

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO ESPÍRITO
SANTO
CURSO SUPERIOR DE BACHARELADO EM SISTEMAS DE INFORMAÇÃO

ANTÔNIO CARLOS DURÃES DA SILVA

**SISTEMA DE APOIO AO DIAGNÓSTICO DE PNEUMONIA BASEADO EM
SEGMENTAÇÃO SEMÂNTICA EM IMAGENS DE RAIO-X USANDO REDES
NEURAIS PROFUNDAS**

Serra
2021

ANTÔNIO CARLOS DURÃES DA SILVA

**SISTEMA DE APOIO AO DIAGNÓSTICO DE PNEUMONIA BASEADO EM
SEGMENTAÇÃO SEMÂNTICA EM IMAGENS DE RAIO-X USANDO REDES
NEURAIS PROFUNDAS**

Trabalho de Conclusão de Curso apresentado à
Coordenadoria do Curso de Bacharelado em Sistemas
de Informação do Instituto Federal de Educação,
Ciência e Tecnologia do Espírito Santo, Campus Serra,
como requisito parcial para a obtenção do título de
Bacharel em Sistemas de Informação.

Orientador: Prof. Dr. Fidelis Zanetti de Castro

Serra
2021

ANTÔNIO CARLOS DURÃES DA SILVA

**SISTEMA DE APOIO AO DIAGNÓSTICO DE PNEUMONIA BASEADO EM
SEGMENTAÇÃO SEMÂNTICA EM IMAGENS DE RAIO-X USANDO REDES
NEURAIS PROFUNDAS**

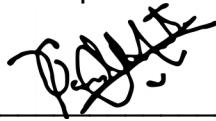
Trabalho de Conclusão de Curso apresentado à
Coordenadoria do Curso de Sistemas de
Informação do Instituto Federal do Espírito Santo,
Campus Serra, como requisito parcial para a
obtenção do título de Bacharel em Sistemas de
Informação.

Aprovado em 08 de abril de 2021

COMISSÃO EXAMINADORA

Prof. Dr. Fidelis Zanetti de Castro
Instituto Federal do Espírito Santo
Campus Serra

Prof. Dr. Daniel Cruz Cavalieri
Instituto Federal do Espírito Santo
Campus Serra



Prof. Dr. David Ernesto Caro Contreras
Universidade Estadual de Campinas

Prof. Dr. Filipe Wall Mutz
Instituto Federal do Espírito Santo
Campus Serra



Emitido em 08/04/2021

FOLHA ANALÍTICA Nº 1/2021 - SER-CCTII (11.02.32.01.08.02.10)

(Nº do Protocolo: NÃO PROTOCOLADO)

(Assinado digitalmente em 04/05/2021 23:28)

DANIEL CRUZ CAVALIERI

PROFESSOR DO ENSINO BASICO TECNICO E TECNOLOGICO

SER-CCECA (11.02.32.01.08.02.03)

Matrícula: 1986870

(Assinado digitalmente em 04/05/2021 20:17)

FIDELIS ZANETTI DE CASTRO

PROFESSOR DO ENSINO BASICO TECNICO E TECNOLOGICO

SER-CCTII (11.02.32.01.08.02.10)

Matrícula: 1527537

(Assinado digitalmente em 05/05/2021 18:25)

FILIPE WALL MUTZ

PROFESSOR DO ENSINO BASICO TECNICO E TECNOLOGICO

SER-CCSI (11.02.32.01.08.02.04)

Matrícula: 2154884

Para verificar a autenticidade deste documento entre em <https://sipac.ifes.edu.br/documentos/> informando seu número: 1, ano: 2021, tipo: **FOLHA ANALÍTICA**, data de emissão: **04/05/2021** e o código de verificação: **d5f992cecd**

*Dedico este trabalho a todos os amigos que me
acompanharam durante esse momento da vida e à minha
família, por ser uma de minhas principais motivações para
trilhar a carreira acadêmica.*

AGRADECIMENTOS

Agradeço ao professor Dr. Fidelis Zanetti de Castro por me orientar, motivar e me auxiliar na obtenção de recursos para execução desse trabalho, além dos professores Daniel Cruz Cavalieri e Filipe Wall Mutz, por contribuírem, respectivamente, com sugestões e a preparação do ambiente para execução dos experimentos computacionais.

Não é possível entrar duas vezes no mesmo rio.
(HERÁCLITO DE ÉFESO)

RESUMO

Neste trabalho de conclusão de curso propomos um sistema de apoio ao diagnóstico de pneumonia baseado em segmentação semântica e classificação em imagens de raio-X por meio das redes neurais profundas U-Net e DeepLabV3+. Aplicamos técnicas de pré-processamento de imagens que envolvem equalização de histograma e combinações lineares de operadores morfológicos, as quais permitiram uma redução das diferenças acentuadas no conjunto de dados e a criação de duas vertentes investigativas. Variações em parâmetros e hiper-parâmetros dos modelos forneceram um amplo cenário de pesquisa que gerou reflexões sobre o uso de redes neurais profundas voltadas para o apoio à tomada de decisão inteligente no contexto do uso da computação cognitiva apoiada na segmentação de imagens biomédicas para o pré-diagnóstico de doenças. Experimentos computacionais realizados tiveram a finalidade de identificar, em milhares de imagens, as regiões pulmonares em que houvesse a manifestação de pneumonia causada por três agentes patológicos, a saber: bactérias, o Sars-CoV-2 (coronavírus) e outras causas virais. Tanto a rede neural U-Net como a rede neural DeepLabV3+, consideradas o estado da arte na tarefa de segmentação semântica, apresentaram resultados significativos e promissores.

Palavras-chave: Segmentação de imagens biomédicas. Imagens de raio-X. Pneumonia. Aprendizado profundo. U-Net. DeepLab.

ABSTRACT

In this undergraduate thesis we propose a support system for the diagnosis of pneumonia disease based on semantic segmentation and classification on X-ray images using U-Net and Deeplabv3+ neural networks. We applied image pre-processing techniques including histogram equalization, as well as linear combinations of morphological operators, which allowed a reduction of the marked differences in the dataset, and the prospection of two investigative ways. Variations in parameters and hyper-parameters of the models provided a broad research scenario that generated reflections about the use of neural networks based on deep learning in the context of the use of segmentation of biomedical images for the pre-diagnosis of diseases. Computational experiments carried out allowed identifying, in thousands of images, the pulmonary regions where there was the manifestation of pneumonia caused by three pathological agents, namely: bacteria, Sars-CoV-2 (corona-virus) and other viral causes. Both the U-Net neural network and the DeepLabV3+ neural network, considered the state of the art in the task of semantic segmentation, presented significant and promising results.

Keywords: Biomedical images segmentation. X-ray images. Pneumonia. Deep learning. U-Net. DeepLab.

LISTA DE FIGURAS

| | |
|---|----|
| Figura 1 – Interesse de pesquisa no <i>Google</i> nos termos “U-Net” e “DeepLab” nos últimos cinco anos. | 14 |
| Figura 2 – Países que mais buscaram sobre as duas redes neurais nos últimos cinco anos. | 14 |
| Figura 3 – Popularidade do termo “image segmentation” nos últimos cinco anos. . | 14 |
| Figura 4 – Países e a busca pelo termo “image segmentation” nos últimos cinco anos. . | 15 |
| Figura 5 – Representação do modelo neuronal de McCulloch e Pitts. | 19 |
| Figura 6 – Neurônio biológico e o modelo matemático de McCulloch e Pitts. | 21 |
| Figura 7 – Arquitetura de rede neural artificial <i>feedforward</i> multicamadas com retropropagação de erro. | 22 |
| Figura 8 – Arquitetura de rede neural recorrente simples. | 22 |
| Figura 9 – Correspondências hierárquicas entre camadas de uma RNC e regiões do cérebro de interesse ao longo da via visual ventral. | 24 |
| Figura 10 – Cubo com <i>stride</i> igual a 2. | 25 |
| Figura 11 – Funções populares usadas em redes neurais profundas. | 26 |
| Figura 12 – Exemplo de <i>poolings</i> máximo e médio. | 27 |
| Figura 13 – Exemplo de classificação, detecção de objeto e segmentação de imagem. . | 28 |
| Figura 14 – Exemplo de segmentação semântica, de instâncias e panóptica. | 29 |
| Figura 15 – Arquitetura da rede <i>U-Net</i> | 30 |
| Figura 16 – Segmentação de órgãos contidos na caixa torácica. | 32 |
| Figura 17 – Segmentação de baço, fígado, pâncreas e rins. | 32 |
| Figura 18 – Segmentação de vasos sanguíneos da retina. | 33 |
| Figura 19 – Exemplo de convolução atrosa para um filtro de 3×3 | 34 |
| Figura 20 – Resultados após iterações do CACTC | 35 |
| Figura 21 – Exemplo de <i>Pooling</i> de Pirâmide Espacial Atrosa. | 35 |
| Figura 22 – Fluxo dos módulos com convolução atrosa (PPEA). | 36 |
| Figura 23 – Arquitetura da rede <i>DeepLab V3+</i> | 37 |
| Figura 24 – Segmentação semântica de ultrassonografias dos rins | 38 |
| Figura 25 – Segmentação semântica de lesões dermatológicas. | 39 |
| Figura 26 – Segmentação semântica de células cervicais. | 40 |
| Figura 27 – Exemplo de tomografias torácicas computadorizadas. | 41 |
| Figura 28 – Exemplo de imagem de um paciente com pneumonia (causada pelo Sars-CoV-2) | 46 |
| Figura 29 – Exemplo de imagem de um paciente com pneumonia viral (não Sars-CoV-2) | 47 |
| Figura 30 – Exemplo de predição assertiva em pulmões com pneumonia (Sars-cov 2) | 51 |
| Figura 31 – Exemplo de predição incorreta para caso de pneumonia viral | 51 |
| Figura 32 – Melhores cenários para etapa de treino | 57 |
| Figura 33 – Melhores cenários para etapa de validação | 57 |

Figura 34 – Melhores cenários para etapa de teste 58

LISTA DE SIGLAS

| | |
|-------|--|
| AM | Aprendizado de Máquina |
| AP | Aprendizado Profundo |
| CACTC | Campo Aleatório Condisional Totalmente Conectado |
| CPU | Central Processing Unit |
| GPU | Graphics Processing Unit |
| IA | Inteligência Artificial |
| ISBI | International Symposium on Biomedical Imaging |
| MIT | Massachusetts Institute of Technology |
| PPEA | Pooling de Pirâmide Espacial Atrosa |
| RDM | Matriz de Dissimilaridade Representacional |
| RNA | Rede Neural Artificial |
| RNC | Rede Neural Convolucional |
| RNCP | Redes Neurais Convolucionais Profundas |
| RNP | Rede Neural Profunda |
| RNR | Rede Neural Recorrente |
| ReLU | Rectified Linear Unit |

SUMÁRIO

| | | |
|--------------|---|-----------|
| 1 | INTRODUÇÃO | 11 |
| 1.1 | Objetivos | 15 |
| 1.2 | Metodologia do Trabalho | 16 |
| 1.3 | Estrutura do trabalho | 17 |
| 2 | REFERENCIAL TEÓRICO | 18 |
| 2.1 | Processo de aprendizagem | 18 |
| 2.2 | Modelo de neurônio artificial de McCulloch e Pitts | 19 |
| 2.3 | Redes neurais: Biológicas e artificiais | 20 |
| 2.3.1 | Redes neurais profundas | 21 |
| 2.4 | Aprendizado profundo | 23 |
| 2.5 | Redes neurais convolucionais | 23 |
| 2.6 | Classificação, detecção e segmentação de objetos | 27 |
| 2.6.1 | Tipos de segmentação de imagens | 28 |
| 2.7 | Redes neurais do tipo encoder-decoder | 29 |
| 2.7.1 | A Rede neural <i>U-Net</i> | 29 |
| 2.7.1.1 | Aplicações da rede neural <i>U-Net</i> em segmentação de imagens biomédicas . | 31 |
| 2.7.2 | A rede neural <i>DeepLab</i> | 33 |
| 2.7.2.1 | Versões e aprimoramentos da rede neural <i>Deeplab</i> | 35 |
| 2.7.2.2 | Aplicações da rede neural <i>Deeplab</i> em segmentação de imagens biomédicas | 38 |
| 2.8 | A Pneumonia e a sua dimensão | 40 |
| 3 | MATERIAIS E MÉTODOS | 43 |
| 3.1 | Métricas para avaliação dos modelos | 43 |
| 3.2 | Conjunto de dados | 44 |
| 3.3 | Pré-processamento dos dados | 45 |
| 3.4 | Definição de parâmetros e hiperparâmetros | 47 |
| 3.4.1 | Recursos computacionais | 49 |
| 4 | EXPERIMENTOS, RESULTADOS E DISCUSSÃO | 50 |
| 4.1 | Resultados: Etapa de treino | 52 |
| 4.2 | Resultados: Etapa de validação | 53 |
| 4.3 | Resultados: Etapa de teste | 55 |
| 5 | CONSIDERAÇÕES FINAIS | 59 |
| 5.1 | Trabalhos futuros | 60 |
| | REFERÊNCIAS | 61 |

1 INTRODUÇÃO

A ideia de instruir uma máquina com um conjunto de dados de entrada-saída a fim de que ela seja capaz de identificar padrões em dados semelhantes não é nova. Um dos programas de computador pioneiros no uso de sistemas de aprendizado de máquina foi desenvolvido em 1959, por Arthur Lee Samuel¹, e sua proposta era aplicar tal método para “ensinar” os computadores a jogarem uma versão digital do jogo de tabuleiro damas (SAMUEL, 1959).

Um dos paradigmas mais promissores para a tarefa de “ensinar” um computador é o *Aprendizado Profundo* (AP). Seu surgimento foi inspirado no funcionamento do sistema visual humano e suas estruturas neuro-sinápticas (CHEN; LIN, 2014). Tal abordagem de aprendizado ganhou notoriedade no meio acadêmico devido à sua acurácia em tarefas envolvendo visão computacional, bem como em reconhecimento de fala (bastante utilizado por assistentes virtuais como a Siri, da Apple, ou a Alexa, da Amazon), entre outras (DENG; YU, 2014).

Diversas técnicas compõem as estratégias metodológicas que envolvem o aprendizado profundo: há propostas voltadas para o aprendizado supervisionado, o não supervisionado, o semi-supervisionado e o por reforço. O aprendizado supervisionado é aquele que ocorre com o auxílio de um agente externo, denominado professor, o qual, dentre várias possibilidades de atuação, pode realizar anotações em dados ou interferir no modo como a rede é treinada. Já o aprendizado não supervisionado trabalha sem auxílio de um agente externo e não faz uso de anotações em imagens, por exemplo. Nesse sentido, todas as informações adquiridas pela máquina são exclusivas de cada entrada, ou seja, não há armazenamento de informações de entradas anteriores ou etapas de treinamento (ZHOU, 2017). Os métodos semi-supervisionados funcionam com etapas que mesclam as duas abordagens supramencionadas, e o último trabalha com a ideia de aprendizado baseado em “tentativa e erro”, com recompensa após cada ação assertiva do sistema, ou punição, para resultados que a distanciam de um objetivo final estabelecido *a priori* (KAELBLING; LITTMAN; MOORE, 1996).

Com os recentes e contínuos avanços em AP, surgiram aplicações nas mais diversas áreas, como biomedicina, farmacologia, guiagem autônoma, processamento de linguagem natural, imagem e vídeo. Aliper e demais autores, por exemplo, descrevem como esse tipo de abordagem foi utilizado para testar propriedades farmacológicas de diversos medicamentos em representações de genes, que por sua vez, foram submetidos a condições e sistemas biológicos distintos (ALIPER et al., 2016). Já Huval e outros pesquisadores, exploraram a

¹ Nascido em 1901, em Kansas, nos Estados Unidos, Arthur Lee Samuel foi físico, instrutor no Instituto de Tecnologia de Massachusetts (MIT) e colaborador da Bell Labs. É reconhecido como um pioneiro na área de IA e iniciou na área com seus trabalhos sobre aprendizado de máquina utilizando o jogo de damas e estratégias de jogadores profissionais. Disponível em: <<https://cs.stanford.edu/memoriam/professor-arthur-samuel>>. Acesso em: 8 jul. 2020.

capacidade de se utilizar o AP para identificação, em tempo real, de placas de veículos em rodovias, e sugerem sua aplicação no desenvolvimento de um sistema para guiagem autônoma (HUVAL et al., 2015).

No survey “*Big Data Deep Learning: Challenges and Perspective*”, Chen e Lin tratam o AP como um paradigma que desempenha um papel crucial para análise preditiva de informação na era da explosão de dados, além de indicarem o volume de dados e o aumento de poder computacional como fatores para o aperfeiçoamento das técnicas inerentes a esse paradigma de aprendizagem (CHEN; LIN, 2014). A chamada era do *Big Data* também alimenta usos ambiciosos do AP para o reconhecimento de padrão nas pesquisas, compras e interesses dos consumidores na internet, a fim de tornar campanhas publicitárias, anúncios e recomendações mecanismos mais assertivos e atrativos para seus potenciais clientes (ELKAHKY; SONG; HE, 2015).

Embora tenha atingido êxito nas tarefas citadas anteriormente, pelo menos dois fatores limitavam a expansão de aplicações das redes neurais baseadas em AP, em seu início: A quantidade pequena de dados disponíveis para treinamento da rede e a limitação no número de camadas das próprias redes (RONNEBERGER; FISCHER; BROX, 2015). Graças ao rompimento dessas duas barreiras, as redes profundas passaram também a serem empregadas com sucesso na tarefa de classificação de imagens, isto é, tornaram-se capazes de calcular a probabilidade de uma imagem pertencer a uma determinada classe e estimar similaridades entre objetos apresentados sob diferentes formatos, como em imagem, áudio e vídeo, por exemplo.

Os trabalhos de Ciresan e outros pesquisadores auxiliaram na expansão do horizonte do AP para além da classificação de imagens, possibilitando a determinação da delimitação e contorno da área da imagem que continha objeto a ser classificado (CIRESAN et al., 2012). Em contrapartida, a maneira como essa inovação foi implementada trazia algumas desvantagens. A primeira delas era a utilização de uma janela deslizante para predizer o rótulo de cada pixel, e a segunda era utilizar uma região adjacente a cada *pixel* como entrada para classificação, o que, como Ronneberger, Fischer e Brox observam, demanda processamento desnecessário devido a redundância e sobreposição de *pixels* adjacentes, ou seja, um mesmo conjunto de dados pode ser processado diversas vezes sem a necessidade disto (RONNEBERGER; FISCHER; BROX, 2015).

Após uma análise dos problemas mencionados, Ronneberger e demais pesquisadores, desenvolveram um modelo de rede neural com arquitetura própria, a rede neural *U-Net*, criada com objetivo de ser *simples* - do ponto de vista arquitetural e da sequência de operações utilizadas -, e *eficiente* - ao lidar com um número menor de imagens para treino do que os modelos anteriores que propunham realizar tarefas semelhantes (RONNEBERGER; FISCHER; BROX, 2015). Em essência, o uso da *U-Net* se dá majoritariamente em um

problema fundamental na área de Visão Computacional: a tarefa de segmentação de imagens, isto é, a classificação de um ou mais objetos e a identificação de suas delimitações na imagem de entrada (GWET et al., 2018).

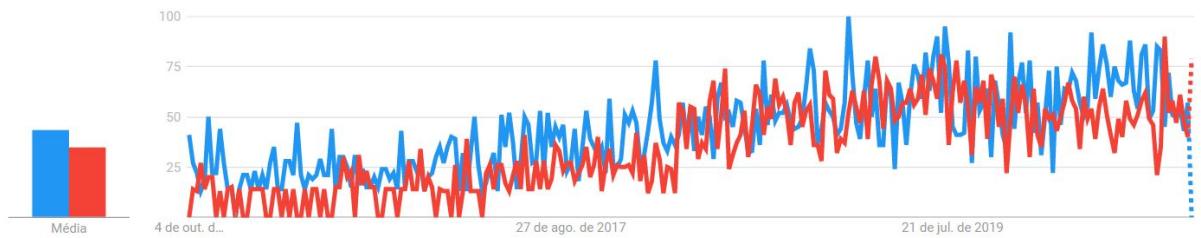
Os idealizadores da *U-Net* a propuseram para extração e identificação de padrões em imagens biomédicas (RONNEBERGER; FISCHER; BROX, 2015). Essa linha investigativa é fortalecida pelo trabalho de He e seus colegas de pesquisa, que fizeram uso da rede para segmentar a região da próstata em imagens obtidas por tomografia computadorizada (HE et al., 2020), e o de Akkus e demais pesquisadores, que não apenas fizeram uso da rede para mapeamento de áreas cerebrais (na possibilidade de detecção de anomalias como tumores) como compararam o desempenho da *U-Net* com várias outras redes neurais, o que a incluiu na lista das dez redes neurais mais performáticas para a tarefa (AKKUS et al., 2017).

Apresentando um campo mais abrangente de aplicações, a *DeepLab* traz como diferencial o uso de recursos (detalhados em outras seções deste trabalho) que tornam o processo de segmentação mais rápido, além de manter desempenho satisfatório. O desempenho da *DeepLab* em segmentação de imagens foi responsável por destacar a rede no desafio de segmentação de imagens PASCAL VOC 2012 (CHEN et al., 2014).

Um grupo de pesquisadores empregaram a rede neural *DeepLab* para segmentar imagens de alimentos que constituíam uma refeição a fim de avaliar a qualidade dos alimentos e estimar a quantidade de calorias de uma dada refeição (MYERS et al., 2015). A rede neural também foi utilizada para auxiliar na identificação de pessoas e veículos sob o risco de incêndios e alagamentos (GIANNAKERIS et al., 2018), além de ter sido aplicada na segmentação de imagens de pólipos colorretais (XIAO; CHANG; LIU, 2018), um pequeno grupo de células que podem se tornar cancerígenas.

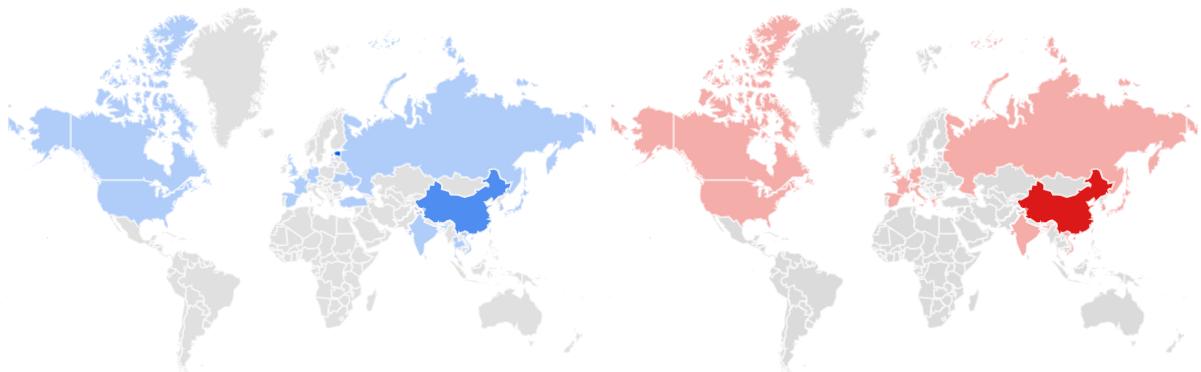
Para ter ciência da popularidade das duas redes, o gráfico apresentado na Figura 1 mostra o interesse crescente dos usuários do buscador *Google* em pesquisar pelos termos “DeepLab” e “U-Net” nos últimos cinco anos, ou seja, desde o ano em que a primeira versão de ambas as redes já haviam sido publicadas na literatura. É importante notar que, embora o termo “DeepLab” tenha sido menos buscado, o grupo de países que o pesquisaram é composto por quase os mesmos que buscaram sobre a *U-Net*, como mostra o *heatmap* apresentado na Figura 2.

Figura 1 – Interesse nos termos de busca “U-Net” (azul) e “DeepLab” (vermelho) nos últimos cinco anos.



Fonte: *Google Trends* (<www.google.com/trends>). Acesso em: 30 set. 2020.

Figura 2 – Países que mais buscaram sobre as duas redes neurais nos últimos cinco anos.



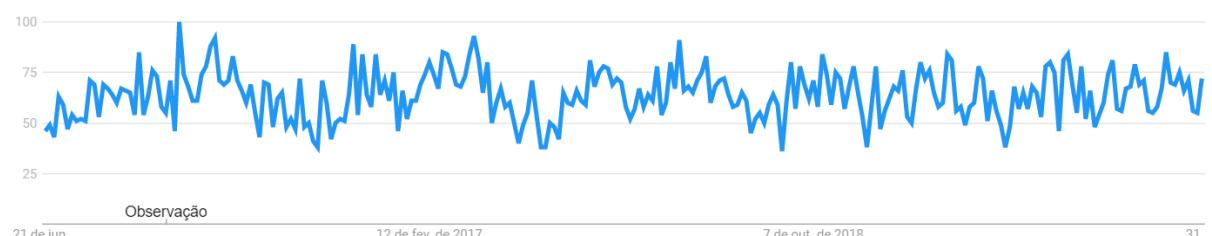
(a) Países que buscaram pela *U-Net*.

(b) Países que buscaram pela *DeepLab*.

Fonte: *Google Trends* (<www.google.com/trends>). Acesso em: 30 set. 2020.

Por sua vez, a segmentação de imagens também é um tema com alta taxa de pesquisa no *Google* e que manteve seu interesse constante nos últimos anos, conforme ilustrado na Figura 3. A segmentação de imagens destaca-se como uma área de pesquisa bastante frequente em buscas na China, Coreia do Sul, Índia, Paquistão e Singapura, como pode ser visto no gráfico do tipo *heatmap* da Figura 4. Destacamos que para obtenção do gráfico buscamos o nível de interesse pelo termo “image segmentation”.

Figura 3 – Popularidade do termo “image segmentation” nos últimos cinco anos.



Fonte: *Google Trends* (<www.google.com/trends>). Acesso em: 26 jun. 2020.

Figura 4 – Países e a busca pelo termo “image segmentation” nos últimos cinco anos.



Fonte: *Google Trends* (<www.google.com/trends>). Acesso em: 26 jun. 2020.

Diante do amplo espectro de aplicações das redes neurais *U-Net* e *DeepLab* delimitamos este trabalho de conclusão de curso ao contexto da *Computação Cognitiva* voltada para o pré-diagnóstico da doença Pneumonia e a identificação de seus agentes causadores. Nesse intuito, propomos um sistema que utiliza como base as redes neurais *U-Net* e *DeepLab* para realizar segmentação semântica em imagens de raio-X.

A seguir, apresentamos os objetivos geral e específicos inerentes a este trabalho.

1.1 Objetivos

Este trabalho tem como objetivo geral:

- Desenvolver um sistema de apoio ao diagnóstico de pneumonia que possibilite a identificação de agentes causadores da doença por meio da segmentação semântica em imagens de raio-X usando as redes neurais *U-Net* e *DeepLab*.

Para atingir o propósito acima, faz-se necessário contemplar os seguintes objetivos específicos:

- Organizar e sanitizar uma base de dados composta por imagens de raio-X de pneumonia provocada por três patógenos: bactérias, vírus gerais e especificamente o vírus Sars-CoV-2, causador da COVID-19;
- Criar, a partir da base de dados original, outras duas bases de dados utilizando as seguintes técnicas de pré-processamento: aplicação de equalização de histograma e uso de operadores morfológicos;
- Adaptar implementações computacionais das redes neurais *U-Net* e *DeepLab* ao contexto do nosso problema de investigação;

- Realizar análises e discussões sobre a viabilidade da solução proposta levando-se em conta diferentes tipos de ajustes paramétricos e hiperparamétricos nos modelos de redes neurais utilizados;
- Disponibilizar para a comunidade as bases de dados previamente organizadas bem como os algoritmos adaptados para este trabalho por meio da plataforma de gerenciamento de projetos *Github*.

1.2 Metodologia do Trabalho

A pesquisa a que nos propomos neste projeto visa, por um lado, realizar um estudo sistemático das redes neurais *U-Net* e *Deeplab* e a adaptação de implementações computacionais voltadas para a tarefa de segmentação semântica de imagens biomédicas, especificamente imagens de raio-X. Por outro lado, visa a revisão e a implementação de técnicas de pré-processamento de imagens envolvendo equalização de histograma e a aplicação de composições de operadores morfológicos específicos. Além disso, envolve a integração de diversos tipos de estruturas que permitem desenvolver um sistema que recebe como entrada uma imagem de raio-X na qual os pulmões estão sujeitos a pneumonia provocada por algum dos patógenos mencionados anteriormente, e como saída uma imagem de segmentação dos pulmões e a identificação do patógeno.

A pesquisa caracteriza-se como teórica-aplicada no sentido de que apresentamos um desenvolvimento sobre o funcionamento das redes neurais e um amplo rol de aplicações, além de adaptarmos implementações computacionais, aplicarmos técnicas de pré-processamento em imagens de raio-X e estudarmos os papéis de diferentes parâmetros e hiperparâmetros dos modelos.

A análise de dados é quantitativa e qualitativa, pois apresenta resultados numéricos representados em tabelas e gráficos (valores das diferentes métricas escolhidas para avaliar os resultados) e uma análise do que foi obtido e observado a partir dos resultados encontrados, bem como em função de todo aprendizado construído ao longo do processo de desenvolvimento deste trabalho.

A seguir apresentamos a lista de atividades que compõe o nosso percurso metodológico com fins de atingir os objetivos geral e específicos supramencionados.

- Realizar uma revisão de literatura sobre a rede neural *U-Net*;
- Realizar uma revisão de literatura sobre a rede neural *DeepLabV3+*;
- Selecionar um conjunto de imagens de radiografias torácicas de pessoas com pneumonia;

- Organizar e sanitizar as bases de dados a serem utilizadas nos experimentos computacionais via técnicas de pré-processamento de imagens;
- Adaptar uma implementação computacional da rede neural *U-Net* ao nosso caso de estudo;
- Adaptar uma implementação computacional da rede neural *DeepLabV3+* ao nosso caso de estudo;
- Analisar o desempenho das redes neurais propostas na tarefa de segmentação semântica de imagens de radiografias torácicas de pessoas com pneumonia causada por bactérias, vírus e pelo vírus específico Sars-CoV-2, causador da COVID-19.

1.3 Estrutura do trabalho

A organização do restante deste trabalho segue a seguinte estrutura:

- Capítulo 2 – Fundamentação teórica: Seção em que são apresentados os conceitos fundamentais para compreensão do funcionamento de uma rede neural convolucional e das tarefas que compõem a área de Visão Computacional;
- Capítulo 3 – Experimentos: Nesta seção é apresentada a proposta de aplicação para segmentação semântica de imagens com as redes *U-Net* e *DeepLabV3+*;
- Capítulo 4 – Resultados: Seção em que são apresentados e analisados os resultados dos experimentos computacionais descritos no Capítulo 3;
- Capítulo 5 – Conclusão: Seção em que oportunidades de aperfeiçoamento e margens para trabalhos futuros são abordadas.

2 REFERENCIAL TEÓRICO

Neste capítulo são apresentadas ideias essenciais para compreensão do funcionamento de uma rede neural artificial (RNA), suas arquiteturas, estruturas, operações e mais importante, suas formas de adaptarem-se e se reorganizarem a cada nova experiência a partir de estímulos de entrada. Também são abordados problemas de pesquisa clássicos que integram a área de Visão Computacional, bem como duas das redes neurais utilizadas na literatura para performar a tarefa de segmentação semântica. Além disso, são apresentadas, de forma breve, características da pneumonia e os impactos causados pela doença.

2.1 Processo de aprendizagem

Thomas J. Shuell, no paper *Cognitive Conceptions of Learning*, em que disserta como o processo de aprendizado é pensado sob a ótica da Psicologia Cognitiva, identificou a presença de três componentes nas definições de aprendizagem estudadas por ele (SHUELL, 1986, p. 412): Mudança no comportamento ou capacidade do indivíduo executar uma tarefa; Convicção de que essa mudança deve ser resultado de algum tipo de prática ou experiência; Convicção de que a mudança deve ser duradoura.

Anderson, em tradução livre, diz que “[...] Aprendizagem é qualquer processo que modifica um sistema de modo a melhorar, de forma mais ou menos irreversível, seu desempenho subsequente da mesma tarefa ou de tarefas retiradas da mesma população” (ANDERSON, 2013, p. 367). Diante do exposto, podemos inferir a partir de ambos os trabalhos que o aprendizado é uma mudança duradoura, adquirida por meio de experiências e que beneficia o executor de uma tarefa a realizar trabalhos de mesma classe.

No domínio da Neurociência, o aprendizado pode ser visto como a indução de alterações de forças sinápticas entre neurônios motivada por estímulos conscientes e inconscientes e sua memorização por períodos curtos ou duradouros é dependente da estabilidade dessas alterações (FANSELOW; POULOS, 2005, p. 208). Essa capacidade de adaptação chamamos de neuroplasticidade cerebral, que pode produzir tanto impactos positivos quanto negativos nas atividades cerebrais (CRAMER et al., 2011).

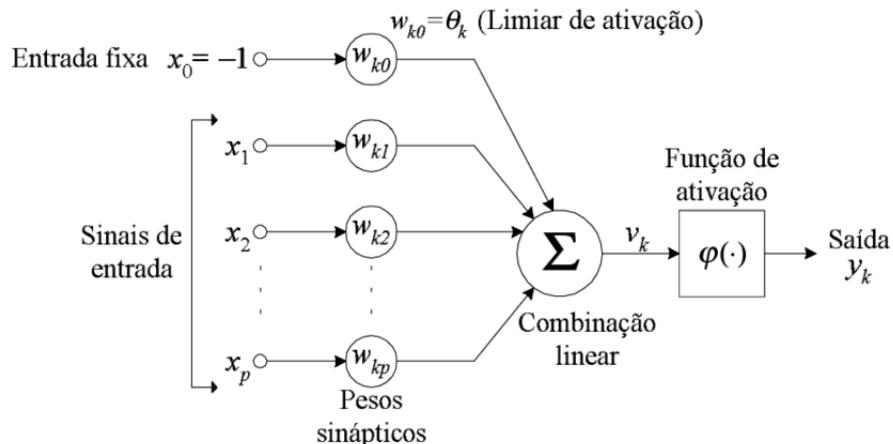
Tendo em vista que a noção de aprendizado permite certo espectro de significados, variando de acordo com a área de conhecimento em que é submetida, trazemos uma definição mais específica de *Thomas Haykin*, no contexto das RNAs, ao dizer que “[...] aprendizagem é um processo pelo qual os parâmetros livres de uma rede neural são adaptados por meio de uma prática de estimulação pelo ambiente no qual a rede está inserida” (HAYKIN, 2007, p. 75).

2.2 Modelo de neurônio artificial de McCulloch e Pitts

Fruto da investigação do funcionamento das atividades nervosas, o modelo matemático de um neurônio proposto pelo lógico e cientista cognitivo Walter Pitts e pelo neurocientista Warren Sturgis McCulloch (MCCULLOCH; PITTS, 1943), em 1943, foi um marco para o início da modelagem de estruturas cerebrais em outras áreas de conhecimento científico fora da Biologia.

Um neurônio biológico realiza conexões com outros a partir de impulsos elétricos e químicos, chamados sinapses (PEREDA, 2014). Cada uma dessas conexões pode ser representada em uma RNA como um valor numérico, um peso numérico, assim como os sinais de entrada. Quanto maiores forem esses valores, mais intenso serão os sinais transmitidos. Após essa transmissão, todos os valores são somados e o resultado é submetido a uma função de ativação, que determinará se o valor recebido é suficientemente alto para disparar um impulso excitatório ou inibitório, representados. Essa sequência de passos pode ser vista na representação gráfica do modelo proposto por McCulloch e Pitts, na Figura 5.

Figura 5 – Representação do modelo neuronal de McCulloch e Pitts.



Fonte: Adaptada de Fernandes (1999)

Posterior à introdução do modelo de McCulloch e Pitts, diversos trabalhos sobre outros modelos de neurônios foram realizados durante as décadas do século XX, contribuindo de forma teórica e prática para o desenvolvimento e o aperfeiçoamento das redes neurais artificiais. A fim de atingir os objetivos propostos para este trabalho, optou-se por saltar tais décadas, com foco na abordagem de alguns modelos de redes neurais profundas, assim, convida-se os leitores interessados na evolução histórica dos modelos neuronais e de redes neurais a consultarem os trabalhos publicados por Haykin (HAYKIN, 2007).

2.3 Redes neurais: Biológicas e artificiais

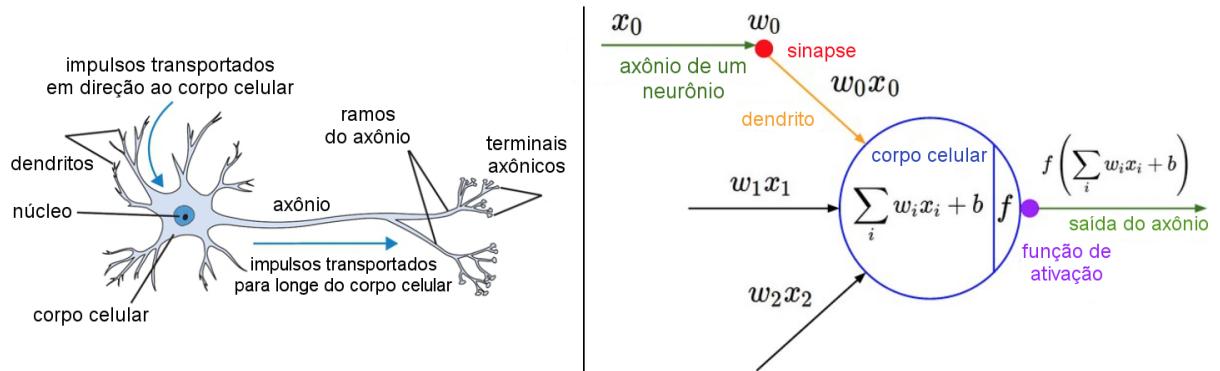
Haykin compara o cérebro biológico a uma unidade não linear, de alta complexidade e paralelismo para processamento de dados (HAYKIN, 2007). O êxito no reconhecimento de padrões realizado pelo cérebro biológico o autor atribui à sua capacidade de aprendizado, isto é, à constante construção e reformulação de normas para performar as tarefas já abordadas.

Para Haykin, uma RNA pode ser compreendida como “um processador maciçamente e paralelamente distribuído, constituído de unidades de processamento simples que têm a propensão natural de armazenar conhecimento experimental e torná-lo disponível para o uso” (HAYKIN, 2007, p. 28). Esse conhecimento experimental seria resultado do processo de aprendizado com uma gama de dados distintos e produziria como impacto não só ajustes incrementais nos pesos sinápticos, mas também o desaparecimento de antigos neurônios e o desaparecimento ou fortalecimento de conexões sinápticas, como ocorre no cérebro humano.

Wang e outros pesquisadores, estabelecem um comparativo entre as estruturas de um neurônio biológico com os componentes de um neurônio artificial implementado sob um resistor (WANG et al., 2018). A Figura 6 exemplifica a relação proposta pelos autores, que compararam:

- O peso da conexão com a sinapse, responsável por conectar os dendritos de um neurônio a outro;
- A entrada ponderada com os dendritos, responsáveis por conectar a entrada ao corpo celular (também chamado de soma);
- O corpo celular (ou soma) com o somatório das entradas ponderadas acrescidas de um valor chamado viés (ou *bias*);
- Os terminais axônicos com a saída obtida a partir da aplicação de uma função de ativação sob o valor resultante da etapa anterior.

Figura 6 – Neurônio biológico e o modelo matemático de McCulloch e Pitts.



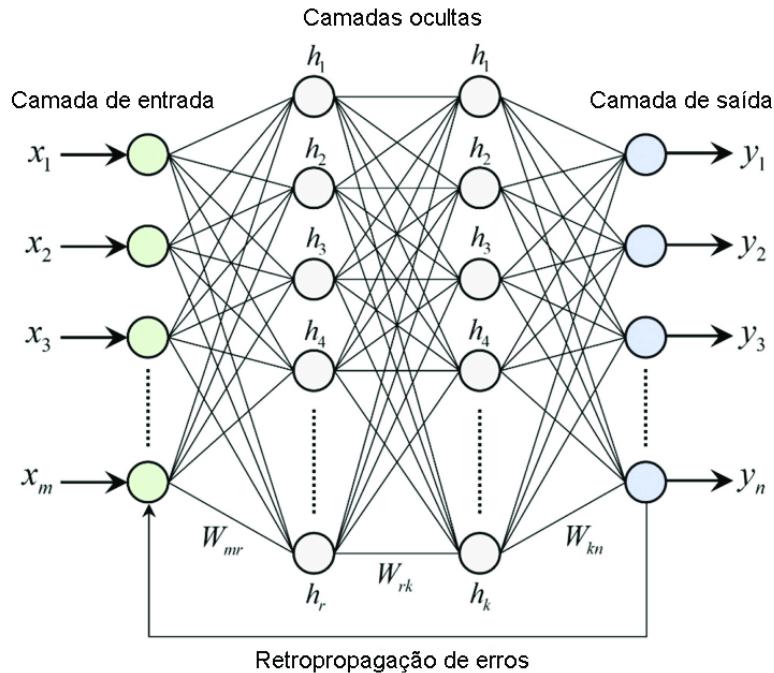
Fonte: Adaptada de Franchini, Burgio e Zanni (2019).

2.3.1 Redes neurais profundas

As redes neurais profundas (RNP) dividem-se em duas categorias principais: Redes neurais recorrentes (RNRs) e redes neurais *feedforward* (YIN et al., 2017). O que distingue os dois tipos de redes é a maneira como se dá o fluxo de dados.

Redes neurais artificiais em que as informações fluem apenas em um sentido, comumente da camada de entrada para as camadas ocultas, por fim, para a camada de saída, são chamadas redes *feedforward*, por possuírem a alimentação de suas camadas sempre fluindo para o fim da rede (SIDHU; SINGH; SACHDEV, 1995, p. 705), sem formar ciclos, como pode ser visto na arquitetura da rede artificial ilustrada na Figura 7.

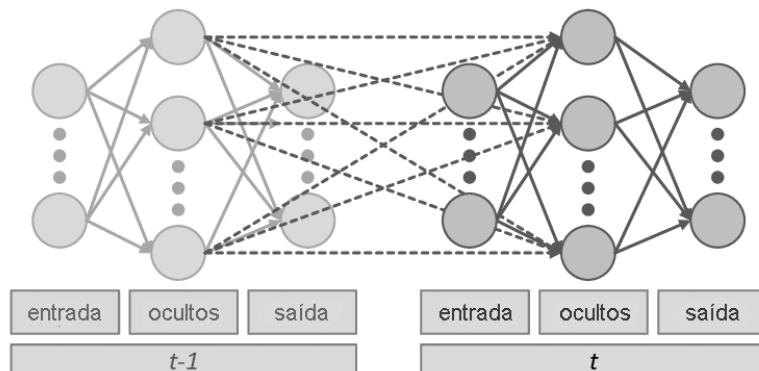
Figura 7 – Arquitetura de rede neural artificial *feedforward* multicamadas com retropropagação de erro.



Fonte: Adaptada de Fernández-Cabán, Masters e Phillips (2018).

Enquanto as redes *feedforward* revertem o fluxo de dados apenas para correção dos pesos de suas conexões, as redes recorrentes apresentam funcionamento dependente de estruturas que direcionam os dados processados tanto para a saída quanto para a própria rede novamente. Essa categoria de redes neurais possui a capacidade de armazenar estados internos, o que lhe propicia a possibilidade de relacionar conjuntos temporalmente próximos de entradas e dessa forma, obter desempenho significativo em tarefas que dependam de processamento de dados sequenciais, como processamento de fala, por exemplo (HOCHREITER; SCHMIDHUBER, 1997, p. 1). A Figura 8, ilustra um exemplo de uma RNR típica.

Figura 8 – Arquitetura de rede neural recorrente simples.



Fonte: Adaptada de Lingenfelser et al. (2016).

2.4 Aprendizado profundo

O Aprendizado Profundo (AP) é o paradigma de aprendizado de máquina (AM) que trabalha sob a composição de camadas de processamento dos dados de entrada. Tal funcionamento permitiu que essa abordagem obtivesse excelência em tarefas que operam com dados sequenciais como processamento de texto e áudio (LECUN; BENGIO; HINTON, 2015). Lecun, Bengio, Hinton (2015) ainda relatam a importância do volume de dados para aperfeiçoamento do aprendizado por meio da correção de seus parâmetros e propagação desses ajustes por todas as camadas da rede.

Deve-se salientar que a precisão desse aprendizado e o custo de processamento aumentam junto ao número de camadas da rede, o que, hoje, não é um empecilho de grande impacto devido recursos como *multithreading*, otimizações em CPUs e GPUs, além da possibilidade de se utilizar *clusters* de computadores. Aliado ao progresso do poder computacional, o significativo volume de dados gerados na atualidade, fazem desse paradigma uma opção de excelência para o aprendizado de máquina.

2.5 Redes neurais convolucionais

Descritas em 1998, com a finalidade de reconhecer caracteres em imagens (LECUN et al., 1998), a origem do nome dessas redes advém do termo “convolução”, originário da matemática e que faz referência a uma função resultado do produto de outras duas funções (HIRSCHMAN; WIDDER, 2012, p. 4). Neste contexto, a convolução faz referência à transformação aplicada sob uma imagem de entrada por meio de uma função *kernel* (também chamada de filtro).

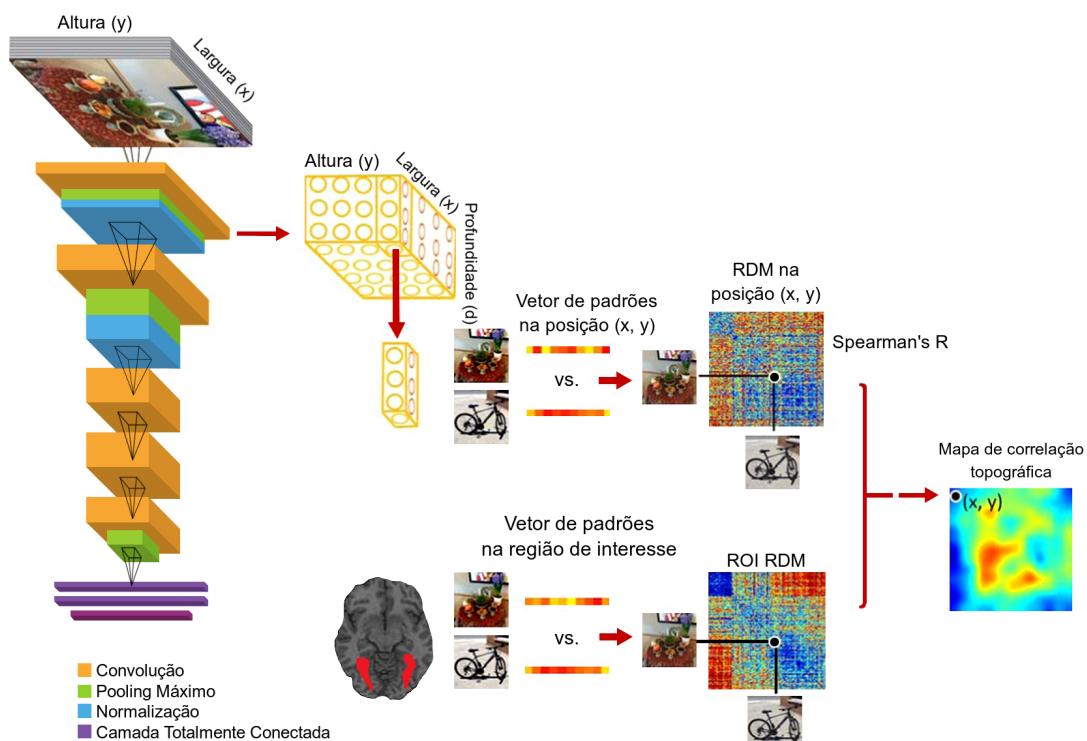
A inspiração para o projeto das Redes Neurais Convolucionais teve origem na observação biológica, a saber o estudo do funcionamento do sistema visual animal. Hubel e Wiesel, após investigarem o córtex visual primário de um gato, observaram uma variedade de tipos de células, todas podendo ser agrupadas em dois tipos principais, simples e complexas, de acordo com sua reação ao estímulo de fontes luminosas de formatos diversos (HUBEL; WIESEL, 1962). O primeiro grupo de células apresenta uma resposta de intensidade dependente da orientação espacial do objeto luminoso; já o segundo grupo, apresenta maior invariância de sua resposta em relação à orientação do objeto.

Os estudos de Hubel, Wiesel e outros neurocientistas, impulsionou a descoberta dos distintos papéis atribuídos às diferentes regiões do sistema visual. Estimulado por esse campo do conhecimento, Lecun e colaboradores projetaram e interligaram sequencialmente as camadas que compõem sua rede neural convolucional no sentido de cada uma desempenhar a extração de características de níveis distintos.

A Figura 9 apresenta uma forma de gerar mapas de correlação topográfica entre a matriz

de dissimilaridade representacional (RDM) construída por uma rede neural convolucional e a construída pelo cérebro humano (MOHSENZADEH et al., 2020, p. 4). A construção da RDM pela rede convolucional ocorre por meio de um vetor de padrões extraídos de cada imagem de entrada, processo que imita o que é realizado pelo cérebro. Contudo, o sistema visual humano captura regiões interesse (ROI) em vez de receberem uma posição específica com valores de X e Y (WANG et al., 2014). Para calcular a similaridade entre as matrizes, utiliza-se o coeficiente de correlação de Spearman ou ρ de Spearman.

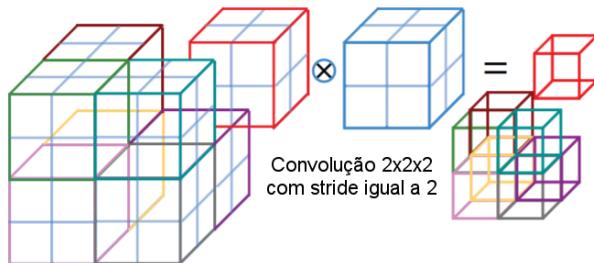
Figura 9 – Correspondências hierárquicas entre camadas de uma RNC e regiões do cérebro de interesse ao longo da via visual ventral.



Fonte: Adaptada de Mohsenzadeh et al. (2020).

Tendo em vista que as transformações realizadas por uma RNC ocorrem em submatrizes de pontos da imagem original, o *stride* representa o salto em *pixels* horizontais e verticais que é dado entre o bloco atual e o próximo a ser processado. A Figura 10 exemplifica o caso de um *stride* com valor de dois *pixels*, neste caso, em um cubo.

Figura 10 – Cubo com *stride* igual a 2.



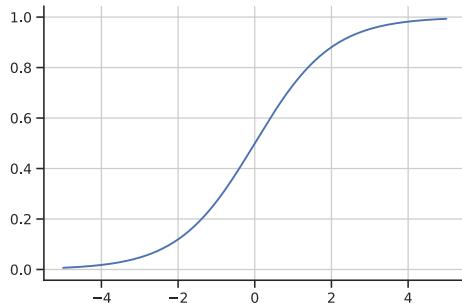
Fonte: Adaptada de Milletari, Navab e Ahmadi (2016).

Ainda que uma rede neural profunda possa conter dezenas de camadas de neurônios, se o funcionamento da rede dependesse apenas de operações lógicas básicas, com emissão de valores binários, seu uso para segmentação, classificação de imagens ou tarefas relacionadas, com foco em acurácia, seu uso seria inviável. Uma rede que possuísse todos seus neurônios capazes apenas de emitir saídas binárias, do tipo 0 ou 1, sem qualquer variação entre esses dois valores reais, estaria fadada à solução de problemas de cunho linear (GOODFELLOW; BENGIO; COURVILLE, 2016, p. 167-170). Para contornar esse dilema, funções de ativação são aplicadas nos potenciais de ativação com a finalidade de adicionar comportamentos não lineares às redes neurais artificiais. Há várias funções que cumprem esse propósito, dessa forma, alguns critérios são estabelecidos para selecionar a mais adequada a um modelo, dentre eles estão: a acurácia obtida e o custo computacional para execução da função.

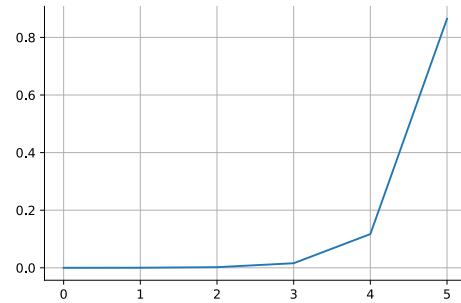
A família “sigmoide” de funções de ativação, as quais apresentam êxito em problemas que envolvem classificações binárias, se destacam ao serem aplicadas tanto em redes neurais rasas (redes com poucas camadas) (NWANKPA et al., 2018, p. 5-6) como em redes neurais recorrentes (GODIN et al., 2018, p. 1). A função *softmax*, por sua vez, é utilizada em problemas que envolvem multi-classificação. A soma dos seus valores de saída resulta no valor 1, diferentemente das funções anteriores. Para além das funções citadas, há também a tangente hiperbólica, a qual pode ser utilizada quando o valor de um neurônio inativo está próximo do valor -1 ao invés de 0 (como é caso da sigmoide), e a função *Rectified Linear Units* (ReLU), conhecida por sua simplicidade, velocidade de execução e popularidade em problemas de classificação de imagens. A função ReLU produz saída zero em metade do seu domínio e é retorna a identidade na outra metade, que possui derivada igual a 1. Como consequência, a ReLU evita a saturação em grande parte dos neurônios de uma rede, ao contrário de outras funções. A Figura 11 apresenta uma ilustração gráfica das funções mencionadas neste parágrafo.

Figura 11 – Funções populares usadas em redes neurais profundas.

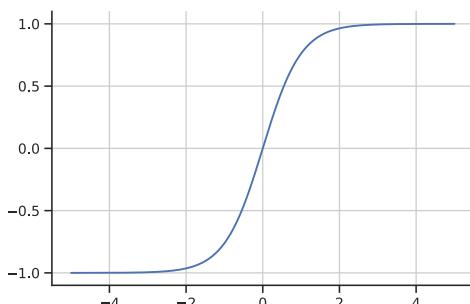
(a) Gráfico: Sigmóide.



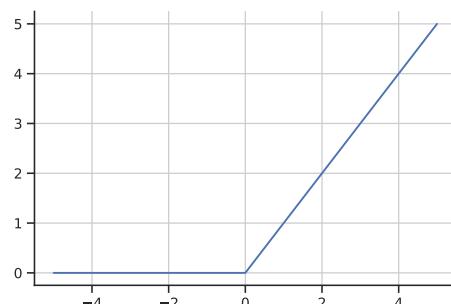
(b) Gráfico: Softmax.



(c) Gráfico: Tangente hiperbólica.



(d) Gráfico: ReLU.

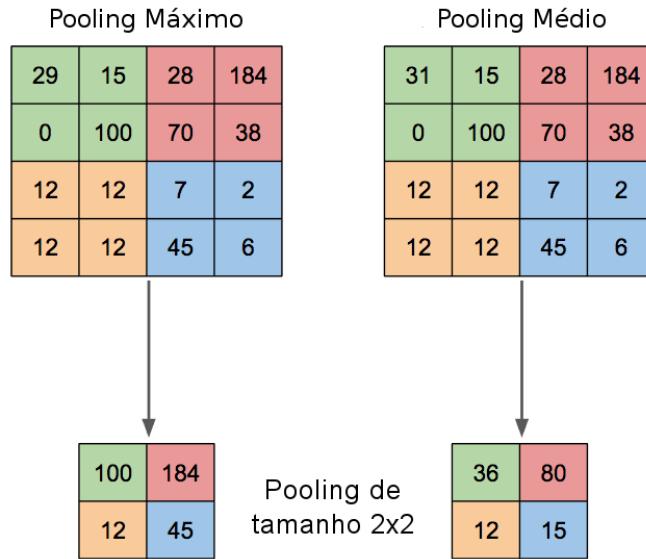


Fonte: Elaborada pelo autor (2021).

Cada convolução resulta em uma coleção de características, isto é, um conjunto de valores que identificam parte da entrada processada. Em uma coleção de imagens contendo pessoas, por exemplo, cada entrada pode comportar características intrínsecas de cada indivíduo, o que não é desejado na construção de uma rede para classificação de inúmeras entradas distintas. A operação de *pooling* é responsável por agrupar a coleção de características de uma convolução em um conjunto menor de valores (BOUREAU; PONCE; LECUN, 2010, p. 1). Além de mitigar o problema da especificidade da rede em relação aos dados de entrada, essa compactação também propicia desempenho à rede neural.

Há duas versões principais da função de *pooling*: O *pooling* médio e o máximo. Seus funcionamentos estão exemplificados na Figura 12. Em uma de suas publicações, Boureau junto a Ponce e Lecun, recomendam a combinação de ambas as versões da operação para minimizar taxas de erro das redes (BOUREAU; PONCE; LECUN, 2010).

Figura 12 – Exemplo de *poolings* máximo e médio.



Fonte: Adaptada de Motalebi (2020).

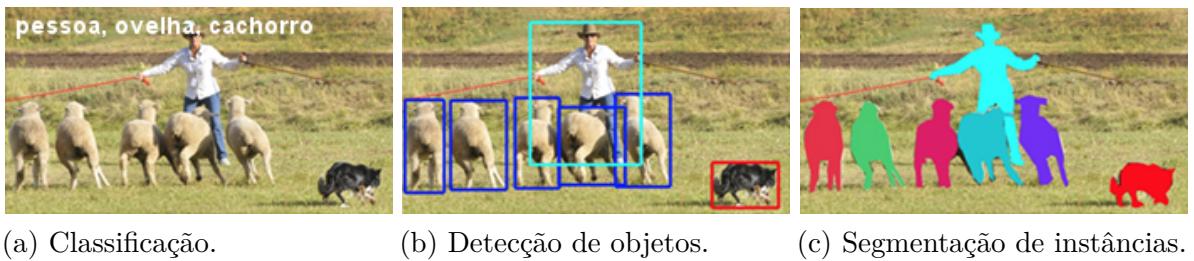
2.6 Classificação, detecção e segmentação de objetos

As tarefas de classificação e segmentação de imagens, bem como a detecção de objetos são processos que integram a área de Visão Computacional (WRIGHT et al., 2010; SZEGEDY et al., 2016), a qual tem como principal objetivo fornecer a capacidade da máquina receber e extrair informações de estímulos gráficos (SONKA; HLAVAC; BOYLE, 2013).

A classificação de imagens é o processo que consiste em calcular a probabilidade de uma imagem de entrada conter um objeto pertencente a uma determinada categoria e é, portanto, uma atividade chave para as tarefas de segmentação e detecção.

Já o processo de detecção de objetos visa aplicar um classificador suprarreferido para identificar múltiplos objetos relacionados a uma ou mais categorias e delimitar a área ocupada por objeto, marcando-a, geralmente com um contorno retangular, também chamado *bounding box* (FELZENSZWALB et al., 2010, p. 1). Com funcionamento semelhante, a segmentação de imagens também busca classificar objetos e destacar sua localização, contudo, em vez de trabalhar com delimitadores retangulares, esse método utiliza a marcação de cada *pixel* que compõe o objeto identificado (CHEN et al., 2018, p. 1). As três atividades de processamento de imagem são ilustradas na Figura 13.

Figura 13 – Exemplo de classificação, detecção de objeto e segmentação de imagem.



Fonte: Adaptada de Lin et al. (2014).

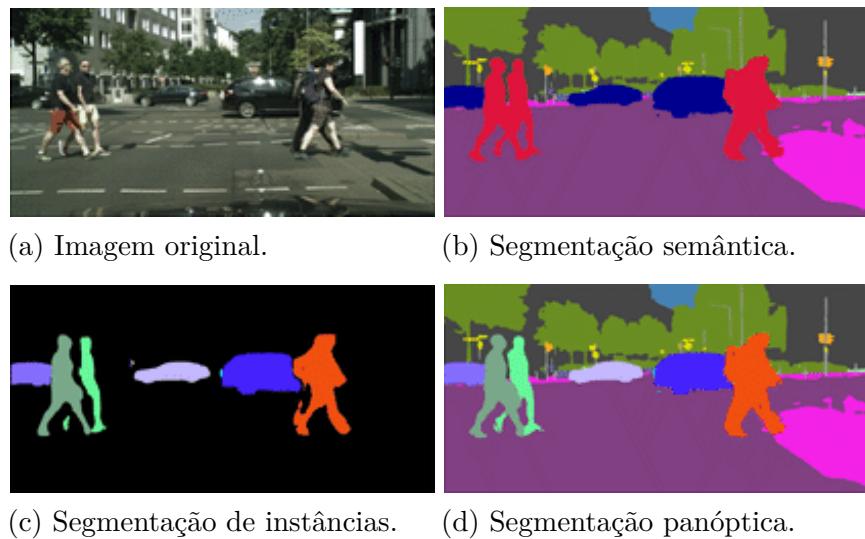
2.6.1 Tipos de segmentação de imagens

Quando há interesse em rotular os objetos pertencentes a um conjunto de categorias, porém sem distinguir objetos de uma mesma categoria, o uso da *segmentação semântica* em relação a outras técnicas é preferível. Na Figura 14b, os objetos pertencentes à categoria **pessoa** estão marcados de uma mesma cor, o mesmo vale para os veículos, já que pertencem à categoria **carro**.

A *segmentação de instâncias* atende a necessidade de diferenciar objetos de uma mesma categoria (DENG et al., 2018, p. 6774), portanto pode beneficiar áreas como videovigilância e robótica, entre outras (LIU et al., 2018, p. 1). Como exemplo, na Figura 14c, embora haja vários itens compondo o grupo **pessoas** há uma marcação de cor diferente para cada indivíduo desse grupo.

Já a *segmentação panóptica* é uma combinação das duas anteriores e seu uso tem o foco em tarefas em que é necessário não só segmentar e diferenciar instâncias, mas também rotular partes do ambiente, como construções, solo, estradas e vegetações (veja Figura 14d) presentes na imagem, como é o caso de problemas que envolvem guiagem autônoma ou realidade aumentada, por exemplo (KIRILLOV et al., 2019).

Figura 14 – Exemplo de segmentação semântica, de instâncias e panóptica.



Fonte: Adaptada de Costea, Petrovai e Nedevschi (2018)

2.7 Redes neurais do tipo encoder-decoder

Tanto a rede neural *DeepLab* como a rede neural *U-Net* são redes do tipo *encoder-decoder*, isto é, possuem um módulo **codificador** que é responsável por receber a entrada, extrair um mapa de características e reduzi-lo (etapa de contração de dados da entrada, em que a resolução é reduzida), e também apresentam conexão com um módulo **decodificador**, que tem como papel receber um mapa de características de baixa resolução e gerar um mapa com resolução alta (equivalente à da imagem original) como saída (BADRINARAYANAN; KENDALL; CIPOLLA, 2017, p. 1). Essa característica compartilhada e o fato de ambas serem projetadas para a segmentação semântica de imagem fazem dessas duas redes objetos interessantes para uma investigação performática comparativa.

2.7.1 A Rede neural *U-Net*

A primeira referência sobre a rede neural *U-Net* surgiu no *paper U-Net: Convolutional Networks for Biomedical Image Segmentation*, introduzido na literatura em 2015, o qual descreve o funcionamento da rede e os fatores que a fizeram vencer o desafio de rastreamento de células do Simpósio Internacional de Imagem Biomédica (ISBI) (RONNEBERGER; FISCHER; BROX, 2015, p. 1).

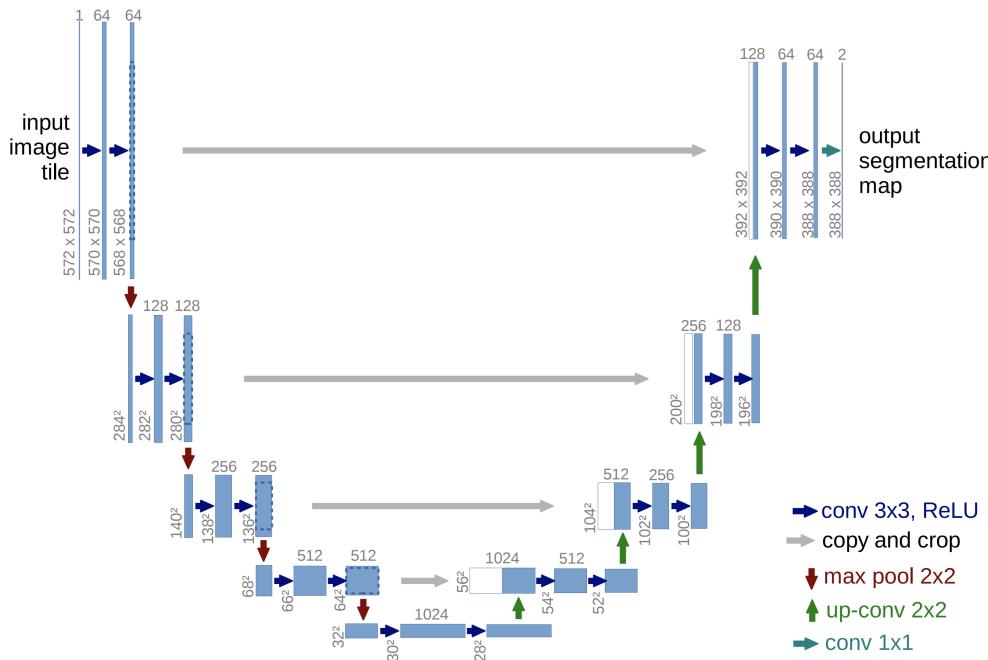
Uma das características marcantes da rede neural é sua estrutura totalmente convolucional, o que implica dizer que a *U-Net*, em suas camadas ocultas, aproveita-se de informações contextuais e locais de cada bloco processado, e não de informações globais, que geralmente são obtidas quando a entrada é tratada integralmente, sem redução de resolução durante seu processamento. Redes com essa particularidade são capazes de receber uma entrada de qualquer tamanho e retornar uma saída com dimensão correspondente, de forma eficiente

(FU et al., 2016, p. 699). Dolz, Desrosiers e Ayed indicam que essa eficiência é consequência das redes totalmente convolucionais utilizarem apenas uma etapa para produzir o mapa de probabilidades de classe, comportamento que dispensaria operações redundantes de convolução, como ocorre em redes convolucionais tradicionais (DOLZ; DESROSIERS; AYED, 2018).

Em síntese, a primeira metade da arquitetura da rede é composta por uma sequência de *poolings* máximo, usados para capturar características mais acentuadas de um fragmento processado, enquanto sua segunda metade é definida por uma sucessão de *upsampling* com convolução, a fim de expandir a resolução da saída. O mapa de características gerado por uma etapa de contração é usado como entrada por uma etapa de expansão correspondente, que é responsável por combinar o mapa recebido com o resultado de uma convolução ascendente (convolução transposta ou deconvolução) de dois *pixels* de largura e altura.

A arquitetura da *U-Net*, apresentada na Figura 15, possui simetria aproximada entre o número de etapas de contração e de expansão. Essa característica confere à arquitetura um formato de “U”, a qual é autoexplicativa em termos do nome da rede neural.

Figura 15 – Arquitetura da rede *U-Net*.



Fonte: Ronneberger, Fischer e Brox (2015).

Embora essa rede neural tenha se originado e se destacado no contexto de segmentação semântica de imagens biomédicas, suas aplicações não se restringem somente a essa categoria de imagens, como demonstram os seguintes autores e suas contribuições:

- Nguyen e outros pesquisadores, em 2018, com o auxílio de um pequeno veículo aéreo

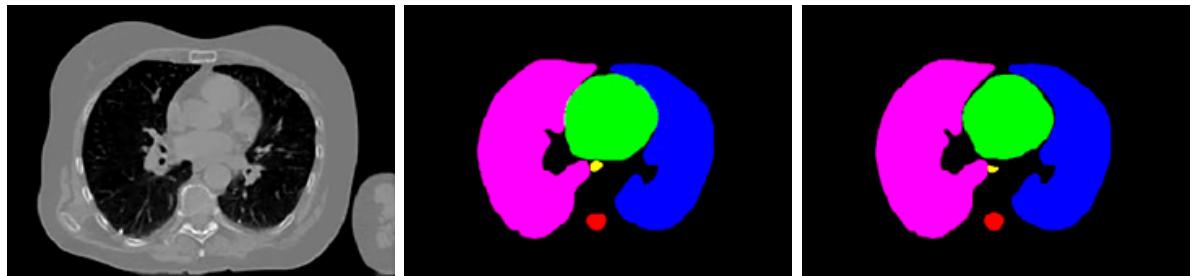
para captura de imagens, empregaram a rede totalmente convolucional para segmentar áreas de corrosão em infraestruturas de difícil acesso humano, que apresentem riscos (NGUYEN et al., 2018). Os autores propuseram a integração da solução desenvolvida a projetos semelhantes para a criação de uma ferramenta de inspeção automatizada e em tempo real, a fim de reduzir ou até dispensar a necessidade da presença humana para avaliar a formação de corrosão, ferrugem ou rachaduras em barragens, condutos e demais estruturas que exigem inspeções periódicas;

- Collett, Flood e Watson, em 2019, realizaram a segmentação de áreas cobertas por árvores e arbustos de grande porte em aerofotografias obtidas por satélites do estado Queensland, na Austrália (FLOOD; WATSON; COLLETT, 2019). Um dos principais objetivos dos pesquisadores era automatizar a geração de mapas com a delimitação dessas áreas, para que essas informações pudessem ser utilizadas por órgãos governamentais na implementação de políticas ambientais em reação ao desmatamento e incêndios de causa natural. Tais objetivos são factíveis e alimentados pela precisão aproximada de 90% obtida pela rede nesse caso de estudo;
- Um grupo de pesquisadores utilizou a rede para realizar a segmentação semântica de raízes de plantas a fim de propor uma alternativa para facilitar e automatizar parte do processo de fenotipagem (SMITH et al., 2020), etapa em que ocorre análise de características físicas do organismo com foco na seleção de genótipos que proporcionarão plantas com maior adaptabilidade a ambientes com limitação de recursos (FIORANI; SCHURR, 2013).

2.7.1.1 Aplicações da rede neural *U-Net* em segmentação de imagens biomédicas

Para realizar a segmentação semântica de órgãos que compõem a região do tórax, em 2019, Dong e companheiros de pesquisa empregaram uma modificação da rede neural *U-Net* e a alimentaram com um conjunto de tomografias computadorizadas da caixa torácica de 35 pacientes (DONG et al., 2019), a fim de destacar a área de cada órgão de interesse para facilitar o processo de planejamento de tratamentos como radioterapia. Um de seus resultados é apresentado na Figura 16.

Figura 16 – Segmentação dos pulmões (em azul e magenta), coração (em verde), esôfago (em amarelo) e medula espinhal (em vermelho).

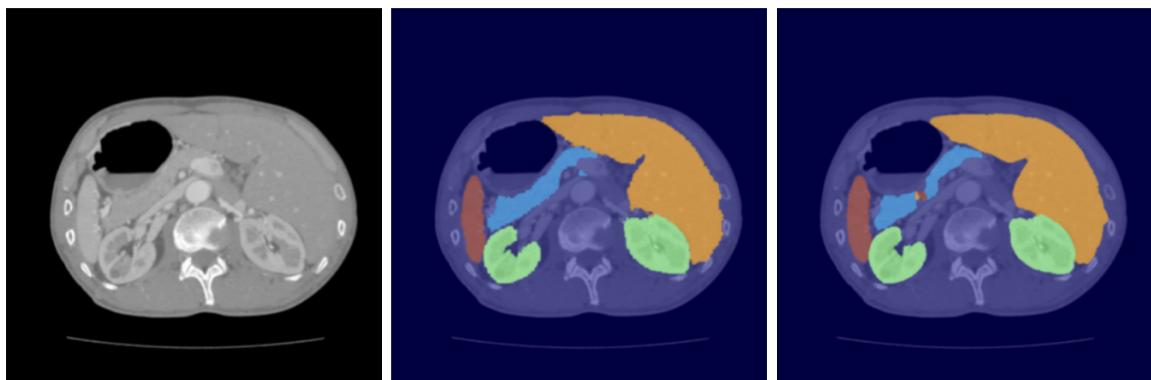


(a) Imagem original. (b) Rotulada manualmente. (c) Segmentada pela RNA.

Fonte: Adaptada de Dong et al. (2019).

Em 2018, Chen e outros pesquisadores propuseram a comparação de versões da *U-Net* e um método próprio para segmentar o baço, fígado, pâncreas e rins (um resultado pode ser visto na Figura 17), além de estruturas cerebrais como o córtex e ventrículos em tomografias (CHEN et al., 2018). A segmentação desses elementos e estruturas não só agiliza o processo de localização das regiões de interesse como também possibilita a avaliação de um especialista por meio da observação de possíveis danos que podem estar presentes nos órgãos analisados.

Figura 17 – Segmentação de baço (marrom), fígado (laranja), pâncreas (azul-claro) e rins (verde).

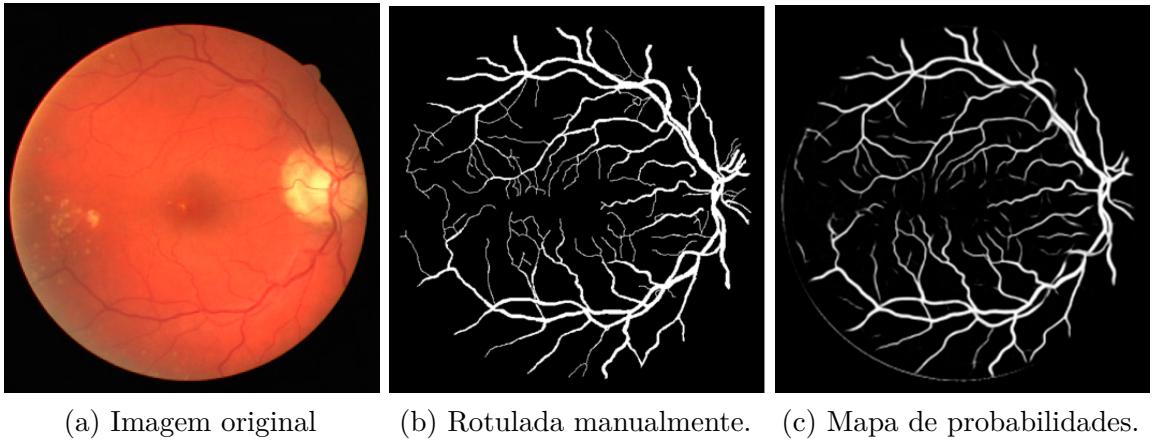


(a) Imagem original. (b) Rotulada manualmente. (c) Segmentada pela RNA.

Fonte: Chen et al. (2018)

Wang e colegas, em 2019, compararam a *Dense U-Net* com a rede neural original na tarefa de segmentar vasos sanguíneos da retina, como pode ser exemplificado pela Figura 18 (WANG et al., 2019). Ambos os modelos apresentaram excelência com desempenho igual ou superior a 95% em diversas métricas de avaliação. A localização e posterior análise desses vasos podem auxiliar no diagnóstico de doenças cardiovasculares, hipertensão, glaucoma e retinopatia diabética (ABRAMOFF; GARVIN; SONKA, 2010).

Figura 18 – (a) Imagem original sem processamento. (b) Imagem rotulada manualmente pelos autores. (c) Mapa de probabilidades gerado pela *Dense U-Net*.



Fonte: Wang et al. (2019).

2.7.2 A rede neural *DeepLab*

Projetada por pesquisadores da *Google* e divulgada no *paper Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs* (CHEN et al., 2014), a rede neural *DeepLab* é reconhecida como uma das redes que compõem o estado da arte em segmentação semântica. Por ser uma rede com potencial e ter sua implementação em código aberto, com o passar dos anos, alterações foram sugeridas, realizadas e geraram algumas versões da rede que serão abordadas nas seções a seguir.

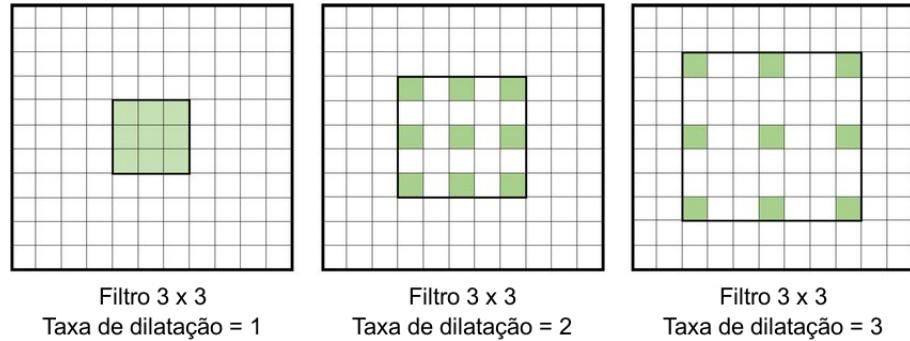
Segundo os idealizadores da rede neural *DeepLab*, há dois grandes complicadores para rotulação de imagens com redes neurais convolucionais profundas (RNCP): A redução da resolução da imagem original ao longo do processamento pela rede e a necessidade de se obter invariância sobre as dimensões de um objeto (invariância espacial). Para minimizar os impactos desses problemas no processo de segmentação semântica, os autores incluíram respectivamente o uso de uma operação denominada **convolução atrosa** e do **campo aleatório condicional totalmente conectado** (CACTC) na primeira versão da rede.

Como visto na arquitetura da rede neural *U-Net*, a aplicação sequencial de operações de *pooling* reduz a resolução da imagem ao longo do processo de *encoder*. Essa alteração nos limites da imagem pode resultar na perda de informações pelos mapas de características extraídos a cada etapa de convolução. A rede neural *U-Net* contorna esse problema com a aplicação de uma quantidade semelhante de etapas de expansão (chamadas de *up-sampling*), que requer recursos adicionais, como processamento e memória.

Conhecida também como convolução dilatada, a convolução atrosa é a solução incluída na implementação da rede neural *DeepLab* e é responsável por expandir o mapa de características extraído de um segmento de uma imagem. A dilatação ocorre por meio da

inserção de “buracos” (valores zero) entre valores diferentes de zero, dada uma taxa de dilatação, como pode ser visto na Figura 19.

Figura 19 – Exemplo de convolução atrosa para um filtro de 3×3 .

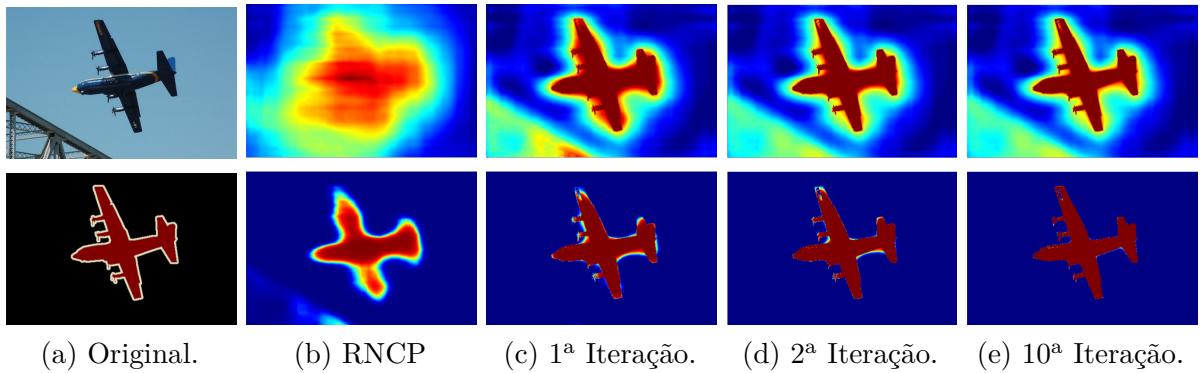


Fonte: Adaptada de Morales et al. (2018).

A fim de obter invariância para a extração de informações espaciais, os autores da rede neural *DeepLab* propuseram para última camada antes da saída o uso do CACTC, o qual explora a total conectividade entre os nós da camada para extrair detalhes finos das saídas geradas pelas camadas convolucionais. O campo é capaz de combinar as propriedades capturadas localmente, geralmente *pixel a pixel*, durante as etapas de convolução e combiná-las para gerar um mapa de características completo e preciso. Essa extração global de características é um comportamento herdado de camadas totalmente conectadas (KRÄHENBÜHL; KOLTUN, 2012).

Há casos em que o processo de segmentação semântica resulta em um modelo ruidoso, com *pixels* espacialmente próximos recebendo o mesmo rótulo de classificação, ainda que não pertençam ao mesmo objeto identificado, portanto, outra aplicação para o CACTC é a suavização de modelos ruidosos. Algumas iterações sucessivas do Campo Aleatório são suficientes para gerar refinamentos visíveis e significativos no resultado da segmentação semântica, como pode ser observado na Figura 20, exemplo dado pelos autores da rede.

Figura 20 – (a) Imagem original e imagem rotulada manualmente. (b) Saída de uma rede neural convolucional profunda. (c) Saída da primeira iteração do CACTC. (d) Saída da segunda iteração do CACTC. (e) Saída após a décima iteração do CACTC.



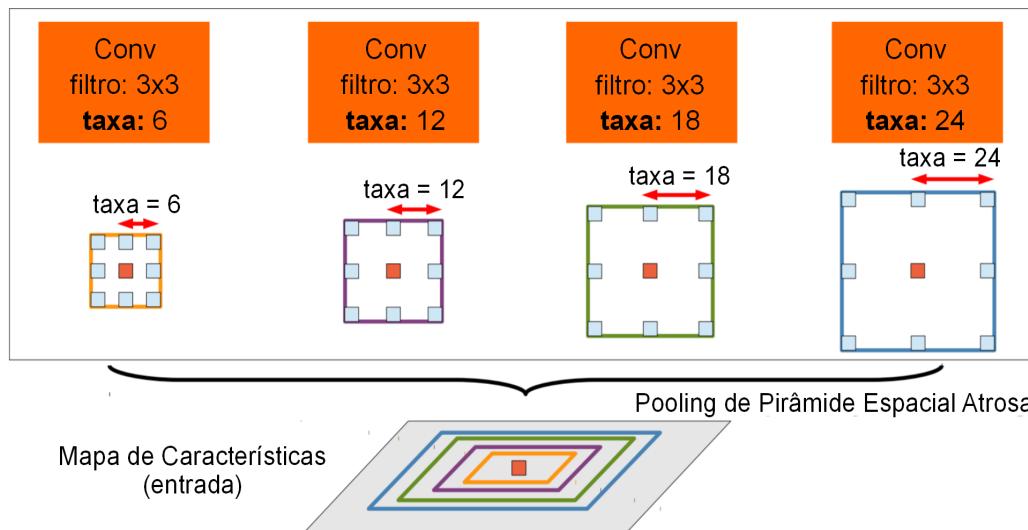
Fonte: Adaptada de Chen et al. (2014).

2.7.2.1 Versões e aprimoramentos da rede neural *DeepLab*

A popularidade da primeira versão da rede motivou a continuação de seu desenvolvimento e a implementação de melhorias, fatores que foram cruciais para o surgimento de outras duas variações principais da rede neural *DeepLab*.

Um dos principais problemas enfrentados pela rede neural *DeepLab* em sua primeira versão era a segmentação de objetos de mesma classe, porém em dimensões distintas. A segunda versão da rede neural trouxe como principal melhoria o uso do *Pooling de Pirâmide Espacial Atrosa* (PPEA), o qual é responsável por combinar as características extraídas de convoluções com diversas taxas de amostragem (CHEN et al., 2016). Esse processo está ilustrado na Figura 21.

Figura 21 – Exemplo de *Pooling de Pirâmide Espacial Atrosa*.

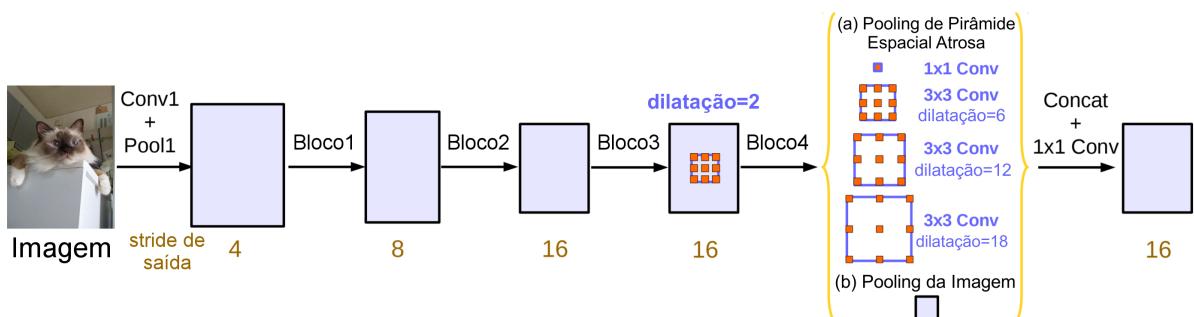


Fonte: Adaptada de Morales et al. (2018).

Durante a trajetória de evolução da rede, recursos não só foram adicionados, mas também removidos, como é o caso do campo aleatório condicional, não incluso na terceira versão. Além disso, a DeepLabV3 trouxe diversas modificações como:

- Inclusão de *Multi-grid* para definir a taxa de dilatação usada pela convolução atrosa. Exemplo: Para um *stride* de saída igual a 16 e *multi-grid* igual a (1, 2, 4), as taxas para convolução atrosa são o dobro do *multi-grid*, isto é, a sequência (2, 4, 8) (CHEN et al., 2017, p. 4);
- Uso da normalização em lote no PPEA, alteração essa que como os autores da normalização apuraram, não só possibilita reduzir o número de passos da etapa de treinamento como também pode maximizar a acurácia da rede;
- Combinação de uma convolução 1x1 *pixel* com outras três convoluções 3x3 (com taxa de dilatação 6, 12 e 18 unidades, respectivamente, considerando *stride* de saída igual a 16 *pixels*) na composição do PPEA. Veja a Figura 22;
- Concatenação do resultado do PPEA com uma convolução 1x1, antes de gerar os *logits* finais. Processo exemplificado na Figura 22;
- Inclusão de um *pooling* médio global sobre o último mapa de características da rede a fim de incorporar informações de contexto global;

Figura 22 – Fluxo dos módulos com convolução atrosa (PPEA).



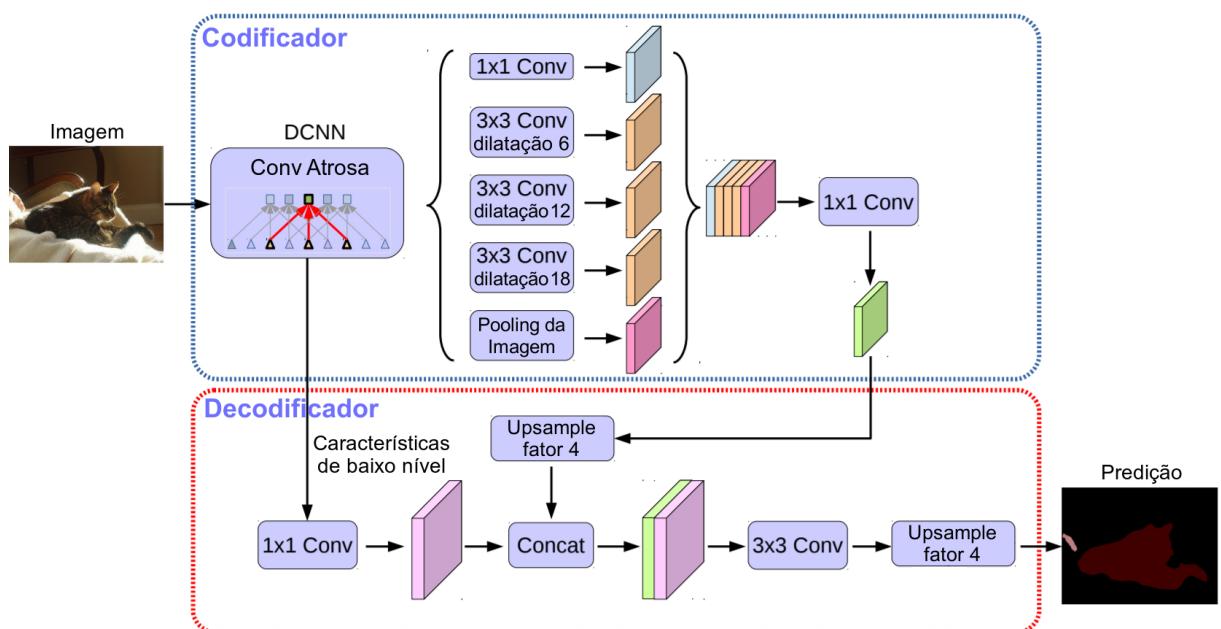
Fonte: Adaptada de Chen et al. (2017).

Nos experimentos computacionais deste trabalho utilizaremos a versão DeepLabV3+, que como o nome indica, é uma extensão da terceira versão da rede (CHEN et al., 2018). Uma representação simplificada de sua arquitetura está disponível na Figura 23, além de a seguir estarem listados alguns de seus principais aperfeiçoamentos:

- Reestruturação dos componentes codificador-decodificador, trazendo um codificador mais robusto e a simplificação do módulo decodificador;

- Domínio sobre a resolução do mapa de características extraídos pela convolução atrosa, o qual torna possível optar por priorizar precisão ou tempo de execução do processo;
- Inclusão e modificação do modelo *Aligned Xception* (que também é uma versão alterada do modelo *Xception*) para reduzir o tempo de computação. As principais alterações realizadas pelos autores foram: Substituição de todas operações de *pooling* máximo por convoluções separáveis em profundidade e uso das operações de normalização em lote e a aplicação da função de ativação ReLU após cada convolução em profundidade de dimensão 3x3.

Figura 23 – Arquitetura da rede *DeepLabV3+*.



Fonte: Adaptada de Chen et al. (2018).

Por ter atingido o estado da arte na tarefa de segmentação semântica e ter sua implementação acessível, o conjunto de versões da *DeepLab* influenciou o desenvolvimento de diversas aplicações:

- Alahari, Schmid e Tokmakov empregaram a segunda versão da *DeepLab* junto a um módulo de memória (componente convolucional recorrente, capaz de relacionar estados com instantes no tempo) para compor uma rede capaz de realizar a tarefa de segmentação semântica de objetos que apresentam movimentos em vídeo (TOKMAKOV; SCHMID; ALAHARI, 2018);
- A fim de facilitar a identificação de doenças reprodutivas e fatores que podem causar estresse em vacas leiteiras, pesquisadores propuseram uma aplicação que faz uso de segmentação semântica (realizada pela *DeepLabV3+*) combinada com outros algoritmos,

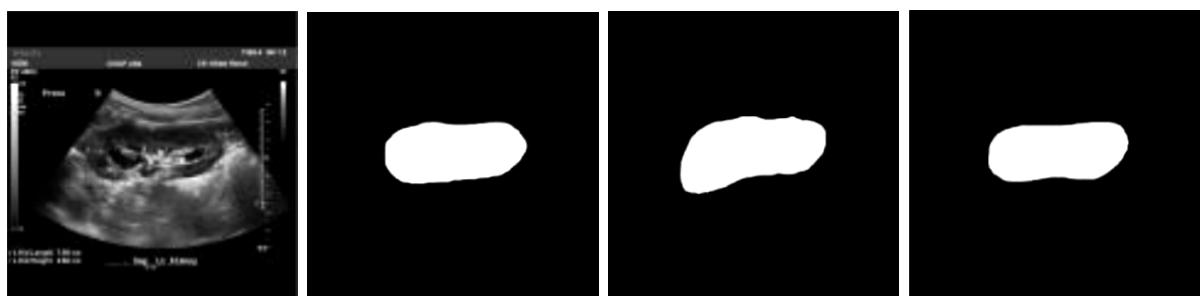
para estimar a taxa respiratória dos animais, que pode ser utilizada como indicador dos problemas citados (WU et al., 2020);

- De posse de milhares de tomografias computadorizadas do tórax, tanto Burgos-Artizzu quanto o grupo com os pesquisadores El-Bana, Al-Kabbany, Sharkas exploraram respectivamente, os modelos da *DeepLabV2* (BURGOS-ARTIZZU, 2020) e da *DeepLabV3+* (EL-BANA; AL-KABBANY; SHARKAS, 2020), para trabalhar no desenvolvimento de uma aplicação capaz de identificar a manifestação da doença **COVID-19**, a fim de oferecer um segundo diagnóstico, preliminar e digital, para auxiliar os profissionais que diagnosticam a enfermidade.

2.7.2.2 Aplicações da rede neural *DeepLab* em segmentação de imagens biomédicas

Como já dito, a gama de aplicações das versões da rede neural *DeepLab* é bastante diversificada. Embora seu desenvolvimento não tenha sido focado na segmentação de imagens biomédicas, há várias aplicações que evidenciam o sucesso da rede também com esse tipo de dados, como é o caso apresentado por Yin e pesquisadores associados, em que uma derivação da *DeepLabv2* foi utilizada para segmentação de imagens de rins em imagens de ultrassonografias (YIN et al., 2019), como é exemplificado na Figura 24. Os autores relatam que a segmentação desses órgãos é um recurso amplamente utilizado para auxiliar a diagnosticar doenças renais, principalmente aquelas que apresentam indícios anatômicos, e que o cálculo das dimensões do órgão é comumente acompanhado de segmentações manuais ou de confiabilidade reduzida.

Figura 24 – (a) Imagem original da ultrassonografia. (b) Imagem rotulada manualmente pelos autores. (c) Segmentação obtida pela *DeepLab* versão 2. (d) Segmentação obtida pela versão customizada da *DeepLabv2*.



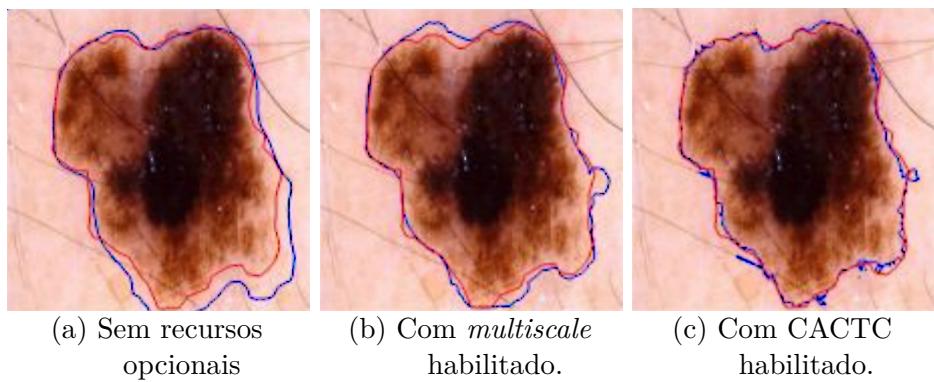
(a) Imagem original (b) Rotulagem manual (c) DeepLab V2 (d) Solução proposta

Fonte: Yin et al. (2019).

A união de características peculiares dos dados com os recursos opcionais adicionados a cada versão da rede possibilitam identificar a combinação que mais favorece a segmentação de um conjunto específico de imagens, como demonstram os experimentos realizados por Deng e outros pesquisadores, em que o uso ou não de alguns recursos fornecidos pela rede

neural *DeepLab* favorece a segmentação de lesões na pele humana (DENG et al., 2017), como é exemplificado na Figura 25. Neste caso, a segmentação das lesões é uma das etapas chave do processo de dermatoscopia, uma das técnicas não invasivas mais utilizadas para diagnosticar o melanoma, considerado uma das formas mais mortais do câncer de pele (JERANT et al., 2000).

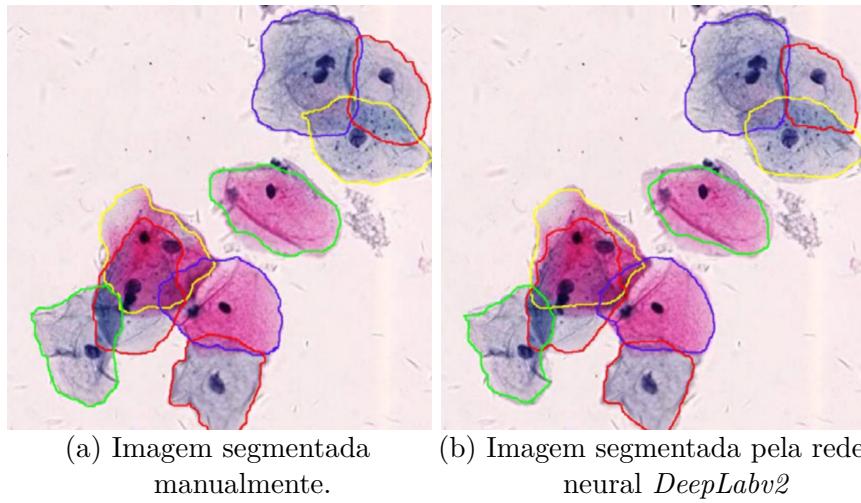
Figura 25 – Em vermelho está a delimitação realizada pelos autores e em azul a realizada pela rede neural. (a) Segmentação realizada pela rede neural *DeepLabv1* sem uso de recursos opcionais adicionais. (b) Segmentação realizada pela rede neural *DeepLabv1* com o recurso *multiscale* habilitado. (c) Segmentação realizada pela rede neural *DeepLabv1* com recurso campo aleatório condicional habilitado.



Fonte: Deng et al. (2017)

Além de aplicações para segmentação de imagens de tumores, lesões e órgãos, a rede neural *DeepLab* junto com a rede neural *U-Net* também ocupam um importante papel na segmentação a nível celular. Wan e outros autores relatam a importância do processo de segmentação semântica celular precisa para o diagnóstico de lesões pré-cancerosas cervicais auxiliado por computador (WAN et al., 2019). Os pesquisadores também ressaltam a importância da segmentação assertiva para automação do processo de análise de imagens citológicas a fim de agilizar e maximizar a detecção precoce das lesões e consequentemente, aumento da eficácia do tratamento. Ainda que haja problemas como sobreposição de células, formatos diversos ou fundos ruidosos, a versão modificada da rede neural *DeepLabv2* apresentou resultados promissores como os da saída presente na Figura 26.

Figura 26 – (a) Imagem rotulada manualmente pelos autores. (b) Segmentação realizada pela rede neural *DeepLabv2* com o recurso de campo aleatório condicional habilitado.



Fonte: Wan et al. (2019).

2.8 A Pneumonia e a sua dimensão

A pneumonia pode ser compreendida como a resposta inflamatória de um hospedeiro à invasão do parênquima pulmonar por micro-organismos (bactérias, vírus, fungos), com sua intensidade dependente da capacidade de reprodução do patógeno e do estado do sistema imunológico do hospedeiro (JOSEPH et al., 2010).

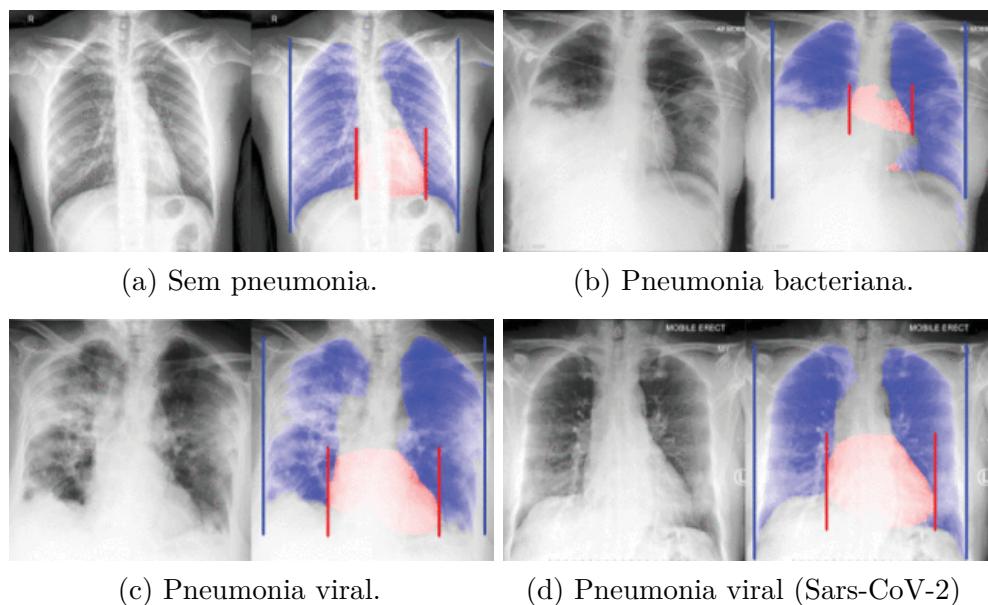
Descoberta durante a década de 80 do século 19, a bactéria *Streptococcus pneumoniae* foi um dos primeiros causadores da pneumonia a serem identificados. Os pesquisadores responsáveis pelo feito foram o químico francês *Louis Pasteur* (DION; ASHURST, 2020) e o médico *George Miller Sternberg* (STERNBERG, 1881). Embora haja diversos gatilhos e origens para a doença, sua versão bacteriana e a viral são as duas formas da pneumonia que mais levam pessoas a óbito globalmente, sendo as crianças de até 5 anos e os idosos com mais de 75 os grupos mais acometidos (RUUSKANEN et al., 2011).

As regiões mais afetadas pela enfermidade no mundo são os locais que compõem a África Subsaariana e o sudeste asiático. A maioria das nações que compõem essas regiões atingem ou superam a taxa de mortalidade de 100 pessoas por 100 mil habitantes. É importante mencionar que no ano de 2017 a taxa de mortalidade infantil causada pela pneumonia na Índia, Nigéria, Paquistão, República Democrática do Congo e Etiópia representou mais da metade dos óbitos de crianças no mundo inteiro pela doença (DADONAITÉ, 2018).

Alguns de seus sintomas, incluindo náuseas, vômito e diarreia, manifestam-se em uma quantidade menor de hospedeiros (ORTQVIST; HEDLUND; KALIN, 2005), diferentemente da dor no peito, falta de ar, tosse frequente, suores noturnos, febre e respiração ofegante, que

acometem uma quantidade maior de indivíduos (HOARE; LIM, 2006; NAIR; NIEDERMAN, 2011). É importante ressaltar que os sintomas da pneumonia viral e bacteriana são semelhantes. Em casos raros, essa manifestação pode variar de acordo com a espécie do agente patológico. Vítimas da doença *COVID-19*, por exemplo, sofrem com uma incidência maior de sintomas digestivos (PAN et al., 2020). Além disso, os sintomas da pneumonia podem ser confundidos com os de outras doenças, como tuberculose. Dessa forma, o diagnóstico definitivo deve ocorrer com suporte de uma radiografia do tórax (NAIR; NIEDERMAN, 2011), processo representado na Figura 27.

Figura 27 – Exemplos de radiografias torácicas de pessoas com e sem pneumonia (de diversas origens) e a segmentação da área pulmonar: (a) Pulmões sem manifestação da doença. (b) Manifestação bacteriana da pneumonia, uma de suas principais características é a opacidade generalizada de uma região, que acaba por dificultar a segmentação completa dos pulmões. (c) Pneumonia viral, apresenta manifestação bilateral, responsáveis pelas pequenas deformações na segmentação de ambos pulmões. (d) Pneumonia causada pelo vírus da *COVID-19*, manifesta-se bilateralmente como a causada por outros vírus, o que dificulta a distinção entre a doença causada pelo *Sars-CoV-2* ou outro agente viral.



Fonte: Oh, Park e Ye (2020).

Nair e Niederman descrevem a relevância da radiografia torácica para acompanhar o progresso da inflamação e avaliar qual local mais adequado para coleta de líquido pleural (NAIR; NIEDERMAN, 2011, p. 1148). El-Bana, Al-Kabbany e Sharkas fizeram uso das radiografias para distinguir as manifestações da pneumonia bacteriana, viral e a causada pelo recente vírus *Sars-CoV-2*, além de segmentarem regiões mais acentuadas da inflamação (EL-BANA; AL-KABBANY; SHARKAS, 2020). A existência desses objetivos complementares e o vasto acervo de radiografias torácicas acessíveis encorajam o

desenvolvimento de trabalhos e ferramentas capazes de não só distinguir o agente causador da pneumonia, mas também identificar seu avanço, de modo que os profissionais de saúde possam se precaver com tratamentos precoces para os casos atenuados e concentrarem seus esforços e infraestrutura hospitalar para manifestações da patologia em estado avançado.

3 MATERIAIS E MÉTODOS

Esta seção é dedicada a tratar adaptações realizadas nas RNAs selecionadas, seleção de subconjunto de dados utilizados e métricas já conhecidas para avaliar a eficiência não só de redes neurais artificiais como de algoritmos utilizados no campo da Inteligência Artificial.

3.1 Métricas para avaliação dos modelos

Há diversas formas e métricas para mensurar o desempenho de um algoritmo computacional, as fórmulas apresentadas abaixo são bastante conhecidas no campo da segmentação de objetos, seja semântica ou de instâncias. As métricas a serem apresentadas possuem como comportamento em comum o cálculo de similaridade entre conjuntos de elementos. Ao mensurar resultados oriundos de uma classificação de elementos, há de considerar verdadeiros positivos (VP), falsos positivos (FP), verdadeiros negativos (VN) e falsos negativos (FN), que no cenário da tarefa de segmentação, rotulam os *pixels*.

Segundo Murphy, a entropia pode ser compreendida como uma medida de incerteza sobre uma variável aleatória, ou mais especificamente, como o número médio de bits que é preciso para codificar dados de uma fonte com distribuição P usando um modelo P (MURPHY, 2012, p. 56).

A entropia cruzada seria o número médio de bits para codificar dados de uma fonte de distribuição P usando um modelo Q, ou seja, um modelo que não seja o esperado (ou ótimo), e que, portanto, acarretaria em um número maior de bits para codificação (MURPHY, 2012, p. 58). Tendo em vista que a entropia é uma medida de incerteza e trabalha com distribuições, é possível, calcular a dissimilaridade entre uma distribuição P, esperada, e outra, Q, que poderia ser obtida como resultado do aprendizado de máquina.

A **entropia cruzada** é representada pela seguinte equação:

$$H(p, q) = - \sum_k p_k \log_2 q_k \quad (1)$$

Li e Lee foram os responsáveis por propor a utilização da minimização da entropia cruzada como alternativa à limiarização por equilíbrio do histograma para critério de parada para a segmentação de imagem (LI; LEE, 1993).

O índice de Jaccard ou **interseção sobre a união** (IoU) é um métrica proposta pelo botânico Paul Jaccard sob o nome de “coeficiente de comunidade” (JACCARD, 1912, p. 41), esse índice apresenta amplo uso na avaliação do desempenho das RNA com foco em segmentação de imagens. Sua representação matemática, disposta na equação abaixo, se dá pela interseção entre a imagem esperada com o resultado da segmentação sobre a união entre as duas, por isso o título adaptado para Interseção sobre União em trabalhos

mais recentes. É importante notar que dentre todos índices que serão apresentados neste trabalho, esse é responsável por impor maior penalização à rede, uma vez que divide seu valor de verdadeiros positivos por sua soma com falsos positivos e falsos negativos.

$$\text{Interseção sobre união} = \frac{VP}{VP + FP + FN} \quad (2)$$

Representada pela soma de acertos sobre o todo, como descrito na equação abaixo, a **acurácia** tem o papel de mensurar o quanto a rede neural errou, semelhante ao índice IoU, contudo considerando também valores de verdadeiros negativos.

$$\text{Acurácia} = \frac{VP + VN}{VP + FP + VN + FN} \quad (3)$$

A medida de **sensibilidade** ou *recall*, representada pela equação a seguir, pode ser compreendida como a divisão do número de resultados relevantes retornados pela quantidade total de elementos relevantes que deveriam ser retornados (ESTRADA; JEPSON, 2009, p. 171). No campo da segmentação de imagem, a métrica pode ser entendida como o número de *pixels* segmentados corretamente pela RNA dividido pelo número total de *pixels* que deveriam ser segmentados.

$$\text{Sensibilidade} = \frac{VP}{VP + FN} \quad (4)$$

Com funcionamento semelhante a medida de sensibilidade, a **precisão** tem como finalidade mensurar a proporção de elementos relevantes em relação ao total de elementos resultantes, ou o número de *pixels* segmentados corretamente dividido pelo número total de *pixels* segmentados (ESTRADA; JEPSON, 2009, p. 171). Sua representação matemática pode ser conferida abaixo:

$$\text{Precisão} = \frac{VP}{VP + FP} \quad (5)$$

Além dos índices anteriores, há também o *f1-score*, também conhecido como coeficiente de Dice, sua fórmula é definida como a média harmônica entre precisão e a sensibilidade (KOSMAN; LEONARD, 2005, p. 417), sua equação é representada a seguir:

$$\text{Dice} = \frac{2 \times VP}{2 \times VP + FP + FN} \quad (6)$$

3.2 Conjunto de dados

Os dados utilizados neste trabalho foram obtidos a partir de um serviço online de anotações de imagens chamado V7 Darwin³, nessa coleção há mais de 6300 imagens pulmonares

³ <<https://github.com/v7labs/covid-19-xray-dataset>>. Acesso em: 31 jan. 2021

segmentadas e distribuídas nas categorias: pessoas saudáveis e com pneumonia (de origem bacteriana, viral, covid-19, fúngica). Segundo os responsáveis por compilar as imagens, tal conjunto de dados foi construído a partir de imagens de outras duas principais fontes, em seguida foram anotados e revisados manualmente por pessoas relacionadas ao serviço (V7) de anotações.

As imagens possuem dimensões variáveis, sendo as menores com 156 *pixels* de largura e altura, e as maiores podendo atingir 5700 *pixels* de altura e 4700 de largura. A variação de cores e resolução se dá pelo fato da coleção unificar imagens coletadas de dispositivos diversos de raio-X. Por meio da plataforma V7 Darwin é possível filtrar as imagens de pacientes pelo sexo, localidade, se houve intubação e sobrevivência.

3.3 Pré-processamento dos dados

Após realizar uma sequência de experimentos computacionais, observar os resultados obtidos e realizar um comparativo com valores comuns na literatura, identificou-se a necessidade de aplicar algumas transformações aos dados de entrada, de modo a realçar as características intrínsecas de cada causa de pneumonia.

As transformações que foram aplicadas e que serão apresentadas derivam de duas operações fundamentais do campo da morfologia matemática, a operação de dilatação (representada pelo símbolo \oplus) e a erosão (representada pelo símbolo \ominus). Ambas operações estão denotadas abaixo e utilizam um elemento estruturante (comumente uma matriz chamada de *kernel*), representado pela letra **B**. Para as equações abaixo, obtidas a partir do trabalho de Ji e Piper, considere A e B subconjuntos do espaço euclidiano, representado pela letra **E** (JI; PIPER; TANG, 1989).

$$\text{Dilatação } (A \oplus B) = x \in E^n \mid x = a + b, \quad (7)$$

para algum a pertencente a A e b pertencente a B

$$\text{Erosão } (A \ominus B) = x \in E^n \mid x = a - b, \quad (8)$$

para cada b em B , existe um a em A tal que $x = a - b$

Segundo Masad, Button e Papagiannakis, a aplicação da operação de erosão resulta na remoção de *pixels* de uma imagem dependendo das cores de seus *pixels* vizinhos, esse efeito tende a simplificar e refinar objetos contidos na imagem (MASAD; BUTTON; PAPAGIANNAKIS, 2000). Ainda segundo os autores, a operação de dilatação possui funcionamento oposto ao da erosão, uma vez que utiliza a adição de *pixels* para gerar uma nova saída.

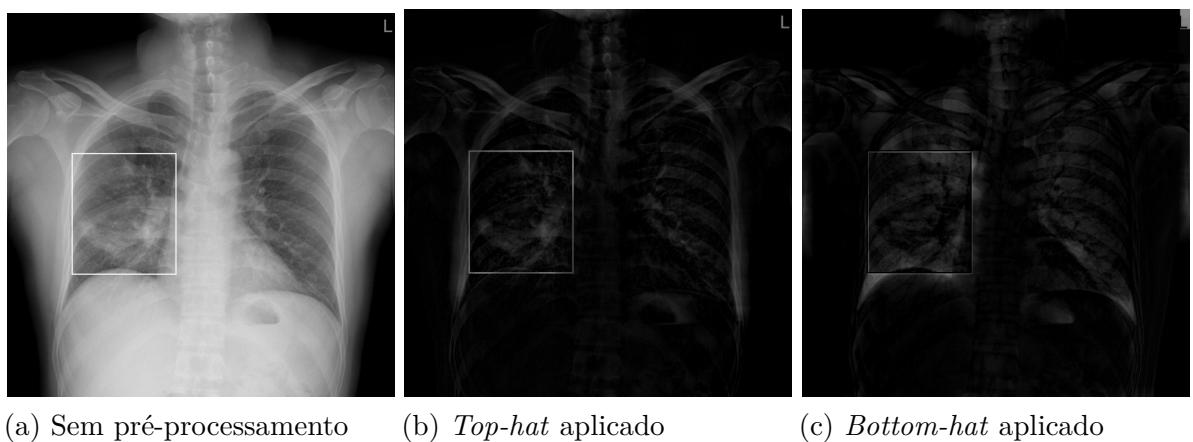
Há diversas derivações criadas partir das operações morfológicas citadas, duas amplamente conhecidas em trabalhos envolvendo processamento digital de imagens, são as operações de abertura (*opening*) e fechamento (*closing*), a primeira consiste na aplicação de uma função de dilatação sob a erosão da imagem de entrada, já a segunda é aplicação da erosão sob a dilatação da imagem (KUSHOL et al., 2017). As equações de ambas operações são apresentadas a seguir.

$$\text{Abertura } (A \circ B) = (A \ominus B) \oplus B \quad (9)$$

$$\text{Fechamento } (A \bullet B) = (A \oplus B) \ominus B \quad (10)$$

Mehrotra junto a outros pesquisadores, propõe a utilização da combinação de outras duas transformações para a segmentação de vasos sanguíneos a partir de imagens de retinas, *top-hat* e *bottom-hat*, a primeira é obtida a partir da diferença entre a imagem de entrada e sua abertura, já a segunda, é resultado da diferença entre a imagem de entrada e seu fechamento (MEHROTRA et al., 2014). As transformações citadas são exemplificadas na Figura 28, a dimensão usada para o elemento estruturante foi de 32 *pixels* de altura e largura.

Figura 28 – Exemplo de imagem de um paciente com pneumonia (causada pelo Sars-CoV-2)



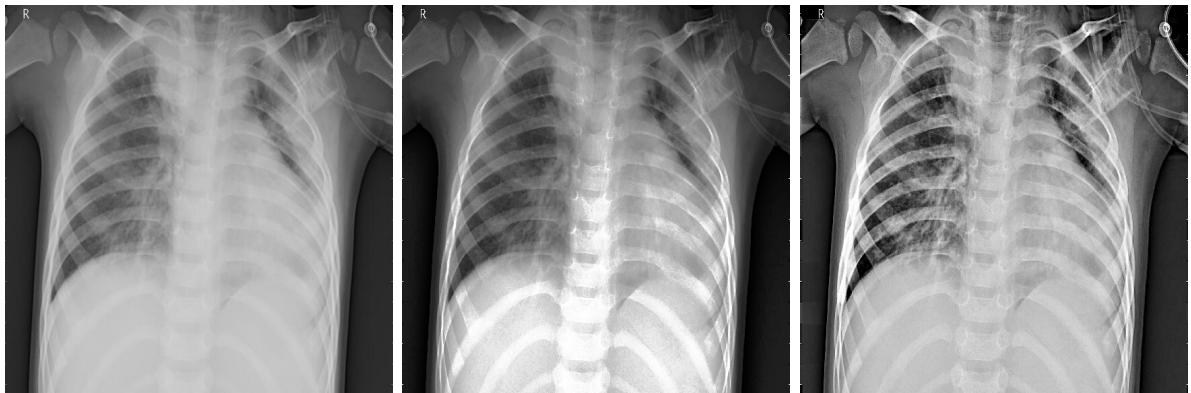
Fonte: Adaptada de Kermany (2018)

Um outro método amplamente utilizado para realizar o pré-processamento digital de imagens é a equalização de histograma, operação que consiste em utilizar a distribuição de probabilidade dos níveis de cinza da imagem de entrada para gerar uma nova imagem com a escala de cinza com maior nivelamento, o que geralmente resulta em imagens com contraste favorável para extração de características (CHEN; RAMLI, 2003).

As duas técnicas de pré-processamentos selecionadas foram a equalização de histograma e a diferença entre a soma do *top-hat* com a imagem de entrada e o *bottom-hat*, para

exemplificar os impactos visuais das transformações, a Figura 29 apresenta três versões de uma imagem do conjunto de dados utilizado nos experimentos computacionais.

Figura 29 – Exemplo de imagem de um paciente com pneumonia viral (não Sars-CoV-2)



(a) Sem pré-processamento (b) Equalização de histograma (c) Transformação morfológica

Fonte: Adaptada de Kermany (2018)

Aliado ao pré-processamento descrito, todas as imagens foram redimensionadas para 448 *pixels* de altura e largura, e suas anotações (máscaras), originalmente de pulmões demarcados pela cor branca, receberam uma cor para cada categoria de pneumonia analisada para que as redes neurais pudessem distinguir a anotação pertencente a cada classe.

3.4 Definição de parâmetros e hiperparâmetros

Durante a execução dos experimentos para definir uma combinação eficaz de valores para os parâmetros, diversas resoluções foram utilizadas, valores iguais para largura e altura, múltiplos de 32, com variação entre 192 e 640, além 768 *pixels*. Além disso, foram testados lotes (*batches*) de imagens contendo 1, 2, 4 e 8 elementos, contudo, para realização dos experimentos que serão apresentados, após considerar limites de *hardware*, fixamos o valor em 4 imagens por lote.

Por meio de experimentos computacionais que antecederam aos que serão apresentados, observamos que o aumento da resolução das imagens, bem como o aumento do tamanho dos lotes, resultavam em saídas majoritária e significativamente melhores do que com resoluções e lotes menores, consequentemente, o consumo de recursos como memória da GPU, o tempo de treinamento, validação e teste eram maximizados, assim, impondo a necessidade de encontrar valores que proporcionassem um equilíbrio satisfatório entre uso de recursos e aprendizado.

Com o objetivo de minimizar mais rapidamente a taxa de erro de uma RNA, são utilizados otimizadores, os quais são empregados na tarefa de encontrar o valor mínimo global da

função de perda, esse processo de otimização é conhecido como descida do gradiente. Um dos principais problemas nessa tarefa é a possibilidade da rede estacionar em um mínimo local. Para contornar o problema dos mínimos locais há técnicas como *Momentum*, que armazena dados (como a média móvel dos valores do gradiente) sobre mínimos anteriores para decidir em qual sentido e velocidade o gradiente deve ser percorrido, dessa maneira, a técnica acelera o encontro do mínimo global (RUDER, 2017).

Mesmo após encontrar o mínimo global, *Momentum* ainda realiza algumas verificações com variações pequenas de valores para confirmar se o estado atual não é um mínimo local (RUDER, 2017), essas oscilações são dispensadas pelo otimizador *Root Mean Square Propagation* (RMSProp), que também utiliza a média móvel, porém quadrática de valores de gradientes anteriores, para auxiliar na atualização de pesos sinápticos da RNA. Combinando a aceleração para encontrar o mínimo global do *Momentum* e a velocidade do RMSProp para se estabilizar após encontrar tal mínimo, surge o otimizador *Adaptive Moment Estimation* (Adam), que segundo os pesquisadores responsáveis por sua implementação, é computacionalmente eficiente, exige baixo consumo de memória e é adequado para problemas com gradientes esparsos (KINGMA; BA, 2015).

Um dos parâmetros mais relevantes para determinar a velocidade de aprendizado de uma RNA, a taxa inicial de aprendizado, foi definida como 1×10^{-4} , após identificar resultados satisfatórios em experimentos realizados pelo autor e por outros pesquisadores do campo de segmentação semântica como Vesal, Ravikumar, Maier (VESAL; RAVIKUMAR; MAIER, 2018), e Dolz, Desrosiers e Ayed (DOLZ; DESROSIERS; AYED, 2019), que ao utilizarem valores próximos de 1×10^{-4} para o otimizador Adam, registraram maior acurácia da rede. As investigações realizadas por Mohagheghi, Foruzan, Chen (MOHAGHEGHI; FORUZAN; CHEN, 2020) e por Lotter, Sorensen e Cox (LOTTER; SORENSEN; COX, 2017), indicam que a taxa inicial de aprendizado escolhida para Adam também é promissora para o otimizador RMSProp.

Para reduzir a probabilidade ou postergar a ocorrência de sobre-ajuste (*overfitting*), cenário em que a RNA especializou-se de forma extrema na compreensão de dados de treino e reduziu sua capacidade de generalização, adicionamos camadas de *dropout* após cada Convolução 2D realizadas pela DeepLab e também cada *max pooling* 2D e concatenação realizada pela U-Net. Essa camada adicional tem como tarefa eliminar um conjunto de neurônios artificiais dada uma taxa de probabilidade, que definimos sendo 20%, com base em trabalhos semelhantes realizados por Fu (FU et al., 2017), Wang (WANG et al., 2018), Colman (COLMAN et al., 2021) e pesquisadores associados.

Como função de perda, utilizada para mensurar se a RNA está aprendendo com os dados ou indicar se ela está cegamente dependente dos dados de treinamento, utilizamos a função de entropia cruzada categórica, semelhante a sua versão já explicada, porém voltada para

classificação multi-classe.

3.4.1 Recursos computacionais

Os experimentos foram executados em dois computadores, a configuração mínima entre as máquinas foi:

- GPU: Nvidia GeForce RTXTM 2060 (6GB GDDR6, 1.680 GHz)
- CPU: Intel® CoreTM i9-9900KF (3.6GHz, 8 núcleos, 16 *threads*)
- RAM: 32GB (DDR4, 2.666 GHz)

É necessário ressaltar que do *hardware* utilizado, a GPU foi o recurso de uso mais intenso durante os experimentos, enquanto a memória RAM e a CPU tiveram maior tempo de ociosidade e uso de uma parte pequena do potencial disponível. Por meio da observação do consumo desses recursos ao longo das execuções, estimamos que um computador com 6GB de memória RAM, uma CPU com 4 núcleos e uma GPU Nvidia (com suporte à tecnologia CUDA) com 6GB de memória atenda os requisitos mínimos para reprodução dos experimentos.

Para realizar os testes foi utilizado a linguagem de programação Python 3, os *frameworks* Tensorflow e Keras, as bibliotecas opencv-python, numpy, pandas, além de um conjunto de dependências para modelos que trabalham com classificação e segmentação de imagens. Todas as dependências citadas além das utilizadas para gerência dos conjuntos de dados e resultados, estão listadas no arquivo “requirements.txt” do repositório deste trabalho, no Github⁴, plataforma que hospeda também as implementações das redes DeepLab⁵ e U-Net⁶, adaptadas e utilizadas neste trabalho.

⁴ <<https://github.com/duraes-antonio/tcc>>. Acesso em: 22 fev. 2021

⁵ <<https://github.com/sainatarajan/deeplabv3plus-keras-jupyter-notebook>>. Acesso em: 14 mar. 2021

⁶ <https://github.com/qubvel/segmentation_models>. Acesso em: 14 mar. 2021

4 EXPERIMENTOS, RESULTADOS E DISCUSSÃO

Neste capítulo será apresentado o desempenho das RNAs na tarefa de segmentação de 4686 imagens de tomografias da caixa torácica. Nas tabelas abaixo, cada linha representa todos os valores obtidos de acordo com o parâmetro definido pelas colunas da célula. Todos os índices exibidos a seguir tem valores reais entre zero (pior) e um (melhor), os números destacados em negrito representam os melhores resultados obtidos.

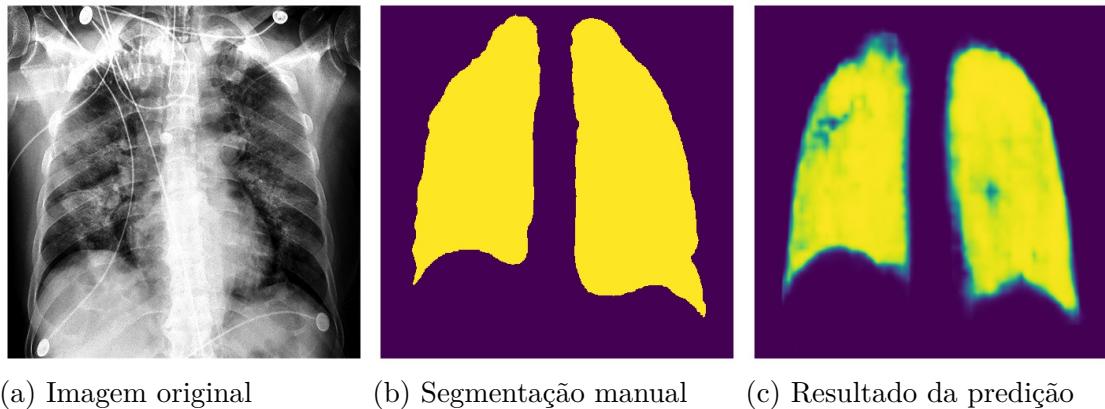
Para mensurar o desempenho das RNAs durante a execução dos experimentos, realizamos cinco execuções de cada caso (combinação de parâmetros e hiperparâmetros) de teste, obtemos o valor de cada métrica (apresentadas no capítulo anterior) da última época e calculamos a média desses valores, considerando as cinco execuções. Nas análises a seguir, para comparar os resultados obtidos pelas redes será utilizado o módulo da diferença entre os valores obtidos por cada uma.

O total de épocas foi definido como 35 para o treino de ambas redes, a rede *backbone* (rede de arquitetura mais simples e de propósito menos específico, responsável por extrair características básicas de uma imagem) utilizada pela U-Net foi a VGG-19, e pela DeepLab, em detrimento das limitações de recursos computacionais, optamos pela MobileNetV2 e *stride* de saída igual à 16.

Visando obter um desempenho médio das redes neurais em relação às métricas utilizadas, realizamos os experimentos em todos os seus 120 cenários, cinco vezes cada um e registramos os valores médios nas tabelas apresentadas nas subseções abaixo. Como o desvio padrão obtido ao longo das realizações indicou baixa variabilidade nos resultados de todas as métricas ao longo das execuções, optamos por não aumentar o número global de repetições do experimento. Os melhores resultados foram destacados em negrito em cada caso.

Após realizar o treinamento de uma das redes e armazenar os pesos de saída do modelo treinado é possível obter representações visuais das predições como as mostradas abaixo, tais saídas usam a técnica de mapa de calor para indicar com cores quentes os *pixels* que pertencem a uma determinada causa da doença. A Figura 30, dispõe o resultado de uma predição assertiva em pulmões acometidos pela pneumonia causada pelo agente viral Sars-cov 2.

Figura 30 – Exemplo de predição assertiva em pulmões com pneumonia (Sars-cov 2)



(a) Imagem original

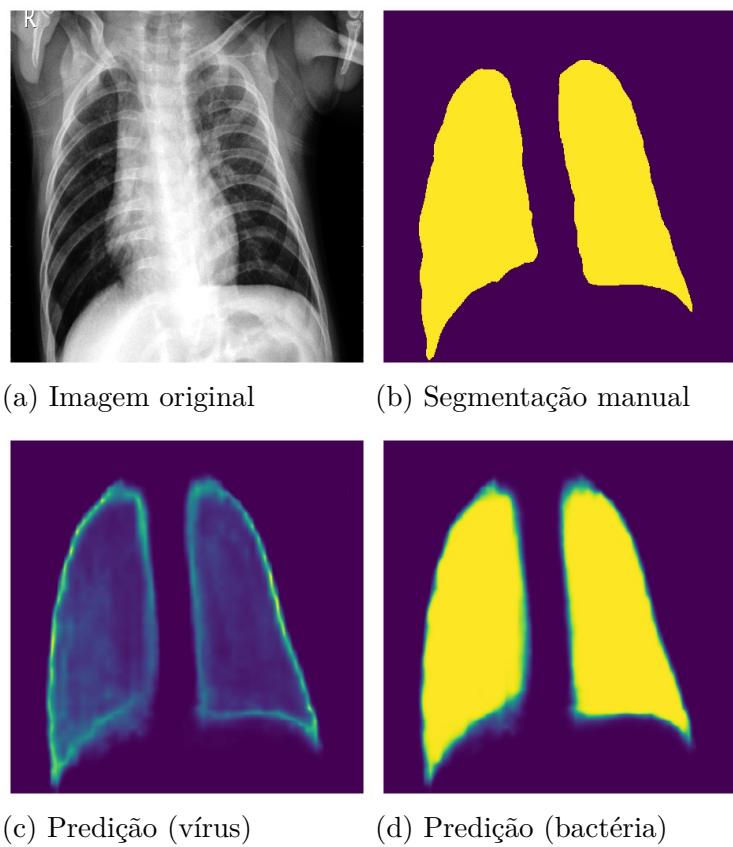
(b) Segmentação manual

(c) Resultado da predição

Fonte: Elaborada pelo autor (2021)

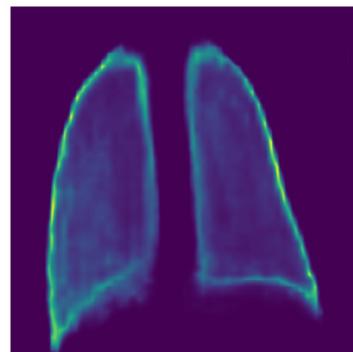
Há também casos em que as redes não conseguiram distinguir com competência as causas da doença, porém segmentaram de forma adequada grande parte da região de interesse (como é exemplificado na Figura 31), casos assim resultam em um menor valor para o índice de IoU médio (média realizada considerando os valores IoU por classe) e não necessariamente reduzem os outros índices.

Figura 31 – Exemplo de predição incorreta para caso de pneumonia viral

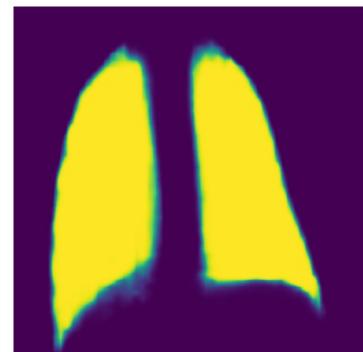


(a) Imagem original

(b) Segmentação manual



(c) Predição (vírus)



(d) Predição (bactéria)

Fonte: Elaborado pelo autor (2021)

4.1 Resultados: Etapa de treino

Tanto a Tabela 1 quanto a Tabela 2 apresentam os resultados obtidos na etapa de treinamento com o conjunto de dados particionado em 70% para treino, 20% validação e 10% teste, da DeepLab e U-Net, respectivamente. Ao analisar os índices de acurácia, precisão e *f1-score*, nota-se uma vantagem pouco significativa da U-Net, em contrapartida, a DeepLab, sem o recurso de *dropout*, apresenta desempenho superior nas métricas de IoU e sensibilidade.

| Métrica | Histograma equalizado | | | | Transformação morfológica | | | |
|----------|-----------------------|---------|---------|---------|---------------------------|---------|---------------|---------|
| | Adam | | RMSProp | | Adam | | RMSProp | |
| | s/ drop | c/ drop | s/ drop | c/ drop | s/ drop | c/ drop | s/ drop | c/ drop |
| Acurácia | 0,9792 | 0,9720 | 0,9781 | 0,9695 | 0,9798 | 0,9729 | 0,9789 | 0,9714 |
| F1-score | 0,9091 | 0,8894 | 0,9093 | 0,8887 | 0,9097 | 0,8936 | 0,9115 | 0,8971 |
| IoU | 0,5848 | 0,5432 | 0,5845 | 0,5462 | 0,5877 | 0,5327 | 0,5887 | 0,5561 |
| Precisão | 0,9167 | 0,9030 | 0,9161 | 0,9009 | 0,9170 | 0,9096 | 0,9179 | 0,9089 |
| Recall | 0,9017 | 0,8762 | 0,9027 | 0,8769 | 0,9024 | 0,8782 | 0,9052 | 0,8856 |

Tabela 1 – DeepLab: Valores para treino e divisão de dados 70%, 20% e 10%

| Métrica | Histograma equalizado | | | | Transformação morfológica | | | |
|----------|-----------------------|---------|---------------|---------|---------------------------|---------|---------|---------|
| | Adam | | RMSProp | | Adam | | RMSProp | |
| | s/ drop | c/ drop | s/ drop | c/ drop | s/ drop | c/ drop | s/ drop | c/ drop |
| Acurácia | 0,9817 | 0,9685 | 0,9820 | 0,9665 | 0,9809 | 0,9690 | 0,9791 | 0,9714 |
| F1-score | 0,9129 | 0,8994 | 0,9137 | 0,8971 | 0,9111 | 0,8978 | 0,9111 | 0,8972 |
| IoU | 0,5812 | 0,5539 | 0,5822 | 0,5488 | 0,5746 | 0,5479 | 0,5766 | 0,5478 |
| Precisão | 0,9243 | 0,9125 | 0,9250 | 0,9102 | 0,9225 | 0,9120 | 0,9226 | 0,9104 |
| Recall | 0,9016 | 0,8867 | 0,9026 | 0,8844 | 0,9000 | 0,8841 | 0,8999 | 0,8844 |

Tabela 2 – U-Net: Valores para treino e divisão de dados 70%, 20% e 10%

Semelhante ao cenário acima, porém com 80% de dados dedicados ao treino e 10% para as outras etapas, a Tabela 3 mostra a manutenção de algumas vantagens da DeepLab sobre a U-Net, nos índices de sensibilidade e IoU, em que a rede apresenta diferença máxima aproximada de 0.89% e 1.81% pras métricas respectivas, e quando há uso de *dropout*, em todos os índices para quase todas configurações. Quando não há uso de *dropout*, a U-Net se sobressai com pequenas diferenças positivas de até 0.44%, 0.12% e 0.41%, respectivamente, para as métricas de acurácia, *f1-score* e precisão, como mostra a Tabela 4. É importante notar que com esse particionamento de dados, ambas redes usufruíram de melhorias, principalmente nos índices IoU, precisão e sensibilidade, esse benefício pode ser produto

de uma maior capacidade de generalização, aprimorada ao trabalhar maior fatia de dados para treinamento.

| Métrica | Histograma equalizado | | | | Transformação morfológica | | | |
|----------|-----------------------|---------|---------|---------|---------------------------|---------|---------------|---------|
| | Adam | | RMSProp | | Adam | | RMSProp | |
| | s/ drop | c/ drop | s/ drop | c/ drop | s/ drop | c/ drop | s/ drop | c/ drop |
| Acurácia | 0,9788 | 0,9720 | 0,9783 | 0,9702 | 0,9794 | 0,9722 | 0,9773 | 0,9734 |
| F1-score | 0,9266 | 0,9088 | 0,9236 | 0,9091 | 0,9264 | 0,9095 | 0,9280 | 0,9095 |
| IoU | 0,6179 | 0,5825 | 0,6111 | 0,5847 | 0,6189 | 0,5825 | 0,6209 | 0,5810 |
| Precisão | 0,9341 | 0,9230 | 0,9312 | 0,9226 | 0,9345 | 0,9230 | 0,9347 | 0,9223 |
| Recall | 0,9192 | 0,8949 | 0,9161 | 0,8960 | 0,9184 | 0,8965 | 0,9213 | 0,8970 |

Tabela 3 – DeepLab: Valores para treino e divisão de dados 80%, 10% e 10%

| Métrica | Histograma equalizado | | | | Transformação morfológica | | | |
|----------|-----------------------|---------|---------------|---------|---------------------------|---------|---------|---------|
| | Adam | | RMSProp | | Adam | | RMSProp | |
| | s/ drop | c/ drop | s/ drop | c/ drop | s/ drop | c/ drop | s/ drop | c/ drop |
| Acurácia | 0,9826 | 0,9668 | 0,9827 | 0,9673 | 0,9814 | 0,9682 | 0,9803 | 0,9655 |
| F1-score | 0,9250 | 0,9096 | 0,9248 | 0,9077 | 0,9239 | 0,9087 | 0,9235 | 0,9060 |
| IoU | 0,6046 | 0,5712 | 0,6052 | 0,5675 | 0,6033 | 0,5697 | 0,6028 | 0,5641 |
| Precisão | 0,9359 | 0,9224 | 0,9353 | 0,9200 | 0,9356 | 0,9218 | 0,9348 | 0,9192 |
| Recall | 0,9144 | 0,8972 | 0,9146 | 0,8957 | 0,9125 | 0,8960 | 0,9124 | 0,8933 |

Tabela 4 – U-Net: Valores para treino e divisão de dados 80%, 10% e 10%

4.2 Resultados: Etapa de validação

A Tabela 5 e 6 demonstra que durante a etapa de validação, considerando o uso ou não de *dropout*, a U-Net tem majoritariamente os melhores valores para todos os índices, com exceção do índice de interseção sobre união, em que ambas redes apresentam resultados competitivos, contudo a DeepLab tem maior número de valores superiores nessa última métrica. A U-Net chegou a apresentar a acurácia, *f1-score*, IoU, precisão e sensibilidade com diferença superior de até, respectivamente, 1.87%, 0.99%, 1.5%, 0.94% e 1.04%, enquanto a DeepLab entregou valores superiores de até 1.30% e 0.52% para IoU e sensibilidade.

| Métrica | Histograma equalizado | | | | Transformação morfológica | | | |
|----------|-----------------------|---------|---------|---------|---------------------------|---------------|---------------|---------|
| | Adam | | RMSProp | | Adam | | RMSProp | |
| | s/ drop | c/ drop | s/ drop | c/ drop | s/ drop | c/ drop | s/ drop | c/ drop |
| Acurácia | 0,8492 | 0,8348 | 0,8471 | 0,8475 | 0,8429 | 0,8529 | 0,8426 | 0,8480 |
| F1-score | 0,9097 | 0,8901 | 0,9099 | 0,8895 | 0,9103 | 0,8944 | 0,9121 | 0,8978 |
| IoU | 0,5860 | 0,5447 | 0,5857 | 0,5477 | 0,5889 | 0,5343 | 0,5897 | 0,5575 |
| Precisão | 0,9172 | 0,9036 | 0,9166 | 0,9015 | 0,9176 | 0,9102 | 0,9184 | 0,9095 |
| Recall | 0,9024 | 0,8770 | 0,9033 | 0,8778 | 0,9031 | 0,8791 | 0,9058 | 0,8864 |

Tabela 5 – DeepLab: Valores para etapa de validação e divisão de dados 70%, 20% e 10%

| Métrica | Histograma equalizado | | | | Transformação morfológica | | | |
|----------|-----------------------|---------|---------------|---------|---------------------------|---------|---------|---------|
| | Adam | | RMSProp | | Adam | | RMSProp | |
| | s/ drop | c/ drop | s/ drop | c/ drop | s/ drop | c/ drop | s/ drop | c/ drop |
| Acurácia | 0,8597 | 0,8492 | 0,8657 | 0,8620 | 0,8568 | 0,8553 | 0,8600 | 0,8598 |
| F1-score | 0,9134 | 0,9000 | 0,9142 | 0,8978 | 0,9117 | 0,8985 | 0,9117 | 0,8979 |
| IoU | 0,5826 | 0,5554 | 0,5836 | 0,5502 | 0,5759 | 0,5493 | 0,5781 | 0,5494 |
| Precisão | 0,9248 | 0,9130 | 0,9255 | 0,9107 | 0,9229 | 0,9125 | 0,9231 | 0,9109 |
| Recall | 0,9024 | 0,8874 | 0,9033 | 0,8852 | 0,9007 | 0,8849 | 0,9006 | 0,8852 |

Tabela 6 – U-Net: Valores para etapa de validação e divisão de dados 70%, 20% e 10%

Os resultados obtidos durante a fase de validação ratifica as melhorias trazidas com aumento do conjunto de treinamento, seus resultados representados na Tabela 7 e 8 indicam valores substancialmente melhores para ambas redes e para a maioria das configurações. As tabelas também mostram que a U-Net tem melhor desempenho quando consideramos valores de acurácia, em que a rede se destacou com valores superiores com diferença de 3.47%, enquanto a DeepLab entregou resultados com diferença superior de até 0.44%, 1.78%, 0.32% e 0.87% para *f1-score*, IoU, precisão e sensibilidade.

| Métrica | Histograma equalizado | | | | Transformação morfológica | | | |
|----------|-----------------------|---------|---------|---------|---------------------------|---------|---------------|---------|
| | Adam | | RMSProp | | Adam | | RMSProp | |
| | s/ drop | c/ drop | s/ drop | c/ drop | s/ drop | c/ drop | s/ drop | c/ drop |
| Acurácia | 0,8957 | 0,8798 | 0,8914 | 0,8780 | 0,8917 | 0,8717 | 0,8881 | 0,8710 |
| F1-score | 0,9272 | 0,9095 | 0,9243 | 0,9098 | 0,9270 | 0,9103 | 0,9286 | 0,9102 |
| IoU | 0,6192 | 0,5841 | 0,6126 | 0,5863 | 0,6203 | 0,5840 | 0,6221 | 0,5826 |
| Precisão | 0,9347 | 0,9237 | 0,9318 | 0,9232 | 0,9350 | 0,9236 | 0,9352 | 0,9229 |
| Recall | 0,9199 | 0,8958 | 0,9168 | 0,8969 | 0,9192 | 0,8974 | 0,9220 | 0,8979 |

Tabela 7 – DeepLab: Valores para etapa de validação e divisão de dados 80%, 10% e 10%

| Métrica | Histograma equalizado | | | | Transformação morfológica | | | |
|----------|-----------------------|---------|---------------|---------|---------------------------|---------|---------|---------|
| | Adam | | RMSProp | | Adam | | RMSProp | |
| | s/ drop | c/ drop | s/ drop | c/ drop | s/ drop | c/ drop | s/ drop | c/ drop |
| Acurácia | 0,9082 | 0,9064 | 0,9074 | 0,9038 | 0,9023 | 0,9064 | 0,8986 | 0,8980 |
| F1-score | 0,9257 | 0,9104 | 0,9255 | 0,9085 | 0,9247 | 0,9095 | 0,9242 | 0,9068 |
| IoU | 0,6062 | 0,5728 | 0,6067 | 0,5692 | 0,6049 | 0,5713 | 0,6043 | 0,5658 |
| Precisão | 0,9365 | 0,9230 | 0,9359 | 0,9206 | 0,9362 | 0,9224 | 0,9354 | 0,9198 |
| Recall | 0,9152 | 0,8981 | 0,9154 | 0,8966 | 0,9134 | 0,8969 | 0,9132 | 0,8941 |

Tabela 8 – U-Net: Valores para etapa de validação e divisão de dados 80%, 10% e 10%

4.3 Resultados: Etapa de teste

Na etapa de teste, seus resultados representados nas Tabelas 9 e 10, demonstram que com exceção de poucos casos (quando há uso do otimizador Adam e de *dropout*), a U-Net apresentou diferenças positivas substanciais de até 3.4%, 0.99%, 1.51%, 0.94% e 1.04% para respectivamente, acurácia, *f1-score*, IoU, precisão e sensibilidade. É possível notar que a U-Net se destaca com o conjunto de dados em que o histograma foi equalizado, e a DeepLab, com o que sofreu transformações morfológicas, apresentando uma quantidade maior de valores melhores do que os obtidos pela U-Net na métrica IoU.

| Métrica | Histograma equalizado | | | | Transformação morfológica | | | |
|----------|-----------------------|---------|---------|---------|---------------------------|---------|---------------|---------|
| | Adam | | RMSProp | | Adam | | RMSProp | |
| | s/ drop | c/ drop | s/ drop | c/ drop | s/ drop | c/ drop | s/ drop | c/ drop |
| Acurácia | 0,8971 | 0,9136 | 0,9185 | 0,9033 | 0,9072 | 0,9047 | 0,9188 | 0,8918 |
| F1-score | 0,9095 | 0,8900 | 0,9097 | 0,8893 | 0,9101 | 0,8942 | 0,9119 | 0,8977 |
| IoU | 0,5855 | 0,5442 | 0,5852 | 0,5474 | 0,5884 | 0,5341 | 0,5894 | 0,5571 |
| Precisão | 0,9170 | 0,9034 | 0,9165 | 0,9013 | 0,9174 | 0,9101 | 0,9182 | 0,9094 |
| Recall | 0,9022 | 0,8769 | 0,9031 | 0,8776 | 0,9029 | 0,8789 | 0,9057 | 0,8862 |

Tabela 9 – DeepLab: Valores para etapa de teste e divisão de dados 70%, 20% e 10%

| Métrica | Histograma equalizado | | | | Transformação morfológica | | | |
|----------|-----------------------|---------|---------------|---------|---------------------------|---------|---------|---------|
| | Adam | | RMSProp | | Adam | | RMSProp | |
| | s/ drop | c/ drop | s/ drop | c/ drop | s/ drop | c/ drop | s/ drop | c/ drop |
| Acurácia | 0,9311 | 0,9296 | 0,9276 | 0,9307 | 0,9197 | 0,9272 | 0,9260 | 0,9224 |
| F1-score | 0,9133 | 0,8999 | 0,9141 | 0,8976 | 0,9115 | 0,8984 | 0,9116 | 0,8979 |
| IoU | 0,5824 | 0,5553 | 0,5834 | 0,5501 | 0,5756 | 0,5492 | 0,5778 | 0,5494 |
| Precisão | 0,9246 | 0,9128 | 0,9253 | 0,9105 | 0,9228 | 0,9123 | 0,9229 | 0,9109 |
| Recall | 0,9023 | 0,8873 | 0,9032 | 0,8851 | 0,9006 | 0,8848 | 0,9005 | 0,8853 |

Tabela 10 – U-Net: Valores para etapa de teste e divisão de dados 70%, 20% e 10%

Com um conjunto maior para treino, os resultados (exibidos nas Tabela 11 e 12) da etapa de teste para a métrica de precisão foram bem próximos para ambas redes, enquanto que acurácia, quando obtida pela U-Net apresentou uma diferença superior de até aproximadamente 4.67%, do que os valores obtidos pela DeepLab. Para os valores de *f1-score*, IoU e sensibilidade, a DeepLab destacou-se por pequenos ganhos e atingiu diferenças superiores de até cerca 0.44%, 1.76% e 0.87% para as métricas citadas.

| Métrica | Histograma equalizado | | | | Transformação morfológica | | | |
|----------|-----------------------|---------|---------------|---------|---------------------------|---------|---------------|---------|
| | Adam | | RMSProp | | Adam | | RMSProp | |
| | s/ drop | c/ drop | s/ drop | c/ drop | s/ drop | c/ drop | s/ drop | c/ drop |
| Acurácia | 0,8893 | 0,9072 | 0,9183 | 0,8965 | 0,9063 | 0,8717 | 0,9126 | 0,8849 |
| F1-score | 0,9272 | 0,9095 | 0,9242 | 0,9098 | 0,9269 | 0,9102 | 0,9285 | 0,9101 |
| IoU | 0,6188 | 0,5841 | 0,6125 | 0,5860 | 0,6202 | 0,5837 | 0,6219 | 0,5825 |
| Precisão | 0,9346 | 0,9236 | 0,9317 | 0,9231 | 0,9349 | 0,9235 | 0,9352 | 0,9228 |
| Recall | 0,9199 | 0,8958 | 0,9168 | 0,8968 | 0,9191 | 0,8973 | 0,9219 | 0,8978 |

Tabela 11 – DeepLab: Valores para etapa de teste e divisão de dados 80%, 10% e 10%

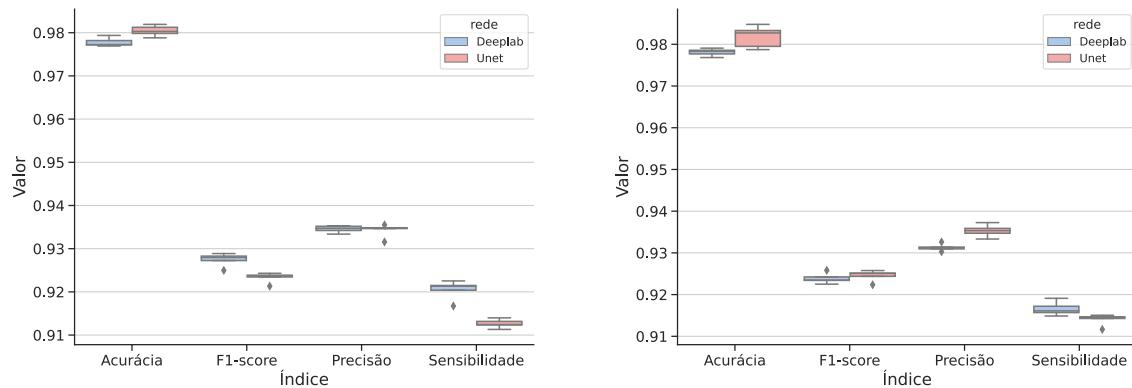
| Métrica | Histograma equalizado | | | | Transformação morfológica | | | |
|----------|-----------------------|---------|---------------|---------------|---------------------------|---------|---------|---------|
| | Adam | | RMSProp | | Adam | | RMSProp | |
| | s/ drop | c/ drop | s/ drop | c/ drop | s/ drop | c/ drop | s/ drop | c/ drop |
| Acurácia | 0,9243 | 0,9278 | 0,9260 | 0,9304 | 0,9209 | 0,9184 | 0,9231 | 0,9267 |
| F1-score | 0,9257 | 0,9103 | 0,9255 | 0,9085 | 0,9246 | 0,9095 | 0,9241 | 0,9068 |
| IoU | 0,6062 | 0,5729 | 0,6067 | 0,5692 | 0,6048 | 0,5714 | 0,6043 | 0,5657 |
| Precisão | 0,9364 | 0,9229 | 0,9358 | 0,9206 | 0,9362 | 0,9224 | 0,9353 | 0,9197 |
| Recall | 0,9152 | 0,8981 | 0,9154 | 0,8967 | 0,9134 | 0,8969 | 0,9132 | 0,8942 |

Tabela 12 – U-Net: Valores para etapa de teste e divisão de dados 80%, 10% e 10%

Após observar os resultados das redes para todos os cenários, nota-se que os melhores valores obtidos pela U-Net foram obtidos, quase que integralmente, com uso do conjunto de imagens de histograma equalizado, enquanto as melhores saídas geradas pela DeepLab decorreram predominantemente do uso do conjunto com transformações morfológicas aplicadas, ambas RNAs concentraram a maioria dos melhores valores (destacados em negrito) quando o otimizador RMSProp foi utilizado. As Figuras 32, 33 e 34 mostram, com exceção do IoU médio (descartado por não apresentar variação relevante), os resultados de ambas redes em seus melhores cenários com o conjunto de dados particionado em 80%, 10%, 10%, para as etapas, respectivamente, de treino, validação e teste.

Figura 32 – Melhores cenários para etapa de treino

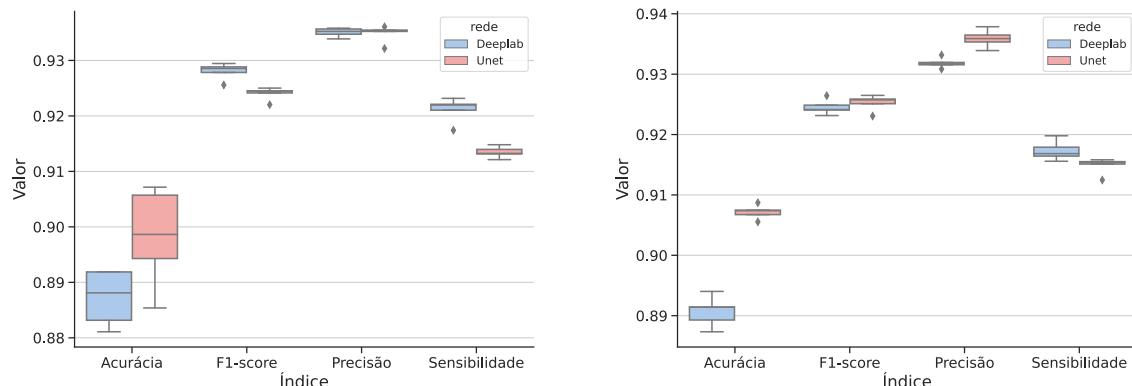
(a) Uso de transformações morfológicas (b) Uso de equalização de histograma



Fonte: Elaborada pelo autor (2021)

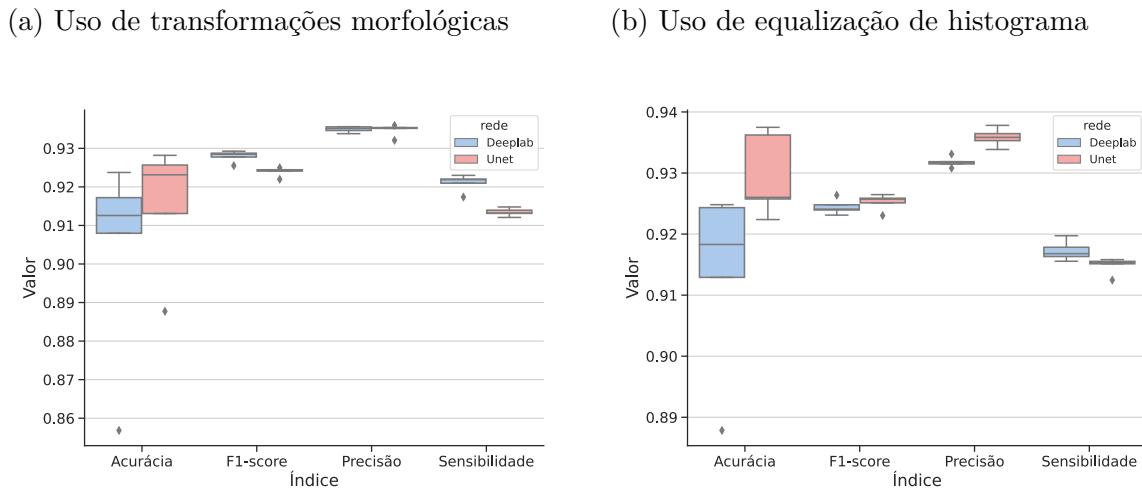
Figura 33 – Melhores cenários para etapa de validação

(a) Uso de transformações morfológicas (b) Uso de equalização de histograma



Fonte: Elaborada pelo autor (2021)

Figura 34 – Melhores cenários para etapa de teste



Fonte: Elaborada pelo autor (2021)

5 CONSIDERAÇÕES FINAIS

No decorrer deste trabalho realizamos uma investigação sobre o uso da técnica de segmentação semântica, com duas redes neurais artificiais consideradas estado da arte nessa tarefa, como método complementar para auxiliar profissionais de saúde na identificação da causa de manifestações de casos de pneumonia.

De posse da revisão de literatura sobre a técnica supracitada, comprehende-se a relevância e o amplo uso desse processo nos últimos anos para o prognóstico de patologias e em tarefas que demandam não só a classificação de elementos contidos em uma imagem mas também o realce de regiões de interesse para realização de análises e tratamentos mais intensos, como a deterioração de estruturas em uma construção, áreas de desmatamento e alagamento, por exemplo. Durante a realização de tal revisão foi encontrado uma quantidade significativa de trabalhos publicados no ano de 2020 que propusessem o uso de tal tipo de segmentação para auxiliar na tarefa de identificar agentes causadores de pneumonia e priorizar recursos, como atendimentos e infraestrutura, para casos do novo coronavírus (sars-cov 2), dessa forma, atribuindo à técnica o papel de uma ferramenta de suporte à tomada de decisão em situação de pandemia.

Os experimentos computacionais realizados mostraram que ambas redes neurais apresentaram resultados compatíveis com os obtidos por trabalhos semelhantes, isto é, pesquisas que investigaram o uso de segmentação semântica aplicada no prognóstico de doenças. Tais experimentos mostraram que os resultados obtidos pela rede U-Net foram majoritariamente maiores do que a outra rede em cenários que o conjunto de dados foi particionado em 70%, 20% e 10% para respectivamente, treino, validação e teste, e quando houve equalização de histograma, enquanto que a DeepLab apresentou resultados mais competitivos com o conjunto que passou por transformações morfológicas e que foi particionado em 80%, 10% e 10%. Vale ressaltar que as duas redes se beneficiaram com o aumento da quantidade de dados para treino e do uso do otimizador RMSProp.

Com a revisão de trabalhos relacionados foi possível obter um conjunto de valores para parâmetros e hiperparâmetros como base, já o trabalho colaborativo e aberto à comunidade possibilitou por meio de ferramentas como o Github e Darwin da V7Labs, ter acesso ao código-fonte de ambas redes e ao conjunto de dados, o que foi fundamental para a conquista dos objetivos estabelecidos no início deste trabalho.

Em suma, esse trabalho apresenta as seguintes contribuições: Desenvolvimento de um sistema baseado em duas arquiteturas de redes neurais que prevê o diagnóstico de pneumonia causada por três tipos de agentes, o uso das técnicas de morfologia matemática e equalização de histograma, análise quantitativa e qualitativa em função dos diferentes parâmetros e hiperparâmetros (a saber, otimizadores, emprego de *dropout* e particionamento

de dados), apresentação de uma ampla revisão de literatura sobre o uso dessa técnica em aplicações diversas.

5.1 Trabalhos futuros

Tendo em vista a limitação imposta pelos recursos computacionais disponíveis para a execução dos experimentos, consequentemente impactos em decisões de implementação e o custo intrínsecos de alguns valores de parâmetros e hiperparâmetros, estabelecemos que com o aumento da oferta de tais recursos, as seguintes ambições para futuros estudos devem ser atendidas:

- Uso de imagens de resoluções superiores a 512 *pixels* por altura e largura;
- Uso de *backbones* como Xception para DeepLab e investigação de outros como variantes da família ResNet, VGGNet e Inception;
- Ajuste automático de hiperparâmetros;
- Execução com maior número de épocas e uso de outras métricas populares em trabalhos relacionados;
- Aplicação de outras transformações no conjunto de dados;
- Uso de redes com arquiteturas desenvolvidas ou adaptadas para resolução de problemas de segmentação semântica mais específicos, como a CovidNet.

REFERÊNCIAS

- ABRAMOFF, M. D.; GARVIN, M. K.; SONKA, M. Retinal imaging and image analysis. *IEEE Reviews in Biomedical Engineering*, Institute of Electrical and Electronics Engineers (IEEE), v. 3, p. 169–208, 2010. Disponível em: <<https://doi.org/10.1109/rbme.2010.2084567>>.
- AKKUS, Z. et al. Deep learning for brain mri segmentation: state of the art and future directions. *Journal of digital imaging*, Springer, v. 30, n. 4, p. 449–459, 2017.
- ALIPER, A. et al. Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data. *Molecular Pharmaceutics*, American Chemical Society (ACS), v. 13, n. 7, p. 2524–2530, jun. 2016. Disponível em: <<https://doi.org/10.1021/acs.molpharmaceut.6b00248>>.
- ANDERSON, J. *Cognitive Skills and Their Acquisition*. Taylor & Francis, 2013. (Carnegie Mellon Symposia on Cognition Series). ISBN 9781135830953. Disponível em: <<https://books.google.com.br/books?id=PDzFBQAAQBAJ>>.
- BADRINARAYANAN, V.; KENDALL, A.; CIPOLLA, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 39, n. 12, p. 2481–2495, 2017.
- BOUREAU, Y.-L.; PONCE, J.; LECUN, Y. A theoretical analysis of feature pooling in visual recognition. In: *Proceedings of the 27th International Conference on International Conference on Machine Learning*. Madison, WI, USA: Omnipress, 2010. (ICML'10), p. 111–118. ISBN 9781605589077. Disponível em: <<https://www.di.ens.fr/willow/pdfs/icml2010b.pdf>>.
- BURGOS-ARTIZZU, X. P. Computer-aided covid-19 patient screening using chest images (x-ray and ct scans). *medRxiv*, Cold Spring Harbor Laboratory Press, 2020. Disponível em: <<https://www.medrxiv.org/content/early/2020/07/17/2020.07.16.20155093>>.
- CHEN, L. et al. Drinet for medical image segmentation. *IEEE Transactions on Medical Imaging*, v. 37, n. 11, p. 2453–2462, 2018.
- CHEN, L.-C. et al. Semantic image segmentation with deep convolutional nets and fully connected crfs. 2014.
- CHEN, L.-C. et al. *DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs*. 2016.
- CHEN, L.-C. et al. *Rethinking Atrous Convolution for Semantic Image Segmentation*. 2017.
- CHEN, L.-C. et al. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: FERRARI, V. et al. (Ed.). *Computer Vision – ECCV 2018*. Cham: Springer International Publishing, 2018. p. 833–851. ISBN 978-3-030-01234-2.
- CHEN, S.-D.; RAMLI, A. R. Contrast enhancement using recursive mean-separate histogram equalization for scalable brightness preservation. *IEEE Transactions on Consumer Electronics*, v. 49, n. 4, p. 1301–1309, 2003.

- CHEN, X.; LIN, X. Big data deep learning: Challenges and perspectives. *IEEE Access*, v. 2, p. 514–525, 2014.
- CIRESAN, D. C. et al. Deep neural networks segment neuronal membranes in electron microscopy images. In: *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 2*. Red Hook, NY, USA: Curran Associates Inc., 2012. (NIPS'12), p. 2843–2851.
- COLMAN, J. et al. Dr-unet104 for multimodal mri brain tumor segmentation. Springer International Publishing, Cham, p. 410–419, 2021.
- COSTEA, A.; PETROVAI, A.; NEDEVSCHI, S. Fusion scheme for semantic and instance-level segmentation. In: . [S.l.: s.n.], 2018. p. 3469–3475.
- CRAMER, S. C. et al. Harnessing neuroplasticity for clinical applications. *Brain*, Oxford University Press (OUP), v. 134, n. 6, p. 1591–1609, abr. 2011. Disponível em: <<https://doi.org/10.1093/brain/awr039>>.
- DADONAITÉ, B. Pneumonia. *Our World in Data*, 2018. <Https://ourworldindata.org/pneumonia>.
- DENG, D. et al. *PixelLink: Detecting Scene Text via Instance Segmentation*. 2018. Disponível em: <<https://arxiv.org/abs/1801.01315>>.
- DENG, L.; YU, D. Deep learning: Methods and applications. *Foundations and Trends® in Signal Processing*, v. 7, n. 3–4, p. 197–387, 2014. ISSN 1932-8346. Disponível em: <<http://dx.doi.org/10.1561/2000000039>>.
- DENG, Z. et al. Segmentation of dermoscopy images based on fully convolutional neural network. In: *2017 IEEE International Conference on Image Processing (ICIP)*. [S.l.: s.n.], 2017. p. 1732–1736.
- DION, C. F.; ASHURST, J. V. *Streptococcus Pneumoniae*. Treasure Island (FL): StatPearls Publishing, 2020. Disponível em: <<http://www.ncbi.nlm.nih.gov/books/NBK470537/>>.
- DOLZ, J.; DESROSIERS, C.; AYED, I. B. 3d fully convolutional networks for subcortical segmentation in mri: A large-scale study. *NeuroImage*, Elsevier BV, v. 170, p. 456–470, Apr 2018. ISSN 1053-8119. Disponível em: <<http://dx.doi.org/10.1016/j.neuroimage.2017.04.039>>.
- DOLZ, J.; DESROSIERS, C.; AYED, I. B. Ivd-net: Intervertebral disc localization and segmentation in mri with a multi-modal unet. In: ZHENG, G. et al. (Ed.). *Computational Methods and Clinical Applications for Spine Imaging*. Cham: Springer International Publishing, 2019. p. 130–143. ISBN 978-3-030-13736-6.
- DONG, X. et al. Automatic multiorgan segmentation in thoraxCTimages using u-net-GAN. *Medical Physics*, Wiley, v. 46, n. 5, p. 2157–2168, mar. 2019. Disponível em: <<https://doi.org/10.1002/mp.13458>>.
- EL-BANA, S.; AL-KABBANY, A.; SHARKAS, M. A multi-task pipeline with specialized streams for classification and segmentation of infection manifestations in covid-19 scans. *medRxiv*, Cold Spring Harbor Laboratory Press, 2020. Disponível em: <<https://www.medrxiv.org/content/early/2020/06/26/2020.06.24.20139238>>.

- ELKAHKY, A. M.; SONG, Y.; HE, X. A multi-view deep learning approach for cross domain user modeling in recommendation systems. In: *Proceedings of the 24th International Conference on World Wide Web*. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, 2015. (WWW '15), p. 278–288. ISBN 9781450334693. Disponível em: <<https://doi.org/10.1145/2736277.2741667>>.
- ESTRADA, F. J.; JEPSON, A. D. Benchmarking image segmentation algorithms. *International Journal of Computer Vision*, Springer Science and Business Media LLC, v. 85, n. 2, p. 167–181, maio 2009. Disponível em: <<https://doi.org/10.1007/s11263-009-0251-z>>.
- FANSELOW, M. S.; POULOS, A. M. The neuroscience of mammalian associative learning. *Annual Review of Psychology*, Annual Reviews, v. 56, n. 1, p. 207–234, fev. 2005. Disponível em: <<https://doi.org/10.1146/annurev.psych.56.091103.070213>>.
- FELZENSZWALB, P. F. et al. Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 32, n. 9, p. 1627–1645, 2010.
- FERNANDES, M. A. C. Redes neurais artificiais aplicadas à detecção inteligente de sinais. Unpublished, 1999. Disponível em: <<http://rgdoi.net/10.13140/RG.2.2.17177.93281>>.
- FERNÁNDEZ-CABÁN, P. L.; MASTERS, F. J.; PHILLIPS, B. M. Predicting roof pressures on a low-rise structure from freestream turbulence using artificial neural networks. *Frontiers in Built Environment*, v. 4, 11 2018.
- FIORANI, F.; SCHURR, U. Future scenarios for plant phenotyping. *Annual Review of Plant Biology*, Annual Reviews, v. 64, n. 1, p. 267–291, abr. 2013. Disponível em: <<https://doi.org/10.1146/annurev-aplant-050312-120137>>.
- FLOOD, N.; WATSON, F.; COLLETT, L. Using a u-net convolutional neural network to map woody vegetation extent from high resolution satellite imagery across queensland, australia. *International Journal of Applied Earth Observation and Geoinformation*, v. 82, p. 101897, 2019. ISSN 0303-2434. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0303243419302041>>.
- FRANCHINI, G.; BURGIO, P.; ZANNI, L. Artificial neural networks: The missing link between curiosity and accuracy. In: *Advances in Intelligent Systems and Computing*. Springer International Publishing, 2019. p. 1025–1034. Disponível em: <https://doi.org/10.1007/978-3-030-16660-1_100>.
- FU, H. et al. Retinal vessel segmentation via deep learning network and fully-connected conditional random fields. In: *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*. [S.l.: s.n.], 2016. p. 698–701.
- FU, J. et al. Densely connected deconvolutional network for semantic segmentation. In: *2017 IEEE International Conference on Image Processing (ICIP)*. [S.l.: s.n.], 2017. p. 3085–3089.
- GIANNAKERIS, P. et al. People and vehicles in danger - a fire and flood detection system in social media. In: *2018 IEEE 13th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*. [S.l.: s.n.], 2018. p. 1–5.

- GODIN, F. et al. Dual rectified linear units (drelus): A replacement for tanh activation functions in quasi-recurrent neural networks. *Pattern Recognition Letters*, Elsevier BV, v. 116, p. 8–14, Dec 2018. ISSN 0167-8655. Disponível em: <<http://dx.doi.org/10.1016/j.patrec.2018.09.006>>.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: MIT Press, 2016. <<http://www.deeplearningbook.org>>.
- GWET, D. L. L. et al. A review on image segmentation techniques and performance measures. *International Journal of Computer and Information Engineering*, World Academy of Science, Engineering and Technology, v. 12, n. 12, p. 1107 – 1117, 2018. ISSN eISSN: 1307-6892. Disponível em: <<https://publications.waset.org/vol/144>>.
- HAYKIN, S. *Redes Neurais: Princípios e Prática*. Artmed, 2007. ISBN 9788577800865. Disponível em: <<https://books.google.com.br/books?id=bhMwDwAAQBAJ>>.
- HE, K. et al. Hf-unet: Learning hierarchically inter-task relevance in multi-task u-net for accurate prostate segmentation. *arXiv preprint arXiv:2005.10439*, 2020.
- HIRSCHMAN, I.; WIDDER, D. *The Convolution Transform*. Dover Publications, 2012. (Dover Books on Mathematics). ISBN 9780486154565. Disponível em: <<https://books.google.com.br/books?id=TAqbPjWNQEJC>>.
- HOARE, Z.; LIM, W. S. Pneumonia: update on diagnosis and management. *BMJ*, BMJ, v. 332, n. 7549, p. 1077–1079, maio 2006. Disponível em: <<https://doi.org/10.1136/bmj.332.7549.1077>>.
- HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. *Neural Computation*, MIT Press - Journals, v. 9, n. 8, p. 1735–1780, nov. 1997. Disponível em: <<https://doi.org/10.1162/neco.1997.9.8.1735>>.
- HUBEL, D. H.; WIESEL, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, Wiley, v. 160, n. 1, p. 106–154, jan. 1962. Disponível em: <<https://doi.org/10.1113/jphysiol.1962.sp006837>>.
- HUVAL, B. et al. An empirical evaluation of deep learning on highway driving. *CoRR*, abs/1504.01716, 2015. Disponível em: <<http://arxiv.org/abs/1504.01716>>.
- 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>>.
- JERANT, A. et al. Early detection and treatment of skin cancer. *American Family Physician*, American Academy of Family Physicians, v. 62, n. 2, jul. 2000. ISSN 0002-838X.
- JI, L.; PIPER, J.; TANG, J.-Y. Erosion and dilation of binary images by arbitrary structuring elements using interval coding. *Pattern Recognition Letters*, v. 9, n. 3, p. 201–209, 1989. ISSN 0167-8655. Disponível em: <<https://www.sciencedirect.com/science/article/pii/016786558990055X>>.
- JOSEPH, N. M. et al. Ventilator-associated pneumonia: A review. *European Journal of Internal Medicine*, Elsevier BV, v. 21, n. 5, p. 360–368, out. 2010. Disponível em: <<https://doi.org/10.1016/j.ejim.2010.07.006>>.

- KAELBLING, L. P.; LITTMAN, M. L.; MOORE, A. W. Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, AI Access Foundation, v. 4, p. 237–285, maio 1996. Disponível em: <<https://doi.org/10.1613/jair.301>>.
- KERMANY, D. *Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification*. Mendeley, 2018. Disponível em: <<https://data.mendeley.com/datasets/rscbjbr9sj/2>>.
- KINGMA, D. P.; BA, J. Adam: A method for stochastic optimization. In: BENGIO, Y.; LECUN, Y. (Ed.). *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*. [s.n.], 2015. Disponível em: <<http://arxiv.org/abs/1412.6980>>.
- KIRILLOV, A. et al. Panoptic segmentation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. [s.n.], 2019. Disponível em: <https://openaccess.thecvf.com/content_CVPR_2019/html/Kirillov_Panoptic_Segmentation_CVPR_2019_paper.html>.
- KOSMAN, E.; LEONARD, K. J. Similarity coefficients for molecular markers in studies of genetic relationships between individuals for haploid, diploid, and polyploid species. *Molecular Ecology*, v. 14, n. 2, p. 415–424, 2005. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-294X.2005.02416.x>>.
- KRÄHENBÜHL, P.; KOLTUN, V. *Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials*. 2012.
- KUSHOL, R. et al. Contrast enhancement by top-hat and bottom-hat transform with optimal structuring element: Application to retinal vessel segmentation. In: KARRAY, F.; CAMPILHO, A.; CHERIET, F. (Ed.). *Image Analysis and Recognition*. Cham: Springer International Publishing, 2017. p. 533–540. ISBN 978-3-319-59876-5.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *Nature*, v. 521, p. 436–44, 05 2015.
- LECUN, Y. et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, v. 86, n. 11, p. 2278–2324, Nov 1998. ISSN 1558-2256.
- LI, C.; LEE, C. Minimum cross entropy thresholding. *Pattern Recognition*, v. 26, n. 4, p. 617 – 625, 1993. ISSN 0031-3203. Disponível em: <<http://www.sciencedirect.com/science/article/pii/003132039390115D>>.
- LIN, T.-Y. et al. Microsoft coco: Common objects in context. In: FLEET, D. et al. (Ed.). *Computer Vision – ECCV 2014*. Cham: Springer International Publishing, 2014. p. 740–755. ISBN 978-3-319-10602-1.
- LINGENFELSER, F. et al. Asynchronous and event-based fusion systems for affect recognition on naturalistic data in comparison to conventional approaches. *IEEE Transactions on Affective Computing*, PP, p. 1–1, 12 2016.
- LIU, S. et al. *Path Aggregation Network for Instance Segmentation*. 2018. Disponível em: <<https://arxiv.org/abs/1803.01534>>.

- LOTTER, W.; SORENSEN, G.; COX, D. A multi-scale CNN and curriculum learning strategy for mammogram classification. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer International Publishing, 2017. p. 169–177. Disponível em: <https://doi.org/10.1007/978-3-319-67558-9_20>.
- MASAD, E.; BUTTON, J. W.; PAPAGIANNAKIS, T. Fine-aggregate angularity: Automated image analysis approach. *Transportation Research Record*, v. 1721, n. 1, p. 66–72, 2000. Disponível em: <<https://doi.org/10.3141/1721-08>>.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, Springer Science and Business Media LLC, v. 5, n. 4, p. 115–133, dez. 1943. Disponível em: <<https://doi.org/10.1007/bf02478259>>.
- MEHROTRA, A. et al. Blood vessel extraction for retinal images using morphological operator and kcn clustering. In: *2014 IEEE International Advance Computing Conference (IACC)*. [S.l.: s.n.], 2014. p. 1142–1146.
- MILLETARI, F.; NAVAB, N.; AHMADI, S.-A. *V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation*. 2016.
- MOHAGHEGHI, S.; FORUZAN, A. H.; CHEN, Y.-W. Improving the performance of deep cnns in medical image segmentation with limited resources. In: _____. *Deep Learning in Healthcare: Paradigms and Applications*. Cham: Springer International Publishing, 2020. p. 79–94. ISBN 978-3-030-32606-7. Disponível em: <https://doi.org/10.1007/978-3-030-32606-7_5>.
- MOHSENZADEH, Y. et al. Emergence of visual center-periphery spatial organization in deep convolutional neural networks. *Scientific Reports*, Springer Science and Business Media LLC, v. 10, n. 1, mar. 2020. Disponível em: <<https://doi.org/10.1038/s41598-020-61409-0>>.
- MORALES, G. et al. Automatic segmentation of mauritia flexuosa in unmanned aerial vehicle (UAV) imagery using deep learning. *Forests*, MDPI AG, v. 9, n. 12, p. 736, nov. 2018. Disponível em: <<https://doi.org/10.3390/f9120736>>.
- MOTALEBI, S. *Analysis of Cosmic Microwave Background maps containing different levels of local non-Gaussianities using DeepSphere*. Tese (Doutorado), 02 2020. Disponível em: <https://www.researchgate.net/publication/339412541_Analysis_of_Cosmic_Microwave_Background_mapsContaining_different_levels_of_local_non-Gaussianities_using_DeepSphere>.
- MURPHY, K. *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012. (Adaptive Computation and Machine Learning series). ISBN 9780262018029. Disponível em: <<https://books.google.com.br/books?id=NZP6AQAAQBAJ>>.
- MYERS, A. et al. Im2calories: Towards an automated mobile vision food diary. In: *2015 IEEE International Conference on Computer Vision (ICCV)*. [S.l.: s.n.], 2015. p. 1233–1241.
- NAIR, G. B.; NIEDERMAN, M. S. Community-acquired pneumonia: An unfinished battle. *Medical Clinics of North America*, Elsevier BV, v. 95, n. 6, p. 1143–1161, nov. 2011. Disponível em: <<https://doi.org/10.1016/j.mcna.2011.08.007>>.

- NGUYEN, T. et al. *U-Net for MAV-based Penstock Inspection: an Investigation of Focal Loss in Multi-class Segmentation for Corrosion Identification*. 2018.
- NWANKPA, C. et al. *Activation Functions: Comparison of trends in Practice and Research for Deep Learning*. 2018.
- OH, Y.; PARK, S.; YE, J. C. Deep learning covid-19 features on cxr using limited training data sets. *IEEE Transactions on Medical Imaging*, v. 39, n. 8, p. 2688–2700, 2020.
- ORTQVIST, A.; HEDLUND, J.; KALIN, M. Streptococcus pneumoniae: epidemiology, risk factors, and clinical features. *Seminars in respiratory and critical care medicine*, v. 26, n. 6, p. 563—574, December 2005. ISSN 1069-3424. Disponível em: <<https://doi.org/10.1055/s-2005-925523>>.
- PAN, L. et al. Clinical characteristics of COVID-19 patients with digestive symptoms in hubei, china. *The American Journal of Gastroenterology*, Ovid Technologies (Wolters Kluwer Health), v. 115, n. 5, p. 766–773, maio 2020. Disponível em: <<https://doi.org/10.14309/ajg.00000000000000620>>.
- PEREDA, A. E. Electrical synapses and their functional interactions with chemical synapses. *Nature Reviews Neuroscience*, Springer Science and Business Media LLC, v. 15, n. 4, p. 250–263, mar. 2014. Disponível em: <<https://doi.org/10.1038/nrn3708>>.
- RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015. Disponível em: <<http://arxiv.org/abs/1505.04597>>.
- RUDER, S. *An overview of gradient descent optimization algorithms*. 2017.
- RUUSKANEN, O. et al. Viral pneumonia. *The Lancet*, Elsevier BV, v. 377, n. 9773, p. 1264–1275, abr. 2011. Disponível em: <[https://doi.org/10.1016/s0140-6736\(10\)61459-6](https://doi.org/10.1016/s0140-6736(10)61459-6)>.
- SAMUEL, A. L. Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, v. 3, n. 3, p. 210–229, 1959.
- SHUELL, T. J. Cognitive conceptions of learning. *Review of Educational Research*, v. 56, n. 4, p. 411–436, 1986. Disponível em: <<https://doi.org/10.3102/00346543056004411>>.
- SIDHU, T. S.; SINGH, H.; SACHDEV, M. S. Design, implementation and testing of an artificial neural network based fault direction discriminator for protecting transmission lines. *IEEE Transactions on Power Delivery*, v. 10, n. 2, p. 697–706, 1995.
- SMITH, A. G. et al. Segmentation of roots in soil with u-net. *Plant Methods*, Springer Science and Business Media LLC, v. 16, n. 1, fev. 2020. Disponível em: <<https://doi.org/10.1186/s13007-020-0563-0>>.
- SONKA, M.; HLAVAC, V.; BOYLE, R. *Image Processing, Analysis and Machine Vision*. Springer US, 2013. (Chapman and Hall computing series). ISBN 9781489932167. Disponível em: <<https://books.google.com.br/books?id=yA7yBwAAQBAJ>>.
- STERNBERG, G. *A Fatal Form of Septicaemia in the Rabbit Produced by the Subcutaneous Injection of Human Saliva: An Experimental Research*. John Murphy & Company, 1881. Disponível em: <<https://books.google.com.br/books?id=eEtGjwEACAAJ>>.

- SZEGEDY, C. et al. Rethinking the inception architecture for computer vision. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2016. p. 2818–2826. ISSN 1063-6919.
- TOKMAKOV, P.; SCHMID, C.; ALAHARI, K. Learning to segment moving objects. *International Journal of Computer Vision*, Springer Science and Business Media LLC, v. 127, n. 3, p. 282–301, set. 2018. Disponível em: <<https://doi.org/10.1007/s11263-018-1122-2>>.
- VESAL, S.; RAVIKUMAR, N.; MAIER, A. SkinNet: A deep learning framework for skin lesion segmentation. In: *2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC)*. IEEE, 2018. Disponível em: <<https://doi.org/10.1109/nssmic.2018.8824732>>.
- WAN, T. et al. Accurate segmentation of overlapping cells in cervical cytology with deep convolutional neural networks. *Neurocomputing*, v. 365, p. 157 – 170, 2019. ISSN 0925-2312. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0925231219310008>>.
- WANG, C. et al. *A two-stage 3D Unet framework for multi-class segmentation on full resolution image*. 2018.
- WANG, C. et al. Dense u-net based on patch-based learning for retinal vessel segmentation. *Entropy*, v. 21, n. 2, 2019. ISSN 1099-4300. Disponível em: <<https://www.mdpi.com/1099-4300/21/2/168>>.
- WANG, J. J. et al. Handwritten-digit recognition by hybrid convolutional neural network based on HfO₂ memristive spiking-neuron. *Scientific Reports*, Springer Science and Business Media LLC, v. 8, n. 1, ago. 2018. Disponível em: <<https://doi.org/10.1038/s41598-018-30768-0>>.
- WANG, L. et al. Probabilistic maps of visual topography in human cortex. *Cerebral Cortex*, Oxford University Press (OUP), v. 25, n. 10, p. 3911–3931, dez. 2014. Disponível em: <<https://doi.org/10.1093/cercor/bhu277>>.
- WRIGHT, J. et al. Sparse representation for computer vision and pattern recognition. *Proceedings of the IEEE*, v. 98, n. 6, p. 1031–1044, 2010.
- WU, D. et al. Detection of the respiratory rate of standing cows by combining the deeplab v3+ semantic segmentation model with the phase-based video magnification algorithm. *Biosystems Engineering*, v. 192, p. 72 – 89, 2020. ISSN 1537-5110. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1537511020300246>>.
- XIAO, W.; CHANG, L.; LIU, W. Semantic segmentation of colorectal polyps with deeplab and lstm networks. In: *2018 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)*. [S.l.: s.n.], 2018. p. 1–2.
- YIN, S. et al. Fully-automatic segmentation of kidneys in clinical ultrasound images using a boundary distance regression network. In: *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. [S.l.: s.n.], 2019. p. 1741–1744.
- YIN, W. et al. *Comparative Study of CNN and RNN for Natural Language Processing*. 2017.

ZHOU, Z.-H. A brief introduction to weakly supervised learning. *National Science Review*, v. 5, n. 1, p. 44–53, 08 2017. ISSN 2095-5138. Disponível em: <<https://doi.org/10.1093/nsr/nwx106>>.