Universidade Federal de São Carlos– UFSCar
Centro de Ciências Exatas e de Tecnologia– CCET
Departamento de Computação– DC
Programa de Pós-Graduação em Ciência da Computação– PPGCC

Wesley Nogueira Galvão

Segmentação de vasos sanguíneos com poucos dados via transferência de informação de forma

São Carlos
2025

Wesley Nogueira Galvão

Segmentação de vasos sanguíneos com poucos dados via transferência de informação de forma

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Centro de Ciências Exatas e de Tecnologia da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação.

Área de concentração: Visão Computacional

Orientador: Prof. Dr. Cesar Henrique Comin

São Carlos
2025

Galvão, Wesley Nogueira

Segmentação de vasos sanguíneos com poucos dados via transferência de informação de forma / Wesley Nogueira Galvão -- 2025.
76f.

Dissertação (Mestrado) - Universidade Federal de São Carlos, campus São Carlos, São Carlos
Orientador (a): Cesar Henrique Comin
Banca Examinadora: Agma Juci Machado Traina, Jurandy Gomes de Almeida Junior
Bibliografia

1. Segmentação de vasos sanguíneos. 2. Transferência de domínio. 3. Aprendizado com poucos exemplos. I. Galvão, Wesley Nogueira. II. Título.

Ficha catalográfica desenvolvida pela Secretaria Geral de Informática (SIn)

DADOS FORNECIDOS PELO AUTOR

Bibliotecário responsável: Arildo Martins - CRB/8 7180

UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia
Programa de Pós-Graduação em Ciência da Computação

Relatório de Defesa de Dissertação
Candidato: Wesley Nogueira Galvão

Aos 24/10/2025, às 14:00, realizou-se na Universidade Federal de São Carlos, nas formas e termos do Regimento Interno do Programa de Pós-Graduação em Ciência da Computação, a defesa de dissertação de mestrado sob o título: Segmentação de vasos sanguíneos com poucos dados via transferência de informação de forma, apresentada pelo candidato Wesley Nogueira Galvão. Ao final dos trabalhos, a banca examinadora reuniu-se em sessão reservada para o julgamento, tendo os membros chegado ao seguinte resultado:

Participantes da Banca	Função	Instituição	Resultado	Resultado Final
Prof. Dr. Cesar Henrique Comin	Presidente	UFSCar	APROVADO	APROVADO
Profa. Dra. Agma Juci Machado Traina	Titular	USP	APROVADO	
Prof. Dr. Jurandy Gomes de Almeida Junior	Titular	UFSCar	APROVADO	

Parecer da Comissão Julgadora*:

Encerrada a sessão reservada, o presidente informou ao público presente o resultado. Nada mais havendo a tratar, a sessão foi encerrada e, para constar, eu, Ivan R Silva, representante do Programa de Pós-Graduação em Ciência da Computação, lavrei o presente relatório, assinado por mim e pelos membros da banca examinadora.

gov.br Documento assinado digitalmente
CESAR HENRIQUE COMIN
Data: 25/10/2025 15:10:57-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Cesar Henrique Comin

Representante do PPG: Ivan R Silva

Profa. Dra. Agma Juci Machado Traina

Prof. Dr. Jurandy Gomes de Almeida Junior

Certifico que a defesa realizou-se com a participação à distância do(s) membro(s) Cesar Henrique Comin, Agma Juci Machado Traina, Jurandy Gomes de Almeida Junior e, depois das arguições e deliberações realizadas, o(s) participante(s) à distância está(ao) de acordo com o conteúdo do parecer da banca examinadora redigido neste relatório de defesa.

gov.br Documento assinado digitalmente
CESAR HENRIQUE COMIN
Data: 25/10/2025 15:11:33-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Cesar Henrique Comin

(X) Não houve alteração no título ( ) Houve alteração no título. O novo título passa a ser:

Observações:
a) Se o candidato for reprovado por algum dos membros, o preenchimento do parecer é obrigatório.
b) Para gozar dos direitos do título de Mestre ou Doutor em Ciência da Computação, o candidato ainda precisa ter sua dissertação ou tese homologada pelo Conselho de Pós-Graduação da UFSCar.

Dedico este trabalho aos meus pais, Ivaldo Galvão e Rosângela Galvão, por todo apoio incondicional e por terem acreditado na educação dos seus filhos como um instrumento de transformação de vida.

Agradecimentos

Agradeço a Deus pela oportunidade de seguir adiante com a minha capacitação acadêmica e profissional.

Aos meus pais que, novamente, foram pontos de apoio e incentivo, o que tornou esse mestrado possível.

Aos meus irmãos, William e Wender, que contribuíram com momentos de descontração durante esse período.

A minha querida companheira Raissa, que me encorajou e apoiou em todos os momentos, com sua paciência, compreensão e carinho.

Ao meu professor e orientador Dr. César Henrique Comin, pela confiança em me entregar mais este projeto e por me orientar para os melhores caminhos, não apenas neste trabalho, mas em minha formação como pesquisador.

"Nós só conseguimos ver um pouco adiante, mas podemos ver que há muito a fazer."
(Alan Turing)

Resumo

A segmentação de vasos sanguíneos é uma tarefa essencial em diversas áreas da biomedicina, como oftalmologia e neurologia. No entanto, esta tarefa impõe um importante desafio devido à escassez de grandes conjuntos de dados anotados, o que, consequentemente, torna oneroso o trabalho de anotação manual de novas amostras para o treinamento de modelos de aprendizado profundo eficientes. Técnicas de transferência de aprendizado têm-se mostrado promissoras para superar essas limitações, permitindo o reaproveitamento de representações aprendidas em um domínio para melhorar o desempenho em outro, mesmo quando há poucos exemplos disponíveis. Nesse contexto, este trabalho investiga a transferência de representações de forma a partir de um domínio sintético para a segmentação de imagens médicas em regime de poucas amostras. Nossa hipótese central é que o aprendizado de padrões de formas similares às dos vasos, como as características tubulares e ramificações, pode levar a modelos mais robustos e eficientes em termos de dados. Para isso, introduzimos o VessShape, uma metodologia para gerar um conjunto de dados sintético em larga escala, projetado para inserir um forte viés de forma em modelos de segmentação. As imagens do VessShape combinam geometrias tubulares geradas proceduralmente com uma ampla variedade de texturas, incentivando os modelos a aprenderem a forma ao invés de características de aparência. Redes neurais pré-treinadas com o VessShape foram refinadas e avaliadas em dois conjuntos de dados reais de domínios distintos. Os resultados demonstraram que a abordagem aprimora o desempenho de segmentação em cenários com poucos exemplos. Adicionalmente, os modelos demonstram uma significativa capacidade em *zero-shot learning*, sendo aptos a segmentar vasos em domínios não vistos sem qualquer treinamento específico no domínio de destino. Tais resultados sustentam que o pré-treinamento com um forte viés de forma constitui uma estratégia efetiva para contornar a escassez de dados e aprimorar a generalização na segmentação de vasos sanguíneos.

Palavras-chave: Segmentação de vasos sanguíneos; viés de forma; transferência de do-

mínio; aprendizado com poucos exemplos.

Abstract

Blood vessel segmentation is an essential task in various areas of biomedicine, such as ophthalmology and neurology. However, this task poses a significant challenge due to the scarcity of large annotated datasets, which consequently makes the manual annotation of new samples for training efficient deep learning models a costly endeavor. Transfer learning techniques have shown promise in overcoming these limitations, allowing the reuse of representations learned in one domain to improve performance in another, even when few examples are available in the target domain. The central hypothesis indicates that leveraging a shape prior of vessel-like forms, such as their tubular and branching characteristics, can lead to more robust and data-efficient models. In this context, this dissertation investigates the transfer of shape representations from a synthetic domain to the segmentation of medical images in a few-shot regime. To this end, we introduce VessShape, a methodology for generating a large-scale synthetic dataset designed to instill a strong shape bias in segmentation models. VessShape images combine procedurally generated tubular geometries with a wide variety of textures, encouraging models to learn shape cues over appearance features. The models pre-trained with VessShape were then fine-tuned and evaluated on two real-world datasets from different domains. The results demonstrate that the approach achieves strong segmentation performance in few-shot scenarios, requiring only a small number of samples for fine-tuning. Additionally, the models demonstrate a significant zero-shot learning capability, proving able to segment vessels in unseen domains without any target-specific training. These results support that pre-training with a strong shape bias constitutes an effective strategy to overcome data scarcity and enhance generalization in blood vessel segmentation.

Keywords: Blood vessel segmentation, shape bias, domain adaptation, few-shot learning.

Lista de ilustrações

Lista de tabelas

Lista de siglas

CNN Rede Neural Convolucional

CNNs Redes Neurais Convolucionais

DRIVE Digital Retinal for Vessel Extraction

FCNN Rede Neural Totalmente Convolucional

FT Refinamento de Parâmetros

FSL Aprendizado com Poucos Exemplos

GPU Unidade de Processamento Gráfico

MRI Imagem por Ressonância Magnética

OSL Aprendizado com um Único Exemplo

ResNets Redes Neurais Residuais

SIN Stylized-ImageNet

TL Transferência de Aprendizado

WMH Hiperintensidade de Matéria Branca

ZSL Aprendizado com Zero Exemplo

Sumário

Capítulo 1

Introdução

1.1 Contextualização e problemática

Vasos sanguíneos são estruturas tubulares que desempenham um papel importante na manutenção da vida de muitos seres vivos, sendo responsáveis pela distribuição de oxigênio e nutrientes para os tecidos do corpo. A análise de imagens de regiões vascularizadas é uma tarefa essencial em diversas áreas da medicina, como oftalmologia, neurologia e cardiologia, auxiliando no diagnóstico e tratamento de doenças vasculares. Diversos trabalhos têm sido desenvolvidos para a segmentação e análises avançadas de vasos sanguíneos, e essas aplicações não se restringem apenas às estruturas vasculares do fundo de retina (QIN; CHEN, 2024), mas exploram também outras regiões anatômicas, como o cérebro (GONI et al., 2022) e artérias coronárias (LIANG et al., 2020). Tampouco se limitam ao escopo do corpo humano, abrangendo outras espécies, como camundongos (OUELLETTE et al., 2020).

Dada a demanda por análises precisas e automatizadas de imagens médicas, a segmentação semântica dessas estruturas tem sido um tópico de pesquisa ativo, com o objetivo de desenvolver métodos eficazes para a identificação automática de vasos sanguíneos em distintos tecidos do corpo humano e de outros animais. No entanto, essa tarefa apresenta desafios consideráveis, especialmente pelo oneroso trabalho de rotulação manual das máscaras de segmentação, comumente chamado de padrão-ouro, nas imagens, que muitas vezes requer o empenho de especialistas.

Consequentemente, há uma escassez de grandes conjuntos de dados com amostras anotadas, o que limita a capacidade de treinamento de modelos de aprendizado profundo para a segmentação de vasos sanguíneos. Por exemplo, conjuntos como Digital Retinal for Vessel Extraction (DRIVE) (STAAL et al., 2004) e CHASE_DB1 (FRAZ et al.,

2012) possuem algumas dezenas de imagens anotadas. Por outro lado, trabalhos mais recentes, como os de (SILVA et al., 2025) e (FREITAS-ANDRADE et al., 2022), têm contribuído para o desenvolvimento de novos conjuntos de dados de vasos sanguíneos, mas ainda há uma carência de amostras anotadas em larga escala. Adicionalmente, a eficácia dos modelos de segmentação é limitada por notáveis mudanças de domínio que ocorrem entre diferentes modalidades de imagem. Por exemplo, características como textura, densidade e calibre vascular variam significativamente entre fotografias de retina e imagens de microscopia do córtex de camundongos, o que dificulta a generalização de um modelo treinado em um domínio quando este é aplicado a outro.

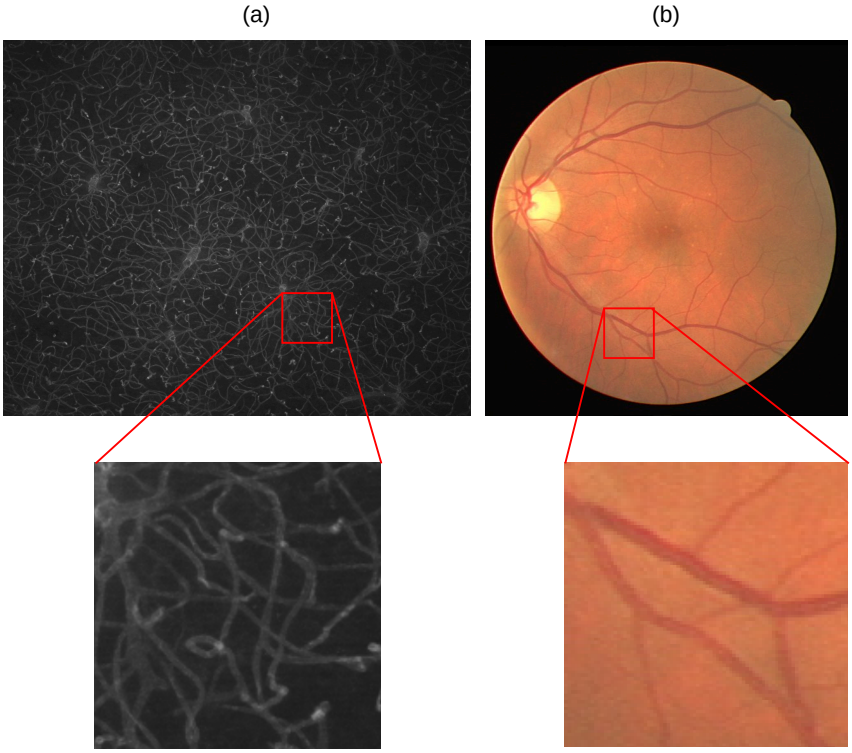
Técnicas de Transferência de Aprendizado (TL) (do inglês, ***T****ransfer* ***L****earning*) mostram-se promissoras para superar essas limitações, pois permitem o reaproveitamento de representações aprendidas em um domínio para melhorar o desempenho em outro, mesmo quando há poucos exemplos disponíveis no domínio de destino. Tarefas de adaptação de domínio, juntamente com Refinamento de Parâmetros (FT) (do inglês, ***F****ine-****T****uning*), são capazes de adaptar modelos pré-treinados, aproveitando as diversas representações de alto nível aprendidas em tarefas de classificação de imagens naturais, como *ImageNet* (Jia Deng et al., 2009), para a segmentação de imagens biomédicas, como proposto por (ZOETMULDER et al., 2022).

Dentre as possíveis representações para a segmentação de vasos sanguíneos, a sua forma, especificamente o seu aspecto tubular e ramificado, destaca-se por ser uma característica invariante que persiste através de diferentes domínios. Por exemplo, embora haja notáveis diferenças de textura e aparência entre imagens de retina e do córtex cerebral, a forma tubular característica dos vasos se mantém consistente em ambos, como pode ser observado na Figura 1. Essa consistência geométrica beneficia diretamente a tarefa de segmentação, tornando a forma uma pista fundamental de grande utilidade.

Estudos como os de (GEIRHOS et al., 2019) e (ISLAM et al., 2021) mostraram que as CNNs (do inglês, ***C****onvolutional* ***N****eural* ***N****etworks*) tendem a aprender características de textura em detrimento de formas geométricas, o que pode limitar sua capacidade de generalização entre diferentes domínios de imagem. Ao propor uma abordagem que estiliza as texturas das imagens, (GEIRHOS et al., 2019) sugere que redes treinadas com ênfase em formas geométricas tendem a ter melhor desempenho e maior robustez. Com isso, a transferência de representações de forma previamente adquiridas em distintos domínios de imagem pode ser uma estratégia para melhorar a segmentação de vasos sanguíneos quando há poucos exemplos disponíveis, e isso é o que este trabalho se propõe a investigar.

Com base nessa observação, é importante que modelos de segmentação de vasos sanguíneos sejam invariantes à textura e outros aspectos de aparência, mas sensíveis a formas geométricas, para obter uma boa generalização entre diferentes domínios de imagem. Neste sentido, propomos a seguinte hipótese:

Figura 1 – Exemplos de vasos sanguíneos em imagens de microscopia do córtex de camundongos (a) e fotografia da retina (b). Apesar das diferenças de textura, densidade e tortuosidade, a forma característica dos vasos sanguíneos se mantém similar.



Fonte: Próprio autor

> *Modelos treinados em domínios de origem com ênfase em forma, e refinados no domínio de imagens de vasos sanguíneos, terão um desempenho da métrica Dice superior aos modelos treinados do zero no domínio de imagens de vasos sanguíneos, devido à reutilização de representações geométricas adquiridas previamente.*

Portanto, para investigar a efetividade da transferência de representações de forma em regime de poucos dados disponíveis, este trabalho apresenta uma contribuição central: o VessShape, uma metodologia para gerar conjuntos de dados sintéticos 2D em larga escala e dotados de formas similares às de vasos sanguíneos. O VessShape foi desenvolvido para o pré-treinamento de modelos de segmentação semântica, combinando geometrias tubulares com uma ampla variedade de texturas para instilar um viés de forma e desencorajar a memorização de padrões de textura. A investigação é validada por um protocolo sistemático de refinamento e avaliação, que mede o desempenho dos modelos pré-treinados em conjuntos reais de vasos sanguíneos com uma abordagem de inserção progressiva de

dados.

1.2 Objetivos

O objetivo geral deste trabalho é investigar a efetividade da transferência de representações de forma previamente adquiridas por redes neurais treinadas em um domínio de imagens sintéticas e refinadas em imagens naturais, próprias para a tarefa de segmentação semântica de vasos sanguíneos. Pretende-se avaliar a capacidade de generalização dessas redes na identificação de estruturas tubulares e alongadas, que são características inerentes aos vasos sanguíneos, independentemente da textura ou aparência das imagens, mas concentrando-se em vieses de forma. Para alcançar esse objetivo, propõe-se a realização dos seguintes objetivos específicos:

1. **Estudo de técnicas de transferência de aprendizado com poucos exemplos**: Nesta tarefa, buscou-se compreender o funcionamento de técnicas de transferência de aprendizado, em especial, aquelas que lidam com poucos exemplos, com a finalidade de identificar as melhores práticas para a aplicação dessas técnicas nesta pesquisa.

2. **Desenvolvimento de um conjunto de dados sintéticos que enaltece o viés de forma de estruturas tubulares**:

   Propõe-se a criação de um conjunto de dados sintéticos de imagens, denominado *VessShape*, com ênfase na geometria de vasos sanguíneos, para ser utilizado em tarefas de transferência de aprendizado. Assim, será possível avaliar a contribuição desse conjunto na generalização de modelos de segmentação semântica de vasos sanguíneos, facilitando o aprendizado de características de forma, especialmente em cenários com poucos exemplos.

3. **Desenvolvimento de uma metodologia para avaliar a efetividade da transferência de representações geométricas de vasos sanguíneos**:

   Neste objetivo, propõe-se a criação de uma metodologia para avaliar a capacidade da transferência de representações geométricas de vasos sanguíneos nas seguintes frentes:

   - ❏ Diferentes arquiteturas codificador-decodificador, comparando sua capacidade de generalização.
   - ❏ A influência do domínio de origem sintético (*VessShape*), avaliando como a transferência de conhecimento de forma impacta o desempenho na segmentação de vasos sanguíneos, em contraste com um modelo treinado diretamente nos dados de destino.

- ❏ Refinamento progressivo dos modelos, partindo de técnicas de aprendizado com um único exemplo e com poucos exemplos.

1.3 Organização do trabalho

Este trabalho está organizado nos seguintes capítulos:

O Capítulo 2 apresenta os conceitos de aprendizado profundo, arquiteturas de redes neurais, como CNNs e ResNet. Também são abordados conceitos de transferência de aprendizado e métricas de avaliação de segmentação semântica de vasos sanguíneos.

O Capítulo 3 explora estudos relacionados ao tema deste trabalho, destacando trabalhos sobre aprendizado com poucos exemplos e transferência de aprendizado em imagens médicas.

No Capítulo 4 são descritos o ferramental tecnológico, as bases de dados e a metodologia proposta, com ênfase no detalhamento dos processos de treinamento e refinamento dos modelos.

O Capítulo 5, por sua vez, apresenta as avaliações quantitativas e qualitativas das tarefas de treinamento e refinamento dos modelos de segmentação.

Por fim, o Capítulo 6, apresenta as considerações finais da pesquisa, consolidando os resultados obtidos e apontando as direções para trabalhos futuros.

Capítulo 2

Fundamentação Teórica

2.1 Aprendizado profundo

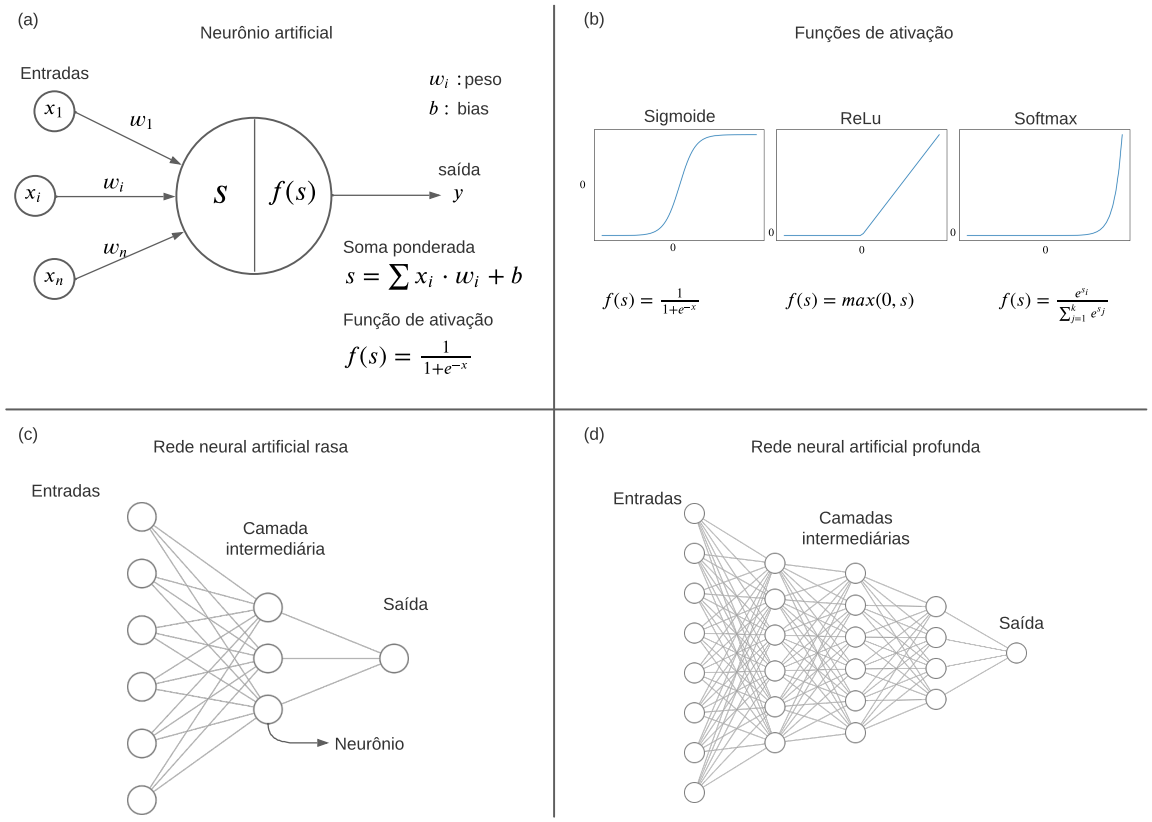
O Aprendizado Profundo (AP), comumente conhecido com *Deep Learning*, é uma subárea da inteligência artificial, especificamente do escopo de Aprendizado de Máquina, que tem atraído crescente interesse devido a excelentes resultados em diferentes aplicações, como Visão computacional, Processamento de Linguagem Natural, sistemas de recomendação e, mais recentemente, modelos generativos capazes de gerar imagens, textos complexos e áudio.

Fundamentalmente, o Aprendizado Profundo baseia a sua estruturação em Redes Neurais Artificiais (ANN), que são estruturas de mapeamento (LEK; PARK, 2008) bioinspiradas. Essas redes têm como unidade computacional não linear o neurônio. Como mostrado na Figura 2 (a), o neurônio é uma função que recebe como entrada um conjunto de valores escalares, os quais passam por uma operação de soma ponderada e uma função de ativação (Figura 2 (b)), que por sua vez gera uma saída.

Os neurônios podem ser arranjados em sucessivas camadas, gerando a arquitetura Perceptron Multicamada (MLP). Sendo assim, na arquitetura MLP a rede pode possuir uma camada intermediária — popularmente conhecida também como camada oculta — ou até múltiplas camadas intermediárias com diversos neurônios por camada, tornando a rede profunda, como mostrado respectivamente pelas Figuras 2 (c) e (d).

A arquitetura apresentada permite o fluxo de informação de forma unidirecional (LEK; PARK, 2008), partindo da camada de entrada até a camada de saída através das camadas intermediárias. Esse passo é chamado também de *foward-propagation*. Nessa configuração, cada neurônio de uma camada está completamente conectado com os neurônios da camada adjacente. Essas conexões são equivalentes aos pesos utilizados no processo com-

Figura 2 – Conceitos básicos de Redes Neurais. (a) Representação de um neurônio artificial, detalhando suas entradas, pesos, soma ponderada e função de ativação. (b) Exemplos das funções de ativação. (c) Arquitetura de uma rede neural rasa, com uma única camada intermediária. (d) Arquitetura de uma rede neural profunda, com múltiplas camadas intermediárias.



Fonte: Adaptado de (DONG; WANG; ABBAS, 2021)

putacional.

A tarefa de aprendizado pode acontecer de forma supervisionada ou não supervisionada. No aprendizado supervisionado, a rede neural realiza o ajuste do modelo sobre os dados de entrada que possuem as saídas desejadas conhecidas. Durante o treinamento, a rede passa por iterações do cálculo da saída de acordo com os dados na camada de entrada, assim como o reajuste dos pesos a cada passo iterativo, uma vez que se conheçam as saídas calculadas e desejadas (RUMELHART; HINTON; WILLIAMS, 1986). O objetivo do reajuste é minimizar o efeito do erro sobre a saída final, e isso é feito em cada neurônio de forma a encontrar o menor erro associado a sua saída. Esse processo é chamado de *backpropagation*.

Portanto, o método *backpropagation* calcula retroativamente — da camada de saída em direção à camada de entrada—- o sinal de erro para ajustar os parâmetros da rede por meio do cômputo do gradiente de cada neurônio em relação a sua entrada (LECUN; BENGIO; HINTON, 2015). Essa técnica é uma forma de derivação automática (GOODFELLOW;

BENGIO; COURVILLE, 2016) que permite que a rede neural seja capaz de produzir mapas de características a partir dos parâmetros reajustados.

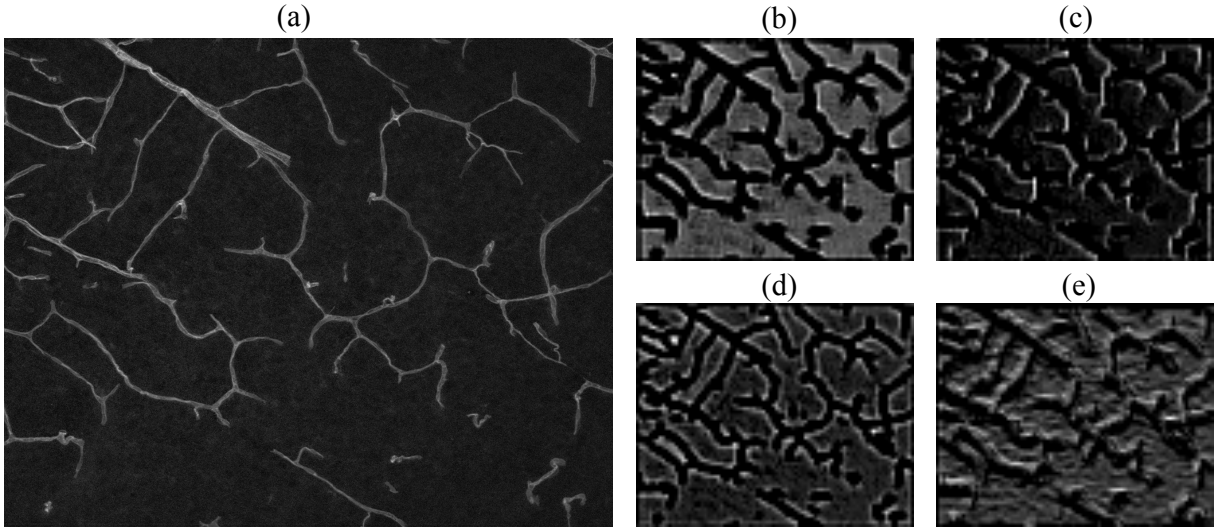
2.2 Arquiteturas de Redes Neurais Profundas

2.2.1 Redes Neurais Convolucionais

As redes neurais profundas são conhecidas por serem eficientes extratoras de características e padrões durante o processo de treinamento. Essa propriedade foi explorada pelas Redes Neurais Convolucionais (CNNs) (do inglês, ***C**onvolutional **N**eural **N**etworks*), amplamente utilizadas para detecção automática de objetos, classificação de imagens, entre outras aplicações. Trabalhos como os de (FUKUSHIMA, 1980) e (LECUN et al., 1989) foram importantes na demonstração da efetividade de CNNs para extração de características e classificação de imagens.

As CNNs são compostas por camadas convolucionais que efetuam a operação de convolução entre uma imagem de entrada e os filtros de convolução, ou pesos, que são matrizes numéricas que se reajustam durante o treinamento para se tornarem melhores extratoras de características, a fim de contribuir com a saída esperada da rede. Essa operação resulta em uma nova matriz chamada de mapa de ativação, ou *feature map* do inglês, que são representações que contêm as características extraídas da imagem de entrada, como bordas, texturas ou formas, como mostrado na Figura 3. Uma rede pode ter diversas camadas convolucionais localizadas nas camadas intermediárias, que recebem como entrada os mapas de ativação das camadas anteriores e geram outros mapas com características mais complexas ao longo da profundidade da rede.

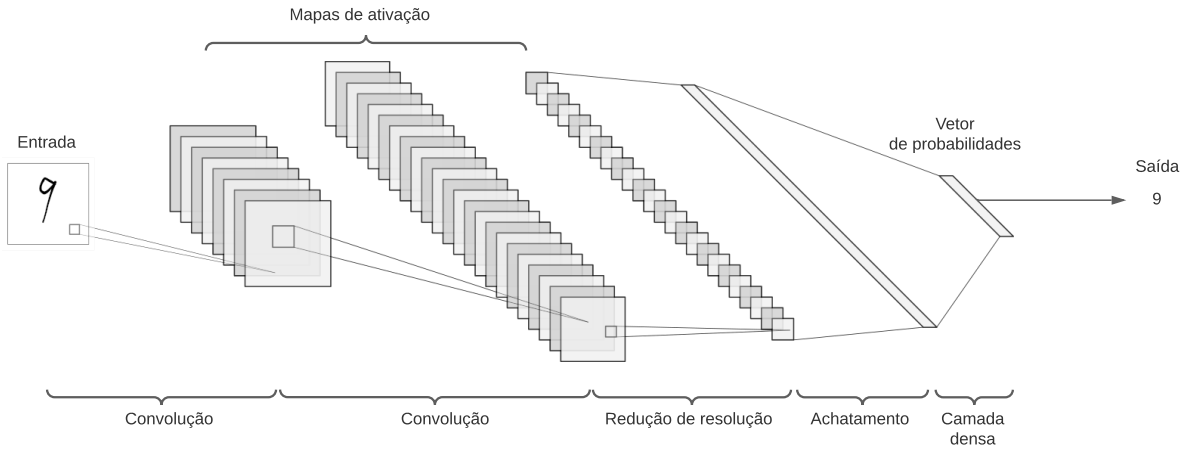
Figura 3 – Exemplo de mapas de ativação de uma U-Net. (a) Imagem original. (b)-(e) Mapas de ativação de alguns filtros da rede



Fonte: Próprio autor

A Figura 4 exemplifica uma CNN que realiza classificação de dígitos manuscritos, dada uma imagem de entrada. Além das camadas de convolução, há as operações de redução de dimensão, também conhecidas como *pooling*, que fazem a amostragem dos mapas mantendo os aspectos mais importantes. Isso ajuda a rede a lidar com o problema de *overfitting* e a aprender características em diferentes escalas da imagem.

Figura 4 – Exemplo de uma CNN para classificação de dígitos manuscritos



Fonte: Próprio autor

Ao final da rede, encontra-se a camada densa, ou totalmente conectada, crucial em tarefas de classificação. Esta camada é alimentada por um vetor de características, que é gerado a partir do achatamento (*flattening*) dos mapas de ativação da camada anterior. A função da camada densa é gerar um vetor de pontuações brutas, conhecidas como *logits*, onde cada pontuação corresponde a uma das classes de saída. Para que essas pontuações sejam convertidas em uma distribuição de probabilidades, aplica-se uma função de ativação, como a *Softmax*, na saída desta camada. O resultado é o vetor de probabilidades final, utilizado para classificar a imagem de entrada.

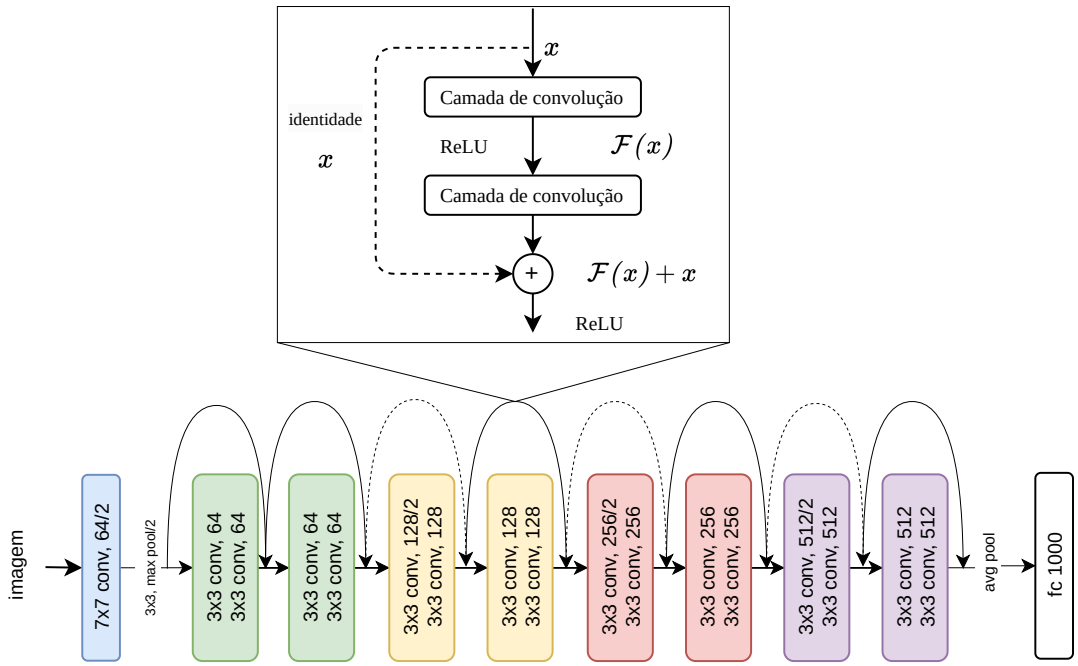
2.2.2 Redes Neurais Residuais

As CNNs se mostraram muito promissoras ao longo dos anos pelo bom desempenho em tarefas de visão computacional. Ao passo que os problemas se tornaram mais desafiadores, a necessidade de redes mais profundas para extração de características mais complexas se tornou evidente. Contudo, na prática, redes muito profundas tendem a apresentar o fenômeno de degradação de desempenho, que ocorre quando a adição de novas camadas não resulta em uma melhoria do desempenho, mas sim o aumento do erro de treinamento. Isso parece contraditório, uma vez que a adição de camadas deveria permitir melhor capacidade de representação do modelo.

O fenômeno ocorre devido à dificuldade de otimização de redes profundas, que sofrem com o desaparecimento do gradiente durante o treinamento. Isto é, o gradiente tende a

se tornar pequeno à medida que se propaga para as camadas iniciais da rede, dificultando a convergência do modelo. Estes problemas foram endereçados por (HE et al., 2016), que propuseram a arquitetura ResNets (do inglês, ***Res**idual **Net**works*). O fator de inovação das ResNets é a introdução de blocos residuais, que são módulos que permitem que a rede propague de maneira mais eficiente o gradiente durante o treinamento ao utilizar atalhos, do inglês *skip connections*, contornando as operações não-lineares das camadas intermediárias, evitando o desaparecimento do gradiente. A Figura 5 ilustra a arquitetura de uma ResNet, com destaque para os fluxos de atalho e o bloco residual.

Figura 5 – Representação da ResNet18



Fonte: Próprio autor

O bloco residual é o elemento central das ResNets, no qual, ao invés da rede aprender a se ajustar à função $\mathcal{H}(x)$, a rede aprende a função residual $\mathcal{F}(x) = \mathcal{H}(x) - x$, onde a saída é dada por:

$$y = \mathcal{F}(x, \mathcal{W}) + x \tag{1}$$

sendo $\mathcal{F}(x, \mathcal{W})$ a função residual a ser aprendida com os pesos $\mathcal{W}$, e x a entrada do bloco residual.

Dessa forma, o aprendizado residual proporciona melhor estabilidade aos gradientes, permitindo a construção de redes mais profundas. Além disso, os blocos residuais se ajustam à função identidade $F(x) \approx 0$, o que significa que a rede pode aprender a ignorar o bloco residual, caso não haja melhoria no desempenho, fazendo com que $y = x$.

As ResNets têm sido amplamente usadas como *backbones*, servindo como redes codificadoras pré-treinadas no conjunto de dados *ImageNet* (Jia Deng et al., 2009), graças a sua capacidade de retenção de representações, que podem ser usadas em diversas tarefas de visão computacional.

2.2.3 Modelos Codificador-Decodificador

Modelos codificador-decodificador, ou *encoder-decoder*, são arquiteturas de redes neurais profundas que são comumente utilizadas em tarefas de visão computacional. Uma outra tradução aceitável para esse termo é a rede de compressão-expansão. A arquitetura é composta por dois módulos principais: o codificador e o decodificador. O codificador é responsável por receber a imagem de entrada e gerar uma representação intermediária, isto é, um vetor de características que contém informações relevantes da imagem. O decodificador, por sua vez, recebe essa representação, realiza a expansão ou interpolação da imagem e gera a saída desejada, geralmente de mesma dimensão da imagem de entrada.

2.2.3.1 U-Net

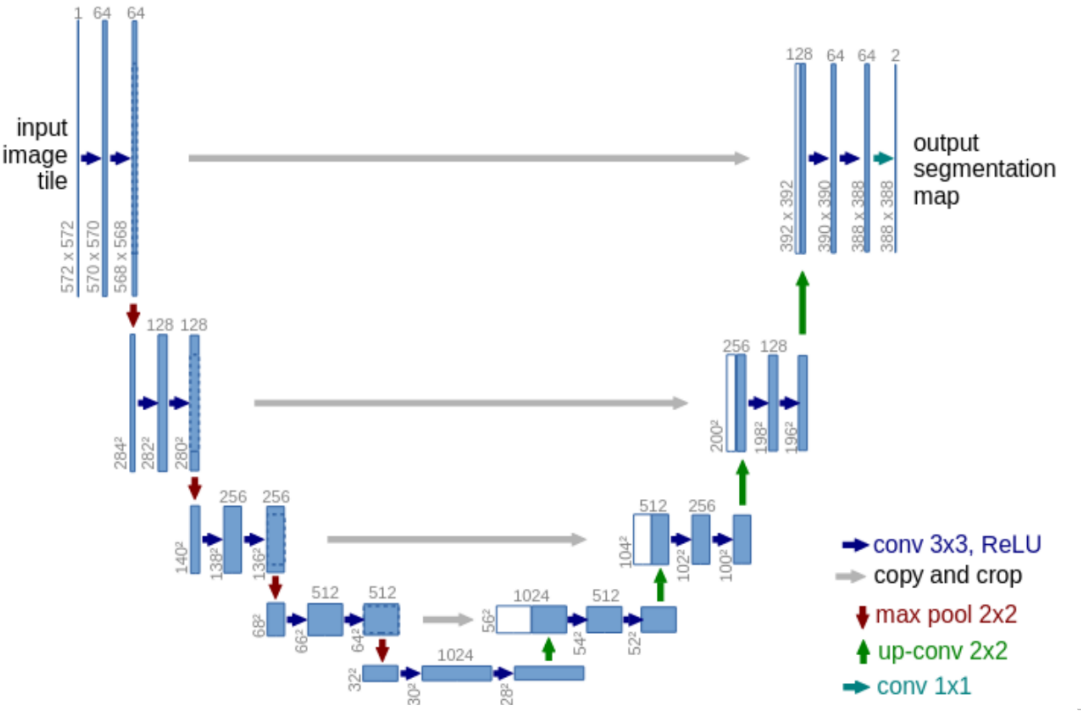
A arquitetura U-Net é comumente aplicada na segmentação de imagens biomédicas, e é uma Rede Neural Totalmente Convolucional (FCNN) (do inglês, ***F**ully **C**onvolutional **N**eural **N**etwork*) composta por dois estágios principais: a etapa de codificação (também chamada de contração) e a etapa de decodificação (ou expansão). Essa arquitetura tem sido efetiva em várias aplicações, pois permite que informações de diferentes níveis sejam combinadas, e seu uso tem levado a resultados promissores com baixo número de dados anotados. Sua arquitetura lembra o formato da letra "U", o que inspirou o nome, como mostrado na Figura 6.

A etapa de codificação na arquitetura U-Net envolve operações de convolução sucessivas, gerando mapas de ativação que contêm atributos informativos das imagens. A resolução dos mapas de ativação é reduzida pela metade por meio de operações de *max pooling* ou convoluções com passo 2, com duplicação do número de canais em cada redução de resolução.

Já a etapa de decodificação da U-Net interpola a saída da etapa de codificação, adicionando convoluções para gerar atributos mais ricos, e é realizada por meio de convoluções transpostas. Além disso, durante esse processo, os mapas de ativação intermediários produzidos na etapa de codificação são concatenados aos mapas intermediários originados na decodificação, permitindo uma segmentação com preservação de detalhes.

A Figura 7 apresenta um exemplo de segmentação de uma imagem de microscopia de vasos sanguíneos realizada por uma rede U-Net. A imagem de entrada é representada pela Figura 7 (a), sendo que o resultado esperado da segmentação é mostrado na Figura 7 (b). O resultado esperado é chamado de padrão ouro, ou imagem referência. Os pixels

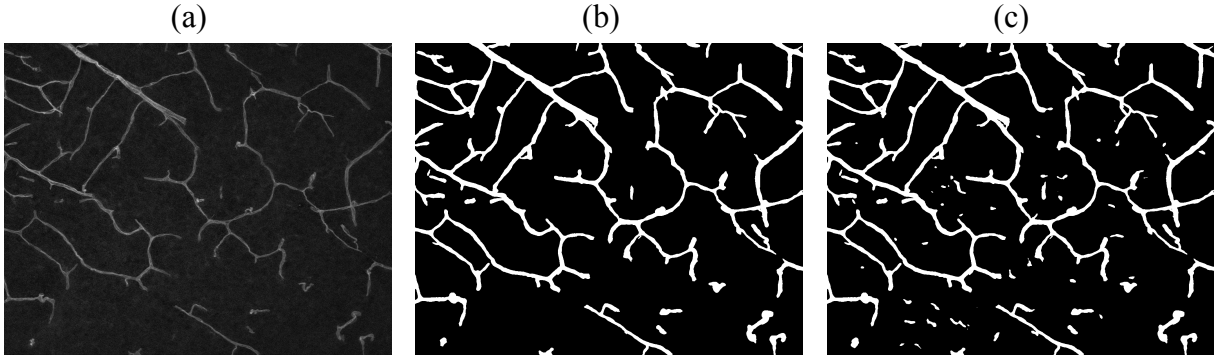
Figura 6 – Arquitetura U-Net



Fonte: (RONNEBERGER; FISCHER; BROX, 2015)

do padrão ouro possuem apenas dois possíveis valores, 0 (fundo) ou 1 (vaso sanguíneo). A Figura 7 (c) mostra o resultado da segmentação realizada pela U-Net.

Figura 7 – Exemplo de segmentação resultante da aplicação de uma rede U-Net a uma imagem de microscopia de vasos sanguíneos. (a) representa a imagem de entrada, (b) o padrão-ouro e (c) a segmentação resultante de uma U-Net.



Fonte: Próprio autor

Nota-se que a rede foi capaz de segmentar com qualidade praticamente toda a estrutura dos vasos sanguíneos. Contudo, há segmentações incorretas em áreas que correspondem ao plano de fundo da imagem original.

2.3 Transferência de Aprendizado

Historicamente, os modelos de aprendizado de máquina (AM) foram concebidos para aprender a realizar tarefas específicas em um dado domínio. Segundo (ZHUANG et al., 2020), muitos modelos de AM estão sob uma suposição comum, de que os dados de treinamento e teste são provenientes do mesmo espaço de características e distribuição de probabilidade. No entanto, coletar e rotular dados para treinar esses modelos é uma tarefa custosa e demorada (PAN; YANG, 2009) e, portanto, a TL (do inglês ***T****ransfer* ***L****earning*) surge como uma alternativa para melhorar o desempenho de modelos de AM em tarefas específicas, mesmo quando há poucos dados anotados disponíveis.

O aproveitamento das representações internas de uma tarefa particular pode ser útil para melhorar o desempenho de uma tarefa relacionada ou diferente. Então, a TL serve a este propósito, pois permite que um modelo treinado em uma tarefa específica seja adaptado para uma nova tarefa, melhorando o desempenho da nova tarefa com um conjunto de dados menor.

2.3.1 Notações e definições

A TL, embora seja uma técnica amplamente utilizada e com ideia intuitiva, não possui uma definição única. No entanto, é possível encontrar definições formais em trabalhos como (PAN; YANG, 2009) e (ZHUANG et al., 2020), que fornecem uma base sólida para a compreensão do conceito. A mesma notação e definições serão utilizadas neste trabalho para facilitar a compreensão do conceito.

Definição 1 (Domínio) O domínio $\mathcal{D}$ é composto de duas componentes: o espaço de instâncias, ou características, $\mathcal{X}$ e a distribuição de probabilidade $\mathcal{P}(x) = P(X = x)$, onde $X = \{x_1, x_2, ..., x_n\}$ é um vetor de características pertencente a $\mathcal{X}$. Assim, $\mathcal{D} = \{\mathcal{X}, \mathcal{P}(x)\}$.

Definição 2 (Conjunto de dados) O conjunto de dados, ou amostra, $D = D_{treino}, D_{teste}$ consiste de duas partes: o conjunto de treinamento, e conjunto de teste. O conjunto de treinamento $D_{treino} = \{(x_i, y_i)\}_{i=1}^{p}$ é composto por p pares de instâncias e rótulos, onde $x_i \in \mathcal{X}^{treino}$ e $y_i \in \mathcal{Y}^{treino}$. O conjunto de teste $D_{teste} = \{(x_i, y_i)\}_{i=1}^{m}$ é composto por m pares de instâncias e rótulos, onde $x_i \in \mathcal{X}^{teste}$, $y_i \in \mathcal{Y}^{teste}$.

Definição 3 (Tarefa) A tarefa $\mathcal{T}$ consiste do espaço de rótulos $\mathcal{Y}$ e da função objetivo f. Logo, $\mathcal{T} = \{\mathcal{Y}, f\}$. Os parâmetros θ da função f, por sua vez, podem ser determinados a partir do conjunto D_{treino}. Em outras palavras, a função f pode ser treinada para realizar a predição de um rótulo y, ou $f(x)$, a partir de instâncias x. Em uma notação de probabilidade, $f(x) = P(y|x)$, o que vale para modelos cuja a saída é uma distribuição da probabilidade condicional das instâncias.

Definição 4 (Transferência de Aprendizado) Dados os domínios fonte $\mathcal{D}_S$ e alvo $\mathcal{D}_T$, e as suas respectivas tarefas fonte $\mathcal{T}_S$ e alvo $\mathcal{T}_T$, a transferência de aprendizado tem como objetivo melhorar o aprendizado da tarefa alvo $\mathcal{T}_T$ no domínio alvo $\mathcal{D}_T$ por meio do conhecimento adquirido na tarefa fonte $\mathcal{T}_S$ e domínio fonte $\mathcal{D}_S$, onde $\mathcal{D}_S \neq \mathcal{D}_T$.

Trazendo esse entendimento para o contexto de segmentação de vasos sanguíneos, a TL pode ser utilizada para melhorar o desempenho de um modelo de segmentação refinado com um conjunto relativamente pequeno de imagens de vasos sanguíneos, mas aproveitando o conhecimento de representações internas de um modelo pré-treinado com um conjunto de um domínio diferente e, possivelmente, para uma tarefa de classificação.

Como exemplo, formalmente pode-se definir o domínio fonte $\mathcal{D}_S$ como um conjunto de imagens naturais diversas, tal como a base de imagens *ImageNet* (Jia Deng et al., 2009), que por sua vez foi utilizado para treinar uma CNN especializada na tarefa $\mathcal{T}_S$ de classificação de imagens. O domínio alvo $\mathcal{D}_T$ é um conjunto de imagens de vasos sanguíneos, e a tarefa alvo $\mathcal{T}_T$ é a segmentação dessas imagens. A CNN original pode ser adaptada para segmentar imagens de vasos sanguíneos, aproveitando as representações previamente aprendidas durante o treinamento utilizando $\mathcal{D}_S$.

Mesmo que as tarefas sejam distintas, as características inerentes às imagens naturais, tais como bordas, texturas e formas, podem ser úteis para a tarefa de segmentação. E isso foi experimentado em trabalhos como (JIANG et al., 2018), que adaptaram uma CNN baseada na arquitetura *AlexNet* (KRIZHEVSKY; SUTSKEVER; HINTON, 2017), pré-treinada com o conjunto *ImageNet*, mas refinada e testada com conjuntos de imagens de vasos sanguíneos, tais como DRIVE (STAAL et al., 2004), *STARE* (HOOVER, 2000) e CHASE_DB1 (FRAZ et al., 2012).

2.3.2 Refinamento de parâmetros

Do inglês ***F**ine-**T**uning* , o FT é uma técnica que consiste em reajustar os pesos, ou parâmetros, de uma rede neural pré-treinada em um novo conjunto de dados. A ideia é que a rede pré-treinada, que aprendeu as representações a partir de um conjunto de dados grande e se especializou na tarefa $\mathcal{T}_S$, possa ser adaptada para uma nova tarefa $\mathcal{T}_T$, com um conjunto de dados menor, por meio do ajuste dos pesos da rede. Isso é possível devido ao método de compartilhamento de parâmetros, que permite a transferência de conhecimento entre tarefas em nível paramétrico (ZHUANG et al., 2020).

Na prática, usa-se um modelo pré-treinado, como uma CNN, congela-se os parâmetros das camadas iniciais - em outras palavras, não se permite que esses parâmetros sejam reajustados - e, em seguida, as últimas camadas são reajustadas para a nova tarefa $\mathcal{T}_T$ ao serem treinadas com o novo conjunto de dados. Essas últimas camadas podem ser substituídas por outras, de modo a se adequarem à nova tarefa, como feito em (JIANG

et al., 2018), que realizou o ajuste fino da AlexNet, ao substituir as últimas camadas de classificação por camadas de segmentação.

2.4 Métricas para avaliação de segmentação de vasos sanguíneos

Nesta seção, serão apresentadas algumas métricas para avaliação de segmentação de imagens. O objetivo é fornecer uma visão geral dessas métricas e mostrar como essas podem ser utilizadas para avaliar e comparar métodos de segmentação.

Na área de segmentação de vasos sanguíneos, há as seguintes definições de avaliação que serão necessárias para a compreensão das medidas de qualidade do resultado:

1. Verdadeiro Positivo (VP): quando a segmentação resultante prediz corretamente o pixel ou conjunto de pixels que estão associados a vasos sanguíneos.

2. Verdadeiro Negativo (VN): quando a segmentação resultante prediz corretamente o pixel ou conjunto de pixels que estão associados ao fundo da imagem.

3. Falso Positivo (FP): quando um pixel ou conjunto de pixels associados ao fundo da imagem são classificados incorretamente como vasos sanguíneos.

4. Falso Negativo (FN): quando um pixel ou conjunto de pixels associados a vasos sanguíneos são classificados incorretamente como fundo da imagem.

2.4.1 Precisão

De todos os pixels que foram segmentados como vasos sanguíneos, dada uma imagem, a precisão informa qual a taxa de pixels corretamente segmentados em relação ao padrão ouro.

$$prec = \frac{VP}{VP + FP} \tag{2}$$

2.4.2 Sensibilidade

A sensibilidade, ou *recall*, avalia a capacidade do modelo em segmentar com sucesso os pixels rotulados no padrão ouro como vasos sanguíneos.

$$recall = \frac{VP}{VP + FN} \tag{3}$$

2.4.3 F1-Score

Por último, F1-Score é a média harmônica entre precisão e sensibilidade, evidenciada na Equação 4.

$$F1 = 2 * \frac{(prec * recall)}{(prec + recall)} \tag{4}$$

2.4.4 Medidas de similaridade

Coeficiente Dice

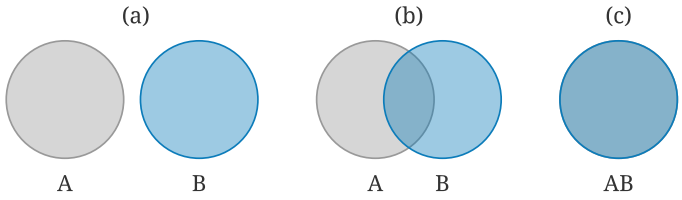
O Coeficiente de Similaridade Dice (DICE) é um índice de sobreposição espacial entre duas segmentações (ZOU et al., 2004), também definido como uma medida de concordância entre as regiões de cada classe (GIGANDET; CUADRA; THIRAN, 2004). Esse índice permite quantificar a qualidade da segmentação proporcionada por um algoritmo de segmentação.

Seja A o conjunto de pixels da imagem referência e B o conjunto de pixels da imagem segmentada pelo algoritmo, então o coeficiente DICE é calculado como

$$DSC = \frac{2(A \cap B)}{A + B}. \tag{5}$$

$(A \cap B)$ representa a intersecção entre os pixels positivos das duas imagens. $A + B$ representa o número total de pixels positivos em ambas as imagens. A Figura 8 ilustra o domínio de valores do coeficiente DICE, sendo que na Figura 8 (a) não há sobreposição dos pixels, então DICE=0. Quando ocorre sobreposição parcial na Figura 8(b), 0<DICE<1. Finalmente, no cenário de completa sobreposição na Figura 8(c), DICE=1.

Figura 8 – Ilustração da sobreposição de dois conjuntos de segmentação e respectivos valores do coeficiente DICE.



Fonte: Próprio autor

Reescrevendo a Equação 5 em termos das definições de VP, FN e FP:

$$DSC = \frac{2VP}{2VP + FN + FP} \tag{6}$$

Nota-se que o coeficiente DICE é matematicamente equivalente ao F1-Score. Expandindo a Equação 4:

$$F1 = \frac{2 \cdot prec \cdot recall}{prec + recall} = \frac{2 \cdot \frac{VP}{VP+FP} \cdot \frac{VP}{VP+FN}}{\frac{VP}{VP+FP} + \frac{VP}{VP+FN}} = \frac{2VP}{2VP + FN + FP} = DSC \tag{7}$$

Essa equivalência demonstra que ambas as métricas representam a média harmônica entre precisão e sensibilidade, sendo o DICE mais utilizado no contexto de segmentação de imagens enquanto o F1-Score é amplamente empregado em tarefas de classificação.

Coeficiente IoU

O coeficiente de interseção sobre união (IoU), ou índice de Jaccard, é também uma métrica de similaridade. Sua equação é descrita assim:

$$IoU = \frac{(A \cap B)}{(A \cup B)} = \frac{VP}{VP + FN + FP} \tag{8}$$

É importante notar que as métricas IoU e DICE estão diretamente relacionadas. Ou seja, quando uma aumenta, a outra também aumenta. É possível converter algebricamente uma na outra:

$$DSC = \frac{2 \cdot IoU}{1 + IoU} \quad \text{e} \quad IoU = \frac{DSC}{2 - DSC} \tag{9}$$

Capítulo 3

Revisão bibliográfica

3.1 Aprendizado com poucos exemplos

Um dos principais desafios em aprendizado de máquina é a necessidade de grandes quantidades de dados anotados para treinar modelos com desempenho satisfatório. No entanto, em muitas aplicações, como em imagens médicas, a coleta e anotação de dados é uma tarefa custosa e demorada. Adicionalmente, mesmo que haja disponibilidade de um grande volume de dados anotados, treinar modelos de AP requer um poder computacional significativo, principalmente pela necessidade do uso de Unidade de Processamento Gráfico (GPU)(do inglês, ***G****raphics* ***P****rocessing* ***U****nit*) para acelerar o treinamento, e isso pode ser um impeditivo para muitos pesquisadores e profissionais.

Nesse cenário, modelos capazes de generalizar o conhecimento prévio com poucos exemplos anotados em novas tarefas supervisionadas são desejáveis, e uma das abordagens para lidar com esse problema é a técnica de Aprendizado com Poucos Exemplos (FSL) (do inglês, ***F****ew-****S****hot* ***L****earning*) (WANG et al., 2020). Segundo os mesmos autores, o FSL é motivado por três justificativas principais:

- **Aproximação da inteligência humana**: o FSL busca emular a capacidade humana de aprender rapidamente a partir de poucos exemplos;
- **Viabilização do aprendizado em casos raros**: a técnica permite a aprendizagem de tarefas com poucos exemplos anotados, nas quais a coleta de dados é custosa ou inviável;
- **Redução do esforço de coleta de dados e custo computacional**: o FSL oferece uma alternativa ao treinamento tradicional, que requer grandes volumes de

dados anotados e poder computacional.

O estudo de (WANG et al., 2020) oferece um panorama sobre FSL, e o define como um problema de AM, em que a experiência (E) disponível, ou conjunto de dados de treinamento, é limitada a um pequeno número de exemplos supervisionados para efetuar uma tarefa (T), que por sua vez é avaliada por uma métrica de desempenho (P). Adicionalmente, (SONG et al., 2022) introduz a notação *N-way-K-shot* para descrever o problema de FSL, onde N é o número de classes e K é o número de exemplos por classe. Ambos os trabalhos destacam os casos especiais de FSL, sendo o Aprendizado com um Único Exemplo (OSL) (do inglês ***O**ne-**S**hot **L**earning*), denotado por *N-way-1-shot*, e o Aprendizado com Zero Exemplo (ZSL) (do inglês ***Z**ero-**S**hot **L**earning*), denotado por *0-way-K-shot*. OSL é um caso especial de FSL em que apenas um exemplo por classe é fornecido para treinamento. Já ZSL, proposto por (LAMPERT; NICKISCH; HARMELING, 2009), é o caso em que o modelo deve ser capaz de generalizar para classes não vistas durante o treinamento.

O trabalho de (WANG et al., 2020) propõe que a questão central do aprendizado supervisionado em FSL é a não confiabilidade do minimizador de risco empírico. Em outras palavras, com poucas amostras de treinamento, o modelo tende a se ajustar excessivamente a esses dados, causando o *overfitting* e resultando em desempenho ruim em dados não vistos. O artigo explica o fenômeno em termos de erro de estimação, que mede o quão bem o risco empírico, calculado nos dados de treinamento, se aproxima do risco real, o erro que o modelo teria em todos os dados possíveis. Com poucas amostras de treinamento, o erro de estimação pode ser alto, tornando o minimizador de risco empírico não confiável, uma vez que com a escassez de dados anotados, o FSL pode ter dificuldade na generalização.

A taxonomia de FSL discutida no mesmo trabalho categoriza três abordagens para lidar com a questão principal do FSL, com base em como o conhecimento prévio é utilizado para melhorar a generalização do modelo. São elas:

- **Dados**: o conhecimento prévio é aproveitado para aumentar ou enriquecer o conjunto de dados de treinamento, expandindo o volume amostral disponível. Assim, o aprendizado pode tornar-se mais robusto e confiável. Técnicas como *data augmentation* e transformações sobre o conjunto de dados de treinamento, e sobre conjuntos não rotulados podem ser utilizadas para aumentar a diversidade dos dados de treinamento.

- **Modelo**: nesta perspectiva, o conhecimento prévio é utilizado para restringir a complexidade do espaço de hipótese do modelo, ou seja, o espaço das possíveis funções que o modelo pode aprender. Isso tem como objetivo simplificar o problema de aprendizado. Técnicas como aprendizado multitarefa, aprendizado de represen-

tações e compartilhamento de parâmetros são exemplos de como o conhecimento prévio pode ser usado para restringir o espaço de hipótese.

- **Algoritmo**: o conhecimento prévio é incorporado ao próprio algoritmo de aprendizado, modificando a forma como o modelo busca a melhor hipótese dentro do espaço de hipóteses, otimizando o processo de busca pela melhor hipótese, ou seja, a função que melhor se ajusta aos dados, mesmo com poucos exemplos, tornando o aprendizado mais eficiente, evitando assim o *overfitting*. Técnicas como meta-aprendizado, aprendizado de otimizadores e refinamento de parâmetros existentes contribuem para a restrição do espaço de busca durante o treinamento.

Por sua vez, (SONG et al., 2022) acrescenta novas camadas de desafios que o FSL enfrenta, além do problema da não confiabilidade do minimizador de risco empírico. Os autores introduzem novos desafios, como:

- **Avaliação Imprecisa da Distribuição de Dados**: refere-se à dificuldade em estimar com precisão a distribuição de dados subjacente em cenários de FSL, onde o número de exemplos é limitado. Isso pode levar a modelos que não generalizam bem para novos dados, especialmente quando a distribuição dos dados de teste difere da distribuição dos dados de treinamento. A avaliação imprecisa da distribuição de dados pode levar a um desempenho insatisfatório do modelo em situações do mundo real, onde a distribuição dos dados pode variar. Técnicas como aumento de dados, aprendizado de domínio adversarial e meta-aprendizado podem ajudar a lidar com esse desafio, tornando os modelos mais robustos a variações na distribuição de dados.

- **Sensibilidade ao Reuso de Características**: a transferência de conhecimento prévio de grandes conjuntos de dados para tarefas de FSL pode ser desafiadora devido à sensibilidade ao reuso de características. Características aprendidas em um domínio ou tarefa podem não ser diretamente aplicáveis a outros, exigindo mecanismos de adaptação para garantir uma transferência eficaz. Abordagens como adaptação de domínio, meta-aprendizado e aprendizado de representações robustas podem ajudar a superar esse desafio, permitindo que o modelo se adapte melhor às características da nova tarefa.

- **Generalidade de tarefas futuras**: a capacidade de generalizar não apenas para novas classes dentro de uma mesma tarefa, mas também para tarefas completamente novas e não vistas durante o treinamento, é crucial para a aplicabilidade do FSL em cenários reais. A falta dessa generalidade limita o potencial do FSL em lidar com a diversidade de tarefas que podem surgir. Abordagens como meta-aprendizado, aprendizado multitarefa e aprendizado de representações abstratas e compostas podem contribuir para aumentar a generalidade dos modelos de FSL.

- **Limitação da representação unimodal**: a dependência exclusiva de uma única modalidade de informação (como imagens ou texto) pode restringir a capacidade do modelo de aprender representações ricas e completas, especialmente em tarefas complexas. A combinação de múltiplas modalidades pode fornecer informações complementares e aprimorar o desempenho do modelo. Métodos como aprendizado multimodal, fusão de informações e atenção multimodal podem ser explorados para superar essa limitação, permitindo que o modelo se beneficie de diferentes fontes de informação.

A respeito da relação do FSL com TL, (WANG et al., 2020) posicionam a TL como um problema de aprendizado de máquina relacionado, mas distinto do FSL. Contudo, os autores mencionam que TL é comumente utilizado em problemas de FSL para melhorar o desempenho do modelo, ao aproveitar o conhecimento de modelos pré-treinados. Essa argumentação fica mais evidente quando o trabalho aprofunda a discussão sobre técnicas de FT dentro do escopo de algoritmo, segundo a sua taxonomia proposta, tais como FT utilizando regularização, ou FT de parâmetros existentes com novos parâmetros.

Por outro lado, (SONG et al., 2022) entendem que o TL é um ferramental fundamental para superar um dos quatro principais desafios do FSL, a "sensibilidade ao reuso de características", mencionada acima. A transferência de conhecimento de modelos pré-treinados, como através do FT da *backbone* da rede, permite a reutilização de características aprendidas em tarefas com abundância de dados para novas tarefas com poucos exemplos.

3.1.1 Segmentação de imagens médicas com transferência de aprendizado

A transferência de conhecimento de modelos pré-treinados tem sido amplamente explorada em tarefas de visão computacional, permitindo a reutilização de características aprendidas em grandes conjuntos de dados, como o *ImageNet*, para tarefas específicas, tais como segmentação de imagens médicas. Ainda assim, a aplicação de TL em imagens médicas apresenta desafios específicos, devido à lacuna entre as características aprendidas em imagens naturais e as características relevantes para o domínio biomédico (SONG et al., 2022), trazendo a necessidade de adaptações sobre a arquitetura da rede, ou transformações e enriquecimento do conjunto de dados de treinamento para reduzir a discrepância entre os domínios.

Os autores (JIANG et al., 2018) propuseram uma abordagem para explorar a técnica de TL no domínio de imagens biomédicas, com foco na segmentação de imagens de vasos sanguíneos da retina. O estudo foi inspirado pelo trabalho de (SHELHAMER; LONG; DARRELL, 2017) que utiliza uma FCNN para segmentação de imagens naturais. A metodologia proposta consiste em adaptar um *backbone* da *AlexNet* (KRIZHEVSKY;

SUTSKEVER; HINTON, 2017) pré-treinada no conjunto de imagens *ImageNet*, mas com a substituição das últimas camadas totalmente conectadas por camadas convolucionais, permitindo a geração de mapas de características de saída com a mesma resolução da imagem de entrada. A partir disso, a rede passa para um processo de FT, na qual as últimas camadas são treinadas com um conjunto de dados de imagens de vasos sanguíneos da retina para a adaptação do modelo ao novo domínio, mas aproveitando os pesos e representações previamente abstraídas pelo backbone da *AlexNet* a partir da *ImageNet*.

Além disso, devido à escassez de dados anotados em imagens médicas, o trabalho propõe a combinação de diferentes fontes de dados de imagens de retina, como DRIVE, STARE e CHASE_DB1, e também a técnica de *data augmentation* para aumentar a diversidade do conjunto de dados de treinamento. Essas abordagens, semelhantes às que foram tratadas no assunto de FSL, permitiram a melhoria do desempenho do modelo, reduzindo a lacuna entre as características aprendidas em imagens naturais e as características relevantes para a segmentação de vasos sanguíneos da retina, e consequentemente, elevando a acurácia da segmentação, comparado a outros trabalhos contemporâneos.

Corroborando com essa perspectiva, (GHAFOORIAN et al., 2017) exploraram a adaptação de domínio sobre o desafio de segmentação da Hiperintensidade de Matéria Branca (WMH) (do inglês, ***W****hite* ***M****atter* ***H****yperintensity*) em Imagem por Ressonância Magnética (MRI) (do inglês, ***M****agnetic* ***R****esonance* ***I****maging*). Os autores propuseram uma abordagem de TL, adaptando uma CNN pré-treinada em um conjunto de dados base de origem, para um conjunto de dados destino. O modelo foi submetido a um processo de FT, permitindo a adaptação do mesmo às particularidades do novo cenário.

Os autores experimentaram diferentes configurações conjugadas de refinamento e modelos. Sendo que um exemplar do modelo foi treinado do zero com o conjunto alvo, enquanto os outros foram refinados com diferentes combinações de camadas congeladas e número de amostras de treinamento. Os resultados demostraram que, com poucos dados no domínio alvo, ajustar apenas as últimas camadas da rede mostrou-se mais eficaz, evitando overfitting. Contudo, ao passo que o número de amostras de treinamento aumentava, o refinamento de mais camadas da rede resultou em melhor desempenho.

Capítulo 4

Material e métodos

4.1 Ferramentas e tecnologias utilizadas

As ferramentas e bibliotecas utilizadas para desenvolvimento deste trabalho, são estas:

- Linguagem de programação: *Python*
- Framework para desenvolvimento dos modelos: *PyTorch*
- Observabilidade dos *logs* dos experimentos: *Weights & Biases* [1]
- Biblioteca para composição de modelo codificador-decodificador: *Segmentation Models Pytorch* [2]
- Biblioteca com utilitários para treinamento dos modelos *PyTorch*: *TorchTrainer*[3]
- API para processamento dos tensores: CUDA

A plataforma de computação utilizada consiste da seguinte configuração:

- Processador: Intel i9-12900KF
- GPU: GeForce RTX 3090 24GB
- Memória RAM: 64 GB

[1] Weights & Biases <https://wandb.ai/>

[2] Pacote baseado em Pytorch para composição de modelos de segmentação de imagens com codificador-decodificador e disponibilizado no repositório: <https://github.com/qubvel-org/segmentation_models.pytorch>

[3] Pacote de utilitários e funções para treinamento de modelos, desenvolvido pelo professor e orientador Dr. Cesar Henrique Comin e disponibilizado no repositório <https://github.com/chcomin/torchtrainer>

4.2 Conjuntos de dados

Neste projeto, foram utilizados conjuntos de dados de diferentes domínios, organizados em dois grupos principais: o de origem ($\mathcal{D}_S$), destinado ao pré-treinamento dos modelos, e o de destino ($\mathcal{D}_T$), utilizado para o refinamento e avaliação. O domínio de destino é composto pelas coleções de imagens médicas DRIVE e VessMap, que servem para especializar os modelos na tarefa final.

Para o domínio de origem, o foco é o conjunto de dados sintético *VessShape*, que foi desenvolvido especificamente para o treinamento dos modelos base. O *ImageNet*, por sua vez, foi empregado com o único propósito de fornecer uma vasta biblioteca de texturas. Essas texturas são aplicadas proceduralmente durante a geração das imagens do *VessShape*, garantindo assim a diversidade visual do conjunto de treinamento, sem que o *ImageNet* seja diretamente utilizado para treinar os modelos.

A Tabela 1 apresenta um resumo dos conjuntos de dados, destacando suas características e o papel que desempenham nesta pesquisa.

Tabela 1 – Conjuntos de dados utilizados neste trabalho

Conjunto de dados	**Domínio**	**Tarefa**	**Origem**	**Propósito**
ImageNet	$\mathcal{D}_S$	Classificação	Natural	Texturização
VessShape	$\mathcal{D}_S$	Segmentação	Sintético	Treinamento
VessMap	$\mathcal{D}_T$	Segmentação	Natural	Refinamento
DRIVE	$\mathcal{D}_T$	Segmentação	Natural	Refinamento

Fonte: Próprio autor

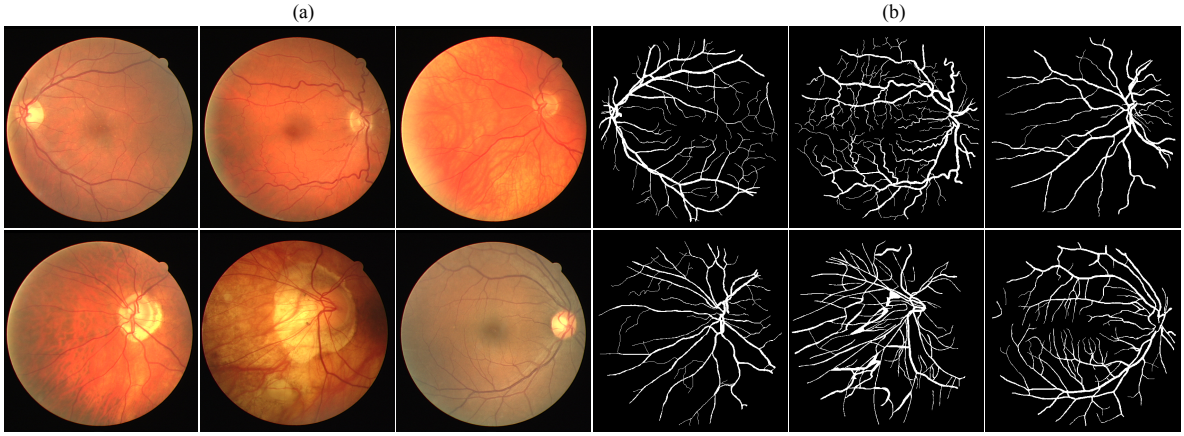
4.2.1 DRIVE

O conjunto de dados DRIVE (STAAL et al., 2004) é uma referência clássica na avaliação de algoritmos para segmentação de vasos sanguíneos em imagens de retina. A coleção consiste em 40 retinografias coloridas, com resolução de 584×565 pixels, já pré-divididas em conjuntos de 20 imagens para treinamento e 20 para teste. Algumas características visuais são importantes nesse conjunto. Primeiro, a presença de muitos vasos de calibre fino, que representam um desafio significativo para os métodos de segmentação. Adicionalmente, os vasos aparecem como estruturas escuras sobre um fundo mais claro, como demonstrado na Figura 9. Para os propósitos deste trabalho, todas as imagens foram convertidas para tons de cinza antes de serem utilizadas nas etapas de treinamento, validação e teste.

4.2.2 Vasos sanguíneos do córtex de camundongos

A coleção de imagens de vasos sanguíneos do córtex cerebral de camundongos corresponde a um conjunto de microscopia confocal, provenientes de um trabalho colaborativo

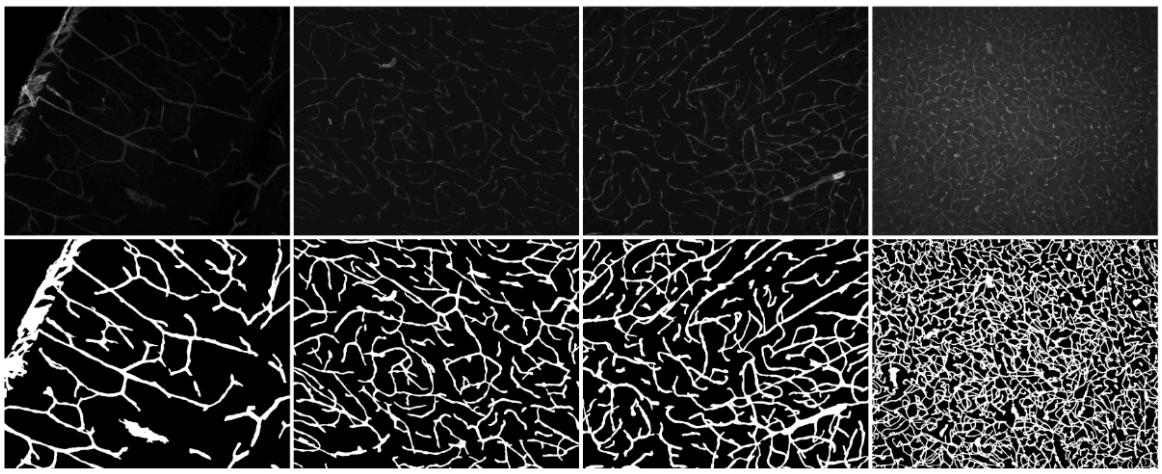
Figura 9 – Amostra de imagens do conjunto de treino do DRIVE. Em (a), são exibidas as retinografias, que apresentam notável variabilidade em termos de iluminação, contraste dos vasos e a presença de estruturas patológicas, como as lesões amareladas na imagem central inferior. Em (b), os respectivos mapas de segmentação (padrão-ouro), que ilustram a complexa árvore vascular a ser segmentada.



Fonte: Próprio autor

com o Professor Baptiste Lacoste, da Universidade de Ottawa, Canadá. Esse conjunto é caracterizado por imagens com tamanho variado, com resoluções de 1376×1104 e 2499×2005 pixels, e com diferentes níveis de ruído, contraste e calibre. As amostras foram coletadas em distintos cenários experimentais (LACOSTE et al., 2014; OUELLETTE et al., 2020; MCDONALD et al., 2021). Cada imagem tem o seu respectivo mapa de segmentação, ou padrão-ouro, que indica a localização dos vasos sanguíneos. A Figura 10 apresenta exemplos de imagens desse conjunto de dados.

Figura 10 – Amostra de imagens de vasos sanguíneos do córtex de camundongos e seus respectivos mapas de segmentação.



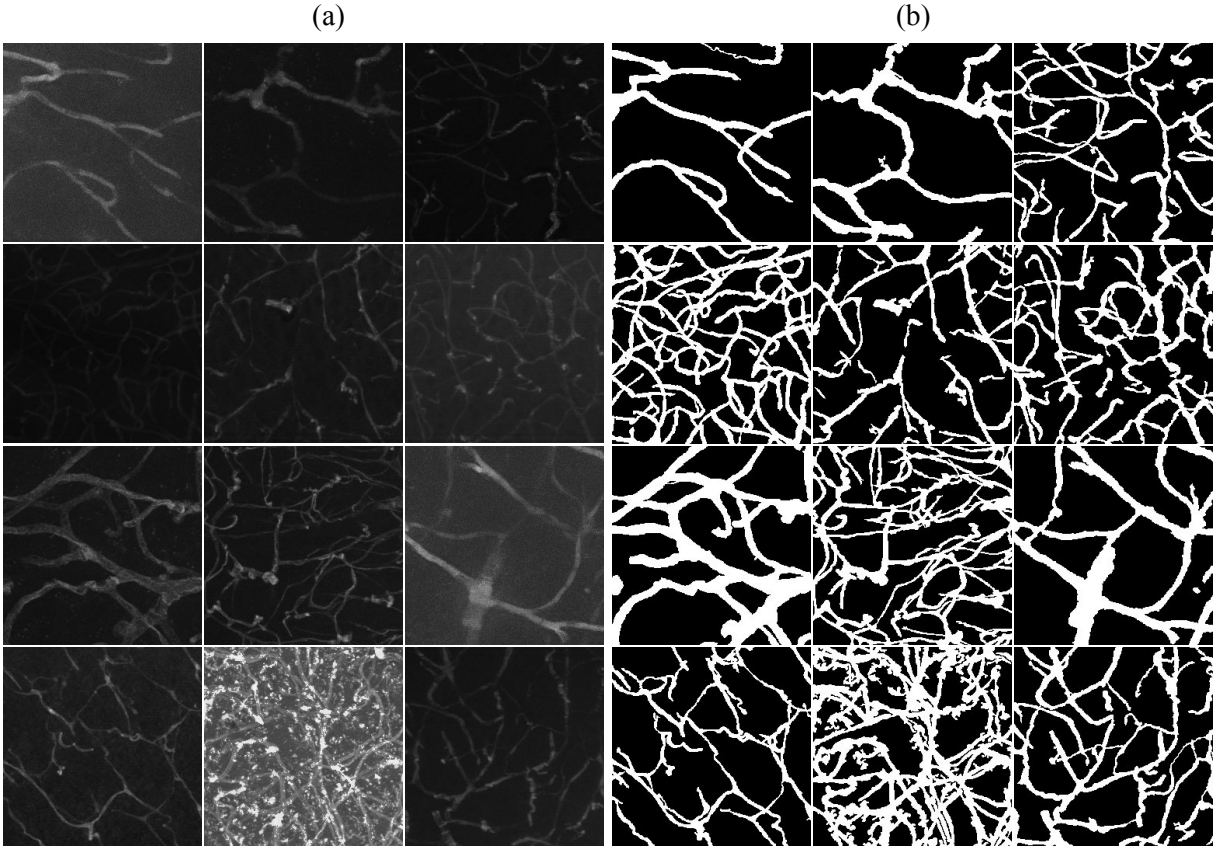
Fonte: Próprio autor

4.2.2.1 VessMap

Derivado do trabalho anterior, o VessMap (SILVA et al., 2025) é um subconjunto de 100 imagens, cada uma segmentada manualmente com precisão em nível de pixel. As imagens do VessMAP foram selecionadas com base em um processo que mapeou o conjunto de imagens original em um espaço n-dimensional, e selecionou 100 pontos do espaço de forma uniforme. O processo garante a heterogeneidade das amostras, isto é, a diversidade de características presentes nas imagens, tais como o contraste, nível de ruído e densidade de vasos sanguíneos.

Diferentemente de métodos tradicionais de seleção de imagens, que valorizam o volume de dados, o VessMap possui amostras com características típicas e atípicas, o que pode ser útil para treinar modelos de segmentação de vasos sanguíneos mais robustos em termos de generalização. Algumas amostras desse conjunto são mostradas na Figura 11.

Figura 11 – Amostra de imagens do conjunto de dados VessMap (a) e seus respectivos mapas de segmentação (b). É possível observar a diversidade de características presentes nas imagens, como contraste, nível de ruído e densidade de vasos sanguíneos.



Fonte: Próprio autor

4.2.3 ImageNet

O *ImageNet* (Jia Deng et al., 2009) é uma base de dados de imagens naturais largamente utilizada como referência (*benchmark*) para treinamento de modelos de aprendizado profundo. O conjunto de dados é composto por mais de 1 milhão de imagens, divididas em 1000 categorias e hierarquicamente organizadas de acordo com a taxonomia da WordNet (CHRISTIANE, 1998).

Além de ser um dos maiores e mais populares conjuntos de dados de imagens disponíveis para tarefas de classificação e detecção de objetos, é também frequentemente utilizado como base de dados de origem para treinamento de *backbones* de redes neurais profundas, como *ResNet* e *VGG*, que são posteriormente refinados para tarefas específicas, como segmentação semântica. A escolha do *ImageNet* como conjunto de dados de origem se deve à sua diversidade de imagens e categorias, exemplificado na Figura 12, o que permite que os modelos aprendam representações de alto nível úteis para TL.

Figura 12 – Amostra de imagens do conjunto de validação da *ImageNet* agrupadas por similaridade após aplicação do algoritmo T-SNE.

Fonte: (KARPATHY, 2014)

4.2.4 VessShape - Geração sintética de imagens com forma de vasos sanguíneos

Este trabalho introduz o *VessShape*, um conjunto de dados gerado de forma sintética para o pré-treinamento de modelos de segmentação de vasos sanguíneos. A metodologia consiste em combinar formas geométricas similares às de vasos sanguíneos com uma vasta gama de texturas aplicadas ao primeiro plano e ao fundo. O propósito central é compor imagens que forcem o modelo a aprender as características geométricas dos vasos, variando sistematicamente a textura para desencorajar o aprendizado de pistas de aparência e textura. Em outras palavras, tornar o modelo invariante às texturas. A principal vantagem é a capacidade de gerar uma coleção massiva de imagens, induzindo um forte viés de forma sem a necessidade de coletar e anotar dados reais.

A escolha dessa abordagem é inspirada no trabalho de (GEIRHOS et al., 2019), que demonstrou que modelos treinados com o conjunto de dados SIN se tornam mais robustos ao aprenderem a reconhecer objetos pela forma, e não pela textura. Contudo, em vez de estilizar imagens existentes, como no caso do SIN, nossa abordagem foca na composição de imagens a partir de geometrias geradas proceduralmente, que são então preenchidas com texturas aleatórias do conjunto *ImageNet* com diferentes categorias para o primeiro plano e o fundo, a fim de preservar a forma geométrica dos objetos.

O processo de geração de cada imagem do *VessShape* ocorre em duas etapas principais: a criação da geometria vascular por meio de uma máscara binária e a composição da imagem final com texturas.

Uma biblioteca em Python foi desenvolvida e disponibilizada em um repositório público para que o método seja reprodutível[4].

4.2.4.1 Geração da Geometria Vascular

A estrutura geométrica dos vasos é definida proceduralmente utilizando curvas de Bézier, que oferecem uma representação flexível e controlada de formas tubulares. Cada segmento vascular é descrito por uma curva de Bézier de ordem n, definida pela Equação 10, onde $\{\mathbf{p}_i\}_{i=0}^{n}$ são os pontos de controle e $t \in [0, 1]$.

$$\mathbf{c}(t) = \sum_{i=0}^{n} \binom{n}{i} (1-t)^{n-i} t^i \mathbf{p}_i \tag{10}$$

Para gerar uma curva, os pontos de controle inicial ($\mathbf{p}_0$) e final ($\mathbf{p}_n$) são amostrados aleatoriamente no domínio da imagem. Os pontos intermediários são posicionados sobre a reta que conecta $\mathbf{p}_0$ e $\mathbf{p}_n$ e, em seguida, deslocados por uma distância aleatória δ ao longo de um vetor normal a essa reta. Esse processo introduz tortuosidade aos segmentos, gerando geometrias mais realistas.

[4] Respositório do VessShape: <https://github.com/galvaowesley/vess-shape-dataset>

Uma vez definida a curva, ela é discretizada para formar uma polilinha de 1 pixel de espessura na grade da imagem. Para conferir espessura tubular, aplica-se uma operação de dilatação morfológica com um elemento estruturante em disco de raio r_0. O resultado é uma máscara binária M que representa a geometria vascular. Para garantir a diversidade estrutural do conjunto de dados, os parâmetros-chave da geração são amostrados aleatoriamente de intervalos pré-definidos, conforme resume a Tabela 2.

Tabela 2 – Principais parâmetros para a geração de geometrias no VessShape.

Parâmetro	**Intervalo**	**Descrição**
Número de curvas (K)	$[1, 20]$	Quantidade de segmentos vasculares por imagem.
Pontos de controle ($n + 1$)	$[2, 20]$	Complexidade da curva de Bézier.
Escala de deslocamento (δ)	$[50, 150]$ (px)	Controla a curvatura/tortuosidade dos segmentos.
Raio inicial (r_0)	$[1, 5]$ (px)	Raio basal do vaso antes da operação de suavização.

Fonte: Próprio autor

4.2.4.2 Composição com texturas

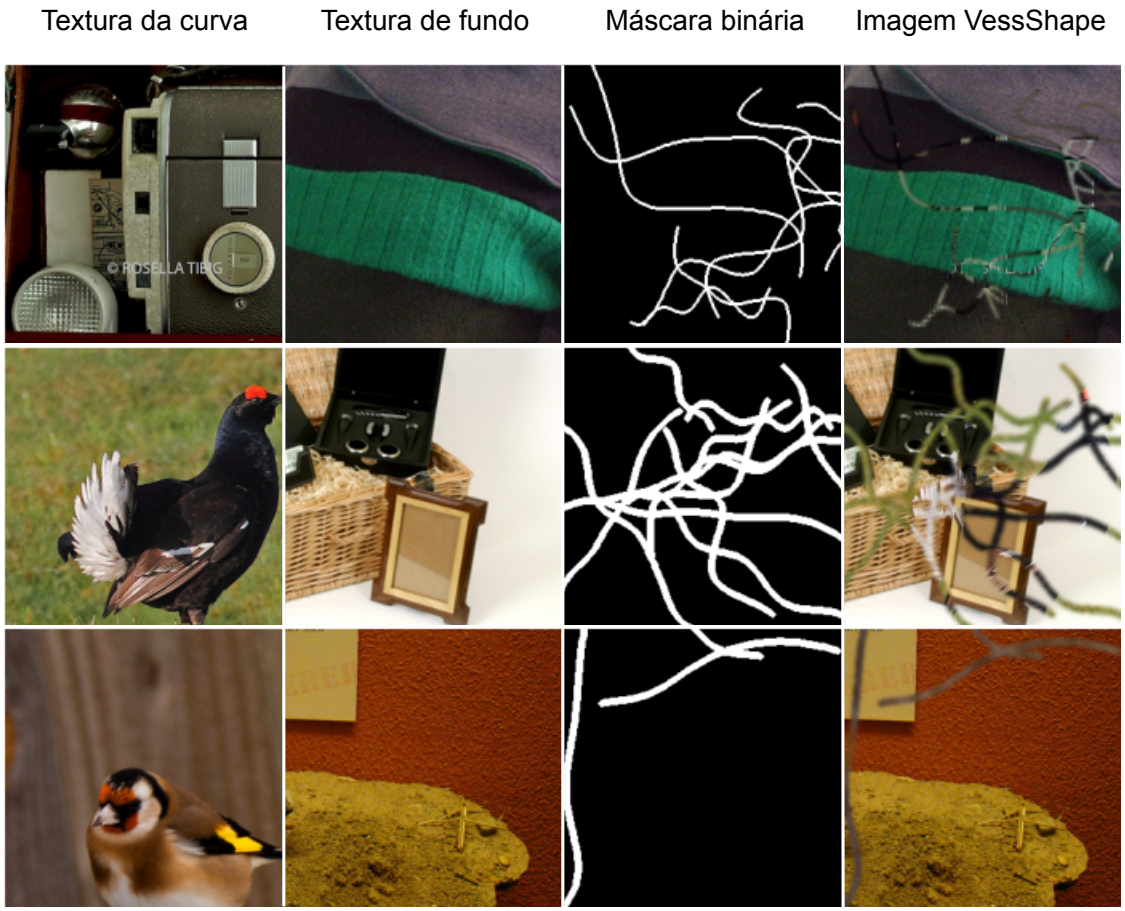
A segunda etapa consiste em aplicar texturas à máscara binária M para criar a imagem final I. Para cada máscara, duas imagens são selecionadas aleatoriamente de classes distintas do *ImageNet*, uma para servir como textura de primeiro plano (F) e outra para o fundo (B).

Para garantir uma transição suave entre o vaso e o fundo, a máscara binária M é suavizada com um filtro Gaussiano de desvio-padrão σ, gerando uma máscara alfa (A) com valores normalizados em $[0, 1]$. A imagem final é, então, composta pela mescla ponderada das texturas, como descrito na Equação 11.

$$I = A \cdot F + (1 - A) \cdot B \tag{11}$$

Essa operação de mescla assegura que as regiões de vasos ($A \approx 1$) preservem a textura de primeiro plano, enquanto o restante da imagem ($A \approx 0$) mantenha a textura de fundo, como ilustrado na Figura 13. O parâmetro σ, que controla a suavidade das bordas, também é amostrado aleatoriamente no intervalo $[1, 2]$. O resultado é um conjunto de dados sintético com formas geométricas consistentes, mas com uma variedade de aparências que força o modelo a focar na estrutura em vez da textura.

Figura 13 – Exemplos do processo de geração do *VessShape*. As texturas são amostradas do *ImageNet* e mescladas conforme máscaras binárias geradas proceduralmente para criar as imagens finais.



Fonte: Próprio autor

4.3 Transferência de aprendizado de forma para segmentação de vasos sanguíneos

Para alcançar os objetivos propostos, a abordagem de TL é empregada para transferir o conhecimento adquirido no domínio de origem sintético, $\mathcal{D}_S$, para os domínios de destino com imagens naturais, $\mathcal{D}_T$. A estratégia central consiste em pré-treinar modelos base no conjunto *VessShape*, que foi projetado para a tarefa de segmentação ($\mathcal{T}_S$) com um forte viés de forma. A hipótese é que, ao serem treinados com o *VessShape*, os modelos aprendem representações geométricas robustas que, subsequentemente, são adaptadas para a tarefa de destino ($\mathcal{T}_T$) — a segmentação de vasos sanguíneos — nos conjuntos DRIVE e *VessMap* ($\mathcal{D}_T$).

Para validar a efetividade da transferência de conhecimento, a metodologia de avaliação foi estruturada em dois cenários de treinamento distintos. O primeiro, que serve

como base de comparação, consiste no treinamento dos modelos inteiramente do zero, utilizando apenas os conjuntos de dados de destino, DRIVE e *VessMap*.

O segundo cenário investiga a transferência de aprendizado propriamente dita. Nele, os modelos são pré-treinados com o conjunto sintético *VessShape* e, em seguida, avaliados nos dados de destino de duas maneiras: por meio de FT em um regime de FSL e diretamente em um cenário de ZSL, sem qualquer refinamento. Essa abordagem comparativa permite quantificar o ganho de desempenho e a eficiência de dados proporcionados pelo pré-treinamento com viés de forma.

Os códigos-fonte produzidos para essa tarefa estão disponíveis em um repositório público[5].

4.3.1 Arquitetura dos modelos

A arquitetura escolhida para todos os modelos de segmentação foi a U-Net (RONNEBERGER; FISCHER; BROX, 2015), uma rede do tipo codificador-decodificador com conexões de atalho entre os estágios correspondentes. Essa arquitetura é amplamente reconhecida por sua eficácia em tarefas de segmentação de imagens biomédicas, especialmente em cenários com dados limitados.

Como codificadores, foram utilizadas duas variantes da arquitetura ResNet (HE et al., 2016): a ResNet18 e a ResNet50. A ResNet50 é uma rede consideravelmente mais profunda e com maior número de parâmetros que a ResNet18, o que lhe confere uma maior capacidade de aprendizado. A escolha de duas arquiteturas com capacidades distintas é proposital e serve a um objetivo central da avaliação: analisar como a complexidade do modelo interage com a eficácia da transferência de conhecimento do domínio sintético.

Essa comparação permite analisar o balanço entre a capacidade do modelo e sua adequação a cenários com dados limitados. Modelos maiores, como a ResNet50, têm o potencial de aprender representações mais complexas. No entanto, quando o volume de dados de treinamento é pequeno, eles se tornam mais suscetíveis ao sobreajuste, fenômeno no qual o modelo memoriza os exemplos em vez de generalizar o aprendizado. Em contrapartida, modelos mais simples, como a ResNet18, tendem a ser mais robustos à escassez de dados. O custo computacional também é um fator relevante nesta escolha, uma vez que modelos menores exigem menos tempo e recursos para o treinamento.

4.3.2 Pré-treinamento no VessShape

A primeira etapa consiste no pré-treinamento dos modelos com o conjunto de dados *VessShape*. O objetivo desta fase é expor os modelos a uma vasta diversidade de geometrias tubulares, tratando a textura como uma característica secundária. Para isso, cada

[5] Código (pré-treinamento *VessShape* + *fine-tuning* com poucos exemplos): <https://github.com/galvaowesley/vess-shape-experiments>

Tabela 3 – Hiperparâmetros de pré-treinamento usados no VessShape. O termo *Weight decay* corresponde ao Decaimento de Peso, uma técnica de regularização.

Hiperparâmetro	**VSUNet50**	**VSUNet18**
Tamanho do batch	96	192
Taxa de aprendizado	10^{-3}	10^{-2}
Weight decay	10^{-4}	0.0

Fonte: Próprio autor

amostra de treinamento é gerada sob demanda, um processo que promove um embaralhamento contínuo da aparência visual. Essa estratégia, ao combinar regras geométricas estáveis com texturas sempre novas, é projetada para injetar um forte viés de forma nos modelos e, ao mesmo tempo, desencorajar a memorização de padrões de textura.

Durante esta etapa, os modelos são otimizados para minimizar a perda de entropia cruzada sobre o conjunto sintético, que é efetivamente infinito. Foram pré-treinados dois modelos base, denominados VSUNet18 e VSUNet50, que utilizam a arquitetura U-Net com codificadores ResNet18 e ResNet50, respectivamente. O treinamento do VSUNet18 abrangeu aproximadamente 7,1 milhões de imagens sintéticas ao longo de 8,6 horas. Já o VSUNet50, de maior capacidade, foi treinado com cerca de 53,0 milhões de imagens por 78,3 horas.

Como pré-processamento, todas as imagens de entrada foram normalizadas por canal, utilizando as estatísticas do *ImageNet*. Para o monitoramento do desempenho durante o treinamento, foram utilizados conjuntos de validação e teste pré-gerados a partir do *VessShape*, contendo 9.000 e 200 imagens, respectivamente, como representado pela Figura 14. A Tabela 3 detalha os hiperparâmetros utilizados nesta etapa.

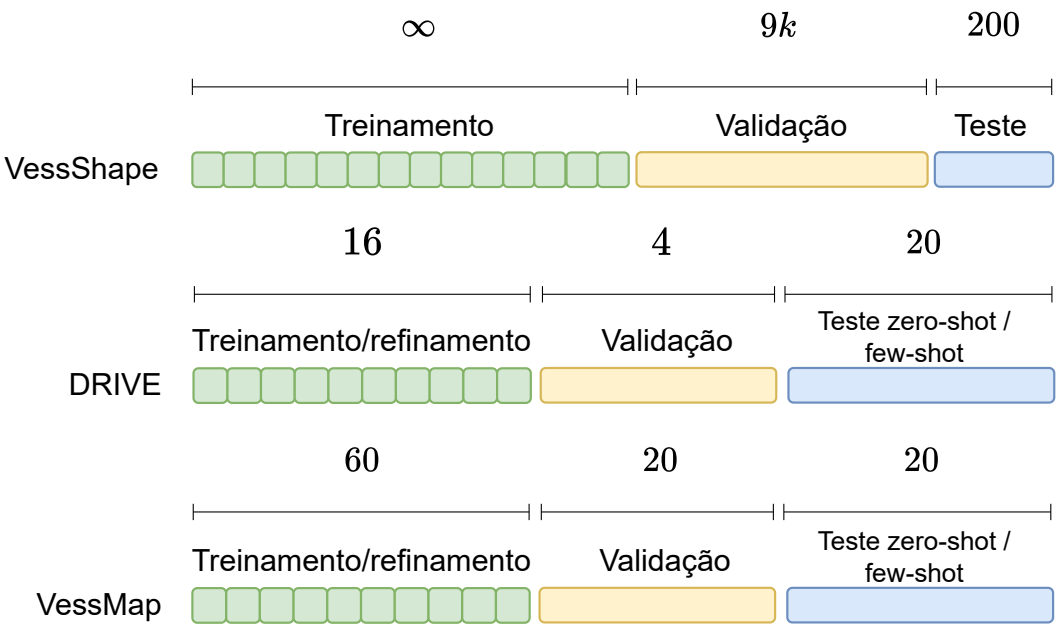
4.3.3 Refinamento dos modelos

Uma vez pré-treinados no *VessShape*, os modelos passam por um processo de refinamento para se especializarem nos domínios de destino. Para isso, foi desenvolvido um protocolo sistemático de FT para o regime de poucas amostras (FSL), aplicado aos conjuntos de dados DRIVE e *VessMap*. O objetivo central é quantificar o ganho de desempenho dos modelos à medida que o número de exemplos rotulados, utilizados para a adaptação, aumenta progressivamente. Em todos os experimentos de refinamento, os modelos (VSUNet18 e VSUNet50) partem dos pesos previamente otimizados com o *VessShape*.

4.3.3.1 Organização dos conjuntos de dados de destino

Para cada um dos conjuntos de dados de destino D, as imagens foram divididas em três subconjuntos disjuntos, como representado pela Figura 14:

Figura 14 – Representação da separação dos conjuntos de imagens usados neste trabalho: *VessShape*, DRIVE e *VessMap*. Para o *VessShape*, o subconjunto de treinamento é indicado com o tamanho infinito, pois, na prática, o volume de imagens sintéticas e diversas geradas depende da quantidade de épocas definida para o treinamento.



Fonte: Próprio autor

- $\mathcal{V}_{\text{train}}$: conjunto de imagens rotuladas para treino do modelo e usado no processo amostragem progressiva no regime FSL.
- $\mathcal{V}_{\text{val}}$: conjunto de validação utilizado para o monitoramento da métrica de desempenho Dice durante o treinamento e ajuste dos hiperparâmetros.
- $\mathcal{V}_{\text{test}}$: conjunto de teste, mantido isolado para a avaliação final do desempenho dos modelos.

Para o DRIVE, foram utilizadas 16 imagens para $\mathcal{V}_{\text{train}}$, 4 para $\mathcal{V}_{\text{val}}$ e as 20 imagens do conjunto de teste oficial para $\mathcal{V}_{\text{test}}$. Para o *VessMap*, a divisão foi de 60 imagens para $\mathcal{V}_{\text{train}}$, 20 para $\mathcal{V}_{\text{val}}$ e 20 para $\mathcal{V}_{\text{test}}$.

4.3.3.2 Protocolo de refinamento com amostragem progressiva

O refinamento é conduzido por meio de um processo de amostragem progressiva, no qual os modelos são treinados com um número crescente de exemplos. Para garantir a reprodutibilidade dos experimentos, todas as sementes de aleatoriedade nos processos de amostragem e treinamento foram fixadas. Para cada conjunto de dados, foi definido

um conjunto ordenado de tamanhos de amostra $\mathcal{N} = \{n_1, n_2, \ldots, n_K\}$, com $n_1 = 1$ e n_K sendo o número total de imagens no subconjunto de treinamento. As sequências utilizadas foram $\mathcal{N} = \{1, 2, 4, \ldots, 16\}$ para o DRIVE e $\mathcal{N} = \{1, 2, 4, \ldots, 20\}$ para o *VessMap*.

Para cada tamanho de amostra $n \in \mathcal{N}$, foram realizadas $R = 5$ execuções independentes para garantir a robustez estatística dos resultados. Em cada execução r, um subconjunto de treinamento $\mathcal{V}_{\text{train}}^{(n,r)}$ é amostrado sem reposição a partir do conjunto $\mathcal{V}_{\text{train}}$:

$$\mathcal{V}_{\text{train}}^{(n,r)} = \text{sample}(\mathcal{V}_{\text{train}}, n) \tag{12}$$

A fim de assegurar a diversidade entre execuções para cada n, mantém-se um registro das combinações já selecionadas e, nas execuções subsequentes, a amostragem ocorre sem reposição no nível da combinação. Somente subconjuntos inéditos são admitidos enquanto houver combinações disponíveis. Uma combinação previamente utilizada só pode reaparecer quando o espaço de combinações sem reposição estiver integralmente esgotado para aquele ou quando o limite operacional de tentativas para localizar uma nova combinação for alcançado, cenário mais provável quando n é elevado ou igual ao total de amostras disponíveis.

Dentro de cada uma dessas execuções, o processo de refinamento é repetido $S = 3$ vezes. Em cada repetição, otimiza-se o modelo para minimizar a perda de entropia cruzada sobre o subconjunto $\mathcal{V}_{\text{train}}^{(n,r)}$. O modelo resultante da última época (*checkpoint)* é, então, avaliado no conjunto de teste $\mathcal{V}_{\text{test}}$. Essa abordagem permite decompor a variância dos resultados, isolando a variabilidade decorrente do processo de treinamento daquela causada pelas diferentes combinações de amostras.

Escolhe-se o *checkpoint* da última época porque, no protocolo adotado, o treinamento entra em platô e o desempenho do modelo na validação estabiliza, fazendo com que o estado final seja indistinguível (ou muito próximo) do melhor *checkpoint* anterior.

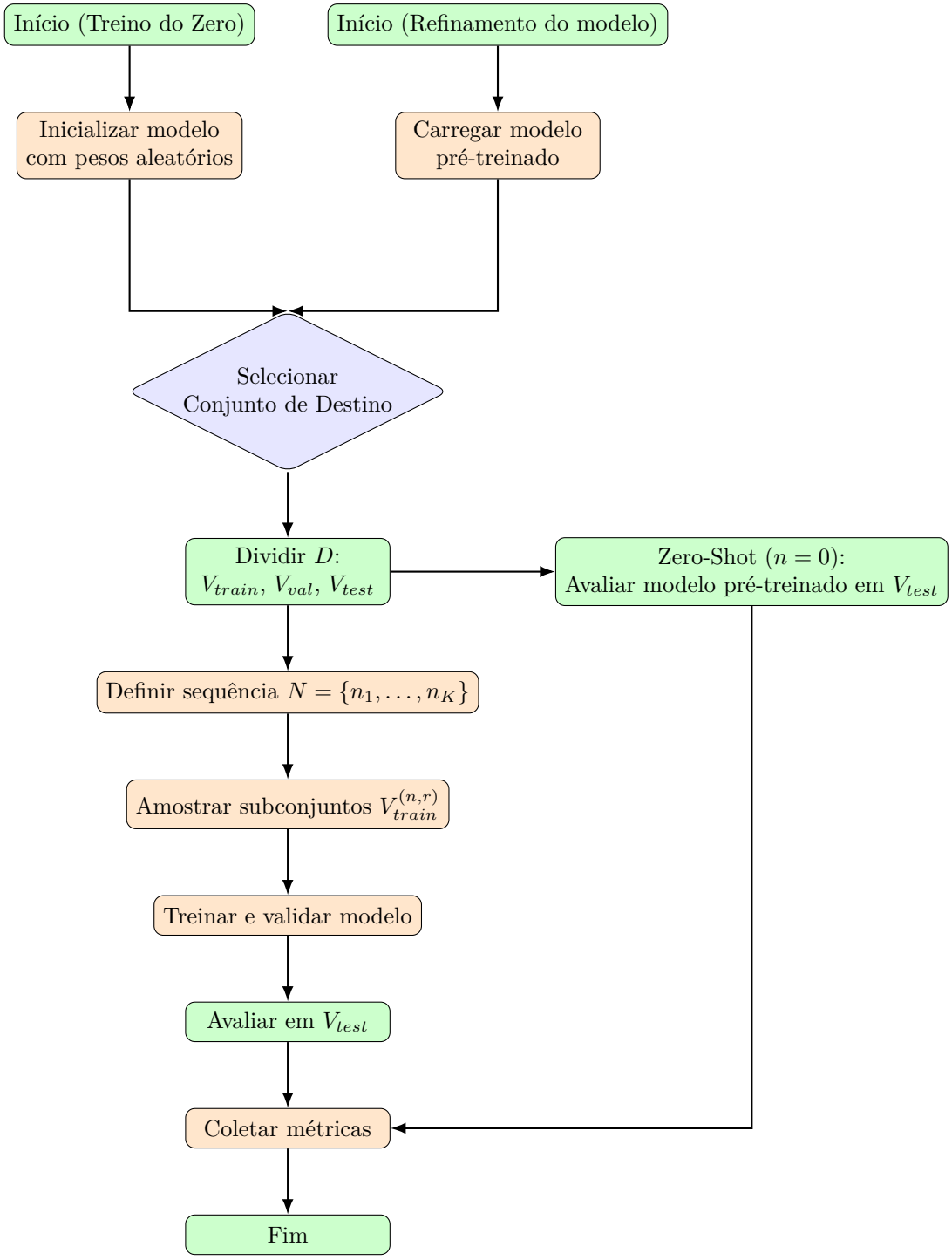
Adicionalmente, o protocolo de avaliação inclui o caso de ZSL ($n = 0$), no qual o desempenho do modelo pré-treinado é medido diretamente no conjunto de teste $\mathcal{V}_{\text{test}}$ sem qualquer etapa de adaptação ao domínio de destino D.

A Figura 15 ilustra o fluxo do processo experimental, e a Figura 16 detalha o protocolo de amostragem progressiva utilizado no treinamento do zero e também no refinamento dos modelos.

4.3.3.3 Treinamento do zero nos conjuntos de destino

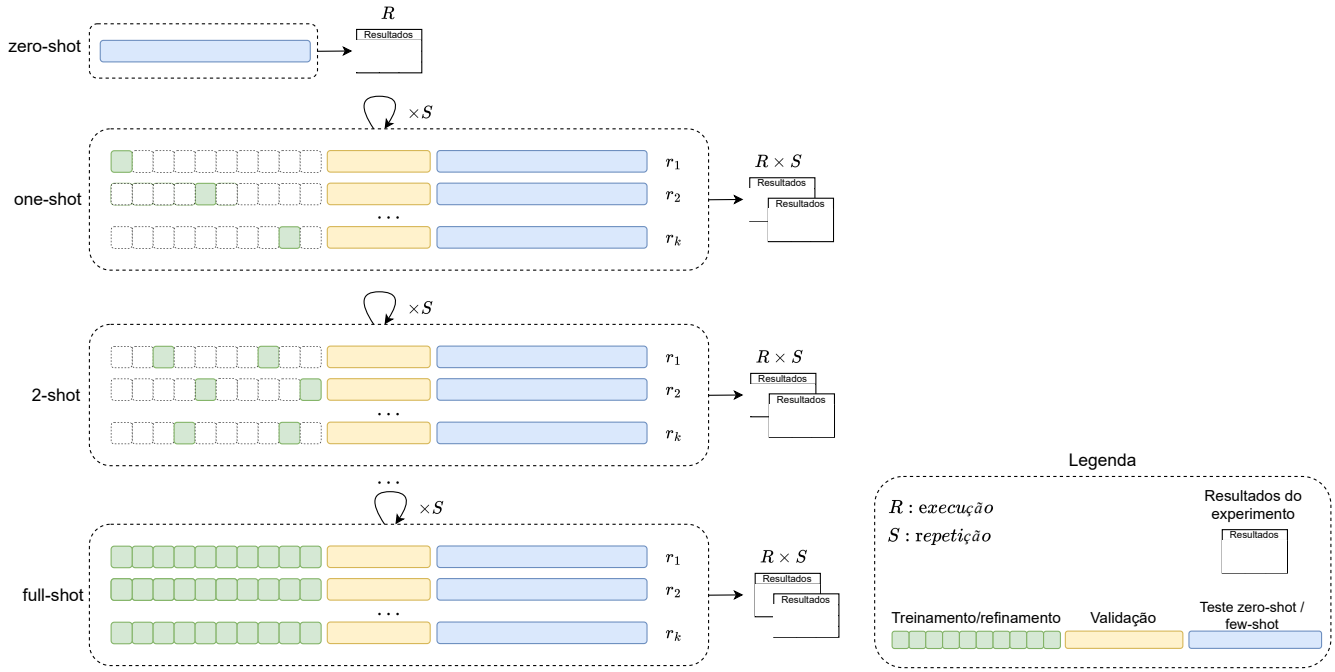
Para estabelecer uma base de comparação e quantificar o impacto do pré-treinamento, foram treinados modelos diretamente nos conjuntos de dados de destino, DRIVE e *VessMap*, sem qualquer etapa de pré-treinamento sintético. Estes modelos, denominados U-Net18 e U-Net50, seguem exatamente o mesmo protocolo de treinamento progressivo utilizado no refinamento, garantindo que a única variável de diferença seja a ausência dos pesos pré-treinados com o *VessShape*, como mostrado na Figura 15.

Figura 15 – Fluxograma dos dois cenários experimentais. O ramo da esquerda representa o treinamento do zero, enquanto o da direita ilustra o processo de transferência de aprendizado e a avaliação *zero-shot*.



Fonte: Próprio autor

Figura 16 – O diagrama ilustra os dois cenários experimentais: o treinamento do zero (U-Net) e o refinamento (VSUNet). Ambos os cenários são submetidos ao mesmo protocolo de treinamento com um número crescente de amostras (n) dos domínios de destino (DRIVE ou VessMap), permitindo a avaliação em regimes de FSL (one-shot até full-shot). Para garantir a robustez estatística, para cada tamanho de amostra n, o protocolo é executado $R = 5$ vezes (execuções independentes com diferentes amostras). Dentro de cada uma dessas execuções, o treinamento é repetido $S = 3$ vezes para avaliar a variabilidade do próprio processo de otimização. O cenário ZSL é uma avaliação direta do modelo VSUNet pré-treinado no conjunto de teste.



Fonte: Próprio autor

Neste cenário, o aprendizado do modelo é exposto a uma fase inicial mais desafiadora, pois depende exclusivamente das características, incluindo as de textura, presentes nas poucas amostras dos dados de destino. É importante notar que, para estes modelos, não se aplica o caso de ZSL, uma vez que não há um estado inicial com conhecimento prévio antes de o modelo observar ao menos uma imagem rotulada.

A comparação direta entre o desempenho dos modelos VSUNet (pré-treinados) e U-Net (treinados do zero) permite, portanto, isolar e medir o ganho proporcionado pelo viés de forma transferido a partir do *VessShape*, mantendo todos os outros fatores experimentais controlados.

4.4 Avaliação dos experimentos

A avaliação dos experimentos foi projetada para fornecer uma análise abrangente e comparativa do desempenho dos modelos, focando tanto em aspectos quantitativos quanto qualitativos. A metodologia permite medir rigorosamente o impacto da transferência de aprendizado com viés de forma.

A metodologia de avaliação, portanto, foi estruturada para responder a questões de pesquisa fundamentais, guiadas pela hipótese central de que os modelos refinados a partir de um pré-treinamento com viés de forma terão desempenho superior. As principais perguntas a serem respondidas são:

1. Os modelos pré-treinados com o *VessShape* e subsequentemente refinados (*VSUNet*) apresentam um desempenho superior aos modelos treinados inteiramente do zero (*U-Net*) nos conjuntos de destino?

2. Qual arquitetura de rede neural (ResNet18 vs. ResNet50) oferece o melhor balanço entre desempenho, custo computacional e capacidade de generalização em cenários com poucos dados para a tarefa?

3. Qual a contribuição do pré-treinamento com o *VessShape* para a adaptação aos domínios de destino, e como a customização de seus parâmetros de geração poderia impactar os resultados, dada a disparidade visual entre o conjunto sintético e os conjuntos de dados de destino?

4.4.1 Análise quantitativa

A análise quantitativa é realizada com base em um conjunto de métricas consolidadas para tarefas de segmentação. Como métrica principal, adota-se o Coeficiente de Similaridade Dice, por ser um índice amplamente utilizado e estabelecido na literatura para avaliação de segmentação de imagens médicas. O Dice é complementado por outras métricas-chave, como Acurácia (Acc), Intersecção sobre União (IoU), Precisão (Prec) e Revocação (Rec), que em conjunto oferecem uma visão detalhada do desempenho.

A avaliação consiste em analisar o desempenho em função do número de amostras de treinamento. Para isso, são geradas curvas de aprendizado que apresentam o valor médio do Dice em relação ao tamanho n do subconjunto de treinamento. Essa abordagem é fundamental para quantificar a eficiência de dados e a velocidade de convergência de cada modelo, revelando quão rapidamente eles se adaptam ao domínio de destino com um número limitado de exemplos.

Dado que o protocolo experimental inclui múltiplas execuções ($R = 5$) e repetições ($S = 3$) para cada tamanho de amostra n, uma grande quantidade de dados é coletada em todas as rodadas. Isso permite não apenas o cálculo da média de desempenho, mas

também a análise da sua variância (desvio-padrão). A análise da variância é crucial para avaliar a estabilidade e a confiabilidade dos modelos, indicando o quão consistentes são seus resultados diante de diferentes subconjuntos de dados de treinamento. Os resultados agregados são apresentados de duas formas: por meio de gráficos, que ilustram as tendências e as curvas de aprendizado, e por meio de tabela, que fornece os valores numéricos para comparações detalhadas.

4.4.2 Análise qualitativa

Além da avaliação quantitativa, a metodologia inclui uma análise qualitativa para complementar os resultados numéricos. Esta etapa consiste na inspeção visual das máscaras de segmentação geradas pelos diferentes modelos (VSUNet e U-Net) nos distintos regimes de treinamento ZSL e FSL. A comparação direta das imagens de saída com o padrão-ouro e entre os diferentes cenários permite identificar pontos fortes e fracos de cada abordagem, como a capacidade de delinear vasos de baixo calibre, a preservação da continuidade das estruturas vasculares e a ocorrência de falsos positivos ou negativos em regiões específicas.

Capítulo 5

Resultados

Este capítulo apresenta os resultados dos experimentos descritos na metodologia, com o objetivo de responder às questões de pesquisa levantadas. A análise é dividida em duas frentes: uma avaliação quantitativa, baseada em métricas de desempenho, e uma avaliação qualitativa, focada na inspeção visual das segmentações produzidas. Os dados foram agregados e são apresentados por meio de gráficos de curvas de aprendizado e tabelas comparativas.

5.1 Análise quantitativa

A análise quantitativa foca no desempenho dos modelos em função do número de amostras utilizadas para o treinamento ou refinamento.

5.1.1 Desempenho comparativo: Pré-treinamento vs. treinamento do zero

Primeiramente, analisou-se as métricas de desempenho do pré-treinamento dos modelos VSUNet, presentes na Tabela 4. Ambas as variantes apresentaram performance comparável, com um valor de Dice ligeiramente melhor para o modelo de maior profundidade, VSUNet50. Os modelos conseguiram segmentar com sucesso as imagens do VessShape, o que indica que as características de forma foram aprendidas.

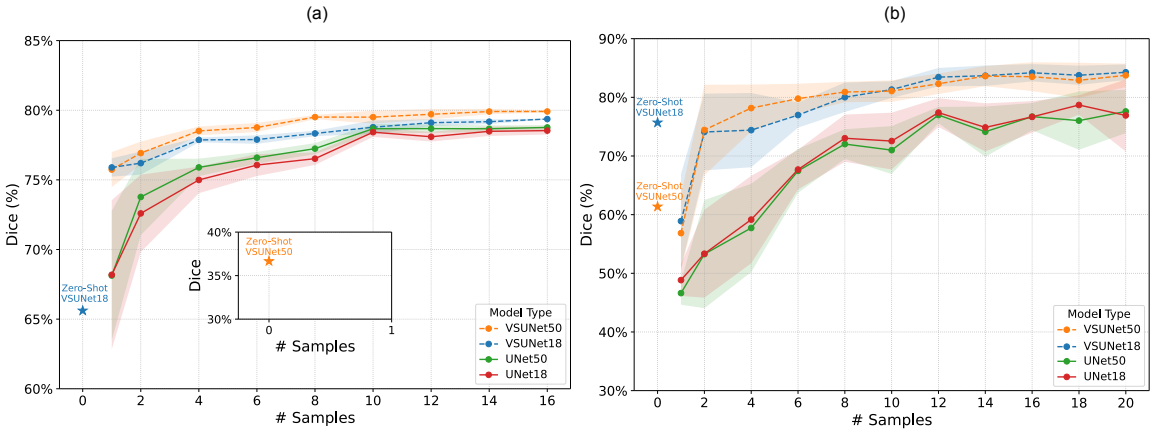
Para responder à primeira pergunta de pesquisa, comparou-se diretamente o desempenho dos modelos pré-treinados (VSUNet) com os modelos treinados inteiramente do zero (U-Net). As curvas de aprendizado na Figura 17 revelam diferenças notáveis no comportamento dos dois cenários.

Tabela 4 – Desempenho das variantes VSUNet após pré-treinamento no *VessShape*. Valores, em porcentagem, são média ± desvio-padrão no conjunto de teste fixo do *VessShape*.

Métrica	**VSUNet50**	**VSUNet18**
Dice	86.1 ± 2.2	85.9 ± 7.7
Acc	96.0 ± 0.8	95.6 ± 3.7
IoU	75.8 ± 3.2	76.1 ± 9.6
Prec	78.0 ± 3.7	77.4 ± 9.6
Rec	96.4 ± 1.2	97.4 ± 1.8

Fonte: Próprio autor

Figura 17 – Desempenho em Dice para poucos exemplos e *zero-shot* em (a) DRIVE e (b) *VessMAP*. As curvas mostram a média do Dice sobre R=5 execuções e S=3 repetições para cada tamanho n. As áreas sombreadas representam o desvio-padrão entre execuções. O quadro inserido em (a) mostra o Dice *zero-shot* (n=0) para VSUNet50 em DRIVE, muito inferior aos demais cenários.



Fonte: Próprio autor

Em ambos os conjuntos de destino, DRIVE e *VessMap*, os modelos VSUNet demonstram uma vantagem significativa no regime de poucas amostras. Com apenas uma amostra de treinamento, os modelos pré-treinados alcançaram um desempenho no Dice de 7 a 10 pontos percentuais acima das variantes U-Net, conforme detalhado na Tabela 5. Além disso, as curvas dos modelos VSUNet apresentam uma convergência mais rápida em comparação com os modelos U-Net. Essa diferença é acentuada pela análise da variância — áreas sombreadas no gráfico da Figura 17 — que tende a ser menor para os modelos pré-treinados, indicando maior estabilidade e robustez no treinamento.

Mesmo quando todo o conjunto de treinamento é utilizado, uma lacuna de desempenho, ainda que pequena, permanece entre os dois cenários, sugerindo que o viés de forma adquirido com o *VessShape* continua a oferecer benefícios mesmo em regimes com mais dados disponíveis. Esses resultados respondem à primeira pergunta, confirmando que o

Tabela 5 – Segmentação com poucos exemplos e em zero-shot no VessMAP e DRIVE. Valores, em porcentagem, são média ± desvio-padrão sobre execuções repetidas avaliadas no subconjunto de teste de cada conjunto de imagens de destino. Avaliações zero-shot decorrem de única inferência e, portanto, têm desvio-padrão zero. Adicionalmente, para facilitar a leitura, escolheu-se apenas os tamanhos de exemplos 0, 1 e tamanho total do conjunto de imagens correspondente.

Dataset	**#Exemplos**	**Modelo**	**Dice**	**Acc**	**IoU**	**Prec**	**Rec**
DRIVE	0	VSUNet18	**65.6** ± 0.0	90.7 ± 0.0	49.0 ± 0.0	62.9 ± 0.0	69.9 ± 0.0
		VSUNet50	36.7 ± 0.0	88.8 ± 0.0	23.0 ± 0.0	72.8 ± 0.0	27.5 ± 0.0
	1	VSUNet18	**75.9** ± 0.7	94.1 ± 0.2	61.2 ± 0.9	78.7 ± 2.1	74.1 ± 1.9
		VSUNet50	75.7 ± 1.3	93.9 ± 0.6	61.1 ± 1.6	77.3 ± 4.6	75.4 ± 3.9
		UNet18	68.2 ± 5.4	91.9 ± 1.6	52.3 ± 5.8	71.7 ± 7.9	69.0 ± 11.8
		UNet50	68.1 ± 4.6	91.6 ± 3.1	52.3 ± 5.0	72.2 ± 9.9	69.0 ± 11.0
	16	VSUNet18	79.4 ± 0.0	95.0 ± 0.0	65.8 ± 0.0	83.3 ± 0.4	76.2 ± 0.3
		VSUNet50	**79.9** ± 0.1	95.2 ± 0.0	66.6 ± 0.1	84.6 ± 0.3	76.2 ± 0.4
		UNet18	78.5 ± 0.3	94.6 ± 0.1	64.7 ± 0.4	79.5 ± 0.5	78.1 ± 0.4
		UNet50	78.8 ± 0.2	94.7 ± 0.1	65.0 ± 0.2	80.7 ± 0.6	77.4 ± 0.4
VessMAP	0	VSUNet18	**75.7** ± 0.0	88.6 ± 0.0	61.6 ± 0.0	84.6 ± 0.0	69.6 ± 0.0
		VSUNet50	61.4 ± 0.0	81.7 ± 0.0	47.2 ± 0.0	74.6 ± 0.0	60.5 ± 0.0
	1	VSUNet18	**58.9** ± 8.0	70.5 ± 20.5	45.5 ± 8.8	67.5 ± 20.5	73.8 ± 14.8
		VSUNet50	56.9 ± 6.3	70.8 ± 20.3	43.9 ± 7.1	68.3 ± 20.8	70.0 ± 16.7
		UNet18	48.8 ± 2.7	67.4 ± 18.4	36.2 ± 3.2	68.2 ± 20.4	62.2 ± 21.0
		UNet50	46.6 ± 2.0	66.5 ± 18.1	34.3 ± 2.5	67.4 ± 20.6	60.4 ± 21.5
	20	VSUNet18	**84.3** ± 1.3	90.1 ± 2.1	73.9 ± 1.8	82.5 ± 4.9	88.8 ± 4.8
		VSUNet50	83.7 ± 2.0	90.5 ± 2.5	73.2 ± 2.6	85.1 ± 3.2	85.0 ± 2.7
		UNet18	76.9 ± 6.2	88.0 ± 3.4	64.5 ± 7.5	86.4 ± 4.5	73.7 ± 9.4
		UNet50	77.6 ± 3.7	88.0 ± 3.5	65.3 ± 4.6	85.5 ± 4.1	75.2 ± 5.1

Fonte: Próprio autor

pré-treinamento com o *VessShape* leva a um desempenho superior na segmentação dos vasos sanguíneos, avaliado pela métrica Dice.

5.1.2 Análise das Arquiteturas: ResNet18 vs. ResNet50

A segunda pergunta de pesquisa aborda o balanço entre a complexidade da arquitetura, o desempenho e o custo computacional. A análise dos resultados mostra que o modelo menor, VSUNet18, oferece um equilíbrio notavelmente eficiente.

Conforme a Tabela 5, o VSUNet18 apresentou um desempenho em ZSL superior ao VSUNet50 em ambos os conjuntos de destino. Apesar de ser uma arquitetura menor e ter sido treinado com um volume de dados sintéticos significativamente inferior, o VSUNet18 se manteve competitivo com o VSUNet50 em todos os cenários de FSL. Este achado sugere que, para a tarefa de transferir um padrões de forma, uma arquitetura mais enxuta pode ser mais vantajosa em tarefas de transferência de conhecimento quando o $\mathcal{D}_T$ possui poucos exemplos anotados, oferecendo um excelente balanço entre eficiência computacional e

capacidade de generalização.

5.1.3 Contribuição do Pré-treinamento e o impacto da disparidade de domínio

A terceira pergunta de pesquisa investiga a contribuição do pré-treinamento e o efeito da disparidade visual entre os domínios. Um dos resultados mais relevantes é a capacidade de generalização dos modelos VSUNet, evidenciada pelo seu desempenho em ZSL. Os modelos conseguem segmentar vasos em domínios visualmente distintos — *VessMap*, com vasos claros sobre fundo escuro, e DRIVE, com vasos escuros sobre fundo claro — sem qualquer treinamento específico no $\mathcal{D}_T$. Essa habilidade sugere que os modelos de fato aprenderam representações de forma robustas, que são aplicáveis independentemente das características de aparência da imagem.

Contudo, um fenômeno contraintuitivo é observado na Figura 17(b), no conjunto *VessMap*: o desempenho dos modelos VSUNet diminui ao passar do cenário ZSL para OSL ($n = 1$), antes de se recuperar com mais amostras. Esse comportamento pode ser atribuído ao esquecimento catastrófico (MCCLOSKEY; COHEN, 1989), no qual o refinamento com um único exemplo, muito específico, leva o modelo a perder parte do conhecimento generalista previamente adquirido. No entanto, a recuperação do modelo sugere que o viés de forma inicial é robusto e pode ser reforçado.

Essa queda de desempenho no cenário OSL não é um fenômeno isolado e já foi documentada em modelos de grande escala, como o CLIP (RADFORD et al., 2021). A forte capacidade ZSL de modelos como o CLIP deriva do aprendizado de regras gerais que podem ser ativadas por meio de instruções textuais (do inglês, *prompts*), como "uma imagem de {rótulo da classe}". De forma análoga, o pré-treinamento com o *VessShape* neste trabalho induz um forte viés de forma, que funciona como um conhecimento explícito sobre a geometria vascular. É esse conhecimento prévio que permite ao modelo realizar inferências eficazes no regime ZSL, buscando por esses padrões geométricos mesmo em domínios não vistos.

Contudo, o refinamento com uma única amostra força o modelo a otimizar seus pesos com base em informação muito restrita. Essa informação contém não apenas a forma desejada, mas também detalhes intrínsecos daquela instância particular, como seus padrões de ruído e textura. Como consequência, o modelo tende a memorizar esses detalhes específicos em vez de aprender os atributos gerais do novo domínio, provocando então um sobreajuste.

Curiosamente, esse fenômeno não ocorre no conjunto DRIVE. Uma possível explicação é que a disparidade de domínio entre o *VessShape* e o DRIVE é maior, possivelmente porque as geometrias genéricas do *VessShape* — vasos mais curvilíneos e espessos — são visualmente menos alinhadas às dos vasos de baixo calibre e densamente ramificados do

DRIVE. Nesse caso, a primeira amostra do DRIVE fornece informação nova e útil que guia o modelo na direção correta, em vez de causar sobreajuste. Isso indica que a customização dos parâmetros do *VessShape* para mimetizar um domínio alvo é uma direção promissora para otimizar ainda mais a transferência de conhecimento.

5.2 Análise qualitativa

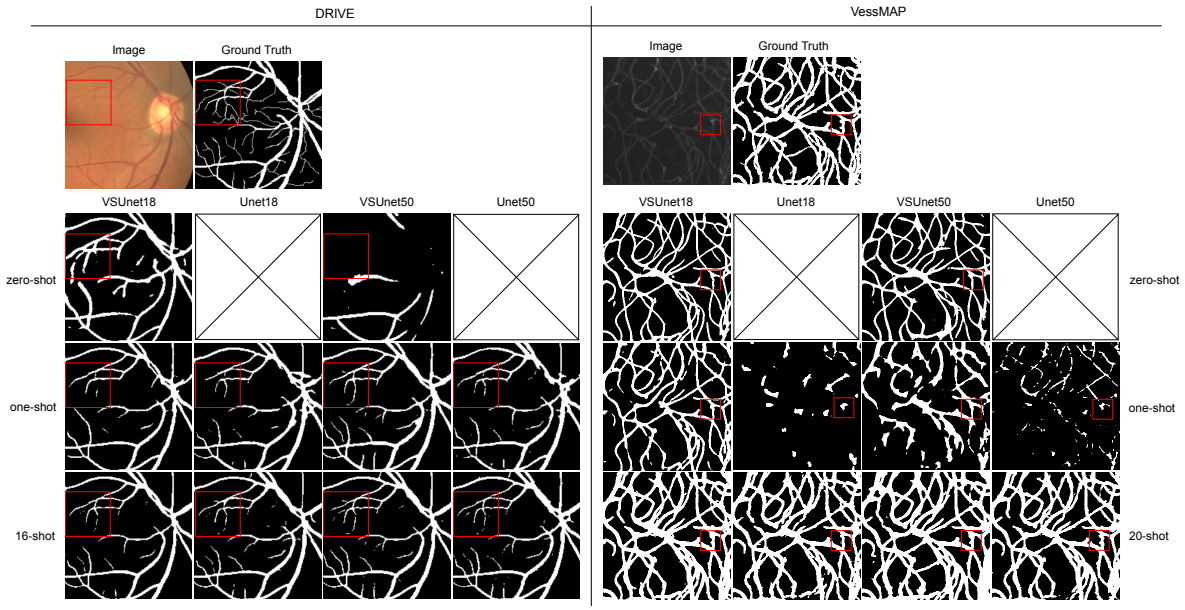
Para complementar a análise quantitativa, foi realizada uma inspeção visual das segmentações geradas. A Figura 18 detalha as diferenças entre os resultados dos modelos e o impacto do viés de forma em relação à qualidade de segmentação em regiões desafiadoras, como vasos de baixo calibre e ramificações complexas. No caso das U-Nets, não é apresentada visualização em regime de ZSL porque esses modelos foram treinados do zero neste estudo, sem carregamento de pesos pré-treinados. Consequentemente, não há configuração ZSL aplicável para essas arquiteturas dentro do protocolo adotado.

No conjunto DRIVE, observa-se que o VSUNet18 em modo ZSL é capaz de segmentar a estrutura vascular principal, embora o VSUNet50 apresente muitas falhas de falsos negativos. Com apenas uma amostra, ambos os modelos VSUNet se adaptam rapidamente, e o VSUNet50 passa a segmentar com mais qualidade os vasos finos. Semelhantemente, as variantes U-Net também segmentam a região de interesse com razoável qualidade no exemplo avaliado. Com 16 amostras, por sua vez, as diferenças visuais entre todos os modelos tornam-se mínimas.

No *VessMap*, a análise visual confirma os achados quantitativos, especialmente em uma região de interesse destacada que contém uma dilatação vascular com mudança brusca de intensidade. No regime ZSL, os modelos VSUNets são capazes de capturar a topologia geral da vasculatura, mas falham em delinear a forma específica da dilatação. O cenário OSL evidencia a queda de desempenho observada nos gráficos, com estruturas descontínuas nos VSUNets, enquanto os modelos U-Net, sem conhecimento prévio, produzem segmentações de baixíssima qualidade. Ainda assim, é notável que o VSUNet18 já apresenta um refinamento inicial da geometria da dilatação com apenas um exemplo. Com 20 amostras, o desempenho de todos os modelos melhora significativamente, embora os U-Net ainda exibam pequenas falhas de continuidade em comparação aos mapas mais coesos e completos gerados pelos VSUNets.

De forma geral, a análise qualitativa reforça a capacidade de generalização dos modelos pré-treinados, um dos resultados mais relevantes deste trabalho. Fica evidente a habilidade dos modelos VSUNets de segmentar vasos em domínios visualmente distintos, o que sugere o aprendizado de representações de forma robustas e aplicáveis a diferentes características visuais. Destaca-se também o desempenho do VSUNet18 que, apesar de ser uma arquitetura menor, apresentou-se competitivo em todos as avaliações, corroborando com o balanço entre desempenho e custo computacional.

Figura 18 – Comparação visual qualitativa do desempenho de segmentação em DRIVE e VessMAP. A figura mostra as saídas das variantes VSUNet e U-Net em diferentes regimes: zero-shot, one-shot e few-shot (16 para DRIVE e 20 para VessMAP). Para U-Nets, não há saída zero-shot. Em vermelho destacamos regiões de interesse para facilitar a comparação: uma área com vasos de baixo calibre no DRIVE e uma dilatação vascular no VessMAP.



Fonte: Próprio autor

Capítulo 6

Considerações Finais

A segmentação de vasos sanguíneos é uma tarefa fundamental na análise de imagens médicas, mas seu avanço é frequentemente limitado pela escassez de dados anotados e pela dificuldade de generalização dos modelos entre diferentes modalidades de imagem. A tendência das redes neurais convolucionais em aprender características de textura, em detrimento da forma, agrava esse desafio, limitando o desempenho quando os modelos são aplicados a novos domínios com aparências distintas.

Diante desse contexto, este trabalho investigou a hipótese de que a transferência de aprendizado, a partir de um domínio sintético projetado para instilar um forte viés de forma, poderia levar a modelos mais robustos e eficientes em termos de dados. O objetivo central foi, portanto, desenvolver e avaliar uma metodologia que explorasse padrões geométricos para superar as limitações dos conjuntos de dados reais.

Para que este objetivo fosse alcançado, o trabalho foi composto por etapas cruciais. A principal contribuição foi o desenvolvimento do *VessShape*, um gerador de imagens sintéticas que combina geometrias tubulares, geradas proceduralmente, com uma ampla variedade de texturas. Essa abordagem foi projetada para forçar os modelos a aprenderem representações de forma, e não de aparência. Subsequentemente, foi estabelecido um protocolo de avaliação, comparando modelos pré-treinados no *VessShape* (VSUNet) com modelos treinados do zero (U-Net) em dois conjuntos reais e distintos, DRIVE (retinografia) e *VessMap* (microscopia do córtex), em regimes de aprendizado com zero (ZSL) e poucos exemplos (FSL).

Os resultados confirmaram a hipótese central deste trabalho. Os modelos VSUNet apresentaram um desempenho consistentemente superior aos U-Net, especialmente em cenários com pouquíssimas amostras, onde a vantagem no Coeficiente de Dice chegou a ser de 10 pontos percentuais com apenas um exemplo. Além disso, os modelos pré-

treinados demonstraram uma notável capacidade de generalização no regime ZSL, sendo capazes de segmentar vasos em domínios visualmente distintos sem qualquer treinamento prévio, validando que o aprendizado de representações de forma é uma estratégia robusta e aplicável a diferentes características visuais. A análise também revelou que a arquitetura menor VSUNet18, ofereceu um balanço entre desempenho e eficiência, superando o modelo de maior profundidade e parâmetros, VSUNet50, em cenários de ZSL.

Portanto, este trabalho demonstra que, para tarefas de segmentação com viés de forma fortes e consistentes, focar no aprendizado de características geométricas a partir de dados sintéticos é uma estratégia de pré-treinamento mais eficaz e generalizável do que depender exclusivamente dos poucos dados disponíveis no domínio de destino.

6.1 Trabalhos futuros

Com o objetivo de expandir o trabalho realizado, há diversas propostas para investigações futuras e aprimoramento da metodologia aqui apresentada.

1. A geração geométrica do *VessShape*, atualmente baseada em curvas de Bézier, pode ser estendida para incluir topologias vasculares mais complexas e biologicamente plausíveis, como bifurcações e estruturas em rede.

2. A metodologia pode ser expandida para o domínio 3D, permitindo a aplicação da estratégia de pré-treinamento a modalidades de imagem volumétricas, como Tomografia Computadorizada (TC) e Imagem por Ressonância Magnética (RM).

3. O pré-treinamento focado em forma pode ser explorado para a segmentação de outras estruturas biológicas tubulares, como neurônios, ductos ou vias aéreas.

4. Investigar o impacto de refinar apenas camadas específicas do codificador, em vez do modelo inteiro. Essa abordagem, conhecida como *layer freezing*, poderia otimizar a transferência de conhecimento, preservando as características de baixo nível aprendidas no domínio de origem.

5. Avaliar a eficácia do pré-treinamento com o *VessShape* no refinamento de modelos de arquitetura *Transformer*, como o *Segment Anything Model* (SAM) e suas variantes especializadas para o domínio médico, como o MedSAM.

Referências

CHRISTIANE, F. Wordnet: an electronic lexical database. **Computational Linguistics**, p. 292–296, 1998.

DONG, S.; WANG, P.; ABBAS, K. A survey on deep learning and its applications. **Computer Science Review**, v. 40, p. 100379, 2021. ISSN 1574-0137. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1574013721000198>.

FRAZ, M. M. et al. An ensemble classification-based approach applied to retinal blood vessel segmentation. **IEEE Transactions on Biomedical Engineering**, v. 59, n. 9, p. 2538–2548, 2012. ISSN 00189294.

FREITAS-ANDRADE, M. et al. Unbiased analysis of mouse brain endothelial networks from two-or three-dimensional fluorescence images. **Neurophotonics**, Society of Photo-Optical Instrumentation Engineers, v. 9, n. 3, p. 031916–031916, 2022.

FUKUSHIMA, K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. **Biological cybernetics**, Springer, v. 36, n. 4, p. 193–202, 1980. Disponível em: <https://link.springer.com/article/10.1007/BF00344251>.

GEIRHOS, R. et al. Imagenet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. In: **International Conference on Learning Representations**. [s.n.], 2019. Disponível em: <https://openreview.net/forum?id=Bygh9j09KX>.

GHAFOORIAN, M. et al. Transfer Learning for Domain Adaptation in MRI: Application in Brain Lesion Segmentation. In: **Medical Image Computing and Computer Assisted Intervention - MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III**. Springer-Verlag, 2017. p. 516–524. ISBN 978-3-319-66178-0. Disponível em: <https://doi.org/10.1007/978-3-319-66179-7_59>.

GIGANDET, X.; CUADRA, M. B.; THIRAN, J. **Satellite image segmentation and classification**. [S.l.], 2004. Disponível em: <https://infoscience.epfl.ch/server/api/core/bitstreams/c035ab08-fa7f-4f4a-ac94-94e5dcd05ed5/content>.

GONI, M. R. et al. Brain Vessel Segmentation Using Deep Learning—A Review. **IEEE Access**, v. 10, p. 111322–111336, 2022. ISSN 2169-3536.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. [S.l.]: MIT Press, 2016. <http://www.deeplearningbook.org>.

HE, K. et al. Deep residual learning for image recognition. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [s.n.], 2016. p. 770–778. Disponível em: <https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html>.

HOOVER, A. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. **IEEE Transactions on Medical Imaging**, IEEE, v. 19, n. 3, p. 203–210, 2000. ISSN 02780062.

ISLAM, A. et al. **SHAPE OR TEXTURE: UNDERSTANDING DISCRIMINATIVE FEATURES IN CNNS**. 2021.

Jia Deng et al. ImageNet: A large-scale hierarchical image database. In: **2009 IEEE Conference on Computer Vision and Pattern Recognition**. IEEE, 2009. p. 248–255. ISBN 978-1-4244-3992-8. Disponível em: <http://www.image-net.org.https://ieeexplore.ieee.org/document/5206848>.

JIANG, Z. et al. Retinal blood vessel segmentation using fully convolutional network with transfer learning. **Computerized Medical Imaging and Graphics**, Pergamon, v. 68, p. 1–15, sep 2018. ISSN 0895-6111.

KARPATHY, A. **T-SNE Visualization of CNN Codes**. 2014. Disponível em: <https://cs.stanford.edu/people/karpathy/cnnembed/>.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. **Communications of the ACM**, AcM New York, NY, USA, v. 60, n. 6, p. 84–90, 2017. Disponível em: <https://dl.acm.org/doi/abs/10.1145/3065386>.

LACOSTE, B. et al. Sensory-Related Neural Activity Regulates the Structure of Vascular Networks in the Cerebral Cortex. Elsevier, v. 83, n. 5, p. 1117–1130, 2014. ISSN 0896-6273. Disponível em: <https://www.cell.com/neuron/abstract/S0896-6273(14)00645-X>.

LAMPERT, C. H.; NICKISCH, H.; HARMELING, S. Learning to detect unseen object classes by between-class attribute transfer. IEEE, p. 951–958, 2009. Disponível em: <https://ieeexplore.ieee.org/document/5206594/>.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **nature**, Nature Publishing Group UK London, v. 521, n. 7553, p. 436–444, 2015. Disponível em: <https://www.nature.com/articles/nature14539>.

LECUN, Y. et al. Backpropagation Applied to Handwritten Zip Code Recognition. **Neural Computation**, MIT Press - Journals, v. 1, n. 4, p. 541–551, dec 1989. ISSN 0899-7667. Disponível em: <https://ieeexplore.ieee.org/abstract/document/6795724>.

LEK, S.; PARK, Y. S. Artificial Neural Networks. **Encyclopedia of Ecology, Five-Volume Set**, Academic Press, p. 237–245, jan 2008.

LIANG, D. et al. Coronary angiography video segmentation method for assisting cardiovascular disease interventional treatment. **BMC Medical Imaging**, v. 20, n. 1, p. 65, jun. 2020. ISSN 1471-2342.

MCCLOSKEY, M.; COHEN, N. J. Catastrophic Interference in Connectionist Networks: The Sequential Learning Problem. In: BOWER, G. H. (Ed.). Academic Press, 1989, (Psychology of Learning and Motivation, v. 24). p. 109–165. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0079742108605368>.

MCDONALD, M. W. et al. An exercise mimetic approach to reduce poststroke deconditioning and enhance stroke recovery. v. 35, n. 6, p. 471–485, 2021. ISSN 1552-6844. Disponível em: <https://journals.sagepub.com/doi/full/10.1177/15459683211005019>.

OUELLETTE, J. et al. Vascular contributions to 16p11. 2 deletion autism syndrome modeled in mice. **Nature Neuroscience**, Nature Publishing Group US New York, v. 23, n. 9, p. 1090–1101, 2020. Disponível em: <https://www.nature.com/articles/s41593-020-0663-1>.

PAN, S. J.; YANG, Q. A Survey on Transfer Learning. 2009. Disponível em: <http://socrates.acadiau.ca/courses/comp/dsilver/NIPS95>.

QIN, Q.; CHEN, Y. A review of retinal vessel segmentation for fundus image analysis. v. 128, p. 107454, 2024. ISSN 0952-1976. Disponível em: <https://www.sciencedirect.com/science/article/pii/S095219762301638X>.

RADFORD, A. et al. Learning transferable visual models from natural language supervision. In: PMLR. **International conference on machine learning**. 2021. p. 8748–8763. Disponível em: <https://proceedings.mlr.press/v139/radford21a>.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: NAVAB, N. et al. (Ed.). **Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015**. Cham: Springer International Publishing, 2015. v. 9351, p. 234–241. ISBN 9783319245737 9783319245744. Disponível em: <http://link.springer.com/10.1007/978-3-319-24574-4_28>.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. **Nature**, v. 323, n. 6088, p. 533–536, out. 1986. ISSN 1476-4687. Disponível em: <https://www.nature.com/articles/323533a0>.

SHELHAMER, E.; LONG, J.; DARRELL, T. Fully Convolutional Networks for Semantic Segmentation. v. 39, n. 4, p. 640–651, 2017. ISSN 1939-3539.

SILVA, M. Viana da et al. A new dataset for measuring the performance of blood vessel segmentation methods under distribution shifts. **Plos one**, Public Library of Science San Francisco, CA USA, v. 20, n. 5, p. e0322048, 2025. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0322048>.

SONG, Y. et al. A Comprehensive Survey of Few-shot Learning: Evolution, Applications, Challenges, and Opportunities. 2022.

STAAL, J. et al. Ridge-based vessel segmentation in color images of the retina. **IEEE Transactions on Medical Imaging**, v. 23, n. 4, p. 501–509, apr 2004. ISSN 02780062.

WANG, Y. et al. Generalizing from a few examples: A survey on few-shot learning. **ACM computing surveys (csur)**, ACM New York, NY, USA, v. 53, n. 3, p. 1–34, 2020. Disponível em: <https://dl.acm.org/doi/abs/10.1145/3386252>.

ZHUANG, F. et al. A comprehensive survey on transfer learning. **Proceedings of the IEEE**, Ieee, v. 109, n. 1, p. 43–76, 2020. Disponível em: <https://ieeexplore.ieee.org/abstract/document/9134370>.

ZOETMULDER, R. et al. Domain- and task-specific transfer learning for medical segmentation tasks. v. 214, p. 106539, 2022. ISSN 0169-2607. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0169260721006131>.

ZOU, K. H. et al. Statistical validation of image segmentation quality based on a spatial overlap index1: scientific reports. **Academic Radiology**, v. 11, n. 2, p. 178–189, 2004. ISSN 1076-6332. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1076633203006718>.