Rede especialista em segmentação automática da fossa craniana posterior na população pediátrica

José Luiz Maciel Pimenta

Dissertação apresentada
ao
Instituto de Matemática e Estatística
da
Universidade de São Paulo
para
obtenção do título
de
Mestre em Ciências

Programa: Ciência da Computação

Orientador: Prof. Dr. Roberto Marcondes Cesar Junior

O desenvolvimento do trabalho contou com o apoio financeiro de: CAPES, FAPESP #2015/22308-2, #2017/50236-1 e #2022/15304-4, CNPq, FINEP e MCTI PPI-SOFTEX (TIC 13 DOU 01245.010222/2022-44).

São Paulo, Maio de 2023

Rede especialista em segmentação automática da fossa craniana posterior na população pediátrica

Esta é a versão original da dissertação elaborada pelo
candidato José Luiz Maciel Pimenta, tal como
submetida à Comissão Julgadora.

Ficha catalográfica elaborada com dados inseridos pelo(a) autor(a)
Biblioteca Carlos Benjamin de Lyra
Instituto de Matemática e Estatística
Universidade de São Paulo

Maciel Pimenta, José Luiz
Rede especialista em segmentação automática da fossa craniana posterior na população pediátrica / José Luiz Maciel Pimenta; orientador, Roberto Marcondes Cesar Junior. - São Paulo, 2023.
76 p.: il.

Dissertação (Mestrado) - Programa de Pós-Graduação em Ciência da Computação / Instituto de Matemática e Estatística / Universidade de São Paulo.
Bibliografia
Versão original

1. VISÃO COMPUTACIONAL. 2. RESSONÂNCIA MAGNÉTICA. 3. REDES NEURAIS. I. Marcondes Cesar Junior, Roberto. II. Título.

Bibliotecárias do Serviço de Informação e Biblioteca
Carlos Benjamin de Lyra do IME-USP, responsáveis pela
estrutura de catalogação da publicação de acordo com a AACR2:
Maria Lúcia Ribeiro CRB-8/2766; Stela do Nascimento Madruga CRB 8/7534.

Agradecimentos

Gostaria de expressar minha profunda gratidão e reconhecimento a todas as pessoas e instituições que contribuíram para a realização da minha dissertação de mestrado.

Em primeiro lugar, agradeço a Deus por me conceder força, sabedoria e perseverança para superar os desafios e obstáculos que encontrei durante todo o processo.

À equipe do laboratório e do Hospital das Clínicas, agradeço pela disponibilidade, dedicação e colaboração na coleta e fornecimento das imagens que foram fundamentais para o desenvolvimento da minha pesquisa.

De forma especial, ao meu orientador, expresso minha profunda gratidão pelo tempo, conhecimento e orientação valiosa que me proporcionou ao longo deste trabalho. Seus conselhos, críticas construtivas e orientações foram essenciais para que eu pudesse desenvolver minhas ideias e melhorar minha pesquisa.

Agradeço também à Capes, pela concessão de bolsa de estudos e pelo apoio financeiro que permitiram a realização deste trabalho de forma integral.

Por fim, mas não menos importante, expresso minha gratidão à Fapesp, por fornecer recursos e financiamento para a realização deste estudo. Sua contribuição foi essencial para o sucesso da minha pesquisa.

Deixo aqui registrado meu sincero agradecimento a todos os envolvidos nesta jornada, que foi longa e desafiadora, mas que me proporcionou grandes aprendizados e crescimento pessoal e profissional.

Os autores agradecem à FAPESP (grants #2015/22308-2, #2017/50236-1, #2022/15304-4), CNPq, CAPES, FINEP and MCTI PPI-SOFTEX (TIC 13 DOU 01245.010222/2022-44).

Resumo

Maciel, J. L. **Rede especialista em segmentação automática da fossa craniana posterior na população pediátrica**. 2010. X f. Dissertação (Mestrado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2022.

As diferenças entre os encéfalos de adultos e de crianças causam padrões visuais distintos nas imagens adquiridas utilizando a ressonância magnética. Isso se deve, principalmente, por existirem diferentes estágios de mielinização no encéfalo pediátrico. Assim, apesar de haver algoritmos para auxiliar o diagnóstico e o acompanhamento dos pacientes por meio da segmentação das cavidades cranianas, eles são frequentemente incapazes de lidar com a variabilidade interindividual. Além disso, muitas vezes esses algoritmos não são adaptados a casos patológicos e pediátricos. Uma das formas para tentar contornar esses problemas se dá por meio de modelos como as *Fully Convolutional Network* (FCN) e as *Convolutional Neural Network* (CNN), que se tornaram viáveis para tarefas de segmentação volumétrica devido ao alto poder computacional atual e novos métodos de treinamento. Assim, esta pesquisa busca propor um novo método de utilização desses modelos dentro de uma pipeline de segmentação automática da fossa posterior pediátrica, em um treinamento supervisionado empregando diferentes arquiteturas. De forma mais específica, o método proposto busca utilizar os conceitos de rede generalista e rede especialista, na qual a primeira faz uma segmentação inicial usando o volume completo. A segunda, composta por duas redes distintas em que cada uma utiliza uma parte da segmentação anterior, realiza uma segmentação mais específica no local. A primeira fase do pipeline da segmentação automática é o pré-processamento das imagens volumétricas, sendo essa composta de três etapas: primeiro é utilizada uma ferramenta para a extração do objeto de interesse (i.e. o encéfalo) (*Brain Extraction Tool* (BET)). Depois é utilizada a normalização da intensidade dos *pixels*. Por fim, é realizada uma correção do sinal de campo de polarização (*Bias-Field Correction* (BFC)). Seguindo o pipeline de segmentação, a segunda etapa é a segmentação das áreas de interesse (i.e. cerebelo, IV ventrículo e tronco cerebelar) pela rede generalista e pelas redes especialistas. Essas redes foram treinadas e validadas utilizando o *5-fold cross validation* dos dados segmentados manualmente. Diferentes arquiteturas foram aplicadas durante essa etapa. Por fim, o último procedimento desse pipeline é a realização de uma fusão entre as duas redes especialistas utilizando um algoritmo de *late fusion*. Para essa tarefa, as imagens por *Magnetic Resonance Imaging* (MRI) escolhidas são de ponderação T2 de crianças entre 0 e 18 anos adquiridas em exames clínicos

realizados com o Hospital das Clínicas da USP. Um total de 32 imagens foram segmentadas manualmente por um grupo de especialistas, nas quais anotações de três regiões diferentes foram feitas na fossa posterior, delimitando assim, as áreas do cerebelo, do IV ventrículo e do tronco cerebelar. Essas segmentações manuais foram utilizadas para treinar e validar as redes neurais generalistas e especialistas. A metodologia proposta alcançou um valor médio de 0,857 no coeficiente Dice durante o teste com apenas 32 imagens volumétricas rotuladas e utilizadas durante o treinamento e validação. Além disso, as distâncias médias entre as superfícies segmentadas, de maneira automática e manual, permaneceram em torno de 1 mm para as três estruturas.

Palavras-chave: Segmentação Semântica, Ressonância Magnética, Redes Especialistas, Fossa Craniana Posterior, Redes Neurais Convolucionais.

Abstract

Maciel, J. L.**Specialist network in automatic segmentation of the posterior cranial fossa in the pediatric population**. 2022. X f. Dissertação (Mestrado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2010.

Differences between adult and child brains cause distinct visual patterns in images acquired using MRI. This is mainly due to the existence of different stages of myelination in the pediatric brain. Thus, although methods exist to support the diagnosis and follow-up of patients through segmentation of the cranial cavities, they are often unable to deal with inter-individual variability. Furthermore, these methods are often not adapted to pathological and pediatric cases. Thus, a possible approach to solve such problems is based on models such as *Fully Convolutional Network* (FCN) and *Convolutional Neural Network* (CNN), which have become viable for volumetric segmentation tasks due to the current high computational power and new methods of training. Thus, this thesis proposes a new method of using these models for automatic segmentation of the pediatric posterior fossa, by supervised training using different architectures. More specifically, the proposed method explores the concepts of generalist network and specialist network, in which the first makes an initial segmentation using the complete volume. The second, made up of two distinct networks that uses a part of the previous segmentation, performs a more specific segmentation on the region. The first phase of the automatic segmentation pipeline is the pre-processing of volumetric images, which consists of three steps: first, a tool is used to extract the object of interest (i.e. the brain) (*Brain Extraction Tool* (BET)). Then the normalization of the intensity of the pixels is applied. Finally, a correction of the polarization field signal (*Bias-Field Correction* (BFC)) is performed. Following the segmentation pipeline, the second step is the segmentation of the areas of interest (i.e. cerebellum, IV ventricle and cerebellar trunk) by the generalist network and by the specialist networks. These models were trained and validated using *5-fold cross validation* of manually segmented data. Different architectures were investigated for this step. Finally, the last procedure of this pipeline is to perform a fusion between the two expert networks using a late fusion algorithm. In order to validate the proposed approach, we have explored a dataset of T2 MRI images of children between 0 and 18 years old acquired in clinical examinations carried out at the Hospital das Clínicas at USP. A total of 32 images were manually segmented by a group of specialists, in which three annotations were made in the posterior fossa, thus delimiting the areas of the cere-

bellum, IV ventricle and cerebellar trunk. These manual segmentations were used to train and validate generalist and specialist neural networks. The proposed methodology reached an average value of 0.857 in the Dice score during the test with only 32 volumetric images labeled and used during training and validation. In addition, the mean distances between the automatically and manually segmented surfaces remained around 1 mm for the three structures.

Keywords: Semantic Segmentation, Magnetic Resonance, Expert Networks, Posterior Cranial Fossa, Convolutional Neural Networks.

Sumário

Lista de Abreviaturas **ix**

Lista de Figuras **xi**

Lista de Tabelas **xiii**

1 Introdução **1**
1.1 Motivação . . . 1
1.2 Objetivos . . . 3
1.3 Contribuições . . . 3
1.4 Organização do trabalho . . . 4

2 Conceitos e Revisão Bibliográfica **7**
2.1 Fossa Craniana Posterior . . . 7
2.2 Ressonância Magnética . . . 8
2.3 Redes Neurais Artificiais . . . 11
2.3.1 Perceptron . . . 13
2.3.2 Multi-Layer Perceptron . . . 16
2.3.3 Redes Neurais Convolucionais . . . 17
2.3.4 Camadas de uma Rede Convolucional . . . 18
2.3.5 Redes generalistas e especialistas . . . 25
2.4 Segmentação de imagens . . . 26
2.4.1 Segmentação Profunda . . . 27
2.4.2 Unet . . . 29
2.4.3 Vnet . . . 29
2.4.4 HighResNet . . . 31
2.5 Aumento de Dados . . . 31
2.5.1 Transformações Afins . . . 32
2.5.2 Transformações Elásticas . . . 32
2.5.3 Transformações em Nível de Pixel . . . 32

3 Método proposto **35**
3.1 Pipeline da Pesquisa 35
3.2 Coleta dos Dados e Segmentação Manual 36
3.3 Método Proposto 37
3.3.1 Rede Generalista e Especialista 38
3.3.2 Treinamento e Avaliação 41
3.4 Implementação 42

4 Resultados e Discussão **43**
4.1 Configuração das redes 43
4.2 Resultados das Redes Generalistas 44
4.3 Resultados das Redes Especialistas e Votação Majoritária 45
4.4 Segmentação Manual 49
4.4.1 Discussão dos Resultados 51

5 Conclusão **53**
5.1 Comentários Finais 53
5.2 Trabalhos Futuros 54

Referências Bibliográficas **55**

Lista de Abreviaturas

NN *Neural Network*

MLP *Multi-Layer Perceptron*

MSE *Mean Squared Error*

MRI *Magnetic Resonance Imaging*

BFC *Bias-Field Correction*

BET *Brain Extraction Tool*

CNN *Convolutional Neural Network*

FCN *Fully Convolutional Network*

ReLU *Rectified Linear Unit*

USP Universidade de São Paulo

IME Instituto de Matemática e Estatística

DSC *Dice similarity Coefficient*

ROI *Region of Interest*

BraTS *Brain Tumor Segmentation*

GPU *Graphics Processing Units*

GAN *Generative Adversarial Network*

HC Hospital das Clínicas

IoU *Intersection Over Union*

Lista de Figuras

1.1 MRIs paciente 0 anos . . . 2
1.2 MRIs paciente 17 anos . . . 3
1.3 Método . . . 4

2.1 Fossas Cranianas . . . 8
2.2 Encéfalo - Visão Lateral . . . 8
2.3 MRI Encéfalo Ponderações T1 e T2 . . . 10
2.4 MRI Encéfalo Diferentes Orientações . . . 10
2.5 Representação de um Neurônio Natural[1]. . . . 11
2.6 Neurônio Artificial Básico . . . 12
2.7 Funções de Ativação . . . 14
2.8 Exemplo Separação Linear *The Perceptron* . . . 15
2.9 MLP . . . 16
2.10 LeNet . . . 18
2.11 Relu . . . 19
2.12 Correlação 2D . . . 20
2.13 Dilatação do *Kernel* . . . 21
2.14 Pooling 2D . . . 21
2.15 AlexNet . . . 22
2.16 VGG-16 . . . 23
2.17 Módulo *Inception* . . . 24
2.18 GoogleLeNet . . . 24
2.19 Bloco residual . . . 25
2.20 Pipeline de segmentação da Med3D . . . 26
2.21 Ressonância Encéfalo Ponderação T2 . . . 27
2.22 CNN x FCN . . . 28
2.23 Arquitetura Encoder-Decoder . . . 29
2.24 Arquitetura 3D Unet . . . 30
2.25 Arquitetura 3D Vnet . . . 30
2.26 Arquitetura HighResNet . . . 31
2.27 Aumento de dados em segmentação de imagens por MRI . . . 32

2.28 Transformação afins em imagens por MRI 33
2.29 Transformação elástica em imagens por MRI 34

3.1 Fluxograma da pesquisa 35
3.2 Histograma das Idades dos Pacientes 36
3.3 Fluxograma Fase I 37
3.4 Pré-processamento pipeline 37
3.5 3D-Slicer 38
3.6 Rede Generalista e Rede Especialista 39

4.1 Resultados Detalhados 45
4.2 Resultados Detalhados 46
4.3 Resultados Detalhados 48
4.4 Resultados Detalhados 49
4.5 Resultados Detalhados 49
4.6 Resultados Detalhados 50
4.7 Resultados Detalhados 50
4.8 Slicer3D 51

5.1 O conjunto de dados foi expandido para 118 imagens anotadas para os segmentos da fossa posterior. A anotação agora inclui rótulos para normais/patologias e para tumores. 54

Lista de Tabelas

4.1 Configuração hiper parâmetros 43
4.2 Pontuação Rede Generalista *Dice similarity Coefficient*(DSC) e *Intersection Over Union* (IoU) 44
4.3 Distâncias μSD e HD95 45
4.4 Pontuação DSC 47
4.5 Distâncias μSD e HD95 47

Capítulo 1

Introdução

1.1 Motivação

O desenvolvimento do encéfalo humano se inicia a partir da terceira semana de gravidez e chega a, aproximadamente, 90% do volume do encéfalo adulto quando a criança completa 6 anos de idade (STILES; JERNIGAN, 2010). Durante esse período, pode-se utilizar algumas das técnicas para a aquisição de imagens do encéfalo como a tomografia computadorizada e a ressonância magnética (*Magnetic Resonance Imaging* (MRI)). Essas técnicas são bastante conhecidas dentro da medicina e normalmente são utilizadas pelos radiologistas quando há suspeita de alguma irregularidade durante o exame clínico. Apesar de existir diferenças entre as duas técnicas, há uma preferência pela MRI devido à não utilização de raios ionizantes e por ser uma técnica não invasiva, não sendo assim um processo inerentemente perigoso para o paciente ou à equipe da ressonância magnética (SPRAWLS, 2000).

A aquisição de imagens utilizando MRI tem por finalidade monitorar o desenvolvimento e a identificação da presença, ou não, de alguma patologia. Assim, é de grande importância a aquisição das imagens do encéfalo pediátrico. Essa aquisição deve acontecer principalmente tratando-se de uma suspeita relacionada a fossa craniana posterior em crianças, pois é onde se encontra a maioria das doenças graves (CATALA, 2015). Dessa forma, a segmentação de imagens tem sido um passo fundamental utilizado para o acompanhamento do desenvolvimento do encéfalo pediátrico (GUI et al., 2012), bem como para o planejamento pré-operatório e avaliação pós-operatório (ZHANG et al., 2007).

Diferentes estágios de mielinização no encéfalo pediátrico resultam em conteúdo distinto de água, lipídios e colesterol na substância branca, produzindo padrões visuais distintos na MRI (GUI et al., 2012). Quando os axônios são mielinizados, o conteúdo de água na substância branca diminui, enquanto o conteúdo de lipídios e colesterol aumenta. Isso explica por que a intensidade é muito diferente da substância branca amielínica para a que é observada em adultos.

A segmentação do encéfalo em crianças é um desafio não apenas devido ao menor tamanho (i.e. que pode resultar em efeito de volume parcial), mas também por causa de

seu contraste particular entre a substância branca e cinzenta (GUI et al., 2012; TADEUSIEWICZ; OGIELA; SZCZEPANIAK, 2009; THOMPSON et al., 2011). Além disso, a falta de mielinização em crianças menores de 2 anos de idade também tem implicações importantes devido ao baixo contraste de imagem entre a substância branca de aparência normal e anormal. As dificuldades relacionadas ao processo de mielinização no encéfalo pediátrico impactam tanto a segmentação auxiliada por computador quanto os métodos manuais, levando a uma discordância potencialmente significativa entre os especialistas (MOREL et al., 2016).

A Figura 1.1 e a Figura 1.2 ilustram os resultados das aquisições de imagens utilizando MRIs de pacientes com idades distintas em duas diferentes orientações. A Figura 1.1 mostra uma imagem de ressonância de um bebê com menos de 1 ano. A Figura 1.2 mostra a imagem de um adolescente de 17 anos. Como pode-se notar, há diferentes padrões visuais entre as duas aquisições em que, por exemplo, o líquido cefalorraquidiano presente no IV-ventrículo na Figura 1.1 apresenta bem menos brilho comparado com esse liquido presente no IV-ventrículo presente na Figura 1.2. Isso se deve aos diferentes estágios de mielinização no encéfalo pediátrico.

(a) *Axial* **(b)** *Sagital*

Figura 1.1: *MRIs do encéfalo pediátrico de um paciente de menos de 1 ano em duas diferentes orientações, na qual (a) está orientado com o plano axial e (b) orientado com o plano sagital. Imagens do conjunto de dados da colaboração com o ICr-HC-USP.*

Essa complexidade representa um desafio significativo para a maioria dos algoritmos de segmentação automática e semiautomática no contexto da segmentação cerebral (DESPOTOVIĆ; GOOSSENS; PHILIPS, 2015). Muitos desses métodos não consideram especificamente encéfalos pediátricos e se baseiam em abordagens de atlas (GOUSIAS et al., 2012; CARDOSO et al., 2013) e *shallow learning* (MOESKOPS et al., 2015; WANG et al., 2015; THOMPSON et al., 2011). No entanto, essas abordagens frequentemente não conseguem lidar adequadamente com a variabilidade interindividual e geralmente não são adaptadas para casos patológicos.

Além disso, não há conjuntos de dados públicos na literatura que abranja todo o espectro da infância (0–18 anos), nem que contenham tarefas relacionadas diretamente à segmentação

(a) *Axial* **(b)** *Sagital*

Figura 1.2: *MRIs do encéfalo pediátrico de um paciente de 17 anos em duas diferentes orientações, na qual (a) está orientado com o plano axial e (b) orientado com o plano sagital. Imagens do conjunto de dados da colaboração com o ICr-HC-USP.*

de estruturas da fossa posterior. Fazendo-se necessário a construção de um conjunto de dados para o projeto no qual este trabalho está inserido.

Portanto, é necessária a utilização de métodos que tenham uma maior capacidade de generalização. Como demonstrado por alguns trabalhos recentes (MILLETARI; NAVAB; AHMADI, 2016; CHEN; MA; ZHENG, 2019; BUI; SHIN; MOON, 2017; LI et al., 2017), as Redes Neurais Convolucionais (*Convolutional Neural Networks Convolutional Neural Network* (CNN)) e as Redes Totalmente Convolucionais (*Fully Convolutional Network* (FCN)), vêm conseguindo resultados expressivos em tarefas de segmentação automática de imagens médicas volumétricas mesmo tendo acesso a poucos dados. Esse é o contexto descrito neste documento.

1.2 Objetivos

O objetivo desta pesquisa é desenvolver um novo método baseado em CNNs e FCNs dentro de um *pipeline* de segmentação automática da fossa posterior do encéfalo pediátrico (Figura 1.3). O novo método utiliza um conceito de rede generalista e redes especialistas, no qual a primeira rede é responsável pela segmentação utilizando o volume completo enquanto as redes especialistas se especializam em partes específicas do volume aproveitando da segmentação anterior para o estabelecimento das regiões de interesse (*Region of Interest* (ROI)). A Figura 1.3 ilustra o pipeline proposto.

1.3 Contribuições

As principais contribuições deste trabalho de mestrado são:

1. Este trabalho contribui para uma elaboração de um pipeline para a segmentação de dados volumétricos baseado nos conceitos de rede generalista e rede especialista. Tendo

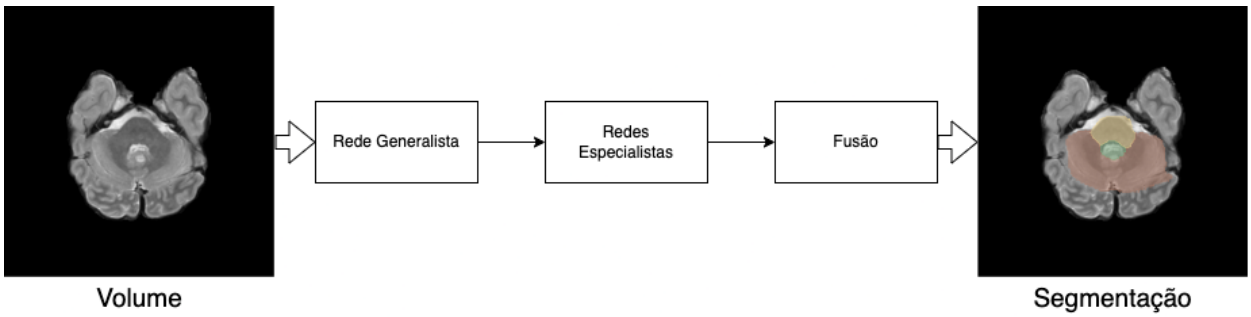


Figura 1.3: *Imagem que ilustra o objetivo desta pesquisa, que é desenvolver um método para realizar a segmentação automática da fossa posterior do encéfalo pediátrico de imagens volumétricas obtidas utilizando uma máquina de ressonância magnética.*

como desafio a segmentação da estrutura supramencionada, foram estudadas diversas formas para a elaboração de um método único que segmentasse encéfalos tanto de crianças quanto de adolescentes entre 0 e 18 anos. Dessa forma, é apresentada neste trabalho uma forma de segmentação que une conceitos distintos de redes dentro de uma mesma pipeline na qual, os resultados devem auxiliar no acompanhamento e possível detecção de anomalias e/ou más-formações nestas áreas.

2. A segunda contribuição é relacionada à formação do conjunto de dados necessários para o desenvolvimento deste trabalho. Em colaboração com o Hospital das Clínicas da USP, foi elaborado um conjunto de dados volumétrico de encéfalos pediátricos (0–18 anos) dos quais 32 volumes foram anotados. Todas as anotações realizadas foram supervisionadas e revisadas por especialistas, prezando dessa forma pela corretude das anotações.

3. Proposição das redes especialistas para segmentação das ROIs.

4. Estudo do desempenho das diferentes arquiteturas de redes de segmentação para segmentação das imagens volumétricas.

5. Uma versão preliminar do método foi publicado em (OLIVEIRA et al., 2021a).

1.4 Organização do trabalho

Este trabalho está estruturado em 5 capítulos, sendo este o Capítulo 1. Neste capítulo, está elaborada uma justificativa para o desenvolvimento desta pesquisa, assim como os seus objetivos. O Capítulo 2, apresenta a base teórica necessária para o entendimento e desenvolvimento deste trabalho. O método desenvolvido é detalhado no Capítulo 3. Também são apresentadas as ferramentas que foram utilizadas, ou seja, a linguagem computacional em conjunto com suas principais bibliotecas utilizadas nesta pesquisa, a arquitetura computacional e os dados utilizados assim como sua aquisição e preparação. São esclarecidas, ainda, as arquiteturas das redes e suas adaptações para este projeto e o método de segmentação

proposto por esta pesquisa. Já no Capítulo 4 são apresentados os experimentos realizados, e os resultados obtidos, em cada uma das arquiteturas utilizando o método proposto. Por fim, no Capítulo 5 é feito discussões sobre os resultados dos experimentos e, finalmente, são expostos os trabalhos futuros.

Capítulo 2

Conceitos e Revisão Bibliográfica

Este capítulo contém seções em que são visitados alguns aspectos teóricos que fundamentam o desenvolvimento desta pesquisa. Ele é dividido em duas partes principais: na primeira parte, estão os aspectos relacionados com a parte teórica da neurociência e radiologia e, na segunda parte, são apresentados os aspectos fundamentais da área de computação.

2.1 Fossa Craniana Posterior

O crânio humano possui uma estrutura chamada de cavidade craniana, responsável por abrigar e proteger o encéfalo. O fundo dessa cavidade é constituída por três fossas (Figura 2.1), sendo que cada uma delas abriga uma parte do encéfalo. A maior dessas três fossas e também a mais anatomicamente complexa é a fossa craniana posterior (A.; A.L., 2015) localizada próximo da parte inferior do crânio. As outras duas fossas são: a fossa craniana anterior, que abriga os lobos frontais do cérebro e o primeiro par de nervos cranianos (DERKOWSKI; KEDZIA; GLONEK, 2003) e a fossa craniana média que abriga na sua porção central a glândula pituitária e nas suas porções laterais os lobos temporais do cérebro (DJALILIAN et al., 2007).

A fossa posterior abriga o tronco cerebelar que é composto pelo mesencéfalo, *pons* (ponte) e o bulbo, fazendo uma conexão entre o telencéfalo com a medula espinhal (SINGH, 2014), mais o cerebelo formado por dois hemisférios (os Hemisférios Cerebelares) e uma parte central chamada de Vérmis Cerebelar. Entre o cerebelo e o tronco cerebelar há uma cavidade preenchida por líquido cefalorraquidiano, o IV-ventrículo. Essa cavidade é uma das quatro cavidades que formam o sistema ventricular, as outras cavidades são: ventrículos laterais, III-ventrículo e o aqueduto cerebral. A Figura 2.2 contém o encéfalo de um adulto saudável, no qual é indicada a localização tanto do IV-ventrículo quanto do cerebelo e do tronco cerebelar.

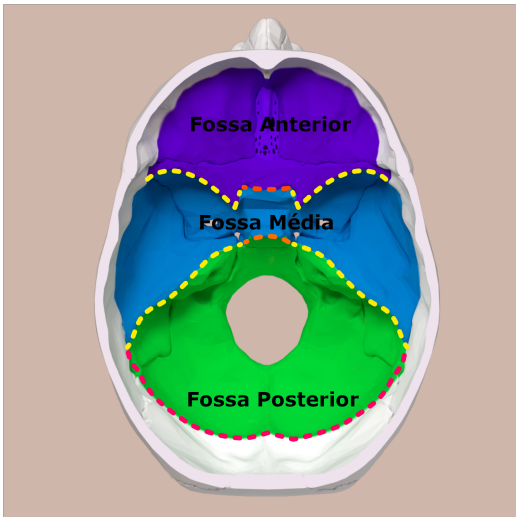


Figura 2.1: *Visão superior das fossas cranianas, cada uma delas abriga uma parte do encéfalo: a fossa craniana anterior, abriga os lobos frontais do cérebro e o primeiro par de nervos cranianos; a fossa craniana média que abriga na sua porção central a glândula pituitária e nas suas porções laterais os lobos temporais do cérebro; a fossa posterior abriga o tronco cerebelar e o o cerebelo. Figura retirada e adaptada de* [1].

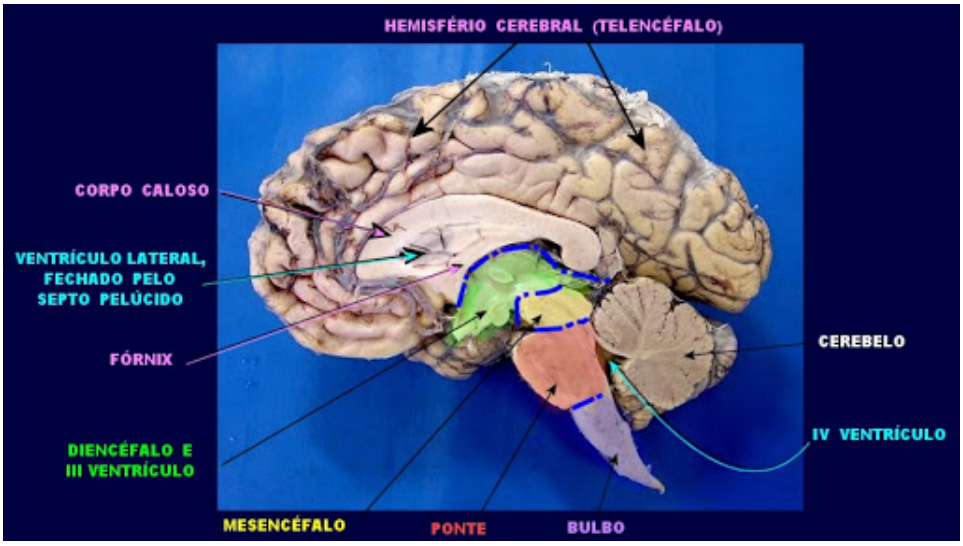


Figura 2.2: *Visão lateral de um encéfalo adulto. Nela é possível notar a cavidade que forma o IV-ventrículo, a localização do cerebelo, a composição e a localização do tronco cerebelar, além de outras partes do encéfalo. Figura retirada de* [2].

2.2 Ressonância Magnética

A ressonância magnética (MRI) é uma técnica de aquisição de imagens que utiliza de campos magnéticos, ondas de rádio e gradientes de campo para formar as imagens. Essa técnica segue, basicamente, três etapas (JÚNIOR; YAMASHITA, 2001), sendo utilizada na radiologia com o objetivo de construir imagens da anatomia e de processos fisiológicos do corpo. As três etapas, na ordem, são:

[1] Adaptado de: <https://commons.wikimedia.org/wiki/File:Cranial_fossae_boundaries.svg>. Acessado em: 30/08/2023.

[2] Retirado de: <http://anatpat.unicamp.br/bineucerebrosagital1.html>. Acessado em: 30/08/2023.

- **Alinhamento:** O alinhamento se refere à propriedade magnética de núcleos de alguns átomos que tendem a se orientar paralelamente a um campo magnético (JÚNIOR; YAMASHITA, 2001). Apesar de outros núcleos possuírem as propriedades necessárias para a utilização em MRIs, o núcleo de hidrogênio (próton) é escolhido, pois é o mais abundante no corpo humano. Existem diferenças significativas na reação pela MRI do hidrogênio no tecido normal e no tecido patológico e, por possuir o maior momento magnético, é o mais sensível ao MRI (MAZZOLA, 2009). Assim, para direcionar esses átomos é necessário um campo magnético. Habitualmente, se usa o valor de 1,5 T[3], aproximadamente (JÚNIOR; YAMASHITA, 2001).

- **Excitação:** A excitação se refere à fase em que o aparelho de MRI emite uma onda eletromagnética para os átomos de hidrogênio, transferindo assim, energia para esses átomos; esse fenômeno é conhecido como ressonância (JÚNIOR; YAMASHITA, 2001). A frequência que deve ser emitida pelo aparelho, chamada de frequência de precessão (MAZZOLA, 2009), é dada pela equação de Larmor (Equação 2.1). Considerando portanto, um campo de 1,5 T e, como a razão giromagnética é de 42,58 MHz/T para o hidrogênio, temos que a frequência de precessão será de 63,87 MHz.

- **Detecção de Radiofrequência:** A detecção de radiofrequência é a etapa de formação da imagem. Após um certo período de tempo da excitação dos átomos, que por sua vez ficaram instáveis dado a energia recebida, eles retornam ao estado habitual e emitem ondas eletromagnéticas na mesma frequência (JÚNIOR; YAMASHITA, 2001). Essa frequência é detectada pelo aparelho e, por meio da transformada de Fourier (MAZZOLA, 2009), determina a intensidade do sinal em um plano bidimensional, que então, é mostrado normalmente como uma imagem em níveis de cinza.

A Equação de Larmor é dada por:

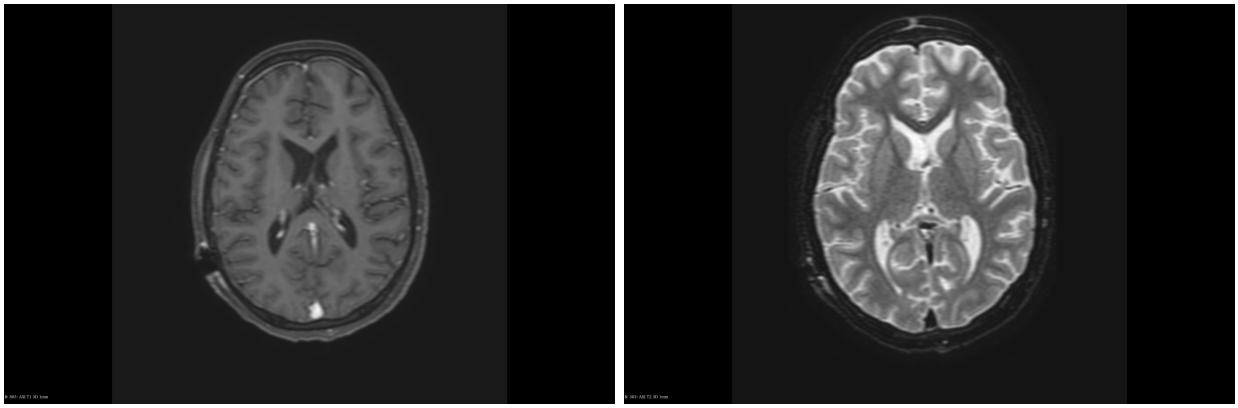
$$\omega = \gamma \mathcal{B}_0, \tag{2.1}$$

em que, w é a frequência precessional, γ é a razão giromagnética e $\mathcal{B}_0$ é a potência do campo magnético.

Variando a sequência de pulsos radio frequência emitidos pelo aparelho (etapa de excitação) e coletados (etapa de detecção de radiofrequência), temos que diferentes tipos de imagens são geradas (MAZZOLA, 2009), cada uma sendo mais sensível a diferentes propriedades do tecido (JÚNIOR; YAMASHITA, 2001). Para dois desses tempos é utilizada a nomenclatura de ponderação em T1 e ponderação em T2. Mazzola (2009) explica de forma mais detalhada os fenômenos físicos e as diferenças de tempo para a formação dessas imagens.

[3]Tesla é uma unidade de medida usada normalmente para descrever campos magnéticos (1T = 10.000 Gauss). E, em MRIs a intensidade do campo magnético interfere na qualidade da imagem gerada, sendo que valores maiores tendem a gerar imagens com uma qualidade maior.

A Figura 2.3 apresenta a aquisição feita utilizando as ponderações T1 e T2. A Figura 2.3a trata da ponderação T1 na qual a substância branca é mais clara que a cinzenta e áreas com alto conteúdo proteico e tecido adiposo em geral tem maior sinal (JÚNIOR; YAMASHITA, 2001). A Figura 2.3b trata da ponderação T2 na qual os líquidos (liquor), desmielinização e áreas de edema no tecido cerebral se mostram mais claros (JÚNIOR; YAMASHITA, 2001).

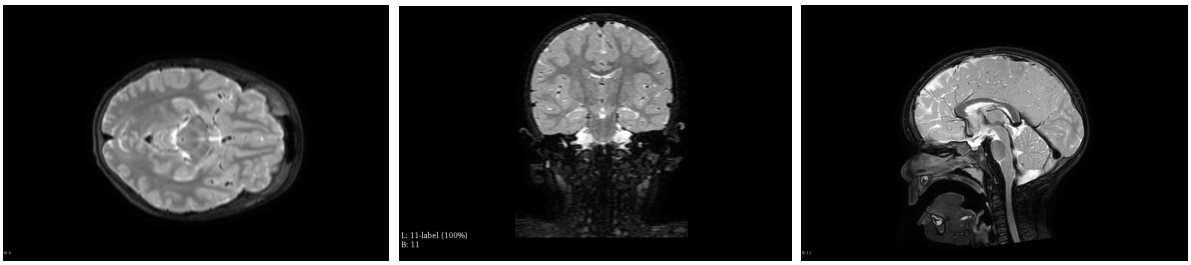


(a) *T1* **(b)** *T2*

Figura 2.3: *Ressonâncias magnéticas do mesmo encéfalo orientada no plano axial, sendo que, (a) foi adquirida usando ponderação T1 e (b) utilizando ponderação T2. Imagens do conjunto de dados da colaboração com o ICr-HC-USP.*

É importante salientar que as imagens da MRI podem ser geradas basicamente em três orientações do plano anatômico:

- **Axial:** O Plano axial, ou transversal, divide o corpo nas porções cranial (superior) e caudal (inferior) (Figura 2.4a).
- **Coronal:** O Plano coronal, ou frontal, divide o corpo nas porções anterior (frente) e posterior (costas) (Figura 2.4b).
- **Sagital:** O Plano sagital divide o corpo nas porções esquerda e direita (Figura 2.4c).



(a) *Axial* **(b)** *Coronal* **(c)** *Sagital*

Figura 2.4: *MRI do encéfalo pediátrico em três diferentes orientações, na qual (a) está orientado com o plano axial, (b) está orientado com o plano coronal e (c) orientado com o plano sagital. Imagens do conjunto de dados da colaboração com o ICr-HC-USP.*

2.3 Redes Neurais Artificiais

As redes neurais artificiais ou simplesmente Redes Neurais (*Neural Network* (NN)) (WANG, 2003) são uma importante área de pesquisa dentro da área de aprendizagem de máquina (GARDNER; DORLING, 1998). Apesar da rede artificial ser apenas um vislumbre em alto nível da rede biológica, essa inspiração é de grande importância para compreensão do seu funcionamento (GUPTA et al., 2013).

No cérebro biológico, o processamento da informação dentro de um neurônio natural, ilustrado pela Figura 2.5, é um estímulo recebido pela sinapse de outros neurônios conectados pelos dendritos. A força dessa conexão determina a importância dessa informação. Esse estímulo é então processado por um corpo celular que integra sinais coletados e gera um sinal de resposta ao longo do axônio ramificado que, por sua vez, distribui a resposta por meio da sinapse que faz contatos com árvores dendríticas de muitos outros neurônios (DONGARE et al., 2012).

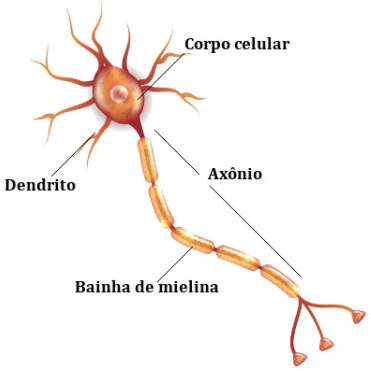


Figura 2.5: *Representação de um Neurônio Natural*[4].

Para modelar o neurônio artificial, fazendo um paralelo entre as redes biológicas, existem três principais conceitos a serem modelados (DONGARE et al., 2012), ilustrados pela Figura 2.6. Primeiro, é modelada a sinapse como pesos para cada valor de entrada. Segundo, é modelado o corpo celular como um somatório entre as conexões ponderadas pelos pesos atribuídos a cada uma delas. Por fim, é modelado o axônio ramificado como uma função de ativação que gera uma resposta que é então distribuída por outras conexões (DONGARE et al., 2012). A realização do processamento da informação de cada um desses neurônios artificiais é descrito pela Equação 2.2:

$$y_k = \varphi(\sum_{j=1}^{m} (w_{kj} x_j) + b_k), \tag{2.2}$$

sendo m o tamanho do sinal de entrada e k o neurônio em questão, em que, $X = [x_1, x_2, x_3, ..., x_m]$ é o vetor de entrada, $W_k = [w_{k1}, w_{k2}, w_{k3}, ..., w_{km}]$ o vetor de pesos, b_k uma variável escalar

[3]Adaptado de: <https://case.edu/med/neurology/NR/MRI%20Basics.htm>. Acessado em: 30/08/2030.
[4]Retirado de: <https://brasilescola.uol.com.br/o-que-e/biologia/o-que-e-neuronio.htm>. Acessado em: 30/08/2023.

(o *bias*) e φ uma função de ativação.

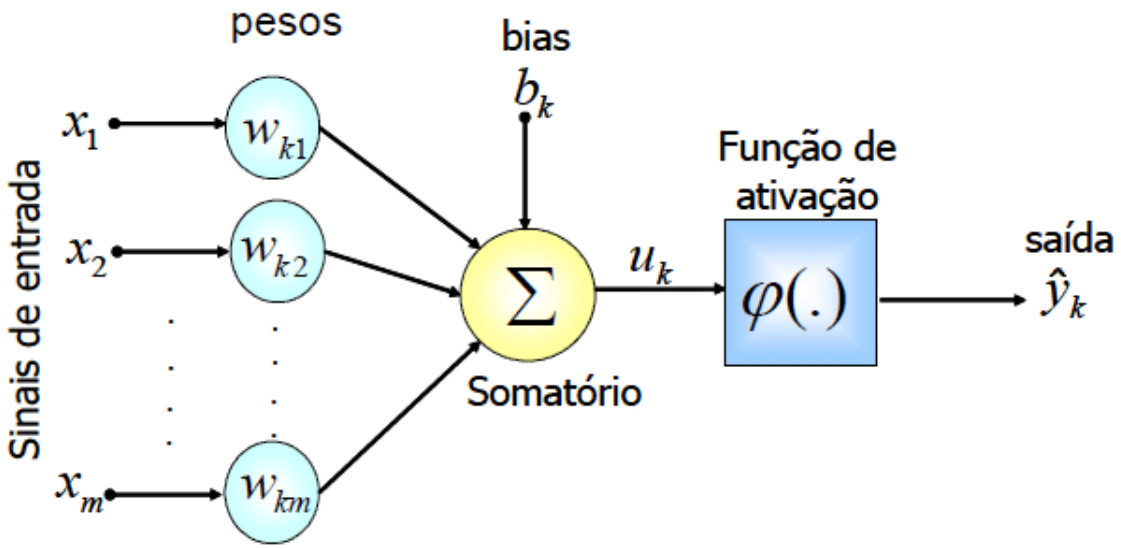


Figura 2.6: *Figura adaptada de (SOARES; SILVA, 2011) e ilustra o funcionamento de um neurônio artificial básico.*

De maneira semelhante ao processo biológico, o aprendizado da NN acontece com o fortalecimento ou enfraquecimento das conexões. Portanto, isso deve ocorrer nos ajustes dos pesos W_k. Esses ajustes, normalmente, utilizam uma função de perda e derivadas parciais para a decida de gradiente.

Uma *função de perda* (AGGARWAL et al., 2018; GOODFELLOW; BENGIO; COURVILLE, 2016), também chamada *função de erro* (WANG, 2003; RASCHKA; MIRJALILI, 2019) e algumas vezes de *função custo*, é uma função cujo o principal objetivo é mensurar a qualidade da predição apresentada pelo algoritmo. Valores menores são preferíveis aos valores maiores. Assim, o aprendizado da rede se torna uma tarefa de otimizar a função de perda (AGGARWAL et al., 2018; GOODFELLOW; BENGIO; COURVILLE, 2016; WANG, 2003; RASCHKA; MIRJALILI, 2019).

Há diversas dessas funções, sendo algumas mais apropriadas para uma determinada tarefa (AGGARWAL et al., 2018). Por exemplo, a função *erro quadrático médio* (*Mean Squared Error* (MSE)), é definida pela Equação 2.3:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \tag{2.3}$$

em que, y_i é o valor esperado e $\hat{y}_i$ o valor predito pela rede. Essa função pode ser utilizada tanto para tarefas de reconstrução quanto para tarefas de regressão, que são treinamentos do tipo não-supervisionado e supervisionado respectivamente.

Como o objetivo é minimizar essa função custo por meio dos ajustes dos pesos da rede, uma maneira é usar o gradiente local para encontrar a direção em que esse custo diminui

[4]Retirado de: <https://brasilescola.uol.com.br/o-que-e/biologia/o-que-e-neuronio.htm>. Acessado em: 12/03/2022

mais rapidamente, técnica conhecida como gradiente descendente (GARDNER; DORLING, 1998; AGGARWAL et al., 2018). No contexto de NN, o algoritmo de *backpropagation* (RUMELHART; HINTON; WILLIAMS, 1985) é o normalmente escolhido para realizar esse procedimento. Utilizando informações do gradiente local, ele atualiza cada um dos pesos de maneira individual e gradativamente a partir do *output* até os neurônios mais internos. Diferentes livros apresentam as derivações completas para diversas funções de perda utilizadas no contexto de NNs (BISHOP et al., 1995; GOODFELLOW; BENGIO; COURVILLE, 2016; AGGARWAL et al., 2018).

Por fim, a função de ativação de uma NN, presente ao final de cada camada, é a transformação não linear após o processamento sobre o sinal de entrada. Sem a função de ativação o sinal de saída seria somente uma transformada linear sobre os dados de entrada, o que limitaria muito a rede (pois o resultado final seria somente transformações lineares sobre o dado de entrada).

2.3.1 Perceptron

A arquitetura mais básica para uma NN, chamada de *perceptron*, inicialmente proposta por Rosenblatt (1958), é uma rede constituída por somente um neurônio (o *perceptron*) e seu funcionamento é dado pela Equação 2.4 (AGGARWAL et al., 2018):

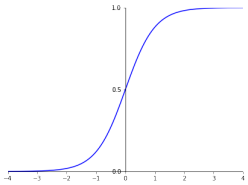
$$\hat{y} = sign(\overline{W}.\overline{X}) = sign(\sum_{j=1}^{d} w_j x_j), \tag{2.4}$$

em que, sendo d o número de *features* (tamanho da entrada), $\overline{X} = [x_1, x_2, x_3, ..., x_d]$ o vetor de entrada (ou vetor de características), $\overline{W} = [w_1, w_2, w_3, ..., w_d]$ o vetor de pesos, $\hat{y}$ o valor predito pela rede (Equação 2.4) e *sign* sua função de ativação dada pela Equação 2.5:

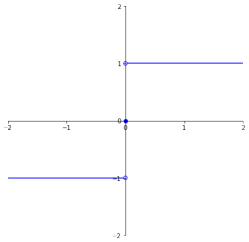
$$sign(x) = \begin{cases} 1 & \text{Se } x > 0 \\ -1 & \text{Se } x < 0 \\ 0 & \text{Se } x = 0 \end{cases}. \tag{2.5}$$

Outras funções de ativação podem ser utilizadas no lugar da *sign* como a *sigmoide* e a *tangente hiperbólica* (WANG, 2003) mudando o escopo dos valores de saída. A Figura 2.7 ilustra cada uma dessas três funções.

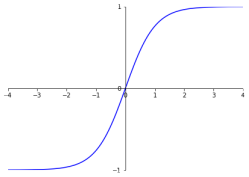
Por ser uma arquitetura desenvolvida em *hardware*, formalmente, Rosenblatt (1958) não propôs uma função de perda, otimizando a rede (reduzindo o erro entre $\hat{y}$ e y) utilizando métodos heurísticos. Contudo, pode-se simular a forma heurística em que era feita a minimização do erro utilizando os mínimos quadrados, de acordo com a Equação 2.6 retirada de



(a) *Função Sigmoide.*



(b) *Função Sign.*



(c) *Função Tangente Hiperbólica.*

Figura 2.7: *Funções de ativação geralmente utilizadas na arquitetura Perceptron.*

(AGGARWAL et al., 2018):

$$Minimize_{\overline{W}} L = \sum_{(\overline{X},y)\in\mathcal{D}} (y - \hat{y})^2 = \sum_{(\overline{X},y)\in\mathcal{D}} (y - sign(\overline{W}.\overline{X}))^2, \tag{2.6}$$

em que $\mathcal{D}$ é composto de pares ordenados $(\overline{X}, y)$ sendo $\overline{X} = [x_1, x_2, x_3, ..., x_d]$ o vetor de características, y o valor observado (valor atribuído como rótulo de $\overline{X}$) e $\hat{y}$ o valor predito pela rede (Equação 2.5).

Definida a função de perda, a forma mais utilizada para sua otimização é descobrir o seu gradiente local e então realizar atualizações graduais em direção ao ótimo (global ou local), o gradiente descendente. Entretanto, como *sign* é uma função não diferenciável, é utilizada a *regra de aprendizado do perceptron* (RASCHKA; MIRJALILI, 2019; AGGARWAL et al., 2018), na qual se usa uma forma suavizada desta função para gerar o gradiente descrito pela Equação 2.7 retirado de (AGGARWAL et al., 2018):

$$\nabla L_{suavizada} = \sum_{(\overline{X},y)\in\mathcal{D}} (y - \hat{y})\overline{X}. \tag{2.7}$$

Assim, a atualização dos pesos $\overline{W}$ em sua forma estocástica (AGGARWAL et al., 2018) é descrita pela Equação 2.8:

$$\overline{W} \leftarrow \overline{W} + \alpha(y - \hat{y})\overline{X}, \tag{2.8}$$

em que, α é um valor escalar que determina a taxa de aprendizagem da rede.

Dessa forma, é possível definir um pseudo código (Algoritmo 1) para realizar o treinamento dessa rede. Observe que esse algoritmo retorna $\overline{W}$. Ele é a *hipótese final* da rede, em que neste caso é uma função linear que tenta aproximar a *função alvo* desconhecida. A Figura 2.8 ilustra o resultado da execução desse algoritmo, em que os pontos são os dados de treinamento, a linha pontilhada é a *função alvo* (Equação 2.9), a linha continua é a *hipótese*

final da rede (após 50 épocas) e a área sombreada é o erro de generalização da rede.

$$y = \begin{cases} 1 & \text{Se } 4x_1 + 3x_2 > 0 \\ -1 & \text{Se } 4x_1 + 3x_2 \leq 0 \end{cases}. \tag{2.9}$$

Algoritmo 1 *Perceptron*

Procedimento PERCEPTRON:$(\mathcal{X} \in \mathcal{R}^{nxd}, \overline{Y} \in \{-1, 1\}^n, \alpha, max\_epochs)$
  $\overline{W} = 0_d$ ▷ vetor de 0's com dimensão d
  para $i \leftarrow 0$ **até** max_epochs **faça**
    para $j \leftarrow 0$ **até** n **faça**
      $\overline{X} = \mathcal{X}_i$ ▷ i-ésima linha de $\mathcal{X}$
      $y = \overline{Y}_i$ ▷ i-ésimo elemento de $\overline{Y}$
      $\hat{y} = sign(\overline{W}.\overline{X})$
      se $y \neq \hat{y}$ **então**
        $\overline{W} \leftarrow \overline{W} + \alpha(y - \hat{y})\overline{X}$ ▷ Atualizar W
      fim se
    fim para
  fim para
  devolve $\overline{W}$
fim Procedimento

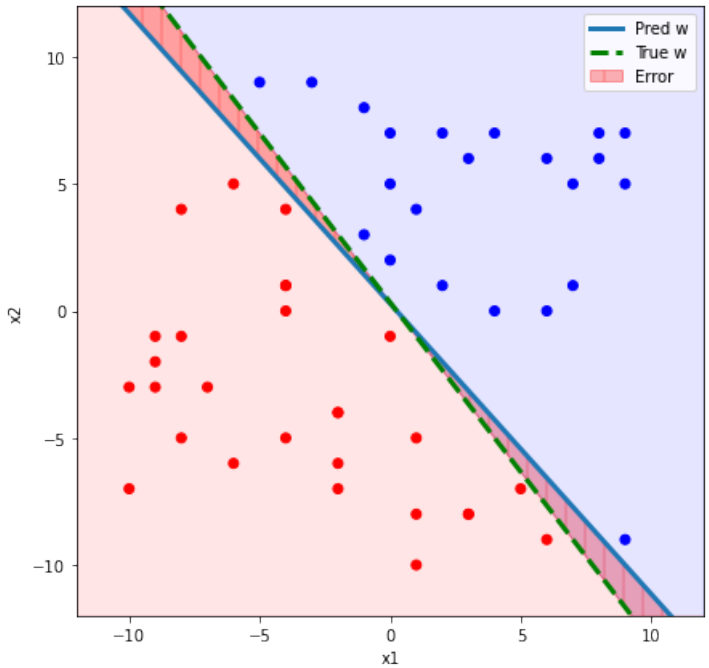


Figura 2.8: *Exemplo de separação linear utilizando The Perceptron após 50 épocas.*

Como pode ser observado, a arquitetura *perceptron* define um hiperplano para a separação dos dados. Não sendo, portanto, adequada para dados não linearmente separáveis.

2.3.2 Multi-Layer Perceptron

Uma *Multi-Layer Perceptron* (MLP) é uma rede baseada em camadas, sendo cada camada totalmente conectada com a camada seguinte. A primeira camada dessa rede é chamada de *input layer*, as camadas intermediárias de *hidden layers* e a camada final de *output layer*. Cada uma dessas camadas é constituída por um conjunto de um ou mais neurônios (GARDNER; DORLING, 1998). A Figura 2.9 ilustra uma MLP com três camadas. Observe que a *input layer* não é contabilizada, pois ela não realiza nenhum cálculo sobre os dados, transferindo-os apenas à camada seguinte (AGGARWAL et al., 2018).

Esses neurônios individualmente funcionam da mesma forma da arquitetura *perceptron*, daí o nome MLP. Contudo, apesar de individualmente funcionarem da mesma forma, essa arquitetura em camadas possui a capacidade de aproximar funções não lineares. Além disso, foi demonstrado por Hornik, Stinchcombe e White (1989) que uma MLP pode aproximar qualquer função suave e mensurável entre os vetores de entrada e saída.

Durante a construção de uma MLP, os neurônios dentro de uma mesma camada, normalmente, possuem a mesma função de ativação (AGGARWAL et al., 2018). Assim, ao invés de serem representados individualmente, os neurônios são representados por um vetor por camada. Por exemplo, o vetor $\overline{h}_1$ na Figura 2.9 é um vetor com dimensão 6 representando a saída dos neurônios da primeira *hidden layer*.

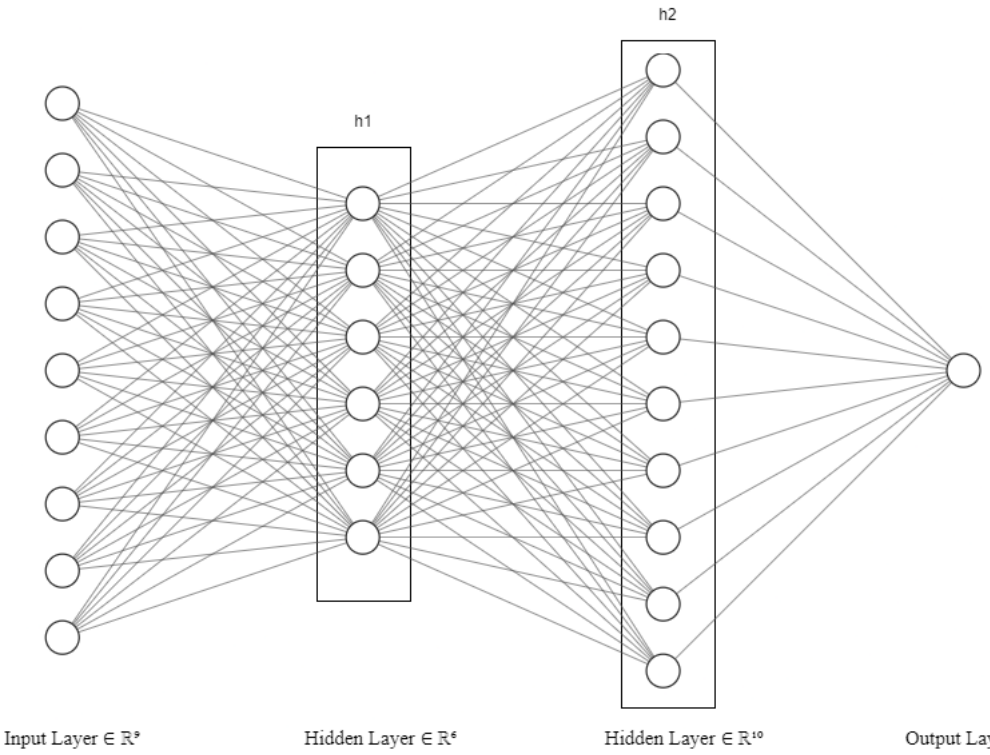


Figura 2.9: *Multi-Layer Perceptron com 3 camadas em que a input layer não é contabilizada por não realizar nenhum cálculo sobre os dados.*

Para realizar uma predição $\overline{o}$ a partir de uma entrada $\overline{x}$, uma rede MLP processa essa entrada em cada uma de suas camadas de forma sequencial. De forma mais precisa, seja k o

número de *hidden layers* de uma determinada rede MLP e $\overline{h}_i$ com $1 \leq i \leq k$ um vetor que contém as saídas da *hidden layer* i. Assim, temos uma matriz W_j como os pesos dessa rede, com $1 \leq j \leq k+1$. A dimensionalidade de W_j é igual a $p_{j-1} \times p_j$ tal que $p_0 = d$, p_i, com $1 \leq i \leq k$, o número de *perceptrons* na *hidden layer* i e p_{k+1} igual a dimensionalidade da *output layer*. Dessa forma, o vetor de entrada $\overline{x}$ é transformado no vetor de saída $\overline{o}$ seguindo de forma recursiva a Equação 2.10[5]:

$$\begin{aligned} &\overline{h}_1 = \varphi_1(W_1^T \overline{x}), \\ &\overline{h}_{p+1} = \varphi_{p+1}(W_{p+1}^T \overline{h}_p) \quad \forall_p \in \{1...k-1\}, \\ &\overline{o} = \varphi_{k+1}(W_{k+1}^T \overline{h}_k), \end{aligned} \tag{2.10}$$

em que, φ_i $1 \leq i \leq k+1$ são as ativações dos neurônios da camada i. Outra forma de escrever essa equação é utilizar mais de um vetor de entrada, *batch* de entrada: seja $\mathcal{X}$ o conjunto de dados de entrada e $X \subseteq \mathcal{X}$ uma matriz de dimensões $n \times d$, sendo n o número de vetores $\overline{x}$, a Equação 2.11 representa, também de maneira recursiva, a equação anterior na sua forma *batch*:

$$\begin{aligned} &H_1 = \varphi_1(X.W_1), \\ &H_{p+1} = \varphi_{p+1}(H_p.W_{p+1}) \quad \forall_p \in \{1...k-1\}, \\ &O = \varphi_{k+1}(H_k.W_{k+1}), \end{aligned} \tag{2.11}$$

em que, H_i tal que $1 \leq i \leq k$ de dimensões $n \times p_i$ e O uma matriz de dimensões $n \times p_{k+1}$.

Assim, para realizar o treinamento dessa arquitetura, é necessário definir como ela atualiza os pesos de cada camada. Como antes citado, normalmente utiliza-se uma técnica de descida de gradiente, em que, a partir do cálculo da *função de perda*, é determinado o gradiente local. Esse gradiente é utilizado para atualizar os pesos da última camada e das camadas anteriores utilizando de derivadas parciais sobre o erro calculado por meio do algoritmo de *backpropagation*.

2.3.3 Redes Neurais Convolucionais

As Redes Neurais Convolucionais (CNN), inicialmente propostas por LeCun et al. (1989), trouxeram uma grande revolução para a área de reconhecimento de padrões na última década (ALBAWI; MOHAMMED; AL-ZAWI, 2017; OLIVEIRA et al., 2021b). Isso se deve, principalmente, à utilização do compartilhamento de parâmetros que reduz drasticamente o número de pesos treináveis, além de a tornar espacialmente invariante (LECUN et al., 1989). Dessa forma, foi possível desenvolver modelos maiores tornando possível a sua utilização em tarefas bastante complexas que não eram possíveis em NN clássicas (ALBAWI; MOHAMMED; AL-ZAWI, 2017). Como exemplo, tarefas de reconhecimento de imagens (FANG et al., 2018) e vídeos (FAN et al., 2016), predição de series temporais financeiras (CHEN et al.,

[5]Equação adaptada do livro (AGGARWAL et al., 2018, pg. 19).

2016), entre outras se tornaram viáveis com a elaborações dessas redes (ALBAWI; MOHAMMED; AL-ZAWI, 2017). Contudo, é importante ressaltar que mesmo com essa redução de pesos treináveis as CNNs são bastante custosas em termos de uso de memória da GPU e deve-se dessa forma ter muito cuidado com a eficiência computacional durante o desenho dessas CNNs.

De maneira semelhante a uma MLP, as CNNs, são organizadas em camadas, sendo que elas possuem uma combinação inicial de camadas de convolução seguidas por uma camada de *pooling* para reduzir a dimensionalidade. Diversas sequências dessas podem ser utilizadas e, ao final da rede, ela é conectada com camadas de neurônios totalmente conectados (BITTEL et al., 2015).

Apesar de a primeira CNN, a LeNet (LECUN et al., 1998) ilustrada pela Figura 2.10, ter sido apresentada no final dos anos 90, as CNNs passaram a ter mais destaque no início da última década quando novas arquiteturas foram apresentadas em conjunto com um melhor poder computacional, um maior conjunto de dados (RONNEBERGER; FISCHER; BROX, 2015) e melhores métodos para o seu treinamento.

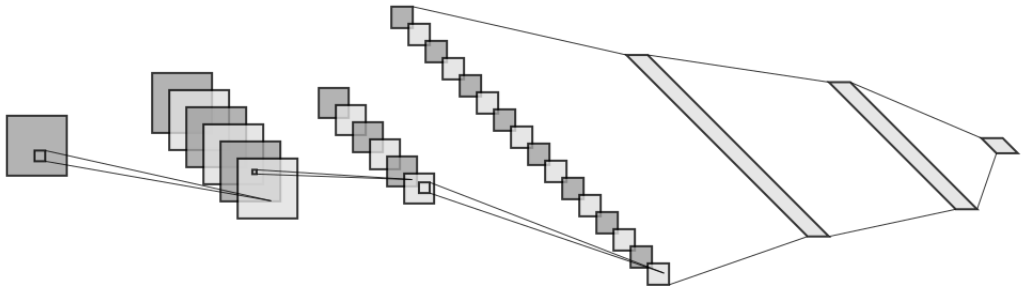


Figura 2.10: *Arquitetura LeNet proposta por LeCun et al. (1998) utilizando de camadas de convolução e avarage pooling seguidas por camadas de neurônios totalmente conectados.*

$$f(x) = max(0, x). \tag{2.12}$$

2.3.4 Camadas de uma Rede Convolucional

Para um melhor entendimento do funcionamento das redes convolucionais, a seguir é feita uma breve explicação sobre a camada de convolução e a camada de *pooling*.

Camada de Convolução

A convolução é uma operação matemática linear (GOODFELLOW; BENGIO; COURVILLE, 2016) em que, nas camadas das redes convolucionais, são realizadas por meio de

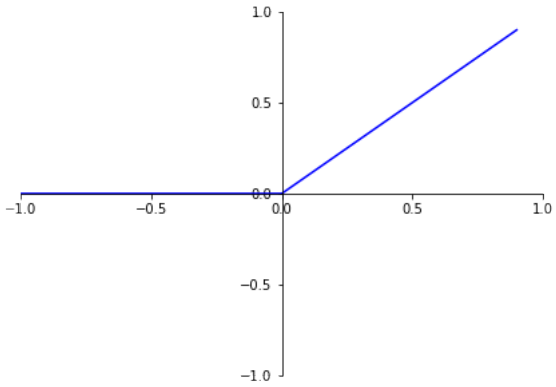


Figura 2.11: *Função de ativação ReLU que é bastante utilizada em arquiteturas de redes convolucionais.*

kernels (ou filtros), sendo cada elemento desse filtro um neurônio (Figura 2.6). Esses filtros deslizam sobre a entrada realizando convoluções que geram uma saída chamada de *feature map*. A equação de convolução sobre uma entrada de duas dimensões e um *kernel* também de duas dimensões é descrita pela Equação 2.13 retirada de (GOODFELLOW; BENGIO; COURVILLE, 2016), que para facilitar foi expressa como um somatório infinito, considerando que todos os pontos, exceto os pontos da entrada e do *kernel*, têm valores nulos:

$$S(i,j) = (I*K)(i,j) = \sum_m \sum_n I(m,n)K(i-m,j-n) = \sum_m \sum_n I(i-m,j-n)K(i,j), \quad (2.13)$$

em que S é a *feature map* resultante, I é *feature map* ou a matriz de entrada, $0 \leq i \leq d_1'$ e $0 \leq i \leq d_2$, no qual, d_1' e d_2' são os tamanhos da *feature map* S obtidos pelas equações:

$$d_1' = d_1 - k + 1 \text{ e } d_2' = d_2 - k + 1,$$

em que d_1 e d_2 são os tamanhos da entrada I. Apesar de ser muito utilizada quando se deseja realizar uma prova, normalmente, as implementações utilizam, apesar de nomeá-las de convolução, uma outra operação matemática chamada de correlação que é bem similar (Equação 2.14), porém, não sendo necessário girar o filtro.

$$S(i,j) = (I \circ K)(i,j) = \sum_m \sum_n I(i+m,j+n)K(i,j). \quad (2.14)$$

A Figura 2.12 ilustra essa operação aplicada a uma matriz I_{7x7} por um filtro K_{3x3}.

É comum uma camada possuir mais que um desses filtros de convolução gerando diversas *feature maps*. Assim, cada filtro da camada seguinte K deve possuir uma terceira dimensão d equivalente ao número de filtros da camada anterior. Isso também ocorre entre a entrada e a segunda camada. Por exemplo, se uma imagem de entrada for da forma RGB, então os filtros da segunda camada deverão ter dimensões $k_1 \times k_2 \times 3$.

Além disso, após a convolução ser realizada, é comum ela ser seguida por uma função de

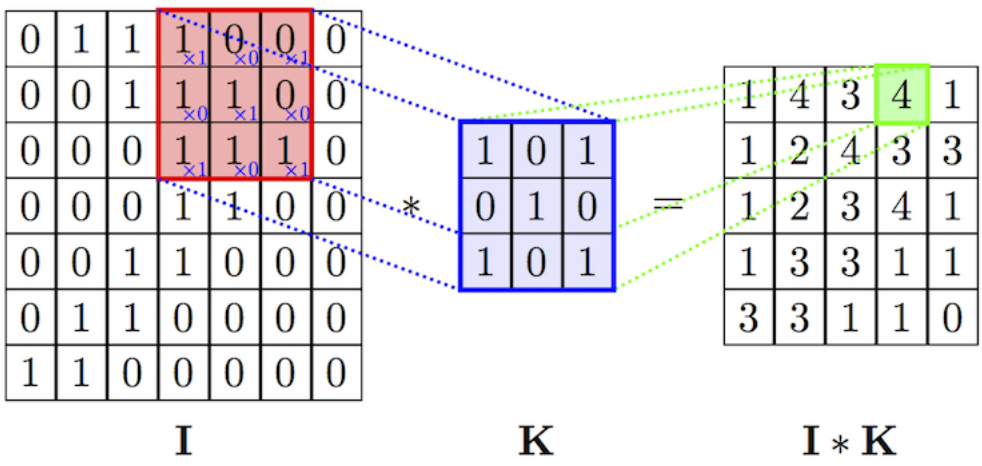


Figura 2.12: *Operação de correlação sobre uma matriz de 2 dimensões retirada de (MOHAMED, 2017). Observe que para este filtro a convolução teria o mesmo resultado.*

ativação introduzindo não linearidade (GOODFELLOW; BENGIO; COURVILLE, 2016). Por fim, pode ser desejável que a matriz resultante depois da operação de convolução tenha o mesmo tamanho da matriz de entrada, ou que o filtro dê um "passo" maior em cada interação. Para atender a essas necessidades é usual utilizar-se das técnicas *padding* e *stride*.

Outra técnica bastante utilizada com o objetivo de cobrir mais informação em cada operação de convolução, com o mesmo custo computacional, é a técnica de dilatação, ou expansão, do *kernel* inserindo buracos entre seus elementos consecutivos. Como pode ser observado pela Figura 2.13, utilizando esse método mais informação é obtida (devido ao maior campo de visão) sem aumentar o número de parâmetros do *kernel*.

Camada de *Pooling*

A camada de *pooling* tem por objetivo diminuir a dimensionalidade da entrada, por meio de uma função de agrupamento que substitui a saída, da camada anterior, em um determinado local, por uma estatística resumida das saídas próximas (GOODFELLOW; BENGIO; COURVILLE, 2016). Ela funciona de forma semelhante à convolução, contudo, não possui neurônios. Um exemplo dessas funções é a Max-Pooling, proposta por Zhou et al. (1988), que escolhe o maior valor dentro de um retângulo que também desliza sobre a entrada. Outras funções de agrupamento populares incluem a média de uma vizinhança retangular, a norma L2 de uma vizinhança retangular ou uma média ponderada com base na distância do pixel central (GOODFELLOW; BENGIO; COURVILLE, 2016). A Figura 2.14 ilustra duas dessas funções com o retângulo de tamanho 2x2 e *stride* de 2.

AlexNet

A arquitetura AlexNet, desenvolvida por Krizhevsky, Sutskever e Hinton (2012) e ilustrada pela Figura 2.15, é constituída por cinco camadas convolucionais e três camadas totalmente conectada, utilizando em suas primeiras camadas convolucionais *kernels* de tamanho

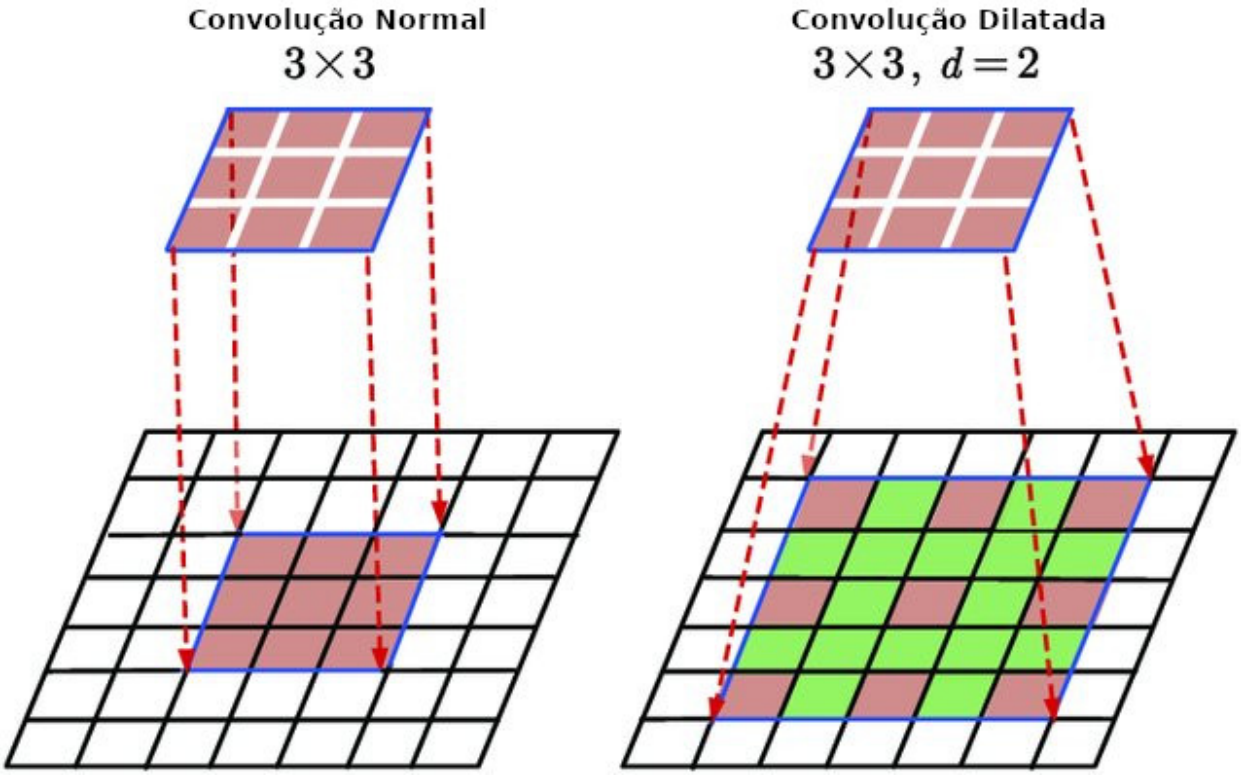


Figura 2.13: *Distinção no campo visual entre uma convolução dilatada e uma convolução normal. Os quadrados vermelhos representam as áreas onde o kernel executa suas operações convolucionais convencionais, enquanto os quadrados verdes representam os espaços vazios introduzidos durante a convolução dilatada. Esses espaços vazios permitem que o kernel abranja uma área mais ampla, permitindo uma percepção mais abrangente das características da entrada.*

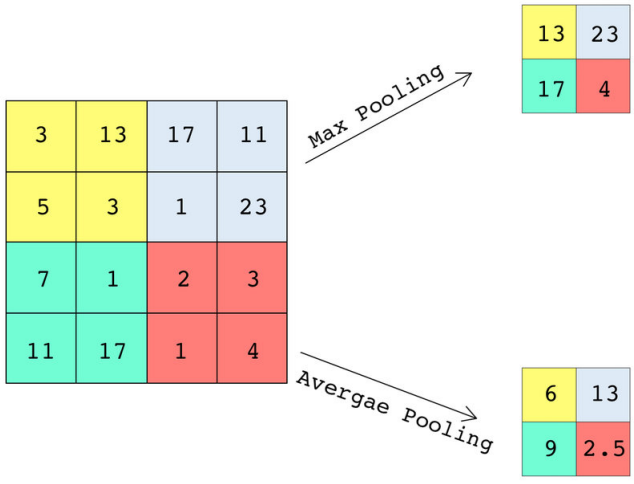


Figura 2.14: *Exemplos de dois pooling amabas com o kernel de tamanho 2x2 e tamanho de passo também igual a 2 (figura retirada de (ALJAAFARI, 2018)).*

maiores do que das camadas mais profundas. Além disso utilizou uma diferente função de ativação, a *Rectified Linear Unit* (ReLU), ilustrada pela Figura 2.11 e descrita pela Equação 2.12, que havia sido demonstrada por Glorot, Bordes e Bengio (2011) ser melhor para o treinamento de redes mais profundas permitindo dessa forma uma melhor escalabilidade do

que as primeiras CNNs (OLIVEIRA et al., 2021b).

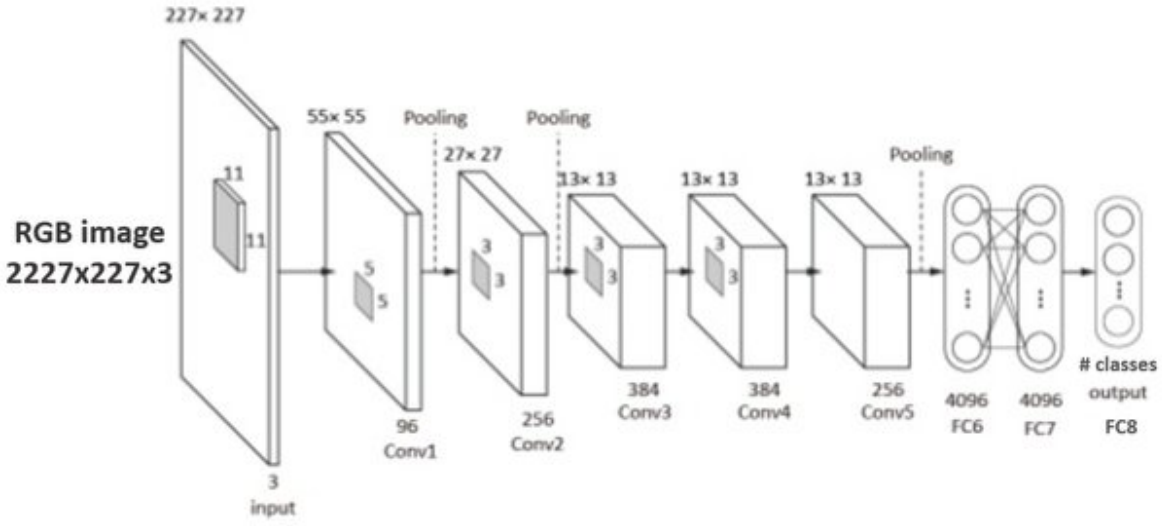


Figura 2.15: *Arquitetura de uma CNN, AlexNet, desenvolvida por Krizhevsky, Sutskever e Hinton (2012). Ela utiliza de kernels de tamanho* 5×5 *em suas camadas iniciais e de tamanho* 3×3 *nas suas camadas mais profundas. Figura retirada de (KHVOSTIKOV et al., 2018).*

VGG

A arquitetura VGG (SIMONYAN; ZISSERMAN, 2014) simplificou as CNNs utilizando *kernels* de tamanho fixo, 3×3, tanto em suas convoluções quanto nas camadas de *pooling*. Apesar do campo receptivo ser mínimo, com *kernels* deste tamanho, a VGG compensou esse efeito aumentando o número de camadas de convolução da rede. Por exemplo, a VGG-16 possui um total de 16 camadas sendo 13 dessas camadas convolucionais e 3 camadas de neurônios totalmente conectados. Essa forma de construir a rede mostrou ser altamente eficiente, pois não utiliza de *kernels* grandes, e alcançou ótimos resultados superando a AlexNet em *benchmarks* conhecidos de classificação e detecção de imagens como o ImageNet[6].

GoogleNet

A arquitetura GoogleNet (SZEGEDY et al., 2015) consiste em uma rede com 22 camadas sendo 9 dessas camadas um novo modulo, o *Inception*, que utiliza diversas convoluções e *pooling*, em paralelo, de *kernels* de tamanhos distintos. Essa arquitetura foi responsável por definir o novo estado da arte para classificação e detecção de imagens no ano de 2014.

O módulo *inception* ilustrado pela Figura 2.17, foi elaborado com o objetivo de resolver o problema da escolha correta do tamanho do *kernel*, pois um *kernel* maior é preferível para informações distribuídas mais globalmente enquanto que um *kernel* menor é preferível para informações distribuídas mais localmente.

[6]Disponível em: <https://image-net.org/>.

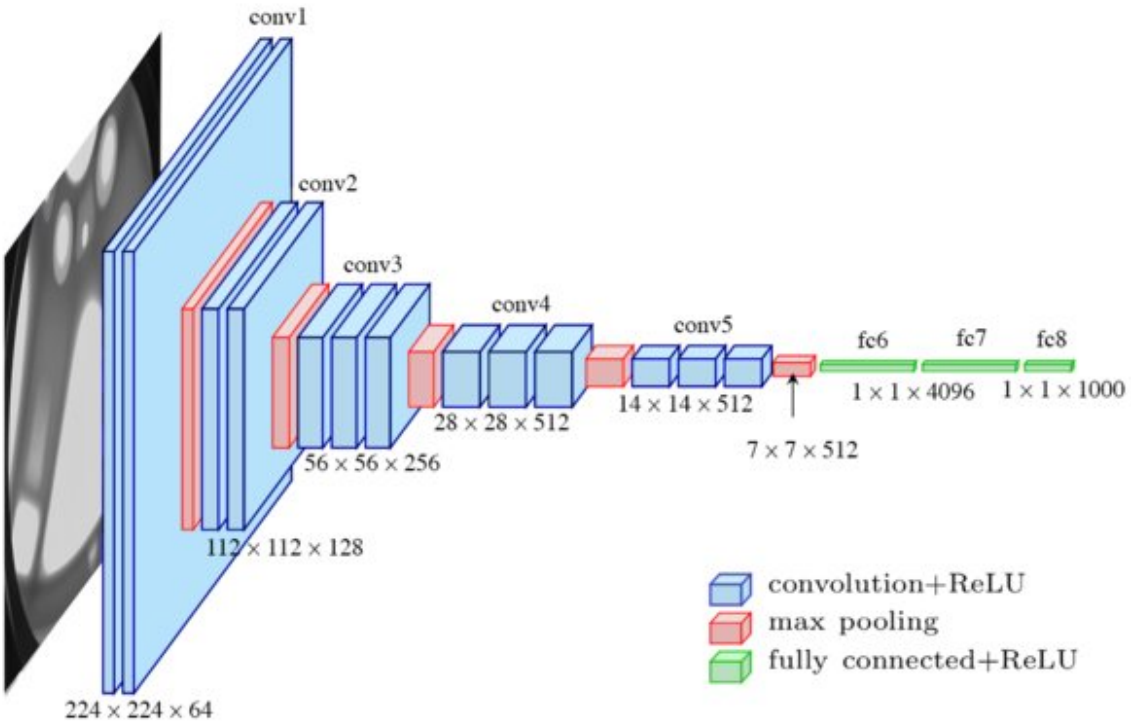


Figura 2.16: *Arquitetura de uma CNN, VGG-16, desenvolvida por Simonyan e Zisserman (2014). Ela utiliza de kernels de tamanho fixo,* 3×3 *em suas 13 camadas convolucionais e utiliza 3 camadas de neurônios totalmente conectados. Figura retirada de (FERGUSON et al., 2017).*

Contudo, como qualquer arquitetura muito profunda, ela está sujeita ao *vanishing gradient problem*[8]. Na tentativa de solucionar este problema, os autores introduziram dois classificadores auxiliares que essencialmente representam uma aplicação do *softmax* após dois desses módulos *inception*. Dessa forma, eles puderam computar duas funções de perda auxiliares em que esses resultados seriam somados à função de perda final da rede de maneira ponderada[9]. A arquitetura apresentada pelos autores é ilustrada pela Figura 2.18.

ResNet

Desde a arquitetura AlexNet as CNNs estão ficando cada vez mais profundas (i.e. de 5 camadas com a AlexNet à um total de 22 camadas com a GoogleLeNet). Contudo, quanto mais as redes ficavam profundas mais era difícil o seu treinamento devido principalmente ao *vanishing gradient problem*. Como foi mostrado, Szegedy et al. (2015) para lidar com esse problema adicionou funções de perda auxiliares no meio da rede, e como esses pesquisadores diversos outros tentaram lidar com esse problema de formas distintas. Entretanto, nenhuma

[9]Figura retirada de: <https://sites.google.com/site/aidysft/objectdetection/recent-list-items>. Acessado em 20 de abril de 2022.

[8]Como o gradiente é retropropagado para as camadas anteriores, a multiplicação repetida pode tornar o gradiente infinitamente pequeno.

[9]No *paper* os autores utilizaram um peso de 0,3 para cada uma das funções de perda auxiliares.

[10]Figura retirada de: <https://sites.google.com/site/aidysft/objectdetection/recent-list-items>. Acessado em 20 de abril de 2022.

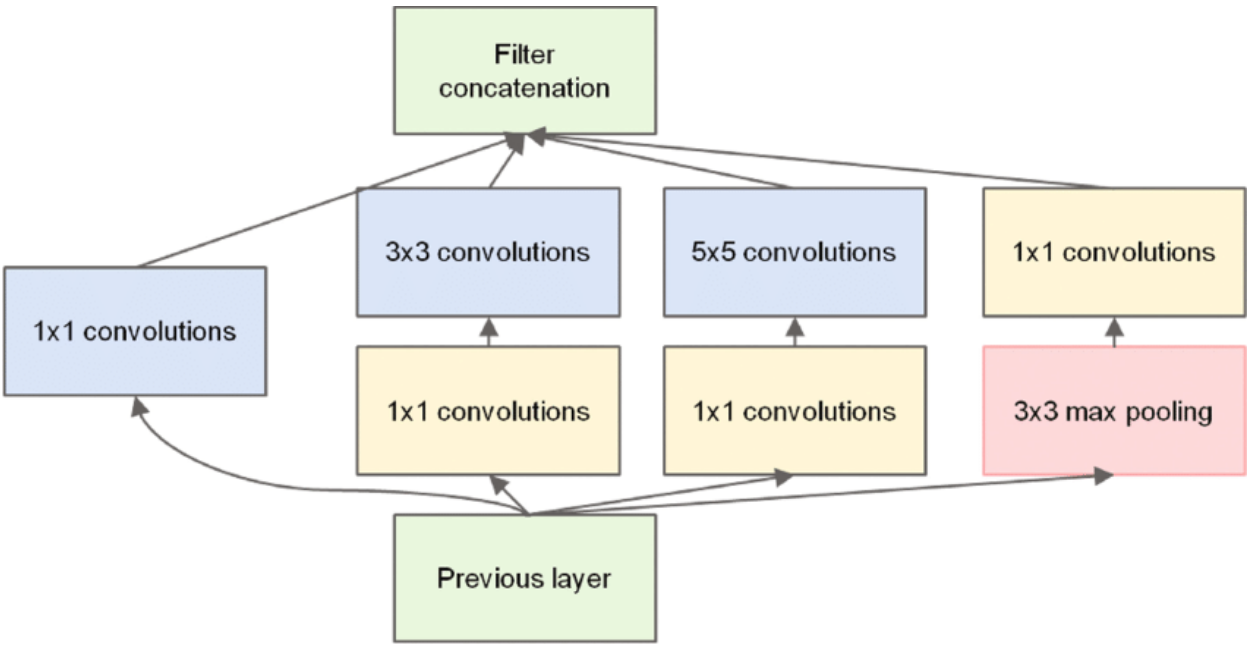


Figura 2.17: *Módulo Inception, que utiliza diversas convoluções e pooling, em paralelo, de kernels de tamanhos distintos com o objetivo de resolver o problema da escolha correta do tamanho do kernel durante a construção de uma CNN*[7].

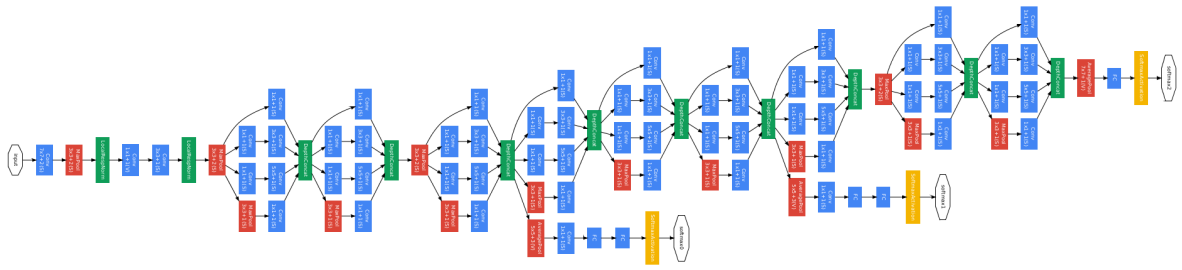


Figura 2.18: *Arquitetura de uma CNN, GoogleNet ou GoogleLeNet, vencedora do desafio ImageNet Large Scale Visual Recognition Challenge (ILSVRC). Arquitetura de 22 camadas que utiliza em sua construção módulos inception e de funções de perda auxiliares*[10].

dessas soluções apresentadas pareceu ser definitiva.

Dessa forma, a ideia principal da ResNet (HE et al., 2016), que tem como principal objetivo lidar com esse problema, é a utilização dos *identity shortcut connection*[11], ilustrado pela Figura 2.19, no qual é realizado uma operação entre um *feature map* de uma camada mais rasa com o *feature map* de uma camada mais profunda.

Na verdade, ResNet não foi a primeira a utilizar de conexões de atalho; sendo a Highway Network (SRIVASTAVA; GREFF; SCHMIDHUBER, 2015) a primeira a introduzir essa conexão, que por sua vez a utiliza com um parâmetro que controla o fluxo de dados, essa conexão é chamada de *gated shortcut connections*. Portanto, a rede ResNet pode ser pensado como um caso especial da Highway Network.

[11] *Conexão de atalho de identidade* em tradução direta.

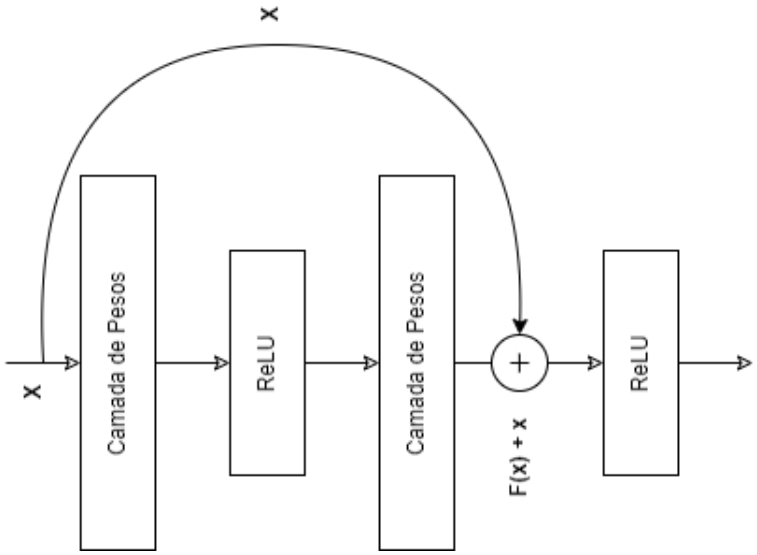


Figura 2.19: *Bloco residual no qual é realizado uma operação entre um feature map de uma camada mais rasa com um feature map de uma camada mais profunda.*

2.3.5 Redes generalistas e especialistas

Um ponto central para esta dissertação se refere aos conceitos de redes generalista e especialista. Embora o levantamento bibliográfico não tenha identificado exatamente essa abordagem, alguns trabalhos apresentam esquemas com alguma analogia. Por exemplo, o trabalho Chen, Ma e Zheng (2019) utiliza um *pipeline* de segmentação bastante semelhante à apresentada por este trabalho. Nesse trabalho, os autores separam a tarefa de segmentação em duas partes: na primeira parte é realizada uma segmentação mais grosseira da estrutura definindo uma área de corte, enquanto a segunda parte utiliza essa região recortada para o refinamento da segmentação reaproveitando o *backbone* da rede inicial. A Figura 2.20 ilustra a arquitetura dessa *pipeline*.

A estrutura Med3D e a arquitetura proposta neste trabalho compartilham uma ideia geral, porém apresentam três distinções fundamentais. Em primeiro lugar, a arquitetura Med3D confia amplamente na aprendizagem de transferência de tarefas correlatas, ao passo que a metodologia proposta não enxerga benefícios no pré-treinamento com ressonâncias magnéticas de adultos do conjunto de dados BraTS2018 (MENZE et al., 2014). Isso pode ser atribuído às consideráveis discrepâncias nos padrões visuais devido aos diferentes estágios de desenvolvimento da mielinização.

Em segundo lugar, ao contrário da Med3D, o método empregado neste trabalho não executa uma adaptação da rede genérica para tarefas especializadas (i.e. não aproveita dos pesos da rede genérica nas redes especialistas). Embora essa abordagem tenha a vantagem de utilizar todo o espectro de dados de ressonância magnética, incluindo aqueles fora das áreas de interesse dos especialistas, ela também obriga a rede a aprender informações em várias escalas, o que pode potencialmente afetar seu desempenho.

Por último, a terceira diferença é que em vez de reconfigurar a arquitetura da rede, a

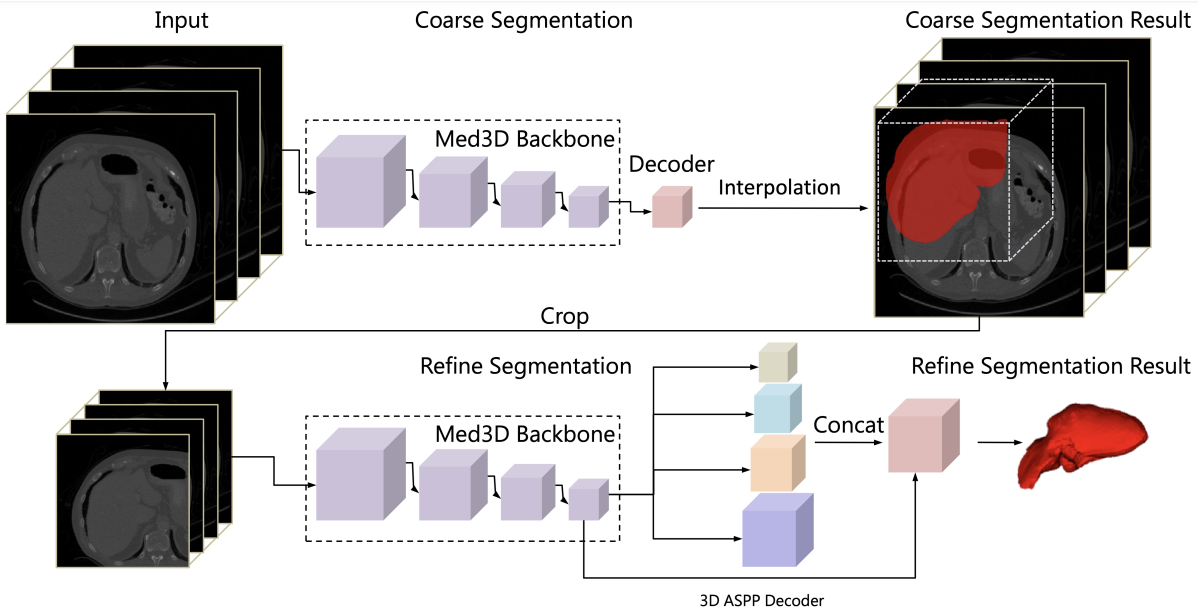


Figura 2.20: *Pipeline de segmentação da Med3D. Esse pipeline é dividido em duas etapas na qual a primeira etapa busca encontrar uma área de corte em volta da estrutura de interesse, por meio de uma segmentação da estrutura, enquanto a segunda etapa realiza o refinamento da primeira segmentação. Figura retirada de (CHEN; MA; ZHENG, 2019).*

metodologia deste trabalho tira proveito das vantagens da combinação de resultados por meio de uma fusão tardia entre três arquiteturas distintas, amplamente reconhecidas. Detalhes completos sobre essa abordagem serão apresentados no Capítulo 3.

Ao realçar esses pontos de divergência, é evidente que embora haja semelhanças entre a Med3D e a arquitetura proposta, há nuances cruciais que delineiam suas abordagens individuais, demonstrando as contribuições únicas do presente trabalho..

2.4 Segmentação de imagens

Em visão computacional, a segmentação de imagens é uma tarefa na qual busca formar grupos de *pixels* (SZELISKI, 2010) que pertencem a uma determinada classe, tipicamente utilizada para delimitar bordas e objetos em uma imagem (TAN, 2016). Pode-se formular o problema da seguinte forma (GARCIA-GARCIA et al., 2017): atribuir um rótulo do espaço de rótulos $\mathcal{S} = \{s_1, s_2, ..., s_k\}$ para cada um dos elementos de um conjunto de variáveis aleatórias $\mathcal{X} = \{x_1, x_2, ..., x_n\}$. O espaço $\mathcal{S}$ é o número de classes possíveis, normalmente adicionando um para incluir a classe *fundo* (*background*), $\mathcal{X}$ é uma imagem 2D (3D para imagens volumétricas), em que, os *pixels* (ou *voxels* para as imagens 3D) são representados pela variável x sendo n o número de *pixels* da imagem.

Esse problema é estudado há muito tempo. Dessa forma, existem diversas formulações que apresentam diferentes restrições à tarefa, como a apresentada pelo Zucker (1976). Assim, diferentes algoritmos já foram propostos, desde os mais clássicos, como o *WaterShed* (SERRA, 1982) e Crescimento de Regiões (ZUCKER, 1976), até os métodos no atual *es-*

tado da arte em segmentação de imagens naturais (LONG; SHELHAMER; DARRELL, 2015; RONNEBERGER; FISCHER; BROX, 2015; MILLETARI; NAVAB; AHMADI, 2016; WANG et al., 2018; LI et al., 2017) baseados principalmente em FCN[12].

A Figura 2.21 ilustra um exemplo de segmentação realizada sobre uma determinada fatia (*slice*) de uma ressonância do encéfalo. Nessa figura foram utilizadas quatro classes o IV-ventrículo, cerebelo, tronco cerebelar e o fundo (*background*).

Figura 2.21: *Exemplo de segmentação de um determinado slice da ressonância do encéfalo em ponderação T2. São ilustradas quatro classes o IV-ventrículo (segmentado com a cor verde), cerebelo (segmentado com a cor marrom), tronco cerebelar (segmentado com a cor amarela) e o fundo (background). Imagens do conjunto de dados da colaboração com o ICr-HC-USP.*

2.4.1 Segmentação Profunda

As CNNs alcançaram resultados expressivos em diversas tarefas relacionadas à imagem, como classificação (KRIZHEVSKY; SUTSKEVER; HINTON, 2012), detecção de objetos, predição de pontos chaves etc. Assim, rapidamente essas redes foram adaptadas para lidar com anotações mais densas (LONG; SHELHAMER; DARRELL, 2015), como é o caso de segmentação de imagens. As que obtiveram melhores resultados utilizaram de um treinamento baseados em *patch* rotulando cada pixel com a classe de seu objeto ou região envolvente. Contudo, foi mostrado por Long, Shelhamer e Darrell (2015) que utilizar arquiteturas totalmente convolucionais, as FCNs, era mais eficiente, e também mais acurado, do que um treinamento baseado em *patches*.

Além disso, qualquer CNN pode ser convertida em uma FCN (LONG; SHELHAMER; DARRELL, 2015), sendo necessário somente a substituição das camadas densas presentes no final da rede por uma interpolação bilinear seguida por camadas de convolução (OLIVEIRA

[12] As FCNs são redes convolucionais que não possuem camadas de neurônios totalmente conectados, assim, possuem somente camadas de *pooling* e camadas de convolução (LONG; SHELHAMER; DARRELL, 2015).

et al., 2021b). Dessa forma, a reinterpretação das CNNs como FCNs, permite a utilização de *transferência de aprendizagem* entre uma representação e outra realizando apenas um ajuste fino de suas representações aprendidas (LONG; SHELHAMER; DARRELL, 2015). Isso propicia o treinamento da rede em grandes *datasets* de classificação bem como utilizando *datasets* menores de segmentação semântica. Essa reinterpretação é ilustrada pela Figura 2.22. Atualmente, essa técnica de *transferência de aprendizagem* vem sendo também complementada por um Treinamento Auto-Supervisionado (Self-Supervised Learning) (**??**). De forma bem resumida, é uma forma de treinamento na qual a rede aprende padrões utilizando dados não rotulados e, após isso, aproveitando dos pesos que foram treinados dessa forma, aprende a rotular os dados em um treinamento supervisionado, essa etapa conhecida como *fine-tunning*.

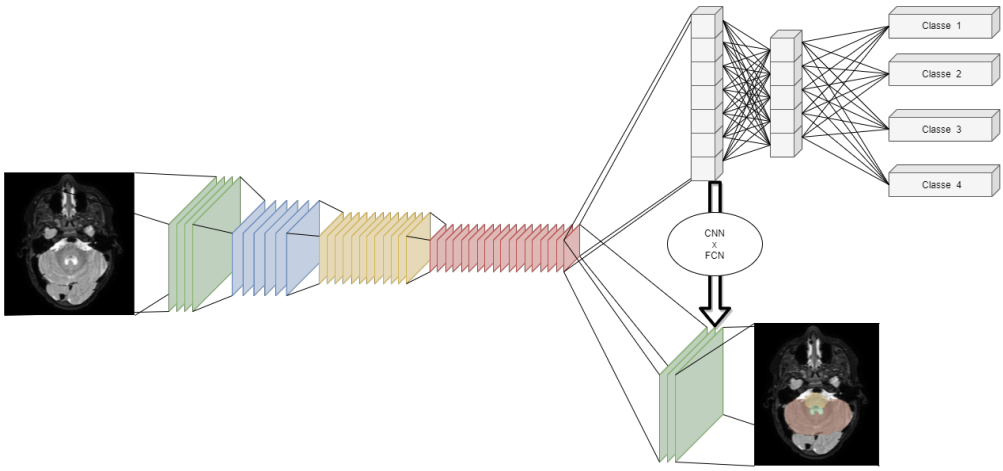


Figura 2.22: *Exemplo da reinterpretação de uma CNN para uma FCN substituindo as camadas totalmente conectadas por uma camada de upsampling seguida por camadas de convolução.*

Mais recentemente, em tarefas de segmentação semântica, são amplamente utilizadas arquiteturas do tipo Encoders-Decoders (YE; SUNG, 2019) que são arquiteturas simétricas geralmente mais complexas compostas por três partes: *Encoder*, *Bottleneck* e *Decoder* ilustrados pela Figura 2.23.

Arquiteturas Encoders-Decoders não são restritas às CNNs, podendo ser empregadas em diferentes NNs como por exemplo as MLPs. Nas arquiteturas convolucionais, pode-se utilizar convoluções transpostas (ZEILER et al., 2010) que, apesar de serem bem similares às convoluções, não diminuem (ou mantêm) a resolução espacial da entrada, realizando um *upsampling* espacial aprendível, permitindo assim arquiteturas simétricas (OLIVEIRA, 2020). Outra forma, é realizar o *upsampling* por meio de algoritmos de interpolação bilinear, que não são aprendíveis, seguidos por uma convolução *regular* (BADRINARAYANAN; KENDALL; CIPOLLA, 2017).

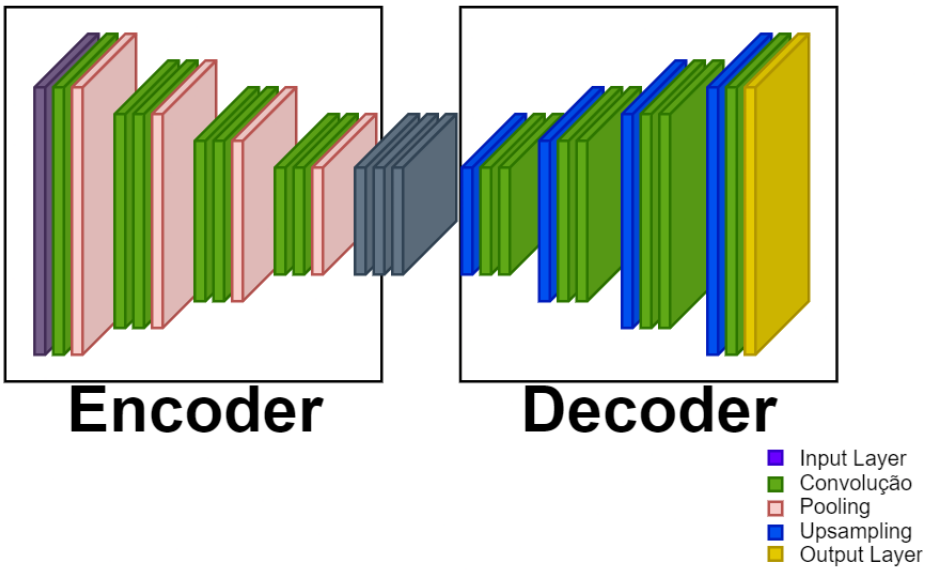


Figura 2.23: *Arquitetura simétrica do tipo Encoder-Decoder.*

2.4.2 Unet

A Unet é uma rede do tipo *encoder-decoder* proposta por Ronneberger, Fischer e Brox (2015) que faz uma boa utilização do aumento de dados, para usar as amostras anotadas disponíveis mais eficientemente. A sua principal contribuição reside na parte de *decoder*, que utiliza uma grande quantidade de canais de *features* permitindo, assim, propagar informações de contexto para camadas de resolução mais alta. Além disso, ela utiliza grande variedade de conexões de salto por meio de concatenação que liga camadas simétricas rasas e profundas para produzir segmentações de alta resolução.

A Unet inicialmente proposta foi desenvolvida para imagens bidimensionais. Dessa forma, nesta pesquisa foi utilizada uma reinterpretação dessa arquitetura, ilustrada pela Figura 2.24, para trabalhar com imagens volumétricas. Essa reinterpretação é atualmente utilizada em inúmeros trabalhos ligados a imagens médicas (QAMAR et al., 2020).

2.4.3 Vnet

A rede Vnet foi desenvolvida por Milletari, Navab e Ahmadi (2016) especificamente para tarefas de segmentação de imagens médicas volumétricas. Sua arquitetura é do tipo *encoder-decoder* e sua maior contribuição foi a utilização de uma nova função de perda, a Dice-*loss*, que considera tanto informações de perda locais quanto globais, o que é crítico para uma alta precisão.

A Equação 2.15 descreve a função de perda Dice-*loss*:

$$\mathcal{L}_{DSC} = \frac{2 * \sum_i^N p_i g_i}{\sum_i^N p_i + \sum_i^N g_i}, \tag{2.15}$$

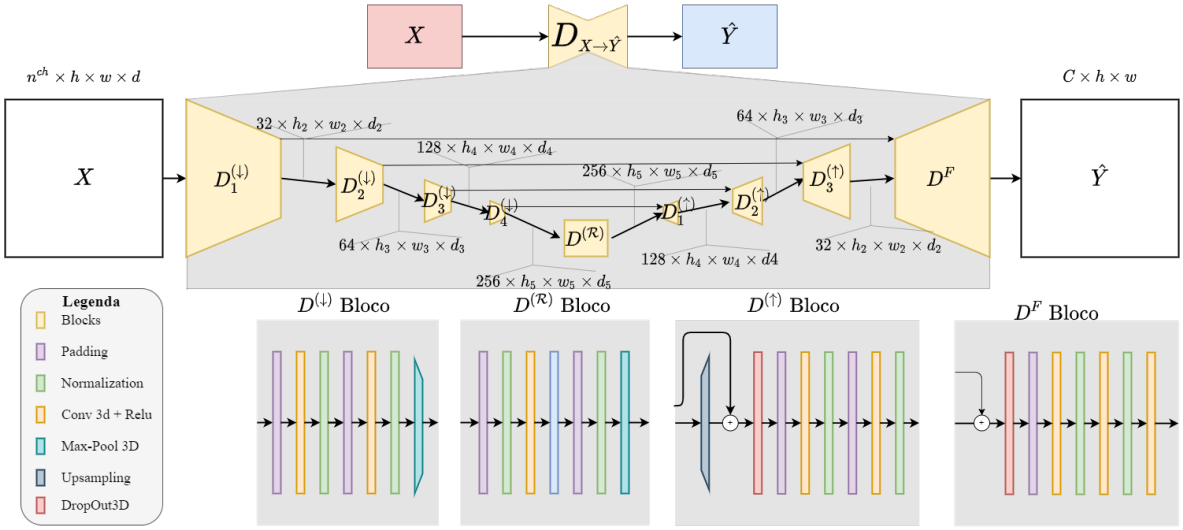


Figura 2.24: *Arquitetura Unet para imagens volumétricas.*

em que, as somas correm sobre os N *voxels*, do volume da predição de segmentação binária $p_i \in P$ e do volume do *ground truth* da segmentação $g_i \in G$.

Em sua arquitetura, a Vnet, de modo semelhante à Unet, utiliza grande variedade de conexões de salto para ligar camadas simétricas rasas e profundas, para produzir segmentações de alta resolução. Contudo, diferentemente da Unet que utiliza concatenação, a Vnet utiliza a soma nessas conexões de salto, reduzindo dessa forma o número de parâmetros treináveis (veja Seção 2.3.4). Além disso, ela utiliza convoluções e convoluções transpostas 2×2 com stride 2 para realizarem um *downsampling* (redução de dimensionalidade) e um upsampling (aumento de dimensionalidade) respectivamente.

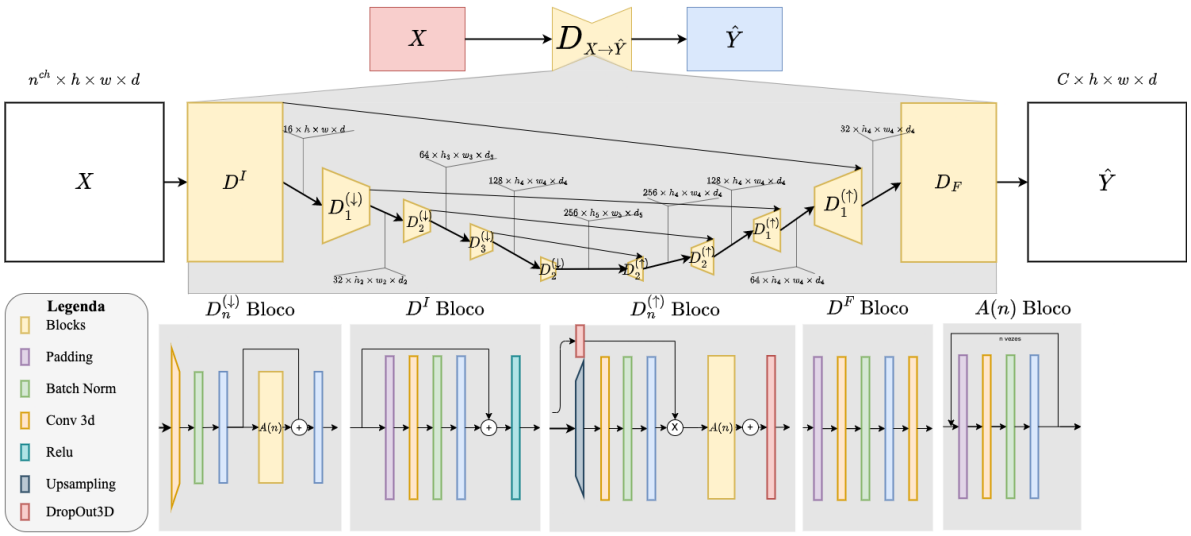


Figura 2.25: *Arquitetura Vnet para imagens volumétricas.*

2.4.4 HighResNet

A HighResNet (LI et al., 2017) foi desenvolvida para a segmentação de imagens volumétricas e apresentou um bom desempenho na segmentação de estruturas neuroanatômicas a partir de imagens de MRI do cérebro. Ela explora os campos receptivos variados de convoluções dilatadas juntamente com conexões residuais para produzir uma arquitetura altamente versátil e mais compacta, sem perder desempenho quando comparada com outras redes na literatura.

Ao utilizar convoluções dilatadas, a arquitetura se beneficia em calcular as *features* da imagem com uma alta resolução espacial e o tamanho do campo receptivo pode ser aumentado arbitrariamente (LI et al., 2017). Se somando a isso, as conexões residuais mostraram tornar a propagação da informação mais suave e melhorar a velocidade do treinamento (HE et al., 2016). A arquitetura HighResNet utilizada nesta pesquisa é ilustrada pela Figura 2.26.

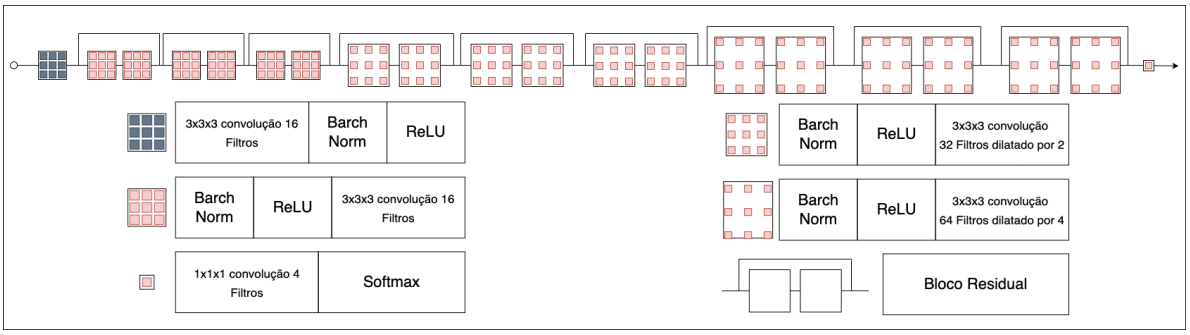


Figura 2.26: *Arquitetura HighResNet para imagens volumétricas.*

2.5 Aumento de Dados

O aumento de dados é uma técnica de crescimento sintético do tamanho dos conjuntos de treinamento, teste e validação. Sendo largamente utilizada durante o treinamento de NNs em geral, ela ajuda a evitar o *overfitting* e melhora a generalização dos modelos. Para tarefas de segmentação de imagens por MRI, a forma de utilização da técnica de aumento de dados pode ser divididas em duas categorias principais(NALEPA; MARCINKIEWICZ; KAWULOK, 2019): 1) transformações na imagem original, como transformações afins, transformações elásticas e transformações em nível de pixel; e 2) geração artificial de dados. Essa categorização é ilustrada pela Figura 2.27.

As transformações na imagem original realizam alterações utilizando as imagens já presentes no conjunto de dados destinados ao treinamento do modelo. Enquanto a geração artificial de dados, busca gerar novos dados utilizando principalmente de Redes Adversárias Generativas (*Generative Adversarial Networks* (GANs)). Porém, na área médica, a qualidade e a precisão dos dados são de extrema importância, já que os resultados podem afetar diretamente as decisões clínicas. GANs podem produzir artefatos ou informações imprecisas

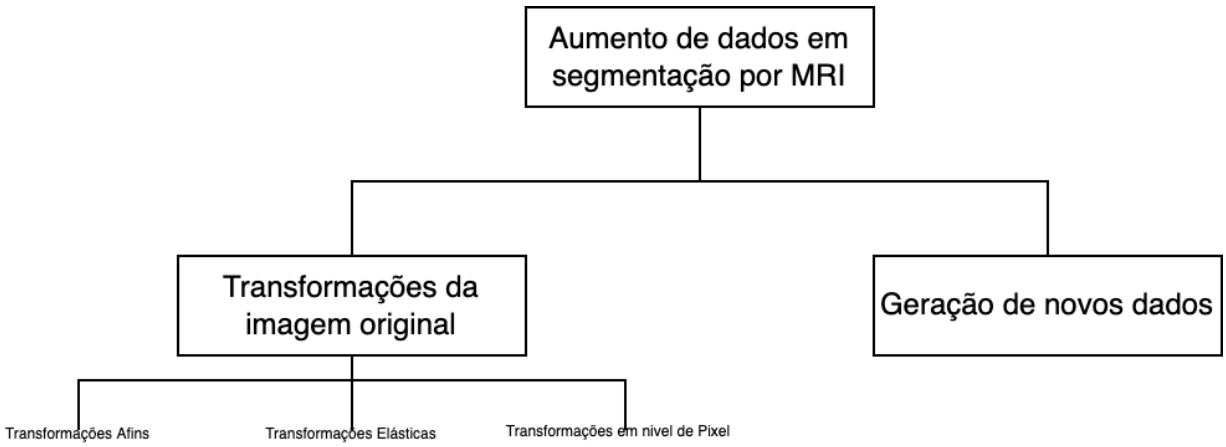


Figura 2.27: *Diagrama das técnicas utilizadas para o aumento de dados em imagens adiquiridas utilizando MRI. Imagem adaptada de (NALEPA; MARCINKIEWICZ; KAWULOK, 2019)*

que seriam inaceitáveis em um contexto médico. A confiabilidade dos dados gerados é crucial e a capacidade das GANs de garantir isso pode ser limitada (**??**).

2.5.1 Transformações Afins

O aumento de dados em imagens utilizando transformações afins envolve operações como rotação, translação, corte, *flipping* e cisalhamento, dentre outras (NALEPA; MARCINKIEWICZ; KAWULOK, 2019; PEREIRA et al., 2016). A Figura 2.28 ilustra algumas dessas transformações aplicadas a uma imagem de ressonância magnética do cérebro.

Porém, é importante destacar que o uso indiscriminado de transformações afins pode gerar imagens que não são anatomicamente precisas, como é o caso da transformação de cisalhamento na imagem exemplificada na figura.

2.5.2 Transformações Elásticas

As transformações elásticas podem modificar bastante as imagens originais. No caso de imagens volumétricas por MRI, essas transformações podem gerar imagens sintéticas totalmente irreais como mostrado por Mok e Chung (2018) e ilustrado pela Figura 2.29. Essas imagens sintéticas podem ser prejudiciais no treinamento de modelos baseados em NN (LORENZO et al., 2019). Contudo, o trabalho desenvolvido por Chaitanya et al. (2019) indica que imagens sintéticas não realistas podem melhorar o desempenho na segmentação de MRI cardíacas, deixando a questão em aberto.

2.5.3 Transformações em Nível de Pixel

As transformações em nível de pixel não alteram a geometria ou a forma da imagem. Quando utilizada para o aumento de dados, essas transformações buscam alterar a intensidade do pixel da imagem. Esse tipo de operação é bastante útil em tarefas de segmentação de

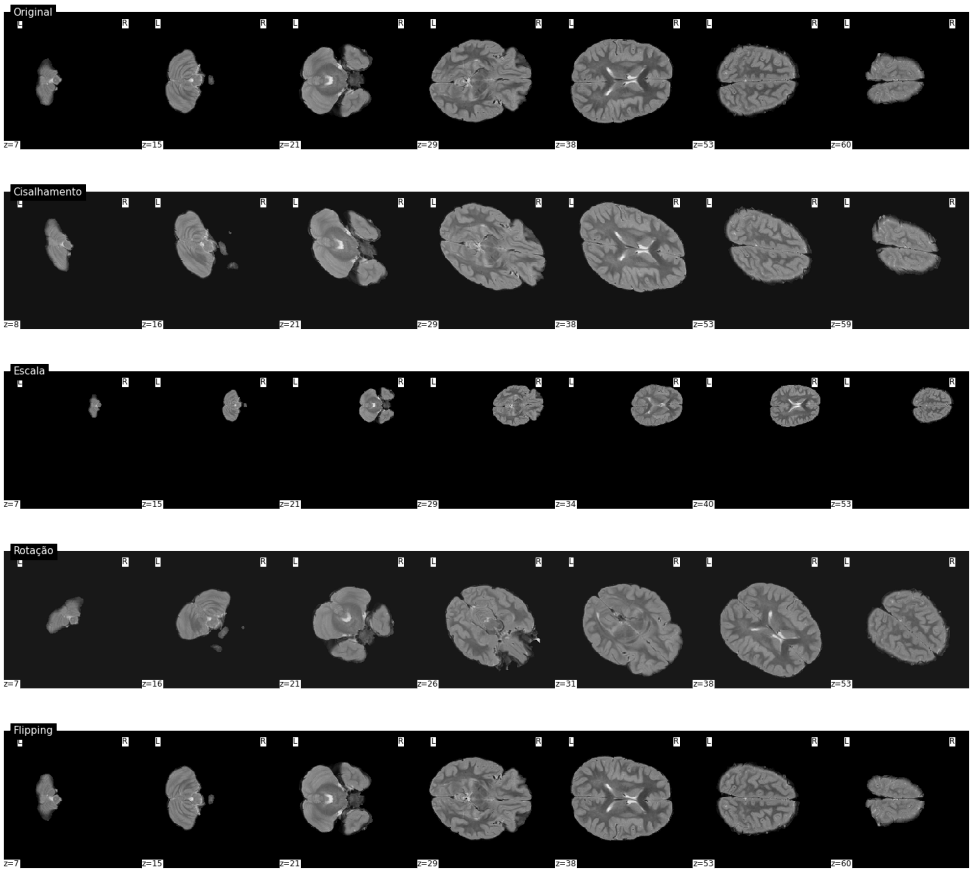


Figura 2.28: *Esta imagem ilustra os resultados de transformações afins aplicadas a imagens de ressonância magnética do encéfalo. Na primeira linha da figura, podemos observar as imagens originais em diferentes cortes no plano axial. Na segunda linha, as imagens resultantes são geradas a partir das transformações afins aplicadas à imagem original, sendo respectivamente: cisalhamento, escala, rotação e flipping.*

imagens por MRI. Diferentes equipamentos para aquisição dessas imagens (*scanners*), além de locais diferentes, podem gerar imagens heterogêneas em intensidades de pixel, gradientes de intensidade ou saturação.

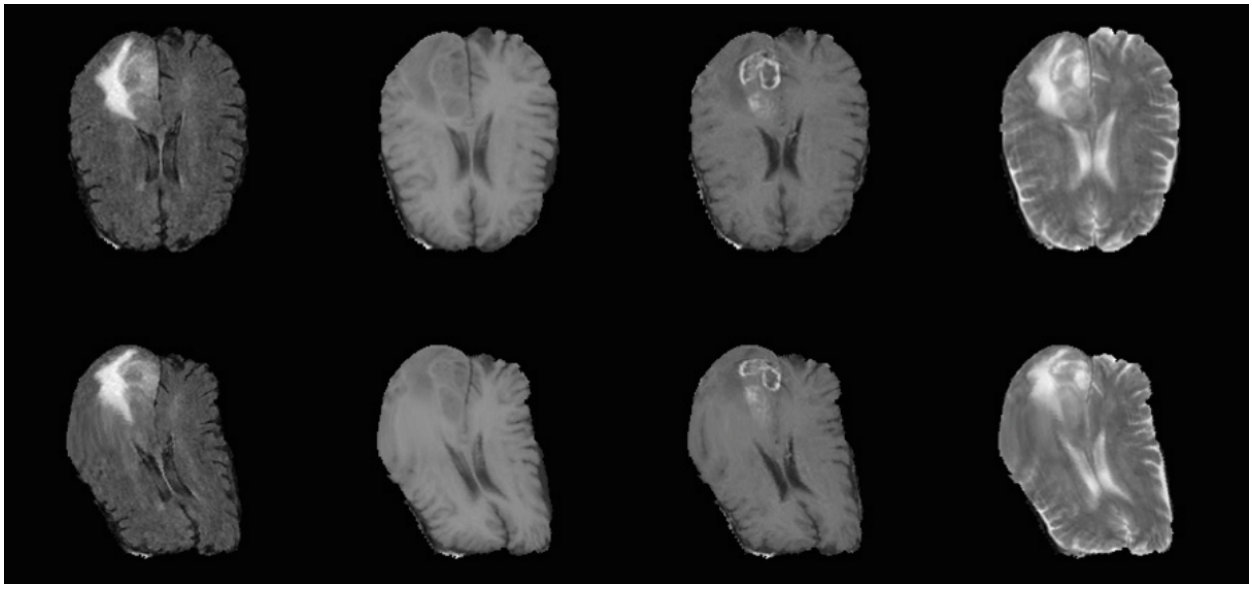


Figura 2.29: *Resultado de transformações elásticas aplicadas a imagens por MRI do encéfalo. As figuras na primeira linha são as imagens originais e na segunda linha é a mesma imagem depois da transformação elástica. Imagem adaptada de (MOK; CHUNG, 2018)*

Capítulo 3

Método proposto

Neste capítulo é apresentado, inicialmente, o *pipeline* da pesquisa na qual este trabalho está inserido. Na sequência, são apresentados os procedimentos utilizados para a aquisição, anotação, tratamento e anonimização dos dados, seguidos pelo método proposto nesta pesquisa. Por fim, são discutidos os materiais que estão sendo utilizados no desenvolvimento deste trabalho.

3.1 Pipeline da Pesquisa

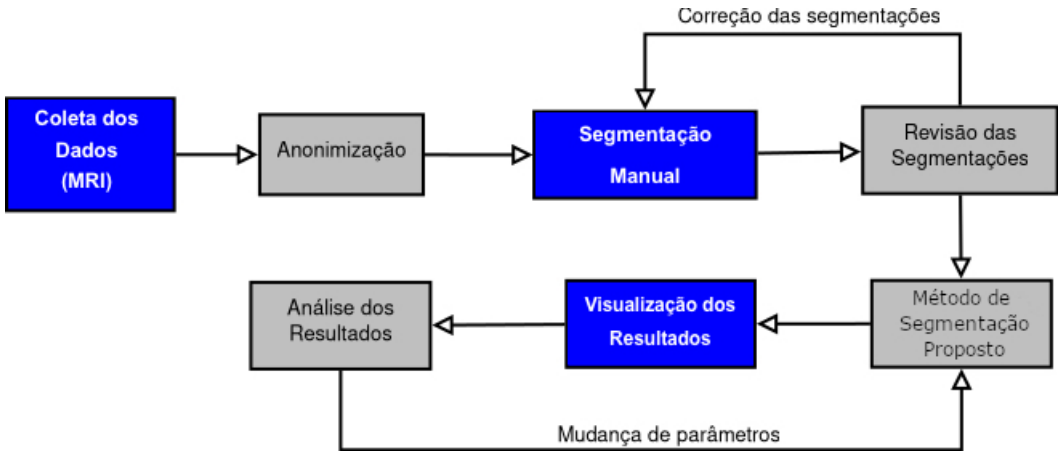


Figura 3.1: *Fluxograma mostrando os passos do desenvolvimento do projeto do qual a pesquisa faz parte.*

Esta pesquisa faz parte de um projeto no qual participam diversos especialistas. A metodologia da pesquisa realizada é ilustrada na Figura 3.1. O pipeline indica que o processo começa com a coleta de dados, realizada por colaboradores do HC-USP. Os radiologistas selecionaram um conjunto de imagens de MRI pediátricas com as características definidas no projeto original. As imagens passam pelos processos padrão de anonimização do HC-USP. As imagens foram anotadas por estudantes e especialistas de supervisão. Foram então usadas pelo método de segmentação e validação proposto, que inclui ferramentas para visualização

e análise dos resultados. As etapas da formação do conjunto de dados são discutidas na Seção 3.2. O método de segmentação proposto é detalhado na Seção 3.3. Por fim, os resultados experimentais são apresentados no Capítulo 4.

3.2 Coleta dos Dados e Segmentação Manual

Nota-se uma falta de conjuntos de dados de ressonância magnética neonatal e pediátrica disponíveis publicamente. Além disso, os conjuntos de dados públicos existentes não são adequados para a tarefa proposta, uma vez que possuem dados anotados para outros fins. A exemplo disso, o desafio de segmentação de tumor cerebral (*Brain Tumor Segmentation* (BraTS))[1] (MENZE et al., 2014) contém dados para segmentação de tumor em ressonâncias magnéticas de adultos e iSeg[2] (SUN et al., 2020) consiste em dados para segmentação de tecido cerebral de neonatos com idades variando de 2 semanas a 12 meses.

Portanto, foi criado um novo conjunto de dados por meio de volumes de ressonância magnética ponderados em T2 adquiridos com um *scanner 1.5T Philips Ingenia* em parceria com o Hospital das Clínicas (HC) da USP. O tamanho do conjunto de dados utilizados consiste em 32 volumes, sendo 11 deles obtidos no plano axial e os outros 21 no plano sagital. Os tamanhos dos *voxels* variaram de $0,5 \times 0,5 \times 0,9$ a $0,9 \times 0,9 \times 2$ em mm. As idades dos pacientes no conjunto de dados variam de alguns meses (0 anos) a 18 anos. Essa distribuição é ilustrada pela Figura 3.2 e, como pode ser observado, há uma predominância de pacientes entre 0 e 4 anos de idade.

A primeira fase (ilustrada pela Figura 3.3), na qual este projeto está inserido, se inicia na aquisição, pelos radiologistas, dessas imagens durante o exame utilizando a ressonância magnética. Após essa coleta, é realizada a anonimização das imagens de modo que seja impossível relacionar as imagens com o paciente.

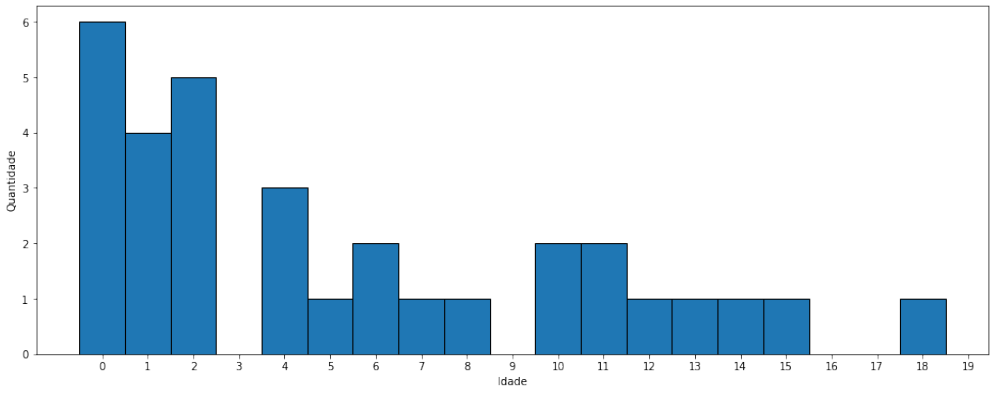


Figura 3.2: *Histograma das idades dos pacientes no conjunto de dados.*

Os dados, após a anonimização, foram padronizados para a visualização do plano axial, transpondo as 21 amostras sagitais. Após isso, foi realizado um pré-processamento das ima-

[1]Disponível em: <http://braintumorsegmentation.org/>.
[2]Disponível em: <https://iseg2017.web.unc.edu/>.

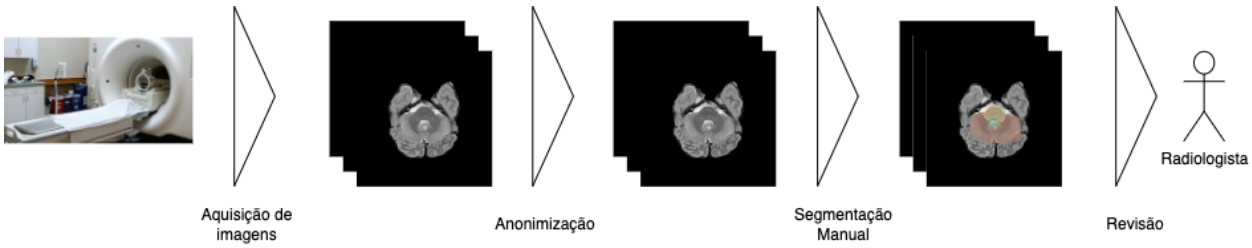


Figura 3.3: *Fluxograma detalhando a Fase I do projeto. Essa fase se inicia na aquisição das imagens na clínica, passa pela anonimização dos dados e termina na segmentação realizada manualmente.*

gens volumétricas, sendo essa composta de três etapas: primeiro é utilizada uma ferramenta para a extração do objeto de interesse (i.e. o encéfalo) (*Brain Extraction Tool* (BET)). Depois foi utilizada a normalização da intensidade dos *pixels*. Por fim, é realizada uma correção do sinal de campo de polarização (*Bias-Field Correction* (BFC)). Esses passos são ilustrados pela Figura 3.4.

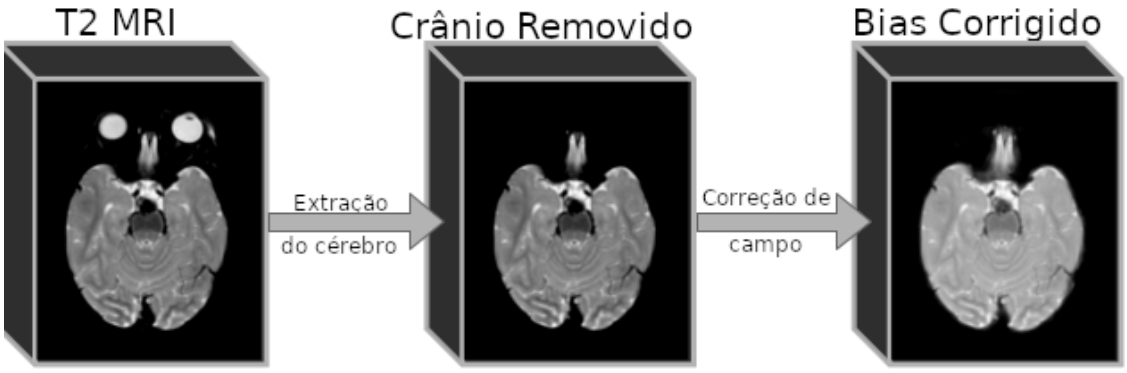


Figura 3.4: *Passos realizados para o pré-processamento no MRI dataset utilizado.*

Por fim, foi realizada a segmentação manual das áreas de interesse (IV-ventrículo, tronco cerebelar e cerebelo) das imagens do conjunto de dados. Essas segmentações foram realizadas por médicos residentes do HC e pelo autor. Cada uma dessas segmentações foram revisadas e corrigidas por radiologistas do HC. A Figura 4.8 contém o *software* utilizado para realizar essas segmentações manuais, a Seção 3.4 traz mais informações sobre esse *software*.

3.3 Método Proposto

O método proposto por esta pesquisa se baseia em CNNs e FCNs dentro de um *pipeline* de segmentação automática da fossa posterior do encéfalo pediátrico (Figura 1.3). O novo método utiliza um conceito de rede generalista e redes especialistas, no qual a primeira rede é responsável por uma segmentação utilizando o volume completo enquanto as redes especialistas se especializam em partes específicas do volume aproveitando da segmentação anterior, realizada pela rede generalista, para o estabelecimento das regiões de interesse (ROI).

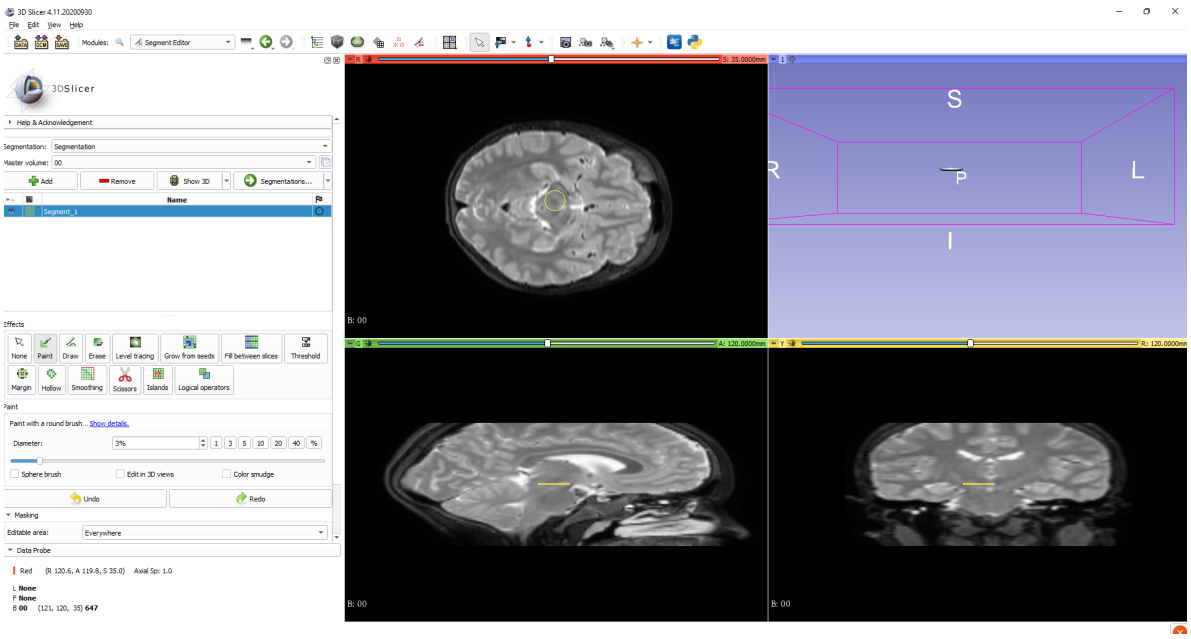


Figura 3.5: *Software utilizado para realizar a segmentação manual.*

Para a condução deste estudo, optamos por selecionar algumas arquiteturas previamente validadas em tarefas semelhantes de segmentação de imagens médicas volumétricas, cada uma trazendo suas próprias características distintas, bem como pontos fortes e limitações intrínsecas. Especificamente, neste trabalho, incluímos as arquiteturas Vnet, Unet e HighResNet, conforme discutidas no Capítulo 2. Vale ressaltar que existe a possibilidade de incorporar outras redes à essa abordagem de pipeline. Para uma análise mais detalhada do método proposto, por favor, consulte a Subseção 3.3.1.

3.3.1 Rede Generalista e Especialista

Os primeiros experimentos mostraram que *patching* pequenos de volumes das ressonâncias magnéticas prejudicava seriamente o desempenho dos modelos, pois a segmentação dessas imagens médicas mostrou ser altamente dependente do contexto espacial dos *voxels*. Em outras palavras, as propriedades do tecido circundante não imediato são muito importantes para a classe de segmentação prevista. Dessa forma, na tentativa de contornar esse problema e evitar a utilização de *patching*, foi utilizada uma arquitetura composta de duas etapas: 1) uma rede generalista ($\mathcal{G}$) para realizar uma segmentação aproximada de cada uma das classes; 2) duas redes especialistas ($\mathcal{S}_1$ e $\mathcal{S}_2$) para realizar uma segmentação mais fina nos locais de interesse. A mesma arquitetura foi utilizada tanto para a rede generalista ($\mathcal{G}$) quanto para as redes especialistas ($\mathcal{S}_1$ e $\mathcal{S}_2$), realizando experimentos com cada uma das redes apresentadas na Seção 3.3. A Figura 3.6 ilustra esse pipeline de redes para realizar a segmentação dos volumes.

De maneira específica, primeiro a rede generalista ($\mathcal{G}$) realizará uma segmentação das áreas de interesse (cerebelo, IV-ventrículo e tronco cerebelar) utilizando o redimensionamento do volume completo para o tamanho $128 \times 128 \times 64$. Esse tamanho de volume foi

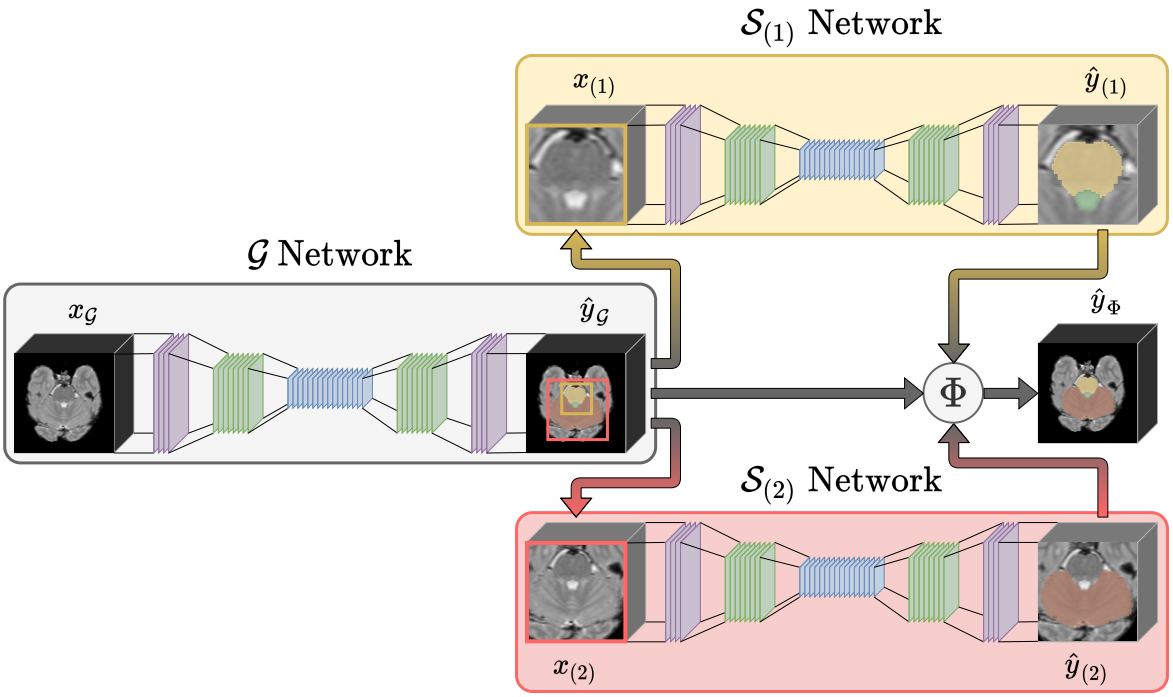


Figura 3.6: *Pipeline de redes proposto para a segmentação das estruturas. O volume completo $x_{\mathcal{G}}$ primeiro passa pela rede generalista ($\mathcal{G}$), no qual realiza uma predição aproximada $\hat{y}_{\mathcal{G}}$. De $\hat{y}_{\mathcal{G}}$ são extraídos dois volumes: $x_{(1)}$, que serve de entrada para a rede $\mathcal{S}_1$ para realizar as predições $\hat{y}_{(1)}$ do IV-ventrículo e do tronco cerebelar; e $x_{(2)}$, entrada para a rede $\mathcal{S}_2$ que realiza uma segmentação no cerebelo $\hat{y}_{(2)}$. Após isso, os resultados das predições são combinados por meio da função Φ, gerando como resultado a segmentação $\hat{y}_{\Phi}$.*

escolhido porque experimentos exploratórios mostraram que ele oferecia um compromisso aceitável entre o custo de memória[3] e o tamanho do volume. Após essa segmentação, foram utilizadas duas redes especialistas ($\mathcal{S}_1$ e $\mathcal{S}_2$), pois, foi observado que o tamanho do volume $128 \times 128 \times 64$ após simplesmente redimensionar os volumes era muito baixo para atingir limites de segmentação suaves nas bordas entre uma classe e outra.

Assim, a rede $\mathcal{S}_1$ é especializada na região dentro e ao redor do tronco cerebral e 4^{th} ventrículo, aprendendo a segmentar apenas uma pequena ROI prevista por $\mathcal{G}$ e redimensionado para $64 \times 64 \times 128$ *voxels*. Já a rede $\mathcal{S}_2$ aprende a realizar segmentação refinada no cerebelo, utilizando a ROI prevista por $\mathcal{G}$ e redimensionada para um volume de $64 \times 128 \times 64$. A escolha dos tamanhos diferentes de entrada entre uma rede especialista e outra, está relacionada com a forma das estruturas.

Cada uma das ROIs alimentadas para ambas especialistas são preenchidas com 5 *voxels* em cada lado, antes de realizar o redimensionamento das regiões, para incluir as estruturas completas mesmo no caso em que $\mathcal{G}$ perca grandes regiões nos limites das estruturas desejadas. Após isso, os resultados das predições são combinados por meio da função Φ, gerando como resultado a segmentação $\hat{y}_{\Phi}$. O Algoritmo 2, descreve o pipeline de segmentação após o pré-processamento. No primeiro momento, o objetivo é iniciar de forma pseudo-aleatória

[3]Esse custo está relacionado com as memórias disponíveis nas unidades de processamento gráfico (*Graphics Processing Units* (GPU)) utilizadas nos experimentos deste trabalho.

as redes generalistas e as especialistas (dos modelos citados (INI_RANDOM)). Então, é realizado o treinamento conforme descrito acima para cada um dos modelos ($f^i_{\mathcal{G}}$). Após, são definidas as regiões de corte para a utilização das redes especialistas durante o treinamento. Finalizado o treinamento da rede generalista e das redes especialistas, é realizada a validação dessas redes, primeiramente realizando essa validação na rede generalista ($\hat{\mathbf{y}}^{ts}_{\mathcal{G}} \leftarrow f^i_{\mathcal{G}}(\mathbf{x}^{ts})$) então esse resultado é utilizado para obter as regiões para a validação das redes especialistas. A fusão dos resultados é então realizada sobre os resultados dessas predições da validação utilizando uma prioridade das classes menores sobre as classes maiores no estágio 1 da fusão entre as redes especialistas. Por fim, com o objetivo de aproveitar cada um dos modelos treinados, é realizada uma votação majoritária sobre as regiões segmentadas por cada um dos pipelines de cada um dos modelos, durante a validação, a fim de obter somente um volume segmentado.

Algoritmo 2 Algoritmo de fusão tardia para redes especializadas e generalistas para um conjunto de arquiteturas de segmentação em uma única divisão entre os conjuntos de treinamento ($\{\mathbf{x}^{tr}, \mathbf{y}^{tr}\}$) e teste ($\{\mathbf{x}^{ts}\}$, $\{\mathbf{y}^{ts}\}$). E o conjunto das arquiteturas $\mathbf{f}$.

Procedimento SEGMENTAÇÃO_PIPELINE($\mathbf{x}^{tr}, \mathbf{y}^{tr}, \mathbf{x}^{ts}, \mathbf{y}^{ts}, \mathbf{f}_{\mathcal{G}}, \mathbf{f}_{\mathcal{S}_{(1)}}, \mathbf{f}_{\mathcal{S}_{(2)}}$)
  INI_RANDOM($\mathbf{f}_{\mathcal{G}}, \mathbf{f}_{\mathcal{S}_{(1)}}, \mathbf{f}_{\mathcal{S}_{(2)}}$) ▷ Inicialização pseudo aleatória
  Para cada: $f^i_{\mathcal{G}}, f^i_{\mathcal{S}_{(1)}}, f^i_{\mathcal{S}_{(2)}} \in \mathbf{f}_{\mathcal{G}}, \mathbf{f}_{\mathcal{S}_{(1)}}, \mathbf{f}_{\mathcal{S}_{(2)}}$ **faça** ▷ Treinamento
    Treinar $f^i_{\mathcal{G}}$ sobre $\{\mathbf{x}^{tr}, \mathbf{y}^{tr}\}$
    Computar predições $\hat{\mathbf{y}}^{tr}_{\mathcal{G}} \leftarrow f^i_{\mathcal{G}}(\mathbf{x}^{\mathbf{tr}}_{\mathcal{G}})$
    Obter $\mathbf{x}^{tr}_{(1)}$ e $\mathbf{x}^{tr}_{(2)}$ pelo corte de $\mathbf{x}^{tr}$ de acordo com $\hat{\mathbf{y}}^{tr}_{\mathcal{G}}$ ▷ Obtenção dos cortes
    Obter $\mathbf{y}^{tr}_{(1)}$ e $\mathbf{y}^{tr}_{(2)}$ pelo corte de $\hat{\mathbf{y}}^{tr}_{\mathcal{G}}$ ▷ para o treinamento das redes especialistas
    Treinar $f^i_{\mathcal{S}_{(1)}}$ sobre $\{\mathbf{x}^{tr}_{(1)}, \mathbf{y}^{tr}_{(1)}\}$
    Treinar $f^i_{\mathcal{S}_{(2)}}$ sobre $\{\mathbf{x}^{tr}_{(2)}, \mathbf{y}^{tr}_{(2)}\}$
  fim para
  $\hat{\mathbf{Y}} \leftarrow \{\}$ ▷ Lista das Predições
  Para cada: $f^i_{\mathcal{G}}, f^i_{\mathcal{S}_{(1)}}, f^i_{\mathcal{S}_{(2)}} \in \mathbf{f}_{\mathcal{G}}, \mathbf{f}_{\mathcal{S}_{(1)}}, \mathbf{f}_{\mathcal{S}_{(2)}}$ **faça** ▷ Validação
    Computar predições $\hat{\mathbf{y}}^{ts}_{\mathcal{G}} \leftarrow f^i_{\mathcal{G}}(\mathbf{x}^{ts})$
    Obter $\mathbf{x}^{ts}_{(1)}$ e $\mathbf{x}^{ts}_{(2)}$ pelo corte $\mathbf{x}^{ts}$ de acordo com $\hat{\mathbf{y}}^{ts}_{\mathcal{G}}$
    Computar $\hat{\mathbf{y}}^{ts}_{\mathcal{S}_{(1)}} \leftarrow f^i_{\mathcal{S}_{(1)}}(\mathbf{x}^{ts}_{(1)})$ e $\hat{\mathbf{y}}^{ts}_{\mathcal{S}_{(2)}} \leftarrow f^i_{\mathcal{S}_{(2)}}(\mathbf{x}^{ts}_{(2)})$
    $\hat{\mathbf{y}}^{ts}_{\Phi} \leftarrow \Phi(\hat{\mathbf{y}}^{ts}_{(1)}, \hat{\mathbf{y}}^{ts}_{(2)})$ ▷ Estagio 1 fusão.
    $\hat{\mathbf{Y}} \leftarrow \hat{\mathbf{Y}} + \hat{\mathbf{y}}^{ts}_{\Phi}$
  fim para
  $\hat{\mathbf{y}}^{ts}_{maj} \leftarrow majority\_voting(\hat{\mathbf{Y}})$ ▷ Voto majoritário entre as arquiteturas
  Devolve $\hat{\mathbf{y}}^{ts}_{maj}$
fim Procedimento

Dessa forma, além de preservar o contexto espacial, o redimensionamento e a utilização de redes especialistas também é mais simples e barato computacionalmente, de se implementar para ser aplicado em amostras com resoluções e tamanhos de *voxels* variados como os dados adquiridos para essa pesquisa. Também é mais barato mesclar previsões para segmentações de ROI de redes especializadas em uma única previsão global do que juntar um grande conjunto de pequenos patches sobrepostos. Esse processo muitas vezes requer uma

heurística de pós-processamento (por exemplo, filtragem de moda, morfologia de imagem ou mesmo Bayesiana adicional modelos) a ser aplicada no volume previsto. Por fim, outra vantagem do treinamento especializado é a mitigação inerente do desequilíbrio do rótulo ao usar volumes menores em torno do ROI nos especialistas pois, as estruturas ocupam tamanhos relativamente maiores em comparação com o volume.

3.3.2 Treinamento e Avaliação

Para evitar que alguma das arquiteturas obtivesse resultados melhores, nos experimentos realizados, simplesmente pela quantidade de parâmetros, foi modificada, cada uma das arquiteturas (i.e. número de filtros em cada camada, número de camadas, etc.). Dessa forma, as arquiteturas foram ajustadas de modo que um *mini-batch* contendo 2 amostras se adequassem a uma única GPU com 8 GB de memória ou em algumas GPUs com tamanho de *mini-batch* de 4.

De modo semelhante, os hiperparâmetros também foram padronizados para todas as arquiteturas. Assim, o treinamento foi realizado com a duração de 400 épocas usando o otimizador Adam (KINGMA; BA, 2014), um inicial *learning rate* de 1×10^{-2} reduzindo-o pela metade a cada 80 épocas, um regularização $L2$ de 5×10^{-5} e 0.5 de *momentum*. Por fim, a função de perda (L_T) escolhida para os experimentos realizados foi uma combinação entre a *Cross-Entropy* (L_{CE}) e a *Dice* (L_{Dice}) descrita pela Equação 3.1:

$$\mathcal{L}_T = \mathcal{L}_{CE} + \mathcal{L}_{Dice}. \tag{3.1}$$

Com o objetivo de quantificar os erros de segmentação, foram utilizadas três métricas de avaliação: 1) o coeficiente Sørensen–Dice (DSC); 2) a distância entre as duas superfícies, como a distância média (μSD) e 95^0 percentil da distância de Hausdorff ($HD95$). As distâncias entre as superfícies podem estimar os limites superiores das distâncias entre os objetos segmentados manualmente e as previsões da rede enquanto que para uma avaliação por voxel objetiva a medida Dice (DSC) é menos cara de se calcular e é largamente utilizada com esse propósito.

Como o conjunto contém poucos dados rotulados, a configuração experimental utilizada teve como objetivo usar a maior quantidade possível de amostras no treinamento, ao mesmo tempo que aproveita todo o conjunto rotulado para realizar a avaliação. Portanto, foi empregado um esquema de validação cruzada (*cross-validation*) de 5 *folds* nas 32 amostras de nosso conjunto de dados. Para o DSC foi calculada a média entre os *folds* de validação como um todo. Contudo, as distâncias da superfície e os volumes da estrutura são calculados separadamente para cada amostra, pois são anotações inerentemente no nível da instância.

3.4 Implementação

Para os experimentos realizados nesta pesquisa, foram utilizados os computadores do laboratório de visão computacional do Instituto de Matemática e Estatística (IME) da Universidade de São Paulo (USP).

O trabalho foi desenvolvido usando a linguagem *python* em sua versão 3.7. Os motivos da escolha dessa linguagem são tanto por sua facilidade de escrita, que aumenta assim a produtividade (BORGES, 2014), quanto por ela apresentar um grande número de *frameworks* desenvolvidos por terceiros que facilitam seu uso, adicionando assim algumas funcionalidades para as estruturas de dados. A seguir serão detalhadas as principais bibliotecas e ferramentas utilizadas para o desenvolvimento desta pesquisa:

- **PyTorch:** PyTorch[4] é uma biblioteca de aprendizado de máquina de código aberto muito utilizada em áreas como visão computacional e processamento de imagens naturais. Foi inicialmente desenvolvida pelo laboratório de IA do Facebook (FAIR) tendo como base a biblioteca Torch. Oferece computação de tensores com forte aceleração por meio de unidades de processamento gráfico (GPU) e fornece uma forma de construir arquiteturas de redes neurais profundas em um sistema de diferenciação automática [5].

- **Scikit-image:** Scikit-image[6] (WALT et al., 2014) é uma coleção de algoritmos para processamento de imagens desenvolvida para o python. Ela é uma biblioteca de código aberto mantida por um grande número de voluntários.

- **Nibabel:** Nibabel[7] é um pacote para a linguagem Python que fornece funções de leitura e escrita para alguns dos mais comuns formatos de arquivos médicos e de neuroimagem, incluindo: ANALYZE (*plain*, SPM99, SPM2 e posterior), GIFTI, NIfTI1, NIfTI2, CIFTI-2, MINC1, MINC2, AFNI BRIK / HEAD, MGH e ECAT, bem como Philips PAR/REC. A licensa de uso desse pacote é a MIT[8] com alguns códigos incluído ao pacote licenciados sobre BSD[9], portanto, de livre utilização e de código aberto.

- **3DSlicer:** 3DSlicer[10] é um pacote de *software* gratuito, de código aberto e multiplataforma amplamente usado para pesquisas médicas, biomédicas e de imagens relacionadas. Nesta pesquisa esse *software* foi utilizado para realizar a segmentação manual das imagens volumétricas.

[4] <https://pytorch.org/>
[5] Ver mais sobre diferenciação automática em *machine learning* em (BAYDIN et al., 2018)
[6] <https://scikit-image.org/>
[7] <https://nipy.org/nibabel/>
[8] <https://opensource.org/licenses/mit-license.php>
[9] <https://opensource.org/licenses/BSD-3-Clause>
[10] <https://www.slicer.org/>

Capítulo 4

Resultados e Discussão

Este capítulo traz os resultados obtidos pelos experimentos descritos no Capítulo 3. Inicialmente, são apresentados os resultados quantitativos e qualitativos obtidos pela rede generalista. Depois, são apresentados os resultados obtidos ao adicionar as redes especialistas na pipeline de segmentação. Por fim, são apresentados os resultados obtidos utilizando a votação majoritária entre as diferentes arquiteturas.

4.1 Configuração das redes

Conforme descrito no Capítulo 3, foram utilizadas três redes diferentes introduzidas no *pipeline* de segmentação. A Tabela 4.1 apresenta os hiperparâmetros utilizados para o treinamento de cada uma das arquiteturas. Foram utilizadas 400 épocas de treinamento, com uma taxa de aprendizagem de 0,01. O valor de decaimento de peso (*weight decay*) foi de 0,00005 e o valor de momentum foi de 0,5. O valor de step foi de 80 e o valor de gamma foi de 0,5. Esses hiperparâmetros foram selecionados com base em experimentações anteriores e considerados adequados para a tarefa em questão.

Para fins de comparação, os resultados quantitativos de IoU e DSC obtidos pela rede MED3D foram incorporados na mesma pipeline de segmentação, seguindo uma abordagem semelhante à das outras redes. Além disso, os mesmos hiperparâmetros foram utilizados. Vale ressaltar que os resultados da rede MED3D não foram considerados na votação majoritária entre os diferentes conjuntos de resultados.

Tabela 4.1: *Hiper parâmetros utilizados para cada uma das arquiteturas durante o seu treinamento.*

Parâmetro	**Valor**
Número de épocas	400
Taxa de aprendizagem	0,01
Weigth decay	0,00005
momentum	0,5
Step	80
Gamma	0,5

4.2 Resultados das Redes Generalistas

As redes generalistas representam as arquiteturas inseridas no *pipeline* de segmentação após o pré-processamento das imagens volumétricas e utilizadas para definir as regiões de corte. Tais regiões são então utilizadas pelas redes especialistas para uma segmentação mais específica. Os resultados referentes às generalistas são apresentados nesta seção.

A Tabela 4.2 apresenta os resultados da pontuação de DSC e IoU obtidos pelo pipeline de segmentação utilizando as redes generalistas. Os valores apresentados estão na forma de média ± e desvio padrão, considerando a divisão 5-fold descrita no Capítulo 3. Os melhores resultados médios estão destacados.

Os resultados mostram que a rede Vnet obteve a melhor pontuação média de DSC (0,824 ± 0,024), seguida pela Unet (0,822 ± 0,017) e pela HR3N (0,781 ± 0,041). Já para o IoU, a Unet obteve a melhor pontuação média (0,7322 ± 0,027), seguida pela Vnet (0,7260 ± 0,024) e pela HR3N (0,6713 ± 0,041). Neste experimento, a MED3D apresentou os piores resultados tanto em DSC quanto IoU.

Esses resultados indicam que as redes generalistas são capazes de segmentar as estruturas da fossa craniana posterior com um desempenho razoável, sendo a Vnet e a Unet as arquiteturas mais adequadas para este problema.

Tabela 4.2: *Pontuação DSC e IoU para as predições realizadas pela pipeline de segmentação utilizando as redes generalistas. Os valores apresentados estão da forma $\mu \pm \sigma$ sobre a divisão 5-fold descrita no Capítulo 3. Os melhores resultados médios estão destacados.*

Arquitetura	DSC ↑	IoU ↑
Vnet	**0,824 ± 0,024**	0,7260 ± 0,024
Unet	0,822 ± 0,017	**0,7322 ± 0,027**
HR3N	0,781 ± 0,041	0,6713 ± 0,041
MED3D	0,767 ± 0,022	0.6520 ± 0,025

Para uma melhor comparação entre as três arquiteturas, a Tabela 4.3 apresenta as métricas de distância obtidas para as classes IV-ventrículo, cerebelo e tronco cerebelar.

Essa tabela apresenta os resultados das distâncias μSD e HD95 para as predições realizadas pelas redes generalistas, sobre as estruturas dos 4 ventrículos, tronco e cerebelo. É possível observar que a Unet obteve os melhores resultados médios para as três estruturas avaliadas, com distâncias μSD menores do que as demais arquiteturas. Além disso, os valores de HD95 são maiores para todas as estruturas, o que indica que há valores discrepantes na predição de algumas imagens. Globalmente, as arquiteturas apresentaram desempenho semelhante, e a Unet teve um desempenho ligeiramente melhor em relação às outras arquiteturas.

Alguns resultados qualitativos obtidos pelas redes generalistas podem ser observados na Figura 4.1. É possível notar que a rede Vnet apresenta menos erros de segmentação evidentes quando comparada às redes Unet e HR3N, como pode ser observado nos cortes 15 e 31, respectivamente.

Tabela 4.3: *Distâncias μSD e HD95 para as predições realizadas pelas redes generalistas. Os valores apresentados estão da forma $\mu \pm \sigma$ sobre a divisão 5-fold descrita no Capítulo 3. Os melhores resultados médios estão destacados.*

Estr.	Arq.	$\mu SD \downarrow$			**HD95** $\downarrow$		
		4 Ventrículo	Tronco	Cerebelo	4 Ventriculo	Tronco	Cerebelo
$\mathcal{G}$	Unet	**1,70 ± 3,24**	**1,29 ± 0,71**	**1,02 ± 0,46**	**7,98 ± 8,93**	**6,29 ± 3,80**	**3,82 ± 1,71**
	Vnet	1,71 ± 2,17	2,37 ± 4,84	2,52 ± 5,27	12,46±16,34	13,06±18,53	15,54±24,28
	HR3N	**1,70 ± 2,93**	1,80 ± 1,84	1,59 ± 1,15	10,18±12,05	10,83±9,83	10,83±10,03

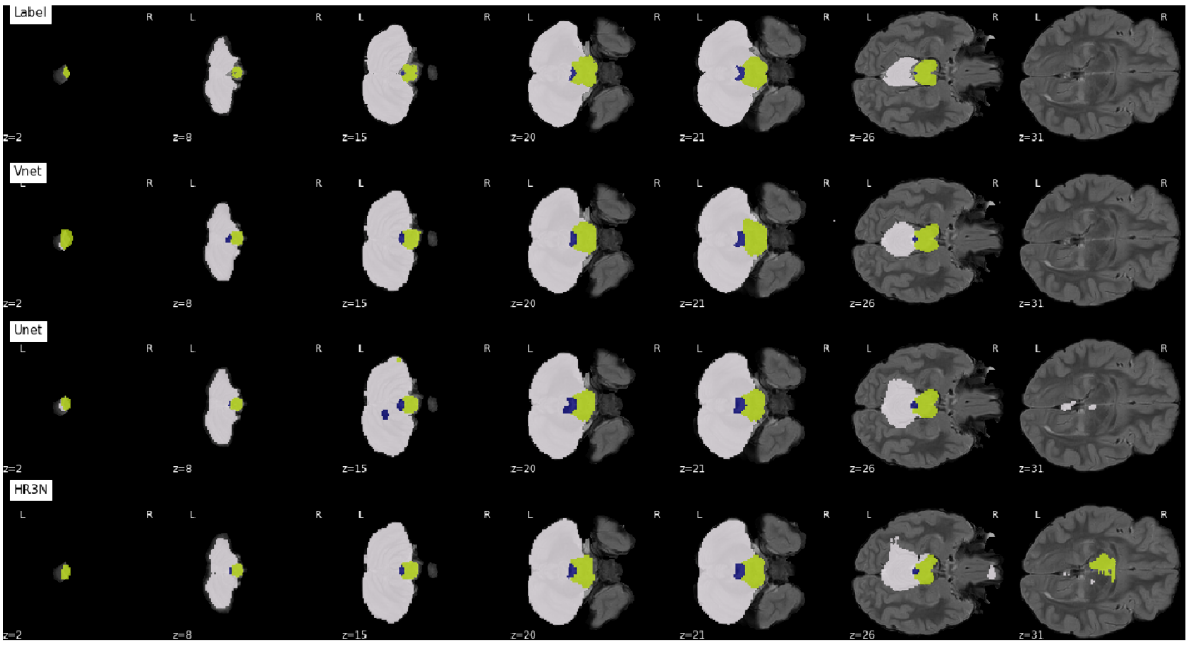


Figura 4.1: *Corte no plano axial para visualização dos resultados obtidos pelas redes generalistas em comparação com a segmentação de refeência Label.*

De maneira semelhante, a Figura 4.2 também mostra que as redes Unet e HR3N apresentam erros claros de segmentação nos cortes -169 e -117, respectivamente, enquanto a rede Vnet apresenta resultados mais precisos.

Esses exemplos qualitativos confirmam a avaliação inicial obtida pelos resultados quantitativos apresentados, ou seja, a rede Vnet obteve um melhor desempenho nesse conjunto de dados quando comparada às redes Unet e HR3N.

4.3 Resultados das Redes Especialistas e Votação Majoritária

A Tabela 4.4 apresenta os resultados de pontuação DSC obtidos por quatro estratégias de segmentação, comparando o desempenho de três diferentes arquiteturas de redes neurais: Vnet, Unet e HR3N. As quatro estratégias avaliadas foram: a rede generalista $\mathcal{G}$, a rede especialista 1 em conjunto com a rede generalista $\Phi(\mathcal{S}1, \mathcal{G})$, a rede especialista 2 em conjunto com a rede generalista $\Phi(\mathcal{G}, \mathcal{S}2)$ e a combinação das redes especialistas 1 e 2 $\Phi(\mathcal{S}1, \mathcal{S}2)$. Adicionalmente, a tabela apresenta os resultados obtidos pela votação majoritária utilizando

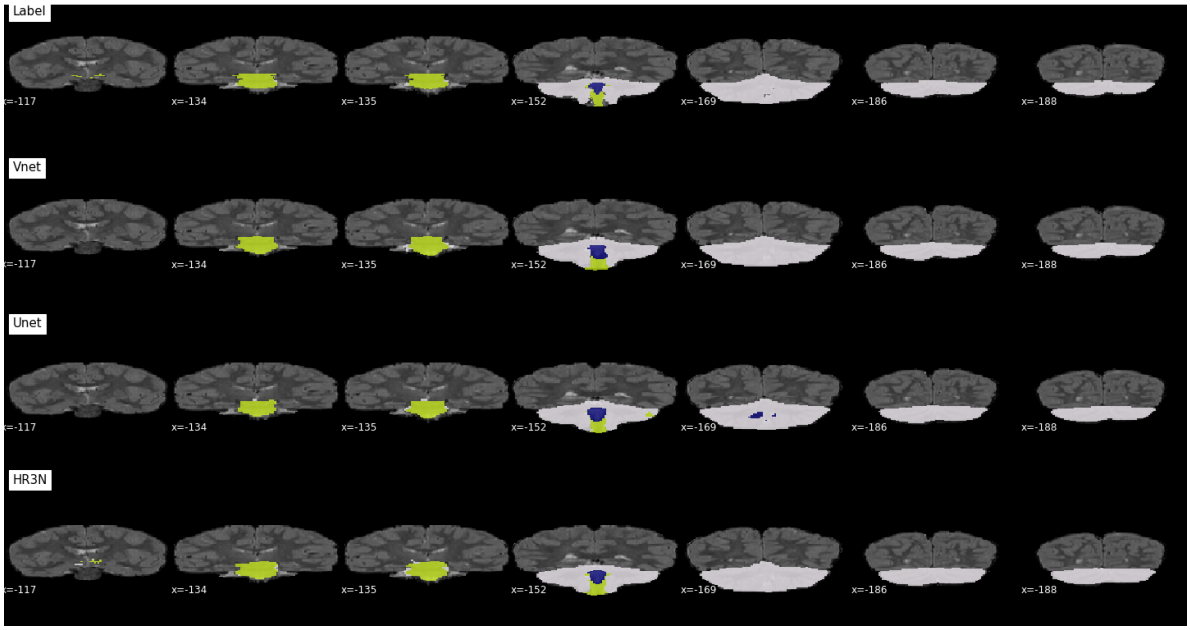


Figura 4.2: *Corte no plano coronal para visualização dos resultados obtidos pelas redes generalistas em comparação com a segmentação de refeência Label.*

as 3 redes em diferentes estratégias, bem como o resultado da votação majoritária entre as estratégias $\mathcal{G}$, $\mathcal{S}1$ e $\mathcal{S}2$ para cada uma das três arquiteturas.

Observa-se que a estratégia que combina ao menos uma das redes especialistas obteve os melhores resultados em todas as arquiteturas testadas. Em particular, a rede especialista 1 com a rede especialista 2 atingiu a maior pontuação DSC média de todas as estratégias e arquiteturas testadas (0,857 $\pm$ 0,021), seguidas pela combinação entre a rede especialista 2 com a rede generalista ($\Phi(\mathcal{G}, \mathcal{S}_2)$) com uma pontuação média de 0,825 $\pm$ 0,018.

Com exceção da estratégia $\Phi(\mathcal{S}_1, \mathcal{S}_2)$ utilizando a Vnet, a votação majoritária entre as três redes apresentou um desempenho médio ainda melhor. A votação majoritária dos resultados obtidos pelas redes ($\mathcal{G}$, $\mathcal{S}1$ e $\mathcal{S}2$) nas diferentes arquiteturas obteve uma pontuação média de 0,8478 $\pm$ 0,019. Esses resultados indicam que combinar as saídas das diferentes estratégias de segmentação pode levar a um melhor desempenho do pipeline de segmentação.

Os resultados das métricas de distância obtidos para avaliar os experimentos estão apresentados na Tabela 4.5. Os resultados estão separados pelos resultados obtidos nas métricas de distância em cada classe segmentada. Observa-se uma melhoria na distância de superfície média (μSD) ao utilizar redes especialistas para segmentar o quarto ventrículo nas arquiteturas Vnet e HighResNet, sendo que a melhor combinação foi a Vnet com ambas as redes especialistas. No entanto, a utilização de redes especialistas resultou em pior desempenho para a arquitetura Unet na segmentação do quarto ventrículo.

Quanto às outras duas classes (tronco e cerebelo), a métrica μSD apresentou melhorias em todas as arquiteturas com o uso de redes especialistas. Destaca-se a arquitetura Unet, que obteve os melhores resultados em relação a essa métrica.

Em relação à métrica de distância de Hausdorff (HD95) apresentada na Tabela 4.5,

Tabela 4.4: *Resultados da pontuação DSC obtidos por quatro estratégias de segmentação utilizando três arquiteturas de redes neurais diferentes (Vnet, Unet e HR3N). As quatro estratégias testadas foram: a rede generalista $\mathcal{G}$, a rede especialista 1 com a rede generalista $\Phi(\mathcal{S}1, \mathcal{G})$, a rede especialista 2 com a rede generalista $\Phi(\mathcal{G}, \mathcal{S}2)$ e a rede especialista 1 com a rede especialista 2 $\Phi(\mathcal{S}1, \mathcal{S}2)$. Como forma de comparação, foi adicionado os resultados obtidos pela rede Med3D dentro da mesma pipeline de segmentação (os resultados dela não foram incluídos na votação majoritária). Os valores apresentados na tabela estão da forma $\mu \pm \sigma$ sobre a divisão 5-fold descrita na seção de método do trabalho. Os resultados destacados em negrito são àqueles que desempenharam melhor na sua respectiva arquitetura.*

Estratégia	Vnet	Unet	HR3N	Todas	Med3D
$\mathcal{G}$	0,824 ± 0,024	0,822 ± 0,017	0,781 ± 0,041	0,8132 ± 0,0483	0,767 ± 0,022
$\Phi(\mathcal{S}_1, \mathcal{G})$	0,848 ± 0,022	0,802 ± 0,044	0,814 ± 0,032	0,8339 ± 0,025	0,825 ± 0,024
$\Phi(\mathcal{G}, \mathcal{S}_2)$	0,821 ± 0,029	**0,825 ± 0,018**	0,817 ± 0,017	0,8359 ± 0,022	0,784 ± 0,021
$\Phi(\mathcal{S}_1, \mathcal{S}_2)$	**0,857 ± 0,021**	0,811 ± 0,042	**0,820 ± 0,036**	0,8382 ± 0,029	**0,839 ± 0,018**
Majority	-	-	-	**0,8478 ± 0,019**	-

houve melhoria em todas as classes com o uso de redes especialistas nas arquiteturas Vnet e HighResNet. No entanto, a arquitetura Unet apresentou melhoria apenas na segmentação do cerebelo, obtendo o melhor resultado entre todas as arquiteturas do experimento.

Na votação majoritária, os resultados obtidos superaram quaisquer outras estratégias nas três classes tanto na métrica μSD quanto na métrica HD95, mostrando que a votação majoritária melhora a estrutura geral da predição ao combinar os diferentes resultados obtidos pelas redes.

Tabela 4.5: *Distâncias μSD e HD95 para as predições realizadas pela pipeline de segmentação. Os valores apresentados estão da forma $\mu \pm \sigma$ sobre a divisão 5-fold descrita no Capítulo 3.*

Estr.	Arq.	$\mu SD \downarrow$			HD95 ↓		
		4 Ventrículo	Tronco	Cerebelo	4 Ventriculo	Tronco	Cerebelo
$\mathcal{G}$	Unet	1,70 ± 3,24	1,29 ± 0,71	1,02 ± 0,46	7,98 ± 8,93	6,29 ± 3,80	3,82 ± 1,71
	Vnet	1,71 ± 2,17	2,37 ± 4,84	2,52 ± 5,27	12,46±16,34	13,06±18,53	15,54±24,28
	HR3N	1,70 ± 2,93	1,80 ± 1,84	1,59 ± 1,15	10,18±12,05	10,83±9,83	10,83±10,03
$\mathbf{\Phi}(\mathcal{S}_1, \mathcal{G})$	Unet	1,80 ± 2,50	1,28 ± 0,80	1,02 ± 0,46	8,24±7,10	6,41±4,45	3,79±1,71
	Vnet	1,31 ± 1,02	1.30 ± 0,91	2,52 ± 5,28	7,14±4,87	6,24±3,71	15,57±24,24
	HR3N	1,66 ± 1,36	1,46 ± 0,98	1,58 ± 1,15	9,35±5,64	8,45±4,55	10,81±9,93
$\mathbf{\Phi}(\mathcal{G}, \mathcal{S}_2)$	Unet	1,74 ± 3,25	1,27 ± 0,70	0,94 ± 0,58	8,34±8,83	6,47±3,83	3,72±2,29
	Vnet	1,77 ± 2,59	2,35 ± 4,85	1,01 ± 1,07	11,96±16,66	13.04±18,83	4,14±3,64
	HR3N	1,64 ± 2,92	1,74 ± 1,83	1,62 ± 1,46	9,55±11,93	10,38±10,11	9,92±5,74
$\mathbf{\Phi}(\mathcal{S}_1, \mathcal{S}_2)$	Unet	1,80 ± 2,50	1,28 ± 0,80	0,93 ± 0,59	8,24±7,10	6,41±4,45	3,64±2,30
	Vnet	1,31 ± 1,02	1,30 ± 0,91	1,00 ± 1,07	7,14±4,87	6,24±3,71	4,01±3,62
	HR3N	1,66 ± 1,36	1,46 ± 0,98	1,60 ± 1,44	9,35±5,64	8,45±4,55	9,91±5,73
Majority	Todas	**1,11 ± 1,60**	**1,06 ± 0,55**	**0,83 ± 0,47**	**5,82 ± 6,05**	**4,82 ± 2,64**	**2,94 ± 1,74**

A Figura 4.3 ilustra um exemplo de como a partir da predição da rede generalista, ocorre a separação da ROI e então com os resultados obtidos pelas redes especialistas ocorre a fusão por meio da função Φ.

Para uma melhor visualização e comparação dos resultados obtidos, a Figura 4.4 apresenta as diferentes estratégias utilizadas pelas redes generalistas e especialistas. É possível observar no corte $z = 8$ que as redes generalistas Vnet e Unet não foram capazes de segmentar adequadamente o tronco cerebelar, porém este problema foi corrigido com o uso das

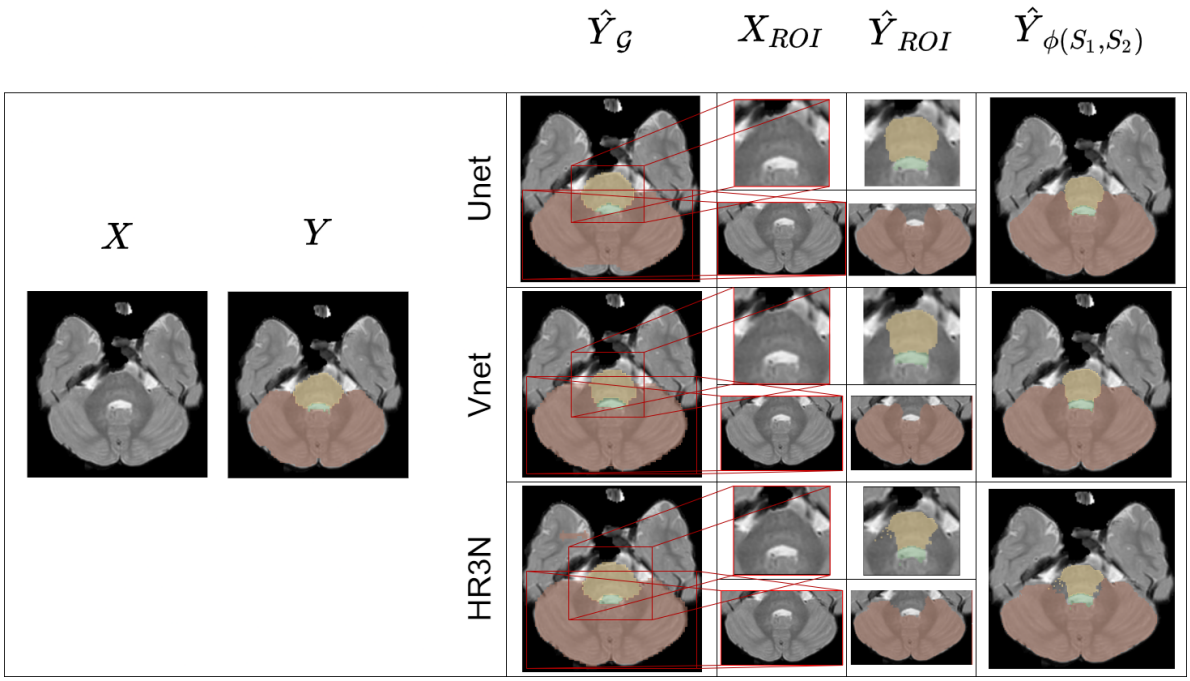


Figura 4.3: *Corte no plano axial para visualização dos resultados obtidos pelas redes, sendo que, X é o volume de entrada, Y o ground truth, $\hat{Y}_{\mathcal{G}}$ resultados obtidos pela rede generalista e $\hat{Y}_{\phi(S_1,S_2)}$ são os resultados obtidos depois da fusão das duas redes especialistas.*

redes especialistas S1. É interessante notar que, para esse mesmo exemplo, a rede H3NR apresentou um desempenho pior quando utilizada em conjunto com alguma rede especialista. Por exemplo, no corte $z = 21$ a rede especialista S2, utilizando a arquitetura H3NR, resultou em alguns buracos na estrutura do cerebelo, e no corte $z = 32$, apresentou também alguns buracos na estrutura do tronco cerebelar. É importante salientar que esse buracos apresentados na segmentação da rede utilizando as redes especialistas podem ser preenchidos durante o pós-processamento.

De forma geral, é possível observar que a utilização das redes especialistas nas arquiteturas Unet, principalmente com a rede especialista S1, e a votação majoritária, proporcionaram resultados semelhantes ao *ground truth* em comparação com as demais estratégias.

A Figura 4.5 apresenta os resultados obtidos pelas votação majoritária em comparação com os resultados da fusão das redes especialistas $\mathbf{\Phi}(\mathcal{S}_1, \mathcal{S}_2)$ em cada uma das arquiteturas (Unet, Vnet, HR3N).

Por fim, as figuras 4.6 e 4.7 apresentam os resultados do mesmo volume apresentado pela Figura 4.5 contudo, utilizando os cortes no plano coronal e sagital respectivamente.

Na Figura 4.6, é possível observar que, visualmente, todos os métodos tiveram bom desempenho na segmentação das três estruturas. No entanto, é perceptível na rede HR3N, nos cortes $x = -251$ e $x = -311$, erros mais evidentes na segmentação do tronco cerebelar e do cerebelo, respectivamente. Situação semelhante ocorreu durante a inspeção visual na Figura 4.7, nos cortes $y = -91$, $y = -103$ e $y = -114$.

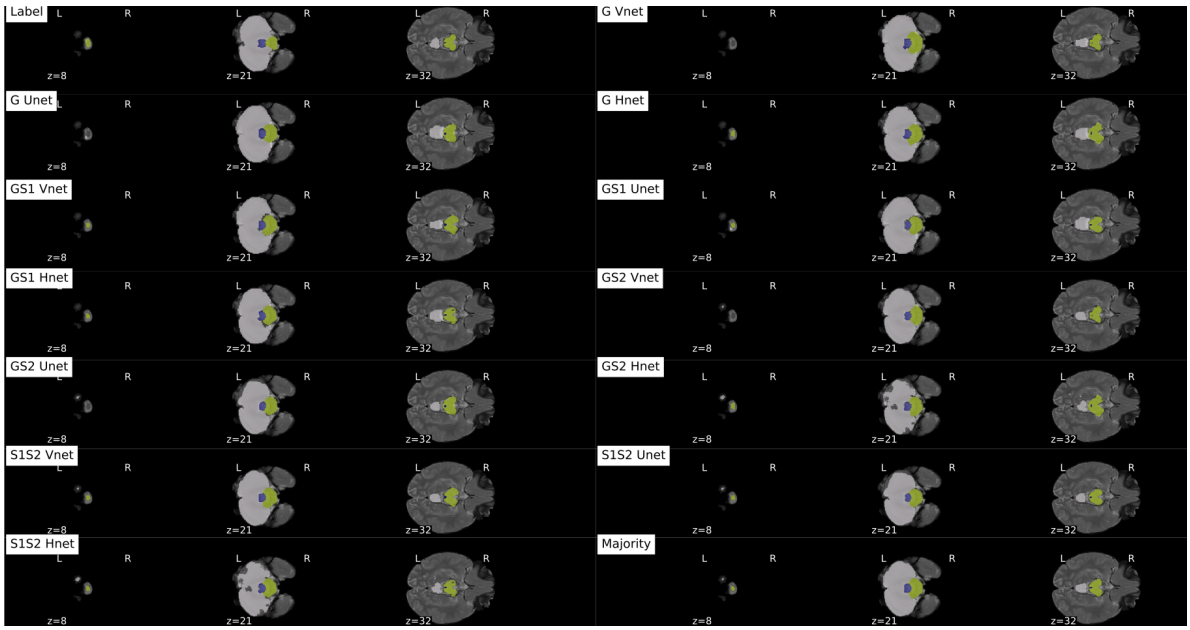


Figura 4.4: *Corte no plano axial para visualização dos resultados obtidos pelas redes, sendo que, Label é o ground truth, G resultados obtidos pela rede generalista, GSX são os resultados obtidos depois da fusão da rede generalista com a rede especialista X e S1S2 são os resultados obtidos após a fusão das duas redes especialistas.*

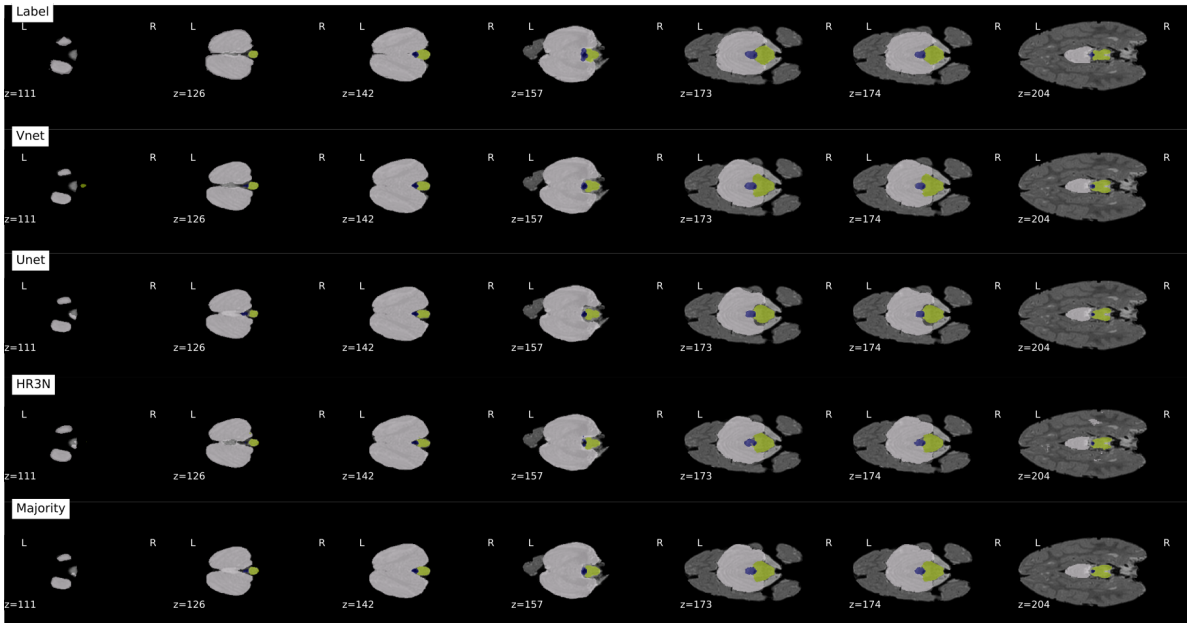


Figura 4.5: *Corte no plano axial para visualização dos resultados obtidos pelas redes especialistas, sendo que, Label é o ground truth, Unet, Vnet e Hnet são os resultados obtidos pela fusão das redes especialistas S1S2 e majority o resultado obtido pela votação majoritária entre G, S1 e S2 de cada uma das arquiteturas.*

4.4 Segmentação Manual

Durante o desenvolvimento desta pesquisa, o autor adquiriu uma habilidade adicional que pode contribuir para avançar os resultados obtidos - a capacidade de realizar segmentações manuais nas partes da fossa craniana posterior mencionadas anteriormente. Com a ajuda

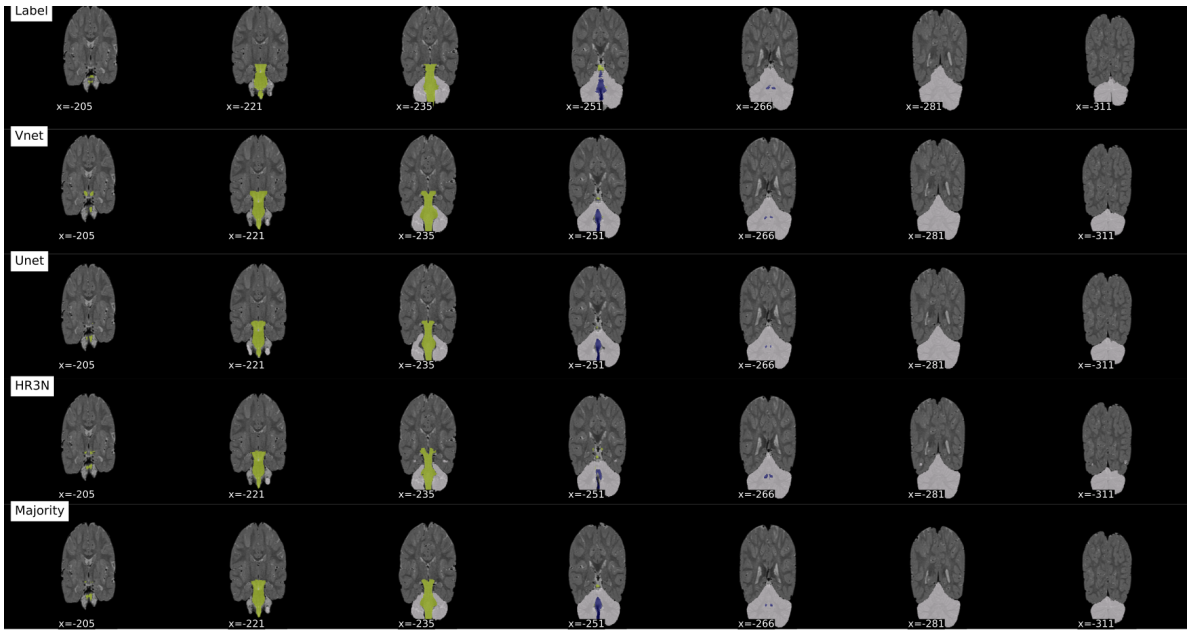


Figura 4.6: *Corte no plano coronal para visualização dos resultados obtidos pelas redes especialistas, sendo que, Label é o ground truth, Unet, Vnet e Hnet são os resultados obtidos pela fusão das redes especialistas S1S2 e majority o resultado obtido pela votação majoritária entre G, S1 e S2 de cada uma das arquiteturas.*

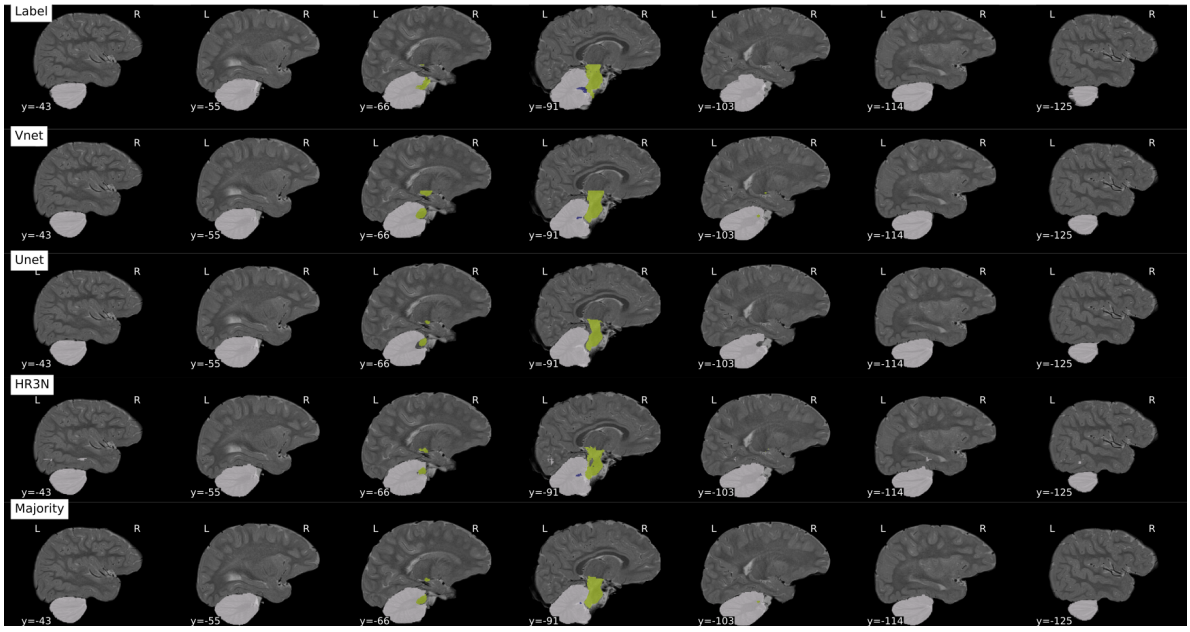


Figura 4.7: *Corte no plano sagital para visualização dos resultados obtidos pelas redes especialistas, sendo que, Label é o ground truth, Unet, Vnet e Hnet são os resultados obtidos pela fusão das redes especialistas S1S2 e majority o resultado obtido pela votação majoritária entre G, S1 e S2 de cada uma das arquiteturas.*

e supervisão de dois radiologistas do HC, foram realizadas várias anotações nos volumes obtidos. A importância da formação desse conjunto de dados e da sua anotação foi discutida anteriormente na Seção 3.2. As anotações foram realizadas utilizando o software Slicer-3D[1], e algumas delas são ilustradas na Figura 4.8.

[1] <https://www.google.com/search?client=firefox-b-d&q=Slice+3d>

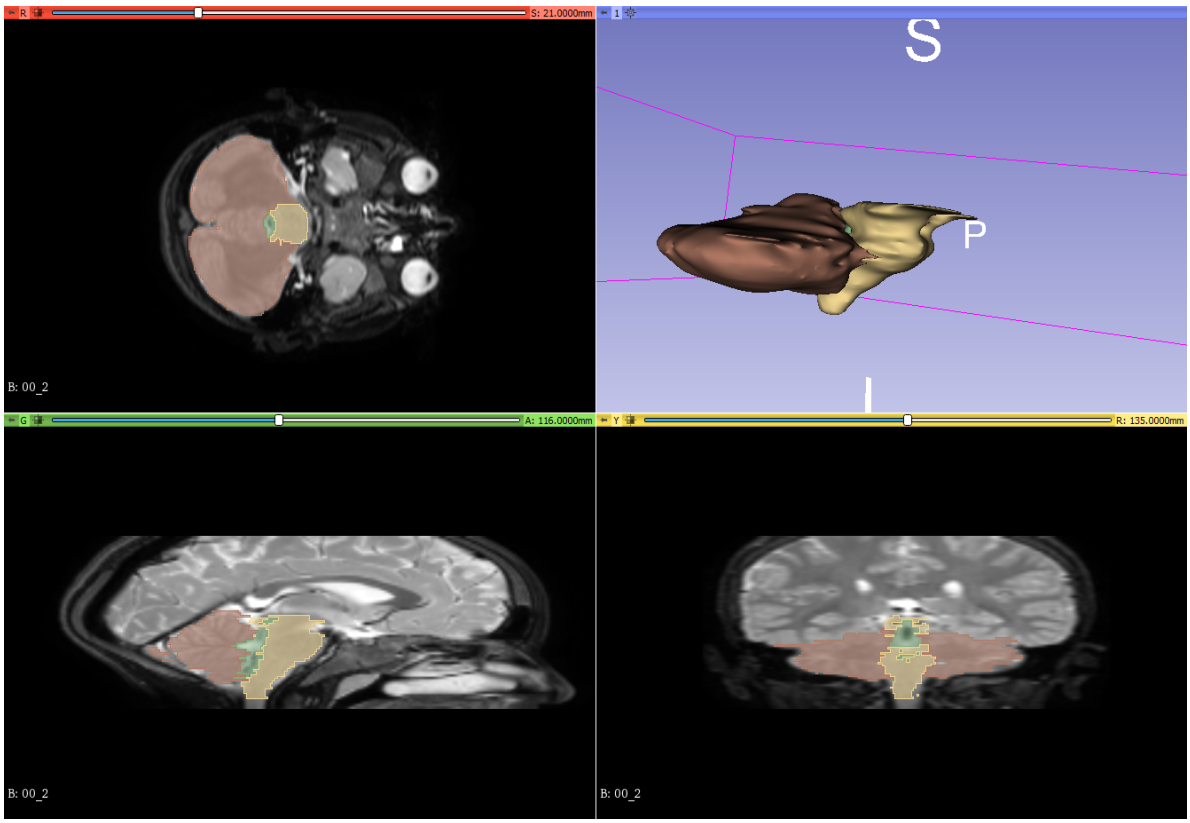


Figura 4.8: *Slicer 3D foi a ferramenta utilizada para a avaliação qualitativa e para segmentação manual das imagens volumétricas. Essa imagem ilustra uma forma de visualização na qual, é possível visualizar o corte axial, o corte coronal, o corte sagital e a forma tridimensional das estruturas segmentadas.*

A ferramenta Slicer-3D também foi utilizada para inspeção visual dos resultados. Com essa ferramenta, é possível visualizar a segmentação em diferentes orientações e inspecionar sua forma em 3D. Essa forma de representação da anotação auxilia os radiologistas na validação dos resultados.

4.4.1 Discussão dos Resultados

Os resultados obtidos pelas redes generalistas demonstram que as três arquiteturas são capazes de segmentar as estruturas posteriores da fossa craniana com um desempenho razoável, obtendo valores médios de DSC e IoU acima de 0,78 e 0,67, respectivamente. No entanto, tanto o Vnet quanto o Unet apresentaram um desempenho ligeiramente melhor do que o HR3N.

Além disso, ao avaliar as métricas de distância, pode-se observar que a Unet obteve os melhores resultados médios para as três estruturas avaliadas, com distâncias médias e desvios padrão inferiores aos de outros projetos arquiteturais. No entanto, as diferenças nas arquiteturas não são muito significativas, indicando que todas são capazes de segmentar as estruturas-alvo com um desempenho razoável.

Apesar dos resultados indicarem um desempenho razoável das redes generalistas, a in-

clusão das redes especialistas na pipeline de segmentação apresentou melhorias tanto nos valores médios DSC e IoU quanto nas métricas de distância média e HD95. Isso demonstra que a utilização das redes especialistas aprimora a segmentação das estruturas propostas.

Por fim, a votação majoritária se mostrou vantajosa, principalmente ao verificar as métricas de distância utilizadas com o destaque para o HD95. Isso mostra que ela tem uma menor discrepância em relação a segmentação de referência. Portanto, a votação majoritária pode ser considerada vantajosa quando se busca uma maior precisão na segmentação de imagens.

Capítulo 5

Conclusão

5.1 Comentários Finais

Esta dissertação propôs um novo método de segmentação automática da fossa posterior pediátrica baseado nos conceitos de rede generalista e rede especialista. A rede generalista é responsável pela segmentação inicial usando o volume completo. A rede especialista, composta por duas redes distintas em que cada uma utiliza uma parte da segmentação anterior, realiza uma segmentação mais específica no local.

O pipeline proposto inclui um pré-processamento das imagens volumétricas, sendo esse composta por três etapas: primeiro é utilizada uma ferramenta para a extração do objeto de interesse (i.e. o encéfalo) (BET). Depois, é utilizada a normalização da intensidade dos *pixels*. Por fim, é realizada uma correção do sinal de campo de polarização (BFC). Seguindo o pipeline de segmentação, a segunda etapa é a segmentação das áreas de interesse (i.e. cerebelo, IV ventrículo e tronco cerebelar) pela rede generalista e pelas redes especialistas. Diferentes arquiteturas foram aplicadas e avaliadas para essa etapa. Por fim, o último procedimento desse pipeline é a realização de uma fusão entre as duas redes especialistas utilizando um algoritmo de *late fusion*.

Foram apresentados resultados experimentais usando imagens MRI de ponderação T2 de crianças entre 0 e 18 anos, adquiridas em exames clínicos realizados com o Hospital das Clínicas da USP. Um total de 32 imagens foram segmentadas manualmente por um grupo de especialistas para geração do *ground-truth*. Três anotações foram feitas na fossa posterior, delimitando assim, as áreas do cerebelo, do IV ventrículo e do tronco cerebelar. Essas segmentações manuais foram utilizadas para treinar e validar as redes neurais generalistas e especialistas. O método proposto alcançou um valor médio de 0,857 no coeficiente Dice. Além disso, as distâncias médias entre as superfícies segmentadas, de maneira automática e manual, permaneceram em torno de 1 mm para as três estruturas. Uma versão preliminar do método foi publicado em (OLIVEIRA et al., 2021a).

5.2 Trabalhos Futuros

A anotação do conjunto de dados da colaboração foi estendida recentemente para incluir patologias em um conjunto de dados maior (Figure 5.1). 118 imagens anotadas por 4 residentes do HC-USP. A anotação agora inclui rótulos para normal (69 casos) / patologias (49 casos) e para tumores (12 casos). Assim, um trabalho futuro importante é avaliar o método proposto nesse dataset estendido, bem como avaliá-lo em situações que aprensentem patologias como tumores. Um outro aspecto interessante é adaptá-lo para incluir outras classes de segmentação, como os próprios tumores.

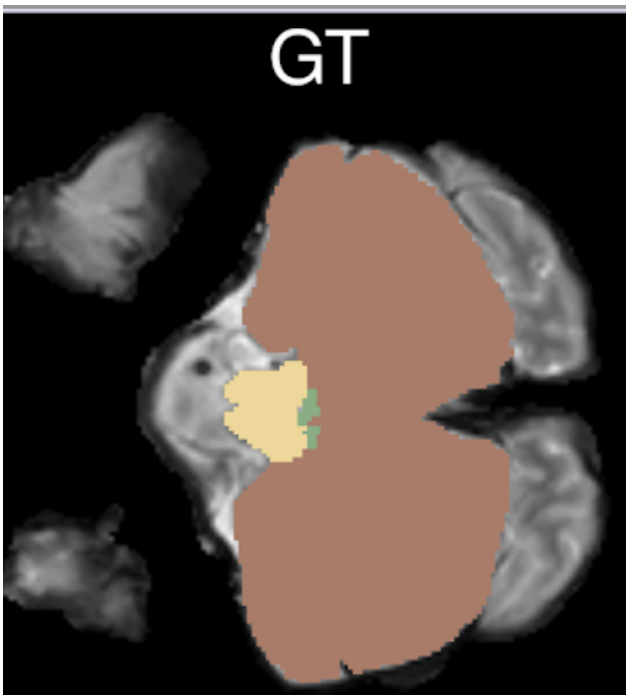


Figura 5.1: *O conjunto de dados foi expandido para 118 imagens anotadas para os segmentos da fossa posterior. A anotação agora inclui rótulos para normais/patologias e para tumores.*

Outra aplicação importante que pode ser explorada futuramente diz respeito à extração de medidas de forma (*shape features*) das estruturas segmentadas. Por último, mencionamos a incorporação e avaliação de métodos para tratar conjuntos de dados fracamente anotados (*weakly supervised*) como *meta-learning* (OLIVEIRA et al., 2022).

Referências Bibliográficas

A. Şeker; A.L., R. The anatomy of the posterior cranial fossa. In: M. Özek et al. (Ed.). *Posterior Fossa Tumors in Children*. [S.l.]: Springer, Cham, 2015. 7

AGGARWAL, C. C. et al. Neural networks and deep learning. *Springer*, Springer, v. 10, p. 978–3, 2018. 12, 13, 14, 16, 17

ALBAWI, S.; MOHAMMED, T. A.; AL-ZAWI, S. Understanding of a convolutional neural network. In: *2017 International Conference on Engineering and Technology (ICET)*. [S.l.: s.n.], 2017. p. 1–6. 17, 18

ALJAAFARI, N. *Ichthyoplankton Classification Tool using Generative Adversarial Networks and Transfer Learning*. Tese (Doutorado), 02 2018. 21

BADRINARAYANAN, V.; KENDALL, A.; CIPOLLA, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 39, n. 12, p. 2481–2495, 2017. 28

BAYDIN, A. G. et al. Automatic differentiation in machine learning: a survey. *Journal of machine learning research*, Journal of Machine Learning Research, v. 18, 2018. 42

BISHOP, C. M. et al. *Neural networks for pattern recognition*. [S.l.]: Oxford university press, 1995. 13

BITTEL, S. et al. Pixel-wise segmentation of street with neural networks. 11 2015. 18

BORGES, L. E. *Python para desenvolvedores: aborda Python 3.3*. [S.l.]: Novatec Editora, 2014. 42

BUI, T. D.; SHIN, J.; MOON, T. 3d densely convolutional networks for volumetric segmentation. *arXiv preprint arXiv:1709.03199*, 2017. 3

CARDOSO, M. J. et al. Adapt: an adaptive preterm segmentation algorithm for neonatal brain mri. *NeuroImage*, Elsevier, v. 65, p. 97–108, 2013. 2

CATALA, M. Development of the posterior fossa structures. In: *Posterior fossa tumors in children*. [S.l.]: Springer, 2015. p. 61–73. 1

CHAITANYA, K. et al. Semi-supervised and task-driven data augmentation. In: SPRINGER. *International conference on information processing in medical imaging*. [S.l.], 2019. p. 29–41. 32

CHEN, J.-F. et al. Financial time-series data analysis using deep convolutional neural networks. In: IEEE. *2016 7th International conference on cloud computing and big data (CCBD)*. [S.l.], 2016. p. 87–92. 18

CHEN, S.; MA, K.; ZHENG, Y. Med3d: Transfer learning for 3d medical image analysis. *arXiv preprint arXiv:1904.00625*, 2019. 3, 25, 26

DERKOWSKI, W.; KEDZIA, A.; GLONEK, M. Clinical anatomy of the human anterior cranial fossa during the prenatal period. *Folia Morphologica*, v. 62, n. 3, p. 271–273, 2003. ISSN 1644-3284. Disponível em: <https://journals.viamedica.pl/folia_morphologica/article/view/16369>. 7

DESPOTOVIĆ, I.; GOOSSENS, B.; PHILIPS, W. Mri segmentation of the human brain: challenges, methods, and applications. *Computational and Mathematical Methods in Medicine*, Hindawi, v. 2015, 2015. 2

DJALILIAN, D. H. R. et al. A study of middle cranial fossa anatomy and anatomic variations. *Ear, Nose & Throat Journal*, v. 86, n. 8, p. 474–481, 2007. PMID: 17915670. Disponível em: <https://doi.org/10.1177/014556130708600813>. 7

DONGARE, A. et al. Introduction to artificial neural network. *International Journal of Engineering and Innovative Technology (IJEIT)*, Citeseer, v. 2, n. 1, p. 189–194, 2012. 11

FAN, Y. et al. Video-based emotion recognition using cnn-rnn and c3d hybrid networks. In: *Proceedings of the 18th ACM international conference on multimodal interaction*. [S.l.: s.n.], 2016. p. 445–450. 17

FANG, W. et al. A method for improving cnn-based image recognition using dcgan. *Computers, Materials and Continua*, v. 57, n. 1, p. 167–178, 2018. 17

FERGUSON, M. et al. Automatic localization of casting defects with convolutional neural networks. In: . [S.l.: s.n.], 2017. p. 1726–1735. 23

GARCIA-GARCIA, A. et al. A review on deep learning techniques applied to semantic segmentation. *CoRR*, abs/1704.06857, 2017. Disponível em: <http://arxiv.org/abs/1704.06857>. 26

GARDNER, M.; DORLING, S. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric Environment*, v. 32, n. 14, p. 2627–2636, 1998. ISSN 1352-2310. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1352231097004470>. 11, 13, 16

GLOROT, X.; BORDES, A.; BENGIO, Y. Deep sparse rectifier neural networks. In: JMLR WORKSHOP AND CONFERENCE PROCEEDINGS. *Proceedings of the fourteenth international conference on artificial intelligence and statistics*. [S.l.], 2011. p. 315–323. 21

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: MIT Press, 2016. <http://www.deeplearningbook.org>. 12, 13, 18, 19, 20

GOUSIAS, I. S. et al. Magnetic resonance imaging of the newborn brain: manual segmentation of labelled atlases in term-born and preterm infants. *Neuroimage*, Elsevier, v. 62, n. 3, p. 1499–1509, 2012. 2

GUI, L. et al. Morphology-driven automatic segmentation of mr images of the neonatal brain. *Medical Image Analysis*, Elsevier, v. 16, n. 8, p. 1565–1579, 2012. 1, 2

GUPTA, N. et al. Artificial neural network. *Network and Complex Systems*, v. 3, n. 1, p. 24–28, 2013. 11

HE, K. et al. Deep residual learning for image recognition. In: *CVPR*. [S.l.: s.n.], 2016. p. 770–778. 24, 31

HORNIK, K.; STINCHCOMBE, M.; WHITE, H. Multilayer feedforward networks are universal approximators. *Neural networks*, Elsevier, v. 2, n. 5, p. 359–366, 1989. 16

JÚNIOR, E. A.; YAMASHITA, H. Aspectos básicos de tomografia computadorizada e ressonância magnética. *Brazilian Journal of Psychiatry*, SciELO Brasil, v. 23, p. 2–3, 2001. 8, 9, 10

KHVOSTIKOV, A. et al. 3d cnn-based classification using smri and md-dti images for alzheimer disease studies. 01 2018. 22

KINGMA, D. P.; BA, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 41

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. *NIPS*, v. 25, p. 1097–1105, 2012. 20, 22, 27

LECUN, Y. et al. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, v. 1, n. 4, p. 541–551, 1989. 17

LECUN, Y. et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, IEEE, v. 86, n. 11, p. 2278–2324, 1998. 18

LI, W. et al. On the compactness, efficiency, and representation of 3d convolutional networks: brain parcellation as a pretext task. In: SPRINGER. *International conference on information processing in medical imaging*. [S.l.], 2017. p. 348–360. 3, 27, 31

LONG, J.; SHELHAMER, E.; DARRELL, T. Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2015. p. 3431–3440. 27, 28

LORENZO, P. R. et al. Segmenting brain tumors from flair mri using fully convolutional neural networks. *Computer methods and programs in biomedicine*, Elsevier, v. 176, p. 135–148, 2019. 32

MAZZOLA, A. A. Ressonância magnética: princípios de formação da imagem e aplicações em imagem funcional. *Revista brasileira de física médica*, v. 3, n. 1, p. 117–129, 2009. 9

MENZE, B. H. et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE TMI*, IEEE, v. 34, n. 10, p. 1993–2024, 2014. 25, 36

MILLETARI, F.; NAVAB, N.; AHMADI, S.-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: IEEE. *3D Vision (3DV), 2016 Fourth International Conference on*. [S.l.], 2016. p. 565–571. 3, 27, 29

MOESKOPS, P. et al. Automatic segmentation of mr brain images of preterm infants using supervised classification. *NeuroImage*, Elsevier, v. 118, p. 628–641, 2015. 2

MOHAMED, I. S. *Detection and Tracking of Pallets using a Laser Rangefinder and Machine Learning Techniques*. Tese (Doutorado), 09 2017. 20

MOK, T. C.; CHUNG, A. Learning data augmentation for brain tumor segmentation with coarse-to-fine generative adversarial networks. In: SPRINGER. *International MICCAI Brainlesion Workshop*. [S.l.], 2018. p. 70–80. 32, 34

MOREL, B. et al. Neonatal brain mri: how reliable is the radiologist's eye? *Neuroradiology*, Springer, v. 58, n. 2, p. 189–193, 2016. 2

NALEPA, J.; MARCINKIEWICZ, M.; KAWULOK, M. Data augmentation for brain-tumor segmentation: a review. *Frontiers in computational neuroscience*, Frontiers Media SA, v. 13, p. 83, 2019. 31, 32

OLIVEIRA, H. et al. Domain generalization in medical image segmentation via meta-learners. In: IEEE. *2022 35th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. [S.l.], 2022. v. 1, p. 288–293. 54

OLIVEIRA, H. et al. Automatic segmentation of posterior fossa structures in pediatric brain mris. In: IEEE. *2021 34th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. [S.l.], 2021. p. 121–128. 4, 53

OLIVEIRA, H. et al. Fully convolutional open set segmentation. *Machine Learning*, Springer, p. 1–52, 2021. 17, 22, 28

OLIVEIRA, H. N. *Semantic Segmentation with Multi-Source Domain Adaptation for Radiological Images*. Tese (Doutorado), 07 2020. 28

PEREIRA, S. et al. Brain tumor segmentation using convolutional neural networks in mri images. *IEEE transactions on medical imaging*, IEEE, v. 35, n. 5, p. 1240–1251, 2016. 32

QAMAR, S. et al. A variant form of 3d-unet for infant brain segmentation. *Future Generation Computer Systems*, Elsevier, v. 108, p. 613–623, 2020. 29

RASCHKA, S.; MIRJALILI, V. *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2*. [S.l.]: Packt Publishing Ltd, 2019. 12, 14

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: SPRINGER. *International Conference on Medical image computing and computer-assisted intervention*. [S.l.], 2015. p. 234–241. 18, 27, 29

ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, American Psychological Association, v. 65, n. 6, p. 386, 1958. 13

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. *Learning internal representations by error propagation*. [S.l.], 1985. 13

SERRA, J. Image analysis and mathematical morphol-ogy. Academic press, 1982. 26

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 22, 23

SINGH, V. *Textbook of Anatomy Head, Neck, and Brain; Volume III*. [S.l.]: Elsevier Health Sciences, 2014. v. 3. 7

SOARES, P.; SILVA, J. da. Aplicação de redes neurais artificiais em conjunto com o método vetorial da propagação de feixes na análise de um acoplador direcional baseado em fibra Ótica. *Revista Brasileira de Computação Aplicada*, v. 3, 12 2011. 12

SPRAWLS, P. *Magnetic resonance imaging: principles, methods, and techniques*. [S.l.]: Medical Physics Publishing Madison, WI, 2000. 1

SRIVASTAVA, R. K.; GREFF, K.; SCHMIDHUBER, J. Training very deep networks. *Advances in neural information processing systems*, v. 28, 2015. 24

STILES, J.; JERNIGAN, T. L. The basics of brain development. *Neuropsychology review*, Springer, v. 20, n. 4, p. 327–348, 2010. 1

SUN, Y. et al. Multi-site infant brain segmentation algorithms: The iseg-2019 challenge. *arXiv preprint arXiv:2007.02096*, 2020. 36

SZEGEDY, C. et al. Going deeper with convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2015. p. 1–9. 22, 23

SZELISKI, R. *Computer vision: algorithms and applications*. [S.l.]: Springer Science & Business Media, 2010. 26

TADEUSIEWICZ, R.; OGIELA, M.; SZCZEPANIAK, P. Notes on a linguistic description as the basis for automatic image understanding. *International Journal of Applied Mathematics and Computer Science*, Sciendo, v. 19, n. 1, p. 143–150, 2009. 2

TAN, Y. Chapter 11 - applications. In: TAN, Y. (Ed.). *Gpu-Based Parallel Implementation of Swarm Intelligence Algorithms*. Morgan Kaufmann, 2016. p. 167–177. ISBN 978-0-12-809362-7. Disponível em: <https://www.sciencedirect.com/science/article/pii/B978012809362750011X>. 26

THOMPSON, D. K. et al. Characterization of the corpus callosum in very preterm and full-term infants utilizing mri. *Neuroimage*, Elsevier, v. 55, n. 2, p. 479–490, 2011. 2

WALT, S. van der et al. scikit-image: image processing in Python. *PeerJ*, v. 2, p. e453, 6 2014. ISSN 2167-8359. Disponível em: <https://doi.org/10.7717/peerj.453>. 42

WANG, G. et al. Interactive medical image segmentation using deep learning with image-specific fine tuning. *IEEE transactions on medical imaging*, IEEE, v. 37, n. 7, p. 1562–1573, 2018. 27

WANG, L. et al. Links: Learning-based multi-source integration framework for segmentation of infant brain images. *NeuroImage*, Elsevier, v. 108, p. 160–172, 2015. 2

WANG, S.-C. Artificial neural network. In: *Interdisciplinary computing in java programming*. [S.l.]: Springer, 2003. p. 81–100. 11, 12, 13

YE, J. C.; SUNG, W. K. Understanding geometry of encoder-decoder cnns. In: PMLR. *International Conference on Machine Learning*. [S.l.], 2019. p. 7064–7073. 28

ZEILER, M. D. et al. Deconvolutional networks. In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2010. p. 2528–2535. 28

ZHANG, Y. et al. A novel medical image segmentation method using dynamic programming. In: IEEE. *International Conference on Medical Information Visualisation-BioMedical Visualisation (MediVis 2007)*. [S.l.], 2007. p. 69–74. 1

ZHOU, Y.-T. et al. Image restoration using a neural network. *IEEE transactions on acoustics, speech, and signal processing*, IEEE, v. 36, n. 7, p. 1141–1151, 1988. 20

ZUCKER, S. W. Region growing: Childhood and adolescence. *Computer Graphics and Image Processing*, v. 5, n. 3, p. 382–399, 1976. ISSN 0146-664X. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0146664X76800147>. 26