
Segmentação Automática de Canais Vasculares em Imagens Histológicas de Tecido Ósseo Utilizando Redes Neurais Completamente Convolucionais

Igor Gonçalves Ribeiro Silva



UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Igor Gonçalves Ribeiro Silva

**Segmentação Automática de Canais Vasculares em
Imagens Histológicas de Tecido Ósseo Utilizando
Redes Neurais Completamente Convolucionais**

Dissertação de mestrado apresentada ao Programa de Pós-graduação da Faculdade de Computação da Universidade Federal de Uberlândia como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação.

Área de concentração: Ciência da Computação

Orientador: Bruno Augusto Nassif Travençolo

Agradecimentos

Agradeço à Universidade Federal de Uberlândia, em especial a Faculdade de Computação, todo o seu corpo docente, direção e administração por esta oportunidade de crescimento pessoal, profissional e acadêmico.

Ao meu orientador Bruno Augusto Nassif Travençolo por todo o suporte e incentivo e compreensão presentes nas incontáveis aulas e reuniões e todas as suas correções e ensinamentos que muito contribuíram para minha formação como pessoa e pesquisador.

À professora Paula Dechichi pela disponibilização do material biológico utilizado no desenvolvimento do trabalho, bem como o incentivo e apoio para o desenvolvimento dele.

A minha esposa Camila que, além da participação ativa no desenvolvimento deste trabalho, pelo incentivo em retornar à academia e todo o apoio, amor e compreensão durante esse período. Sem ela, por diversos motivos, este trabalho não seria realizado.

A minha mãe Maria de Fátima, que nunca mediou esforços nem incentivos para que eu pudesse construir uma vida digna através dos estudos.

A minha tia Geralda Gislene deu todo o apoio e carinho durante minha graduação, tornando possível tal importante passo em minha carreira.

Resumo

Este estudo se propõe a validar um método de segmentação de canais ósseos em imagens histológicas de lâmina inteira, tal método utiliza redes neurais completamente convolucionais para a realização da tarefa. Foi utilizada uma rede neural originalmente desenvolvida com o intuito de segmentar tumores derivados da cavidade oral em imagens histológicas coradas com Hematoxilina e Eosina. O conjunto de dados foi composto por 65 imagens de lâmina inteira, as quais foram coradas com Hematoxilina e Eosina e extraídas a partir do fêmur de ratos saudáveis da linhagem Wistar. Com a ajuda de um especialista em histologia tais imagens foram analisadas e seus canais ósseos foram manualmente marcados. Em seguida as imagens foram preprocessadas para se gerar imagens binárias destacando a região de interesse e descartando o fundo. As imagens de ambos os conjuntos (imagens originais e binárias) foram então quebradas em sub-imagens de 640x640 pixels de tamanho. A rede foi treinada e validada com 1722 sub-imagens. O treinamento contou ainda com uma estratégia de aumento de dados com 7 possíveis variações das imagens. O método foi avaliado comparando-se as regiões segmentadas pela rede com as marcações do especialista. Foram calculadas a acurácia, especificidade, sensibilidade, precisão, Intersecção sobre União e $f1$ -score das segmentações resultantes. Além disso, foi feita uma comparação com outro método de segmentação automática de canais ósseos encontrada na literatura. O método validado por este trabalho mostrou-se eficiente e superior ao método com o qual foi comparado, apresentando $f1$ -score de 84,9% e Intersecção sobre União de 73,7%.

Palavras-chave: Segmentação de imagens. Canais Ósseos. Redes Neurais Convolucionais.

Abstract

This study proposes to validate a method that use fully convolutional neural networks for segmentation of bone canals on whole slide images. A neural network developed originally for segmenting tumors derived from the oral cavity in histological images stained with Hematoxylin and Eosin was used. The dataset was composed by 65 whole slide images images extracted from the femur of healthy Wistar rats. With the help of a histology expert, 65 whole slide images images were analyzed and their bone canals were manually labeled. These images were preprocessed in order to generate binary images highlighting the region of interest and discarding the background. Then the images of both sets (original and binary images) were broken into patches of 640x640 pixels. The network was trained and validated with 1722 patches. During the training a data augmentation strategy was used with 7 possible variations of the images. The method was evaluated by comparing the regions segmented by the network with the manual labels. The accuracy, specificity, precision, sensitivity, Intersection over Union and f1-score of the resulting segmented images were calculated. Furthermore, the method was compared with another method for automatic segmentation of bone canals found in the literature. The method proved to be efficient and superior to the method that was compared with, reaching a f1-score close of 84,9% and Intersection over Union of 73,7%.

Keywords: Image Segmentation. Bone Canals. Convolutional Neural Networks.

Lista de ilustrações

Figura 1 – Estrutura básica de um osso longo.	16
Figura 2 – Estrutura básica de um <i>perceptron</i>	18
Figura 3 – Estrutura básica de uma rede neural convolucional.	19
Figura 4 – Exemplo de convolução entre duas matrizes bidimensionais.	19
Figura 5 – Agrupamento utilizando uma função de <i>max-pooling</i> .	20
Figura 6 – Estrutura básica da rede U-Net.	21
Figura 7 – Exemplo de entrada e saída do método proposto por (GONDIM et al., 2021)	24
Figura 8 – Exemplo de marcações feitas pelo especialista em (GONDIM et al., 2021)	25
Figura 9 – Exemplos de saídas do método proposto em (JULIA et al., 2021).	25
Figura 10 – Principais elementos gráficos presentes no método proposto por (SANTOS et al., 2022).	26
Figura 11 – Exemplo de imagem utilizada no método proposto.	28
Figura 12 – Marcação do especialista para o método proposto.	28
Figura 13 – Arquitetura da rede utilizada	29
Figura 14 – Região de imagem marcada manualmente e sua respectiva máscara binária.	30
Figura 15 – Transformações utilizadas no aumento de dados.	31
Figura 16 – Diagrama do método proposto.	33
Figura 17 – Curvas de acurácia e perda ao logo do treinamento.	34
Figura 18 – Saída da rede neural.	35
Figura 19 – Métricas obtidas na análise por pixel.	36
Figura 20 – Comparação entre marcação manual feita por especialista e saída do método.	37
Figura 21 – Imagem marcada pela rede com regiões ampliadas	37
Figura 22 – Métricas obtidas na análise por canal.	39
Figura 23 – Exemplos de componentes conectados obtidos pelo método proposto.	39

Figura 24 – Comparaço de regies ampliadas das marcações manuais feita por especialista e de saídas do metodo	40
Figura 25 – Comparação entre marcação manual feita por especialista e saída do metodo.	41
Figura 26 – Exemplos de componentes conectados obtidos pelo metodo proposto por (GONDIM et al., 2021).	42
Figura 27 – Exemplo de falha do algoritmo proposto por (GONDIM et al., 2021).	43

Lista de tabelas

Tabela 1 – Médias de acurácia, precisão, <i>f1-score</i> , sensibilidade e especificidade para cada <i>threshold p</i> testado na análise por pixel.	36
Tabela 2 – Médias de acurácia, precisão, <i>f1-score</i> , sensibilidade e especificidade para cada <i>threshold p</i> testado na análise por canal.	38
Tabela 3 – Médias de precisão, <i>f1-score</i> , sensibilidade e Intersecção sobre União para cada os métodos testados.	43

Lista de abreviaturas

RNC Redes Neurais Convolucionais

RNCC Redes Neurais Completamente Convolucionais

HE Hematoxilina e Eosina

Sumário

1	INTRODUÇÃO	11
1.1	Motivação	12
1.2	Objetivo	13
1.3	Hipótese	13
1.4	Contribuições	14
2	FUNDAMENTAÇÃO TEÓRICA	15
2.1	O tecido ósseo e a rede vascular	15
2.2	Aprendizado de máquina	16
2.3	Aprendizado Profundo	17
2.4	Redes Neurais Convolucionais	18
2.4.1	Camadas convolucionais	19
2.4.2	Camadas de agrupamento	20
2.4.3	Camadas de achatamento	20
2.4.4	Camada densa ou conectada	20
2.5	U-Net	20
2.6	Métricas de validação	22
2.6.1	Acurácia	22
2.6.2	Especificidade	22
2.6.3	Precisão	23
2.6.4	Sensibilidade	23
2.6.5	<i>F1-Score</i>	23
2.6.6	Intersecção sobre União	23
2.7	Trabalhos Relacionados	24
3	MATERIAIS E MÉTODOS	27
3.1	Conjunto de dados	27
3.2	Arquitetura da Rede Neural	29

3.3	Método	30
3.3.1	Ambiente de execução	31
4	RESULTADOS E DISCUSSÃO	34
4.0.1	Treinamento da Rede	34
4.1	Análise por pixel	35
4.2	Análise por canal	38
4.2.1	Comparação	42
5	CONCLUSÃO	44
5.0.1	Contribuições em Produções Bibliográficas	45
	REFERÊNCIAS	46

CAPÍTULO 1

Introdução

A histologia é a área da biologia que estuda tecidos e composição de órgãos. Ela possui grande relevância na área acadêmica pois tem como proposta estudar a estrutura e a função dos tecidos, realizar análises quantitativas, avaliar respostas a tratamentos e investigar aspectos do desenvolvimento embrionário e da fisiologia. Outras aplicações importantes da histologia são encontradas na medicina, em que é empregada em diagnósticos médicos, auxiliando patologistas a identificar e caracterizar doenças, tumores, infecções e outras alterações nos tecidos (JUNQUEIRA; CARNEIRO, 1985).

As análises histológicas são comumente feitas a partir de imagens denominadas cortes histológicos, que podem ser obtidas em laboratório por meio de um longo processo de preparo. A análise desses imagens de lâmina inteira muitas vezes são demoradas e demandam grande esforço de um especialista (LINHARES et al., 2022). Dada a importância da análise histológica, é interessante desenvolver métodos que possam facilitá-la, tornando-a mais rápida e acessível para os pesquisadores e profissionais da saúde.

O tecido ósseo é um tecido conjuntivo que tem grande relevância em pesquisas da área da histologia devido à sua estrutura, composição e propriedades regenerativas. Desenvolver formas de automatizar a análise de imagens histológicas de tecido ósseo pode auxiliar estudos relacionados à estrutura do tecido, sua regeneração e efeitos de tratamentos de doenças (LINHARES et al., 2019). Porém é um desafio automatizar tais análises devido a vários fatores como o tamanho das imagens, complexidade e quantidade de estruturas, irregularidades e rasgos no tecido decorrentes do processo de preparo do corte histológico (GONDIM et al., 2021).

Nos últimos anos o aprendizado de máquina tem ganhado destaque tanto no meio acadêmico como no corporativo por apresentar resultados satisfatórios na execução de tarefas complexas, inclusive em relação ao processamento de imagens (MIKLOSIK; EVANS, 2020). Uma das abordagens do aprendizado de máquina é o aprendizado supervisionado, cujo princípio é o uso de uma base de dados para treinar um modelo de forma que o mesmo aprenda a realizar uma determinada tarefa. Esse conjunto de dados deve estar organizado de forma a conter exemplos diversos de entradas de dados e suas respectivas

saídas esperadas (MONARD; BARANAUSKAS, 2003).

Entretanto, ao se trabalhar com modelos de aprendizado de máquina, o conjunto de dados a ser utilizado pode ser um fator limitante, pois é preciso um conjunto de dados bem estruturado para se realizar a tarefa desejada a fim de se realizar um treinamento que não seja enviesado e que apresente bons resultados (PAULLADA et al., 2021). Dessa forma, para se criar e treinar um modelo de aprendizado de máquina para a execução de uma tarefa, muitas vezes é necessário um árduo trabalho prévio de elaboração de um conjunto de dados que viabilize o treinamento.

Conjuntos de imagens são amplamente utilizados em técnicas de aprendizado profundo (do inglês *deep learning*), uma subárea do aprendizado de máquina. O uso de redes neurais, modelos de aprendizado de máquina que tentam simular estruturas do cérebro humano, tem apresentado resultados interessantes no campo da visão computacional, especialmente as Redes Neurais Convolucionais (RNC). Tais redes se baseiam na forma como os seres humanos percebem e aprendem características chave das imagens (RAWAT; WANG, 2017).

O uso do aprendizado profundo na área médica vêm ganhando força e importância especialmente desde o surgimento das redes do tipo U-Net, arquitetura proposta em 2015 por (RONNEBERGER; FISCHER; BROX, 2015) que vem sido amplamente utilizada para segmentação de imagens biomédicas graças à sua capacidade de realizar classificação a nível de pixel. A utilização de tais técnicas pode ser de grande ajuda para especialistas em análises clínicas, diagnósticos e pesquisas, tornando o processo mais acessível e ágil (ESTEVA et al., 2021).

1.1 Motivação

Sabendo da importância da visão computacional nas ciências biomédicas, e do importante papel que o aprendizado profundo vem desempenhando na área de visão computacional, é interessante oferecer ferramentas que façam algum tipo de processamento automático em imagens histológicas, tais como classificação de imagens ou segmentação semântica.

Existem alguns métodos que visam resolver problemas específicos no processamento automático de imagens, como segmentação de alguma estrutura específica a partir da aplicação de determinados procedimentos nas imagens, como o proposto em (GONDIM et al., 2021). Porém tais métodos muitas vezes não apresentam invariância, ou seja, não costumam reagir bem a variações nas entradas, como tamanho das imagens, ruídos, variações de cor e posicionamento. Isso dificulta a utilização dessas ferramentas, visto que a parametrização ideal pode ser diferente para cada entrada (LINHARES et al., 2022), o que levou ao questionamento sobre a possibilidade de se criar uma ferramenta de processamento de imagens que possa se ajustar automaticamente para a realização de uma determinada tarefa independentemente das características específicas de cada imagem a

ser analisada.

Para a realização desse tipo de trabalho, as RNCs têm apresentado ótimos resultados, pois por meio da convolução aprendem padrões de imagens em um nível local e os identificam na imagem independente de posição, tamanho ou rotação (MUELLER; MAS-SARON, 2019). Portanto deve ser possível utilizar uma RNC que realize segmentação de estruturas de interesse em imagens histológicas de tecido ósseo com performance e precisão satisfatórias.

Entretanto, como já mencionado, utilizar uma rede neural para este tipo de tarefa requer um conjunto de dados estruturado de forma adequada. Existem vários conjuntos de dados de imagens médicas disponíveis para uso, como o ALL-IDB (Labati; Piuri; Scotti, 2011) e ErythrocytesIDB (GONZÁLEZ-HIDALGO et al., 2015), que são conjuntos de amostras de sangue; e o COVID-19 Radiography Database (CHOWDHURY et al., 2020), conjunto de imagens de raio-x de pulmões de pacientes com COVID-19. Porém observou-se uma escassez de dados destinado especificamente à segmentação de imagens histológicas de tecido ósseo.

1.2 Objetivo

Partindo da ideia apresentada, o objetivo deste trabalho é validar o uso de uma RNC como método de segmentação de canais vasculares em imagens histológicas de tecido ósseo. O método testado deve alcançar resultados aceitáveis e comparáveis com métodos já existentes.

Um desafio deste trabalho é a preparação do *dataset*, que será feito por meio da marcação da região de interesse em várias imagens de lâmina inteira. Tais marcações serão realizadas por um especialista em histologia.

1.3 Hipótese

Este trabalho busca validar a seguinte hipótese:

É possível utilizar Redes Neurais Convolucionais para realizar a segmentação de canais ósseos em imagens histológicas com precisão aceitável em relação à posição e à forma das estruturas em questão.

1.4 Contribuições

A principal contribuição esperada deste trabalho é demonstrar a viabilidade do uso de Redes Neurais Convolucionais para a tarefa de segmentação de canais ósseos em imagens histológicas, abrindo assim portas para o uso de RNCs em outras tarefas da histologia relacionadas a imagens de tecido ósseo e para o desenvolvimento de novas tecnologias que possam contribuir para pesquisas dessa área.

Outra contribuição esperada é a disponibilização pública de um conjunto de imagens histológicas de tecido ósseo, que pode vir a ser útil para outros pesquisadores na elaboração e desenvolvimento de suas pesquisas. Tal conjunto pode ser evoluído em trabalhos posteriores a para que seja possível realizar outras tarefas de visão computacional em imagens de tecido ósseo como classificação ou segmentação de outras estruturas presentes nesse tipo de tecido.

CAPÍTULO 2

Fundamentação Teórica

Neste capítulo são abordados os conceitos fundamentais que embasam o desenvolvimento desta pesquisa e os principais trabalhos correlatos.

2.1 O tecido ósseo e a rede vascular

O tecido ósseo é um dos objetos de estudo da histologia e chama a atenção de pesquisadores e profissionais da área médica especialmente devido às suas propriedades regenerativas. Macroscopicamente os ossos longos podem ser divididos em duas regiões: diáfise e epífise. As epífises são as extremidades do osso e são compostas por osso esponjoso, ou seja, com muitas cavidades intercomunicantes (JUNQUEIRA; CARNEIRO, 1985). Já a diáfise é a região intermediária do osso, sendo mais fina e apresentando osso compacto – sem cavidades – em sua maior parte (JUNQUEIRA; CARNEIRO, 1985). A Figura 1 (a) mostra tais regiões e suas respectivas composições.

A maior parte da diáfise é composta túneis longitudinais formados por lamelas concêntricas, tais túneis são conhecidos como sistema de Havers e formam em seus centros canais que são percorridos por vasos sanguíneos, linfáticos e nervos. Também existem canais transversais que conectam canais de Havers adjacentes chamados canais de Volkmann, formando assim a rede de canais ósseos (JUNQUEIRA; CARNEIRO, 1985). A Figura 1 (b) ilustra as estruturas mencionadas destacando os canais de Havers e de Volkmann.

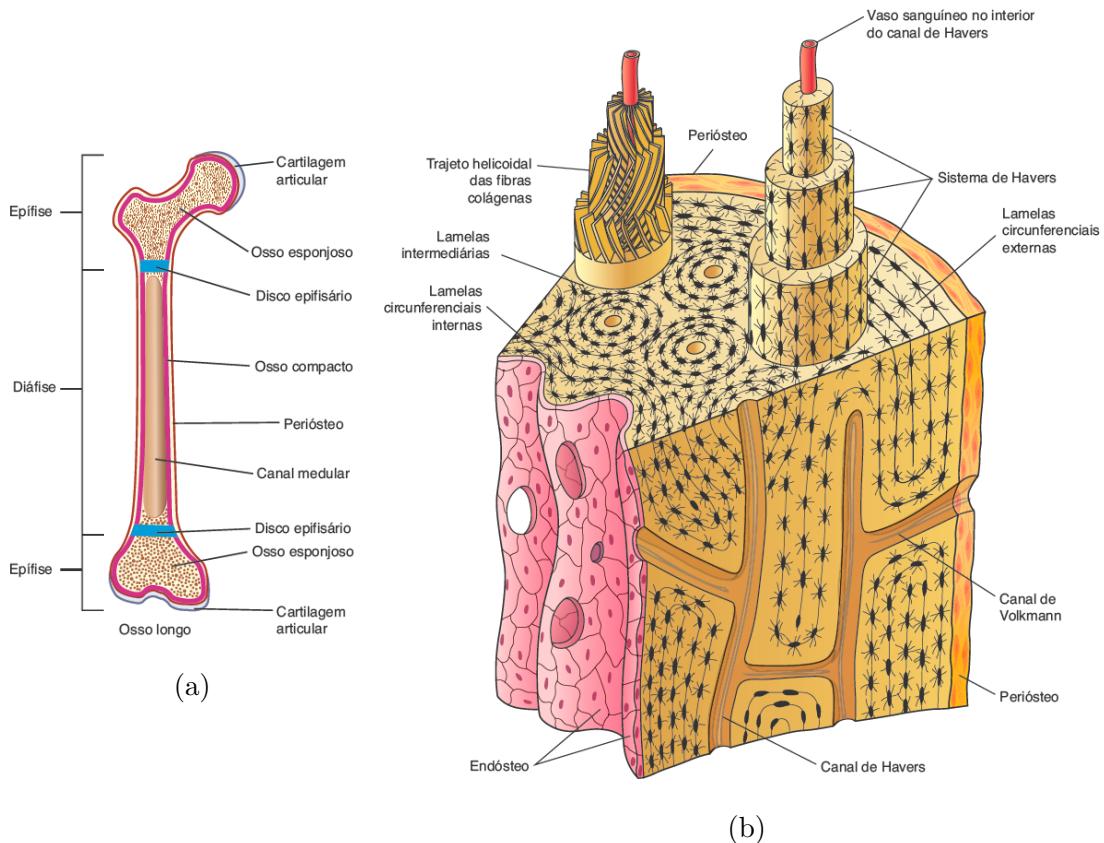


Figura 1 – Estrutura básica de um osso longo. Em (a) os principais componentes anatômicos de um osso longo. Em (b) os canais de Havers, que compõem o osso compacto, principal componente da diáfise, região intermediária de ossos longos. Fonte: (JUNQUEIRA; CARNEIRO, 1985)

Os vasos sanguíneos e linfáticos que percorrem os canais nutrem os osteócitos, que são estruturas que cumprem um importante papel na manutenção e integridade da matriz óssea. Dessa forma a rede de canais ósseos cumpre um importante papel de nutrição do tecido contribuindo para seu desenvolvimento e no processo de reparo ósseo (JUNQUEIRA; CARNEIRO, 1985).

2.2 Aprendizado de máquina

Ainda em 1959 Arthur Lee Samuel definiu o aprendizado de máquina como “o campo de estudo que dá aos computadores a habilidade de aprender sem serem explicitamente programados” (SIMON, 2013). De fato o aprendizado de máquinas automatizado permite que computadores aprendam a realizar tarefas analisando uma base de dados previamente rotulada, além de aprimorar seu aprendizado à medida que novos dados lhe são apresentados (MONARD; BARANAUSKAS, 2003).

O aprendizado de máquina é um área da inteligência artificial que se apoia em técnicas e conceitos da matemática e estatística. Esse aprendizado é dividido em quatro tipos: aprendizado supervisionado, não supervisionado, auto supervisionado e por reforço. Todos

possuem uma variedades de algoritmos, sendo que, para cada problema a ser resolvido, um determinado algoritmo pode ser mais adequado que os demais (MUELLER; MASSARON, 2019).

Como mencionado acima, uma das técnicas de aprendizado de máquina é o aprendizado supervisionado. Nessa abordagem utiliza-se uma base de dados na qual para cada entrada o resultado desejado é conhecido. As entradas com seus respectivos resultados esperados, também chamados de rótulos, são apresentados ao algoritmo em uma etapa conhecida como etapa de treinamento, em que o algoritmo ajusta, com base nos dados fornecidos, uma série de parâmetros internos a fim de encontrar padrões que o permitam estimar resultados para novas entradas (MONARD; BARANAUSKAS, 2003). A saída pode ser qualitativa, tratando-se assim de uma tarefa de classificação; ou quantitativa, tratando-se assim de uma tarefa de regressão (MUELLER; MASSARON, 2019).

2.3 Aprendizado Profundo

Uma subárea do aprendizado de máquina é o aprendizado profundo, que também trabalha com grandes conjuntos de dados para aprender a realizar tarefas. A grande diferença é que, ao invés de métodos e modelos estatísticos, o aprendizado profundo faz uso de apenas uma técnica que simula o funcionamento do cérebro humano: as redes neurais (MUELLER; MASSARON, 2019).

Redes neurais podem possuir diversas camadas e arquiteturas, sendo que cada uma pode ser mais ou menos indicada para cada tipo de problema a ser tratado. Tais redes são compostas por estruturas chamadas neurônios, que simulam os neurônios biológicos e se ligam uns aos outros por meio de ligações com diferentes pesos que simulam as sinapses cerebrais (MUELLER; MASSARON, 2019).

A rede neural mais simples possível é o *perceptron*, que pode ser utilizada para tarefas de classificação binária ou regressão linear. O *perceptron* é composto por apenas uma camada de um único neurônio. Ele recebe uma entrada X de tamanho m que é multiplicada por um vetor de pesos W também de tamanho m . Em seguida é aplicada sobre o produto escalar $X \cdot W$ uma função de ativação a fim de produzir um único resultado de saída (BLOCK, 1962). A Figura 2 (a) ilustra a estrutura de um *perceptron*.

O processo de treinamento de uma rede neural consiste justamente em realizar, com base no conjunto de dados de entrada, a parametrização do vetor de pesos W das ligações entre os componentes da entrada X e o neurônio (MUELLER; MASSARON, 2019).

A partir do *perceptron*, novas arquiteturas de redes neurais podem ser formuladas adicionando mais camadas, mais neurônios, funções de retro-propagação de erro dentre outras estratégias que permitem que as redes neurais resolvam problemas mais complexos. A Figura 2 (b) mostra um exemplo de rede multi-camadas, contendo uma camada de entrada, uma camadas escondida e uma camada de saída composta por 4 neurônios,

dessa forma, uma rede como essa seria capaz de realizar tarefas de classificação com 4 classes.

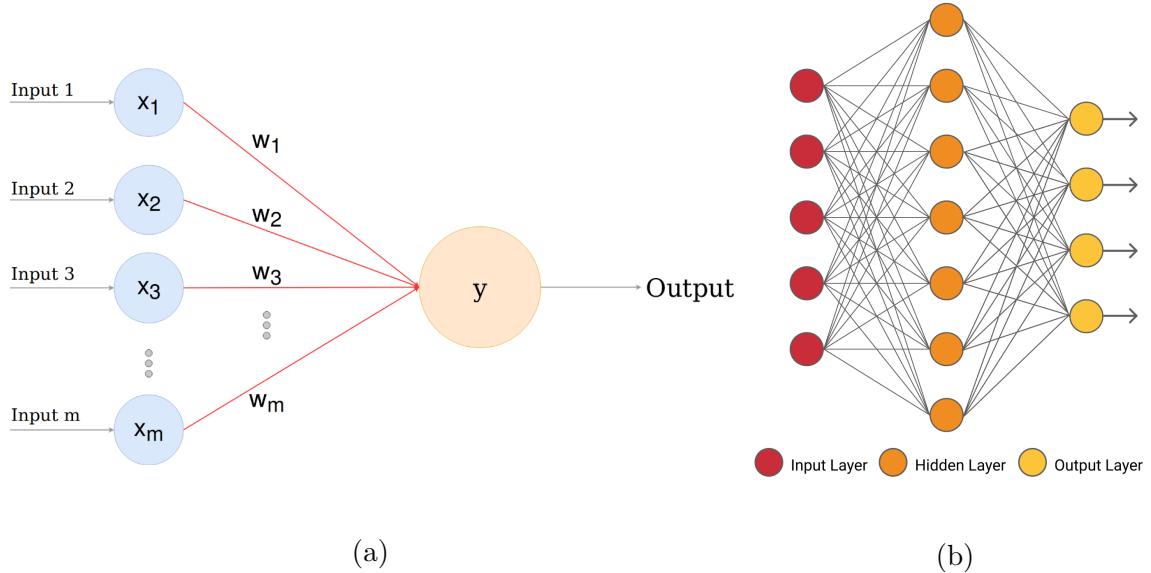


Figura 2 – Em (a) um Exemplo da estrutura básica de um *perceptron*. Fonte: (ARC, 2018). Em (b) exemplo de uma rede de perceptrans multicamada. Fonte: (VADAPALLI,)

2.4 Redes Neurais Convolucionais

Desde o ano de 2012, quando a arquitetura *AlexNet* venceu a competição *ImageNet Large Scale Visual Recognition Challenge (ILSVRC)*, as Redes Neurais Convolucionais (RNC) vêm ganhando popularidade na execução de tarefas relacionadas a processamento de imagens, vídeos e até mesmo voz (VARGAS; PAES; VASCONCELOS, 2016). Esse tipo de rede neural é uma variação das redes de *perceptron* multicamada, e se baseia no princípio biológico da percepção visual dos seres humanos a fim de minimizar o processamento dos dados para obter o resultado esperado (MUELLER; MASSARON, 2019). São compostas por camadas convolucionais, de agrupamento, achatamento e conectada.

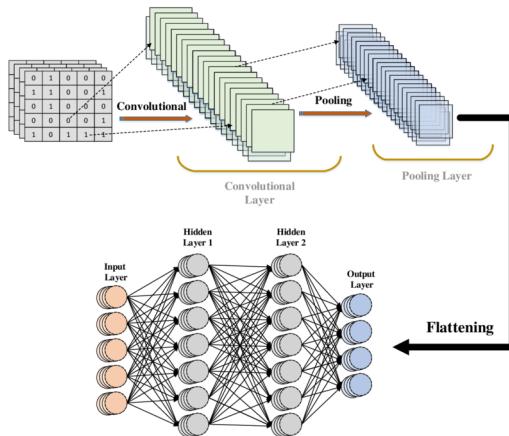


Figura 3 – Exemplo de estrutura básica de uma rede neural convolucional. Fonte: (SRI-NIVASAN et al., 2021)

2.4.1 Camadas convolucionais

Na arquitetura de uma RNC, ilustrada na Figura 3, ao receber uma imagem como entrada são utilizadas as ditas camadas convolucionais. Tais camadas aplicam filtros (também conhecidos por *kernels*), representados por matrizes tridimensionais, sobre parte da imagem de entrada (RAWAT; WANG, 2017).

A aplicação desses filtros se dá por um algoritmo de janela deslizante, que percorre a imagem em seus três canais de cores utilizando um passo de determinado tamanho (também conhecido como *stride*) realizando uma operação de produto escalar a fim de gerar uma nova matriz cujas dimensões podem ser menores ou iguais à anterior dependendo do tamanho do filtro e do *stride*. As matrizes resultantes da convolução entre os canais de cor e os filtros são chamadas de mapas de características, e suas dimensões variam de acordo com o tamanho do filtro (*kernel*) e do passo da janela deslizante (*stride*). Estes são importantes parâmetros que devem ser determinados durante o projeto da rede. Esse é o processo chamado convolução, que é ilustrado pela Figura 4 e é o mais importante processo das RNCs (GÉRON, 2019).

$$\begin{array}{c}
 \begin{array}{|c|c|c|c|c|c|c|c|c|c|} \hline
 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ \hline
 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ \hline
 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ \hline
 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ \hline
 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ \hline
 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ \hline
 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ \hline
 \end{array} \\
 \mathbf{I}
 \end{array}
 \quad
 \begin{array}{c}
 \begin{array}{|c|c|c|} \hline
 1 & 0 & 1 \\ \hline
 0 & 1 & 0 \\ \hline
 1 & 0 & 1 \\ \hline
 \end{array} \\
 \mathbf{K}
 \end{array}
 \quad
 \begin{array}{c}
 \begin{array}{|c|c|c|c|c|} \hline
 1 & 4 & 3 & 4 & 1 \\ \hline
 1 & 2 & 4 & 3 & 3 \\ \hline
 1 & 2 & 3 & 4 & 1 \\ \hline
 1 & 3 & 3 & 1 & 1 \\ \hline
 3 & 3 & 1 & 1 & 0 \\ \hline
 \end{array} \\
 \mathbf{I} * \mathbf{K}
 \end{array}$$

Figura 4 – Exemplo de convolução entre duas matrizes bidimensionais. O filtro K percorre a matriz I com um passo de uma unidade calculando os produtos escalares entre a respectiva região da matriz I e o filtro K, gerando assim uma nova matriz $I * K$ que é o mapa de características, resultado da convolução entre I e K. Fonte: (RIEBESELL,)

2.4.2 Camadas de agrupamento

Após cada convolução é executada outra etapa importante do processamento em RNCs que é o agrupamento. Para esta tarefa existem as camadas de agrupamento (do inglês *pooling*) cuja função é diminuir o tamanho do mapa de características visando agilidade e invariância espacial (RAWAT; WANG, 2017). Essa etapa consiste na utilização de uma janela deslizante para aplicar sobre as janelas uma função que seleciona um único valor para representá-la, conforme ilustrado pela Figura 5. O uso de funções de valor máximo, mínimo e médio é bastante comum nesta etapa (GÉRON, 2019).

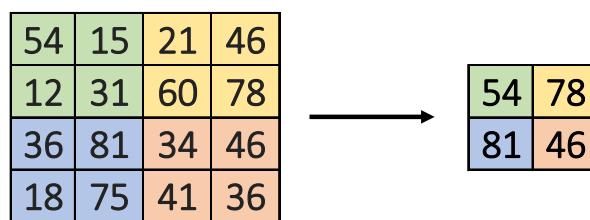


Figura 5 – Exemplo de agrupamento utilizando uma função de *max-pooling*. O mapa de características é percorrido por uma janela deslizante que seleciona apenas o maior valor da região para compôr um mapa de características menor.

Devido ao fato de diminuir o tamanho da imagem de entrada, esta etapa do processo – composta por camadas de convolução seguidas de camadas de agrupamento – também é conhecida como caminho de contração (GÉRON, 2019).

2.4.3 Camadas de achatamento

Após as etapas de convolução é executada uma camada de achatamento (do inglês *flatten layer*), na qual os mapas de características, que são matrizes, são transformados em um vetor que será usado para alimentar uma rede de neurônios multicamada.

2.4.4 Camada densa ou conectada

A camada que comporta a rede de neurônios é chamada camada completamente conectada (ou densa) e tem por objetivo traçar a decisão a partir dos mapas de características. Esse tipo de camada é muito utilizado em tarefas de classificação (RAWAT; WANG, 2017).

Uma RNC pode apresentar outros tipos de camadas além desses tipos básicos, ou ainda não apresentar alguma das camadas mencionadas acima dependendo do objetivo da rede (MUELLER; MASSARON, 2019).

2.5 U-Net

Proposta em 2015 por (RONNEBERGER; FISCHER; BROX, 2015), as redes U-Net vêm sendo aplicadas em diversos trabalhos referentes à segmentação de imagens médicas, uma

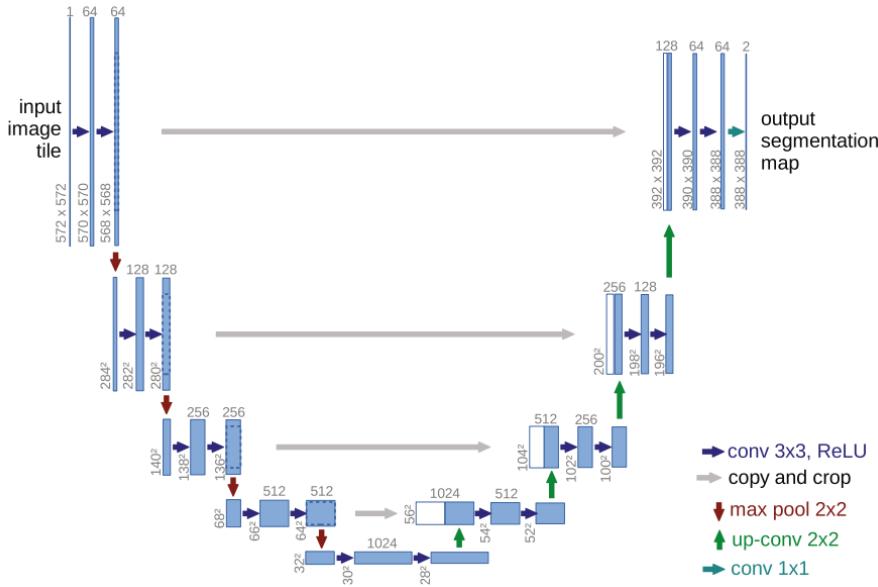


Figura 6 – Estrutura básica da rede U-Net. Fonte: (RONNEBERGER; FISCHER; BROX, 2015)

vez que as redes de classificação, como as RNCs tradicionais, não conseguem trazer informações contextuais a nível de pixel, o que é essencial em tarefas de segmentação, especialmente ao se trabalhar com imagens médicas (SIDDIQUE et al., 2021).

As redes U-Net se diferem das RNCs de classificação principalmente devido à adição de uma etapa de decodificação. Dessa forma após passar pelo caminho de contração o vetor de saída é novamente transformado em uma matriz e esta é ampliada por algumas camadas de aumento até chegar ao tamanho original da entrada, sendo a imagem resultante uma máscara que representa a região de interesse segmentada. Este processo também é conhecido como caminho de expansão. A combinação do caminho de contração com o caminho de expansão resulta em uma rede quase simétrica, cujo formato lembra a letra U, conforme ilustrado pela Figura 6, dando assim o nome a esse tipo de rede (RONNEBERGER; FISCHER; BROX, 2015).

Outra interessante característica deste tipo de rede é a ausência de uma camada conectada, sendo que esta é substituída por camadas convolucionais que processam os mapas de características com mais eficiência que uma camada conectada, conforme descrito em (LONG; SHELHAMER; DARRELL, 2015). A ausência da camada conectada, dá o nome à arquitetura das Redes Neurais Completamente Convolucionais (RNCC).

Assim, a arquitetura básica de uma U-Net, conforme foi originalmente proposta, consiste em um caminho de contração e um caminho de expansão. Durante a contração são aplicadas repetidamente duas convoluções com um filtro 3x3 seguidas de uma função de ativação *ReLU* e um *max-pooling* de tamanho 2x2, diminuindo em duas vezes e aumentando em duas vezes, respectivamente: o tamanho e o número dos mapas de características da saída da camada anterior (SIDDIQUE et al., 2021). Na expansão é feito o

processo inverso: a cada camada é feito o aumento de resolução do mapa de características com uma convolução 2x2 que diminui pela metade os canais de características e dobra as dimensões da imagem (*up-convolution*), uma concatenação com o mapa de características correspondente da etapa de contração e duas convoluções 3x3 seguidas por uma função de ativação *ReLU*. Por fim é feita uma convolução 1x1 para mapear vetores de características para as classes desejadas. A rede possui um total de 23 convoluções (RONNEBERGER; FISCHER; BROX, 2015).

2.6 Métricas de validação

Esta seção apresenta e explica as principais métricas utilizadas para a validação do método proposto neste trabalho. Para isso, consideremos a seguinte nomenclatura:

- VP - Verdadeiros positivos
- VN - Verdadeiros negativos
- FP - Falsos positivos
- FN - Falsos negativos

2.6.1 Acurácia

A acurácia é a relação entre a quantidade de elementos classificados corretamente e a quantidade total de elementos analisados. Muitas vezes a acurácia não é a medida mais confiável para se avaliar modelos de aprendizado de máquina, pois é um desafio encontrar (e até mesmo construir) conjuntos de dados balanceados, ou seja, com aproximadamente a mesma quantidade de elementos de cada classe (UDDIN, 2019). Dessa forma, se um conjunto de dados para classificação binária possui um número muito mais elevado de elementos que não pertencem a uma classe C1 do que elementos que pertencem a C1 o modelo pode classificar corretamente os elementos que não pertencem a C1 e incorretamente uma boa parte dos elementos de C1 e ainda assim apresentar uma boa acurácia, visto que os elementos que não pertencem a C1 compõem a maior parte do conjunto de dados. A acurácia é calculada segundo a Equação 1.

$$\text{acurácia} = \frac{VP + VN}{VP + FN + FP + VN} \quad (1)$$

2.6.2 Especificidade

A especificidade é a razão entre os elementos negativos que foram classificados corretamente, ou seja, os verdadeiros negativos VN; e os elementos que são de fato negativos, ou seja, a soma entre os verdadeiros negativos VN e falsos positivos FP. Novamente, se

o número de elementos negativos do conjunto de dados for muito superior ao número de elementos positivos a especificidade também pode ser uma métrica enviesada. A especificidade é calculada segundo a Equação 2.

$$\text{especificidade} = \frac{VN}{VN + FP} \quad (2)$$

2.6.3 Precisão

A precisão é a razão entre o número de elementos classificados corretamente como positivos e o número de elementos que foram classificados como positivos, conforme mostra a Equação 3.

$$\text{precisão} = \frac{VP}{VP + FP} \quad (3)$$

2.6.4 Sensibilidade

A sensibilidade (também conhecida como *recall*) é a razão entre o número de elementos positivos classificados corretamente e o número total de elementos que são de fato positivos. A sensibilidade é calculada segundo a Equação 4.

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

2.6.5 F1-Score

O *f1-score* (também conhecida como *Dice Coefficient*) é a média harmônica entre precisão e a sensibilidade. Calculada conforme a Equação 5, busca representar a qualidade geral do modelo e é a métrica mais confiável quando se busca obter um modelo que minimize tanto os falsos positivos como os falsos negativos.

$$f1\text{-score} = \frac{2 * \text{precision} * \text{sensitivity}}{\text{precision} + \text{sensitivity}} \quad (5)$$

2.6.6 Intersecção sobre União

A Intersecção sobre União (*Intersetion over Union – IoU*, também conhecida como *Jaccard Index*) é uma medida para calcular a similaridade entre conjuntos e é dada pela Equação 6.

$$IoU = \frac{VP}{VP + FP + FN} \quad (6)$$

2.7 Trabalhos Relacionados

Esta seção apresenta os principais trabalhos relacionados a esta pesquisa e que mostram técnicas de segmentação em imagens histológicas.

A literatura apresenta poucos trabalhos focados em segmentação de canais ósseos em imagens de lâmina inteira. Em (LIU et al., 1999) a segmentação foi realizada em imagens micro-radiográficas usando conhecimento em imagens ósseas e algoritmos de clusterização. As imagens de lâmina inteira se diferem das imagens micro-radiográficas especialmente em tamanho e resolução já que podem assumir ordens de giga-pixel e apresentar cores enquanto as imagens utilizadas por (LIU et al., 1999) possuíam dimensões de 512 por 576 pixels e se apresentavam em escala de cinza. A alta resolução das imagens de lâmina inteira é interessante na análise histológica devido à riqueza de detalhes que proporciona ao especialista. Porém trabalhar com tais imagens é uma tarefa desafiadora devido ao seu tamanho, complexidade e quantidade das estruturas em relação às imagens micro-radiográficas (GONDIM et al., 2021).

Um método para segmentação automática de canais ósseos em imagens histológicas de tecido ósseo foi proposto por (GONDIM et al., 2021). Tal método processa as imagens de entrada, resumidamente, em quatro etapas: remoção de artefatos externos à matriz óssea, remoção de artefatos internos à matriz óssea, segmentação da rede vascular e remoção de falsos positivos. A cada etapa são utilizadas diversas técnicas de processamento de imagens (tais como abertura e fechamento morfológico, binarização, deconvolução de cor, etc.) em sequência e até mesmo um algoritmo de aprendizado de máquina não supervisionado denominado *K-means*. Isso torna o método bastante complexo e implica em uma grande quantidade de parâmetros cujos valores devem ser fixos e obtidos empiricamente, o que leva a questionar sobre sua sensibilidade a variações na imagem. A Figura 7 mostra um exemplo de entrada e saída do método proposto por (GONDIM et al., 2021).

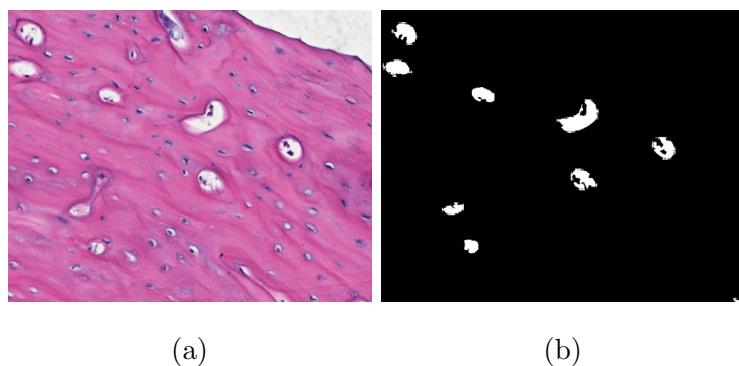


Figura 7 – Exemplo de entrada e saída do método proposto por (GONDIM et al., 2021). Em (a) imagem de entrada. Em (b) a saída do método para a entrada em (a).

Para a avaliação do método, as regiões segmentadas por (GONDIM et al., 2021) foram comparadas com marcações feitas por um especialista. Como as marcações manuais não seguiam a forma exata das estruturas, para comparar a segmentação do método foi

considerado que se ao menos 20% dos *pixels* de uma marcação feita pelo especialista estivessem contidos na região segmentada pelo método, então a estrutura foi corretamente segmentada. Isso deixa uma lacuna em relação à forma da região segmentada pois tais formas não puderam ser validadas devido às marcações manuais, pois estas foram feitas colocando um círculo envolvendo cada canal, como mostra a Figura 8.

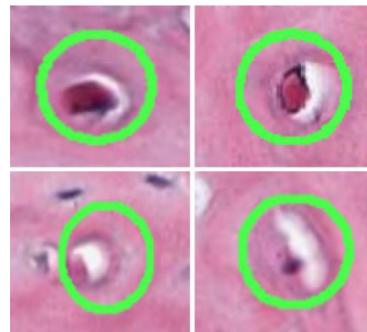


Figura 8 – Exemplo de marcações feitas pelo especialista. Fonte (GONDIM et al., 2021).

Outra abordagem foi testada por (JULIA et al., 2021), que, na intenção de segmentar canais ósseos em imagens histológicas geradas a partir do fêmur de ratos, modificou uma rede neural convolucional do tipo *U-Net* que originalmente havia sido feita para segmentação de anormalidades FLAIR em imagens de ressonância magnética cerebral. Entretanto os resultados da marcação automática feita pela RNC implementada não foram satisfatórios devido à falta de precisão das marcações manuais feitas nas imagens, o que é ilustrado pela Figura 9. O modelo treinado apresentou sobreajuste e alcançou um *Dice Coefficient* de apenas 20%.

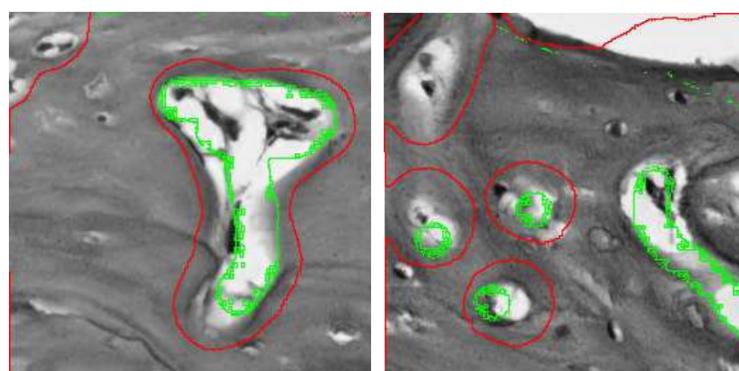


Figura 9 – Exemplos de saídas do método proposto. Em verde a saída da rede neural, em vermelho a marcação provida pelo especialista. Imagem adaptada de (JULIA et al., 2021).

Como proposta de melhoria em seu trabalho, (JULIA et al., 2021) sugere o uso de um conjunto de imagens com marcações mais precisas, a fim de melhorar os resultados da RNC.

Em ambos os trabalhos observa-se que a qualidade do conjunto de imagens teve impacto negativo na avaliação dos resultados da segmentação. Acredita-se que um conjunto

de imagens com marcações mais precisas e confiáveis em relação ao tamanho, forma e posição das estruturas de interesse pode levar a um resultado mais preciso, e o uso de redes neurais convolucionais pode resultar em um método mais robusto quanto à invariância, apresentando assim mais flexibilidade ao tratar imagens com características diferentes.

Pensando nos bons resultados que as redes neurais convolucionais têm apresentado no processamento de imagens nos últimos anos, (SANTOS et al., 2021) utiliza redes neurais convolucionais para segmentar regiões tumorais em imagens de lâmina inteira de tecido oral, obtendo bons resultados, como cerca de 97.6% de acurácia, 98.4% de especificidade e 92.9% de sensibilidade. Foi utilizada uma RNCC baseada no modelo *U-Net* treinada com uma base de imagens, desenvolvida pelos autores, que continha imagens histológicas de tecidos da cavidade oral corados com Hematoxilina e Eosina (HE).

Antes do treinamento foi feita uma etapa de pre-processamento para identificar as regiões de tecido, descartando áreas desinteressantes para o trabalho. As imagens eram então quebradas em sub-imagens e apresentadas à rede para o treinamento, que também contava com uma estratégia de aumento de dados. O resultado da rede era uma imagem em escala de cinza onde era calculada para cada pixel da imagem a probabilidade p de tal pixel pertencer ou não à região de interesse. Por fim eram marcados como positivos os pixels que apresentassem p acima de um *threshold* de 50%, valor este que foi escolhido empiricamente. A Figura 10 apresenta exemplos dos principais elementos gráficos presentes no método proposto por (SANTOS et al., 2022).

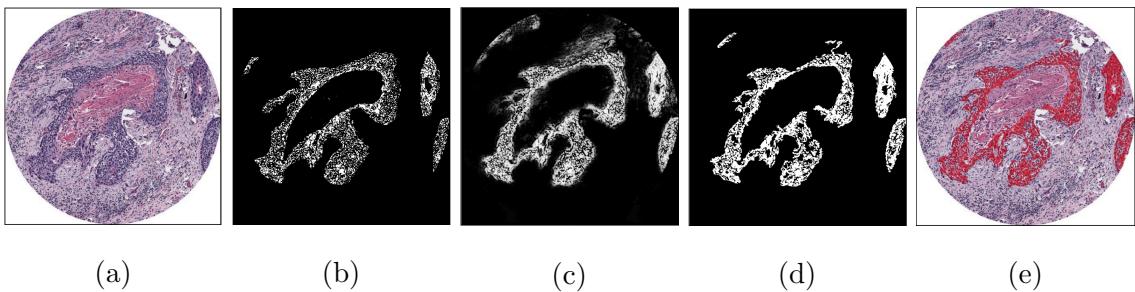


Figura 10 – Exemplos dos principais elementos gráficos presentes no método proposto por (SANTOS et al., 2022). Em (a) uma imagem de entrada. Em (b) um exemplo de marcação padrão ouro. Em (c) um exemplo da saída da rede, em que cada pixel é preenchido de acordo com a probabilidade de pertencer à região de interesse. Em (d) a marcação final do método, com *threshold* de 50%. Em (e) a região com tumor representada em vermelho na imagem original. Imagem adaptada de (SANTOS et al., 2022).

O método também foi validado em outras bases de imagens presentes na literatura, e apresentou um *f1-score* de 90% na base desenvolvida pelos autores e média geral de 83%. Além disso o método foi testado com diferentes tamanhos de sub-imagens – o melhor resultado foi obtido para sub-imagens de tamanho 640x640 pixels.

CAPÍTULO 3

Materiais e Métodos

Este capítulo irá descrever o conjunto de dados elaborado, a arquitetura da rede neural e o método de segmentação utilizado.

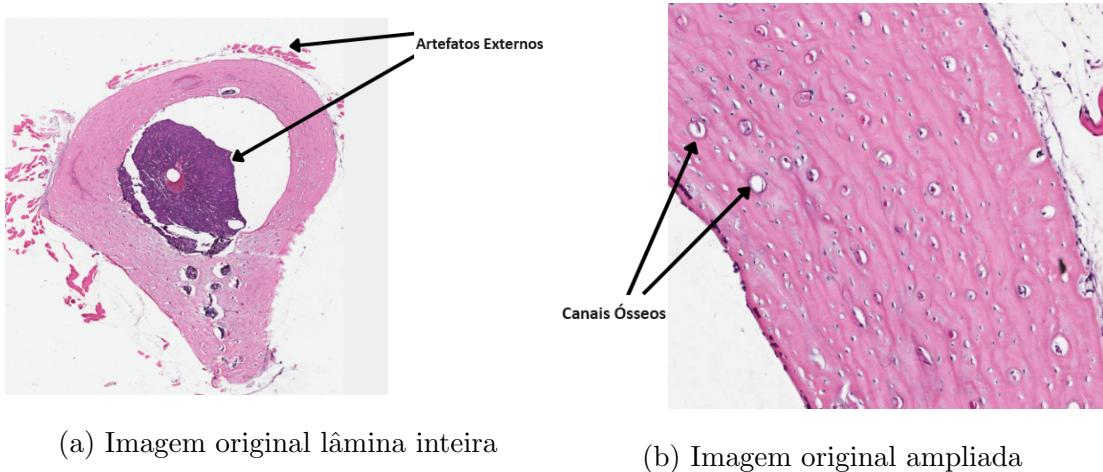
3.1 Conjunto de dados

As imagens utilizadas foram obtidas a partir do fêmur esquerdo de um rato *Rattus norvegicus* da linhagem *Wistar* saudável com 200 - 250g. O animal foi sacrificado e os fêmures foram removidos e fixados em formaldeído 10%, tamponado e desmineralizado em EDTA 4,13%. A diáfise, região intermediária do fêmur, foi incluída em parafina e obteve-se cortes histológicos seriados de cerca de 5 micrômetros de espessura corados em HE, que é o tipo de coragem mais amplamente utilizado na histologia devido a sua simplicidade e ao grande número de estruturas do tecido que permite visualizar (FELDMAN; WOLFE, 2014).

Todos os procedimentos para se obter este material foram executados de acordo com as normas do Colégio Brasileiro de Experimentação Animal (COBEA), com aprovação do Comitê de Ética no Uso de Animais da Universidade Federal de Uberlândia (CEUA-UFU) – Protocolo 060/09.

Neste estudo foram analisados 65 cortes histológicos de fêmur esquerdo. Foi utilizado um ScanScope AT Turbo®Scanner (Leica Biosystems, Nussloch, Alemanha) para digitalizar as imagens com uma ampliação efetiva de 20 \times . As imagens foram exportadas em arquivos TIFF de alta resolução. A Figura 11 mostra um exemplo de corte histológico para o fêmur irradiado. O tamanho da imagem na Figura 11 (a) é de 7.940 \times 9.051 pixels, a largura e altura do pixel são 0,502 μm . Nessa resolução é possível identificar os canais vasculares (Figura 11(b)).

As imagens coletadas foram marcadas por um especialista em histologia. A marcação foi feita por meio do software Photoshop, versão 2016 (Adobe Inc., 2016), contornando manualmente os canais ósseos com a ferramenta ‘Laço’. Em seguida foram exportadas em formato JPEG. Nesta etapa também foi feita a remoção do fundo da imagem, removendo



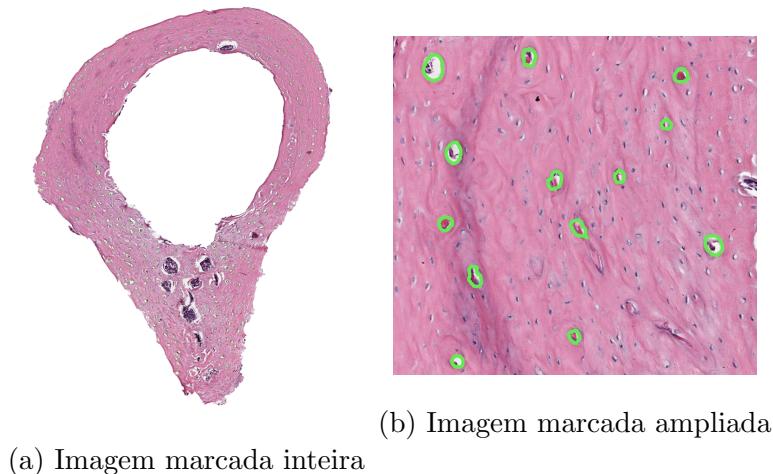
(a) Imagem original lâmina inteira

(b) Imagem original ampliada

Figura 11 – Exemplo de imagem utilizada no método proposto. Em (a) imagem da secção inteira destacando artefatos externos que não são de interesse na análise, em (b) imagem ampliada destacando os canais vasculares.

artefatos externos que não eram interessantes para a análise. Os cortes histológicos marcados manualmente foram posteriormente divididos em 1722 sub-imagens de dimensões 640x640 pixels conforme será descrito adiante no texto. A Figura 12(a) apresenta um exemplo de imagem marcada, e a Figura 12(b) uma região ampliada da mesma imagem evidenciando as marcações.

As imagens de lâmina inteira bem como as sub-imagens utilizadas neste trabalho estão disponíveis publicamente no endereço: <https://github.com/igorgonribs/histology-bone-canals-imageset>.



(b) Imagem marcada ampliada

(a) Imagem marcada inteira

Figura 12 – Imagem marcada por especialista. Em (a) imagem inteira. Em (b) imagem ampliada destacando os canais ósseos, contornados em verde.

3.2 Arquitetura da Rede Neural

A rede utilizada por este método foi a mesma desenvolvida por (SANTOS et al., 2022) com os mesmos parâmetros internos. Tal rede por sua vez é baseada na rede U-Net originalmente proposta por (RONNEBERGER; FISCHER; BROX, 2015) com algumas modificações que estão destacadas na Figura 13 e serão descritas a seguir:

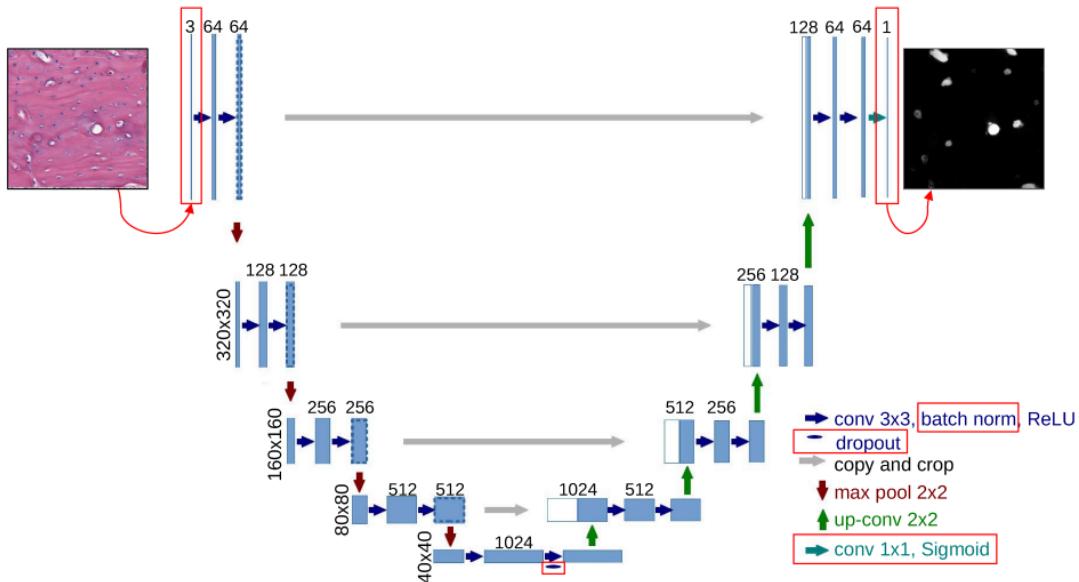


Figura 13 – Arquitetura da rede utilizada baseada na rede U-Net original com modificações destacadas em vermelho. Imagem adaptada de (SANTOS et al., 2022).

- **Normalização em lote:** A normalização em lote normaliza as entradas evitando valores muito baixos que possam gerar gradiente de fuga e fazer com que os pesos não se atualizem (ou se atualizem muito lentamente) durante o treinamento. A técnica também evita valores muito altos que possam gerar explosão de gradiente fazendo com que os pesos assumam valores muito altos rapidamente prejudicando o treinamento (GÉRON, 2019). A normalização em lote foi aplicada após cada camada convolucional.
- **Dropout:** O *dropout* é uma técnica que consiste em desativar parte dos neurônios da camada conectada durante algumas etapas do treinamento a fim de evitar o sobreajuste, tornando a rede mais robusta e generalizável (GÉRON, 2019). Foi utilizada uma taxa de *dropout* de 0,5. Dessa forma a cada passo do treinamento metade dos neurônios eram desativados aleatoriamente para minimizar as chances de sobreajuste.
- **Convolução 1x1 com função de ativação Sigmoid:** Ao final do processamento, foi adicionada à última convolução (1x1) uma função de ativação Sigmoid a fim de que, ao invés de uma imagem binária, a saída da rede fosse uma imagem em escala de

cinza onde quanto mais próxima de branco a cor de um pixel maior a probabilidade de o mesmo pertencer à região de interesse.

3.3 Método

A fim de reaproveitar a rede neural proposta por (SANTOS et al., 2022), foi necessário transformar as imagens marcadas manualmente em imagens binárias que destacassem as regiões de interesse. A Figura 14 mostra uma marcação manual e a respectiva máscara binária gerada.

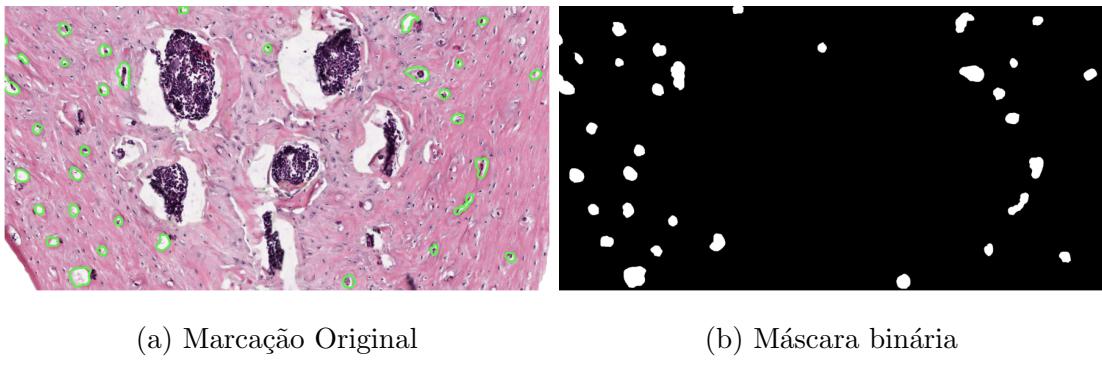


Figura 14 – Região de imagem marcada manualmente (a) e sua respectiva máscara binária (b).

O fundo das imagens originais (sem marcações) foi removido utilizando do processo descrito por (SANTOS et al., 2022): diminuindo-se a escala das imagens por um fator de 32x, destacando-se a região com tecido por meio de um filtro de cor e mapeando-se a região selecionada na imagem em tamanho real.

Ainda como foi feito por (SANTOS et al., 2022), são geradas sub-imagens a partir das imagens originais, desconsiderando-se as regiões que não contém tecido. O tamanho de sub-imagens utilizado foi 640x640 pixels, visto que foi o tamanho que apresentou melhores resultados em (SANTOS et al., 2022). As sub-imagens que continham apenas fundo foram descartadas.

O treinamento ainda conta com a estratégia de aumento de dados utilizada por (SANTOS et al., 2023), em que a cada época é aplicada uma transformação selecionada a partir de uma lista pré-definida. Para este aumento de dados, as seguintes operações foram consideradas: Inversão Horizontal, Inversão Vertical, Rotação, Tranposição, Transformação elástica, Distorção de grade e Distorção ótica. A Figura 16 mostra exemplos de cada uma dessas operações em comparação com um exemplar de sub-imagem original.

O conjunto de dados de treinamento foi composto por 80% das sub-imagens, destas, 20% foram usados para validação da rede durante o treinamento. Os 20% restantes compuseram o conjunto de teste.

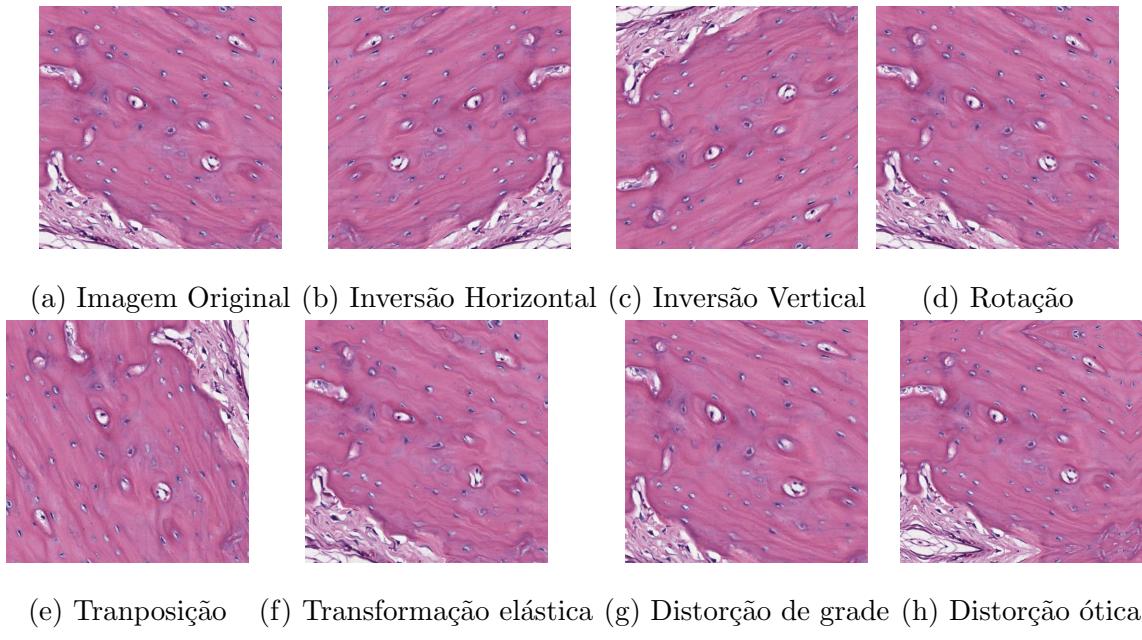


Figura 15 – Transformações utilizadas no aumento de dados.

Após o treinamento foi realizada a etapa de inferência, na qual são apresentadas à rede imagens do conjunto de testes, ou seja, que não foram utilizadas na etapa de treinamento, para que a segmentação dos canais seja feita e as métricas sejam calculadas.

Nesta etapa é calculada para cada pixel uma probabilidade de que o mesmo pertença a um canal. Caso a probabilidade seja maior que uma probabilidade limiar p , entende-se que aquele pixel faz parte da região de interesse.

Para determinar qual o valor a ser usado como probabilidade limiar foram testados para este parâmetro valores de 0% a 95% com um passo de 5%. Para cada imagem do conjunto de testes foram calculados acurácia, precisão, *f1-score*, sensibilidade e especificidade. A média das métricas citadas acima foi calculada para cada valor de p testado a fim de identificar empiricamente qual o melhor valor de p a ser utilizado para a segmentação final.

Também foi feita uma análise das medidas precisão, *f1-score*, sensibilidade e intersecção sobre união para cada canal segmentado a fim de se comparar os resultados com os obtidos pelo método proposto por (GONDIM et al., 2021).

3.3.1 Ambiente de execução

Os testes mencionados foram realizados em uma máquina com as seguintes especificações:

- Processador:** Intel(R) Core(TM) i7-12700 2.10 GHz de 12^a geração
- Memória RAM:** 16 Gb
- Sistema operacional:** Windows 11 Pro, versão 22H2

❑ **GPU:** NVIDIA GeForce RTX 3060

Para o desenvolvimento da rede neural e demais elementos de código utilizados no processamento dos dados foi criado um ambiente de desenvolvimento utilizando o gerenciador de pacotes e ambientes Anaconda e a linguagem Python na versão 3.10.11. As principais bibliotecas utilizadas foram:

- ❑ **Numpy** versão 1.24.3
- ❑ **Opencv** versão 4.6.0
- ❑ **Pandas** versão 1.5.3
- ❑ **Pillow** versão 9.4.0
- ❑ **Pytorch** versão 2.0.1
- ❑ **Scikit Learn** versão 1.2.2

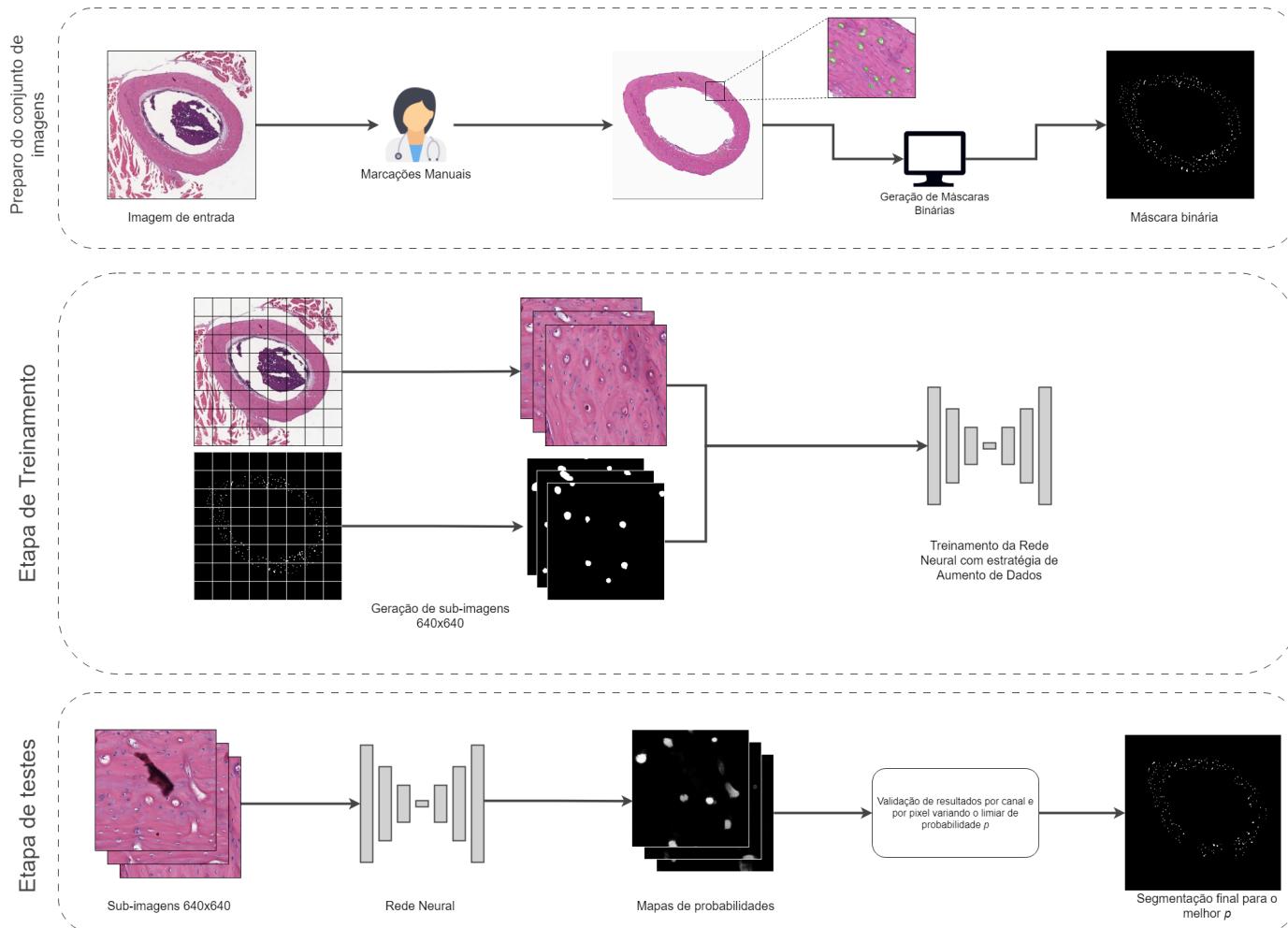


Figura 16 – Diagrama da metodologia utilizada. Na etapa de preparo do conjunto de dados as imagens foram manualmente, e foram geradas máscaras binárias destacando as estruturas de interesse. Na etapa de treinamento a rede foi treinada com as sub-imagens originais máscaras; o treinamento contou com uma estratégia de aumento de dados. Na etapa de testes a rede computou mapas de probabilidades; foram testados vários valores de limiar de probabilidade em dois tipos de análise: pixel a pixel e canal a canal; a segmentação final foi feita melhor o limiar de probabilidade observado.

CAPÍTULO 4

Resultados e Discussão

Este capítulo apresenta os resultados obtidos nos experimentos descritos no capítulo anterior.

4.0.1 Treinamento da Rede

Para a fase de treinamento, obteve-se uma acurácia acima de 98% para o conjunto de treinamento e em torno de 97% para o conjunto de validação. Já para o erro obteve-se 2% para o conjunto de treinamento e 12% para o conjunto de validação. A Figura 17 mostra as curvas de perda e acurácia para ambos os conjuntos de dados ao longo do treinamento.

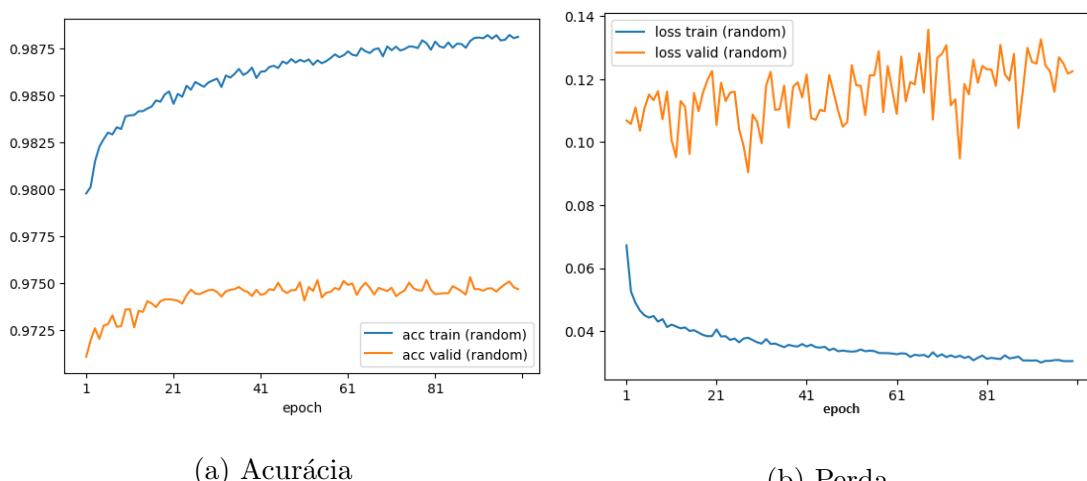


Figura 17 – Curvas de acurácia e perda ao longo das 100 épocas do treinamento. Em (a) a acurácia das previsões da rede aplicadas ao conjunto de treinamento (em azul) e ao conjunto de validação (em laranja). Em (b) a perda das previsões da rede aplicadas ao conjunto de treinamento (em azul) e ao conjunto de validação (em laranja).

A acurácia apresenta valores elevados principalmente devido ao desbalanceamento do conjunto de dados, pois as imagens possuem muito mais pixels negativos do que positivos,

visto que os canais vasculares são estruturas pequenas que ocupam pouco espaço nas imagens quando comparamos à matriz óssea, por exemplo.

4.1 Análise por pixel

Em seguida foi executada a inferência sobre as imagens do conjunto de testes. Nesta etapa a saída da rede neural é uma imagem em escala de cinza onde cada pixel possui um valor entre 0 e 255 e quanto maior a probabilidade de um pixel pertencer à região de interesse maior será o seu valor. Portanto, visualmente, quanto mais próximo da cor branca for o *pixel* na imagem de saída maior a probabilidade de o mesmo pertencer à região de interesse, conforme ilustrado pela Figura 18.

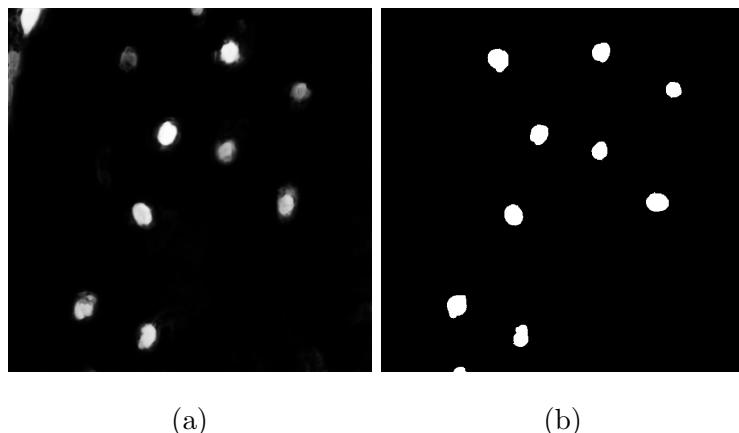


Figura 18 – Saída da rede neural. Em (a) um exemplo de uma saída da rede neural. Em (b) a respectiva máscara manualmente marcada pelo especialista.

Dessa forma, um *pixel* é considerado um positivo se a probabilidade de o mesmo pertencer à região de interesse for maior que um limiar p .

$$Y(x) = \begin{cases} 0, & \text{se } \frac{x}{255} < p \\ 1, & \text{caso contrário} \end{cases} \quad (7)$$

Para validar o melhor valor para tal limiar foram testados vários valores de p , variando-o de 0,0 até 0,95 com um passo de 0,05. Para cada valor testado foram calculados os valores médios de acurácia, precisão, *f1-score*, sensibilidade e especificidade. A Tabela 1 e o gráfico da Figura 19 mostram os valores médios obtidos para cada valor de p .

Após o teste foi observado que para $p=0,4$ obteve-se o melhor valor de *f1-score*, logo, a melhor relação entre precisão e sensibilidade. A especificidade e acurácia se mantiveram elevadas durante todo o teste devido ao alto volume de *pixels* negativos corretamente classificados.

Utilizando o resultado acima foi feita a segmentação das imagens do conjunto de testes utilizando $p=0,4$. A Figura 20 compara a máscara gerada pela rede com a máscara gerada a partir das marcações do especialista, e a Figura 21 mostra uma imagem da lâmina inteira

Limiar p	Acurácia	Precisão	F1-Score	Sensibilidade	Especificidade
0.00	0.801	0.154	0.257	0.968	0.796
0.05	0.944	0.416	0.548	0.907	0.953
0.10	0.958	0.505	0.615	0.873	0.969
0.15	0.965	0.563	0.651	0.844	0.977
0.20	0.968	0.608	0.673	0.817	0.982
0.25	0.970	0.642	0.685	0.794	0.985
0.30	0.971	0.674	0.693	0.769	0.987
0.35	0.972	0.701	0.697	0.743	0.989
0.40	0.973	0.726	0.696	0.716	0.991
0.45	0.973	0.747	0.693	0.689	0.992
0.50	0.973	0.768	0.684	0.657	0.993
0.55	0.973	0.789	0.671	0.621	0.995
0.60	0.973	0.810	0.652	0.580	0.996
0.65	0.972	0.829	0.628	0.537	0.996
0.70	0.971	0.840	0.592	0.484	0.997
0.75	0.970	0.858	0.547	0.426	0.998
0.80	0.969	0.873	0.487	0.360	0.998
0.85	0.967	0.882	0.415	0.290	0.999
0.90	0.965	0.884	0.303	0.195	0.999
0.95	0.962	0.780	0.135	0.078	0.999

Tabela 1 – Médias de acurácia, precisão, *f1-score*, sensibilidade e especificidade para cada *threshold p* testado na análise por pixel.

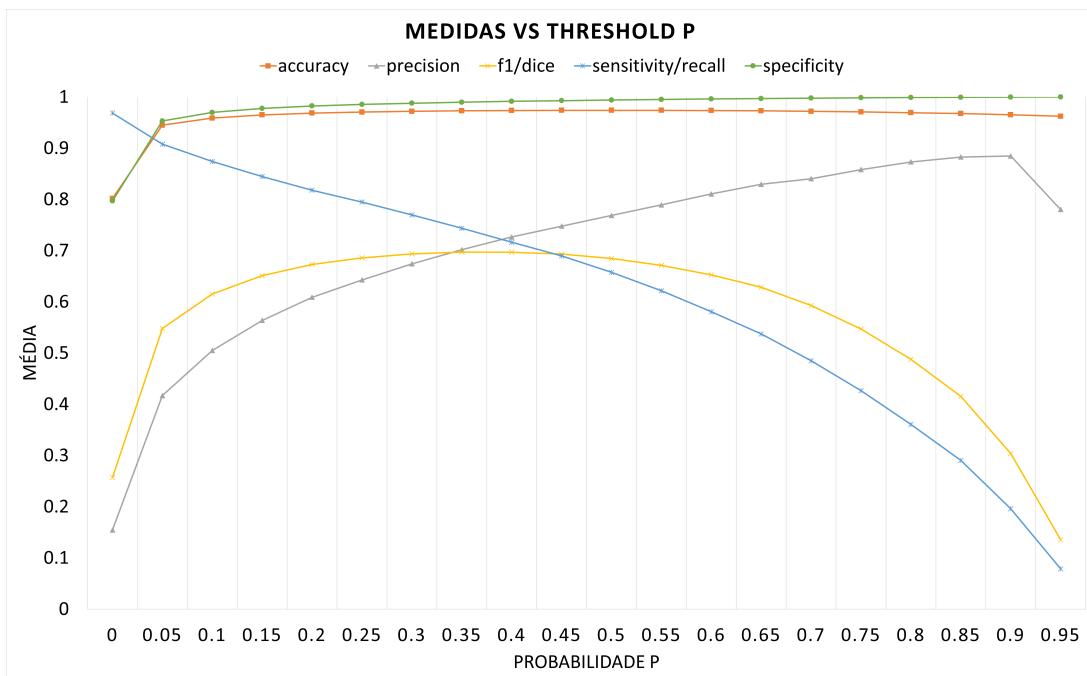


Figura 19 – Acurácia, *f1-score*, precisão, sensibilidade e especificidade em função do *threshold p* na análise por pixel.

marcada a partir da máscara gerada pela rede e algumas regiões ampliadas para detalhar os resultados, onde ainda é possível observar alguns falsos negativos .

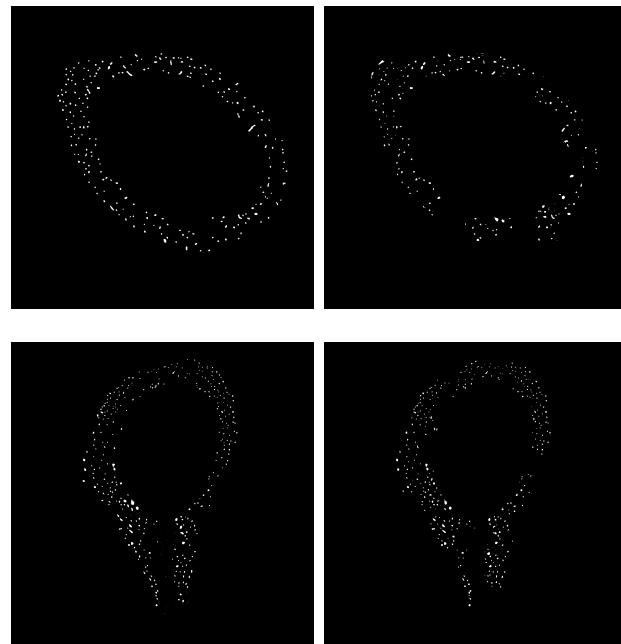


Figura 20 – Comparação entre marcação manual feita por especialista e saída do método.
Na primeira coluna exemplos de marcações manuais, na segunda coluna as máscaras geradas pelo método utilizando um limiar $p=0,4$.

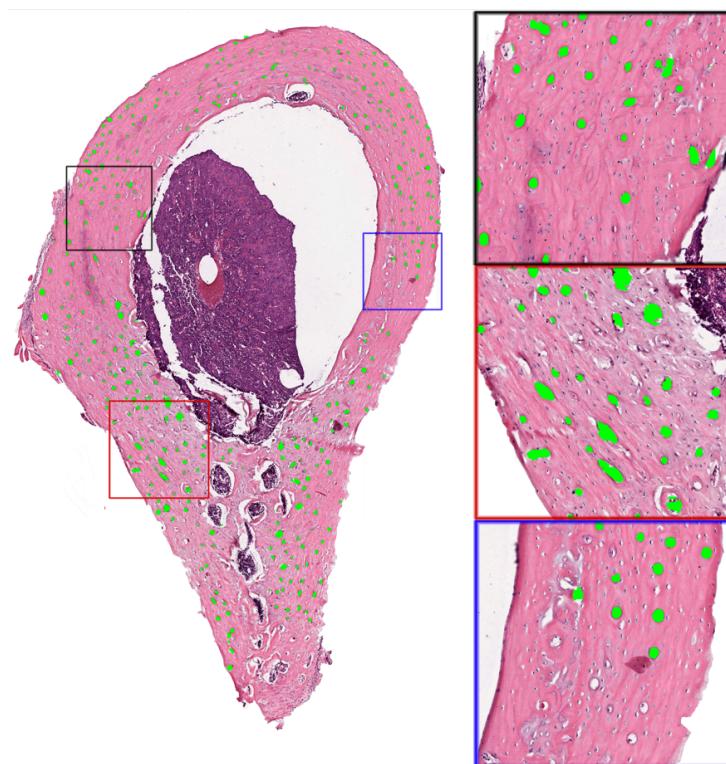


Figura 21 – Imagem de lâmina inteira com as marcações feitas pela rede em verde e com três regiões ampliadas. Observa-se, principalmente na região delimitada pela cor azul, a presença de alguns falsos negativos.

4.2 Análise por canal

Nesta análise, para cada componente conectado da máscara de referência foi extraída uma sub-imagem de acordo com as coordenadas x_i, x_f, y_i, y_f do componente. Extraía-se também uma sub-imagem da respectiva região delimitada por x_i, x_f, y_i, y_f na imagem segmentada pelo método para verificar se havia ou não um canal naquela região. Em seguida foi calculada a intersecção entre as duas sub-imagens. Foram considerados como verdadeiros positivos os componentes cuja intersecção coincidiram ao menos 70% com o canal da máscara de referência. A Figura 23 mostra exemplos de canais na máscara de referência, na segmentação da rede e a intersecção entre os dois canais.

Em seguida foram extraídas da imagem segmentada sub-imagens representando os componentes conectados que não estavam presentes na imagem de referência, identificando, portanto, os falsos positivos. Nesta análise não é possível contabilizar verdadeiros negativos.

Para essa nova análise foram testados novamente os valores do limiar de probabilidade p . Por não termos o número de verdadeiros negativos as métricas utilizadas foram: Precisão, *f1-score*, sensibilidade e Intersecção sobre União. O melhor resultado foi obtido para $p = 0,10$. A Tabela 2 mostra o resultados da análise para cada valor de p .

Threshold p	Precisão	F1-Score	Sensibilidade	IoU
0.00	0.772	0.837	0.965	0.718
0.05	0.805	0.831	0.896	0.718
0.10	0.836	0.838	0.866	0.737
0.15	0.844	0.826	0.835	0.720
0.20	0.846	0.812	0.806	0.702
0.25	0.844	0.796	0.778	0.679
0.30	0.846	0.780	0.748	0.660
0.35	0.837	0.756	0.713	0.631
0.40	0.827	0.730	0.678	0.595
0.45	0.813	0.698	0.637	0.555
0.50	0.792	0.659	0.591	0.508
0.55	0.770	0.614	0.538	0.452
0.60	0.748	0.563	0.477	0.396
0.65	0.712	0.503	0.415	0.340
0.70	0.660	0.434	0.347	0.277
0.75	0.609	0.368	0.284	0.218
0.80	0.535	0.291	0.217	0.163
0.85	0.420	0.200	0.142	0.104
0.90	0.246	0.099	0.067	0.046
0.95	0.055	0.017	0.011	0.008

Tabela 2 – Médias de acurácia, precisão, *f1-score*, sensibilidade e especificidade para cada *threshold p* testado na análise por canal.

A análise descrita acima foi realizada sobre as imagens do conjunto de testes segmentadas com o limiar $p = 0,10$. Foram obtidos bons resultados resultados. O *f1-score* alcançou 84,8%, um ganho significativo em relação ao melhor valor obtido na análise por pixel, e a Intersecção sobre União foi de cerca de 73,7%. A Figura 22 mostra um gráfico com as medidas calculadas para cada valor de p testado.

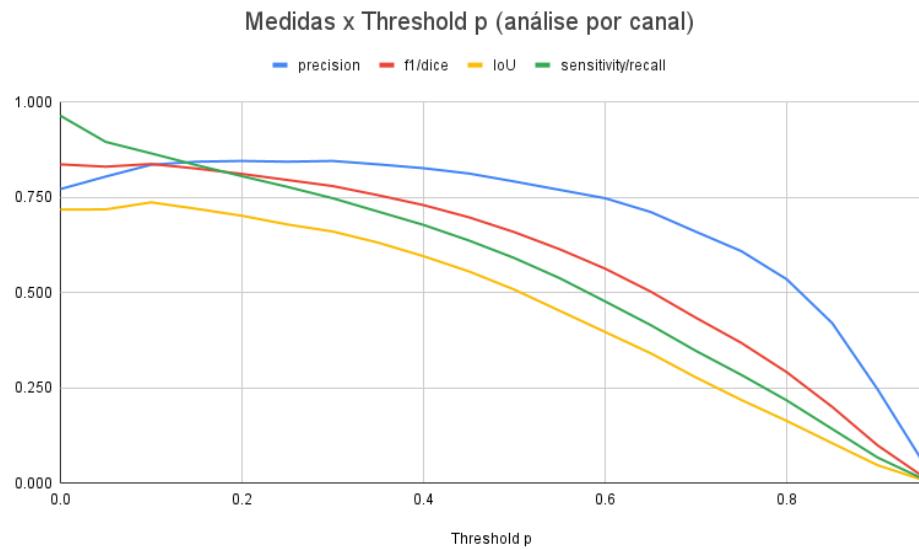


Figura 22 – Precisão, *f1-score*, Intersecção sobre União e sensibilidade em função do *threshold p* na análise por canal.

A Figura 23 mostra alguns exemplos de sub-imagens desta análise para as imagens segmentadas pela rede neural.

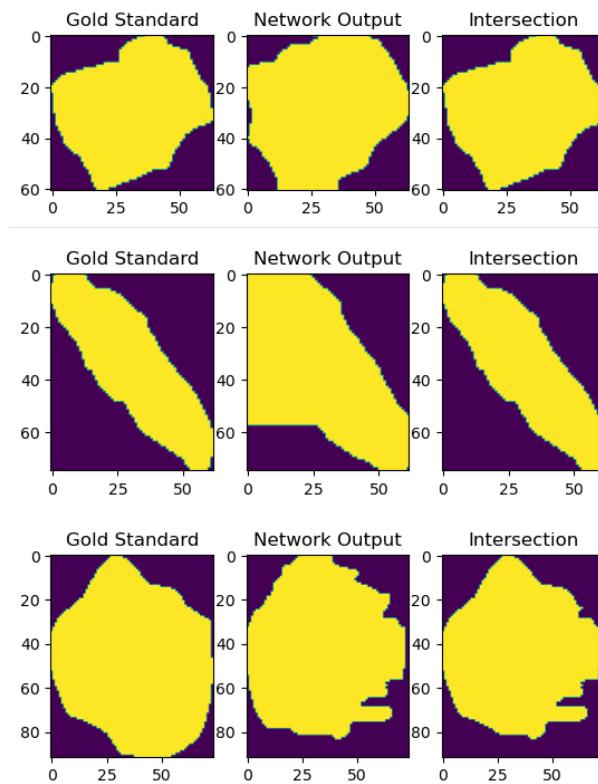


Figura 23 – Três exemplos de componentes conectados analisados durante a análise por canal. Na primeira coluna o componente na máscara de referência em amarelo (padrão ouro). Na segunda coluna a segmentação da rede neural para o respectivo componente. Na terceira coluna a intersecção entre a primeira e a segunda coluna.

A Figura 25 mostra a comparação entre máscaras geradas a partir da marcação manual do especialista e máscaras geradas pelo método com $p=0,1$. Já a Figura 24 mostra algumas regiões ampliadas para melhor visualização dos detalhes das marcações.

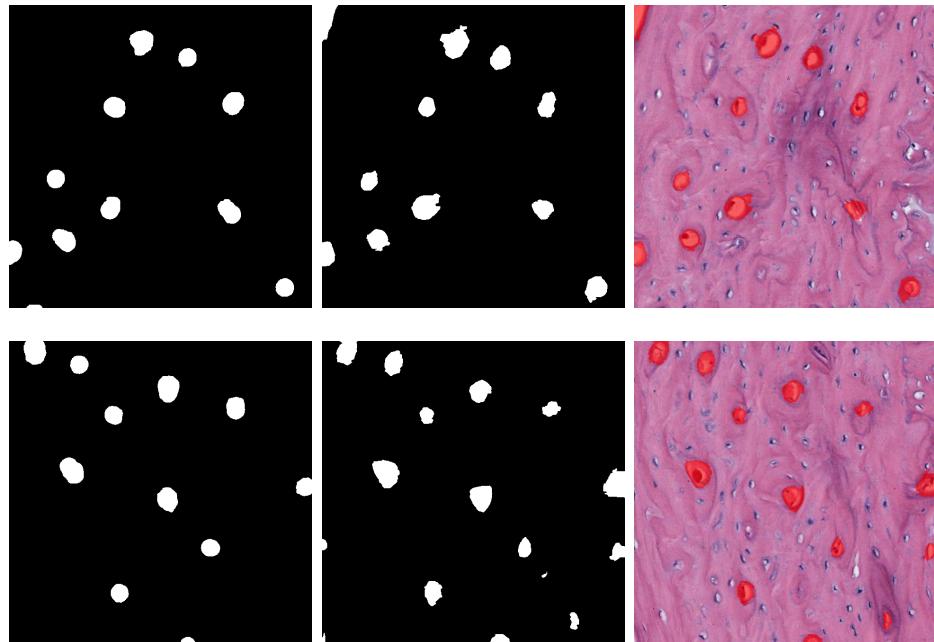


Figura 24 – Comparação de regiões ampliadas das marcações manuais feita por especialista e de saídas do método. Na primeira coluna regiões ampliadas de marcações manuais, na segunda coluna as respectivas regiões nas máscaras geradas pelo método utilizando um limiar $p=0,1$, e na terceira coluna as respectivas regiões na imagens marcadas a partir das máscaras geradas pela rede.

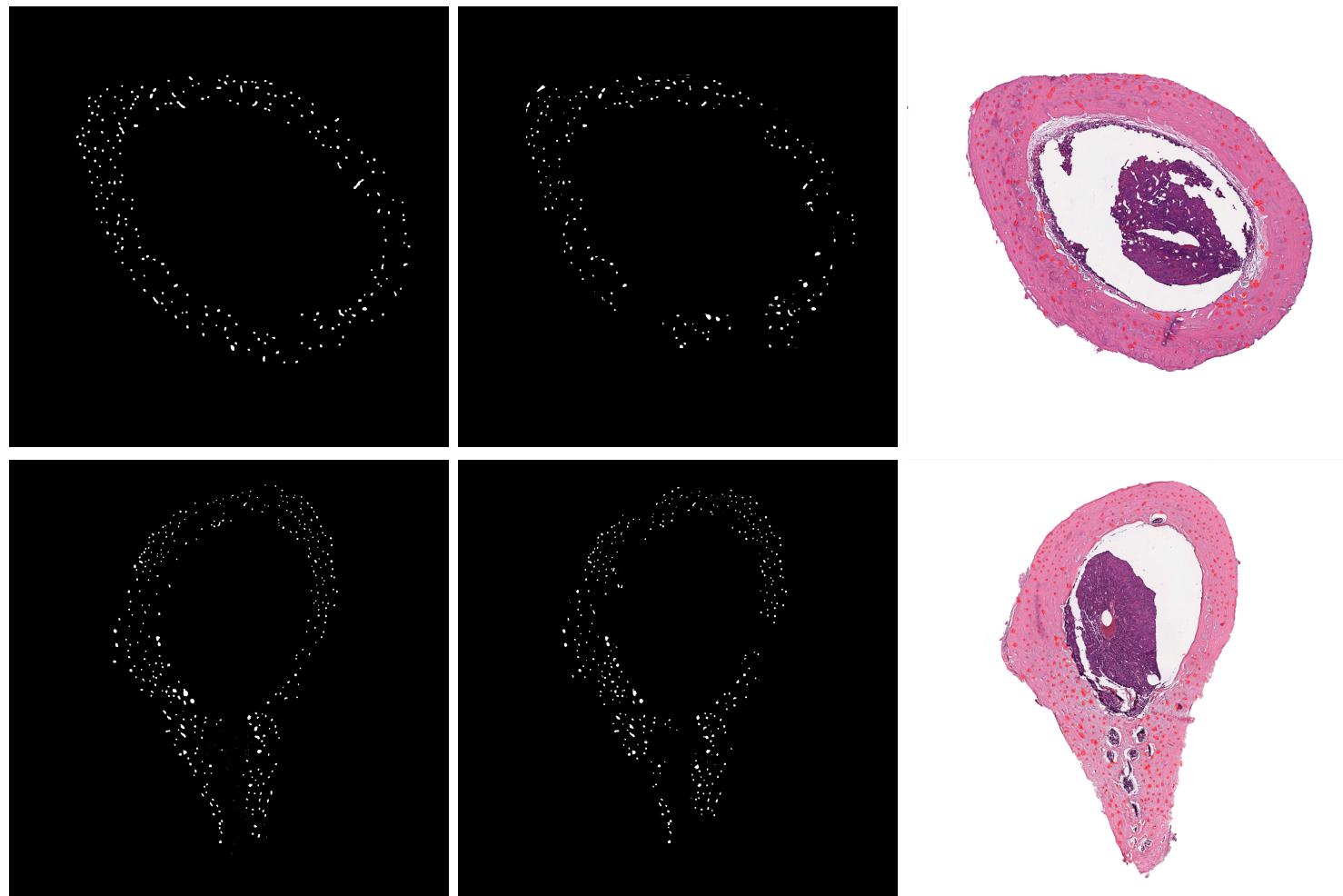


Figura 25 – Comparação entre marcação manual feita por especialista e saída do método. Na primeira coluna exemplos de marcações manuais, na segunda coluna as máscaras geradas pelo método utilizando um limiar $p=0,1$, e na terceira coluna a imagem marcada a partir da máscara gerada pela rede.

4.2.1 Comparação

A fim de comparação, foi feita uma implementação do algoritmo proposto por (GONDIM et al., 2021) utilizando a linguagem de programação Python na versão 3.10. As imagens de lâmina inteira do conjunto de testes foram então segmentadas e foi realizada a análise por canal conforme descrito anteriormente. A Figura 26 mostra exemplos de componentes conectados obtidos na análise por canal do algoritmo.

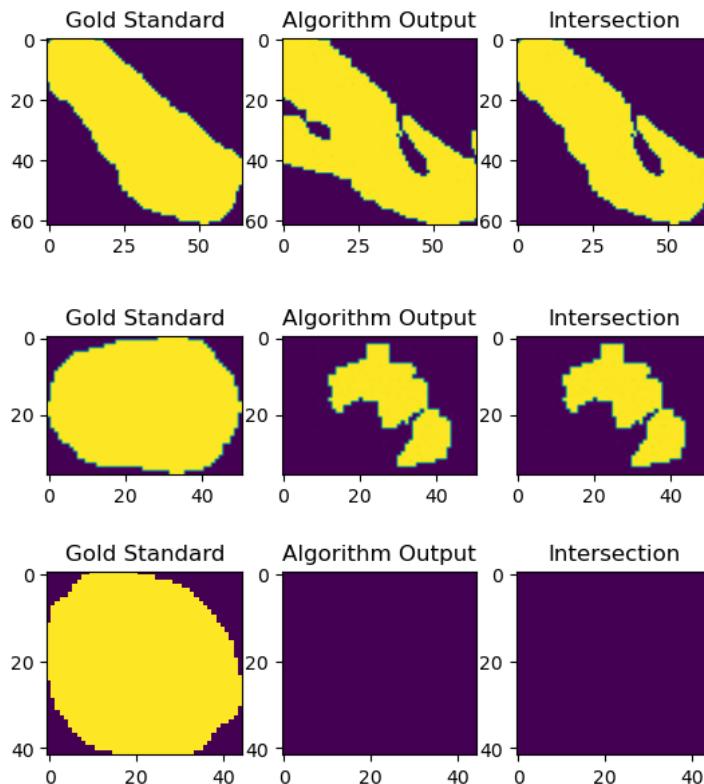


Figura 26 – Três exemplos de componentes conectados analisados. Na primeira coluna o componente na máscara de referência (padrão ouro). Na segunda coluna a segmentação do algoritmo de (GONDIM et al., 2021) para o respectivo componente. Na terceira coluna a intersecção entre a primeira e a segunda coluna.

Foi observado que no método proposto neste trabalho era menos comum a aparição de falsos negativos. Além disso, algumas imagens do conjunto de testes apresentavam concavidades no formato da matriz óssea e o método de (GONDIM et al., 2021) não reagiu bem para tais imagens, conforme mostrado na Figura 27.

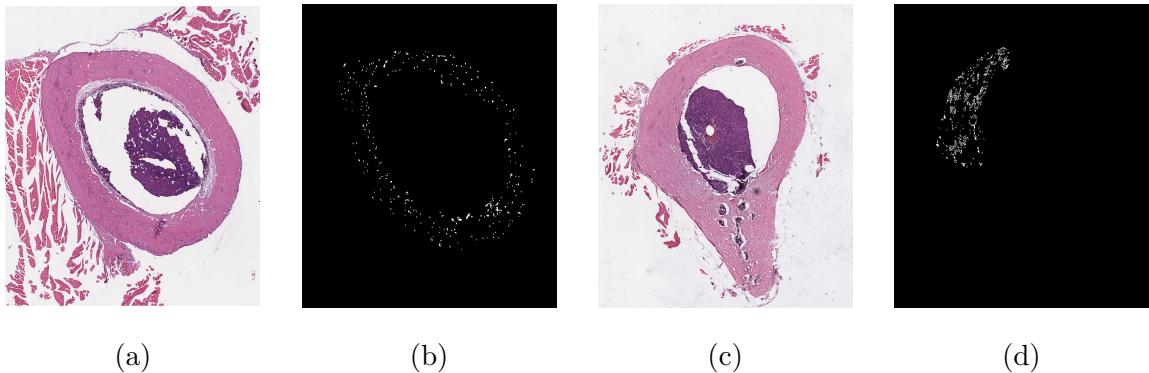


Figura 27 – Exemplo de falha do algoritmo proposto por (GONDIM et al., 2021). Apesar de funcionar para imagens com matriz óssea em um formato convexo (sub-figuras (a) e (b)) o método falha ao trabalhar com imagens que apresentem concavidades na matriz óssea (sub-figuras (c) e (d)).

Os resultados obtidos em (GONDIM et al., 2021) mostram acurácia e especificidade próximas a 96%, sensibilidade acima de 80% e cerca de 90% de *f1-score*. Porém as marcações dos especialistas não seguiram a forma exata das estruturas, assim, foram considerados como verdadeiros positivos os canais cuja marcação do especialista estivesse ao menos 20% contida na marcação feita pelo método.

Não fica claro se as imagens utilizadas em (GONDIM et al., 2021) continham imagens de diferentes regiões da diáfise femoral – região intermediária do fêmur. Devido à anatomia do osso, imagens provenientes da região central da diáfise resultam em uma matriz óssea mais arredondada, semelhante à Figura 27 (a), enquanto cortes mais próximos às extremidades da diáfise podem resultar em imagens cuja matriz óssea apresenta outros formatos, como o observado na Figura 27 (c).

A análise por canal – descrita na seção 4.2 – da reprodução do algoritmo proposto por (GONDIM et al., 2021) apresentou um *f1-score* de aproximadamente 17,5% e uma Intersecção sobre União de 9,6%. A Tabela 3 mostra os resultados obtidos pela análise por canal para cada um dos métodos.

Método	Precisão	F1-Score	Sensibilidade	IoU
FCN ($p = 0.10$)	0.826	0.849	0.872	0.737
Algoritmo (GONDIM et al., 2021)	0.619	0.175	0.102	0.095

Tabela 3 – Médias de precisão, *f1-score*, sensibilidade e Intersecção sobre União para cada os métodos testados.

CAPÍTULO 5

Conclusão

Neste trabalho foi explorado um novo método de segmentação de canais vasculares em imagens histológicas de tecido ósseo utilizando uma rede neural completamente convolucional. A rede foi treinada com um conjunto de imagens próprio criado a partir de marcações manuais de um especialista em histologia. Além disso, o conjunto de imagens utilizado foi disponibilizado publicamente.

O método testado mostrou-se capaz de executar a tarefa de segmentação de canais vasculares em imagens histológicas WSI de tecido ósseo. Apresentou bons valores de acurácia e especificidade e valores superiores de precisão, sensibilidade e *f1-score* e mostrou-se mais robusto e preciso em comparação com outros trabalhos encontrados na literatura. Entretanto o trabalho apresenta algumas limitações.

Não foram testados outros tamanhos para a quebra da imagem WSI em imagens menores. Uma quebra em tamanhos menores possivelmente melhoraria os resultados visto que poderiam haver mais sub-imagens que não apresentam canais. Dessa forma poderíamos descartá-las e treinar a rede com um conjunto de dados mais equilibrado entre *pixels* que são canais e *pixels* que não são canais.

Outra limitação do conjunto de dados foi o fato de não ter sido feita uma análise de concordância sobre as marcações realizadas pelo especialista. Uma análise de concordância com o apoio de um segundo especialista na área seria interessante pois, ao realizar as marcações, o especialista traz consigo um viés de subjetividade. Com a análise de concordância poderíamos ter um conjunto de dados com marcações mais confiáveis, o que traria um resultado mais próximo da realidade.

Conclui-se portanto que a segmentação de canais vasculares em imagens histológicas de tecido ósseo em imagens de lâmina inteira pode ser realizada utilizando redes neurais completamente convolucionais desde que haja um conjunto de imagens apropriado para o treinamento. Para trabalhos futuros almeja-se aprimorar o método focando nas limitações apresentadas acima. Além disso, pretende-se utilizar o método para segmentar canais vasculares em imagens sequenciais de tecido ósseo a fim de se modelar a rede vascular desse tecido tornando viável a realização de análises mais profundas sobre a rede vascular

óssea para pesquisadores da área da histologia.

5.0.1 Contribuições em Produções Bibliográficas

O artigo *Automatic Segmentation of Bone Canals on Histological Images Using Fully Convolutional Neural Networks* foi submetido para a revista Social *Biomedical Signal Processing and Control*. Nesse artigo foram reunidas as principais propostas e resultados descritos nesta dissertação.

Referências

Adobe Inc. **Adobe Photoshop**. 2016. Disponível em: <<https://www.adobe.com/products/photoshop.html>>.

ARC. **The Perceptron**. 2018. Accessed: 2023-12-03. Disponível em: <<https://towardsdatascience.com/the-perceptron-3af34c84838c>>.

BLOCK, H.-D. The perceptron: A model for brain functioning. In: **Reviews of Modern Physics**, APS, v. 34, n. 1, p. 123, 1962.

CHOWDHURY, M. E. H.; RAHMAN, T.; KHANDAKAR, A.; MAZHAR, R.; KADIR, M. A.; MAHBUB, Z. B.; ISLAM, K. R.; KHAN, M. S.; IQBAL, A.; EMADI, N. A.; REAZ, M. B. I.; ISLAM, M. T. Can ai help in screening viral and covid-19 pneumonia? **IEEE Access**, v. 8, p. 132665–132676, 2020.

ESTEVA, A.; CHOU, K.; YEUNG, S.; NAIK, N.; MADANI, A.; MOTTAGHI, A.; LIU, Y.; TOPOL, E.; DEAN, J.; SOCHER, R. Deep learning-enabled medical computer vision. **NPJ digital medicine**, Nature Publishing Group, v. 4, n. 1, p. 1–9, 2021.

FELDMAN, A. T.; WOLFE, D. Tissue processing and hematoxylin and eosin staining. **Histopathology: methods and protocols**, Springer, p. 31–43, 2014.

GÉRON, A. **Mãos à Obra: Aprendizado de Máquina com Scikit-Learn & TensorFlow**. [S.l.]: Alta Books, 2019.

GONDIM, P. H. C. C.; LIMIRIO, P. H. J. O.; ROCHA, F. S.; BATISTA, J. D.; DECHICHI, P.; TRAVENÇOLO, B. A. N.; BACKES, A. R. Automatic segmentation of bone canals in histological images. **Journal of Digital Imaging**, Springer, v. 34, n. 3, p. 678–690, 2021.

GONZÁLEZ-HIDALGO, M.; GUERRERO-PEÑA, F. A.; HEROLD-GARCÍA, S.; CAPÓ, A. Jaume-i; MARRERO-FERNÁNDEZ, P. D. Red blood cell cluster separation from digital images for use in sickle cell disease. **IEEE Journal of Biomedical and Health Informatics**, v. 19, n. 4, p. 1514–1525, 2015.

JULIA, R. S. et al. Histological image segmentation of the bone vascular network through deep learning. Universidade Federal de Uberlândia, 2021.

JUNQUEIRA, L. C.; CARNEIRO, J. Histologia básica. In: **Histologia básica**. [S.l.: s.n.], 1985. p. 512–512.

- Labati, R. D.; Piuri, V.; Scotti, F. All-idb: The acute lymphoblastic leukemia image database for image processing. In: **2011 18th IEEE International Conference on Image Processing**. [S.l.: s.n.], 2011. p. 2045–2048.
- LINHARES, C. R. B. et al. Melhoria da regeneração óssea em ratos diabéticos tipo i submetidos à oxigenoterapia hiperbárica. Universidade Federal de Uberlândia, 2019.
- LINHARES, C. R. B.; RABELO, G. D.; LIMIRIO, P. H. J. O.; VENÂNCIO, J. F.; SILVA, I. G. R.; DECHICHI, P. Automated bone healing evaluation: New approach to histomorphometric analysis. **Microscopy Research and Technique**, Wiley Online Library, v. 85, n. 10, p. 3339–3346, 2022.
- LIU, Z.-Q.; LIEW, H. L.; CLEMENT, J. G.; THOMAS, C. D. L. Bone image segmentation. **IEEE Transactions on biomedical engineering**, IEEE, v. 46, n. 5, p. 565–573, 1999.
- LONG, J.; SHELHAMER, E.; DARRELL, T. Fully convolutional networks for semantic segmentation. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2015. p. 3431–3440.
- MIKLOSIK, A.; EVANS, N. Impact of big data and machine learning on digital transformation in marketing: A literature review. **Ieee Access**, IEEE, v. 8, p. 101284–101292, 2020.
- MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. **Sistemas inteligentes-Fundamentos e aplicações**, Manole, v. 1, n. 1, p. 32, 2003.
- MUELLER, J. P.; MASSARON, L. **Deep Learning for dummies**. [S.l.]: John Wiley & Sons, 2019.
- PAULLADA, A.; RAJI, I. D.; BENDER, E. M.; DENTON, E.; HANNA, A. Data and its (dis) contents: A survey of dataset development and use in machine learning research. **Patterns**, Elsevier, v. 2, n. 11, p. 100336, 2021.
- RAWAT, W.; WANG, Z. Deep convolutional neural networks for image classification: A comprehensive review. **Neural computation**, MIT Press, v. 29, n. 9, p. 2352–2449, 2017.
- RIEBESELL, J. **Convolution Operator**. <<https://tikz.net/conv2d/>>. Accessed: 2023-12-03.
- RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: SPRINGER. **Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18**. [S.l.], 2015. p. 234–241.
- SANTOS, D. F. D. d. et al. Automated segmentation of tumor regions from oral histological whole slide images using fully convolutional neural networks. Universidade Federal de Uberlândia, 2022.
- SANTOS, D. F. D. dos; FARIA, P. R. de; TRAVENÇOLO, B. A. N.; NASCIMENTO, M. Z. do. Influence of data augmentation strategies on the segmentation of oral histological images using fully convolutional neural networks. **Journal of Digital Imaging**, Springer Science and Business Media LLC, apr 2023.

- SANTOS, D. F. dos; FARIA, P. R. de; TRAVENÇOLO, B. A.; NASCIMENTO, M. Z. do. Automated detection of tumor regions from oral histological whole slide images using fully convolutional neural networks. **Biomedical Signal Processing and Control**, Elsevier BV, v. 69, p. 102921, aug 2021.
- SIDDIQUE, N.; PAHEDING, S.; ELKIN, C. P.; DEVABHAKTUNI, V. U-net and its variants for medical image segmentation: A review of theory and applications. **Ieee Access**, IEEE, v. 9, p. 82031–82057, 2021.
- SIMON, P. **Too Big to Ignore: The Business Case for Big Data**. Wiley, 2013. (Wiley and SAS Business Series). ISBN 9781118642108. Disponível em: <<https://books.google.com.br/books?id=Dn-Gdoh66sgC>>.
- SRINIVASAN, S.; VINAYAKUMAR, R.; ARUNACHALAM, A.; ALAZAB, M.; SOMAN, K. Durld: Malicious url detection using deep learning-based character level representations. In: **Malware analysis using artificial intelligence and deep learning**. [S.l.]: Springer, 2021. p. 535–554.
- UDDIN, M. F. Addressing accuracy paradox using enhanched weighted performance metric in machine learning. In: **2019 Sixth HCT Information Technology Trends (ITT)**. [S.l.: s.n.], 2019. p. 319–324.
- VADAPALLI, P. **Ultimate Guide for Deep Learning with Neural Network in 2023**. <<https://www.upgrad.com/blog/guide-for-deep-learning-with-neural-network/>>. Accessed: 2023-12-03.
- VARGAS, A. C. G.; PAES, A.; VASCONCELOS, C. N. Um estudo sobre redes neurais convolucionais e sua aplicação em detecção de pedestres. In: SN. **Proceedings of the XXIX conference on graphics, patterns and images**. [S.l.], 2016. v. 1, n. 4.