

**UNIVERSIDADE FEDERAL DE SÃO CARLOS**

CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA

BACHARELADO EM ENGENHARIA DA COMPUTAÇÃO

**SEGMENTAÇÃO DE VASOS SANGUÍNEOS  
UTILIZANDO REDES NEURAIS  
CONVOLUCIONAIS - VISUALIZAÇÃO E  
ANÁLISE DE CORRELAÇÃO DOS MAPAS DE  
ATIVACÃO**

**WESLEY NOGUEIRA GALVÃO**

**ORIENTADOR: PROF. DR. CESAR HENRIQUE COMIN**

São Carlos – SP

2023



**UNIVERSIDADE FEDERAL DE SÃO CARLOS**

CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA

BACHARELADO EM ENGENHARIA DA COMPUTAÇÃO

**SEGMENTAÇÃO DE VASOS SANGUÍNEOS  
UTILIZANDO REDES NEURAIS  
CONVOLUCIONAIS - VISUALIZAÇÃO E  
ANÁLISE DE CORRELAÇÃO DOS MAPAS DE  
ATIVACÃO**

**WESLEY NOGUEIRA GALVÃO**

Trabalho de Conclusão de Curso apresentado ao curso de Engenharia de 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 Bacharel em Engenharia de Computação.

Orientador: Prof. Dr. Cesar Henrique Comin

São Carlos – SP

2023





**Wesley Nogueira Galvão**

**Segmentação de vasos sanguíneos utilizando Redes Neurais  
Convolucionais - Visualização e análise de correlação dos  
mapas de ativação**

Trabalho de Conclusão de Curso apresentado ao curso de Engenharia de 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 Bacharel em Engenharia de Computação.

Trabalho aprovado. São Carlos – SP, 29 de março de 2023:

---

**Prof. Dr. Cesar Henrique Comin**  
Orientador  
Universidade Federal de São Carlos

---

**Prof. Dr. Alan Demétrius Baria Valejo**  
Universidade Federal de São Carlos

---

**Prof. Dr. Alexandre Luis Magalhães Levada**  
Universidade Federal de São Carlos

São Carlos – SP  
2023

*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 primeiramente a Deus pela fé e força para seguir adiante na realização de um sonho.

Aos meus pais que não mediram esforços para me incentivar e dar o suporte no momento que mais precisei, dando sempre muito amor e sabedoria para que eu pudesse correr atrás dos meus objetivos.

Aos meus familiares que contribuíram com a minha educação por todos esses anos, e sempre acreditaram no meu sucesso.

Aos amigos que fiz durante essa jornada, com quem compartilho minhas risadas e alegrias.

A minha querida e amada Raissa Camelo, com quem compartilhei minhas dores, que me encorajou nos momentos de incerteza e me dedicou a sua atenção, fazendo os meus dias mais alegres.

Aos professores e professoras da Universidade Federal de São Carlos que me enriqueceram em conhecimento. Em especial, ao meu orientador Dr. César Henrique Comin, que me deu a oportunidade de desenvolver este projeto, compartilhou de forma exemplar seus conhecimentos e me orientou nas melhores decisões.

A vocês, meu muito obrigado!

*A ciência consiste em substituir o saber que parecia seguro por uma teoria, ou seja, por algo problemático.*  
*(José Ortega y Gasset)*



# RESUMO

As redes neurais convolucionais (CNN) são responsáveis pelo avanço da automatização de tarefas do campo da visão computacional, com classificação de imagens, reconhecimento de objetos e segmentação de imagens, pois se faz presente em diferentes nichos de pesquisa e da indústria. Em especial, as CNNs possibilitaram diversos avanços na área biomédica, que faz uso de CNNs para segmentação de imagens de raio-x e microscopia de células e vasos sanguíneos. Essas redes são conhecidas por serem profundas, compostas por várias camadas que extraem características das imagens, gerando os chamados mapas de ativação, e definem modelos complexos com milhões de parâmetros. Logo, existe a necessidade de estudar meios de visualização e interpretação dos mapas de ativação gerados. Além disso, investigar o nível de similaridade entre os mapas de ativação pode levar a uma melhor compreensão de como as redes geram suas representações de informações. Este estudo investiga formas de extrair representações aprendidas por CNNs e desenvolve métodos e ferramentas para visualizar mapas de ativação, além de calcular e analisar a correlação desses mapas em redes de segmentação com diferentes profundidades. Compreender a quantidade de mapas correlacionados e identificá-los pode ser útil para a aplicação de métodos que tornem as redes mais eficientes computacionalmente sem diminuir sua precisão. O trabalho também propõe investigar o grau de correlação e similaridade entre diferentes mapas produzidos pelas redes, seja por uma única rede ou comparando redes de diferentes profundidades. Dentre os principais resultados, foi verificado que mapas nas camadas mais profundas tendem a possuir maior redundância do que os demais mapas da rede, tendendo a formar blocos de mapas altamente correlacionados nas últimas camadas.

**Palavras-chave:** redes neurais convolucionais, segmentação de imagens, mapa de ativação, vasos sanguíneos, similaridade, correlação.





# ABSTRACT

Convolutional neural networks (CNN) are responsible for advancing the automation of tasks in the field of computer vision, with image classification, object recognition and image segmentation, as it is present in different research and industry areas. In particular, the biomedical area, which makes use of CNNs for segmentation of x-ray images and microscopy of cells and blood vessels. These networks are known to be deep, composed of several layers that extract features, called activation maps, from the images and generate complex models with millions of parameters. Therefore, it is important to develop approaches for visualization and interpretation of the generated activation maps. Furthermore, investigating the degree of similarity between representations can lead to a better understanding of how networks generate their information representations. This study investigates ways to extract representations produced by CNNs and develop methods and tools to visualize activation maps, in addition to calculating and analyzing the correlation of these maps in segmentation networks with different depths. Understanding the degree of correlation among the maps and identifying highly correlated maps can be useful for applying methods that make networks more computationally efficient without decreasing their accuracy. This work also proposes to investigate the degree of correlation and similarity between different maps produced by networks, either by a single network or by comparing networks of different depths. The study carried out found that CNNs produce redundant maps as they get deeper, tending to form blocks of highly correlated maps in the last layers.

**Keywords:** convolutional neural networks, image segmentation, feature map, blood vessels, similarity, correlation.



## LISTA DE FIGURAS

---

---

Figura 1 – Exemplo de mapas de ativação de uma U-Net. . . . .	15
Figura 2 – Conceitos básicos de Redes Neurais . . . . .	19
Figura 3 – Exemplo de uma CNN para classificação de dígitos manuscritos . . . . .	20
Figura 4 – Arquitetura da U-Net . . . . .	22
Figura 5 – Exemplo de segmentação resultante da aplicação de uma rede U-Net a uma imagem de microscopia de vasos sanguíneos . . . . .	22
Figura 6 – Ilustração da sobreposição de dois conjuntos de segmentação e respectivos valores do coeficiente DICE. . . . .	24
Figura 7 – Exemplo de comparação da qualidade de duas segmentações de um vaso sanguíneo. . . . .	25
Figura 8 – Exemplos de imagens e seus respectivos padrão-ouro amostrados. . . . .	34
Figura 9 – Imagem ruidosa . . . . .	34
Figura 10 – Representação de uma CNN para segmentação de imagens composta por $N$ camadas intermediárias. . . . .	35
Figura 11 – Representação simplificada da biblioteca CNN Explorer. Principais atributos e métodos . . . . .	39
Figura 12 – Esquemático do pipeline do cálculo de correlação . . . . .	40
Figura 13 – Exemplo de visualização dos mapas bem correlacionados. O número acima de cada figura indica o ID do mapa na respectiva camada. . . . .	43
Figura 14 – Exemplo de gráfico para visualizar proporção de mapas bem correlacionados por camada . . . . .	44
Figura 15 – Qualidade das segmentações dos modelos treinados . . . . .	47
Figura 16 – Amostra de imagens utilizadas na análise de correlação . . . . .	48
Figura 17 – Nível de correlação absoluta entre mapas de uma mesma CNN. Visualização por profundidade da rede . . . . .	49
Figura 18 – Proporção de mapas bem correlacionados na mesma CNN. Visualização por profundidade da rede . . . . .	50
Figura 19 – Nível de correlação absoluta entre mapas de CNNs distintas. Visualização por profundidade da rede . . . . .	51
Figura 20 – Proporção de mapas bem correlacionados entre CNNs distintas. Visualização por profundidade da rede . . . . .	52

Figura 21 – Mapas de ativação extraídos da CNN6. Bordas coloridas indicam exemplos de mapas similares na primeira e última camada. . . . . 53

## LISTA DE TABELAS

---

---

Tabela 1 – Exemplo de tabela de correlações entre mapas de uma rede com dois módulos convolucionais . . . . .	42
Tabela 2 – Exemplo de tabela de máximas correlações entre mapas de uma rede com dois módulos convolucionais . . . . .	43
Tabela 3 – Exemplo de tabela pivotada de máximas correlações entre mapas de uma rede com dois módulos convolucionais . . . . .	44
Tabela 4 – Informações do treinamento dos modelos . . . . .	46
Tabela 5 – Avaliação da qualidade das segmentações dos modelos treinados . . . . .	46

## LISTA DE SIGLAS

---

---

ANN	Rede Neural Artificial
CCA	Análise de Correlação Canônica
CKA	Índice de Alinhamento Centralizado do Kernel
CNN	Rede Neural Convolucional
CPU	Unidade Central de Processamento
DICE	Coefficiente de Similaridade de Sørensen
FCN	Rede Totalmente Convolucional
FN	Falso Negativo
FP	Falso Positivo
GPU	Unidade Gráfica de Processamento
GUI	Interface Gráfica de Usuário
IOU	Intersecção sobre União - Índice de Jaccard
MLP	Perceptron Multicamada
ReLU	Unidade Linear Retificada
TEA	Transtorno do Espectro Autista
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo

# SUMÁRIO

---

---

<b>CAPÍTULO 1–INTRODUÇÃO</b> . . . . .	<b>14</b>
1.1 Contexto e motivação . . . . .	14
1.2 Objetivos . . . . .	16
1.2.1 Geral . . . . .	16
1.2.2 Específicos . . . . .	16
1.3 Organização . . . . .	16
<b>CAPÍTULO 2–FUNDAMENTAÇÃO TEÓRICA</b> . . . . .	<b>18</b>
2.1 Aprendizado Profundo e Redes Neurais Convolucionais . . . . .	18
2.2 Segmentação de imagens utilizando CNNs . . . . .	20
2.2.1 Métricas para avaliação de segmentação de vasos sanguíneos . . . . .	23
2.2.1.1 Precisão . . . . .	23
2.2.1.2 Sensibilidade . . . . .	23
2.2.1.3 F1-Score . . . . .	23
2.2.1.4 Medidas de similaridade . . . . .	24
2.2.1.4.1 Coeficiente Dice . . . . .	24
2.2.1.4.2 Coeficiente IoU . . . . .	24
2.2.1.4.3 Coeficiente cIDice . . . . .	25
2.2.1.5 Medidas de distância e correlação entre mapas de ativação . . . . .	26
2.2.1.5.1 Distância euclidiana . . . . .	26
2.2.1.5.2 Similaridade cosseno . . . . .	27
2.2.1.5.3 Correlação de Pearson . . . . .	27
<b>CAPÍTULO 3–TRABALHOS RELACIONADOS</b> . . . . .	<b>28</b>
3.1 Interpretação de mapas de ativação . . . . .	28
3.2 Similaridade entre mapas de ativação de CNNs . . . . .	29
<b>CAPÍTULO 4–METODOLOGIA E DESENVOLVIMENTO</b> . . . . .	<b>32</b>
4.1 Natureza da pesquisa . . . . .	32
4.2 Ferramentas . . . . .	32
4.3 Conjunto de dados . . . . .	33
4.3.1 Seleção de um subconjunto de imagens . . . . .	33
4.4 Planejamento das arquiteturas de CNNs . . . . .	35
4.5 Estratégias de comparação de similaridade entre mapas de ativações . . . . .	36
4.5.1 Extração dos mapas de ativação . . . . .	36
4.5.2 Escolha da medida de similaridade . . . . .	36

4.5.3	Cálculo de similaridade entre mapas de ativação . . . . .	37
4.6	Desenvolvimento e organização de um pacote para exploração e análise de CNN	38
4.6.1	Arquitetura da biblioteca . . . . .	38
4.6.1.1	Classe ExtractCNNLayers . . . . .	38
4.6.1.2	Classe ExtractCNNMaps . . . . .	39
4.6.1.3	Classe CorrelationExplorer . . . . .	40
4.7	Exploração dos mapas de ativações de CNNs . . . . .	41
<b>CAPÍTULO 5–RESULTADOS . . . . .</b>		<b>45</b>
5.1	Treinamento e análise das CNNs . . . . .	45
5.2	Análise de correlação de mapas de ativação . . . . .	47
5.2.1	Correlação na mesma rede . . . . .	48
5.2.2	Correlação entre redes distintas . . . . .	50
<b>CAPÍTULO 6–CONCLUSÃO . . . . .</b>		<b>54</b>
6.1	Trabalhos futuros . . . . .	55
<b>REFERÊNCIAS . . . . .</b>		<b>57</b>



# Capítulo 1

## INTRODUÇÃO

---

---

### 1.1 Contexto e motivação

Vasos sanguíneos são estruturas importantes e vitais para seres vertebrados, pois realizam o transporte de gases e propagação de nutrientes para os tecidos celulares desses organismos (PITTMAN, 2011). Isto posto, há diversos estudos dessas estruturas publicados na literatura, tais como o aprimoramento de diagnóstico para detecção de doenças através de exames, como imagens de fundo de retina (KLUFAS et al., 2015), e pesquisas de fronteira em biologia. Por exemplo, a hipótese da relação entre o espectro autista e as alterações na formação de vasos sanguíneos do córtex, o que ocasiona obstrução do fluxo sanguíneo, proposta por (AZMITIA et al., 2016). Recentemente, essa hipótese foi verificada no trabalho publicado por (OUELLETTE et al., 2020), o qual sugere que a remoção do pedaço de cromossomo 16p11.2, característica observada no transtorno do espectro autista (TEA), pode causar alterações na estrutura dos vasos sanguíneos no córtex de camundongos.

Visto que há grande importância na análise de vasos sanguíneos, surge a necessidade de desenvolvimento e melhoria de métodos para captura e categorização dessas estruturas em imagens geradas por técnicas de imageamento, tal como a microscopia. Em consonância com as técnicas de obtenção e processamento de imagens, há a necessidade da utilização de abordagens computacionais, como aprendizado profundo por meio de Redes Neurais Convolucionais (CNNs), que servem de apoio ao especialista para a detecção automática de patologias em vasos sanguíneos.

As CNNs ganharam enorme importância na área de Visão Computacional na última década por serem muito eficientes nas tarefas de detecção de objetos, classificação e segmentação de imagens. Também se mostraram ferramentas promissoras para segmentação de vasos sanguíneos (KAPLAN et al., 2020). Essas redes são capazes de extrair características de uma imagem de entrada utilizando convoluções e funções de ativações ao longo das suas camadas intermediárias e retornar uma imagem segmentada, isto é, com os vasos sanguíneos identificados.

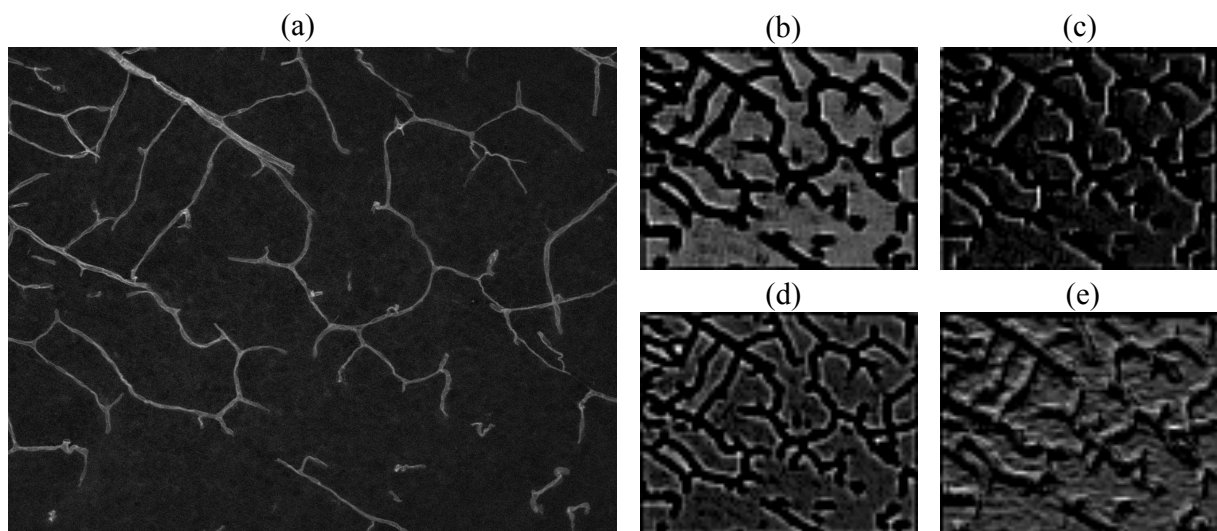
Existem variações importantes em relação às arquiteturas de aprendizado profundo para

segmentação de vasos sanguíneos. O trabalho de (LISKOWSKI; KRAWIEC, 2016), por exemplo, utiliza técnicas como *downsampling* e camadas totalmente conectadas, enquanto (FU et al., 2016) baseia a sua rede numa U-Net (RONNEBERGER et al., 2015) com supervisões diretas nas camadas intermediárias. Por outro lado, e mais recentemente, (TETTEH et al., 2020) propôs uma arquitetura especializada em segmentação de vasos sanguíneos que não utiliza a estratégia de *downsampling* e *upsampling*, comumente visto em arquiteturas baseadas na U-Net.

Contudo, na grande maioria das aplicações, as CNNs ainda são utilizadas como uma “caixa-preta”, ou seja, estritamente como uma função desconhecida que mapeia uma imagem de entrada em uma categoria ou imagem de rótulos. Portanto, visualizar e interpretar as características aprendidas pela rede ao longo das suas camadas pode auxiliar na compreensão dos processos de ativação e nível de discriminabilidade dos mapas de ativação.

Para ilustrar, a Figura 1 exibe alguns exemplos de mapas de ativação extraídos de U-Net especializada em segmentação de imagens microscópicas de vasos sanguíneos. A Figura 1 (a) é a imagem original, enquanto as Figuras 1 (b), (c), (d) e (e) representam quatro mapas de ativação produzidos em uma camada intermediária da U-Net.

**Figura 1 – Exemplo de mapas de ativação de uma U-Net.**



**Fonte: Próprio autor**

Uma outra questão que surge é o quão similares e correlacionadas são as representações extraídas de camadas intermediárias da rede. Isto é, se a rede aprende e gera representações que são redundantes, tornando-a mais complexa e custosa computacionalmente do que o necessário para a solução de determinado problema.

Diante da importância da segmentação de vasos sanguíneos em diversas aplicações, torna-se relevante o desenvolvimento de ferramentas que possibilitem a melhoria da eficiência de CNNs utilizadas nesse processo. Para tanto, é necessário extrair as representações da rede, visualizá-las e medir sua correlação ou similaridade, a fim de encontrar mapas de ativação

similares. Tal compreensão possibilita utilizar técnicas para reduzir o número de parâmetros da rede, mas manter o mesmo nível de eficiência da rede original.

Nesse contexto, o presente trabalho tem como motivação estudar meios de extrair as representações aprendidas por CNNs e desenvolver métodos para a visualização e análise de correlação dessas representações por meio de uma ferramenta que permita identificar e medir mapas bem correlacionados, que podem ser podados da rede sem diminuir a sua acurácia. Isso pode levar ao desenvolvimento de arquiteturas mais eficientes e leves, o que possibilita a aplicação em maiores quantidades de dados.

## 1.2 Objetivos

### 1.2.1 Geral

Mensurar e analisar o nível de correlação dos mapas de ativação de camadas intermediárias de CNNs treinadas para a segmentação de vasos sanguíneos.

### 1.2.2 Específicos

Os objetivos específicos do trabalho são:

1. Desenvolver um conjunto de ferramentas que permitam extrair, visualizar e comparar os mapas de ativação gerados por uma CNN.
2. Desenvolver metodologias para a quantificação sistemática da correlação dos mapas de ativação.
3. Verificar se mapas de ativação intermediários de CNNs possuem propriedades distintas dependendo da profundidade da rede.

## 1.3 Organização

Este trabalho está organizado da seguinte forma:

- Capítulo 2: são apresentados a fundamentação teórica e conceitos necessários para compreensão dos assuntos mencionados no decorrer do trabalho, incluindo: (i) aprendizado profundo e redes neurais convolucionais, (ii) Segmentação de imagens utilizando CNNs e (iii) métricas para avaliação da qualidade das segmentações e medidas de similaridade.
- Capítulo 3: neste capítulo é realizada uma revisão bibliográfica de trabalhos relacionados, incluindo: (i) Interpretação de mapas de ativação e (ii) Similaridade entre mapas de ativação de CNNs.

- Capítulo 4: o capítulo descreve e estabelece a metodologia e técnicas utilizadas no trabalho para alcançar os objetivos propostos.
- Capítulo 5: são descritos os experimentos e análises dos resultados.
- Capítulo 6: apresenta as considerações finais e propostas de trabalhos futuros.

# Capítulo 2

## FUNDAMENTAÇÃO TEÓRICA

---

---

### 2.1 Aprendizado Profundo e Redes Neurais Convolucionais

O Aprendizado Profundo, comumente conhecido com *Deep Learning*, é uma subárea da inteligência artificial, especificamente do escopo de Aprendizado de Máquina, o qual tem atraído crescente interesse devido às diversas aplicações, como Visão computacional, Processamento de Linguagem Natural, sistemas de recomendação e, mais recentemente, modelos generativos capazes de gerar imagens, textos complexos e áudio.

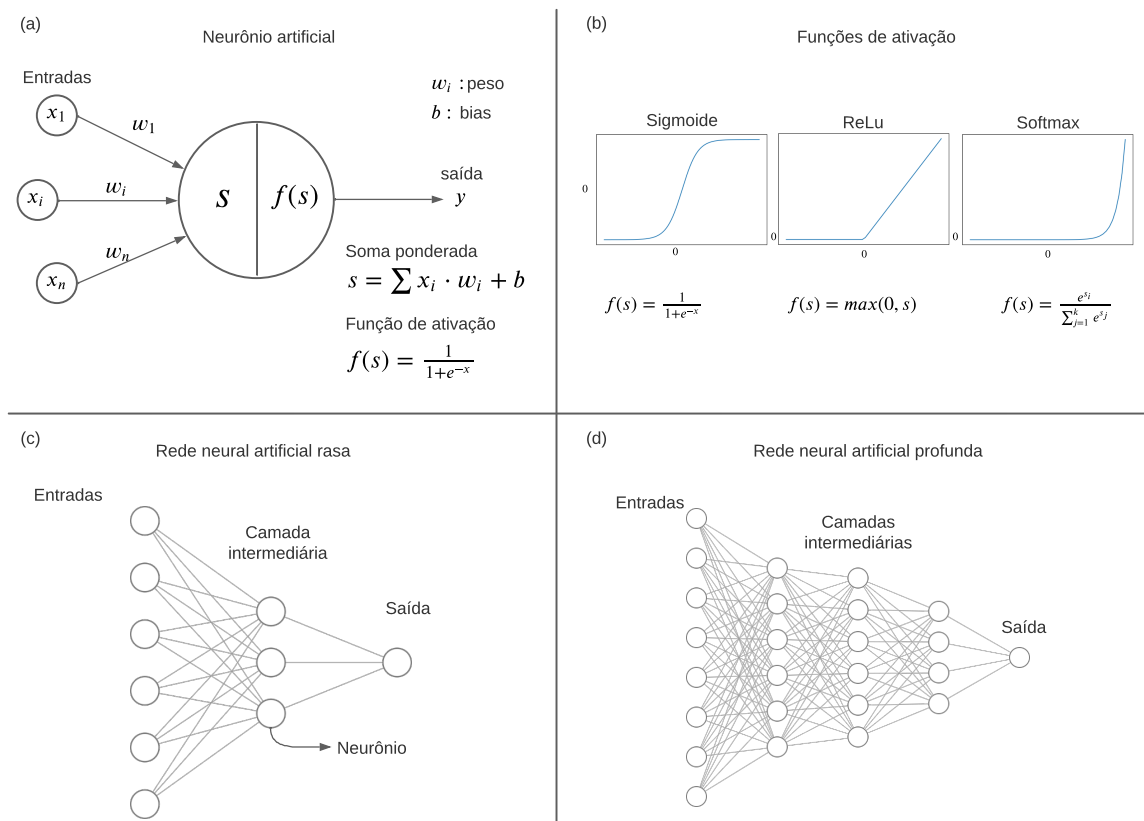
Fundamentalmente, o Aprendizado Profundo baseia a sua estruturação em Redes Neurais Artificiais (ANN), que são estruturas de mapeamento (LEK; PARK, 2008) bioinspiradas. Essas redes têm como unidade computacional não linear o neurônio. Como mostrado na Figura 2 (a), o neurônio é uma função que recebe 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 *forward-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 computacional.

A tarefa de aprendizado pode acontecer de forma supervisionada e 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

Figura 2 – Conceitos básicos de Redes Neurais



Fonte: Adaptado de (DONG et al., 2021)

reajuste dos pesos a cada passo iterativo, uma vez que se conheçam as saídas calculadas e desejadas (RUMELHART et al., 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*.

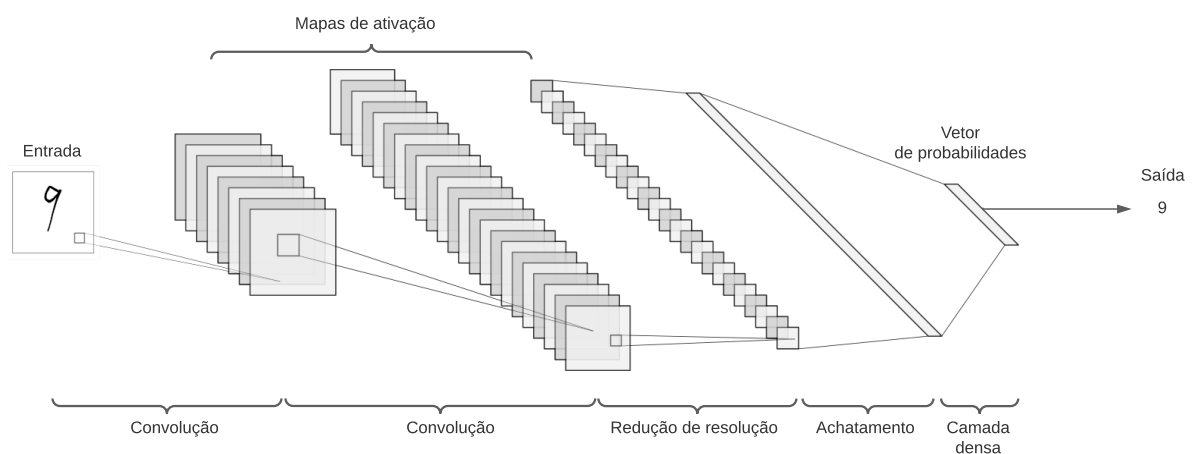
Portanto, o método *backpropagation* calcula retroativamente - da camada de saída em direção à camada de entrada - o sinal de erro para ajustar os parâmetros da rede, como pesos e *bias*, por meio do cálculo do gradiente de cada neurônio em relação a sua entrada (LECUN et al., 2015). Essa técnica é uma forma de derivação automática (GOODFELLOW et al., 2016) que permite que a rede neural seja capaz de produzir os mapas de características a partir dos parâmetros reajustados.

As redes neurais profundas são conhecidas por serem eficientes extratoras de características e padrões durante o processo de treinamento. Essa propriedade foi explorada pelas CNNs, amplamente utilizadas para detecção automática de objetos, classificação e imagens, entre outras aplicações. Trabalhos como os de (FUKUSHIMA, 1980) e (LECUN et al., 1989a) 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 1. 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.

A Figura 3 exemplifica uma CNN hipotética que realiza classificação de dígitos manuscritos, dada uma imagem de entrada. Além das camadas de convolução, há as operações de redução de dimensão, também conhecida como *pooling*, que fazem 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 3 – Exemplo de uma CNN para classificação de dígitos manuscritos**



**Fonte: Próprio autor**

Há também a camada densa, geralmente utilizada em tarefas de classificação, que é localizada no fim da rede. Esta camada é alimentada com uma matriz densa proveniente dos mapas da camada anterior. A camada densa é uma rede neural totalmente conectada que gera o vetor de probabilidades utilizado na classificação final da imagem de entrada.

## 2.2 Segmentação de imagens utilizando CNNs

A segmentação de imagens é uma tarefa importante em muitas aplicações, como reconhecimento de objetos, diagnósticos biomédicos, carros autônomos, entre outros. O avanço das CNNs permitiu que os modelos de aprendizado profundo pudessem realizar a segmentação semântica das imagens, isto é, identificar e separar os diferentes objetos presentes em uma ima-

gem de forma semântica, contextualizada pelas suas classes ou categorias. Em outras palavras, as redes de segmentação são capazes de atribuir a cada pixel da imagem a classe a qual este pertence.

Uma das principais arquiteturas é do tipo Rede Totalmente Convolutiva, do inglês *Fully Convolutional Network* (FCN) (LONG et al., 2015), que substitui as camadas totalmente conectadas das CNNs por camadas convolucionais, o que permite a entrada de imagens com diferentes tamanhos e a geração de segmentação com a mesma resolução da imagem de entrada.

Tradicionalmente, as CNNs exigem grandes conjuntos de dados para o treinamento, o que pode ser um desafio na prática, uma vez que a anotação manual é um processo caro e demorado. Portanto, há uma crescente demanda por técnicas de treinamento eficientes em relação ao número de imagens anotadas disponíveis. Nesse contexto, a arquitetura de rede neural U-Net (RONNEBERGER et al., 2015) tem se destacado como uma opção viável para o treinamento com poucas imagens, produzindo bons resultados em tarefas de segmentação de imagens.

A arquitetura U-Net é comumente aplicada na segmentação de imagens biomédicas, e é uma rede neural totalmente convolutiva 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 4.

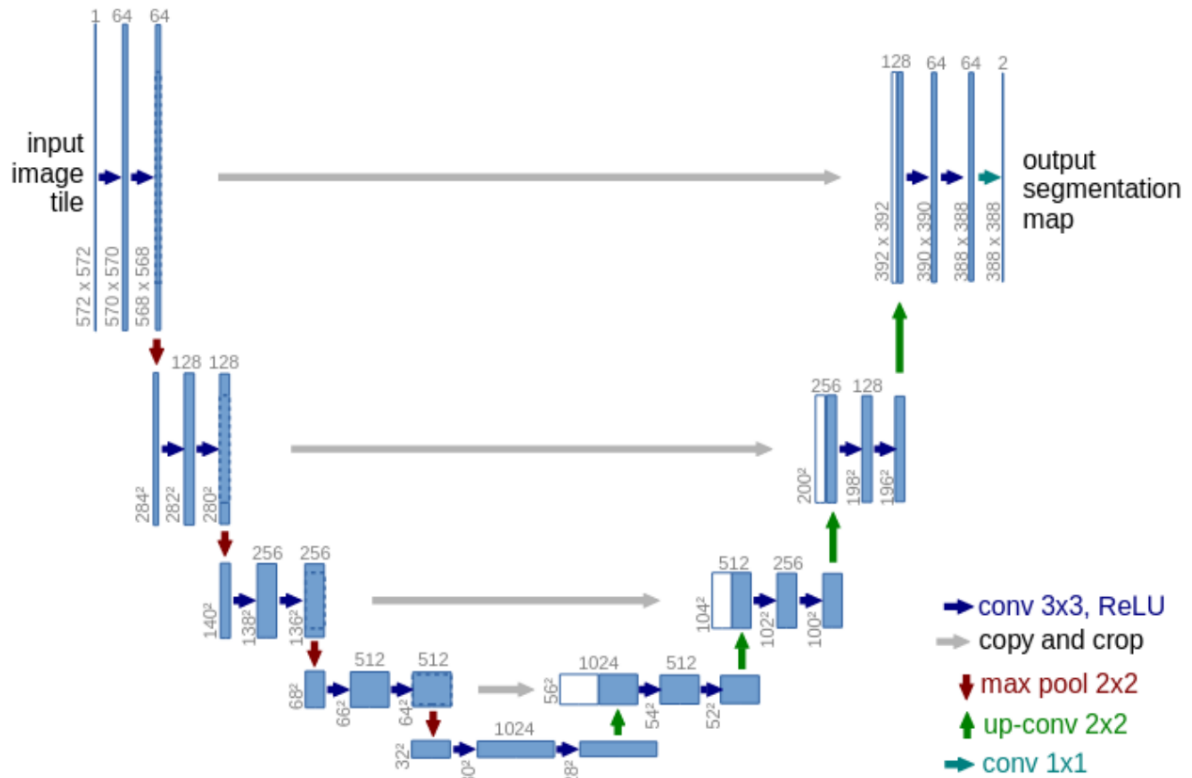
A etapa de codificação na arquitetura U-Net envolve operações de convolução sucessivas, gerando mapas de ativação que contêm atributos informativos das imagens. A resolução dos mapas de ativação é reduzida pela metade por meio de operações de *max pooling* ou convoluções com passo 2, com duplicação do 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. Sendo assim, a arquitetura U-Net é amplamente utilizada na área de segmentação de imagens biomédicas.

A técnica de *data augmentation* é comumente utilizada para compensar a falta de dados anotados disponíveis para treinamento de modelos de aprendizado profundo, incluindo as CNNs. Essa técnica consiste em gerar novas imagens a partir das imagens originais através da aplicação de transformações como variação de brilho, rotação, reflexão e adição de ruído. A utilização de *data augmentation* tem se mostrado eficaz na melhoria da generalização dos modelos de CNNs e na redução do sobreajuste, além de permitir um uso mais eficiente dos dados anotados disponíveis.



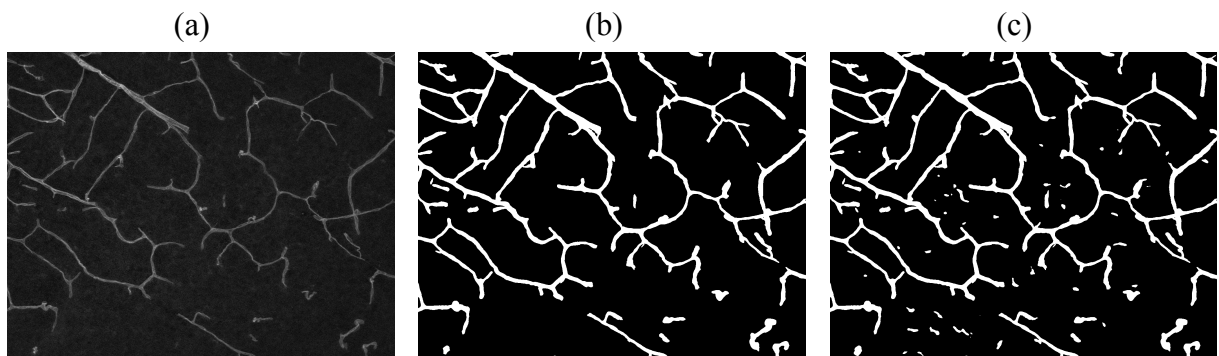
Figura 4 – Arquitetura da U-Net



Fonte: (RONNEBERGER et al., 2015)

A Figura 5 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 5 (a), sendo que o resultado esperado da segmentação é mostrado na Figura 5 (b). O resultado esperado é chamado de padrão ouro, ou imagem referência. Os pixels do padrão ouro podem possuir apenas dois valores, 0 (fundo) ou 1 (vaso sanguíneo). A Figura 5 (c) mostra um exemplo de segmentação realizado por uma rede U-Net.

Figura 5 – Exemplo de segmentação resultante da aplicação de uma rede U-Net a uma imagem de microscopia de vasos sanguíneos



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.

## 2.2.1 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.

### 2.2.1.1 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.

$$prec = \frac{VP}{VP + FP} \quad (2.1)$$

### 2.2.1.2 Sensibilidade

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

$$recall = \frac{VP}{VP + FN} \quad (2.2)$$

### 2.2.1.3 F1-Score

Por último, F1-Score é a média harmônica entre precisão e recall.

$$F1 = 2 * \frac{(prec * recall)}{(pre + recall)} \quad (2.3)$$

#### 2.2.1.4 Medidas de similaridade

##### 2.2.1.4.1 Coeficiente Dice

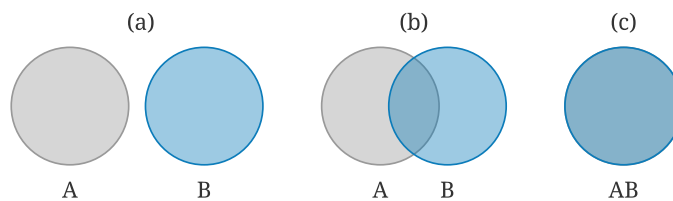
O Coeficiente de Similaridade Dice (DICE) é um índice de sobreposição espacial entre duas segmentações (ZOU et al., 2004), também definido como uma medida de concordância entre as regiões de cada classe (GIGANDET et al., 2004). Esse índice permite quantificar a qualidade da segmentação proporcionada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

Seja a 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

$$DSC = \frac{2(A \cap B)}{A + B}. \quad (2.4)$$

$(A \cap B)$  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 6 ilustra o domínio de valores do coeficiente DICE, sendo que na Figura 6 (a) não há sobreposição dos pixels, então  $DICE=0$ . Quando ocorre sobreposição parcial na Figura 6(b),  $0 < DICE < 1$ . Finalmente, no cenário de completa sobreposição na Figura 6(c),  $DICE=1$ .

**Figura 6 – Ilustração da sobreposição de dois conjuntos de segmentação e respectivos valores do coeficiente DICE.**



**Fonte: Próprio autor**

Reescrevendo a Equação 2.4 em termos das definições de VP, FN e FP:

$$DSC = \frac{2VP}{2VP + FN + FP} \quad (2.5)$$

##### 2.2.1.4.2 Coeficiente IoU

O coeficiente de intersecção sobre união (IoU), ou índice de Jaccard, pode ser utilizado semelhantemente ao coeficiente Dice. Sua equação é descrita assim:

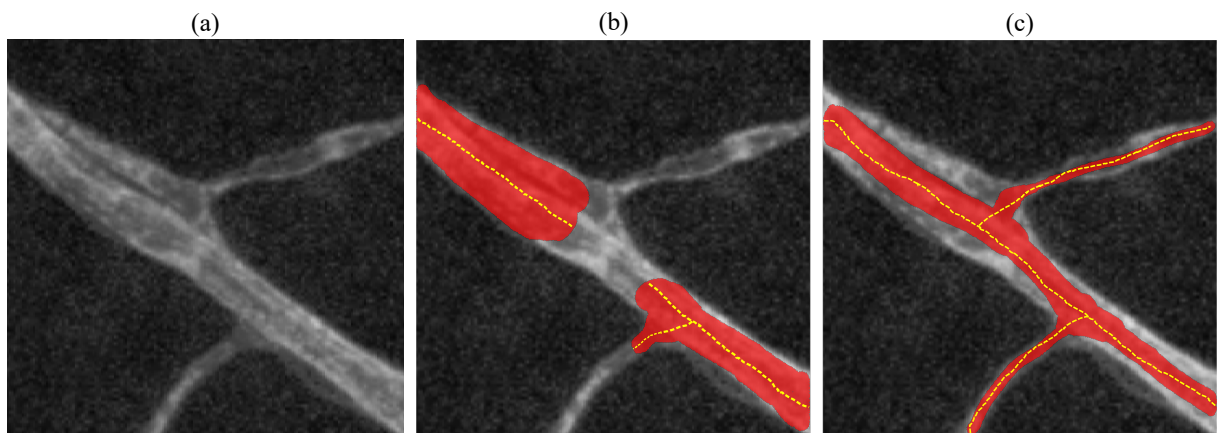
$$IoU = \frac{(A \cap B)}{(A \cup B)} = \frac{VP}{VP + FN + FP}. \quad (2.6)$$

### 2.2.1.4.3 Coeficiente *clDice*

O coeficiente de similaridade DICE é amplamente utilizado para avaliar a eficácia das técnicas de segmentação, mas para segmentações de estruturas tubulares, como vasos sanguíneos, pode não ser adequado, pois não leva em conta a continuidade dessas estruturas. Dado que os vasos sanguíneos são importantes para o transporte de nutrientes e oxigênio, é importante que a imagem segmentada preserve a integridade dessas estruturas, pois a presença de quebra compromete a interpretação da dinâmica do fluxo sanguíneo (SHIT et al., ).

Supondo que o coeficiente DICE seja o mesmo para as segmentações mostradas nas Figuras 7 (b) e (c). A primeira segmentação cobriu quase toda a área do vaso sanguíneo, mas gerou uma descontinuidade. Por outro lado, a segunda segmentação não abrangeu corretamente a área do vaso sanguíneo, ou seja, o raio da estrutura foi subestimado. Contudo, manteve sua propriedade de continuidade.

**Figura 7 – Exemplo de comparação da qualidade de duas segmentações de um vaso sanguíneo.**



**Fonte: Próprio autor**

A medida *centerlineDice* (*clDice*) foi proposta como um coeficiente de similaridade para algoritmos de segmentação de estruturas tubulares, pois leva em consideração a interseção entre as linhas mediais das estruturas e a segmentação das mesmas, facilitando a preservação da topologia de sistemas compostos por essas estruturas. Pode ser utilizado como uma função de perda, dado que a otimização desse índice é importante para garantir a precisão e a confiabilidade dos resultados obtidos.

Para realizar o cálculo da medida, consideram-se duas máscaras binárias. Sendo a máscara binária padrão ouro ( $V_L$ ) e a máscara binária de segmentação predita ( $V_P$ ). De ambas as máscaras, são extraídas respectivamente as linhas mediais (TAGLIASACCHI et al., 2016)

$S_L$  e  $S_P$ , o que possibilita calcular as estatísticas de Sensibilidade Topológica  $Tsens(S_L, V_P)$  e Precisão Topológica  $Tprec(S_P, V_L)$ .

A Sensibilidade Topológica é definida pela divisão da interseção entre  $S_L$  e  $V_P$  por  $S_L$ :

$$Tsens(S_L, V_P) = \frac{|S_L \cap V_P|}{|S_L|}$$

Basicamente, a Sensibilidade Topológica quantifica se o resultado da segmentação tende a possuir os píxeis das linhas mediais dos vasos sanguíneos indicados no padrão ouro. Essa medida possui paralelo com a medida de revocação, que quantifica a fração de verdadeiros positivos em relação ao número total de itens positivos.

De forma similar, a Precisão Topológica é dada por

$$Tprec(S_P, V_L) = \frac{|S_P \cap V_L|}{|S_P|}$$

A Precisão Topológica possui paralelo com a medida de precisão, que quantifica o número de verdadeiros positivos em relação ao número de itens positivos identificados pelo algoritmo.

Com isso, o coeficiente de similaridade  $clDice$  é dado pela média harmônica das duas estatísticas anteriores, ou seja:

$$clDice(V_P, V_L) = 2 \times \frac{Tprec(S_P, V_L) \times Tsens(S_L, V_P)}{Tprec(S_P, V_L) + Tsens(S_L, V_P)}$$

### 2.2.1.5 Medidas de distância e correlação entre mapas de ativação

Neste trabalho foram utilizadas diferentes medidas para quantificar a similaridade entre mapas de ativação de CNNs. Apresentaremos a seguir as medidas utilizadas. O objetivo das medidas é comparar dois vetores de atributos amostrados, sendo  $x = \{x_1, x_2, \dots, x_n\}$  e  $y = \{y_1, y_2, \dots, y_n\}$ .

#### 2.2.1.5.1 Distância euclidiana

A distância euclidiana é uma medida de proximidade entre dois pontos  $(x_1, y_1)$  e  $(x_2, y_2)$  no espaço bi-dimensional, e é dada por:

$$d = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$$

Generalizando para  $n$  pontos no espaço:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

### 2.2.1.5.2 Similaridade cosseno

Esta medida calcula a similaridade entre dois vetores pelo cosseno do ângulo  $\theta$  entre ambos.

$$\cos(\theta) = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|}$$

De outra forma

$$\cos(\theta) = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}}$$

O índice de similaridade compreende valores de 0 à 1.

- Se  $\theta = 0$ , então  $\cos(0) = 1$ . Os dois vetores são paralelos, portanto, similares.
- Se  $\theta = 90$ , então  $\cos(90) = 0$ . Os dois vetores são ortogonais, portanto, não similares.

### 2.2.1.5.3 Correlação de Pearson

O coeficiente de correlação de Pearson mede a relação linear entre duas variáveis. Esse índice compreende valores de -1 à 1, sendo 1 uma perfeita correlação linear positiva, -1 uma perfeita correlação linear negativa, e 0 indica que não há correlação linear entre as variáveis. Contudo, o valor 0 não quer dizer que não existam outros graus de correlações não lineares entre os valores observados.

$$\text{corr}(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{\text{cov}(x, y)}{\sqrt{\text{var}(x)\text{var}(y)}}$$

# Capítulo 3

## TRABALHOS RELACIONADOS

---

---

Nesta seção serão explorados alguns estudos da literatura relacionados ao tema desenvolvido neste trabalho.

### 3.1 Interpretação de mapas de ativação

A interpretabilidade dos modelos de aprendizado profundo é importante para a compreensão dos processos internos dos modelos. Nesse sentido, a visualização e interpretação dos mapas de ativação de CNNs tem sido objeto de estudo e de desenvolvimento de técnicas e ferramentas.

Em (ERHAN et al., 2009) os autores apresentaram uma técnica de visualização dos mapas de ativação baseada em uma metodologia de otimização que identifica os padrões que maximizam as ativações dos mapas, e que pode ser generalizada a todos os mapas de uma camada intermediária.

Com essa técnica, foi possível reconhecer características estruturais e hierárquicas que são aprendidas pela rede. Os autores afirmam que as redes geram representações mais simples nas primeiras camadas e que se tornam mais complexas nas camadas mais profundas.

Essa constatação também foi verificada nos trabalhos de (YOSINSKI et al., 2015) e (OLAH et al., 2017), focados em criar ferramentas de visualização de representações das redes de aprendizado profundo baseados em otimização. Os mapas extraídos pelas ferramentas indicam que a rede gera representações a nível hierárquico ao longo da sua profundidade, partindo da camada inicial com reconhecimento de contornos, seguindo com aprendizado das texturas, depois reconhecimento de padrões, e então detecção de características mais complexas, como partes dos objetos e os objetos em si.

O trabalho de (ZEILER; FERGUS, 2014) oferece uma importante contribuição para o campo de CNNs, pois apresenta uma série de técnicas para visualização e interpretação de como ocorre o processamento e extração de informações de imagens. Os autores propuseram a utilização de uma *Deconvolutional Network* (deconvnet) que utiliza os mesmos componentes de uma CNN, como camadas de filtragem e *pooling*, mas em ordem reversa, possibilitando o

mapeamentos das ativações, ou características extraídas, de volta aos pixels de entrada. Com isso, é possível identificar quais píxeis de uma dada imagem foram mais importantes para a classificação. A deconvnet é anexada a cada camada da rede convolucional.

Para realizar os experimentos, foram utilizadas CNNs como as definidas por (LECUN et al., 1989b) e (KRIZHEVSKY et al., 2017), e treinadas para classificar imagens do conjunto de dados ImageNet (DENG et al., 2009). Os resultados obtidos e apresentados mostram que essa abordagem de visualização pode ser utilizada com o intuito de melhor entendimento do funcionamento de CNNs, o que contribui para tarefas de depuração e melhorias na rede. Além disso, a técnica pode ser aplicada a outras CNNs.

Já no artigo de (WANG et al., 2020), é apresentada uma ferramenta interativa para auxiliar a compreender o funcionamento de CNNs. Como motivação, os autores explicaram que o processo de interpretação do treinamento de CNNs pode ser complexo, mas a ferramenta chamada de “*CNN Explainer*”<sup>1</sup> permite aos usuários interagir com uma rede neural previamente treinada, a fim de visualizar a nível de neurônio como as representações aprendidas pela rede são processadas.

A ferramenta oferece uma gama de visualizações dinâmicas e interativas, tais como a exibição de camadas individuais, análise de características aprendidas, bem como dos termos das equações de ativação. A aplicação pode ser executada via navegador de internet e, segundo os autores, é um instrumento que pode auxiliar usuários a entender como as CNNs funcionam ao longo das camadas intermediárias e serve até mesmo aos leigos no assunto de aprendizado profundo.

Os trabalhos revisados corroboraram com a hipótese do aprendizado hierárquico das representações, isto é, a medida que a rede se aprofunda, esta é capaz de identificar características mais abstratas, tornando-se menos sensível às variações e mais focadas nas características mais importantes dos dados de entrada.

## 3.2 Similaridade entre mapas de ativação de CNNs

Uma das questões fundamentais em relação à pesquisa e desenvolvimento de redes neurais profundas é a necessidade de melhor compreensão de como as representações aprendidas por essas redes variam com diferentes arquiteturas. Isso gera um segundo questionamento de se essas representações, situadas em diferentes níveis de profundidade de uma mesma rede, são similares entre si.

No trabalho intitulado “*Similarity of Neural Network Representations Revisited*” de (KORNBLITH et al., 2019), os autores examinaram de forma sistemática como as representações, ou mapas de ativação, podem ser similares entre camadas da mesma rede, entre camadas de

<sup>1</sup> Website CNN Explainer: <https://poloclub.github.io/cnn-explainer/>



redes com diferentes arquiteturas, e também comparando os mapas de uma mesma rede, mas treinada com diferentes conjuntos de dados. Para medir o nível de similaridade entre mapas de ativação, foi proposta a medida *Centered Kernel Alignment* (CKA), que foi comparada com outros índices de similaridade, como regressão linear e *Canonical Correlation Analysis* (CCA). Nesse quesito, os autores destacam as implicações da escolha de uma métrica adequada para medida de similaridade entre representações de redes neurais, pois ela pode afetar os resultados da análise.

De acordo com o estudo, a medida CKA se mostrou promissora em relação às outras, pois é capaz de determinar correspondências entre as camadas intermediárias de redes neurais treinadas a partir inicializações randômicas. Também foi demonstrado que a acurácia das redes cresce a medida que se tornam mais profundas, mas chegam a um platô de acurácia máxima alcançada. Ao realizar o cálculo de similaridade de mapas de ativação na rede de maior profundidade que alcançou o platô de ganho de acurácia, verificou-se que mais da metade das representações são muito similares com os mapas da última camada.

Além disso, o trabalho verificou que diferentes redes neurais são mais semelhantes com outras redes com as mesmas larguras, ou quantidade de representações por camada. Por fim, o experimento resultou que uma mesma rede treinada com diferentes conjunto de dados, CIFAR-10 e CIFAR-100 (KRIZHEVSKY et al., 2009), é capaz de desenvolver representações similares.

Em um outro trabalho relacionado, intitulado "*Do wide and deep networks learn the same things? Uncovering how neural network representations vary with width and depth*" de (LEE et al., 2020), os autores investigaram a hipótese de que redes neurais com diferentes arquiteturas, variando a largura ou profundidade, são capazes de aprender as mesmas representações. Para investigar essa hipótese, os autores implementaram vários experimentos para comparar as representações das redes, em profundidade e largura. Os conjuntos de dados utilizados para treinamento foram as bases CIFAR-10, CIFAR-100, além da ImageNet. A arquitetura escolhida foi a ResNet (HE et al., 2016), uma CNN com blocos residuais. O método de medição de similaridade entre mapas de ativação foi baseado no índice CKA (KORNBLITH et al., 2019) pois, de acordo com os autores, é uma medida confiável para avaliar as correspondências entre ativações das camadas intermediárias.

Nos resultados dos experimentos de cálculo de similaridade dos mapas de ativação entre camadas consecutivas, verificou-se que a medida que as redes neurais se tornam mais largas ou profundas, as representações extraídas nas camadas intermediárias se comportam como blocos de mapas fortemente similares. Isso fica mais evidente nas últimas camadas da rede. Em concordância com o trabalho de (KORNBLITH et al., 2019), os autores verificaram que modelos menores não apresentam a estrutura em bloco de mapas bem correlacionados, mas os maiores e mais profundo sim, mesmo com inicializações aleatórias.

Por fim, dado que exista redundância de representações nas camadas intermediárias da rede, isso possibilita reduzir o custo computacional da rede através da redução de parâmetros,

mas mantendo o desempenho de acurácia aceitável. Essa proposta foi testada, e os autores aplicaram a técnica de poda de estrutura (LI et al., 2016) dos blocos com representações similares no fim de cada estágio residual das redes. Não houve impacto significativo na acurácia do modelo sobre o conjunto de teste. Essa abordagem se mostra promissora na definição de modelos de redes mais compactos.

# Capítulo 4

## METODOLOGIA E DESENVOLVIMENTO

---

---

Este capítulo descreve os dados, ferramentas e processos utilizados no desenvolvimento da pesquisa deste trabalho, de maneira a justificar a metodologia utilizada nos experimentos que visam calcular o nível de correlação entre os mapas de ativação de CNNs. O capítulo está dividido da seguinte forma

1. Natureza da pesquisa
2. Ferramentas utilizadas
3. Conjunto de dados
4. Planejamento das arquiteturas de CNNs
5. Estratégias de comparação de similaridade entre mapas de ativações
6. Desenvolvimento e organização de um pacote para exploração e análise de CNNs
7. Exploração dos mapas de ativação das CNNs

### 4.1 Natureza da pesquisa

O trabalho desenvolvido é de natureza experimental, pois tem o intuito de gerar novos conhecimentos que sejam úteis para a área de interpretação de CNNs e o avanço do estado da arte, além de realizar simulações a fim de obter dados experimentais para serem avaliados durante o ciclo de pesquisa.

### 4.2 Ferramentas

As ferramentas e bibliotecas utilizadas para desenvolvimento e treinamento dos modelos das CNNs, bem como as análises estatísticas dos experimentos são *open-source* e gratuitas. São estas:

- Linguagem de programação: *Python 3.9*
- Framework para desenvolvimento das CNNs: *PyTorch*
- Utilitários para treinamento das CNNs: *TorchTrainer*<sup>1</sup>
- API para processamento dos tensores: *CUDA*
- Computação numérica: *NumPy*
- Análise e manipulação dos dados: *Pandas*
- Plot de gráficos e imagens: *Seaborn* e *Matplotlib*

Dada a necessidade de treinamento de redes neurais com imagens de alta resolução, o computador disponibilizado pelo laboratório *Image and Network Analysis Group* (INAG) do Departamento de Computação (DC) da Universidade Federal de São Carlos (UFSCar) tem estas especificações:

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

## 4.3 Conjunto de dados

A base de dados utilizada durante o projeto é composta por 2800 imagens 2D de microscopia obtidas de amostras de vasos sanguíneos do córtex cerebral de camundongos, possuindo a dimensão  $1376 \times 1104$  pixels. Cada imagem da base possui sua respectiva imagem padrão-ouro ou, do inglês, *ground-truth*, rotulada manualmente por especialistas, ou por meios semi-automáticos. O padrão-ouro corresponde à segmentação esperada. Essa amostra foi coletada pelo Prof. Baptiste Lacoste da Universidade de Ottawa, Canadá. O método de coleta é semelhante ao descrito e publicado por (OUELLETTE et al., 2020).

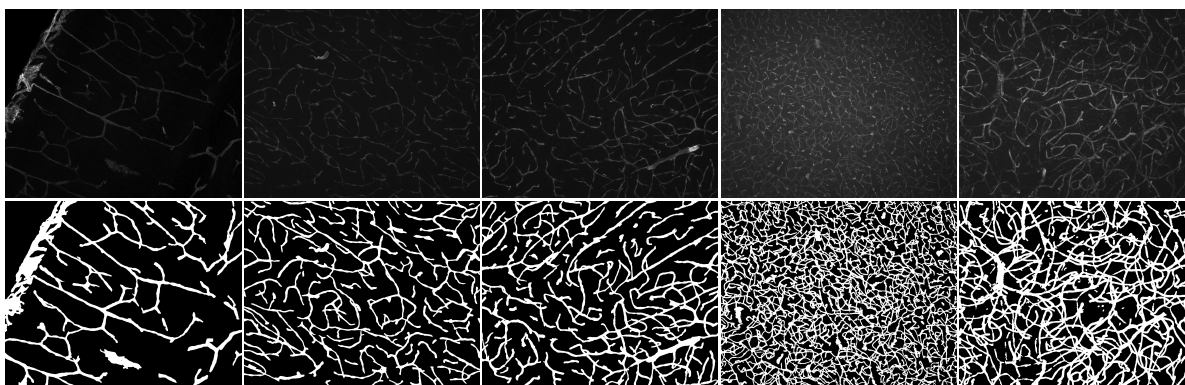
### 4.3.1 Seleção de um subconjunto de imagens

Diante do volume de imagens presentes no conjunto disponível, foi necessária a seleção de uma amostra de imagens para este trabalho. A seleção se deu de forma manual, seguindo o critério de variabilidade de imagens, isto é, a consideração da nitidez e qualidade do imageamento dos vasos sanguíneos, a escala e quantidade de vasos sanguíneos presentes na imagem, além

<sup>1</sup> Pacote de utilitários e funções para treinar CNNs, desenvolvido pelo professor e orientador Dr. Cesar Henrique Comin e disponibilizado no repositório <https://github.com/chcomin/torchtrainer>

da variação de posicionamento e direção dos vasos. A Figura 8 exemplifica a diversidade das imagens selecionadas e suas respectivas segmentações padrão-ouro. Ao todo, a amostra contou com 296 imagens que serão utilizadas para treino e validação dos modelos. Além disso, um outro conjunto de 20 imagens foi selecionado para servir apenas como conjunto de teste, seguindo os mesmos critérios.

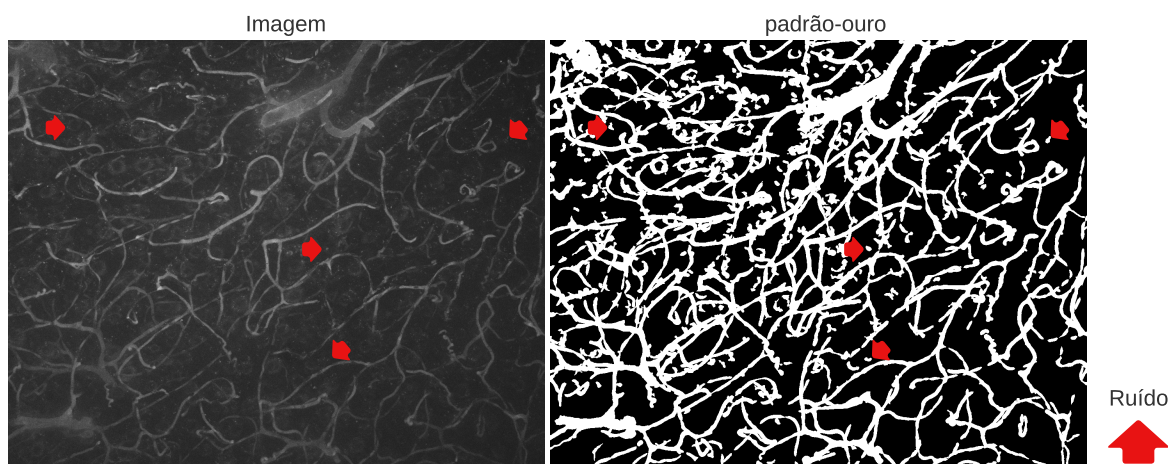
**Figura 8 – Exemplos de imagens e seus respectivos padrão-ouro amostrados.**



**Fonte: Próprio autor**

Além disso, um critério importante adotado foi a qualidade da segmentação do padrão-ouro. Algumas imagens apresentaram aparentes ruídos em regiões de fundo, e esses pixels foram rotulados como vasos sanguíneos no padrão ouro, devido ao processo semi-automático de anotação. Isso está demonstrado na Figura 9, onde há o indicativo de alguns exemplos de ruídos que foram anotados incorretamente.

**Figura 9 – Imagem ruidosa**



**Fonte: Próprio autor**

Essas imagens de baixa qualidade foram descartadas, pois poderiam afetar o aprendizado

das redes, uma vez que as regiões de ruídos teriam chances de sofrer ativação e então serem segmentadas erroneamente, influenciando negativamente na avaliação do desempenho das redes.

## 4.4 Planejamento das arquiteturas de CNNs

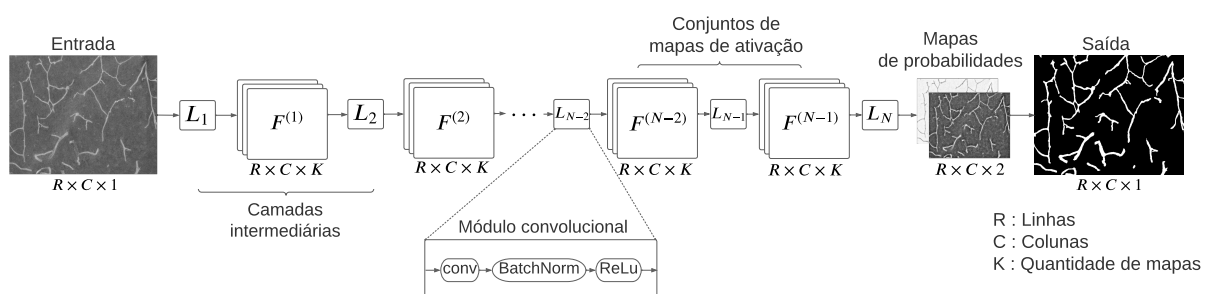
Este trabalho propõe fazer a comparação entre mapas de ativações presentes nas camadas intermediárias de diferentes CNNs e diferentes profundidades de camadas de convolução. Assim sendo, um passo importante é a definição das arquiteturas que serão utilizadas na concepção das CNNs.

Em análises preliminares, a rede U-Net (RONNEBERGER et al., 2015), considerada uma rede eficiente para segmentação de imagens biomédica, foi utilizada. Contudo, devido à existência das etapas de codificação e decodificação, que reduz e amplia a resolução dos mapas de ativação, escolheu-se padronizar os experimentos com uma arquitetura mais simples de se reproduzir.

Portanto, as redes desenvolvidas no projeto são do tipo FCN (LONG et al., 2015) e não possuem a etapa de redução de resolução. Ou seja, tanto os mapas de ativação quanto a imagem segmentada de saída têm a mesma resolução da imagem original. O intuito dessa abordagem é preservar as informações espaciais dos vasos sanguíneos ao longo da rede, sem que haja a necessidade de interpolação dos mapas para recuperar a resolução original. A Figura 10 ilustra a arquitetura escolhida para este trabalho.

Seja uma rede CNN dotada de  $N$  camadas de convolução intermediárias e denotadas por  $L = \{L_1, L_2, \dots, L_N\}$ . Cada camada  $L_i$  gera um conjunto de mapas de ativação  $F^i = \{F_1^{(i)}, F_2^{(i)}, \dots, F_k^{(i)}\}$ , que por sua vez contém as representações das características extraídas por meio das operações de convolução e ativação na camada  $L_i$  sobre o conjunto de mapas  $F_k^{(i-1)}$ , sendo  $k$  a quantidade de mapas de ativação e  $i$  o índice da camada  $L_i$  à qual o conjunto de mapas pertence.

**Figura 10 – Representação de uma CNN para segmentação de imagens composta por  $N$  camadas intermediárias.**



**Fonte: Próprio autor**

Em relação às camadas intermediárias, a rede gera 64 mapas de ativação a cada camada

de convolução até a camada  $L_{N-1}$ , tornando a largura da rede constante. Apenas a última camada  $L_N$  tem como saída dois mapas de ativação que representam, respectivamente, a probabilidade dos pixels serem fundo, portanto classe 0, e a probabilidade de ser vaso sanguíneo, classe 1.

Sabendo disso, considere a nomenclatura “módulos convolucionais” para as camadas de  $L_1$  à  $L_{N-1}$ , pois cada módulo convolucional é composto de: (i) uma operação de convolução, tendo os filtros de convolução de tamanho  $3 \times 3$ ; (ii) uma operação regularizadora *Batch Normalization*; (iii) uma função de ativação ReLu, que gera novos mapas para serem processados na camada adjacente.

Para os experimentos, foram consideradas 3 redes com profundidades distintas, contendo respectivamente 1, 6 e 10 módulos convolucionais.

## 4.5 Estratégias de comparação de similaridade entre mapas de ativações

Uma vez que as redes estejam treinadas, os conjuntos de mapas de ativações podem ser extraídos para que sejam feitos cálculos de similaridade entre eles. Contudo, é importante traçar algumas estratégias, como a forma de obtenção desses mapas, a escolha da medida de similaridade a ser utilizada no processo, e quais camadas serão consideradas no processo.

### 4.5.1 Extração dos mapas de ativação

Como visto na seção 3.1, alguns trabalhos desenvolveram métodos para obtenção das ativações de camadas intermediárias. Neste trabalho, o método escolhido está implementado na biblioteca *PyTorch*, e é chamado de método de gancho (*hook*). Mais especificamente `torch.nn.modules.module.register_module_forward_hook(hook)`.

O *hook* é uma função que é chamada quando uma propagação direta da rede atinge a camada em que o *hook* está registrado. Logo, o *PyTorch hooks* se torna uma ferramenta útil para extrair os mapas de ativação, pois permite acessar as saídas intermediárias quando a camada que se deseja observar é ativada. Isso permite adicionar funções personalizadas para serem executadas com o propósito de capturar e salvar as saídas intermediárias da camada para posterior análise.

### 4.5.2 Escolha da medida de similaridade

A escolha de uma medida de similaridade adequada é fundamental para garantir que as comparações entre os mapas de ativação sejam significativas, entretanto, não é uma tarefa simples, visto que não existe uma medida universal para aferir redundância de aprendizado em CNNs.

No trabalho de Iniciação Científica (IC) “Interpretação e Refinamento de Redes Neurais Convolucionais para a Segmentação de Vasos Sanguíneos”<sup>2</sup> realizou-se testes de similaridade entre mapas com as medidas de correlação de Pearson, similaridade cosseno e distância euclidiana. A correlação de Pearson mostrou resultados promissores nos experimentos, quando considerados os valores numéricos da correlação linear absoluta e as similaridades visuais entre os pares de mapas comparados, sendo então essa a medida escolhida e utilizada na IC e neste trabalho.

Esse indicador foi capaz de retornar mapas bem correlacionados que tivessem fortes correspondentes lineares. Contudo, a correlação de Pearson pode ser deficiente em mensurar relações não lineares entre os mapas que possuam um bom grau de similaridade, de acordo com outras medidas. Ademais, a métrica CKA utilizada nos trabalhos citados na seção 3.2 não foi empregada neste projeto, entretanto, é considerada um indicador elegível para ser implementado na atualização da biblioteca que será tratada na seção 4.6.

### 4.5.3 Cálculo de similaridade entre mapas de ativação

Como forma de verificar o nível de correlação que os mapas de uma mesma CNN e de CNNs distintas têm entre si, seguiu-se as seguintes estratégias:

Seja  $A$  uma CNN possuindo  $N - 1$  módulos convolucionais. Sejam também seus conjuntos de mapas de ativação extraídos dos módulos convolucionais  $\{F_a^{(1)}, F_a^{(2)}, \dots, F_a^{(N-1)}\}$ , onde  $F_a^{(i)}$  é o conjunto de todos os mapas gerados pelo  $i$ -ésimo módulo. Par-a-par, é computada a correlação entre os mapas contidos nos conjuntos, sendo permitido o cálculo de correlação entre mapas de um mesmo conjunto. Assim, a quantidade de combinações de correlações possíveis é dada pelo arranjo com repetições  $A = (N - 1)^2$ .

De forma semelhante, para calcular a correlação entre dois modelos,  $A$  e  $B$ , possuindo respectivamente  $N - 1$  e  $M - 1$  módulos convolucionais, considere os seus conjuntos de mapas de ativação sendo respectivamente:  $\{F_a^{(1)}, F_a^{(2)}, \dots, F_a^{(N-1)}\}$  e  $\{F_b^{(1)}, F_b^{(2)}, \dots, F_b^{(M-1)}\}$ . Logo, o cálculo de correlação gera uma quantidade de combinações possíveis  $A = (N - 1) \times (M - 1)$ .

Devido ao fato de que as CNNs podem produzir centenas de mapas de ativação, e a análise de correlação aqui abordada tem complexidade quadrática, decidiu-se aplicar o cálculo apenas nas camadas pares dos módulos convolucionais das redes analisadas, incluindo o primeiro e o último módulo, respectivamente  $L_1$  e  $L_{N-1}$ . Por exemplo, se a rede tem 7 camadas de convolução, mas 6 módulos convolucionais, as camadas consideradas na análise de correlação são:  $L_1, L_2, L_4, L_6$ .

<sup>2</sup> Iniciação Científica produzida pelo autor deste Trabalho de Conclusão de Curso (TCC) e também sob a orientação do mesmo professor orientador deste TCC.



## 4.6 Desenvolvimento e organização de um pacote para exploração e análise de CNN

As CNNs são compostas por várias camadas que realizam operações complexas nas imagens de entrada, produzindo mapas de ativação como saídas que representam as características extraídas pela rede. Essas características, em diferentes níveis de abstração, são importantes para entender como a rede processa as informações da imagem e efetua a tarefa de segmentação.

Contudo, esse modelos podem gerar centenas ou até milhares de mapas, o que torna a análise manual dessas representações impraticável. Além do mais, computacionalmente pode ser um desafio processar esses mapas e organizar os resultados de forma intuitiva.

Logo, é importante que haja ferramentas que automatizem o processo para obtenção, análise e visualização de mapas de ativação. Sendo assim, esses instrumentos de análise possibilitam a identificação de padrões nos mapas de ativação. Podem também ajudar em descobertas sobre como as redes operam, o entendimento das características aprendidas, o nível de discriminabilidade nas ativações e - o objeto de pesquisa deste trabalho - o grau de correlação entre diferentes mapas da rede.

A fim de alcançar esses objetivos, desenvolveu-se um pacote em *Python* intitulado "*CNN Explorer*" (GALVAO, 2023)<sup>3</sup>, capaz de extrair os mapas de ativação de um modelo de CNN pré-treinado e também os seus filtros, além de oferecer a visualização dos dados extraídos. A biblioteca permite também o cálculo de correlação entre mapas de diferentes camadas e modelos, produzindo no fim do cômputo estatísticas valiosas para análise descritiva e visual das correlações.

### 4.6.1 Arquitetura da biblioteca

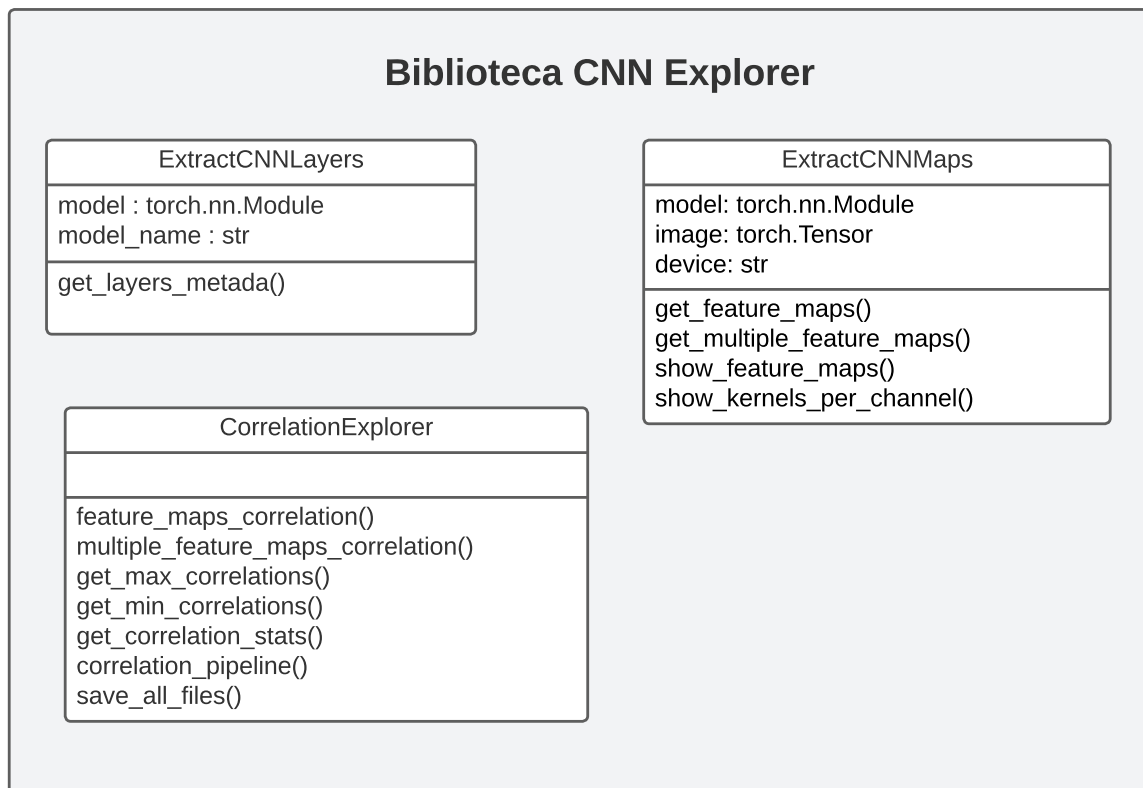
A *CNN Explorer* é formada por três classes responsáveis por gerar metadados das camadas, extrair e exibir os mapas de características e filtros das camadas que se deseja observar e, por último, uma classe responsável por computar a correlação entre os mapas de ativação de uma ou duas CNNs e gerar estatísticas a partir dos resultados. A Figura 11 exibe uma representação simplificada do diagrama de classes da biblioteca, com os principais atributos e métodos. A seguir será apresentada uma breve explicação do propósito e funcionamento de cada classe.

#### 4.6.1.1 Classe `ExtractCNNLayers`

Esta classe é responsável por extrair as informações sobre as camadas de uma CNN *PyTorch*. O objetivo é receber o modelo da rede neural como entrada e extrair informações sobre a estrutura das camadas e o número de mapas de características de cada camada.

<sup>3</sup> Repositório do pacote *CNN Explorer*: [https://github.com/wesleygalvao/cnn\\_explorer](https://github.com/wesleygalvao/cnn_explorer)

**Figura 11 – Representação simplificada da biblioteca CNN Explorer. Principais atributos e métodos**



**Fonte: Próprio autor**

O método `get_layers` dessa classe recebe uma lista de caminhos para as camadas desejadas pertencentes ao modelo e retorna um dicionário de listas contendo informações (metadados) sobre cada camada, incluindo seu caminho como submódulo do modelo, número de mapas de ativações gerados pela camada e a estrutura da camada.

#### 4.6.1.2 Classe `ExtractCNNMaps`

Classe que contém métodos para extração dos mapas de ativação e filtros de uma CNN, assim como métodos para visualização desses dados. Aqui serão tratados apenas métodos referentes aos mapas de ativação.

O construtor da classe recebe como parâmetros um modelo CNN, uma imagem ou um conjunto de imagens, e o dispositivo onde o modelo deve ser executado, podendo ser na GPU (Unidade Gráfica de Processamento) ou CPU (Unidade Central de Processamento).

A função `get_feature_maps` é responsável por obter os mapas de uma determinada camada da rede, dada uma imagem de entrada, utilizando um *hook* do *PyTorch*. O método recebe como parâmetros o índice da imagem e a camada desejada, e então retorna os mapas de ativação capturados no *hook*.

Já o método `get_multiple_feature_maps` é semelhante ao método anterior, mas permite a extração de mapas de características para várias camadas da rede. Portanto, este recebe uma lista com as camadas desejadas. Após o processamento, retorna uma lista com os mapas de ativação extraídos, respectivos a cada camada.

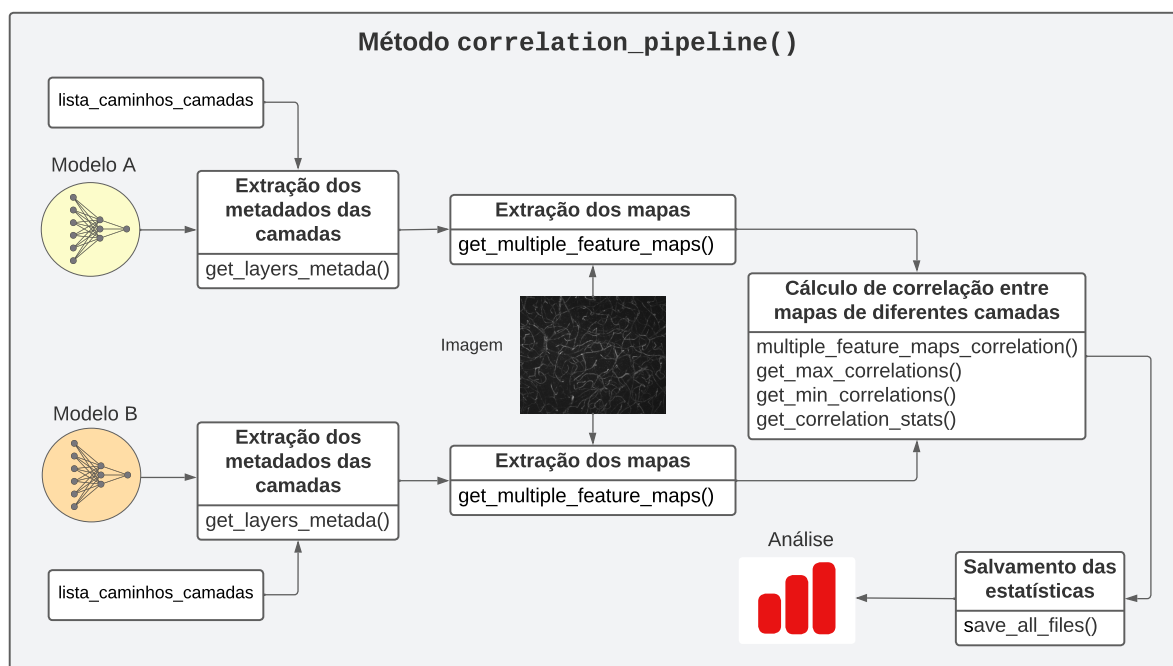
Para exibição dos mapas extraídos, há o método `show_feature_maps`, que plota os mapas para uma determinada camada. O usuário pode passar uma lista com as camadas desejadas, uma lista com os mapas de características previamente extraídos, ou uma lista de índices de mapas que se deseja visualizar.

#### 4.6.1.3 Classe CorrelationExplorer

A classe `CorrelationExplorer` tem como objetivo fornecer métodos que auxiliem no cálculo da correlação entre mapas de ativação de dois modelos pré-treinados, além da obtenção de estatísticas que são úteis para a análise de correlação.

Essa classe pode ser melhor compreendida pelo método `correlation_pipeline`, pois implementa o fluxo necessário para a obtenção das estatísticas sobre a correlação entre mapas de duas CNNs e em diferentes camadas. Como ilustrado na Figura 12, o método recebe dois modelos e a lista de camadas desejadas, respectivas aos modelos, faz a extração dos metadados dessas camadas, seguido da extração dos mapas contidos nas camadas especificadas, e o cálculo de similaridade entre os mapas das redes. Por fim, são gerados e salvos arquivos contendo estatísticas sobre as correlações, para que possam ser carregados em futuras análises.

**Figura 12 – Esquemático do pipeline do cálculo de correlação**



Fonte: Próprio autor

O método `feature_maps_correlation` é o principal método da classe, que recebe os mapas previamente extraídos das camadas correspondentes de cada modelo, além da identificação dessas camadas. Tendo esses dados, o método calcula a correlação de Pearson entre pares de mapas e armazena em um *Data Frame*.

Para contemplar todas as camadas e conjuntos de mapas de ambos os modelos, usa-se o método `multiple_feature_maps_correlation`, que recebe uma lista de mapas de características para cada um dos modelos, bem como os metadados necessários, e então itera chamando o método `feature_maps_correlation` até que todas as combinações de correlações sejam feitas.

Uma vez que todas as combinações de pares de mapas de ativação estejam calculadas e devidamente identificadas pelos índices desses mapas, e também a quais redes e camadas pertencem, os métodos `get_max_correlations` e `get_min_correlations` são responsáveis por calcular as estatísticas de máxima e mínima correlação que cada um dos mapas obteve nas combinações previamente calculadas.

Já o método `get_correlation_stats` produz algumas estatísticas sobre os dados provenientes das funções `get_max_correlations` e `get_min_correlations`. Por exemplo, são calculados valores de frequência absoluta e relativa das ocorrências dos mapas que foram considerados como os pares com a máxima correlação. Ou seja, esses índices podem revelar o nível de repetição de um conjunto de mapas nas máximas correlações verificadas. Além disso, o método também produz estatísticas sobre a correlação média e desvio padrão. Analogamente, o mesmo ocorre para valores das mínimas correlações.

Por fim, o método `save_all_files` salva todas os artefatos estatísticos em um diretório especificado. O usuário pode escolher salvar os arquivos em *csv* ou *json*.

## 4.7 Exploração dos mapas de ativações de CNNs

Como mencionado na seção anterior, a biblioteca *CNN Explorer* proporciona a visualização de mapas de ativações, bem como a produção de arquivos com estatísticas relevantes para a análise. Contudo, os arquivos possuem uma quantidade massiva de informações, devido às milhares de combinações entre pares de mapas presentes nas redes. Portanto, a análise dos dados requer algumas estratégias de agregação para que sejam possíveis a análise descritiva e também a criação de visualizações que sejam representativas e eficazes em explicar as relações entre mapas de diferentes profundidades.

Para ilustrar o desafio da análise exploratória dos dados, considere uma CNN com dois módulos convolucionais. Essa rede passou pelo *pipeline* de correlação, e dela foram extraídos os seus mapas e todas as correlações entre as representações das duas camadas consideradas. A Tabela 1 exemplifica as correlações obtidas de uma rede chamada *cnn2*. Comparou-se a rede

com ela mesma para produzir os resultados. Nota-se o grande volume de registros produzidos pela computação das correlações, com mais de 16300 registros para uma rede de apenas dois módulos convolucionais, que gera 64 mapas na saída desses módulos.

Na tabela, cada mapa é identificado pela rede e seu índice, de 0 à 63, além da camada a qual pertence. A interpretação da tabela pode ser generalizada da seguinte forma:

- *modelName\_map\_id* : para cada um dos modelos, identifica o nome do modelo e os IDs dos mapas.
- *correlation*: valor absoluto da correlação de Pearson.
- *modelName\_zero\_flag*: para cada um dos modelos, sinaliza se o mapa correspondente é formado majoritariamente por zeros, sem ativações.
- *modelName\_layer*: para cada um dos modelos, identifica a qual módulo convolucional o mapa pertence.

**Tabela 1 – Exemplo de tabela de correlações entre mapas de uma rede com dois módulos convolucionais**

index	cnn2(1)_map_id	cnn2(2)_map_id	correlation	cnn2(1)_zero_flag	cnn2(2)_zero_flag	cnn2(1)_layer	cnn2(2)_layer
0	0	0	0.99	0	0	conv_module1	conv_module1
1	0	1	0.50	0	0	conv_module1	conv_module1
2	0	2	0.10	0	0	conv_module1	conv_module1
3	0	3	0.38	0	0	conv_module1	conv_module1
4	0	4	0.18	0	0	conv_module1	conv_module1
...	...	...	...	...	...	...	...
16379	63	59	0.77	0	0	conv_module2	conv_module2
16380	63	60	0.20	0	0	conv_module2	conv_module2
16381	63	61	0.60	0	0	conv_module2	conv_module2
16382	63	62	0.14	0	0	conv_module2	conv_module2
16383	63	63	1.0	0	0	conv_module2	conv_module2

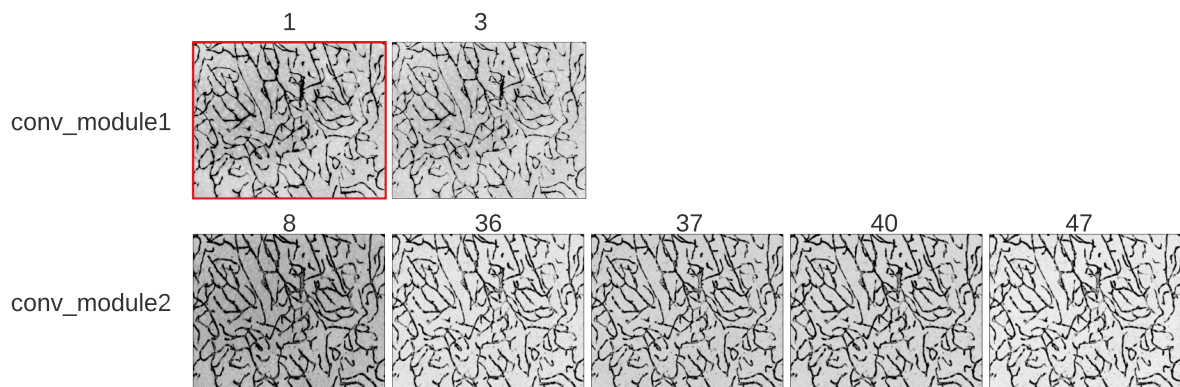
O primeiro passo para a redução da dimensão da tabela de correlação é utilizar a estratégia de capturar a máxima correlação que os mapas obtiveram ao compará-los com os seus pares, agregando por camada. Por exemplo, na Tabela 2, o mapa de ID 2 obteve a máxima correlação com o mapa de ID 38, ambos contidos na mesma camada *conv\_module1*. Isso facilita a análise, ao passo que as correlações muito pequenas são ignoradas nesse momento, reduzindo a tabela resultante para 256 registros. É importante destacar que, para correlação entre mapas da mesma rede, a máxima correlação obtida é entre pares distintos. Isto é, não são considerados valores de correlação iguais a 1 quando um mapa é comparado com ele mesmo.

Uma outra forma de visualizar é plotando os mapas bem correlacionados com o auxílio da função de *plot* de mapas presente na biblioteca *CNN Explorer*. Como visto na Tabela 2, alguns mapas podem aparecer mais de uma vez como sendo os correspondentes que apresentaram a máxima correlação. A Figura 13 exibe uma amostra desse comportamento. O mapa de ID 1, presente na camada *conv\_module1*, é o correspondente que apresentou a máxima correlação

**Tabela 2 – Exemplo de tabela de máximas correlações entre mapas de uma rede com dois módulos convolucionais**

index	cnn2(1)_map_id	cnn2(2)_map_id	max_correlation	cnn2(1)_zero_flag	cnn2(2)_zero_flag	cnn2(1)_layer	cnn2(2)_layer
0	0	1	0.50	0	0	conv_module1	conv_module1
1	1	3	0.93	0	0	conv_module1	conv_module1
2	2	38	0.29	0	0	conv_module1	conv_module1
3	3	1	0.93	0	0	conv_module1	conv_module1
4	4	15	0.67	0	0	conv_module1	conv_module1
...	...	...	...	...	...	...	...
251	59	53	0.93	0	0	conv_module2	conv_module2
252	60	53	0.52	0	0	conv_module2	conv_module2
253	61	47	0.96	0	0	conv_module2	conv_module2
254	62	30	0.58	0	0	conv_module2	conv_module2
255	63	6	0.92	0	0	conv_module2	conv_module2

maior que 0.8 com os outros mapas exibidos na figura. Ou seja, o mapa 1 teve mais de uma ocorrência nas máximas correlações.

**Figura 13 – Exemplo de visualização dos mapas bem correlacionados. O número acima de cada figura indica o ID do mapa na respectiva camada.**

**Fonte: Próprio autor. Adaptação do plot da biblioteca CNN Explorer**

Em relação à representatividade de mapas bem correlacionados, escolheu-se um limiar de 0.80 no valor das máximas correlações. Utilizando o método de tabela pivotada e agregando pelas camadas, é possível fazer a contagem de quantas vezes cada mapa foi considerado sendo o correspondente com a máxima correlação, e então ter as estatísticas da quantidade de mapas bem correlacionados entre camadas, assim como a correlação média, mediana, desvio padrão e proporção de mapas bem correlacionados em relação às camadas observadas.

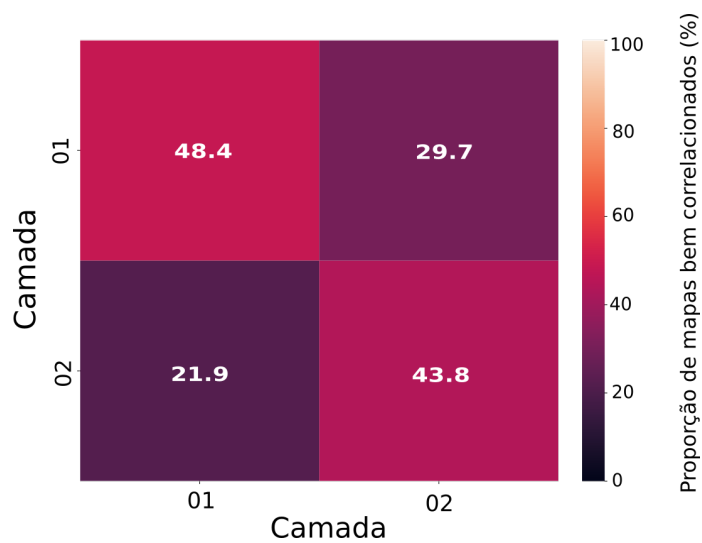
Por exemplo, na Tabela 3 vê-se que há 31 mapas que foram identificados como os pares com a máxima correlação maiores ou iguais a 0.80, ao comparar todos os mapas da camada *conv\_module1*. Isso indica que a camada possui 48,43% dos seus mapas bem correlacionados com os mapas da mesma camada.

Essas relações também podem ser visualizadas por um gráfico de mapa de calor, como mostrado pela Figura 14. Dessa forma, visualmente se torna intuitivo, para redes maiores, verificar se há uma tendência de crescimento na proporção de mapas bem correlacionados a

**Tabela 3 – Exemplo de tabela pivotada de máximas correlações entre mapas de uma rede com dois módulos convolucionais**

index	cnn2(2)_layer	cnn2(1)_layer	count_correlation	mean_correlation	median_correlation	std_correlation	%_count
0	conv_module1	conv_module1	31	0.90	0.91	0.056	48.43
1	conv_module1	conv_module2	14	0.88	0.89	0.046	21.87
2	conv_module2	conv_module1	19	0.85	0.85	0.043	29.68
3	conv_module2	conv_module2	28	0.92	0.93	0.043	43.75

medida que a rede se aprofunda.

**Figura 14 – Exemplo de gráfico para visualizar proporção de mapas bem correlacionados por camada**

**Fonte: Próprio autor**

Outras formas de visualização, como histograma, também são possíveis para analisar a distribuição das correlações numa determinada camada. As estratégias de agrupamento e visualizações apresentadas nesta seção foram importantes para as análises realizadas sobre os resultados, que serão apresentadas no próximo capítulo.

# Capítulo 5

## RESULTADOS

---

---

Neste capítulo são apresentados os resultados obtidos a partir dos experimentos descritos no capítulo anterior. Portanto, aqui serão explorados resultados do treinamento das redes propostas e também a análise das correlações entre os mapas de ativação dessas redes.

### 5.1 Treinamento e análise das CNNs

Para o treinamento dos modelos, algumas decisões foram tomadas a fim de padronizar os resultados dos experimentos e torná-los reproduzíveis. Portanto, os critérios considerados para o treinamento dos modelos foram os seguintes:

1. Divisão do dataset: 80% para treino e 20% para validação, com a semente do gerador de números aleatórios fixa para evitar aleatorização;
2. Não utilização de *Data Augmentation* para evitar transformações que causem aleatorização na imagem resultante;
3. Normalização das intensidades dos valores das imagens;
4. Imagens de entrada com resolução original:  $1376 \times 1104$ ;
5. Taxa de aprendizado igual a 0.01;
6. 250 épocas de treinamento;
7. O modelo com maior valor de IoU encontrado durante o treinamento é utilizado;
8. Tamanho do batch variável, de acordo com a profundidade da rede. Por exemplo, a rede com 10 camadas precisou ser treinada com tamanho de batch igual a 1, pois a alocação de recursos na memória da GPU excedia a sua capacidade caso o tamanho do batch fosse maior que 1.
9. Todos os valores de performance são calculados utilizando o conjunto de validação.



Dadas as definições de critérios, os modelos foram treinados e avaliados. A Tabela 4 exibe algumas informações relevantes sobre os modelos. Primeiro, os modelos convergiram para o melhor IoU próximos ao total de épocas configurado. A medida que a profundidade da rede cresce, a quantidade de parâmetros gerados também cresce numa grande velocidade, sendo que os modelos CNN6 e CNN10 estão na faixa dos milhões de parâmetros. Cresce também a capacidade do modelo em segmentar as regiões de vasos e não segmentar erroneamente as regiões de fundo, o que pode ser observado pelo IoU. Por último, uma outra medida considerada foi o índice *clDice*, que indica a qualidade da manutenção da continuidade das estruturas dos vasos sanguíneos segmentados. Vê-se que essa medida tende a crescer conforme a rede gera mais representações, contudo, nessa análise preliminar, as duas redes mais profundas aparentam ter desempenho similar.

**Tabela 4 – Informações do treinamento dos modelos**

Modelo	Camadas	Módulos convolucionais	Tamanho do Batch	Tempo de treinamento	Parâmetros	Melhor época	<i>IOU</i>	<i>clDice</i>
CNN1	2	1	5	2h40min	3.3K	221	0.81	0.88
CNN6	7	6	3	11h40min	1.01M	237	0.91	0.97
CNN10	11	10	1	17h46min	1.81M	220	0.92	0.98

A avaliação qualitativa também é importante para verificar a qualidade visual do modelo em segmentar os vasos sanguíneos. Então, considere a Tabela 5 e as informações correspondentes na Figura 15. Essa avaliação leva em conta três imagens distintas, com diferentes escalas e densidade de vasos sanguíneos. Um destaque para a Figura 15 (Imagem 2), que possui ruídos nas regiões de fundo que foram evidenciados após o processo de normalização das imagens.

Como se vê, o modelo CNN1 consegue segmentar imagens com baixa densidade de vasos, mantendo um bom *clDice*. Mas à proporção que a rede de vasos sanguíneos se torna mais complexa e densa, o modelo CNN1 passa a errar na predição dos pixels de fundo, aumentando a taxa de falso positivo. Isso pode ser visto nas imagens 2 e 3 da Figura 15. Nessas mesmas imagens, o modelo teve dificuldades em manter a estrutura dos vasos sanguíneos, provocando um decaimento significativo do *clDice*.

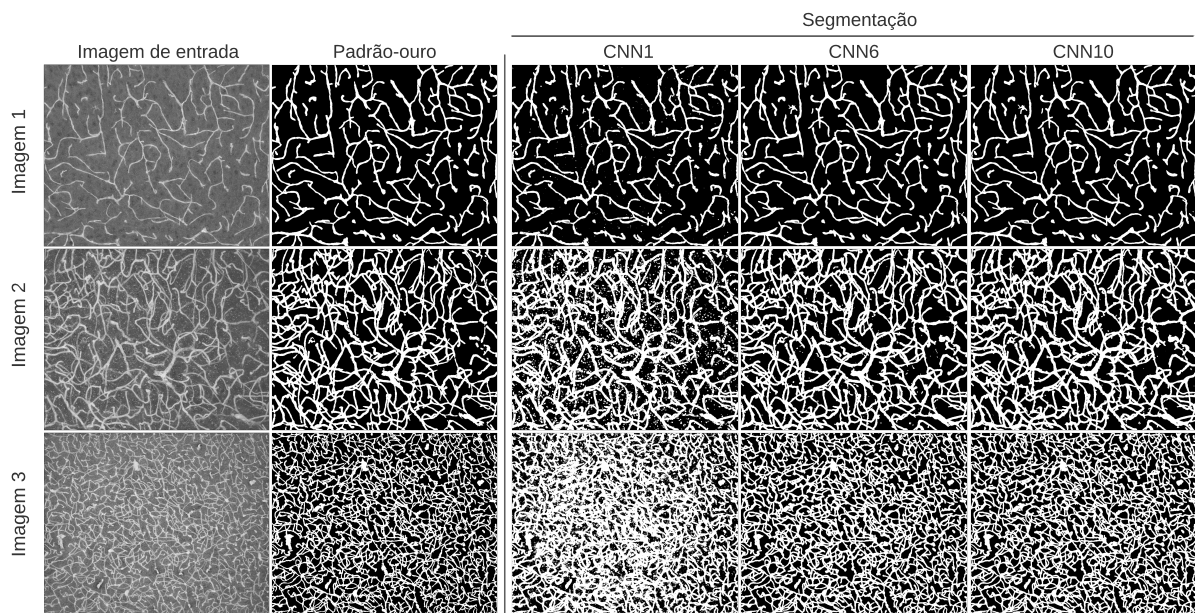
**Tabela 5 – Avaliação da qualidade das segmentações dos modelos treinados**

Modelo	Imagem	<i>IOU</i>	<i>clDice</i>
CNN1	1	0.89	0.95
	2	0.75	0.83
	3	0.61	0.74
CNN6	1	0.95	0.99
	2	0.86	0.96
	3	0.79	0.97
CNN10	1	0.96	0.99
	2	0.83	0.96
	3	0.79	0.97

Já os modelos CNN6 e CNN10 apresentaram os melhores resultados métricos e visuais. Ambos tiveram bom desempenho visual em segmentar tanto regiões de vasos sanguíneos quanto

regiões de fundo, mesmo nas imagens com redes complexas de vasos. As segmentações desses modelos são menos ruidosas que as do modelo CNN1, gerando menos falsos positivos e falsos negativos. Também em relação à capacidade de preservar as estruturas e continuidade houve um bom desempenho, alcançando valores de *cIDice* maiores que 0.95 para as imagens avaliadas.

**Figura 15 – Qualidade das segmentações dos modelos treinados**



**Fonte: Próprio autor**

De acordo com as métricas de treinamento dos modelos e avaliação das segmentações, verificou-se que os modelos mais profundos e, portanto, com maior quantidade de mapas de ativação gerados, tendem a produzir melhores resultados quantitativos e qualitativos na tarefa de segmentação. Essa propriedade da relação proporcional entre a profundidade das CNNs, volume de parâmetros e acurácia já foi estudada e verificada nos trabalhos de (BA; CARUANA, 2013) e (TAN; LE, 2019).

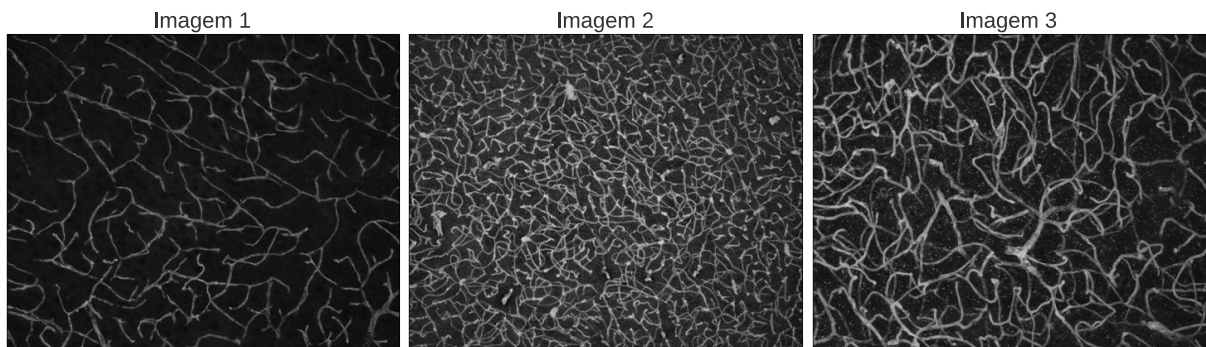
O bom desempenho de redes profundas pode ser explicado pela quantidade de mapas de ativação produzidos por estas, isto é, um grande volume de representações do aprendizado da rede que foram gerados durante o processo de convolução e ativação. A variabilidade dos mapas pode contribuir para uma rede mais discriminativa nas características que precisam ser ativadas, contudo, conforme verificado pelas literaturas apresentadas na seção 3.2, as CNNs podem produzir representações redundantes. Isso será explorado na próxima seção.

## 5.2 Análise de correlação de mapas de ativação

Tendo os modelos treinados, as redes passaram pelo pipeline de correlação esquematizado na Figura 12. Logo, os mapas dessas redes foram extraídos e as correlações de mapas na

mesma rede e entre redes foi devidamente calculada. Para tal, utilizou-se de uma amostra de três imagens, mostradas na Figura 16, que serviram de insumo para as ativações dos modelos.

**Figura 16 – Amostra de imagens utilizadas na análise de correlação**



**Fonte: Próprio autor**

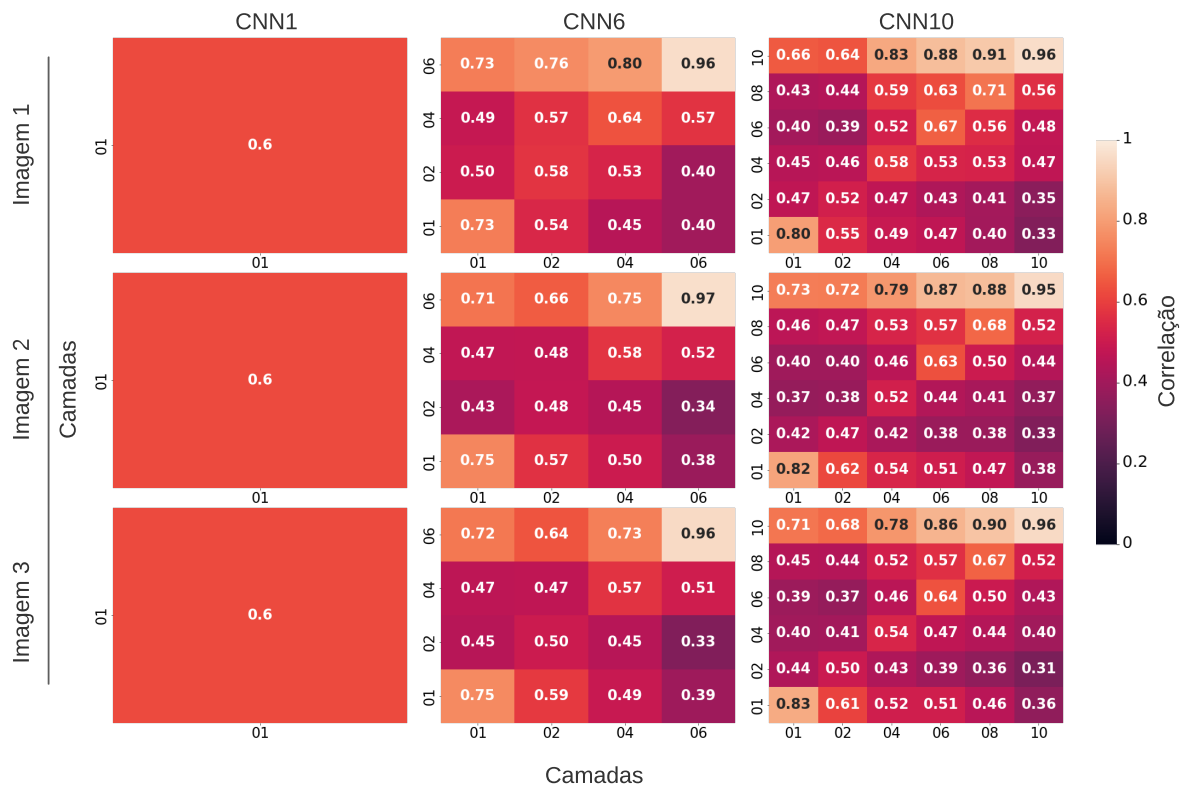
Os experimentos foram divididos em duas etapas: (i) análise da correlação de mapas presentes na mesma rede; (ii) análise de correlação entre mapas de redes distintas. Essas análises foram processadas sobre os dados de máxima correlação entre os pares de mapas, como descrito na seção 4.7.

### 5.2.1 Correlação na mesma rede

O objetivo deste estudo é verificar se uma CNN pode produzir representações redundantes nas suas camadas intermediárias, além de investigar se existe alguma tendência na produção de mapas correlacionados a medida que a CNN se torna mais profunda, mesmo gerando ativações com imagens de entrada diferentes.

Sendo assim, para descrever a análise, a Figura 17 está organizada de forma a mostrar para cada uma das imagens de entrada, e para cada um dos modelos, o resultado da média das máximas correlações dos mapas agregadas por camadas. Pelos mapas de calor, nota-se um padrão de nível de correlação em todos os modelos. A primeira camada quando comparada a si própria obteve um nível de correlação média considerável em todos os modelos, indicado que há representações similares na primeira camada de ativações. Conforme a rede se torna profunda, há uma aparente variabilidade das representações aprendidas nas camadas intermediárias. Contudo, a última camada das redes CNN6 e CNN10 começam a apresentar forte correlação linear quando os mapas são comparados com seus pares nas últimas camadas.

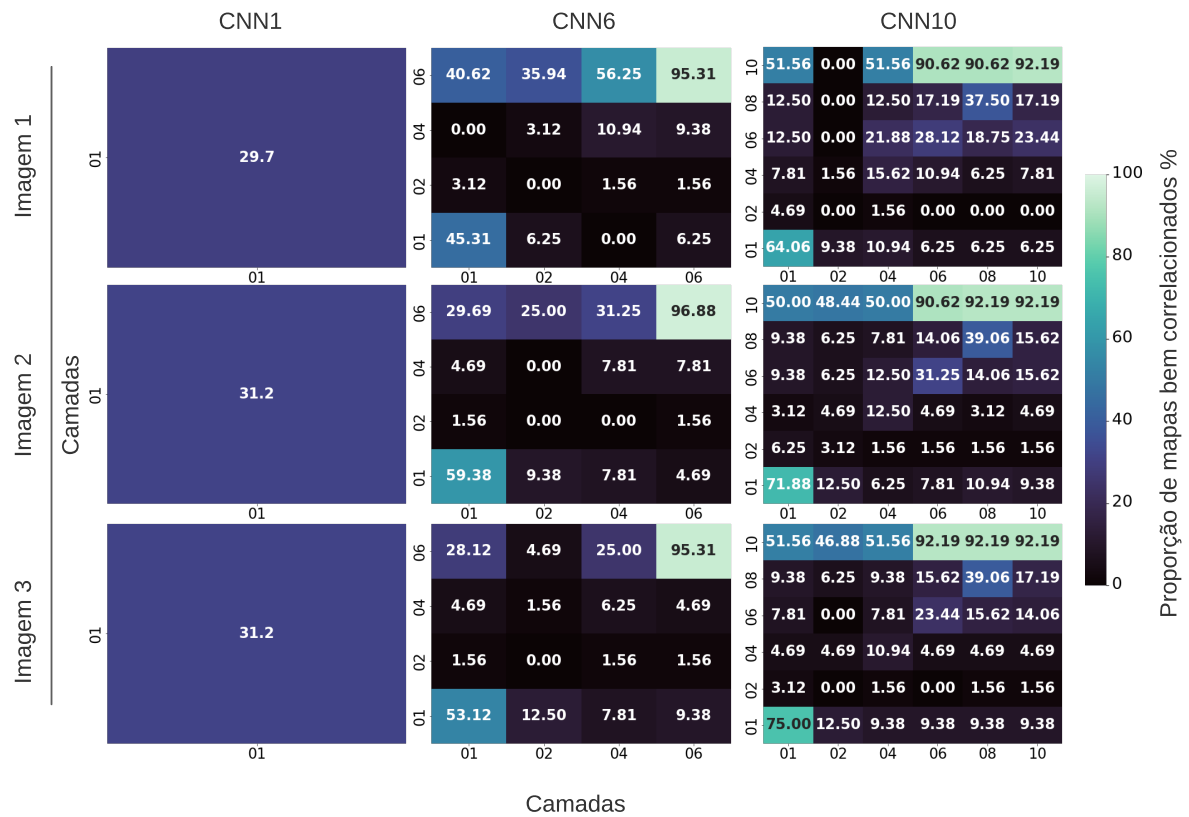
**Figura 17 – Nível de correlação absoluta entre mapas de uma mesma CNN. Visualização por profundidade da rede**



**Fonte: Próprio autor**

Uma outra análise pertinente é calcular a representatividade dos mapas bem correlacionados. Considerou-se mapas bem correlacionados os que apresentaram a máxima correlação maior ou igual à 0.8. Na Figura 18 é possível confirmar que as maiores proporções de mapas bem correlacionados se encontram justamente na primeira camada e nas duas últimas camadas das redes mais profundas.

**Figura 18 – Proporção de mapas bem correlacionados na mesma CNN. Visualização por profundidade da rede**



Fonte: Próprio autor

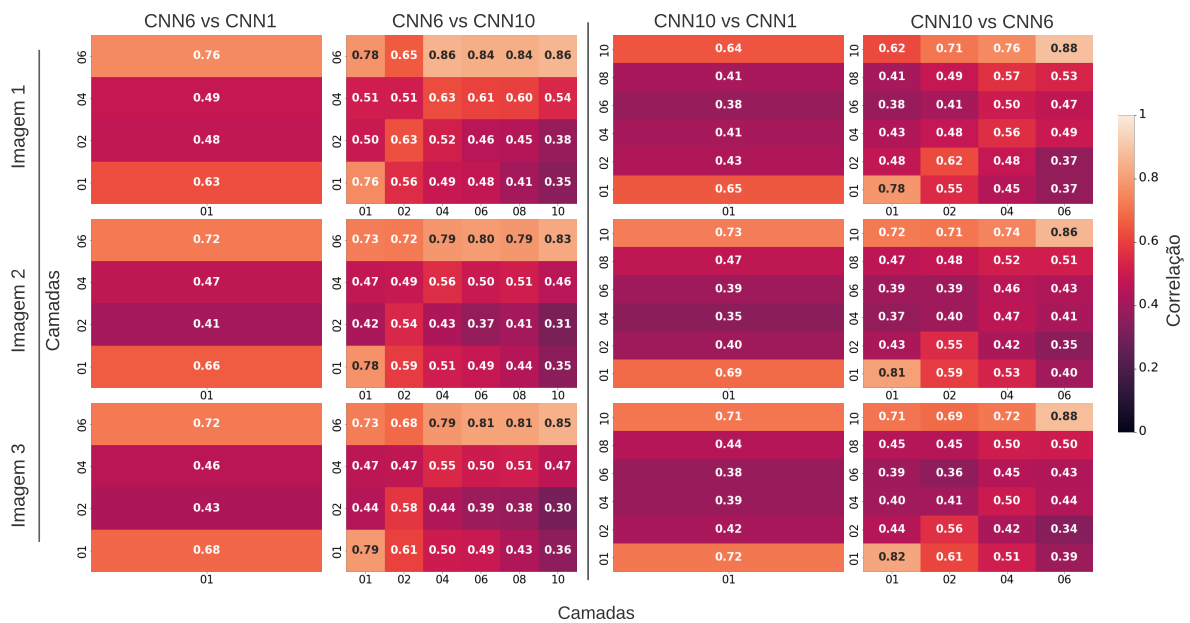
A CNN1 possui cerca de 31% dos seus mapas bem correlacionados, e a camada 2 das redes CNN6 e CNN10 tendem a ter menores ocorrências de mapas bem correlacionados. Tal padrão sugere que os mapas são mais variados nessa camada.

### 5.2.2 Correlação entre redes distintas

Nesta etapa, as análises são realizadas sobre o nível de correlação entre mapas de diferentes redes. De forma semelhante ao que foi feito na etapa anterior, o estudo é realizado sobre a média das máximas correlações e proporção de mapas bem correlacionados. Aqui são examinadas as interações: CNN6 vs CNN1, CNN6 vs CNN10, CNN10 vs CNN1 e CNN10 vs CNN6.

A Figura 19 evidencia o padrão observado na análise anterior, de que as camadas das extremidades da rede apresentam mapas com alto grau de correlação média, mesmo entre modelos com diferentes profundidades. De acordo com os gráficos, as camadas intermediárias, por sua vez, podem gerar mapas menos similares, reduzindo a redundância de aprendizado.

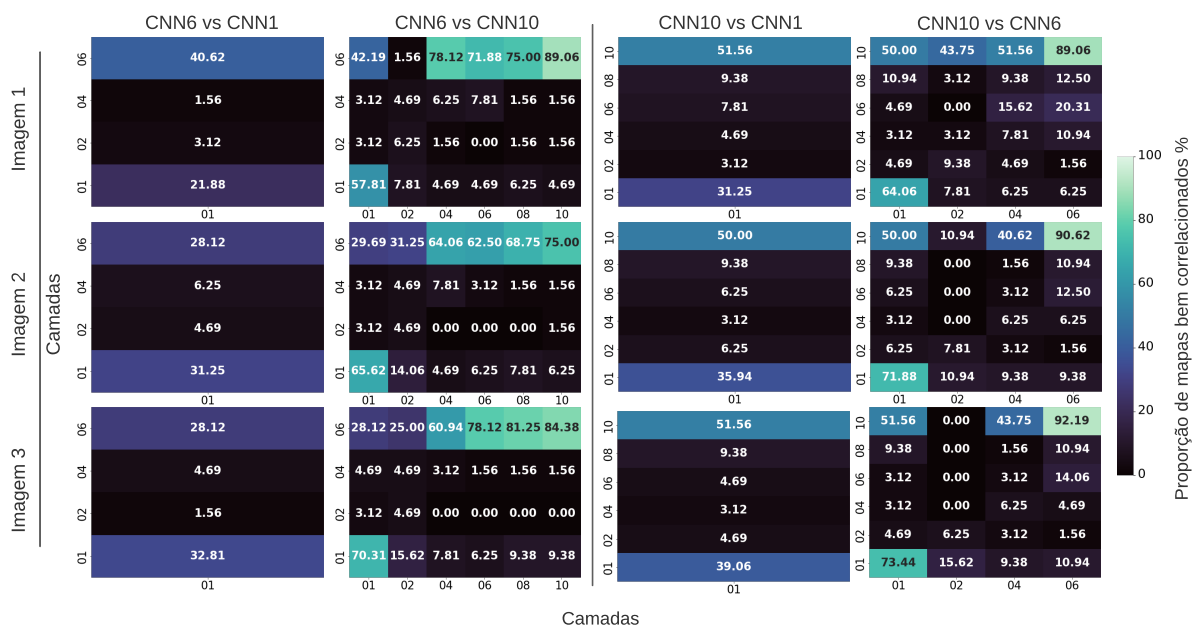
**Figura 19 – Nível de correlação absoluta entre mapas de CNNs distintas. Visualização por profundidade da rede**



**Fonte: Próprio autor**

Na análise de representatividade dos mapas bem correlacionados, também se verificou que estes aparecem nas extremidades da rede, contudo, mais fortemente nas últimas camadas das redes mais profundas. Com esses dados, há evidências de que redes diferentes em profundidade também podem aprender as mesmas características e produzir mapas de ativação semelhantes. Essa verificação é corroborada também pelos trabalhos (KORNBLITH et al., 2019) e (LEE et al., 2020).

**Figura 20 – Proporção de mapas bem correlacionados entre CNNs distintas. Visualização por profundidade da rede**



Fonte: Próprio autor

Em suma, este estudo explorou a correlação de mapas de ativação em CNNs de diferentes profundidades, com o objetivo de entender melhor como as representações são aprendidas em diferentes camadas e como a correlação pode ocorrer entre redes distintas. Os resultados sugerem que as camadas intermediárias das redes tendem a produzir representações mais variadas e menos redundantes, enquanto as camadas das extremidades tendem a produzir representações mais similares e correlacionadas.

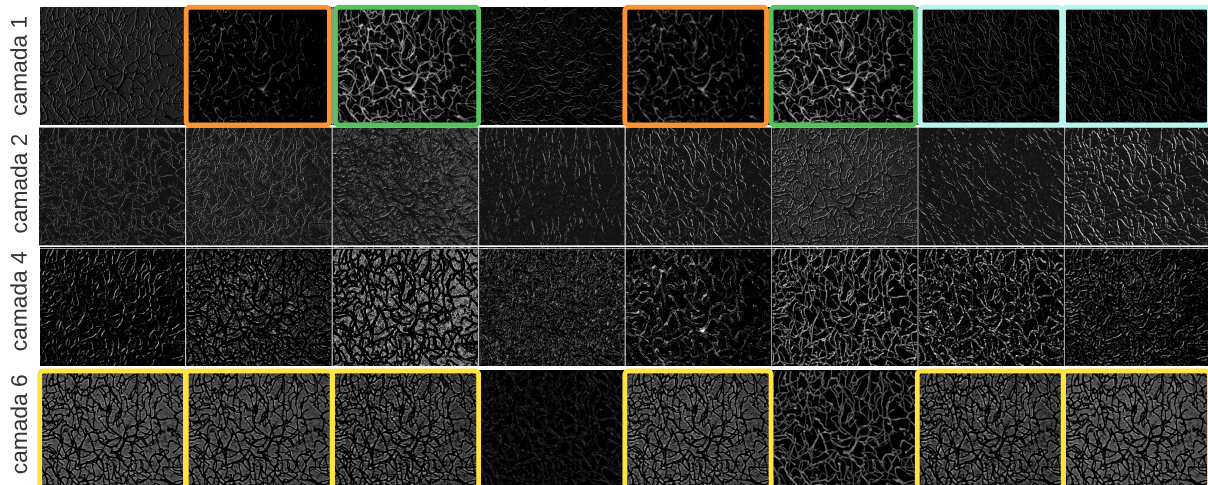
Uma hipótese para essa constatação é que a rede é treinada com dados padronizados na entrada e saída. Ou seja, para a primeira camada, durante o treinamento a rede utiliza o mesmo conjunto de imagens, e essa camada é responsável por assimilar características mais simples dos objetos presentes nas imagens, como já foi abordado na seção 3.1. Por outro lado, a última camada produz mapas de características que podem possuir valores que se aproximam do padrão-ouro, pois esses mapas precisam ser combinados para posteriormente gerar os mapas de probabilidade, que originam finalmente a imagem segmentada ao passarem pela última função de ativação da rede. Portanto, isso limita o espaço de possibilidades dos valores que os mapas podem ter na primeira e última camada, aumentando a probabilidade de produção de mapas similares.

Em contrapartida, as camadas intermediárias extraem características mais complexas por conta do aprendizado hierárquico, tendo maior variabilidade de valores nos mapas de ativação, conseqüentemente, tornando-os menos similares. Para ilustrar essa hipótese, considere o experimento ilustrado na Figura 21. Foram extraídos de forma aleatória 8 mapas de ativação



das camadas 1, 2, 4 e 6 da rede CNN6. Nota-se que as camadas 1 e principalmente a camada 6 apresentam mapas visualmente similares localizados pelas cores destacadas. Enquanto isso, as duas camadas intermediárias possuem mapas visualmente diferentes, com características mais complexas.

**Figura 21 – Mapas de ativação extraídos da CNN6. Bordas coloridas indicam exemplos de mapas similares na primeira e última camada.**



**Fonte: Próprio autor**

Além disso, foi observado que redes diferentes em profundidade podem aprender características semelhantes e produzir mapas de ativação correlacionados, como foi verificado também pelas literaturas revisadas. Essas descobertas têm implicações importantes para o desenvolvimento de redes neurais mais eficientes, pois mostra a oportunidade de se utilizar técnicas de poda estrutural para remover representações redundantes, a fim de retrainar a rede, mas manter o mesmo desempenho esperado.



# Capítulo 6

## CONCLUSÃO

---

---

O estudo da similaridade entre mapas de ativação das CNNs é fundamental para investigar a redundância de aprendizado que essas redes produzem. Compreender como as diferentes camadas das redes estão relacionadas pode levar à implementação de abordagens para aprimorar a arquitetura dos modelos e melhorar a eficiência das redes em termos de custo de computação, reduzindo a quantidade de parâmetros. Além disso, a visualização das camadas e análise de similaridade ajudam a entender melhor como as redes de aprendizado profundo assimilam as características e representam informações. Entretanto, o cálculo de informações massivas que as redes produzem não é uma tarefa trivial e é inviável realizá-la manualmente devido à alta dimensionalidade das representações.

Diante desse contexto, neste trabalho foi apresentada a proposta de investigação do grau de correlação e similaridade entre diferentes mapas produzidos pelas redes, seja por uma única rede ou pela comparação entre redes de diferentes profundidade. Para que este objetivo pudesse ser alcançado, o trabalho foi formado por etapas importantes de estudos e desenvolvimento que contribuíram para a obtenção dos resultados. Portanto, houve a necessidade da compreensão de conceitos fundamentais de aprendizado profundo e como as suas representações são formadas, a revisão bibliográfica de trabalhos relevantes que já haviam se debruçado sobre os assuntos abordados e, provavelmente o passo mais importante, a definição de uma metodologia para sumarizar a relação entre mapas de ativação de redes. Além disso, o desenvolvimento da ferramenta *CNN Explorer* foi fundamental para automatizar a extração e visualização dos artefatos produzidos pelas CNNs, que foram necessários para as análises empregadas.

A respeito da análise de correlação de mapas de ativação de redes neurais convolucionais, foi possível verificar que as CNNs produzem mapas redundantes a medida que se aprofundam, tendendo a formar blocos de mapas bem correlacionadas nas últimas camadas. Esse padrão foi verificado tanto no experimento de correlação entre mapas da mesma rede quanto entre mapas de redes de diferentes profundidade, combinando as diferentes camadas existentes. É importante ressaltar que esse comportamento foi observado utilizando distintas imagens de entrada. Portanto, como proposto pelos trabalhos das literaturas aqui revisadas, CNNs distintas podem compartilhar as mesmas características aprendidas das imagens de entrada, principalmente nas camadas mais

profundas. No entanto, notou-se uma variabilidade das representações aprendidas nas camadas intermediárias, pois obtiveram a menor correlação média e também a menor taxa de mapas bem correlacionados.

A ferramenta *CNN Explorer* (GALVAO, 2023) permitiu conhecer quais mapas de ativação são redundantes e onde estes estão localizados, o que abre margem para a abordagem de poda das redes, a fim de torná-las computacionalmente mais eficientes, mas mantendo a mesma qualidade das segmentações esperadas. O pacote foi disponibilizado publicamente em um repositório do GitHub . Ademais, a escolha da métrica da correlação de Pearson como medida de similaridade se mostrou acertada para encontrar mapas linearmente bem correlacionados. No entanto, esta métrica pode ocultar relações não lineares entre mapas que são similares, impossibilitando a localização dos conjuntos de mapas similares que não atendem ao critério de correspondência linear. Logo, faz-se necessário experimentar outras métricas difundidas e com bons resultados já demonstrados por outros trabalhos, como o índice CKA.

## 6.1 Trabalhos futuros

Com o objetivo de expandir o trabalho realizado, há algumas propostas para melhorias da biblioteca desenvolvida e aprofundamento nas investigações sobre a similaridade dos mapas e técnicas para aprimoramento das arquiteturas das redes neurais convolucionais.

1. Estudar meios de tornar as redes mais eficientes computacionalmente pelo método de poda estrutural, uma vez que já é possível identificar quais camadas e mapas produzem correspondentes redundantes.
2. Investigar a relação entre a etapa *backpropagation* e a geração de mapas similares.
3. Empregar meios de redução de dimensionalidade dos mapas de características, como a técnica Aproximação e Projeção de Manifold Uniforme para Redução de Dimensões (UMAP), para clusterizar esses mapas e então visualizá-los em termos de similaridade por vizinhança no espaço de valores.
4. Verificar como ocorre a distribuição de mapas similares nas CNNs com supervisão profunda, já que estas produzem mapas mais discriminativos, como descrito por (LEE et al., 2015) e (FU et al., 2016).
5. Implementar na *CNN Explorer* outras medidas de similaridade, como similaridade cosseno e CKA. Permitir ao usuário escolher qual métrica será utilizada no processamento.
6. Atualmente o pipeline de correlação permite processar a correlação entre os mapas para uma única imagem de entrada por vez. Possibilitar esse cálculo para múltiplas imagens, de forma a extrair as médias e desvio padrão dos valores para se obter uma sumarização de correlação para um conjunto de imagens de entrada.

7. Implementar também formas de visualização dos mapas e gráficos de correlação mais interativos por meio de interface gráfica de usuário (GUI).

## REFERÊNCIAS

---

- AZMITIA, E. C.; SACCOMANO, Z. T.; ALZOOBAEE, M. F.; BOLDRINI, M.; WHITAKER-AZMITIA, P. M. Persistent Angiogenesis in the Autism Brain: An Immunocytochemical Study of Postmortem Cortex, Brainstem and Cerebellum. *Journal of Autism and Developmental Disorders*, v. 46, n. 4, p. 1307–1318, apr 2016. ISSN 0162-3257. Disponível em: <<http://link.springer.com/10.1007/s10803-015-2672-6>>. Citado na página 14.
- BA, L. J.; CARUANA, R. Do deep nets really need to be deep? *Advances in Neural Information Processing Systems*, v. 27, 12 2013. Disponível em: <<http://arxiv.org/abs/1312.6184>>. Citado na página 47.
- DENG, J.; DONG, W.; SOCHER, R.; LI, L.-J.; LI, K.; FEI-FEI, L. Imagenet: A large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2009. p. 248–255. ISSN 1063-6919. Citado na página 29.
- 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>>. Citado na página 19.
- ERHAN, D.; BENGIO, Y.; COURVILLE, A.; VINCENT, P. Visualizing higher-layer features of a deep network. *University of Montreal*, v. 1341, p. 1, 2009. Citado na página 28.
- FU, H.; XU, Y.; LIN, S.; WONG, D. W. K.; LIU, J. *DeepVessel: Retinal Vessel Segmentation via Deep Learning and Conditional Random Field*. 2016. 132-139 p. Citado 2 vezes nas páginas 15 e 55.
- 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. Citado na página 19.
- GALVAO, W. N. *CNN Explorer: A library to extract and visualize feature maps from convolutional neural networks, and analysis the correlation level between these maps*. 2023. Disponível em: <[https://github.com/wesleygalvao/cnn\\_explorer](https://github.com/wesleygalvao/cnn_explorer)>. Citado 2 vezes nas páginas 38 e 55.
- GIGANDET, X.; CUADRA, M. B.; THIRAN, J. *Satellite image segmentation and classification*. [S.l.], 2004. Citado na página 24.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: MIT Press, 2016. <<http://www.deeplearningbook.org>>. Citado na página 19.
- HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 770–778. Citado na página 30.

KAPLAN, L.; CHOW, B. W.; GU, C. Neuronal regulation of the blood–brain barrier and neurovascular coupling. *Nature Reviews Neuroscience*, Nature Publishing Group, v. 21, n. 8, p. 416–432, 2020. Citado na página 14.

KLUFAS, M. A.; PATEL, S. N.; RYAN, M. C.; GUPTA, M. P.; JONAS, K. E.; OSTMO, S.; MARTINEZ-CASTELLANOS, M. A.; BERROCAL, A. M.; CHIANG, M. F.; CHAN, R. P. Influence of fluorescein angiography on the diagnosis and management of retinopathy of prematurity. *Ophthalmology*, Elsevier, v. 122, n. 8, p. 1601–1608, 2015. Citado na página 14.

KORNBLITH, S.; NOROUZI, M.; LEE, H.; HINTON, G. Similarity of neural network representations revisited. *36th International Conference on Machine Learning, ICML 2019*, International Machine Learning Society (IMLS), v. 2019-June, p. 6156–6175, 5 2019. Disponível em: <<https://arxiv.org/abs/1905.00414v4>>. Citado 3 vezes nas páginas 29, 30 e 51.

KRIZHEVSKY, A.; HINTON, G. et al. Learning multiple layers of features from tiny images. Toronto, ON, Canada, 2009. Citado na página 30.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. *Commun. ACM*, Association for Computing Machinery, New York, NY, USA, v. 60, n. 6, p. 84–90, may 2017. ISSN 0001-0782. Disponível em: <<https://doi.org/10.1145/3065386>>. Citado na página 29.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *nature*, Nature Publishing Group UK London, v. 521, n. 7553, p. 436–444, 2015. Citado na página 19.

LECUN, Y.; BOSER, B.; DENKER, J. S.; HENDERSON, D.; HOWARD, R. E.; HUBBARD, W.; JACKEL, L. D. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation*, MIT Press - Journals, v. 1, n. 4, p. 541–551, dec 1989. ISSN 0899-7667. Citado na página 19.

LECUN, Y.; BOSER, B.; DENKER, J. S.; HENDERSON, D.; HOWARD, R. E.; HUBBARD, W.; JACKEL, L. D. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, v. 1, n. 4, p. 541–551, 1989. Citado na página 29.

LEE, C.-Y.; XIE, S.; GALLAGHER, P.; ZHANG, Z.; TU, Z. Deeply-supervised nets. In: LEBANON, G.; VISHWANATHAN, S. V. N. (Ed.). PMLR, 2015. v. 38, p. 562–570. Disponível em: <<https://proceedings.mlr.press/v38/lee15a.html>>. Citado na página 55.

LEE, C.-Y.; XIE, S.; GALLAGHER, P.; ZHANG, Z.; TU, Z. Do wide and deep networks learn the same things? uncovering how neural network representations vary with width and depth. 10 2020. Disponível em: <<https://arxiv.org/abs/2010.15327v2>>. Citado 2 vezes nas páginas 30 e 51.

LEK, S.; PARK, Y. S. Artificial Neural Networks. *Encyclopedia of Ecology, Five-Volume Set*, Academic Press, p. 237–245, jan 2008. Citado na página 18.

LI, H.; KADAV, A.; DURDANOVIC, I.; SAMET, H.; GRAF, H. P. Pruning filters for efficient convnets. 8 2016. Citado na página 31.

LISKOWSKI, P.; KRAWIEC, K. Segmenting retinal blood vessels with deep neural networks. *IEEE Transactions on Medical Imaging*, v. 35, p. 2369–2380, 11 2016. ISSN 0278-0062. Citado na página 15.

- LONG, J.; SHELHAMER, E.; DARRELL, T. Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2015. p. 3431–3440. Citado 2 vezes nas páginas 21 e 35.
- OLAH, C.; MORDVINTSEV, A.; SCHUBERT, L. Feature visualization. *Distill*, v. 2, 11 2017. ISSN 2476-0757. Citado na página 28.
- OUELLETTE, J.; TOUSSAY, X.; COMIN, C. H.; COSTA, L. d. F.; HO, M.; LACALLE-AURIOLES, M.; FREITAS-ANDRADE, M.; LIU, Q. Y.; LECLERC, S.; PAN, Y.; LIU, Z.; THIBODEAU, J.-F.; YIN, M.; CARRIER, M.; MORSE, C. J.; DYKEN, P. V.; BERGIN, C. J.; BAILLET, S.; KENNEDY, C. R.; TREMBLAY, M.-; BENOIT, Y. D.; STANFORD, W. L.; BURGER, D.; STEWART, D. J.; LACOSTE, B. Vascular contributions to 16p11.2 deletion autism syndrome modeled in mice. *Nature Neuroscience*, v. 23, n. 9, p. 1090–1101, set. 2020. ISSN 1097-6256, 1546-1726. Disponível em: <<http://www.nature.com/articles/s41593-020-0663-1>>. Citado 2 vezes nas páginas 14 e 33.
- PITTMAN, R. N. The circulatory system and oxygen transport. In: *Regulation of Tissue Oxygenation*. [S.l.]: Morgan & Claypool Life Sciences, 2011. Citado na página 14.
- RONNEBERGER, O.; FISCHER, P.; BROX, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: NAVAB, N.; HORNEGGER, J.; WELLS, W. M.; FRANGI, A. F. (Ed.). *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Cham: Springer International Publishing, 2015. v. 9351, p. 234–241. ISBN 9783319245737 9783319245744. Disponível em: <[http://link.springer.com/10.1007/978-3-319-24574-4\\_28](http://link.springer.com/10.1007/978-3-319-24574-4_28)>. Citado 4 vezes nas páginas 15, 21, 22 e 35.
- RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. *nature*, Nature Publishing Group UK London, v. 323, n. 6088, p. 533–536, 1986. Citado na página 19.
- SHIT, S.; PAETZOLD, J. C.; EZHOV, I.; SEKUBOYINA, A.; UNGER, A.; ZHYLKA, A.; PLUIM, J. P.; BAUER, U.; MENZE, B. H. cldice-a novel topology-preserving loss function for tubular structure segmentation. Citado na página 25.
- TAGLIASACCHI, A.; DELAME, T.; SPAGNUOLO, M.; AMENTA, N.; TELEA, A. 3d skeletons: A state-of-the-art report. In: WILEY ONLINE LIBRARY. *Computer Graphics Forum*. [S.l.], 2016. v. 35, n. 2, p. 573–597. Citado na página 25.
- TAN, M.; LE, Q. V. Efficientnet: Rethinking model scaling for convolutional neural networks. 5 2019. Disponível em: <<http://arxiv.org/abs/1905.11946>>. Citado na página 47.
- TETTEH, G.; EFREMOV, V.; FORKERT, N. D.; SCHNEIDER, M.; KIRSCHKE, J.; WEBER, B.; ZIMMER, C.; PIRAUD, M.; MENZE, B. H. DeepVesselNet: Vessel Segmentation, Centerline Prediction, and Bifurcation Detection in 3-D Angiographic Volumes. *Frontiers in Neuroscience*, v. 14, p. 592352, dez. 2020. ISSN 1662-453X. Disponível em: <<https://www.frontiersin.org/articles/10.3389/fnins.2020.592352/full>>. Citado na página 15.
- WANG, Z. J.; TURKO, R.; SHAIKH, O.; PARK, H.; DAS, N.; HOHMAN, F.; KAHNG, M.; CHAU, D. H. P. Cnn explainer: learning convolutional neural networks with interactive visualization. *IEEE Transactions on Visualization and Computer Graphics*, IEEE, v. 27, n. 2, p. 1396–1406, 2020. Citado na página 29.

YOSINSKI, J.; CLUNE, J.; NGUYEN, A.; FUCHS, T.; LIPSON, H. Understanding neural networks through deep visualization. 6 2015. Disponível em: <<https://arxiv.org/abs/1506.06579v1>>. Citado na página 28.

ZEILER, M. D.; FERGUS, R. *Visualizing and Understanding Convolutional Networks*. 2014. 818-833 p. Citado na página 28.

ZOU, K. H.; WARFIELD, S. K.; BHARATHA, A.; TEMPANY, C. M.; KAUS, M. R.; HAKER, S. J.; WELLS, W. M.; JOLESZ, F. A.; KIKINIS, R. 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>>. Citado na página 24.