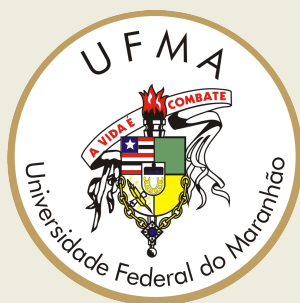




UNIVERSIDADE FEDERAL DO MARANHÃO

Programa de Pós-Graduação em Ciência da Computação

Arthur Guilherme Santos Fernandes

***Segmentação de Traqueia em Tomografia
Computadorizada Utilizando Efficient-Deeplab***

**São Luís
2024**

Arthur Guilherme Santos Fernandes

Segmentação de Traqueia em Tomografia Computadorizada Utilizando Efficient-Deeplab

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

Programa de Pós-Graduação em Ciência da Computação

Universidade Federal do Maranhão

Orientador: Prof. Dr. Geraldo Braz Junior

Coorientador: Prof. Dr. João Otávio Bandeira Diniz

São Luís - MA

2024

Arthur Guilherme Santos Fernandes

Segmentação de Traqueia em Tomografia Computadorizada Utilizando Efficient-Deeplab

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

Trabalho aprovado. São Luís - MA, 28 de Junho de 2024:

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

Prof. Dr. João Otávio Bandeira Diniz
Coorientador
Universidade Federal do Maranhão

**Prof. Dr. João Dallyson Sousa de
Almeida**
Examinador Interno
Universidade Federal do Maranhão

**Prof. Dr. António Manuel Trigueiros
da Silva Cunha**
Examinador Externo
Universidade de Trás-os-Montes e Alto
Douro

São Luís - MA
2024

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

Santos Fernandes, Arthur Guilherme.

Segmentação de Traqueia Em Tomografia Computadorizada
Utilizando Efficient-deeplab / Arthur Guilherme Santos
Fernandes. - 2024.

58 f.

Coorientador(a) 1: João Otávio Bandeira Diniz.

Orientador(a): Geraldo Braz Junior.

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

1. Aprendizado Profundo. 2. Segmentação de Imagens.
3. Redes Neurais. 4. Tomografia Computadorizada. 5.
Radioterapia. I. Bandeira Diniz, João Otávio. II. Braz
Junior, Geraldo. III. Título.

*Dedico este trabalho à minha mãe, Antônia de Fátima, que dedicou a sua vida inteira para
que eu pudesse chegar a este momento.*

Agradecimentos

Agradeço ao meu orientador Prof. Dr. Geraldo Braz Junior pela dedicação, paciência e por ser uma inspiração para mim. Agradeço também aos companheiros de laboratório que compartilharam seus conhecimentos e ouviram minhas ideias durante inúmeros almoços no R.U e operações de aquisição de café no NCA.

Our smartest machines are still blind.

(Fei-Fei Li)

Resumo

A segmentação de Órgãos em Risco (OAR) é crucial no planejamento da radioterapia. Seu objetivo é demarcar os tecidos saudáveis para que a radiação ionizante seja direcionada apenas às células cancerosas. Para isso, médicos realizam a demarcação dos órgãos manualmente, o que torna o processo demorado e propenso a erros. Portanto, metodologias de segmentação automática que utilizam aprendizado profundo podem acelerar a delimitação dos órgãos durante o planejamento da radioterapia. As metodologias existentes possuem muitos parâmetros, o que deixa o modelo pesado e caro de ser disponibilizado como um serviço. Este trabalho propõe o uso de uma arquitetura de rede neural convolucional denominada *EfficientDeeplab*, treinada em exames de tomografia computadorizada para realizar a segmentação de traqueia. O modelo diferencia-se de outras arquiteturas por possuir menor quantidade de parâmetros, o que a torna ideal para aplicações em serviços de saúde em larga escala. Para obter baixa quantidade de parâmetros e uma segmentação de qualidade foram aplicadas as convoluções atrous e a arquitetura EfficientNet. Foram realizados testes no conjunto de dados SEGTHOR, que obtiveram um *dice score* de 82,21%.

Palavras-chave: Aprendizado Profundo, Segmentação de Imagens, Redes Neurais, Tomografia Computadorizada, Radioterapia.

Abstract

Organs at Risk (OAR) segmentation is crucial in radiotherapy planning. Its objective is to demarcate healthy tissues so that ionizing radiation is directed only to cancer cells. To do this, doctors demarcate the organs manually, which makes the process very time-consuming and prone to errors. Therefore, automatic segmentation methodologies using deep learning can accelerate organ delineation during radiotherapy planning. Existing methodologies have many parameters, which makes the model cumbersome and expensive to make available as a service. This work proposes the use of a convolutional neural network architecture called *EfficientDeeplab*, trained on computed tomography scans to perform trachea segmentation. The model differs from other architectures by having fewer parameters, which makes it ideal for applications in large-scale healthcare services. To obtain a low number of parameters and quality segmentation, atrous convolutions and the EfficientNet architecture were applied. Tests were carried out on the SEGTHOR dataset, which obtained a *dice score* of 82.21%.

Keywords: Deep Learning, Image Segmentation, Neural Networks, Computerized Tomography, Radiotherapy.

Lista de ilustrações

Figura 1 – Cortes em TC a partir de diferentes ângulos: a) Axial; b) Coronal; c) Sagital.	13
Figura 2 – Planejamento Radioterápico com Auxílio de TC.	18
Figura 3 – Aplicação de Janelamento.	20
Figura 4 – Exemplo de segmentação de imagens.	21
Figura 5 – Segmentação de órgãos em TC.	21
Figura 6 – Neurônio Artificial.	24
Figura 7 – Rede Neural MLP.	25
Figura 8 – Operação de Convolução.	27
Figura 9 – Arquitetura Lenet.	27
Figura 10 – Representação Genérica das Redes EfficientNet.	28
Figura 11 – Diagrama do Bloco MBConv.	29
Figura 12 – Diagrama da Rede DeepLabV3+.	30
Figura 13 – Convoluções Atrous	30
Figura 14 – Módulo ASPP.	31
Figura 15 – Arquitetura das Redes FCN.	32
Figura 16 – Arquitetura da Rede U-net.	33
Figura 17 – Filtro de Convolução Atrous.	34
Figura 18 – Gargalos Residuais.	34
Figura 19 – Diagrama do Método Proposto.	38
Figura 20 – Demarcação de Órgãos Feita por Profissional.	39
Figura 21 – Aplicação do Janelamento de Intensidade.	40
Figura 22 – Diagrama da Rede DeepLabV3+.	41
Figura 23 – Visão Geral da Arquitetura do Modelo Proposto.	42
Figura 24 – Comparativo Após o Pós-Processamento.	42
Figura 25 – Falsos Positivos na Segmentação.	47
Figura 26 – Falso Positivo em Fatia.	47
Figura 27 – Porção Central da Traqueia.	48
Figura 28 – Divisão da Traqueia.	48

Lista de tabelas

Tabela 1 – Escala Hounsfield.	19
Tabela 2 – Janelas Radiológicas.	20
Tabela 3 – Resumo dos Trabalhos Relacionados.	37
Tabela 4 – Experimentos com Diferentes Backbones.	45
Tabela 5 – Comparação de Desempenho com Modelos U-net.	45
Tabela 6 – Comparação de Janelamento de Intensidade com EfficientDeeplab - B2.	46
Tabela 7 – Comparação de Resultados com Volumes Inteiros.	47
Tabela 8 – Comparação com a Literatura.	49

Lista de abreviaturas e siglas

ASPP	<i>Atrous Spatial Pyramid Pooling</i> (Pirâmide de Pooling Espacial Atrous).
CNN	<i>Convolutional Neural Networks</i> (Redes Neurais Convolucionais).
CRF	<i>Conditional Random Fields</i> (Campos Aleatórios Condicionais).
DICE	<i>Dice Score</i> .
DNA	<i>Deoxyribonucleic Acid</i> (Ácido Desoxirribonucleico).
FCN	<i>Fully Convolutional Networks</i> (Redes Completamente Convolucionais).
GAN	<i>Generative Adversarial Networks</i> (Redes Adversárias Generativas).
GPU	<i>Graphics Processing Unit</i> (Unidade de Processamento Gráfico).
HU	<i>Hounsfield Units</i> (Unidades de Hounsfield).
IOU	<i>Intersection Over Union</i> (Intersecção Sobre União).
MLP	<i>Multilayered Perceptron</i> (Perceptron de Múltiplas Camadas).
OR	Órgãos em Risco.
SGD	<i>Stochastic Gradient Descent</i> (Gradiente Descendente Estocástico).
TC	Tomografia Computadorizada.

Sumário

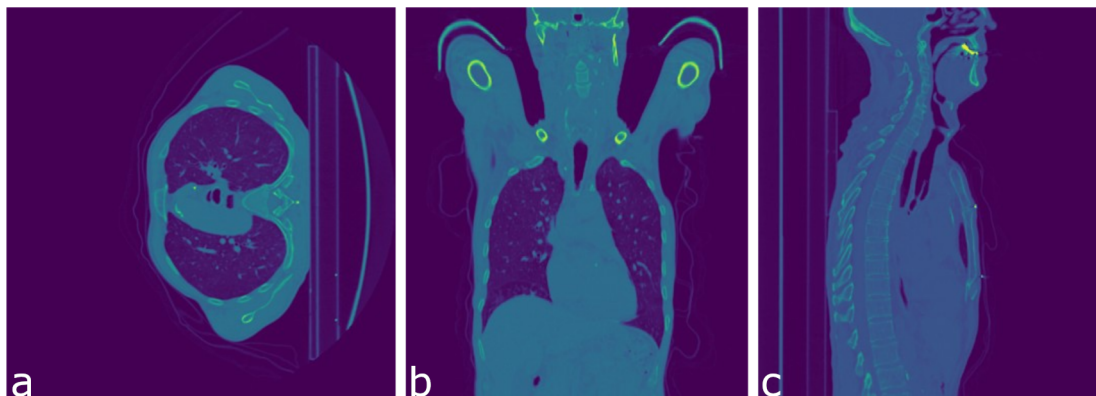
1	INTRODUÇÃO	13
1.1	Hipóteses de Pesquisa	14
1.2	Objetivo	15
1.2.1	Objetivos Específicos	15
1.3	Contribuições	15
1.3.1	Organização do Trabalho	16
2	FUNDAMENTAÇÃO TEÓRICA	17
2.1	Visão Computacional em Imagens Médicas	18
2.1.1	Tomografia computadorizada	18
2.1.2	Janelamento de Intensidade	19
2.1.3	Segmentação de Imagens	20
2.2	Aprendizado de Máquina	22
2.2.1	Redes Neurais	23
2.2.2	Redes Neurais Convolucionais	26
2.2.3	EfficientNet	28
2.2.4	DeepLab	29
2.3	Considerações Finais	30
3	TRABALHOS RELACIONADOS	32
3.1	Segmentação Semântica	32
3.2	Segmentação Semântica de Órgãos em Risco	35
4	MATERIAIS E MÉTODO	38
4.1	Aquisição da Imagens	38
4.2	Pré-Processamento	38
4.2.1	Janelamento de Intensidade	39
4.2.2	Redimensionamento e Normalização	40
4.3	Construção do Modelo	40
4.4	Pós Processamento	42
4.5	Considerações Finais	43
5	RESULTADOS	44
5.1	Ambiente de Execução	44
5.2	Treinamento do Modelo com Convoluções Atrous e DeepLab	44
5.3	Utilização de Janelamento de Intensidade	46

5.4	Estudo de Casos com Volumes Inteiros	46
5.5	Considerações	48
6	CONCLUSÃO	50
	REFERÊNCIAS	51

1 Introdução

A radioterapia é uma forma de tratamento do câncer onde doses substanciais de radiação ionizante são direcionadas às células malignas, destruindo-as ou reduzindo o tamanho do tumor. A fase inicial do planejamento da radioterapia gira em torno da identificação precisa da localização do tumor e dos órgãos adjacentes. Essa etapa é fundamental para impedir que radiação seja direcionada aos tecidos saudáveis próximos ao tumor, mitigando os seus efeitos adversos sobre as células não cancerígenas. Os órgãos que circundam o tumor são comumente denominados **Órgãos em Risco (OR)** (NOEL et al., 2016; DINIZ et al., 2021). Normalmente, durante o planejamento radioterápico, os ORs são delineados manualmente a partir de imagens adquiridas por **Tomografia Computadorizada (TC)**: uma técnica de aquisição de imagens de tecidos moles e duros no corpo humano que ajuda os profissionais a avaliarem a condição de órgãos e suas estruturas (WEBER et al., 2014). O resultado do exame é um volume 3D analisado pelos médicos por meio de cortes 2D (também chamados de *slices*). A Figura 1 apresenta exemplos de cortes a serem analisados por médicos. Da esquerda para a direita os ângulos de cortes são denominados: Axial, Coronal, Corte Sagital.

Figura 1 – Cortes em TC a partir de diferentes ângulos: **a)** Axial; **b)** Coronal; **c)** Sagital.



Fonte: Lambert et al. (2019).

Demarcar os ORs de maneira precisa é de grande importância para o planejamento da radioterapia, pois mapear os tecidos saudáveis ao redor do tumor permite que os especialistas direcionem o feixe de radiação exclusivamente ao tecido canceroso, minimizando a exposição desnecessária aos raios ionizantes e reduzindo a probabilidade de efeitos adversos da radioterapia, que podem incluir: inflamação, fibrose, úlcera e, em casos graves, falência de órgãos (TEKATLI et al., 2018). No entanto, a qualidade da demarcação dos órgãos depende da experiência clínica do médico, de maneira que um médico experiente terá maior precisão, resultando em pacientes com menos efeitos colaterais e mais qualidade

de vida (LAMBERT et al., 2019). Esta é uma tarefa demorada, que está suscetível aos erros humanos. Órgãos menores, como a traqueia, apresentam desafios únicos para a segmentação devido ao seu tamanho e similaridades texturais com tecidos circundantes. Tendo em vista estas dificuldades da segmentação manual de ORs, pesquisas recentes buscaram desenvolver métodos automatizados que acelerem o processo e eliminem o fator de erro humano (DINIZ et al., 2021), (FU et al., 2021), (FENG et al., 2020), (ZHAO; CHEN; LU, 2020).

Ao analisar os modelos de Redes Neurais utilizados nas metodologias mais bem sucedidas da literatura (ZHAO; CHEN; LU, 2020), (YAN et al., 2022), observa-se que possuem uma grande quantidade de parâmetros. Isso significa um maior uso de memória do computador. Pensando na utilização em larga escala de tais modelos, mais parâmetros pode significar mais gastos operacionais (plataforma de nuvem, máquinas, GPU, etc), o que por sua vez, ocasiona na baixa aplicabilidade deste tipo de metodologia.

Neste contexto, esta dissertação de mestrado visa propor uma arquitetura de Rede Neural Convolutiva para segmentação de traqueia que possua menos parâmetros e capacidade similar aos melhores modelos da literatura, facilitando a sua implementação em sistemas médicos de larga escala.

1.1 Hipóteses de Pesquisa

Para tanto, foram estabelecidas as seguintes hipóteses de pesquisa que norteiam este trabalho:

1. As convoluções *atrous*, quando aplicadas ao problema, contribuem para a construção de um modelo compacto e, ao mesmo tempo, capaz de generalizar;
2. A segmentação semântica da traqueia pode ser obtida com precisão via uma arquitetura de rede neural convolutiva que utiliza como backbone uma EfficientNet e como decoder uma DeepLabV3;
3. A segmentação da área da traqueia pode ser mais precisa caso seja determinada uma janela radiográfica como pré-processamento, específica para o problema;
4. Aplicar o algoritmo de componentes conectadas como pós-processamento consegue separar a região de traqueia dos falsos positivos de segmentação produzidos pelo modelo.

1.2 Objetivo

Esta dissertação de mestrado tem por objetivo avaliar metodologias de segmentação baseadas em Redes Neurais Convolucionais aplicadas à segmentação de traqueia para desenvolver um modelo que apresente um bom equilíbrio entre a quantidade de parâmetros treináveis e qualidade de segmentação e seja adequado para implementação em dispositivos de memória reduzida.

1.2.1 Objetivos Específicos

Especificamente para a segmentação de traqueia, este trabalho possui os seguintes objetivos:

- Desenvolver um modelo de CNN que combina as redes *EfficientNet* (TAN; LE, 2019) e *DeepLab* (CHEN et al., 2018b) para realizar a segmentação de traqueia;
- Avaliar a utilização de convoluções dilatáveis (também chamadas convoluções *atrous*) (HOLSCHNEIDER et al., 1990) para aumentar o campo de visão dos filtros de convolução sem adicionar parâmetros ao modelo;
- Analisar o impacto do janelamento de intensidade de *pixel* como pré-processamento no resultado da segmentação;
- Avaliar a acurácia de segmentação utilizando *Dice Score* e a quantidade de parâmetros treináveis do modelo final, comparando com as melhores arquiteturas disponíveis na literatura.

1.3 Contribuições

Este trabalho agrega valor na pesquisa de segmentação de órgãos em risco a partir das seguintes contribuições:

- Avaliação da rede *EfficientNet* (TAN; LE, 2019) como *backbone* extrator de características, em conjunto com a estrutura da *DeepLab* (CHEN et al., 2018b) para a segmentação de traqueia;
- Desenvolvimento de um modelo de segmentação automática de traqueia baseado em Redes Neurais Convolucionais que consome consideravelmente menos memória e possui menos parâmetros que o estado da arte, o que pode ser especialmente útil em ambientes de computação de memória limitada;
- Estudo de intervalos adequados para o janelamento de intensidade de *pixel* como pré-processamento no resultado da segmentação de traqueia.

1.3.1 Organização do Trabalho

Os capítulos subsequentes estão estruturados da seguinte maneira:

- O Capítulo 2 apresenta a fundamentação teórica em relação aos conceitos de Órgãos em Risco, planejamento radioterápico e também sobre as técnicas de Aprendizado Profundo e Redes Neurais Convolucionais utilizadas.
- O Capítulo 3 descreve os trabalhos relacionados à pesquisa atual de metodologias automáticas de segmentação de traqueia que utilizam Redes Neurais Convolucionais.
- O Capítulo 4 descreve todas as etapas que compõem a metodologia proposta para segmentação automática de traqueia.
- O Capítulo 5 apresenta os resultados obtidos a partir de experimentos, bem como análises de casos em que o modelo obteve boa e má precisão, buscando entender os motivos e fornecer soluções para o comportamento observado.
- O Capítulo 6 apresenta as considerações finais sobre os resultados, ideias para trabalhos futuros e os artigos científicos desenvolvidos durante a pesquisa.

2 Fundamentação Teórica

Neste capítulo são apresentados os conceitos fundamentais utilizados para o desenvolvimento desta pesquisa. Estes conceitos abrangem desde conhecimentos sobre câncer e radioterapia, até o processo de análise de imagens médicas e aprendizado profundo com redes neurais convolucionais.

A radioterapia é um dos principais métodos de tratamento do câncer e pode ser aplicada em diversos casos, embora sua eficácia dependa de uma análise individualizada do paciente. Este método pode ser eficaz tanto em estágios iniciais, quanto avançados da doença. No primeiro caso, este pode ser o tratamento definitivo, evitando uma cirurgia e preservando o órgão alvo. Já no caso de estágios avançados de câncer, a radioterapia pode ser um grande aliado da cirurgia e dos cuidados paliativos (VAIDYA, 2021).

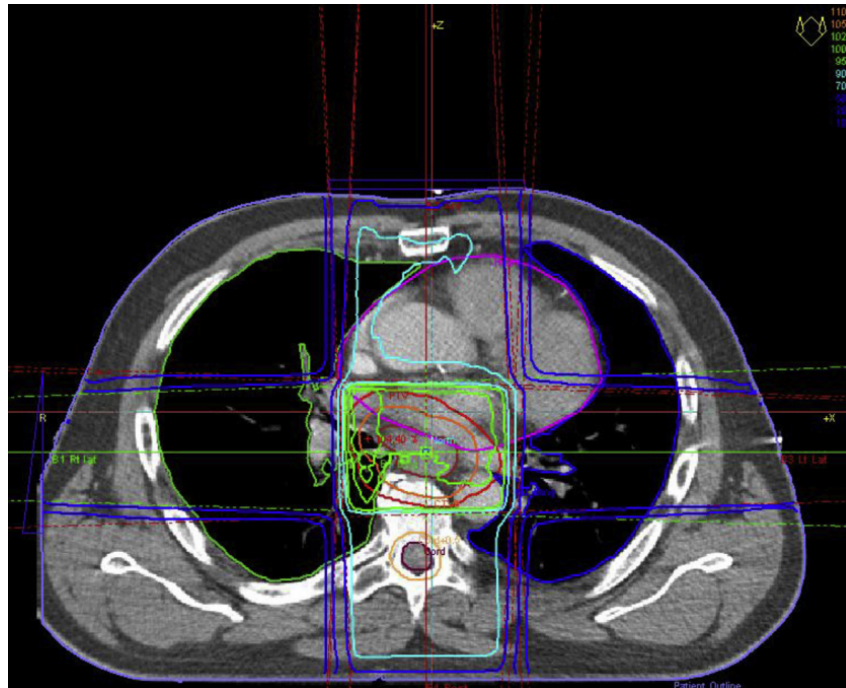
Como sugere o nome, este tratamento consiste em bombardear tecidos cancerígenos com raios ionizantes, sendo a Radiação Gama (γ) e os Raios-x os dois tipos de radiação mais utilizados na medicina (MEHTA et al., 2010). Ao entrarem em contato com tecido biológico, a radiação ionizante consegue remover elétrons das órbitas dos átomos, o que a nível molecular significa a quebra de porções de DNA das células (VAIDYA, 2021). O corpo possui mecanismos de reparo automático de DNA, porém o processo de regeneração é consideravelmente mais lento nas células cancerígenas, daí a eficácia do tratamento.

Os efeitos colaterais da radiação ionizante em tecido saudável podem ser divididos em dois tipos: imediatos e tardios (BARAZZUOL; COPPES; LUIJK, 2020). O primeiro tipo pode ser sentido dias ou semanas após o tratamento e pode se manifestar na forma de irritação no tecido, pequenos sangramentos e dores localizadas, sendo que estes sintomas são revertidos em pouco tempo. Já os efeitos tardios são crônicos e muitas vezes progressivos, levando a condições como: inflamação crônica, fibrose e úlcera (TEKATLI et al., 2018).

Para garantir que a radioterapia seja eficaz sem causar danos aos órgãos saudáveis próximos ao tumor, chamados de **Órgãos em Risco** (ORs), é realizado o planejamento radioterápico. Nessa etapa, os ORs são identificados e delimitados, pois, eles é que determinam a dose máxima de radiação que pode ser aplicada ao tumor alvo (EVANS; STAFFURTH, 2018). O objetivo é maximizar a dose no tumor e minimizar nos ORs. Após a aprovação do médico, os dados são transferidos para a máquina de tratamento. A Figura 2 ilustra o contorno dos órgãos e o direcionamento dos feixes de radiação. Nela é possível verificar quatro feixes de radiação: pela frente, por trás, pela esquerda e pela direita. O traçado vermelho ao centro da imagem é a região com câncer.

A delimitação de órgãos se torna então uma tarefa fundamental para o tratamento por radioterapia. O restante deste capítulo trata do referencial teórico utilizado para esta

Figura 2 – Planejamento Radioterápico com Auxílio de TC.



Fonte: [Evans e Staffurth \(2018\)](#).

finalidade.

2.1 Visão Computacional em Imagens Médicas

Por visão computacional, entende-se a área de estudo que busca extrair informações úteis de imagens (majoritariamente digitais) ([PRINCE, 2012](#)). Para alcançar este fim, são utilizados diversos algoritmos e técnicas de processamento de imagens, seja para restauração, melhoramento, segmentação, agrupamento de regiões, etc.

As aplicações de visão computacional estendem-se às muitas áreas do conhecimento. Alguns exemplos são: processamento de imagens via satélite, geração de imagens médicas por Raios-x e, até mesmo imagens geradas por fontes não convencionais, como ultrassom ([GONZALEZ; WOODS, 2008](#)). Dentro da medicina, a visão computacional tem ganhado muito destaque ao ajudar médicos em diagnósticos por imagem ou acelerar processos demorados como a demarcação de Órgãos em Risco. Esta última tarefa utiliza imagens geradas pelo exame de tomografia computadorizada, detalhado na Seção [2.1.1](#)

2.1.1 Tomografia computadorizada

A tomografia é uma técnica de imageamento de um objeto por meio de dados adquiridos ao irradiar o objeto a partir de diferentes direções. Seu impacto na medicina

foi revolucionário, pois permitiu aos médicos visualizarem órgãos internos com precisão e segurança sem precedentes (KAK; SLANEY, 2001).

Segundo Dance et al. (2014), o processo de aquisição de imagens de tomografia computadorizada envolve a irradiação de partículas através do corpo de um paciente e a subsequente medição da transmissão de radiação, feita por um arco detector posicionado atrás do paciente em direção oposta ao dispositivo emissor de radiação. Ao rotacionar o arco detector e o dispositivo emissor em torno do paciente, é possível adquirir um volume de amostras de dados, utilizadas posteriormente para reconstruir uma imagem digital.

A imagem reconstruída possui os valores de intensidade de píxel dentro da Escala Hounsfield (SEERAM, 2016), variando conforme a densidade do material representado. O nome da unidade de medida para estes valores é Unidade de Hounsfield (HU). A Tabela 1 detalha os valores apresentados para alguns tipos de tecido do corpo humano.

Tabela 1 – Escala Hounsfield.

Substancia	Unidades de Hounsfield (HU)
Ar	-1000
Gordura	-100
Água	0
Fluido Cefaloraquídeo	15
Matéria Branca	20 - 30
Massa Cinzenta	36 - 46
Músculo	50
Sangue	56 - 76
Osso Compacto	1000

Fonte: Seeram (2016).

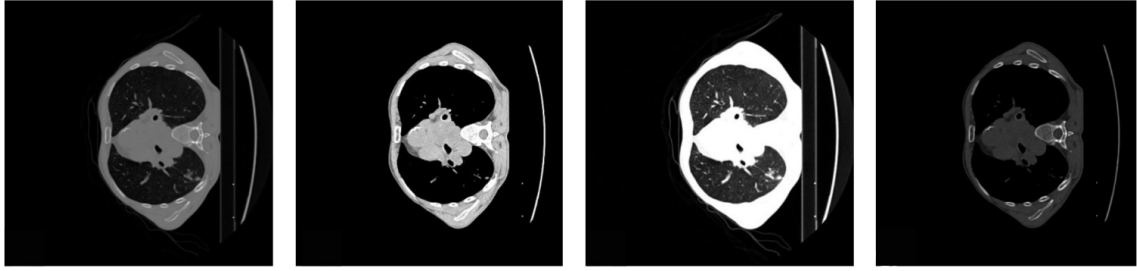
2.1.2 Janelamento de Intensidade

Em processamento de imagens, o pré-processamento é uma das etapas fundamentais de uma metodologia, pois é nela que ocorrem melhoramentos, realces, correção de artefatos, redução de ruído, alinhamento, etc. Esses processos simplificam a análise da imagem, seja por outro algoritmo ou por humanos (GONZALEZ; WOODS, 2008).

O Janelamento de Intensidade é um dos pré-processamentos mais comuns aplicados na medicina diagnóstica por tomografia computadorizada. Consiste em selecionar um determinado intervalo do espectro de intensidades provenientes da imagem tomográfica e limitar os valores de intensidade de píxel a esse intervalo. Estes intervalos são conhecimentos como “janelas” (WEBER et al., 2014) e cada janela serve para melhorar a visualização de determinados tipos de tecido.

A Figura 3 ilustra algumas das janelas mais comuns da radiologia aplicadas à mesma imagem de origem, da esquerda para a direita: Imagem original, Janela de Mediastino,

Figura 3 – Aplicação de Janelamento.



Fonte: Elaborado pelo autor.

Janela Pulmonar (parênquima) e Janela Óssea (CONSIDERA et al., 2006). Já a Tabela 2, sumariza os intervalos em Unidades de Hounsfield (HU) para cada uma destas janelas. Os valores referência para cada intervalo podem ser encontrados em: VICENTE et al. (2004), CONSIDERA et al. (2006) e MARCHIORI et al. (2007).

Tabela 2 – Janelas Radiológicas.

Nome	Aplicação	Intervalo (HU)
Janela de Parênquima	Vis. de pulmões e vias aéreas (traquéia, nasal, etc.)	[-1000, 150]
Janela de Mediastino	Vis. de partes moles (músculos, órgãos abdominais, etc.)	[-160, 240]
Janela Óssea	Vis. de ossos compactos (crânio, fêmur, etc.)	[-1.650, 2350]

Fonte: elaborado pelo autor.

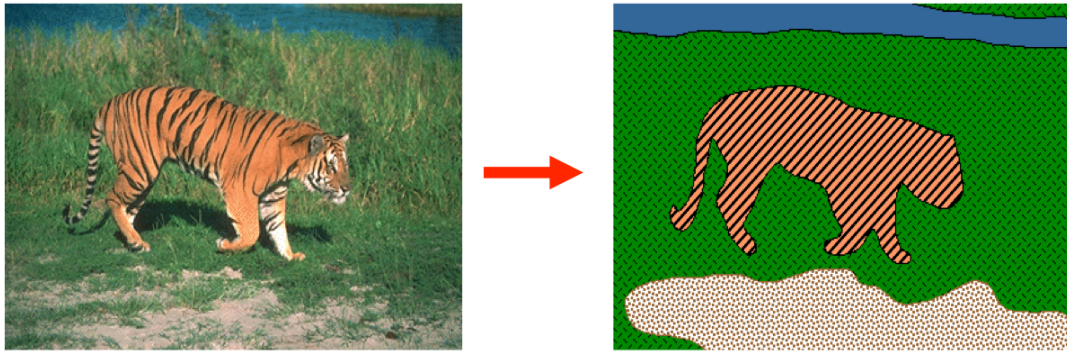
Sejam α e β , respectivamente, o limite inferior e superior do intervalo $[\alpha, \beta]$ escolhido, a operação de janelamento de intensidade pode ser definida como na Equação 2.1:

$$f(x) = \begin{cases} \alpha & \text{se } x \leq \alpha \\ \beta & \text{se } x \geq \beta \\ x & \text{se } x \in [\alpha, \beta] \end{cases} \quad (2.1)$$

2.1.3 Segmentação de Imagens

Uma das tarefas elementares dentro da visão computacional é a segmentação de imagens. Consiste em subdividir uma imagem em suas regiões constituintes (GONZALEZ; WOODS, 2008) para analisar suas características de maneira isolada. De maneira geral, a saída dos algoritmos de segmentação é chamada de máscara de segmentação. Esta contém *labels* que indicam a qual classe pertence cada *pixel*. A Figura 4 ilustra o resultado esperado de uma segmentação de imagens, onde o algoritmo (representado pela seta vermelha) foi capaz de agrupar os *pixels* em quatro classes: animal, terra, água e vegetação

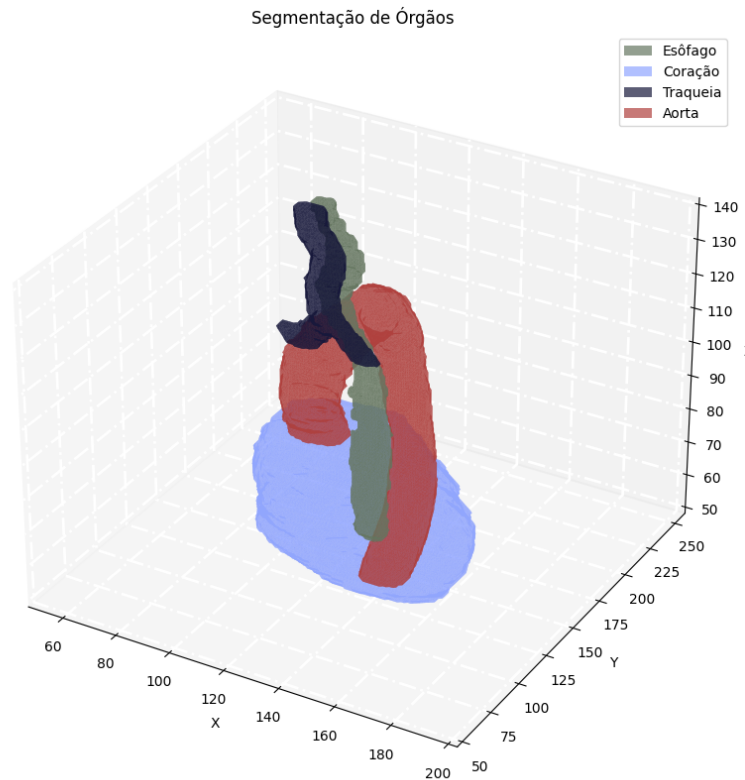
Figura 4 – Exemplo de segmentação de imagens.



Fonte: (YEUNG, 2010).

Na medicina, alguns exemplos de aplicação da segmentação são: a medida de ossos e tecidos (LEE et al., 2022), detecção de estruturas anômalas (THANOON et al., 2023), identificação de músculos lesionados (ACKERMANS et al., 2021) e segmentação de órgãos (LAMBERT et al., 2019). A Figura 5 apresenta a saída esperada de um algoritmo de segmentação de órgãos contendo: demarcação do coração, traqueia, esôfago e artéria aorta.

Figura 5 – Segmentação de órgãos em TC.



Fonte: Elaborado pelo autor.

Para determinar a qualidade de uma segmentação de imagem deve-se comparar

(visual ou numericamente) as máscaras geradas pelo algoritmo, com a segmentação considerada a mais próxima do ideal (também chamada de valor-verdade). O padrão na literatura de segmentação de imagens médicas é a utilização de funções matemáticas de pontuação (métricas) como o *Dice Score* (DICE, 1945) e o *Jaccard Score* (JACCARD, 1912), como evidenciado em Eelbode et al. (2020) e Isensee et al. (2021) por exemplo.

O *Dice-Score* é dado pela Equação 2.2, onde Y_{true} e Y_{out} representam, respectivamente, os conjuntos de *pixels* pertencentes à saída esperada e à saída obtida. A mesma nomenclatura vale para o *Jaccard Score* (também referido por IOU), dado pela Equação 2.3. Ambas as métricas de segmentação serão utilizadas para avaliar o modelo proposto neste trabalho. Os detalhes estão situados na Seção 5.5.

$$D(Y_{true}, Y_{out}) = 2 \cdot \frac{|Y_{true} \cap Y_{out}|}{|Y_{true}| + |Y_{out}|} \quad (2.2)$$

$$J(Y_{true}, Y_{out}) = \frac{|Y_{true} \cap Y_{out}|}{|Y_{true} \cup Y_{out}|} \quad (2.3)$$

O pós-processamento da área segmentada é uma etapa essencial em muitas tarefas de segmentação de imagem (SHIVHARE; GUPTA, 2015; DHAWAN; BODANI; GARG, 2019) incluindo segmentação de órgãos. Após a segmentação inicial, técnicas de pós-processamento são aplicadas para refinar e melhorar os resultados. Isso pode envolver a remoção de pequenos artefatos ou ruídos, a correção de erros de segmentação, a suavização das bordas ou a conexão de regiões segmentadas que fazem parte do mesmo objeto. O pós-processamento é fundamental para garantir que os resultados da segmentação sejam precisos e clinicamente úteis, especialmente em aplicações médicas. Este trabalho utilizou o pós-processamento de regiões conectadas (SAMET; TAMMINEN, 1988) para extrair o maior objeto tridimensional obtido pela segmentação. O algoritmo segue os seguintes passos:

1. Detecção das ilhas de *voxels* da imagem (*voxels* de valor igual que são vizinhos).
2. Contagem da quantidade de *voxels* presente em cada ilha.
3. Seleção da ilha que contém maior quantidade de *voxels*.

2.2 Aprendizado de Máquina

O Aprendizado de Máquina é uma sub-área da Inteligência Artificial que busca desenvolver modelos de processamento de informação que aprendem a partir de observações (dados), melhorando seu desempenho em tarefas futuras. Russell e Norvig (2009) resume

a operação deste tipo de sistema: a partir de uma coleção de pares de entrada e saída, aprende uma função que prevê a saída para novas entradas.

Segundo [Shalev-Shwartz e Ben-David \(2014\)](#), estes modelos são particularmente úteis ao desempenhar tarefas que são muito complexas de serem programadas explicitamente ou que necessitem de maior flexibilidade, adaptando-se às entradas. Nestes casos, é mais viável projetar um modelo de Aprendizado de Máquina do que desenvolver um software que aplique um conjunto de regras de processamento.

Existem diferentes tipos de aprendizado: supervisionado, semi-supervisionado, não supervisionado e aprendizado por reforço. A abordagem de treinamento utilizada neste trabalho é o aprendizado supervisionado. Para [Burkov \(2020\)](#), este tipo de aprendizagem ocorre quando há um modelo preditivo e uma coleção de exemplos rotulados $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, onde x_i é chamado de vetor de características e y_i é o chamado rótulo. O objetivo do aprendizado supervisionado é criar um modelo que, dado um vetor de características x como entrada, seja capaz de deduzir um rótulo y para esse vetor.

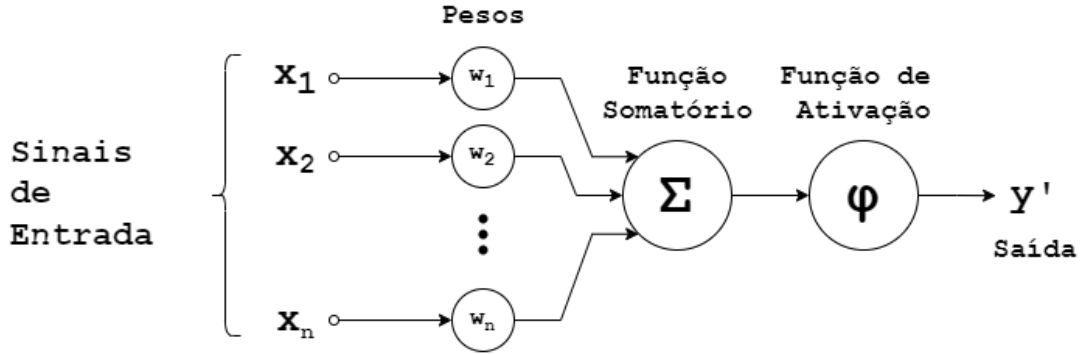
2.2.1 Redes Neurais

As redes neurais são um tipo específico de modelo de aprendizado de máquina. Segundo [Haykin \(2009\)](#), foram desenvolvidas a partir da constatação de que o cérebro humano computa de forma completamente diferente do computador digital e possui a capacidade de construir suas próprias regras de comportamento através do que chamamos de *experiência*. Para [Hecht-Nielsen \(1989\)](#), as redes neurais artificiais são uma estrutura de processamento de informações que funciona de maneira paralela e distribuída, que consiste em unidades básicas de processamento interconectadas por canais de comunicação unidirecional.

As referidas *unidades básicas de processamento* são os neurônios artificiais. Assim como a sua contraparte biológica, eles possuem uma propriedade conhecida como plasticidade ([HAYKIN, 2009](#)), isto é: têm a capacidade de ajustar o seu funcionamento interno conforme as entradas recebidas. A Figura 6 ilustra um neurônio artificial e suas partes constituintes.

O modelo de neurônio artificial apresentado acima foi primeiramente proposto por [Rosenblatt \(1958\)](#), e é chamado de *perceptron*. Suas partes constituintes são: o conjunto de sinais de entrada $X = \{x_1, x_2, \dots, x_n\}$, o conjunto de pesos $W = \{w_1, w_2, \dots, w_n\}$, o *bias* que age como uma entrada de valor fixo 1, a função somatório Σ , função de ativação φ e a saída y_k . Para que o *perceptron* consiga aprender, é executada uma etapa de treinamento, onde os sinais de entrada X são apresentados ao modelo, obtendo o sinal de saída y' . Em seguida, os pesos são atualizados seguindo a regra de aprendizado escolhida ([HECHT-NIELSEN, 1989](#)). A Equação 2.4 detalha a obtenção da saída do *perceptron*, onde y' é

Figura 6 – Neurônio Artificial.



Fonte: Adaptado de (HAYKIN, 2009).

a saída, φ representa uma função de ativação, e, x_i e w_i são o valor de entrada e o peso da conexão, respectivamente. Já as Equações 2.5 e 2.6 detalham a função de ativação e a regra de aprendizado utilizadas no trabalho de Rosenblatt (1958).

$$y' = \varphi\left(\sum_{i=1}^n x_i w_i\right) \quad (2.4)$$

$$\varphi(z) = \begin{cases} 1 & \text{se } z \geq 0 \\ 0 & \text{se } z < 0 \end{cases} \quad (2.5)$$

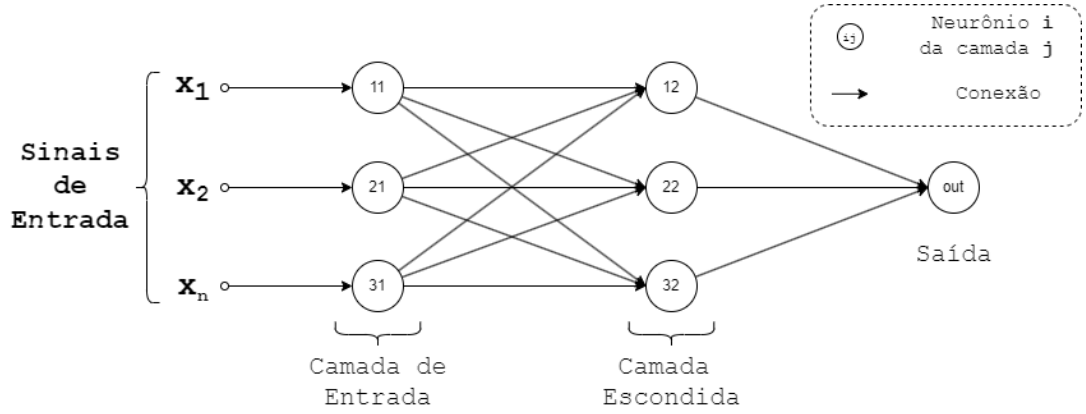
$$w_i = w_i + (y - y')x_i \quad (2.6)$$

Por si só, o *perceptron* já consegue aproximar funções, entretanto, é possível organizar os neurônios em camadas interconectadas, formando as primeiras redes neurais desenvolvidas, conhecidas como *Multi Layered Perceptrons* (MLP). A Figura 2.2.1 detalha a composição de uma rede neural MLP clássica contendo: camada de entrada, conexões sinápticas representadas por setas, os neurônios representados por círculos, a camada intermediária, também chamada de camada escondida e o neurônio de saída.

A partir deste ponto, a teoria de redes neurais recebeu cada vez mais incrementos, sendo um deles o uso do algoritmo de Gradiente Descendente Estocástico (SGD) (ROBBINS; MONRO, 1951). Ele funciona atualizando iterativamente os pesos da rede com base no gradiente da função de erro calculado em um subconjunto aleatório dos dados de treinamento (AMARI, 1967).

Já o algoritmo de retropropagação, também referido por *backpropagation*, foi introduzido por Rumelhart, Hinton e Williams (1986), e tornou-se um marco importante no desenvolvimento das redes neurais. Ele consiste em calcular o gradiente da função de

Figura 7 – Rede Neural MLP.



Fonte: Elaborado pelo autor.

erro em relação aos próprios pesos da rede, permitindo a adição de múltiplas camadas escondidas. Haykin (2009) define os cinco passos que constituem o *backpropagation*:

1. Calcular o erro de saída e_{out} , definido pela função de erro (*loss*) escolhida. Este trabalho utiliza a Entropia Cruzada Binária (GOODFELLOW; BENGIO; COURVILLE, 2016) como função de *loss* definida pela Equação 2.7.

$$e_{out} = - \sum_{i=0}^1 y_{true} \cdot \log(y_{out}) \quad (2.7)$$

2. Calcular a variação dos pesos entre a camada de saída e a camada escondida. Esta operação é dada pela Equação 2.8, onde η é uma constante pré-definida chamada taxa de aprendizado, $\varphi'(y_j)$ é a derivada da função de ativação φ e y_j é a saída da camada escondida.

$$\Delta W_{j,out} = \eta \cdot e_{out} \cdot \varphi'(y_j) \cdot y_j \quad (2.8)$$

3. Calcular erro retropropagado, dado pela Equação 2.9, onde $W_{j,out}$ representa os pesos sinápticos entre a camada escondida e a camada de saída.

$$e_j = - \sum_{i=1}^n e_{out} \cdot \varphi'(y_j) \cdot W_{j,out} \quad (2.9)$$

4. Calcular variação dos pesos entre as camadas escondida e de entrada, através da Equação 2.10. Nela, y_i representa a saída da camada de entrada.

$$\Delta W_{i,j} = -\eta \cdot e_j \cdot \varphi'(y_i) \cdot y_i \quad (2.10)$$

5. Somar os pesos antigos aos ΔW obtidos.

Desta maneira, pode-se ajustar os pesos de uma rede neural com múltiplas camadas, minimizando o erro e consequentemente, tornando-o melhor na tarefa para qual está sendo treinado. Este tipo de rede com várias camadas intermediárias é conhecido como rede neural profunda, e deu origem ao ramo do Aprendizado de Máquina conhecido como Aprendizado Profundo ou *Deep Learning*.

O uso combinado do SGD e da retropropagação permitiu o treinamento eficiente de redes neurais com muitas camadas ocultas, o que antes era considerado inviável devido aos desafios de treinamento. Isso abriu caminho para o surgimento do aprendizado profundo e possibilitou avanços significativos em áreas como reconhecimento de fala, visão computacional, processamento de linguagem natural, etc.

2.2.2 Redes Neurais Convolucionais

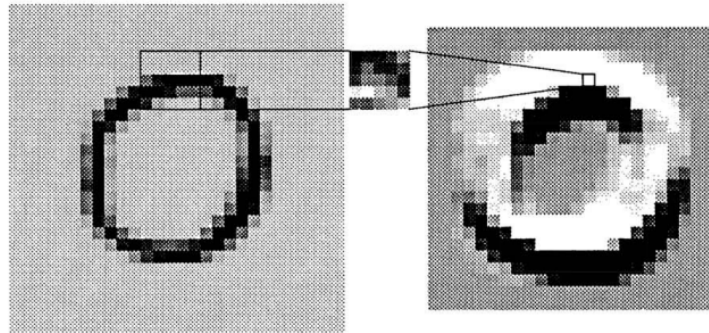
Em visão computacional, utilizar as redes completamente conectadas (MLP) para processar imagens é algo que exige que a rede possua muitos neurônios. Por exemplo, para uma imagem de dimensões 256×256 , têm-se ao total 65.536 neurônios apenas na camada de entrada. Este número aumenta exponencialmente em relação ao tamanho da imagem, tornando a utilização do *perceptron* proibitivamente custosa.

Para solucionar este problema, as Redes Neurais Convolucionais (CNNs) foram especialmente projetadas para processar imagens com eficiência. Ao introduzir esse modelo de rede, [LeCun et al. \(1989\)](#) inspirou-se na ideia do processamento clássico de imagens de detectar e combinar características locais. Portanto, restringiu os neurônios das camadas iniciais para que eles recebessem como entrada uma pequena porção da imagem (por exemplo, uma janela 3×3). Em seguida, desloca-se o neurônio até que ele tenha processado toda a imagem, armazenando o resultado em um espaço de memória que recebe o nome de mapa de características. Esta operação ficou conhecida como convolução e dá o nome a este tipo de rede neural. A Figura 8 ilustra a operação de convolução, contendo a imagem de entrada (à esquerda), os pesos do neurônio (ao meio) e o mapa de características (à direita).

A Figura 9 apresenta a arquitetura proposta por [LeCun et al. \(1989\)](#), com camadas de convolução que transformam a entrada em mapas de características e reduzem a dimensionalidade em direção à camada de saída totalmente conectada. As convoluções reduzem naturalmente o tamanho da entrada, enquanto as camadas de *pooling* também diminuem a dimensionalidade, geralmente usando médias ou máximos em regiões da imagem.

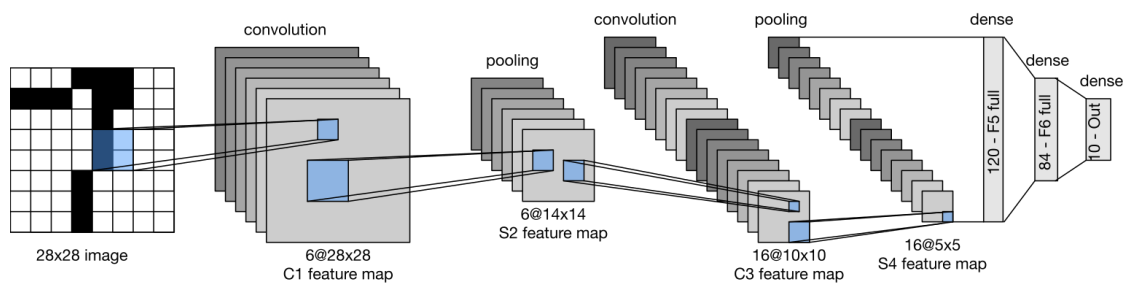
Esta arquitetura descarta a utilização dos neurônios completamente conectados nas camadas iniciais, sendo estes aplicados somente na camada final, que realiza a classificação. Isso diminuiu muito o custo computacional necessário para o processamento de imagens

Figura 8 – Operação de Convolução.



Fonte: Adaptado de [LeCun et al. \(1989\)](#).

Figura 9 – Arquitetura Lenet.



Fonte: Adaptado de [Zhang et al. \(2023\)](#).

com aprendizado profundo, tornando-o eficaz no reconhecimento de objetos, segmentação de imagens, etc.

A partir deste ponto, houve o surgimento de modelos mais profundos ([SIMONYAN; ZISSERMAN, 2015](#)), novos tipos de conexão ([HE et al., 2015](#)) e operações de convolução ([LONG; SHELHAMER; DARRELL, 2014](#)). Estas adições tornaram os modelos ainda mais eficientes e, ao mesmo tempo, mais elaborados em sua arquitetura, dando origem a diversos tipos de redes neurais convolucionais como, por exemplo: as Redes Generativas Adversárias ([GOODFELLOW et al., 2014](#)), Redes Completamente Convolucionais ([LONG; SHELHAMER; DARRELL, 2014](#)) e redes do tipo *Encoder-Decoder* como em [Ronneberger, Fischer e Brox \(2015\)](#).

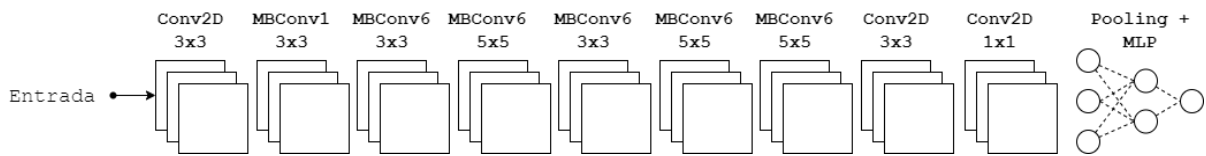
Arquiteturas encoder-decoder são amplamente utilizadas para a tarefa de segmentação semântica, onde um conjunto de características é extraída da imagem original e depois convertida numa máscara de segmentação. Este trabalho avalia a EfficientNet como mecanismo para auxiliar na extração de características juntamente com uma DeepLab

para o refino de características e decodificação na máscara de segmentação.

2.2.3 EfficientNet

A arquitetura *EfficientNet* trata-se de uma família de redes neurais pensadas com o uso eficiente de recursos em mente, escalando as suas camadas conforme a necessidade da aplicação (TAN; LE, 2019). São ao todo sete redes neurais convolucionais que se diferem apenas pela quantidade de parâmetros em cada camada. A Figura 10 ilustra a base de uma rede *EfficientNet*: são nove estágios, sendo três deles convoluções 2D convencionais, seis MBConv e 1 estágio de *pooling* e camadas completamente conectadas para realização de classificação.

Figura 10 – Representação Genérica das Redes EfficientNet.

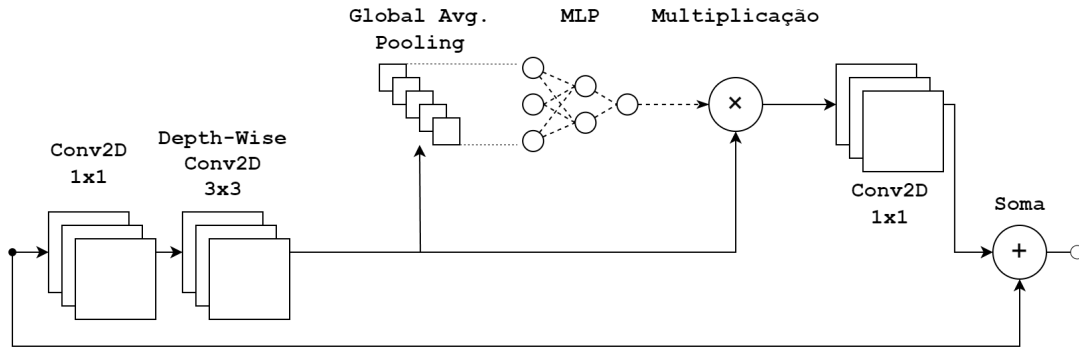


Fonte: Elaborado pelo autor.

As MBConv são partes fundamentais da arquitetura *EfficientNet*. Elas foram propostas pela primeira vez no trabalho de Sandler et al. (2019) e consistem em um gargalo residual invertido: expandem o número de canais nas camadas intermediárias e em seguida reduzem ao tamanho original, além de possuírem uma conexão residual entre o primeiro e o último bloco. A Figura 11 ilustra a estrutura do bloco MBConv, ele consiste em uma convolução 1×1 seguida de uma convolução *depth-wise* (detalhada no parágrafo seguinte), em seguida o resultado é encaminhado para um bloco de compressão-excitação, que consiste em uma camada de *pooling* por média, seguida de uma *mlp* cujo resultado é multiplicado com a entrada. Por fim, a saída deste bloco é encaminhada para mais uma convolução 1×1 e seu resultado é somado à entrada do bloco MBConv, formando a conexão residual e encaminhando a saída aos estágios seguintes de rede neural. O número que segue o nome em MBConv1 e MBConv6 é um multiplicador que indica o número de filtros na camada convolucional 1×1 mais à esquerda. Por exemplo, se o número de canais de entrada for 4, então o número de filtros será 4 no MBConv1 e 24 no MBConv6.

Por sua vez, o bloco de convolução *depth-wise* é um tipo de operação onde cada canal de entrada é processado por um neurônio diferente. De acordo com Chollet (2015), o processamento neste bloco se dá pelos seguintes passos: dividir a entrada em canais individuais, convoluir cada canal com um filtro, concatenar as saídas convolvidas ao longo do eixo dos canais.

Figura 11 – Diagrama do Bloco MBConv.



Fonte: Elaborado pelo autor.

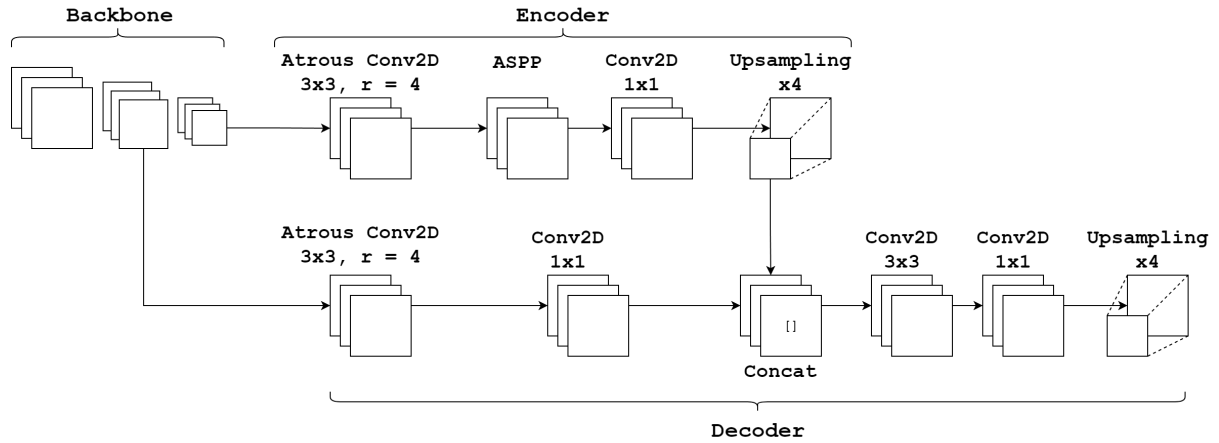
2.2.4 DeepLab

Este trabalho utiliza a estrutura *encoder-decoder*, baseando-se na arquitetura *DeepLabV3+* [Chen et al. \(2018b\)](#). A Figura 12 exibe o diagrama desta arquitetura. O bloco inicial, chamado de *backbone* é uma rede neural convolucional que pode ser modificada sem interferir no funcionamento do *encoder* ou *decoder*. Este trabalho modifica o *backbone* da rede, removendo o proposto no trabalho original, que é uma rede *Resnet* ([HE et al., 2015](#)) e inserindo no lugar a *EfficientNet* ([TAN; LE, 2019](#)) (os detalhes são explicitados na Seção 4.3).

O *encoder* é conectado com a camada final do *backbone*, recebe mapas de características compactos e os expande por meio do módulo ASPP, convoluções *atrous* e *upsampling*. Por sua vez, o *decoder* recebe como entrada mapas de características intermediários do *backbone*, e também a saída do *encoder*, para aplicar uma sequência de convoluções e *upsampling* (interpolação bilinear) ([KIRKLAND, 2010](#)), restaurando a imagem ao seu tamanho original e produzindo como saída uma máscara de segmentação.

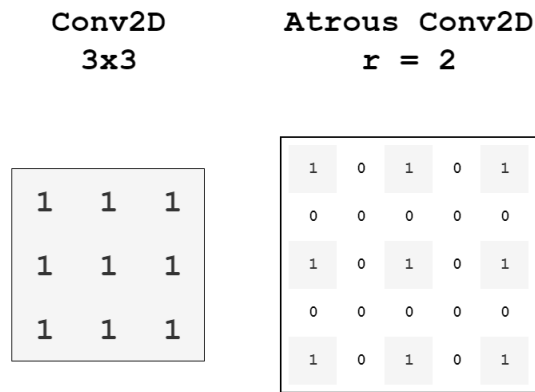
O módulo ASPP (*Atrous Spatial Pyramid Pooling*) consiste em uma pirâmide de *pooling* que consegue extrair e combinar características de diferentes resoluções utilizando convoluções dilatáveis, também chamadas de *atrous convolutions* ([CHEN et al., 2016](#)). Este tipo de convolução tem o objetivo de alcançar um campo de visão maior na extração de características. O seu vetor de pesos (filtro) é preenchido com zeros de acordo com uma taxa de dilatação escolhida (r). A Figura 13 exemplifica o efeito da taxa de dilatação sobre o filtro de convolução. Note que a área coberta pelo filtro passa de 3x3 para 5x5, mas o número de parâmetros permanece o mesmo. Variar a taxa de dilatação nas convoluções ajuda a capturar características em diversas escalas. Já a Figura 14 ilustra como o módulo ASPP utiliza convoluções *atrous* com diferentes taxas de dilatação para extrair *features*.

Figura 12 – Diagrama da Rede DeepLabV3+.



Fonte: Elaborado pelo autor.

Figura 13 – Convoluções Atrous

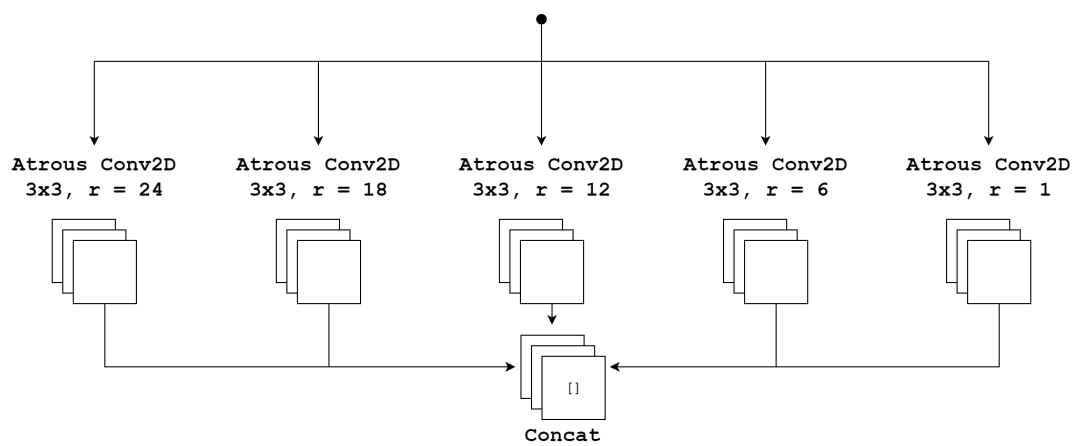


Fonte: Elaborado pelo autor.

2.3 Considerações Finais

Neste capítulo foram descritos conceitos envolvendo radioterapia, a aplicação de visão computacional em imagens médicas, fundamentos de processamento de imagens e segmentação de imagens, aprendizado de máquina, aprendizado profundo e redes neurais convolucionais. Estas informações compõem a base teórica para entendimento da metodologia proposta no Capítulo 4 e dos trabalhos relacionados apresentados no Capítulo 3.

Figura 14 – Módulo ASPP.



Fonte: Elaborado pelo autor.

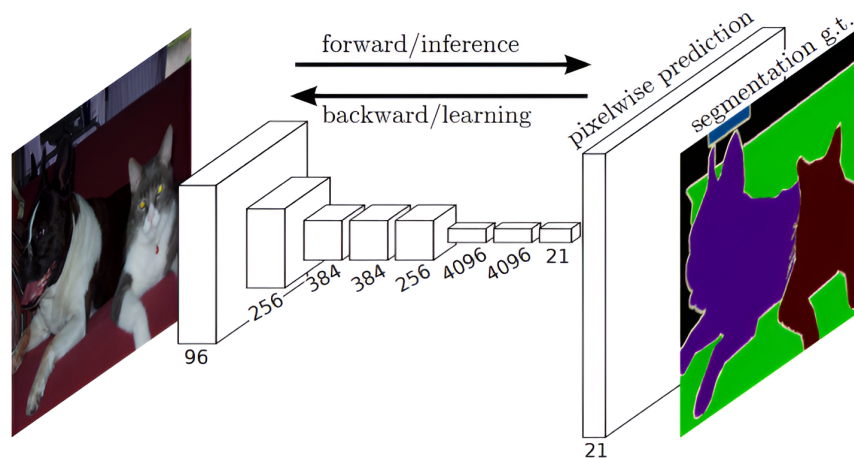
3 Trabalhos Relacionados

Neste capítulo são apresentados os trabalhos que ajudaram a fundamentar ou inspirar esta pesquisa. Eles estão distribuídos tanto na grande área de segmentação de imagens quanto especificamente na segmentação de Órgãos em Risco.

3.1 Segmentação Semântica

O trabalho de (LONG; SHELHAMER; DARRELL, 2014) é de grande importância para a segmentação de imagens via aprendizado profundo. Nele foram propostas as Redes Completamente Convolucionais (*FCN: Fully Convolutional Networks*) para segmentação semântica de imagens, contribuindo com a pesquisa nesta área ao propor uma arquitetura que substituiu as camadas densas (completamente conectadas) por camadas convolucionais e *upsampling* bilinear. Na época eram comuns esquemas de treinamentos apenas com sub-regiões das imagens, mas devido a sua arquitetura inovadora, as redes FCN geram uma saída do mesmo tamanho da entrada, possibilitando o treinamento fim-a-fim com imagens e máscaras de segmentação inteiras. A Figura 15 proporciona um panorama de uma arquitetura FCN, contendo as camadas convolucionais que diminuem sucessivamente a resolução da entrada ao mesmo tempo que extraem características. Ao final, uma camada de *upsampling* consegue transformar as características de baixa resolução de volta ao tamanho original, resultando em uma máscara de segmentação.

Figura 15 – Arquitetura das Redes FCN.

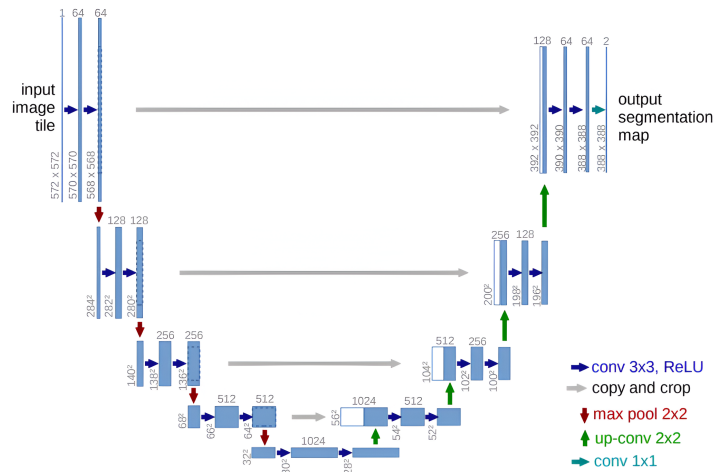


Fonte: Long, Shelhamer e Darrell (2014).

Uma das implementações de rede completamente convolucional com mais aplicações no contexto da segmentação de imagens médicas é a arquitetura U-net. Proposta por

Ronneberger, Fischer e Brox (2015), trouxe inovação para a área ao modificar a FCN, adicionando uma estrutura de expansão de características que é simétrica ao modelo de contração tradicional das redes convolucionais. A expansão ocorre a partir de camadas de convolução combinadas com *upsampling* aplicadas de maneira sucessiva, ficando conhecidas como *up-convolution*. Além disso, *features* de maior escala são concatenadas com a saída do *upsampling*, formando as *skip-connections*, que servem para suplementar o caminho de expansão com as *features* extraídas no caminho de contração da rede. A Figura 15 ilustra a arquitetura de uma rede U-net, que contém sucessivas camadas convolucionais, seguidas de uma estrutura simétrica composta de sucessivas *up-convolution*. A estrutura de contração-expansão é interconectada por *skip-connections* concatenadas com a saída das *up-convolutions*, resultando em uma máscara de segmentação mais precisa.

Figura 16 – Arquitetura da Rede U-net.



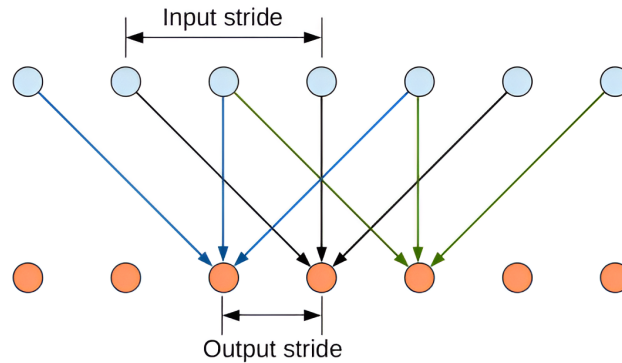
Fonte: Ronneberger, Fischer e Brox (2015).

Outro trabalho que norteou esta pesquisa iniciou-se com (CHEN et al., 2016), que propôs o modelo *DeepLab*. Este modelo aplicou convoluções *atrous* para manter o tamanho da entrada intacto enquanto extrai características de diferentes resoluções. Este tipo de convolução introduz esparsidade nos filtros, fazendo-os agir sobre uma área maior da imagem de entrada. A Figura 17 ilustra um *kernel* de convolução de tamanho 3x3 que produz uma saída de mesmo tamanho ao introduzir esparsidade nos pesos.

Posteriormente, o modelo *DeepLab* foi incrementado em (CHEN et al., 2018a) que adicionou o módulo *ASPP* (*atrous spatial pyramid pooling*) que consiste de convoluções *atrous* executadas paralelamente com diferentes taxas de dilatação aplicadas no mapa de características de entrada. Este módulo ajuda a rede a considerar diferentes tamanhos de objetos para uma mesma classe.

Já o trabalho iniciado por Howard et al. (2017) propõe a criação de redes eficientes voltadas para aplicações móveis, onde existem limitações de recursos como memória e poder

Figura 17 – Filtro de Convolução Atrous.

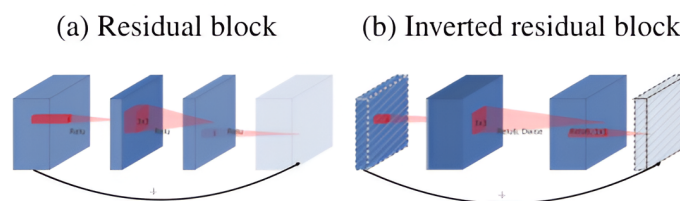


Fonte: [Chen et al. \(2016\)](#).

computacional. Nele foi analisado o uso das chamadas *depthwise separable convolutions*, que são um tipo de convolução subdivididas em: uma convolução *depthwise* (um filtro para cada canal) seguida por uma convolução *pointwise* (convolução convencional com kernel de 1x1). Este tipo inovador de convolução possibilitou a diminuição da quantidade de parâmetros do modelo, bem como a execução de menos operações de soma e multiplicação de matrizes pelo custo de 1% a menos de acurácia em seus testes. esta arquitetura ficou conhecida pelo nome de *MobileNet*.

Algum tempo depois, o trabalho de ([SANDLER et al., 2019](#)) seguiu na mesma linha de raciocínio e propôs a arquitetura *MobileNetV2*, cuja principal contribuição foi a análise da aplicação de gargalos residuais invertidos, que são uma modificação dos gargalos residuais propostos primeiramente em ([HE et al., 2015](#)). Enquanto o último segue uma lógica de conexões residuais entre as camadas de maior número de canais, os gargalos residuais invertidos realizam a conexão entre as camadas de menor número de canais. A Figura 18 ilustra a diferença entre os dois tipos de gargalos. Este tipo de estrutura tem o objetivo de comprimir a informação e em seguida expandi-la, obtendo apenas *features* essenciais para a tarefa, sendo que o gargalo invertido consome menos recursos computacionais e proporciona um modelo com menor quantidade de parâmetros treináveis.

Figura 18 – Gargalos Residuais.



Fonte: [Sandler et al. \(2019\)](#).

Na mesma direção, Tan e Le (2019) em seu trabalho, gerou uma arquitetura compacta que combina os gargalos residuais invertidos (SANDLER et al., 2019) com os blocos de compressão-expansão (HU; SHEN; SUN, 2018). Este trabalho inovou ao propor um método de escalabilidade de redes convolucionais, relacionando a resolução de entrada com a quantidade de canais de maneira proporcional através da escalabilidade composta (*compound scaling*). Dessa maneira foram geradas modelos gradativamente maiores a partir de um modelo-base menor.

3.2 Segmentação Semântica de Órgãos em Risco

Dentre os trabalhos relacionados que buscaram propor metodologias específicas para a segmentação de Órgãos em Risco, uma das abordagens é aproveitar a natureza tridimensional do exame de tomografia computadorizada. Em seu trabalho, Trullo et al. (2017) realizou a segmentação de diversos ORs utilizando uma arquitetura de rede neural completamente convolucional 3D conhecida como *Sharp Mask* (PINHEIRO et al., 2016). Essa arquitetura substitui o *upsampling* bilinear original por convoluções transpostas e adiciona o uso de *Conditional Random Fields* (CRF) (SUTTON; MCCALLUM, 2010) à saída da rede. Roth et al. (2018) analisou a utilização de redes V-net (MILLETARI; NAVAB; AHMADI, 2016) em conjunto com técnicas de pirâmide de resolução (ADELSON et al., 1984) e autocontexto (TU; BAI, 2010) para extrair não só a informação contextual multi-resolução, mas também a informação conjunta dos 3 eixos da TC, obtendo bons resultados na segmentação de órgãos.

Mais tarde, (WANG et al., 2019) aprimorou esta mesma arquitetura adicionando convoluções dilatáveis, pós-processamento e mais um estágio de pirâmide ao modelo anterior. Seu trabalho figura o estado da arte na segmentação de Órgãos em Risco, tendo obtido um Dice Score de 92,17% para a segmentação de traqueia. Ainda se aproveitando do contexto tridimensional da informação contida nos exames, (WANG et al., 2020) desenvolveu um método que, apesar de ser primariamente bidimensional, inclui informações da vizinhança dos *pixels* como segunda entrada para a rede neural *Xception* (CHOLLET, 2017). Além disso, esta rede foi modificada para aplicar convoluções *atrous* e um módulo *ASPP* também foi adicionado.

Outro tipo de abordagem na segmentação de Órgãos em Risco, é a utilização apenas dos *slices* bidimensionais gerados pelo exame, partindo do pressuposto de que a informação contida naquela fatia do exame é suficiente para a segmentação do órgão em questão. Dessa maneira, Feng et al. (2020) explorou como a informação contextual de diferentes camadas de uma rede neural poderiam ajudar na tarefa de segmentação semântica de órgãos em risco. O modelo proposto, nomeado *CPFNet*, trata-se de uma rede U-net onde as *skip connections* foram substituídas por um módulo que utiliza *upsampling*

para concatenar mapas de características de camadas mais abaixo e convoluções dilatáveis para obter *features* de diferentes resoluções. Além disso, em vez de convoluções tradicionais, foram utilizadas as convoluções separáveis propostas por (CHOLLET, 2017), que possuem menor quantidade de parâmetros treináveis.

Outra técnica comum no aprendizado profundo, é o uso de redes generativas adversárias (GAN) para treinar extratores de características e usá-los posteriormente em conjunto com uma rede de segmentação. Na segmentação de ORs, o trabalho que aplicou esta técnica foi o de (ZHAO; CHEN; LU, 2020), que combinou uma GAN pré-treinada, com uma rede U-net. Nesse contexto, a rede generativa adversária funcionou como um suplemento de informações relevantes para a rede U-net, obtendo resultados relevantes na segmentação de órgãos da região torácica.

Um novo tipo de modelo inovador adicionou mais uma possibilidade às redes neurais convolucionais: os Transformers. Inicialmente propostos para uso em processamento de linguagem natural, tratam-se de um tipo de camada de atenção em redes neurais, e foram adaptados para o uso em imagens no trabalho de Vaswani et al. (2017). Alguns trabalhos aplicaram este conceito em suas arquiteturas, como em (YAN et al., 2022), que modificou uma arquitetura U-net, adicionando *transformers* na estrutura de gargalo da rede, possibilitando o estudo sobre o desempenho deste tipo de modelo aplicado às imagens médicas.

A maioria dos trabalhos citados realizaram experimentos em uma base de dados em comum: a SEGTHOR, que foi disponibilizada em (LAMBERT et al., 2019) visando servir como base comparativa entre metodologias de segmentação de órgãos em risco. O trabalho reuniu tomografias computadorizadas de 60 pacientes em tratamento de câncer de pulmão ou linfoma de hodgkin, onde foram realizados experimentos com versões levemente modificadas da U-net (RONNEBERGER; FISCHER; BROX, 2015) que serviram de experimento base para trabalhos futuros. A Tabela 3 sumariza as metodologias estudadas na revisão de literatura realizada durante esta pesquisa, destacando seus respectivos *Dice Score* (métrica de avaliação mais comum na área) para a segmentação de traqueia e as bases de dados utilizadas.

Outra pesquisa que buscou montar um *dataset* referência foi o de (PODOBNIK et al., 2023). Os autores reuniram imagens de cabeça e pescoço de 56 pacientes que foram submetidos a tomografia computadorizada e ressonância magnética. Para cada paciente, foram obtidas segmentações de referência de até 30 Órgãos em Risco (OARs) por especialistas que realizaram a anotação manual dos *pixels* da imagem. Além disso, foram fornecidos resultados preliminares treinando a rede *U-Net* (RONNEBERGER; FISCHER; BROX, 2015) no conjunto de dados, que ficou conhecido como HaN-Seg.

Os trabalhos elencados na Tabela 3 obtiveram bom desempenho, entretanto, possuem uma característica em comum: o alto consumo de recursos computacionais.

Tabela 3 – Resumo dos Trabalhos Relacionados.

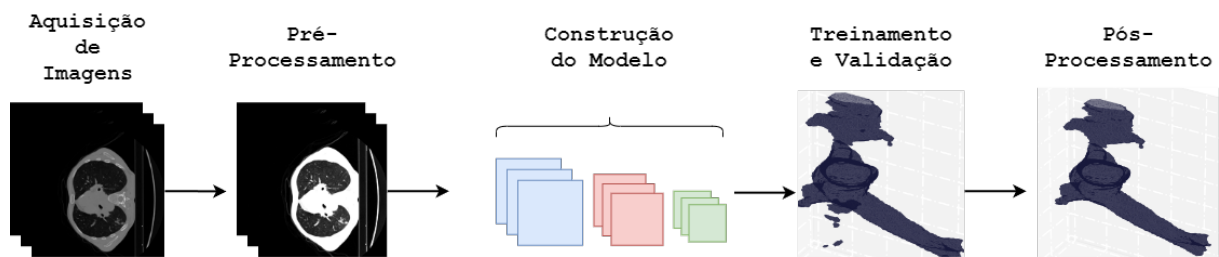
Paper	Method	DATASET	(%) Dice - traqueia	#Params
(TRULLO et al., 2017)	SharpMask	~	83.00	~
(WANG et al., 2019)	3D U-net	SEGTHOR	92,17	120M
(LAMBERT et al., 2019)	Simplified U-net	SEGTHOR	86.00	19M
(WANG et al., 2020)	U-net 2.5D	SEGTHOR	92,56	23M
(ZHAO; CHEN; LU, 2020)	U-net + GAN	SEGTHOR	88.30	~
(FENG et al., 2020)	U-net + SPP	SEGTHOR	89	26M
(DINIZ et al., 2021)	Res-unet + Atlas	LSTSC	~	~
(YAN et al., 2022)	U-net + Transformer	SEGTHOR	91,76	41,5M
(PODOBNIK et al., 2023)	U-net	HaN-Seg	~	24M

Isso acontece, pois todos eles possuem mais de 20 milhões de parâmetros treináveis, o que aumenta a demanda por memória e processamento destes métodos. Este trabalho propõe uma metodologia que equilibra custo computacional e qualidade na segmentação automática de traqueia.

4 Materiais e Método

Este capítulo apresenta detalhes sobre cada uma das etapas do método desenvolvido assim como sua análise para segmentação automática de traqueia com aprendizado profundo (*deep learning*), sendo elas: a aquisição das imagens; pré-processamento; treinamento da CNN, segmentação e pós-processamento. A Figura 19 resume o fluxo geral seguido pelo método proposto.

Figura 19 – Diagrama do Método Proposto.



Fonte: Elaborado pelo autor.

4.1 Aquisição da Imagens

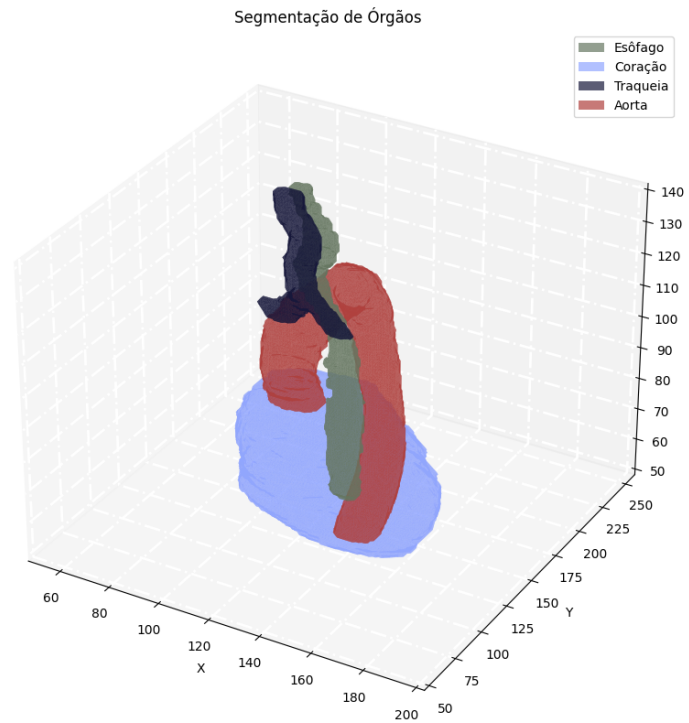
As imagens utilizadas para treinamento do modelo foram obtidas do conjunto de dados SEGTHOR (LAMBERT et al., 2019), que contém 60 volumes de TC, obtidos de pacientes com Câncer de Pulmão ou Linfoma de Hodgkins.

Dos 60 pacientes, 40 foram disponibilizados publicamente pelos autores do conjunto de dados. A partir dos volumes gerados pelos exames, foi possível extrair 7420 cortes transversais, dos quais 1987 imagens contém porções da traqueia, enquanto as 5433 restantes não. Os dados também incluem marcações manuais dos ORs feitas por médicos usadas como *labels* para o treinamento da rede neural. Os órgãos demarcados no dataset são: Esôfago, Traquéia, Coração e Aorta. A Figura 20 representa em 3 dimensões as marcações fornecidas pelos médicos, sendo a traqueia destacada pela cor azul-escuro.

4.2 Pré-Processamento

Adotar medidas de pré-processamento nas imagens antes de avançar para as etapas de treinamento do modelo e segmentação das amostras é essencial para aprimorar a qualidade dos dados de entrada. Nesta etapa busca-se minimizar ruídos, melhorar o contraste e corrigir possíveis defeitos presentes nas imagens, contribuindo para uma

Figura 20 – Demarcação de Órgãos Feita por Profissional.



Fonte: Elaborado pelo autor.

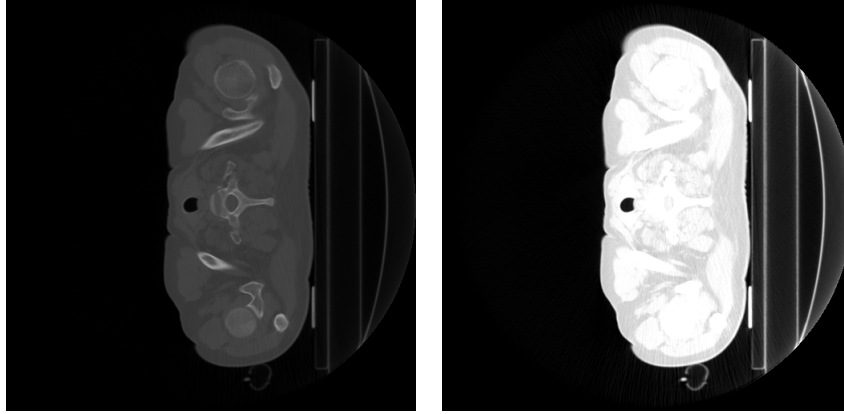
representação mais discriminativa dos objetos contidos na imagem. As subseções seguintes detalham os pré-processamentos executados sobre as imagens.

4.2.1 Janelamento de Intensidade

Uma das hipóteses deste trabalho consiste em avaliar o impacto do janelamento de intensidades como pré-processamento para a segmentação de traqueia. Visando encontrar de maneira automática um intervalo próximo do ótimo para a detecção de traqueia. Foi realizada otimização bayesiana durante 300 épocas utilizando a biblioteca de otimização de hiper-parâmetros *Optuna* (AKIBA et al., 2019). No caso do janelamento de intensidade, os hiper-parâmetros são o limite inferior e superior. Estes foram buscados nos intervalos entre $[-1000, 0]$ e $[10, 100]$, considerando apenas os números inteiros. A otimização resultou em uma janela com limite inferior de -800 e limite superior de 30 .

Em seguida, o janelamento de intensidade foi aplicado sobre os *slices*. A Figura 21 exemplifica a aplicação da janela sobre um recorte. À esquerda, encontra-se o recorte original. Já à direita, está representado o recorte após aplicação da janela $[-800, 30]$ encontrada no experimento. Nota-se o destaque da região de traqueia na imagem após o pré-processamento.

Figura 21 – Aplicação do Janelamento de Intensidade.



Fonte: Elaborado pelo autor.

4.2.2 Redimensionamento e Normalização

As imagens adquiridas do *dataset* SEGTHOR possuem tamanho 512×512 , ocupando considerável espaço na memória da GPU. Dada as limitações de memória do hardware, redimensionar as imagens para um tamanho menor fez com que fosse possível a execução do treinamento da rede neural em tempo hábil. Portanto, o tamanho selecionado foi o de 256×256 , visto que representa um equilíbrio entre a quantidade de dados e a qualidade de representação dos órgãos.

Em seguida foi executada a normalização dos *pixels*, cujos valores de intensidade estão contidos no intervalo $[-800, 30]$, encontrado via otimização de hiper-parâmetros via algoritmo TPE (*Tree-structured Parzen Estimator*) (AKIBA et al., 2019). Esta etapa facilita o aprendizado da rede neural, reduzindo a variação de intensidades entre imagens.

A normalização é muito comum em algoritmos de *Deep Learning* (KOO; CHA, 2017), (PIRES et al., 2020), (LUBANA; DICK; TANAKA, 2021) e serve para homogeneizar os dados de entrada, de maneira que o modelo de aprendizado possa entender as relações entre *pixels* de maneira mais generalizada. Existem diferentes tipos de normalização, algumas das mais comuns são: a normalização por média, média e desvio padrão e normalização mínimo-máximo. O tipo de normalização a ser utilizada neste trabalho é a normalização por média e desvio padrão, dada pela Equação 4.1.

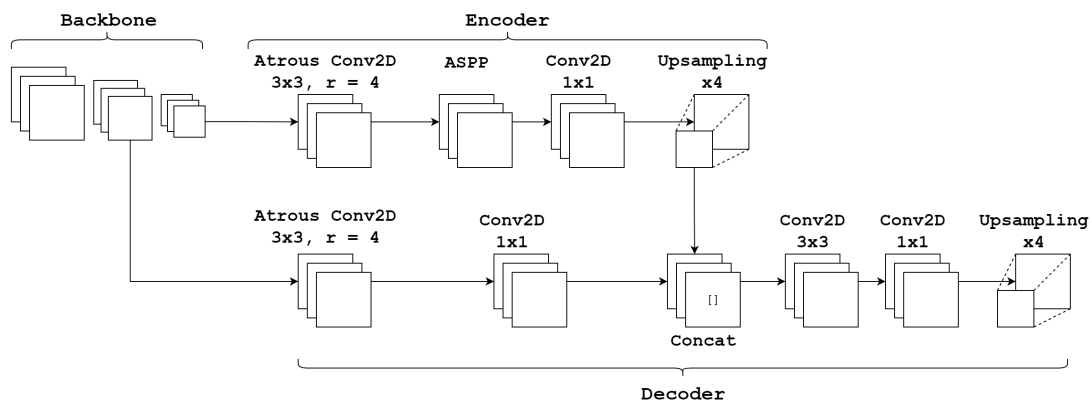
$$x' = \frac{x - x_{mean}}{std(x)} \quad (4.1)$$

4.3 Construção do Modelo

A primeira hipótese norteadora desta pesquisa supõe que a utilização de convoluções *atrous* (HOLSCHNEIDER et al., 1990) possa ajudar a construir um modelo compacto

para tarefas de segmentação de traqueia. Para por esta hipótese em prova, a estrutura *encoder/decoder* da rede *DeepLabV3+* (CHEN et al., 2018b), foi utilizada como base para a construção da arquitetura do modelo proposto, já que ela utiliza as convoluções *atrous* e o módulo ASPP *Atrous Spatial Pyramid Pooling* (HE et al., 2014) para promover uma fusão de características de várias resoluções. A Figura 22 representa o diagrama arquitetural da rede *DeepLabV3+* utilizada.

Figura 22 – Diagrama da Rede *DeepLabV3+*.



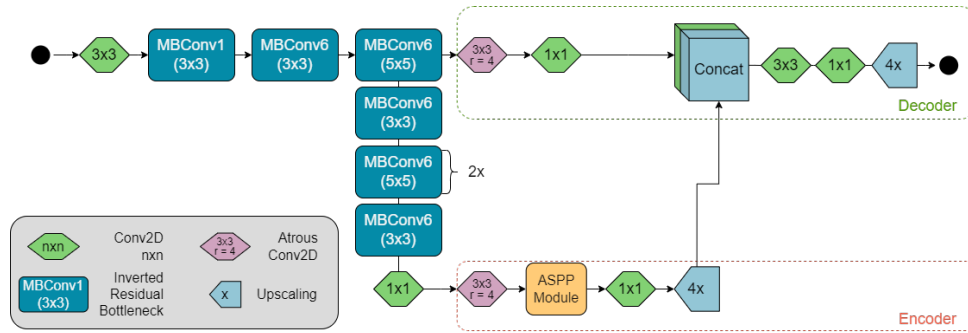
Fonte: Elaborado pelo autor.

Já a segunda hipótese levanta a possibilidade de utilizar a rede *EfficientNet* como *backbone* em conjunto com a rede *DeepLab* para executar a tarefa de segmentação semântica de traqueia. Por isso, o modelo proposto para a segmentação semântica da área da traqueia consiste em uma arquitetura de rede neural baseada em uma combinação das redes *DeepLabV3+* (CHEN et al., 2018b) e *EfficientNet* (TAN; LE, 2019). A escolha foi realizada baseando-se na capacidade da *EfficientNet* de ser um modelo de aprendizado profundo escalável, robusto e compacto em termos de parâmetros, que supera a *ResNet* em *benchmarks* recentes no conjunto de dados *ImageNet* (DENG et al., 2009), um dos principais *datasets* para comparação entre modelos de CNN.

Para a construção do modelo foi realizada uma modificação realizada na arquitetura *DeepLab* original: a substituição do *backbone ResNet* (HE et al., 2015) pela rede *EfficientNet* (TAN; LE, 2019), agregando mapas de características de camadas intermediárias, somando-os e interpolando-os de volta para a resolução de entrada, obtendo assim uma máscara de saída de segmentação. A Figura 23 proporciona uma visão geral da arquitetura do modelo.

Dessa maneira, este trabalho utiliza a estrutura de *encoder/decoder* para extrair características e restaurá-las para a resolução original. O modelo *EfficientNet* serve como *backbone* de extração de características, enquanto o módulo ASPP compõe o *encoder* do modelo e recebe como entrada os mapas de ativação da última camada convolucional da *EfficientNet*. As entradas do *decoder* são os mapas de ativação da terceira camada da *EfficientNet*. Elas são encaminhadas para o *decoder* composto por: uma convolução

Figura 23 – Visão Geral da Arquitetura do Modelo Proposto.



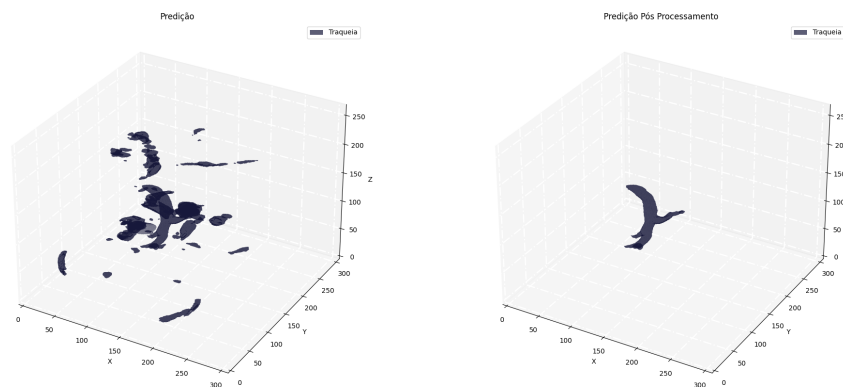
Fonte: Elaborado pelo autor.

separável *atrous*, seguida de uma camada convolucional 1×1 , cujo resultado é concatenado com a saída do *encoder* e enviado para uma convolução 3×3 seguida de outra 1×1 , que por fim, sofre *upsample*, retornando à resolução original da imagem. O resultado deste processo é a máscara de segmentação de órgãos.

4.4 Pós Processamento

Para eliminar os ruídos produzidos pelo modelo, foi aplicado o algoritmo de pós-processamento de componentes conectadas sobre o resultado da predição. Este algoritmo identifica os grupos de *voxels* vizinhos e retorna o grupo que possui maior quantidade de indivíduos. Dessa maneira foi possível eliminar da máscara de segmentação os *voxels* que não fazem parte da traqueia. A Figura 24 ilustra o funcionamento do pós-processamento, sendo que à esquerda está representado o resultado sem pós-processamento e à direita a máscara de segmentação 3D resultante da eliminação dos falsos positivos.

Figura 24 – Comparativo Após o Pós-Processamento.



Fonte: Elaborado pelo autor.

4.5 Considerações Finais

Neste capítulo foram detalhadas as etapas do método proposto para a segmentação automática da traqueia com aprendizado profundo. Isso inclui a aquisição das imagens do conjunto de dados SEGTHOR (Seção 4.1), o pré-processamento das imagens com janelamento de intensidade, redimensionamento e normalização (Seção 4.2), a construção e treinamento da rede neural convolucional com a combinação de *DeepLab V3+* e *EfficientNet* (Seção 4.3).

5 Resultados

Foram realizados experimentos para testar cada uma das hipóteses de pesquisas elencadas na Seção 1.1. Este capítulo apresenta os resultados obtidos em cada um destes experimentos.

5.1 Ambiente de Execução

O ambiente no qual os experimentos foram realizados consiste em um computador com processador Intel Core TMi5-10400F @ 2.90GHz e uma GPU Nvidia @RTX-3060 com 12GB de memória. Para a programação e implementação dos modelos de aprendizagem profunda, utilizou-se a linguagem de programação Python em conjunto com as bibliotecas: Tensorflow, Keras e Tensorflow Advanced Segmentation Models (KEZMANN, 2020).

5.2 Treinamento do Modelo com Convoluções Atrous e DeepLab

O treinamento do modelo *EfficientDeeplab* foi realizado utilizando os exames disponíveis na base de dados SEGTHOR. A separação entre treinamento, teste e validação seguiu o modelo de validação cruzada (*k-fold*) com $k = 3$. Cada *fold* está organizado da seguinte maneira: do total de 40 pacientes, são selecionados aproximadamente 65% para o treinamento, enquanto os 35% restantes são divididos igualmente entre validação e teste. Neste conjunto de dados as fatias que não contém traqueia são a maioria esmagadora. Por isso, elas foram excluídas das fases de treinamento e validação para que fosse obtido um *dataset* relativamente balanceado. Por outro lado, a etapa de testes é executada com todo o volume dos pacientes.

Foram executadas 100 épocas de treinamento, utilizando o otimizador SGD (ROBBINS; MONRO, 1951) em conjunto com o algoritmo de *backpropagation* (RUMELHART; HINTON; WILLIAMS, 1986) e a função de *loss binary crossentropy* para atualizar os neurônios da rede. Ao longo das 100 épocas, foram coletadas as métricas: *Dice Score* e *Jaccard Score*, como definidas na Seção 2.1.3. Ambas são comumente utilizadas para avaliar a sobreposição de duas máscaras de segmentação, neste caso a predita pelo modelo e a anotada pelos médicos. Quanto mais próximas estas medidas estiverem de 1, melhor a qualidade da segmentação de órgãos produzida.

O termo *EfficientNet* denomina uma família de redes derivadas de um modelo base mais compacto, chamado de B0. As redes derivadas variam de B1 a B7, com diferenças principalmente na resolução de entrada e no número de canais e camadas em cada estágio,

de maneira que, quanto maior o número no nome da rede, maior a quantidade de parâmetros treináveis. Para encontrar o modelo mais adequado para atuar em conjunto com a rede *DeepLabV3+* no problema da segmentação de traqueia, o treinamento foi executado para os modelos de B0, B1 e B2. Os modelos a partir de B3 não foram testados devido a limitações de memória. A Tabela 4 sumariza os *dice scores* coletados durante a etapa de validação para cada um dos experimentos realizados seguindo o esquema de validação cruzada.

Tabela 4 – Experimentos com Diferentes Backbones.

Model	Fold:	1	2	3	Mean	Std
EfficientDeeplab - B0	Dice	92,39	93,61	92,93	92,98	0,5
EfficientDeeplab - B1	Dice	93,35	93,24	93,25	93,28	0,04
EfficientDeeplab - B2	Dice	93,39	93,40	93,89	93,56	0,02

O modelo *EfficientDeepLab* que obteve o melhor desempenho foi o que utilizou a rede *EfficientNet B2* como backbone, tendo atingido um *Dice Score* de 93,31%, com desvio padrão de 0,09%. O alto *dice score* sugere ótima capacidade do modelo em segmentar traqueia, enquanto o baixo desvio padrão observado valida sua capacidade de generalização para exemplos ainda não observados. Isso prova verdadeira a hipótese 2, levantada na Seção 1.1. Além disso, observa-se que, dentre os modelos testados, a rede *EfficientNet B2* foi a que melhor trabalhou em conjunto com a arquitetura *DeepLabV3+*.

Para testar a hipótese 1, foram executados treinamentos em outras duas arquiteturas de redes neurais que não utilizam convoluções *atrous*: as redes U-net (RONNEBERGER; FISCHER; BROX, 2015) e SUnet (LAMBERT et al., 2019). Elas foram escolhidas, pois já foram testadas no artigo da própria base de dados SEGTHOR (LAMBERT et al., 2019). Para estes treinamentos foram utilizadas as mesmas configurações e mesma quantidade de épocas que o treinamento das *EfficientDeeplab*. Além disso, foi coletada a quantidade de parâmetros presente em cada um dos modelos testados. Os resultados da comparação estão elencados na Tabela 5.

Tabela 5 – Comparação de Desempenho com Modelos U-net.

Modelo	Dice Médio		Jaccard Médio		Sens. Média		#Params
EfficientDeeplab - B0	92,98	± 0,5	86,20	± 0,003	96,43	± 0,03	6.2 M
EfficientDeeplab - B1	93,28	± 0,04	87,79	± 0,002	99,46	± 0,004	8.7 M
EfficientDeeplab - B2	93,56	± 0,02	88,21	± 0,000	98,94	± 0,01	9.8 M
SUNET	92,73	± 1,14	86,54	± 0,003	96,57	± 0,02	4.8 M
UNET	93,73	± 0,1	88,10	± 0,01	98,69	± 0,01	19 M

Nota-se que a rede SUNET é a que possui menor quantidade de parâmetros e, ao mesmo tempo, o pior *dice score* médio. Enquanto isso, a rede *Unet* obteve a melhor métrica, porém maior quantidade de parâmetros. O modelo *Efficient Deeplab - B2* obtém um excelente equilíbrio entre *dice score* e quantidade de parâmetros, provando verdadeira

a hipótese levantada sobre o uso de convoluções *atrous* para a construção de um modelo compacto e que seja capaz de generalização.

5.3 Utilização de Janelamento de Intensidade

Para avaliar a hipótese 3, o melhor modelo obtido durante os experimentos da Seção 5.2 (*Efficient Deeplab - B2*) foi utilizado em dois testes. O primeiro ocorreu sem a introdução do janelamento de intensidade como pré-processamento. Já o segundo teste utilizou este pré-processamento com valores determinados por otimização bayesiana, como explicado na Seção 4.2.1. O resultado dos testes está resumido na Tabela 6. Observou-se que a adição do pré-processamento não ajudou a melhorar a qualidade da segmentação, chegando a diminuir o *dice score*, *jaccard score* e *sensibilidade* quando comparada ao experimento sem janelamento, tornando falsa a hipótese de pesquisa número 3.

Tabela 6 – Comparação de Janelamento de Intensidade com EfficientDeeplab - B2.

Experimento	Dice Médio	Jaccard Médio	Sens. Média
Sem Janelamento	93,56	88,31	98,94
Com Janelamento	93,31	88,21	98,23

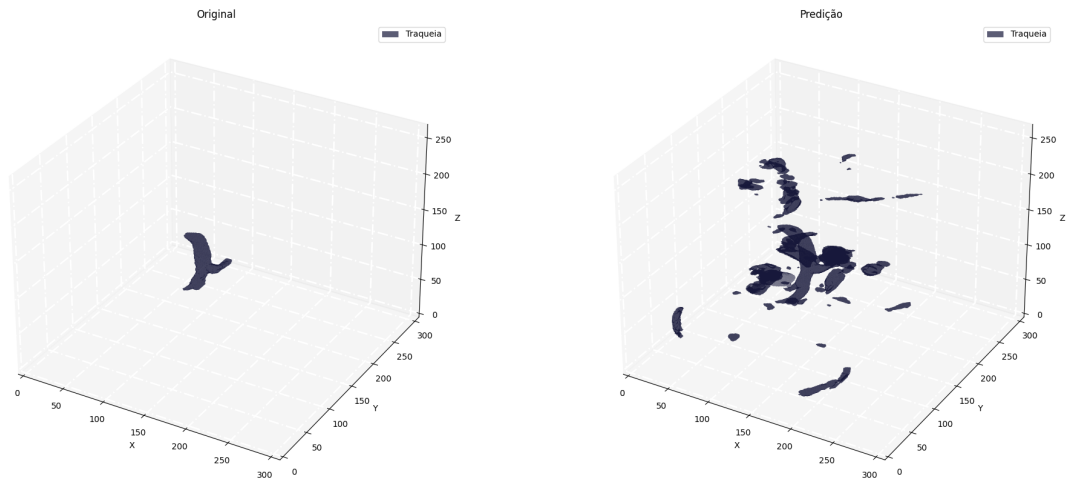
5.4 Estudo de Casos com Volumes Inteiros

Os pacientes separados para testes não tiveram os *slices* sem traqueia descartados, portanto, foi utilizado o volume inteiro obtido do exame de TC. Isto representa um caso de uso real do modelo, onde não se sabe quais *slices* contém traqueia ou não. Neste teste foram obtidas métricas inferiores aos resultados encontrados na validação, chegando a 39,32% de *dice score*, comparado aos 93,56% da fase de validação. Após cuidadosa análise, observou-se que a diminuição do *score* deve-se à presença de falsos positivos na máscara de segmentação de saída. A Figura 25 exibe à esquerda a máscara de segmentação proposta por médicos, e, à direita, a máscara obtida pela segmentação automática.

Isto ocorre porque algumas regiões da TC que não exibem traqueia, possuem características texturais e formas semelhantes ao órgão, levando o modelo a falsos positivos na predição. Na Figura 26 é possível observar uma fatia na qual não há presença da traqueia. No entanto, há outra estrutura corporal com forma e textura semelhantes que confundem o modelo, resultando na marcação equivocada dessa região como traqueia.

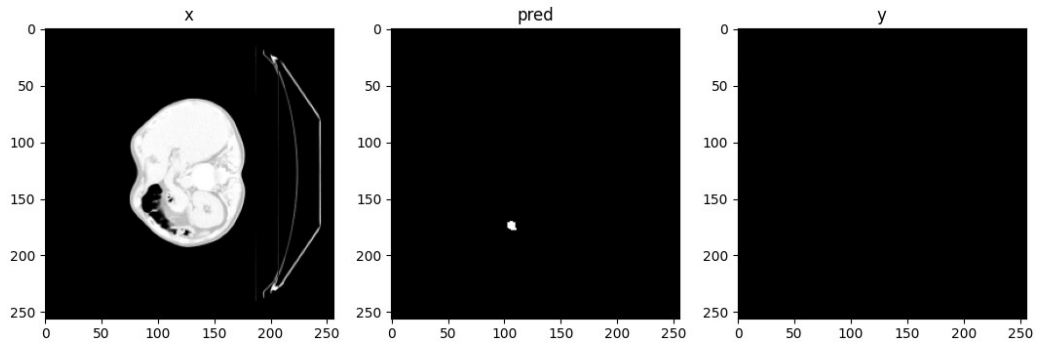
Já nas regiões em que há a presença do órgão, o modelo realiza a segmentação de forma correta, como ilustrado na Figura 27, onde a *EfficientDeeplab-B2* foi capaz de identificar corretamente a forma e a localização do órgão. Em outras regiões, a traqueia divide-se em dois tubos. Nestes casos o modelo também performou de maneira satisfatória, como exemplificado na Figura 28.

Figura 25 – Falsos Positivos na Segmentação.



Fonte: Elaborado pelo autor.

Figura 26 – Falso Positivo em Fatia.



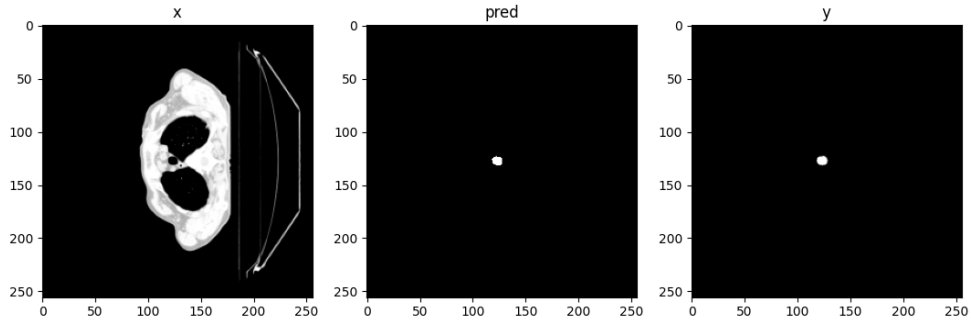
Fonte: Elaborado pelo autor.

Para eliminar os falsos positivos, foi realizada a aplicação do pós-processamento de componentes conectadas descrito na Seção 4.4. Observou-se melhora significativa do *dice score* para a segmentação de traqueia ao utilizar a rede *EfficientDeeplab-B2* seguida pelo pós-processamento de componentes conectadas, provando verdadeira a hipótese de pesquisa número 4. Os resultados estão elencados na Tabela 7.

Tabela 7 – Comparação de Resultados com Volumes Inteiros.

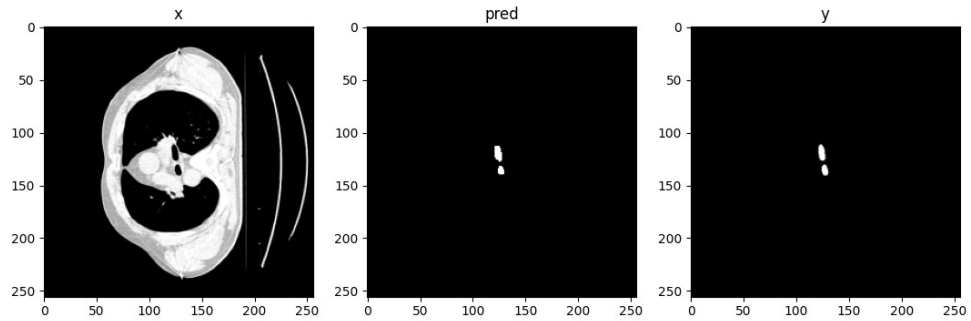
Modelo	Dice Médio	Dice Médio Pós-Proc.	Jaccard Médio	Jaccard Médio Pós-Proc.	Sens. Média	Sens. Média Pós-Proc.
EfficientDeeplab - B0	50,73	82,21	49,08	80,83	93,90	93,56
EfficientDeeplab - B1	42,64	82,08	40,94	80,51	92,73	92,49
EfficientDeeplab - B2	44,75	81,41	43,20	80,05	94,44	94,23

Figura 27 – Porção Central da Traqueia.



Fonte: Elaborado pelo autor.

Figura 28 – Divisão da Traqueia.



Fonte: Elaborado pelo autor.

5.5 Considerações

Nesta seção foram apresentados os resultados obtidos com a arquitetura proposta. A rede *EfficientDeeplab-B2* ofereceu os melhores resultados de validação, enquanto o modelo *B0* obteve as melhores métricas de segmentação em volumes inteiros, pois gera menos falsos positivos. Os resultados obtidos estão próximos aos de outros trabalhos encontrados na literatura, com a vantagem de possuir até 20 vezes menos parâmetros que o melhor modelo encontrado na literatura (WANG et al., 2019).

A arquitetura *EfficientDeeplab* incrementa o módulo *SPP* utilizado em (FENG et al., 2020) ao substituir as suas convoluções convencionais por convoluções *atrous*, que ocupam menos espaço em memória e alcançam resultados similares. Em relação ao pós-processamento, trabalhos como o de (TRULLO et al., 2017) utilizam-se de *conditional random fields* para redução de falsos positivos. Apesar de obter bons resultados, esta técnica adiciona mais complexidade e mais parâmetros ao modelo. Neste trabalho, a utilização do algoritmo de componentes conectadas elimina a necessidade de introduzir mais parâmetros na rede. Além disso, a utilização da rede *Deeplab* (CHEN et al., 2018b) torna a estrutura *encoder/decoder* mais compacta quando comparada às redes baseadas em *U-Net*, devido ao uso das convoluções *atrous* e *upsampling bilinear*, em contraste com as convoluções

transpostas utilizadas na rede *U-net*. Essa economia considerável de parâmetros resulta em uma rede neural compacta e robusta, que consome menos recursos computacionais que modelos semelhantes sem perdas consideráveis na qualidade de segmentação, tornando o método aplicável em ambientes com limitação de recursos. A Tabela 8 realiza uma comparação entre os principais trabalhos relacionados e a metodologia proposta.

Tabela 8 – Comparação com a Literatura.

Paper	Method	DATASET	(%) Dice - traqueia	#Params
(TRULLO et al., 2017)	SharpMask	~	83.00	~
(WANG et al., 2019)	3D U-net	SEGTHOR	92,17	120M
(LAMBERT et al., 2019)	Simplified U-net	SEGTHOR	86.00	19M
(WANG et al., 2020)	U-net 2.5D	SEGTHOR	92,56	23M
(ZHAO; CHEN; LU, 2020)	U-net + GAN	SEGTHOR	88.30	~
(FENG et al., 2020)	U-net + SPP	SEGTHOR	89	26M
(YAN et al., 2022)	U-net + Transformer	SEGTHOR	91,76	41,5M
(PODOBNIK et al., 2023)	U-net	HaN-Seg	~	24M
Este Trabalho	EfficientDeepLab - B0	SEGTHOR	82,21	6M

Por fim, vale destacar que o desenvolvimento deste trabalho resultou em duas produções acadêmicas: (FERNANDES et al., 2023) e (FERNANDES et al., 2024), respectivamente intituladas “EfficientDeepLab For Automated Trachea Segmentation On Medical Images” e “Eff-Unet For Trachea Segmentation On CT Scans”. Ambas as publicações foram concebidas durante etapas intermediárias do desenvolvimento e pesquisa deste projeto de mestrado.

6 Conclusão

Este trabalho propôs um método automatizado para segmentação da traqueia em imagens de TC usando a arquitetura de rede neural *EfficientDeeplab*. Os resultados obtidos são promissores e competitivos. O uso de convoluções dilatadas (*atrous*) e a versão *atrous* do agrupamento espacial de pirâmide permitiram a reconstrução precisa das máscaras de segmentação. A utilização eficiente de recursos computacionais foi possível ao adotar a arquitetura *EfficientNet* como *backbone* para extração de características, resultando em um modelo com apenas 6 milhões de parâmetros, que pode ser utilizado em sistemas embarcados de visão computacional para medicina por ser consideravelmente menor do que os métodos propostos na literatura atual para segmentação da traqueia com aprendizado profundo. O pós-processamento baseado em componentes conectadas melhorou a aplicação da rede neural em volumes inteiros de TC, reduzindo a quantidade de falsos positivos gerados pelo modelo, sendo, portanto, uma técnica adequada para utilização em conjunto com a segmentação via aprendizado profundo.

Por fim, há espaço para melhorias futuras, como a adição de novas técnicas de pré-processamento e a agregar informações 3D sem haver aumento do consumo de recursos. Outra possibilidade é integrar uma arquitetura para redução de falsos positivos que consiga ser treinada em conjunto com o modelo *EfficientDeeplab*. Em suma, a metodologia proposta pode ajudar na obtenção de um planejamento radioterápico mais rápido e, conseqüentemente, melhorar a qualidade de vida dos pacientes de câncer.

Referências

- ACKERMANS, L. L. G. C.; VOLMER, L.; WEE, L.; BRECHEISEN, R.; SÁNCHEZ-GONZÁLEZ, P.; SEIFFERT, A. P.; GÓMEZ, E. J.; DEKKER, A.; BOSCH, J. A. T.; DAMINK, S. M. W. O.; BLOKHUIS, T. J. Deep learning automated segmentation for muscle and adipose tissue from abdominal computed tomography in polytrauma patients. *Sensors*, v. 21, n. 6, 2021. ISSN 1424-8220. Disponível em: <https://www.mdpi.com/1424-8220/21/6/2083>. Citado na página 21.
- ADELSON, E. H.; ANDERSON, C. H.; BERGEN, J. R.; BURT, P. J.; OGDEN, J. M. Pyramid methods in image processing. *RCA engineer*, v. 29, n. 6, p. 33–41, 1984. Citado na página 35.
- AKIBA, T.; SANO, S.; YANASE, T.; OHTA, T.; KOYAMA, M. Optuna: A next-generation hyperparameter optimization framework. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. [S.l.: s.n.], 2019. Citado 2 vezes nas páginas 39 e 40.
- AMARI, S. A theory of adaptive pattern classifiers. *IEEE Transactions on Electronic Computers*, IEEE, n. 3, p. 299–307, 1967. Citado na página 24.
- BARAZZUOL, L.; COPPES, R. P.; LUIJK, P. van. Prevention and treatment of radiotherapy-induced side effects. *Molecular Oncology*, v. 14, n. 7, p. 1538–1554, 2020. Disponível em: <https://febs.onlinelibrary.wiley.com/doi/abs/10.1002/1878-0261.12750>. Citado na página 17.
- BURKOV, A. *Machine Learning Engineering*. True Positive Incorporated, 2020. ISBN 9781999579579. Disponível em: <https://books.google.com.br/books?id=HeXizQEACAAJ>. Citado na página 23.
- CHEN, L.-C.; PAPANDREOU, G.; KOKKINOS, I.; MURPHY, K.; YUILLE, A. L. *Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs*. 2016. Citado 3 vezes nas páginas 29, 33 e 34.
- CHEN, L.-C.; PAPANDREOU, G.; KOKKINOS, I.; MURPHY, K.; YUILLE, A. L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, v. 40, n. 4, p. 834–848, April 2018. ISSN 0162-8828. Disponível em: <https://doi.org/10.1109/TPAMI.2017.2699184>. Citado na página 33.
- CHEN, L.-C.; ZHU, Y.; PAPANDREOU, G.; SCHROFF, F.; ADAM, H. *Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation*. 2018. Citado 4 vezes nas páginas 15, 29, 41 e 48.
- CHOLLET, F. *Keras*. 2015. https://keras.io/api/layers/convolution_layers/depthwise_convolution2d/. Citado na página 28.
- CHOLLET, F. Xception: Deep learning with depthwise separable convolutions. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, 2017. p. 1800–1807. ISSN 1063-6919. Disponível em:

<<https://doi.ieeecomputersociety.org/10.1109/CVPR.2017.195>>. Citado 2 vezes nas páginas 35 e 36.

CONSIDERA, D. P.; MARCHIORI, E.; JR, A. S. S.; ZANETTI, G.; ESCUISSATO, D. L.; GASPARETTO, E. L.; NETO, C. d. A.; LIMA, R. d. S. L.; XAVIER, S. S.; PEDROSA, R. C. A tomografia computadorizada de alta resolução na avaliação da toxicidade pulmonar por amiodarona. *Radiologia Brasileira*, SciELO Brasil, v. 39, p. 113–118, 2006. Citado na página 20.

DANCE, D. R.; CHRISTOFIDES, S.; MAIDMENT, A. D.; MCLEAN, I. D.; NG, K. H. *Diagnostic Radiology Physics*. Vienna: INTERNATIONAL ATOMIC ENERGY AGENCY, 2014. (Non-serial Publications). ISBN 978-92-0-131010-1. Disponível em: <<https://www.iaea.org/publications/8841/diagnostic-radiology-physics>>. Citado na página 19.

DENG, J.; DONG, W.; SOCHER, R.; LI, L.-J.; LI, K.; FEI-FEI, L. Imagenet: A large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2009. p. 248–255. Citado na página 41.

DHAWAN, A.; BODANI, P.; GARG, V. Post processing of image segmentation using conditional random fields. In: *2019 6th International Conference on Computing for Sustainable Global Development (INDIACom)*. [S.l.: s.n.], 2019. p. 729–734. Citado na página 22.

DICE, L. R. Measures of the amount of ecologic association between species. *Ecology*, v. 26, n. 3, p. 297–302, 1945. Disponível em: <<https://esajournals.onlinelibrary.wiley.com/doi/abs/10.2307/1932409>>. Citado na página 22.

DINIZ, J.; FERREIRA, J.; SILVA, G.; QUINTANILHA, D.; SILVA, A.; PAIVA, A. Segmentação de coração em tomografias computadorizadas utilizando atlas probabilístico e redes neurais convolucionais. In: *Anais do XXI Simpósio Brasileiro de Computação Aplicada à Saúde*. Porto Alegre, RS, Brasil: SBC, 2021. p. 83–94. ISSN 2763-8952. Disponível em: <<https://sol.sbc.org.br/index.php/sbcas/article/view/16055>>. Citado 3 vezes nas páginas 13, 14 e 37.

EELBODE, T.; BERTELS, J.; BERMAN, M.; VANDERMEULEN, D.; MAES, F.; BISSCHOPS, R.; BLASCHKO, M. B. Optimization for medical image segmentation: Theory and practice when evaluating with dice score or jaccard index. *IEEE Transactions on Medical Imaging*, v. 39, n. 11, p. 3679–3690, 2020. Citado na página 22.

EVANS, E.; STAFFURTH, J. Principles of cancer treatment by radiotherapy. *Surgery (Oxford)*, v. 36, n. 3, p. 111–116, 2018. ISSN 0263-9319. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0263931917302715>>. Citado 2 vezes nas páginas 17 e 18.

FENG, S.; ZHAO, H.; SHI, F.; CHENG, X.; WANG, M.; MA, Y.; XIANG, D.; ZHU, W.; CHEN, X. Cpfnet: Context pyramid fusion network for medical image segmentation. *IEEE Transactions on Medical Imaging*, v. 39, n. 10, p. 3008–3018, 2020. Citado 5 vezes nas páginas 14, 35, 37, 48 e 49.

FERNANDES, A. G. S.; JUNIOR, G. B.; DINIZ, J. O. B.; SILVA, A. C.; MATOS, C. E. F. Efficientdeeplab for automated trachea segmentation on medical images. In:

- NALDI, M. C.; BIANCHI, R. A. C. (Ed.). *Intelligent Systems*. Cham: Springer Nature Switzerland, 2023. p. 154–166. ISBN 978-3-031-45389-2. Citado na página 49.
- FERNANDES, A. G. S.; JUNIOR, G. B.; DINIZ, J. O. B.; FERREIRA, M. M.; JUNIOR, J. R. D. R.; SILVA, M. L. J. D.; GONÇALVES, L. A. *Eff-Unet For Trachea Segmentation On CT Scans*. 2024. <https://link.springer.com/book/9783031606649>. Citado na página 49.
- FU, Y.; LEI, Y.; WANG, T.; CURRAN, W. J.; LIU, T.; YANG, X. A review of deep learning based methods for medical image multi-organ segmentation. *Physica Medica*, v. 85, p. 107–122, 2021. ISSN 1120-1797. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1120179721001848>. Citado na página 14.
- GONZALEZ, R. C.; WOODS, R. E. *Digital image processing*. Upper Saddle River, N.J.: Prentice Hall, 2008. ISBN 9780131687288 013168728X 9780135052679 013505267X. Disponível em: <http://www.amazon.com/Digital-Image-Processing-3rd-Edition/dp/013168728X>. Citado 3 vezes nas páginas 18, 19 e 20.
- GOODFELLOW, I.; POUGET-ABADIE, J.; MIRZA, M.; XU, B.; WARDE-FARLEY, D.; OZAIR, S.; COURVILLE, A.; BENGIO, Y. Generative adversarial nets. In: GHAHRAMANI, Z.; WELING, M.; CORTES, C.; LAWRENCE, N.; WEINBERGER, K. (Ed.). *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2014. v. 27. Disponível em: https://proceedings.neurips.cc/paper_files/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf. Citado na página 27.
- GOODFELLOW, I. J.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. Citado na página 25.
- HAYKIN, S. S. *Neural networks and learning machines*. Third. Upper Saddle River, NJ: Pearson Education, 2009. Citado 3 vezes nas páginas 23, 24 e 25.
- HE, K.; ZHANG, X.; REN, S.; SUN, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. In: *Computer Vision textendash ECCV 2014*. [S.l.]: Springer International Publishing, 2014. p. 346–361. Citado na página 41.
- HE, K.; ZHANG, X.; REN, S.; SUN, J. *Deep Residual Learning for Image Recognition*. 2015. Citado 4 vezes nas páginas 27, 29, 34 e 41.
- HECHT-NIELSEN, R. *Neurocomputing*. USA: Addison-Wesley Longman Publishing Co., Inc., 1989. ISBN 0201093553. Citado na página 23.
- HOLSCHNEIDER, M.; KRONLAND-MARTINET, R.; MORLET, J.; TCHAMITCHIAN, P. A real-time algorithm for signal analysis with the help of the wavelet transform. In: COMBES, J.-M.; GROSSMANN, A.; TCHAMITCHIAN, P. (Ed.). *Wavelets*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1990. p. 286–297. ISBN 978-3-642-75988-8. Citado 2 vezes nas páginas 15 e 40.
- 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. *CoRR*, abs/1704.04861, 2017. Disponível em: <http://dblp.uni-trier.de/db/journals/corr/corr1704.html#HowardZCKWWAA17>. Citado na página 33.

HU, J.; SHEN, L.; SUN, G. Squeeze-and-excitation networks. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2018. p. 7132–7141. Citado na página 35.

ISENSEE, F.; JÄGER, P. F.; FULL, P. M.; VOLLMUTH, P.; MAIER-HEIN, K. H. nnu-net for brain tumor segmentation. In: CRIMI, A.; BAKAS, S. (Ed.). *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Cham: Springer International Publishing, 2021. p. 118–132. ISBN 978-3-030-72087-2. Citado na página 22.

JACCARD, P. The distribution of the flora in the alpine zone.1. *New Phytologist*, v. 11, n. 2, p. 37–50, 1912. Disponível em: <<https://nph.onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-8137.1912.tb05611.x>>. Citado na página 22.

KAK, A. C.; SLANEY, M. *Principles of Computerized Tomographic Imaging*. Society for Industrial and Applied Mathematics, 2001. Disponível em: <<https://epubs.siam.org/doi/abs/10.1137/1.9780898719277>>. Citado na página 19.

KEZMANN, J.-M. *Tensorflow Advanced Segmentation Models*. [S.l.]: GitHub, 2020. <<https://github.com/JanMarcelKezmann/TensorFlow-Advanced-Segmentation-Models>>. Citado na página 44.

KIRKLAND, E. J. Bilinear interpolation. In: _____. *Advanced Computing in Electron Microscopy*. Boston, MA: Springer US, 2010. p. 261–263. ISBN 978-1-4419-6533-2. Disponível em: <https://doi.org/10.1007/978-1-4419-6533-2_12>. Citado na página 29.

KOO, K.-M.; CHA, E.-Y. Image recognition performance enhancements using image normalization. *Human-centric Computing and Information Sciences*, v. 7, n. 1, p. 33, Nov 2017. ISSN 2192-1962. Disponível em: <<https://doi.org/10.1186/s13673-017-0114-5>>. Citado na página 40.

LAMBERT, Z.; PETITJEAN, C.; DUBRAY, B.; RUAN, S. *SegTHOR: Segmentation of Thoracic Organs at Risk in CT images*. arXiv, 2019. Disponível em: <<https://arxiv.org/abs/1912.05950>>. Citado 8 vezes nas páginas 13, 14, 21, 36, 37, 38, 45 e 49.

LECUN, Y.; BOSER, B.; DENKER, J. S.; HENDERSON, D.; HOWARD, R. E.; HUBBARD, W.; JACKEL, L. D. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, v. 1, n. 4, p. 541–551, 1989. Citado 2 vezes nas páginas 26 e 27.

LEE, C.-T.; KABIR, T.; NELSON, J.; SHENG, S.; MENG, H.-W.; DYKE, T. E. V.; WALJI, M. F.; JIANG, X.; SHAMS, S. Use of the deep learning approach to measure alveolar bone level. *J. Clin. Periodontol.*, Wiley, v. 49, n. 3, p. 260–269, mar. 2022. Citado na página 21.

LONG, J.; SHELHAMER, E.; DARRELL, T. *Fully Convolutional Networks for Semantic Segmentation*. arXiv, 2014. Disponível em: <<https://arxiv.org/abs/1411.4038>>. Citado 2 vezes nas páginas 27 e 32.

LUBANA, E. S.; DICK, R.; TANAKA, H. Beyond batchnorm: Towards a unified understanding of normalization in deep learning. In: RANZATO, M.; BEYGELZIMER, A.; DAUPHIN, Y.; LIANG, P.; VAUGHAN, J. W. (Ed.). *Advances in Neural*

- Information Processing Systems*. Curran Associates, Inc., 2021. v. 34, p. 4778–4791. Disponível em: <https://proceedings.neurips.cc/paper_files/paper/2021/file/2578eb9cdf020730f77793e8b58e165a-Paper.pdf>. Citado na página 40.
- MARCHIORI, E.; GONÇALVES, C. M.; ESCUISSATO, D. L.; TEIXEIRA, K.-I.-S. S.; RODRIGUES, R.; BARRETO, M. M.; ESTEVES, M. Microlitíase alveolar pulmonar: achados na tomografia computadorizada de alta resolução do tórax em 10 pacientes. *Jornal Brasileiro de Pneumologia*, Sociedade Brasileira de Pneumologia e Tisiologia, v. 33, n. 5, p. 552–557, Sep 2007. ISSN 1806-3713. Disponível em: <<https://doi.org/10.1590/S1806-37132007000500010>>. Citado na página 20.
- MEHTA, S.; SUHAG, V.; SEMWAL, M.; SHARMA, N. Radiotherapy: Basic concepts and recent advances. *Medical Journal Armed Forces India*, v. 66, n. 2, p. 158–162, 2010. ISSN 0377-1237. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0377123710801327>>. Citado na página 17.
- MILLETARI, F.; NAVAB, N.; AHMADI, S.-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: *2016 Fourth International Conference on 3D Vision (3DV)*. [S.l.: s.n.], 2016. p. 565–571. Citado na página 35.
- NOEL, G.; ANTONI, D.; BARILLOT, I.; CHAUVET, B. Délinéation des organes à risque et contraintes dosimétriques. *Cancer Radiothérapie*, v. 20, p. S36–S60, 2016. ISSN 1278-3218. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1278321816301676>>. Citado na página 13.
- PINHEIRO, P. O.; LIN, T.-Y.; COLLOBERT, R.; DOLL’AR, P. Learning to refine object segments. In: LEIBE, B.; MATAS, J.; SEBE, N.; WELLING, M. (Ed.). *Computer Vision – ECCV 2016*. Cham: Springer International Publishing, 2016. p. 75–91. ISBN 978-3-319-46448-0. Citado na página 35.
- PIRES, I. M.; HUSSAIN, F.; GARCIA, N. M.; LAMESKI, P.; ZDRAVEVSKI, E. Homogeneous data normalization and deep learning: A case study in human activity classification. *Future Internet*, v. 12, n. 11, 2020. ISSN 1999-5903. Disponível em: <<https://www.mdpi.com/1999-5903/12/11/194>>. Citado na página 40.
- PODOBNIK, G.; STROJAN, P.; PETERLIN, P.; IBRAGIMOV, B.; VRTOVEC, T. Han-seg: The head and neck organ-at-risk ct and mr segmentation dataset. *Medical Physics*, v. 50, n. 3, p. 1917–1927, 2023. Disponível em: <<https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/mp.16197>>. Citado 3 vezes nas páginas 36, 37 e 49.
- PRINCE, S. *Computer Vision: Models Learning and Inference*. [S.l.]: Cambridge University Press, 2012. Citado na página 18.
- ROBBINS, H.; MONRO, S. A stochastic approximation method. *The Annals of Mathematical Statistics*, Institute of Mathematical Statistics, v. 22, n. 3, p. 400 – 407, 1951. Disponível em: <<https://doi.org/10.1214/aoms/1177729586>>. Citado 2 vezes nas páginas 24 e 44.
- RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: NAVAB, N.; HORNEGGER, J.; WELLS, W. M.; FRANGI, A. F. (Ed.). *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Cham: Springer International Publishing, 2015. p. 234–241. ISBN 978-3-319-24574-4. Citado 4 vezes nas páginas 27, 33, 36 e 45.

- ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, v. 65 6, p. 386–408, 1958. Disponível em: <<https://api.semanticscholar.org/CorpusID:12781225>>. Citado 2 vezes nas páginas 23 e 24.
- ROTH, H. R.; SHEN, C.; ODA, H.; SUGINO, T.; ODA, M.; HAYASHI, Y.; MISAWA, K.; MORI, K. A multi-scale pyramid of 3d fully convolutional networks for abdominal multi-organ segmentation. In: FRANGI, A. F.; SCHNABEL, J. A.; DAVATZIKOS, C.; ALBEROLA-LÓPEZ, C.; FICHTINGER, G. (Ed.). *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*. Cham: Springer International Publishing, 2018. p. 417–425. ISBN 978-3-030-00937-3. Citado na página 35.
- RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. *Nature*, v. 323, n. 6088, p. 533–536, Oct 1986. ISSN 1476-4687. Disponível em: <<https://doi.org/10.1038/323533a0>>. Citado 2 vezes nas páginas 24 e 44.
- RUSSELL, S.; NORVIG, P. *Artificial Intelligence: A Modern Approach*. 3rd. ed. USA: Prentice Hall Press, 2009. ISBN 0136042597. Citado na página 22.
- SAMET, H.; TAMMINEN, M. Efficient component labeling of images of arbitrary dimension represented by linear bintrees. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 10, n. 4, p. 579–586, 1988. Citado na página 22.
- SANDLER, M.; HOWARD, A.; ZHU, M.; ZHMOGINOV, A.; CHEN, L.-C. *MobileNetV2: Inverted Residuals and Linear Bottlenecks*. 2019. Citado 3 vezes nas páginas 28, 34 e 35.
- SEERAM, E. *Computed tomography : physical principles, clinical applications, and quality control*. Fourth edition. St. Louis, Missouri: Elsevier St. Louis, Missouri, 2016. ISBN 9780323312882; 0323312888. Citado na página 19.
- SHALEV-SHWARTZ, S.; BEN-DAVID, S. *Understanding Machine Learning - From Theory to Algorithms*. [S.l.]: Cambridge University Press, 2014. I-XVI, 1-397 p. ISBN 978-1-10-705713-5. Citado na página 23.
- SHIVHARE, P.; GUPTA, V. Review of image segmentation techniques including pre & post processing operations. *International Journal of Engineering and Advanced Technology*, Citeseer, v. 4, n. 3, p. 153–157, 2015. Citado na página 22.
- SIMONYAN, K.; ZISSERMAN, A. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. 2015. Citado na página 27.
- SUTTON, C.; MCCALLUM, A. *An Introduction to Conditional Random Fields*. 2010. Citado na página 35.
- TAN, M.; LE, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In: CHAUDHURI, K.; SALAKHUTDINOV, R. (Ed.). *Proceedings of the 36th International Conference on Machine Learning*. PMLR, 2019. (Proceedings of Machine Learning Research, v. 97), p. 6105–6114. Disponível em: <<https://proceedings.mlr.press/v97/tan19a.html>>. Citado 5 vezes nas páginas 15, 28, 29, 35 e 41.
- TEKATLI, H.; DUIJM, M.; HOOP, E. Oomen-de; VERBAKEL, W.; SCHILLEMANS, W.; SLOTMAN, B.; NUYTENS, J.; SENAN, S. Normal tissue complication probability modeling of pulmonary toxicity after stereotactic and hypofractionated radiation therapy

for central lung tumors. *International Journal of Radiation Oncology Biology Physics*, Elsevier, v. 100, n. 3, p. 738–747, 2018. Citado 2 vezes nas páginas 13 e 17.

THANOON, M. A.; ZULKIFLEY, M. A.; ZAINURI, M. A. A. M.; ABDANI, S. R. A review of deep learning techniques for lung cancer screening and diagnosis based on ct images. *Diagnostics (Basel)*, 2023. Citado na página 21.

TRULLO, R.; PETITJEAN, C.; RUAN, S.; DUBRAY, B.; NIE, D.; SHEN, D. Segmentation of organs at risk in thoracic ct images using a sharpmask architecture and conditional random fields. In: *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. [S.l.: s.n.], 2017. p. 1003–1006. Citado 4 vezes nas páginas 35, 37, 48 e 49.

TU, Z.; BAI, X. Auto-context and its application to high-level vision tasks and 3d brain image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 32, n. 10, p. 1744–1757, 2010. Citado na página 35.

VAIDYA, J. S. Principles of cancer treatment by radiotherapy. *Surgery (Oxford)*, v. 39, n. 4, p. 193–201, 2021. ISSN 0263-9319. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0263931921000363>>. Citado na página 17.

VASWANI, A.; SHAZEER, N.; PARMAR, N.; USZKOREIT, J.; JONES, L.; GOMEZ, A. N.; KAISER, L. u.; POLOSUKHIN, I. Attention is all you need. In: GUYON, I.; LUXBURG, U. V.; BENGIO, S.; WALLACH, H.; FERGUS, R.; VISHWANATHAN, S.; GARNETT, R. (Ed.). *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2017. v. 30. Disponível em: <https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf>. Citado na página 36.

VICENTE, A. d. O.; PENIDO, N. d. O.; YAMASHITA, H. K.; ALBERNAZ, P. L. M. Tomografia computadorizada no diagnóstico da otosclerose fenestral. *Revista Brasileira de Otorrinolaringologia*, ABORL-CCF Associação Brasileira de Otorrinolaringologia e Cirurgia Cérvico-Facial, v. 70, n. 1, p. 66–73, Jan 2004. ISSN 0034-7299. Disponível em: <<https://doi.org/10.1590/S0034-72992004000100012>>. Citado na página 20.

WANG, Q.; ZHAO, W.; ZHANG, C.; ZHANG, L.; WANG, C.; LI, Z.; CUI, S.; LI, G. 3d enhanced multi-scale network for thoracic organs segmentation. *SegTHOR@ ISBI*, v. 3, n. 1, p. 1–5, 2019. Citado 4 vezes nas páginas 35, 37, 48 e 49.

WANG, S.; CAO, S.; CHAI, Z.; WEI, D.; MA, K.; WANG, L.; ZHENG, Y. Conquering data variations in resolution: A slice-aware multi-branch decoder network. *IEEE Transactions on Medical Imaging*, v. 39, n. 12, p. 4174–4185, 2020. Citado 3 vezes nas páginas 35, 37 e 49.

WEBER, E.; VILENSKY, J.; CARMICHAEL, S.; LEE, K. *Netter's Concise Radiologic Anatomy E-Book*. Elsevier Health Sciences, 2014. (Netter Basic Science). ISBN 9780323221474. Disponível em: <<https://books.google.com.br/books?id=o63zAgAAQBAJ>>. Citado 2 vezes nas páginas 13 e 19.

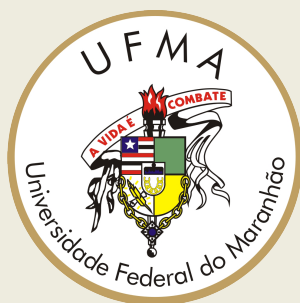
YAN, X.; TANG, H.; SUN, S.; MA, H.; KONG, D.; XIE, X. After-unet: Axial fusion transformer unet for medical image segmentation. In: *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. Los Alamitos, CA, USA: IEEE Computer Society, 2022. p. 3270–3280. Disponível em: <<https://>>

[//doi.ieeecomputersociety.org/10.1109/WACV51458.2022.00333](https://doi.ieeecomputersociety.org/10.1109/WACV51458.2022.00333)>. Citado 4 vezes nas páginas 14, 36, 37 e 49.

YEUNG, S. *Tutorial 3: Image Segmentation*. 2010. <https://ai.stanford.edu/~syyeung/cvweb/tutorial3.html>. Citado na página 21.

ZHANG, A.; LIPTON, Z. C.; LI, M.; SMOLA, A. J. *Dive into Deep Learning*. 2023. Citado na página 27.

ZHAO, W.; CHEN, H.; LU, Y. W-net: A network structure for automatic segmentation of organs at risk in thorax computed tomography. In: *Proceedings of the 2020 2nd International Conference on Intelligent Medicine and Image Processing*. New York, NY, USA: Association for Computing Machinery, 2020. (IMIP 2020), p. 66–69. ISBN 9781450377799. Disponível em: <<https://doi.org/10.1145/3399637.3399642>>. Citado 4 vezes nas páginas 14, 36, 37 e 49.



UNIVERSIDADE FEDERAL DO MARANHÃO

Programa de Pós-Graduação em Ciência da Computação

Arthur Guilherme Santos Fernandes

***Segmentação de Traqueia em Tomografia
Computadorizada Utilizando Efficient-Deeplab***

**São Luís
2024**

Arthur Guilherme Santos Fernandes

Segmentação de Traqueia em Tomografia Computadorizada Utilizando Efficient-Deeplab

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

Programa de Pós-Graduação em Ciência da Computação

Universidade Federal do Maranhão

Orientador: Prof. Dr. Geraldo Braz Junior

Coorientador: Prof. Dr. João Otávio Bandeira Diniz

São Luís - MA

2024

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

Santos Fernandes, Arthur Guilherme.

Segmentação de Traqueia Em Tomografia Computadorizada
Utilizando Efficient-deeplab / Arthur Guilherme Santos
Fernandes. - 2024.

58 f.

Coorientador(a) 1: João Otávio Bandeira Diniz.

Orientador(a): Geraldo Braz Junior.

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

1. Aprendizado Profundo. 2. Segmentação de Imagens.
3. Redes Neurais. 4. Tomografia Computadorizada. 5.
Radioterapia. I. Bandeira Diniz, João Otávio. II. Braz
Junior, Geraldo. III. Título.

*Dedico este trabalho à minha mãe, Antônia de Fátima, que dedicou a sua vida inteira para
que eu pudesse chegar a este momento.*

Agradecimentos

Agradeço ao meu orientador Prof. Dr. Geraldo Braz Junior pela dedicação, paciência e por ser uma inspiração para mim. Agradeço também aos companheiros de laboratório que compartilharam seus conhecimentos e ouviram minhas ideias durante inúmeros almoços no R.U e operações de aquisição de café no NCA.

Our smartest machines are still blind.

(Fei-Fei Li)

Resumo

A segmentação de Órgãos em Risco (OAR) é crucial no planejamento da radioterapia. Seu objetivo é demarcar os tecidos saudáveis para que a radiação ionizante seja direcionada apenas às células cancerosas. Para isso, médicos realizam a demarcação dos órgãos manualmente, o que torna o processo demorado e propenso a erros. Portanto, metodologias de segmentação automática que utilizam aprendizado profundo podem acelerar a delimitação dos órgãos durante o planejamento da radioterapia. As metodologias existentes possuem muitos parâmetros, o que deixa o modelo pesado e caro de ser disponibilizado como um serviço. Este trabalho propõe o uso de uma arquitetura de rede neural convolucional denominada *EfficientDeeplab*, treinada em exames de tomografia computadorizada para realizar a segmentação de traqueia. O modelo diferencia-se de outras arquiteturas por possuir menor quantidade de parâmetros, o que a torna ideal para aplicações em serviços de saúde em larga escala. Para obter baixa quantidade de parâmetros e uma segmentação de qualidade foram aplicadas as convoluções atrous e a arquitetura EfficientNet. Foram realizados testes no conjunto de dados SEGTHOR, que obtiveram um *dice score* de 82,21%.

Palavras-chave: Aprendizado Profundo, Segmentação de Imagens, Redes Neurais, Tomografia Computadorizada, Radioterapia.

Abstract

Organs at Risk (OAR) segmentation is crucial in radiotherapy planning. Its objective is to demarcate healthy tissues so that ionizing radiation is directed only to cancer cells. To do this, doctors demarcate the organs manually, which makes the process very time-consuming and prone to errors. Therefore, automatic segmentation methodologies using deep learning can accelerate organ delineation during radiotherapy planning. Existing methodologies have many parameters, which makes the model cumbersome and expensive to make available as a service. This work proposes the use of a convolutional neural network architecture called *EfficientDeeplab*, trained on computed tomography scans to perform trachea segmentation. The model differs from other architectures by having fewer parameters, which makes it ideal for applications in large-scale healthcare services. To obtain a low number of parameters and quality segmentation, atrous convolutions and the EfficientNet architecture were applied. Tests were carried out on the SEGTHOR dataset, which obtained a *dice score* of 82.21%.

Keywords: Deep Learning, Image Segmentation, Neural Networks, Computerized Tomography, Radiotherapy.

Lista de ilustrações

Figura 1 – Cortes em TC a partir de diferentes ângulos: a) Axial; b) Coronal; c) Sagital.	13
Figura 2 – Planejamento Radioterápico com Auxílio de TC.	18
Figura 3 – Aplicação de Janelamento.	20
Figura 4 – Exemplo de segmentação de imagens.	21
Figura 5 – Segmentação de órgãos em TC.	21
Figura 6 – Neurônio Artificial.	24
Figura 7 – Rede Neural MLP.	25
Figura 8 – Operação de Convolução.	27
Figura 9 – Arquitetura Lenet.	27
Figura 10 – Representação Genérica das Redes EfficientNet.	28
Figura 11 – Diagrama do Bloco MBConv.	29
Figura 12 – Diagrama da Rede DeepLabV3+.	30
Figura 13 – Convoluções Atrous	30
Figura 14 – Módulo ASPP.	31
Figura 15 – Arquitetura das Redes FCN.	32
Figura 16 – Arquitetura da Rede U-net.	33
Figura 17 – Filtro de Convolução Atrous.	34
Figura 18 – Gargalos Residuais.	34
Figura 19 – Diagrama do Método Proposto.	38
Figura 20 – Demarcação de Órgãos Feita por Profissional.	39
Figura 21 – Aplicação do Janelamento de Intensidade.	40
Figura 22 – Diagrama da Rede DeepLabV3+.	41
Figura 23 – Visão Geral da Arquitetura do Modelo Proposto.	42
Figura 24 – Comparativo Após o Pós-Processamento.	42
Figura 25 – Falsos Positivos na Segmentação.	47
Figura 26 – Falso Positivo em Fatia.	47
Figura 27 – Porção Central da Traqueia.	48
Figura 28 – Divisão da Traqueia.	48

Lista de tabelas

Tabela 1 – Escala Hounsfield.	19
Tabela 2 – Janelas Radiológicas.	20
Tabela 3 – Resumo dos Trabalhos Relacionados.	37
Tabela 4 – Experimentos com Diferentes Backbones.	45
Tabela 5 – Comparação de Desempenho com Modelos U-net.	45
Tabela 6 – Comparação de Janelamento de Intensidade com EfficientDeeplab - B2.	46
Tabela 7 – Comparação de Resultados com Volumes Inteiros.	47
Tabela 8 – Comparação com a Literatura.	49

Lista de abreviaturas e siglas

ASPP	<i>Atrous Spatial Pyramid Pooling</i> (Pirâmide de Pooling Espacial Atrous).
CNN	<i>Convolutional Neural Networks</i> (Redes Neurais Convolucionais).
CRF	<i>Conditional Random Fields</i> (Campos Aleatórios Condicionais).
DICE	<i>Dice Score</i> .
DNA	<i>Deoxyribonucleic Acid</i> (Ácido Desoxirribonucleico).
FCN	<i>Fully Convolutional Networks</i> (Redes Completamente Convolucionais).
GAN	<i>Generative Adversarial Networks</i> (Redes Adversárias Generativas).
GPU	<i>Graphics Processing Unit</i> (Unidade de Processamento Gráfico).
HU	<i>Hounsfield Units</i> (Unidades de Hounsfield).
IOU	<i>Intersection Over Union</i> (Intersecção Sobre União).
MLP	<i>Multilayered Perceptron</i> (Perceptron de Múltiplas Camadas).
OR	Órgãos em Risco.
SGD	<i>Stochastic Gradient Descent</i> (Gradiente Descendente Estocástico).
TC	Tomografia Computadorizada.

Sumário

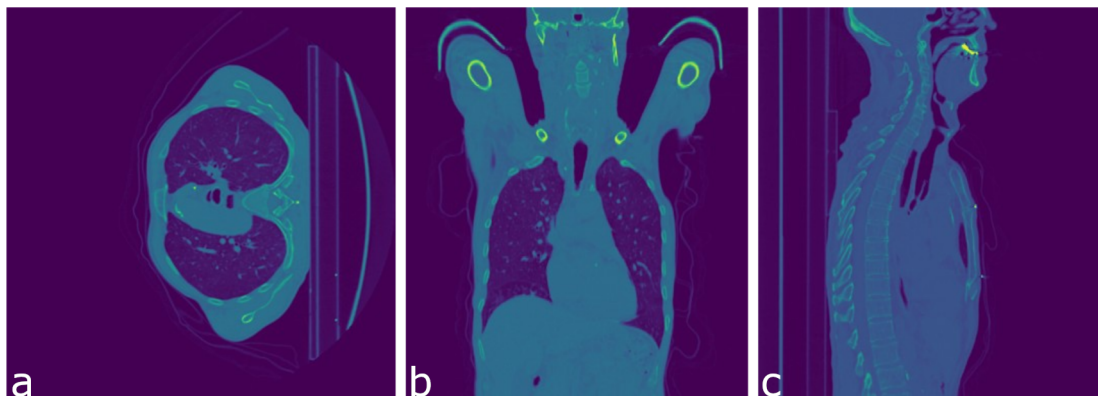
1	INTRODUÇÃO	13
1.1	Hipóteses de Pesquisa	14
1.2	Objetivo	15
1.2.1	Objetivos Específicos	15
1.3	Contribuições	15
1.3.1	Organização do Trabalho	16
2	FUNDAMENTAÇÃO TEÓRICA	17
2.1	Visão Computacional em Imagens Médicas	18
2.1.1	Tomografia computadorizada	18
2.1.2	Janelamento de Intensidade	19
2.1.3	Segmentação de Imagens	20
2.2	Aprendizado de Máquina	22
2.2.1	Redes Neurais	23
2.2.2	Redes Neurais Convolucionais	26
2.2.3	EfficientNet	28
2.2.4	DeepLab	29
2.3	Considerações Finais	30
3	TRABALHOS RELACIONADOS	32
3.1	Segmentação Semântica	32
3.2	Segmentação Semântica de Órgãos em Risco	35
4	MATERIAIS E MÉTODO	38
4.1	Aquisição da Imagens	38
4.2	Pré-Processamento	38
4.2.1	Janelamento de Intensidade	39
4.2.2	Redimensionamento e Normalização	40
4.3	Construção do Modelo	40
4.4	Pós Processamento	42
4.5	Considerações Finais	43
5	RESULTADOS	44
5.1	Ambiente de Execução	44
5.2	Treinamento do Modelo com Convoluções Atrous e DeepLab	44
5.3	Utilização de Janelamento de Intensidade	46

5.4	Estudo de Casos com Volumes Inteiros	46
5.5	Considerações	48
6	CONCLUSÃO	50
	REFERÊNCIAS	51

1 Introdução

A radioterapia é uma forma de tratamento do câncer onde doses substanciais de radiação ionizante são direcionadas às células malignas, destruindo-as ou reduzindo o tamanho do tumor. A fase inicial do planejamento da radioterapia gira em torno da identificação precisa da localização do tumor e dos órgãos adjacentes. Essa etapa é fundamental para impedir que radiação seja direcionada aos tecidos saudáveis próximos ao tumor, mitigando os seus efeitos adversos sobre as células não cancerígenas. Os órgãos que circundam o tumor são comumente denominados **Órgãos em Risco (OR)** (NOEL et al., 2016; DINIZ et al., 2021). Normalmente, durante o planejamento radioterápico, os ORs são delineados manualmente a partir de imagens adquiridas por **Tomografia Computadorizada (TC)**: uma técnica de aquisição de imagens de tecidos moles e duros no corpo humano que ajuda os profissionais a avaliarem a condição de órgãos e suas estruturas (WEBER et al., 2014). O resultado do exame é um volume 3D analisado pelos médicos por meio de cortes 2D (também chamados de *slices*). A Figura 1 apresenta exemplos de cortes a serem analisados por médicos. Da esquerda para a direita os ângulos de cortes são denominados: Axial, Coronal, Corte Sagital.

Figura 1 – Cortes em TC a partir de diferentes ângulos: **a)** Axial; **b)** Coronal; **c)** Sagital.



Fonte: Lambert et al. (2019).

Demarcar os ORs de maneira precisa é de grande importância para o planejamento da radioterapia, pois mapear os tecidos saudáveis ao redor do tumor permite que os especialistas direcionem o feixe de radiação exclusivamente ao tecido canceroso, minimizando a exposição desnecessária aos raios ionizantes e reduzindo a probabilidade de efeitos adversos da radioterapia, que podem incluir: inflamação, fibrose, úlcera e, em casos graves, falência de órgãos (TEKATLI et al., 2018). No entanto, a qualidade da demarcação dos órgãos depende da experiência clínica do médico, de maneira que um médico experiente terá maior precisão, resultando em pacientes com menos efeitos colaterais e mais qualidade

de vida (LAMBERT et al., 2019). Esta é uma tarefa demorada, que está suscetível aos erros humanos. Órgãos menores, como a traqueia, apresentam desafios únicos para a segmentação devido ao seu tamanho e similaridades texturais com tecidos circundantes. Tendo em vista estas dificuldades da segmentação manual de ORs, pesquisas recentes buscaram desenvolver métodos automatizados que acelerem o processo e eliminem o fator de erro humano (DINIZ et al., 2021), (FU et al., 2021), (FENG et al., 2020), (ZHAO; CHEN; LU, 2020).

Ao analisar os modelos de Redes Neurais utilizados nas metodologias mais bem sucedidas da literatura (ZHAO; CHEN; LU, 2020), (YAN et al., 2022), observa-se que possuem uma grande quantidade de parâmetros. Isso significa um maior uso de memória do computador. Pensando na utilização em larga escala de tais modelos, mais parâmetros pode significar mais gastos operacionais (plataforma de nuvem, máquinas, GPU, etc), o que por sua vez, ocasiona na baixa aplicabilidade deste tipo de metodologia.

Neste contexto, esta dissertação de mestrado visa propor uma arquitetura de Rede Neural Convolutacional para segmentação de traqueia que possua menos parâmetros e capacidade similar aos melhores modelos da literatura, facilitando a sua implementação em sistemas médicos de larga escala.

1.1 Hipóteses de Pesquisa

Para tanto, foram estabelecidas as seguintes hipóteses de pesquisa que norteiam este trabalho:

1. As convoluções *atrous*, quando aplicadas ao problema, contribuem para a construção de um modelo compacto e, ao mesmo tempo, capaz de generalizar;
2. A segmentação semântica da traqueia pode ser obtida com precisão via uma arquitetura de rede neural convolutacional que utiliza como backbone uma EfficientNet e como decoder uma DeepLabV3;
3. A segmentação da área da traqueia pode ser mais precisa caso seja determinada uma janela radiográfica como pré-processamento, específica para o problema;
4. Aplicar o algoritmo de componentes conectadas como pós-processamento consegue separar a região de traqueia dos falsos positivos de segmentação produzidos pelo modelo.

1.2 Objetivo

Esta dissertação de mestrado tem por objetivo avaliar metodologias de segmentação baseadas em Redes Neurais Convolucionais aplicadas à segmentação de traqueia para desenvolver um modelo que apresente um bom equilíbrio entre a quantidade de parâmetros treináveis e qualidade de segmentação e seja adequado para implementação em dispositivos de memória reduzida.

1.2.1 Objetivos Específicos

Especificamente para a segmentação de traqueia, este trabalho possui os seguintes objetivos:

- Desenvolver um modelo de CNN que combina as redes *EfficientNet* (TAN; LE, 2019) e *DeepLab* (CHEN et al., 2018b) para realizar a segmentação de traqueia;
- Avaliar a utilização de convoluções dilatáveis (também chamadas convoluções *atrous*) (HOLSCHNEIDER et al., 1990) para aumentar o campo de visão dos filtros de convolução sem adicionar parâmetros ao modelo;
- Analisar o impacto do janelamento de intensidade de *pixel* como pré-processamento no resultado da segmentação;
- Avaliar a acurácia de segmentação utilizando *Dice Score* e a quantidade de parâmetros treináveis do modelo final, comparando com as melhores arquiteturas disponíveis na literatura.

1.3 Contribuições

Este trabalho agrega valor na pesquisa de segmentação de órgãos em risco a partir das seguintes contribuições:

- Avaliação da rede *EfficientNet* (TAN; LE, 2019) como *backbone* extrator de características, em conjunto com a estrutura da *DeepLab* (CHEN et al., 2018b) para a segmentação de traqueia;
- Desenvolvimento de um modelo de segmentação automática de traqueia baseado em Redes Neurais Convolucionais que consome consideravelmente menos memória e possui menos parâmetros que o estado da arte, o que pode ser especialmente útil em ambientes de computação de memória limitada;
- Estudo de intervalos adequados para o janelamento de intensidade de *pixel* como pré-processamento no resultado da segmentação de traqueia.

1.3.1 Organização do Trabalho

Os capítulos subsequentes estão estruturados da seguinte maneira:

- O Capítulo 2 apresenta a fundamentação teórica em relação aos conceitos de Órgãos em Risco, planejamento radioterápico e também sobre as técnicas de Aprendizado Profundo e Redes Neurais Convolucionais utilizadas.
- O Capítulo 3 descreve os trabalhos relacionados à pesquisa atual de metodologias automáticas de segmentação de traqueia que utilizam Redes Neurais Convolucionais.
- O Capítulo 4 descreve todas as etapas que compõem a metodologia proposta para segmentação automática de traqueia.
- O Capítulo 5 apresenta os resultados obtidos a partir de experimentos, bem como análises de casos em que o modelo obteve boa e má precisão, buscando entender os motivos e fornecer soluções para o comportamento observado.
- O Capítulo 6 apresenta as considerações finais sobre os resultados, ideias para trabalhos futuros e os artigos científicos desenvolvidos durante a pesquisa.

2 Fundamentação Teórica

Neste capítulo são apresentados os conceitos fundamentais utilizados para o desenvolvimento desta pesquisa. Estes conceitos abrangem desde conhecimentos sobre câncer e radioterapia, até o processo de análise de imagens médicas e aprendizado profundo com redes neurais convolucionais.

A radioterapia é um dos principais métodos de tratamento do câncer e pode ser aplicada em diversos casos, embora sua eficácia dependa de uma análise individualizada do paciente. Este método pode ser eficaz tanto em estágios iniciais, quanto avançados da doença. No primeiro caso, este pode ser o tratamento definitivo, evitando uma cirurgia e preservando o órgão alvo. Já no caso de estágios avançados de câncer, a radioterapia pode ser um grande aliado da cirurgia e dos cuidados paliativos (VAIDYA, 2021).

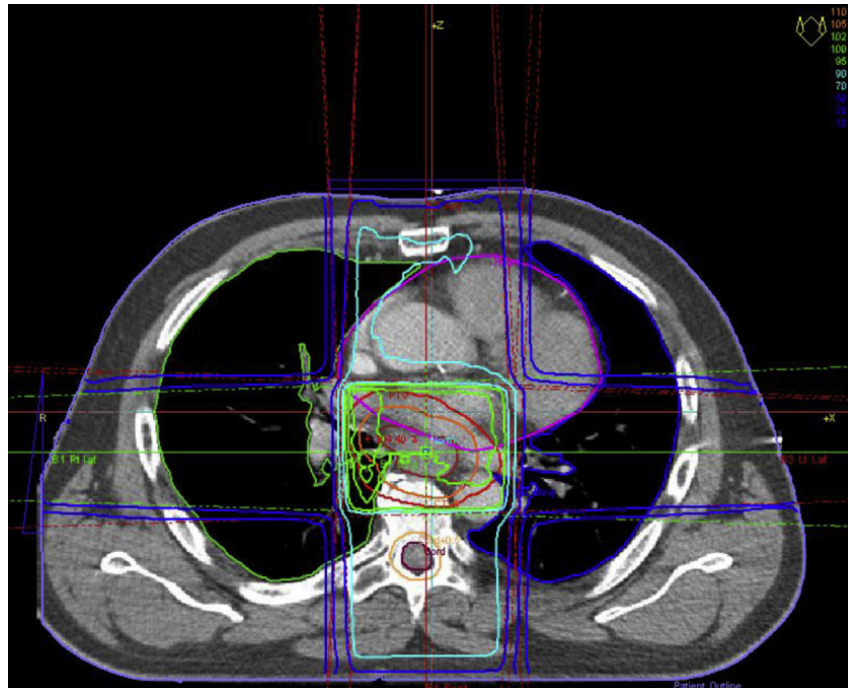
Como sugere o nome, este tratamento consiste em bombardear tecidos cancerígenos com raios ionizantes, sendo a Radiação Gama (γ) e os Raios-x os dois tipos de radiação mais utilizados na medicina (MEHTA et al., 2010). Ao entrarem em contato com tecido biológico, a radiação ionizante consegue remover elétrons das órbitas dos átomos, o que a nível molecular significa a quebra de porções de DNA das células (VAIDYA, 2021). O corpo possui mecanismos de reparo automático de DNA, porém o processo de regeneração é consideravelmente mais lento nas células cancerígenas, daí a eficácia do tratamento.

Os efeitos colaterais da radiação ionizante em tecido saudável podem ser divididos em dois tipos: imediatos e tardios (BARAZZUOL; COPPES; LUIJK, 2020). O primeiro tipo pode ser sentido dias ou semanas após o tratamento e pode se manifestar na forma de irritação no tecido, pequenos sangramentos e dores localizadas, sendo que estes sintomas são revertidos em pouco tempo. Já os efeitos tardios são crônicos e muitas vezes progressivos, levando a condições como: inflamação crônica, fibrose e úlcera (TEKATLI et al., 2018).

Para garantir que a radioterapia seja eficaz sem causar danos aos órgãos saudáveis próximos ao tumor, chamados de **Órgãos em Risco** (ORs), é realizado o planejamento radioterápico. Nessa etapa, os ORs são identificados e delimitados, pois, eles é que determinam a dose máxima de radiação que pode ser aplicada ao tumor alvo (EVANS; STAFFURTH, 2018). O objetivo é maximizar a dose no tumor e minimizar nos ORs. Após a aprovação do médico, os dados são transferidos para a máquina de tratamento. A Figura 2 ilustra o contorno dos órgãos e o direcionamento dos feixes de radiação. Nela é possível verificar quatro feixes de radiação: pela frente, por trás, pela esquerda e pela direita. O traçado vermelho ao centro da imagem é a região com câncer.

A delimitação de órgãos se torna então uma tarefa fundamental para o tratamento por radioterapia. O restante deste capítulo trata do referencial teórico utilizado para esta

Figura 2 – Planejamento Radioterápico com Auxílio de TC.



Fonte: Evans e Staffurth (2018).

finalidade.

2.1 Visão Computacional em Imagens Médicas

Por visão computacional, entende-se a área de estudo que busca extrair informações úteis de imagens (majoritariamente digitais) (PRINCE, 2012). Para alcançar este fim, são utilizados diversos algoritmos e técnicas de processamento de imagens, seja para restauração, melhoramento, segmentação, agrupamento de regiões, etc.

As aplicações de visão computacional estendem-se às muitas áreas do conhecimento. Alguns exemplos são: processamento de imagens via satélite, geração de imagens médicas por Raios-x e, até mesmo imagens geradas por fontes não convencionais, como ultrassom (GONZALEZ; WOODS, 2008). Dentro da medicina, a visão computacional tem ganhado muito destaque ao ajudar médicos em diagnósticos por imagem ou acelerar processos demorados como a demarcação de Órgãos em Risco. Esta última tarefa utiliza imagens geradas pelo exame de tomografia computadorizada, detalhado na Seção 2.1.1

2.1.1 Tomografia computadorizada

A tomografia é uma técnica de imageamento de um objeto por meio de dados adquiridos ao irradiar o objeto a partir de diferentes direções. Seu impacto na medicina

foi revolucionário, pois permitiu aos médicos visualizarem órgãos internos com precisão e segurança sem precedentes (KAK; SLANEY, 2001).

Segundo Dance et al. (2014), o processo de aquisição de imagens de tomografia computadorizada envolve a irradiação de partículas através do corpo de um paciente e a subsequente medição da transmissão de radiação, feita por um arco detector posicionado atrás do paciente em direção oposta ao dispositivo emissor de radiação. Ao rotacionar o arco detector e o dispositivo emissor em torno do paciente, é possível adquirir um volume de amostras de dados, utilizadas posteriormente para reconstruir uma imagem digital.

A imagem reconstruída possui os valores de intensidade de píxel dentro da Escala Hounsfield (SEERAM, 2016), variando conforme a densidade do material representado. O nome da unidade de medida para estes valores é Unidade de Hounsfield (HU). A Tabela 1 detalha os valores apresentados para alguns tipos de tecido do corpo humano.

Tabela 1 – Escala Hounsfield.

Substancia	Unidades de Hounsfield (HU)
Ar	-1000
Gordura	-100
Água	0
Fluido Cefaloraquídeo	15
Matéria Branca	20 - 30
Massa Cinzenta	36 - 46
Músculo	50
Sangue	56 - 76
Osso Compacto	1000

Fonte: Seeram (2016).

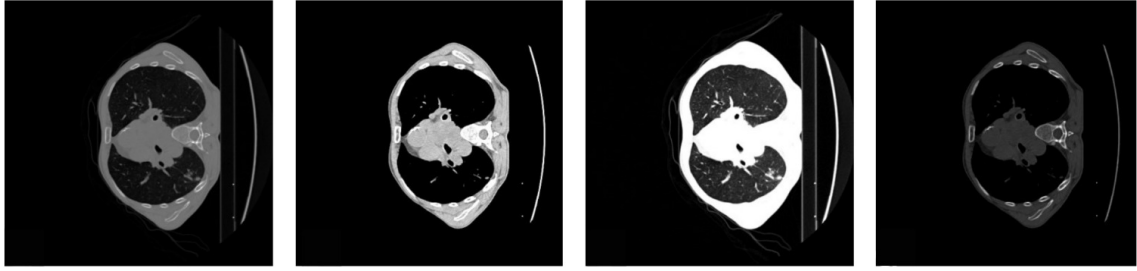
2.1.2 Janelamento de Intensidade

Em processamento de imagens, o pré-processamento é uma das etapas fundamentais de uma metodologia, pois é nela que ocorrem melhoramentos, realces, correção de artefatos, redução de ruído, alinhamento, etc. Esses processos simplificam a análise da imagem, seja por outro algoritmo ou por humanos (GONZALEZ; WOODS, 2008).

O Janelamento de Intensidade é um dos pré-processamentos mais comuns aplicados na medicina diagnóstica por tomografia computadorizada. Consiste em selecionar um determinado intervalo do espectro de intensidades provenientes da imagem tomográfica e limitar os valores de intensidade de píxel a esse intervalo. Estes intervalos são conhecimentos como “janelas” (WEBER et al., 2014) e cada janela serve para melhorar a visualização de determinados tipos de tecido.

A Figura 3 ilustra algumas das janelas mais comuns da radiologia aplicadas à mesma imagem de origem, da esquerda para a direita: Imagem original, Janela de Mediastino,

Figura 3 – Aplicação de Janelamento.



Fonte: Elaborado pelo autor.

Janela Pulmonar (parênquima) e Janela Óssea (CONSIDERA et al., 2006). Já a Tabela 2, sumariza os intervalos em Unidades de Hounsfield (HU) para cada uma destas janelas. Os valores referência para cada intervalo podem ser encontrados em: VICENTE et al. (2004), CONSIDERA et al. (2006) e MARCHIORI et al. (2007).

Tabela 2 – Janelas Radiológicas.

Nome	Aplicação	Intervalo (HU)
Janela de Parênquima	Vis. de pulmões e vias aéreas (traquéia, nasal, etc.)	[-1000, 150]
Janela de Mediastino	Vis. de partes moles (músculos, órgãos abdominais, etc.)	[-160, 240]
Janela Óssea	Vis. de ossos compactos (crânio, fêmur, etc.)	[-1.650, 2350]

Fonte: elaborado pelo autor.

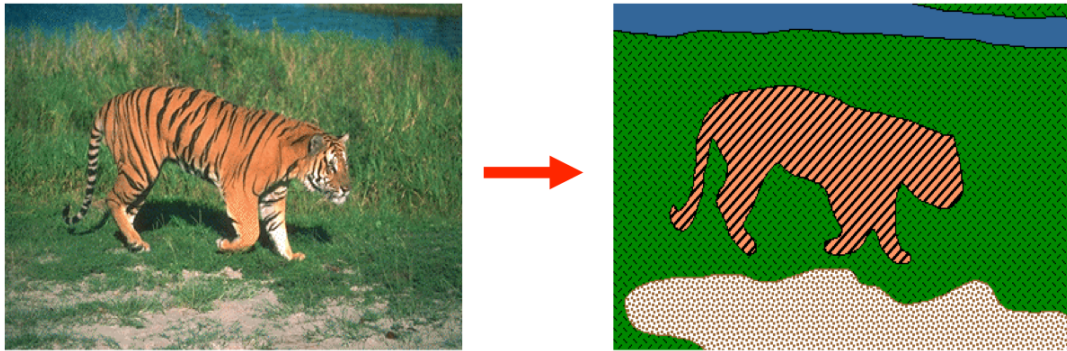
Sejam α e β , respectivamente, o limite inferior e superior do intervalo $[\alpha, \beta]$ escolhido, a operação de janelamento de intensidade pode ser definida como na Equação 2.1:

$$f(x) = \begin{cases} \alpha & \text{se } x \leq \alpha \\ \beta & \text{se } x \geq \beta \\ x & \text{se } x \in [\alpha, \beta] \end{cases} \quad (2.1)$$

2.1.3 Segmentação de Imagens

Uma das tarefas elementares dentro da visão computacional é a segmentação de imagens. Consiste em subdividir uma imagem em suas regiões constituintes (GONZALEZ; WOODS, 2008) para analisar suas características de maneira isolada. De maneira geral, a saída dos algoritmos de segmentação é chamada de máscara de segmentação. Esta contém *labels* que indicam a qual classe pertence cada *pixel*. A Figura 4 ilustra o resultado esperado de uma segmentação de imagens, onde o algoritmo (representado pela seta vermelha) foi capaz de agrupar os *pixels* em quatro classes: animal, terra, água e vegetação

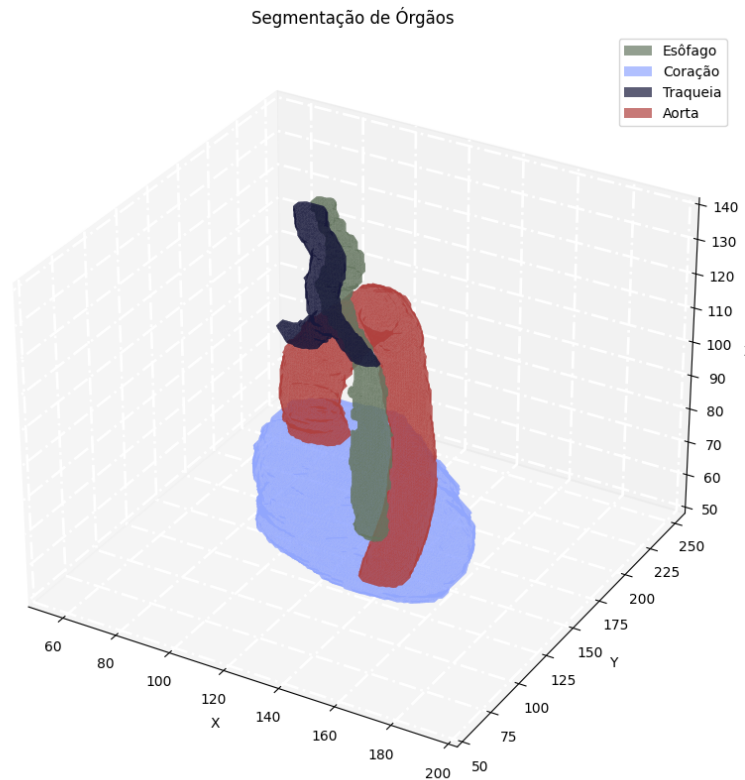
Figura 4 – Exemplo de segmentação de imagens.



Fonte: (YEUNG, 2010).

Na medicina, alguns exemplos de aplicação da segmentação são: a medida de ossos e tecidos (LEE et al., 2022), detecção de estruturas anômalas (THANOON et al., 2023), identificação de músculos lesionados (ACKERMANS et al., 2021) e segmentação de órgãos (LAMBERT et al., 2019). A Figura 5 apresenta a saída esperada de um algoritmo de segmentação de órgãos contendo: demarcação do coração, traqueia, esôfago e artéria aorta.

Figura 5 – Segmentação de órgãos em TC.



Fonte: Elaborado pelo autor.

Para determinar a qualidade de uma segmentação de imagem deve-se comparar

(visual ou numericamente) as máscaras geradas pelo algoritmo, com a segmentação considerada a mais próxima do ideal (também chamada de valor-verdade). O padrão na literatura de segmentação de imagens médicas é a utilização de funções matemáticas de pontuação (métricas) como o *Dice Score* (DICE, 1945) e o *Jaccard Score* (JACCARD, 1912), como evidenciado em Eelbode et al. (2020) e Isensee et al. (2021) por exemplo.

O *Dice-Score* é dado pela Equação 2.2, onde Y_{true} e Y_{out} representam, respectivamente, os conjuntos de *pixels* pertencentes à saída esperada e à saída obtida. A mesma nomenclatura vale para o *Jaccard Score* (também referido por IOU), dado pela Equação 2.3. Ambas as métricas de segmentação serão utilizadas para avaliar o modelo proposto neste trabalho. Os detalhes estão situados na Seção 5.5.

$$D(Y_{true}, Y_{out}) = 2 \cdot \frac{|Y_{true} \cap Y_{out}|}{|Y_{true}| + |Y_{out}|} \quad (2.2)$$

$$J(Y_{true}, Y_{out}) = \frac{|Y_{true} \cap Y_{out}|}{|Y_{true} \cup Y_{out}|} \quad (2.3)$$

O pós-processamento da área segmentada é uma etapa essencial em muitas tarefas de segmentação de imagem (SHIVHARE; GUPTA, 2015; DHAWAN; BODANI; GARG, 2019) incluindo segmentação de órgãos. Após a segmentação inicial, técnicas de pós-processamento são aplicadas para refinar e melhorar os resultados. Isso pode envolver a remoção de pequenos artefatos ou ruídos, a correção de erros de segmentação, a suavização das bordas ou a conexão de regiões segmentadas que fazem parte do mesmo objeto. O pós-processamento é fundamental para garantir que os resultados da segmentação sejam precisos e clinicamente úteis, especialmente em aplicações médicas. Este trabalho utilizou o pós-processamento de regiões conectadas (SAMET; TAMMINEN, 1988) para extrair o maior objeto tridimensional obtido pela segmentação. O algoritmo segue os seguintes passos:

1. Detecção das ilhas de *voxels* da imagem (*voxels* de valor igual que são vizinhos).
2. Contagem da quantidade de *voxels* presente em cada ilha.
3. Seleção da ilha que contém maior quantidade de *voxels*.

2.2 Aprendizado de Máquina

O Aprendizado de Máquina é uma sub-área da Inteligência Artificial que busca desenvolver modelos de processamento de informação que aprendem a partir de observações (dados), melhorando seu desempenho em tarefas futuras. Russell e Norvig (2009) resume

a operação deste tipo de sistema: a partir de uma coleção de pares de entrada e saída, aprende uma função que prevê a saída para novas entradas.

Segundo [Shalev-Shwartz e Ben-David \(2014\)](#), estes modelos são particularmente úteis ao desempenhar tarefas que são muito complexas de serem programadas explicitamente ou que necessitem de maior flexibilidade, adaptando-se às entradas. Nestes casos, é mais viável projetar um modelo de Aprendizado de Máquina do que desenvolver um software que aplique um conjunto de regras de processamento.

Existem diferentes tipos de aprendizado: supervisionado, semi-supervisionado, não supervisionado e aprendizado por reforço. A abordagem de treinamento utilizada neste trabalho é o aprendizado supervisionado. Para [Burkov \(2020\)](#), este tipo de aprendizagem ocorre quando há um modelo preditivo e uma coleção de exemplos rotulados $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, onde x_i é chamado de vetor de características e y_i é o chamado rótulo. O objetivo do aprendizado supervisionado é criar um modelo que, dado um vetor de características x como entrada, seja capaz de deduzir um rótulo y para esse vetor.

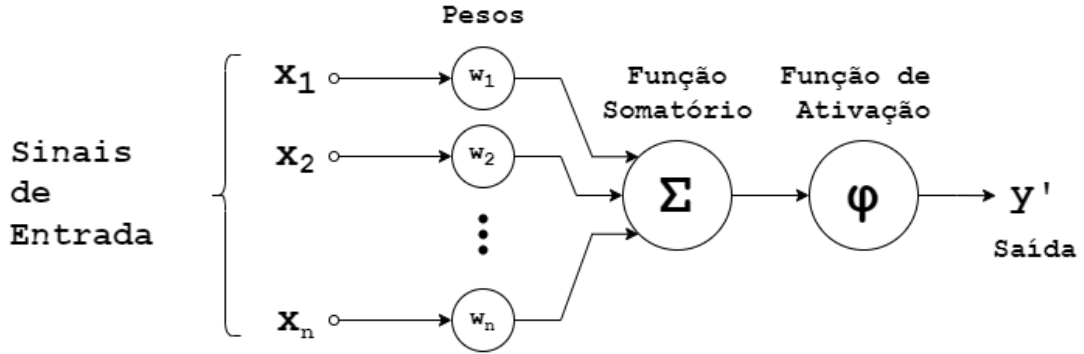
2.2.1 Redes Neurais

As redes neurais são um tipo específico de modelo de aprendizado de máquina. Segundo [Haykin \(2009\)](#), foram desenvolvidas a partir da constatação de que o cérebro humano computa de forma completamente diferente do computador digital e possui a capacidade de construir suas próprias regras de comportamento através do que chamamos de *experiência*. Para [Hecht-Nielsen \(1989\)](#), as redes neurais artificiais são uma estrutura de processamento de informações que funciona de maneira paralela e distribuída, que consiste em unidades básicas de processamento interconectadas por canais de comunicação unidirecional.

As referidas *unidades básicas de processamento* são os neurônios artificiais. Assim como a sua contraparte biológica, eles possuem uma propriedade conhecida como plasticidade ([HAYKIN, 2009](#)), isto é: têm a capacidade de ajustar o seu funcionamento interno conforme as entradas recebidas. A Figura 6 ilustra um neurônio artificial e suas partes constituintes.

O modelo de neurônio artificial apresentado acima foi primeiramente proposto por [Rosenblatt \(1958\)](#), e é chamado de *perceptron*. Suas partes constituintes são: o conjunto de sinais de entrada $X = \{x_1, x_2, \dots, x_n\}$, o conjunto de pesos $W = \{w_1, w_2, \dots, w_n\}$, o *bias* que age como uma entrada de valor fixo 1, a função somatório Σ , função de ativação φ e a saída y_k . Para que o *perceptron* consiga aprender, é executada uma etapa de treinamento, onde os sinais de entrada X são apresentados ao modelo, obtendo o sinal de saída y' . Em seguida, os pesos são atualizados seguindo a regra de aprendizado escolhida ([HECHT-NIELSEN, 1989](#)). A Equação 2.4 detalha a obtenção da saída do *perceptron*, onde y' é

Figura 6 – Neurônio Artificial.



Fonte: Adaptado de (HAYKIN, 2009).

a saída, φ representa uma função de ativação, e, x_i e w_i são o valor de entrada e o peso da conexão, respectivamente. Já as Equações 2.5 e 2.6 detalham a função de ativação e a regra de aprendizado utilizadas no trabalho de Rosenblatt (1958).

$$y' = \varphi\left(\sum_{i=1}^n x_i w_i\right) \quad (2.4)$$

$$\varphi(z) = \begin{cases} 1 & \text{se } z \geq 0 \\ 0 & \text{se } z < 0 \end{cases} \quad (2.5)$$

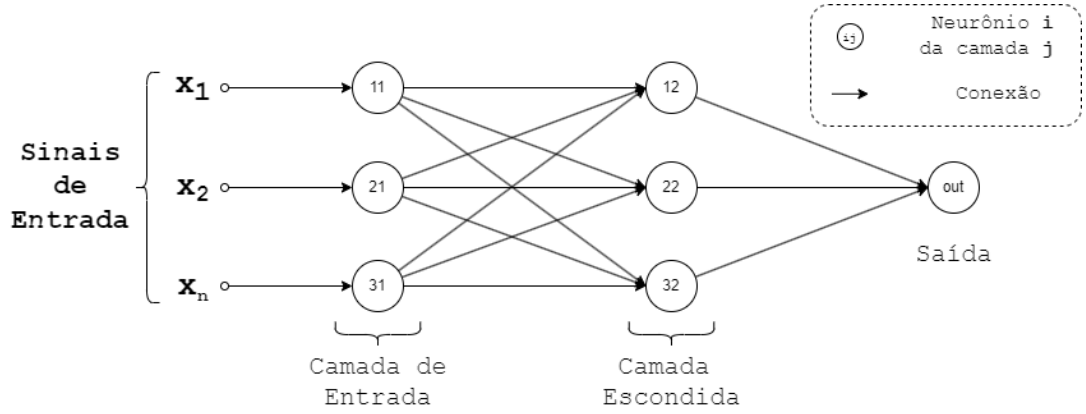
$$w_i = w_i + (y - y')x_i \quad (2.6)$$

Por si só, o *perceptron* já consegue aproximar funções, entretanto, é possível organizar os neurônios em camadas interconectadas, formando as primeiras redes neurais desenvolvidas, conhecidas como *Multi Layered Perceptrons* (MLP). A Figura 2.2.1 detalha a composição de uma rede neural MLP clássica contendo: camada de entrada, conexões sinápticas representadas por setas, os neurônios representados por círculos, a camada intermediária, também chamada de camada escondida e o neurônio de saída.

A partir deste ponto, a teoria de redes neurais recebeu cada vez mais incrementos, sendo um deles o uso do algoritmo de Gradiente Descendente Estocástico (SGD) (ROBBINS; MONRO, 1951). Ele funciona atualizando iterativamente os pesos da rede com base no gradiente da função de erro calculado em um subconjunto aleatório dos dados de treinamento (AMARI, 1967).

Já o algoritmo de retropropagação, também referido por *backpropagation*, foi introduzido por Rumelhart, Hinton e Williams (1986), e tornou-se um marco importante no desenvolvimento das redes neurais. Ele consiste em calcular o gradiente da função de

Figura 7 – Rede Neural MLP.



Fonte: Elaborado pelo autor.

erro em relação aos próprios pesos da rede, permitindo a adição de múltiplas camadas escondidas. Haykin (2009) define os cinco passos que constituem o *backpropagation*:

1. Calcular o erro de saída e_{out} , definido pela função de erro (*loss*) escolhida. Este trabalho utiliza a Entropia Cruzada Binária (GOODFELLOW; BENGIO; COURVILLE, 2016) como função de *loss* definida pela Equação 2.7.

$$e_{out} = - \sum_{i=0}^1 y_{true} \cdot \log(y_{out}) \quad (2.7)$$

2. Calcular a variação dos pesos entre a camada de saída e a camada escondida. Esta operação é dada pela Equação 2.8, onde η é uma constante pré-definida chamada taxa de aprendizado, $\varphi'(y_j)$ é a derivada da função de ativação φ e y_j é a saída da camada escondida.

$$\Delta W_{j,out} = \eta \cdot e_{out} \cdot \varphi'(y_j) \cdot y_j \quad (2.8)$$

3. Calcular erro retropropagado, dado pela Equação 2.9, onde $W_{j,out}$ representa os pesos sinápticos entre a camada escondida e a camada de saída.

$$e_j = - \sum_{i=1}^n e_{out} \cdot \varphi'(y_j) \cdot W_{j,out} \quad (2.9)$$

4. Calcular variação dos pesos entre as camadas escondida e de entrada, através da Equação 2.10. Nela, y_i representa a saída da camada de entrada.

$$\Delta W_{i,j} = -\eta \cdot e_j \cdot \varphi'(y_i) \cdot y_i \quad (2.10)$$

5. Somar os pesos antigos aos ΔW obtidos.

Desta maneira, pode-se ajustar os pesos de uma rede neural com múltiplas camadas, minimizando o erro e consequentemente, tornando-o melhor na tarefa para qual está sendo treinado. Este tipo de rede com várias camadas intermediárias é conhecido como rede neural profunda, e deu origem ao ramo do Aprendizado de Máquina conhecido como Aprendizado Profundo ou *Deep Learning*.

O uso combinado do SGD e da retropropagação permitiu o treinamento eficiente de redes neurais com muitas camadas ocultas, o que antes era considerado inviável devido aos desafios de treinamento. Isso abriu caminho para o surgimento do aprendizado profundo e possibilitou avanços significativos em áreas como reconhecimento de fala, visão computacional, processamento de linguagem natural, etc.

2.2.2 Redes Neurais Convolucionais

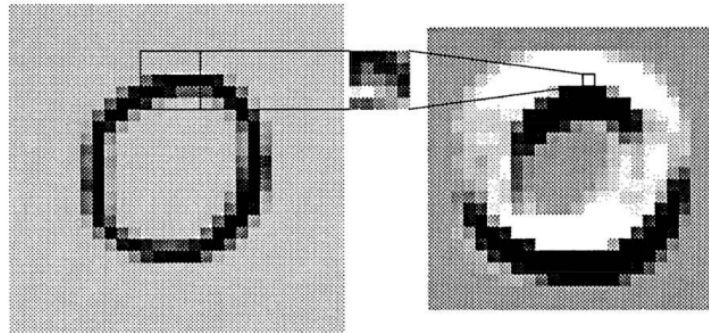
Em visão computacional, utilizar as redes completamente conectadas (MLP) para processar imagens é algo que exige que a rede possua muitos neurônios. Por exemplo, para uma imagem de dimensões 256×256 , têm-se ao total 65.536 neurônios apenas na camada de entrada. Este número aumenta exponencialmente em relação ao tamanho da imagem, tornando a utilização do *perceptron* proibitivamente custosa.

Para solucionar este problema, as Redes Neurais Convolucionais (CNNs) foram especialmente projetadas para processar imagens com eficiência. Ao introduzir esse modelo de rede, [LeCun et al. \(1989\)](#) inspirou-se na ideia do processamento clássico de imagens de detectar e combinar características locais. Portanto, restringiu os neurônios das camadas iniciais para que eles recebessem como entrada uma pequena porção da imagem (por exemplo, uma janela 3×3). Em seguida, desloca-se o neurônio até que ele tenha processado toda a imagem, armazenando o resultado em um espaço de memória que recebe o nome de mapa de características. Esta operação ficou conhecida como convolução e dá o nome a este tipo de rede neural. A Figura 8 ilustra a operação de convolução, contendo a imagem de entrada (à esquerda), os pesos do neurônio (ao meio) e o mapa de características (à direita).

A Figura 9 apresenta a arquitetura proposta por [LeCun et al. \(1989\)](#), com camadas de convolução que transformam a entrada em mapas de características e reduzem a dimensionalidade em direção à camada de saída totalmente conectada. As convoluções reduzem naturalmente o tamanho da entrada, enquanto as camadas de *pooling* também diminuem a dimensionalidade, geralmente usando médias ou máximos em regiões da imagem.

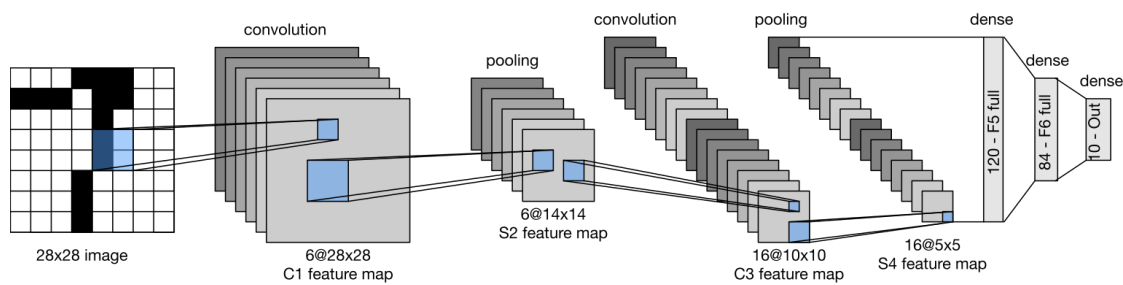
Esta arquitetura descarta a utilização dos neurônios completamente conectados nas camadas iniciais, sendo estes aplicados somente na camada final, que realiza a classificação. Isso diminuiu muito o custo computacional necessário para o processamento de imagens

Figura 8 – Operação de Convolução.



Fonte: Adaptado de [LeCun et al. \(1989\)](#).

Figura 9 – Arquitetura Lenet.



Fonte: Adaptado de [Zhang et al. \(2023\)](#).

com aprendizado profundo, tornando-o eficaz no reconhecimento de objetos, segmentação de imagens, etc.

A partir deste ponto, houve o surgimento de modelos mais profundos ([SIMONYAN; ZISSERMAN, 2015](#)), novos tipos de conexão ([HE et al., 2015](#)) e operações de convolução ([LONG; SHELHAMER; DARRELL, 2014](#)). Estas adições tornaram os modelos ainda mais eficientes e, ao mesmo tempo, mais elaborados em sua arquitetura, dando origem a diversos tipos de redes neurais convolucionais como, por exemplo: as Redes Generativas Adversárias ([GOODFELLOW et al., 2014](#)), Redes Completamente Convolucionais ([LONG; SHELHAMER; DARRELL, 2014](#)) e redes do tipo *Encoder-Decoder* como em [Ronneberger, Fischer e Brox \(2015\)](#).

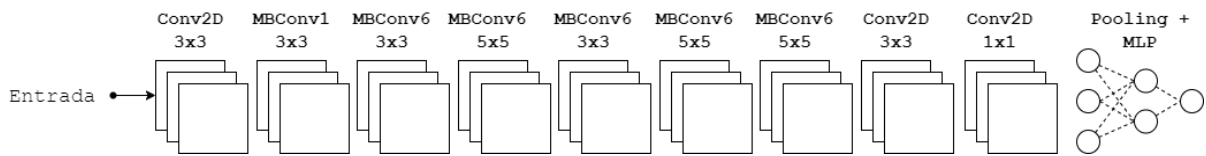
Arquiteturas encoder-decoder são amplamente utilizadas para a tarefa de segmentação semântica, onde um conjunto de características é extraída da imagem original e depois convertida numa máscara de segmentação. Este trabalho avalia a EfficientNet como mecanismo para auxiliar na extração de características juntamente com uma DeepLab

para o refino de características e decodificação na máscara de segmentação.

2.2.3 EfficientNet

A arquitetura *EfficientNet* trata-se de uma família de redes neurais pensadas com o uso eficiente de recursos em mente, escalando as suas camadas conforme a necessidade da aplicação (TAN; LE, 2019). São ao todo sete redes neurais convolucionais que se diferem apenas pela quantidade de parâmetros em cada camada. A Figura 10 ilustra a base de uma rede *EfficientNet*: são nove estágios, sendo três deles convoluções 2D convencionais, seis MBConv e 1 estágio de *pooling* e camadas completamente conectadas para realização de classificação.

Figura 10 – Representação Genérica das Redes EfficientNet.

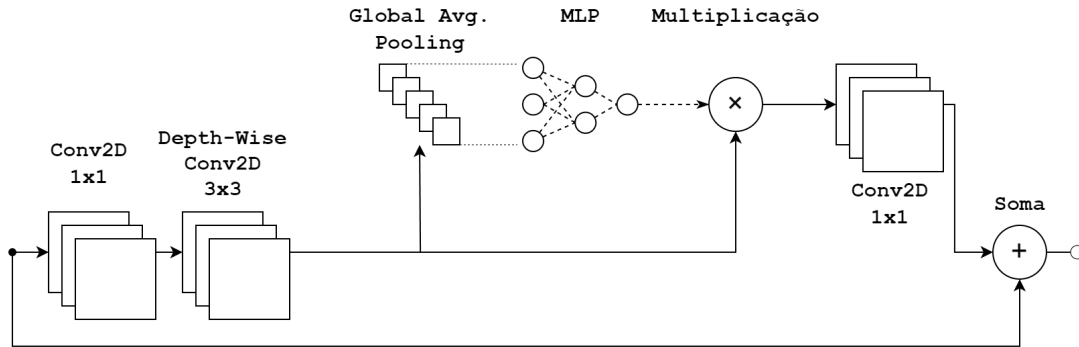


Fonte: Elaborado pelo autor.

As MBConv são partes fundamentais da arquitetura *EfficientNet*. Elas foram propostas pela primeira vez no trabalho de Sandler et al. (2019) e consistem em um gargalo residual invertido: expandem o número de canais nas camadas intermediárias e em seguida reduzem ao tamanho original, além de possuírem uma conexão residual entre o primeiro e o último bloco. A Figura 11 ilustra a estrutura do bloco MBConv, ele consiste em uma convolução 1×1 seguida de uma convolução *depth-wise* (detalhada no parágrafo seguinte), em seguida o resultado é encaminhado para um bloco de compressão-excitação, que consiste em uma camada de *pooling* por média, seguida de uma *mlp* cujo resultado é multiplicado com a entrada. Por fim, a saída deste bloco é encaminhada para mais uma convolução 1×1 e seu resultado é somado à entrada do bloco MBConv, formando a conexão residual e encaminhando a saída aos estágios seguintes de rede neural. O número que segue o nome em MBConv1 e MBConv6 é um multiplicador que indica o número de filtros na camada convolucional 1×1 mais à esquerda. Por exemplo, se o número de canais de entrada for 4, então o número de filtros será 4 no MBConv1 e 24 no MBConv6.

Por sua vez, o bloco de convolução *depth-wise* é um tipo de operação onde cada canal de entrada é processado por um neurônio diferente. De acordo com Chollet (2015), o processamento neste bloco se dá pelos seguintes passos: dividir a entrada em canais individuais, convoluir cada canal com um filtro, concatenar as saídas convolvidas ao longo do eixo dos canais.

Figura 11 – Diagrama do Bloco MBConv.



Fonte: Elaborado pelo autor.

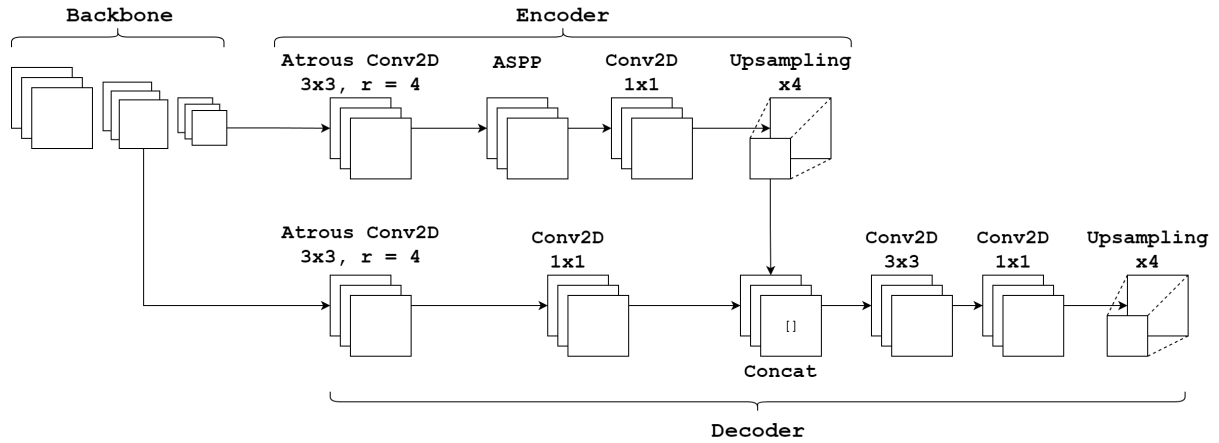
2.2.4 DeepLab

Este trabalho utiliza a estrutura *encoder-decoder*, baseando-se na arquitetura *DeepLabV3+* [Chen et al. \(2018b\)](#). A Figura 12 exibe o diagrama desta arquitetura. O bloco inicial, chamado de *backbone* é uma rede neural convolucional que pode ser modificada sem interferir no funcionamento do *encoder* ou *decoder*. Este trabalho modifica o *backbone* da rede, removendo o proposto no trabalho original, que é uma rede *Resnet* ([HE et al., 2015](#)) e inserindo no lugar a *EfficientNet* ([TAN; LE, 2019](#)) (os detalhes são explicitados na Seção 4.3).

O *encoder* é conectado com a camada final do *backbone*, recebe mapas de características compactos e os expande por meio do módulo ASPP, convoluções *atrous* e *upsampling*. Por sua vez, o *decoder* recebe como entrada mapas de características intermediários do *backbone*, e também a saída do *encoder*, para aplicar uma sequência de convoluções e *upsampling* (interpolação bilinear) ([KIRKLAND, 2010](#)), restaurando a imagem ao seu tamanho original e produzindo como saída uma máscara de segmentação.

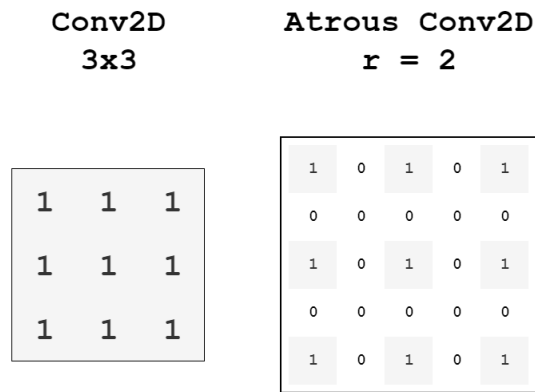
O módulo ASPP (*Atrous Spatial Pyramid Pooling*) consiste em uma pirâmide de *pooling* que consegue extrair e combinar características de diferentes resoluções utilizando convoluções dilatáveis, também chamadas de *atrous convolutions* ([CHEN et al., 2016](#)). Este tipo de convolução tem o objetivo de alcançar um campo de visão maior na extração de características. O seu vetor de pesos (filtro) é preenchido com zeros de acordo com uma taxa de dilatação escolhida (r). A Figura 13 exemplifica o efeito da taxa de dilatação sobre o filtro de convolução. Note que a área coberta pelo filtro passa de 3x3 para 5x5, mas o número de parâmetros permanece o mesmo. Variar a taxa de dilatação nas convoluções ajuda a capturar características em diversas escalas. Já a Figura 14 ilustra como o módulo ASPP utiliza convoluções *atrous* com diferentes taxas de dilatação para extrair *features*.

Figura 12 – Diagrama da Rede DeepLabV3+.



Fonte: Elaborado pelo autor.

Figura 13 – Convoluções Atrous

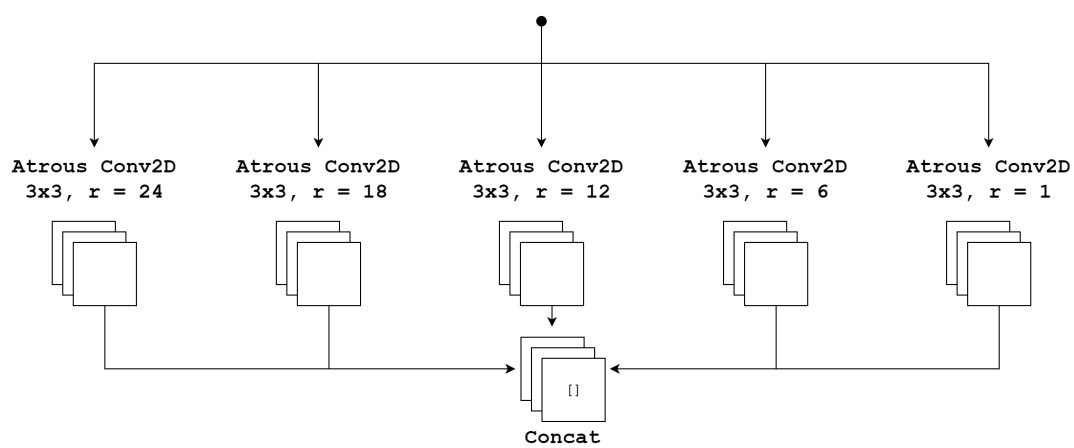


Fonte: Elaborado pelo autor.

2.3 Considerações Finais

Neste capítulo foram descritos conceitos envolvendo radioterapia, a aplicação de visão computacional em imagens médicas, fundamentos de processamento de imagens e segmentação de imagens, aprendizado de máquina, aprendizado profundo e redes neurais convolucionais. Estas informações compõem a base teórica para entendimento da metodologia proposta no Capítulo 4 e dos trabalhos relacionados apresentados no Capítulo 3.

Figura 14 – Módulo ASPP.



Fonte: Elaborado pelo autor.

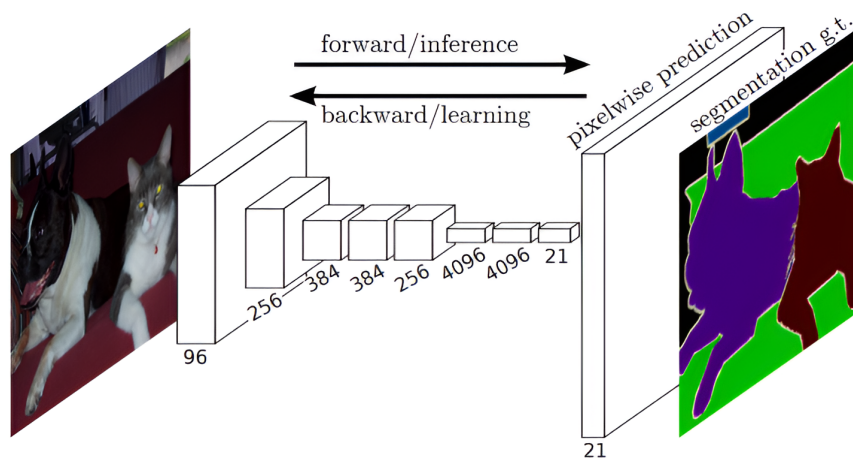
3 Trabalhos Relacionados

Neste capítulo são apresentados os trabalhos que ajudaram a fundamentar ou inspirar esta pesquisa. Eles estão distribuídos tanto na grande área de segmentação de imagens quanto especificamente na segmentação de Órgãos em Risco.

3.1 Segmentação Semântica

O trabalho de (LONG; SHELHAMER; DARRELL, 2014) é de grande importância para a segmentação de imagens via aprendizado profundo. Nele foram propostas as Redes Completamente Convolucionais (*FCN: Fully Convolutional Networks*) para segmentação semântica de imagens, contribuindo com a pesquisa nesta área ao propor uma arquitetura que substituiu as camadas densas (completamente conectadas) por camadas convolucionais e *upsampling* bilinear. Na época eram comuns esquemas de treinamentos apenas com sub-regiões das imagens, mas devido a sua arquitetura inovadora, as redes FCN geram uma saída do mesmo tamanho da entrada, possibilitando o treinamento fim-a-fim com imagens e máscaras de segmentação inteiras. A Figura 15 proporciona um panorama de uma arquitetura FCN, contendo as camadas convolucionais que diminuem sucessivamente a resolução da entrada ao mesmo tempo que extraem características. Ao final, uma camada de *upsampling* consegue transformar as características de baixa resolução de volta ao tamanho original, resultando em uma máscara de segmentação.

Figura 15 – Arquitetura das Redes FCN.

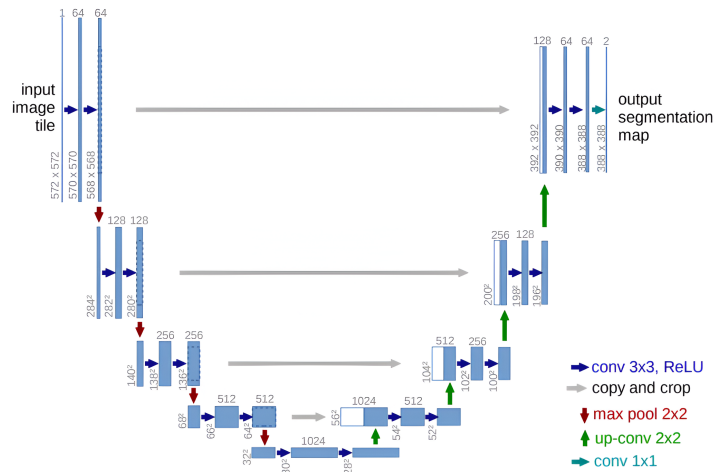


Fonte: Long, Shelhamer e Darrell (2014).

Uma das implementações de rede completamente convolucional com mais aplicações no contexto da segmentação de imagens médicas é a arquitetura U-net. Proposta por

Ronneberger, Fischer e Brox (2015), trouxe inovação para a área ao modificar a FCN, adicionando uma estrutura de expansão de características que é simétrica ao modelo de contração tradicional das redes convolucionais. A expansão ocorre a partir de camadas de convolução combinadas com *upsampling* aplicadas de maneira sucessiva, ficando conhecidas como *up-convolution*. Além disso, *features* de maior escala são concatenadas com a saída do *upsampling*, formando as *skip-connections*, que servem para suplementar o caminho de expansão com as *features* extraídas no caminho de contração da rede. A Figura 15 ilustra a arquitetura de uma rede U-net, que contém sucessivas camadas convolucionais, seguidas de uma estrutura simétrica composta de sucessivas *up-convolution*. A estrutura de contração-expansão é interconectada por *skip-connections* concatenadas com a saída das *up-convolutions*, resultando em uma máscara de segmentação mais precisa.

Figura 16 – Arquitetura da Rede U-net.



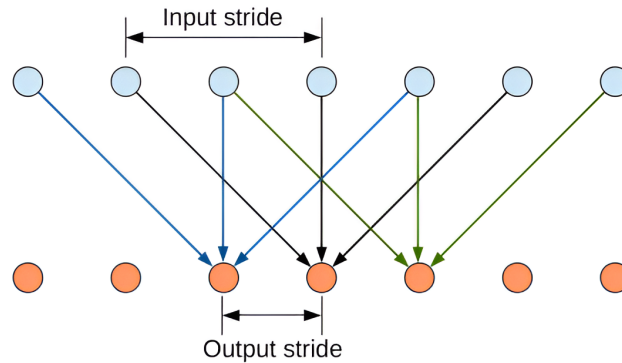
Fonte: Ronneberger, Fischer e Brox (2015).

Outro trabalho que norteou esta pesquisa iniciou-se com (CHEN et al., 2016), que propôs o modelo *DeepLab*. Este modelo aplicou convoluções *atrous* para manter o tamanho da entrada intacto enquanto extrai características de diferentes resoluções. Este tipo de convolução introduz esparsidade nos filtros, fazendo-os agir sobre uma área maior da imagem de entrada. A Figura 17 ilustra um *kernel* de convolução de tamanho 3x3 que produz uma saída de mesmo tamanho ao introduzir esparsidade nos pesos.

Posteriormente, o modelo *DeepLab* foi incrementado em (CHEN et al., 2018a) que adicionou o módulo *ASPP* (*atrous spatial pyramid pooling*) que consiste de convoluções *atrous* executadas paralelamente com diferentes taxas de dilatação aplicadas no mapa de características de entrada. Este módulo ajuda a rede a considerar diferentes tamanhos de objetos para uma mesma classe.

Já o trabalho iniciado por Howard et al. (2017) propõe a criação de redes eficientes voltadas para aplicações móveis, onde existem limitações de recursos como memória e poder

Figura 17 – Filtro de Convolução Atrous.

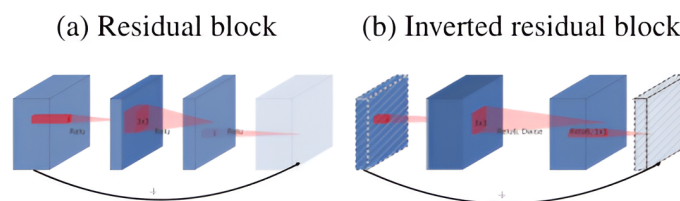


Fonte: [Chen et al. \(2016\)](#).

computacional. Nele foi analisado o uso das chamadas *depthwise separable convolutions*, que são um tipo de convolução subdivididas em: uma convolução *depthwise* (um filtro para cada canal) seguida por uma convolução *pointwise* (convolução convencional com kernel de 1x1). Este tipo inovador de convolução possibilitou a diminuição da quantidade de parâmetros do modelo, bem como a execução de menos operações de soma e multiplicação de matrizes pelo custo de 1% a menos de acurácia em seus testes. esta arquitetura ficou conhecida pelo nome de *MobileNet*.

Algum tempo depois, o trabalho de ([SANDLER et al., 2019](#)) seguiu na mesma linha de raciocínio e propôs a arquitetura *MobileNetV2*, cuja principal contribuição foi a análise da aplicação de gargalos residuais invertidos, que são uma modificação dos gargalos residuais propostos primeiramente em ([HE et al., 2015](#)). Enquanto o último segue uma lógica de conexões residuais entre as camadas de maior número de canais, os gargalos residuais invertidos realizam a conexão entre as camadas de menor número de canais. A Figura 18 ilustra a diferença entre os dois tipos de gargalos. Este tipo de estrutura tem o objetivo de comprimir a informação e em seguida expandi-la, obtendo apenas *features* essenciais para a tarefa, sendo que o gargalo invertido consome menos recursos computacionais e proporciona um modelo com menor quantidade de parâmetros treináveis.

Figura 18 – Gargalos Residuais.



Fonte: [Sandler et al. \(2019\)](#).

Na mesma direção, Tan e Le (2019) em seu trabalho, gerou uma arquitetura compacta que combina os gargalos residuais invertidos (SANDLER et al., 2019) com os blocos de compressão-expansão (HU; SHEN; SUN, 2018). Este trabalho inovou ao propor um método de escalabilidade de redes convolucionais, relacionando a resolução de entrada com a quantidade de canais de maneira proporcional através da escalabilidade composta (*compound scaling*). Dessa maneira foram geradas modelos gradativamente maiores a partir de um modelo-base menor.

3.2 Segmentação Semântica de Órgãos em Risco

Dentre os trabalhos relacionados que buscaram propor metodologias específicas para a segmentação de Órgãos em Risco, uma das abordagens é aproveitar a natureza tridimensional do exame de tomografia computadorizada. Em seu trabalho, Trullo et al. (2017) realizou a segmentação de diversos ORs utilizando uma arquitetura de rede neural completamente convolucional 3D conhecida como *Sharp Mask* (PINHEIRO et al., 2016). Essa arquitetura substitui o *upsampling* bilinear original por convoluções transpostas e adiciona o uso de *Conditional Random Fields* (CRF) (SUTTON; MCCALLUM, 2010) à saída da rede. Roth et al. (2018) analisou a utilização de redes V-net (MILLETARI; NAVAB; AHMADI, 2016) em conjunto com técnicas de pirâmide de resolução (ADELSON et al., 1984) e autocontexto (TU; BAI, 2010) para extrair não só a informação contextual multi-resolução, mas também a informação conjunta dos 3 eixos da TC, obtendo bons resultados na segmentação de órgãos.

Mais tarde, (WANG et al., 2019) aprimorou esta mesma arquitetura adicionando convoluções dilatáveis, pós-processamento e mais um estágio de pirâmide ao modelo anterior. Seu trabalho figura o estado da arte na segmentação de Órgãos em Risco, tendo obtido um Dice Score de 92,17% para a segmentação de traqueia. Ainda se aproveitando do contexto tridimensional da informação contida nos exames, (WANG et al., 2020) desenvolveu um método que, apesar de ser primariamente bidimensional, inclui informações da vizinhança dos *pixels* como segunda entrada para a rede neural *Xception* (CHOLLET, 2017). Além disso, esta rede foi modificada para aplicar convoluções *atrous* e um módulo *ASPP* também foi adicionado.

Outro tipo de abordagem na segmentação de Órgãos em Risco, é a utilização apenas dos *slices* bidimensionais gerados pelo exame, partindo do pressuposto de que a informação contida naquela fatia do exame é suficiente para a segmentação do órgão em questão. Dessa maneira, Feng et al. (2020) explorou como a informação contextual de diferentes camadas de uma rede neural poderiam ajudar na tarefa de segmentação semântica de órgãos em risco. O modelo proposto, nomeado *CPFNet*, trata-se de uma rede U-net onde as *skip connections* foram substituídas por um módulo que utiliza *upsampling*

para concatenar mapas de características de camadas mais abaixo e convoluções dilatáveis para obter *features* de diferentes resoluções. Além disso, em vez de convoluções tradicionais, foram utilizadas as convoluções separáveis propostas por (CHOLLET, 2017), que possuem menor quantidade de parâmetros treináveis.

Outra técnica comum no aprendizado profundo, é o uso de redes generativas adversárias (GAN) para treinar extratores de características e usá-los posteriormente em conjunto com uma rede de segmentação. Na segmentação de ORs, o trabalho que aplicou esta técnica foi o de (ZHAO; CHEN; LU, 2020), que combinou uma GAN pré-treinada, com uma rede U-net. Nesse contexto, a rede generativa adversária funcionou como um suplemento de informações relevantes para a rede U-net, obtendo resultados relevantes na segmentação de órgãos da região torácica.

Um novo tipo de modelo inovador adicionou mais uma possibilidade às redes neurais convolucionais: os Transformers. Inicialmente propostos para uso em processamento de linguagem natural, tratam-se de um tipo de camada de atenção em redes neurais, e foram adaptados para o uso em imagens no trabalho de Vaswani et al. (2017). Alguns trabalhos aplicaram este conceito em suas arquiteturas, como em (YAN et al., 2022), que modificou uma arquitetura U-net, adicionando *transformers* na estrutura de gargalo da rede, possibilitando o estudo sobre o desempenho deste tipo de modelo aplicado às imagens médicas.

A maioria dos trabalhos citados realizaram experimentos em uma base de dados em comum: a SEGTHOR, que foi disponibilizada em (LAMBERT et al., 2019) visando servir como base comparativa entre metodologias de segmentação de órgãos em risco. O trabalho reuniu tomografias computadorizadas de 60 pacientes em tratamento de câncer de pulmão ou linfoma de hodgkin, onde foram realizados experimentos com versões levemente modificadas da U-net (RONNEBERGER; FISCHER; BROX, 2015) que serviram de experimento base para trabalhos futuros. A Tabela 3 sumariza as metodologias estudadas na revisão de literatura realizada durante esta pesquisa, destacando seus respectivos *Dice Score* (métrica de avaliação mais comum na área) para a segmentação de traqueia e as bases de dados utilizadas.

Outra pesquisa que buscou montar um *dataset* referência foi o de (PODOBNIK et al., 2023). Os autores reuniram imagens de cabeça e pescoço de 56 pacientes que foram submetidos a tomografia computadorizada e ressonância magnética. Para cada paciente, foram obtidas segmentações de referência de até 30 Órgãos em Risco (OARs) por especialistas que realizaram a anotação manual dos *pixels* da imagem. Além disso, foram fornecidos resultados preliminares treinando a rede *U-Net* (RONNEBERGER; FISCHER; BROX, 2015) no conjunto de dados, que ficou conhecido como HaN-Seg.

Os trabalhos elencados na Tabela 3 obtiveram bom desempenho, entretanto, possuem uma característica em comum: o alto consumo de recursos computacionais.

Tabela 3 – Resumo dos Trabalhos Relacionados.

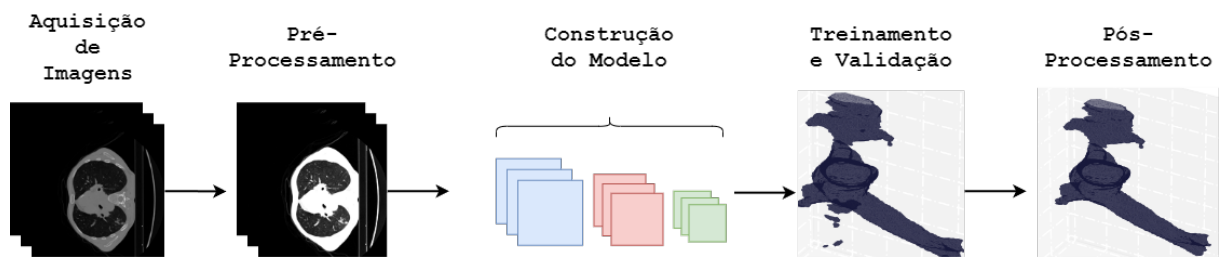
Paper	Method	DATASET	(%) Dice - traqueia	#Params
(TRULLO et al., 2017)	SharpMask	~	83.00	~
(WANG et al., 2019)	3D U-net	SEGTHOR	92,17	120M
(LAMBERT et al., 2019)	Simplified U-net	SEGTHOR	86.00	19M
(WANG et al., 2020)	U-net 2.5D	SEGTHOR	92,56	23M
(ZHAO; CHEN; LU, 2020)	U-net + GAN	SEGTHOR	88.30	~
(FENG et al., 2020)	U-net + SPP	SEGTHOR	89	26M
(DINIZ et al., 2021)	Res-unet + Atlas	LSTSC	~	~
(YAN et al., 2022)	U-net + Transformer	SEGTHOR	91,76	41,5M
(PODOBNIK et al., 2023)	U-net	HaN-Seg	~	24M

Isso acontece, pois todos eles possuem mais de 20 milhões de parâmetros treináveis, o que aumenta a demanda por memória e processamento destes métodos. Este trabalho propõe uma metodologia que equilibra custo computacional e qualidade na segmentação automática de traqueia.

4 Materiais e Método

Este capítulo apresenta detalhes sobre cada uma das etapas do método desenvolvido assim como sua análise para segmentação automática de traqueia com aprendizado profundo (*deep learning*), sendo elas: a aquisição das imagens; pré-processamento; treinamento da CNN, segmentação e pós-processamento. A Figura 19 resume o fluxo geral seguido pelo método proposto.

Figura 19 – Diagrama do Método Proposto.



Fonte: Elaborado pelo autor.

4.1 Aquisição da Imagens

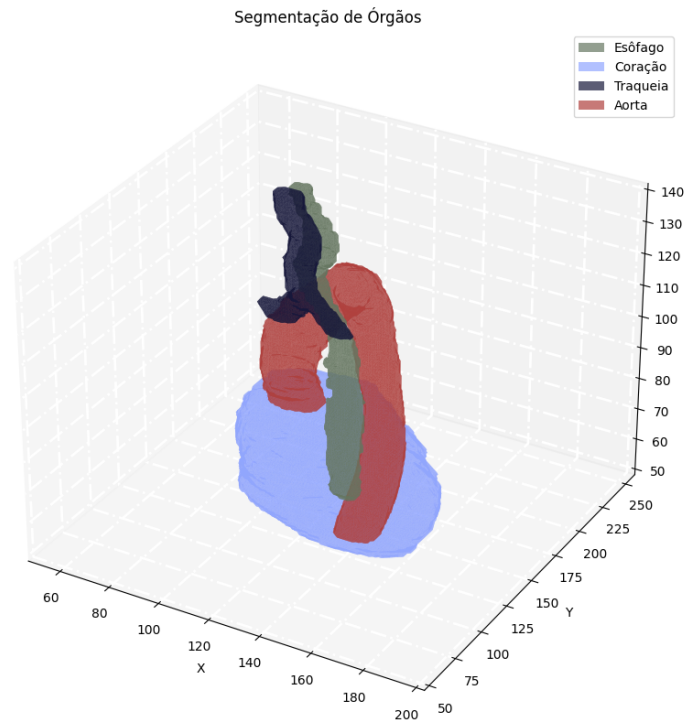
As imagens utilizadas para treinamento do modelo foram obtidas do conjunto de dados SEGTHOR (LAMBERT et al., 2019), que contém 60 volumes de TC, obtidos de pacientes com Câncer de Pulmão ou Linfoma de Hodgkins.

Dos 60 pacientes, 40 foram disponibilizados publicamente pelos autores do conjunto de dados. A partir dos volumes gerados pelos exames, foi possível extrair 7420 cortes transversais, dos quais 1987 imagens contém porções da traqueia, enquanto as 5433 restantes não. Os dados também incluem marcações manuais dos ORs feitas por médicos usadas como *labels* para o treinamento da rede neural. Os órgãos demarcados no dataset são: Esôfago, Traquéia, Coração e Aorta. A Figura 20 representa em 3 dimensões as marcações fornecidas pelos médicos, sendo a traqueia destacada pela cor azul-escuro.

4.2 Pré-Processamento

Adotar medidas de pré-processamento nas imagens antes de avançar para as etapas de treinamento do modelo e segmentação das amostras é essencial para aprimorar a qualidade dos dados de entrada. Nesta etapa busca-se minimizar ruídos, melhorar o contraste e corrigir possíveis defeitos presentes nas imagens, contribuindo para uma

Figura 20 – Demarcação de Órgãos Feita por Profissional.



Fonte: Elaborado pelo autor.

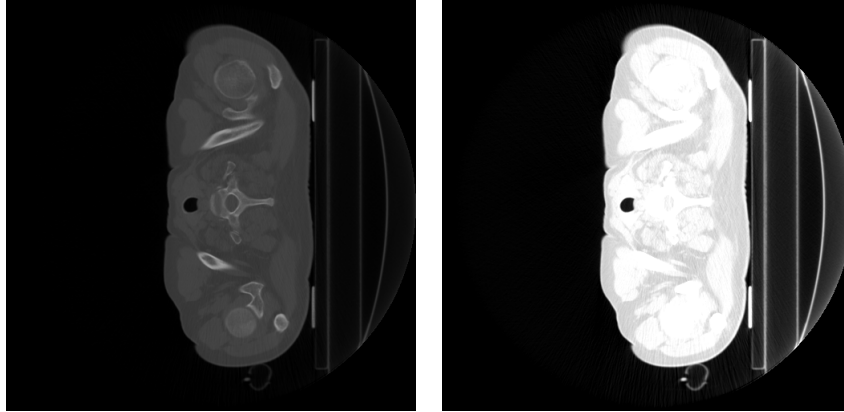
representação mais discriminativa dos objetos contidos na imagem. As subseções seguintes detalham os pré-processamentos executados sobre as imagens.

4.2.1 Janelamento de Intensidade

Uma das hipóteses deste trabalho consiste em avaliar o impacto do janelamento de intensidades como pré-processamento para a segmentação de traqueia. Visando encontrar de maneira automática um intervalo próximo do ótimo para a detecção de traqueia. Foi realizada otimização bayesiana durante 300 épocas utilizando a biblioteca de otimização de hiper-parâmetros *Optuna* (AKIBA et al., 2019). No caso do janelamento de intensidade, os hiper-parâmetros são o limite inferior e superior. Estes foram buscados nos intervalos entre $[-1000, 0]$ e $[10, 100]$, considerando apenas os números inteiros. A otimização resultou em uma janela com limite inferior de -800 e limite superior de 30 .

Em seguida, o janelamento de intensidade foi aplicado sobre os *slices*. A Figura 21 exemplifica a aplicação da janela sobre um recorte. À esquerda, encontra-se o recorte original. Já à direita, está representado o recorte após aplicação da janela $[-800, 30]$ encontrada no experimento. Nota-se o destaque da região de traqueia na imagem após o pré-processamento.

Figura 21 – Aplicação do Janelamento de Intensidade.



Fonte: Elaborado pelo autor.

4.2.2 Redimensionamento e Normalização

As imagens adquiridas do *dataset* SEGTHOR possuem tamanho 512×512 , ocupando considerável espaço na memória da GPU. Dada as limitações de memória do hardware, redimensionar as imagens para um tamanho menor fez com que fosse possível a execução do treinamento da rede neural em tempo hábil. Portanto, o tamanho selecionado foi o de 256×256 , visto que representa um equilíbrio entre a quantidade de dados e a qualidade de representação dos órgãos.

Em seguida foi executada a normalização dos *pixels*, cujos valores de intensidade estão contidos no intervalo $[-800, 30]$, encontrado via otimização de hiper-parâmetros via algoritmo TPE (*Tree-structured Parzen Estimator*) (AKIBA et al., 2019). Esta etapa facilita o aprendizado da rede neural, reduzindo a variação de intensidades entre imagens.

A normalização é muito comum em algoritmos de *Deep Learning* (KOO; CHA, 2017), (PIRES et al., 2020), (LUBANA; DICK; TANAKA, 2021) e serve para homogeneizar os dados de entrada, de maneira que o modelo de aprendizado possa entender as relações entre *pixels* de maneira mais generalizada. Existem diferentes tipos de normalização, algumas das mais comuns são: a normalização por média, média e desvio padrão e normalização mínimo-máximo. O tipo de normalização a ser utilizada neste trabalho é a normalização por média e desvio padrão, dada pela Equação 4.1.

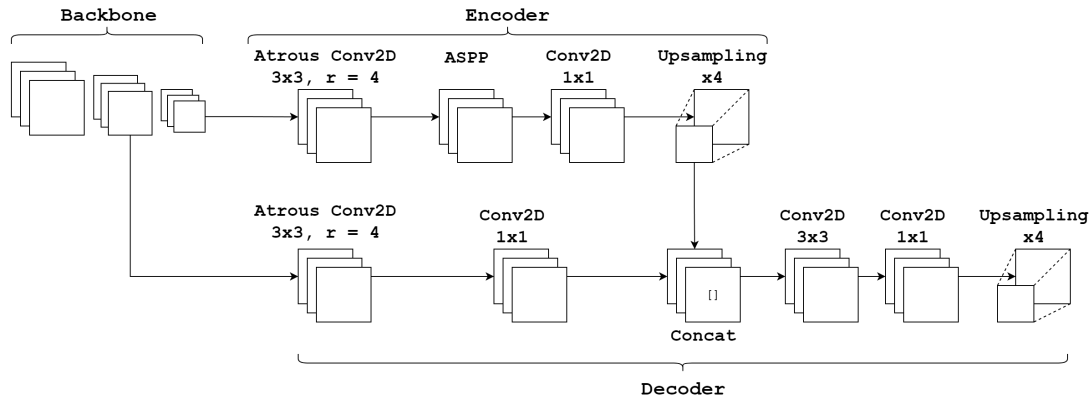
$$x' = \frac{x - x_{mean}}{std(x)} \quad (4.1)$$

4.3 Construção do Modelo

A primeira hipótese norteadora desta pesquisa supõe que a utilização de convoluções *atrous* (HOLSCHNEIDER et al., 1990) possa ajudar a construir um modelo compacto

para tarefas de segmentação de traqueia. Para por esta hipótese em prova, a estrutura *encoder/decoder* da rede *DeepLabV3+* (CHEN et al., 2018b), foi utilizada como base para a construção da arquitetura do modelo proposto, já que ela utiliza as convoluções *atrous* e o módulo ASPP *Atrous Spatial Pyramid Pooling* (HE et al., 2014) para promover uma fusão de características de várias resoluções. A Figura 22 representa o diagrama arquitetural da rede *DeepLabV3+* utilizada.

Figura 22 – Diagrama da Rede *DeepLabV3+*.



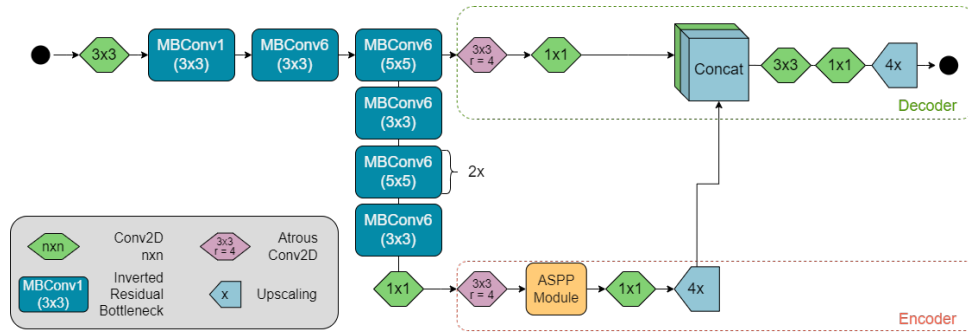
Fonte: Elaborado pelo autor.

Já a segunda hipótese levanta a possibilidade de utilizar a rede *EfficientNet* como *backbone* em conjunto com a rede *DeepLab* para executar a tarefa de segmentação semântica de traqueia. Por isso, o modelo proposto para a segmentação semântica da área da traqueia consiste em uma arquitetura de rede neural baseada em uma combinação das redes *DeepLabV3+* (CHEN et al., 2018b) e *EfficientNet* (TAN; LE, 2019). A escolha foi realizada baseando-se na capacidade da *EfficientNet* de ser um modelo de aprendizado profundo escalável, robusto e compacto em termos de parâmetros, que supera a *ResNet* em *benchmarks* recentes no conjunto de dados *ImageNet* (DENG et al., 2009), um dos principais *datasets* para comparação entre modelos de CNN.

Para a construção do modelo foi realizada uma modificação realizada na arquitetura *DeepLab* original: a substituição do *backbone ResNet* (HE et al., 2015) pela rede *EfficientNet* (TAN; LE, 2019), agregando mapas de características de camadas intermediárias, somando-os e interpolando-os de volta para a resolução de entrada, obtendo assim uma máscara de saída de segmentação. A Figura 23 proporciona uma visão geral da arquitetura do modelo.

Dessa maneira, este trabalho utiliza a estrutura de *encoder/decoder* para extrair características e restaurá-las para a resolução original. O modelo *EfficientNet* serve como *backbone* de extração de características, enquanto o módulo ASPP compõe o *encoder* do modelo e recebe como entrada os mapas de ativação da última camada convolucional da *EfficientNet*. As entradas do *decoder* são os mapas de ativação da terceira camada da *EfficientNet*. Elas são encaminhadas para o *decoder* composto por: uma convolução

Figura 23 – Visão Geral da Arquitetura do Modelo Proposto.



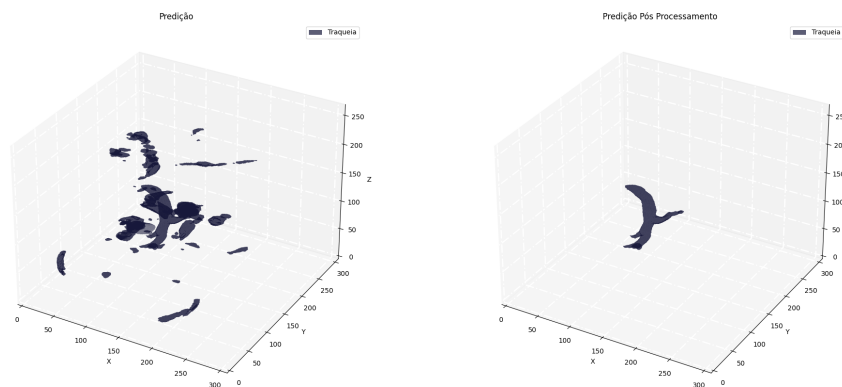
Fonte: Elaborado pelo autor.

separável *atrous*, seguida de uma camada convolucional 1×1 , cujo resultado é concatenado com a saída do *encoder* e enviado para uma convolução 3×3 seguida de outra 1×1 , que por fim, sofre *upsample*, retornando à resolução original da imagem. O resultado deste processo é a máscara de segmentação de órgãos.

4.4 Pós Processamento

Para eliminar os ruídos produzidos pelo modelo, foi aplicado o algoritmo de pós-processamento de componentes conectadas sobre o resultado da predição. Este algoritmo identifica os grupos de *voxels* vizinhos e retorna o grupo que possui maior quantidade de indivíduos. Dessa maneira foi possível eliminar da máscara de segmentação os *voxels* que não fazem parte da traqueia. A Figura 24 ilustra o funcionamento do pós-processamento, sendo que à esquerda está representado o resultado sem pós-processamento e à direita a máscara de segmentação 3D resultante da eliminação dos falsos positivos.

Figura 24 – Comparativo Após o Pós-Processamento.



Fonte: Elaborado pelo autor.

4.5 Considerações Finais

Neste capítulo foram detalhadas as etapas do método proposto para a segmentação automática da traqueia com aprendizado profundo. Isso inclui a aquisição das imagens do conjunto de dados SEGTHOR (Seção 4.1), o pré-processamento das imagens com janelamento de intensidade, redimensionamento e normalização (Seção 4.2), a construção e treinamento da rede neural convolucional com a combinação de *DeepLab V3+* e *EfficientNet* (Seção 4.3).

5 Resultados

Foram realizados experimentos para testar cada uma das hipóteses de pesquisas elencadas na Seção 1.1. Este capítulo apresenta os resultados obtidos em cada um destes experimentos.

5.1 Ambiente de Execução

O ambiente no qual os experimentos foram realizados consiste em um computador com processador Intel Core TMi5-10400F @ 2.90GHz e uma GPU Nvidia @RTX-3060 com 12GB de memória. Para a programação e implementação dos modelos de aprendizagem profunda, utilizou-se a linguagem de programação Python em conjunto com as bibliotecas: Tensorflow, Keras e Tensorflow Advanced Segmentation Models (KEZMANN, 2020).

5.2 Treinamento do Modelo com Convoluções Atrous e DeepLab

O treinamento do modelo *EfficientDeeplab* foi realizado utilizando os exames disponíveis na base de dados SEGTHOR. A separação entre treinamento, teste e validação seguiu o modelo de validação cruzada (*k-fold*) com $k = 3$. Cada *fold* está organizado da seguinte maneira: do total de 40 pacientes, são selecionados aproximadamente 65% para o treinamento, enquanto os 35% restantes são divididos igualmente entre validação e teste. Neste conjunto de dados as fatias que não contém traqueia são a maioria esmagadora. Por isso, elas foram excluídas das fases de treinamento e validação para que fosse obtido um *dataset* relativamente balanceado. Por outro lado, a etapa de testes é executada com todo o volume dos pacientes.

Foram executadas 100 épocas de treinamento, utilizando o otimizador SGD (ROBBINS; MONRO, 1951) em conjunto com o algoritmo de *backpropagation* (RUMELHART; HINTON; WILLIAMS, 1986) e a função de *loss binary crossentropy* para atualizar os neurônios da rede. Ao longo das 100 épocas, foram coletadas as métricas: *Dice Score* e *Jaccard Score*, como definidas na Seção 2.1.3. Ambas são comumente utilizadas para avaliar a sobreposição de duas máscaras de segmentação, neste caso a predita pelo modelo e a anotada pelos médicos. Quanto mais próximas estas medidas estiverem de 1, melhor a qualidade da segmentação de órgãos produzida.

O termo *EfficientNet* denomina uma família de redes derivadas de um modelo base mais compacto, chamado de B0. As redes derivadas variam de B1 a B7, com diferenças principalmente na resolução de entrada e no número de canais e camadas em cada estágio,

de maneira que, quanto maior o número no nome da rede, maior a quantidade de parâmetros treináveis. Para encontrar o modelo mais adequado para atuar em conjunto com a rede *DeepLabV3+* no problema da segmentação de traqueia, o treinamento foi executado para os modelos de B0, B1 e B2. Os modelos a partir de B3 não foram testados devido a limitações de memória. A Tabela 4 sumariza os *dice scores* coletados durante a etapa de validação para cada um dos experimentos realizados seguindo o esquema de validação cruzada.

Tabela 4 – Experimentos com Diferentes Backbones.

Model	Fold:	1	2	3	Mean	Std
EfficientDeeplab - B0	Dice	92,39	93,61	92,93	92,98	0,5
EfficientDeeplab - B1	Dice	93,35	93,24	93,25	93,28	0,04
EfficientDeeplab - B2	Dice	93,39	93,40	93,89	93,56	0,02

O modelo *EfficientDeepLab* que obteve o melhor desempenho foi o que utilizou a rede *EfficientNet B2* como backbone, tendo atingido um *Dice Score* de 93,31%, com desvio padrão de 0,09%. O alto *dice score* sugere ótima capacidade do modelo em segmentar traqueia, enquanto o baixo desvio padrão observado valida sua capacidade de generalização para exemplos ainda não observados. Isso prova verdadeira a hipótese 2, levantada na Seção 1.1. Além disso, observa-se que, dentre os modelos testados, a rede *EfficientNet B2* foi a que melhor trabalhou em conjunto com a arquitetura *DeepLabV3+*.

Para testar a hipótese 1, foram executados treinamentos em outras duas arquiteturas de redes neurais que não utilizam convoluções *atrous*: as redes U-net (RONNEBERGER; FISCHER; BROX, 2015) e SUnet (LAMBERT et al., 2019). Elas foram escolhidas, pois já foram testadas no artigo da própria base de dados SEGTHOR (LAMBERT et al., 2019). Para estes treinamentos foram utilizadas as mesmas configurações e mesma quantidade de épocas que o treinamento das *EfficientDeeplab*. Além disso, foi coletada a quantidade de parâmetros presente em cada um dos modelos testados. Os resultados da comparação estão elencados na Tabela 5.

Tabela 5 – Comparação de Desempenho com Modelos U-net.

Modelo	Dice Médio		Jaccard Médio		Sens. Média		#Params
EfficientDeeplab - B0	92,98	± 0,5	86,20	± 0,003	96,43	± 0,03	6.2 M
EfficientDeeplab - B1	93,28	± 0,04	87,79	± 0,002	99,46	± 0,004	8.7 M
EfficientDeeplab - B2	93,56	± 0,02	88,21	± 0,000	98,94	± 0,01	9.8 M
SUNET	92,73	± 1,14	86,54	± 0,003	96,57	± 0,02	4.8 M
UNET	93,73	± 0,1	88,10	± 0,01	98,69	± 0,01	19 M

Nota-se que a rede SUNET é a que possui menor quantidade de parâmetros e, ao mesmo tempo, o pior *dice score* médio. Enquanto isso, a rede *Unet* obteve a melhor métrica, porém maior quantidade de parâmetros. O modelo *Efficient Deeplab - B2* obtém um excelente equilíbrio entre *dice score* e quantidade de parâmetros, provando verdadeira

a hipótese levantada sobre o uso de convoluções *atrous* para a construção de um modelo compacto e que seja capaz de generalização.

5.3 Utilização de Janelamento de Intensidade

Para avaliar a hipótese 3, o melhor modelo obtido durante os experimentos da Seção 5.2 (*Efficient Deeplab - B2*) foi utilizado em dois testes. O primeiro ocorreu sem a introdução do janelamento de intensidade como pré-processamento. Já o segundo teste utilizou este pré-processamento com valores determinados por otimização bayesiana, como explicado na Seção 4.2.1. O resultado dos testes está resumido na Tabela 6. Observou-se que a adição do pré-processamento não ajudou a melhorar a qualidade da segmentação, chegando a diminuir o *dice score*, *jaccard score* e *sensibilidade* quando comparada ao experimento sem janelamento, tornando falsa a hipótese de pesquisa número 3.

Tabela 6 – Comparação de Janelamento de Intensidade com EfficientDeeplab - B2.

Experimento	Dice Médio	Jaccard Médio	Sens. Média
Sem Janelamento	93,56	88,31	98,94
Com Janelamento	93,31	88,21	98,23

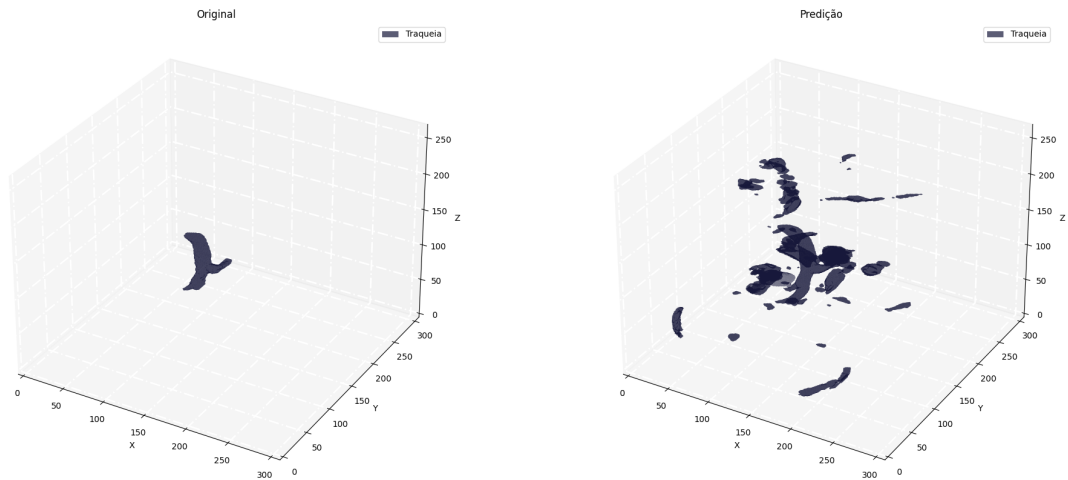
5.4 Estudo de Casos com Volumes Inteiros

Os pacientes separados para testes não tiveram os *slices* sem traqueia descartados, portanto, foi utilizado o volume inteiro obtido do exame de TC. Isto representa um caso de uso real do modelo, onde não se sabe quais *slices* contém traqueia ou não. Neste teste foram obtidas métricas inferiores aos resultados encontrados na validação, chegando a 39,32% de *dice score*, comparado aos 93,56% da fase de validação. Após cuidadosa análise, observou-se que a diminuição do *score* deve-se à presença de falsos positivos na máscara de segmentação de saída. A Figura 25 exibe à esquerda a máscara de segmentação proposta por médicos, e, à direita, a máscara obtida pela segmentação automática.

Isto ocorre porque algumas regiões da TC que não exibem traqueia, possuem características texturais e formas semelhantes ao órgão, levando o modelo a falsos positivos na predição. Na Figura 26 é possível observar uma fatia na qual não há presença da traqueia. No entanto, há outra estrutura corporal com forma e textura semelhantes que confundem o modelo, resultando na marcação equivocada dessa região como traqueia.

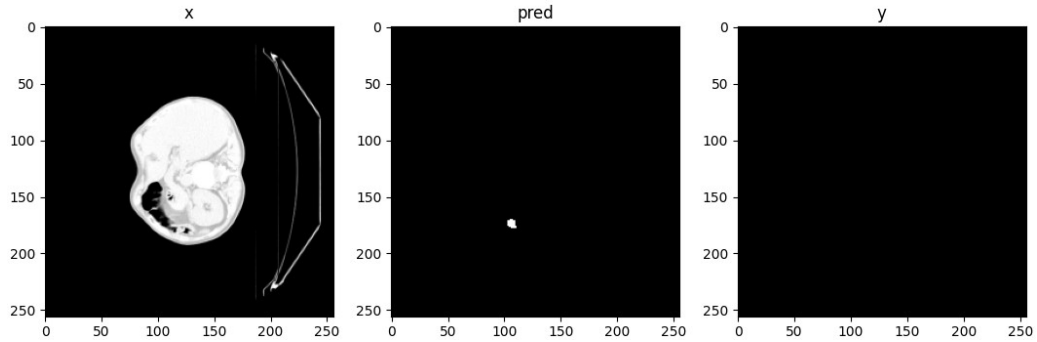
Já nas regiões em que há a presença do órgão, o modelo realiza a segmentação de forma correta, como ilustrado na Figura 27, onde a *EfficientDeeplab-B2* foi capaz de identificar corretamente a forma e a localização do órgão. Em outras regiões, a traqueia divide-se em dois tubos. Nestes casos o modelo também performou de maneira satisfatória, como exemplificado na Figura 28.

Figura 25 – Falsos Positivos na Segmentação.



Fonte: Elaborado pelo autor.

Figura 26 – Falso Positivo em Fatia.



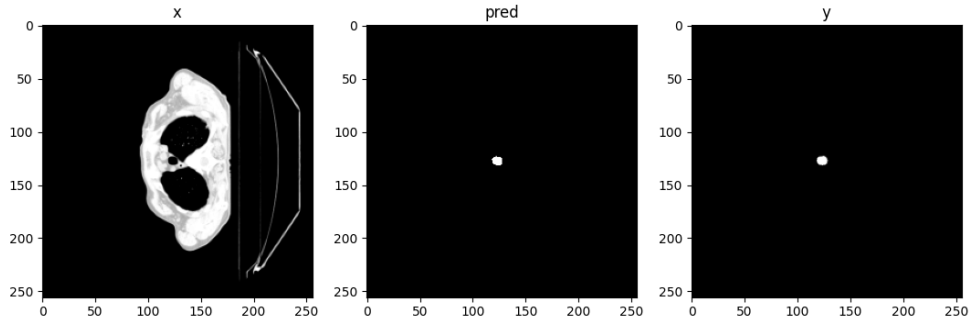
Fonte: Elaborado pelo autor.

Para eliminar os falsos positivos, foi realizada a aplicação do pós-processamento de componentes conectadas descrito na Seção 4.4. Observou-se melhora significativa do *dice score* para a segmentação de traqueia ao utilizar a rede *EfficientDeeplab-B2* seguida pelo pós-processamento de componentes conectadas, provando verdadeira a hipótese de pesquisa número 4. Os resultados estão elencados na Tabela 7.

Tabela 7 – Comparação de Resultados com Volumes Inteiros.

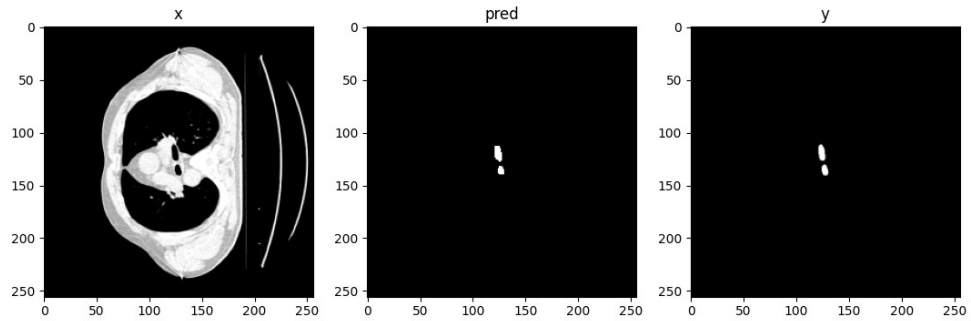
Modelo	Dice Médio	Dice Médio Pós-Proc.	Jaccard Médio	Jaccard Médio Pós-Proc.	Sens. Média	Sens. Média Pós-Proc.
EfficientDeeplab - B0	50,73	82,21	49,08	80,83	93,90	93,56
EfficientDeeplab - B1	42,64	82,08	40,94	80,51	92,73	92,49
EfficientDeeplab - B2	44,75	81,41	43,20	80,05	94,44	94,23

Figura 27 – Porção Central da Traqueia.



Fonte: Elaborado pelo autor.

Figura 28 – Divisão da Traqueia.



Fonte: Elaborado pelo autor.

5.5 Considerações

Nesta seção foram apresentados os resultados obtidos com a arquitetura proposta. A rede *EfficientDeeplab-B2* ofereceu os melhores resultados de validação, enquanto o modelo *B0* obteve as melhores métricas de segmentação em volumes inteiros, pois gera menos falsos positivos. Os resultados obtidos estão próximos aos de outros trabalhos encontrados na literatura, com a vantagem de possuir até 20 vezes menos parâmetros que o melhor modelo encontrado na literatura (WANG et al., 2019).

A arquitetura *EfficientDeeplab* incrementa o módulo *SPP* utilizado em (FENG et al., 2020) ao substituir as suas convoluções convencionais por convoluções *atrous*, que ocupam menos espaço em memória e alcançam resultados similares. Em relação ao pós-processamento, trabalhos como o de (TRULLO et al., 2017) utilizam-se de *conditional random fields* para redução de falsos positivos. Apesar de obter bons resultados, esta técnica adiciona mais complexidade e mais parâmetros ao modelo. Neste trabalho, a utilização do algoritmo de componentes conectadas elimina a necessidade de introduzir mais parâmetros na rede. Além disso, a utilização da rede *Deeplab* (CHEN et al., 2018b) torna a estrutura *encoder/decoder* mais compacta quando comparada às redes baseadas em *U-Net*, devido ao uso das convoluções *atrous* e *upsampling bilinear*, em contraste com as convoluções

transpostas utilizadas na rede *U-net*. Essa economia considerável de parâmetros resulta em uma rede neural compacta e robusta, que consome menos recursos computacionais que modelos semelhantes sem perdas consideráveis na qualidade de segmentação, tornando o método aplicável em ambientes com limitação de recursos. A Tabela 8 realiza uma comparação entre os principais trabalhos relacionados e a metodologia proposta.

Tabela 8 – Comparação com a Literatura.

Paper	Method	DATASET	(%) Dice - traqueia	#Params
(TRULLO et al., 2017)	SharpMask	~	83.00	~
(WANG et al., 2019)	3D U-net	SEGTHOR	92,17	120M
(LAMBERT et al., 2019)	Simplified U-net	SEGTHOR	86.00	19M
(WANG et al., 2020)	U-net 2.5D	SEGTHOR	92,56	23M
(ZHAO; CHEN; LU, 2020)	U-net + GAN	SEGTHOR	88.30	~
(FENG et al., 2020)	U-net + SPP	SEGTHOR	89	26M
(YAN et al., 2022)	U-net + Transformer	SEGTHOR	91,76	41,5M
(PODOBNIK et al., 2023)	U-net	HaN-Seg	~	24M
Este Trabalho	EfficientDeepLab - B0	SEGTHOR	82,21	6M

Por fim, vale destacar que o desenvolvimento deste trabalho resultou em duas produções acadêmicas: (FERNANDES et al., 2023) e (FERNANDES et al., 2024), respectivamente intituladas “EfficientDeepLab For Automated Trachea Segmentation On Medical Images” e “Eff-Unet For Trachea Segmentation On CT Scans”. Ambas as publicações foram concebidas durante etapas intermediárias do desenvolvimento e pesquisa deste projeto de mestrado.

6 Conclusão

Este trabalho propôs um método automatizado para segmentação da traqueia em imagens de TC usando a arquitetura de rede neural *EfficientDeeplab*. Os resultados obtidos são promissores e competitivos. O uso de convoluções dilatadas (*atrous*) e a versão *atrous* do agrupamento espacial de pirâmide permitiram a reconstrução precisa das máscaras de segmentação. A utilização eficiente de recursos computacionais foi possível ao adotar a arquitetura *EfficientNet* como *backbone* para extração de características, resultando em um modelo com apenas 6 milhões de parâmetros, que pode ser utilizado em sistemas embarcados de visão computacional para medicina por ser consideravelmente menor do que os métodos propostos na literatura atual para segmentação da traqueia com aprendizado profundo. O pós-processamento baseado em componentes conectadas melhorou a aplicação da rede neural em volumes inteiros de TC, reduzindo a quantidade de falsos positivos gerados pelo modelo, sendo, portanto, uma técnica adequada para utilização em conjunto com a segmentação via aprendizado profundo.

Por fim, há espaço para melhorias futuras, como a adição de novas técnicas de pré-processamento e a agregar informações 3D sem haver aumento do consumo de recursos. Outra possibilidade é integrar uma arquitetura para redução de falsos positivos que consiga ser treinada em conjunto com o modelo *EfficientDeeplab*. Em suma, a metodologia proposta pode ajudar na obtenção de um planejamento radioterápico mais rápido e, consequentemente, melhorar a qualidade de vida dos pacientes de câncer.

Referências

- ACKERMANS, L. L. G. C.; VOLMER, L.; WEE, L.; BRECHEISEN, R.; SÁNCHEZ-GONZÁLEZ, P.; SEIFFERT, A. P.; GÓMEZ, E. J.; DEKKER, A.; BOSCH, J. A. T.; DAMINK, S. M. W. O.; BLOKHUIS, T. J. Deep learning automated segmentation for muscle and adipose tissue from abdominal computed tomography in polytrauma patients. *Sensors*, v. 21, n. 6, 2021. ISSN 1424-8220. Disponível em: <<https://www.mdpi.com/1424-8220/21/6/2083>>. Citado na página 21.
- ADELSON, E. H.; ANDERSON, C. H.; BERGEN, J. R.; BURT, P. J.; OGDEN, J. M. Pyramid methods in image processing. *RCA engineer*, v. 29, n. 6, p. 33–41, 1984. Citado na página 35.
- AKIBA, T.; SANO, S.; YANASE, T.; OHTA, T.; KOYAMA, M. Optuna: A next-generation hyperparameter optimization framework. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. [S.l.: s.n.], 2019. Citado 2 vezes nas páginas 39 e 40.
- AMARI, S. A theory of adaptive pattern classifiers. *IEEE Transactions on Electronic Computers*, IEEE, n. 3, p. 299–307, 1967. Citado na página 24.
- BARAZZUOL, L.; COPPES, R. P.; LUIJK, P. van. Prevention and treatment of radiotherapy-induced side effects. *Molecular Oncology*, v. 14, n. 7, p. 1538–1554, 2020. Disponível em: <<https://febs.onlinelibrary.wiley.com/doi/abs/10.1002/1878-0261.12750>>. Citado na página 17.
- BURKOV, A. *Machine Learning Engineering*. True Positive Incorporated, 2020. ISBN 9781999579579. Disponível em: <<https://books.google.com.br/books?id=HeXizQEACAAJ>>. Citado na página 23.
- CHEN, L.-C.; PAPANDREOU, G.; KOKKINOS, I.; MURPHY, K.; YUILLE, A. L. *Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs*. 2016. Citado 3 vezes nas páginas 29, 33 e 34.
- CHEN, L.-C.; PAPANDREOU, G.; KOKKINOS, I.; MURPHY, K.; YUILLE, A. L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, v. 40, n. 4, p. 834–848, April 2018. ISSN 0162-8828. Disponível em: <<https://doi.org/10.1109/TPAMI.2017.2699184>>. Citado na página 33.
- CHEN, L.-C.; ZHU, Y.; PAPANDREOU, G.; SCHROFF, F.; ADAM, H. *Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation*. 2018. Citado 4 vezes nas páginas 15, 29, 41 e 48.
- CHOLLET, F. *Keras*. 2015. <https://keras.io/api/layers/convolution_layers/depthwise_convolution2d/>. Citado na página 28.
- CHOLLET, F. Xception: Deep learning with depthwise separable convolutions. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, 2017. p. 1800–1807. ISSN 1063-6919. Disponível em:

<<https://doi.ieeecomputersociety.org/10.1109/CVPR.2017.195>>. Citado 2 vezes nas páginas 35 e 36.

CONSIDERA, D. P.; MARCHIORI, E.; JR, A. S. S.; ZANETTI, G.; ESCUISSATO, D. L.; GASPARETTO, E. L.; NETO, C. d. A.; LIMA, R. d. S. L.; XAVIER, S. S.; PEDROSA, R. C. A tomografia computadorizada de alta resolução na avaliação da toxicidade pulmonar por amiodarona. *Radiologia Brasileira*, SciELO Brasil, v. 39, p. 113–118, 2006. Citado na página 20.

DANCE, D. R.; CHRISTOFIDES, S.; MAIDMENT, A. D.; MCLEAN, I. D.; NG, K. H. *Diagnostic Radiology Physics*. Vienna: INTERNATIONAL ATOMIC ENERGY AGENCY, 2014. (Non-serial Publications). ISBN 978-92-0-131010-1. Disponível em: <<https://www.iaea.org/publications/8841/diagnostic-radiology-physics>>. Citado na página 19.

DENG, J.; DONG, W.; SOCHER, R.; LI, L.-J.; LI, K.; FEI-FEI, L. Imagenet: A large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2009. p. 248–255. Citado na página 41.

DHAWAN, A.; BODANI, P.; GARG, V. Post processing of image segmentation using conditional random fields. In: *2019 6th International Conference on Computing for Sustainable Global Development (INDIACom)*. [S.l.: s.n.], 2019. p. 729–734. Citado na página 22.

DICE, L. R. Measures of the amount of ecologic association between species. *Ecology*, v. 26, n. 3, p. 297–302, 1945. Disponível em: <<https://esajournals.onlinelibrary.wiley.com/doi/abs/10.2307/1932409>>. Citado na página 22.

DINIZ, J.; FERREIRA, J.; SILVA, G.; QUINTANILHA, D.; SILVA, A.; PAIVA, A. Segmentação de coração em tomografias computadorizadas utilizando atlas probabilístico e redes neurais convolucionais. In: *Anais do XXI Simpósio Brasileiro de Computação Aplicada à Saúde*. Porto Alegre, RS, Brasil: SBC, 2021. p. 83–94. ISSN 2763-8952. Disponível em: <<https://sol.sbc.org.br/index.php/sbcas/article/view/16055>>. Citado 3 vezes nas páginas 13, 14 e 37.

EELBODE, T.; BERTELS, J.; BERMAN, M.; VANDERMEULEN, D.; MAES, F.; BISSCHOPS, R.; BLASCHKO, M. B. Optimization for medical image segmentation: Theory and practice when evaluating with dice score or jaccard index. *IEEE Transactions on Medical Imaging*, v. 39, n. 11, p. 3679–3690, 2020. Citado na página 22.

EVANS, E.; STAFFURTH, J. Principles of cancer treatment by radiotherapy. *Surgery (Oxford)*, v. 36, n. 3, p. 111–116, 2018. ISSN 0263-9319. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0263931917302715>>. Citado 2 vezes nas páginas 17 e 18.

FENG, S.; ZHAO, H.; SHI, F.; CHENG, X.; WANG, M.; MA, Y.; XIANG, D.; ZHU, W.; CHEN, X. Cpfnet: Context pyramid fusion network for medical image segmentation. *IEEE Transactions on Medical Imaging*, v. 39, n. 10, p. 3008–3018, 2020. Citado 5 vezes nas páginas 14, 35, 37, 48 e 49.

FERNANDES, A. G. S.; JUNIOR, G. B.; DINIZ, J. O. B.; SILVA, A. C.; MATOS, C. E. F. Efficientdeeplab for automated trachea segmentation on medical images. In:

- NALDI, M. C.; BIANCHI, R. A. C. (Ed.). *Intelligent Systems*. Cham: Springer Nature Switzerland, 2023. p. 154–166. ISBN 978-3-031-45389-2. Citado na página 49.
- FERNANDES, A. G. S.; JUNIOR, G. B.; DINIZ, J. O. B.; FERREIRA, M. M.; JUNIOR, J. R. D. R.; SILVA, M. L. J. D.; GONÇALVES, L. A. *Eff-Unet For Trachea Segmentation On CT Scans*. 2024. <https://link.springer.com/book/9783031606649>. Citado na página 49.
- FU, Y.; LEI, Y.; WANG, T.; CURRAN, W. J.; LIU, T.; YANG, X. A review of deep learning based methods for medical image multi-organ segmentation. *Physica Medica*, v. 85, p. 107–122, 2021. ISSN 1120-1797. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1120179721001848>. Citado na página 14.
- GONZALEZ, R. C.; WOODS, R. E. *Digital image processing*. Upper Saddle River, N.J.: Prentice Hall, 2008. ISBN 9780131687288 013168728X 9780135052679 013505267X. Disponível em: <http://www.amazon.com/Digital-Image-Processing-3rd-Edition/dp/013168728X>. Citado 3 vezes nas páginas 18, 19 e 20.
- GOODFELLOW, I.; POUGET-ABADIE, J.; MIRZA, M.; XU, B.; WARDE-FARLEY, D.; OZAIR, S.; COURVILLE, A.; BENGIO, Y. Generative adversarial nets. In: GHAHRAMANI, Z.; WELLING, M.; CORTES, C.; LAWRENCE, N.; WEINBERGER, K. (Ed.). *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2014. v. 27. Disponível em: https://proceedings.neurips.cc/paper_files/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf. Citado na página 27.
- GOODFELLOW, I. J.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. Citado na página 25.
- HAYKIN, S. S. *Neural networks and learning machines*. Third. Upper Saddle River, NJ: Pearson Education, 2009. Citado 3 vezes nas páginas 23, 24 e 25.
- HE, K.; ZHANG, X.; REN, S.; SUN, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. In: *Computer Vision textendash ECCV 2014*. [S.l.]: Springer International Publishing, 2014. p. 346–361. Citado na página 41.
- HE, K.; ZHANG, X.; REN, S.; SUN, J. *Deep Residual Learning for Image Recognition*. 2015. Citado 4 vezes nas páginas 27, 29, 34 e 41.
- HECHT-NIELSEN, R. *Neurocomputing*. USA: Addison-Wesley Longman Publishing Co., Inc., 1989. ISBN 0201093553. Citado na página 23.
- HOLSCHNEIDER, M.; KRONLAND-MARTINET, R.; MORLET, J.; TCHAMITCHIAN, P. A real-time algorithm for signal analysis with the help of the wavelet transform. In: COMBES, J.-M.; GROSSMANN, A.; TCHAMITCHIAN, P. (Ed.). *Wavelets*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1990. p. 286–297. ISBN 978-3-642-75988-8. Citado 2 vezes nas páginas 15 e 40.
- 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. *CoRR*, abs/1704.04861, 2017. Disponível em: <http://dblp.uni-trier.de/db/journals/corr/corr1704.html#HowardZCKWWAA17>. Citado na página 33.

- HU, J.; SHEN, L.; SUN, G. Squeeze-and-excitation networks. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2018. p. 7132–7141. Citado na página 35.
- ISENSEE, F.; JÄGER, P. F.; FULL, P. M.; VOLLMUTH, P.; MAIER-HEIN, K. H. nnu-net for brain tumor segmentation. In: CRIMI, A.; BAKAS, S. (Ed.). *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Cham: Springer International Publishing, 2021. p. 118–132. ISBN 978-3-030-72087-2. Citado na página 22.
- JACCARD, P. The distribution of the flora in the alpine zone.1. *New Phytologist*, v. 11, n. 2, p. 37–50, 1912. Disponível em: <<https://nph.onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-8137.1912.tb05611.x>>. Citado na página 22.
- KAK, A. C.; SLANEY, M. *Principles of Computerized Tomographic Imaging*. Society for Industrial and Applied Mathematics, 2001. Disponível em: <<https://epubs.siam.org/doi/abs/10.1137/1.9780898719277>>. Citado na página 19.
- KEZMANN, J.-M. *Tensorflow Advanced Segmentation Models*. [S.l.]: GitHub, 2020. <<https://github.com/JanMarcelKezmann/TensorFlow-Advanced-Segmentation-Models>>. Citado na página 44.
- KIRKLAND, E. J. Bilinear interpolation. In: _____. *Advanced Computing in Electron Microscopy*. Boston, MA: Springer US, 2010. p. 261–263. ISBN 978-1-4419-6533-2. Disponível em: <https://doi.org/10.1007/978-1-4419-6533-2_12>. Citado na página 29.
- KOO, K.-M.; CHA, E.-Y. Image recognition performance enhancements using image normalization. *Human-centric Computing and Information Sciences*, v. 7, n. 1, p. 33, Nov 2017. ISSN 2192-1962. Disponível em: <<https://doi.org/10.1186/s13673-017-0114-5>>. Citado na página 40.
- LAMBERT, Z.; PETITJEAN, C.; DUBRAY, B.; RUAN, S. *SegTHOR: Segmentation of Thoracic Organs at Risk in CT images*. arXiv, 2019. Disponível em: <<https://arxiv.org/abs/1912.05950>>. Citado 8 vezes nas páginas 13, 14, 21, 36, 37, 38, 45 e 49.
- LECUN, Y.; BOSER, B.; DENKER, J. S.; HENDERSON, D.; HOWARD, R. E.; HUBBARD, W.; JACKEL, L. D. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, v. 1, n. 4, p. 541–551, 1989. Citado 2 vezes nas páginas 26 e 27.
- LEE, C.-T.; KABIR, T.; NELSON, J.; SHENG, S.; MENG, H.-W.; DYKE, T. E. V.; WALJI, M. F.; JIANG, X.; SHAMS, S. Use of the deep learning approach to measure alveolar bone level. *J. Clin. Periodontol.*, Wiley, v. 49, n. 3, p. 260–269, mar. 2022. Citado na página 21.
- LONG, J.; SHELHAMER, E.; DARRELL, T. *Fully Convolutional Networks for Semantic Segmentation*. arXiv, 2014. Disponível em: <<https://arxiv.org/abs/1411.4038>>. Citado 2 vezes nas páginas 27 e 32.
- LUBANA, E. S.; DICK, R.; TANAKA, H. Beyond batchnorm: Towards a unified understanding of normalization in deep learning. In: RANZATO, M.; BEYGELZIMER, A.; DAUPHIN, Y.; LIANG, P.; VAUGHAN, J. W. (Ed.). *Advances in Neural*

- Information Processing Systems*. Curran Associates, Inc., 2021. v. 34, p. 4778–4791. Disponível em: <https://proceedings.neurips.cc/paper_files/paper/2021/file/2578eb9cdf020730f77793e8b58e165a-Paper.pdf>. Citado na página 40.
- MARCHIORI, E.; GONÇALVES, C. M.; ESCUISSATO, D. L.; TEIXEIRA, K.-I.-S. S.; RODRIGUES, R.; BARRETO, M. M.; ESTEVES, M. Microlitíase alveolar pulmonar: achados na tomografia computadorizada de alta resolução do tórax em 10 pacientes. *Jornal Brasileiro de Pneumologia*, Sociedade Brasileira de Pneumologia e Tisiologia, v. 33, n. 5, p. 552–557, Sep 2007. ISSN 1806-3713. Disponível em: <<https://doi.org/10.1590/S1806-37132007000500010>>. Citado na página 20.
- MEHTA, S.; SUHAG, V.; SEMWAL, M.; SHARMA, N. Radiotherapy: Basic concepts and recent advances. *Medical Journal Armed Forces India*, v. 66, n. 2, p. 158–162, 2010. ISSN 0377-1237. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0377123710801327>>. Citado na página 17.
- MILLETARI, F.; NAVAB, N.; AHMADI, S.-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: *2016 Fourth International Conference on 3D Vision (3DV)*. [S.l.: s.n.], 2016. p. 565–571. Citado na página 35.
- NOEL, G.; ANTONI, D.; BARILLOT, I.; CHAUVET, B. Délinéation des organes à risque et contraintes dosimétriques. *Cancer Radiothérapie*, v. 20, p. S36–S60, 2016. ISSN 1278-3218. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1278321816301676>>. Citado na página 13.
- PINHEIRO, P. O.; LIN, T.-Y.; COLLOBERT, R.; DOLL’AR, P. Learning to refine object segments. In: LEIBE, B.; MATAS, J.; SEBE, N.; WELLING, M. (Ed.). *Computer Vision – ECCV 2016*. Cham: Springer International Publishing, 2016. p. 75–91. ISBN 978-3-319-46448-0. Citado na página 35.
- PIRES, I. M.; HUSSAIN, F.; GARCIA, N. M.; LAMESKI, P.; ZDRAVEVSKI, E. Homogeneous data normalization and deep learning: A case study in human activity classification. *Future Internet*, v. 12, n. 11, 2020. ISSN 1999-5903. Disponível em: <<https://www.mdpi.com/1999-5903/12/11/194>>. Citado na página 40.
- PODOBNIK, G.; STROJAN, P.; PETERLIN, P.; IBRAGIMOV, B.; VRTOVEC, T. Han-seg: The head and neck organ-at-risk ct and mr segmentation dataset. *Medical Physics*, v. 50, n. 3, p. 1917–1927, 2023. Disponível em: <<https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/mp.16197>>. Citado 3 vezes nas páginas 36, 37 e 49.
- PRINCE, S. *Computer Vision: Models Learning and Inference*. [S.l.]: Cambridge University Press, 2012. Citado na página 18.
- ROBBINS, H.; MONRO, S. A stochastic approximation method. *The Annals of Mathematical Statistics*, Institute of Mathematical Statistics, v. 22, n. 3, p. 400 – 407, 1951. Disponível em: <<https://doi.org/10.1214/aoms/1177729586>>. Citado 2 vezes nas páginas 24 e 44.
- RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: NAVAB, N.; HORNEGGER, J.; WELLS, W. M.; FRANGI, A. F. (Ed.). *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Cham: Springer International Publishing, 2015. p. 234–241. ISBN 978-3-319-24574-4. Citado 4 vezes nas páginas 27, 33, 36 e 45.

- ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, v. 65 6, p. 386–408, 1958. Disponível em: <<https://api.semanticscholar.org/CorpusID:12781225>>. Citado 2 vezes nas páginas 23 e 24.
- ROTH, H. R.; SHEN, C.; ODA, H.; SUGINO, T.; ODA, M.; HAYASHI, Y.; MISAWA, K.; MORI, K. A multi-scale pyramid of 3d fully convolutional networks for abdominal multi-organ segmentation. In: FRANGI, A. F.; SCHNABEL, J. A.; DAVATZIKOS, C.; ALBEROLA-LÓPEZ, C.; FICHTINGER, G. (Ed.). *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*. Cham: Springer International Publishing, 2018. p. 417–425. ISBN 978-3-030-00937-3. Citado na página 35.
- RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. *Nature*, v. 323, n. 6088, p. 533–536, Oct 1986. ISSN 1476-4687. Disponível em: <<https://doi.org/10.1038/323533a0>>. Citado 2 vezes nas páginas 24 e 44.
- RUSSELL, S.; NORVIG, P. *Artificial Intelligence: A Modern Approach*. 3rd. ed. USA: Prentice Hall Press, 2009. ISBN 0136042597. Citado na página 22.
- SAMET, H.; TAMMINEN, M. Efficient component labeling of images of arbitrary dimension represented by linear bintrees. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 10, n. 4, p. 579–586, 1988. Citado na página 22.
- SANDLER, M.; HOWARD, A.; ZHU, M.; ZHMOGINOV, A.; CHEN, L.-C. *MobileNetV2: Inverted Residuals and Linear Bottlenecks*. 2019. Citado 3 vezes nas páginas 28, 34 e 35.
- SEERAM, E. *Computed tomography : physical principles, clinical applications, and quality control*. Fourth edition. St. Louis, Missouri: Elsevier St. Louis, Missouri, 2016. ISBN 9780323312882; 0323312888. Citado na página 19.
- SHALEV-SHWARTZ, S.; BEN-DAVID, S. *Understanding Machine Learning - From Theory to Algorithms*. [S.l.]: Cambridge University Press, 2014. I-XVI, 1-397 p. ISBN 978-1-10-705713-5. Citado na página 23.
- SHIVHARE, P.; GUPTA, V. Review of image segmentation techniques including pre & post processing operations. *International Journal of Engineering and Advanced Technology*, Citeseer, v. 4, n. 3, p. 153–157, 2015. Citado na página 22.
- SIMONYAN, K.; ZISSERMAN, A. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. 2015. Citado na página 27.
- SUTTON, C.; MCCALLUM, A. *An Introduction to Conditional Random Fields*. 2010. Citado na página 35.
- TAN, M.; LE, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In: CHAUDHURI, K.; SALAKHUTDINOV, R. (Ed.). *Proceedings of the 36th International Conference on Machine Learning*. PMLR, 2019. (Proceedings of Machine Learning Research, v. 97), p. 6105–6114. Disponível em: <<https://proceedings.mlr.press/v97/tan19a.html>>. Citado 5 vezes nas páginas 15, 28, 29, 35 e 41.
- TEKATLI, H.; DUIJM, M.; HOOP, E. Oomen-de; VERBAKEL, W.; SCHILLEMANS, W.; SLOTMAN, B.; NUYTENS, J.; SENAN, S. Normal tissue complication probability modeling of pulmonary toxicity after stereotactic and hypofractionated radiation therapy

- for central lung tumors. *International Journal of Radiation Oncology Biology Physics*, Elsevier, v. 100, n. 3, p. 738–747, 2018. Citado 2 vezes nas páginas 13 e 17.
- THANOON, M. A.; ZULKIFLEY, M. A.; ZAINURI, M. A. A. M.; ABDANI, S. R. A review of deep learning techniques for lung cancer screening and diagnosis based on ct images. *Diagnostics (Basel)*, 2023. Citado na página 21.
- TRULLO, R.; PETITJEAN, C.; RUAN, S.; DUBRAY, B.; NIE, D.; SHEN, D. Segmentation of organs at risk in thoracic ct images using a sharpmask architecture and conditional random fields. In: *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. [S.l.: s.n.], 2017. p. 1003–1006. Citado 4 vezes nas páginas 35, 37, 48 e 49.
- TU, Z.; BAI, X. Auto-context and its application to high-level vision tasks and 3d brain image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 32, n. 10, p. 1744–1757, 2010. Citado na página 35.
- VAIDYA, J. S. Principles of cancer treatment by radiotherapy. *Surgery (Oxford)*, v. 39, n. 4, p. 193–201, 2021. ISSN 0263-9319. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0263931921000363>>. Citado na página 17.
- VASWANI, A.; SHAZEER, N.; PARMAR, N.; USZKOREIT, J.; JONES, L.; GOMEZ, A. N.; KAISER, L. u.; POLOSUKHIN, I. Attention is all you need. In: GUYON, I.; LUXBURG, U. V.; BENGIO, S.; WALLACH, H.; FERGUS, R.; VISHWANATHAN, S.; GARNETT, R. (Ed.). *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2017. v. 30. Disponível em: <https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf>. Citado na página 36.
- VICENTE, A. d. O.; PENIDO, N. d. O.; YAMASHITA, H. K.; ALBERNAZ, P. L. M. Tomografia computadorizada no diagnóstico da otosclerose fenestral. *Revista Brasileira de Otorrinolaringologia*, ABORL-CCF Associação Brasileira de Otorrinolaringologia e Cirurgia Cérvico-Facial, v. 70, n. 1, p. 66–73, Jan 2004. ISSN 0034-7299. Disponível em: <<https://doi.org/10.1590/S0034-72992004000100012>>. Citado na página 20.
- WANG, Q.; ZHAO, W.; ZHANG, C.; ZHANG, L.; WANG, C.; LI, Z.; CUI, S.; LI, G. 3d enhanced multi-scale network for thoracic organs segmentation. *SegTHOR@ ISBI*, v. 3, n. 1, p. 1–5, 2019. Citado 4 vezes nas páginas 35, 37, 48 e 49.
- WANG, S.; CAO, S.; CHAI, Z.; WEI, D.; MA, K.; WANG, L.; ZHENG, Y. Conquering data variations in resolution: A slice-aware multi-branch decoder network. *IEEE Transactions on Medical Imaging*, v. 39, n. 12, p. 4174–4185, 2020. Citado 3 vezes nas páginas 35, 37 e 49.
- WEBER, E.; VILENSKY, J.; CARMICHAEL, S.; LEE, K. *Netter's Concise Radiologic Anatomy E-Book*. Elsevier Health Sciences, 2014. (Netter Basic Science). ISBN 9780323221474. Disponível em: <<https://books.google.com.br/books?id=o63zAgAAQBAJ>>. Citado 2 vezes nas páginas 13 e 19.
- YAN, X.; TANG, H.; SUN, S.; MA, H.; KONG, D.; XIE, X. After-unet: Axial fusion transformer unet for medical image segmentation. In: *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. Los Alamitos, CA, USA: IEEE Computer Society, 2022. p. 3270–3280. Disponível em: <<https://>>

[//doi.ieeecomputersociety.org/10.1109/WACV51458.2022.00333](https://doi.ieeecomputersociety.org/10.1109/WACV51458.2022.00333)>. Citado 4 vezes nas páginas 14, 36, 37 e 49.

YEUNG, S. *Tutorial 3: Image Segmentation*. 2010. <https://ai.stanford.edu/~syyeung/cvweb/tutorial3.html>. Citado na página 21.

ZHANG, A.; LIPTON, Z. C.; LI, M.; SMOLA, A. J. *Dive into Deep Learning*. 2023. Citado na página 27.

ZHAO, W.; CHEN, H.; LU, Y. W-net: A network structure for automatic segmentation of organs at risk in thorax computed tomography. In: *Proceedings of the 2020 2nd International Conference on Intelligent Medicine and Image Processing*. New York, NY, USA: Association for Computing Machinery, 2020. (IMIP 2020), p. 66–69. ISBN 9781450377799. Disponível em: <<https://doi.org/10.1145/3399637.3399642>>. Citado 4 vezes nas páginas 14, 36, 37 e 49.

Arthur Guilherme Santos Fernandes

Segmentação de Traqueia em Tomografia Computadorizada Utilizando Efficient-Deeplab

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

Trabalho aprovado. São Luís - MA, 28 de Junho de 2024:

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

Prof. Dr. João Otávio Bandeira Diniz
Coorientador
Universidade Federal do Maranhão

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

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

São Luís - MA
2024