

UNIVERSIDADE FEDERAL DO AMAZONAS
FACULDADE DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

MAKOTO MIYAGAWA

SEGMENTAÇÃO DO LÚMEN E IDENTIFICAÇÃO DE REGIÃO DE
BIFURCAÇÕES EM IMAGENS DE TOMOGRAFIA DE COERÊNCIA
ÓPTICA INTRAVASCULAR UTILIZANDO REDES NEURAIAS
CONVOLUTIVAS

MANAUS
2019

UNIVERSIDADE FEDERAL DO AMAZONAS
FACULDADE DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

MAKOTO MIYAGAWA

SEGMENTAÇÃO DO LÚMEN E IDENTIFICAÇÃO DE REGIÃO DE
BIFURCAÇÕES EM IMAGENS DE TOMOGRAFIA DE COERÊNCIA
ÓPTICA INTRAVASCULAR UTILIZANDO REDES NEURAIS
CONVOLUTIVAS

Dissertação apresentada ao Curso de Mestrado
em Engenharia Elétrica, área de concentração
Engenharia de Automação de Sistema do
Programa de Pós-Graduação em Engenharia
Elétrica da Universidade Federal do
Amazonas.

Orientador: Prof. Dr. Cícero Ferreira Fernandes Costa Filho
Co-Orientadora: Prof^a. Dr^a. Marly Guimarães Fernandes Costa

MANAUS
2019

Ficha Catalográfica

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

M685s Miyagawa, Makoto
Segmentação do lúmen e identificação de região de bifurcações em imagens de tomografia de coerência óptica intravascular utilizando redes neurais convolutivas / Makoto Miyagawa. 2019 88 f.: il. color; 31 cm.

Orientador: Cícero Ferreira Fernandes Costa Filho
Coorientadora: Marly Guimarães Fernandes Costa
Dissertação (Mestrado em Engenharia Elétrica) - Universidade Federal do Amazonas.

1. Doenças cardiovasculares. 2. Tomografia por coerência óptica intravascular. 3. Redes neurais convolutivas. 4. Segmentação do lúmen. 5. Detecção de bifurcação. I. Costa Filho, Cícero Ferreira Fernandes II. Universidade Federal do Amazonas III. Título

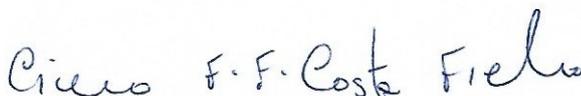
MAKOTO MIYAGAWA

**SEGMENTAÇÃO DO LÚMEN E IDENTIFICAÇÃO DE REGIÃO DE
BIFURCAÇÕES EM IMAGENS DE TOMOGRAFIA DE COERÊNCIA ÓPTICA
INTRAVASCULAR UTILIZANDO REDES NEURAIS CONVOLUTIVAS**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Amazonas, como requisito parcial para obtenção do título de Mestre em Engenharia Elétrica na área de concentração Controle e Automação de Sistemas.

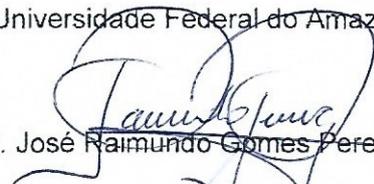
Aprovado em 30 de janeiro de 2019.

BANCA EXAMINADORA



Prof. Dr. Cícero Ferreira Fernandes Costa Filho, Presidente

Universidade Federal do Amazonas



Prof. Dr. José Raimundo Gomes Pereira, Membro

Universidade Federal do Amazonas



Prof. Dr. Marco Antônio Gutierrez, Membro

Universidade de São Paulo

AGRADECIMENTOS

Agradeço aos meus orientadores, Prof. Dr. Cícero Ferreira Fernandes Costa Filho e Prof^a. Dr^a. Marly Guimarães Fernandes Costa, pela orientação dada, pelo aprendizado adquirido e, principalmente, pela confiança em mim depositada.

Aos meus pais e irmão, pelo apoio e suporte constante, por estarem sempre ao lado nos bons e maus momentos.

À Thais Lemos de Mattos e família, que tem me acompanhado desde antes do mestrado, torcendo e apoiando de perto.

Aos meus amigos Bruno Magnani Degan e Sonny de Souza Costa Miranda, que juntos nós nos apoiamos e nos incentivamos a concluir nossos respectivos mestrados.

À Universidade Federal do Amazonas e, em especial, ao Centro de Tecnologia Eletrônica e da Informação – CETELI – pela concessão de toda infraestrutura para realização deste trabalho.

RESUMO

A utilização da tomografia por coerência óptica intravascular (IVOCT) permite que especialistas possam avaliar lesões coronarianas em alta resolução. A automatização de algumas etapas da análise poderia beneficiá-los, uma vez que a análise visual dos *frames* em um *pullback* é trabalhosa e consome muito tempo. Mesmo com a crescente popularidade das redes neurais convolutivas (CNN) na área médica, ainda há poucos trabalhos aplicados a imagens IVOCT para segmentação do lúmen e classificação de região de bifurcação. Neste trabalho, avaliamos três arquiteturas de CNN para a tarefa de segmentação do lúmen e quatro arquiteturas para a tarefa de classificação de região de bifurcação, utilizando um conjunto de imagens IVOCT de nove *pullbacks* de nove diferentes pacientes. Em relação à segmentação do lúmen, foram avaliadas redes diretas e de grafos acíclicos direcionados (DAG) em diferentes bases de dados, variando a resolução espacial, sistema de coordenadas e espaço de cores. Em relação à classificação de região de bifurcação, além das variações em sistema de coordenadas e espaço de cores nas bases de dados, foram utilizadas técnicas de *data augmentation* para balanceá-las, de forma a compensar a menor quantidade de imagens de bifurcação, além de utilizar transferência de conhecimento em algumas das redes avaliadas, aplicando a aprendizagem originada de uma das redes de segmentação. Nossos resultados são comparáveis aos outros trabalhos encontrados na literatura, apresentando, para a segmentação, melhores resultados em acurácia, coeficiente Dice e Jaccard acima de 99%, 98% e 97%, respectivamente. Na classificação, apresentou melhores resultados em *score* F1 (99,68%) e AUC (99,72%) obtidos por uma rede CNN com conhecimento transferido.

Palavras-chave: doenças cardiovasculares; tomografia por coerência óptica intravascular; redes neurais convolutivas; transferência de aprendizado; segmentação do lúmen; detecção de bifurcação.

ABSTRACT

Intravascular Optical Coherence Tomography (IVOCT) technology enables the experts to analyze coronary lesions from high-resolution images. Some level of automation could benefit experts since visual analysis of pullback frames is a laborious and time-consuming task. Even with the growing popularity of Convolutional Neural Networks (CNN) in the medical area, there are few works in the literature applying them to lumen segmentation and classification of bifurcation regions tasks. In this work, we evaluated three CNN architectures for the lumen segmentation task, and four architectures for bifurcation region classification, using an IVOCT image set of nine pullbacks from nine different patients. Regarding lumen segmentation task, direct networks and direct acyclic graph (DAG) networks were evaluated in different datasets, varying spatial resolution, coordinate systems, and color space. Regarding bifurcation region classification, besides variations in the coordinate systems and color space of the datasets, data augmentation techniques were used to balance them, in order to compensate for the smaller number of bifurcation images, besides using transfer learning in some of the evaluated networks, applying knowledge acquired from one of the segmentation networks. Our results are comparable to the other works found in the literature, presenting, for segmentation, best results in accuracy, Dice coefficient, and Jaccard over 99%, 98%, and 97%, respectively. In classification, better results were presented in F1 score (99,68%), and AUC (99,72%) obtained by a CNN with transferred knowledge.

Keywords: cardiovascular diseases; intravascular optical coherence tomography; convolutional neural networks; transfer learning; lumen segmentation; bifurcation detection.

LISTA DE ILUSTRAÇÕES

Figura 1 – Ilustração de uma artéria saudável e outra com estreitamento causado por uma placa aterosclerótica. Fonte: (NEW BEGINNINGS HEALTH CARE, 2014).	10
Figura 2 – Diagramas de uma artéria muscular, apresentando suas camadas. (a) Vista longitudinal da artéria; (b) Vista transversal. Fonte: Adaptado de (JUNQUEIRA e CARNEIRO, 2013).	11
Figura 3 – Ilustração de uma região de bifurcação, indicando regiões de alta e baixa tensão de cisalhamento (<i>high shear region</i> e <i>low shear region</i> , respectivamente). Fonte: (ZARINS, GIDDENS e GLAGOV, 1983).	12
Figura 4 – Elementos da artéria visíveis em imagens IVOCT. (a) lúmen (L), parede do vaso (W), túnica adventícia (AD) e vasa vasorum (V), (b) lâmina elástica interna (IEL), túnica média (M) e lâmina elástica externa (EEL). Fonte: adaptado de (ROLEDER et al., 2015).	14
Figura 5 – Exemplo de uma imagem IVOCT com uma bifurcação (<i>side branch</i>), o cateter-guia (<i>guide wire</i>), cateter de imageamento (<i>imaging catheter</i>), bainha protetora (<i>protective sheath</i>) e um <i>stent</i> (<i>metallic stent strut</i>) identificados. Fonte: adaptado de (WANG et al., 2014).	14
Figura 6 – Imagens IVOCT ilustrando desafios da segmentação automática do lúmen. (a) Irregularidades no lúmen causadas por placas e lesões (a1), stents e bifurcações (a2). (b) Sombra do cateter-guia. (c) Artefatos gerados por sangue no interior da artéria. Fonte: adaptado de (CAO et al., 2017).	18
Figura 7 – Exemplo de rede neural com uma camada de entrada com uma unidade, duas camadas ocultas com três unidades cada e uma camada de saída com duas unidades.	29
Figura 8 – Exemplo de convolução do <i>kernel</i> K sobre a imagem I	34
Figura 9 – Exemplo de <i>max pooling</i>	34
Figura 10 – Representação da convolução da Figura 8 em forma de matrizes.	35
Figura 11 – Convolução transposta da imagem C pelo kernel K representado por uma multiplicação de matrizes.	36
Figura 12 – Exemplo da ação da camada de <i>dropout</i> . A rede da direita é o resultado da aplicação do <i>dropout</i> na rede da esquerda. Fonte: (SRIVASTAVA et al., 2014).	36
Figura 13 – Exemplo de uma CNN para segmentação semântica. Fonte: (LONG, SHELHAMER e DARRELL, 2015).	38
Figura 14 – Arquitetura de segmentação semântica proposta por Noh, Hong e Han (2015). Fonte: (NOH, HONG e HAN, 2015).	38

Figura 15 – Quatro passos da fase de experimentação. 1) Criação das bases de dados; 2) Definição das redes utilizadas; 3) Configuração do roteiro de experimentação; 4) Configuração do treino e teste.....	41
Figura 16 – Processo de pré-processamento das imagens.....	43
Figura 17 – Processo de criação das bases de dados para os experimentos de segmentação do lúmen.	43
Figura 18 – Processo de criação do padrão-ouro das bases de dados para os experimentos de segmentação do lúmen.....	44
Figura 19 – Processo de criação da base de dados para a classificação de região de bifurcação. (a) As imagens de bifurcação são rotacionadas em incrementos de 45 graus cada. (b) Quatro bases definidas com diferentes sistemas de coordenadas e espaço de cores.	45
Figura 20 – Arquitetura das três redes convolutivas utilizadas na segmentação do lúmen.....	48
Figura 21 – Arquitetura das quatro redes convolutivas utilizadas na classificação de região de bifurcação. CNN-C2, CNN-C3 e CNN-C4 utilizam técnicas de <i>transfer learning</i> para auxiliar no desempenho da rede.....	50
Figura 22 – Validação cruzada em cinco partições	54
Figura 23 – Diagrama de Venn ilustrando os possíveis resultados entre o padrão-ouro e o que foi predito.	54
Figura 24 – Comparativo dos experimentos de classificação de região de bifurcação em relação a cada métrica. (a) Acurácia, (b) Sensibilidade, (c) Especificidade, (d) Coeficiente Dice, (e) Coeficiente Jaccard e (f) Score F1.....	59
Figura 25 – O resultado da segmentação do lúmen de uma imagem IVOCT em cada experimento. A linha verde representa o padrão-ouro e a área azul representa o resultado da segmentação.....	62
Figura 26 – Amostras de falsos negativos (a-b) e falsos positivos (c-e) da rede treinada no Experimento C4.....	65
Figura 27 – Amostras de falso negativo (a) e falsos positivos (b-c) da rede treinada no Experimento C9. A área destacada em verde mostra a porção da imagem localizada sobre a borda ao convertê-la em coordenadas polares.	66
Figura 28 – Curvas ROC dos experimentos de classificação de região de bifurcação. (a-b) Desempenho dos classificadores com base de dados baseados em coordenadas cartesianas. (c-d) Desempenho dos classificadores com base de dados baseados em coordenadas polares. (e-f) Desempenho dos classificadores com rede CNN-S4.	69

Figura 29 – Comparativo dos experimentos de classificação de região de bifurcação em relação a cada métrica. (a) Taxa de falso positivo, (b) Sensibilidade, (c) Acurácia, (d) Especificidade, (e) Precisão, (f) Taxa de alarme falso, (g) *Score* F1 e (h) Área sobre a curva ROC. 70

LISTA DE TABELAS

Tabela 1 – Características do ultrassom intravascular e da tomográfica de coerência óptica intravascular. Fonte: Adaptado de (RATHOD et al., 2015).....	13
Tabela 2 – Bases de dados utilizados nas experimentações de segmentação do lúmen.....	44
Tabela 3 – Bases de dados utilizados nas experimentações de classificação de região de bifurcação	45
Tabela 4 – Parâmetros utilizados na etapa de treinamento em todos os experimentos	51
Tabela 5 – Descrição dos 21 experimentos conduzidos para a segmentação do lúmen e o tempo decorrido para o treinamento de cada um deles.	52
Tabela 6 – Descrição dos dez experimentos conduzidos para a classificação de região de bifurcação e o tempo decorrido para o treinamento de cada um deles.....	53
Tabela 7 – Resultados obtidos na segmentação do lúmen da base de dados BD-S1.	57
Tabela 8 – Resultados obtidos na segmentação do lúmen da base de dados BD-S2.	57
Tabela 9 – Resultados obtidos na segmentação do lúmen da base de dados BD-S3.	57
Tabela 10 – Resultados obtidos na segmentação do lúmen da base de dados BD-S4.	57
Tabela 11 – Resultados obtidos na segmentação do lúmen da base de dados BD-S5.	57
Tabela 12 – Resultados obtidos na segmentação do lúmen da base de dados BD-S6.	58
Tabela 13 – Resultados obtidos na segmentação do lúmen da base de dados BD-S7.	58
Tabela 14 – Comparativo dos resultados de trabalhos anteriores de segmentação do lúmen com o deste trabalho.	60
Tabela 15 – Tempo médio de processamento por imagem nas redes de cada um dos experimentos na tarefa de segmentação.	63
Tabela 16 – Resultados obtidos na classificação de região de bifurcação com bases de dados baseados em coordenadas cartesianas (BD-C1 e BC-C2).....	64
Tabela 17 – Resultados obtidos na classificação de região de bifurcação com bases de dados baseados em coordenadas polares (BD-C3 e BC-C4).....	65
Tabela 18 – Comparativo dos resultados entre trabalhos anteriores de classificação de região de bifurcação com o deste trabalho	67
Tabela 19 – Tempo médio de processamento por imagem nas redes de cada um dos experimentos na tarefa de classificação.....	67

LISTA DE QUADROS

Quadro 1 – Síntese da revisão bibliográfica da segmentação automática do lúmen em imagens IVOCT	24
Quadro 2 – Síntese da revisão bibliográfica da classificação de região de bifurcação em imagens intravasculares.	28
Quadro 3 – Métricas utilizadas na avaliação da segmentação do lúmen.....	55
Quadro 4 – Métricas utilizadas na avaliação da classificação de região de bifurcação	56

LISTA DE SIGLAS

ACC	Acurácia
AUC	Área sob a curva ROC
CNN	Redes Neurais Convolutivas
D	Coefficiente Dice
DAG	Grafos Acíclicos Direcionados
DICOM	<i>Digital Imaging and Communications in Medicine</i>
DWPF	Discrete Wavelet Packet Frame
F1	Score F ₁
FA	Taxa de alarme falso
FN	Falso negativo
FNAF	Fração da área falso negativo
FP	Falso positivo
FPR	Taxa de falso positivo
FPS	Quadros por segundo
FROLS	<i>Forward Regression with Orthogonal Least Squares</i>
GAN	Redes Adversárias Generativas
GRBF	Função de base radial gaussiana
ICC	Correlação intra-classe
InCor	Instituto do Coração
IVOCT	Tomografia por Coerência Óptica Intravascular
IVUS	Ultrassom Intravascular
J	Coefficiente Jaccard
LBP	Padrões binários locais
LDA	Análise de Discriminantes Lineares

MADA	Diferença da área absoluta média
MFROLS	<i>Multiple FROLS</i>
MSE	Erro quadrático médio
MSSL	<i>Multiscale Stacked Sequential Learning</i>
ORL	Mínimos quadrados ortogonal
PCA	Análise de Componentes Principais
PNG	<i>Portable Network Graphics</i>
PRE	Precisão
ReLU	Unidade Linear Retificadora
RMS	<i>Root mean square</i>
RMSSSD	<i>RMS symmetric surface distance</i>
RNA	Redes Neurais Artificiais
ROC	<i>Receiver Operating Characteristic</i>
SEN	Sensibilidade
SFS	Seleção escalar de características
SGD	Gradiente descendente estocástico
SPE	Especificidade
SVM	Máquina de vetores de suporte
TN	Verdadeiro negativo
TNAF	Fração da área verdadeiro negativo
TP	Verdadeiro positivo
TPAF	Fração da área verdadeiro positivo
TPR	Taxa de verdadeiro positivo
VOE	<i>Volume overlap error</i>

SUMÁRIO

1	INTRODUÇÃO	10
1.1	OBJETIVO GERAL.....	15
1.2	OBJETIVOS ESPECÍFICOS	15
1.3	ORGANIZAÇÃO DO TRABALHO	16
2	REVISÃO BIBLIOGRÁFICA	18
2.1	SEGMENTAÇÃO AUTOMÁTICA DO LÚMEN EM IMAGENS IVOCT	19
2.2	IDENTIFICAÇÃO DE REGIÕES DE BIFURCAÇÃO EM IMAGENS INTRAVASCULARES.....	25
3	REFERENCIAL TEÓRICO	29
3.1	REDES NEURAIIS ARTIFICIAIS	29
3.1.1	Funções de ativações.....	30
3.1.2	Treinamento da rede.....	30
3.2	REDES NEURAIIS CONVOLUTIVAS	31
3.2.1	Camada convolutiva.....	33
3.2.2	Unidades retificadoras lineares (ReLU).....	34
3.2.3	Camada de <i>pooling</i>	34
3.2.4	Convolução transposta	34
3.2.5	Camada de <i>dropout</i>	36
3.2.6	Camada de normalização em lote (<i>batch normalization</i>)	36
3.2.7	Segmentação semântica	37
3.2.8	Aumento de dados (<i>Data augmentation</i>)	39
3.2.9	Transferência de aprendizado (<i>transfer learning</i>).....	39
4	MATERIAS E MÉTODOS	41
4.1	CONJUNTO DE DADOS	42
4.1.1	Conjunto de dados para a segmentação do lúmen	43
4.1.2	Conjunto de dados para classificação de região de bifurcação	44
4.2	PROPOSTAS DAS REDES NEURAIIS CONVOLUTIVAS.....	45
4.2.1	Arquiteturas propostas para a segmentação do lúmen	45
4.2.2	Arquitetura proposta para a classificação de região de bifurcação	48
4.3	EXPERIMENTOS	51

4.3.1	Divisão da base para treinamento e teste	53
4.4	MÉTRICAS DE DESEMPENHO	54
4.4.1	Métricas utilizadas na segmentação do lúmen	54
4.4.2	Métricas utilizadas na classificação de região de bifurcação	55
5	RESULTADOS E DISCUSSÕES	57
5.1	RESULTADOS DA SEGMENTAÇÃO DO LÚMEN.....	57
5.2	RESULTADOS DA CLASSIFICAÇÃO DE REGIÃO DE BIFURCAÇÃO	63
6	CONCLUSÕES	71
	REFERÊNCIAS.....	74
7	APÊNDICE I	79

1 INTRODUÇÃO

De acordo com a Organização Mundial de Saúde, doenças cardiovasculares são a principal causa de mortes atualmente, correspondendo a 31% das mortes em todo o mundo em 2015 (WHO, 2015). A aterosclerose, um dos principais causadores de derrames e infartos, consiste no acúmulo de colesterol nas paredes das artérias, ocasionando o estreitamento da passagem do sangue e lesões (ateromas), podendo ocasionar uma obstrução total dos vasos (MENDIS, PUSKA e NORRVING, 2011). A Figura 1 ilustra uma artéria saudável e outra com estreitamento causado por uma placa de ateroma.

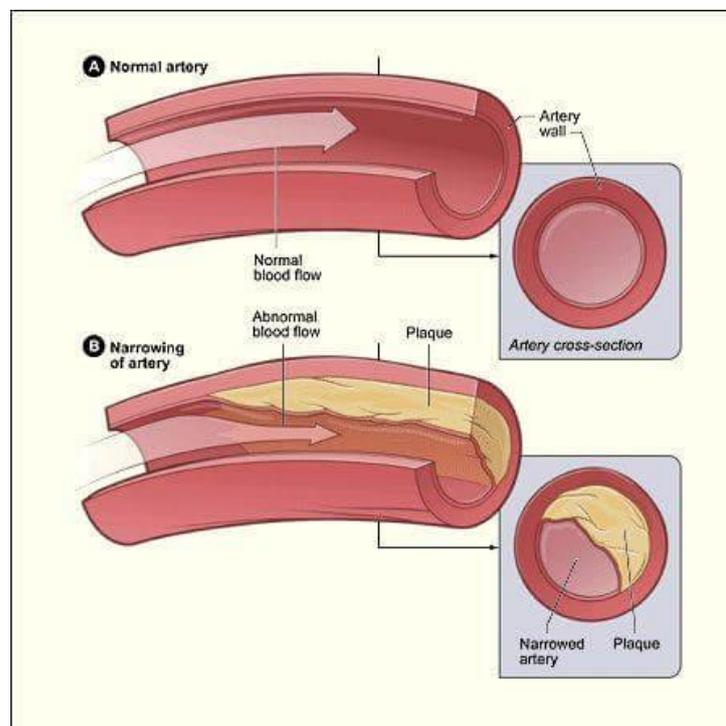


Figura 1 – Ilustração de uma artéria saudável e outra com estreitamento causado por uma placa aterosclerótica.
Fonte: (NEW BEGINNINGS HEALTH CARE, 2014).

A estrutura de um vaso sanguíneo é, em geral, dotada de três camadas denominadas túnicas, apresentados na Figura 2, e a cavidade interna por onde sangue circula é chamada de luz ou lúmen. As túnicas são caracterizadas da seguinte maneira (JUNQUEIRA e CARNEIRO, 2013):

- **Túnica íntima:** Camada mais interna do vaso constituída por duas camadas de tecido: uma camada epitelial que circunda o lúmen (área interna do vaso) e uma

camada subendotelial, que pode conter células musculares lisas. Em artérias, uma lâmina elástica interna separa a túnica íntima da túnica média;

- **Túnica média:** Camada intermediária, consiste principalmente de camadas de células musculares lisas em artérias musculares, enquanto em artérias elásticas é formada por camadas de músculo liso intercaladas por lâminas elásticas. Em artérias musculares menos calibradas, uma lâmina elástica externa separa a túnica média com a túnica adventícia;
- **Túnica adventícia:** Camada externa, consiste principalmente de fibras elásticas e colágeno, fundindo-se com o tecido conjuntivo do órgão pelo qual o vaso está passando. Em vasos grandes, normalmente estão presentes arteríolas, capilares e vênulas chamadas de *vasa vasorum*.

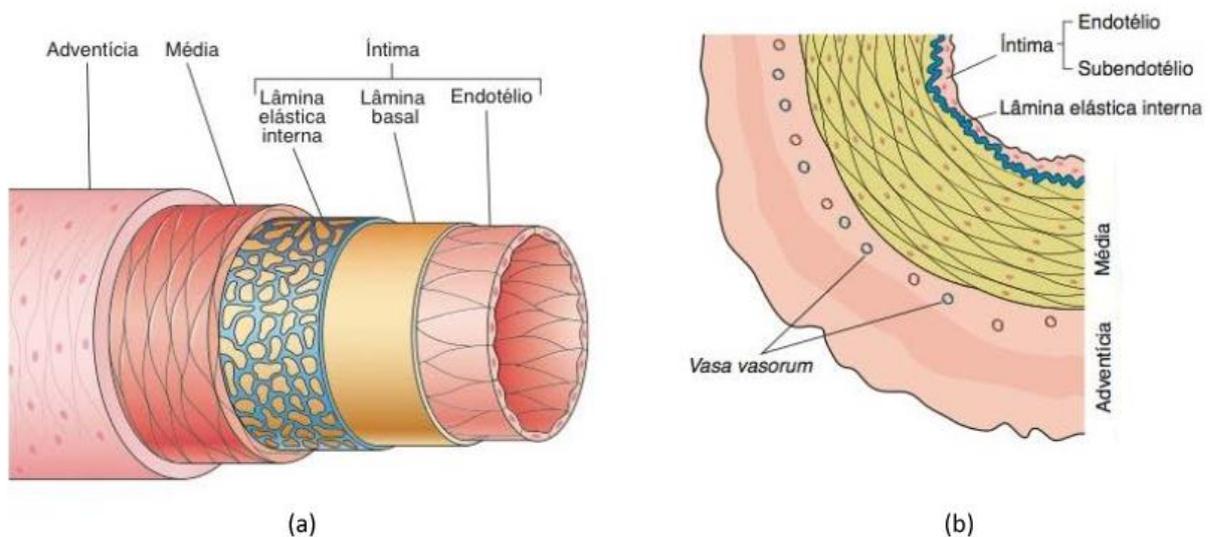


Figura 2 – Diagramas de uma artéria muscular, apresentando suas camadas. (a) Vista longitudinal da artéria; (b) Vista transversal. Fonte: Adaptado de (JUNQUEIRA e CARNEIRO, 2013).

Lesões ateroscleróticas geralmente ocorrem na túnica íntima e são categorizados em três estágios (SURI, KATHURIA e MORILARI, 2011):

- **Estria lipídica:** Estágio assintomático ocasionado principalmente por lipídeos e macrófagos. Neste primeiro momento ocasiona o estreitamento da artéria;

- **Placa fibrosa:** Consiste em uma camada geralmente localizada nos mesmos lugares que as estrias lipídicas, formada por células musculares lisas cobertas, lipídeos e macrófagos cobertos por uma capa fibrosa. A placa cresce em direção ao lúmen, causando estreitamento da artéria;
- **Lesão complicada:** Surge por conta de necrose, calcificação, ulceração na superfície e hemorragia de pequenos vasos que cresceram na lesão. Embora as lesões deste estágio sejam oclusivas, o maior risco está na ruptura da lesão ocasionando a formação do trombo, podendo causar infartos e derrames.

Além disso, há uma maior incidência de placas e de espessamento das paredes das artérias em áreas das principais bifurcações e segmentos de curvas, pois são regiões de separação do fluxo, turbulência e variações na velocidade do fluxo e na tensão de cisalhamento (ZARINS et al., 1983). O maior acúmulo de placas em regiões de bifurcações geralmente ocorre nas paredes laterais à bifurcação, onde a tensão de cisalhamento é menor, enquanto um menor acúmulo é presenciado no ponto de bifurcação, onde a tensão é maior. A Figura 3 ilustra o fluxo em uma região de bifurcação, indicando as regiões de menor tensão de cisalhamento (*low shear region*) e de maior tensão (*high shear region*).

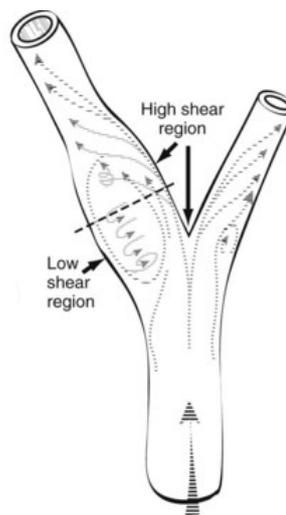


Figura 3 – Ilustração de uma região de bifurcação, indicando regiões de alta e baixa tensão de cisalhamento (*high shear region* e *low shear region*, respectivamente). Fonte: (ZARINS, GIDDENS e GLAGOV, 1983).

Para o diagnóstico e tratamento dessas lesões, métodos não invasivos, como a ressonância magnética, ou invasivos, como a angiografia, o ultrassom intravascular (*Intravascular Ultrasound*, IVUS) e a tomografia de coerência óptica intravascular (*Intravascular Optical Coherence Tomography*, IVOCT), oferecem a possibilidade de analisar as artérias em busca de anomalias. O IVOCT destaca-se entre as modalidades intravasculares pela maior resolução e a capacidade de detectar alguns tipos de lesões que não podem ser vistos por IVUS como, por exemplo, hiperplasia intimal, trombos intraluminais e fibroateroma (RATHOD et al., 2015).

A tomografia de coerência óptica (OCT) utiliza-se de luzes próximas ao infravermelho para a aquisição de imagens de alta resolução (12 a 15 μm) se comparado a imagens de ultrassom (100 a 150 μm). Dessa forma, a OCT permite visualizar detalhes de estruturas não visíveis por ultrassom como características de placas coronarianas, de forma a distinguir placas calcificadas, fibrosas ou ricas em lipídeos, por exemplo (RATHOD et al., 2015). Porém a profundidade de penetração da OCT é limitada, alcançando 1 a 2 mm contra 4 a 8 mm do ultrassom. Além disso, a luz infravermelho não penetra no sangue, prejudicando a aquisição da imagem por OCT, sendo necessário a retirada do sangue, seja por interrupção do fluxo por um cateter-balão ou por infusão de uma solução viscosa isosmolar (PRATI et al., 2011). A Tabela 1 apresenta as principais características do IVUS e IVOCT.

Tabela 1 – Características do ultrassom intravascular e da tomográfica de coerência óptica intravascular. Fonte: Adaptado de (RATHOD et al., 2015).

	IVUS	IVOCT
Tecnologia	Ultrassom	Próximo do infravermelho
Resolução axial (μm)	100–150	12–15
Resolução lateral (μm)	150–300	19
<i>Frame rate</i> (fps)	30	100
Velocidade de <i>pullback</i> (mm/s)	0,5–2,0	10–15
Penetração do tecido (mm)	4–8	1–2
Imagem através do sangue	Sim	Não
Remoção do sangue por contraste	Não	Sim
Comprimento de onda	1,3 μm	10–40MHz

Na Figura 4, é possível visualizar na imagem IVOCT os vários elementos de uma artéria: o lúmen (L), a parede do vaso (W) composta pela a lâmina elástica interna (IEL), túnica média (M), lâmina elástica externa (EEL), túnica adventícia (AD) e o *vasa vasorum* (V). Na Figura 5 está identificado outros elementos como uma bifurcação (*side branch*), o cateter de imagem (*imaging catheter*), a bainha protetora (*protective sheath*) e o cateter-guia (*guide wire*) do instrumento e um *stent* (*metallic stent strut*).

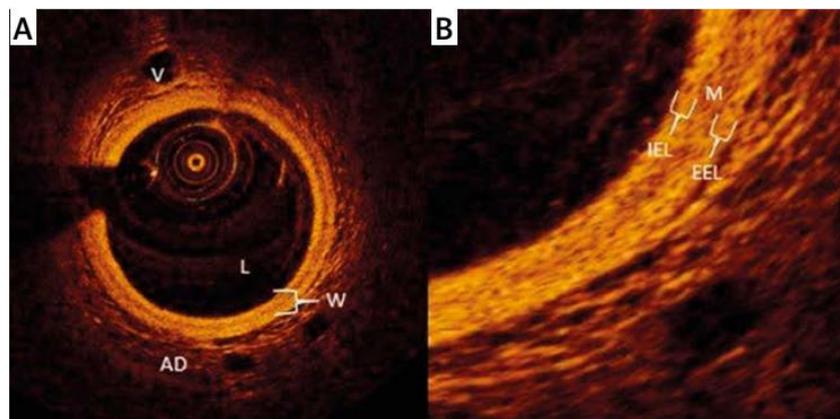


Figura 4 – Elementos da artéria visíveis em imagens IVOCT. (a) lúmen (L), parede do vaso (W), túnica adventícia (AD) e *vasa vasorum* (V), (b) lâmina elástica interna (IEL), túnica média (M) e lâmina elástica externa (EEL). Fonte: adaptado de (ROLEDER et al., 2015).

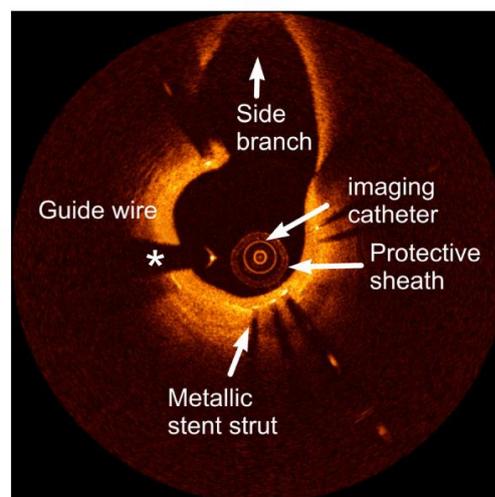


Figura 5 – Exemplo de uma imagem IVOCT com uma bifurcação (*side branch*), o cateter-guia (*guide wire*), cateter de imageamento (*imaging catheter*), bainha protetora (*protective sheath*) e um *stent* (*metallic stent strut*) identificados. Fonte: adaptado de (WANG et al., 2014).

Em técnicas recentes, onde dispensa-se a utilização de cateter-balão para bloquear a passagem de sangue, o processo de aquisição de imagens em um exame IVOCT ocorre da seguinte forma: o cateter-guia é inserido na artéria até a região de interesse, onde é então substituído pelo cateter de imagem e o processo de *pullback* (operação em que as imagens são capturadas enquanto o cateter é puxado de dentro para fora da artéria) é feito na maior velocidade disponível enquanto é injetado uma solução isosmolar viscosa para remover o sangue da área de imagem (PRATI et al., 2011). Neste processo, o cateter é puxado de volta numa velocidade de 10 a 15 mm/s, gerando cerca de 100 imagens por segundo.

Portanto, em um único exame IVOCT são geradas centenas de imagens e o tempo consumido para analisá-las torna o processo demorado e cansativo para o profissional, além destas análises estarem sujeitas a variabilidade intra-observador (CAO et al., 2017). Assim, a utilização de ferramentas que aplicam técnicas de processamento de imagens pode auxiliá-los no diagnóstico preciso. Atualmente, a utilização de redes neurais convolutivas (CNN) vem se destacando em diversas aplicações relacionados à imagens na área médica, como na segmentação de estruturas neuronais (RONNEBERGER, FISCHER e BROX, 2015) e na classificação de câncer de pele (ESTEVA et al., 2017). Portanto, este trabalho busca contribuir com o estudo em CNNs para a segmentação do lúmen e identificação de regiões de bifurcação em imagens IVOCT.

1.1 OBJETIVO GERAL

Propor e implementar arquiteturas de redes neurais convolutivas para a segmentação automática do lúmen e identificação de região de bifurcação em imagens de tomografia de coerência óptica intravascular.

1.2 OBJETIVOS ESPECÍFICOS

1. Propor arquiteturas de redes neurais convolutivas diretas e de grafos acíclicos direcionados na tarefa de segmentação do lúmen em imagens IVOCT;

2. Analisar a influência de características como tamanho, espaço de cores e sistema de coordenadas da base de dados na tarefa de segmentação do lúmen em imagens IVOCT;
3. Propor arquiteturas de redes neurais convolutivas diretas na tarefa de identificação de bifurcações em imagens IVOCT;
4. Analisar a influência de características como tamanho, espaço de cores e sistema de coordenadas da base de dados na tarefa de identificação de região de bifurcação em imagens IVOCT;
5. Analisar a influência da transferência de aprendizagem na tarefa de identificação de região de bifurcação em imagens IVOCT;
6. Avaliar o desempenho das arquiteturas propostas e suas parametrizações utilizando métricas quantitativas;
7. Comparar os resultados com os métodos presentes na literatura.

1.3 ORGANIZAÇÃO DO TRABALHO

A organização do trabalho segue as divisões a seguir:

- Introdução (Capítulo 1);
- Revisão Bibliográfica (Capítulo 2);
- Referencial Teórico (Capítulo 3);
- Materiais e Métodos (Capítulo 4);
- Resultados e Discussões (Capítulo 5);
- Conclusão (Capítulo 6);
- Apêndice I.

O capítulo 2 apresenta trabalhos desenvolvidos recentemente na área de segmentação do lúmen e de classificação de bifurcações em imagens IVOCT, utilizando tanto por técnicas de processamento de imagens quanto por reconhecimento de padrões.

No capítulo 3, são apresentados fundamentos das CNNs: seu modo de funcionamento, características das principais camadas utilizadas nas redes, aspectos das arquiteturas utilizadas na tarefa de segmentação de imagens em geral e a utilização de transferência de conhecimento em CNNs.

O capítulo 4 aborda os materiais e métodos utilizados neste trabalho. Descreve as características das bases de dados utilizadas, a configuração das arquiteturas das redes escolhidas para os experimentos e suas parametrizações, assim como a forma em que foi conduzida o processo de treinamento das redes.

O capítulo 5 apresenta as métricas utilizadas para avaliar o desempenho das redes, os resultados obtidos e suas discussões, comparando com outros trabalhos presentes na literatura e o capítulo 6 apresenta as conclusões acerca deste trabalho.

Por fim, no apêndice I encontra-se uma cópia do artigo “*Lumen Segmentation in Optical Coherence Tomography Images using Convolutional Neural Network*”, de Makoto Miyagawa, Marly Guimarães Fernandes Costa, Marco Antônio Gutierrez, João Pedro C. F. Costa e Cícero Ferreira Fernandes Costa Filho e apresentado no *40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, de 17 a 21 de julho de 2018 em Honolulu, Haváí, Estados Unidos.

2 REVISÃO BIBLIOGRÁFICA

Em um exame IVOCT, são geradas centenas de imagens da seção transversal da artéria a cada *pullback*, que serão posteriormente analisadas por um profissional a fim de encontrar pontos de interesse para o diagnóstico de doenças cardiovasculares. A segmentação do lúmen, por exemplo, é um passo importante para o processo de detecção de tais pontos de interesse, porém é uma tarefa que pode consumir horas de um profissional se feita manualmente, além da possibilidade de haver problemas de variabilidade intra-observador (CAO et al., 2017).

Portanto, a segmentação automática do lúmen é desejável, mas é desafiadora em três aspectos. Como ilustra a Figura 6, o primeiro se refere ao formato irregular do lúmen quando este possui placas e lesões (a1), *stents*, ou quando a região é uma bifurcação (a2). O segundo é relacionado à sombra do cateter-guia (b), que gera uma descontinuidade do lúmen na imagem. Por fim, o terceiro se refere a artefatos gerados na imagem por conta do sangue no interior da artéria (c) (CAO et al., 2017).

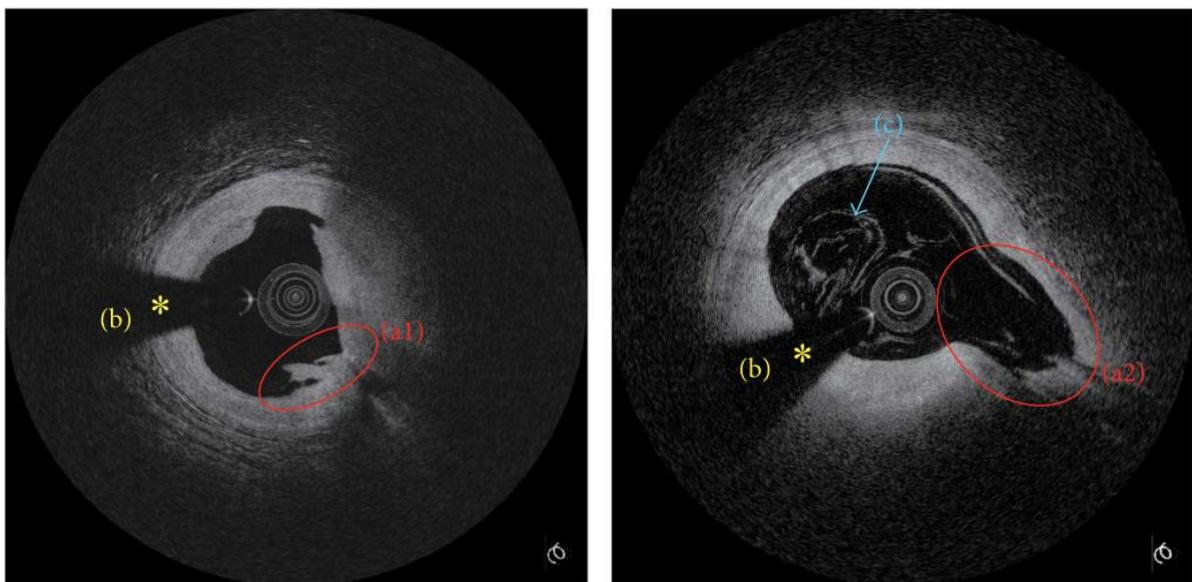


Figura 6 – Imagens IVOCT ilustrando desafios da segmentação automática do lúmen. (a) Irregularidades no lúmen causadas por placas e lesões (a1), *stents* e bifurcações (a2). (b) Sombra do cateter-guia. (c) Artefatos gerados por sangue no interior da artéria. Fonte: adaptado de (CAO et al., 2017).

Esta seção busca discorrer sobre as principais técnicas descritas na literatura no que refere a segmentação automática do lúmen em imagens IVOCT, bem como os seus resultados.

2.1 SEGMENTAÇÃO AUTOMÁTICA DO LÚMEN EM IMAGENS IVOCT

O trabalho de Tsantis et al. (2012) aborda detecção de *stents*, mas eles descrevem o método utilizado para a segmentação do lúmen. No pré-processamento, a bainha é detectada e removida da imagem através da Transformada de Hough e, utilizando Campo Aleatório de Markov (MRF), eles combinaram informações condicionais (distribuição de intensidade local) com contextuais (similaridade de intensidade dentro da vizinhança) para segmentar o lúmen. 2710 imagens IVOCT em coordenadas cartesianas foram utilizadas no experimento, resultando no Coeficiente Jaccard de $93,7\% \pm 4,5\%$.

Moraes, Cardenas e Furuie (2013) utilizam uma base de dados diferenciada dos demais por conter 290 imagens IVOCT de dois pacientes, dois porcos e um coelho, para representar uma variedade de condições como irregularidades no lúmen e diferentes graus de contraste na parede dos vasos. O trabalho utiliza *Discrete Wavelet Packet Frame* (DWPF) para detectar a área do tecido, selecionando o *wavelet* e o coeficiente de decomposição com maior correlação com as informações do tecido. No pré-processamento, é feita a remoção do cateter da imagem e, com um filtro de mediana, atenuam as marcas de alinhamento e ruídos, além de converter a imagem para coordenadas polares. Depois, o lúmen é localizado através da extração das características por DWPF e a imagem é binarizada através do método de Otsu. Por fim, são feitas correções na imagem com operações morfológicas, a sombra do cateter-guia é detectada e removida e a imagem é convertida de volta para coordenadas cartesianas. As métricas utilizadas foram a fração da área verdadeiro positivo (TPAF), fração da área verdadeiro negativo (TNAF), fração da área falso negativo (FNAF), Coeficiente Jaccard e Dice com resultados $99,29\% \pm 2,96\%$, $3,69\% \pm 2,88\%$, $0,71\% \pm 2,96\%$, $95,4\% \pm 4,8\%$ e $97,8\% \pm 2,16\%$, respectivamente.

Kim et al. (2014) apresenta uma abordagem que une resultados de duas técnicas para maior precisão da segmentação: binarização por Otsu e aproximação por curvas de intensidade.

Primeiramente, uma fase de pré-processamento reduz as imagens de 1024×1024 para 512×512 *pixels* em escala de cinza, são removidos os pixels do anel do cateter e um filtro de mediana remove as marcas de alinhamento. Por fim, a imagem é convertida para coordenadas polares para melhorar a eficiência na detecção do lúmen. Na primeira técnica, a imagem é dividida em quatro segmentos e o algoritmo de Otsu é utilizado em cada um deles. Depois, operações morfológicas reduzem os ruídos e refinam a imagem. A segunda técnica divide a imagem em 64 segmentos e cada um deles são analisados por curvas de intensidade, determinando posição do lúmen conforme o valor de intensidade. O lúmen nas regiões de descontinuidade é estimado por interpolação. Por fim, os dois resultados são unidos através de um processo de correção de posição dos pontos encontrados, resultando no contorno do lúmen. Neste trabalho, os autores utilizaram como métrica TPAF, TNAF, FNAF, especificidade e sensibilidade, resultando em $99,21\% \pm 0,21\%$, $0,30\% \pm 0,15\%$, $0,79\% \pm 0,51\%$, $99,2\%$ e $99,7\%$, respectivamente.

Xu et al. (2015) categoriza as imagens em cinco grupos, conforme as características das lesões presentes na imagem: normal (sem lesão), placa fibrosa, fibroateroma, ruptura da placa e placa fibrocalcificante. Sua base de dados possui 500 imagens, onde cada grupo está representado em quantidades iguais. A segmentação é feita através de uma representação da imagem em forma de grafos, onde os vértices (*pixels*) são conectados através de arestas com pesos, calculados com base em gradiente de intensidade. Utilizando algum algoritmo de caminho ótimo como o de Dijkstra, a borda do lúmen é definida como o caminho de menor peso. Um algoritmo de detecção do cateter-guia, que utiliza transformada de Hough, é utilizado para que a região do lúmen oculta pela sombra seja estimada através de uma linha reta. Como métrica, utilizaram o coeficiente Dice em comparação a padrões-ouro definidos por dois especialistas, E1 e E2, obtendo, dentre as variações do experimento, o melhor resultado de $94,31\%$ para E1 e $94,46\%$ para E2.

Macedo et al. (2015) volta-se para o problema de identificação de bifurcações, onde utiliza diversas técnicas de aprendizado de máquina juntamente com extratores de características. Dentre o método proposto, há a proposta de um algoritmo de segmentação do lúmen. No trabalho, utilizou-se imagens IVOCT de 9 *pullbacks* de 9 pacientes diferentes. Dessas imagens, foram geradas três bases de dados, onde a primeira base é composta de imagens ausentes de anomalias tais como lesões complexas, presença de sangue residual no lúmen e dissecções. A segunda base é composta pelas imagens da primeira base acrescentadas de imagens com lesões complexas e dissecções, e a terceira é composta da segunda base com as outras anomalias que não foram incluídas anteriormente. Ao todo, foram utilizadas 1460 imagens, sendo 1282 sem bifurcação e 178 com bifurcação.

Na etapa de pré-processamento, as imagens foram convertidas para coordenadas polares, houve a remoção do cateter, filtragem de ruídos por filtro bilateral, limiarização por Otsu e remoção do cateter-guia. Através da técnica por varredura por linha, preenche-se a região da íntima com o lúmen e esta imagem é subtraída da imagem que contém apenas a região da íntima (da última etapa do pré-processamento), de forma a identificar a região do lúmen. Operações morfológicas são utilizadas para corrigir buracos e outras anomalias na imagem e também é localizado e preenchido a lacuna gerada pela sombra do cateter-guia. Os resultados da segmentação da terceira base de dados resultou em *volume overlap error (VOE)* = $95,1 \pm 4,41$ %, coeficiente Dice = $97,4 \pm 2,67$ %, *RMS symmetric surface distance (RMSSSD)* = $0,04 \pm 0,04$ mm, diferença de área absoluta média (MADA) = $0,13 \pm 0,25$ mm², distância de Hausdorff = $0,17 \pm 0,22$ mm e acurácia = $99,88 \pm 0,25$ %.

Roy et al. (2016) explora o problema através da representação da imagem através de grafos, incorporando informações de atenuação de intensidade dos fótons em imagens de tomografia. O grafo é representado por uma matriz laplaciana e a região segmentada é determinada pela sua solução ótima. Para isso, utilizaram o algoritmo de passeio aleatório com

sementes na região entre o lúmen e a túnica íntima determinadas através de um modelo de *optical backscattering* nas imagens utilizadas, de dimensão 512×512 *pixels* em coordenadas cartesianas. As métricas utilizadas na segmentação foram coeficiente Kappa de Cohen ($97,86\% \pm 0,61\%$) e raiz do erro quadrático médio do contorno ($8,4655 \pm 3,8540$).

Wang e Tang (2016) segmentam a imagem utilizando um algoritmo de programação dinâmica. Na etapa de pré-processamento são feitas as detecções e remoções de alguns elementos como o cateter, a sombra do cateter-guia e o *stent*. Para a detecção do lúmen, um filtro de mediana e de borda são utilizados e o gradiente da imagem é tratado com uma matriz de custo, onde o menor caminho, pode ser encontrado dinamicamente. A sombra do cateter-guia é detectada ao analisar os níveis de intensidade e a borda do lúmen ofuscada por essa área é determinado por interpolação.

O enfoque de Menguy et al. (2016) está na detecção de *stents*, mas descreve o uso do algoritmo *Fast Marching* para segmentação do lúmen. Na etapa de pré-processamento, o cateter é removido e o ruído é suavizado das imagens em coordenadas polares, e então, o algoritmo *Fast Marching*, encarrega-se de segmentar a região através da busca, de dentro para fora do lúmen, da borda de transição entre o lúmen e o tecido. Os resultados preliminares deles indicam que a segmentação ainda não é satisfatória na área do cateter-guia e em bifurcações da artéria, com Coeficiente de Tanimoto resultando em 90,5%.

Cao et al. (2017) propõem uma abordagem utilizando *level set* para a segmentação do lúmen. Primeiramente, a imagem é convertida para coordenadas polares através da Transformada de Hough, a sombra do cateter-guia é removida e a imagem é suavizada por um filtro gaussiano estreito $N \times 1$ proposto no artigo, evitando problemas na segmentação se comparado a um filtro gaussiano tradicional $N \times N$. Após a suavização, aplica-se o *level set* para segmentar o lúmen, a área do cateter-guia é separado novamente e, com a fronteira determinada pelo *level set*, é possível propor uma curva polinomial de ordem M a partir de K pontos de cada

lado do cateter-guia. O experimento foi conduzido em 880 imagens de 5 pacientes, onde o coeficiente Dice médio resultante foi de $98,1\% \pm 1,1\%$.

Mesmo com a crescente utilização de redes convolutivas na área médica, ainda não há uma quantidade relevante de trabalhos referentes a segmentação do lúmen em imagens IVOCT, grande parte devido ao fato das bases de dados utilizadas serem privadas, ao contrário de aplicações envolvendo análises de melanomas e do ventrículo esquerdo, por exemplo, que possuem bases de acesso públicas (ISIC, 2018; RADAU et al., 2009).

Dentre os trabalhos revisados, apenas Yong et al. (2017) utiliza redes convolutivas, baseado em resultados positivos obtidos na segmentação do ventrículo esquerdo (TAN et al., 2017). O trabalho utiliza uma abordagem em que a rede é treinada para realizar uma regressão linear, inferindo a posição radial de 100 pontos equidistantes que determinam a borda do lúmen na imagem. Fora a subamostragem, que reduziu as imagens de 984×496 para 488×248 *pixels*, não houve nenhum pré ou pós-processamento adicional, pois espera-se que estes comportamentos surjam de forma implícita na rede durante a etapa de treinamento. Os resultados foram $20,87\mu\text{m}$, $98,5\%$ e $97,0\%$ para o erro médio absoluto, Coeficiente Dice e Coeficiente Jaccard, respectivamente.

O Quadro 1 contém os principais pontos de cada referência abordada, sintetizando esta seção.

Quadro 1 – Síntese da revisão bibliográfica da segmentação automática do lúmen em imagens IVOCT

Autores	Base de dados	Sistema de coordenadas	Pré-processamento	Técnica Utilizada	Métricas	Resultados
Tsantis et al. (2012)	10 <i>pullbacks in-vivo</i> com 271 <i>frames</i> cada	Cartesiano	Transformada de Hough para remoção de artefatos	Campo aleatório de Markov (MRF)	Coefficiente Jaccard	93,7% ± 4,5%
Moraes, Cardenas e Furuie (2013)	290 imagens de 2 pacientes, 2 porcos e 1 coelho	Polar	Remoção de artefatos, filtro de mediana	Discrete Wavelet Packet Frame (DWPF), Otsu, Reconstrução morfológica binária	Fração da área verdadeiro positivo (TPAF) Fração da área falso positivo (TNAF) Fração da área falso negativo (FNAF) Coefficiente Jaccard Coefficiente Dice	99,29% ± 2,96% 3,69% ± 2,88% 0,71% ± 2,96% 95,4% ± 4,8% 97,8% ± 2,16%
Kim et al. (2014)	30 Imagens de seis pacientes	Polar	Remoção de artefatos, filtro de mediana	Binarização por Otsu e Curva de Intensidade	Fração da área verdadeiro positivo (TPAF) Fração da área falso positivo (FPAF) Fração da área falso negativo (FNAF) Sensibilidade Especificidade	99,21% ± 0,21% 0,30% ± 0,15% 0,79% ± 0,51% 99,2% 99,7%
Xu et al. (2015)	500 imagens, divididos em 5 categorias	Polar	Transformada de Hough para remoção de artefatos	Segmentação baseada em grafo	Coefficiente Dice comparado a dois especialistas (E1 e E2)	(E1) 94,31% (E2) 94,46%
Macedo et al. (2015)	9 <i>pullbacks</i> de 9 pacientes, 1460 imagens ao todo	Polar	Remoção de artefatos, filtro bilinear, limiarização por Otsu	Varredura por linha, subtração de imagens, operações morfológicas	<i>Volume Overlap</i> (VOE) Coefficiente Dice <i>RMS Symmetric surface distance</i> (RMSSSD) Diferença de área absoluta média (MADA) Distância de Hausdorff Acurácia	95,1 ± 4,41% 97,4 ± 2,67% 0,04 ± 0,04mm 0,13 ± 0,25mm ² 0,17 ± 0,22mm 99,88 ± 0,25%
Roy et al. (2016)	15 <i>pullbacks in-vitro</i> e 6 <i>in-vivo</i> , com 150 a 200 <i>frames</i> cada	Cartesiano	limiarização, reconstrução geodésica	Representação em grafos e otimização por passeio aleatório	Coefficiente Kappa de Cohen Erro RMS do contorno	97,8% ± 0,61% 8.4655 ± 3.8540
Wang e Tang (2016)	25 bases de dados <i>in-vivo</i> , 23 com 271 <i>frames</i> e 2 com 541	Polar	Binarização, filtro de borda vertical, filtro de mediana, remoção de artefatos	Detecção do contorno do lúmen por programação dinâmica	Comparação entre dois especialistas (GT1 e GT2) do coeficiente de Correlação intra-classe (ICC) e diferença média de tamanho (Diff)	(GT1) ICC: 0,940; Diff: -2,56 ± 12,21 (GT2) ICC: 0,974; Diff: 1,49 ± 7,27
Menguy et al. (2016)	9 <i>pullbacks</i> , 390 imagens ao todo	Polar	Transformada de Hough para remoção de artefatos, limiarização e operações morfológicas	Gradiente, filtro gaussiano, algoritmo <i>fast marching</i>	Coefficiente de Tanimoto	90,5%
Cao et al. (2017)	880 imagens de 5 pacientes	Polar	Remoção de artefatos, filtro gaussiano.	<i>Level Set</i>	Coefficiente Dice	98,1% ± 1,1%
Yong et al. (2017)	64 <i>pullbacks</i> de 28 pacientes, com 155 a 375 <i>frames</i> cada	Polar	–	Redes convolutivas para regressão linear	Coefficiente Dice Coefficiente Jaccard Erro médio absoluto	98,5% 97% 20,87µm

2.2 IDENTIFICAÇÃO DE REGIÕES DE BIFURCAÇÃO EM IMAGENS INTRAVASCULARES

Alberti et al. (2012) utiliza características de textura para avaliar três classificadores (*Adaboost*, *Random Forest* e SVM) em imagens IVUS. A sua base de dados consiste em 22 *pullbacks* de 22 pacientes diferentes de onde foram extraídas características, sendo as principais: características de textura do tecido a partir filtros de Gabor e de padrões binários locais (LBP), e características de textura na região do sangue a partir de técnicas de auto correlação e correlação cruzada. A classificação ocorre em duas etapas, em um esquema denominado *Multiscale Stacked Sequential Learning* (MSSL): as características utilizadas na primeira classificação, juntamente com a sua margem de classificação, são utilizadas como insumo para um segundo classificador. A margem de classificação é convertida em informações de estimativa do *frame* pertencer ou não a uma região de bifurcação, de forma que o segundo classificador aproveita o contexto espacial presente em sequências de *frames* para melhorar o desempenho. Ao final, um refinamento do resultado é feito, de forma a remover falsos positivos derivados de sombras do cateter-guia e outros artefatos. Assim, os melhores resultados são do *Adaboost*, apresentado acurácia = $95,42 \pm 2,05\%$, sensibilidade = $80,66 \pm 14,90\%$, especificidade = $95,62 \pm 2,05\%$, precisão = $94,40 \pm 2,65\%$, taxa de alarme falso (FAR) = $4,38 \pm 2,05\%$ e *score* F1 = $86,35 \pm 9,28\%$.

Wang et al. (2014) analisa a distância do centro do vaso para a borda do lúmen em imagens IVOCT para classificar regiões de bifurcação. Sua base de dados consiste em 25 *pullbacks*, onde os frames passam por uma etapa de remoção de artefatos (remoção do cateter, do cateter-guia e da bainha). Em seguida, uma etapa de detecção da borda do lúmen é feita a partir do algoritmo de Dijkstra, de forma a encontrar o melhor caminho (a borda do tecido) sobre o gradiente da imagem, dado por operador de Prewitt. Após isso, analisa-se a distância do centro do lúmen à borda, em que uma variação abrupta da distância pode indicar uma região

de bifurcação. Falsos positivos são tratados nos casos em áreas do vaso não mapeadas, que ocorre por obstrução quando o cateter está muito próximo da parede, e em situações causadas por sombras geradas por *stents*. As métricas utilizadas foram taxa de verdadeiro positivo (TPR) e taxa de falso positivo (FPR), com 94,0% e 5,0%, respectivamente.

Macedo et al. (2015) utiliza-se de descritores de formato para classificação, avaliando três classificadores: SVM, *Adaboost* e *Random Forest*. Como dito anteriormente na seção 2.1, foram utilizados 9 *pullbacks* IVOCT de 9 diferentes pacientes, gerando três bases de dados conforme os tipos de lesões presentes nos *frames*. As imagens passam por uma etapa de pré-processamento com o objetivo de detectar o contorno do lúmen. A partir do contorno, são definidas extraídas as seguintes características geométricas: distância do lúmen para o seu centroide, circularidade, energia de dobra (*bending energy*), curvatura, variância radial, razão do eixo, variância e desvio padrão da assinatura do contorno (*contour signature*), diferença de área entre frames e área triangular média. Duas técnicas de seleção de características baseadas em mínimos quadrados ortogonal (ORL) são avaliadas neste trabalho: algoritmo *Forward Regression with Orthogonal Least Squares* (FROLS) e *Multiple FROLS* (MFROLS). Os melhores resultados são obtidos com *Adaboost* sem seleção de características com TPR = $27,48 \pm 13,87\%$, FPR = $98,99 \pm 0,67\%$, acurácia = $95,58 \pm 2,08\%$, especificidade = $72,52 \pm 13,87\%$, precisão = $96,17 \pm 1,86\%$, taxa de alarme falso = $3,97 \pm 1,98\%$, *score* F1 = $97,51 \pm 1,15\%$ e AUC = $96,41 \pm 3,18\%$.

Porto et al. (2017) aproveita-se da mesma base de dados e do conjunto de características de Macedo et al. (2015) para avaliar o desempenho de redes neurais e SVM. Foram avaliadas redes neurais de quatro camadas, de diferentes quantidades de neurônios nas camadas ocultas e comparando o desempenho dos seguintes critérios de parada: erro quadrático médio (MSE), regularização e parada antecipada. Para o SVM, foram avaliados o desempenho de dois *kernels* diferentes: polinomial e função de base radial gaussiana (GRBF). Além disso, foram utilizados

três diferentes critérios de seleção de características durante os experimentos: seleção escalar de características com correlação (SFS), análise de discriminantes lineares (LDA) e análise de componentes principais (PCA). Em geral, os resultados obtidos com redes neurais foram melhores que os obtidos com SVM, sendo que o melhor classificador foi a rede neural de 104-8-8-1 neurônios, utilizando MSE com $FPR = 1,7 \pm 1,47\%$, $TPR = 97,36 \pm 0,81\%$, acurácia = $98,15 \pm 0,79\%$, especificidade = $98,83 \pm 1,47\%$, precisão = $98,7 \pm 1,56\%$, taxa de falso alarme = $1,37 \pm 1,73\%$, *score* F1 = $97,93 \pm 0,81\%$ e Área sob a curva ROC = $99,0 \pm 1,0\%$.

O Quadro 2 contém os principais pontos de cada referência abordada, sintetizando esta seção.

Quadro 2 – Síntese da revisão bibliográfica da classificação de região de bifurcação em imagens intravasculares.

Autor	Base de dados	Pré-processamento	Características	Seleção de características	Técnica Utilizada	Métricas	Resultados
Alberti et al. (2012)	22 <i>pullbacks</i> IVUS	Divisão por setores angulares	Descritores de texturas	Seleção por análise de pesos	<i>Adaboost, Random Forest, SVM</i>	Acurácia Sensibilidade Especificidade Precisão Taxa de alarme falso <i>Score F1</i>	95,42 ± 2,05% 80,66 ± 14,90% 95,62 ± 2,05% 94,40 ± 2,65% 4,38 ± 2,05% 86,35 ± 9,28%
Wang et al. (2014)	25 <i>pullbacks</i> IVOCT, 23 com 271 <i>frames</i> e 2 com 541 <i>frames</i>	Remoção de artefatos, detecção do lúmen	–	–	Análise da distância entre o centro e a borda do lúmen	Taxa de verdadeiro positivo Taxa de falso positivo	94,0% 5,0%
Macedo et al. (2015)	9 <i>pullbacks</i> IVOCT, 1460 imagens	Conversão para coordenadas polares, remoção de artefatos, detecção do lúmen	Descritores de formato	FROLS, MFROLS	<i>Adaboost, Random Forest, SVM</i>	Taxa de falso positivo Taxa de verdadeiro positivo Acurácia Especificidade Precisão Taxa de alarme falso <i>Score F1</i> AUC	27,48 ± 13,87% 98,99 ± 0,67% 95,58 ± 2,08% 72,52 ± 13,87% 96,17 ± 1,86% 3,97 ± 1,98% 97,51 ± 1,15% 96,41 ± 3,18%
Porto et al. (2017)	9 <i>pullbacks</i> IVOCT, 1460 imagens	Remoção de artefatos	–	SFS, LDA, PCA	Redes Neurais, SVM	Taxa de falso positivo Taxa de verdadeiro positivo Acurácia Especificidade Precisão Taxa de alarme falso <i>Score F1</i> AUC	1,7 ± 1,47% 97,36 ± 0,81% 98,15 ± 0,79% 98,83 ± 1,47% 98,7 ± 1,56% 1,37 ± 1,73% 97,93 ± 0,81% 99,0 ± 1,0%

3 REFERENCIAL TEÓRICO

3.1 REDES NEURAIIS ARTIFICIAIS

Redes neurais artificiais (RNA) são redes inspiradas na forma que um cérebro desempenha uma tarefa. Estas redes são compostas por unidades de processamento chamadas de neurônios e são interligadas entre si. A capacidade de uma rede aprender a desempenhar uma determinada tarefa é baseada na força dessas interconexões, chamadas de pesos sinápticos (HAYKIN, 2008). A Figura 7 exemplifica uma rede neural com uma unidade de entrada na camada de entrada e duas camadas ocultas com três unidades cada e uma camada de saída com duas unidades de saída.

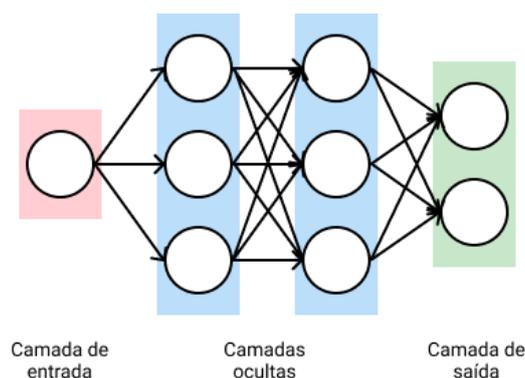


Figura 7 – Exemplo de rede neural com uma camada de entrada com uma unidade, duas camadas ocultas com três unidades cada e uma camada de saída com duas unidades.

Assim como as redes biológicas, além da possibilidade de aprender a partir de amostras de treinamento, as RNAs também possuem a capacidade de generalizar, ou seja, uma rede pode dar resultados satisfatórios para problemas similares de mesma natureza que foram apresentados durante o treino. Dessa forma, eles também apresentam ter um alto nível de tolerância a falhas contra entradas ruidosas (KRIESEL, 2005).

Um neurônio artificial consiste em uma combinação linear de entradas x_1, x_2, \dots, x_D na forma:

$$a = \sum_{i=1}^D w_i x_i + w_0 \quad (1)$$

Onde w_i são chamados de pesos e w_0 é chamado de viés (*bias*). A saída z deste neurônio é dado por uma função de ativação diferenciável não linear $h(\cdot)$ sobre a :

$$z = h(a) \quad (2)$$

Podemos definir que a saída computada z da i -ésima unidade computacional de uma camada l uma rede neural de $(L + 1)$ camadas como:

$$z_i^{(l)} = h(a_i^{(l)}), \text{ com } a_i^{(l)} = \sum_{k=1}^{m^{(l-1)}} w_{ik}^{(l)} z_i^{(l-1)} + w_{i0}^{(l)} \quad (3)$$

Onde a saída da i -ésima unidade da camada $(l - 1)$ é ponderado pelo peso $w_{ik}^{(l)}$ da k -ésima unidade da camada (l) e $m^{(l)}$ indica a quantidade de neurônios na camada l (STUTZ, 2014).

3.1.1 Funções de ativações

Para que a rede adquira uma capacidade de representação mais generalizada, a função de ativação $h(\cdot)$ precisa ser não linear. Para que o treinamento de uma rede seja possível essa função deverá ser diferenciável (BISHOP, 1994). Por essas razões, comumente utiliza-se a função tangente hiperbólica:

$$\tanh a = \frac{e^a - e^{-a}}{e^a + e^{-a}} \quad (4)$$

Outra função muito utilizada é a função sigmoide:

$$\sigma(a) = \frac{1}{1 + e^{-a}} \quad (5)$$

3.1.2 Treinamento da rede

O treinamento de uma rede busca encontrar um conjunto de pesos para a rede de tal forma a minimizar uma determinada função que indica o erro entre o valor desejado t_n e a

saída da rede $y(x_n; w)$, dado um valor de entrada x_n de um conjunto de treinamento de $n = 1, 2, \dots, N$ elementos. Uma função de erro muito utilizada é a soma dos quadrados do erro:

$$E(w) = \sum_{n=1}^N E_n(w) = \sum_{n=1}^N \sum_{k=1}^C (y_k(x_n; w) - t_{nk})^2 \quad (6)$$

Muitos dos algoritmos utilizados para a minimização de funções de erro baseiam-se no uso de derivadas na forma do vetor gradiente $\nabla E(w)$ da função de erro (BISHOP, 1994). Uma técnica muito eficiente para calcular este gradiente em redes neurais é chamada de *backpropagation*.

De forma resumida, o algoritmo de *backpropagation* é dada da seguinte forma (BISHOP, 2006):

1. Encontrar as ativações de todas as unidades escondidas e de saída aplicando um vetor de entrada x_n a rede;
2. Encontrar o valor de δ_k para todas as unidades utilizando

$$\delta_k = y_k - t_k; \quad (7)$$

3. Obter δ_j de cada unidade escondida através do *backpropagation* dos δ através da fórmula

$$\delta_j = h'(a_j) \sum_k w_{kj} \delta_k; \quad (8)$$

4. Obter as derivadas necessárias utilizando

$$\frac{\partial E_n}{\partial w_{ji}} = \delta_j z_i. \quad (9)$$

3.2 REDES NEURAIAS CONVOLUTIVAS

Redes neurais convolutivas (CNN) é uma abordagem proposta por LeCun et al. (1989, 1998) que se diferenciam das redes neurais tradicionais por combinar três características: campo receptivo local, pesos compartilhados e subamostragem (LECUN et al., 1998). Sua

aplicação, em sua origem, é voltada para processamento de imagens, pois buscou solucionar alguns desafios devido às características desse tipo de dado.

Utilizar imagens, que geralmente possuem milhares de variáveis (*pixels*), como entrada em uma rede com unidades totalmente conectadas significa que, cada unidade computacional da camada de entrada, por exemplo, estaria conectada a cada uma dessas variáveis, totalizando centenas de milhares de pesos a serem computados somente nesta camada. Utilizar um campo receptivo local é uma estratégia eficiente para reduzir a quantidade de pesos, pois cada unidade, organizada em planos chamados de mapa de características, estaria sujeita a uma pequena sub-região da imagem. Uma região de 5×5 *pixels*, por exemplo, significaria que o mapa de características teria 25 pesos e um *bias*.

Ao compartilhar um mesmo conjunto de pesos entre as unidades de uma camada, a ativação dessas unidades reproduz um comportamento equivalente a convolução entre a imagem e um *kernel*. Essa propriedade torna a rede menos susceptível a variações por translação e distorções na imagem (BISHOP, 2006). Em camadas mais iniciais, cada unidade pode extrair características simples como cantos e bordas em diversas orientações e, conforme essas características são combinadas em camadas mais profundas, características mais complexas são extraídas.

A propriedade de subamostragem ou *pooling* procura reduzir a sensibilidade a pequenos deslocamentos e distorções na imagem (BISHOP, 2006). Isto é feito nas camadas de *pooling*, que reduzem a resolução espacial do mapa de características. Uma camada de *pooling*, por exemplo, pode ter unidades com campos receptivos 2×2 posicionados continuamente e sem sobreposição que aplicam a média das quatro entradas, reduzindo pela metade a dimensão do mapa.

Em uma arquitetura típica, uma rede pode contém várias camadas convolutivas acompanhadas de camadas de ativação não-linear e algumas camadas de *pooling*, reduzindo

gradativamente a resolução espacial enquanto aumenta a concentração de características. Ao final, dependendo da tarefa da rede, pode haver camadas totalmente conectadas (*fully-connected layers*) e uma camada *softmax* em uma tarefa de classificação, por exemplo.

3.2.1 Camada convolutiva

A convolução é uma operação entre duas funções definida por:

$$s(t) = (x * w)(t) = \int x(a)w(t - a)da \quad (10)$$

Sendo que sua versão discreta é definida por:

$$s(t) = (x * w)(t) = \sum_{a=-\infty}^{\infty} x(a)w(t - a) \quad (11)$$

Para aplicações em CNNs, geralmente aplica-se a convolução em mais de um eixo, portanto, para uma aplicação bidimensional de uma imagem I sobre um kernel K , temos a equação:

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(m, n)K(i - m, j - n) \quad (12)$$

Como a convolução é uma operação comutativa, podemos reescrever a equação acima como:

$$S(i, j) = (K * I)(i, j) = \sum_m \sum_n I(i - m, j - n)K(m, n) \quad (13)$$

Por fim, é comum que bibliotecas de redes neurais implementem a função correlação cruzada ao invés da convolução, pois a propriedade comutativa obtida ao virar o *kernel* em relação a imagem não é muito importante na implementação da rede (GOODFELLOW, BENGIO e COURVILLE, 2016). Portanto, a equação da convolução em aplicações de redes neurais é:

$$S(i, j) = (K * I)(i, j) = \sum_m \sum_n I(i + m, j + n)K(m, n) \quad (14)$$

A Figura 8 exemplifica uma operação de convolução de um kernel 2×2 sobre uma imagem 3×3 :

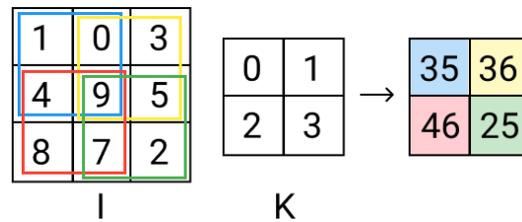


Figura 8 – Exemplo de convolução do *kernel* K sobre a imagem I.

3.2.2 Unidades retificadoras lineares (ReLU)

Em redes tradicionais comumente é utilizado a função sigmoide como função de ativação. Em CNNs, por outro lado, a recomendação é utilizar a função ReLU, definida por:

$$g(z) = \max\{0, z\}, \quad (15)$$

pois funções como o sigmoide e a tangente hiperbólica tornam o processo de treinamento da rede mais lento, uma vez que seus gradientes são fracos próximos da saturação (DENG e YU, 2013).

3.2.3 Camada de *pooling*

A camada de *pooling* reduz a resolução espacial da rede a partir de operações de média (*mean pooling*) ou de máxima (*max pooling*). A operação de *max pooling*, por exemplo, seleciona o valor máximo dentro da vizinhança, como exemplifica a Figura 9.

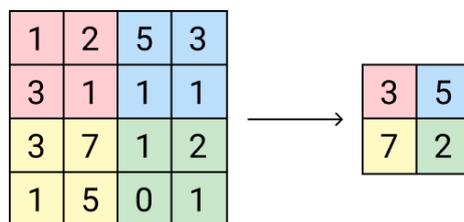


Figura 9 – Exemplo de *max pooling*.

3.2.4 Convolução transposta

A convolução transposta, chamado também de deconvolução, é uma operação que funciona de forma contrária a convolução, utilizado para a sobreamostragem da rede. Primeiramente, podemos expressar a convolução como uma multiplicação de matrizes, onde,

por exemplo, expressamos a convolução da Figura 8 na forma mostrada na Figura 10, em que a figura B é convertida para um vetor (9×1) e cada linha da matriz A representa uma operação de convolução utilizando o *kernel* K . Por fim, o vetor C resultante da operação pode ser convertido de volta para uma matriz (2×2), sendo assim o resultado da convolução de K sobre B .

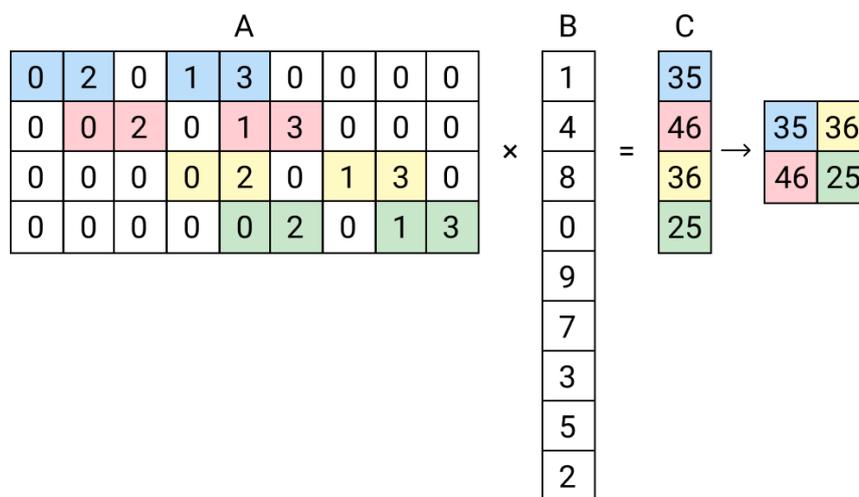


Figura 10 – Representação da convolução da Figura 8 em forma de matrizes.

Da mesma forma, podemos expressar a convolução transposta na forma de multiplicação de matrizes. Com base no exemplo dado na Figura 10, a convolução é uma multiplicação de matrizes que resulta em um vetor (4×1) e é dada por:

$$A_{(4 \times 9)} B_{(9 \times 1)} = C_{(4 \times 1)} \quad (16)$$

A convolução transposta é dada ao reescrever a equação acima de forma que a operação resulte em um vetor (9×1) a partir do vetor (4×1):

$$(A^T)_{(9 \times 4)} C_{(4 \times 1)} = B_{(9 \times 1)} \quad (17)$$

Ou seja, podemos sobreamostrar uma imagem $C_{(2 \times 2)}$, convertendo-o em um vetor (4×1) e multiplicando-o pela transposta da matriz A . O vetor resultante B é convertido de volta para uma matriz (3×3), sendo esta a imagem de saída da convolução transposta. Na Figura 11,

vemos um exemplo de convolução transposta, onde utilizamos o mesmo kernel K em uma imagem C (2×2), resultando em uma matriz de saída B (3×3).

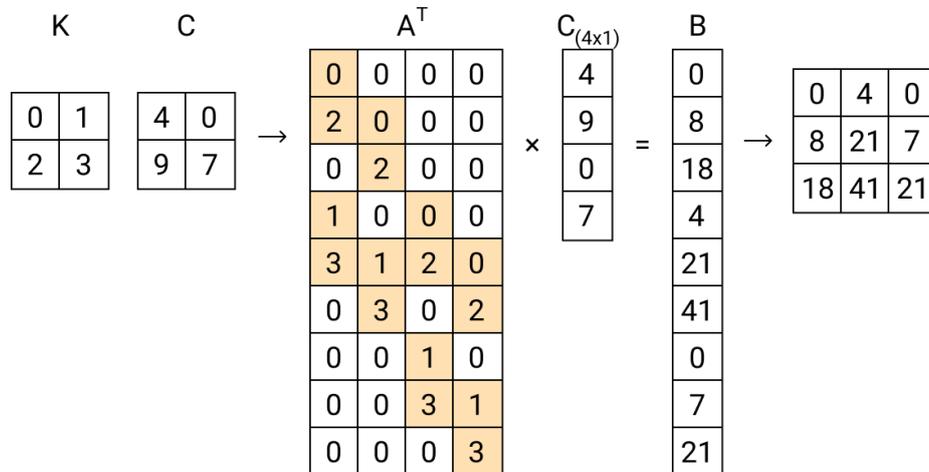


Figura 11 – Convolução transposta da imagem C pelo kernel K representado por uma multiplicação de matrizes.

3.2.5 Camada de *dropout*

Uma forma de evitar *overfitting* da rede, além de técnicas como parada antecipada e regularização L_1 e L_2 , é o *dropout*. Durante o treinamento, as unidades possuem uma probabilidade p de não serem desativadas. As unidades desativadas não participam da rede, como mostra a Figura 12.

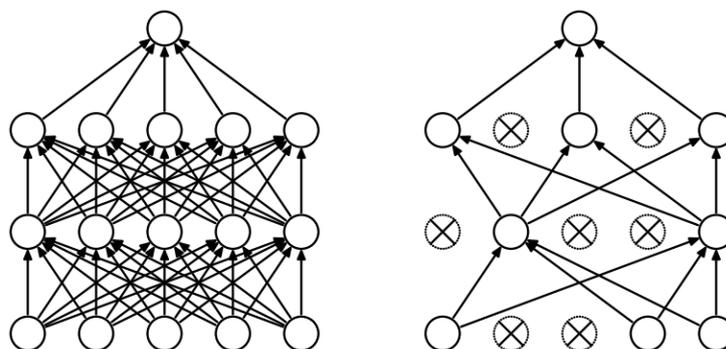


Figura 12 – Exemplo da ação da camada de *dropout*. A rede da direita é o resultado da aplicação do *dropout* na rede da esquerda. Fonte: (SRIVASTAVA et al., 2014).

3.2.6 Camada de normalização em lote (*batch normalization*)

As camadas de *batch normalization* solucionam um problema onde, durante o treinamento, as mudanças dos parâmetros da rede causam mudanças na distribuição das ativações da rede nas camadas ocultas da rede. Este problema, definido como *internal covariate*

shift, é solucionado através de uma etapa de normalização que corrige a média e a variância da saída da camada anterior. O *batch normalization* permite taxas de aprendizado maiores e contribui na normalização do modelo, sendo que em redes que utilizam esta camada, a camada de *dropout* pode deixar de ser utilizada ou ter sua influência reduzida (IOFFE e SZEGEDY, 2015).

O *batch normalization* adiciona dois parâmetros γ e β a serem aprendidos durante o treinamento. Portanto, a saída torna-se:

$$y = \gamma \hat{x} + \beta, \quad (18)$$

sendo que \hat{x} é a entrada normalizada. Para calcular \hat{x} em um mini lote B de tamanho m , $B = \{x_1, x_2, \dots, x_m\}$, calcula-se primeiramente a média e a variância do mini lote:

$$\mu_B = \frac{1}{m} \sum_{i=1}^m x_i \quad (19)$$

$$\sigma_B^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \mu_B)^2 \quad (20)$$

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \quad (21)$$

onde a constante ϵ tem a função de estabilidade numérica.

3.2.7 Segmentação semântica

Redes neurais convolutivas podem, além de classificar imagens, desempenhar tarefas de classificação local. Na segmentação semântica, cada pixel da imagem de entrada é classificado em uma das classes definidas, como exemplifica a Figura 13.

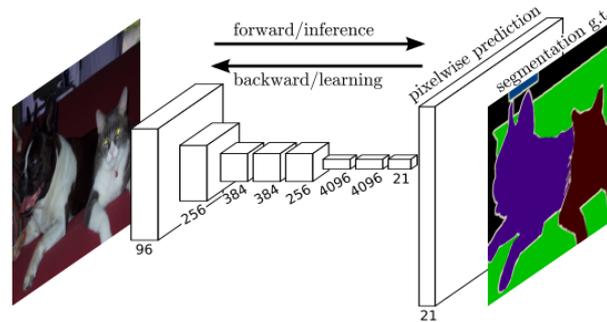


Figura 13 – Exemplo de uma CNN para segmentação semântica. Fonte: (LONG, SHELHAMER e DARRELL, 2015).

No início, adaptava-se redes de classificação como AlexNet (KRIZHEVSKY, SUTSKEVER e GEOFFREY E., 2012), VGGNet (SIMONYAN e ZISSERMAN, 2014) e GoogleNet (SZEGEDY et al., 2015), como proposto por Long, Shelhamer e Darrell (2015), porém essa abordagem sofre com algumas limitações, como a incapacidade de classificar elementos pequenos e a perda de detalhes nos objetos classificados por conta do processo simplificado de sobreamostragem (NOH, HONG e HAN, 2015).

Arquiteturas mais modernas utilizam etapas de múltiplas camadas de sobreamostragem e ativações não lineares para que o objeto seja reconstruído de forma mais gradual, obtendo melhores resultados. A Figura 14 mostra a arquitetura proposta por Noh, Hong e Han (2015), em que a rede apresenta a etapa de sobreamostragem como espelhamento da etapa de subamostragem. A rede cresce em tamanho espacial gradativamente através de camadas de *unpooling* e de convolução transposta.

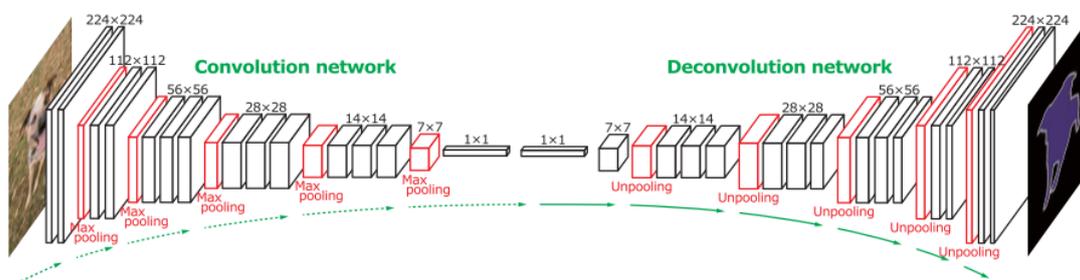


Figura 14 – Arquitetura de segmentação semântica proposta por Noh, Hong e Han (2015). Fonte: (NOH, HONG e HAN, 2015).

3.2.8 Aumento de dados (*Data augmentation*)

Data augmentation consiste em aumentar artificialmente a base de dados através de técnicas de processamento de imagens de forma a ajustar uma base desbalanceada e evitar o *overfitting*. Base de dados especializados, como os da área médica, geralmente sofrem com a quantidade limitada de imagens disponíveis para classificação, apoiando-se em técnicas de aumento de dados para melhorar o desempenho de seus classificadores (VASCONCELOS e VASCONCELOS, 2017).

Geralmente, as técnicas mais utilizadas são de transformação geométrica, seja por espelhamento, rotação, distorção, translação e corte da imagem; e por transformação de cores, alterando a paleta de cores da imagem. Recentemente, Redes Adversárias Generativas (GANs), também têm sido utilizados para o aumento de dados, gerando novas imagens através de transferência de estilo onde, por exemplo, a rede aplica um estilo de pintura à uma imagem de entrada (PEREZ e WANG, 2017).

3.2.9 Transferência de aprendizado (*transfer learning*)

A transferência de aprendizado busca reduzir o esforço de aprendizagem aplicando o conhecimento adquirido em um treinamento anterior, ajudando a solucionar novos problemas. Pan e Yang (2010) classificam a transferência de aprendizado em três tipos:

- **Transferência indutiva:** onde os domínios de aprendizado da origem e do destino são iguais, enquanto as tarefas de origem e destino são diferentes;
- **Transferência transdutora:** onde os domínios de aprendizado da origem e do destino são diferentes, enquanto as tarefas de origem e destino são iguais;
- **Transferência não-supervisionada:** similar a transferência indutiva, mas seu foco é em melhorar o aprendizado não-supervisionado no domínio de destino. Neste caso, as bases de treinamento tanto da origem quanto do destino não são rotuladas.

A abordagem tradicional de se transferir conhecimento é copiando as primeiras n camadas de uma rede treinada para as n primeiras camadas da rede de destino. As camadas restantes são inicializadas de forma aleatória e treinadas normalmente. O critério de propagar o erro de treinamento às camadas já treinadas de forma a sofrerem um ajuste fino à tarefa, ou de manter os pesos “congelados” depende do tamanho da base de dados do destino e da quantidade de parâmetros das primeiras n camadas. Se o tamanho da base for muito pequeno e a quantidade de parâmetros forem grandes, há uma chance de ocorrer *overfitting*, portanto os pesos são mantidos sem alterações. Caso contrário, a rede pode se beneficiar do ajuste fino e melhorar seu desempenho (YOSINSKI et al., 2014).

4 MATERIAS E MÉTODOS

Para a implementação deste trabalho, foram empregadas a linguagem de programação Python 3.5 com as bibliotecas OpenCV3 e *Python Imaging Library* para a preparação das bases de dados e o ambiente de desenvolvimento Matlab versão 2017b para a execução e avaliação das redes convolutivas. Este ambiente foi executado em um computador Intel i5-7600 com 3,50 GHz, 16Gb de memória RAM e uma GPU GeForce GTX1070 com 8Gb e 1920 núcleos CUDA. A seguir, serão descritas a montagem das bases de dados utilizadas e as redes convolutivas propostas neste experimento. A fase de experimentação será dada conforme a Figura 15 para ambas as etapas de segmentação e classificação.

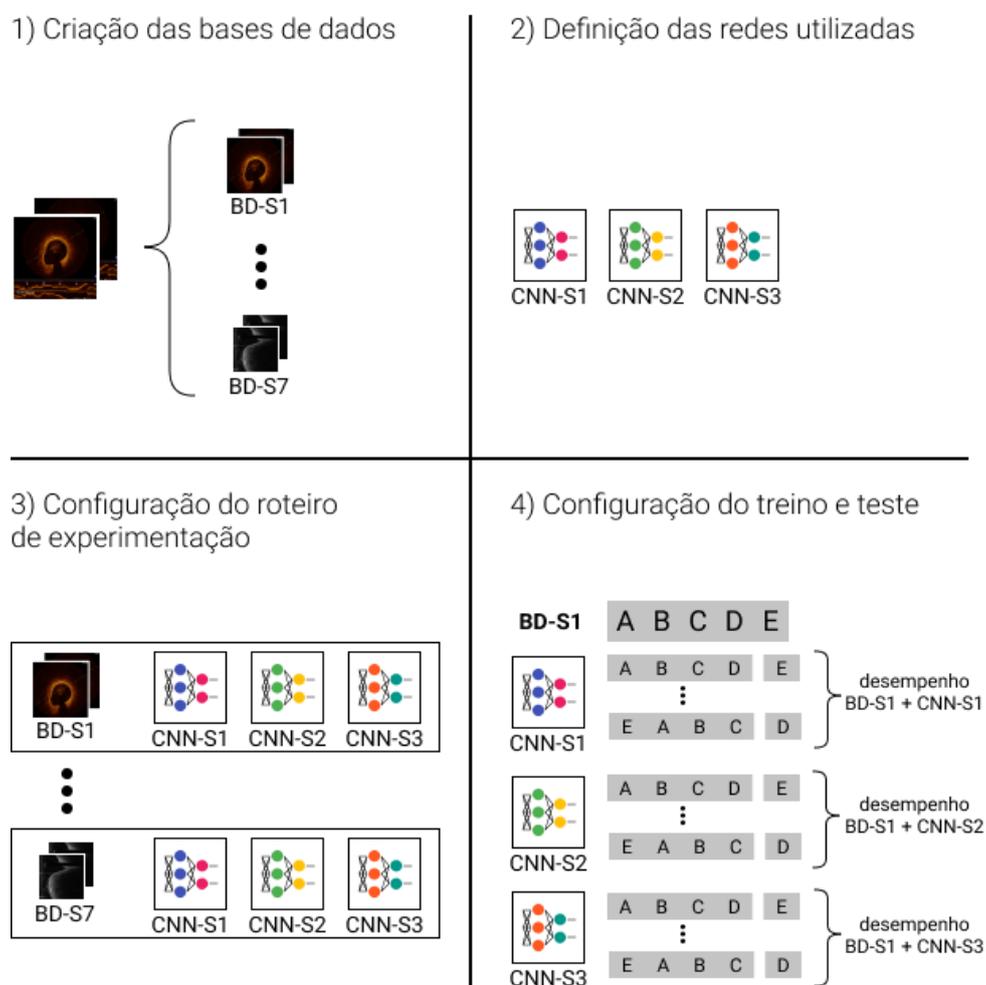


Figura 15 – Quatro passos da fase de experimentação. 1) Criação das bases de dados; 2) Definição das redes utilizadas; 3) Configuração do roteiro de experimentação; 4) Configuração do treino e teste.

Para a segmentação, a primeira fase consiste na criação de sete conjuntos de dados com base nas imagens IVOCT adquiridas para este trabalho. A seguir, configuram-se as três arquiteturas definidas no ambiente de desenvolvimento. O roteiro de experimentação consiste na combinação dos sete conjuntos de dados com as três redes. Por fim, configura-se o roteiro de execução individual, onde é feito o treinamento e teste por validação cruzada, conforme o conjunto de dados e a rede dados pelo roteiro de execução.

Para a classificação, a primeira fase consiste na criação de quatro conjunto de dados e são configuradas quatro arquiteturas na segunda fase. O roteiro de experimentação consiste em dez experimentos e o roteiro de treino e teste seguem de forma similar à segmentação.

4.1 CONJUNTO DE DADOS

O conjunto de dados utilizado é proveniente de um trabalho anterior feito por Macedo et al. (2015) e consiste em 1689 imagens IVOCT, de 9 *pullbacks* de 9 diferentes pacientes, cedidas pelo Instituto do Coração (InCor), do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, juntamente com uma lista de coordenadas que determinam o contorno do lúmen de cada imagem do conjunto. O projeto de pesquisa que originou o conjunto de imagens foi devidamente aprovado pelo Comitê de Ética da instituição proponente, assegurando que o caráter anônimo dos participantes está sendo mantido e suas identidades devidamente protegidas. As imagens foram cedidas em arquivos DICOM (*Digital Imaging and Communications in Medicine*) e foram extraídos e convertidos para imagens PNG em tamanho original, 1024×1024 *pixels*, em cores (RGB), sendo 157 imagens de bifurcação e 1204 de não-bifurcação.

A Figura 16 mostra o processo de pré-processamento das imagens, onde um detector de círculos foi utilizado para isolar a seção transversal do vaso sanguíneo de cada imagem. O algoritmo binariza a imagem com um limiar $T = 1$ determinado experimentalmente, de forma

a separar quase toda a imagem do vaso do fundo da imagem. As bordas dos elementos são destacadas através de uma operação morfológica de fechamento seguido de um gradiente morfológico. Por fim, a transformada de Hough para detecção de círculos é utilizada e a área detectada é cortada, resultando em imagens 768×768 pixels.

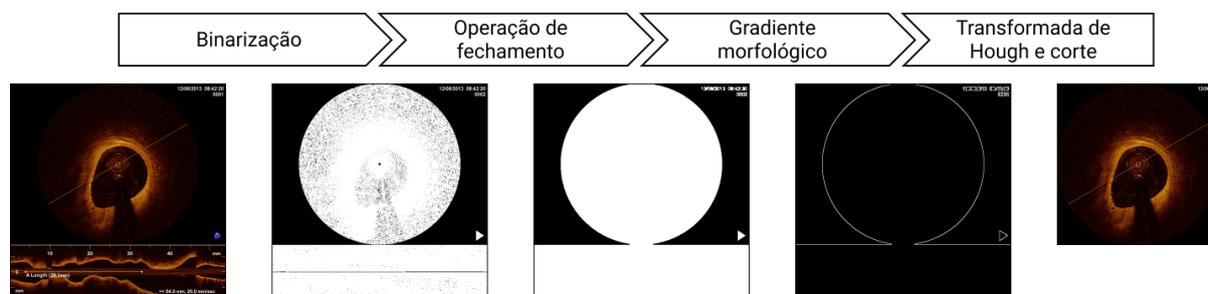


Figura 16 – Processo de pré-processamento das imagens

4.1.1 Conjunto de dados para a segmentação do lúmen

A partir deste ponto, foram montadas seis bases de dados para as experimentações conforme a Tabela 2. As funções *resize* e *linearpolar* do OpenCV3 para Python foram utilizados para redimensionar e converter as imagens para coordenadas polares, respectivamente. A Figura 17 exemplifica o processo de criação das bases de dados.

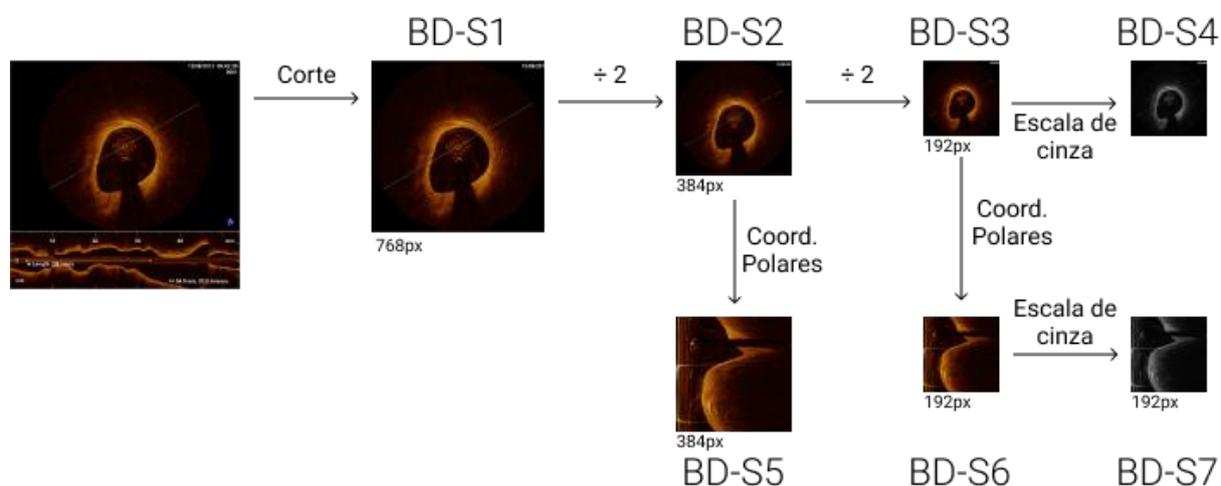


Figura 17 – Processo de criação das bases de dados para os experimentos de segmentação do lúmen.

Tabela 2 – Bases de dados utilizados nas experimentações de segmentação do lúmen

Base de dados	Espaço de cores	Tamanho (pixels)	Sistema de coordenadas
BD-S1	RGB	768×768	Cartesianas
BD-S2	RGB	384×384	Cartesianas
BD-S3	RGB	192×192	Cartesianas
BD-S4	Escala de cinza	192×192	Cartesianas
BD-S5	RGB	384×384	Polares
BD-S6	RGB	192×192	Polares
BD-S7	Escala de cinza	192×192	Polares

Para montar os conjuntos de imagens do padrão-ouro, utilizou-se a função *ImageDraw.Draw* do *Python Imaging Library* para desenhar a área do lúmen conforme as coordenadas do contorno e preencher a área do lúmen. Como exemplificado na Figura 18, o mesmo processamento de corte, redimensionamento e conversão de sistema de coordenadas foram feitos nessas imagens, conforme as respectivas bases de dados.

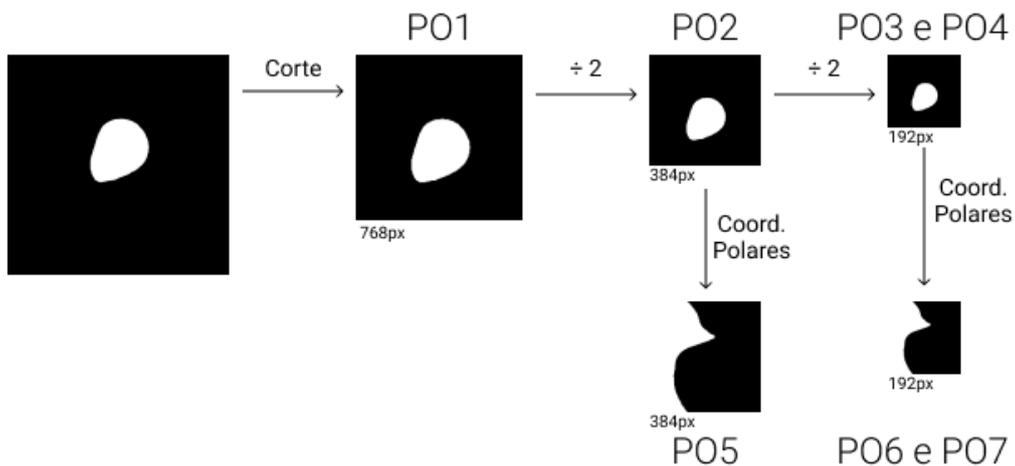


Figura 18 – Processo de criação do padrão-ouro das bases de dados para os experimentos de segmentação do lúmen.

4.1.2 Conjunto de dados para classificação de região de bifurcação

Para a tarefa de classificação, foram criadas quatro bases de dados, conforme a Tabela 3. O processo de redimensionamento e conversão foram conduzidos da mesma forma que na Seção 4.1.1. Como a quantidade de imagens rotuladas como bifurcação correspondem a apenas

11,5% do total de imagens, há um problema de viés de classificação. Assim, acrescentou-se uma etapa de *data augmentation*, onde as imagens de bifurcação foram rotacionadas em incrementos de 45 graus sete vezes como ilustra a Figura 19, adicionando 1099 novas imagens, num total de 2460 imagens em cada base de dados.

Tabela 3 – Bases de dados utilizados nas experimentações de classificação de região de bifurcação

Base de dados	Sistema de coordenadas	Espaço de cores	Classe bifurcação	Classe não-bifurcação	Total
BD-C1	Cartesianas	RGB			
BD-C2	Cartesianas	Escala de cinza	1256	1204	2460
BD-C3	Polares	RGB			
BD-C4	Polares	Escala de cinza			

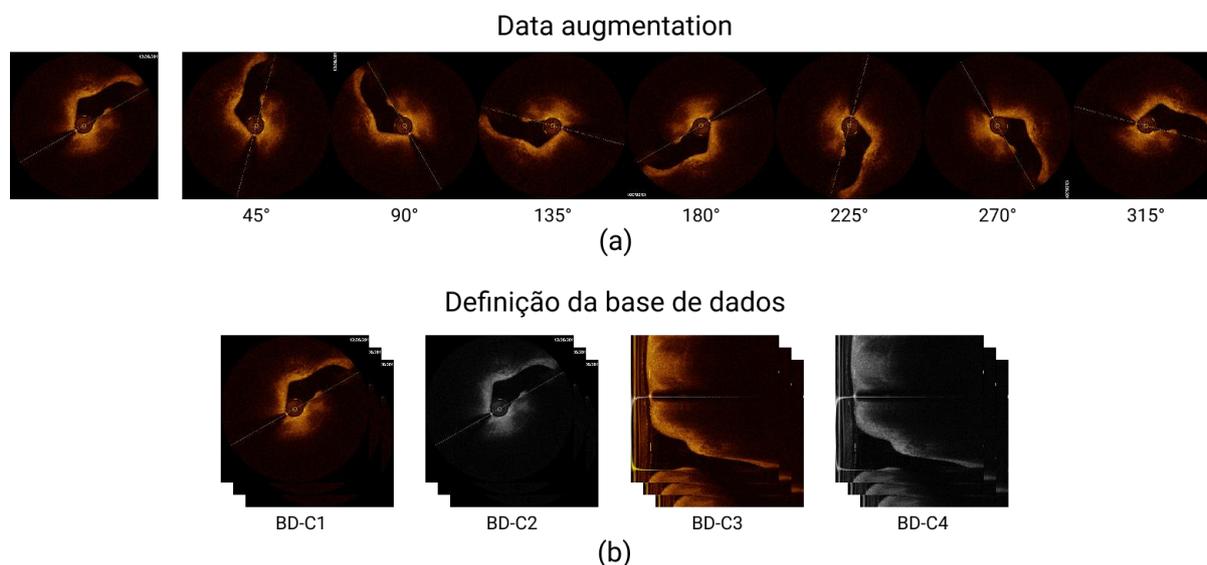


Figura 19 – Processo de criação da base de dados para a classificação de região de bifurcação. (a) As imagens de bifurcação são rotacionadas em incrementos de 45 graus cada. (b) Quatro bases definidas com diferentes sistemas de coordenadas e espaço de cores.

4.2 PROPOSTAS DAS REDES NEURAIIS CONVOLUTIVAS

4.2.1 Arquiteturas propostas para a segmentação do lúmen

Para a segmentação do lúmen, foram propostas três arquiteturas, descritas a seguir. As duas primeiras redes são inspiradas no modelo proposto por Noh, Hong e Han (2015) e a terceira se inspira no modelo U-Net de Ronneberger, Fischer e Brox (2015). As arquiteturas propostas, ilustradas na Figura 20, compartilham da mesma estrutura-base dos trabalhos anteriores, que consistem na alteração da resolução espacial de forma gradual em ambos os

blocos de subamostragem e sobreamostragem e, para a terceira rede, conexões entre os blocos presentes em cada etapa de alteração da resolução espacial. Em todos os casos, a quantidade de camadas, bem como seus hiperparâmetros diferem das arquiteturas-modelo, conforme descrito a seguir.

De forma geral, todas as redes apresentam seqüências de camada convolutiva 3×3 , *batch normalization* e ReLU. Esta seqüência (Conv 3×3 – Batch – ReLU) é repetida algumas vezes e, ao final, uma camada convolutiva 1×1 precede uma etapa de sub ou sobreamostragem. Essa camada reduz ou aumenta a profundidade da rede, desonerando o processamento na camada de convolução seguinte (SZEGEDY et al., 2015). Nas etapas de subamostragem, cada camada de *max pooling* possui filtro de tamanho 2×2 , que reduz a altura e largura da rede pela metade. De forma análoga, as etapas de sobreamostragem possuem uma camada de convolução transposta, com filtro de tamanho 4×4 , que dobra o tamanho espacial da rede, aumentando gradativamente até atingir o mesmo tamanho da imagem utilizada na entrada. Além disso, uma camada de *dropout*, com parâmetro $p = 0,3$, está presente entre a última etapa de subamostragem e o início da primeira etapa de sobre amostragem. No final da rede, as camadas *softmax* e de classificação atribuem 0 ou 1 para cada pixel da imagem, de forma a classificá-los em uma das duas classes definidas: lúmen ou fundo (*background*).

A primeira arquitetura (CNN-S1) é uma rede direta de 26 camadas, com quatro etapas de subamostragem, seguido de uma camada de *dropout* e quatro etapas de sobreamostragem. Como dito anteriormente, sobreamostragem é feita de forma gradual, dividido em várias camadas, de forma a preservar detalhes, como em (NOH, HONG e HAN, 2015).

A segunda arquitetura (CNN-S2) é uma variação da CNN-S1 com camadas adicionais, Convolutiva de *kernel* 3×3 – *Batch Normalization* – ReLU, e Convolutiva de *kernel* 1×1 entre as etapas de sobreamostragem, totalizando 41 camadas. Se comparado a CNN-S1, as camadas

ReLU adicionadas na segunda metade da rede devem conferir maior não linearidade e capacidade de abstração durante cada etapa de sobreamostragem.

A terceira arquitetura (CNN-S3) é uma variação da CNN-S2, com a adição de grafos acíclicos direcionados (DAG), ou seja, informações das etapas de subamostragem são passadas mais adiante para etapas de sobreamostragem análogas. Essas informações são unidas com os dados vindos do fluxo direto por camadas de concatenação. Este tipo de técnica pretende combinar as características de resolução mais alta detectados durante a subamostragem com as características encontradas na sobreamostragem, de forma a identificar maiores detalhes na saída (RONNEBERGER, FISCHER e BROX, 2015).

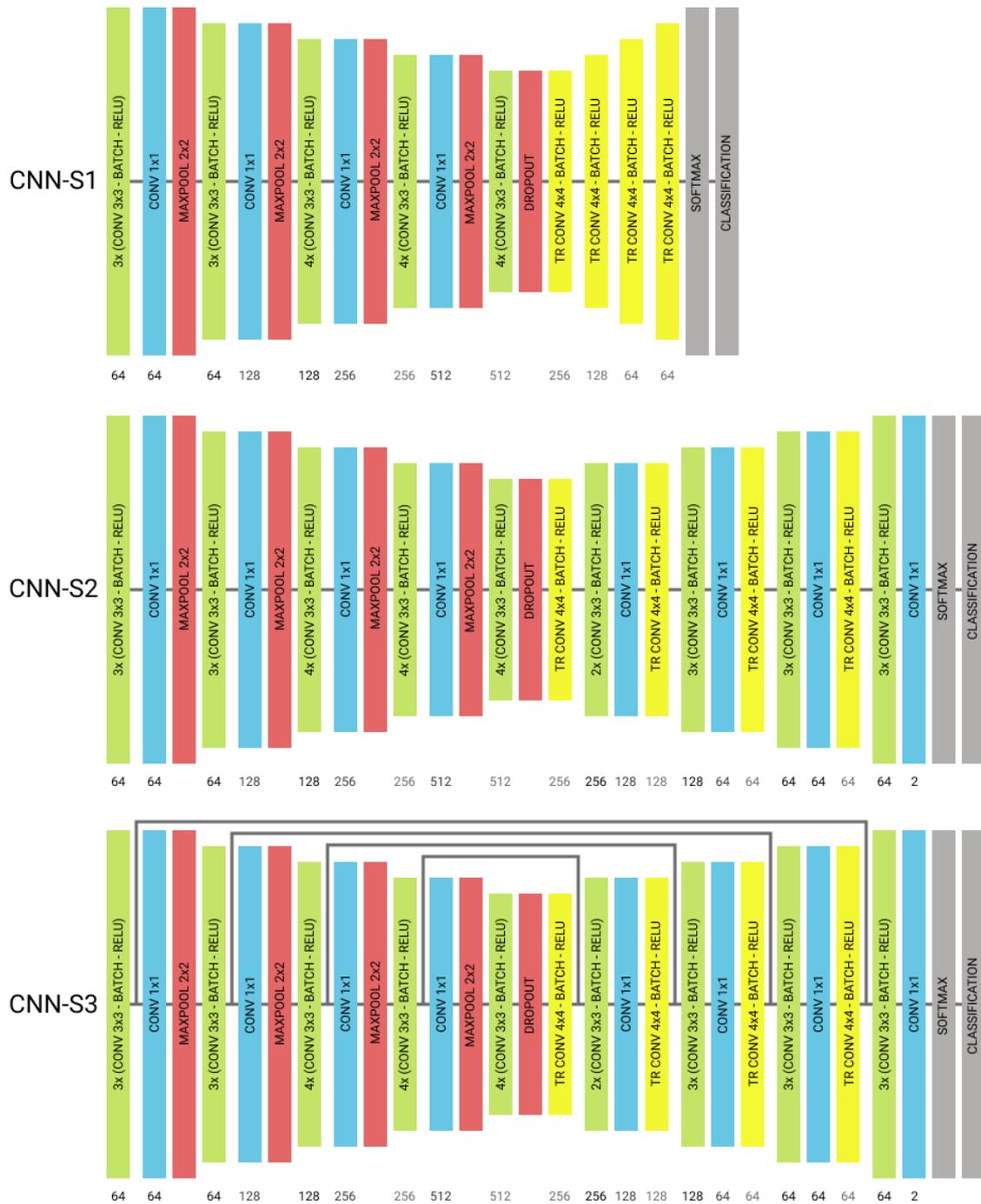


Figura 20 – Arquitetura das três redes convolutivas utilizadas na segmentação do lúmen.

4.2.2 Arquitetura proposta para a classificação de região de bifurcação

Para a classificação de região de bifurcação, quatro redes foram avaliadas, conforme a Figura 21. A primeira rede, CNN-C1, possui 30 camadas e é composta por 5 etapas de quatro conjuntos de camadas, Convolutiva de $kernel\ 3\times 3$ – *Batch Normalization* – ReLU, e Convolutiva de $kernel\ 1\times 1$ seguido de uma camada *Max Pool* para subamostragem. Ao final,

uma camada de *dropout*, três camadas totalmente conectadas de 1024 saídas e uma camada totalmente conectada de duas saídas ligada a camadas *softmax* e de classificação, atribuindo 1 para imagens com bifurcação e 0 para sem bifurcação.

As outras três arquiteturas utilizam estratégias de *transfer learning* para melhorar a tarefa de classificação. Por conta dos resultados obtidos e por ser uma rede direta, a rede CNN-S2 foi escolhida como rede de origem do conhecimento, ou seja, utilizou-se os pesos da rede treinada originalmente na tarefa de segmentação de lúmen para auxiliar no treinamento de uma nova rede de classificação de região de bifurcação.

A rede CNN-C2 consiste em ligar a saída da rede CNN-S2 à entrada da rede CNN-C1, com o objetivo de treinar a rede CNN-C1 somente com imagens segmentadas originadas do CNN-S2. A rede CNN-C3 possui uma camada de concatenação entre as sub redes, de forma a alimentar a rede CNN-C1 com ambas as imagens de entrada e sua contrapartida segmentada, originada da saída da rede CNN-S2. Por fim, a rede CNN-C4 utiliza parte da rede CNN-S2, reaproveitando todas as características aprendidas nas etapas de subamostragem, que são ligadas a três camadas totalmente conectadas para fins de classificação. Em todas estas redes, os pesos oriundos da CNN-S2 foram congelados, de forma que o aprendizado ocorra apenas no lado direito das redes.

4.3 EXPERIMENTOS

Para a segmentação, as três arquiteturas foram testadas em cada uma das sete bases de dados. Com o intuito de comparar o desempenho de cada arquitetura em cada base de dados, os hiperparâmetros de treinamento foram mantidos em todo o processo de experimentação. Para todo o processo de treinamento, foi utilizado o gradiente descendente estocástico (BOTTOU, 2010) com os parâmetros de treinamento ajustados conforme mostrado na Tabela 4.

Tabela 4 – Parâmetros utilizados na etapa de treinamento em todos os experimentos

	Segmentação	Classificação
Algoritmo de treinamento	SGD	SGD
Taxa de aprendizagem inicial	0,001	0,001
Fator de queda de taxa de aprendizagem	0,1	0,5
Período de queda da taxa de aprendizagem	10 épocas	40 épocas
Tamanho do lote	1	32
Condição de parada	50000 iterações	5000 iterações

Analisando os pixels das imagens, vimos que há uma forte predominância da classe *background*, configurando um problema de viés de classificação. De forma a corrigir isso, foram atribuídos pesos para ponderar a função de perda durante o treinamento e equilibrar a influência das duas classes sobre o treinamento. O cálculo do peso w para cada classe k se dá pela Equação (22):

$$w_k = \frac{\text{total de pixels}}{\text{total de pixels da classe } k}, \quad k = \text{lúmen, background} \quad (22)$$

A Tabela 5 apresenta uma descrição de cada experimento e o tempo de treinamento decorrido em cada um deles. Não houve o treinamento S3 por memória insuficiente para a realização do experimento.

Tabela 5 – Descrição dos 21 experimentos conduzidos para a segmentação do lúmen e o tempo decorrido para o treinamento de cada um deles.

Experimento	Base de dados	Arquitetura	Tempo de treinamento
S1	BD-S1	CNN-S1	35:41:19
S2		CNN-S2	69:29:37
S3		CNN-S3	– ⁽¹⁾
S4	BD-S2	CNN-S1	14:45:04
S5		CNN-S2	28:37:01
S6		CNN-S3	23:44:18
S7	BD-S3	CNN-S1	13:05:04
S8		CNN-S2	24:28:07
S9		CNN-S3	17:01:15
S10	BD-S4	CNN-S1	12:54:31
S11		CNN-S2	23:12:23
S12		CNN-S3	16:57:59
S13	BD-S5	CNN-S1	15:15:58
S14		CNN-S2	27:33:49
S15		CNN-S3	23:51:46
S16	BD-S6	CNN-S1	13:11:11
S17		CNN-S2	24:12:16
S18		CNN-S3	17:33:49
S19	BD-S7	CNN-S1	13:40:29
S20		CNN-S2	16:31:41
S21		CNN-S3	16:53:43

⁽¹⁾ Não houve experimento.

Para a classificação, as quatro arquiteturas foram avaliadas nas bases de dados descritas anteriormente, conforme ilustrado na Tabela 6. Da mesma forma que na segmentação, com o propósito de comparação de desempenho, os mesmos hiperparâmetros de treinamento foram mantidos para todas as redes, como apresentadas na Tabela 4. O tempo decorrido para o treinamento de cada experimento também é apresentado na Tabela 6.

Tabela 6 – Descrição dos dez experimentos conduzidos para a classificação de região de bifurcação e o tempo decorrido para o treinamento de cada um deles.

Experimento	Base de dados	Arquitetura	Tempo de treinamento
C1	BD-C1	CNN-C1	04:56:10
C2		CNN-C2	17:41:22
C3		CNN-C3	18:01:38
C4		CNN-C4	15:46:06
C5	BD-C2	CNN-C4	15:27:36
C6	BD-C3	CNN-C1	04:58:03
C7		CNN-C2	17:43:08
C8		CNN-C3	18:02:55
C9		CNN-C4	15:40:16
C10	BD-C4	CNN-C4	15:11:16

4.3.1 Divisão da base para treinamento e teste

Utilizamos a técnica de validação cruzada para conduzir a etapa de treinamento e teste. Para a segmentação, cada experimento teve sua base de dados dividida em cinco partes. O treinamento é feito com quatro delas e o teste com a quinta parte, de forma que 80% da base fosse utilizada para treino e 20% para teste. Esta etapa é repetida mais quatro vezes, de forma que os testes sejam feitos com cada uma das partições. Por fim, é feita a média aritmética dos resultados obtidos em cada iteração, resultando no desempenho médio do treinamento daquele experimento, como mostra a Figura 22. Para a classificação, em cada experimento a base de dados foi dividida em dez partes, de forma que 90% da base fosse utilizada para treino e 10% para teste.

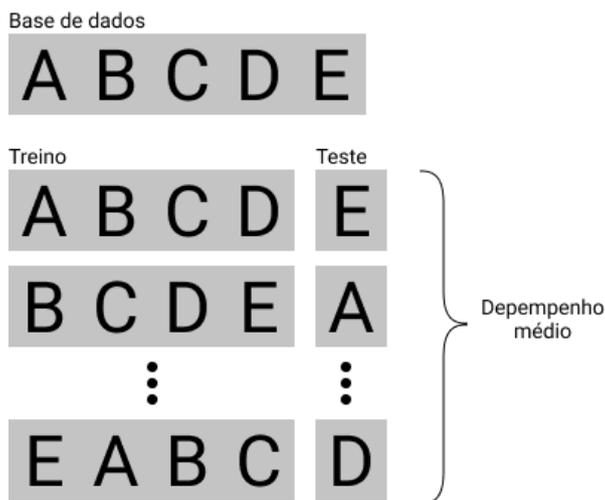


Figura 22 – Validação cruzada em cinco partições

4.4 MÉTRICAS DE DESEMPENHO

4.4.1 Métricas utilizadas na segmentação do lúmen

As seguintes métricas foram utilizadas para a análise do desempenho na tarefa de segmentação do lúmen: acurácia (ACC), sensibilidade (SEN), especificidade (SPE), coeficiente Dice (D), coeficiente Jaccard (J) e *score* F1 (F1).

Em uma predição onde pode-se obter resultados positivos ou negativos, temos quatro possíveis resultados, ilustrados no diagrama de Venn da Figura 23. O conjunto A representa o padrão-ouro, a classificação que nos é dada como correta, enquanto o conjunto B representa o resultado predito pelo experimento. Neste experimento, consideramos como positivo a classe lúmen e como negativo a classe *background*.

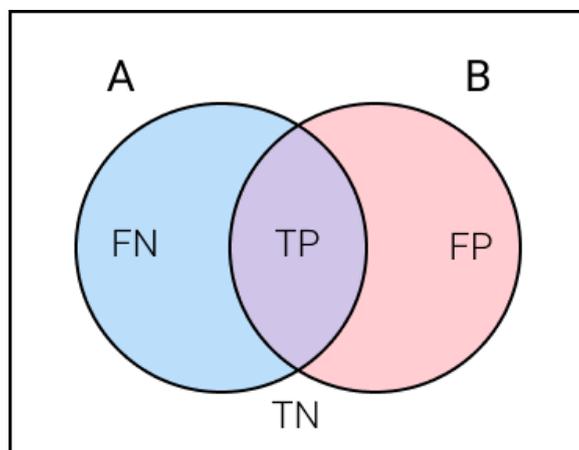


Figura 23 – Diagrama de Venn ilustrando os possíveis resultados entre o padrão-ouro e o que foi predito.

Temos um verdadeiro positivo (TP) quando um *pixel* da classe lúmen é corretamente classificado pelo preditor. Um falso positivo (FP) se um *pixel* da classe *background* é classificado de forma incorreta como lúmen. Um verdadeiro negativo (TN) caso um *pixel* da classe *background* seja corretamente classificado como tal e, por último, temos um falso negativo (FN) se um *pixel* da classe lúmen é classificado incorretamente como *background*.

Desta forma, as métricas podem ser definidas conforme mostrado no Quadro 3:

Quadro 3 – Métricas utilizadas na avaliação da segmentação do lúmen.

Acurácia	$ACC = \frac{TP + TN}{TP + FP + TN + FN}$ (23)
Sensibilidade	$SEN = \frac{TP}{TP + FN}$ (24)
Especificidade	$SPE = \frac{TN}{TN + FP}$ (25)
Coefficiente Dice	$D = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}$ (26)
Coefficiente Jaccard	$J = \frac{TP}{TP + FP + FN}$ (27)
Score F ₁	$F1 = 2 \cdot \frac{SEN \cdot PRE}{SEN + PRE}$ (28)

Onde PRE é a precisão e pode ser definido como:

$$PRE = \frac{TP}{TP + FP} \quad (29)$$

Embora a avaliação com ambos os coeficientes Dice e Jaccard não adicionem informações complementares entre si, já que os dois medem similaridade de conjuntos e, inclusive, há uma relação entre eles onde:

$$J = \frac{D}{2 - D} \quad (30)$$

Ambos os coeficientes estão presentes no objetivo de comparar este trabalho com outros que foram avaliados somente em uma das duas métricas.

4.4.2 Métricas utilizadas na classificação de região de bifurcação

As seguintes métricas foram utilizadas para a análise do desempenho na tarefa de classificação de região de bifurcação: taxa de falso positivo (FPR), taxa de verdadeiro positivo ou sensibilidade (SEN), acurácia (ACC), especificidade (SPE), precisão (PRE), taxa de alarme falso (FA), *score* F1 (F1) e área sob a curva ROC (AUC).

Consideramos como resultado positivo a classe bifurcação e negativo a classe não-bifurcação. Assim, as métricas são definidas Quadro 4, onde TP = verdadeiro positivo, TN = verdadeiro negativo, FP = falso positivo, FN = falso negativo, f é a curva ROC e x é a taxa de falso positivo.

Quadro 4 – Métricas utilizadas na avaliação da classificação de região de bifurcação.

Taxa de falso positivo	$FPR = \frac{FP}{TN + FP} \quad (31)$
Sensibilidade	$SEN = \frac{TP}{TP + FN} \quad (32)$
Acurácia	$ACC = \frac{TP + TN}{TP + FP + TN + FN} \quad (33)$
Especificidade	$SPE = \frac{TN}{TN + FP} \quad (34)$
Precisão	$PRE = \frac{TP}{TP + FP} \quad (35)$
Taxa de alarme falso	$FA = \frac{FP}{TP + FN} \quad (36)$
Score F ₁	$F1 = 2 \cdot \frac{SEN \cdot PRE}{SEN + PRE} \quad (37)$
Área sob a Curva ROC	$AUC = \int_0^1 f(x)dx \quad (38)$

5 RESULTADOS E DISCUSSÕES

5.1 RESULTADOS DA SEGMENTAÇÃO DO LÚMEN

Nessa seção, são apresentados os resultados das três arquiteturas propostas testadas em cada uma das cinco bases de dados apresentadas na seção 4.1. As Tabelas 7–13 mostram a média e o desvio padrão dos resultados obtidos em cada base de dados utilizada na fase experimental. A Figura 24 mostra, de forma consolidada e comparativa, os resultados de todos os experimentos em relação às métricas apresentadas no Quadro 3.

A Tabela 7 mostra a média e o desvio padrão dos resultados obtidos com a base de dados BD-S1. Como dito anteriormente, o experimento S3 não pôde ser feito por insuficiência de memória no *hardware* utilizado.

Tabela 7 – Resultados obtidos na segmentação do lúmen da base de dados BD-S1.

Experimento	ACC ⁽¹⁾ (%)	SEN ⁽²⁾ (%)	SPE ⁽³⁾ (%)	D ⁽⁴⁾ (%)	J ⁽⁵⁾ (%)	F1 ⁽⁶⁾ (%)
S1	98,69 ± 2,38	97,45 ± 3,84	98,77 ± 2,60	93,42 ± 12,05	89,35 ± 15,16	83,46 ± 22,46
S2	98,93 ± 2,07	96,92 ± 3,75	99,11 ± 2,19	94,34 ± 11,71	90,80 ± 13,92	88,88 ± 17,34
S3 ⁽⁷⁾	–	–	–	–	–	–

⁽¹⁾ACC = Acurácia. ⁽²⁾SEN = Sensibilidade. ⁽³⁾SPE = Especificidade. ⁽⁴⁾D = Coeficiente Dice. ⁽⁵⁾J = Coeficiente Jaccard. ⁽⁶⁾F1 = Score F₁. ⁽⁷⁾Não houve experimento.

Tabela 8 – Resultados obtidos na segmentação do lúmen da base de dados BD-S2.

Experimento	ACC ⁽¹⁾ (%)	SEN ⁽²⁾ (%)	SPE ⁽³⁾ (%)	D ⁽⁴⁾ (%)	J ⁽⁵⁾ (%)	F1 ⁽⁶⁾ (%)
S4	99,21 ± 1,63	98,38 ± 4,40	99,30 ± 1,75	95,30 ± 8,52	91,96 ± 11,17	88,27 ± 19,46
S5	99,35 ± 1,57	97,71 ± 3,30	99,55 ± 1,47	95,35 ± 10,09	92,30 ± 11,92	92,62 ± 13,54
S6	98,81 ± 2,68	97,59 ± 4,37	98,99 ± 2,67	92,50 ± 15,36	88,53 ± 17,66	88,16 ± 17,94

⁽¹⁾ACC = Acurácia. ⁽²⁾SEN = Sensibilidade. ⁽³⁾SPE = Especificidade. ⁽⁴⁾D = Coeficiente Dice. ⁽⁵⁾J = Coeficiente Jaccard. ⁽⁶⁾F1 = Score F₁.

Tabela 9 – Resultados obtidos na segmentação do lúmen da base de dados BD-S3.

Experimento	ACC ⁽¹⁾ (%)	SEN ⁽²⁾ (%)	SPE ⁽³⁾ (%)	D ⁽⁴⁾ (%)	J ⁽⁵⁾ (%)	F1 ⁽⁶⁾ (%)
S7	99,45 ± 0,70	98,19 ± 3,27	99,65 ± 0,67	95,67 ± 4,71	92,06 ± 7,63	90,63 ± 15,81
S8	99,57 ± 0,62	97,08 ± 3,45	99,38 ± 0,53	96,73 ± 3,23	93,83 ± 5,18	94,85 ± 9,68
S9	99,18 ± 1,81	97,37 ± 5,16	99,47 ± 1,50	93,42 ± 13,98	89,82 ± 16,19	89,34 ± 17,33

⁽¹⁾ACC = Acurácia. ⁽²⁾SEN = Sensibilidade. ⁽³⁾SPE = Especificidade. ⁽⁴⁾D = Coeficiente Dice. ⁽⁵⁾J = Coeficiente Jaccard. ⁽⁶⁾F1 = Score F₁.

Tabela 10 – Resultados obtidos na segmentação do lúmen da base de dados BD-S4.

Experimento	ACC ⁽¹⁾ (%)	SEN ⁽²⁾ (%)	SPE ⁽³⁾ (%)	D ⁽⁴⁾ (%)	J ⁽⁵⁾ (%)	F1 ⁽⁶⁾ (%)
S10	99,48 ± 0,63	98,29 ± 3,10	99,66 ± 0,62	95,82 ± 4,44	92,30 ± 7,29	91,04 ± 15,04
S11	99,59 ± 0,49	96,89 ± 3,21	99,86 ± 0,42	96,76 ± 2,90	93,86 ± 4,70	95,07 ± 8,87
S12	99,35 ± 1,31	97,65 ± 3,33	99,54 ± 1,28	94,40 ± 12,14	91,01 ± 14,10	89,82 ± 16,45

⁽¹⁾ACC = Acurácia. ⁽²⁾SEN = Sensibilidade. ⁽³⁾SPE = Especificidade. ⁽⁴⁾D = Coeficiente Dice. ⁽⁵⁾J = Coeficiente Jaccard. ⁽⁶⁾F1 = Score F₁.

Tabela 11 – Resultados obtidos na segmentação do lúmen da base de dados BD-S5.

Experimento	ACC ⁽¹⁾ (%)	SEN ⁽²⁾ (%)	SPE ⁽³⁾ (%)	D ⁽⁴⁾ (%)	J ⁽⁵⁾ (%)	F1 ⁽⁶⁾ (%)
S13	99,42 ± 1,00	99,66 ± 1,48	99,33 ± 1,19	98,74 ± 1,55	97,57 ± 2,81	95,35 ± 7,76
S14	99,57 ± 0,96	99,47 ± 1,46	99,62 ± 1,08	98,72 ± 1,43	97,51 ± 2,58	97,09 ± 4,66
S15	99,52 ± 1,10	99,53 ± 1,29	99,53 ± 1,37	98,63 ± 1,70	97,35 ± 2,99	96,66 ± 5,27

⁽¹⁾ACC = Acurácia. ⁽²⁾SEN = Sensibilidade. ⁽³⁾SPE = Especificidade. ⁽⁴⁾D = Coeficiente Dice. ⁽⁵⁾J = Coeficiente Jaccard. ⁽⁶⁾F1 = Score F₁.

Tabela 12 – Resultados obtidos na segmentação do lúmen da base de dados BD-S6.

Experimento	ACC ⁽¹⁾ (%)	SEN ⁽²⁾ (%)	SPE ⁽³⁾ (%)	D ⁽⁴⁾ (%)	J ⁽⁵⁾ (%)	F1 ⁽⁶⁾ (%)
S16	99,47 ± 8,50	99,58 ± 1,55	99,45 ± 0,79	98,80 ± 1,26	97,65 ± 2,32	96,96 ± 4,92
S17	99,59 ± 0,90	99,15 ± 1,90	99,81 ± 0,41	98,71 ± 1,24	97,48 ± 2,25	97,04 ± 4,21
S18	99,60 ± 1,04	99,24 ± 1,90	99,79 ± 0,72	98,70 ± 1,41	97,48 ± 2,53	97,32 ± 4,36

⁽¹⁾ ACC = Acurácia. ⁽²⁾ SEN = Sensibilidade. ⁽³⁾ SPE = Especificidade. ⁽⁴⁾ D = Coeficiente Dice. ⁽⁵⁾ J = Coeficiente Jaccard. ⁽⁶⁾ F1 = Score F₁.

Tabela 13 – Resultados obtidos na segmentação do lúmen da base de dados BD-S7.

Experimento	ACC ⁽¹⁾ (%)	SEN ⁽²⁾ (%)	SPE ⁽³⁾ (%)	D ⁽⁴⁾ (%)	J ⁽⁵⁾ (%)	F1 ⁽⁶⁾ (%)
S19	99,49 ± 0,81	99,60 ± 1,44	99,47 ± 0,79	98,82 ± 1,23	97,70 ± 2,25	97,09 ± 4,82
S20	99,60 ± 0,85	99,15 ± 1,81	99,81 ± 0,38	98,71 ± 1,19	97,48 ± 2,15	97,01 ± 4,12
S21	99,59 ± 1,14	99,17 ± 2,01	99,80 ± 0,88	98,68 ± 1,50	97,44 ± 2,64	97,22 ± 4,34

⁽¹⁾ ACC = Acurácia. ⁽²⁾ SEN = Sensibilidade. ⁽³⁾ SPE = Especificidade. ⁽⁴⁾ D = Coeficiente Dice. ⁽⁵⁾ J = Coeficiente Jaccard. ⁽⁶⁾ F1 = Score F₁.

De forma geral, os melhores resultados são obtidos com as bases de dados de imagens de menor tamanho (BD-S3, BD-S4, BD-S6 e BD-S7). Embora a rede CNN-S3 seja mais sofisticada que as outras duas redes, o desempenho dela é, em geral, pior do que o desempenho da rede CNN-S2.

Embora a conversão para coordenadas polares não seja uma prática adotada em CNNs, já que esta conversão pode ser operacionalizada intrinsecamente pela própria rede durante o processo de treinamento, os resultados apresentaram melhor desempenho quando comparados com suas contrapartidas em coordenadas cartesianas. Converter previamente a imagem para coordenadas polares possivelmente beneficia o treinamento da mesma maneira que em técnicas de processamento de imagens tradicionais, onde a conversão aumenta a eficiência na detecção do lúmen ao simplificar os descritores de imagem.

Além disso, imagens IVOCT utilizam pseudo-cores, geralmente tons sépia, para mapear a intensidade do sinal. Portanto, uma prática comum é converter a imagem em tons de cinza antes de manipulá-las. Em termos de segmentação, a diferença de desempenho entre os experimentos com bases de dados de diferentes espaços de cores (BD-S3 contra BD-S4 e BD-S6 contra BD-S7) é negligenciável. Embora este resultado fosse previsível no sentido de que pseudo-cores não carregam informações adicionais que os tons de cinza, certas anotações presentes na imagem geradas pelo dispositivo de imageamento são coloridas, portanto, poderia haver algum benefício neste aspecto.

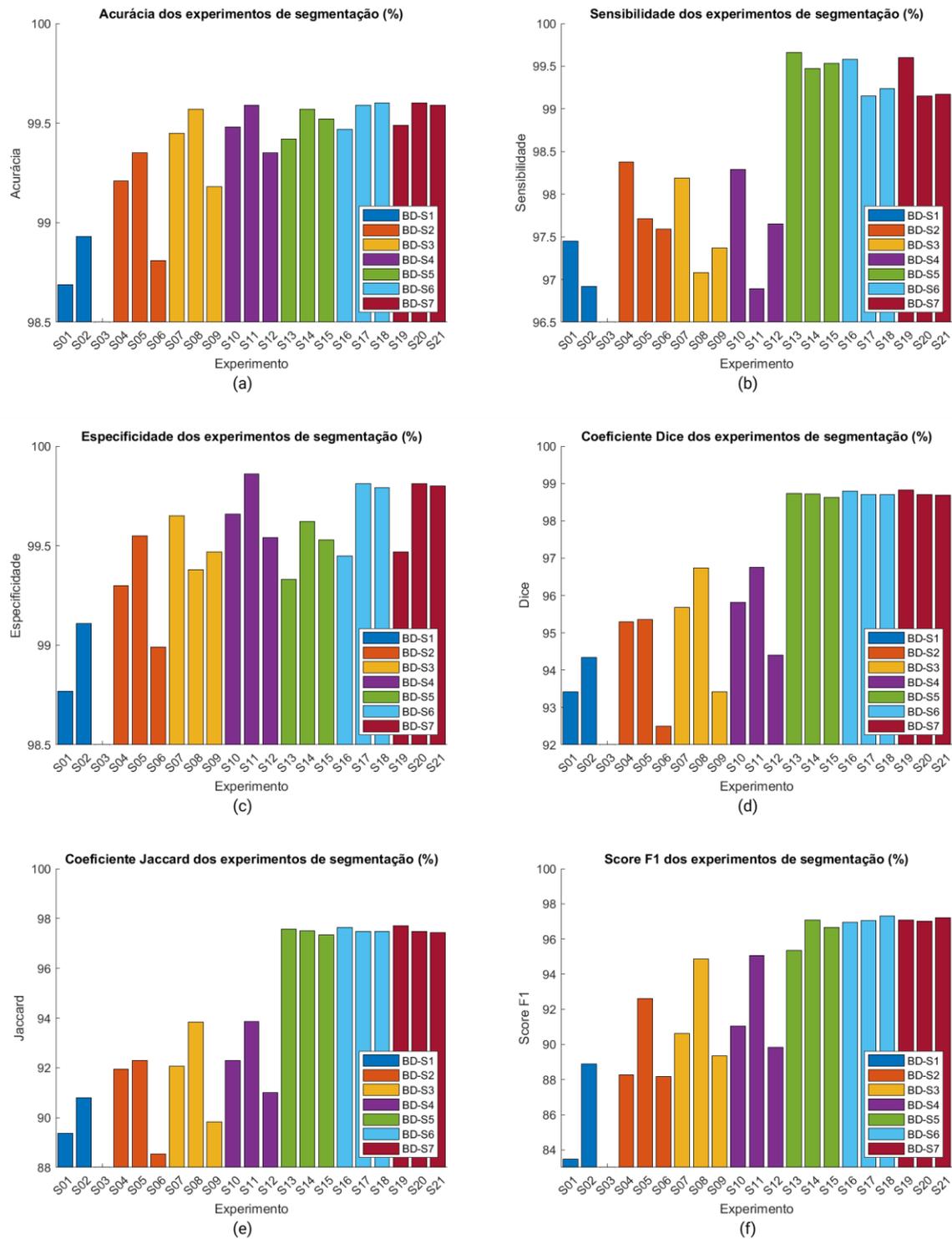


Figura 24 – Comparativo dos experimentos de classificação de região de bifurcação em relação a cada métrica. (a) Acurácia, (b) Sensibilidade, (c) Especificidade, (d) Coeficiente Dice, (e) Coeficiente Jaccard e (f) Score F1.

Na Tabela 14 comparamos os resultados de trabalhos anteriores com o experimento S17 deste trabalho, selecionado pelo seu melhor desempenho geral dentre os experimentos com

imagens em cores e de arquitetura CNN-S2. Esta arquitetura foi utilizada posteriormente para *transfer learning* na tarefa de classificação.

Embora uma comparação direta só possa ser feita com Macedo et al. (2015), já que compartilhamos o mesmo conjunto de dados, os resultados obtidos, em geral, são comparáveis aos encontrados na literatura, apresentando resultados próximos aos de Macedo et al. (2015), com acurácia e coeficiente Dice 99,88% e 97,4%, respectivamente, e Kim et al. (2014), com sensibilidade = 99,2% e especificidade = 99,7%.

O coeficiente Dice é uma métrica presente em quase todos os trabalhos, de forma que podemos comparar outros artigos em função da mesma. Além de Macedo et al. (2015), o coeficiente Dice obtido também é superior aos de Moraes, Cardenas e Furuie (2013), Cao et al.(2017) e Yong et al.(2017), que obtiveram valores de 97,8%, 98,1% e 97%, respectivamente.

Devemos destacar que as imagens do experimento tiveram sua resolução espacial reduzida e, portanto, apresentam menores detalhes se comparados às imagens de tamanho original. Esta redução pode prejudicar a utilização da imagem segmentada em algumas aplicações, no sentido de que os detalhes perdidos podem dificultar a tarefa de estimar as dimensões de um *stent* para implantação, ou reduzir a visibilidade de texturas que podem caracterizar certos tipos de lesões, por exemplo.

Tabela 14 – Comparativo dos resultados de trabalhos anteriores de segmentação do lúmen com o deste trabalho.

Artigo	ACC ⁽¹⁾ (%)	SEN ⁽²⁾ (%)	SPE ⁽³⁾ (%)	D ⁽⁴⁾ (%)	J ⁽⁵⁾ (%)	F1 ⁽⁶⁾ (%)
Tsantis et al. (2012)	–	–	–	–	93,7 ± 4,5	–
Moraes, Cardenas e Furuie (2013)	–	–	–	97,8 ± 2,16	95,4 ± 4,8	–
Kim et al. (2014)	–	99,2	99,7	–	–	–
Macedo et al. (2015)	99,88 ± 0,25	–	–	97,4 ± 2,67	–	–
Cao et al. (2017)	–	–	–	98,1 ± 1,1	–	–
Yong et al. (2017)	–	–	–	97	98,5	–
Este trabalho	99,59 ± 0,90	99,15 ± 1,90	99,81 ± 0,41	98,71 ± 1,24	97,48 ± 2,25	97,04 ± 4,21

⁽¹⁾ ACC = Acurácia. ⁽²⁾ SEN = Sensibilidade. ⁽³⁾ SPE = Especificidade. ⁽⁴⁾ D = Coeficiente Dice. ⁽⁵⁾ J: Coeficiente Jaccard. ⁽⁶⁾ F1: Score F₁.

A Figura 25 mostra o resultado da segmentação de uma imagem em cada um dos experimentos. A linha verde indica o padrão-ouro, enquanto a área em azul indica o resultado da classificação rotulada como lúmen. Observamos que houve poucos erros na segmentação

de imagens em coordenadas polares e nas imagens da base BD-S3. Ao contrário do que foi observado com as imagens da base BD-S1. Acreditamos que este tipo de erro possa ser evitado utilizando *kernels* maiores nas camadas de convolução.

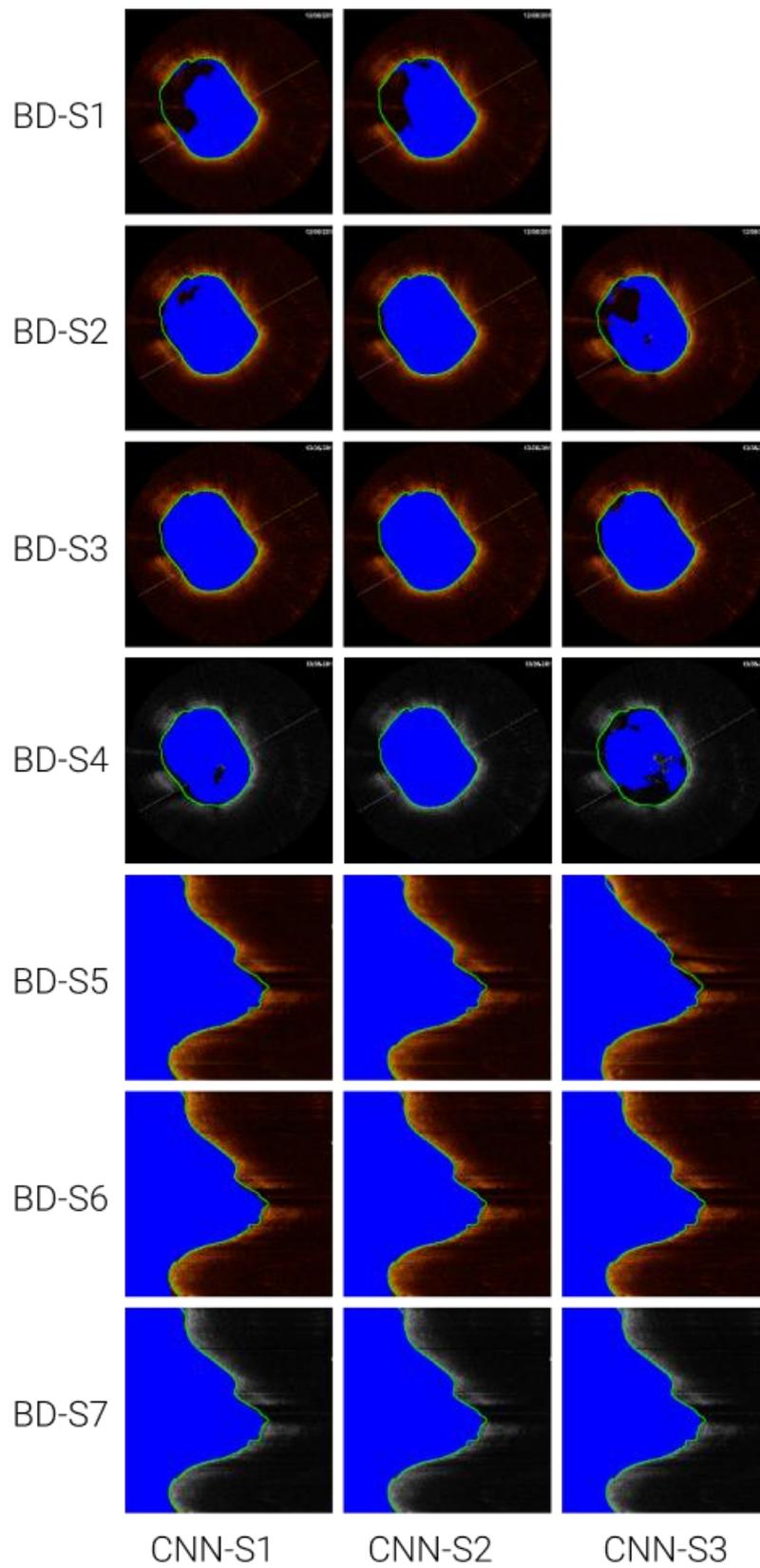


Figura 25 – O resultado da segmentação do lúmen de uma imagem IVOCT em cada experimento. A linha verde representa o padrão-ouro e a área azul representa o resultado da segmentação.

A Tabela 15 mostra o tempo médio de processamento por imagem nas redes de cada um dos experimentos. O tempo de processamento é, em geral, maior em arquiteturas mais complexas (CNN-S3) e em base de dados de maior resolução. O espaço de cores e de coordenadas pouco influenciaram no tempo de processamento.

Tabela 15 – Tempo médio de processamento por imagem nas redes de cada um dos experimentos na tarefa de segmentação.

Experimento	Tempo de processamento por imagem (ms)
S1	$197,7 \pm 8,7$
S2	$345,6 \pm 25,0$
S3	– ⁽¹⁾
S4	$51,0 \pm 4,2$
S5	$75,4 \pm 6,4$
S6	$82,9 \pm 3,9$
S7	$22,1 \pm 4,4$
S8	$27,0 \pm 2,4$
S9	$34,3 \pm 2,7$
S10	$21,0 \pm 3,7$
S11	$27,8 \pm 6,3$
S12	$32,8 \pm 2,1$
S13	$50,3 \pm 4,6$
S14	$72,2 \pm 4,7$
S15	$84,2 \pm 3,3$
S16	$21,7 \pm 5,2$
S17	$26,5 \pm 2,6$
S18	$33,4 \pm 2,5$
S19	$20,1 \pm 2,4$
S20	$26,8 \pm 3,1$
S21	$32,6 \pm 2,0$

⁽¹⁾ Não houve experimento.

5.2 RESULTADOS DA CLASSIFICAÇÃO DE REGIÃO DE BIFURCAÇÃO

A Tabela 16 mostra os resultados dos experimentos apresentados na Tabela 6 para a classificação de região de bifurcação, quando se utiliza as bases de dados baseadas em coordenadas cartesianas (BD-C1 e BD-C2). A Tabela 17 apresenta os resultados quando se utiliza as bases de dados baseadas em coordenadas polares (BD-C3 e BC-C4). A Figura 28

apresenta as curvas ROC. Por fim, a Figura 29 mostra, de forma consolidada e comparativa, os resultados de todos os experimentos em relação às métricas apresentadas no Quadro 4.

De forma geral, assim como na segmentação, os resultados por experimentos que utilizaram coordenadas polares apresentaram desempenhos ligeiramente superiores se comparados aos que utilizaram coordenadas cartesianas. Porém, as diferenças entre as métricas não foram muito discrepantes. Enquanto isso, o comparativo entre espaços de cores diferentes apresentou resultados melhores para imagens com cores (Experimentos C4 e C9), indicando que, ao contrário da segmentação, por mais que imagens IVOCT utilizem pseudo-cores, elas possuem detalhes importantes para a classificação que são perdidos na conversão para escala de cinza. Acreditamos que isso ocorre por conta de que algumas marcações feitas pelo dispositivo de imageamento são mais facilmente detectadas quando se utiliza uma representação com cores.

Os experimentos feitos com a rede CNN-C2 (C2 e C7) apresentaram os piores resultados, indicando que somente a imagem segmentada não é uma fonte confiável para a classificação de região de bifurcação. Ao contrário, ao unir-se a imagem de entrada com informação de segmentação, como é o caso da rede CNN-S3, obteve-se resultados melhores que a rede CNN-S1, que foi treinada totalmente do início, ou seja, sem *transfer learning*.

Os experimentos feitos com a rede CNN-S4 mostraram os melhores resultados, tanto em coordenadas polares quanto em coordenadas cartesianas, indicando que as características adquiridas por transferência de conhecimento influenciaram positivamente o aprendizado da rede. As suas curvas ROC podem ser comparadas na Figura 28(e–f).

Tabela 16 – Resultados obtidos na classificação de região de bifurcação com bases de dados baseados em coordenadas cartesianas (BD-C1 e BC-C2).

Exp.	FPR ⁽¹⁾ (%)	SEN ⁽²⁾ (%)	ACC ⁽³⁾ (%)	SPE ⁽⁴⁾ (%)	PRE ⁽⁵⁾ (%)	FA ⁽⁶⁾ (%)	F1 ⁽⁷⁾ (%)	AUC ⁽⁸⁾ (%)
C1	3,34 ± 1,74	98,53 ± 1,32	97,60 ± 1,17	98,51 ± 1,42	96,75 ± 1,60	3,32 ± 1,72	99,43 ± 0,28	99,59 ± 0,19
C2	6,39 ± 2,85	95,57 ± 1,58	94,55 ± 1,42	98,41 ± 1,79	93,73 ± 2,77	6,49 ± 3,13	98,81 ± 0,69	99,57 ± 0,28
C3	3,01 ± 1,24	98,06 ± 1,55	97,52 ± 1,19	98,02 ± 1,64	97,02 ± 1,42	3,02 ± 1,48	99,46 ± 0,59	99,72 ± 0,22
C4	2,64 ± 1,42	98,48 ± 0,98	97,93 ± 1,00	98,41 ± 1,11	97,46 ± 1,30	2,58 ± 1,34	99,68 ± 0,31	99,72 ± 0,17
C5	3,24 ± 1,93	98,54 ± 1,29	97,64 ± 1,20	98,50 ± 1,36	96,81 ± 1,85	3,27 ± 1,98	99,51 ± 0,40	99,66 ± 0,15

⁽¹⁾ FPR = Taxa de falso positivo. ⁽²⁾ SEN = Sensibilidade. ⁽³⁾ ACC = Acurácia. ⁽⁴⁾ SPE = Especificidade. ⁽⁵⁾ PRE = Precisão. ⁽⁶⁾ FA = Taxa de alarme falso. ⁽⁷⁾ F1: Score F1. ⁽⁸⁾ Área sob a curva ROC.

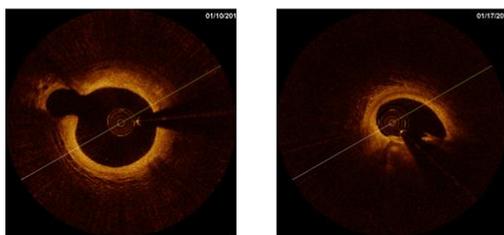
Tabela 17 – Resultados obtidos na classificação de região de bifurcação com bases de dados baseados em coordenadas polares (BD-C3 e BC-C4).

Exp.	FPR ⁽¹⁾ (%)	SEN ⁽²⁾ (%)	ACC ⁽³⁾ (%)	SPE ⁽⁴⁾ (%)	PRE ⁽⁵⁾ (%)	FA ⁽⁶⁾ (%)	F1 ⁽⁷⁾ (%)	AUC ⁽⁸⁾ (%)
C6	3,14 ± 1,50	98,93 ± 0,80	97,89 ± 0,85	98,93 ± 0,75	96,88 ± 1,54	3,21 ± 1,63	99,60 ± 0,37	99,70 ± 0,16
C7	3,77 ± 1,83	97,37 ± 1,32	96,78 ± 1,15	97,23 ± 1,45	96,34 ± 1,73	3,72 ± 1,81	99,23 ± 0,39	99,34 ± 0,24
C8	2,59 ± 1,71	98,56 ± 1,56	97,97 ± 1,42	98,50 ± 1,77	97,42 ± 1,93	2,64 ± 2,04	99,43 ± 0,36	99,62 ± 0,20
C9	2,06 ± 1,11	98,17 ± 1,11	98,05 ± 0,89	98,09 ± 1,19	97,99 ± 1,10	2,02 ± 1,12	99,32 ± 0,37	99,69 ± 0,17
C10	2,70 ± 1,57	98,13 ± 2,04	97,68 ± 1,41	97,98 ± 2,40	97,36 ± 1,50	2,68 ± 1,53	99,40 ± 0,47	99,72 ± 0,12

⁽¹⁾ FPR = Taxa de falso positivo. ⁽²⁾ SEN = Sensibilidade. ⁽³⁾ ACC = Acurácia. ⁽⁴⁾ SPE = Especificidade. ⁽⁵⁾ PRE = Precisão. ⁽⁶⁾ FA = Taxa de alarme falso. ⁽⁷⁾ F1: Score F₁. ⁽⁸⁾ Área sob a curva ROC.

A Figura 26 apresenta algumas imagens incorretamente classificadas pelo classificador do Experimento C4. Aparentemente a rede possui uma maior dificuldade em classificar quando o cateter está posicionado próximo à parede do vaso (Figura 26(b–e)).

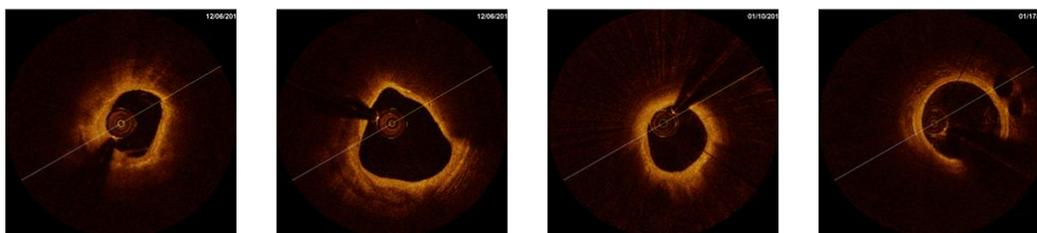
Falso negativo



(a)

(b)

Falso positivo



(c)

(d)

(e)

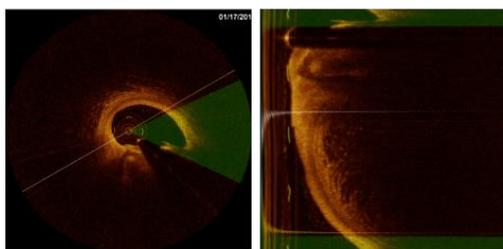
(e)

Figura 26 – Amostras de falsos negativos (a-b) e falsos positivos (c-e) da rede treinada no Experimento C4.

A Figura 27 apresenta algumas imagens incorretamente classificadas pelo classificador do Experimento C9. Observa-se uma similaridade entre as imagens onde o cateter está posicionado mais à esquerda do vaso, gerando uma depressão ao converter a imagem para coordenadas polares, destacado em verde. Acreditamos que, por conta disso, há dificuldade em reconhecer a bifurcação, como mostra a Figura 27(a), e a irregularidade do lúmen colabora para a classificação incorreta na Figura 27(b). Na Figura 27(c), há uma particularidade: a

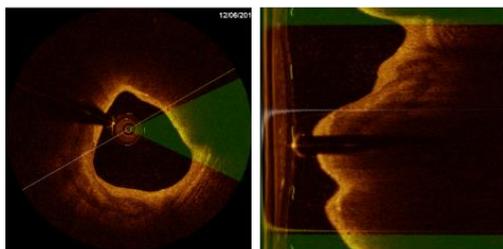
sombra do cateter-guia se localiza sobre a borda da imagem em coordenadas polares, confundindo o classificador. Portanto, ao converter uma imagem em coordenadas polares, a região da imagem que corresponde às bordas da imagem convertida em coordenadas polares, impacta de forma decisiva em uma classificação correta.

Falso negativo

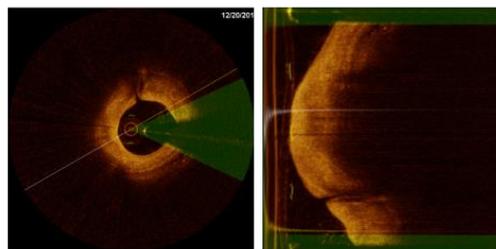


(a)

Falso positivo



(b)



(c)

Figura 27 – Amostras de falso negativo (a) e falsos positivos (b-c) da rede treinada no Experimento C9. A área destacada em verde mostra a porção da imagem localizada sobre a borda ao convertê-la em coordenadas polares.

A Tabela 18 mostra um comparativo entre os trabalhos anteriores na área com os resultados deste trabalho. Os resultados obtidos são comparáveis com os resultados presentes na literatura, especialmente comparados a Macedo et al. (2015) e Porto et al. (2017), já que utilizam o mesmo conjunto de imagens. Os resultados obtidos superam os de Macedo et al. (2015), que utiliza *Adaboost*, em todas as métricas, com exceção da sensibilidade (98,99%). Além disso, observa-se uma redução acentuada do desvio padrão em algumas métricas como o TPR, de 13,87 para 1,42, e SPE, de 13,87 para 1,11. Quando comparado com o método de Porto et al. (2017), os resultados são mais próximos, apresentando melhores resultados para os parâmetros SEN, F1 e AUC, com 98,48%, 99,68% e 99,72%, respectivamente.

Em relação aos trabalhos de Alberti et al. (2012) e Wang et al. (2014), por utilizarem um conjunto de imagens diferentes do nosso, as comparações podem não serem justas, mas o desempenho obtido no presente trabalho foi superior à ambos com a sensibilidade, por exemplo, apresentando 98,48% contra 80,66% e 94,0%, respectivamente. Observamos também que o desvio padrão observado em Alberti et al. (2012) é maior que o apresentado pelo nosso trabalho, apresentando, por exemplo, acentuados 14,90% e 9,28% contra nossos 0,98% e 0,31% em sensibilidade e *score* F₁, respectivamente.

Tabela 18 – Comparativo dos resultados entre trabalhos anteriores de classificação de região de bifurcação com o deste trabalho

Artigo	FPR ⁽¹⁾ (%)	SEN ⁽²⁾ (%)	ACC ⁽³⁾ (%)	SPE ⁽⁴⁾ (%)	PRE ⁽⁵⁾ (%)	FA ⁽⁶⁾ (%)	F1 ⁽⁷⁾ (%)	AUC ⁽⁸⁾ (%)
Alberti et al. (2012)	–	80,66 ± 14,90	95,21 ± 2,28	95,38 ± 2,29	94,60 ± 2,65	4,62 ± 2,29	86,35 ± 9,28	–
Wang et al. (2014)	5,0	94,0	–	–	–	–	–	–
Macedo et al. (2015)	27,48 ± 13,87	98,99 ± 0,67	95,58 ± 2,08	72,52 ± 13,87	96,17 ± 1,86	3,97 ± 1,98	97,51 ± 1,15	96,41 ± 3,18
Porto et al. (2017)	1,7 ± 1,47	97,36 ± 0,81	98,15 ± 0,79	98,83 ± 1,47	98,7 ± 1,56	1,37 ± 1,73	97,93 ± 0,81	99,0 ± 1,0
Este trabalho	2,64 ± 1,42	98,48 ± 0,98	97,93 ± 1,00	98,41 ± 1,11	97,46 ± 1,30	2,58 ± 1,34	99,68 ± 0,31	99,72 ± 0,17

⁽¹⁾ FPR = Taxa de falso positivo. ⁽²⁾ SEN = Sensibilidade. ⁽³⁾ ACC = Acurácia. ⁽⁴⁾ SPE = Especificidade. ⁽⁵⁾ PRE = Precisão. ⁽⁶⁾ FA = Taxa de alarme falso. ⁽⁷⁾ F1: Score F₁. ⁽⁸⁾ Área sob a curva ROC.

A Tabela 19 mostra o tempo médio de processamento por imagem para as redes de cada experimento. As arquiteturas mais profundas, CNN-C2 e CNN-C3, possuem um maior tempo de processamento que as demais arquiteturas. Assim como na segmentação, o espaço de cores e de coordenadas pouco influenciaram no tempo de processamento.

Tabela 19 – Tempo médio de processamento por imagem nas redes de cada um dos experimentos na tarefa de classificação.

Experimento	Tempo médio de processamento por imagem (ms)
C1	5,8 ± 1,8
C2	25,7 ± 1,5
C3	26,6 ± 3,3
C4	13,8 ± 5,1
C5	12,8 ± 3,8
C6	5,9 ± 1,7
C7	25,6 ± 1,3
C8	26,1 ± 1,5
C9	13,4 ± 3,8
C10	12,9 ± 4,0

Em resumo, podemos dizer que o método proposto neste trabalho para classificação obteve melhor desempenho em comparação aos outros trabalhos encontrados na literatura. Essa afirmativa é baseada em uma avaliação do *score* F_1 , que incorpora a média harmônica da sensibilidade e precisão e, portanto, pode ser utilizado como uma métrica para avaliar o desempenho geral do método, já que a sensibilidade, que expressa a proporção de imagens de bifurcação corretamente identificadas como tal, e a precisão, que expressa a proporção de imagens identificadas como bifurcação que foram corretamente classificadas, são as duas métricas mais relevantes em um problema de detecção (ALBERTI et al., 2012). O método proposto obteve um *score* F_1 de 99,68%, contra 97,93% de Porto et al. (2017), 97,51% de Macedo et al. (2015) e 86,35% de Alberti et al. (2012).

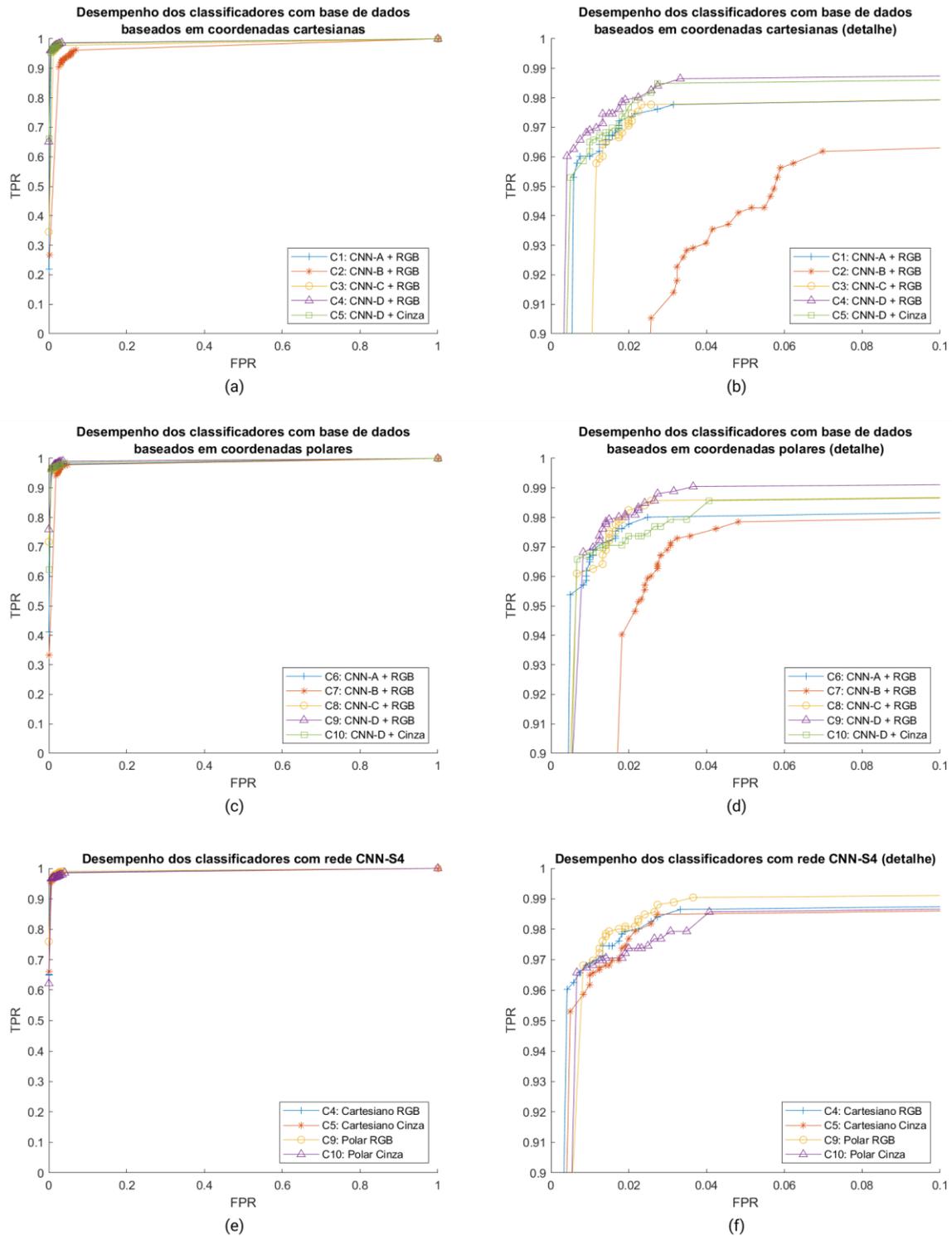


Figura 28 – Curvas ROC dos experimentos de classificação de região de bifurcação. (a-b) Desempenho dos classificadores com base de dados baseados em coordenadas cartesianas. (c-d) Desempenho dos classificadores com base de dados baseados em coordenadas polares. (e-f) Desempenho dos classificadores com rede CNN-S4.

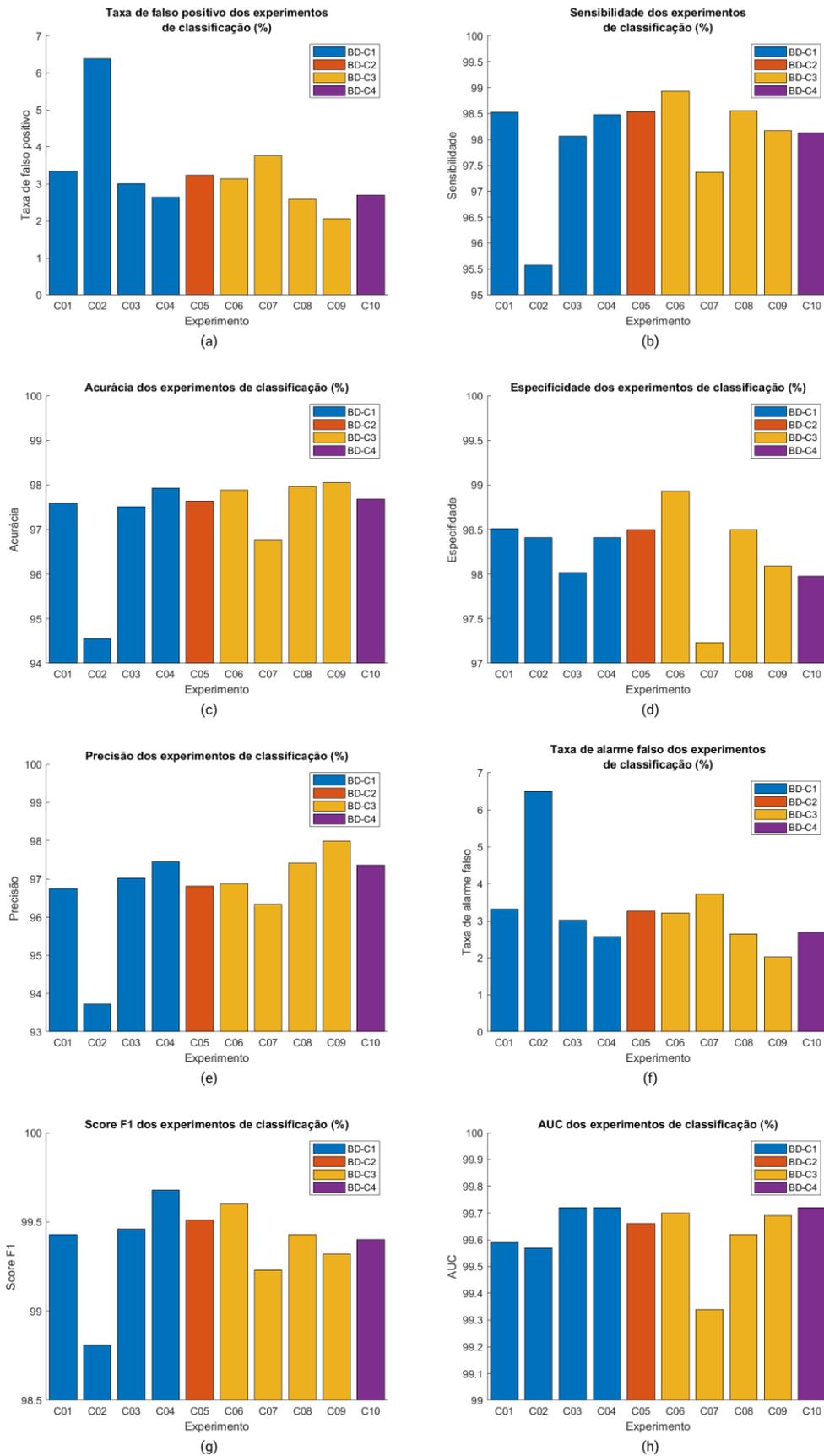


Figura 29 – Comparativo dos experimentos de classificação de região de bifurcação em relação a cada métrica. (a) Taxa de falso positivo, (b) Sensibilidade, (c) Acurácia, (d) Especificidade, (e) Precisão, (f) Taxa de alarme falso, (g) *Score F1* e (h) Área sobre a curva ROC.

6 CONCLUSÕES

A revisão bibliográfica permitiu visualizar o estado da arte no que se refere a tarefas de segmentação automática do lúmen e classificação de região de bifurcação em imagens IVOCT. Confrontando com esse contexto, os resultados desse trabalho permitiram comparar o desempenho do estado da arte com o desempenho de redes neurais convolutivas.

Neste trabalho utilizamos o mesmo conjunto de imagens de Macedo et al. (2015), de forma que é possível comparar de forma justa os resultados obtidos aqui tanto com os mesmos, quanto com os de Porto et al. (2017), que compartilha do mesmo conjunto. Na tarefa de classificação, para evitar que um possível viés de classificação pudesse prejudicar os resultados, foram utilizadas técnicas de *data augmentation* na construção das bases de dados.

No quesito segmentação do lúmen, foram avaliadas três redes convolutivas com sete diferentes bases de dados, num total de 21 experimentos. O tamanho espacial das imagens que compõem a base dados influenciam enormemente no desempenho das redes, de forma que os melhores resultados foram obtidos com as bases com imagens de menor tamanho espacial. A redução nos detalhes, aliado ao tamanho do campo receptivo utilizado nos experimentos ser relativamente grande em imagens menores são possíveis fatores que atuaram no desempenho positivo desses experimentos.

Em relação às coordenadas espaciais das imagens, experimentos que utilizaram bases com imagens em coordenadas polares tiveram, no geral, melhor desempenho que as bases em coordenadas cartesianas. Já em relação ao espaço de cor, não há diferenças acentuadas de desempenho ao utilizar imagens em cores ou em tons de cinza. Também foram feitas comparações entre redes diretas e de grafos acíclicos diretos (DAG). De forma geral, o desempenho de redes DAG foi inferior ao desempenho de redes diretas.

Os resultados obtidos nesta etapa são comparáveis com os encontrados na literatura, apresentando melhores resultados em especificidade e coeficiente Dice, com 99,81% e 98,71%,

respectivamente. Vale notar que, mesmo apresentando bons resultados, a redução das imagens pode prejudicar a utilização da imagem segmentada em algumas aplicações, onde detalhes da imagem, perdidos na etapa de redução, possuem forte influência sobre o desempenho.

A menor resolução das imagens não impede em utilizar, através do *transfer learning*, as redes treinadas na etapa de segmentação, para o auxílio na tarefa de classificação de região de bifurcação. Dessa forma, a transferência de aprendizagem teve forte impacto positivo no desempenho das redes, sendo que os melhores resultados foram oriundos de experimentos com a rede CNN-C4. Ao aumentar a quantidade de imagens utilizadas no treinamento, a técnica de *data augmentation* também teve sua parcela de contribuição no desempenho geral dos experimentos, pois, além de corrigir o viés de classificação, auxiliou numa melhor generalização da rede por conta da maior massa de dados.

Em relação às coordenadas espaciais e espaço de cores, a utilização de coordenadas polares apresentou um desempenho ligeiramente melhor, mas não muito acentuadas, se comparado às coordenadas cartesianas, enquanto que a utilização de imagens em cores, que, embora sejam falsas, propiciou um melhor desempenho frente ao desempenho dos experimentos realizados com imagens em tons de cinza. Os experimentos, em geral, apresentaram um baixo desvio padrão, principalmente se comparados à FPR e SPE de Macedo et al. (2015), por exemplo. Os melhores resultados obtidos neste trabalho comparados com os encontrados na literatura foram *score* F_1 e AUC com 99,68% e 99,72%, respectivamente.

A grande popularidade das redes neurais convolutivas implica no surgimento constante de novas técnicas e arquiteturas. Aplicá-las no contexto abordado neste trabalho, objetivando a busca de melhores desempenhos, pode se tornar uma linha de pesquisa para trabalhos futuros. Uma oportunidade especial seria explorar a necessidade de buscar um melhor desempenho para a tarefa de segmentação, sem a necessidade de reduzir as imagens, preservando as características que seriam perdidas com uma menor resolução espacial.

Outra área de trabalhos futuros seria um estudo aprofundado do *transfer learning*, estudando os limites da influência de redes treinadas em outras arquiteturas ou na tarefa de segmentação, por exemplo, já que foi um tópico não explorado neste trabalho. Devemos lembrar também que é possível receber conhecimento inclusive de domínios não necessariamente voltados às imagens IVOCT. Portanto, uma possível área a ser estudada também seria a utilização de conhecimentos de outras aplicações para o uso neste contexto.

Outro tópico explorável seria a utilização de GAN's para *data augmentation*. Um estudo de viabilidade poderia ser conduzido para verificar a validade da utilização dessas redes para a geração imagens de bifurcação, por exemplo.

Os resultados aqui apresentados pelas redes neurais convolutivas sugerem que, à semelhança do que já ocorre em outras áreas de aplicações das redes convolutivas, as mesmas se tornarão a principal referência no que tange às tarefas de segmentação do lúmen e classificação de região de bifurcação em imagens IVOCT.

REFERÊNCIAS

- ALBERTI, M. et al. Automatic Bifurcation Detection in Coronary IVUS Sequences. **IEEE Transactions of Biomedical Engineering**, v. 59, n. 4, p. 1022–1031, 2012.
- BISHOP, C. M. Neural networks and their applications. **Rev. Sci. Instrm.**, v. 65, n. 6, p. 1803–1832, 1994.
- BISHOP, C. M. **Pattern Recognition and Machine Learning**. 1. ed. New York: Springer, 2006.
- BOTTOU, L. Large-Scale Machine Learning with Stochastic Gradient Descent. **Proceedings of COMPSTAT'2010**, p. 177–186, 2010.
- CAO, Y. et al. Automatic Lumen Segmentation in Intravascular Optical Coherence Tomography Images Using Level Set. **Computational and Mathematical Methods in Medicine**, v. 2017, p. 4710305:1-4710305:11, 2017.
- DENG, L.; YU, D. Deep Learning: Methods and Applications. **Foundations and Trends® in Signal Processing**, v. 7, n. 3–4, p. 197–387, 2013.
- ESTEVA, A. et al. Dermatologist-level classification of skin cancer with deep neural networks. **Nature**, v. 542, n. 7639, p. 115–118, 2017.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. Deep learning. **Healthcare Informatics Research**, v. 22, n. 4, p. 351–354, 2016.
- GUHA ROY, A. et al. Lumen Segmentation in Intravascular Optical Coherence Tomography Using Backscattering Tracked and Initialized Random Walks. **IEEE Journal of Biomedical and Health Informatics**, v. 20, n. 2, p. 606–614, 2016.
- HAYKIN, S. **Neural Networks and Learning Machines**. 3. ed. New Jersey: Pearson Education, 2008.
- IOFFE, S.; SZEGEDY, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. 2015.
- ISIC. **International Skin Imaging Collaboration: Melanoma Project Website**. Disponível em: <<https://www.isic-archive.com>>. Acesso em: 30 set. 2018.

JUNQUEIRA, L. C. U.; CARNEIRO, J. **Histologia Básica**. 12. ed. Rio de Janeiro: Editora Guanabara Koogan LTDA, 2013.

KIM, H. M. et al. Automatic lumen contour detection in intravascular OCT images using Otsu binarization and intensity curve. **2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society**, p. 178–181, 2014.

KRIESEL, D. **A Brief Introduction to Neural Networks**. Disponível em: <http://www.dkriesel.com/en/science/neural_networks>. Acesso em: 31 jul. 2018.

KRIZHEVSKY, A.; SUTSKEVER, I.; GEOFFREY E., H. ImageNet Classification with Deep Convolutional Neural Networks. **Advances in Neural Information Processing Systems 25 (NIPS2012)**, p. 1–9, 2012.

LECUN, Y. et al. Backpropagation Applied to Handwritten Zip Code Recognition. **Neural Computation**, v. 1, n. 4, p. 541–551, 1989.

LECUN, Y. et al. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, v. 86, n. 11, p. 2278–2323, 1998.

LONG, J.; SHELHAMER, E.; DARRELL, T. Fully convolutional networks for semantic segmentation. **Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition**, v. 07–12–June, p. 3431–3440, 2015.

MACEDO, M. M. G. et al. A bifurcation identifier for IV-OCT using orthogonal least squares and supervised machine learning. **Computerized Medical Imaging and Graphics**, v. 46, p. 237–248, 2015.

MENDIS, S.; PUSKA, P.; NORRVING, B. (EDS.). **Global Atlas on Cardiovascular disease prevention and control**. [s.l.] World Health Organization, 2011.

MENGUY, P. Y. et al. Preliminary results for the supervised detection of lumen and stent from OCT pullbacks. **Irbm**, v. 37, n. 5–6, p. 271–275, 2016.

MORAES, M. C.; CARDENAS, D. A. C.; FURUIE, S. S. Automatic lumen segmentation in IVOCT images using binary morphological reconstruction. **BioMedical Engineering Online**, v. 12, n. 1, p. 1–17, 2013.

NEW BEGINNINGS HEALTH CARE. **What is atherosclerosis?** Disponível em: <<http://www.newbeginningshealthcare.net/articles/cardiovascular-risk-and-support/atherosclerosis/>>. Acesso em: 30 jul. 2018.

NOH, H.; HONG, S.; HAN, B. Learning deconvolution network for semantic segmentation. **Proceedings of the IEEE International Conference on Computer Vision**, v. 2015 Inter, p. 1520–1528, 2015.

PAN, S. J.; YANG, Q. A survey on transfer learning. **IEEE Transactions on Knowledge and Data Engineering**, v. 22, n. 10, p. 1345–1359, 2010.

PEREZ, L.; WANG, J. The Effectiveness of Data Augmentation in Image Classification using Deep Learning. **arXiv preprint arXiv:1712.04621**, 2017.

PORTO, C. D. N. et al. Classification of bifurcations regions in IVOCT images using support vector machine and artificial neural network models. **Progress in Biomedical Optics and Imaging - Proceedings of SPIE**, v. 10134, n. March 2017, 2017.

PRATI, F. et al. Intracoronary optical coherence tomography, basic theory and image acquisition techniques. **Int J Cardiovasc Imaging**, v. 27, n. 2, p. 251–8, 2011.

RADAU, P. et al. Evaluation Framework for Algorithms Segmenting Short Axis Cardiac MRI. **The MIDAS Journal: Cardiac MR Left Ventricle Segmentation Challenge**, 2009.

RATHOD, K. S. et al. Diagnosis & Imaging Intravascular Ultrasound Versus Optical Coherence Tomography for Coronary Artery Imaging – Apples and Oranges? Intravascular Ultrasound Versus Optical Coherence Tomography for Coronary Artery Imaging. **Interventional Cardiology Review**, v. 10, n. 1, p. 2–9, 2015.

ROLEDER, T. et al. The basics of intravascular optical coherence tomography. **Postepy w Kardiologii Interwencyjnej**, v. 11, n. 2, p. 74–83, 2015.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. **CoRR**, v. abs/1505.0, 2015.

SIMONYAN, K.; ZISSERMAN, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. p. 1–14, 2014.

SRIVASTAVA, N. et al. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. **Journal of Machine Learning Research**, v. 15, p. 1929–1958, 2014.

STUTZ, D. Understanding Convolutional Neural Networks. **Nips 2016**, n. 3, p. 1–23, 2014.

SURI, J. S.; KATHURIA, C.; MORILARI, F. (EDS.). **Atherosclerosis Disease Management**. 1st. ed. New York: Springer, 2011.

SZEGEDY, C. et al. Going deeper with convolutions. **Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition**, v. 07–12–June, p. 1–9, 2015.

TAN, L. K. et al. Convolutional neural network regression for short-axis left ventricle segmentation in cardiac cine MR sequences. **Medical Image Analysis**, v. 39, p. 78–86, 2017.

TSANTIS, S. et al. Automatic vessel lumen segmentation and stent strut detection in intravascular optical coherence tomography. **Medical Physics**, v. 39, n. 1, p. 503–513, 2012.

VASCONCELOS, C. N.; VASCONCELOS, B. N. Convolutional Neural Network Committees for Melanoma Classification with Classical And Expert Knowledge Based Image Transforms Data Augmentation. **CoRR**, [abs/1702.07025](https://arxiv.org/abs/1702.07025), v. 1, 2017.

WANG, A. et al. Fully automated side branch detection in intravascular optical coherence tomography pullback runs. **Biomedical Optics Express**, v. 5, n. 9, p. 3160, 2014.

WANG, A.; TANG, X. Automatic segmentation of vessel lumen in intravascular optical coherence tomography images. **2016 IEEE International Conference on Mechatronics and Automation, IEEE ICMA 2016**, p. 948–953, 2016.

WHO. **Cardiovascular diseases (CVDs)**. Disponível em: <http://www.who.int/mediacentre/factsheets/fs317/en/>. Acesso em: 13 jun. 2017.

XU, M. et al. Graph Based Lumen Segmentation in Optical Coherence Tomography Images. **2015 10th International Conference on Information, Communications and Signal Processing (ICICS)**, p. 1–5, 2015.

YONG, Y. L. et al. Linear-regression convolutional neural network for fully automated coronary lumen segmentation in intravascular optical coherence tomography. **Journal of**

Biomedical Optics, v. 22, n. 12, p. 1, 2017.

YOSINSKI, J. et al. How transferable are features in deep neural networks? v. 27, 2014.

ZARINS, C. et al. Carotid Bifurcation Atherosclerosis: Quantitative Correlation of Plaque Localization with Flow Velocity Profiles and Wall Shear Stress. **Circ. Res.**, v. 53, p. 502–514, 1983.

ZARINS, C. K.; GIDDENS, D. P.; GLAGOV, S. Atherosclerotic plaque distribution and flow velocity profiles in the carotid bifurcation. In: **Cerebrovascular insufficiency**. [s.l.] Grune and Stratton, New York, 1983. p. 19–30.

7 APÊNDICE I

Lumen Segmentation in Optical Coherence Tomography Images using Convolutional Neural Network

M. Miyagawa, M. G. F. Costa, *Member, IEEE*, M. A. Gutierrez, *J. P. G. F. Costa* and C. F. F. Costa Filho, *Member, IEEE*

Abstract— Lumen segmentation in Optical Coherence Tomography (OCT) images is a very important step to analyze points of interest that may help on atherosclerosis diagnostic and treatment. Past studies use many different methods to segment the lumen in IVOCT images, like level set, morphological reconstruction, Markov random fields, and Otsu binarization. Despite Convolutional Neural Networks (CNN) have shown promising results in the image processing area, we did not identify, in the literature, works applying CNN in IVOCT images. In this paper, we present the lumen segmentation using CNN. We evaluated three different CNN architectures. The CNNs were evaluated using three versions from the image dataset, differing from each other by image size (768x768 pixels and 192x192 pixels), and by coordinate system representation (Cartesian and polar). The best results, Accuracy, Dice index and Jaccard index of over 99%, 98% and 97%, respectively, were obtained with the smallest size images represented by polar coordinate system.

I. INTRODUCTION

Atherosclerosis, one of the main responsible for heart attacks and strokes, is a pathological process where cholesterol builds up inside the vessels wall, narrowing the blood flow and occasioning lesions (atheroma), and even full obstruction of the vessel [1]. Analyzing images from intravascular optical coherence tomography (IVOCT) or intravascular ultrasound (IVUS) may help with diagnostics and treatment of this disease, detecting stenosis, atherosclerotic plaques, and vessel bifurcations. Bifurcations are suggested in studies as areas with higher probability of occur lesions [2]. Therefore, lumen segmentation is a very important step to assist the detection of these interest points.

In the last years, the Convolutional Neural Networks (CNN) have shown interesting results in the image segmentation area [3]–[6]. However, there is no work applying CNN for semantic segmentation of IVOCT images. Past studies use many different methods to segment the lumen in IVOCT images, like level set [7], Otsu binarization and intensity curves [8], morphological reconstruction [9], and Markov random fields [10]. More recently, Yong et al. [11], instead of using a per-pixel classification approach to segment the lumen, used linear regression CNN to estimate the radial distances from the center of the catheter to the vessel wall.

This work has the objective of evaluating lumen segmentation in IVOCT images using different CNN

architectures. Additionally, we evaluate the influence of image size and image coordinate system in the results.

The sections of this paper are structured as follows: Section 2 details the used dataset, presents the network architectures and their configurations, and defines the evaluation metrics used. The results and discussion are in Section 3. Finally, in Section 4, we show the conclusions.

The study protocol, under register n. 0243/08., was approved by the Institutional Review Board (CAPPesq) of the Heart Institute (InCor), Clinics Hospital, University of Sao Paulo Medical School (HC FMUSP).

II. METODOLOGY

A. Dataset

The dataset consists in 1,689 RGB IVOCT images from nine pullbacks. This image dataset belongs to Heart Institute, University of Sao Paulo (INCOR-USP). The ground truth is comprised of lumen regions segmented by specialists. Fig. 1(a) shows an image from this dataset. Through a pre-processing, three image sets were obtained from original image dataset. In order to isolate the transversal section of the vessel, we cropped the original images, resulting in images sized 768x768 (set #1), as shown in Fig. 1(b). This first image set was resized to 192x192, by factor of 4 downsampling in both directions, to compose the set #2, as shown in Fig. 1(c). In the third set, the images from set #2 were converted to polar coordinates, Fig. 1(d). The same steps above were applied to the ground truth, generating the sets used in #1, #2, and #3.

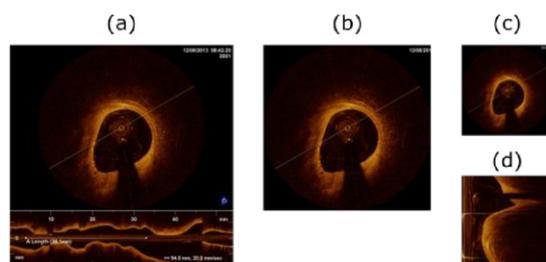


Fig. 1. Image pre-processing steps: (a) original image; (b) set #1: cropped image sized 768x768 pixels; (c) set #2: images of set #1 resized to 192x192 pixels by factor of 4 downsampling; (d) set #3: images of set #2 converted to polar coordinates.

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (2)$$

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (3)$$

$$\text{Dice} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \quad (4)$$

$$\text{Jaccard} = \frac{TP}{TP+FP+FN} \quad (5)$$

$$\text{F1 score} = 2 \cdot \frac{\text{sensitivity} \cdot \text{specificity}}{\text{sensitivity} + \text{specificity}} \quad (6)$$

III. RESULTS AND DISCUSSIONS

Table 1 shows the segmentation results obtained with CNN1 and CNN2, using images from set #1. Table 2 shows the segmentation results obtained with CNN1, CNN2 and CNN3, using images from set #2. Finally, Table 3 shows the segmentation results obtained with CNN1, CNN2 and CNN3, using images from set #3.

In a general way, Table 1 shows the results with worse metric values and large standard deviations (images sized 768x768). Table 3 shows the results with best metric performances (images sized 192x192 in polar coordinates). Although CNN3 have a more sophisticated architecture than the other two networks, the differences between the results of the three networks are not significant. We did not make experiments with CNN3 networks using images of set #1, because of the limited GPU memory size.

Fig. 3 shows segmentation results for two images, image 1 (shown in (a)) and image 2 (shown in (b)), for the three CNN architectures. In Lines (i), (ii), and (iii), these images were sampled from image sets #1, #2, and #3, respectively. As observed, segmentations shown in Fig. 3(b) (i) present errors. The errors observed in Fig. 3(b) (i), that used images sized 768x768, are corrected in Fig. 3(b) (ii) and in Fig. 2(b) (iii), when images sized 192x192 are used. We believe that the

errors shown in Fig. 3(b) (i) could also be avoided if larger filter sizes were used in convolution layers.

Although we cannot do a direct comparison with other works, since the datasets used are different, the results presented here are similar to the ones obtained in other works: sensitivity = 99.2% and specificity = 99.7% in [8]. Jaccard index = 95.7% and Dice index = 97.8% in [9]. Dice index = 98.1% in [7]. Jaccard index = 93.7% in [10]. Dice index = 95.5% and Jaccard index = 97% in [11].

Most of the methods described in the literature use grayscale images and pre-processing techniques to remove alignment marks, speckle noise, and guide wire to obtain an accurate segmentation of the images [7]–[10]. The CNN-based approach proposed in this study does not need to rely on these steps and outperformed the results presented in [11], that used linear regression CNN to estimate the radial distances from the center of the catheter to the vessel wall .

IV. CONCLUSION

In this work, we evaluated the use of CNNs for lumen segmentation of IVOCT images. We trained three architecture networks, two feedforward static networks and one DAG network, and used three images sets: 768x768 images, 192x192 images, and 192x192 polar coordinate images. DAG network shown the same performance of feedforward static networks. The results shown better networks performances when small sized polar coordinates images were used. Although we cannot do a direct comparison with other works, the best results obtained for accuracy, Jaccard index and Dice index were as good as the best results presented in the literature. We believe that the segmentation errors found in some large sized images are due to the small filter size (3x3) used in convolutional layers. To verify this hypothesis, future studies evaluating network segmentation using convolutional layers with larger size filters are necessary.

TABLE I SEGMENTATION RESULTS OBTAINED WITH CNN1 AND CNN2, USING THE IMAGES OF SET #1.

Network	Accuracy	Sensitivity	Specificity	Dice index	Jaccard index	F1 score
CNN1	98.69 ± 2.38	97.45 ± 3.84	98.77 ± 2.60	93.42 ± 12.05	89.35 ± 15.16	83.46 ± 22.46
CNN2	98.93 ± 2.07	96.92 ± 3.75	99.11 ± 2.19	94.34 ± 11.71	90.80 ± 13.92	88.88 ± 17.34

TABLE II SEGMENTATION RESULTS OBTAINED WITH CNN1, CNN2 AND CNN3 USING THE IMAGES OF SET #2.

Network	Accuracy	Sensitivity	Specificity	Dice index	Jaccard index	F1 score
CNN1	99.45 ± 0.70	98.19 ± 3.27	99.65 ± 0.67	95.67 ± 4.71	92.06 ± 7.63	90.63 ± 15.81
CNN2	99.57 ± 0.62	97.08 ± 3.45	99.38 ± 0.53	96.73 ± 3.23	93.83 ± 5.18	94.85 ± 9.68
CNN3	99.18 ± 1.81	97.37 ± 5.16	99.47 ± 1.50	93.42 ± 13.98	89.82 ± 16.19	89.34 ± 17.33

TABLE III SEGMENTATION RESULTS OBTAINED WITH CNN1, CNN2 AND CNN3, USING THE IMAGES OF SET #3.

Network	Accuracy	Sensitivity	Specificity	Dice index	Jaccard index	F1 score
CNN1	99.47 ± 8.50	99.58 ± 1.55	99.45 ± 0.79	98.80 ± 1.26	97.65 ± 2.32	96.96 ± 4.92
CNN2	99.59 ± 0.90	99.15 ± 1.90	99.81 ± 0.41	98.71 ± 1.24	97.48 ± 2.25	97.04 ± 4.21
CNN3	99.60 ± 1.04	99.24 ± 1.90	99.79 ± 0.72	98.70 ± 1.41	97.48 ± 2.53	97.32 ± 4.36

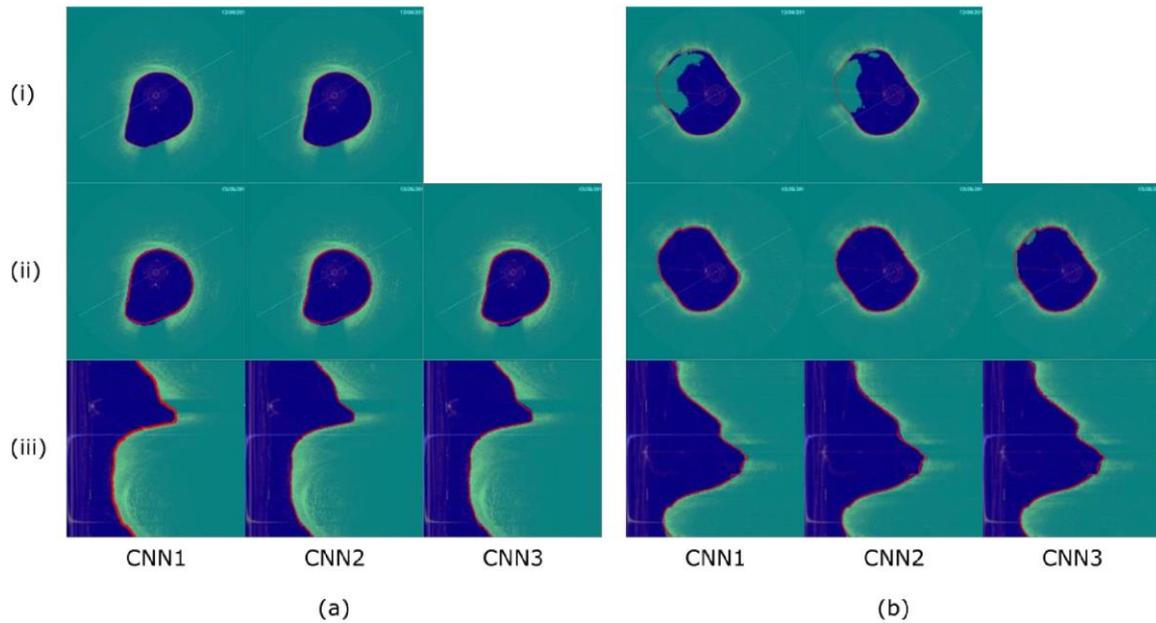


Fig. 3. Segmentation results for two images, image 1 (shown in (a)) and image 2 (shown in (b)), for the three CNN architectures. In Lines (i), (ii) and (iii), these images were sampled from sets #1, #2 and #3, respectively

ACKNOWLEDGMENTS

We would like to thank CNPq (INCT-MACC, 2015-2021), for financial support.

REFERENCES

- [1] P. Puska, B. Norrving, and World Health Organization, "Global Atlas on Cardiovascular disease prevention and control," *Glob. atlas Cardiovasc. Dis. Prev. Control.*, p. 155, 2011.
- [2] M. M. G. de Macedo, W. V. N. Guimaraes, M. Z. Galon, C. K. Takimura, P. A. Lemos, M. A. Gutierrez, "A bifurcation identifier for IV-OCT using orthogonal least squares and supervised machine learning", *Comput Med Imaging Graph*, vol. 46, Pt. 2, pp. 237-248, 2015.
- [3] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June, pp. 3431-3440, 2015.
- [4] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2015 Inter, pp. 1520-1528, 2015.
- [5] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *CoRR*, vol. abs/1505.0, 2015.
- [6] A. Kendall, V. Badrinarayanan, and R. Cipolla, "Bayesian SegNet: Model Uncertainty in Deep Convolutional Encoder-Decoder Architectures for Scene Understanding," *arXiv Prepr. arXiv1511.02680*, 2015.
- [7] Y. Cao *et al.*, "Automatic Lumen Segmentation in Intravascular Optical Coherence Tomography Images Using Level Set," *Comput. Math. Methods Med.*, vol. 2017, p. 4710305:1-4710305:11, 2017.
- [8] H. M. Kim, S. H. wan Lee, C. Lee, J. W. Ha, and Y. R. Yoon, "Automatic lumen contour detection in intravascular OCT images using Otsu binarization and intensity curve," *2014 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pp. 178-181, 2014.
- [9] S. Tsantis, G. C. Kagadis, K. Katsanos, D. Karnabatidis, G. Bourantas, and G. C. Nikiiforidis, "Automatic vessel lumen segmentation and stent strut detection in intravascular optical coherence tomography," *Med. Phys.*, vol. 39, no. 1, pp. 503-513, 2012.
- [10] Y. L. Yong, L. K. Tan, R. A. McLaughlin, K. H. Chee, and Y. M. Liew, "Linear-regression convolutional neural network for fully automated coronary lumen segmentation in intravascular optical coherence tomography," *J. Biomed. Opt.*, vol. 22, no. 12, p. 1, 2017.
- [11] L. K. Tan, Y. M. Liew, E. Lim, and R. A. McLaughlin, "Convolutional neural network regression for short-axis left ventricle segmentation in cardiac cine MR sequences," *Med. Image Anal.*, vol. 39, pp. 78-86, 2017.
- [12] L. K. Tan, Y. M. Liew, E. Lim, and R. A. McLaughlin, "Convolutional neural network regression for short-axis left ventricle segmentation in cardiac cine MR sequences," *Med. Image Anal.*, vol. 39, pp. 78-86, 2017.