente pela Equação 4:

$$a_j^l = \sigma(\sum_k w_{jk}^l a_k^{l-1} + b_j^l) \quad (4)$$

onde,

a_j^l representa o peso do j -ésimo neurônio na l -ésima camada;

σ é a função de ativação escolhida para o modelo;

w_{jk}^l pondera a relevância da saída do k -ésimo neurônio na camada $(l - 1)$ para o j -ésimo neurônio na l -ésima camada;

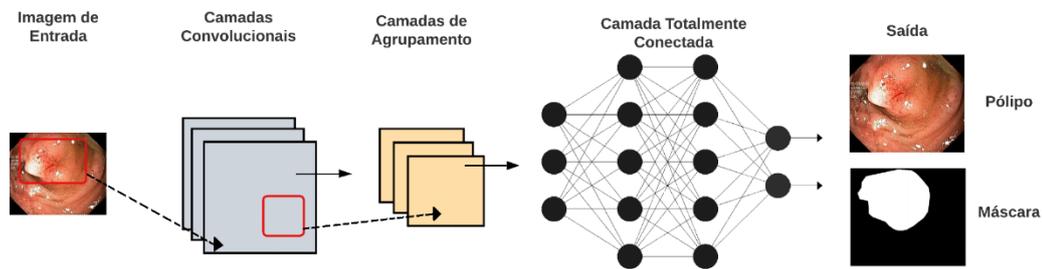
b_j^l representa o *bias* do j -ésimo neurônio na l -ésima camada.

A camada de entrada tem tantos neurônios quanto o número de dados de entrada. O número de neurônios na camada de saída depende do tipo de tarefa em que a rede está sendo utilizada. Quanto maior o número de camadas ocultas, maior será a capacidade da rede de estimar e realizar tarefas mais complexas a partir de maior volume de dados. Contudo, isso demanda alto custo computacional, influenciando diretamente no tempo de treinamento da rede que pode causar problemas como *overfitting* (ocorre quando uma rede tem bom desempenho durante o treinamento, mas não consegue repetir o mesmo desempenho no teste com imagens novas e desconhecidas para o modelo). A rede neural, portanto, pode não ser capaz de generalizar para dados não vistos previamente. Existem técnicas de regularização que podem evitar isso (KUKACKA et al. 2017). O algoritmo de *backpropagation* (RUMELHART et al., 1986), bastante implementado até os dias atuais, possibilita aprimorar a resposta gerada pelo modelo a partir dos ajustes dos pesos.

2.1.1 Redes Neurais Convolucionais

As Redes Neurais Convolucionais (CNN) possuem camadas que realizam operações de convolução (YAMASHITA et al., 2018). Suas maiores aplicações são em tarefas de visão computacional. Segundo Goodfellow, Bengio e Courville (2016), uma rede convolucional é capaz de reconhecer objetos e características nas imagens analisadas desde as básicas (bordas, formatos e tamanhos) às mais complexas (classificar e segmentar determinado objeto conforme seu aprendizado). A Figura 6 exibe a arquitetura básica de uma rede convolucional.

Figura 6 – Rede neural convolucional.



Fonte: O autor.

A CNN é composta por três tipos de camadas: convolucionais, de agrupamento e densas (ou totalmente conectada).

2.1.1.1 Camada Convolucional

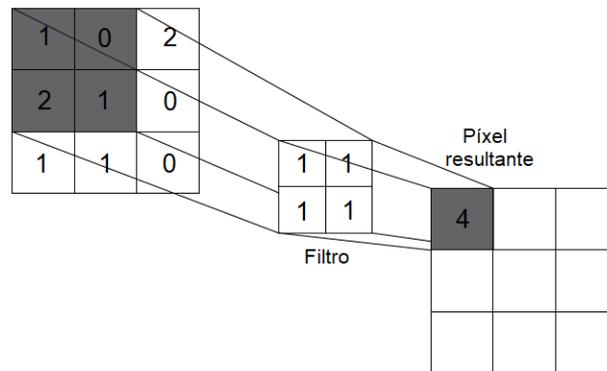
As camadas convolucionais são as responsáveis por extrair as características das imagens de entrada e aprender os padrões necessários para a realização de determinada tarefa (ERAZO, 2021). A operação de convolução pode ser descrita pela Equação 5.

$$y = x * w \quad (5)$$

Em que x e w são matrizes existentes que representam, respectivamente, a imagem de entrada e o filtro que percorre a imagem; y é o resultado da convolução. Os filtros ou *kernels* que percorrem as imagens são matrizes de valores escolhidos de acordo com o que se pretende filtrar nos dados com um determinado passo (ou *stride*) que estabelece o deslocamento do filtro sobre a imagem na operação de convolução (ERAZO, 2021). A Figura 7 ilustra o processo de convolução a partir da sobreposição do filtro com passo igual a 1.

Cada pixel é multiplicado por um peso, sendo a saída dada pela soma destes produtos. Os pesos e as dimensões da matriz variam de acordo com a função do filtro.

Figura 7 – Camada convolucional com passo igual a 1.



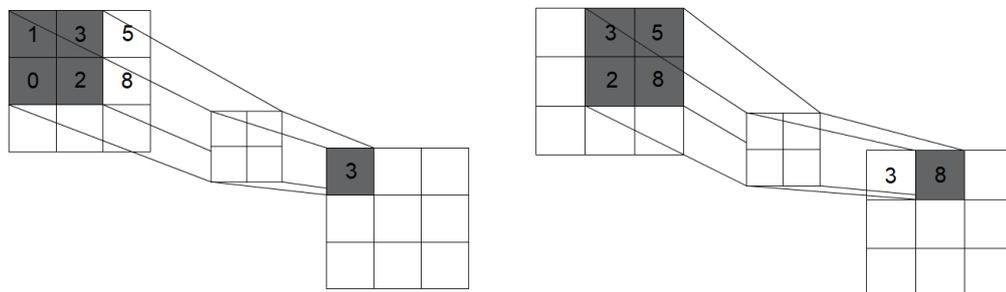
Fonte: O autor.

As saídas de uma camada convolucional são chamadas de mapas de recursos com as diversas características extraídas pelo modelo e indica os padrões encontrados nos dados analisados.

2.1.1.2 Camada de Agrupamento

As camadas de agrupamento ou *pooling* objetivam reduzir o tamanho dos mapas de recursos gerados pelas camadas convolucionais pela média (*average-pooling*) ou pelo valor máximo (*max-pooling*) dos pixels de determinada região das imagens. A função *max-pooling* (Figura 8) analisa uma região de pixels e retorna, na saída, o pixel de maior valor entre os analisados.

Figura 8 – Exemplo de operação de *max-pooling* para uma matriz 2x2.



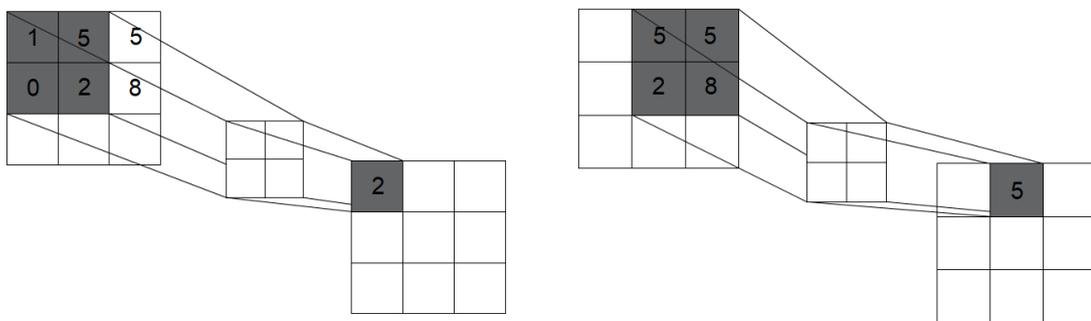
Fonte: O autor.

As posições nas quais se localizam os pixels de maior valor são onde se encontram as características importantes para a tarefa. Com a aplicação dessa técnica, partes da imagem que

não possuem relevância são desprezadas pela rede. Segundo Wu (2017), uma vantagem das camadas de *pooling* é que, reduzindo os mapas de recursos, reduz-se a quantidade de parâmetros da rede, tornando o modelo menos susceptível a ruídos, mais fácil de treinar e reduz a possibilidade da ocorrência de *overfitting*, nome dado comportamento quando a rede atinge bom desempenho no treinamento, porém não consegue replicar o mesmo desempenho em dados desconhecidos nos conjuntos de teste e validação, então não consegue generalizar.

A função *average-pooling* (Figura 9) analisa uma região de pixels e retorna na saída o valor médio entre os pixels sobrepostos.

Figura 9 – Exemplo de operação de *average-pooling* para uma matriz 2x2.



Fonte: O autor.

Implementar o *average-pooling* também reduz o efeito de características menos importantes, atribuindo igual peso para todos os elementos da região de pixels analisados (NIRTHIKA et al., 2022).

2.1.1.3 Camada Totalmente Conectada (Fully Connected)

As camadas totalmente conectadas, também chamadas de camadas densas, estão localizadas no final da rede e têm a função de tomar a decisão da tarefa realizada. Antes dos dados entrarem nessa camada, ocorre a transformação dos mapas de recursos resultantes das camadas de *pooling*, que até então eram representados na forma matricial, em vetores colunas através de uma operação denominada achatamento (*flattening*). Dessa forma, a camada densa conecta todas as características extraídas e aprendidas dos dados e gera a saída com base nos níveis de semelhança entre as imagens do treinamento e as imagens analisadas (MASUCATTO, 2018).

2.1.2 Redes com Transferência de Aprendizado

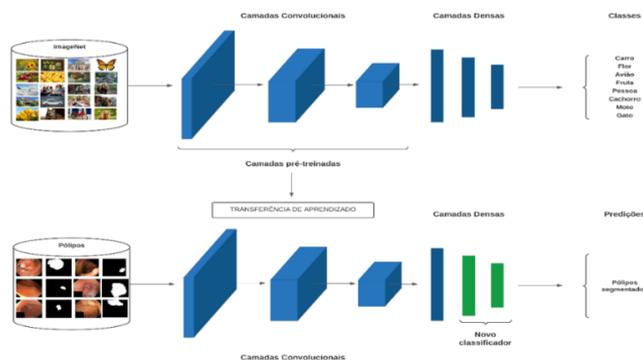
Desde o início da década passada, uma das competições mais famosas no ramo de detecção e classificação de objetos é a *ImageNet Large-Scale Visual Recognition Competition* (ILSVRC). O banco de dados ImageNet possui mais de 14 milhões de imagens rotuladas. Vários modelos de redes neurais convolucionais bastante complexas empregam o conjunto de dados ImageNet para classificar imagens com muito sucesso; essas redes foram treinadas por máquinas de alto poder computacional e por longas horas, conseguindo aprender diversas características das milhões de imagens disponíveis.

De acordo com Zhuang et al. (2021), o aprendizado por transferência pode ser implementado utilizando rede neural convolucional da ImageNet e seus pesos pré-treinados. Os modelos pré-treinados com o banco de dados ImageNet visam classificar imagens simples e naturais ou identificar formas e objetos básicos presentes nestas imagens. Portanto, faz-se necessário ajustar os parâmetros da rede para que o aprendizado possa ser implementado com precisão em imagens médicas complexas.

Conforme Cheplygina, De Bruijne e Pluim (2019), o ImageNet é o conjunto de dados mais utilizado para *transfer learning* em tarefas de análise de imagens médicas.

Em pesquisas mais recentes (ALZUBAIDI et al., 2020) (KARIMI; WARFIELD; GHOLIPOUR, 2021), a aprendizagem por transferência permitiu obter desempenho superior em tarefas de segmentação ou classificação no processamento de bancos de dados com poucas imagens médicas. As imagens médicas e as imagens do banco ImageNet diferem em muitos aspectos; logo, o ajuste dos parâmetros é essencial para que o aprendizado por transferência (Figura 10) apresente bons resultados.

Figura 10 – Aprendizado por transferência.



Fonte: O autor.

As camadas convolucionais da estrutura de rede apresentada na Figura 10 foram treinadas com o ImageNet por longas horas e têm a capacidade de reconhecer características gerais de imagens; isto é, formatos, bordas, texturas. Esse aprendizado será aproveitado nesta pesquisa. As últimas camadas, as camadas densas e totalmente conectadas, aprenderam características e detalhes extremamente específicos das imagens originais de treinamento que não são úteis para auxiliar na segmentação de pólipos por serem imagens e classes completamente diferentes. Sendo assim, as últimas camadas são substituídas por novas camadas e por um classificador binário para realizar a segmentação na saída da rede.

2.1.3 Aprendizado Profundo

O aprendizado profundo é um dos ramos do aprendizado de máquina que foi desenvolvido com base em redes neurais artificiais. Mais recentemente, as técnicas de *Deep Learning* tornaram-se o estado da arte para muitas tarefas de processamento de imagens e sinais, pois nesta modalidade, o aprendizado da rede neural ocorre através do uso e integração de diversas camadas ocultas que são responsáveis por extrair e aprender os padrões presentes nos dados de entrada e modelar e classificar estas características extraídas. Os métodos de aprendizado profundo são capazes de descobrir relações e dependências de imagem que são difíceis de serem percebidas por um humano e, até mesmo, que podem passar despercebidos em algoritmos de aprendizado de máquina menos robustos. O termo "profundo" significa que o número de camadas ocultas é muito maior do que aquelas usadas anteriormente e, por isso, conseguem extrair e aprender informações mais complexas dos dados.

Redes neurais convolucionais são algoritmos representativos de aprendizado profundo muito utilizados no processamento de imagens.

2.2 SEGMENTAÇÃO SEMÂNTICA

A segmentação semântica é um processo avançado de processamento de imagem que envolve a classificação de pixels de uma imagem em diferentes classes, com base em sua semântica ou significado. Em outras palavras, cada imagem que é analisada pelas camadas de entrada de uma rede neural é examinada em detalhes e cada pixel é atribuído a uma classe específica, que pode ser representada por um valor binário 0 ou 1. A partir da atribuição desses valores, a extração de características das imagens pode ser realizada de maneira

eficiente, permitindo que a rede neural produza uma saída que corresponde à segmentação da imagem em diferentes partes, com base em sua semântica. Esse processo é especialmente útil em aplicações de reconhecimento de objetos e detecção de cena, em que é necessário identificar áreas específicas de uma imagem com base em sua função ou significado.

Esta modalidade de segmentação tem sido amplamente utilizada, pois reduz significativamente a complexidade das tarefas e o custo computacional demandado pelo treinamento das redes. A segmentação semântica pode ser realizada por métodos clássicos de aprendizado de máquina e com redes neurais em aprendizado profundo.

2.2.1 Segmentação semântica com métodos clássicos

Alguns algoritmos foram propostos para aprimorar a segmentação. Dentre eles, tem-se o limiar de imagem (*image thresholding*).

O algoritmo de limiarização é o mais simples utilizado na segmentação de imagens (JYOTHI; K.BHARGAVI, 2014)(GURUNG; TAMANG, 2019). Esta técnica consiste em dividir uma imagem em duas classes de pixels, geralmente chamadas de primeiro plano e plano de fundo. O processo de limiarizar uma imagem ou binarização ocorre a partir da conversão de uma imagem original em uma segunda em tons de preto e branco (Figura 11), definindo um valor limite para classificação dos pixels. Os pixels da imagem analisada (originalmente no intervalo de 0 a 255) são convertidos em valores binários (0 ou 1). Dessa forma, se o valor de intensidade de um pixel for menor que o valor de limiar, ele é convertido para 1 (branco). Se o valor de um pixel for maior que o valor limite, o pixel será convertido em 0 (preto). Em aplicações simples, o limiar pode ser definido pelo desenvolvedor do algoritmo. Esse limiar é determinado com o objetivo de reduzir a variância intra-classes de pixels ou maximizar a variância entre as classes, sendo definida como a soma ponderada das variâncias das duas classes.

Figura 11 – Limiarização.



Fonte: O autor.

Outra técnica bastante implementada é o *clustering* em que os pixels com cores semelhantes são atribuídos à uma mesma classe (XU; TIAN, 2015). Dessa forma, ao analisar todas as imagens disponíveis, o modelo divide os pixels em classes de acordo com as suas similaridades, segmentando a imagem em regiões (AHMED; SERAJ; ISLAM, 2020).

Esses algoritmos tradicionais de segmentação de imagem não podem ser aplicados diretamente a casos e tarefas muito complexas por apresentarem limitações para analisarem grande quantidade de dados, informações e características das imagens. Assim, o surgimento do *Deep Learning* (DL) possibilitou a segmentação de imagens médicas de forma eficiente e eficaz.

2.2.2 Segmentação semântica com aprendizado profundo

Algoritmos de segmentação de imagens baseados em redes neurais convolucionais (CNNs) vêm sendo largamente utilizados devido à alta capacidade apresentada teórico e experimentalmente (GARCIA-GARCIA et al., 2017).

Inúmeras técnicas de segmentação semântica com redes neurais presentes na literatura seguem a estrutura da rede apresentada em Long et al. (2015), uma das primeiras propostas para segmentação de imagens com uma rede totalmente convolucional. O interesse em aplicar de aprendizado profundo para classificação é a possibilidade de substituir algumas camadas por um mapa de características extraído das imagens de entrada, sendo possível indicar a qual classe cada pixel da imagem pertence e assim, realizar a segmentação com o auxílio de operações de convolução (MICLEA; NEDEVSKI, 2020).

Os principais atuantes de uma rede são as camadas de convolução, de agrupamento e a totalmente conectada (Seção 2.1.1); a de convolução é o componente principal.

Jiang et al. (2018) propuseram uma rede capaz de segmentar radiografias do tórax de pacientes de um banco de dados público em diferentes classes para identificar o coração e os pulmões, tendo obtido resultados promissores.

A segmentação semântica com redes neurais pode ser aplicada em diversos campos e tarefas. As redes implementadas nesta pesquisa serão testadas em imagens médicas e serão apresentadas e discutidas nas seções subsequentes para melhor entendimento.

2.3 TRABALHOS RELACIONADOS

Pesquisas na área de segmentação de pólipos colorretais têm sido cada vez mais frequentes. Quando os métodos tradicionais são implementados para obter a segmentação, utiliza-se, geralmente, informações básicas presentes nas imagens, como os formatos e as cores dos pólipos, porém sem combiná-los. Em Hwang et al. (2007), os autores propuseram algoritmo para a detecção e segmentação dos pólipos que considera que a grande maioria dos pólipos possui formato elíptico; sendo assim, o algoritmo calcula a elipse da área de cada pólipo e não pólipo, realizando a identificação pelo contorno. Entretanto, por ser um método que generaliza os pólipos como elípticos, a taxa de erro é alta, pois acaba não reconhecendo muitos pólipos que não possuem esse formato. Ganz et al. (2012) utilizaram a Transformada de Hough para encontrar regiões em que os pólipos estão localizados, como também, um algoritmo que se baseia no formato e nos contornos dos pólipos. Ameling et al. (2009) propuseram um método que analisa a textura dos pólipos nas imagens com o auxílio dos níveis de cinza das imagens e os valores dos pixels de uma imagem segmentada, avaliando cada pixel normalizado em um intervalo de 0 a 1. Em Mamonov et al. (2014) e Tajbakhsh, Gurudu e Liang (2015), classificadores foram treinados para distinguir o pólipo do resto da imagem; porém, o desempenho foi aquém do desejado.

Apesar de serem abordagens interessantes, o formato e tamanho dos pólipos são bastante variados. Portanto, os métodos tradicionais são limitados e não conseguem identificá-los com muita eficiência, apresentando baixo desempenho na tarefa de detecção e segmentação. As pesquisas mais recentes passaram a utilizar redes neurais convolucionais (CNN) e aprendizado profundo, visto que esses modelos de rede possuem grande capacidade e recursos para extrair uma quantidade muito maior de informações das imagens. Assim, estes métodos geram segmentações muito melhores, tendo obtido vitórias em competições e desafios públicos. Shin et al. (2018) implementaram uma CNN capaz de detectar automaticamente os pólipos em imagens e vídeos extraídos de exames de colonoscopia baseando-se na região de cada pólipo; realizaram pré-processamento para melhorar a qualidade das imagens submetidas a processamento pela rede. Estes autores utilizaram as arquiteturas Inception e ResNet, tendo obtido resultados promissores. Jha et al. (2020) apresentaram o banco de dados Kvasir-SEG e utilizaram uma arquitetura de rede ResUnet (ZHANG; DOLWANI; YE, 2017). Os autores implementaram uma rede mais leve, com menor custo computacional; ainda assim, alcançaram resultados bastante satisfatórios de

segmentação. Para avaliação do desempenho, foram utilizadas as métricas IoU e DSC (Seção 4.5.1) que apresentaram valores iguais a 0,777 e 0,787, respectivamente, ao serem testados em novas imagens.

Com o sucesso da U-Net, pesquisadores (LI et al., 2022);(SIDDIQUE et al., 2021);(DIAKOGIANNIS et al., 2020) observaram que a integração de uma rede convolucional com a U-Net elevava a capacidade de reconhecimento e segmentação dos modelos, pois era possível extrair e aprender uma quantidade muito maior de características das imagens.

Algumas pesquisas (BRANDAO et al., 2017; AKBARI et al., 2018) utilizam um modelo de rede pré-treinada com o banco de dados ImageNet (DENG et al., 2009) que possui mais de 14 milhões de imagens. Assim, o modelo aprende diversas características e alcançam bom desempenho na tarefa de segmentação

A partir desses avanços, Jha et al. (2019) aprimoraram a rede e apresentaram a ResU-net++. Nesta, novos blocos e camadas foram adicionadas à rede; assim, essa se tornou mais sensível às características mais relevantes das imagens, relevando características menos importantes, proporcionando melhor qualidade da segmentação de novas imagens. Foi ainda possível reduzir o custo computacional, pois o modelo se tornou mais simples. Além das imagens utilizadas na pesquisa anterior, foram acrescentadas novas imagens do banco de dados CVC-ClinicDB (BERNAL et al., 2015), o qual possui diversas imagens de pólipos e segmentações manuais destes. Novamente, as métricas DSC e IoU foram implementadas, atingindo valores de 0,8133 e 0,792, respectivamente. Mostraram assim, que os aprimoramentos propostos pelos autores tiveram bons resultados em relação à pesquisa anterior.

O modelo proposto em Huang, Wu e Lin (2021) consiste-se em um codificador e decodificador (*encoder-decoder*) responsáveis por aprender as características das imagens e utilizá-las para gerar a segmentação como saída. Como codificador, os autores optaram por utilizar parte de outra rede já pré-treinada, a HarDNet (CHAO et al., 2019). Os testes foram realizados para cinco banco de dados diferentes, todos contendo imagens de exames de colonoscopia. Durante esses testes, diversos hiperparâmetros e técnicas de processamento das imagens foram avaliados, tendo gerado resultados bem satisfatórios e superiores aos resultados de outras pesquisas utilizadas para comparação.

Fang et al. (2019) propuseram uma rede que agregava as características das imagens analisadas com foco nas áreas e bordas dos pólipos existentes, obtendo resultados muito bons.

Porém, por ser uma rede muito robusta, o treinamento era extremamente lento; muitas vezes os resultados sofriam *overfitting* (quando um modelo tem bom desempenho com os dados do treinamento, porém não consegue repetir essa mesma qualidade na testagem em dados novos e desconhecidos).

Alam et al. (2020) também utilizaram uma rede neural pré-treinada como codificador da U-Net visando uma melhor extração de características das imagens. A rede escolhida possui blocos residuais que auxiliam a extrair as características mais importantes das imagens de entrada que são repassadas para o decodificador.

Diferentemente do modelo ResU-net++, a rede PraNet (FAN et al., 2020) busca a segmentação de toda a área do pólipo, ignorando assim, a restrição de limite de área, não definindo muito bem as bordas nos pólipos. Isto acabou depreciando a performance da rede visto que essas duas características dos pólipos auxiliam bastante a diferenciar tecidos normais dos tecidos com pólipos.

Pesquisas como as realizadas por (FANG et al., 2020) e (MAHMUD; PAUL; FATTAH, 2021) adicionaram novos blocos à cada camada da arquitetura da U-Net, aumentando a velocidade dos modelos e a precisão dos resultados obtidos, pois conseguiram extrair muitas características das imagens e combiná-las de diversas maneiras; o modelo aprendeu muitas variações importantes para reconstruir as imagens e realizar a segmentação.

Apesar de bastante pesquisada e estudada atualmente, a IA aplicada à medicina ainda é um grande desafio por conta de questões éticas, pela autorização ou não do uso das imagens de pacientes e pela ausência de bancos de dados suficientemente volumosos. Além disso, os dados existentes geralmente são desbanlaceados; isto é, com muitas informações de determinadas características e poucas de outras.

3 METODOLOGIA

Neste capítulo, serão descritos e analisados os *datasets* utilizados na pesquisa, os métodos de pré-processamento de imagem que serão implementados, as redes e suas arquiteturas, as métricas que auxiliam a melhorar os resultados obtidos e outras etapas e informações fundamentais para o desenvolvimento deste estudo. Nesta dissertação, a U-Net foi escolhida como rede padrão, pois é uma das mais utilizadas em tarefas de segmentação por conta de seu desempenho. Seu codificador será implementado com duas redes diferentes na etapa de treinamento: a EfficientNetB0 e a MobileNet-V2. Estas duas redes foram escolhidas pela eficiência e velocidade de processamento, ambas disponíveis a partir da biblioteca de código aberto Keras.

3.1 BANCOS DE DADOS UTILIZADOS

Para uma análise mais representativa, serão utilizados bancos de dados públicos, sendo eles, Kvasir-SEG (JHA et al., 2019), CVC-ClinicDB (BERNAL et al., 2015), ETIS-Larib Polyp DB (SILVA et al., 2014). Além disso, será utilizado também o banco de dados privados PICCOLO (SÁNCHEZ-PERALTA et al., 2020), o qual está disponível no site dos proprietários desses dados (<https://www.biobancovasco.org/en/Sample-and-data-catalog/Databases/PD178-PICCOLO-EN.html>), sendo sua utilização autorizada através de um formulário. Todos os *datasets* utilizados na pesquisa possuem dois tipos de imagem: uma imagem original da colonoscopia e outra já segmentada por profissionais da área da saúde, sendo esta última denominada máscara.

A Figura 12 apresenta uma amostra das imagens que serão fundamentais para o desenvolvimento do estudo.

Figura 12 – Imagem original e máscara correspondente.



Fonte: Dataset PICCOLO.

Os dados foram divididos em subconjuntos de treinamento, validação e teste. A grande maioria das imagens foi destinada para o subconjunto de treinamento para que a rede possa analisar uma maior quantidade de dados e aprender mais características importantes, sendo 80% do total; os subconjuntos de validação e teste contém 20% das imagens. O Quadro 1 abaixo apresenta as características dos bancos de dados utilizados na pesquisa.

Quadro 1 – Informações dos bancos de dados

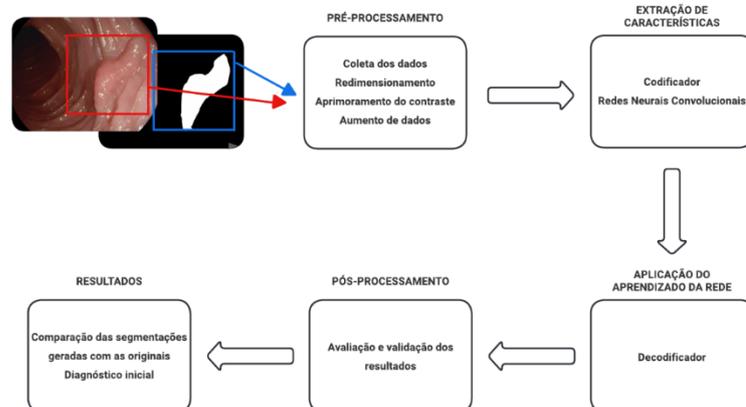
Banco de dados	Número de imagens	Resolução
Kvasir-SEG	1000	Variadas
PICCOLO	897	854 x 480
CVC-ClinicDB	612	384 x 288
ETIS-LaribPolypDB	196	1225 x 966

Fonte: O autor.

3.2 PROCEDIMENTO PROPOSTO

Antes da implementação de um método de segmentação de imagem, é necessário considerar os diferentes tipos de tarefas e dados utilizados. Imagens de diferentes conjuntos de dados com diferentes características e resoluções serão processadas da mesma forma para garantir reprodutibilidade e confiabilidade. O diagrama de blocos da Figura 13 apresenta as diversas etapas do processamento das imagens necessárias para a detecção dos pólipos nas imagens utilizadas na pesquisa. As próximas seções justificam e descrevem as técnicas utilizadas.

Figura 13 – Diagrama de blocos do treinamento das redes.



Fonte: O autor.

3.3 PRÉ-PROCESSAMENTO

De acordo com San, Win e Thant (2021), a implementação de técnicas de pré-processamento é uma etapa fundamental quando se trabalha com imagens nas áreas de visão computacional, aprendizado de máquina e aprendizado profundo. O pré-processamento reduz ou filtra informações das imagens que podem prejudicar o desempenho da rede na tarefa de segmentação. Estas técnicas incluem desde operações básicas para aumento ou redução do brilho, aumento da nitidez da imagem, redução de ruído, alteração da resolução das imagens até operações mais avançadas como a expansão do banco de dados.

A implementação do código utilizado nesta pesquisa foi com a linguagem de programação Python e suas bibliotecas que aumentam a produtividade do trabalho de processamento de imagens.

3.3.1 Redimensionamento

As redes neurais convolucionais apresentam restrição em relação ao tamanho das imagens de entrada, pois a grande maioria é projetada para aceitar apenas imagens de dimensões fixas e quadradas (TALEBI; MILANFAR, 2021).

As imagens dos bancos de dados utilizados nesta dissertação foram obtidas em aparelhos de alta resolução e possuem diferentes dimensões, o que cria desafios durante a preparação dos dados para alimentar a camada de entrada das redes, pois a redução dos tamanhos para dimensões padrões de 320x320 pode acarretar perda de informações importantes nas imagens, como a parte de um pólipo e, conseqüentemente, depreciar o desempenho das redes (LUKE; JOSEPH; BALAJI, 2019). Em contrapartida, imagens com altas resoluções demandam maior custo computacional e tempo de processamento, podendo tornar inviável o treinamento das redes dependendo da máquina utilizada.

Com exceção do banco de dados CVC-ClinicDB, todos os outros tiveram suas imagens redimensionadas para 320x320; o banco em questão teve suas imagens redimensionadas para 320x288 pelo fato de algumas de suas imagens possuírem dimensões menores. Assim, não houve perda significativa de informações nas imagens e nem tornou o treinamento lento por conta de altas resoluções. Na Figura 14 são apresentadas duas amostras da mesma imagem, porém em resoluções diferentes após o redimensionamento.

Figura 14 – Técnica de redimensionamento.



Fonte: Dataset PICCOLO.

3.3.2 Normalização

Em tarefas que envolvem o pré-processamento de imagens biomédicas, a normalização é amplamente utilizada. O objetivo é normalizar os dados amostrados com diferentes equipamentos e profissionais para que tenham uma média nula e um desvio padrão unitário. Isto porque essa prática geralmente acelera o processo de aprendizado da rede e leva a uma convergência em menor tempo (YIN et al., 2017).

A técnica de normalização *Z-score* tem a função de garantir que determinada imagem possua pixels com valores que estão dentro da média do conjunto de dados (ASHRAF et al., 2017). Nesta pesquisa, essa técnica foi aplicada nas imagens originais e nas suas máscaras correspondentes com os pólipos, incluindo os pixels em que há informação e aqueles que fazem parte do *background* e são definidos como 0 (Equação 6).

$$Z = X[X > 0] = \frac{(X[X > 0] - \mu)}{\sigma} \quad (6)$$

onde,

Z representa a pontuação *Z-score* encontrada;

$X[X > 0]$ é a máscara dos pólipos;

μ representa a média dos pixels das máscaras dos pólipos;

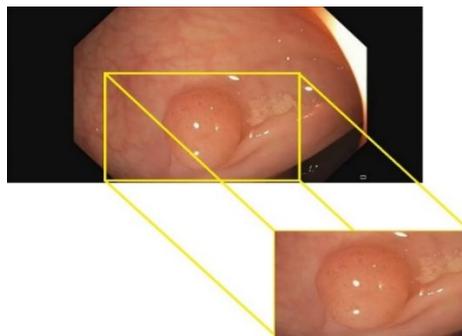
σ representa o desvio padrão dos pixels das máscaras dos pólipos.

3.3.3 Expansão das imagens do banco de dados

Para trabalhar com redes neurais, é necessária uma grande quantidade de dados para que seja possível realizar treinamentos adequados para se ter bons resultados. Entretanto, quando a tarefa é realizada com imagens médicas, a disponibilidade de uma grande quantidade de dados nem sempre é possível, pois é preciso muito esforço para coletar os dados e, em seguida, rotular estas imagens, fazendo com que muitos dos conjuntos de dados disponíveis sejam limitados. Para lidar com este problema, é possível expandir o número de imagens existentes em determinados *datasets* gerando novas imagens a partir das que já existem, tanto da imagem quanto da máscara (para o caso da segmentação); o modelo é então alimentado com estes novos dados gerados. De maneira geral, o aumento de dados tem como objetivo melhorar o desempenho, diminuindo possível *overfitting* da rede utilizada e permitir que o modelo treine com variações de orientação das imagens (SHORTEN; KHOSHGOFTAAR, 2019).

Segundo Hernández-García e König (2018), essas variações das imagens originais não alteram a região que interessa, mas fornece uma nova perspectiva de captura da imagem (Figura 15) para que a rede possa aprender novas características.

Figura 15 – Região de Interesse.



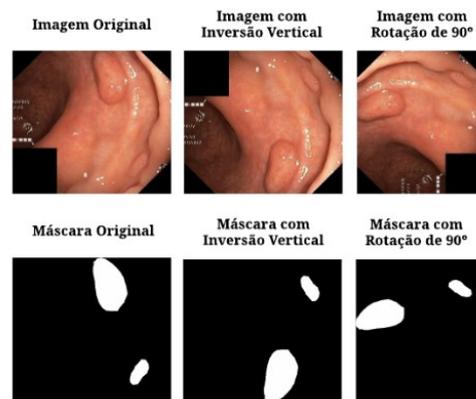
Fonte: O autor.

A expansão do banco de dados pode ser realizada com o auxílio de diversas bibliotecas disponíveis na linguagem de programação. Nesta dissertação, a biblioteca utilizada foi a *Albumentations*, pois além de possuir comandos que simplificam o processo de aumento dos dados, ainda implementa com alta eficiência diversas operações de transformação de imagem, bem como possui ferramentas que aprimoram a gestão dos novos dados gerados. As técnicas utilizadas nessa pesquisa para criar variações das imagens são:

- Rotação
- Zoom
- Inversão Vertical
- Inversão Horizontal
- Corte Central

A Figura 16 apresenta o resultado da implementação de duas técnicas de *data augmentation* em imagens utilizadas nesta pesquisa.

Figura 16 – Aumento de dados em imagens e máscaras.



Fonte: O autor.

3.3.4 Divisão dos dados

Após as etapas de pré-processamento para redimensionamento e normalização das imagens, bem como, aumento dos bancos de dados, foi realizada a divisão dos dados entre conjuntos de treinamento, validação e teste. A proporção determinada foi 80/10/10, em que 80% foram destinados ao treinamento e os 20% restantes divididos igualmente para os conjuntos de validação e teste, valores esses frequentemente utilizados na literatura.

3.4 EXTRAÇÃO DE CARACTERÍSTICAS

Nesta dissertação, foram escolhidas duas redes pré-treinadas (EfficientNetB0 e MobileNetV2) no banco de imagens ImageNet e disponíveis na biblioteca *Keras* (KERAS, 2015) para serem implementadas como o codificador da rede principal U-Net. Aplica-se,

portanto, transferência de aprendizado utilizando modelos inúmeras vezes treinados em conjuntos com muitos dados, permitindo que estes modelos aprendessem diversas características de diferentes tipos de imagens.

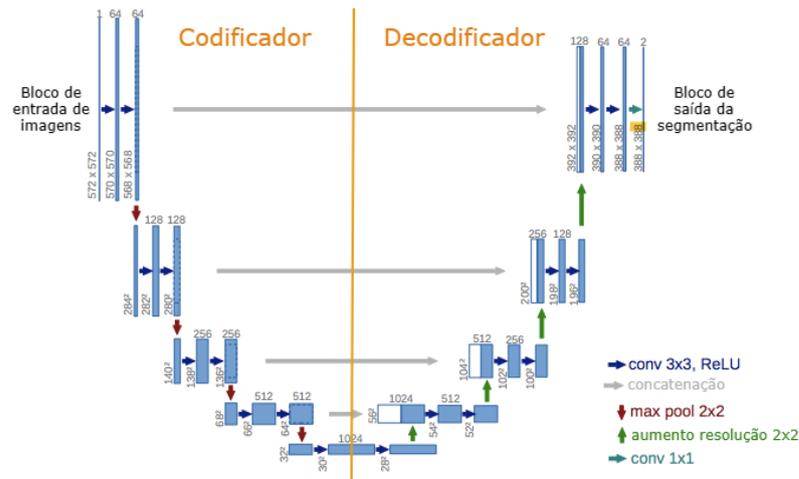
3.4.1 U-Net

Uma das redes mais amplamente utilizada para a segmentação de diferentes tipos de imagens é a U-Net (RONNEBERGER; FISCHER; BROX, 2015), tendo sido também empregada para segmentação de imagens médicas. Trata-se de uma rede totalmente convolucional (*Fully Convolutional Network*); ou seja, possui apenas camadas convolucionais sem nenhuma camada densa (LONG; SHELHAMER; DARRELL, 2015).

A U-Net é constituída por duas cadeias de processamento que se comunicam (Figura 17), compondo um diagrama em formato de “U” que a denomina. Da esquerda para a direita, estes dois ramos de processamento são chamados de codificador (ou caminho de contração) e decodificador (ou caminho de expansão). As duas cadeias desta rede são constituídas de vários blocos e camadas convolucionais e de *Max-Pooling* (Seção 2.1.1.2), responsáveis por reduzir a resolução das imagens (NIRTHIKA et al., 2022).

Na Figura 17, tem-se o codificador responsável por extrair e aprender as características das imagens de entrada através de operações convolucionais que reduzem a dimensão das imagens (*downsampling*) e realizam a extração de recursos. O ramo da direita (Figura 15), simétrica à primeira, realiza as mesmas operações convolucionais para recompor o tamanho das imagens que foram inicialmente inseridas na rede e utilizam as características aprendidas pelo codificador para reconstruir as imagens até a sua resolução original; procede então, e as segmentações a partir das imagens de entrada para comparação com suas respectivas máscaras (*ground-truth*) na saída da rede. Entre o codificador e o decodificador ocorrem operações de concatenação (da esquerda para a direita) que permitem a combinação de informações e características das camadas anteriores que possibilitam uma segmentação muito mais precisa. Basicamente, a concatenação é uma operação que une as informações de duas camadas diferentes da rede. Na U-Net, isso é feito adicionando as informações de uma camada em outra através e unindo as informações de uma camada de "downsampling" (diminuição da resolução espacial da imagem) com uma camada de "upsampling" (aumento da resolução espacial da imagem). A importância da concatenação entre os lados da rede se dá pela redução de resolução devido às operações de convolução.

Figura 17 – Arquitetura U-Net.



Fonte: (RONNEBERGER; FISCHER; BROX, 2015).

3.4.2 EfficientNetB0

As EfficientNets (TAN; LE, 2019) são uma família de redes neurais convolucionais de alto desempenho utilizadas em tarefas de classificação e segmentação de imagens. As EfficientNets otimizam os resultados, mantendo a qualidade mesmo com modelo menos robusto.

Existem oito modelos diferentes da EfficientNet, nomeados de EfficientNetB0 a EfficientNetB7. Nesta dissertação, será utilizado o modelo mais fundamental, a EfficientNetB0, pelo fato de ter menor custo computacional e menor tempo de treinamento.

É necessário prestar atenção em um ponto quando for implementá-la. As imagens devem possuir dimensões divisíveis por múltiplos de 8; caso contrário algumas camadas não conseguirão implementar a redução da resolução e extração de características, pois não receberão dados, aumentando o custo computacional de maneira desnecessária. Caso contrário, algumas camadas podem não receber informações suficientes, aumentando o custo computacional do processo de maneira desnecessária, como por exemplo a EfficientNetB0 também implementada nesta dissertação e que possui dimensão de entrada mínima de 224x224x3.

A arquitetura dessa rede é baseada em um modelo de escalabilidade composto por diferentes níveis de resolução de imagem, em que ela é capaz de ajustar a arquitetura com base no tamanho das imagens de entrada.

Após a rede ser alimentada com as imagens, essas são pré-processadas e passadas por camadas convolucionais que extraem as características básicas da imagem.

Cada bloco da EfficientNetB0 utiliza um mecanismo de atenção (técnica que permite que a rede se concentre nas características mais relevantes da imagem e também que a rede ajuste dinamicamente o peso que é dado a cada uma das características da imagem). Assim, a rede é capaz de se concentrar nas características mais importantes para a detecção e segmentação corretas. Devido à complexidade de técnica, esta não é descrita aqui por razões de concisão. Maior detalhamento pode ser encontrado em Hu et al. (2022).

Durante o treinamento, a rede é ajustada para maximizar a correção das respostas, com base nos dados fornecidos. A rede utiliza uma função de perda para calcular o erro de previsão e um ajuste de pesos da rede durante o treinamento.

Por fim, existe a camada de saída que realiza a classificação dos pixels para segmentar as imagens. A EfficientNetB0 também utiliza técnicas de regularização, como normalização em lote, para evitar *overfitting* durante o treinamento.

Outra característica da EfficientNetB0 é a presença dos blocos convolucionais chamados de *MBCConv*. Este módulo é similar aos blocos convolucionais presentes na MobileNetV2 (também será utilizada nesta pesquisa). Estes blocos conectam o fim de um bloco convolucional ao seu início, realizando a realimentação de informações. Assim, a extração e aprendizado de características das imagens de entrada é maior, sendo capaz de formar maiores mapas de recursos que serão repassados para o resto da rede gerar os resultados (AHN; YIM, 2020). Outra vantagem desse tipo de conexões entre os blocos internos da rede é a diminuição do número total de operações necessárias, bem como do tamanho do modelo (TU; ZHANG, 2017).

3.4.3 MobileNetV2

Proposta por Sandler et al (2018), a MobileNetV2 é uma versão aprimorada da MobileNetV1 (HOWARD et al., 2017); pertence a uma família de redes neurais desenvolvida para auxiliar em tarefas de segmentação, classificação e detecção. A MobileNetV2 possui implementação de camadas lineares, denominadas de *bottleneck layers*, que reduzem fluxo de dados na rede, tornando-a menos custosa computacionalmente, aumentando sua velocidade de processamento (CHOLLET, 2017).

Esta rede possui também conexões entre as camadas que auxiliam a rede na busca do melhor resultado durante o processo de *backpropagation*. Assim, o gradiente consegue atualizar os pesos das camadas anteriores para que os valores de perda sejam menores e as métricas de validação maiores (DROZDZAL et al., 2016).

A utilização desta rede pré treinada possui duas vantagens que foram muito importantes nesta pesquisa:

- Possui menor número de parâmetros treináveis, sendo assim, mais fácil e rápida de treinar;
- O modelo pré-treinado terá melhor desempenho e convergirá mais rápido do que a rede não treinada anteriormente.

Apesar de possuir menor número de parâmetros treináveis quando comparada com outras redes, a MobileNetV2 contém um grande número de camadas.

3.5 TREINAMENTO

A partir das imagens e suas máscaras, realiza-se o treinamento das redes para que seja possível segmentar novas imagens. Para avaliar o desempenho do treinamento e a capacidade das arquiteturas de segmentarem essas imagens é necessário utilizar algumas métricas de avaliação.

3.5.1 Hiperparâmetros

A técnica de ajuste de parâmetros para uma determinada tarefa utilizando rede neural é denominada de otimização de hiperparâmetros. Serão realizadas análises das duas diferentes redes utilizadas nesta pesquisa para compreender como os parâmetros escolhidos influenciam nos valores das métricas, no tempo de treinamento e no tempo de predição dessas redes.

Os treinamentos foram realizados a uma taxa de aprendizado fixa de 0,001, mas que decai com o passar das iterações quando o modelo percebe que não está havendo ganho nas métricas monitoradas, evitando o *overfitting* e garantindo que o modelo continue se ajustando aos dados.

Outro hiperparâmetro fixado é tamanho do lote (*batch size*), que é o número de imagens utilizadas para treinar a rede em uma única iteração (KANDEL; CASTELLI, 2020). De acordo com Goodfellow, Bengio e Courville (2016), tamanhos de lote muito pequenos podem exigir uma pequena taxa de aprendizado para manter um treinamento estável e contínuo, pois o modelo sofre variação na estimativa do mínimo global; porém, valores menores de lote apresentam uma convergência mais rápida (KANDEL; CASTELLI, 2020). Por outro lado, a utilização de um tamanho de lote muito grande conduzirá a rede a uma generalização ruim e de baixa eficiência na geração dos resultados, comprometendo o desempenho do modelo, que não consegue aprender muitas características das imagens (KESKAR et al., 2017)(KANDEL; CASTELLI, 2020). Na literatura, há uma grande discussão entre os diferentes tamanhos de lote e suas vantagens e desvantagens. Valores comumente utilizados para este parâmetro variam de 8 a 1024. Para esta pesquisa, o valor escolhido foi 16 para que o treinamento fosse mais estável em combinação com a pequena taxa de aprendizado e buscar convergência mais rápida.

De acordo com (TAYLOR et al., 2022), um dos principais objetivos dos algoritmos de aprendizado de máquina e aprendizado profundo é otimizar o processo de aprendizado e, conseqüentemente, reduzir os valores da função perda e elevar as métricas monitoradas. Reduzir a função perda é fundamental, pois a rede precisa aprender os pesos que melhor indicam a correlação entre os dados analisados. Esses pesos formarão o modelo capaz de gerar respostas e previsões quando apresentada a conjuntos de testes desconhecidos após serem treinadas. Para isso, existe alguns otimizadores que podem ser implementados nas redes. Nesta pesquisa, o otimizador utilizado foi o Adam (KINGMA; BA, 2014), pois é um método que auxilia o aprendizado da rede com base em estimativas atualizadas a cada momento do treinamento; não possui alta demanda de memória, tem baixo custo computacional e apresenta bom desempenho em tarefas com grande quantidade de dados (KINGMA; BA, 2014). O Adam calcula as taxas de aprendizados individuais que se adaptam a cada época para cada parâmetro da rede, otimizando todos em busca do mínimo.

O número de épocas é um parâmetro muito importante, pois determina a quantidade de ciclos com a qual rede irá analisar todo o conjunto de dados (DHANDE; SHAIKH, 2019). A escolha do número de épocas tende a evitar problemas como *underfitting* e *overfitting* que depreciam o desempenho do modelo. Estes problemas podem ocorrer por ausência ou excesso de épocas, pois um pequeno número de épocas pode fazer com que o modelo não aprenda o suficiente sobre os dados; por outro lado, se o modelo for treinado com um grande número de

épocas, este pode aprender características totalmente indesejadas que estão presentes nas imagens (ZHANG; ZHANG; JIANG, 2019). Nesta dissertação, todos os modelos foram executados para 70 épocas devido à limitação de recursos computacionais ocasionada pela pandemia, porém alguns convergiram com menor número. O Quadro 2 a seguir apresenta todos os hiperparâmetros utilizados e seus valores.

Quadro 2 – Valores dos hiperparâmetros.

Hiperparâmetro	Valor
Taxa de Aprendizado	0,001
Tamanho de Lote	16
Otimizador	Adam
Épocas	70

Fonte: O autor.

3.5.2 Métricas de Avaliação

3.5.2.1 Dice Similarity Coefficient (DSC)

Para analisar a qualidade da segmentação de imagens médicas gerada por uma rede, três métricas são bastante implementadas. A principal e mais difundida para problemas desta natureza, é o *Dice Similarity Coefficient* (DSC). Esta métrica indica a quantidade de sobreposição espacial entre dois conjuntos de dados, podendo também ser tratada como uma proporção de concordância e semelhança entre os conjuntos analisados (SHAMIR et al., 2018). Seu valor varia de 0 a 1, indicando que não há sobreposição espacial ou sobreposição completa, respectivamente. O DSC é obtido pela Equação 7.

$$DSC = \frac{2 \cdot \|X \cap Y\|}{\|X\| + \|Y\|} \quad (7)$$

onde,

X e Y são dois conjuntos de dados ou duas imagens;

Um conjunto com barras verticais (exemplo: $\|X\|$) refere-se à cardinalidade do conjunto, ou seja, o número de elementos contidos nesse conjunto;

\cap indica a interseção entre dois conjuntos e retorna os elementos comuns a ambos.

Outra forma de calcular e analisar o DSC é em termos da matriz de confusão, em que são mostrados os valores de Verdadeiros Positivos e Negativos e os Falsos Positivos e Negativos. Uma matriz de confusão é uma tabela que permite analisar o desempenho de um algoritmo ao realizar determinada predição (SALMON et al., 2015). De maneira geral, cada linha da matriz possui os valores relacionados às predições e cada coluna representa a classe real que é a referência para o treinamento do modelo. Neste segundo modo, o cálculo do DSC é realizado através da Equação 8.

$$DSC = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \quad (8)$$

onde,

True Positive (TP) indica o número de amostras que o modelo prediz corretamente a classe positiva;

False Positive (FP) indica o número de amostras que o modelo prediz incorretamente a classe positiva;

False Negative (FN) indica o número de amostras que o modelo prediz incorretamente a classe negativa.

3.5.2.2 Intersection over Union (IoU)

A métrica *Intersection over Union (IoU)*, também denominada de *Jaccard Index*, é quase sempre utilizada para avaliar métodos de segmentação. Esta é capaz de calcular e indicar a similaridade entre uma segmentação realizada e a sua máscara original (*ground-truth*), a referência no treinamento da rede (VAN BEERS et al., 2019). Assim como o DSC, os valores de IoU também variam de 0 (quando não há sobreposição) a 1 (quando a imagem segmentada corresponde perfeitamente à máscara original). A IoU é definida pela Equação 9.

$$IoU = \frac{\|X \cap Y\|}{\|X \cup Y\|} \quad (9)$$

Assim como a DSC, a IoU também pode ser reescrita em termos da matriz de confusão (Equação 10).

$$IoU = \frac{TP}{TP + FP + FN} \quad (10)$$

3.5.2.3 Soft Dice (Perda de Dados)

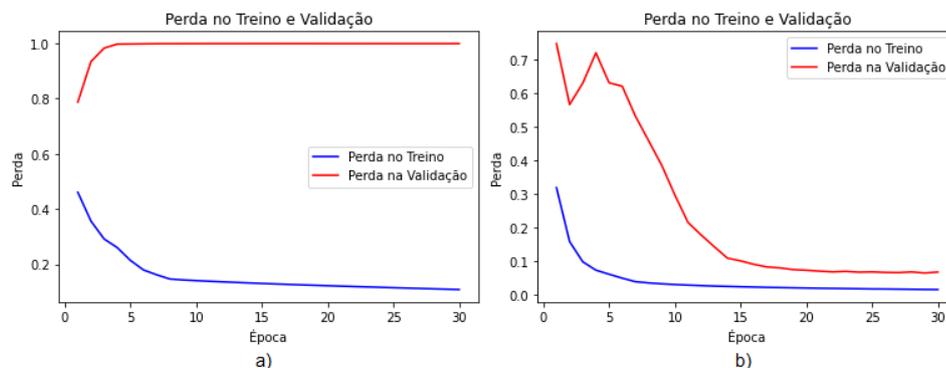
Outra métrica bastante utilizada em tarefas de segmentação de imagens médicas é a perda de dados. Basicamente, o valor da função perda gerado pela rede ao final do treinamento indica a diferença entre o valor da saída da rede e o valor que era esperado para uma determinada entrada, ou seja, a diferença entre as imagens de saída e entrada. A *Soft Dice*, conhecida como perda suave de dados, é derivada do DSC (Equação 11).

$$soft\ Dice = 1 - DSC \quad (11)$$

3.5.3 Comparação dos métodos de inicialização dos pesos

Essa seção tem o objetivo de comparar as duas redes investigadas neste trabalho com dois tipos de inicialização de pesos: com pesos pré-treinados no banco de dados ImageNet e com os pesos aleatórios. Os testes foram realizados alterando apenas a rede e a inicialização dos pesos, porém todas as outras configurações e hiperparâmetros foram mantidos iguais. O banco de dados utilizado para a comparação foi o Kvasir-SEG e o modelo foi treinado com apenas 30 épocas para análise do comportamento inicial. A Figura 18 apresenta as curvas obtidas com a MobileNetV2.

Figura 18 – Comparação de desempenho do modelo com codificador da MobileNetV2 a) com pesos aleatórios e b) com pesos pré-treinados.

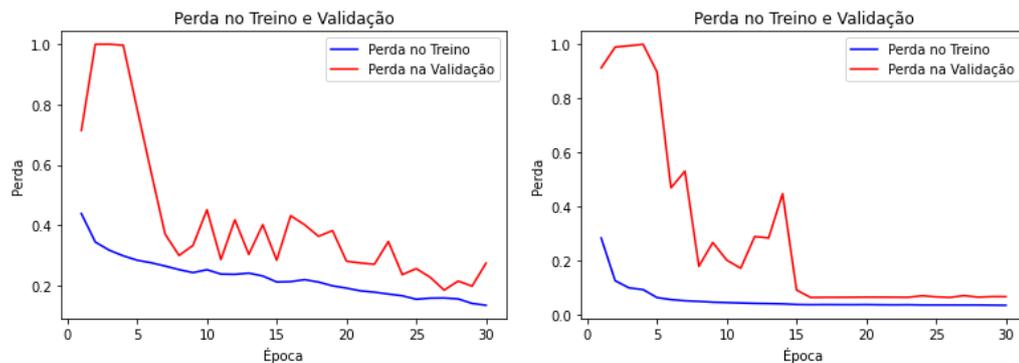


Fonte: O autor.

É possível observar que os dois casos geraram curvas muito diferentes. No primeiro teste, com os pesos aleatórios, ocorreu *overfitting*; o modelo assimilou detalhes e características particulares das imagens de treinamento, não conseguindo generalizar a segmentação para imagens novas do conjunto de validação, depreciando o desempenho. Já na iniciação com os pesos pré-treinados a métrica convergiu de maneira muito mais rápida.

O mesmo procedimento foi realizado com a rede EfficientNetB0. A Figura 19 apresenta as curvas de aprendizagem dessa rede com os dois métodos de inicialização dos pesos.

Figura 19 – Comparação de desempenho do modelo com codificador da EfficientNetB0 a) com pesos aleatórios e b) com pesos pré-treinados.



Fonte: O autor.

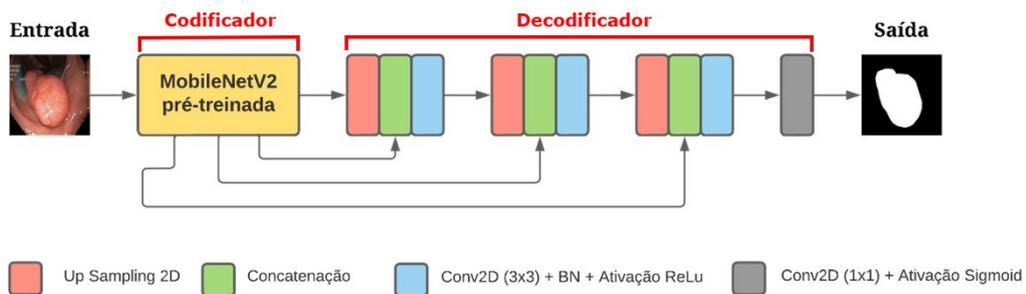
Após análise das curvas de ambas arquiteturas, fica evidente que utilizar as redes com os pesos pré-treinados melhora o desempenho das duas redes, pois é possível verificar uma grande variação na perda quando a rede possui pesos aleatórios (a), comportamento que é eliminado quando a rede é iniciado com os pesos já pré-treinados, melhorando e facilitando a convergência a uma menor perda e melhores resultados. Sendo assim, os treinamentos e testes nessa dissertação foram realizados com esta metodologia.

Ambos os modelos testados tiveram tempo de inferência aproximados, cerca de 150 minutos para as 30 épocas.

3.5.4 U-Net com MobileNetV2 pré-treinada

Nesta primeira proposta investigada, o caminho de contração da U-Net foi substituído pelo modelo MobileNetV2 pré-treinado para segmentar as regiões dos pólipos nas imagens de colonoscopia. A representação do modelo proposto é apresentada na Figura 20.

Figura 20 – MobileNetV2 + U-Net.



Fonte: O autor.

Ao usar o modelo pré-treinado e combinado com a U-Net, uma conexão horizontal é implementada para conectar a parte do codificador, que irá aprender as características das imagens originais e reduzir a resolução, à parte do decodificador. Dessa forma, as informações extraídas das imagens analisadas serão utilizadas no processo de segmentação final. As imagens são inseridas no modelo com resolução $320 \times 320 \times 3$, em que os dois primeiros valores representam a largura e comprimento da imagem e o último indica os três canais (RGB) da imagem colorida. Os quatro blocos convolucionais da primeira metade da rede reduzem a resolução da imagem com proporções 2, 4, 8 e 16, respectivamente, e repassam para a segunda metade da rede. O decodificador também possui quatro blocos convolucionais, em que cada bloco dispõe de uma operação de reconstrução das imagens seguidas por uma operação de concatenação que auxilia a reconstrução das imagens repassando o aprendizado obtido pela rede no codificador.

A segunda metade da rede, já com o decodificador original da U-Net, recebe as imagens com resolução reduzida com dimensões $20 \times 20 \times 3$ e inicia o processo de recuperação e aumento das dimensões. Nesta segunda parte da rede, o *batch normalization* é aplicado após cada camada convolucional e antes da função de ativação ReLU (*Rectified Linear Unit*). Essa técnica normaliza as saídas das camadas anteriores para que as entradas para as camadas seguintes estejam em uma escala mais próxima, ajudando a regularizar a rede, reduzir o

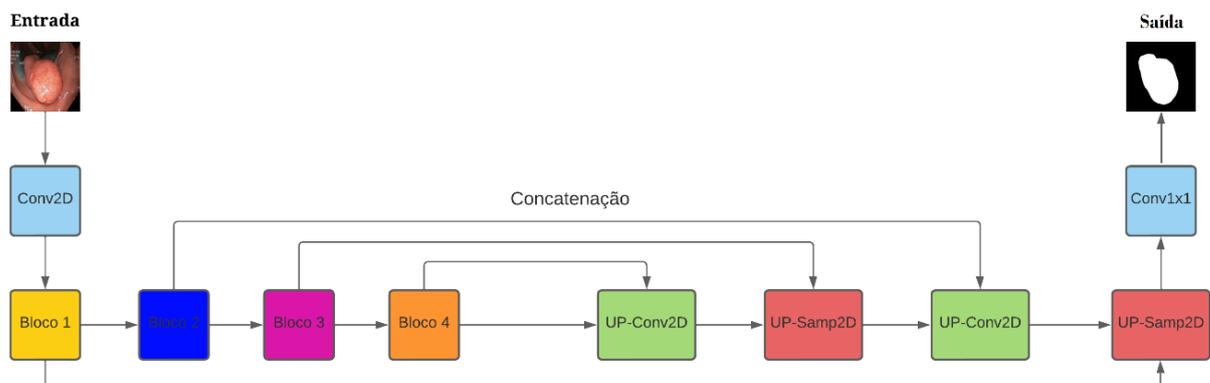
overfitting e tornar a rede mais robusta a variações nos dados de entrada e a melhorar a velocidade e estabilidade do treinamento. A função ReLU é geralmente aplicada após cada camada convolucional e comumente usada na rede U-Net para introduzir não-linearidade nas camadas convolucionais. Ao fazer isso, permite que a rede aprenda recursos mais complexos e abstratos nas imagens, o que pode levar a uma melhor segmentação de objetos em imagens médicas, por exemplo. Além disso, a função ReLU é computacionalmente eficiente e fácil de implementar, o que a torna uma escolha popular em muitas redes neurais convolucionais, incluindo a rede U-Net.

Da mesma forma e com algumas operações inversas, dados são repassados para o decodificador através das concatenações de informações, o que possibilita que as informações detalhadas das camadas de baixo nível sejam transmitidas para as camadas de alto nível, permitindo que a rede aprenda características precisas das imagens e segmenta regiões com maior acurácia, auxiliando diretamente na realização da predição pela rede após a reconstrução das imagens até sua dimensão original. A camada final é uma camada de convolução com um kernel de tamanho 1x1, que atua como um classificador, atribuindo uma probabilidade para cada pixel da imagem e gera a máscara segmentada como saída.

3.5.5 U-Net com EfficientNetB0 pré-treinada

Outra proposta investigada neste trabalho consiste na utilização da rede EfficientNetB0 pré-treinada como o codificador da U-Net. A representação deste modelo é apresentada na Figura 21.

Figura 21 – EfficientNetB0 + U-Net.



Fonte: O autor.

Como comentado na subseção 3.4.2, uma característica da EfficientNetB0 é a presença dos blocos convolucionais chamados de *MBCConv*. Os blocos numerados de 1 a 4 no lado esquerdo são os blocos que realizam as operações de concatenação entre os dois lados da rede e auxiliam na extração das características das imagens. Da mesma forma que realizado com a MobileNetV2, as imagens de entrada alimentam a rede com resolução $320 \times 320 \times 3$ e a segunda metade da rede possui o decodificador padrão da U-Net com o processo de recuperação da resolução e geração das segmentações.

3.5.6 Recursos Computacionais

Todos os modelos desta dissertação foram implementados utilizando a biblioteca Keras, a qual é uma biblioteca de alto nível para implementação de redes neurais.

Sabe-se que o aprendizado profundo possui grande custo computacional. Portanto, para esta dissertação foi utilizada a plataforma *Google Colab* com a linguagem de programação Python, pois é bastante interativa e fácil de compilar os códigos, além de possuir maior poder de processamento com as unidades de processamento gráfico (GPUs), pois sem essas unidades o treinamento pode ser extremamente lento. Sendo assim, nesta dissertação foi utilizada a GPU Tesla K80 com 12 GB de memória RAM e mais uma CPU AMD Ryzen 5 2500U com Radeon Vega Mobile Gfx com 2GB de RAM, auxiliando a redução do tempo de treinamento.

4 RESULTADOS

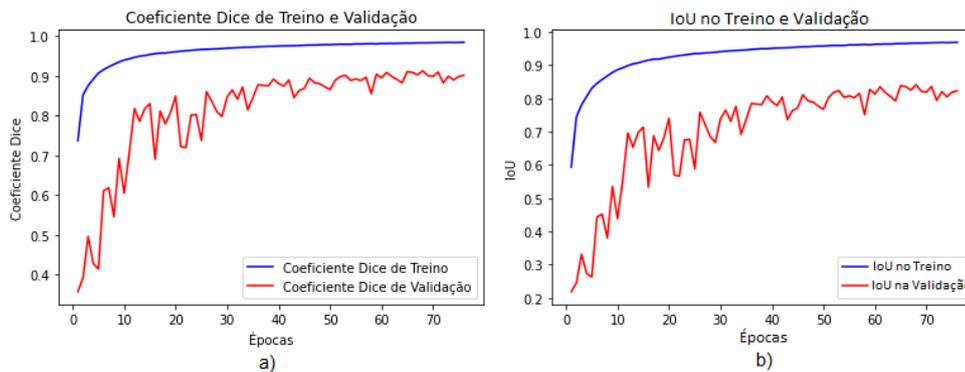
Este capítulo apresenta os resultados obtidos nesta dissertação a partir das abordagens com as duas redes para codificação. Os resultados serão apresentados em seções divididas de acordo com a rede e o banco de dados utilizado. Dessa forma, cada subseção apresentará os resultados de cada banco de dados utilizando as duas redes, em que serão apresentadas as métricas e curvas de aprendizagem.

4.1 ABORDAGEM UTILIZANDO A MOBILENETV2

4.1.1 Kvasir-SEG

O primeiro teste realizado com a rede MobileNetV2 foi com o banco de dados Kvasir-SEG, que possui 1000 imagens e 1000 máscaras correspondentes; após a implementação das técnicas de aumento de dados, o total passou a ser de 7000 imagens. A Figura 22 apresenta as curvas das métricas DSC e IoU que foram monitoradas.

Figura 22 – Curvas das métricas DSC e IoU no KvasirSEG com a rede MobileNetV2.



Fonte: O autor.

Como esperado, há uma semelhança entre as curvas das duas métricas com pequenas diferenças nos valores numéricos alcançados por cada uma. Observa-se também que as curvas do treinamento possuem uma subida suavizada, isso por conta da implementação da técnica chamada *ReduceLRonPlateau*, responsável por reduzir a taxa de aprendizagem padrão sempre que a curva chegar no seu platô e a métrica deixar de ser aprimorada (LEWKOWYCZ, 2021).

A Figura 23 apresenta a performance das curvas de perda do treino e da validação.

Figura 23 – Curva da função perda de dados no treino e validação no Kvasir-SEG com a rede MobileNetV2+U-Net.



Fonte: O autor.

Após o treinamento, foi realizada a inferência no conjunto de teste, obtendo as métricas finais e as predições da rede. A Tabela 1 apresenta as métricas obtidas nos conjuntos de treino, validação e teste.

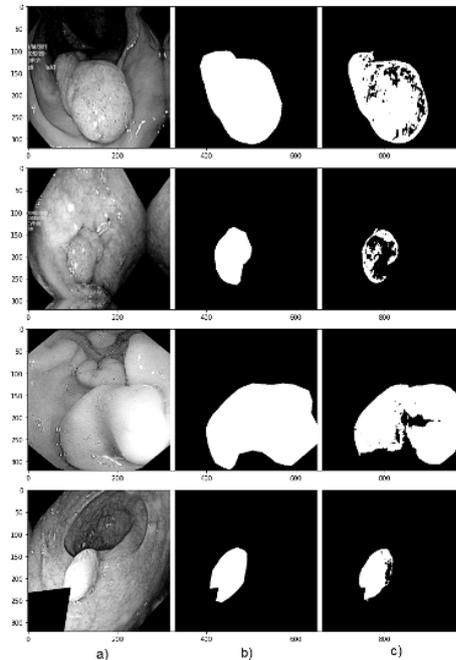
Tabela 1 – Métricas da arquitetura MobileNetV2+U-Net no banco de dados KvasirSEG.

Coeficiente DSC			IoU		
Treinamento	Validação	Teste	Treinamento	Validação	Teste
0,964	0,901	0,897	0,931	0,822	0,816

Fonte: O autor.

A partir destes valores, é possível verificar que o modelo se adaptou bem aos dados de todos os três conjuntos e não houve a ocorrência de *overfitting*. No conjunto de teste para inferência final do desempenho da rede, foram alcançados valores satisfatórios para as duas métricas. Os resultados obtidos demonstraram que as regiões segmentadas pelo modelo correspondem, em média, a 89% e 81% da região dos pólipos presentes nas máscaras originais, quando avaliadas pelo DSC e IoU, respectivamente. Além disso, foram realizadas comparações qualitativas entre as máscaras geradas pelo modelo e as máscaras originais (Figura 24).

Figura 24 – Comparação dos resultados das segmentações: a) imagens originais; b) máscaras originais; c) máscara gerada pela rede MobileNetV2+U-Net com o banco de dados Kvasir-SEG.



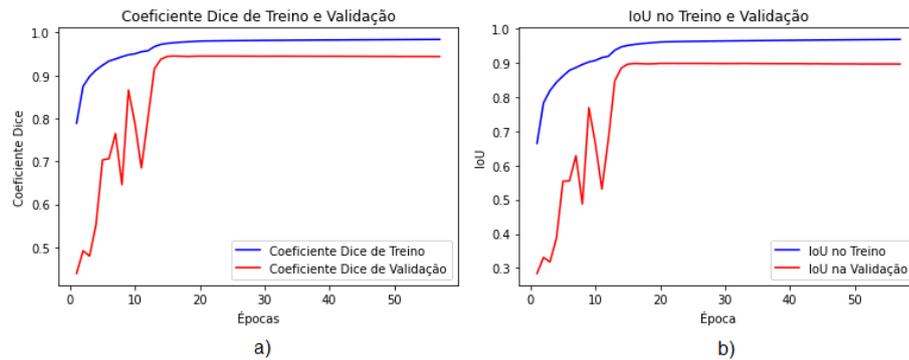
Fonte: O autor.

Como é possível observar na Figura 24, em algumas imagens há uma pequena perda de informação que acaba gerando um erro de detecção no centro ou na extremidade dos pólipos. Contudo, o modelo reconheceu muito bem os diferentes formatos e tamanhos dos pólipos, conseguindo segmentar de forma suficiente e satisfatória as anomalias presentes nas imagens e auxiliando na sua detecção automática.

4.1.2 PICCOLO

O segundo teste realizado com esta arquitetura utilizou o banco de dados PICCOLO. Foram aplicadas a 2203 imagens deste banco, técnicas de aumento de dados para que o total de imagens para o teste fosse 7000 imagens. A Figura 25 apresenta as curvas das métricas DSC e IoU obtidas neste banco.

Figura 25 – Curvas das métricas DSC e IoU no PICCOLO com a rede MobileNetV2.



Fonte: O autor.

Novamente, observa-se semelhança entre as curvas. A curva do treinamento possui uma subida mais suavizada que a da validação, o que pode ser explicado pela adaptação do modelo aos dados de treinamento; porém, encontrou algumas dificuldades no conjunto de validação, apesar de ter convergido a um valor satisfatório para as duas métricas. Para este banco de dados, é possível também notar que o modelo teve um menor tempo de processamento, alcançando o valor final em 57 épocas; a técnica *EarlyStopping* faz com que as iterações sejam interrompidas quando as métricas monitoradas deixam de evoluir por algumas épocas, evitando um possível *overfitting*. A Figura 26 apresenta a curva de perda do treino e da validação que comprovam o bom desempenho do modelo nos dois primeiros subconjuntos.

Figura 26 – Curva da função perda de dados no treino e validação no PICCOLO com a rede MobileNetV2+U-Net.



Fonte: O autor.

Foi também realizada a inferência no conjunto de teste para verificar o desempenho real do modelo em dados desconhecidos. A Tabela 2 apresenta as métricas obtidas para este banco de dados nos conjuntos de treino, validação e teste.

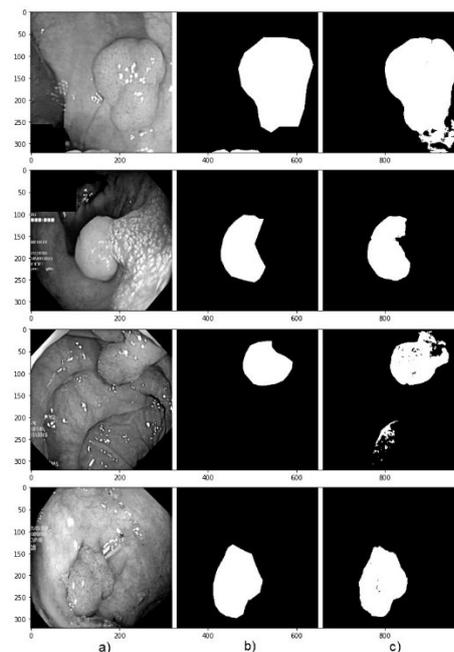
Tabela 2 – Métricas da arquitetura MobileNetV2+U-Net no banco de dados PICCOLO.

Coeficiente DSC			IoU		
Treinamento	Validação	Teste	Treinamento	Validação	Teste
0,979	0,944	0,946	0,960	0,896	0,900

Fonte: O autor.

Os valores das métricas confirmam melhor desempenho desta arquitetura para este banco de dados quando comparado ao primeiro, pois mesmo com menos iterações, obteve melhor generalização do modelo no conjunto de teste; ou seja, aprendeu melhor as características das imagens e foi capaz de gerar resultados mais confiáveis ao analisar o conjunto das imagens não vistas antes pela rede. As segmentações geradas neste teste são apresentadas na Figura 27.

Figura 27 – Comparação dos resultados das segmentações: a) imagens originais; b) máscaras originais; c) máscara gerada pela rede MobileNetV2+U-Net com o banco de dados PICCOLO.



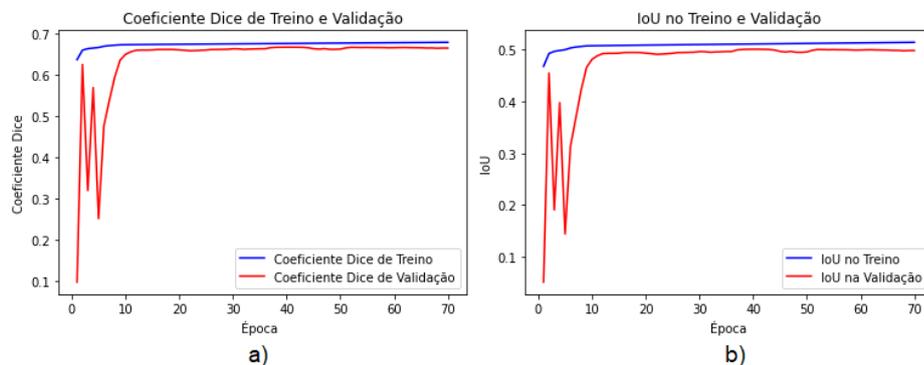
Fonte: O autor.

Diferente do que se observou nas segmentações geradas no banco de dados Kvasir-SEG, não houve perda de informações nas imagens, mas sim o entendimento do modelo de que os pólipos tinham tamanho ou formato maior do que o da segmentação geral.

4.1.3 CVC-ClinicDB

O banco CVC-ClinicDB possui 612 imagens com as máscaras segmentadas para servir de base ao aprendizado. Após a aplicação das técnicas de aumento de dados, o número de imagens disponíveis passou a ser 3672. Vale ressaltar que este banco de dados foi o único em que as imagens não puderam ser redimensionadas para o tamanho padrão de 320x320, pois algumas imagens do conjunto tinham uma das dimensões um pouco menor. Portanto, todas foram redimensionadas para 288x384. A Figura 28 apresenta as curvas das métricas DSC e IoU obtidas neste banco.

Figura 28 – Curvas das métricas DSC e IoU no CVC-ClinicDB com a rede MobileNetV2.



Fonte: O autor.

Apesar das curvas obtidas terem convergido, demonstrando uma progressiva adaptação do modelo às imagens e aprendendo informações dos dados, o resultado ao final das épocas não foi o esperado. O DSC não superou 70% de acerto e a IoU foi ligeiramente maior que 50%, valores muito abaixo do esperado quando comparados com as métricas alcançadas pela mesma rede com outros bancos de dados nesta dissertação.

A Figura 29 apresenta a curva de perda do treino e da validação, mostrando o baixo desempenho da rede com esses dados.

Figura 29 – Curva da função perda de dados no treino e validação no CVC-ClinicDB com a rede MobileNetV2+U-Net.



Fonte: O autor.

As métricas do modelo não foram satisfatórias, comprovando que a rede não se adaptou bem aos dados deste conjunto, sendo incapaz de aprender as características das imagens e aplicá-las no processo de segmentação. A Tabela 3 apresenta estas métricas.

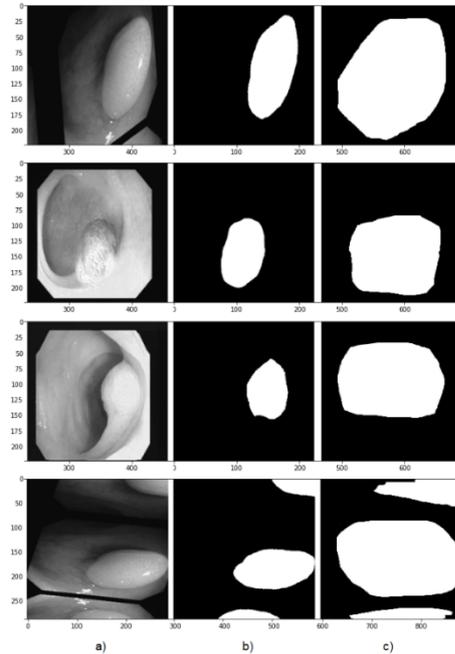
Tabela 3 – Métricas da arquitetura MobileNetV2+U-Net no banco de dados CVC-ClinicDB.

Coeficiente DSC			IoU		
Treinamento	Validação	Teste	Treinamento	Validação	Teste
0,677	0,669	0,674	0,512	0,503	0,509

Fonte: O autor.

As segmentações geradas pelo modelo são apresentadas na Figura 30 na qual fica evidenciada as interpretações incorretas que o modelo teve a partir das imagens que analisou.

Figura 30 – Comparação dos resultados das segmentações: a) imagens originais; b) máscaras originais; c) máscara gerada pela rede MobileNetV2+U-Net com o banco de dados CVC-ClinicDB.



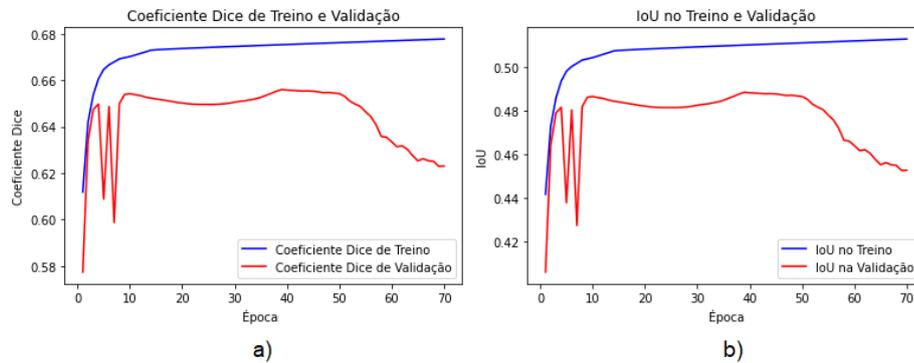
Fonte: O autor.

A partir da análise das imagens, é possível concluir que o modelo foi capaz de segmentar corretamente as áreas em que os pólipos estavam localizados. Entretanto, segmentou incorretamente muitas áreas em que não existiam pólipos e o tecido estava saudável. Dessa forma, a taxa de falsos positivos presente nas equações das métricas aumentou consideravelmente, o que depreciou ambas.

4.1.4 ETIS-LaribPolypDB

O banco de dados ETIS-LaribPolypDB possui apenas 196 imagens e suas máscaras correspondentes. A aplicação das técnicas para o aumento de dados expandiu-o para 1176 imagens. A Figura 31 apresenta as curvas das métricas DSC e IoU obtidas neste banco.

Figura 31 – Curvas das métricas DSC e IoU no ETIS-LaribPolypDB com a rede MobileNetV2+U-Net.



Fonte: O autor.

Diferentemente da tendência de comportamento observada nos outros testes, este sofreu *overfitting* e *underfitting*, exemplificado pelo comportamento das curvas de validação que após aumentarem, passaram a cair a partir da 50^a época, elevando a diferença entre as duas curvas da imagem. Os valores de métricas alcançados nas curvas de treinamento foram baixos. Já era esperado um baixo desempenho em decorrência do pequeno número de amostras disponíveis para alimentar a rede.

Figura 32 – Curva da função perda de dados no treino e validação no ETIS-LaribPolypDB com a rede MobileNetV2+U-Net.



Fonte: O autor.

O mal desempenho do modelo neste teste foi ratificado com a obtenção e análise das métricas monitoradas durante o processo (Tabela 4).

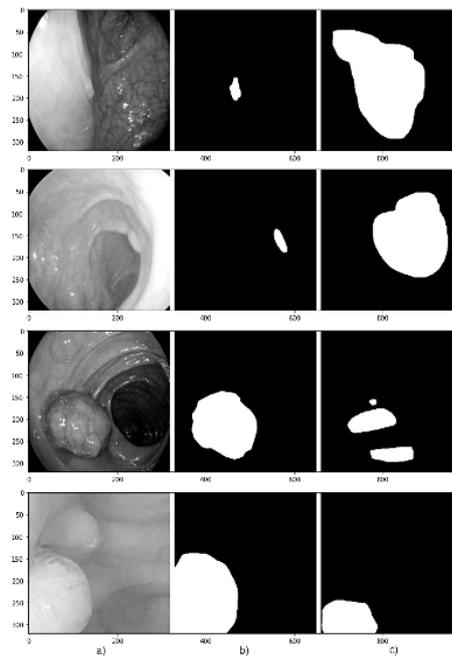
Tabela 4 – Métricas da arquitetura MobileNet+U-Net no banco de dados ETIS-LaribPolypDB.

Coeficiente DSC			IoU		
Treinamento	Validação	Teste	Treinamento	Validação	Teste
0,677	0,669	0,674	0,5126	0,503	0,509

Fonte: O autor.

Pelo baixo valor obtido pelas métricas monitoradas, era esperado que as segmentações tivessem grande quantidade de erros e falsos positivos e negativos (Figura 33).

Figura 33 – Comparação dos resultados das segmentações: a) imagens originais; b) máscaras originais; c) máscara gerada pela rede MobileNetV2+U-Net com o banco de dados ETIS-LaribPolypDB.



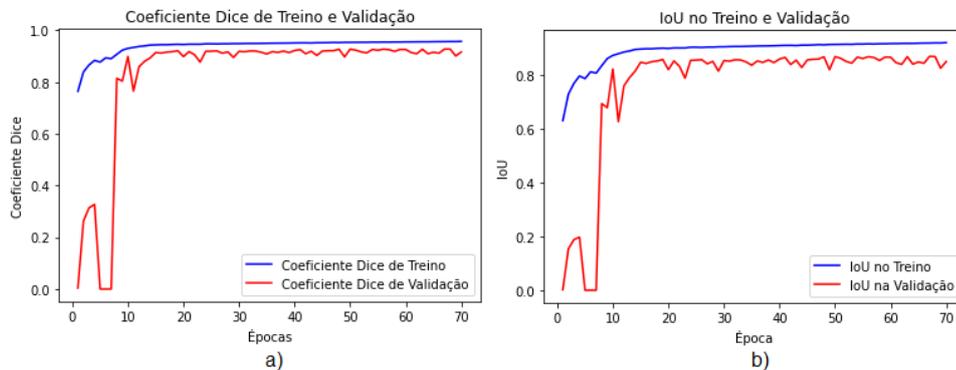
Fonte: O autor.

4.2 ABORDAGEM UTILIZANDO A EFFICIENTNETB0

4.2.1 Kvasir-SEG

Os resultados obtidos para esta rede também foram satisfatórios. A Figura 34 apresenta as curvas das métricas DSC e IoU para esta arquitetura.

Figura 34 – Curvas das métricas DSC e IoU no KvasirSEG com a rede EfficientNetB0+U-Net.



Fonte: O autor.

As curvas possuem semelhança e a curva do treinamento possui uma subida mais suavizada do que a da validação, o que pode ser explicado pela boa adaptação do modelo aos dados de treinamento; porém, encontrou algumas dificuldades no conjunto de validação. Observa-se também que nas primeiras iterações; a curva de validação sofreu grandes variações, o que pode ser explicado pela alta taxa de aprendizado inicial, levando o modelo a não encontrar o mínimo global. No decorrer das iterações, essas variações continuaram, porém de forma atenuada, mostrando que a aplicação da técnica *ReduceLRonPlateau* auxiliou o modelo a reduzir essas variações e se ajustar aos dados ao longo do treinamento, alcançando os resultados satisfatórios ao final. A Figura 35 apresenta a curva de perda do treino e da validação, comprovando o bom desempenho deste teste.

Figura 35 – Curva da função perda de dados no treino e validação no Kvasir-SEG com a rede EfficientNetB0+U-Net.



Fonte: O autor.

Repetindo os passos das outras simulações, o desempenho do modelo foi inferido no conjunto de teste para verificar a generalização do modelo e sua capacidade de segmentar os pólipos. A Tabela 5 apresenta as métricas.

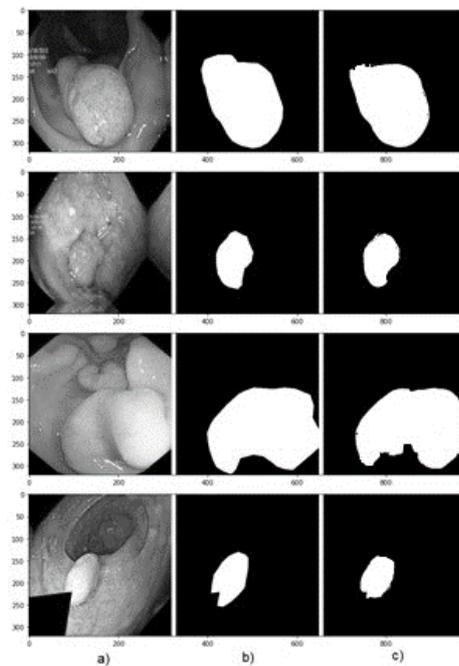
Tabela 5 – Métricas da arquitetura EfficientNet+U-Net no banco de dados Kvasir-SEG.

Coeficiente DSC			IoU		
Treinamento	Validação	Teste	Treinamento	Validação	Teste
0,955	0,925	0,910	0,915	0,864	0,838

Fonte: O autor.

Os valores alcançados pelas métricas ratificam que este modelo teve um bom desempenho sem apresentar *overfitting*. A capacidade do modelo de segmentar corretamente as imagens analisadas foi inferida pelo conjunto de teste, em que as métricas DSC e IoU tiveram valores de, respectivamente, 91% e 83% (Tabela 5). As comparações qualitativas para este modelo são apresentadas na Figura 36.

Figura 36 – Comparação dos resultados das segmentações: a) imagens originais; b) máscaras originais; c) máscara gerada pela rede EfficientNetB0+U-net com o banco de dados Kvasir-SEG.

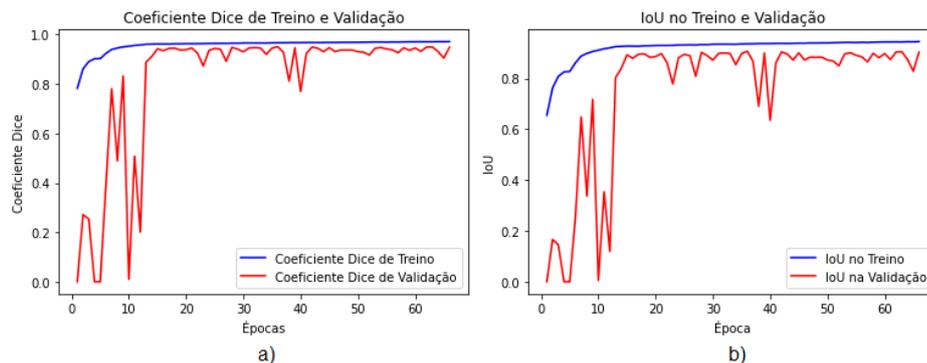


Fonte: O autor.

4.2.2 Piccolo

A rede EfficientNetB0 foi testada para banco de dados PICCOLO. A Figura 37 apresenta as curvas das métricas DSC e IoU obtidas.

Figura 37 – Curvas das métricas DSC e IoU no PICCOLO com a rede EfficientNetB0+U-Net.



Fonte: O autor.

A partir da análise destes resultados, é possível inferir que, até a época 12, o modelo não estava tendo êxito em se adaptar aos dados analisados e aprender sobre estes, pois as métricas estavam variando muito entre valores altos e baixos. Este comportamento pode ser explicado pela alta taxa de aprendizado inicial que pode atrapalhar a rede a encontrar o valor ótimo das métricas. Contudo, a partir da época 15, essas variações foram atenuadas pela técnica *ReduceLRonPlateau*. A Figura 38 apresenta a curva de perda do treino e da validação.

Figura 38 – Curva da função perda de dados no treino e validação no PICCOLO com a rede EfficientNetB0+U-Net.



Fonte: O autor.

A Tabela 6 a seguir expõe as métricas finais obtidas após o treinamento, validação e teste neste banco de dados.

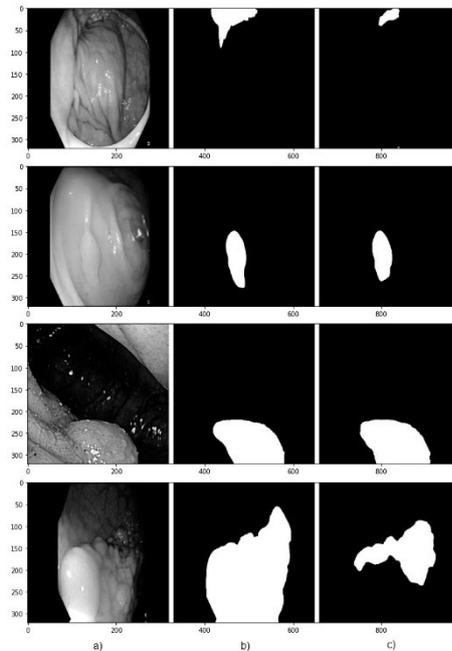
Tabela 6 – Métricas da arquitetura EfficientNet+U-Net no banco de dados PICCOLO.

Coeficiente DSC			IoU		
Treinamento	Validação	Teste	Treinamento	Validação	Teste
0,971	0,948	0,941	0,945	0,903	0,892

Fonte: O autor.

Os valores obtidos no conjunto de teste para o DSC (94%) e IoU (89%) mostram que o modelo teve desempenho muito bom na segmentação das imagens analisadas apesar das grandes variações iniciais observadas nas Figuras 37 e 38. O modelo não sofreu *overfitting* e nem *underfitting*. As imagens de saída da rede após este teste são apresentadas na Figura 39.

Figura 39 – Comparação dos resultados das segmentações: a) imagens originais; b) máscaras originais; c) máscara gerada pela rede EfficientNetB0+U-Net com o banco de dados PICCOLO.



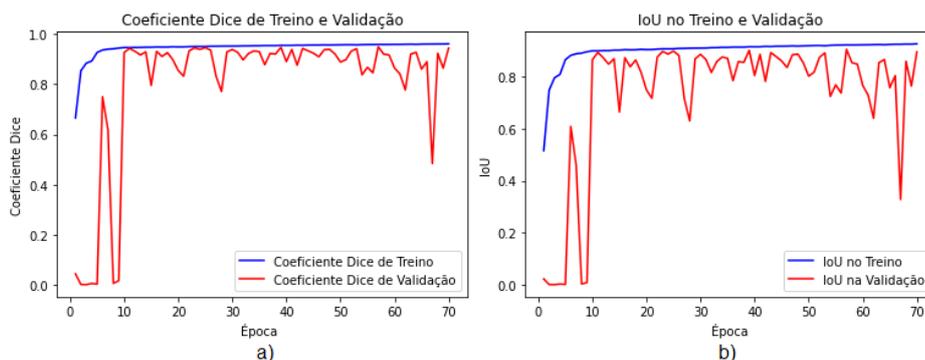
Fonte: O autor.

Apesar das métricas satisfatórias, não são todas as segmentações que apresentaram boa taxa de acerto, como é possível observar nas primeira e quarta amostras. Entretanto, as amostras com boa segmentação foram maioria, não permitindo maior depreciação das métricas.

4.2.3 CVC-ClinicDB

A Figura 40 apresenta as curvas das métricas DSC e IoU geradas pelo modelo utilizado.

Figura 40 – Curvas das métricas DSC e IoU no CVC-ClinicDB com a rede EfficientNetB0+U-Net.



Fonte: O autor.

Apesar do bom desempenho no conjunto de treinamento, o modelo teve bastante dificuldade para avaliar os dados do conjunto de validação e sofreu muitas variações.

Figura 41 – Curva da função perda de dados no treino e validação no CVC-ClinicDB com a rede EfficientNetB0+U-Net.



Fonte: O autor.

A Tabela 7 resume as métricas finais obtidas após o treinamento, validação e teste neste banco de dados.

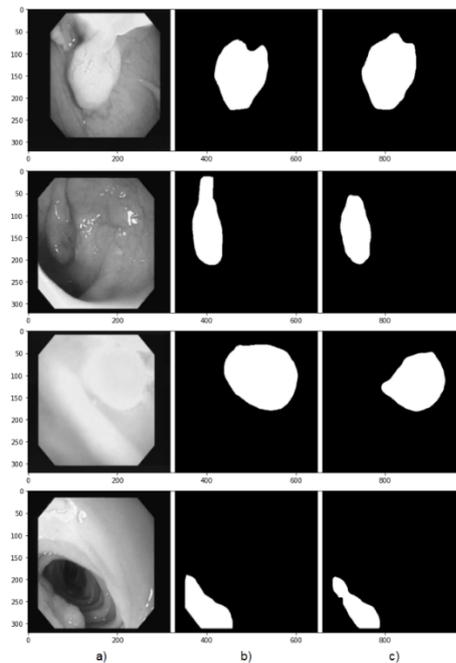
Tabela 7 – Métricas da arquitetura EfficientNet+U-Net no banco de dados CVC-ClinicDB.

Coeficiente DSC			IoU		
Treinamento	Validação	Teste	Treinamento	Validação	Teste
0,952	0,941	0,920	0,937	0,929	0,891

Fonte: O autor.

Ao final do treinamento, os valores atingidos nos três conjuntos foram bastante satisfatórios, mostrando bom desempenho da rede neste conjunto de dados. Ao inferir a capacidade do modelo nas imagens de teste, o DSC foi de 92% e o IoU de 89% mesmo com a variação negativa nas épocas iniciais. As imagens de saída da rede após este teste são apresentadas na Figura 42.

Figura 42 – Comparação dos resultados das segmentações: a) imagens originais; b) máscaras originais; c) máscara gerada pela rede EfficientNetB0+U-Net com o banco de dados CVC-ClinicDB.

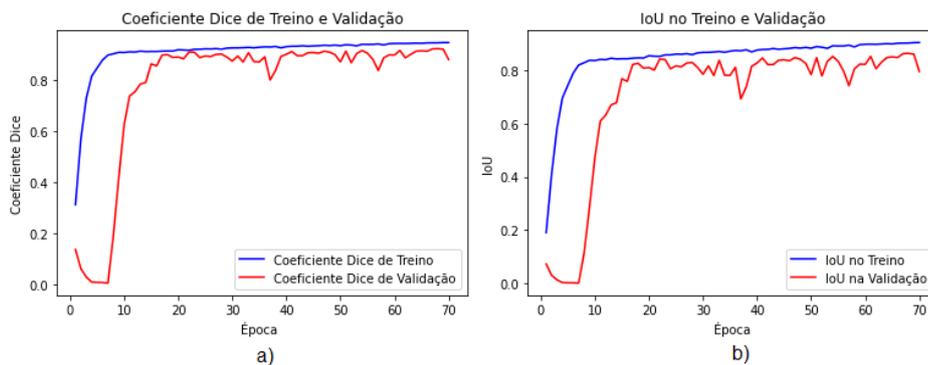


Fonte: O autor.

4.2.4 ETIS-LaribPolypDB

A Figura 43 apresenta as curvas das métricas DSC e IoU obtidas nesta variação de rede e dados.

Figura 43 – Curvas das métricas DSC e IoU no ETIS-LaribPolypDB com a rede EfficientNetB0+U-Net.



Fonte: O autor.

Observa-se um comportamento esperado pelo pequeno número de imagens divididas entre os subconjuntos de treino, validação e teste. O modelo teve bastante dificuldade para extrair características das imagens no início do treinamento e aplicar este conhecimento na validação, visto que as curvas das duas métricas tiveram seus valores iniciais muito menores do que os iniciais dos testes anteriores. Apesar disso, as técnicas que melhoram a taxa de aprendizado monitorando as métricas fez com que o modelo se ajustasse melhor aos dados a partir da décima época (Figura 44).

Figura 44 – Curva da função perda de dados no treino e validação no ETIS-LaribPolypDB com a rede EfficientNetB0+U-Net.



Fonte: O autor.

A Tabela 8 apresenta as métricas obtidas após as etapas de treinamento, validação e teste com este banco de dados.

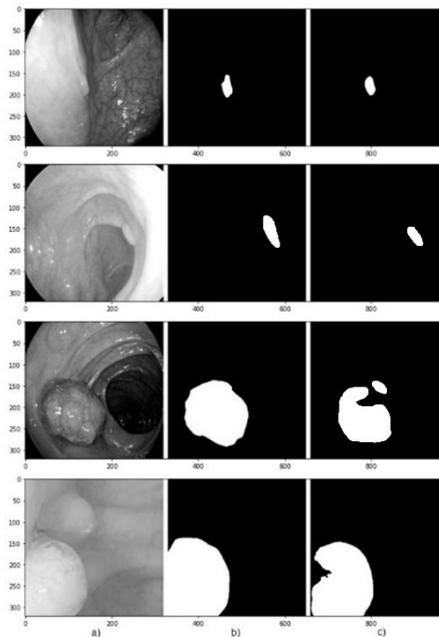
Tabela 8 – Métricas da arquitetura EfficientNet+U-Net no banco de dados ETIS-LaribPolypDB.

Coeficiente DSC			IoU		
Treinamento	Validação	Teste	Treinamento	Validação	Teste
0,948	0,882	0,911	0,902	0,794	0,839

Fonte: O autor.

Os valores da Tabela 8 confirmam que as métricas atingidas nesta simulação foram menores que as anteriores nas etapas de treinamento e validação; isto se deve ao pequeno número de dados inseridos na rede. Apesar de serem menores que os demais, o DSC e o IoU atingiram valores bons para a tarefa ao se testar o modelo no conjunto de dados desconhecidos, sendo 91% e 83%, respectivamente. As imagens geradas pelo modelo são apresentadas na Figura 45.

Figura 45 – Comparação dos resultados das segmentações: a) imagens originais; b) máscaras originais; c) máscara gerada pela rede EfficientNetB0+U-Net com o banco de dados ETIS-LaribPolypDB.



Fonte: O autor.

4.3 COMPARAÇÃO ENTRE AS ABORDAGENS

Após a realização de todos os experimentos, foi possível notar que a rede MobileNetV2 apresentou diversas deficiências em relação ao desempenho da EfficientNetB0.

Enquanto a EfficientNetB0 se ajustou muito bem a todos os bancos de dados utilizados nesta dissertação, aprendendo características e alcançando boas métricas e segmentações, a MobileNetV2 não teve a mesma capacidade quando teve de enfrentar dois problemas: o pequeno número de imagens fornecidas pelo conjunto mesmo após a aplicação de técnicas de aumento de dados e imagens com resoluções menores, o que acaba por alimentar o modelo com imagens com menos informações e a rede não consegue aprender de maneira suficiente para que consiga realizar as segmentações necessárias. Sendo assim, a EfficientNetB0 mostrou-se mais promissora que a outra rede para implementações como codificador da U-Net ou alguma outra rede para tarefas de segmentação.

4.4 COMPARAÇÃO COM A LITERATURA

Esta seção realiza uma comparação quantitativa dos métodos propostos nesta pesquisa com outros métodos de segmentação de pólipos existentes no estado da arte. Contudo, realizar tal comparação é algo desafiador devido alguns fatores.

Primeiramente, os detalhes de implementação de alguns trabalhos não são fornecidos, por exemplo, hiperparâmetros, número de amostras de imagens fornecidas à rede. Logo, é difícil garantir a reprodução dos resultados tal que se possa compará-los.. Outros autores utilizam, no conjunto de testes, alguns dados que já foram utilizados na etapa de treinamento, o que influencia negativamente no desempenho. Muitas pesquisas utilizam métricas diferentes, o que impede uma comparação exata. Portanto, a comparação inclui apenas trabalhos de segmentação de pólipos que utilizaram os índices DSC ou IoU como métricas. Como a rede EfficientNetB0 foi a que apresentou os melhores resultados em todos os bancos de dados utilizados, ela será a única a ser comparada com os demais estudos.

As Tabelas 9 e 10 comparam nossos resultados com métodos que utilizaram os bancos de dados CVC-ClinicDB e Kvasir-SEG, respectivamente. Os resultados apresentados mostram que nosso modelo supera a maioria das outras pesquisas quando comparamos as métricas. No CVC-ClinicDB, nosso teste com a rede EfficientNetB0 como codificador da U-Net atinge um coeficiente DSC de 92% e um IoU de 89,1% (Tabela 9).

Tabela 9 – Comparação das métricas obtidas no CVC-ClinicDB com a rede EfficientNet+U-Net com outras pesquisas.

Método	Coefficiente DSC	IoU
Yeung et al. (2021)	0,938	0,889
Fan D.-P. et al. (2020)	0,899	0,849
Jha et al. (2019)	0,795	0,796
EfficientNetB0 + U-Net	0,920	0,891

Fonte: O autor.

Uma comparação entre nosso método e o modelo proposto por Yeung et al. (2021) mostra que nosso método obteve valor ligeiramente inferior para a métrica DSC de 92% em comparação a 93,8%, porém teve melhor desempenho na segunda métrica. A razão para isso pode ser o fato de que o autor utilizou uma rede mais robusta e teve maior tempo de processamento. Ao compararmos o método escolhido nesta pesquisa com os estudos apresentados por D.-P. Fan et al. (2020) e Jha et al. (2019), observa-se a superioridade do nosso método em 2,1% e 12,5% para o DSC e 4,2% e 9,5%, respectivamente (Tabela 9).

Realizando a comparação entre as pesquisas em outro conjunto de dados, no Kvasir-SEG, nosso modelo foi capaz de obter um coeficiente DSC de 91% e um IoU de 83,8%. Novamente, ao comparar com outros estudos, nosso modelo teve performance superior na maioria (Tabela 10).

Tabela 10 – Comparação das métricas obtidas no Kvasir-SEG com a rede EfficientNet+U-Net com outras pesquisas.

Método	Coefficiente DSC	IoU
Yeung et al. (2021)	0,910	0,853
Fan D.-P. et al. (2020)	0,898	0,840
Jha et al. (2019)	0,813	0,792
EfficientNetB0 + U-Net	0,910	0,838

Fonte: O autor.

Com base em uma comparação aos outros modelos em termos de métricas de desempenho, podemos ver que as abordagens propostas nesta dissertação foram capazes de

detectar maior taxa de verdadeiros positivos e segmentou menos áreas com falsos negativos. Apesar do método proposto ter tido desempenho inferior em 1,5% quando comparada a métrica IoU com a pesquisa de Yeung et al. (2021), o restante obtido foi satisfatório e bastante competitivo, com boas margens de superação.

Para a tarefa de segmentação de pólipos, é fundamental melhorar a taxa de segmentação, pois os modelos são sensíveis a alterações bruscas e isso é um fator primordial que influencia diretamente no desempenho das redes.

5 CONCLUSÃO

Nesta dissertação, foram implementadas redes neurais para realizar a segmentação semântica de pólipos em imagens de colonoscopia, em que se utilizou três redes, sendo a U-Net como rede principal e a MobileNetV2 e a EfficientNetB0 como codificadores da U-Net a fim de realizar testes de desempenho, aplicando técnicas de processamento digital de imagens para melhorar a eficiência do algoritmo, que foi um fator significativo para obter os melhores resultados.

A construção deste trabalho possibilitou explorar e analisar o comportamento e o treinamento das redes ao serem unidas como codificador e decodificador, ao realizar a personalização e escolha dos hiperparâmetros em busca das melhores métricas.

Com a realização de todos os testes, que permitiram comparar as duas redes utilizadas entre si e também com outras redes e bancos de dados presentes na literatura, foi possível concluir que os resultados obtidos foram satisfatórios e promissores, pois apresentaram melhorias nas métricas monitoradas e nas segmentações geradas. Apesar das limitações computacionais enfrentadas pela pesquisa, os modelos foram capazes de realizar segmentações com qualidades superiores com menor complexidade quando comparadas a outras pesquisas existentes na literatura, condição que era buscada desde o início desta pesquisa.

Para os trabalhos futuros, recomenda-se a realização de treinamentos com novos bancos de dados e com as redes que já foram testadas neste trabalho, mas também a implementação de novas redes principais ou como codificadores da U-Net, a fim de possuir uma maior gama de resultados para comparações futuras, bem como de obter melhores métricas de segmentação e um modelo cada vez mais confiável. Além disso, seria interessante aperfeiçoar os algoritmos para reduzir tempo de treinamento em máquinas com maior capacidade computacional, sendo possível realizar diversos testes em menos tempo e com modelos mais robustos capazes de entregar resultados cada vez melhores.

REFERÊNCIAS

- ADATRAO, S.; MITTAL, M. An analysis of different image preprocessing techniques for determining the centroids of circular marks using hough transform. **International Conference on Frontiers of Signal Processing (ICFSP)**, 2016.
- AGGARWAL, C. C. **Neural Networks and Deep Learning: A Textbook**. New York: Springer, p. 497, 2018.
- AGÊNCIA INTERNACIONAL DE PESQUISA EM CÂNCER. Ficha informativa sobre câncer colorretal. **Agência Internacional de Pesquisa em Câncer**, 2018.
- AHMED, M.; SERAJ, R.; ISLAM, S. M. S. The k-means Algorithm: A Comprehensive Survey and Performance Evaluation. **Electronics**, v. 9, n. 8, p. 1295, 2020.
- AHN, H.; YIM, C. Convolutional Neural Networks Using Skip Connections with Layer Groups for Super-Resolution Image Reconstruction Based on Deep Learning. **Applied Sciences**, v. 10, n. 6, p. 1959, 2020.
- AKBARI, M.; MOHREKESH, M.; NARS-ESFAHANI, E.; SOROUSHMEHR, S.M.R.; KARIMI, N.; SAMAVI, S.; NAJARIAN, K., Polyp Segmentation in Colonoscopy Images Using Fully Convolutional Network. **IEEE Engineering in Medicine and Biology**, p. 69-72, 2018.
- ALAM, S.; TOMAR, N. K.; THAKUR, A.; JHA, D.; RAUNIYAR, A. Automatic Polyp Segmentation using U-Net-ResNet50. **Mediaeval Medico 2020 Automatic Polyp Segmentation Challenge**, 2020.
- ALLOGHANI, M.; AL-JUMEILY, D.; MUSTAFINA, J.; HUSSAIN, A.; ALJAAF, A. J. **A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science**, SPRINGER, p. 3-21, 2020.
- ALZUBAIDI, L.; FADHEL, M. A.; AL-SHAMMA, O.; ZHANG, J.; DUAN, Y. Deep Learning Models for Classification of Red Blood Cells in Microscopy Images to Aid in Sickle Cell Anemia Diagnosis. **Electronics**, v. 9, n. 3, p. 427, 2020.
- AMELING, S.; WIRTH, S.; PAULUS, D.; LACEY, G.; VILARINO, F. **Texture-Based Polyp Detection in Colonoscopy**, SPRINGER, p. 346, 2009.
- AMERICAN CANCER SOCIETY. **Colorectal Cancer Stages**. 2021. Disponível em: <https://www.cancer.org/cancer/colon-rectal-cancer/detection-diagnosisstaging/staged.html>. Acesso em: 03 jan. 2022.
- ASHRAF, S. M.; CHAKRABARTI, S. Voltage stability monitoring of power systems using reduced network and artificial neural network. **International Journal of Electrical Power & Energy Systems**, v. 87, p. 43–51, 2017.

BARROWCLOUGH, O. J. D.; MUNTINGH, G.; NAINAMALAI, V.; STANGEBY, I. Binary segmentation of medical images using implicit spline representations and deep learning. **Computer Aided Geometric Design**, v. 85, p. 101972, 2021.

BERNAL, J.; SÁNCHEZ, F. J.; FERNÁNDEZ-ESPARRACH, G.; GIL, D.; RODRIGUEZ, C.; VILARIÑO, F. WM-DOVA. Maps for Accurate Polyp Highlighting in Colonoscopy: Validation Vs. Saliency Maps from Physicians. **Computerized Medical Imaging and Graphics**, v. 43, p. 99–111, 2015.

BRAY, F.; FERLAY, J.; SOERJOMATARAM, I.; SIEGEL, R. L.; TORRE, L. A.; JEMAL, A. Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. **A Cancer Journal for Clinicians**, v. 68, n. 6, p. 394–424, 2018.

BRANDAO, P.; MAZOMENOS, E. B.; CIUTI, G.; CALIÒ, R.; BIANCHI, F.; MENCIASSI, A.; DARIO, P.; KOULAOUZIDIS, A.; AREZZO, A.; STOYANOV, D. Fully convolutional neural networks for polyp segmentation in colonoscopy. **Medical Imaging: Computer-Aided Diagnosis**, 2017.

CHAO, P.; KAO, C-Y.; RUAN, Y-S.; HUANG, C-H.; LIN, Y-L. HarDNet: A Low Memory Traffic Network. **IEEE International Conference on Computer Vision (ICCV)**, p. 3551-3560, 2019.

CHEPLYGINA, V.; DE BRUIJNE, M.; PLUIM, J. P. W. Not-so-supervised: A survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. **Medical Image Analysis**, v. 54, p. 280–296, 2019.

CHOLLET, F. Xception: Deep Learning with Depthwise Separable Convolutions. **IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**, 2017.

DENG, J.; DONG, W.; SOCHER, R.; LI, L-J.; LI, K.; FEI-FEI, L. ImageNet: A large-scale hierarchical image database. **IEEE Conference on Computer Vision and Pattern Recognition**, 2009.

DHANDE, G.; SHAIKH, Z. Analysis of Epochs in Environment based Neural Networks Speech Recognition System. **International Conference on Trends in Electronics (ICOEI)**, p. 605-608, 2019.

DIAKOGIANNIS, F. I.; WALDNER, F.; CACCETTA, P.; WU, C. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 162, p. 94–114, 2020.

DOUBENI, C. A.; CORLEY, D. A.; QUINN, V. P.; JENSEN, C. D.; ZAUBER, A. G.; GOODMAN, M.; JOHNSON, J. R.; MEHTA, S. J.; BECERRA, T. A.; ZHAO, W. K.; SCHOTTINGER, J.; DORIA-ROSE, V. P.; LEVIN, T. R.; WEISS, N. S.; FLETCHER, R. H. Effectiveness of Screening Colonoscopy in Reducing the Risk of Death from Right and Left Colon Cancer: A Large Community-Based Study. **Gut**, v. 67, n. 2, p. 291–298, 2018.

ERAZO, J. J. M. **Desenvolvimento de um sistema de contagem e classificação de carros utilizando redes neurais convolucionais**. Dissertação (Mestrado em Engenharia de Automação e Sistemas) – Universidade Federal de Santa Catarina, Florianópolis, 2021.

FAN, D.-P.; JI, G.-P.; ZHOU, T.; CHEN, G.; FU, H.; SHEN, J. SHAO, L. PraNet: Parallel Reverse Attention Network for Polyp Segmentation. **Medical Image Computing Assisted Intervention (MICCAI)**, p. 263-273, 2020.

FANG, Y.; ZHU, D.; YAO, J.; YUAN, Y.; TONG, K-Y. ABC-Net: Area-Boundary Constraint Network with Dynamical Feature Selection for Colorectal Polyp Segmentation. **IEEE Sensors Journal**, v. 21, n. 10, p. 11799–11809, 2020.

FANG, Y.; CHEN, C.; YUAN, Y. TONG, K-Y. Selective feature aggregation network with area-boundary constraints for polyp segmentation. **Medical Image Computing and Computer Assisted Intervention (MICCAI)**, p. 302-310, 2019.

GANZ, M.; XIAOYUN YANG, NULL; SLABAUGH, G. Automatic segmentation of polyps in colonoscopic narrow-band imaging data. **IEEE Transactions on Bio-Medical Engineering**, v. 59, n. 8, p. 2144–2151, 2012.

GARCIA-GARCIA, A.; ORTS-ESCOLANO, S.; OPREA, S.; VILLENA-MARTINEZ, V.; MARTINEZ-GONZALEZ, P. GARCIA-RODRIGUEZ, J. A Review on Deep Learning Techniques Applied to Semantic Segmentation. **Applied Soft Computing**, v. 70, p. 41-65, 2017

GASTROCLÍNICA FLORIANÓPOLIS. **Colonoscopia: O que você precisa saber sobre este exame**. 2020. Disponível em: <http://www.gastroflorianopolis.com.br/arquivos-novidades/colonoscopia-2/>. Acesso em: 05 jan. 2022.

GERSTNER, W.; KISTLER, W. M.; NAUD, R.; PANINSKI, L. **Neuronal Dynamics: From Single Neurons to Networks and Models of Cognition**. CAMBRIDGE UNIVERSITY PRESS, 2014.

GONZALES, R. C., WOODS, R. E. **Processamento de imagens digitais**. São Paulo: PEARSON PRENTICE HALL, 2010.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. USA: MIT Press, 2016.

GURUNG, A.; TAMANG, S. L. **Image Segmentation using Multi-Threshold technique by Histogram Sampling**, 2019.

HARI, D. M.; LEUNG, A. M.; LEE, J.-H.; SIM, M.-S.; VUONG, B.; CHIU, C. G.; BILCHIK, A. J. AJCC-7TH Edition Staging Criteria for Colon Cancer: Do the Complex Modifications Improve Prognostic Assessment? **Journal of the American College of Surgeons**, v. 217, n. 2, p. 181–190, 2013.

HASSINGER, J. P.; HOLUBAR, S. D.; PENDLIMARI, R.; DOZOIS, E. J.; LARSON, D. W.; CIMA, R. R. Effectiveness of a multimedia-based educational intervention for improving

colon cancer literacy in screening colonoscopy patients. **Diseases of the Colon and Rectum**, v. 53, n. 9, p. 1301–1307, 2010.

HAYKIN, S. S. **Neural Networks and Learning Machines**, PEARSON EDUCATION, 2009.

HERNÁNDEZ-GARCÍA, A.; KÖNIG, P. **Further advantages of data augmentation on convolutional neural networks**, LECTURE NOTES IN COMPUTER SCIENCE, v. 11139, p. 95–103.

HERTZ, J.; KROGH, A.; PALMER, R. G. **Introduction to the Theory of Neural Computation**. CRC PRESS, 2019.

HOWARD, A. G.; ZHU, M.; CHEN, B.; KALENICHENKO, D.; WANG, W.; WEYAND, T.; ANDREETTO, M.; ADAM, H. **MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications**, 2017.

HU, B.; TANG, J.; WU, J.; QING, J. An Attention EfficientNet-Based Strategy for Bearing Fault Diagnosis under Strong Noise. **Sensors**, v. 22, n. 17, p. 6570, 2022.

HUANG, C.-H.; WU, H.-Y.; LIN, Y.-L. HarDNet-MSEG: A Simple Encoder-Decoder Polyp Segmentation Neural Network that Achieves over 0.9 Mean Dice and 86 FPS, **DBLP Journals**, 2021.

HWANG, S.; OH, J.; TAVANAPONG, W.; WONG, J. Polyp detection in colonoscopy video using elliptical shape feature: **IEEE International Conference on Image Processing (ICIP)**. p. 465–468, 2007.

JHA, D.; SMEDSRUD, P. H.; RIEGLER, M. A.; HALVORSEN, P.; LANGE, T.; JOHANSEN, D.; JOHANSER, H. D. Kvasir-SEG: A Segmented Polyp Dataset. **International Conference on Multimedia Modeling**, 2019.

JHA, D.; SMEDSRUD, P. H.; RIEGLER, M. A.; HALVORSEN, P.; LANGE, T.; JOHANSEN, D.; JOHANSER, H. D. ResUNet++: An Advanced Architecture for Medical Image Segmentation. **IEEE International Symposium on Multimedia (ISM)**, pp. 225, 2019.

JIANG, F.; GRIGOREV, A.; RHO, S.; TIAN, Z.; FU, Y.; JIFARA, W.; ADIL, K.; LIU, S. Medical Image Semantic Segmentation Based On Deep Learning. **Neural Computing and Applications**, v. 29, n. 5, p. 1257–1265, 2018.

JORDAN, M. I.; MITCHELL, T. M. Machine learning: Trends, perspectives, and prospects. **Science**, v. 349, n. 6245, p. 255–260, 2015.

JYOTHI, S.; K.BHARGAVI. A Survey on Threshold Based Segmentation Technique in Image Processing, **International Journal of Innovative Research and Development**, 2014.

KANDEL, I.; CASTELLI, M. The Effect of Batch Size on the Generalizability of the Convolutional Neural Networks on a Histopathology Dataset. **ICT Express**, v. 6, n. 4, p. 312–315, 2020.

KARIMI, D.; WARFIELD, S. K.; GHOLIPOUR, A. Transfer Learning in Medical Image Segmentation: New Insights from Analysis of the Dynamics of Model Parameters and Learned Representations. **Artificial Intelligence in Medicine**, v. 116, p. 102078, 2021.

KERAS. Keras applications. <https://keras.io/applications/>, 2015.

KESKAR, N. S.; MUDIGERE, D.; NOCEDAL, J.; SMELYANSKIY, M.; TANG, P. T. P. On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima. **International Conference on Learning Representations (ICLR)**, 2017.

KINGMA, D. P.; BA, J. Adam: A Method for Stochastic Optimization, 2014.

KOTSIANTIS, S. B.; ZAHARAKIS, I. D.; PINTELAS, P. E. Machine Learning: A Review of Classification and Combining Techniques. **Artificial Intelligence Review**, v. 26, n. 3, p. 159–190, 2006.

KUKACKA, J.; GOLKOV, V.; CREMERS, D. Regularization for Deep Learning: A Taxonomy. **International Conference on Learning Representations (ICLR)**, 2017.

LEE, S. B.; KIM, H.; KIM, J.; CHUNG, I. Y.; KIM, H. J.; KO, B. S.; LEE, J. W.; AHN, S. H.; SON, B. H. Prognosis according to clinical and pathologic lymph node status in breast cancer patients who underwent sentinel lymph node biopsy alone after neoadjuvant therapy. **PLoS One**, v. 16, n. 5, p. e0251597, 2021.

LEE, T. J. W.; NAIR, S.; BENTARIS, I.; RUTTER, M. D. Recent advances in colonoscopy. **F1000Research**, v. 5, 2016.

LEWKOWYCZ, A. **How to decay your learning rate**, 2021.

LI, S.; SULTONOV, F.; YE, Q.; BAI, Y.; PARK, J-H.; YANG, C.; SONG, M.; KOO, S.; KANG, J-M. TA-Unet: Integrating Triplet Attention Module for Drivable Road Region Segmentation. **Sensors**, v. 22, n. 12, p. 4438, 2022.

LIU, J.; CHEN, Y.; LAN, L.; LIN, B.; CHEN, W.; WANG, M.; LI, R.; YANG, Y.; ZHAO, B.; HU, Z.; DUAN, Y. Prediction of Rupture Risk in Anterior Communicating Artery Aneurysms with a Feed-Forward Artificial Neural Network. **European Radiology**, v. 28, n. 8, p. 3268–3275, 2018.

LIU, X.; SONG, L.; LIU, S.; ZHANG, Y. A Review of Deep-Learning-Based Medical Image Segmentation Methods. **Sustainability**, v. 13, n. 3, p. 1224, 2021.

MASSUCATTO, J. D. P. **Aplicação de conceitos de redes neurais convolucionais na classificação de imagens de folhas**. Trabalho de Conclusão de Curso (Bacharel em Engenharia da Computação) – Universidade Tecnológica Federal do Paraná, Pato Branco, 2018.

MESTHA, L. K.; DIANAT, S. A. **Control of Color Imaging Systems: Analysis and Design**. CRC PRESS, 696 p., 2009.

MICLEA, V.-C.; NEDEVSCI, S. Real-Time Semantic Segmentation-Based Stereo Reconstruction. **IEEE Transactions on Intelligent Transportation Systems**, v. 21, n. 4, p. 1514–1524, 2020.

MONTEIRO, M.; NEWCOMBE, V. F. J.; MATHIEU, F.; ADATIA, K.; KAMNITSAS, K.; FERRANTE, E.; DAS, T.; WHITEHOUSE, D.; RUECKERT, D.; MENON, D. K.; GLOCKER, B. Multiclass semantic segmentation and quantification of traumatic brain injury lesions on head CT using deep learning: an algorithm development and multicentre validation study. **The Lancet Digital Health**, v. 2, n. 6, p. e314–e322, 2020.

NATARAJAN, N.; DHILLON, I. S.; RAVIKUMAR, P. Learning with Noisy Labels. **Advances in Neural Information Processing Systems**, 2013.

NIRTHIKA, R.; MANIVANNAN, S.; RAMANAN, A.; WANG, R. Pooling in convolutional neural networks for medical image analysis: a survey and an empirical study. **Neural Computing and Applications**, v. 34, n. 7, p. 5321–5347, 2022.

OBARO, A. E.; BURLING, D. N.; PLUMB, A. A. Colon cancer screening with CT colonography: logistics, cost-effectiveness, efficiency and progress. **The British Journal of Radiology**, v. 91, n. 1090, p. 20180307, 2018.

ORGANIZAÇÃO PANAMERICANA DE SAÚDE (OPAS). **Câncer**, 2019. Disponível em: <https://www.paho.org/pt/topicos/cancer>. Acesso em: 12 jan. 2022.

PENTAX MEDICAL. **Colonoscopes**. 2018. Disponível em: <https://www.pentaxmedical.com/pentax/pt/100/3/Colonoscopes>. Acesso em: 05 jan. 2022.

RONNEBERGER, O.; FISCHER, P.; BROX, T. **U-Net: Convolutional Networks for Biomedical Image Segmentation**, 2015.

ROSENFELD, A. Picture Processing by Computer. **ACM Comput**, v. 1, p. 147-176, 1969.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. **Nature**, v. 323, n. 6088, p. 533–536, 1986.

SALMON, B. P.; KLEYNHANS, W.; SCHWEGMANN, C. P.; OLIVIER, J. C. Proper comparison among methods using a confusion matrix. **IEEE International Geoscience and Remote Sensing Symposium (IGARSS)**, p. 3057-3060, 2015.

SÁNCHEZ-PERALTA, L. F.; PAGADOR, J. B.; PICON, A.; CALDERON, A. J.; POLO, F.; ANDRAKA, N.; BILBAO, R.; GLOVER, B.; SARATXAGA, C. L.; SANCHEZ-MARGALLO, F. M. PICCOLO White-Light and Narrow-Band Imaging Colonoscopic Dataset: A Performance Comparative of Models and Datasets. **Applied Sciences**, v. 10, n. 23, p. 8501, 2020.

SANDLER, M.; HOWARD, A. G.; ZHU, M.; ZHMOGINOV, A.; CHEN, L-C.; MobileNetV2: Inverted Residuals and Linear Bottlenecks. **IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**, p. 4510-4520, 2018.

SANN, S. S.; WIN, S. S.; THANT, Z. M. An analysis of various Image Pre-Processing techniques in Butterfly Image. **International Journal for Advance Research and Development**, v. 6, n. 1, p. 1-4, 2021.

SHAMIR, R. R.; DUCHIN, Y.; KIM, J.; SAPIRO, G.; HAREL, N. Continuous Dice Coefficient: a Method for Evaluating Probabilistic Segmentations. **Bioengineering**, 2018.

SHARMA, N.; SHARMA, R.; JINDAL, N. Machine Learning and Deep Learning Applications-A Vision. **International Conference on Advances in Information, Computing and Trends in Data Engineering (AICDE)**. v. 2, n. 1, p. 24-28, 2021.

SHIN, Y.; QADIR, H. A.; AABAKKEN, L.; BERGSLAND, J.; BALASINGHAM, I. Automatic colon polyp detection using region based deep cnn and post learning approaches. **IEEE Access**, v. 6, pp. 40950-40962, 2018.

SHORTEN, C.; KHOSHGOFTAAR, T. M. A survey on Image Data Augmentation for Deep Learning. **Journal of Big Data**, v. 6, n. 1, p. 60, 2019.

SIDDIQUE, N.; PAHEDING, S.; ELKIN, C. P.; DEVABHAKTUNI, V. U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications. **IEEE Access**, v. 9, p. 82031-82057, 2021.

SILVA, J.; HISTACE, A.; ROMAIN, O.; DRAY, Z.; GRANADO, B. Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer. **International Journal of Computer Assisted Radiology and Surgery**, v. 9, n. 2, p. 283-293, 2014.

SUTTON, Richard S.; BARTO, Andrew G. **Reinforcement Learning: An Introduction**. 2. ed. CAMBRIDGE: MIT PRESS, p. 526, 2018.

TALEBI, H.; MILANFAR, P. Learning to Resize Images for Computer Vision Tasks. **International Conference on Computer Vision (ICCV)**, 2021.

TAN, M.; LE, Q. V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. **International Conference on Machine Learning (ICML)**, 2020.

TAYLOR, J.; WANG, W.; BALA, B.; BEDNARZ, T. **Optimizing the optimizer for data driven deep neural networks and physics informed neural networks**, p. 23, 2022.

TOMAR, N. K.; JHA, D.; RIEGLER, M. A.; DAGENBORG, H. J.; JOHANSEN, D.; RITTSCHER, J.; HALVORSEN, P.; ALI, S. **FANet: A Feedback Attention Network for Improved Biomedical Image Segmentation**, 2022.

TRESCA, Amber J. **Stages of Colon and Rectal Cancer**. 2021. Disponível em: <https://www.verywellhealth.com/what-are-the-stages-of-colon-and-rectal-cancer-1941590>. Acesso em: 05 jan. 2022.

- TU, M.; ZHANG, X. Speech enhancement based on Deep Neural Networks with skip connections. **IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)**, 2017.
- VAN BEERS, F.; LINDSTROM, A.; OKAFOR, E.; WIERING, M. A. Deep Neural Networks with Intersection over Union Loss for Binary Image Segmentation. **International Conference on Pattern Recognition Applications and Methods (ICPRAM)**, 2019.
- VAN RIJN, J. C.; REITSMA, J. B.; STOKER, J.; BOSSUYT, P. M.; DEVENTER, S. J. V.; DEKKER, E. Polyp miss rate determined by tandem colonoscopy: a systematic review. **The American Journal of Gastroenterology**, v. 101, n. 2, p. 343–350, 2006.
- WEHLE, H.-D. **Machine Learning, Deep Learning, and AI: What's the Difference?**, 2017.
- XU, D.; TIAN, Y. A Comprehensive Survey of Clustering Algorithms. **Annals of Data Science**, v. 2, n. 2, p. 165–193, 2015.
- YAMASHITA, R.; NISHIO, M.; DO, R. K. G.; TOGASHI, K. Convolutional neural networks: an overview and application in radiology. **Insights into Imaging**, v. 9, n. 4, p. 611–629, 2018.
- YEUNG, M.; SALA, E.; SCHONLIEB, C-B.; RUNDO, L. Focus U-Net: A novel dual attention-gated CNN for polyp segmentation during colonoscopy. **Computers in Biology and Medicine**, v. 137, p. 104815, 2021.
- YIN, Z.; WAN, B.; YUAN, F.; XIA, X.; SHI, A. J. A Deep Normalization and Convolutional Neural Network for Image Smoke Detection. **IEEE Access**, v. 5, p. 18429–18438, 2017.
- YOUNG, I. T. **Fundamentals of Image Processing**. Delft University Of Technology, 110 p., 1995.
- ZHANG, H.; ZHANG, L.; JIANG, Y. Overfitting and Underfitting Analysis for Deep Learning Based End-to-end Communication Systems. **International Conference on Wireless Communications and Signal Processing (WCSP)**, p. 1-6, 2019.
- ZHANG, L.; DOLWANI, S.; YE, X. Automated Polyp Segmentation in Colonoscopy Frames Using Fully Convolutional Neural Network and Textons. **Medical Image Understanding and Analysis**, 2017.
- ZHANG, Y.; LIU, S.; LI, C.; WANG, J. Rethinking the Dice Loss for Deep Learning Lesion Segmentation in Medical Images. **Journal of Shanghai Jiaotong University**, v. 26, n. 1, p. 93–102, 2021.
- ZHUANG, F.; QI, Z.; DUAN, K.; XI, D.; ZHU, Y.; ZHU, H.; XIONG, H.; HE, Q. A Comprehensive Survey on Transfer Learning. **Proceedings of the IEEE**, v. 109, n. 1, p. 43–76, 2021.